TEAM 7 – AIRLINE DATASET

JANET HUANG

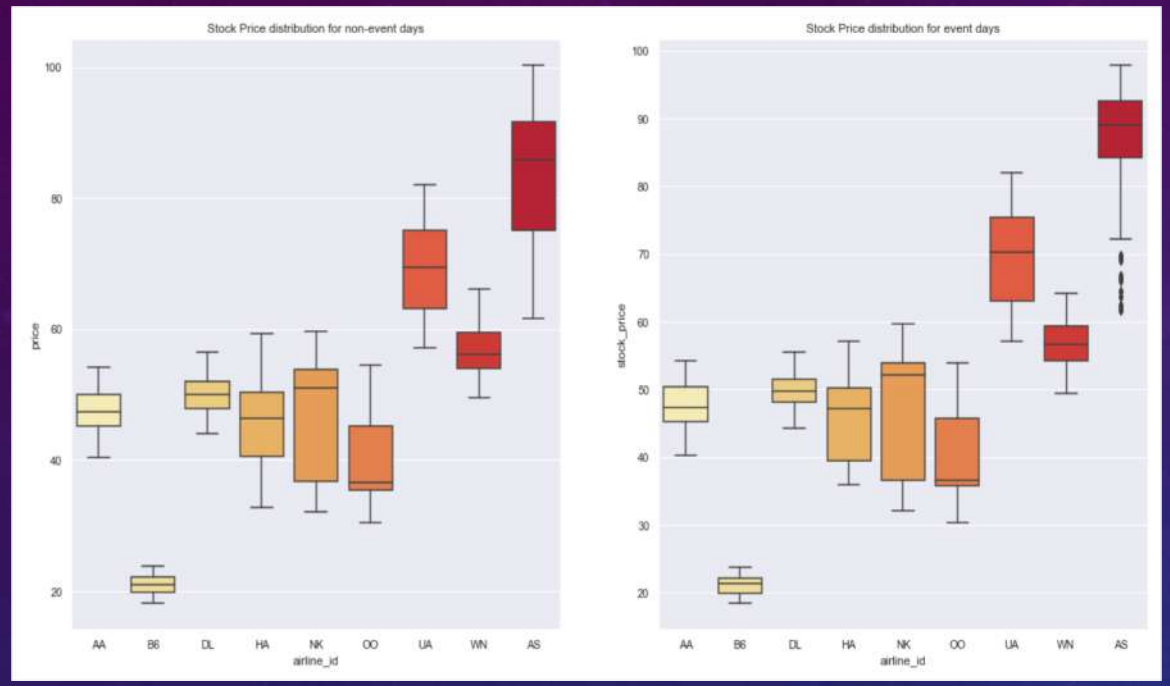
NEHA SARAF

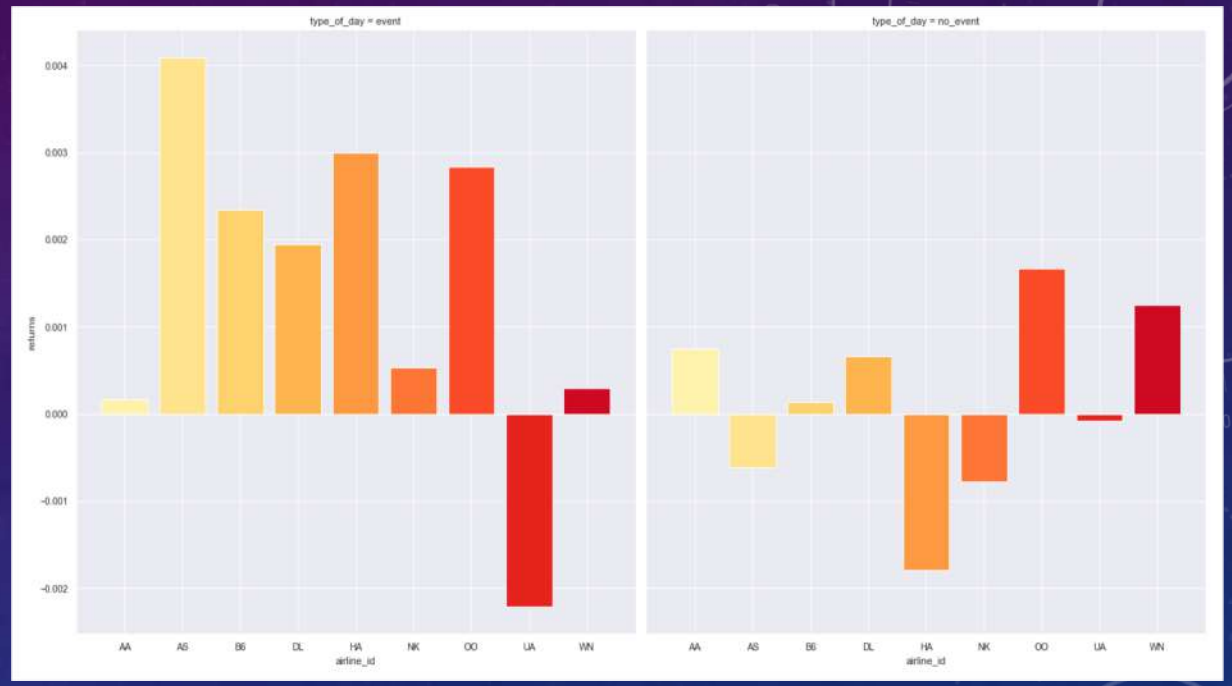
SHIQI DUAN

PROJECT OUTLINE

- **Problem Statement**: We wish to assess the viability of a short-term trading strategy for airline stocks based on knowledge of upcoming events and weather conditions.
- **Key Questions**:
  - Are cancellations/delays that occur on event days more likely to depress stock price/returns than cancellations/delays that occur on non-event days?
  - Do weather conditions at scheduled departure influence the probability of having cancellation/delays and/or the severity of delays?
  - Are certain companies better/worse at navigating bad weather on event days i.e. do they cancel less/more flights than their peers?

