**MAHRUKH NISAR**

**Big Data Analyst**

**Final Paper**

**Tennis Dataset:**

Dataset Link:

<https://www.kaggle.com/datasets/dissfya/atp-tennis-2000-2023daily-pull>

About Dataset:

This dataset contains 60k+ ATP matches (2000-2024)

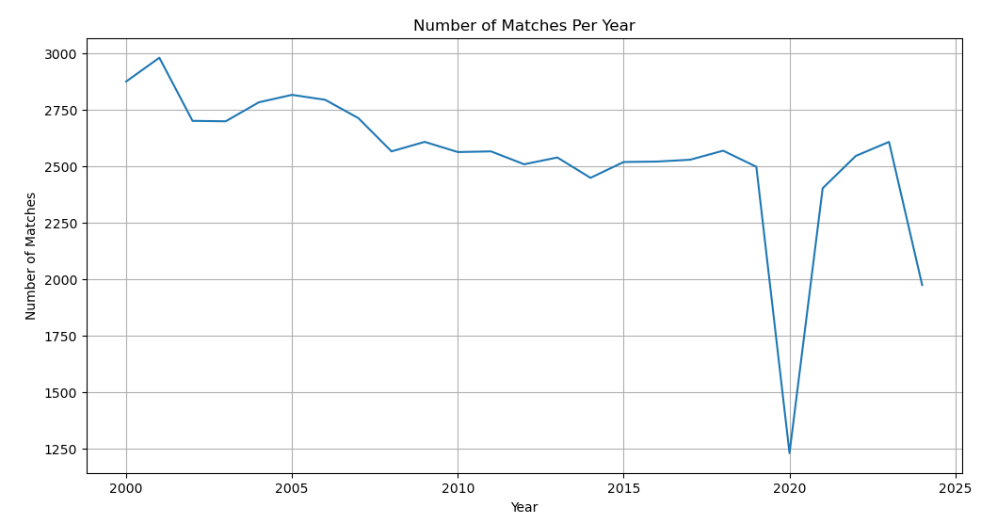
Each match includes the tournament, date, type of tournament, whether the match was played indoors or not, type of court surface, round of the match, maximum number of sets in the match, participants in the match, winner, ranks of the participants, odds to win and match score.

**Query no 01:** **Number of Matches Per Year**

This code calculates and visualizes the number of tennis matches per year. It first converts the 'Date' column to datetime format and extracts the year. It then counts the matches for each year and plots the data using a line chart to show trends over time.



