---

Title: "DSC520 final"

author: "xin tang"

date: "`r Sys.Date()`"

output: html\_document

output:

pdf\_document: default

bibliography: flight\_citations.bib

editor\_options:

markdown:

wrap: 72

**My home airport is Austin, Texas (AUS), due to work, I fly to bay area in California frequently. I can choose from 3 airprots to use: San Francisco airport (SFO), SanJose airport (SJC) and Oakland airport(OJC). I am interested to find how to pick the flight out and back flight to avoid delay, with least cancellation risk and possible best fare.I only studied direct flight since it will be most time saving options.**

**First I will do an analysis on delay using flight data.**

A screenshot of a computer

Description automatically generated

A graph of a graph showing the airport and delay count

Description automatically generated with medium confidence

### Now analysis delay condition from those airports

A screenshot of a computer

Description automatically generated

### From this analysis, no matter as Origin or Destination airport, OAK airport have twice as much delay as other 2 airports. so it is not a good choice to fly into

### The rest analysis will not include OAK, since it is not a preferred place to fly to/from.

### Now look at the delay from other 2 aspects( by daily and by airline).

### looks like Friday may have more chances to delay but every Airline have similar chance of delay

A graph with colorful dots

Description automatically generated

A graph with a number of points

Description automatically generated with medium confidence

### Excluding early arrival, now answer the question: who may have least delay.

#### **Airline OO (skywest) has least delay and rest are about same when fly into Austin.**

#### **Airline UA (United) has worst delay when fly into bay area.**

#### **Fly into and from SJC is always a good choice.**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

### Check which delay factor is the most significant when fly to bay area

### it shows the prior flight delay and control tower caused most delay, which I can not control

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

### Get histogram of delay, inbound and outbound each, check if there are normal distribution.

### A graph of a flight delay Description automatically generated

### A graph of a time line Description automatically generated

### Now check correlation between flight time and different delay factors. using flight data 'Austin to SFO'

**Again the air control caused delay is more important factor**

### A number of numbers and a carrier Description automatically generated with medium confidence

### Now I want to look at cancellation data. Data is from Jan-2022 to April-2023

#### **first check monthly cancel rate, then check airline performance.**

looks like month Jan'2022, Feb'2022 and Dec'2022 are outliners (we know it from News) Airline JetBlue, Southwest, Spirit and United have some bad records.

### A graph of different colored dots Description automatically generated

### A screenshot of a computer Description automatically generated

### A graph with different colored dots Description automatically generated

## Next remove outliner month and try to rank the best and worst airline in term of cancellation.

### Now looks like Hawaiian is the best, but it do not fly to Austin. the United airline looks better now, ranked #4.

A screenshot of a phone

Description automatically generated

### Last, review the fare data. which contains fare info in whole year 2022

#### **looks like in Q4 ticket is the most expensive, while Q1 is the cheapest**

A graph with numbers and points

Description automatically generated

## Conclusion

### from Austin to Bay area and back. looks like taking skywest (now is called United Express) is likely had minimum delay and average risk of cancellation. the best airport to use is San Jose airport (SJC). Flying in Q1 is most likely to get cheap ticket.

# References

@Flight Dataset @bureau\_of\_transportation\_statistics\_quarterly\_nodate @bureau\_of\_transportation\_statistics\_airline\_nodate