Midterm Project

Predicting flight delays

A machine learning challenge

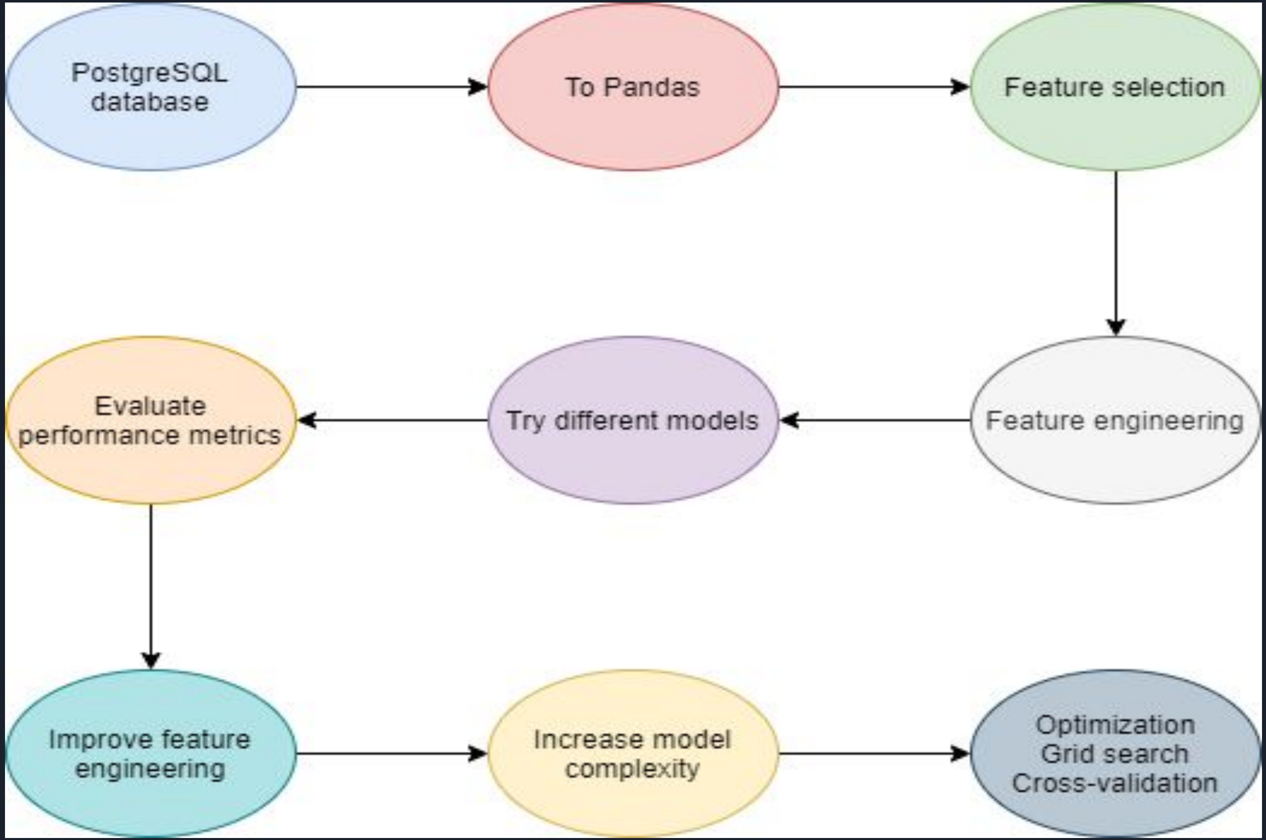
By: Titania Yan & William Li

Why predict flight delays a week in advance?

