DEPARTMENT OF COMPUTER SCIENCE
BANGALORE YESWANTHPUR CAMPUS

M.Sc. DATA SCIENCE

Project Report
on

CUSTOMER CHURN ANALYSIS

Submitted by
Blessy Louis
M.Sc. Data Science 'B'
2348416

MAY 2024

TABLE OF CONTENTS

ABSTRACT

This project aims to address the challenge of customer churn in the banking industry by analysing a dataset and developing predictive models to identify customers likely to churn. We first perform data pre-processing tasks, including handling missing values and feature scaling, to prepare the dataset. Through exploratory data analysis (EDA), we visualize the distribution of various features and prominently display the overall churn rate to highlight its significance. Subsequently, we train logistic regression and random forest models to predict churn, evaluating their performance using accuracy, precision, recall, and F1 score metrics. Finally, we create a user-friendly interface for predicting churn based on customer features, empowering stakeholders to proactively retain at-risk customers. This project offers valuable insights and predictive capabilities to assist businesses in mitigating churn and enhancing customer retention strategies in the banking sector.

AIM

The aim of this project is to develop a predictive model to identify customers at risk of churning in the banking industry, with a focus on enhancing customer retention strategies. Through comprehensive data analysis and machine learning techniques, the project aims to provide actionable insights into customer behaviour and factors influencing churn. By leveraging predictive models, the project seeks to enable businesses to proactively address churn, thereby improving customer satisfaction and loyalty while reducing revenue loss associated with customer attrition.

Objective

The objectives of this project are as follows:

- Data Acquisition and Pre-processing: Gather and preprocess customer data to ensure data quality and consistency for analysis.
- Exploratory Data Analysis (EDA): Conduct an in-depth analysis of customer demographics, behaviours, and churn patterns to uncover insights and trends.
- Model Building and Evaluation: Develop machine learning models, including Logistic Regression and Random Forest, to predict customer churn based on relevant features extracted from the data. Evaluate model performance using metrics such as accuracy, precision, recall, and F1 score.

- Customer Prediction: Implement a user-friendly interface for users to input customer information and obtain predictions regarding the likelihood of churn, enabling businesses to take proactive retention actions.
- Documentation and Reporting: Compile a detailed document or report summarizing the project's aim, methodology, code implementation, results, and insights derived from the analysis.

CHAPTER – 01

INTRODUCTION

This project focuses on analysing customer churn in a financial services context, aiming to provide insights into customer behaviour and facilitate proactive retention strategies. Customer churn, the phenomenon where customers discontinue their relationship with a company, poses significant challenges to businesses, including revenue loss and decreased customer satisfaction. Through this project, we aim to understand the factors influencing customer churn and develop predictive models to identify customers at risk of churning. By leveraging machine learning techniques such as Logistic Regression and Random Forest, we seek to build accurate models that can forecast customer churn based on various features such as credit score, tenure, balance, and demographic information. The project involves comprehensive data acquisition, preprocessing, exploratory data analysis, model building, and evaluation. Ultimately, the findings and models generated will empower businesses to proactively engage with at-risk customers, optimize retention efforts, and enhance overall customer satisfaction and loyalty.

CHAPTER – 02

DATA DESCRIPTION

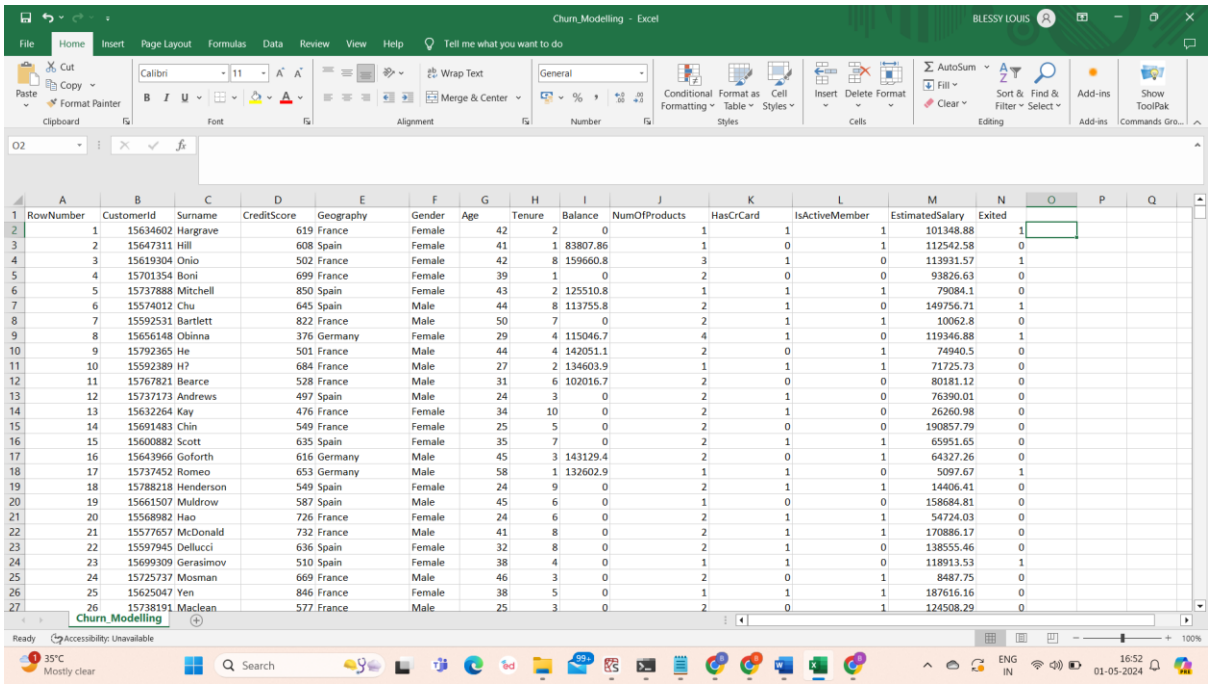
The dataset was taken from “Kaggle.com” for this particular project, the data can be verified from https://www.kaggle.com/code/d4rklucif3r/churn-modelling-deployment-luciferml/input

This dataset comprises customer information from a financial institution, including demographic details such as gender, age, and geographical location, as well as financial attributes like credit score, account balance, and tenure with the institution. The dataset also includes binary indicators for whether the customer holds a credit card, is an active member, and has churned or exited the relationship with the institution. With these features, the dataset offers insights into customer behaviour, preferences, and churn patterns, making it suitable for various analytical tasks such as customer segmentation, predictive modelling, and churn analysis.

FEATURES:

1. RowNumber: This represents the row number in the dataset, serving as a unique identifier for each record.
2. CustomerId: This is a unique identifier assigned to each customer, enabling the tracking of individual customer profiles.
3. Surname: This feature denotes the surname or last name of the customer, providing additional personal identification information.
4. CreditScore: This numeric attribute reflects the credit score of each customer, which is a measure of their creditworthiness based on various factors such as credit history and financial behaviour.
5. Geography: This categorical variable indicates the geographical location of the customer, providing insight into regional demographics and preferences.
6. Gender: This categorical attribute specifies the gender of the customer, allowing for gender-based analysis and segmentation.
7. Age: This numeric attribute represents the age of the customer, which is a crucial demographic factor influencing financial behaviour and preferences.
8. Tenure: This numeric feature indicates the number of years the customer has been associated with the financial institution, reflecting customer loyalty and longevity.

9. Balance: This numeric attribute denotes the account balance of the customer, representing the amount of funds available in their account.
10. NumOfProducts: This numeric feature indicates the number of financial products or services held by the customer, providing insights into their level of engagement with the institution.
11. HasCrCard: This binary attribute indicates whether the customer holds a credit card with the institution (1 for yes, 0 for no), offering insights into their banking behavior and preferences.
12. IsActiveMember: This binary feature indicates whether the customer is an active member of the institution (1 for yes, 0 for no), reflecting their level of engagement and participation.
13. EstimatedSalary: This numeric attribute represents the estimated salary of the customer, providing insights into their financial capacity and spending potential.
14. Exited: This binary target variable indicates whether the customer has churned or exited the relationship with the institution (1 for churned, 0 for retained), serving as the outcome variable for predictive modelling and analysis.



RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
1	15634602	Hargrave	619	France	Female	42	2	0	1	1	1	101348.88	1
2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
3	15619304	Onio	502	France	Female	42	8	159660.8	3	1	0	113931.57	1
4	15701354	Boni	699	France	Female	39	1	0	2	0	0	93826.63	0
5	15737888	Mitchell	850	Spain	Female	43	2	125510.8	1	1	1	79084.1	0
6	15574012	Chu	645	Spain	Male	44	8	113755.8	2	1	0	149756.71	1
7	15592531	Bartlett	822	France	Male	50	7	0	2	1	1	10062.8	0
8	15656148	Obinna	376	Germany	Female	29	4	115046.7	4	1	0	119346.88	1
9	15792365	He	501	France	Male	44	4	142051.1	2	0	1	74940.5	0
10	15592389	H?	684	France	Male	27	2	134603.9	1	1	1	71725.73	0
11	15767821	Bearce	528	France	Male	31	6	102016.7	2	0	0	80181.12	0
12	15737173	Andrews	497	Spain	Male	24	3	0	2	1	0	76390.01	0
13	15632264	Kay	476	France	Female	34	10	0	2	1	0	26260.98	0
14	15691483	Chin	549	France	Female	25	5	0	2	0	0	190857.79	0
15	15600882	Scott	635	Spain	Female	35	7	0	2	1	1	65951.65	0
16	15643966	Goforth	616	Germany	Male	45	3	143129.4	2	0	1	64327.26	0
17	15737452	Romeo	653	Germany	Male	58	1	132602.9	1	1	0	5097.67	1
18	15788218	Henderson	549	Spain	Female	24	9	0	2	1	1	14406.41	0
19	15661507	Muldrow	587	Spain	Male	45	6	0	1	0	0	158684.81	0
20	15568982	Hao	726	France	Female	24	6	0	2	1	1	54724.03	0
21	15577657	McDonald	732	France	Male	41	8	0	2	1	1	170886.17	0
22	15597945	Dellucci	636	Spain	Female	32	8	0	2	1	0	138555.46	0
23	15699309	Gerasimov	510	Spain	Female	38	4	0	1	1	0	118913.53	1
24	15725737	Mosman	669	France	Male	46	3	0	2	0	1	8487.75	0
25	15625047	Yen	846	France	Female	38	5	0	1	1	1	187616.16	0
26	15738191	Maclean	577	France	Male	25	3	0	2	0	1	124508.29	0

CHAPTER – 03

PROCEDURE

The procedure involved in this project can be elaborated as follows:

- **Data Acquisition and Pre-processing:** The initial step involves acquiring the dataset containing customer information from a financial institution. The dataset is then pre-processed to handle missing values, which are often filled using mean or median imputation methods. Categorical variables are encoded using one-hot encoding to convert them into a numerical format suitable for analysis. Additionally, feature scaling is applied to standardize numerical features, ensuring that they are on a similar scale.
- **Exploratory Data Analysis (EDA):** Following pre-processing, exploratory data analysis (EDA) is conducted to gain insights into the dataset's characteristics. Various visualizations such as pie charts, histograms, and box plots are utilized to explore the distributions and relationships between different variables. EDA helps in identifying trends, patterns, outliers, and potential correlations within the data.
- **Model Building and Evaluation**: After gaining a comprehensive understanding of the dataset, predictive models are built to analyse customer churn. Two popular algorithms, namely Logistic Regression and Random Forest Classifier, are trained using the pre-processed data. The models are evaluated using performance metrics such as accuracy, precision, recall, and F1 score, along with a confusion matrix to assess their predictive capabilities.
- **Customer Prediction**: In this phase, the trained models are applied to make predictions on new customer data. Users can input customer attributes such as credit score, tenure, and account balance through a user-friendly interface. The models predict the likelihood of a customer churning based on these input features, providing valuable insights for customer retention strategies.

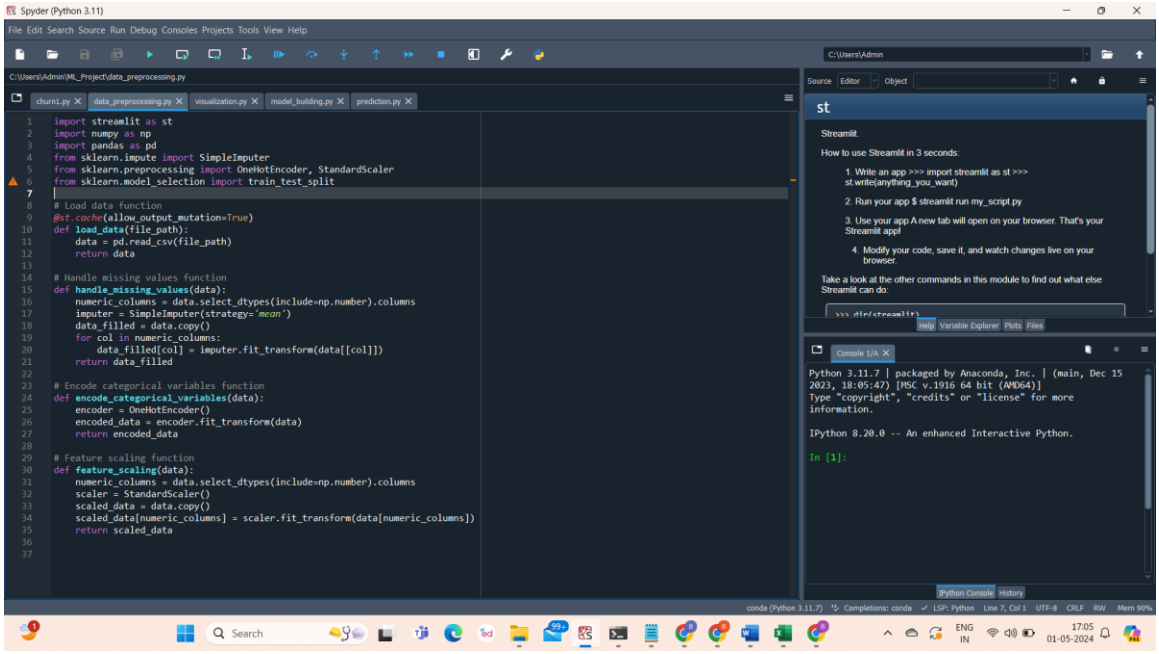
Throughout the procedure, the project emphasizes a systematic approach, leveraging both descriptive and predictive analytics techniques to analyse customer churn and inform decision-making in financial institutions.

CHAPTER – 04

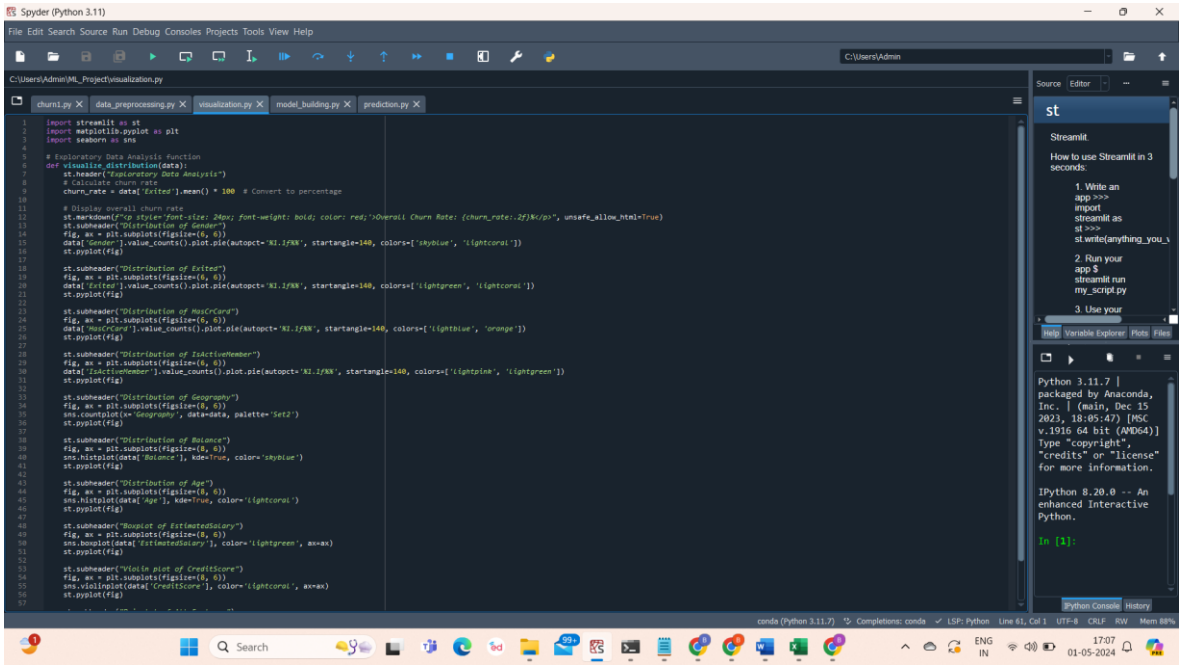
CODE

The project involves a Modular approach of the Churn prediction Analysis:

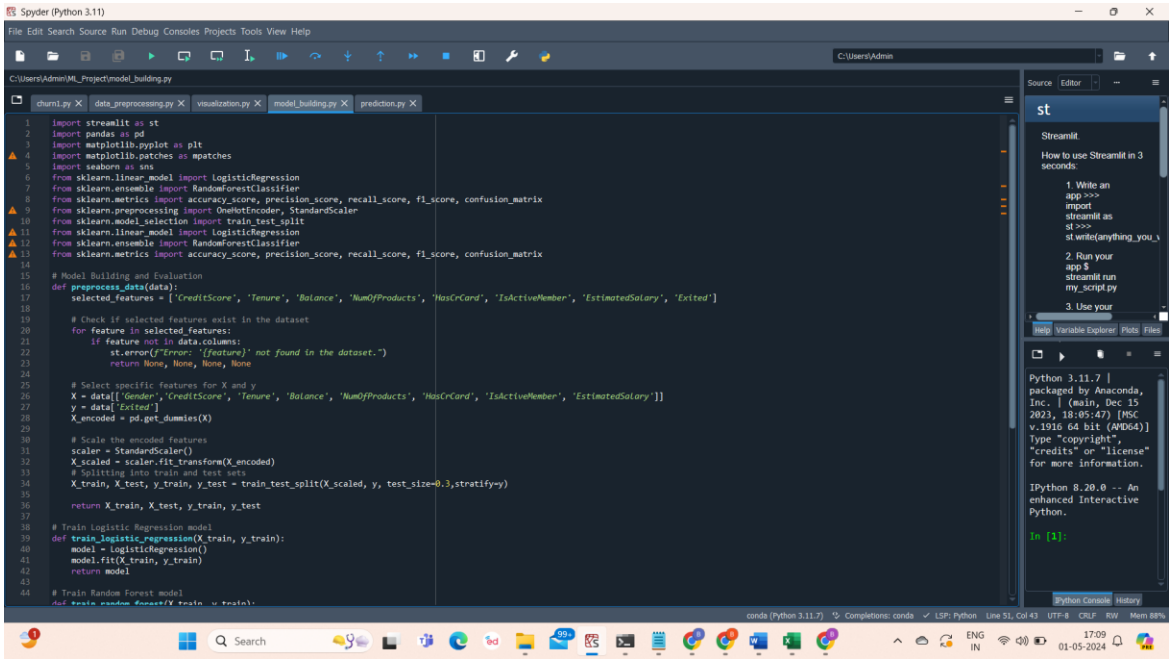
Data_preprocessing.py:



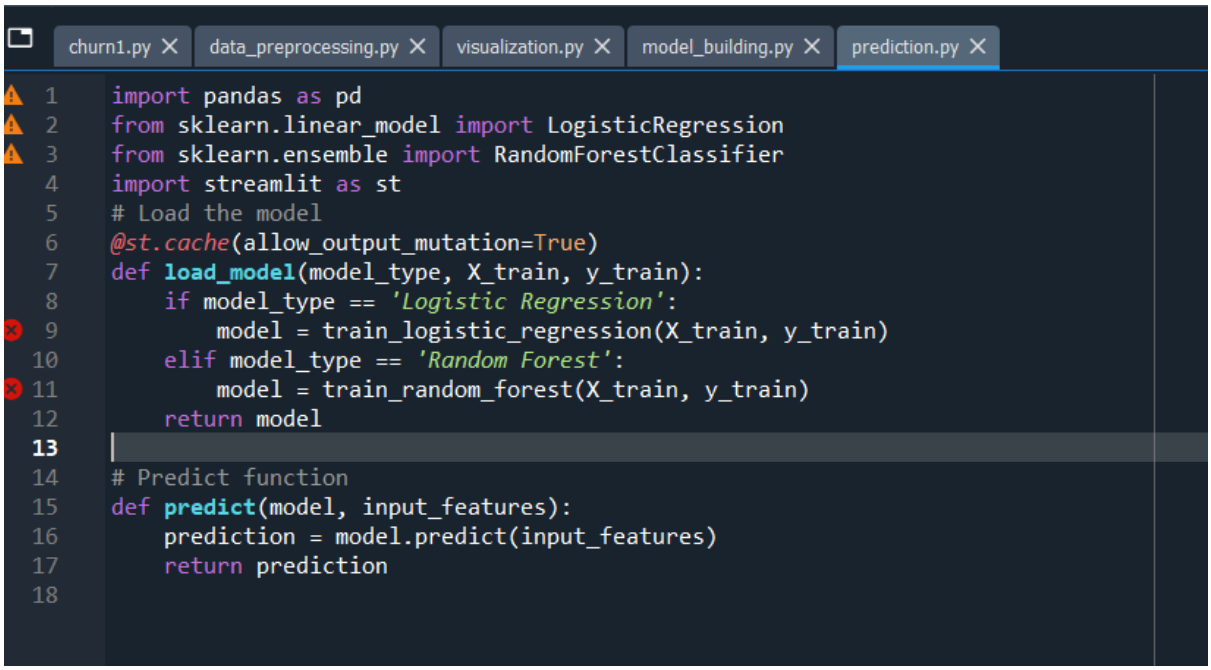
Visualization.py:



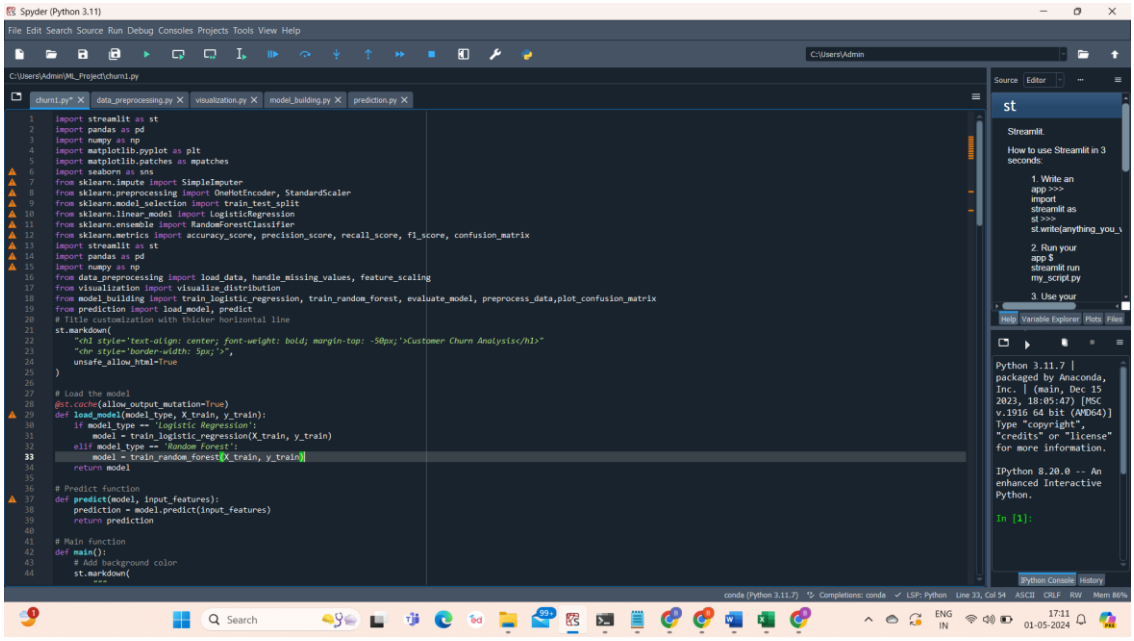
Model_building.py:

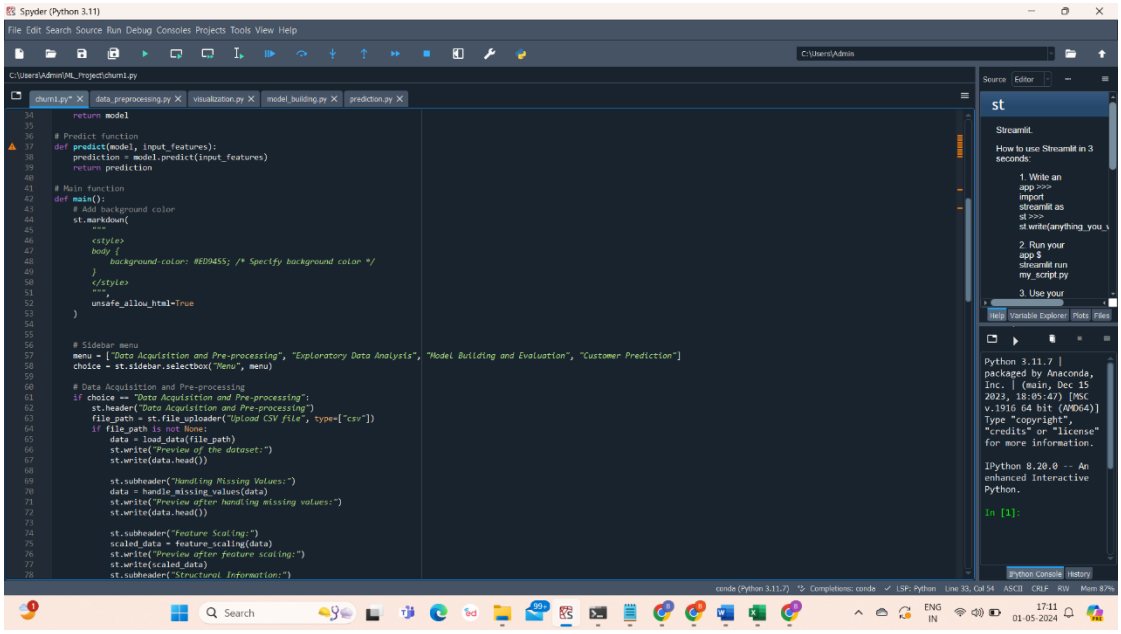


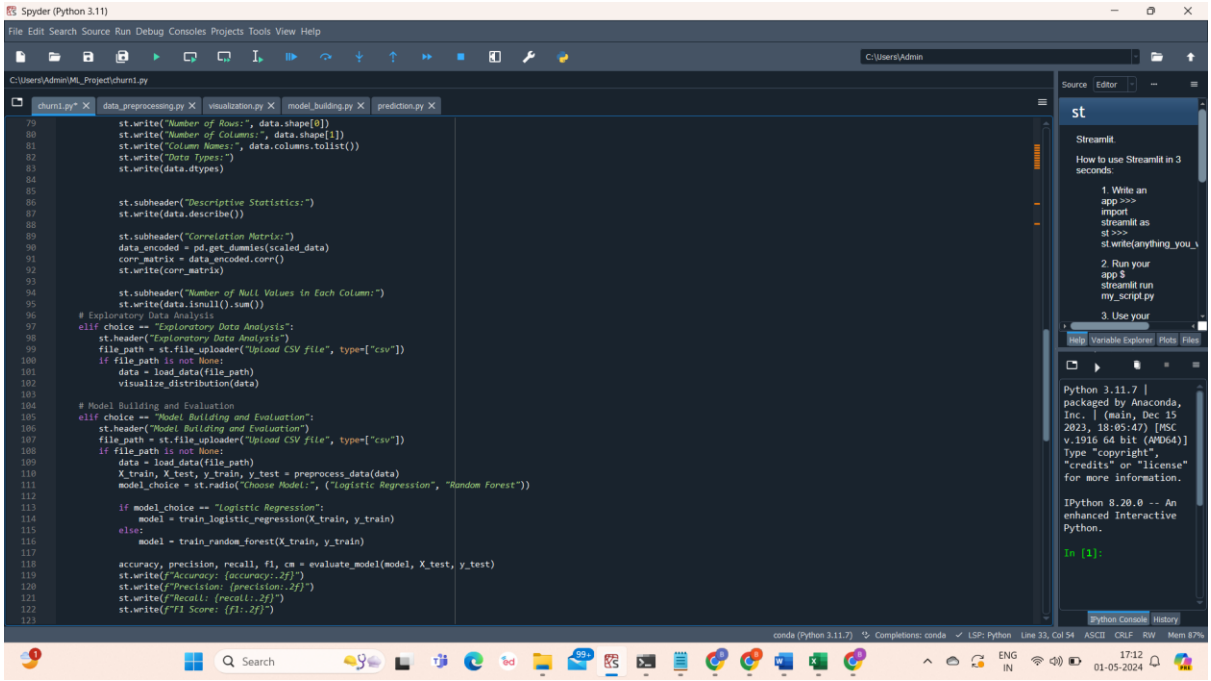
Prediction.py:

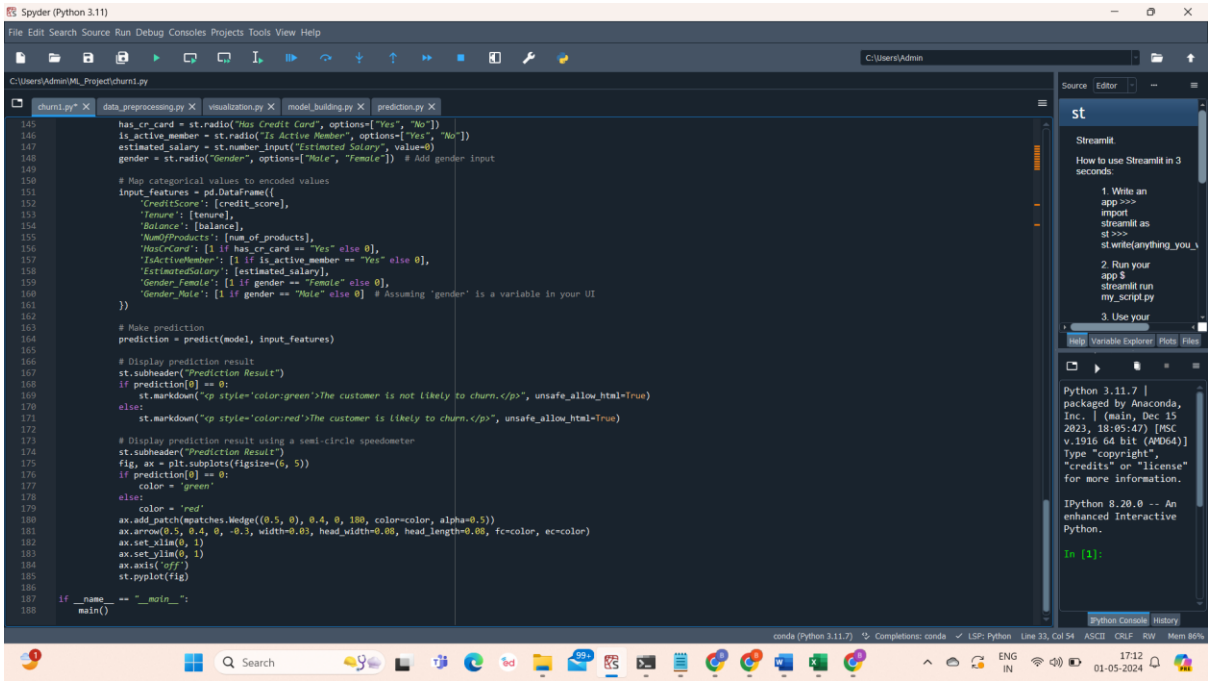


Churn1.py: (Main)





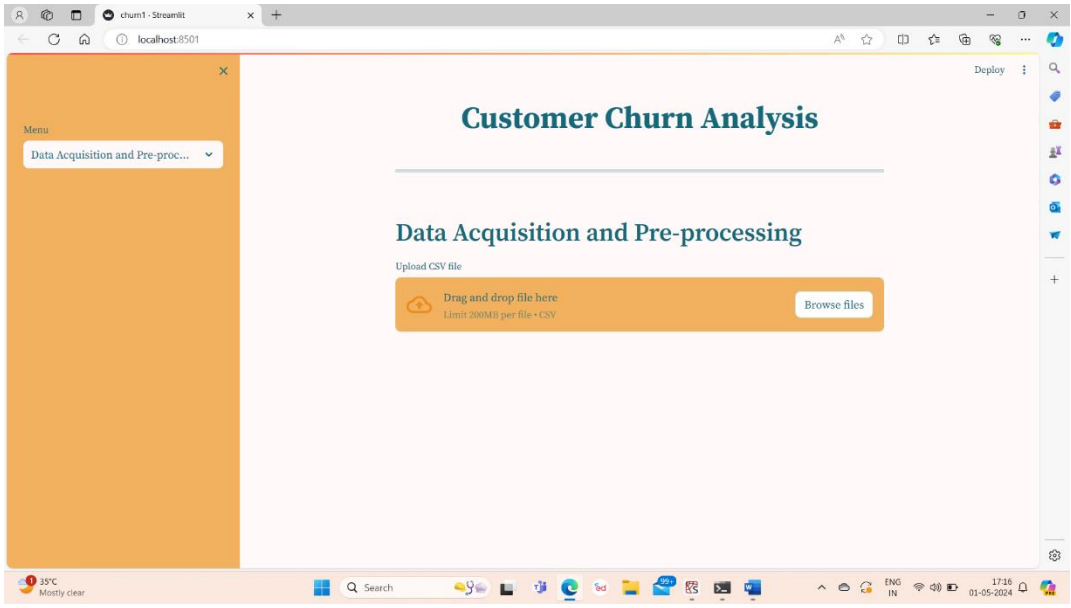

Spyder (Python 3.11)
File Edit Search Source Run Debug Consoles Projects Tools View Help
C:\Users\Admin\ML_Project\churn1.py
churn1.py* data_preprocessing.py visualization.py model_building.py prediction.py
st.write("Number of Rows:", data.shape[0])
st.write("Number of Columns:", data.shape[1])
st.write("Column Names:", data.columns.tolist())
st.write("Data Types:")
st.write(data.dtypes)
st.subheader("Descriptive Statistics:")
st.write(data.describe())
st.subheader("Correlation Matrix:")
data_encoded = pd.get_dummies(scaled_data)
corr_matrix = data_encoded.corr()
st.write(corr_matrix)
st.subheader("Number of Null Values in Each Column:")
st.write(data.isnull().sum())
# Exploratory Data Analysis
elif choice == "Exploratory Data Analysis":
st.header("Exploratory Data Analysis")
file_path = st.file_uploader("Upload CSV file", type=["csv"])
if file_path is not None:
data = load_data(file_path)
visualize_distribution(data)
# Model Building and Evaluation
elif choice == "Model Building and Evaluation":
st.header("Model Building and Evaluation")
file_path = st.file_uploader("Upload CSV file", type=["csv"])
if file_path is not None:
data = load_data(file_path)
X_train, X_test, y_train, y_test = preprocess_data(data)
model_choice = st.radio("Choose Model:", ("Logistic Regression", "Random Forest"))
if model_choice == "Logistic Regression":
model = train_logistic_regression(X_train, y_train)
else:
model = train_random_forest(X_train, y_train)
accuracy, precision, recall, f1, cm = evaluate_model(model, X_test, y_test)
st.write(f"Accuracy: {accuracy:.2f}")
st.write(f"Precision: {precision:.2f}")
st.write(f"Recall: {recall:.2f}")
st.write(f"F1 Score: {f1:.2f}")
st
Streamlit
How to use Streamlit in 3 seconds:
1. Write an app >>> import streamlit as st >>> st.write(anything_you_
2. Run your app $ streamlit run my_script.py
3. Use your
Python 3.11.7 | packaged by Anaconda, Inc. | (main, Dec 15 2023, 18:05:47) [MSC v.1916 64 bit (AMD64)] Type "copyright", "credits" or "license" for more information.
IPython 8.20.0 -- An enhanced Interactive Python.
In [1]:
Search
ENG IN
17:12 01-05-2024

