<u>Telecom Churn Analysis</u>

Gaurav Yogeshwar

ABSTRACT

While customer acquisition and retention is a major concern for many, with the rapid growth of the telecom industry, service providers are keen to expand the customer base.

To meet the need to survive in the competitive environment, retaining existing customers has become a major challenge, it is cheaper to retain a loyal customer then acquire a new one.

Therefore, gathering insights from telecom industries can help predict customer engagement, whether they leave the company or not.

This paper focuses on exploratory data analysis to identify potential customers, classify them based on usage patterns, and analyze the data to discover key factors responsible for customer churn and come up with ways to ensure customer retention., so that the telecom industry can stabilize its market value to acquire its associated customers to take necessary steps.

Keywords: Churn Analysis, Churn Rate, Dimensionality reduction, Classified labels, One-hot encoding, Exploratory data analysis, Feature engineering.

DEFINITION OF CHURN

Customer churn is the term used when an existing customer stops using a company's services and/or stops buying their products. In other words, the customer chooses to cut his ties with the company.

Customer churn rate

is defined as the proportion of customers who stopped using a particular company's products or services during a definite time frame. Mathematically,

$$C(T) = \frac{A(T)}{B(T)} \times 100$$

where,

C represents the churn % for a time frame T,

A(T) represents the total number of customers after time T,

B(T) represents the total number of customers before time T.

PROBLEM STATEMENT

Orange S.A., formerly France Télécom S.A., is a French multinational telecommunications corporation. The Orange Telecom's Churn Dataset, consists of cleaned customer activity data (features), along with a churn label specifying whether a customer cancel the subscription.

Our goal is to explore and analyze the data to discover key factors responsible for customer churn and come up with ways/recommendations to ensure customer retention.

operate in. They, in turn, search for cabs from various service providers and provide the best option to their clients across available options. They have been in operation for a little less than a year now. During this period, they have captured surge pricing types from the service providers.

INTEGRAL METHODOLOGY

The entire Analysis is divided into the following phases: Dataset Description, Breakdown of Datasets, Examining the null values, Data Cleaning, pre-processing and Feature engineering followed by Exploratory Data Analysis by and applying different models. First, we collect the data from Alma's better dashboard. Thereafter we did basic data cleaning and data visualization. After visualizing the data set, we removed some unnecessary features and made it ready for analyzing the data set using different plots. Next, we conduct data modeling by using Bar plot graphs, violin plots, histogram, etc. Finally, we narrate the analysis results to provide a clear vision of the relationship among the areas of interest.

DATASET DESCRIPTION

Let's take a look at the data, which consists of :T elecomChurn.csv. There are 20 columns that give information about customers.

About Dataset Most regularly a dataset relates to the matter of the single database table, or the single factual information framework, where each segment of the table speaks to a specific variable, and each column compares to a given individual from the informational collection being referred to. In this project, I have analyzed all these various columns of the dataset.

Column	**Explanation**
Account Length	Length of The Account
Number vmail messages	Number of Voicemail Messages
Total day minutes	Total Number of Minutes Spent By Customers in Day (before evening).
Total day calls	Total Number of Calls made by Customer in Day (before evening).
Total day charge	Total Charge to the Customers in day time.
Total eve minutes	Number of Minutes Spent By Customers in Evening.
Total eve calls	Total Number of Calls made by Customer in Evening
Total eve charge	Total Charge to the Customers at Night.
Total intl minutes	Total Number of Minutes Spent By Customers in international calls.
Total intl calls	Total Number of International calls made by Customers.
Total intl charge	Total charge to Customers in international calls.
Customer service calls	Total number of Calls by Customer to service Center.
State	51 Unique States in United States of America
Area Code	3 unique area codes
International Plans	Yes Indicate International Plan is Present and No Indicates no subscription for International Plan
Voicemail Plan	Yes Indicates Voicemail Plan is Present and No Indicates no subscription for Voicemail Plan

Churn:	Whether he customer churned or not (True or False)

