# Bank Customers Segmentation

By
Mercy Jhansi Bai

# Credit Card Spending Data

- 7 numerical variables of float data type
- We don't have any missing values for the given dataset
- **spending**: Amount spent by the customer per month (in 1000s)
- **advance_payments**: Amount paid by the customer in advance by cash (in 100s)
- **probability_of_full_payment**: Probability of payment done in full by the customer to the bank
- **current_balance**: Balance amount left in the account to make purchases (in 1000s)
- **credit_limit**: Limit of the amount in credit card (10000s)
- **min_payment_amt**: minimum paid by the customer while making payments for purchases made monthly (in 100s)
- **max_spent_in_single_shopping**: Maximum amount spent in one purchase (in 1000s)

Shape: (210, 7)

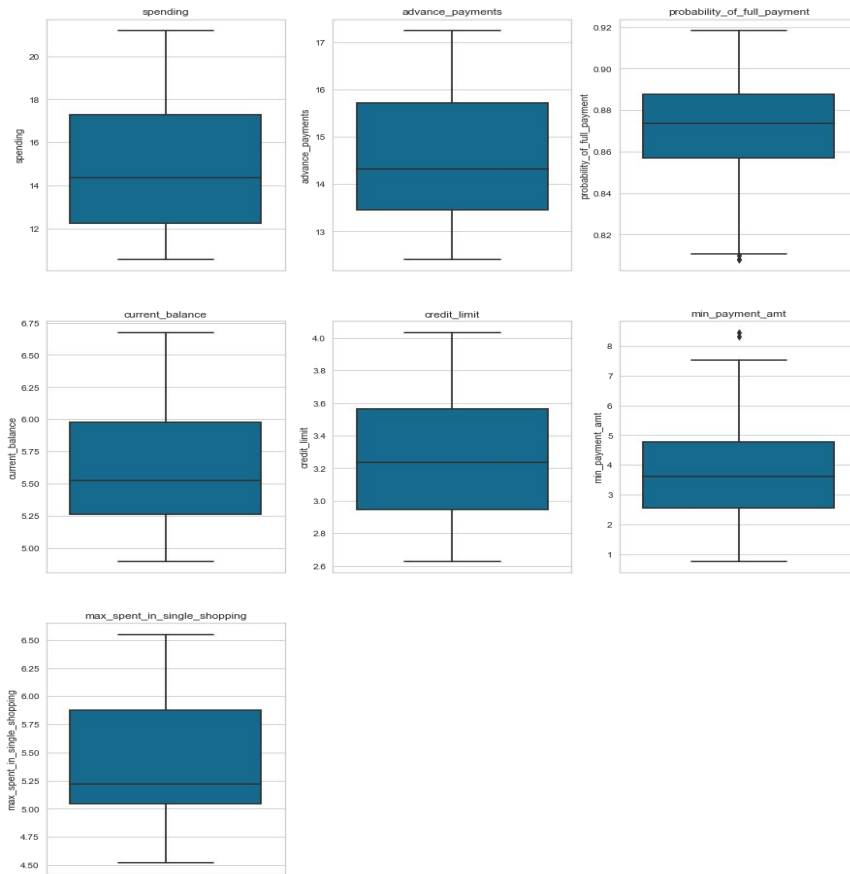| | spending | advance_payments | probability_of_full_payment | current_balance | credit_limit | min_payment_amt | max_spent_in_single_shopping |
|---|---|---|---|---|---|---|---|
| 0 | 19.94 | 16.92 | 0.8752 | 6.675 | 3.763 | 3.252 | 6.550 |
| 1 | 15.99 | 14.89 | 0.9064 | 5.363 | 3.582 | 3.336 | 5.144 |
| 2 | 18.95 | 16.42 | 0.8829 | 6.248 | 3.755 | 3.368 | 6.148 |
| 3 | 10.83 | 12.96 | 0.8099 | 5.278 | 2.641 | 5.182 | 5.185 |
| 4 | 17.99 | 15.86 | 0.8992 | 5.890 | 3.694 | 2.068 | 5.837 |

