PROJECT REPORT

Data Visualization and Analytics for

Comprehensive Flight Data from Priceline Dashboard

Rup Kiran Boyana

Contents

[**1.** **INTRODUCTION – ABOUT PROJECT** 3](#_Toc133419674)

[1.1 Problem Statement 3](#_Toc133419675)

[1.2 About Data 3](#_Toc133419676)

[**2.** **DATA DISCOVERY** 4](#_Toc133419677)

[2.1 Understanding Data 4](#_Toc133419678)

[2.2 Data Visualizations and answers to the business questions 4](#_Toc133419679)

[2.2.1 Visualizing Aggregate no of passengers and Destination Airports. 4](#_Toc133419680)

[2.2.2 Visualizing Ticket Price based on Arrival Time and Destination Airport. 5](#_Toc133419681)

[2.2.3 Visualizing the factors that are most Important for Priceline customers based Number of Stoppages. 5](#_Toc133419682)

[2.2.4 Visualizing the factors that are most Important for Priceline customers based on multiple dimensions. 6](#_Toc133419683)

[2.2.5 Visualizing particular routes that are more popular or more expensive than others 8](#_Toc133419684)

[2.2.6 Visualizing average duration of the flight and how does it vary by destination. 9](#_Toc133419685)

[2.2.7 Visualizing departure time affects the flight ticket price. 10](#_Toc133419686)

[2.2.8 Visualizing the relationship between the distance of the flight and the ticket price. 11](#_Toc133419687)

[2.2.9 Visualizing the number of stops affect the price of a flight ticket. 11](#_Toc133419688)

[2.2.10 Visualizing the relationship between the popularity of destination and the price of flight ticket. 12](#_Toc133419689)

[**3.** **DATA STANDARDIZATION AND PROCESSING** 13](#_Toc133419690)

[3.1 Data Scrubbing 13](#_Toc133419691)

[3.2 Data Processing 13](#_Toc133419692)

[**4.** **CALCULATIONS AND PARAMETERS** 13](#_Toc133419693)

[4.1 Row level calculations 13](#_Toc133419694)

[4.2 Aggregate level Calculations 14](#_Toc133419695)

[4.3 Level Of Detail 14](#_Toc133419696)

[4.4 Table level calculations 15](#_Toc133419697)

[**5.** **DASHBOARD** 15](#_Toc133419698)

[5.1 Dashboard 1 15](#_Toc133419699)

[5.2 Main Dashboard 16](#_Toc133419700)

[**6.** **Conclusion** 17](#_Toc133419701)

# **INTRODUCTION – ABOUT PROJECT**

**Project Title: COMPREHENSIVE FLIGHT DATA FROM PRICELINE – Analytics and Dashboard**

**Project Description**:

Users of Priceline's online travel agency can look for and reserve flights, lodging, rental cars, and other travel-related services. On Priceline, users can enter their trip information, such as the departure and arrival locations, the dates of travel, and the number of passengers. Following a search of its database of accessible flights, Priceline presents consumers with a list of possibilities that includes flight information such as schedules, airlines, pricing, and other specifics.

The data file contains flight details from Priceline, with each row representing a single flight record. The columns in the file provide information on various flight details, including the airline name, travel time, number of stoppages, ticket price, departure airport, departure time, stoppage details (if any), destination airport, arrival time, and arrival date.

## **Problem Statement**

Priceline’s initial goal was to increase the customers and accommodate the flights from vivid airlines with higher capacity of passengers by offering discounts and naming their own price. This dataset can be useful for analysis and visualization of pricing trends and flight patterns, which can be useful for various purposes such as market research, business strategy development, and customer behavior analysis. Here, data discovery and visualization help in predicting flight prices or identifying patterns in customer behavior. Overall, the Priceline flight dataset provides a valuable resource for those interested in exploring and analyzing the travel industry.

**REAL LIFE EXAMPLE:**

## **About Data**

To illustrate, a customer might enter their travel information into the search tool of the travel agency, including their departure and arrival cities, travel dates, and the number of passengers. The search tool would then use Priceline's API to acquire thorough flight data from several airlines and present the consumer with a list of available flights along with their costs, travel durations, layover locations, and other specifics.

