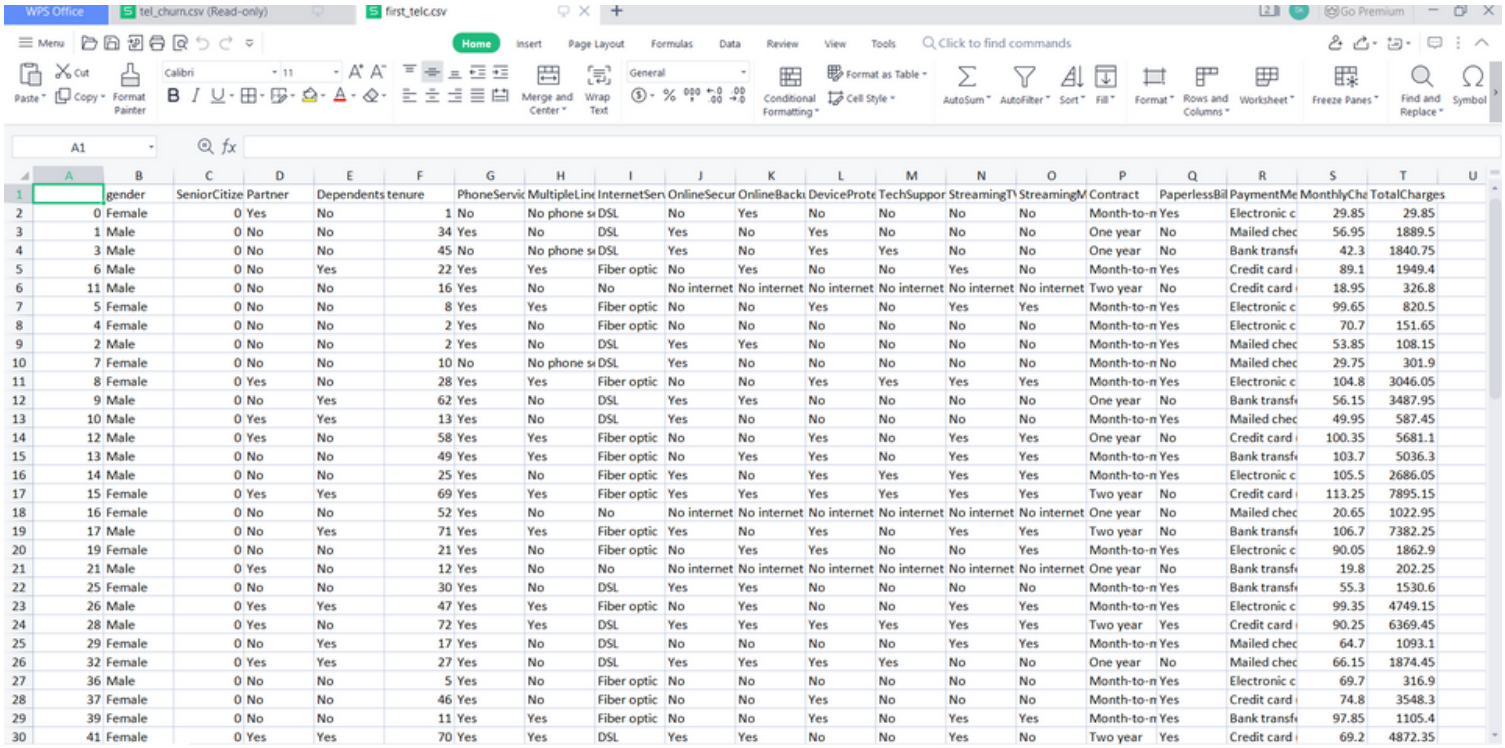
MACHINE LEARNING END-TO-END PROJECT

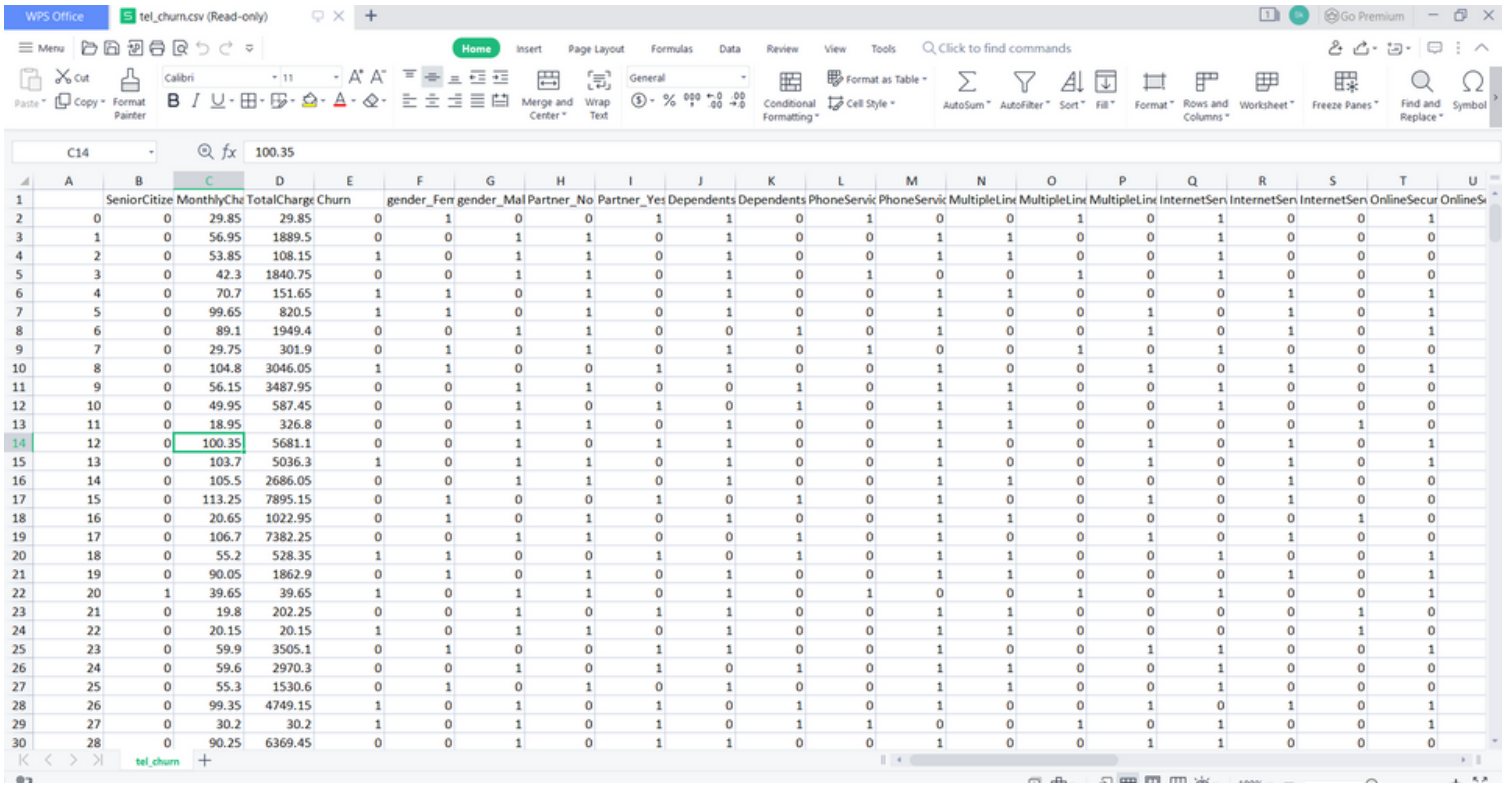
CUSTOMER CHURN PREDICTION

-

TELECOM



	gender	SeniorCitize	Partner	Dependents	tenure	PhoneServic	MultipleLine	InternetSer	OnlineSecur	OnlineBacku	DeviceProte	TechSuppor	StreamingT	StreamingM	Contract	PaperlessBil	PaymentMe	MonthlyCha	TotalCharges
0	Female	0	Yes	No	1	No	No phone s	DSL	No	Yes	No	No	No	No	Month-to-m	Yes	Electronic c	29.85	29.85
1	Male	0	No	No	34	Yes	No	DSL	Yes	No	Yes	No	No	No	One year	No	Mailed chec	56.95	1889.5
3	Male	0	No	No	45	No	No phone s	DSL	Yes	No	Yes	Yes	No	No	One year	No	Bank transfe	42.3	1840.75
6	Male	0	No	Yes	22	Yes	Yes	Fiber optic	No	Yes	No	No	Yes	No	Month-to-m	Yes	Credit card	89.1	1949.4
11	Male	0	No	No	16	Yes	No	No	No internet	No internet	No internet	No internet	No internet	No internet	Two year	No	Credit card	18.95	326.8
5	Female	0	No	No	8	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Yes	Month-to-m	Yes	Electronic c	99.65	820.5
4	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No	No	No	No	Month-to-m	Yes	Electronic c	70.7	151.65
2	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	No	No	No	No	Month-to-m	Yes	Mailed chec	53.85	108.15
7	Female	0	No	No	10	No	No phone s	DSL	Yes	No	No	No	No	No	Month-to-m	No	Mailed chec	29.75	301.9
8	Female	0	Yes	No	28	Yes	Yes	Fiber optic	No	No	Yes	Yes	Yes	Yes	Month-to-m	Yes	Electronic c	104.8	3046.05
9	Male	0	No	Yes	62	Yes	No	DSL	Yes	Yes	No	No	No	No	One year	No	Bank transfe	56.15	3487.95
10	Male	0	Yes	Yes	13	Yes	No	DSL	Yes	No	No	No	No	No	Month-to-m	Yes	Mailed chec	49.95	587.45
12	Male	0	Yes	No	58	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Yes	One year	No	Credit card	100.35	5681.1
13	Male	0	No	No	49	Yes	Yes	Fiber optic	No	Yes	Yes	No	Yes	Yes	Month-to-m	Yes	Bank transfe	103.7	5036.3
14	Male	0	No	No	25	Yes	No	Fiber optic	Yes	No	Yes	Yes	Yes	Yes	Month-to-m	Yes	Electronic c	105.5	2686.05
15	Female	0	Yes	Yes	69	Yes	Yes	Fiber optic	Yes	Yes	Yes	Yes	Yes	Yes	Two year	No	Credit card	113.25	7895.15
16	Female	0	No	No	52	Yes	No	No	No internet	No internet	No internet	No internet	No internet	No internet	One year	No	Mailed chec	20.65	1022.95
17	Male	0	No	Yes	71	Yes	Yes	Fiber optic	Yes	No	Yes	No	Yes	Yes	Two year	No	Bank transfe	106.7	7382.25
19	Female	0	No	No	21	Yes	No	Fiber optic	No	Yes	Yes	No	No	Yes	Month-to-m	Yes	Electronic c	90.05	1862.9
21	Male	0	Yes	No	12	Yes	No	No	No internet	No internet	No internet	No internet	No internet	No internet	One year	No	Bank transfe	19.8	202.25
