Optimal Shot Sequences by Court Position and Player Types

Preprint Submission

By Shane Liyanage (19956089)

Submitted as partial fulfilment of the Master of Sports Analytics

Supervisors: Dr David Carey and Dr Minh Huynh

Industry Primary Partner: Tennis Australia

Subsidiary Partner: Data Driven Sports Analytics

La Trobe University

Sport and Exercise Science

School of Allied Health, Human Services and Sport

Word Count: 8696 (excluding SOA, Acknowledgments, Terminology and Appendix)

Submitted Date: 7/11/2021

Example Citation:

Liyanage, S. (2022). Optimal Shot Sequences by Court Position and Player Types, SportRxiv

Contents

Statement of Authorship

This is to certify that to the best of my knowledge; the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes. I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources has been appropriately acknowledged.

Shane Liyanage

Shane Liyanage

Dated: 7 / 11 / 2021

Acknowledgments

I want to thank my supervising professor’s Dr David Carey and Dr Minh Huynh for providing guidance and direction for the duration of my project.

I would also like to acknowledge the support of the primary Industry Partner – Tennis Australia, in particular Dr Machar Reid, Olivia Cant, Nicole Pratt and Marc Sophoulis for the access to data and systems as well as providing domain expert knowledge and decisional support on key aspects of the project.

Further, the contributions and perspective of tour coaches Anton Dubrov, Issam Jellali, Tiago Leivas and Federico Ricci, and the players they are working with were greatly appreciated.

Data Driven Sports Analytics also provided access to various software licences and GPU infrastructure which was important to the delivery of the overall project.

Finally, my wife played a tremendous role in providing me the time and space to focus on this project.

Shane Liyanage

7 November 2021

Abstract

The thesis aimed to find optimal shot sequences using sequential market basket analysis from different court positions using domain expert agreed court tessellations (Court Segments) and for different player types determined by using unsupervised learning (clustering) (Figure 1).

Accurately determining this information can provide benefits to coaches and tennis academies to:

- better setup training for specific players
- be used alongside talent identification models to help nurture players unique game styles from early in their development
- scout opponents and strategise game plans

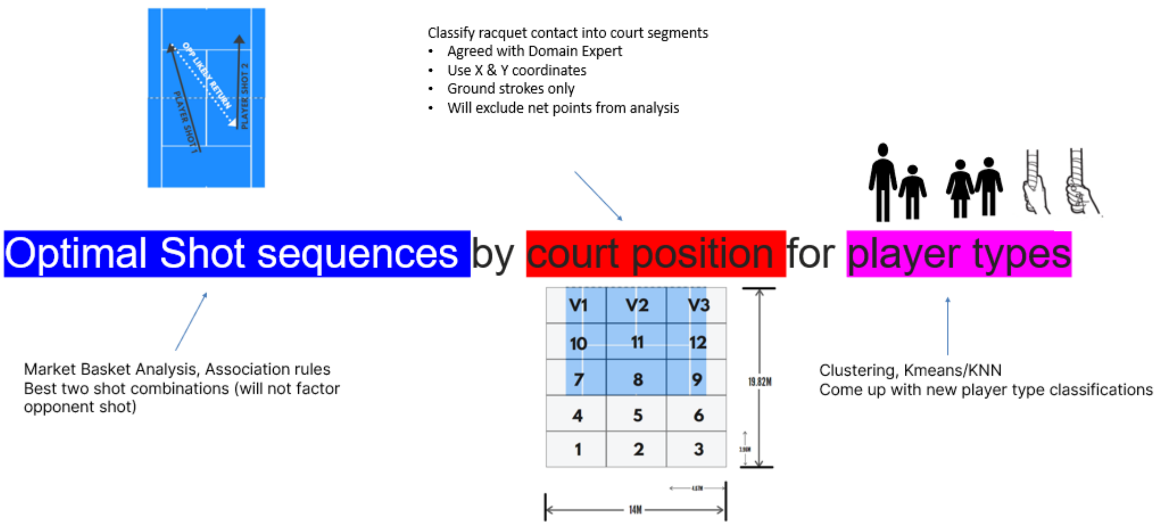


Figure 1: High Level project components – 1. Optimal Shot Sequences 2. Court Position 3. Player Types

The K-Prototypes clustering algorithm was used to infer player types. The algorithm was selected over other algorithms by a majority of domain experts. The cluster group sizes were determined using the knee/elbow method and the cluster cohesion and separation was assessed using silhouette coefficient testing.

Sequential Market Basket Analysis was applied to spatio-temporal data with the player clusters to determine shot sequences, which were then pruned using multi-metric thresholds including support and lift. Shot Sequences were evaluated using a derived weighted per point outcome and compared to various intuitive baselines.

Terminology

Shot Sequence, Shot Chain & Shot Combination: In the tennis community shot sequences is used interchangeably with shot chains and shot combinations, for the purposes of this thesis we will refer to this as shot sequence. The shot sequence definition encompasses the ground stroke shot being used (e.g. forehand or backhand) as well as the direction type (e.g. cross court, inside-out etc). An example of shot sequence could look like: Backhand-Line → Forehand middle.

Serve: The serve (or service) is the first shot played in each game. The serving player tosses the ball in the air and hits it to land in the service box on the opposite – diagonal – side of the court. On each point, if a player misses their first serve, they may play a second serve.[1]

Groundstroke: The shot played after the ball has hit the ground. Groundstrokes are often played nearer the baseline, but can happen anywhere on the court. [1]

Volley: This shot is when the ball is hit by the player before it hits the ground. Volleys are often played close to the net, but can happen anywhere on the court. [1]

Forehand: A groundstroke or volley played with the palm of the racket hand facing the direction of the shot. Forehand groundstrokes are typically referred to as 'forehands'. [1]

Backhand: A shot played with the back of the racket hand facing forward toward the direction of the shot. Backhand groundstrokes are typically referred to as 'backhands'. [1]

Ad side of the court: The half of the court, where the server hits his serve on odd number scoring. It can be used as a reference to where the player make contact with the ball.

Deuce side of the court: The half of the court, where the server hits his serve on even number scoring. It can be used as a reference to where the player make contact with the ball.

Introduction

The 2019 Wimbledon Final was memorable. Novak Djokovic saved match points winning the final in the event's first super tiebreak. However, could it have been different? Late in the final set, Roger Federer had a match point and chose to serve out wide to the ad-court, following this up with an inside in forehand and advances up the court – He ends up losing the point when Novak Djokovic hits a cross court passing shot. Roger Federer would end up losing the match.

Figure 2: Wimbledon 2019 Final – Match point #2 for Roger Federer – Was his shot combination appropriate?

Many observing the final would question this moment:

- Did Federer make the right shot selection to approach?
- Could he have set up his shot sequences to prevent his opponent a forehand cross court passing opportunity?
- From the start of the match what shot sequences should he have played more often to win the match?

An understanding of shot sequences from court positions for player types would go some way to answering the above.

Importance of shot sequences in tennis

Tennis is an open skill sport with many shot choices and combinations available to a player from different court positions.[2] The ability to evaluate and predict best shot sequences from specific court positions can help coaches, players and academies maximise opportunities. For instance:

- **When scouting opponents and strategizing -** understanding how to retain ascendancy in a rally by hitting a certain combination of shots that maximise the chances of winning a point; or to manoeuvre their opponent to uncomfortable locations can be the difference in a sport with fine margins between winning and losing.[3]
- **Helping develop junior athletes in a tailored manner specific to their player type.** This can fast track a player understanding their strengths and weaknesses from different court locations. This information combined with existing talent identification models in tennis[4] that can predict a junior players anthropometric characteristic when adults, will allow the

coaching team and player to work on shot sequences that will provide them the best benefit as an adult player. It can also help identify changes required early on in relation to their technical aspects – e.g. grip or backhand type to give them the best opportunity to play a certain way.

The problem and the complexity around it

Currently shot sequence information for specific player types, whilst sought after, is not available from peer reviewed research-based sources. As a result, there is an over reliance on anecdote to get this information. A key challenge to addressing the problem can be summarised into three distinct components:

a) Determining different player types.
b) Tessellating the court in line with the contemporary game
c) Finding Shot Sequences that work best for different player types from court locations.

The aim of the project

The research project will use anthropometric and technical player data from the top 100 ranked Association of Tennis Professional (ATP) and Women's Tennis Association (WTA) players to derive player type groupings using the unsupervised learning technique of clustering.

A contemporary tessellation of the tennis court into segments is agreed with domain experts.

The project will then use Hawk-eye system (spatio-temporal) data and apply sequential market basket analysis to determine shot sequences that are more associated and successful for different player types from different court segments.

Literature Review

Aim and Scope of Review

The scope of this literature review will be research papers relating to:

a) Finding Shot Sequence that work best for different player types from court locations.
b) Determining different player types.

The aim of the literature review is to ascertain how much of the required analysis components have already been addressed by existing work, and to determine what gaps remain. For a detailed understanding of the Literature review strategy – see Appendix A.

Results

The literature review results will be set out using a topical order addressing the key sub-problems and additional literature relevant for small components of the research problem.

a) Finding Shot Sequences

In relation to finding shot sequences, theoretical areas of association rule mining and sequential market basket analysis will be explored.

<u>Association Rule Mining and Market Basket Analysis</u>

Association rule mining is a well-established key data mining technique that has been used for knowledge discovering of associations between variables of datasets.[5] Rules in an itemset are determined using *if-then* associations between items *(also known as antecedent-consequent associations)*.[6] For example- *If* Backhand Cross is hit, *then* Backhand Line is hit.

The rule mining can show how frequently certain events occur with other events. The application of association rule mining is found in the *Market Basket Analysis (MBA)* technique.[7] MBA has come to prominence due to retailers using it to uncover associations between items in large transaction datasets. Using a similar approach in the proposed research, MBA will be used to find association between tennis shots – e.g. Forehand Cross and Backhand Line.

Three common metrics are used to evaluate the strength of the association rules – Support, Confidence and Lift highlighted in Equation 1, Figure 3 and Equation 2.[7]

$$Support(\{X\} \rightarrow \{Y\}) = \frac{Transactions\ containing\ both\ X\ and\ Y}{Total\ number\ of\ transactions}$$

Equation 1: Support is the percentage of transactions that contain all the items in a target itemset

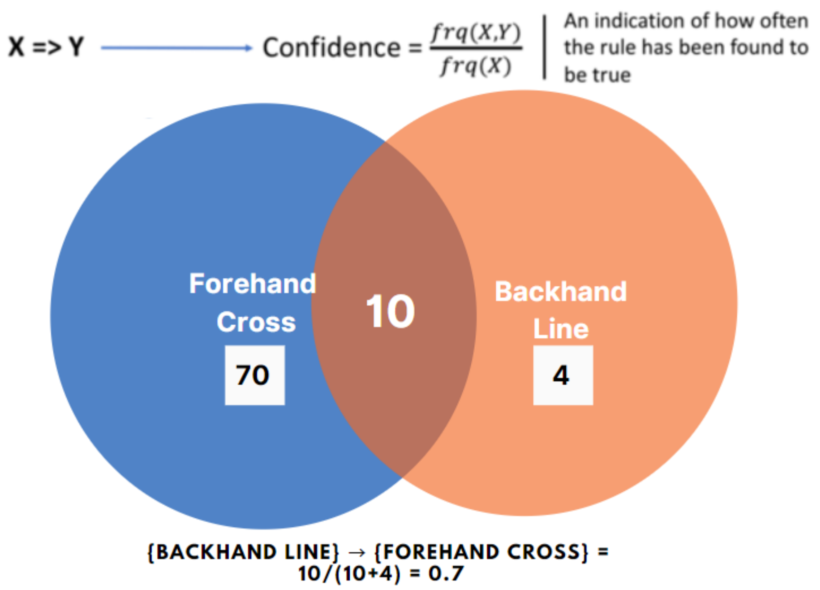


Figure 3: Confidence is the probability that a transaction that contains the items on the antecedent (left hand side) of the rule also contains the item on the consequent (right hand side) of the rule.

$$\text{Lift} = \frac{Support}{Supp(X)\ X\ Supp(Y)}$$

Greater lift values (>1) indicate stronger associations between X and Y and they depend on one another

Equation 2: The probability of all of the items in a rule occurring together.

MBA was applied in sport in *Wenninger et al*[8] in the context of finding volley ball hit patterns that could provide a tactical advantage in elite competition. The research used data collected from 413 men's and 552 women's top-level matches at world tour events which were annotated by professional volleyball analysts. The data quality used in the research met a minimum Cohen's k-statistic threshold to ensure there was substantial agreement between annotators.[8] Annotated rally *transactions* were created based on each unique time step of the rally with all indicators captured (see Figure 4) and item sets rules were created – e.g. Block_Pos_Line => Def_Pos_Line.

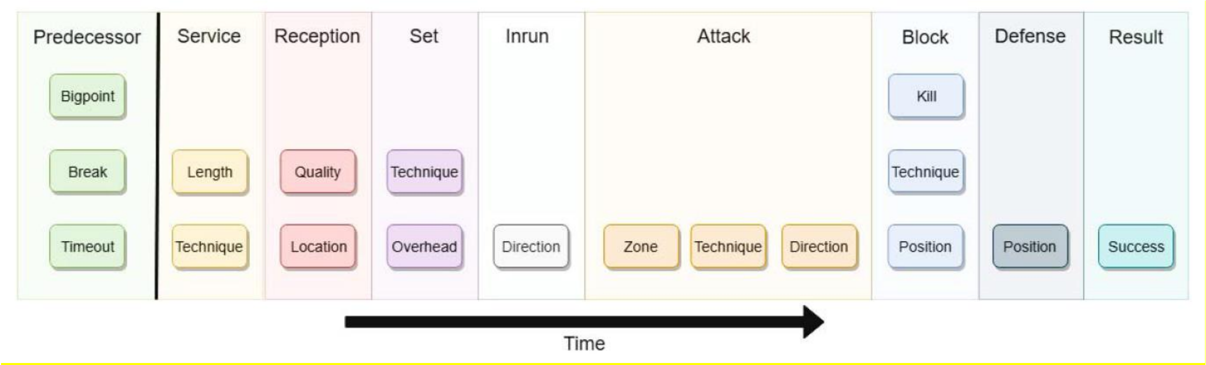


Figure 4: Rally Indicators annotated in Wenninger study

To refine the item sets the study uses a variation of the *apriori* algorithm [7] which specifies a minimum confidence as well as support to prune rules.

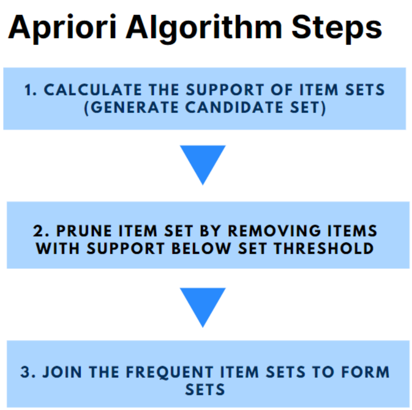


Figure 5: Apriori Algorithm steps

For the *Wenninger et al*[8] study the following thresholds for individual player analysis were set:

- a minimum support of 5% and
- minimum confidence of 50%.

The reason for these thresholds was the desire to find interesting rules for individuals but still refine the rules down to a manageable amount to analyse only important ball hit sequences. With more shot types in tennis than volleyball, there will be more permutations of shot sequences, hence setting a slightly lower support and confidence is required.

Sequential Market Basket Analysis (SMBA)

In elite sport often the events sequence is important, something which traditional MBA does not accurately measure.*[9]* The *Kamakura[9]* paper introduces the concept of *Gain* and *Relation* to MBA which helps measure the impact of sequence on the strength of association where additional information like a transaction time variable is also available.*[9]* Gain directly measures how the prior selection of one item (e.g. Forehand Cross) affects the likelihood that another item (e.g. Backhand Cross) is subsequently selected. This is important as Traditional MBA measures only imply joint occurrence of the events.*[9]*

$$\mathrm{GAIN}[A \rightarrow B] = \frac{P[A-B]}{P[B]} - 1$$

Equation 3: Gain calculation in SMBA

The *Kamakura[9]* paper also introduces the metric *Relation* which uses conditional probability to evaluate the order of transactions, and categories the following types of association pairs:

Weak a→b sequence—if $p(A \rightarrow B)p(B)$ and $p(B \rightarrow A) \sim p(A)$

Strong A→B sequence—if $p(A \rightarrow B)p(B)$ and $p(B \rightarrow A)p(A)$

Complements— if $p(A \rightarrow B)p(B)$ and $p(B \rightarrow A)p(A)$

Substitutes—if $p(A \rightarrow B)p(B)$ and $p(B \rightarrow A)p(A)$

Independents—otherwise

Equation 4: Various Relation classifications from Kamakura Paper of SMBA

The application of SMBA will be important to finding shots that are associated together and occur in a specific order.

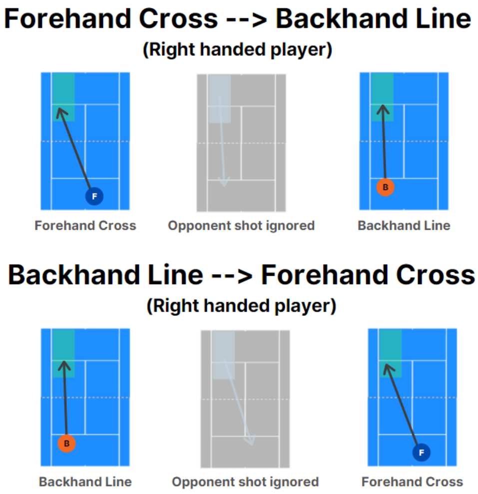


Figure 6: Sequence of shots is important (the rule in the top diagram is different to the bottom)

Same item in Antecedent and Consequent

In *Wenninger et al*[8] by incorporating timestep, an item order variables into an analysis, the rules are able to have the same *item* in the antecedent and consequent – e.g. forehand Line (shot 1)→ Forehand Line (shot 2) – because shot 1 occurs before shot 2 is considered the antecedent.

Application to Tennis data

Weidner et al[10] applied association rule mining with spatio temporal tennis data to find maximal non-redundant association rules in tennis data.

The research uses variables including tennis shot type (e.g. forehand, backhand), ball trajectories , ball bounce location, shot result (e.g. winner, error) and time of shot to create association rules which are pruned using the apriori algorithm.*[10]* Before the association rules are created, ball bounce coordinates are mapped to N x M court tessellation regions in Figure7. This is done because the research correctly identifies the probability of exact coordinates occuring on ball bounces is low, and as a result too many rules would be created wihout mapping.*[10]* While the research left it open to choose different N and M region amounts, it made no justification on why the court tessellation regions used in the final results were selected – e.g. it does not follow existing tennis literature on tennis court regions or have a domain expert citing that the regions used are appropriate.

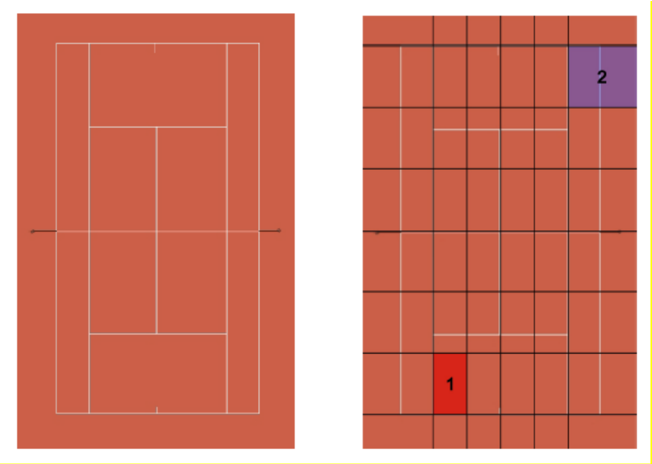


Figure 7: Weidner - Tessellation of tennis court

All the association rules created are then narrowed by removing redundant and non–maximal association rules.

b) Determining Player Types using clustering

Clustering is an unsupervised learning technique that divides the data points into different groups with similar traits and assigns them into clusters. It was used in *Wenninger* as the authors believed the simplicity of the technique allowed it to be communicated better to non-technical audiences.[8]

Clustering of tennis players anthropometric and individual features into groups with different performance was used in *Cui[11]*. Variables used for clustering in the *Cui* research included - match statistics and anthropometric features of 1,188 male players competing during 2015–2017 Grand Slam singles events. Some of these variables included - height, weight, experience, playing hand, and backhand style. This study did not go far enough in relation to variables used to distinguish techniques related to tennis players. *Mehaffey*[12] alludes to the way the racquet is gripped by a player, impacts their swing paths and consequently their game style.[12] Therefore not having this technique related variable may prevent optimal player clusters forming in the *Cui* study. Further, a gap the *Cui* research was that it only used data collected on male athletes. The cluster formed in this research may not be representative of female tennis players.

The *Cui* study used a combination of data provided from official tournaments (including Hawkeye data), and other data captured from the official Association of Tennis Professionals (ATP) website.*[11]*

The study then used a two-step cluster analysis with log-likelihood as the distance measure undertaken to classify players according to height, weight, handedness, backhand style and professional experience. The original data was grouped into pre-clusters by constructing a cluster features tree, then the standard hierarchical clustering algorithm on the pre-clusters was used and provided a range of solutions with different numbers of clusters.*[11]*

Finding the optimal amount of clusters

Cui uses the Schwartz's Bayesian criterion (BIC), the BIC change value, the ratios of BIC changes and ratios of distance measures to find the optimal number of clusters. It purports that the better clusters will have*[11]*:

- A lower BIC value;
- Larger value of BIC change;
- Larger ratio of BIC; and
- Larger ration of distance measures

The study results found the optimal number of clusters was four by comparing fifteen clustering solutions.

Figure 8: Cui – Optimal amount - Four Player clusters

Quality of clustering model and most important variables

Cui[11] evaluates the quality of the clustering model using average silhouette coefficient, which is a measure of both cohesion and separation, with the following mapping of scores:

- −1.0 – 0.2: Poor model;
- 0.2 – 0.5: Moderate to fair model;
- > 0.5 very good model

The study also ranked the most important variables according to predictive importance and found backhand style and handedness were the most important predictors, while professional experience was the weakest predictor (see Figure 9).

Performance characteristics of the player clusters evaluated

Once the player groups were formed, a multivariate analysis of variance (MANOVA) was used to evaluate differences of all 29 performance variables among them.

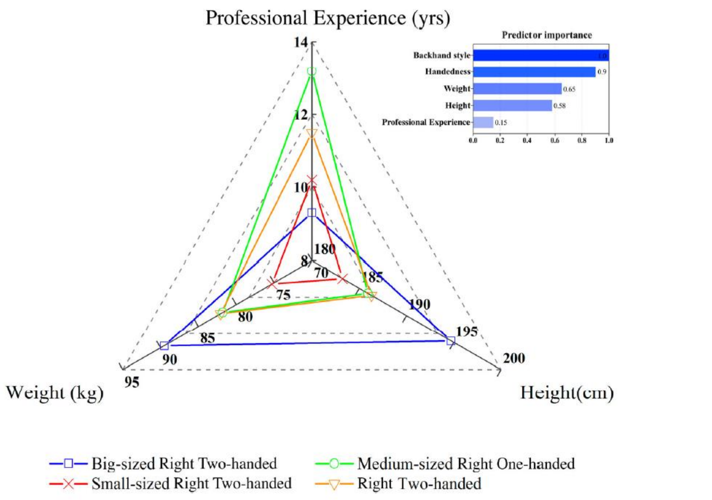


Figure 9: From Cui: Predictor importance of input variables from two-step cluster analysis and description of player groups by cluster analysis.

Conclusion

The reviewed literature can go part of the way in addressing the problem statement.

Finding optimal shot sequences from court locations

The research in *Wenninger[8]* highlighted that association rule mining, in particular MBA can be used in an elite sports setting to provide a strategic advantage.[8]

The SMBA technique in *Kamakura[9]* can be applied with tennis data using time ordered shot data so that the importance of shot sequence order is not missed. The approach in *Weidner* of mapping hawk-eye ball bounce coordinate data into tessellated court regions prior to finding association rules is a good approach to minimise rules and find high quality associations, however a more appropriately tessellated court should be developed with a domain expert before ball coordinates are mapped.

Determine player types

Different player types can be determined by adopting the clustering approach in *Cui[11]* using anthropometric and individual player feature variables while selecting an optimal amount of player clusters using BIC. Different algorithms types can be trialled, not just replicating the hierachicial clustering algorithm in that paper. Additional variables should be used for clustering which could improve the quality of the clusters, namely shot grip variables collected using a similar process to that in *Eng & Hagler[13]*, where 6-8 high quality images are used to annotate the player’s grip (See Appendix 2 for more detail).

The proposed research would aim to address some of the gaps in the problem not resolved by the research above, including:

- Working with domain experts to determine an appropriate amount number of shot combinations to develop association rules for;
- Determine an appropriate court tessellation for contemporary tennis playing styles;
- Develop a tool, like a drill through dashboard to allow non-technical users (tennis coaches) to use and interact with the information from the study at both a macro and micro level.

Methods

The research involves deriving player clusters and then performing sequential market basket analysis (See Figure 10).

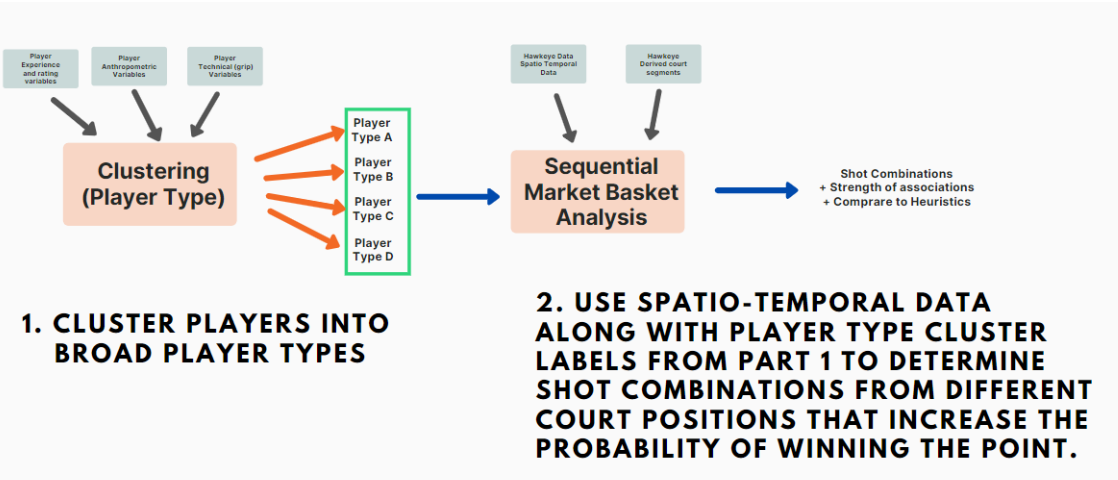


Figure 10: Proposed Research Steps

Data Collection

The following broad types of data will be collected for the project:

- Player Anthropometric
- Technical player grips
- Player positional and ball tracking

Anthropometric Data

Data related to anthropometric characteristics of tennis players was scraped from player tour websites including the WTA, ATP, and ITF websites. At the time of web scraping (04/21/2021) the terms and conditions of the respective organisations and relevant ethical considerations were complied with.

Player Grip Data

Player forehand grip (Figure 11) and backhand grip (Figure 12) was manually coded according to how players held the racquet bevel by viewing players hitting shots in videos and high-definition images. Manual classification was conducted by two separate persons with tennis certifications, including one who has a Tennis Australia High-Performance Coaching qualification (Domain Expert). The inter-rater agreement was 96% and Cohen's Kappa of 0.89 which indicated strong agreement[14].

All variances in classification were within half a grip (e.g. Eastern vs Half Continental-Eastern). Where the two manual coders had different classifications, the high-performance coach classifications took precedence.

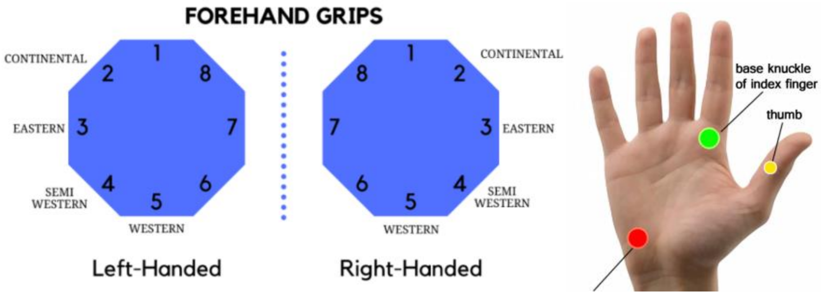


Figure 11: Forehand Grip Diagram – Where base knuckle of index hand should sit

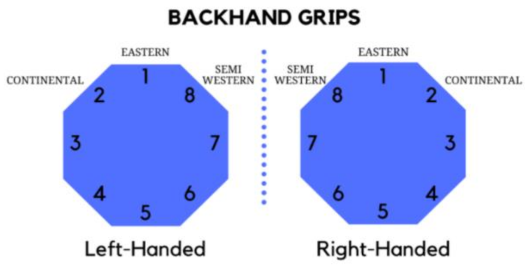


Figure 12: Backhand Grip Diagram – Where the base knuckle of the index hand should sit

An alternative deep learning approach to code grips was developed but ultimately not used (more info in Appendix 3).

Hawkeye player positional and ball tracking data

Structured Hawkeye data which included shot location, ball trajectory, racquet connection point, match score and shot type variables was provided by Tennis Australia. As part of the Extract Transform Load (ETL) process some variable reduction was performed to ensure only useful variables were used (More information in the Data Analysis section)

Domain Expert Reliance

Various components of the project relied on input, guidance and decision making by all or some of the domain experts listed in Table 1.

