Telecom_Churn_Model

June 8, 2019

1 Index

2 1.Data Import and Cleanup (index)

[35]:
```
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
from IPython.display import display # Allows the use of display() for␣
 ↪DataFrames
%matplotlib inline
```

2.0.1 1.1 Data Overview (index)

[36]:
```
#path to data source file
source_path = '/home/gino/Documents/Udacity/MLEnano/machine-learning-master/
↪projects/capstone/Final/WA_Fn-UseC_-Telco-Customer-Churn.csv'

raw_data = pd.read_csv(source_path)

print raw_data.shape
raw_data
```

```
(7043, 21)
```

[36]:

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	\
0	7590-VHVEG	Female	0	Yes	No	1	
1	5575-GNVDE	Male	0	No	No	34	
2	3668-QPYBK	Male	0	No	No	2	
3	7795-CFOCW	Male	0	No	No	45	
4	9237-HQITU	Female	0	No	No	2	
5	9305-CDSKC	Female	0	No	No	8	
6	1452-KIOVK	Male	0	No	Yes	22	
7	6713-OKOMC	Female	0	No	No	10	
8	7892-POOKP	Female	0	Yes	No	28	
9	6388-TABGU	Male	0	No	Yes	62	
10	9763-GRSKD	Male	0	Yes	Yes	13	
11	7469-LKBCI	Male	0	No	No	16	
12	8091-TTVAX	Male	0	Yes	No	58	
13	0280-XJGEX	Male	0	No	No	49	
14	5129-JLPIS	Male	0	No	No	25	
15	3655-SNQYZ	Female	0	Yes	Yes	69	
16	8191-XWSZG	Female	0	No	No	52	
17	9959-WOFKT	Male	0	No	Yes	71	
18	4190-MFLUW	Female	0	Yes	Yes	10	
19	4183-MYFRB	Female	0	No	No	21	
20	8779-QRDMV	Male	1	No	No	1	
21	1680-VDCWW	Male	0	Yes	No	12	
22	1066-JKSGK	Male	0	No	No	1	
23	3638-WEABW	Female	0	Yes	No	58	
24	6322-HRPFA	Male	0	Yes	Yes	49	
25	6865-JZNKO	Female	0	No	No	30	
26	6467-CHFZW	Male	0	Yes	Yes	47	
27	8665-UTDHZ	Male	0	Yes	Yes	1	
28	5248-YGIJN	Male	0	Yes	No	72	
29	8773-HHUOZ	Female	0	No	Yes	17	
...	...	...	...	...	...	...	
7013	1685-BQULA	Female	0	No	No	40	
7014	9053-EJUNL	Male	0	No	No	41	
7015	0666-UXTJO	Male	1	Yes	No	34	

```
7016  1471-GIQKQ  Female              0      No         No       1
7017  4807-IZYOZ  Female              0      No         No      51
7018  1122-JWTJW    Male              0     Yes        Yes       1
7019  9710-NJERN  Female              0      No         No      39
7020  9837-FWLCH    Male              0     Yes        Yes      12
7021  1699-HPSBG    Male              0      No         No      12
7022  7203-OYKCT    Male              0      No         No      72
7023  1035-IPQPU  Female              1     Yes         No      63
7024  7398-LXGYX    Male              0     Yes         No      44
7025  2823-LKABH  Female              0      No         No      18
7026  8775-CEBBJ  Female              0      No         No       9
7027  0550-DCXLH    Male              0      No         No      13
7028  9281-CEDRU  Female              0     Yes         No      68
7029  2235-DWLJU  Female              1      No         No       6
7030  0871-OPBXW  Female              0      No         No       2
7031  3605-JISKB    Male              1     Yes         No      55
7032  6894-LFHLY    Male              1      No         No       1
7033  9767-FFLEM    Male              0      No         No      38
7034  0639-TSIQW  Female              0      No         No      67
7035  8456-QDAVC    Male              0      No         No      19
7036  7750-EYXWZ  Female              0      No         No      12
7037  2569-WGERO  Female              0      No         No      72
7038  6840-RESVB    Male              0     Yes        Yes      24
7039  2234-XADUH  Female              0     Yes        Yes      72
7040  4801-JZAZL  Female              0     Yes        Yes      11
7041  8361-LTMKD    Male              1     Yes         No       4
7042  3186-AJIEK    Male              0      No         No      66

     PhoneService     MultipleLines InternetService       OnlineSecurity  \
0              No  No phone service             DSL                   No
1             Yes                No             DSL                  Yes
2             Yes                No             DSL                  Yes
3              No  No phone service             DSL                  Yes
4             Yes                No     Fiber optic                   No
5             Yes               Yes     Fiber optic                   No
6             Yes               Yes     Fiber optic                   No
7              No  No phone service             DSL                  Yes
8             Yes               Yes     Fiber optic                   No
9             Yes                No             DSL                  Yes
10            Yes                No             DSL                  Yes
11            Yes                No              No  No internet service
12            Yes               Yes     Fiber optic                   No
13            Yes               Yes     Fiber optic                   No
14            Yes                No     Fiber optic                  Yes
15            Yes               Yes     Fiber optic                  Yes
16            Yes                No              No  No internet service
17            Yes               Yes     Fiber optic                  Yes
```

```
18          Yes                No          DSL                   No
19          Yes                No  Fiber optic                   No
20           No  No phone service          DSL                   No
21          Yes                No           No  No internet service
22          Yes                No           No  No internet service
23          Yes               Yes          DSL                   No
24          Yes                No          DSL                  Yes
25          Yes                No          DSL                  Yes
26          Yes               Yes  Fiber optic                   No
27           No  No phone service          DSL                   No
28          Yes               Yes          DSL                  Yes
29          Yes                No          DSL                   No
...         ...               ...          ...                  ...
7013        Yes               Yes  Fiber optic                   No
7014        Yes               Yes  Fiber optic                   No
7015        Yes                No  Fiber optic                   No
7016        Yes                No          DSL                   No
7017        Yes                No           No  No internet service
7018        Yes                No  Fiber optic                   No
7019        Yes                No           No  No internet service
7020        Yes                No           No  No internet service
7021        Yes                No          DSL                   No
7022        Yes               Yes  Fiber optic                   No
7023        Yes               Yes  Fiber optic                   No
7024        Yes               Yes  Fiber optic                  Yes
7025        Yes               Yes  Fiber optic                   No
7026        Yes                No          DSL                   No
7027        Yes                No          DSL                   No
7028        Yes                No          DSL                   No
7029         No  No phone service          DSL                   No
7030        Yes                No           No  No internet service
7031        Yes               Yes          DSL                  Yes
7032        Yes               Yes  Fiber optic                   No
7033        Yes                No  Fiber optic                   No
7034        Yes               Yes  Fiber optic                  Yes
7035        Yes                No  Fiber optic                   No
7036         No  No phone service          DSL                   No
7037        Yes                No           No  No internet service
7038        Yes               Yes          DSL                  Yes
7039        Yes               Yes  Fiber optic                   No
7040         No  No phone service          DSL                  Yes
7041        Yes               Yes  Fiber optic                   No
7042        Yes                No  Fiber optic                  Yes

      ...     DeviceProtection          TechSupport          StreamingTV  \
0     ...                   No                   No                   No
1     ...                  Yes                   No                   No
```

2	...	No	No	No
3	...	Yes	Yes	No
4	...	No	No	No
5	...	Yes	No	Yes
6	...	No	No	Yes
7	...	No	No	No
8	...	Yes	Yes	Yes
9	...	No	No	No
10	...	No	No	No
11	...	No internet service	No internet service	No internet service
12	...	Yes	No	Yes
13	...	Yes	No	Yes
14	...	Yes	Yes	Yes
15	...	Yes	Yes	Yes
16	...	No internet service	No internet service	No internet service
17	...	Yes	No	Yes
18	...	Yes	Yes	No
19	...	Yes	No	No
20	...	Yes	No	No
21	...	No internet service	No internet service	No internet service
22	...	No internet service	No internet service	No internet service
23	...	No	Yes	No
24	...	No	Yes	No
25	...	No	No	No
26	...	No	No	Yes
27	...	No	No	No
28	...	Yes	Yes	Yes
29	...	No	No	Yes
...	...	...	...	...
7013	...	Yes	No	Yes
7014	...	No	No	Yes
7015	...	Yes	No	Yes
7016	...	No	No	No
7017	...	No internet service	No internet service	No internet service
7018	...	No	No	No
7019	...	No internet service	No internet service	No internet service
7020	...	No internet service	No internet service	No internet service
7021	...	No	Yes	Yes
7022	...	Yes	No	Yes
7023	...	Yes	No	Yes
7024	...	Yes	No	No
7025	...	Yes	Yes	No
7026	...	No	No	No
7027	...	No	Yes	Yes
7028	...	No	Yes	Yes
7029	...	No	No	Yes
7030	...	No internet service	No internet service	No internet service

```
7031  ...                   No                   No                   No
7032  ...                   No                   No                   No
7033  ...                   No                   No                   No
7034  ...                  Yes                   No                  Yes
7035  ...                   No                   No                  Yes
7036  ...                  Yes                  Yes                  Yes
7037  ...  No internet service  No internet service  No internet service
7038  ...                  Yes                  Yes                  Yes
7039  ...                  Yes                   No                  Yes
7040  ...                   No                   No                   No
7041  ...                   No                   No                   No
7042  ...                  Yes                  Yes                  Yes

          StreamingMovies        Contract PaperlessBilling  \
0                      No  Month-to-month              Yes
1                      No        One year               No
2                      No  Month-to-month              Yes
3                      No        One year               No
4                      No  Month-to-month              Yes
5                     Yes  Month-to-month              Yes
6                      No  Month-to-month              Yes
7                      No  Month-to-month               No
8                     Yes  Month-to-month              Yes
9                      No        One year               No
10                     No  Month-to-month              Yes
11    No internet service        Two year               No
12                    Yes        One year               No
13                    Yes  Month-to-month              Yes
14                    Yes  Month-to-month              Yes
15                    Yes        Two year               No
16    No internet service        One year               No
17                    Yes        Two year               No
18                     No  Month-to-month               No
19                    Yes  Month-to-month              Yes
20                    Yes  Month-to-month              Yes
21    No internet service        One year               No
22    No internet service  Month-to-month               No
23                     No        Two year              Yes
24                     No  Month-to-month               No
25                     No  Month-to-month              Yes
26                    Yes  Month-to-month              Yes
27                     No  Month-to-month               No
28                    Yes        Two year              Yes
29                    Yes  Month-to-month              Yes
...                   ...             ...              ...
7013                   No  Month-to-month              Yes
7014                   No  Month-to-month              Yes
```

7015	No	Month-to-month	Yes
7016	No	Month-to-month	No
7017	No internet service	Two year	No
7018	No	Month-to-month	Yes
7019	No internet service	Two year	No
7020	No internet service	Month-to-month	Yes
7021	No	One year	Yes
7022	Yes	One year	Yes
7023	Yes	Month-to-month	Yes
7024	No	Month-to-month	Yes
7025	Yes	Month-to-month	Yes
7026	No	Month-to-month	Yes
7027	Yes	Month-to-month	No
7028	No	Two year	No
7029	Yes	Month-to-month	Yes
7030	No internet service	Month-to-month	Yes
7031	No	One year	No
7032	No	Month-to-month	Yes
7033	No	Month-to-month	Yes
7034	No	Month-to-month	Yes
7035	No	Month-to-month	Yes
7036	Yes	One year	No
7037	No internet service	Two year	Yes
7038	Yes	One year	Yes
7039	Yes	One year	Yes
7040	No	Month-to-month	Yes
7041	No	Month-to-month	Yes
7042	Yes	Two year	Yes

	PaymentMethod	MonthlyCharges	TotalCharges	Churn
0	Electronic check	29.85	29.85	No
1	Mailed check	56.95	1889.5	No
2	Mailed check	53.85	108.15	Yes
3	Bank transfer (automatic)	42.30	1840.75	No
4	Electronic check	70.70	151.65	Yes
5	Electronic check	99.65	820.5	Yes
6	Credit card (automatic)	89.10	1949.4	No
7	Mailed check	29.75	301.9	No
8	Electronic check	104.80	3046.05	Yes
9	Bank transfer (automatic)	56.15	3487.95	No
10	Mailed check	49.95	587.45	No
11	Credit card (automatic)	18.95	326.8	No
12	Credit card (automatic)	100.35	5681.1	No
13	Bank transfer (automatic)	103.70	5036.3	Yes
14	Electronic check	105.50	2686.05	No
15	Credit card (automatic)	113.25	7895.15	No
16	Mailed check	20.65	1022.95	No

