

**A PROJECT REPORT ON**

# **BI Analytics**

**SUBMITTED BY**

**MR. ROHAN MAHENDRA JOSHI**

**FOR THE YEAR**

**2019-2020**

**UNDER THE GUIDANCE OF**

**PROF. MR. ABHIJEET GOLE**

**SUBMITTED IN PARTIAL FULFILLMENTS OF**

**REQUIREMENTS OF QUALIFYING**

**M.SC. (COMPUTER SCIENCES)**

**S P MANDALI'S**

**RAMNARAIN RUIA AUTONOMOUS COLLEGE,**

**MATUNGA, MUMBAI - 400 019**

**RE-ACCREDITED GRADE 'A+' BY NAAC**

## **Acknowledgement**

I owe my deep gratitude to our project guide Prof. Mr. Abhijeet Gole who took keen interest in my project work and guided me to complete this project successfully. I would like to extend my sincere thanks to all staff members at ITCS Department for their timely support. I would like to express gratitude towards my Prof. and Head of the department Mrs. Megha Swant for giving me opportunity to undertake this wonderful project. I am also thankful to Prof. Mrs. Rasika Mundhe and Mrs. Edith Juni for helping me to overcome obstacles in this project.

## **Declaration by the student**

I hereby declare that the project entitled “**BI Analytics**” which is being submitted as a project of 3<sup>rd</sup> semester of Masters in computer science to Ramnarain Ruia Autonomous College, is an authenticate record of my genuine work done under the guidance of Prof. Mr.Abhijeet Gole, Department of Computer Science, Ramnarain Ruia Autonomous College.

**Rohan Joshi**

## **Preface**

This project is an Analytics Dashboard based on Dummy Bank Data using Python Programming Language. For bank to know it's customers which is very important and know their needs what think they want and what are their expectations from bank. It also helps for strategy making and marketing of bank. The project objective is to deliver the analytics to the bank authority.

In this project all analysis is done on stored data. And data is generated randomly, regarding basic discussion with the bank managers and IT guys in bank.

# Index

<b>Chapter 1 .....</b>	<b>8</b>
<b>1.2 Description of system.....</b>	<b>9</b>
<b>1.3 Limitation of existing system .....</b>	<b>10</b>
<b>1.4 Problem Statement.....</b>	<b>11</b>
<b>1.5 Proposed system and advantages.....</b>	<b>12</b>
<b>1.6 Project Scope.....</b>	<b>13</b>
<b>1.7 Feasibility study .....</b>	<b>14</b>
<b>1.8 Stakeholders .....</b>	<b>15</b>
<b>1.9 Technologies used in the Project.....</b>	<b>16</b>
<b>Chapter 2 .....</b>	<b>17</b>
<b>System Workflow.....</b>	<b>17</b>
<b>Process: -.....</b>	<b>17</b>
<b>1. Data Generation.....</b>	<b>17</b>
<b>2. Data Cleaning.....</b>	<b>17</b>
<b>3. Data Preprocessing .....</b>	<b>17</b>
<b>4. Data Analysis.....</b>	<b>17</b>
<b>1. Data Generation:.....</b>	<b>17</b>
<b>2. Data Cleaning:.....</b>	<b>17</b>
<b>3. Data Preprocessing:.....</b>	<b>17</b>
<b>4. Data Analysis:.....</b>	<b>17</b>
<b>Chapter 3 .....</b>	<b>18</b>
<b>System Interpretation .....</b>	<b>18</b>
• <b>Credit Transactions .....</b>	<b>18</b>
• <b>Services.....</b>	<b>18</b>
• <b>Insurance .....</b>	<b>18</b>
• <b>Gold Investment .....</b>	<b>18</b>
• <b>SIP Investment .....</b>	<b>18</b>
• <b>Overview:.....</b>	<b>19</b>
• <b>Saving Accounts:.....</b>	<b>19</b>
• <b>Current Accounts: .....</b>	<b>19</b>
• <b>Loan Account: .....</b>	<b>19</b>
• <b>FD and RD Accounts:.....</b>	<b>19</b>

• <b>Debit and Credit Transactions:</b> .....	20
• <b>Services:</b> .....	20
• <b>Insurance:</b> .....	20
• <b>Gold Investment:</b> .....	20
• <b>SIP Investment:</b> .....	20
• <b>Social Media:</b> .....	20
<b>Chapter 4 .....</b>	21
<b>System implementation .....</b>	21
<b>Overview:.....</b>	21
<b>Saving Account: .....</b>	24
<b>Current Account: .....</b>	28
<b>Loan Account: .....</b>	33
<b>FD Accounts: .....</b>	40
<b>RD Accounts:.....</b>	44
<b>Debit Transactions:.....</b>	48
<b>Credit Transactions: .....</b>	51
<b>Insurance: .....</b>	53
<b>Services: .....</b>	57
<b>Gold Investments: .....</b>	59
<b>SIP Investments: .....</b>	61
<b>Social Media: .....</b>	63
<b>Mobile View :- .....</b>	68
<b>ML Models: - .....</b>	70
• <b>Customer Income Prediction Model .....</b>	70
• <b>Customer Churn Detection Model .....</b>	70
• <b>Evaluating Social Networking ads base customer purchasing policy's .....</b>	70
• <b>Credit Card Froud Detection.....</b>	70
• <b>Credit Risk Prediction .....</b>	70
• <b>Automobile Price Prediction .....</b>	70
• <b>House Rate Prediction .....</b>	70
<b>AI Model:.....</b>	80
<b>Chapter 6 .....</b>	84
<b>Summary.....</b>	84
<b>Chapter 7 .....</b>	85
<b>References and Bibliography .....</b>	85

<b>Chapter 8 .....</b>	86
<b>Index &amp; Acronyms.....</b>	86

# **Chapter 1**

## **Preliminary investigation**

### **1.1 Organizational overview:**

In the Bank there are several departments has Operation department, FDR department, Credit department, Insurance and Investments department, ATM department. Each Department have different tasks and different types of data generated by each department daily. All records are stored in traditional system in SQL database.

## **1.2 Description of system**

To Start the Analysis first step is to clean the data. After cleaning of the data, we do the aggregation of data also grouping data and get the frequency count of it. We do the grouping based on the gender, region, years, days and months and other attributes specifically in data. Also finds some facts in a data, we also consider some cities at some instance. We also try to understand customers more deeply and grow rapidly in banking sector. We make some machine learning models and implement artificial intelligence in system.

To get more details of a process of developing the system and ML models and implementation of AI in Chapter No 2.

### **1.3 Limitation of existing system**

The current system is analytics based on SQL Queries. So, it gives following disadvantages.

1. Many organizations do not have the data analysis feature in their system.
2. SQL has limited set of queries to analyze the data.
3. SQL queries are not made for data analysis.
4. Doing analysis with PL/SQL. It is time consuming and difficult in terms of line of code.
5. All the SQL analysis is summary based analysis.

## **1.4 Problem Statement**

Create an Analytics Dash board on web. Analysis should answer all the facts and queries in a data. Data should be presented in a website which is responsive and mobile friendly.

System should identify the customer churn behaviour, and anomaly detections and evaluating risk while giving a loan, predict the income of a person and also calculate values of customer assets for mortgage. Also find which region marketing strategies are working and what are chances whether we get a customer or not.

## **1.5 Proposed system and advantages**

In the proposed System analytics done using Pandas and matplotlib and scikit-learn in python.

For representation we use HTML5 and W3-CSS as CSS library. Because of this my project satisfies all the requirements and overcome the limitation of existing system. The Proposed system is sample or demonstration of actual system. The analytics dashboard is based on web and it's mobile friendly and hosted on the Cloud. We implement of Microsoft Azure for ML and google colabs for AI model creation and training purpose.

### **Advantages:**

- I cover all possible major aspects of analysis in given data.
- Implementation of ML Model
- Implementation of AI
- Web Interface for Access
- Mobile friendly and Responsive Website.
- Google Assistant Support

## **1.6 Project Scope**

The scope of the project is to do data analysis of given data and covers maximum all aspect in datasets, Mostly all major ones. And create a ML model and test it on data and try to get maximum accuracy from it. Building the customer behaviour detection system is the goal of the project for the growth of the bank.

## **1.7 Feasibility study**

I fulfil the 95% wish of stake holder. And the remaining 5% will be completed in the next release. The Project is 100% completed and working in all platforms.

Website is very easy to use. It has attractive UI. Any person can easily operate this application. The application is easy to scaling and easy to maintain as it includes low maintenance cost.

## **1.8 Stakeholders**

The major stake holder of this project is Bank Managers, CEO and Director of the bank.

But Marketing team and Strategy building team is also a stake holder of application as well as Loan department.

This project will be beneficial to the bank to understand the customer and get more customer based analytics and detect customer behaviour and evaluating the customer applications for loan using AI and ML Models.

## **1.9 Technologies used in the Project**

- Python Complete project done in python is the primary programming language. Apart from python R programing is used for data cleaning in some ML model.
- For Representation purpose HTML5 and W3 CSS as CSS library is used.
- For Analysis NumPy and Pandas library is used.
- For Graph plotting Matplotlib and Seaborn is used.
- For Map Folium is python library is used.

## Chapter 2

### System Workflow

**Process:** -

- 1. Data Generation**
- 2. Data Cleaning**
- 3. Data Preprocessing**
- 4. Data Analysis**

#### **1. Data Generation:**

Before generating data, I discuss this project topic with Bank manager to get more details. According to their guidance I generate the data using python random function. All the data stored it in CSV files.

#### **2. Data Cleaning:**

For data cleaning, I have different date formatted dates so I want to split the date in three different columns, day, month, year I split date using pandas and lambda function in python.

#### **3. Data Preprocessing:**

After data cleaning, data aggregation take place it means grouping the data, calculating mean and frequency count from data. This process makes easier to day, today analysis. For data preprocessing I used pandas and NumPy libraries.

#### **4. Data Analysis:**

After data preprocessing, we get summarize data in dataframe format. So, it is not much effectively understandable as compare to graphical format. According to the dataframe format I choose appropriate graph to represent it, for representation in graph I used matplotlib and seaborn library. After that I want to represent it on web so I used HTML5 and CSS.

## **Chapter 3**

### **System Interpretation**

- **Overview**
- **Saving Accounts**
- **Current Accounts**
- **Loan Account**
- **FD Accounts**
- **RD Accounts**
- **Debit Transactions**
- **Credit Transactions**
- **Services**
- **Insurance**
- **Gold Investment**
- **SIP Investment**

- **Overview:**

In overview first we show the current status of the bank. It means we count total saving accounts, current accounts, FDR accounts, count of insurance sold, and Gold and SIP investment doing by customers.

I also represent the geographic location of each branch in a map, also show each region contribution in the bank in tabular form.

It also shows the current stock state in graphical format in timeseries graph ( randomly graph plotting because web scraping is not allowed and it is illegal for stock market ).

I also show the upcoming business year's targets.

- **Saving Accounts:**

In Saving Account, we are grouping the data in region, gender, year, month and day. We also categorize the data based on their basic details like having a car, having a home, having a bike and their education and many more.

Based on this categorical data we plot analyze the customers and their needs more efficiently. Based on their details we can predict when and at what time the customers want loan from us and other services like Insurance, SIP Investment etc. We also find out some amazing facts about top five cities which have maximum CASA deposit. We also find according to the age group how much account balance they have and also, we find at what age customers are opening their account in our bank. It helps us to target the customers by age group.

- **Current Accounts:**

In the current account customer is a businessman. So current account doesn't have any rate of interest unlike saving account. In current account we are grouping the data in region, gender, year, month and day. We also categorize the data based on their basic details like occupation. We find out some facts like maximum net worth by occupation.

- **Loan Account:**

In loan account we classify it by region, gender, purpose, rate of interest and also by tenure and mortgage type. We find some amazing facts about the people that is what region pays maximum EMI and which region has maximum loan distribution.

- **FD and RD Accounts:**

It is classified by region, gender, purpose, and also by tenure and loan taken against FD and RD. It is also classified as per maturity date also. So, we can set the target for a particular year. We also find the facts about the region which had maximum FD and RD amounts

- **Debit and Credit Transactions:**

We monitor Debit and Credit transactions by sources like cash withdrawal, cash deposits, NEFT, RTGS, Rupay Debit Card and UPI etc. we also monitor the transactions by month, year and date. So, we can plan which day and month is having more load than other days, Like festival days.

- **Services:**

We also do analysis on services availed by bank customer, we classify it by region, gender and usage of each service.

- **Insurance:**

Bank also provide insurance facility to their customers; we classify sold insurance by region gender and the frequent purchasing of it. Based on it we recommend the insurance to the users.

- **Gold Investment:**

Bank also provide direct gold investment plans for their customer. We classify by region, gender, amount and tenure. we find some facts that maximum amt purchased by which gender and maximum gold purchase by which region.

- **SIP Investment:**

Bank also provide SIP Investment for their customers. We classify it by region, gender and tenure of it. We also find some facts that most amt invested by region and which gender prefers SIP investments.

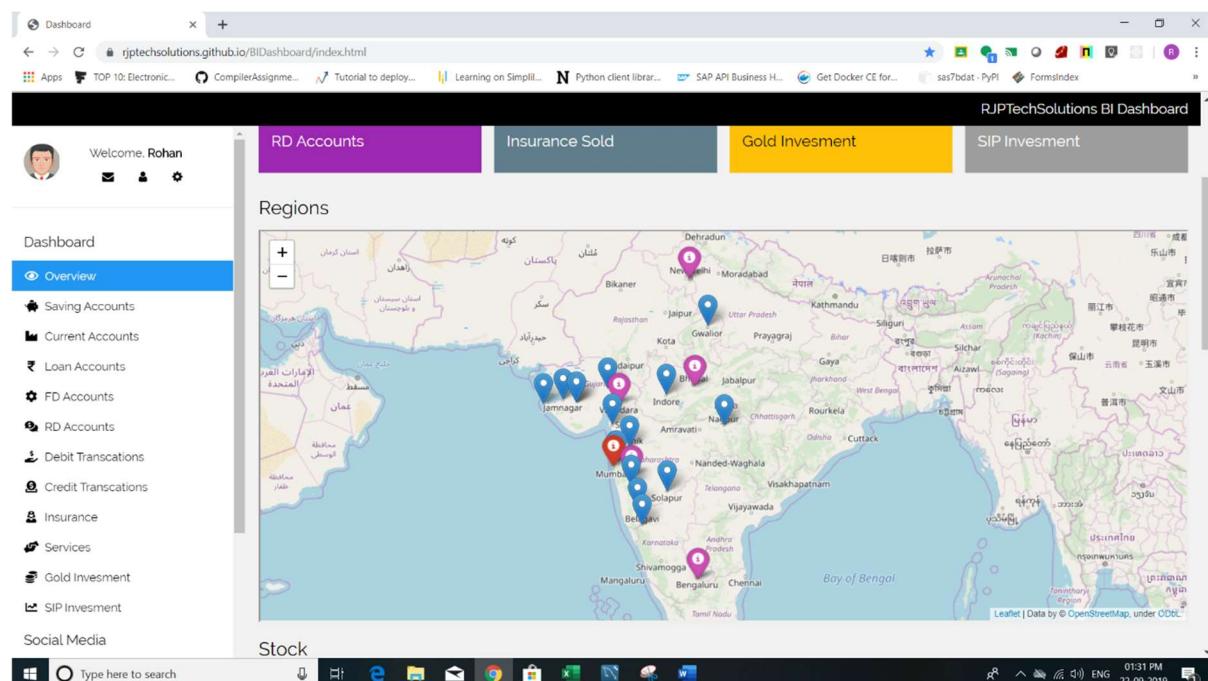
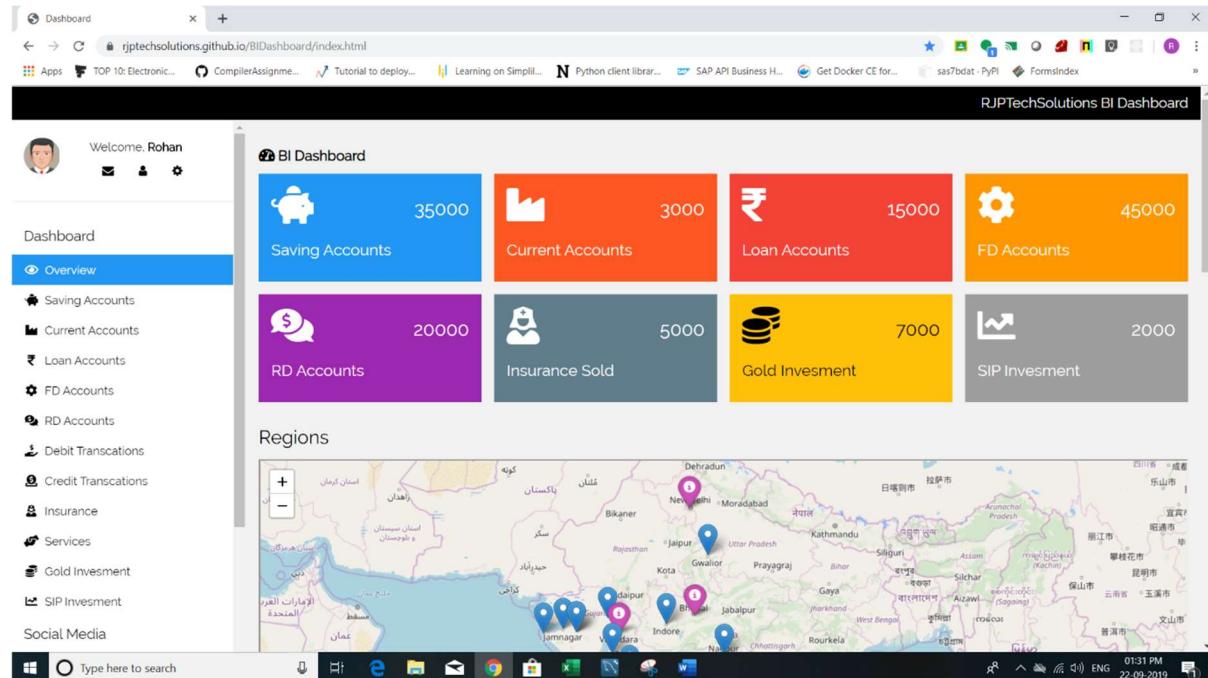
- **Social Media:**

From the bank's point of view social media is a key point of marketing and growing rapidly. I don't have official API of Social Media so I just give the link of it but in future updates we can add social media handers console in these tabs.

# Chapter 4

## System implementation

### Overview:



Dashboard x

rjptechsolutions.github.io/BIDashboard/index.html

Apps TOP 10: Electronic... CompilerAssignment... Tutorial to deploy... Learning on Simpli... Python client libr... SAP API Business H... Get Docker CE for... sas7bdat - PyPI FormsIndex

RJPTechSolutions BI Dashboard

Welcome, Rohan

Dashboard

Overview

- Saving Accounts
- Current Accounts
- Loan Accounts
- FD Accounts
- RD Accounts
- Debit Transactions
- Credit Transactions
- Insurance
- Services
- Gold Investment
- SIP Investment
- Social Media

Stock

Stock Rate

Tragets

	45000
	5000
	18000
	50000

CanvasJS.com

Type here to search

01:31 PM 22-09-2019

Dashboard x

rjptechsolutions.github.io/BIDashboard/index.html

Apps TOP 10: Electronic... CompilerAssignment... Tutorial to deploy... Learning on Simpli... Python client libr... SAP API Business H... Get Docker CE for... sas7bdat - PyPI FormsIndex

RJPTechSolutions BI Dashboard

Loan Accounts

FD Accounts

RD Accounts

Debit Transactions

Credit Transactions

Insurance

Services

Gold Investment

SIP Investment

Social Media

Facebook

Instagram

Twitter

LinkedIn

Github

RJPTechSolutions

Dev Instagram

Tragets

	45000
	5000
	18000
	50000
	22000
	7000
	10000
	3000

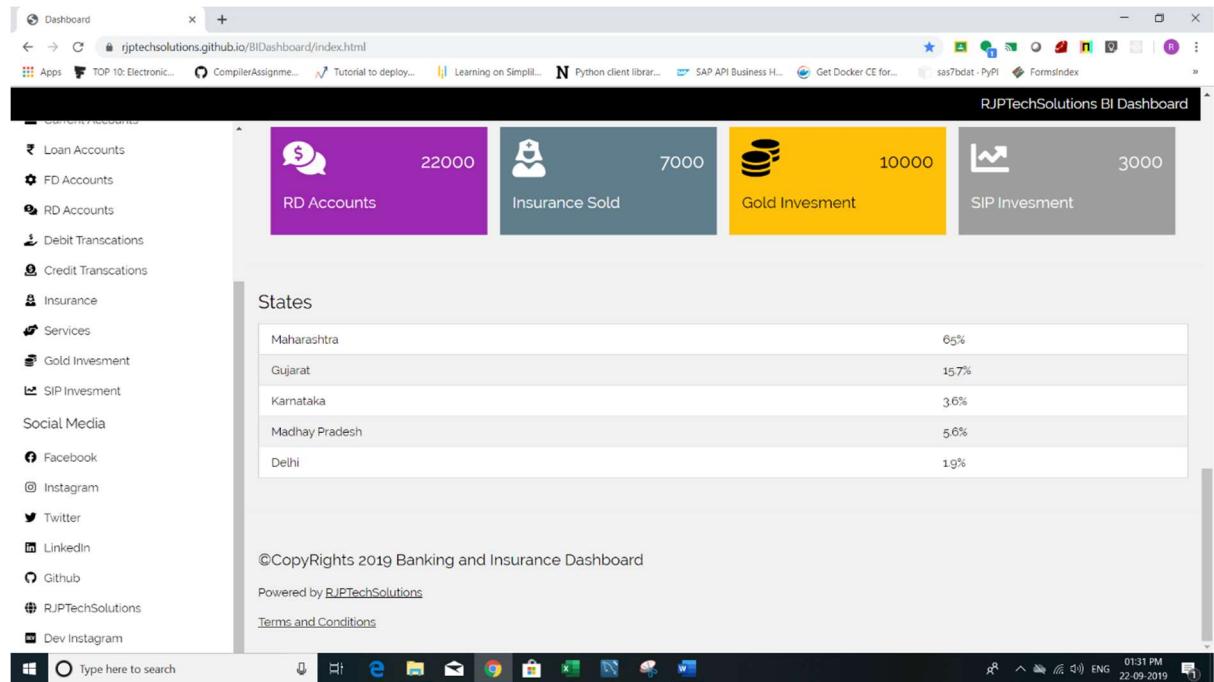
States

Maharashtra	65%
Gujarat	157%
Karnataka	3.6%

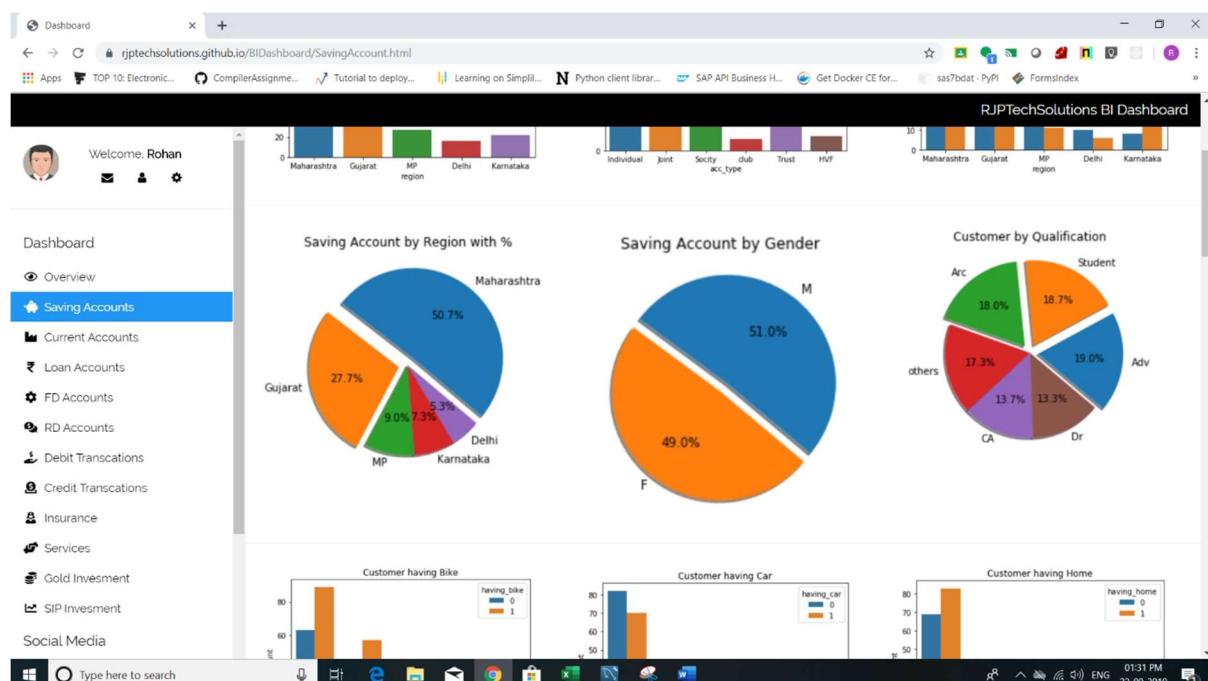
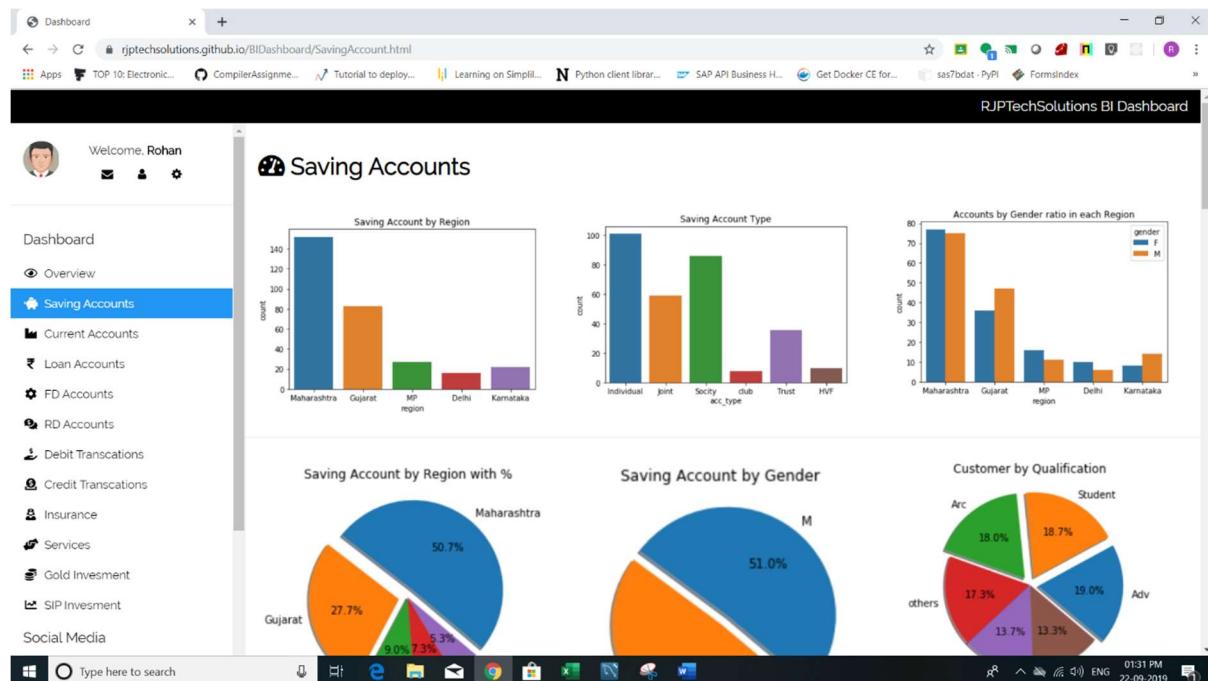
CanvasJS.com

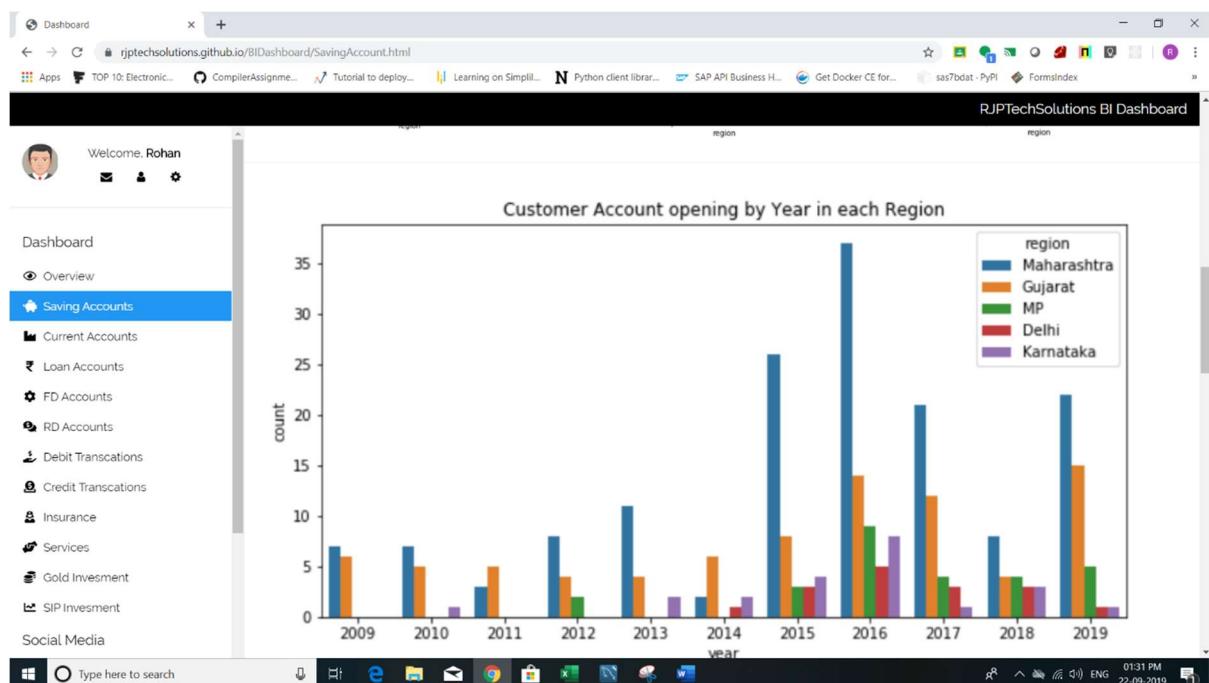
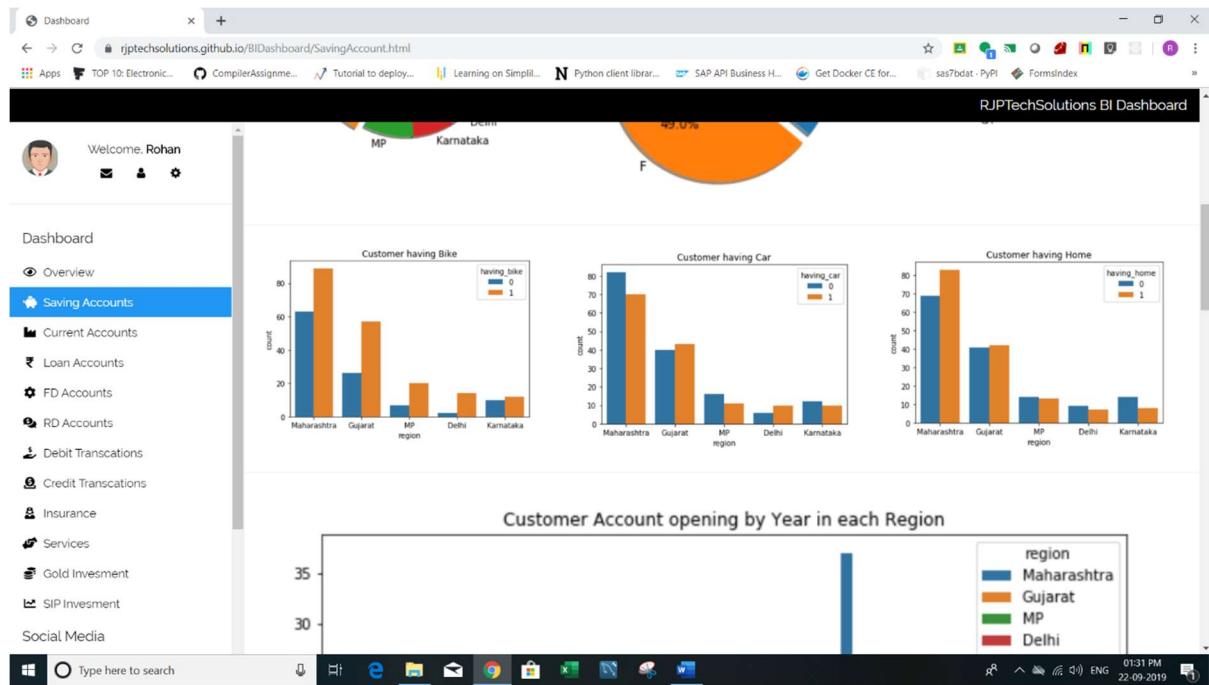
Type here to search

