Analysis of International Airline Ticket Prices

FROM norwegian

PREDICTING INT'L AIRFARE – SUMMARY

FLIGHT DATA

90
days starting

June 2, 2019

4,400+
flights

PRODUCTION

Predicting Airfare
via
Linear Reg. Model

Producing:

Flask

NORWEGIAN AIR FLIGHT ITINERARIES

I ✈ OAK

JFK

LAX
Los Angeles World Airports

CPH
Copenhagen Airport

Gatwick
LONDON AIRPORT

PARiS AÉROPORT

Stockholm
Arlanda Airport
SWEDAVIA SWEDISH AIRPORTS

Aeropuerto de Barcelona-El Prat

OVERVIEW – DATA COLLECTION

STANDARD PAGE FORMAT

Flights to Copenhagen from New York–JFK

Select departure and start reservation. All fares per person in USD. Restrictions, baggage fee and credit card surcharge

Outbound: New York-JFK - Copenhagen — Wednesday 5. Jun 2019

Flights to Copenhagen from Los Angeles

Select departure and start reservation. All fares per person in USD. Restrictions, baggage fee and credit card surchar

Outbound: Los Angeles - Copenhagen — Sunday 9. Jun 2019

WARNING:

Site structure sometimes changes….

[Notice only 3 price columns here]

DATA EXPLORATORY ANALYSIS

DATA SUMMARY

UNFILTERED
4,847 x 34

CLEANED
4,413 x 9

3,600
Connecting

813
Direct Flights

orig_port_coded	dest_port_coded	days_to_flight	depart_yr	depart_mo	depart_day	duration_total_min	stops	prices_lowfare_USD
0	0	1.0	2019	6	3	690	1	1172.9
0	0	1.0	2019	6	3	750	1	772.9
0	0	1.0	2019	6	3	760	1	910.9
0	0	1.0	2019	6	3	920	1	1172.9
0	0	1.0	2019	6	3	980	1	772.9
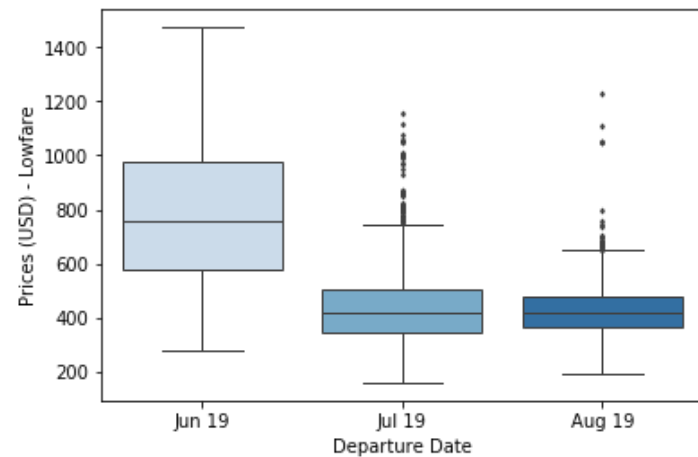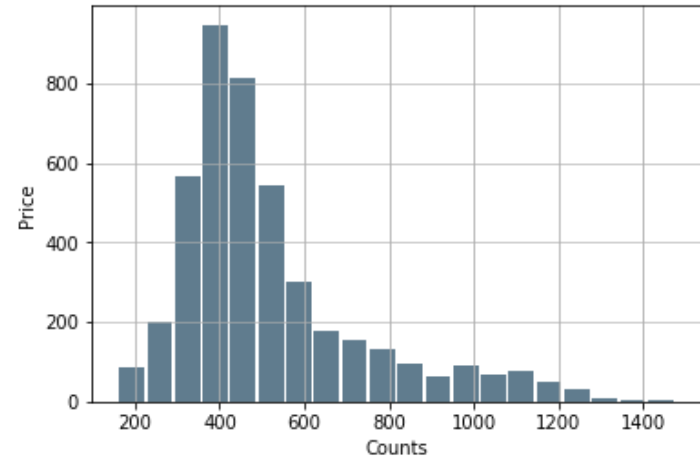
STOP LOCATION

London	**1,109**	Madrid	**225**
Oslo	**840**	Amsterdam	**129**
Stockholm	**532**	Paris	**101**
Copenhagen	**361**	Rome	**64**
Barcelona	**239**		

PRICES

PREDICTIVE LINEAR MODEL & FLASK API

**Airfare Price
(Lowfare USD)**

Destination Airport	Origin Airport
Days Until Flight	Departure Year
Departure Month	Departure Day
Total Duration (Minutes)	# Stops