STOCK PRICE / RETURN DISTRIBUTION ON EVENT AND NON-EVENT DAYS





The mean stock price distribution remains almost the same for every flight on event/non-event days. But the mean stock return is different for every flight when seen for event vs non-event days.
The return rate for Jetblue, Delta, Hawaiian, Spirit, Alaska and Skywest airlines increases when there is an event but for American, United and Southwest airlines it decreases.

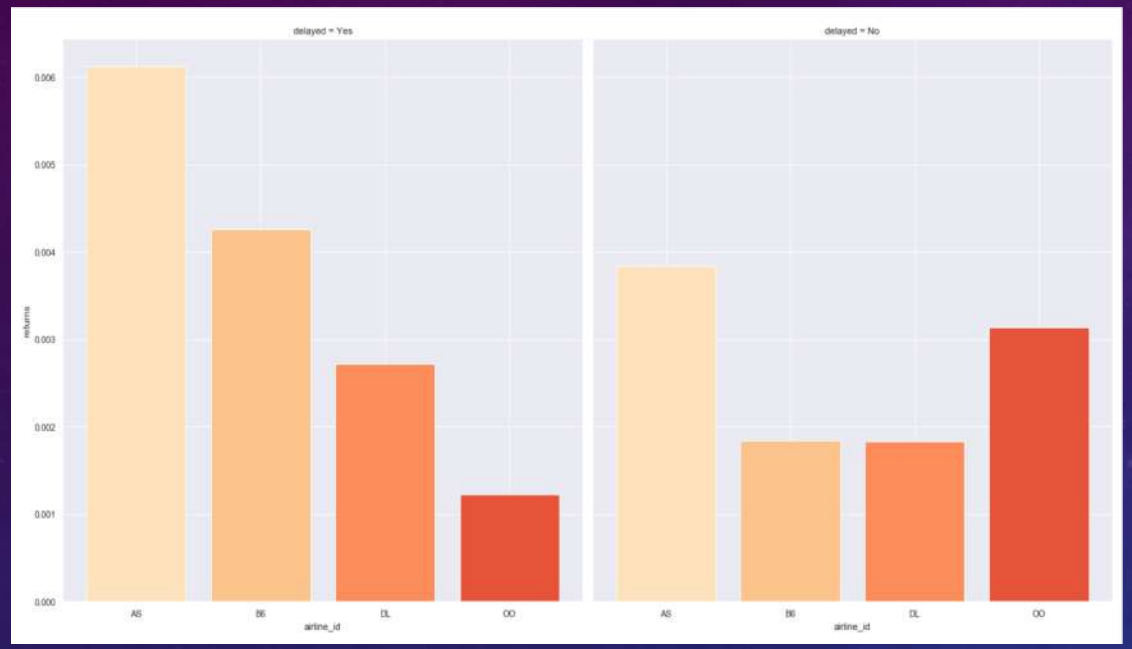
QUESTION : Is the mean stock return of the event and non-event days groups significantly different?

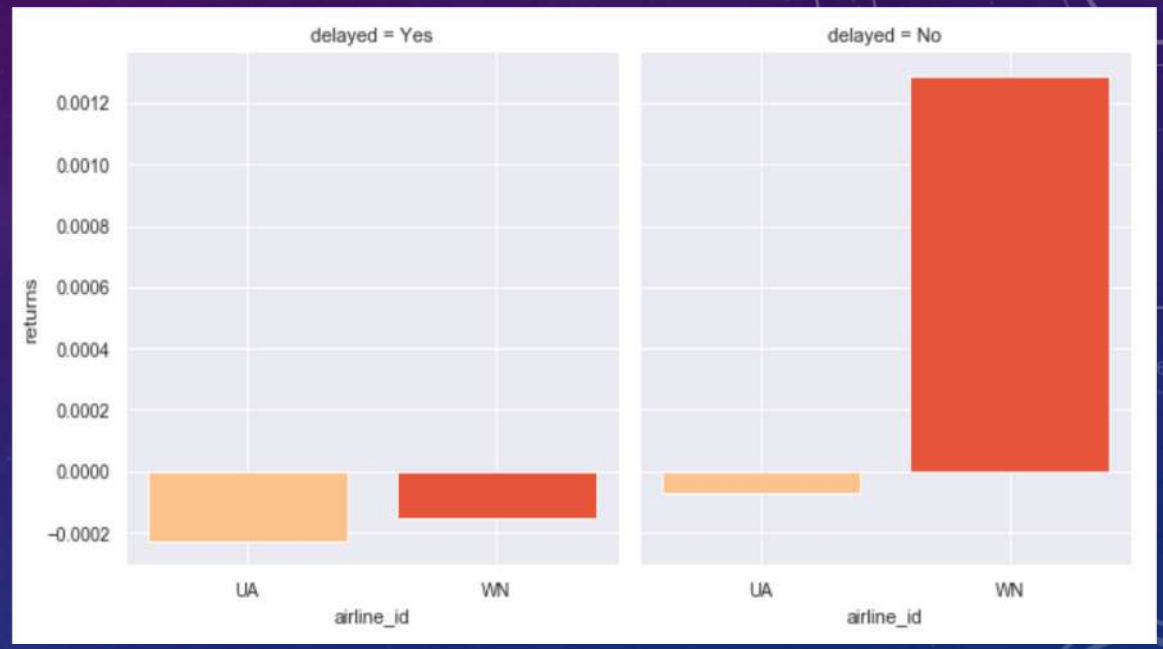
Methodology used: Hypothesis testing

Airlines	P-value	Statistically different?
JetBlue (B6)	5.987237629108067e-08	Yes
Delta (DL)	5.9519231968561 52e-07	Yes
Hawaiian (HA)	0.22609560431867037	No
Spirit (NK)	0.12588633046689335	No
Alaska (AS)	2.1663681359397 07e-06	Yes
American (AA)	0.09424409127591074	No
Skywest (OO)	0.025225205181498924	Yes
United (UA)	0.00073677292258 62906	Yes
Southwest (WN)	2.3297185332180505e-06	Yes

The mean stock return is statistically different on event days and non-event days for JetBlue, Delta, Alaska, Skywest, United and Southwest airlines. It is recommended for investors to invest in JetBlue, Delta, Alaska, Skywest airlines when there are events going on and to invest in United/Southwest airlines on non-event days.

