An Analysis of Badminton

I. Executive Summary

Badminton has been an Olympic sport since 1992 and is most popular in Asia and some parts of Europe. China has seen remarkable success in the sport, taking home 20 gold medals (Olympics.com, 2022). Badminton is primarily a casual sport outside Asia and Europe, often played at family gatherings. Multiple sources call badminton the second most popular participatory sport in the world, after soccer, with an estimated 220 million players yearly (Where does badminton rank, 2022) (Clement, 2004). Although well-known in the United States, it receives less attention than other professional sports such as basketball and football.

Here is an overview of gameplay according to official rules. Players compete either in singles or doubles matches. A team or player wins a match by winning two games. Sometimes, they accomplish this in only two games, but often it takes three games. The first player or team to score 21 points wins the game unless the score is 21-20, in which case play continues until one player has a two-point lead. A player scores a point when an opponent fails to return the shuttlecock, also known as the birdie, from a valid shot. It is also possible to earn points if the opponent faults, for example, by making contact with the net. The point winner earns the next serve, as stated by the Badminton World Federation (2015).

The dataset consists of data for all BWF World Tour tournaments from 2018 to 2021. The HSBC (Hong Kong and Shanghai Banking Corporation) Badminton World Federation (BWF) World Tour, commonly known as BWF World Tour, attracts the world's top-ranked badminton men and women's singles and doubles. It features different tour types based on different levels of skill. From most prestigious to least prestigious, the tours are Super 1000 (four tournaments), Super 750 (six tournaments), Super 500 (nine tournaments), Super 300 (eleven tournaments), and Super 100 (BWF World Tour). The first to third-place winners in each tour earn points, allowing them to progress to higher-level competitions. The Super 1000 also includes a cash prize. The BWF Tour Super 100 serves as an entry point, and winners of this tour gain access to compete in the broader BWF World Tours (BWF World Tour).

The growing audience for badminton makes analytics for this sport relevant. The amount of prize money for the BWF World Tour now has a payout of $2 million compared to the $1.5 million payout five years ago (Mikkelsen, 2023). This upward trend in prize money may continue as the sport's popularity grows.

Another reason why this project is relevant is the relative dearth of existing analytics in the sport. Professional badminton player Shlok Ramchandran emphasized the need for enhanced analytics for badminton. Ramchandran acknowledged the absence of statistical breakdowns in badminton, like the number of smashes, a downward hit to the opponent's side, or points lost or won on serves (Ramchandran, 2022). Ramchandran believes implementing these sports statistics will provide insights on how to increase a player's chances of winning and offer badminton fans a more informative and engaging experience, further boosting badminton's popularity (Ramchandran, 2022).

Even though professional badminton has yet to gain popularity in the United States, there is potential for change. In September 2022, the USA Badminton and the Badminton World Federation announced a partnership to prepare for the 2028 Olympics. This collaboration aims to boost the sport's popularity and commercial growth in the United States (USA Badminton, 2022). USA Badminton CEO Linda French stated that their plan involves funding athletic opportunities for badminton in schools (USA Badminton, 2022). This partnership can potentially elevate the status of professional badminton in the United States, contributing to my interest in analyzing this data.

Although the analytics for badminton appear less developed compared to more popular sports, the sport's growing global appeal and its potential rise in North America highlight the increasing importance of my analysis. The dataset, though limited in player and game information, provides a foundation for future badminton analytics that can grow with additional match data. Even from this limited dataset, I have gained insights into the winners from different countries, the influence of homecourt advantage, and the significance of consecutive points during a game. These insights highlight the value of gathering additional data on players and games for future analysis.

II. Data Description

My project's data management journey involved several key steps. The ERD focused on the most crucial elements, like entities (the main data categories) and relationships between them. Subsequently, I collaborated to merge these individual CSVs into one, leveraging the best aspects from each design.

In designing the database, emphasis was on clarity and efficiency. I aimed to ensure that the data was organized in a way that was easy to understand and performs well when I run queries (searching for specific information). This involved making certain choices about how I store and link data. I opted for a balance between normalization, a method that minimizes data repetition, and query performance. I chose to normalize the database up to the 2nd Normal Form, which allows for efficient data retrieval while still maintaining some data consistency safeguards.

My database structure comprised three entities: 'tournament,' 'matches,' and 'player.' Each of these entities was interconnected to accurately represent how data related to one another. For example, I ensured that each match was associated with a specific tournament, and each match had a winner and a loser. Players could be linked to multiple matches as either winners or losers, creating distinct one-to-many relationships.

Furthermore, I attentively monitored data uniformity, including standardizing city names. After successfully normalizing my data, I loaded it into the database. A few issues surfaced with post-loading, such as incorrect dates and column order. These matters were promptly resolved, ensuring my data was error-free and ready for application in my project.

My database consists of three main entities: 'tournament,' 'matches,' and 'player,' each of which has a unique primary key. In the 'tournament' entity, 'tour_id' serves as the primary key, while the 'player' entity employs 'player_id' as its primary key. In the 'matches' entity, a composite primary key is used, which includes 'match_id,' 'tour_id,' 'winner_id,' and 'loser_id.' There are many characteristics of the data.

