Airline Safety from 1985 to 2014



**Julio Corteza**

September 2020

Applied Data Science Capstone

# INTRODUCTION

As air travel rebounds from the effects of the Covid-19 or SARS-CoV-2 worldwide pandemic, airline safety in terms of past events which can be categorized into incidents, fatal accidents, and fatalities are still important to explore. I will examine airline event data from 1985-2014 from 56 airline carriers worldwide. This report’s target audience will be the air traveler and perhaps will assist in the selection of which airline to choose. It is commonly known that higher profile airline disasters result in less demand to travel with that specific airline.

# PROBLEMS

1. What significant information can be gathered from past airline incidents or accidents?
2. Can a traveler who is concerned about safety make a decision about which airline to use from past airline safety data?

# DATA

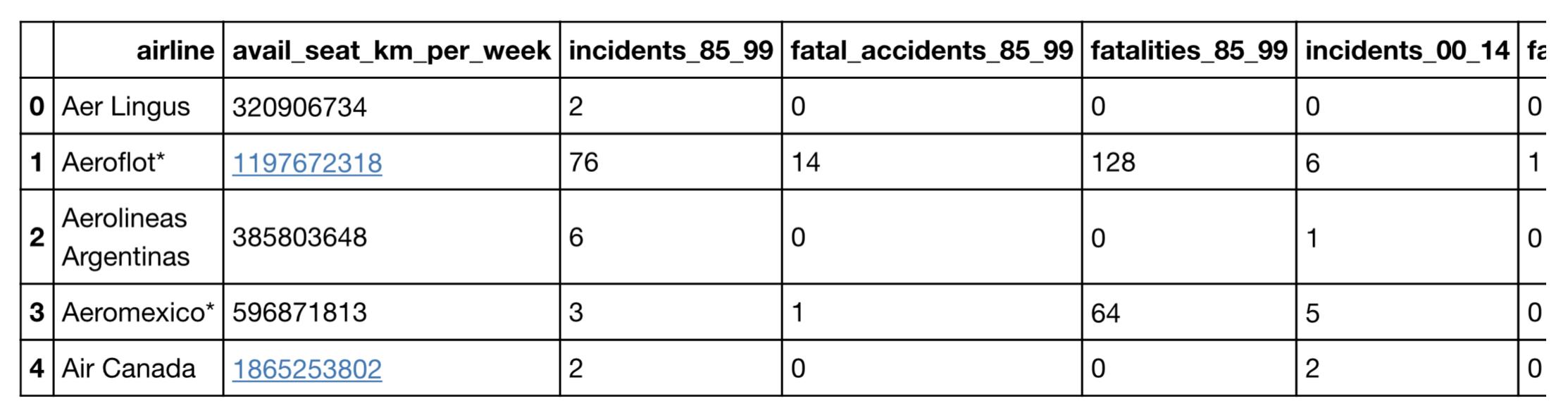
Data for airline safety from 1985 to 2014 was taken from the url link below. The dataset was inputted into a Jupyter notebook using IBM Watson Studio and Python programming language.

<https://raw.githubusercontent.com/fivethirtyeight/data/master/airline-safety/airline-safety.csv>

