**CAPSTONE PROJECT PROGRESS REPORT**

**Title**: Trip by Distance.

**Name**: Nurul Hakimah Mohd Zaki.

**Type**: Time Series Analysis.

**Date**: 9 June 2024.

SECTION 1: SUMMARY

A dataset, called Trip by Distance has been presented & approved during Part 2 lightning talk on 2 June as my project. The dataset is obtained from Kaggle & US government site. It is a time series dataset, with number of trips and distance travelled by counties and states.

SECTION 2: ACTIVITIES & ACTION TAKEN

During part 2 lightning talk (2 June), I have proposed a problem statement & import the data. The data is huge, well organized, with manageable null values.

I continue with data cleaning, dropping null values, change dates to DateTimeIndex, filtered DataFrame to Cook County in Illinois only. I then plot the data in time series format.

The data is stationary as checked by Augmented Dickey-Fuller test.

ARIMA model was performed in the final year of the data ( 1 June 2022 – 1 June 2023).

I performed train/test on the model.

SECTION 3: ISSUES

Initially I had issue with problem statement, that I have revised. Upon performing Augmented Dickey-Fueller (ADF), I received memory error trying to allocate a large array. I realized the data is too big, and I have chosen to specify the model to Chicago city, in Cook county, Illinois as it is the second-most congested city in the world in 2022.

SECTION 4: FUTURE PLAN

After stationary check, I plan to transform the data using Differencing and Least Square removal methods. Next, I will perform ARIMA modelling and Exponential Smoothing method for forecasting. I plan to use Cross validation & RMSE as error metrics. I plan to research more on other modelling & forecasting methods.