Domain Expert Type	Coach Certification Authority	Number of persons
High Performance Coach	Tennis Australia	3
Tennis Australia Bio-mechanics Expert	Tennis Australia	1
High Performance	Tennis Canada	2
WTA Coach	ITF	2
WTA Player	N/A	2
ATP Player	N/A	1

Table 1: Domain Experts used for components of the project

Data Analysis

Player Type Clustering

Tennis players have various physical differences due to gender, height and weight [15]; and technical differences due to open and closed grips[16] , backhand types [17] and player handedness[18] which influence how they play the game including shot options from different court positions.

Before determining optimal shot sequences from different court positions, it's important to group players into player types, with a focus on finding groups that play differently from different parts of the court.

Benefits of clustering include:

- Dealing with limited player match samples
- Context for results to coaches
- Useful filter in practical tools like a dashboard

Why cluster with anthropometric and technical data only?

Approaching clustering only using anthropometric and technical data ensures scalability in application to more players, as this data is much more accessible than detailed playing statistics. This approach is taken factoring in the lack of spatio-temporal data available for lower tier professional events outside certain junior grand slams.

Variable Selection

Data Set variables used for clustering

The variables used for the clustering include**:** Height, Weight, Right/Left, Backhand Type, Forehand Grip, Dominant Hand Backhand

A	B	C	D	E	F	G	H	I	J	K	L	M
Tour	Rank	Name	Countr	DOB	Height	Weight	BMI	Age	Right/L	Backhand Type	Forehand Gr	Dominant Hand Backhand
ATP	1	Novak Djokovic	SRB	22/05/1987	188	88	24.89814	34	R	Two-Handed Backhand	SW	C
ATP	2	Daniil Medvedev	RUS	11/02/1996	198	81	20.66116	25	R	Two-Handed Backhand	E	E
ATP	4	Dominic Thiem	AUT	3/09/1993	185	82	23.95909	27	R	One-Handed Backhand	SW	C
ATP	5	Stefanos Tsitsipas	GRE	12/08/1998	193	83	22.28248	22	R	One-Handed Backhand	E	C
ATP	6	Roger Federer	SUI	8/08/1981	185	85	24.83565	39	R	One-Handed Backhand	E	C
ATP	7	Alexander Zverev	GER	20/04/1997	198	86	21.93654	24	R	Two-Handed Backhand	E-SW	C
ATP	8	Andrey Rublev	RUS	20/10/1997	188	68	19.23947	23	R	Two-Handed Backhand	SW	C
ATP	9	Diego Schwartzman	ARG	16/08/1992	170	64	22.14533	28	R	Two-Handed Backhand	W	C
ATP	10	Matteo Berrettini	ITA	12/04/1996	196	95	24.72928	25	R	Two-Handed Backhand	SW	C
ATP	12	Roberto Bautista Agut	ESP	14/04/1988	183	76	22.69402	33	R	Two-Handed Backhand	E	C
ATP	13	David Goffin	BEL	7/12/1990	180	69	21.2963	30	R	Two-Handed Backhand	SW	C
ATP	14	Gael Monfils	FRA	1/09/1986	193	84	22.55094	34	R	Two-Handed Backhand	SW	C
ATP	15	Pablo Carreno Busta	ESP	12/07/1991	188	78	22.06881	30	R	Two-Handed Backhand	SW	C
ATP	16	Grigor Dimitrov	BUL	16/05/1991	191	80	21.92922	30	R	One-Handed Backhand	E-SW	C
ATP	17	Fabio Fognini	ITA	24/05/1987	180	74	22.83951	34	R	Two-Handed Backhand	SW	C
ATP	18	Felix Auger-Aliassime	CAN	8/08/2000	188	84	23.76641	20	R	Two-Handed Backhand	E-SW	C
ATP	19	Milos Raonic	CAN	27/12/1990	196	90	23.42774	30	R	Two-Handed Backhand	SW	C
ATP	20	Cristian Garin	CHI	30/05/1996	185	85	24.83565	25	R	Two-Handed Backhand	SW	E
ATP	21	Stan Wawrinka	SUI	28/03/1985	183	79	23.58984	36	R	One-Handed Backhand	SW	C
ATP	22	Karen Khachanov	RUS	21/05/1996	198	88	22.44669	25	R	Two-Handed Backhand	W	C
ATP	23	Alex De Minaur	AUS	17/02/1999	180	69	21.2963	22	R	Two-Handed Backhand	E	C
ATP	24	Borna Coric	CRO	14/11/1996	185	78	22.79036	24	R	Two-Handed Backhand	SW-W	C
ATP	25	Casper Ruud	NOR	22/12/1998	183	77	22.99262	22	R	Two-Handed Backhand	SW	C

Figure 13: Screen shot of dataset highlight the variables used for Clustering

Clustering Algorithm

Due to the significant difference in psychological, anatomical and biomechanics between male and female players [19], the clustering dataset was split on gender.

Additionally because player handedness differences cause differences in serve[18] and shot options for players[20] with opposing playing hands, the data set was further split on handedness.

Separate clustering processes were run for all these partitioned data sets and compared to clustering where no partitioning was done on the dataset, with the former being preferred by a majority (8/9) of domain experts reviewing the cluster partitions. A union of the separately generated cluster sets was performed to have a single consolidated file - See Figure 14 for process.

Country, age, ranking data removed as relevance to clustering into groups for the purpose of finding differences in shot options from court positions was considered low or detrimental. BMI was not used because it is derived off a calculation using the Height and weight variables.

The clustering process was trialled with and without technical player grip variables (Forehand Grip and Dominant Hand Backhand Grip). The resulting player clusters were evaluated by domain experts with 8 out of 9 selecting player clusters generated with technical player grips being used.

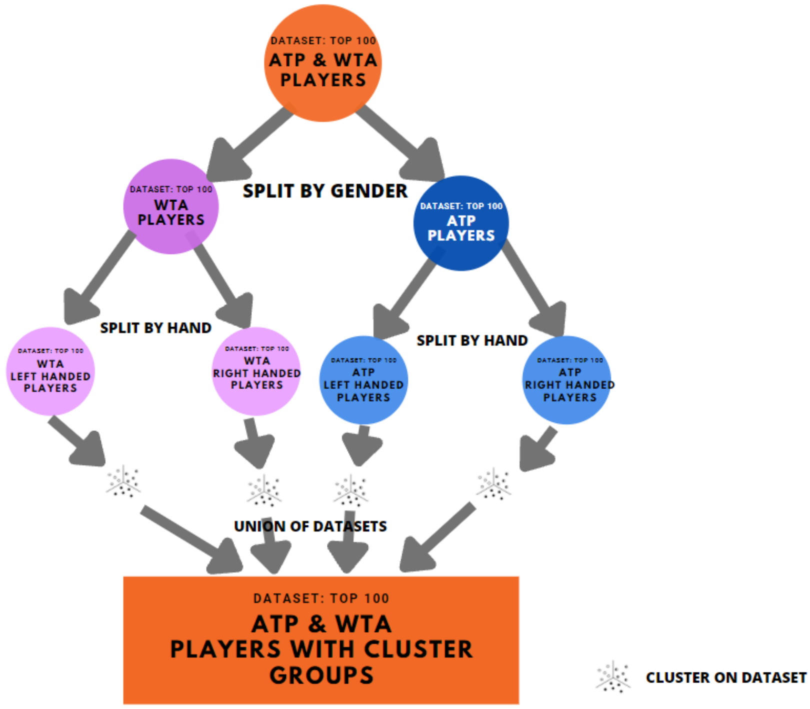


Figure 14: Clustering - Pre clustering dataset partitioning and Post clustering consolidation process

Clustering algorithms
The following clustering algorithms were used to develop various cluster sets:

1. K Means[21] - converting categorical variables into integers
2. K Mode[22] - converting numerical variables into categorical bins (e.g. Height Bins)
3. K Prototypes using
   a. Huang initialisation[23] (frequent categorisation to initial k-modes) – *K Prototypes X*; and
   b. Cao Initialisation [24] (density of the data point and the dissimilarity value categorisation) – *K Prototypes*
4. Agglomerative Hierarchical[25] by converting the categorical variables into integers and using the *dynamic cut method* to draw appropriate points to cut off clusters on dendrogram

Variable data type conversions
For K Means, K Prototypes and Agglomerative Hierarchical clustering methods, the categorical variables of *Forehand Grip* and the *Dominant Hand Grip* were converted into continuous variables as the order of the grips have an inherent meaning, e.g. a continental grip is considered 'more open' and further in distance from a Western grip than an Eastern grip is from a Western grip.

This data type change was also actioned when using the K Prototype algorithm despite it being capable of handling categorical variables, as the algorithm performs distance-based partitioning of continuous variables better than dissimilarity-based partitioning of categorical variables on a data set of the size used in this project.[26]

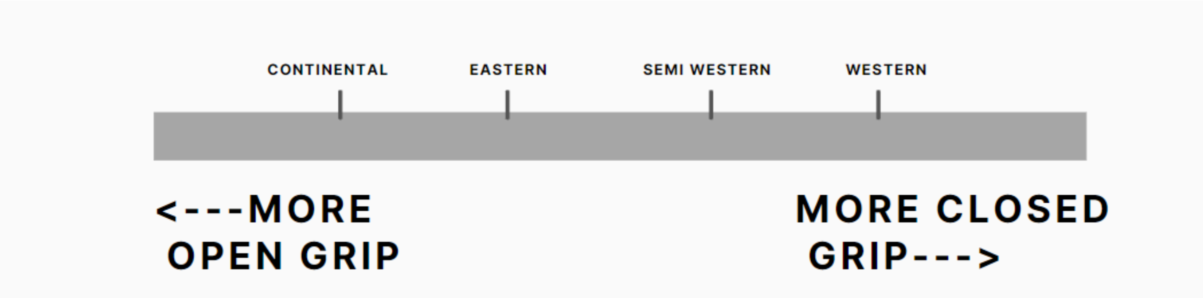


Figure 15: Grip Scale

For K-Means and Agglomerative Hierarchical clustering methods the categorical variable, *Backhand Type* (single handed vs double handed) which does not have an inherent order required the creation of a new derived variable – *Backhand Type Grip* (More info in Appendix 4).

All continuous variables were standardised prior to clustering.[27]

Optimal Cluster Size

Optimal cluster sizes for the above cluster sets were evaluated using

- Knee/Elbow Method[28] by visually comparing the difference of the sum of square error (SSE) for each cluster.
- Silhouette Coefficient (SC) cohesion and separation testings[29] which factors how close the data point is to other points in the cluster, and how far away the data point is from points in other clusters. SC scores range from -1 to 1, with scores above 0.25 aimed for.[30]
- Akaike's Information Criteria (AIC)[31] – goodness of fit for datapoints in a cluster using a penalty based approach
- Bayesian information criterion (BIC)[31] – similar to AIC but using a larger penalty term.

Domain Expert Cluster Selection

Nine Domain experts ranked the groupings generated by the above clustering algorithms in a template (Figure 16) based on criteria in Figure 17. They were also asked for a qualitative opinion on if certain cluster sizes should be increased.

Algorithm Used	Order of Preference
K Prototypes	*2*
K Prototypes X	*1 (best)*
Agglomerative Hierarchical	*3*
K Means	*4*
K Mode	*5 (Worst)*

Figure 16: Ranking Template for generated clusters – with examples in italics entered

1. **Do the players in a particular cluster play similar shots from the same parts of the court? -** *eg; do these players have the same shot options/shot limitations when considering them from the same part of the court*
2. **Do the players sitting in different clusters play differently from the same part of the court -** *eg do players in other clusters have different shot options from the same part of the court*
3. **If you detached the names of the players and simply looked at Height, Weight, Playing hand, Backhand Type and Playing grips what groups make the most sense considering criteria above**
4. **Any other critical consideration that would support players being in the same cluster or sitting in different clusters**

Figure 17: Criteria guide sent to coaches to evaluate cluster ranked order

Variance Influence on cluster makeup

Using a random forest classifier, variable influence was determined for the selected K-Prototypes algorithm.

Analysis of Clusters

A multivariate analysis of variance (MANOVA) was used to evaluate the differences of all 10-surface rating and contextual performance variables among distinct player groups.[11] To conduct the MANOVA analysis:

- surface specific Elo ratings were derived based on match results.

- key contextual performance variables were scraped from ATP and WTA statistics pages. These include – Aces, First Serve Points won, Second Serve Points won, First Return Points won, and Second Return Points won

- Linear and Lateral movement averages were acquired from the Data Driven Sports Analytics match database

Court Segment Determination

In line with the domain experts understanding of the modern game, the Court was tessellated into segments that grouped where shots were being hit from (see Figure 18).

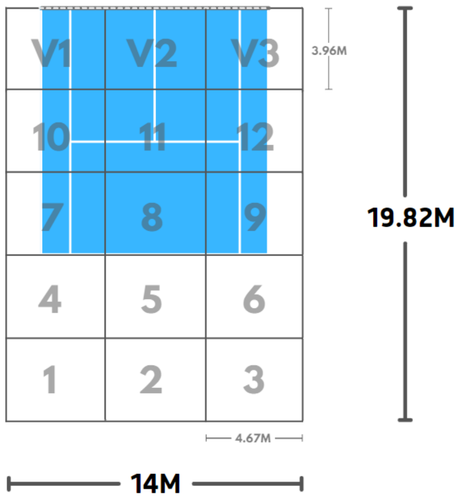


Figure 18: Court Tessellation of agreed court segments

Tennis Shot Classification and Level of Detail

The shot classification took into account where the player hit the shot from and where it landed on the other side of the court. Using the *hitpoint, IsGrounstroke, Short Arc Start Y* and *ProjectedBallMark Y* variables tennis shots with a level of detail describing the direction was classified in the dataset (see Figure 19 – Forehand and Figure 20 – Backhand). A domain expert reviewed and endorsed the approach.

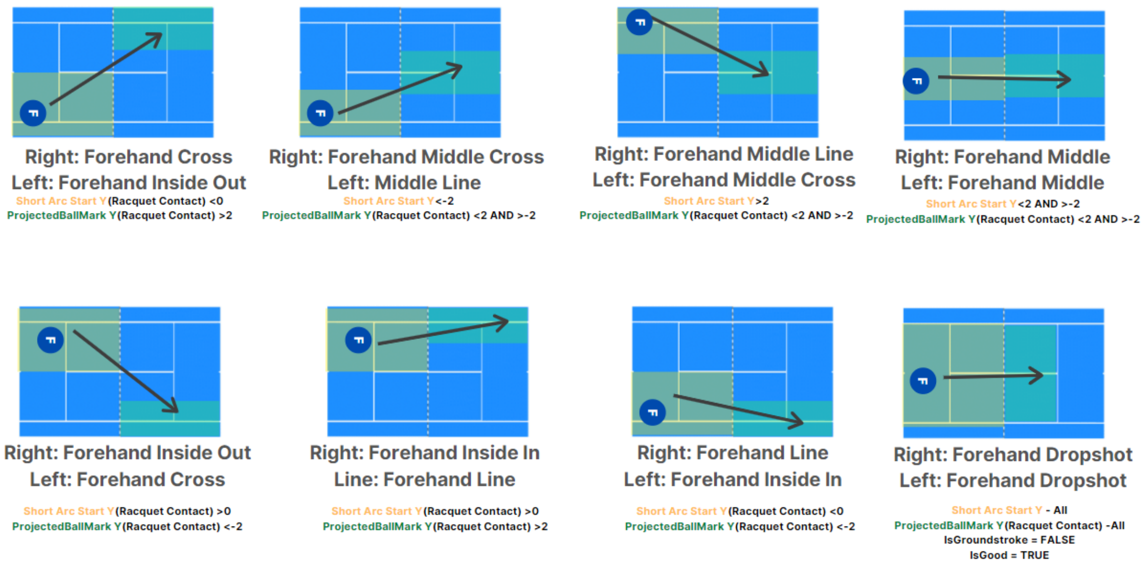


Figure 19: Level of Detail for Forehand Shot classifications based on Hawkeye

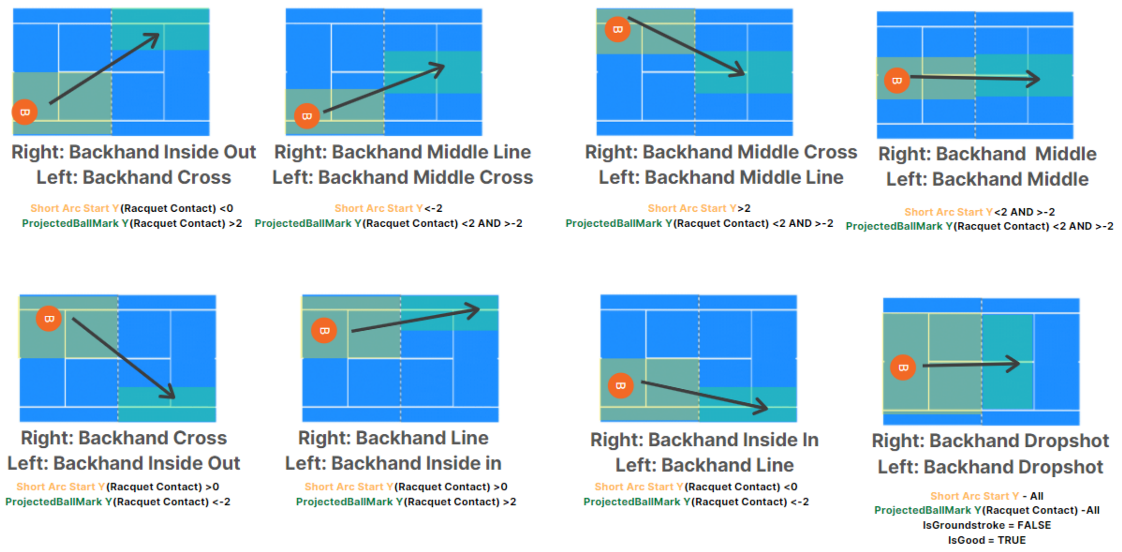


Figure 20: Level of Detail for Backhand Shot classifications based on Hawkeye

Shot sequence analysis - Sequential Market Basket Analysis (SMBA)

Data Structuring in ETL tool

In the Tableau Prep Extract Transform Load (ETL) tool, shot sequences were derived first by joining the dataset with itself and matching on the *shot ID* with the derived *Impact Player Next shot ID* (*Shot ID* -2) and *Unique Point ID*, a two-shot sequence can be created (see Figure 21 for data model and applied join clause). Using the ETL tool, data was restructured (See Figure 22) to input into python to perform SMBA.

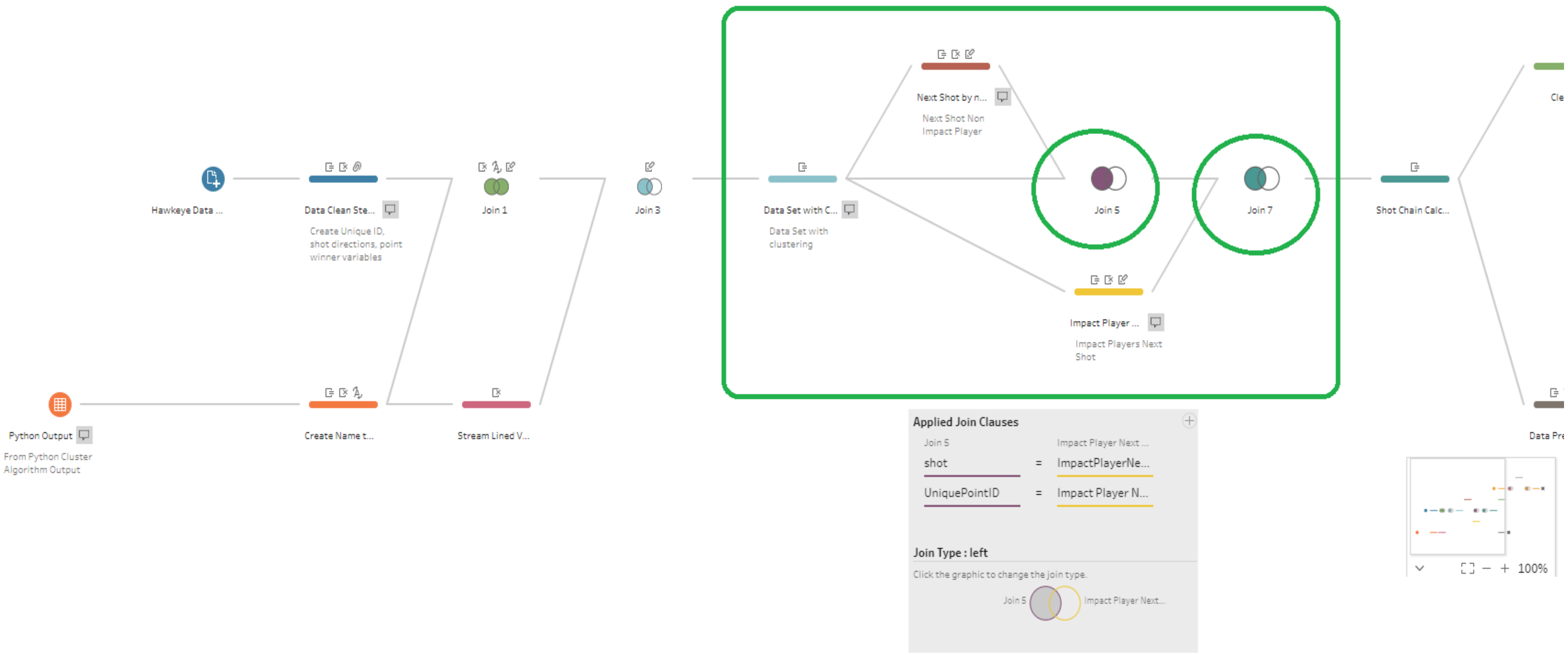


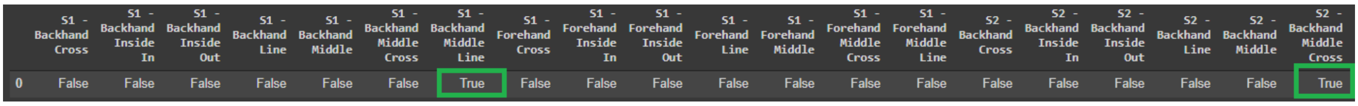
Figure 21: ETL Tool data model – Showing where dataset joined on itself to create shot sequences

	A	B	C	D	E	F	G	H	I	J	K
1	UniqueChainID	ChainShot2	ChainShot1	Chain Outcome Detail	Court Segments	Impact Player Role	WhichServe	impact.player	Impact Player K Protype Algorithm Clusters	Non_Impact Name	Non_Impact K Protype Algorithm Clusters
2	AO2017-MS112-1-1-2-1-1_01_02_1_213946-2	S2 - Forehand Cross	S1 - Forehand Middle Line	Lost Immediately	8	Returner	First Serve	R Federer	R-1 (ATP)	Jurgen Melzer	L-0 (ATP)
3	AO2017-MS112-1-1-3-1-1_01_03_1_214005-2	S2 - Forehand Cross	S1 - Backhand Inside Out	Lost Immediately	5	Returner	First Serve	R Federer	R-1 (ATP)	Jurgen Melzer	L-0 (ATP)
4	AO2017-MS112-1-2-1-2-1_02_01_2_214126-2	S2 - Backhand Line	S1 - Backhand Middle Cross	Lost Eventually	9	Returner	Second Serve	J Melzer	L-0 (ATP)	Roger Federer	R-1 (ATP)
5	AO2017-MS112-1-2-1-2-1_02_01_2_214126-3	S2 - Backhand Middle Cross	S1 - Forehand Middle	Won Eventually	5	Server	Second Serve	R Federer	R-1 (ATP)	Jurgen Melzer	L-0 (ATP)

Figure 22: Player Cluster and Spatio Temporal data restructured to perform Sequential Market Basket Analysis

Finding shot chains to evaluate

Rules (shot chains) were created by restructuring the data in python using one-hot encoding for all detailed shot options (see Figure 23) and running rule mining code. A sequential rule - A => C is created based on shot 1(S1) and shot 2(S2) in the dataset. S1 is the antecedent(A), and S2 is the consequent(C).



	S1 - Backhand Cross	S1 - Backhand Inside In	S1 - Backhand Inside Out	S1 - Backhand Line	S1 - Backhand Middle	S1 - Backhand Middle Cross	S1 - Backhand Middle Line	S1 - Forehand Cross	S1 - Forehand Inside In	S1 - Forehand Inside Out	S1 - Forehand Line	S1 - Forehand Middle	S1 - Forehand Middle Cross	S1 - Forehand Middle Line	S2 - Backhand Cross	S2 - Backhand Inside In	S2 - Backhand Inside Out	S2 - Backhand Line	S2 - Backhand Middle	S2 - Backhand Middle Cross
0	False	False	False	False	False	False	True	False	False	False	False	False	False	False	False	False	False	False	False	True

Figure 23: One hot encoding of all first and second shot options

To assist in the analysis process of all shot sequences the following metrics were used or derived to help with multi-metric filtering - See Figure 24:

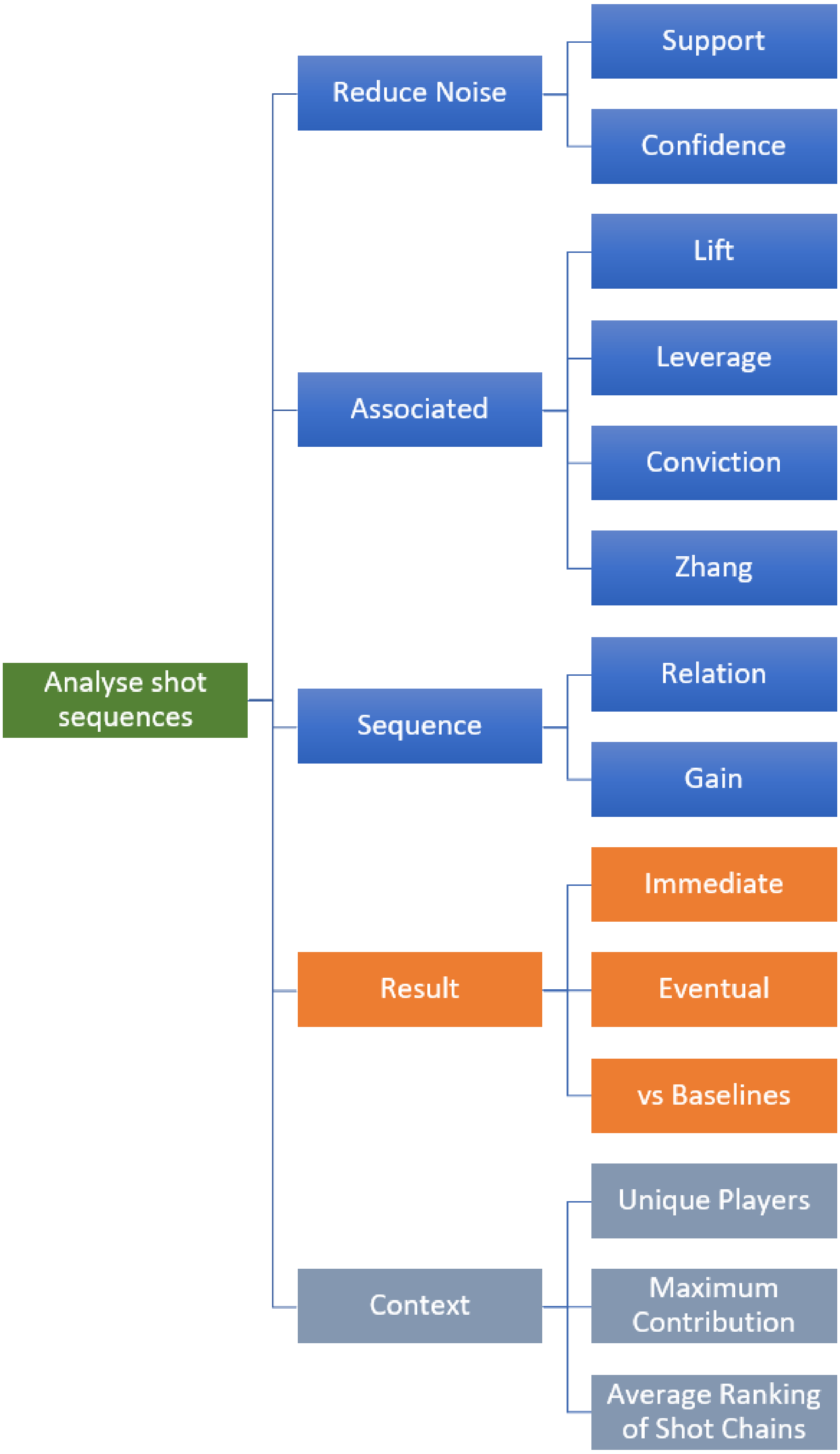


Figure 24: Shot Chain Suitability evaluation (Blue), Success evaluation (Orange) and Context providing metrics (Grey)

Shot Chain Suitability Evaluation – More meaningful rules to analyse

Support is used to measure the frequency of an itemset in a database (Formula in Equation 5). E.g. how frequently *Backhand Middle* and *Forehand Middle* occur together. The higher the support percentage, the more frequently the itemset occurs.

$$\mathbf{Support(A \rightarrow C)} = \frac{\mathbf{P(A,C)}}{\mathbf{Total}}$$

Equation 5: Support Formula

Confidence is used to determine the probability a transaction that contains the items on the left-hand side of the rule also contains the item on the right-hand side (Formula in Equation 6). This metric is not symmetrical, hence Forehand middle → Backhand Middle is different to Backhand Middle → Forehand Middle.

$$\mathbf{Confidence(A \rightarrow C)} = \frac{\mathbf{P(A,C)}}{\mathbf{P(A)}}$$

Equation 6: Confidence Formula

Support and *Confidence* will be used reduce the shot chains analysed. A support of 3%, as well as a confidence of 10% is set by considering the shot sequence permutations and by reference to other Market Basket Analysis with similar data set size and transaction permutations.[8]

Lift is used help determine which shot sequences are more associated and less random by looking at how much more often the antecedent and consequent of a rule A→C occur together than if they were statistically independent (Formula in Equation 7). For example, how much more likely is backhand middle and forehand middle to occur together than random. Lift scores are interpreted as follows:

- Lift = 1; implies relationship between A and C is random
- Lift > 1; implies that there is a positive relationship between A and C (occurs more than random)
- Lift < 1; implies that there is a negative relationship between A and C (occur less than random)

A minimum lift of 1.4 is used in this analysis.

$$\mathbf{Lift(A \rightarrow C) = \frac{Confidence}{P(\mathit{C})}}$$

Equation 7: Lift Formula [Range 0, ∞]

Leverage is derived to compute the difference between the observed frequency of A and C appearing together and the frequency that would be expected if A and C were independent.[5] It can help determine how many more shot sequences occur together than independently (Formula in Equation 8). A minimum leverage of 2% is used in this analysis.

$$\mathbf{Leverage(A \rightarrow C) = support(A \rightarrow C) - support(A) \times support(C)}$$

Equation 8: Lift Formula [Range -1, 1]

Conviction is calculated to determine how dependent the consequent is on the antecedent and measures the implication strength of the rule from statistical independence[32] (Equation 9 formula). Conviction helps isolate rules with high occurring shots like backhand middle cross. Backhand middle cross can occur with almost any other shot, but that does not always imply their relationship. By using conviction, it ensures considering all the other combinations that backhand middle cross is paired with to weight your association.

