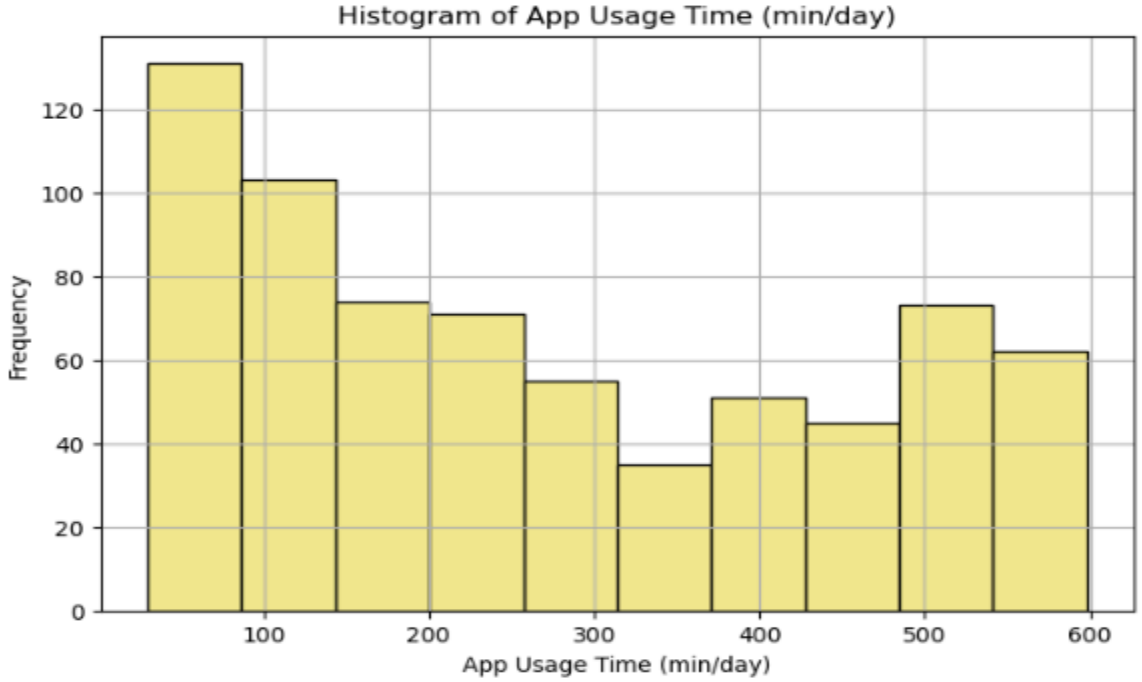
Name: Awaiz Hussain
Email: ah24adq@herts.ac.uk
UH-ID: 23100185
Assignment: Clustering and Fitting

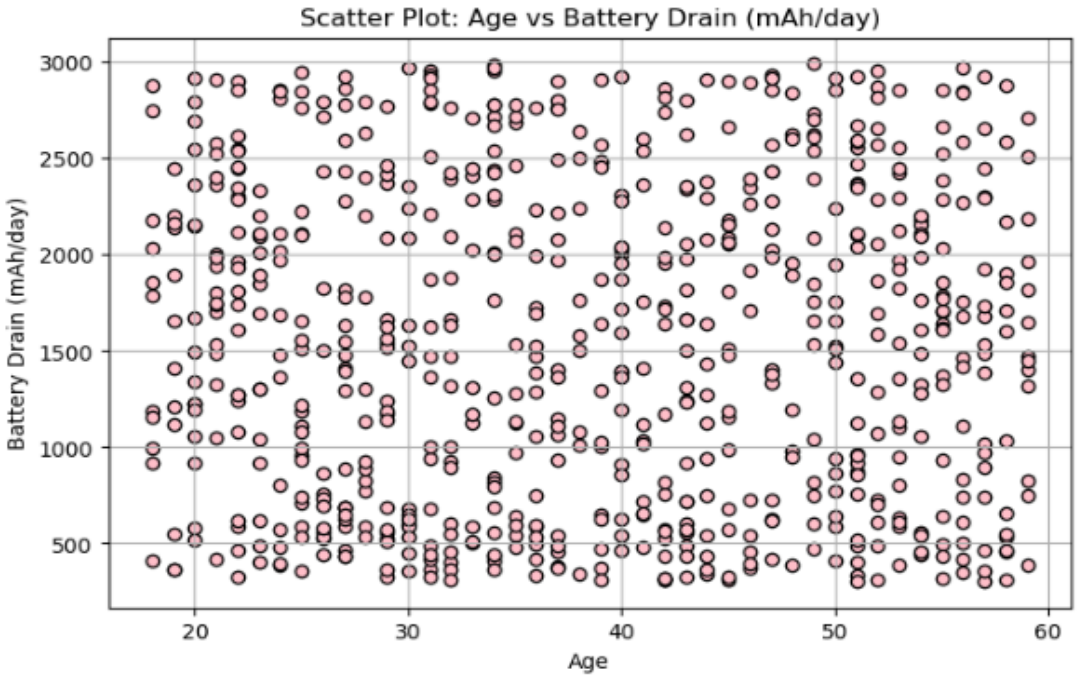
I chose the Mobile Device Usage and User Behavior Dataset from Kaggle. This dataset provides a comprehensive analysis of mobile device usage patterns and user behavior classification. It contains 700 samples of user data, including metrics such as app usage time, screen-on time, battery drain, and data consumption. Each entry is categorized into one of five user behavior classes, ranging from light to extreme usage, allowing for insightful analysis and modeling. My code and report can be found on Github.



