

**Predicting credit consumption**

**for Customers of a**

**leading bank**

By Anouska Priya

MAY-JULY 2023

# **CERTIFICATE**

This is to certify that Ms. Anouska Priya has successfully completed 2 months of Internship from 17th May 2023 to 17th July 2023 under Innovation Cell, ICT Division. During the period she worked on multiple sets of data and successfully modelled on python platform. During the period of his internship with us she was found punctual, hard-working, and creative.

**Mr. Ashish Das Mr. B. A. Reddy**

**Manager (Mining) GM (ICT), CMPDI**

**Innovation Cell/ICT Division**

# **Acknowledgement**

I would like to express my heartfelt gratitude to Mr. R. K. Mohapatro, GM (HRD), and Mr. B. A Reddy, GM (ICT), for their unwavering support and guidance during my training. Their expertise, encouragement, and constant availability have been instrumental in shaping my professional growth and contributing to the successful completion of my training.

I am also deeply grateful to Mr. Suresh Behera for his invaluable support throughout this journey. Furthermore, I would like to extend my appreciation to my esteemed mentors, Mr. Ashish Das and Mr. Himanshu Shekhar Sharma, whose thoughtful reviews and valuable guidance have played a crucial role in my development during this training period.

I am sincerely thankful to all the individuals mentioned above and to anyone else who has supported me throughout my Industrial Training.

Sincerely,

Anouska Priya

**Contents**

[**CERTIFICATE** 2](#_Toc140525177)

[**Acknowledgement** 3](#_Toc140525178)

[**Introduction:** 5](#_Toc140525179)

[**Problem Statement** 5](#_Toc140525180)

[**Objective:** 5](#_Toc140525181)

[**Data Availability:** 6](#_Toc140525182)

[**Data Dictionary** 6](#_Toc140525183)

[**a. CustomerDemographics.csv** 6](#_Toc140525184)

[**b. CustomerBehaviorData.csv** 6](#_Toc140525185)

[**c. CreditConsumptionData.csv** 6](#_Toc140525186)

[**Flow chart:** 6](#_Toc140525187)

[**Uploading the data:** 7](#_Toc140525188)

[**Data Visualization** 7](#_Toc140525189)

[**Correlation analysis** 7](#_Toc140525190)

[**Account type based on gender** 8](#_Toc140525191)

[**Credit card usage in 3 months by gender and account type** 8](#_Toc140525192)

[**Box plot for credit card consumption by gender** 9](#_Toc140525193)

[**Data Age bins** 10](#_Toc140525194)

[**Mean of total credit and debit consumption for April, May and June.** 10](#_Toc140525195)

[**Accessing models on data to predict credit card consumption** 12](#_Toc140525196)

[**1)Linear model** 12](#_Toc140525197)

[**2)Decision tree regressor** 13](#_Toc140525198)

[**3)Random Forest regressor** 13](#_Toc140525199)

[**Final Output: prediction of credit card consumption** 14](#_Toc140525200)

[**Suggestions** 14](#_Toc140525201)

[**Learnings** 15](#_Toc140525202)

[**Bibliography** 16](#_Toc140525203)

# **Introduction:**

In today's data-driven landscape, analytics has become a driving force across industries, enabling businesses to gather and analyze information from diverse sources to better understand customer needs and preferences. The Credit Card industry, being data-rich, can leverage this vast amount of information in countless ways to gain insights into customer behavior. By examining consumer types and their business spending behaviors, companies can develop targeted marketing campaigns that directly address consumer preferences. This, in turn, leads to improved sales and revenue growth.

A key aspect of customer relationship management in the Credit Card industry is understanding individual consumption patterns. This understanding enables banks to personalize their offerings and create strategic marketing plans tailored to the unique needs of consumers. Therefore, studying the relationship between consumer characteristics and their consumption patterns becomes crucial for financial institutions aiming to enhance customer satisfaction and loyalty.

By leveraging analytics, the Credit Card industry can unlock valuable insights and drive innovation by making informed decisions based on comprehensive data analysis. This empowers financial institutions to better understand their customers and develop effective strategies that meet their evolving needs.