25	Female	0	No	No	30	Yes	No	DSL	Yes	Yes	No	No	No	No	Month-to-m	Yes	Bank transfe	55.3	1530.6
26	Male	0	Yes	Yes	47	Yes	Yes	Fiber optic	No	Yes	No	No	Yes	Yes	Month-to-m	Yes	Electronic c	99.35	4749.15
28	Male	0	Yes	No	72	Yes	Yes	DSL	Yes	Yes	Yes	Yes	Yes	Yes	Two year	Yes	Credit card	90.25	6369.45
29	Female	0	No	Yes	17	Yes	No	DSL	No	No	No	No	Yes	Yes	Month-to-m	Yes	Mailed chec	64.7	1093.1
32	Female	0	Yes	Yes	27	Yes	No	DSL	Yes	Yes	Yes	Yes	No	No	One year	No	Mailed chec	66.15	1874.45
36	Male	0	No	No	5	Yes	No	Fiber optic	No	No	No	No	No	No	Month-to-m	Yes	Electronic c	69.7	316.9
37	Female	0	No	No	46	Yes	No	Fiber optic	No	No	Yes	No	No	No	Month-to-m	Yes	Credit card	74.8	3548.3
39	Female	0	No	No	11	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Yes	Month-to-m	Yes	Bank transfe	97.85	1105.4
41	Female	0	Yes	Yes	70	Yes	Yes	DSL	Yes	Yes	No	No	Yes	No	Two year	Yes	Credit card	69.2	4872.35



	SeniorCitize	MonthlyCha	TotalCharge	Churn	gender_Fen	gender_Mal	Partner_No	Partner_Yes	Dependents	Dependents	PhoneServic	PhoneServic	MultipleLine	MultipleLine	MultipleLine	InternetSer	InternetSer	InternetSer	OnlineSecur	OnlineSe
0	0	29.85	29.85	0	1	0	0	1	1	0	1	0	0	1	0	1	0	0	1	
1	0	56.95	1889.5	0	0	1	1	0	1	0	0	1	1	0	0	1	0	0	0	
2	0	53.85	108.15	1	0	1	1	0	1	0	0	1	1	0	0	1	0	0	0	
3	0	42.3	1840.75	0	0	1	1	0	1	0	1	0	0	1	0	1	0	0	0	
4	0	70.7	151.65	1	1	0	1	0	1	0	0	1	1	0	0	0	1	0	1	
5	0	99.65	820.5	1	1	0	1	0	1	0	0	1	0	0	1	0	1	0	1	
6	0	89.1	1949.4	0	0	1	1	0	0	1	0	1	0	0	1	0	1	0	1	
7	0	29.75	301.9	0	1	0	1	0	1	0	1	0	0	1	0	1	0	0	0	
8	0	104.8	3046.05	1	1	0	0	1	1	0	0	1	0	0	1	0	1	0	1	
9	0	56.15	3487.95	0	0	1	1	0	0	1	0	1	1	0	0	1	0	0	0	
10	0	49.95	587.45	0	0	1	0	1	0	1	0	1	1	0	0	1	0	0	0	
11	0	18.95	326.8	0	0	1	1	0	1	0	0	1	1	0	0	0	0	1	0	
12	0	100.35	5681.1	0	0	1	0	1	1	0	0	1	0	0	1	0	1	0	1	
13	0	103.7	5036.3	1	0	1	1	0	1	0	0	1	0	0	1	0	1	0	1	
14	0	105.5	2686.05	0	0	1	1	0	1	0	0	1	1	0	0	1	0	0	0	
15	0	113.25	7895.15	0	1	0	0	1	0	1	0	1	0	0	1	0	1	0	0	
16	0	20.65	1022.95	0	1	0	1	0	1	0	0	1	1	0	0	0	0	1	0	
17	0	106.7	7382.25	0	0	1	1	0	0	1	0	1	0	0	1	0	1	0	0	
18	0	55.2	528.35	1	1	0	0	1	0	1	0	1	1	0	0	1	0	0	1	
19	0	90.05	1862.9	0	1	0	1	0	1	0	0	1	1	0	0	0	1	0	1	
20	1	39.65	39.65	1	0	1	1	0	1	0	1	0	0	1	0	1	0	0	1	
21	0	19.8	202.25	0	0	1	0	1	1	0	0	1	1	0	0	0	0	1	0	
22	0	20.15	20.15	1	0	1	1	0	1	0	0	1	1	0	0	0	0	1	0	
23	0	59.9	3505.1	0	1	0	0	1	1	0	0	1	0	0	1	1	0	0	1	
24	0	59.6	2970.3	0	0	1	0	1	0	1	0	1	1	0	0	1	0	0	0	
25	0	55.3	1530.6	0	1	0	1	0	1	0	0	1	1	0	0	1	0	0	0	
26	0	99.35	4749.15	1	0	1	0	1	0	1	0	1	0	0	1	0	1	0	1	
27	0	30.2	30.2	1	0	1	0	1	0	1	1	0	0	1	0	1	0	0	1	
28	0	90.25	6369.45	0	0	1	0	1	1	0	0	1	0	0	1	1	0	0	0	