A minimum conviction of 1 is used in this analysis.

$$\mathbf{Conviction(A \rightarrow C) = \frac{1 - support(C)}{1 - confidence\ (A \rightarrow C)}}$$

Equation 9: Conviction Formula [Range 0, ∞]

The Zhang metric is also calculated to help determine the level of association or disassociation for shot chains[33] (Formula in Equation 10). Zhang scores are interpreted as follows:

- +1 indicates perfect association
- −1 indicates perfect dissociation
- 0 indicates no association

A minimum Zhang metric of 0.5 is used in this analysis.

$$\mathbf{Zhang}(\mathbf{A} \rightarrow \mathbf{C}) = \frac{\mathbf{support}(\mathbf{A} \rightarrow \mathbf{C}) - \mathbf{support}(\mathbf{A})\mathbf{Support}(\mathbf{C})}{\boldsymbol{Max}[\boldsymbol{Support}(\boldsymbol{AC})(\mathbf{1} - \boldsymbol{Support}(\boldsymbol{A})), \boldsymbol{Support}(\boldsymbol{A})(\boldsymbol{Support}(\boldsymbol{C}) - \boldsymbol{Support}(\boldsymbol{AC}))]}$$

Equation 10: Lift Formula [Range -1, 1]

Shots occurring in a sequence

The Gain statistic is calculated to determine the gain (loss) in probability of shot C occurring after a previous shot A, relative to the unconditional probability P(C).[9] It shows the increase/decrease in odds of shot C being hit after shot A. It is based on sequence, a point of difference to previously defined lift (Formula in Equation 11).

$$\mathbf{GAIN}\ [\boldsymbol{A} \rightarrow \boldsymbol{C}] = \frac{\boldsymbol{P}[\boldsymbol{A} - \boldsymbol{C}]}{\boldsymbol{P}[\boldsymbol{C}]} - 1$$

Equation 11: Gain Formula

Using Gain we can calculate Relation classifications, of which we only rely on sequences that are classified as occurring in an ‘A→C sequence’ or are classified as ‘Complements’ for this project.

- ***A→C sequence*** *– Where GAIN[A→C] > 0 and GAIN[C→A] <=0*
- ***Complements -*** *Where GAIN[A→C] > 0 and GAIN[C→A] >0*
- ***C→A sequence*** *– Where GAIN[A→C] <= 0 and GAIN[C→A] >0*
- ***Substitutes -*** *Where GAIN[A→C] < 0 and GAIN[C→A] <0*
- ***Independent*** *Where GAIN[A→C] = 0 and GAIN[C→A] =0*

Equation 12: Relation Classifications

Success Evaluation Metrics

Shot chain success is evaluated by breaking up the shot chains results being played by:

- Point Won Immediately
- Point Won Eventually
- Point Lost Immediately
- Point Lost Eventually

A weighted per point outcome is derived using the formula in Equation 13 with assigned weights of 1.5 for both Immediate Point Wins and Losses based on domain expert guidance.

$$\text{Weighted Per Point Outcome} = \frac{(\theta \times \text{No. Immediate Point Wins} + \text{No. Eventual Point Wins}) - (\delta \times \text{No. Immediate Point Loss} + \text{No. Eventual Point Wins})}{\text{Total Observed Shot Chains}}$$

θ = Assigned Weighting to Immediate point win
δ = Assigned Weighting to Immediate point loss

Equation 13: Weighted Per Point Outcome (WPPO) Formula

Shot chains success was also evaluated against the following baselines – a) High level Shot chains (e.g. Forehand → Forehand); b) Any two shots being hit.

Context Providing Metrics

The best performing shot sequences will be presented with:

- Number of unique players hitting the shot sequence
- Maximum contribution percentage by one player to the total of the shot sequences
- Average ranking of the executed shot chain

$$\text{Average Ranking of Excecuted Shot chain} = \frac{\text{P1 Rank} \times \text{P Sequences Hit} + \cdots + \text{P}(\text{n}-1)\ \text{Rank} \times \text{P}(\text{n}-1)\ \text{Sequences Hit}}{\text{Total Shot Sequences}}$$

Equation 14: Average Ranking of Executed Shot Formula

Results

Clustering

Algorithm Selection

The clustering generated by the K-Prototypes algorithm with *Cao* initialisation was selected from 5 different clustering solutions based on majority selection of Domain experts (6 of 9) and the best average ranking.

Utilised Algorithm	No. of First Preference	Average Ranking
K Prototypes	6	1.44
K Prototypes X	1	2.33
Agglomerative Hierarchical	2	2.44
K Means	0	3.89
K Modes	0	4.89

Figure 25: Cluster Selection by Domain Experts

Cluster Size

Table 2 and Table 3 show the cluster sizes selected for each of the cluster algorithms used on the ATP and WTA datasets respectively.

Algorithm Used	Right Handed Clusters	Left Handed Clusters	Total Clusters
K-Means	5	4	9
K-Mode	5	3	8
K-Prototypes	6	5	11
K-Prototypes X	6	4	10
Agglomerative	6	4	10

Table 2: ATP Cluster Sizes

Algorithm Used	Right Handed Clusters	Left Handed Clusters	Total Clusters
K-Means	4	3	7
K-Mode	5	4	9
K-Prototypes	4	3	7
K-Prototypes X	5	3	8
Agglomerative	6	3	9

Table 3: WTA Cluster Sizes

The cluster size for the selected clustering algorithm (K-Prototypes with Cao Initialisation) were determined by reference to the Elbow chart with bend values pin-pointed by the python Knee locator package.

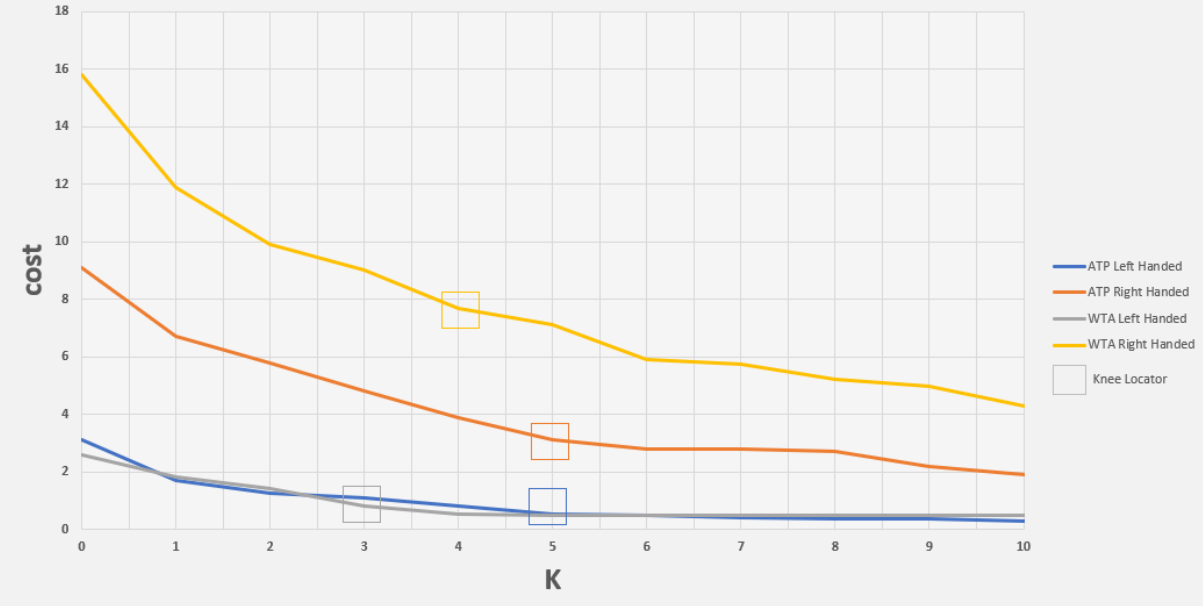


Figure 26: Elbow Chart for ATP & WTA left and right-handed data partitions

Based on the domain expert's recommendation the cluster size for right-handed ATP players was increased from the knee locator selection of 5 to 6 for the K-Prototypes algorithm.

The silhouette coefficient at the selected cluster sizes is presented in the table below. All numbers are above 0.3 indicating acceptable cluster models.[30]

Data Partition	Cluster Size	Silhouette Coefficient
ATP Right Handers	6	0.36
ATP Left Handers	5	0.37
WTA Right Handers	4	0.30
WTA Left Handers	3	0.31

Table 4: Silhouette Coefficient Scores for selected cluster sizes of K-Prototype algorithm on relevant data partition

Clustering - Variable Influence

The variable influence in determining the cluster groups is shown in Table 5 with *Height* and *Weight* being the most influential. For the men the next most influential variable was backhand type in determining cluster group, whereas for the women the *Dominant-Hand Backhand Grip* or *Forehand Grip* played a more influential role.

Influence Ranking	ATP Right	ATP Left	WTA Right	WTA Left
1	Weight (0.33)	Weight (0.28)	Height (0.26)	Weight (0.29)
2	Height (0.25)	Height (0.28)	Weight (0.18)	Height (0.28)
3	Backhand Type (0.12)	Backhand Type (0.14)	Dominant Hand Backhand Grip – *Eastern* (0.18)	Forehand Grip – *Western* (0.13)
4	Dominant Hand Backhand Grip – *Eastern* (0.10)	Forehand Grip – *Western* (0.14)	Forehand Grip – *Semi-Western* (0.12)	Forehand Grip – *Semi-Western – Western* (0.1)
5	Forehand Grip – *Western* (0.09)	Forehand Grip – *Semi-Western* (0.13)	Forehand Grip – *Western* (0.08)	Dominant Hand Backhand Grip – *Eastern* (0.07)
6	Forehand Grip – *Semi-Western – Western* (0.04)	Forehand Grip – *Semi-Western – Western* (0.03)	Dominant Hand Backhand Grip – *Eastern -Semi Western* (0.06)	Forehand Grip – *Semi-Western* (0.06)
7	Forehand Grip – *Semi-Western* (0.04)	Dominant Hand Backhand Grip – *Eastern* (0.02)	Forehand Grip – *Semi-Western – Western* (0.05)	Forehand Grip – *Eastern-Semi Western* (0.05)
8	Forehand Grip – *Eastern* (0.02)	Forehand Grip – *Eastern* (0.02)	Dominant Hand Backhand Grip – *Semi-Western* (0.03)	Dominant Hand Backhand Grip – *Continental - Eastern* (0.03)
9	Forehand Grip – *Eastern-Semi Western* (0.01)	Forehand Grip – *Eastern-Semi Western* (0.01)	Dominant Hand Backhand Grip – *Continental - Eastern* (0.02)	Backhand Type (0.02)
10	Dominant Hand Backhand Grip – *Continental* (0.01)	Dominant Hand Backhand Grip – *Continental* (0.01)	Forehand Grip – *Eastern-Semi Western* (0.02)	Dominant Hand Backhand Grip – *Continental* (0.01)

Table 5: Top 10 Variable influence ranking for selected K Prototype algorithm on relevant data partition

Analysis of clusters against key statistics

The height and weight relationship between the player clusters is presented in Figure 27 and Figure 28. The Ranked order of ATP and WTA players based on surface specific ELO rating and their tour rankings is shown in Table 6 and Table 7 respectively. R-1(ATP) and R-2(WTA) player clusters had the highest average ATP and WTA tour ranking respectively, and also had the highest average Elo ratings for all three distinct playing surfaces.

The results of MANOVA shows that there was some significant effect of player groups on surface specific Elo ratings & contextual performance variables (see Table 8 – ATP, Table 9 - WTA). The *P* value on the dependent variable *Playercluster* is less than 0.001 (*Pillai's trace*) indicating a significant effect of player types on the performance variables. The results of post hoc tests are presented in Tables 10 (ATP) and Table 11 (WTA) along with the descriptive statistics of some key surface specific rating variables and contextual performance variables. The detailed player cluster names with an example of a player in the cluster is presented in Table 12.

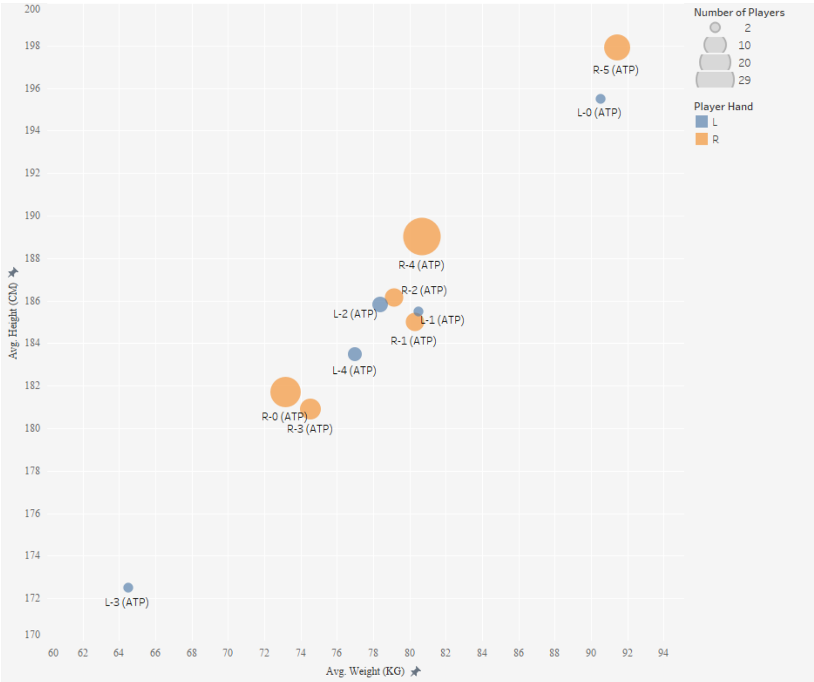


Figure 27: Height vs Weight ATP Players - K Prototypes Clustering Algorithm

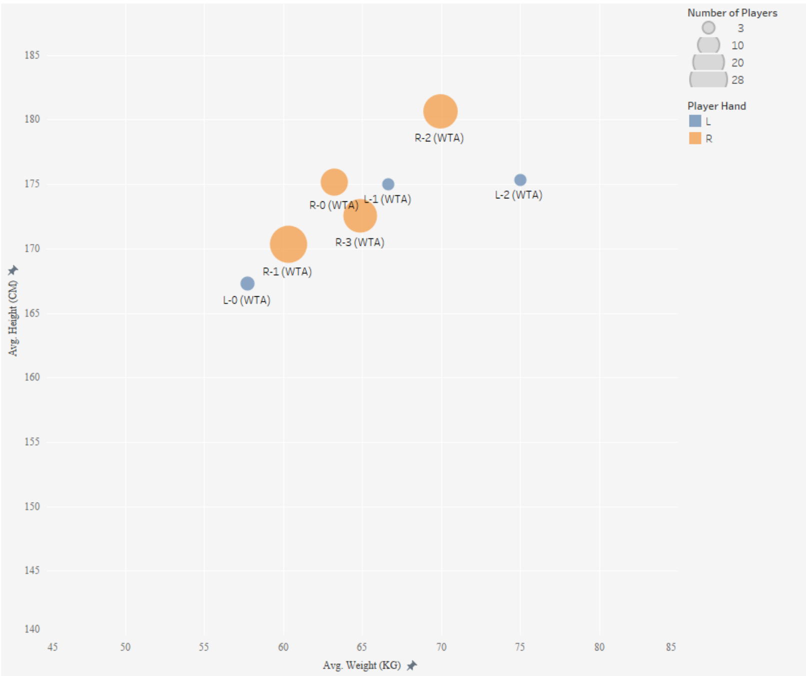


Figure 28: Height vs Weight WTA Players - K Prototypes Clustering Algorithm

Ranking Order	Hard Court	Clay Court	Grass Court	ATP Tour Rank
1	R-1(ATP) - 1870	R-1(ATP) - 1853	R-1(ATP) - 1749	R-1(ATP) - 15
2	L-2 (ATP) - 1804	L-2 (ATP) - 1793	R-5 (ATP) -1740	L-1(ATP) - 38
3	R-5 (ATP) -1789	R-5 (ATP) -1772	L-2 (ATP) - 1737	L-2 (ATP) - 43
4	R-4 (ATP) – 1755	R-2 (ATP) -1765	L-1 (ATP) - 1728	R-5 (ATP) -45
5	R-2 (ATP) -1739	R-3 (ATP) - 1735	R-2 (ATP) -1668	L-4 (ATP) - 52
6	L-3 (ATP) -1735	L-1 (ATP) - 1714	R-4 (ATP) – 1666	R-3 (ATP) -52
7	L-1 (ATP) - 1724	R-4 (ATP) – 1713	R-0 (ATP) – 1629	R-4 (ATP) -55
8	R-3 (ATP) - 1713	R-0 (ATP) – 1713	R-3 (ATP) - 1621	R-2 (ATP) -56
9	R-0 (ATP) – 1710	L-0 (ATP) -1692	L-3 (ATP) -1597	R-0 (ATP) -57
10	L-4 (ATP) -1657	L-3 (ATP) -1672	L-4 (ATP) -1591	L-3 (ATP) - 71
11	L-0 (ATP) -1636	L-4 (ATP) -1652	L-0 (ATP) -1563	L-0 (ATP) - 76

Table 6: ATP Ranked Player clusters- Average Surface Specific ELO and Average Ranking

Ranking Order	Hard Court	Clay Court	Grass Court	WTA Tour Rank
1	R-2 (WTA) -1787	R-2 (WTA) -1696	R-2 (WTA) -1573	R-2 (WTA) -35
2	L-1 (WTA) -1775	L-1 (WTA) -1695	R-3 (WTA) -1571	L-1 (WTA) -43
3	R-3 (WTA) -1760	R-0 (WTA) -1666	L-2 (WTA) -1553	R-3 (WTA) -45
4	L-0 (WTA) -1730	L-2 (WTA) -1665	L-1 (WTA) -1514	L-2 (WTA) -59
5	R-1 (WTA) -1696	R-3 (WTA) -1611	R-0 (WTA) -1496	R-0 (WTA) -61
6	L-2 (WTA) -1687	R-1 (WTA) -1609	R-1 (WTA) -1453	R-1 (WTA) -61
7	R-0 (WTA) -1681	L-0 (WTA) -1537	L-0 (WTA) -1337	L-0 (WTA) -64

Table 7: WTA Ranked Player clusters- Average Surface Specific ELO and Average Ranking

Multivariate Linear Model - ATP MANOVA analysis

	Value	Num DF	Den DF	F Value	Pr > F
Pillai's Trace - Intercept	0.9511	10.0000	80.0000	155.4751	0.0000
Pillai's Trace - PlayerCluster	2.1229	100.0000	890.0000	2.3986	0.0000

Table 8: ATP MANOVA Analysis for statistical variables

Multivariate Linear Model - WTA MANOVA analysis

	Value	Num DF	Den DF	F Value	Pr > F
Pillai's Trace - Intercept	0.9667	10.0000	84.0000	244.1517	0.0000
Pillai's Trace - PlayerCluster	1.3982	60.0000	534.0000	2.7041	0.0000

Table 9: WTA MANOVA Analysis for statistical variables

Surface Specific Elo and Contexual Performance Variables	K Prototype Algorithm ATP Player Cluster Grouping												
	R-0 (ATP) (n=19)	R-1 (ATP) (n=7)	R-2 (ATP) (n=7)	R-3 (ATP) (n=9)	R-4 (ATP) (n=29)	R-5 (ATP) (n=14)	L-0 (ATP) (n=2)	L-1 (ATP) (n=2)	L-2 (ATP) (n=5)	L-3 (ATP) (n=2)	L-4 (ATP) (n=4)	F	*PR(>F)*
Clay Elo	1713.12 (107.36)	1852.84 (131.83)	1765.17 (74.35)	1735.34 (102.53)	1713.19 (123.69)	1771.89 (129.56)	1691.80 (120.49)	1714.15 (100.62)	1792.62 (176.10)	1671.80 (30.55)	1652.48 (76.80)	1.44	0.17485
Grass Elo	1628.59 (114.12)	1749.11 (123.32)	1667.67 (131.26)	1620.67 (107.50)	1665.81 (111.04)	1739.86 (113.38)	1562.75 (118.30)	1727.80 (49.36)	1736.58 (86.52)	1597.35 (18.03)	1590.88 (73.64)	2.06	0.03562
Hard Elo	1709.56 (123.43)	1869.77 (126.87)	1738.64 (166.86)	1712.74 (106.51)	1754.95 (109.55)	1788.76 (117.78)	1635.60 (25.46)	1723.90 (114.27)	1803.88 (124.43)	1735.20 (3.96)	1656.55 (64.12)	1.76	0.07954
Aces Per Match	3.88 (1.95)	6.17 (1.54)	4.94 (1.97)	4.23 (1.57)	6.96 (2.82)	11.17 (4.00)	5.66 (2.80)	9.44 (1.52)	5.18 (1.71)	1.75 (0.07)	4.48 (0.73)	9.07	3.82E-10
First Serve Win %	68.62 (4.34)	72.26 (4.34)	70.48 (4.42)	68.94 (4.06)	71.40 (4.04)	77.23 (2.51)	69.71 (3.47)	76.69 (0.74)	71.87 (3.26)	64.09 (2.01)	69.09 (1.31)	6.00	6.72E-07
Second Serve Win %	50.11 (3.99)	53.17 (3.19)	51.31 (1.59)	52.74 (2.45)	49.82 (4.18)	52.37 (2.90)	50.18 (1.05)	51.83 (0.38)	51.58 (1.61)	46.94 (3.68)	51.27 (1.14)	1.54	0.13641
First Return Win %	30.53 (2.75)	29.94 (1.19)	31.10 (1.66)	29.70 (2.99)	29.53 (2.82)	26.68 (4.62)	29.94 (4.46)	27.72 (1.99)	30.39 (3.05)	29.86 (5.06)	28.13 (23.34)	1.80	0.07168
Second Return Win %	49.55 (4.16)	49.16 (3.17)	50.43 (2.62)	49.76 (2.56)	48.43 (3.44)	46.92 (3.78)	50.51 (4.16)	46.52 (3.08)	50.66 (4.73)	45.70 (6.56)	47.52 (5.52)	1.09	0.37976
Lateral Positioning (+/- m from Centre)	0.54 (0.32)	-0.44 (0.95)	-0.79 (1.06)	0.57 (0.31)	-0.91 (0.83)	-0.06 (0.15)	0.38 (0.45)	0.82 (0.24)	0.28 (0.33)	0.40 (0.45)	0.33 (0.12)	16.14	1.42E-16
Linear Positioning (+/- m from baseline)	0.24 (0.54)	0.22 (0.67)	-0.09 (0.54)	-1.25 (0.21)	0.18 (0.43)	0.92 (0.47)	0.43 (0.30)	0.54 (0.62)	-0.17 (0.30)	-1.42 (0.48)	-0.72 (0.15)	10.27	1.56E-11

Table 10: Means (standard deviations) of Surface specific - and contextual performance variables for ATP Player clusters and MANOVA results of pairwise comparisons

Surface Specific Elo and Contexual Performance Variables	K Prototype Algorithm WTA Player Cluster Grouping								
	R-0 (WTA) (n=15)	R-1 (WTA) (n=28)	R-2 (WTA) (n=23)	R-3 (WTA) (n=23)	L-0 (WTA) (n=4)	L-1 (WTA) (n=3)	L-2 (WTA) (n=4)	F	*PR(>F)*
Clay Elo	1666.43 (132)	1608.7 (129.01)	1701.78 (93.19)	1611.07 (132.79)	1536.67 (103.25)	1694.96 (52.03)	1638.55 (155.51)	2.21	0.04921
Grass Elo	1496.19 (116.63)	1453.46 (125.7)	1575.69 (101.57)	1570.75 (107.88)	1337.2 (1337.2)	1514.26 (283.91)	1545.05 (144.49)	4.50	0.00048
Hard Elo	1681.4 (103.29)	1695.73 (131.72)	1796.52 (105.55)	1759.93 (133.4)	1729.52 (146.16)	1774.56 (134.33)	1657.3 (163.91)	2.34	0.03799
Aces Per Match	2.23 (1.51)	2.02 (0.96)	3.48 (1.34)	2.89 (1.6)	2.01 (1)	1.58 (0.42)	7.07 (4.83)	7.67	1.00E-06
First Serve Win %	62.2 (4.81)	62.32 (3.42)	65.91 (3.63)	64.78 (5.49)	62.97 (4.23)	63.14 (1.88)	68.71 (4.42)	2.90	1.23E-02
Second Serve Win %	44.52 (3.59)	45.85 (3.08)	45.55 (3.11)	46.26 (3.64)	45.17 (2.93)	44.8 (3.58)	44.38 (4.24)	0.56	0.76266
First Return Win %	37 (2.29)	36.79 (2.65)	36.53 (2.91)	35.69 (3.31)	35.82 (0.79)	36.53 (2.76)	31.04 (7.49)	2.38	0.0346
Second Return Win %	54.36 (2.27)	54.43 (2.45)	54.59 (3.5)	54.81 (2.71)	54.18 (2.73)	55.09 (2.04)	51.88 (3.98)	0.66	0.68535
Lateral Positioning (+/- m from Centre)	-0.64 (0.36)	-0.82 (0.56)	-0.37 (0.84)	-0.05 (0.85)	0.5 (0.15)	0.18 (0.16)	0.26 (0.09)	5.39	8.10E-05
Linear Positioning (+/- m from baseline)	-0.21 (0.55)	-0.78 (1.07)	-5.38 (6.25)	0.01 (0.41)	-0.63 (0.67)	-0.04 (0.16)	0.76 (0.42)	8.04	5.30E-07

Table 11: Means (standard deviations) of Surface specific - and contextual performance variables for WTA Player clusters and MANOVA results of pairwise comparison

Player Cluster	Cluster Name	Example of player in the cluster
R-0 (ATP)	Smaller body frame right-handed male	David Goffin
R-1 (ATP)	Single handed backhand right-handed male	Roger Federer
R-2 (ATP)	Two handed Eastern grip backhand right-handed male	Christian Garin
R-3 (ATP)	Extreme closed grip forehand, right-handed male	John Millman
R-4 (ATP)	Mid-size body frame right-handed male	Aslan Karatsev
R-5 (ATP)	Big bodied right-handed male	John Isner
L-0 (ATP)	Bigger body frame left-handed male	Jiri Vesely
L-1 (ATP)	Single handed backhand left-handed male	Denis Shapovalov
L-2 (ATP)	Mid-size open grip backhand, left-handed male	Cameron Norrie
L-3 (ATP)	Smaller bodied left-handed male	Yoshihito Nishioka
L-4 (ATP)	Mid-size body frame closed forehand grip, left-handed male	Guido Pella
R-0 (WTA)	Mid-size closed extreme grip forehands and semi open to closed backhand grip, right-handed female	Danielle Collins
R-1 (WTA)	Mid-size closed forehand grip and open grip backhand, right-handed female	Iga Swiatek
R-2 (WTA)	Bigger body frame right-handed female	Karolina Pliskova
R-3 (WTA)	Semi open forehand and backhand grip, all body-size right-handed female	Belinda Bencic
L-0 (WTA)	Smaller body-frame left-handed female	Leylah Fernandez
L-1 (WTA)	Mid-size body frame left-handed female	Angelique Kerber
L-2 (WTA)	Big body-frame left-handed female	Petra Kvitova

Table 12: Cluster Names and examples of players within a cluster.

