# Traffic prediction Froth draft

**Project Overview**

You are working on a traffic prediction project using historical traffic data. Your primary goal is to develop a model that can accurately predict traffic volume based on various features.

**Data Overview**

Your dataset contains the following columns:

* **Date**: Date of the observation.
* **Road/Intersection Name**: The name of the road or intersection.
* **Traffic Volume**: The target variable for prediction.
* **Average Speed**: Average speed on that day.
* **Travel Time Index**: Indicator of travel time.
* **Congestion Level**: Level of congestion.
* **Incident Reports**: Number of incidents reported.
* **Environmental Impact**: Impact of the environment on traffic.
* **Public Transport Usage**: Usage of public transport.
* **Pedestrian and Cyclist Count**: Count of pedestrians and cyclists.
* **Weather Conditions**: Weather conditions on that day.
* **Roadwork and Construction Activity**: Ongoing roadwork or construction.
* **Day of Week**: Day of the week.

**Steps Implemented So Far**

1. **Data Cleaning**:
   * Removed outliers and handled missing values.
2. **Feature Engineering**:
   * **Lag Features**: Created lagged versions of traffic volume and possibly other features.
   * **Date Features**: Extracted useful time-based features (e.g., day of the week).
   * Normalized numerical features using MinMaxScaler.
3. **Baseline Model**:
   * Developed a simple regression model to establish a performance benchmark.
   * Evaluated the baseline model using RMSE and MAE:
     + RMSE: 0.1835
     + MAE: 0.1536
   * This model provided a baseline for comparison with future models.
4. **Next Steps Planned**:
   * **Time-Series Specific Features**: Prepare data for LSTM by creating time-series sequences and potentially additional features.
   * **LSTM Model Development**: Build and train an LSTM model for predicting traffic volume.

**Next Steps to Implement**

1. **Reshape Data for LSTM**: Create sequences of data in the format required for LSTM input.
2. **Train-Test Split**: Chronologically split the data into training, validation, and test sets.
3. **Model Definition**: Define the LSTM model architecture using Keras or TensorFlow.
4. **Training**: Train the LSTM model and evaluate its performance.