



How to (not) arrive on Time

Analyzing and Predicting US Flight Delays

Created by Manuel Klein
01/15/2020

» neue fische
School and Pool for Digital Talent

Picture Source:

<https://static.independent.co.uk/s3fs-public-thumbnails/image/2019/04/30/19/fly.jpg>



Manuel Klein
01/15/2020

2

→ neue fische
School and Pool for Digital Talent

Picture Source:

https://www.toonpool.com/user/1233/files/airport_delays_2277305.jpg



Table of Contents

- 1. Problem Statement
- 2. Business Value
- 3. Methodology
- 4. Dataset Overview
- 5. Aircraft Delays
 - a. Time Perspective
 - b. Airline Perspective
 - c. Airport Perspective
 - d. Flightroute Perspective
- 6. Delay Causes
- 7. Predictive Modeling
- 8. Summary
- 9. Future Work



Problem Statement

- When looking for flights you want...
 - to have an idea which factors are relevant for aircraft delays
 - to know in advance how high the delay probability for a specific flight is



Business Value

- Understanding influencing factors for delays helps customers to choose a flight with a low delay probability
- > Can e.g. be offered as a service on flight search websites



Methodology

- Data Understanding
- Data Cleaning
- Data Visualization
- Predictive Modeling

The methodology is based on the Data Science Lifecycle



Dataset Overview

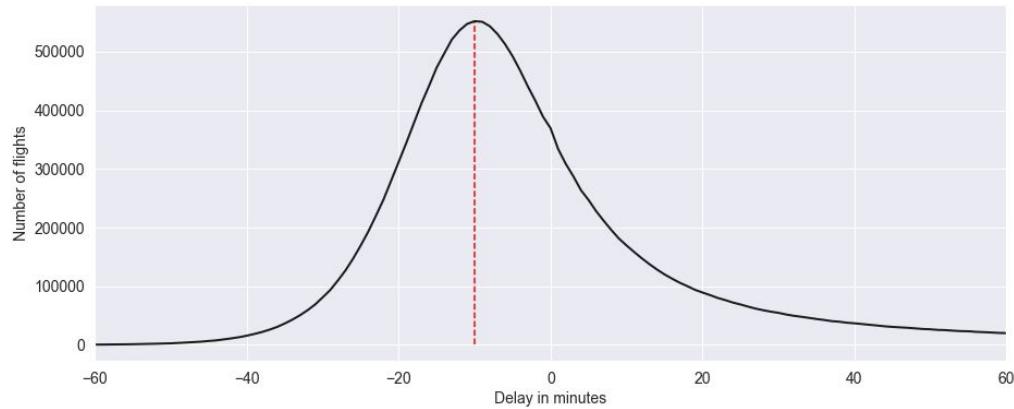
- All US Flights between 2016 and 2018
- 18.2 Million flights in total
- 3302 Flight Routes
- 358 Airports
- 18 Airlines

The dataset used for this project can be found on the following website (only the files for 2016, 2017 and 2018 were used):

<https://www.kaggle.com/yuanyuwendymu/airline-delay-and-cancellation-data-2009-2018>



Aircraft Delay Distribution



Manuel Klein
01/15/2020

8

→ neue fische
School and Pool for Digital Talent

The delay distribution is almost a normal distribution with a peak value for -10 minutes. It can clearly be seen that it is way more likely to arrive too early than arriving too late.



Aircraft Delay – Time Perspective



Manuel Klein
01/15/2020

9

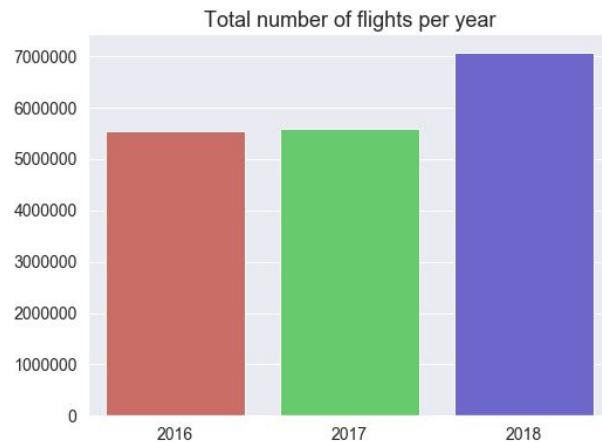
→ neue fische
School and Pool for Digital Talent

Picture Source:

<https://cdn2.iconfinder.com/data/icons/meeting-13/64/flight-time-travel-plane-worldwide-512.png>



Aircraft Delay – Time Perspective



Manuel Klein
01/15/2020

10

→ neue fische
School and Pool for Digital Talent

The total number of flights is growing worldwide every year. This is also valid for the US.



Aircraft Delay – Time Perspective

Percentage of delayed flights from 2016 to 2018 (Delay definition: 0 minutes)



While the total number of flights grows, the percentage of delayed flights also increases year by year, at least during 2016 and 2018 in the US.



Aircraft Delay – Time Perspective



Manuel Klein
01/15/2020

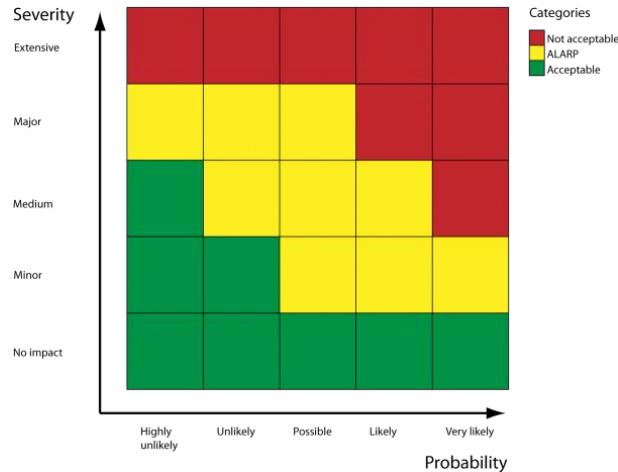
12

→ neue fische
School and Pool for Digital Talent

Not just the percentage of delayed flights increases over the years, but also the amount of delay.



Aircraft Delay Rating Concept



Manuel Klein
01/15/2020

13

→ neue fische
School and Pool for Digital Talent

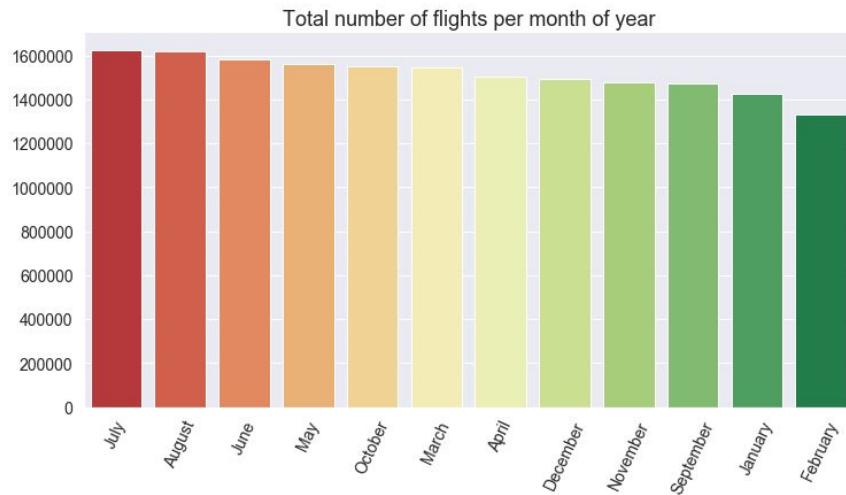
Picture source:

https://www.cgerisk.com/knowledgebase/images/thumb/4/4f/Risk_matrix.png/600px-Risk_matrix.png

The risk matrix from the risk management is adapted to show the probability of delay and the amount of delay (=severity) for several different influencing factors on the following slides.