# **Problem Statement**

Banks and financial institutions in the credit card industry lack a comprehensive understanding of the relationship between consumer characteristics and their credit card consumption patterns. This knowledge gap hinders their ability to customize marketing strategies, offer personalized services, and optimize credit card product offerings. As a result, banks struggle to enhance customer satisfaction, loyalty, and profitability. Additionally, there is a need to develop accurate predictive models to forecast credit card consumption based on the available data. Addressing these challenges is crucial for financial institutions to better serve their customers, drive business growth, and make informed decisions in the dynamic credit card industry.

# **Objective:**

The objective of this study is to investigate the relationship between consumer characteristics and credit card consumption patterns in order to develop a deeper understanding of individual consumer behavior. By achieving this objective, banks can customize marketing strategies, offer personalized services, and optimize credit card product offerings based on consumer preferences and requirements. This will ultimately lead to improved customer relationship management, enhanced customer satisfaction and loyalty, increased profitability, and a competitive advantage in the credit card industry.

# **Data Availability:**

Below data is available:

a. Customer Demographics

b. Customer Behavioural data (information on liabilities, assets and history of transactions with

the bank for each customer). Data has been provided for a particular set of customers' credit

card spend in the previous 3 months (April, May & June) and their expected average spend

in the coming 3 months (July, August & September)

c. Credit consumption

# **Data Dictionary**

## **a. CustomerDemographics.csv**

ID – Customer ID - Unique ID for every Customer

Account\_type - Account Type (current or saving)

Gender- Gender of customer (M or F)

Age - Age of customer

Income – Income Levels (High/Medium/Low)

Emp\_Tenure\_Years – Experience – Employment Tenure of customer in Years

Tenure\_with\_Bank – Number of years with bank

Region\_code Code assigned to region of residence (has order)

NetBanking\_Flag – Whether customer is using net banking for the transactions

Avg\_days\_between\_transaction – Average days between two transactions

## **b. CustomerBehaviorData.csv**

ID – Customer ID - Unique ID for every Customer

CC\_cons\_apr - Credit card spend in April

DC\_cons\_apr - Debit card spend in April

CC\_cons\_may - Credit card spend in May

DC\_cons\_may - Debit card spend in May

CC\_cons\_jun - Credit card spend in June

DC\_cons\_jun - Debit card spend in June

CC\_count\_apr - Number of credit card transactions in April

CC\_count\_may - Number of credit card transactions in May

CC\_count\_jun - Number of credit card transactions in June

DC\_count\_apr - Number of debit card transactions in April

DC\_count\_may - Number of debit card transactions in May

DC\_count\_jun - Number of debit card transactions in June

Card\_lim - Maximum Credit Card Limit allocated

Personal\_loan\_active - Active personal loan with other bank

Vehicle\_loan\_active - Active Vehicle loan with other bank

Personal\_loan\_closed - Closed personal loan in last 12 months

Vehicle\_loan\_closed - Closed vehicle loan in last 12 months

Investment\_1 - DEMAT investment in june

Investment\_2 - Fixed deposit investment in june

Investment\_3 - Life Insurance investment in June

Investment\_4 - General Insurance Investment in June

Debit\_amount\_apr - Total amount debited for April

Credit\_amount\_apr - Total amount credited for April

Debit\_count\_apr- Total number of times amount debited in april

Credit\_count\_apr - Total number of times amount credited in april

Max\_credit\_amount\_apr - Maximum amount credited in April

Debit\_amount\_may - Total amount debited for May

Credit\_amount\_may - Total amount credited for May

Credit\_count\_may - Total number of times amount credited in May

Debit\_count\_may - Total number of times amount debited in May

Max\_credit\_amount\_may - Maximum amount credited in May

Debit\_amount\_jun - Total amount debited for June

Credit\_amount\_jun - Total amount credited for June

Credit\_count\_jun - Total number of times amount credited in June

Debit\_count\_jun - Total number of times amount debited in June

Max\_credit\_amount\_jun - Maximum amount credited in June

Loan\_enq - Loan enquiry in last 3 months (Y or N)

