

Customer Churn Analysis

May 19, 2024

1 Customer Churn Analysis

This is a dataset of an ecommerce company and we have some customers who are churning (leaving).

1.1 Goals

1. Perform exploratory analysis of the provided customer data to share insights of the behavior and characteristics of the customers. Make suggestions to help the company with customer retention
2. Build a predictive model to identify customers who are at risk of leaving the company (churn) based on the provided variables. This can help the company take proactive steps to retain these customers and reduce the rate of churn

1.2 Data description

1. CustomerID
2. Churn: Churn Flag
3. Tenure: in months
4. PreferredLoginDevice
5. CityTier
6. WarehouseToHome: Distance in between warehouse to home of customer
7. PreferredPaymentMode
8. Gender
9. HourSpendOnApp
10. NumberOfDeviceRegistered
11. PreferredOrderCat
12. SatisfactionScore
13. MaritalStatus
14. NumberOfAddress
15. OrderAmountHikeFromlastYear: Percentage increases in order from last year
16. CouponUsed: Total number of coupon has been used in last month

17. OrderCount: Total number of orders has been places in last month
18. DaySinceLastOrder
19. CashbackAmount: Average cashback in last month
20. Complain: Complain flag - if the customer ever had a complain

1.3 Tips

1. Sharing your thoughts and reasoning as you go will help!
2. Feel free to use any libraries like scikit-learn, use stackoverflow. Don't use ChatGPT and similar.
3. If you are unable to complete any step I can provide help towards the answer. Demonstrating understanding of the solution and result will earn some points!

1.4 Import the libraries

```
[1]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

2 1. Data overview

Question: Read the sheet named 'E Comm' from file 'E Commerce Dataset.xlsx' saved in current directory into df variable. Print first 5 rows of the dataframe

```
[2]: df = pd.read_excel('E Commerce Dataset.xlsx')
df.head(5)
```

```
[2]:
```

	CustomerID	Churn	Tenure	PreferredLoginDevice	CityTier	WarehouseToHome	\
0	50001	1	4.0	Mobile Phone	3	6.0	
1	50002	1	NaN	Phone	1	8.0	
2	50003	1	NaN	Phone	1	30.0	
3	50004	1	0.0	Phone	3	15.0	
4	50005	1	0.0	Phone	1	12.0	

	PreferredPaymentMode	Gender	HourSpendOnApp	NumberOfDeviceRegistered	\
0	Debit Card	Female	3.0	3	
1	UPI	Male	3.0	4	
2	Debit Card	Male	2.0	4	
3	Debit Card	Male	2.0	4	
4	CC	Male	NaN	3	

	PreferredOrderCat	SatisfactionScore	MaritalStatus	NumberOfAddress	\
0	Laptop & Accessory	2	Single	9	
1	Mobile	3	Single	7	
2	Mobile	3	Single	6	
3	Laptop & Accessory	5	Single	8	

	4	Mobile	5	Single	3
		Complain	OrderAmountHikeFromlastYear	CouponUsed	OrderCount \
0		1	11.0	1.0	1.0
1		1	15.0	0.0	1.0
2		1	14.0	0.0	1.0
3		0	23.0	0.0	1.0
4		0	11.0	1.0	1.0

		DaySinceLastOrder	CashbackAmount
0		5.0	159.93
1		0.0	120.90
2		3.0	120.28
3		3.0	134.07
4		3.0	129.60

```
[3]: df.shape
```

```
[3]: (5630, 20)
```

```
[4]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5630 entries, 0 to 5629
Data columns (total 20 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   CustomerID                            5630 non-null   int64
1   Churn                                  5630 non-null   int64
2   Tenure                                 5366 non-null   float64
3   PreferredLoginDevice                  5630 non-null   object
4   CityTier                              5630 non-null   int64
5   WarehouseToHome                       5379 non-null   float64
6   PreferredPaymentMode                  5630 non-null   object
7   Gender                                5630 non-null   object
8   HourSpendOnApp                        5375 non-null   float64
9   NumberOfDeviceRegistered              5630 non-null   int64
10  PreferedOrderCat                      5630 non-null   object
11  SatisfactionScore                     5630 non-null   int64
12  MaritalStatus                         5630 non-null   object
13  NumberOfAddress                       5630 non-null   int64
14  Complain                              5630 non-null   int64
15  OrderAmountHikeFromlastYear           5365 non-null   float64
16  CouponUsed                            5374 non-null   float64
17  OrderCount                            5372 non-null   float64
18  DaySinceLastOrder                     5323 non-null   float64
19  CashbackAmount                        5630 non-null   float64
```

```
dtypes: float64(8), int64(7), object(5)
memory usage: 879.8+ KB
```

Question: How many unique values are in each column?

```
[5]: for columns in df:
      print(columns,df[columns].nunique())
```

```
CustomerID 5630
Churn 2
Tenure 36
PreferredLoginDevice 3
CityTier 3
WarehouseToHome 34
PreferredPaymentMode 7
Gender 2
HourSpendOnApp 6
NumberOfDeviceRegistered 6
PreferedOrderCat 6
SatisfactionScore 5
MaritalStatus 3
NumberOfAddress 15
Complain 2
OrderAmountHikeFromlastYear 16
CouponUsed 17
OrderCount 16
DaySinceLastOrder 22
CashbackAmount 2586
```

Question: Calculate average churn rate Note: Churn = 1 means customer has churned

```
[6]: df[df['Churn'] == 1].count()
```

```
[6]: CustomerID          948
      Churn              948
      Tenure            867
      PreferredLoginDevice 948
      CityTier          948
      WarehouseToHome    864
      PreferredPaymentMode 948
      Gender            948
      HourSpendOnApp     890
      NumberOfDeviceRegistered 948
      PreferedOrderCat   948
      SatisfactionScore  948
      MaritalStatus      948
      NumberOfAddress    948
      Complain           948
```

```
OrderAmountHikeFromlastYear    934
CouponUsed                      940
OrderCount                     930
DaySinceLastOrder              894
CashbackAmount                 948
dtype: int64
```

```
[7]: #check if there is any n/a
df['Churn'].isna().nunique()
```

```
[7]: 1
```

```
[8]: #churn rate
df[df['Churn'] == 1].shape[0]/df.shape[0]*100
```

```
[8]: 16.838365896980463
```

Question: How many missing values / nulls are there in each column?

```
[9]: for col in df.columns:
      print(col, df[col].isna().sum())
```

```
CustomerID 0
Churn 0
Tenure 264
PreferredLoginDevice 0
CityTier 0
WarehouseToHome 251
PreferredPaymentMode 0
Gender 0
HourSpendOnApp 255
NumberOfDeviceRegistered 0
PreferedOrderCat 0
SatisfactionScore 0
MaritalStatus 0
NumberOfAddress 0
Complain 0
OrderAmountHikeFromlastYear 265
CouponUsed 256
OrderCount 258
DaySinceLastOrder 307
CashbackAmount 0
```

3 2. Exploratory Data Analysis

3.1 Univariate analysis

Here we will understand select variables

3.1.1 1. Numeric variables

Question: Histogram Show histograms for all numeric columns. Describe the each variable's distribution briefly to a business stakeholder

```
[10]: numerical_df =  
↳ df[['CustomerID', 'Churn', 'Tenure', 'CityTier', 'WarehouseToHome', 'HourSpendOnApp', 'NumberOfDe',  
↳ 'OrderAmountHikeFromlastYear', 'CouponUsed', 'OrderCount', 'DaySinceLa',  
numerical_df.head(5)
```

```
[10]: CustomerID  Churn  Tenure  CityTier  WarehouseToHome  HourSpendOnApp  \  
0      50001      1    4.0        3           6.0           3.0  
1      50002      1    NaN        1           8.0           3.0  
2      50003      1    NaN        1          30.0           2.0  
3      50004      1    0.0        3          15.0           2.0  
4      50005      1    0.0        1          12.0           NaN  
  
    NumberOfDeviceRegistered  SatisfactionScore  NumberOfAddress  Complain  \  
0                3                2                9            1  
1                4                3                7            1  
2                4                3                6            1  
3                4                5                8            0  
4                3                5                3            0  
  
    OrderAmountHikeFromlastYear  CouponUsed  OrderCount  DaySinceLastOrder  \  
0                11.0            1.0            1.0            5.0  
1                15.0            0.0            1.0            0.0  
2                14.0            0.0            1.0            3.0  
3                23.0            0.0            1.0            3.0  
4                11.0            1.0            1.0            3.0  
  
    CashbackAmount  
0      159.93  
1      120.90  
2      120.28  
3      134.07  
4      129.60
```

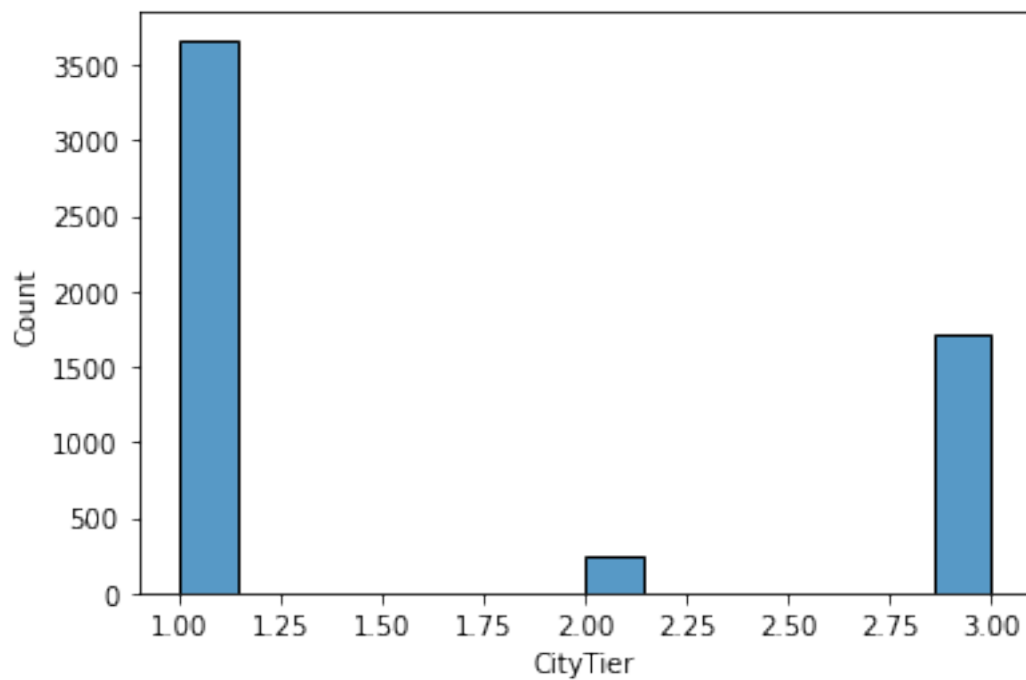
```
[11]: categorical_df =  
↳ df[['PreferredLoginDevice', 'PreferredPaymentMode', 'Gender', 'PreferedOrderCat', 'MaritalStatu',  
categorical_df.head(5)
```

