Churning of Customers

Every bank wants to hold their customers for sustaining their business and thus this Anonymous Multinational bank. You have customer data of account holders at Anonymous Multinational Bank with the aim of understanding

Exploring the correlation between variables such as credit score, age, tenure, balance, and geography with customer churn. Assess the impact of demographic factors like gender and the presence of credit cards on churn rates.

Additionally, analyze customer satisfaction scores and complaint resolutions to identify areas for service improvement. Utilize your analytics skills to find factors contributing to potential churn based. This project provides an opportunity to enhance customer retention strategies by uncovering patterns and insights within the dataset.

In [169…
```
#importing libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [170…
```
#loading data
!gdown 1q1Mh3Mm4kv1LitxWcdY6--gNHVmuAfPP
```

```
Downloading...
From: https://drive.google.com/uc?id=1q1Mh3Mm4kv1LitxWcdY6--gNHVmuAfPP
To: /content/Bank-Records.csv
100% 837k/837k [00:00<00:00, 107MB/s]
```

In [171…
```
bank = pd.read_csv("Bank-Records.csv")
```

In [172…
```
bank.head()
```

Out[172]:

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited	Complain	Satisfaction Score	Card Type	Point Earned
0	1	15634602	Hargrave	619	France	Female	42	2	0.00	1	1	1	101348.88	1	1	2	DIAMOND	464
1	2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0	1	3	DIAMOND	456
2	3	15619304	Onio	502	France	Female	42	8	159660.80	3	1	0	113931.57	1	1	3	DIAMOND	377
3	4	15701354	Boni	699	France	Female	39	1	0.00	2	0	0	93826.63	0	0	5	GOLD	350
4	5	15737888	Mitchell	850	Spain	Female	43	2	125510.82	1	1	1	79084.10	0	0	5	GOLD	425

In [173…
```
#Shape
bank.shape
```

Out[173]:
```
(10000, 18)
```

Observation

There are 10000 rows and 18 columns in the dataset.

In [174…
```
#info
bank.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 18 columns):
 #   Column              Non-Null Count  Dtype  
---  ------              --------------  -----  
 0   RowNumber           10000 non-null  int64  
 1   CustomerId          10000 non-null  int64  
 2   Surname             10000 non-null  object 
 3   CreditScore         10000 non-null  int64  
 4   Geography           10000 non-null  object 
 5   Gender              10000 non-null  object 
 6   Age                 10000 non-null  int64  
 7   Tenure              10000 non-null  int64  
 8   Balance             10000 non-null  float64
 9   NumOfProducts       10000 non-null  int64  
 10  HasCrCard           10000 non-null  int64  
 11  IsActiveMember      10000 non-null  int64  
 12  EstimatedSalary     10000 non-null  float64
 13  Exited              10000 non-null  int64  
 14  Complain            10000 non-null  int64  
 15  Satisfaction Score  10000 non-null  int64  
 16  Card Type           10000 non-null  object 
 17  Point Earned        10000 non-null  int64  
dtypes: float64(2), int64(12), object(4)
memory usage: 1.4+ MB
```

Observartion

There are 10000 values in each and every column of the data. From this we can say there are no null values in the dataset.

In [175…
```
bank.duplicated().any()
```
Out[175]:
```
False
```

Observation

There are no duplicated values in the data.

In [176…
```
bank.head()
```
Out[176]:

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited	Complain	Satisfaction Score	Card Type	Point Earned
0	1	15634602	Hargrave	619	France	Female	42	2	0.00	1	1	1	101348.88	1	1	2	DIAMOND	464
1	2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0	1	3	DIAMOND	456
2	3	15619304	Onio	502	France	Female	42	8	159660.80	3	1	0	113931.57	1	1	3	DIAMOND	377
3	4	15701354	Boni	699	France	Female	39	1	0.00	2	0	0	93826.63	0	0	5	GOLD	350
4	5	15737888	Mitchell	850	Spain	Female	43	2	125510.82	1	1	1	79084.10	0	0	5	GOLD	425

In [177…
```
gender_cnt = bank["Gender"].value_counts()
gender_cnt
```
Out[177]:
```
Male      5457
Female    4543
Name: Gender, dtype: int64
```

Observation

There are total 5457 male and 4543 female customers in the entire dataset

In [178…
```
country = bank["Geography"].value_counts()
country
```

Out[178]:
```
France     5014
Germany    2509
Spain      2477
Name: Geography, dtype: int64
```

Observation

There are 3 geographical regions in the data. They are France, Germany, Spain with a value of 5014, 2509 and 2477 respectively

In [179…
```
age = bank["Age"].value_counts()
age.head(20)
```

Out[179]:
```
37    478
38    477
35    474
36    456
34    447
33    442
40    432
39    423
32    418
31    404
41    366
29    348
30    327
42    321
43    297
28    273
44    257
45    229
46    226
27    209
Name: Age, dtype: int64
```

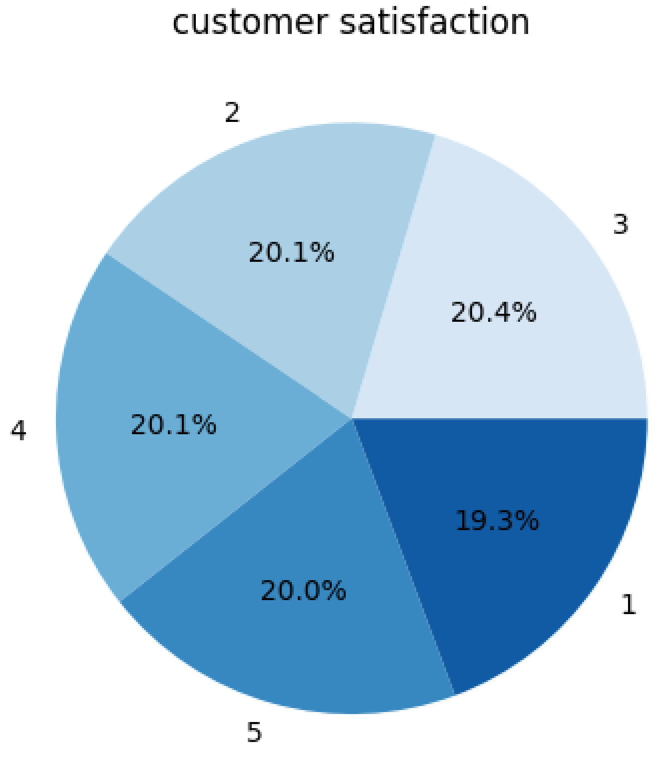
In [180…
```
sat_score = bank['Satisfaction Score'].value_counts()
sat_score
```

Out[180]:
```
3    2042
2    2014
4    2008
5    2004
1    1932
Name: Satisfaction Score, dtype: int64
```

Observation

There are a total 5 categories of Satisfaction Score. All the categories with their respected values were reported. It is evident that satisfaction score 3 has value count of 2042. A pieplot was plotted below for the same.

In [181…
```
colors = sns.color_palette('Blues', 5)
plt.pie(x= bank["Satisfaction Score"].value_counts(), labels= ["3","2","4","5","1"], autopct='%1.1f%%', colors = colors)
plt.title("customer satisfaction")
plt.show()
```



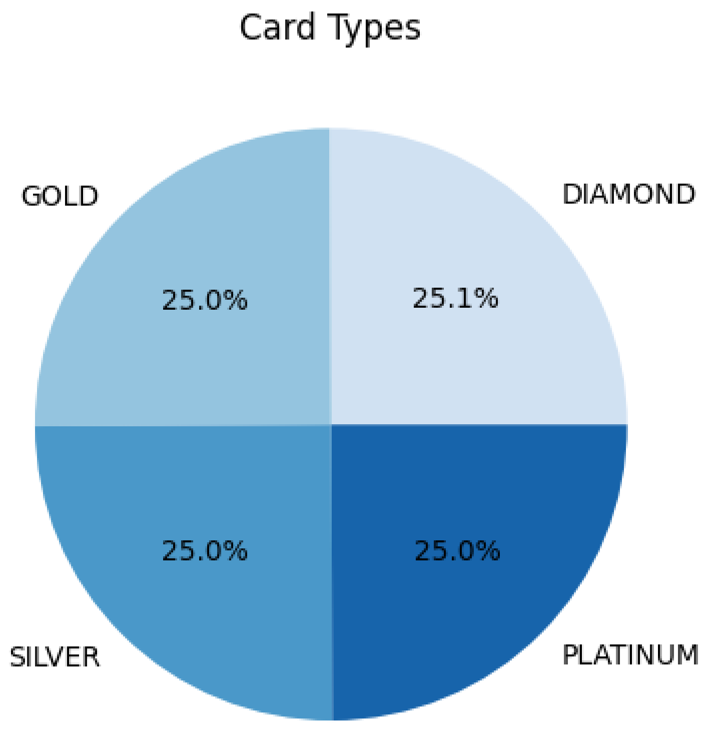
In [182…
```
cards =  bank["Card Type"].value_counts()
cards
```

Out[182]:
```
DIAMOND     2507
GOLD        2502
SILVER      2496
PLATINUM    2495
Name: Card Type, dtype: int64
```

Observation

Types of credit cards with there respective values are calculated. A lot of customers were found to hold Diamond and Gold CreditCards.

In [183…
```
colors = sns.color_palette('Blues', 4)
plt.pie(x= bank["Card Type"].value_counts(), labels = ["DIAMOND", "GOLD", "SILVER", "PLATINUM"],autopct='%1.1f%%', colors = colors)
plt.title("Card Types")
plt.show()
```



In [184…

