

ATP TENNIS ANALYSIS

BY INDIA FLORES

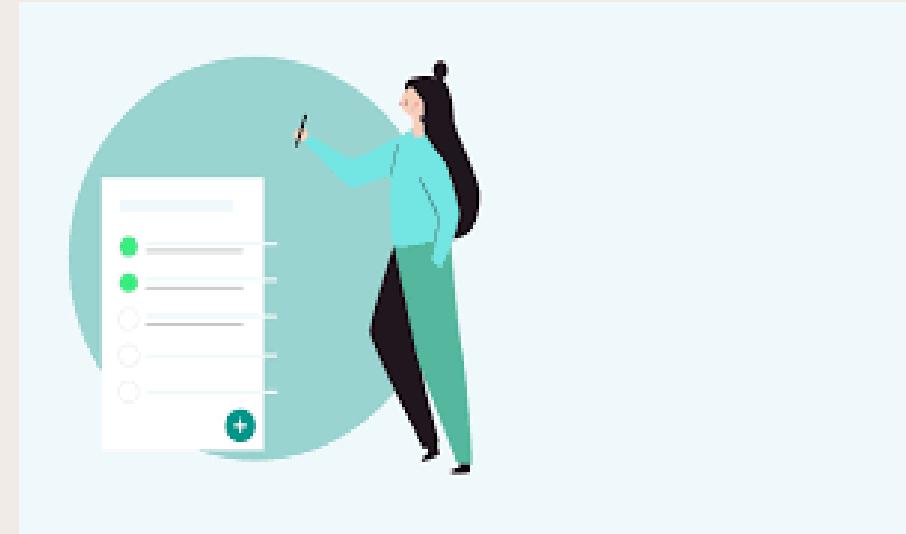
Demonstrates data cleaning, SQL analysis in BigQuery, and Tableau visualization



OBJECTIVES

Analyze trends in the data by answering:

1. How many ATP matches are played on each surface type?
2. Which players win the most matches on each surface?
3. Which surface produces the most physically demanding matches?
4. How many matches does each player play on each surface?



