



Analysis of Tennis Matches

Group 9
July 2, 2024

Introduction

The dataset includes player information, statistics and results of matches, and various match-specific metrics in Parquet file format for professional tennis tournaments in May 2024.

Purpose of Analysis:

1. How many tennis players are included in the dataset?
2. What is the average height of the players?
3. Which player has the highest number of wins?
4. What is the longest match recorded in terms of duration?
5. How many sets are typically played in a tennis match?
6. Which country has produced the most successful tennis players?
7. What is the average number of aces per match?
8. Is there a difference in the number of double faults based on gender?
9. Which player has won the most tournaments in a single month?
10. Is there a correlation between a player's height and their ranking?
11. What is the average duration of matches?
12. What is the average number of games per set in men's matches compared to women's matches?
13. What is the distribution of left-handed versus right-handed players?
14. What is the most common type of surface used in tournaments?
15. How many distinct countries are represented in the dataset?
16. Which player has the highest winning percentage against top 10 ranked opponents?
17. What is the average number of breaks of serve per match?

Methodology

Step 1

Data Extraction:

Extracting the files related to each schema, concating and saving them under the related schema name to generate data frames based on them.

Step 2

Data Understanding and Cleansing:

- Removing duplicate data;
- Filling Nan data.

Step 3

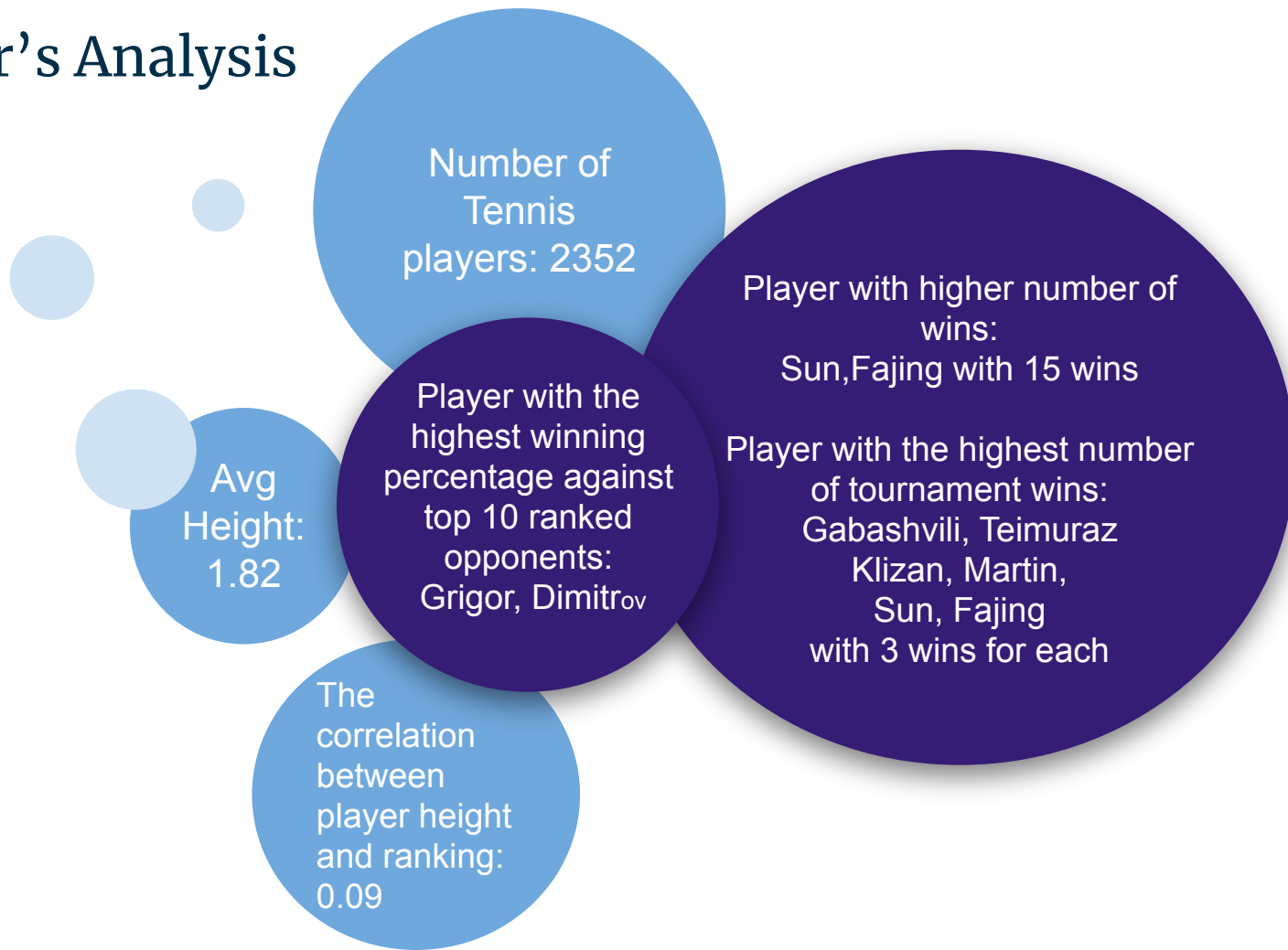
- Answering the questions based on the generated data frames and generating additional questions to analysis other aspects of data;
- Providing data visualization.



