

Project Title:

Airport Performance Analysis

Team Member:

Chen, Yibo	yc3224@columbia.edu
Hu, Yuxin	yh2874@columbia.edu
Qi, Zeyu	zq2144@columbia.edu
Su. Xianzheng	xs2285@columbia.edu

Abstract

Ever since Wright brother invented the first airplane in 1903, aviation is one of the major methods of transport for the public. In March 2016, Port Authority announced a \$6 billion plan to rebuild the terminals in LaGuardia and Newark, which was aimed to adapt to the increasing number of airplane passengers. In June 2017, New York's Stewart Airport will begin to offer transatlantic flights to Ireland, Scotland and other European countries. Reporters claimed that the new airway will offer New Yorkers more economic and flexible flights options. However, how much benefits (or detriments) could the new airports and airways brought to the region and vicinity are not clear. We did not find many statistical analysis or prediction on the aviation traffic and airport locations from existing literatures. The purpose of this project is to understand the relationship between air terminals (such as location, size, and aircraft type) and its economic contributions (such as passenger flows and profits). We plan to answer the following questions in this project:

- What factors determine the passenger flows of an airport?
- How does the density of airports in a geographic location relate to the population of residents in the region?
- What are the key determinants of airport profits?
- Does the distance from the geographic (economic) hub matter for an airport?
- What is the relationship between enplanement (passenger boarding) and conditions of airport (size, location, year built etc)

In terms of the instruments and methods of analysis, we are going to use ggplot to display some summary information about airports and airways in the United States. Then, we will mark airports in a map using their geographic coordinates. A text analysis might be used to examine some features of these airports. Last, since flights connects airports in different cities, we may draw a network analysis to find clusters and associations.

This preliminary plan only exposes some degrees of analysis; we look forward to finding more interesting patterns and relationships after visualizing the dataset.

Data

Our data mainly focus on three areas: airport, flight and customer.

Airport: the basic information about airports such as location, number of terminals and gates.

Flight: date, time, airplane type, on-time rate.

Custom: number of customers by airports, customer review on the airport.

So far, we have collected the airport data. We are trying to determine the time period for flight information data and the sources of customer review. Besides, we plan to collect data on city locations and total traveler amounts. Thus, we can tell how many cities the airport covers, the distance between airport and city center, and the popularity of cities.

Obstacles

We decided to focus on domestic flights in USA. However, it is hard to distinguish international and domestic travelers. Also, the time schedule of domestic flight might be influenced by international flight. This might lead to a bias on our analysis.

Dataset Website:

<http://ourairports.com/data/>

<https://www.transtats.bts.gov/ONTIME/>

<http://openflights.org/data.html>

https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/passenger/

Possible Work Cited:

<http://www.nbcnewyork.com/news/local/Stewart-International-Airport-New-York-Norwegian-Transatlantic-Flights-Ireland-Scotland--414604033.html>

<http://www.lohud.com/story/money/business/2017/02/23/stewart-international-airport-ireland-scotland/98294176/>

http://www.nj.com/traffic/index.ssf/2016/03/ga_port_authoritys_6b_plan_for_new_airport_terminals.html