

Introduction to Data Management

PROJECT REPORT

(Project Semester August-December 2018)

**DATA ANALYSIS IN AIRLINES IN INDIA**

Submitted By

Amit Sadarang

Registration No. 11612527

Programme and Section: P132: B.Tech.(CSE), KEM45

Course Code: INT217

Under the Guidance of

Mr. Hargobind Singh, 23599

Lovely School of Computer Science & Engineering

Lovely Professional University, Phagwara

**CERTIFICATE**

This is to certify that T. Amit Sadarang bearing Registration no. 11612527 has completed INT 217 project titled, **“Data Analysis in Airlines in India”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort and study.

**Signature and Name of the Supervisor**

**Designation of the Supervisor**

**School of Computer Science & Engineering**

Lovely Professional University

Phagwara, Punjab.

Date:

**DECLARATION**

I, Amit Sadarang student of P132: B.Tech. (Computer Science & Engineering) under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 17-11-2018

Signature

Amit Sadarang

11612527

ACKNOWLEDGEMENT

I have taken efforts in this course. However, it would not have been possible without the kind support and help of many individuals and University. I would like to extend my sincere thanks to all of them.

I am highly indebted to Mr. Hargobind Singh Sir for his guidance and constant supervision as well as for providing necessary information for the course & also for his support in completing this course.

I would like to express my gratitude towards my parents Lovely Professional University for their kind co-operation and encouragement.

I would like to express my special gratitude and thanks to my faculty for giving me such attention and time.

My thanks and appreciations also go to my colleague and people who have willingly helped me out with their abilities.

## **TABLE OF CONTENTS**

1. Introduction

2. Scope of the Analysis

3. Existing System

1. Drawbacks or limitations of existing system

4. Source of dataset

5. ETL process

6. Analysis on dataset (for each analysis)

1. Introduction
2. General Description
3. Specific Requirements, functions and formulas
4. Analysis results
5. Visualization

7. List of Analysis with results

8. Future scope

9. References

10. Bibliography

INTRODUCTION

When writing a airline report that can be easily understood, it's important to create a clear report that gives critical details. Having a sales report is beneficial to showcase accurate airlines’s passengers tracking. To write a successful airline’s passengers report, you must think about your audience and what information to share, and choose a specific time period and the right visuals.

Before drafting a airline report, think about your audience and consider the information they need. Information that would intrigue the airlines marketing will be different than what the airlines financial would be interested to know how to make decision. The airline company of marketing will want to know how well ticket sales to passenger representatives are converting leads into sales, along with details regarding marketing campaigns that have the most conversion rates and return of investment. The airlines company will want to known the major sales of tickets with the help of numbers of passengers and expenses.

First, you must identify the main audience for your airlines report. Next, decide what data will create a clear picture of how the airlines is performing. Choose specific information to share, such as meeting sales goals; revenue and expenses within a certain time period, services that are selling the most tickets to passenger ; number of flight booking forecasts for the next month and quarter; potential areas for improvement and opportunities; and any challenges.

*To choosing best time period for Flight Booking and offers:*Rather than sharing all of the sales numbers from the inception of the Airlines company, choose a time period to focus on. For example, choose a monthly or quarterly report, or a yearly review. The report will be easier for the audience to understand if it is time-focused and presents an accurate comparison factor like flight booking and which flight is chosen by passenger in that month.

*Now including right visuals for analysis of Airline data*:It's important to focus on specific engaging information. Think of how you want to present the data in a way that will capture the attention of your audience. Visuals are a great way to do this. The best way to include visuals in a Airline report is to make sure the graphics are actionable, digestible and understandable. For example, a bar graph can show how specific products or services are doing over a certain amount of time because it's easy to read and straightforward

A sales report should include dates of the period that will be covered, including dates of Airline within that specific time. Think about the main accomplishment and start the report with the most significant numbers. Follow this with a description of how much the target or goal has been met or even exceeded. The sales numbers should also be included in a daily, weekly, monthly, quarterly or yearly report. This figure includes a summary of how a sales number increased or decreased in comparison to the previous number. Be sure to include relevant statistics that show increases or decreases, along with any problems throughout the sales period.

SCOPE OF ANALYSIS

The Data Analysis in Airlines with the data sets and information related to passenger that traveled airline so that we analyse the data of Airline in order to increase Ticket sale,Booking analysis and customer satisfaction as well as to reduce sales barriers and low revenue levels .They create standardized and customized reports That analyse everything from quantitative data to sales funnel flow to future need forecasts.

The scope of analysing the data that I have taken is the following:-

we required the data set which have the information related to the number of passengers going out from the India to other country and number of passengers coming from other country to India with specific month.

Next the information related to the number of passengers going out from the India to other country and number of passengers coming from other country to India with specific country.

The analysis on the data set of the Airline give overview on the number of passenger traveled by the airline in each month between the India and other countries.This analysis give idea in which month the more number of passengers travels through the airline .

The analysis on the data set of the Airline give overview on specific airport with the number of passenger traveled by the airline in between the India and other countries.This analysis give idea in which month the more number of passengers travels through the airline

In this analysis, we required the data set which have the information related to the number of passengers going out from the India to other country and number of passengers coming from other country to India with specific airport of India.

so that in future we can study on Booking analysis of Airlines segment customers, target with personalized offers, optimize pricing in real-time using predictive analytics techniques such as modelling and forecasting.

To analyse a dataset using Pivot table, user needs to simply drag and drop the relevant data in the appropriate cells. The tool itself re arrange the data enabling the user to play around with the summarized or reorganized data to discover and identify data trends and patterns. During data analysis following features offered by excel pivot tables are of great help:

**Grouping:** groups the data according to the Header values. Any field which is added in to the Pivot table              either as row or column is can be grouped using the Grouping option. By default, the numeric values,                    including date and time fields, are grouped by pivot table

**Filters:** While doing an analysis on a large data it is often required to narrow down the data based on field            values to bring out trends in the data. This can be effectively achieved using the Filters feature of the Pivot            table.

**Slicers:** Slicers are just another form of filtering Slicers provided set of buttons based on data values of the             field selected for slicer. There are no drop downs in this case. Slicers can be defined for any type of field in             the pivot table

**Custom Calculation:** Very often there is requirement to change the way field values are displayed. For                 instance, instead of showing the sum or count of the numeric value it is required to display the numeric                 data in the form of percentages. This can be achieved using the Custom Calculation feature offered by Pivot           table.

