UNIVERSITY OF SKÖVDE

AIRLINE ON-TIME PERFORNAMCE VISUALIZATION USING TABLEAU

IT730A Business Intelligence A1F
Student Name: Welemhret Welay Baraki
Email: a20welba@student.his.se

Course Instructor: Mikael Berndtsson

APRIL 11, 2022

1. Introduction

Nowadays, Business Intelligence Technologies (Chaudhuri, Dayal & Narasayya, 2011) are widely deployed and used in businesses and industries to be able to analyze large volumes of data and extract insights to support knowledge workers for their decision-making capabilities. Particularly, the airline business sector (Fatima, 2022) uses different Business Intelligence technologies to analyze and monitor climate change, control and analyze flights, as well as forecasting weather conditions to detect risks. Generally, this project deals with data collection and data preparation using Python (i.e. Pandas) for the Tableau BI Tool, and finally the required parameters and features are analyzed, interpreted and reported with the visuals generated. In this project business intelligence analytical questions were answered and ethical challenges and dilemmas accessed.

2. BI Analysis and Visualization of years 1989 and 1999

2.1. Data Collection and Preparation

The US airline flight data was downloaded and prepared based on the (Harvard Dataverse ,2008) for the years 1989 and 1999. The dataset of the two years has 10 million records and 29 features. The features that are null values have excluded and finally the 1989 and 1999 records were combined and exported to comma separated values (CSV values) to be able to be visualized in the Tableau Desktop software. Additionally, airport names, region/state/ names, and unique carrier/airline names/ were added in separate columns to the dataset to use an intuitive visual display instead of using IATA Codes which are difficult to read and understand easily for users.

2.2. What characterize flights that are on-time?

The on-time performance analysis in the US domestic flights are characterized based on regions/states/, airline carriers and airports. The tableau visualizations are publicly available as a story[1] in the tableau public repository. To characterize the on-time flights, as an expert in the business intelligence area I have formulated questions in the business domain that are going to be answered using visualizations. Based on users preference of the years, quarters, months and weekdays many questions can be formulated to extract business insights based on maps, bar charts and line charts. The states that have high flight status (i.e. on time, delayed and cancelled flights) are shaded in the map with SteelBlue colors and low flight status are represented with light blue as shown in the figure 1. Based on the figure 1 below, in 1989, California and Texas are the states that shows high on-time number of flights with 315,024 and 294,510 flights respectively. In 1999, California, and Texas are the two top states that score high on the on-time flights performance with 380,663 and 348,115 flights respectively.

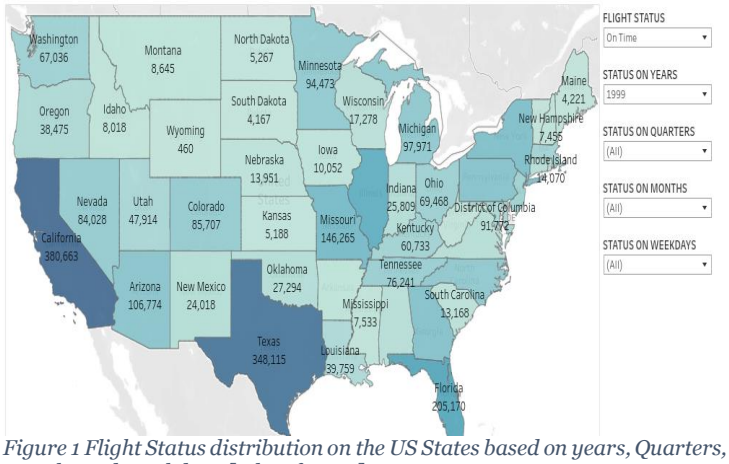

Figure 1 Flight Status distribution on the US States based on years, Quarters, Months and Weekdays [taken from[1]]

