Make A Smart Choice On Booking Your Flight!

Yu-Hsiang Sun[1], Jiayang Sun[2], Bo Hu[3]
Case Western Reserve University, Department of Statistics

Abstract

Nobody wants to gamble by choosing a flight that will have a high likelihood to be late. By utilizing the graphical techniques, we analyze the 2009 Expo data set to discover some patterns about the delayed flights according to time, locations, and different carriers, among other factors. We try to show a reasonable way to pick a flight based on different criteria. Visualizing the airline on-time performance can give an interested traveler information in making a smart decision on booking, and motivate airlines and airports to find some ways to improve their performance.

Profile of Data Expo 2009 Data Set

Name : On-Time Performance

Time Period : 1987/10 - 2009/05

Description :

This on-time arrival data set is for non-stop domestic flights by major air carriers, and provides such additional items as departure and arrival delays, origin and destination airports, flight numbers, scheduled and actual departure and arrival times, cancelled or diverted flights, taxi-out and taxi-in times, air time, and non-stop distance.

Summary of The Influential Factors On "On-Time Performance" :

Factors To Be Considered When Making A Reservation On A Flight

Time's Factors	Fixed - No Choice
	Flexible : **Month** , **Day of Week** , **Departure Time** , **Arrival Time**
Location's Factors	**Departure Airport** - Large / Medium / Small Hub , Non-Hub
	Arrival Airport - Large / Medium / Small Hub , Non-Hub
Airline's Factors	**On-Time Performance**
	Price / Comfort / Service - Not Included In Data Expo 2009 Data Set
Other Factors	**Distance** , **Taxi-Out / Taxi-In**

Causes of Flight Delays and Cancellations

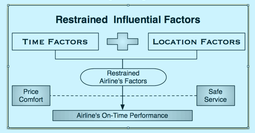
- **Air Carrier**: Due to circumstances within the airline's control (e.g. maintenance or crew problems, aircraft cleaning, baggage loading, fueling, etc.).
- **Extreme Weather**: Significant meteorological conditions (actual or forecasted) that, in the judgment of the carrier, delays or prevents the operation of a flight such as tornado, blizzard or hurricane.
- **National Aviation System** : Attributable to the national aviation system that refer to a broad set of conditions, such as non-extreme weather conditions, airport operations, heavy traffic volume, and air traffic control.
- **Late-Arriving Aircraft**: A previous flight with same aircraft arrived late, causing the present flight to depart late.
- **Security**: Delays or cancellations caused by evacuation of a terminal or concourse, re-boarding of aircraft because of security breach, inoperative screening equipment and/or long lines in excess of 29 minutes at screening areas.

Note :

- Taxi-Out/Taxi-In Time : The time elapsed between departure from the origin airport gate and wheels off / wheels down and arrival at the destination airport gate.
- We **mainly focus on the airlines (WN / AA / UA / US / CO / DL / NW /AS)** in this study. Because we only have complete data set for the eight airlines.
- We **mainly focus on Large/Medium/Small Hubs** in this study. Because not all airports report the "On-Time Performance Records" every year.

Basic Assumption

Influential Factors When Making A Flight's Reservation Are Restrained



Objectives

Find out the relationship between arrival on-time performance with the following factors:

1. Time's factors;
2. Location's factors;
3. Airline's Factors, Airline's Factors x Time's factors , Airline's Factors x Distance.

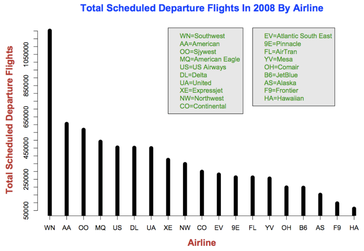
Methods

By showing the domestic market share of each airline by scheduled departure flights and the distribution of routes' distance for each airline, we can have a basic idea for each airline. (*Figure 1,Figure 2, and Figure 3*)

- As for the effect of the **Time's Factors** on arrival on-time performance. We can calculate the percentage of arrival on-time flights, or percentage of cancelled flights, or percentage of diverted flights according to different time frame. (*Figure 4 and Figure 5*)
- As for the effect of the **Location's Factors** on arrival on-time performance, within each kind of hub(Large/Medium/Small Hub), we can mark the departure airports which have better/ worse arrival on-time performance under each kind of hub. (*Figure 9 and Figure 10)*
- As for the effect of the **Airline's Factors** on arrival on-time performance, we can find out the airlines which have better/worse arrival on-time performance according to different time period or different distance's group. (*Figure 6, Figure 7, and Figure 8)*

Results

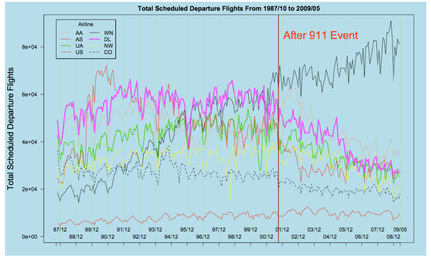
[Figure 1] - Market Share On 2008 For 16 Airlines



Note :

Southwest Airline has the largest domestic market share by the "Number of Scheduled Departure Flights" In 2008. From the legend, we can learn the abbreviation code for each airline.