**Calculated Field:** In case there is a requirement to do some runtime calculation using the Pivot tables                   fields then calculated fields are used. These type of fields allows to write custom formulas in the Pivot table           using already present fields in Pivot table.

EXISTING SYSTEM

The Existing system of the Airlines is given as follows:

# **System Name:** India: an analysis of the air transport market

**Source of the System:**

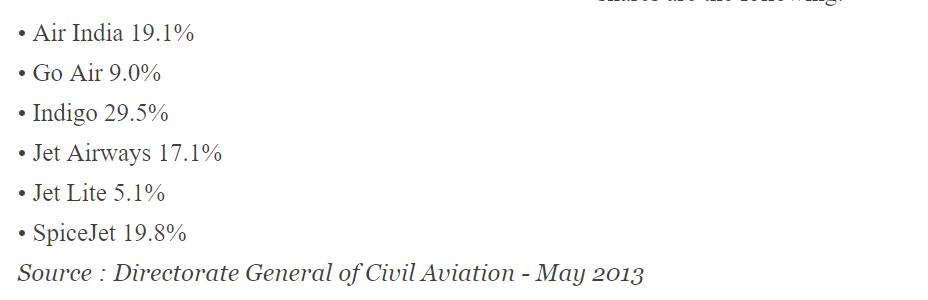
<https://www.linkedin.com/pulse/20141024053713-22815357-india-an-analysis-of-the-air-transport-market>

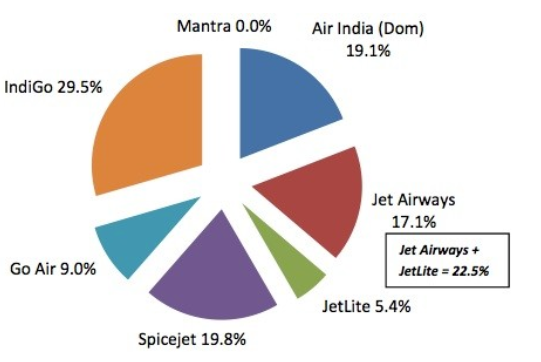
**Outcomes of the System:**

The best airline in which the more number of passengers are traveled in the India.

The show the visual graphical representation of airline data .

The more popular airport and number of passenger traveled in airlines.





**Drawbacks of the Existing System:**

The Existing system just gave the analysis and relation between passenger and Airline.The factors and the causes behind this were not explained.

No clear analysis of the number of passenger traveled is describe there for Airline.

The optimal solutions and the way to control it is not represented.

There are only limited number of data sets and charts in the system.

**Source of Database**

All the dataset that is used in this analysis is available in [https://www.kaggle.com/](https://www.kaggle.com/rajanand/international-air-traffic-from-and-to-india)

Also all the attributes that are present in specific file is mentioned in the description here I am not referring all the links for the excel sheet that are used in this analysis process but you can find all of them in this link: <https://www.kaggle.com/rajanand/international-air-traffic-from-and-to-india>.

It contain all the dataset in specific order that is used here, time period of data may differs form year to year of them are of 2015 to 2017 and month to month of them are of 2003 to 2009.

The data refers to country/city wise statistics of Airline. Dataset contains information like Total Number of passengers coming to India , Total Number of passengers going out from India, different Number of Airline in which Number of passenger travelled in India, different Number of Airport in which Number of passenger travelled from India.

**ETL Process**

This is very important part of the whole analysis process, here we extract, transform to useful one and then load to the visualization from different sources in this project the data is extracted from [https://www.kaggle.com/](https://www.kaggle.com/rajanand/international-air-traffic-from-and-to-india)

After extraction the available data was 6-8 different excel sheets, later I combined all the excel sheet in one sheet in which the rest of the analysis is done. There are several attributes that are present in data that are not required in the result for that I made several pivot tables to concise the data which is important for the analysis. For few results in which I require data from different tables I used Vlookup and Hlookup to get data from different tables.

Next part was to create different visualization form the concise data that is obtained from the pivot tables, Visualization part is very important as it is the result of all the process that can be understood by other people.

Pivot table helps in collecting important data that is going to used to create graphs and charts. A **pivot table** is a table of statistics that summarize the data of a more extensive table (such as from a database, spreadsheet, or business intelligence program). This summary might include sums, averages, or other statistics, which the pivottable groups together in a meaningful way. Which clears the raw data very much that can be used to show to the customer.

**ANALYSIS ON EACH DATA SET**

1. **Analysis the variation of the total number of passenger of traveled in each month:-**

1)Introduction:-

The analysis on the data set of the Airline give overview on the number of passenger traveled by the airline in each month between the India and other countries.This analysis give idea in which month the more number of passengers travels through the airline so that in future we can study on Booking analysis of Airlines segment customers, target with personalized offers, optimize pricing in real-time using predictive analytics techniques such as modelling and forecasting.

2)General Description:-

In this analysis, we required the data set which have the information related to the number of passengers going out from the India to other country and number of passengers coming from other country to India with specific month.

So we apply the filter to get specific data set from original data and the divide the data set on the bases of month and the total number passengers travelled inn that month.We also apply filter to analysis on the basis of each year.Pivot table is made and the bar graph chart is made for the visualisation. At last, a pie chart is made to show the percentage of the total passenger traveled in each month.

