DSE Term Project Proposal

Project Title: Analysis of Credit Card Customer Information Dataset

Project Team:

	Name	ASU ID	Email ID
1.	**Nayan Bhiwapurkar**	1229979233	nbhiwapu@asu.edu
2.	**Kaumudi Patil**	1230036771	kspatil1@asu.edu
3.	**Shreyas Hingmire**	1230415136	shingmir@asu.edu
4.	**Sharan Patel**	1230596772	spate294@asu.edu

Project Objective:
To analyze and derive insights from a dataset containing credit card customer information.

Project Overview and Introduction to Dataset:

We propose to conduct a comprehensive analysis of a credit card customer information dataset. This dataset contains various attributes related to credit card customers, including their age, income, credit limit, credit card type, and transaction history. Our primary goal is to extract valuable insights from this data to help make informed decisions and improve our understanding of customer behavior. The dataset provides a subset of credit card customer information in 2018 and 2019.

Credit card customers are beneficial to banks as they generate revenue through various fees like annual fees and late payment penalties. Additionally, credit card usage often leads to increased customer engagement, potentially leading to cross-selling of other financial products and services, further boosting the bank's profitability. Lastly, responsible credit card usage can build long-term customer relationships, fostering loyalty and trust.

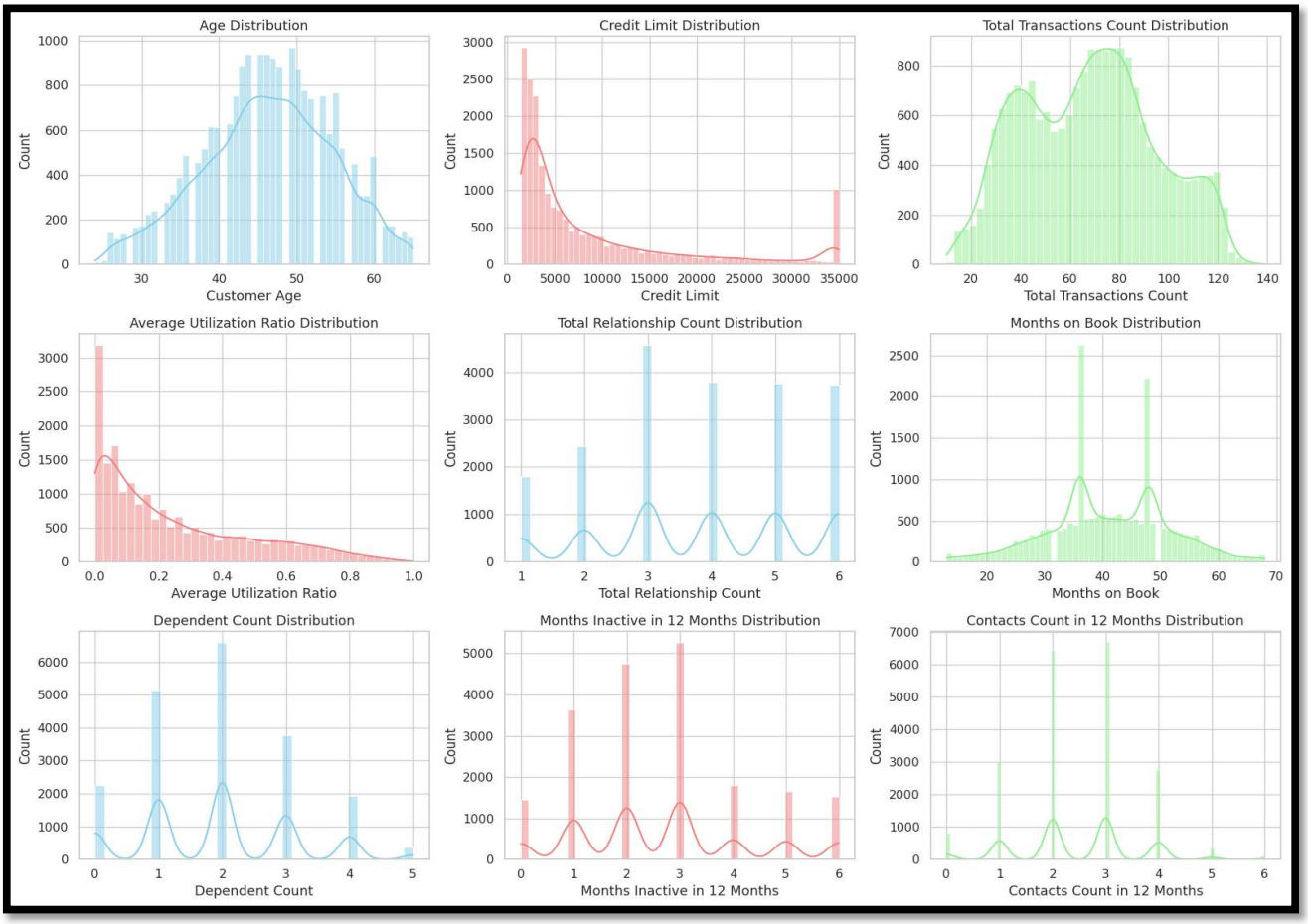
The purpose of this dataset could be to better understand the spending habits of credit card users and potentially identify trends or patterns in credit card usage, and lastly to find what factors lead to customers to leave or stop using credit cards.