BREAKDOWN OF DATASETS

In order to go ahead for data visualization upon key factors we need to go for certain extra steps before proceeding to the main segment. In this part we are going with the following steps:

1. Mounting drive and installing a few libraries.
2. Importing Analytical necessary library classes for future analysis.
3. Reading the csv data file from Google drive.
4. Visualizing all the columns of the respective Data frame.
5. Viewing all data information.
6. Checking the Unique values in the columns and duplicate rows in the dataframe.
7. Combination of some feature to get new features
8. Get a quick statistical summary of the numeric columns.
9. Converting the data types to similar objects as the Analysis Demands.
10. Set default grid type, font size and palette colors.
11. Eradicating special characters from the dataset columns.

EXAMINING NULL VALUES

The most critical thing from which we can draw some observations is Dataset, however data comes with unexpected values too i.e. sometimes it may be Null or missing in other words the space might be blank. But in this dataset we are working in there is no missing value.

DATA MINING

Data mining is the process of extracting and discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems.

DATA CLEANING

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

It is one of the most essential subtask of any data science project. Although it can be a very tedious process, it's worth should never be undermined.

We know that area code is not a hierarchical variable and data has 3 distinct unique values without any hierarchy, better convert it from numeric to object data type.

DATA PREPROCESSING AND FEATURE ENGINEERING.

It is the process of using domain knowledge to extract features from raw data via data mining technique.

There are Three general approaches:

- Extracting Information
- Combining Information
- Transforming Information

Extracting Information

Creating a new feature by extracting any hidden information from the given data.

I created a new feature which is a churn rate for each customer according to his state.

Combining Information

Creating a new feature by combining two or more features by some mathematical, logical or any other operation.

I created some new feature ie. charge per minute by dividing total charges to total minutes for different time zones as well as for international calls.

Transforming Information

Transform one type of data into different types of data that contain the same information. eg..

1. One-hot encoding.
2. Ordinal / numeric encoding.

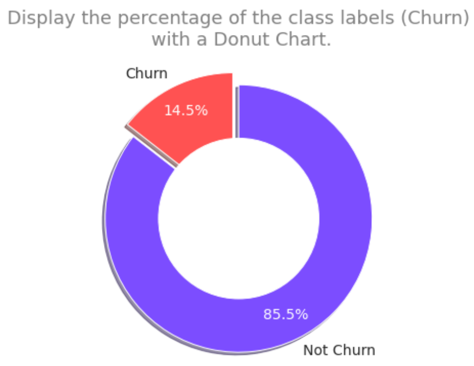
Before performing any mathematical operation with the categorical data we have done a hot encoding on it. eg. on Area code, International Plan and Voicemail plan.

DATA VISUALIZATIONS

Orange S.A., formerly France Télécom S.A. is a French multinational telecommunications corporation. It has 266 million customers worldwide.
We have 3,333 customer datasets with certain attributes, and using this data we will do data visualization with python libraries and get informative insights that will eventually help with some recommendations to prevent customer churn.

Observation-1

Before improving the churn rate of the customers, we have to know what the churn rate of our customers is. Let's check the balance of class labels (churning) with a count plot, and we'll use a donut chart to visualize the percentages..



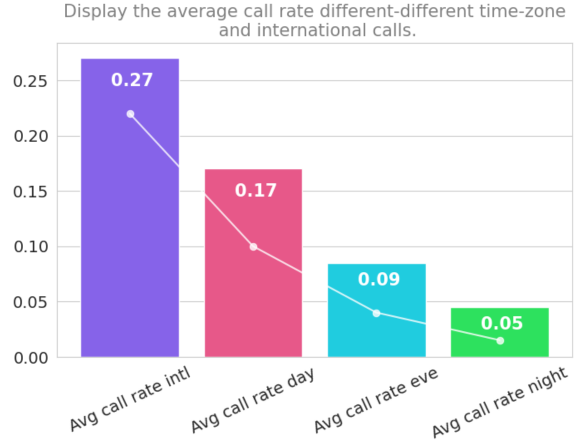
Display the percentage of the class labels (Churn) with a Donut Chart.

