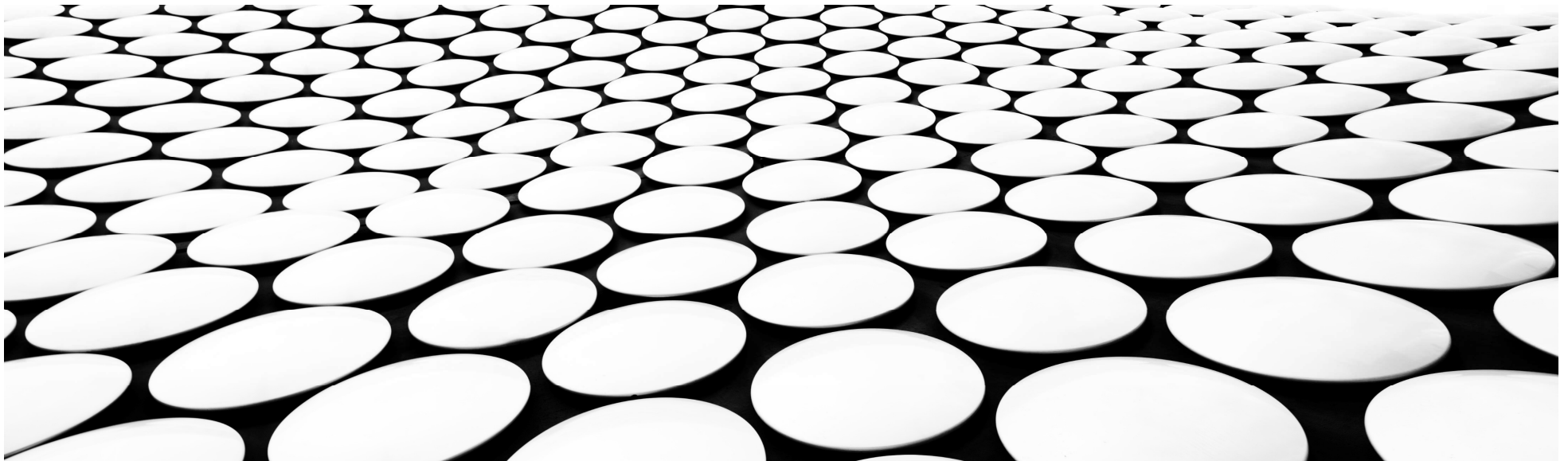# CAPSTONE PROJECT

TRIP BY DISTANCE

PART 2 PRESENTATION

NURUL HAKIMAH MOHD ZAKI

TOPICS CHOSEN

- **1. Trip by Distance (Time-Series)** ⬅
- 2. Fedex Data (Classification)
- 3. Earthquake Magnitude, Damage and Impact (Multi-Classification)

TOPIC: TRIP BY DISTANCE

- Problem Statement

    - Traffic congestion in the US has gotten worse over the past year. According to Inrix, the average driver in the US spent time 29% more than in 2021, equivalent to $134 more spent on fuel in a year. This problem is expected as US population has steadily increased by 300 million people every year. By modelling the population movements in distance and frequencies over time, the pattern can be analyzed to determine the areas with significant changes that require attention by the town ship planner. Modelling and forecasting can be done by counties, states and nationwide level. Necessary development such as increasing public transport coverage and frequencies, adding more highways or smart traffic lights shall be evaluated for future needs. Chicago will be a focused city, as it is the second-most congested city in the world in 2022.

    - S – I will use ARIMA in Time Series analysis, and develop a prediction model using Seasonal and Trend decomposition using Loess & Granger Causality Analysis

    - M – using Cross Validation and RMSE as a metric to measure accuracy

    - A – Yes it is achievable, data is organized and abundant with small amount of dependant variables.

    - R – It is relevant to the:

        1. Local and federal town planners can plan for commercial/residential areas,

        2. Marketing analyst and businesses understand changes in consumers' needs and its relevance (supermarket/childcare/setting up new   offices).

        3. Value added in making decisions:

            For example: Should a company setup and office close to the residential area or it doesn't matter where if people are willing to travel.

                Should district office add more roads/highways/public transports as people prefer to travel more?

    - T – I plan to finish by 1st of July.