Data Integrity: That includes data integrity. It shows that primary keys play a critical role in maintaining data integrity by ensuring that each record is uniquely identifiable within its respective entity.

Relationships: The presence of 'tour_id,' in the 'matches' entity suggests relationships between entities. For example, 'tour_id' links matches to tournaments. These relationships allow for the representation of information like which players participated in which tournaments and the outcomes of matches. This easily shows the relationships between the entities.

Complexity: The use of a composite primary key in the 'matches' entity adds a level of complexity, which may require careful handling in database operations and queries.

Data Modelling: The structure is clear and straightforward, allowing for efficient data modeling and database design.

Flexibility: The design supports the flexibility of recording matches with various participants in different tournaments while maintaining data integrity and uniqueness.

See **Figure 1** for data dictionary.

I obtained the entire raw dataset from 2018 to 2021 from the Kaggle platform. From this dataset, I specifically selected men's and women's singles matches, excluding doubles and mixed matches, for my analytical purposes. It is worth noting that Juan Liong is the primary contributor to this dataset on Kaggle. I sourced this data from the official website of BWF tournaments, known for its comprehensive collection of badminton-related data (Liong, n.d.).

This section provides an overview of the decisions made regarding selecting and altering columns in the badminton dataset. This section explores why I removed specific columns and renamed others.

During data preparation and cleaning, I removed several columns from the dataset for specific reasons:

- I eliminated the 'number of sets' column because it was derived. I can compute the required information using a query, which reduces inaccuracy.
- I excluded the 'retired' column due to its inadequate information on retirements, which account for only around 1.5% of total matches, making it less significant for analysis.
- I removed the 'winner' column, which indicates the winning team, because I could derive it from the team game score columns.

During the data modification, I made several changes to enhance the dataset's consistency:

- I standardized the date column, converting all rows to the mm/dd/yyyy format to eliminate inconsistencies in the original dataset.

- I transformed the ‘team one player’ and ‘team two player’ columns to ‘winner ID’ and ‘loser ID’ to limit data explosion during normalization.
- I also renamed the ‘team game score’ columns to ‘winner game score’ and ‘loser game score’ to align logically with the previous Winner and Loser ID changes.

To further enhance the dataset’s quality, I carried out additional removals:

- I removed the columns for the team’s game points, the team’s most consecutive points, and the team’s total points. I made this decision because the data for the team’s game points did not accurately align with the derived team total points, eliminating potential inconsistencies.
- After removing the team’s game points, the column for the team’s most consecutive points became unnecessary.
- I also removed the draw column to prevent null values in the winner ID and loser ID columns in the match table. I had modified these columns to serve as foreign keys referencing the player table.

Using the final data set, I was interested in a few key insights. To investigate top players between 2018 and 2021 by matches won and then determining how many matches players overall. I was interested in seeing if the winners of each game had a higher streak of consecutive points. And, I was clear about my interest in determining if there was a home-court advantage and quickly identified many obstacles to presenting that based on the limitations of my dataset. Intrigued by the number of matches per year and how COVID shutting down most of the tournaments that year affected this. Lastly, I wanted to know the total number of matches played by each player.

III. Database Design

Given that the primary focus of my project is analytical rather than operational, I emphasized simplicity and clarity in my database design. I consciously opted for a trade-off, sacrificing some degree of normalization to enhance query performance. Specifically, I decided to normalize my database up to the 2nd Normal Form, minimizing data redundancy while still offering some protection against inconsistencies. I felt this was appropriate, especially since I anticipated minimal updates to the database post-initial data upload.

The database schema includes three tables: `tournament,`, `matches`, and `player.` Each tournament hosted one to many matches. Each match occurred at one and only one tournament. In each match, one player won, and one player lost[1]. Conversely, each player can be associated with multiple wins and losses, forming separate one-to-many relationships between `player_id` in the `player` entity and `Winner_id` and `Loser_id` in the `matches` entity. As seen in **Figure 2**, the `tournament` table is connected to the `matches` table via the foreign key `tour_id`.

IV. Data Analysis

For the first query, I found the top player between 2018 and 2021 by matches won and then determined how many matches players won overall. I found the count of p.player_id as WinCount sourcing from the players (p) table and joining matches (m) where winner_id = player_id. Additionally, I grouped the data by player names and sorted it in descending order by WinCount to identify the top players in the output. From the result shown in **Figure 3**, I exported to CSV and performed the visualizations shown in **Figures 4 and 5** attached in the Appendix. Using these visualizations, I displayed the difference in win counts amongst the top players and showed the number of players who achieved such scores versus the majority. There is a clear difference

[1] As mentioned previously. Draws were removed from the dataset.

between the vast majority of players, mostly scoring wins in under seven matches, versus the top 50 players scoring at least 38 wins.