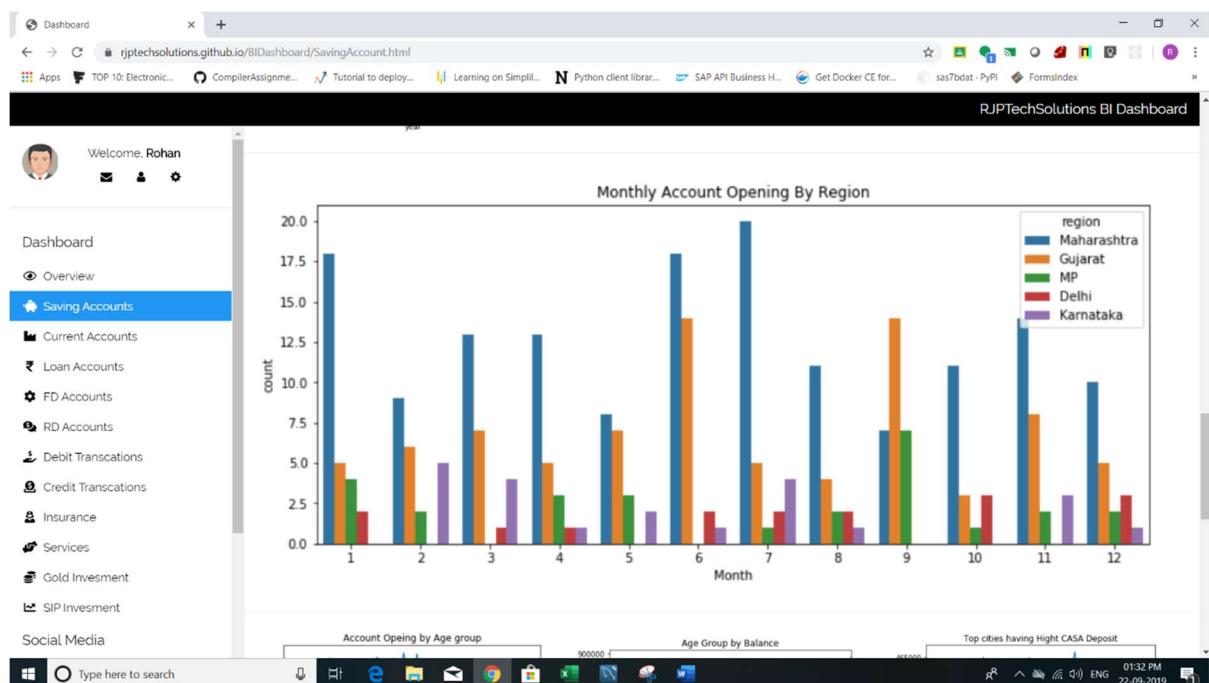
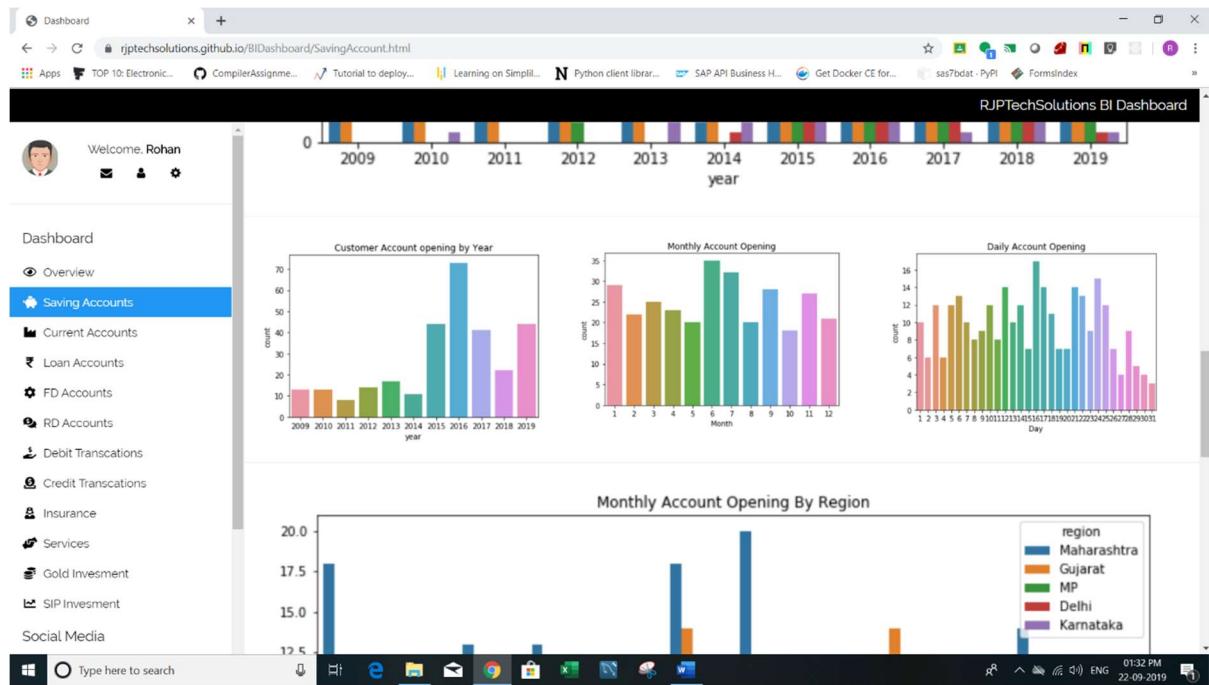
01:31 PM 22-09-2019

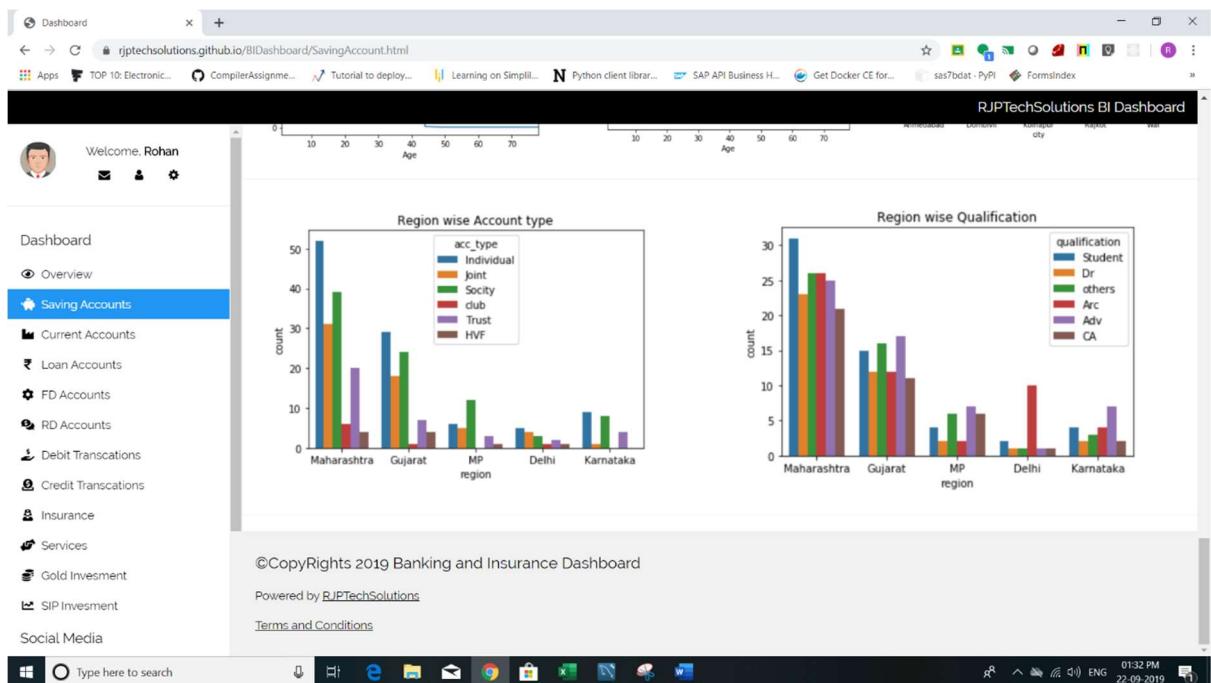
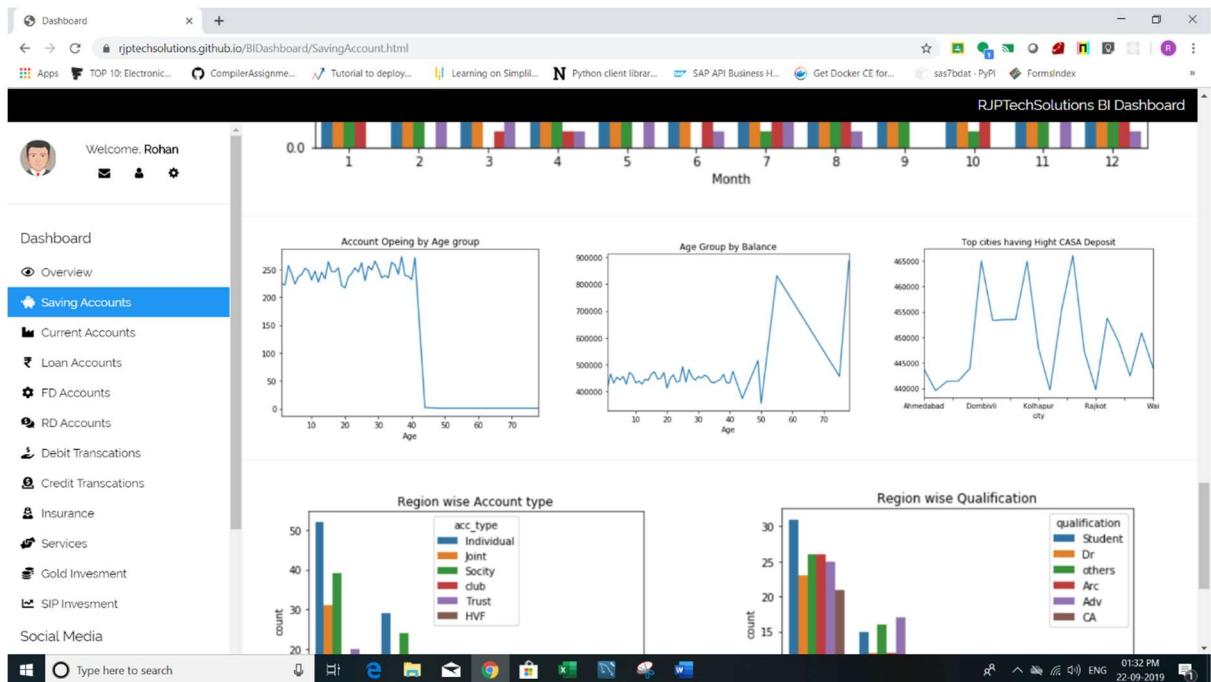


## Saving Account:

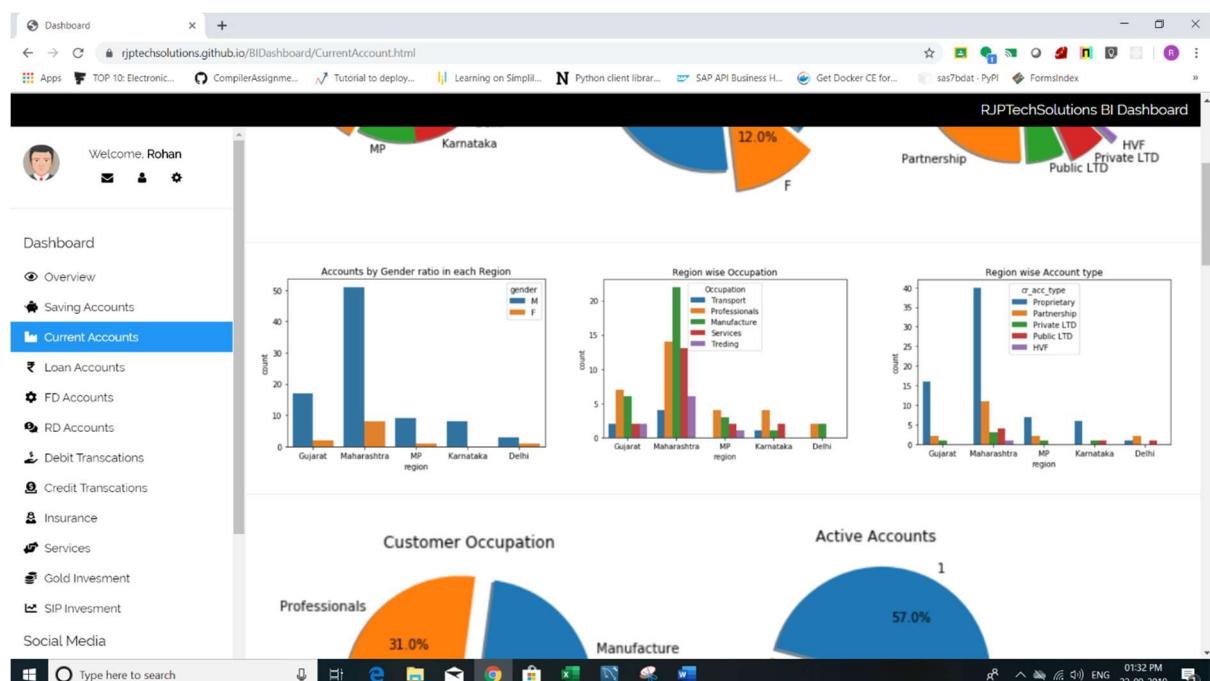
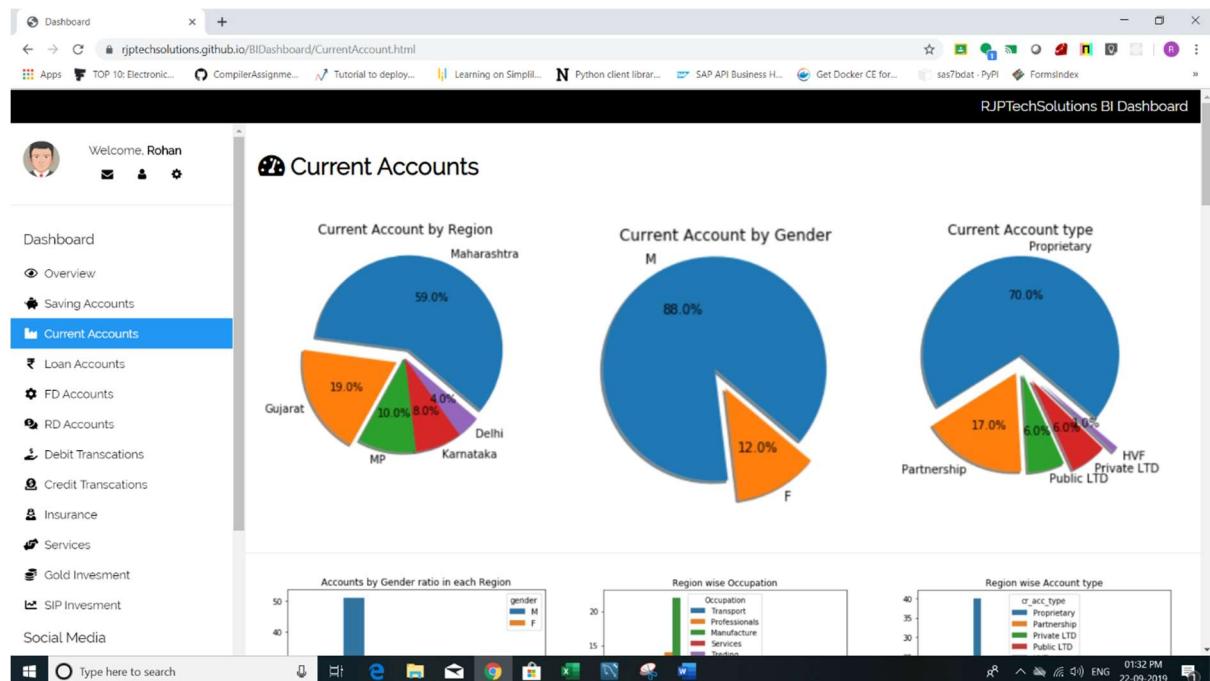


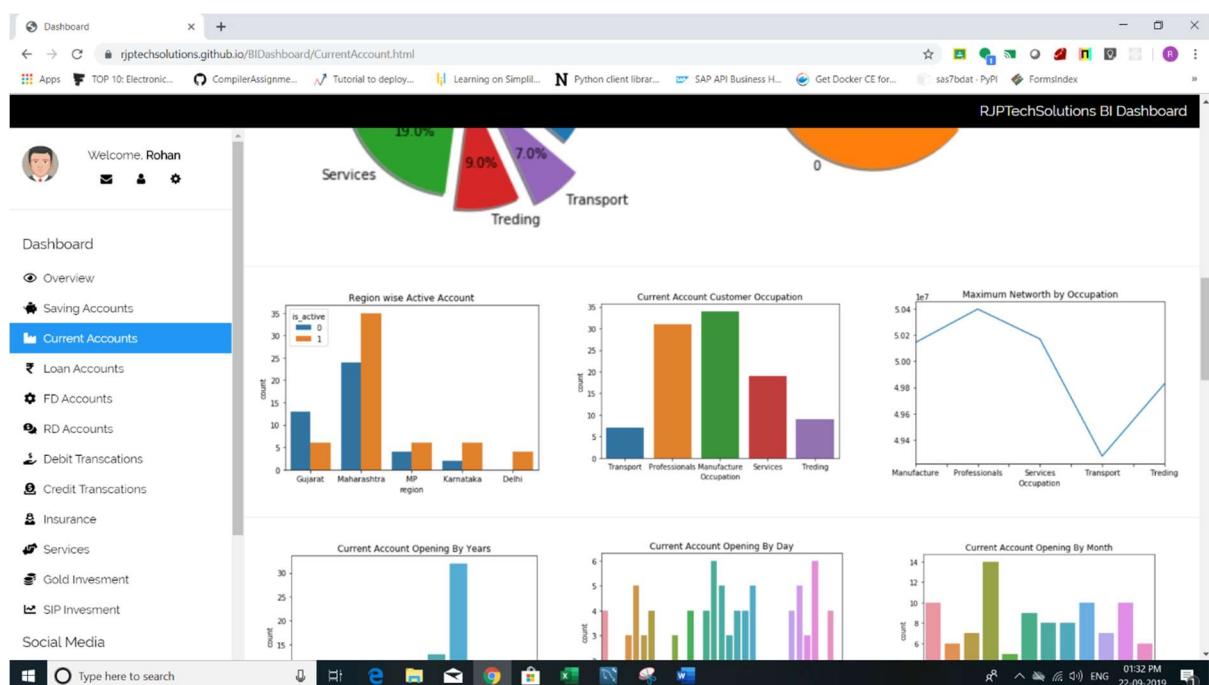
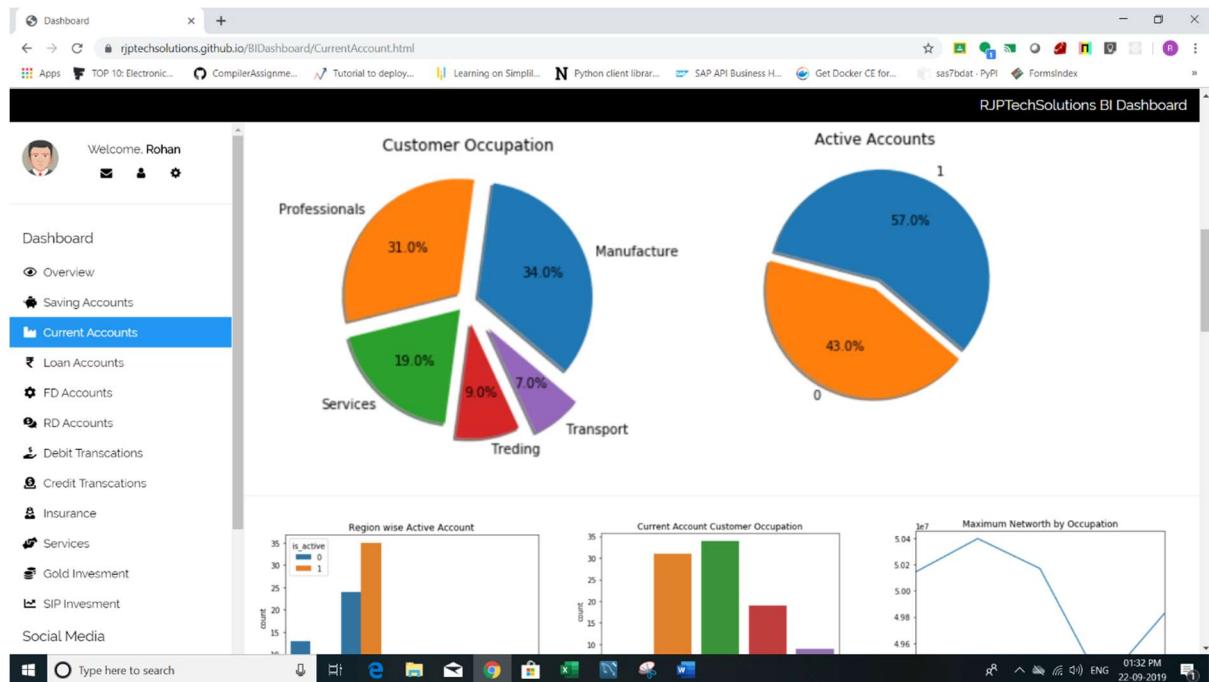


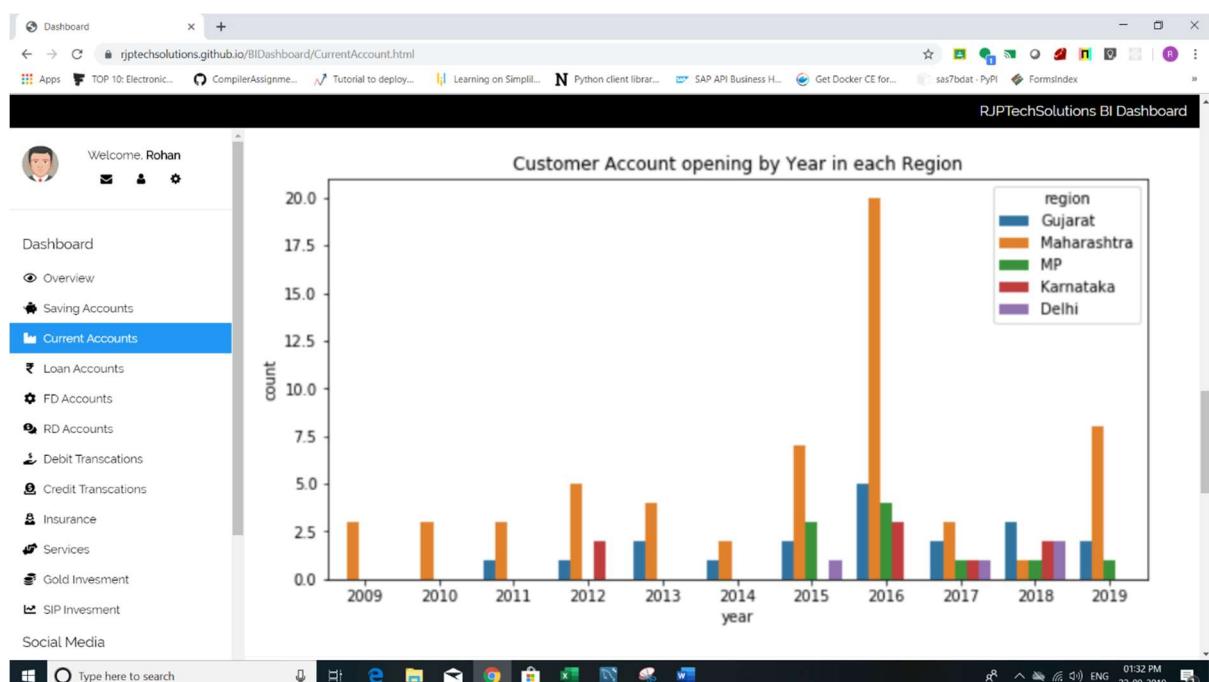
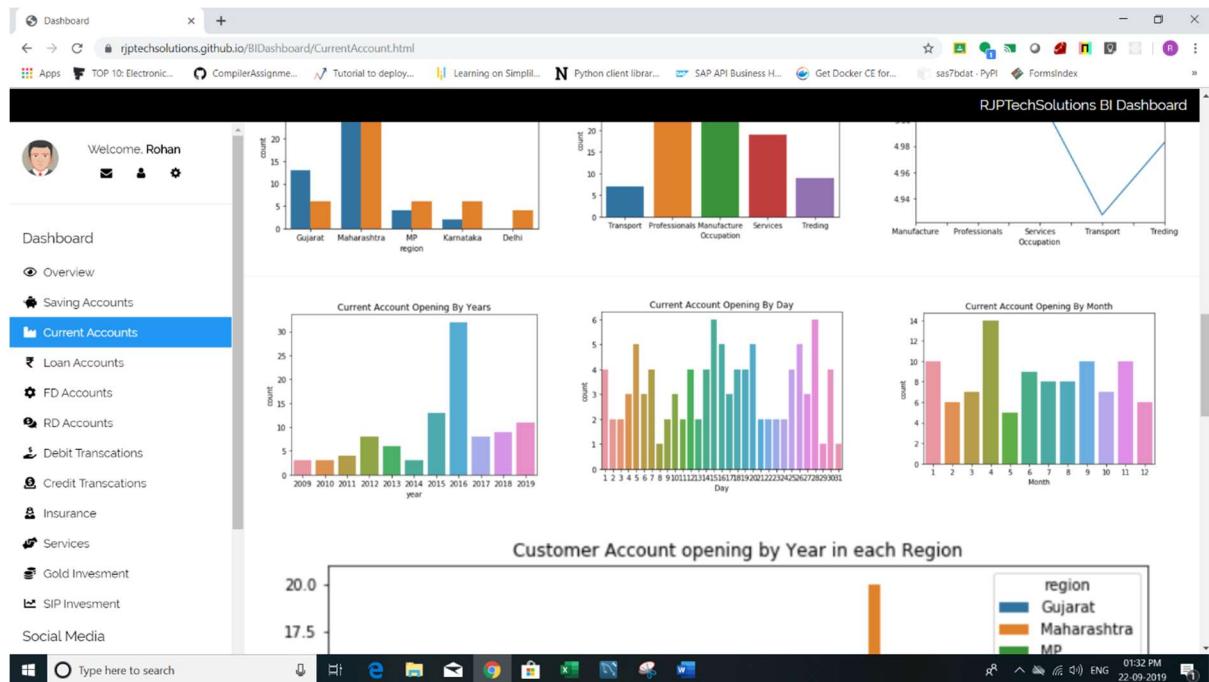


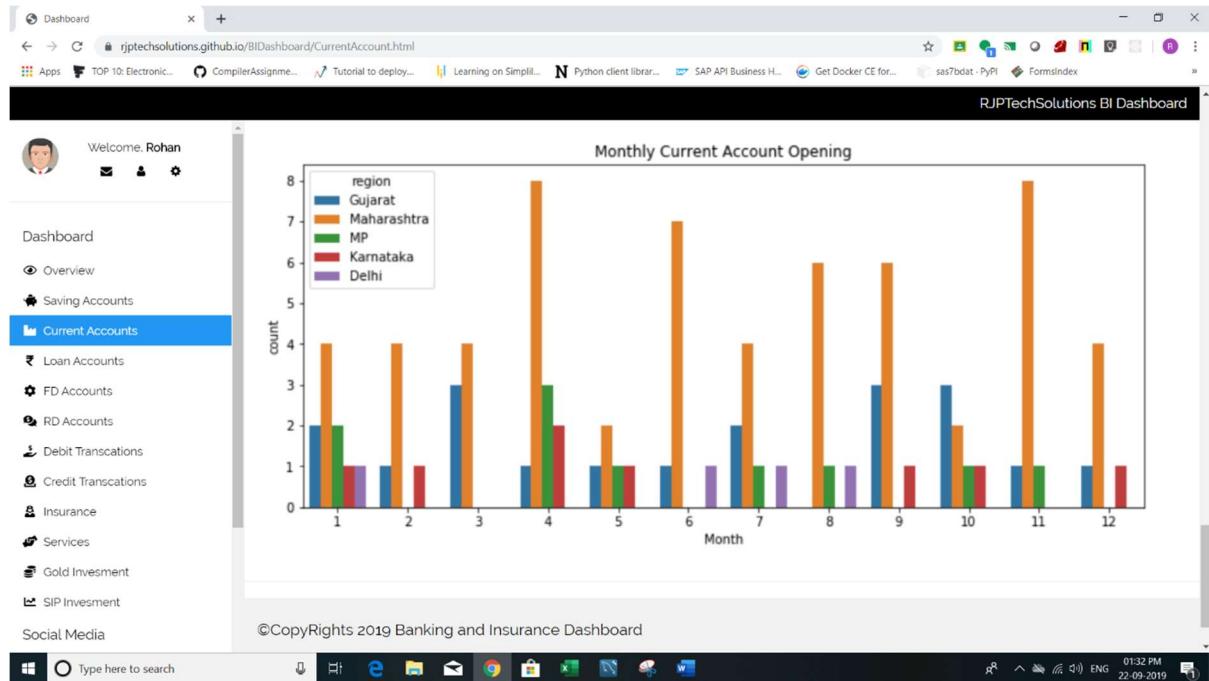
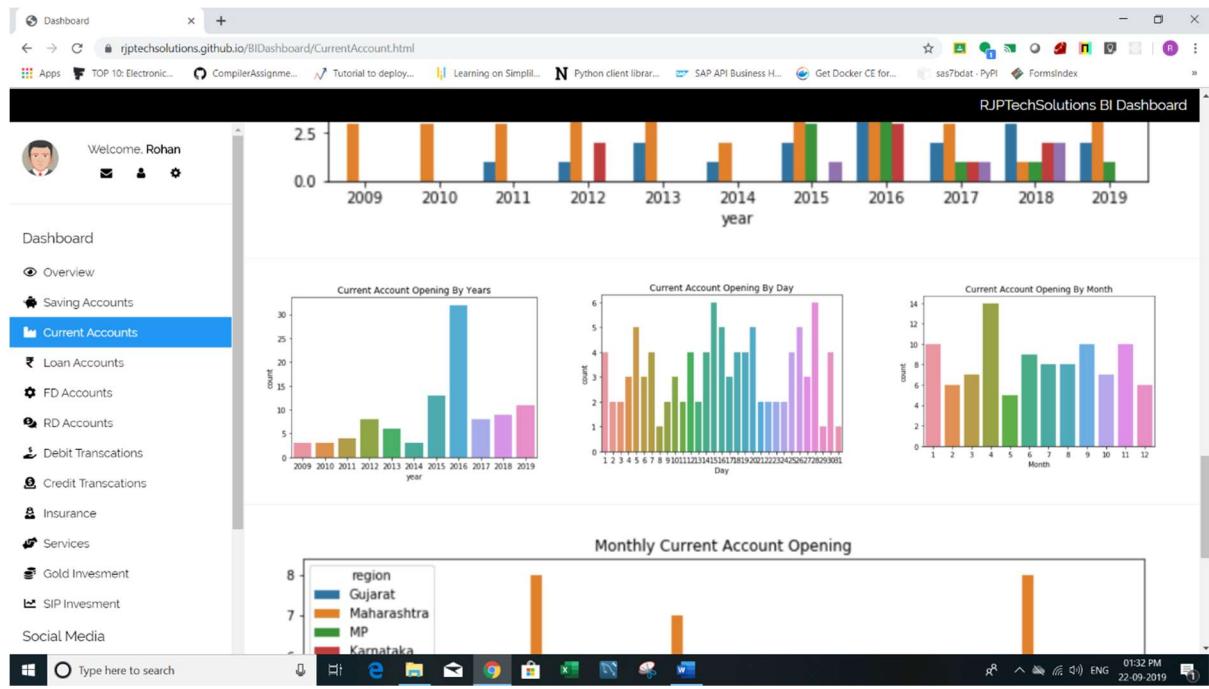


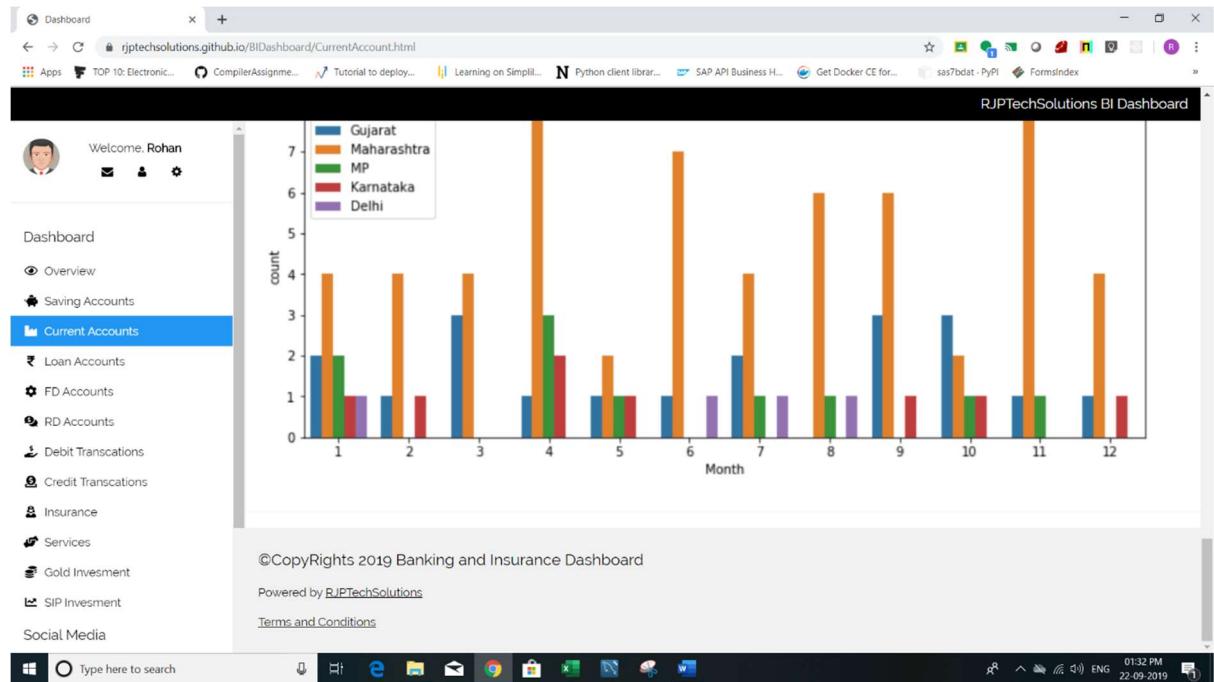
## Current Account:



