SEKT EL-SALAMA

Before you crash ,
we predict !!

DATA SCIENCE PROCESS OVERVIEW

Problem Understanding

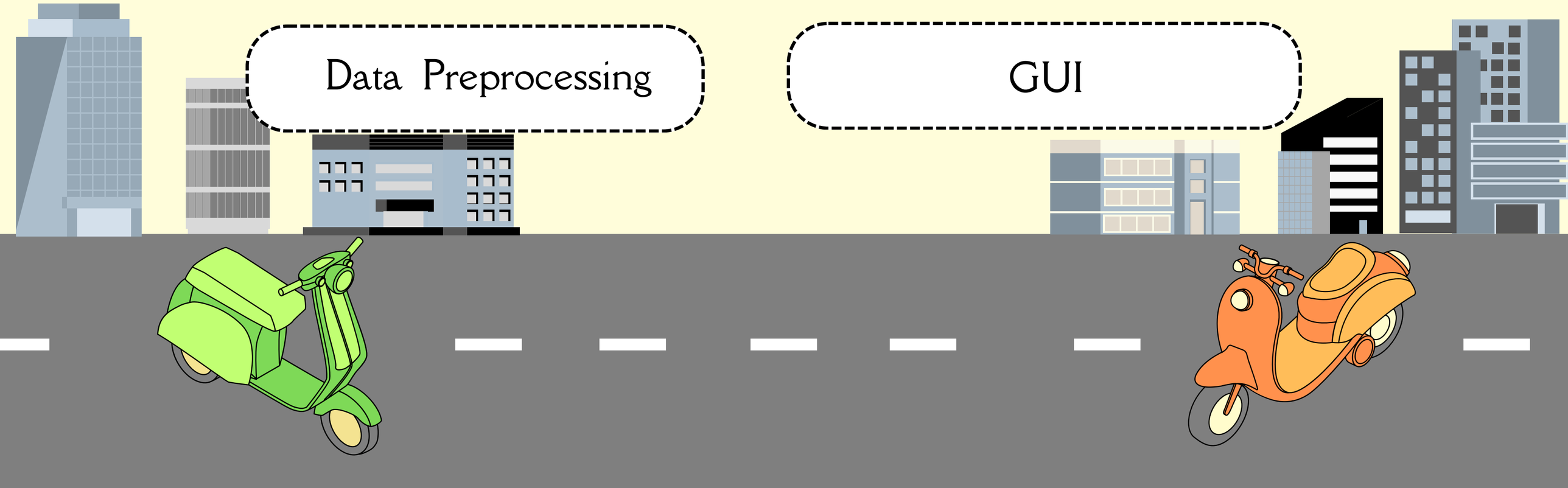
Exploratory Data Analysis

Data Collection

Model Selection & Training

Data Preprocessing

GUI

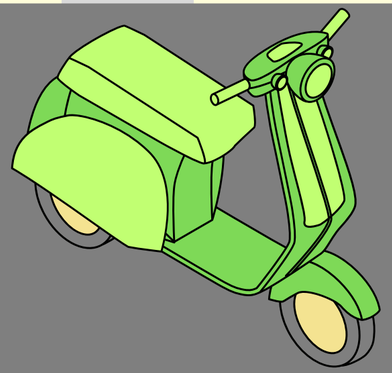


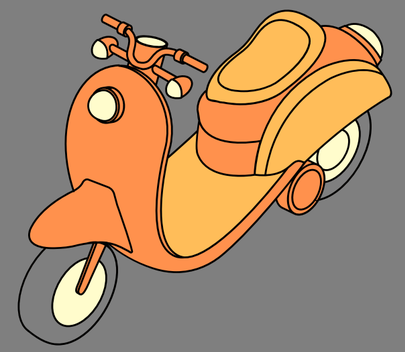
PROBLEM UNDERSTANDING

The main idea of this project is to predict the likelihood of a traffic accident based on real-world road and traffic data.

We use several features like traffic speed, number of vehicles, road occupancy, traffic lights, and weather conditions to train a machine learning model that can tell us if an accident is likely to happen.