17	Bank transfer (automatic)	106.70	7382.25	No
18	Credit card (automatic)	55.20	528.35	Yes
19	Electronic check	90.05	1862.9	No
20	Electronic check	39.65	39.65	Yes
21	Bank transfer (automatic)	19.80	202.25	No
22	Mailed check	20.15	20.15	Yes
23	Credit card (automatic)	59.90	3505.1	No
24	Credit card (automatic)	59.60	2970.3	No
25	Bank transfer (automatic)	55.30	1530.6	No
26	Electronic check	99.35	4749.15	Yes
27	Electronic check	30.20	30.2	Yes
28	Credit card (automatic)	90.25	6369.45	No
29	Mailed check	64.70	1093.1	Yes
...	...	...	...	...
7013	Bank transfer (automatic)	93.40	3756.4	No
7014	Electronic check	89.20	3645.75	No
7015	Credit card (automatic)	85.20	2874.45	No
7016	Electronic check	49.95	49.95	No
7017	Bank transfer (automatic)	20.65	1020.75	No
7018	Mailed check	70.65	70.65	Yes
7019	Mailed check	20.15	826	No
7020	Electronic check	19.20	239	No
7021	Electronic check	59.80	727.8	Yes
7022	Electronic check	104.95	7544.3	No
7023	Electronic check	103.50	6479.4	No
7024	Credit card (automatic)	84.80	3626.35	No
7025	Bank transfer (automatic)	95.05	1679.4	No
7026	Bank transfer (automatic)	44.20	403.35	Yes
7027	Mailed check	73.35	931.55	No
7028	Bank transfer (automatic)	64.10	4326.25	No
7029	Electronic check	44.40	263.05	No
7030	Mailed check	20.05	39.25	No
7031	Credit card (automatic)	60.00	3316.1	No
7032	Electronic check	75.75	75.75	Yes
7033	Credit card (automatic)	69.50	2625.25	No
7034	Credit card (automatic)	102.95	6886.25	Yes
7035	Bank transfer (automatic)	78.70	1495.1	No
7036	Electronic check	60.65	743.3	No
7037	Bank transfer (automatic)	21.15	1419.4	No
7038	Mailed check	84.80	1990.5	No
7039	Credit card (automatic)	103.20	7362.9	No
7040	Electronic check	29.60	346.45	No
7041	Mailed check	74.40	306.6	Yes
7042	Bank transfer (automatic)	105.65	6844.5	No

[7043 rows x 21 columns]

[3]:
```
raw_data.describe()
```

[3]:

	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

3 2.Data Manipulation (index)

3.0.1 Blanks and NA's in the data

[4]:
```
#Look for NA's, missing datapoints.
display(raw_data.isna().any())

#count of all non-NA values
display(raw_data.count())
```

```
customerID          False
gender              False
SeniorCitizen       False
Partner             False
Dependents          False
tenure              False
PhoneService        False
MultipleLines       False
InternetService     False
OnlineSecurity      False
OnlineBackup        False
DeviceProtection    False
TechSupport         False
StreamingTV         False
StreamingMovies     False
Contract            False
PaperlessBilling    False
PaymentMethod       False
MonthlyCharges      False
TotalCharges        False
Churn               False
dtype: bool
```

```
customerID          7043
```

```
gender              7043
SeniorCitizen       7043
Partner             7043
Dependents          7043
tenure              7043
PhoneService        7043
MultipleLines       7043
InternetService     7043
OnlineSecurity      7043
OnlineBackup        7043
DeviceProtection    7043
TechSupport         7043
StreamingTV         7043
StreamingMovies     7043
Contract            7043
PaperlessBilling    7043
PaymentMethod       7043
MonthlyCharges      7043
TotalCharges        7043
Churn               7043
dtype: int64
```

[5]:
```
raw_data.columns
```

[5]:
```
Index([u'customerID', u'gender', u'SeniorCitizen', u'Partner', u'Dependents',
       u'tenure', u'PhoneService', u'MultipleLines', u'InternetService',
       u'OnlineSecurity', u'OnlineBackup', u'DeviceProtection', u'TechSupport',
       u'StreamingTV', u'StreamingMovies', u'Contract', u'PaperlessBilling',
       u'PaymentMethod', u'MonthlyCharges', u'TotalCharges', u'Churn'],
      dtype='object')
```

[6]:
```
#Find indexes of empty cells
blanks = np.where(raw_data.applymap(lambda x: x==' '))
print np.array(blanks).reshape(2, len(blanks[0]))

#11 blanks on column 19, Total Charges
#drop blank rows
raw_data.drop(blanks[0], inplace = True)
raw_data.reset_index(drop=True, inplace = True)

#drop customerID column
raw_data.drop('customerID', axis=1, inplace = True)

#change data types on some columns
cleaned_data = raw_data.astype({'TotalCharges':'float64', 'tenure':'int64',␣
 ↪'SeniorCitizen':'object',
        'gender':'object', 'SeniorCitizen':'int64', 'Partner':'object',␣
 ↪'Dependents':'object',
```

```
'PhoneService':'object', 'MultipleLines':'object', 'InternetService':
↪'object',
      'OnlineSecurity':'object', 'OnlineBackup':'object', 'DeviceProtection':
↪'object', 'TechSupport':'object',
      'StreamingTV':'object', 'StreamingMovies':'object', 'Contract':'object',␣
↪'PaperlessBilling':'object',
      'PaymentMethod':'object', 'Churn':'category'})

cleaned_data.info()
cleaned_data.describe()
cleaned_data
```

```
[[ 488  753  936 1082 1340 3331 3826 4380 5218 6670 6754]
 [  19   19   19   19   19   19   19   19   19   19   19]]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7032 entries, 0 to 7031
Data columns (total 20 columns):
gender              7032 non-null object
SeniorCitizen       7032 non-null int64
Partner             7032 non-null object
Dependents          7032 non-null object
tenure              7032 non-null int64
PhoneService        7032 non-null object
MultipleLines       7032 non-null object
InternetService     7032 non-null object
OnlineSecurity      7032 non-null object
OnlineBackup        7032 non-null object
DeviceProtection    7032 non-null object
TechSupport         7032 non-null object
StreamingTV         7032 non-null object
StreamingMovies     7032 non-null object
Contract            7032 non-null object
PaperlessBilling    7032 non-null object
PaymentMethod       7032 non-null object
MonthlyCharges      7032 non-null float64
TotalCharges        7032 non-null float64
Churn               7032 non-null category
dtypes: category(1), float64(2), int64(2), object(15)
memory usage: 1.0+ MB
```

[6]:

```
      gender  SeniorCitizen Partner Dependents  tenure PhoneService  \
0     Female              0     Yes         No       1           No
1       Male              0      No         No      34          Yes
2       Male              0      No         No       2          Yes
3       Male              0      No         No      45           No
4     Female              0      No         No       2          Yes
5     Female              0      No         No       8          Yes
```

6	Male	0	No	Yes	22	Yes
7	Female	0	No	No	10	No
8	Female	0	Yes	No	28	Yes
9	Male	0	No	Yes	62	Yes
10	Male	0	Yes	Yes	13	Yes
11	Male	0	No	No	16	Yes
12	Male	0	Yes	No	58	Yes
13	Male	0	No	No	49	Yes
14	Male	0	No	No	25	Yes
15	Female	0	Yes	Yes	69	Yes
16	Female	0	No	No	52	Yes
17	Male	0	No	Yes	71	Yes
18	Female	0	Yes	Yes	10	Yes
19	Female	0	No	No	21	Yes
20	Male	1	No	No	1	No
21	Male	0	Yes	No	12	Yes
22	Male	0	No	No	1	Yes
23	Female	0	Yes	No	58	Yes
24	Male	0	Yes	Yes	49	Yes
25	Female	0	No	No	30	Yes
26	Male	0	Yes	Yes	47	Yes
27	Male	0	Yes	Yes	1	No
28	Male	0	Yes	No	72	Yes
29	Female	0	No	Yes	17	Yes
...	...	...	...	...	...	...
7002	Female	0	No	No	40	Yes
7003	Male	0	No	No	41	Yes
7004	Male	1	Yes	No	34	Yes
7005	Female	0	No	No	1	Yes
7006	Female	0	No	No	51	Yes
7007	Male	0	Yes	Yes	1	Yes
7008	Female	0	No	No	39	Yes
7009	Male	0	Yes	Yes	12	Yes
7010	Male	0	No	No	12	Yes
7011	Male	0	No	No	72	Yes
7012	Female	1	Yes	No	63	Yes
7013	Male	0	Yes	No	44	Yes
7014	Female	0	No	No	18	Yes
7015	Female	0	No	No	9	Yes
7016	Male	0	No	No	13	Yes
7017	Female	0	Yes	No	68	Yes
7018	Female	1	No	No	6	No
7019	Female	0	No	No	2	Yes
7020	Male	1	Yes	No	55	Yes
7021	Male	1	No	No	1	Yes
7022	Male	0	No	No	38	Yes
7023	Female	0	No	No	67	Yes

```
7024    Male              0      No         No      19           Yes
7025  Female              0      No         No      12            No
7026  Female              0      No         No      72           Yes
7027    Male              0     Yes        Yes      24           Yes
7028  Female              0     Yes        Yes      72           Yes
7029  Female              0     Yes        Yes      11            No
7030    Male              1     Yes         No       4           Yes
7031    Male              0      No         No      66           Yes

         MultipleLines InternetService       OnlineSecurity  \
0     No phone service             DSL                   No
1                   No             DSL                  Yes
2                   No             DSL                  Yes
3     No phone service             DSL                  Yes
4                   No     Fiber optic                   No
5                  Yes     Fiber optic                   No
6                  Yes     Fiber optic                   No
7     No phone service             DSL                  Yes
8                  Yes     Fiber optic                   No
9                   No             DSL                  Yes
10                  No             DSL                  Yes
11                  No              No  No internet service
12                 Yes     Fiber optic                   No
13                 Yes     Fiber optic                   No
14                  No     Fiber optic                  Yes
15                 Yes     Fiber optic                  Yes
16                  No              No  No internet service
17                 Yes     Fiber optic                  Yes
18                  No             DSL                   No
19                  No     Fiber optic                   No
20    No phone service             DSL                   No
21                  No              No  No internet service
22                  No              No  No internet service
23                 Yes             DSL                   No
24                  No             DSL                  Yes
25                  No             DSL                  Yes
26                 Yes     Fiber optic                   No
27    No phone service             DSL                   No
28                 Yes             DSL                  Yes
29                  No             DSL                   No
...                ...             ...                  ...
7002               Yes     Fiber optic                   No
7003               Yes     Fiber optic                   No
7004                No     Fiber optic                   No
7005                No             DSL                   No
7006                No              No  No internet service
7007                No     Fiber optic                   No
```

```
7008                No              No  No internet service
7009                No              No  No internet service
7010                No             DSL                   No
7011               Yes     Fiber optic                   No
7012               Yes     Fiber optic                   No
7013               Yes     Fiber optic                  Yes
7014               Yes     Fiber optic                   No
7015                No             DSL                   No
7016                No             DSL                   No
7017                No             DSL                   No
7018  No phone service             DSL                   No
7019                No              No  No internet service
7020               Yes             DSL                  Yes
7021               Yes     Fiber optic                   No
7022                No     Fiber optic                   No
7023               Yes     Fiber optic                  Yes
7024                No     Fiber optic                   No
7025  No phone service             DSL                   No
7026                No              No  No internet service
7027               Yes             DSL                  Yes
7028               Yes     Fiber optic                   No
7029  No phone service             DSL                  Yes
7030               Yes     Fiber optic                   No
7031                No     Fiber optic                  Yes

             OnlineBackup     DeviceProtection          TechSupport  \
0                     Yes                   No                   No
1                      No                  Yes                   No
2                     Yes                   No                   No
3                      No                  Yes                  Yes
4                      No                   No                   No
5                      No                  Yes                   No
6                     Yes                   No                   No
7                      No                   No                   No
8                      No                  Yes                  Yes
9                     Yes                   No                   No
10                     No                   No                   No
11    No internet service  No internet service  No internet service
12                     No                  Yes                   No
13                    Yes                  Yes                   No
14                     No                  Yes                  Yes
15                    Yes                  Yes                  Yes
16    No internet service  No internet service  No internet service
17                     No                  Yes                   No
18                     No                  Yes                  Yes
19                    Yes                  Yes                   No
20                     No                  Yes                   No
```

```
21    No internet service  No internet service  No internet service
22    No internet service  No internet service  No internet service
23                    Yes                   No                  Yes
24                    Yes                   No                  Yes
25                    Yes                   No                   No
26                    Yes                   No                   No
27                    Yes                   No                   No
28                    Yes                  Yes                  Yes
29                     No                   No                   No
...                   ...                  ...                  ...
7002                  Yes                  Yes                   No
7003                  Yes                   No                   No
7004                   No                  Yes                   No
7005                  Yes                   No                   No
7006  No internet service  No internet service  No internet service
7007                   No                   No                   No
7008  No internet service  No internet service  No internet service
7009  No internet service  No internet service  No internet service
7010                   No                   No                  Yes
7011                  Yes                  Yes                   No
7012                  Yes                  Yes                   No
7013                   No                  Yes                   No
7014                   No                  Yes                  Yes
7015                   No                   No                   No
7016                  Yes                   No                  Yes
7017                  Yes                   No                  Yes
7018                   No                   No                   No
7019  No internet service  No internet service  No internet service
7020                  Yes                   No                   No
7021                   No                   No                   No
7022                   No                   No                   No
7023                  Yes                  Yes                   No
7024                   No                   No                   No
7025                  Yes                  Yes                  Yes
7026  No internet service  No internet service  No internet service
7027                   No                  Yes                  Yes
7028                  Yes                  Yes                   No
7029                   No                   No                   No
7030                   No                   No                   No
7031                   No                  Yes                  Yes

              StreamingTV      StreamingMovies        Contract  \
0                      No                   No  Month-to-month
1                      No                   No        One year
2                      No                   No  Month-to-month
3                      No                   No        One year
4                      No                   No  Month-to-month
```

