CUSTOMER CHURN IN AIRLINES

ABSTRACT:

This report provides an analysis on customer churn out and retention on the airlines industry. This research has been done because of the rapid increasing of airline service usage. In today’s time where the competition is strict, knowing how to acquire and retained customers is very important. As customer churn is a serious problem, which appears to be a threat to the company in the long run, if no precautions are taken. It shows the percentage of the customers who stop doing business with a company for different reasons-unsatisfied needs, high prices, bad service etc. The high rate of customer churn isn’t just a problem that needs to be solved, but also a symptom showing the company doesn’t know its clients, their needs and how to satisfy them. Hence, these project objectives are to find out what are the factors that affect customer’s churn.

TABLE OF CONTENTS:

Abstract

Objective

Descriptive and Diagnostic Analysis

Different types of churn

Various scenarios for customer churn

Sample Analysis

-Different model evaluations

Understanding the problem statement

Preparing the data

Data Cleansing

Exploratory Data Analysis

-Business moments

Correlation

-Pearson

-Contingency table

-chi-square test

-Spearman

Visualizations

Building the model

Validating the model (training and testing)

Conclusion

OBJECTIVE:

To predict the reasons for the customer churn and also on what factors it is mostly dependent in the Airlines industry.

DESCRIPTIVE AND DIAGNOSTIC ANALYSIS:

Most Important factors that customer’s look out while selecting the Airlines.

1. low prices

2. Available Non Stop Flights

3. Convenient Routes and Schedules

4. Quality customer service and Onboard Service

5. Ease of e-ticketing and offers

6. Boarding and clearance Time

7. Capacity of Luggage

8. Past Experience.

DIFFERENT TYPES OF CHURN:

1. INVOLUNTARY CHURN: This occurs when customers fail to pay their bill and as a result ,the provider terminates the service.

2. INEVITABLE CHURN: This occurs when customers die or migrate that results in not considering him into the market.

Above two categories together contribute less than 5% and airlines cannot implement any strategies in order to decrease the churn rate.

3. VOLUNTARY CHURN: This occurs when customer prefer to switch into another service because of extra features or benefits.

Analyzing this strategies plays a major role in deciding the profits for a firm and thereby increasing the retention on the Airlines industry.

VARIOUS SCENARIOS FOR CUSTOMER CHURN:  
1.BAGGAGE ISSUES: Many business flyers travel for the purpose of convenience. Hence, they don’t want any hassle i.e; they don’t want to drag their bags from one terminal to another.

-LOST BAGGAGE: Lost baggage is regarded by many customers as a violation of their personal property, and causes major upset to many.

-BAGGAGE CAPACITY: Maximum of the people who are settlers in the destination have so much luggage to carry and if the baggage capacity is less, there is a chance of customer’s iteration

2. PRICING: If the difference between services offered by the company and those being offered in market is perceived to be small, then the customers mostly choose one with low price.

-EXPENSIVE TICKETS: Tickets can be expensive due to many reasons(High fuel rate, fastest etc) but as everyone cannot afford it, they may drop off. For example-some airlines provides free food for small routes also which makes their tickets costly and generally travellers do not care much about food all they need is cheap flights.

-AVAILABILITY OF DISCOUNTS: Customers mostly look after the discounts,by the services that airlines can provide.

3. POOR SERVICE: Due to security procedures at check-in, cramped seating, inconvenient schedules etc. when the list of customer complaints are more.

-DELAYED FLIGHTS: Time is more valuable than anything in the world. Many people choose airway for the time constraint. If we have delayed flights very common, then definitely it makes the no.1 reason for standing out of customers.

-CABIN CLEANLINESS: Today’s airline customers seem to be much more aware about cleanliness and hygiene. With increased aircraft utilization an important factor for airlines, shorter turnaround times often don’t allow sufficient time for thorough cleaning. But there are large complaints about poor cabin condition, dirty aircraft and cleanliness of cabin washrooms being a prime source of angst.

-UNCOMFORTABLE SEATS: A common cause for complaint, almost evenly distributed between long haul and short haul airline seats. Whilst many airlines have introduced new slimline seats that use a smaller seat pitch and are supposed to offer equal or better “personal living space” for customers.

4. FREQUENCY: Metropolitan cities like Delhi, Hyderabad have flights with more frequencies. If in case the customer misses a flight, he can have alternate immediate flight in a less time. Therefore, they prefer airline with more frequency of flights.

5. CONNECTIVITY: Customers choose airline having many places connected. This means you need to have connectivity all the nearby towns connected with you to have better and more number of customers.

