

Flight Delay Predictions



Structure



Our Stakeholders

Problem introduction

**The data: Flights &
Delay**

**Prediction I:
Flight delay: Yes or No?**

**Prediction II:
Flight delay: How
much?**



Conclusion

Stakeholder:



Tunisair is the flag carrier airline of Tunisia.

Operates scheduled international services to four continents.

As of April 2018 The fleet currently consists of 29 Airbus and Boeing aircrafts.

Problem introduction

Nobody likes flight delays as they bring many unfortunate things with them:

- Distressful waiting for passengers
- Missed transfer opportunities
- Additional costs for the airline
- Loss of customers
- Lower efficiency for airports (More organizational effort)

DEPARTURES			
9:00	MONTREAL	TU 0564	DELAYED
9:20	ROTTERDAM	TU 8956	DELAYED
9:50	LUANDA	TU 0399	DELAYED
10:00	ANTALYA	UG 1301	DELAYED
10:20	MADRID	UG 0008	DELAYED
10:45	JEDDAH	TU 0722	DELAYED
11:15	HAMBURG	TU 0250	DELAYED
12:10	KAIRO	TU 0614	DELAYED



Tunisair Dataset

Records of national and international flights:

- **Totally ~ 100k Flights**
- **3 Years of Flight Data (Jan. 2016 - Dec. 2018)**
- **8 Various details (Features) about the flights**
- **Target variable - Delay time of flights**
- **Dataset for geographic data about the airports**



Data insights: Flights & Delay

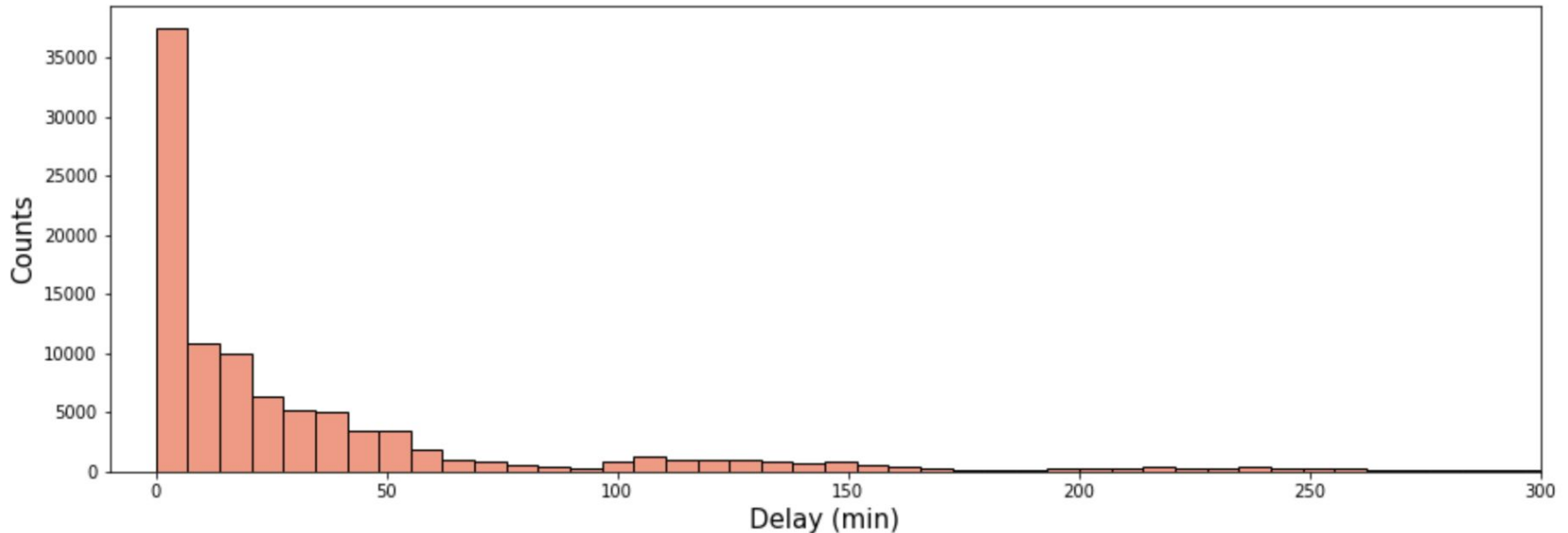
Why are the flights delayed?

