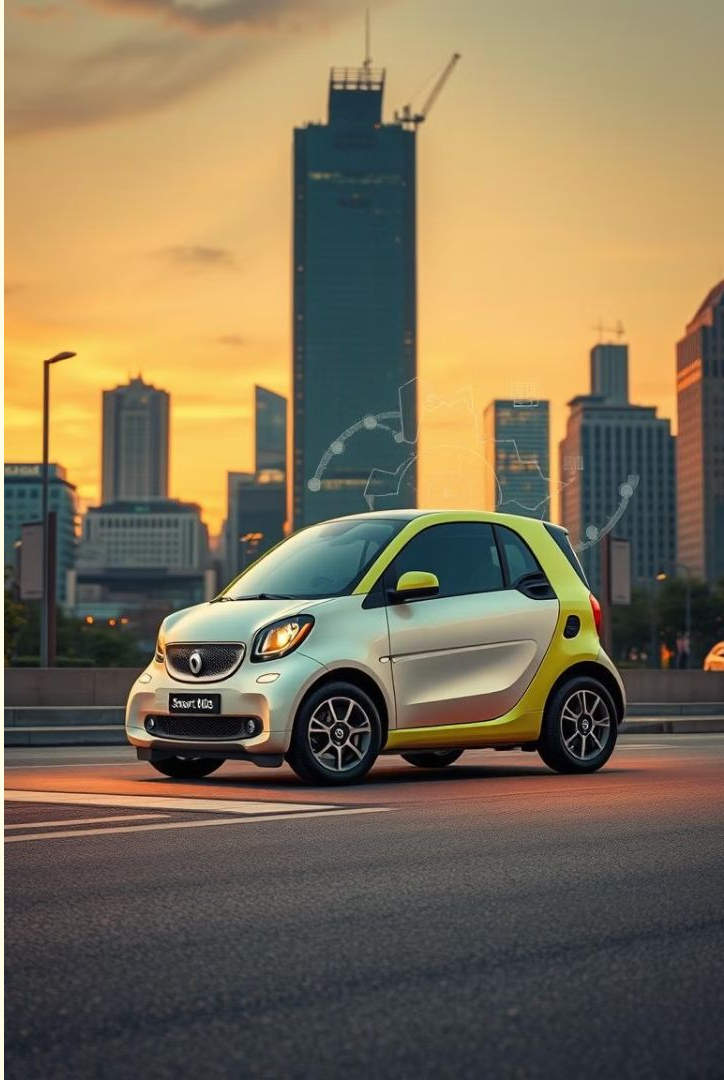
Optimizing Smart Rides Services with Data Science

This presentation explores the application of data science techniques to enhance Smart Rides services. By leveraging advanced algorithms and comprehensive data analysis, we aim to optimize various aspects of the platform, from ride type prediction to fraud detection and pricing strategies. The project addresses inefficiencies and provides actionable insights for improved operational efficiency and customer satisfaction.

We will walk through the project including the data description, preprocessing and EDA as well as algorithms and techniques. By the end of this presentation, you will be confident in our team and our approach.



Introduction: Smart Rides and Data Science

Smart Rides Overview

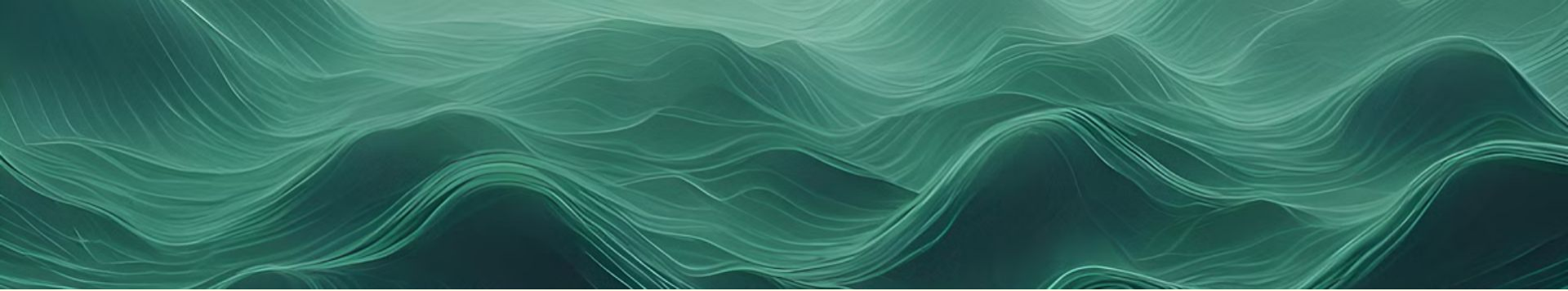
Smart Rides aims to provide convenient and efficient transportation solutions through its ride-sharing platform. By integrating technology with transportation, it offers a seamless experience for both drivers and passengers.

