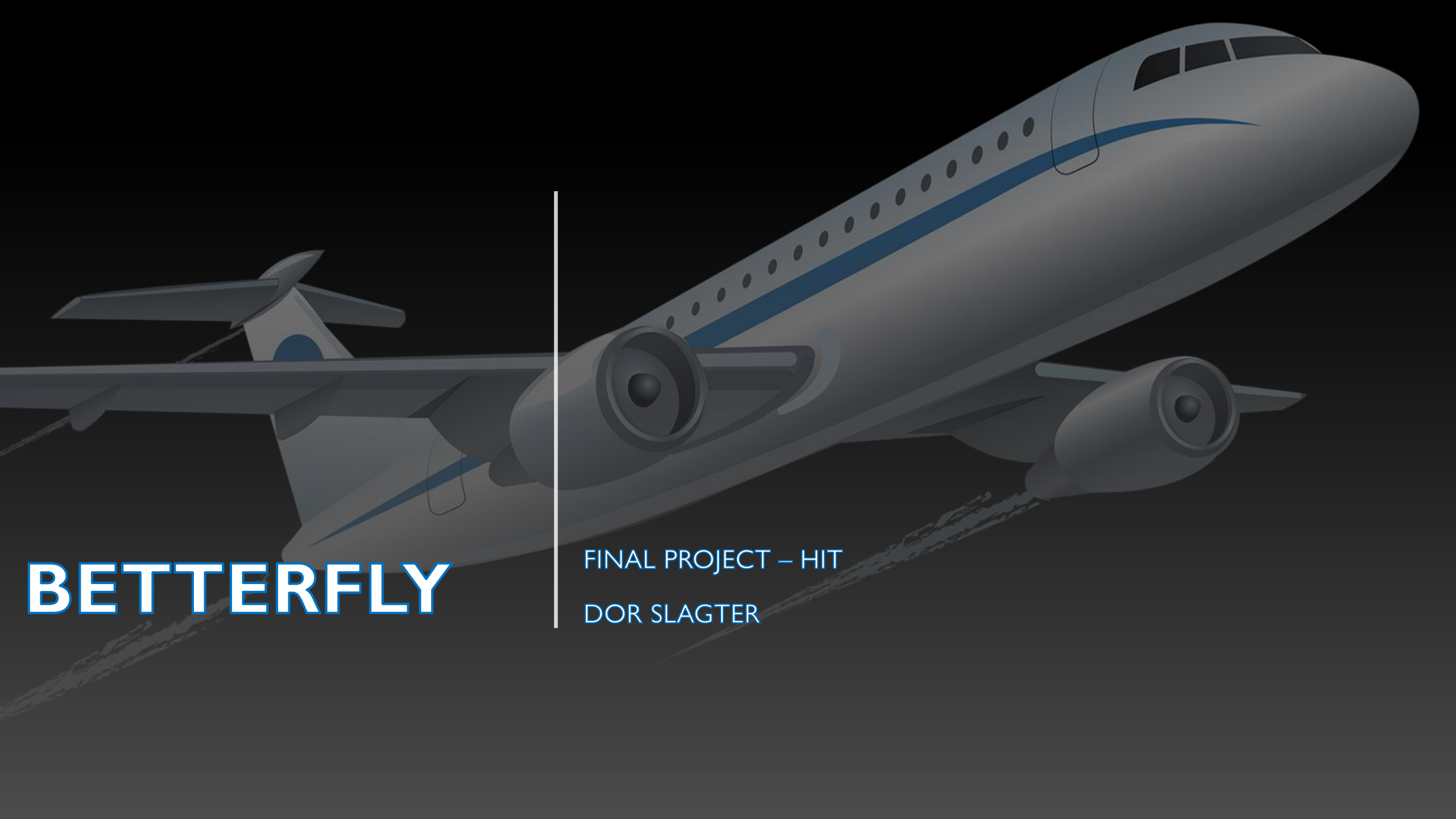


BETTERFLY

FINAL PROJECT – HIT

DOR SLAGTER

RESEARCH QUESTION

Is it possible to predict the price of a flight on a variety of sites like kayak using the characteristics of the flight and the seasons of the year?

RESEARCH QUESTION

Is it possible to predict the price of a flight on a variety of sites like kayak using the characteristics of the flight and the seasons of the year?