Churn Prediction | Home Page - Select or create a n | Project Churn Analysis - Jupyter

localhost:8888/notebooks/Project%20Churn%20Analysis.ipynb

jupyter **Project Churn Analysis** Last Checkpoint: 08/19/2023 (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Run Code

In [1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.ticker as mtick
import matplotlib.pyplot as plt
%matplotlib inline
```

In [4]:
```
telco_base_data = pd.read_csv("C:/Users/DELL/Downloads/WA_Fn-UseC_-Telco-Customer-Churn.csv")
```

In [5]:
```
telco_base_data.head()
```

Out[5]:

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechSupp
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	No	
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	Yes	
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	No	
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	Yes	
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	No	

5 rows × 21 columns

In [6]:
```
telco_base_data.shape
```

Out[6]:
```
(7043, 21)
```

In [7]:
```
telco_base_data.columns.values
```

Out[7]:
```
array(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents',
       'tenure', 'PhoneService', 'MultipleLines', 'InternetService',
       'OnlineSecurity', 'OnlineBackup', 'DeviceProtection',
       'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract',
       'PaperlessBilling', 'PaymentMethod', 'MonthlyCharges',
       'TotalCharges', 'Churn'], dtype=object)
```

In [8]:
```
telco_base_data.dtypes
```

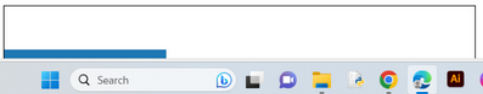
Out[8]:
```
customerID          object
gender              object
SeniorCitizen        int64
Partner             object
Dependents          object
tenure               int64
PhoneService        object
MultipleLines       object
InternetService     object
OnlineSecurity      object
OnlineBackup        object
DeviceProtection    object
TechSupport         object
StreamingTV         object
StreamingMovies     object
Contract            object
PaperlessBilling    object
PaymentMethod       object
MonthlyCharges     float64
TotalCharges        object
Churn               object
dtype: object
```

In [9]:
```
telco_base_data.describe()
```

Out[9]:

	SeniorCitizen	tenure	MonthlyCharges
count	7043.000000	7043.000000	7043.000000
mean	0.162147	32.371149	64.761692
std	0.368612	24.559481	30.090047
min	0.000000	0.000000	18.250000
25%	0.000000	9.000000	35.500000
50%	0.000000	29.000000	70.350000
75%	0.000000	55.000000	89.850000
max	1.000000	72.000000	118.750000

In [10]:
```
telco_base_data['Churn'].value_counts().plot(kind='barh', figsize=(8, 6))
plt.xlabel("Count", labelpad=14)
plt.ylabel("Target Variable", labelpad=14)
plt.title("Count of TARGET Variable per category", y=1.02);
```



31°C Mostly cloudy | Search | ENG IN | 12:51 04-09-2023

Churn Prediction | Home Page - Select or create a n | Project Churn Analysis - Jupyter

localhost:8888/notebooks/Project%20Churn%20Analysis.ipynb

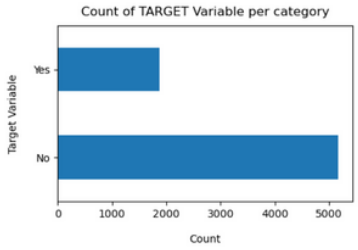
Letter and Symbol... | Applications of Mac... | What is Operating... | IMARTICUS Batch 2... | Imarticus Learning | Coursera | Online C... | Ds Programming-A... | Gmail | YouTube | Maps | Statistics - InClass -...

jupyter **Project Churn Analysis** Last Checkpoint: 08/19/2023 (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help | Not Trusted | Python 3 (ipykernel)

Run | Code

In [9]:
```
telco_base_data['Churn'].value_counts().plot(kind='barh', figsize=(5, 3))
plt.xlabel("Count", labelpad=14)
plt.ylabel("Target Variable", labelpad=14)
plt.title("Count of TARGET Variable per category", y=1.02);
```



In [10]:
```
100*telco_base_data['Churn'].value_counts()/len(telco_base_data['Churn'])
```

Out[10]:
```
No     73.463013
Yes    26.536987
Name: Churn, dtype: float64
```

In [12]:
```
telco_base_data['Churn'].value_counts()
```

Out[12]:
```
No     5174
```

31°C Mostly cloudy | Search | ENG IN | 12:52 04-09-2023

Churn Prediction | Home Page - Select or create a n | Project Churn Analysis - Jupyter

localhost:8888/notebooks/Project%20Churn%20Analysis.ipynb

Letter and Symbol... | Applications of Mac... | What is Operating... | IMARTICUS Batch 2... | Imarticus Learning | Coursera | Online C... | Ds Programming-A... | Gmail | YouTube | Maps | Statistics - InClass -...

jupyter **Project Churn Analysis** Last Checkpoint: 08/19/2023 (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help | Not Trusted | Python 3 (ipykernel)

Run | Code