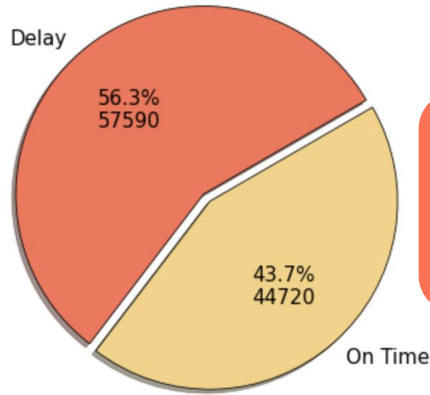
Hypotheses:

- I. More international flights than domestic flights are affected by delays.**
- II. The season, day of the week and time of day play a role in the delay.**
- III. The delay is due to an individual aircraft.**

Data insights: Flights & Delay

Overview of the time distribution of delays: Most flights are delayed less than one hour.

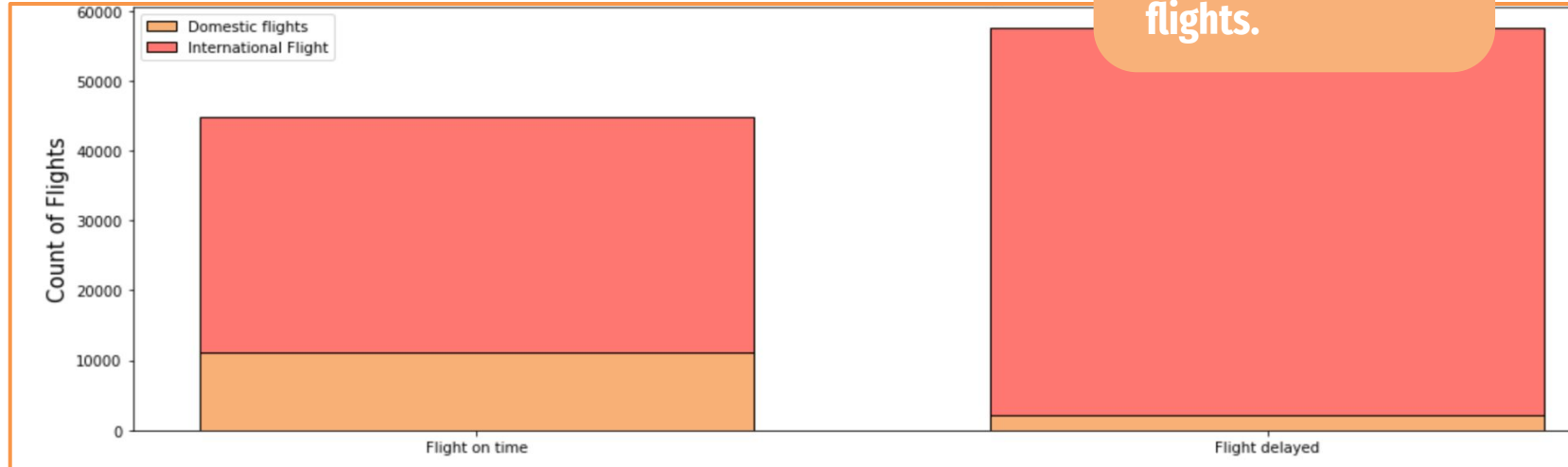




46.37% of our flights are on time and 53.62% are delayed.