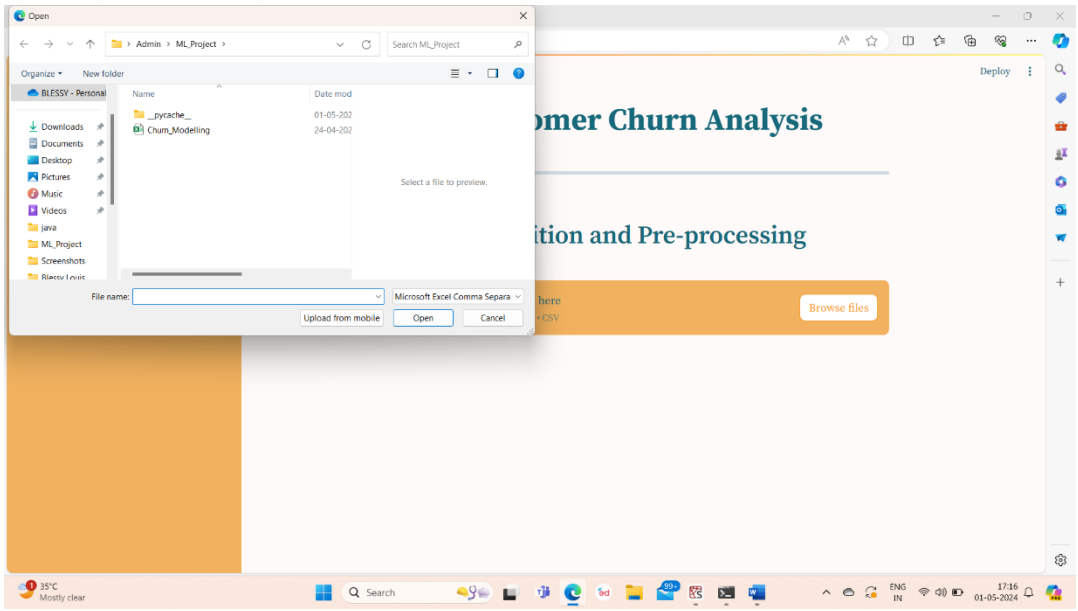

Spyder (Python 3.11)
File Edit Search Source Run Debug Consoles Projects Tools View Help
C:\Users\Admin\ML_Project\churn1.py
churn1.py* data_preprocessing.py visualization.py model_building.py prediction.py
has_cr_card = st.radio("Has Credit Card", options=["Yes", "No"])
is_active_member = st.radio("Is Active Member", options=["Yes", "No"])
estimated_salary = st.number_input("Estimated Salary", value=0)
gender = st.radio("Gender", options=["Male", "Female"]) # Add gender input
# Map categorical values to encoded values
input_features = pd.DataFrame({
'CreditScore': [credit_score],
'Tenure': [tenure],
'Balance': [balance],
'NumOfProducts': [num_of_products],
'HasCrCard': [1 if has_cr_card == "Yes" else 0],
'IsActiveMember': [1 if is_active_member == "Yes" else 0],
'EstimatedSalary': [estimated_salary],
'Gender_Female': [1 if gender == "Female" else 0],
'Gender_Male': [1 if gender == "Male" else 0] # Assuming 'gender' is a variable in your UI
})
# Make prediction
prediction = predict(model, input_features)
# Display prediction result
st.subheader("Prediction Result")
if prediction[0] == 0:
st.markdown(" The customer is not likely to churn. ", unsafe_allow_html=True)
else:
st.markdown(" The customer is likely to churn. ", unsafe_allow_html=True)
# Display prediction result using a semi-circle speedometer
st.subheader("Prediction Result")
fig, ax = plt.subplots(figsize=(6, 5))
if prediction[0] == 0:
color = 'green'
else:
color = 'red'
ax.add_patch(mpatches.Wedge((0.5, 0), 0.4, 0, 180, color=color, alpha=0.5))
ax.arrow(0.5, 0.4, 0, -0.3, width=0.03, head_width=0.08, head_length=0.08, fc=color, ec=color)
ax.set_xlim(0, 1)
ax.set_ylim(0, 1)
ax.axis('off')
st.pyplot(fig)
if __name__ == "__main__":
main()
st
Streamlit
How to use Streamlit in 3 seconds:
1. Write an app >>> import streamlit as st >>> st.write(anything_you_
2. Run your app $ streamlit run my_script.py
3. Use your
Python 3.11.7 | packaged by Anaconda, Inc. | (main, Dec 15 2023, 18:05:47) [MSC v.1916 64 bit (AMD64)] Type "copyright", "credits" or "license" for more information.
IPython 8.20.0 -- An enhanced Interactive Python.
In [1]:
Search
ENG IN
17:12 01-05-2024

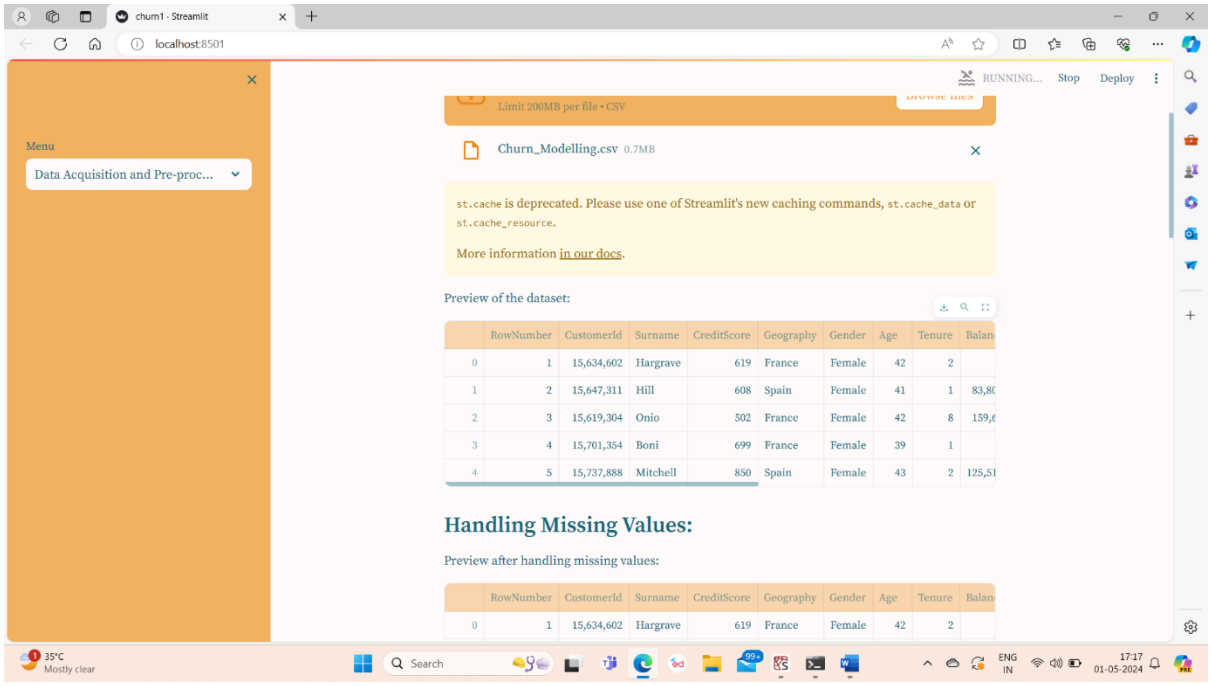
CHAPTER – 05

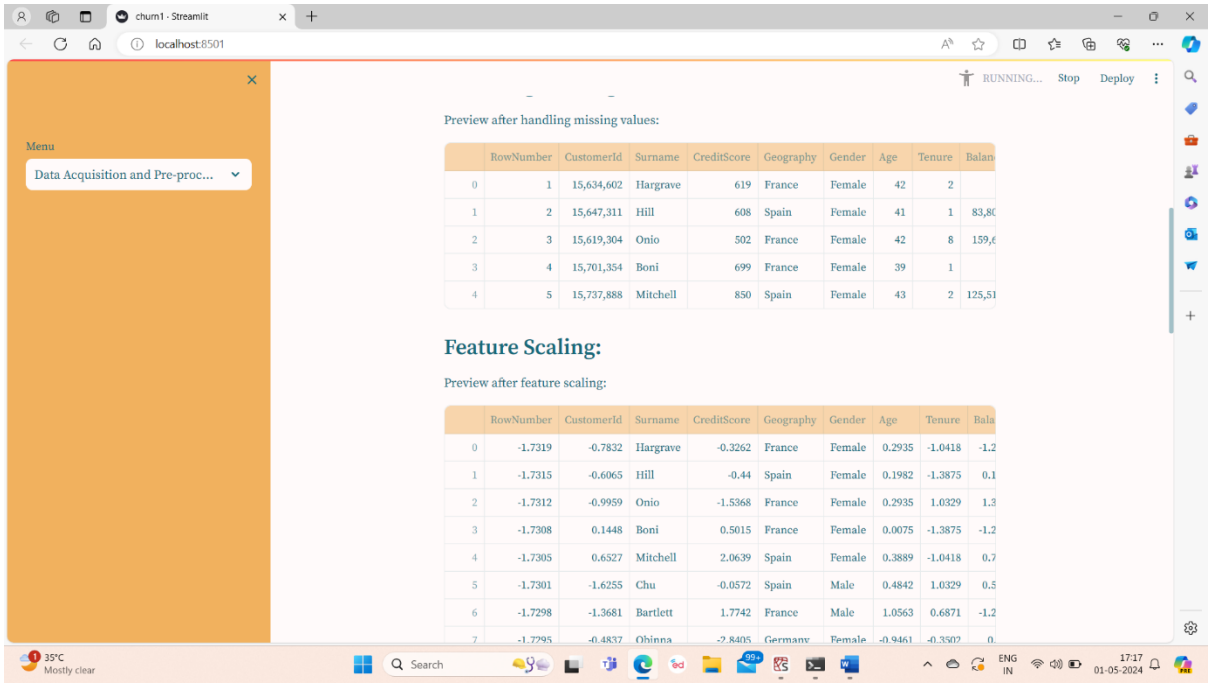
OUTPUT

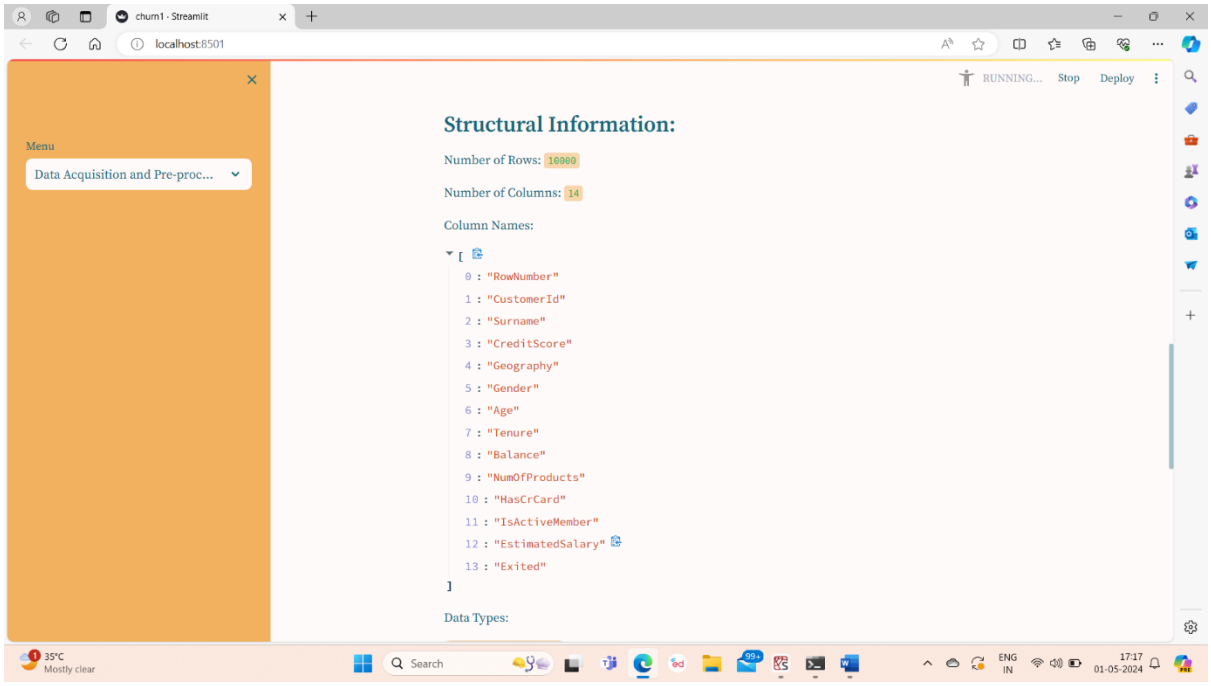
```
(base) C:\Users\Admin\ML_Project>streamlit run churn1.py

  You can now view your Streamlit app in your browser.

  Local URL: http://localhost:8501
  Network URL: http://192.168.0.103:8501
```