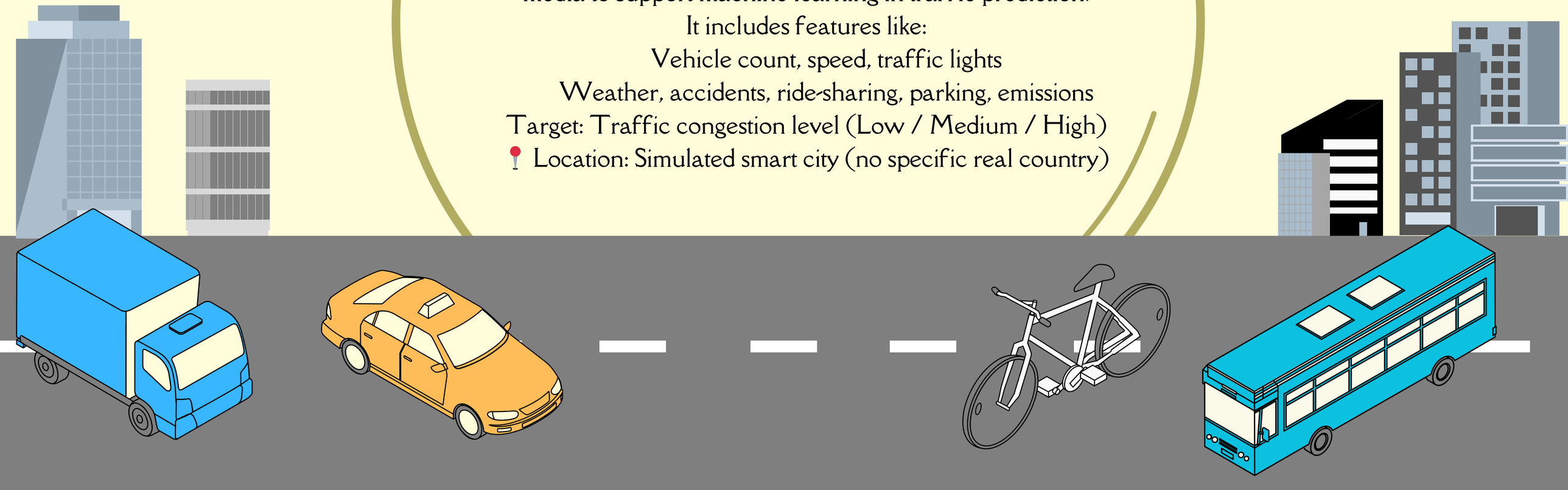
This prediction can help improve road safety, assist traffic management systems, and reduce response time in emergencies





DATA COLLECTION

We used a dataset from Kaggle called:
Smart Mobility and Traffic Optimization Dataset
It contains simulated smart city data that reflects real-world traffic, weather, and environmental conditions.
The dataset combines information from sensors and social media to support machine learning in traffic prediction.
It includes features like:
Vehicle count, speed, traffic lights
Weather, accidents, ride-sharing, parking, emissions
Target: Traffic congestion level (Low / Medium / High)
📍 Location: Simulated smart city (no specific real country)



