**PYTHON PROJECT**

-Arjuna Menon Puzhankarai

The goal of this project is to implement a predictive churn model, which is a system that determines the factors responsible for a customer leaving a product or service, and to what extent.

**CODE:**

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

# 1. Data collection

data=pd.read\_csv('Churn\_Modelling.csv')

data.head()

# 2. Question considered - What factors affect the chances of chances of a customer leaving the #company and to what level?

# 3. Clean the data

#Dropping the columns RowNumber, CustomerId and Surname since they aren't necessary for #determining the answer to the posed question.

data=data.drop(['RowNumber','CustomerId','Surname'],axis=1)

data.head()

#Mapping non-numerical data to numerical values: Convert female and male to 0 and 1 in Gender #column and France, Spain and Germany to 0,1 & 2 in Geography column.

data['Gender']=data['Gender'].map({'Female':0,'Male':1})

data['Geography']=data['Geography'].map({'France':0,'Spain':1,'Germany':2})

data.head()

#4. Data Analysing

#Box Plots

fig1=plt.figure(figsize=(18,10))

plt.subplot(2,3,1)

sns.boxplot(x='Exited',y='CreditScore',data=data)

plt.title("Does credit score influence exit of a customer?")

plt.subplot(2,3,2)

sns.boxplot(x='Exited',y='Age',data=data)

plt.title("Does age influence exit of a customer?")

plt.subplot(2,3,3)

sns.boxplot(x='Exited',y='Tenure',data=data)

plt.title("Does tenure influence exit of a customer?")

plt.subplot(2,3,4)

sns.boxplot(x='Exited',y='Balance',data=data)

plt.title("Does balance influence exit of a customer?")

plt.subplot(2,3,5)

sns.boxplot(x='Exited',y='EstimatedSalary',data=data)

plt.title("Does salary influence exit of a customer?")

plt.show()

#Count Plots

fig2=plt.figure(figsize=(17,10))

plt.subplot(2,3,1)

sns.countplot(x="Exited", hue="Geography", data=data, order=[0,1]).set\_title("Does geography affect exit?")

plt.subplot(2,3,2)

sns.countplot(x="Exited", hue="Gender", data=data, order=[0,1]).set\_title("Does gender affect exit?")

plt.subplot(2,3,3)

sns.countplot(x='Exited', hue="HasCrCard", data=data, order=[0,1]).set\_title("Does having a CR card affect exit?")

plt.subplot(2,3,4)

sns.countplot(x='Exited', hue="IsActiveMember", data=data, order=[0,1]).set\_title("Does being an active member affect exit?")

plt.subplot(2,3,5)

sns.countplot(x='Exited', hue="NumOfProducts", data=data, order=[0,1]).set\_title("Does number of products affect exit?")

plt.show()

# Correlation and heat map

print(data.corr())

sns.heatmap(data.corr())

**OUTPUTS AND EXPLANATION:**

**Correlation results:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **CreditScore** | **Age** | **Tenure** | **Balance** | **EstimatedSalary** |
| **Exited** | **-0.027** | **0.285** | **-0.014** | **0.118** | **0.012** |

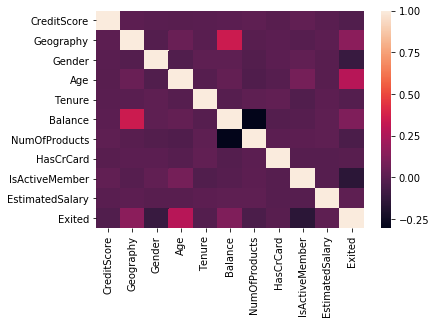
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Geography** | **Gender** | **HasCrCard** | **IsActiveMember** | **NumOfProducts** |
| **Exited** | **0.154** | **-0.106** | **-0.007** | **-0.156** | **-0.048** |

Positive correlation: As the x values increase, y values should also increase.

Negative correlation: As the x values increase, y values decrease.

Hence, here positive correlation means as x values increase, the chances of exiting (1) should increase.

