**Name: Rania Ramadan Gad**

**Project Report: Credit Card Customer Segmentation**

1. Executive Summary

* Brief Overview: This project aimed to segment credit card customers based on their behavioral data to identify distinct customer groups. The analysis used a dataset of credit card customer transactions and balances.
* Key Finding: The analysis identified three distinct customer segments, demonstrating varying patterns of card usage, spending behavior, and credit management.
* High-Level Segment Profiles: The three segments are:
  + Engaged Transactors (Cluster 0): High-spending customers who use their cards frequently for purchases and pay their balances responsibly.
  + High Balance Revolvers (Cluster 1): Customers with high revolving balances and a reliance on cash advances.
  + Low Spenders (Cluster 2): Customers with lower spending and credit limits, who primarily use their cards for occasional purchases.
* Impact: This segmentation provides valuable insights for tailoring marketing strategies, optimizing credit risk management, and personalizing customer interactions, ultimately leading to increased profitability and customer satisfaction. The results of this analysis can be directly applied to improve business outcomes.

2. Data Exploration and Preprocessing

* Dataset Description: The dataset contains credit card customer behavioral data, including balances, purchase amounts, payment history, credit limits, and cash advance usage. It comprises 8950 customers and 17 features.
* Initial Data Quality:
  + Missing values were found in the MINIMUM\_PAYMENTS and CREDIT\_LIMIT columns. These were imputed with the median values for each column, respectively.
  + Duplicate rows were removed to ensure data integrity.
* Outlier Management:
  + Outliers were detected in several features, including BALANCE, PURCHASES, and CASH\_ADVANCE. For example, the 1.5\*IQR method identified 695 outliers in the BALANCE column, indicating a significant deviation from the typical data distribution.
  + Winsorizing was used to handle these outliers, limiting extreme values to the 1.5\*IQR range. This method was preferred over removal to retain information about potentially valuable but infrequent customer behaviors, which could represent distinct customer segments. Winsorizing minimizes the influence of outliers on the subsequent clustering process while preserving the data points.
* Feature Scaling: StandardScaler was applied to scale all features to have zero mean and unit variance. This was necessary to ensure that all features contribute equally to the distance calculations in the K-Means clustering algorithm, preventing features with larger scales from dominating the clustering process.

3. Determining Optimal Clusters

* Objective: To identify the optimal number of clusters (K) that best represent the distinct customer segments in the data.
* Methods Used:
  + Elbow Method: This method plots the Within-Cluster Sum of Squares (WCSS) against different values of K. The optimal K is typically located at the "elbow" point, where the rate of decrease in WCSS diminishes.
  + Silhouette Score: This method measures the similarity of each data point to its own cluster compared to other clusters. A higher Silhouette Score indicates better-defined clusters.
* Visualization and Interpretation:
  + Elbow Method Plot: The plot shows a decreasing WCSS as K increases. The "elbow" is observed at K=3. This suggests that adding more clusters beyond 3 provides diminishing returns in terms of reducing the overall variance within the clusters.
  + Silhouette Score Plot: The plot shows the highest Silhouette Score at K=3. A peak at K=3 indicates that this number of clusters provides the best separation between the clusters, with data points being more similar to their own cluster than to neighboring clusters.
  + Conclusion on K: Based on both the Elbow Method and the Silhouette Score, the optimal number of clusters was determined to be K=3. Both methods indicate that 3 clusters provide a good balance of minimizing within-cluster variance and maximizing cluster separation, leading to the most meaningful and interpretable segmentation of the customer base. The choice of K=3 is further supported by the business interpretability of the resulting clusters, as they align with distinct and actionable customer behaviors.

4. Customer Segmentation

* Clustering Algorithm: K-Means clustering was used to partition the data into the identified optimal number of clusters (K=3). K-Means iteratively assigns data points to the nearest cluster centroid and updates the centroids until convergence, aiming to minimize the distance between data points and their respective cluster centers.
* Clustering Result: The K-Means algorithm successfully grouped the customers into three distinct clusters, each representing a different segment of credit card users with unique behavioral patterns.

5. Cluster Profiles

* Cluster 0: Engaged Transactors (33.0%): This segment comprises customers with the highest overall purchase activity, including both one-off and installment purchases. They have the highest purchase frequency and transaction count, indicating active card usage. They also have the highest credit limits and consistently make full payments, demonstrating responsible credit behavior.
* Cluster 1: High Spenders (20.0%): This segment comprises customers with high overall purchase activity, including both one-off and installment purchases. They have high purchase frequency and transaction count, indicating active card usage. They also have high credit limits and consistently make full payments, demonstrating responsible credit behavior.
* Cluster 2: Low Spenders (47.0%): This segment comprises customers with low purchase activity. They have low purchase frequency and transaction count, indicating less active card usage. They also have low credit limits and low payments.

7. Business Insights and Recommendations

* Engaged Transactors (Cluster 0):
  + Persona: Rania (Engaged Transactor)
  + Characteristics: High credit limit, frequent purchases, pays balance in full.
  + Recommendations:
    - Offer premium rewards programs (e.g., travel points, cashback on preferred categories) to incentivize continued high spending and loyalty.
    - Proactively offer credit limit increases to accommodate their spending and reinforce their positive credit behavior.
    - Provide VIP services, such as dedicated customer support or exclusive access to events, to enhance their experience.
    - Provide personalized recommendations and offers based on their purchase history to further increase their spending.
* High Balance Revolvers (Cluster 1):
  + Persona: Mohamed (High Balance Revolver)
  + Characteristics: High credit utilization, relies on cash advances, makes minimum payments.
  + Recommendations:
    - Implement targeted financial education programs to promote responsible credit management and the risks of high-interest debt.
    - Offer balance transfer options with lower interest rates to help consolidate their debt and reduce their overall interest burden.
    - Implement stricter controls on cash advance limits and fees to discourage this high-cost behavior and mitigate risk.
    - Develop risk models to predict which customers in this segment are most likely to default and implement proactive collection strategies for those at higher risk.
* Low Spenders (Cluster 2):
  + Persona: Sarah (Low Spender)
  + Characteristics: Low credit limit, infrequent purchases, some installment usage.
  + Recommendations:
    - Develop targeted marketing campaigns to increase card usage for everyday purchases, highlighting rewards and benefits.
    - Offer promotions on installment purchases to capitalize on their existing behavior and encourage more significant purchases.
    - Provide educational materials on the benefits of using their card for a wider range of transactions and responsible credit management.
    - Consider gradual credit limit increases for customers who demonstrate consistent and responsible usage to support increased spending.

This analysis successfully identified three distinct customer segments within the credit card customer base. These segments exhibit unique behavioral patterns and financial characteristics, providing valuable insights for developing targeted business strategies. By understanding the needs and behaviors of each segment, the credit card company can improve customer satisfaction, optimize marketing effectiveness, and mitigate credit risk, leading to increased profitability and sustainable growth. The segmentation allows for a shift from a one-size-fits-all approach to a more personalized and effective customer management strategy.

* + An interactive dashboard was created using Dash and Plotly. This dashboard allows stakeholders to select a cluster and view a scatter plot of the clusters, a histogram of feature distributions for the selected cluster, and a table of descriptive statistics. This provides a dynamic way to explore the customer segments and their characteristics, enabling more informed decision-making. The dashboard enhances the usability of the analysis and facilitates communication of the findings to non-technical audiences.