Video Presentation

Telecommunications Company Customer Churn Analysis

CSCA 5622 Supervised Learning Final Project
David Emilio Vega Bonza, david.vegabonza@colorado.edu
ID: 80215162, Bogotá. Colombia

March 2025

AGENDA ITEMS

1. Project Topic
2. Data
3. Data Cleaning
4. Exploratory Data Analysis
5. Models
6. Results and Analysis
7. Demo
8. Conclusions

1. Project Topic

1.1 Telecommunications Company Customer Churn Analysis

1.2 Customer churn analysis in **telecom companies focuses on identifying customers who are likely to leave the service provider**. This project involves analyzing customer data to predict churn and understand the factors contributing to it.

In this project we will review how **"Customer churn analysis"** is crucial for telecom companies for several reasons: Revenue Loss Prevention, Customer Retention, Competitive Edge, Customer Satisfaction, Marketing Optimization , Product Development.

1.3 The main **goal is to develop predictive models that can accurately identify potential churners** and **provide insights to help telecom companies** retain their customers for a much longer.

2. Data

Data Source Description:

Context Data Source:

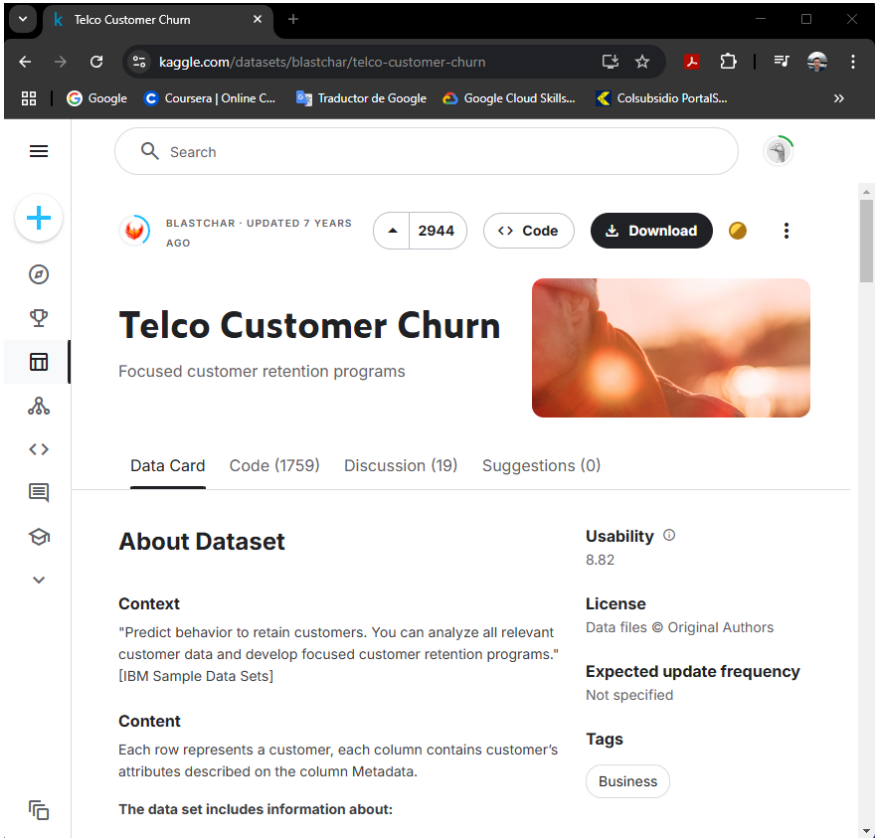
- **Dataset Name:** Telco Customer Churn (focused customer retention programs)
- **Url:** https://www.kaggle.com/datasets/blastchar/telco-customer-churn
- **Motivation:** "Predict behavior to retain customers. You can analyze all relevant customer data and develop focused customer retention programs." [IBM Sample Data Sets]

Content:

Each row represents a customer, each column contains customer's attributes described on the column Metadata.