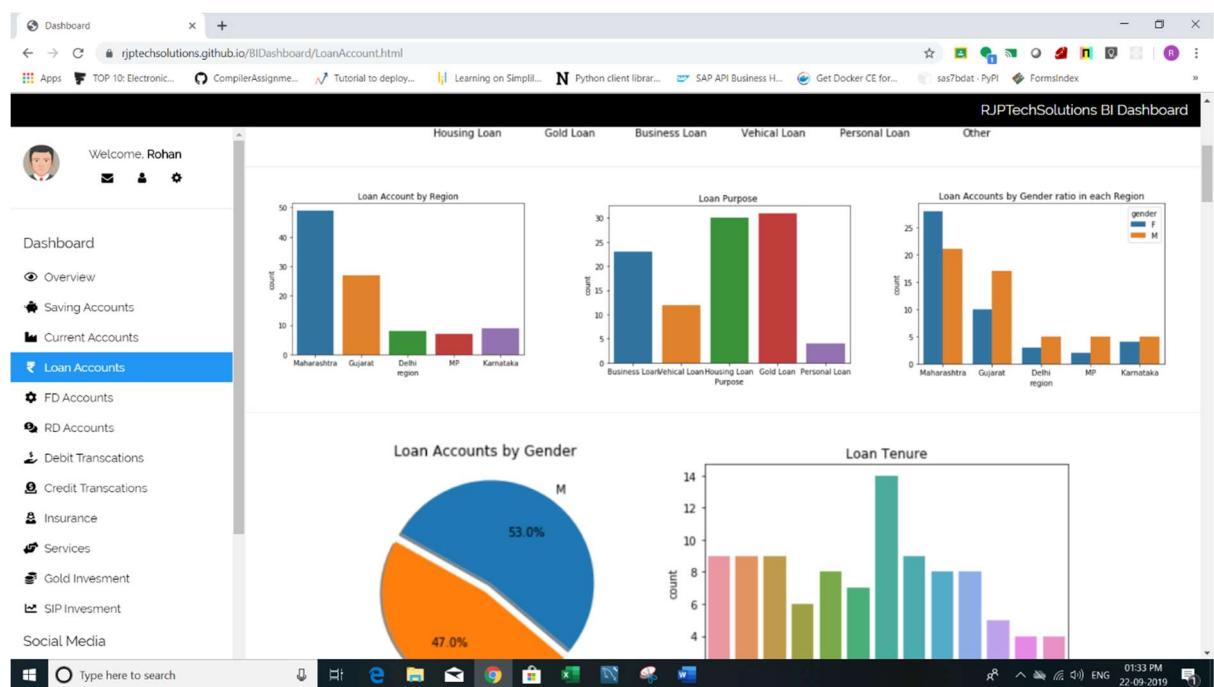
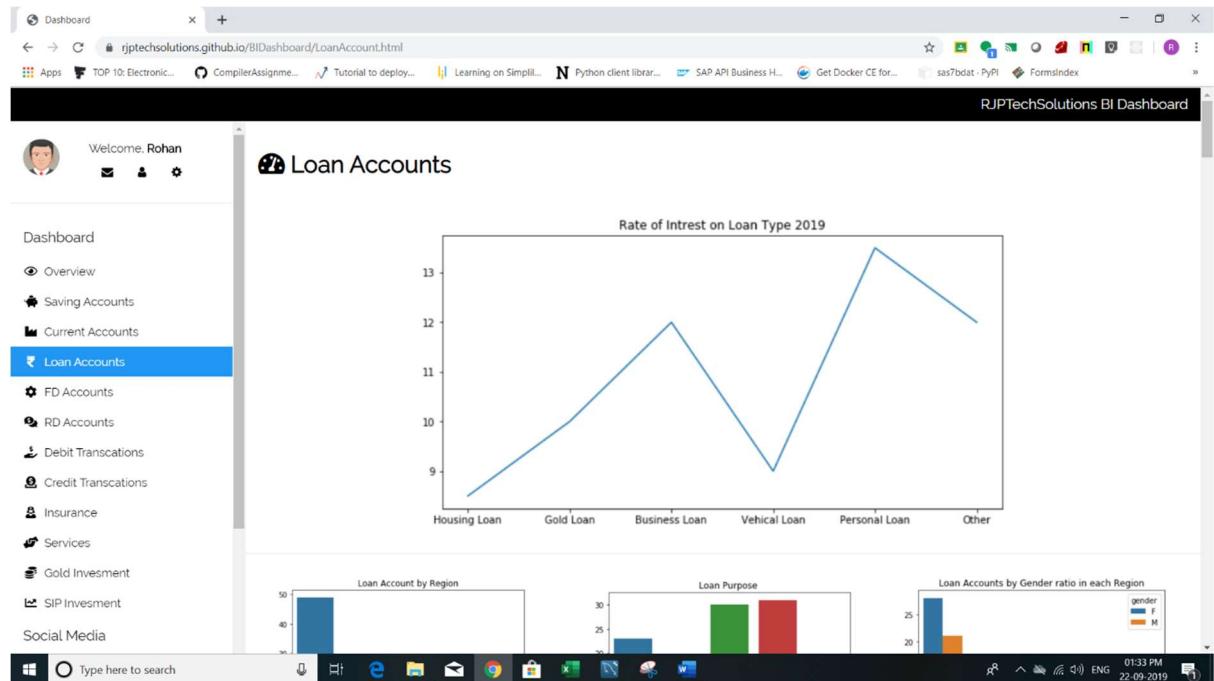


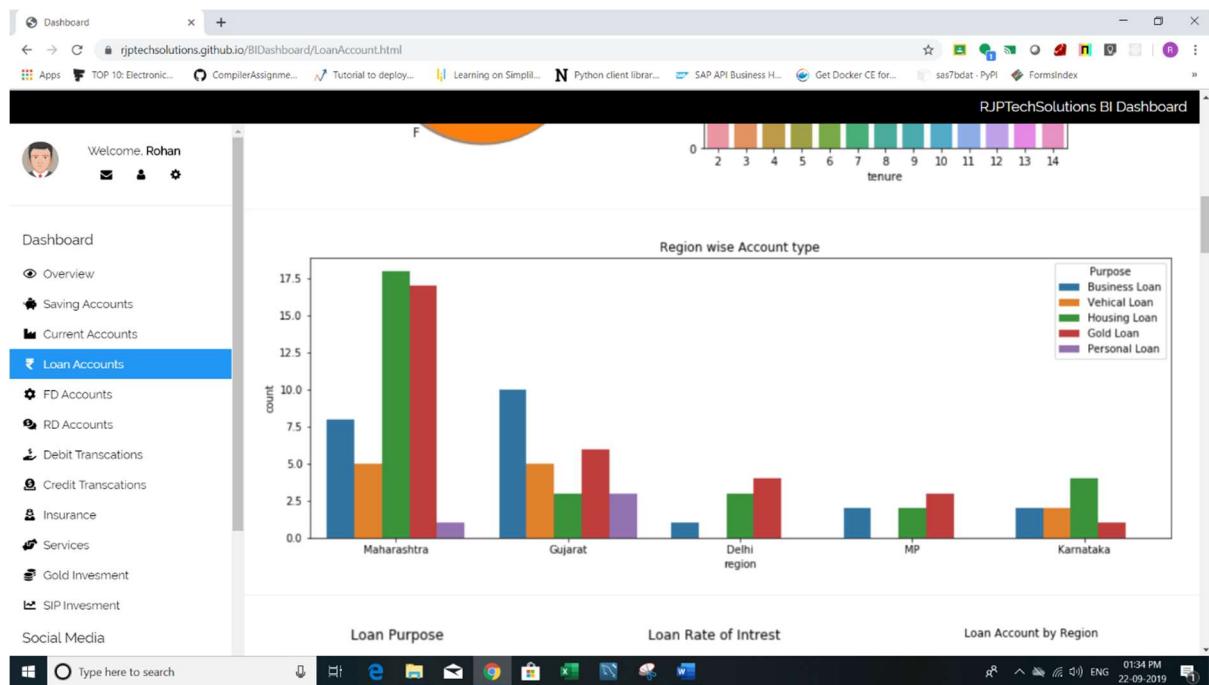
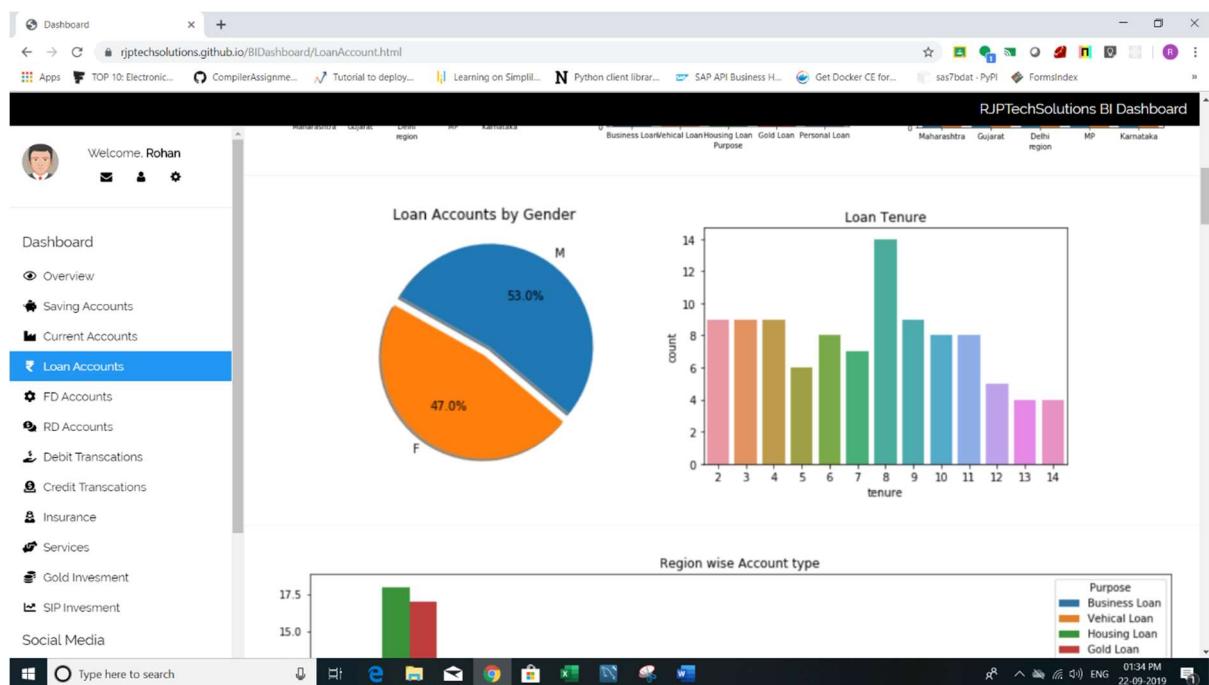


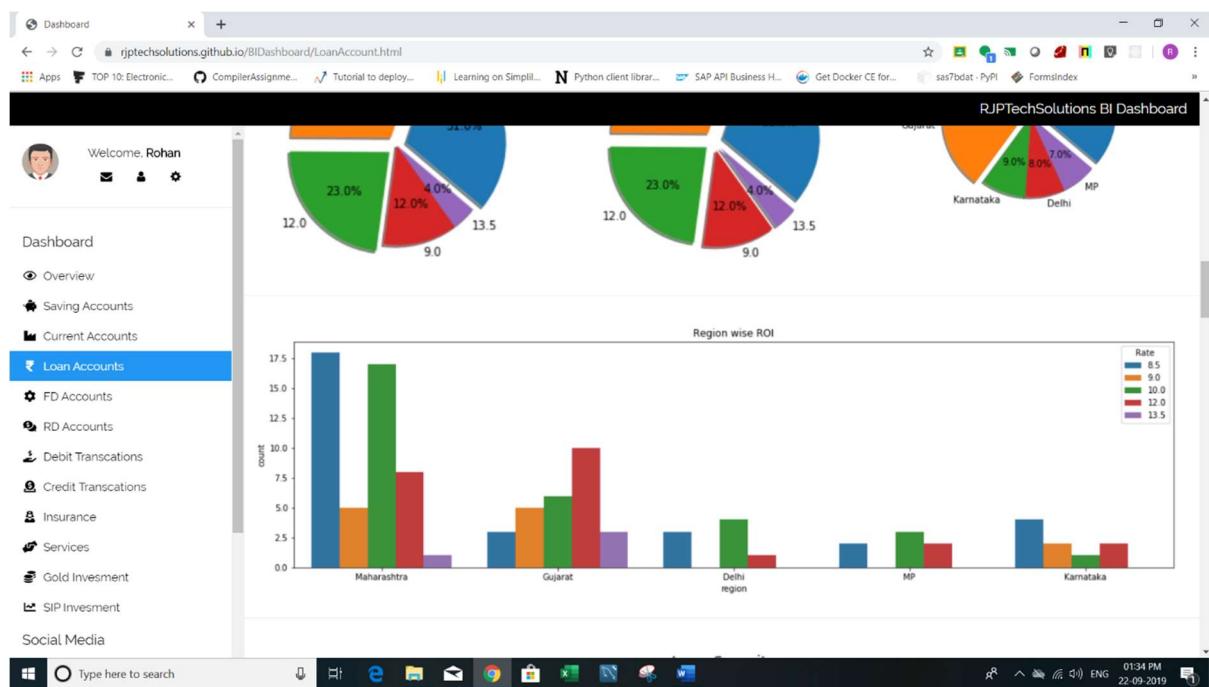
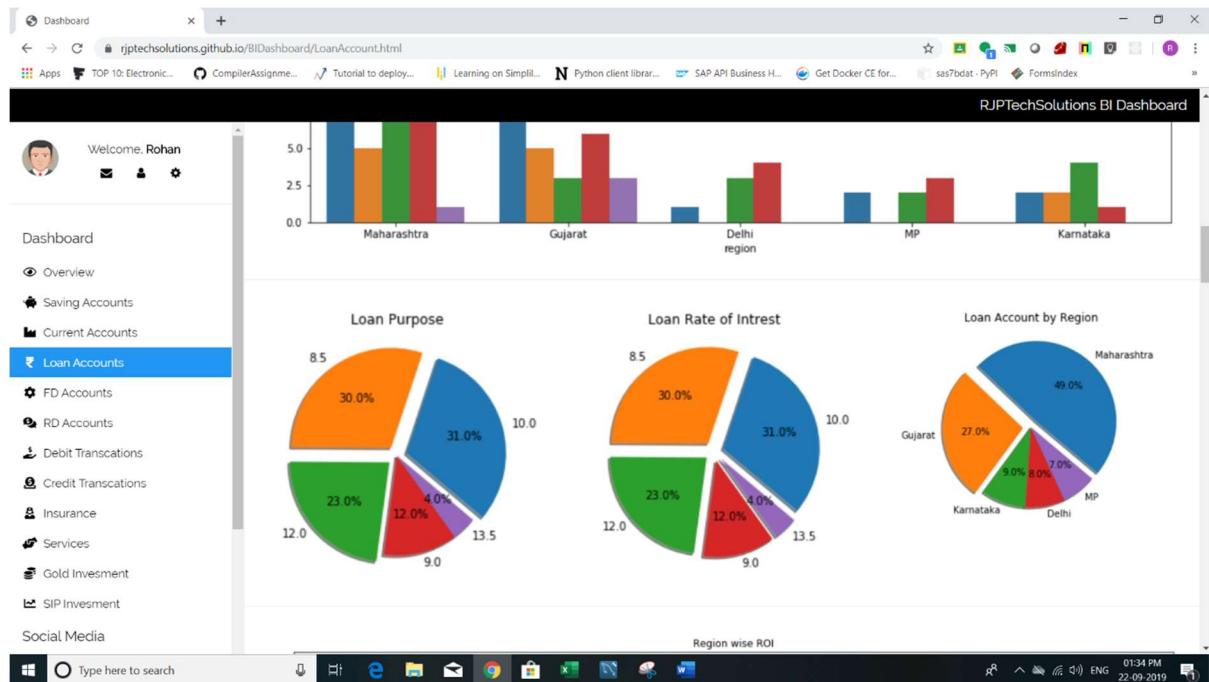


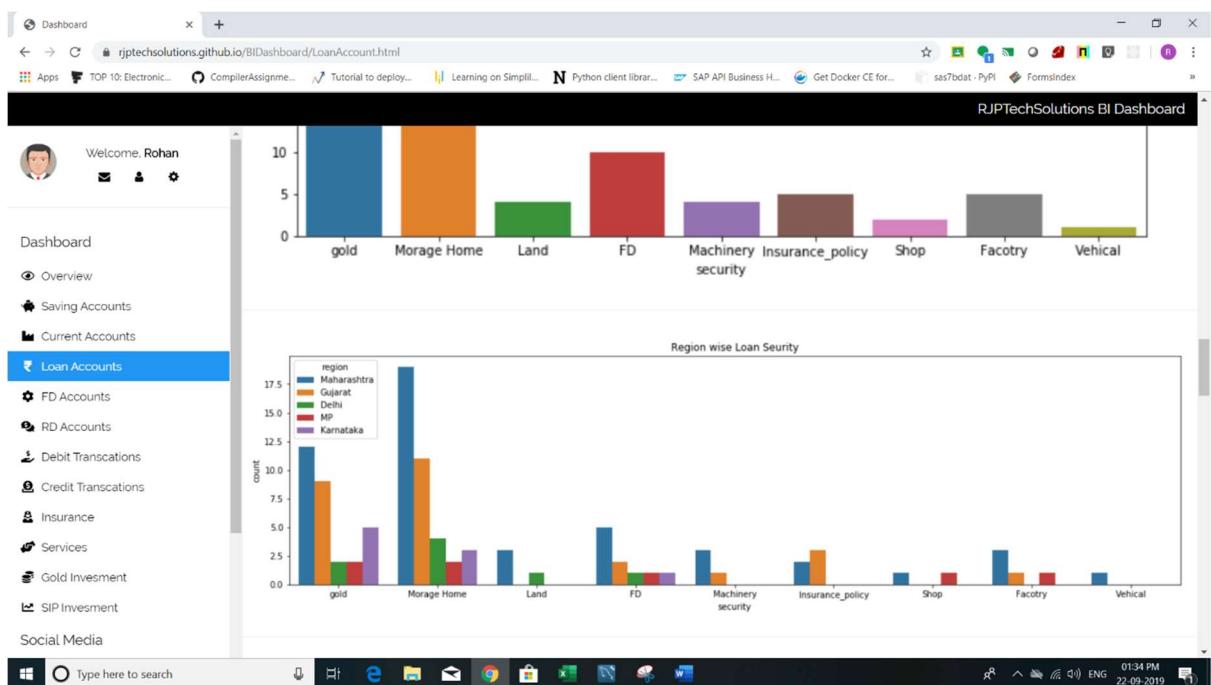
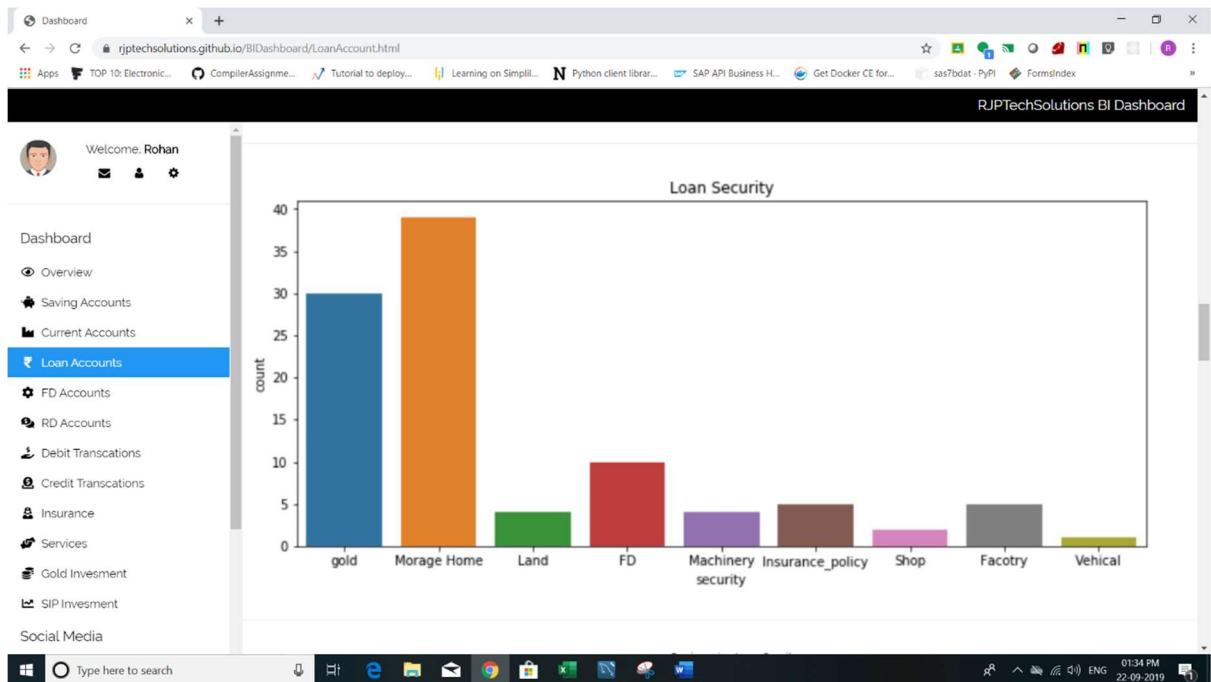


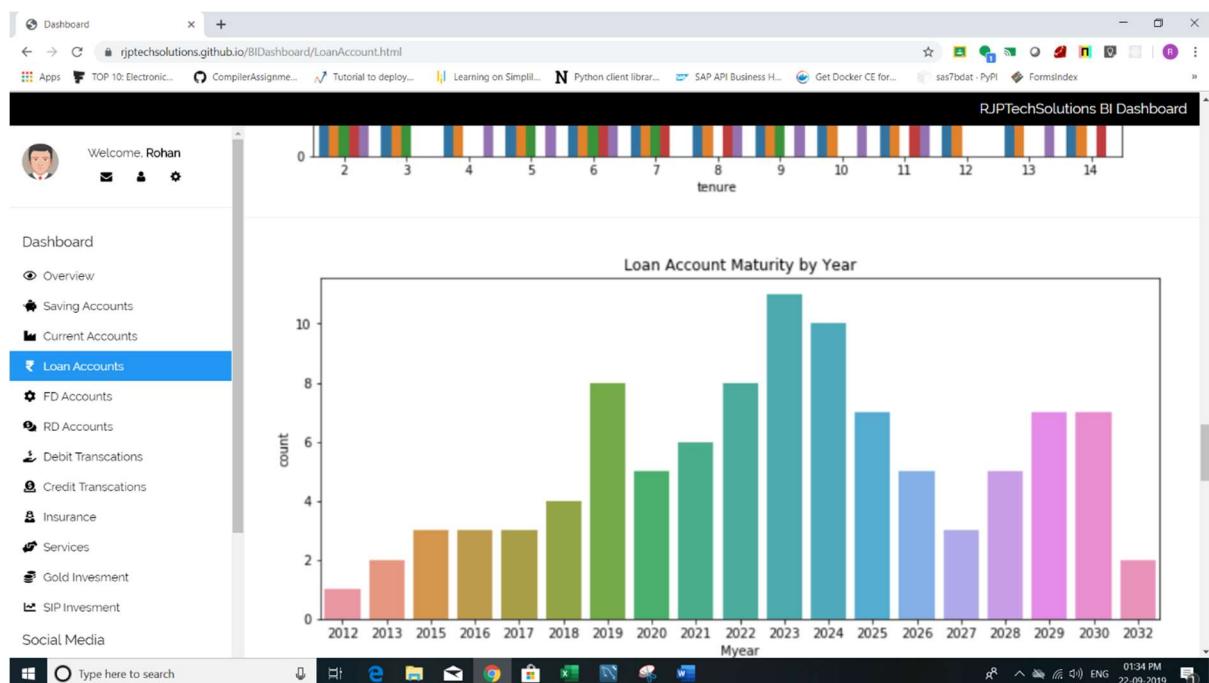
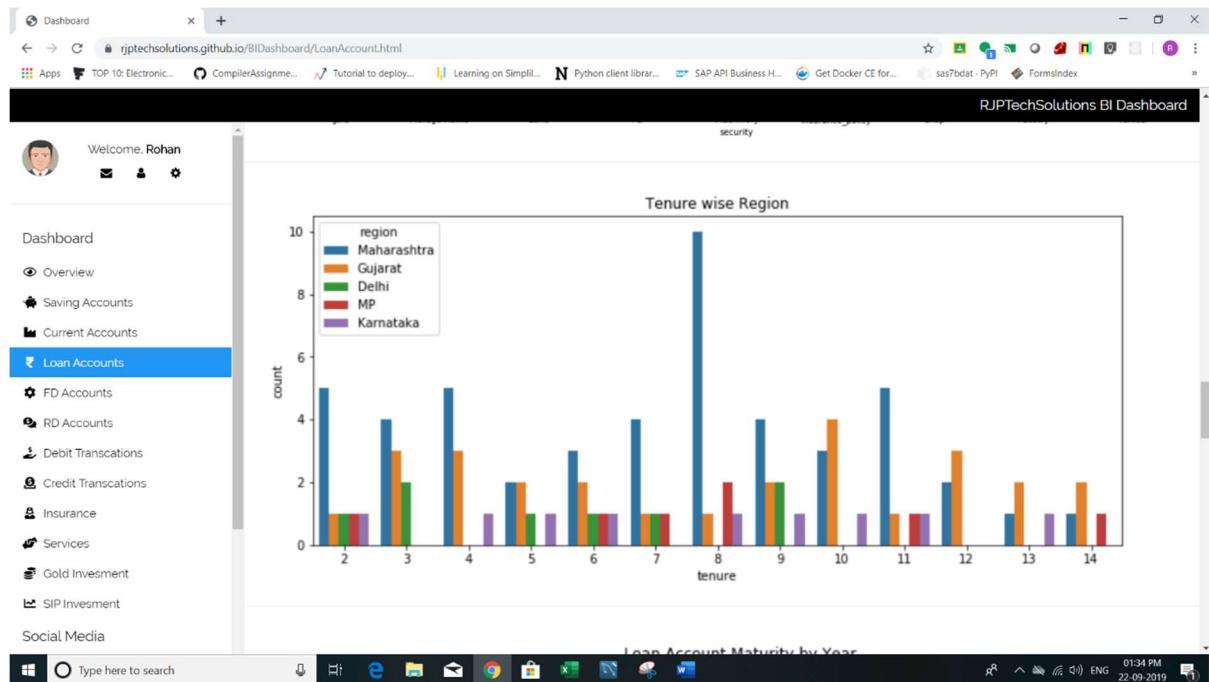
## Loan Account:

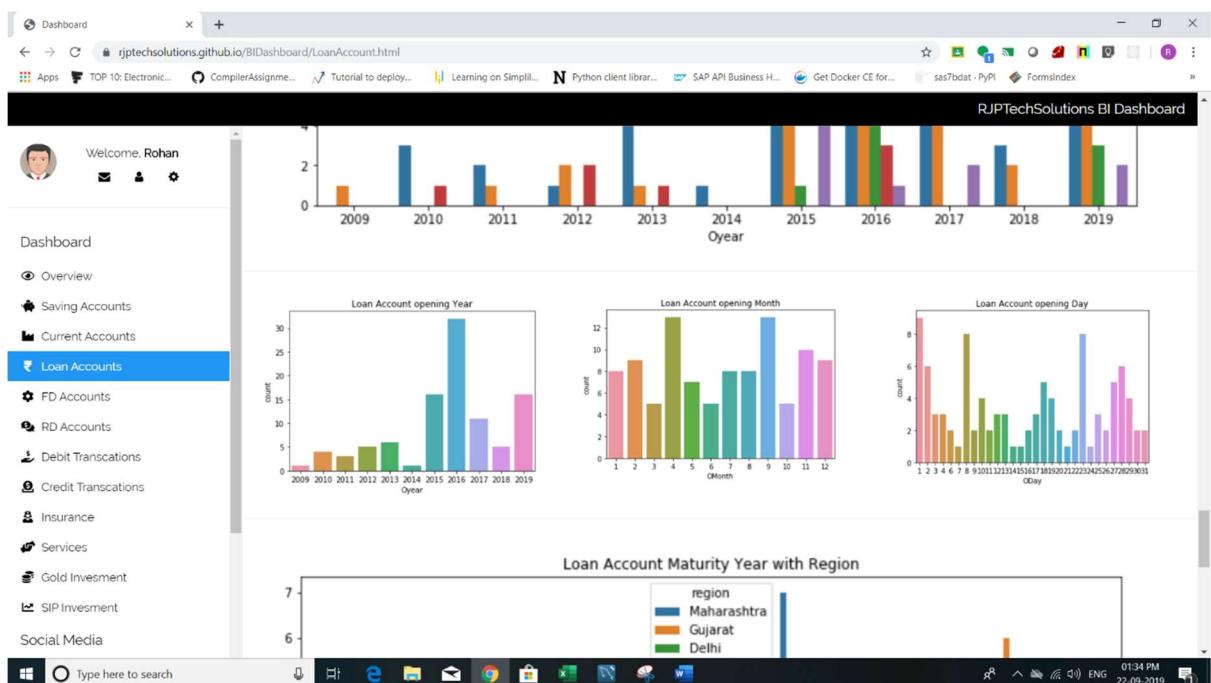
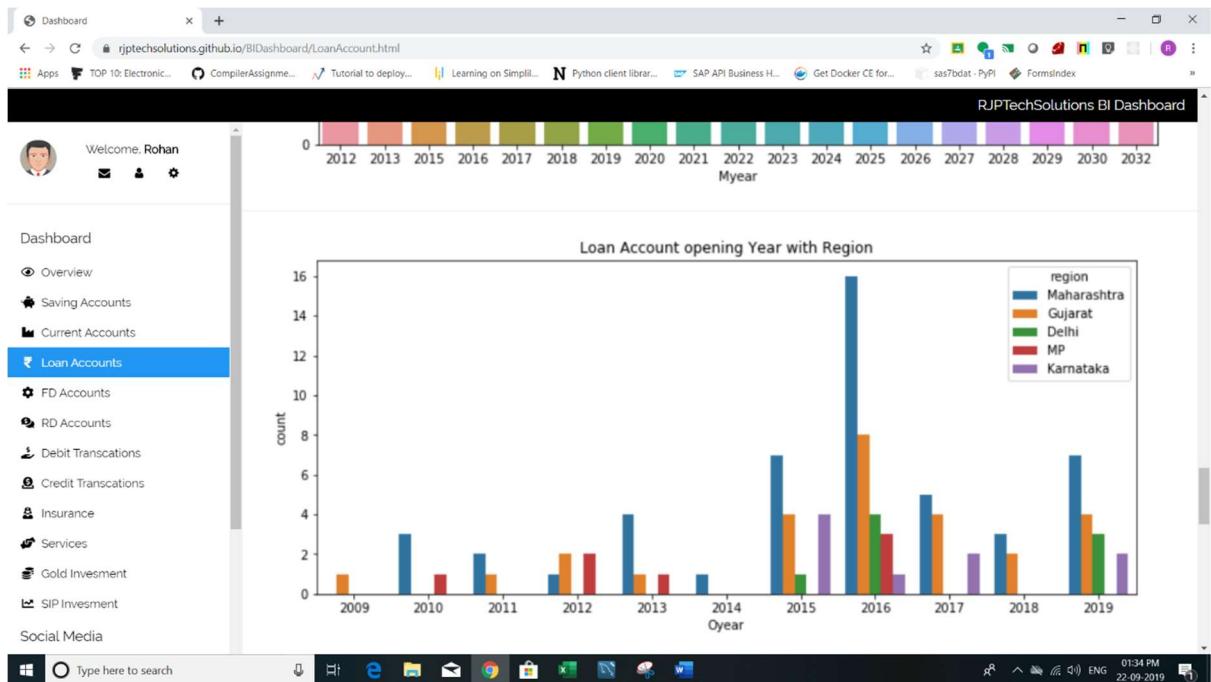


