

Best airline of the month

CSE 140 Final Project

Yepingzhi Kuang

Research questions:

1. Which airline has the lowest delay percentage?
I will compute the number of total flights of each airline and the number of delayed flights to get percentages. Results will be a 10-element-list of ranks from lowest percentage to highest.
2. Which airline has the lowest delay time?
I will compute the total delay time of each airline and report a 10-element-rank from lowest to highest.
3. Which airline has the largest number of air routes?
I will compute number of routes of each airline and report a 10-element-rank from highest to lowest.
4. Which airline has the longest total traveled distance?
I will compute the total distances traveled of each airline and report a 10-element-rank from highest to lowest.
5. Which airline is the best based on the information above?
I will compute the average rank of each airline based on four ranks above and report a 10-element-rank from lowest to highest.

Motivation and background:

Airplane has become one of the most important transportations now which is closely related to people's life. Although people are benefit from its convenience and rapidness, the delay of flights is really one of the most common issues which make people disappoint with air transportation. It is interesting to know how much time was wasted on flight delays, how many flights are delayed among all the flights and other information about different airlines. After we have all these information, we can have an overall impression of an airline which can help us to choose the right airline company.

Datasets:

A .csv file contains: Date of the flights, Flight carriers, Origin City names, Destination city names, Scheduled Arrival time, Actual arrival time, Scheduled departure time, Actual arrival

time (times are showed in military time format), delay checking (0 for on-time flight, 1 for delayed flight), distance traveled for every flight. Data set only contains information of one month.

Methodology:

1. Things to be clarified: There are two parts in a single flight that should be considered, the arriving and departure. If one of these two is delayed, this flight is considered as a delayed flight. Each part has three situations that should be considered differently: delayed, on time and ahead of time. If one part of the flight is delayed, the time gap between the scheduled time and actual time should be greater than zero. If one part of the flight is ahead of time, the time gap between the scheduled time and actual time should be less than zero. If it is on time, then the gap is zero of course. The total delayed time of a flight should equal to the sum of the time gaps of two parts. It can be positive, negative or zero. Delayed check information which shows whether a part is delayed or not is provided in the data set (0 for not delayed, 1 for delayed). The time gap is always less than 24 hours. Unfortunately there are some broken data in the columns that show time and delayed checks that have to be take care of. If the numbers in the time columns are less than 3 digits or greater 4 digits, this will be considered as a broken data and the whole flight should be skipped. If the delayed checks are not provided in a flight, the flight should also be considered as a broken data and be skipped.
2. For question 1, first count the total number of flights of each airline and the number of delayed flights among the total flights. Then, divide the number of delayed flights by the total number of flights to get percentages. To report the results, sort the list of results in ascending order and only report first 10 of them.
3. Question 2 is a little bit challenging. If the time from the data set is three digits, the first digit should be hour and last two should be minute. If it is four digits, the first two should be hour and last two should be minute. The actual time could be either later than or earlier than the scheduled time, and these two situations should be considered differently depends on whether the flight is delayed. Time gap should be calculated in minute.

Examples:

If delayed:

Sch: 120, Act: 130 → gap:10 min

Sch: 120, Act: 115 → gap:1435 min

If no delayed:

Sch: 120, Act: 110 → gap: -10 min

Sch: 230, Act: 320 → gap: -1390min

Add up all the time gap of flights of each airline and report 10 airlines that have lowest

total delaying time in ascending order.

4. In question 3, a route should be considered as a pair of two different cities. For example, flight from Seattle to Vancouver should be considered as a route. But the flight from Vancouver to Seattle should be considered as a different flight. Count the total number of flights of each airline and report 10 airlines that have largest number of routes in descending order.
5. For question 4, add up all the distance traveled of each airline and report 10 airlines that have largest traveled distance in descending order. Distances of same routes should be also added up.
6. For question 5, gather up all four of the rank information and calculate the average rank of each airline and report 10 best airlines in ascending order. For example, if airline DL has ranks: 1, 3, 4, 2, the average rank for DL should be 2.5. If airline US has ranks: 2, 2, 2, 1, the average rank for US should be 1.75 and it is doing better than DL.

Results:

1.

Top 10 airlines of lowest delay percentage

No.1 Airline: VX Delay Percentage: 31.76%
No.2 Airline: DL Delay Percentage: 36.23%
No.3 Airline: 9E Delay Percentage: 38.27%
No.4 Airline: HA Delay Percentage: 38.34%
No.5 Airline: AS Delay Percentage: 40.06%
No.6 Airline: YV Delay Percentage: 40.31%
No.7 Airline: US Delay Percentage: 44.63%
No.8 Airline: MQ Delay Percentage: 45.86%
No.9 Airline: OO Delay Percentage: 46.86%
No.10 Airline: B6 Delay Percentage: 49.2%

2.

Top 10 airlines of lowest delay time (in minutes)

No.1 Airline: VX Delay Time: 332642
No.2 Airline: HA Delay Time: 335444
No.3 Airline: F9 Delay Time: 526674
No.4 Airline: YV Delay Time: 661379
No.5 Airline: AS Delay Time: 764882
No.6 Airline: 9E Delay Time: 1369789
No.7 Airline: B6 Delay Time: 1622532
No.8 Airline: US Delay Time: 2365748
No.9 Airline: UA Delay Time: 2832513

No.10 Airline: OO Delay Time: 3350313

3.

Top 10 airlines of longest total traveled distances (in miles)

No.1 Airline: WN Distance Traveled: 58546415
No.2 Airline: UA Distance Traveled: 49302836
No.3 Airline: DL Distance Traveled: 48801468
No.4 Airline: AA Distance Traveled: 45190062
No.5 Airline: US Distance Traveled: 28441957
No.6 Airline: EV Distance Traveled: 25912688
No.7 Airline: OO Distance Traveled: 22428272
No.8 Airline: B6 Distance Traveled: 19521507
No.9 Airline: MQ Distance Traveled: 18057489
No.10 Airline: AS Distance Traveled: 14167745

4.

Top 10 airlines of largest number of flight routes

No.1 Airline: EV Number of routes: 1055
No.2 Airline: WN Number of routes: 1008
No.3 Airline: OO Number of routes: 810
No.4 Airline: DL Number of routes: 608
No.5 Airline: 9E Number of routes: 534
No.6 Airline: UA Number of routes: 502
No.7 Airline: MQ Number of routes: 374
No.8 Airline: AA Number of routes: 327
No.9 Airline: US Number of routes: 292
No.10 Airline: B6 Number of routes: 245

5.

Top 10 Best Airlines of the Month

No.1 Airline: DL Comprehensive ranking: 5.0
No.2 Airline: 9E Comprehensive ranking: 6.25
No.3 Airline: VX Comprehensive ranking: 7.0
No.4 Airline: OO Comprehensive ranking: 7.25
No.5 Airline: US Comprehensive ranking: 7.25
No.6 Airline: UA Comprehensive ranking: 7.5
No.7 Airline: EV Comprehensive ranking: 7.75
No.8 Airline: AS Comprehensive ranking: 8.0
No.9 Airline: WN Comprehensive ranking: 8.0
No.10 Airline: B6 Comprehensive ranking: 8.75

Reproducing results:

Simply put the Flight info.csv file and the Airline analysis.py file in the same directory and run the program. Results will be printed out in the console.

Collaboration: None

Reflection: I learned where to find data and statistics on line. I didn't know that I could find so many data on line in the past. I want to know more about graphing class in python, because I will use this class a lot in my project.