# Distribution Plots
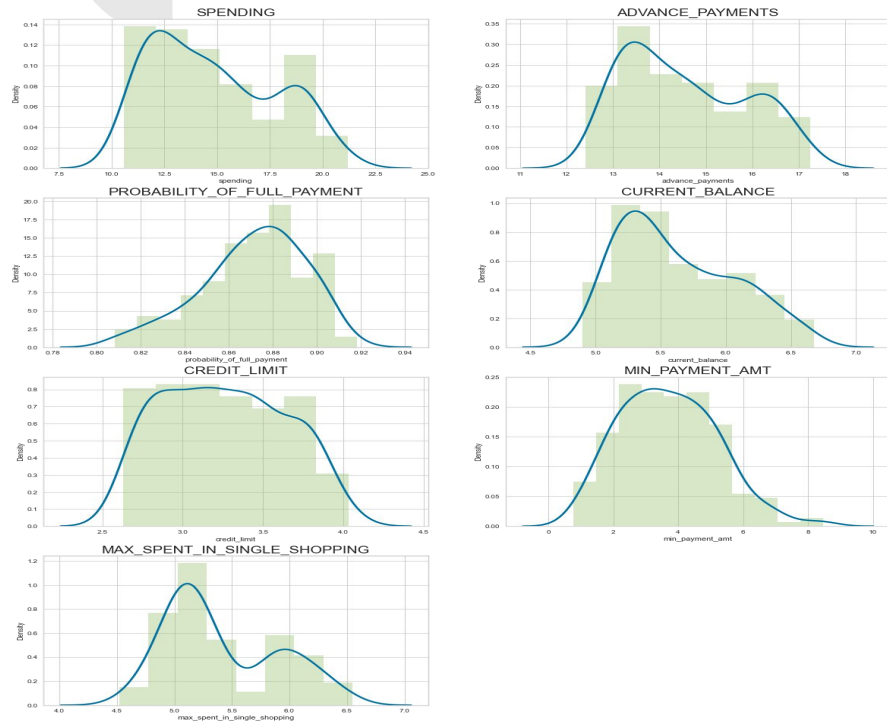


- The distribution of spending, advance payments and max spent in single shopping are bimodal.
- The distribution of credit limit is plateau-shaped.
- The current balance and min payment amount distributions are close to normal
- The probability of full payment distribution is slightly left-skewed.
- The min-payment amount is slightly right-skewed

# Box Plots



- **Probability of full payment**
    - Few lower bound outliers are present - 1.43%
- **Min Payment**
    - Has few upper bound outliers - 0.95%
- **Max_spent_in_single_shopping**
    - moderately right skewed as it's skewness is within .5 and 1 range.
- The skewness is within the -0.5 to +0.5 range for all features except max_spent_in_single_shopping, which indicates they all are approximately symmetrical.

# Heatmap



Moderate to Very Strong Correlation
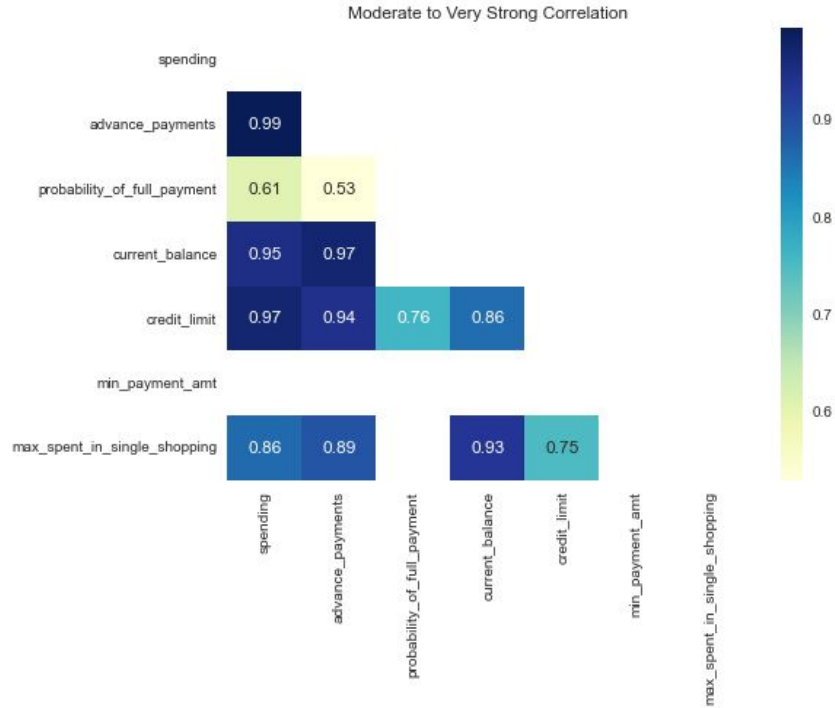
- There exists a very strong (>0.8) positive correlation between
  - spending and advance payments, current balance, credit limit, max spent in single shopping
  - current balance and advance payments, credit limit, max spent in single shopping
  - credit limit and advance payments
- There also exists a strong and moderate positive correlation (0.4 - 0.8) between certain features.
- This shows most of the features are highly correlated with each other.
- Hence we are using PCA to remove the redundant features.

