Comprehensive Banking Analytics

Presented by

Nandhagopal S

Problem Definition:

- To enhance the Decision Making in loan approval process, this project is aims to address the customer segmentation and Credit Risk Assessment.

Solution:

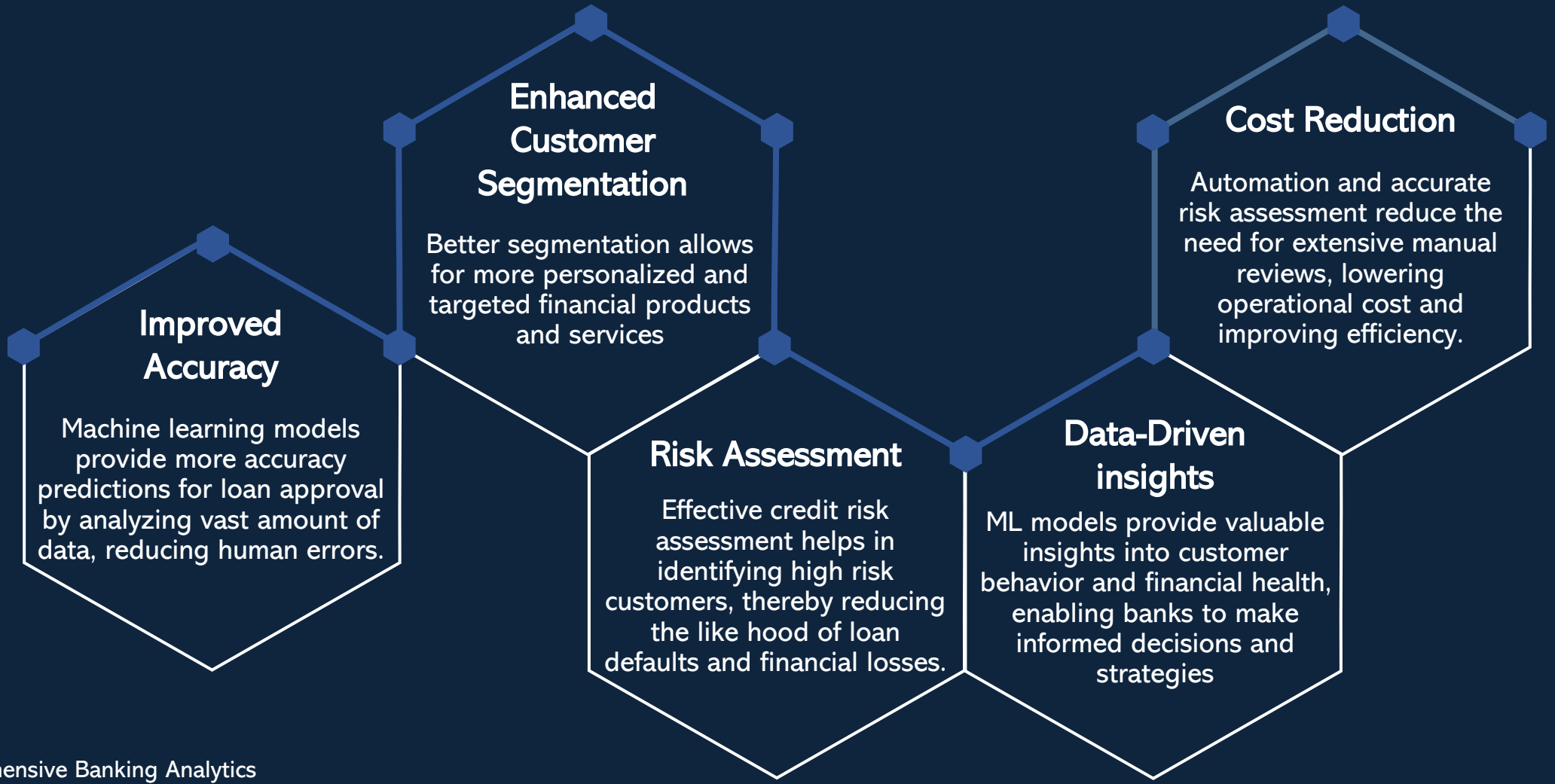
1. Customer Segmentation

   - Grouping Customers based on banking behaviours, Transaction histories, and demographics to understand their financial profiles better

2. Credit Risk Assessment

   - Developing a robust credit scoring system that evaluates loan approval based on the historical data, clustering analysis, and predictive modelling for creditworthiness and late payment risks.