5	Yes	Yes	Month-to-month
6	Yes	No	Month-to-month
7	No	No	Month-to-month
8	Yes	Yes	Month-to-month
9	No	No	One year
10	No	No	Month-to-month
11	No internet service	No internet service	Two year
12	Yes	Yes	One year
13	Yes	Yes	Month-to-month
14	Yes	Yes	Month-to-month
15	Yes	Yes	Two year
16	No internet service	No internet service	One year
17	Yes	Yes	Two year
18	No	No	Month-to-month
19	No	Yes	Month-to-month
20	No	Yes	Month-to-month
21	No internet service	No internet service	One year
22	No internet service	No internet service	Month-to-month
23	No	No	Two year
24	No	No	Month-to-month
25	No	No	Month-to-month
26	Yes	Yes	Month-to-month
27	No	No	Month-to-month
28	Yes	Yes	Two year
29	Yes	Yes	Month-to-month
...	...	...	...
7002	Yes	No	Month-to-month
7003	Yes	No	Month-to-month
7004	Yes	No	Month-to-month
7005	No	No	Month-to-month
7006	No internet service	No internet service	Two year
7007	No	No	Month-to-month
7008	No internet service	No internet service	Two year
7009	No internet service	No internet service	Month-to-month
7010	Yes	No	One year
7011	Yes	Yes	One year
7012	Yes	Yes	Month-to-month
7013	No	No	Month-to-month
7014	No	Yes	Month-to-month
7015	No	No	Month-to-month
7016	Yes	Yes	Month-to-month
7017	Yes	No	Two year
7018	Yes	Yes	Month-to-month
7019	No internet service	No internet service	Month-to-month
7020	No	No	One year
7021	No	No	Month-to-month
7022	No	No	Month-to-month

```
7023                  Yes                  No  Month-to-month
7024                  Yes                  No  Month-to-month
7025                  Yes                 Yes        One year
7026  No internet service  No internet service        Two year
7027                  Yes                 Yes        One year
7028                  Yes                 Yes        One year
7029                   No                  No  Month-to-month
7030                   No                  No  Month-to-month
7031                  Yes                 Yes        Two year

     PaperlessBilling              PaymentMethod  MonthlyCharges  \
0                 Yes           Electronic check           29.85
1                  No               Mailed check           56.95
2                 Yes               Mailed check           53.85
3                  No  Bank transfer (automatic)           42.30
4                 Yes           Electronic check           70.70
5                 Yes           Electronic check           99.65
6                 Yes    Credit card (automatic)           89.10
7                  No               Mailed check           29.75
8                 Yes           Electronic check          104.80
9                  No  Bank transfer (automatic)           56.15
10                Yes               Mailed check           49.95
11                 No    Credit card (automatic)           18.95
12                 No    Credit card (automatic)          100.35
13                Yes  Bank transfer (automatic)          103.70
14                Yes           Electronic check          105.50
15                 No    Credit card (automatic)          113.25
16                 No               Mailed check           20.65
17                 No  Bank transfer (automatic)          106.70
18                 No    Credit card (automatic)           55.20
19                Yes           Electronic check           90.05
20                Yes           Electronic check           39.65
21                 No  Bank transfer (automatic)           19.80
22                 No               Mailed check           20.15
23                Yes    Credit card (automatic)           59.90
24                 No    Credit card (automatic)           59.60
25                Yes  Bank transfer (automatic)           55.30
26                Yes           Electronic check           99.35
27                 No           Electronic check           30.20
28                Yes    Credit card (automatic)           90.25
29                Yes               Mailed check           64.70
...               ...                        ...             ...
7002              Yes  Bank transfer (automatic)           93.40
7003              Yes           Electronic check           89.20
7004              Yes    Credit card (automatic)           85.20
7005               No           Electronic check           49.95
7006               No  Bank transfer (automatic)           20.65
```

```
7007                Yes              Mailed check           70.65
7008                 No              Mailed check           20.15
7009                Yes          Electronic check           19.20
7010                Yes          Electronic check           59.80
7011                Yes          Electronic check          104.95
7012                Yes          Electronic check          103.50
7013                Yes   Credit card (automatic)           84.80
7014                Yes Bank transfer (automatic)           95.05
7015                Yes Bank transfer (automatic)           44.20
7016                 No              Mailed check           73.35
7017                 No Bank transfer (automatic)           64.10
7018                Yes          Electronic check           44.40
7019                Yes              Mailed check           20.05
7020                 No   Credit card (automatic)           60.00
7021                Yes          Electronic check           75.75
7022                Yes   Credit card (automatic)           69.50
7023                Yes   Credit card (automatic)          102.95
7024                Yes Bank transfer (automatic)           78.70
7025                 No          Electronic check           60.65
7026                Yes Bank transfer (automatic)           21.15
7027                Yes              Mailed check           84.80
7028                Yes   Credit card (automatic)          103.20
7029                Yes          Electronic check           29.60
7030                Yes              Mailed check           74.40
7031                Yes Bank transfer (automatic)          105.65

      TotalCharges Churn
0            29.85    No
1          1889.50    No
2           108.15   Yes
3          1840.75    No
4           151.65   Yes
5           820.50   Yes
6          1949.40    No
7           301.90    No
8          3046.05   Yes
9          3487.95    No
10          587.45    No
11          326.80    No
12         5681.10    No
13         5036.30   Yes
14         2686.05    No
15         7895.15    No
16         1022.95    No
17         7382.25    No
18          528.35   Yes
19         1862.90    No
```

```
20          39.65   Yes
21         202.25    No
22          20.15   Yes
23        3505.10    No
24        2970.30    No
25        1530.60    No
26        4749.15   Yes
27          30.20   Yes
28        6369.45    No
29        1093.10   Yes
...           ...   ...
7002      3756.40    No
7003      3645.75    No
7004      2874.45    No
7005        49.95    No
7006      1020.75    No
7007        70.65   Yes
7008       826.00    No
7009       239.00    No
7010       727.80   Yes
7011      7544.30    No
7012      6479.40    No
7013      3626.35    No
7014      1679.40    No
7015       403.35   Yes
7016       931.55    No
7017      4326.25    No
7018       263.05    No
7019        39.25    No
7020      3316.10    No
7021        75.75   Yes
7022      2625.25    No
7023      6886.25   Yes
7024      1495.10    No
7025       743.30    No
7026      1419.40    No
7027      1990.50    No
7028      7362.90    No
7029       346.45    No
7030       306.60   Yes
7031      6844.50    No

[7032 rows x 20 columns]
```

[7]:

```
#Simplify columns by replacing 'No Internet Service' to 'No' for
 ↪'OnlineSecurity', 'OnlineBackup',
# 'DeviceProtection','TechSupport','StreamingTV','StreamingMovies'
```

```
cols = ['OnlineSecurity', 'OnlineBackup',␣
↪'DeviceProtection','TechSupport','StreamingTV', 'StreamingMovies']

for i in cols:
    cleaned_data[i] = cleaned_data[i].replace({'No internet service':'No'})

cleaned_data
```

[7]:

```
      gender  SeniorCitizen Partner Dependents  tenure PhoneService  \
0     Female              0     Yes         No       1           No
1       Male              0      No         No      34          Yes
2       Male              0      No         No       2          Yes
3       Male              0      No         No      45           No
4     Female              0      No         No       2          Yes
5     Female              0      No         No       8          Yes
6       Male              0      No        Yes      22          Yes
7     Female              0      No         No      10           No
8     Female              0     Yes         No      28          Yes
9       Male              0      No        Yes      62          Yes
10      Male              0     Yes        Yes      13          Yes
11      Male              0      No         No      16          Yes
12      Male              0     Yes         No      58          Yes
13      Male              0      No         No      49          Yes
14      Male              0      No         No      25          Yes
15    Female              0     Yes        Yes      69          Yes
16    Female              0      No         No      52          Yes
17      Male              0      No        Yes      71          Yes
18    Female              0     Yes        Yes      10          Yes
19    Female              0      No         No      21          Yes
20      Male              1      No         No       1           No
21      Male              0     Yes         No      12          Yes
22      Male              0      No         No       1          Yes
23    Female              0     Yes         No      58          Yes
24      Male              0     Yes        Yes      49          Yes
25    Female              0      No         No      30          Yes
26      Male              0     Yes        Yes      47          Yes
27      Male              0     Yes        Yes       1           No
28      Male              0     Yes         No      72          Yes
29    Female              0      No        Yes      17          Yes
...      ...            ...     ...        ...     ...          ...
7002  Female              0      No         No      40          Yes
7003    Male              0      No         No      41          Yes
7004    Male              1     Yes         No      34          Yes
7005  Female              0      No         No       1          Yes
7006  Female              0      No         No      51          Yes
7007    Male              0     Yes        Yes       1          Yes
7008  Female              0      No         No      39          Yes
```

```
7009    Male              0     Yes        Yes       12           Yes
7010    Male              0      No         No       12           Yes
7011    Male              0      No         No       72           Yes
7012  Female              1     Yes         No       63           Yes
7013    Male              0     Yes         No       44           Yes
7014  Female              0      No         No       18           Yes
7015  Female              0      No         No        9           Yes
7016    Male              0      No         No       13           Yes
7017  Female              0     Yes         No       68           Yes
7018  Female              1      No         No        6            No
7019  Female              0      No         No        2           Yes
7020    Male              1     Yes         No       55           Yes
7021    Male              1      No         No        1           Yes
7022    Male              0      No         No       38           Yes
7023  Female              0      No         No       67           Yes
7024    Male              0      No         No       19           Yes
7025  Female              0      No         No       12            No
7026  Female              0      No         No       72           Yes
7027    Male              0     Yes        Yes       24           Yes
7028  Female              0     Yes        Yes       72           Yes
7029  Female              0     Yes        Yes       11            No
7030    Male              1     Yes         No        4           Yes
7031    Male              0      No         No       66           Yes

         MultipleLines InternetService OnlineSecurity OnlineBackup  \
0     No phone service             DSL             No          Yes
1                   No             DSL            Yes           No
2                   No             DSL            Yes          Yes
3     No phone service             DSL            Yes           No
4                   No     Fiber optic             No           No
5                  Yes     Fiber optic             No           No
6                  Yes     Fiber optic             No          Yes
7     No phone service             DSL            Yes           No
8                  Yes     Fiber optic             No           No
9                   No             DSL            Yes          Yes
10                  No             DSL            Yes           No
11                  No              No             No           No
12                 Yes     Fiber optic             No           No
13                 Yes     Fiber optic             No          Yes
14                  No     Fiber optic            Yes           No
15                 Yes     Fiber optic            Yes          Yes
16                  No              No             No           No
17                 Yes     Fiber optic            Yes           No
18                  No             DSL             No           No
19                  No     Fiber optic             No          Yes
20    No phone service             DSL             No           No
21                  No              No             No           No
```

```
22                  No           No        No        No
23                 Yes          DSL        No       Yes
24                  No          DSL       Yes       Yes
25                  No          DSL       Yes       Yes
26                 Yes  Fiber optic        No       Yes
27    No phone service          DSL        No       Yes
28                 Yes          DSL       Yes       Yes
29                  No          DSL        No        No
...                ...          ...       ...       ...
7002               Yes  Fiber optic        No       Yes
7003               Yes  Fiber optic        No       Yes
7004                No  Fiber optic        No        No
7005                No          DSL        No       Yes
7006                No           No        No        No
7007                No  Fiber optic        No        No
7008                No           No        No        No
7009                No           No        No        No
7010                No          DSL        No        No
7011               Yes  Fiber optic        No       Yes
7012               Yes  Fiber optic        No       Yes
7013               Yes  Fiber optic       Yes        No
7014               Yes  Fiber optic        No        No
7015                No          DSL        No        No
7016                No          DSL        No       Yes
7017                No          DSL        No       Yes
7018  No phone service          DSL        No        No
7019                No           No        No        No
7020               Yes          DSL       Yes       Yes
7021               Yes  Fiber optic        No        No
7022                No  Fiber optic        No        No
7023               Yes  Fiber optic       Yes       Yes
7024                No  Fiber optic        No        No
7025  No phone service          DSL        No       Yes
7026                No           No        No        No
7027               Yes          DSL       Yes        No
7028               Yes  Fiber optic        No       Yes
7029  No phone service          DSL       Yes        No
7030               Yes  Fiber optic        No        No
7031                No  Fiber optic       Yes        No

     DeviceProtection TechSupport StreamingTV StreamingMovies        Contract  \
0                  No          No          No              No  Month-to-month
1                 Yes          No          No              No        One year
2                  No          No          No              No  Month-to-month
3                 Yes         Yes          No              No        One year
4                  No          No          No              No  Month-to-month
5                 Yes          No         Yes             Yes  Month-to-month
```