From the customer churn perspective it's still high.Our job will be to reduce it as much as possible, for every single churn customer we have 5.9 retain customers.

Observation-2

We have information for day, night and evening calls as well as international calls according to different time zones of the customers, as we mentioned in the data description.
Using the same features, we will make some new features and check the rates of each and every year so that we can do a comparative study of them.

$$\text{Average call Rate} = \frac{Total\ charge}{Total\ minutes}$$



Display the average call rate different-different time-zone and international calls.

We know that international calls are the most expensive because we are talking in other countries but the thing to note here is that the day time calls are very expensive compared to evening and nights.

And also we observe that customers prefer to talk more at night and evening than during the day.

Observation-3

By plotting the histogram of each of the numerical features we found that most of the features approximately follow a normal distribution, except total international calls and Customer service calls are positively skewed.

Another thing we noticed is that most people do not send voice messages.

Observation-4

Correlation check

To get the correlation matrix, we do the hot encoding of the Area code. Since we changed the international plan and voice message plan to boolean datatype, there was no need to encode those features and we will drop State column because there are 51 state and if we do one hot encoding of all 51 state then it will create total 50 new column and bunch of zeros in our dataframe.

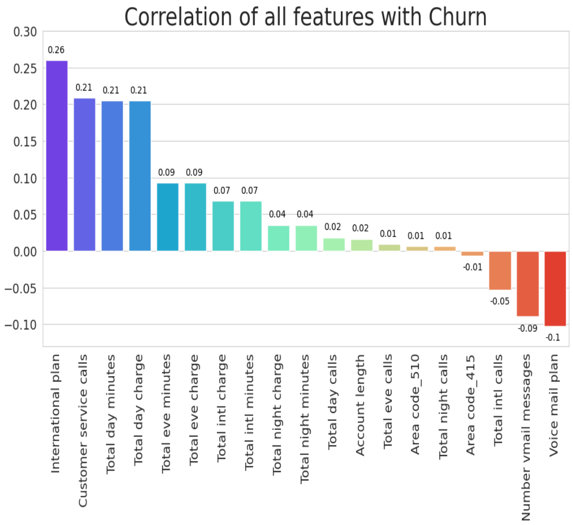
After getting the correlation matrix we will plot them in a heat map for visualization.

We found that the number of Voicemail is highly correlated with voicemail plans.

Total day minutes is perfectly correlated with total day charge, and it follows the same pattern in evening, nights minutes and international minutes with charges.

In this matrix, the column with correlation with churn was the most important, so we extracted it and got its bar plot.

From this plot we got a brief idea of which feature is responsible for churn, which is responsible for retention and which feature has no effect on churn.

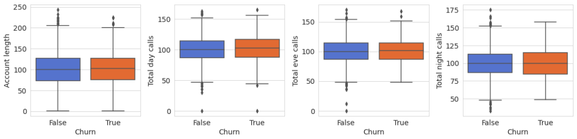


Then I plot all the numeric features with churn through boxplot so that we get a brief idea, And we will notice that this plot also matches with our bar plot.

Observation-5

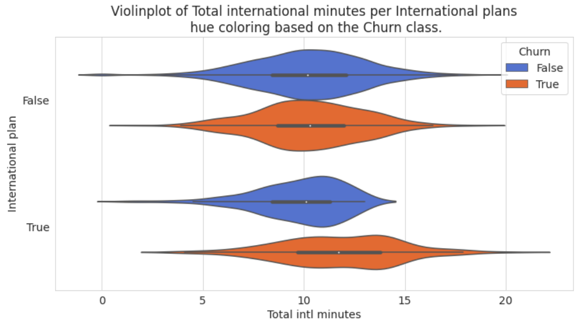
First we plot behavior with features that have no correlation with the target variable.

We analyzed the account length, total calls to all three categories through box-plot and histogram and then analyzed area code through bar-plot and as we have seen in the Correlation plot these have no effect on churn.



