**PROGRESS REPORT 1**

**CAPSTONE PROJECT DATA SCIENCE AND ANALYTICAL COHORT 3**

NAJIBAH BINTI ALI

Time -Series Analysis (Arima Model)

Predicting The Usage of Household Power Consumption for Years to Come

|  |  |  |
| --- | --- | --- |
| **No.** | **Points** | **Remarks** |
| **1** | Do you have data fully in hand and if not, what blockers are you facing? | **First Data:**   * Data on USA historical flood *records* from 1980 – 2015 * Data has many blank rows and not complete * Dataset is not relevant with initial objective * Blockers: Data does not provide solution for problem statement * Decision to change dataset   **Second Data:**   * Source out data for a household power consumption for 2007 * Data is assumed to be in complete form with manageable nulls and format * Dataset is relevant with objective * Dataset is chosen for project |
| **2** | Have you done a full EDA on all your data? | **Univariate**  Graphical:  Non-graphical |
| **Bivariate**  Graphical:  Non-graphical: |
| **3** | Have you begun the modeling process? How accurate are your predictions so far? | Time Series Analysis  ARIMA Model  Stages:  1.  2.  3. |

|  |  |  |
| --- | --- | --- |
| **4** | What blockers are you facing, including processing power, data acquisition, modeling difficulties, data cleaning, etc.? How can we help you overcome those challenges? | **Problem**  Data acquisition:  Modeling difficulties:  Data cleaning:  Processing power: |
| **Overcoming**  Data acquisition:  Modeling difficulties:  Data cleaning:  Processing power: |
| **5** | Have you changed topics since your lightning talk? Since you submitted your Problem Statement? If so, do you have the necessary data in hand (and the requisite EDA completed) to continue moving forward? | **Topic changed**  **Data 1**: Predicting the risk of getting flood  **Data 2**: Predicting the household power consumption  Data 2 is a complete dataset, and the EDA is completed on the dataset. |
| **6** | What is your timeline for the next week and a half? What do you have to get done versus what would you like to get done? | Focus on modeling the prediction model (ARIMA) and handling error and accuracy issue for the model  **What have to be done:**   1. Clean the data 2. Prepare data to be appropriate for analysis 3. EDA data 4. Understanding data   **What would like to do:**   1. Complete the modeling stage 2. Run the prediction on testing model 3. Handling all errors |
| **7** | What topics do you want to discuss during your 1:1? | **Topics:**   1. How to know the bivariate or multivariate EDA is correct 2. EDA output accuracy 3. Null handling 4. Understanding the graphical EDA 5. How to reduce error 6. Method to tackle the error |