In next query, I set out to determine whether a higher streak of consecutive points during the game would lead to a win. To achieve this, I created a subquery that labeled the winners of each game who had a higher streak as ‘Yes’ and ‘No’ if this was false. The subquery was then combined into a single table, making it easy to identify that the winners generally had a longer streak of consecutive points during the game. As seen in Figure 6, I ran this query for each game. The adjustment for game three involves score consideration. By incorporating the condition 'WHERE Winner_game_3_Score > 0', SQL only counted matches in which a game three occurred. The significance of this query is that it gives insights into predicting game outcomes based on who has the higher streak of consecutive points.

For each game, most winners had a higher streak of consecutive points than the losers. Additionally, the percentage of winners with a higher streak increased with each game. In game one, only 65% of winners had a higher streak. In game two, this increased to 70%. Finally, 75% of the winners in game three had a higher streak of consecutive points than the losers. I created a visual representation of the query to showcase findings more effectively (see **Figure 7**).

In another query, created analysis to explore tournament-type participation trends from 2018 to 2021. It involves selecting the year, along with the selection of the tournament types. I then computed the number of matches by counting the match IDs. The tournament ID column joined and linked the matches and tournament tables. Lastly, the results were grouped with the year and tournament type columns in the SELECT statement (See Appendix **Figures 8, 9,** and **10** for query and visualizations). This query forms the foundation for temporal analysis of badminton tournament participation trends.

The analysis showed that the BWF Tour Super 100 tournament featured 2,089 matches over four years, surpassing the HSBC BWF World Tour Finals by 240.12%. The BWF Tour Super 100 accounted for 32.21% of all matches, demonstrating its significance in badminton. While Super 100 tournaments are vital, higher-tier events like the BWF World Tour offer more rewards and prestige. The data also revealed a wide range of match counts across different tournament types, reflecting the sport's diversity and competitiveness.

However, the dataset also revealed the disruption brought about by the COVID-19 pandemic. In 2020, the HSBC BWF World Tour Finals faced an unusual interruption, leading to the tournament cancellation for that year as you can see in the **Figure 8** visualization. However, the sport bounced back as competitions resumed in 2021, highlighting its popularity and adaptability.

In my next analysis, I investigated whether badminton has a home-court advantage. For each tournament, I have calculated the ratio of home-country players to total players and the ratio of wins by home-country players to total matches (**Figure 12)**. For example, in the 2019 China Open, 13 of the 61 players were from China, approximately 20%. Chinese players won 12 of the 64 matches played during the tournament, about 20%. The ratio of home-country wins to home-country players is 1, suggesting no home-court advantage or disadvantage for this tournament. I then calculated the country-level average of these tournament-specific ratios to assess the presence of a home-court advantage or disadvantage across multiple tournaments hosted in the same country.

Some countries, specifically Japan, Denmark, China, and France, demonstrate a competitive edge when playing on their home courts. However, more countries appear to underperform in their home tournaments, indicating that factors other than location likely influence the outcomes.

For my last query I chose to investigate the total matches played by each player. The count of p.player_id as TotalMatchesPlayed taken from the players (p) table and joining matches (m) where player_id = winner_id OR player_id = loser_id, in addition I grouped by the player names and then ordered by winner_id, so then I get total matches played by each player.

From the result shown in the, I exported to csv and performed the visualizations shown in **Figure 14** and **Figure 15**. Using these visualizations, I was hoping to display the difference in count amongst the total games played by each player. The highest number of plays was 104 and lowest number of plays was 32. The query is designed to find out how many matches each player in a database has participated in, whether they won or lost. It counts the total number of matches for each player and shows their names. Total matches played by each player is important because it helps assess player performance, identify talent, and provides valuable insights for players, coaches, fans, and other stakeholders in the world of badminton.

References

Badminton World Federation. (2015). Simplified Rules of Badminton. BWF. https://system.bwfbadminton.com/documents/folder_1_81/Regulations/Simplified-Rules/Simplified%20Rules%20of%20Badminton%20-%20Dec%202015.pdf

Clement, B. (2004, July 23). *Badminton second to soccer in participation worldwide*. ESPN. https://www.espn.com/olympics/summer04/badminton/news/story?id=1845228

HSBC. (n.d.). *BWF World Tour*. HSBC BWF World Tour. https://bwfworldtour.bwfbadminton.com/about/

Liong, J. (n.d.). Analyzing Tennis Match Data: A Focus on Men's and Women's Singles (2018-2021) from Kaggle. https://www.kaggle.com/datasets/sanderp/badminton-bwf-world-tour

Mikkelsen, S. (2023, August 24). *What has impacted the way badminton is played today? Olympics.com takes a look at the development of the sport.* Badminton's Evolution: A historical glimpse into the sport's changing face over the decades. https://olympics.com/en/news/badminton-evolution-historical-glimpse-sport-over-decades

Olympics.com. (2022, January 1). *How to play badminton: Rules, scoring system and equipment - Olympics.com*. Badminton: How to play, rules, and all you need to know. https://olympics.com/en/news/badminton-guide-how-to-play-rules-olympic-history

Ramchandran, S. (2022, June 30). *Shuttle zone: Badminton could really use a data revolution, starting from broadcast to coaching*. Scroll.in. https://scroll.in/field/1027235/shuttle-zone-badminton-could-really-use-a-data-revolution-starting-from-broadcast-to-coaching