	Court Segments - Where Shots are hit from															
Player Clusters	1	2	3	4	5	6	7	8	9	10	11	12	V1	V2	V3	Out of segment range	Total
R-0 (ATP)	3 (0.21%)	5 (0.35%)	9 (0.64%)	349 (24.72%)	399 (28.26%)	223 (15.79%)	139 (9.84%)	124 (8.78%)	98 (6.94%)	12 (0.85%)	32 (2.27%)	7 (0.5%)	2 (0.14%)	9 (0.64%)	(0%)	1 (0.07%)	1412 (100%)
R-1 (ATP)	13 (0.1%)	8 (0.06%)	18 (0.14%)	2759 (21.53%)	2960 (23.1%)	1688 (13.17%)	1737 (13.56%)	1792 (13.98%)	962 (7.51%)	141 (1.1%)	479 (3.74%)	86 (0.67%)	24 (0.19%)	117 (0.91%)	22 (0.17%)	8 (0.06%)	12814 (100%)
R-2 (ATP)	2 (0.22%)	2 (0.22%)	5 (0.56%)	208 (23.19%)	256 (28.54%)	158 (17.61%)	80 (8.92%)	99 (11.04%)	54 (6.02%)	4 (0.45%)	11 (1.23%)	9 (1%)	4 (0.45%)	3 (0.33%)	(0%)	2 (0.22%)	897 (100%)
R-3 (ATP)	3 (0.4%)	(0%)	7 (0.92%)	178 (23.51%)	236 (31.18%)	106 (14%)	63 (8.32%)	69 (9.11%)	61 (8.06%)	5 (0.66%)	21 (2.77%)	5 (0.66%)	(0%)	2 (0.26%)	(0%)	1 (0.13%)	757 (100%)
R-4 (ATP)	10 (0.61%)	6 (0.37%)	6 (0.37%)	413 (25.14%)	504 (30.68%)	279 (16.98%)	146 (8.89%)	140 (8.52%)	67 (4.08%)	20 (1.22%)	24 (1.46%)	14 (0.85%)	3 (0.18%)	7 (0.43%)	2 (0.12%)	2 (0.12%)	1643 (100%)
R-5 (ATP)	2 (0.15%)	1 (0.07%)	1 (0.07%)	277 (20.32%)	431 (31.62%)	205 (15.04%)	141 (10.34%)	157 (11.52%)	76 (5.58%)	10 (0.73%)	42 (3.08%)	9 (0.66%)	1 (0.07%)	6 (0.44%)	3 (0.22%)	1 (0.07%)	1363 (100%)
L-0 (ATP)	1 (0.18%)	(0%)	1 (0.18%)	62 (11.36%)	115 (21.06%)	97 (17.77%)	45 (8.24%)	64 (11.72%)	59 (10.81%)	6 (1.1%)	65 (11.9%)	7 (1.28%)	6 (1.1%)	15 (2.75%)	3 (0.55%)	(0%)	546 (100%)
L-1 (ATP)	(0%)	(0%)	(0%)	36 (13.79%)	72 (27.59%)	59 (22.61%)	20 (7.66%)	33 (12.64%)	28 (10.73%)	3 (1.15%)	7 (2.68%)	1 (0.38%)	1 (0.38%)	1 (0.38%)	(0%)	(0%)	261 (100%)
L-2 (ATP)	13 (2.48%)	38 (7.24%)	5 (0.95%)	106 (20.19%)	153 (29.14%)	126 (24%)	13 (2.48%)	25 (4.76%)	29 (5.52%)	3 (0.57%)	4 (0.76%)	(0%)	(0%)	(0%)	(0%)	10 (1.9%)	525 (100%)
L-3 (ATP)																	
L-4 (ATP)																	
R-0 (WTA)	3 (0.09%)	1 (0.03%)	2 (0.06%)	768 (22.19%)	1069 (30.89%)	837 (24.18%)	250 (7.22%)	321 (9.27%)	164 (4.74%)	5 (0.14%)	25 (0.72%)	9 (0.26%)	(0%)	(0%)	2 (0.06%)	5 (0.14%)	3461 (100%)
R-1 (WTA)	10 (0.54%)	3 (0.16%)	5 (0.27%)	549 (29.74%)	565 (30.61%)	435 (23.56%)	74 (4.01%)	89 (4.82%)	68 (3.68%)	6 (0.33%)	23 (1.25%)	8 (0.43%)	(0%)	3 (0.16%)	(0%)	8 (0.43%)	1846 (100%)
R-2 (WTA)	2 (0.04%)	2 (0.04%)	1 (0.02%)	931 (20.79%)	1348 (30.1%)	1016 (22.69%)	332 (7.41%)	470 (10.5%)	299 (6.68%)	6 (0.13%)	54 (1.21%)	3 (0.07%)	3 (0.07%)	4 (0.09%)	1 (0.02%)	6 (0.13%)	4478 (100%)
R-3 (WTA)	1 (0.01%)	3 (0.03%)	5 (0.05%)	1912 (17.97%)	3686 (34.65%)	1509 (14.18%)	833 (7.83%)	1160 (10.9%)	1238 (11.64%)	35 (0.33%)	158 (1.49%)	37 (0.35%)	7 (0.07%)	36 (0.34%)	12 (0.11%)	7 (0.07%)	10639 (100%)
L-0 (WTA)																	
L-1 (WTA)																	
L-2 (WTA)	(0%)	(0%)	(0%)	19 (11.05%)	51 (29.65%)	27 (15.7%)	25 (14.53%)	29 (16.86%)	18 (10.47%)	1 (0.58%)	1 (0.58%)	(0%)	(0%)	1 (0.58%)	(0%)	(0%)	172 (100%)
Grand Total	63 (0.15%)	69 (0.17%)	65 (0.16%)	8567 (20.99%)	11845 (29.02%	6765 (16.58%)	3898 (9.55%)	4572 (11.2%)	3221 (7.89%)	257 (0.63%)	946 (2.32%)	195 (0.48%)	51 (0.12%)	204 (0.5%)	45 (0.11%)	51 (0.12%)	40814 (100%)

Table 13: Shots hit from court segment by Player Type Cluster (Percentage of shots hit by Player Type Cluster in court segment)

	Court Segments - Where Ground Stroke Only Shot chains are hit from															
Player Clusters	1	2	3	4	5	6	7	8	9	10	11	12	V1	V2	V3	Out of segment range	Total
R-0 (ATP)	(0%)	1 (0.21%)	2 (0.43%)	112 (24.4%)	136 (29.62%)	83 (18.08%)	53 (11.54%)	47 (10.23%)	24 (5.22%)	1 (0.21%)	(0%)	(0%)	(0%)	(0%)	(0%)	(0%)	459 (100%)
R-1 (ATP)	1 (0.02%)	1 (0.02%)	6 (0.14%)	1032 (24.21%)	1198 (28.11%)	585 (13.72%)	632 (14.83%)	528 (12.39%)	269 (6.31%)	1 (0.02%)	6 (0.14%)	1 (0.02%)	(0%)	(0%)	(0%)	1 (0.02%)	4261 (100%)
R-2 (ATP)	1 (0.35%)	1 (0.35%)	(0%)	65 (22.8%)	100 (35.08%)	49 (17.19%)	21 (7.36%)	39 (13.68%)	8 (2.8%)	(0%)	(0%)	(0%)	(0%)	1 (0.35%)	(0%)	(0%)	285 (100%)
R-3 (ATP)	(0%)	(0%)	1 (0.43%)	49 (21.49%)	91 (39.91%)	34 (14.91%)	20 (8.77%)	21 (9.21%)	12 (5.26%)	(0%)	(0%)	(0%)	(0%)	(0%)	(0%)	(0%)	228 (100%)
R-4 (ATP)	2 (0.35%)	2 (0.35%)	1 (0.17%)	164 (29.02%)	187 (33.09%)	83 (14.69%)	52 (9.2%)	52 (9.2%)	21 (3.71%)	(0%)	1 (0.17%)	(0%)	(0%)	(0%)	(0%)	(0%)	565 (100%)
R-5 (ATP)	(0%)	(0%)	(0%)	97 (20.81%)	179 (38.41%)	64 (13.73%)	55 (11.8%)	48 (10.3%)	21 (4.5%)	1 (0.21%)	1 (0.21%)	(0%)	(0%)	(0%)	(0%)	(0%)	466 (100%)
L-0 (ATP)	1 (0.69%)	(0%)	(0%)	19 (13.28%)	47 (32.86%)	30 (20.97%)	12 (8.39%)	16 (11.18%)	17 (11.88%)	(0%)	1 (0.69%)	(0%)	(0%)	(0%)	(0%)	(0%)	143 (100%)
L-1 (ATP)	(0%)	(0%)	(0%)	12 (12.5%)	30 (31.25%)	29 (30.2%)	7 (7.29%)	11 (11.45%)	7 (7.29%)	(0%)	(0%)	(0%)	(0%)	(0%)	(0%)	(0%)	96 (100%)
L-2 (ATP)	4 (2.19%)	15 (8.24%)	(0%)	26 (14.28%)	65 (35.71%)	48 (26.37%)	3 (1.64%)	7 (3.84%)	12 (6.59%)	(0%)	(0%)	(0%)	(0%)	(0%)	(0%)	2 (1.09%)	182 (100%)
L-3 (ATP)																	
L-4 (ATP)																	
R-0 (WTA)	(0%)	1 (0.05%)	(0%)	412 (21.33%)	615 (31.84%)	475 (24.59%)	159 (8.23%)	140 (7.25%)	128 (6.62%)	(0%)	1 (0.05%)	(0%)	(0%)	(0%)	(0%)	(0%)	1931 (100%)
R-1 (WTA)	4 (0.55%)	2 (0.27%)	2 (0.27%)	232 (32.13%)	249 (34.48%)	164 (22.71%)	29 (4.01%)	24 (3.32%)	14 (1.93%)	(0%)	(0%)	(0%)	(0%)	(0%)	(0%)	2 (0.27%)	722 (100%)
R-2 (WTA)	(0%)	(0%)	(0%)	451 (20.26%)	692 (31.08%)	528 (23.71%)	192 (8.62%)	188 (8.44%)	175 (7.86%)	(0%)	(0%)	(0%)	(0%)	(0%)	(0%)	(0%)	2226 (100%)
R-3 (WTA)	(0%)	1 (0.01%)	(0%)	932 (18.63%)	1813 (36.25%)	735 (14.69%)	332 (6.63%)	558 (11.15%)	627 (12.53%)	1 (0.01%)	1 (0.01%)	(0%)	(0%)	(0%)	(0%)	1 (0.01%)	5001 (100%)
L-0 (WTA)																	
L-1 (WTA)																	
L-2 (WTA)	(0%)	(0%)	(0%)	6 (10.16%)	22 (37.28%)	12 (20.33%)	6 (10.16%)	8 (13.55%)	5 (8.47%)	(0%)	(0%)	(0%)	(0%)	(0%)	(0%)	(0%)	59 (100%)
Grand Total	13 (0.07%)	24 (0.14%)	12 (0.07%)	3609 (21.7%)	5424 (32.62%)	2919 (17.55%)	1573 (9.46%)	1687 (10.14%)	1340 (8.06%)	4 (0.02%)	11 (0.06%)	1 (0%)	0 (0%)	1 (0%)	0 (0%)	6 (0.03%)	16624 (100%)

Table 14: Shots Chains (Ground Stoke only) hit from court segment by Player Type Cluster (Percentage of shot chains hit by Player Type Cluster in court segment)

Sequential Market Basket Analysis

Of the 47 matches with 61 unique players in the dataset - 40,814 ground stroke shots were observed, 11,845 (29%) shots were hit from segment 5; 8,567 (21%) from Segment 4; and 6,765 (17%) from Segment 6. The most shots captured in the dataset were from R1(ATP) player cluster (12,814 shots) and R3(WTA) player cluster (10,639 shots) (See Table 13, Table 1615 and Figure 30).

A total of 16,624 shot chain with ground strokes only were observed, with 9,939 from WTA players, and 6,685 from ATP players. The majority of shot chains hit from Segment 5 (5,424 – 32.6%), Segment 4 (3,609 -21.7%) and Segment 6 (2,919 – 17.6%). The R1(ATP) player cluster (4,261 shot chains) and R3(WTA) player cluster (5,001 shot chains) had the most observed ground stroke only shot chains (See Table 14 and Table 16).

	Male (ATP)	**Female (WTA)**	**Total**
Matches in Data set	22	25	47
Unique Players	31	30	61
Ground Strokes Observed	20,218	20,596	40,814
Ground Stroke Only Shot chains	6,685	9,939	16,624

Table 16: High level statistics of the data set used in the project

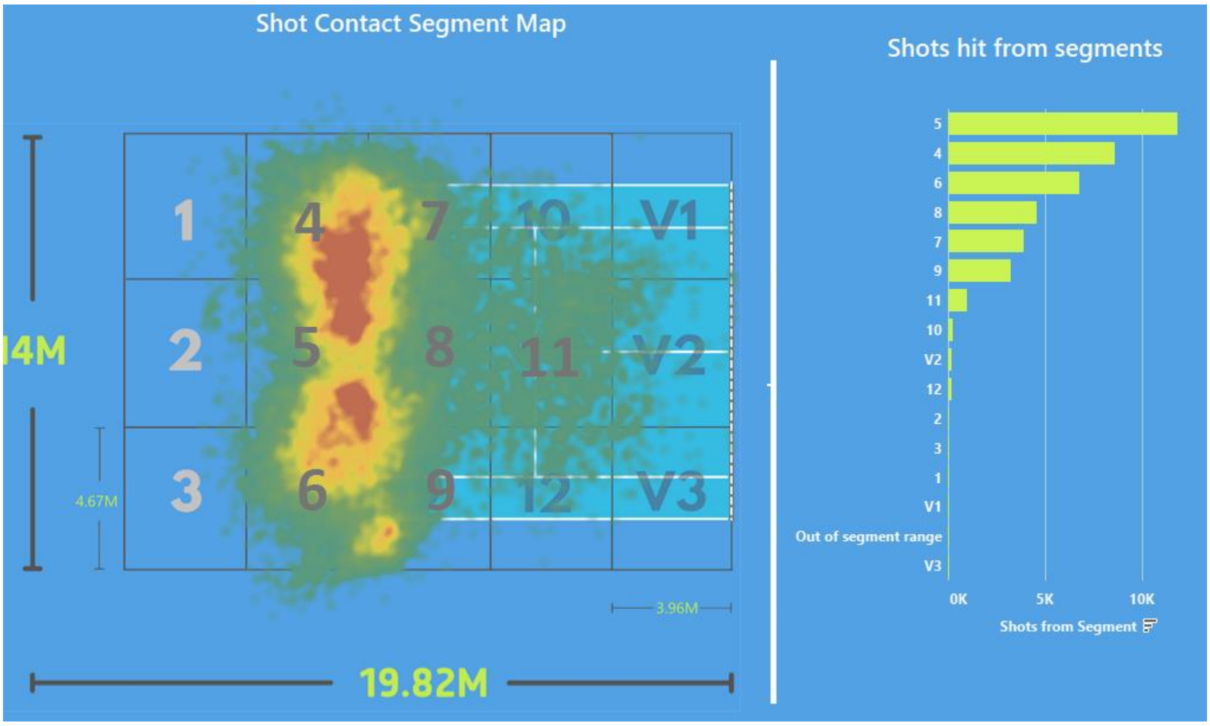


Figure 29: Heat map of Ground Stroke Shots from segments- Dashboard developed for tennis coaches and analysts

The results presented in Tables 18 to 41 indicate there are differences in shot sequences used and found success by different player types from different court positions. Results by segment are presented next.

Result Summary

Table 17 provides a high-level result summary of the six player clusters that had over 500 shot sequences in aggregate meeting suitability thresholds. The table shows most used, most successful, hidden gem (selected as most surprising by domain expert) and band aid sequences (best sequence from a bad segment for the player types).

Player Type	Most Used Shot Sequence	Most Successful Shot Sequence	Hidden Gem Sequence	Band aid Sequence
R-1(ATP) [4086 Sequences]	Backhand Middle Cross → Forehand Middle Cross (Segment 4)	Forehand Inside-In → Forehand Line (Segment 8)	Forehand Inside-In → Forehand Line (Segment 8)	Backhand Middle Cross → Forehand Middle Cross (Segment 4)
R-3 (WTA) [2050 Sequences]	Backhand Middle→ Backhand Cross (Segment 5)	Forehand Line → Forehand Cross (Segment 9)	Backhand Middle → Backhand Line (Segment 8)	Backhand Cross → Backhand Line (Segment 4)
R-2 (WTA) [892 Sequences]	Forehand Middle Cross → Backhand Cross (Segment 6)	Forehand Line → Backhand Cross (Segment 9)	Forehand Line → Backhand Line (Segment 4)	Backhand Middle → Forehand Middle Cross (Segment 5)
R-1 (WTA) [682 Sequences]	Backhand Middle→ Forehand Middle Cross (Segment 5)	Backhand Cross→ Backhand Cross (Segment 7)	Backhand Line→ Forehand Cross (Segment 4)	Forehand Cross→ Forehand Line (Segment 6)
R-0 (WTA) [590 Sequences]	Backhand Middle Cross → Forehand Middle Cross (Segment 4)	Forehand Cross → Forehand Line (Segment 8)	Backhand Line→ Forehand Line (Segment 4)	Backhand Middle → Forehand Middle Cross (Segment 5)
R-4 (ATP) [560 Sequences]	Backhand Cross→ Backhand Cross (Segment 7)	Backhand Cross → Forehand Inside In (Segment 4)	Backhand Cross → Forehand Inside In (Segment 4)	Backhand Cross → Forehand Inside In (Segment 4)

Table 17: High Level summary of shot sequences for the most observed player types

Court Segment 5

Table 18 shows the shot chains that met the minimum suitability evaluation metrics thresholds set for the project. Table 19 evaluates the success of these shot chains against the project's success evaluation metrics. Table 20 shows the shot chains that have a positive per point outcome weighting by player cluster.

The most utilised shot sequence (56 times) was the Backhand middle → Backhand Cross by players in R-3(WTA) with a weighted per point outcome of -0.45. When used it had a 35.71%-point success, none of that success was from points being won immediately when the sequence was used.

Additionally, when used, the point was lost immediately 32.14% of the time. This sequence performed worse against both baseline comparisons (Table 19).

The Forehand Line → Forehand Inside Out hit by players in R-3(ATP) had the highest per point weighted outcome. This chain had a lift score of 2.74 (274% more likely than random) and a Zhang score of 0.75 indicating high associative properties between antecedent and consequent shots.

This shot sequence was executed three times by two unique players, with the maximum contribution by a single player being 66.7% of all these sequences. The average ranking of the executed shot chain being 65.0 (Table 21). Table 20 indicates only four player clusters had shot chains that generated a positive weighted per point outcome (WPPO) shot sequence.

Equivalent Tables in sequential order are presented for each of the following court segments.

Court Segment 4

The most used shot sequence in this segment was the Backhand Middle Cross → Forehand Middle Cross – 66 times by the R-1(ATP) player cluster (Table 22). This sequence had a negative WPPO and performed worse against both baseline comparisons (Table 23). Nine shot sequences had a positive WPPO (Table 20 and Table 21). The Backhand Line → Forehand Line sequence was successful for both the R-0(WTA) and R-2(WTA) player clusters, with both performing at least 10% above all respective baseline comparisons (Table 23). In the R-0(WTA) cluster, 4 unique players hit this shot sequence with the maximum contribution by a single player 74.07%, while in the R-2(WTA) cluster 6 unique player hit this sequence with the maximum contribution of 35.71%. When comparing the median rankings of both clusters to the average ranking of the executed shot chain – the R-0(WTA) was in line with the clusters median ranking, whereas R- 2(WTA) was considerably higher (See Table 7 vs Table 25).

Court Segment 6

The Forehand Middle Cross →Backhand Cross sequence was used by players:

- 41 times in R-0(WTA)
- 29 times in R-2(WTA)
- 17 times in R-0(WTA)
- 13 times in R-1(WTA)

For both R-0(WTA) and R-2(WTA) this sequence had a positive WPPO (Table 27 and Table 28)

The Forehand Cross → Forehand Line was a shot sequence that four player clusters had positive WPPO from this court segment (Table 28 and Table 29).

The R-0(ATP) cluster have three shot sequences with a positive WPPO, the most of all the player clusters (Table 28).

Court Segment 8

The Forehand Cross → Forehand Line had 4 unique players using it from R-2(WTA) (see Table 33). The sequence was used on 8 occasions with a high lift of 3.06 and Zhang metric of 0.78 (Table 30) and outperformed the two baseline comparisons (Table 31). The R-2(WTA) cluster have five shot sequences with a positive WPPO, the most of all the player clusters (Table 32).

Court Segment 7

The 20 Backhand Cross → Backhand Cross shot sequences hit by R-3(WTA) player cluster is the most utilised from this segment (Table 34) and has a positive WPPO (Table 35). Table 34 shows that it was used by two unique players, with one of the players (Serena Williams) contributing to 94.4% of the sequences. The R-3(WTA) had four shot sequences with a positive WPPO, the most of all the player clusters, with R-0(WTA) having three (Table 36).

Court Segment 9

The Forehand Middle Cross → Backhand Middle shot sequences hit by R-2(WTA) player cluster is the most utilised from this segment (Table 38). It has a positive WPPO and outperforms against the two baselines (Table 39). The sequence is used by four unique players, with one player contributing 80% of shot sequences used (Table 41).

While R-3(WTA) had four shot sequences with a positive WPPO (Table 40), the sequences were only utilised by one unique player – Serena Williams (Table 41).

Sequential Market basket analysis statistics for selected shot sequence pairs from Court Segment 5 (Ordered by Frequency of A & C together)

antecedents	consequents	Player Cluster	frequency A	frequency C	frequency A&C	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhang	P(A-->C)	gain A-->C	P(C-->A)	*gain_C-->A*	*Relation*
S1 - Backhand Middle	S2 - Backhand Cross	R-3 (WTA)	215	146	56	0.227	0.154	0.059	0.260	1.688	0.024	1.144	0.527	0.207	0.339	0.143	1.703	Complements
S1 - Backhand Middle	S2 - Forehand Middle Cross	R-2 (WTA)	83	62	27	0.218	0.163	0.071	0.325	1.999	0.035	1.241	0.639	0.245	0.508	0.091	0.385	Complements
S1 - Backhand Middle	S2 - Forehand Middle Cross	R-0 (WTA)	62	49	22	0.207	0.164	0.074	0.355	2.165	0.040	1.296	0.679	0.262	0.598	0.000	-1.000	A-->C
S1 - Forehand Middle	S2 - Backhand Line	R-0 (WTA)	79	34	15	0.264	0.114	0.050	0.190	1.670	0.020	1.094	0.545	0.160	0.403	0.000	-1.000	A-->C
S1 - Backhand Middle	S2 - Forehand Middle Cross	R-1 (WTA)	44	32	15	0.187	0.136	0.064	0.341	2.504	0.038	1.311	0.739	0.254	0.867	0.000	-1.000	A-->C
S1 - Forehand Inside Out	S2 - Backhand Middle Cross	R-1 (WTA)	19	53	11	0.081	0.226	0.047	0.579	2.567	0.029	1.839	0.664	0.367	0.626	0.000	-1.000	A-->C
S1 - Backhand Middle	S2 - Forehand Cross	R-5 (ATP)	38	25	10	0.222	0.146	0.058	0.263	1.800	0.026	1.159	0.571	0.208	0.425	0.077	2.288	Complements
S1 - Backhand Middle	S2 - Forehand Middle Cross	R-5 (ATP)	38	22	9	0.222	0.129	0.053	0.237	1.841	0.024	1.142	0.587	0.191	0.488	0.000	-1.000	A-->C
S1 - Forehand Middle	S2 - Forehand Middle Cross	R-0 (ATP)	31	15	8	0.237	0.115	0.061	0.258	2.254	0.034	1.193	0.729	0.205	0.791	0.000	-1.000	A-->C
S1 - Backhand Middle	S2 - Backhand Cross	R-0 (ATP)	22	22	8	0.168	0.168	0.061	0.364	2.165	0.033	1.308	0.647	0.267	0.588	0.077	1.519	Complements
S1 - Forehand Middle	S2 - Forehand Cross	R-2 (ATP)	22	19	7	0.224	0.194	0.071	0.318	1.641	0.028	1.182	0.504	0.241	0.245	0.000	-1.000	A-->C
S1 - Forehand Cross	S2 - Forehand Cross	L-2 (ATP)	12	10	6	0.200	0.167	0.100	0.500	3.000	0.067	1.667	0.833	0.333	1.000	0.333	1.000	Complements
S1 - Forehand Line	S2 - Backhand Cross	R-5 (ATP)	20	22	6	0.117	0.129	0.035	0.300	2.332	0.020	1.245	0.647	0.231	0.794	0.000	-1.000	A-->C
S1 - Backhand Cross	S2 - Backhand Line	L-2 (ATP)	11	14	6	0.183	0.233	0.100	0.545	2.338	0.057	1.687	0.701	0.353	0.513	0.000	-1.000	A-->C
S1 - Forehand Middle	S2 - Forehand Middle Cross	R-3 (ATP)	13	14	6	0.157	0.169	0.072	0.462	2.736	0.046	1.544	0.752	0.316	0.872	0.000	-1.000	A-->C
S1 - Forehand Inside In	S2 - Forehand Line	R-0 (ATP)	11	21	6	0.084	0.160	0.046	0.545	3.403	0.032	1.847	0.771	0.353	1.202	0.077	0.679	Complements
S1 - Forehand Inside Out	S2 - Forehand Inside Out	R-0 (ATP)	21	12	5	0.160	0.092	0.038	0.238	2.599	0.023	1.192	0.733	0.192	1.099	0.192	1.099	Complements
S1 - Backhand Middle	S2 - Forehand Line	R-2 (ATP)	21	10	5	0.214	0.102	0.051	0.238	2.333	0.029	1.179	0.727	0.192	0.885	0.100	0.089	Complements
S1 - Backhand Cross	S2 - Backhand Middle Cross	R-3 (ATP)	10	13	4	0.120	0.157	0.048	0.400	2.554	0.029	1.406	0.692	0.286	0.824	0.000	-1.000	A-->C
S1 - Forehand Line	S2 - Forehand Inside Out	R-3 (ATP)	13	7	3	0.157	0.084	0.036	0.231	2.736	0.023	1.190	0.752	0.188	1.223	0.125	-0.308	A-->C
S1 - Forehand Inside In	S2 - Backhand Middle Cross	R-2 (ATP)	5	14	3	0.051	0.143	0.031	0.600	4.200	0.023	2.143	0.803	0.375	1.625	0.000	-1.000	A-->C
S1 - Backhand Cross	S2 - Forehand Middle	R-2 (ATP)	10	4	3	0.102	0.041	0.031	0.300	7.350	0.026	1.370	0.962	0.231	4.654	0.083	-0.417	A-->C
S1 - Backhand Line	S2 - Backhand Line	L-0 (ATP)	9	5	3	0.196	0.109	0.065	0.333	3.067	0.044	1.337	0.838	0.250	1.300	0.250	1.300	Complements

Table 18: Key Sequential Market Basket Evaluation Statistics for Court Segment 5 Shot Sequences. Support, Confidence, Lift, Leverage, Conviction, Zhang and Relation thresholds applied

Sequential Market basket analysis - shot sequence success evaluation in Court Segment 5 (Ordered by Weighted Per Point Outcome)									
Shot Sequence	Player Cluster	Point Win %	Point Won Immediately %	Point Won Eventually %	Point Lost Immediately %	High Level Chain (HLC) Comparator	Baseline Comparison 1: Difference to HLC	Baseline Comparison 2: Difference to Any Two Shots	Weighted Per Point Outcome
S1 - Forehand Line-->S2 - Forehand Inside Out	R-3 (ATP)	66.67%	33.33%	33.33%	33.33%	Two Forehands	1.96%	26.91%	0.33
S1 - Forehand Middle-->S2 - Backhand Line	R-0 (WTA)	60.00%	0.00%	60.00%	6.67%	Forehand THEN Backhand	16.72%	7.83%	0.17
S1 - Forehand Inside Out-->S2 - Forehand Inside Out	R-0 (ATP)	60.00%	0.00%	60.00%	20.00%	Two Forehands	4.44%	21.07%	0.10
S1 - Forehand Middle-->S2 - Forehand Cross	R-2 (ATP)	57.14%	0.00%	57.14%	14.29%	Two Forehands	28.57%	23.47%	0.07
S1 - Forehand Cross-->S2 - Forehand Cross	L-2 (ATP)	50.00%	0.00%	50.00%	16.67%	Two Forehands	0.00%	3.33%	-0.08
S1 - Backhand Middle-->S2 - Forehand Middle Cross	R-5 (ATP)	44.44%	0.00%	44.44%	11.11%	Backhand THEN Forehand	17.62%	7.60%	-0.17
S1 - Backhand Middle-->S2 - Forehand Middle Cross	R-0 (WTA)	50.00%	0.00%	50.00%	36.36%	Backhand THEN Forehand	6.72%	-2.17%	-0.18
S1 - Backhand Middle-->S2 - Forehand Middle Cross	R-2 (WTA)	44.44%	0.00%	44.44%	37.04%	Backhand THEN Forehand	5.35%	-3.06%	-0.30
S1 - Forehand Inside Out-->S2 - Backhand Middle Cross	R-1 (WTA)	36.36%	0.00%	36.36%	9.09%	Forehand THEN Backhand	-5.30%	-8.74%	-0.32
S1 - Forehand Line-->S2 - Backhand Cross	R-5 (ATP)	33.33%	0.00%	33.33%	0.00%	Forehand THEN Backhand	6.50%	-3.51%	-0.33
S1 - Forehand Inside In-->S2 - Backhand Middle Cross	R-2 (ATP)	33.33%	0.00%	33.33%	0.00%	Forehand THEN Backhand	4.76%	-0.34%	-0.33
S1 - Backhand Cross-->S2 - Forehand Middle	R-2 (ATP)	33.33%	0.00%	33.33%	0.00%	Backhand THEN Forehand	4.76%	-0.34%	-0.33
S1 - Backhand Cross-->S2 - Backhand Line	L-2 (ATP)	33.33%	0.00%	33.33%	16.67%	Two Backhands	-16.67%	-13.33%	-0.42
S1 - Forehand Middle-->S2 - Forehand Middle Cross	R-3 (ATP)	33.33%	0.00%	33.33%	16.67%	Two Forehands	-31.37%	-6.43%	-0.42
S1 - Backhand Middle-->S2 - Backhand Cross	R-3 (WTA)	35.71%	0.00%	35.71%	32.14%	Two Backhands	-18.17%	-14.18%	-0.45
S1 - Forehand Middle-->S2 - Forehand Middle Cross	R-0 (ATP)	25.00%	0.00%	25.00%	25.00%	Two Forehands	-30.56%	-13.93%	-0.63
S1 - Backhand Line-->S2 - Backhand Line	L-0 (ATP)	33.33%	0.00%	33.33%	66.67%	Two Backhands	4.76%	2.90%	-0.67
S1 - Backhand Middle-->S2 - Forehand Middle Cross	R-1 (WTA)	20.00%	0.00%	20.00%	20.00%	Backhand THEN Forehand	-21.67%	-25.11%	-0.70
S1 - Backhand Middle-->S2 - Forehand Cross	R-5 (ATP)	20.00%	0.00%	20.00%	50.00%	Backhand THEN Forehand	-6.83%	-16.84%	-0.85
S1 - Backhand Middle-->S2 - Backhand Cross	R-0 (ATP)	12.50%	0.00%	12.50%	25.00%	Two Backhands	-43.06%	-26.43%	-0.88
S1 - Forehand Inside In-->S2 - Forehand Line	R-0 (ATP)	16.67%	0.00%	16.67%	50.00%	Two Forehands	-38.89%	-22.26%	-0.92
S1 - Backhand Middle-->S2 - Forehand Line	R-2 (ATP)	0.00%	0.00%	0.00%	20.00%	Backhand THEN Forehand	-28.57%	-33.67%	-1.10
S1 - Backhand Cross-->S2 - Backhand Middle Cross	R-3 (ATP)	0.00%	0.00%	0.00%	25.00%	Two Backhands	-64.71%	-39.76%	-1.13