Effect of delays/no-delays on stock return for an event/non-event days





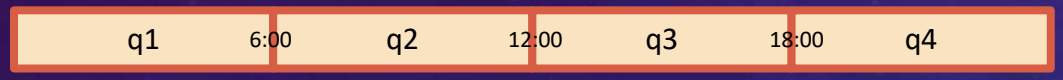
From the visualization, on event-days with delays the stock return is higher as compared to no delays. And on the non-event days, the stock return is lower when the airlines have delays. To verify this difference in observations, is only due to sampling variability, we conducted the same series of test and found that all the null-hypothesis were accepted and hence there is no effect of delays on the mean stock return for an event/non-event days

Analyzing the Relationship Between Weather and Flight Cancellations/Delays

METHODOLOGY

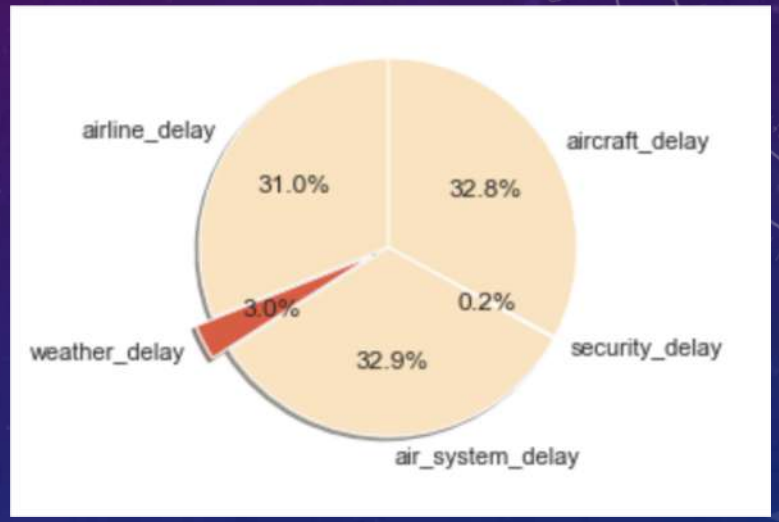
Joining Weather Data with Flight Traffic Data

- Divided the weather measurements into 4 buckets (q1-q4) based on time of day – did the same for flight traffic scheduled departure times



- Matched the origin airport, date, and q1-q4 bucket of the each flight entry with the airport id, date, and q1-q4 bucket of the weather measurements

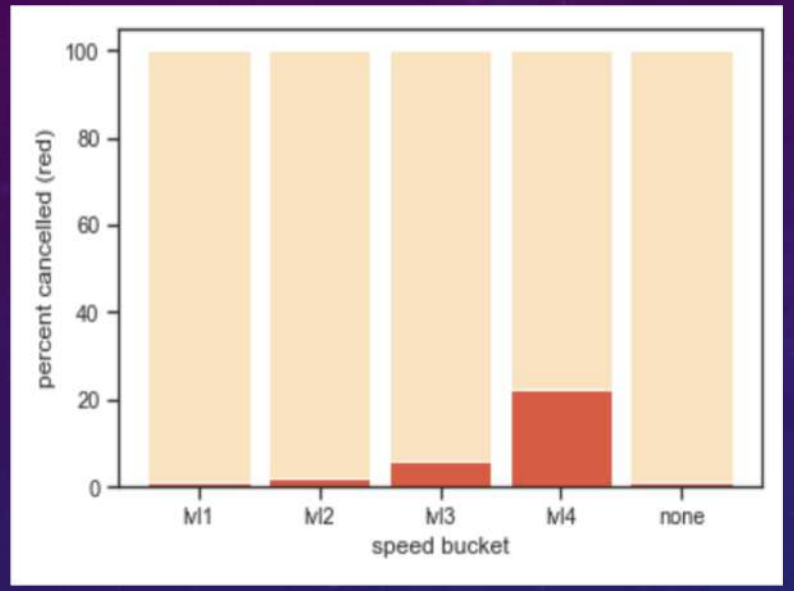
Determining How to Attribute Delays to Weather



- BTS only labels a delay as a weather delay (~3% of delays) under extreme weather conditions
- will examine the relationship between weather and overall delay (sum of all delays)

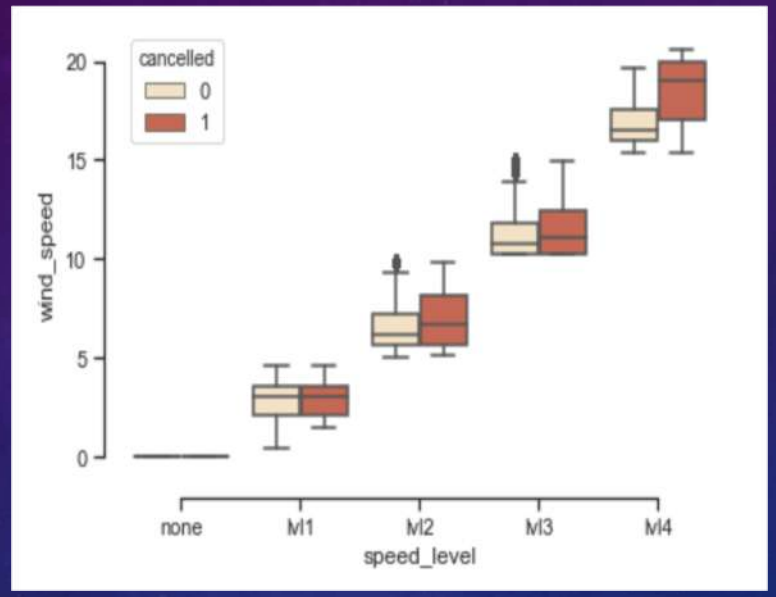
Wind Speed Analysis Highlights

prep steps: removed a series of outliers with wind speed = 999.9mps & broke down wind speed measurements into 4 levels: lvl1: 0-5mps, lvl2: 5-10mps, lvl3: 10-15mps, lvl4: 15+ mps



