Flights Price Prediction

Ali Albayat – Mozah Alkhaldi

Introduction

Backstory

- New airline company want to operate its flight from different countries of Europe to Riyadh.
- Need to analyse the cost of existing airlines.
- Look at important factors on flight price.

Data Set & Tools

https://www.almosafer.com/en

- Scraping Data from Almosafer.
- 11 Airport to Riyadh.
- First week of December.
- 3986 rows .
- 13 column .

Tools

- Selenium
- BeautifulSoup
- SQL
- Pandas
- Matplotlib
- Seaborn
- Sklearn
- Numpy

Data Cleaning

01 **Cheack for NaN and duplicate**

02 **Drop unwanted columns**

DESTINATION
DATE
FLIGHT NUMBER

03 **Cheack for outliers**

Features Engineering

01

Convert categorical to numerical

Price
SAR1,808 → 1808
DURATION
09h 25m → 565

02

Convert time to AM\PM

Time of take off
Time of arrival
02:45pm → pm

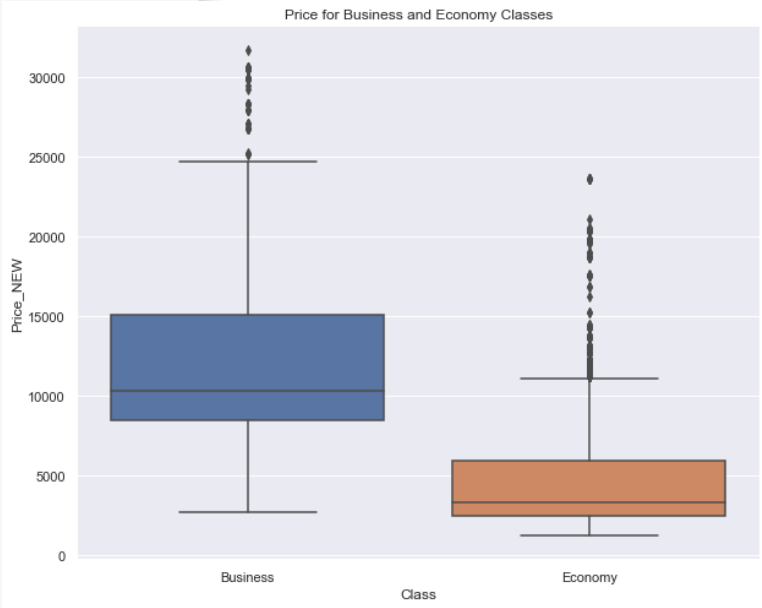
03

Convert day to day type

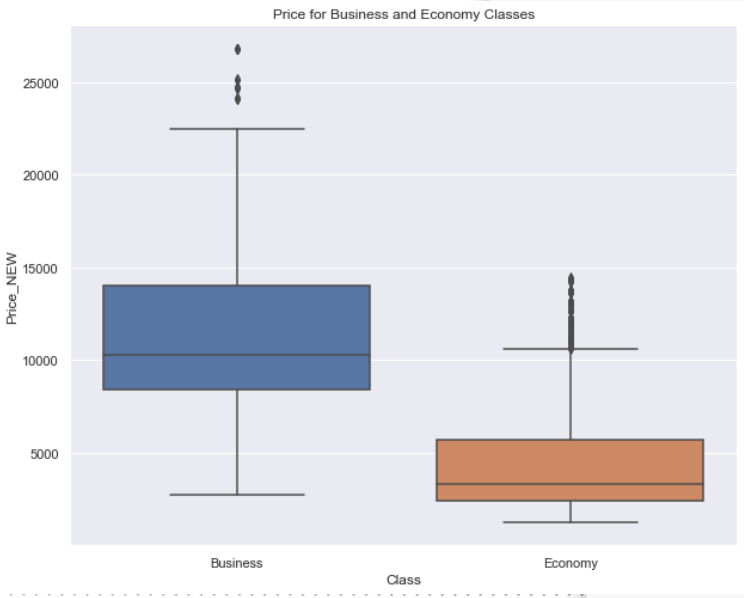
Wednesday → WeekDay
Friday → WeekEnd

DATA ANALYSIS

Outliers

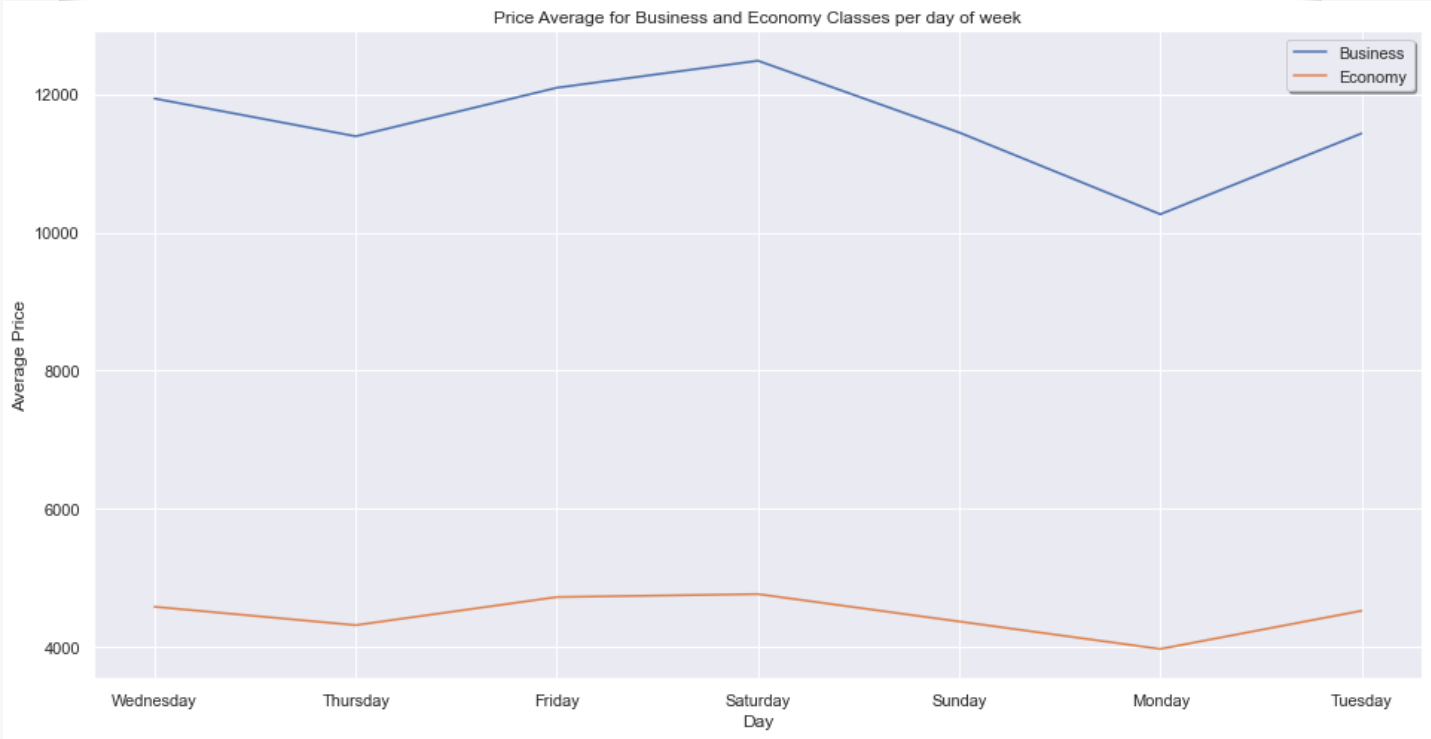


Before