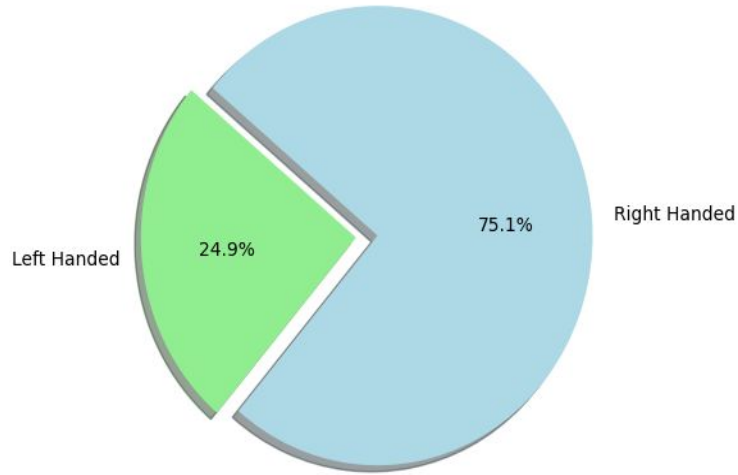
Data Analysis Findings

- Player Analysis
- Match Analysis
- Country Analysis

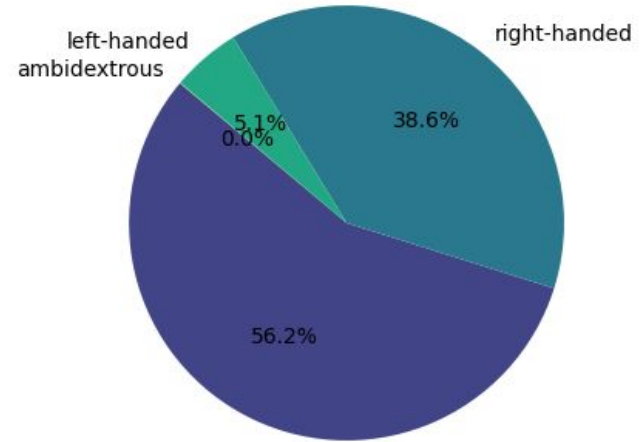
Player's Analysis



The distribution of left-handed versus right-handed players

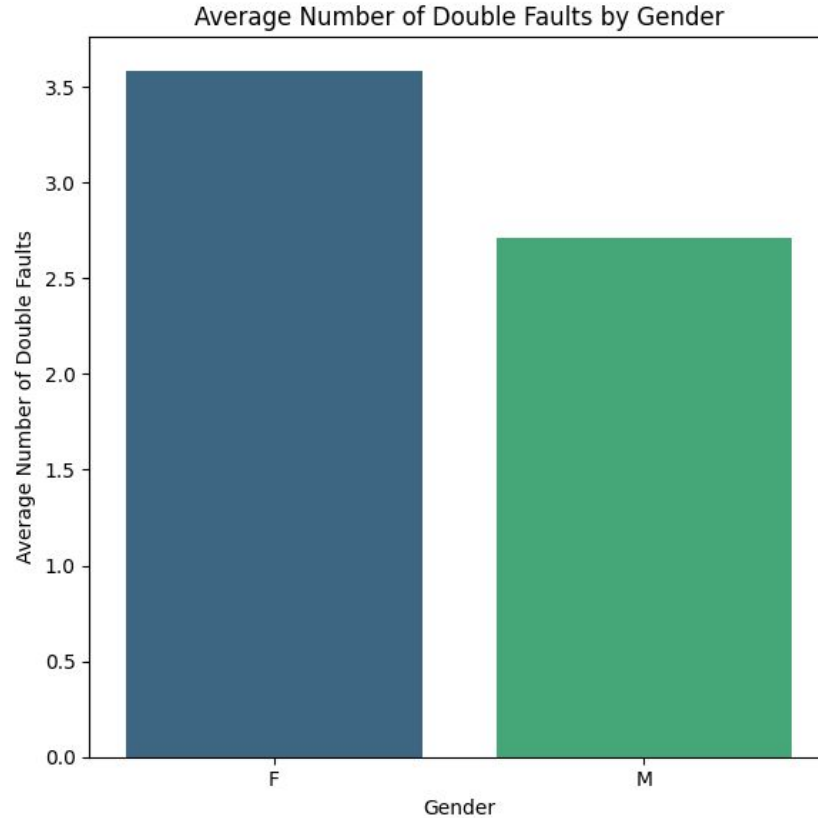


After Data Cleansing



Before Data Cleansing

The difference in the number of double faults based on gender



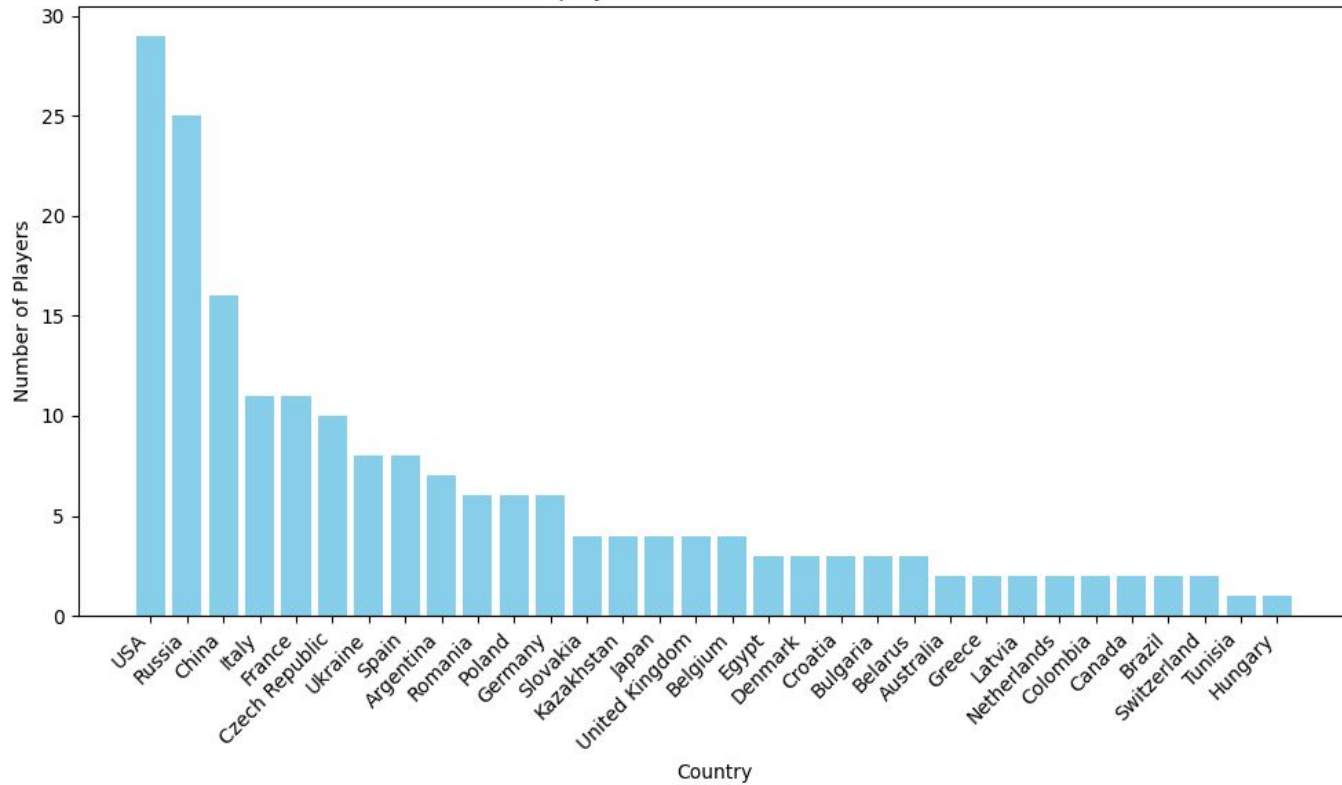
Match Analysis

- The match with the longest duration is ID 12346747.0, lasting 47.54 hours.
- The typical number of sets played in a match is 2.
- The average number of aces per match is 4.05.
- The average duration of matches is 127.03 minutes.
- The most common tournament surface is red clay.
- The average number of breaks of serve per match is 7.25.
- The average number of games per set is 9.22 in men's matches and 8.93 in women's matches.

