Uncovering the Secrets of the 2024 ATP

Data analysis project

By:

Mahdieh Alavi

Zahra Mahmoodabadi

Introduction

- The **2024 ATP** is one of the most prestigious tennis tournaments on the ATP Tour
- Deep into the data and uncover the hidden stories behind the matches played at this event.

Exploring Players Profile and participated countries Information

- Let's start by taking a closer look at the players who competed in the tournament

- The data reveals that the tournament featured a diverse mix of players, including stars like **Ujichima**.
- Players from **80** different **countries** participated in this competition

- The total number of players played in the tournament was **2352.**

- Interestingly, the average height of the players is **1.82** meters,

- By analyzing the players' total prize money, we can see that **Djokovic**, **Novak, Nadal Rafael** and **Murray, Andy** have the 3 highest career earnings

full_name	total_prize
Djokovic, Novak	455008686
Nadal, Rafael	224653134
Murray, Andy	107560350

This slide covers the answers of question numbers 1, 2, 13, Ex. Q. 3

…Exploring Players Profile Information

- The player with the most winnings in matches was **Uchijima M**. with player id 253356 from Japan. The number of winnings for the player was **15 times.**

- The player who wins the most tournaments was **Paquet C.**

- The player with the highest winning percentage against top 10 ranked opponents is: **Świątek I.** with 16.6 percent win
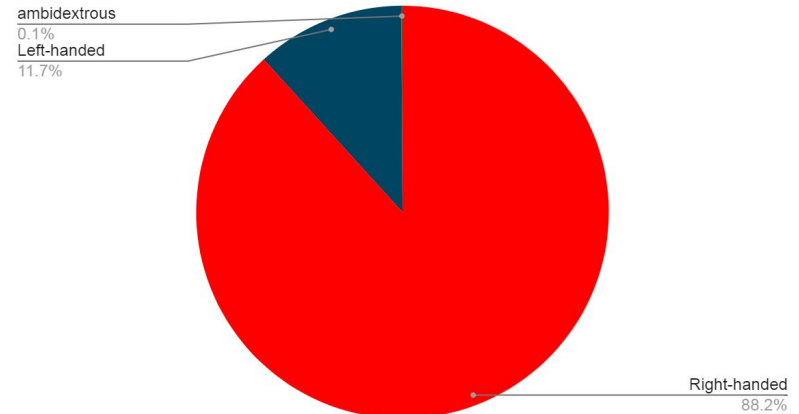
player_name	no_win_against_top10_rank_players
Świątek I.	3
Rublev A.	1
Cerundolo F.	1
Lehečka J.	1

This slide covers the answers of question numbers 3, 15, 16

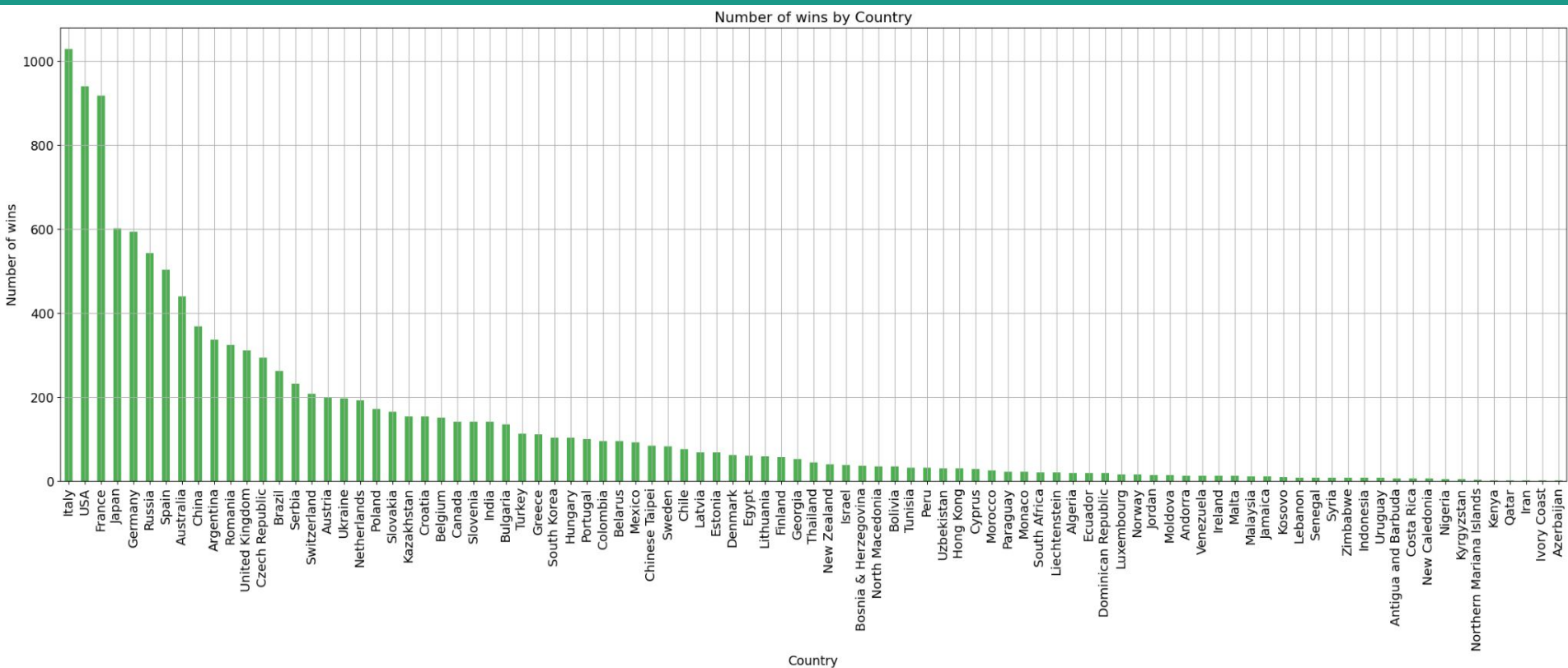## …Exploring Players Profile Information