```
exit_cnt = bank["Exited"].value_counts()
exit_cnt
```

Out[184]:

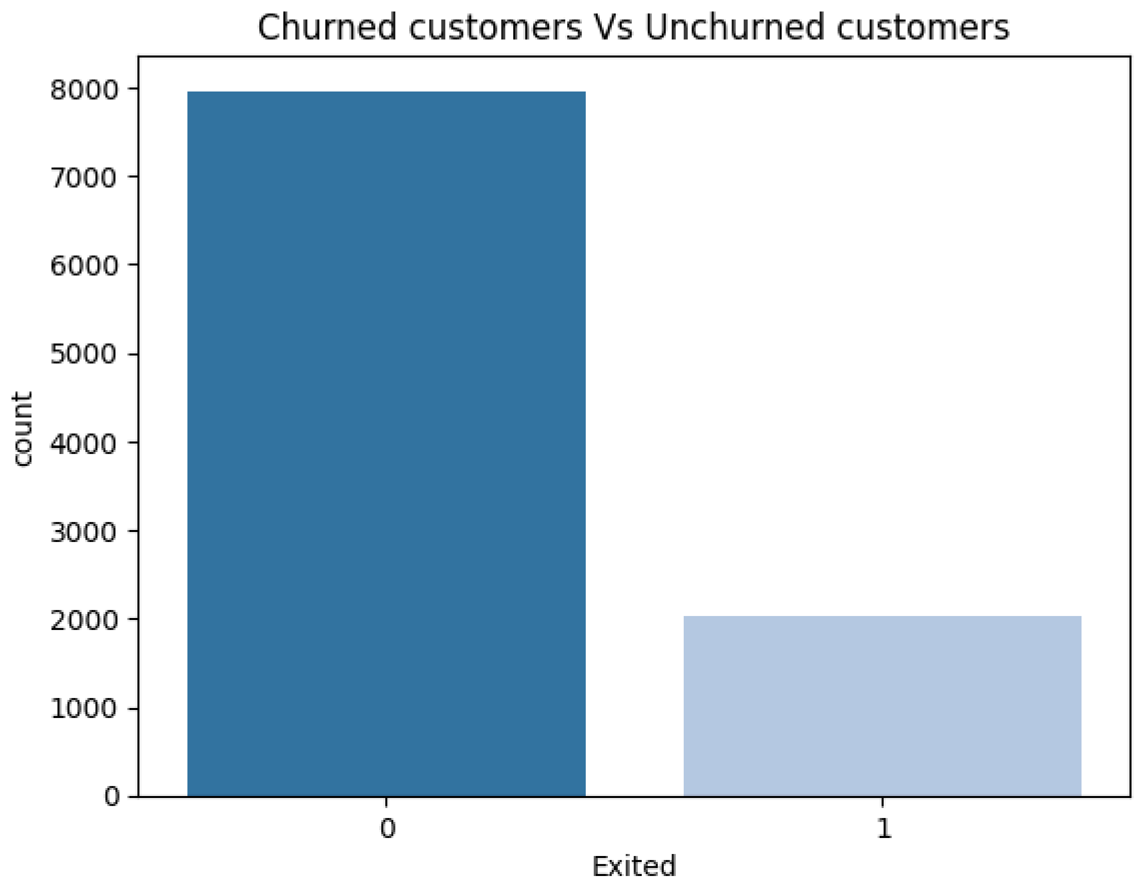
```
0    7962
1    2038
Name: Exited, dtype: int64
```

Observation

From the above observation we can say 2038 customers has exited from the bank.

In [185…

```
sns.countplot(data= bank, x= "Exited", hue = 'Exited', palette="tab20", legend = False)
plt.title("Churned customers Vs Unchurned customers")
plt.show()
```

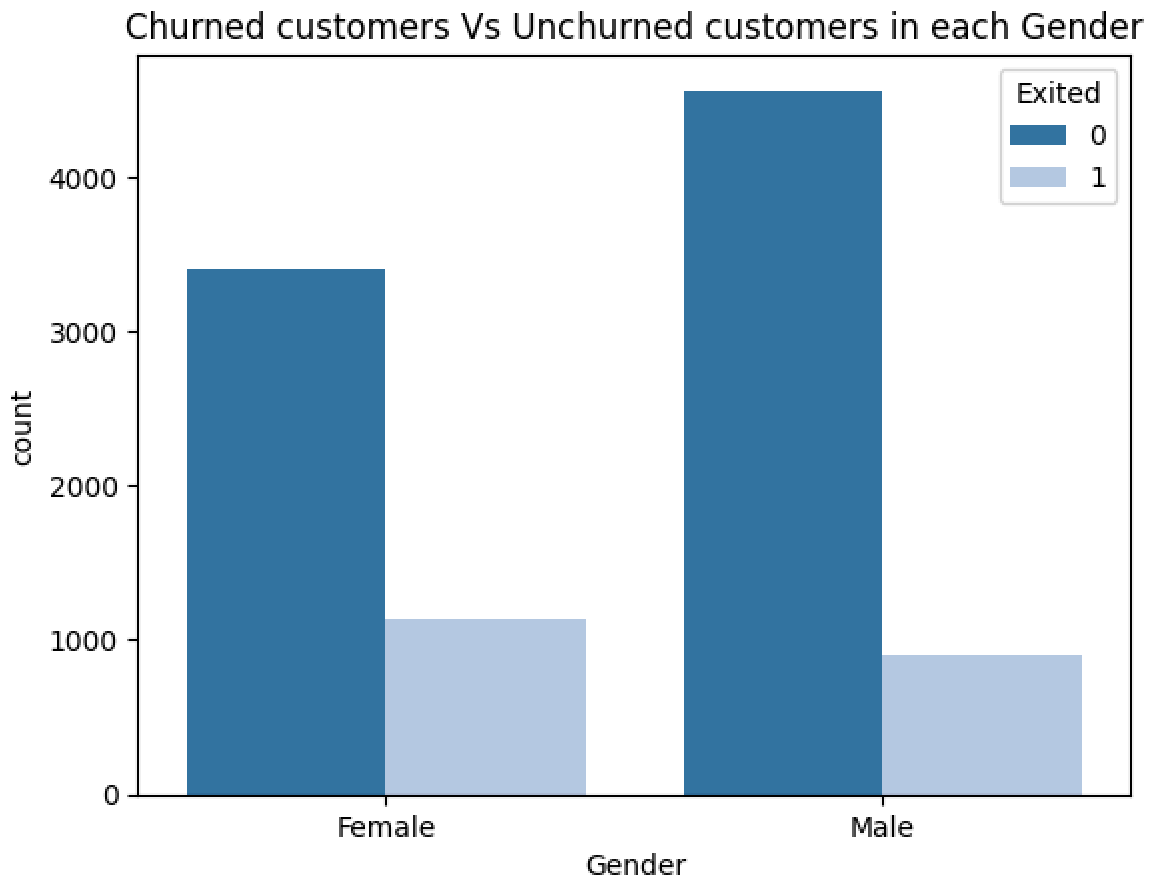


In [186…

```
sns.countplot(data = bank, x= "Gender", hue = 'Exited', palette="tab20")
plt.title("Churned customers Vs Unchurned customers in each Gender")
```

Out[186]:

```
Text(0.5, 1.0, 'Churned customers Vs Unchurned customers in each Gender')
```



In [187…

```
#Top 10 customers with highest credit score
credit_score = bank[["CustomerId", "Surname", "CreditScore"]]
credit_score = credit_score.sort_values("CreditScore", ascending = False).reset_index()
credit_score = credit_score.drop("index", axis =1)
credit_score
```

Out[187]:

	CustomerId	Surname	CreditScore
0	15707473	Summers	850
1	15672640	Kambinachi	850
2	15771580	Davison	850
3	15719793	Watson	850
4	15604536	Vachon	850
...	...	...	...
9995	15765173	Lin	350
9996	15685372	Azubuike	350
9997	15668309	Maslow	350
9998	15758813	Campbell	350
9999	15803202	Onyekachi	350

10000 rows × 3 columns

Observation

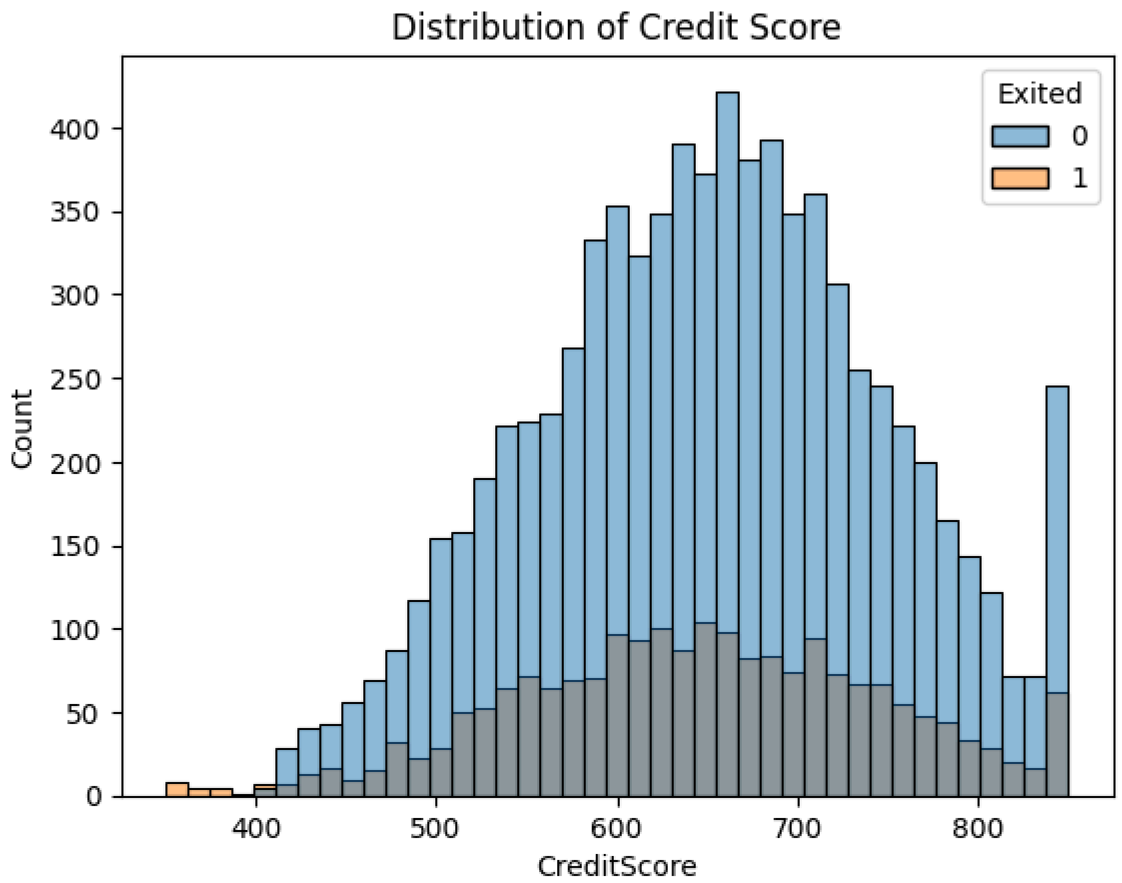
from the above table we can observe that the highest credit score value in the data is 850.

In [188…

```
sns.histplot(data =bank, x="CreditScore", hue = "Exited")
plt.title("Distribution of Credit Score")
```

Out[188]:

```
Text(0.5, 1.0, 'Distribution of Credit Score')
```



Observation

Distribution of credit score of the banking data was plotted using a histplot and found to be have a large number of customers with a score between 580 to 740.

In [189…