6. SAFETY AND SECURITY: Safety and security always place a major in travelling from place to another, customers mostly check for secured airlines by considering the previous major past incidents inorder to avoid all the disturbances.

-PLANE CRASHES AND ACCIDENTS: People avoid negativity in spread things. Similarly, if an airline has repetitive no of plane crashes and accidents customers generally step back.

7. AIRLINE MEALS: Most of the customer complaints about airline meals refers to special order meals; like vegetarian, gluten free, child meals etc. Prime discontent is when pre-ordered meals have not been loaded onto the flight, or they have been served to the wrong customer.

8. UNAVAILABILITY OF SEATS: If it is a mismatch in demand and availability of the airlines. For instance, there was an over provisioning of wide body aircraft where it didn’t have required number of customers and viceversa.

9. EXTERNAL EVENTS: Airline industry is particularly vulnerable to external events such as terrorism, political instabilities and natural disaster, which can drastically affect their operations and passenger demand.

10. REVIEWS AND RATINGS:4 out of 5 people go with the rating and review of the previous customer before doing anything. Therefore, it is really important to maintain healthy relation and make the customer feel satisfied as much as possible.

11.OTHER FACTORS:

-Non-seasonal time: Obviously at a busy time which could be holidays, sporting or religious events then tickets sell very quick and this means a flight would seem more expensive in that case.

-More attractive features(fare, facilities )in other airlines when compared with each other, as the customer mostly crosschecks more airlines in order to book for the best one with maximum benefits.

-No proper Business strategy: Most of top management are still following the old way of administration while things have changed a lot in recent years which are unacceptable and unimaginable. They should try in different perspective in order to gain customer attention.

-not very well practice of marketing: It is like not promoting their airlines up to the point that includes not having any clear insights why customer should choose their company, not increase in existing customer base, monitoring results by previous ones.

SAMPLE ANALYSIS:

As one of the Airlines industry was losing its customers consistently and to predict the customer churn, we have considered one of the sample data set in order to leverage customer analytics to

to minimize the gaps in quality of service and improve offerings to ensure long-term customer relationship.

After analyzing the sample data, we can tell that customers left in flight distance, Gate location, Departure delay and arrival delay in minutes are more in case than the other factors. Plotting the actual target variable using pie chart, has got exactly 50% of churners and non-churner, that means we can grab more efficient in sights in the data.



Three different algorithms (logistic regression, random forest and decision tree) are used for model evaluation, as it is the process of choosing between models, tuning parameters and features. Better evaluation process lead to better, more accurate models in the applications.

MODEL: LOGISTIC REGRESSION

Pie chart:



Confusion matrix:

[[1630 364]

[ 336 1662]]

Roc curve:



MODEL: RANDOM FOREST

Pie chart: 

Confusion matrix:

[[1870 124]

[ 93 1905]]

Bar graph:



ROC Curve:



**MODEL:DECISION TREE**

Visualization of decision tree

****

**Pie chart:**

****

Confusion matrix:

[[1731 263]

[ 297 1701]]

Roc curve:

****

PROBLEM STATEMENT:

Acquiring a new customer is equally important with retention to the existing customer. As companies are struggling to retain customers because of cut-throat competition in airlines, customer churn plays a vital role in this rivalry. By this we are going to predict the churn rate and the elements affecting over it.

PREPARING THE DATA:

According to the recent study, preparing data and sampling requires 80% of whole time and effort. A data sample was given, around 10,000 records of 9 fields. In predictive+ diagnostic analytics, there is one more step involved in data preparation- creating the target variable. Here ( is\_ inactive )is the dependent or target variable, whereas rest of fields considered as independent variable.

DATA CLEANSING:

Missing Values: In this missing or null values are replaced in dataset with the mean values in case of continuous data and mode values in case of categorical data.

As per the given data set, there are no null values.

Ambiguous Values: In case of any cells with ambiguous entries, for example (calfornia, CA, cal) kind of different identifiers represent the ambiguous entries.

As per the given data set, there are no ambiguous entries.

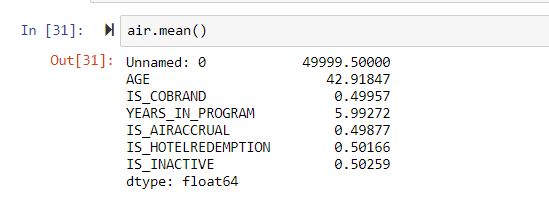
Unique Values: The set of values in a column are defined in another column of table with unique values.