Benefits



Comprehensive Banking Analytics

Data Collection: Feature details of Given dataset (train.csv)

01. ID	02. Customer ID	03. Month	04. Name	05.Age
ID number for customer's	Customer ID for each customer's	Indicating the month of transaction	Name of the Customer's	Age of the Customer's

06. SSN	07. Occupation	08. Annual Income	09. Monthly in hand salary	10. No. of Bank Accounts
SSN number of customer's	Customer's occupation detail	Customer's Annual Salary income information	Customer's Monthly in hand salary information	Total no. of banks owned by customer

11.Num_Credit_Card	12. Interest_Rate	13. Num_of_Loan	14. Type of loan	15. Delay_from_due_date
Total no. of credit card owned by customer	Interest rate of loan	Total no. of loan with customer	Type of loan customer purchased	No. of delay payment from due date

16. Num_of_Delayed_Payment	17. Changed_Credit_Limit	18. Num_Credit_Inquiries	19. Credit_mix	20. Outstanding_Debt
No. of delay payment done by customer	The difference between the previous and new credit limits.	Credit inquiries from customer	Credit score based on the variety of credit accounts that a customer has.	Outstanding_Debt of a credit accounts

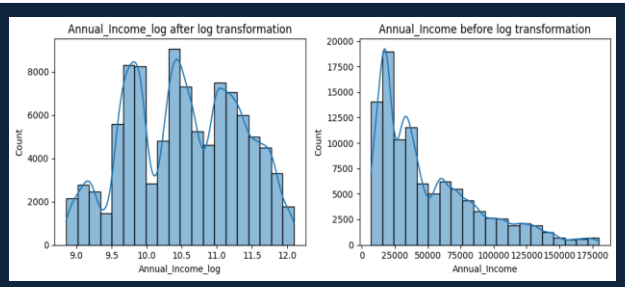
21. Credit_Utilization_Ratio	22. Credit_History_Age	23. Payment_of_Min_Amount	24. Total_EMI_per_month	25. Amount_invested_monthly
Credit_Utilization_Ratio of credit accounts	Credit history age of customer	Labeled feature if customer likely have min amount for credit account	Total amount of EMI need to pay for the customer	Invested amount by customer

26. Payment_Behaviour	27. Monthly_Balance	28. Credit_Score
Labeled feature of customer payment behavior's	Monthly balance amount after all the debit by customer	Labeled feature of Credit Behavior