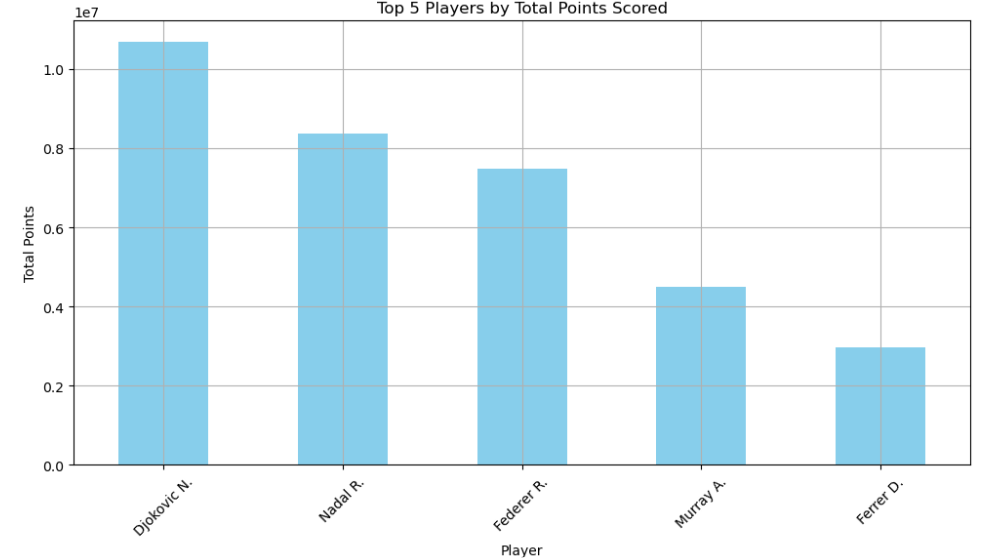


**Output:** 

**Query no 02: Total Points Scored by Each Player**

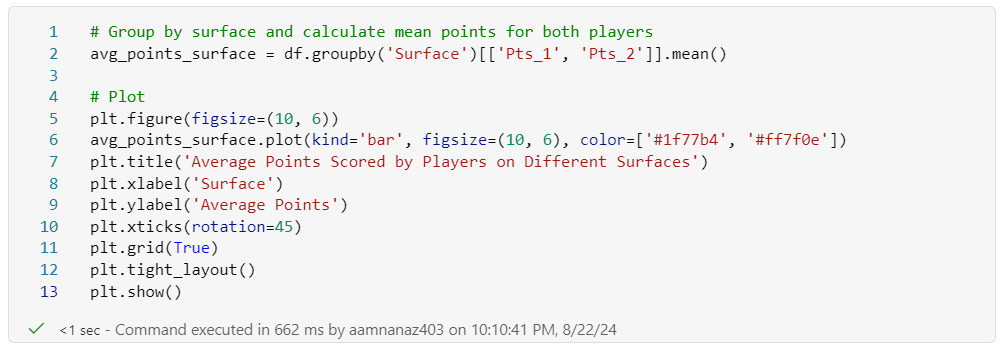
This code aggregates total points scored by each player from both Player\_1 and Player\_2 columns, combines and sums these points, then identifies the top 5 players with the highest total points.



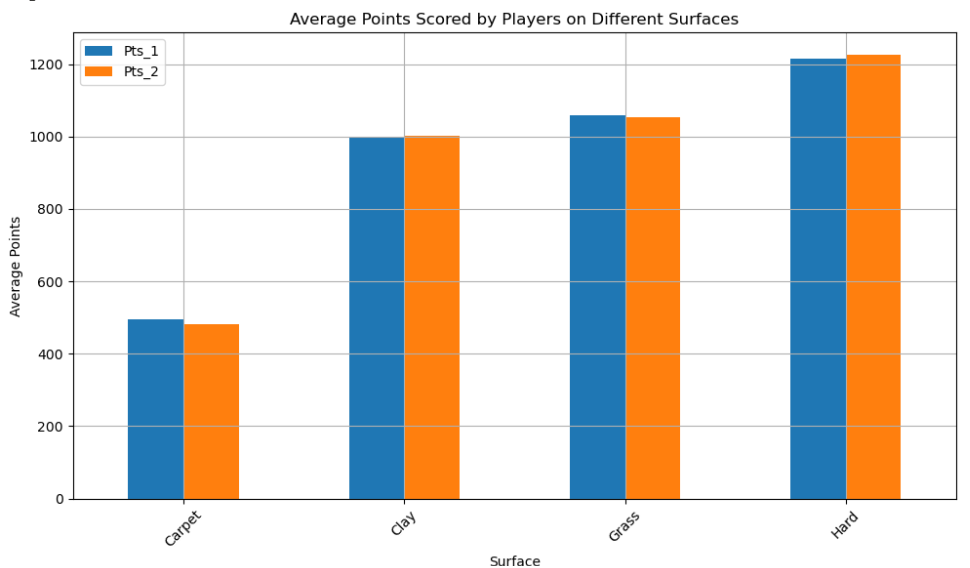
**Output:** 

**Query no 03: Average Points by Court Surface**

This visualization will show the average points scored by players on different court surfaces, broken down by Player 1 and Player 2.