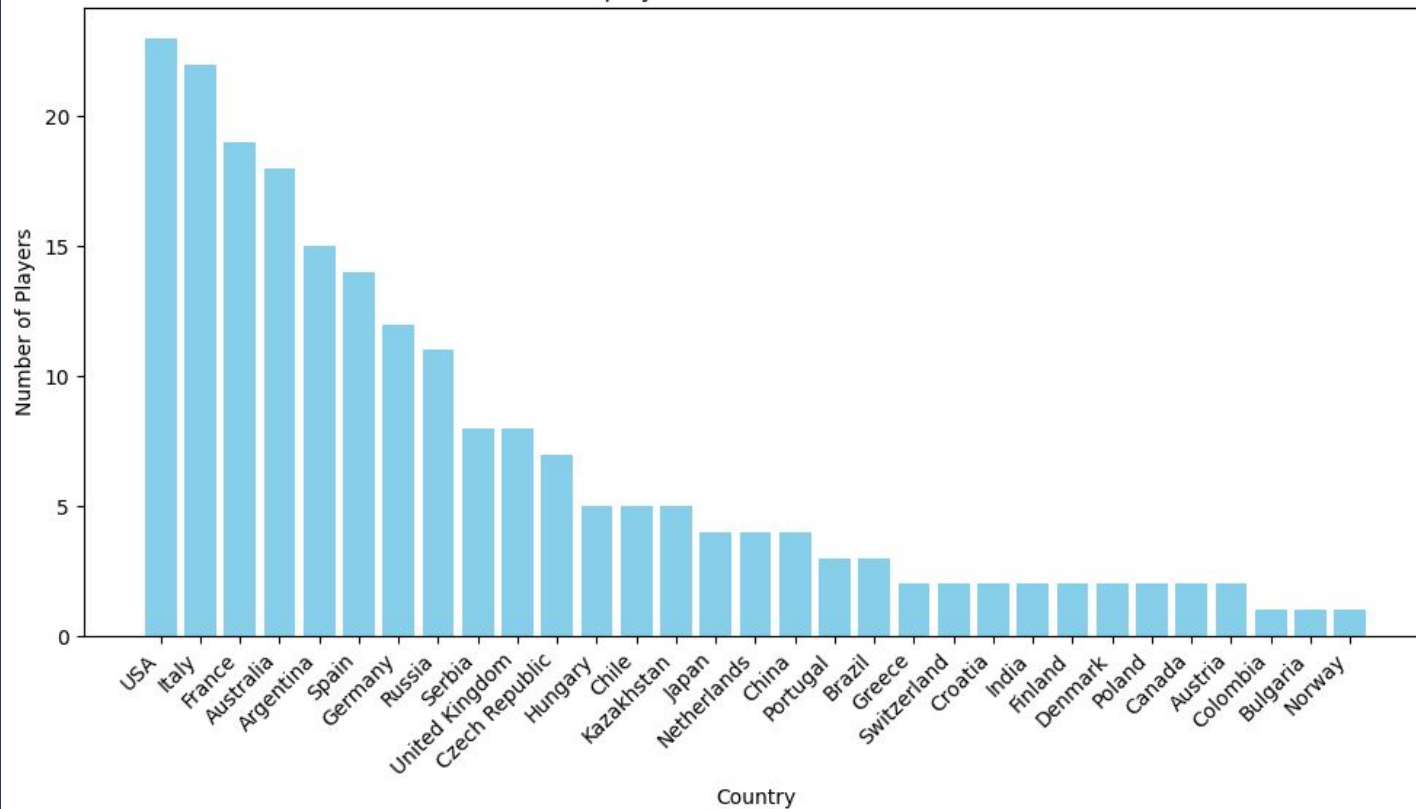
Country Analysis

- The dataset includes players from 99 unique countries.
- The country with the most successful tennis players is the USA.

Number of players with ranks under 100 - Female



Number of players with ranks under 100 - male



Thanks For Your Attention