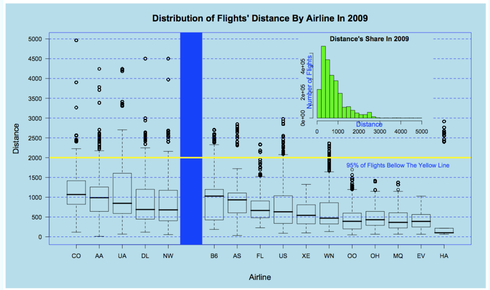
[Figure 2] - Historical Trend On Domestic Market Share



Note :

After 9/11, only Southwest Airline develops stably on the "Number of Scheduled Departure Flights". Other seven airlines wither in a very significant tendency.

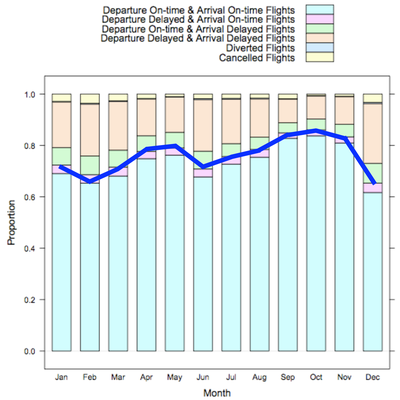
[Figure 3] - Distribution of Routes' Distance For Each Airline



Note :

On the left side of the blue bar, the five airlines serve the routes with distance greater than 3000 miles. As for Hawaiian Airline (HA), it only provides the routes within two distance groups (Distance < 200 and 2250 < Distance < 3000 Miles.)

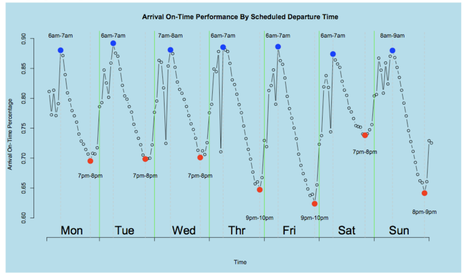
[Figure 4] - Month V.S. Arrival On-Time Performance



Note :

- The blue line show the trend of arrival on-time performance according to different month. There exists a monthly effect on arrival on-time performance in every year. So we just used the 2008's data set as the proxy of other year to make the plot.
- **Months Which Have Bad Arrival On-Time Performance :**
  June : Summer Vacation (Popular Travel Season)
  December : Winter Vacation (Popular Travel Season) + Terrible Weather In Winter
  January - March : Terrible Weather In Winter
- **Higher cancelled flights % than any other months**
  December - March : Terrible Weather In Winter

[Figure 5] - Day of Week+Departure Time V.S. Arrival On-Time Performance

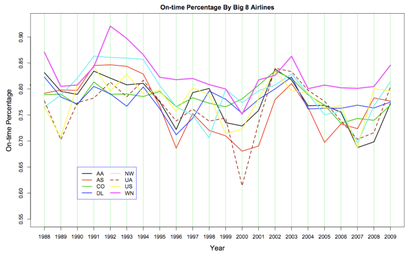


[2008's Data Set As The Proxy of Other Year.]

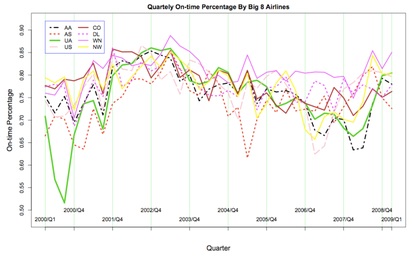
Note :

- [Earlier Flights Really Guarantee Better Arrival On-Time Performance ?]
  - Workers/crews who served in a airport/airline maybe still felt asleep before 6AM
- The blue/red points show the best/worst arrival on-time performance according to different day of week and departure time.
- There exists a significant downward trend from blue point to red point. From 7AM to 9AM, rush hour made several arriving-late flights in some certain airports first, and a butterfly effect causes the subsequent flights after 9AM with worse arrival on-time performance.
- Better arrival on-time performance usually occurred from Monday to Wednesday.
- Worse/The worst arrival on-time performance occurred on Thursday/Friday. Maybe more/most travelers started their trip (for visiting family/friends/or fun) on Thursday/Friday. Higher volume on these two days caused worse/the worst butterfly effect. Vice versa, same situation happened on Sunday after their trips.
- On Saturday, the butterfly effect was slightest than any other day. Maybe most travelers preferred to stay with their family on this precious day.

[Figure 6] - Yearly Arrival On-Time Performance By 8 Airlines



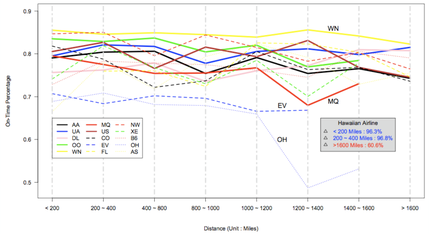
[Figure 7] - Quarterly Arrival On-Time Performance By 8 Airlines



Note :

- Either from Figure 7 or Figure 8, we can see Southwest Airline(WN) performed very well in its arrival on-time performance. One reason is from the strategy on selection of their departure and arrival airports. They prefer to choose secondary airports (lower volume) instead of primary airports to reduce the chance to have a arriving-late flight.
- From the Figure 7, after the fourth quarter of 2007(Financial Crisis), due to consistently decreasing scheduled departure flight, all airlines apparently can provide better arrival on-time performance. Economy's condition is also an important influential factors on arrival on-time performance.
- American Airline (AA) and United Airline (UA) may be not a best choice if you still have other selections on choosing a flight.