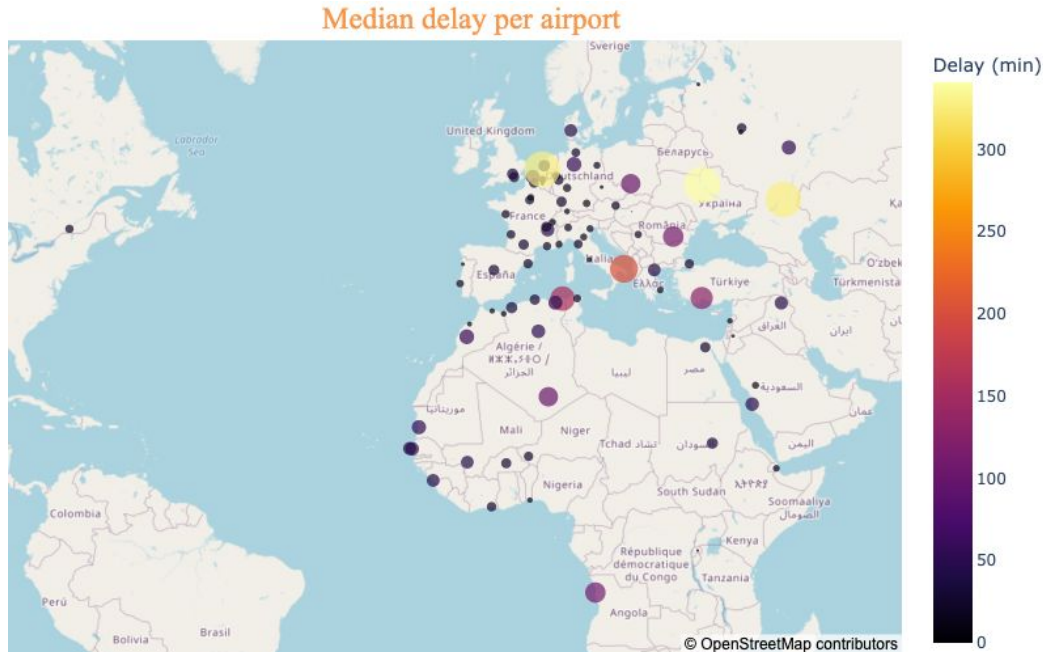
H1: More international than domestic flights are affected by delay.

95.77% of the delayed flights are international flights.



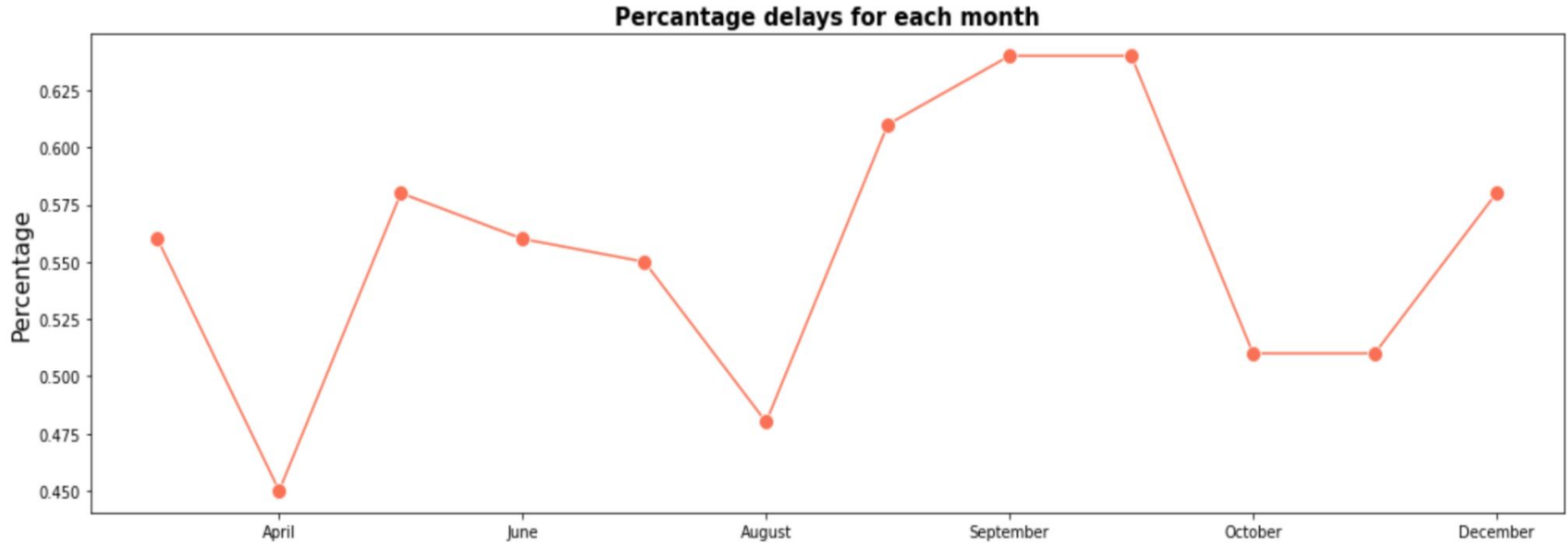
H1: More international than domestic flights are affected by delay.