6	No	No	Yes	No	Month-to-month
7	No	No	No	No	Month-to-month
8	Yes	Yes	Yes	Yes	Month-to-month
9	No	No	No	No	One year
10	No	No	No	No	Month-to-month
11	No	No	No	No	Two year
12	Yes	No	Yes	Yes	One year
13	Yes	No	Yes	Yes	Month-to-month
14	Yes	Yes	Yes	Yes	Month-to-month
15	Yes	Yes	Yes	Yes	Two year
16	No	No	No	No	One year
17	Yes	No	Yes	Yes	Two year
18	Yes	Yes	No	No	Month-to-month
19	Yes	No	No	Yes	Month-to-month
20	Yes	No	No	Yes	Month-to-month
21	No	No	No	No	One year
22	No	No	No	No	Month-to-month
23	No	Yes	No	No	Two year
24	No	Yes	No	No	Month-to-month
25	No	No	No	No	Month-to-month
26	No	No	Yes	Yes	Month-to-month
27	No	No	No	No	Month-to-month
28	Yes	Yes	Yes	Yes	Two year
29	No	No	Yes	Yes	Month-to-month
...	...	...	...	...	...
7002	Yes	No	Yes	No	Month-to-month
7003	No	No	Yes	No	Month-to-month
7004	Yes	No	Yes	No	Month-to-month
7005	No	No	No	No	Month-to-month
7006	No	No	No	No	Two year
7007	No	No	No	No	Month-to-month
7008	No	No	No	No	Two year
7009	No	No	No	No	Month-to-month
7010	No	Yes	Yes	No	One year
7011	Yes	No	Yes	Yes	One year
7012	Yes	No	Yes	Yes	Month-to-month
7013	Yes	No	No	No	Month-to-month
7014	Yes	Yes	No	Yes	Month-to-month
7015	No	No	No	No	Month-to-month
7016	No	Yes	Yes	Yes	Month-to-month
7017	No	Yes	Yes	No	Two year
7018	No	No	Yes	Yes	Month-to-month
7019	No	No	No	No	Month-to-month
7020	No	No	No	No	One year
7021	No	No	No	No	Month-to-month
7022	No	No	No	No	Month-to-month
7023	Yes	No	Yes	No	Month-to-month

```
7024               No           No          Yes               No  Month-to-month
7025              Yes          Yes          Yes              Yes        One year
7026               No           No           No               No        Two year
7027              Yes          Yes          Yes              Yes        One year
7028              Yes           No          Yes              Yes        One year
7029               No           No           No               No  Month-to-month
7030               No           No           No               No  Month-to-month
7031              Yes          Yes          Yes              Yes        Two year

     PaperlessBilling              PaymentMethod  MonthlyCharges  \
0                 Yes           Electronic check           29.85
1                  No               Mailed check           56.95
2                 Yes               Mailed check           53.85
3                  No  Bank transfer (automatic)           42.30
4                 Yes           Electronic check           70.70
5                 Yes           Electronic check           99.65
6                 Yes    Credit card (automatic)           89.10
7                  No               Mailed check           29.75
8                 Yes           Electronic check          104.80
9                  No  Bank transfer (automatic)           56.15
10                Yes               Mailed check           49.95
11                 No    Credit card (automatic)           18.95
12                 No    Credit card (automatic)          100.35
13                Yes  Bank transfer (automatic)          103.70
14                Yes           Electronic check          105.50
15                 No    Credit card (automatic)          113.25
16                 No               Mailed check           20.65
17                 No  Bank transfer (automatic)          106.70
18                 No    Credit card (automatic)           55.20
19                Yes           Electronic check           90.05
20                Yes           Electronic check           39.65
21                 No  Bank transfer (automatic)           19.80
22                 No               Mailed check           20.15
23                Yes    Credit card (automatic)           59.90
24                 No    Credit card (automatic)           59.60
25                Yes  Bank transfer (automatic)           55.30
26                Yes           Electronic check           99.35
27                 No           Electronic check           30.20
28                Yes    Credit card (automatic)           90.25
29                Yes               Mailed check           64.70
...               ...                        ...             ...
7002              Yes  Bank transfer (automatic)           93.40
7003              Yes           Electronic check           89.20
7004              Yes    Credit card (automatic)           85.20
7005               No           Electronic check           49.95
7006               No  Bank transfer (automatic)           20.65
7007              Yes               Mailed check           70.65
```

7008	No	Mailed check	20.15
7009	Yes	Electronic check	19.20
7010	Yes	Electronic check	59.80
7011	Yes	Electronic check	104.95
7012	Yes	Electronic check	103.50
7013	Yes	Credit card (automatic)	84.80
7014	Yes	Bank transfer (automatic)	95.05
7015	Yes	Bank transfer (automatic)	44.20
7016	No	Mailed check	73.35
7017	No	Bank transfer (automatic)	64.10
7018	Yes	Electronic check	44.40
7019	Yes	Mailed check	20.05
7020	No	Credit card (automatic)	60.00
7021	Yes	Electronic check	75.75
7022	Yes	Credit card (automatic)	69.50
7023	Yes	Credit card (automatic)	102.95
7024	Yes	Bank transfer (automatic)	78.70
7025	No	Electronic check	60.65
7026	Yes	Bank transfer (automatic)	21.15
7027	Yes	Mailed check	84.80
7028	Yes	Credit card (automatic)	103.20
7029	Yes	Electronic check	29.60
7030	Yes	Mailed check	74.40
7031	Yes	Bank transfer (automatic)	105.65

	TotalCharges	Churn
0	29.85	No
1	1889.50	No
2	108.15	Yes
3	1840.75	No
4	151.65	Yes
5	820.50	Yes
6	1949.40	No
7	301.90	No
8	3046.05	Yes
9	3487.95	No
10	587.45	No
11	326.80	No
12	5681.10	No
13	5036.30	Yes
14	2686.05	No
15	7895.15	No
16	1022.95	No
17	7382.25	No
18	528.35	Yes
19	1862.90	No
20	39.65	Yes

```
21          202.25    No
22           20.15   Yes
23         3505.10    No
24         2970.30    No
25         1530.60    No
26         4749.15   Yes
27           30.20   Yes
28         6369.45    No
29         1093.10   Yes
...            ...   ...
7002       3756.40    No
7003       3645.75    No
7004       2874.45    No
7005         49.95    No
7006       1020.75    No
7007         70.65   Yes
7008        826.00    No
7009        239.00    No
7010        727.80   Yes
7011       7544.30    No
7012       6479.40    No
7013       3626.35    No
7014       1679.40    No
7015        403.35   Yes
7016        931.55    No
7017       4326.25    No
7018        263.05    No
7019         39.25    No
7020       3316.10    No
7021         75.75   Yes
7022       2625.25    No
7023       6886.25   Yes
7024       1495.10    No
7025        743.30    No
7026       1419.40    No
7027       1990.50    No
7028       7362.90    No
7029        346.45    No
7030        306.60   Yes
7031       6844.50    No

[7032 rows x 20 columns]
```

4 3. Data Preparation and Exploratory Data Analysis (index)

[8]:
```
# separate features and labels dataframe
y = cleaned_data['Churn']
X = cleaned_data.drop('Churn', axis = 1)
```

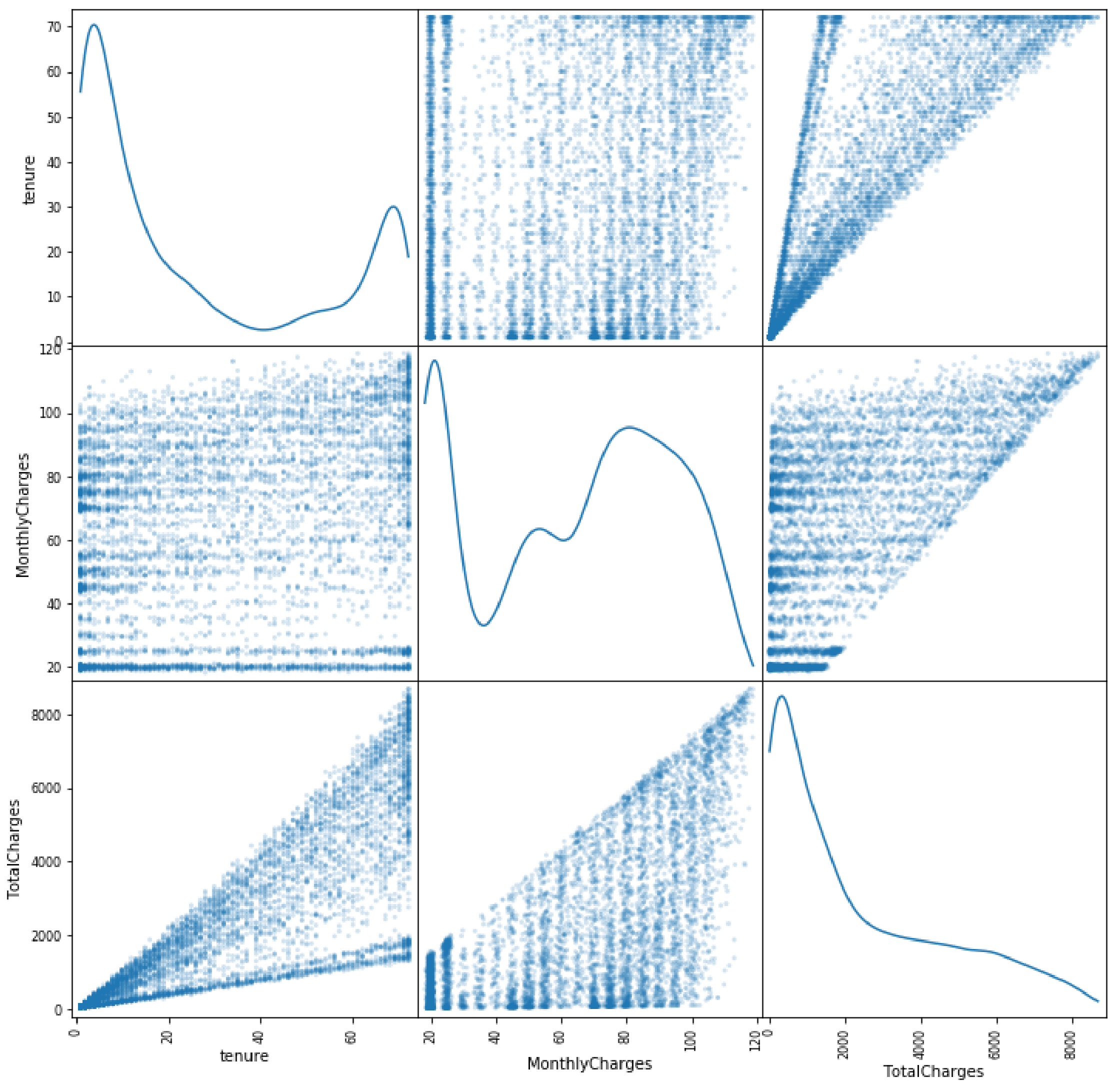
[9]:
```
#display some basic statistics
display(cleaned_data.describe())
display(X.drop(['SeniorCitizen','tenure','MonthlyCharges','TotalCharges'],␣
 ↪axis=1).describe())
display(pd.DataFrame(y, columns=['Churn']).describe())
```

