**Project Documentation: Predictive Analysis of iPhone Purchases**

**Introduction**

This project aims to analyze a dataset containing information about individuals, including their gender, age, salary, and iPhone purchase behaviour. The goal is to build a predictive model to understand the factors influencing iPhone purchases and present the insights through Jupyter Notebook analysis.

**Jupyter Notebook Analysis**

**Data Import and Preprocessing:**

In the Jupyter Notebook, the basic necessary packages, including Pandas, NumPy, Seaborn, and Matplotlib, were imported. The dataset, consisting of 400 rows and 4 columns (Gender, Age, Salary, Purchase\_iPhone), was loaded and copied for analysis. During preprocessing, the 'Gender' column was converted to a numerical format using one-hot encoding to prepare it for the analysis.

**Exploratory Data Analysis (EDA):**

The initial data exploration involved checking for null values (none found) and understanding the correlation between variables using a heatmap. Notably, the 'Purchase\_iPhone' variable exhibited a higher correlation with age compared to other variables.

**Outlier Detection and Imputation:**

Box plots were employed to identify outliers in the distribution of iPhone purchases by age and salary. Outliers were imputed to ensure the robustness of

the analysis. The resulting box plots were visualized to observe the impact of imputation.

**Model Planning and Execution:**

The analysis involved designing a predictive model using the K-Nearest Neighbours (KNN) algorithm. The appropriate value of k was determined using an error plot, and the dataset was split into training and testing sets. The model achieved an accuracy of 88.75% on the test set. A confusion matrix was plotted to evaluate the model's performance.

**Conclusion**

In summary, the analysis revealed key insights into iPhone purchase behaviour. Age emerged as a significant factor, showcasing a correlation with purchases. Outlier handling enhanced data reliability. The K-Nearest Neighbours (KNN) model achieved an 88.75% accuracy in predicting purchases.