SCRAPING

SCRAPING TOOL

User Input

In [ ]:

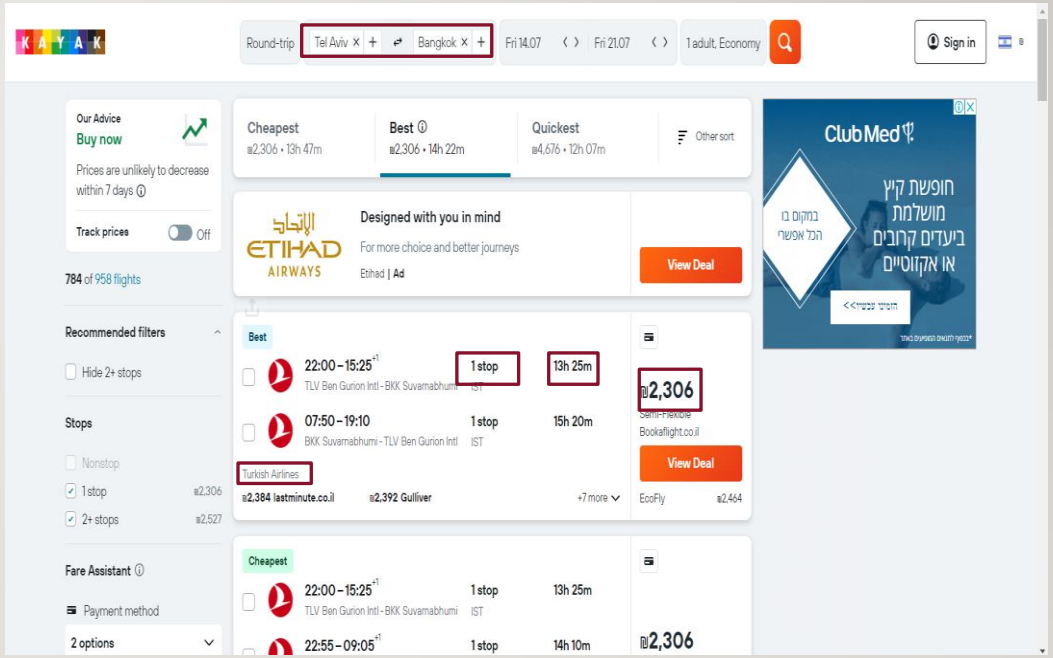
```
# get user input for routes
sources = []
destinations = []
print("Please enter -1 when done.")
print("-"*10)
while True:
    sources.append(input("From which city?\n"))
    if "-1" in sources:
        sources.pop(-1)
        break
    destinations.append(input("Where to?\n"))
    if "-1" in destinations:
        sources.pop(-1)
        destinations.pop(-1)
        break
    print("-"*10)

print("\nRoutes:")
for i in range(len(sources)):
    print(f"{sources[i]} => {destinations[i]}")
```

In [ ]:

```
# get user input for period (start and end date)
start_date = np.datetime64(input('Start Date, Please use YYYY-MM-DD format only '))
end_date = np.datetime64(input('End Date, Please use YYYY-MM-DD format only '))
days = end_date - start_date
num_days = days.item().days
```