Table 19: Shot Sequence success evaluation for Court Segment 5. Support, Confidence, Lift, Leverage, Conviction, Zhang and Relation thresholds applied

No. of Positive Chains	Shot Sequences With Positive Weighted Per Point Outcome from Court Segment 5			
	R-0 (ATP)	R-2 (ATP)	R-3 (ATP)	R-0 (WTA)
1	S1 - Forehand Inside Out-->S2 - Forehand Inside Out	S1 - Forehand Middle-->S2 - Forehand Cross	S1 - Forehand Line-->S2 - Forehand Inside Out	S1 - Forehand Middle-->S2 - Backhand Line

Table 20: Shot Sequences with a positive weighted per point outcome from Court Segment 5 by Player Cluster Groups

Sequential Market basket analysis - Unique player contribution and Average ranking of executed shot sequences				
Shot Sequence	Player Cluster	No of Unique Players Hitting Chain	Max Percentage Contribution by a single player to Shot	Average Ranking of Executed shot chain
S1 - Forehand Line-->S2 - Forehand Inside Out	R-3 (ATP)	2	66.67%	65.00
S1 - Forehand Middle-->S2 - Backhand Line	R-0 (WTA)	1	100.00%	7.00
S1 - Forehand Inside Out-->S2 - Forehand Inside Out	R-0 (ATP)	3	40.00%	87.40
S1 - Forehand Middle-->S2 - Forehand Cross	R-2 (ATP)	3	57.14%	94.29

Table 21: Positive Weighted Shot Sequence - Unique player contribution, maximum contribution percentage by a player and average ranking of executed shot sequence from Court Segment 5

Sequential Market basket analysis statistics for selected shot sequence pairs from Court Segment 4 (Ordered by Frequency of A & C together)

antecedents	consequents	Player Cluster	frequency A	frequency C	frequency A&C	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhang	P(A-->C)	gain A-->C	P(C-->A)	*gain_C-->A*	*Relation*
S1 - Backhand Middle Cross	S2 - Forehand Middle Cross	R-1 (ATP)	409	106	66	0.416	0.108	0.067	0.161	1.495	0.022	1.064	0.567	0.139	0.287	0.000	0.000	A-->C
S1 - Backhand Cross	S2 - Backhand Line	R-3 (WTA)	196	78	53	0.358	0.143	0.097	0.270	1.896	0.046	1.175	0.737	0.213	0.493	0.047	-0.591	A-->C
S1 - Backhand Line	S2 - Forehand Line	R-2 (WTA)	44	47	28	0.180	0.192	0.114	0.636	3.317	0.080	2.222	0.851	0.389	1.027	0.000	0.000	A-->C
S1 - Backhand Middle Cross	S2 - Forehand Middle Cross	R-0 (WTA)	100	36	28	0.503	0.181	0.141	0.280	1.548	0.050	1.138	0.711	0.219	0.209	0.000	0.000	A-->C
S1 - Backhand Line	S2 - Forehand Line	R-0 (WTA)	33	43	27	0.166	0.216	0.136	0.818	3.786	0.100	4.312	0.882	0.450	1.083	0.000	0.000	A-->C
S1 - Backhand Line	S2 - Forehand Line	R-3 (WTA)	81	57	25	0.148	0.104	0.046	0.309	2.962	0.030	1.296	0.778	0.236	1.263	0.000	0.000	A-->C
S1 - Backhand Middle Cross	S2 - Forehand Middle Cross	R-2 (WTA)	110	35	24	0.449	0.143	0.098	0.218	1.527	0.034	1.096	0.627	0.179	0.254	0.000	0.000	A-->C
S1 - Backhand Cross	S2 - Backhand Line	R-2 (WTA)	86	32	17	0.351	0.131	0.069	0.198	1.513	0.024	1.084	0.523	0.165	0.264	0.000	-1.000	A-->C
S1 - Backhand Middle Cross	S2 - Forehand Cross	R-0 (ATP)	55	18	13	0.505	0.165	0.119	0.236	1.431	0.036	1.093	0.608	0.191	0.158	0.000	0.000	A-->C
S1 - Backhand Cross	S2 - Backhand Cross	R-0 (WTA)	65	18	10	0.327	0.090	0.050	0.154	1.701	0.021	1.075	0.612	0.133	0.474	0.133	0.474	Complements
S1 - Backhand Middle Cross	S2 - Forehand Middle Cross	R-5 (ATP)	40	14	10	0.426	0.149	0.106	0.250	1.679	0.043	1.135	0.704	0.200	0.343	0.000	0.000	A-->C
S1 - Backhand Line	S2 - Forehand Cross	R-1 (WTA)	31	29	9	0.138	0.129	0.040	0.290	2.242	0.022	1.227	0.643	0.225	0.738	0.000	0.000	A-->C
S1 - Backhand Cross	S2 - Backhand Line	R-1 (WTA)	49	19	9	0.219	0.085	0.040	0.184	2.165	0.022	1.121	0.689	0.155	0.829	0.031	-0.720	A-->C
S1 - Backhand Middle Cross	S2 - Forehand Middle Cross	R-2 (ATP)	26	9	9	0.406	0.141	0.141	0.346	2.462	0.083	1.314	1.000	0.257	0.829	0.000	0.000	A-->C
S1 - Backhand Cross	S2 - Forehand Inside Out	R-5 (ATP)	39	10	8	0.415	0.106	0.085	0.205	1.928	0.041	1.124	0.823	0.170	0.600	0.200	0.175	Complements
S1 - Backhand Cross	S2 - Backhand Line	R-2 (ATP)	24	10	8	0.375	0.156	0.125	0.333	2.133	0.066	1.266	0.850	0.250	0.600	0.100	-0.086	A-->C
S1 - Backhand Cross	S2 - Backhand Line	R-0 (WTA)	65	9	8	0.327	0.045	0.040	0.123	2.721	0.025	1.089	0.939	0.110	1.423	0.029	-0.675	A-->C
S1 - Backhand Cross	S2 - Forehand Inside In	R-4 (ATP)	61	9	7	0.377	0.056	0.043	0.115	2.066	0.022	1.067	0.827	0.103	0.853	0.000	-1.000	A-->C
S1 - Backhand Cross	S2 - Forehand Inside In	R-5 (ATP)	39	9	6	0.415	0.096	0.064	0.154	1.607	0.024	1.069	0.645	0.133	0.393	0.000	-1.000	A-->C
S1 - Backhand Middle Cross	S2 - Backhand Cross	R-3 (ATP)	23	6	5	0.479	0.125	0.104	0.217	1.739	0.044	1.118	0.816	0.179	0.429	0.211	0.263	Complements
S1 - Backhand Cross	S2 - Forehand Middle Line	R-0 (ATP)	32	5	4	0.294	0.046	0.037	0.125	2.725	0.023	1.090	0.896	0.111	1.422	0.167	0.298	Complements
S1 - Backhand Cross	S2 - Backhand Middle Cross	R-3 (ATP)	15	8	4	0.313	0.167	0.083	0.267	1.600	0.031	1.136	0.545	0.211	0.263	0.179	0.429	Complements
S1 - Backhand Cross	S2 - Forehand Middle	R-3 (ATP)	15	3	3	0.313	0.063	0.063	0.200	3.200	0.043	1.172	1.000	0.167	1.667	0.000	0.000	A-->C
S1 - Forehand Middle Cross	S2 - Forehand Middle Cross	L-0 (ATP)	13	3	3	0.684	0.158	0.158	0.231	1.462	0.050	1.095	1.000	0.188	0.188	0.188	0.188	Complements

Table 22: Key Sequential Market Basket Evaluation Statistics for Court Segment 4 Shot Sequences. Support, Confidence, Lift, Leverage, Conviction, Zhang and Relation thresholds applied

Sequential Market basket analysis - shot sequence success evaluation in Court Segment 4 (Ordered by Weighted Per Point Outcome)									
Shot Sequence	Player Cluster	Point Win %	Point Won Immediately %	Point Won Eventually %	Point Lost Immediately %	High Level Chain (HLC) Comparator	Baseline Comparison 1: Difference to HLC	Baseline Comparison 2: Difference to Any Two Shots	Weighted Per Point Outcome
S1 - Backhand Cross-->S2 - Forehand Middle	R-3 (ATP)	100.00%	0.00%	100.00%	0.00%	Backhand THEN Forehand	58.33%	62.50%	1.00
S1 - Backhand Line-->S2 - Forehand Cross	R-1 (WTA)	77.78%	33.33%	44.44%	22.22%	Backhand THEN Forehand	35.35%	35.81%	0.61
S1 - Backhand Cross-->S2 - Forehand Inside Out	R-5 (ATP)	75.00%	12.50%	62.50%	0.00%	Backhand THEN Forehand	33.33%	36.70%	0.56
S1 - Backhand Line-->S2 - Forehand Line	R-0 (WTA)	77.78%	18.52%	59.26%	22.22%	Backhand THEN Forehand	23.70%	28.03%	0.54
S1 - Backhand Cross-->S2 - Forehand Inside In	R-4 (ATP)	71.43%	14.29%	57.14%	14.29%	Backhand THEN Forehand	29.66%	26.98%	0.43
S1 - Forehand Middle Cross-->S2 - Forehand Middle Cross	L-0 (ATP)	66.67%	0.00%	66.67%	0.00%	Two Forehands	41.67%	29.82%	0.33
S1 - Backhand Line-->S2 - Forehand Line	R-2 (WTA)	67.86%	14.29%	53.57%	25.00%	Backhand THEN Forehand	10.81%	20.10%	0.30
S1 - Backhand Cross-->S2 - Backhand Line	R-2 (ATP)	50.00%	12.50%	37.50%	12.50%	Two Backhands	22.22%	18.75%	0.00
S1 - Backhand Cross-->S2 - Forehand Middle Line	R-0 (ATP)	50.00%	0.00%	50.00%	0.00%	Backhand THEN Forehand	15.00%	12.39%	0.00
S1 - Backhand Cross-->S2 - Backhand Line	R-3 (WTA)	47.17%	11.32%	35.85%	13.21%	Two Backhands	0.43%	1.47%	-0.07
S1 - Backhand Cross-->S2 - Backhand Cross	R-0 (WTA)	40.00%	10.00%	30.00%	0.00%	Two Backhands	-14.07%	-9.75%	-0.15
S1 - Backhand Middle Cross-->S2 - Forehand Middle Cross	R-2 (WTA)	45.83%	0.00%	45.83%	16.67%	Backhand THEN Forehand	-11.22%	-1.92%	-0.17
S1 - Backhand Middle Cross-->S2 - Forehand Middle Cross	R-1 (ATP)	46.97%	0.00%	46.97%	22.73%	Backhand THEN Forehand	-8.77%	-6.19%	-0.17
S1 - Backhand Middle Cross-->S2 - Forehand Middle Cross	R-0 (WTA)	42.86%	0.00%	42.86%	25.00%	Backhand THEN Forehand	-11.22%	-6.89%	-0.27
S1 - Backhand Middle Cross-->S2 - Forehand Cross	R-0 (ATP)	38.46%	7.69%	30.77%	23.08%	Backhand THEN Forehand	3.46%	0.85%	-0.31
S1 - Backhand Middle Cross-->S2 - Backhand Cross	R-3 (ATP)	40.00%	0.00%	40.00%	40.00%	Two Backhands	-1.67%	2.50%	-0.40
S1 - Backhand Cross-->S2 - Backhand Line	R-1 (WTA)	33.33%	11.11%	22.22%	33.33%	Two Backhands	-9.09%	-8.63%	-0.44
S1 - Backhand Cross-->S2 - Backhand Line	R-0 (WTA)	37.50%	12.50%	25.00%	62.50%	Two Backhands	-16.57%	-12.25%	-0.50
S1 - Backhand Middle Cross-->S2 - Forehand Middle Cross	R-5 (ATP)	30.00%	0.00%	30.00%	50.00%	Backhand THEN Forehand	-11.67%	-8.30%	-0.65
S1 - Backhand Line-->S2 - Forehand Line	R-3 (WTA)	24.00%	4.00%	20.00%	48.00%	Backhand THEN Forehand	-22.74%	-21.70%	-0.74
S1 - Backhand Cross-->S2 - Forehand Inside In	R-5 (ATP)	16.67%	0.00%	16.67%	33.33%	Backhand THEN Forehand	-25.00%	-21.63%	-0.83
S1 - Backhand Cross-->S2 - Backhand Line	R-2 (WTA)	11.76%	11.76%	0.00%	47.06%	Two Backhands	-45.29%	-35.99%	-0.94
S1 - Backhand Middle Cross-->S2 - Forehand Middle Cross	R-2 (ATP)	11.11%	0.00%	11.11%	44.44%	Backhand THEN Forehand	-16.67%	-20.14%	-1.00
S1 - Backhand Cross-->S2 - Backhand Middle Cross	R-3 (ATP)	0.00%	0.00%	0.00%	0.00%	Two Backhands	-41.67%	-37.50%	-1.00

Table 23: Shot Sequence success evaluation for Court Segment 4. Support, Confidence, Lift, Leverage, Conviction, Zhang and Relation thresholds applied

No. of Positive Chains	Shot Sequences With Positive Weighted Per Point Outcome from Court Segment 4								
	R-0 (ATP)	R-2 (ATP)	R-3 (ATP)	R-4 (ATP)	R-5 (ATP)	L-0 (ATP)	R-0 (WTA)	R-1 (WTA)	R-2 (WTA)
1	S1 - Backhand Cross-->S2 - Forehand Middle Line	S1 - Backhand Cross-->S2 - Backhand Line	S1 - Backhand Cross-->S2 - Forehand Middle	S1 - Backhand Cross-->S2 - Forehand Inside In	S1 - Backhand Cross-->S2 - Forehand Inside Out	S1 - Forehand Middle Cross-->S2 - Forehand Middle Cross	S1 - Backhand Line-->S2 - Forehand Line	S1 - Backhand Line-->S2 - Forehand Cross	S1 - Backhand Line-->S2 - Forehand Line

Table 24: Shot Sequences with a positive weighted per point outcome from Court Segment 4 by Player Cluster Groups

Sequential Market basket analysis - Unique player contribution and Average ranking of executed shot sequences

Shot Sequence	Player Cluster	No of Unique Players Hitting Chain	Max Percentage Contribution by a single player to Shot Chain total	Average Ranking of Executed shot chain
S1 - Backhand Cross-->S2 - Forehand Middle	R-3 (ATP)	2	66.67%	59.33
S1 - Backhand Line-->S2 - Forehand Cross	R-1 (WTA)	3	55.56%	38.78
S1 - Backhand Cross-->S2 - Forehand Inside Out	R-5 (ATP)	3	50.00%	21.00
S1 - Backhand Line-->S2 - Forehand Line	R-0 (WTA)	4	74.07%	58.98
S1 - Backhand Cross-->S2 - Forehand Inside In	R-4 (ATP)	3	42.86%	30.29
S1 - Forehand Middle Cross-->S2 - Forehand Middle Cross	L-0 (ATP)	2	75.00%	50.75
S1 - Backhand Line-->S2 - Forehand Line	R-2 (WTA)	6	35.71%	53.87
S1 - Backhand Cross-->S2 - Backhand Line	R-2 (ATP)	4	50.00%	51.77
S1 - Backhand Cross-->S2 - Forehand Middle Line	R-0 (ATP)	4	25.00%	77.43

Table 25: Positive Weighted Shot Sequence- Unique player contribution, maximum contribution percentage by a player and average ranking of executed shot sequence from Court Segment 4

antecedents	consequents	Player Cluster	frequency A	frequency C	frequency A&C	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhang	P(A-->C)	gain A-->C	P(C-->A)	*gain_C-->A*	*Relation*
S1 - Forehand Middle Cross	S2 - Backhand Cross	R-3 (WTA)	130	76	41	0.349	0.204	0.110	0.315	1.548	0.039	1.163	0.543	0.240	0.177	0.000	0.000	A-->C
S1 - Forehand Cross	S2 - Forehand Line	R-1 (ATP)	135	63	33	0.243	0.114	0.059	0.244	2.153	0.032	1.173	0.708	0.196	0.730	0.070	-0.307	A-->C
S1 - Forehand Middle Cross	S2 - Backhand Cross	R-2 (WTA)	111	44	29	0.466	0.185	0.122	0.261	1.413	0.036	1.103	0.548	0.207	0.120	0.000	0.000	A-->C
S1 - Forehand Cross	S2 - Forehand Line	R-3 (WTA)	123	42	26	0.330	0.113	0.070	0.211	1.877	0.033	1.125	0.697	0.174	0.550	0.116	0.004	Complements
S1 - Forehand Line	S2 - Backhand Line	R-3 (WTA)	84	57	21	0.225	0.153	0.056	0.250	1.636	0.022	1.130	0.502	0.200	0.309	0.000	0.000	A-->C
S1 - Forehand Middle Cross	S2 - Backhand Cross	R-0 (WTA)	89	23	17	0.481	0.124	0.092	0.191	1.536	0.032	1.082	0.673	0.160	0.290	0.000	0.000	A-->C
S1 - Forehand Cross	S2 - Forehand Line	R-2 (WTA)	54	25	14	0.227	0.105	0.059	0.259	2.468	0.035	1.208	0.769	0.206	0.960	0.088	-0.222	A-->C
S1 - Forehand Middle Cross	S2 - Backhand Cross	R-1 (WTA)	71	19	13	0.473	0.127	0.087	0.183	1.446	0.027	1.069	0.585	0.155	0.222	0.000	0.000	A-->C
S1 - Forehand Cross	S2 - Forehand Middle Cross	R-2 (WTA)	54	31	12	0.227	0.130	0.050	0.222	1.706	0.021	1.118	0.535	0.182	0.396	0.075	-0.339	A-->C
S1 - Forehand Cross	S2 - Forehand Cross	R-2 (WTA)	54	27	11	0.227	0.113	0.046	0.204	1.796	0.020	1.113	0.573	0.169	0.492	0.169	0.492	Complements
S1 - Forehand Line	S2 - Backhand Line	R-0 (ATP)	25	17	10	0.316	0.215	0.127	0.400	1.859	0.058	1.308	0.676	0.286	0.328	0.000	0.000	A-->C
S1 - Forehand Cross	S2 - Forehand Line	R-0 (WTA)	32	25	10	0.173	0.135	0.054	0.313	2.312	0.031	1.258	0.686	0.238	0.762	0.098	0.134	Complements
S1 - Forehand Line	S2 - Backhand Middle Cross	R-0 (ATP)	25	15	8	0.316	0.190	0.101	0.320	1.685	0.041	1.191	0.595	0.242	0.277	0.000	0.000	A-->C
S1 - Forehand Cross	S2 - Forehand Line	R-1 (WTA)	20	24	8	0.133	0.160	0.053	0.400	2.500	0.032	1.400	0.692	0.286	0.786	0.059	-0.198	A-->C
S1 - Forehand Cross	S2 - Forehand Middle Cross	R-0 (WTA)	32	23	8	0.173	0.124	0.043	0.250	2.011	0.022	1.168	0.608	0.200	0.609	0.053	-0.385	A-->C
S1 - Forehand Line	S2 - Forehand Middle	R-0 (WTA)	46	9	7	0.249	0.049	0.038	0.152	3.128	0.026	1.122	0.905	0.132	1.715	0.000	0.000	A-->C
S1 - Forehand Middle Cross	S2 - Forehand Middle Cross	R-4 (ATP)	40	8	6	0.482	0.096	0.072	0.150	1.556	0.026	1.063	0.690	0.130	0.353	0.130	0.353	Complements
S1 - Forehand Line	S2 - Forehand Inside Out	R-1 (WTA)	48	6	5	0.320	0.040	0.033	0.104	2.604	0.021	1.072	0.906	0.094	1.358	0.000	0.000	A-->C
S1 - Forehand Middle Cross	S2 - Forehand Cross	R-5 (ATP)	24	8	5	0.387	0.129	0.081	0.208	1.615	0.031	1.100	0.621	0.172	0.336	0.000	-1.000	A-->C
S1 - Backhand Middle Cross	S2 - Backhand Middle Cross	L-1 (ATP)	13	7	5	0.448	0.241	0.172	0.385	1.593	0.064	1.233	0.675	0.278	0.151	0.278	0.151	Complements
S1 - Forehand Line	S2 - Backhand Middle Cross	R-2 (ATP)	16	7	5	0.340	0.149	0.106	0.313	2.098	0.056	1.238	0.794	0.238	0.599	0.000	0.000	A-->C
S1 - Forehand Middle Cross	S2 - Forehand Middle Cross	R-0 (ATP)	32	6	5	0.405	0.076	0.063	0.156	2.057	0.033	1.095	0.864	0.135	0.779	0.135	0.779	Complements
S1 - Forehand Cross	S2 - Forehand Cross	R-0 (ATP)	21	6	4	0.266	0.076	0.051	0.190	2.508	0.030	1.141	0.819	0.160	1.107	0.160	1.107	Complements
S1 - Forehand Line	S2 - Backhand Line	R-5 (ATP)	23	4	4	0.371	0.065	0.065	0.174	2.696	0.041	1.132	1.000	0.148	1.296	0.000	0.000	A-->C
S1 - Backhand Cross	S2 - Forehand Cross	L-2 (ATP)	9	12	4	0.196	0.261	0.087	0.444	1.704	0.036	1.330	0.514	0.308	0.179	0.000	0.000	A-->C
S1 - Backhand Middle Cross	S2 - Backhand Line	L-2 (ATP)	20	5	4	0.435	0.109	0.087	0.200	1.840	0.040	1.114	0.808	0.167	0.533	0.125	-0.425	A-->C
S1 - Forehand Cross	S2 - Forehand Line	R-0 (ATP)	21	6	4	0.266	0.076	0.051	0.190	2.508	0.030	1.141	0.819	0.160	1.107	0.000	-1.000	A-->C
S1 - Forehand Cross	S2 - Backhand Middle	R-5 (ATP)	14	7	4	0.226	0.113	0.065	0.286	2.531	0.039	1.242	0.781	0.222	0.968	0.000	0.000	A-->C
S1 - Forehand Middle Cross	S2 - Forehand Middle Cross	R-5 (ATP)	24	4	3	0.387	0.065	0.048	0.125	1.938	0.023	1.069	0.789	0.111	0.722	0.111	0.722	Complements
S1 - Forehand Line	S2 - Forehand Inside Out	R-5 (ATP)	23	3	3	0.371	0.048	0.048	0.130	2.696	0.030	1.094	1.000	0.115	1.385	0.000	0.000	A-->C
S1 - Backhand Middle Cross	S2 - Forehand Middle Line	L-1 (ATP)	13	3	3	0.448	0.103	0.103	0.231	2.231	0.057	1.166	1.000	0.188	0.813	0.000	-1.000	A-->C
S1 - Backhand Line	S2 - Forehand Cross	L-0 (ATP)	8	7	3	0.267	0.233	0.100	0.375	1.607	0.038	1.227	0.515	0.273	0.169	0.000	0.000	A-->C
S1 - Forehand Line	S2 - Forehand Inside Out	R-4 (ATP)	19	4	3	0.229	0.048	0.036	0.158	3.276	0.025	1.130	0.901	0.136	1.830	0.000	0.000	A-->C
S1 - Backhand Line	S2 - Forehand Cross	L-1 (ATP)	7	3	3	0.241	0.103	0.103	0.429	4.143	0.078	1.569	1.000	0.300	1.900	0.000	0.000	A-->C
S1 - Forehand Cross	S2 - Forehand Middle Cross	R-2 (ATP)	16	5	3	0.340	0.106	0.064	0.188	1.763	0.028	1.100	0.656	0.158	0.484	0.118	0.382	Complements

Table 26: Key Sequential Market Basket Evaluation Statistics for Court Segment 6 Shot Sequences. Support, Confidence, Lift, Leverage, Conviction, Zhang and Relation thresholds applied

Sequential Market basket analysis - shot sequence success evaluation in Court Segment 6 (Ordered by Weighted Per Point Outcome)									
Shot Sequence	Player Cluster	Point Win %	Point Won Immediately %	Point Won Eventually %	Point Lost Immediately %	High Level Chain (HLC) Comparator	Baseline Comparison 1: Difference to HLC	Baseline Comparison 2: Difference to Any Two Shots	Weighted Per Point Outcome
S1 - Forehand Middle Cross-->S2 - Backhand Cross	R-0 (WTA)	94.12%	23.53%	70.59%	5.88%	Forehand THEN Backhand	69.12%	50.33%	0.97
S1 - Forehand Cross-->S2 - Forehand Line	R-1 (ATP)	60.61%	24.24%	36.36%	15.15%	Two Forehands	24.24%	9.98%	0.26
S1 - Forehand Line-->S2 - Backhand Middle Cross	R-0 (ATP)	62.50%	0.00%	62.50%	0.00%	Forehand THEN Backhand	62.50%	21.99%	0.25
S1 - Forehand Cross-->S2 - Forehand Line	R-1 (WTA)	62.50%	12.50%	50.00%	12.50%	Two Forehands	5.36%	25.17%	0.25
S1 - Forehand Cross-->S2 - Forehand Cross	R-0 (ATP)	50.00%	50.00%	0.00%	0.00%	Two Forehands	50.00%	9.49%	0.25
S1 - Forehand Middle Cross-->S2 - Backhand Cross	R-2 (WTA)	62.07%	17.24%	44.83%	24.14%	Forehand THEN Backhand	12.07%	17.11%	0.21
S1 - Forehand Cross-->S2 - Forehand Line	R-3 (WTA)	53.85%	11.54%	42.31%	15.38%	Two Forehands	41.35%	12.02%	0.06
S1 - Forehand Line-->S2 - Backhand Line	R-5 (ATP)	50.00%	0.00%	50.00%	0.00%	Forehand THEN Backhand	50.00%	24.19%	0.00
S1 - Backhand Cross-->S2 - Forehand Cross	L-2 (ATP)	50.00%	0.00%	50.00%	0.00%	Backhand THEN Forehand	10.00%	4.35%	0.00
S1 - Backhand Middle Cross-->S2 - Backhand Line	L-2 (ATP)	50.00%	0.00%	50.00%	0.00%	Two Backhands	10.00%	4.35%	0.00
S1 - Forehand Cross-->S2 - Forehand Line	R-0 (ATP)	50.00%	0.00%	50.00%	0.00%	Two Forehands	50.00%	9.49%	0.00
S1 - Forehand Line-->S2 - Forehand Inside Out	R-1 (WTA)	40.00%	20.00%	20.00%	20.00%	Two Forehands	-17.14%	2.67%	-0.20
S1 - Forehand Line-->S2 - Forehand Middle	R-0 (WTA)	28.57%	28.57%	0.00%	0.00%	Two Forehands	3.57%	-15.21%	-0.29
S1 - Forehand Line-->S2 - Backhand Line	R-0 (ATP)	40.00%	10.00%	30.00%	30.00%	Forehand THEN Backhand	40.00%	-0.51%	-0.30
S1 - Forehand Middle Cross-->S2 - Forehand Cross	R-5 (ATP)	40.00%	0.00%	40.00%	20.00%	Two Forehands	40.00%	14.19%	-0.30
S1 - Forehand Middle Cross-->S2 - Backhand Cross	R-3 (WTA)	39.02%	0.00%	39.02%	17.07%	Forehand THEN Backhand	26.52%	-2.80%	-0.30
S1 - Forehand Middle Cross-->S2 - Forehand Middle Cross	R-5 (ATP)	33.33%	0.00%	33.33%	0.00%	Two Forehands	33.33%	7.53%	-0.33
S1 - Forehand Line-->S2 - Forehand Inside Out	R-5 (ATP)	33.33%	0.00%	33.33%	0.00%	Two Forehands	33.33%	7.53%	-0.33
S1 - Backhand Middle Cross-->S2 - Forehand Middle Line	L-1 (ATP)	33.33%	0.00%	33.33%	0.00%	Backhand THEN Forehand	15.15%	9.20%	-0.33
S1 - Forehand Line-->S2 - Backhand Line	R-3 (WTA)	28.57%	9.52%	19.05%	14.29%	Forehand THEN Backhand	16.07%	-13.25%	-0.45
S1 - Forehand Cross-->S2 - Backhand Middle	R-5 (ATP)	25.00%	0.00%	25.00%	0.00%	Forehand THEN Backhand	25.00%	-0.81%	-0.50
S1 - Forehand Cross-->S2 - Forehand Middle Cross	R-2 (ATP)	33.33%	0.00%	33.33%	33.33%	Two Forehands	1.33%	-2.84%	-0.50
S1 - Forehand Cross-->S2 - Forehand Middle Cross	R-2 (WTA)	33.33%	0.00%	33.33%	41.67%	Two Forehands	-16.67%	-11.62%	-0.54
S1 - Backhand Middle Cross-->S2 - Backhand Middle Cross	L-1 (ATP)	20.00%	0.00%	20.00%	0.00%	Two Backhands	1.82%	-4.14%	-0.60
S1 - Backhand Line-->S2 - Forehand Cross	L-0 (ATP)	33.33%	0.00%	33.33%	66.67%	Backhand THEN Forehand	-2.38%	-16.67%	-0.67
S1 - Forehand Cross-->S2 - Forehand Line	R-2 (WTA)	21.43%	0.00%	21.43%	21.43%	Two Forehands	-28.57%	-23.53%	-0.68
S1 - Forehand Line-->S2 - Backhand Middle Cross	R-2 (ATP)	20.00%	0.00%	20.00%	20.00%	Forehand THEN Backhand	-12.00%	-16.17%	-0.70
S1 - Forehand Middle Cross-->S2 - Forehand Middle Cross	R-4 (ATP)	16.67%	0.00%	16.67%	16.67%	Two Forehands	16.67%	-23.09%	-0.75
S1 - Forehand Middle Cross-->S2 - Backhand Cross	R-1 (WTA)	15.38%	0.00%	15.38%	23.08%	Forehand THEN Backhand	-41.76%	-21.95%	-0.81
S1 - Forehand Cross-->S2 - Forehand Line	R-0 (WTA)	10.00%	0.00%	10.00%	20.00%	Two Forehands	-15.00%	-33.78%	-0.90
S1 - Forehand Cross-->S2 - Forehand Cross	R-2 (WTA)	18.18%	9.09%	9.09%	72.73%	Two Forehands	-31.82%	-26.78%	-0.95
S1 - Forehand Line-->S2 - Forehand Inside Out	R-4 (ATP)	0.00%	0.00%	0.00%	0.00%	Two Forehands	0.00%	-39.76%	-1.00
S1 - Backhand Line-->S2 - Forehand Cross	L-1 (ATP)	0.00%	0.00%	0.00%	0.00%	Backhand THEN Forehand	-18.18%	-24.14%	-1.00
S1 - Forehand Cross-->S2 - Forehand Middle Cross	R-0 (WTA)	12.50%	0.00%	12.50%	75.00%	Two Forehands	-12.50%	-31.28%	-1.13
S1 - Forehand Middle Cross-->S2 - Forehand Middle Cross	R-0 (ATP)	0.00%	0.00%	0.00%	40.00%	Two Forehands	0.00%	-40.51%	-1.20

