CHURN RATE

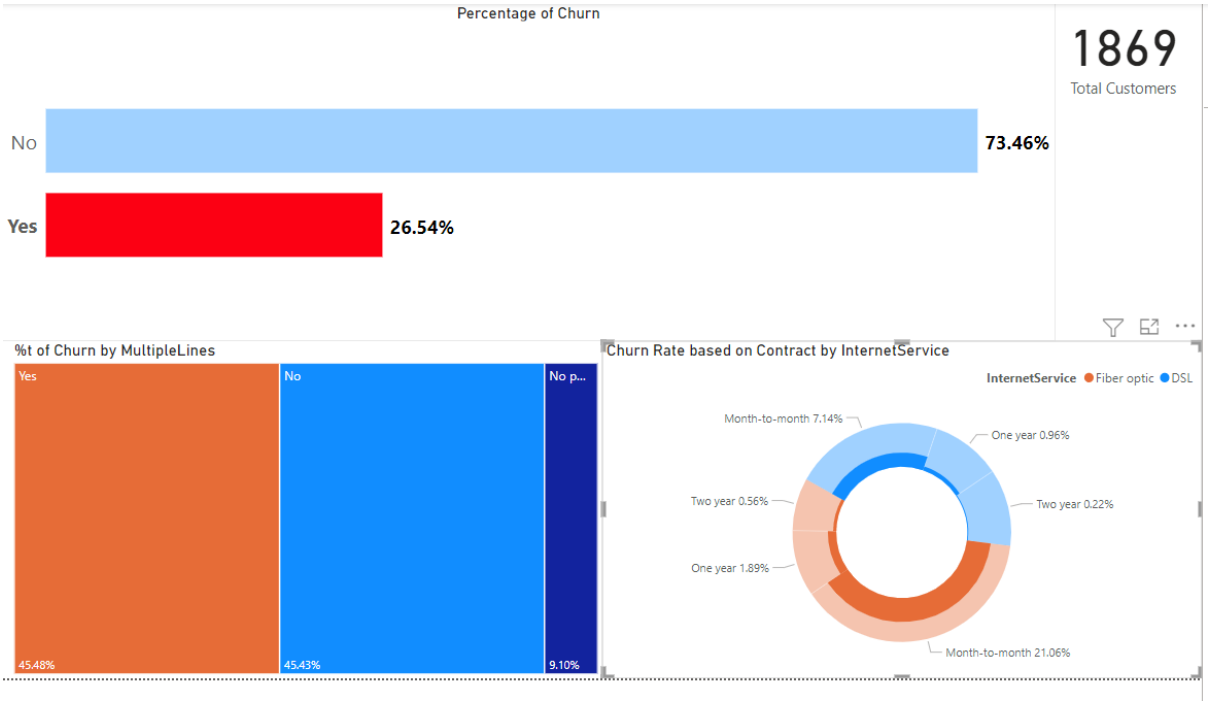
By:
Nikhilesh Vashisht

Review Churn

1. Provide brief description of method you used to bring data to your chosen tool.

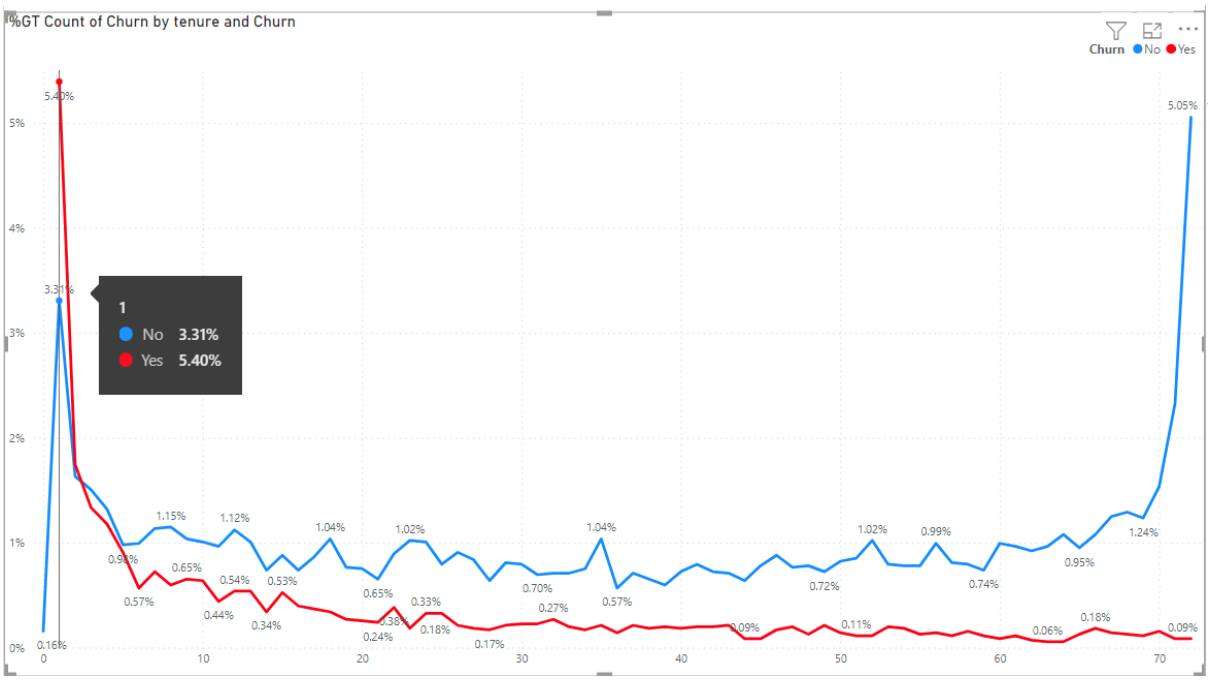
I have chosen PowerBi to make the visuals and python for churn prediction. The dataset was available on Kaggle (https://www.kaggle.com/datasets/blastchar/telco-customer-churn). I imported the CSV file on PowerBi using the import text file feature. And I imported the dataset in python using pandas. Pandas has a function (read_csv()) to import the CSV and convert it into data frame.

2. Chart or Graph for Churn Rate by month by product (Internet service – DSL vs Fiber Optic vs Other), multiple lines – Yes vs No. Briefly provide your observation. Highlight any cleaning job you needed to perform and what was your approach.



The chart above depicts that around 26.5% of the total customers are churned. Further, when we check on customers who avail of the internet service on a contract basis, we can see that around 21% of the customers who use fiber optics on a month-to-month contract are churned. There is almost an equal split of customers churned between the customers who have availed multiple lines is ~ 45%. Blank values from the dataset were removed as a data cleaning process

3. Explore the data and provide a summary of the Impact of Tenure on Churn with at least 1 chart to demonstrate your insight.