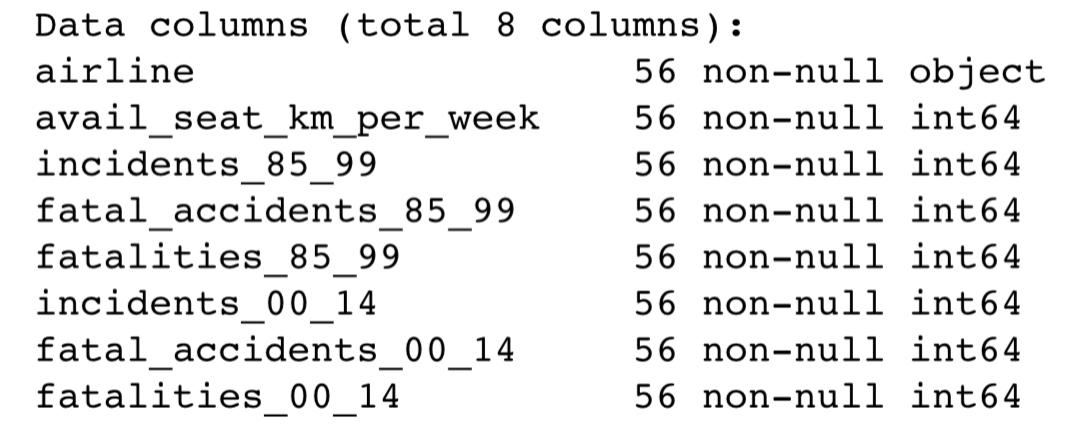
Figure 1: Part of Data Table Output



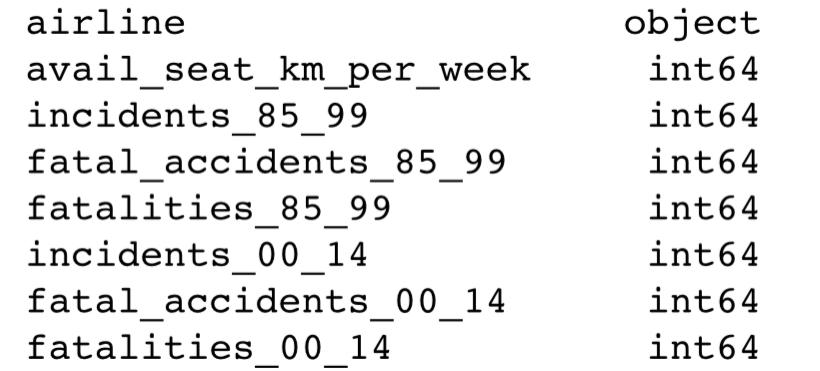
***Figure 2: Part of Data Header Output***