```
age_cat = bank["Age"].value_counts().head(10)
age_cat
```

Out[189]:

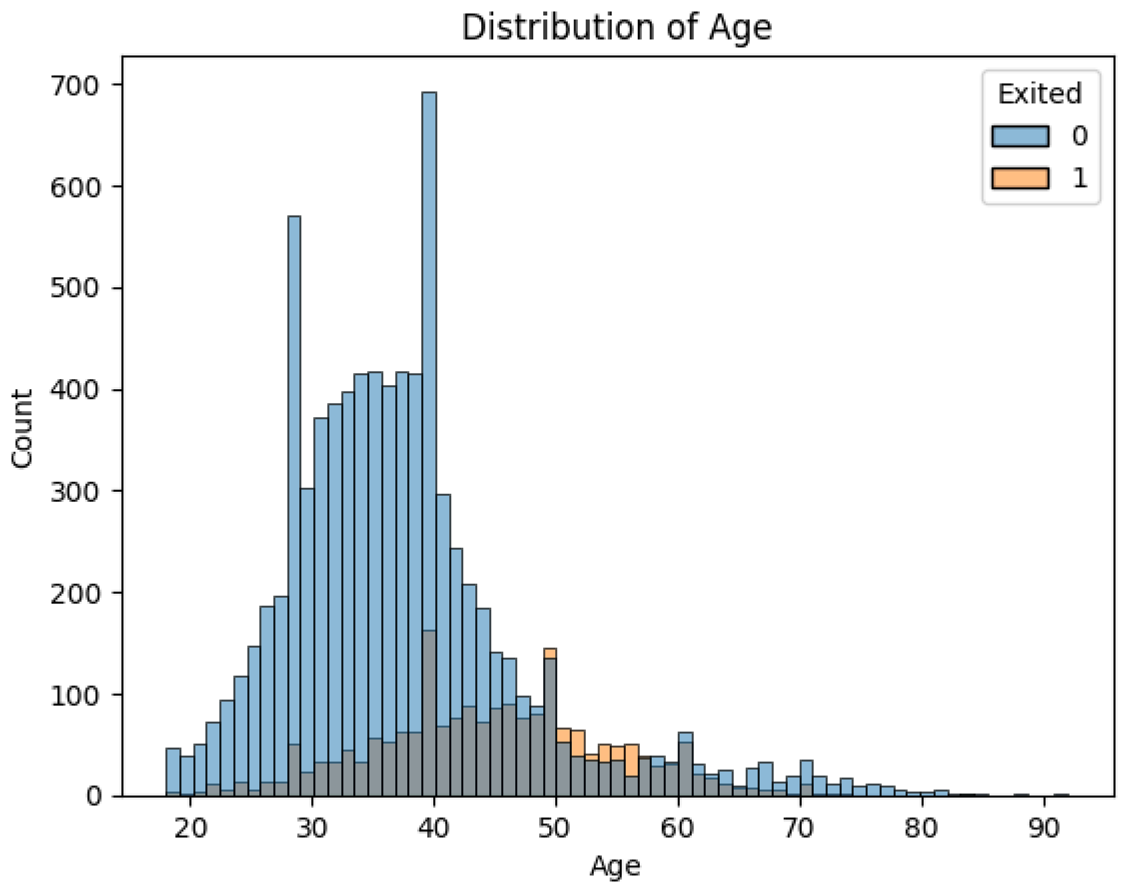
```
37    478
38    477
35    474
36    456
34    447
33    442
40    432
39    423
32    418
31    404
Name: Age, dtype: int64
```

In [190…

```
sns.histplot(data = bank,x="Age", hue ="Exited")
plt.title("Distribution of Age")
```

Out[190]:

```
Text(0.5, 1.0, 'Distribution of Age')
```



Observation

Distribution of age was plotted using histplot and found that most of the customers are with age 37.

In [191…

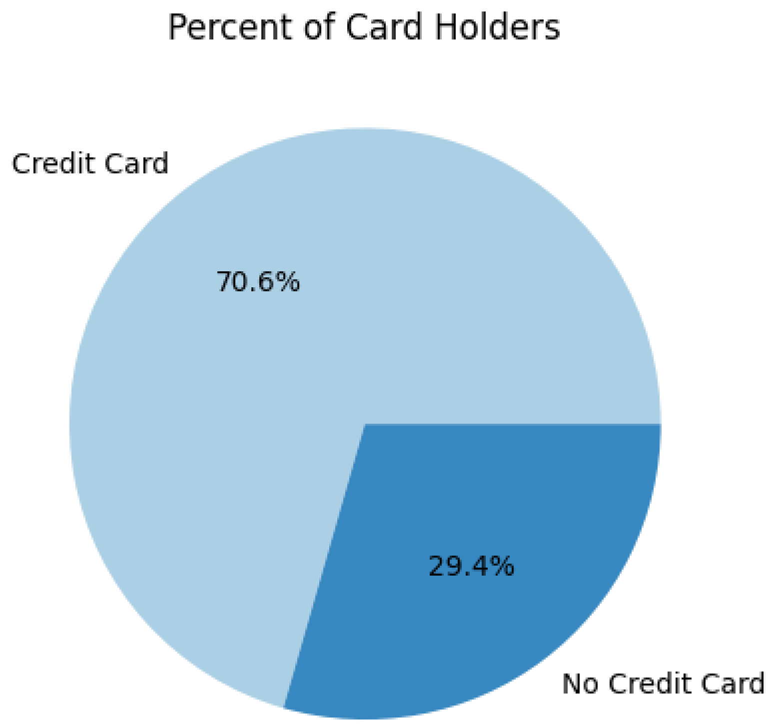
```
card_holders = bank["HasCrCard"].value_counts()
card_holders
```

Out[191]:

```
1    7055
0    2945
Name: HasCrCard, dtype: int64
```

In [192…

```
colors = sns.color_palette('Blues', 2)
plt.pie(data = bank, x= bank["HasCrCard"].value_counts(), labels=["Credit Card", "No Credit Card"], autopct='%1.1f%%', colors = colors)
plt.title("Percent of Card Holders")
plt.show()
```



Observation

A pie plot of customers having credit card and not having customers was plotted and was found 70% of the customers are using credit card

In [193…

```
bank.head()
```

Out[193]:

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited	Complain	Satisfaction Score	Card Type	Point Earned
0	1	15634602	Hargrave	619	France	Female	42	2	0.00	1	1	1	101348.88	1	1	2	DIAMOND	464
1	2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0	1	3	DIAMOND	456
2	3	15619304	Onio	502	France	Female	42	8	159660.80	3	1	0	113931.57	1	1	3	DIAMOND	377
3	4	15701354	Boni	699	France	Female	39	1	0.00	2	0	0	93826.63	0	0	5	GOLD	350
4	5	15737888	Mitchell	850	Spain	Female	43	2	125510.82	1	1	1	79084.10	0	0	5	GOLD	425

In [194…

```
complaint_cnt = bank.groupby(["Complain"])["Exited"].value_counts()
complaint_cnt
```

Out[194]:

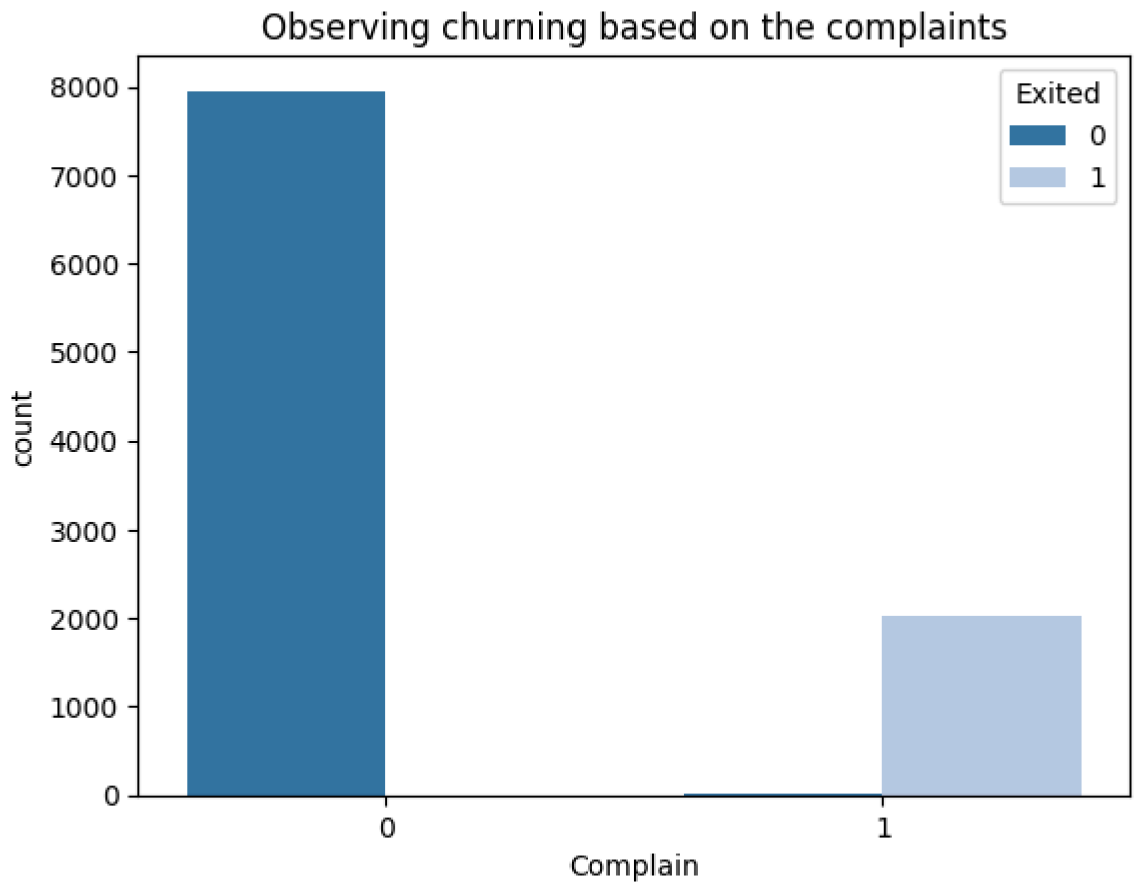
```
Complain  Exited
0         0         7952
          1            4
1         1         2034
          0           10
Name: Exited, dtype: int64
```

In [195…

```
sns.countplot(data = bank, x= "Complain", hue = "Exited", palette="tab20")
plt.title("Observing churning based on the complaints")
```

Out[195]:

```
Text(0.5, 1.0, 'Observing churning based on the complaints')
```



In [196…

```
card_user_cnt = bank.groupby(["Card Type"])["Exited"].value_counts()
card_user_cnt
```

Out[196]:

```
Card Type  Exited
DIAMOND    0         1961
           1          546
GOLD       0         2020
           1          482
PLATINUM   0         1987
           1          508
SILVER     0         1994
           1          502
Name: Exited, dtype: int64
```

Observation

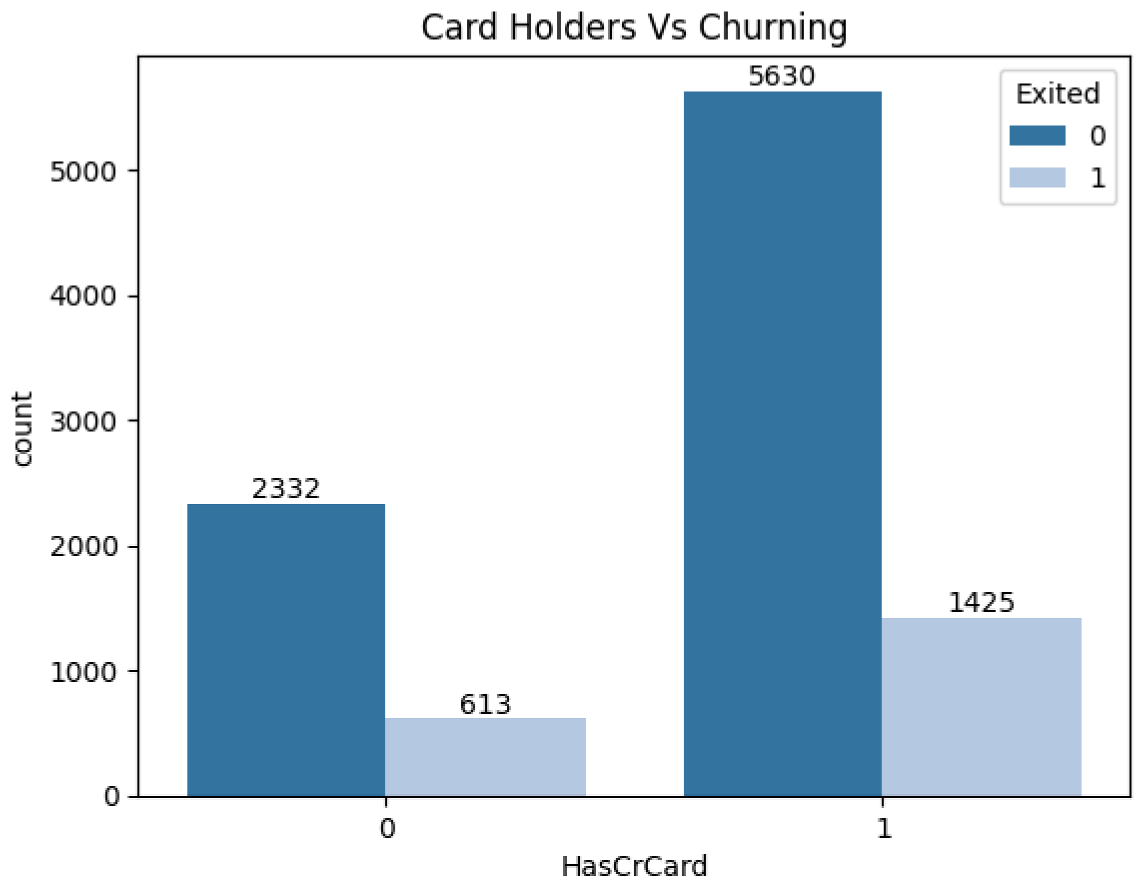
Number of customers who are still a customer of the bank in each credit card type was calculated. we can observe that 546,484, 508 and 502 customers have left the bank from Diamond, Gold, Platinum, and Sliver credit card types. A countplot was plotted below for the same.

In [197…

```
card_holder = sns.countplot(data = bank, x= "HasCrCard", hue= "Exited", palette="tab20")
for i in card_holder.containers:
      card_holder.bar_label(i)
plt.title("Card Holders Vs Churning")
```

Out[197]:

```
Text(0.5, 1.0, 'Card Holders Vs Churning')
```



Observation

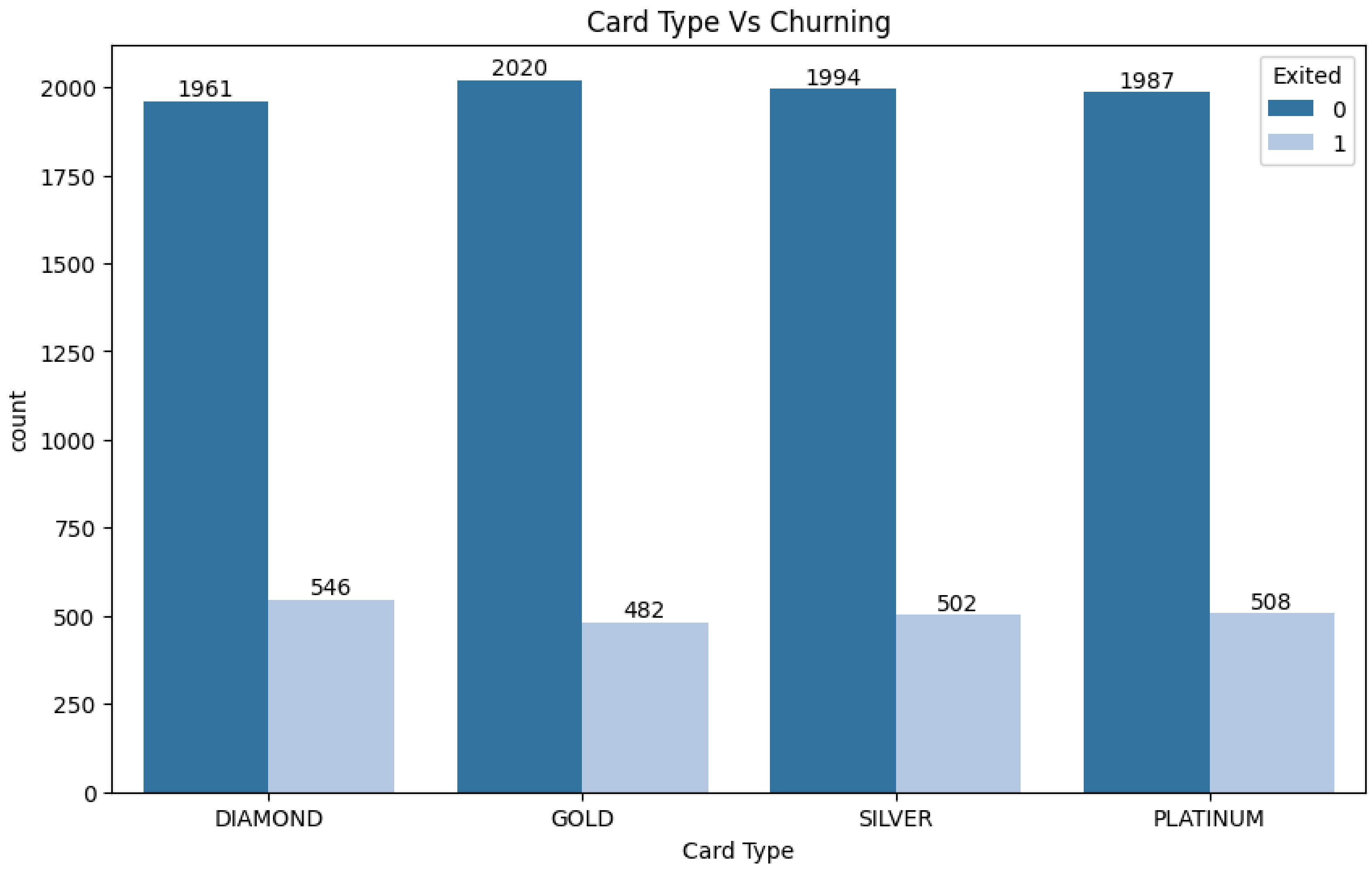
Customers holding credit card was plotted as countplot. Nearly 7000 customers hold a credit card and out of which 1425 customers left the bank. There are 2332 customers who are still a part of the bank have no cards.

In [198…

```
plt.figure(figsize = (10,6))
card_users = sns.countplot(data = bank, x= "Card Type", hue= 'Exited', palette="tab20")
for container in card_users.containers:
    card_users.bar_label(container)
plt.title("Card Type Vs Churning")
```

Out[198]:

```
Text(0.5, 1.0, 'Card Type Vs Churning')
```



Observation

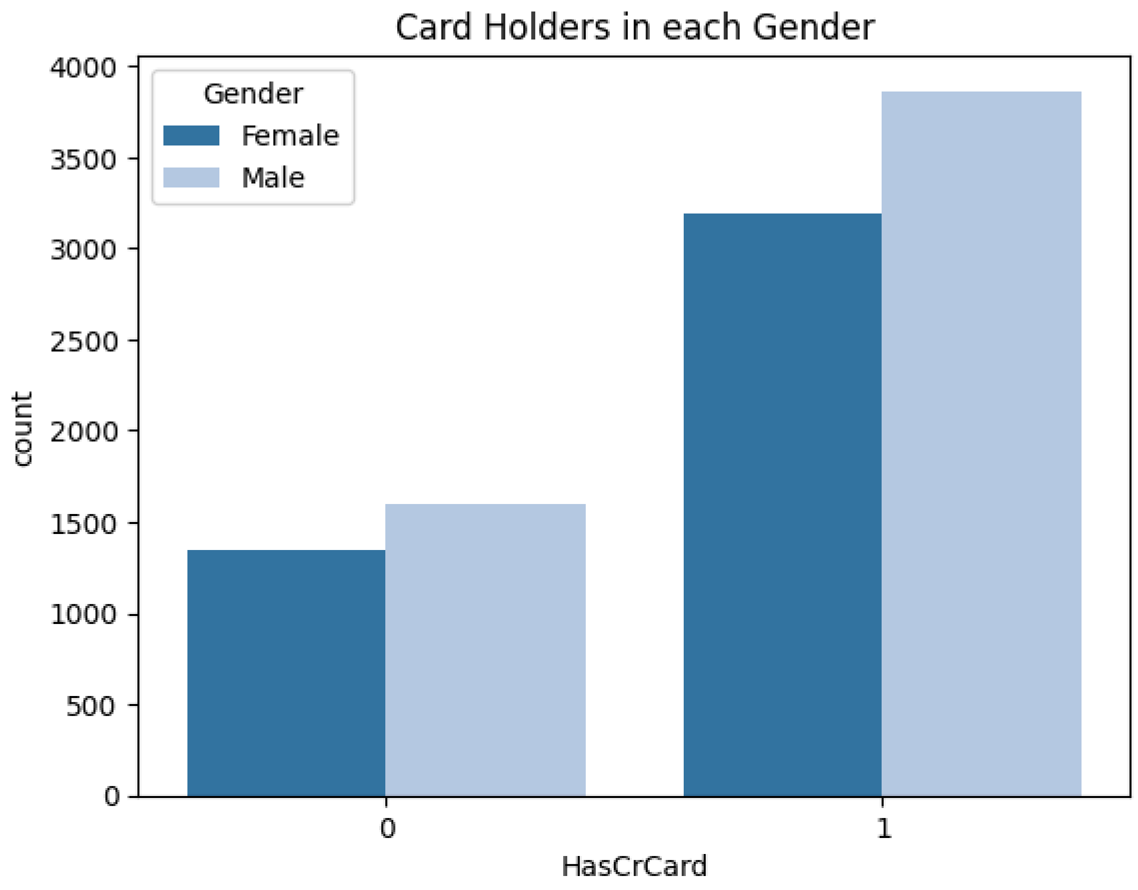
A countplot representing the number of customers in each card type who are still a customer at the bank and number of customers left the bank. We can say Gold has the highest customer count at the bank and Diamond has the highest count of customers left the bank.

In [199…