Summary Statistics:

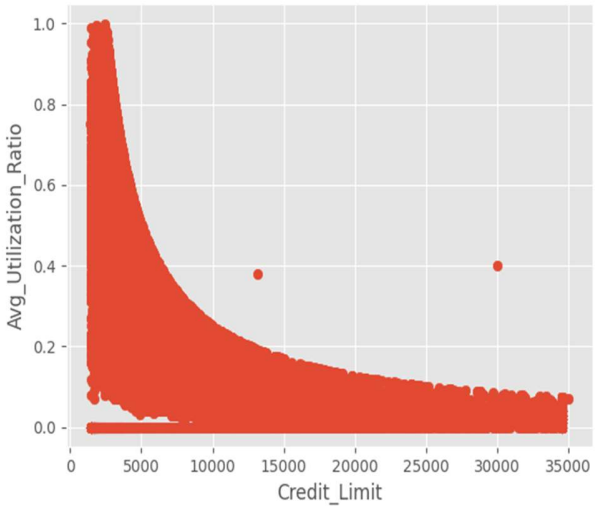
Column	Description	Variable Type	mean	std	min	max
CLIENTNUM	Client number. Unique identifier for the customer holding the account	Numerical				
Attrition_Flag	customer activity variable - if the account is closed then "Attrited Customer" else "Existing Customer"	Categorical				
Customer_Age	Age in Years	Numerical	46	8	24	65
Gender	Gender of the account holder - M / F	Categorical				
Dependent_count	Number of dependents	Numerical	2	1	0	5
Education_Level	Educational Qualification of the account holder - College, Doctorate, Graduate, High School, Post-Graduate, Uneducated	Categorical				
Marital_Status	Marital Status of the account holder - Divorced, Married, Single	Categorical				
Income_Category	Annual Income Category of the account holder - Less than $40K, $40K - $60K, $60K - $80K, $80K - $120K, $120K +	Categorical				
Card_Category	Type of Card - Blue, Silver, Gold, Platinum	Categorical				
Months_on_book	Period of relationship with the bank	Numerical	41	10	13	68
Total_Relationship_Count	Total no. of products held by the customer	Numerical	4	2	1	6
Months_Inactive_12_mon	No. of months inactive in the last 12 months	Numerical	3	2	0	6
Contacts_Count_12_mon	No. of Contacts between the customer and bank in the last 12 months	Numerical	2	1	0	6
Credit_Limit	Credit Limit on the Credit Card	Numerical	8637	9084	1400	35000
Total_Revolving_Bal	The balance that carries over from one month to the next is the revolving balance	Numerical	1011	658	0	12080
Avg_Open_To_Buy	Open to Buy refers to the amount left on the credit card to use (Average of last 12 months)	Numerical	7480	9103	3	34516
Total_Trans_Ct	Total Transaction Count (Last 12 months)	Numerical	68	27	10	139
Avg_Utilization_Ratio	Represents how much of the available credit the customer spent	Numerical	0	0	0	1
Quarter	Attrition Quarter - none, Q1, Q2, Q3, Q4	Categorical				
Year	Attrition Year - 2018, 2019	Categorical				
Date_Leave	Attrition Date - Quarter, Year	Categorical				