ABOUT THE DATA

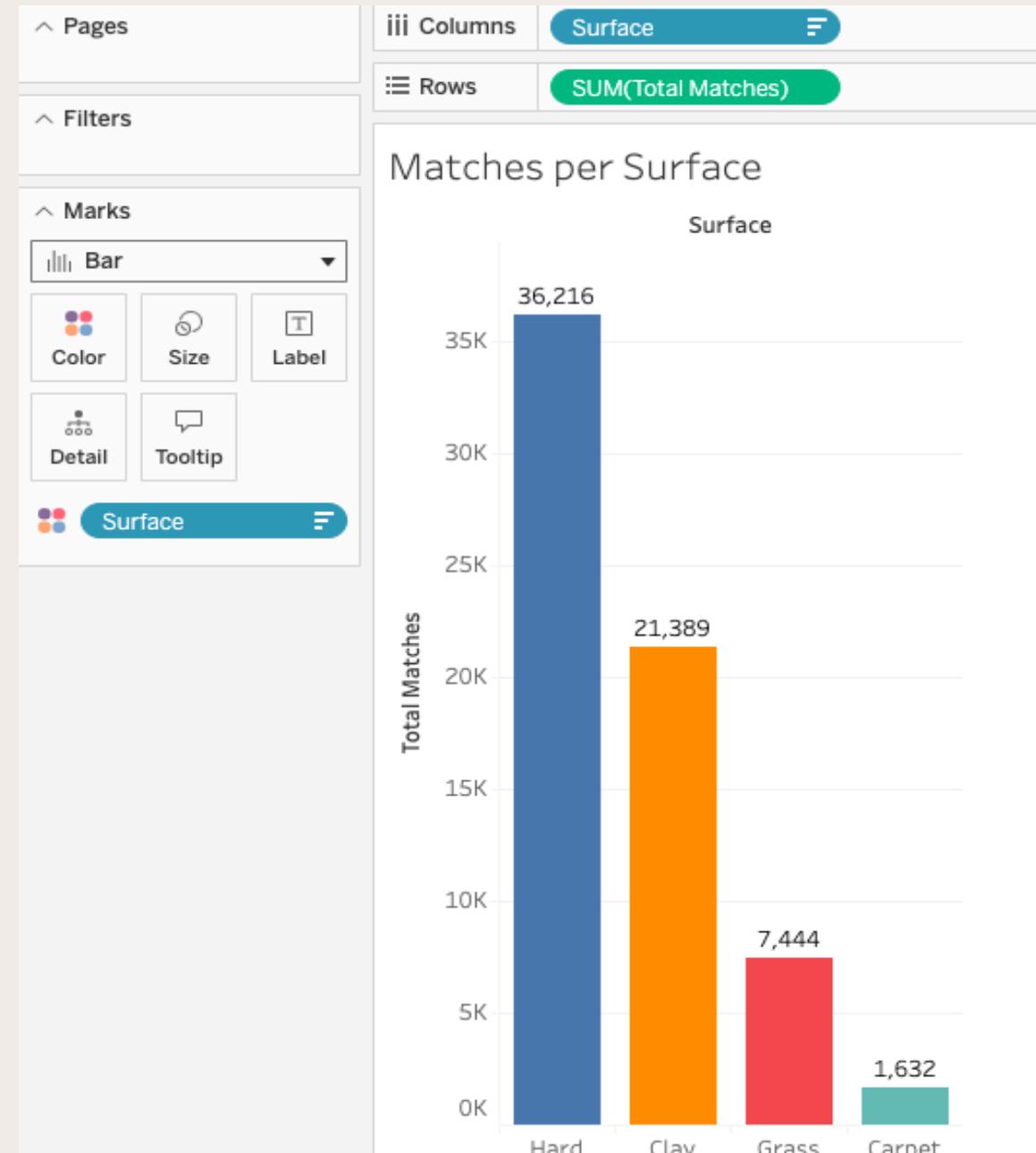
- Includes 66,682 rows with information on tournaments, surface type, players, winners, and scores
- Downloaded from Kaggle in a CSV file to Excel
- Cleaned the data by splitting score fields into separate columns, standardizing surface names, converting dates to a consistent format, and removing duplicate records
- Analysis focuses on main-draw ATP matches from 2000–2025
- Full dataset: [ATP Tennis 2000-2025](#)