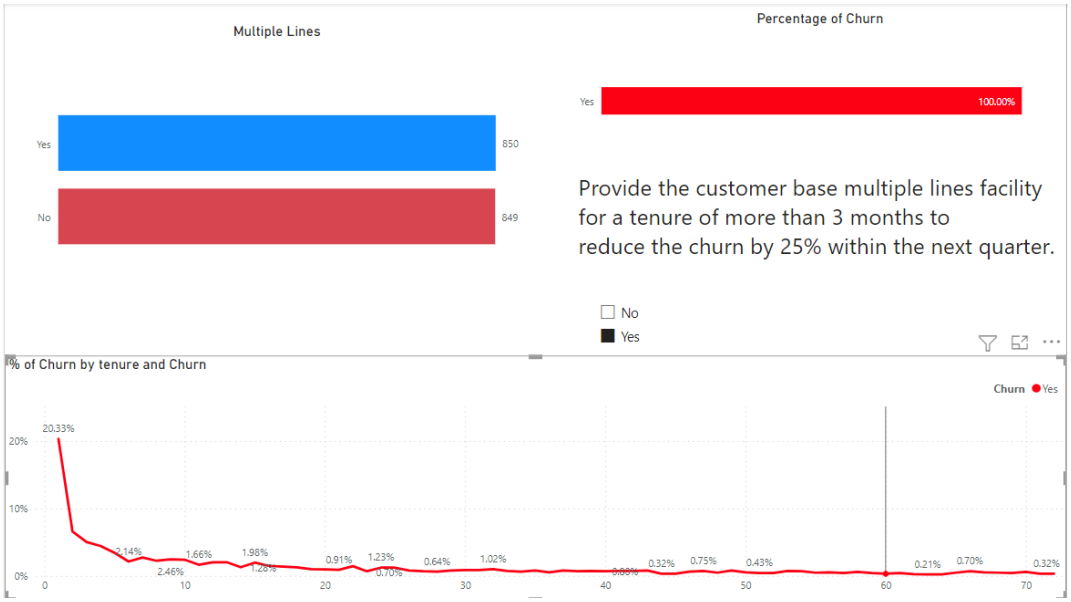


The churn rate is high when the tenure of the service availed is less. The churn rate for a customer with a tenure of less than 2 is greater than the ones who have been using the services for a long time. As we can see the blue line steeply rising at the end of the graph, this implies that customers who have a higher tenure is less likely to churn the services than the ones with shorter tenures.

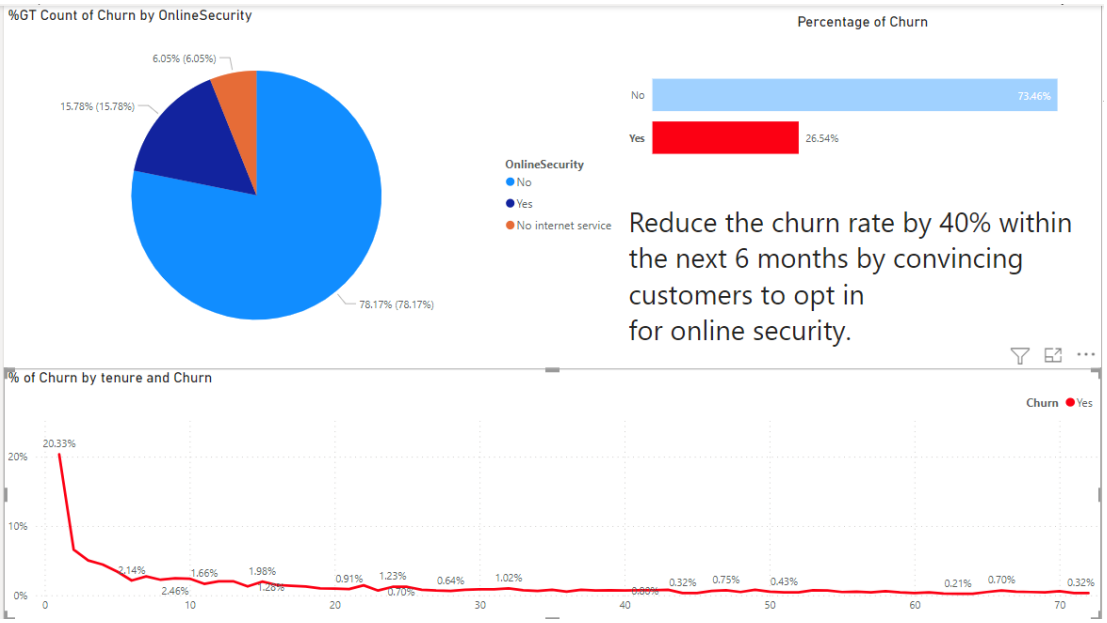
4. Given the company's primary objective is to reduce overall Churn - Come up with 2 examples of ANY SMART Objectives. Explain why these are SMART. At least 1 should be trackable using the data provided.