- A range of playing styles, including both right-handed and left-handed players.
  - right-handed **88.24%**
  - left-handed **11.66%**
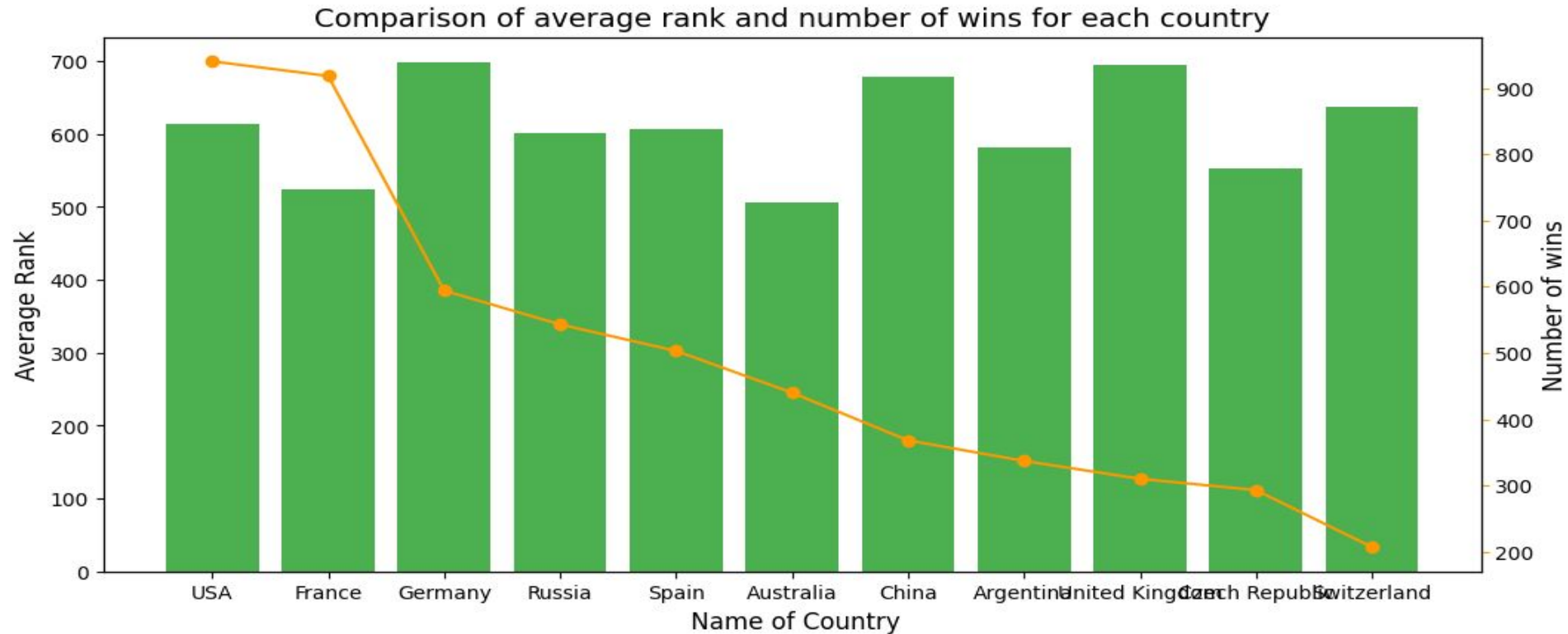 - ambidextrous **0.09%**