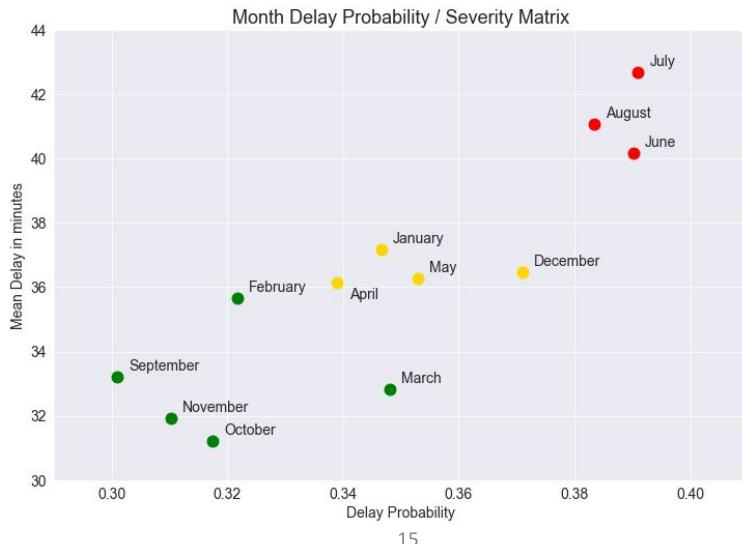
Aircraft Delay – Time Perspective



Most flights in the US take place during the summer months June, July and August.



Aircraft Delay – Time Perspective



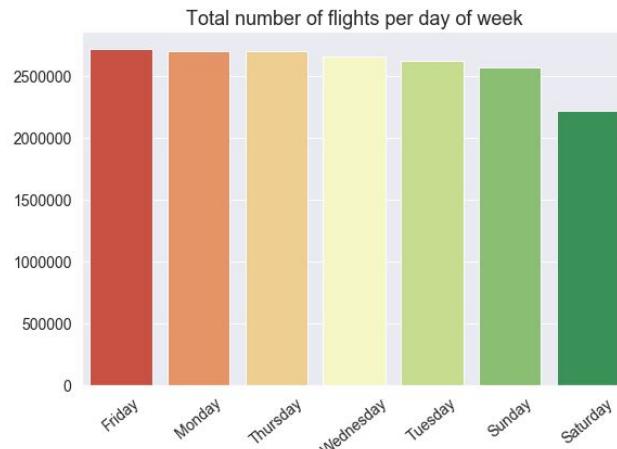
Manuel Klein
01/15/2020

→ neue fische
School and Pool for Digital Talent

During these summer months, also the percentage of delay and the amount of delay is very high. Avoid flying in the summer months if you want to arrive on time!



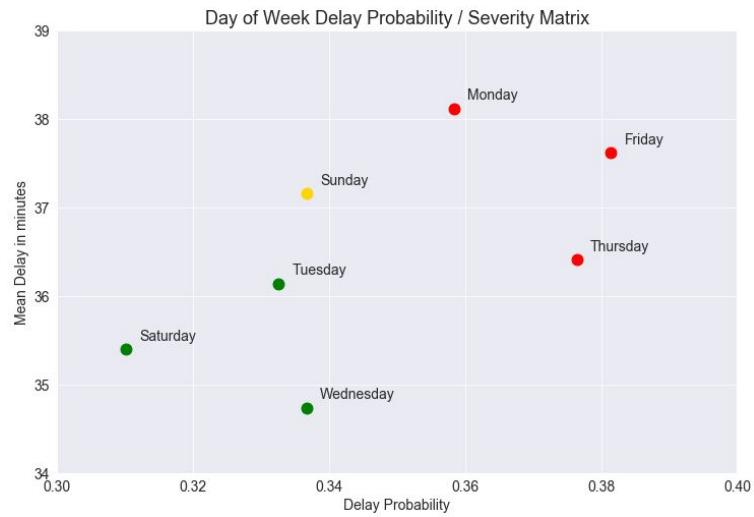
Aircraft Delay – Time Perspective



Most flights take place on the days of the week when usually business people fly:
Monday, Thursday and Friday



Aircraft Delay – Time Perspective



Manuel Klein
01/15/2020

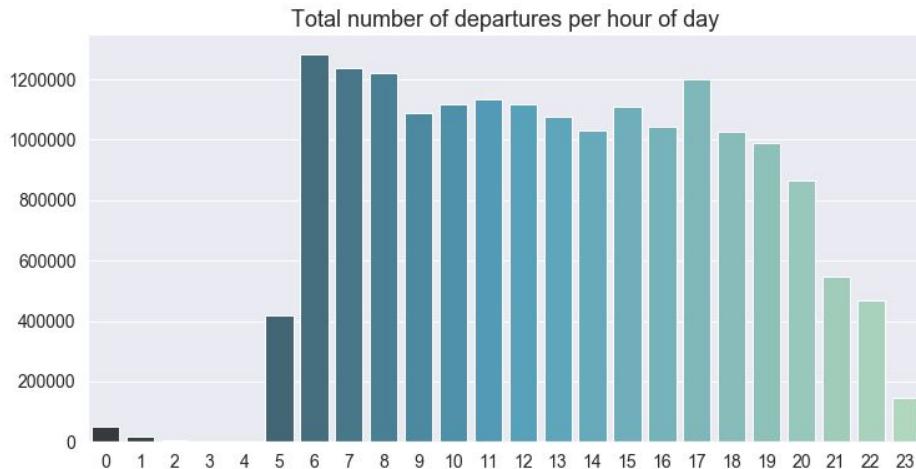
17

→ neue fische
School and Pool for Digital Talent

The product of delay probability and mean delay is for these three days of the week higher than for the rest of the days. Avoid flying on these days if you want to avoid delays!



Aircraft Delay – Time Perspective



Manuel Klein
01/15/2020

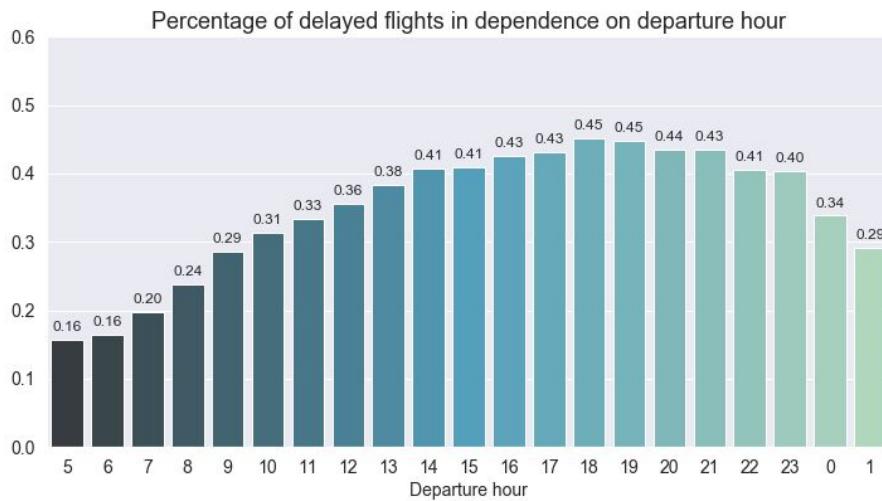
18

→ neue fische
School and Pool for Digital Talent

Flight operations start usually at around 5am in the morning. They stay on a relatively high level till round about 7pm. Then they decrease again.