- Easy booking process
- Diverse vehicle options
- Real-time tracking

The Role of Data Science

Data science plays a crucial role in optimizing Smart Rides' services by analyzing vast amounts of data. This analysis helps identify inefficiencies, predict trends, and improve decision-making processes.

- Enhanced operational efficiency
- Improved customer satisfaction
- Data-driven decision-making



Dataset Description and Features

The dataset offers a comprehensive view of Smart Rides' operations through a variety of key features. These features are essential for identifying inefficiencies and developing data-driven solutions to enhance the platform's overall performance.

Ride Characteristics

Date and Time, Pickup and Drop Locations, Ride Distance

Customer Information

Customer ID, Customer Ratings

Booking Details

Booking ID, Booking Status, Booking Value

Vehicle Information

Vehicle Type, Vehicle Turnaround Time

Preprocessing Techniques

Handling Missing Values

Imputation with mean/mode for numerical/categorical features, and dropping rows with critical missing values to ensure data completeness.

Data Cleaning

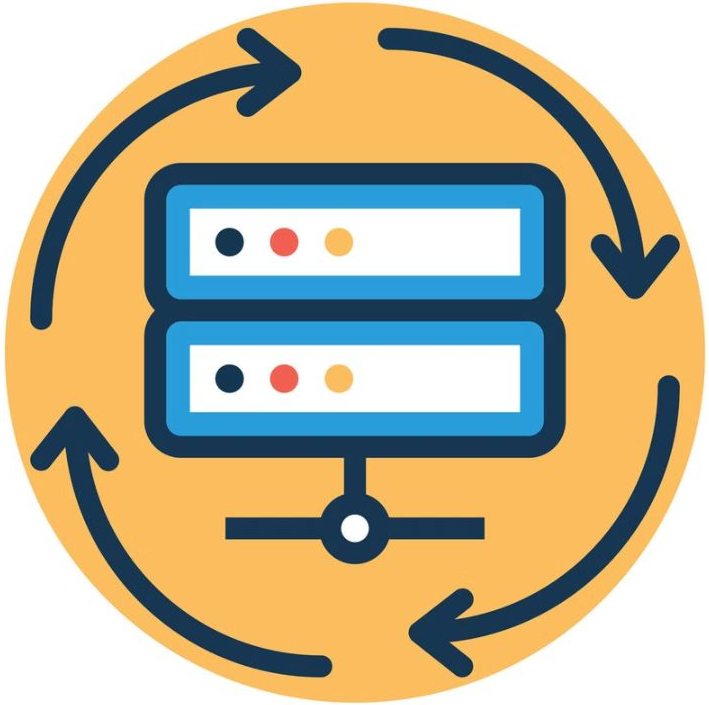
Removal of duplicate records and correction of inconsistent or erroneous entries in columns like dates and distances for data accuracy.

Feature Engineering

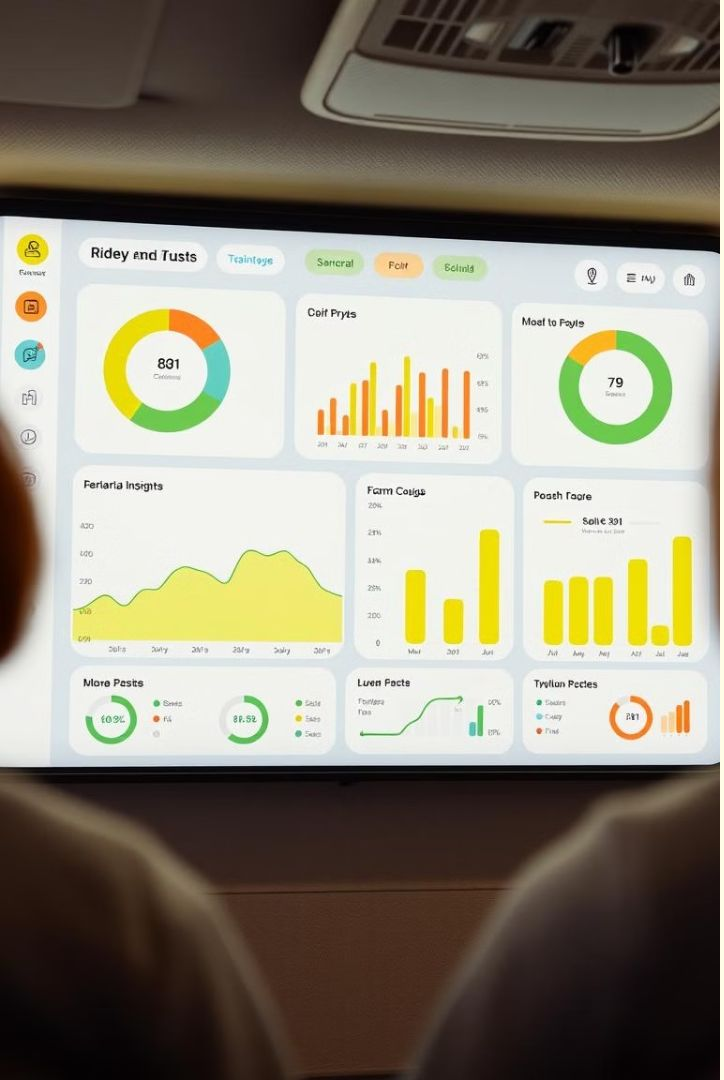
Extraction of new features such as ride duration and time of day, along with categorization of ride distance into short, medium, and long for enhanced data insights.