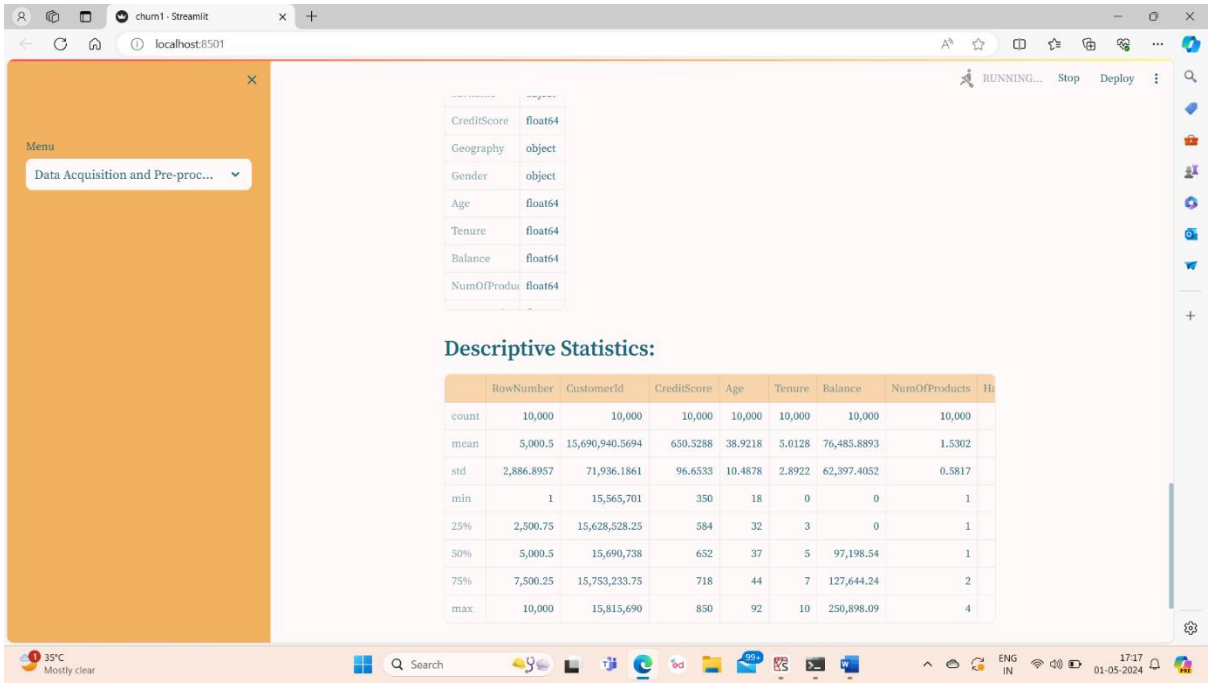


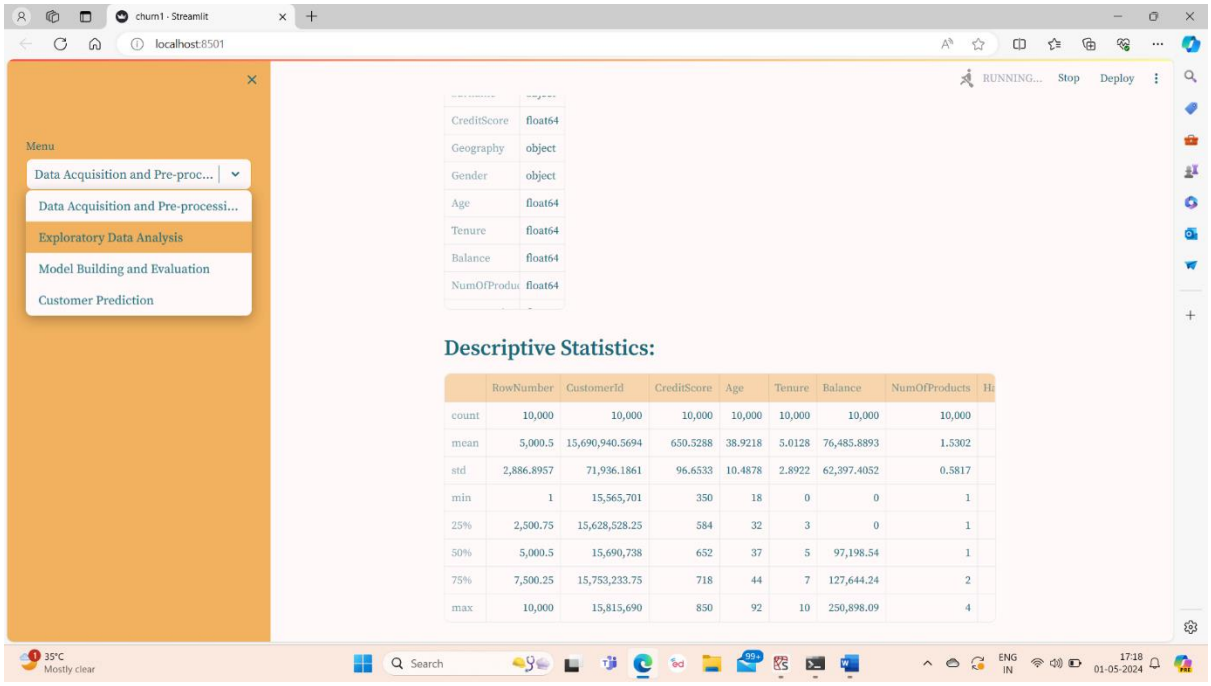


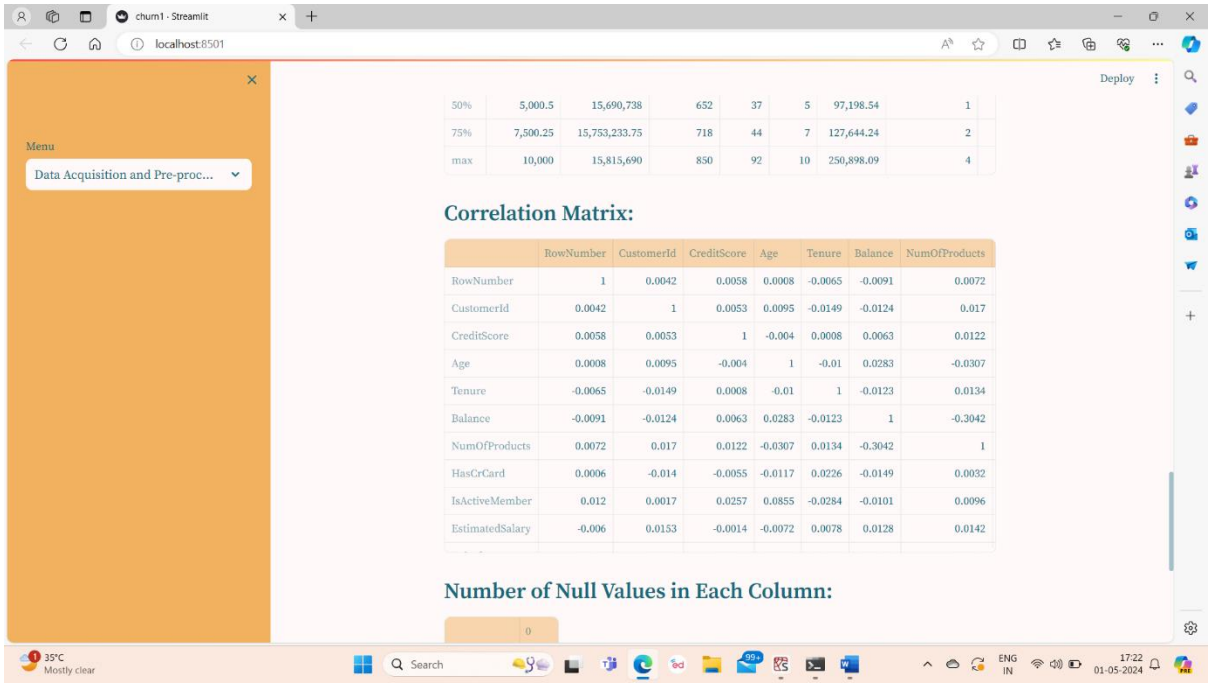

churn1 - Streamlit
localhost:8501
RUNNING...
Stop
Deploy
Menu
Data Acquisition and Pre-proc...
Limit 200MB per file • CSV
Churn_Modelling.csv 0.7MB
st.cache is deprecated. Please use one of Streamlit's new caching commands, st.cache_data or st.cache_resource.
More information in our docs.
Preview of the dataset:
RowNumber CustomerId Surname CreditScore Geography Gender Age Tenure
0 1 15,634,602 Hargrave 619 France Female 42 2
1 2 15,647,311 Hill 608 Spain Female 41 1
2 3 15,619,304 Onio 502 France Female 42 8
3 4 15,701,354 Boni 699 France Female 39 1
4 5 15,737,888 Mitchell 850 Spain Female 43 2
Handling Missing Values:
Preview after handling missing values:
RowNumber CustomerId Surname CreditScore Geography Gender Age Tenure
0 1 15,634,602 Hargrave 619 France Female 42 2
35°C
Mostly clear
Search
ENG
IN
17:17
01-05-2024

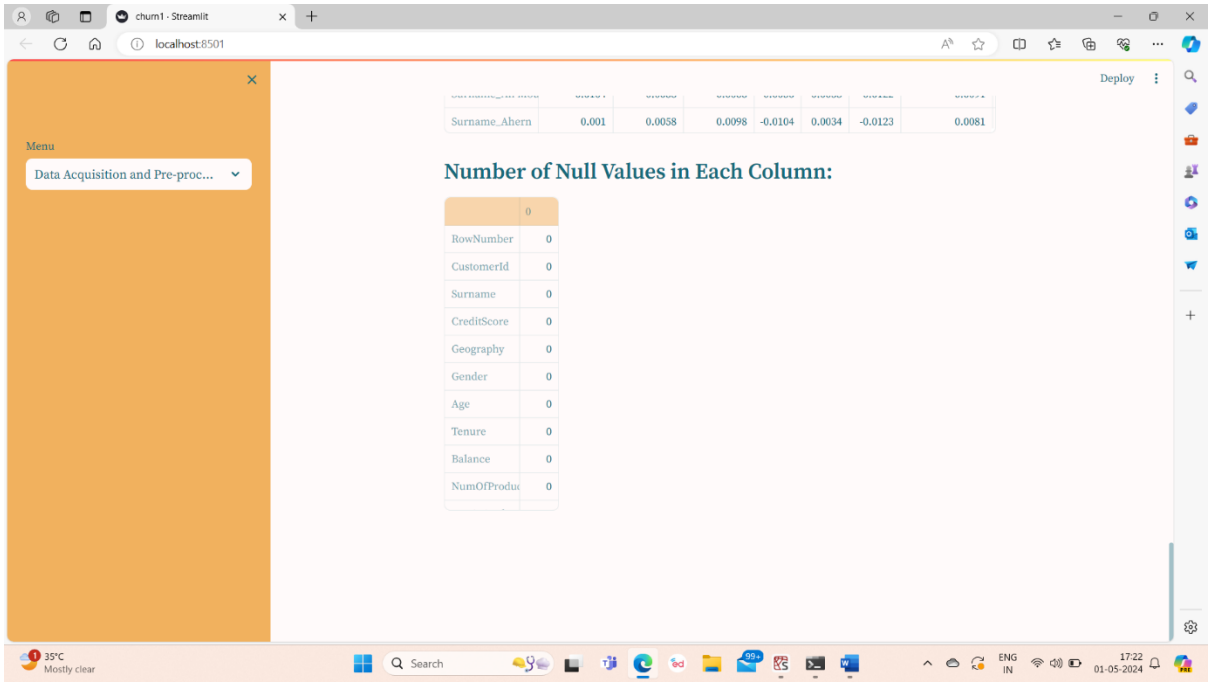

churn1 - Streamlit
localhost:8501
RUNNING...
Stop
Deploy
Menu
Data Acquisition and Pre-proc...
Preview after handling missing values:
RowNumber CustomerId Surname CreditScore Geography Gender Age Tenure
0 1 15,634,602 Hargrave 619 France Female 42 2
1 2 15,647,311 Hill 608 Spain Female 41 1
2 3 15,619,304 Onio 502 France Female 42 8
3 4 15,701,354 Boni 699 France Female 39 1
4 5 15,737,888 Mitchell 850 Spain Female 43 2
Feature Scaling:
Preview after feature scaling:
RowNumber CustomerId Surname CreditScore Geography Gender Age Tenure
0 -1.7319 -0.7832 Hargrave -0.3262 France Female 0.2935 -1.0418
1 -1.7315 -0.6065 Hill -0.44 Spain Female 0.1982 -1.3875
2 -1.7312 -0.9959 Onio -1.5368 France Female 0.2935 1.0329
3 -1.7308 0.1448 Boni 0.5015 France Female 0.0075 -1.3875
4 -1.7305 0.6527 Mitchell 2.0639 Spain Female 0.3889 -1.0418
5 -1.7301 -1.6255 Chu -0.0572 Spain Male 0.4842 1.0329
6 -1.7298 -1.3681 Bartlett 1.7742 France Male 1.0563 0.6871
35°C
Mostly clear
Search
ENG
IN
17:17
01-05-2024

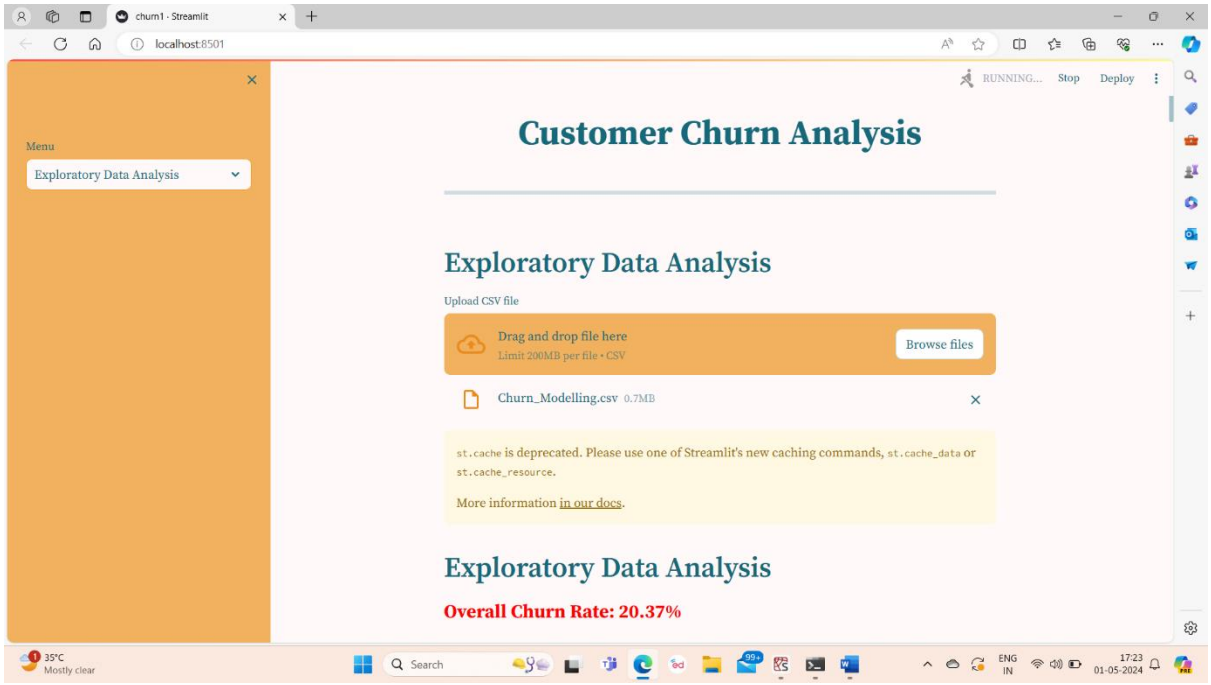

churn1 - Streamlit
localhost:8501
RUNNING...
Stop
Deploy
Menu
Data Acquisition and Pre-proc...
Structural Information:
Number of Rows: 10000
Number of Columns: 14
Column Names:
[
0 : "RowNumber"
1 : "CustomerId"
2 : "Surname"
3 : "CreditScore"
4 : "Geography"
5 : "Gender"
6 : "Age"
7 : "Tenure"
8 : "Balance"
9 : "NumOfProducts"
10 : "HasCrCard"
11 : "IsActiveMember"
12 : "EstimatedSalary"
13 : "Exited"
]
Data Types:
35°C
Mostly clear
Search
ENG
IN
17:17
01-05-2024

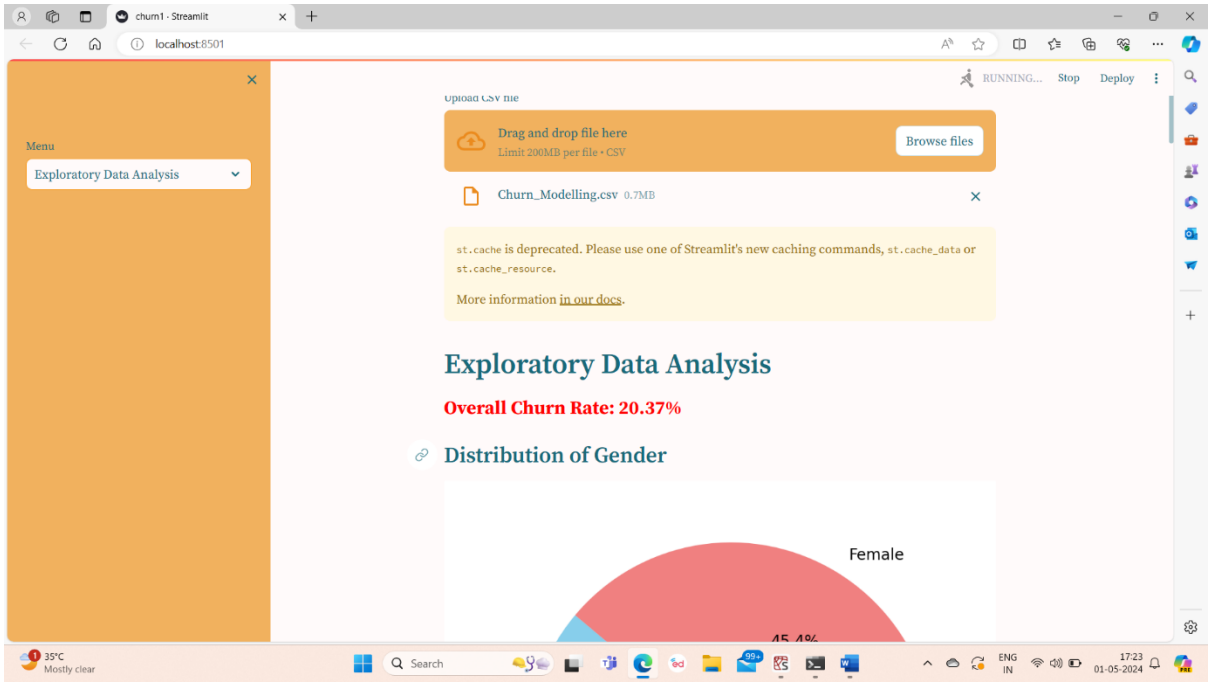

churn1 - Streamlit
localhost:8501
RUNNING...
Stop
Deploy
Menu
Data Acquisition and Pre-proc...
CreditScore float64
Geography object
Gender object
Age float64
Tenure float64
Balance float64
NumOfProduc float64
Descriptive Statistics:
RowNumber CustomerId CreditScore Age Tenure Balance NumOfProducts
count 10,000 10,000 10,000 10,000 10,000 10,000 10,000
mean 5,000.5 15,690,940.5694 650.5288 38.9218 5.0128 76,485.8893 1.5302
std 2,886.8957 71,936.1861 96.6533 10.4878 2.8922 62,397.4052 0.5817
min 1 15,565,701 350 18 0 0 1
25% 2,500.75 15,628,528.25 584 32 3 0 1
50% 5,000.5 15,690,738 652 37 5 97,198.54 1
75% 7,500.25 15,753,233.75 718 44 7 127,644.24 2
max 10,000 15,815,690 850 92 10 250,898.09 4
35°C
Mostly clear
Search
ENG
IN
17:17
01-05-2024