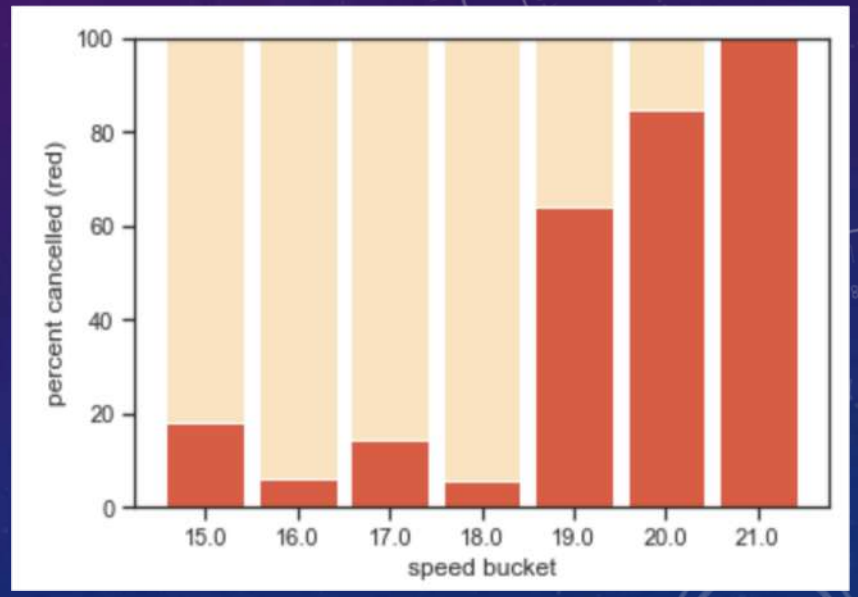
percent of flights cancelled for each speed bucket

notable spike in cancellation rates for flights facing lvl4 winds



wind speed distribution for cancelled vs not cancelled

notable disparity in cancelled vs non-cancelled for lvl4 speeds



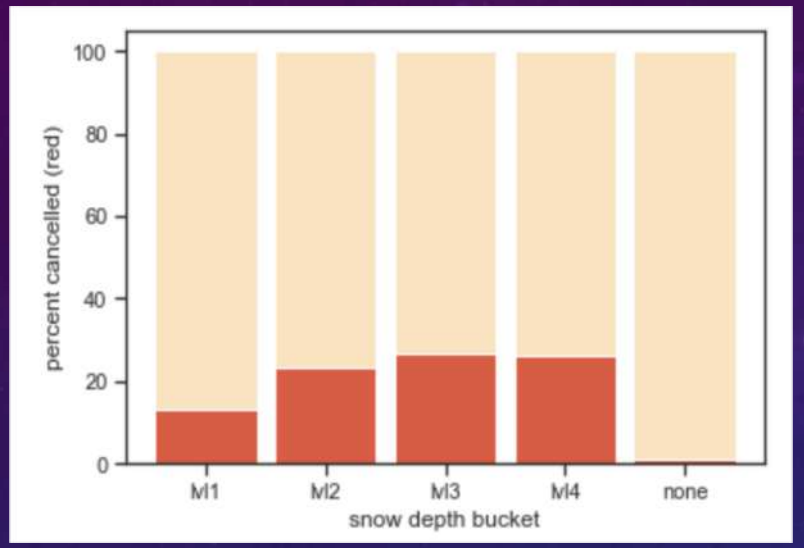
percent of flights cancelled for each speed bucket in lvl4

cancellation rate suddenly jumps once wind speed reaches 19mps

KEY INSIGHT: if wind speed >= 19 mps, flight will most likely be cancelled

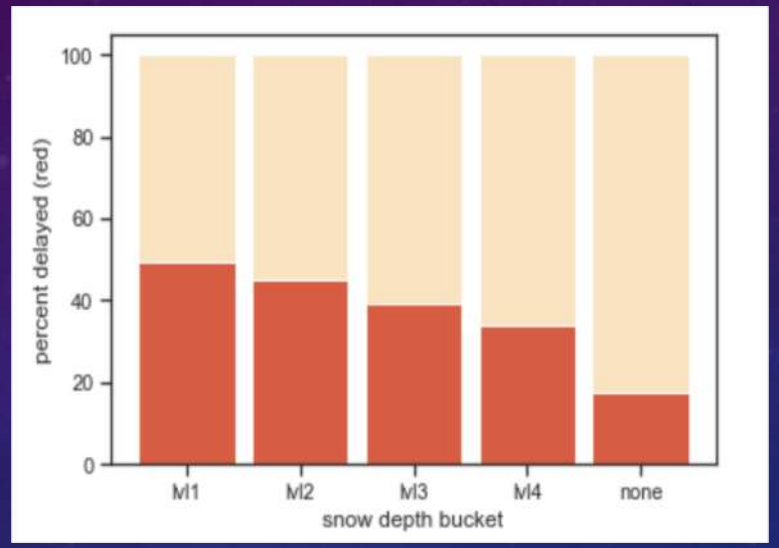
Snow Depth Analysis Highlights

prep steps: broke down snow depth measurements into 4 levels: lvl1: 0-4cm, lvl2: 4-8cm, lvl3: 8-12cm, lvl4: 12+ cm