**Output:**

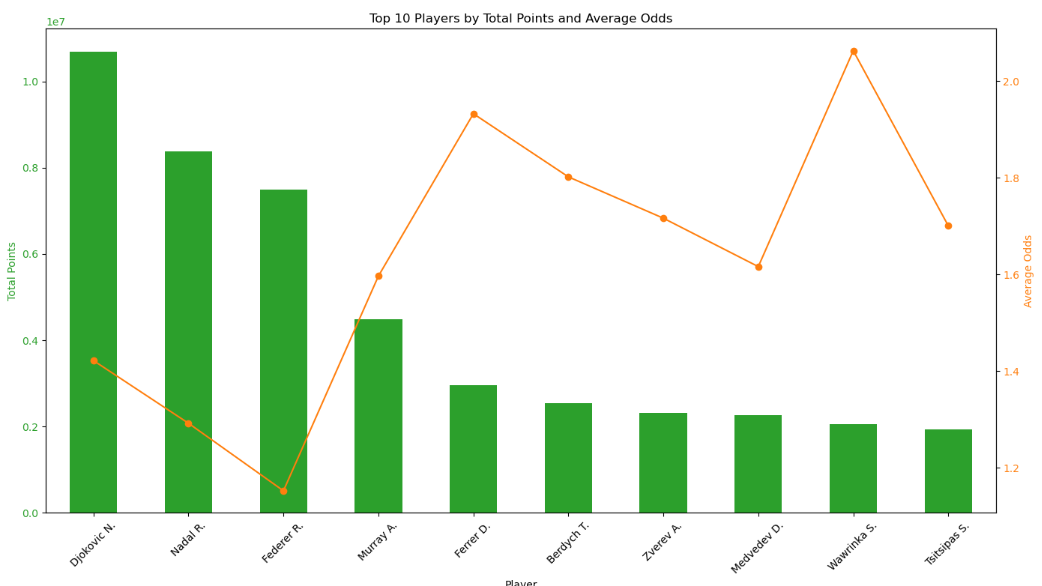


### Query no 04: Points Distribution by Surface and Court

Visualize the total points scored on different surfaces and courts.



**Output:**

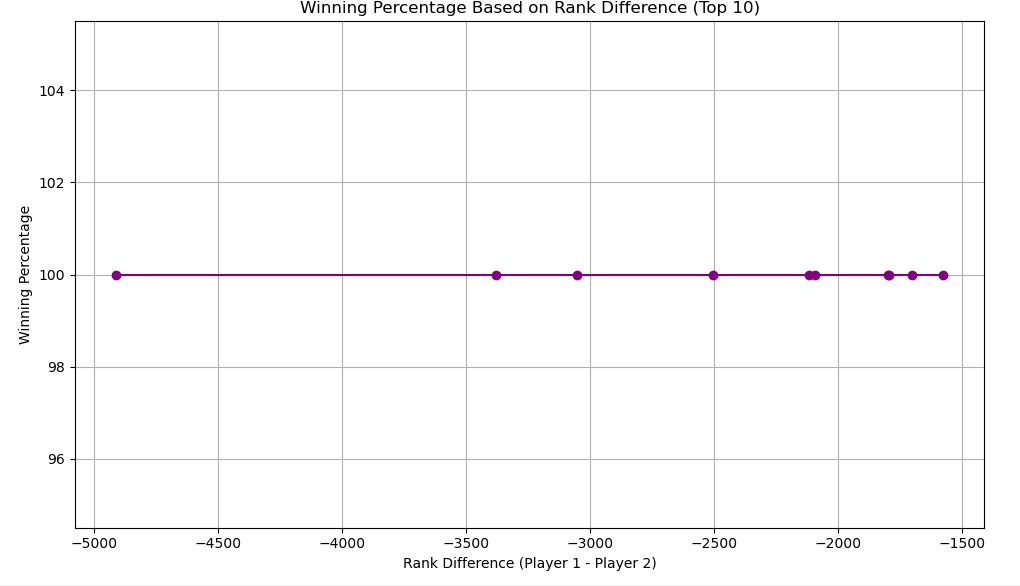


### ****Query no 05: Rank Difference and Winning Percentage****

### This visualization explores how rank difference between two players affects the percentage of times the higher-ranked player wins.