# Pairplot & Data Pre-processing Steps

- Almost all the features share a linear relationship with each other except min payment amount and probability of full payment.
- Scaling is done as we are using distance based algorithms
  - Standard Scaler is used
- PCA is done to remove the redundant features
  - More than 95% variance is explained
  - We reduced 7 features to 3 features
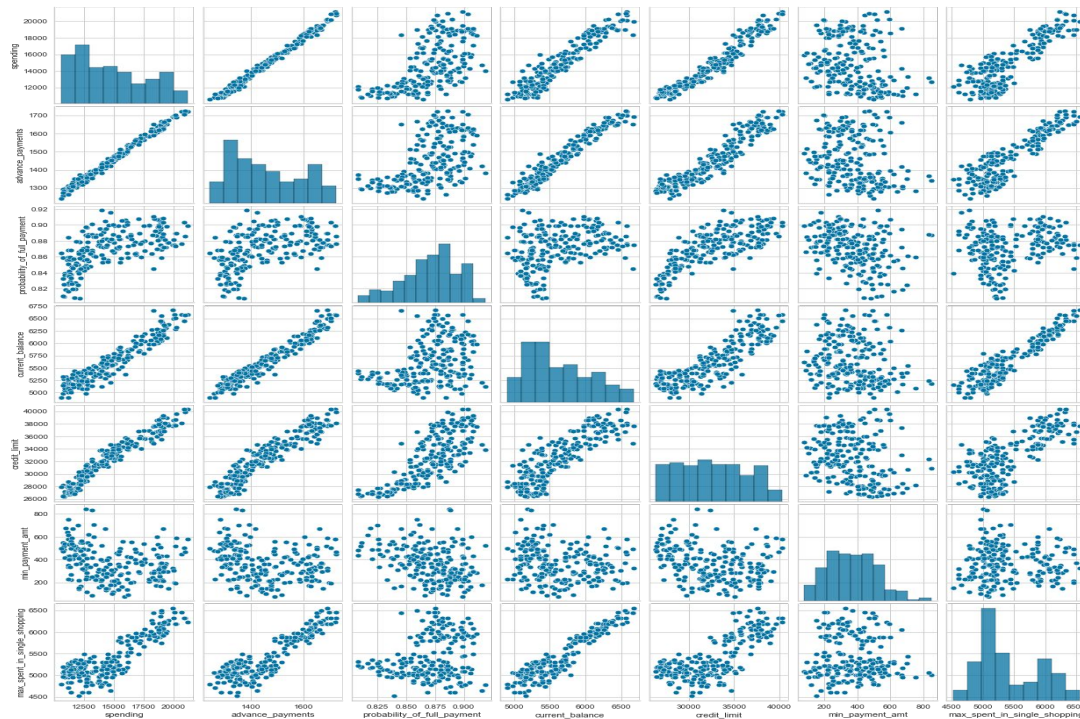
# Hierarchical Clustering - Dendrogram