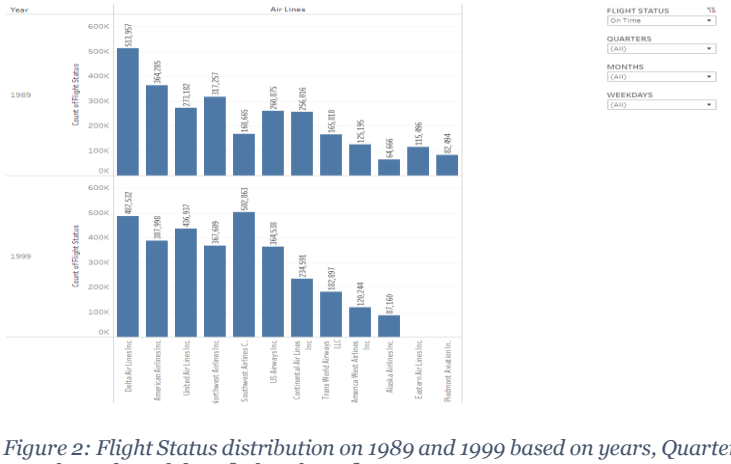

Figure 2: Flight Status distribution on 1989 and 1999 based on years, Quarters, Months and Weekdays [taken from[1]]

[1] This projects airline on-time performance visualization is publicly available in URL: Visualization of Airline Ontime Performance | Tableau Public

***Note**: Some of the visualizations are not complete to illustrate in this short report, so that I recommend referring the website for More.*

In the year 1989, the states that have high cancellation of flights are Georgia with 6,593, California with 6,305, Florida with 6,281, Illinois with 5,696, New York with 5,452 and Texas with 5,276 flights cancellations. In 1999, Illinois, California and Texas are the top states that score high cancelations of US domestic flights. The number of delayed flights by region in 1989, in states of California and Texas are recorded as 248,994 and 231,978 respectively based on my visualization. Whereas in the year 1999, the highest number of delayed flights are in California with 259,279 and Texas with 255,084 flights. The delayed and on-time shows similar pattern that California and Texas have high number of delayed or on-time flights. Whereas the cancelled flights in the two years showed different patterns. The cancellation of flights in the year 1999 have increased in the states of Illinois, California and Texas. This can be due to the air pollution and other environmental effects.

Figure 2 shows the distribution of on-time flights of 1989 and 1999 based on the twelve airline carriers in the US domestic flights. Generally, the Delta Airlines and American Airlines are the top two airlines that arrive/depart/ on-time in the two years of 1989 and 1999. In 1989, Delta Airline is the leading airline carrier on arriving and departing on-time. Whereas in the year 1999, the Southwest Airlines, Delta Airlines and United Airlines are the top three airlines that have high on-time flights in the US domestic flights. Generally, Delta Airline can be the most recommended one to customers based on their highest on-time performance. Eastern Airlines ("Eastern Air Lines - Wikipedia", 2022) are dissolved and closed in 1991, and the Piedmont Airlines was merged within the brand of American Airlines ("Piedmont Airlines (1948–1989) - Wikipedia", 2022). As a result, there is no separate record of flights in 1999 for these two airline carriers. In the year 1989, the Eastern Airlines have the highest flight cancellations. Whereas in the year 1999, US Airways have the highest flight cancellations. US Airways have the highest number of flight delays in 1989 and the Delta Airlines have the highest number of flight delay in 1999. Considering the average arrival and departure delay for the delta airlines shows an average of 7.5 Minutes of delaying. Possibly this airline can among the comfortable and preferable airline in the US.

We can also characterize on-time performance based on a particular airport. Hartsfield–Jackson Atlanta International Airport have the highest on-time flight status followed by Dallas Fort Worth International Airport and Chicago O'Hare International Airport in 1989 as shown in the figure 3 below. The leading airports in the on-time flight status in the year 1999 are Chicago O'Hare International Airport, Dallas Fort Worth International Airport and Hartsfield–Jackson Atlanta International in their respected order. Punctuality pays off, so that the airports that have high on-time flights are more preferable by the customers and partners and are the more likely satisfy their customers and partners.

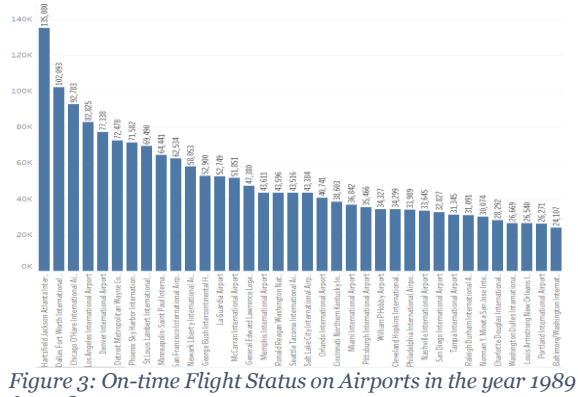

Figure 3: On-time Flight Status on Airports in the year 1989 [taken from[1]]