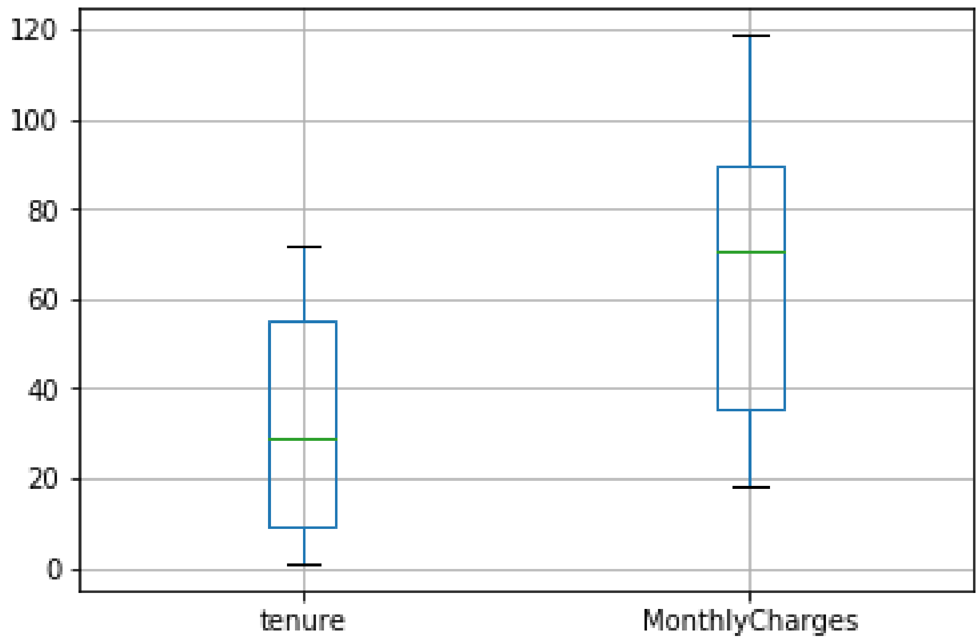
```
       SeniorCitizen       tenure  MonthlyCharges  TotalCharges
count    7032.000000  7032.000000     7032.000000   7032.000000
mean        0.162400    32.421786       64.798208   2283.300441
std         0.368844    24.545260       30.085974   2266.771362
min         0.000000     1.000000       18.250000     18.800000
25%         0.000000     9.000000       35.587500    401.450000
50%         0.000000    29.000000       70.350000   1397.475000
75%         0.000000    55.000000       89.862500   3794.737500
max         1.000000    72.000000      118.750000   8684.800000
```

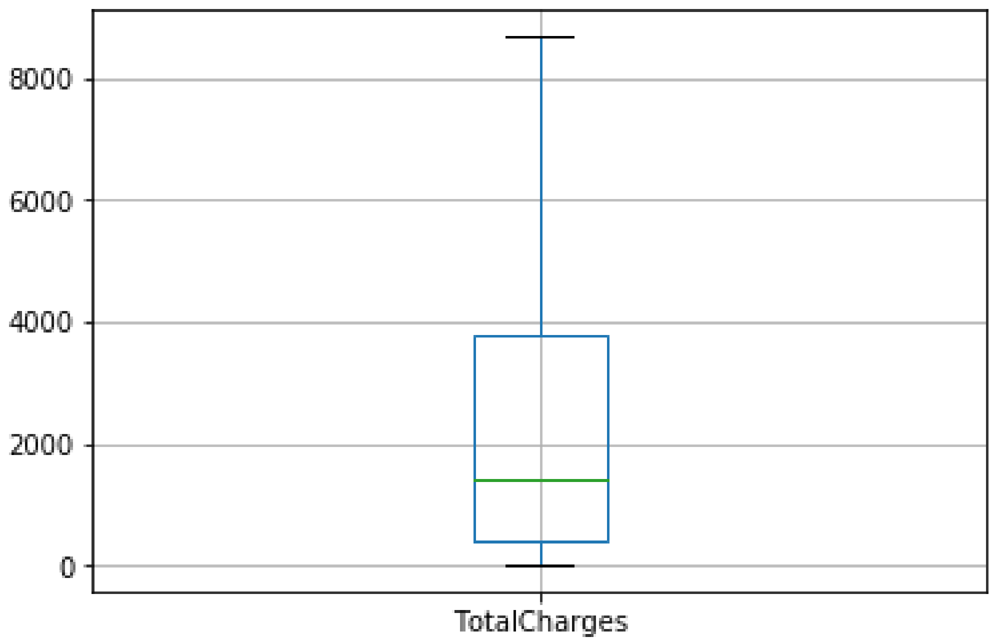
```
       gender Partner Dependents PhoneService MultipleLines InternetService  \
count    7032    7032       7032         7032          7032            7032
unique      2       2          2            2             3               3
top      Male      No         No          Yes            No     Fiber optic
freq     3549    3639       4933         6352          3385            3096

       OnlineSecurity OnlineBackup DeviceProtection TechSupport StreamingTV  \
count            7032         7032             7032        7032        7032
unique              2            2                2           2           2
top                No           No               No          No          No
freq             5017         4607             4614        4992        4329

       StreamingMovies        Contract PaperlessBilling     PaymentMethod
count             7032            7032             7032              7032
unique               2               3                2                 4
top                 No  Month-to-month              Yes  Electronic check
freq              4301            3875             4168              2365
```

```
       Churn
count   7032
unique     2
top       No
freq    5163
```

[10]:
```
#explore data distribution
num_cols = ['tenure','MonthlyCharges','TotalCharges']
pd.plotting.scatter_matrix(cleaned_data[num_cols], alpha = 0.2,
                           figsize = (12,12), diagonal = 'kde')
plt.show()

# inspect for outliers,  boxplots for numerical fields
cleaned_data.boxplot(column = ['tenure','MonthlyCharges'])
plt.show()
cleaned_data.boxplot(column = ['TotalCharges'])
plt.show()
```







[11]:
```
#scale numerical fields
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
```

```
X[num_cols] = scaler.fit_transform(X[num_cols])
```

```
/home/gino/anaconda2/lib/python2.7/site-
packages/sklearn/preprocessing/data.py:625: DataConversionWarning: Data with
input dtype int64, float64 were all converted to float64 by StandardScaler.
  return self.partial_fit(X, y)
/home/gino/anaconda2/lib/python2.7/site-packages/sklearn/base.py:462:
DataConversionWarning: Data with input dtype int64, float64 were all converted
to float64 by StandardScaler.
  return self.fit(X, **fit_params).transform(X)
```

[12]:
```
#One-Hot encode relevant fields

from sklearn.preprocessing import LabelEncoder

#transform y to binary
lb = LabelEncoder()

y = lb.fit_transform(y).ravel()
#convert back to DF
y = pd.DataFrame(y,columns = ['Churn'])


#transform Binary category columns of X to 1's and 0's
bin_cols = X.nunique()[X.nunique() == 2].keys().tolist()

for i in bin_cols:
    X[i] = lb.fit_transform(X[i])

#transform Non-Binary category columns of X to one hot encode
non_bin_cols = ['MultipleLines','InternetService','Contract','PaymentMethod']

X = pd.get_dummies(data = X, columns = non_bin_cols)
```

[13]:
```
#view newly created one-hot features
X.describe()
```

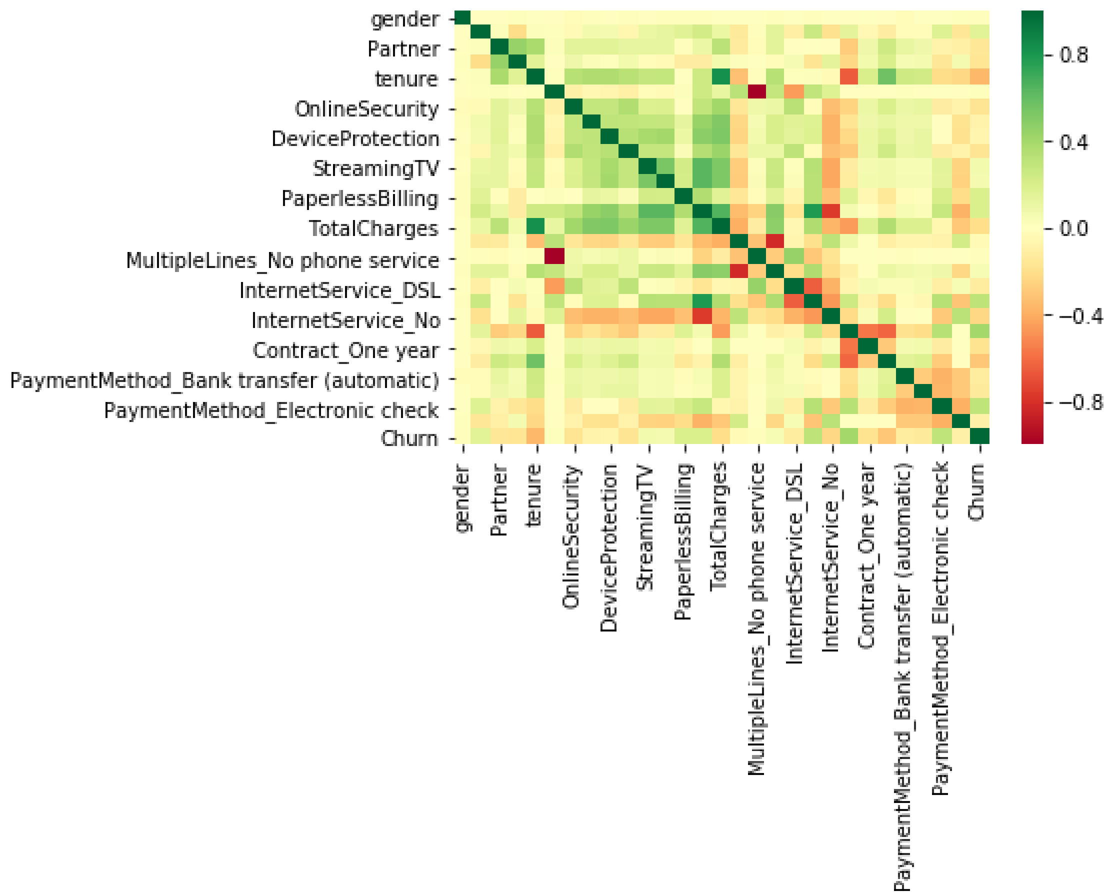
[13]:
```
            gender  SeniorCitizen      Partner   Dependents        tenure  \
count  7032.000000    7032.000000  7032.000000  7032.000000  7.032000e+03
mean      0.504693       0.162400     0.482509     0.298493 -1.214741e-16
std       0.500014       0.368844     0.499729     0.457629  1.000071e+00
min       0.000000       0.000000     0.000000     0.000000 -1.280248e+00
25%       0.000000       0.000000     0.000000     0.000000 -9.542963e-01
50%       1.000000       0.000000     0.000000     0.000000 -1.394171e-01
75%       1.000000       0.000000     1.000000     1.000000  9.199259e-01
max       1.000000       1.000000     1.000000     1.000000  1.612573e+00
```

```
       PhoneService  OnlineSecurity  OnlineBackup  DeviceProtection  \
count   7032.000000     7032.000000   7032.000000       7032.000000
mean       0.903299        0.286547      0.344852          0.343857
std        0.295571        0.452180      0.475354          0.475028
min        0.000000        0.000000      0.000000          0.000000
25%        1.000000        0.000000      0.000000          0.000000
50%        1.000000        0.000000      0.000000          0.000000
75%        1.000000        1.000000      1.000000          1.000000
max        1.000000        1.000000      1.000000          1.000000

       TechSupport             ...            InternetService_DSL  \
count  7032.000000             ...                    7032.000000
mean      0.290102             ...                       0.343572
std       0.453842             ...                       0.474934
min       0.000000             ...                       0.000000
25%       0.000000             ...                       0.000000
50%       0.000000             ...                       0.000000
75%       1.000000             ...                       1.000000
max       1.000000             ...                       1.000000

       InternetService_Fiber optic  InternetService_No  \
count                  7032.000000         7032.000000
mean                      0.440273            0.216155
std                       0.496455            0.411650
min                       0.000000            0.000000
25%                       0.000000            0.000000
50%                       0.000000            0.000000
75%                       1.000000            0.000000
max                       1.000000            1.000000

       Contract_Month-to-month  Contract_One year  Contract_Two year  \
count              7032.000000        7032.000000        7032.000000
mean                  0.551052           0.209329           0.239619
std                   0.497422           0.406858           0.426881
min                   0.000000           0.000000           0.000000
25%                   0.000000           0.000000           0.000000
50%                   1.000000           0.000000           0.000000
75%                   1.000000           0.000000           0.000000
max                   1.000000           1.000000           1.000000

       PaymentMethod_Bank transfer (automatic)  \
count                              7032.000000
mean                                  0.219283
std                                   0.413790
min                                   0.000000
25%                                   0.000000
50%                                   0.000000
```

```
75%                                       0.000000
max                                       1.000000

       PaymentMethod_Credit card (automatic)  PaymentMethod_Electronic check  \
count                            7032.000000                     7032.000000
mean                                0.216297                        0.336320
std                                 0.411748                        0.472483
min                                 0.000000                        0.000000
25%                                 0.000000                        0.000000
50%                                 0.000000                        0.000000
75%                                 0.000000                        1.000000
max                                 1.000000                        1.000000

       PaymentMethod_Mailed check
count                 7032.000000
mean                     0.228100
std                      0.419637
min                      0.000000
25%                      0.000000
50%                      0.000000
75%                      0.000000
max                      1.000000

[8 rows x 28 columns]
```