The data set includes information about:

- Customers who left within the last month – the column is called Churn
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
- Demographic info about customers – gender, age range, and if they have partners and dependents



2. Data

Dataset Variables

#	VARIABLES
1	**customerID:** Customer ID
2	**gender:** Whether the customer is a male or a female
3	**SeniorCitizen:** Whether the customer is a senior citizen or not (1, 0)
4	**Partner:** Whether the customer has a partner or not (Yes, No)
5	**Dependents:** Whether the customer has dependents or not (Yes, No)
6	**tenure:** Number of months the customer has stayed with the company
7	**PhoneService:** Whether the customer has a phone service or not (Yes, No)
8	**MultipleLines:** Whether the customer has multiple lines or not (Yes, No, No phone service)
9	**InternetService:** Customer's internet service provider (DSL, Fiber optic, No)
10	**OnlineSecurity:** Whether the customer has online security or not (Yes, No, No internet service)

#	VARIABLES
11	**OnlineBackup:** Whether the customer has online backup or not (Yes, No, No internet service)
12	**DeviceProtection:** Whether the customer has device protection or not (Yes, No, No internet service)
13	**TechSupport:** Whether the customer has tech support or not (Yes, No, No internet service)
14	**StreamingTV:** Whether the customer has streaming TV or not (Yes, No, No internet service)
15	**StreamingMovies:** Whether the customer has streaming movies or not (Yes, No, No internet service)
16	**Contract:** The contract term of the customer (Month-to-month, One year, Two year)
17	**PaperlessBilling:** Whether the customer has paperless billing or not (Yes, No)
18	**PaymentMethod:** The customer's payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))
19	**MonthlyCharges:** The amount charged to the customer monthly
20	**TotalCharges:** The total amount charged to the customer
21	**Churn:** Whether the customer churned or not (Yes or No)

3. Data Cleaning

Data Overview:

- Reading the dataset
- Understanding the shape of the dataset
- Checking the data types
- Checking for missing values
- Checking for duplicated values
- Identification of categorical and numerical features

```
df=pd.read_csv("./data/telco-customer-churn.csv")
```

```
# Head info
df.head()
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService
0	7590-VHVEG	Female	0	Yes	No	1	No
1	5575-GNVDE	Male	0	No	No	34	Yes
2	3668-QPYBK	Male	0	No	No	2	Yes
3	7795-CFOCW	Male	0	No	No	45	No
4	9237-HQITU	Female	0	No	No	2	Yes

5 rows × 21 columns

```
# Head info
print(df.shape)
```

```
(7043, 21)
```

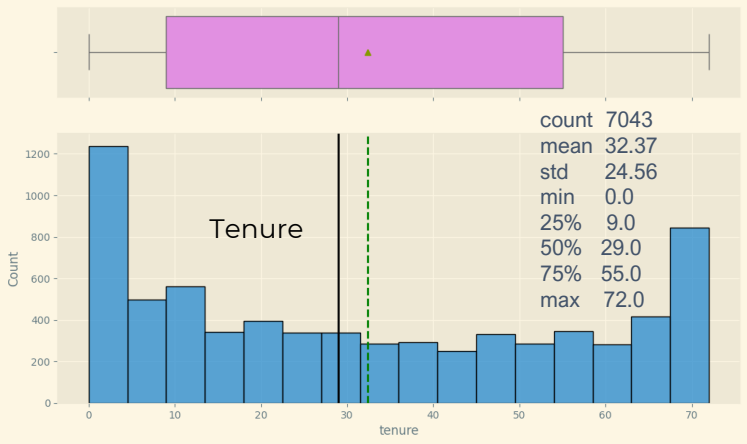
3. Data Cleaning:

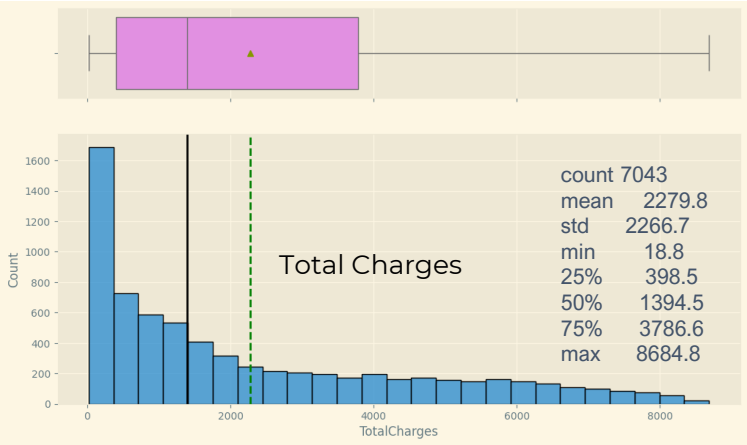
```
# 1. Drop 'customerID' as this column does not gives important information
# in the data set
df.drop('customerID', axis=1, inplace=True)