```
[11]: PreferredLoginDevice PreferredPaymentMode  Gender  PreferedOrderCat  \  
0      Mobile Phone      Debit Card  Female  Laptop & Accessory  
1              Phone      UPI      Male      Mobile  
2              Phone      Debit Card  Male      Mobile  
3              Phone      Debit Card  Male  Laptop & Accessory  
4              Phone      CC      Male      Mobile
```

```
MaritalStatus
0      Single
1      Single
2      Single
3      Single
4      Single
```

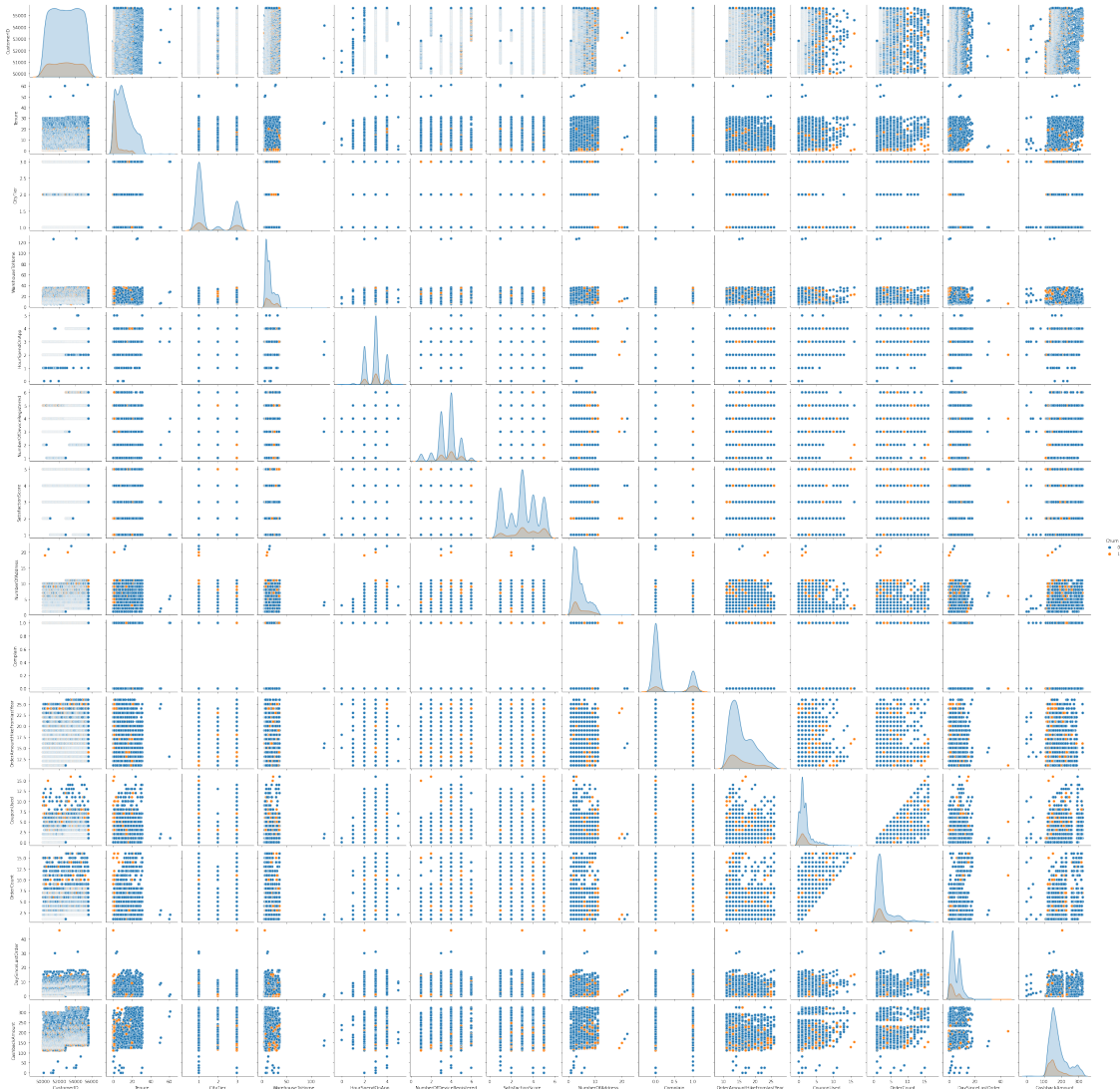
```
[12]: sns.histplot(data=numerical_df, x="CityTier")
```

```
[12]: <AxesSubplot:xlabel='CityTier', ylabel='Count'>
```



```
[13]: sns.pairplot(numerical_df, hue="Churn")
```

```
[13]: <seaborn.axisgrid.PairGrid at 0x7fd73b7e36a0>
```



3.1.2 2. Non numeric columns

Cleaning (ignore)

```
[14]: #As mobile phone and phone are both same so we have merged them
df.loc[df['PreferredLoginDevice'] == 'Phone', 'PreferredLoginDevice'] = 'Mobile Phone'
df.loc[df['PreferredOrderCat'] == 'Mobile', 'PreferredOrderCat'] = 'Mobile Phone'
#as cod is also cash on delivery
#as cc is also credit card so i merged them
df.loc[df['PreferredPaymentMode'] == 'COD', 'PreferredPaymentMode'] = 'Cash on Delivery'
df.loc[df['PreferredPaymentMode'] == 'CC', 'PreferredPaymentMode'] = 'Credit Card'
```


Question: Show the unique values of each variable and print the number of occurrences of each value Describe these results briefly

```
[15]: for col in df.columns:
      if df[col].dtype == object:
          print(df[col].unique(), df[col].value_counts())

['Mobile Phone' 'Computer'] Mobile Phone    3996
Computer    1634
Name: PreferredLoginDevice, dtype: int64
['Debit Card' 'UPI' 'Credit Card' 'Cash on Delivery' 'E wallet'] Debit Card
2314
Credit Card    1774
E wallet    614
Cash on Delivery    514
UPI    414
Name: PreferredPaymentMode, dtype: int64
['Female' 'Male'] Male    3384
Female    2246
Name: Gender, dtype: int64
['Laptop & Accessory' 'Mobile Phone' 'Others' 'Fashion' 'Grocery'] Mobile Phone
2080
Laptop & Accessory    2050
Fashion    826
Grocery    410
Others    264
Name: PreferredOrderCat, dtype: int64
['Single' 'Divorced' 'Married'] Married    2986
Single    1796
Divorced    848
Name: MaritalStatus, dtype: int64
```

3.2 Analysing the Churn by select variables

Provide business recommendation for each of the below

Question: Relation between complains and churn

```
[16]: df['Churn'].corr(df['Complain'])

[16]: 0.25018825469703104

[20]: churn_rate_complain = df[df['Complain'] == 1]['Churn'].mean()
      churn_rate_no_complain = df[df['Complain'] == 0]['Churn'].mean()

[21]: print("Churn rate for Complain =", churn_rate_complain)
      print("Churn rate for No Complain =", churn_rate_no_complain)
```

Churn rate for Complain = 0.3167082294264339
 Churn rate for No Complain = 0.1092896174863388

3.3 Correlation matrix

Visualize the correlation between all variables

Question: Do we need to do any preprocessing on categorical variables before calculating correlation?

```
[22]: #One hot coding for categorical dataset
from sklearn.preprocessing import OneHotEncoder
enc = OneHotEncoder()
new_df = enc.fit_transform(categorical_df)
df_ohc = pd.DataFrame(new_df.toarray(), columns=enc.get_feature_names_out(),
dtype=int)
```

```
[23]: df_ohc.head(5)
```

```
[23]: PreferredLoginDevice_Computer PreferredLoginDevice_Mobile Phone \
0 0 1
1 0 0
2 0 0
3 0 0
4 0 0

PreferredLoginDevice_Phone PreferredPaymentMode_CC \
0 0 0
1 1 0
2 1 0
3 1 0
4 1 1

PreferredPaymentMode_COD PreferredPaymentMode_Cash on Delivery \
0 0 0
1 0 0
2 0 0
3 0 0
4 0 0

PreferredPaymentMode_Credit Card PreferredPaymentMode_Debit Card \
0 0 1
1 0 0
2 0 1
3 0 1
4 0 0

PreferredPaymentMode_E wallet PreferredPaymentMode_UPI ... Gender_Male \
```

0	0	0 ...	0
1	0	1 ...	1
2	0	0 ...	1
3	0	0 ...	1
4	0	0 ...	1

	PreferedOrderCat_Fashion	PreferedOrderCat_Grocery \
0	0	0
1	0	0
2	0	0
3	0	0
4	0	0

	PreferedOrderCat_Laptop & Accessory	PreferedOrderCat_Mobile \
0	1	0
1	0	1
2	0	1
3	1	0
4	0	1

	PreferedOrderCat_Mobile Phone	PreferedOrderCat_Others \
0	0	0
1	0	0
2	0	0
3	0	0
4	0	0

	MaritalStatus_Divorced	MaritalStatus_Married	MaritalStatus_Single
0	0	0	1
1	0	0	1
2	0	0	1
3	0	0	1
4	0	0	1

[5 rows x 21 columns]

```
[24]: df_final = pd.concat([numerical_df, df_ohc], axis=1)
df_final.head(5)
```

```
[24]: CustomerID  Churn  Tenure  CityTier  WarehouseToHome  HourSpendOnApp \
0      50001      1    4.0      3           6.0           3.0
1      50002      1   NaN      1           8.0           3.0
2      50003      1   NaN      1          30.0           2.0
3      50004      1    0.0      3          15.0           2.0
4      50005      1    0.0      1          12.0           NaN

NumberOfDeviceRegistered  SatisfactionScore  NumberOfAddress  Complain \
```

0		3		2		9		1
1		4		3		7		1
2		4		3		6		1
3		4		5		8		0
4		3		5		3		0

	...	Gender_Male	PreferedOrderCat_Fashion	PreferedOrderCat_Grocery	\
0	...	0	0	0	
1	...	1	0	0	
2	...	1	0	0	
3	...	1	0	0	
4	...	1	0	0	

		PreferedOrderCat_Laptop & Accessory	PreferedOrderCat_Mobile	\
0		1	0	
1		0	1	
2		0	1	
3		1	0	
4		0	1	

		PreferedOrderCat_Mobile Phone	PreferedOrderCat_Others	\
0		0	0	
1		0	0	
2		0	0	
3		0	0	
4		0	0	

		MaritalStatus_Divorced	MaritalStatus_Married	MaritalStatus_Single
0		0	0	1
1		0	0	1
2		0	0	1
3		0	0	1
4		0	0	1

[5 rows x 36 columns]