******

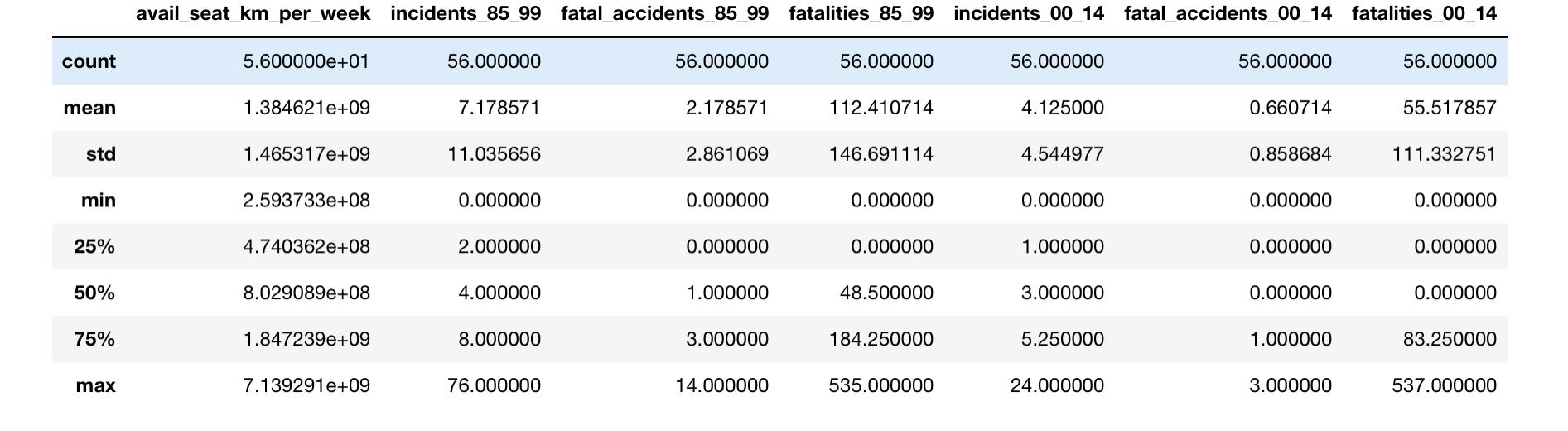
***Figure 3: Data Columns Output***



***Figure 4: Data Type Output***



***Figure 5: Data Describe Output***

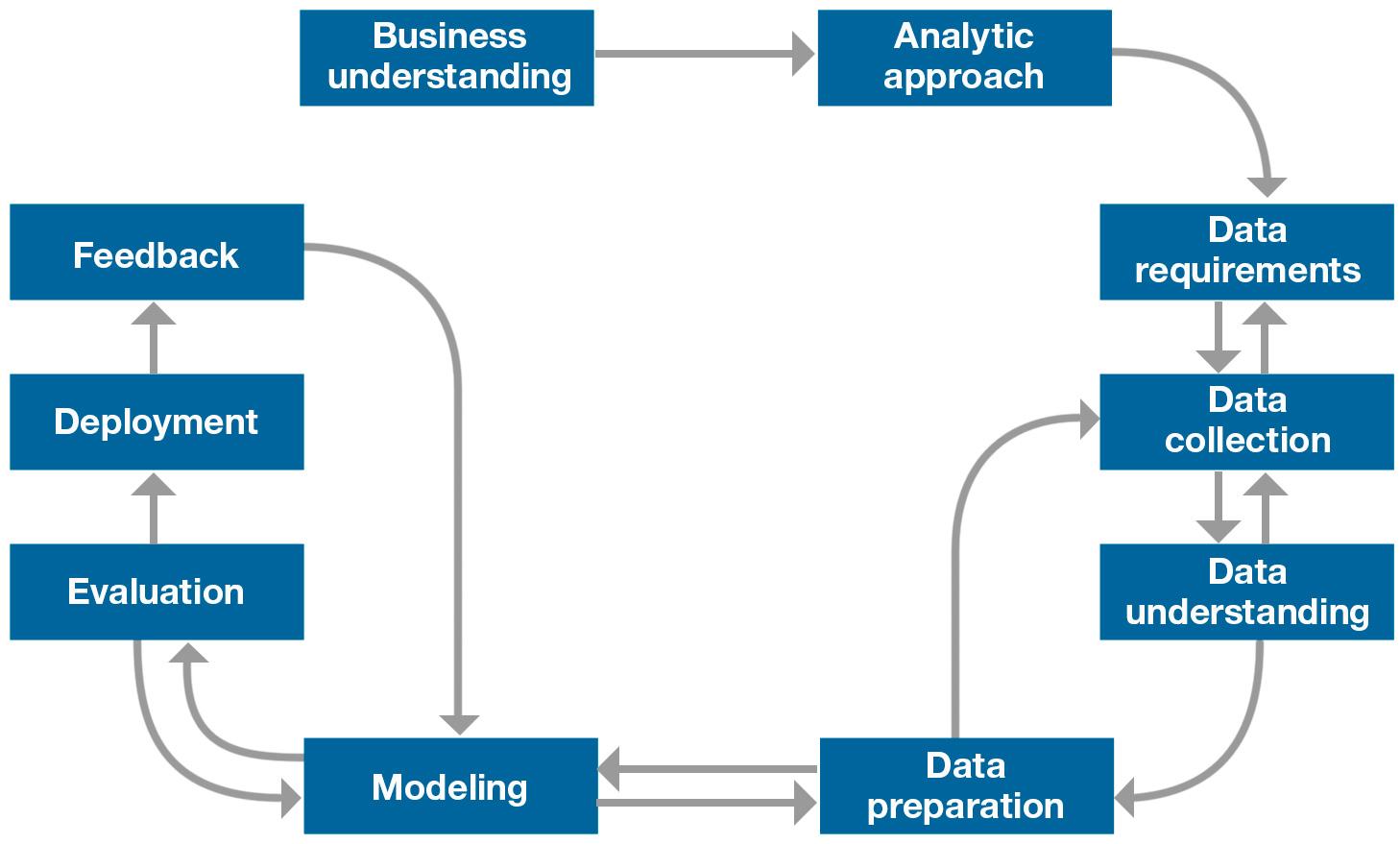


***Preliminary Data Discussion***

The output results show eight columns with 7 integer features and 1 categorical feature with a total of 56 airlines worldwide represented and 8 total features. The data set appears to be clean with no requirement