Normalization & Scaling

Using MinMaxScaler and StandardScaler for scaling numerical features like ride distance and booking value to ensure fair algorithm treatment.



Data processing



Exploratory Data Analysis (EDA)

Descriptive Statistics

Analyzing mean, median, and standard deviation to understand data distribution.

Visualizations

Using box plots, heatmaps, and histograms to reveal data patterns and correlations.

Key Insights

Identifying ride cancellations during peak hours and fare variance in long-distance rides.

EDA provides critical insights into the dataset, revealing patterns and relationships that inform subsequent modeling stages. The visualizations and statistical analyses help uncover key trends and anomalies that can be addressed with targeted solutions.

Algorithms and Techniques Used

Smart Rides uses different data science algorithms to address different questions. Here's a summary of each:

Ride Type & Incomplete Ride Prediction

Predicting ride types and identifying incomplete rides.

- XGBoost
- Logistic Regression
- ARIMA

Ride Fare Prediction

Estimating ride fares based on various factors.

- Linear Regression
- K-Means Clustering

Anomaly Detection

Detecting anomalies and unusual patterns in ride data.

- Isolation Forest
- DBSCAN

Association Rules

Discovering associations and relationships between different variables.

- Apriori Algorithm
- FP-Growth

Comparative Evaluation Measures

1

Classification Metrics

Accuracy, Precision, Recall, F1-Score

2

Regression Metrics

MAE, RMSE

3

Clustering Metrics

Silhouette Score

4

Association Rule Evaluation

Lift, Support

A range of evaluation measures are used to assess the performance of different models. These metrics provide a quantitative basis for comparing models and selecting the most effective solutions for each task, ensuring reliable and accurate results.

EXAMPLE-

Evaluation Results: Predicting Ride Type

XGBoost Model

Accuracy: 92.4%

Precision: 91.8%

Recall: 90.5%

F1-Score: 91.1%

Significantly outperforms Logistic Regression.

Logistic Regression

Accuracy: 82.6%

Precision: 80.4%

Recall: 78.9%

F1-Score: 79.6%

Decent performance, but struggles with non-linear relationships.

XGBoost proves to be a superior choice for predicting ride types, offering higher accuracy and recall compared to Logistic Regression. Its ability to handle complex relationships in the data makes it ideal for this task.