Comprehensive Banking Analytics

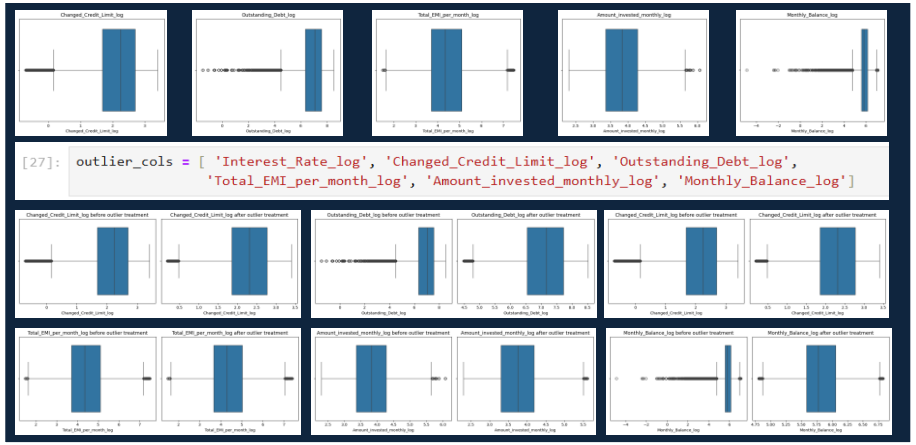
EDA : Exploratory Data Analysis

Skewed Data after log transformation Comparison



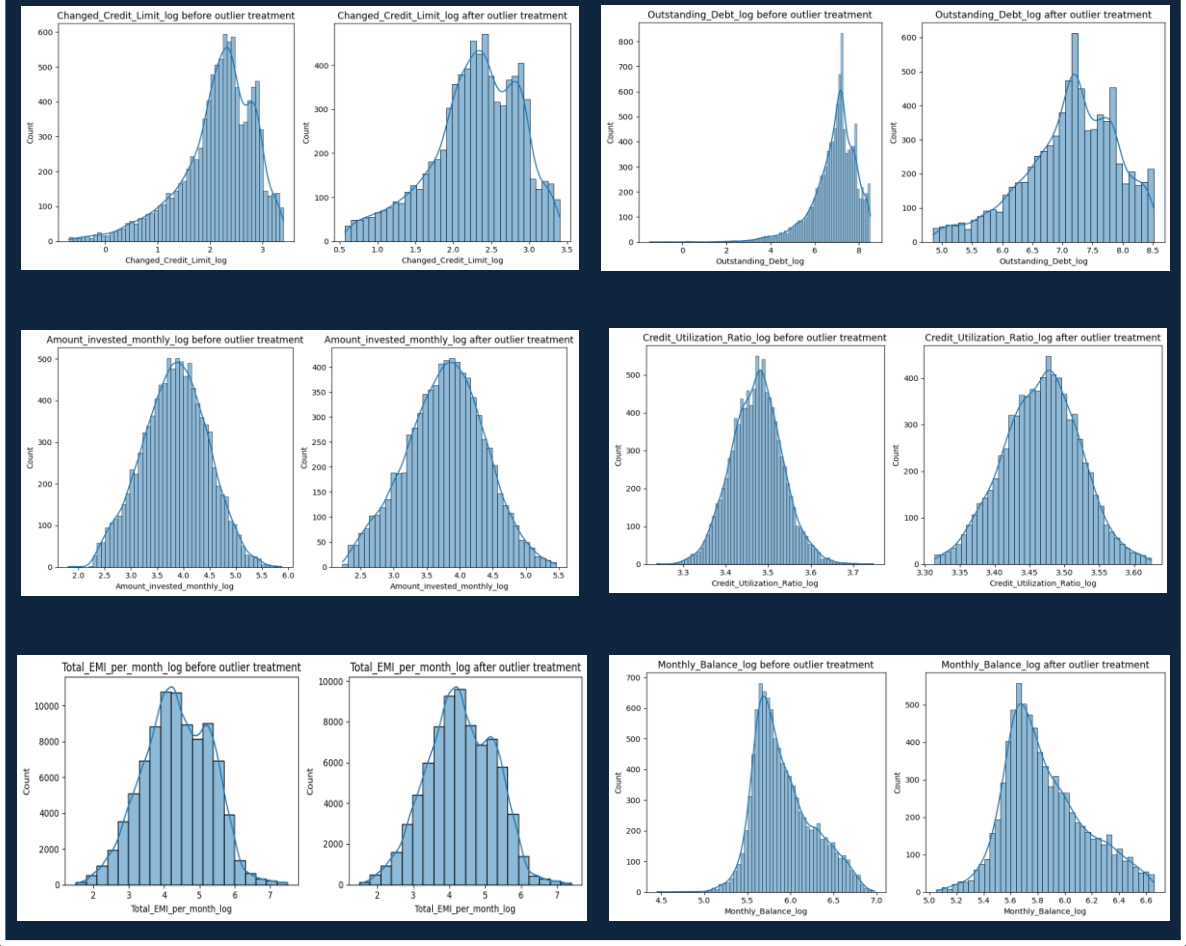
Our Data is Positively skewed (right Skewed) and handled by log transformation to compress the data and brings distribution closer to normal. This makes ml algorithm that assume a normal distribution.

Outliers and treatment (Using IQR Method)



```
[27]: outlier_cols = [ 'Interest_Rate_log', 'Changed_Credit_Limit_log', 'Outstanding_Debt_log',
                       'Total_EMI_per_month_log', 'Amount_invested_monthly_log', 'Monthly_Balance_log']
```


Comparison of Distribution after outlier treatment Using IQR Method



Summary

Method Used:
Interquartile Range (IQR) Method

IQR = Q3 – Q1

- **Lower Bound:**
  Q1 - 1.5 * IQR
- **Upper Bound:**
  Q3 + 1.5 * IQR

Action Taken:

Outliers beyond the lower and upper bounds were identified and removed.

Outcome:

After handling outliers, the data is cleaned and ready for the next step in the machine learning process.