Observation-6

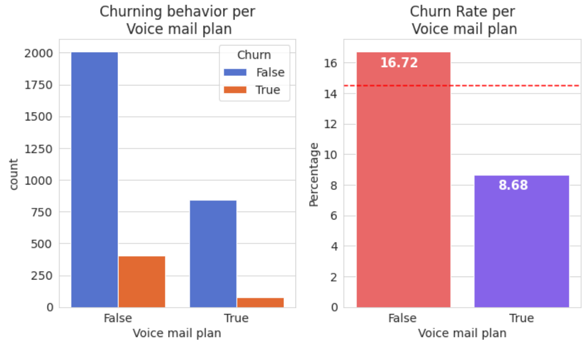
We noticed that the International plan deals with the most churn, so we start with that, and add up the total international calls and minutes as well.



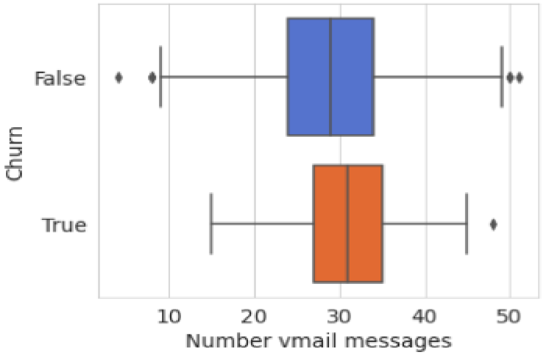
In the Violinplot, we have seen that customers who talk more along with taking an international plan are more likely to churn.

Observation-7

Our next analysis after the International plan will be with the Voicemail plan.



We saw in the correlation plot that the voicemail scheme is negatively correlated with churn, and now we found the same in the barplot.



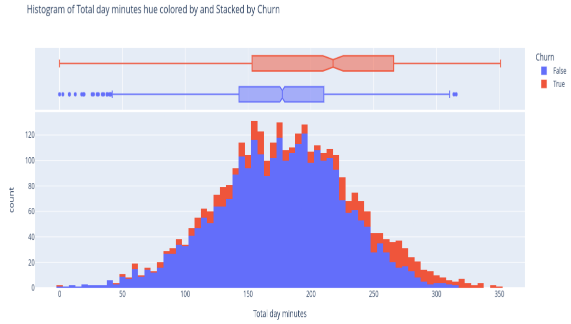
Although if seen, there is less chance of churn in people who have taken voicemail plans, but when the same people send more voicemail plans, then more churn has been seen in them.

Observation-8

Let's dive deep down into total day, evening and night minutes.

We saw in the box plot and histogram that there is not that much difference of night calls, there is a slight difference in evening calls, but there is a big difference in day calls.

Those who talk more during the day have a higher churn rate than those who talk less.

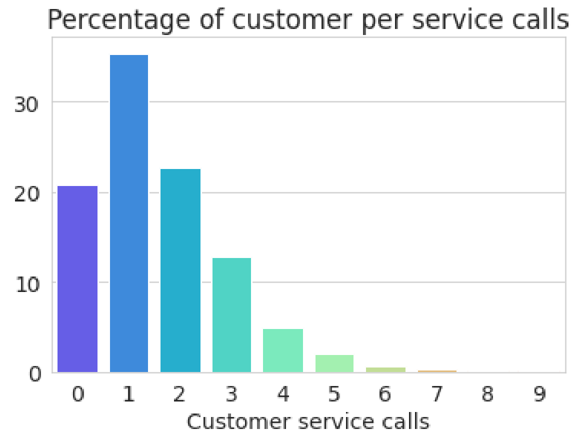


It can be seen that the median value is 217.6 minutes for churning customer and 177.2 minutes for non- churning customers.