Median delay per airport



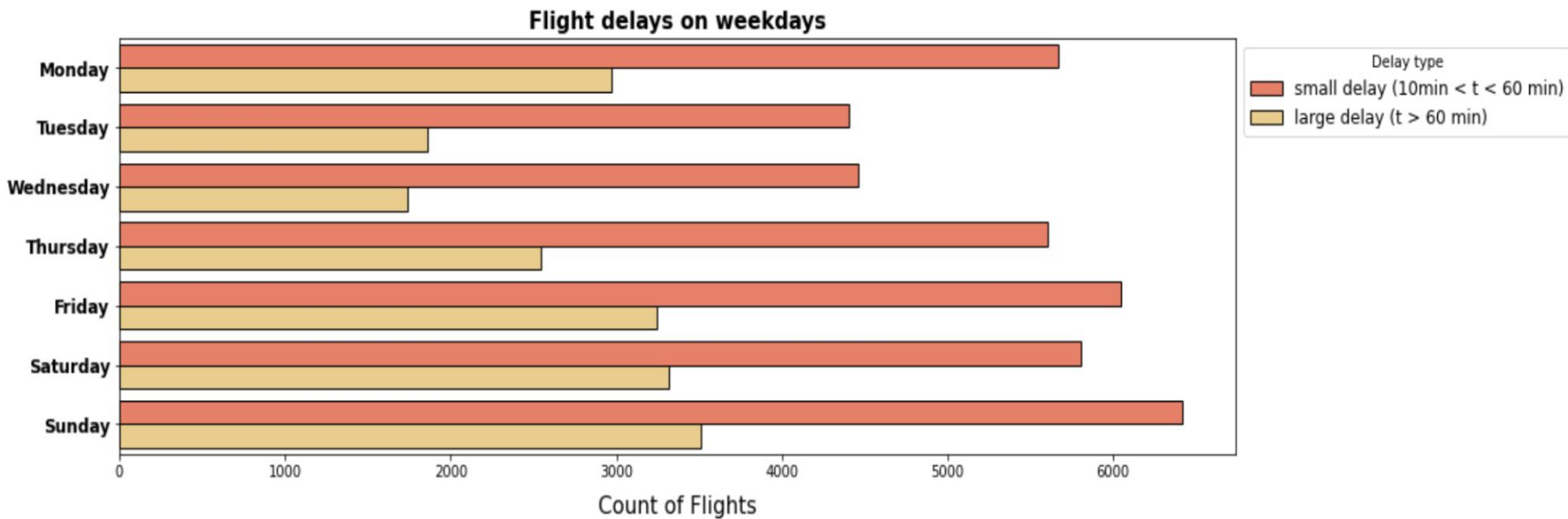
Airports with largest departure delays:
Volgograd (Russia)
Rotterdam (Netherlands)
Boryspil (Ukraine)

H2: The season, day of the week and time of day play a role in the delay.