SCRAPING

- Source
- Destination
- Date
- Price
- Duration
- Total stops
- Airline



SCRAPED ROUTES

GRU

BKK

TLV

LUX

PAR

MEX

YYZ

OSL

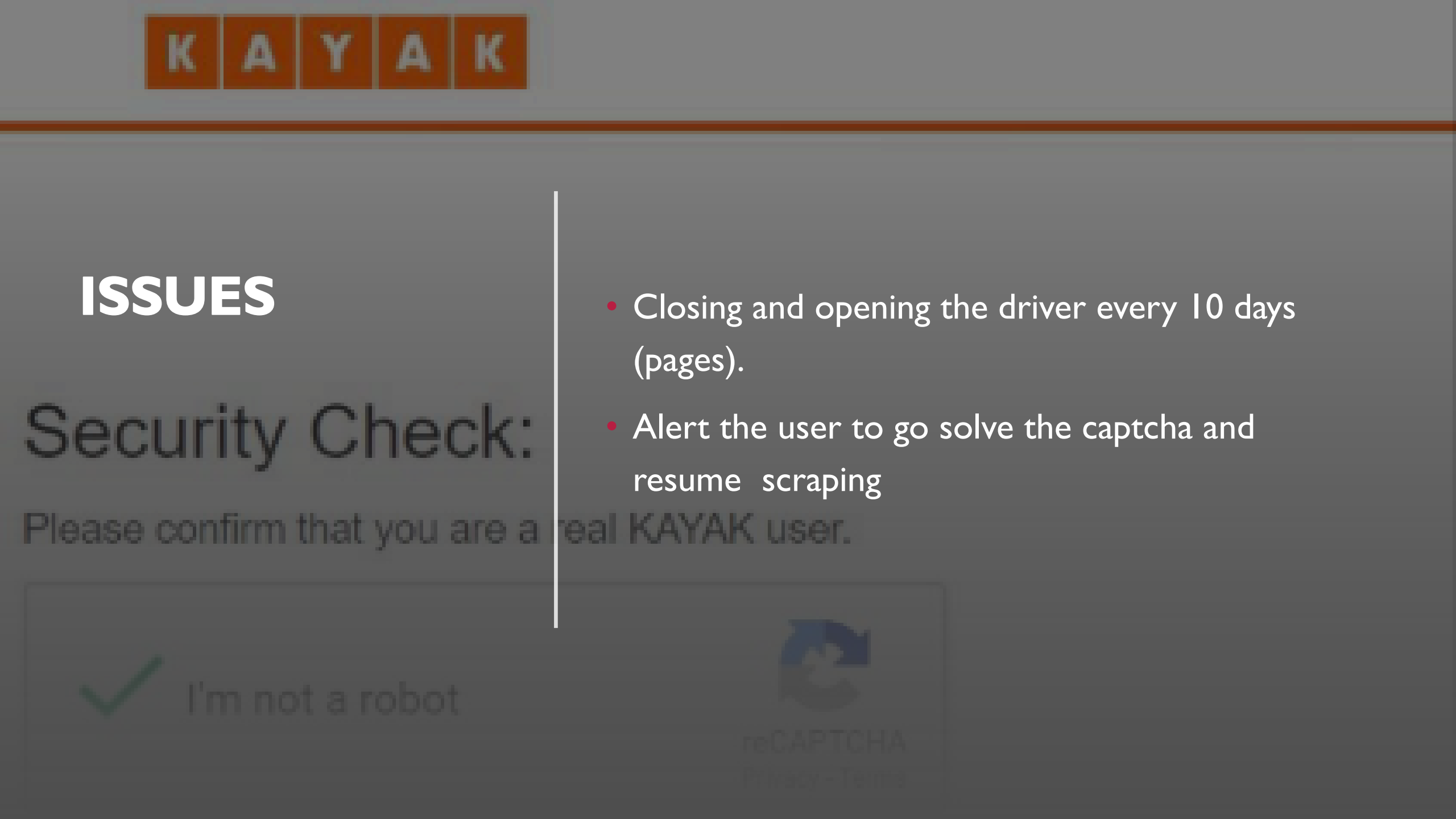
FCO

AMS

NYC

ISSUES

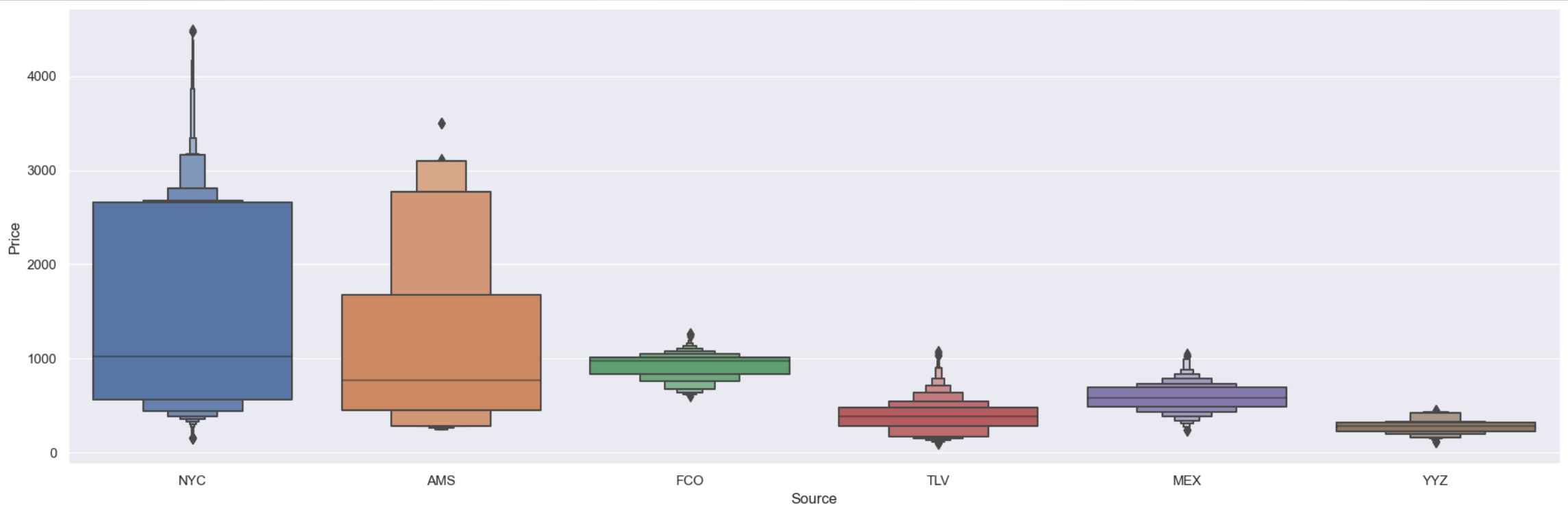
- Closing and opening the driver every 10 days (pages).
- Alert the user to go solve the captcha and resume scraping

