Task 3 – Customer Segmentation / Clustering

1. Import the datasets
2. Group the transaction data by customer ID and aggregate the grouped data.
3. Merge both data frames based on the customer ID.
4. Perform preprocessing steps including converting the sign-update column to Date Time format. Extract the year and month from the signup-date column and store it in a new column.
5. Drop unnecessary columns.
6. Import the necessary algorithms. select features for clustering and convert categorical values into numeric values using the get dummies function and performing feature scaling.
7. Fit the means model to the scaled customer data, assign the cluster labels to each customer, storing them in a new column Cluster in the customer data DataFrame.
8. Calculating the Davies-Bouldin Index for the clustering. Lower values indicate better clustering, as they represent a lower average similarity ratio of each cluster with the cluster most similar to it.
9. Initializes a PCA object to reduce the data dimensions to 2 components. Fits the PCA model to the scaled customer data and transforms it into a 2-dimensional space. The resulting array customer_data_2d contains the 2 principal components for each customer.
10. **Cluster Separation**:
    The data points are color-coded according to different clusters, with distinct groups visible in the plot. This indicates that the clustering algorithm has effectively segmented the customers into distinct groups based on their characteristics.
    Cluster Distribution:
    The distribution of points within each cluster varies. Some clusters are more densely packed, indicating homogeneity within those groups, while others are more spread out, suggesting diverse customer profiles within those clusters.
    Cluster Overlap:
    There may be some overlap between clusters, where data points from different clusters are close to each other. This overlap could indicate similarities between certain customer segments or transitional behavior between clusters.
    Outliers:
    Any data points that are far from the main cluster centers can be considered outliers. These customers may exhibit unique behavior that sets them apart from the majority of the customer base.