SMART objectives are Specific, Measurable, Achievable, Relevant, and Time-bound

SMART OBJECTIVE 1: Provide the customer base multiple lines facility for a tenure of more than 3 months to reduce the churn by 25% within the next quarter. This is a SMART objective as we are specific on where the improvement needs to be done. Showing how much churn rate we can decrease which is measurable. We are showing it is achievable by doing the said activity within the next quarter. It is relevant to the industry and the time frame is 3 months.



SMART OBJECTIVE 2: Reduce the churn rate by 40% within the next 6 months by convincing customers to opt-in for online security. This is a SMART objective as we are specific on where the improvement needs to be done. Showing how much churn rate we can decrease which is measurable. We are showing it is achievable by doing the said activity to reduce the churn rate. It is relevant to the industry and the time frame is 6 months.



2. Balance Scorecard and Visualization

1. You are asked to design a balance scorecard to track churn. Provide a mock up of a Balance Scorecard to track Churn. Identify what dataset is required or needs to be sourced to complete your proposed Balance Scorecard.

A balance scorecard is a tool that is used to track and measure the performance of a business or organization. It provides a holistic view of an organization's performance by looking at four key areas: financial performance, customer satisfaction, internal processes, and learning and growth.

To design a balance scorecard to track churn, we would need to identify the key metrics that are important for understanding and measuring churn. Some potential metrics that could be included in the balance scorecard are:

Customer churn rate: This is the percentage of customers who have stopped doing business with the company over a given period of time. This is an important metric to track because it provides a high-level overview of how many customers are leaving the company and can be used to identify trends and patterns over time.

Customer lifetime value: This is the estimated value of a customer over the course of their relationship with the company. This is important because it helps to understand how much revenue a customer is likely to generate for the company over time, which can be used to prioritize resources and efforts to retain valuable customers.

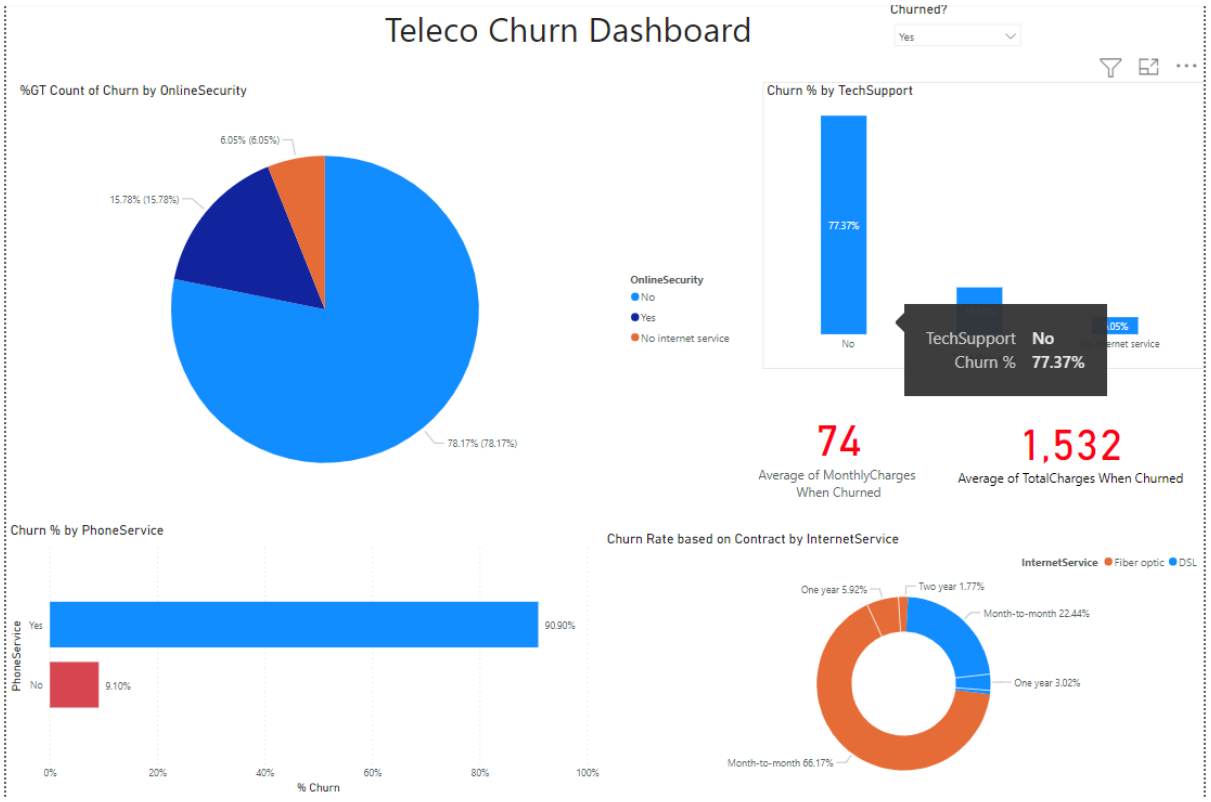
Customer satisfaction score: This is a measure of how happy customers are with the company's products or services. This is important because customer satisfaction is closely tied to churn, as unhappy customers are more likely to leave the company.

