

|  |
| --- |
| Assignment 1  PS1 – Airline Data Analysis |
|  |
| October 20  Group 23  Authored by:  Snigdha Tarua – 2019ab04171 Mahesh Sai – 2019ab04135 Chandra Sekhar Gupta – 2019ab04187 |

# Airline Data Project

# The first step of a data visualization project should start with identification of visualization context.

|  |
| --- |
| Visualization Context  Before we begin with data visualization it is important to get our context correct. It mainly focusses on answering three right questions. |
| 1. Who are our audience? 2. What we want our audience to do? 3. How to do that with support from our data? |
| Given Situation:  US commercial airline industry is one of the most diverse, dynamic and perplexing in the world. It is fast evolving, labor intensive, capital intensive, hyper competitive and highly susceptible to ebb and flow of business cycles as well as being among the most regulated of deregulated businesses.  Airline Data Project (ADP) was established by MIT Global Industry Program to better understand the opportunities, risks facing in the vital industry.  Data has to be analyzed to present a view of industry and its important trends, as well as to identify fundamental drivers of success - and early cases of failure.  Context in our case:   * Who are our audience?   Airline Data Project need to talk about opportunities, risks in the industry and trends, fundamental drivers of success and also failures.  So, audience in our case can be:  People of MIT Global Airline who are trying to understand the important aspects of data and use it accordingly.   * What we want our audience to do?   We want them to understand any positive/ negative trend and try to take decision accordingly and improve its services.  Basically, to understand the intuition about what is happening in terms of profit or expenses with data given.   * How to do that with support from our data?   By extracting insights hidden within the raw numbers and presenting in an interpretable form. |

Exploring our airline data

After playing around with the data that is provided over different domains related to problem statement, we came up with four questions which gave an advantage to exploit some insights in the data.

* 1. What are the expense trends of the airlines with respect to different narrow-bodies (based on plane sizes) over the past year?
  2. What is the expense trends of the airlines over their employee category in last few years?
  3. What is the relation between traffic and capacity by operating regions over the years for United Airlines? Are there any unusual behavior in the above values in past year?
  4. Do the airline services have an ideal load factor?

For all of the provided questions we tried to answer the following questions:

* What type of viz did you create?
* Why did you select the viz that you did?
* Identify Gestalt principles employed.
* Mention the pre-attentive attributes used.

Question 1:

Problem Statement:

1. What are the expense trends of the airlines with respect to different narrow-bodies (based on plane sizes) over the past year?

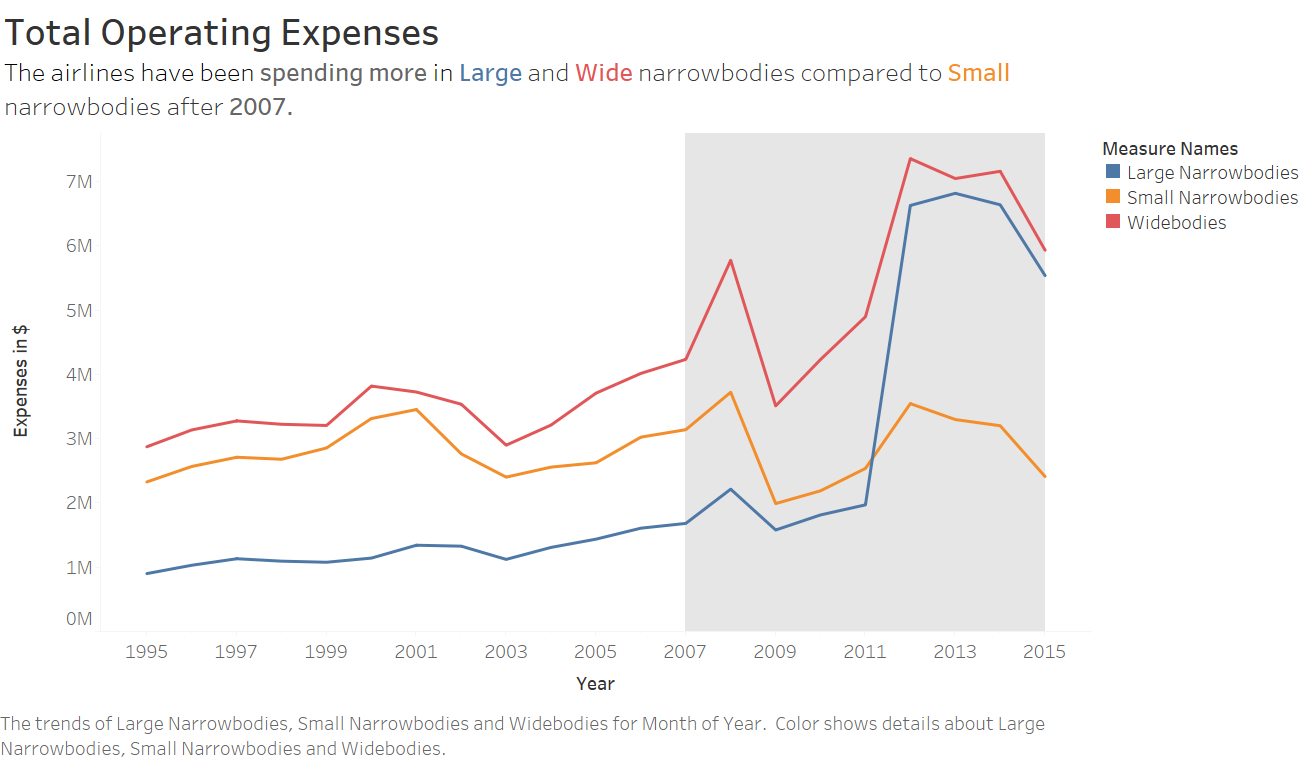
Domain Information:

In keeping with the ADP’s efforts to aggregate data to ensure meaningful comparisons, fleets are broken down into the following categories:

* Small narrow body aircraft: Typically, 150 seats or less in a two-class configuration (e.g. Boeing 737-700, Airbus A319)
* Large narrow body aircraft: Typically, 151 seats or more in a two-class configuration (e.g. Boeing 737-800/900/Max 8/Max 9, Boeing 757, Airbus A321/A320 NEO/A321NEO)
* Widebody aircraft: Two-aisle configuration.

The above information is extracted from reference link provided.

Data Visualization:



* What type of viz did you create?

A graph containing three-line charts that analyze total expenditure spent on three different plane categories.

* Why did you select the viz that you did?

Main objective of the graph was to point out increase in expenditure over the years for different plane categories and also to understand steep increase in large and wide narrow bodies after 2007.

Even though bar charts can help in visualizing increasing trend, line charts portray increase in steepness better than bar charts.

* Identify Gestalt principles employed.

Principle of Similarity:

The principle of similarity states that when things appear to be similar to each other, we group them together.

In the above graph, same colour is used for respective line graphs and description in the text.

Principle of Enclosure:

This principle points out that any form of visual enclosure causes us to see the enclosed objects as a group.

In the above graph, even though we are not having any border it is understandable that graph is a whole component.

Principle of Figure and Ground:

The figure-ground principle states that people instinctively perceive objects as either being in the foreground or the background. They either stand out prominently in the front (the figure) or recede into the back (the ground).

In the above graph, the shaded region of the graph draws user attention there by pushing the trend to the front.

Principle of Proximity:

The principle of proximity states that things that are close together appear to be more related than things that are spaced farther apart.

The axes titles and axes are close to each other so that there is no need to explicitly mention these two belong to same category.

* Mention the pre-attentive attributes used.

Colour:

Used grey colour to drew user’s attention to required part of the graph.

This actually helps to grab immediate attention in required part of the graph.

Size:

Used font with bigger size to relate colour and related line graph also to notice the trend of increasing nature in the graph.

Spatial positioning:

Finally, a top-down and left-right approach is being used to position the graphs to provide a visual hierarchy of information that will help make it clear to the audience how they should interact with the information that is being provided.

Question 2:

Problem Statement:

1. What is the expense trends of the airlines over their employee category in last few years?

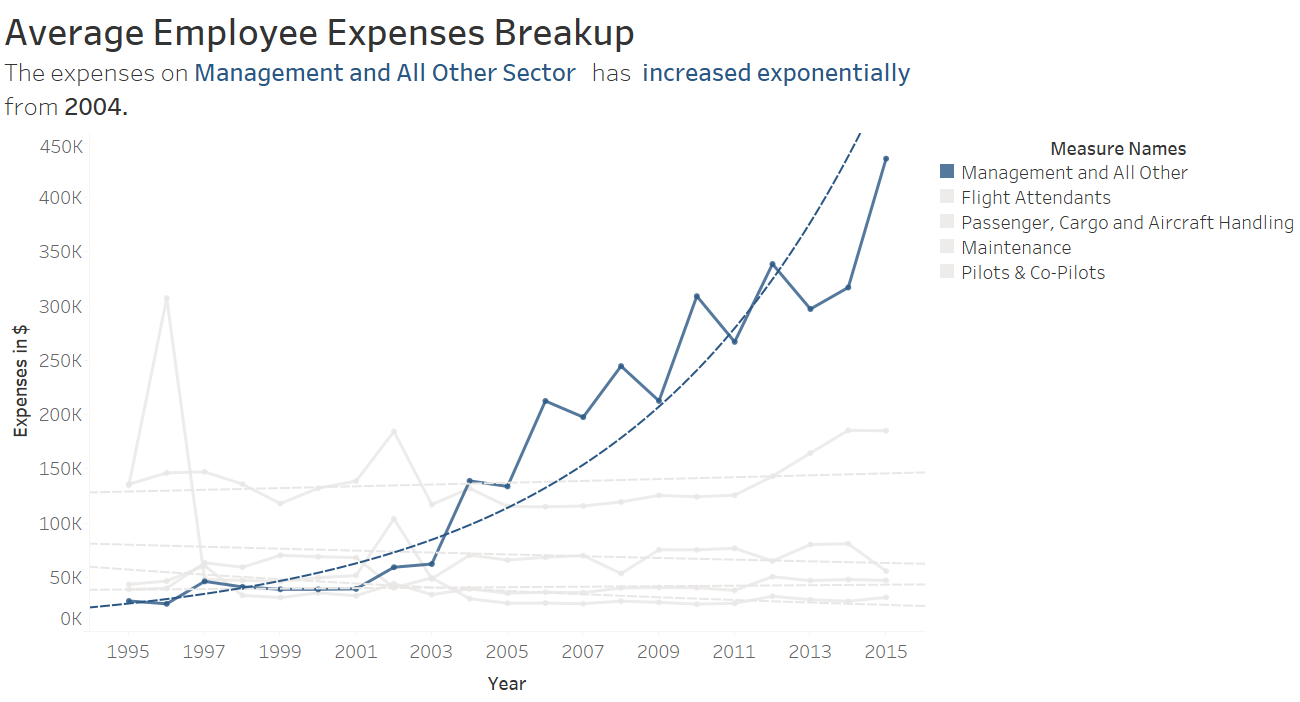
Domain Information:

Employees are divided majorly into 5 categories:

1. Management and All Other  
2. Flight Attendants  
3. Passenger, Cargo, and Aircraft Handling  
4. Maintenance  
5. Pilots & Co-Pilots

Each employee will have a salary and its benefits and expenses like Productive Measures. All together gives employee expenses to the company.

Data Visualization:



* What type of viz did you create?

We used multi-line chart to compare employee expenses by the company over different categories and to identify which category is getting highly beneficial.

* Why did you select the viz that you did?

Line graphs help us to understand trends over time.

Multi-line chart helps the users to compare the trend across different categories and to understand which category is out-performing.

* Identify Gestalt principles employed.

Principle of Similarity:

The principle of similarity states that when things appear to be similar to each other, we group them together.

Similar colour is used for denoting Management in text and line graph so that the viewers can relate the measure to their respective meaning from the text. Using the principle, there is not requirement of a separate legend for the graphs.

Principle of Closure:

The principle of closure states that when we look at a complex arrangement of visual elements, we tend to look for a single, recognizable pattern.

The trend line shown is a dashed line and on seeing the it we can perceive to be a connected line.

Principle of Focal point:

The focal point principle states that whatever stands out visually will capture and hold the viewer's attention first.

Management line is brought to foreground and rest all are faded out, this helps the user to focus on Management expenses and able to find insights faster.

Principle of Proximity:

The principle of proximity states that things that are close together appear to be more related than things that are spaced farther apart.

The axes titles and axes are close to each other so that there is no need to explicitly mention these two belong to same category.

* Mention the pre-attentive attributes used.

Colour:

Using the blue for indicating valuable information and to draw the audience's eye to the most important part of the visual. This shows on what audience should focus on.

Size:

The eye catches the title (font is bigger than that which is around it, also the bold is a signal that it's important) and scans it so viewers know what they are looking at.