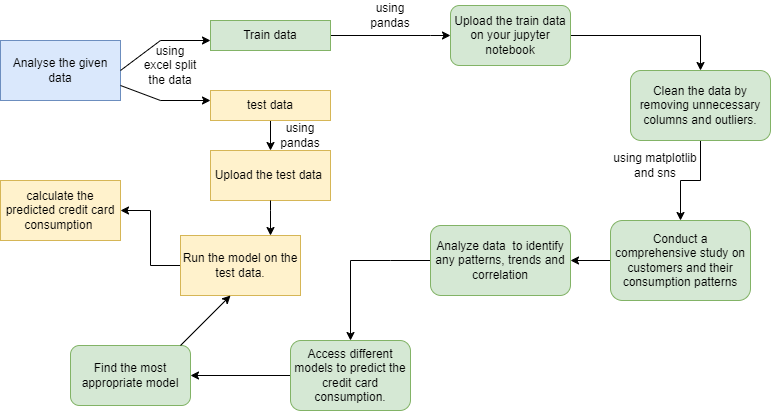
Emi\_active - Monthly EMI paid to other bank for active loans

## **c. CreditConsumptionData.csv**

ID – Customer ID - Unique ID for every Customer

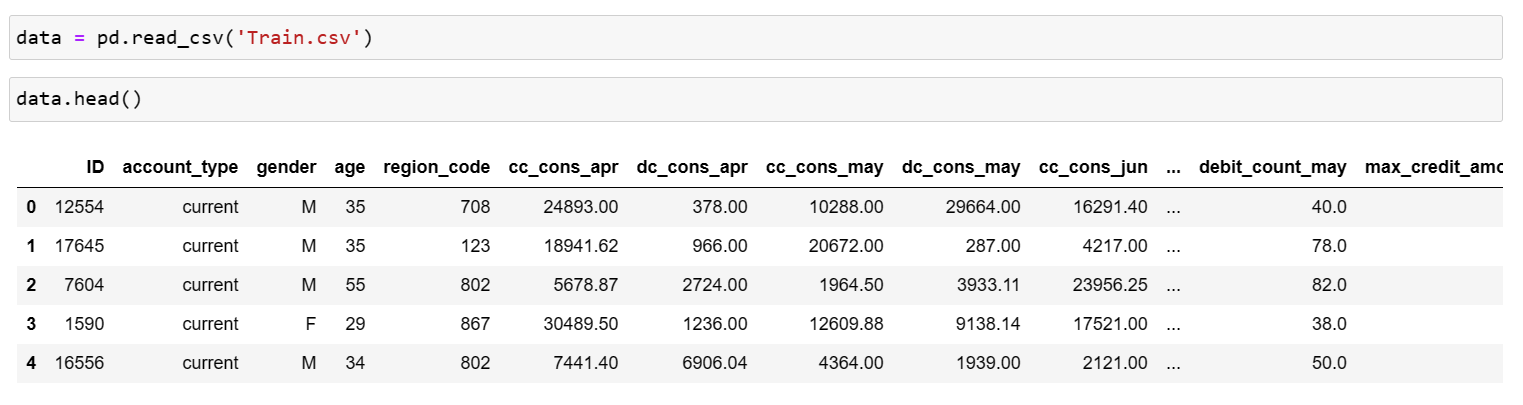
cc\_cons (Target) - Average Credit Card Spend in next three months

