Data Scientist Take-Home Test

Time Limit

Expected completion time: 3-5 hours
Deadline: Within 72 hours of receiving this test

Submission Format

- One Jupyter notebook (or Google Colab link)
- A brief PDF summary (max 2 pages) explaining your approach, findings, and recommendations

Problem Statement

You are working as a data scientist at a ride-hailing platform. You've been provided with a sample of trip data. Your goal is to:

1. Understand user behavior
2. Detect potential operational inefficiencies
3. Recommend data-driven solutions

Dataset

rides.csv: Contains ride-level data with columns like:
- ride_id, user_id, driver_id, start_time, end_time
- pickup_location, dropoff_location
- fare, distance_km, rating, cancelled

1. Exploratory Data Analysis (EDA)

- Describe key characteristics of the dataset.
- Identify any issues (missing values, outliers, duplicates).
- Visualize key insights (top locations, ride volumes, etc.).

2. User Behavior Analysis

- Segment users by ride frequency and spending.
- Analyze trends in cancellations and low ratings.
- Identify churn indicators (users who stop using the service).

3. Predictive Modeling

- Build a binary classification model to predict ride cancellation.
- Include feature engineering.
- Evaluate using precision, recall, and F1-score.

4. Operational Insights

- Identify underperforming or overworked drivers.
- Spot peak demand times and locations.
- Recommend data-driven actions to improve efficiency or user retention.

Data Scientist Take-Home Test

5. Bonus (Optional)

- Deploy a simple dashboard using Streamlit or similar.
- Conduct a basic time series forecast of demand.

Evaluation Criteria

- Data cleaning & EDA: 20%
- Analytical reasoning: 20%
- Feature engineering: 15%
- Model quality & metrics: 20%
- Interpretation & insights: 15%
- Clarity of communication: 10%