```
[25]: df_final = df_final.dropna()
df_final.head(5)
```

```
[25]: CustomerID  Churn  Tenure  CityTier  WarehouseToHome  HourSpendOnApp  \
0      50001      1    4.0      3      6.0      3.0
3      50004      1    0.0      3     15.0      2.0
5      50006      1    0.0      1     22.0      3.0
11     50012      1   11.0      1      6.0      3.0
12     50013      1    0.0      1     11.0      2.0
```

		NumberOfDeviceRegistered	SatisfactionScore	NumberOfAddress	Complain	\
--	--	--------------------------	-------------------	-----------------	----------	---

0	3	2	9	1
3	4	5	8	0
5	5	5	2	1
11	4	3	10	1
12	3	3	2	1

	Gender_Male	PreferedOrderCat_Fashion	PreferedOrderCat_Grocery	\
0	0	0	0	
3	1	0	0	
5	0	0	0	
11	1	1	0	
12	1	0	0	

	PreferedOrderCat_Laptop & Accessory	PreferedOrderCat_Mobile	\
0	1	0	
3	1	0	
5	0	0	
11	0	0	
12	0	1	

	PreferedOrderCat_Mobile Phone	PreferedOrderCat_Others	\
0	0	0	
3	0	0	
5	1	0	
11	0	0	
12	0	0	

	MaritalStatus_Divorced	MaritalStatus_Married	MaritalStatus_Single
0	0	0	1
3	0	0	1
5	0	0	1
11	0	0	1
12	0	0	1

[5 rows x 36 columns]

Question: Plot correlation matrix Discuss a few significant correlations

```
[26]: #CustomerID column as it's not relevant for correlation
df_final = df_final.drop(columns=['CustomerID'])
df_final.head(5)
```

	Churn	Tenure	CityTier	WarehouseToHome	HourSpendOnApp	\
0	1	4.0	3	6.0	3.0	
3	1	0.0	3	15.0	2.0	
5	1	0.0	1	22.0	3.0	
11	1	11.0	1	6.0	3.0	

12	1	0.0	1	11.0	2.0
----	---	-----	---	------	-----

	NumberOfDeviceRegistered	SatisfactionScore	NumberOfAddress	Complain	\
0	3	2	9	1	
3	4	5	8	0	
5	5	5	2	1	
11	4	3	10	1	
12	3	3	2	1	

	OrderAmountHikeFromlastYear	...	Gender_Male	PreferedOrderCat_Fashion	\
0	11.0	...	0	0	
3	23.0	...	1	0	
5	22.0	...	0	0	
11	13.0	...	1	1	
12	13.0	...	1	0	

	PreferedOrderCat_Grocery	PreferedOrderCat_Laptop & Accessory	\
0	0	1	
3	0	1	
5	0	0	
11	0	0	
12	0	0	

	PreferedOrderCat_Mobile	PreferedOrderCat_Mobile Phone	\
0	0	0	
3	0	0	
5	0	1	
11	0	0	
12	1	0	

	PreferedOrderCat_Others	MaritalStatus_Divorced	MaritalStatus_Married	\
0	0	0	0	
3	0	0	0	
5	0	0	0	
11	0	0	0	
12	0	0	0	

	MaritalStatus_Single
0	1
3	1
5	1
11	1
12	1

[5 rows x 35 columns]