# **Flow chart:**



# **Uploading the data:**

Using pandas we upload the train data

****

# **Data Visualization**

## **Correlation analysis**

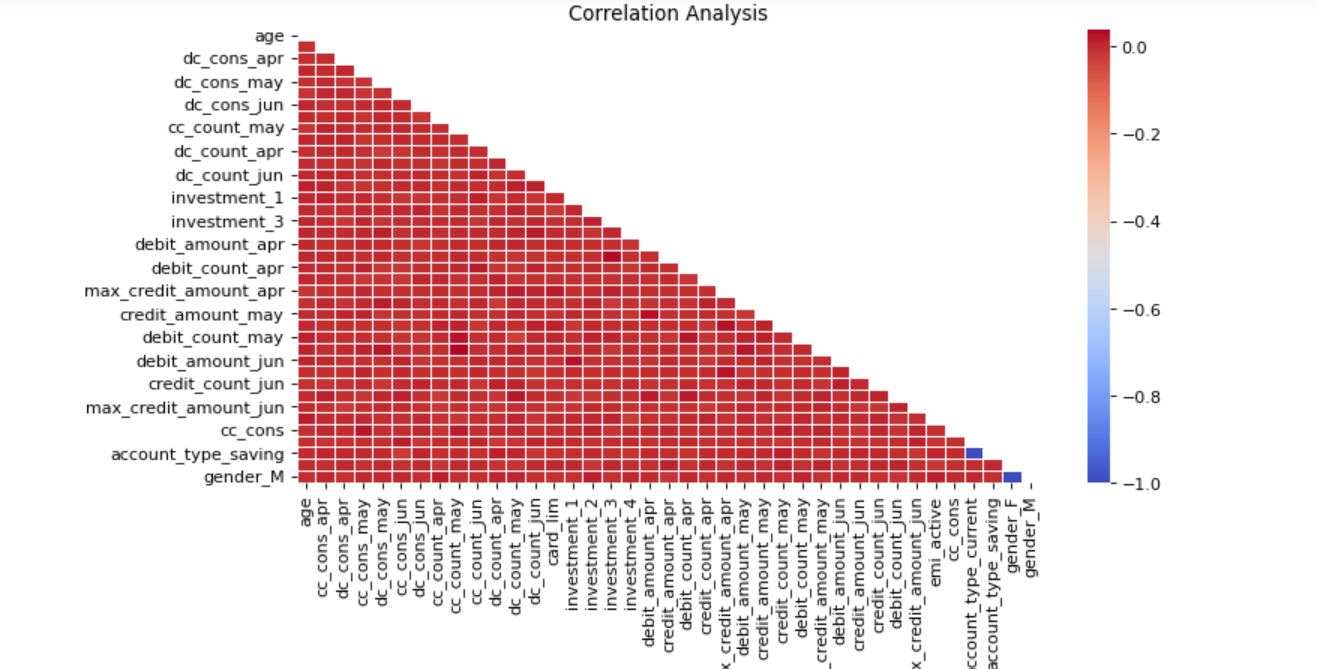
****

Figure 1-Correlation

A correlation close to zero indicates that there is no linear relationship between the variables. However, it does not necessarily mean that there is no relationship at all. Nonlinear relationships or other complex interactions might still exist, which are not captured by the correlation coefficient

## **Account type based on gender**

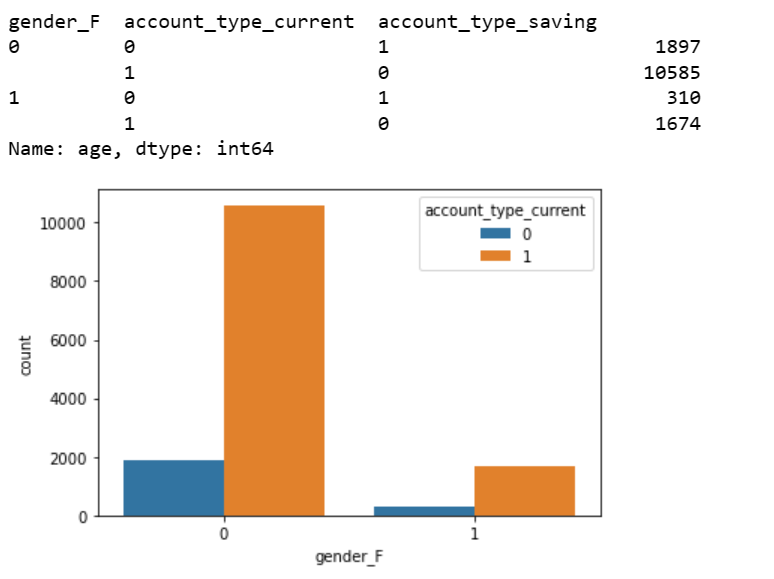


Figure 2-Bar graph based on gender

We can see here that in both male and female current account is more popular.

