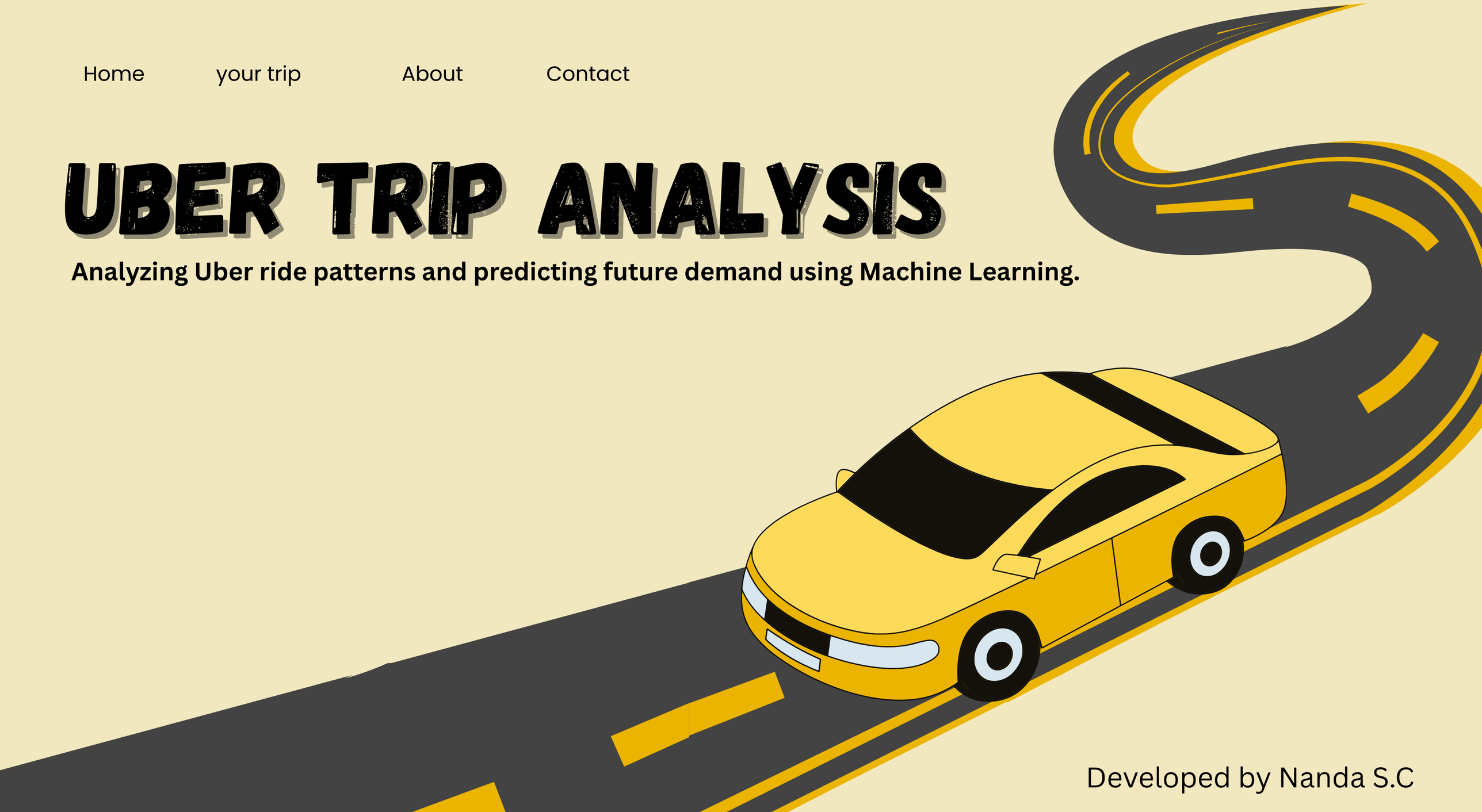
Home your trip About Contact

UBER TRIP ANALYSIS

Analyzing Uber ride patterns and predicting future demand using Machine Learning.



Developed by Nanda S.C

THE PROBLEM

With the growing popularity of Uber, understanding and predicting ride demand has become crucial for optimising driver allocation and improving customer satisfaction.

However, daily trip patterns fluctuate significantly due to factors such as weekdays, weekends, holidays, and time of month.

The challenge is to analyse historical trip data, uncover hidden patterns, and forecast future trip counts accurately using Machine Learning techniques.

