Predictive Analytics & BI for Smarter Taxi Pricing

Business Context

Accurate fare prediction is critical for taxi and ride-hailing companies. Overcharging leads to customer dissatisfaction, while undercharging erodes revenue. This project demonstrates how **data analytics, machine learning, and business intelligence tools** can be combined to improve fare prediction accuracy, reduce pricing errors, and build customer trust.

Objective

Predict taxi fares based on trip details (pickup/dropoff location, distance, time, passenger count) and deliver **interactive Power BI dashboards** for business teams to explore fare patterns, demand trends, and model performance.

Methodology & Workflow

1. **Data Cleaning & EDA** – Processed 1M+ records, handled outliers, missing values, and anomalies.
2. **Feature Engineering** – Derived distance, speed, time-based variables (hour, day-of-week, weekend), and categorical features.
3. **ML Modeling** – Built and compared Linear Regression, Random Forest, and XGBoost models.
4. **Model Evaluation** – Assessed performance using MAE and RMSE metrics.
5. **Visualization in Power BI** – Created dashboards for fare distribution, demand patterns, and error analysis.

Key Results

- **Model Performance**
  - Linear Regression → RMSE: 5.64
  - Random Forest → RMSE: 5.28
  - XGBoost → RMSE: 5.02 (**Best**)
- **Insights**
  - Fares strongly correlated with distance and time of day.
  - Prediction errors are higher during late nights and weekends.
  - Business dashboards enabled non-technical users to identify demand surges and pricing anomalies interactively.

Tech Stack

- **Python**: pandas, scikit-learn, XGBoost
- **Power BI**: Interactive visualizations and dashboards

End-to-End Capabilities Showcased

- Data Cleaning & Feature Engineering
- Machine Learning Modeling & Evaluation
- Business-focused Visualization
- Clear Documentation & Project Delivery