USA Badminton. (2022, September 15). *USA Badminton announces LA 2028 Legacy partnership with BWF - USA Badminton*. USA Badminton announces LA 2028 legacy partnership with BWF. https://usabadminton.org/usa-badminton-announces-la-2028-legacy-partnership-with-bwf/

Where does badminton rank in popularity among other sports? [answered]. Where Does Badminton Rank in Popularity Among Other Sports? [Answered] – More Than Sport. (2022, July 28). https://www.morethansport.com/where-does-badminton-rank-in-popularity-among-other-sports-answered/

Appendix

Figure 1: Data Dictionary

Column Name	Definition	Data Type	Keys
Tournament			
tour _id	This column represents a unique identifier for each tournament.	INT	Primary key
tournament	The name or title of the tournament.	VARCHAR 64	
country	The country where the tournament is held.	VARCHAR 32	
city	The city where the tournament is located.	VARCHAR 32	
tour_type	The type or category of the tournament	VARCHAR 64	
Match			
match_id	A unique identifier for each match.	INT	Primary key
tour_id	An identifier for the tournament or event where the match took place.	INT	Foreign key
Winner_id	The unique identifier of the winning player or team.	INT	Foreign key
Loser_id	The unique identifier of the losing player or team	INT	Foreign key
discipline	The type of sport or game discipline in which the match was played.	VARCHAR 32	
round	The round or stage of the tournament in which the match occurred.	VARCHAR 32	
date	The date on which the match was played.	DATE	
Winner_most_consecutive_points_game_1	The highest streak of consecutive points won by the winner in the first game.	INT	
Loser_most_consecutive_points_game_1	The highest streak of consecutive points won by the loser in the first game.	INT	
Winner_most_consecutive_points_game_2	The highest streak of consecutive points won by the winner in the second game.	INT	
Loser_most_consecutive_points_game_2	The highest streak of consecutive points won by the loser in the second game.	INT	
Winner_most_consecutive_points_game_3	The highest streak of consecutive points won by the winner in the third game.	INT	
Loser_most_consecutive_points_game_3	The highest streak of consecutive points won by the loser in the third game.	INT	
Winner_game_1_score	The score achieved by the winner in the first game.	INT	
Loser_game_1_score	The score achieved by the loser in the first game.	INT	
Winner_game_2_score	The score achieved by the winner in the second game.	INT	
Loser_game_2_score	The score achieved by the loser in the second game.	INT	
Winner_game_3_score	The score achieved by the winner in the third game.	INT	
Loser_game_3_score	The score achieved by the loser in the third game.	INT	
Player			
player_id	A unique identifier for each player	INT	Primary key
name	This column stores the name of the player	VARCHAR 32	
nationality	This column stores the nationality of the player	VARCHAR 32	