# 2. Convert Yes o No to binary (1,0)
df['Churn'] = df['Churn'].map({'Yes': 1, 'No': 0})

# 3. Data cleaning for the column 'TotalCharges'
blank_indexes = df[df['TotalCharges'] == ' '].index.values.tolist()
#blank_indexes , _ = df[df['TotalCharges'] == ' '].index
#print(blank_indexes)

# As we found blank string values for some 'TotalCharges' it will be useful
# to assign the value as 'MonthlyCharges' exist
for index in blank_indexes:
    df.loc[index, 'TotalCharges'] = df.loc[index, 'MonthlyCharges']
    #print(df.iloc[index])

#Convert TotalCharges to numerical
df["TotalCharges"] = pd.to_numeric(df["TotalCharges"])
```

Setting categorical and numerical variables

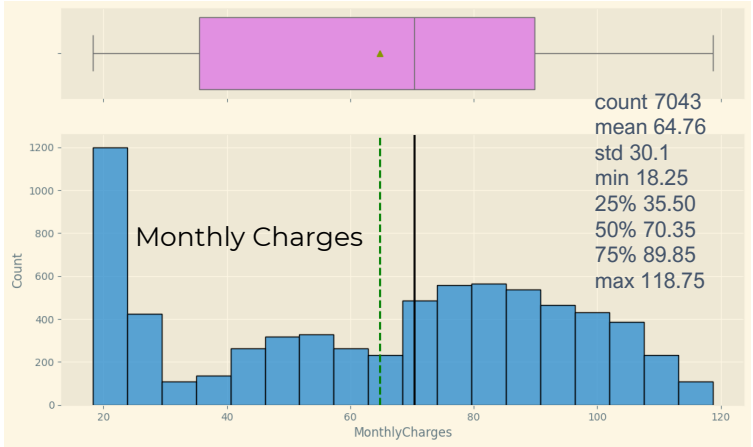
```
# Categorical variables
categorical_cols = ['gender', 'SeniorCitizen', 'Partner', 'Dependents', 'PhoneService', 'MultipleLines',
'InternetService', 'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport', 'StreamingTV',
'StreamingMovies', 'Contract', 'PaperlessBilling', 'PaymentMethod']