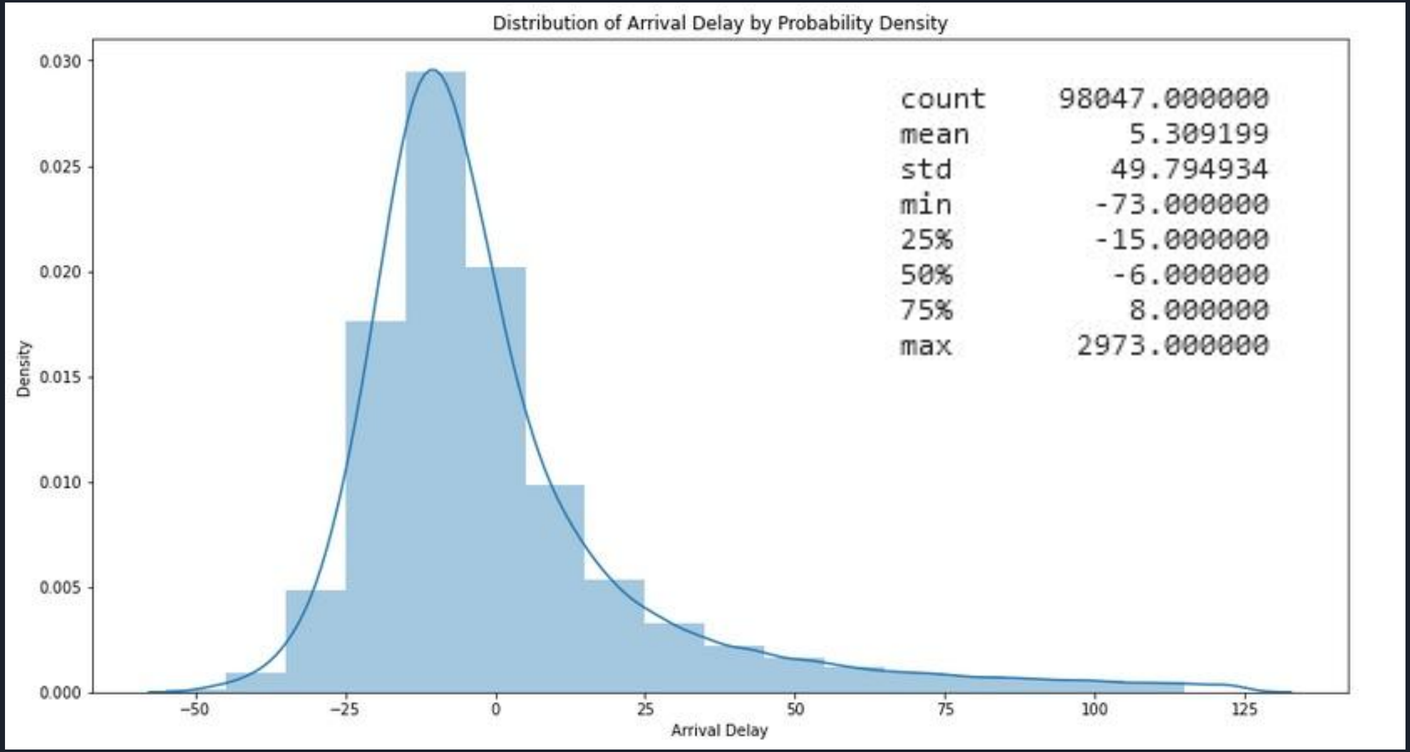
- Help airlines plan for scheduling changes
- Help airports navigate traffic
- Customer satisfaction

Especially challenging since many factors contributing to delays on the day of flight are not known!