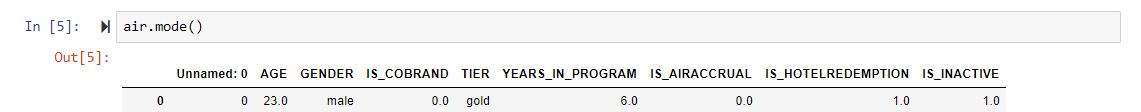
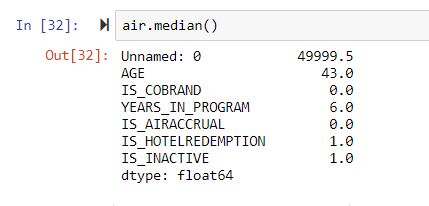
We have gender: male, female and unknown. we replace unknown with mode!!

EXPLORATORY DATA ANALYSIS:

1st Business moment:

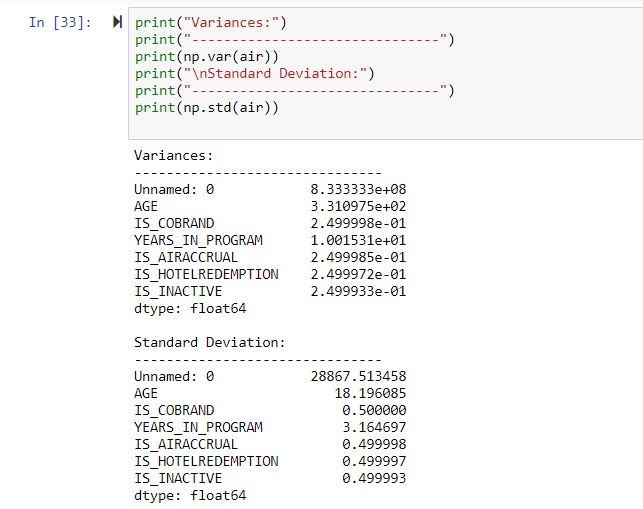
MEAN:

MEDIAN:

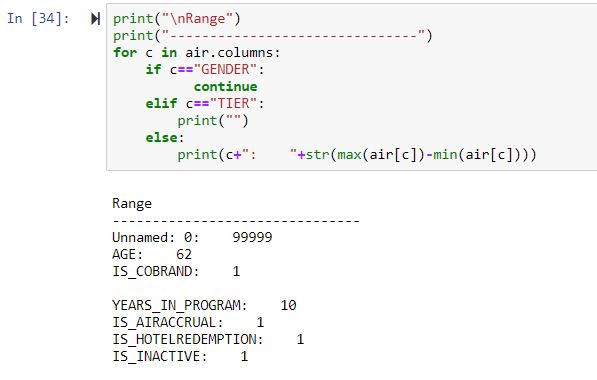
MODE: 

2ND Business Moment:

Standard deviation and variance:

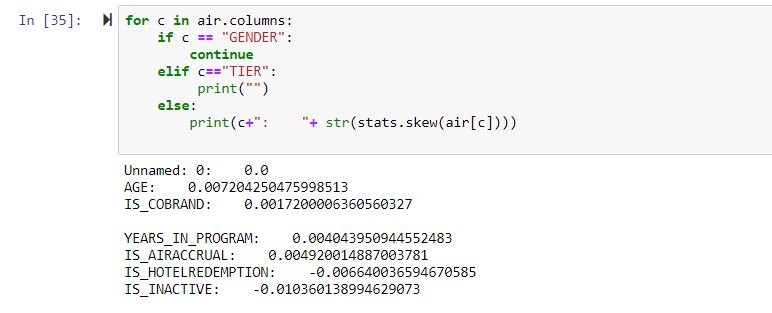


Range:



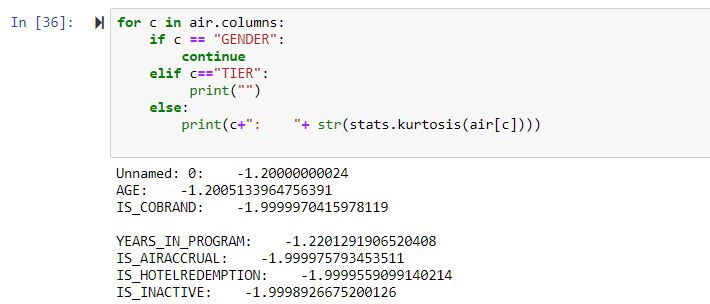
3rd Business moment:

Skewness: We observe that, AGE, IS\_COBRAND, IS\_AIRACCRUAL, YEARS\_IN\_PROGRAM are right skewed since they are positive. Is\_hotel\_redemption is slightly negative skewed.