3)Specific Requirements, functions and formulas:-

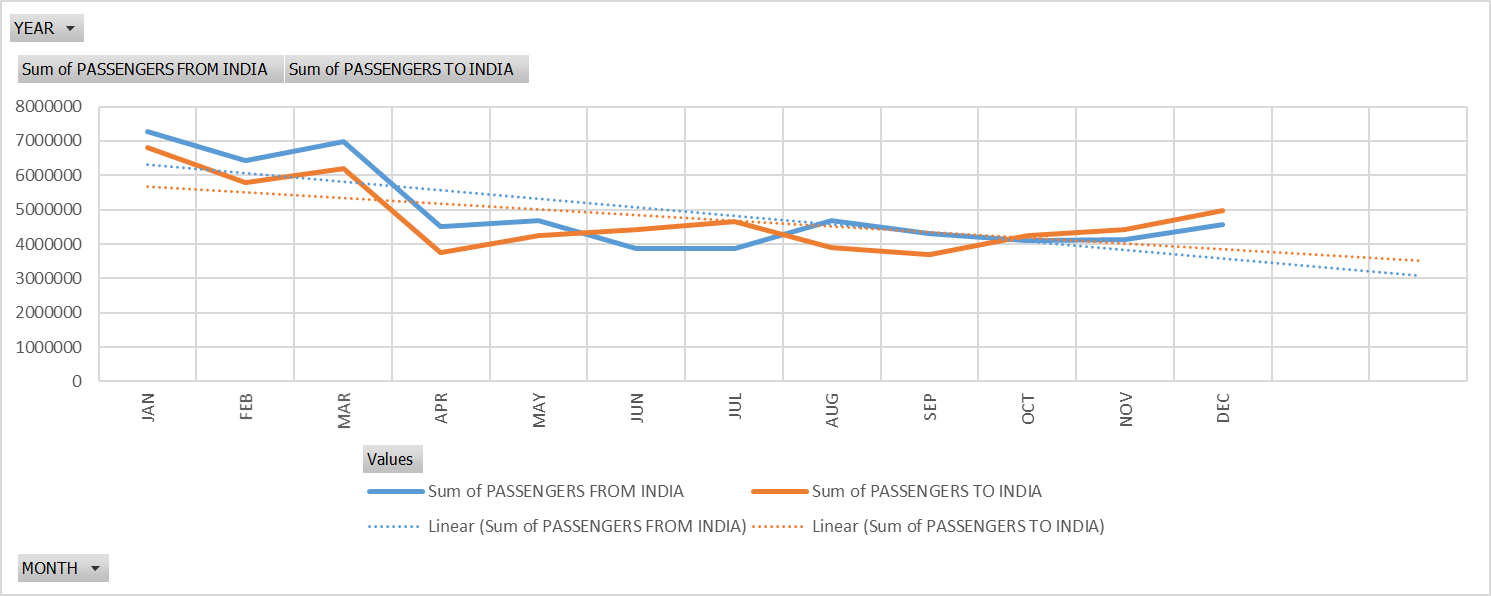
The analysis requires Microsoft excel 2010 or above, it also require pivoted table and different type of graphs.The function and formula of sum and count is also required in the calculation for analysis the total number of passengers.

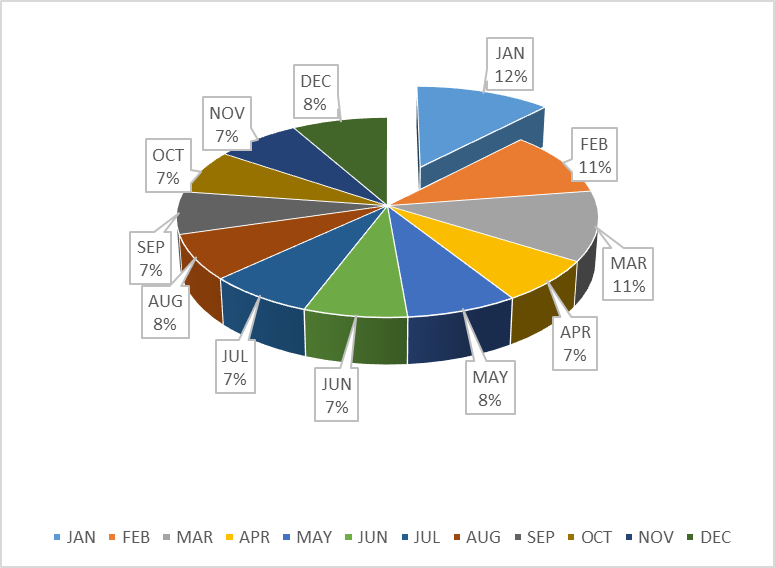
4)Analysis results:-

Analysis of this result is that we the highest number of passengers traveled in January month with 12%,February with 11%,March with 11%, April with 7%, May with 8%, June with 7%, July with 7%, August with 8%, September with 7%, November with 7, December with 8%.

So we can predict that January is the best month in which most of the passenger traveled from airline.February and March is also the second highest month in which most of the passenger traveled from airline.Airlines segment customers, target with personalized offers, optimize pricing in real-time using predictive analytics techniques such as modelling and forecasting according to above months.

1. Visualization:-





1. **Analysis the top most country from where the most number of passenger of traveled between India and other countries:-**

1)Introduction:-

The analysis on the data set of the Airline give overview on the number of passenger traveled by the airline in between the India and other countries.This analysis give idea from which foreign country, the more number of passengers travels through the airline so that in India so that in future we can study on Foreign tourism and Booking analysis of Airlines segment customers, target with personalized offers, optimize pricing in real-time using predictive analytic techniques such as modelling and forecasting.

2)General Description:-

In this analysis, we required the data set which have the information related to the number of passengers going out from the India to other country and number of passengers coming from other country to India with specific country.

So we apply the filter to get specific data set from original data and the divide the data set on the bases of country and the total number passengers travelled in that country.We also apply filter to analysis on the basis of each year.Pivot table is made and the bar graph chart is made for the visualisation. At last, a pie chart is made to show the percentage of the total passenger traveled in top most airlines in India.