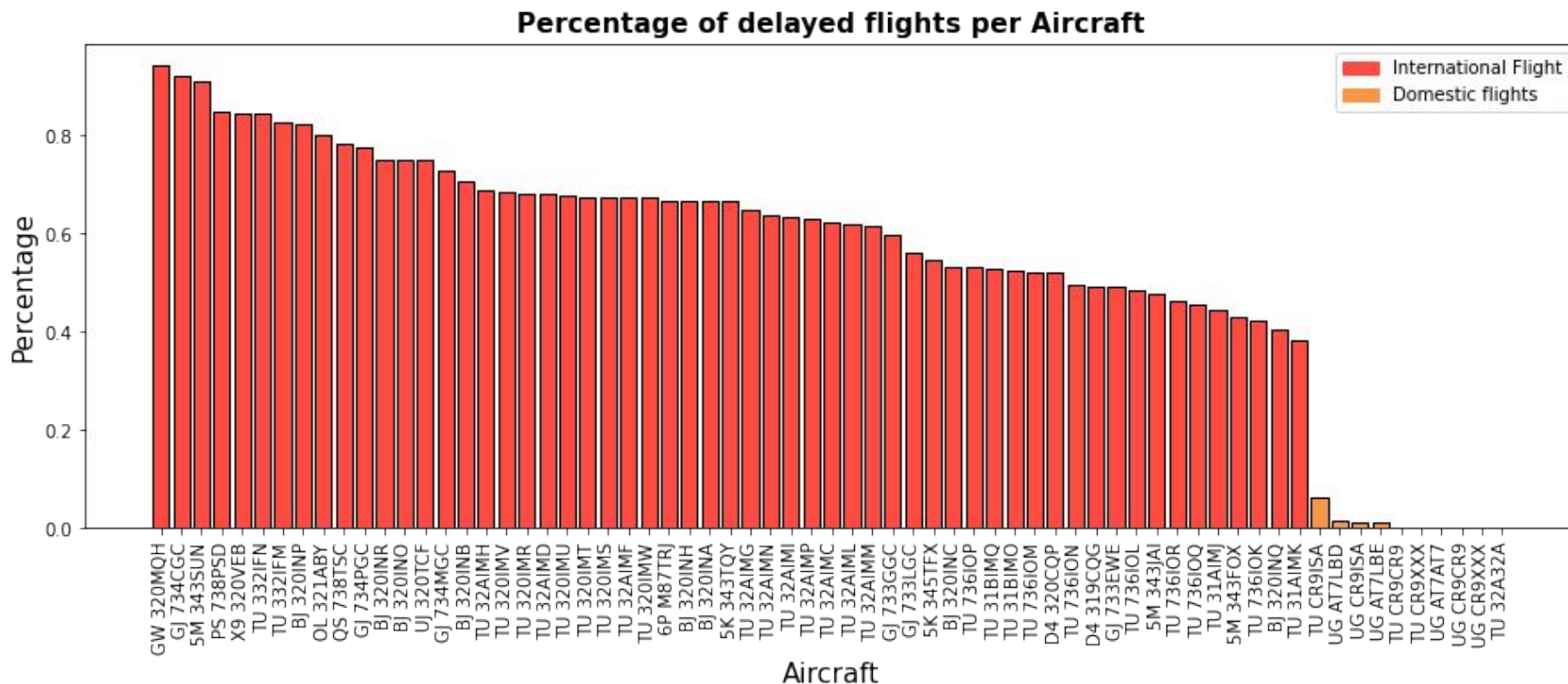
We can see a peak in percentage delays August/ September.

H2: The season, day of the week and time of day play a role in the delay.



We can observe higher counts of flights that are delayed on the weekend and the beginning of the week.

H3: The delay is due to an individual aircraft.



Six aircrafts with more than 80% delayed flights

Modeling



01

Classification

Is a flight delayed
or not



02

Regression

How much delay
does a flight have



03

Prediction

Combining both to
predict delay

Baseline Model

Baseline Model: Classification

Our baseline classifies flights with an accuracy of 58%.