Figure 2: ERD

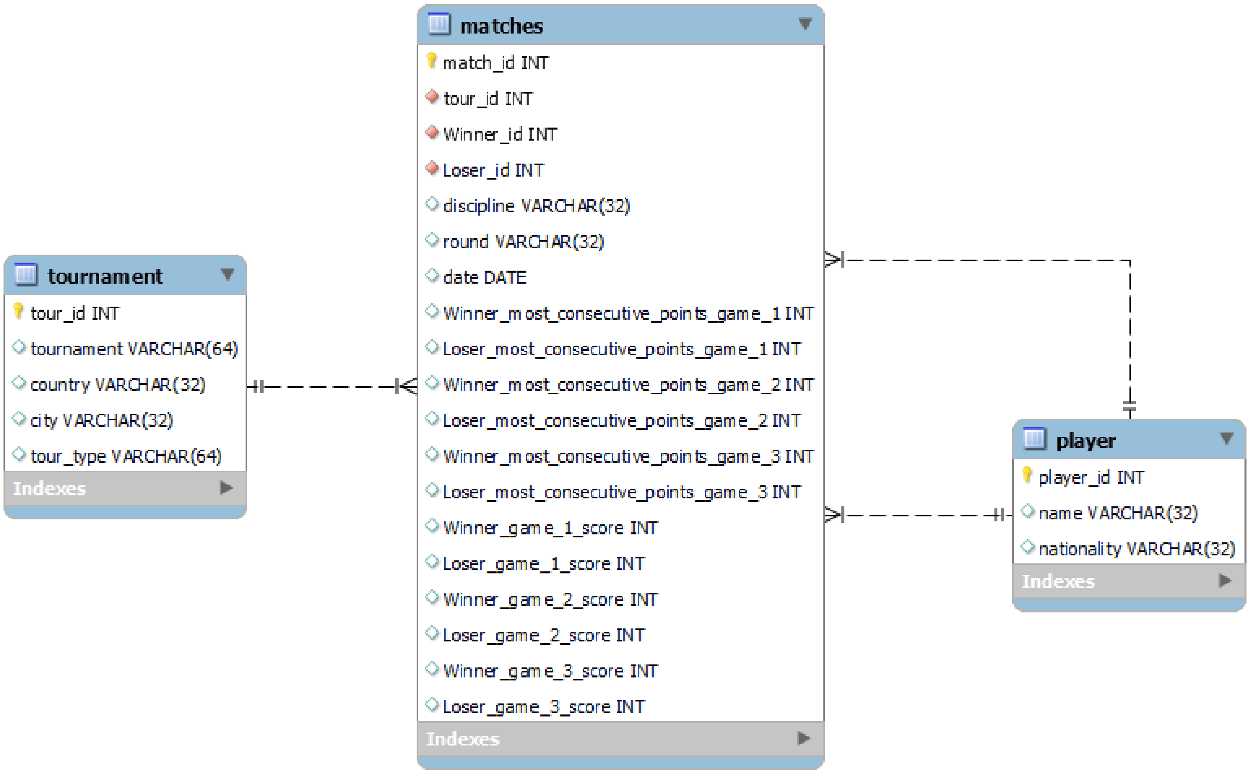
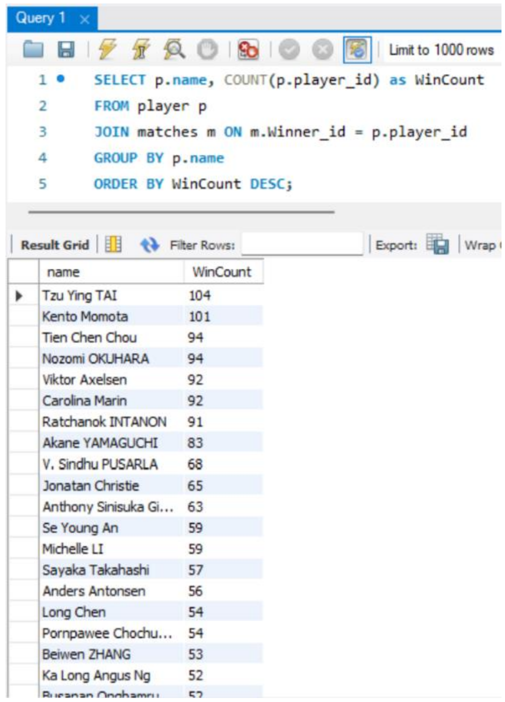
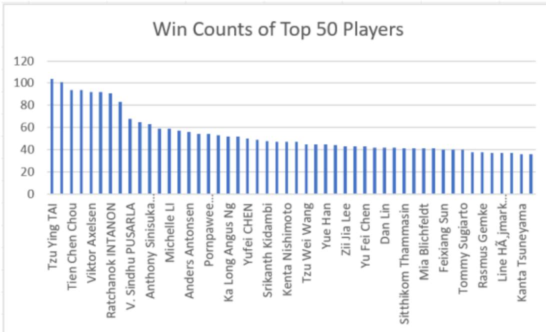

matches
match_id INT
tour_id INT
Winner_id INT
Loser_id INT
discipline VARCHAR(32)
round VARCHAR(32)
date DATE
Winner_most_consecutive_points_game_1 INT
Loser_most_consecutive_points_game_1 INT
Winner_most_consecutive_points_game_2 INT
Loser_most_consecutive_points_game_2 INT
Winner_most_consecutive_points_game_3 INT
Loser_most_consecutive_points_game_3 INT
Winner_game_1_score INT
Loser_game_1_score INT
Winner_game_2_score INT
Loser_game_2_score INT
Winner_game_3_score INT
Loser_game_3_score INT
Indexes
tournament
tour_id INT
tournament VARCHAR(64)
country VARCHAR(32)
city VARCHAR(32)
tour_type VARCHAR(64)
Indexes
player
player_id INT
name VARCHAR(32)
nationality VARCHAR(32)
Indexes

Figure 3: SQL Query



WinCount SQL Code:
SELECT p.name, COUNT(p.player_id) as WinCount
FROM player p
JOIN matches m ON m.Winner_id = p.player_id
GROUP BY p.name
ORDER BY WinCount DESC;

Figure 4: Win Count by Player and Figure 5: Histogram of WinCount:



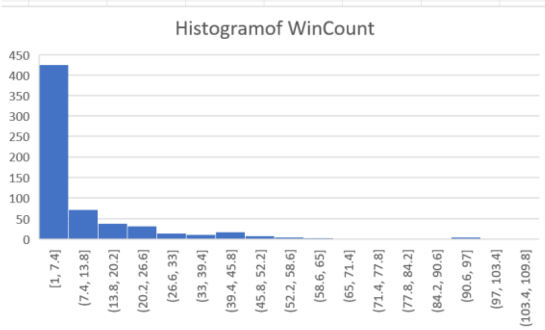


Figure 6: SQL Query

```
-- For Game 1
SELECT HigherStreakConsecPtsLeadToWinG1,
COUNT('Yes''No') AS Count
FROM (SELECT CASE WHEN Winner_most_consecutive_points_game_1 >
Loser_most_consecutive_points_game_1
THEN 'Yes' ELSE 'No'
END AS HigherStreakConsecPtsLeadToWinG1
FROM matches) s GROUP BY HigherStreakConsecPtsLeadToWinG1;
-- For Game 2
SELECT HigherStreakConsecPtsLeadToWinG2,
COUNT('Yes''No') AS Count
FROM (SELECT CASE WHEN Winner_most_consecutive_points_game_2 >
Loser_most_consecutive_points_game_2
THEN 'Yes' ELSE 'No'
END AS HigherStreakConsecPtsLeadToWinG2
FROM matches) s GROUP BY HigherStreakConsecPtsLeadToWinG2;
-- For Game 3
SELECT HigherStreakConsecPtsLeadToWinG3,
COUNT('Yes''No') AS Count
FROM (SELECT CASE WHEN Winner_most_consecutive_points_game_3 >
Loser_most_consecutive_points_game_3
THEN 'Yes' ELSE 'No'
END AS HigherStreakConsecPtsLeadToWinG3, Winner_game_3_Score
```