# Numerical variables
numerical_cols = ['tenure', 'MonthlyCharges', 'TotalCharges']
```

4. Exploratory Data Analysis



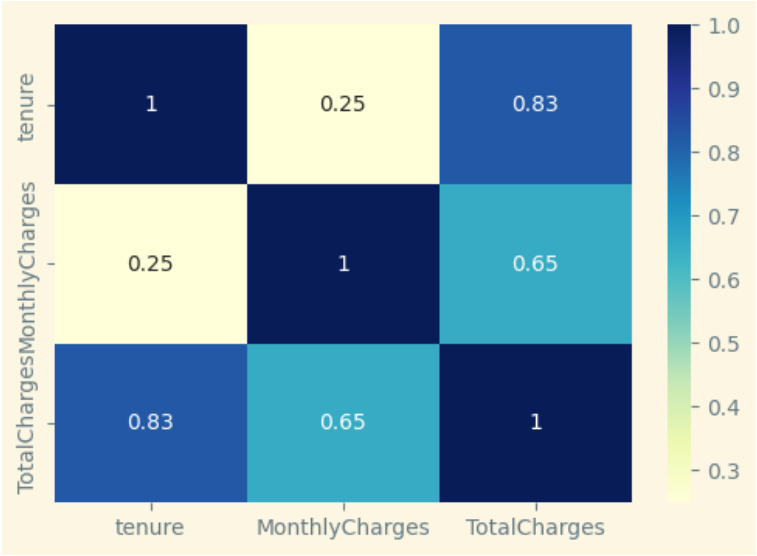


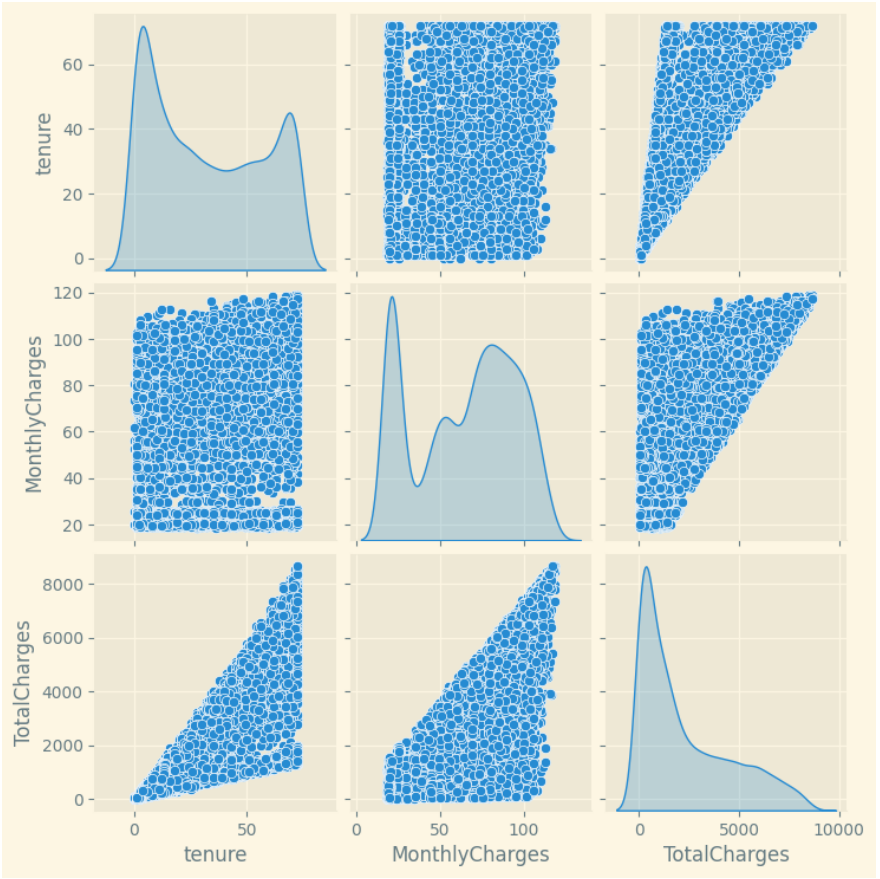
Boxplot and histogram combined analysis