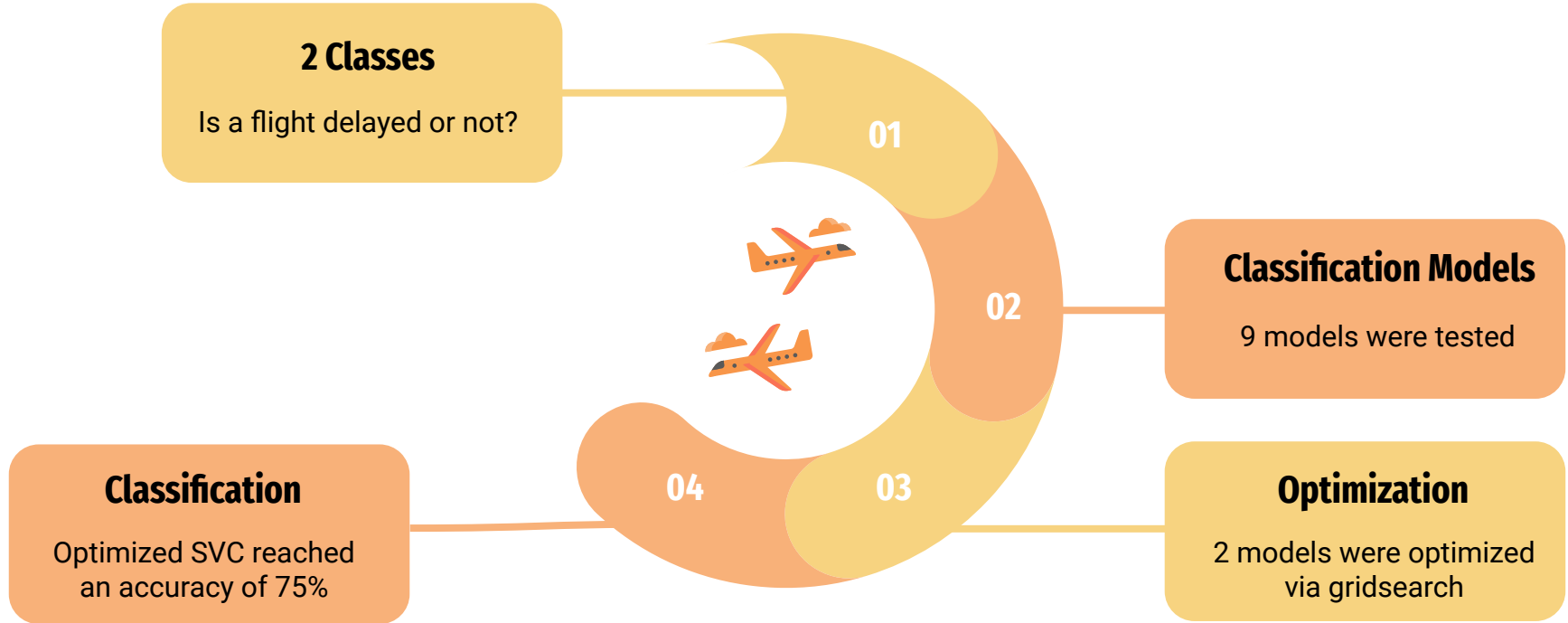
Baseline Model: Regression

Our baseline model predicts a delay with a root mean squared error of 140.09 minutes.



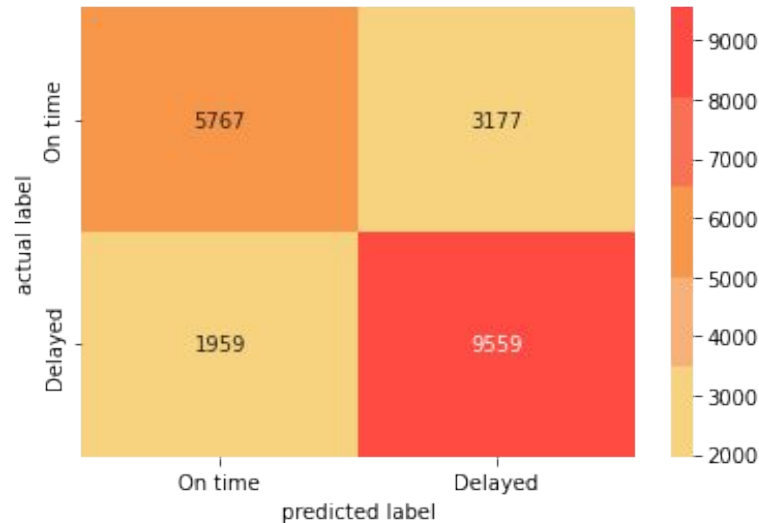
Those values we have to beat
with our sophisticated models.