Spatial positioning:

Finally, a top-down and left-right approach is being used to position the graphs to provide a visual hierarchy of information that will help make it clear to the audience how they should interact with the information that is being provided.

Question 3 & 4:

Problem Statement:

1. What is the relation between traffic and capacity by operating regions over the years for United Airlines? Are there any unusual behavior in the above values in past year?
2. Do the airline services have an ideal load factor?

Domain Information:

The most fundamental data to any analysis of the airline industry are traffic, capacity and the relationship of one to the other:

1. Traffic, measured by Revenue Passenger Miles (RPMs),
2. Capacity, measured by Available Seat Miles (ASMs),

are together used to determine:

1. Average Load Factor (ALF) – the proportion of airline output that is actually sold.

 In turn, these metrics provide the basis for measures of unit revenue and unit costs.

A revenue passenger mile (RPM) is a transportation industry metric that shows the number of miles traveled by paying passengers and is typically an airline traffic statistic. Revenue passenger miles are calculated by multiplying the number of paying passengers by the distance traveled. For example, an airplane with 100 passengers that flies 250 miles has generated 25,000 RPM.

Available seat miles (ASM) is a measure of an airplane’s carrying capacity available to generate revenues. Available seat miles refers to how many seat miles are actually available for purchase on an airline. Seat miles are calculated by multiplying the number of miles that a given airplane will be flying by the number of seats available for a given flight.

The load factor (LF) is a metric used in the airline industry that measures the percentage of available seating capacity that has been filled with passengers.

Data Available:

Information is for United Airlines.

Features :

Atlantic Operations

Latin America Operations

Pacific Operations

Domestic Operations

International Operations

Sub-Features :

|  |
| --- |
| Sum of Pax Rev (000) |
| Sum of Total Rev (000) |
| Sum of RPM |
| Sum of ASM |
| Sum of AC Op Expense (000) |

Total

International Operation Percentages

Atlantic Operations Percentages

Latin America Operations Percentages

Pacific Operations Percentages

Domestic Operations Percentages

Atlantic Operations Percentage of International

Latin America Operations Percentage of International

Pacific Operations Percentage of International

Yield - Pax Revenue / RPM

RASM - Pax Revenue / ASM

CASM Aircraft Operations - Expenses / ASM

Region-Wise :

|  |
| --- |
| Atlantic |
| Latin America |
| Pacific |
| Domestic |
| International |
| Total |