In order to help consumers make educated decisions when booking their tickets, the travel firm might leverage Priceline's API to offer them additional information, such as baggage allowances, in-flight amenities, and seat availability. Customers could compare and pick the best flights depending on their tastes and budget using this information.

Overall, integrating detailed flight data from Priceline can assist travel companies in developing unique travel applications that offer their clients' customers a smooth and individualized booking experience, as well as a large selection of flights and affordable prices.

# **DATA DISCOVERY**

## **Understanding Data**

Before even starting to make meaningful visualizations, data analysts must spend 60–80% of their time gathering, organizing, and cleaning the data. That is certainly the case and then some for these datasets. It appears that much deeper exploration of the data is necessary before it can be completely utilized to construct a compelling story. For this project, we simply "cleansed and re-fluffed" a portion of the data in order to address the primary questions we want to address; the other data gaps will be analyzed and cleaned up in the future.

To begin sorting and cleaning up the data, first, we used filtering and sorting techniques in an Excel spreadsheet to locate the "largest" gaps in the data. For example, we have almost worked on each and every column beginning from the column 1 Airline name we have deleted all the blanks from the column to get the exact passengers count. Columns Departure time and Arrival time are converted to AM, PM formats.

Also column Departure Date has been calculated using the columns Travel time and Arrival Date.

## **Data Visualizations and answers to the business questions**

### Visualizing Aggregate no of passengers and Destination Airports.

This visualization explains what are the most popular destinations for travelers booking flights through Priceline? for this we have created No of passengers field using row level calculations count([Airline name]) to find out aggregate of total no of passengers. Later this is measured field is used against dimension Destination Airport field to find out which destination airport has the highest no of passengers traveled.

Horizontal Bar Chart1 -

A picture containing graphical user interface

Description automatically generated

### Visualizing Ticket Price based on Arrival Time and Destination Airport.

This visualization explains how the Ticket Prices differ based on the Arrival Time and Destination Airports. For this we have considered two dimensions Arrival Time and Destination Airport the SUM of Ticket Prices is the measured field. Using this we can understand how the prices peak during particular hour of the day for each destination airport and what is the least cost during particular hour of the day for any particular destination airport.

Stacked Bar Chart -

Chart, bar chart

Description automatically generated

### Visualizing the factors that are most Important for Priceline customers based Number of Stoppages.

This visualization is used to find the most Important factors for the priceline customers based on Number of stoppages and Aggregate of No of passengers having 1 Stop, 2 Stops, 3 Stops and No Stops.

Using dimension Number of stoppage and measured field No of passengers we are able to get an insight on the No of passengers travelled from departure airport to destination airport with 1 Stop, 2 Stops, 3 Stops and No Stops. From the visualization we can understand that highest no of passengers travelled with 1 stop.

Pie Chart -

Chart, pie chart

Description automatically generated

### Visualizing the factors that are most Important for Priceline customers based on multiple dimensions.

This visualization is used to find the most Important factors for the priceline customers based on multiple dimensions. Here we have used three dimensions Airline Name, Number of stoppage and Travel Group. Travel Group is created using row level calculations and other calculated field Travel Hour.

**Travel Hour 1**

SPLIT([Travel Time], ':', 1)

Travel Group –

IF INT([Travel Hour 1]) <= 24 THEN "1day"

ELSEIF INT([Travel Hour 1]) <=48 THEN "2days"

Else ">2days"

END

Also used Level of detail to find the percentage of Total No of passengers. To find out the percentage of passengers it is converted to continuous variable and then using the quick table calculations from the drop down and selecting the percent of total we can obtain the individual percentages of passengers.

Table

Description automatically generated with medium confidence

Here dimension Airline Name has been grouped into four categories and named as group A, B, C and D

The total number of unique Airlines are 56 and are divided into four Groups consisting of 14 airlines including each group.

Group A has the Airlines with highest percentage of passengers travelled and Group D has the Airlines with least percentage of passengers travelled. Below are the insights on the individual percentage of passengers travelled in each group.