```
Name: Churn, dtype: int64
```

In [13]:
```
telco_base_data.info(verbose = True)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   customerID        7043 non-null   object
 1   gender            7043 non-null   object
 2   SeniorCitizen     7043 non-null   int64
 3   Partner           7043 non-null   object
 4   Dependents        7043 non-null   object
 5   tenure            7043 non-null   int64
 6   PhoneService      7043 non-null   object
 7   MultipleLines     7043 non-null   object
 8   InternetService   7043 non-null   object
 9   OnlineSecurity    7043 non-null   object
 10  OnlineBackup      7043 non-null   object
 11  DeviceProtection  7043 non-null   object
 12  TechSupport       7043 non-null   object
 13  StreamingTV       7043 non-null   object
 14  StreamingMovies   7043 non-null   object
 15  Contract          7043 non-null   object
 16  PaperlessBilling  7043 non-null   object
 17  PaymentMethod     7043 non-null   object
 18  MonthlyCharges    7043 non-null   float64
 19  TotalCharges      7043 non-null   object
 20  Churn             7043 non-null   object
dtypes: float64(1), int64(2), object(18)
memory usage: 1.1+ MB
```

In [27]:
```
missing = pd.DataFrame((telco_base_data.isnull().sum())*100/telco_base_data.shape[0]).reset_index()
plt.figure(figsize=(16,5))
```

31°C Mostly cloudy | Search | ENG IN | 12:52 04-09-2023

Churn Prediction | Home Page - Select or create a n | Project Churn Analysis - Jupyter

localhost:8888/notebooks/Project%20Churn%20Analysis.ipynb

Letter and Symbol... | Applications of Mac... | What is Operating... | IMARTICUS Batch 2... | Imarticus Learning | Coursera | Online C... | Ds Programming-A... | Gmail | YouTube | Maps | Statistics - InClass -...

jupyter **Project Churn Analysis** Last Checkpoint: 08/19/2023 (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help | Not Trusted | Python 3 (ipykernel)

Run | Code

In [17]:
```
telco_data = telco_base_data.copy()
```

In [18]:
```
telco_data.TotalCharges = pd.to_numeric(telco_data.TotalCharges, errors='coerce')
telco_data.isnull().sum()
```

Out[18]:
```
customerID          0
gender              0
SeniorCitizen       0
Partner             0
Dependents          0
tenure              0
PhoneService        0
MultipleLines       0
InternetService     0
OnlineSecurity      0
OnlineBackup        0
DeviceProtection    0
TechSupport         0
StreamingTV         0
StreamingMovies     0
Contract            0
PaperlessBilling    0
PaymentMethod       0
MonthlyCharges      0
TotalCharges       11
Churn               0
dtype: int64
```

In [19]:
```
telco_data.loc[telco_data ['TotalCharges'].isnull() == True]
```

Out[19]:

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechS
488	4472-LVYGI	Female	0	Yes	Yes	0	No	No phone service	DSL	Yes	...	Yes	

31°C Mostly cloudy | Search | ENG IN | 12:52 04-09-2023

Churn Prediction | Home Page - Select or create a n | Project Churn Analysis - Jupyter

localhost:8888/notebooks/Project%20Churn%20Analysis.ipynb

jupyter **Project Churn Analysis** Last Checkpoint: 08/19/2023 (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help | Not Trusted | Python 3 (ipykernel)

Run Code

In [19]:
```
telco_data.loc[telco_data ['TotalCharges'].isnull() == True]
```

Out[19]:

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechS
488	4472-LVYGI	Female	0	Yes	Yes	0	No	No phone service	DSL	Yes	...	Yes	
753	3115-CZMZD	Male	0	No	Yes	0	Yes	No	No	No internet service	...	No internet service	No i
936	5709-LVOEQ	Female	0	Yes	Yes	0	Yes	No	DSL	Yes	...	Yes	
1082	4367-NUYAO	Male	0	Yes	Yes	0	Yes	Yes	No	No internet service	...	No internet service	No i
1340	1371-DWPAZ	Female	0	Yes	Yes	0	No	No phone service	DSL	Yes	...	Yes	
3331	7644-OMVMY	Male	0	Yes	Yes	0	Yes	No	No	No internet service	...	No internet service	No i
3826	3213-VVOLG	Male	0	Yes	Yes	0	Yes	Yes	No	No internet service	...	No internet service	No i
4380	2520-SGTTA	Female	0	Yes	Yes	0	Yes	No	No	No internet service	...	No internet service	No i
5218	2923-ARZLG	Male	0	Yes	Yes	0	Yes	No	No	No internet service	...	No internet service	No i
6670	4075-WKNIU	Female	0	Yes	Yes	0	Yes	Yes	DSL	No	...	Yes	
6754	2775-SEFEE	Male	0	No	Yes	0	Yes	Yes	DSL	Yes	...	No	

11 rows × 21 columns

31°C Mostly cloudy | Search | ENG IN | 12:52 04-09-2023

Churn Prediction | Home Page - Select or create a n | Project Churn Analysis - Jupyter

localhost:8888/notebooks/Project%20Churn%20Analysis.ipynb

jupyter **Project Churn Analysis** Last Checkpoint: 08/19/2023 (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help | Not Trusted | Python 3 (ipykernel)

Run Code

```
72
```

In [22]:
```
labels = ["{0} - {1}".format(i, i + 11) for i in range(1, 72, 12)]