**Heat map:**



This is a heatmap of the Pearson correlation coefficients between the variables.

**Box plots:**

**Interpretation of box plots:**

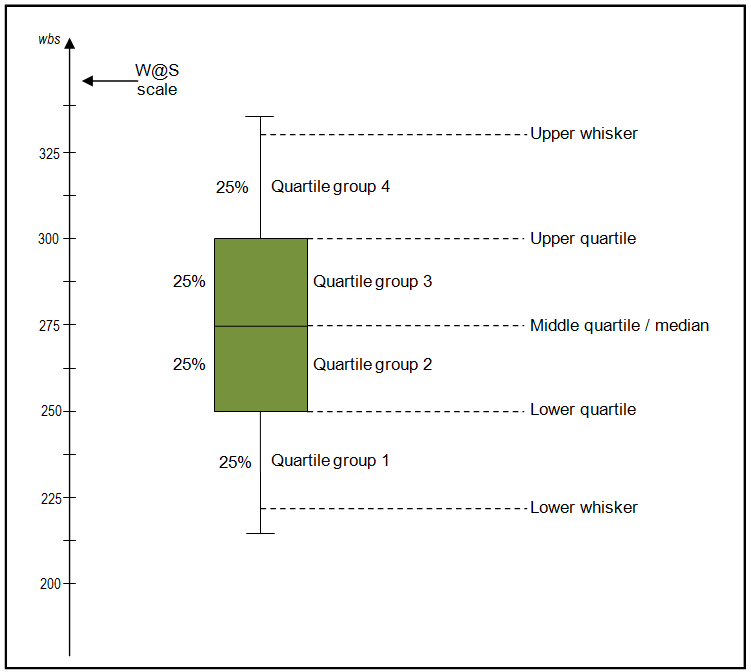
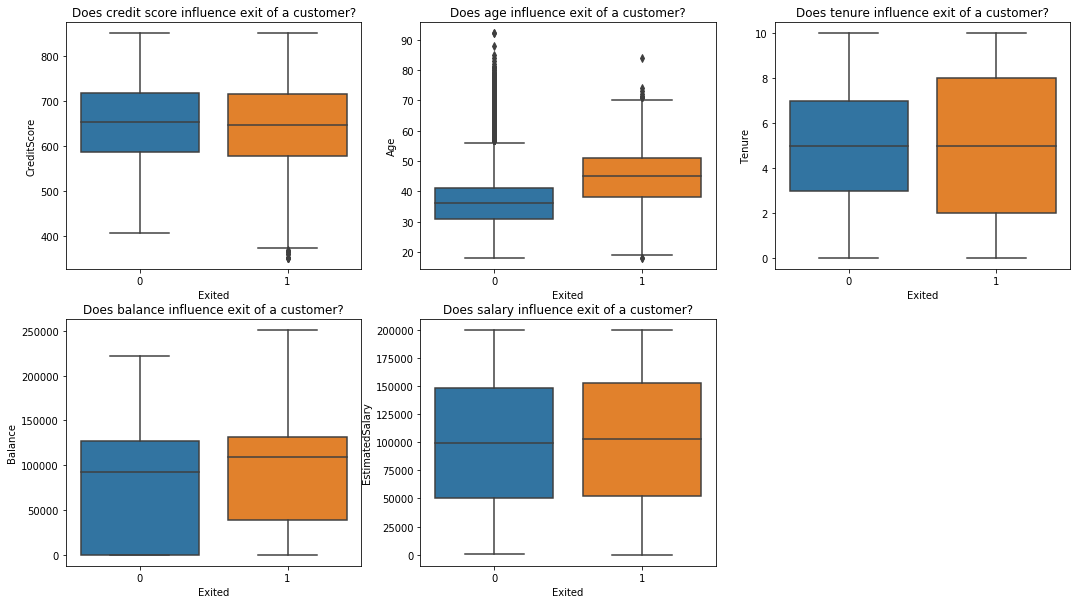


Image obtained from https://www.wellbeingatschool.org.nz/information-sheet/understanding-and-interpreting-box-plots.