Group A - 36.55

Group B - 25.85

Group C - 21.89

Group D - 15.25

Also, the concept of parameters has been used to visually include a reference line to easily identify the current values that are greater than 11 percent.

Graphical user interface, application

Description automatically generated

From this visualization we can draw the insights on the factors that are most important for the Priceline customers are based on the multiple dimensions such as Airline name, Number of stoppage and Travel Group and by calculating the percentage of the passengers travelled.

Horizontal Bar Chart 4 -

Chart, bar chart

Description automatically generated

### Visualizing particular routes that are more popular or more expensive than others

This visualization is used to get insights on particular routes that are more popular or more expensive than others. From the visualization we can understand Route DXB 🡪 BKK has the highest No of passengers and the average ticket price of $588 which is most popular among the other routes. Also routes DXB 🡪 MLE, BOM 🡪 LHR and BOM 🡪 DXB have only single passenger with highest average ticket price of $3,188 and $3,166 has considered to be the most expensive routes to the other routes.

Thereby the correlation between the routes and the average Ticket price are inversely proportional.

Here the dimension route has been created using the row level calculations.

**Route**

[Departure Airport]+'---> '+[Destination Airport]

As we have included more than one measured field in the columns two parallel horizontal graphs are shown one provides the insights on aggregate of no of passengers travelled in each route and the second horizontal graph provides insights Average Ticket Price for each route.

Side by side Horizontal Bar Charts -

Chart, bar chart

Description automatically generated

### Visualizing average duration of the flight and how does it vary by destination.

The visualization is used to get insights on the average duration of the flight and how it varies by the destinations. From the table data obtained we can visualize the unique Airlines and their travelled destinations along with the number of minutes travelled from the departure to the destination airport.

Hence it is easy to extract data and estimate the total travel time based on the Airlines name and the destination airport.

The dimensions and measures created using the row level calculations are.

**Travel Hour 1**

SPLIT([Travel Time], ':', 1)

**Travel Min 1**

SPLIT([Travel Time], ':', 2)

**No of Mins Travelled**

INT([Travel Hour 1])\*60+INT([Travel Min 1 ])

Cathay Pacific Airlines has travelled 3,945 average no of minutes to the destination airport BOM which the highest travelled time among the airlines and destinations.

Text Table Chart -

Graphical user interface, application, table

Description automatically generated

### Visualizing departure time affects the flight ticket price.

We have used visualization to obtain the insights on departure time effects on the flight ticket price. From this visualization we can understand that departure times 2:10 AM, 4:45 AM, 9:20 AM, 1:35 AM and 5:40 AM have the highest Average of Ticket Prices ranging from $3,042 and $2,534.

For this visualization we have considered dimension Departure Time and measure Avg. Ticket Price.

Packed Bubbles Chart -

Chart, bubble chart

Description automatically generated

From the color pattern it is easy to understand the darker theme is resembling the highest Avg. Ticket Price and lighter theme is resembling the lowest Avg. Ticket Price.

### Visualizing the relationship between the distance of the flight and the ticket price.

This visualization draws the relationship between the distance of the flight and the Average Ticket Price.

The line graph illustrates the correlation between Average Ticket price and the No of mins travelled. Here the correlation is inversely proportional to the other as the travel time increases the average ticket price decreases and if the travel time decreases the average ticket price increases.

From the visualization Average Ticket Price for 1861 minutes travel time is $7,608 and the Average Ticket Price for 4930 minutes travel time is $840.

For this visualization we have considered two measures No of Mins Travelled and the Average Ticket price.

Line Graph1 -

Chart

Description automatically generated

### Visualizing the number of stops affects the price of a flight ticket.

Visualization using the tree map explains how the number of stops are affecting the price of a flight ticket.

From the visualization most of passengers have opted for flights with more than 2 stops as the Ticket Price decrease based on the average time travelled.

Hence, we can understand that 3 stops and 2 stops have the highest average Ticket Price in comparison to 1 stop and Nonstop flights as more no of people opted for more than 2 stops the average has been picked.

Here we have used dimension Number of Stoppage and the measured field Average Ticket Price in the Tree map visualization.