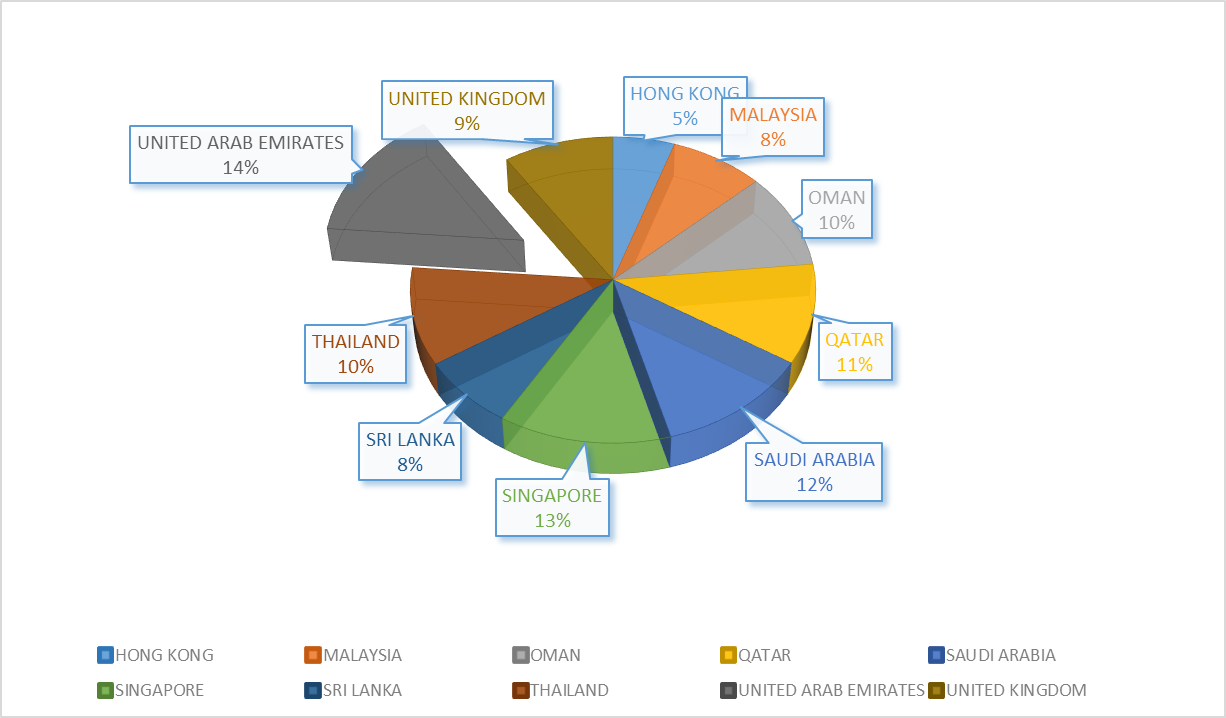
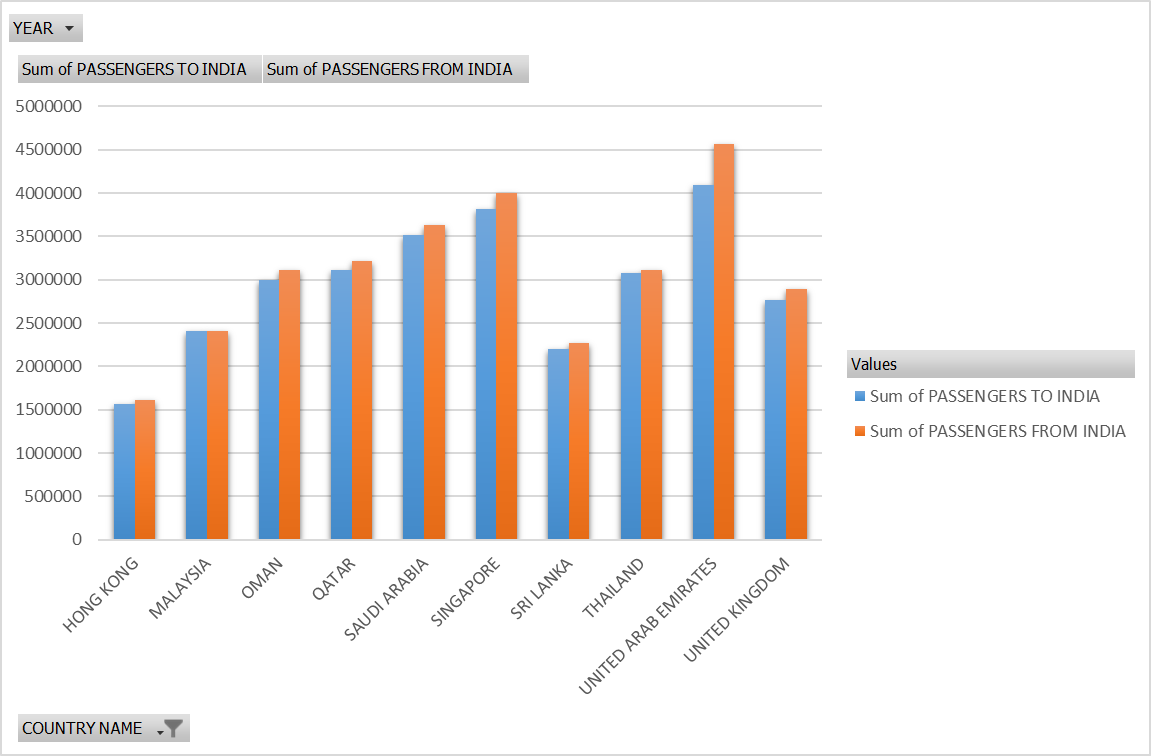
3)Specific Requirements, functions and formulas:-

The analysis requires Microsoft excel 2010 or above, it also require pivoted table and different type of graphs.The function and formula of sum and count is also required in the calculation for analysis the total number of passengers.

1. Analysis results:-

Analysis of this result is that we the highest number of passengers traveled in top 10 countries like United Arab Emirates with 14%,Singapore with 13%, Saudi Arab with 12%,Omen with 10%,Thailand with 10%, Malaysia with 8%,Srilanka with 8%,and Hong kong with 5%.

So we can predict that United Arab Emirates is the best country in which most of the passenger traveled from airline.Singapore,Thailand and Omen is also the country in which most of the passenger traveled from airline.Airlines segment customers passengers, target with personalized offers, optimize pricing in real-time using predictive analytics techniques such as modelling and forecasting according to above country which they visited the most.

1. Visualization:-
2. **Analysis the best Airline chosen by the number of passenger of traveled between the India and other countries:-**

1)Introduction:-

The analysis on the data set of the Airline give overview on the number of passenger traveled by the airline in each month between the India and other countries.This analysis give idea in which month the more number of passengers travels through the airline so that in future we can study on Booking analysis of Airlines segment customers, target with personalized offers, optimize pricing in real-time using predictive analytics techniques such as modelling and forecasting.

2)General Description:-

In this analysis, we required the data set which have the information related to the number of passengers going out from the India to other country and number of passengers coming from other country to India with specific airline.