Name Box	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Tournament	Date	Series	Court	Surface	Round	Best of	Player_1	Player_2	Winner	Rank_1	Rank_2	Pts_1	Pts_2	Odd_1	Odd_2	Score
Australian Ha	1/3/00	International	Outdoor	Hard	1st Round	3	Dosedel S.	Ljubicic I.	Dosedel S.	63	77	-1	-1	-1	-1	-1 6-46-2
Australian Ha	1/3/00	International	Outdoor	Hard	1st Round	3	Clement A.	Enqvist T.	Enqvist T.	56	5	-1	-1	-1	-1	-1 3-63-6
Australian Ha	1/3/00	International	Outdoor	Hard	1st Round	3	Escude N.	Baccanello P	Escude N.	40	655	-1	-1	-1	-1	-1 6-77-56-3
Australian Ha	1/3/00	International	Outdoor	Hard	1st Round	3	Knippschild J	Federer R.	Federer R.	87	65	-1	-1	-1	-1	-1 1-64-6
Australian Ha	1/3/00	International	Outdoor	Hard	1st Round	3	Fromberg R.	Woodbridge	Fromberg R.	81	198	-1	-1	-1	-1	-1 7-65-76-4
Australian Ha	1/3/00	International	Outdoor	Hard	1st Round	3	Arthur W.	Gambill J.M.	Gambill J.M.	105	58	-1	-1	-1	-1	-1 6-36-74-6
Australian Ha	1/3/00	International	Outdoor	Hard	1st Round	3	Grosjean S.	Ilie A.	Grosjean S.	26	51	-1	-1	-1	-1	-1 6-26-1
Australian Ha	1/3/00	International	Outdoor	Hard	1st Round	3	Balcells J.	Henman T.	Henman T.	218	11	-1	-1	-1	-1	-1 4-66-7
Australian Ha	1/3/00	International	Outdoor	Hard	1st Round	3	Hewitt L.	Woodforde M	Hewitt L.	24	129	-1	-1	-1	-1	-1 6-32-66-1
Australian Ha	1/3/00	International	Outdoor	Hard	1st Round	3	Tebbutt M.	Lisnard J.	Lisnard J.	351	134	-1	-1	-1	-1	-1 6-77-64-6
Australian Ha	1/3/00	International	Outdoor	Hard	1st Round	3	Martin A.	Spadea V.	Martin A.	55	21	-1	-1	-1	-1	-1 6-26-2
Australian Ha	1/3/00	International	Outdoor	Hard	1st Round	3	Draper S.	Norman M.	Norman M.	155	15	-1	-1	-1	-1	-1 4-64-6
Australian Ha	1/3/00	International	Outdoor	Hard	1st Round	3	Petrovic D.	Huet S.	Petrovic D.	208	122	-1	-1	-1	-1	-1 3-66-36-3
Australian Ha	1/3/00	International	Outdoor	Hard	1st Round	3	Pozzi G.	Sekulov J.	Sekulov J.	79	142	-1	-1	-1	-1	-1 4-66-22-6
Australian Ha	1/3/00	International	Outdoor	Hard	1st Round	3	Stoltenberg J	Hrbaty D.	Stoltenberg J	84	20	-1	-1	-1	-1	-1 6-36-1
Australian Ha	1/3/00	International	Outdoor	Hard	1st Round	3	Goldstein P.	Vinck C.	Vinck C.	88	227	-1	-1	-1	-1	-1 6-33-64-6
Australian Ha	1/3/00	International	Outdoor	Hard	2nd Round	3	Enqvist T.	Federer R.	Enqvist T.	5	65	-1	-1	-1	-1	-1 7-66-4
Australian Ha	1/3/00	International	Outdoor	Hard	2nd Round	3	Dosedel S.	Escude N.	Escude N.	63	40	-1	-1	-1	-1	-1 6-74-6