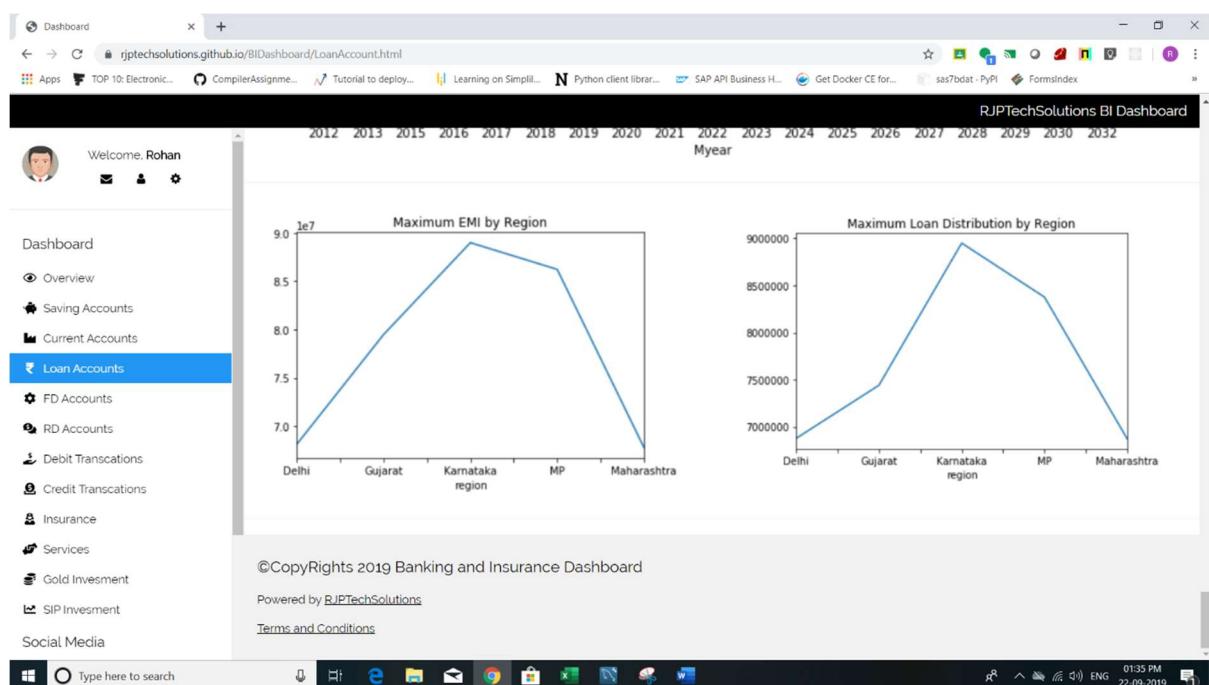




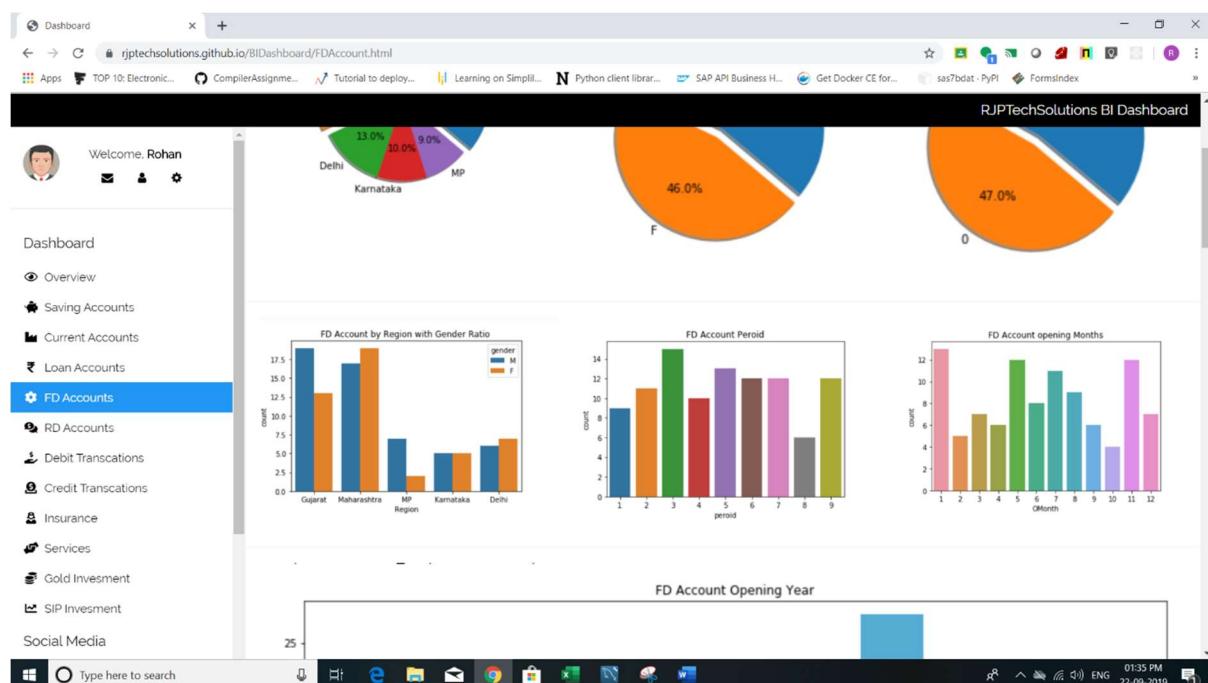
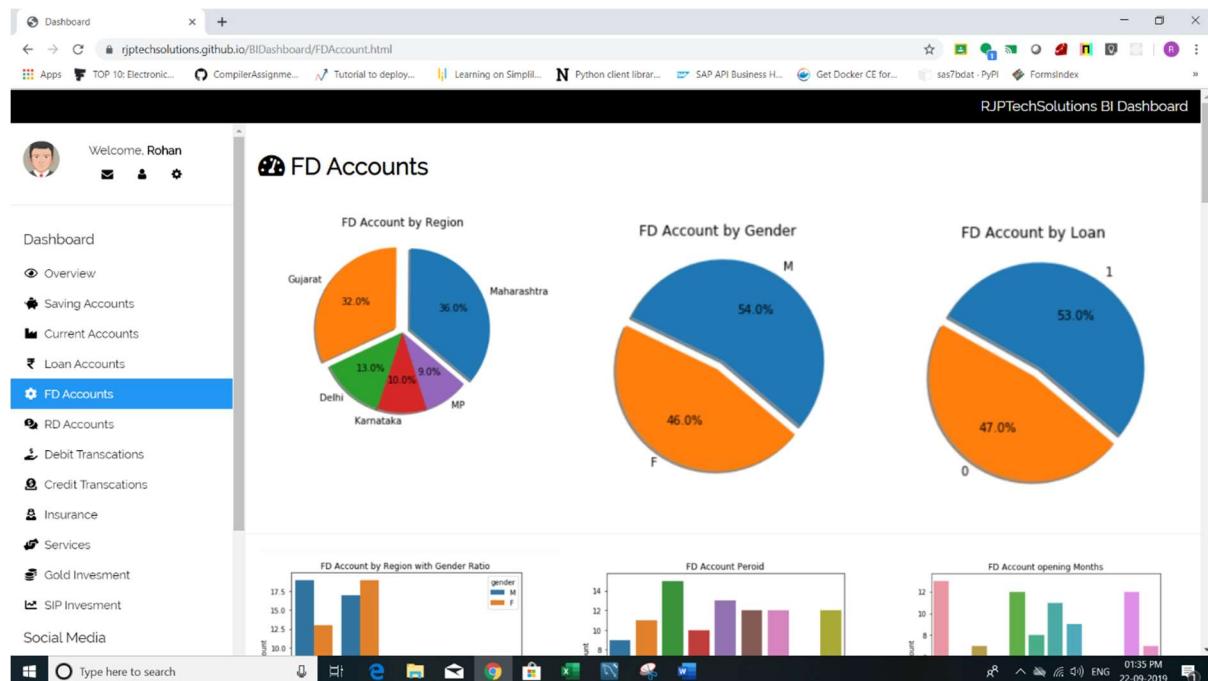


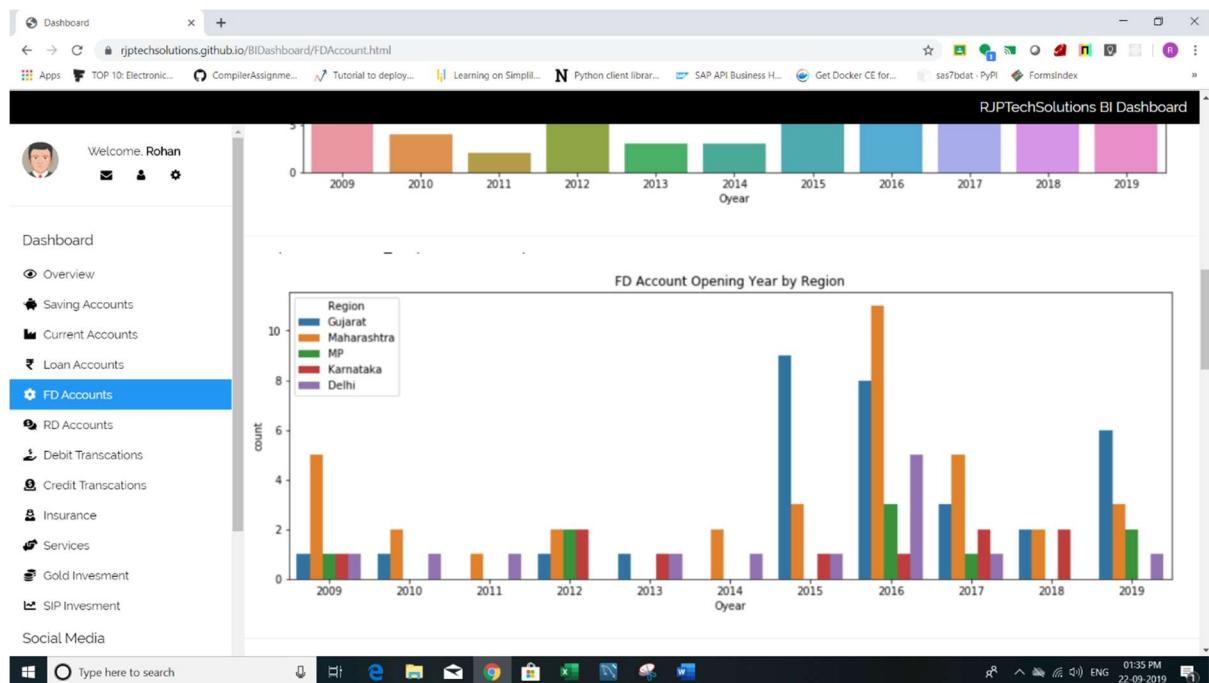
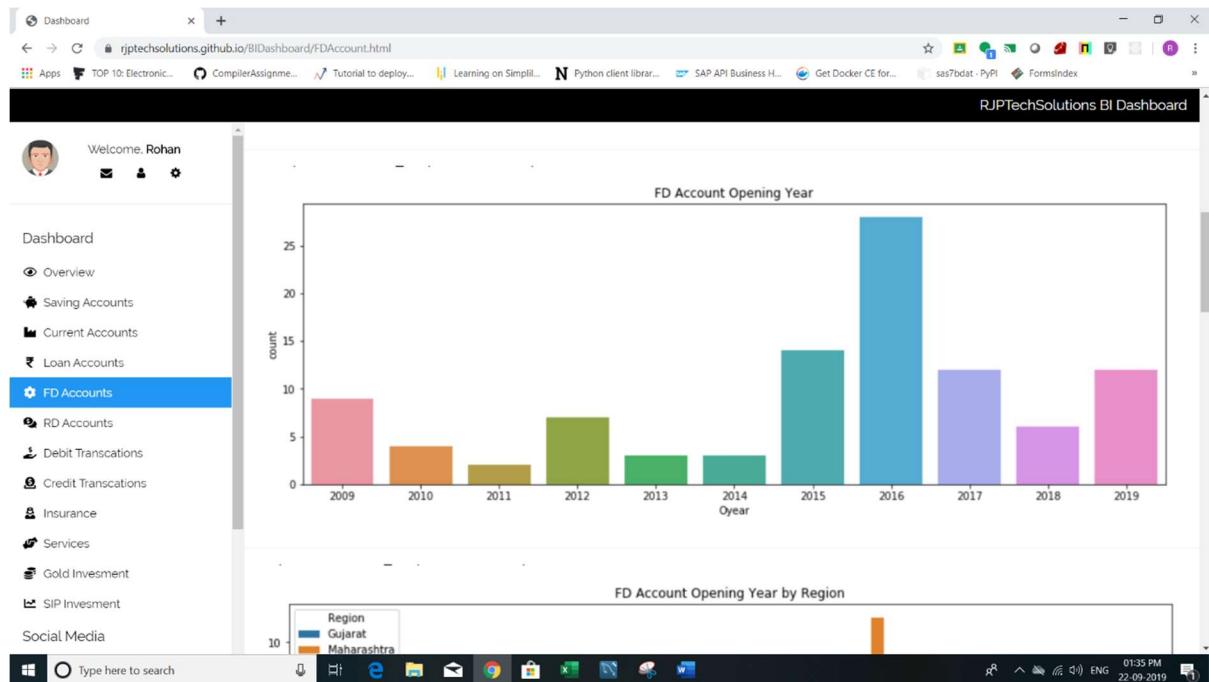


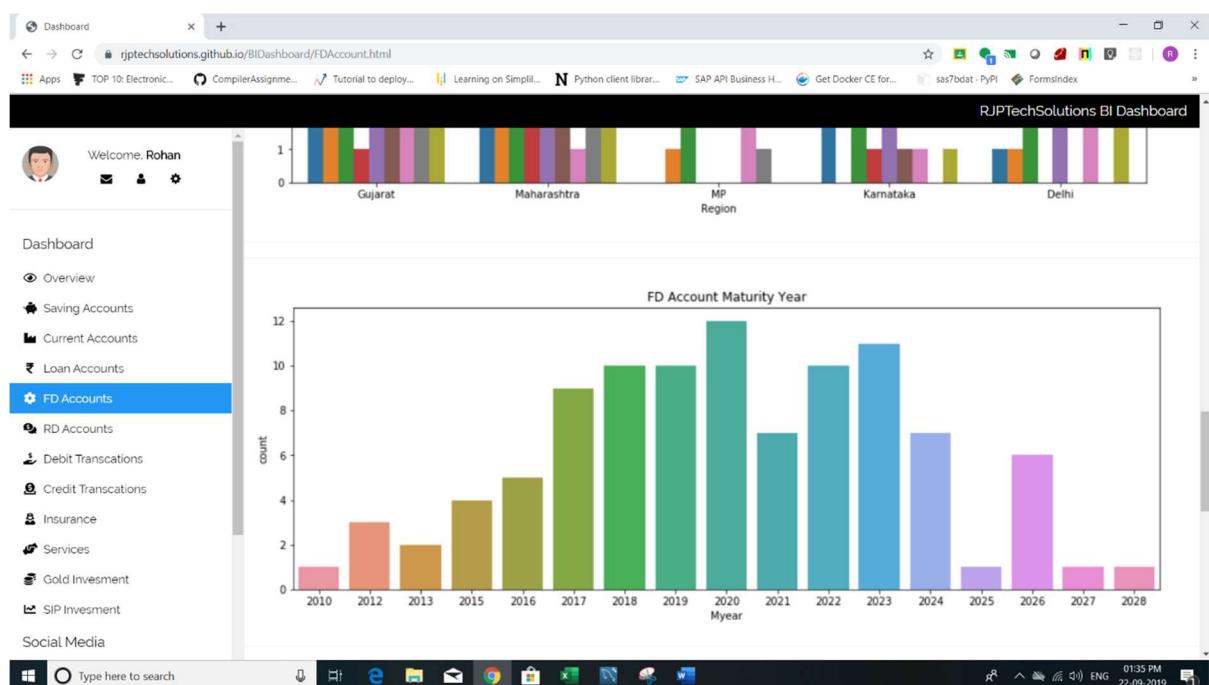
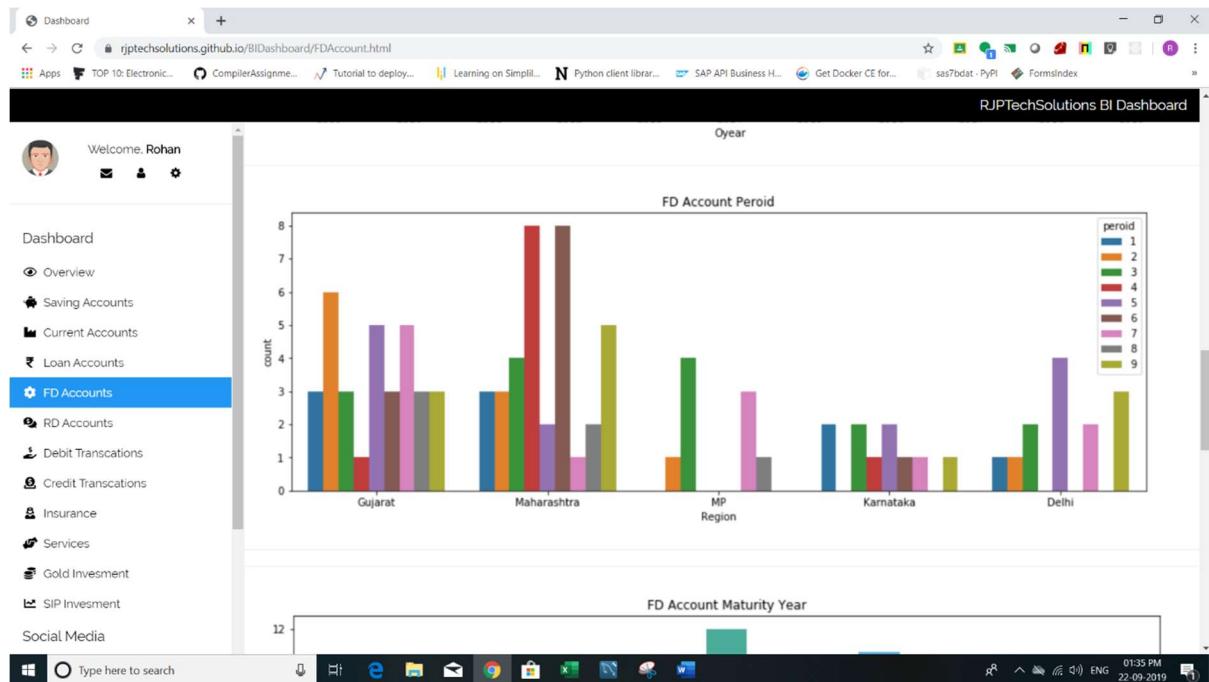




## **FD Accounts:**







RJPTechSolutions BI Dashboard

Welcome, Rohan

- Dashboard
- Overview
- Saving Accounts
- Current Accounts
- Loan Accounts
- FD Accounts**
- RD Accounts
- Debit Transactions
- Credit Transactions
- Insurance
- Services
- Gold Investment
- SIP Investment
- Social Media

FD Account Maturity Year By Region

Myear	Gujarat	Maharashtra	MP	Karnataka	Delhi
2010	1	0	0	0	0
2012	0	1	1	0	0
2013	0	1	0	0	1
2015	0	2	1	0	1
2016	2	3	0	0	0
2017	3	3	0	1	2
2018	3	3	1	1	2
2019	2	4	2	2	0
2020	3	4	1	2	1
2021	2	2	0	1	1
2022	4	1	0	0	1
2023	6	2	1	1	2
2024	3	3	1	0	0
2025	0	0	0	0	0
2026	2	1	2	0	1
2027	1	0	0	0	0
2028	0	1	0	0	0

FD Account opening Days

Day	Count
1	1
2	3
3	5
4	6
5	7
6	5
7	4
8	5
9	6
10	5
11	4
12	3
13	5
14	5
15	4
16	3
17	5
18	4
19	3
20	2
21	3
22	2
23	3
24	2
25	1
26	2
27	1
28	2
29	1
30	1
31	1

Maximum Amt of FD Deposit by Region

Region	Max Amt (e7)
Delhi	4.6
Gujarat	4.5
Karnataka Region	4.4
MP	4.5
Maharashtra	4.6

Rate of Interest on FD

Year	Rate of Interest
2009	7.5
2010	8.0
2011	8.5
2012	9.0
2013	7.5
2014	9.5
2015	10.0
2016	9.0
2017	8.5
2018	8.0
2019	7.5

RJPTechSolutions BI Dashboard

Welcome, Rohan

- Dashboard
- Overview
- Saving Accounts
- Current Accounts
- Loan Accounts
- FD Accounts**
- RD Accounts
- Debit Transactions
- Credit Transactions
- Insurance
- Services
- Gold Investment
- SIP Investment
- Social Media

FD Account Maturity Year By Region

Myear	Gujarat	Maharashtra	MP	Karnataka	Delhi
2010	1	0	0	0	0
2012	0	1	1	0	0
2013	0	1	0	0	1
2015	0	2	1	0	1
2016	2	0	0	0	0
2017	3	3	0	1	2
2018	3	3	1	1	2
2019	2	4	2	2	0
2020	3	4	1	2	1
2021	2	2	0	1	1
2022	1	0	0	0	1
2023	0	0	1	1	2
2024	0	0	1	0	0
2025	0	0	0	0	0
2026	2	1	2	0	1
2027	1	0	0	0	0
2028	0	1	0	0	0

FD Account opening Days

Day	Count
1	1
2	3
3	5
4	6
5	7
6	5
7	4
8	5
9	6
10	5
11	4
12	3
13	5
14	5
15	4
16	3
17	5
18	4
19	3
20	2
21	3
22	2
23	3
24	2
25	1
26	2
27	1
28	2
29	1
30	1
31	1

Maximum Amt of FD Deposit by Region

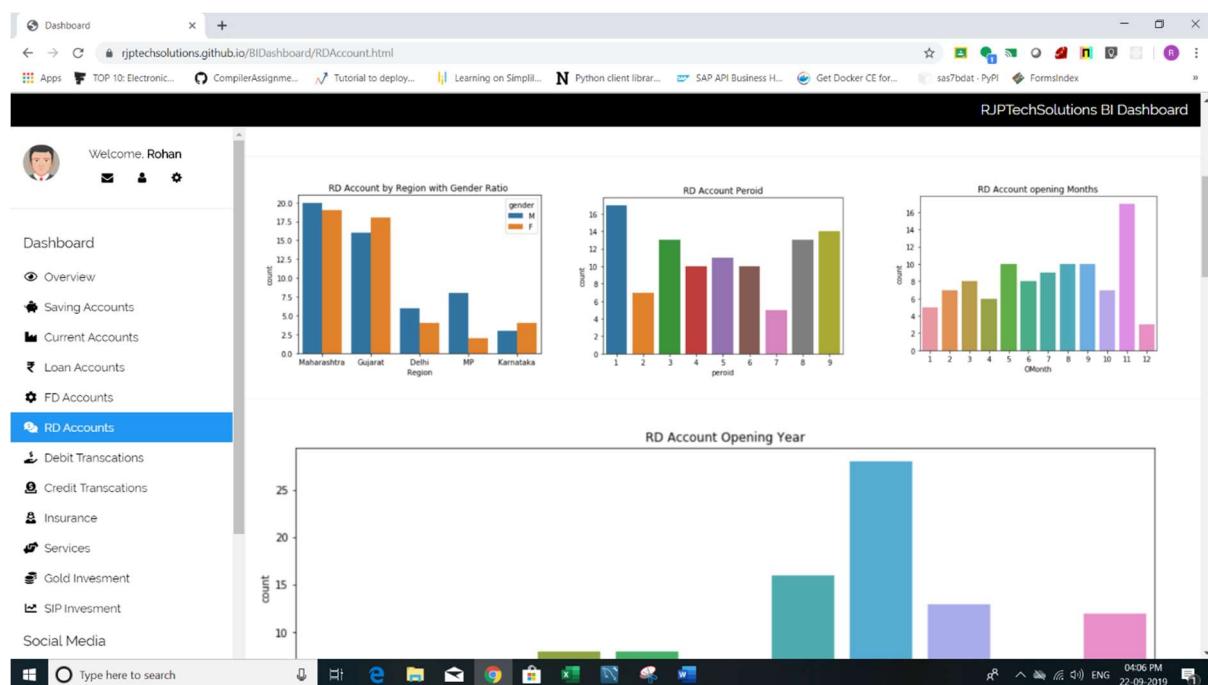
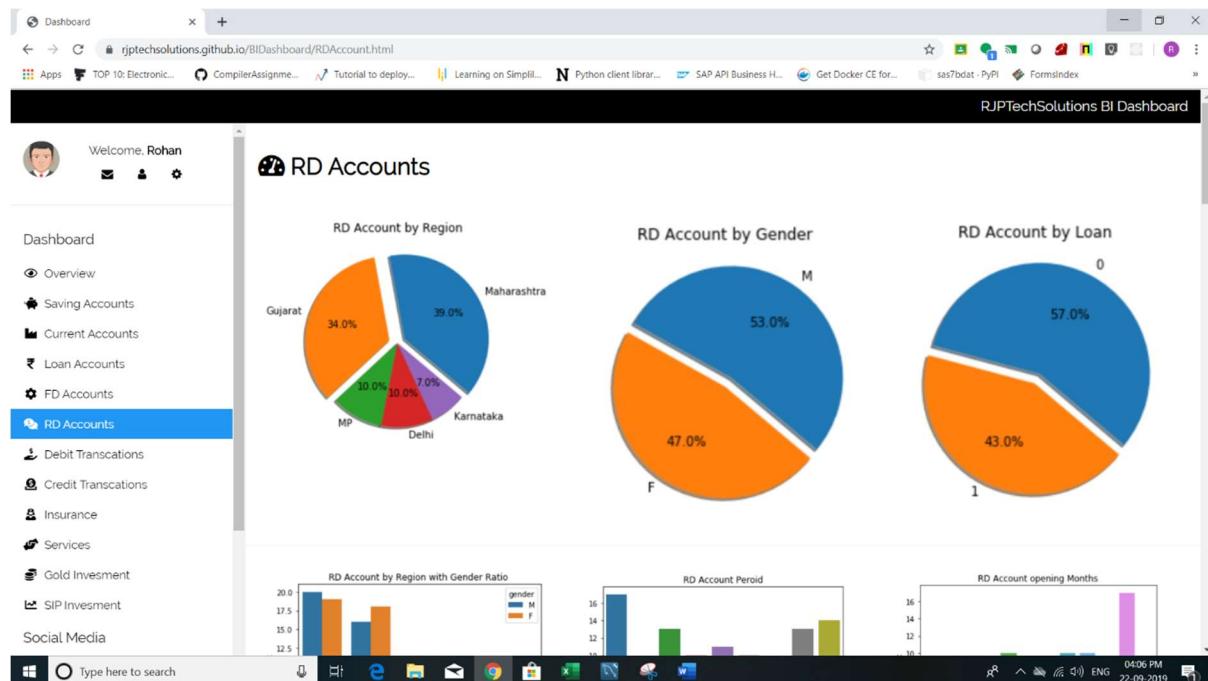
Region	Max Amt (e7)
Delhi	4.6
Gujarat	4.5
Karnataka Region	3.4
MP	4.5
Maharashtra	4.6

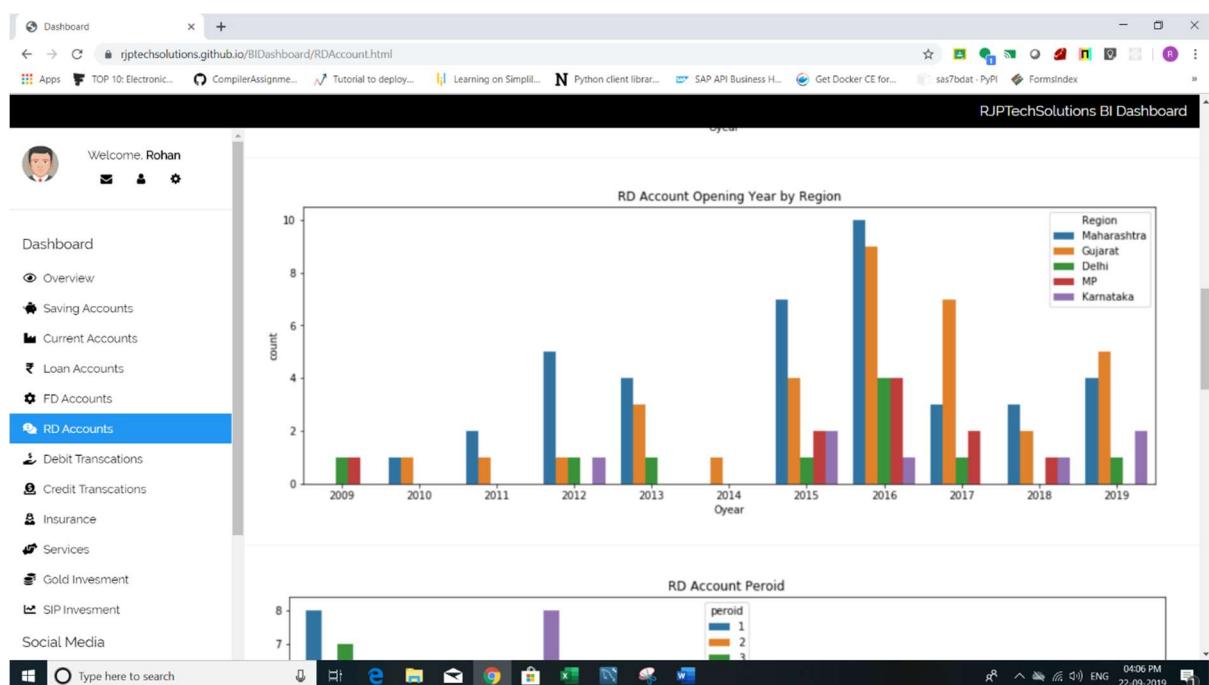
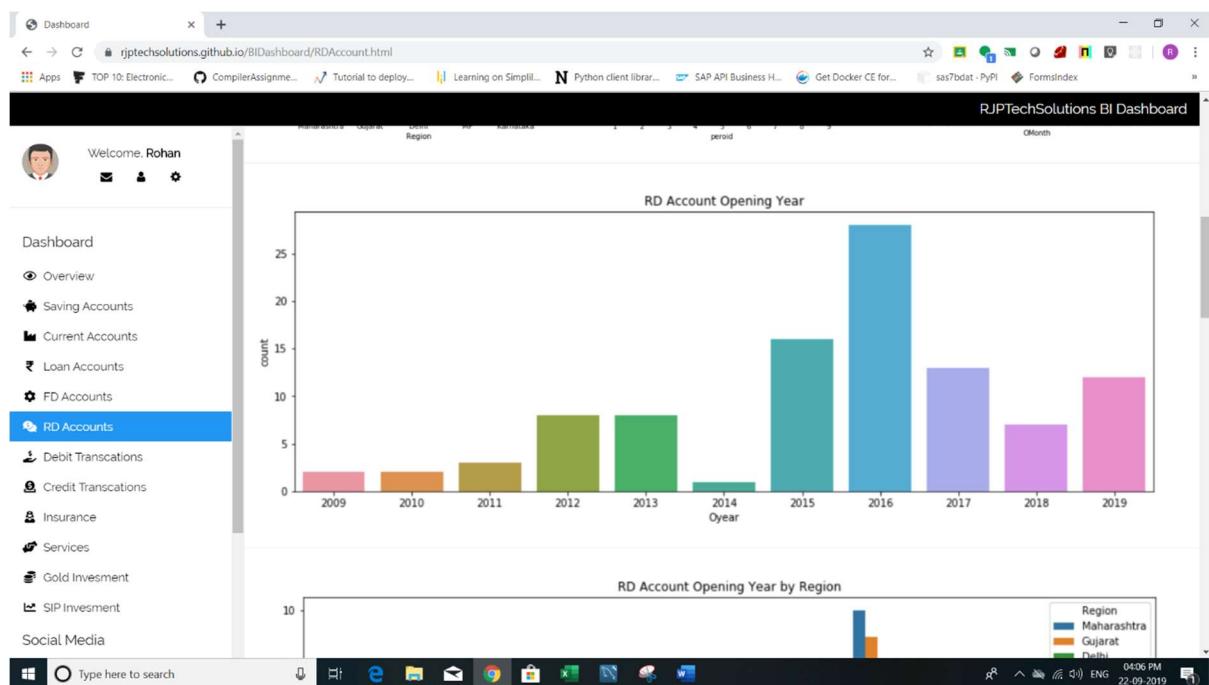
Rate of Interest on FD

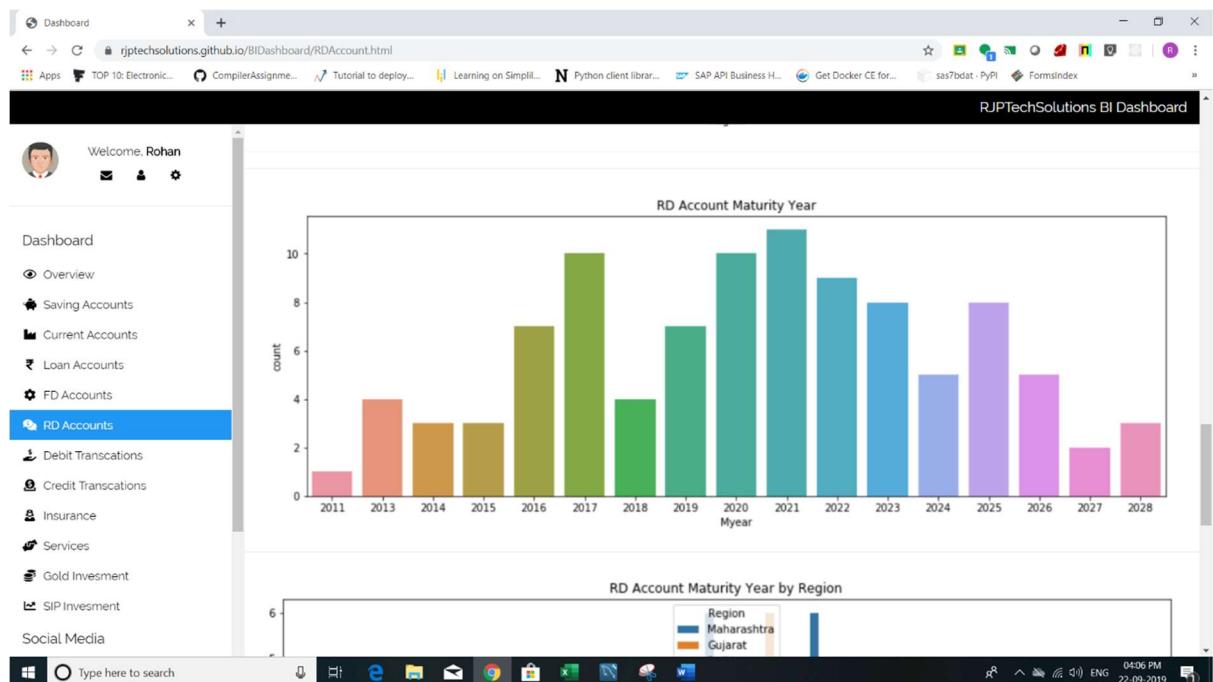
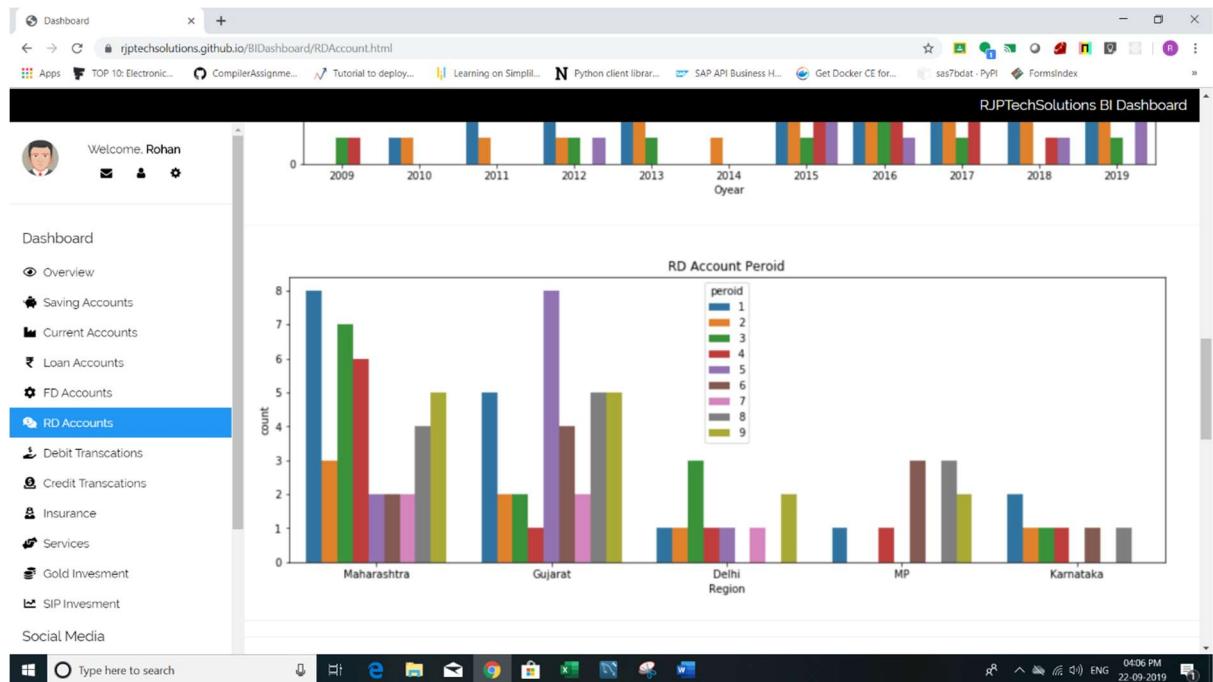
Year	Rate of Interest
2009	7.5
2010	8.0
2011	8.5
2012	9.0
2013	7.5
2014	9.5
2015	10.0
2016	9.0
2017	8.5
2018	8.0
2019	7.5