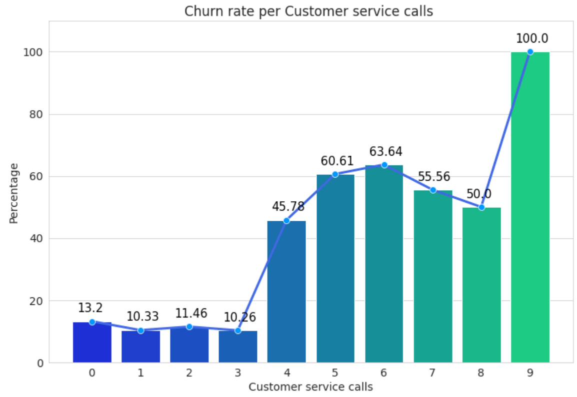
from this, we can conclude that our day plans are not that good and unable to satisfy our customers.

Observation-9

Now we will see what is the effect of customer service calls on churn.



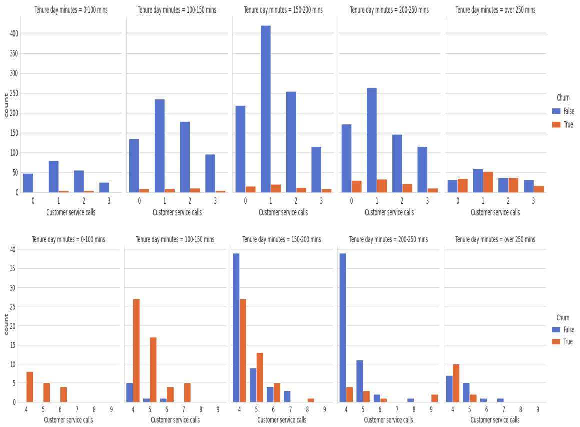
The number of customer service calls greater than 3 is significantly low, and also the churn rate increases significantly for 4 or more calls to the customer service.



Then we created a broader Cohort Groups based on the total day minutes column values because this is field that we are more concerned, we create a new column called Tenure day minutes that creates 5 separate categories:

- ➢ 0-100 minutes
- ➢ 100-160 minutes
- ➢ 160-220 minutes
- ➢ 220-250 minutes
- ➢ above 250 minutes

Now we create a grid of Count Plots showing counts per Service calls, separated out by Tenure day minutes and colored by the Churn hue.

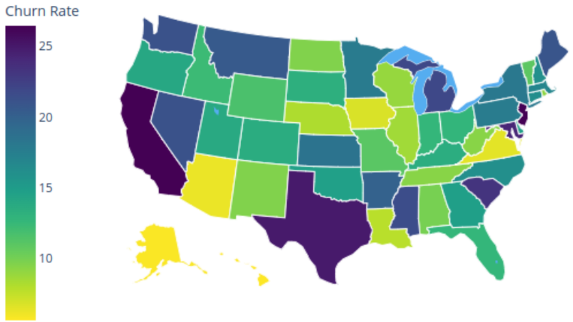


We have seen that customers who talk less in a day (below 200) do not churn so easily until their problem is solved, they are trying to solve their problems by talking to the service center as much as possible, even after that if their problems are not resolved then they churn.

And customers who talk more during the day have higher churn rate even in low service calls, which means they are not happy with the high call rate during the day.

Observation-10

Now we have done data preprocessing for the Churning rate with various states of the USA , then plot choromap.



This behavior shows that some regions have high churn rate and some have low.

One of the reasons for this could be that the service in those areas is not good, other, there may be other competitors in those areas that provide better service.

Then I created a special dataframe for only those areas where the Churan rate is more than 20.

We observed that except a few, almost all features follow the same churning pattern but with amplified magnitude.

The people of these areas are not as happy with the night facilities as the people of the whole country. the company can launch better night calls tariff for some selected states.

CURSE OF DIMENSIONALITY

Curse of Dimensionality refers to a set of problems that arise when working with high-dimensional data. The dimension of a dataset corresponds to the number of attributes/features that exist in a dataset. A dataset with a large number of attributes.

Some of the difficulties that come with high dimensional data manifest during analyzing or visualizing the data to identify patterns, and some manifest while training machine learning models.