Model Training – Clustering(Unsupervised learning)

Feature-1

```
cluster_ss_score(df8_scaled[features_1],k)
```

	SS_Cluster_2	SS_Cluster_3	SS_Cluster_4	SS_Cluster_5
0	0.216033	0.178831	0.161741	0.153662

Feature-2

```
cluster_ss_score(df8_scaled[features_2],k)
```

	SS_Cluster_2	SS_Cluster_3	SS_Cluster_4	SS_Cluster_5
0	0.204655	0.184658	0.165221	0.157254

Feature-3

```
cluster_ss_score(df8_scaled[features_3],k)
```

	SS_Cluster_2	SS_Cluster_3	SS_Cluster_4	SS_Cluster_5
0	0.221306	0.183258	0.16509	0.159165

Feature-4

```
cluster_ss_score(df8_scaled[features_4],k)
```

	SS_Cluster_2	SS_Cluster_3	SS_Cluster_4	SS_Cluster_5
0	0.223161	0.200094	0.185516	0.172263

Feature-5

```
cluster_ss_score(df8_scaled[features_5],k)
```

	SS_Cluster_2	SS_Cluster_3	SS_Cluster_4	SS_Cluster_5
0	0.217391	0.21631	0.208808	0.18785

Feature-6

```
cluster_ss_score(df8_scaled[features_6],k)
```

	SS_Cluster_2	SS_Cluster_3	SS_Cluster_4	SS_Cluster_5
0	0.234503	0.232507	0.228636	0.209537

Feature-7

```
cluster_ss_score(df8_scaled[features_7],k)
```

	SS_Cluster_2	SS_Cluster_3	SS_Cluster_4	SS_Cluster_5
0	0.284632	0.276759	0.284767	0.256027

Feature-8

```
cluster_ss_score(df8_scaled[features_8],k)
```

	SS_Cluster_2	SS_Cluster_3	SS_Cluster_4	SS_Cluster_5
0	0.294805	0.262094	0.251958	0.269809

Feature-9

```
cluster_ss_score(df8_scaled[features_9],k)
```

	SS_Cluster_2	SS_Cluster_3	SS_Cluster_4	SS_Cluster_5
0	0.345421	0.265471	0.253078	0.26176

Feature-10

```
cluster_ss_score(df8_scaled[features_10],k)
```

	SS_Cluster_2	SS_Cluster_3	SS_Cluster_4	SS_Cluster_5
0	0.440031	0.390063	0.428756	0.455061

Feature-11

```
cluster_ss_score(df8_scaled[features_11],k)
```

	SS_Cluster_2	SS_Cluster_3	SS_Cluster_4	SS_Cluster_5
0	0.346433	0.340537	0.360176	0.339808

Feature-12

```
cluster_ss_score(df8_scaled[features_12],k)
```

	SS_Cluster_2	SS_Cluster_3	SS_Cluster_4	SS_Cluster_5
0	0.387328	0.373105	0.399633	0.411922

Feature-13

```
cluster_ss_score(df8_scaled[features_13],k)
```

	SS_Cluster_2	SS_Cluster_3	SS_Cluster_4	SS_Cluster_5
0	0.387153	0.435742	0.495394	0.549055

Insights

- Among the 13 feature sets feature_13 is performing well with 5 clusters
- Based on the silhouette score feature_13 with n_clusters = 5 having the silhouette score of 0.549055 within the range of 0.50 to 0.75 which means Good Clustering

Model Selection & Training (Classification):

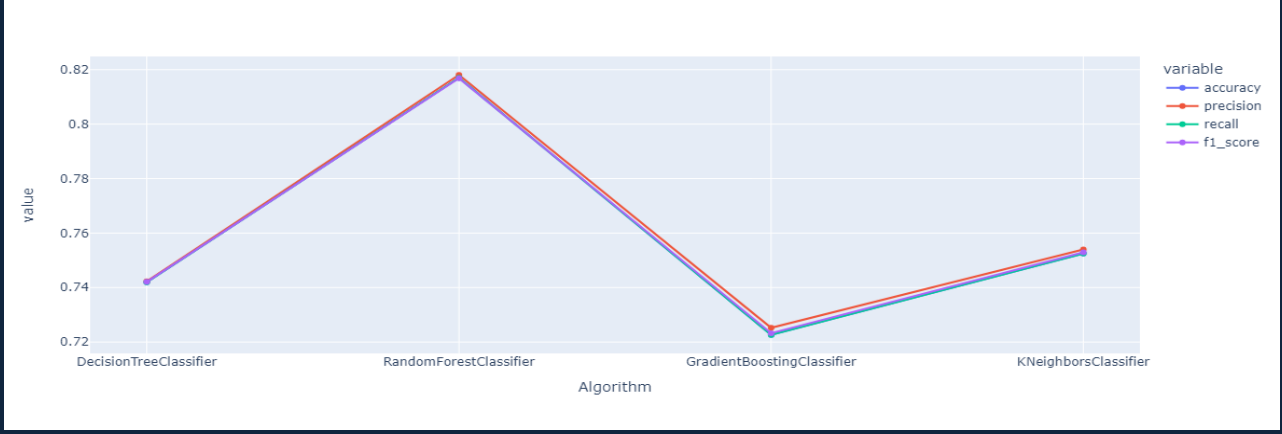
Our Model is Multi class classification(Supervised learning) to find the credit score of customer based on banking transaction behavior. Algorithms for our problem is