Load Factor - RPM/ASM

Sum of RPM

Sum of ASM

Sum of Pax Rev

Sum of Aircraft Expense

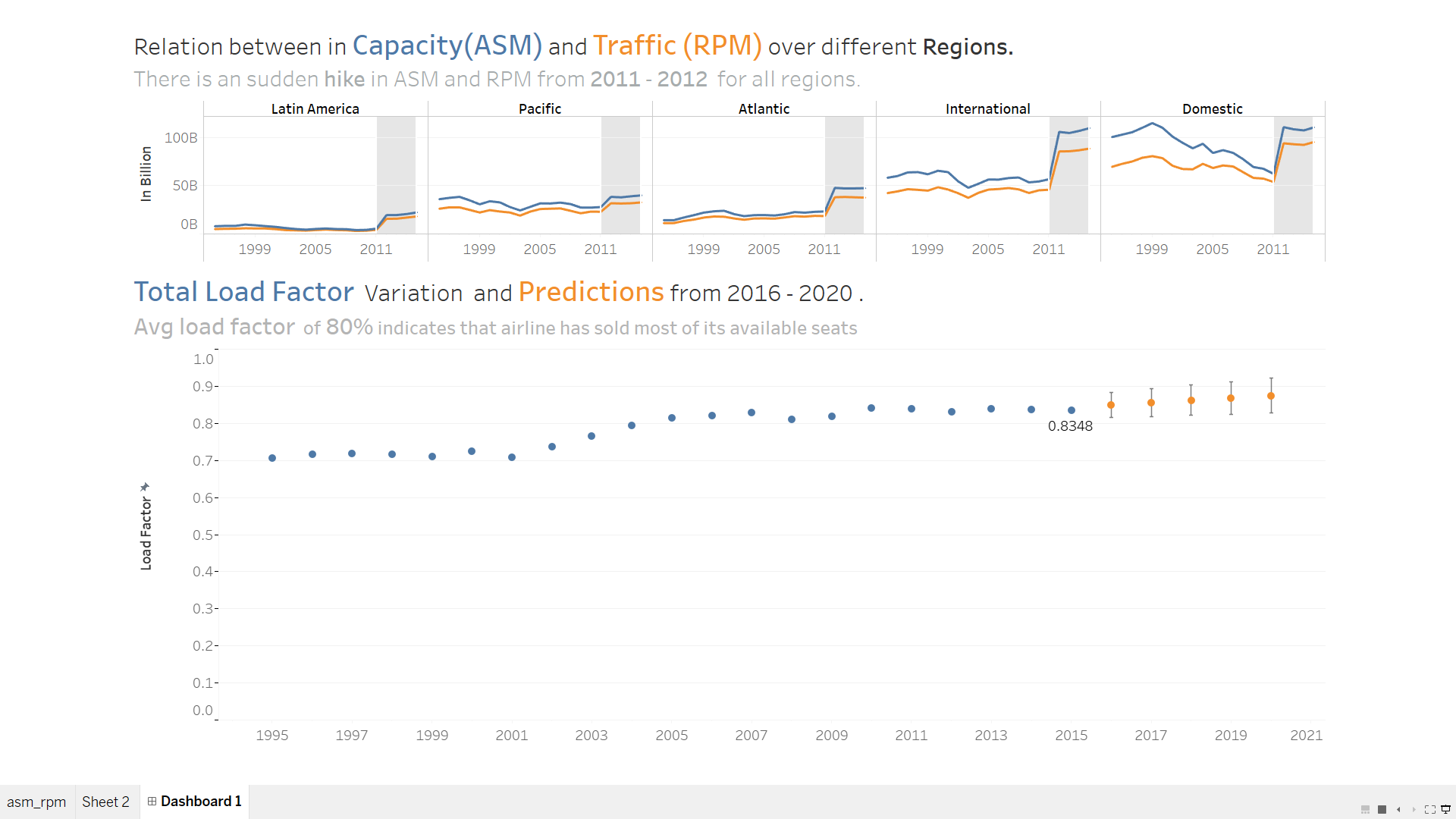
Aircraft Expense Percentage

ASM Percentage

Pax Rev Percentage

RPM Percentage

Data Visualization:



* What type of viz did you create?

1. A line chart is used to track both ASM and RPM measures over the period 1995 – 2015 in different operating regions.
2. Analyze Load Factor over the past year for United Airlines and its prediction for the upcoming five years using a scatterplot.

* Why did you select the viz that you did?

1. Line graphs are used to track changes over short and long periods of time. When smaller changes exist, line graphs are better to use than bar graphs. Line graphs can also be used to compare changes over the same period of time for more than one group.
2. Scatterplots are used to plot data points on a horizontal and a vertical axis in the attempt to show how much one variable is affected by another.

* Identify Gestalt principles employed.

Principle of Similarity:

The principle of similarity states that when things appear to be similar to each other, we group them together.

Similar color is used for denoting Capacity/ASM and Traffic/RPM in text and line graph so that the viewers can relate the measure to there respective meaning from the text. Using the principle, there is not requirement of a separate legend for the graphs.

Principle of Enclosure:

This principle points out that any form of visual enclosure causes us to see the enclosed objects as a group.

To show the hikes, the line has been enclosed within the shaded region. This is to draw attention of the users on unusual pattern over the year.

Principle of Continuity:

The principle of continuity states that elements that are arranged on a line or curve are perceived to be more related than elements not on the line or curve.

The dots for the Load factor, the elements that are arranged on a line or curve are perceived to be more related than elements not on the line or curve.

Principle of Proximity:

The principle of proximity states that things that are close together appear to be more related than things that are spaced farther apart.

The axes titles and axes are close to each other so that there is no need to explicitly mention these two belong to same category.

* Mention the pre-attentive attributes used.

Colour:

Using the color orange and blue for indicating valuable information and to draw the audience's eye to the most important part of the visual. This shows on what audience should focus on.

Orange is used for predictions of load factor because it increases cheerfulness , optimism and stimulates logic of center of brain .Since the load factor values are increasing we need to drive the attention of the viewers to the valuable information.

Size:

The eye catches the title (font is bigger than that which is around it, also the bold is a signal that it's important) and scans it so viewers know what they are looking at.

Spatial Positioning:

Finally a top -down and left-right approach is being used to position the graphs to provide a visual hierarchy of information that will help make it clear to the audience how they should interact with the information that is being provided.