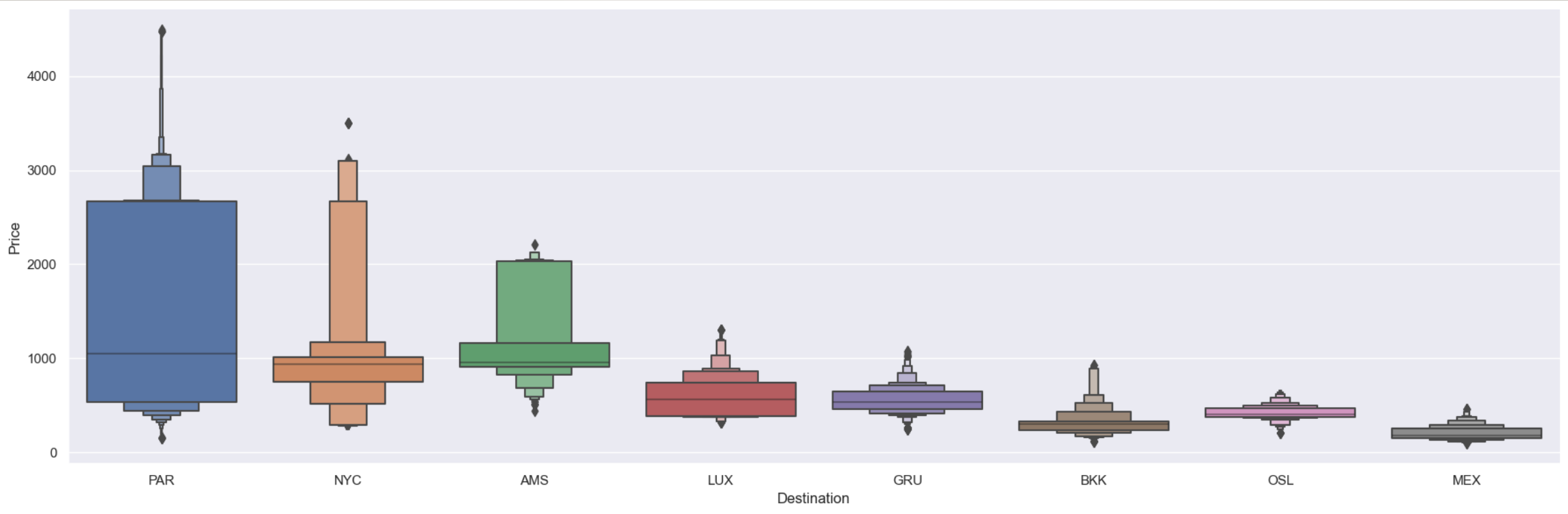


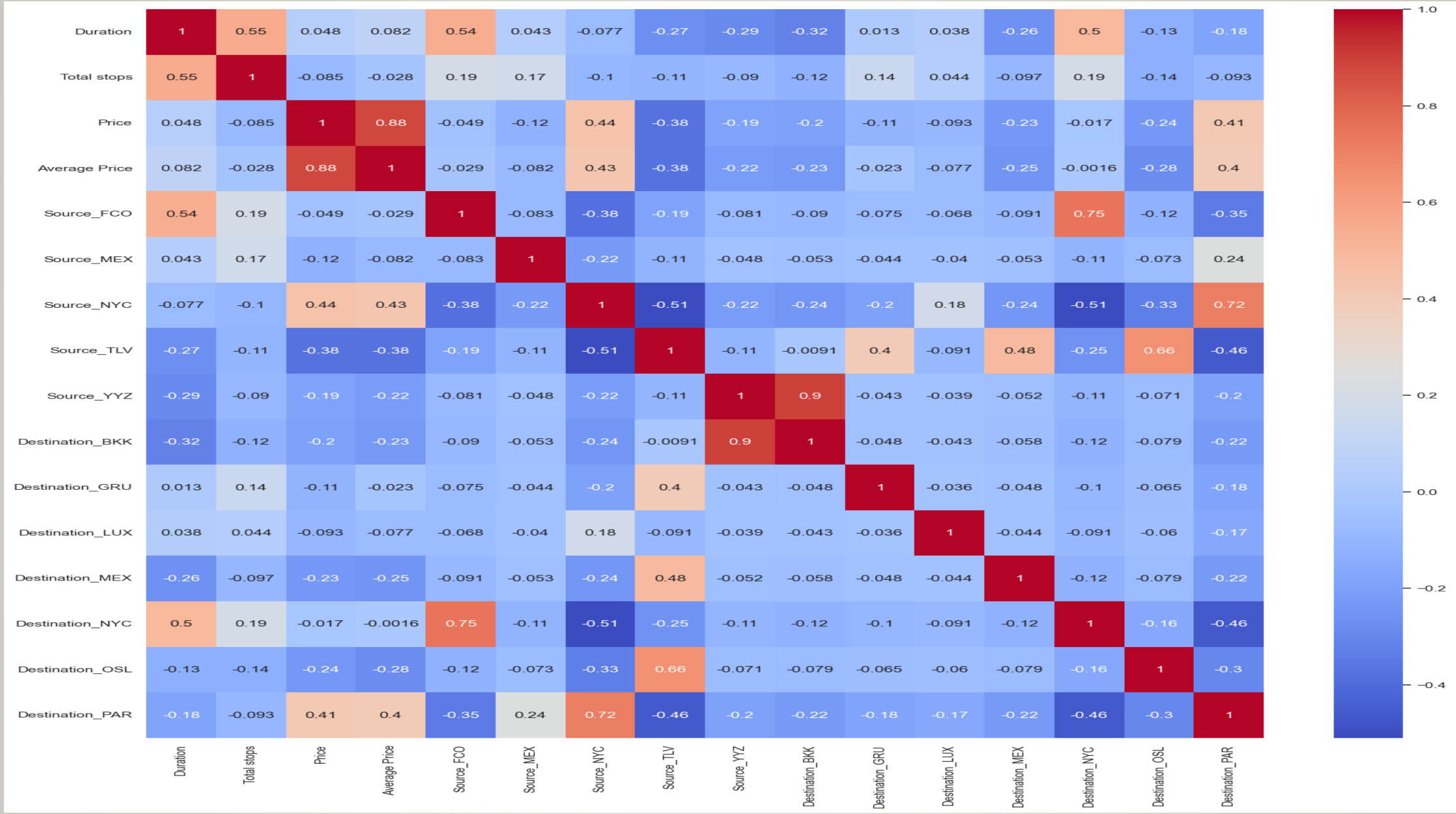
THE DATASET

	Airline	Source	Destination	Duration	Total stops	Price	Date	Average Price
0	Aeroflot	AMS	NYC	615	nonstop	70.69	2023-03-01	67.915656
1	Aeroflot	AMS	NYC	615	nonstop	66.06	2023-03-01	67.915656
2	Aeroflot	AMS	NYC	615	nonstop	66.06	2023-03-01	67.915656
3	Aeroflot	AMS	NYC	615	nonstop	70.69	2023-03-02	67.915656
4	Aeroflot	AMS	NYC	615	nonstop	66.06	2023-03-02	67.915656
...	...	...	...	...	...	...	...	...
2174	Air France, airBaltic	YYZ	BKK	655	3 stops	295.73	2023-04-09	283.559226
2175	Air France, Pobeda	YYZ	BKK	400	1 stop	286.67	2023-04-03	273.818786
2176	Transavia France, Finnair	YYZ	BKK	695	2 stops	202.13	2023-04-05	215.548002
2177	Norwegian, LOT	YYZ	BKK	825	2 stops	192.53	2023-04-09	251.441247
2178	SWISS, Pobeda	YYZ	BKK	520	2 stops	285.87	2023-04-11	672.917667