4. Exploratory Data Analysis

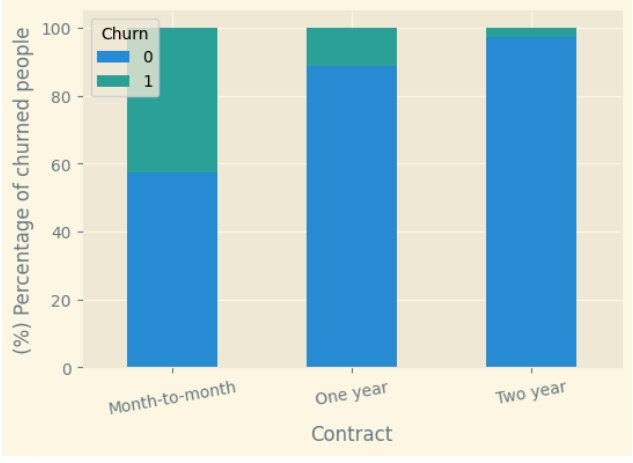
Multivariate Analysis

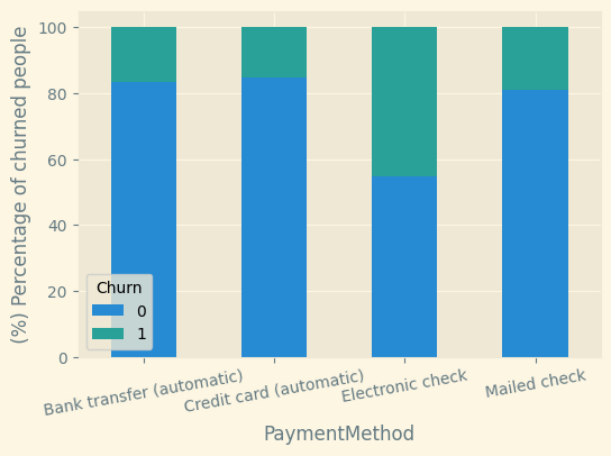


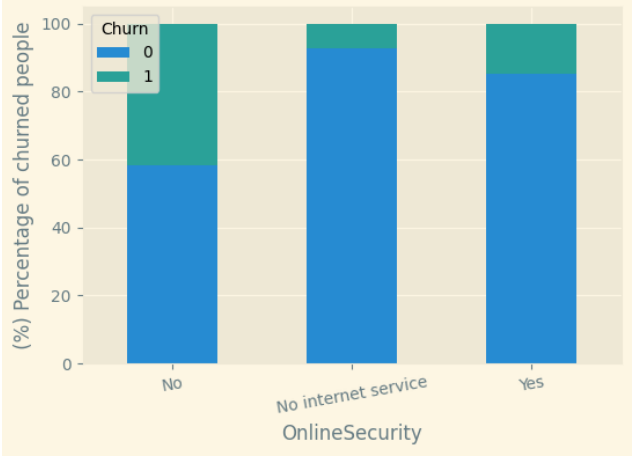


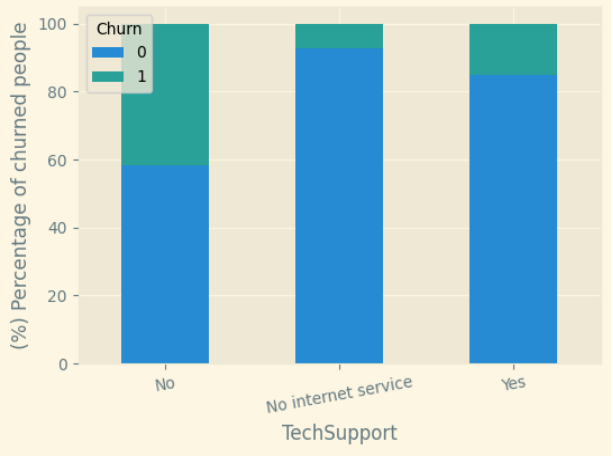
4. Exploratory Data Analysis