1. Decision Tree Classifier
2. Random Forest Classifier
3. KNN Classifier
4. Gradient Booster Classifier (GBM)

Algorithm Metrics for our trained model

	Algorithm	accuracy	precision	recall	f1_score
0	DecisionTreeClassifier	0.741948	0.742277	0.741948	0.742097
1	RandomForestClassifier	0.817012	0.818069	0.817012	0.816845
2	GradientBoostingClassifier	0.722611	0.725225	0.722611	0.723228
3	KNeighborsClassifier	0.752466	0.753998	0.752466	0.752914

Visualization of Algorithm performance metrics



Insights

- **Best Overall Model:** The Random Forest Classifier stands out as the best performing model with the highest accuracy (0.8170), precision (0.8181), recall (0.8170), and F1 score (0.8168), indicating that it is the best performing model for this dataset. The consistency across these metrics suggests it is a reliable and balanced model.
- **Moderate Performers:** The Decision Tree and K-Nearest Neighbors classifiers exhibit moderate performance.
- **Lower Performance:** The Gradient Boosting Classifier has the lowest performance metrics in this comparison.

Hyperparameter Tunning: Grid Search CV (To avoid under fitting and over fitting the model we will find the best parameters for our model)

```
" Best parameters for RandomForestClassifier: {'bootstrap': True, 'max_features': 'log2', 'n_estimators': 100}"
```

Customer Segmentation-Cluster-1-Insights

Descriptive and Perspective Analysis

- Low Income: Poor Credit Score: Customers in Cluster 1 have a low mean annual income of approximately 16,021.36, with a minimum of 7,064.39 and a maximum of 28,062.39. This indicates that they are in a lower income bracket. Additionally, all customers in this cluster have been categorized as having a poor credit score.
- Financial Vulnerability: The combination of low income and poor credit score suggests that customers in Cluster 1 may be financially vulnerable. They may have limited access to credit or financial resources, making them more sensitive to economic fluctuations or unexpected expenses.
- Risk Assessment: From a risk assessment perspective, customers in Cluster 1 may present a higher risk for default or late payments. We need to carefully evaluate their creditworthiness and may consider offering lower credit limits or higher interest rates to mitigate potential risks.
- Targeted Financial Products: Given their financial situation, customers in Cluster 1 may benefit from tailored financial products and services that cater to individuals with low income or poor credit history. This could include microloans, financial literacy programs, or credit-building products.
- Support and Assistance: Providing financial education and support to customers in Cluster 1 could help improve their financial well-being. This could include budgeting tips, debt management strategies, or access to financial counseling services.
- Overall, Cluster 1 represents a segment of customers who are financially vulnerable and may require specialized financial products and support to improve their financial stability and creditworthiness.

Cluster-1

Cluster_1

```
df_cluster_1['Credit_Score'].value_counts()
```

```
Credit_Score
Poor    1191
Name: count, dtype: int64
```

```
df_cluster_1[['Annual_Income']].describe().loc[['min','max','mean']].T
```