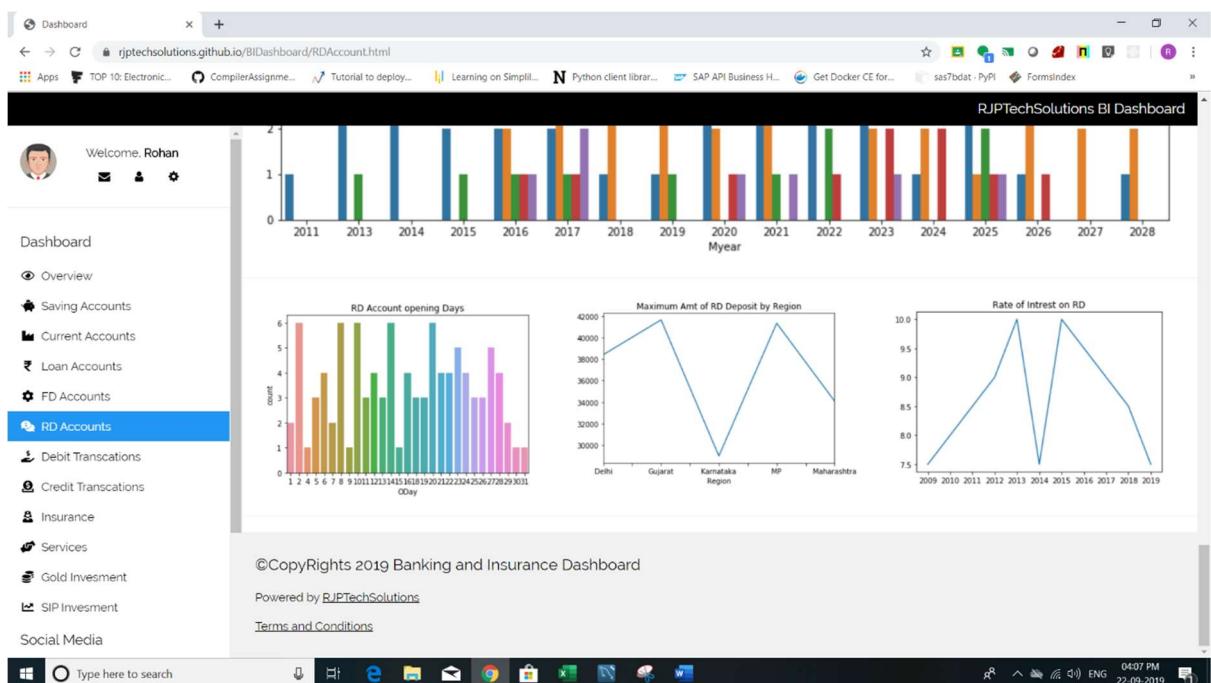
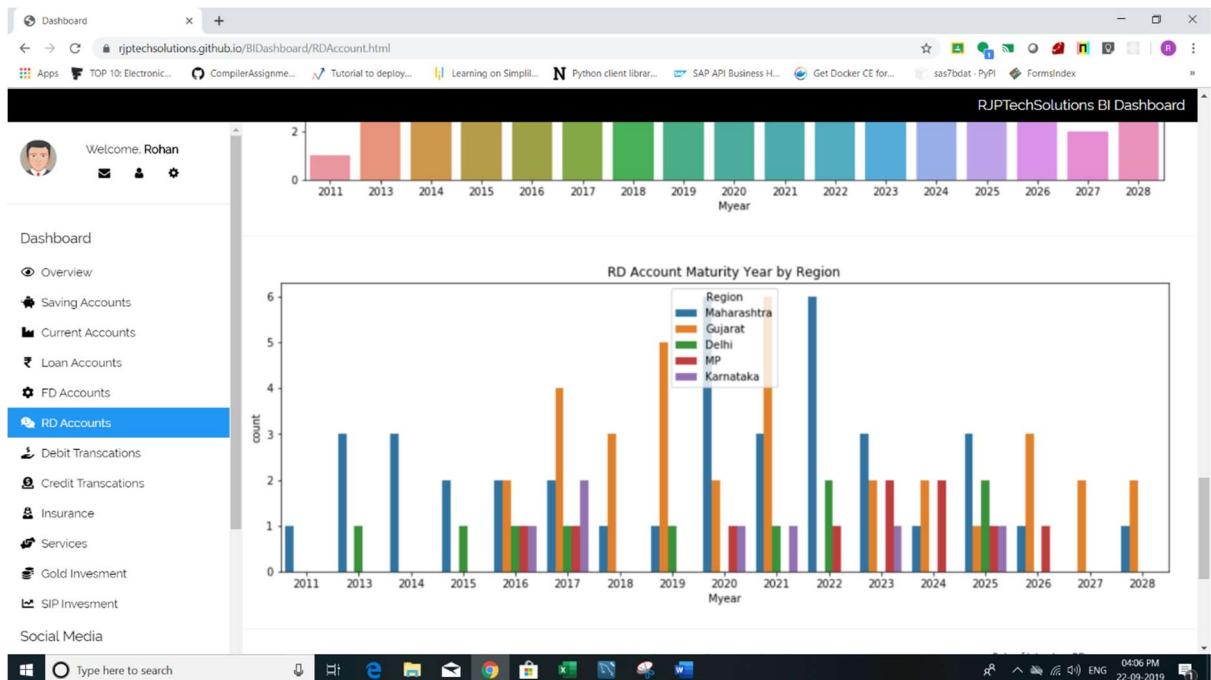
©CopyRights 2019 Banking and Insurance Dashboard  
Powered by RJPTechSolutions  
[Terms and Conditions](#)

## RD Accounts:

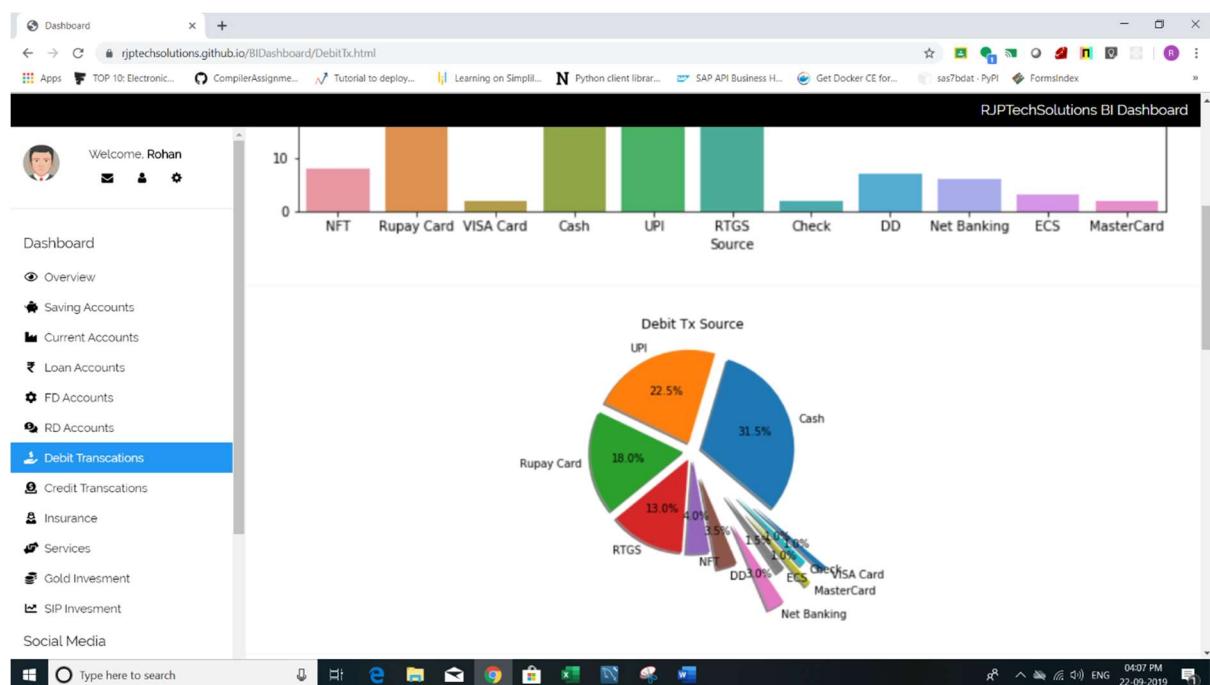
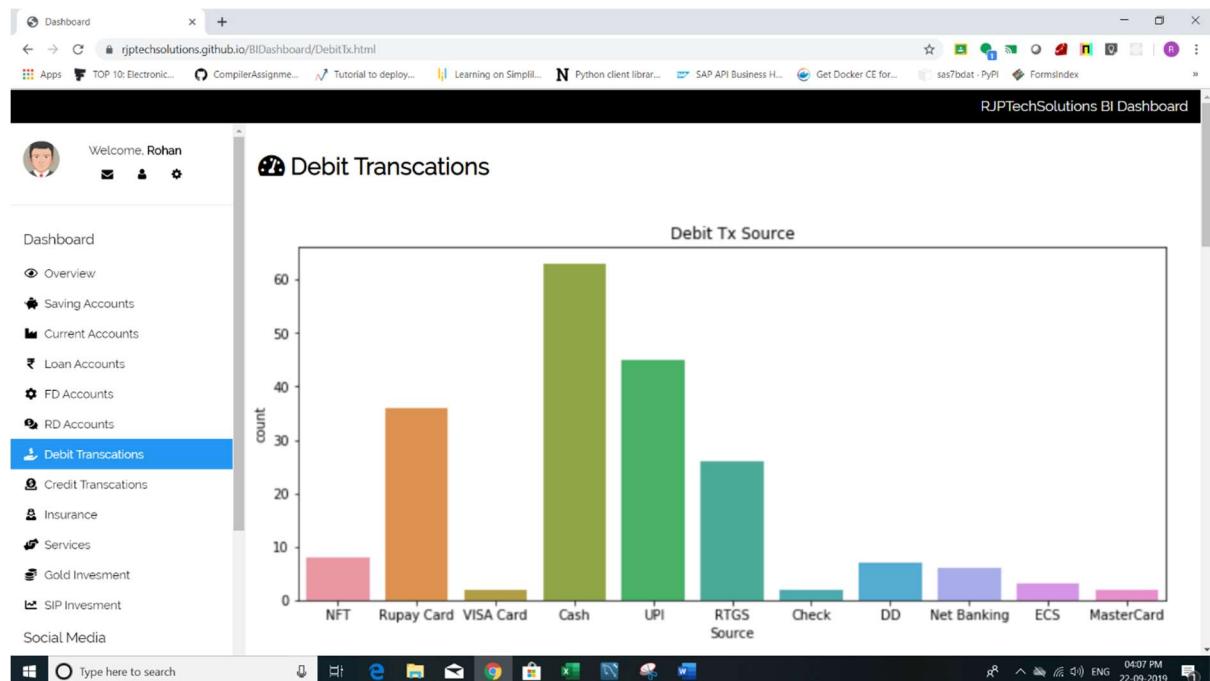


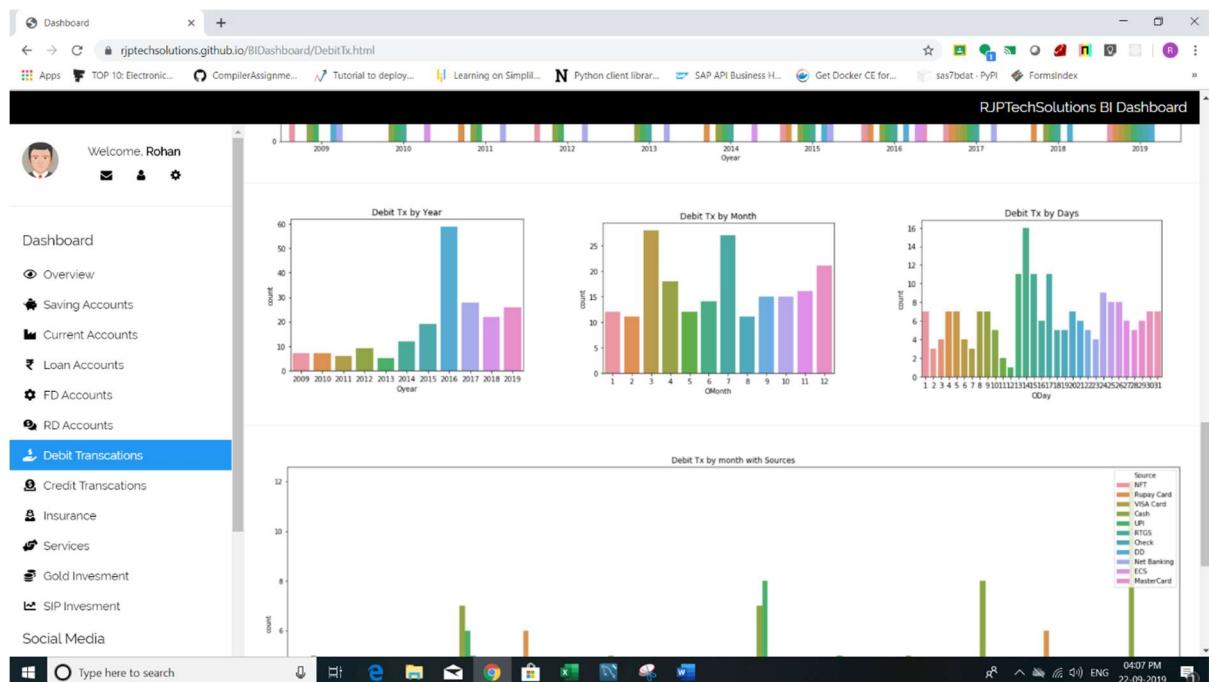
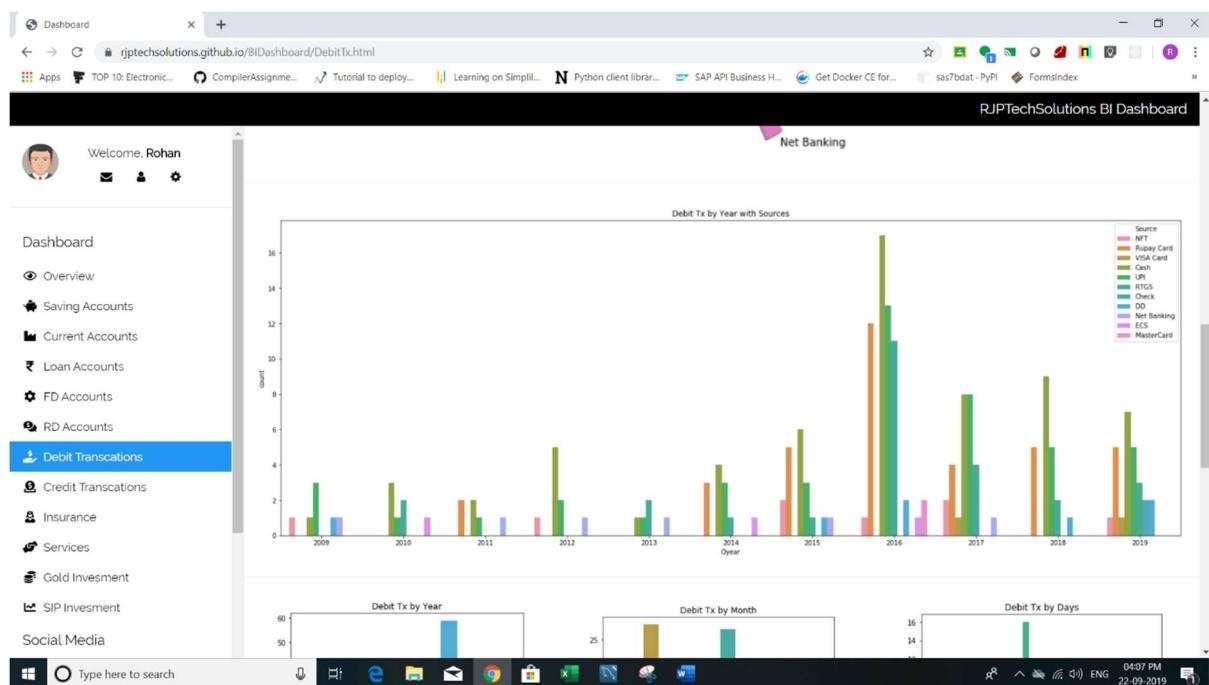


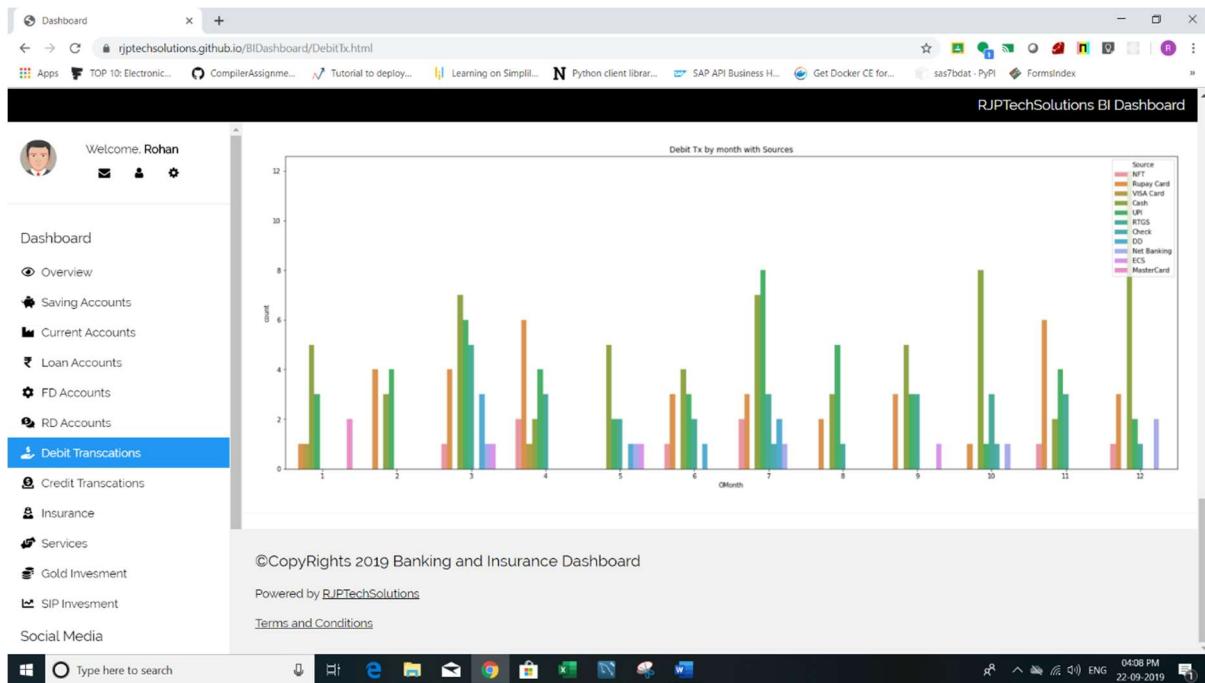




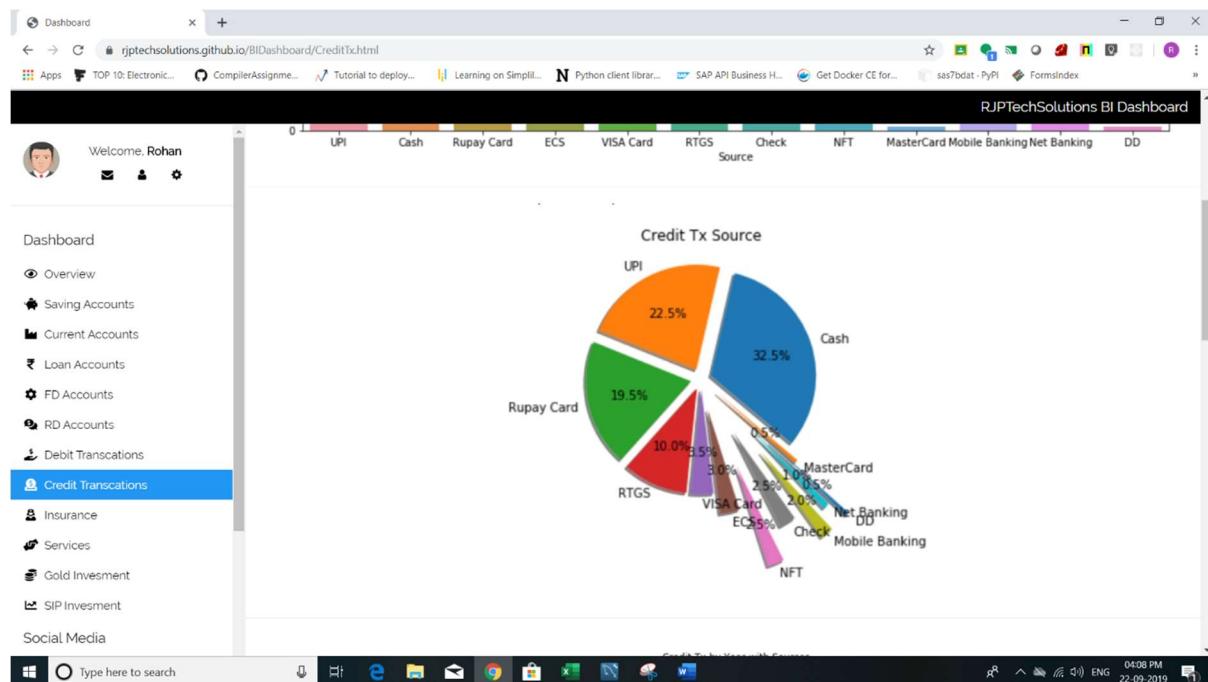
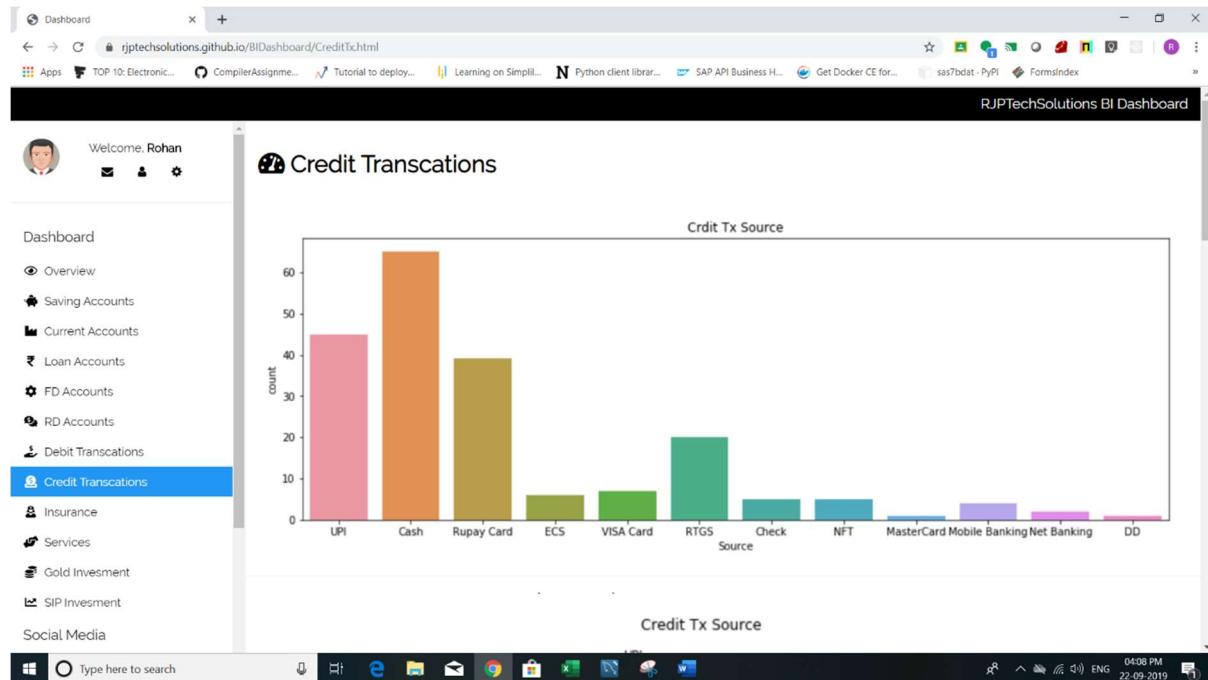
## Debit Transactions:

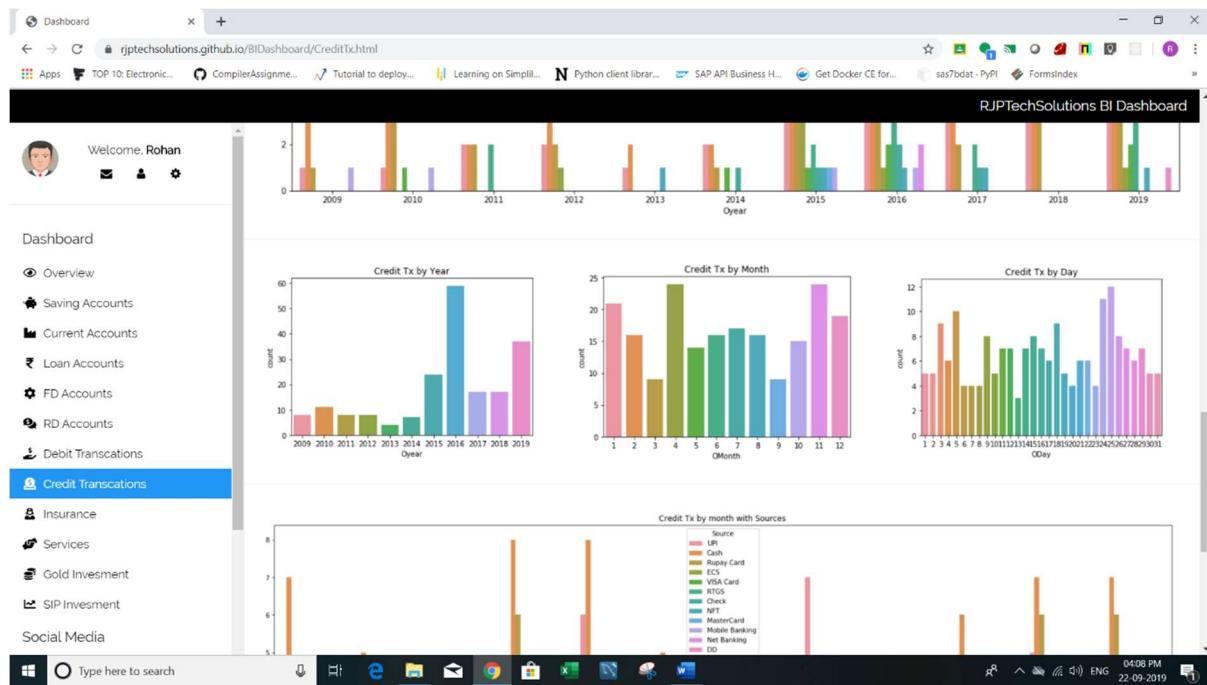
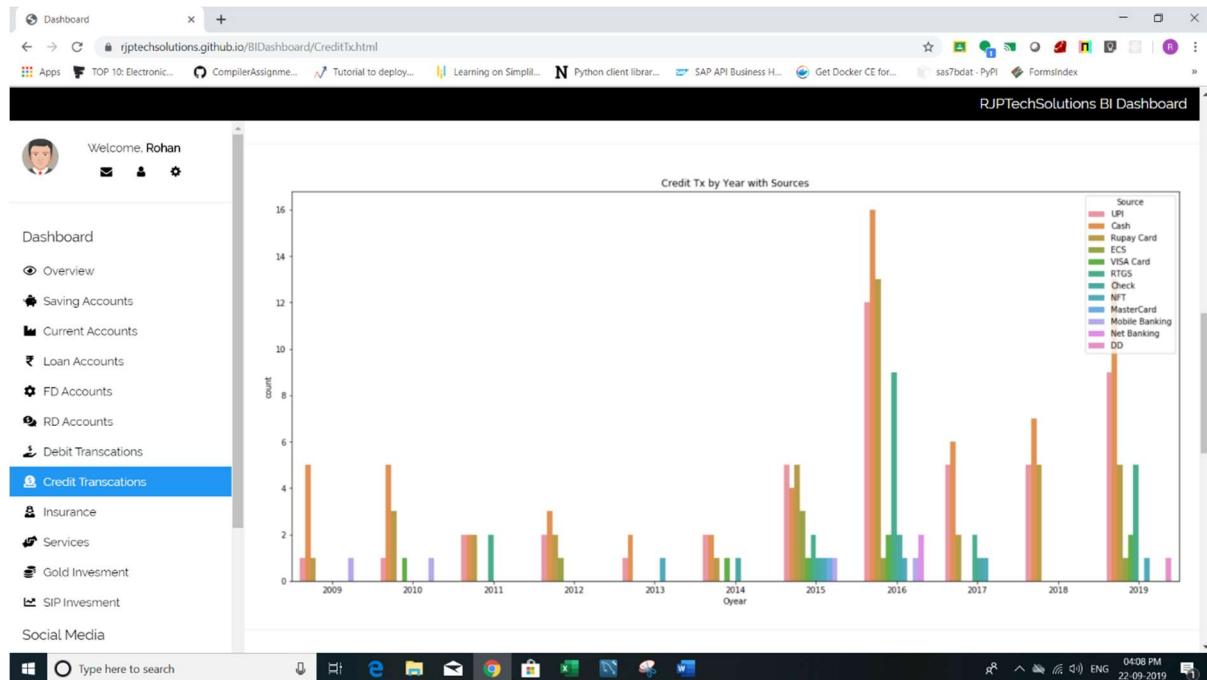


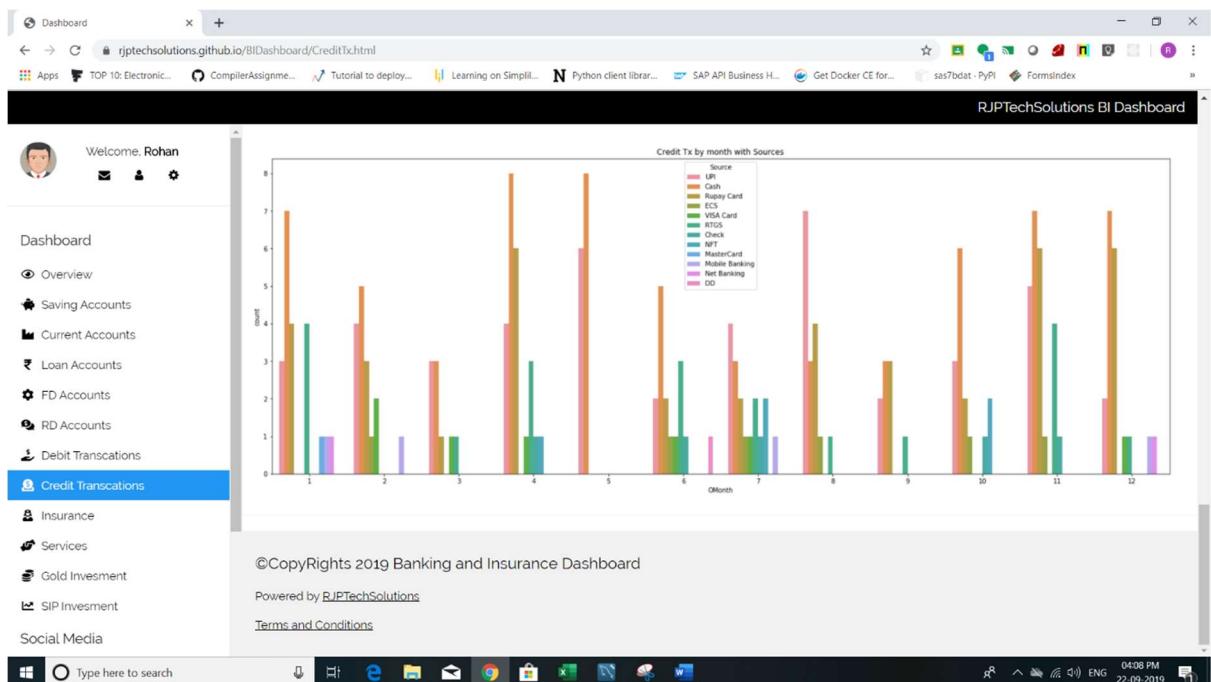
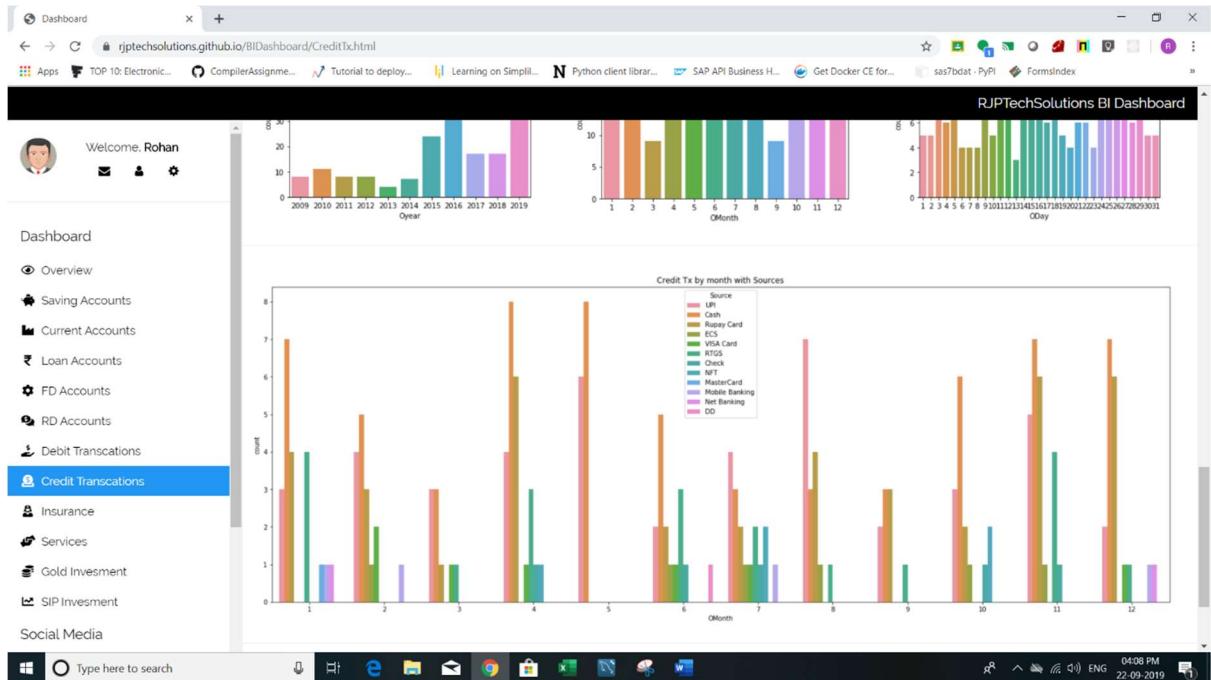