4.0.1 3.1 Correlation Matrix (index)

[14]:
```
import seaborn as sns

joined = X.join(y)
gr = sns.heatmap(joined.corr(),annot=False,cmap="RdYlGn")
sns.set(rc={'figure.figsize':(22,14)})
plt.show()
```

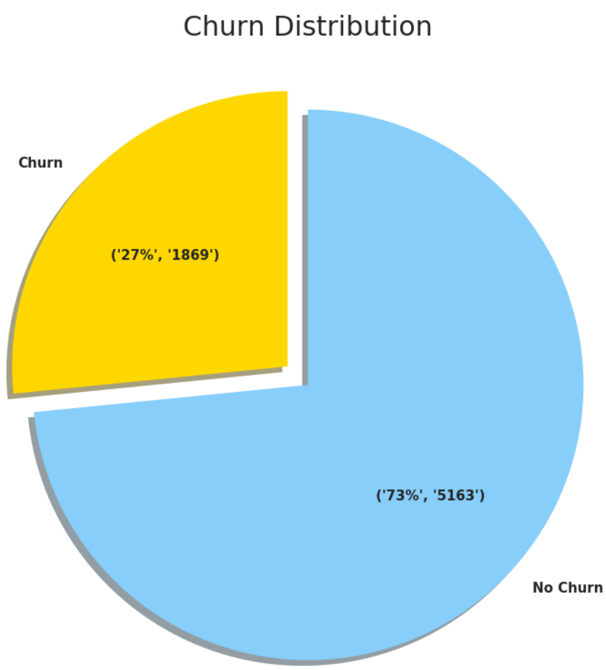


4.0.2 3.2 Variables Distribution (index)

[15]:

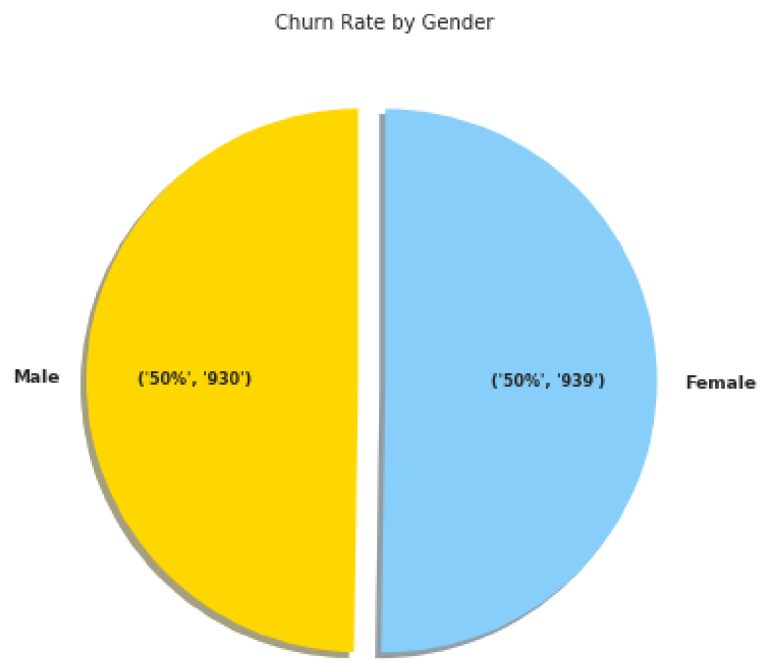
```
# default plot settings
labels = 'Churn', 'No Churn'
colors = ['gold', 'lightskyblue']
explode = (0.1, 0)  # explode 1st slice
labelsize = 15
titlesize = 30
valuesize = 15

#function for pie chart to show BOTH value and %
def absolute_value(val):
    a  = np.round(val/100.*np.array(values).sum(), 0)
    return '%.0f%%' % val, '%.0f' % a

#function to create pie chart
def plot_pie(values, colors, title):
```

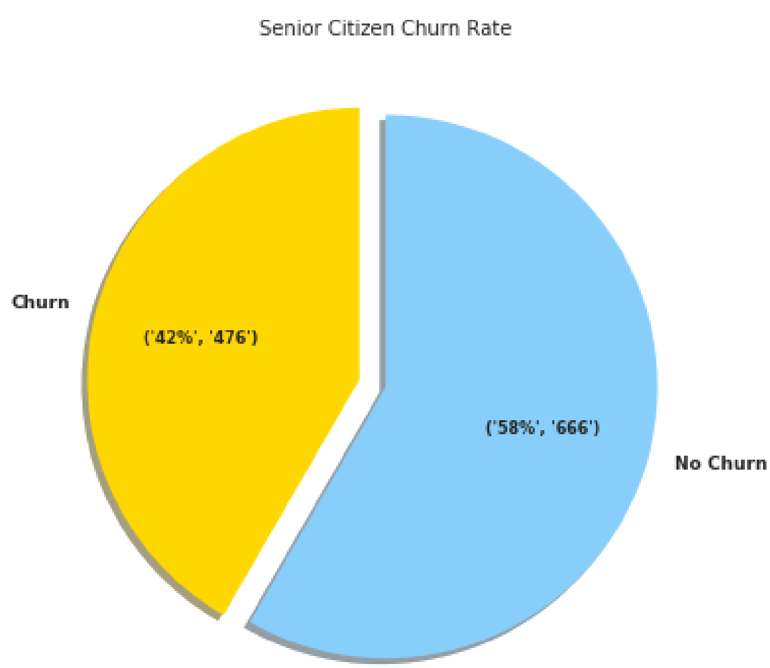
```
    plt.title(title, size = titlesize)
    plt.axis('equal')
    font = {'family' : 'Sans Serif',
        'weight' : 'bold',
        'size'   : valuesize}
    plt.rc('xtick', labelsize=labelsize)
    plt.rc('font', **font)
    plt.pie(values, explode=explode, labels=labels, colors=colors,
        autopct=absolute_value, shadow=True, startangle=90)
    return plt.show()

# Plot Data

values = [y.sum(), len(y) - y.sum()]  #[churn, no churn]
plot_pie(values, colors, 'Churn Distribution')
```

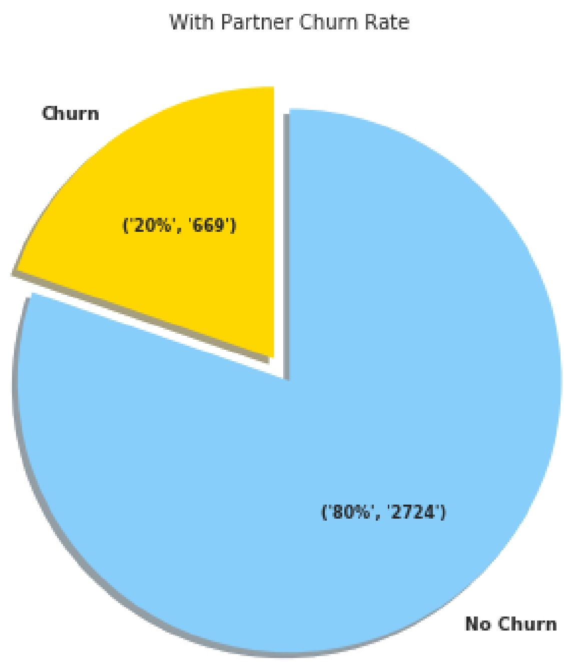


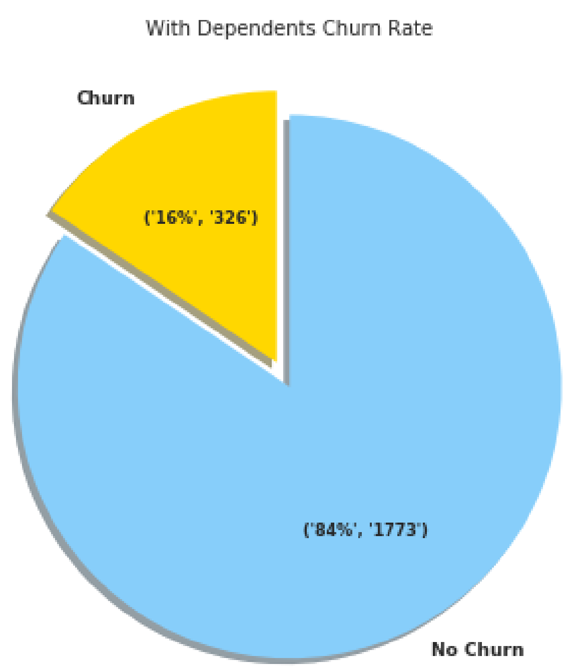
[16]:
```
# default plot settings
labels = 'Churn', 'No Churn'
colors = ['gold', 'lightskyblue']
explode = (0.1, 0)  # explode 1st slice
labelsize = 9
titlesize = 10
valuesize = 8
```

```
grid = plt.GridSpec(2,2)

#by Gender
male_churn = y['Churn'][X['gender'] == 1].sum()
female_churn = y['Churn'][X['gender'] == 0].sum()

# by Gender
plt.subplot(grid[0, 0])
values = [male_churn,female_churn]
labels = 'Male', 'Female'
plot_pie(values, colors, 'Churn Rate by Gender')



# by Senior Citizen
plt.subplot(grid[0, 1])
senior = pd.DataFrame()
senior['Churn'] = y['Churn'][X['SeniorCitizen'] == 1]
values = [senior['Churn'][senior['Churn'] == 1].
 ↪sum(),len(senior['Churn'][senior['Churn'] == 0])]
labels = 'Churn', 'No Churn'
plot_pie(values, colors, 'Senior Citizen Churn Rate')

# by Partner
plt.subplot(grid[1, 0])
partner = pd.DataFrame()
partner['Churn'] = y['Churn'][X['Partner'] == 1]
values = [partner['Churn'][partner['Churn'] == 1].
 ↪sum(),len(partner['Churn'][partner['Churn'] == 0])]
labels = 'Churn', 'No Churn'
plot_pie(values, colors, 'With Partner Churn Rate')


# by Dependents
plt.subplot(grid[1, 1])
dependents = pd.DataFrame()
dependents['Churn'] = y['Churn'][X['Dependents'] == 1]
values = [dependents['Churn'][dependents['Churn'] == 1].sum(),␣
 ↪len(dependents['Churn'][dependents['Churn'] == 0])]
labels = 'Churn', 'No Churn'
plot_pie(values, colors, 'With Dependents Churn Rate')
```