playing style

ambidextrous
0.1%
Left-handed
11.7%

Right-handed
88.2%

This slide covers the answers of question numbers 9

…Exploring Players Profile Information

- If most successful means **most winning** ,The country which produced the most successful tennis players was Italy with maximum number of wins .

Number of wins by Country

…Exploring Players Profile Information

- According to the below diagram,If most successful means **most winning and top ranking, USA** with maximum number of wins and average rank is the country which produced the most successful tennis players .

Comparison of average rank and number of wins for each country

…Exploring Players Profile Information

- The country with the most number of players is **USA** with **208** players (`Table 2)`
- The number of winning matches for each player has been shown in `Table 1`

player_name	number_of_winns
Uchijima M.	**15**
Sherif M.	13
Sun F.	12
Urhobo A.	12
Wiskandt M.	12

Table 1

country	number_of_players
USA	**208**
Italy	181
France	153
Japan	122
Russia	118

Table 2

This slide covers the answers of question numbers Ex. Q. 5, Ex. Q. 6
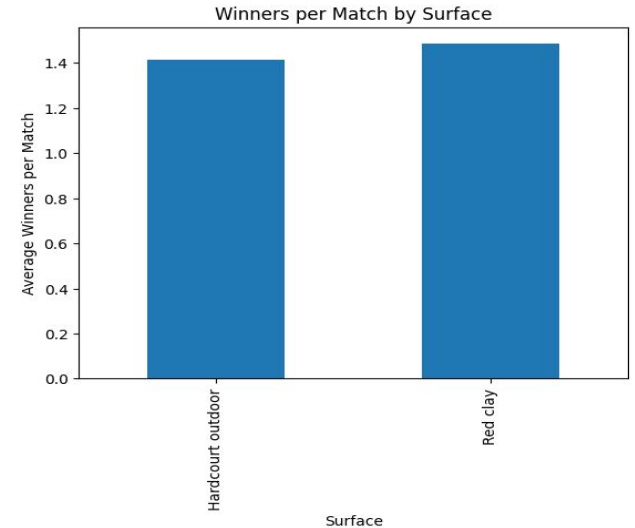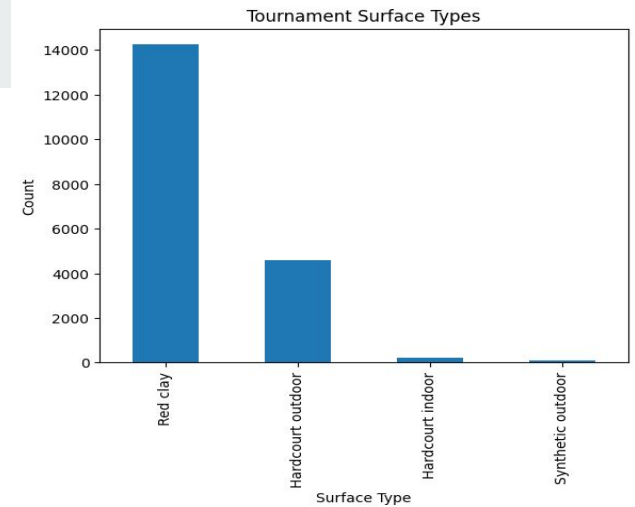
Analysing Match Duration

- According to the given information ,the longest match recorded in terms of duration is the match with match-id: **12346747** (47.54 h) and between players:**Vulpitta G. & Pieri S.**

match_id	period_1	period_2	period_3	period_4	period_5	Current_period_start_timestamp	Sumation_time_periods_match
12346747	167761	3392	0	0	0	1.72E+09	171153

- The average duration of matches was **182.77 minutes**.