## **Credit card usage in 3 months by gender and account type**

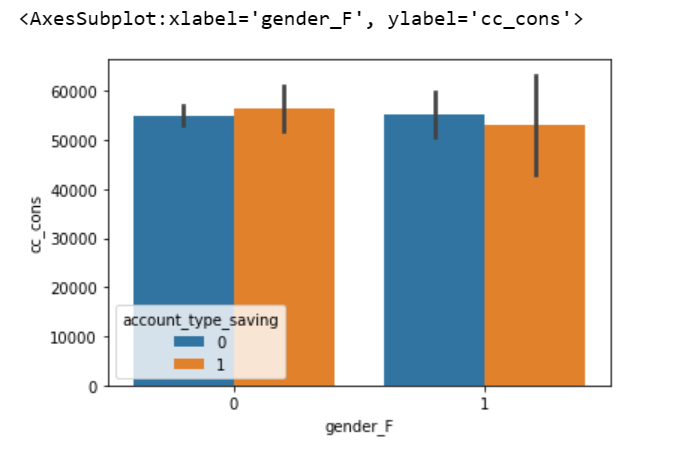


Figure 3-credit card usage by gender and account type

The above graph shows the relation between gender, account type and credit card consumption.

**Box plot for credit card consumption by gender**

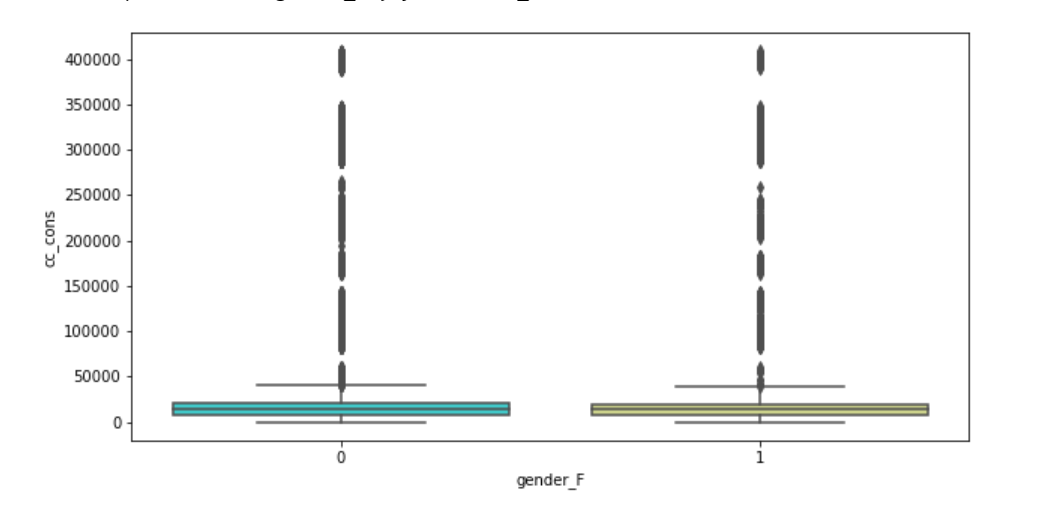


Figure 4-box plot for cc cons based on gender

Based on the box plot graph, we observe that the majority of credit card consumption values fall within the range of 0 to 50,000. However, removing the outliers from the dataset might not be appropriate since there is a substantial number of data points classified as outliers.

## **Data Age bins**

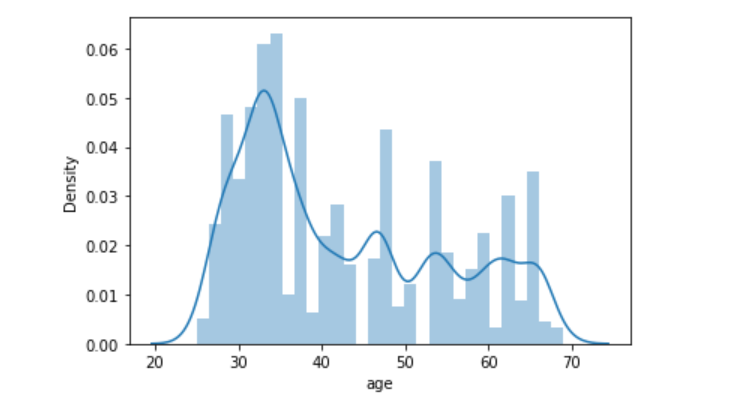
****

Figure 5-age bins

Here we can see that the maximum data lies between the age of 30 to 40.