DIMENSIONALITY REDUCTION

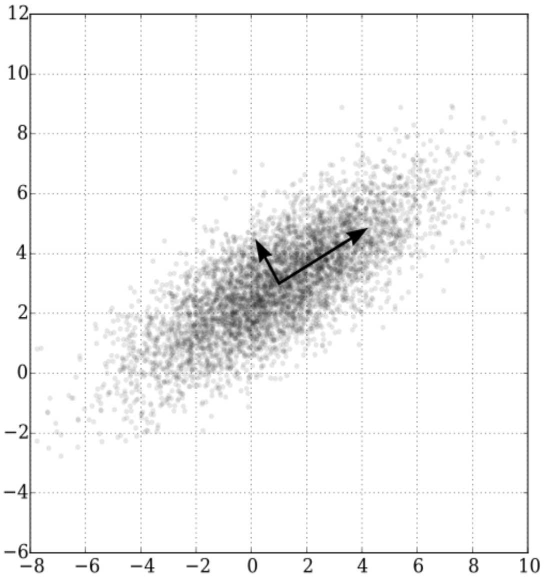
Dimensionality reduction, or dimension reduction, is the transformation of data from a high-dimensional space into a low-dimensional space so that the low-dimensional representation retains some meaningful properties of the original data, ideally close to its intrinsic dimension.

FOR VISUALIZATION

One of the most important aspects of Dimensionality reduction, it is Data Visualization. Having to drop the dimensionality down to two or three, make it possible to visualize the data on a 2d or 3d plot, meaning important insights can be gained by analyzing these patterns in terms of clusters and much more.

PRINCIPAL COMPONENT ANALYSIS

Principal component analysis, or PCA, is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.



STEP 1: STANDARDIZATION

Standardization is a scaling technique where the values are centered around the mean with a unit standard deviation. This means that the mean of the attribute becomes zero and the resultant distribution has a unit standard deviation.

Mathematically, this can be done by subtracting the mean and dividing by the standard deviation for each value of each variable.

$$z = \frac{x - \mu_x}{\sigma_x}$$

z *is transformed data.*

μ_x is the mean of feature x.

σ_x is the Standard deviation of feature x.

STEP 2: COVARIANCE MATRIX COMPUTATION

The covariance matrix is a n × n symmetric matrix (where n is the number of dimensions) that has as entries the covariances associated with all possible pairs of the initial variables.

Covariance Matrix Formula,

$$\begin{bmatrix} Var(X_1) & Cov(X_1, X_2) & .. & .. & Cov(X_1, X_n) \\ Cov(X_2, X_1) & Var(X_2) & .. & .. & Cov(X_2, X_n) \\ : & : & .. & .. & : \\ : & : & .. & .. & : \\ Cov(X_1, X_n) & Cov(X_2, X_n) & .. & .. & Var(X_n) \end{bmatrix}$$

STEP 3: COMPUTE THE EIGENVECTORS AND EIGENVALUES OF THE COVARIANCE MATRIX TO IDENTIFY THE PRINCIPAL COMPONENTS

Finding eigenvalues for matrix Covariance Matrix A using equation

$$|A - \lambda I| = 0$$

A is the matrix for which we want to find eigenvalue.
λ is variable for eigenvalue in the equation.
I is an identity matrix of the same order of A.

After finding eigenvalues, we have to find eigenvectors using equation,

$$AX = \lambda X$$

Using the above equation we can find n eigenvectors for n eigenvalues.

Out of n eigenvectors we choose only those p numbers of eigenvectors for which we have largest eigenvalues.

Now these p eigenvectors are our new dimensions and we've reduced it from n to p dimension.

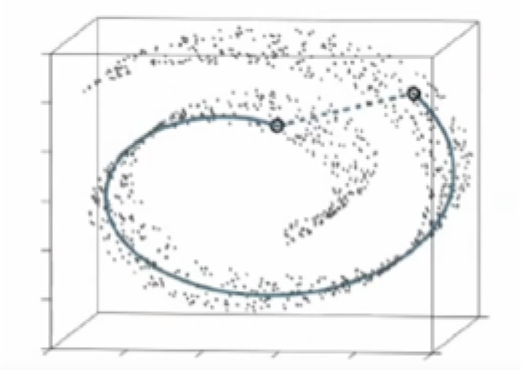
KERNEL PCA:

PCA is a linear method. That is it can only be applied to datasets which are linearly separable. It does an excellent job for datasets, which are linearly separable. But, if we use it for non-linear datasets, we might get a result which may not be the optimal dimensionality reduction. Kernel PCA uses a kernel function to project a dataset into a higher dimensional feature space, where it is linearly separable. It is similar to the idea of Support Vector Machines.