```
[28]: corr_matrix = df_final.corr()
      corr_matrix
```

```
[28]:
```

	Churn	Tenure	CityTier	\
Churn	1.000000	-0.340013	0.073858	
Tenure	-0.340013	1.000000	-0.057414	
CityTier	0.073858	-0.057414	1.000000	
WarehouseToHome	0.087318	-0.016353	0.004457	
HourSpendOnApp	0.060845	-0.029818	-0.070035	
NumberOfDeviceRegistered	0.149041	-0.058752	-0.008616	
SatisfactionScore	0.095759	-0.009972	-0.019494	
NumberOfAddress	0.076336	0.196547	-0.046406	
Complain	0.238137	-0.035228	-0.006122	
OrderAmountHikeFromlastYear	0.017193	0.012197	-0.027628	
CouponUsed	0.010982	0.075959	0.013313	
OrderCount	-0.001962	0.112339	0.035018	
DaySinceLastOrder	-0.139254	0.118932	0.014304	
CashbackAmount	-0.058756	0.213852	0.146569	
PreferredLoginDevice_Computer	0.040890	-0.010308	-0.000184	
PreferredLoginDevice_Mobile Phone	-0.079120	0.085754	0.098949	
PreferredLoginDevice_Phone	0.053023	-0.096773	-0.125199	
PreferredPaymentMode_CC	0.030729	-0.032474	-0.073109	
PreferredPaymentMode_COD	0.059430	-0.027497	-0.049111	
PreferredPaymentMode_Cash on Delivery	-0.006498	0.069208	-0.004977	
PreferredPaymentMode_Credit Card	-0.029392	-0.024243	-0.150637	
PreferredPaymentMode_Debit Card	-0.039232	0.040958	-0.127528	
PreferredPaymentMode_E wallet	0.043972	0.016709	0.503165	
PreferredPaymentMode_UPI	0.000938	-0.043942	-0.036088	
Gender_Female	-0.033792	0.052123	0.018072	
Gender_Male	0.033792	-0.052123	-0.018072	
PreferedOrderCat_Fashion	0.013086	0.130259	0.083035	
PreferedOrderCat_Grocery	-0.017880	0.091251	-0.001750	
PreferedOrderCat_Laptop & Accessory	-0.184584	0.104482	0.245824	
PreferedOrderCat_Mobile	0.016675	-0.033466	-0.097493	
PreferedOrderCat_Mobile Phone	0.181683	-0.211477	-0.281079	
PreferedOrderCat_Others	0.010208	0.103702	-0.011248	
MaritalStatus_Divorced	-0.013023	0.015788	0.019181	
MaritalStatus_Married	-0.159808	0.085990	-0.029876	
MaritalStatus_Single	0.179481	-0.103150	0.017370	

	WarehouseToHome	HourSpendOnApp	\
Churn	0.087318	0.060845	
Tenure	-0.016353	-0.029818	
CityTier	0.004457	-0.070035	
WarehouseToHome	1.000000	0.052731	
HourSpendOnApp	0.052731	1.000000	
NumberOfDeviceRegistered	0.024582	0.293021	

SatisfactionScore	0.000434	0.039879
NumberOfAddress	0.003422	0.124962
Complain	0.003829	0.020413
OrderAmountHikeFromlastYear	0.031975	0.096827
CouponUsed	-0.014311	0.160239
OrderCount	0.008800	0.092601
DaySinceLastOrder	0.010646	0.019526
CashbackAmount	-0.012433	0.131281
PreferredLoginDevice_Computer	0.022101	-0.004911
PreferredLoginDevice_Mobile Phone	-0.002403	0.106765
PreferredLoginDevice_Phone	-0.022497	-0.129642
PreferredPaymentMode_CC	0.007388	-0.054905
PreferredPaymentMode_COD	-0.004197	-0.004520
PreferredPaymentMode_Cash on Delivery	-0.007623	0.016056
PreferredPaymentMode_Credit Card	0.036141	0.033820
PreferredPaymentMode_Debit Card	-0.041475	0.020115
PreferredPaymentMode_E wallet	0.033578	-0.039522
PreferredPaymentMode_UPI	-0.021470	-0.030052
Gender_Female	-0.035581	0.009952
Gender_Male	0.035581	-0.009952
PreferedOrderCat_Fashion	-0.009435	-0.147830
PreferedOrderCat_Grocery	0.003512	-0.017382
PreferedOrderCat_Laptop & Accessory	0.043231	-0.130819
PreferedOrderCat_Mobile	-0.015441	-0.112896
PreferedOrderCat_Mobile Phone	-0.028939	0.281309
PreferedOrderCat_Others	-0.035530	0.023109
MaritalStatus_Divorced	-0.013650	0.018024
MaritalStatus_Married	0.038336	0.022257
MaritalStatus_Single	-0.030496	-0.037134

	NumberOfDeviceRegistered \
Churn	0.149041
Tenure	-0.058752
CityTier	-0.008616
WarehouseToHome	0.024582
HourSpendOnApp	0.293021
NumberOfDeviceRegistered	1.000000
SatisfactionScore	-0.017788
NumberOfAddress	0.066684
Complain	0.018881
OrderAmountHikeFromlastYear	0.083342
CouponUsed	0.114066
OrderCount	0.069559
DaySinceLastOrder	-0.049817
CashbackAmount	0.113504
PreferredLoginDevice_Computer	0.025628
PreferredLoginDevice_Mobile Phone	0.042948

PreferredLoginDevice_Phone	-0.084051
PreferredPaymentMode_CC	-0.036118
PreferredPaymentMode_COD	0.002875
PreferredPaymentMode_Cash on Delivery	-0.011988
PreferredPaymentMode_Credit Card	0.000782
PreferredPaymentMode_Debit Card	0.007461
PreferredPaymentMode_E wallet	-0.008889
PreferredPaymentMode_UPI	0.010839
Gender_Female	0.026706
Gender_Male	-0.026706
PreferedOrderCat_Fashion	-0.091604
PreferedOrderCat_Grocery	0.022548
PreferedOrderCat_Laptop & Accessory	-0.080608
PreferedOrderCat_Mobile	-0.125343
PreferedOrderCat_Mobile Phone	0.188563
PreferedOrderCat_Others	0.035364
MaritalStatus_Divorced	-0.001099
MaritalStatus_Married	-0.064517
MaritalStatus_Single	0.069345

	SatisfactionScore	NumberOfAddress \
Churn	0.095759	0.076336
Tenure	-0.009972	0.196547
CityTier	-0.019494	-0.046406
WarehouseToHome	0.000434	0.003422
HourSpendOnApp	0.039879	0.124962
NumberOfDeviceRegistered	-0.017788	0.066684
SatisfactionScore	1.000000	0.054627
NumberOfAddress	0.054627	1.000000
Complain	-0.044533	-0.017363
OrderAmountHikeFromlastYear	-0.008143	0.010275
CouponUsed	0.006687	-0.017851
OrderCount	-0.004622	-0.079413
DaySinceLastOrder	0.012950	-0.148784
CashbackAmount	0.012061	0.096752
PreferredLoginDevice_Computer	-0.035021	-0.024398
PreferredLoginDevice_Mobile Phone	-0.007731	0.078405
PreferredLoginDevice_Phone	0.050272	-0.071176
PreferredPaymentMode_CC	0.013975	0.018801
PreferredPaymentMode_COD	0.011974	-0.009988
PreferredPaymentMode_Cash on Delivery	0.005600	-0.004053
PreferredPaymentMode_Credit Card	0.043184	0.003998
PreferredPaymentMode_Debit Card	-0.073269	0.013272
PreferredPaymentMode_E wallet	0.026658	-0.027925
PreferredPaymentMode_UPI	0.009321	0.007556
Gender_Female	0.032398	0.030980
Gender_Male	-0.032398	-0.030980

PreferedOrderCat_Fashion	0.016587	-0.040136
PreferedOrderCat_Grocery	0.027097	-0.005968
PreferedOrderCat_Laptop & Accessory	-0.025448	0.059515
PreferedOrderCat_Mobile	-0.006238	-0.031229
PreferedOrderCat_Mobile Phone	0.011297	-0.023102
PreferedOrderCat_Others	0.030401	-0.004349
MaritalStatus_Divorced	0.051528	-0.024802
MaritalStatus_Married	-0.012167	-0.001463
MaritalStatus_Single	-0.025659	0.020125

	Complain	OrderAmountHikeFromlastYear \
Churn	0.238137	0.017193
Tenure	-0.035228	0.012197
CityTier	-0.006122	-0.027628
WarehouseToHome	0.003829	0.031975
HourSpendOnApp	0.020413	0.096827
NumberOfDeviceRegistered	0.018881	0.083342
SatisfactionScore	-0.044533	-0.008143
NumberOfAddress	-0.017363	0.010275
Complain	1.000000	0.003260
OrderAmountHikeFromlastYear	0.003260	1.000000
CouponUsed	0.005472	0.048959
OrderCount	-0.001917	0.042950
DaySinceLastOrder	-0.056315	0.017716
CashbackAmount	0.010389	0.014215
PreferredLoginDevice_Computer	-0.014878	-0.000220
PreferredLoginDevice_Mobile Phone	0.012571	0.004871
PreferredLoginDevice_Phone	0.001261	-0.005919
PreferredPaymentMode_CC	0.019183	-0.028536
PreferredPaymentMode_COD	-0.017256	-0.059288
PreferredPaymentMode_Cash on Delivery	-0.013373	0.000700
PreferredPaymentMode_Credit Card	0.003623	0.036106
PreferredPaymentMode_Debit Card	-0.002413	0.010242
PreferredPaymentMode_E wallet	0.001807	-0.036139
PreferredPaymentMode_UPI	0.012421	0.033331
Gender_Female	0.039522	-0.010081
Gender_Male	-0.039522	0.010081
PreferedOrderCat_Fashion	-0.000022	-0.051335
PreferedOrderCat_Grocery	0.004534	0.041476
PreferedOrderCat_Laptop & Accessory	-0.018126	-0.024964
PreferedOrderCat_Mobile	-0.005327	-0.067089
PreferedOrderCat_Mobile Phone	0.019793	0.080301
PreferedOrderCat_Others	0.007866	0.021090
MaritalStatus_Divorced	0.012777	-0.014517
MaritalStatus_Married	-0.000364	0.014894
MaritalStatus_Single	-0.009180	-0.004949

	... Gender_Male \
Churn	... 0.033792
Tenure	... -0.052123
CityTier	... -0.018072
WarehouseToHome	... 0.035581
HourSpendOnApp	... -0.009952
NumberOfDeviceRegistered	... -0.026706
SatisfactionScore	... -0.032398
NumberOfAddress	... -0.030980
Complain	... -0.039522
OrderAmountHikeFromlastYear	... 0.010081
CouponUsed	... -0.041493
OrderCount	... -0.035418
DaySinceLastOrder	... -0.018478
CashbackAmount	... -0.020263
PreferredLoginDevice_Computer	... 0.025482
PreferredLoginDevice_Mobile Phone	... -0.037886
PreferredLoginDevice_Phone	... 0.018568
PreferredPaymentMode_CC	... 0.010948
PreferredPaymentMode_COD	... 0.027718
PreferredPaymentMode_Cash on Delivery	... -0.009100
PreferredPaymentMode_Credit Card	... -0.028839
PreferredPaymentMode_Debit Card	... 0.004968
PreferredPaymentMode_E wallet	... -0.022827
PreferredPaymentMode_UPI	... 0.039949
Gender_Female	... -1.000000
Gender_Male	... 1.000000
PreferedOrderCat_Fashion	... -0.009374
PreferedOrderCat_Grocery	... 0.005292
PreferedOrderCat_Laptop & Accessory	... -0.043367
PreferedOrderCat_Mobile	... -0.001884
PreferedOrderCat_Mobile Phone	... 0.053910
PreferedOrderCat_Others	... -0.006532
MaritalStatus_Divorced	... -0.012541
MaritalStatus_Married	... 0.035047
MaritalStatus_Single	... -0.027833
	PreferedOrderCat_Fashion \
Churn	0.013086
Tenure	0.130259
CityTier	0.083035
WarehouseToHome	-0.009435
HourSpendOnApp	-0.147830
NumberOfDeviceRegistered	-0.091604
SatisfactionScore	0.016587
NumberOfAddress	-0.040136
Complain	-0.000022

OrderAmountHikeFromlastYear	-0.051335
CouponUsed	0.078762
OrderCount	0.154599
DaySinceLastOrder	0.157977
CashbackAmount	0.455466
PreferredLoginDevice_Computer	-0.054930
PreferredLoginDevice_Mobile Phone	0.188988
PreferredLoginDevice_Phone	-0.176046
PreferredPaymentMode_CC	-0.035283
PreferredPaymentMode_COD	-0.064820
PreferredPaymentMode_Cash on Delivery	0.193691
PreferredPaymentMode_Credit Card	-0.039491
PreferredPaymentMode_Debit Card	-0.004245
PreferredPaymentMode_E wallet	0.074168
PreferredPaymentMode_UPI	-0.016992