I - Classification



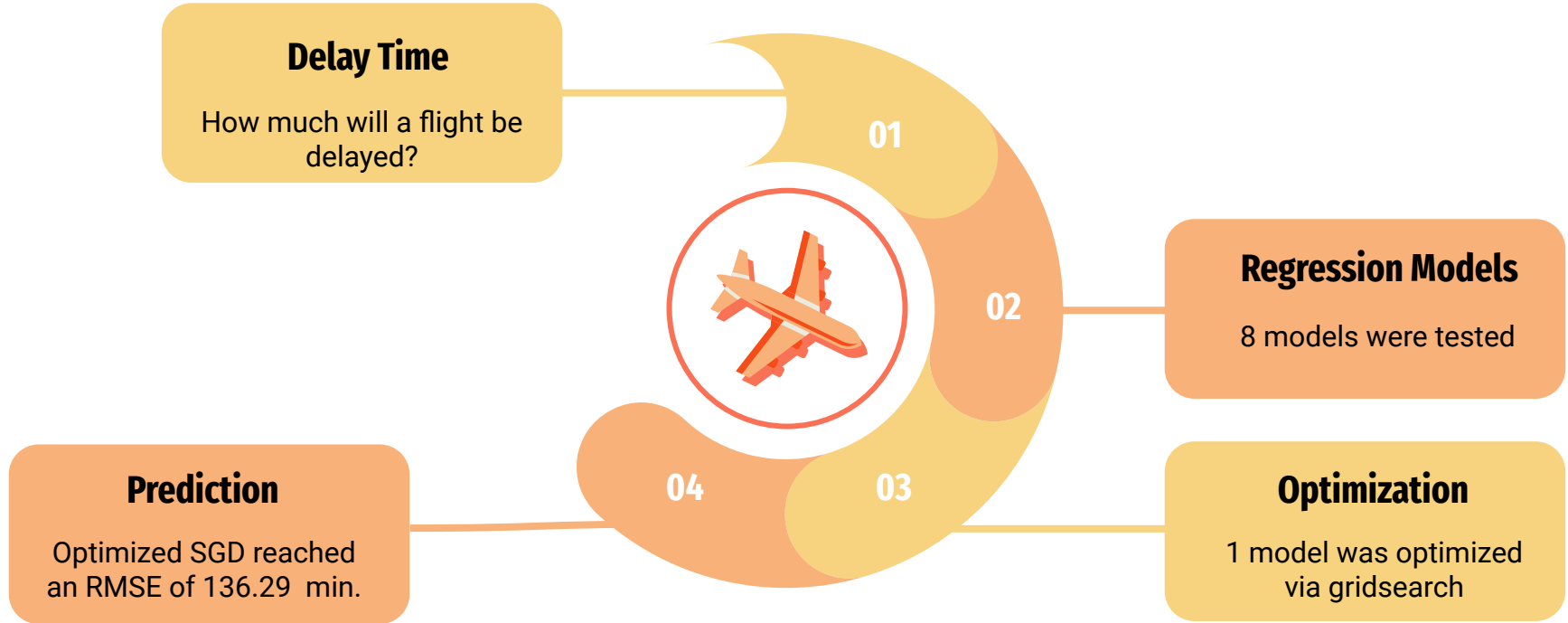
I - Classification

The best model classifies flights with an accuracy of 75%



83% of delayed flights are detected

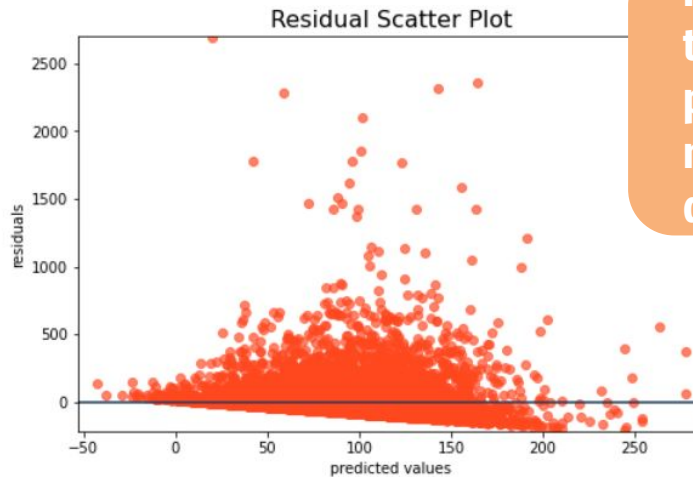
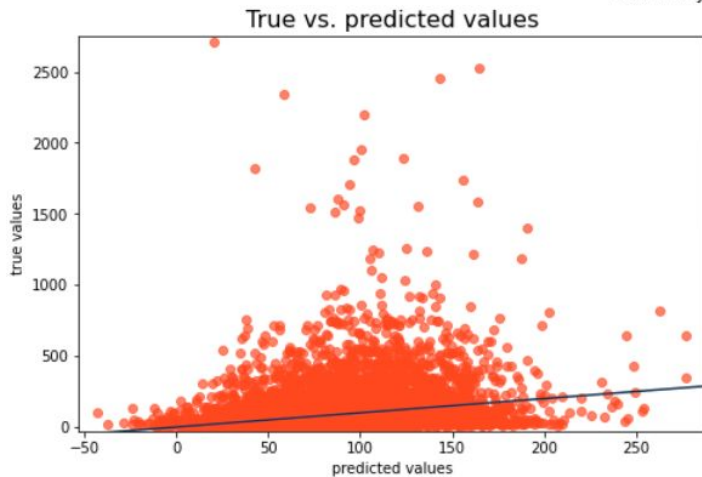
II - Regression



Prediction II: How much is the delay?

Our best model predicts the true value of the delay with an root mean squared error of 136.29 minutes.

Error Analysis



Here we also see the model still predicts bad on most of the delays.

Conclusions



Predicting whether a flight is delayed or not works reasonably well from the existing data

Predicting exact delay times is not possible. Further data and analysis will be needed

Customer satisfaction and profits could be increased by deploying an automated delay predictor.

Thank you for listening!