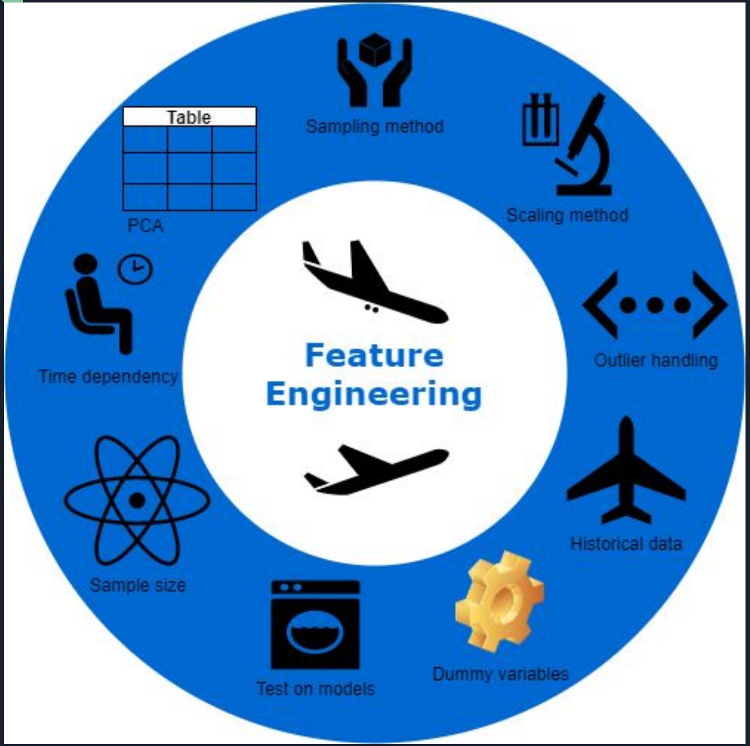
Workflow of project

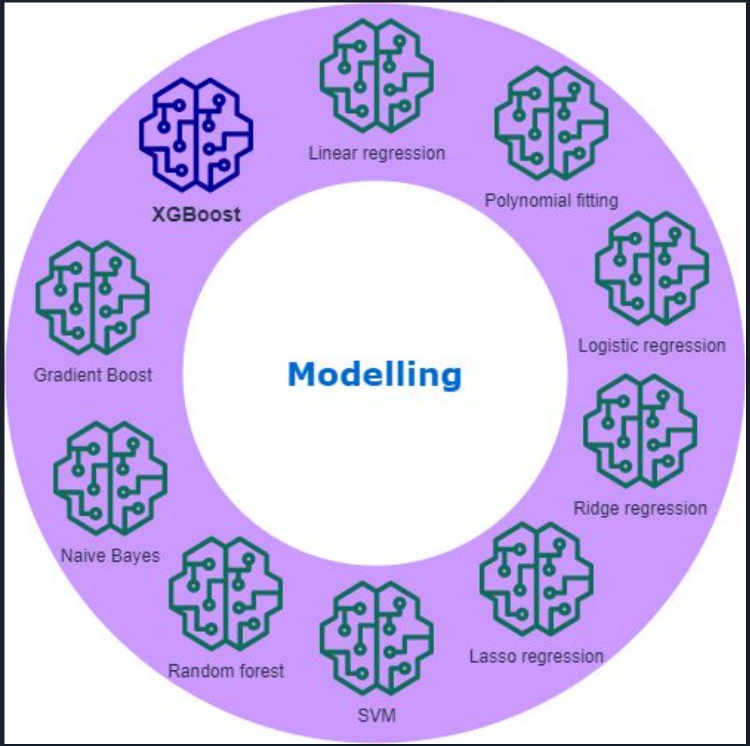


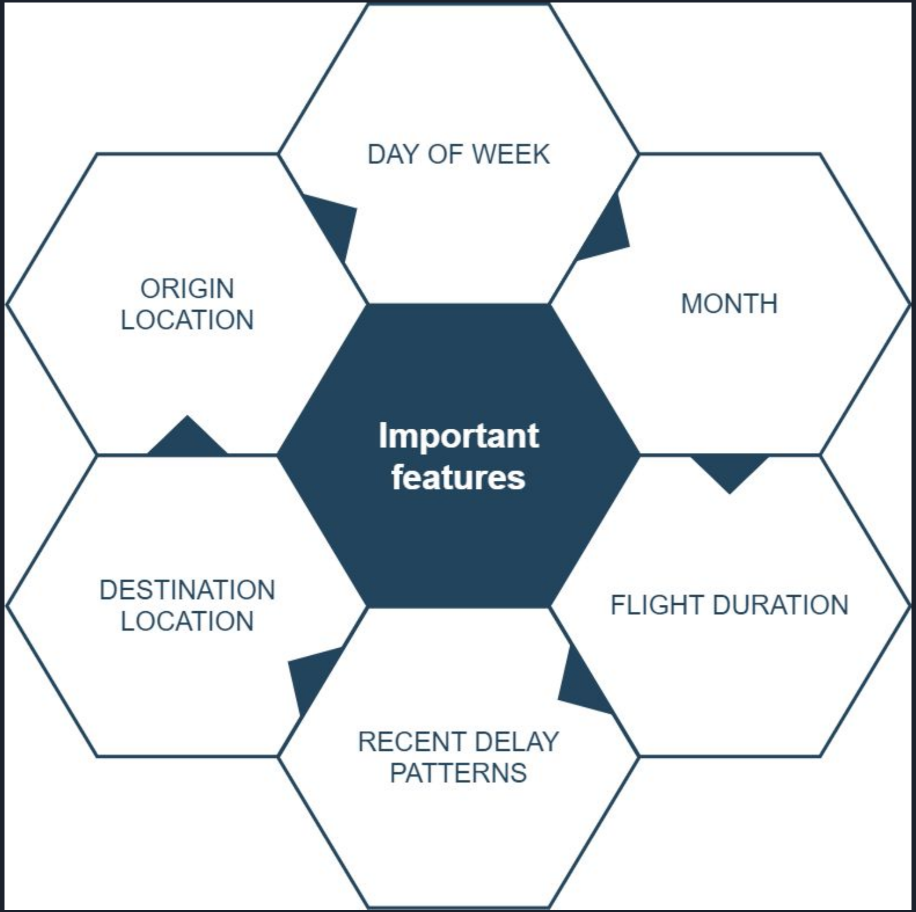
What We're Trying to Predict
How are flight arrival times distributed?



Feature engineering and modelling







These features correlated with flight arrival delay

Choosing Features : Verifying Dependencies ANOVA Analysis

One Way ANOVA Results:

Choosing Features : Verifying Dependencies ANOVA Analysis

<u>One Way ANOVA Results:</u>

	N	Mean	SD	SE	95% Conf.	Interval
month						
1	7555	3.7463	53.0309	0.6101	2.5503	4.9423
2	7014	5.8205	50.2671	0.6002	4.6439	6.9971
3	8271	2.4336	41.8342	0.4600	1.5319	3.3353
4	7999	4.7847	52.6396	0.5886	3.6310	5.9385
5	8454	6.6266	50.3646	0.5478	5.5528	7.7003
6	8377	11.0054	54.9273	0.6001	9.8290	12.1818
7	8662	7.6240	51.9228	0.5579	6.5304	8.7176
8	8653	8.9021	49.6534	0.5338	7.8558	9.9485
9	8185	1.3827	40.9655	0.4528	0.4950	2.2703
10	8505	3.2188	44.0544	0.4777	2.2824	4.1552
11	8040	3.2833	58.3320	0.6505	2.0081	4.5586
12	8332	4.3986	46.2035	0.5062	3.4064	5.3908

Choosing Features : Verifying Dependencies ANOVA Analysis

One Way ANOVA Results:

F statistic=27.16
Pvalue=1.93 * 10^(-57)