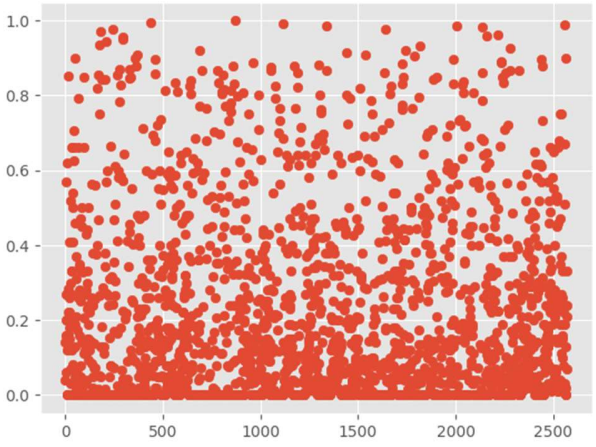
Data Visualization:

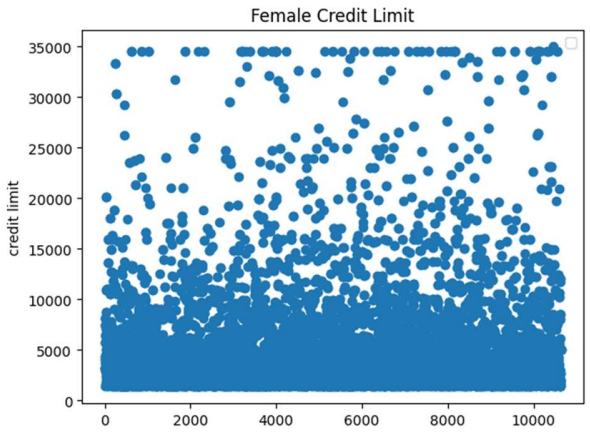
The visualizations above provide insights into key attributes of the dataset, such as age distribution, credit limit variability, and transaction behavior. They help in understanding the range and patterns within these variables, which is essential for data exploration and analysis.



Hypothesis Formulation:

Hypothesis 1	
Null Hypothesis (H0): "The proportion of people with a credit limit greater than $13,000 and a utilization ratio less than 0.2 is very close to 1."	
Alternative Hypothesis (H1): "There is a significant difference in the proportion of people with a credit limit greater than $13,000 and a utilization ratio less than 0.2 compared to the assumed proportion of very close to 1."	

Hypothesis 2	
Null Hypothesis (H0): "The proportion of attrited customers with an average utilization ratio less than 0.5 is more than 0.70.”	
Alternative Hypothesis (H1): “The proportion of attrited customers with an average utilization ratio less than 0.5 is less than 0.70."	

Hypothesis 3	
Null Hypothesis (H0): "The proportion of people with a credit limit greater than $25,000 who are women is 0.10"	
Alternative Hypothesis (H1): "The proportion of people with a credit limit greater than $25,000 who are women is much more than the estimated value of 0.10"	

Hypothesis 4 - Total Transaction Count and Attrition flag:

Null Hypothesis (H0): “Customers having lower total transaction counts have higher chance of attrition.”

Alternative Hypothesis (H1): “Customers having lower total transaction counts do not have higher chance of attrition.”

Hypothesis 5 - Age and Credit Limit:

Null Hypothesis (H0): “There is no significant correlation between customer age and credit limit.”

Alternative Hypothesis (H1): “Older customers tend to have higher credit limits.”

Hypothesis 6 - Utilization Ratio and Credit Limit:

Null Hypothesis (H0): “There is no significant correlation between average utilization ratio and credit limit.”

Alternative Hypothesis (H1): “Customers with higher credit limits tend to have a lower utilization ratio.”

Hypothesis 7 - Average Utilization and Gender

Null Hypothesis (H0): “The proportion of people with an average utilization ratio greater than 0.75 who are female greater than 0.7”

Alternative Hypothesis (H1): "The proportion of people with an average utilization ratio greater than 0.75 who are female is less than 0.7"