TOPIC: TRIP BY DISTANCE

- **Proposed method and models**
  - Time Series Model,
  - Time Series Visualization,
  - Autoregressive Integrated Moving Average (ARIMA) (models the next value in a time series based on linear combination of its own past values and past forecast errors)
  - Forecasting:
    1. Seasonal and Trend decomposition using Loess (modeling and forecasting each component separately),
    2. Granger Causality Analysis (determines whether one time series can predict future values of another time series)

- **Risks and assumptions**
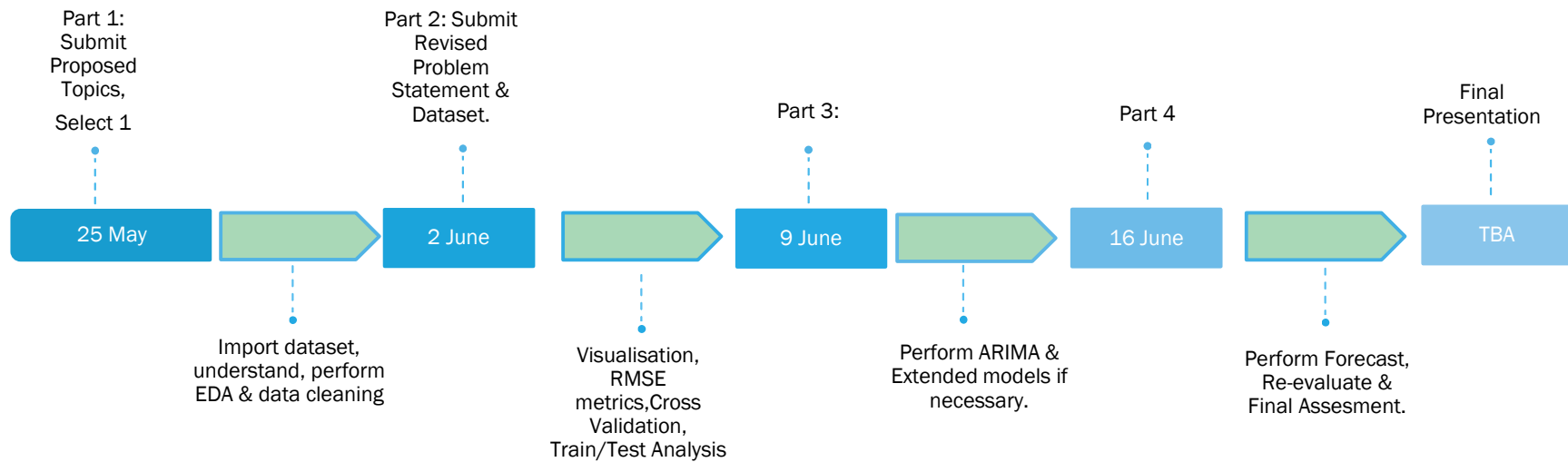 - Steep learning curve
  - Time constraint

- **Revised initial goals and success criteria**
  - Initial goals to have a time series model of accuracy >90%, less than 1 MAE, with as low as possible RMSE.

- **Data source**
  - Trips by Distance (US) (kaggle.com)
    - Bureau of Transportation Statistics by the Maryland Transportation Institute and Center for Advanced Transportation Technology Laboratory at the University of Maryland
  - Time Series Analysis and Forecasting – GeeksforGeeks
  - The Complete Guide to Time Series Forecasting Models | by Peter Wainaina | Medium
  - How much did traffic congestion cost US last year? (https://www.govtech.com/question-of-the-day/how-much-did-traffic-congestion-cost-the-u-s-last-year#:~:text=Traffic%20congestion%20in%20the%20U.S.,to%20%24869%20in%20lost%20time.)

ROADMAP

Part 1: Submit Proposed Topics, Select 1

Part 2: Submit Revised Problem Statement & Dataset.

Part 3:

Part 4

Final Presentation

| 25 May | | 2 June | | 9 June | | 16 June | | TBA |

Import dataset, understand, perform EDA & data cleaning

Visualisation, RMSE metrics,Cross Validation, Train/Test Analysis

Perform ARIMA & Extended models if necessary.

Perform Forecast, Re-evaluate & Final Assesment.

THANK YOU FOR TODAY

KEEP THE QUESTIONS COMING