Project Title:
Smart Ride Fare Analysis & Prediction System

---

1. Introduction

In recent years, smart transportation services have witnessed significant growth, leveraging dynamic pricing to provide a balanced and fair experience for both drivers and passengers. This pricing is influenced by multiple factors such as traffic congestion, weather conditions, vehicle status, trip distance, and timing.

Despite the importance of these factors, many companies struggle to *accurately understand the real impact of each factor* and determine whether current fares truly reflect the trip conditions.

This project aims to build an integrated system for data analysis and insight extraction, along with developing a predictive model that helps companies optimize pricing mechanisms and achieve higher operational efficiency.

---

2. Problem Statement

Smart transportation platforms face several challenges related to fare pricing, including:

- Identifying which factors have the greatest impact on trip fares.
- Determining whether a predictive model can estimate fares before the trip starts.
- Assessing whether current pricing is fair and aligned with trip conditions.
- Representing data spatially and temporally to uncover new insights.
- Understanding how fares vary across drivers, passengers, or congested areas.
- Leveraging insights to improve driver performance and service quality.

These questions reveal a gap that can be addressed through intelligent data-driven analysis.

---

3. Research Questions

To strengthen the academic rigor of the project, we aim to answer the following:

1. **RQ1:** Which factors directly affect trip fare values?
2. **RQ2:** How do weather and traffic congestion influence fare increases or decreases?
3. **RQ3:** Are there significant differences in pricing across geographic regions?
4. **RQ4:** Can a predictive model accurately forecast trip fares?
5. **RQ5:** At what times do fares peak, and why?
6. **RQ6:** How does vehicle condition affect fares and passenger experience?

Addressing these questions will enhance the project's appeal to evaluators.

---

4. Project Objectives

Main Objectives:

1. Conduct a comprehensive analysis of fare-determining factors.
2. Build an AI model capable of accurately predicting trip fares.
3. Develop an interactive dashboard for trip behavior analysis.

Sub-Objectives:

- Identify the most congested and high-fare areas.
- Analyze driver performance and generate performance reports.
- Study the impact of weather and traffic on trip demand.
- Design a performance evaluation system to improve service management.

---

5. Dataset Description

The dataset contains 26 features, including:

- Pickup and drop-off coordinates.
- Distance between pickup and drop-off points.
- Weather conditions (Clear, Rainy, Foggy, etc.).
- Traffic conditions (Light, Medium, Heavy).
- Vehicle condition (Excellent, Good, Poor).
- Temporal data (Hour, Day, Month, Year, Weekday).
- Distance from airports (JFK, LGA, EWR).
- Fare amount.

Strengths of the Dataset:

- Rich and diverse.
- Includes environmental, temporal, and spatial data.
- Contains pre-engineered features ready for machine learning.

---

6. Methodology

Phase 1: Data Cleaning

- Handling missing values.
- Removing outliers.
- Formatting temporal and spatial columns.
- Encoding categorical variables.
- Standardization & normalization.

Phase 2: Exploratory Data Analysis (EDA)

- Distribution plots.
- Correlation matrix to study relationships between variables.
- Impact of traffic and weather on fares.
- Time series trends analysis.
- Heatmaps for geographic patterns.

Phase 3: Feature Engineering

- Extracting peak hours.
- Converting vehicle and weather conditions to numerical features.
- Creating speed feature = Distance / Time.
- Identifying high-demand zones.

Phase 4: Predictive Modeling

Several models will be tested for fare prediction:

- Linear Regression
- Random Forest Regressor
- XGBoost Regressor
- Gradient Boosting Regressor

The best model will be selected based on:

- RMSE (Root Mean Squared Error)
- MAE (Mean Absolute Error)

- R^2 Score

Phase 5: Dashboard Development Using Power BI

The dashboard will include:

- Interactive map of pickup and drop-off locations.
- Fare analysis by time, day, and month.
- Impact of traffic and weather conditions on fares.
- Driver performance analysis.
- Hotspot and geographic demand analysis.

---

7. Expected Outcomes

- A predictive model capable of forecasting fares with accuracy exceeding 85%.
- Identification of the top 10 factors influencing fares.
- A professional, interactive Power BI dashboard.
- Improved company ability to plan and make data-driven decisions.

---

8. Innovation Points

- Integration of environmental data (weather) with traffic data.
- Geospatial trend analysis (GIS-like insights).
- Practical recommendations for improving dynamic pricing.
- Driver performance evaluation system.
- Potential future development into a pricing engine API.

---

9. Project Deliverables

- Comprehensive PDF report.
- Python notebook.
- Machine learning model file.
- Power BI dashboard.
- Presentation slides.
- Well-organized GitHub repository.

---

10. Conclusion

This project represents a robust application of data science in real-world scenarios, particularly in the smart transportation sector. It combines statistical analysis, machine learning, and spatiotemporal analysis to deliver a comprehensive system that helps companies make data-driven decisions, enhance service quality, and increase customer satisfaction.