Predicting Flight Ticket Prices

Regression Project Presentation | Machine Learning CA1

EDUARDO J. MATOS ROMERO

20 NOVEMBER 2023

Goal

- To build a reliable machine learning model capable of accurately predicting flight ticket prices.

Steps to achieve it

- Exploratory Data Analysis of features.
- Modelling testing
- Hyperparameter Tuning
- Modelling
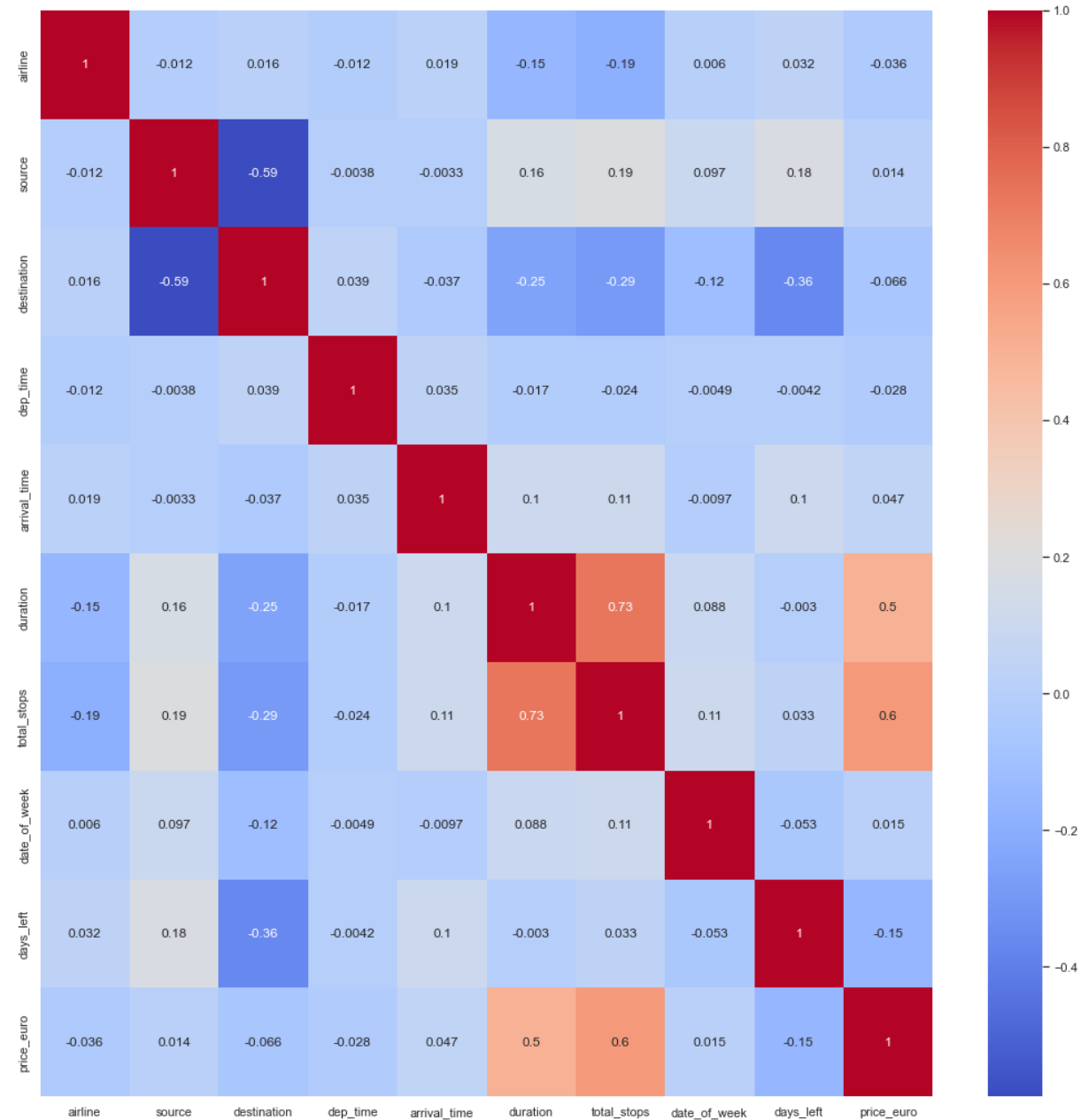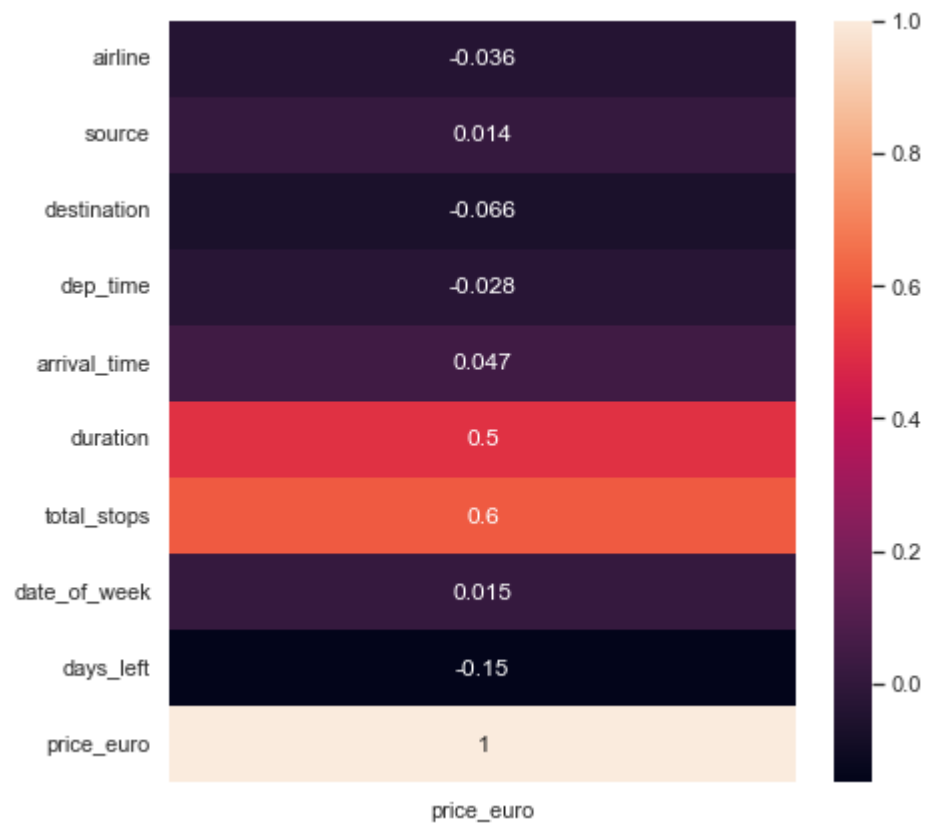
Exploratory Data Analysis