Gender_Female	0.009374
Gender_Male	-0.009374
PreferedOrderCat_Fashion	1.000000
PreferedOrderCat_Grocery	-0.014552
PreferedOrderCat_Laptop & Accessory	-0.379275
PreferedOrderCat_Mobile	-0.065803
PreferedOrderCat_Mobile Phone	-0.253118
PreferedOrderCat_Others	-0.025246
MaritalStatus_Divorced	0.004190
MaritalStatus_Married	0.010466
MaritalStatus_Single	-0.014253

	PreferedOrderCat_Grocery \
Churn	-0.017880
Tenure	0.091251
CityTier	-0.001750
WarehouseToHome	0.003512
HourSpendOnApp	-0.017382
NumberOfDeviceRegistered	0.022548
SatisfactionScore	0.027097
NumberOfAddress	-0.005968
Complain	0.004534
OrderAmountHikeFromlastYear	0.041476
CouponUsed	0.078182
OrderCount	0.126821
DaySinceLastOrder	0.035204
CashbackAmount	0.014736
PreferredLoginDevice_Computer	0.032595
PreferredLoginDevice_Mobile Phone	-0.014342
PreferredLoginDevice_Phone	-0.019492
PreferredPaymentMode_CC	-0.003861
PreferredPaymentMode_COD	-0.011748

PreferredPaymentMode_Cash on Delivery	0.054822
PreferredPaymentMode_Credit Card	0.032185
PreferredPaymentMode_Debit Card	-0.019560
PreferredPaymentMode_E wallet	-0.014552
PreferredPaymentMode_UPI	-0.011405
Gender_Female	-0.005292
Gender_Male	0.005292
PreferedOrderCat_Fashion	-0.014552
PreferedOrderCat_Grocery	1.000000
PreferedOrderCat_Laptop & Accessory	-0.041501
PreferedOrderCat_Mobile	-0.007200
PreferedOrderCat_Mobile Phone	-0.027697
PreferedOrderCat_Others	-0.002762
MaritalStatus_Divorced	-0.016429
MaritalStatus_Married	0.024625
MaritalStatus_Single	-0.013853

	PreferedOrderCat_Laptop & Accessory \
Churn	-0.184584
Tenure	0.104482
CityTier	0.245824
WarehouseToHome	0.043231
HourSpendOnApp	-0.130819
NumberOfDeviceRegistered	-0.080608
SatisfactionScore	-0.025448
NumberOfAddress	0.059515
Complain	-0.018126
OrderAmountHikeFromlastYear	-0.024964
CouponUsed	-0.038409
OrderCount	-0.018610
DaySinceLastOrder	0.122257
CashbackAmount	0.101787
PreferredLoginDevice_Computer	0.056688
PreferredLoginDevice_Mobile Phone	0.093415
PreferredLoginDevice_Phone	-0.183909
PreferredPaymentMode_CC	-0.089557
PreferredPaymentMode_COD	-0.053643
PreferredPaymentMode_Cash on Delivery	-0.065981
PreferredPaymentMode_Credit Card	-0.033678
PreferredPaymentMode_Debit Card	-0.001249
PreferredPaymentMode_E wallet	0.141394
PreferredPaymentMode_UPI	0.003837
Gender_Female	0.043367
Gender_Male	-0.043367
PreferedOrderCat_Fashion	-0.379275
PreferedOrderCat_Grocery	-0.041501
PreferedOrderCat_Laptop & Accessory	1.000000

PreferedOrderCat_Mobile	-0.187659
PreferedOrderCat_Mobile Phone	-0.721851
PreferedOrderCat_Others	-0.071997
MaritalStatus_Divorced	-0.000340
MaritalStatus_Married	0.057496
MaritalStatus_Single	-0.060812

	PreferedOrderCat_Mobile \
Churn	0.016675
Tenure	-0.033466
CityTier	-0.097493
WarehouseToHome	-0.015441
HourSpendOnApp	-0.112896
NumberOfDeviceRegistered	-0.125343
SatisfactionScore	-0.006238
NumberOfAddress	-0.031229
Complain	-0.005327
OrderAmountHikeFromlastYear	-0.067089
CouponUsed	-0.088426
OrderCount	-0.082487
DaySinceLastOrder	-0.082597
CashbackAmount	-0.232654
PreferredLoginDevice_Computer	-0.036701
PreferredLoginDevice_Mobile Phone	-0.185187
PreferredLoginDevice_Phone	0.277125
PreferredPaymentMode_CC	0.504562
PreferredPaymentMode_COD	0.053221
PreferredPaymentMode_Cash on Delivery	-0.020480
PreferredPaymentMode_Credit Card	-0.094304
PreferredPaymentMode_Debit Card	0.026243
PreferredPaymentMode_E wallet	-0.065803
PreferredPaymentMode_UPI	-0.034355
Gender_Female	0.001884
Gender_Male	-0.001884
PreferedOrderCat_Fashion	-0.065803
PreferedOrderCat_Grocery	-0.007200
PreferedOrderCat_Laptop & Accessory	-0.187659
PreferedOrderCat_Mobile	1.000000
PreferedOrderCat_Mobile Phone	-0.125238
PreferedOrderCat_Others	-0.012491
MaritalStatus_Divorced	-0.009681
MaritalStatus_Married	-0.028832
MaritalStatus_Single	0.037872

	PreferedOrderCat_Mobile Phone \
Churn	0.181683
Tenure	-0.211477

CityTier	-0.281079
WarehouseToHome	-0.028939
HourSpendOnApp	0.281309
NumberOfDeviceRegistered	0.188563
SatisfactionScore	0.011297
NumberOfAddress	-0.023102
Complain	0.019793
OrderAmountHikeFromlastYear	0.080301
CouponUsed	0.001006
OrderCount	-0.083855
DaySinceLastOrder	-0.221627
CashbackAmount	-0.401303
PreferredLoginDevice_Computer	-0.012669
PreferredLoginDevice_Mobile Phone	-0.162324
PreferredLoginDevice_Phone	0.220374
PreferredPaymentMode_CC	-0.067153
PreferredPaymentMode_COD	0.085898
PreferredPaymentMode_Cash on Delivery	-0.068682
PreferredPaymentMode_Credit Card	0.098421
PreferredPaymentMode_Debit Card	-0.008097
PreferredPaymentMode_E wallet	-0.175791
PreferredPaymentMode_UPI	0.024280
Gender_Female	-0.053910
Gender_Male	0.053910
PreferedOrderCat_Fashion	-0.253118
PreferedOrderCat_Grocery	-0.027697
PreferedOrderCat_Laptop & Accessory	-0.721851
PreferedOrderCat_Mobile	-0.125238
PreferedOrderCat_Mobile Phone	1.000000
PreferedOrderCat_Others	-0.048049
MaritalStatus_Divorced	0.001864
MaritalStatus_Married	-0.062740
MaritalStatus_Single	0.065241

PreferedOrderCat_Others \

Churn	0.010208
Tenure	0.103702
CityTier	-0.011248
WarehouseToHome	-0.035530
HourSpendOnApp	0.023109
NumberOfDeviceRegistered	0.035364
SatisfactionScore	0.030401
NumberOfAddress	-0.004349
Complain	0.007866
OrderAmountHikeFromlastYear	0.021090
CouponUsed	0.082800
OrderCount	0.118690

DaySinceLastOrder	0.071340
CashbackAmount	0.443994
PreferredLoginDevice_Computer	0.005916
PreferredLoginDevice_Mobile Phone	0.021285
PreferredLoginDevice_Phone	-0.033815
PreferredPaymentMode_CC	-0.006698
PreferredPaymentMode_COD	-0.020380
PreferredPaymentMode_Cash on Delivery	0.060785
PreferredPaymentMode_Credit Card	-0.019855
PreferredPaymentMode_Debit Card	0.028682
PreferredPaymentMode_E wallet	-0.001349
PreferredPaymentMode_UPI	-0.019785
Gender_Female	0.006532
Gender_Male	-0.006532
PreferedOrderCat_Fashion	-0.025246
PreferedOrderCat_Grocery	-0.002762
PreferedOrderCat_Laptop & Accessory	-0.071997
PreferedOrderCat_Mobile	-0.012491
PreferedOrderCat_Mobile Phone	-0.048049
PreferedOrderCat_Others	1.000000
MaritalStatus_Divorced	0.004273
MaritalStatus_Married	0.019615
MaritalStatus_Single	-0.024032

	MaritalStatus_Divorced \
Churn	-0.013023
Tenure	0.015788
CityTier	0.019181
WarehouseToHome	-0.013650
HourSpendOnApp	0.018024
NumberOfDeviceRegistered	-0.001099
SatisfactionScore	0.051528
NumberOfAddress	-0.024802
Complain	0.012777
OrderAmountHikeFromlastYear	-0.014517
CouponUsed	0.043312
OrderCount	0.022122
DaySinceLastOrder	-0.007111
CashbackAmount	0.013801
PreferredLoginDevice_Computer	-0.016560
PreferredLoginDevice_Mobile Phone	0.005117
PreferredLoginDevice_Phone	0.012652
PreferredPaymentMode_CC	0.015132
PreferredPaymentMode_COD	0.006594
PreferredPaymentMode_Cash on Delivery	0.007005
PreferredPaymentMode_Credit Card	-0.001500
PreferredPaymentMode_Debit Card	-0.010594

PreferredPaymentMode_E wallet	0.020559
PreferredPaymentMode_UPI	-0.017968
Gender_Female	0.012541
Gender_Male	-0.012541
PreferedOrderCat_Fashion	0.004190
PreferedOrderCat_Grocery	-0.016429
PreferedOrderCat_Laptop & Accessory	-0.000340
PreferedOrderCat_Mobile	-0.009681
PreferedOrderCat_Mobile Phone	0.001864
PreferedOrderCat_Others	0.004273
MaritalStatus_Divorced	1.000000
MaritalStatus_Married	-0.432989
MaritalStatus_Single	-0.288871

	MaritalStatus_Married \
Churn	-0.159808
Tenure	0.085990
CityTier	-0.029876
WarehouseToHome	0.038336
HourSpendOnApp	0.022257
NumberOfDeviceRegistered	-0.064517
SatisfactionScore	-0.012167
NumberOfAddress	-0.001463
Complain	-0.000364
OrderAmountHikeFromlastYear	0.014894
CouponUsed	0.010976
OrderCount	0.043676
DaySinceLastOrder	0.034002
CashbackAmount	0.034234
PreferredLoginDevice_Computer	-0.006364
PreferredLoginDevice_Mobile Phone	0.030007
PreferredLoginDevice_Phone	-0.030677
PreferredPaymentMode_CC	-0.018716
PreferredPaymentMode_COD	-0.029529
PreferredPaymentMode_Cash on Delivery	0.017955
PreferredPaymentMode_Credit Card	0.003140
PreferredPaymentMode_Debit Card	-0.011571
PreferredPaymentMode_E wallet	0.017060
PreferredPaymentMode_UPI	0.024753
Gender_Female	-0.035047
Gender_Male	0.035047
PreferedOrderCat_Fashion	0.010466
PreferedOrderCat_Grocery	0.024625
PreferedOrderCat_Laptop & Accessory	0.057496
PreferedOrderCat_Mobile	-0.028832
PreferedOrderCat_Mobile Phone	-0.062740
PreferedOrderCat_Others	0.019615

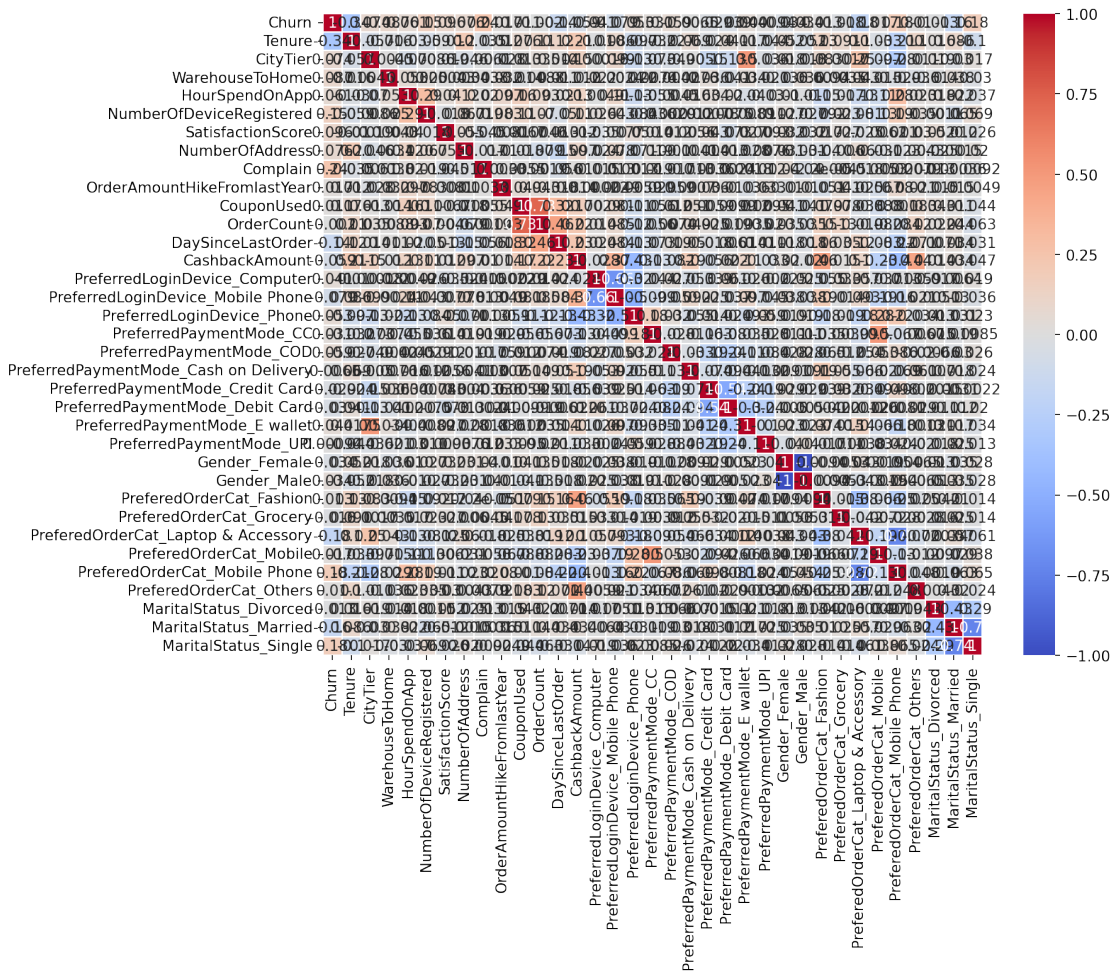
MaritalStatus_Divorced	-0.432989
MaritalStatus_Married	1.000000
MaritalStatus_Single	-0.737892

	MaritalStatus_Single
Churn	0.179481
Tenure	-0.103150
CityTier	0.017370
WarehouseToHome	-0.030496
HourSpendOnApp	-0.037134
NumberOfDeviceRegistered	0.069345
SatisfactionScore	-0.025659
NumberOfAddress	0.020125
Complain	-0.009180
OrderAmountHikeFromlastYear	-0.004949
CouponUsed	-0.044087
OrderCount	-0.062952
DaySinceLastOrder	-0.030789
CashbackAmount	-0.046693
PreferredLoginDevice_Computer	0.019158
PreferredLoginDevice_Mobile Phone	-0.035702
PreferredLoginDevice_Phone	0.023109
PreferredPaymentMode_CC	0.008548
PreferredPaymentMode_COD	0.026426
PreferredPaymentMode_Cash on Delivery	-0.024315
PreferredPaymentMode_Credit Card	-0.002212
PreferredPaymentMode_Debit Card	0.020221
PreferredPaymentMode_E wallet	-0.033513
PreferredPaymentMode_UPI	-0.012836
Gender_Female	0.027833
Gender_Male	-0.027833
PreferedOrderCat_Fashion	-0.014253
PreferedOrderCat_Grocery	-0.013853
PreferedOrderCat_Laptop & Accessory	-0.060812
PreferedOrderCat_Mobile	0.037872
PreferedOrderCat_Mobile Phone	0.065241
PreferedOrderCat_Others	-0.024032
MaritalStatus_Divorced	-0.288871
MaritalStatus_Married	-0.737892
MaritalStatus_Single	1.000000

[35 rows x 35 columns]