Aircraft Delay – Time Perspective



Manuel Klein
01/15/2020

19

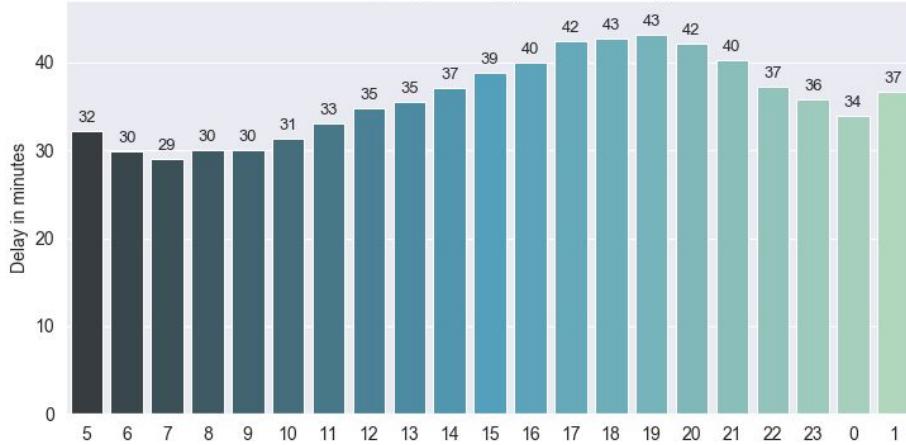
→ neue fische
School and Pool for Digital Talent

The percentage of delayed flights increases during the day and reaches its maximum at around 6pm, before decreasing again.



Aircraft Delay – Time Perspective

Average flight delay per hour of day



Manuel Klein
01/15/2020

20

→ neue fische
School and Pool for Digital Talent

The amount of delay also increases during the day. Therefore, fly early if possible to avoid delays!



Aircraft Delay – Airline Perspective



Manuel Klein
01/15/2020

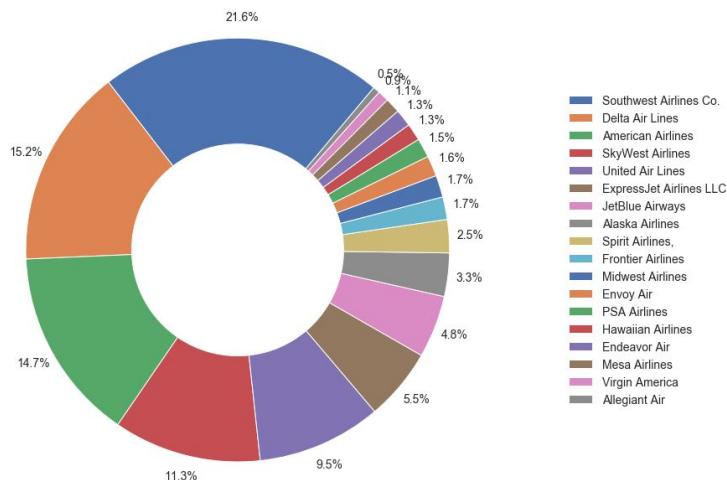
21

→ neue fische
School and Pool for Digital Talent



Aircraft Delay – Airline Perspective

Market share of US airlines



Manuel Klein
01/15/2020

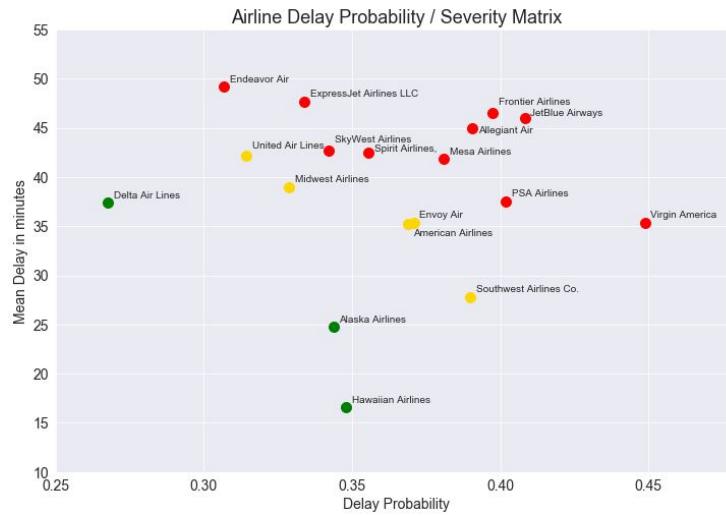
22

→ neue fische
School and Pool for Digital Talent

The three biggest US airlines have a market share of around 50% in the US.



Aircraft Delay – Airline Perspective



Manuel Klein
01/15/2020

23

→ neue fische
School and Pool for Digital Talent

One of the three biggest airlines, Delta Airlines, is very good rated and should be chosen when you have the choice between the three biggest airlines. This is quite often the case as the big airlines cover the majority of the available flight routes.



