**BUSI 650 - Final Project**

**Weightage:** 20% of the final grade

**Business Analytics Data Processing**

The goal of this project is to process and analyze real-world business data using Python and Tableau.

**Step 1: Data Loading and Analysis**

* Download the " finalproject\_dataset\_group#" dataset provided on Moodle.

**Step 2: Data Visualization and Cleaning**

* Examine the data structure and contents. Plot the data points on a graph and examine the trend over time. For your plot, consider Xlable, Ylable, and title.
* Identify and handle missing values with an appropriate technique. Present 'after' plots of the dataset to demonstrate the effectiveness of your technique. Explain how many missing values you have and describe the technique you used to handle missing values.
* Identify and describe the outliers on the cleaned dataset.
* Perform correlation analysis. Identify relevant variables and calculate their correlation coefficients. Interpret the correlation coefficients to understand the relationships between variables.

**Step 3: Regression Modeling**

* What would be the appropriate variables for regression analysis? Define the dependent and independent variables and provide your rationale. (use the results of correlation analysis)
* Export the cleaned dataset to an Excel file using the following code in colab:

df.to\_excel('/content/cleaned\_dataset.csv', index=False)

* Download your Python code in .ipynb format, as well as your cleaned dataset in a CSV file.

**Step 4: Interactive Visualizations by Tableau**

* Import the cleaned dataset into Tableau.
* Create a scatter plot of each feature in Tableau. Scatter plots typically involve two variables (x and y) to visualize the relationship between them. However, in order to create a scatter plot of each single feature in this part, you can create a calculated field by a constant. In the Data pane, right-click on “cleaned\_dataset.csv” and select “Create Calculated Field.” Name the calculated field (e.g., "Time").

Replace the formula with the following number: 1.

Now, you can create a scatter plot:

* Drag each feature to the Columns shelf.
* Drag the "Time" calculated field to the Rows shelf.
* For each feature, apply appropriate filter to remove the outliers and present 'before' and 'after' plots of the features to demonstrate the effectiveness of your technique.

**Step 5: Presentation**

* Summarize key project steps, highlighting the results and techniques in data exploration, cleaning, regression, and Tableau visualization. Provide clear and concise explanations for each step during the presentation with a total presentation time of under 3 minutes. Record your video using PowerPoint of Teams' recording feature. Ensure your face is clear and visible during the presentation.

**Submission:**

Submit:

* A PDF file containing all the solutions, explanations and figures requested. Save your file as **FirstName-LastName.PDF.**
* A Google Colab notebook containing your Python code, analysis, and visualizations (in. ipynb format).
* A video file (Mp4, PPT with recording, etc.) containing your presentation.

Good luck