So we apply the filter to get specific data set from original data and the divide the data set on the bases of airline and the total number passengers travelled in that month.We also apply filter to analysis on the basis of each year.Pivot table is made and the bar graph chart is made for the visualisation. At last, a pie chart is made to show the percentage of the total passenger traveled from each airlines.

3)Specific Requirements, functions and formulas:-

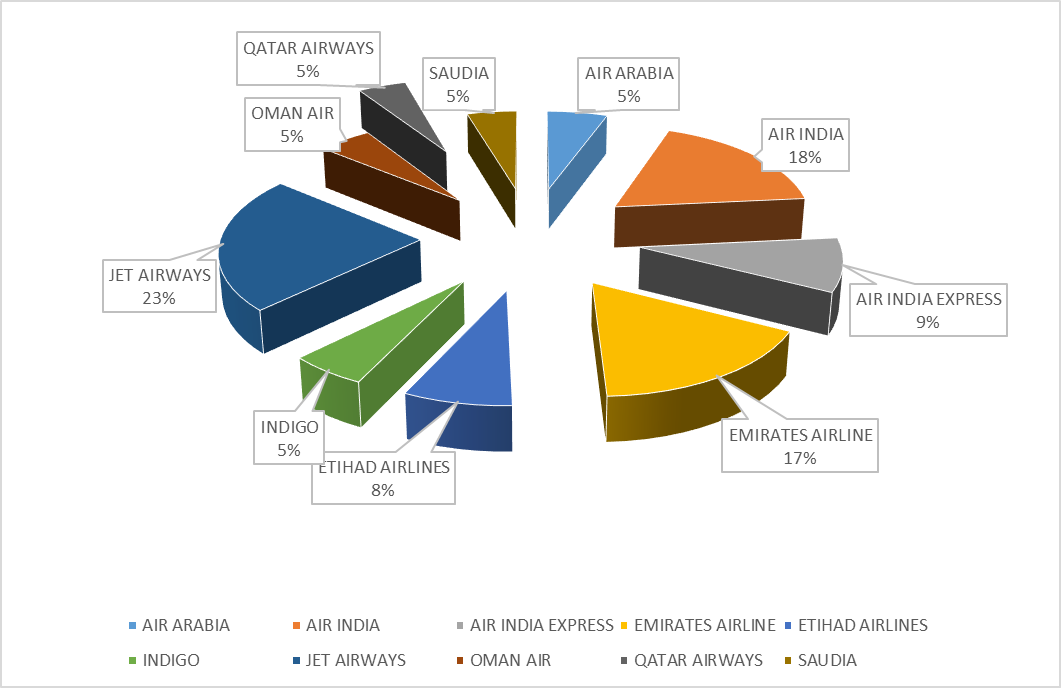
The analysis requires Microsoft excel 2010 or above, it also require pivoted table and different type of graphs.The function and formula of sum and count is also required in the calculation for analysis the total number of passengers.

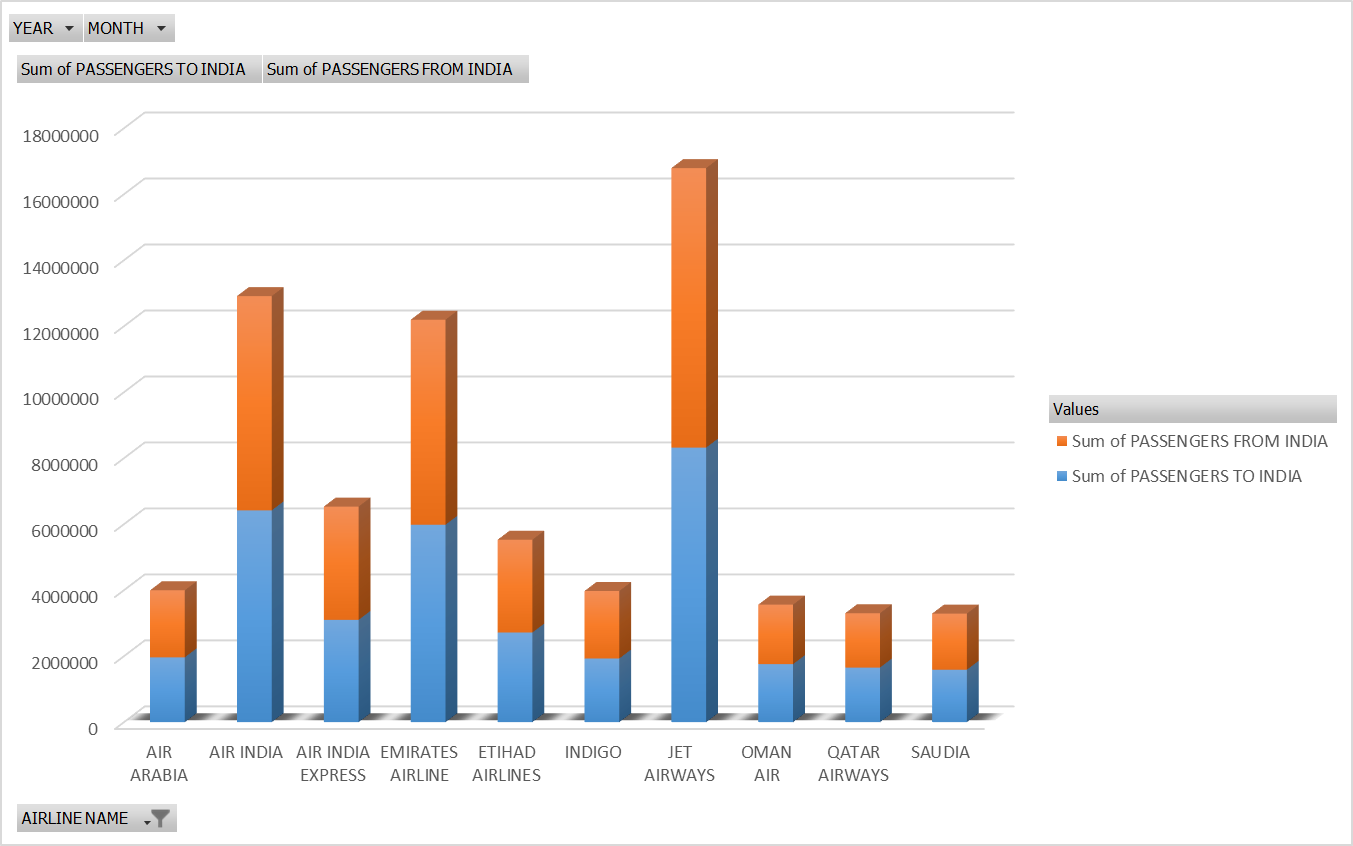
1. Analysis results:-

Analysis of this result is that Top 10 Airline in which the highest number of passengers traveled in like Jet Airways with 23%, Air India with 18%, Emirates Airline with 17%, Air India Express with 9%, Indigo with 5%, and Omen airways with 5%.

