SMART SERVICE VS CHURN - CHURNSERVICE

Author: Ali Sedaghat
Date: 2022-06-13
Version: 1.0
Tags: Churn
Project: Churn Analysis
Notebook: Churn

Import Libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import StandardScaler, MinMaxScaler, LabelEncoder
from sklearn.model_selection import train_test_split, GridSearchCV, RandomizedSearchCV
from sklearn.metrics import confusion_matrix, classification_report, roc_curve, roc_auc_score, auc
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTree, DecisionTreeClassifier, plot_tree
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.svm import SVC
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.pipeline import Pipeline
from sklearn.feature_selection import SelectKBest, chi2, f_classif
from sklearn.decomposition import PCA
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
from scipy.stats import chi2_contingency, ttest_ind, mannwhitneyu
import warnings
warnings.filterwhere('ignore')
%matplotlib inline
%matplotlib inline
```

Load Data

```
df = pd.read_csv('./data/churn.csv')
```

```
df.shape
```

```
(7043, 21)
```

```
df.info()
```

```
<class 'DataFrame'>
RangeIndex: 7043 entries, 0 to 7043
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
dtype: int64, object: 18, int64: 3
memory usage: 1.1+MB
```

```
df.head(5)
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	Yes	No	No	No	No	Month-to-month	Yes	Electronic check	29.85	29.85	No
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	No	Yes	No	No	No	One year	No	Mailed check	56.95	1889.5	No
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	No	No	No	No	Month-to-month	Yes	Mailed check	53.85	108.15	Yes
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	No	Yes	Yes	No	No	One year	No	Bank transfer (automatic)	42.3	1840.75	No
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No	No	No	No	Month-to-month	Yes	Electronic check	70.7	151.65	Yes

```
df.tail(5)
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
7039	2234-XADUH	Female	0	Yes	Yes	72	Yes	Yes	Fiber optic	No	Yes	Yes	No	Yes	Yes	One year	Yes	Credit card (automatic)	103.2	7362.9	No
7040	4801-JZAZL	Female	0	Yes	Yes	11	No	No phone service	DSL	Yes	No	No	No	No	No	Month-to-month	Yes	Electronic check	29.65	314.75	No
7041	8779-QRDMV	Male	1	No	No	39	Yes	No	Fiber optic	No	No	Yes	No	Yes	Yes	Month-to-month	Yes	Electronic check	74.4	3214.8	Yes
7042	1313-GNXLE	Male	0	Yes	Yes	55	Yes	Yes	Fiber optic	Yes	Yes	Yes	Yes	Yes	Yes	One year	No	Bank transfer (automatic)	118.75	6544.65	No
7043	3186-AJIEK	Male	0	No	No	14	No	No phone service	DSL	Yes	Yes	Yes	Yes	Yes	Yes	Two year	Yes	Bank transfer (automatic)	105.65	1419.4	No

```
df.sample(5)
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	No	Yes	No	No	No	One year	No	Mailed check	56.95	1889.5	No
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	Yes	No	No	No	No	Month-to-month	Yes	Mailed check	53.85	108.15	Yes
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	No	Yes	Yes	No	No	One year	No	Bank transfer (automatic)	42.3	1840.75	No
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No	No	No	No	Month-to-month	Yes	Electronic check	70.7	151.65	Yes
5	9305-CDSKC	Female	0	No	No	8	Yes	Yes	Fiber optic	No	No	Yes	No	Yes	Yes	Month-to-month	Yes	Electronic check	99.65	820.5	Yes

```
df.columns
```

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

```
df.dtypes
```

```
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: int64, object: 18, int64: 3
```

Data Cleaning

```
df.isnull()
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False	False

```
df.isnull().sum()
```

```
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      0
Churn             0
dtype: int64
```

```
df.duplicated()
```

```
0    False
1    False
2    False
3    False
4    False
dtype: bool
```

```
df.duplicated().sum()
```

```
0
```

```
df.nunique()
```

```
customerID        7043
gender            2
SeniorCitizen     2
Partner           2
Dependents        2
tenure            73
PhoneService      2
MultipleLines     3
InternetService   3
OnlineSecurity    3
OnlineBackup      3
DeviceProtection  3
TechSupport       3
StreamingTV       3
StreamingMovies   3
Contract          3
PaperlessBilling  2
PaymentMethod     4
MonthlyCharges    1585
TotalCharges      6531
Churn             2
dtype: int64
```