```
sns.countplot(data= bank,x = "HasCrCard", hue = "Gender", palette="tab20")
plt.title("Card Holders in each Gender")
```

Out[199]:

```
Text(0.5, 1.0, 'Card Holders in each Gender')
```



Observation

A countplot on customers holding a credit card and not holding a credit card for both males and females was plotted and we can observe males have both higher number of credit card users and have no credit card compared to females

In [200…

```
users_per_region =  bank["Geography"].value_counts()
users_per_region
```

Out[200]:

```
France     5014
Germany    2509
Spain      2477
Name: Geography, dtype: int64
```

Observation

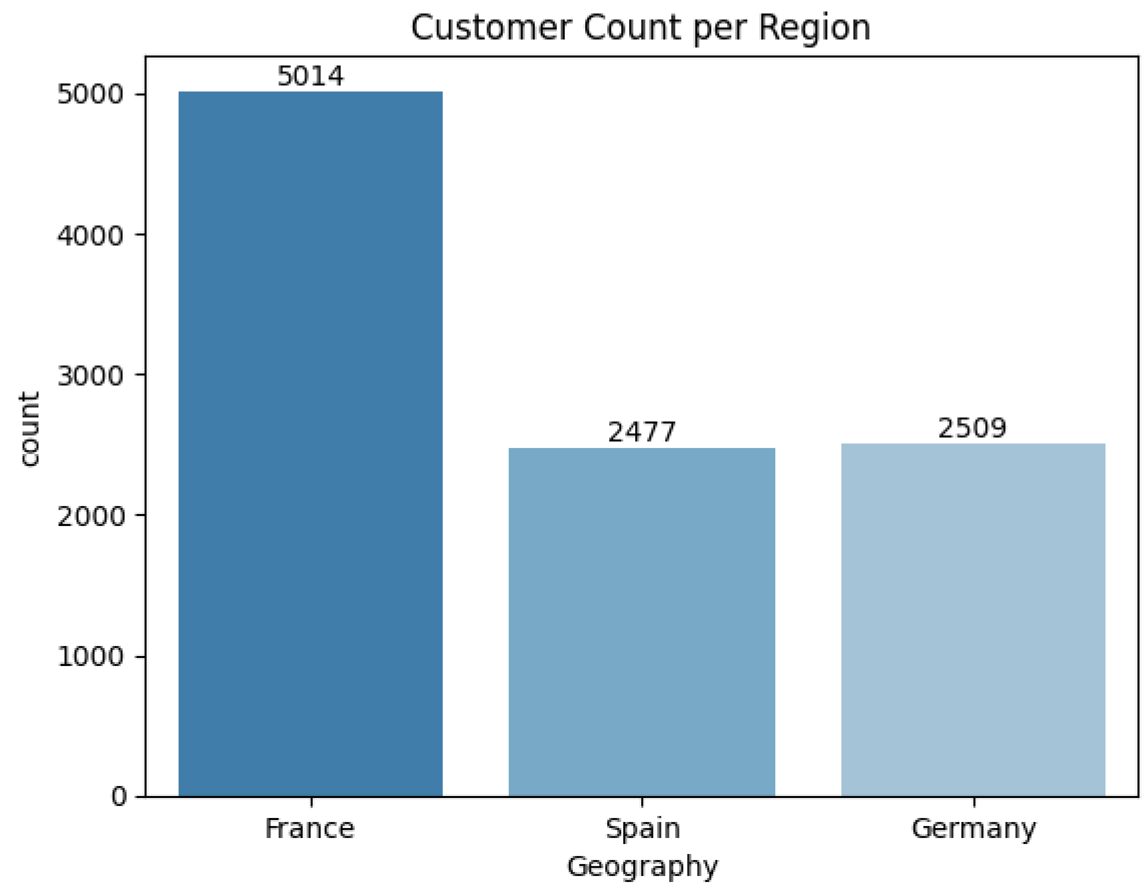
Number of customers from each region was calculated. France have the highest number of customers (double the customers number of Germany).

In [201…

```
region_cnt = sns.countplot(data = bank, x= "Geography", hue = "Geography", palette="tab20c")
for i in region_cnt.containers:
     region_cnt.bar_label(i)
plt.title("Customer Count per Region")
```

Out[201]:

```
Text(0.5, 1.0, 'Customer Count per Region')
```



Observation

A countplot was plotted for the customers count per region. France have the highest number of customers

In [202…

```
bank.groupby("Geography")["Gender"].value_counts()
```

Out[202]:

```
Geography  Gender
France     Male      2753
           Female    2261
Germany    Male      1316
           Female    1193
Spain      Male      1388
           Female    1089
Name: Gender, dtype: int64
```

In [203…

```
cardholders_per_region =bank.groupby(["Geography","Gender"])["HasCrCard"].value_counts(normalize = True).reset_index(name = "probability_count")
cardholders_per_region
```

Out[203]:

	Geography	Gender	HasCrCard	probability_count
0	France	Female	1	0.697921
1	France	Female	0	0.302079
2	France	Male	1	0.713767
3	France	Male	0	0.286233
4	Germany	Female	1	0.706622
5	Germany	Female	0	0.293378
6	Germany	Male	1	0.720365
7	Germany	Male	0	0.279635
8	Spain	Female	1	0.707989
9	Spain	Female	0	0.292011
10	Spain	Male	1	0.684438
11	Spain	Male	0	0.315562

Observation

Number of female and male customers holding a credit card from each region were identified. The probability of card holders in both males and females from Spain is same . Also, the probability of customers with card is also same

In [204…

```
bank.groupby("Geography")["Exited"].value_counts()
```

Out[204]:

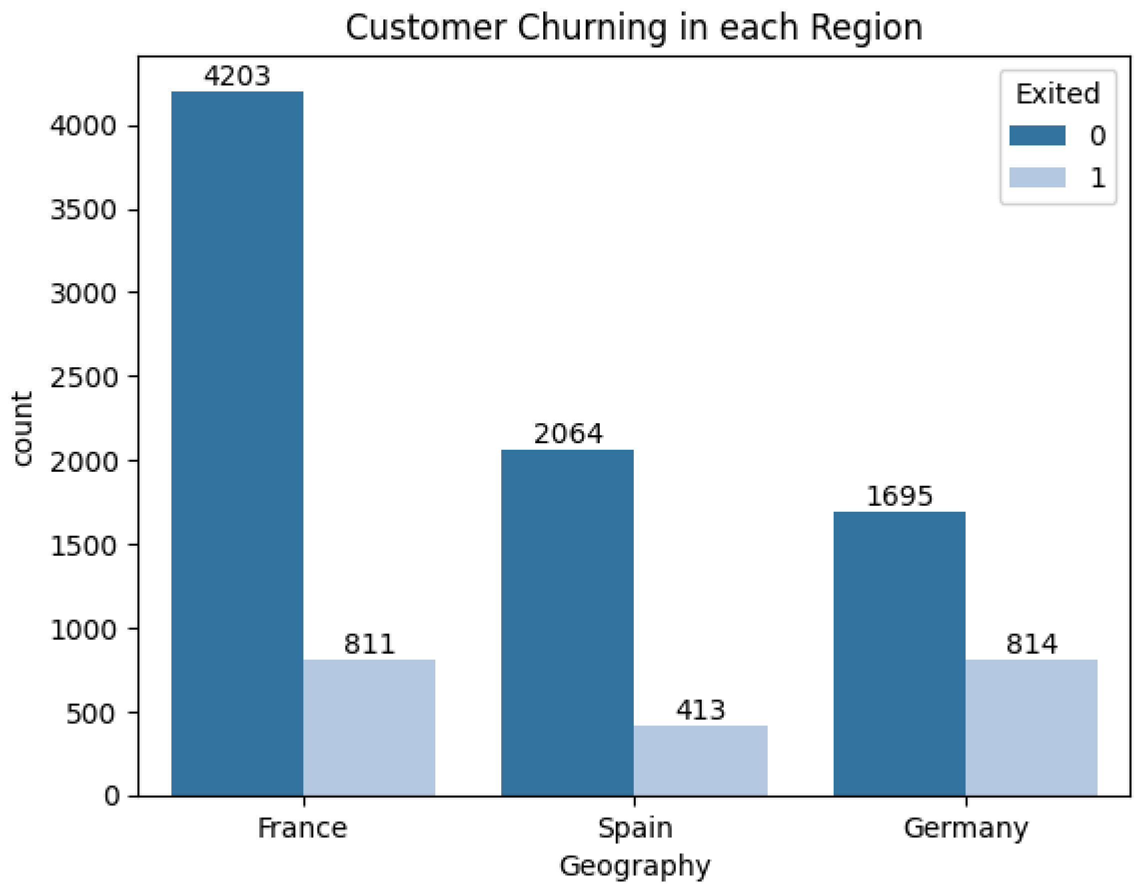
```
Geography  Exited
France     0         4203
           1          811
Germany    0         1695
           1          814
Spain      0         2064
           1          413
Name: Exited, dtype: int64
```

In [205…