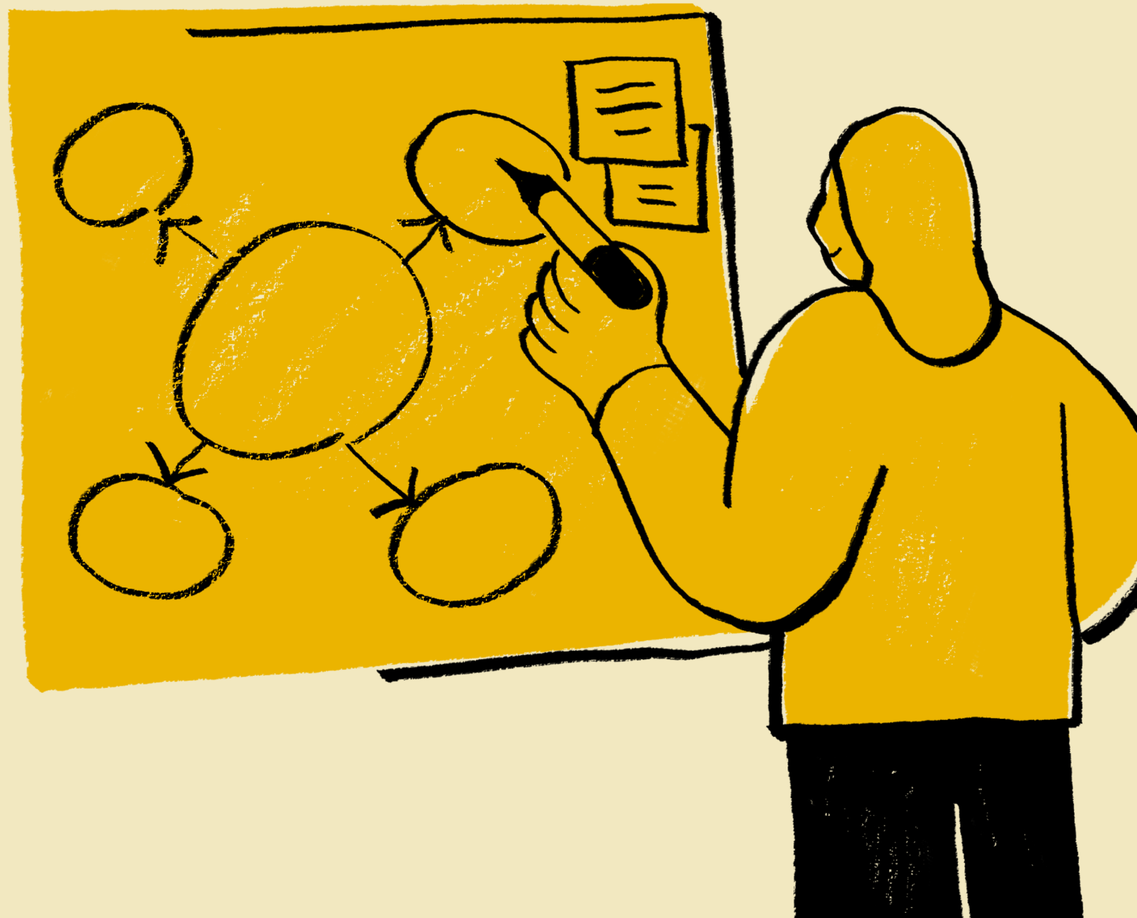
OBJECTIVE

- Analyze historical Uber trip data (Jan–Feb) to identify patterns and demand trends.
- Build a Machine Learning model to forecast future trip demand accurately.
- Develop an interactive Streamlit dashboard for real-time data visualization and prediction.

“From raw data to real-time prediction — a complete machine learning pipeline.”

DATA CLEANING & PREPROCESSING

1. Loaded Raw Data
   - Imported Uber Jan–Feb dataset
   - Inspected data types and missing values
2. Handled Missing & Invalid Entries
   - Removed duplicates
   - Filled/cleaned null or inconsistent records
3. Datetime Conversion
   - Converted date column to proper datetime format
   - Extracted day, month, weekday features
4. Column Renaming & Selection
   - Standardized column names → dispatching_base_number, date, active_vehicles, trips
5. Data Aggregation
   - Aggregated hourly data → daily totals
   - Summed total trips per day
6. Saved Cleaned Data
   - Exported to data/processed/uber_hourly.csv for further analysis



FEATURE ENGINEERING

1. **Time-based Features:** Extracted Day, Month, Weekday from the datetime column. Helped identify weekday vs weekend demand trends.
2. **Lag Features:** Created Lag_1 (previous day trips) and Lag_7 (previous week trips). Enabled the model to recognize short-term and weekly patterns
3. **Rolling Statistics:** Added 7-day rolling average to smooth fluctuations and highlight trend continuity.
4. **Feature Scaling (if applicable):** Normalized numeric features to keep all variables on a similar scale.
5. **Final Feature Set:** Stored the engineered features for model training and evaluation.