```
df.drop(columns=['customerID'])
```

```
df.shape
```

```
(7043, 20)
```

```
df.columns
```

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

```
df.dtypes
```

```
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: int64, object: 18, int64: 3
```

```
df.TotalCharges
```

```
0       29.85
1      1889.5
2      108.15
3     1840.75
4      151.65
...       ...
7039   7362.9
7040   314.75
7041   3214.8
7042   6544.65
7043   1419.4
Name: TotalCharges, dtype: object
```

```
df.TotalCharges.unique()
```

```
array[' 29.85', '1889.5', '108.15', '1840.75', '151.65', '820.5',
       '1949.4', '301.9', '3046.05', '3487.95', '587.45', '326.8',
       '5681.1', '5036.3', '2070.6', '843.9', '2547.15', '2837.75',
       '202.25', '401.45', '1023.1', '7362.9', '314.75', '3214.8',
       '6544.65', '1419.4', '2683.3', '908.15', '1837.75', '267.4',
       '799.9', '2844.5', '3035.5', '235.9', '748.35', '987.65',
       '148.35', '742.9', '1873.7', '4749.15', '2422.1', '889.5',
       '268.55', '271.75', '1372.9', '610.1', '3505.1', '2993.7',
       '2181.3', '1411.15', '1077.6', '7076.35', '1357.5', '907.1',
       '787.3', '362.05', '135.75', '820.75', '989.35', '1490.8',
       '7931.05', '1026.1', '381.9', '5778.5', '62.1', '305.5',
       '3475.8', '742.15', '2361.25', '132.15', '1724.8', '411.35',
       '1320.75', '1062.7', '3731.7', '2608.9', '2994.25', '393.1',
       '7899.2', '1099.9', '686.35', '349.1', '3772.9', '2995.7',
       '565.25', '2901.25', '3586.35', '5681.1', '1843.2', '2051.3',
       '2035.05', '577.25', '8684.8', '394.75', '503.65', '116.8',
       '1022.95', '1627.65', '916.7', '953.7', '745.3', '863.3',
       '744.05', '116.15', '863.65', '304.2', '6811.9', '3751.05',
       '1990.5', '7362.05', '1678.6', '7392.35', '334.35', '2848.35',
       '1465.4', '8672.45', '1022.4', '799.75', '772.3', '5370.35',
       '1203.4', '267.75', '569.85', '2660.4', '844.65', '3054.1',
       '466.3', '3608.6', '7944.55', '1414.4', '3557.8', '2214.15',
       '2210.9', '2315.9', '548.15', '837.85', '2869.85', '764.75',
       '6500.2', '1430.1', '592.2', '965.8', '884.35', '890.35',
       '261.3', '493.45', '636.65', '573.65', '465.75', '914.1',
       '322.35', '306.6', '575.05', '860.75', '414.95', '1115.4',
       '773.25', '777.9', '307.3', '335.35', '905.2', '70.35',
       '1239.9', '571.45', '2041.7', '2351.95', '775.85', '2044.75',
       '477.5', '1051.1', '351.65', '350.35', '133.1', '475.05',
       '1239.4', '1132.55', '844.2', '354.6', '214.9', '918.15',
       '2683.05', '430.35', '1743.4', '779.9', '854.5', '5723.05',
       '490.15', '1334.3', '3382.25', '805.2', '3099.1', '2190.15',
       '2775.1', '545.2', '908.75', '2304.6', '2049.3', '486.85',
       '1990.75', '4931.3', '1818.1', '587.9', '326.15', '2137.8',
       '8672.8', '748.8', '387.35', '999.9', '43.9', '884.88',
       '83.95', '1778.65', '167.1', '239.3', '488.35', '334.65',
       '233.1', '405.3', '1318.1', '274.75', '2054.35', '2475.05',
       '6773.7', '133.35', '3394.8', '1082.35', '9095.1', '156.65',
       '2624.3', '2659.9', '273.35', '3181.7', '931.7', '1884.8',
       '1147.4', '100.35', '8352.9', '2594.9', '573.9', '2070.2',
       '1346.1', '544.15', '6766.95', '2001.1', '482.09', '1743.7',
       '1331.15', '805.35', '3370.35', '386.05', '1016.7', '2608.25',
       '571.3', '2384.8', '909.75', '1357.9', '364.05', '1024.1',