```
geo_count = sns.countplot(data = bank, x= 'Geography', hue = 'Exited', palette="tab20")
for i in geo_count.containers:
    geo_count.bar_label(i)
plt.title("Customer Churning in each Region")
```

Out[205]:

```
Text(0.5, 1.0, 'Customer Churning in each Region')
```



Observation

France have highest customers 4203 who are still part of the bank out of 5014. 811 customers from France region are no more having account in the bank. on the other hand Germany has very low number of customers (2509 overall). Out of which 814 customers left the bank.

In [206…

```
gender_per_region = bank.groupby("Geography")["Gender"].value_counts()
gender_per_region
```

Out[206]:

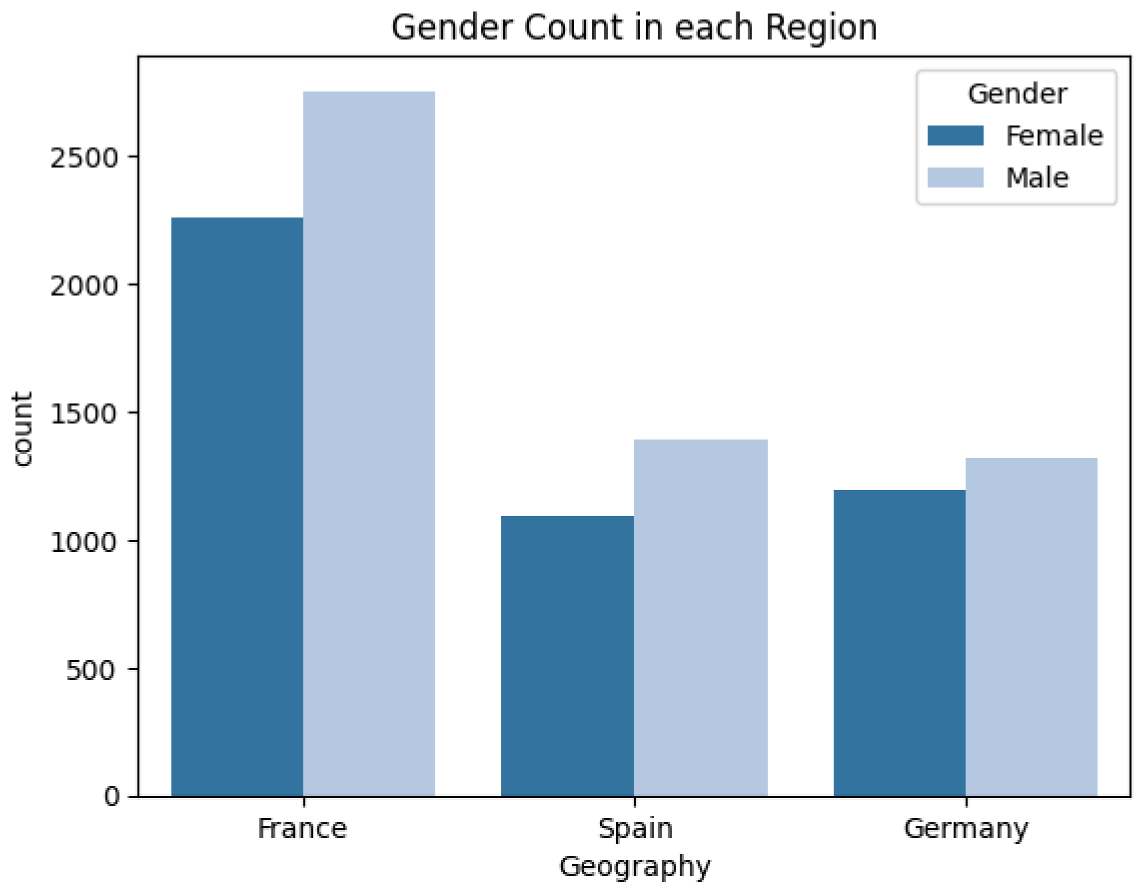
```
Geography  Gender
France     Male      2753
           Female    2261
Germany    Male      1316
           Female    1193
Spain      Male      1388
           Female    1089
Name: Gender, dtype: int64
```

In [207…

```
sns.countplot(data =bank, x= "Geography", hue = "Gender", palette="tab20")
plt.title("Gender Count in each Region")
```

Out[207]:

```
Text(0.5, 1.0, 'Gender Count in each Region')
```



Observation

The number of male customers per region is found to be greater than females. we can also observe that France have greatest number of customers when compared overall.

In [208…

```
gender_per_geography_exited= bank.groupby(["Geography", "Gender"])["Exited"].value_counts(normalize = True).reset_index(name = "cnt")
gender_per_geography_exited["cnt"] = gender_per_geography_exited["cnt"]*100
gender_per_geography_exited
```

Out[208]:

	Geography	Gender	Exited	cnt
0	France	Female	0	79.655020
1	France	Female	1	20.344980
2	France	Male	0	87.250272
3	France	Male	1	12.749728
4	Germany	Female	0	62.447611
5	Germany	Female	1	37.552389
6	Germany	Male	0	72.188450
7	Germany	Male	1	27.811550
8	Spain	Female	0	78.787879
9	Spain	Female	1	21.212121
10	Spain	Male	0	86.887608
11	Spain	Male	1	13.112392

Observation

Number of female and male customers from each region who left the bank was calculated above. From the table, it is clearly evident that the more number of custoemrs who left the bank were found to be females.

In [209…

```
age_count_per_region = bank.groupby(["Geography"])["Age"].value_counts().reset_index(name = "cnt")
age_count_per_region["rank"] = age_count_per_region.groupby("Geography")["cnt"].rank(method = "dense", ascending = False).astype(int)
age_count_per_region = age_count_per_region[age_count_per_region["rank"]<=5]
age_count_per_region_5 = age_count_per_region[age_count_per_region["rank"]<=5]
age_count_per_region_5 = age_count_per_region_5.reset_index()
age_count_per_region_5 = age_count_per_region_5.drop("index", axis = 1)
age_count_per_region_5
```

Out[209]:

	Geography	Age	cnt	rank
0	France	38	249	1
1	France	34	247	2
2	France	37	244	3
3	France	33	235	4
4	France	40	233	5
5	Germany	37	128	1
6	Germany	35	113	2
7	Germany	40	106	3
8	Germany	33	103	4
9	Germany	41	102	5
10	Spain	35	131	1
11	Spain	38	130	2
12	Spain	36	128	3
13	Spain	39	121	4
14	Spain	34	109	5

Observation

Top 5 age groups from each region with highest number of customers are identified using dense rank function. France has the highest customers in age group of 38, 34, 37. Germany has the highest customers in age group of 37, 34, 40.

In [210…

```
customer_satisfaction = bank.groupby("Satisfaction Score")["Exited"].value_counts().reset_index(name = "count")
customer_satisfaction
```

Out[210]:

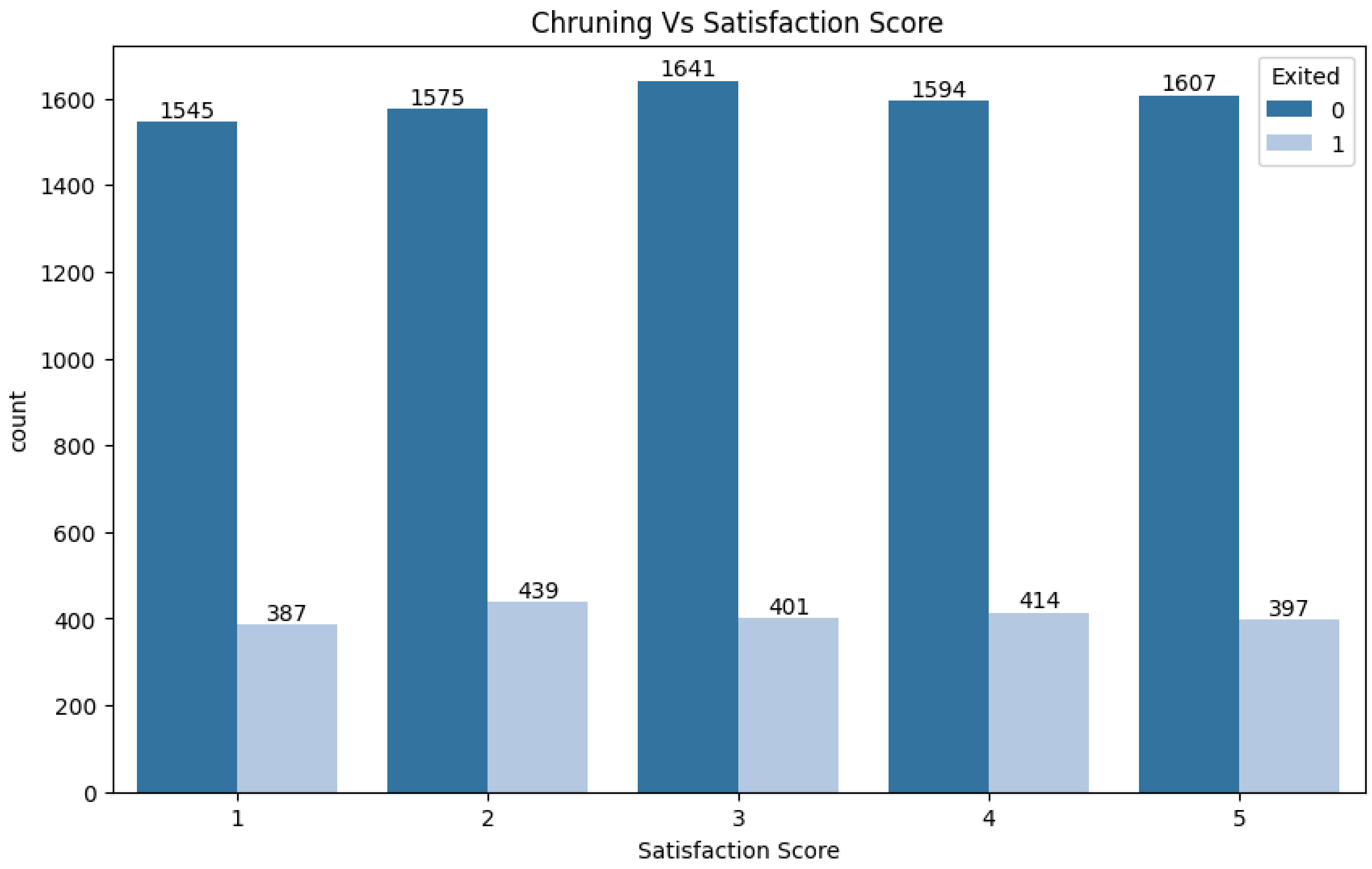
	Satisfaction Score	Exited	count
0	1	0	1545
1	1	1	387
2	2	0	1575
3	2	1	439
4	3	0	1641
5	3	1	401
6	4	0	1594
7	4	1	414
8	5	0	1607
9	5	1	397

In [211…