1. The credit scores for 0 (not exited) and 1 (exited) are not all that different but we can see that the mean of credit scores for 1 is a little lower so they have a slight negative correlation (-0.03).

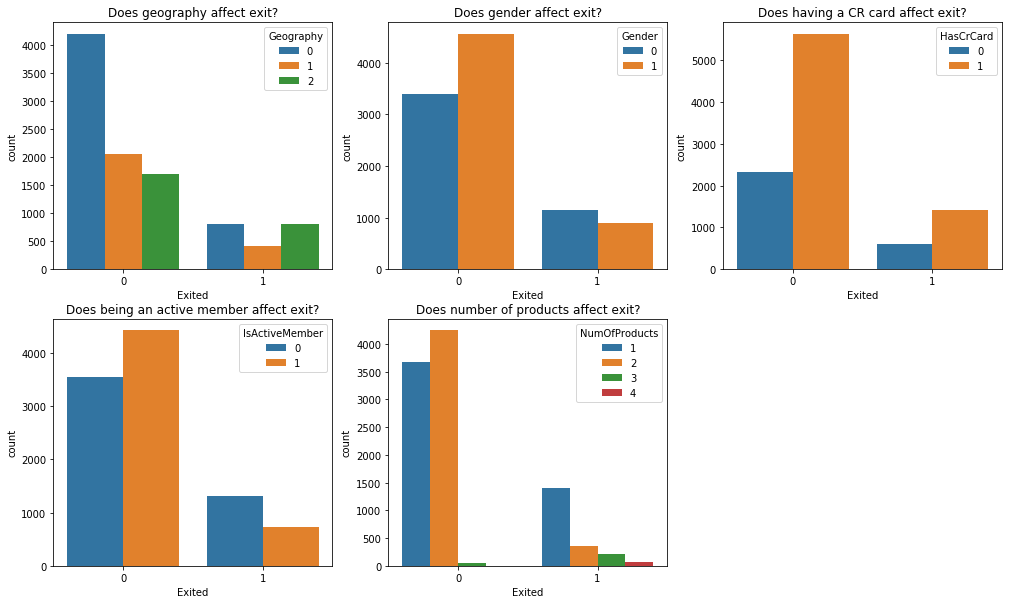
2. It is clear from this plot that higher the age, higher the chances (mean for age is higher for 1) of the person exiting the company. Thus, they have a positive correlation (0.28).

3. While the means of tenure are almost equal here for 0 and 1, hence almost 0 correlation (-0.01). Negative since mean is actually the slightest bit lower for 1.

4. The balance mean is higher for 1 and hence, higher balance means there’s a higher chance of the person exiting the company. Thus, they have a positive correlation (0.12). Also, the lower whisker is not visible here since 25% of the data has a balance of 0 which can be seen from the table.

5. The mean for estimated salary is slightly higher for 1 and hence, they have positive correlation (0.01).

**Count plots:**



1. 0-France, 1-Spain, 2-Germany

For Geography = 1, more chance of not exiting (0) than for exiting (1) while for Geography = 2, more chance of exiting (1) than for not exiting (0). Hence, positive correlation (0.15).

2. 0-Female, 1-Male

For Gender = 0, more chance of exiting (1) than not exiting (0) while for Gender = 1, more chance of not exiting (0) than for exiting (1). Hence, negative correlation (-0.11).

3. No correlation according to this for HasCrCard and exited, and from the correlation matrix, we find there’s a correlation of -0.01 which is nearly zero.

4. For IsActiveMember = 0, more chance of exiting (1) than not exiting (0) while for IsActiveMember = 1, more chance of not exiting (0) than for exiting (1). Hence, negative correlation (-0.16).

5. For NumOfProducts = 1, more chance of exiting (1) than not exiting (0) while for NumOfProducts = 2, more chance of not exiting (0) than for exiting (1). Hence, negative correlation (-0.05).