- Tickets prices are in general:
    - More expensive if you fly on a Friday.
    - Cheaper if you fly on Monday.

- Tickets prices are in general:
    - More expensive if booked on the same day.
    - Cheaper if booked more than 30 days in advance.

Days Left For Departure vs Ticket Price
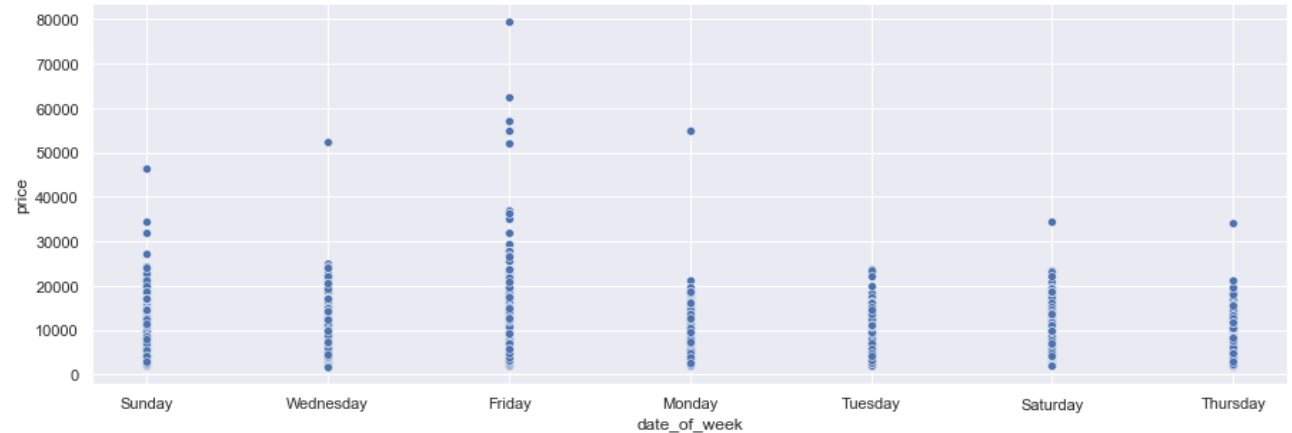
Exploratory Data Analysis

- Features

Modelling Testing

80 - 20%

Model	Train Score	Mean Squared Error	R-squared	Mean Absolute Error
Linear Regression	0.483939	1268.64	0.453352	27.1936
Ridge Regression	0.479757	1262.83	0.455858	27.0855
Lasso Regression	0.379593	1416.3	0.389727	28.2187
Decision Tree Regressor	0.995216	499.741	0.784666	8.19945
Random Forest Regressor	0.930258	266.978	0.884961	10.166
Support Vector Regressor	0.410312	1312.39	0.4345	24.5688
K-Nearest Neighbours Regressor	0.815295	765.724	0.670056	17.5633

70 - 30%

Model	Train Score	Mean Squared Error	R-squared	Mean Absolute Error
Decision Tree Regressor	0.994739	479.743	0.821956	8.01483
Random Forest Regressor	0.929308	328.007	0.878269	9.6725

Best performance
- Random forest
- Decision tree
- Splitting 70-30%

90 - 10%

Model	Train Score	Mean Squared Error	R-squared	Mean Absolute Error
Decision Tree Regressor	0.994643	221.628	0.914388	6.9058
Random Forest Regressor	0.92791	229.46	0.911363	9.38698

Hyperparameter Tuning - Modelling

Before

Model	Train Score	Mean Squared Error	R-squared	Mean Absolute Error
Decision Tree Regressor	0.994739	479.743	0.821956	8.01483
Random Forest Regressor	0.929308	328.007	0.878269	9.6725

- Increasing in the training R-squared
    o Improved model fit.

- Decreasing in MSE
    o Better model's predictions.

- Decreasing in R-squared
    o Minor reduction in generalization performance.

After

Model	Train Score	Mean Squared Error	R-squared	Mean Absolute Error
Random Forest Regressor	0.972614	304.811	0.884481	9.02995

- Decreasing in MAE
    o Improved accuracy.

Results

	Predicted Price	Actual Price
0	76.476407	76.571
1	173.038760	166.485
2	192.940346	204.050
3	83.839941	86.152
4	43.219783	42.438
5	79.428250	79.519
6	114.144733	114.048
7	122.979821	127.556
8	65.434395	46.849
9	101.228138	107.734
10	85.554659	84.590
11	141.719825	141.878
12	112.023285	119.284
13	65.278671	68.145
14	87.815620	82.335

$\approx$

Mean error

14.52 €