This slide covers the answers of question numbers 4, 11
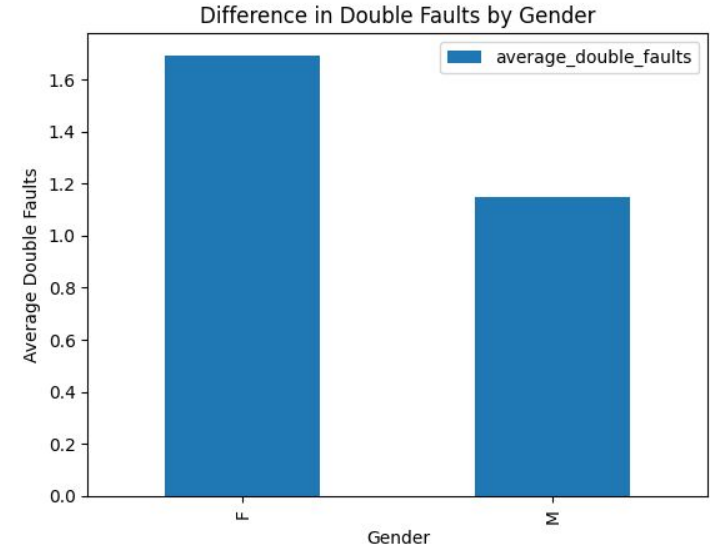
Exploring Tournament Dynamic

- Now, let's take a closer look at the tournament dynamics and how ground type may have influenced the outcomes.

- The data shows that the tournament was played entirely on **Red clay courts**, which is known to favor certain playing styles and strategies.

- The average number of winners per match differ between hard and clay court surfaces

- Results shows that **'Red clay'** ground provides **more winners** in compare with 'HardCourt outdoor'.

This slide covers the answers of question numbers 14, Ex Q. 4

10

Uncovering Insights from the Scoreboard

- Finally, let's dive into the scoreboard data and see what insights we can uncover about the team performances.

- The average number of **aces per match** was **6.71**

- There is a difference in the number of **double faults** based on gender, as we can see in the following **plot**

Difference in Double Faults by Gender

This slide covers the answers of question numbers 7, 8

…Uncovering Insights from the Scoreboard

- Most matches typically had **2 sets** played in a tennis match. Some of matches played 3 sets.

```
match_id
12260075     2
12260076     3
12260077     3
12260078     3
12260080     2
             ..
12384789     2
12384806     3
12384892     2
12384975     3
12385017     3
Name: set_num,
Length: 6658
```

This slide covers the answers of question numbers 5
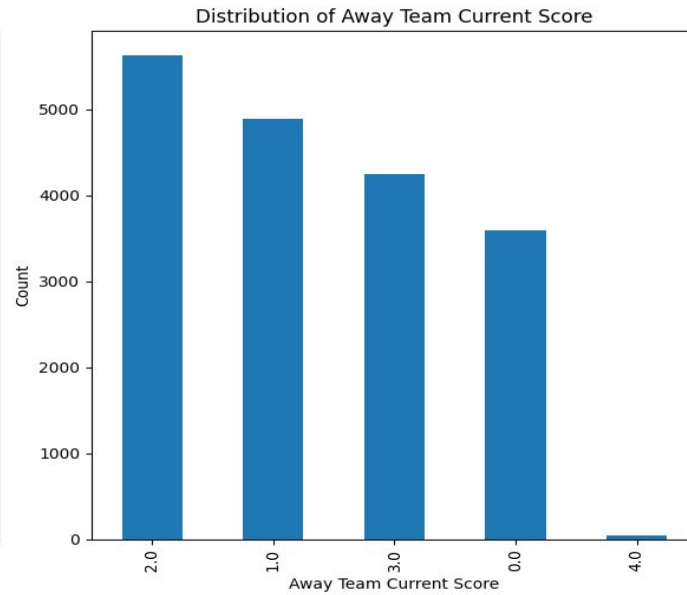
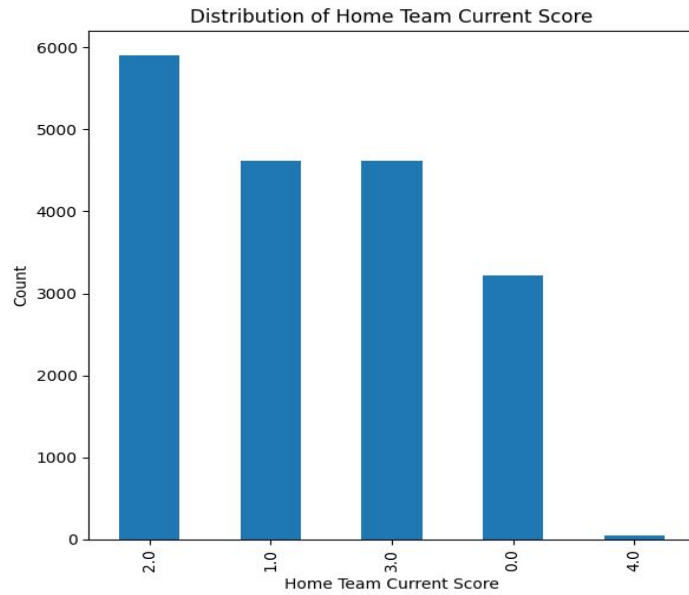## …Uncovering Insights from the Scoreboard