Australian Ha	1/3/00	International	Outdoor	Hard	2nd Round	3	Grosjean S.	Gambill J.M.	Grosjean S.	26	58	-1	-1	-1	-1	-1 3-67-66-2
Australian Ha	1/3/00	International	Outdoor	Hard	2nd Round	3	Lisnard J.	Henman T.	Henman T.	134	11	-1	-1	-1	-1	-1 6-34-62-6
Australian Ha	1/3/00	International	Outdoor	Hard	2nd Round	3	Hewitt L.	Petrovic D.	Hewitt L.	24	208	-1	-1	-1	-1	-1 6-06-2
Australian Ha	1/3/00	International	Outdoor	Hard	2nd Round	3	Sekulov J.	Martin A.	Martin A.	142	55	-1	-1	-1	-1	-1 6-76-7
Australian Ha	1/3/00	International	Outdoor	Hard	2nd Round	3	Norman M.	Fromberg R.	Norman M.	15	81	-1	-1	-1	-1	-1 7-56-4
Australian Ha	1/3/00	International	Outdoor	Hard	2nd Round	3	Vinck C.	Stoltenberg J	Stoltenberg J	227	84	-1	-1	-1	-1	-1 6-34-64-6
Australian Ha	1/3/00	International	Outdoor	Hard	Quarterfinals	3	Enqvist T.	Grosjean S.	Enqvist T.	5	26	-1	-1	-1	-1	-1 7-67-5
Australian Ha	1/3/00	International	Outdoor	Hard	Quarterfinals	3	Henman T.	Escude N.	Escude N.	11	40	-1	-1	-1	-1	-1 3-61-6
Australian Ha	1/3/00	International	Outdoor	Hard	Quarterfinals	3	Hewitt L.	Stoltenberg J	Hewitt L.	24	84	-1	-1	-1	-1	-1 6-26-2
Australian Ha	1/3/00	International	Outdoor	Hard	Quarterfinals	3	Martin A.	Norman M.	Norman M.	55	15	-1	-1	-1	-1	-1 4-63-6
Australian Ha	1/3/00	International	Outdoor	Hard	Semifinals	3	Enqvist T.	Norman M.	Enqvist T.	5	15	-1	-1	-1	-1	-1 3-66-26-2
Australian Ha	1/3/00	International	Outdoor	Hard	Semifinals	3	Escude N.	Hewitt L.	Hewitt L.	40	24	-1	-1	-1	-1	-1 6-31-64-6
Australian Ha	1/3/00	International	Outdoor	Hard	The Final	3	Hewitt L.	Enqvist T.	Hewitt L.	24	5	-1	-1	-1	-1	-1 3-66-36-2
Gold Flake O	1/3/00	International	Outdoor	Hard	1st Round	3	Kumar S.	Agenor R.	Agenor R.	1198	94	-1	-1	-1	-1	-1 4-64-6
Gold Flake O	1/3/00	International	Outdoor	Hard	1st Round	3	Behrend T.	Popp A.	Behrend T.	116	124	-1	-1	-1	-1	-1 6-44-66-1
Gold Flake O	1/3/00	International	Outdoor	Hard	1st Round	3	Rochus C.	Damm M.	Damm M.	120	104	-1	-1	-1	-1	-1 3-60-6
Gold Flake O	1/3/00	International	Outdoor	Hard	1st Round	3	Delgado J.	Black B.	Delgado J.	162	67	-1	-1	-1	-1	-1 6-27-6
Gold Flake O	1/3/00	International	Outdoor	Hard	1st Round	3	Spottl M.	Golmard J.	Golmard J.	200	38	-1	-1	-1	-1	-1 3-64-6
Gold Flake O	1/3/00	International	Outdoor	Hard	1st Round	3	Hantschik M.	Tieleman L.	Hantschik M.	128	89	-1	-1	-1	-1	-1 7-66-3
Gold Flake O	1/3/00	International	Outdoor	Hard	1st Round	3	Koslak J.	Jonsson F.	Jonsson F.	103	127	-1	-1	-1	-1	-1 3-63-6