FROM matches) s
WHERE Winner_game_3_Score > 0
GROUP BY HigherStreakConsecPtsLeadToWinG3;

Figure 7: Higher Streak of Consecutive Points Lead to a Win Visualization

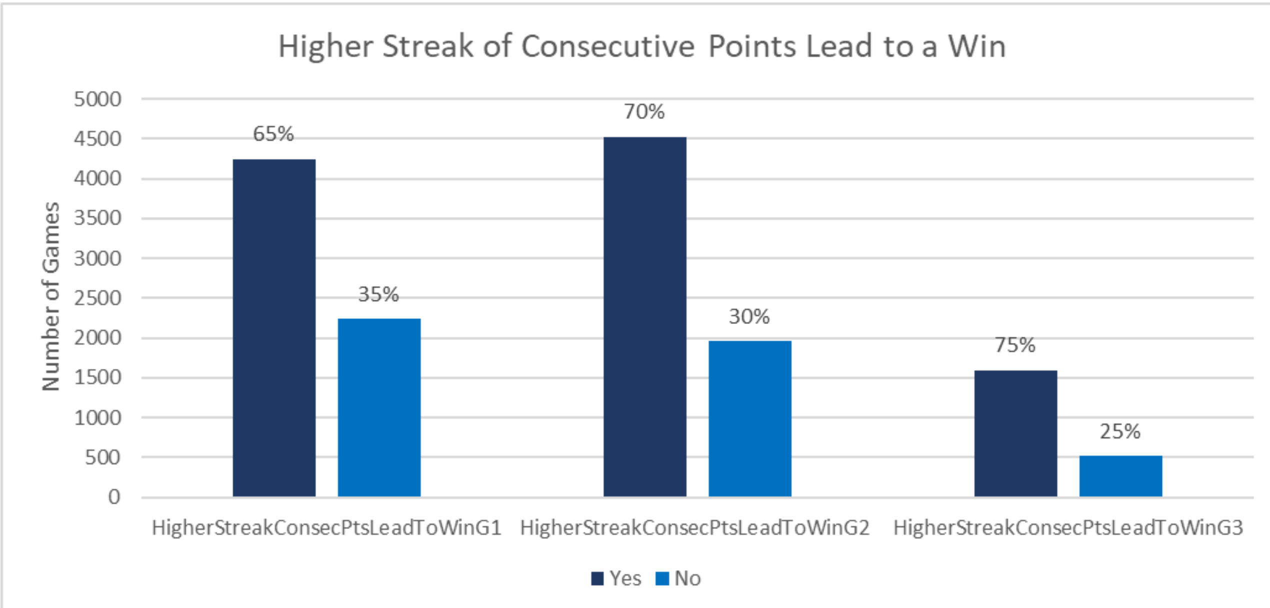


Figure 8: SQL Query

```
select
YEAR(matches.date) as years,
tournament.tour_type as tournament_types,
count(matches.match_id) as matches_conducted
from matches
join
tournament on matches.tour_id = tournament.tour_id
group by
1,2
order by
tournament.tour_type
```

Result Grid | Filter Rows: | Export: | Wrap Cell Cont

years	tournament_types	matches_conducted
2018	BWF Tour Super 100	1035
2019	BWF Tour Super 100	874
2021	BWF Tour Super 100	106
2020	BWF Tour Super 100	74
2019	HSBC BWF World Tour Finals	29
2018	HSBC BWF World Tour Finals	28
2021	HSBC BWF World Tour Finals	30
2020	HSBC BWF World Tour Super 1000	60

Figure 9: Participation Trends in Badminton from 2018 to 2018

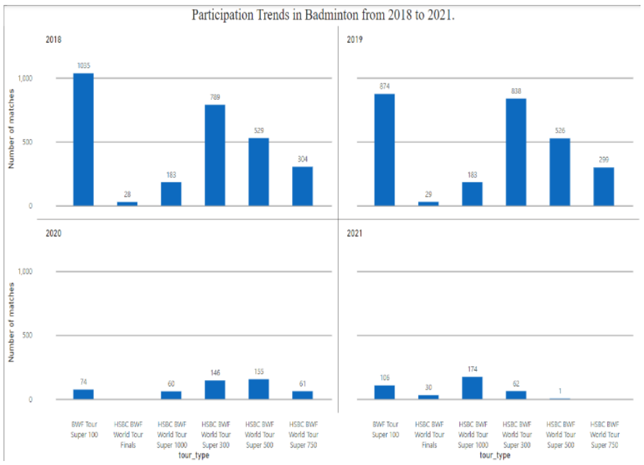


Figure 10: Overall Participation Trends in Tournament Types on Span of Four Years