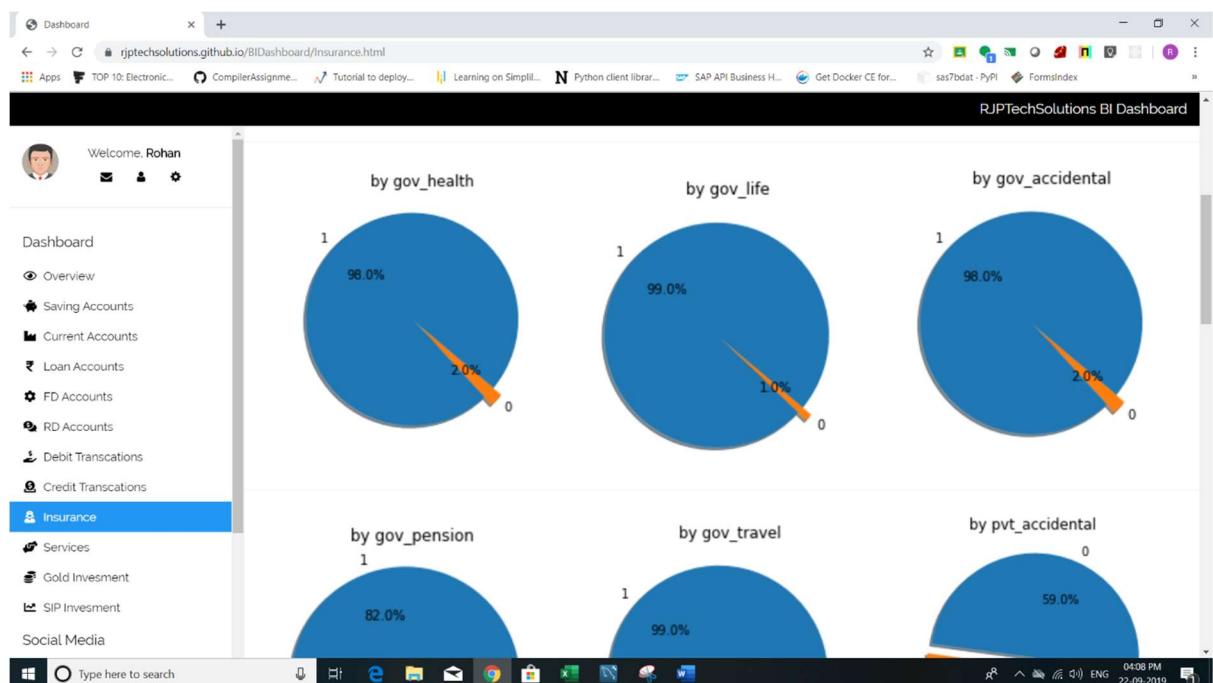
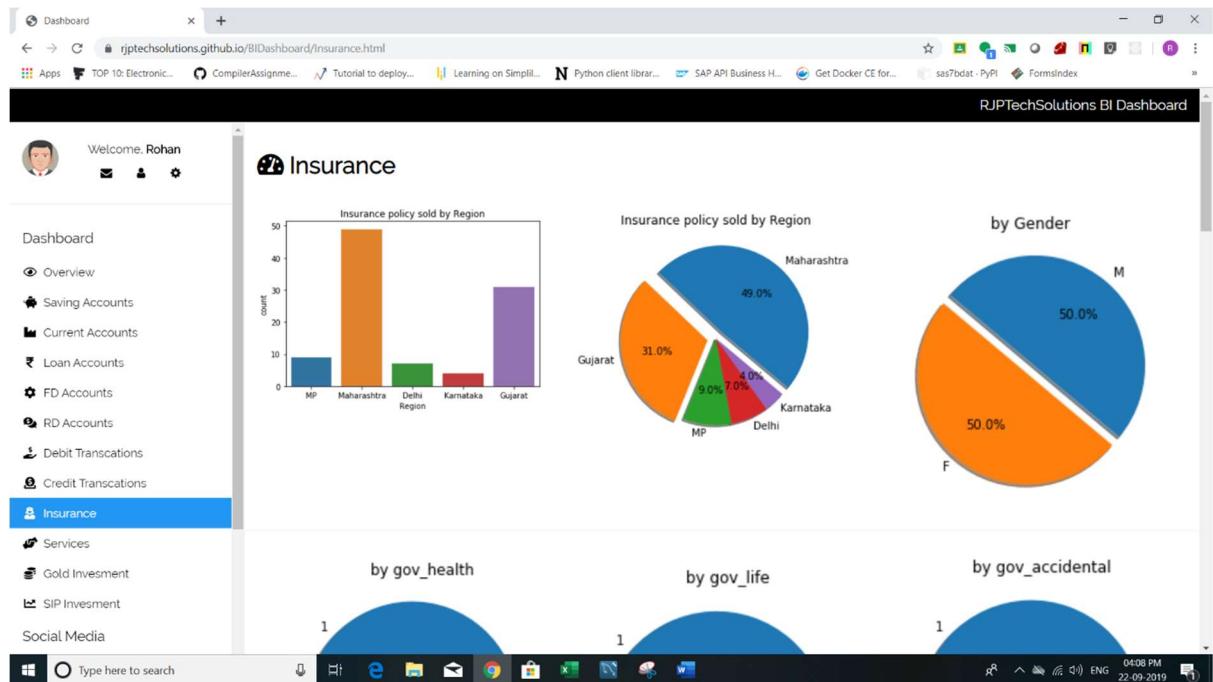
## Credit Transactions:

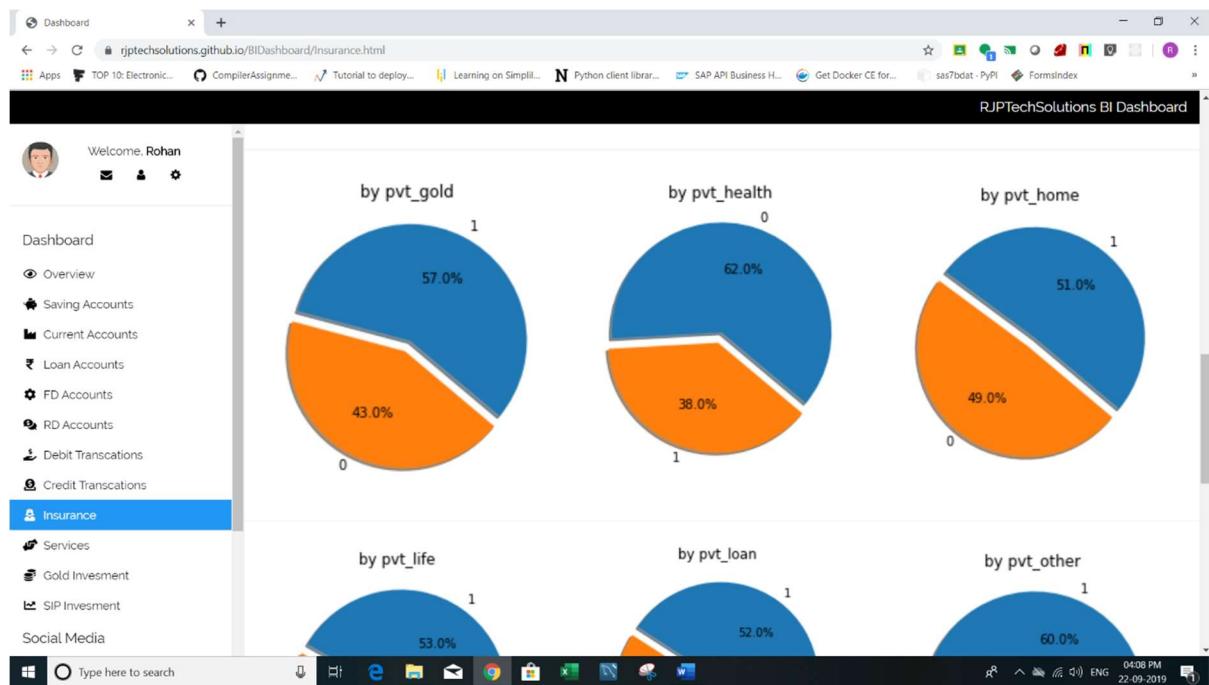
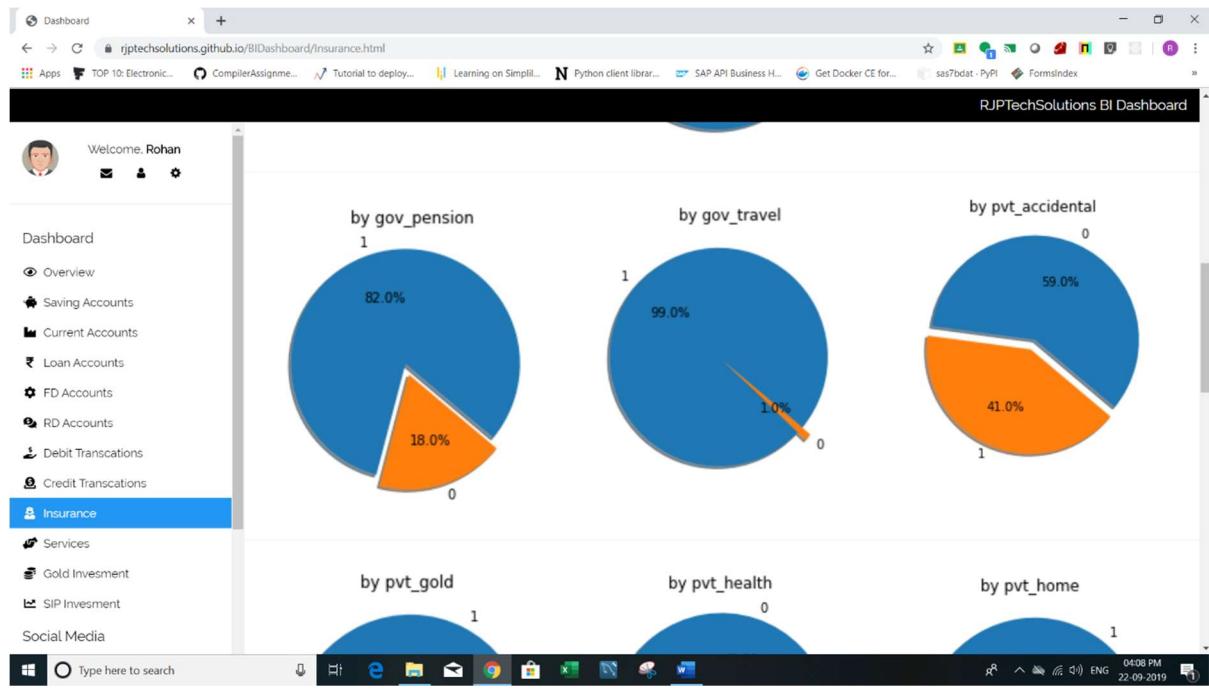


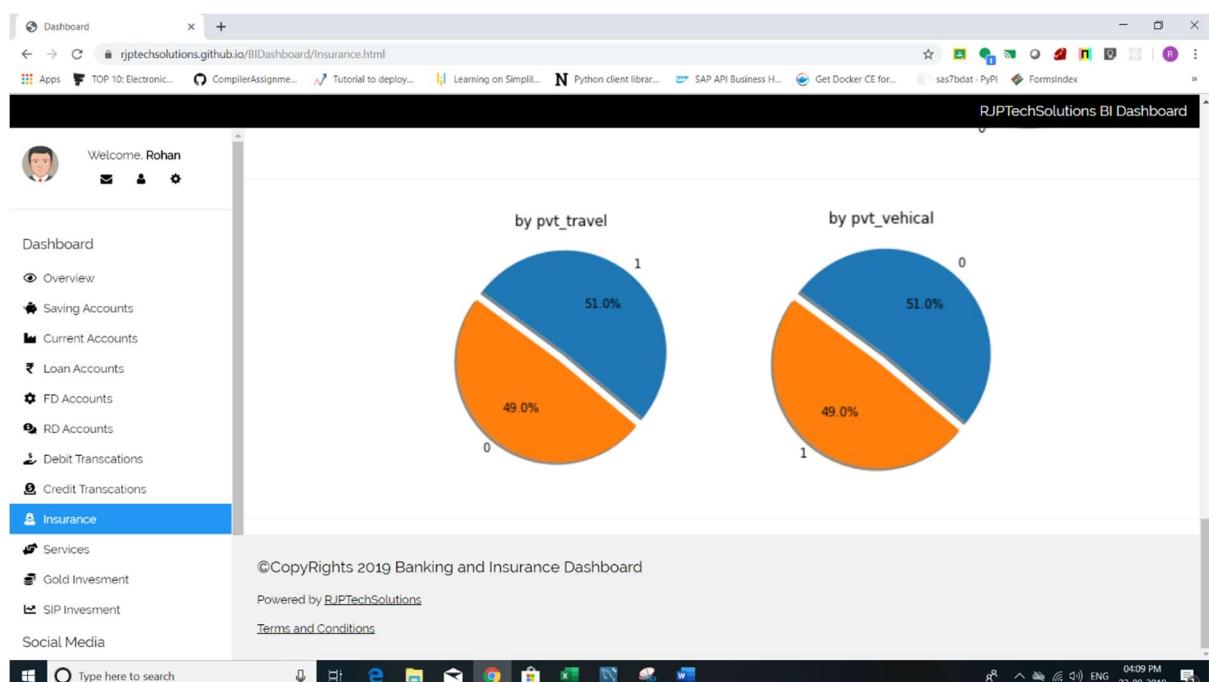




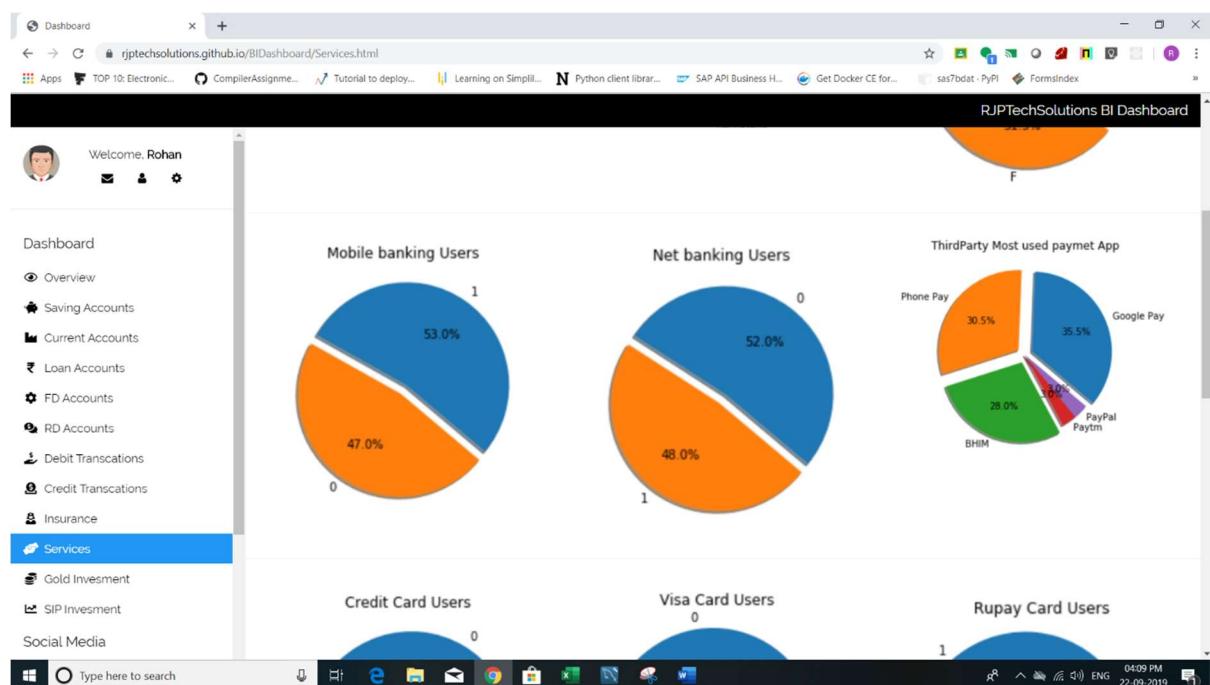
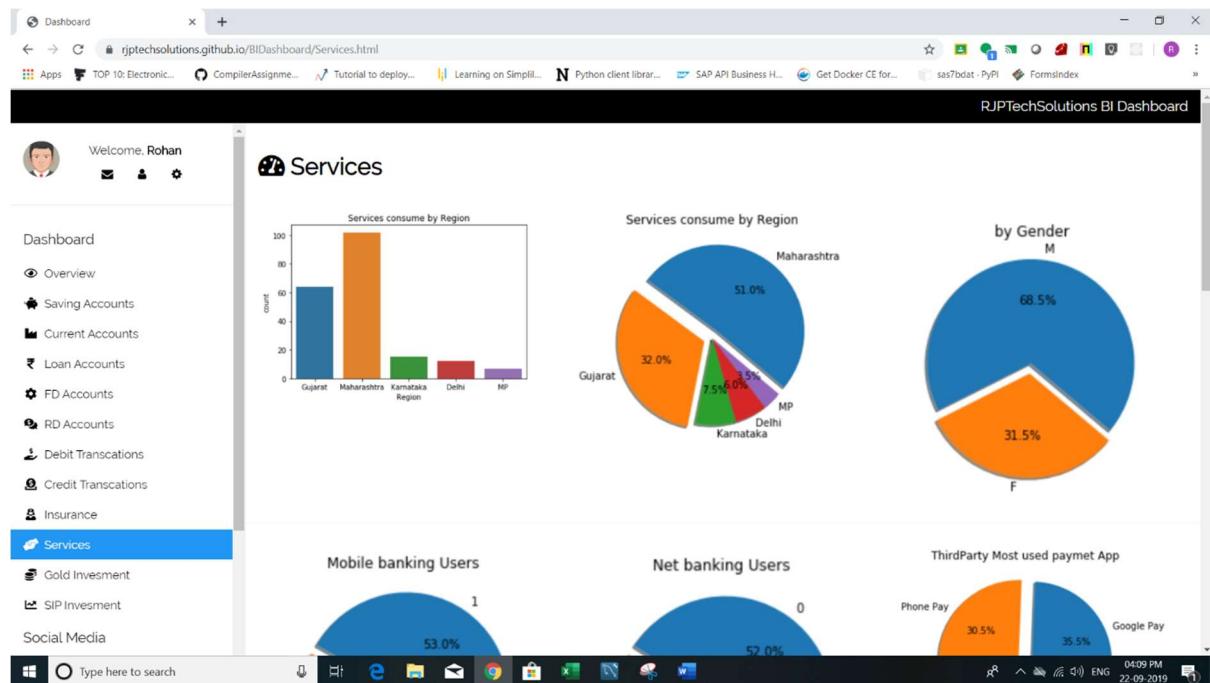
## Insurance:

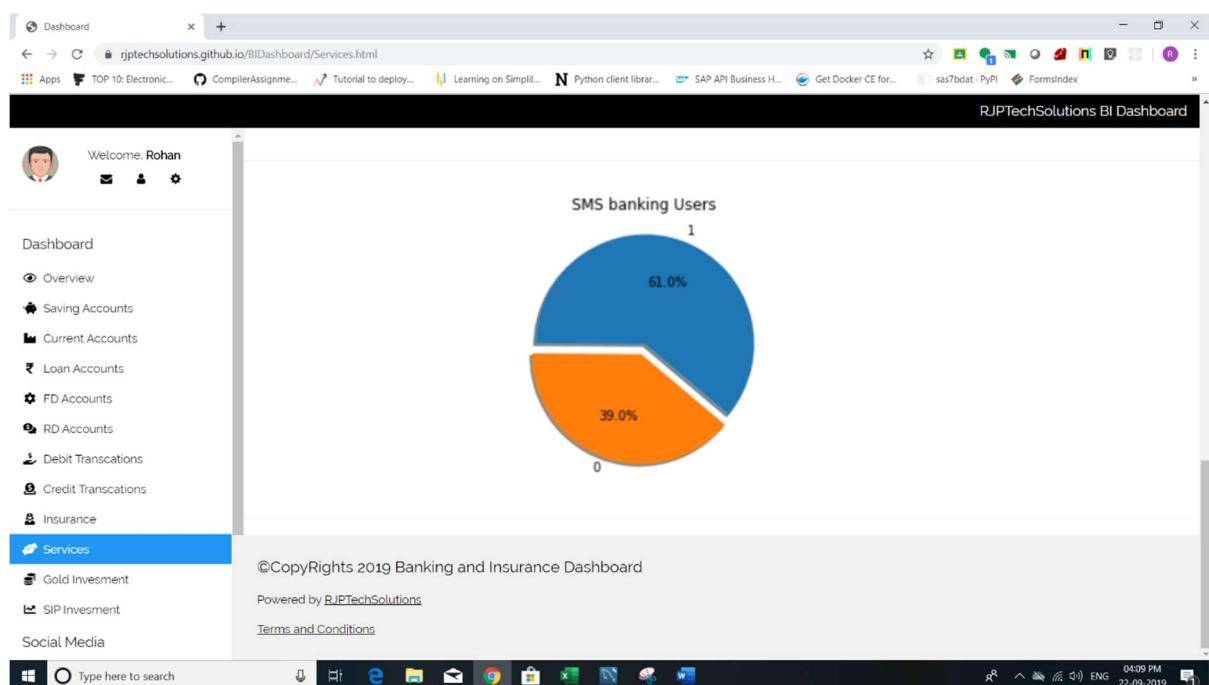
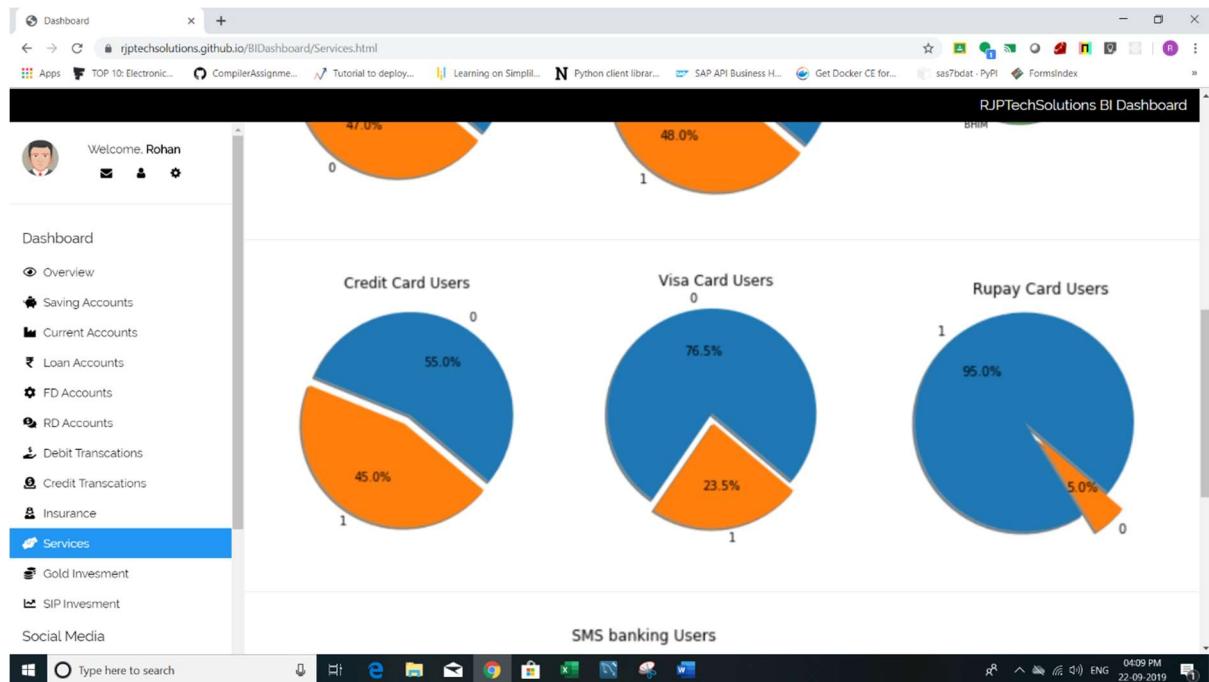




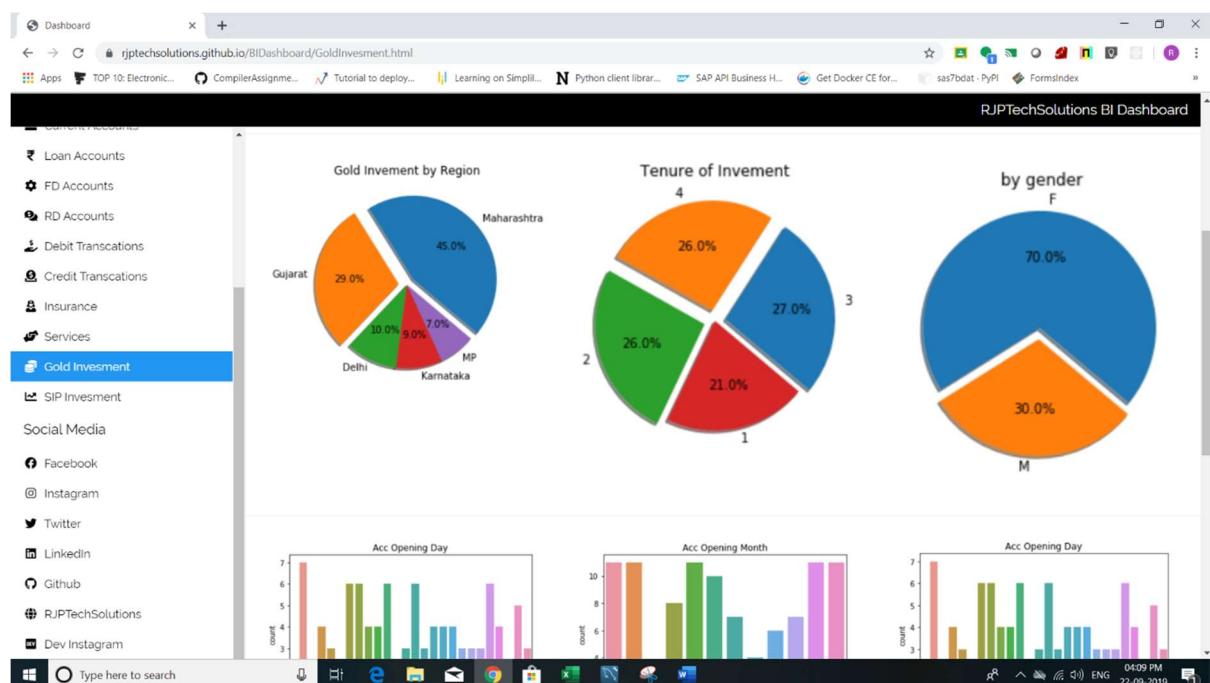
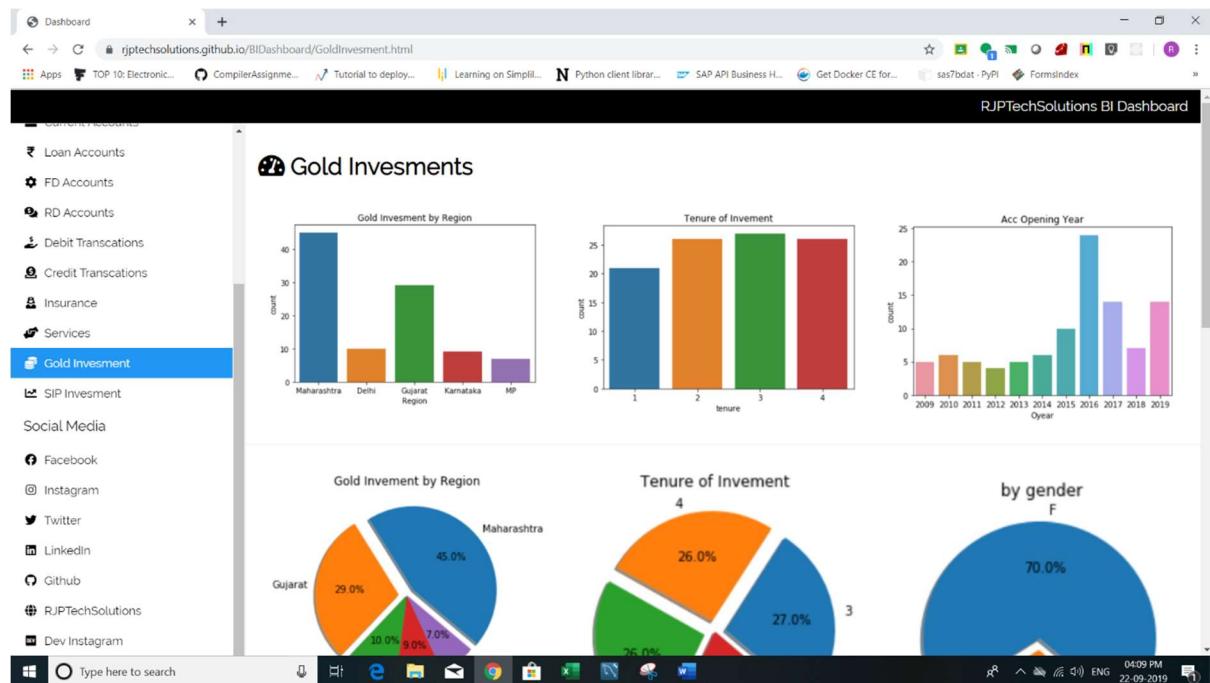


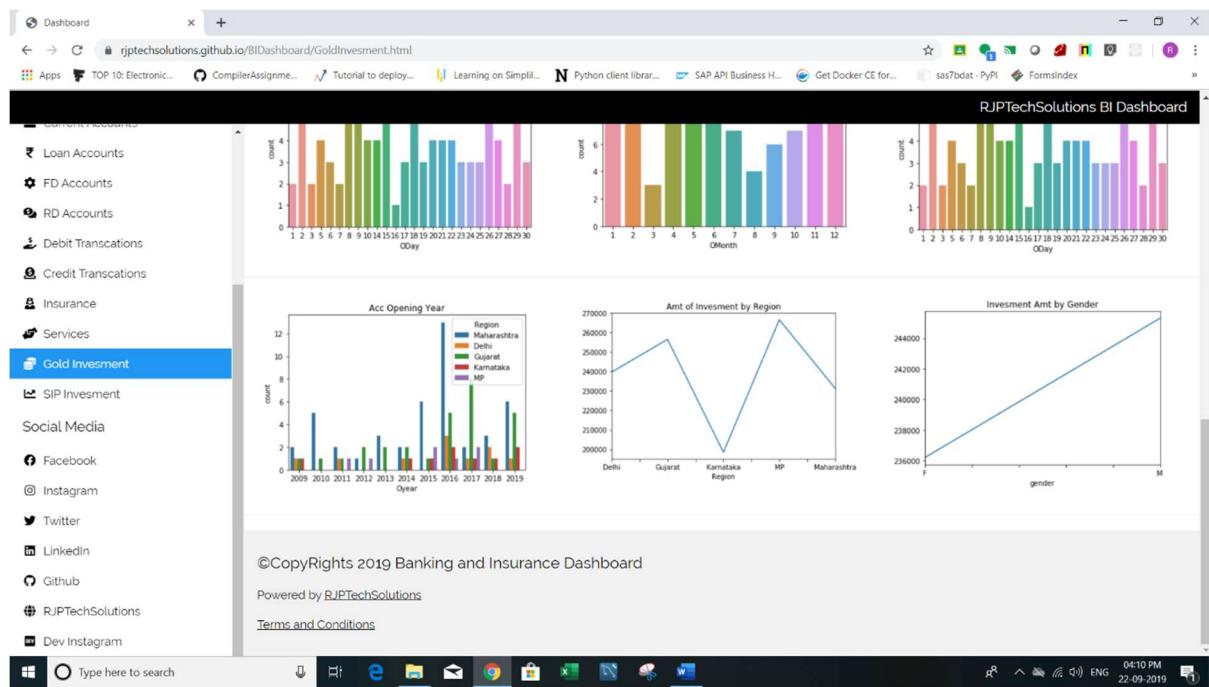
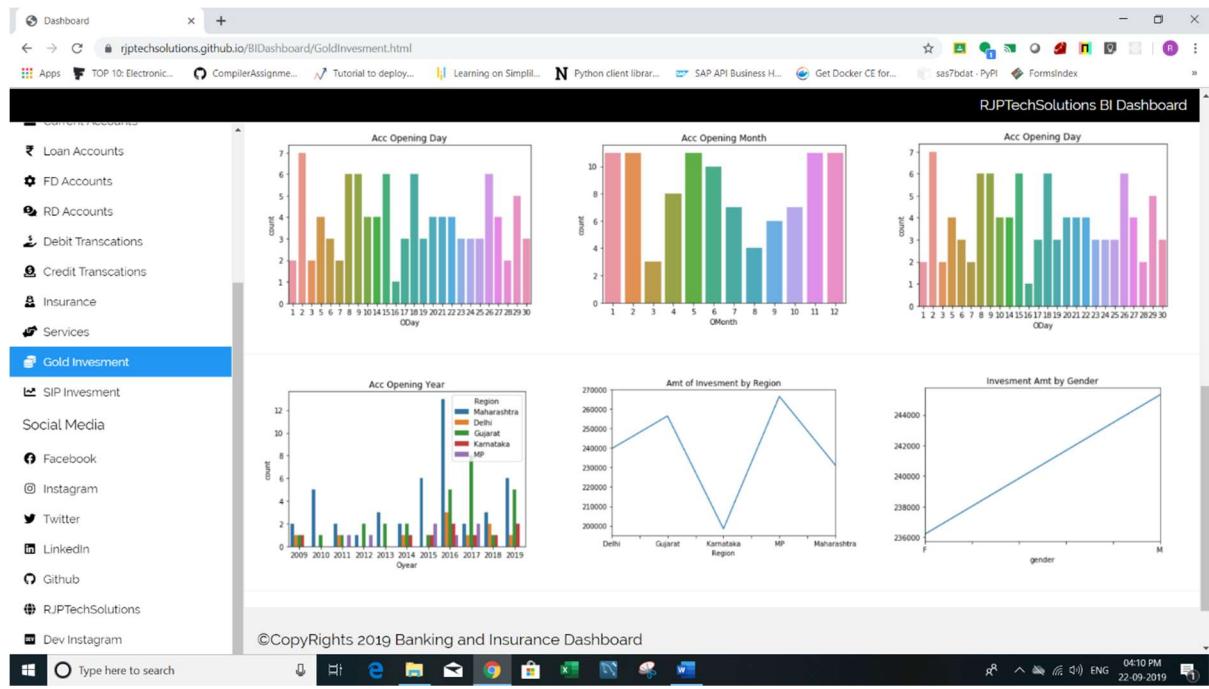
## Services:



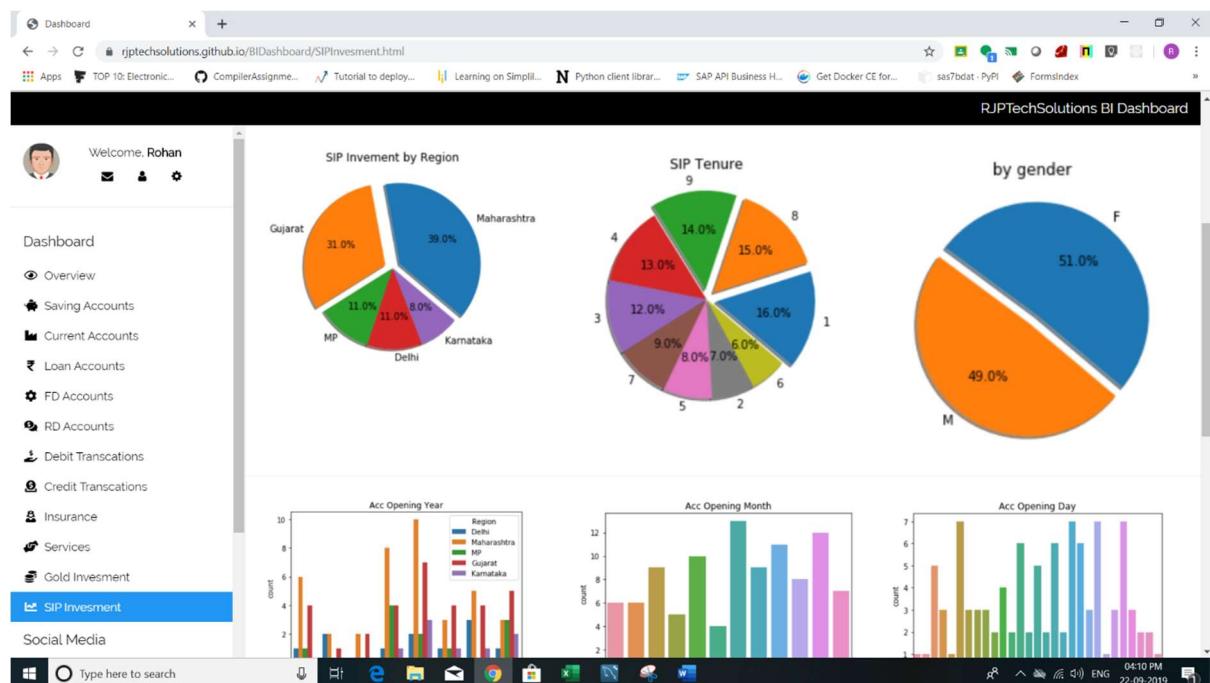
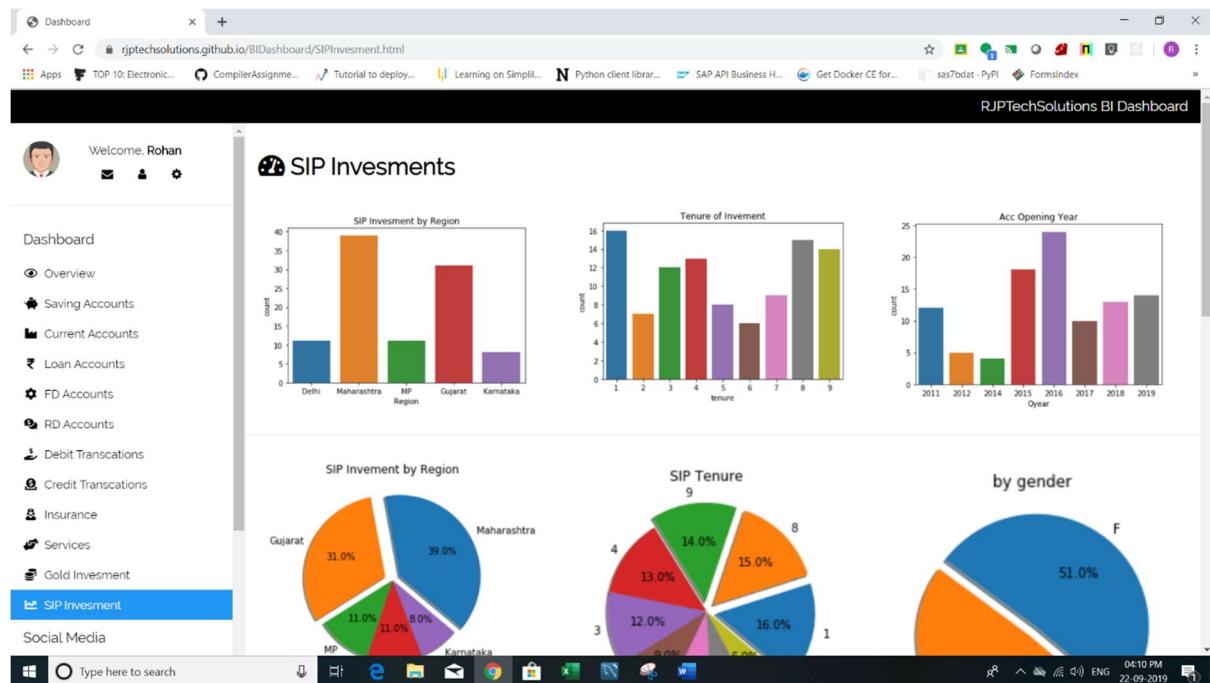


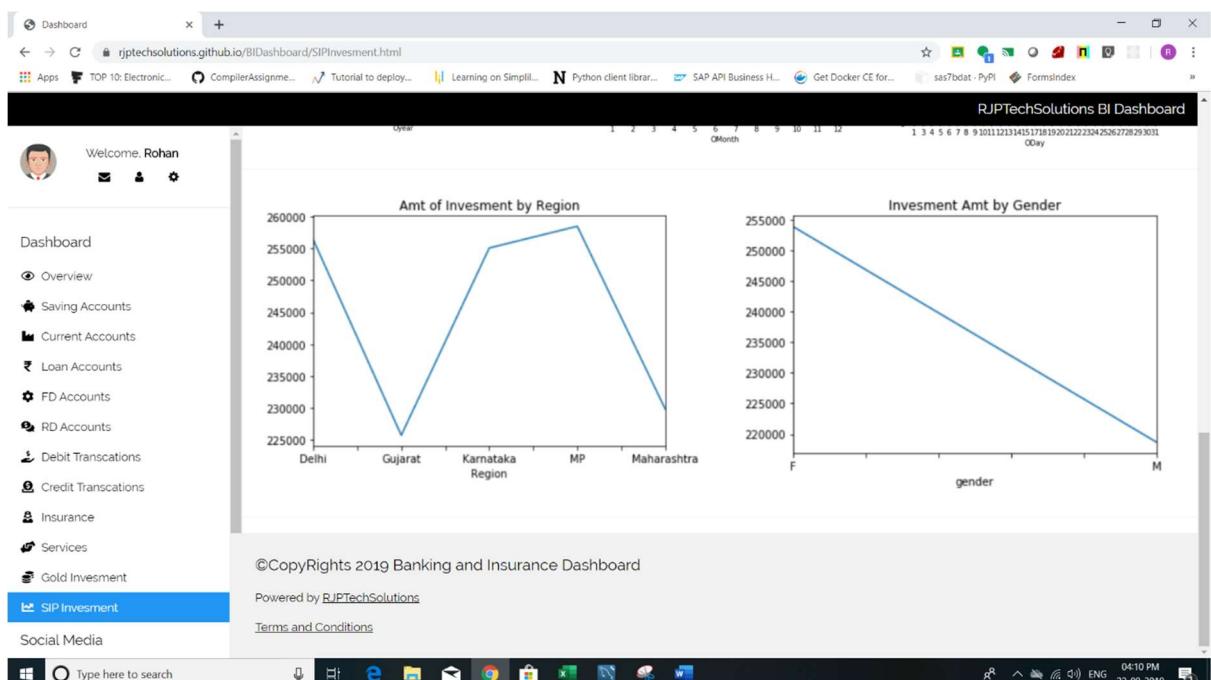
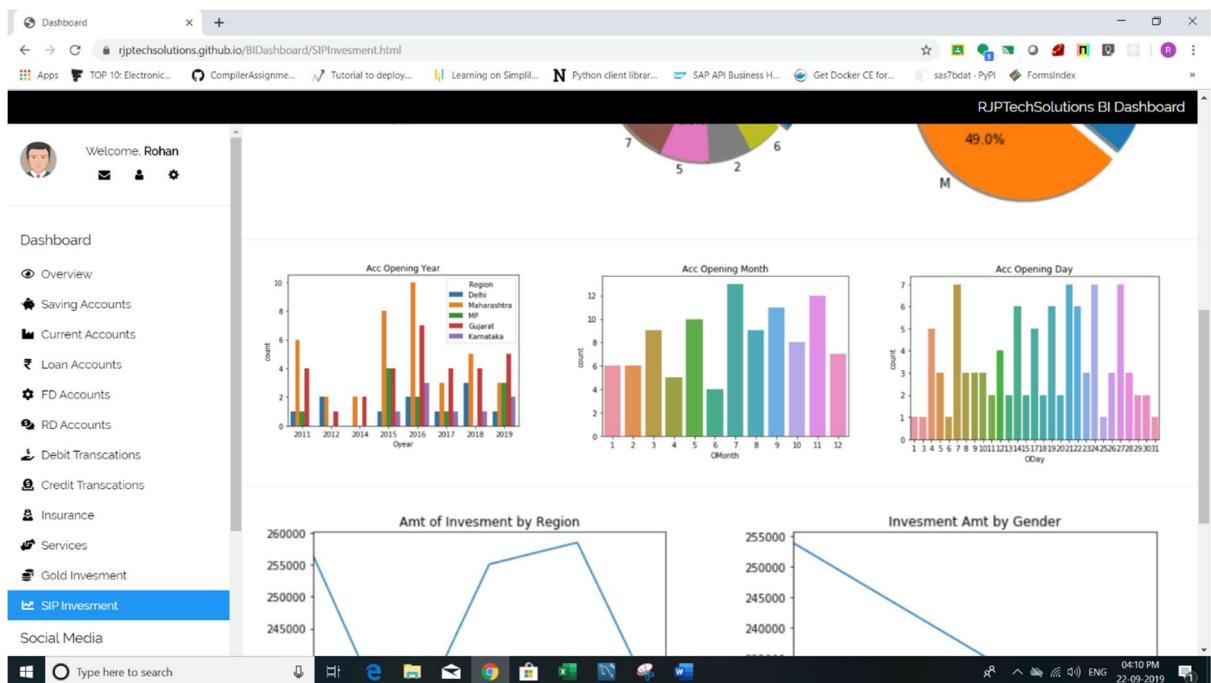
## Gold Investments:



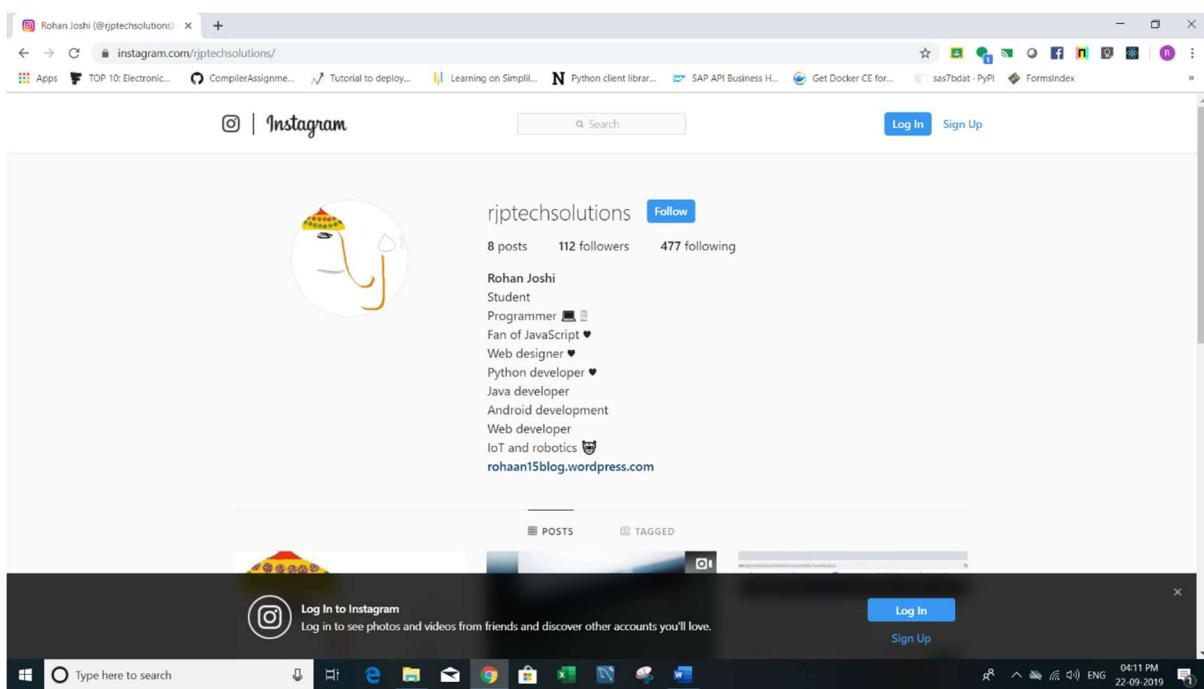
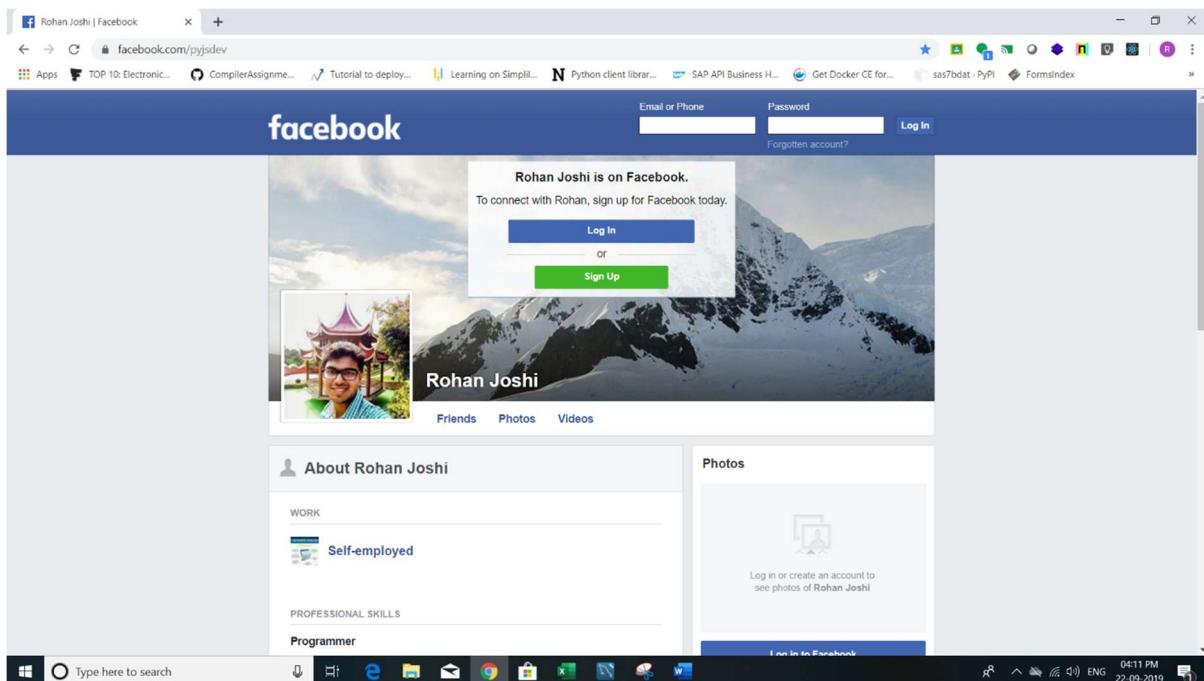


## SIP Investments:





## Social Media:



R.js (@Rohan83670129) | Twitter

Tweets 27 Following 386 Followers 9 Likes 157

R.js @Rohan83670129 - 23h #DevFest19

Want to take advantage of all the new Twitter features? It's simple – just log in.

Log in Sign up

Worldwide trends

#DaughtersDay 29.6K Tweets  
#Ghili 14.9K Tweets  
#WeAdmireKavin 206K Tweets

(2) Rohan Joshi | LinkedIn

Rohan Joshi

Doing Master's in Computer science from Ramnarain Ruia college working on IoT & Machine learning & Business intelligence

Mumbai, Maharashtra, India • 184 connections • Contact info

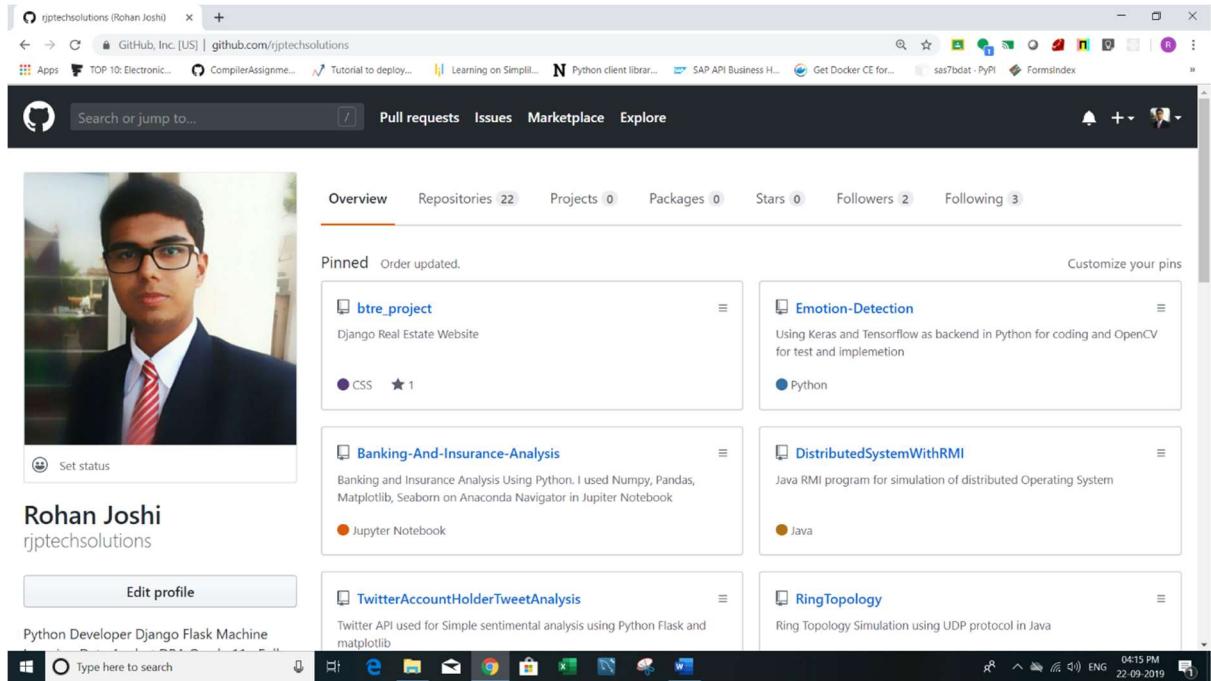
Add profile section More...

freshers Ramnarain Ruia College

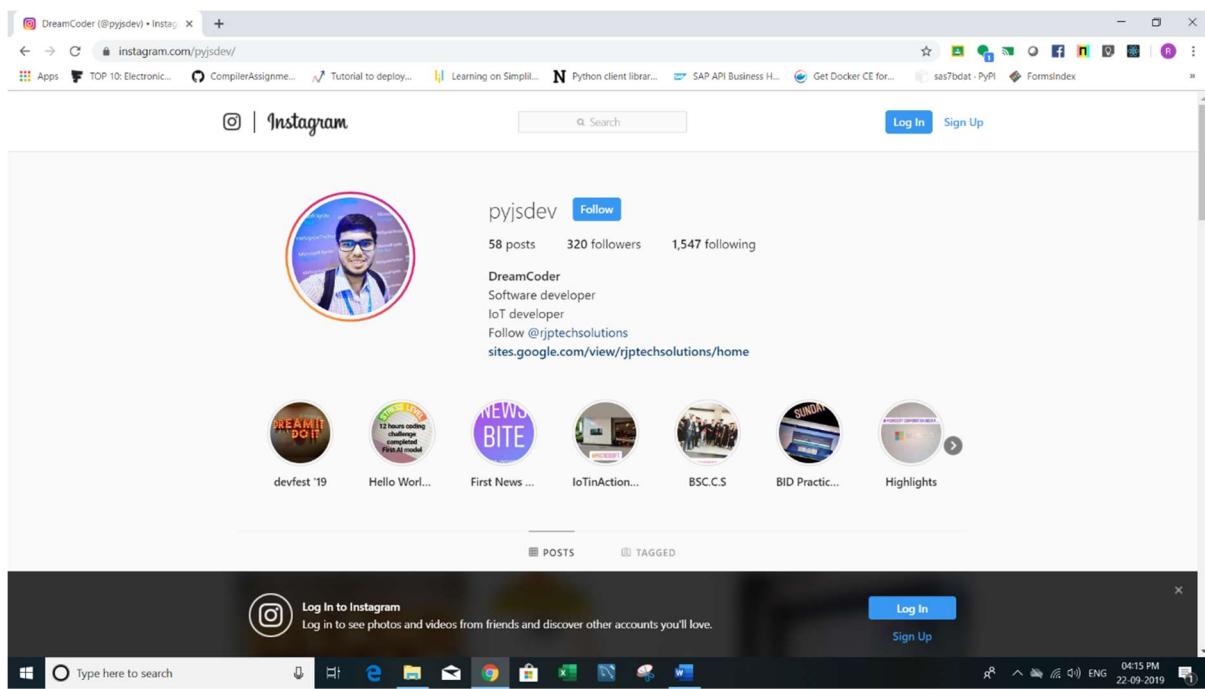
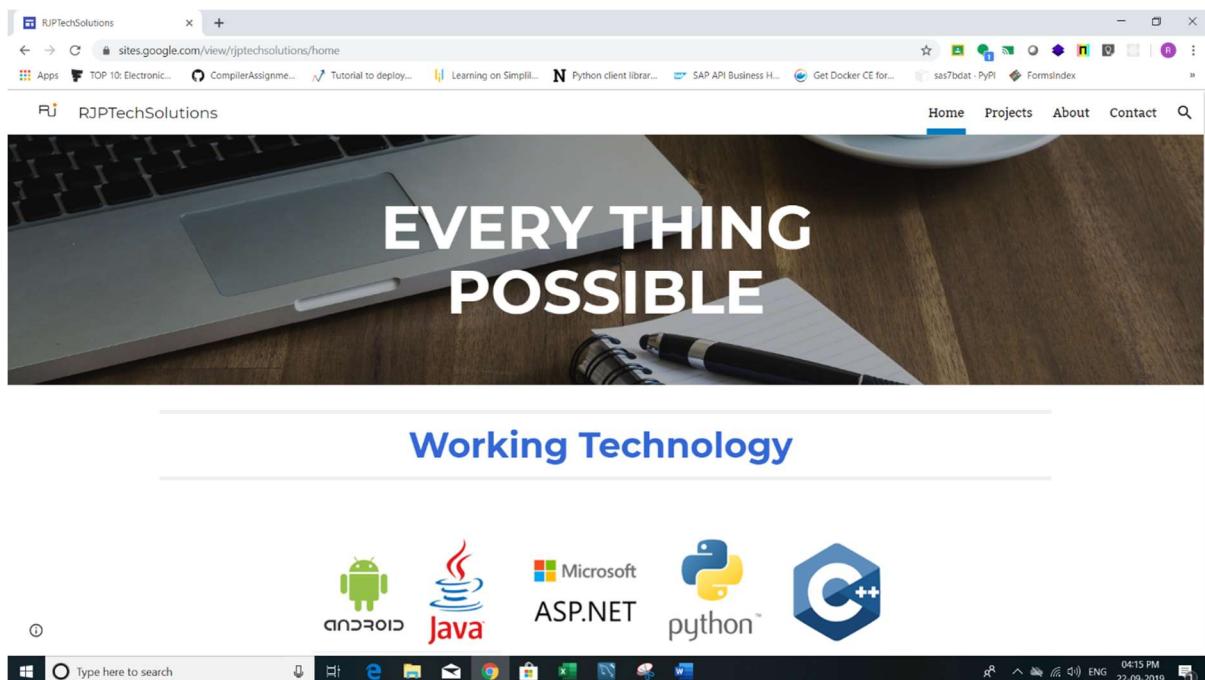
Strengthen your profile

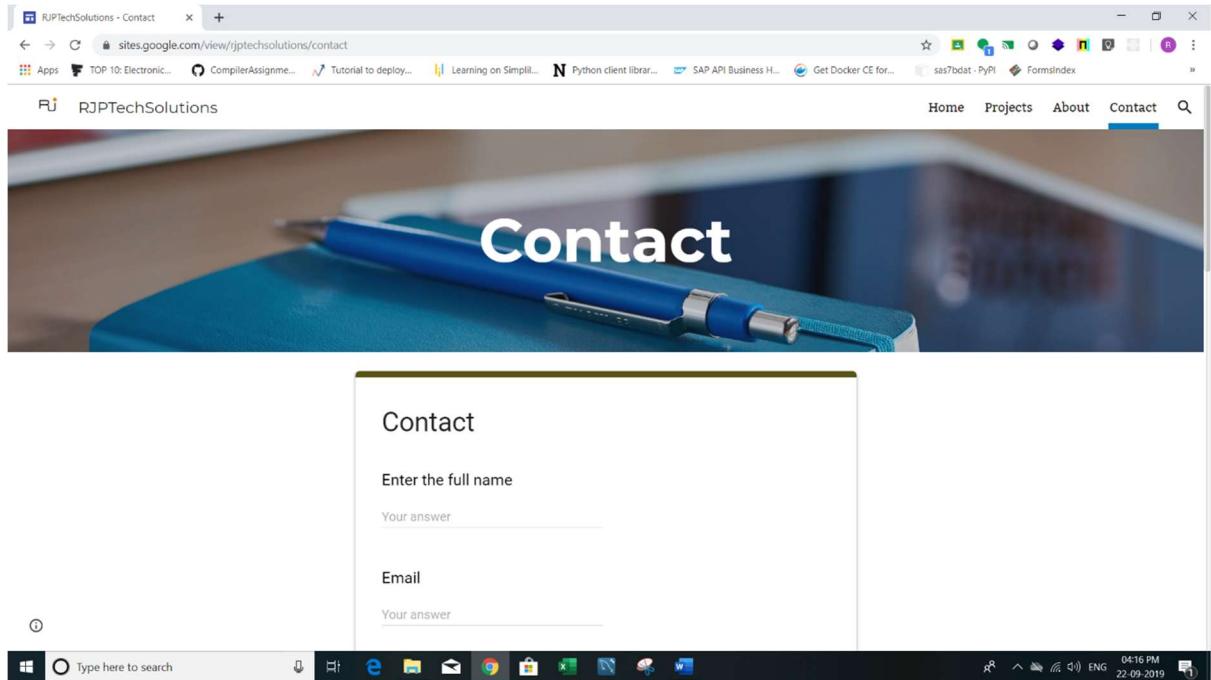
Rohan, help hiring managers and old coworkers find you Add a past position so others can easily find and connect with you Not now Add past position

Kalyani Gadre You: Congrats on your work anniv...  
Rishi Ganguly You: 😊  
Nitin Solanki You: Yes, thanks  
Shivani Singh You: Nice, But I am not a content ...  
Cassie from LinkedIn Premium LinkedIn Offer • Thanks for being...  
Muriel Fernandes Munet: Thank you  
Apoorva Mahale You: Thanks for connecting! Hope...  
Ainish Orine Mathias You: Thanks for connecting! Hope...  
Madhura Hete Madhura: Thanks! You, too  
Marina Saji Marina: Thanks! You, too

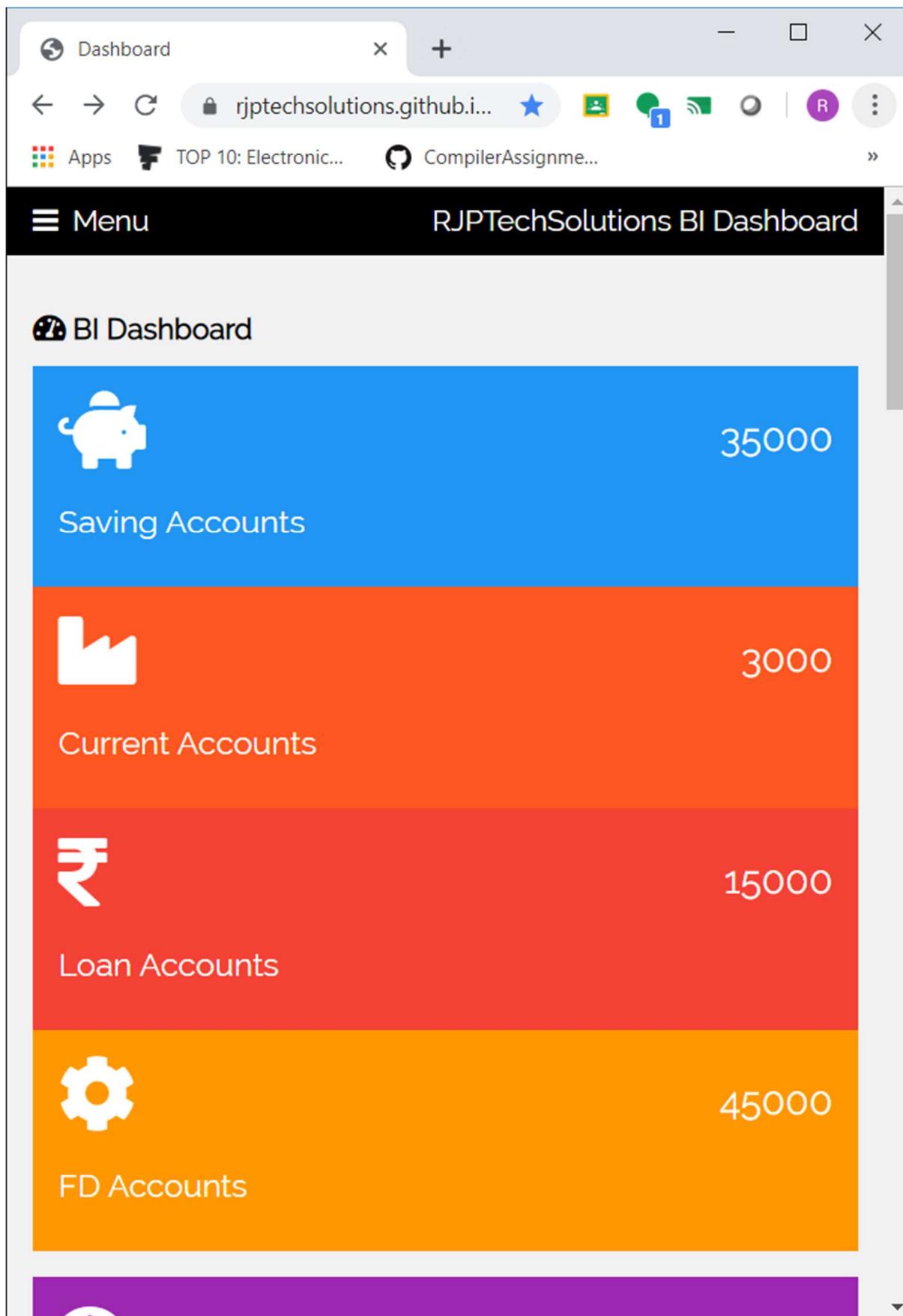


## Developer Support:





## Mobile View :-



Dashboard x +

← → C 🔒 rjptechsolutions.github.i... ⭐ 📁 1 💬 1 🌐 ⚙ R ⋮

Apps TOP 10: Electronic... CompilerAssignment... »

☰ Menu RJPTechSolutions BI Dashboard

Welcome, Rohan

✉️ 🚪 ⚙

Dashboard

✖ Close Menu

👁 Overview

🏦 Saving Accounts

💳 Current Accounts

₹ Loan Accounts

⚙ FD Accounts

₹ RD Accounts

🏧 Debit Transcations

₹ Credit Transcations

👤 Insurance

The dashboard displays a vertical bar chart with four distinct segments. The top segment is blue and labeled '35000'. The second segment is orange and labeled '3000'. The third segment is red and labeled '15000'. The bottom segment is brown and labeled '45000'.

