گزارش پروژه شماره ۱ دیتا آنالیز

بوتکمپ دیتاساینس دانشکار

گروه 8

اعضای گروه:

Arman Aghania

Reza Abedi

Fateme Nobakht

Behnam Mohseni

Aref Shojaei

تاریخ ارائه گزارش: ۱۴۰۲/۱۱/۲۵

This dataset offers a comprehensive collection of data points spanning multiple facets of professional tennis matches. It is meticulously structured into 15 relational dataframes derived from parquet files across 6 distinct folders. These dataframes collectively provide a multidimensional view of tennis matches, covering player profiles, match scores, event details, betting odds, point-by-point actions, tournament rounds, seasonal information, statistical analyses, time durations, tournament specifics, venue details, and audience engagement metrics.

the end reading and analyzing each one by the team, respectively, to review and analyze the questions and We went to the requirements of the project.

تحلیل و نتیجه‌گیری‌ها:

Player Profiles (HomeTeamInfo and AwayTeamInfo):

These tables detail profiles for players in home and away teams, encompassing personal demographics (e.g., birthplace, residence), professional attributes (e.g., plays, turned pro), and performance metrics (e.g., current rank, total prize).

Match Scores (HomeTeamScore and AwayTeamScore):

Document scoring dynamics for both teams, capturing scores by period and detailing tie-break outcomes, thus providing a snapshot of competitive flow across match segments.

Event Details (Event):

This table is the nucleus of the dataset, chronicling essential match events, participant details, match outcomes, and timings, enriched with unique identifiers and datetime stamps for comprehensive event tracking.

Betting Odds (Odds):

Offers a glimpse into the betting landscape prior to matches, with detailed odds and predictions, serving as a valuable resource for understanding market expectations versus actual outcomes.

Point-by-Point Analysis (PointbyPoint):

Provides a granular breakdown of scoring for each point, illuminating strategies, scoring patterns, and serving turns, essential for in-depth performance analysis.

Power Metrics (Power):

Focuses on the intensity and strategic breaks within games and sets, offering insights into the pivotal moments that could shift match momentum.

Round Progression (Round):

Details the stages of tournament progression, from preliminary rounds to the climactic finals, categorized by unique round identifiers and types, essential for understanding the competitive hierarchy.

Seasonal Context (Season) :

Aligns matches within their respective seasons, providing a temporal framework that is crucial for trend analysis and seasonal performance assessment.

Statistical Deep Dive (Statistics:

A comprehensive compilation of match statistics, from aces to double faults, segmented by home and away participants, offering a quantitative foundation for performance evaluation.

Timing Insights (Time):

Tracks the duration of each set, adding a layer of analysis that can correlate time factors with player performance and endurance.

Tournament Insights (Tournament):

Expands on tournament specifics, including location, surface type, and categories, providing a rich backdrop against which matches unfold, influencing player performance and strategy.

Venue Details (Venue):

Specifies match venues, offering insights into geographical and infrastructural factors that may influence match dynamics and player adaptability.

Public Engagement (Votes):

Reflects public sentiment through vote counts for home and away teams, serving as an indicator of audience preferences and perceptions.

شرح وظایف:

Arman

- Data loading
- Data visualization
- Data cleaning
- Analyze and answer questions
- Analyze the main questions using SQL queries
- Creating and cleaning the database and loading data on the database

Reza

- Data visualization
- Data cleaning
- Analyze and answer questions
- Report writing

Fateme

- Git
- Data visualization
- Data cleaning
- Analyze and answer questions
- Report writing