```
[31]: import matplotlib.pyplot as plt
from matplotlib.pyplot import figure
figure(figsize=(10, 8), dpi=150)
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', linewidths=0.8)
```

```
plt.show()
```



Optional: What is the correlation of each feature with target Sort the correlation in descending order

```
[35]: corr_matrix.sort_values(by = 'Churn', ascending=False)
```

```
[35]:
```

	Churn	Tenure	CityTier	\
Churn	1.000000	-0.340013	0.073858	
Complain	0.238137	-0.035228	-0.006122	
PreferredOrderCat_Mobile Phone	0.181683	-0.211477	-0.281079	
MaritalStatus_Single	0.179481	-0.103150	0.017370	
NumberOfDeviceRegistered	0.149041	-0.058752	-0.008616	
SatisfactionScore	0.095759	-0.009972	-0.019494	
WarehouseToHome	0.087318	-0.016353	0.004457	
NumberOfAddress	0.076336	0.196547	-0.046406	

CityTier	0.073858	-0.057414	1.000000
HourSpendOnApp	0.060845	-0.029818	-0.070035
PreferredPaymentMode_COD	0.059430	-0.027497	-0.049111
PreferredLoginDevice_Phone	0.053023	-0.096773	-0.125199
PreferredPaymentMode_E wallet	0.043972	0.016709	0.503165
PreferredLoginDevice_Computer	0.040890	-0.010308	-0.000184
Gender_Male	0.033792	-0.052123	-0.018072
PreferredPaymentMode_CC	0.030729	-0.032474	-0.073109
OrderAmountHikeFromlastYear	0.017193	0.012197	-0.027628
PreferedOrderCat_Mobile	0.016675	-0.033466	-0.097493
PreferedOrderCat_Fashion	0.013086	0.130259	0.083035
CouponUsed	0.010982	0.075959	0.013313
PreferedOrderCat_Others	0.010208	0.103702	-0.011248
PreferredPaymentMode_UPI	0.000938	-0.043942	-0.036088
OrderCount	-0.001962	0.112339	0.035018
PreferredPaymentMode_Cash on Delivery	-0.006498	0.069208	-0.004977
MaritalStatus_Divorced	-0.013023	0.015788	0.019181
PreferedOrderCat_Grocery	-0.017880	0.091251	-0.001750
PreferredPaymentMode_Credit Card	-0.029392	-0.024243	-0.150637
Gender_Female	-0.033792	0.052123	0.018072
PreferredPaymentMode_Debit Card	-0.039232	0.040958	-0.127528
CashbackAmount	-0.058756	0.213852	0.146569
PreferredLoginDevice_Mobile Phone	-0.079120	0.085754	0.098949
DaySinceLastOrder	-0.139254	0.118932	0.014304
MaritalStatus_Married	-0.159808	0.085990	-0.029876
PreferedOrderCat_Laptop & Accessory	-0.184584	0.104482	0.245824
Tenure	-0.340013	1.000000	-0.057414

	WarehouseToHome	HourSpendOnApp \
Churn	0.087318	0.060845
Complain	0.003829	0.020413
PreferedOrderCat_Mobile Phone	-0.028939	0.281309
MaritalStatus_Single	-0.030496	-0.037134
NumberOfDeviceRegistered	0.024582	0.293021
SatisfactionScore	0.000434	0.039879
WarehouseToHome	1.000000	0.052731
NumberOfAddress	0.003422	0.124962
CityTier	0.004457	-0.070035
HourSpendOnApp	0.052731	1.000000
PreferredPaymentMode_COD	-0.004197	-0.004520
PreferredLoginDevice_Phone	-0.022497	-0.129642
PreferredPaymentMode_E wallet	0.033578	-0.039522
PreferredLoginDevice_Computer	0.022101	-0.004911
Gender_Male	0.035581	-0.009952
PreferredPaymentMode_CC	0.007388	-0.054905
OrderAmountHikeFromlastYear	0.031975	0.096827
PreferedOrderCat_Mobile	-0.015441	-0.112896

PreferedOrderCat_Fashion	-0.009435	-0.147830
CouponUsed	-0.014311	0.160239
PreferedOrderCat_Others	-0.035530	0.023109
PreferredPaymentMode_UPI	-0.021470	-0.030052
OrderCount	0.008800	0.092601
PreferredPaymentMode_Cash on Delivery	-0.007623	0.016056
MaritalStatus_Divorced	-0.013650	0.018024
PreferedOrderCat_Grocery	0.003512	-0.017382
PreferredPaymentMode_Credit Card	0.036141	0.033820
Gender_Female	-0.035581	0.009952
PreferredPaymentMode_Debit Card	-0.041475	0.020115
CashbackAmount	-0.012433	0.131281
PreferredLoginDevice_Mobile Phone	-0.002403	0.106765
DaySinceLastOrder	0.010646	0.019526
MaritalStatus_Married	0.038336	0.022257
PreferedOrderCat_Laptop & Accessory	0.043231	-0.130819
Tenure	-0.016353	-0.029818

	NumberOfDeviceRegistered \
Churn	0.149041
Complain	0.018881
PreferedOrderCat_Mobile Phone	0.188563
MaritalStatus_Single	0.069345
NumberOfDeviceRegistered	1.000000
SatisfactionScore	-0.017788
WarehouseToHome	0.024582
NumberOfAddress	0.066684
CityTier	-0.008616
HourSpendOnApp	0.293021
PreferredPaymentMode_COD	0.002875
PreferredLoginDevice_Phone	-0.084051
PreferredPaymentMode_E wallet	-0.008889
PreferredLoginDevice_Computer	0.025628
Gender_Male	-0.026706
PreferredPaymentMode_CC	-0.036118
OrderAmountHikeFromlastYear	0.083342
PreferedOrderCat_Mobile	-0.125343
PreferedOrderCat_Fashion	-0.091604
CouponUsed	0.114066
PreferedOrderCat_Others	0.035364
PreferredPaymentMode_UPI	0.010839
OrderCount	0.069559
PreferredPaymentMode_Cash on Delivery	-0.011988
MaritalStatus_Divorced	-0.001099
PreferedOrderCat_Grocery	0.022548
PreferredPaymentMode_Credit Card	0.000782
Gender_Female	0.026706

PreferredPaymentMode_Debit Card	0.007461
CashbackAmount	0.113504
PreferredLoginDevice_Mobile Phone	0.042948
DaySinceLastOrder	-0.049817
MaritalStatus_Married	-0.064517
PreferedOrderCat_Laptop & Accessory	-0.080608
Tenure	-0.058752

	SatisfactionScore	NumberOfAddress \
Churn	0.095759	0.076336
Complain	-0.044533	-0.017363
PreferedOrderCat_Mobile Phone	0.011297	-0.023102
MaritalStatus_Single	-0.025659	0.020125
NumberOfDeviceRegistered	-0.017788	0.066684
SatisfactionScore	1.000000	0.054627
WarehouseToHome	0.000434	0.003422
NumberOfAddress	0.054627	1.000000
CityTier	-0.019494	-0.046406
HourSpendOnApp	0.039879	0.124962
PreferredPaymentMode_COD	0.011974	-0.009988
PreferredLoginDevice_Phone	0.050272	-0.071176
PreferredPaymentMode_E wallet	0.026658	-0.027925
PreferredLoginDevice_Computer	-0.035021	-0.024398
Gender_Male	-0.032398	-0.030980
PreferredPaymentMode_CC	0.013975	0.018801
OrderAmountHikeFromlastYear	-0.008143	0.010275
PreferedOrderCat_Mobile	-0.006238	-0.031229
PreferedOrderCat_Fashion	0.016587	-0.040136
CouponUsed	0.006687	-0.017851
PreferedOrderCat_Others	0.030401	-0.004349
PreferredPaymentMode_UPI	0.009321	0.007556
OrderCount	-0.004622	-0.079413
PreferredPaymentMode_Cash on Delivery	0.005600	-0.004053
MaritalStatus_Divorced	0.051528	-0.024802
PreferedOrderCat_Grocery	0.027097	-0.005968
PreferredPaymentMode_Credit Card	0.043184	0.003998
Gender_Female	0.032398	0.030980
PreferredPaymentMode_Debit Card	-0.073269	0.013272
CashbackAmount	0.012061	0.096752
PreferredLoginDevice_Mobile Phone	-0.007731	0.078405
DaySinceLastOrder	0.012950	-0.148784
MaritalStatus_Married	-0.012167	-0.001463
PreferedOrderCat_Laptop & Accessory	-0.025448	0.059515
Tenure	-0.009972	0.196547

	Complain	OrderAmountHikeFromlastYear \
Churn	0.238137	0.017193

Complain	1.000000	0.003260
PreferedOrderCat_Mobile Phone	0.019793	0.080301
MaritalStatus_Single	-0.009180	-0.004949
NumberOfDeviceRegistered	0.018881	0.083342
SatisfactionScore	-0.044533	-0.008143
WarehouseToHome	0.003829	0.031975
NumberOfAddress	-0.017363	0.010275
CityTier	-0.006122	-0.027628
HourSpendOnApp	0.020413	0.096827
PreferredPaymentMode_COD	-0.017256	-0.059288
PreferredLoginDevice_Phone	0.001261	-0.005919
PreferredPaymentMode_E wallet	0.001807	-0.036139
PreferredLoginDevice_Computer	-0.014878	-0.000220
Gender_Male	-0.039522	0.010081
PreferredPaymentMode_CC	0.019183	-0.028536
OrderAmountHikeFromlastYear	0.003260	1.000000
PreferedOrderCat_Mobile	-0.005327	-0.067089
PreferedOrderCat_Fashion	-0.000022	-0.051335
CouponUsed	0.005472	0.048959
PreferedOrderCat_Others	0.007866	0.021090
PreferredPaymentMode_UPI	0.012421	0.033331
OrderCount	-0.001917	0.042950
PreferredPaymentMode_Cash on Delivery	-0.013373	0.000700
MaritalStatus_Divorced	0.012777	-0.014517
PreferedOrderCat_Grocery	0.004534	0.041476
PreferredPaymentMode_Credit Card	0.003623	0.036106
Gender_Female	0.039522	-0.010081
PreferredPaymentMode_Debit Card	-0.002413	0.010242
CashbackAmount	0.010389	0.014215
PreferredLoginDevice_Mobile Phone	0.012571	0.004871
DaySinceLastOrder	-0.056315	0.017716
MaritalStatus_Married	-0.000364	0.014894
PreferedOrderCat_Laptop & Accessory	-0.018126	-0.024964
Tenure	-0.035228	0.012197

	... Gender_Male \	
Churn	... 0.033792	
Complain	... -0.039522	
PreferedOrderCat_Mobile Phone	... 0.053910	
MaritalStatus_Single	... -0.027833	
NumberOfDeviceRegistered	... -0.026706	
SatisfactionScore	... -0.032398	
WarehouseToHome	... 0.035581	
NumberOfAddress	... -0.030980	
CityTier	... -0.018072	
HourSpendOnApp	... -0.009952	
PreferredPaymentMode_COD	... 0.027718	

PreferredLoginDevice_Phone	...	0.018568
PreferredPaymentMode_E wallet	...	-0.022827
PreferredLoginDevice_Computer	...	0.025482
Gender_Male	...	1.000000
PreferredPaymentMode_CC	...	0.010948
OrderAmountHikeFromlastYear	...	0.010081
PreferedOrderCat_Mobile	...	-0.001884
PreferedOrderCat_Fashion	...	-0.009374
CouponUsed	...	-0.041493
PreferedOrderCat_Others	...	-0.006532
PreferredPaymentMode_UPI	...	0.039949
OrderCount	...	-0.035418
PreferredPaymentMode_Cash on Delivery	...	-0.009100
MaritalStatus_Divorced	...	-0.012541
PreferedOrderCat_Grocery	...	0.005292
PreferredPaymentMode_Credit Card	...	-0.028839
Gender_Female	...	-1.000000
PreferredPaymentMode_Debit Card	...	0.004968
CashbackAmount	...	-0.020263
PreferredLoginDevice_Mobile Phone	...	-0.037886
DaySinceLastOrder	...	-0.018478
MaritalStatus_Married	...	0.035047
PreferedOrderCat_Laptop & Accessory	...	-0.043367
Tenure	...	-0.052123

	PreferedOrderCat_Fashion \	
Churn		0.013086
Complain		-0.000022
PreferedOrderCat_Mobile Phone		-0.253118
MaritalStatus_Single		-0.014253
NumberOfDeviceRegistered		-0.091604
SatisfactionScore		0.016587
WarehouseToHome		-0.009435
NumberOfAddress		-0.040136
CityTier		0.083035
HourSpendOnApp		-0.147830
PreferredPaymentMode_COD		-0.064820
PreferredLoginDevice_Phone		-0.176046
PreferredPaymentMode_E wallet		0.074168
PreferredLoginDevice_Computer		-0.054930
Gender_Male		-0.009374
PreferredPaymentMode_CC		-0.035283
OrderAmountHikeFromlastYear		-0.051335
PreferedOrderCat_Mobile		-0.065803
PreferedOrderCat_Fashion		1.000000
CouponUsed		0.078762