Net promoter score: This is a measure of how likely customers are to recommend the company's products or services to others. This is important because customers who are happy with the company and willing to recommend it to others are less likely to churn.

In addition to these metrics, the balance scorecard could also include metrics that relate to internal processes, such as the number of customer service inquiries or the percentage of customer service inquiries that are resolved within a certain timeframe. This can help to identify potential areas for improvement that could reduce churn.

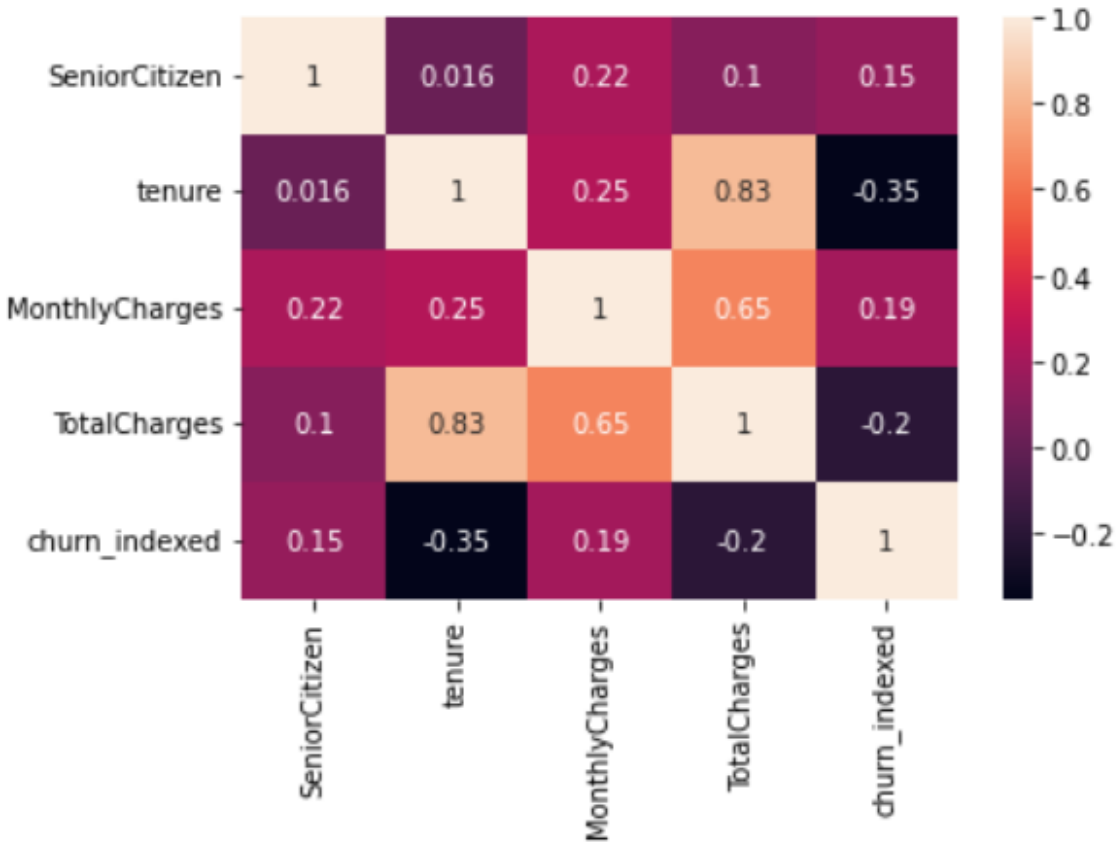
To complete the proposed balance scorecard, we would need to source data on the above metrics. This data could be collected through customer surveys, customer service logs, and other sources. By tracking these metrics over time, we can identify trends and patterns in churn and take action to reduce it.