So we can predict that Jet Airways is the best airline in which most of the passenger traveled from airline.Air India and Emirates Airline is also the second and third highest Airline in which most of the passenger traveled from airline.Airlines segment customers passengers, target with personalized offers, optimize pricing in real-time using predictive analytics techniques such as modelling and forecasting according to above Airline in which they chose the most.

1. Visualization:-





1. **Analysis the top best Airport visited by the passengers who traveled between India and the other country :-**

1)Introduction:-

The analysis on the data set of the Airline give overview on specific airport with the number of passenger traveled by the airline in between the India and other countries.This analysis give idea in which month the more number of passengers travels through the airline so that in future we can study on Booking analysis of Airlines segment customers, target with personalized offers, optimize pricing in real-time using predictive analytics techniques such as modelling and forecasting.

2)General Description:-

In this analysis, we required the data set which have the information related to the number of passengers going out from the India to other country and number of passengers coming from other country to India with specific airport of India.

So we apply the filter to get specific data set from original data and the divide the data set on the bases of month and the total number passengers travelled inn that month.We also apply filter to analysis on the basis of each year.Pivot table is made and the bar graph chart is made for the visualisation. At last, a pie chart is made to show the percentage of the total passenger visited specific airport.

3)Specific Requirements, functions and formulas:-

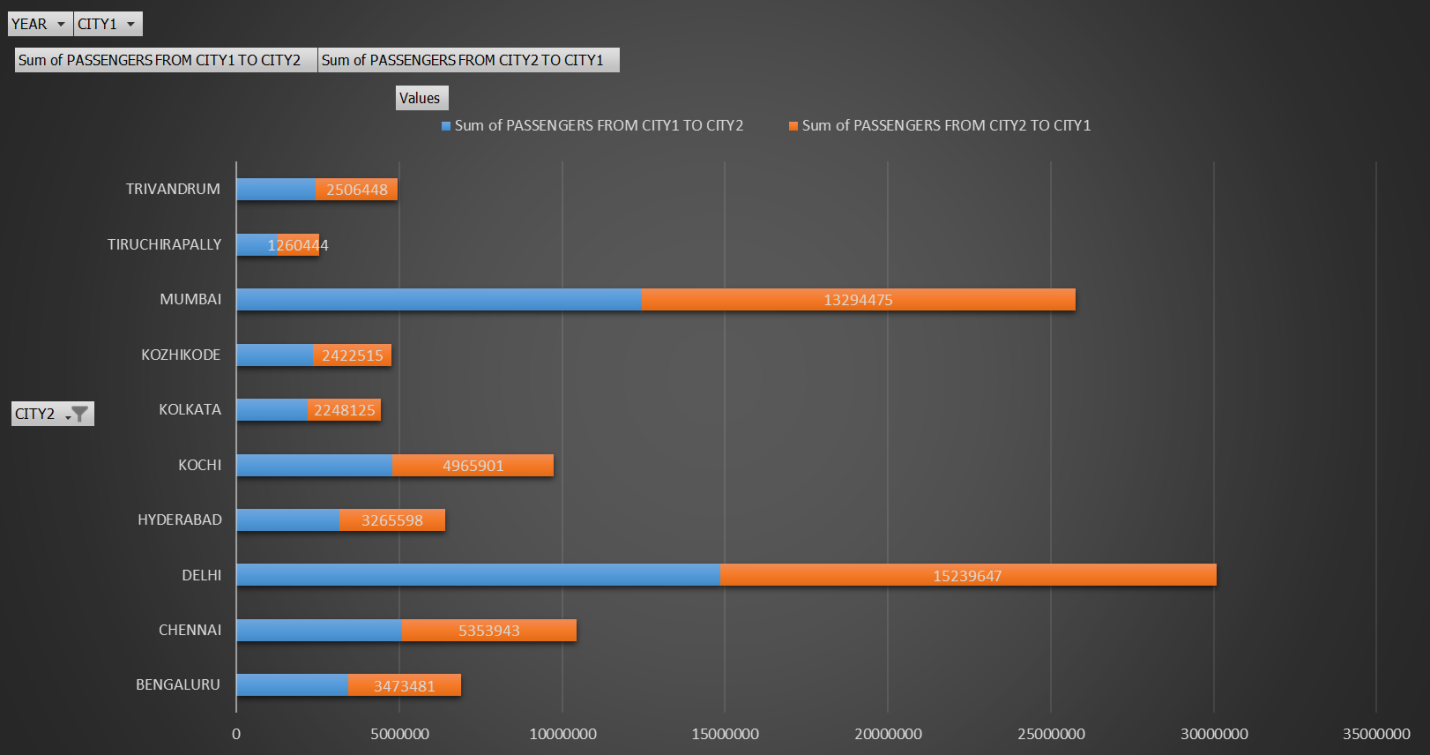
The analysis requires Microsoft excel 2010 or above, it also require pivoted table and different type of graphs.The function and formula of sum and count is also required in the calculation for analysis the total number of passengers.

