**Forecasting Future Walmart Sales Using Python**

*Muhammad Salman – 100995527*

This research project was undertaken by Muhammad Salman, who is pursuing a Masters of Engineering in Electrical and Computer Engineering, with a Specialization in Data Science. He has a background in Electrical Engineering and a passion for programming and writing scripts for embedded systems. Recently, he has gained a passion for data science, which led to his pursuit for graduate studies.

**Introduction:**

Using a provided Walmart dataset from Kaggle, data engineers can predict future sales that the megastore will generate. The information in these datasets comprise of a wide range of information regarding 45 stores and their localities. These include departments per store, weekly sales, sales during weeks of major holidays, physical size of the store, CPI for the local population, etc.

This is quite significant as it would allow the store to predict which departments within certain stores would yield greater sales depending on the locality and time of year. In a realistic sense, this would allow Walmart to calculate how much and what kind of inventory they would require, leading to a growth of sales and minimizing of costs

**Methodology:**

As mentioned before, the dataset used for this study was imported from Kaggle. The dataset comprised of CSV spreadsheets which are easily accessible in Excel. All data analysis for this project past this point was done using Python3.9. Within Python, the Pandas library was used to attain quick access to CSV files and build a virtual data frame using the given the data.

The various data frames from each CSV file were first cleaned, as is done in every data science project. In simple terms, data that is N/A can be filled with 0 or any other value using Pandas’ commands. Once that was complete, the CSV files were merged into one large data frame for combined access.

The next stage exploring the data, and observing how certain features relate to others. For each comparison, a certain type of plot was generate using the Matplotlib and Seaborn libraries. An example of this was a bar graph that was generated to see the difference in sales between holiday weekly sales and standard weeks.

The final stages will include modelling the data, training the model, and verifying the model. Although this is yet to be performed, a regression model will be used, since the predictable value is a dependent variable, weekly sales, based on a series of independent variable, such as store size, holidays, locality statistics, etc.

**Conclusion:**

From performing this analysis on the given dataset, I hope to test the training model and yield passing results. As a hypothesis, we would assume a lower rate of unemployment, holiday weeks, and certain other factors would yield a greater amount of weekly sales. However, it is desirable to generate specific dollar amounts for this project.