```
plt.figure(figsize = (10,6))
satisfaction_scores = sns.countplot(data = bank, x= "Satisfaction Score", hue= "Exited",palette="tab20")
for i in satisfaction_scores.containers:
        satisfaction_scores.bar_label(i)
plt.title("Chruning Vs Satisfaction Score")
```

Out[211]:

```
Text(0.5, 1.0, 'Chruning Vs Satisfaction Score')
```



Observation

A countplot was plotted on the satisfaction score and the number of customers. We can assume that, irrespective of the customer satisfaction score the number of customers leaving the bank were found to be similar in number.

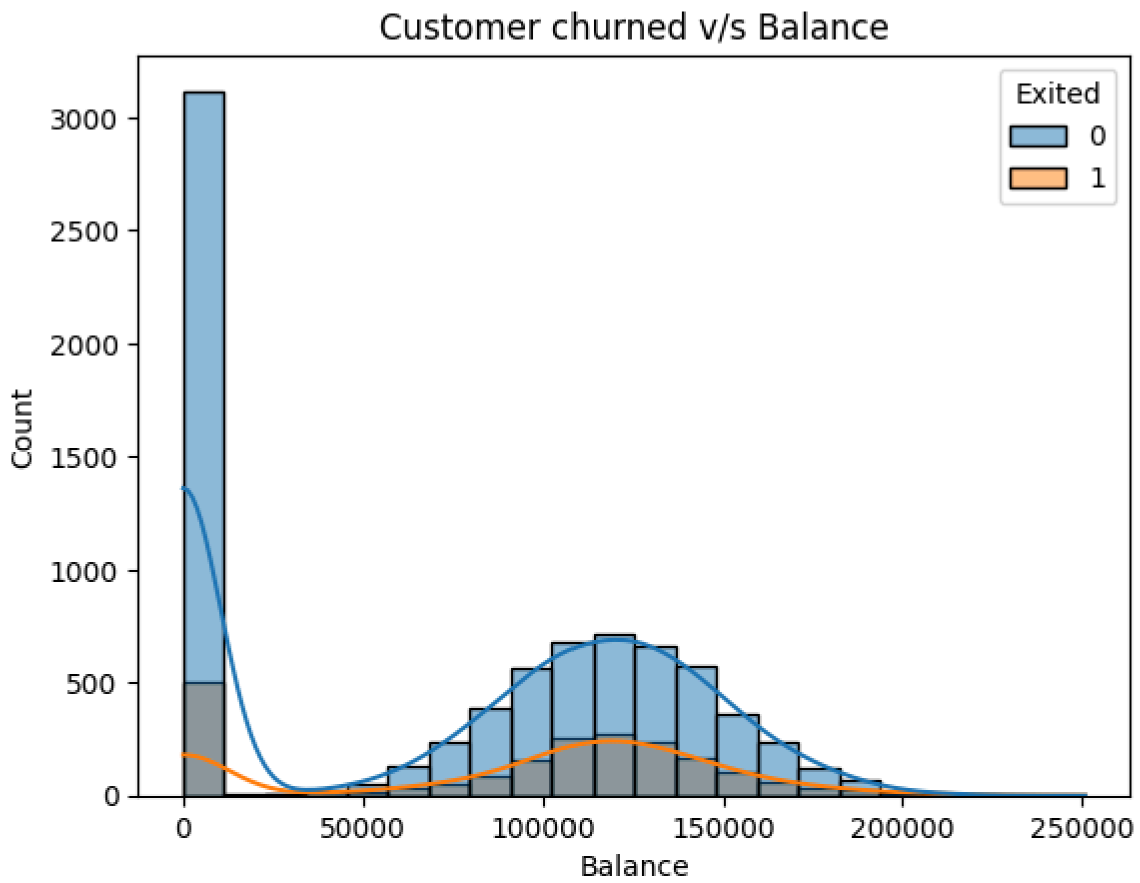
In [212…
```
bank.head()
```

Out[212]:

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited	Complain	Satisfaction Score	Card Type	Point Earned
0	1	15634602	Hargrave	619	France	Female	42	2	0.00	1	1	1	101348.88	1	1	2	DIAMOND	464
1	2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0	1	3	DIAMOND	456
2	3	15619304	Onio	502	France	Female	42	8	159660.80	3	1	0	113931.57	1	1	3	DIAMOND	377
3	4	15701354	Boni	699	France	Female	39	1	0.00	2	0	0	93826.63	0	0	5	GOLD	350
4	5	15737888	Mitchell	850	Spain	Female	43	2	125510.82	1	1	1	79084.10	0	0	5	GOLD	425

In [213…
```
sns.histplot(data = bank, x= 'Balance',hue ='Exited',kde =True)
plt.title('Customer churned v/s Balance')
```

Out[213]:
```
Text(0.5, 1.0, 'Customer churned v/s Balance')
```



In [214…
```
bank.head()
```

Out[214]:

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited	Complain	Satisfaction Score	Card Type	Point Earned
0	1	15634602	Hargrave	619	France	Female	42	2	0.00	1	1	1	101348.88	1	1	2	DIAMOND	464
1	2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0	1	3	DIAMOND	456
2	3	15619304	Onio	502	France	Female	42	8	159660.80	3	1	0	113931.57	1	1	3	DIAMOND	377
3	4	15701354	Boni	699	France	Female	39	1	0.00	2	0	0	93826.63	0	0	5	GOLD	350
4	5	15737888	Mitchell	850	Spain	Female	43	2	125510.82	1	1	1	79084.10	0	0	5	GOLD	425

In [215…

```
customer_tenure= bank.groupby("Tenure")["Exited"].value_counts().reset_index(name ="Count").sort_values("Tenure", ascending = False)
customer_tenure = customer_tenure.reset_index()
customer_tenure= customer_tenure.drop("index", axis =1)
customer_tenure
```

Out[215]:

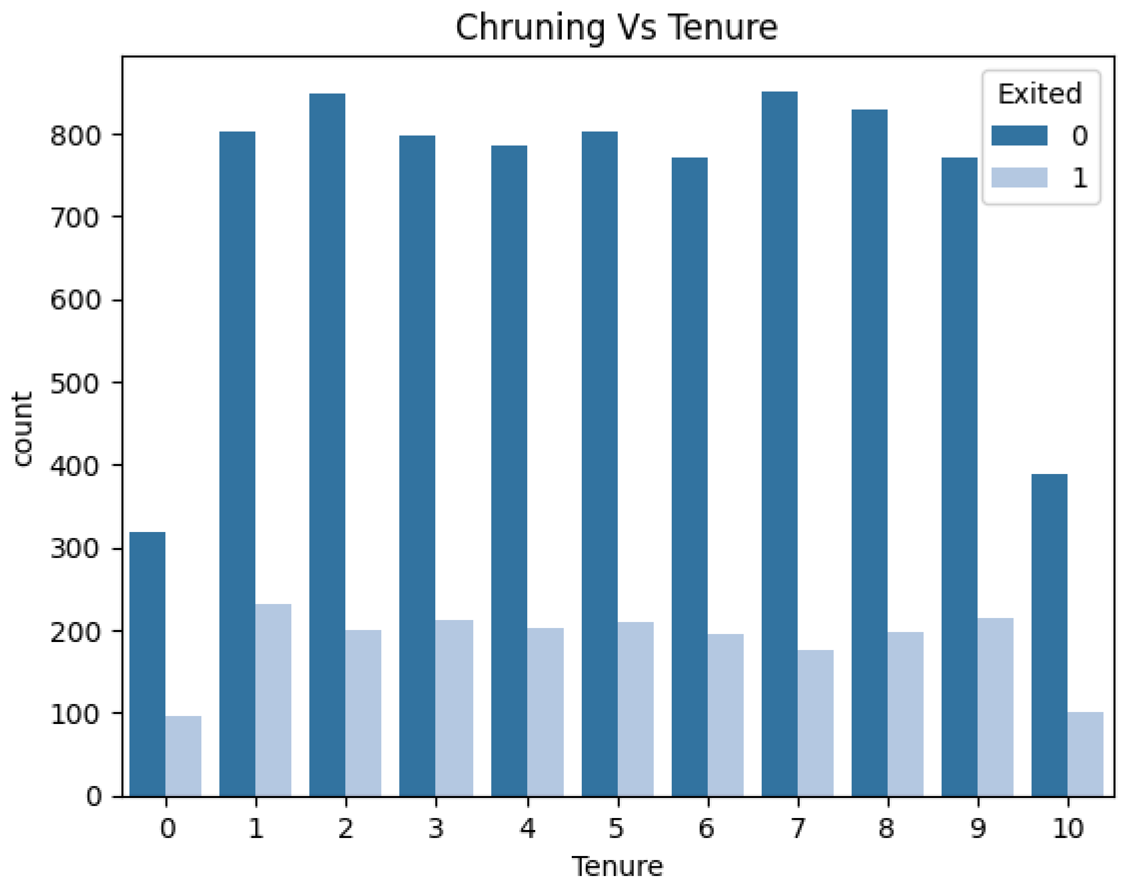
	Tenure	Exited	Count
0	10	1	101
1	10	0	389
2	9	1	214
3	9	0	770
4	8	1	197
5	8	0	828
6	7	1	177
7	7	0	851
8	6	0	771
9	6	1	196
10	5	1	209
11	5	0	803
12	4	1	203
13	4	0	786
14	3	1	213
15	3	0	796
16	2	1	201
17	2	0	847
18	1	1	232
19	1	0	803
20	0	1	95
21	0	0	318

In [216…

```
sns.countplot(data = bank, x= "Tenure", hue = "Exited", palette = 'tab20')
plt.title("Chruning Vs Tenure")
```

Out[216]:

```
Text(0.5, 1.0, 'Chruning Vs Tenure')
```



In [217…

```
customer_tenure_exited = customer_tenure[customer_tenure["Exited"]==1]
customer_tenure_exited
```

Out[217]:

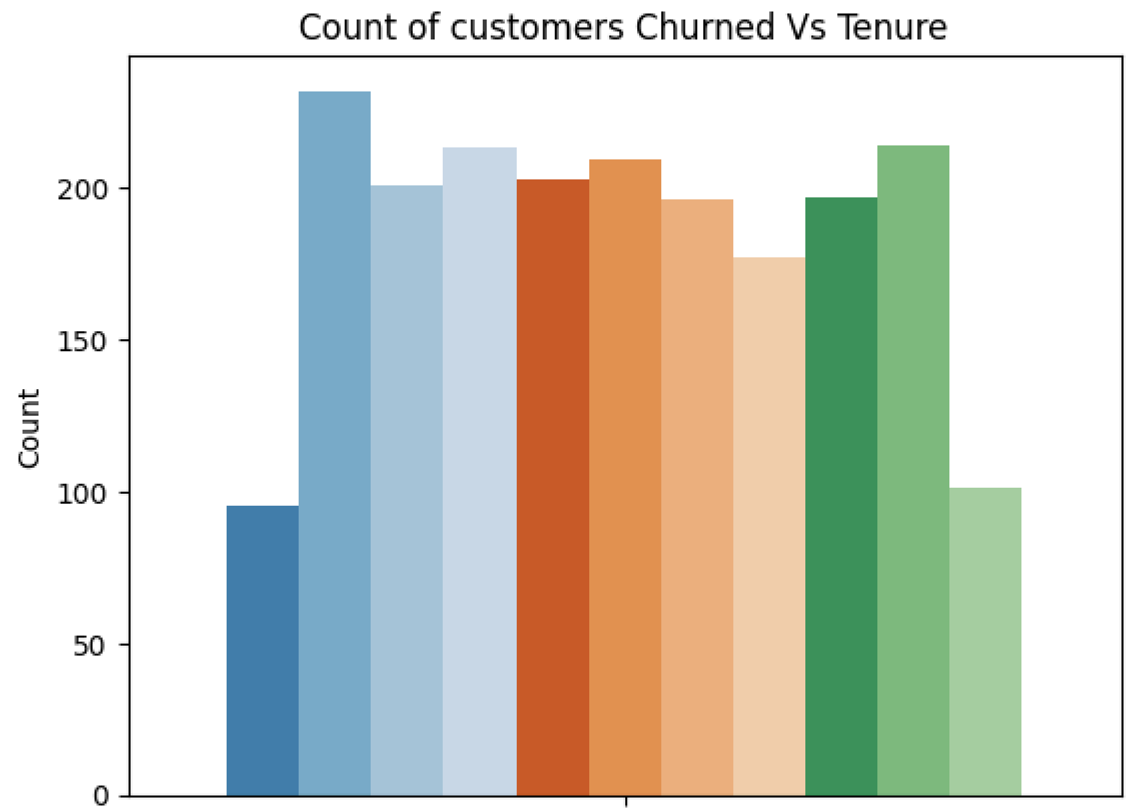
	Tenure	Exited	Count
0	10	1	101
2	9	1	214
4	8	1	197
6	7	1	177
9	6	1	196
10	5	1	209
12	4	1	203
14	3	1	213
16	2	1	201
18	1	1	232
20	0	1	95

In [218…

```
x= customer_tenure_exited["Tenure"]
y = customer_tenure_exited["Count"]
sns.barplot(data = customer_tenure_exited, hue= "Tenure", y= 'Count', palette= 'tab20c', legend = False)
plt.title("Count of customers Churned Vs Tenure")
```

Out[218]:

```
Text(0.5, 1.0, 'Count of customers Churned Vs Tenure')
```



In [219…
```
avg_mean =  bank["EstimatedSalary"].mean()
bank[(bank["EstimatedSalary"] > avg_mean) & (bank["Exited"]==1)]
```

Out[219]:

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited	Complain	Satisfaction Score	Card Type	Point Earned
0	1	15634602	Hargrave	619	France	Female	42	2	0.00	1	1	1	101348.88	1	1	2	DIAMOND	464
2	3	15619304	Onio	502	France	Female	42	8	159660.80	3	1	0	113931.57	1	1	3	DIAMOND	377
5	6	15574012	Chu	645	Spain	Male	44	8	113755.78	2	1	0	149756.71	1	1	5	DIAMOND	484
7	8	15656148	Obinna	376	Germany	Female	29	4	115046.74	4	1	0	119346.88	1	1	2	DIAMOND	282
22	23	15699309	Gerasimov	510	Spain	Female	38	4	0.00	1	1	0	118913.53	1	1	2	DIAMOND	887
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
9956	9957	15707861	Nucci	520	France	Female	46	10	85216.61	1	1	0	117369.52	1	1	1	GOLD	669
9960	9961	15681026	Lucciano	795	Germany	Female	33	9	104552.72	1	1	1	120853.83	1	1	1	SILVER	381
9975	9976	15666295	Smith	610	Germany	Male	50	1	113957.01	2	1	0	196526.55	1	1	4	SILVER	264
9978	9979	15703563	P'eng	774	France	Male	40	9	93017.47	2	1	0	191608.97	1	0	1	GOLD	354
9982	9983	15768163	Griffin	655	Germany	Female	46	7	137145.12	1	1	0	115146.40	1	1	4	GOLD	591

1043 rows × 18 columns

In [220…
```
avg_bal = bank["Balance"].mean()
bank[(bank["Balance"] > avg_bal) & (bank["Exited"]==1)]
```

Out[220]:

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited	Complain	Satisfaction Score	Card Type	Point Earned
2	3	15619304	Onio	502	France	Female	42	8	159660.80	3	1	0	113931.57	1	1	3	DIAMOND	377
5	6	15574012	Chu	645	Spain	Male	44	8	113755.78	2	1	0	149756.71	1	1	5	DIAMOND	484
7	8	15656148	Obinna	376	Germany	Female	29	4	115046.74	4	1	0	119346.88	1	1	2	DIAMOND	282
16	17	15737452	Romeo	653	Germany	Male	58	1	132602.88	1	1	0	5097.67	1	0	2	SILVER	163
35	36	15794171	Lombardo	475	France	Female	45	0	134264.04	1	1	0	27822.99	1	1	1	DIAMOND	877
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
9975	9976	15666295	Smith	610	Germany	Male	50	1	113957.01	2	1	0	196526.55	1	1	4	SILVER	264
9978	9979	15703563	P'eng	774	France	Male	40	9	93017.47	2	1	0	191608.97	1	0	1	GOLD	354
9981	9982	15672754	Burbidge	498	Germany	Male	42	3	152039.70	1	1	1	53445.17	1	1	3	GOLD	790
9982	9983	15768163	Griffin	655	Germany	Female	46	7	137145.12	1	1	0	115146.40	1	1	4	GOLD	591
9991	9992	15769959	Ajuluchukwu	597	France	Female	53	4	88381.21	1	1	0	69384.71	1	1	3	GOLD	369

1427 rows × 18 columns

In [221…

```
bank[(bank["Balance"] > avg_bal) & (bank["EstimatedSalary"] > avg_mean)& (bank["Exited"]==1)]
```

Out[221]:

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited	Complain	Satisfaction Score	Card Type	Point Earned
2	3	15619304	Onio	502	France	Female	42	8	159660.80	3	1	0	113931.57	1	1	3	DIAMOND	377
5	6	15574012	Chu	645	Spain	Male	44	8	113755.78	2	1	0	149756.71	1	1	5	DIAMOND	484
7	8	15656148	Obinna	376	Germany	Female	29	4	115046.74	4	1	0	119346.88	1	1	2	DIAMOND	282
41	42	15738148	Clarke	465	France	Female	51	8	122522.32	1	0	0	181297.65	1	1	5	SILVER	828
43	44	15755196	Lavine	834	France	Female	49	2	131394.56	1	0	0	194365.76	1	1	2	GOLD	567
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
9956	9957	15707861	Nucci	520	France	Female	46	10	85216.61	1	1	0	117369.52	1	1	1	GOLD	669
9960	9961	15681026	Lucciano	795	Germany	Female	33	9	104552.72	1	1	1	120853.83	1	1	1	SILVER	381
9975	9976	15666295	Smith	610	Germany	Male	50	1	113957.01	2	1	0	196526.55	1	1	4	SILVER	264
9978	9979	15703563	P'eng	774	France	Male	40	9	93017.47	2	1	0	191608.97	1	0	1	GOLD	354
9982	9983	15768163	Griffin	655	Germany	Female	46	7	137145.12	1	1	0	115146.40	1	1	4	GOLD	591

734 rows × 18 columns

In [222…

```
avg_credit_score = bank["CreditScore"].mean()
bank[(bank["CreditScore"] > avg_bal) & (bank["Exited"]==1)]
```

Out[222]:

RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited	Complain	Satisfaction Score	Card Type	Point Earned

Insights:

1. Number of customers in Germany is very less compared to France and Spain. Eventhough the population of Germany much greater than France the number of customers are very less.
2. Considering the size of the customer population, the number of customers who left the bank are more in Germany than the other two regions.
3. The Count of male customers are greater than female customers in all the 3 regions. Consider the individual population size of males and females, we can say customers who are leaving the bank are more in females than that of males.
4. France has the highest customers in age group of 38, 34, 37. Germany has the highest customers in age group of 37, 34, 40.
5. There are 1427 customers with a balance greater than the avg balance have left the bank. This would affect the business of the bank as most of their valuable customers are leaving the bank.
6. Nearly 734 customers who have greater than salary and avg salary and the balance greater than avg balance of customers have left the bank. These customers can be considered as the Premium customers of the bank and would a great impact in terms of the revenue.
7. There no customers with creditscore greater than the avg credit score who has left the bank.
8. Irrespective of the customer satisfaction score the number of customers leaving the bank were found to be similar in number.

Recommendation

1. Most of customers belongs to France region. So, bank has high chance of increase there customers by increasing promotions and also through developing new offers like beneficial Insurances schemes.
2. As the number of the customers in Germany are less out of all the regions, bank can focus more on marketing and special offers like giving coupons or special loan offers in Germany.
3. Nearly 30% of customers doesn't hold a credit card. Bank can provide new offers like providing cashbacks, vouchers, coupons on paying bills using credit card.
4. Majority of the customers belong to the age group of 34 to 41. Bank can provide them loans basing on their credit scores. So that they will be part of the bank for a long period of time.
5. Bank can hold back customers without leaving by introducing new plans like housing loans or personal loans with less rate of interest.
6. They can also offer special tours to their most valubale customers(customers with highest balance and credit score).
7. Customers with 600 to 700 credit score are more likely to churn. Bank can take of these customers by offering special loans because of the good credit scores.

Conclusion

An analysis was performed on the churning of customers from the bank. Analysis was performed using variables like Credit Score ,Age and Geography, Balance Tenure and found that they has no relation with customer who churned.