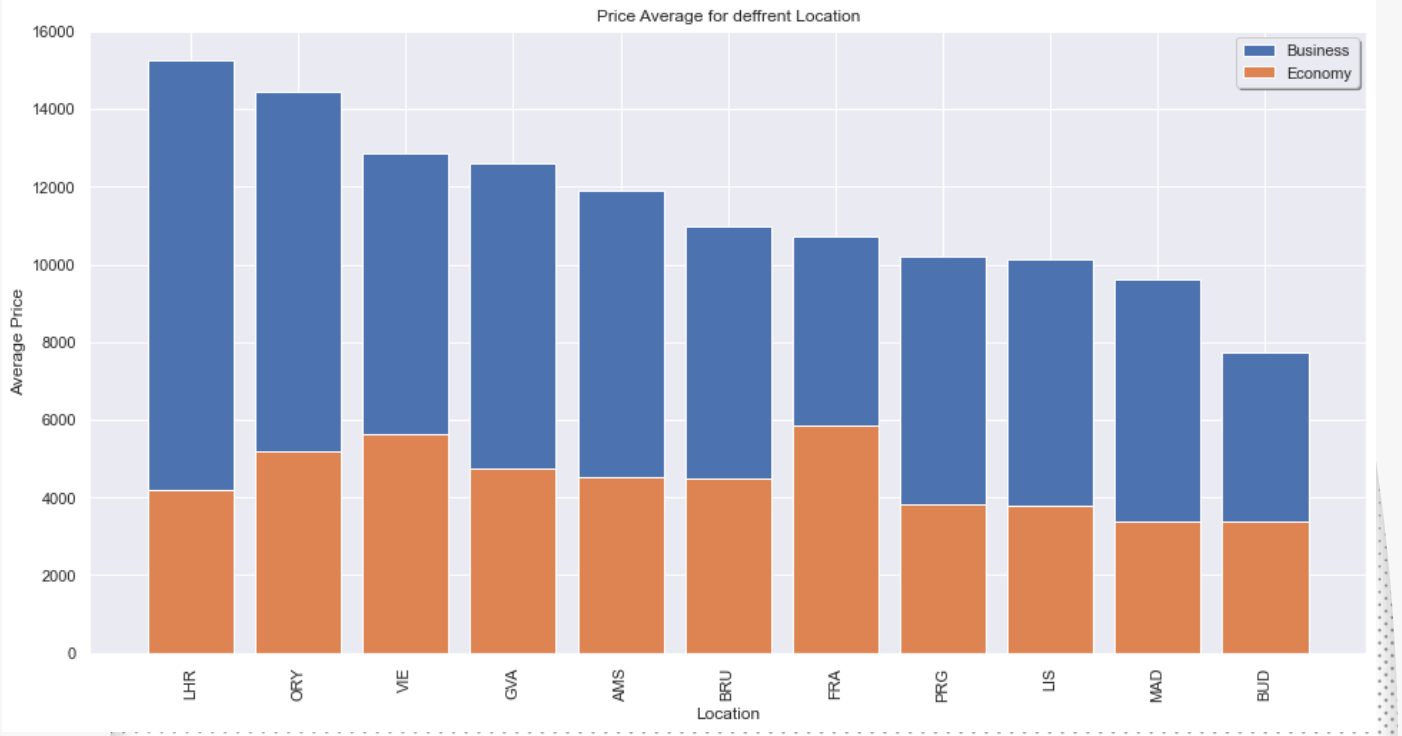


After

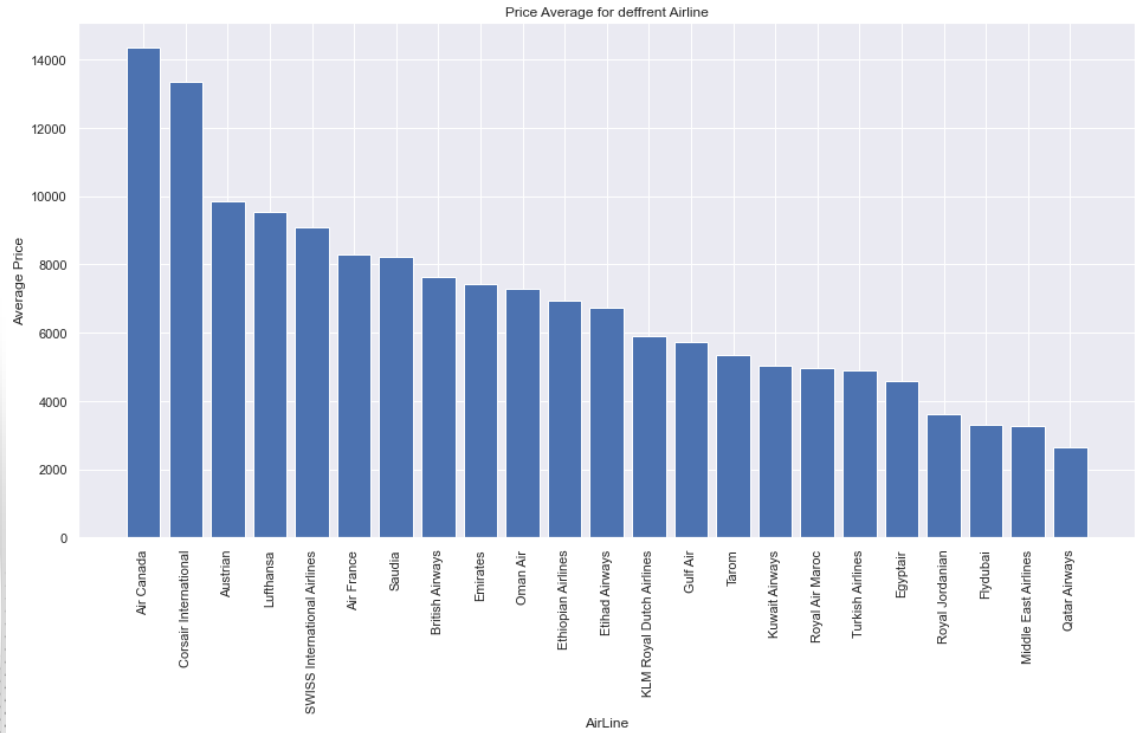
Average Price Per Day



Average Price Per Take-off Location



Average Price Per Airline



Regression Models

Trining

Validation

Testing

Regression Models

Model	Training	Validation	
Linear Regression	0.659	0.654	✓
Degree 2 polynomial Regression	0.827	-4911972.0	✗
Degree 3 polynomial Regression	0.921	-282.521	✗
Ridge Regression ($\lambda = 1$)	0.658	0.657	✓
Lasso Regression ($\lambda = 100$)	0.635	0.645	✓

Regression Models

Cross validation