Churn Rate by Gender
Male
('50%', '930')
('50%', '939')
Female


Senior Citizen Churn Rate
Churn
('42%', '476')
('58%', '666')
No Churn


With Partner Churn Rate
Churn
('20%', '669')
('80%', '2724')
No Churn


With Dependents Churn Rate
Churn
('16%', '326')
('84%', '1773')
No Churn

4.0.3 3.3 Principal Component Analysis (index)

[17]:
```
from sklearn.decomposition import PCA
pca = PCA(n_components = 28, random_state=1)
pca.fit(X)
reduced_data = pca.transform(X)

#convert back to Dataframe
reduced_data = pd.DataFrame(reduced_data)

#show 1st n_components that comprises 98% of variance
print "1st 21 Components represent", pca.explained_variance_ratio_[0:20].
 ↪sum(),'% of Variance'

#use only 1st 21 components
drop = reduced_data.columns[21:].values.tolist()
reduced_data.drop(drop, axis = 1, inplace=True)
```

```
1st 21 Components represent 0.983002583600193 % of Variance
```

5 4. Model Creation and Testing

5.0.1 4.1 Baseline Model (index)

[18]:
```
from sklearn.metrics import fbeta_score, recall_score, precision_score,␣
 ↪accuracy_score

'''NAIVE BASELINE MODEL'''
#Assume all customers will churn

#True total number of Churn customers
num_churn = float(np.sum(y))

#Naive Prediction: Assume all customers will churn
pred = float(len(y))


print "**** BASIC NAIVE Benchmark ****"

naive_acc = num_churn / pred
print "Naive Benchmark Accuracy = ", naive_acc

naive_f2beta = fbeta_score(y, np.ones((len(y),)), beta= 2)
print "Naive Benchmark F2 Beta Score = ", naive_f2beta

naive_recall = recall_score(y , np.ones((len(y),)))
print "Naive Benchmark Recall Score = ", naive_recall
```

```
naive_precision = precision_score(y , np.ones((len(y),)))
print "Naive Benchmark Precision Score = ", naive_precision



'''IMPROVED NAIVE MODEL'''
# predictions of customers with less than 12 months tenure

# scale the tenure value of 12 months
transformed_12 = scaler.transform([[12,0,0]])

tenure = y.copy()

#initiate all values to 0
tenure['Prediction'] = 0

#set prediction to 1 if less than 12 months tenure
tenure['Prediction'][X['tenure'] < transformed_12[0][0]] = 1

acc = accuracy_score(tenure['Churn'], tenure['Prediction'])

print "\n**** IMPROVED NAIVE Benchmark ****"
print "Naive Tenure-based Benchmark Accuracy = ", acc

naive_tenure_f2beta = fbeta_score(tenure['Churn'], tenure['Prediction'], beta=
 ↪2)
print "Naive Tenure-based Benchmark F2 Beta Score = ", naive_tenure_f2beta

naive_tenure_recall = recall_score(tenure['Churn'], tenure['Prediction'])
print "Naive Tenure-based Benchmark Recall Score = ", naive_tenure_recall

naive_tenure_precision = precision_score(tenure['Churn'], tenure['Prediction'])
print "Naive Tenure-based Benchmark Precision Score = ", naive_tenure_precision
```

```
**** BASIC NAIVE Benchmark ****
Naive Benchmark Accuracy =  0.265784982935
Naive Benchmark F2 Beta Score =  0.6441273779983457
Naive Benchmark Recall Score =  1.0
Naive Benchmark Precision Score =  0.26578498293515357

**** IMPROVED NAIVE Benchmark ****
Naive Tenure-based Benchmark Accuracy =  0.7256825938566553
Naive Tenure-based Benchmark F2 Beta Score =  0.5239144115796098
Naive Tenure-based Benchmark Recall Score =  0.5345104333868379
Naive Tenure-based Benchmark Precision Score =  0.48542274052478135
```

[19]:
```python
#create confusion matrix for Improved Naive Benchmark
from sklearn.metrics import confusion_matrix

#conf_matrix_naive = confusion_matrix(y, np.ones((len(y),)))
conf_matrix_naive = confusion_matrix(tenure['Prediction'], tenure['Churn'])

#convert to DF
df_cf_naive = pd.DataFrame(conf_matrix_naive,columns = ['No Churn␣
↪Predicted','Churn Predicted'],
                         index = ['No Churn', 'Churn'])

print '\n    **** Improved Naive Benchmark Confusion Matrix ***'
fig = sns.heatmap(df_cf_naive,annot=True,fmt = 'd',cmap="RdYlGn")
sns.set(rc={'figure.figsize':(12,8)})
```

```
    **** Improved Naive Benchmark Confusion Matrix ***
```

[20]:
```python
from sklearn.model_selection import train_test_split
from sklearn.metrics import make_scorer, accuracy_score
from sklearn.model_selection import GridSearchCV, StratifiedShuffleSplit
import pickle
```

```
# make y into 1d array
y = y.values
y = y.ravel()

#create training and test split
X_train, X_test, y_train, y_test = train_test_split(reduced_data, y,
                                                    test_size=0.15,
                                                    random_state=42)



#Create scorer for Grid Search
f2scorer = make_scorer(fbeta_score, beta = 2)



#create Stratified Shuffle Split for cross validation
cv = StratifiedShuffleSplit(n_splits=5, test_size=0.20, random_state=1)
```

5.0.2 4.2 Support Vector Machine (index)

[21]:
```
from sklearn.svm import SVC
svc = SVC(random_state=1)
svc_parameters = {'kernel':['rbf'],
                  'C':[1e07, 1e09,1000.],
                  'gamma': [1e-09,1e-10]
                 }

# Store in Pickle file.  Uncomment below to rerun classifier with new␣
 ↪parameters
#clf_gs = GridSearchCV(svc, svc_parameters, scoring = f2scorer, cv =cv, n_jobs␣
 ↪= -1)

#clf_gs.fit(X_train, y_train)

#with open('grid_searchcv_svc.pickle','wb') as f:
#         pickle.dump(clf_gs, f)

pickle_in = open('grid_searchcv_svc.pickle','rb')
clf_gs = pickle.load(pickle_in)
print clf_gs.best_estimator_
print clf_gs.best_params_
```

```
SVC(C=10000000.0, cache_size=200, class_weight=None, coef0=0.0,
  decision_function_shape='ovr', degree=3, gamma=1e-10, kernel='rbf',
  max_iter=-1, probability=False, random_state=1, shrinking=True,
  tol=0.001, verbose=False)
{'kernel': 'rbf', 'C': 10000000.0, 'gamma': 1e-10}
```

[22]:
```python
svc = SVC(kernel = clf_gs.best_params_['kernel'],
          C = clf_gs.best_params_['C'],
          gamma = clf_gs.best_params_['gamma'],
          random_state = 1
         )

svc.fit(X_train,y_train)
```

[22]:
```
SVC(C=10000000.0, cache_size=200, class_weight=None, coef0=0.0,
  decision_function_shape='ovr', degree=3, gamma=1e-10, kernel='rbf',
  max_iter=-1, probability=False, random_state=1, shrinking=True,
  tol=0.001, verbose=False)
```

5.0.3 4.3. Decision Tree (index)

[23]:
```python
from sklearn import tree

dt = tree.DecisionTreeClassifier(random_state=1, class_weight= 'balanced')
```

[24]:
```python
#Parameters for Decision Tree

dt_parameters = {'criterion':['gini', 'entropy'],
                 'max_depth':[3, 5, 8,12],
                 'splitter':['random','best'],
                 'min_samples_leaf':[1,2,5,10]
                }


# Store in Pickle file.  Uncomment below to rerun classifier with new␣
 ↪parameters
#clf_gs = GridSearchCV(dt, dt_parameters, scoring = f2scorer, cv =cv, n_jobs =␣
 ↪-1)

#clf_gs.fit(X_train, y_train)

#with open('grid_searchcv_dt.pickle','wb') as f:
#          pickle.dump(clf_gs, f)

pickle_in = open('grid_searchcv_dt.pickle','rb')
clf_gs = pickle.load(pickle_in)
print clf_gs.best_estimator_
print clf_gs.best_params_
```

```
DecisionTreeClassifier(class_weight='balanced', criterion='gini', max_depth=3,
            max_features=None, max_leaf_nodes=None,
            min_impurity_decrease=0.0, min_impurity_split=None,
            min_samples_leaf=10, min_samples_split=2,
            min_weight_fraction_leaf=0.0, presort=False, random_state=1,
```

```
             splitter='random')
{'splitter': 'random', 'criterion': 'gini', 'max_depth': 3, 'min_samples_leaf':
10}
```

[25]:
```
dt = tree.DecisionTreeClassifier(criterion = clf_gs.best_params_['criterion'],
          max_depth = clf_gs.best_params_['max_depth'],
          splitter = clf_gs.best_params_['splitter'],
          min_samples_leaf = clf_gs.best_params_['min_samples_leaf'],
          random_state = 1,
          class_weight= 'balanced'
       )

dt.fit(X_train,y_train)
```

[25]:
```
DecisionTreeClassifier(class_weight='balanced', criterion='gini', max_depth=3,
            max_features=None, max_leaf_nodes=None,
            min_impurity_decrease=0.0, min_impurity_split=None,
            min_samples_leaf=10, min_samples_split=2,
            min_weight_fraction_leaf=0.0, presort=False, random_state=1,
            splitter='random')
```

5.0.4 4.4. XGBoost (index)

[26]:
```
from xgboost import XGBClassifier
xgb = XGBClassifier(random_state=1)
```

[27]:
```
xgb_parameters = {'booster':['gblinear', 'gbtree','dart'],
                 'learning_rate':[0.1, 0.01, 0.3, 0.6],
                  'max_depth': [1,3,9],
                  'scale_pos_weight': [2.76, 5.5, 6.0]
                 }


# Store in Pickle file.  Uncomment below to rerun classifier with new␣
 ↪parameters
#clf_gs = GridSearchCV(xgb, xgb_parameters, scoring = f2scorer, cv = cv, n_jobs␣
 ↪= -1)

#clf_gs.fit(X_train, y_train)

#with open('grid_searchcv_xgb.pickle','wb') as f:
#          pickle.dump(clf_gs, f)

pickle_in = open('grid_searchcv_xgb.pickle','rb')
clf_gs = pickle.load(pickle_in)
print clf_gs.best_estimator_
print clf_gs.best_params_
```

```
XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1,
       colsample_bytree=1, gamma=0, learning_rate=0.3, max_delta_step=0,
       max_depth=1, min_child_weight=1, missing=nan, n_estimators=100,
       n_jobs=1, nthread=None, objective='binary:logistic', random_state=1,
       reg_alpha=0, reg_lambda=1, scale_pos_weight=6.0, seed=None,
       silent=True, subsample=1)
{'scale_pos_weight': 6.0, 'learning_rate': 0.3, 'max_depth': 1, 'booster':
'gbtree'}
```

[28]:
```
xgb = XGBClassifier(booster=clf_gs.best_params_['booster'],
                    learning_rate=clf_gs.best_params_['learning_rate'],
                    max_depth = clf_gs.best_params_['max_depth'],
                    scale_pos_weight=clf_gs.best_params_['scale_pos_weight'],
                    random_state = 1)

xgb.fit(X_train,y_train)
```