TOTAL MATCHES PER SURFACE

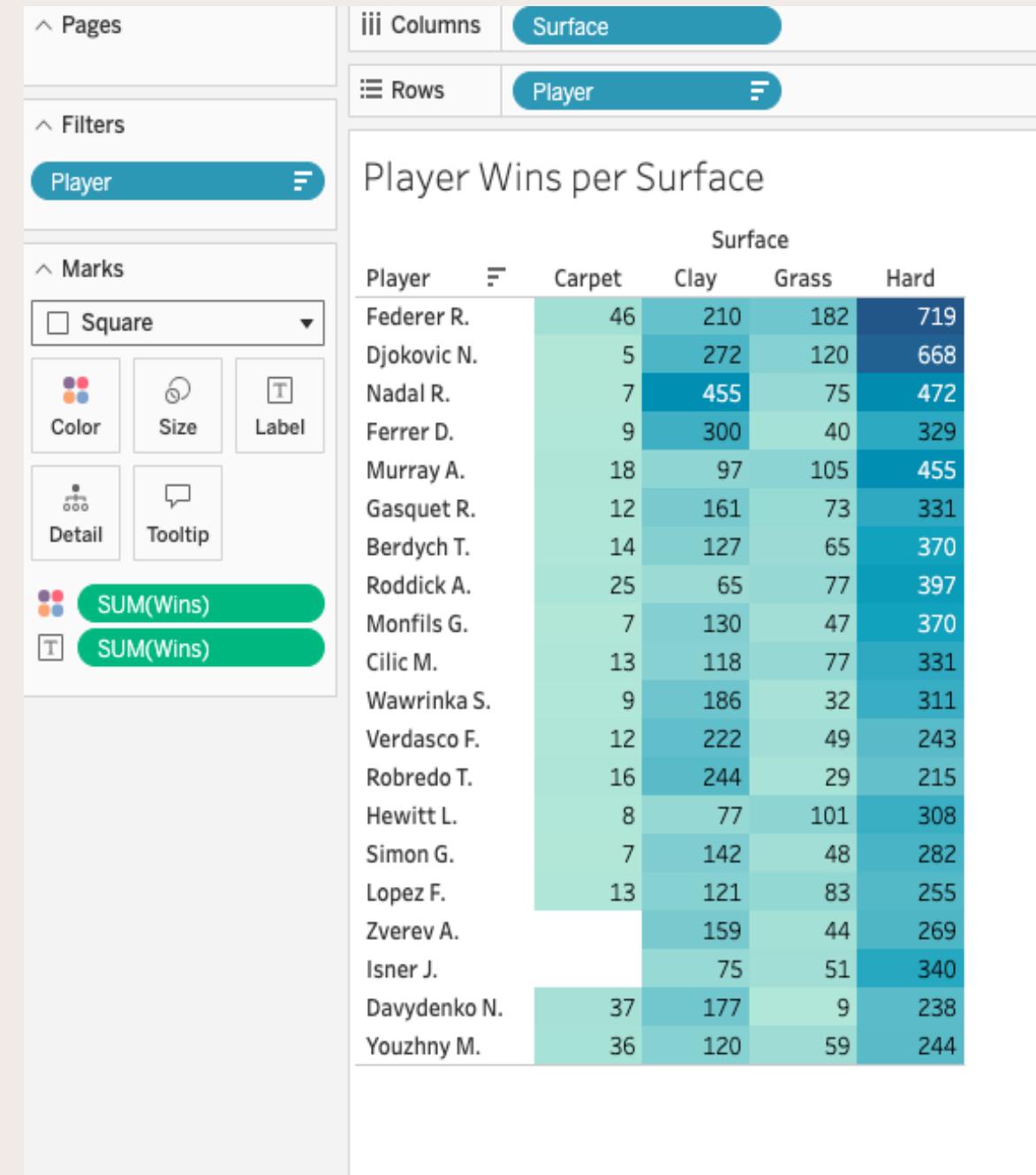
```
1 SELECT Surface, COUNT(*) AS total_matches
2 FROM tennis-analysis-481321.tennis_data.tennis_cleaned
3 GROUP BY Surface
4 ORDER BY total_matches DESC
```