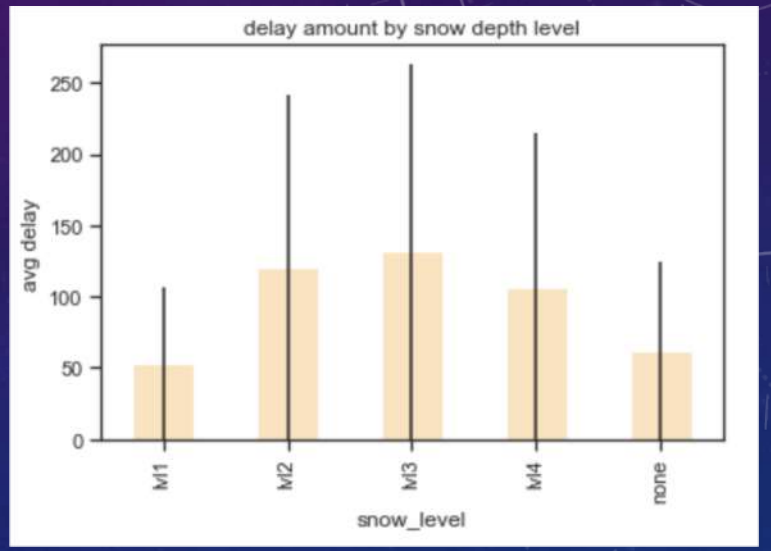
percent of flights cancelled for each snow depth bucket

cancellation rate significantly higher for snow vs. no-snow departures



percent of flights delayed for each snow depth bucket

delays are most likely to occur when snow just begins to fall



average delay (min) by snow depth bucket

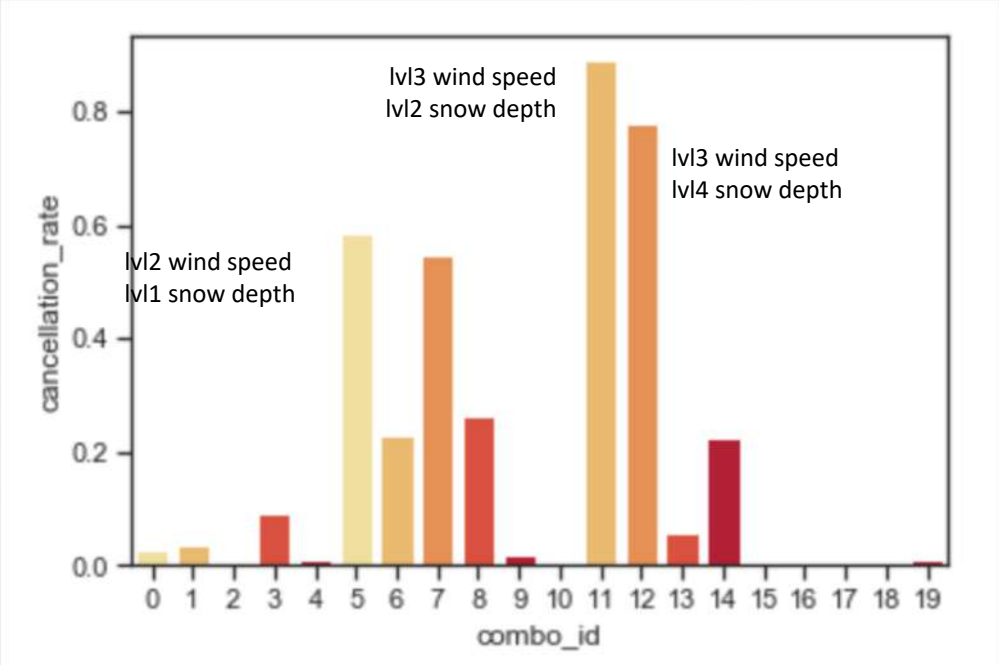
delay will generally increase as snow intensifies through lvl3

KEY INSIGHT: any snow at time of departure is correlated with higher delay and cancellation rate

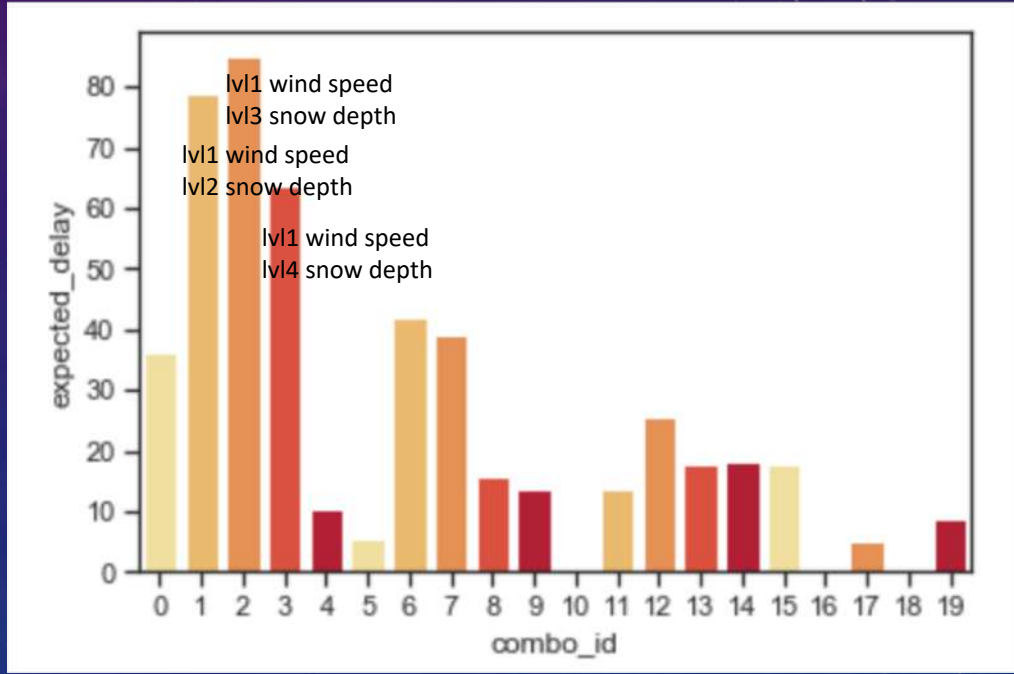
Wind Speed x Snow Depth Highlights

prep steps: assigned a unique id ranging from 0-19 for every possible combination of wind speed x snow depth levels