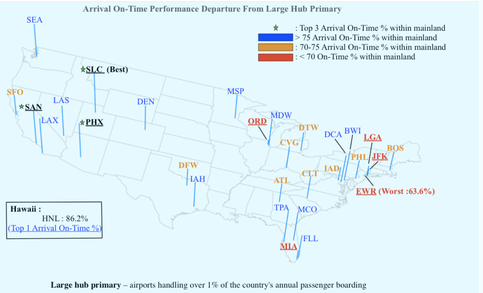
[Figure 8] - Distance V.S. Arrival On-Time Performance By 16 Airlines
On-Time Percentage By Distance Group For 16 Airline In 2009



Note :

- Southwest Airline(WN) performed very well within all distance groups.
- Comair(OH) / Atlantic Southeast(EV) / American Eagle(MQ) /Expressjet(XE) apparently had bad arrival on-time performance within all distance groups.
- Correlation coefficient between Distance and Arrival On-Time percentage is "-0.17 ", which means that "Long-distance flights may have a little bit higher probability to be a arriving-late flight.
- Hawaiian Airline had a apparently worse arrival on-time performance on the distance group(>1600 miles). Because the arrival-delayed flights departured from Newark(EWR) Airport(very bad arrival on-time %).

[Figure 9] - 2008 | Arrival On-Time Performance
Departure At Large Hubs

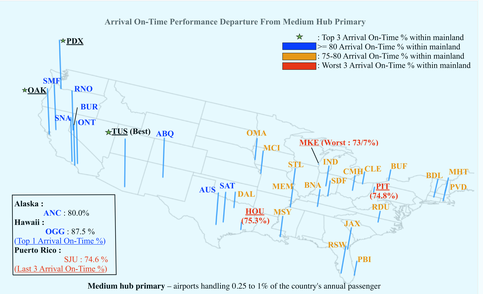


Large hub primary – airports handling over 1% of the country's annual passenger boarding

Note :

- The busiest air route in the U.S. must be around New York area. There are 3 large hubs which locate around New York. Even the total scheduled departure flights within this area is 376,445 flights (LGA is 119,135 flights / in EWR is 138,506 / in JFK is 118,804 flights), which is lower than ATL (414,513 flights). However, especially in winter, the weather in the Northern U.S. is always worse than the weather in the Southern U.S. Weather in winter may be a reasonable reason to explain the bad arrival on-time performance around New York.
- In Chicago, Midway Airport(MDW) performed better in arrival on-time performance than O'Hare Airport(ORD). 2008 scheduled departure flights in : MDW is 87,619 flights / in ORD is 350380 flights. Bad weather in Chicago may be a reason of the arrival delayed flights. However, heavy traffic may be an more important influential factor on arrival on-time performance.
- In Miami Airport(MIA), despite 2008 scheduled departure flights in MIA(62,559) is a little bit smaller than than FLL(67,177), but FLL had better arrival on-time performance(75.6%) than MIA(68.4%). The fact indicted that there must be some other influential factors(ex: service, management , etc.) exist in Miami Airport.

[Figure 10] - 2008 | Arrival On-Time Performance
Departure At Medium Hubs



Medium hub primary – airports handling 0.25 to 1% of the country's annual passenger

Note :

- From Figure 6 and Figure 7, we can see a significant difference. Due to lower volume in an airport, medium hubs had much better arrival on-time performance than large hubs. Even the worst 3 airports had arrival on-time percentage greater than 70%.
- Interesting Observation- For the medium hubs, those airports which locate on the western U.S. performed better than those airports which locate on the eastern U.S.

Conclusions And Discussions

Advanced research is still on-going. We anticipate that these results can give you a general idea on how to book your flight in a smarter way. Time's factors, Location's factors, and Airline's factors all are very relative to the arrival on-time performance. However, there might be still some confounding effects which exist among these factors. On the future study, we will focus on : **1.** Discussing the Route's factor, we will try to find out which routes perform worse in "Arrival on-time performance" and clarifying the reasons for the arriving-late flights (i.e. these delayed flights were caused by their departure airports or by their arrival airports or both); **2.** Discussing the airlines with worse or better arrival on-time performance, and clarifying the real reasons which affect their arrival on-time performance. By discussing the causes of arriving-late flights in a hierarchical way, we can provide an interested traveler the information on how to book a flight in a smarter way.

Acknowledgements

Thanks for the Joe (Laboratory assistance in Department. of Statistics.) for providing me the knowledge about Unix and B.Y Sun (Ph.D in Computer Science) for assisting me in how to develop a useful database with SQLite.