**Full Stack BI Developer Screening Task**

**Name: Pranitha Nekkanti**

**Problem Statement:**  
In modern business intelligence (BI) workflows, the effective integration and visualization of sales data is key in making well-informed data-driven decisions. This project solves the issue of extracting, processing, and visualizing sales data from a PostgreSQL database in an interactive Streamlit dashboard. The goal is to gain real-time insights into sales trends and category performance.

**Technologies Used**

* **Backend**
  + **PostgreSQL**: Database for storing sales data.
  + **SQLAlchemy**: ORM for database interactions.
  + **pandas**: Data extraction and transformation.
  + **Python**: Core programming language for data processing.
* **Frontend**
  + **Streamlit**: Interactive dashboard for visualization.
  + **Seaborn & Matplotlib**: Data visualization.
* **Version Control & CI/CD**
  + **GitHub**: Version control.
  + **Virtual Environments**: venv for dependency isolation.

**Challenges:**

1. Module Import Errors

Issue: ModuleNotFoundError: No module named 'visualizations'

Solution: Ensured that each directory (backend, frontend, visualizations) contains an \_\_init\_\_.py file, making it a Python package.

2. Database Connection Warning

Issue: pandas only supports SQLAlchemy connectable...

Solution: Changed raw psycopg2 connection to an SQLAlchemy engine for compatibility.

3. Non-GUI Backend for Matplotlib

Issue: Matplotlib is currently using agg, which is a non-GUI backend, so cannot show the figure.

Solution: Used st.pyplot(fig) instead of plt.show() to correctly render charts in Streamlit.

4. Dataset Size & Performance

Issue: Fetching large datasets from PostgreSQL slowed down performance.

Solution: Implemented pagination in SQL queries and caching with st.cache\_data().

**Additional Tableau Projects:**

**(link:** [**https://app.mural.co/t/dvfinalproject6417/m/dvfinalproject6417/1680728836627/1f81b1a12f7175f731446d113e920bfd01e565d1**](https://app.mural.co/t/dvfinalproject6417/m/dvfinalproject6417/1680728836627/1f81b1a12f7175f731446d113e920bfd01e565d1)**)**

**Dashboard for Indian Premier League:**

The following dashboard summarizes IPL match data from 2008 to 2020. It focuses on various factors related to players, teams, and matches. This dashboard was created to assist IPL management, analysts, media, betting companies, fantasy sports users, fans, and coaches in making better-informed decisions based on insights and better understanding the game.

**Proof of Accomplishment:**

**Dashboard:** The IPL dashboard presents an interactive way to understand match statistics, achievements of players, and team trends over the years from 2008 to 2020. It assists data-informed analytics in performance analysis across seasons, players, and teams.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Plot 1:** The graph depicts the number of IPL matches hosted in various cities, with the size of elements indicating the volume of matches. This visual approach enables quick and easy comparisons of match counts across different locations.

A screenshot of a computer

Description automatically generated

**Plot 2:** The following chart represents the number of IPL matches held in each season, from the year 2008 to 2020, reflecting the development of the schedule of this tournament. As can be seen from the chart, the match counts of different seasons can be easily compared with one another, hence showing whether more or fewer games have been played.

A screenshot of a computer

Description automatically generated

**Plot 3:** This graph represents the leading run-scorers in the history of the IPL, featuring aggregate runs across different seasons and pointing out peak performances. The seasonal breakdown is given to show consistency, big innings impact, and records of most runs scored in individual seasons.

A screenshot of a computer

Description automatically generated

**Plot 4:** It visualizes the distribution of runs scored by different methods (1s, 2s, 3s, 4s, and 6s), highlighting how players accumulated their total scores. Color and size are used as pre-attentive attributes to distinguish each scoring method and emphasize its frequency.

A screenshot of a computer

Description automatically generated

**Plot 5:** The graph illustrates the various types of dismissals achieved by bowlers, employing length as a pre-attentive attribute to enhance visual differentiation, and understanding of the data.

A screenshot of a computer

Description automatically generated

**Plot 6**: It shows match outcomes based on whether the toss was won or lost, providing insight into the influence of toss results on game success. Color intensity differentiates outcomes, making it easy to assess the impact visually.

A screenshot of a computer

Description automatically generated

**Plot 7:** The graph illustrates the toss decisions made at various venues, utilizing length and color as pre-attentive attributes to enhance visual understanding.

A screenshot of a computer

Description automatically generated

**Summary:**

In the IPL Dashboard project, an overview and analysis of the data from Indian Premier League matches between 2008 and 2020 were presented, supported with a dataset comprising over 800 match records and close to 200,000 ball-by-ball entries. Preprocessing of the data through grouping attributes such as years and teams helped in reaping insightful information concerning the tournament's history. It is used because it can work with big datasets and build an interactive visualization with different plots for analysis, like top scorers, wicket-takers, and venue outcomes. The dashboard will assist different kinds of user management and fans who can derive insights from this dashboard regarding player performance and team performance. Finally, it would introduce an important role of data visualization in sports analytics: to gain better insight into historical trends and to make better decisions based on data. The project thus proved how the use of data will lead to better strategic planning and more audience sports engagement in the competitive landscape. The dashboard allowed complex data to be easily accessed and interpretable; users could create changes in patterns and trends that might have otherwise gone unnoticed and further enrich their engagement with the IPL. Moreover, interactive features embedded within allowed readers to personalize their analyses, thus developing a deeper regard for the game's complexity.