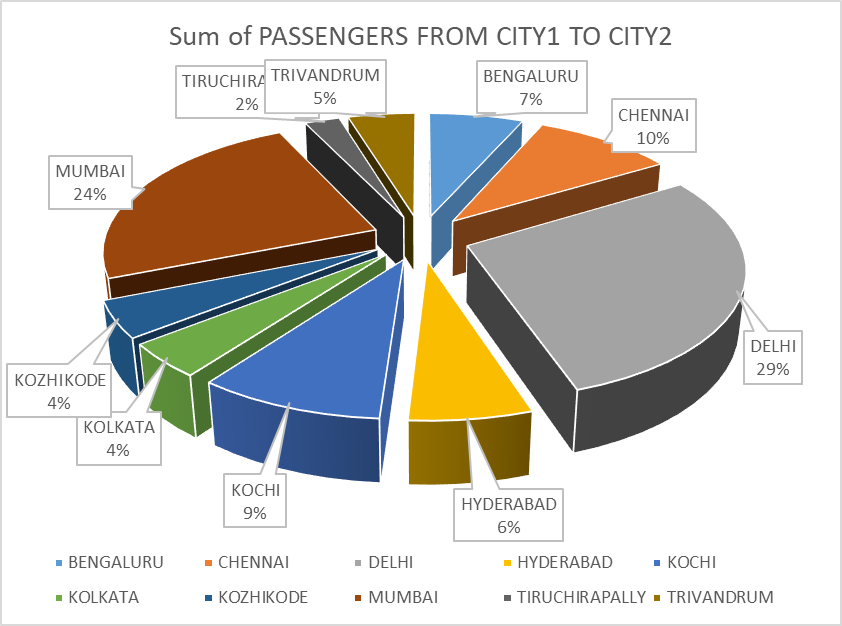
1. Analysis results:-

Analysis of this result is that Top 10 City Airport in which the highest number of passengers traveled in like Delhi with 29%, Mumbai with 23%, Chennai with 10%, Hyderabad with 6%, Trivandrum with 5%, Bengaluru with 7%, Kolkata with 4%, and Triuchirapally 2%.

So we can predict that Delhi is the best airline in which most of the passenger traveled from airline.Mumbai and Chennai Airline is also the second and third highest Airprt in which most of the passenger traveled from airline.Airlines segment customers passengers, target with personalized offers, optimize pricing in real-time using predictive analytics techniques such as modelling and forecasting according to above City Airport in which they visited the most.

1. Visualization:-





1. **Analysis the variation of the number of passengers traveled in each year by every airlines :-**

1)Introduction:-

The analysis on the data set of the Airline give overview on the number of passenger traveled by the airline in each year by every airline between the India and other countries.This analysis give idea in which month the more number of passengers travels through the airline so that in future we can study on Booking analysis of Airlines segment customers, target with personalized offers, optimize pricing in real-time using predictive analytics techniques such as modelling and forecasting.

2)General Description:-

In this analysis, we required the data set which have the information related to the number of passengers going out from the India to other country and number of passengers coming from other country to India with specific month.

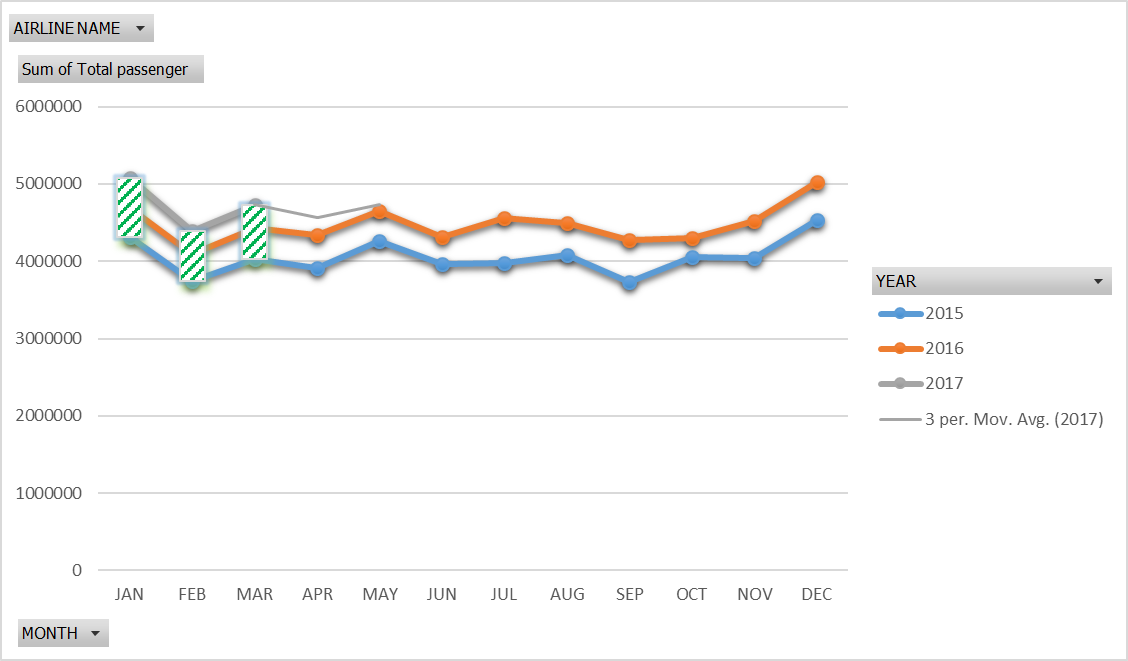
So we apply the filter to get specific data set from original data and the divide the data set on the bases of year and the total number passengers travelled inn that month.We also apply filter to analysis on the basis of each month and airline.Pivot table is made and the bar graph chart is made for the visualisation. At last, a pie chart is made to show the percentage of the total passenger traveled in each year.