4rth Business Moment:

Kurtosis: Kurtosis is less for all the features. This means the data consist of less outliers, less standard deviation. This ensures our data set is good.

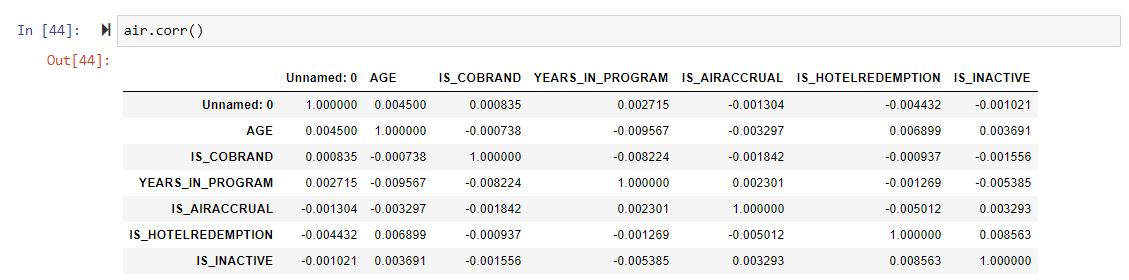


CORRELATION:

Variables within a dataset can be related for lots of reasons.

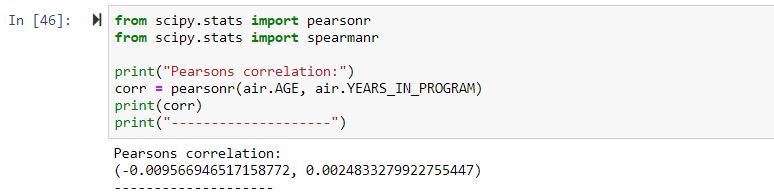
* One variable could cause or depend on the values of another variable.
* One variable could be lightly associated with another variable.
* Two variables could depend on a third unknown variable.

It can be useful in data analysis and modelling to better understand the relationships between variables. The statistical relationship between two variables is referred to as the correlation. A correlation could be positive, meaning both variables move in the same direction, or negative.



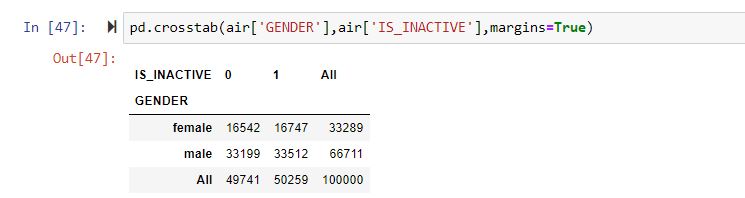
Pearson:

The Pearson’s correlation coefficient is calculated as the covariance of the two variables divided by the product of the standard deviation of each data sample. It is the normalization of the covariance between the two variables to give an interpretable score.



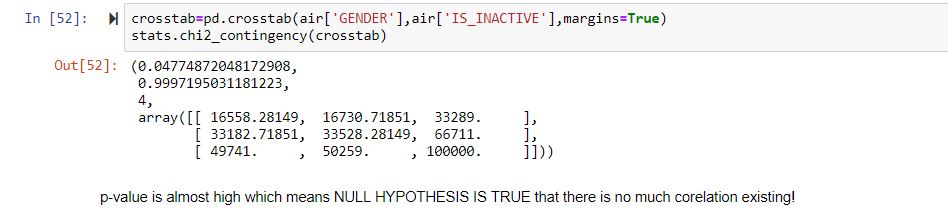
Contingency Table:

**Contingency Table** is one of the techniques for exploring two or even more variables. It is basically a tally of counts between two or more categorical variables.

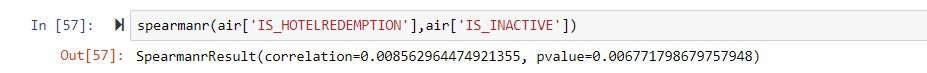


Chi squared Test:

The chi-square statistical test is used to determine whether there's a significant difference between an expected distribution and an actual distribution. It's typically used with categorical data such as educational attainment, colors, or gender