Dendrogram
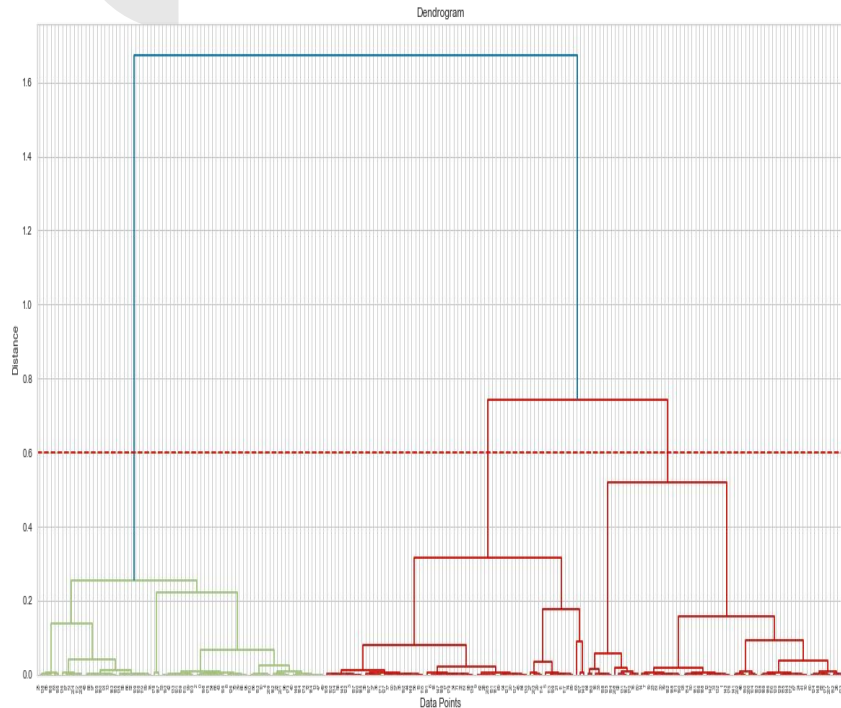
- Cophenetic Correlation coefficient is used
  - To identify the best linkage criterion among single, complete, average, ward, centroid
  - To identify the best distance metric among braycurtis, canberra, chebyshev, cityblock, correlation, cosine, dice, euclidean, hamming, mahalanobis, matching, minkowski, rogerstanimoto, russellrao, seuclidean, sokalmichener, sokalsneath, sqeuclidean, yule
- Best Params:
  - Linkage Criterion: **Average**
  - Distance Metric: **Correlation**
  - Cophenetic Correlation Value: **87.86%**
- The dendrogram suggests 3 as the optimal number of clusters

# Agglomerative Cluster Visualization

We notice 3 groups of customers.