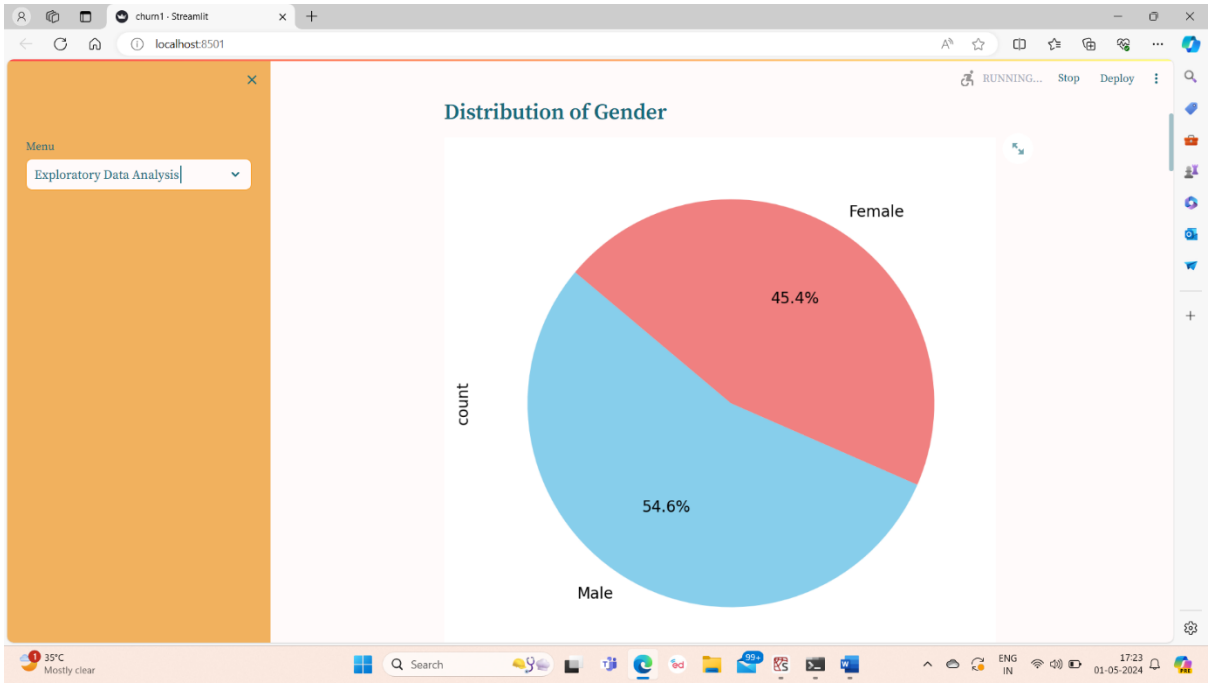

churn1 - Streamlit
localhost:8501
RUNNING...
Stop
Deploy
Menu
Data Acquisition and Pre-proc...
Data Acquisition and Pre-processi...
Exploratory Data Analysis
Model Building and Evaluation
Customer Prediction
CreditScore float64
Geography object
Gender object
Age float64
Tenure float64
Balance float64
NumOfProduc float64
Descriptive Statistics:
RowNumber CustomerId CreditScore Age Tenure Balance NumOfProducts
count 10,000 10,000 10,000 10,000 10,000 10,000 10,000
mean 5,000.5 15,690,940.5694 650.5288 38.9218 5.0128 76,485.8893 1.5302
std 2,886.8957 71,936.1861 96.6533 10.4878 2.8922 62,397.4052 0.5817
min 1 15,565,701 350 18 0 0 1
25% 2,500.75 15,628,528.25 584 32 3 0 1
50% 5,000.5 15,690,738 652 37 5 97,198.54 1
75% 7,500.25 15,753,233.75 718 44 7 127,644.24 2
max 10,000 15,815,690 850 92 10 250,898.09 4
35°C
Mostly clear
Search
ENG
IN
17:18
01-05-2024

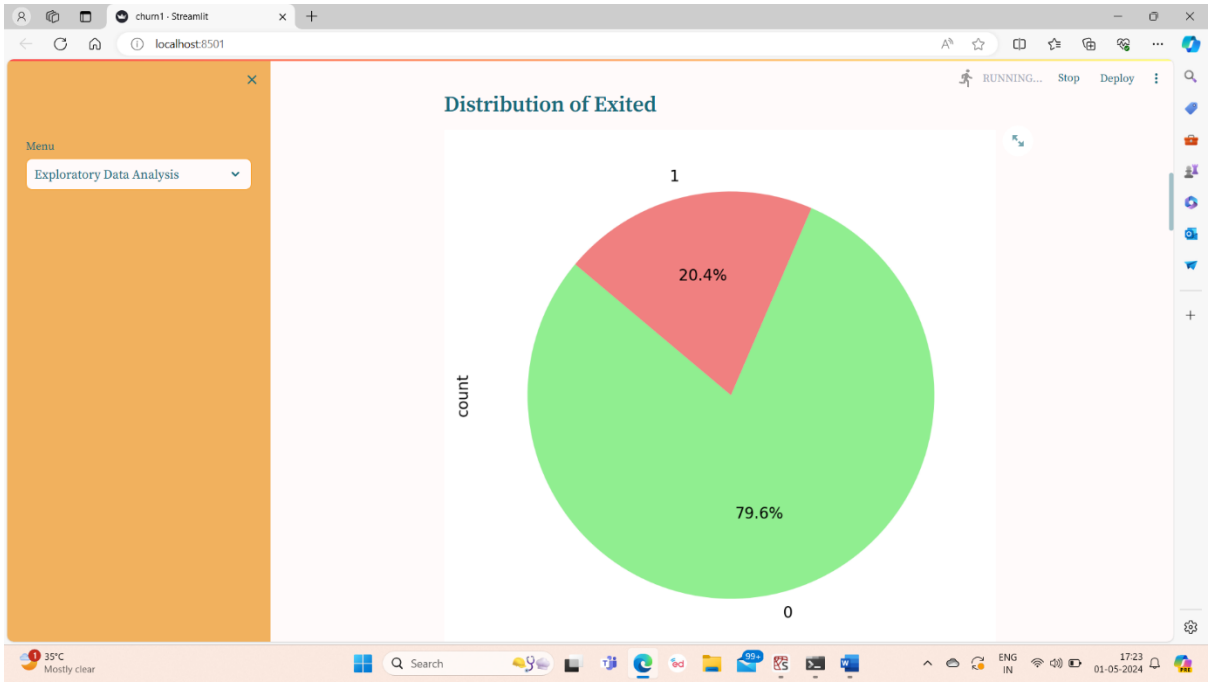

churn1 - Streamlit
localhost:8501
Deploy
Menu
Data Acquisition and Pre-proc...
50% 5,000.5 15,690,738 652 37 5 97,198.54 1
75% 7,500.25 15,753,233.75 718 44 7 127,644.24 2
max 10,000 15,815,690 850 92 10 250,898.09 4
Correlation Matrix:
RowNumber CustomerId CreditScore Age Tenure Balance NumOfProducts
RowNumber 1 0.0042 0.0058 0.0008 -0.0065 -0.0091 0.0072
CustomerId 0.0042 1 0.0053 0.0095 -0.0149 -0.0124 0.017
CreditScore 0.0058 0.0053 1 -0.004 0.0008 0.0063 0.0122
Age 0.0008 0.0095 -0.004 1 -0.01 0.0283 -0.0307
Tenure -0.0065 -0.0149 0.0008 -0.01 1 -0.0123 0.0134
Balance -0.0091 -0.0124 0.0063 0.0283 -0.0123 1 -0.3042
NumOfProducts 0.0072 0.017 0.0122 -0.0307 0.0134 -0.3042 1
HasCrCard 0.0006 -0.014 -0.0055 -0.0117 0.0226 -0.0149 0.0032
IsActiveMember 0.012 0.0017 0.0257 0.0855 -0.0284 -0.0101 0.0096
EstimatedSalary -0.006 0.0153 -0.0014 -0.0072 0.0078 0.0128 0.0142
Number of Null Values in Each Column:
0
35°C
Mostly clear
Search
ENG
IN
17:22
01-05-2024

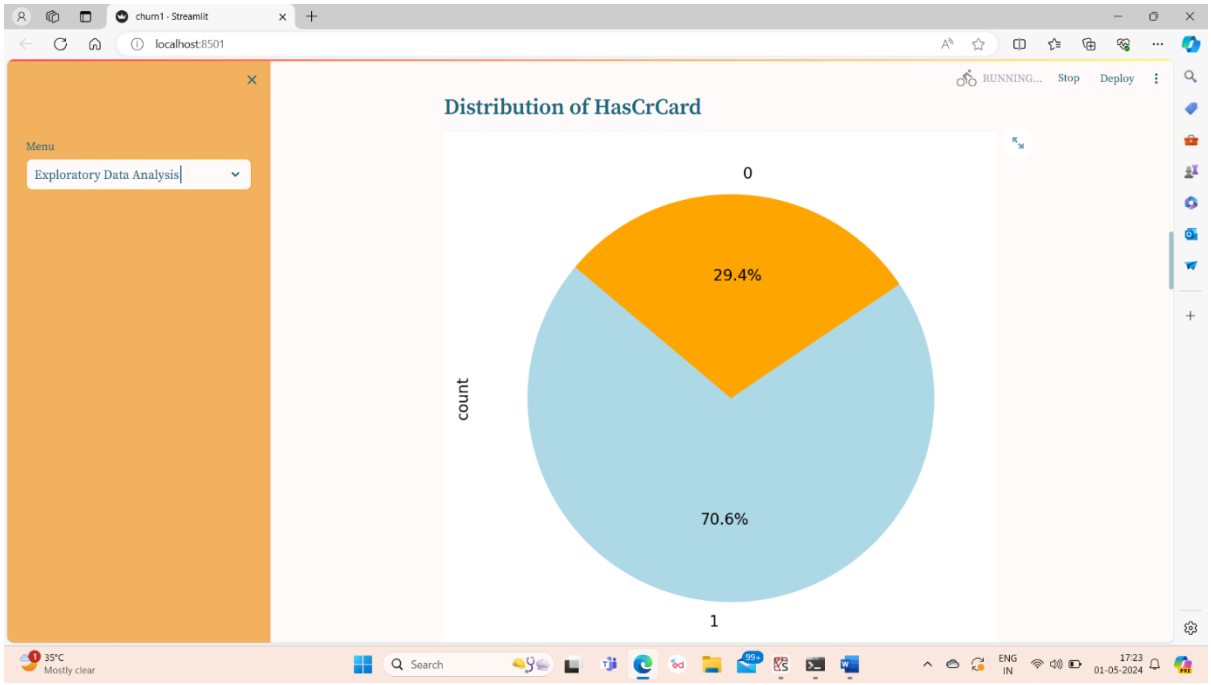

churn1 - Streamlit
localhost:8501
Deploy
Menu
Data Acquisition and Pre-proc...
Surname_Ahern 0.001 0.0058 0.0098 -0.0104 0.0034 -0.0123 0.0081
Number of Null Values in Each Column:
0
RowNumber 0
CustomerId 0
Surname 0
CreditScore 0
Geography 0
Gender 0
Age 0
Tenure 0
Balance 0
NumOfProduc 0
35°C
Mostly clear
Search
ENG
IN
17:22
01-05-2024

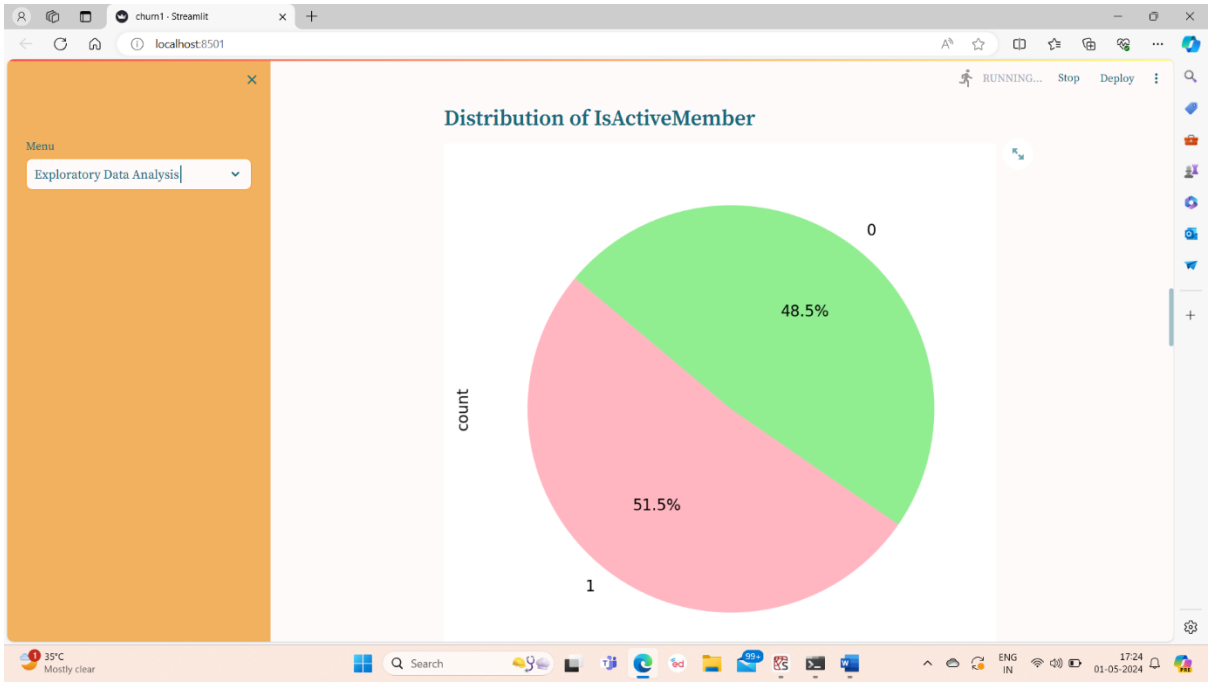

churn1 - Streamlit
localhost:8501
RUNNING...
Stop
Deploy
Menu
Exploratory Data Analysis
Customer Churn Analysis
Exploratory Data Analysis
Upload CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
Churn_Modelling.csv 0.7MB
st.cache is deprecated. Please use one of Streamlit's new caching commands, st.cache_data or st.cache_resource.
More information in our docs.
Exploratory Data Analysis
Overall Churn Rate: 20.37%
35°C
Mostly clear
Search
ENG
IN
17:23
01-05-2024

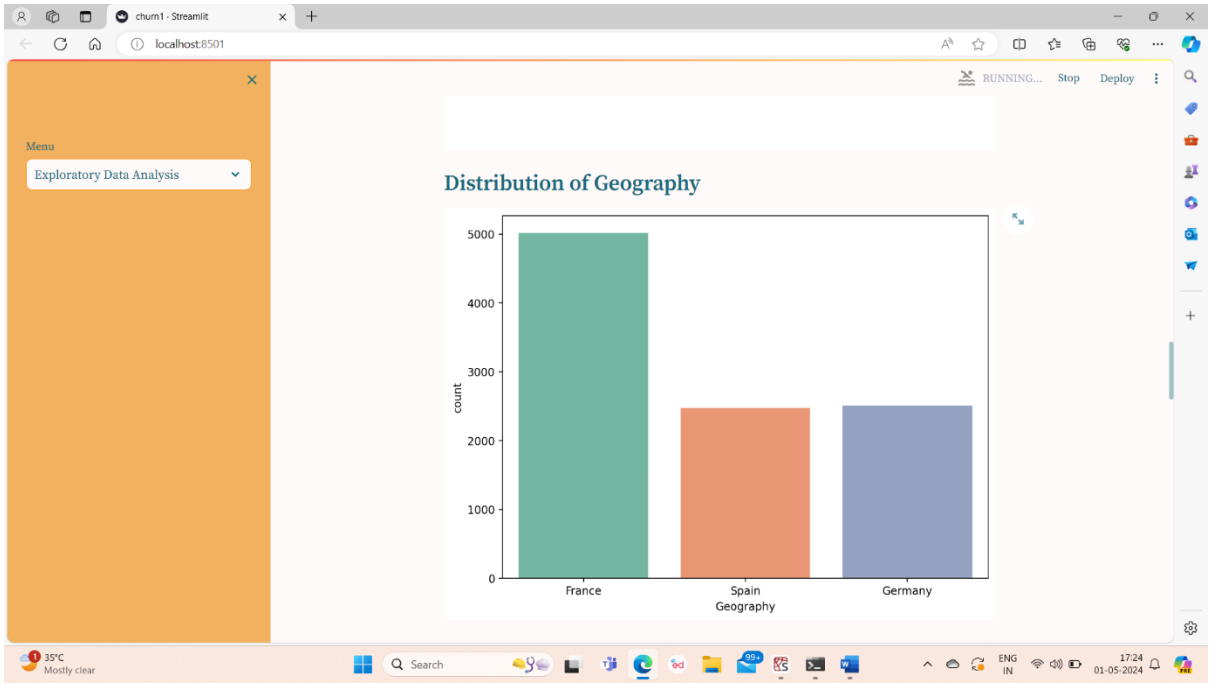

churn1 - Streamlit
localhost:8501
RUNNING...
Stop
Deploy
Menu
Exploratory Data Analysis
Drag and drop file here
Limit 200MB per file • CSV
Browse files
Churn_Modelling.csv 0.7MB
st.cache is deprecated. Please use one of Streamlit's new caching commands, st.cache_data or st.cache_resource.
More information in our docs.
Exploratory Data Analysis
Overall Churn Rate: 20.37%
Distribution of Gender
Female
35°C
Mostly clear
Search
ENG
IN
17:23
01-05-2024

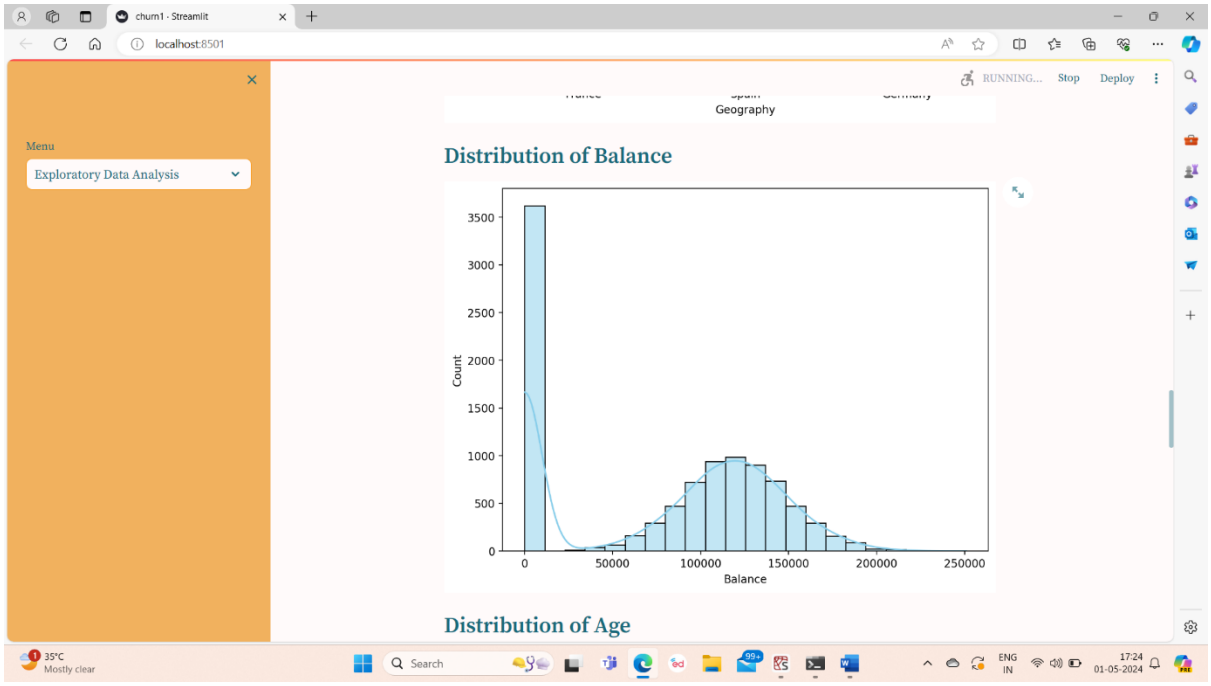

churn1 - Streamlit
localhost:8501
RUNNING...
Stop
Deploy
Menu
Exploratory Data Analysis
Distribution of Gender
Female
45.4%
count
54.6%
Male
35°C
Mostly clear
Search
ENG
IN
17:23
01-05-2024