## **Mean of total credit and debit consumption for April, May and June.**

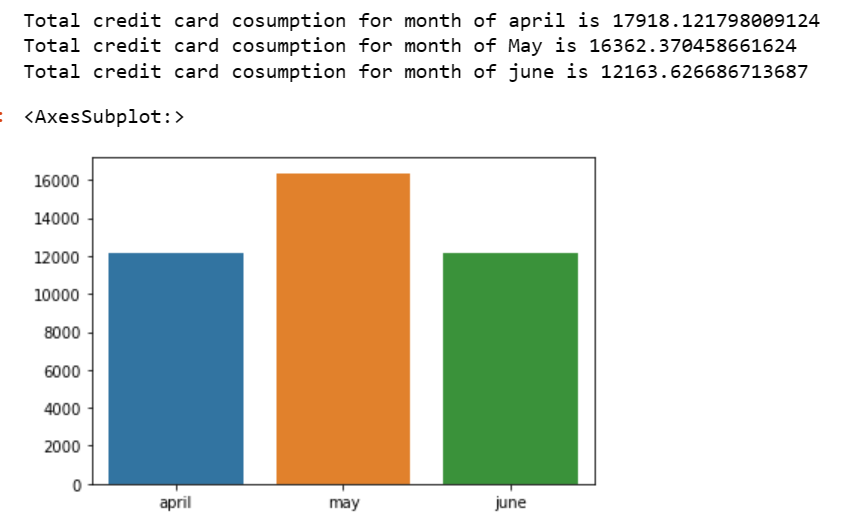


Figure 6-Credit consumption for april, may and june

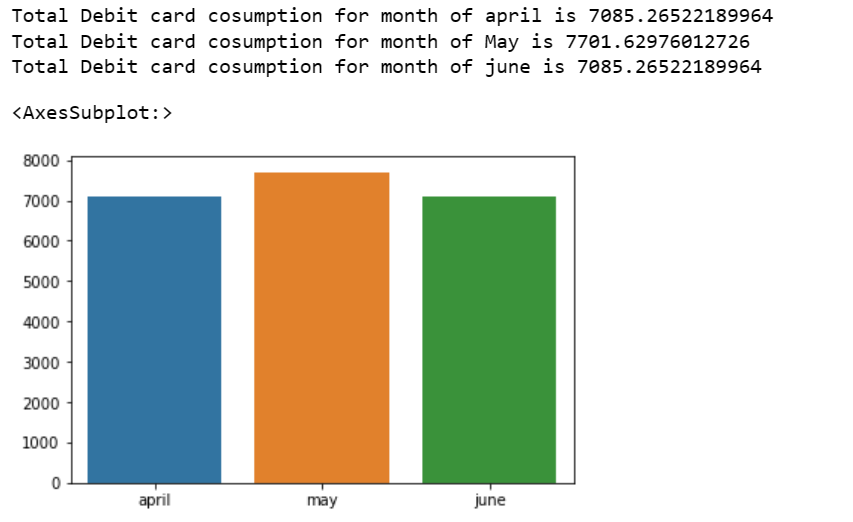


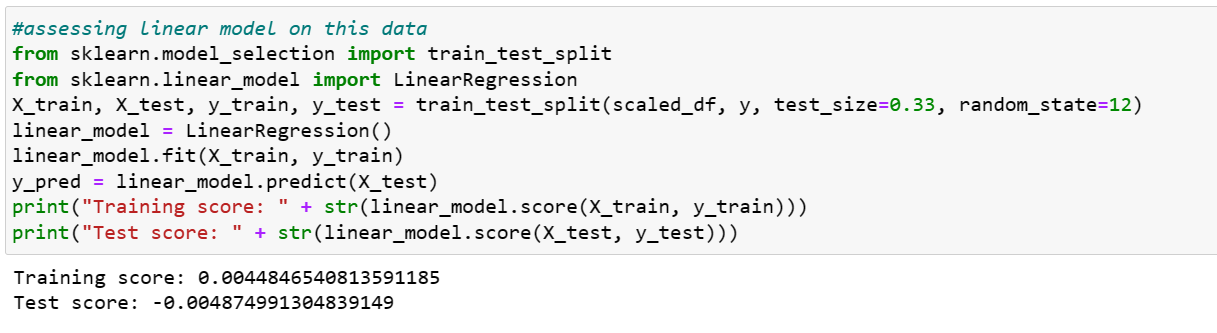
Figure 7-debit consumption for april, may and june

By comparing both graphs, it becomes evident that the credit card usage among customers is considerably higher.