DATA PREPROCESSING

we cleaned and prepared the raw traffic dataset using a custom ETL pipeline

Extract

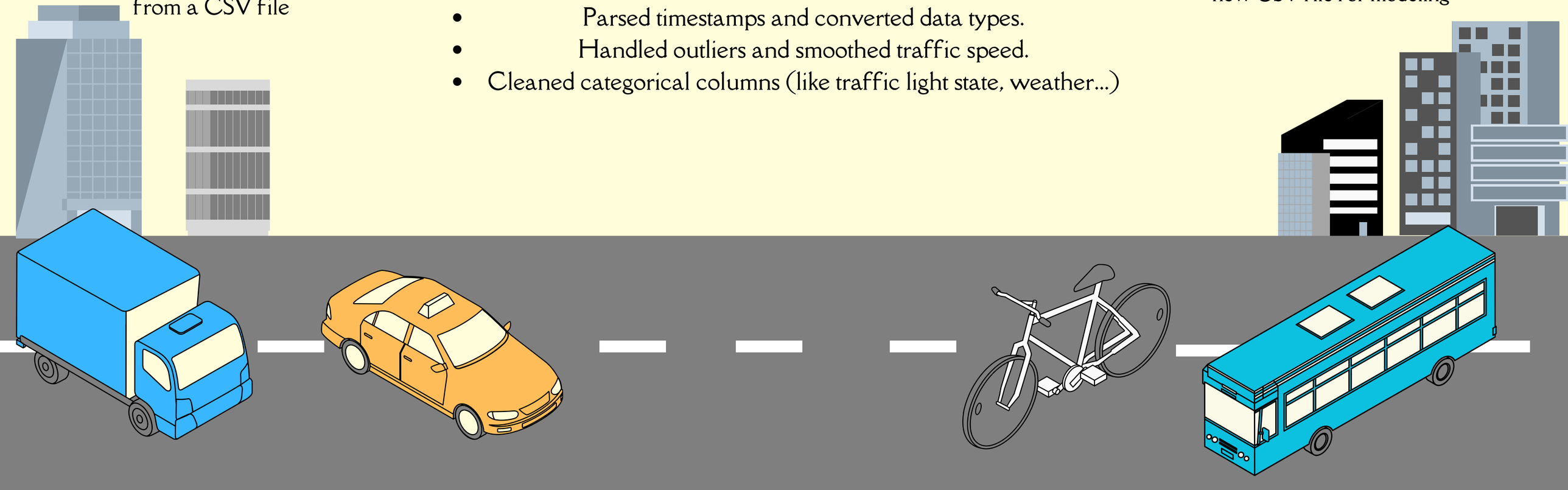
Loaded the raw data from a CSV file

Transform

- Removed duplicates and missing critical values.
- Parsed timestamps and converted data types.
- Handled outliers and smoothed traffic speed.
- Cleaned categorical columns (like traffic light state, weather...)

Load

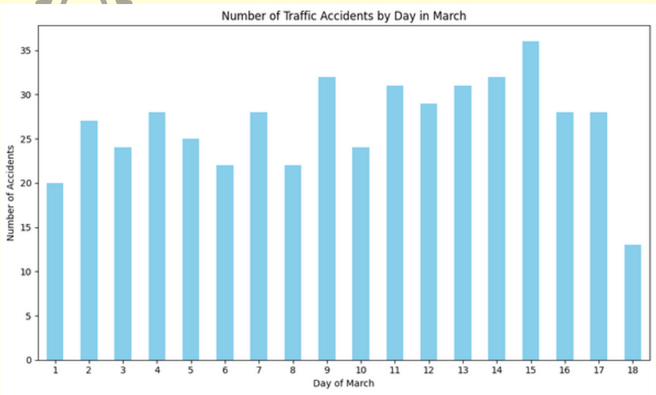
aved the cleaned dataset into a new CSV file for modeling



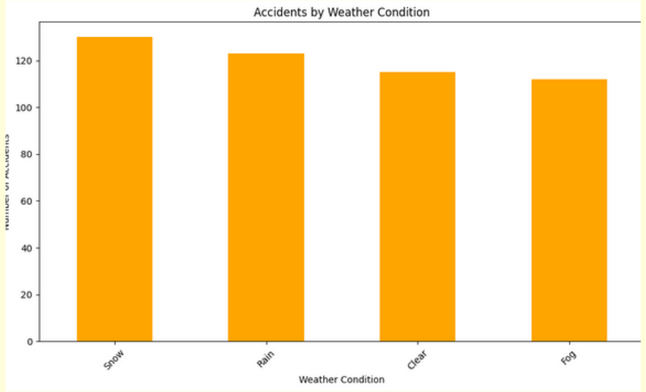
EXPLORATORY DATA ANALYSIS (EDA)

In this step, we analyzed the cleaned traffic accident data using various techniques:

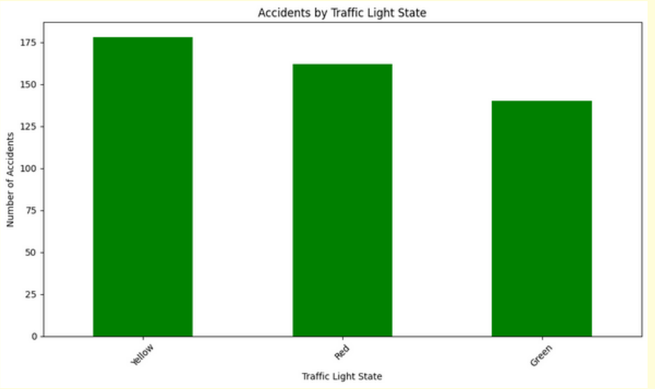
Number of accidents by day in March.



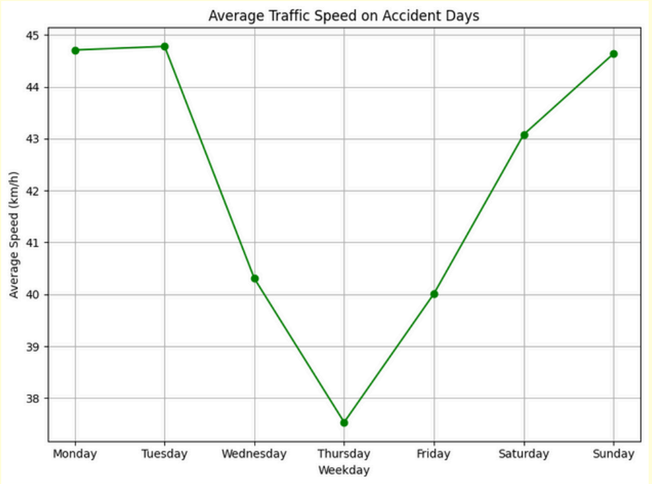
Accidents by weather and traffic light states.



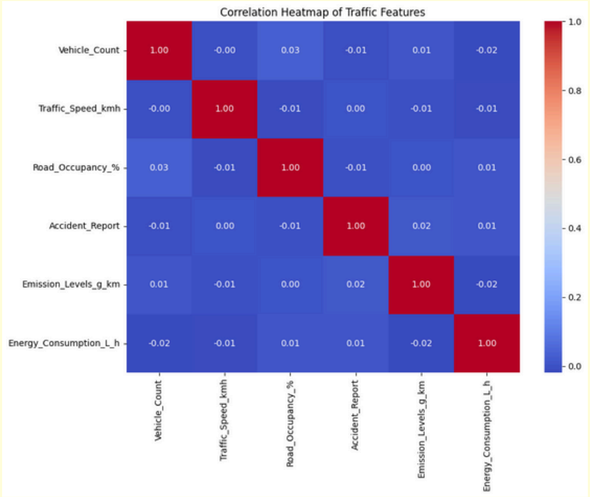
Average traffic speed on accident days.



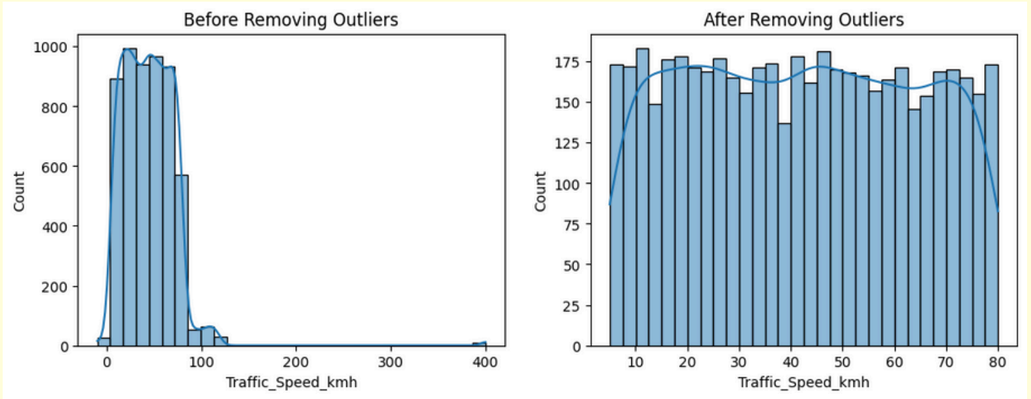
Average Traffic Speed on Accident Days by Weekday



correlations and relationships using a correlation heatmap and boxplots.



accident rates by speed ranges and distribution of traffic speed before and after removing outliers.