Bivariate Analysis



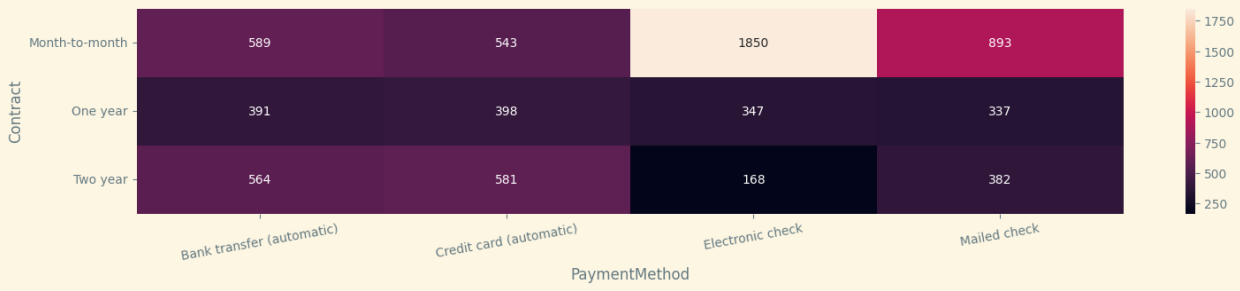


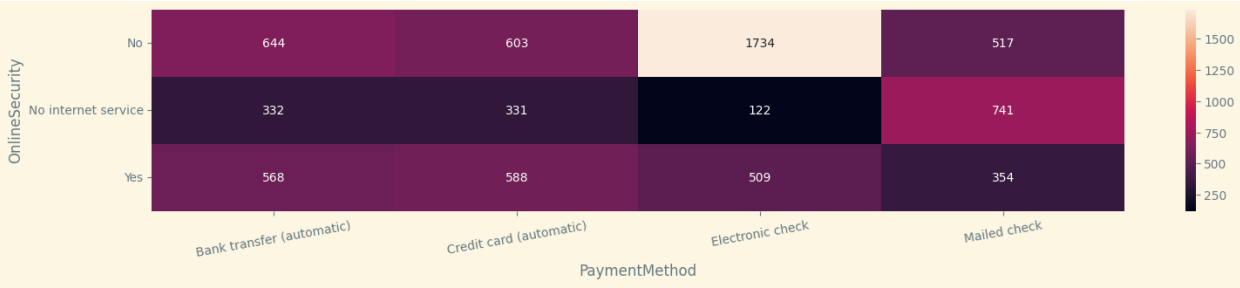


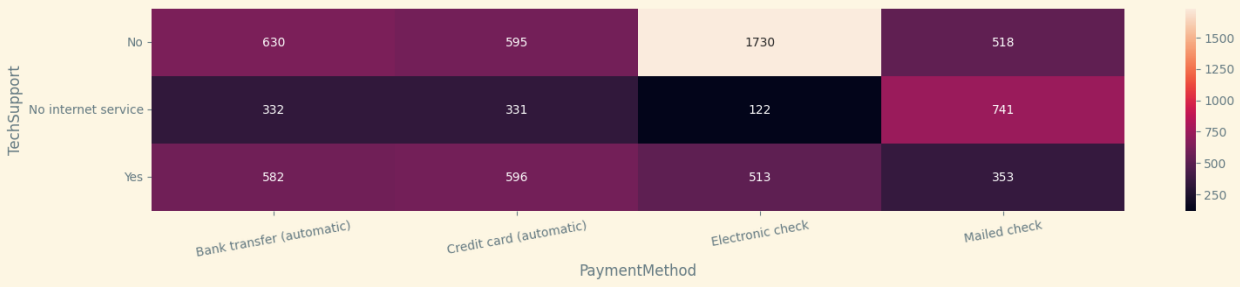


4. Exploratory Data Analysis

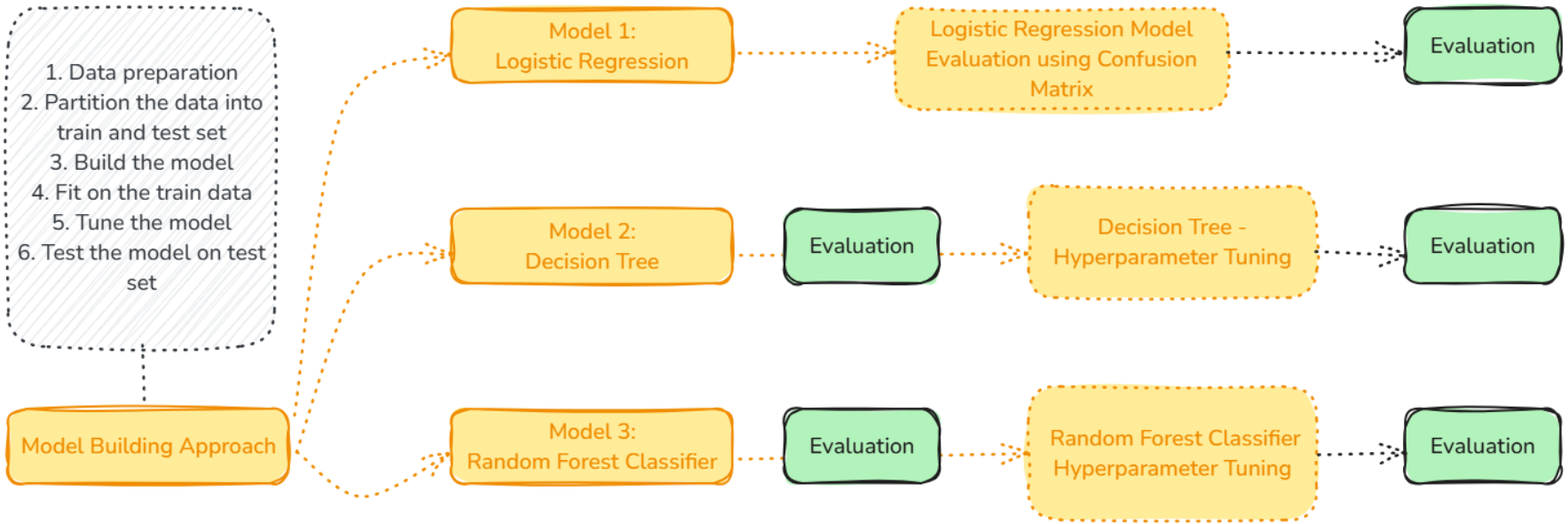