Cancellation rate of flights within each weather combo

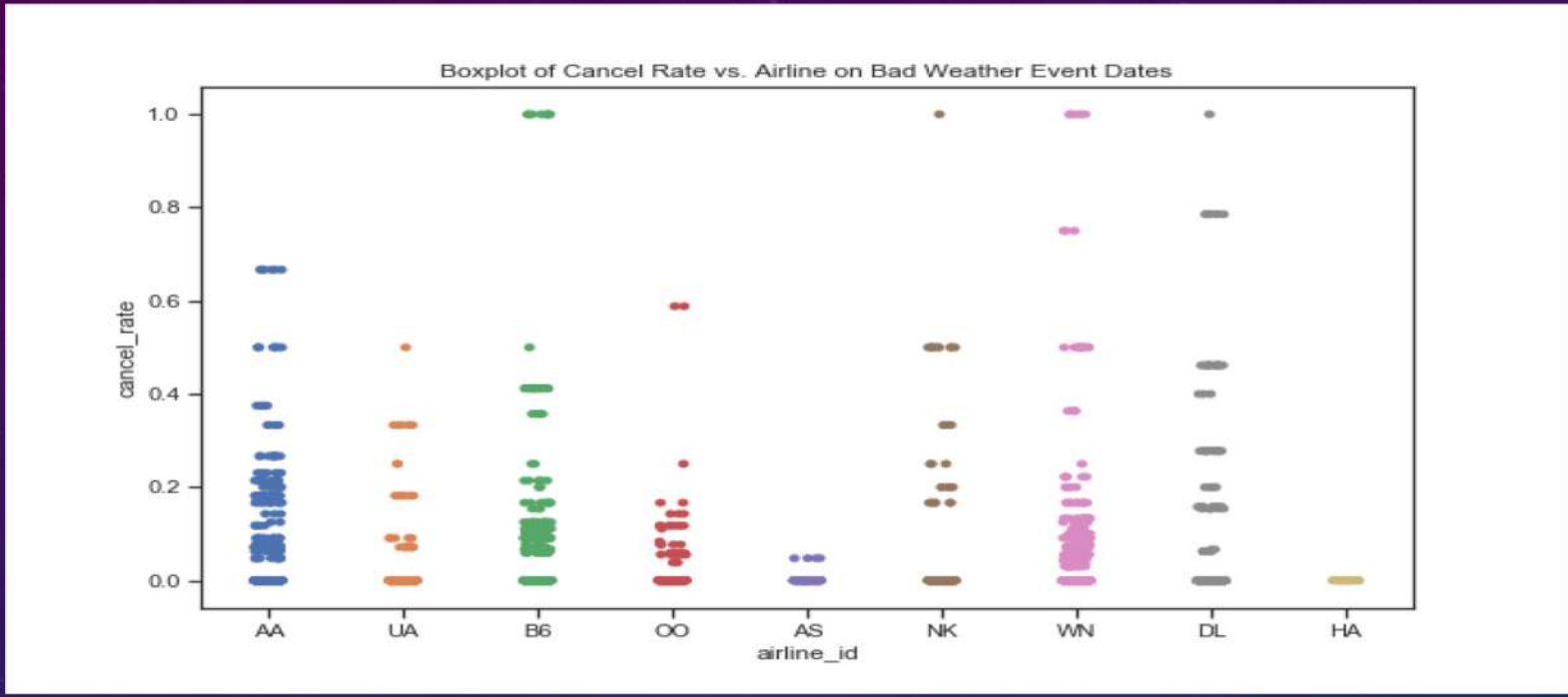


Expected delay of flights within each weather combo



CONCLUSION: while only EDA was used to examine the relationship between weather conditions and flight delays/cancellations, the drastic variation in cancellation rates and expected delays across various weather condition pairings strongly indicate that weather can help predict flight disturbances

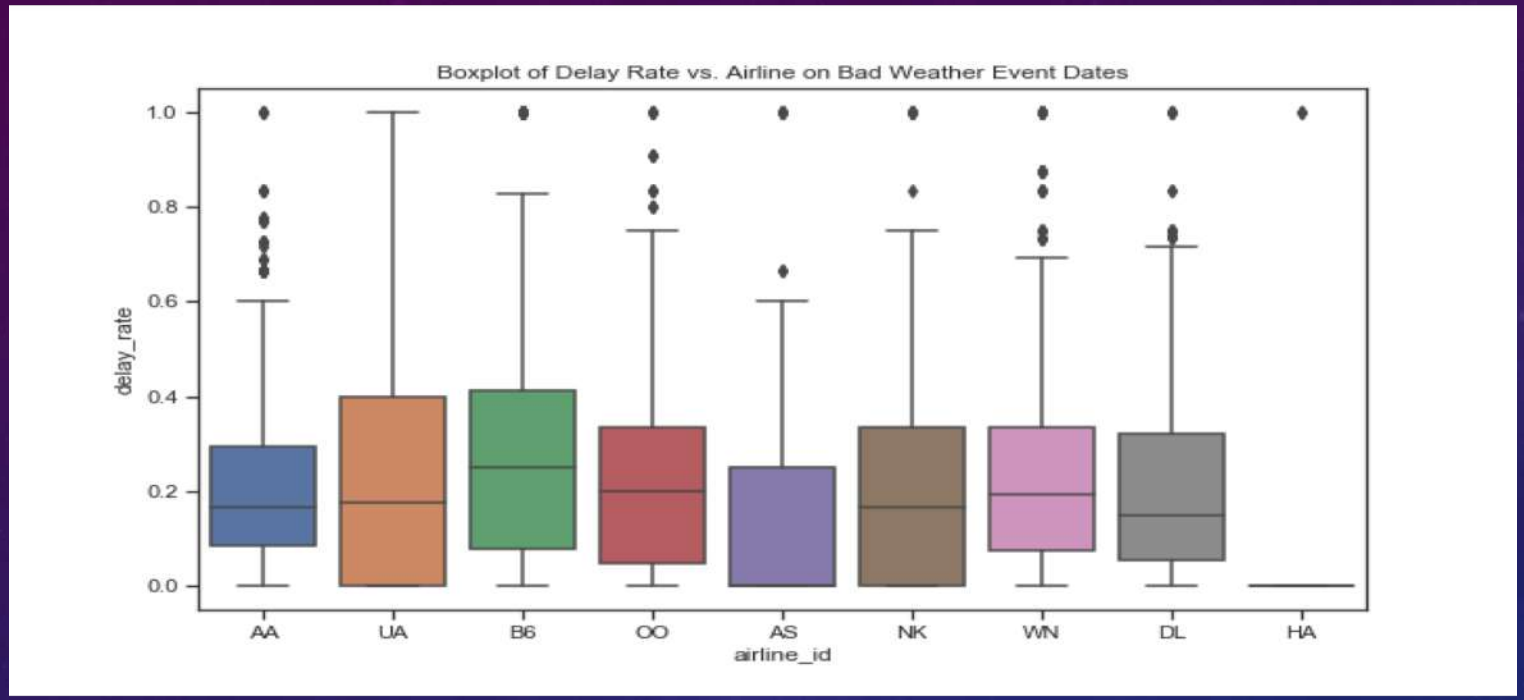
Airline Navigation Performance on Bad Weather Event Days