MODEL SELECTION & TRAINING

Preprocessing

- Feature Engineering: Added time-based features like hour, weekday, rush hour, and interaction terms
- Target Encoding for Weather_Condition.
- Handled skewed features with PowerTransformer and used SMOTE to balance the dataset.

Model Selection

- Used RandomForestClassifier, XGBoost, and SVM.
- Combined models in a VotingClassifier for better performance.

Hyperparameter Tuning

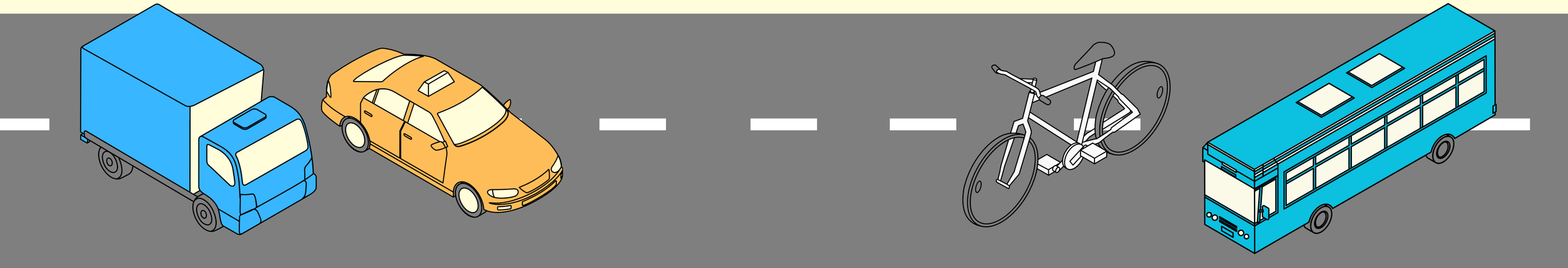
Performed Grid Search on XGBoost to find the best parameters.

Hyperparameter Tuning

- Accuracy and F1-score used for performance evaluation.
- Feature importance analyzed with RandomForest.

Hyperparameter Tuning

Saved the trained Voting Classifier and preprocessing pipeline.