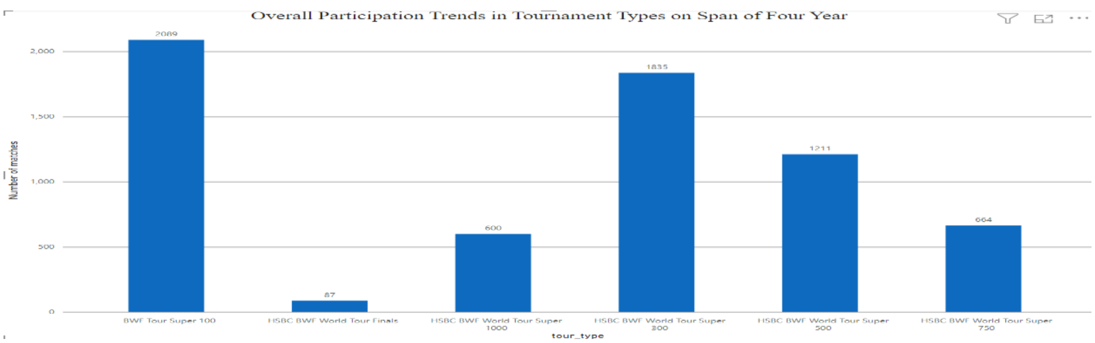


Figure 11: SQL Queries

Home-court advantage query:

```
SELECT sub.country,
   ROUND(AVG(sub.ratio_proportion_home_wins_to_players),2)
avg_ratio_home_wins_to_home_players
FROM (SELECT t.country,
      CASE
         WHEN COUNT(DISTINCT CASE WHEN p.nationality = t.country
THEN p.player_id ELSE NULL END) / COUNT(DISTINCT p.player_id) = 0 OR
            SUM(CASE WHEN p.nationality = t.country AND m.Winner_id = p.player_id
THEN 1 ELSE 0 END) / COUNT(DISTINCT m.match_id) = 0 THEN NULL
         ELSE (SUM(CASE WHEN p.nationality = t.country AND m.Winner_id =
p.player_id
THEN 1 ELSE 0 END) / COUNT(DISTINCT m.match_id)) /
             (COUNT(DISTINCT CASE WHEN p.nationality = t.country
THEN p.player_id ELSE NULL END) / COUNT(DISTINCT p.player_id))
      END AS ratio_proportion_home_wins_to_players
   FROM tournament t
   JOIN matches m ON t.tour_id = m.tour_id
```

```
    JOIN (SELECT DISTINCT Winner_id AS player_id
FROM matches UNION
SELECT DISTINCT Loser_id AS player_id FROM matches) participants
        ON (m.Winner_id = participants.player_id OR m.Loser_id = participants.player_id)
    JOIN player p ON p.player_id = participants.player_id
    GROUP BY t.tour_id, t.country) AS sub
GROUP BY sub.country
ORDER BY AVG(sub.ratio_proportion_home_wins_to_players) DESC;
```

Wins by nationality query:

```
SELECT
    p.nationality,
    COUNT(m.Winner_id) AS total_wins,
    ROUND(SUM(COUNT(m.Winner_id)) OVER (ORDER BY COUNT(m.Winner_id)
DESC ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW) /
(SELECT COUNT(*) FROM matches) * 100, 2) AS cumulative_percent
FROM
    matches m
JOIN
    player p ON m.Winner_id = p.player_id
GROUP BY
    p.nationality
ORDER BY
    total_wins DESC;
```