telco_data['tenure_group'] = pd.cut(telco_data.tenure, range(1, 80, 12), right=False, labels=labels)
```

In [23]:
```
telco_data['tenure_group'].value_counts()
```

Out[23]:
```
1 - 12     2175
61 - 72    1407
13 - 24    1024
25 - 36     832
49 - 60     832
37 - 48     762
Name: tenure_group, dtype: int64
```

In [28]:
```
telco_data.drop(columns= ['customerID','tenure'], axis=1, inplace=True)
telco_data.head()
```

Out[28]:

	gender	SeniorCitizen	Partner	Dependents	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	Strea
0	Female	0	Yes	No	No	No phone service	DSL	No	Yes	No	No	
1	Male	0	No	No	Yes	No	DSL	Yes	No	Yes	No	
2	Male	0	No	No	Yes	No	DSL	Yes	Yes	No	No	
3	Male	0	No	No	No	No phone service	DSL	Yes	No	Yes	Yes	
4	Female	0	No	No	Yes	No	Fiber optic	No	No	No	No	

31°C Mostly cloudy | Search | ENG IN | 12:52 04-09-2023

Churn Prediction | Home Page - Select or create a n | Project Churn Analysis - Jupyter

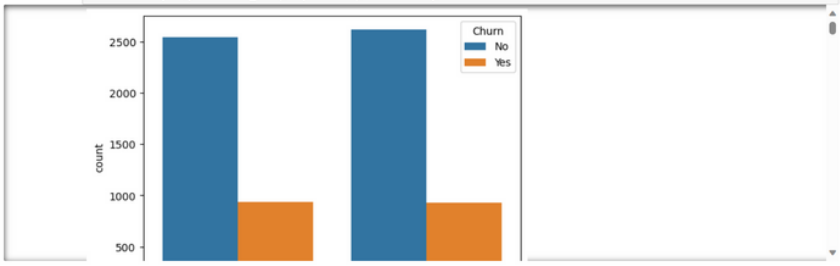
localhost:8888/notebooks/Project%20Churn%20Analysis.ipynb

jupyter **Project Churn Analysis** Last Checkpoint: 08/19/2023 (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help | Not Trusted | Python 3 (ipykernel)

Run Code

In [29]:
```
for i, predictor in enumerate(telco_data.drop(columns=['Churn', 'TotalCharges', 'MonthlyCharges'])):
    plt.figure(i)
    sns.countplot(data=telco_data, x=predictor, hue='Churn')
```



In [30]:
```
telco_data['Churn'] = np.where(telco_data.Churn == 'Yes',1,0)
```

In [31]:
```
telco_data.head()
```

Out[31]:

	gender	SeniorCitizen	Partner	Dependents	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	Strea
0	Female	0	Yes	No	No	No phone service	DSL	No	Yes	No	No	
1	Male	0	No	No	Yes	No	DSL	Yes	No	Yes	No	

31°C Mostly cloudy | Search | ENG IN | 12:52 04-09-2023

Churn Prediction | Home Page - Select or create a n | Project Churn Analysis - Jupyter

localhost:8888/notebooks/Project%20Churn%20Analysis.ipynb

jupyter **Project Churn Analysis** Last Checkpoint: 08/19/2023 (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Run Code

In [30]:
```
telco_data['Churn'] = np.where(telco_data.Churn == 'Yes',1,0)
```

In [31]:
```
telco_data.head()
```

Out[31]:

	gender	SeniorCitizen	Partner	Dependents	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	Strea
0	Female	0	Yes	No	No	No phone service	DSL	No	Yes	No	No	
1	Male	0	No	No	Yes	No	DSL	Yes	No	Yes	No	
2	Male	0	No	No	Yes	No	DSL	Yes	Yes	No	No	
3	Male	0	No	No	No	No phone service	DSL	Yes	No	Yes	Yes	
4	Female	0	No	No	Yes	No	Fiber optic	No	No	No	No	

In [32]:
```
telco_data_dummies = pd.get_dummies(telco_data)
telco_data_dummies.head()
```

Out[32]:

	SeniorCitizen	MonthlyCharges	TotalCharges	Churn	gender_Female	gender_Male	Partner_No	Partner_Yes	Dependents_No	Dependents_Yes	...	PaymentM transfe
0	0	29.85	29.85	0	1	0	0	1	1	0	...	
1	0	56.95	1889.50	0	0	1	1	0	1	0	...	
2	0	53.85	108.15	1	0	1	1	0	1	0	...	
3	0	42.30	1840.75	0	0	1	1	0	1	0	...	
4	0	70.70	151.65	1	1	0	1	0	1	0	...	

5 rows × 51 columns

31°C Mostly cloudy Search ENG IN 12:53 04-09-2023