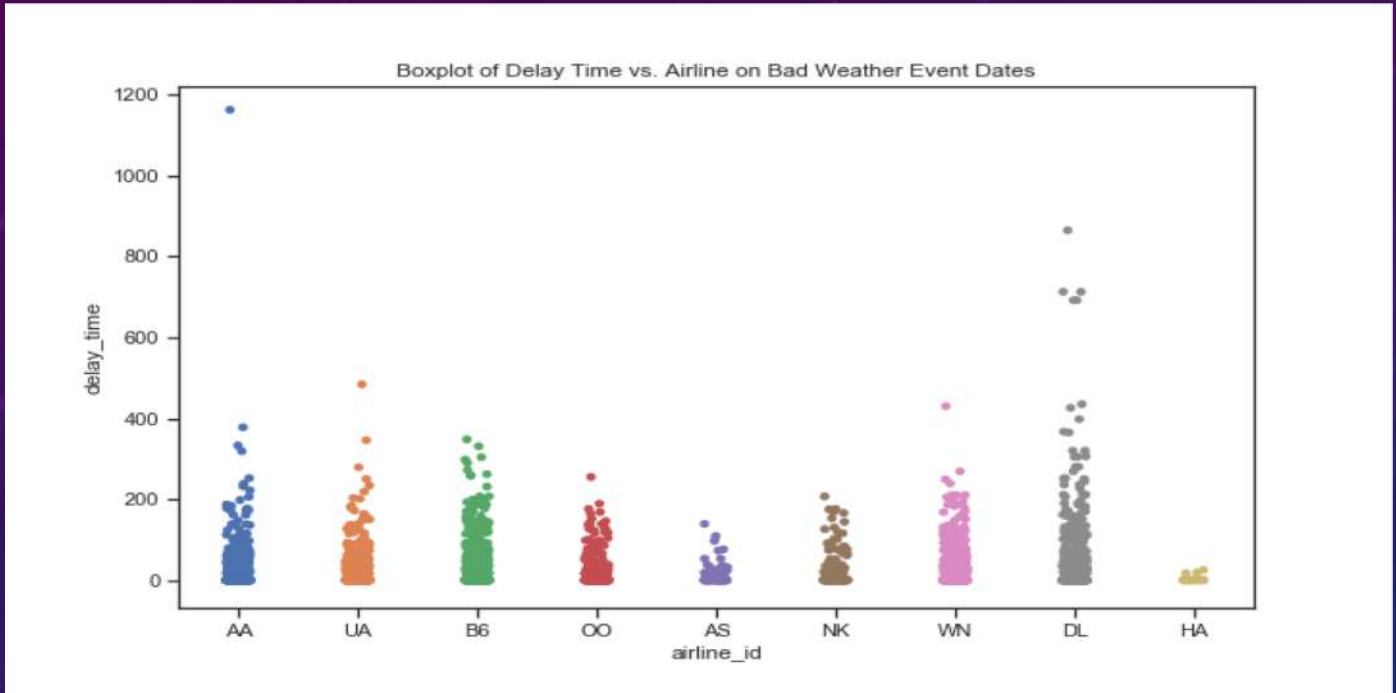
In terms of the Cancellation Rate, AS and HA have lower cancellation rates compared to other airlines. However, the total flights of these two airlines are less than others.

AA, UA, and OO perform not bad as they do not have extreme high cancellation rates and their flight amounts are relatively large.