Figure 4 Monthly average arrival and departure delays 1989 [taken from[1]]

Flight delays have negative economic impacts for passengers, airlines and airports. So as analyzing and visualizing the flight delays is critical in the aviation industry. As illustrated in the line graph of figure 4 of 1989, the month of June have the highest arrival delay, and the month of April records the lowest arrival and departure delays in the US domestic flights. December have the highest departure delays in 1989. In the year 1999(i.e. refer to the online visualization), the month of January and June have high average monthly arrival delays, and November have the lowest average arrival delay. January have the highest departure delay, whereas September have the lowest departure delay in the year 1999. From both years, the month of June shows similar departure and arrival delays. The airports of Gustavus and Petersburg

James Airports have high departure and arrival delays an average of 18 to 19 minutes. These airports are found in Alaska. These delays can be due to the environmental conditions of that particular arctic region.

2.3. Which regions, airline carriers and airports have highest diverted flights?

Diverted flights can be caused due to many reasons. In this project I will visualize diverted flights based on the US states, Airline carriers and different airports. To explore more flexible views, the Tableau Visualizations are publicly available in the public repository in[1]. According the US Regions of figure 5a and diverted analysis, California, Texas and Florida are the top three states that record high number of diverted flights in the 1989 US domestic flights. In 1999, Texas, California and Florida are the states that have high number of flights diverts. Diverted flights for a particular airport as shown in figure 5b are also visualized with maps and bubbles in each of the airports based on the number of diverted flights according to the destination airports. The number of diverted flights is spotted in the airports of the US map as bubbles of circles. Different visualizations can be produced based on the user's preferences of years, months, quarters, and weekdays. To view diverted flights from the perspective of airline carriers is important that customers preferences depends on, bar charts based on the twelve airline carriers are plotted. Based on the visualization of figure 5c below American Airlines have the highest diverted flights and American West Airlines have the lowest diverted flights in both the years of 1989 and 1999.

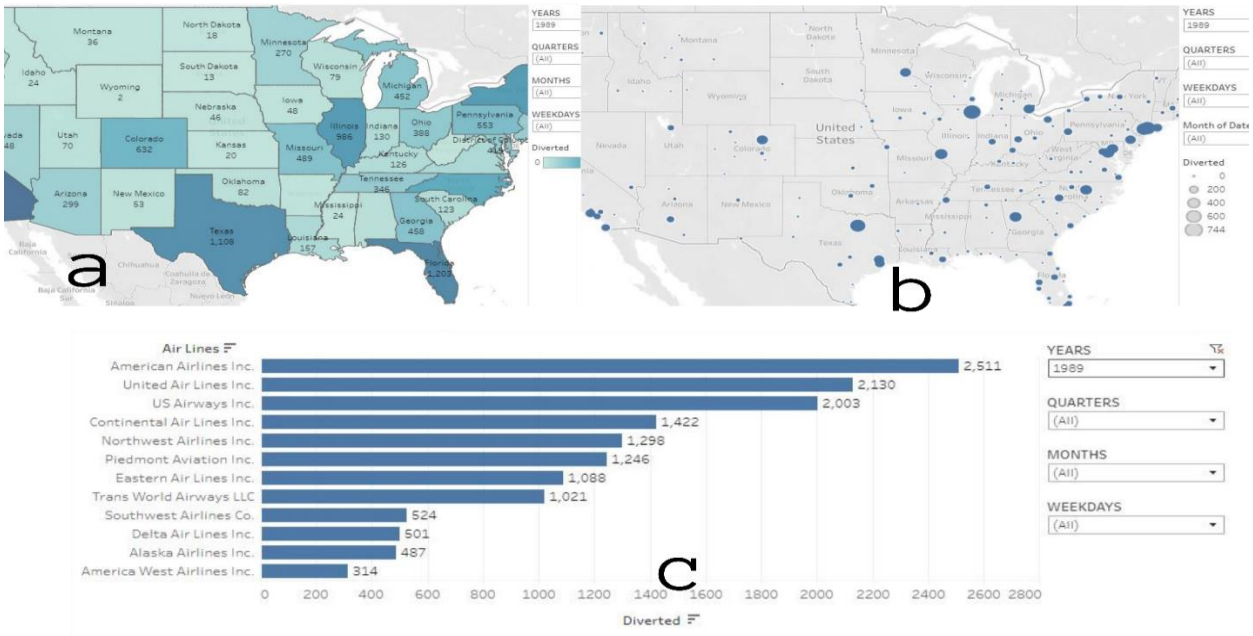

*Figure 5 Visualization of diverted Flights of the year 1989 based on Regions(**a**), Airports(**b**) and Airlines(**c**)*