Tree Map -

Chart, treemap chart

Description automatically generated

### Visualizing the relationship between the popularity of destination and the price of flight ticket.

The Scatter Plot from the visualization explains the correlation between popularity of the destination and the price of the flight ticket. The popularity of the destination is identified by No of Passengers visiting the destination. As the number of passengers visiting the destination is high then the price of flight ticket decreases and as the passengers visiting the destination is low the price of the flight ticket increases dramatically. So, the correlation is inversely proportional between the popularity of destination and the price of flight ticket.

Here Destination Airport BKK has the highest No of Passengers, and the Avg. Ticket Price is $943 and similarly Destination Airport LHR has only 1 passenger and the Avg. Ticket Price is $3,166.

For the visualization we have used dimension Destination Airport and the measured fields Avg. Ticket Price and Aggregate No of passengers.

Scatter Plot -

Chart, scatter chart

Description automatically generated

# **DATA STANDARDIZATION AND PROCESSING**

## **Data Scrubbing**

As mentioned above, a significant portion of the data had to be manually corrected in Excel before being imported into Tableau because we were unable to find an immediate, simple solution to fill the gaps in Tableau. One problem, the time formats, was something we were able to quickly fix in Tableau.

## **Data Processing**

The new files were loaded into Tableau section after the data files were as clean as they would be for this Project. These are the fixed set of files:

# **CALCULATIONS AND PARAMETERS**

## **Row level calculations**

Route - Concatenation of destination airport and arrival airport

[Departure Airport]+'---> '+[Destination Airport]

Travel Group - This is to know the travel time if it is less than 24, 48 hours or greater than 48 hours.

IF INT([Travel Hour 1]) <= 24 THEN "1day"

ELSEIF INT([Travel Hour 1]) <=48 THEN "2days"

Else ">2days"

END

Travel Hour 1 - This is to split the travel time into hours

SPLIT([Travel Time], ':', 1)

Travel MIN 1 - This is to split the travel time into mins

SPLIT([Travel Time], ':', 2)

No of Mins Traveled - This is to convert hours into mins and then sum with mins

INT([Travel Hour 1])\*60+INT([Travel Min 1 ])

No of Passengers - To know the no of passengers in each airline.

count([Airline name])

Color difference –

IF AVG([Ticket price(Doller)]) < AVG([Avg Tckt Price]) then "Green"

ELSE "Red"

END

## **Aggregate level Calculations**

No of Passengers, Ticket Price are the measures which can be used as aggregate.

## **Level Of Detail**

Avg Tckt Price - This is different from aggregate, and it is used to find the total average.

{ FIXED : AVG([Ticket price(Doller)])}

## **Table level calculations**

No of passengers - Used to find the percentage of passengers after aggregation.

# **DASHBOARD**

## **Dashboard 1**

The below dashboard has been created to represent the Visualizing factors that are most Important for Priceline customers based on multiple dimensions. As there are multiple factors that are part of visualization, we have used two graphs and are included in dashboard for a single view of the visualizations.

Graphical user interface

Description automatically generated with medium confidence

## **Main Dashboard**

The main dashboard includes all the charts that are used for the data visualization that provides an overall insight on how the data has been used and analyzed to answer the business questions in an appropriate manner.

Chart

Description automatically generated

# **Conclusion**

In conclusion, the Priceline Comprehensive Flight Data collection offers an enormous amount of data about flight reservations made using the Priceline platform. The dataset contains comprehensive details about flight paths, hours of departure and arrival, airline companies, the cost of airline tickets, and more.

Using this data, we have drawn insights on how the Airlines, Routes, Departure and Destination Airports, Arrival Time and Departure Time, Total Travel Time, Ticket price impact the customer bookings. It also explains what are the factors that are considered by individual customers before proceeding to booking flights.

From the data visualizations and analytics that are performed by using tableau on the Priceline Comprehensive Flight Data. We have gathered insights that help Priceline firm to make better business decisions from the analysis provided, which helps to increase the customers and thereby profits.

Overall, the Comprehensive Flight Data from Priceline dataset is a useful for researchers, analysts, and anyone working in the travel sector, but it should be combined with other data sources and contextual information to give a complete picture of the travel industry.