	min	max	mean
Annual_Income	7064.387207	28062.392578	16021.358398

Customer Segmentation-Cluster-2-Insights

Descriptive and Perspective Analysis

- **Standard Credit Score:** All customers in Cluster 2 have been categorized as having a standard credit score. This indicates that they are likely to have a relatively good credit history and creditworthiness compared to customers with poor or no credit history.
- **Higher Annual Income:** Customers in Cluster 2 have a higher mean annual income of approximately 71,398.48 , with a minimum of 36,707.71 and a maximum of 177,330.19. This suggests that they are in a higher income bracket compared to customers in Cluster 1.
- **Financial Stability:** The combination of a standard credit score and higher annual income indicates that customers in Cluster 2 are likely to be more financially stable. They may have better access to credit and financial resources, making them less sensitive to economic fluctuations.
- **Lower Risk Profile:** From a risk assessment perspective, customers in Cluster 2 may present a lower risk for default or late payments compared to customers in Cluster 1. We may be more inclined to offer them higher credit limits or lower interest rates.
- **Targeted Financial Products:** Customers in Cluster 2 may benefit from a wider range of financial products and services, including higher credit limits, lower interest rates, and premium banking services. Lending institutions may also offer them investment products or wealth management services tailored to their higher income bracket.
- **Financial Planning:** Given their higher income and standard credit score, customers in Cluster 2 may be interested in financial planning services to help them achieve their long-term financial goals. This could include retirement planning, tax planning, and wealth preservation strategies.
- **Overall,** Cluster 2 represents a segment of customers who are more financially stable and have a lower risk profile compared to customers in Cluster 1. They may require different financial products and services that cater to their higher income and creditworthiness.

Cluster-2

Cluster_2

```
df_cluster_2['Credit_Score'].value_counts()
```

```
Credit_Score
Standard    1944
Name: count, dtype: int64
```

```
df_cluster_2[['Annual_Income']].describe().loc[['min','max','mean']].T
```

	min	max	mean
Annual_Income	36707.714844	177330.1875	71398.476562

Customer Segmentation-Cluster-3-Insights

Descriptive and Perspective Analysis

- **Poor Credit Score:** All customers in Cluster 3 have been categorized as having a poor credit score. This suggests that they may have a history of late payments, defaults, or other credit issues that have impacted their creditworthiness.
- **Moderate Annual Income:** Customers in Cluster 3 have a moderate mean annual income of approximately 55,641.20 , with a minimum of 28,105.59 and a maximum of 179,948.78. This indicates that they are in a middle-income bracket.
- **Financial Challenges:** The combination of a poor credit score and moderate income suggests that customers in Cluster 3 may face financial challenges. They may have limited access to credit and financial resources, making it difficult for them to meet their financial obligations.
- **Higher Risk Profile:** From a risk assessment perspective, customers in Cluster 3 may present a higher risk for default or late payments compared to customers in Clusters 1 and 2. We may need to carefully evaluate their creditworthiness and may consider offering them lower credit limits or higher interest rates to mitigate potential risks.
- **Financial Assistance:** Customers in Cluster 3 may benefit from financial assistance programs aimed at improving their creditworthiness and financial stability. This could include credit counseling, debt management plans, or financial education programs.
- **Targeted Financial Products:** Given their financial situation, customers in Cluster 3 may require specialized financial products and services that cater to individuals with poor credit history and moderate income. This could include credit repair services, alternative lending options, or budgeting tools.
- Overall, Cluster 3 represents a segment of customers who are facing financial challenges and may require targeted financial products and support to improve their financial stability and creditworthiness.

Cluster-3

Cluster_3

```
df_cluster_3['Credit_Score'].value_counts()
```

```
Credit_Score
Poor    1576
Name: count, dtype: int64
```

```
df_cluster_3[['Annual_Income']].describe().loc[['min','max','mean']].T
```

	min	max	mean
Annual_Income	28105.59375	179948.78125	55641.199219

Customer Segmentation-Cluster-4-Insights

Descriptive and Perspective Analysis

- **Mixed Credit Scores:** Customers in Cluster 4 have a mix of credit scores. The majority have a standard credit score, with 1,781 customers falling into this category. Additionally, there are 186 customers with a good credit score in this cluster.
- **Moderate Annual Income:** Customers in Cluster 4 have a moderate mean annual income of approximately 21,814.63, with a minimum of 7,023.16 and a maximum of 36,690.32. This suggests that they are in a middle to lower-middle income bracket.
- **Diverse Financial Profiles:** The mix of credit scores in Cluster 4 indicates that customers in this cluster have diverse financial profiles. Some may have a strong credit history and creditworthiness (good credit score), while others may have a more average credit profile (standard credit score).
- **Medium Risk Profile:** From a risk assessment perspective, customers in Cluster 4 may present a medium risk for default or late payments. While the majority have a standard credit score, which indicates a relatively good credit history, there are still customers with a good credit score who may present a lower risk.
- **Financial Stability:** The moderate annual income of customers in Cluster 4 suggests that they are relatively stable financially. They may have decent access to credit and financial resources, making them less sensitive to economic fluctuations compared to customers in Cluster 1 and 3.
- **Targeted Financial Products:** Customers in Cluster 4 may benefit from a range of financial products and services tailored to their diverse financial profiles. This could include credit options for those with standard credit scores and investment or savings products for those with good credit scores.
- Overall, Cluster 4 represents a segment of customers with diverse financial profiles, moderate risk, and moderate financial stability. They may require a mix of financial products and services to meet their needs and manage their financial well-being