Aircraft Delay – Airport Perspective

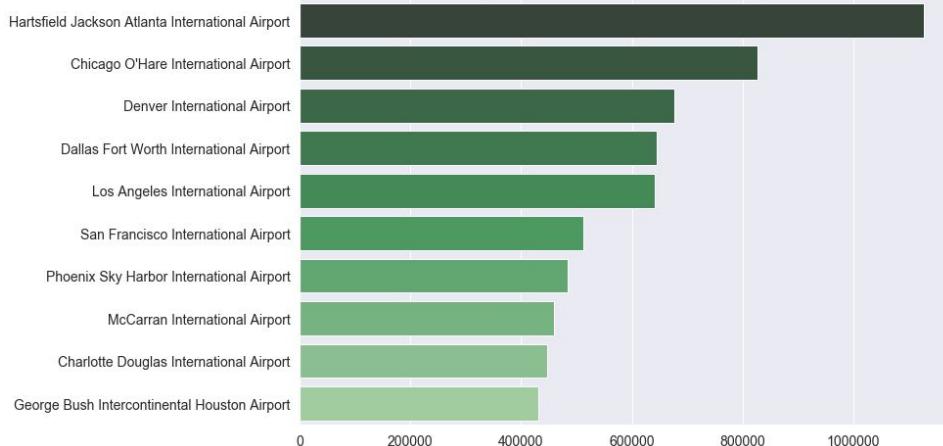


Picture Source: <https://image.flaticon.com/icons/png/512/48/premium/48875.png>



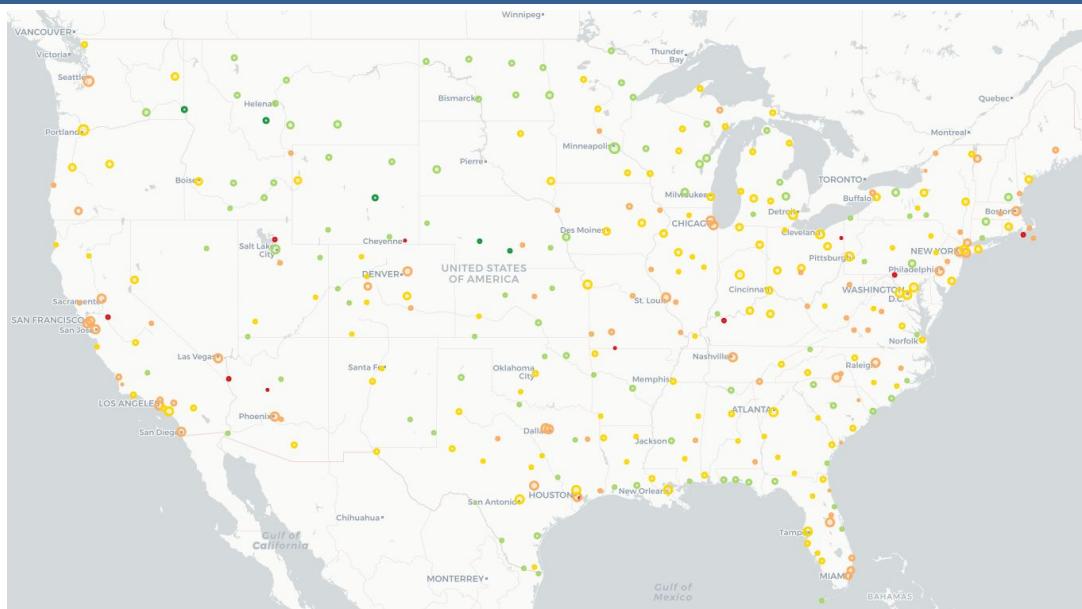
Aircraft Delay – Airport Perspective

Number of flights of 10 most frequented US airports (2016 to 2018)



Atlanta International Airport is the biggest airport in the US and one of the biggest airports in the world with around 1000 departures every single day.

Analyzing and Predicting US Flight Delays



Manuel Klein
01/15/2020

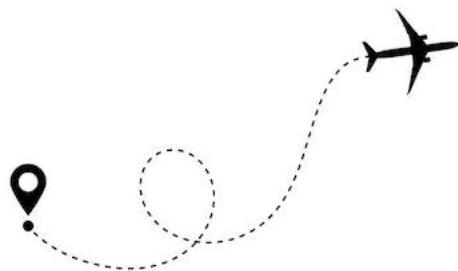
26

→ neue fische
School and Pool for Digital Talent

All US airports have been plotted onto a map with the use of Folium. The size of the circles represents the number of flights (the larger, the more flights), the color represents the percentage of delayed flights (green = low delay, red = high delay).



Aircraft Delay – Flight Route Perspective



Manuel Klein
01/15/2020

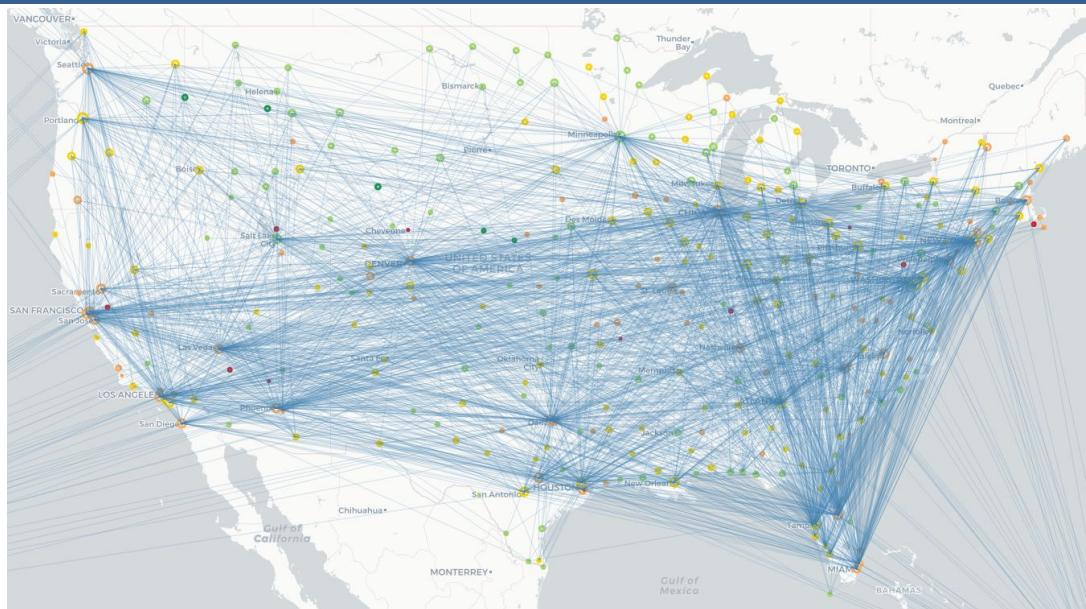
27

→ neue fische
School and Pool for Digital Talent

Picture Source:

<https://image.shutterstock.com/image-vector/airplane-line-path-vector-icon-260nw-1153316677.jpg>

Analyzing and Predicting US Flight Delays



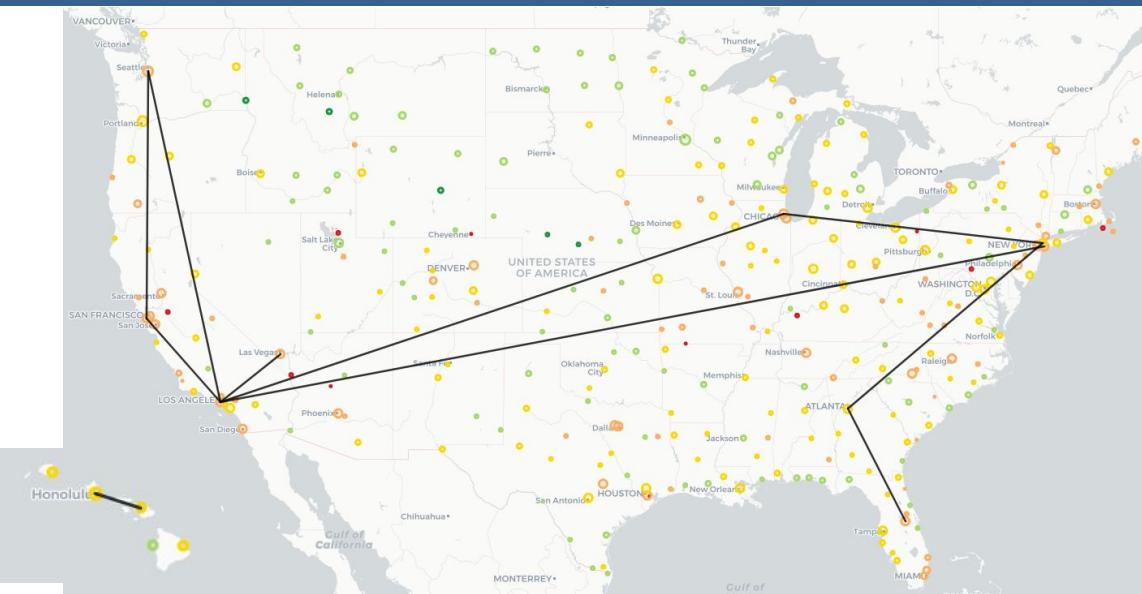
Manuel Klein
01/15/2020

28

→ neue fische
School and Pool for Digital Talent

This map shows all US flight routes that have been available between 2016 and 2018 in the US.