Considering all these points together, we can conclude that credit card usage is considerably higher among customers, and current accounts appear to be the preferred choice for both male and female individuals. The credit card consumption values primarily range from 0 to 50,000, and the age group of 30 to 40 shows the highest concentration of data.

# **Accessing models on data to predict credit card consumption**

## **1)Linear model**

****

The linear model seems to have a poor fit to the data. It explains only a small fraction of the variance in both the training and test datasets, suggesting that the model might not capture the underlying patterns or relationships present in the data effectively.

Plotting the predicted y- values for this model:

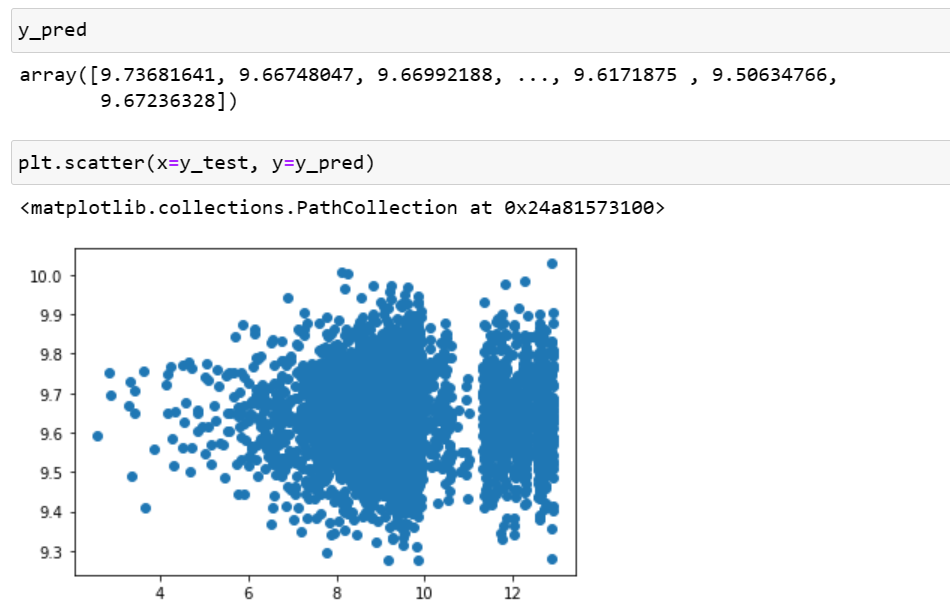
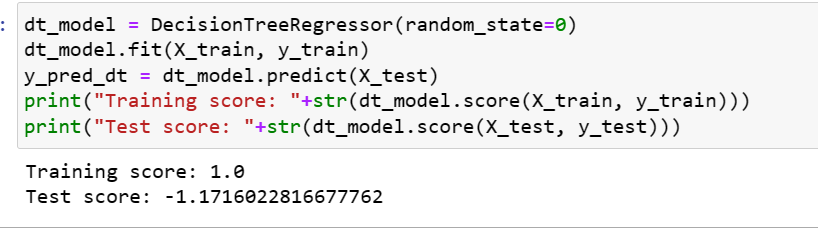


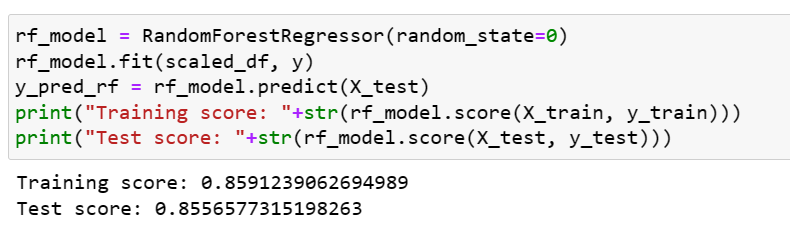
Figure 8-Y prediction plot

## **2)Decision tree regressor**

****

The Decision Tree Regressor model seems to have overfit the training data perfectly, but it performs poorly on unseen data (the test dataset). This suggests that the model has not generalized well and may not be suitable for accurately predicting the target variable on new, unseen instances.