There are various kernel methods like linear, polynomial, and gaussian.

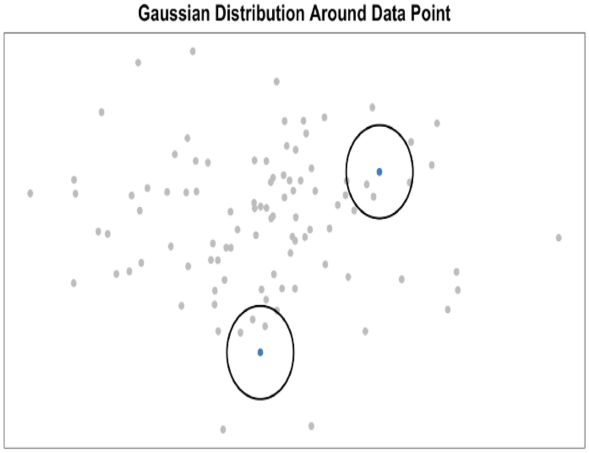
t-SNE(t-distributed Stochastic Neighbor Embedding)

t-SNE differs from PCA by preserving only small pairwise distances or local similarities whereas PCA is concerned with preserving large pairwise distances to maximize variance. Laurens illustrates the PCA and t-SNE approach pretty well using the Swiss Roll dataset in Figure 1. You can see that due to the non-linearity of this toy dataset (manifold) and preserving large distances that PCA would incorrectly preserve the structure of the data.

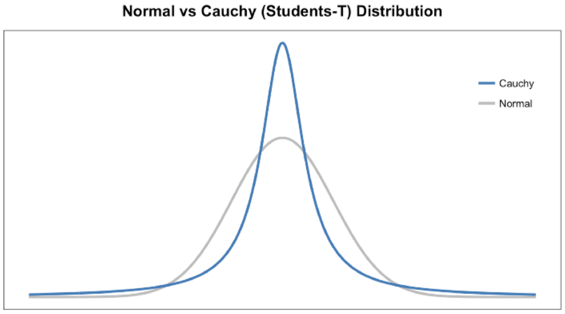


The t-SNE algorithm calculates a similarity measure between pairs of instances in the high dimensional space and in the low dimensional space. It then tries to optimize these two similarity measures using a cost function. Let's break that down into 3 basic steps.

Step 1, measure similarities between points in the high dimensional space. Think of a bunch of data points scattered on a 2D space (Figure). For each data point (xi) we'll center a Gaussian distribution over that point. Then we measure the density of all points (xj) under that Gaussian distribution. Then renormalize for all points. This gives us a set of probabilities (Pij) for all points. Those probabilities are proportional to the similarities. All that means is, if data points x1 and x2 have equal values under this gaussian circle then their proportions and similarities are equal and hence you have local similarities in the structure of this high-dimensional space. The Gaussian distribution or circle can be manipulated using what's called perplexity, which influences the variance of the distribution (circle size) and essentially the number of nearest neighbors. Normal range for perplexity is between 5 and 50 .



Step 2 is similar to step 1, but instead of using a Gaussian distribution you use a Student t-distribution with one degree of freedom, which is also known as the Cauchy distribution (Figure 3). This gives us a second set of probabilities (Qij) in the low dimensional space. As you can see the Student t-distribution has heavier tails than the normal distribution. The heavy tails allow for better modeling of far apart distances.