Analyzing and Predicting US Flight Delays



Manuel Klein
01/15/2020

29

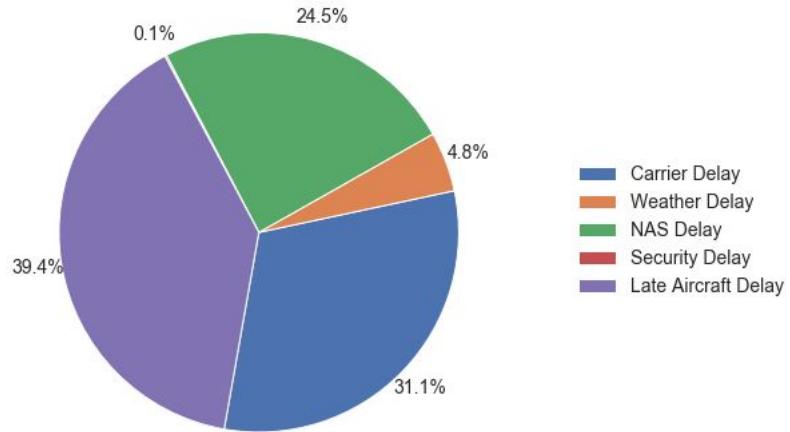
→ neue fische
School and Pool for Digital Talent

On this map, the 10 flightsroutes with the highest number of flights per day have been marked. These are mostly flight routes between the big cities and the islands of Hawaii.



Aircraft Delay – Reasons

Delay Reasons Segmentation



Manuel Klein
01/15/2020

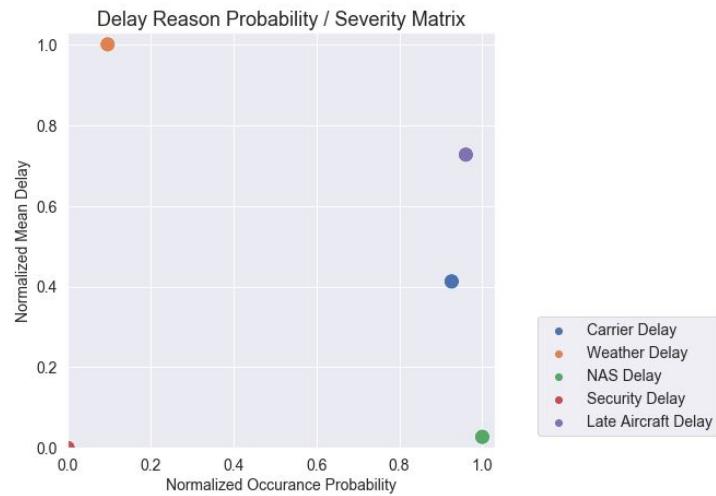
30

→ neue fische
School and Pool for Digital Talent

The major reason for late departures are late arrivals of the same plane. Carrier delay includes all delays originated by the airlines. NAS delay refers to the air traffic control.



Aircraft Delay – Reasons



Manuel Klein
01/15/2020

31

→ neue fische
School and Pool for Digital Talent

Extreme weather events occur not that often but when they occur, they have a massive impact on delays.



Aircraft Delay – Reasons



Manuel Klein
01/15/2020

32

→ neue fische
School and Pool for Digital Talent

Picture Source: <https://www.flightradar24.com/>

This picture from flightradar24 shows planes flying around Iraq and Syria for safety reasons. As these detours are planned, no delay occurs.



Aircraft Delay – Reasons



Manuel Klein
01/15/2020

33

→ neue fische
School and Pool for Digital Talent

Picture Source: <https://youtu.be/99Sx8l1tAK0>

This is an example for NAS delay. The purple lines show flights taking a detour around France. The reason is that the French air traffic controllers went on strike on that day. Thus the air space capacity was very limited on that day over France. As these detours were NOT planned, affected flights were delayed.



Aircraft Delay – Reasons

Extreme Weather Example: Hurricane Irma



Manuel Klein
01/15/2020

34

→ neue fische
School and Pool for Digital Talent

Picture Source:

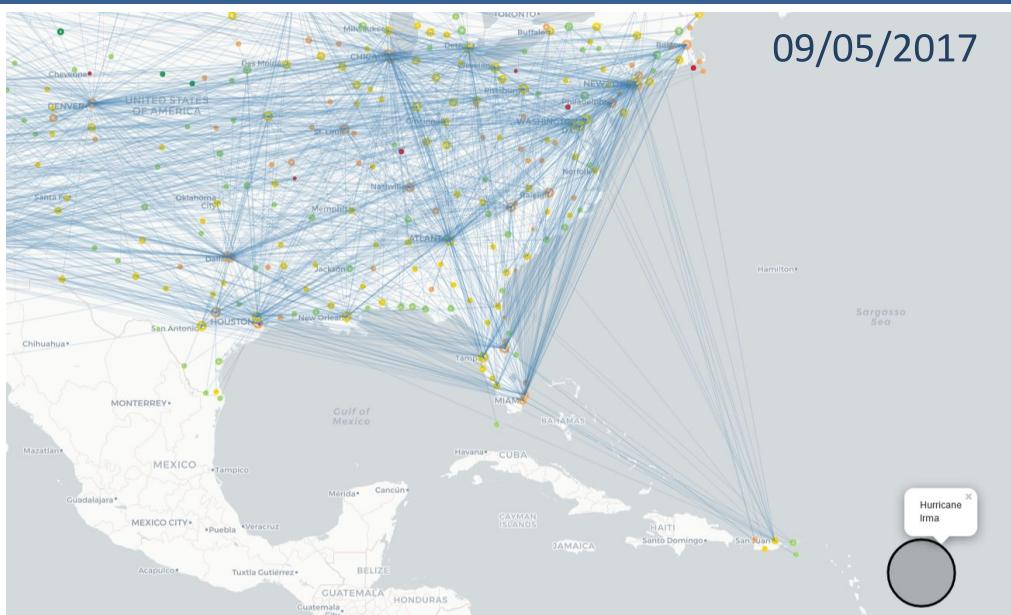
[https://www.baynews9.com/content/dam/News/2019/09/10/irma-iss-nasa-090717-1jpg/jcr:content/renditions/cq5dam.thumbnail.767.431.margin.png?wid=1250&hei=703&\\$wide-bg\\$](https://www.baynews9.com/content/dam/News/2019/09/10/irma-iss-nasa-090717-1jpg/jcr:content/renditions/cq5dam.thumbnail.767.431.margin.png?wid=1250&hei=703&$wide-bg$)

The following slides show the impact of the Hurricane Irma onto the available flightroutes during September 2017.