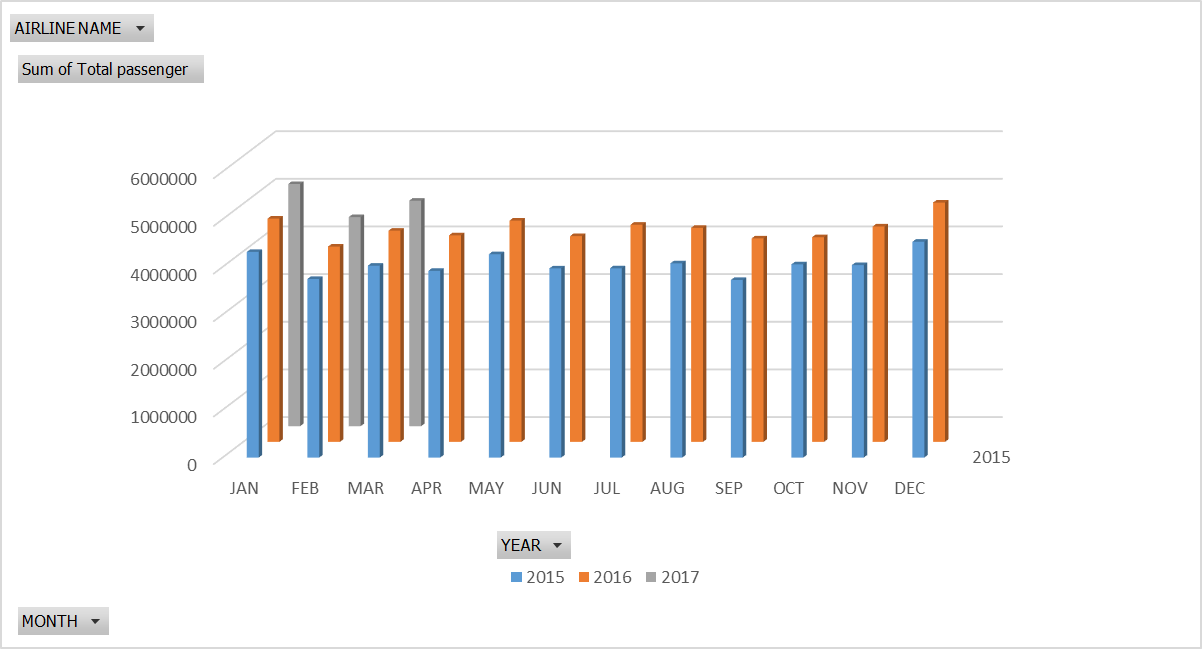
3)Specific Requirements, functions and formulas:-

The analysis requires Microsoft excel 2010 or above, it also require pivoted table and different type of graphs.The function and formula of sum and count is also required in the calculation for analysis the total number of passengers.

1. Analysis results:-

Analysis of this result is that we can predict that each Year the number of passenger traveled from the Airline is increased and best month in January in which most of the passenger traveled from airline.Air India and Emirates Airline is also the second and third highest Airline in which most of the passenger traveled from airline.Airlines segment customers passengers, target with personalized offers, optimize pricing in real-time using predictive analytics techniques such as modelling and forecasting according to increases the number of passengers in each year.

5)Visualization:- 



7}List of Analysis with results:-

1>first we can predict that January is the best month in which most of the passenger traveled from airline.February and March is also the second highest month in which most of the passenger traveled from airline.Airlines segment customers, target with personalized offers, optimize pricing in real-time using predictive analytics techniques such as modelling and forecasting according to above months.

2>Next thing we can predict that United Arab Emirates is the best country in which most of the passenger traveled from airline.Singapore,Thailand and Omen is also the country in which most of the passenger traveled from airline.Airlines segment customers passengers, target with personalized offers, optimize pricing in real-time using predictive analytics techniques such as modelling and forecasting according to above country which they visited the most.

3>Now here we can predict that Jet Airways is the best airline in which most of the passenger traveled from airline.Air India and Emirates Airline is also the second and third highest Airline in which most of the passenger traveled from airline.Airlines segment customers passengers, target with personalized offers, optimize pricing in real-time using predictive analytics techniques such as modelling and forecasting according to above Airline in which they chose the most.

4>Also we can predict that Delhi is the best airline in which most of the passenger traveled from airline.Mumbai and Chennai Airline is also the second and third highest Airprt in which most of the passenger traveled from airline.Airlines segment customers passengers, target with personalized offers, optimize pricing in real-time using predictive analytics techniques such as modelling and forecasting according to above City Airport in which they visited the most.

5>At last we can predict that each Year the number of passenger traveled from the Airline is increased and best month in January in which most of the passenger traveled from airline.Air India and Emirates Airline is also the second and third highest Airline in which most of the passenger traveled from airline.Airlines segment customers passengers, target with personalized offers, optimize pricing in real-time using predictive analytics techniques such as modelling and forecasting according to increases the number of passengers in each year.

8}. Future scope:-

The Future scope of this Data Analysis in Airlines with the data sets and information related to passenger that traveled airline so that we analyse the data of Airline in order to increase Ticket sale,Booking analysis and customer satisfaction as well as to reduce sales barriers and low revenue levels .They create standardized and customized reports That analyse everything from quantitative data to sales funnel flow to future need forecasts.

We required the data set which have the information related to the number of passengers going out from the India to other country and number of passengers coming from other country to India with specific month.

Next the information related to the number of passengers going out from the India to other country and number of passengers coming from other country to India with specific country.

The analysis on the data set of the Airline give overview on the number of passenger traveled by the airline in each month between the India and other countries.This analysis give idea in which month the more number of passengers travels through the airline .

The analysis on the data set of the Airline give overview on specific airport with the number of passenger traveled by the airline in between the India and other countries.This analysis give idea in which month the more number of passengers travels through the airline

In this analysis, we required the data set which have the information related to the number of passengers going out from the India to other country and number of passengers coming from other country to India with specific airport of India.

so that in future we can study on Booking analysis of Airlines segment customers, target with personalized offers, optimize pricing in real-time using predictive analytics techniques such as modelling and forecasting.

**9} References**

1. advanced\_excel\_tutorial.pdf.
2. [pax-monthly-analysis passenger-analysis-aug-2018.pdf](https://www.iata.org/publications/economics/Reports/pax-monthly-analysis/passenger-analysis-aug-2018.pdf)
3. 2016\_eu\_air\_transport\_industry\_analyses\_report.pdf]

**10}Bibliography:**

* https://www.iata.org/publications/economics/Reports/pax-monthly-analysis/passenger-analysis-aug-2018.pdf
* <https://panoply.io/data-warehouse-guide/3-ways-to-build-an-etl-process/>
* <https://www.datasciencecentral.com/profiles/blogs/10-open-source-etl-tools>
* <https://www.kaggle.com/rajanand/international-air-traffic-from-and-to-india/>
* <http://www.airindia.in/index.htm>
* <https://www.tutorialspoint.com/advanced_excel/advanced_excel_tutorial.pdf>
* <https://ec.europa.eu/transport/sites/transport/files/2016_eu_air_transport_industry_analyses_report.pdf>