Model	Validation
Linear Regression	0.64652
Ridge Regression ($\lambda = 1$)	0.64733
Lasso Regression ($\lambda = 100$)	0.54951

Regression Models

LassoCV

Training : 0.65585

Validation : 0.65749

λ : 5.6724

RidgeCV

Training : 0.65530

Validation : 0.65741

λ : 39.6268

Elastic NET

Validation : 0.65750

Regression Models

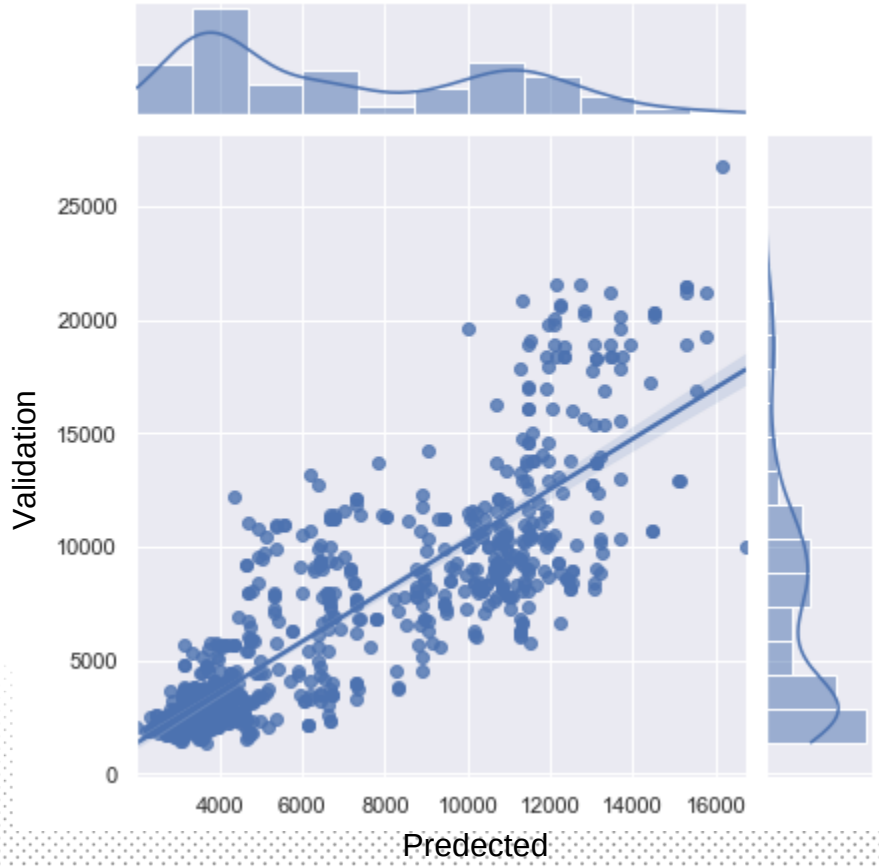
Random Forest

Validation : 0.43793

Gradient Boosting

Validation : 0.69350

Gradient Boosting



Gradient Boosting Test : 0.68298

Conclusion

THANK YOU FOR LISTINING