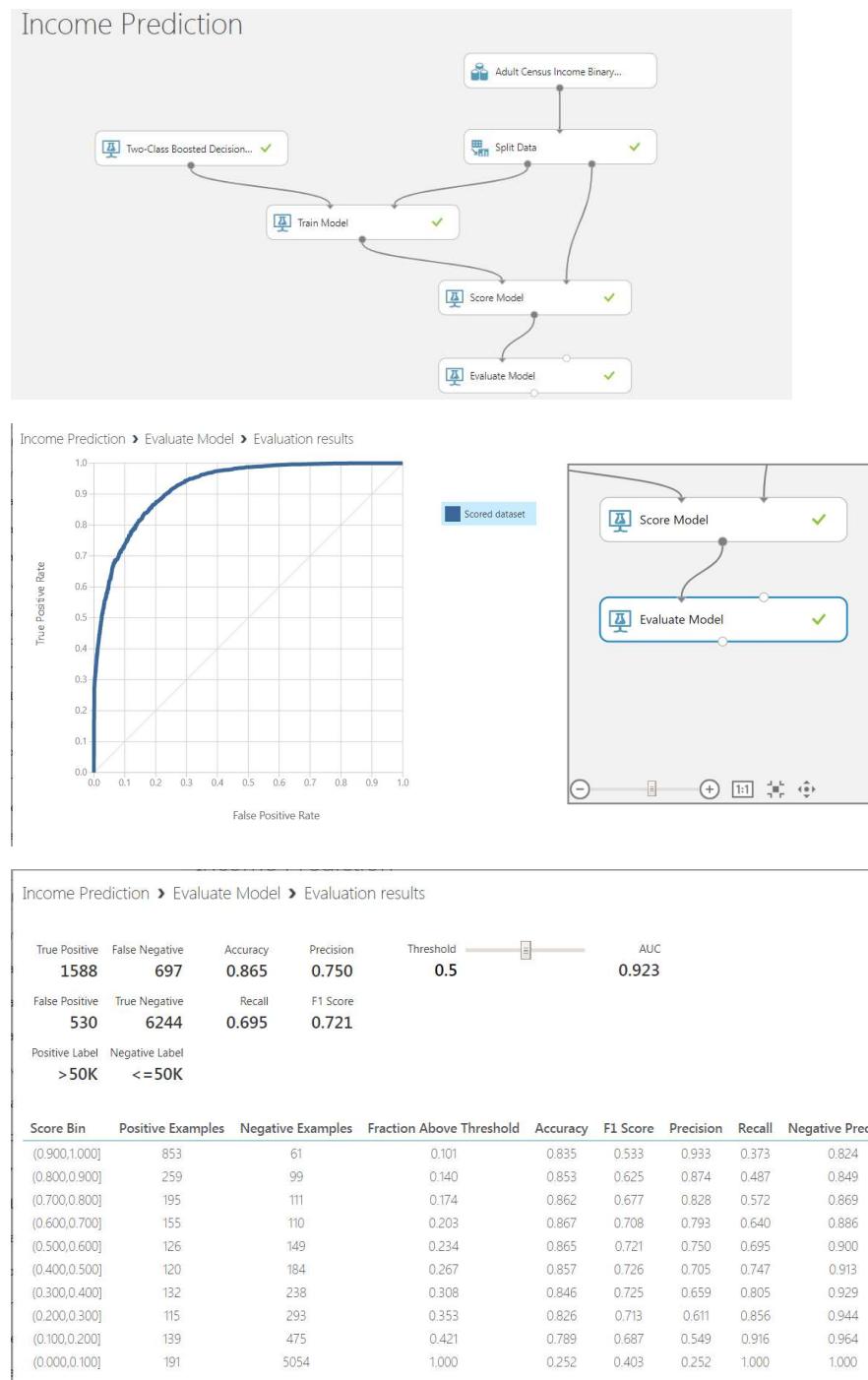
## **ML Models: -**

- **Customer Income Prediction Model**
- **Customer Churn Detection Model**
- **Evaluating Social Networking ads base customer purchasing policy's**
- **Credit Card Fraud Detection**
- **Credit Risk Prediction**
- **Automobile Price Prediction**
- **House Rate Prediction**

## **AI Model:**

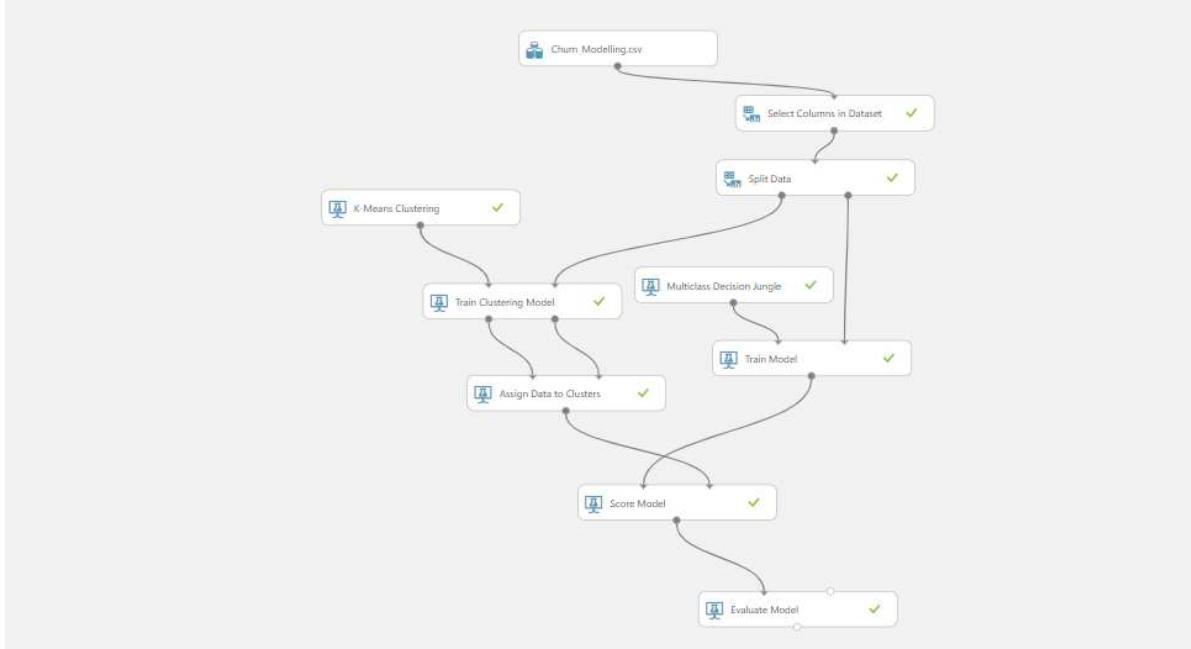
- **Customer Mood Detection**

## Customer Income Prediction Model:



## Customer Churn Detection Model:

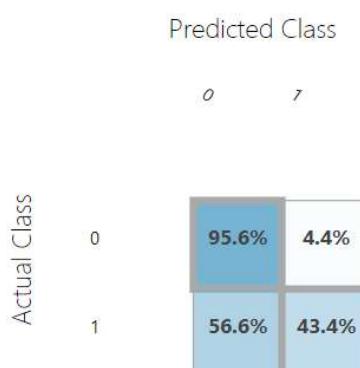
### Customer Churn Detection Model



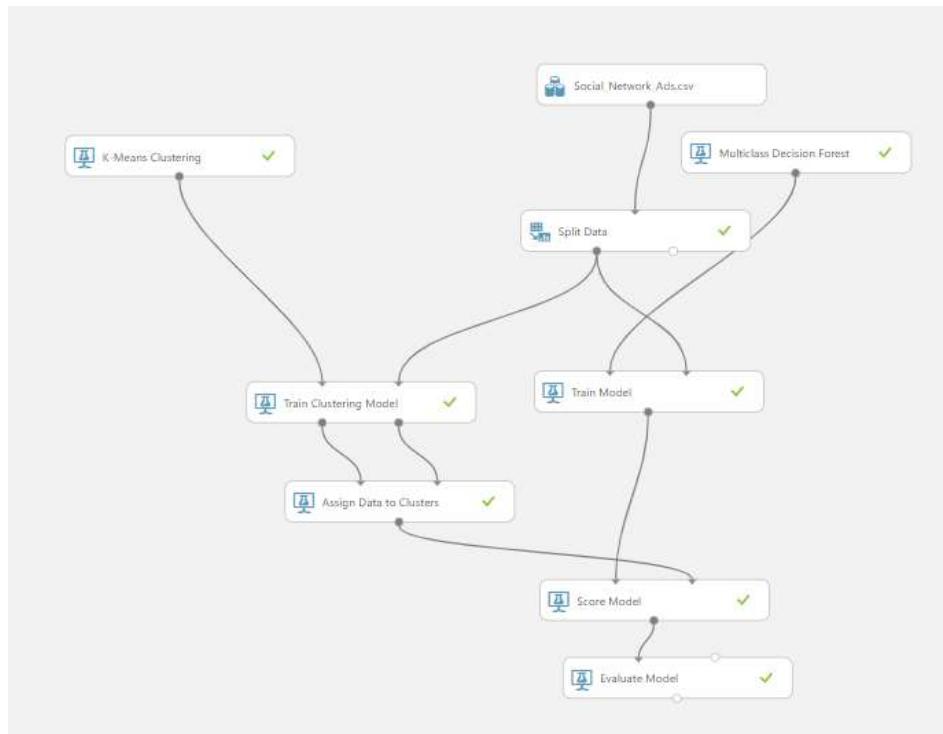
Customer Churn Detection Model ➤ Evaluate Model ➤ Evaluation results

Overall accuracy	0.850375
Average accuracy	0.850375
Micro-averaged precision	0.850375
Macro-averaged precision	0.79281
Micro-averaged recall	0.850375
Macro-averaged recall	0.695315

#### ◀ Confusion Matrix



## Evaluating Social Networking ads base customer purchasing policy's:

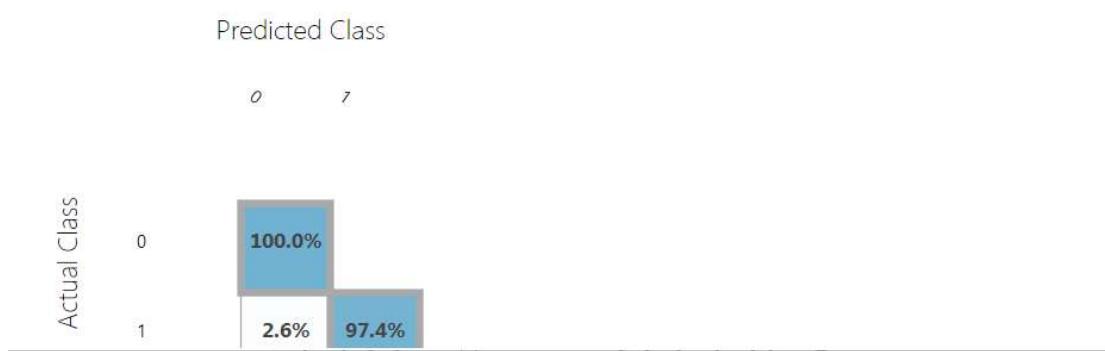


Social networking Ads to know customer purchasi... ➤ Evaluate Model ➤ Evaluation results

### Metrics

Overall accuracy	0.990625
Average accuracy	0.990625
Micro-averaged precision	0.990625
Macro-averaged precision	0.992823
Micro-averaged recall	0.990625
Macro-averaged recall	0.986842

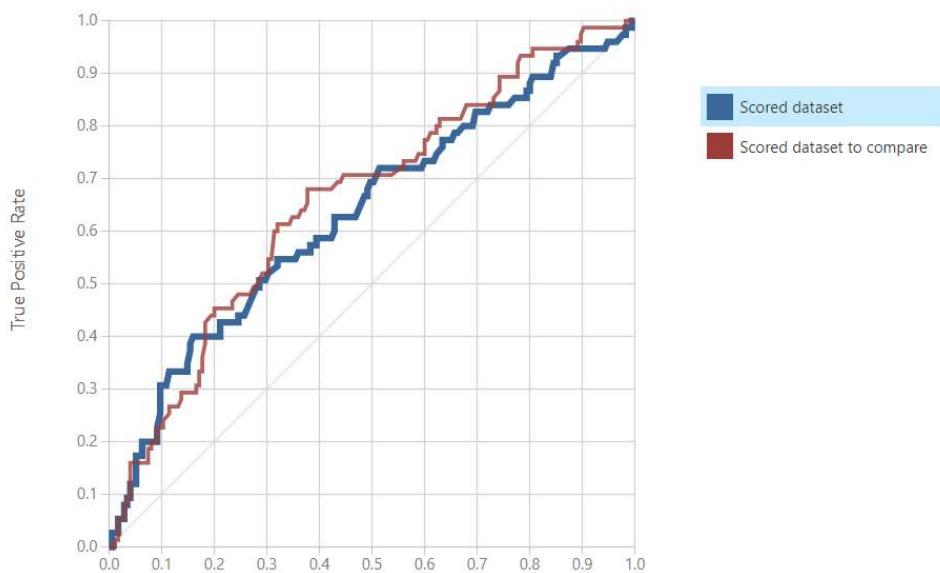
### Confusion Matrix



## Credit Card Fraud Detection: -

Anomaly Detection: Credit Risk > Evaluate Model > Evaluation results

ROC PRECISION/RECALL LIFT



Anomaly Detection: Credit Risk > Evaluate Model > Evaluation results

True Positive	False Negative	Accuracy	Precision	Threshold	AUC
15	60	0.716	0.577	0.5	0.633
False Positive	True Negative	Recall	F1 Score		
11	164	0.200	0.297		
Positive Label	Negative Label				
2	1				

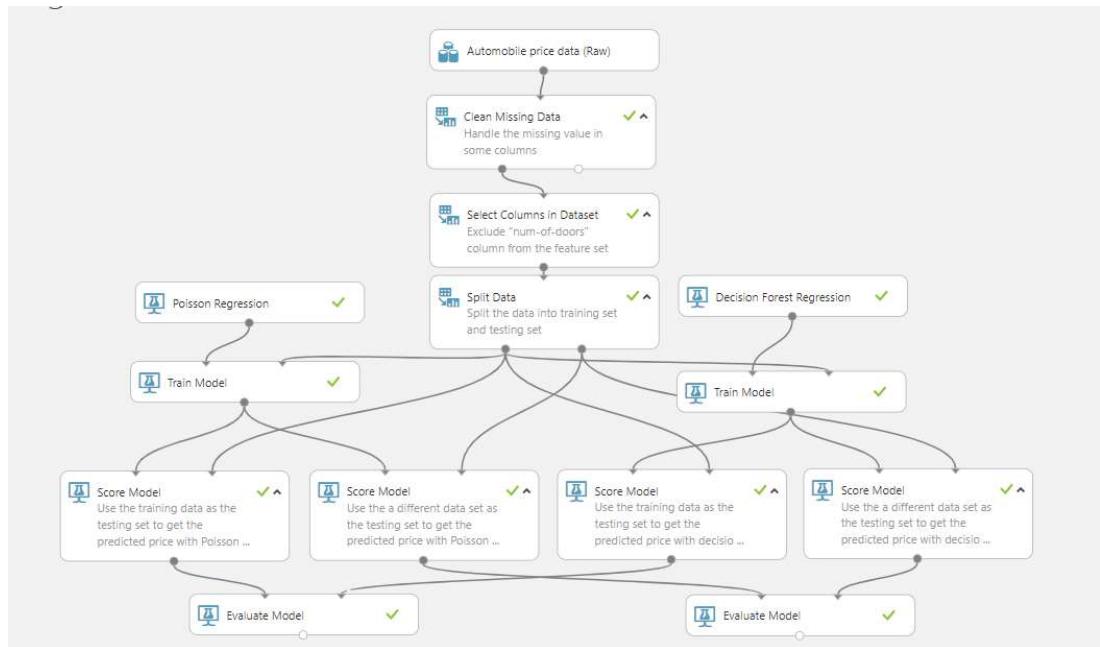
Score Bin	Positive Examples	Negative Examples	Fraction Above Threshold	Accuracy	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Cumulative AUC
(0.900,1.000]	0	0	0.000	0.700	0.000	1.000	0.000	0.700	1.000	0.000
(0.800,0.900]	0	0	0.000	0.700	0.000	1.000	0.000	0.700	1.000	0.000
(0.700,0.800]	1	1	0.008	0.700	0.026	0.500	0.013	0.702	0.994	0.000
(0.600,0.700]	2	2	0.024	0.700	0.074	0.500	0.040	0.705	0.983	0.000
(0.500,0.600]	12	8	0.104	0.716	0.297	0.577	0.200	0.732	0.937	0.005
(0.400,0.500]	27	54	0.428	0.608	0.462	0.393	0.560	0.769	0.629	0.133
(0.300,0.400]	28	85	0.880	0.380	0.475	0.318	0.933	0.833	0.143	0.497
(0.200,0.300]	5	25	1.000	0.300	0.462	0.300	1.000	1.000	0.000	0.633
(0.100,0.200]	0	0	1.000	0.300	0.462	0.300	1.000	1.000	0.000	0.633
(0.000,0.100]	0	0	1.000	0.300	0.462	0.300	1.000	1.000	0.000	0.633

## Credit Risk Prediction:

Binary Classification: Credit risk prediction ➔ Select Columns in Dataset ➔ Results dataset

rows	columns	
4	3	
Algorithm	Training	Accuracy
SVM	weighted	0.704545
SVM	unweighted	0.56
Boosted Decision Tree	weighted	0.659091
Boosted Decision Tree	unweighted	0.582727

## Automobile Price Prediction:



Regression: Automobile Price Prediction > Evaluate Model > Evaluation results

rows    columns  
2        6

	Negative Log Likelihood	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error	Coefficient of Determination
view as						
Infinity	3188.522151	4744.027861	0.454332	0.264362	0.735638	
375.110236	2324.806417	3850.840643	0.331261	0.174187	0.825813	

Regression: Automobile Price Prediction > Evaluate Model > Evaluation results

rows    columns  
2        6

	Negative Log Likelihood	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error	Coefficient of Determination
view as						
Infinity	2669.07619	4178.917271	0.47386	0.294259	0.705741	
1384.706175	1091.958769	1991.194483	0.193863	0.066808	0.933192	

## House Rate Prediction: -

jupyter housing project Last Checkpoint: 07/30/2019 (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 C Logout

```
In [3]: #import libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

In [4]: #import dataset
data = pd.read_csv('housingData-Real.csv')

In [5]: data.head()

Out[5]:
   id      date  price  bedrooms  bathrooms  sqft_living  sqft_lot  floors  waterfront  view ... grade  sqft_above  sqft_basement  yr_built
0  7129300520 20141013T000000 221900.0       3     1.00      1180    5650     1.0        0     0 ...    7    1180        0     1955
1  6414100192 20141209T000000 538000.0       3     2.25      2570    7242     2.0        0     0 ...    7    2170        400    1951
2  5631500400 20150225T000000 180000.0       2     1.00      770    10000     1.0        0     0 ...    6    770        0     1935
3  2487200875 20141209T000000 604000.0       4     3.00      1980    5000     1.0        0     0 ...    7    1050        910    1965
4  1954400510 20150218T000000 510000.0       3     2.00      1680    8080     1.0        0     0 ...    8    1680        0     1987

5 rows × 21 columns
```

```
In [39]: #data.info()

In [7]: # select your columns
livingspace = data['sqft_living']
price = data['price']

In [8]: #convert livingspace into 2D matrix
#X usually is in CAPS
x = np.array(livingspace).reshape(-1, 1)

In [9]: #convert price into 2D matrix
y = np.array(price)

In [10]: y

Out[10]: array([221900., 538000., 180000., ..., 402101., 400000., 325000.])

In [11]: # convert the data into test and training
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=1/3)

In [12]: x_test

Out[12]: array([[3530],
 [2890],
 [1677],
 ...,
 [ 840],
 [4690],
 [1240]], dtype=int64)

In [13]: #pass your data into Linear Regression model
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(x_train, y_train)

Out[13]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
 normalize=False)

In [14]: #create a predictor
predictor = regressor.predict(x_test)

In [15]: predictor
```

jupyter housing project Last Checkpoint: 07/30/2019 (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 C Logout

```
In [9]: #convert price into 2D matrix
y = np.array(price)

In [10]: y

Out[10]: array([221900., 538000., 180000., ..., 402101., 400000., 325000.])

In [11]: # convert the data into test and training
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=1/3)

In [12]: x_test

Out[12]: array([[3530],
 [2890],
 [1677],
 ...,
 [ 840],
 [4690],
 [1240]], dtype=int64)

In [13]: #pass your data into Linear Regression model
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(x_train, y_train)

Out[13]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
 normalize=False)

In [14]: #create a predictor
predictor = regressor.predict(x_test)

In [15]: predictor
```

File Edit View Insert Cell Kernel Widgets Help

Trusted Python 3

```
In [13]: #pass your data into Linear Regression model
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(x_train, y_train)

Out[13]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
normalize=False)
```

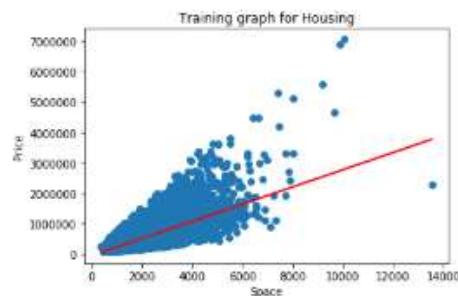
```
In [14]: #create a predictor
predictor = regressor.predict(x_test)
```

```
In [15]: predictor
```

```
Out[15]: array([ 951150.57155768, 770432.35471377, 427914.85935179, ...,
191569.3163856 , 1278702.33958727, 304518.20191305])
```

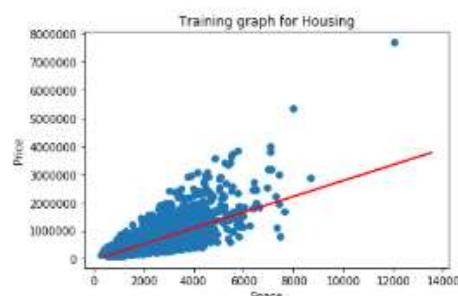
```
In [17]: #plot for training dataset
plt.scatter(x_train, y_train)
plt.plot(x_train, regressor.predict(x_train), color='red')
plt.title('Training graph for Housing')
plt.xlabel('Space')
plt.ylabel('Price')
```

```
Out[17]: Text(0, 0.5, 'Price')
```



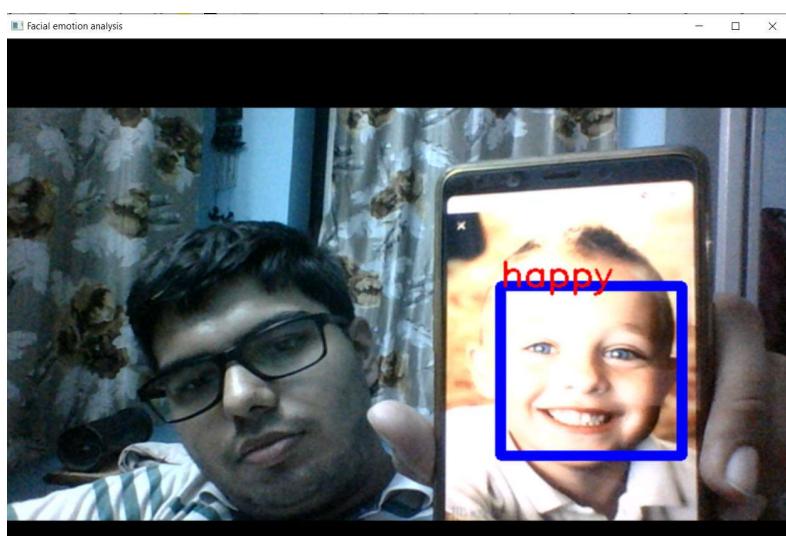
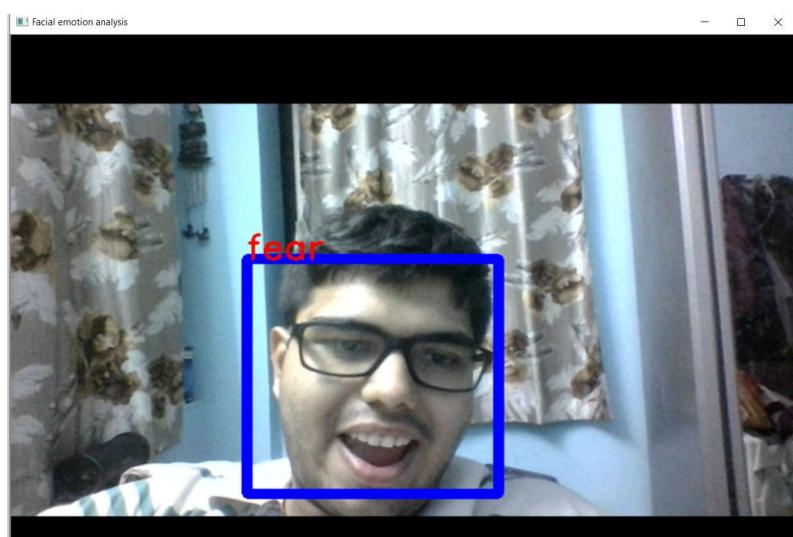
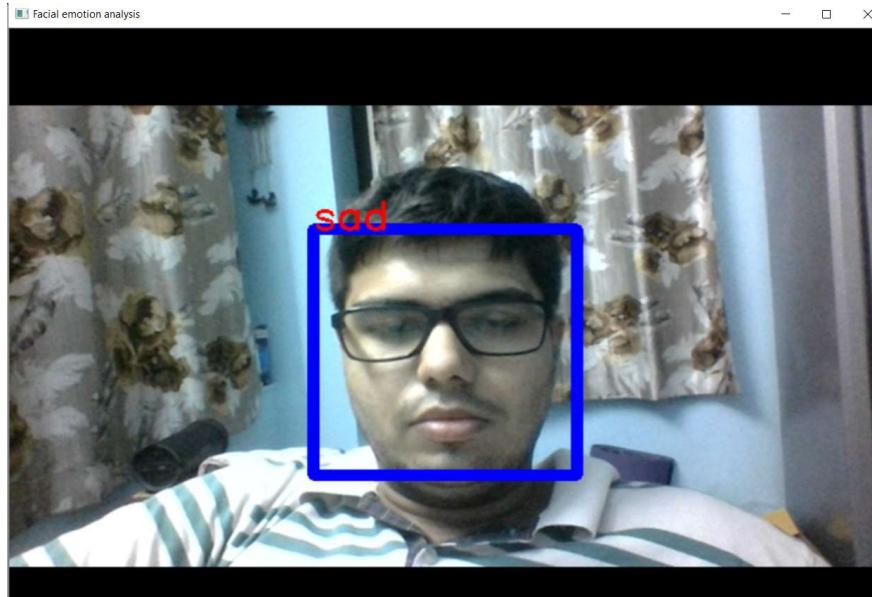
```
In [18]: #plot for test dataset
plt.scatter(x_test, y_test)
plt.plot(x_train, regressor.predict(x_train), color='red')
plt.title('Training graph for Housing')
plt.xlabel('Space')
plt.ylabel('Price')
```

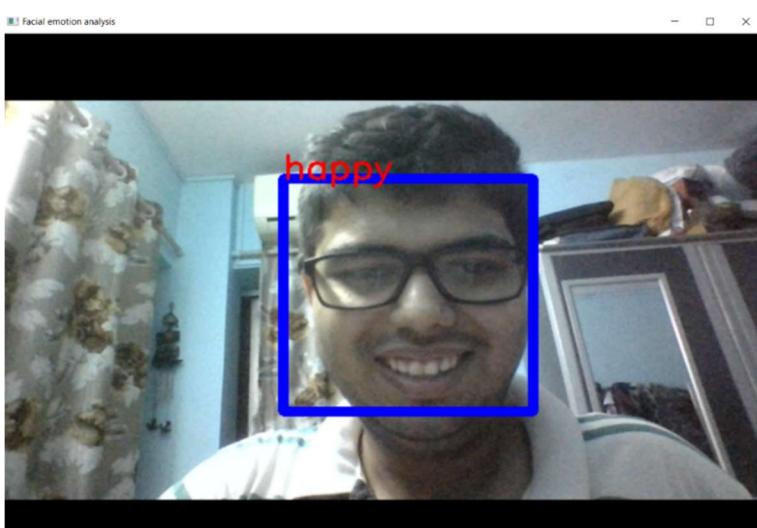
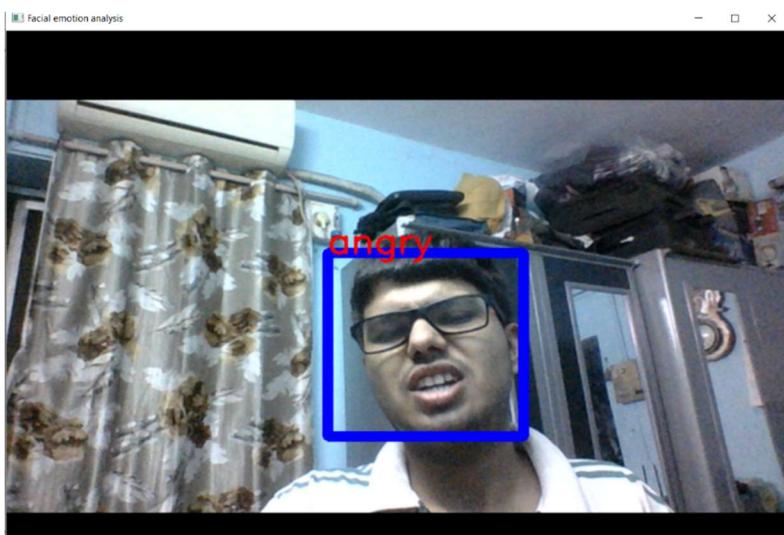
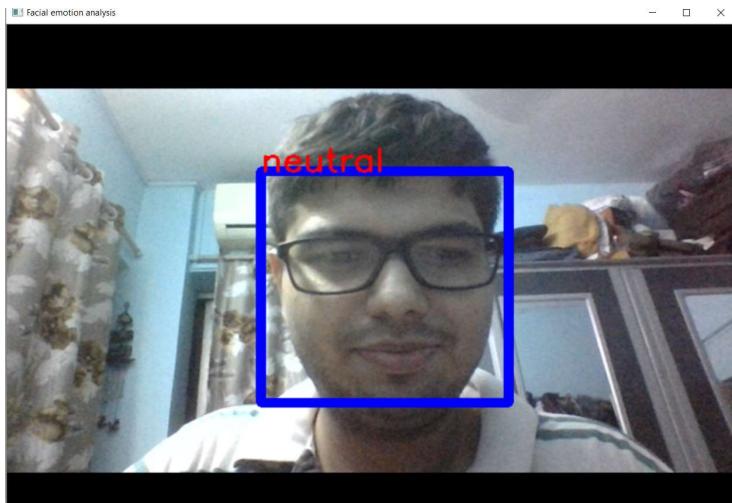
```
Out[18]: Text(0, 0.5, 'Price')
```



```
In [ ]:
```

## AI Model: Customer Mood Detection Using AI





## Google Assistance Query set :-

- Customer preferred duration(Tenure) for deposit
- Customer Tendency while keeping Deposit like FD, RD and daily collection ⓘ
- ▢ Default Fallback Intent
- Default Welcome Intent
- Give details of EMI cards Usage according to Region
- Most used payment gateway
- Tell me deposit status according to region
- Tell me why maximum people stored deposit
- Thank you
- What customer tendency while taking loan
- What is customer house status
- what is loan status with respect to repay
- what is purpose most of the customer usage Credit System or EMI Cards
- What is purpose of customer takes the loan
- What is rate of interest on credit system
- What is the main purpose of most of Transaction
- What types of Loan applicant we have ?
- Which branch customer do maximum digital transaction
- Which income category our customer belongs
- Which rank of Loan Account Customer We have
  
- Which region credit system most used ?
- Which Region have Maximum defaulters in credit system
- Which region have maximum loan count
- Which region used most EMI Card
- Why Gujarat Region customer keep deposit
- why Karnataka Region customers keep deposit
- Why Madhya Pradesh Region (MP) customer keep deposits
- Why Maharashtra Region customer keep Deposit ⓘ



## **Chapter 5**

### **Feature Enhancements**

At this movement my project is divide into four parts. First is Analytics, Second Machine Learning Model and Third is AI Model and last but not least Google Assistants support. After getting actual data from bank the analysis is changed because its dummy data generated by me and many attributes are not cover in this. Also after getting data we need to re-train model or may be do some changes in it.

- Connecting main Banking application to analytics.
- Connecting Analysis + ML Model +AI Model + Google Assistant together in one Application
- Adding Customer Wealth Management Model (currently in Alpha Stage).
- Complete Deployment on Cloud.

## **Chapter 6**

### **Summary**

In this project I have done data analysis, creating Machine Learning and Artificial Intelligence models. This project is fully functional and it is capable of working in live environment.

It is very helpful to understand the customer and their needs also understand the region wise diversity. It helps to plan and building strategies, also to capture market.

These ML and AI models definitely help to understand the customer more deeply and make more grow in the field.

## Chapter 7

### References and Bibliography

- <https://numpy.org/>
- <https://pandas.pydata.org/>
- <https://scikit-learn.org/>
- <https://matplotlib.org/>
- <https://seaborn.pydata.org/>
- <https://www.tensorflow.org/>
- <http://keras.io/>
- [https://opencv-python-tutorials.readthedocs.io/en/latest/py\\_tutorials/py\\_setup/py\\_intro/py\\_intro.html](https://opencv-python-tutorials.readthedocs.io/en/latest/py_tutorials/py_setup/py_intro/py_intro.html)
- <https://www.w3schools.com/w3css/>
- <https://stackoverflow.com/>
- <https://www.google.com/>
- <https://www.youtube.com/>

## Chapter 8

### Index & Acronyms

#### A

**Account** :- an arrangement made with a bank whereby one may deposit and withdraw money and in some cases be paid interest.

**ATM Department** :- It is for resolving queries credit and debit Transactions, Mobile banking, ATM related queries, net banking queries and many more.

#### B

**Balance** :- The amount of money in your bank account.

#### C

**CASA** :- Current and Saving Account Balance

**Credit** :- A credit is an **accounting** entry that either increases a liability or equity account, or decreases an asset or expense account.

**Current Account** :- Business Account only for business purpose and don't have ROI

**Credit Department** :- It involves Loan and Credit System departments.

#### D

**Debit** :- A **debit** is an **accounting** entry that either increases an asset or expense account, or decreases a liability or equity account. It is positioned to the left in an **accounting** entry.

**Deposit** :- A deposit is the act of placing cash with some entity, most commonly with a financial institution such as a bank.

#### E

**ECS** :- ECS stands for Electronic Clearing Service. It is an electronic mode of funds transfer from one bank account to another bank account. It also facilitates electronic credit/debit transaction

associated with customer's account. It is usually used for transactions that are repetitive or periodic in nature.

**EMI** :- An equated monthly instalment (EMI) is a fixed payment amount made by a borrower to a lender at a specified date each calendar month. Equated monthly instalments are used to pay off both interest and principal each month, so that over a specified number of years the loan is paid off in full.

## F

**FD**:- A fixed deposit (FD) is a financial instrument provided by banks or NBFCs which provides investors a higher rate of interest than a regular savings account, until the given maturity date.

**FDR** :- FD and RD Department.

## I

**Insurance** :- an arrangement by which a company or the state undertakes to provide a guarantee of compensation for specified loss, damage, illness, or death in return for payment of a specified premium.

**Investment Department** :- It involves Gold and SIP

**Insurance Department** :- It involves Insurance department.

## L

**Loan** :- A **loan** means borrow money, against property, or other material goods given to another party in exchange for future repayment of the **loan** value or principal amount, along with interest or finance charges.

## N

**Net-worth** :- Current Account have high amount balance or maximum loan take by a business account

## O

**Operational Department :-** involves Saving and Current Accounts.

## R

**ROI :-** Rate of Interest

**Rate :-** Rate of Interest

**RD :-** A recurring deposit is a special kind of term deposit offered by banks which help people with regular incomes to deposit a fixed amount every month into their recurring deposit account and earn interest at the rate applicable to fixed deposits.

## S

**Saving Account :-** For general account type any one open/create a saving account. It has ROI

**SIP :-** A Systematic Investment Plan or SIP is a mode for investing money in mutual funds. SIP allows you to invest a certain pre-determined amount at a regular interval (weekly, monthly, quarterly, etc.).

## T

**Tenure :-** Time Period

**Transaction/Tx :-** a financial event that causes a change in financial position means transferring amt from one account to another count

## W

**Withdrawal:-** Taking money from your bank account