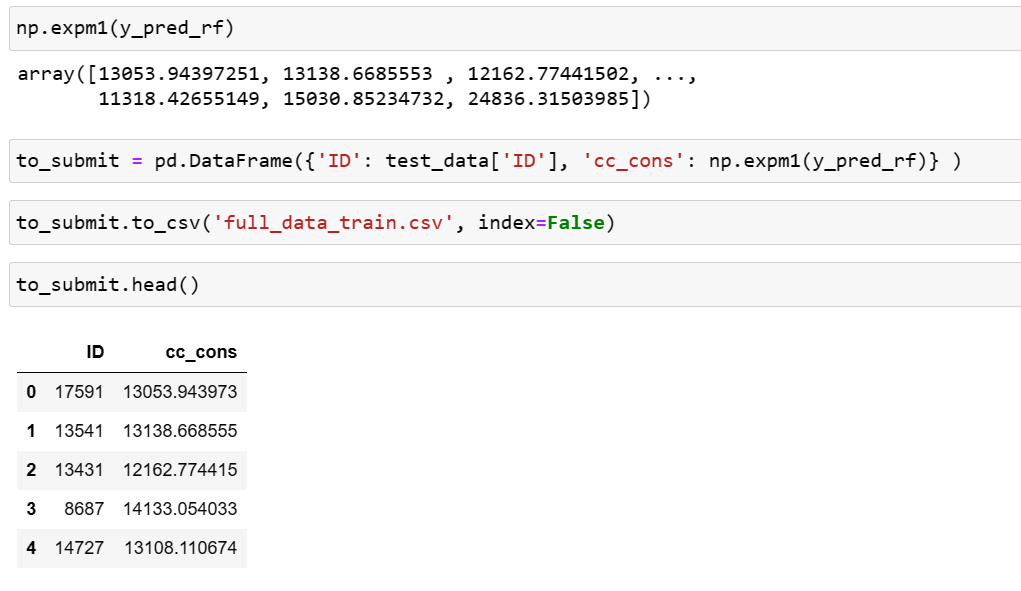
## **3)Random Forest regressor**

****

The Random Forest Regressor model demonstrates strong performance on both the training and test datasets. It explains a significant portion of the variance in the target variable and generalizes well to unseen data. These scores suggest that the Random Forest Regressor model is capable of accurately predicting the target variable and capturing the underlying patterns in the data**.**

**Thus, for our test data we will use random forest to predict the credit card consumption values.**

# **Final Output: prediction of credit card consumption**

****

In the given image we can see that the credit consumption values are predicted using Random Forest regressor.

# **Suggestions**

* Targeted Campaigns: Utilize the knowledge that current accounts are more popular among both male and female individuals. Design targeted marketing campaigns that highlight the benefits of credit cards linked to current accounts, such as convenient payment options, rewards, or exclusive offers for current account holders.
* Personalized Offers: Leverage the predicted credit card consumption values to tailor personalized offers for individual customers.
* Targeted Communication: Craft targeted and personalized communication strategies to engage customers. Utilize customer segmentation based on gender, age, and account type to deliver relevant content through various channels, such as email marketing, social media, or mobile apps.

# **Learnings**

Throughout my journey, I have acquired expertise in utilizing the SAS Visual Analytics tool to create dynamic and interactive dashboards. Working with diverse datasets has honed my skills in data analysis, enabling me to provide effective solutions to various problems. Data cleaning, including the removal of unnecessary columns and handling outliers, has become second nature to me.

By delving into different types of models available in Python, I have expanded my repertoire of analytical techniques. Applying these models to real-world datasets has provided me with valuable insights and the ability to derive actionable recommendations.

A strong emphasis on data visualization has allowed me to effectively communicate complex findings. Visualizing data enables me to uncover meaningful patterns and trends, enhancing my ability to provide impactful suggestions and guide decision-making processes.

In conclusion, my proficiency in SAS Visual Analytics, data analysis, data cleaning, model application, and data visualization empowers me to excel in the field of data analytics. I am well-prepared to offer valuable insights and make a significant impact through my analytical skills and problem-solving capabilities.

# **Bibliography**

[www.kaggle.com](http://www.kaggle.com)

<https://www.w3schools.com/python/default.asp>

An introduction to machine learning by Miroslav Kubat