Review of the relationships of categorical data







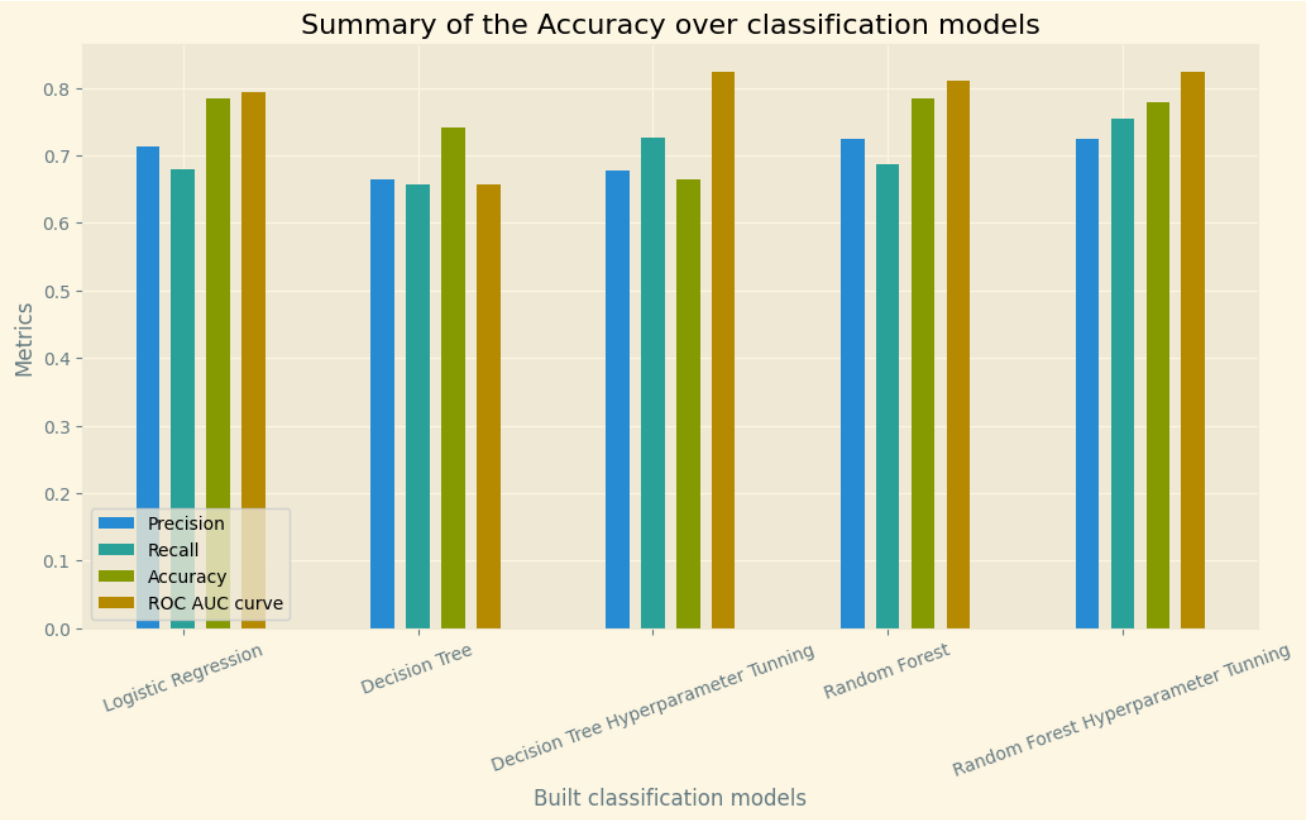
4. Models Building Approach



5. Demo

6. Results and Analysis, and Conclusion

Visualize the built model metrics (Precision, Recall, Accuracy, ROC AUC curve)



Thank you