MODEL TRAINING AND SELECTION

To build and compare multiple regression models for accurate Uber trip demand forecasting.

Random Forest Regressor
Ensemble of decision trees to reduce overfitting.
Good performance, slightly slower

Gradient Boosting Regressor
Sequential trees correcting previous errors.
Improved accuracy

XGBoost Regressor
Optimized boosting algorithm with regularization.
Best performance (lowest MAE & RMSE)

VALIDATION APPROACH:

- Used TimeSeriesSplit (5 folds) for cross-validation.
- Prevented data leakage between time steps.
- Evaluated models using metrics: MAE, RMSE, and MAPE.

Implementation steps:

- Split Data → Train & Test sets.
- Train Models → Fit multiple regressors.
- Validate → Compare performance metrics.
- Select Best Model → Save best_model.joblib.

XGBoost Regressor — achieved lowest RMSE and highest stability across folds.

MODEL EVALUATION

To evaluate model performance using key regression metrics and visualize actual vs predicted Uber trip trends.

Metric	Description	Value
📉 MAE	Mean Absolute Error	1450.67
📈 RMSE	Root Mean Square Error	2105.32
🎯 MAPE	Mean Absolute Percentage Error	6.50%

Interpretation:

- The low MAE and RMSE indicate minimal prediction error.
- MAPE of 6.5% shows high accuracy in trip demand forecasting.

Model closely follows the actual demand trend with minimal deviation.

INTERACTIVE STREAMLIT DASHBOARD

To visualize Uber trip trends, analyze model performance, and predict future trip demand through an interactive web dashboard.

Overview Tab
Displays key model metrics (MAE, RMSE, MAPE) and trip trends over time.

Model Insights Tab
Visualizes feature importance and residuals for model interpretability.

Prediction Tab
Allows users to upload new Uber trip data (CSV) and forecast next-day demand.

Home your trip About

PREDICTION MODULE & RESULTS

Upload File (CSV)

- User uploads a new file containing datetime and Count columns
- Example: new_uber_trip_data.csv

Feature Extraction

- Automatically extracts day, month, weekday, lag features

Prediction Generation

- Pre-trained XGBoost model forecasts the next day's trip demand

Visualization

- Displays predicted value and plots it on a line chart with recent 7 days

"Predicting tomorrow's demand — one click, one file, one forecast."

KEY INSIGHTS & LEARNINGS

-Weekly Demand Pattern

- Uber rides peak during weekends, showing ~25–30% higher demand compared to weekdays.

-Lag Features Matter

- Lag_1 (previous day) and Lag_7 (previous week) are top predictors — confirming strong short-term and weekly correlations.

-Stable Model Performance

- Low MAE (1450) and MAPE (6.5%) indicate the model generalizes well on unseen data.

-Data-Driven Opportunities

- Forecasts can help optimize driver allocation, pricing strategy, and city planning for peak hours.

- Gained hands-on experience in end-to-end data science workflow — from data cleaning to dashboard deployment.
- Improved skills in Python, Pandas, Streamlit, and XGBoost.
- Learned how to derive business insights from data patterns.
- Enhanced presentation and analytical thinking through real-world problem solving.

CONCLUSION

Summary

Successfully built an end-to-end Machine Learning pipeline for Uber trip demand forecasting.

Invitation

Developed an interactive Streamlit dashboard for visualization and prediction.

Achieved

high accuracy (MAPE ~6.5%), capturing weekly demand patterns effectively.

Demonstrated

- strong integration of data analysis, modeling, and visualization skills.

THANK YOU!

“From data to decisions — leveraging Machine Learning for Uber trip forecasting.”

Nanda S.C
nandachowgle@gmail.com
https://github.com/NandaSC833
https://www.linkedin.com/in/nanda-s-c-598358349/