- Used COUNT(*) grouped by surface to calculate the total number of matches played on each surface
- Hard courts had the most matches played, carpet courts had the least
- Hard courts dominate ATP scheduling, reflecting their prevalence in professional tournaments



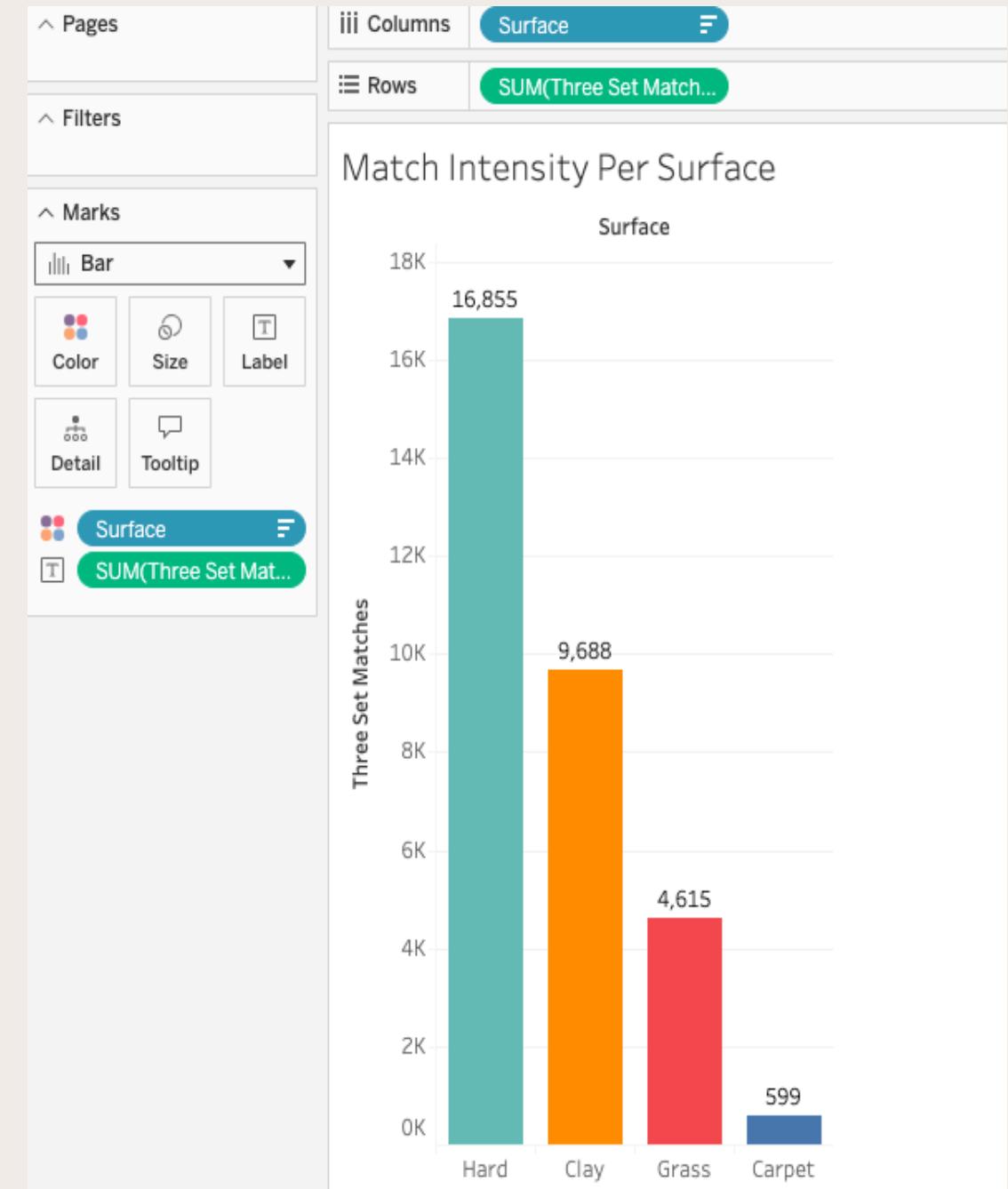
PLAYER SUCCESS BY SURFACE

- Highlighted top 20 performers and their most successful surface type by using color and labels in Tableau
- Federer R. had the most success on hard surface courts with 719 wins
- Rafael Nadal demonstrated exceptional dominance on clay courts, leading all players with 455 wins



MATCH INTENSITY PER SURFACE

- Used SQL to identify matches that required a third set as a signal for match intensity
- 16,855 out of 36,216 hard-court matches required a third set
- Only 599 games played on carpet had a 3rd set



PLAYER WORKLOAD PER SURFACE

- Created a CTE that unions player_1 and player_2 to account for total match participation per player
- Referenced the CTE to find how many matches each player played on each surface type
- This approach ensures each match contributes equally to player workload regardless of match outcome

```
WITH matches AS (
    SELECT Surface, Player_1 AS player
    FROM `tennis-analysis-481321.tennis_data.tennis_cleaned`
    UNION ALL
    SELECT Surface, Player_2 AS player
    FROM `tennis-analysis-481321.tennis_data.tennis_cleaned`
)
SELECT
    Surface,
    player,
    COUNT(*) AS matches_played
FROM matches
GROUP BY Surface, player
ORDER BY Surface, matches_played DESC;
```

INSIGHTS

1. Surface Distribution

Hard courts account for the majority of ATP matches, followed by clay and grass, reflecting both tournament scheduling and surface availability on tour.

2. Player Specialization

Player performance varies significantly by surface, with certain players demonstrating clear surface specialization rather than uniform success across all surfaces.

3. Match Intensity

Clay-court matches show a higher likelihood of extending to a third set, suggesting longer rallies and increased endurance demands compared to faster surfaces.

4. Player Workload

Players accumulate substantially more matches on hard courts, while select players show concentrated participation on specific surfaces, reinforcing strategic surface preferences.



LIMITATIONS AND NEXT STEPS

- Carpet matches represent a small sample size and may limit generalizability
- Match intensity measured only by third-set frequency
- Future analysis could include win percentage by surface or average games per match