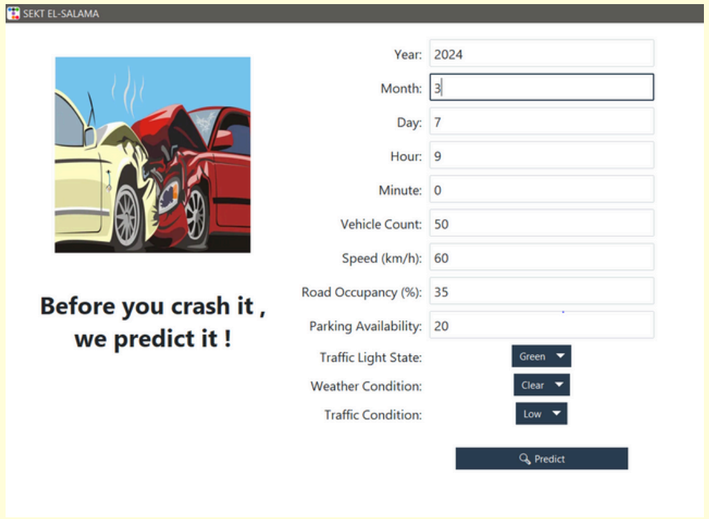
GUI

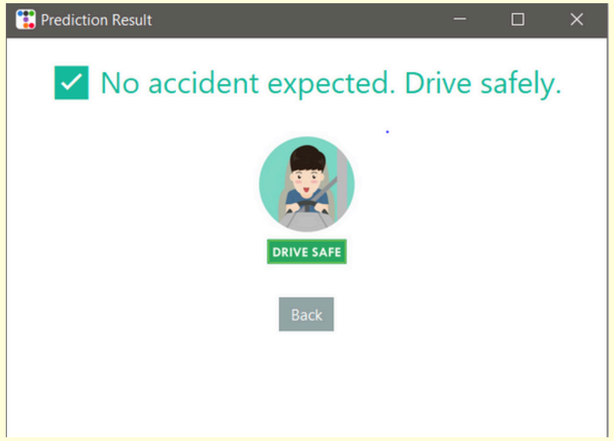
provides a graphical user interface (GUI) where users can input details such as the date, time, vehicle count, traffic speed, weather condition, and more.

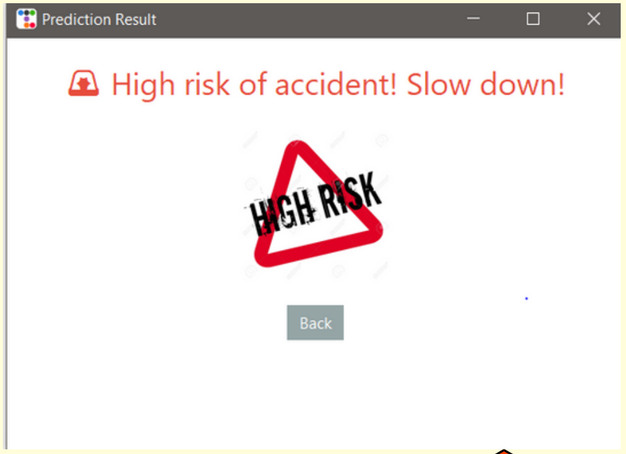
Prediction Logic: After the user fills out the form, the app processes the data and uses the pre-trained model to predict the risk of an accident.

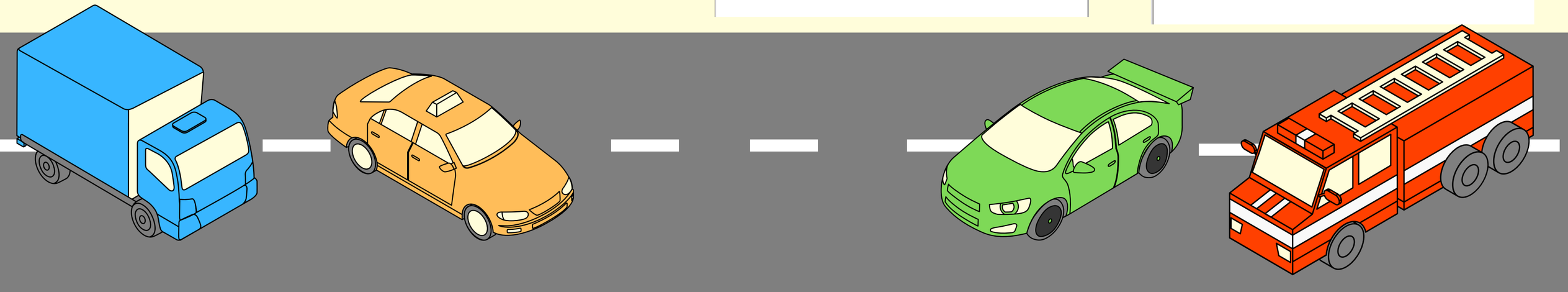
Result Display: The result is shown with a message indicating whether there is a high or low risk of an accident, along with an appropriate image.

Error Handling: Input validation ensures the data is correct, and any errors are displayed in an alert box.









TANK YOU

SAMAR FAWZY ABOUSAMRA

EMAN MOUSA KAMER

EMAN HUSSEIN BATIE

HANA MOHAMED GOHAR

SUPERVISOR
DR:HEBA EL-HADIDI