1. The Group **Low Spending**, has low Spending but high min payment amount, the cluster is dense in the middle and sparse in the upper and lower boundaries of the cluster.
2. The Group **Medium Spending**, the Spending is medium and the min payment is low, the cluster is dense except in the upper boundaries.
3. The Group **High Spending**, the spending is high and the min payment is medium, the cluster is dense except in the upper left corner of the cluster.
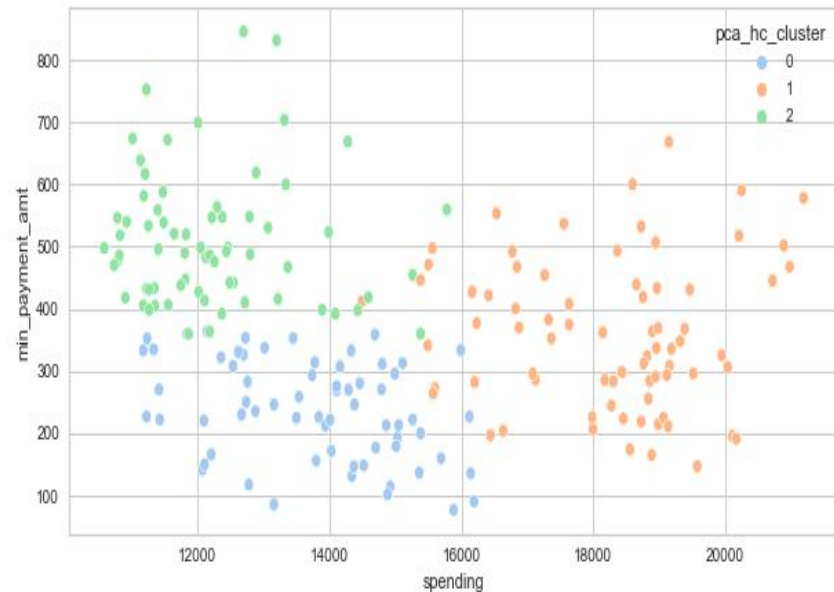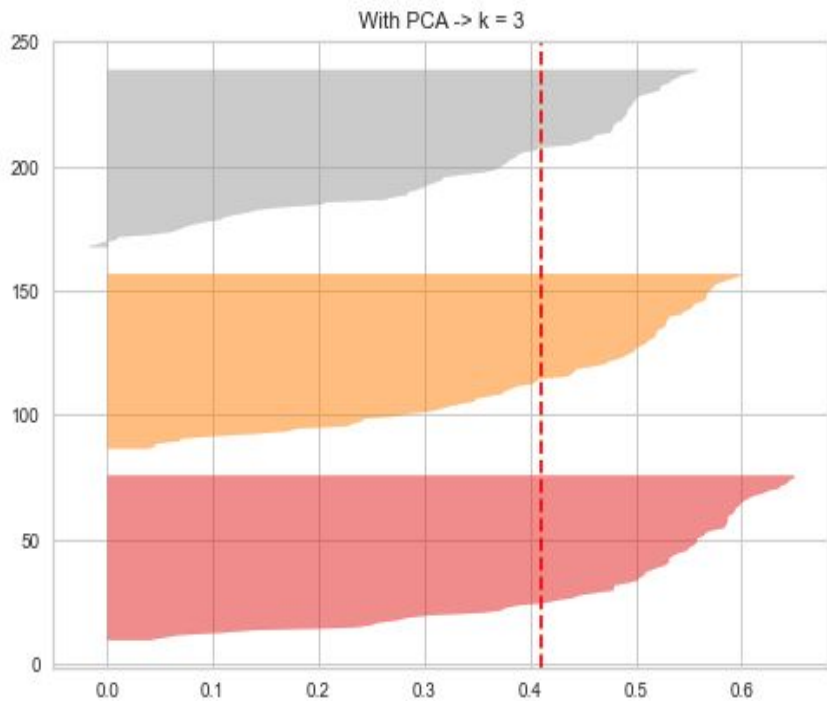
# KMeans Clustering - Elbow Plot & Cluster



Distortion Score Elbow for KMeans Clustering

--- elbow at $k = 3$, $score = 412.339$
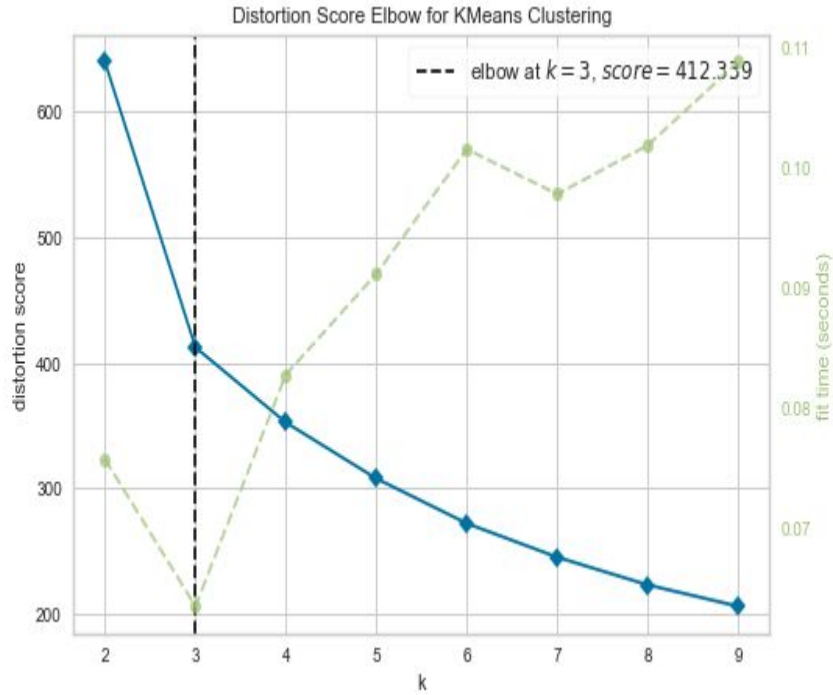
Distortion reduces slowly after 3, hence selecting the optimal value for k as 3