- The average number of **games per set** in **men's matches** was **9.18**, while this value was **8.92** in **women's matches**.

- The average number of **breaks** per match was **16.37**

This slide covers the answers of question numbers 12, 17
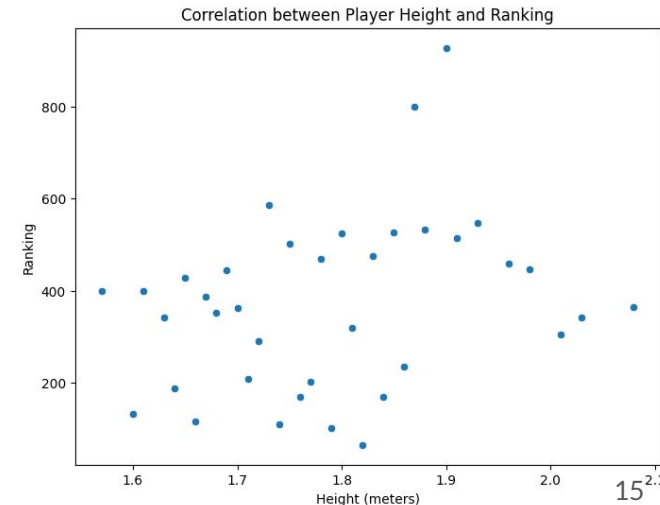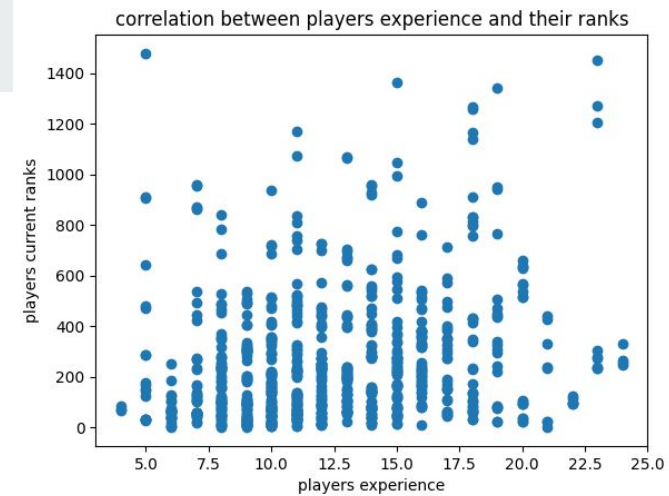
…Uncovering Insights from the Scoreboard