Churn Prediction | Home Page - Select or create a n | Project Churn Analysis - Jupyter

localhost:8888/notebooks/Project%20Churn%20Analysis.ipynb

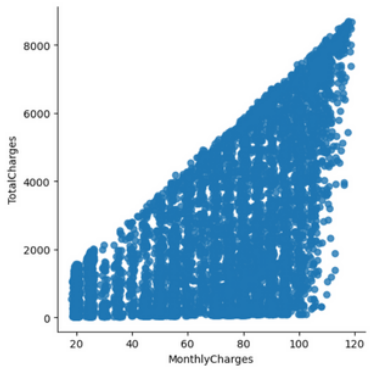
jupyter **Project Churn Analysis** Last Checkpoint: 08/19/2023 (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Run Code

In [33]:
```
sns.lmplot(data=telco_data_dummies, x='MonthlyCharges', y='TotalCharges', fit_reg=False)
```

Out[33]:
```
<seaborn.axisgrid.FacetGrid at 0x24e2fe9cd00>
```



31°C Mostly cloudy Search ENG IN 12:53 04-09-2023

Churn Prediction | Home Page - Select or create a n | Project Churn Analysis - Jupyter

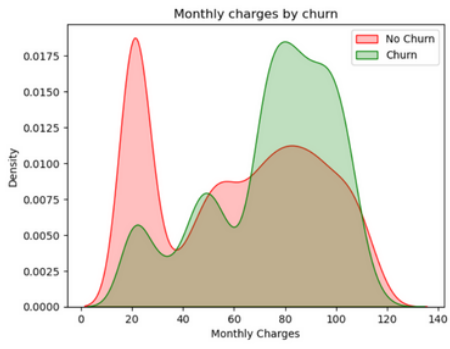
localhost:8888/notebooks/Project%20Churn%20Analysis.ipynb

jupyter **Project Churn Analysis** Last Checkpoint: 08/19/2023 (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Run Code

Out[44]:
```
Text(0.5, 1.0, 'Monthly charges by churn')
```



In [42]:
```
Tot = sns.kdeplot(telco_data_dummies.TotalCharges[(telco_data_dummies["Churn"] == 0) ],
                  color="Red", shade = True)
Tot = sns.kdeplot(telco_data_dummies.TotalCharges[(telco_data_dummies["Churn"] == 1) ],
                  ax =Tot, color="Yellow", shade= True)
```

31°C Mostly cloudy Search ENG IN 12:53 04-09-2023

Churn Prediction | Home Page - Select or create a n | Project Churn Analysis - Jupyter

localhost:8888/notebooks/Project%20Churn%20Analysis.ipynb

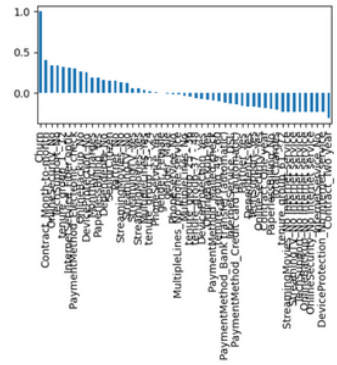
jupyter **Project Churn Analysis** Last Checkpoint: 08/19/2023 (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help | Not Trusted | Python 3 (ipykernel)

Run | Code

In [31]:
```
plt.figure(figsize=(5,2))
telco_data_dummies.corr()['Churn'].sort_values(ascending = False).plot(kind='bar')
```

Out[31]: <Axes: >



31°C Mostly cloudy | Search | ENG IN | 12:53 04-09-2023

Churn Prediction | Home Page - Select or create a n | Project Churn Analysis - Jupyter

localhost:8888/notebooks/Project%20Churn%20Analysis.ipynb

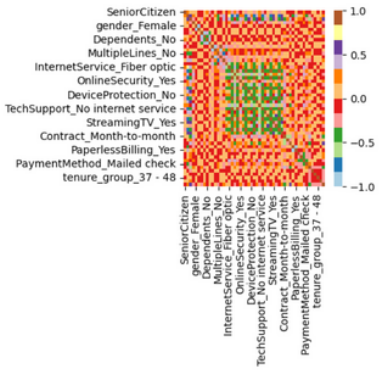
jupyter **Project Churn Analysis** Last Checkpoint: 08/19/2023 (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help | Not Trusted | Python 3 (ipykernel)

Run | Code

In [34]:
```
plt.figure(figsize=(3,3))
sns.heatmap(telco_data_dummies.corr(), cmap="Paired")
```

Out[34]: <Axes: >



31°C Mostly cloudy | Search | ENG IN | 12:54 04-09-2023

Churn Prediction | Home Page - Select or create a n | Project Churn Analysis - Jupyter

localhost:8888/notebooks/Project%20Churn%20Analysis.ipynb

jupyter **Project Churn Analysis** Last Checkpoint: 08/19/2023 (unsaved changes) Logout

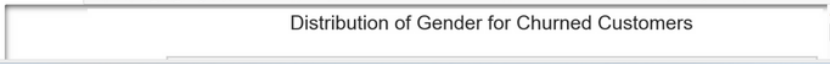
File Edit View Insert Cell Kernel Widgets Help | Not Trusted | Python 3 (ipykernel)

Run | Code

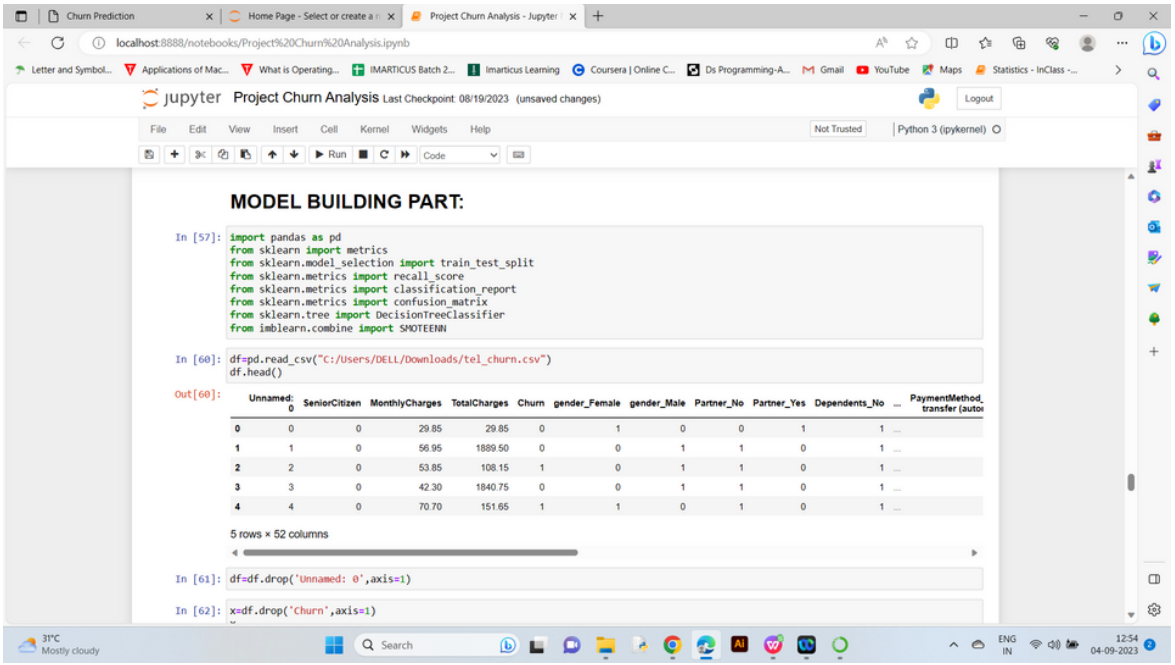
In [33]:
```
new_df1_target0=telco_data.loc[telco_data["Churn"]==0]
new_df1_target1=telco_data.loc[telco_data["Churn"]==1]
```

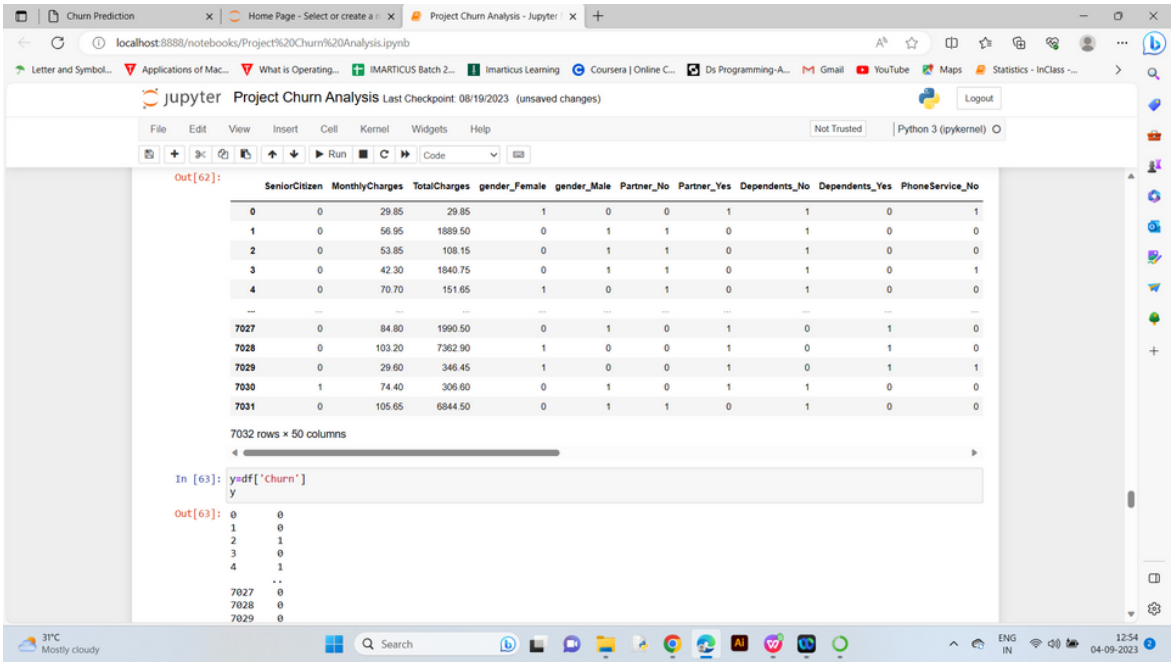
In [46]:
```
def uniplot(df,col,title,hue =None):