Analyzing and Predicting US Flight Delays



09/05/2017



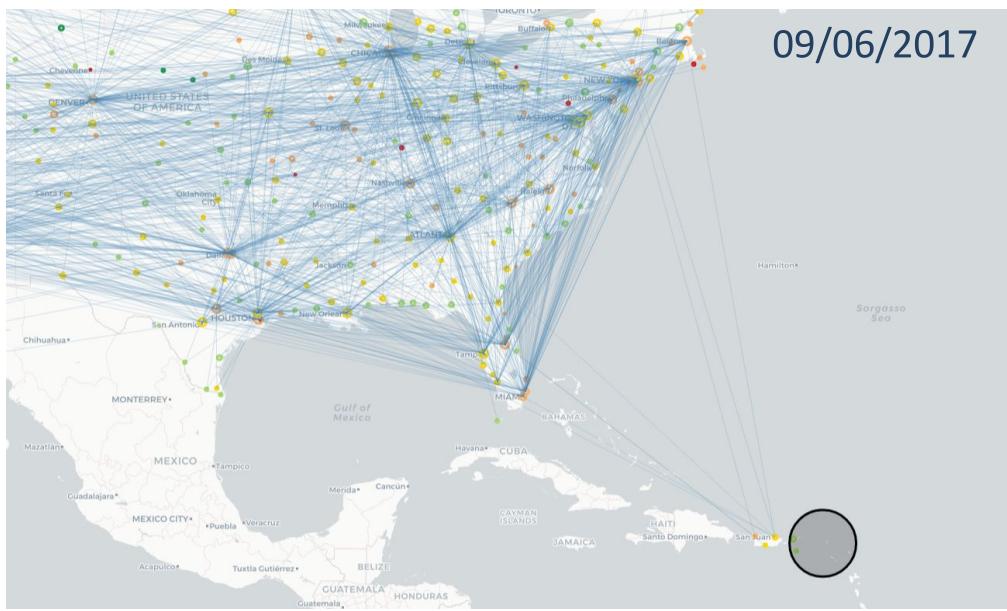
Manuel Klein
01/15/2020

35

→ neue fische
School and Pool for Digital Talent

Analyzing and Predicting US Flight Delays

09/06/2017



Manuel Klein
01/15/2020

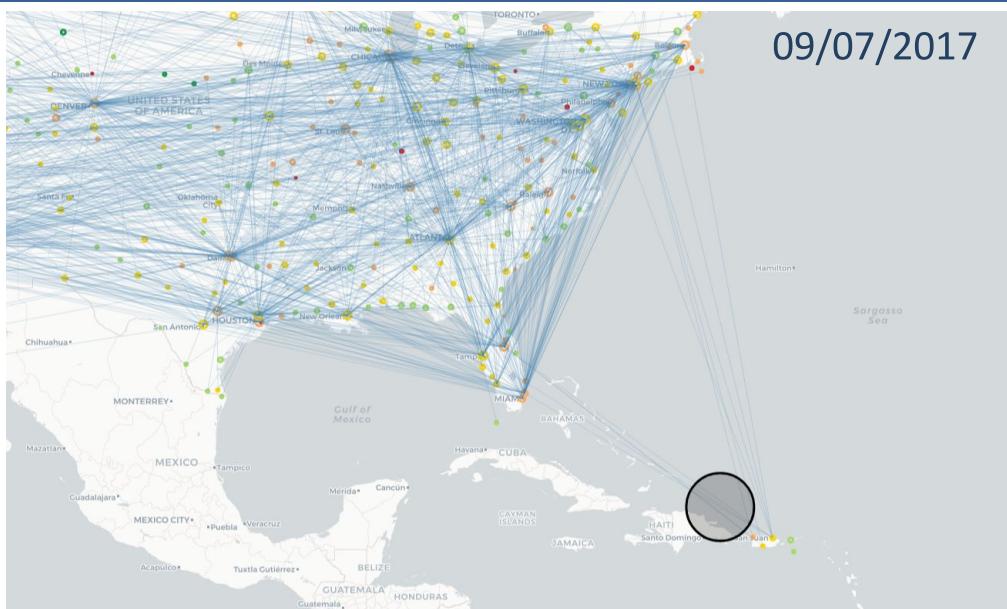
36

→ neue fische
School and Pool for Digital Talent

Analyzing and Predicting US Flight Delays



09/07/2017



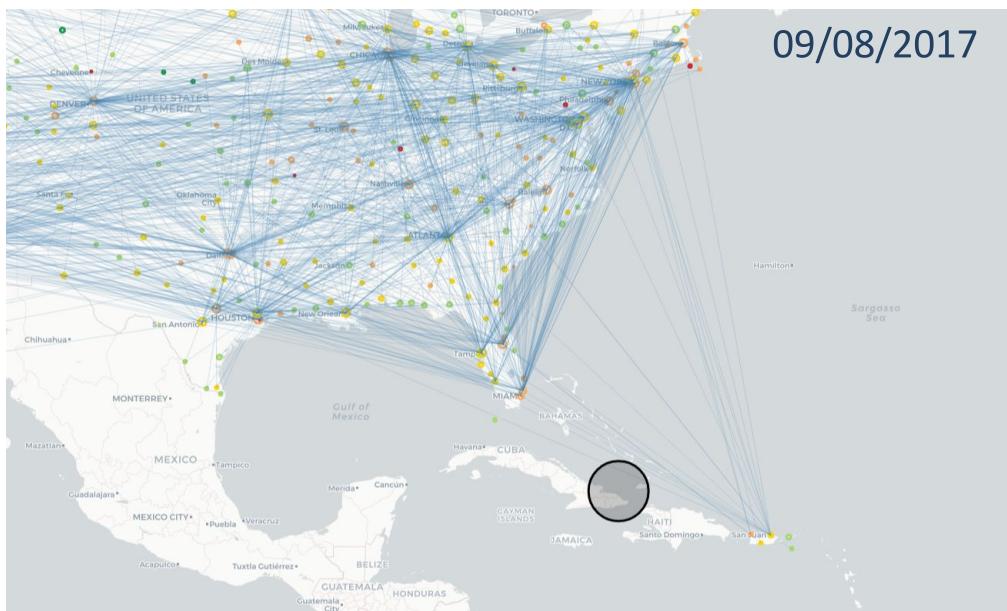
Manuel Klein
01/15/2020

37

→ neue fische
School and Pool for Digital Talent

Analyzing and Predicting US Flight Delays

09/08/2017



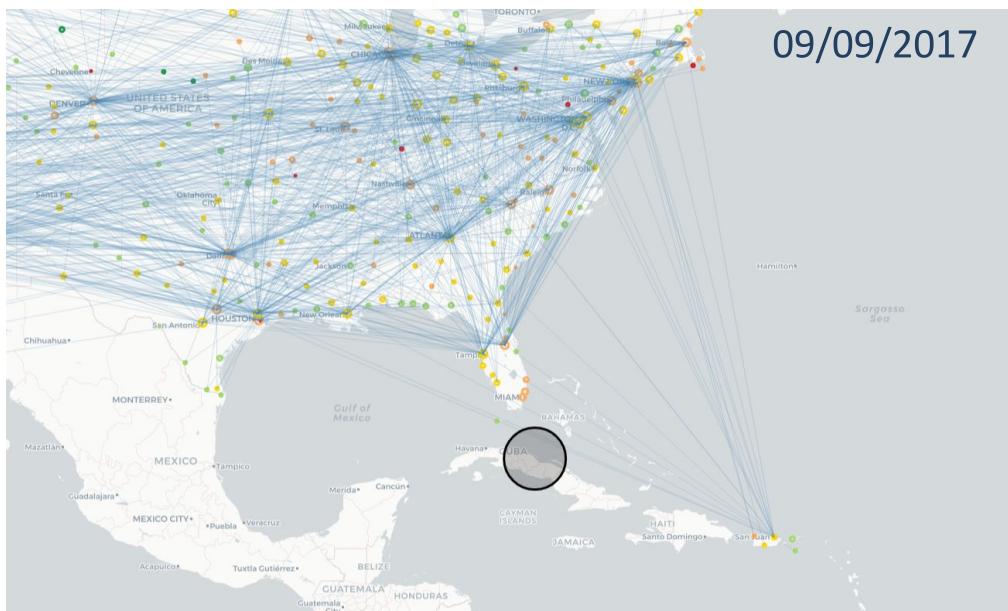
Manuel Klein
01/15/2020

38

→ neue fische
School and Pool for Digital Talent