PreferedOrderCat_Others		-0.025246

PreferredPaymentMode_UPI	-0.016992
OrderCount	0.154599
PreferredPaymentMode_Cash on Delivery	0.193691
MaritalStatus_Divorced	0.004190
PreferedOrderCat_Grocery	-0.014552
PreferredPaymentMode_Credit Card	-0.039491
Gender_Female	0.009374
PreferredPaymentMode_Debit Card	-0.004245
CashbackAmount	0.455466
PreferredLoginDevice_Mobile Phone	0.188988
DaySinceLastOrder	0.157977
MaritalStatus_Married	0.010466
PreferedOrderCat_Laptop & Accessory	-0.379275
Tenure	0.130259

	PreferedOrderCat_Grocery \
Churn	-0.017880
Complain	0.004534
PreferedOrderCat_Mobile Phone	-0.027697
MaritalStatus_Single	-0.013853
NumberOfDeviceRegistered	0.022548
SatisfactionScore	0.027097
WarehouseToHome	0.003512
NumberOfAddress	-0.005968
CityTier	-0.001750
HourSpendOnApp	-0.017382
PreferredPaymentMode_COD	-0.011748
PreferredLoginDevice_Phone	-0.019492
PreferredPaymentMode_E wallet	-0.014552
PreferredLoginDevice_Computer	0.032595
Gender_Male	0.005292
PreferredPaymentMode_CC	-0.003861
OrderAmountHikeFromlastYear	0.041476
PreferedOrderCat_Mobile	-0.007200
PreferedOrderCat_Fashion	-0.014552
CouponUsed	0.078182
PreferedOrderCat_Others	-0.002762
PreferredPaymentMode_UPI	-0.011405
OrderCount	0.126821
PreferredPaymentMode_Cash on Delivery	0.054822
MaritalStatus_Divorced	-0.016429
PreferedOrderCat_Grocery	1.000000
PreferredPaymentMode_Credit Card	0.032185
Gender_Female	-0.005292
PreferredPaymentMode_Debit Card	-0.019560
CashbackAmount	0.014736
PreferredLoginDevice_Mobile Phone	-0.014342

DaySinceLastOrder	0.035204
MaritalStatus_Married	0.024625
PreferedOrderCat_Laptop & Accessory	-0.041501
Tenure	0.091251

	PreferedOrderCat_Laptop & Accessory \
Churn	-0.184584
Complain	-0.018126
PreferedOrderCat_Mobile Phone	-0.721851
MaritalStatus_Single	-0.060812
NumberOfDeviceRegistered	-0.080608
SatisfactionScore	-0.025448
WarehouseToHome	0.043231
NumberOfAddress	0.059515
CityTier	0.245824
HourSpendOnApp	-0.130819
PreferredPaymentMode_COD	-0.053643
PreferredLoginDevice_Phone	-0.183909
PreferredPaymentMode_E wallet	0.141394
PreferredLoginDevice_Computer	0.056688
Gender_Male	-0.043367
PreferredPaymentMode_CC	-0.089557
OrderAmountHikeFromlastYear	-0.024964
PreferedOrderCat_Mobile	-0.187659
PreferedOrderCat_Fashion	-0.379275
CouponUsed	-0.038409
PreferedOrderCat_Others	-0.071997
PreferredPaymentMode_UPI	0.003837
OrderCount	-0.018610
PreferredPaymentMode_Cash on Delivery	-0.065981
MaritalStatus_Divorced	-0.000340
PreferedOrderCat_Grocery	-0.041501
PreferredPaymentMode_Credit Card	-0.033678
Gender_Female	0.043367
PreferredPaymentMode_Debit Card	-0.001249
CashbackAmount	0.101787
PreferredLoginDevice_Mobile Phone	0.093415
DaySinceLastOrder	0.122257
MaritalStatus_Married	0.057496
PreferedOrderCat_Laptop & Accessory	1.000000
Tenure	0.104482

	PreferedOrderCat_Mobile \
Churn	0.016675
Complain	-0.005327
PreferedOrderCat_Mobile Phone	-0.125238
MaritalStatus_Single	0.037872

NumberOfDeviceRegistered	-0.125343
SatisfactionScore	-0.006238
WarehouseToHome	-0.015441
NumberOfAddress	-0.031229
CityTier	-0.097493
HourSpendOnApp	-0.112896
PreferredPaymentMode_COD	0.053221
PreferredLoginDevice_Phone	0.277125
PreferredPaymentMode_E wallet	-0.065803
PreferredLoginDevice_Computer	-0.036701
Gender_Male	-0.001884
PreferredPaymentMode_CC	0.504562
OrderAmountHikeFromlastYear	-0.067089
PreferedOrderCat_Mobile	1.000000
PreferedOrderCat_Fashion	-0.065803
CouponUsed	-0.088426
PreferedOrderCat_Others	-0.012491
PreferredPaymentMode_UPI	-0.034355
OrderCount	-0.082487
PreferredPaymentMode_Cash on Delivery	-0.020480
MaritalStatus_Divorced	-0.009681
PreferedOrderCat_Grocery	-0.007200
PreferredPaymentMode_Credit Card	-0.094304
Gender_Female	0.001884
PreferredPaymentMode_Debit Card	0.026243
CashbackAmount	-0.232654
PreferredLoginDevice_Mobile Phone	-0.185187
DaySinceLastOrder	-0.082597
MaritalStatus_Married	-0.028832
PreferedOrderCat_Laptop & Accessory	-0.187659
Tenure	-0.033466

	PreferedOrderCat_Mobile Phone \
Churn	0.181683
Complain	0.019793
PreferedOrderCat_Mobile Phone	1.000000
MaritalStatus_Single	0.065241
NumberOfDeviceRegistered	0.188563
SatisfactionScore	0.011297
WarehouseToHome	-0.028939
NumberOfAddress	-0.023102
CityTier	-0.281079
HourSpendOnApp	0.281309
PreferredPaymentMode_COD	0.085898
PreferredLoginDevice_Phone	0.220374
PreferredPaymentMode_E wallet	-0.175791
PreferredLoginDevice_Computer	-0.012669

Gender_Male	0.053910
PreferredPaymentMode_CC	-0.067153
OrderAmountHikeFromlastYear	0.080301
PreferedOrderCat_Mobile	-0.125238
PreferedOrderCat_Fashion	-0.253118
CouponUsed	0.001006
PreferedOrderCat_Others	-0.048049
PreferredPaymentMode_UPI	0.024280
OrderCount	-0.083855
PreferredPaymentMode_Cash on Delivery	-0.068682
MaritalStatus_Divorced	0.001864
PreferedOrderCat_Grocery	-0.027697
PreferredPaymentMode_Credit Card	0.098421
Gender_Female	-0.053910
PreferredPaymentMode_Debit Card	-0.008097
CashbackAmount	-0.401303
PreferredLoginDevice_Mobile Phone	-0.162324
DaySinceLastOrder	-0.221627
MaritalStatus_Married	-0.062740
PreferedOrderCat_Laptop & Accessory	-0.721851
Tenure	-0.211477

	PreferedOrderCat_Others \
Churn	0.010208
Complain	0.007866
PreferedOrderCat_Mobile Phone	-0.048049
MaritalStatus_Single	-0.024032
NumberOfDeviceRegistered	0.035364
SatisfactionScore	0.030401
WarehouseToHome	-0.035530
NumberOfAddress	-0.004349
CityTier	-0.011248
HourSpendOnApp	0.023109
PreferredPaymentMode_COD	-0.020380
PreferredLoginDevice_Phone	-0.033815
PreferredPaymentMode_E wallet	-0.001349
PreferredLoginDevice_Computer	0.005916
Gender_Male	-0.006532
PreferredPaymentMode_CC	-0.006698
OrderAmountHikeFromlastYear	0.021090
PreferedOrderCat_Mobile	-0.012491
PreferedOrderCat_Fashion	-0.025246
CouponUsed	0.082800
PreferedOrderCat_Others	1.000000
PreferredPaymentMode_UPI	-0.019785
OrderCount	0.118690
PreferredPaymentMode_Cash on Delivery	0.060785

MaritalStatus_Divorced	0.004273
PreferedOrderCat_Grocery	-0.002762
PreferredPaymentMode_Credit Card	-0.019855
Gender_Female	0.006532
PreferredPaymentMode_Debit Card	0.028682
CashbackAmount	0.443994
PreferredLoginDevice_Mobile Phone	0.021285
DaySinceLastOrder	0.071340
MaritalStatus_Married	0.019615
PreferedOrderCat_Laptop & Accessory	-0.071997
Tenure	0.103702

	MaritalStatus_Divorced \
Churn	-0.013023
Complain	0.012777
PreferedOrderCat_Mobile Phone	0.001864
MaritalStatus_Single	-0.288871
NumberOfDeviceRegistered	-0.001099
SatisfactionScore	0.051528
WarehouseToHome	-0.013650
NumberOfAddress	-0.024802
CityTier	0.019181
HourSpendOnApp	0.018024
PreferredPaymentMode_COD	0.006594
PreferredLoginDevice_Phone	0.012652
PreferredPaymentMode_E wallet	0.020559
PreferredLoginDevice_Computer	-0.016560
Gender_Male	-0.012541
PreferredPaymentMode_CC	0.015132
OrderAmountHikeFromlastYear	-0.014517
PreferedOrderCat_Mobile	-0.009681
PreferedOrderCat_Fashion	0.004190
CouponUsed	0.043312
PreferedOrderCat_Others	0.004273
PreferredPaymentMode_UPI	-0.017968
OrderCount	0.022122
PreferredPaymentMode_Cash on Delivery	0.007005
MaritalStatus_Divorced	1.000000
PreferedOrderCat_Grocery	-0.016429
PreferredPaymentMode_Credit Card	-0.001500
Gender_Female	0.012541
PreferredPaymentMode_Debit Card	-0.010594
CashbackAmount	0.013801
PreferredLoginDevice_Mobile Phone	0.005117
DaySinceLastOrder	-0.007111
MaritalStatus_Married	-0.432989
PreferedOrderCat_Laptop & Accessory	-0.000340

Tenure	0.015788
--------	----------

	MaritalStatus_Married \
Churn	-0.159808
Complain	-0.000364
PreferedOrderCat_Mobile Phone	-0.062740
MaritalStatus_Single	-0.737892
NumberOfDeviceRegistered	-0.064517
SatisfactionScore	-0.012167
WarehouseToHome	0.038336
NumberOfAddress	-0.001463
CityTier	-0.029876
HourSpendOnApp	0.022257
PreferredPaymentMode_COD	-0.029529
PreferredLoginDevice_Phone	-0.030677
PreferredPaymentMode_E wallet	0.017060
PreferredLoginDevice_Computer	-0.006364
Gender_Male	0.035047
PreferredPaymentMode_CC	-0.018716
OrderAmountHikeFromlastYear	0.014894
PreferedOrderCat_Mobile	-0.028832
PreferedOrderCat_Fashion	0.010466
CouponUsed	0.010976
PreferedOrderCat_Others	0.019615
PreferredPaymentMode_UPI	0.024753
OrderCount	0.043676
PreferredPaymentMode_Cash on Delivery	0.017955
MaritalStatus_Divorced	-0.432989
PreferedOrderCat_Grocery	0.024625
PreferredPaymentMode_Credit Card	0.003140
Gender_Female	-0.035047
PreferredPaymentMode_Debit Card	-0.011571
CashbackAmount	0.034234
PreferredLoginDevice_Mobile Phone	0.030007
DaySinceLastOrder	0.034002
MaritalStatus_Married	1.000000
PreferedOrderCat_Laptop & Accessory	0.057496
Tenure	0.085990

	MaritalStatus_Single
Churn	0.179481
Complain	-0.009180
PreferedOrderCat_Mobile Phone	0.065241
MaritalStatus_Single	1.000000
NumberOfDeviceRegistered	0.069345
SatisfactionScore	-0.025659
WarehouseToHome	-0.030496

NumberOfAddress	0.020125
CityTier	0.017370
HourSpendOnApp	-0.037134
PreferredPaymentMode_COD	0.026426
PreferredLoginDevice_Phone	0.023109
PreferredPaymentMode_E wallet	-0.033513
PreferredLoginDevice_Computer	0.019158
Gender_Male	-0.027833
PreferredPaymentMode_CC	0.008548
OrderAmountHikeFromlastYear	-0.004949
PreferedOrderCat_Mobile	0.037872
PreferedOrderCat_Fashion	-0.014253
CouponUsed	-0.044087
PreferedOrderCat_Others	-0.024032
PreferredPaymentMode_UPI	-0.012836
OrderCount	-0.062952
PreferredPaymentMode_Cash on Delivery	-0.024315
MaritalStatus_Divorced	-0.288871
PreferedOrderCat_Grocery	-0.013853
PreferredPaymentMode_Credit Card	-0.002212
Gender_Female	0.027833
PreferredPaymentMode_Debit Card	0.020221
CashbackAmount	-0.046693
PreferredLoginDevice_Mobile Phone	-0.035702
DaySinceLastOrder	-0.030789
MaritalStatus_Married	-0.737892
PreferedOrderCat_Laptop & Accessory	-0.060812
Tenure	-0.103150

[35 rows x 35 columns]

4 3. Modelling

4.1 Prepare data

Fill nulls in each column