Cluster-4

Cluster_4

```
df_cluster_4['Credit_Score'].value_counts()
```

```
Credit_Score
Standard    1781
Good         186
Name: count, dtype: int64
```

```
df_cluster_4[['Annual_Income']].describe().loc[['min','max','mean']].T
```

	min	max	mean
Annual_Income	7023.162598	36690.316406	21814.634766

Customer Segmentation-Cluster-5-Insights

Descriptive and Perspective Analysis

- **Good Credit Score:** All customers in Cluster 5 have been categorized as having a good credit score. This indicates that they have a strong credit history and creditworthiness.
- **Higher Annual Income:** Customers in Cluster 5 have a higher mean annual income of approximately 69,599.71, with a minimum of 23,291.78 and a maximum of 179,072.28. This suggests that they are in a higher income bracket.
- **Financial Stability:** The combination of a good credit score and higher annual income indicates that customers in Cluster 5 are likely to be more financially stable. They may have better access to credit and financial resources, making them less sensitive to economic fluctuations.
- **Lower Risk Profile:** From a risk assessment perspective, customers in Cluster 5 may present a lower risk for default or late payments compared to customers in Clusters 1, 3, and 4. Lending institutions may be more inclined to offer them higher credit limits or lower interest rates.
- **Targeted Financial Products:** Customers in Cluster 5 may benefit from a range of financial products and services tailored to their strong credit profile and higher income. This could include premium banking services, investment products, or wealth management services.
- **Financial Planning:** Given their higher income and good credit score, customers in Cluster 5 may be interested in financial planning services to help them achieve their long-term financial goals. This could include retirement planning, tax planning, and wealth preservation strategies.
- Overall, Cluster 5 represents a segment of customers who are financially stable, have a strong credit profile, and are in a higher income bracket. They may require different financial products and services that cater to their higher income and creditworthiness compared to other clusters.

Cluster-5

Cluster_5

```
df_cluster_5['Credit_Score'].value_counts()
```

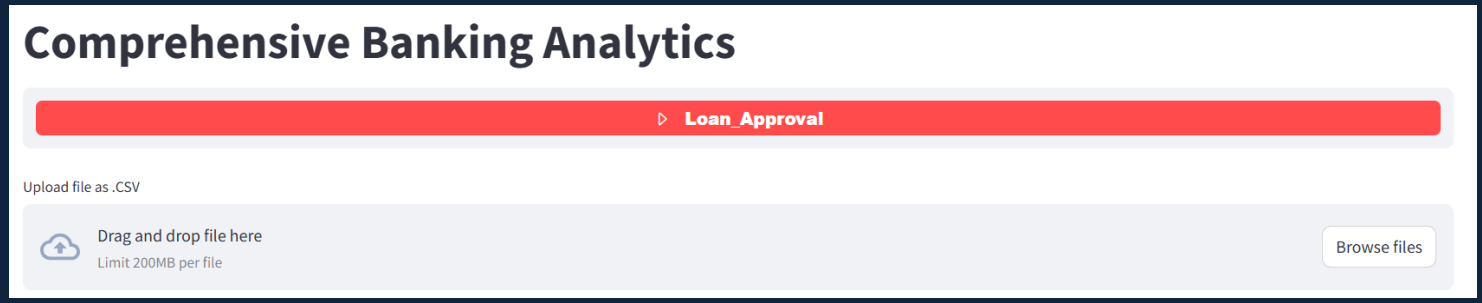
```
Credit_Score
Good    923
Name: count, dtype: int64
```

```
df_cluster_5[['Annual_Income']].describe().loc[['min','max','mean']].T
```

	min	max	mean
Annual_Income	23291.779297	179072.28125	69599.710938

Robust Credit Scoring and Loan Approval System:

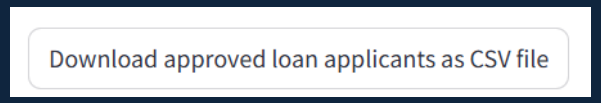
Robust Credit Scoring System



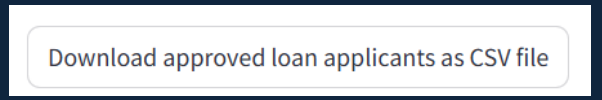
- By uploading the CSV file to predict credit score and this system enables to approve or reject the loan based on the risk assessment and credit score prediction.