for cleaning. There appear to be no missing values and unique values are displayed. It is duly noted that this is a smaller data set with less features, nonetheless valuable and important. The avail\_seat\_km\_per\_week were unadjusted.

# METHODOLOGY

***Figure 6: IBM Foundational methodology for data science***



In the previous Introduction and Problems sections of this report, two problems were presented: 1.) What significant information can be gathered from past airline incidents or accidents? 2.) Can a traveler who is concerned about safety make a decision about which airline to use from past airline safety data? The air traveler would be interested in this project. Data was collected for airline safety from 1985 to 2014 from the following url link:

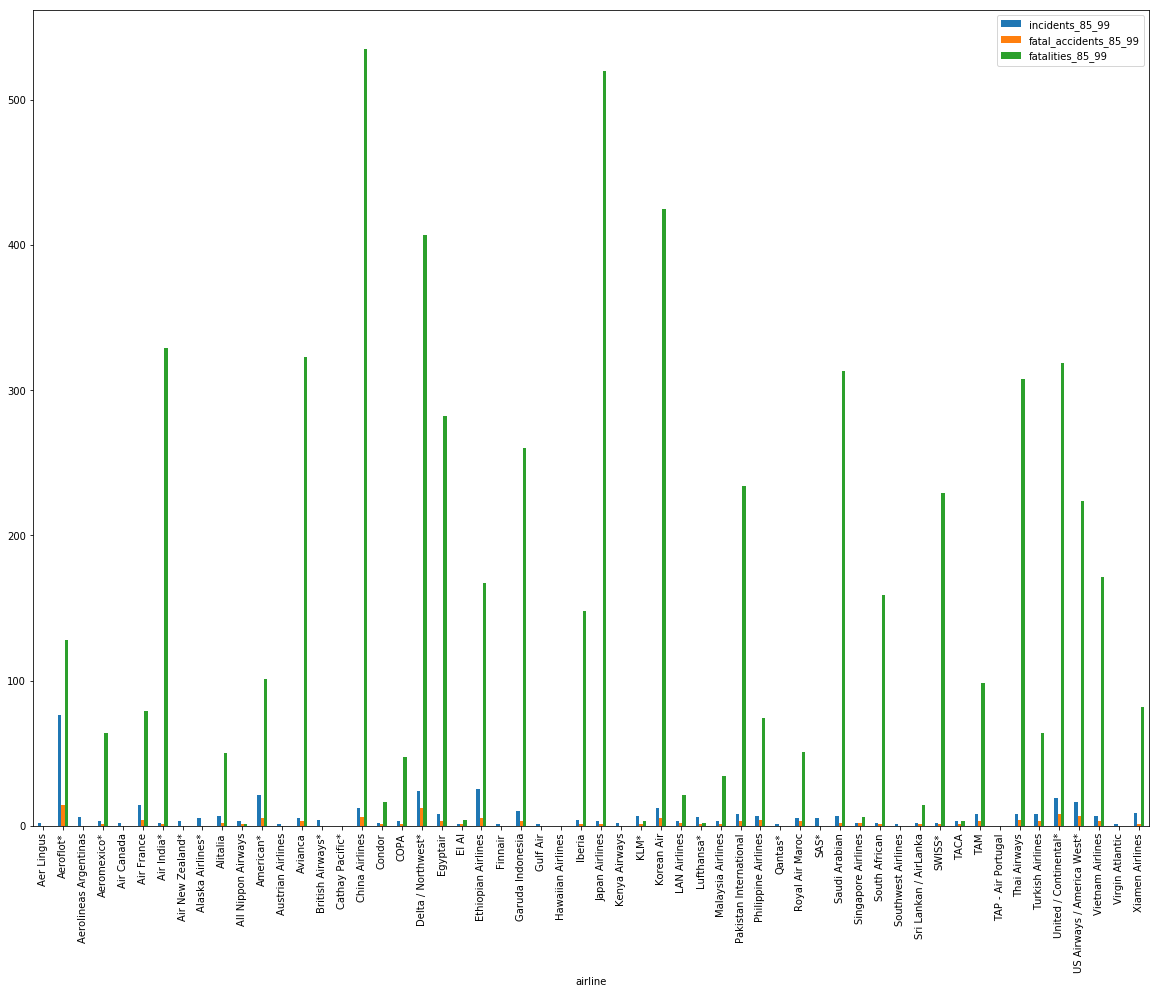
<https://raw.githubusercontent.com/fivethirtyeight/data/master/airline-safety/airline-safety.csv>