EDA


4000
3000
2000
1000
0
Price
NYC
AMS
FCO
TLV
MEX
YYZ
Source


Price
4000
3000
2000
1000
0
PAR
NYC
AMS
LUX
GRU
BKK
OSL
MEX
Destination


Duration 1 0.55 0.048 0.082 0.54 0.043 -0.077 -0.27 -0.29 -0.32 0.013 0.038 -0.26 0.5 -0.13 -0.18
Total stops 0.55 1 -0.085 -0.028 0.19 0.17 -0.1 -0.11 -0.09 -0.12 0.14 0.044 -0.097 0.19 -0.14 -0.093
Price 0.048 -0.085 1 0.88 -0.049 -0.12 0.44 -0.38 -0.19 -0.2 -0.11 -0.093 -0.23 -0.017 -0.24 0.41
Average Price 0.082 -0.028 0.88 1 -0.029 -0.082 0.43 -0.38 -0.22 -0.23 -0.023 -0.077 -0.25 -0.0016 -0.28 0.4
Source_FCO 0.54 0.19 -0.049 -0.029 1 -0.083 -0.38 -0.19 -0.081 -0.09 -0.075 -0.068 -0.091 0.75 -0.12 -0.35
Source_MEX 0.043 0.17 -0.12 -0.082 -0.083 1 -0.22 -0.11 -0.048 -0.053 -0.044 -0.04 -0.053 -0.11 -0.073 0.24
Source_NYC -0.077 -0.1 0.44 0.43 -0.38 -0.22 1 -0.51 -0.22 -0.24 -0.2 0.18 -0.24 -0.51 -0.33 0.72
Source_TLV -0.27 -0.11 -0.38 -0.38 -0.19 -0.11 -0.51 1 -0.11 -0.0091 0.4 -0.091 0.48 -0.25 0.66 -0.46
Source_YYZ -0.29 -0.09 -0.19 -0.22 -0.081 -0.048 -0.22 -0.11 1 0.9 -0.043 -0.039 -0.052 -0.11 -0.071 -0.2
Destination_BKK -0.32 -0.12 -0.2 -0.23 -0.09 -0.053 -0.24 -0.0091 0.9 1 -0.048 -0.043 -0.058 -0.12 -0.079 -0.22
Destination_GRU 0.013 0.14 -0.11 -0.023 -0.075 -0.044 -0.2 0.4 -0.043 -0.048 1 -0.036 -0.048 -0.1 -0.065 -0.18
Destination_LUX 0.038 0.044 -0.093 -0.077 -0.068 -0.04 0.18 -0.091 -0.039 -0.043 -0.036 1 -0.044 -0.091 -0.06 -0.17
Destination_MEX -0.26 -0.097 -0.23 -0.25 -0.091 -0.053 -0.24 0.48 -0.052 -0.058 -0.048 -0.044 1 -0.12 -0.079 -0.22
Destination_NYC 0.5 0.19 -0.017 -0.0016 0.75 -0.11 -0.51 -0.25 -0.11 -0.12 -0.1 -0.091 -0.12 1 -0.16 -0.46
Destination_OSL -0.13 -0.14 -0.24 -0.28 -0.12 -0.073 -0.33 0.66 -0.071 -0.079 -0.065 -0.06 -0.079 -0.16 1 -0.3
Destination_PAR -0.18 -0.093 0.41 0.4 -0.35 0.24 0.72 -0.46 -0.2 -0.22 -0.18 -0.17 -0.22 -0.46 -0.3 1
Duration
Total stops
Price
Average Price
Source_FCO
Source_MEX
Source_NYC
Source_TLV
Source_YYZ
Destination_BKK
Destination_GRU
Destination_LUX
Destination_MEX
Destination_NYC
Destination_OSL
Destination_PAR
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

RESULTS

RESULTS

model / metric	train(60%)	Val(20%)	MAE	MSE	RMSE
Linear regression	0.804	0.789	225.09	152995.68	391.15
Polynomial(5)	0.888	0.871	149.72	93366.04	305.56
Lasso	0.803	0.789	223.49	153097.06	391.28
Ridge	0.804	0.789	225.08	152995.61	391.15
Elastic Net	0.790	0.775	224.41	163034.63	403.78
Random Forest	0.965	0.945	61.72	40035.32	200.09

Actual VS Predicted