LINEAR MODEL FEATURE DESIGN
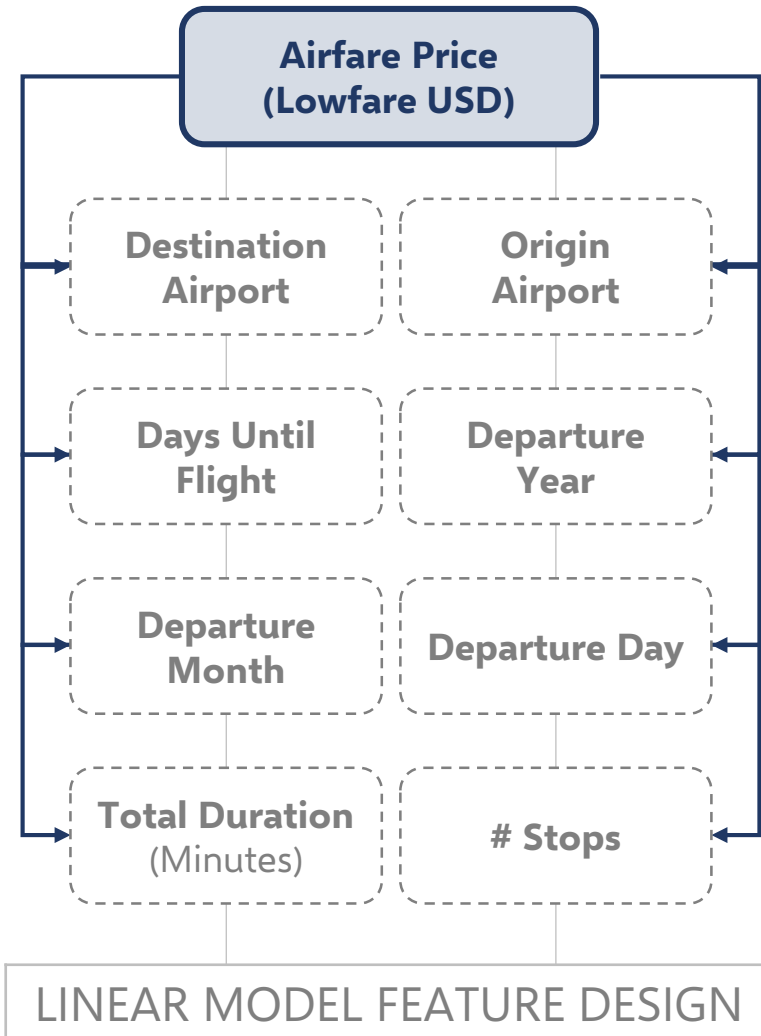
Flask + Amazon EC2

REQUEST USING `curl`

```
curl -H "Content-Type: application/json" -X POST -d
'{"orig_port_coded":"0","dest_port_coded":"0","days_to_flight":"10","depart_yr":"2019","depart_mo":"6","depart_da
y":"13","duration_total_min":"750","stops":"1"}' "http://ec2-54-202-68-148.us-west-
2.compute.amazonaws.com:5000/predict_airfare"
```

LAUNCHING FLASK API

```
[ec2-user@ip-172-31-27-175 docker]$ docker-compose up
Recreating docker_flask_1 ... done
Attaching to docker_flask_1
flask_1    | * Serving Flask app "server" (lazy loading)
flask_1    | * Environment: production
flask_1    |   WARNING: Do not use the development server in a production environment.
flask_1    |   Use a production WSGI server instead.
flask_1    | * Debug mode: on
flask_1    | * Running on http://0.0.0.0:5000/ (Press CTRL+C to quit)
flask_1    | * Restarting with stat
flask_1    | * Debugger is active!
flask_1    | * Debugger PIN: 690-542-898
flask_1    | 72.197.183.122 - - [03/Jun/2019 07:51:28] "GET / HTTP/1.1" 200 -
```

EXAMPLE: REQUEST PREDICTION

```
Amazon Linux AMI

https://aws.amazon.com/amazon-linux-ami/2018.03-release-notes/
[ec2-user@ip-172-31-27-175 ~]$ curl -H "Content-Type: application/json" -X POST -d '{"orig_port_coded":"0","d
est_port_coded":"0","days_to_flight":"10","depart_yr":"2019","depart_mo":"6","depart_day":"13","duration_tota
l_min":"750","stops":"1"}' "http://ec2-54-202-68-148.us-west-2.compute.amazonaws.com:5000/predict_airfare"
{
  "airfare prediction": 432.46229440441675
```

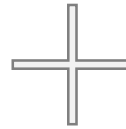# CHALLENGES & FUTURE IMPROVEMENTS

SITE STRUCTURE CHANGES
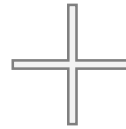
EXTREMELY SLOW SCRAPER
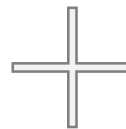
SEMI-CLUNKY CODE

IMPROVEMENTS

+ Add FX comparisons / access FX API

+ Run scraper daily at consistent times to capture changes in ticket prices

+ Get better at **python**

DISCLAIMER

All mentions of Norwegian (Norwegian Air Shuttle) and uses of logos, colors and font schemes are registered trademarks of Norwegian Air Shuttle and subject to all trademark and copyright laws.

All data presented in this analysis is the property of Norwegian Air Shuttle and used for personal academic and educational purposes only.