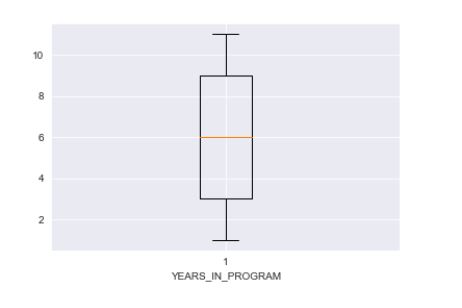
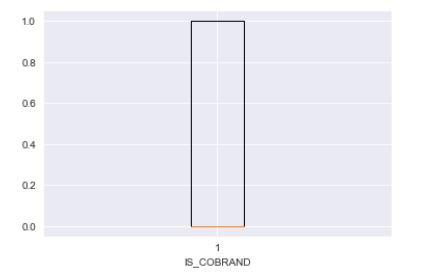
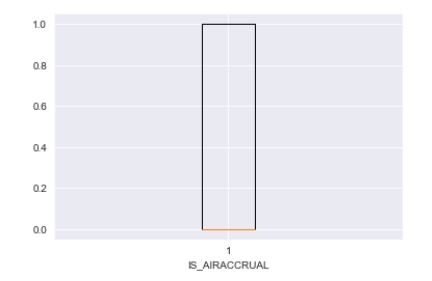
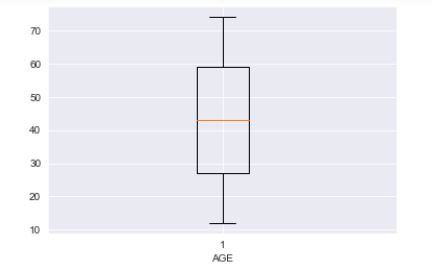


Spearman Correlation:



VISUALISATION:

Boxplots:





CLASSIFICATION MODEL:

Random Forest:

The Random Forest Classifier. Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction

Decision Tree:

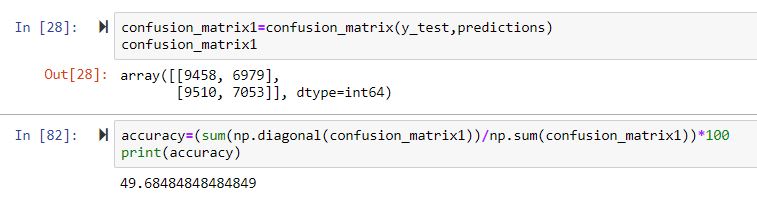
A decision tree is a flowchart-like tree structure where an internal node represents feature (or attribute), the branch represents a decision rule, and each leaf node represents the outcome.

SVM:

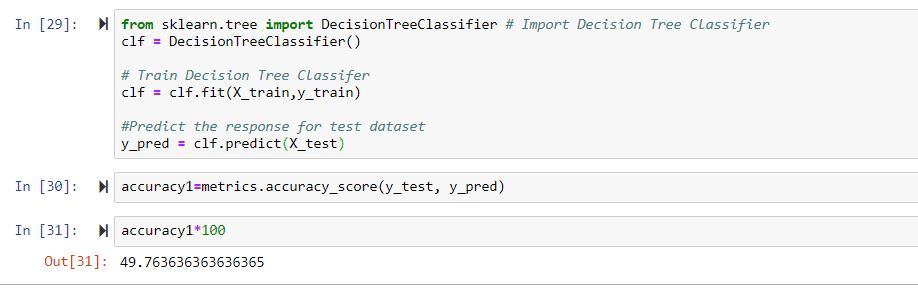
 Support vector machines so called as SVM is a supervised learning algorithm which can be used for classification and regression problems as support vector classification (SVC) and support vector regression (SVR). It is used for smaller dataset as it takes too long to process.

VALIDATING MODEL:

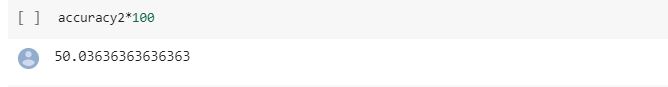
Random Forest:



Decision Tree:



SVM:



SVM is the best model among the three with accuracy of 50%.

CONCLUSION:

Customers satisfaction plays a key role in any firm. If customers are satisfied with all facilities, then we would definitely be on profit scale. Existing customers are very much necessary for our further development.

We observe higher dependency of inactive members on hotel redemption.

1)Hotel redemption facility is one of the most look out factor.

2)The most no of times traveling age is 23. It would be a better kick start if we concentrate on needs of this age group. That means Youth at the age of 23 are tending to be more travelling or explorative. Therefore, we can give them incentives and exciting offers relevant to this age

3)The maximum subscriptions are from gold base; therefore, we can improve the customers in this gold subscription by making them stay loyal.