● The distribution of the **score values** in the home and away teams

This slide covers the answers of question numbers Ex. Q. 8
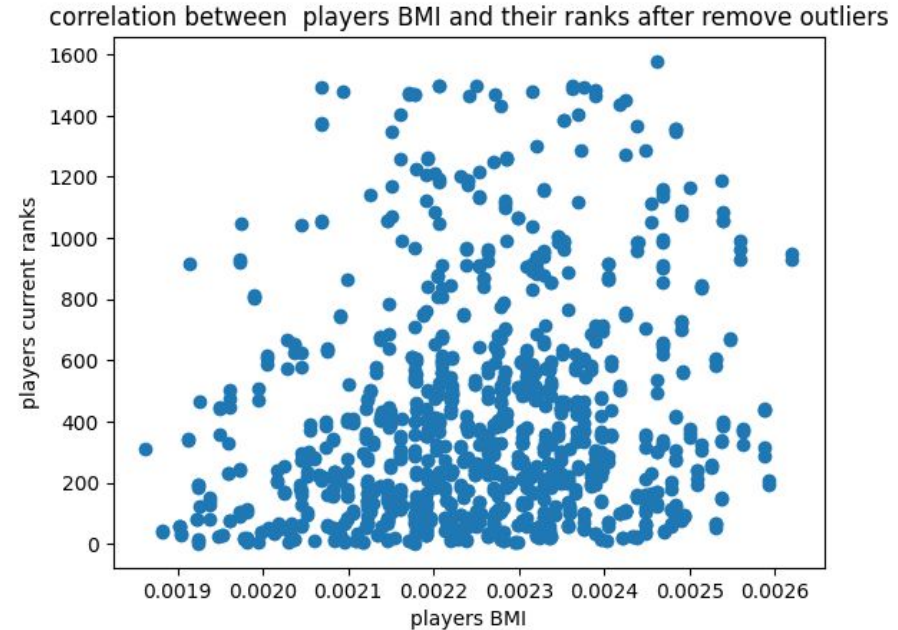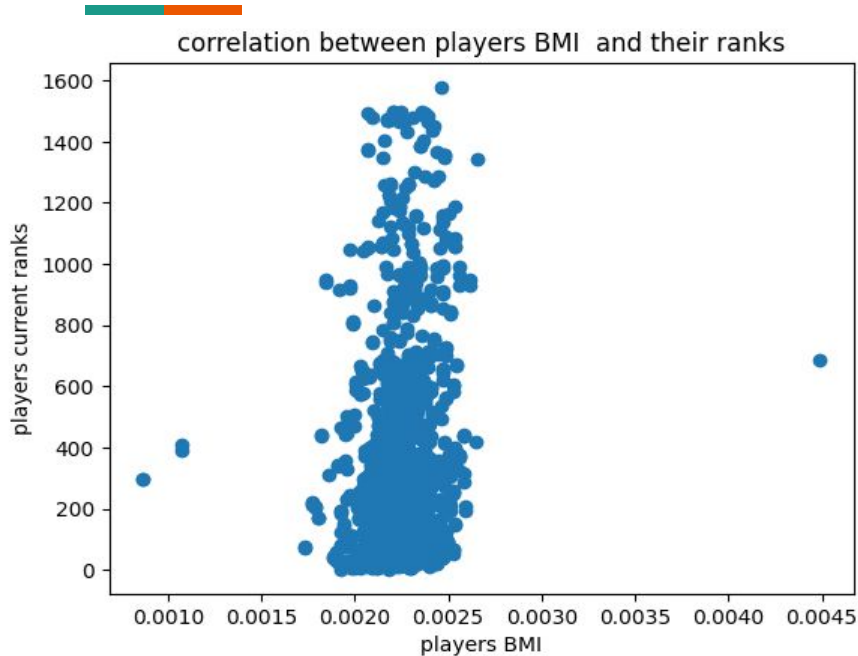
Analyse some parameters correlations

- The correlation between **height** and **ranking** was 0.10, which is **very weak**. The scatter plot verify this event.

- There is a **weak correlation** between players **experience** and their **rank** (0.217).

- There isn't much difference between the **ranks** for right-handed and left-handed players. **Right-handed** players have an average ranking of **359.29**, while **left-handed** players have an average ranking of **367.87**.

This slide covers the answers of question numbers 10, Ex. Q. 1, Ex. Q. 2

correlation between players experience and their ranks

Correlation between Player Height and Ranking

Analyse some parameters correlations

- There is no correlation between **Height** and weight ratio **(BMI)** and ranking value

correlation between players BMI and their ranks

correlation between players BMI and their ranks after remove outliers

This slide covers the answers of question numbers Ex. Q. 7,

Challenges:

- **Reading data**: **os.walk()** method
- **Cleaning data**: **dropduplicate()** method.
    - Also, there were many null and zero data in the tables related to the team information, which were deleted or replaced according to the conditions of use.
    - For example, to calculate the average height of the players
- **The strangeness of some results:** The result obtained from the **duration of the longest match**, **number of games in sets** seems strange and illogical, the reason for which needs further investigation.

Conclusion

- Through this data analysis project, we've uncovered a wealth of insights about the 2024 ATP
- From player profiles to match durations and tournament dynamics, the data has provided a fascinating glimpse into the world of professional tennis
- As we look ahead to future tournaments, this analysis can serve as a valuable resource for players, coaches, and fans alike, helping them better understand the factors that contribute to success on the ATP Tour.

Thank You