	N	Mean	SD	SE	95% Conf.	Interval
month						
1	7555	3.7463	53.0309	0.6101	2.5503	4.9423
2	7014	5.8205	50.2671	0.6002	4.6439	6.9971
3	8271	2.4336	41.8342	0.4600	1.5319	3.3353
4	7999	4.7847	52.6396	0.5886	3.6310	5.9385
5	8454	6.6266	50.3646	0.5478	5.5528	7.7003
6	8377	11.0054	54.9273	0.6001	9.8290	12.1818
7	8662	7.6240	51.9228	0.5579	6.5304	8.7176
8	8653	8.9021	49.6534	0.5338	7.8558	9.9485
9	8185	1.3827	40.9655	0.4528	0.4950	2.2703
10	8505	3.2188	44.0544	0.4777	2.2824	4.1552
11	8040	3.2833	58.3320	0.6505	2.0081	4.5586
12	8332	4.3986	46.2035	0.5062	3.4064	5.3908

Choosing Features : Verifying Dependencies ANOVA Analysis

One Way ANOVA Results:

F statistic=27.16
Pvalue=1.93 * 10^(-57)

Conclusion: There is sufficient evidence to believe mean delay among months are different!

month	N	Mean	SD	SE	95% Conf.	Interval
1	7555	3.7463	53.0309	0.6101	2.5503	4.9423
2	7014	5.8205	50.2671	0.6002	4.6439	6.9971
3	8271	2.4336	41.8342	0.4600	1.5319	3.3353
4	7999	4.7847	52.6396	0.5886	3.6310	5.9385
5	8454	6.6266	50.3646	0.5478	5.5528	7.7003
6	8377	11.0054	54.9273	0.6001	9.8290	12.1818
7	8662	7.6240	51.9228	0.5579	6.5304	8.7176
8	8653	8.9021	49.6534	0.5338	7.8558	9.9485
9	8185	1.3827	40.9655	0.4528	0.4950	2.2703
10	8505	3.2188	44.0544	0.4777	2.2824	4.1552
11	8040	3.2833	58.3320	0.6505	2.0081	4.5586
12	8332	4.3986	46.2035	0.5062	3.4064	5.3908

Choosing Features : Verifying Dependencies ANOVA Analysis

One Way ANOVA Results:

F statistic=27.16
Pvalue=1.93 * 10^(-57)