       '348.85', '462.75', '435.3', '615.45', '039.05', '1610.4',
       '5926.5', '2875.95', '2260.3', '911.1', '2686.05', '1743.3',
       '1873.15', '2790.25', '546.65', '2578.9', '3779.9', '790.75',
       '1165.9', '5305.7', '1686.5', '1863.1', '632.4', '363.4',
       '306.05', '6160.7', '2811.9', '6465.8', '2108.9', '7067.2',
       '687.3', '1593.25', '1469.9', '7267.05', '571.7', '1051.7',
       '500.4', '1747.9', '884.5', '20.2', '1906.5', '307.7',
       '6039.05', '475.3', '2097.15', '1774.25', '472.2', '1090.15',
       '1468.35', '1404.9', '6369.45', '462.3', '435.9', '1962.3',
       '775.5', '7995.7', '1743.41', '3790.55', '326.9', '2445.7',
       '132.45', '819.5', '1964.5', '7569.7', '3096.5', '8564.75',
       '1475.9', '2048.25', '618.3', '1475.45', '3624.1', '574.3',
       '3779.05', '2081.9', '6465.8', '2108.9', '7067.2', '687.3',
       '1593.25', '1469.9', '7267.05', '571.7', '1051.7', '500.4',
       '1747.9', '884.5', '20.2', '1906.5', '307.7', '6039.05',
       '475.3', '2097.15', '1774.25', '472.2', '1090.15', '1468.35',
       '1404.9', '6369.45', '462.3', '435.9', '1962.3', '775.5',
       '7995.7', '1743.41', '3790.55', '326.9', '2445.7', '132.45',
       '819.5', '1964.5', '7569.7', '3096.5', '8564.75', '1475.9',
       '2048.25', '618.3', '1475.45', '3624.1', '574.3', '3779.05',
       '2081.9', '6465.8', '2108.9', '7067.2', '687.3', '1593.25',
       '1469.9', '7267.05', '571.7', '1051.7', '500.4', '1747.9',
       '884.5', '20.2', '1906.5', '307.7', '6039.05', '475.3',
       '2097.15', '1774.25', '472.2', '1090.15', '1468.35', '1404.9',
       '6369.45', '462.3', '435.9', '1962.3', '775.5', '7995.7',
       '1743.41', '3790.55', '326.9', '2445.7', '132.45', '819.5',
       '1964.5', '7569.7', '3096.5', '8564.75', '1475.9', '2048.25',
       '618.3', '1475.45', '3624.1', '574.3', '3779.05', '2081.9',
       '6465.8', '2108.9', '7067.2', '687.3', '1593.25', '1469.9',
       '7267.05', '571.7', '1051.7', '500.4', '1747.9', '884.5',
       '20.2', '1906.5', '307.7', '6039.05', '475.3', '2097.15',
       '1774.25', '472.2', '1090.15', '1468.35', '1404.9', '6369.45',
       '462.3', '435.9', '1962.3', '775.5', '7995.7', '1743.41',
       '3790.55', '326.9', '2445.7', '132.45', '819.5', '1964.5',
       '7569.7', '3096.5', '8564.75', '1475.9', '2048.25', '618.3',
       '1475.45', '3624.1', '574.3', '3779.05', '2081.9', '6465.8',
       '2108.9', '7067.2', '687.3', '1593.25', '1469.9', '7267.05',
       '571.7', '1051.7', '500.4', '1747.9', '884.5', '20.2',
       '1906.5', '307.7', '6039.05', '475.3', '2097.15', '1774.25',
       '472.2', '1090.15', '1468.35', '1404.9', '6369.45', '462.3',
       '435.9', '1962.3', '775.5', '7995.7', '1743.41', '3790.55',
       '326.9', '2445.7', '132.45', '819.5', '1964.5', '7569.7',
       '3096.5', '8564.75', '1475.9', '2048.25', '618.3', '1475.45',
       '3624.1', '574.3', '3779.05', '2081.9', '6465.8', '2108.9',
       '7067.2', '687.3', '1593.25', '1469.9', '7267.05', '571.7',
       '1051.7', '500.4', '1747.9', '884.5', '20.2', '1906.5',
       '307.7', '6039.05', '475.3', '2097.15', '1774.25', '472.2',
       '1090.15', '1468.35', '1404.9', '6369.45', '462.3', '435.9',
       '1962.3', '775.5', '7995.7', '1743.41', '3790.55', '326.9',
       '2445.7', '132.45', '819.5', '1964.5', '7569.7', '3096.5',
       '8564.75', '1475.9', '2048.25', '618.3', '1475.45', '3624.1',
       '574.3', '3779.05', '2081
```