Analyzing and Predicting US Flight Delays

09/09/2017



Manuel Klein
01/15/2020

39

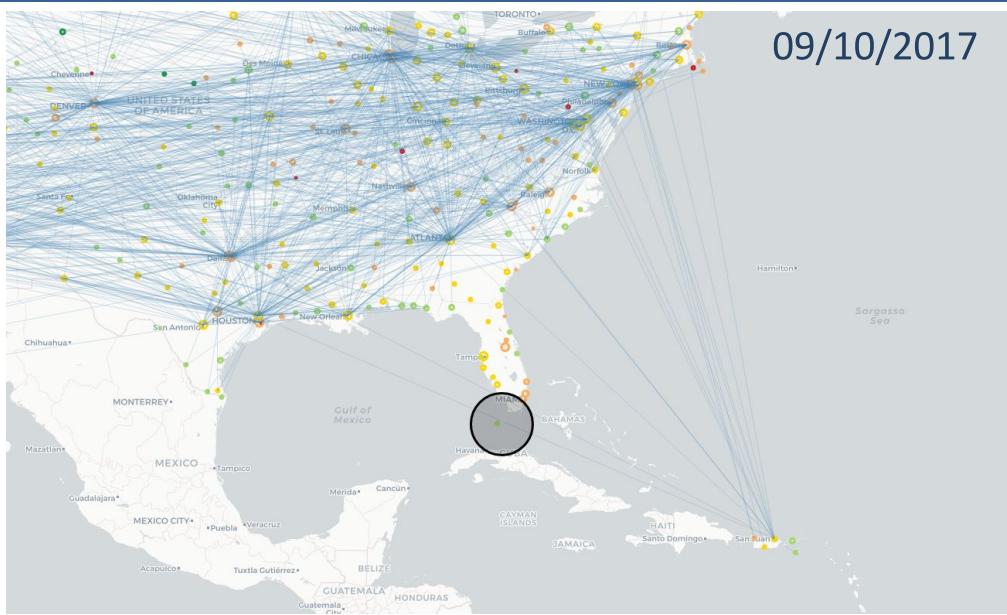
→ neue fische
School and Pool for Digital Talent

As the hurricane reaches the coast of Florida, airports are completely closed.

Analyzing and Predicting US Flight Delays



09/10/2017



Manuel Klein
01/15/2020

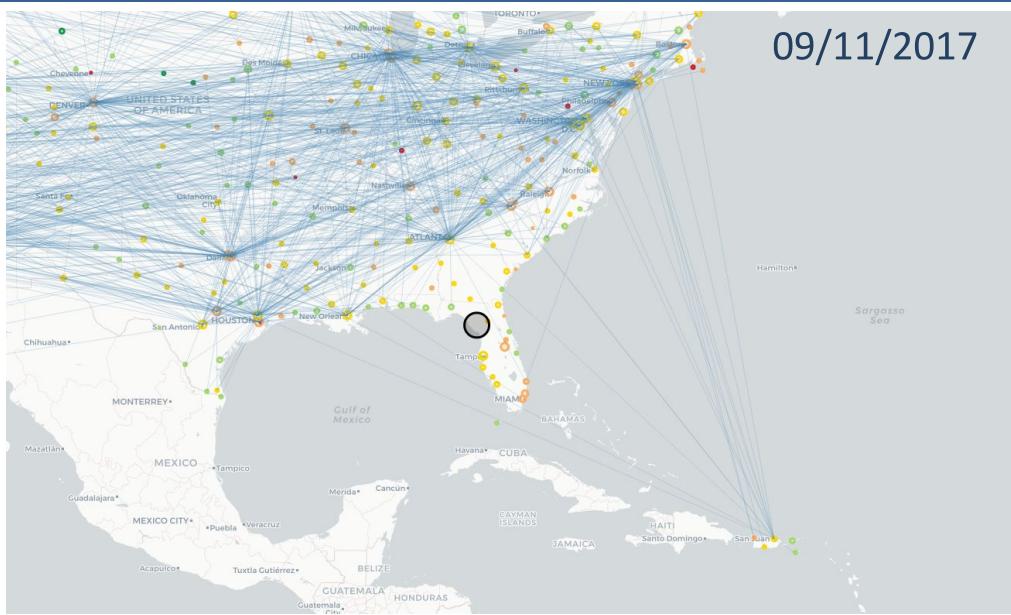
40

→ neue fische
School and Pool for Digital Talent

Analyzing and Predicting US Flight Delays



09/11/2017



Manuel Klein
01/15/2020

41

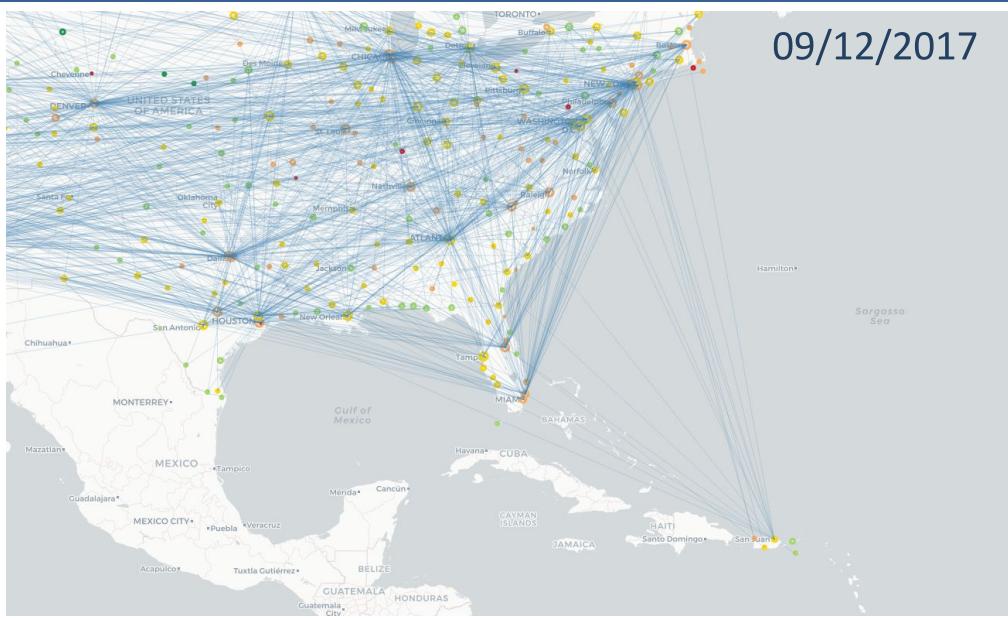
→ neue fische
School and Pool for Digital Talent

As soon as the hurricane hits the land, it becomes weaker rapidly.