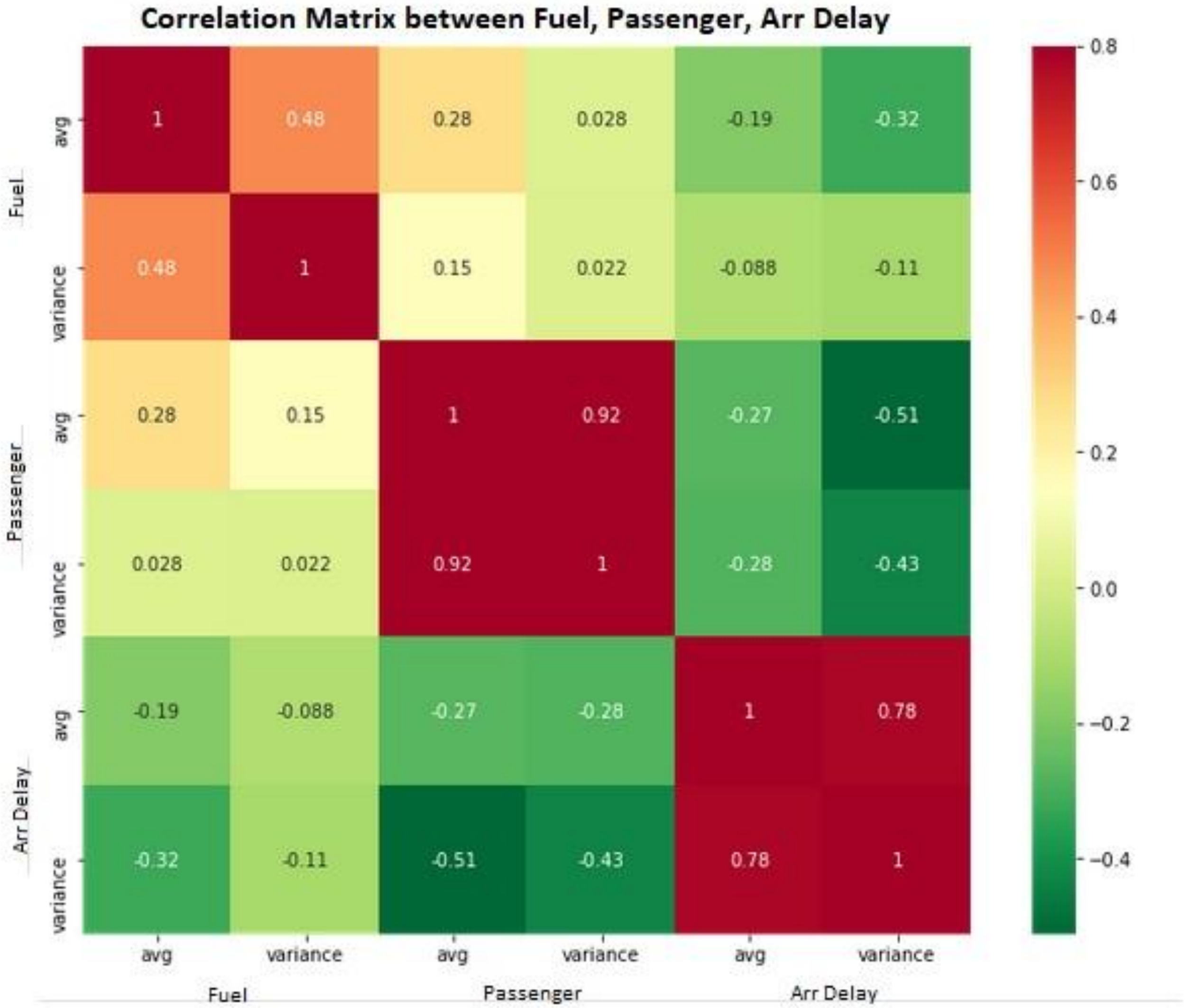
Conclusion: There is sufficient evidence to believe mean delay among months are different!

Task: Figure out what the dependencies are!

month	N	Mean	SD	SE	95% Conf.	Interval
1	7555	3.7463	53.0309	0.6101	2.5503	4.9423
2	7014	5.8205	50.2671	0.6002	4.6439	6.9971
3	8271	2.4336	41.8342	0.4600	1.5319	3.3353
4	7999	4.7847	52.6396	0.5886	3.6310	5.9385
5	8454	6.6266	50.3646	0.5478	5.5528	7.7003
6	8377	11.0054	54.9273	0.6001	9.8290	12.1818
7	8662	7.6240	51.9228	0.5579	6.5304	8.7176
8	8653	8.9021	49.6534	0.5338	7.8558	9.9485
9	8185	1.3827	40.9655	0.4528	0.4950	2.2703
10	8505	3.2188	44.0544	0.4777	2.2824	4.1552
11	8040	3.2833	58.3320	0.6505	2.0081	4.5586
12	8332	4.3986	46.2035	0.5062	3.4064	5.3908

Features Dropped: Features not available one week before a flight, including departure delay

Fuel consumption and passengers were not good predictors of flight delay!