Table 27: Shot Sequence success evaluation for Court Segment 6. Support, Confidence, Lift, Leverage, Conviction, Zhang and Relation thresholds applied

No. of Positive Chains	Shot Sequences With Positive Weighted Per Point Outcome from Court Segment 6							
	R-0 (ATP)	R-1 (ATP)	R-5 (ATP)	L-2 (ATP)	R-0 (WTA)	R-1 (WTA)	R-2 (WTA)	R-3 (WTA)
1	S1 - Forehand Line-->S2 - Backhand Middle Cross	S1 - Forehand Cross-->S2 - Forehand Line	S1 - Forehand Line-->S2 - Backhand Line	S1 - Backhand Cross-->S2 - Forehand Cross	S1 - Forehand Middle Cross-->S2 - Backhand Cross	S1 - Forehand Cross-->S2 - Forehand Line	S1 - Forehand Middle Cross-->S2 - Backhand Cross	S1 - Forehand Cross-->S2 - Forehand Line
2	S1 - Forehand Cross-->S2 - Forehand Cross			S1 - Backhand Middle Cross-->S2 - Backhand				
3	S1 - Forehand Cross-->S2 - Forehand Line							

Table 28: Shot Sequences with a positive weighted per point outcome from Court Segment 6 by Player Cluster Groups

Sequential Market basket analysis - Unique player contribution and Average ranking of executed shot sequences				
Shot Sequence	Player Cluster	No of Unique Players Hitting Chain	Max Percentage Contribution by a single player to Shot Chain total	Average Ranking of Executed shot chain
S1 - Forehand Middle Cross-->S2 - Backhand Cross	R-0 (WTA)	4	70.59%	73.41
S1 - Forehand Cross-->S2 - Forehand Line	R-1 (ATP)	3	84.85%	5.70
S1 - Forehand Line-->S2 - Backhand Middle Cross	R-0 (ATP)	4	50.00%	61.32
S1 - Forehand Cross-->S2 - Forehand Line	R-1 (WTA)	5	37.50%	11.34
S1 - Forehand Cross-->S2 - Forehand Cross	R-0 (ATP)	2	50.00%	49.87
S1 - Forehand Middle Cross-->S2 - Backhand Cross	R-2 (WTA)	3	50.00%	70.21
S1 - Forehand Cross-->S2 - Forehand Line	R-3 (WTA)	4	80.77%	18.42
S1 - Forehand Line-->S2 - Backhand Line	R-5 (ATP)	2	75.00%	17.54
S1 - Backhand Cross-->S2 - Forehand Cross	L-2 (ATP)	1	100.00%	2.00
S1 - Backhand Middle Cross-->S2 - Backhand Line	L-2 (ATP)	1	100.00%	2.00
S1 - Forehand Cross-->S2 - Forehand Line	R-0 (ATP)	3	50.00%	80.33

Table 29: Positive Weighted Shot Sequence- Unique player contribution, maximum contribution percentage by a player and average ranking of executed shot sequence from Court Segment 6

Sequential Market basket analysis statistics for selected shot sequence pairs from Court Segment 8 (Ordered by Frequency of A & C together)

antecedents	consequents	Player Cluster	frequency A	frequency C	frequency A&C	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhang	P(A-->C)	gain A-->C	P(C-->A)	*gain_C-->A*	*Relation*
S1 - Forehand Inside In	S2 - Forehand Line	R-1 (ATP)	56	73	19	0.111	0.144	0.038	0.339	2.352	0.022	1.295	0.646	0.253	0.756	0.111	0.171	Complements
S1 - Forehand Inside In	S2 - Forehand Line	R-3 (WTA)	31	38	17	0.117	0.143	0.064	0.548	3.824	0.047	1.897	0.836	0.354	1.470	0.000	-1.000	A-->C
S1 - Backhand Middle	S2 - Backhand Line	R-3 (WTA)	45	29	12	0.170	0.109	0.045	0.267	2.437	0.027	1.214	0.710	0.211	0.924	0.000	-1.000	A-->C
S1 - Forehand Line	S2 - Forehand Middle	R-3 (WTA)	21	25	9	0.079	0.094	0.034	0.429	4.543	0.026	1.585	0.847	0.300	2.180	0.065	-0.545	A-->C
S1 - Backhand Line	S2 - Forehand Cross	R-0 (WTA)	9	15	8	0.150	0.250	0.133	0.889	3.556	0.096	6.750	0.846	0.471	0.882	0.000	-1.000	A-->C
S1 - Forehand Middle Line	S2 - Backhand Cross	R-3 (WTA)	11	40	8	0.042	0.151	0.030	0.727	4.818	0.024	3.113	0.827	0.421	1.789	0.000	-1.000	A-->C
S1 - Forehand Cross	S2 - Forehand Line	R-2 (WTA)	14	20	8	0.131	0.187	0.075	0.571	3.057	0.050	1.897	0.774	0.364	0.945	0.158	-0.296	A-->C
S1 - Backhand Line	S2 - Forehand Cross	R-2 (WTA)	15	24	8	0.140	0.224	0.075	0.533	2.378	0.043	1.662	0.674	0.348	0.551	0.067	-0.287	A-->C
S1 - Forehand Cross	S2 - Forehand Line	R-0 (WTA)	6	7	5	0.100	0.117	0.083	0.833	7.143	0.072	5.300	0.956	0.455	2.896	0.000	-1.000	A-->C
S1 - Forehand Line	S2 - Backhand Middle Cross	R-2 (WTA)	16	13	5	0.150	0.121	0.047	0.313	2.572	0.029	1.278	0.719	0.238	0.960	0.000	-1.000	A-->C
S1 - Forehand Line	S2 - Backhand Middle Cross	R-0 (WTA)	12	11	5	0.200	0.183	0.083	0.417	2.273	0.047	1.400	0.700	0.294	0.604	0.167	0.429	Complements
S1 - Forehand Middle	S2 - Backhand Middle Cross	R-4 (ATP)	11	10	5	0.216	0.196	0.098	0.455	2.318	0.056	1.474	0.725	0.313	0.594	0.000	-1.000	A-->C
S1 - Forehand Line	S2 - Backhand Line	R-0 (WTA)	12	5	4	0.200	0.083	0.067	0.333	4.000	0.050	1.375	0.938	0.250	2.000	0.000	-1.000	A-->C
S1 - Forehand Cross	S2 - Forehand Line	R-4 (ATP)	7	7	4	0.137	0.137	0.078	0.571	4.163	0.060	2.013	0.881	0.364	1.649	0.000	-1.000	A-->C
S1 - Backhand Middle Cross	S2 - Forehand Middle	R-0 (WTA)	5	7	4	0.083	0.117	0.067	0.800	6.857	0.057	4.417	0.932	0.444	2.810	0.167	-0.091	A-->C
S1 - Backhand Cross	S2 - Backhand Cross	R-2 (ATP)	8	9	4	0.205	0.231	0.103	0.500	2.167	0.055	1.538	0.677	0.333	0.444	0.333	0.444	Complements
S1 - Forehand Line	S2 - Backhand Line	R-2 (WTA)	16	10	4	0.150	0.093	0.037	0.250	2.675	0.023	1.209	0.736	0.200	1.140	0.167	-0.108	A-->C
S1 - Forehand Middle	S2 - Forehand Middle Cross	R-2 (WTA)	18	10	4	0.168	0.093	0.037	0.222	2.378	0.022	1.166	0.697	0.182	0.945	0.000	-1.000	A-->C
S1 - Forehand Middle	S2 - Backhand Cross	R-2 (WTA)	18	9	4	0.168	0.084	0.037	0.222	2.642	0.023	1.178	0.747	0.182	1.162	0.333	1.972	Complements
S1 - Backhand Middle Cross	S2 - Forehand Middle	R-2 (WTA)	6	12	4	0.056	0.112	0.037	0.667	5.944	0.031	2.664	0.881	0.400	2.567	0.053	-0.567	A-->C
S1 - Forehand Inside In	S2 - Forehand Inside Out	R-5 (ATP)	8	8	3	0.178	0.178	0.067	0.375	2.109	0.035	1.316	0.640	0.273	0.534	0.000	-1.000	A-->C
S1 - Backhand Middle	S2 - Forehand Line	R-2 (ATP)	8	6	3	0.205	0.154	0.077	0.375	2.438	0.045	1.354	0.742	0.273	0.773	0.000	-1.000	A-->C
S1 - Forehand Inside Out	S2 - Forehand Inside Out	R-5 (ATP)	8	8	3	0.178	0.178	0.067	0.375	2.109	0.035	1.316	0.640	0.273	0.534	0.273	0.534	Complements
S1 - Backhand Cross	S2 - Backhand Middle Cross	R-5 (ATP)	8	7	3	0.178	0.156	0.067	0.375	2.411	0.039	1.351	0.712	0.273	0.753	0.333	2.750	Complements
S1 - Forehand Inside In	S2 - Forehand Line	R-5 (ATP)	8	9	3	0.178	0.200	0.067	0.375	1.875	0.031	1.280	0.568	0.273	0.364	0.000	-1.000	A-->C
S1 - Forehand Cross	S2 - Forehand Line	R-0 (ATP)	9	7	3	0.196	0.152	0.065	0.333	2.190	0.035	1.272	0.676	0.250	0.643	0.000	-1.000	A-->C
S1 - Backhand Middle	S2 - Forehand Line	R-0 (ATP)	8	7	3	0.174	0.152	0.065	0.375	2.464	0.039	1.357	0.719	0.273	0.792	0.000	-1.000	A-->C

Table 30: Key Sequential Market Basket Evaluation Statistics for Court Segment 8 Shot Sequences. Support, Confidence, Lift, Leverage, Conviction, Zhang and Relation thresholds applied

Sequential Market basket analysis - shot sequence success evaluation in Court Segment 8 (Ordered by Weighted Per Point Outcome)

Shot Sequence	Player Cluster	Point Win %	Point Won Immediately %	Point Won Eventually %	Point Lost Immediately %	High Level Chain (HLC) Comparator	Baseline Comparison 1: Difference to HLC	Baseline Comparison 2: Difference to Any Two Shots	Weighted Per Point Outcome
S1 - Forehand Line-->S2 - Backhand Line	R-2 (WTA)	100.00%	0.00%	100.00%	0.00%	Forehand THEN Backhand	72.22%	51.40%	1.00
S1 - Forehand Cross-->S2 - Forehand Line	R-0 (WTA)	100.00%	0.00%	100.00%	0.00%	Two Forehands	58.33%	36.67%	1.00
S1 - Forehand Line-->S2 - Backhand Line	R-0 (WTA)	100.00%	0.00%	100.00%	0.00%	Forehand THEN Backhand	58.33%	36.67%	1.00
S1 - Backhand Middle-->S2 - Backhand Line	R-3 (WTA)	83.33%	66.67%	16.67%	16.67%	Two Backhands	45.83%	41.07%	0.92
S1 - Forehand Cross-->S2 - Forehand Line	R-2 (WTA)	87.50%	0.00%	87.50%	0.00%	Two Forehands	59.72%	38.90%	0.75
S1 - Forehand Line-->S2 - Backhand Middle Cross	R-2 (WTA)	80.00%	0.00%	80.00%	0.00%	Forehand THEN Backhand	52.22%	31.40%	0.60
S1 - Forehand Line-->S2 - Backhand Middle Cross	R-0 (WTA)	80.00%	0.00%	80.00%	0.00%	Forehand THEN Backhand	38.33%	16.67%	0.60
S1 - Forehand Inside In-->S2 - Forehand Inside Out	R-5 (ATP)	66.67%	33.33%	33.33%	0.00%	Two Forehands	46.67%	35.56%	0.50
S1 - Forehand Middle-->S2 - Forehand Middle Cross	R-2 (WTA)	75.00%	0.00%	75.00%	25.00%	Two Forehands	47.22%	26.40%	0.38
S1 - Forehand Cross-->S2 - Forehand Line	R-4 (ATP)	75.00%	0.00%	75.00%	25.00%	Two Forehands	0.00%	29.90%	0.38
S1 - Forehand Inside In-->S2 - Forehand Line	R-1 (ATP)	68.42%	15.79%	52.63%	15.79%	Two Forehands	17.93%	19.80%	0.37
S1 - Backhand Middle-->S2 - Forehand Line	R-2 (ATP)	66.67%	0.00%	66.67%	33.33%	Backhand THEN Forehand	16.67%	38.46%	0.17
S1 - Forehand Middle-->S2 - Backhand Cross	R-2 (WTA)	50.00%	25.00%	25.00%	25.00%	Forehand THEN Backhand	22.22%	1.40%	0.00
S1 - Forehand Middle-->S2 - Backhand Middle Cross	R-4 (ATP)	40.00%	0.00%	40.00%	0.00%	Forehand THEN Backhand	-35.00%	-5.10%	-0.20
S1 - Backhand Line-->S2 - Forehand Cross	R-0 (WTA)	50.00%	0.00%	50.00%	50.00%	Backhand THEN Forehand	8.33%	-13.33%	-0.25
S1 - Backhand Line-->S2 - Forehand Cross	R-2 (WTA)	50.00%	0.00%	50.00%	50.00%	Backhand THEN Forehand	22.22%	1.40%	-0.25
S1 - Forehand Inside Out-->S2 - Forehand Inside Out	R-5 (ATP)	33.33%	0.00%	33.33%	0.00%	Two Forehands	13.33%	2.22%	-0.33
S1 - Backhand Cross-->S2 - Backhand Middle Cross	R-5 (ATP)	33.33%	0.00%	33.33%	0.00%	Two Backhands	13.33%	2.22%	-0.33
S1 - Forehand Inside In-->S2 - Forehand Line	R-3 (WTA)	5.88%	0.00%	5.88%	0.00%	Two Forehands	-31.62%	-36.38%	-0.88
S1 - Forehand Middle Line-->S2 - Backhand Cross	R-3 (WTA)	0.00%	0.00%	0.00%	0.00%	Forehand THEN Backhand	-37.50%	-42.26%	-1.00
S1 - Backhand Middle Cross-->S2 - Forehand Middle	R-0 (WTA)	0.00%	0.00%	0.00%	0.00%	Backhand THEN Forehand	-41.67%	-63.33%	-1.00
S1 - Forehand Inside In-->S2 - Forehand Line	R-5 (ATP)	0.00%	0.00%	0.00%	0.00%	Two Forehands	-20.00%	-31.11%	-1.00
S1 - Backhand Middle Cross-->S2 - Forehand Middle	R-2 (WTA)	0.00%	0.00%	0.00%	0.00%	Backhand THEN Forehand	-27.78%	-48.60%	-1.00
S1 - Backhand Cross-->S2 - Backhand Cross	R-2 (ATP)	0.00%	0.00%	0.00%	25.00%	Two Backhands	-50.00%	-28.21%	-1.13
S1 - Forehand Line-->S2 - Forehand Middle	R-3 (WTA)	11.11%	0.00%	11.11%	88.89%	Two Forehands	-26.39%	-31.15%	-1.22
S1 - Forehand Cross-->S2 - Forehand Line	R-0 (ATP)	0.00%	0.00%	0.00%	66.67%	Two Forehands	-41.67%	-36.96%	-1.33
S1 - Backhand Middle-->S2 - Forehand Line	R-0 (ATP)	0.00%	0.00%	0.00%	66.67%	Backhand THEN Forehand	-41.67%	-36.96%	-1.33

Table 31: Shot Sequence success evaluation for Court Segment 8. Support, Confidence, Lift, Leverage, Conviction, Zhang and Relation thresholds applied

No. of Positive Chains	R-1 (ATP)	R-2 (ATP)	R-4 (ATP)	R-5 (ATP)	R-0 (WTA)	R-2 (WTA)	R-3 (WTA)	R-3 (WTA)
	Shot Sequences With Positive Weighted Per Point Outcome from Court Segment 8							
1	S1 - Forehand Inside In-->S2 - Forehand Line	S1 - Backhand Middle-->S2 - Forehand Line	S1 - Forehand Cross-->S2 - Forehand Line	S1 - Forehand Inside In-->S2 - Forehand Inside Out	S1 - Forehand Cross-->S2 - Forehand Line	S1 - Forehand Line-->S2 - Backhand Line	S1 - Backhand Middle-->S2 - Backhand Line	S1 - Backhand Middle-->S2 - Backhand Line
2					S1 - Forehand Line-->S2 - Backhand Line	S1 - Forehand Cross-->S2 - Forehand Line		
3					S1 - Forehand Line-->S2 - Backhand Middle Cross	S1 - Forehand Line-->S2 - Backhand Middle Cross		
4						S1 - Forehand Middle-->S2 - Forehand Middle Cross		
5						S1 - Forehand Middle-->S2 - Backhand Cross		

Table 32: Shot Sequences with a positive weighted per point outcome from Court Segment 8 by Player Cluster Groups

Sequential Market basket analysis - Unique player contribution and Average ranking of executed shot sequences

Shot Sequence	Player Cluster	No of Unique Players Hitting Chain	Max Percentage Contribution by a single player to Shot Chain total	Average Ranking of Executed shot chain
S1 - Forehand Line-->S2 - Backhand Line	R-2 (WTA)	1	100.00%	88.00
S1 - Forehand Cross-->S2 - Forehand Line	R-0 (WTA)	2	80.00%	76.98
S1 - Forehand Line-->S2 - Backhand Line	R-0 (WTA)	1	100.00%	68.00
S1 - Backhand Middle-->S2 - Backhand Line	R-3 (WTA)	2	91.67%	10.91
S1 - Forehand Cross-->S2 - Forehand Line	R-2 (WTA)	4	50.00%	47.60
S1 - Forehand Line-->S2 - Backhand Middle Cross	R-2 (WTA)	2	80.00%	67.50
S1 - Forehand Line-->S2 - Backhand Middle Cross	R-0 (WTA)	2	80.00%	72.40
S1 - Forehand Inside In-->S2 - Forehand Inside Out	R-5 (ATP)	2	66.67%	11.32
S1 - Forehand Middle-->S2 - Forehand Middle Cross	R-2 (WTA)	3	50.00%	24.50
S1 - Forehand Cross-->S2 - Forehand Line	R-4 (ATP)	2	75.00%	12.00
S1 - Forehand Inside In-->S2 - Forehand Line	R-1 (ATP)	1	100.00%	3.00
S1 - Backhand Middle-->S2 - Forehand Line	R-2 (ATP)	3	33.33%	73.41
S1 - Forehand Middle-->S2 - Backhand Cross	R-2 (WTA)	3	50.00%	14.70

Table 33: Positive Weighted Shot Sequence- Unique player contribution, maximum contribution percentage by a player and average ranking of executed shot sequence from Court Segment 8

Sequential Market basket analysis statistics for selected shot sequence pairs from Court Segment 7 (Ordered by Frequency of A & C together)

antecedents	consequents	Player Cluster	frequency A	frequency C	frequency A&C	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhang	P(A-->C)	gain A-->C	P(C-->A)	*gain_C-->A*	*Relation*
S1 - Backhand Cross	S2 - Backhand Cross	R-3 (WTA)	100	31	20	0.439	0.136	0.088	0.200	1.471	0.028	1.080	0.570	0.167	0.226	0.167	0.226	Complements
S1 - Backhand Cross	S2 - Backhand Line	R-3 (WTA)	100	27	18	0.439	0.118	0.079	0.180	1.520	0.027	1.075	0.609	0.153	0.288	0.025	-0.816	A-->B
S1 - Backhand Line	S2 - Forehand Cross	R-3 (WTA)	39	51	15	0.171	0.224	0.066	0.385	1.719	0.028	1.262	0.505	0.278	0.242	0.000	0.000	A-->B
S1 - Backhand Middle Cross	S2 - Forehand Middle Cross	R-3 (WTA)	79	18	12	0.346	0.079	0.053	0.152	1.924	0.025	1.086	0.735	0.132	0.670	0.000	0.000	A-->B
S1 - Backhand Line	S2 - Forehand Middle Cross	R-2 (WTA)	23	16	11	0.256	0.178	0.122	0.478	2.690	0.077	1.576	0.844	0.324	0.820	0.000	0.000	A-->B
S1 - Backhand Line	S2 - Forehand Line	R-2 (WTA)	23	17	9	0.256	0.189	0.100	0.391	2.072	0.052	1.333	0.695	0.281	0.489	0.000	0.000	A-->B
S1 - Backhand Cross	S2 - Backhand Cross	R-4 (ATP)	20	14	8	0.392	0.275	0.157	0.400	1.457	0.049	1.209	0.516	0.286	0.041	0.286	0.041	Complements
S1 - Backhand Line	S2 - Forehand Middle Cross	R-0 (WTA)	16	12	8	0.271	0.203	0.136	0.500	2.458	0.080	1.593	0.814	0.333	0.639	0.000	0.000	A-->B
S1 - Backhand Line	S2 - Forehand Line	R-0 (WTA)	16	13	8	0.271	0.220	0.136	0.500	2.269	0.076	1.559	0.767	0.333	0.513	0.000	0.000	A-->B
S1 - Backhand Cross	S2 - Backhand Cross	R-5 (ATP)	18	11	7	0.360	0.220	0.140	0.389	1.768	0.061	1.276	0.679	0.280	0.273	0.280	0.273	Complements
S1 - Backhand Cross	S2 - Forehand Cross	R-2 (WTA)	22	15	7	0.244	0.167	0.078	0.318	1.909	0.037	1.222	0.630	0.241	0.448	0.000	0.000	A-->B
S1 - Backhand Cross	S2 - Backhand Line	R-2 (WTA)	22	12	7	0.244	0.133	0.078	0.318	2.386	0.045	1.271	0.769	0.241	0.810	0.000	-1.000	A-->B
S1 - Backhand Middle Cross	S2 - Forehand Middle Cross	R-5 (ATP)	19	7	6	0.380	0.140	0.120	0.316	2.256	0.067	1.257	0.898	0.240	0.714	0.000	0.000	A-->B
S1 - Backhand Middle Cross	S2 - Backhand Middle Cross	R-2 (WTA)	40	9	6	0.444	0.100	0.067	0.150	1.500	0.022	1.059	0.600	0.130	0.304	0.130	0.304	Complements
S1 - Backhand Middle Cross	S2 - Backhand Middle	R-2 (WTA)	40	6	6	0.444	0.067	0.067	0.150	2.250	0.037	1.098	1.000	0.130	0.957	0.000	0.000	A-->B
S1 - Backhand Cross	S2 - Forehand Cross	R-0 (WTA)	16	7	5	0.271	0.119	0.085	0.313	2.634	0.053	1.282	0.851	0.238	1.007	0.000	0.000	A-->B
S1 - Backhand Cross	S2 - Backhand Line	R-0 (WTA)	16	10	5	0.271	0.169	0.085	0.313	1.844	0.039	1.208	0.628	0.238	0.405	0.000	-1.000	A-->B
S1 - Backhand Cross	S2 - Backhand Middle Cross	R-5 (ATP)	18	6	4	0.360	0.120	0.080	0.222	1.852	0.037	1.131	0.719	0.182	0.515	0.050	-0.773	A-->B
S1 - Backhand Middle Cross	S2 - Forehand Cross	R-0 (ATP)	16	7	4	0.308	0.135	0.077	0.250	1.857	0.036	1.154	0.667	0.200	0.486	0.000	0.000	A-->B
S1 - Forehand Inside Out	S2 - Backhand Line	R-2 (WTA)	5	12	4	0.056	0.133	0.044	0.800	6.000	0.037	4.333	0.882	0.444	2.333	0.042	0.250	Complements
S1 - Backhand Cross	S2 - Backhand Cross	R-1 (WTA)	7	7	4	0.259	0.259	0.148	0.571	2.204	0.081	1.728	0.738	0.364	0.403	0.364	0.403	Complements
S1 - Backhand Middle Cross	S2 - Backhand Middle	R-0 (WTA)	23	4	4	0.390	0.068	0.068	0.174	2.565	0.041	1.128	1.000	0.148	1.185	0.000	0.000	A-->B
S1 - Backhand Middle Cross	S2 - Forehand Cross	R-5 (ATP)	19	5	3	0.380	0.100	0.060	0.158	1.579	0.022	1.069	0.591	0.136	0.364	0.000	0.000	A-->B
S1 - Forehand Inside Out	S2 - Forehand Inside Out	R-0 (ATP)	7	6	3	0.135	0.115	0.058	0.429	3.714	0.042	1.548	0.844	0.300	1.600	0.300	1.600	Complements
S1 - Backhand Middle Cross	S2 - Backhand Middle Cross	R-0 (WTA)	23	4	3	0.390	0.068	0.051	0.130	1.924	0.024	1.072	0.787	0.115	0.702	0.115	0.702	Complements

Table 34: Key Sequential Market Basket Evaluation Statistics for Court Segment 7 Shot Sequences. Support, Confidence, Lift, Leverage, Conviction, Zhang and Relation thresholds applied

Sequential Market basket analysis - shot sequence success evaluation in Court Segment 7 (Ordered by Weighted Per Point Outcome)