A Jupyter notebook was created within IBM Watson Studio using the Python programming language importing primarily NumPy, pandas, plotly, seaborn, and Matplotlib. Data was analyzed by performing an info ( ) function for a concise summary of the data-frame. The dtypes function was then used to determine the data type of each column. The describe ( ) function was performed on the data-frame to view basic statistical details like count, percentile, std, and min. Then the nunique ( ) function was performed to return the number of unique elements. A heat map was created using seaborn. The bar graphs were created to show the airline on the X-axis and the number of incidents, fatal accidents, and fatalities on the Y-axis. A cut function was used to bin or segment and sort values with incidents, fatal accidents, and fatalities . Then the min and max function was used to determine the minimum and maximum values in regards to incidents, fatal accidents, and fatalities. A model/code was created for inputting the airline name, for example Aeroflot which would generate a count of incidents, fatal accidents, and fatalities to help in the deciding whether to use a particular airline for air travel from past events.

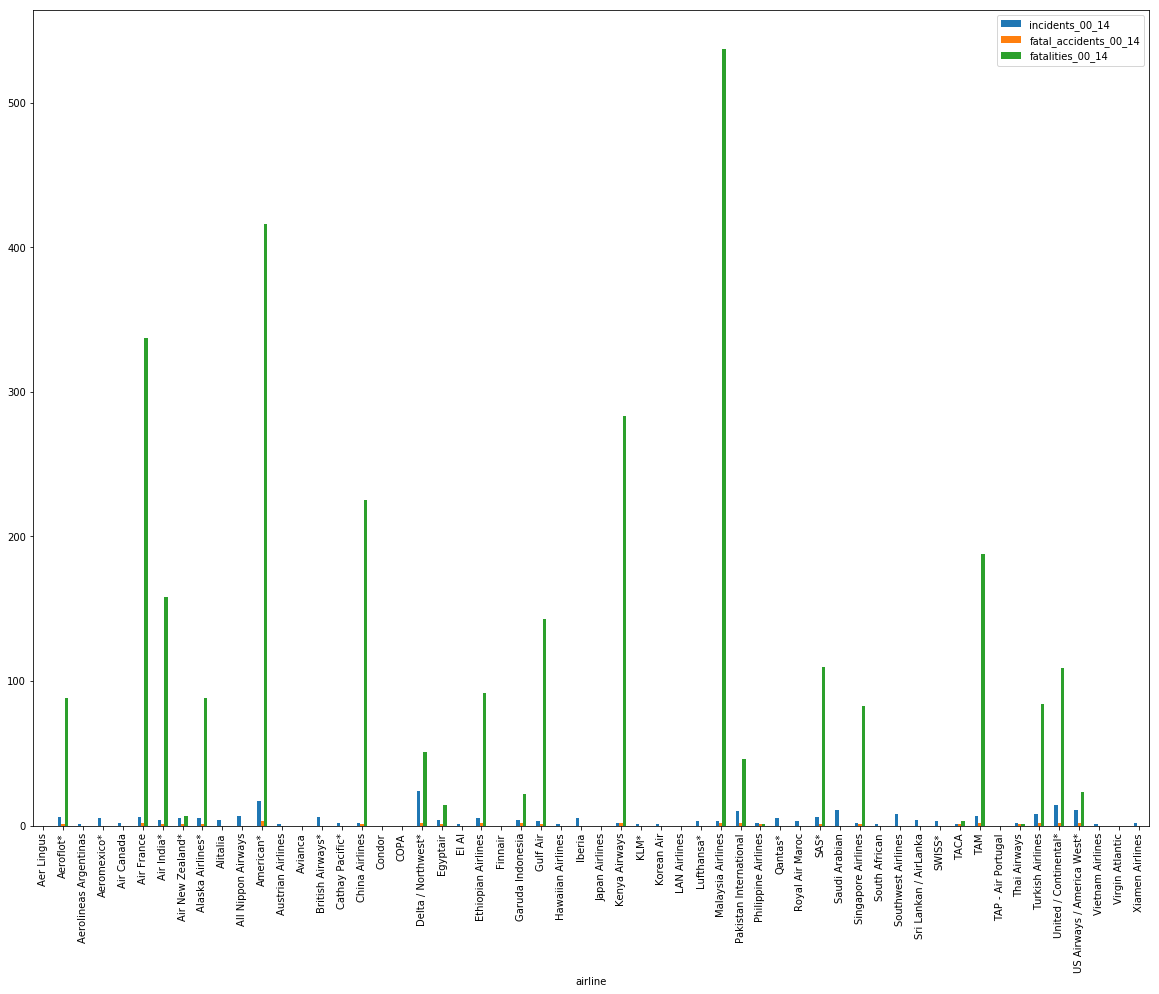
***Figure 7: Heatmap***



***Figure 8: Bar Graph 1985-1999***



***Figure 9: Bar Graph 2000-2014***



# RESULTS

The output results show eight columns with 7 integer features and 1 categorical feature with a total of 56 airlines worldwide represented and 8 total features. The average airline incidents between 1985-1999 is 7, the average fatal accidents is 2, and the average fatalities is 112. The average incidents between 2000-2014 is 4, the average fatal accidents is less than 1, and average fatalities is 55. As a result, between 1985-1999 there were more airline incidents, fatal accidents, and fatalities in comparison to the period of time from 2000-2014. With the heat-map one is able to visualize the intensity of values in our study, which is fatalities. TAP - Air Portugal reported 0 Incidents, Fatal Accidents, and Fatalities for the time periods 1985-1999 and 2000-2014. Aeroflot recorded the most incidents 82 in total for the time periods 1985-1999 and 2000-2014. Malaysia Airlines recorded the most fatalities at 571 for the time periods 1985-1999 and 2000-2014.

DISCUSSION

From the time period of 2000-2014 with lower incidents, fatal accidents, and fatalities in comparison to the time period 1985-1999 are perhaps a result of technological advances in airplanes, enhanced flight crew training, airplanes that are newer, etc.

Some Further Considerations or Weaknesses:

1.) An airline may not fly a specific route to a destination.

2.) An airline may have more flights than another airline.

3.) The reason (e.g. weather, human error, etc.) for the airplane accident/ incident.

4.) What is the impact on the airline’s reputation?

CONCLUSION

It is both possible to find significant information from past airline incidents or accidents and a traveler can make a perhaps more informed decision about which airlines he/she should use for air travel. In particular, it could be wise to reconsider air travel with Aeroflot and Malaysia Airlines. Though there is also evidence that shows incidents, fatal accidents, and fatalities have fallen from the period of 2000-2014. It seems that further work on this project would be interesting and may benefit from examining further considerations and weaknesses.