**Project Proposal**

Based on the feedback received, I have decided to work on **Freight Transport Data**

. This proposal consists of problem identification, initial understanding of the dataset and approach I will be taking to solve this problem.

**Problem Identification**

|  |  |
| --- | --- |
| **Problem Statement Sheet:**  How can we reduce the cost of transportation for ABC company by Traffic flow prediction for Belgium's Highways? | |
| **Contex:**  All owners of Belgian lorries having a Maximum Authorized Mass exceeding 3.5 tonnes must pay a kilometer charge. Every road user who is not exempt from the toll must then install an On-Board Unit (OBU) recording the distance that a lorry travels within Belgium. Our aim is to reduce the cost of transportation for ABC company by accurately predicting traffic flow on Belgium’s Highways. | **Constraints Within the Solution Space:**   1. All the necessary streets from OpenStreetMap. |
| **Criteria of Success:**  Reduce the transportation cost for ABC company by at least 30%. | **Stakeholders to Provide Key Insight:**   1. Alka Luqman - Mentor 2. Freight manager BC company 3. Member from Database team. 4. Senior manager ABC company |
| **Scope of Solution Space:**   1. Time stamp 2. GPS Position (latitude, longitude) 3. Speed (engine) 4. Direction (compass) 5. daily activity (e.g. loading/unloading goods, stopping at depots’ parking slot) | **Key Data Sources:**   1. Kaggle: Freight Transport Data( <https://www.kaggle.com/giobbu/belgium-obu> ) |

**Initial understanding of the dataset**

Each OBU record contains an anonymous Identifier (ID resetting every day at 3 a.m.), the Timestamp, the GPS Position (latitude, longitude), the Speed (engine), and the Direction (compass).

The OBU data require a pre-processing step to predict the traffic conditions at network scale since the trucks recorded positions may refer not only to streets but also to areas where trucks perform the daily activity (e.g. loading/unloading goods, stopping at depots’ parking slot)

**Approach**

Time series

Baseline model

Neural network/ Deep learning

Models Performance Comparison