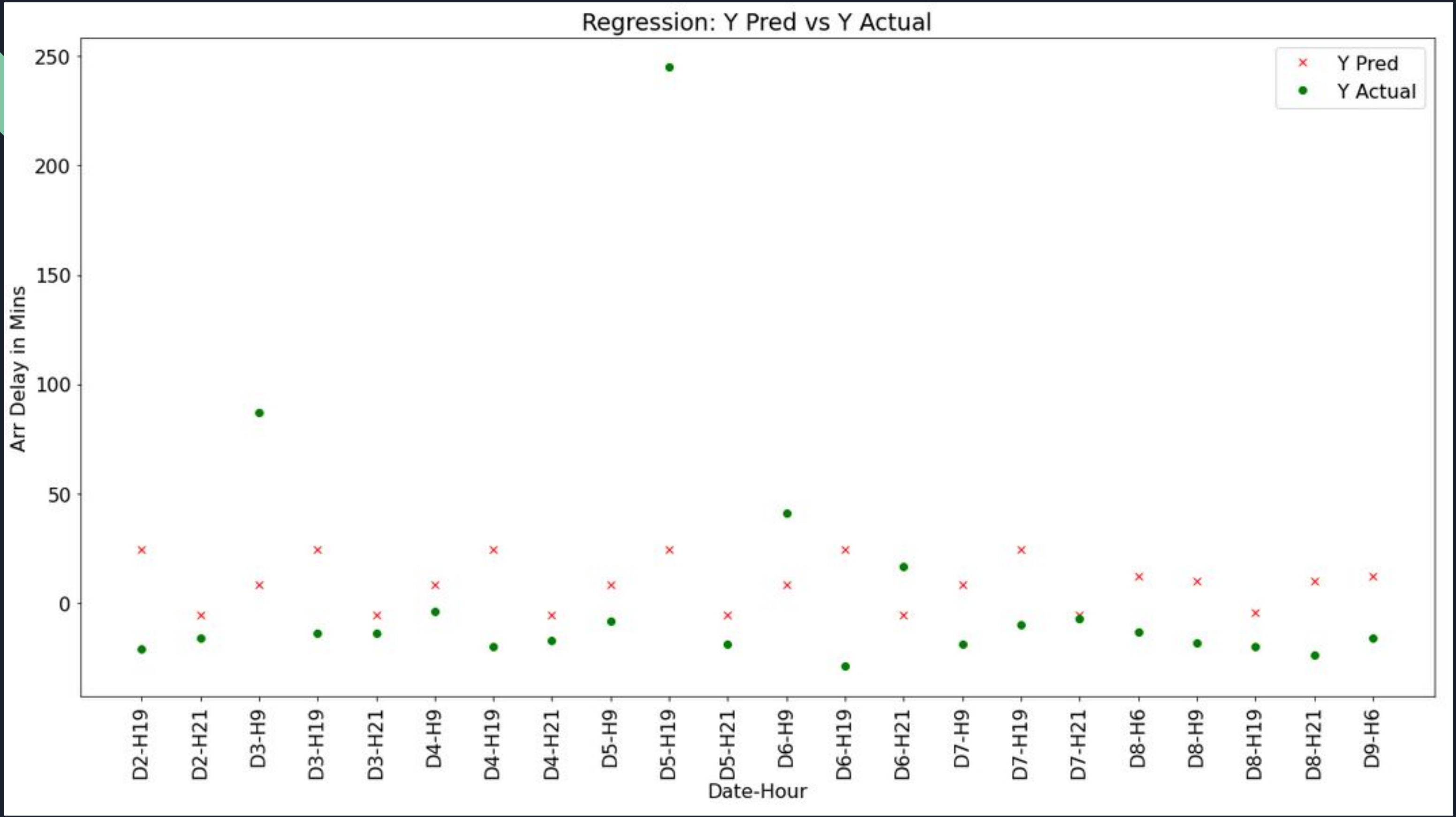
How well can we predict two possibilities?