**Output:**

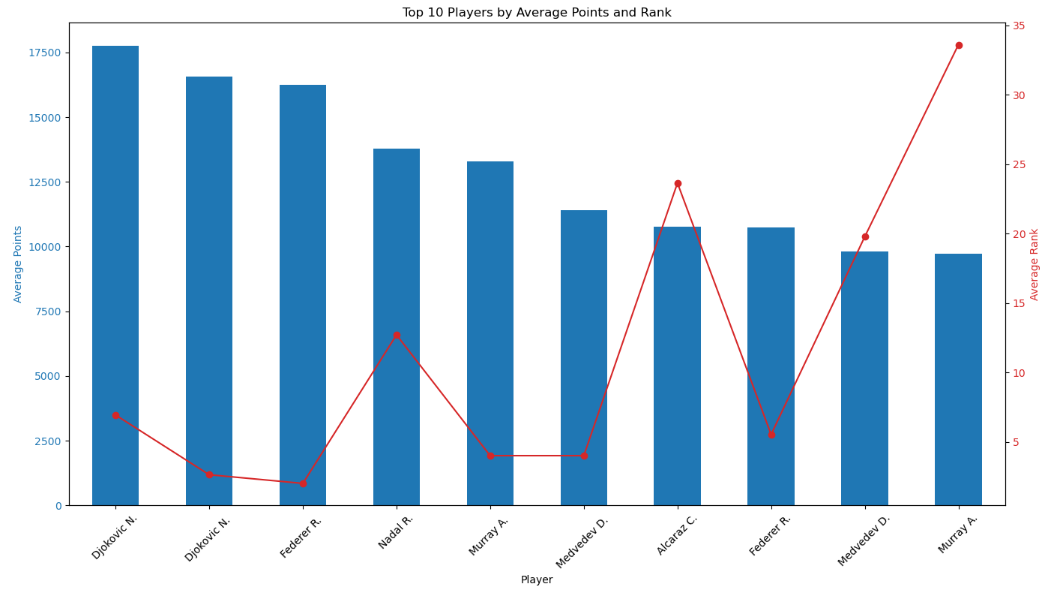


### Query no 06: Top 10 Players by Average Points Scored and Rank

### Visualize the top 10 players based on their average points scored, alongside their average rank.



**Output:**

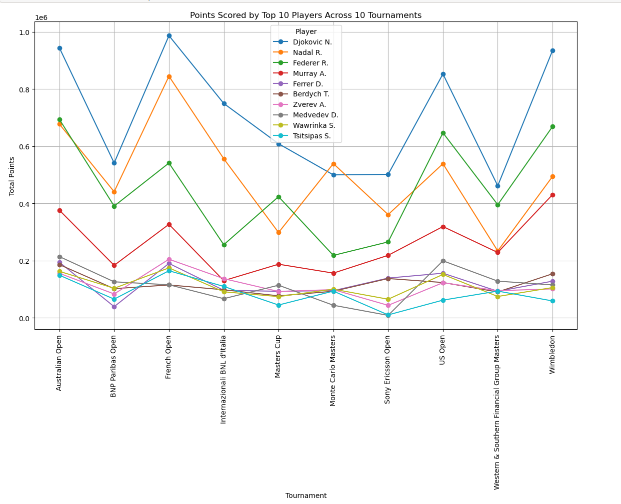


### Query no 07: Comparison of Points Scored by Top 10 Players Across 10 Tournaments

### Visualize the total points scored by the top 10 players across different tournaments.



**Output:**

****