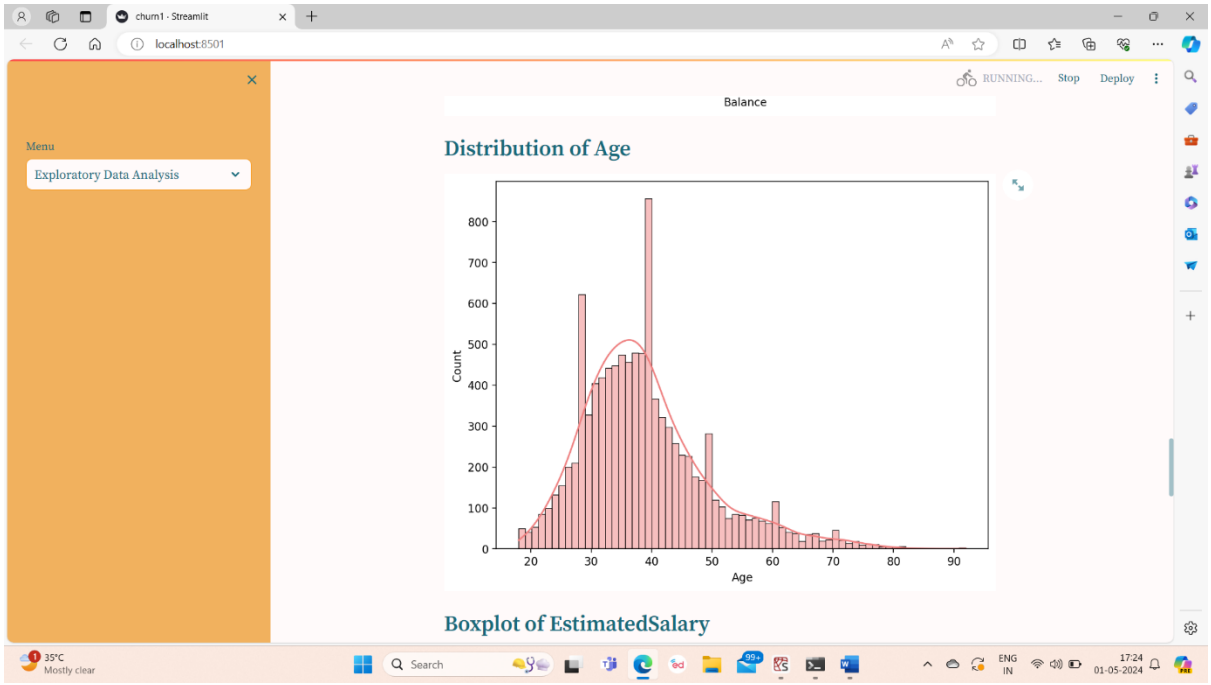

churn1 - Streamlit
localhost:8501
RUNNING...
Stop
Deploy
Menu
Exploratory Data Analysis
Distribution of Exited
1
20.4%
count
79.6%
0
35°C
Mostly clear
Search
ENG
IN
17:23
01-05-2024

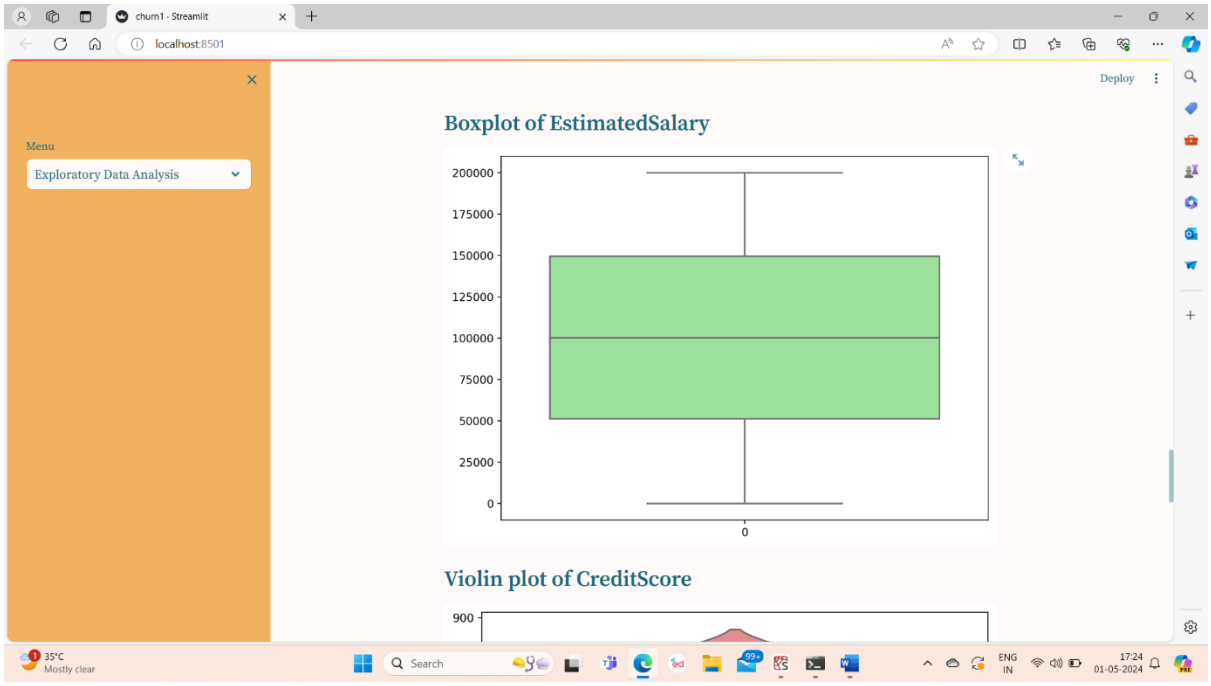

churn1 - Streamlit
localhost:8501
RUNNING...
Stop
Deploy
Menu
Exploratory Data Analysis
Distribution of HasCrCard
0
29.4%
count
70.6%
1
35°C
Mostly clear
Search
ENG
IN
17:23
01-05-2024

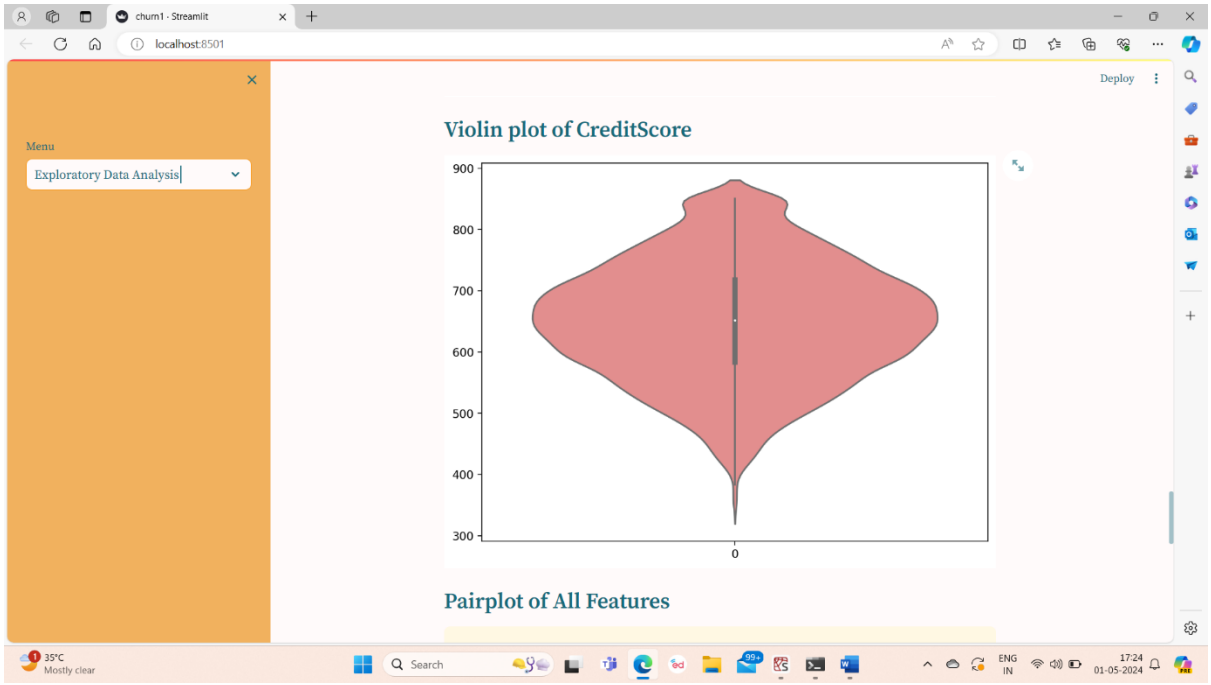

churn1 - Streamlit
localhost:8501
RUNNING...
Stop
Deploy
Menu
Exploratory Data Analysis
Distribution of IsActiveMember
0
48.5%
count
51.5%
1
35°C
Mostly clear
Search
ENG
IN
17:24
01-05-2024

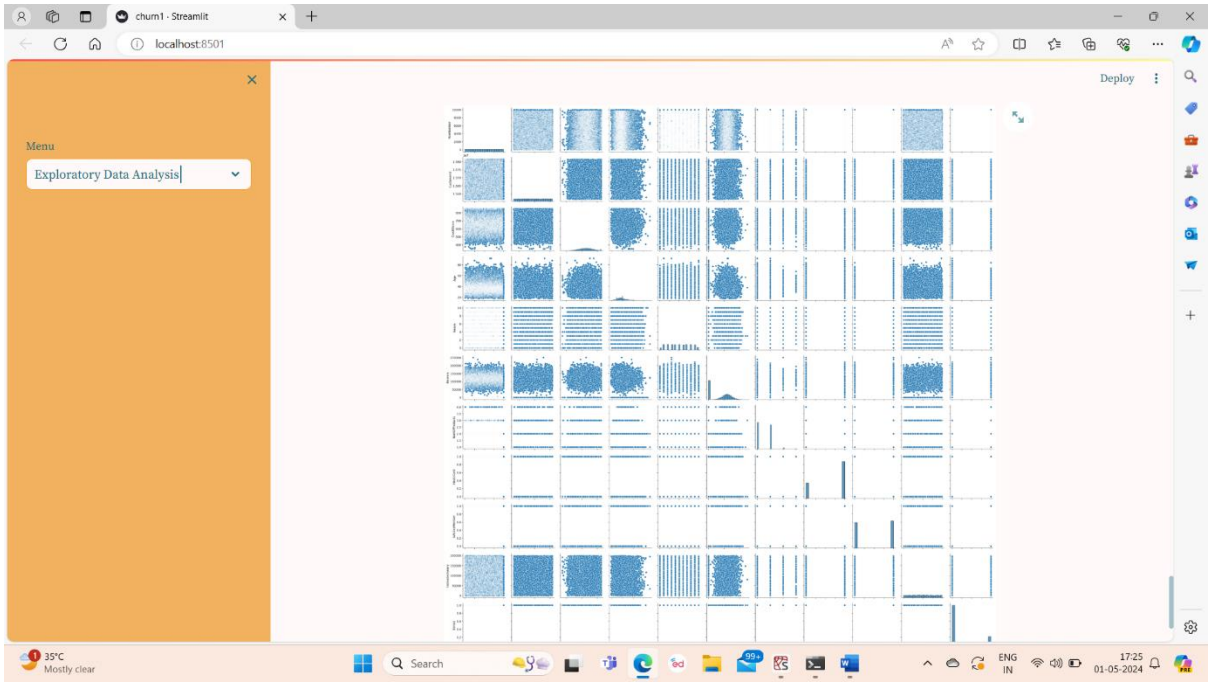

churn1 - Streamlit
localhost:8501
RUNNING...
Stop
Deploy
Menu
Exploratory Data Analysis
Distribution of Geography
5000
4000
3000
2000
1000
0
count
France
Spain
Germany
Geography
35°C
Mostly clear
Search
ENG
IN
17:24
01-05-2024

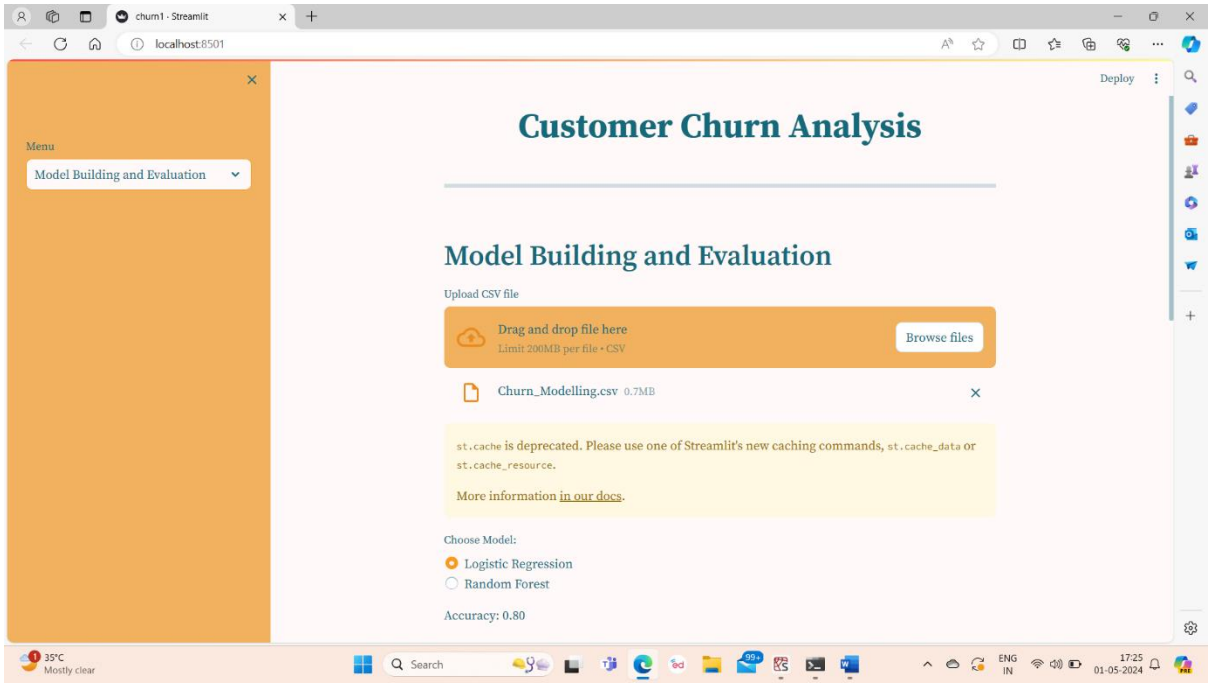

churn1 - Streamlit
localhost:8501
RUNNING...
Stop
Deploy
Menu
Exploratory Data Analysis
Geography
Distribution of Balance
Count
Balance
0
50000
100000
150000
200000
250000
500
1000
1500
2000
2500
3000
3500
Distribution of Age
35°C
Mostly clear
Search
ENG
IN
17:24
01-05-2024

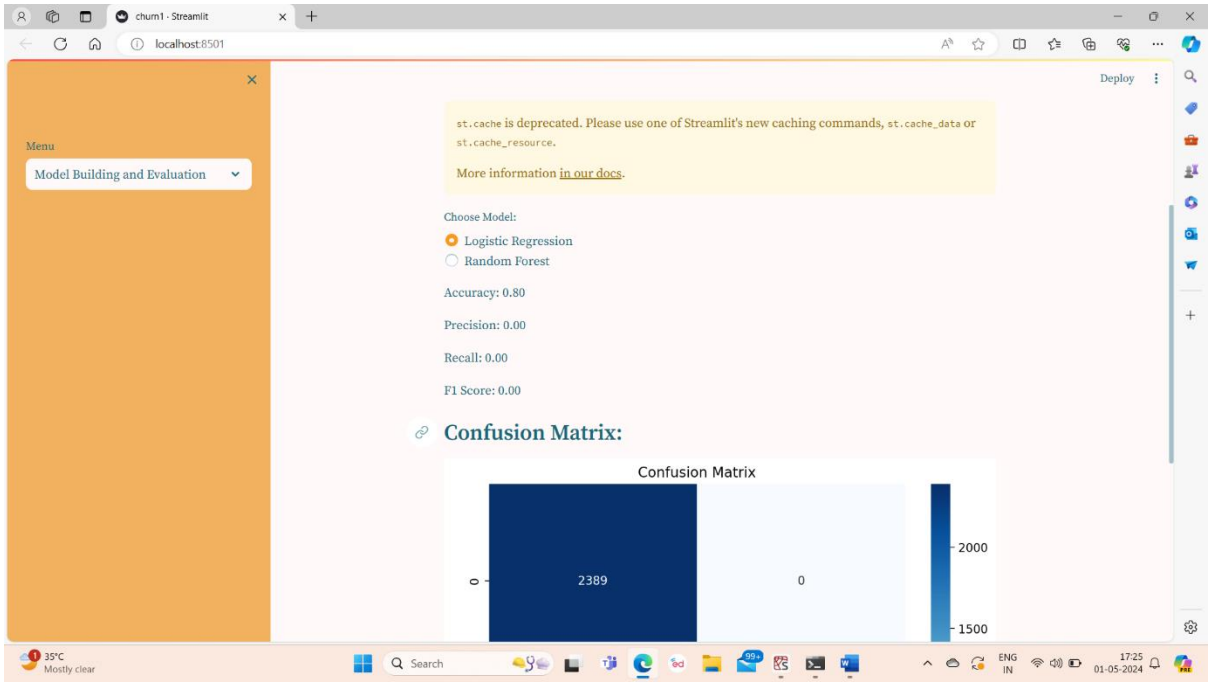

churn1 - Streamlit
localhost:8501
RUNNING...
Stop
Deploy
Menu
Exploratory Data Analysis
Balance
Distribution of Age
Count
Age
0
20
30
40
50
60
70
80
90
100
200
300
400
500
600
700
800
Boxplot of EstimatedSalary
35°C
Mostly clear
Search
ENG
IN
17:24
01-05-2024