3. Ethical dilemmas according to (Solove ,2006)

As Solove (2006), stated in the taxonomy of privacy that helps to classify and comprehend privacy issues and violations during the information collection, information processing, information dissemination, and invasion. In this assignment of on-time airline performance analysis and visualization, all of the four privacy taxonomy stated in (Solove, 2006) can rise different concerns.

The data that is going to collect by the airport is traveler's smart phones, flight plans, passport control, security control, visits to shops and restaurants, and their future plan is to install facial recognition system in shops and restaurants. The airport collects traveler's data without the legal consent of the user. Generally, many big airports collect your personal information from different sources of visits to restaurants and shops, and other airport activities. The traveler´s personal, private and sensitive information are collected by the airport without the any legal agreement with the traveler. This may lead the travelers frustrated, discomfort and insecurity. As a result, the user may not freely shop and visit restaurants. From the organization´s point of view, collecting traveler´s information will definitely benefit in increasing the airport´s profits by using different business analytics methods. From the user's point of

view, the traveler should have to have a legal consent with the airport and get explicit profit of his collected personal information from the airport´s in return. The information collected can be used for surveillance of the traveler. The collected data can be used to restrict passengers, delays, ask extra unnecessary cross-examination questions. This can be caused because of biased stereotypes. As a good example in Ethiopia, Tigrayan's of national ethnic groups faced interrogation in airports and other services throughout the country. The airport can violate traveler's privacy due to the anonymous collection of data from different sources.

Most of the time in information processing private data can be disclosed to different parts that are involved in the processing based on their roles. During information processing, private personal data can be aggregated to extract business insights which will endanger and target individuals or groups based on their activities in the airport. Processing and manipulation of sensitive information like passport control, security control, and smartphone may lead to violation of privacy breaches of insecurity and identification. In the era of technology, most of the data are processed in cloud platforms like amazon, google cloud and other cloud business intelligence services where sensitive data can be processed and interpreted out of context. One issue that may raise in this case is secondary use in which the data collected from restaurants, and shops will be shared with the airport without the explicit permission of the traveler, to remind if they are going to be late.

Information dissemination is the primary source of privacy issue which leads to disclosure and breach of confidentiality, and increased accessibility of private information like passport and security controls. This violates privacy and trust that creates unsafe condition to individuals or groups. Intentionally to harm or not, the disseminated private information can be distorted from different perspectives or personal or group interests. Generally, this may expose individuals or groups to biases, embarrassments, humiliation, suffering, trauma, and disclosure of sensitive private information.

Once the airports, restaurants and shops have access to collect your smartphones data, call history, messages and other private information can be easy to extracted and intrude the person´s activity, location and other personal affairs according to the (Solove, 2006) Solove taxonomy. The collected data can be used to threaten once person to enforce and drive his decision in a way that the airport or government wants to intend.

References

Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications Of The ACM*, *54*(8), 88-98. doi: 10.1145/1978542.1978562

Piedmont Airlines (1948–1989) - Wikipedia. (2022). Retrieved 1 April 2022, from https://en.wikipedia.org/wiki/Piedmont_Airlines_(1948%E2%80%931989)

Eastern Air Lines - Wikipedia. (2022). Retrieved 1 April 2022, from https://en.wikipedia.org/wiki/Eastern_Air_Lines.

Fatima, N., 2022. *How would the Airline Industry use Business Intelligence*. [online] Syntaxtechs.com. Available at: https://www.syntaxtechs.com/blog/business-intelligence-in-airline-industry [Accessed 24 February, 2022].

Harvard Dataverse (2008). Data Expo 2009: Airline on time data [Dataset]. http://doi.org/10.7910/DVN/HG7NV7

Solove, D. (2006). A Taxonomy of Privacy. *University Of Pennsylvania Law Review*, *154*(3), 477. https://doi.org/10.2307/40041279