# Silhouette Visualizer



With PCA -> k = 3
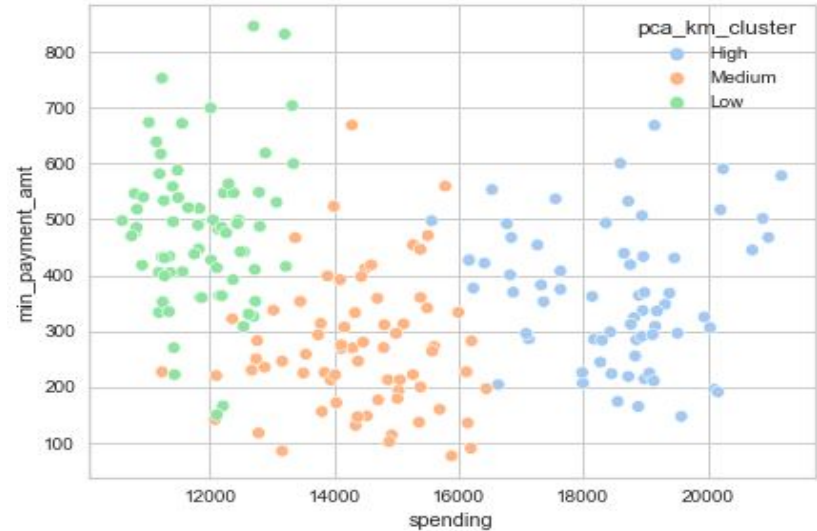
- Though the Silhouette score suggests 2 (Sil Score: 0.47) as the optimal number of clusters as that has the highest silhouette score, if we look at the Silhouette Visualizer for k=2, the cluster 2 is wider than cluster 1, which indicates more data is grouped into 1 cluster.
- Whereas in the plot of k=3, all the clusters are of equal width, which indicates all the data is grouped equally into 3 clusters. Hence, we choose **3 as the optimal number of clusters** which has a **silhouette score of 0.41**.

# Cluster Visualization – Spending Vs Cluster labels

- For cluster **High Spending**, the median line is closer to the top of the rectangle, which indicates the distribution of spending is slightly left skewed.
- For cluster **Medium Spending**, we have few lower bound outliers, which indicates the distribution of spending is slightly left skewed.
- For cluster **Low Spending**, the data is almost equally dispersed around the median, which indicates the distribution is almost symmetric.
- Median could be best measure over mean, to determine the cluster characteristics, as we notice outliers in the clusters in the boxplot visualization

# Cluster Profiling

| n_cluster | spending | advance_payments | probability_of_full_payment | current_balance | credit_limit | min_payment_amt | max_spent_in_single_shopping |
|-----------|----------|------------------|------------------------------|-----------------|--------------|-----------------|------------------------------|
| High | 18750.0 | 1623.0 | 1.0 | 6153.0 | 37190.0 | 362.0 | 6009.0 |
| Low | 11830.0 | 1327.0 | 1.0 | 5226.0 | 28360.0 | 482.0 | 5089.0 |
| Medium | 14405.0 | 1438.0 | 1.0 | 5534.0 | 32515.0 | 261.0 | 5122.0 |

We can categorize the clusters into

1. Low Spending and High Min Payment
2. Medium Spending and Low Min Payment
3. High spending and Medium Min Payment

## Low Spending and High Min Payment Cluster Characteristics

The cluster has low Spending but high min payment amount, which indicates though they spend less, they convert most of their purchases to EMI. This may indicate lower middle-class category of customers. The average spending per month of this cluster is as low as Rs 11,830. The credit limit, advance payment and max spent in single shopping is also low for this cluster. 34% of the customers belong to this cluster.