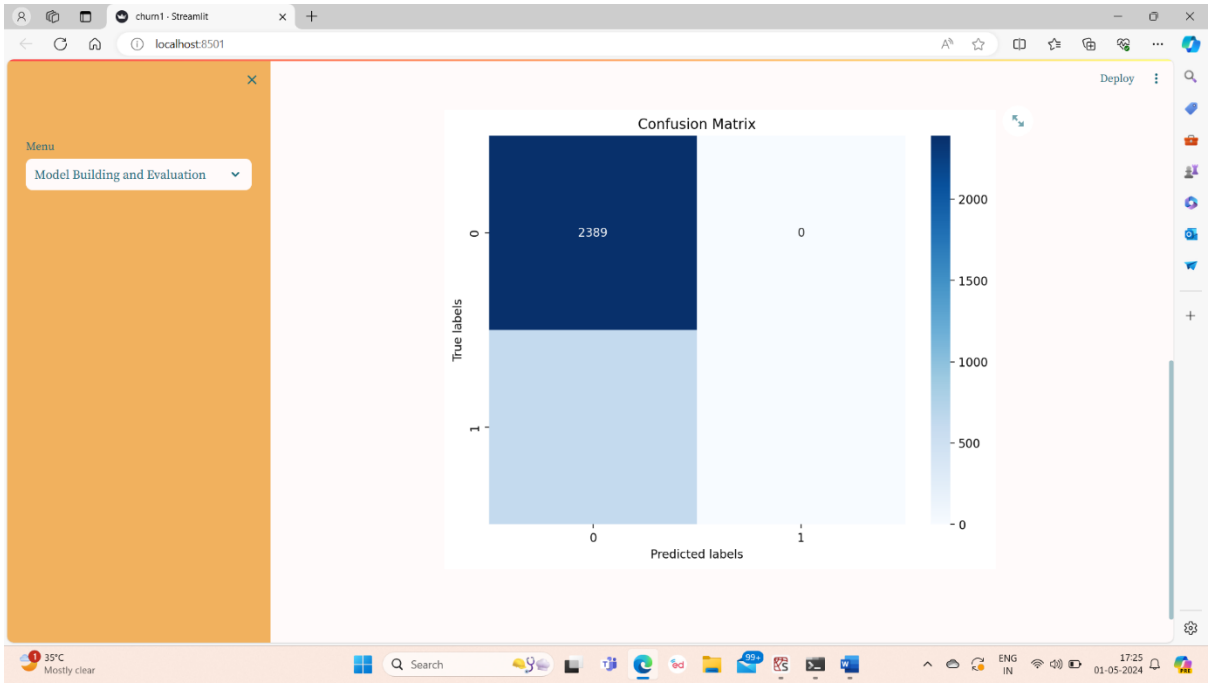

churn1 - Streamlit
localhost:8501
Deploy
Menu
Exploratory Data Analysis
Boxplot of EstimatedSalary
200000
175000
150000
125000
100000
75000
50000
25000
0
0
Violin plot of CreditScore
900
35°C
Mostly clear
Search
ENG
IN
17:24
01-05-2024

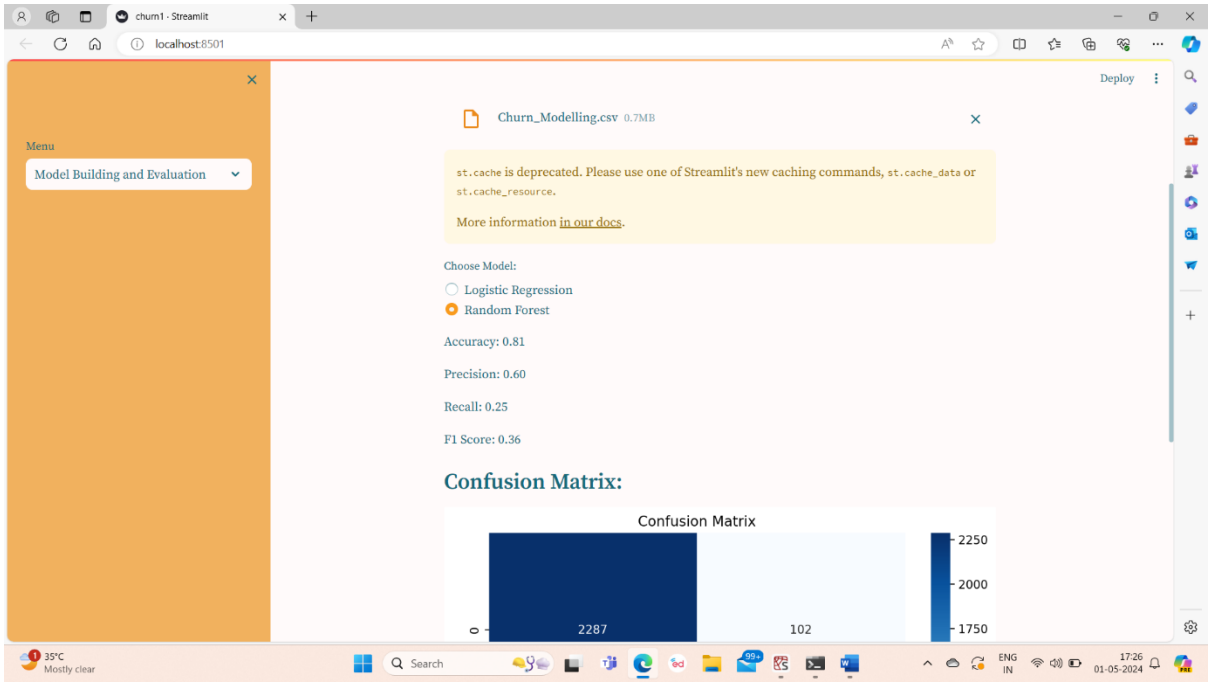

churn1 - Streamlit
localhost:8501
Deploy
Menu
Exploratory Data Analysis
Violin plot of CreditScore
900
800
700
600
500
400
300
0
Pairplot of All Features
35°C
Mostly clear
Search
ENG
IN
17:24
01-05-2024

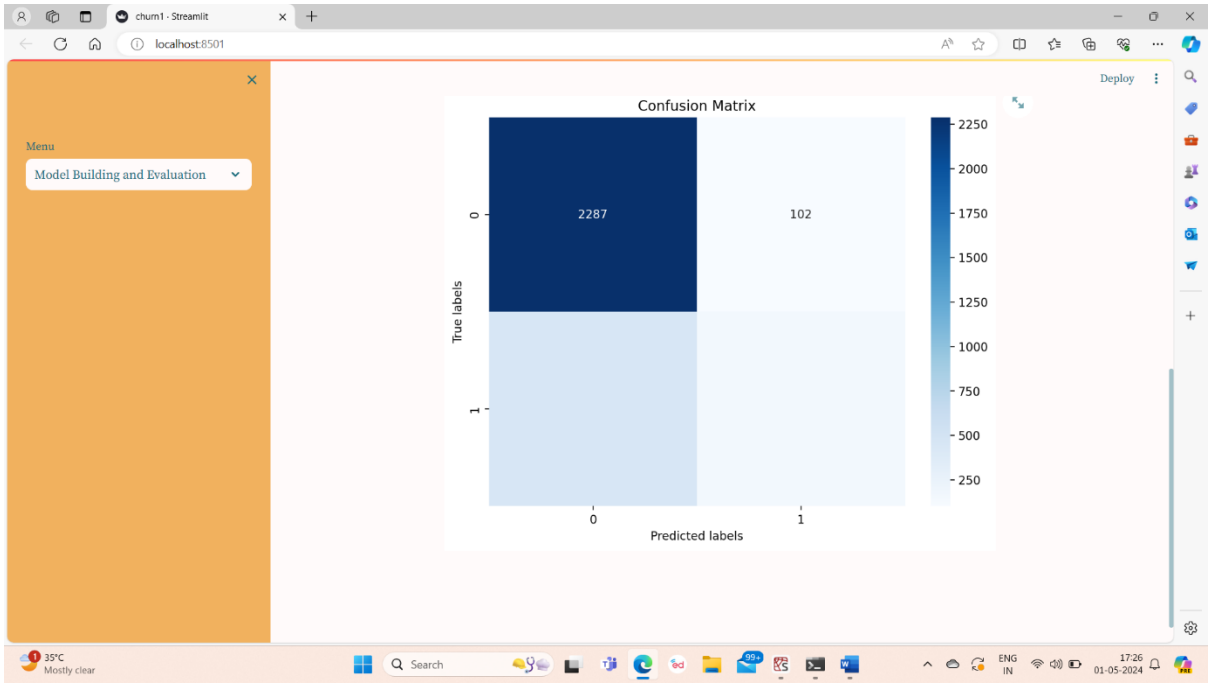

churn1 - Streamlit
localhost:8501
Deploy
Menu
Exploratory Data Analysis
35°C
Mostly clear
Search
ENG
IN
17:25
01-05-2024

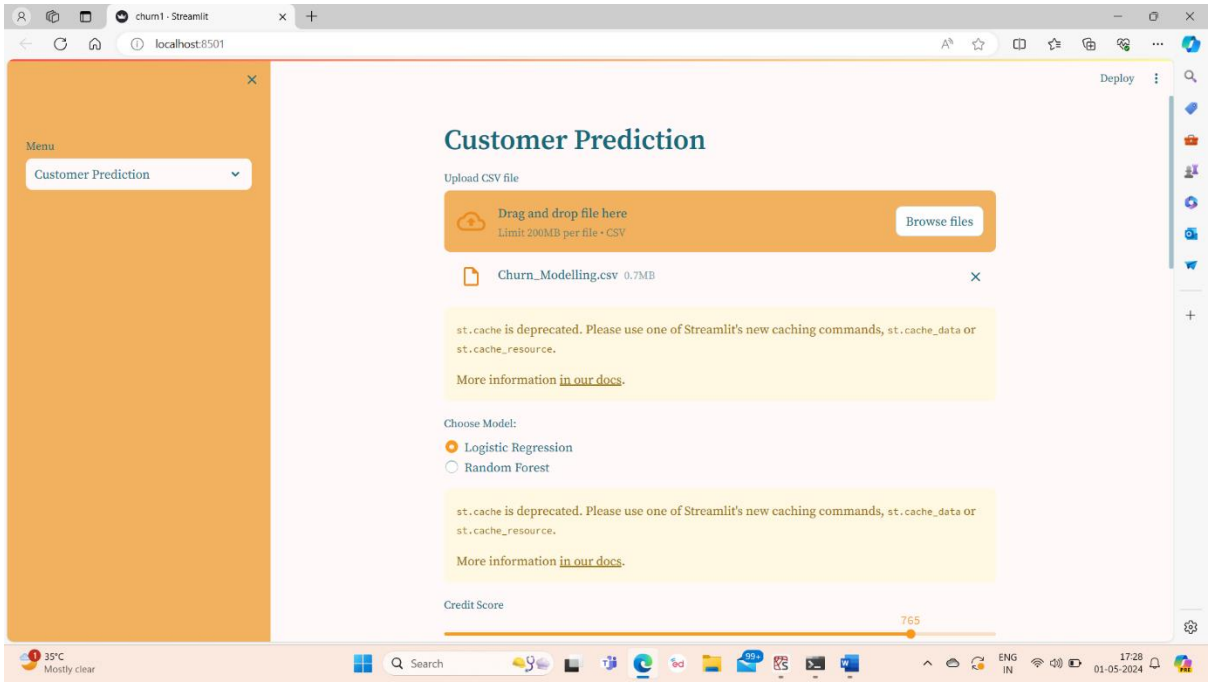

churn1 - Streamlit
localhost:8501
Deploy
Menu
Model Building and Evaluation
Customer Churn Analysis
Model Building and Evaluation
Upload CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
Churn_Modelling.csv 0.7MB
st.cache is deprecated. Please use one of Streamlit's new caching commands, st.cache_data or st.cache_resource.
More information in our docs.
Choose Model:
Logistic Regression
Random Forest
Accuracy: 0.80
35°C
Mostly clear
Search
ENG
IN
17:25
01-05-2024

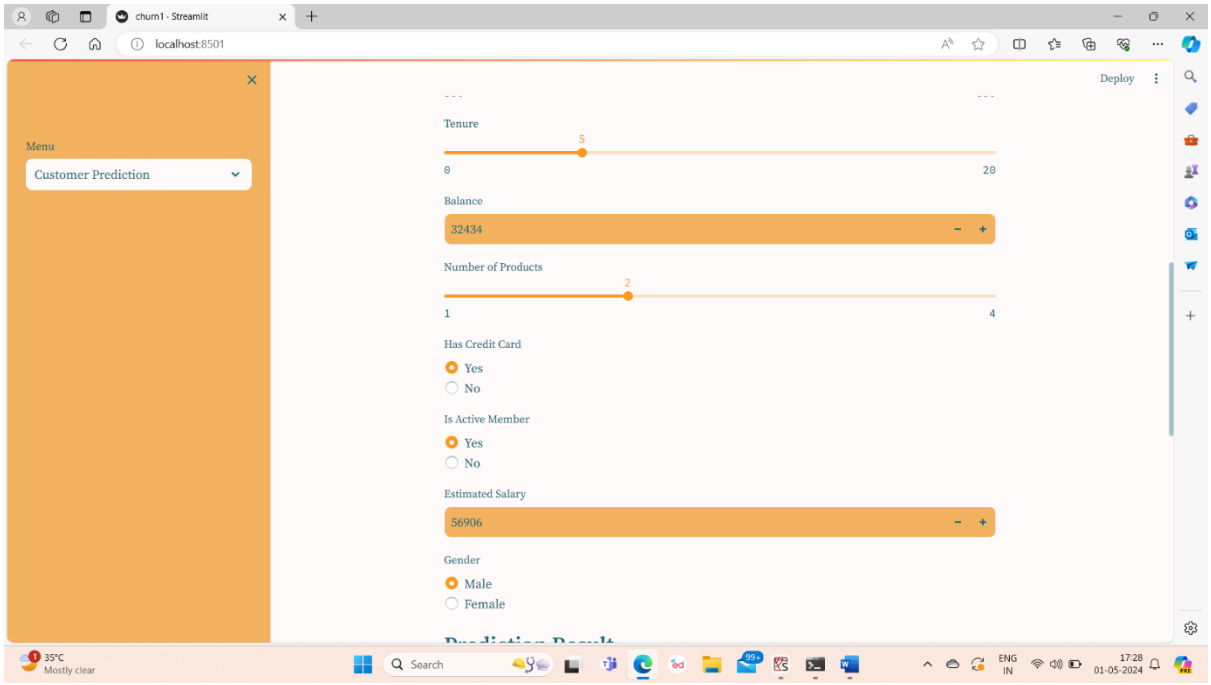

churn1 - Streamlit
localhost:8501
Deploy
Menu
Model Building and Evaluation
st.cache is deprecated. Please use one of Streamlit's new caching commands, st.cache_data or st.cache_resource.
More information in our docs.
Choose Model:
Logistic Regression
Random Forest
Accuracy: 0.80
Precision: 0.00
Recall: 0.00
F1 Score: 0.00
Confusion Matrix:
Confusion Matrix
2389
0
2000
1500
35°C
Mostly clear
Search
ENG
IN
17:25
01-05-2024