```
[36]: from sklearn.impute import KNNImputer
      imputer = KNNImputer(n_neighbors=2)
```

```
[37]: for col in df.columns:
      nullCount = df[col].isnull().sum()
      if nullCount > 0:
          df[col]=imputer.fit_transform(df[[col]])
          print("Filled", nullCount, "nulls in", col)
```

Filled 264 nulls in Tenure
 Filled 251 nulls in WarehouseToHome
 Filled 255 nulls in HourSpendOnApp
 Filled 265 nulls in OrderAmountHikeFromlastYear
 Filled 256 nulls in CouponUsed
 Filled 258 nulls in OrderCount
 Filled 307 nulls in DaySinceLastOrder

Question: Make the data suitable for model training

```
[39]: df_final.head(5)
```

```
[39]:
```

	Churn	Tenure	CityTier	WarehouseToHome	HourSpendOnApp	\
0	1	4.0	3	6.0	3.0	
3	1	0.0	3	15.0	2.0	
5	1	0.0	1	22.0	3.0	
11	1	11.0	1	6.0	3.0	
12	1	0.0	1	11.0	2.0	

	NumberOfDeviceRegistered	SatisfactionScore	NumberOfAddress	Complain	\
0	3	2	9	1	
3	4	5	8	0	
5	5	5	2	1	
11	4	3	10	1	
12	3	3	2	1	

	OrderAmountHikeFromlastYear	...	Gender_Male	PreferedOrderCat_Fashion	\
0	11.0	...	0	0	
3	23.0	...	1	0	
5	22.0	...	0	0	
11	13.0	...	1	1	
12	13.0	...	1	0	

	PreferedOrderCat_Grocery	PreferedOrderCat_Laptop & Accessory	\
0	0	1	
3	0	1	
5	0	0	
11	0	0	
12	0	0	

	PreferedOrderCat_Mobile	PreferedOrderCat_Mobile Phone	\
0	0	0	
3	0	0	
5	0	1	
11	0	0	
12	1	0	

	PreferedOrderCat_Others	MaritalStatus_Divorced	MaritalStatus_Married	\
--	-------------------------	------------------------	-----------------------	---

0	0	0	0
3	0	0	0
5	0	0	0
11	0	0	0
12	0	0	0

MaritalStatus_Single	
0	1
3	1
5	1
11	1
12	1

[5 rows x 35 columns]

```
[40]: from sklearn.linear_model import LogisticRegression
x = df_final.drop(['Churn'],axis = 1)
x.head(5)
y = df_final['Churn']
```

```
[47]: from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2,
↳random_state=42)
```

```
[48]: #check the shape of X_train and X_test
x_train.shape, x_test.shape
```

```
[48]: ((3019, 34), (755, 34))
```

```
[49]: clf = LogisticRegression(random_state=0).fit(x_train,y_train)
```

```
/Users/jacque/opt/anaconda3/lib/python3.9/site-
packages/sklearn/linear_model/_logistic.py:814: ConvergenceWarning: lbfgs failed
to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
```

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

```
n_iter_i = _check_optimize_result(
```

```
[51]: clf.fit(x_train, y_train)
```

```
/Users/jacque/opt/anaconda3/lib/python3.9/site-
packages/sklearn/linear_model/_logistic.py:814: ConvergenceWarning: lbfgs failed
to converge (status=1):
```

STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

```
n_iter_i = _check_optimize_result(
```

```
[51]: LogisticRegression(random_state=0)
```

```
[52]: #predict result
y_pred = clf.predict(x_test)
```

4.2 Model training

Question: Train one or more models and show their performance on training and test data Train and show train & test accuracy

```
[54]: from sklearn.metrics import roc_curve, roc_auc_score, classification_report, \
      accuracy_score, confusion_matrix
train_accuracy = accuracy_score(y_pred, y_test)*100
train_accuracy
```

```
[54]: 89.27152317880794
```

```
[56]: y_pred_train = clf.predict(x_train)
y_pred_train
```

```
[56]: array([0, 0, 1, ..., 0, 1, 0])
```

```
[57]: print('Training-set accuracy score: {0:0.4f}'.format(accuracy_score(y_train, \
      y_pred_train)))
```

Training-set accuracy score: 0.8917

5 4. Evaluation

Feel free to use any libraries / stackoverflow

Question: Show the precision, recall and f1 score for the model with best accuracy

```
[59]: print(classification_report(y_pred, y_test, digits=6))
```

	precision	recall	f1-score	support
0	0.958333	0.920000	0.938776	675
1	0.495327	0.662500	0.566845	80

accuracy			0.892715	755
macro avg	0.726830	0.791250	0.752810	755
weighted avg	0.909273	0.892715	0.899366	755

```
print(classification_report(y_train, y_pred_train, digits=6))
```

5.1 Confusion matrix

Question: Show the confusion matrix Describe the performance of model and steps you could take to improve it

Bonus: Plot the confusion matrix

```
[61]: print('Confusion matrix:\n', confusion_matrix(y_pred, y_test))
```

Confusion matrix:

```
[[621  54]
 [ 27  53]]
```

```
[62]: print('Confusion matrix:\n', confusion_matrix(y_pred_train, y_train))
```

Confusion matrix:

```
[[2404  236]
 [  91  288]]
```

6 Discussion

```
[64]: #check for overfitting and underfitting
print('Training set score: {:.4f}'.format(clf.score(x_train, y_train)))
```

Training set score: 0.8917

```
[65]: print('Test set score: {:.4f}'.format(clf.score(x_test, y_test)))
```

Test set score: 0.8927

The training-set accuracy score is 0.8917 while the test-set accuracy to be 0.8927. These two values are quite comparable. So, there is no question of overfitting.

Next Step, we could perform hyperparameter optimization using gridsearch CV to improve the performance for this particular model

```
[ ]:
```