EXAMPLE -

Evaluation Results: Ride Fare Prediction

₹12.6

MAE

Mean Absolute Error

₹18.2

RMSE

Root Mean Squared Error

85.7%

R^2 Score

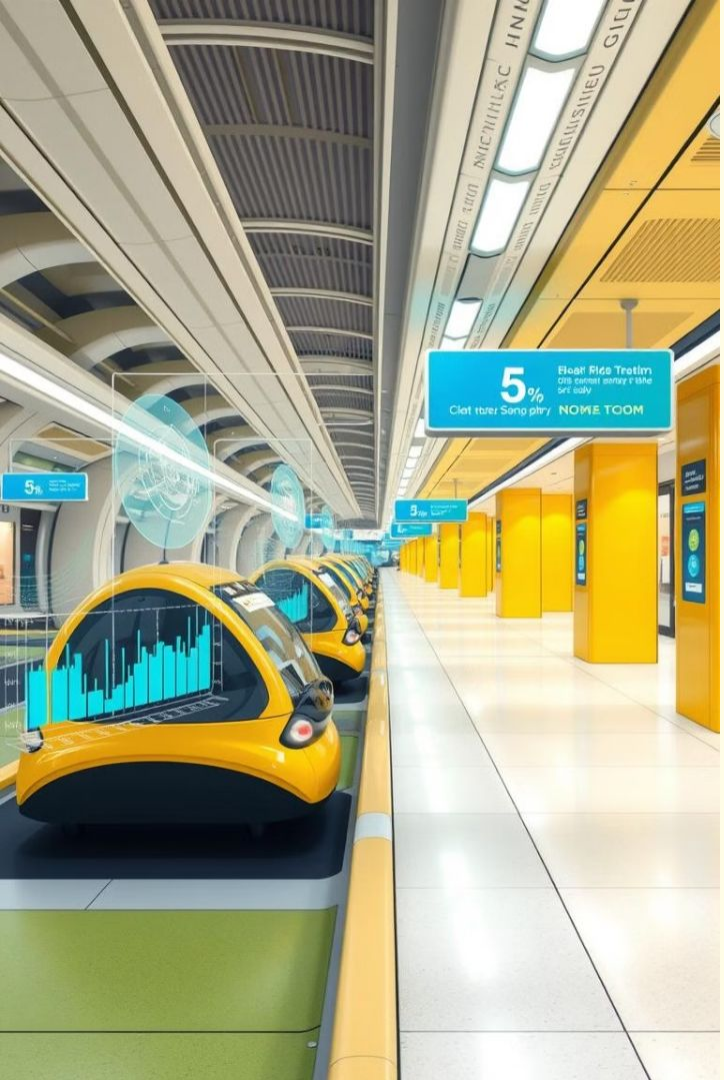
Explains variance in ride fares

0.68

Silhouette Score

Effectively segments rides

Linear Regression provides a high R^2 score, indicating its effectiveness in explaining ride fare variance based on distance and time. K-Means Clustering effectively segments rides into short, medium, and long distances, aiding in price optimization and customer behavior analysis.



Conclusion and Future Work

Ride Categorization Models

Successfully predict ride types with high accuracy.

Clustering Techniques

Identified customer behavior patterns.

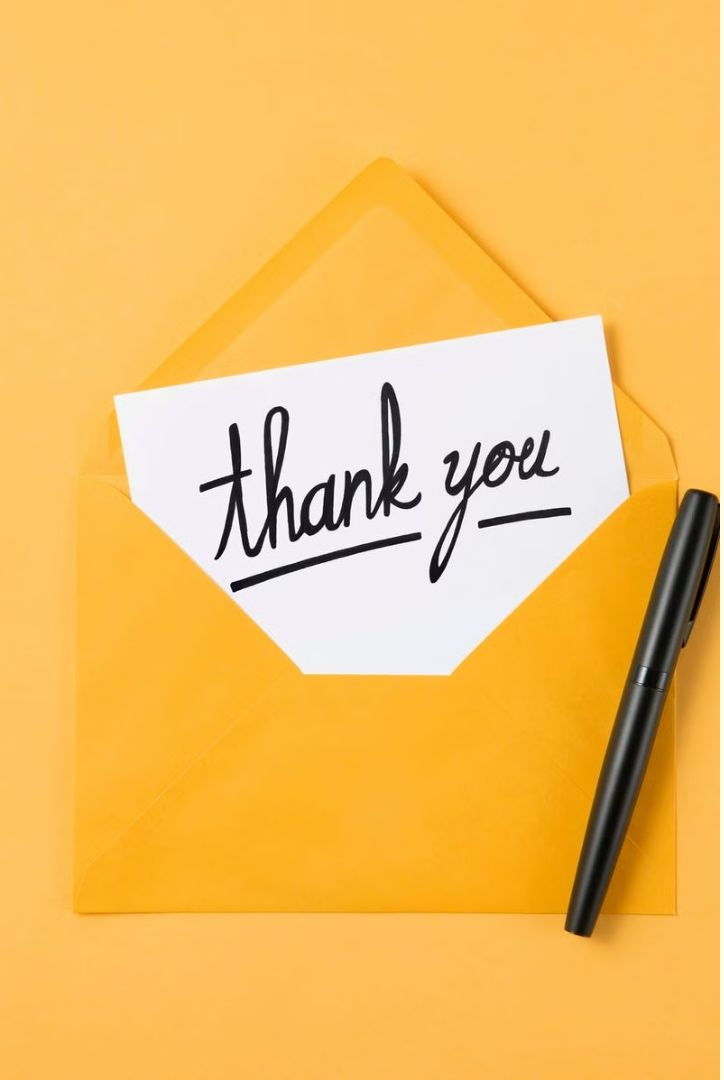
Association Rule Mining

Revealed common ride route patterns.

Anomaly Detection

Identified fraudulent cancellations.

The project successfully leveraged data science techniques to optimize Smart Rides services. Future work includes implementing real-time pricing optimization, using reinforcement learning for dynamic driver allocation, and deploying models in production for live predictions and improved decision-making.



BY-

POOJA MADAVI- 612203108

MEDHA BHARTI -612203115

EKTA MESHRAM- 612203117

MANYA ABROL- 612203112