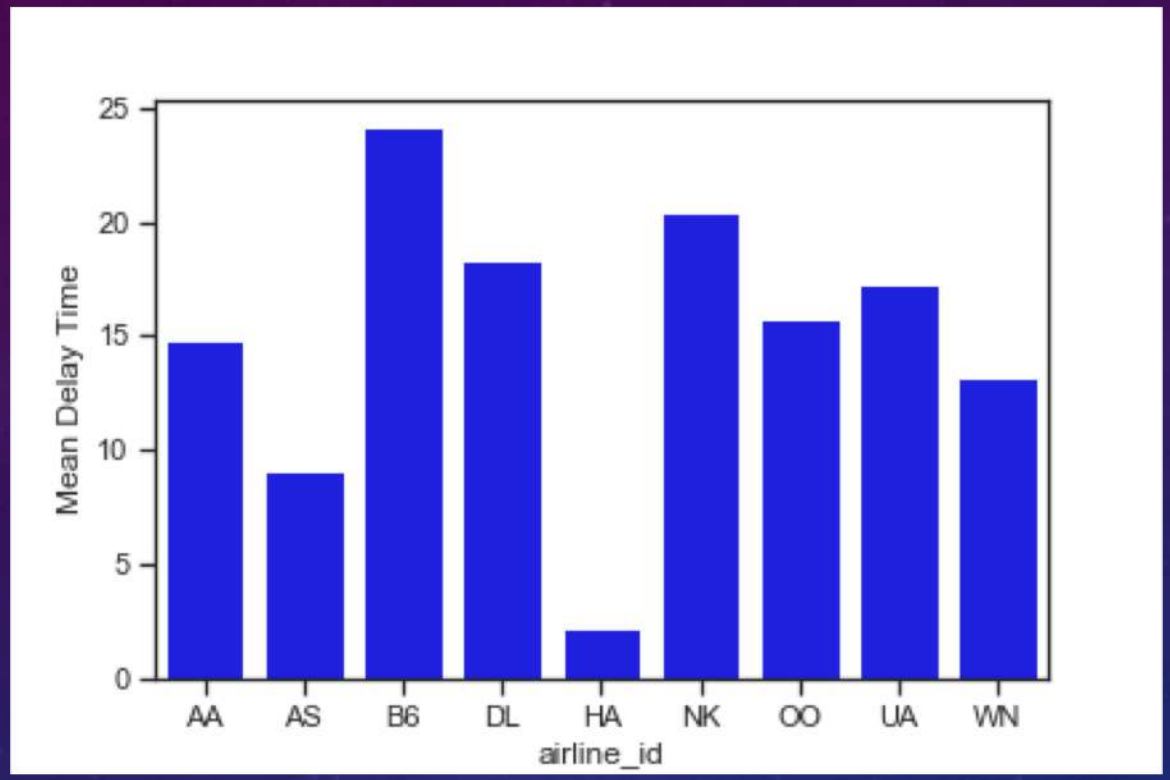
B6, NK, WN, and DL have several extreme high cancellation rate cases. Their performances on cancellation rate are not satisfied.



For the Delay Rate, each airline has an extreme delay rate value of 100%. The mean delay rate of AS and HA are low, but may be due to their small flight amounts.
The mean delay rate of the other airlines are similar, about 20%.
The overall delay rate performance of UA and B6 are relatively bad.There are about 25% of UA flights having a delay rate higher than 40%. B6 is in the same situation. It also has the highest mean delay rate.



- DL has the worst performance on delay time.
- OO and NK are relatively good as their delay time are relatively less.
- AS and HA are good but still may due to the total small amount of flights.
- The other airlines are similar and neutral, though AA has an extreme long delay.



- The mean delay time of AS and HA are relatively low.
- B6 and NK have high mean delay time.
- AA, OO, and WN have a mean delay time less than, or around 15 mins. It shows some different trend as the previous strip plot.

Question: Does delay time differ significantly across airlines?

Hypothesis: the mean delay time is the same for all nine airlines

Analysis of Variance (ANOVA):

	sum_sq	df	F	PR(>F)
education	3.220417e+09	3.0	116.682074	2.849538e-75
Residual	4.159034e+11	45207.0	NaN	NaN

As we can see, looking at the p-value that accompanies the F-statistics, we obtain a strong rejection of the null hypothesis, leading us to conclude that airlines have some differences in the mean of delay time.

However, The ANOVA test does not tell us which pair of groups have means that are different from each other.

Bonferroni Correction

This method divides the significance level by the number of multiple comparisons being performed (36 in this case). So if our alpha=0.05, we will reject the null hypothesis only if the p-value is less than 0.001388.

	Contrast	A	B	Paired	Parametric	T	dof	Tail	p-unc	p-corr	p-adjust	BF10	hedges
7	airline_id	AA	HA	False	True	6.312	179.69	two-sided	2.096083e-09	7.545900e-08	bonf	1.822e+07	0.248
14	airline_id	UA	HA	False	True	7.533	171.44	two-sided	2.720507e-12	9.793827e-11	bonf	3.336e+10	0.364
16	airline_id	B6	AS	False	True	5.555	289.00	two-sided	6.284135e-08	2.262289e-06	bonf	2.842e+05	0.311
18	airline_id	B6	WN	False	True	5.752	1378.12	two-sided	1.085272e-08	3.906980e-07	bonf	5.838e+05	0.267
20	airline_id	B6	HA	False	True	10.514	203.86	two-sided	6.196903e-21	2.230885e-19	bonf	1.218e+21	0.439
25	airline_id	OO	HA	False	True	5.963	222.12	two-sided	9.672700e-09	3.482172e-07	bonf	1.607e+06	0.378
32	airline_id	NK	HA	False	True	5.394	206.00	two-sided	1.884518e-07	6.784266e-06	bonf	6.112e+04	0.460
34	airline_id	WN	HA	False	True	7.352	60.12	two-sided	6.203369e-10	2.233213e-08	bonf	1.569e+10	0.332
35	airline_id	DL	HA	False	True	8.179	172.46	two-sided	5.983616e-14	2.154102e-12	bonf	6.598e+12	0.265

We could conclude that with 95% confidence,
HA has a shorter mean delay time than all other airlines except AS.
B6 has a longer mean delay time than AS, WN, and HA.
We could say that for mean delay time, B6 is the worst, while HA is the best. AS and WN are relatively better than the other airlines.

Conclusions on Airline Navigation Performance on Bad Weather Event Days

- Cancellation Rate: AS, HA > AA, UA, OO > B6, NK, WN, DL
- Delay Rate: AS, HA > AA, OO, NK, WN, DL > UA, B6
- Delay Time:

1. Strip Plot: AS, HA > OO, NK > WN, UA, B6 > AA > DL
2. Histogram of Mean: AS, HA > AA, OO, WN > DL, UA > B6, NK
3. Hypothesis Testing of Mean: HA > AS > WN > AA, UA, DL, NK, OO > B6

Conclusions:

1. As the total flight amount of AS and HA are small, the results may have a high variance.
2. B6 should be the worst one on navigation performance on bad weather event days.
3. OO should be a good choice to avoid bad navigations as it performs well on each aspect.
4. WN and AA are relatively good, though WN performs bad on cancellation rate, and AA has an extreme delay time data point.
5. UA, DL, NK are relatively below average on performance.

THANK YOU