**MEETING MINUTES –**

**[Week 3 MEETING 1 OF INTEGRATED PROJECT\_GROUP #Carlos]**

**MEETING INFORMATION**

***Objective:*** Check Assignment, New Assumptions Teach

***Date:*** 06/03/2024

***Time:*** 8:00 AM EDT ***Meeting Type:*** Virtual Meeting

***Timekeeper:*** Carlos ***Note Taker:***  Luyang

***Attendees:*** Carlos (Mentor), Luyang(Mentee), Orange (Mentee), Ruizhe (Mentee)

**Meeting Documents**

* eda\_visuals.ipynb
* dataset: “mpg”

**Progress Report**

Carlos updated the schedule, starting next week we meet mostly in July, except 2nd of July and 31st of July.

**Special Business**

1. **Exploratory Data Analysis(EDA):**

* Carlos went over the following topics.
* Data in real life is always dirty, we always need to clean it. EDA aims to summarize its features using visuals. EDA helps with data processing and feature selections, identify relationships between variables, detect outliers, and provide insights into further analysis. The steps in EDA are 1. Data Collection(e.g. csv, xlsx, database, etc), 2. Data Cleaning(Carlos stressed the importance of dealing with missing values on a case by case basis), 3. Data Transformation(e.g. min-max normalization, z-score normalization).
* Descriptive Statistics: Count, mean, std, min, 25% 50% 75% quantiles, max.

Example in python: mpg.describe(include=np.number).T. If median < mean, the distribution is right skewed, and vice versa.

* IQR for a standard normal distribution.
* Data visualization:

1. Univariate analysis(e.g.Box plot, histogram, count plot, pie chart(not recommended in good practice)).
2. Bivariate analysis(scatter plot, boxplot, linear correlations)
3. Multivariate Analysis(pairplots, heatmaps)

* How to do EDA with python
* Anscobe’s quartet to illustrate the importance of visualization in analysis.

1. **Statement of Work:**

* Up to introduction are finished
* Objective: look at the first presentation - problem definition
* e.g. Use machine learning techniques to understand what could be the different risk classes.

**Date and Time of the Next Meeting**

The students and mentors will meet on conference call on 06/06/2024 Monday.

**Close**

Assignment for next meeting: show statement of work

* Statement of work to be refined at the next meeting.

Review ASOP 23

* ASOP 23 is on Data quality
* found in the resources folder
* Always keep the client’s original data

June 03, 2024

Recorded by

Luyang