Analyzing and Predicting US Flight Delays



09/12/2017



Manuel Klein
01/15/2020

42

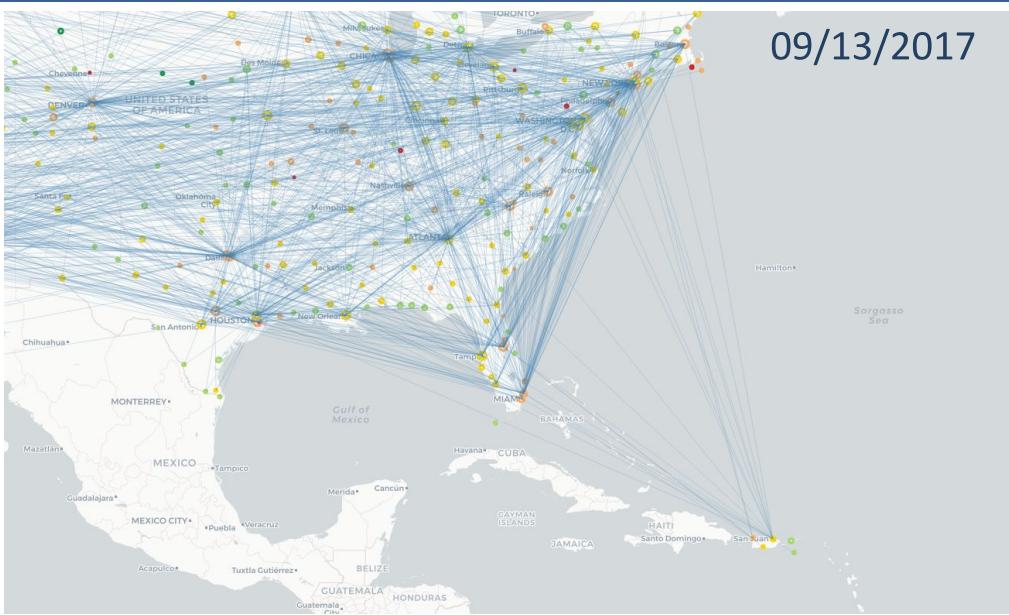
→ neue fische
School and Pool for Digital Talent

One day after the hurricane has disappeared, most flight routes are already operated again.

Analyzing and Predicting US Flight Delays



09/13/2017



Manuel Klein
01/15/2020

43

→ neue fische
School and Pool for Digital Talent



Aircraft Delay – Predictive Modeling

- A predictive model has been set up to predict arrival delays
- Model accuracy is limited (R^2 adjusted ~ 0.48)
- Delay reasons could not be used for prediction
 - Model based on implicit data

A linear regression model was used to set up a flight delay prediction model. The model was integrated into a demo to predict the amount of delay for any future US flight (see US flight delays demo Jupyter Notebook).



Aircraft Delay – Summary

- Avoid flying in the summer months
- Avoid flying on the business days
- Fly early in the morning if possible
- Choose an airline with a low delay score if possible
- Choose an airport / flight route with a low delay score if possible



Aircraft Delay – Future Work

- More detailed data cleaning
- Including features to improve prediction results
- Implementing more models and compare results



Questions?

» neue fische
School and Pool for Digital Talent

Thank you!



» neue fische
School and Pool for Digital Talent

Picture Source:

<https://static.independent.co.uk/s3fs-public-thumbnails/image/2019/04/30/19/fly.jpg>



Aircraft Delay – Appendix

Manuel Klein
01/15/2020

49

→ neue fische
School and Pool for Digital Talent



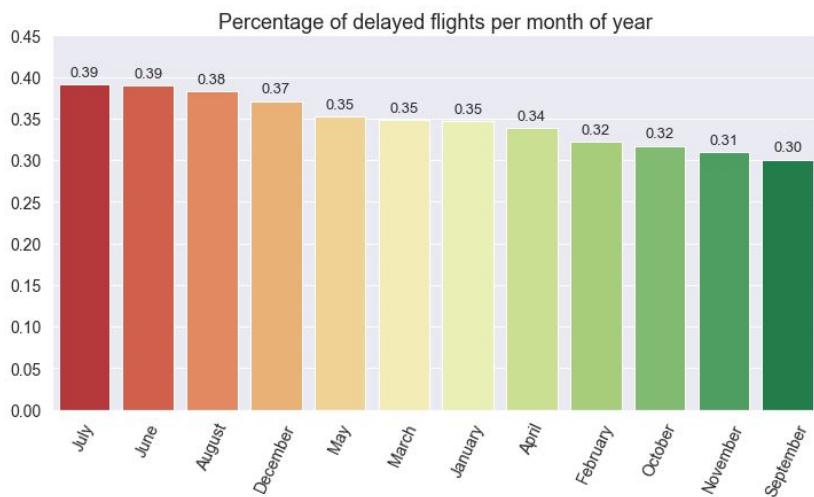
Aircraft Delay – Time Perspective

Percentage of delayed flights from 2016 to 2018 (Delay definition FAA: 15 minutes)



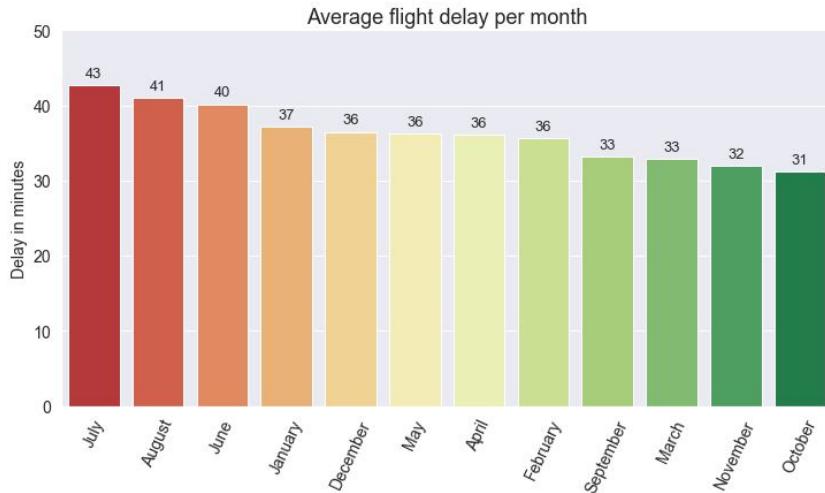


Aircraft Delay – Time Perspective



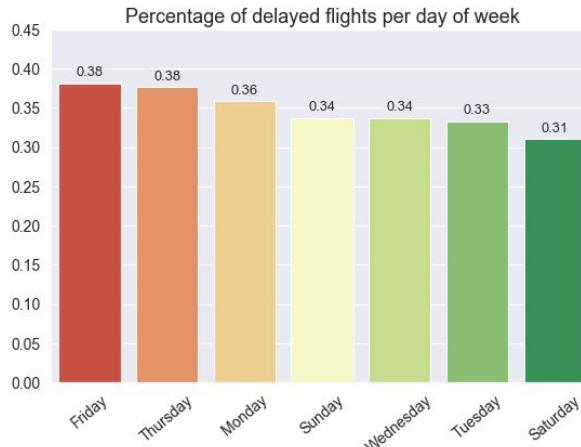


Aircraft Delay – Time Perspective



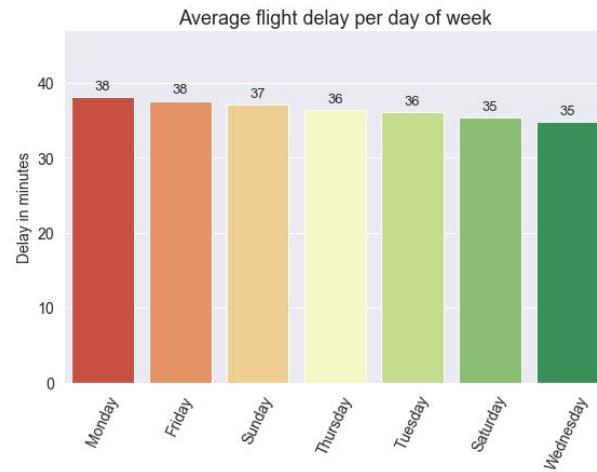


Aircraft Delay – Time Perspective





Aircraft Delay – Time Perspective



Manuel Klein
01/15/2020

→ neue fische
School and Pool for Digital Talent