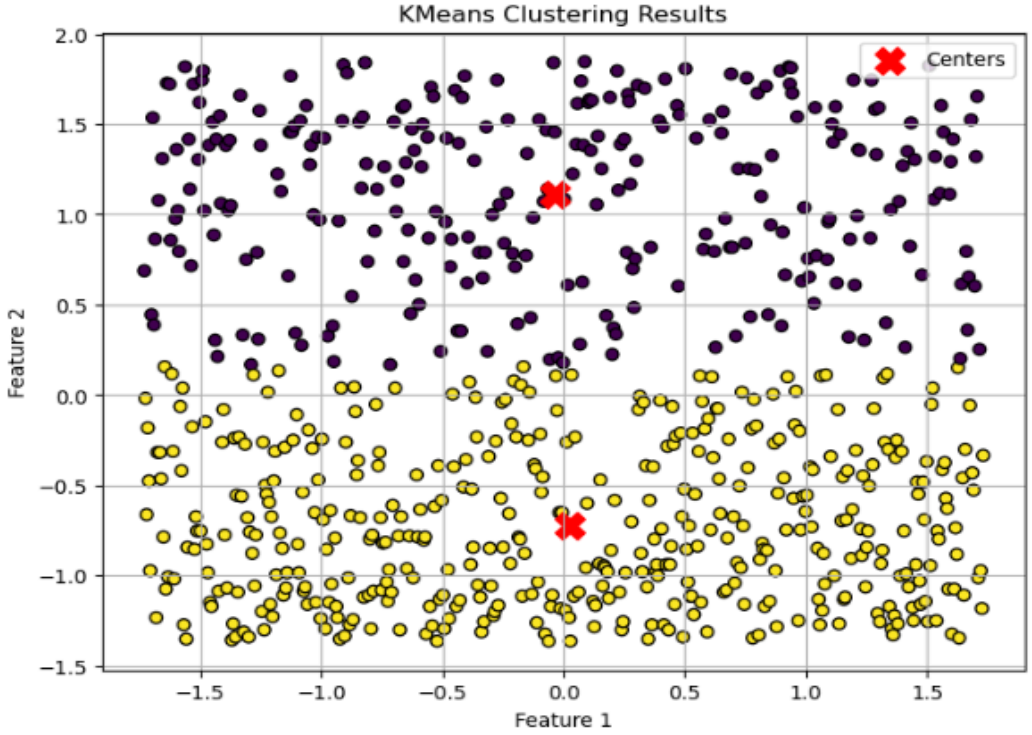
The histogram displays the distribution of App Usage Time (min/day) across the dataset. It provides insights into how app usage varies among users. The data suggests variations in app usage time, with higher values being more common. Peaks in the histogram indicate that certain users spend significantly more time on apps daily.

The scatter plot visualizes the relationship between Age and Battery Drain (mAh/day). There is no immediately strong visual correlation between Age and battery drain. Users across different ages show varying levels of battery consumption. This indicates that age alone may not be a driving factor for battery drain.

The heatmap shows the pairwise correlations among all numeric variables in the dataset. App Usage Time and Battery Drain are positively correlated (value close to 1), which suggests that



higher app usage contributes to greater battery consumption. Screen On Time also has a moderate correlation with Battery Drain. The heatmap highlights key relationships that can be useful for further analysis, such as feature selection for clustering or predictive modeling.



The Elbow method determines the optimal number of clusters (K) for K-Means clustering by examining the inertia (sum of squared distances). The plot shows a significant "elbow" at **K=2**, indicating that two clusters are appropriate for this dataset. On the basis of "Silhouette Scores: [0.45667855391520057, 0.3507617991910084, 0.34282678177737136]" for cluster 2,3,4 respectively we have decided the clusters that will be 2.

K-Means clustering groups data into distinct clusters based on similarities. The resulting scatter plot displays clear separation between clusters. Cluster centers (marked in red) are clearly visible, indicating the centroid of each group. Adding the cluster labels back to the original data provides actionable insights into user segmentation.

The line fit demonstrates the relationship between App Usage Time (min/day) and Battery Drain (mAh/day) using a linear regression model. A clear positive linear relationship exists: as app usage increases, battery drain also rises.The fit captures the trend well, suggesting that app usage is a major predictor of battery consumption.