Figure 12: Home-court advantage visualization

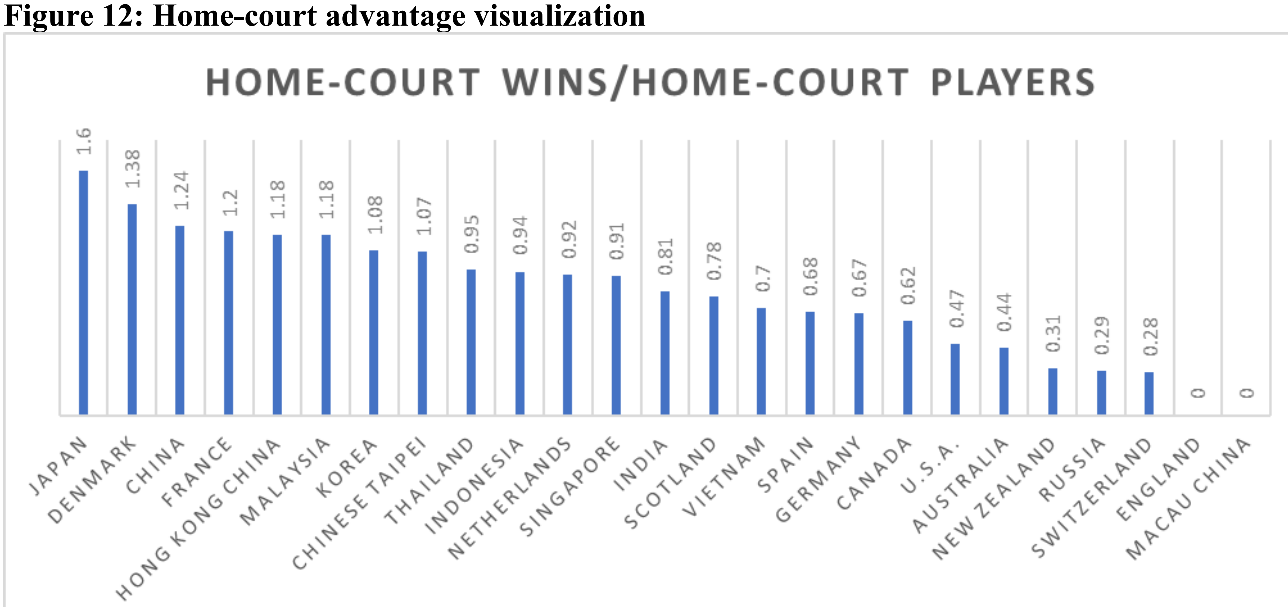
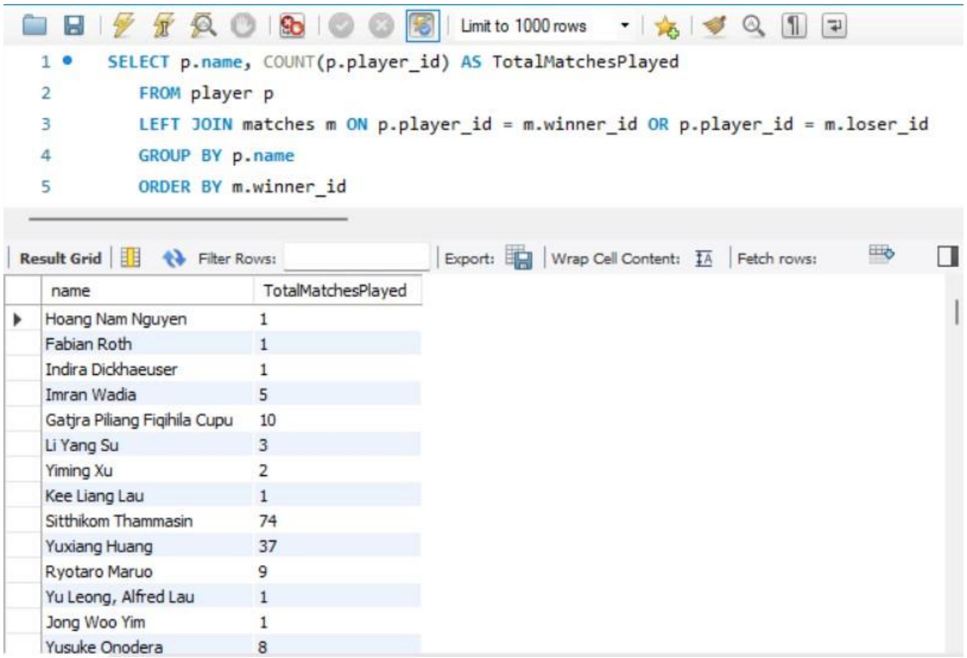


Figure 13: Wins by Nationality

nationality	total_wins	cumulative_percent
China	935	14.42
Japan	782	26.47
India	667	36.76
Indonesia	559	45.37
Thailand	540	53.70
Chinese Taipei	499	61.39
Denmark	438	68.15
Korea	367	73.81
Malaysia	311	78.60
France	180	81.38
Hong Kong C...	173	84.04
Canada	130	86.05
Spain	128	88.02
U.S.A.	98	89.53
Germany	89	90.90
England	66	91.92
Singapore	58	92.82
Russia	58	93.71
Netherlands	49	94.47
Finland	33	94.97
Scotland	32	95.47
Ireland	31	95.95
Vietnam	29	96.39

Figure 14: SQL Query



```
SELECT p.name, COUNT(p.player_id) AS TotalMatchesPlayed
   FROM player p
   LEFT JOIN matches m ON p.player_id = m.winner_id OR p.player_id = m.loser_id
   GROUP BY p.name
   ORDER BY m.winner_id
```

name	TotalMatchesPlayed
Hoang Nam Nguyen	1
Fabian Roth	1
Indira Dickhaeuser	1
Imran Wadia	5
Gatjra Piliang Fiqihila Cupu	10
Li Yang Su	3
Yiming Xu	2
Kee Liang Lau	1
Sitthikom Thammasin	74
Yuxiang Huang	37
Ryotaro Maruo	9
Yu Leong, Alfred Lau	1
Jong Woo Yim	1
Yusuke Onodera	8

SELECT p.name, COUNT(p.player_id) AS TotalMatchesPlayed

FROM player p

LEFT JOIN matches m ON p.player_id = m.winner_id OR p.player_id = m.loser_id

GROUP BY p.name

ORDER BY m.winner_id

Figure 15: Count of match_id by name