Shot Sequence	Player Cluster	Point Win %	Point Won Immediately %	Point Won Eventually %	Point Lost Immediately %	High Level Chain (HLC) Comparator	Baseline Comparison 1: Difference to HLC	Baseline Comparison 2: Difference to Any Two Shots	Weighted Per Point Outcome
S1 - Forehand Inside Out-->S2 - Forehand Inside Out	R-0 (ATP)	100.00%	33.33%	66.67%	0.00%	Two Forehands	69.23%	57.69%	1.17
S1 - Backhand Cross-->S2 - Backhand Cross	R-1 (WTA)	100.00%	25.00%	75.00%	0.00%	Two Backhands	53.85%	44.44%	1.13
S1 - Backhand Middle Cross-->S2 - Backhand Middle Cross	R-0 (WTA)	100.00%	0.00%	100.00%	0.00%	Two Backhands	62.50%	59.32%	1.00
S1 - Backhand Cross-->S2 - Backhand Line	R-2 (WTA)	85.71%	0.00%	85.71%	14.29%	Two Backhands	50.63%	51.27%	0.64
S1 - Backhand Cross-->S2 - Backhand Line	R-0 (WTA)	80.00%	0.00%	80.00%	0.00%	Two Backhands	42.50%	39.32%	0.60
S1 - Backhand Middle Cross-->S2 - Forehand Middle Cross	R-3 (WTA)	83.33%	0.00%	83.33%	16.67%	Backhand THEN Forehand	38.02%	35.96%	0.58
S1 - Backhand Cross-->S2 - Backhand Cross	R-3 (WTA)	80.00%	0.00%	80.00%	5.00%	Two Backhands	34.69%	32.63%	0.58
S1 - Backhand Line-->S2 - Forehand Cross	R-3 (WTA)	73.33%	6.67%	66.67%	13.33%	Backhand THEN Forehand	28.02%	25.96%	0.43
S1 - Backhand Middle Cross-->S2 - Forehand Middle Cross	R-5 (ATP)	66.67%	0.00%	66.67%	0.00%	Backhand THEN Forehand	26.67%	24.67%	0.33
S1 - Backhand Middle Cross-->S2 - Forehand Cross	R-5 (ATP)	66.67%	33.33%	33.33%	33.33%	Backhand THEN Forehand	26.67%	24.67%	0.33
S1 - Backhand Cross-->S2 - Backhand Line	R-3 (WTA)	55.56%	16.67%	38.89%	22.22%	Two Backhands	10.24%	8.19%	0.08
S1 - Backhand Line-->S2 - Forehand Middle Cross	R-0 (WTA)	50.00%	0.00%	50.00%	0.00%	Backhand THEN Forehand	12.50%	9.32%	0.00
S1 - Backhand Line-->S2 - Forehand Middle Cross	R-2 (WTA)	45.45%	0.00%	45.45%	0.00%	Backhand THEN Forehand	10.37%	11.01%	-0.09
S1 - Backhand Cross-->S2 - Backhand Cross	R-5 (ATP)	42.86%	0.00%	42.86%	0.00%	Two Backhands	2.86%	0.86%	-0.14
S1 - Backhand Cross-->S2 - Backhand Cross	R-4 (ATP)	37.50%	0.00%	37.50%	12.50%	Two Backhands	-45.83%	-13.48%	-0.31
S1 - Backhand Middle Cross-->S2 - Forehand Cross	R-0 (ATP)	25.00%	25.00%	0.00%	25.00%	Backhand THEN Forehand	-5.77%	-17.31%	-0.50
S1 - Backhand Cross-->S2 - Forehand Cross	R-2 (WTA)	28.57%	28.57%	0.00%	57.14%	Backhand THEN Forehand	-6.52%	-5.87%	-0.57
S1 - Backhand Cross-->S2 - Backhand Middle Cross	R-5 (ATP)	25.00%	0.00%	25.00%	25.00%	Two Backhands	-15.00%	-17.00%	-0.63
S1 - Backhand Middle Cross-->S2 - Backhand Middle Cross	R-2 (WTA)	16.67%	0.00%	16.67%	33.33%	Two Backhands	-18.42%	-17.78%	-0.83
S1 - Backhand Middle Cross-->S2 - Backhand Middle	R-2 (WTA)	16.67%	16.67%	0.00%	83.33%	Two Backhands	-18.42%	-17.78%	-1.00
S1 - Backhand Cross-->S2 - Forehand Cross	R-0 (WTA)	20.00%	0.00%	20.00%	80.00%	Backhand THEN Forehand	-17.50%	-20.68%	-1.00
S1 - Backhand Line-->S2 - Forehand Line	R-2 (WTA)	0.00%	0.00%	0.00%	44.44%	Backhand THEN Forehand	-35.09%	-34.44%	-1.22
S1 - Backhand Line-->S2 - Forehand Line	R-0 (WTA)	0.00%	0.00%	0.00%	50.00%	Backhand THEN Forehand	-37.50%	-40.68%	-1.25
S1 - Forehand Inside Out-->S2 - Backhand Line	R-2 (WTA)	0.00%	0.00%	0.00%	100.00%	Forehand THEN Backhand	-35.09%	-34.44%	-1.50
S1 - Backhand Middle Cross-->S2 - Backhand Middle	R-0 (WTA)	0.00%	0.00%	0.00%	100.00%	Two Backhands	-37.50%	-40.68%	-1.50

Table 35: Shot Sequence success evaluation for Court Segment 7. Support, Confidence, Lift, Leverage, Conviction, Zhang and Relation thresholds applied

No. of Positive Chains	Shot Sequences With Positive Weighted Per Point Outcome from Court Segment 7					
	R-0 (ATP)	R-5 (ATP)	R-0 (WTA)	R-1 (WTA)	R-2 (WTA)	R-3 (WTA)
1	S1 - Forehand Inside Out-->S2 - Forehand Inside Out	S1 - Backhand Middle Cross-->S2 - Forehand Middle Cross	S1 - Backhand Middle Cross-->S2 - Backhand Middle Cross	S1 - Backhand Cross-->S2 - Backhand Cross	S1 - Backhand Cross-->S2 - Backhand Line	S1 - Backhand Middle Cross-->S2 - Forehand Middle Cross
2		S1 - Backhand Middle Cross-->S2 - Forehand Cross	S1 - Backhand Cross-->S2 - Backhand Line			S1 - Backhand Cross-->S2 - Backhand Cross
3			S1 - Backhand Line-->S2 - Forehand Middle Cross			S1 - Backhand Line-->S2 - Forehand Cross
4						S1 - Backhand Cross-->S2 - Backhand Line

Table 36: Shot Sequences with a positive weighted per point outcome from Court Segment 7 by Player Cluster Groups

Sequential Market basket analysis - Unique player contribution and Average ranking of executed shot sequences				
Shot Sequence	Player Cluster	No of Unique Players Hitting Chain	Max Percentage Contribution by a single player to Shot Chain total	Average Ranking of Executed shot chain
S1 - Forehand Inside Out-->S2 - Forehand Inside Out	R-0 (ATP)	1	100.00%	100.00
S1 - Backhand Cross-->S2 - Backhand Cross	R-1 (WTA)	4	25.00%	33.50
S1 - Backhand Middle Cross-->S2 - Backhand Middle Cross	R-0 (WTA)	3	33.33%	87.50
S1 - Backhand Cross-->S2 - Backhand Line	R-2 (WTA)	3	57.14%	65.40
S1 - Backhand Cross-->S2 - Backhand Line	R-0 (WTA)	2	80.00%	86.50
S1 - Backhand Middle Cross-->S2 - Forehand Middle Cross	R-3 (WTA)	1	100.00%	6.00
S1 - Backhand Cross-->S2 - Backhand Cross	R-3 (WTA)	2	95.00%	8.00
S1 - Backhand Line-->S2 - Forehand Cross	R-3 (WTA)	1	100.00%	6.00
S1 - Backhand Middle Cross-->S2 - Forehand Middle Cross	R-5 (ATP)	5	33.33%	14.80
S1 - Backhand Middle Cross-->S2 - Forehand Cross	R-5 (ATP)	3	33.33%	18.33
S1 - Backhand Cross-->S2 - Backhand Line	R-3 (WTA)	2	94.44%	8.00
S1 - Backhand Line-->S2 - Forehand Middle Cross	R-0 (WTA)	2	50.00%	37.00

Table 37: Positive Weighted Shot Sequence- Unique player contribution, maximum contribution percentage by a player and average ranking of executed shot sequence from Court Segment 7

Sequential Market basket analysis statistics for selected shot sequence pairs from Court Segment 9 (Ordered by Frequency of A & C together)

antecedents	consequents	Player Cluster	frequency A	frequency C	frequency A&C	antecedent support	consequent support	support	confidence	lift	leverage	conviction	zhang	P(A-->C)	gain A-->C	P(C-->A)	*gain_C-->A*	*Relation*
S1 - Forehand Middle Cross	S2 - Backhand Middle	R-2 (WTA)	45	19	15	0.542	0.229	0.181	0.333	1.456	0.057	1.157	0.684	0.250	0.092	0.000	0.000	A-->C
S1 - Forehand Cross	S2 - Backhand Middle Cross	R-3 (WTA)	89	27	15	0.360	0.109	0.061	0.169	1.542	0.021	1.071	0.549	0.144	0.319	0.000	0.000	A-->C
S1 - Forehand Middle Cross	S2 - Backhand Cross	R-3 (WTA)	70	29	14	0.283	0.117	0.057	0.200	1.703	0.023	1.103	0.576	0.167	0.420	0.000	0.000	A-->C
S1 - Forehand Middle Cross	S2 - Backhand Middle	R-0 (WTA)	21	14	13	0.553	0.368	0.342	0.619	1.680	0.139	1.658	0.905	0.382	0.038	0.000	0.000	A-->C
S1 - Backhand Middle Line	S2 - Backhand Middle	R-3 (WTA)	36	46	12	0.146	0.186	0.049	0.333	1.790	0.021	1.221	0.517	0.250	0.342	0.000	0.000	A-->C
S1 - Forehand Line	S2 - Forehand Cross	R-3 (WTA)	37	27	11	0.150	0.109	0.045	0.297	2.720	0.028	1.268	0.744	0.229	1.096	0.183	-0.075	A-->C
S1 - Forehand Cross	S2 - Backhand Inside In	R-3 (WTA)	89	11	10	0.360	0.045	0.040	0.112	2.523	0.024	1.076	0.944	0.101	1.268	0.125	0.144	Complements
S1 - Forehand Cross	S2 - Forehand Line	R-2 (WTA)	12	11	6	0.145	0.133	0.072	0.500	3.773	0.053	1.735	0.859	0.333	1.515	0.000	-1.000	A-->C
S1 - Forehand Line	S2 - Backhand Cross	R-2 (WTA)	10	17	5	0.120	0.205	0.060	0.500	2.441	0.036	1.590	0.671	0.333	0.627	0.000	0.000	A-->C
S1 - Forehand Cross	S2 - Forehand Line	R-0 (WTA)	7	6	5	0.184	0.158	0.132	0.714	4.524	0.102	2.947	0.955	0.417	1.639	0.000	-1.000	A-->C
S1 - Forehand Line	S2 - Backhand Cross	R-0 (WTA)	7	10	4	0.184	0.263	0.105	0.571	2.171	0.057	1.719	0.661	0.364	0.382	0.000	0.000	A-->C
S1 - Backhand Middle Line	S2 - Backhand Middle Cross	R-2 (WTA)	12	14	4	0.145	0.169	0.048	0.333	1.976	0.024	1.247	0.577	0.250	0.482	0.000	0.000	A-->C
S1 - Backhand Middle Cross	S2 - Backhand Line	L-0 (ATP)	8	5	4	0.471	0.294	0.235	0.500	1.700	0.097	1.412	0.778	0.333	0.133	0.143	-0.393	A-->C
S1 - Forehand Middle Cross	S2 - Forehand Cross	R-0 (WTA)	21	5	4	0.553	0.132	0.105	0.190	1.448	0.033	1.073	0.691	0.160	0.216	0.000	-1.000	A-->C
S1 - Forehand Line	S2 - Backhand Middle Cross	R-4 (ATP)	7	3	3	0.350	0.150	0.150	0.429	2.857	0.098	1.488	1.000	0.300	1.000	0.000	0.000	A-->C

Table 38: Key Sequential Market Basket Evaluation Statistics for Court Segment 9 Shot Sequences. Support, Confidence, Lift, Leverage, Conviction, Zhang and Relation thresholds applied

Sequential Market basket analysis - shot sequence success evaluation in Court Segment 9 (Ordered by Weighted Per Point Outcome)									
Shot Sequence	Player Cluster	Point Win %	Point Won Immediately %	Point Won Eventually %	Point Lost Immediately %	High Level Chain (HLC) Comparator	Baseline Comparison 1: Difference to HLC	Baseline Comparison 2: Difference to Any Two Shots	Weighted Per Point Outcome
S1 - Forehand Line-->S2 - Forehand Cross	R-3 (WTA)	100.00%	0.00%	100.00%	0.00%	Two Forehands	42.11%	43.72%	1.00
S1 - Forehand Cross-->S2 - Backhand Inside In	R-3 (WTA)	100.00%	0.00%	100.00%	0.00%	Forehand THEN Backhand	42.11%	43.72%	1.00
S1 - Forehand Line-->S2 - Backhand Cross	R-2 (WTA)	100.00%	0.00%	100.00%	0.00%	Forehand THEN Backhand	75.00%	54.22%	1.00
S1 - Forehand Line-->S2 - Backhand Cross	R-0 (WTA)	100.00%	0.00%	100.00%	0.00%	Forehand THEN Backhand	0.00%	52.63%	1.00
S1 - Forehand Middle Cross-->S2 - Backhand Cross	R-3 (WTA)	85.71%	0.00%	85.71%	0.00%	Forehand THEN Backhand	27.82%	29.44%	0.71
S1 - Backhand Middle Line-->S2 - Backhand Middle	R-3 (WTA)	83.33%	0.00%	83.33%	8.33%	Two Backhands	25.44%	27.06%	0.63
S1 - Forehand Middle Cross-->S2 - Backhand Middle	R-0 (WTA)	69.23%	0.00%	69.23%	30.77%	Forehand THEN Backhand	-30.77%	21.86%	0.23
S1 - Forehand Line-->S2 - Backhand Middle Cross	R-4 (ATP)	66.67%	0.00%	66.67%	33.33%	Forehand THEN Backhand	66.67%	6.67%	0.17
S1 - Forehand Middle Cross-->S2 - Backhand Middle	R-2 (WTA)	60.00%	0.00%	60.00%	33.33%	Forehand THEN Backhand	35.00%	14.22%	0.03
S1 - Backhand Middle Line-->S2 - Backhand Middle Cross	R-2 (WTA)	50.00%	0.00%	50.00%	25.00%	Two Backhands	25.00%	4.22%	-0.13
S1 - Backhand Middle Cross-->S2 - Backhand Line	L-0 (ATP)	25.00%	0.00%	25.00%	25.00%	Two Backhands	8.33%	-4.41%	-0.63
S1 - Forehand Cross-->S2 - Backhand Middle Cross	R-3 (WTA)	20.00%	0.00%	20.00%	60.00%	Forehand THEN Backhand	-37.89%	-36.28%	-0.90
S1 - Forehand Cross-->S2 - Forehand Line	R-0 (WTA)	20.00%	20.00%	0.00%	80.00%	Two Forehands	-80.00%	-27.37%	-0.90
S1 - Forehand Cross-->S2 - Forehand Line	R-2 (WTA)	16.67%	0.00%	16.67%	66.67%	Two Forehands	-8.33%	-29.12%	-1.00
S1 - Forehand Middle Cross-->S2 - Forehand Cross	R-0 (WTA)	0.00%	0.00%	0.00%	100.00%	Two Forehands	-100.00%	-47.37%	-1.50

Table 39: Shot Sequence success evaluation for Court Segment 9. Support, Confidence, Lift, Leverage, Conviction, Zhang and Relation thresholds applied

No. of Positive Chains	Shot Sequences With Positive Weighted Per Point Outcome from Court Segment 9			
	R-4 (ATP)	R-0 (WTA)	R-2 (WTA)	R-3 (WTA)
1	S1 - Forehand Line-->S2 - Backhand Middle Cross	S1 - Forehand Line-->S2 - Backhand Cross	S1 - Forehand Line-->S2 - Backhand Cross	S1 - Forehand Line-->S2 - Forehand Cross
2		S1 - Forehand Middle Cross-->S2 - Backhand Middle	S1 - Forehand Middle Cross-->S2 - Backhand Middle	S1 - Forehand Cross-->S2 - Backhand Inside In
3				S1 - Forehand Middle Cross-->S2 - Backhand Cross
4				S1 - Backhand Middle Line-->S2 - Backhand Middle

Table 40: Shot Sequences with a positive weighted per point outcome from Court Segment 9 by Player Cluster Groups

Sequential Market basket analysis - Unique player contribution and Average ranking of executed shot sequences

Shot Sequence	Player Cluster	No of Unique Players Hitting Chain	Max Percentage Contribution by a single player to Shot Chain total	Average Ranking of Executed shot chain
S1 - Forehand Line-->S2 - Forehand Cross	R-3 (WTA)	1	100.00%	6.00
S1 - Forehand Cross-->S2 - Backhand Inside In	R-3 (WTA)	1	100.00%	6.00
S1 - Forehand Line-->S2 - Backhand Cross	R-2 (WTA)	2	80.00%	73.00
S1 - Forehand Line-->S2 - Backhand Cross	R-0 (WTA)	1	100.00%	88.00
S1 - Forehand Middle Cross-->S2 - Backhand Cross	R-3 (WTA)	1	100.00%	6.00
S1 - Backhand Middle Line-->S2 - Backhand Middle	R-3 (WTA)	1	100.00%	6.00
S1 - Forehand Middle Cross-->S2 - Backhand Middle	R-0 (WTA)	2	92.31%	90.40
S1 - Forehand Line-->S2 - Backhand Middle Cross	R-4 (ATP)	2	66.67%	19.33
S1 - Forehand Middle Cross-->S2 - Backhand Middle	R-2 (WTA)	4	80.00%	73.00

Table 41: Positive Weighted Shot Sequence- Unique player contribution, maximum contribution percentage by a player and average ranking of executed shot sequence from Court Segment 9

Discussion and Conclusion

The aim of the study was to find optimal shot sequences for different player types from different court positions. Due to the small data sample used (47 matches), not enough shot sequence combinations were observed and not enough player type data samples were obtained to draw any substantial conclusions or settle on any findings of optimal shot sequences for player types from court positions.

Nonetheless the interim results show some interesting findings that should warrant further research to investigate further. Primarily, there is some evidence to suggest different player clusters have different shot sequence options from different court segments. Some examples discussed below:

1) *Deeper Court position preferred by taller, male and extreme grip players*

Players in the R-0(ATP) playing cluster, who have a shorter stature had five shot sequence options from Court Segment 4, 5, and 6 that met minimum thresholds of the shot sequence suitability metrics, this was the most of any of the player types. A deeper look at the results showed the contribution to this shot sequence in every instance was from more than one unique player, with no player contributing to more than 50% of the shots to any of these five shot sequences, adding value that these findings are generalisable to the broader group of players falling within that cluster.

An explanation for the success of R-0(ATP) in this analysis can be explained by research which has previously found that shorter stature players generate less ball speed off tennis strokes[34] and generally want more time to neutralise the power of bigger opponents, hence they play from deeper court positions as well as extend the rallies where they have an advantage. Shot sequences like Forehand Line →Backhand Middle Cross and Forehand Cross → Forehand Line from segment 6 will likely get the game to be more laterally movement based, which can create an advantage for players in R-0(ATP).

Another player type R-3(ATP) also had more shot sequence options from Court Segments 4, 5 and 6. This playing cluster contains players with extreme grips who need a bit more time for them to hit balls with their full swing paths[35]. These player types with these grips prefer to take the shot on the move in terms of footwork[36], which is more feasible from deeper court positions. Sequences like Backhand Cross → Forehand Middle can help these play a game which is more about using their grips to help generate added weight of shot (more spin) than hitting outright winners or go close to the lines.

2) *Closer to the baseline preferred by female players and bigger bodied male players*

Female players had more shot sequence options closer to the baseline than from deeper position. In court segments 4,5 and 6 (deeper positions) no female player cluster had more than one shot sequences option, whereas in court segments 7,8 and 9 (closer to the baseline), multiple female player types had more than one shot sequence option. For instance, the three female player clusters with the most data samples R-0(WTA), R-2(WTA) and R-3(WTA) had 8, 8, 10 shot sequence options from segments 7,8 and 9 respectively.

These results can be put into context by existing literature which allude that Women's tennis has:

- Less court coverage[37] and explosive movement[38] compared to men's tennis meaning closer baseline positions are preferred to deeper
- Flatter ground strokes are hit [39]which are executed with better accuracy from closer to the baseline[2, 39]
- More shots hit by opponents on the return, 3rd and 4th shots end up in parts of the court that are better handled by racquet contact in positions closer to the baseline[40]

Bigger body males in R-5(ATP) also have better success from closer to the baseline than deeper, particularly from court segment 7. The time and space pressures[41] they create for opponents from closer outweighs what they can create from deeper. The movement and endurance deficiencies of taller players is also less exposed than if they were playing deeper in the court.

3) Wider vs Central positions

Certain player types like R-3(WTA), had considerably better shot sequence options from wider court positions of the court, than they did from central positions of the court. Players in that cluster have semi-open forehand and backhand grips which explains their ability to hit the forehand line and forehand cross effectively on the first shot of shot sequences initiated from the Deuce side of the court, as well backhand cross and backhand line on the first shot of shot sequences initiated from the Ad side of the court. The time and space pressure created after these first shots allow them to have multiple other shots available to them on the second shot in the sequence.

This advantage is not as great from central positions where they are unable to use the angles as well or change direction on a shot and have the same impact on an opponent. Therefore, from a tactical perspective they should focus on hitting a shot that allows them to move into wider position that allows them to initiate a subsequent shot sequence to take advantage of their strengths.

Other player types like R-2(WTA) had slightly more shot sequence options from central court positions. These players are the bigger body frame female players that can have a lot of good shot combinations initiated from central positions close to the baseline where their power games can hurt their opponents but are more vulnerable from wider positions in the court.

4) Shot Sequences that worked for all

Some shot sequences were used by most of the player types, for instance the Forehand Cross → Forehand Line was used by seven of the player types with five player types having a positive WPPO and point success percentage above 50%.

The most prominent use of this sequence came from court segment 6, where it was used 95 times. When considering putting time and space pressure on opponents this combination allows players to maximise the advantage by playing the Forehand Cross first that biomechanically can give them some extra racquet head speed than most other shots, but maintain a good level of shot accuracy [42], then allow them to continue the rally ascendancy by attacking with the Forehand Line.

Practical Application – Dashboard for Opponent Scouting, Coach Education and Junior player development

A dashboard was developed allowing users to drill into findings further. The users can apply a number of filters to see the results based on certain conditions – e.g. playing only against players from R-3(ATP) Cluster or against left handed player etc. This dashboard can help individualise analysis to specific players or situations – e.g. match scouting an opponent. If used alongside models that predict a junior player adult height and weights, the dashboard has a 'player cluster predictor' view to show which class the players falls into. See Appendix – Dashboard Implementation Section

Extensions and future directions

More Data

More data samples needed – estimated at least 1000 matches, with 15 matches per player cluster with a minimum of 4 different players from each player type. It would also be beneficial to have players at various ends of the ranking spectrum in each player cluster to help evaluate if the chain requires 'higher' skill levels.

By addressing the limited data sample in relation to player clusters, better analysis and comparison can be made between player clusters containing right-handed players vs left handed players. This was not possible in this thesis.

The definition of 'optimal' could be also expanded further to look for shot sequences that are 'hidden gems' – e.g. found infrequently but had a surprising value when played. The greater the dataset the more of these that would be found.

Threshold Review

The thresholds in this paper were selected with a combination of relying on similar SMBA examples and trial and error. The dataset for the project was finalised in the last two weeks of the project timeline, hence the thresholds set could benefit from a deeper review and evaluation going forward to ensure shot sequences being captured were suitable for the analysis and/or shot sequences that were interesting were not being omitted.

Future extensions of this work should consider adding another threshold requiring maximum contributions for shot sequences to be more evenly spread – e.g. under 50% by a single player.

Alternative to WPPO

Improvements could be made to WPPO to be more specific in assigning penalty or reward to immediate point outcomes. For instance, have the shot sequence scored against changes to court position of the player and the opponent rather than point outcomes. This could be a more precise way to attribute a reward or penalty to a shot sequence being played.

Additional Clustering

Clustering was conducted with top 100 ATP and WTA players. This should be expanded to include lower tier players. To be able to scale for the increase in sample size, the CV solution for grip/shot classification should be further developed (Appendix C). With research indicating athletes getting taller and having a higher body mass[43], clustering will need to be revisited from time to time to ensure the feature importance attributed by the algorithms is still consistent with the game.

Further Study – More Shot Types included in Shot Sequence

Table 14 shows that while a similar amount of ground strokes is observed between male and female players, the later ended up with 3,254 more shot sequences that contained ground strokes only. This highlighted some of the differences in the men's game with more points ending with the serve and 1 ground stroke (Serve +1) or with the men using more volleys and drop shots in their shot sequences. These findings align with the *Carboch[19]* article finding men using the serve more to end points, and *Fernandez* [44] finding men are likely to use slice and volley more. A future study could look to expand the analysis to shot sequences involving the serve, return and non-ground strokes like volleys and drop shots.

Other Improvements

This study only considers shots hit by the impact player of the first shot in the sequence. There is research indicating that on court position of a player may influence the anticipation of shot outcome[45] by the opponent, therefore further analysis is merited to look at shots hit in-between the players shot sequence by the opponent and the shots and court position after the shot sequence has been executed. This information is available to users of the dashboard built as part of this project.

As shot sequences may diminish in value over time, a pattern robustness analysis should be considered going forward. This can be useful extra insight from a high-performance perspective to ensure shot sequences do not get overplayed.

Adding quantitative time pressure metrics[41] to shot sequences will provide another way to rank shot sequences other than tying it to point result.

References

1. ITF Federation. Tennis Glossary [Webpage]. London: ITF; 2021 [cited Nov 2021. Available from: https://m.itftennis.com/en/about-us/organisation/tennis-glossary/.
2. Association UST. Tennis Tactics: Winning Patterns of Play. 1 ed. New York: Human Kinetics; 1996. 236 p.
3. Reid M, McMurtrie D, Crespo M. Title: The relationship between match statistics and top 100 ranking in professional men's tennis. International Journal of Performance Analysis in Sport. 2017;10(2):131-8.
4. Buscombe RM, Ward JM, editors. Talent Identification and Development in Tennis2014.
5. Piatetsky-Shapiro G. Discovery, Analysis, and Presentation of Strong Rules. Knowledge Discovery in Databases: AAAI/MIT Press; 1991. p. 229-48.
6. Ivan FV-C, Sebastián AR. Extending market basket analysis with graph mining techniques: A real case. Expert Systems with Applications. 2014;41(4, Part 2):1928-36.
7. Agrawal R, Srikant R. Fast Algorithms for Mining Association Rules. Proc 20th Int Conf Very Large Data Bases VLDB. 2000;1215.
8. Wenninger S, Link D, Lames M. Data Mining in Elite Beach Volleyball – Detecting Tactical Patterns Using Market Basket Analysis. International Journal of Computer Science in Sport. 2019;18(2):1-19.
9. Kamakura WA. Sequential market basket analysis. Marketing Letters. 2012;23(3):505-16.
10. Weidner D, Atzmueller M, Seipel D. Finding Maximal Non-redundant Association Rules in Tennis Data. 2020. p. 59-78.
11. Cui Y, Gomez MA, Goncalves B, Sampaio J. Clustering tennis players' anthropometric and individual features helps to reveal performance fingerprints. Eur J Sport Sci. 2019;19(8):1032-44.
12. Mehaffey EL. The Effect of Teaching selected forehand stances and grips on Tennis achievement by College Men: Indiana University; 1966.
13. Eng D, Hagler D. A novel analysis of grip variations on the two-handed backhand. 2014.
14. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276-82.
15. Krumer A, Rosenboim M, Shapir O. Gender, Competitiveness, and Physical Characteristics: Evidence From Professional Tennis. Journal of Sports Economics. 2014;17.
16. Muhamad TA, Golestani F, Razak MRA. Comparison of Open and Closed Stance Forehand Strokes among Intermediate Tennis Players. International Journal of Kinesiology and Sports Science. 2016;4:26-32.
17. Genevois C, Reid M, Rogowski I, Crespo M. Performance Factors Related to the Different Tennis Backhand Groundstrokes: A Review. Journal of Sport Sciences and Medicine. 2015;14:194-202.
18. Loffing F, Hagemann N, Strauss B. The Serve in Professional Men's Tennis: Effects of Players' Handedness. International Journal of Performance Analysis in Sport. 2009;9:255-74.
19. Carboch J. Comparison of game characteristics of male and female tennis players at grand-slam tournaments in 2016. Trends in sport sciences. 2017;4:151-5.
20. Loffing F, Hagemann N, Strauss B. Left-Handedness in Professional and Amateur Tennis. PloS one. 2012;7:e49325.
21. Li Y, Wu H. A Clustering Method Based on K-Means Algorithm. Physics Procedia. 2012;25:1104-9.
22. Chaturvedi A, Green P, Caroll J. K-modes Clustering. Journal of Classification. 2001;18:35-55.
23. Huang Z. Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. Data Mining and Knowledge Discovery. 1998;2(3):283-304.
24. ji J, Pang W, Zhou C, Han X, Wang Z. A fuzzy k-prototype clustering algorithm for mixed numeric and categorical data. Knowledge-Based Systems. 2013;30:129–35.

25. Day WHE, Edelsbrunner H. Efficient algorithms for agglomerative hierarchical clustering methods. Journal of Classification. 1984;1(1):7-24.
26. Preud'homme G, Duarte K, Dalleau K, Lacomblez C, Bresso E, Smaïl-Tabbone M, et al. Head-to-head comparison of clustering methods for heterogeneous data: a simulation-driven benchmark. Scientific Reports. 2021;11(1):4202.
27. Milligan GW, Cooper MC. A study of standardization of variables in cluster analysis. Journal of Classification. 1988;5(2):181-204.
28. Umargono E, Suseno J, Gunawan SK. K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula. 2020.
29. Wang F, Franco-Penya H-H, Kelleher J, Pugh J, Ross R. An Analysis of the Application of Simplified Silhouette to the Evaluation of k-means Clustering Validity. 2017.
30. Kaufman L, Rousseeuw P. Finding Groups in Data: An Introduction to Cluster Analysis2009.
31. Akogul S, Erişoğlu M. A Comparison of Information Criteria in Clustering Based on Mixture of Multivariate Normal Distributions. Mathematical and Computational Applications. 2016;21:34.
32. Brin S, Motwani R, Ullman JD, Tsur S, editors. Dynamic itemset counting and implication rules for market basket data. SIGMOD '97; 1997.
33. Zhang T, editor Association Rules. PAKDD; 2000.
34. Söğüt M. Stature: Does it really make a difference in match-play outcomes among professional tennis players? International Journal of Performance Analysis in Sport. 2018;18:255-61.
35. Lavoie S. SportsEd TV. 2021. [cited 2021 November 2021]. Available from: https://sportsedtv.com/blog/What-your-forehand-grip-means-for-shot-depth-and-point-of-contact-tennis%C2%A0.
36. Suprenko M. Biomechanical substantiation of motor and punch action formation in tennis by taking into account the formation of promising skills and abilities. Journal of Physical Education and Sport. 2021;21(01).
37. Cross R. Men's tennis vs Women's tennis. ITF Coaching and Sport Science Review. 2014;a(62).
38. Reid M, Morgan S, Whiteside D. Matchplay characteristics of Grand Slam tennis: implications for training and conditioning. Journal of sports sciences. 2016;34:1-8.
39. Antoun R. Women's Tennis Tactics. LTA, editor. London: Human Kinetics Publishers; 2007. 232 p.
40. Born P, Malejka L, Behrens M, Grambow R, Meffert D, Breuer J, et al. Stroke placement in women's professional tennis: What's after the serve? Sport Science. 2021;3:37-45.
41. Mlakar M, Kovalchik SA. Analysing time pressure in professional tennis. Journal of Sports Analytics. 2020;6:147-54.
42. Landlinger J, Stöggl T, Lindinger S, Wagner H, Müller E. Differences in ball speed and accuracy of tennis groundstrokes between elite and high-performance players. European Journal of Sport Science. 2012;12(4):301-8.
43. Gale-Watts A, Nevill A. From endurance to power athletes: The changing shape of successful male professional tennis players. European Journal of Sport Science. 2016;16:1-7.
44. Fernandez-Fernandez J, Mendez-Villanueva A, Pluim B. Intensity of tennis match play * Commentary. British journal of sports medicine. 2006;40:387-91; discussion 91.
45. Loffing F, Hagemann N. On-Court Position Influences Skilled Tennis Players' Anticipation of Shot Outcome. Journal of sport & exercise psychology. 2014;36:14-26.
46. Bollettieri N. Bollettieri's tennis handbook. Champaign, IL: Human Kinetics; 2001.