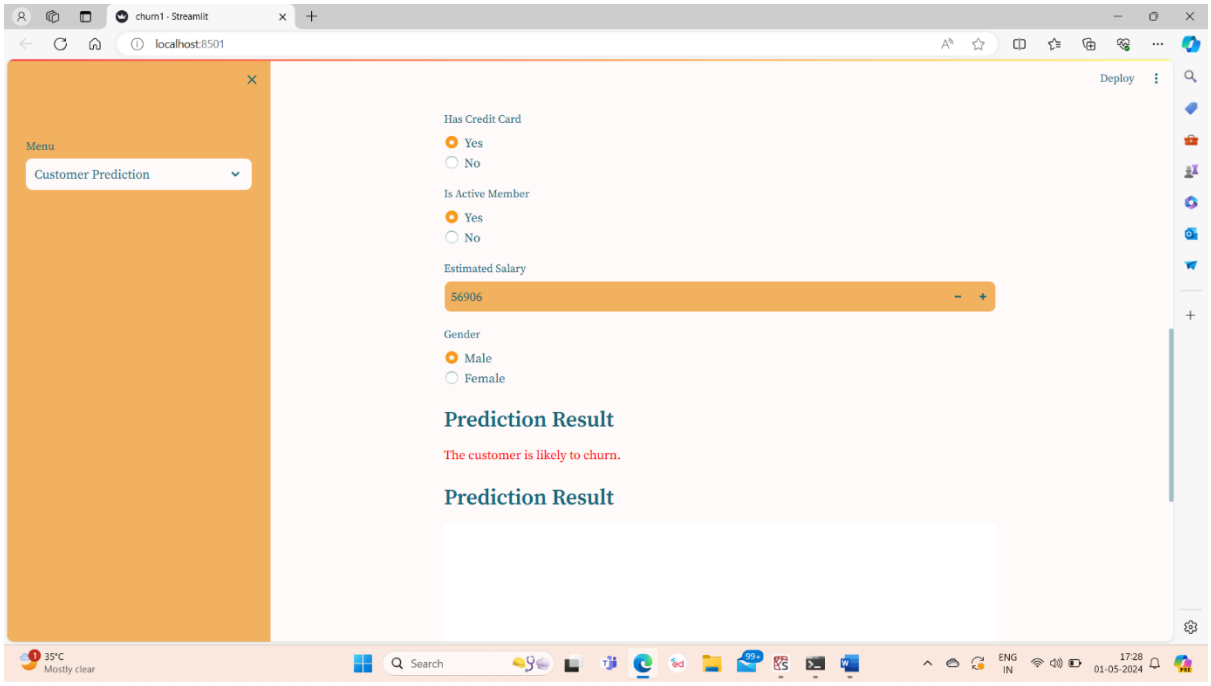

churn1 - Streamlit
localhost:8501
Deploy
Menu
Model Building and Evaluation
Confusion Matrix
2389
0
True labels
Predicted labels
2000
1500
1000
500
0
35°C
Mostly clear
Search
ENG
IN
17:25
01-05-2024

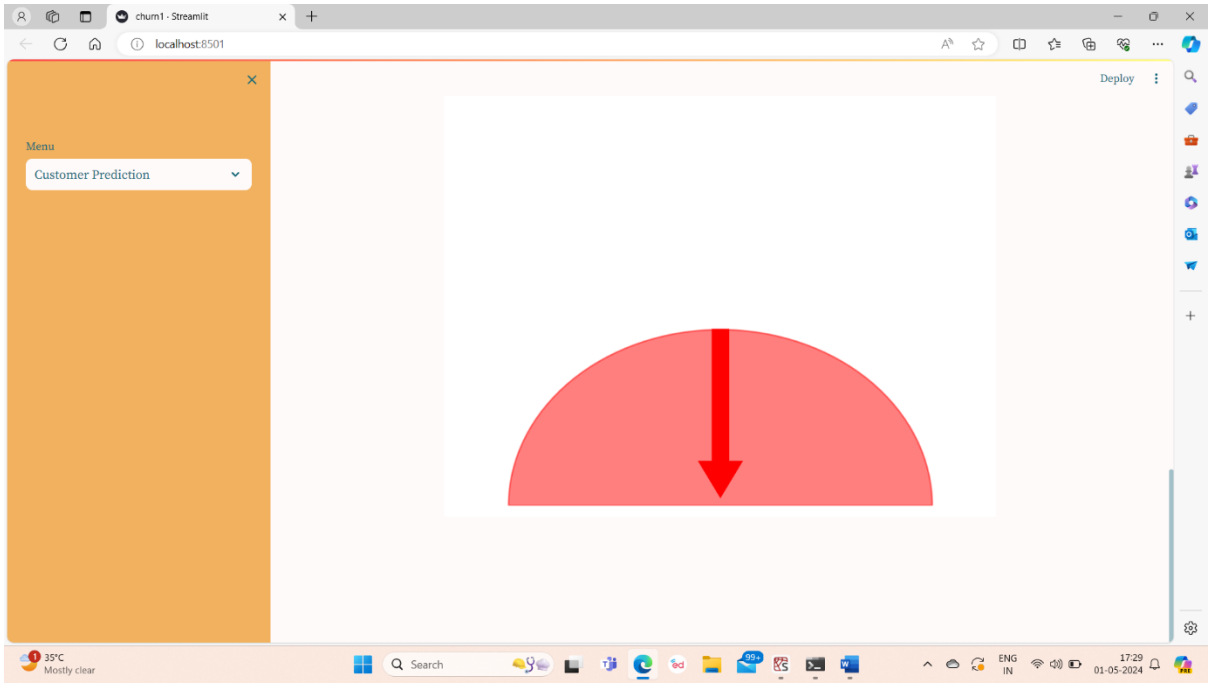

churn1 - Streamlit
localhost:8501
Deploy
Menu
Model Building and Evaluation
Churn_Modelling.csv 0.7MB
st.cache is deprecated. Please use one of Streamlit's new caching commands, st.cache_data or st.cache_resource.
More information in our docs.
Choose Model:
Logistic Regression
Random Forest
Accuracy: 0.81
Precision: 0.60
Recall: 0.25
F1 Score: 0.36
Confusion Matrix:
Confusion Matrix
2287
102
0
2250
2000
1750
35°C
Mostly clear
Search
ENG
IN
17:26
01-05-2024

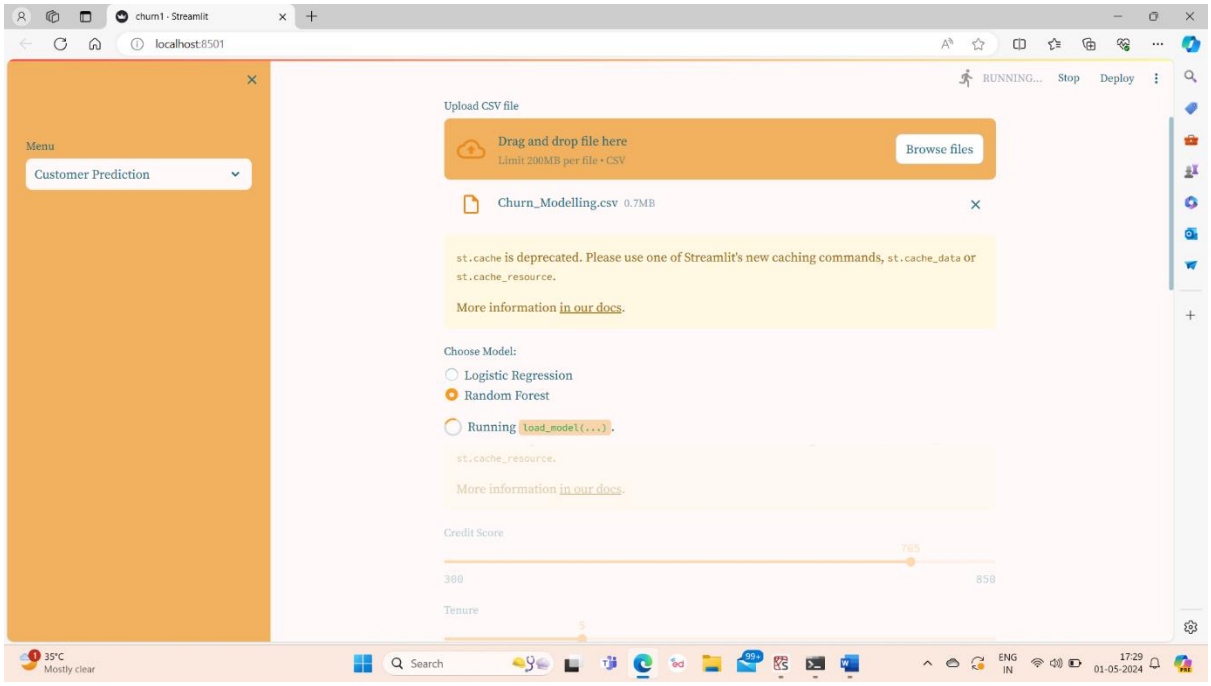

churn1 - Streamlit
localhost:8501
Deploy
Menu
Model Building and Evaluation
Confusion Matrix
2287
102
True labels
Predicted labels
0
1
2250
2000
1750
1500
1250
1000
750
500
250
35°C
Mostly clear
Search
ENG
IN
17:26
01-05-2024

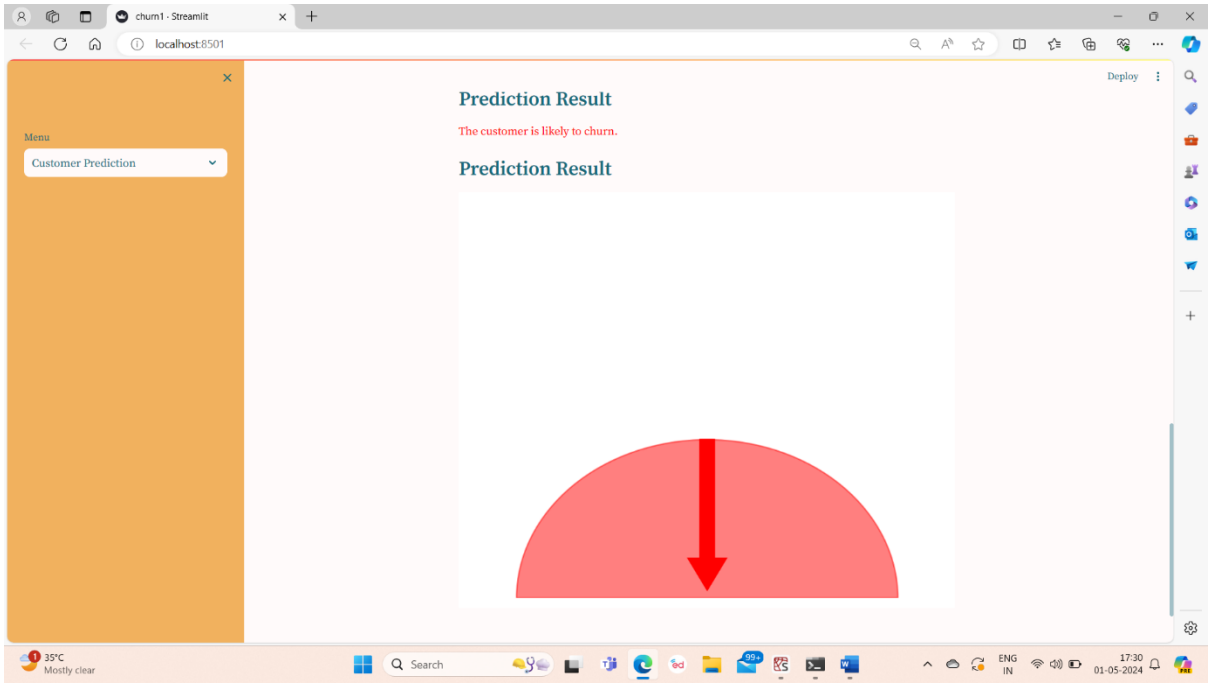

churn1 - Streamlit
localhost:8501
Deploy
Menu
Customer Prediction
Customer Prediction
Upload CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
Churn_Modelling.csv 0.7MB
st.cache is deprecated. Please use one of Streamlit's new caching commands, st.cache_data or st.cache_resource.
More information in our docs.
Choose Model:
Logistic Regression
Random Forest
st.cache is deprecated. Please use one of Streamlit's new caching commands, st.cache_data or st.cache_resource.
More information in our docs.
Credit Score
765
35°C
Mostly clear
Search
ENG
IN
17:28
01-05-2024

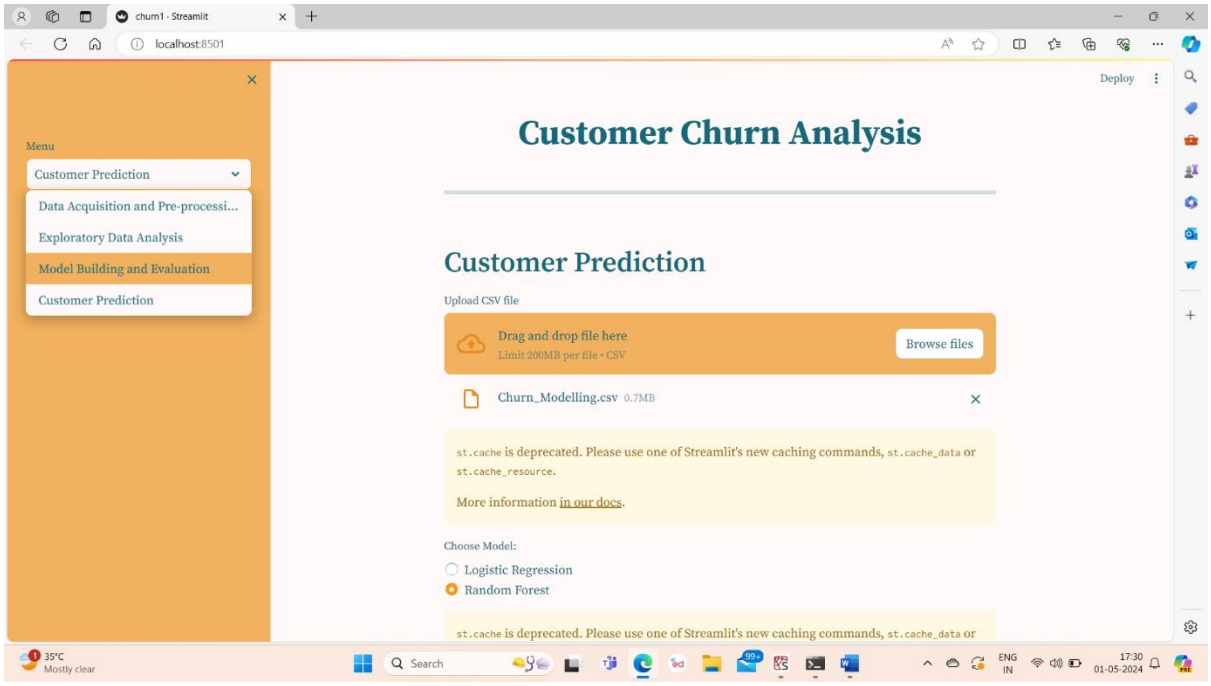

churn1 - Streamlit
localhost:8501
Deploy
Menu
Customer Prediction
Tenure
5
0
20
Balance
32434
Number of Products
2
1
4
Has Credit Card
Yes
No
Is Active Member
Yes
No
Estimated Salary
56906
Gender
Male
Female
35°C
Mostly clear
Search
ENG
IN
17:28
01-05-2024


churn1 - Streamlit
localhost:8501
Deploy
Menu
Customer Prediction
Has Credit Card
Yes
No
Is Active Member
Yes
No
Estimated Salary
56906
Gender
Male
Female
Prediction Result
The customer is likely to churn.
Prediction Result
35°C
Mostly clear
Search
ENG
IN
17:28
01-05-2024


churn1 - Streamlit
localhost:8501
Deploy
Menu
Customer Prediction
35°C
Mostly clear
Search
ENG
IN
17:29
01-05-2024


churn1 - Streamlit
localhost:8501
RUNNING...
Stop
Deploy
Menu
Customer Prediction
Upload CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
Churn_Modelling.csv 0.7MB
st.cache is deprecated. Please use one of Streamlit's new caching commands, st.cache_data or st.cache_resource.
More information in our docs.
Choose Model:
Logistic Regression
Random Forest
Running load_model(...).
st.cache_resource.
More information in our docs.
Credit Score
765
300
850
Tenure
5
35°C
Mostly clear
Search
ENG
IN
17:29
01-05-2024


churn1 - Streamlit
localhost:8501
Deploy
Menu
Customer Prediction
Prediction Result
The customer is likely to churn.
Prediction Result
35°C
Mostly clear
Search
ENG
IN
17:30
01-05-2024


churn1 - Streamlit
localhost:8501
Deploy
Menu
Customer Prediction
Data Acquisition and Pre-processi...
Exploratory Data Analysis
Model Building and Evaluation
Customer Prediction
Customer Churn Analysis
Customer Prediction
Upload CSV file
Drag and drop file here
Limit 200MB per file • CSV
Browse files
Churn_Modelling.csv 0.7MB
st.cache is deprecated. Please use one of Streamlit's new caching commands, st.cache_data or st.cache_resource.
More information in our docs.
Choose Model:
Logistic Regression
Random Forest
st.cache is deprecated. Please use one of Streamlit's new caching commands, st.cache_data or
35°C
Mostly clear
Search
ENG
IN
17:30
01-05-2024

CHAPTER – 05

CONCLSION

In this comprehensive project on customer churn analysis within a financial institution, we embarked on a journey to understand and predict customer behaviour using advanced data analytics techniques. Leveraging a dataset encompassing various customer attributes such as credit score, geography, gender, age, tenure, balance, number of products, credit card status, activity status, estimated salary, and churn status, we aimed to develop models capable of identifying potential churners. Our meticulous approach involved preprocessing the data, which included handling missing values, encoding categorical variables, and scaling features to ensure optimal model performance.

The exploration phase delved deep into the dataset's characteristics, providing insights into the distribution of key features and their relationships. Additionally, we unveiled **the overall churn rate of 20.37%**, shedding light on the prevalence of churn within the institution's customer base. Armed with this knowledge, we proceeded to build and evaluate two predictive models: Logistic Regression and Random Forest Classifier. While **Logistic Regression** yielded an **accuracy of 80% and a precision of 50%,** the **Random Forest Classifier** outperformed it with an **accuracy of 81% and a precision of 59%.** Despite these variations in performance, both models played a crucial role in identifying potential churners, albeit with different levels of precision and recall.

Moreover, our project enabled real-time predictions based on user-input customer attributes, offering actionable insights for proactive churn management strategies. By harnessing the power of data analytics, we empowered financial institutions to anticipate and mitigate customer churn effectively, thereby fostering customer loyalty and sustaining business growth.

In conclusion, our project serves as a testament to the transformative impact of data-driven decision-making in tackling contemporary business challenges.