    sns.set_style('whitegrid')
    sns.set_context('talk')
    plt.rcParams["axes.labelsize"] = 20
    plt.rcParams['axes.titlesize'] = 22
    plt.rcParams['axes.titlepad'] = 30


    temp = pd.Series(data = hue)
    fig, ax = plt.subplots()
    width = len(df[col].unique()) + 7 + 4*len(temp.unique())
    fig.set_size_inches(width , 8)
    plt.xticks(rotation=45)
    plt.yscale('log')
    plt.title(title)
    ax = sns.countplot(data = df, x= col, order=df[col].value_counts().index,hue = hue,palette='bright')

    plt.show()
```

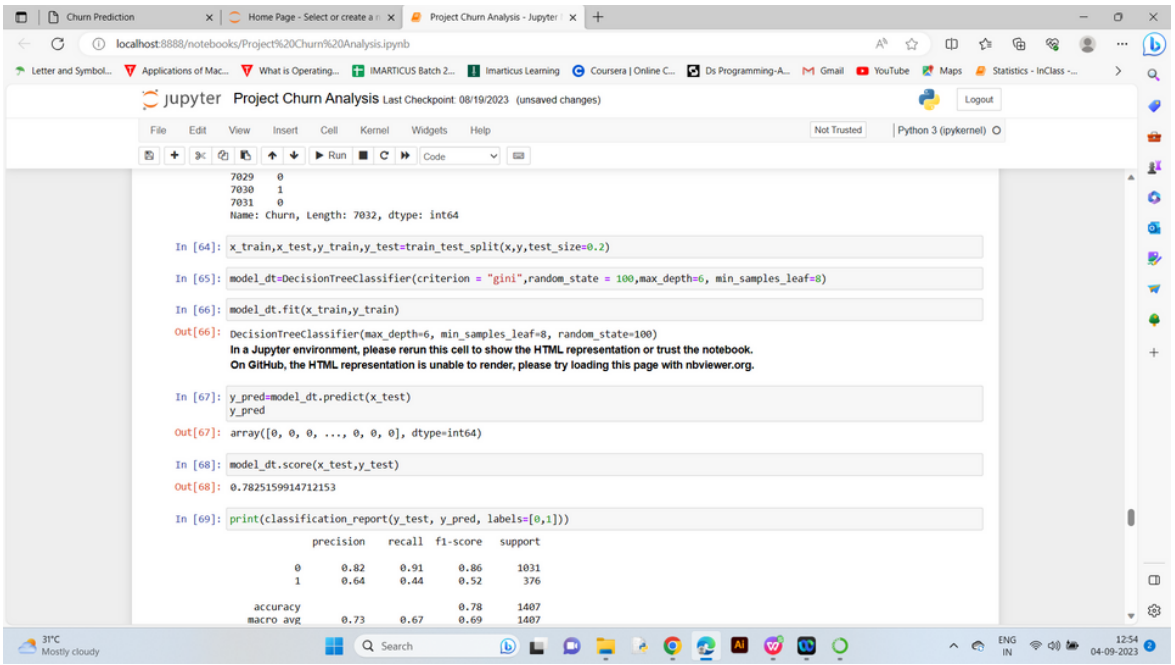
In [55]:
```
uniplot(new_df1_target1,col='Partner',title='Distribution of Gender for Churned Customers',hue='gender')
uniplot(new_df1_target0,col='Partner',title='Distribution of Gender for Non Churned Customers',hue='gender')
uniplot(new_df1_target1,col='PaymentMethod',title='Distribution of PaymentMethod for Churned Customers',hue='gender')
uniplot(new_df1_target1,col='Contract',title='Distribution of Contract for Churned Customers',hue='gender')
uniplot(new_df1_target1,col='TechSupport',title='Distribution of TechSupport for Churned Customers',hue='gender')
```

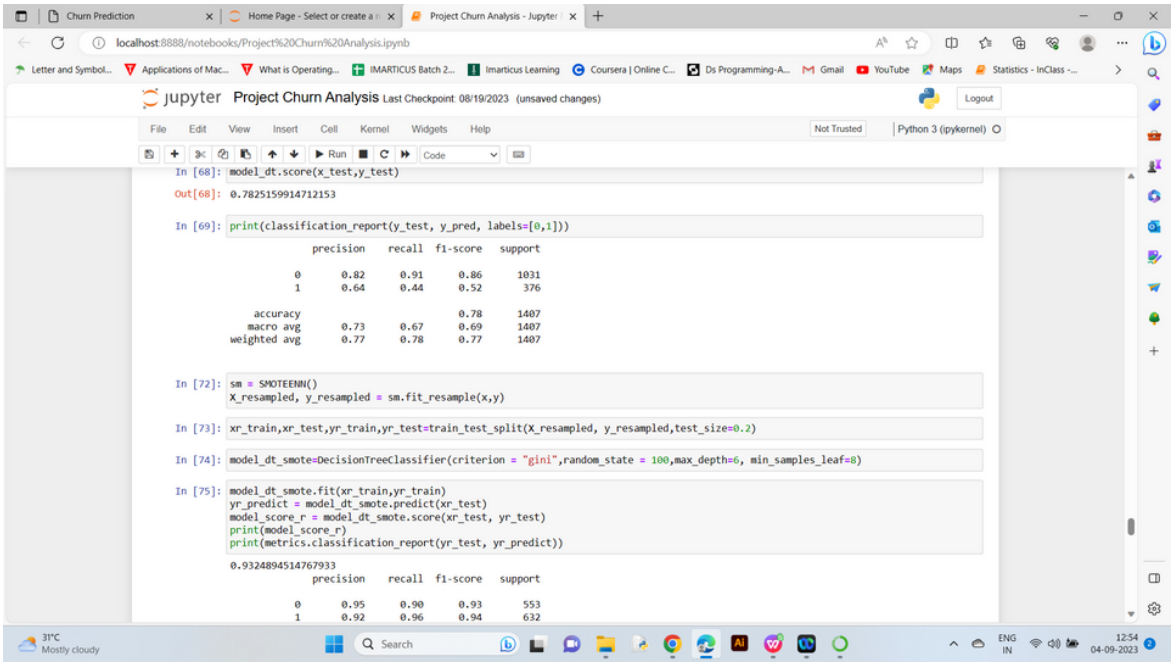


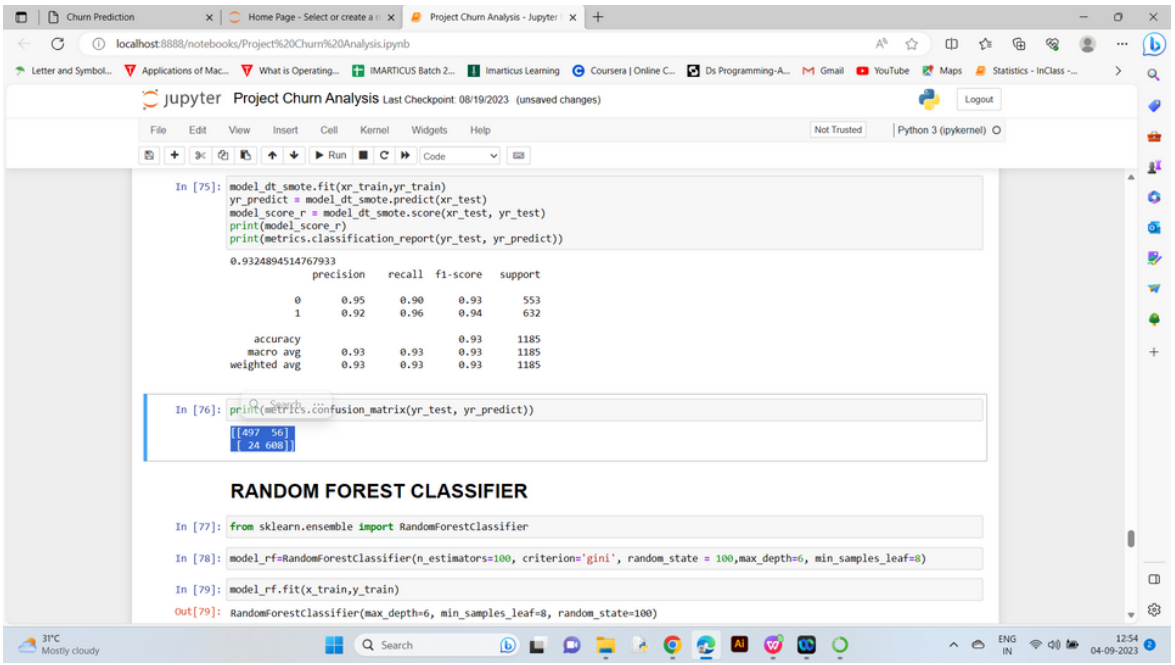
31°C Mostly cloudy | Search | ENG IN | 12:54 04-09-2023

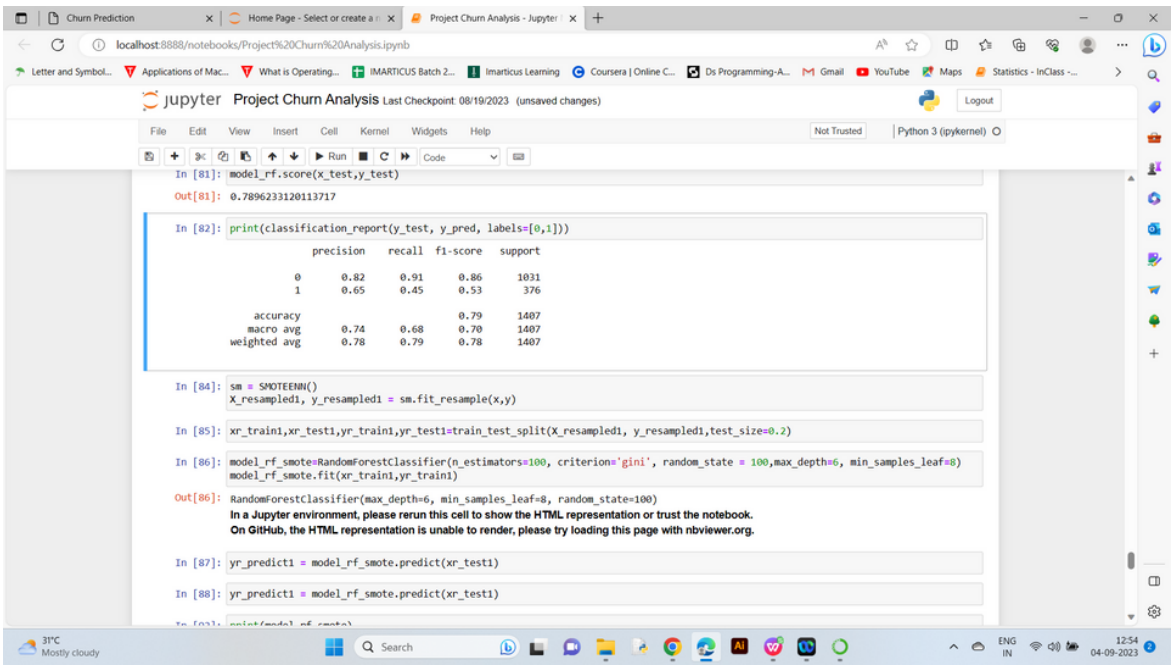

Churn Prediction
Home Page - Select or create a n
Project Churn Analysis - Jupyter
localhost:8888/notebooks/Project%20Churn%20Analysis.ipynb
jupyter Project Churn Analysis Last Checkpoint: 08/19/2023 (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
MODEL BUILDING PART:
In [57]:
import pandas as pd
from sklearn import metrics
from sklearn.model_selection import train_test_split
from sklearn.metrics import recall_score
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sklearn.tree import DecisionTreeClassifier
from imblearn.combine import SMOTEENN
In [60]:
df=pd.read_csv("C:/Users/DELL/Downloads/tel_churn.csv")
df.head()
Out[60]:
Unnamed: 0 SeniorCitizen MonthlyCharges TotalCharges Churn gender_Female gender_Male Partner_No Partner_Yes Dependents_No ... PaymentMethod_ transfer (auto
0 0 0 29.85 29.85 0 1 0 0 1 1 ...
1 1 0 56.95 1889.50 0 0 1 1 0 1 ...
2 2 0 53.85 108.15 1 0 1 1 0 1 ...
3 3 0 42.30 1840.75 0 0 1 1 0 1 ...
4 4 0 70.70 151.65 1 1 0 1 0 1 ...
5 rows × 52 columns
In [61]:
df=df.drop('Unnamed: 0',axis=1)
In [62]:
x=df.drop('Churn',axis=1)
31°C Mostly cloudy
Search
ENG IN
12:54 04-09-2023


Churn Prediction
Home Page - Select or create a n
Project Churn Analysis - Jupyter
localhost:8888/notebooks/Project%20Churn%20Analysis.ipynb
jupyter Project Churn Analysis Last Checkpoint: 08/19/2023 (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
Out[62]:
SeniorCitizen MonthlyCharges TotalCharges gender_Female gender_Male Partner_No Partner_Yes Dependents_No Dependents_Yes PhoneService_No
0 0 29.85 29.85 1 0 0 1 1 0 1
1 0 56.95 1889.50 0 1 1 0 1 0 0
2 0 53.85 108.15 0 1 1 0 1 0 0
3 0 42.30 1840.75 0 1 1 0 1 0 1
4 0 70.70 151.65 1 0 1 0 1 0 0
... ...
7027 0 84.80 1990.50 0 1 0 1 0 1 0
7028 0 103.20 7362.90 1 0 0 1 0 1 0
7029 0 29.60 346.45 1 0 0 1 0 1 1
7030 1 74.40 306.60 0 1 0 1 1 0 0
7031 0 105.65 6844.50 0 1 1 0 1 0 0
7032 rows × 50 columns
In [63]:
y=df['Churn']
y
Out[63]:
0 0
1 0
2 1
3 0
4 1
..
7027 0
7028 0
7029 0
31°C Mostly cloudy
Search
ENG IN
12:54 04-09-2023


Churn Prediction
Home Page - Select or create a n
Project Churn Analysis - Jupyter
localhost:8888/notebooks/Project%20Churn%20Analysis.ipynb
jupyter Project Churn Analysis Last Checkpoint: 08/19/2023 (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
7029 0
7030 1
7031 0
Name: Churn, Length: 7032, dtype: int64
In [64]:
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2)
In [65]:
model_dt=DecisionTreeClassifier(criterion = "gini",random_state = 100,max_depth=6, min_samples_leaf=8)
In [66]:
model_dt.fit(x_train,y_train)
Out[66]:
DecisionTreeClassifier(max_depth=6, min_samples_leaf=8, random_state=100)
In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.
In [67]:
y_pred=model_dt.predict(x_test)
y_pred
Out[67]:
array([0, 0, 0, ..., 0, 0, 0], dtype=int64)
In [68]:
model_dt.score(x_test,y_test)
Out[68]:
0.7825159914712153
In [69]:
print(classification_report(y_test, y_pred, labels=[0,1]))
precision recall f1-score support
0 0.82 0.91 0.86 1031
1 0.64 0.44 0.52 376
accuracy 0.78 1407
macro avg 0.73 0.67 0.69 1407
31°C Mostly cloudy
Search
ENG IN
12:54 04-09-2023


Churn Prediction
Home Page - Select or create a n
Project Churn Analysis - Jupyter
localhost:8888/notebooks/Project%20Churn%20Analysis.ipynb
jupyter Project Churn Analysis Last Checkpoint: 08/19/2023 (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
In [68]: model_dt.score(x_test,y_test)