### Promotional Strategy

- We can verify their recent payslip, and check if we can offer them an increase in their credit limit and cash limit, thereby helping them make more purchases or withdraw more money from their credit card.
- The higher min payment amount indicates they are availing more EMI options, hence we can slash down the interest rates to help them opt for more EMI options which thereby also could increase their max spent in single shopping.
- We can lower the cash advance fee payment, thereby helping them make more cash withdrawals from their credit card.

# Medium Spending and Low Min Payment Cluster Characteristics

## Promotional Strategy

- Can be offered customized credit cards with customized discounts based on the kind of purchases they make the most, to help them save money and thereby increase their frequency of purchases.
- We can verify their recent payslip, and check if we can offer them an increase in their credit limit and cash limit, thereby helping them make more purchases or withdraw more money from their credit card.
- We can provide EMI options for purchases above Rs 3000 with lower interest rates thereby we can increase their max spent in single shopping.
- We can lower the cash advance fee payment, thereby helping them make more cash withdrawals from their credit card.

| _cluster | spending | advance_payments | probability_of_full_payment | current_balance | credit_limit | min_payment_amt | max_spent_in_single_shopping |
|---|---|---|---|---|---|---|---|
| High | 18750.0 | 1623.0 | 1.0 | 6153.0 | 37190.0 | 362.0 | 6009.0 |
| Low | 11830.0 | 1327.0 | 1.0 | 5226.0 | 28360.0 | 482.0 | 5089.0 |
| Medium | 14405.0 | 1438.0 | 1.0 | 5534.0 | 32515.0 | 261.0 | 5122.0 |

The Spending is medium and the min payment is low, which indicates though they spend enough, they don't convert much of their transactions to EMI. This may indicate middle-class group of customers. The average spending per month of this cluster is medium of Rs 14,405. The credit limit and advance payment is also medium for this cluster. The max spent in single shopping is low for this cluster. 34% of the customers belong to this cluster.

# High Spending and Medium Min Payment Cluster Characteristics

| _cluster | spending | advance_payments | probability_of_full_payment | current_balance | credit_limit | min_payment_amt | max_spent_in_single_shopping |
|---|---|---|---|---|---|---|---|
| High | 18750.0 | 1623.0 | 1.0 | 6153.0 | 37190.0 | 362.0 | 6009.0 |
| Low | 11830.0 | 1327.0 | 1.0 | 5226.0 | 28360.0 | 482.0 | 5089.0 |
| Medium | 14405.0 | 1438.0 | 1.0 | 5534.0 | 32515.0 | 261.0 | 5122.0 |

The spending is high and the min payment is medium, which indicate they are spending high and converting some of their transactions to EMI. They may be the elite and high-class group of customers. The average spending per month of this cluster is as high as Rs 18,750. The current balance, credit limit and max spent in single shopping is high for this cluster compared to the other 2 clusters. 32% of the customers belong to this cluster.

## Promotional Strategy

- This group is performing best among all customers, hence giving rewards points for the purchases and providing attractive shopping discount vouchers for the reward points will make them do more and better purchases.
- Can be given customized credit cards with customized discounts based on the kind of purchases they make the most, to help them save money and thereby increase their frequency of purchases.

# What better can be done?

- We can collect more data such as where they are spending more, like in restaurants, or textile shopping, or grocery shopping or online purchases, or in education or if they spending much in tours and travels and offer them customized credit cards options with discounts on the respective category of spending.
- We can capture their credit score, and offer loan facilities for those with good credit score.
- We can capture their monthly average credit limit utilization, and their debit card usage and suggest them ways to enhance their credit card usage, thereby earning reward points.