Model Prediction

	Custome	Name	Age	Occupation	Annual_Income	Monthly_Inhand_Salary	Total_EMI_per_month	Num_of_D	Interest_Rate	Credit_Score_Predicted	Risk_Assessment	Loan_Status
0	16,640	Reema	31	Entrepreneur	19,300.342	1,512.3618	49.5721	18	17	Standard	high_risk	Rejected
1	41,326	Lee Chyenz	41	Media_Manager	10,183.016	1,074.5848	37.5881	21	17	Standard	high_risk	Rejected
2	24,648	Ayeshaz	43	Musician	18,627.648	1,387.3031	65.1395	20	19	Standard	high_risk	Rejected
3	17,068	Lawrencea	37	Musician	15,566.018	1,423.1686	43.0705	20	32	Poor	high_risk	Rejected
4	24,733	Shupingu	27	Architect	14,165.236	1,057.4357	58.8684	24	33	Poor	high_risk	Rejected



- **Approved:** We can download the approved loan applicants list by clicking "Download approved loan applicants as CSV file".



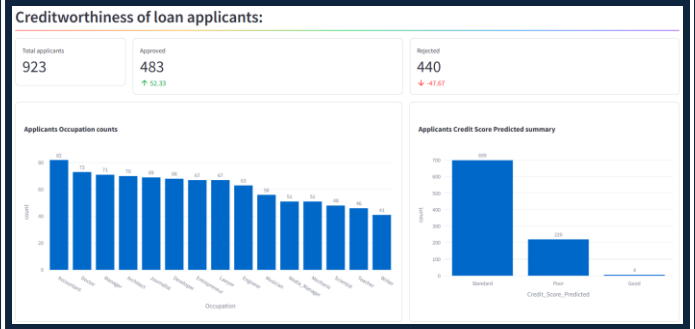
- **Rejected:** We can download the Rejected loan applicants list by clicking "Download Rejected loan applicants as CSV file

Robust Credit Scoring and Loan Approval System:

Loan Applicants in Cluster-1

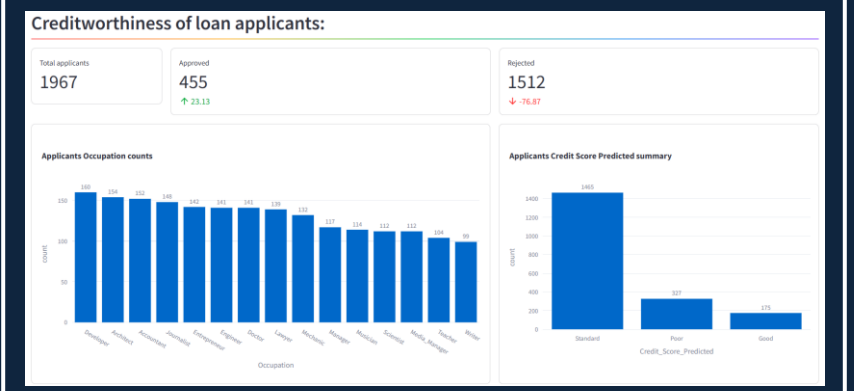
Loan Applicants in Cluster-2

Loan Applicants in Cluster-5



Loan Applicants in Cluster-3

Loan Applicants in Cluster-4



Insights

- **Loan Approval:** Displayed the Credit prediction insights from given CSV file. High risk profiles are getting rejected even though they are with Good or Standard credit score. This system provides the high level risk assessment loan approval process.

Perspective Analysis

- ✓ **Credit Improvement Programs:** Implement programs aimed at helping applicants improve their credit scores. This could include financial education, credit counseling, and credit-building loans.
- ✓ **Diversify Risk Profiles:** Consider approving applicants with slightly higher risk profiles but with solid income and financial stability, perhaps with adjusted loan terms or higher interest rates to offset the risk.




Thank you

Nandhagopal S

nandha2790@gmail.com

LinkedIn Profile