Out[68]: 0.7825159914712153
In [69]: print(classification_report(y_test, y_pred, labels=[0,1]))
precision recall f1-score support
0 0.82 0.91 0.86 1031
1 0.64 0.44 0.52 376
accuracy 0.78 1407
macro avg 0.73 0.67 0.69 1407
weighted avg 0.77 0.78 0.77 1407
In [72]: sm = SMOTEENN()
X_resampled, y_resampled = sm.fit_resample(x,y)
In [73]: xr_train,xr_test,yr_train,yr_test=train_test_split(X_resampled, y_resampled,test_size=0.2)
In [74]: model_dt_smote=DecisionTreeClassifier(criterion = "gini",random_state = 100,max_depth=6, min_samples_leaf=8)
In [75]: model_dt_smote.fit(xr_train,yr_train)
yr_predict = model_dt_smote.predict(xr_test)
model_score_r = model_dt_smote.score(xr_test, yr_test)
print(model_score_r)
print(metrics.classification_report(yr_test, yr_predict))
0.9324894514767933
precision recall f1-score support
0 0.95 0.90 0.93 553
1 0.92 0.96 0.94 632
31°C
Mostly cloudy
Search
ENG
IN
12:54
04-09-2023


Churn Prediction
Home Page - Select or create a n
Project Churn Analysis - Jupyter
localhost:8888/notebooks/Project%20Churn%20Analysis.ipynb
jupyter Project Churn Analysis Last Checkpoint: 08/19/2023 (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
In [75]: model_dt_smote.fit(xr_train,yr_train)
yr_predict = model_dt_smote.predict(xr_test)
model_score_r = model_dt_smote.score(xr_test, yr_test)
print(model_score_r)
print(metrics.classification_report(yr_test, yr_predict))
0.9324894514767933
precision recall f1-score support
0 0.95 0.90 0.93 553
1 0.92 0.96 0.94 632
accuracy 0.93 1185
macro avg 0.93 0.93 0.93 1185
weighted avg 0.93 0.93 0.93 1185
In [76]: print(metrics.confusion_matrix(yr_test, yr_predict))
[[497 56]
[24 608]]
RANDOM FOREST CLASSIFIER
In [77]: from sklearn.ensemble import RandomForestClassifier
In [78]: model_rf=RandomForestClassifier(n_estimators=100, criterion='gini', random_state = 100,max_depth=6, min_samples_leaf=8)
In [79]: model_rf.fit(x_train,y_train)
Out[79]: RandomForestClassifier(max_depth=6, min_samples_leaf=8, random_state=100)
31°C
Mostly cloudy
Search
ENG
IN
12:54
04-09-2023


Churn Prediction
Home Page - Select or create a n
Project Churn Analysis - Jupyter
localhost:8888/notebooks/Project%20Churn%20Analysis.ipynb
jupyter Project Churn Analysis Last Checkpoint: 08/19/2023 (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3 (ipykernel)
Run
Code
In [81]: model_rf.score(x_test,y_test)
Out[81]: 0.7896233120113717
In [82]: print(classification_report(y_test, y_pred, labels=[0,1]))
precision recall f1-score support
0 0.82 0.91 0.86 1031
1 0.65 0.45 0.53 376
accuracy 0.79 1407
macro avg 0.74 0.68 0.70 1407
weighted avg 0.78 0.79 0.78 1407
In [84]: sm = SMOTEENN()
X_resampled1, y_resampled1 = sm.fit_resample(x,y)
In [85]: xr_train1,xr_test1,yr_train1,yr_test1=train_test_split(X_resampled1, y_resampled1,test_size=0.2)
In [86]: model_rf_smote=RandomForestClassifier(n_estimators=100, criterion='gini', random_state = 100,max_depth=6, min_samples_leaf=8)
model_rf_smote.fit(xr_train1,yr_train1)
Out[86]: RandomForestClassifier(max_depth=6, min_samples_leaf=8, random_state=100)
In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.
In [87]: yr_predict1 = model_rf_smote.predict(xr_test1)
In [88]: yr_predict1 = model_rf_smote.predict(xr_test1)
31°C
Mostly cloudy
Search
ENG
IN
12:54
04-09-2023

localhost:8888/notebooks/Project%20Churn%20Analysis.ipynb

Letter and Symbol... | Applications of Mac... | What is Operating... | IMARTICUS Batch 2... | Imarticus Learning | Coursera | Online C... | Ds Programming-A... | Gmail | YouTube | Maps | Statistics - InClass -...

jupyter **Project Churn Analysis** Last Checkpoint: 08/19/2023 (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Run Code

Out[86]:
```
RandomForestClassifier(max_depth=6, min_samples_leaf=8, random_state=100)
```
In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [87]:
```
yr_predict1 = model_rf_smote.predict(xr_test1)
```

In [88]:
```
yr_predict1 = model_rf_smote.predict(xr_test1)
```

In [92]:
```
print(model_rf_smote)
print(metrics.classification_report(yr_test1, yr_predict1))
```

```
RandomForestClassifier(max_depth=6, min_samples_leaf=8, random_state=100)
              precision    recall  f1-score   support

           0       0.95      0.89      0.92       522
           1       0.92      0.96      0.94       652

    accuracy                           0.93      1174
   macro avg       0.94      0.93      0.93      1174
weighted avg       0.93      0.93      0.93      1174
```

In [93]:
```
print(metrics.confusion_matrix(yr_test1, yr_predict1))
```

```
[[467  55]
 [ 24 628]]
```

PCA - PERFORMANCE

In [94]:
```
from sklearn.decomposition import PCA
pca = PCA(0.9)
```

31°C Mostly cloudy

Search

ENG IN 12:54 04-09-2023