2. Build a visualization dashboard which will provide an overview of Telco churn. Maximum 6 Charts. You can re-use charts from Part 1. You can use up to 2 dummy chart



3. Churn Prediction

I have used python and sklearn to predict the churn. As the column “Churn” which we are trying to classify was in a boolean value, I have indexed it to 1 and 0 using Label encoder class in SKLearn module. Next, I have dropped the blank and NA values in all the rows. As there were only 10-11 blank values, we can ignore them and delete them.
I have checked the correlation between numerical variables



It is clearly evident that they are not a strong correlation but still has a effect over the churn rate. This shows that even though these variables have a weak impact on the churn rate, there are other dependencies as well.
Can we predict whether the customer will churn the services based on Total Charges and Tenure. It is very common for a customer to discontinue the services based on the charges he might bear on them.
I have used 2 well known classification techniques Logistc Regression and KNeighborsClassifier in SKLearn.

Using Logistic regression we were able to get an accuracy of **73.5%**

jupyter final_assignment_enterprise_analytics Last Checkpoint: 4 hours ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel)

```
memory usage: 1.2+ MB
```

In [82]:
```
df['TotalCharges'].value_counts(dropna=False)
```

Out[82]:
```
20.20      11
19.75       9
20.05       8
19.90       8
19.65       8
           ..
6849.40     1
692.35      1
130.15      1
3211.90     1
6844.50     1
Name: TotalCharges, Length: 6530, dtype: int64
```

In [118]:
```
logisticRegr.fit(X_train, y_train)
```

Out[118]:
```
LogisticRegression()
```

In [119]:
```
predictions = logisticRegr.predict(X_test)
```

In [120]:
```
score = logisticRegr.score(X_test, y_test)
print(score)
```

```
0.7355136864557412
```

In [121]:
```
import sklearn.metrics as metrics
```

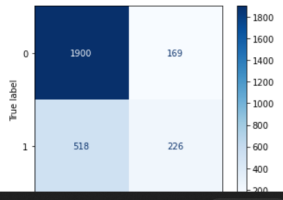
Whereas, using KNeighborsClassifier we were able to get around **75.5%** accuracy.

In [124]:
```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors=8)
knn.fit(X_train, y_train)
```

Out[124]:
```
KNeighborsClassifier(n_neighbors=8)
```

In [125]:
```
knn.score(X_test, y_test)
```

Out[125]:
```
0.7557767507998578
```

In [127]:
```
matrix = metrics.plot_confusion_matrix(knn, X_test, y_test, cmap=plt.cm.Blues)
plt.show()
```

```
C:\ProgramData\Anaconda\lib\site-packages\sklearn\utils\deprecation.py:87: FutureWarning: Function plot_confusion_matrix is dep
recated; Function `plot_confusion_matrix` is deprecated in 1.0 and will be removed in 1.2. Use one of the class methods: Confus
ionMatrixDisplay.from_predictions or ConfusionMatrixDisplay.from_estimator.
  warnings.warn(msg, category=FutureWarning)
```