[28]:
```
XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1,
       colsample_bytree=1, gamma=0, learning_rate=0.3, max_delta_step=0,
       max_depth=1, min_child_weight=1, missing=None, n_estimators=100,
       n_jobs=1, nthread=None, objective='binary:logistic', random_state=1,
       reg_alpha=0, reg_lambda=1, scale_pos_weight=6.0, seed=None,
       silent=True, subsample=1)
```

5.0.5 4.5. LightGBM (index)

[29]:
```
from lightgbm import LGBMClassifier
lgbm = LGBMClassifier(random_state = 1, is_unbalance = True, objective='binary')
```

[30]:
```
lgbm_parameters = {'boosting':['gbdt','goss','dart'],
                   'min_data_per_leaf':[5,10,30,500],
                   'num_leaves':[5,31,80,150],
                   'max_depth': [1,3,7,9],
                   'is_unbalance':[True],
                   'objective':['binary'],
                   'random_state': [1]
                  }

# Store in Pickle file.  Uncomment below to rerun classifier with new␣
 ↪parameters
#clf_gs = GridSearchCV(lgbm, lgbm_parameters, scoring = f2scorer, cv = cv,␣
 ↪n_jobs = -1)

#clf_gs.fit(X_train, y_train)

#with open('grid_searchcv_lgbm.pickle','wb') as f:
#         pickle.dump(clf_gs, f)
```

```
pickle_in = open('grid_searchcv_lgbm.pickle','rb')
clf_gs = pickle.load(pickle_in)
print clf_gs.best_estimator_
print clf_gs.best_params_
```

```
LGBMClassifier(boosting='dart', boosting_type='gbdt', class_weight=None,
        colsample_bytree=1.0, importance_type='split', is_unbalance=True,
        learning_rate=0.1, max_depth=3, min_child_samples=20,
        min_child_weight=0.001, min_data_per_leaf=5, min_split_gain=0.0,
        n_estimators=100, n_jobs=-1, num_leaves=5, objective='binary',
        random_state=1, reg_alpha=0.0, reg_lambda=0.0, silent=True,
        subsample=1.0, subsample_for_bin=200000, subsample_freq=0)
{'num_leaves': 5, 'min_data_per_leaf': 5, 'random_state': 1, 'is_unbalance':
True, 'boosting': 'dart', 'objective': 'binary', 'max_depth': 3}
```

[31]:
```
lgbm = LGBMClassifier(boosting = clf_gs.best_params_['boosting'],
                      num_leaves = clf_gs.best_params_['num_leaves'],
                      max_depth = clf_gs.best_params_['max_depth'],
                      min_data_per_leaf = clf_gs.
 ↪best_params_['min_data_per_leaf'],
                      random_state = 1,
                      is_unbalance = True,
                      objective='binary'
                     )

lgbm.fit(X_train,y_train)
```

[31]:
```
LGBMClassifier(boosting='dart', boosting_type='gbdt', class_weight=None,
        colsample_bytree=1.0, importance_type='split', is_unbalance=True,
        learning_rate=0.1, max_depth=3, min_child_samples=20,
        min_child_weight=0.001, min_data_per_leaf=5, min_split_gain=0.0,
        n_estimators=100, n_jobs=-1, num_leaves=5, objective='binary',
        random_state=1, reg_alpha=0.0, reg_lambda=0.0, silent=True,
        subsample=1.0, subsample_for_bin=200000, subsample_freq=0)
```

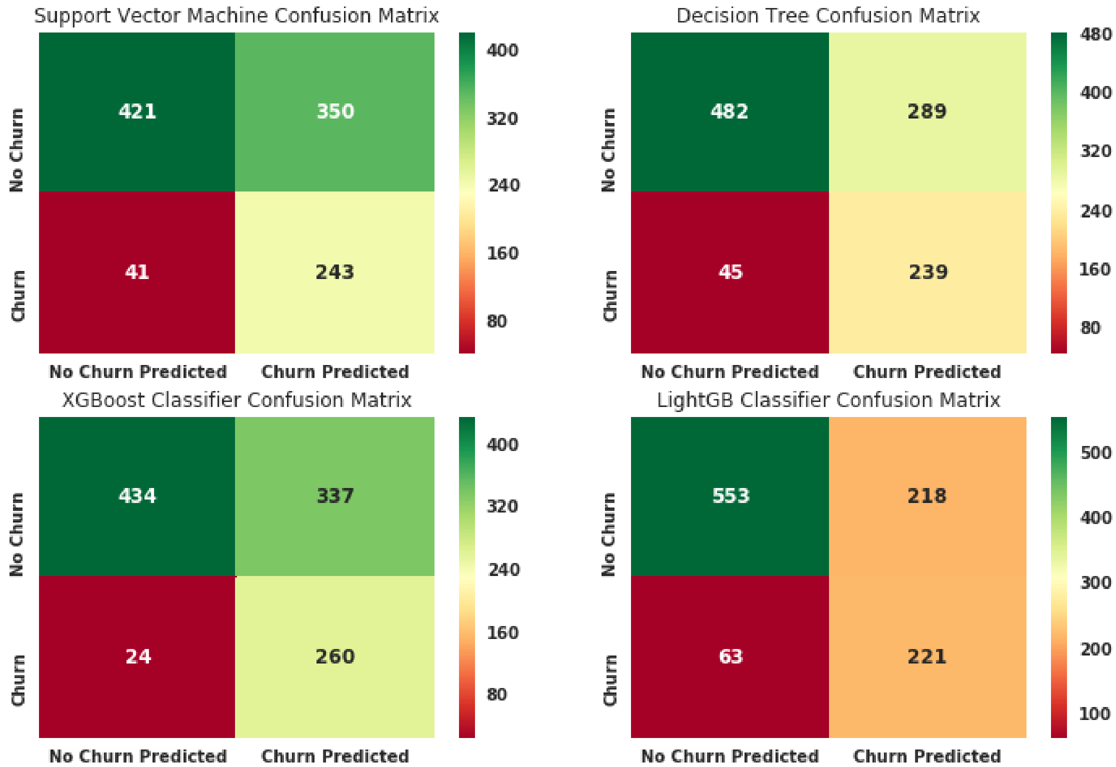
6 5. Model Evaluation and Performance Metrics

6.0.1 5.1 Confusion Matrices for Models (index)

[32]:
```
#create confusion matrices


models = {svc:'Support Vector Machine', dt:'Decision Tree', xgb:'XGBoost␣
 ↪Classifier', lgbm:'LightGB Classifier'}
fig = plt.GridSpec(2,2)

#SVC Confusion Matrix
```

```
model = svc

conf_matrix = confusion_matrix(y_test, model.predict(X_test))
plt.subplot(grid[0,0])
#convert to DF
df_cf = pd.DataFrame(conf_matrix,columns = ['No Churn Predicted','Churn␣
↪Predicted'],
                     index = ['No Churn', 'Churn'])
a = sns.heatmap(df_cf,annot=True,fmt = 'd',cmap="RdYlGn")
sns.set(rc={'figure.figsize':(12,8)})
a.set_title('%s Confusion Matrix' % models[model])

#Record metrics for the model
f2 = fbeta_score(y_test, model.predict(X_test), beta = 2)
recall = recall_score(y_test, model.predict(X_test))
precision = precision_score(y_test, model.predict(X_test))
accuracy = accuracy_score(y_test, model.predict(X_test))
svc_metrics = {'Model': models[model],'F2 Score':[f2], 'Recall': [recall],
               'Precision':[precision], 'Accuracy':[accuracy]}
all_metrics = pd.DataFrame(svc_metrics)




#Decision Tree Confusion Matrix
model = dt

conf_matrix = confusion_matrix(y_test, model.predict(X_test))
plt.subplot(grid[0,1])
#convert to DF
df_cf = pd.DataFrame(conf_matrix,columns = ['No Churn Predicted','Churn␣
↪Predicted'],
                     index = ['No Churn', 'Churn'])
a = sns.heatmap(df_cf,annot=True,fmt = 'd',cmap="RdYlGn")
sns.set(rc={'figure.figsize':(12,8)})
a.set_title('%s Confusion Matrix' % models[model])

#Record metrics for the model
f2 = fbeta_score(y_test, model.predict(X_test), beta = 2)
recall = recall_score(y_test, model.predict(X_test))
precision = precision_score(y_test, model.predict(X_test))
accuracy = accuracy_score(y_test, model.predict(X_test))
data = {'Model': models[model],'F2 Score':[f2], 'Recall': [recall],
               'Precision':[precision], 'Accuracy':[accuracy]}
dt_metrics = pd.DataFrame(data)
all_metrics = all_metrics.append(dt_metrics)
```

```
#XGBoost Confusion Matrix
model = xgb
conf_matrix = confusion_matrix(y_test, model.predict(X_test))
plt.subplot(grid[1,0])
#convert to DF
df_cf = pd.DataFrame(conf_matrix,columns = ['No Churn Predicted','Churn␣
 ↪Predicted'],
                     index = ['No Churn', 'Churn'])
a = sns.heatmap(df_cf,annot=True,fmt = 'd',cmap="RdYlGn")
sns.set(rc={'figure.figsize':(12,8)})
a.set_title('%s Confusion Matrix' % models[model])

#Record metrics for the model
f2 = fbeta_score(y_test, model.predict(X_test), beta = 2)
recall = recall_score(y_test, model.predict(X_test))
precision = precision_score(y_test, model.predict(X_test))
accuracy = accuracy_score(y_test, model.predict(X_test))
data = {'Model': models[model],'F2 Score':[f2], 'Recall': [recall],
               'Precision':[precision], 'Accuracy':[accuracy]}
xgb_metrics = pd.DataFrame(data)
all_metrics = all_metrics.append(xgb_metrics)



#LightBM Confusion Matrix
model = lgbm
conf_matrix = confusion_matrix(y_test, model.predict(X_test))
plt.subplot(grid[1,1])
#convert to DF
df_cf = pd.DataFrame(conf_matrix,columns = ['No Churn Predicted','Churn␣
 ↪Predicted'],
                     index = ['No Churn', 'Churn'])
a = sns.heatmap(df_cf,annot=True,fmt = 'd',cmap="RdYlGn")
sns.set(rc={'figure.figsize':(12,8)})
a.set_title('%s Confusion Matrix' % models[model])

#Record metrics for the model
f2 = fbeta_score(y_test, model.predict(X_test), beta = 2)
recall = recall_score(y_test, model.predict(X_test))
precision = precision_score(y_test, model.predict(X_test))
accuracy = accuracy_score(y_test, model.predict(X_test))
data = {'Model': models[model],'F2 Score':[f2], 'Recall': [recall],
               'Precision':[precision], 'Accuracy':[accuracy]}
lgbm_metrics = pd.DataFrame(data)
all_metrics = all_metrics.append(lgbm_metrics)
```



6.0.2 5.2 Compare Model Metrics (index)

[33]:
```
#Prepare Metric Results to display

#Add Naive Benchmark Metrics
data = {'Model': ['Naive Benchmark'],'F2 Score':naive_f2beta, 'Recall':␣
 ↪naive_recall,
                'Precision':naive_precision, 'Accuracy':naive_acc}
naive_metrics = pd.DataFrame(data)
all_metrics = all_metrics.append(naive_metrics)

#Add Improved Benchmark Metrics
data = {'Model': ['Improved Naive Benchmark'],'F2 Score': naive_tenure_f2beta,␣
 ↪'Recall': naive_tenure_recall,
                'Precision':naive_tenure_precision, 'Accuracy':acc}
imp_naive_metrics = pd.DataFrame(data)
all_metrics = all_metrics.append(imp_naive_metrics)

#order headers in DF
all_metrics = all_metrics[['Model','F2 Score', 'Recall', 'Precision',␣
 ↪'Accuracy']]
```

```python
#Display Table of Metrics
from IPython.display import HTML
HTML(all_metrics.to_html(index=False))
```

[33]:
```
<IPython.core.display.HTML object>
```

[34]:
```python
#plot important features
from xgboost import plot_importance

xgb = xgb.fit(X,y)
plot_importance(xgb, importance_type='gain')
```

[34]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f29fafb5fd0>
```