The last step is that we want these set of probabilities from the low-dimensional space (Qij) to reflect those of the high dimensional space (Pij) as best as possible. We want the two map structures to be similar. We measure the difference between the probability distributions of the two-dimensional spaces using Kullback-Liebler divergence (KL). I won't get too much into KL except that it is an asymmetrical approach that efficiently compares large Pij and Qij values. Finally, we use gradient descent to minimize our KL cost function.

Observation-11

For visual perspective I am reducing the dataset to 2 dimension

As I told earlier that we had calculated the churn rate of each state, now we will pass the churn rate data of every customer corresponding to his state.

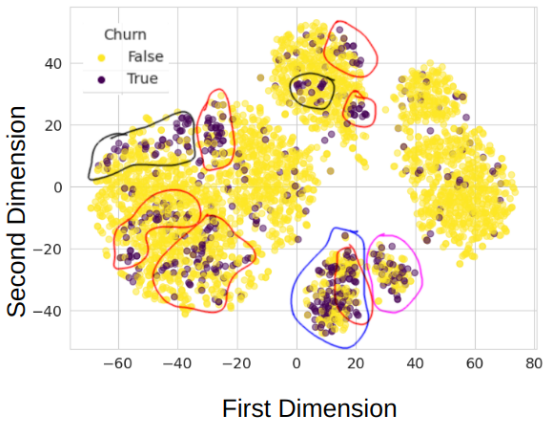
Although it is a personal choice to pass the churn_rate_of_state feature, we can ignore this column while passing the data to the model like we will ignore the state column.

Then we do standardization of the dataset by fit_transform into StandardScaler.

We will pass the scaled data in various kernel PCA like linear, poly, rbf, sigmoid, cosine and t-SNE to reduce the data set into 2 dimensions.

Out of all the decomposition we can see that t-SNE is the best decomposition for our data

When we look at it in the Reduced dimension, we will find that all those dissatisfied customers who have churn are grouped together, and are unhappy with a particular policy.



Blue marked region are customers with International plans but no voicemail plan. These are the customers that are not satisfied by our international plans.

Magenta market region customers with both International and Voicemail plans. These are the customers that are not satisfied by either international plans or Voice mail plans or both.

Red marked regions are the customers who talk more than 230 minutes in a dayThey are not satisfied by our high call rate of the day.

Black marked region are those users that have called 4 or more times in the service center. Their problems are not properly resolved.

Recommendations

After analyzing the above data set we came up with four recommendations.

Recommendations that help to retain the customers.

First Suggestion

Many customers do international conversations, but most of them have not taken any international plan. Do they dislike our international plan? And those who have taken it, their churn rate is very high.

We need a very attractive international plan, which can provide satisfaction to the customers making international calls.

Second Suggestion

By the way, the customer who has taken the voicemail plan has seen a lower Churn rate, which means that our voicemail plan is good, But customers who send more emails have seen higher churn rates.

We need a new voicemail plan along with the old one which is specially designed keeping in mind the more voicemail senders.

Third Suggestion

We noticed that people who call customer service more than 3 times, their churn rate is drastically increased, which means despite calling the service center so many times, their problem could not be solved.

Need to improve the feedback system that doesn't ignore customer problems.

Fourth Suggestion

We noticed that the call rate of the day is high, so the customers who talk more during the day have seen higher churn rate than others. They have higher churn rate even in service calls less than 3, which means they do not like any of our day's plans, they do not want to listen to any excuses.

Need to introduce better tariff plans for day calling which is specially designed for users to talk too much in a day.

Fifth suggestion

The churn rate has also been seen high among the callers in the evening, but it is less than the day.

Therefore, there is a need for improvement in the evening tariff plans as well.

Sixth suggestion

In the states where the churn rate is more than 20, there is a slight dislike for the night calls.

The company can launch better night calls tariff which is specifically designed for specific states.

CONCLUSION

The dataset contains immense possibilities to improve business values and have a positive impact. It is not limited to the problem taken into consideration for this project.

The company can increase its profit keeping in view some of these certain recommendations which have been suggested above.

References-

- **Builitin**
- **tds Towards Data Science**
- **Visualizing Data using t-SNE Journal**