Appendix

Appendix A- Literature Search Strategy

Key words for each of the topics in Table 1 were searched on various research databases including Scopus, PubMed and IEEE Xplore and short listed with a suitability score according to criteria in Figure 30 and then categorised according to project relevance categories in Table 2.

Topic	Description
Sequential Rule mining/Market basket Analysis in Elite sport	Papers that used rule mining to evaluate the association between certain transactions in a game and then value of the sequence of those transactions
Clustering	Papers that used anthropometric and box statistical variables to cluster players into groups
Tennis Grips	Papers that looked at methods of collecting data or labelling of tennis grips based on video vision or still shot images
Tennis Shot Taxonomy	Papers that discuss tennis shot types or clusters
Analysis using Hawkeye Tennis data	Papers that used tennis spatiotemporal Hawkeye data

Table 42 Key Search Topics

- RELEVANCE TO ASSOCIATION RULE MINING, MARKET BASKET ANALYSIS, SEQUENTIAL RULE MINING OR CLUSTERING
- RELEVANCE TO TENNIS
- RELEVANT VARIABLES USED (IF APPLICABLE)
- WAS THE RESEARCH CONDUCTED RELATING TO ELITE SPORT?
- SIMILARITY OF PROBLEM ADDRESSED IN PAPER TO ALL OR PART OF RESEARCH PROBLEM
- ABILITY TO PROVIDE CONTEXT TO RESEARCH PAPER PROBLEM OR FINDINGS
- APPROPRIATE METHODOLOGY/PROCESS USED IN PAPER.

Figure 30:Criteria used to assess the quality and suitability of the literature to the project problem

Research Category	Description
Foundational	Contains an essential theoretical or analytical component required to answer one of more of the problem topics
Replicate Process	Parts of the paper contain processes that can be replicated for answering parts of the problem
Contextual	Research findings that are not directly relevant to the problem topics but can add some context to help interpret results better

Table 43: Research Paper Project Relevance Category

The literature search process is summarised in Figure 32 where the databases searches initially yielded 116 articles, but through various screening and relevance determination processes 9 articles were found as relevant to the project, and 4 articles has significant enough to highlight in this literature review.

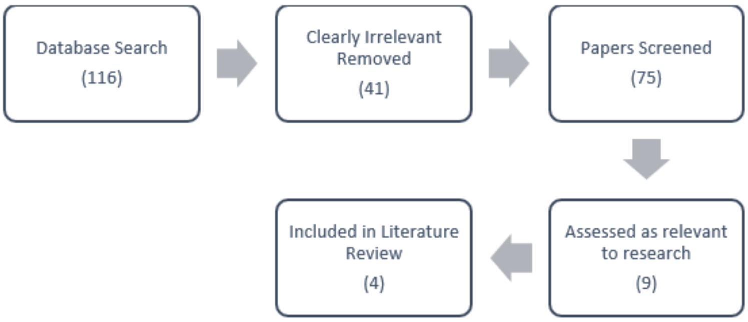


Figure 31: Literature Search Process.

Appendix B- Tennis Grip Study

Tennis Stroke Grips

Tennis coaching expert Nick Bollettieri in his tennis handbook alludes to the significant differences in playing style brought on by grip choices.[46]

Eng & Hagler[13] looked at the variations in grips used on two handed backhands on the top 100 ATP (male) and WTA (female) athletes. In their paper, data collection consisted of manually annotating grip type based on close-up high-quality photographs of 5-8 strokes of each player. The authors independently verified the grips of all players to ensure inter-rater reliability.

The study also included two different variations within grip types: 1. *Precision Grip*; and *2. Power Grips.*

TWO-HANDED GRIPS			
	ATP	WTA	TOTAL
BOTTOM HAND			
Continental Strong Precision	13	15	28
Continental Regular Precision	58	63	121
Eastern Backhand	7	18	25
TOP HAND			
Continental Precision	4	14	18
Eastern Precision, thumb on grip	19	45	64
Eastern Precision, thumb on fingers	25	19	44
Eastern Power	22	8	30
Semi-Western	5	8	13
Continental Power	3	0	3
Number of Players in Top 100	78	96	

Figure 32: From Eng & Hagler - Two handed grips ATP and WTA top 100

The study did not look at forehand grips.

Appendix C - Deep learning computer vision Image classification

An alternative way to collecting the player grip data using a computer vision deep learning transfer model (EfficientNet V2). While the manual approach was used, iwas tested to automatically classify forehand and backhand grips from player images. Code was written to use this approach with a pretrained transfer learning model (EfficientNet V2), however limited sample size and image quality resulted in a sub-standard validation accuracy of 73%, hence this approach was not further used, however work on this is recommended to allow for scalability in future applications of this work.

Appendix D – Backhand Type Grip Variable

The new variable factors in domain expertise which highlights the greater difference in backhand grip handling based on if a single-handed vs two handed backhand was used. E.g. there is a greater distance between single handed and double handed open grip than single handed and double handed closed grips in terms of playing styles. This variable replaces *Backhand Type* and *Dominant Hand Backhand.*

$$\text{Backhand Type Grip} = \frac{\text{Backhand Type}}{\text{Dominant Hand Backhand}}$$

Where Single Handed Backhand = -1, and Two-Handed Backhand = 1

Equation 15 Backhand Type Grip

Dashboard Implementation

A short video highlighting the use of the dashboard can be found at [Dashboard Example - Tableau Tennis - YouTube](https://www.youtube.com/watch?v=ANjPpg8lU9M) (https://www.youtube.com/watch?v=ANjPpg8lU9M)

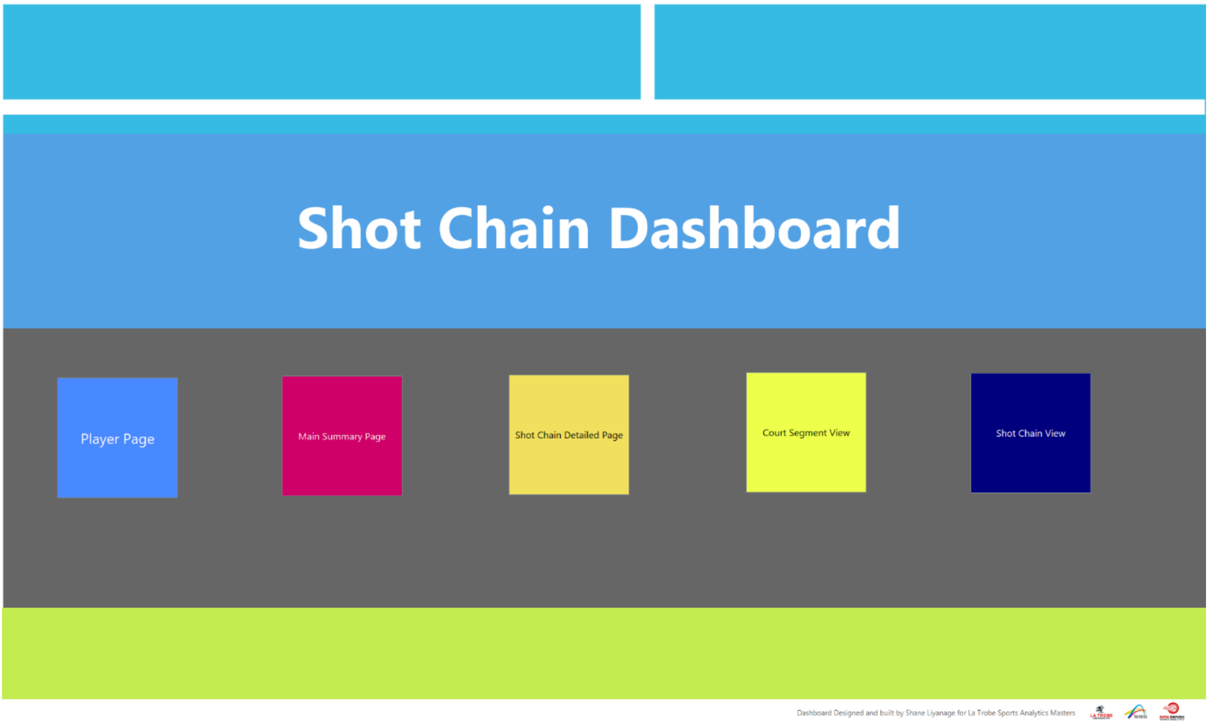


Figure 33 Shot Chain Dashboard Designed for Tennis Australia - Landing Page

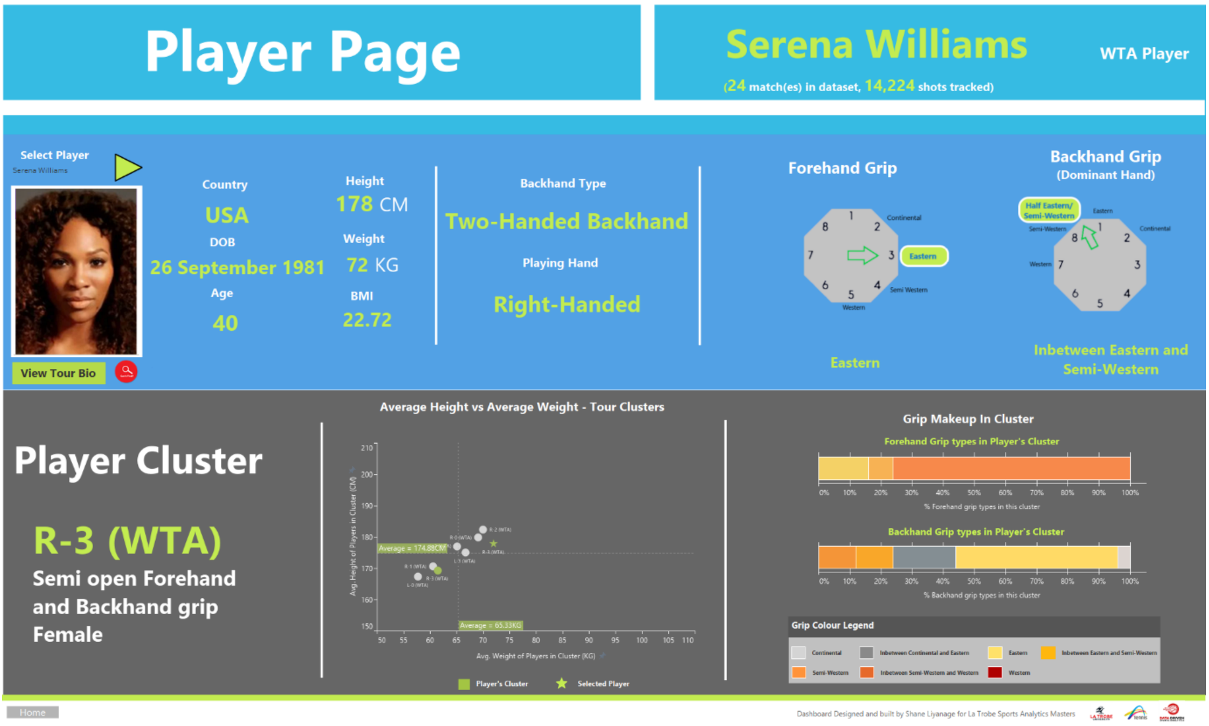


Figure 34 Shot Chain Dashboard Designed for Tennis Australia - Player Selection Page

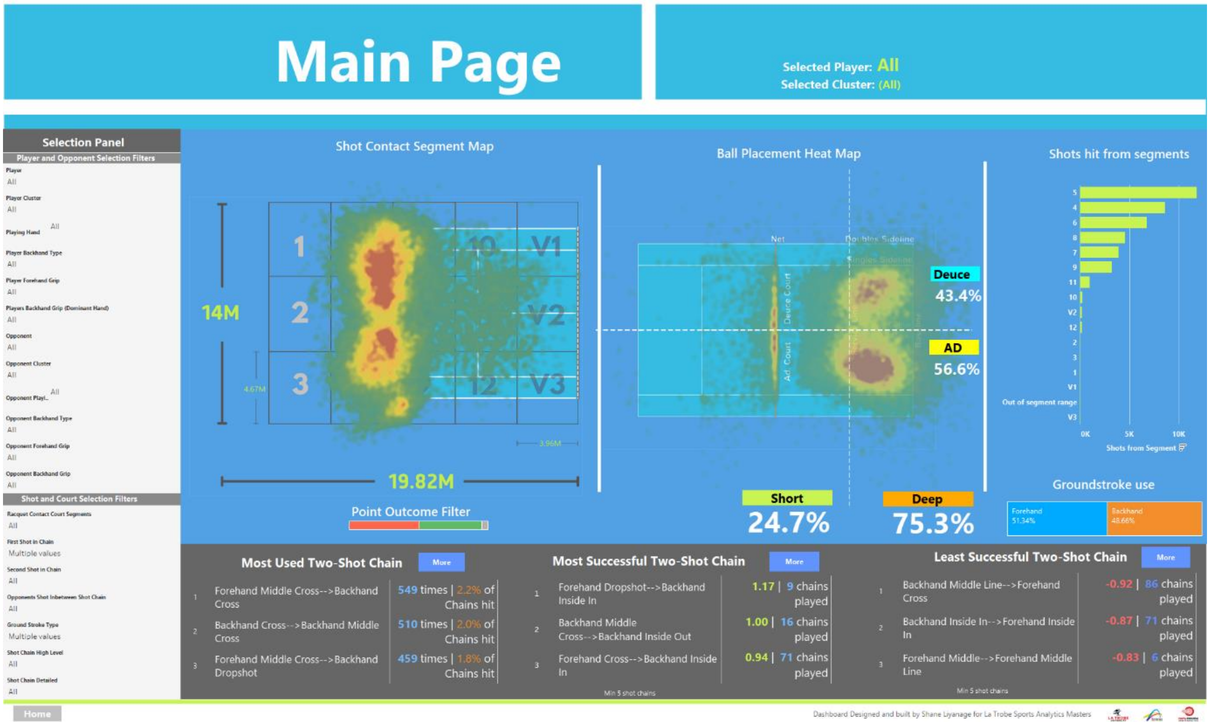


Figure 35 Shot Chain Dashboard Designed for Tennis Australia - Main Summary Page

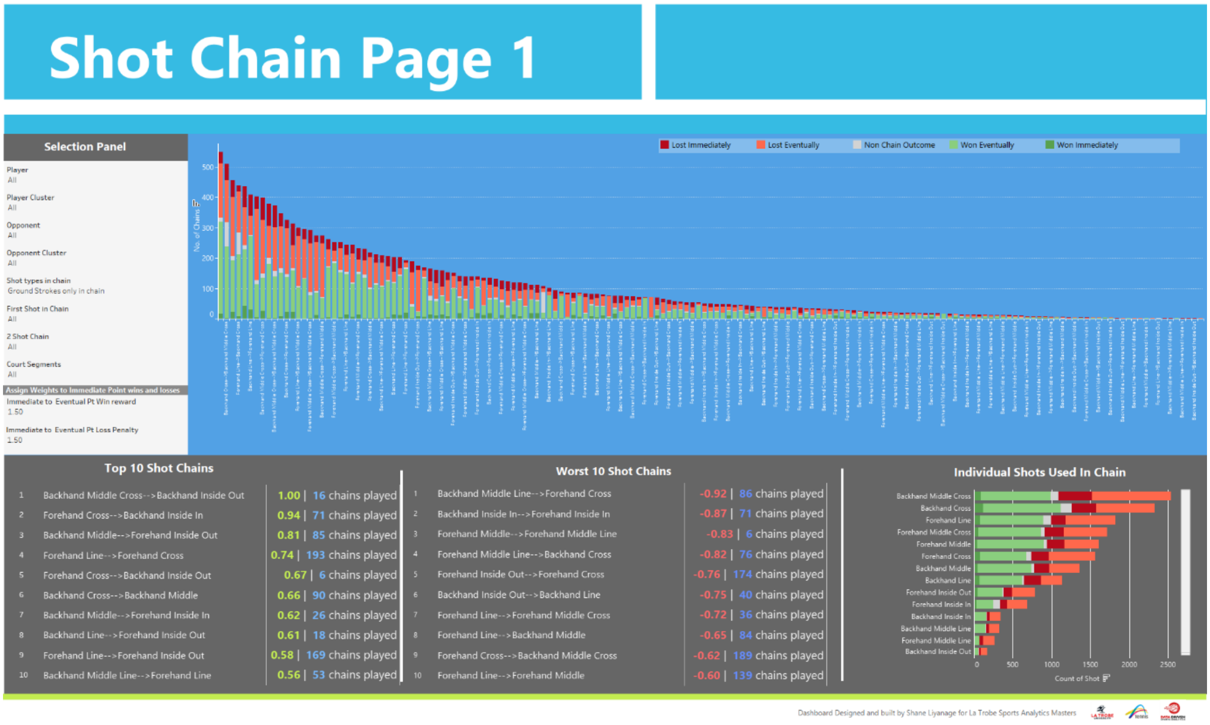


Figure 36 Shot Chain Dashboard Designed for Tennis Australia - Shot Chain Detailed Page

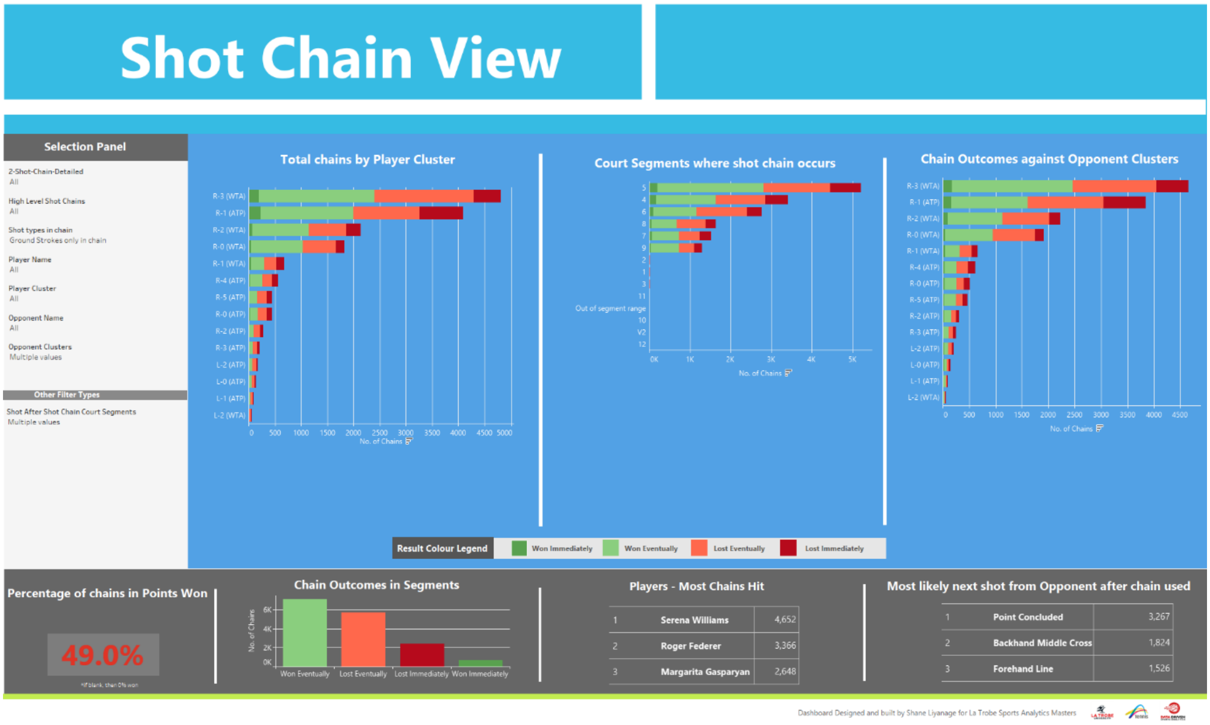


Figure 37 Shot Chain Dashboard Designed for Tennis Australia - Shot Chain Summary Page

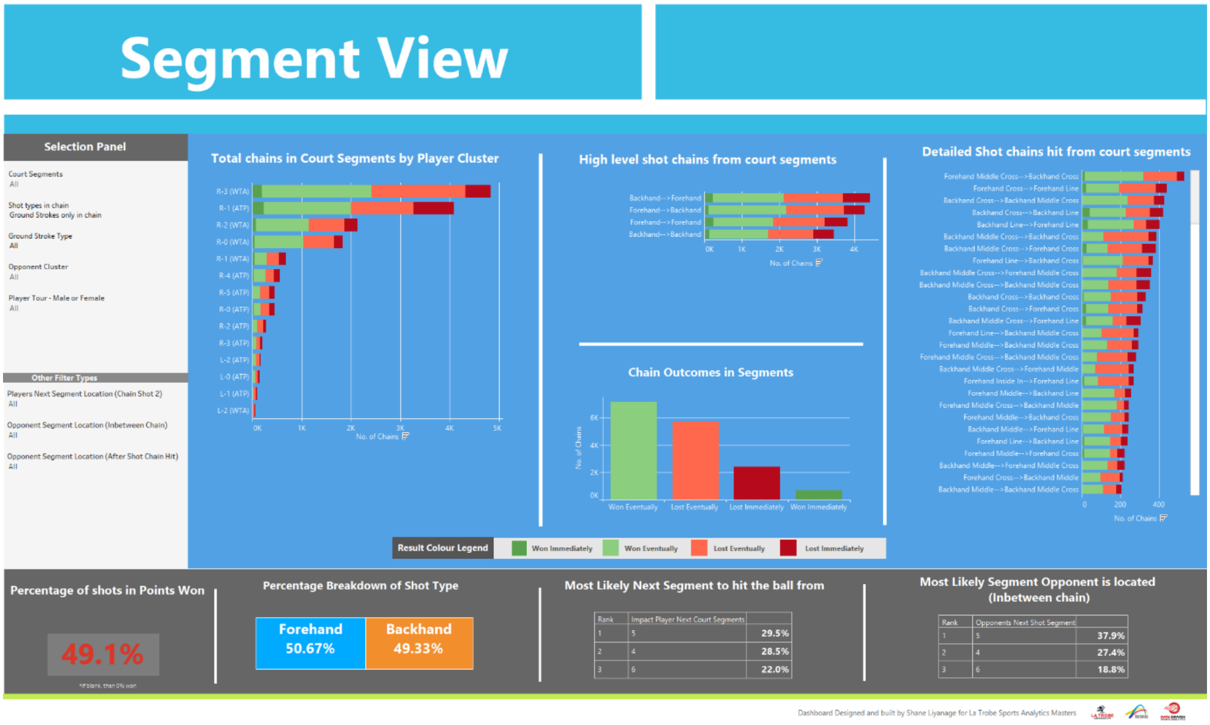


Figure 38 Shot Chain Dashboard Designed for Tennis Australia - Segment View

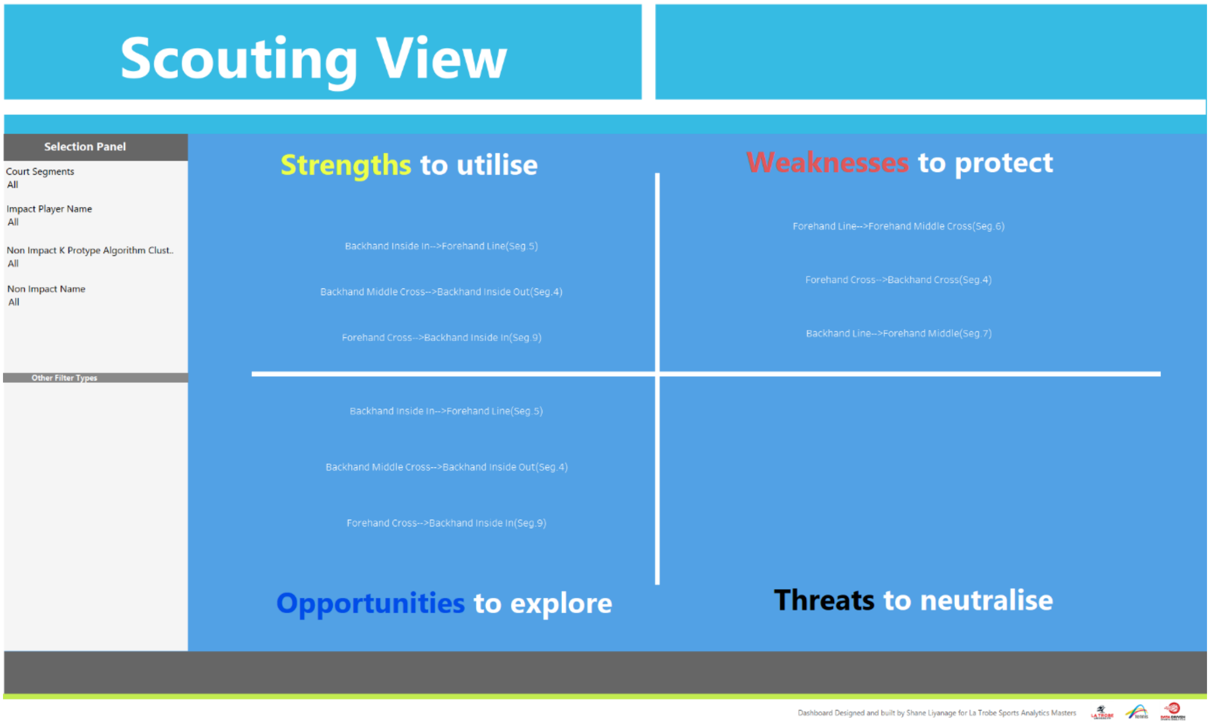


Figure 39 Shot Chain Dashboard Designed for Tennis Australia - Scouting Page

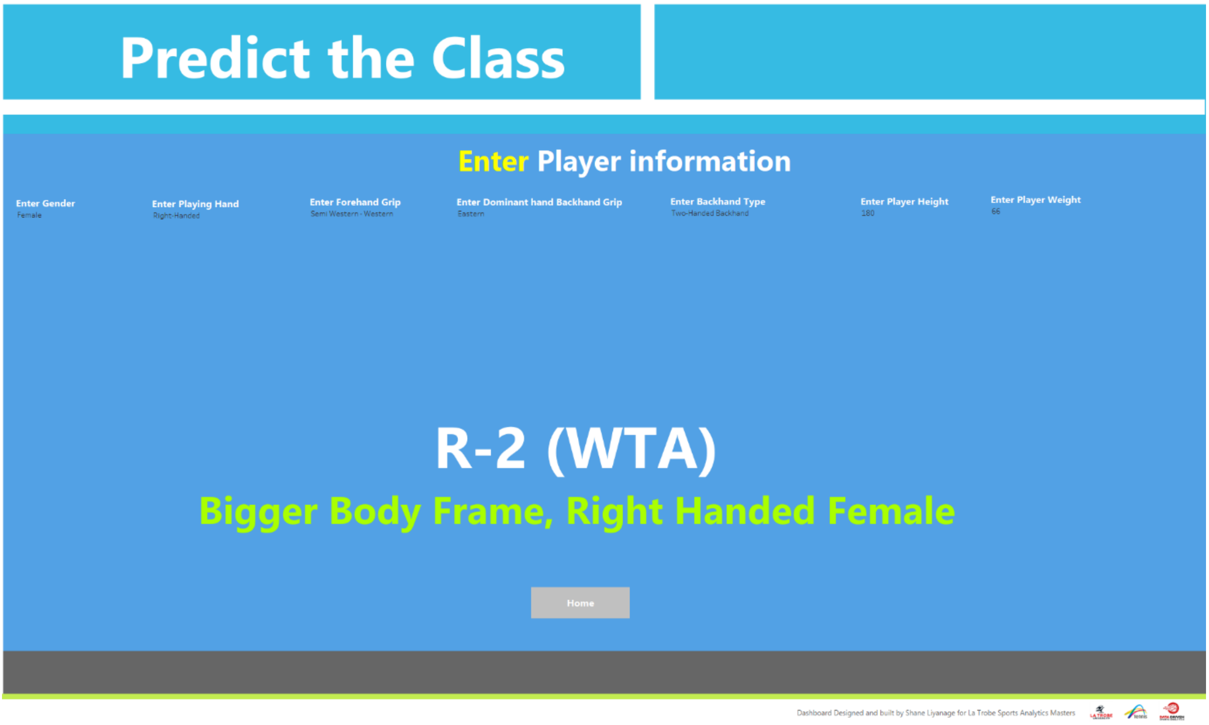


Figure 40 Shot Chain Dashboard Designed for Tennis Australia - Predict the Cluster for players not in database