	Predicted Delay	Predicted No Delay
True Delay	2353	8396
True No Delay	2273	16393

Accuracy	70%
Precision	66%
Recall	88%

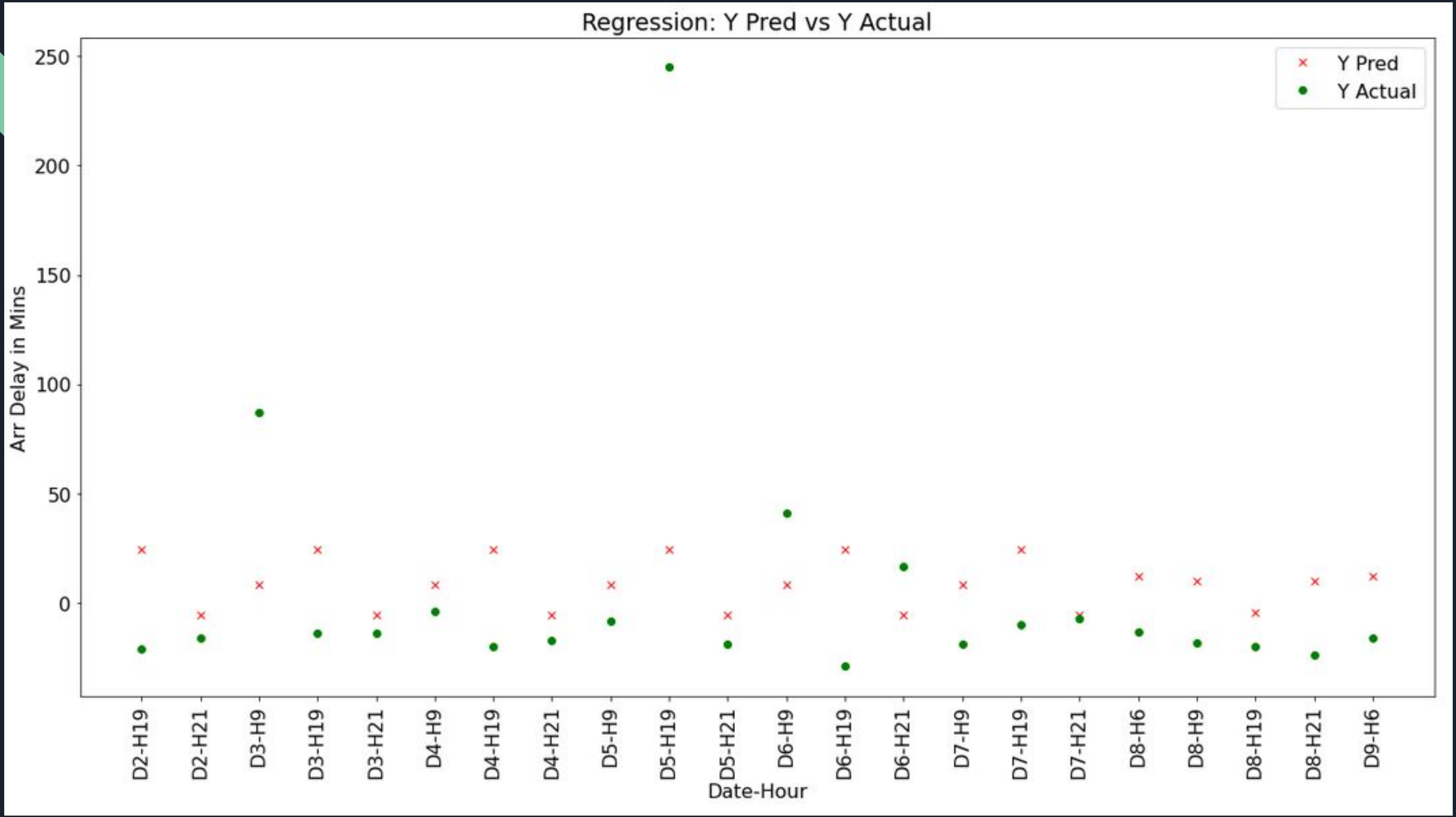
Regression Results

AA: MIA to ATL Route



Regression Results

AA: MIA to ATL Route



RSquare=0.04

Conclusion

- Use patterns found by models
  - Increase Support for busy times and airports
- Eliminate causes of big departure delay further where possible for better metrics
- Continue to strive for better customer satisfaction

THANK YOU!

Questions?