Data Analysis Final Project

 Read each **univariate, bivariate, and multivariate** question inside.

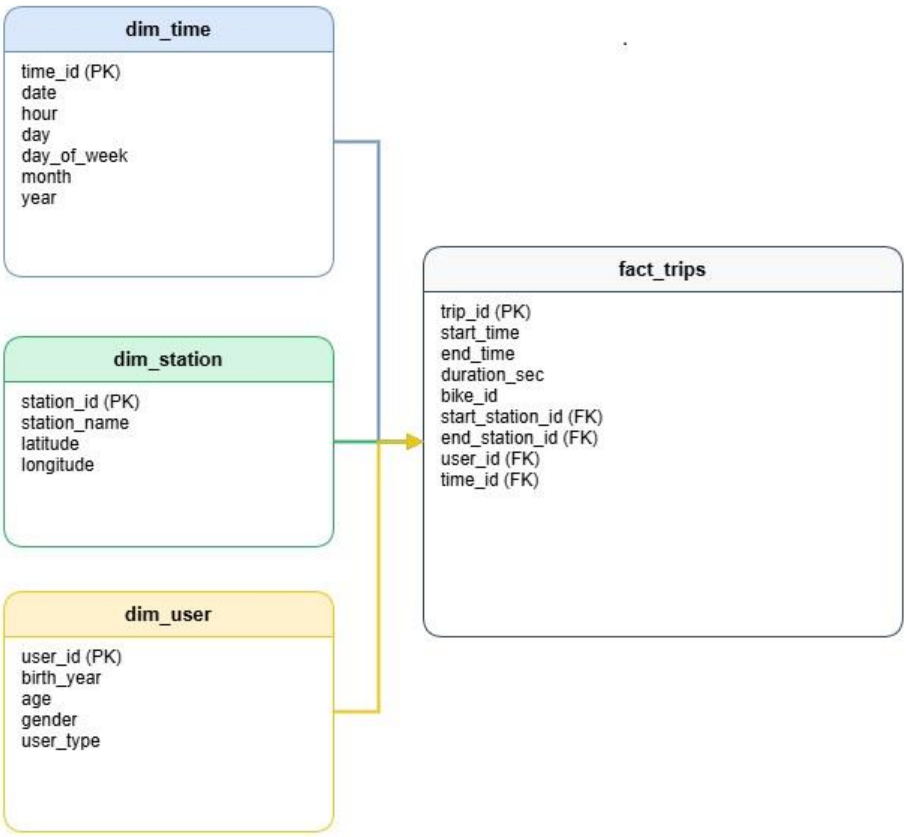
 Write **Python code cells** with the correct analysis & plots.

 Add **professional explanations below each visualization** (like an analyst's interpretation).

 Enhance the notebook so it looks like a **polished EDA report** your students can learn from.

After that

You are given the following **Entity-Relationship Diagram (ERD)** for a **bike-sharing system (Ford GoBike)**. The ERD contains one **fact table** and three-dimensional **tables**:



Part 2 – Preprocessing (Python/Pandas)

1. **Clean data** → fix missing values, remove outliers, standardize categories.
2. **Engineer features** → trip duration in minutes, weekend flag, age groups.
3. **Encode/scale** → convert categorical to numeric, scale numbers if needed.
4. **EDA** → plots: trips by weekday, duration distribution, subscriber vs customer.
5. **Answer every question in a notebook and you can add more**

Part 3 – Interactive Dashboard (Plotly Dash / Streamlit)

1. **Load cleaned data** (after preprocessing).
2. **Add filters/slicers** (e.g., by date, user type, gender, age group).
3. **Visualize key insights**:
    - Trips by weekday. ○ Trip duration distribution.
    - Subscriber vs Customer usage.
    - Top start/end stations.
4. **Make interactive** dropdowns, sliders, checkboxes.
5. **Layout** → divide dashboard into 4 sections:
    - **Overview KPIs** (total trips, avg duration, active users). ○ **Time Analysis** (day/week/month trends). ○ **User Analysis** (age, gender, type).
    - **Station/Trip Analysis** (popular stations, trip flows).

This image looks like international companies did

Try to do: