

SUPERVISED MACHINE LEARNING:

Predicting prices
of flights in India
with regressions

By Tobias Törnblom



THE DATA SET

I found the data set at kaggle.com
Data was collected over 50 days, from
February 11th to March 31st, 2022.
It contained 300153 rows and 11 columns.

Out of these, 9 became my features.
My target: To predict the price of tickets



airline - the name of the airline ex. *AirAsia*

flight - the name of the airplane ex. *I5-764*

source_city - Departure city ex. *Delhi*

departure_time - When did it depart ex. *Early_Morning*

stops - How many stops on the way ex. *one*

arrival_time - When will the plane land ex. *Afternoon*

destination_city - Destination city ex. *Mumbai*

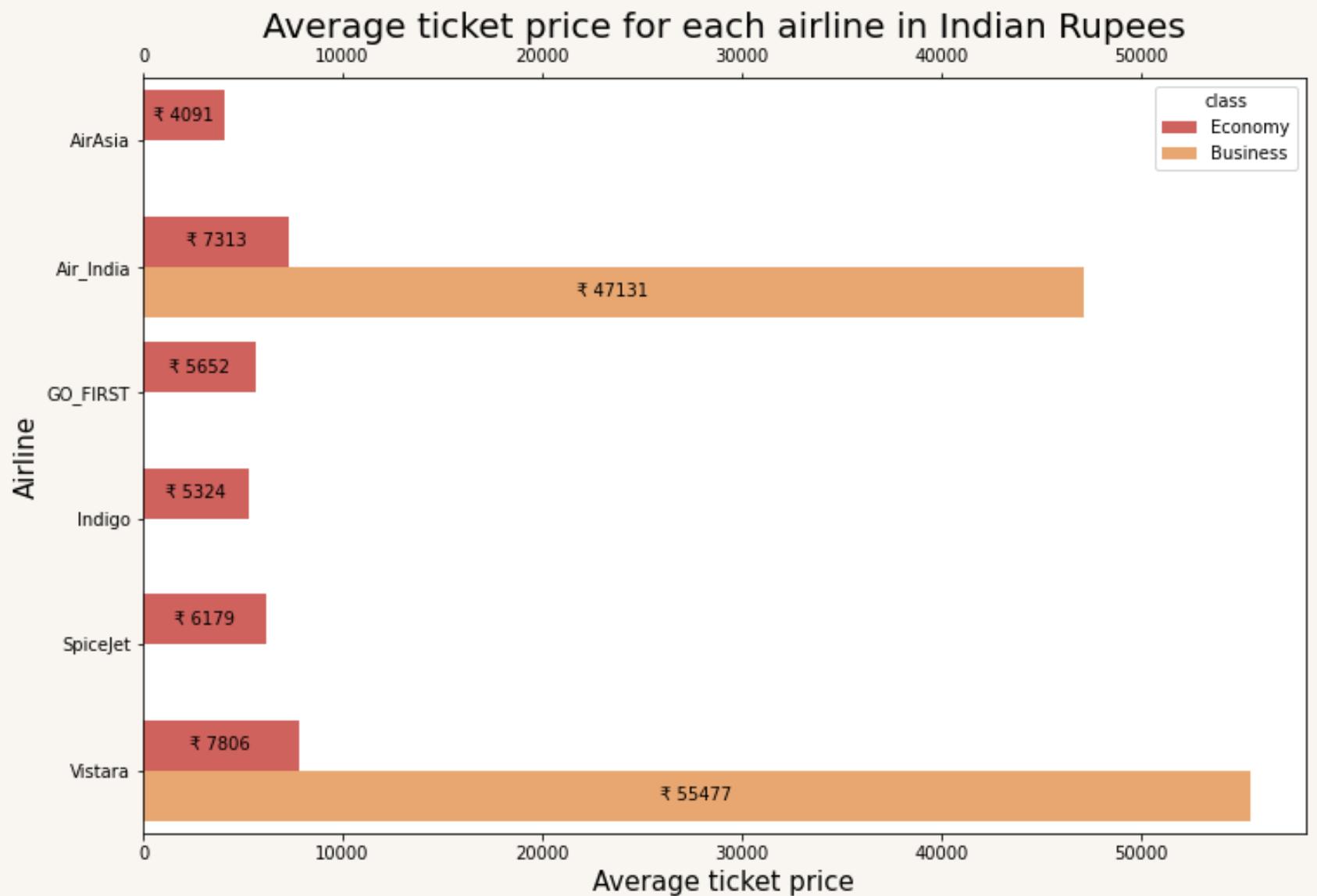
class - Business or Economy

duration - How many hours does the trip take ex. *2.5*

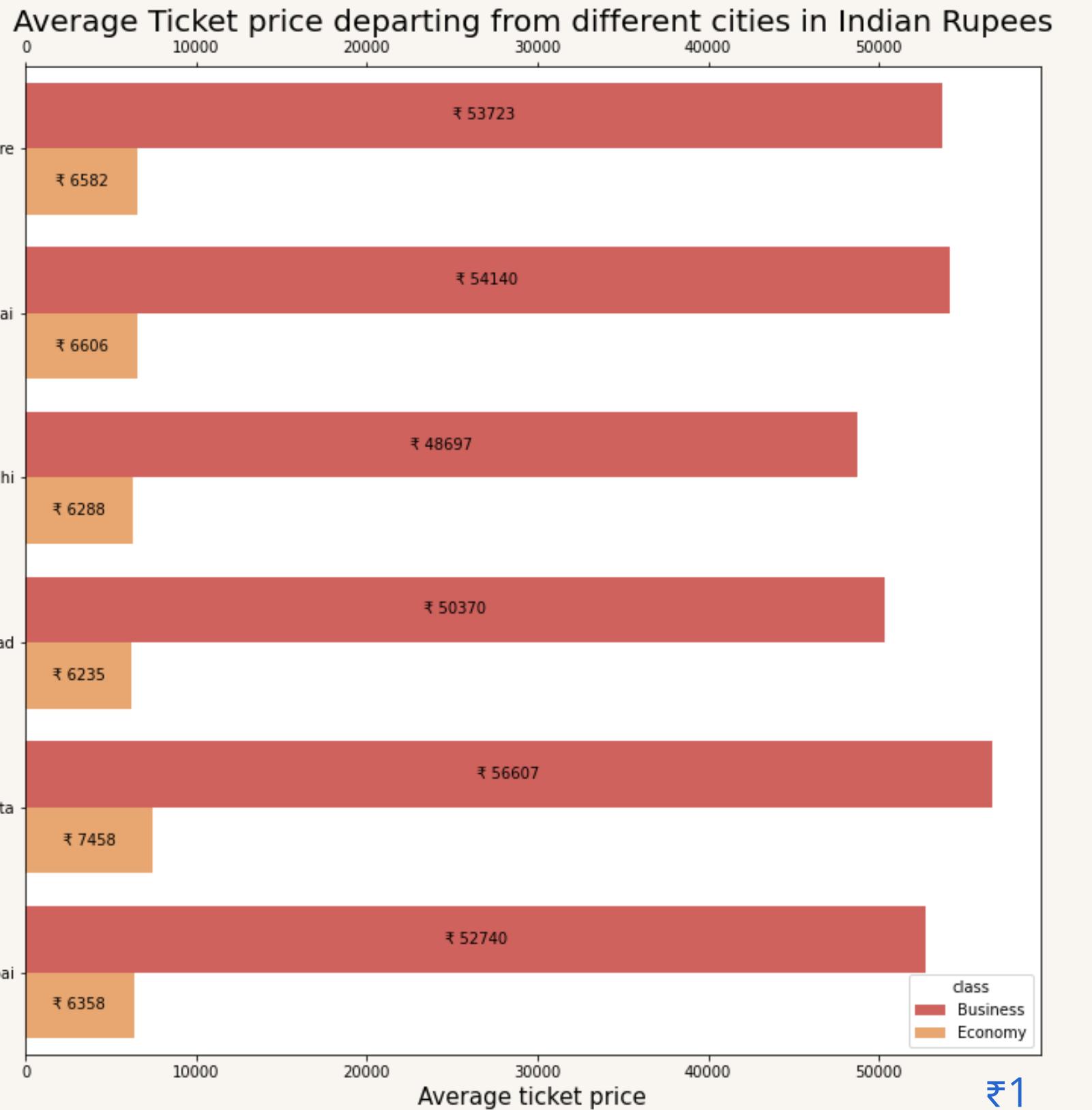
days_left - days until takeoff ex. *23*

price - price for one person ex. *20889*

Airlines



The cities



The data set contains the 6 most densely populated cities of India

New Delhi

Mumbai

Bangalore

Kolkata

Hyderabad

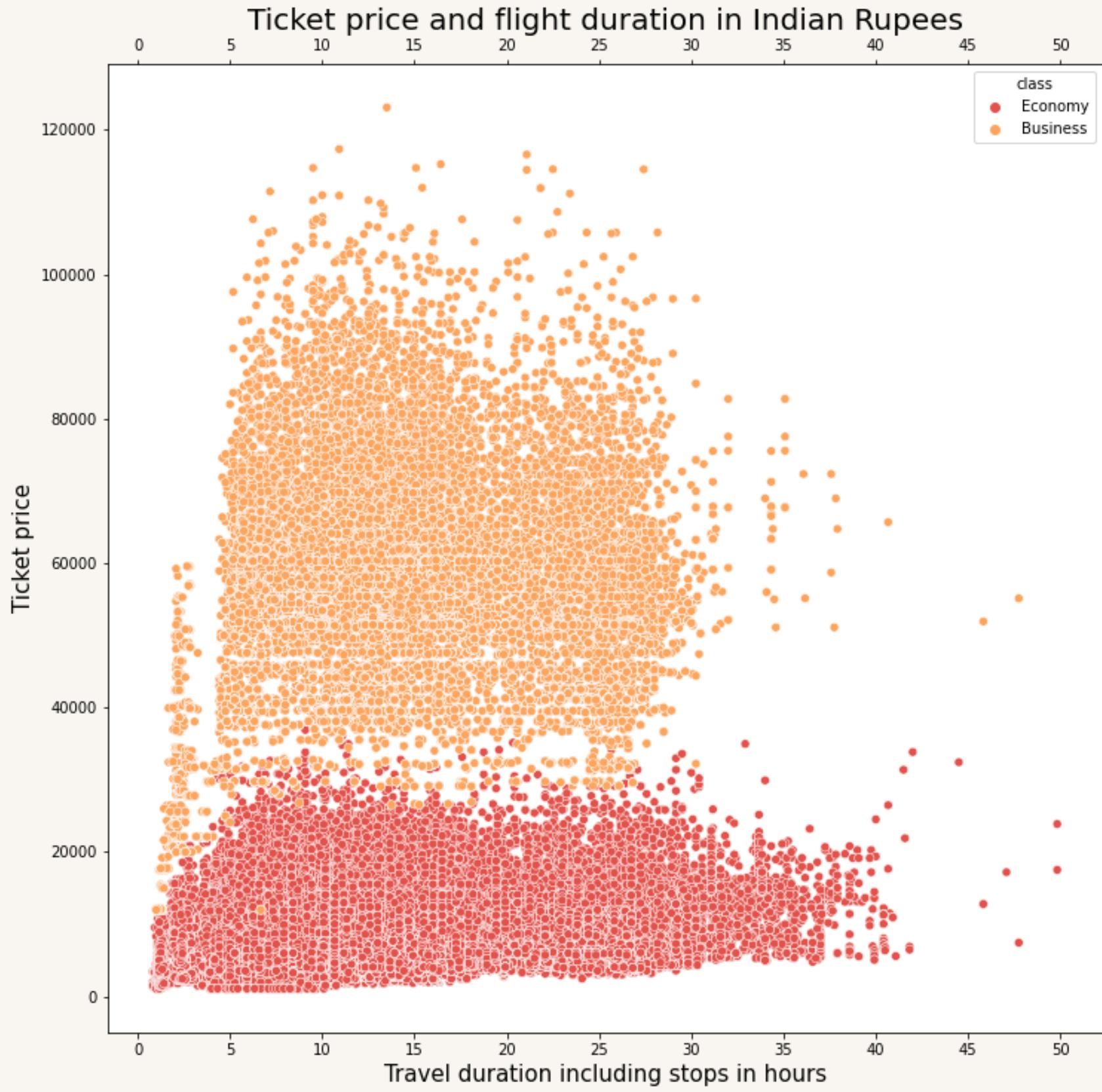
Chennai



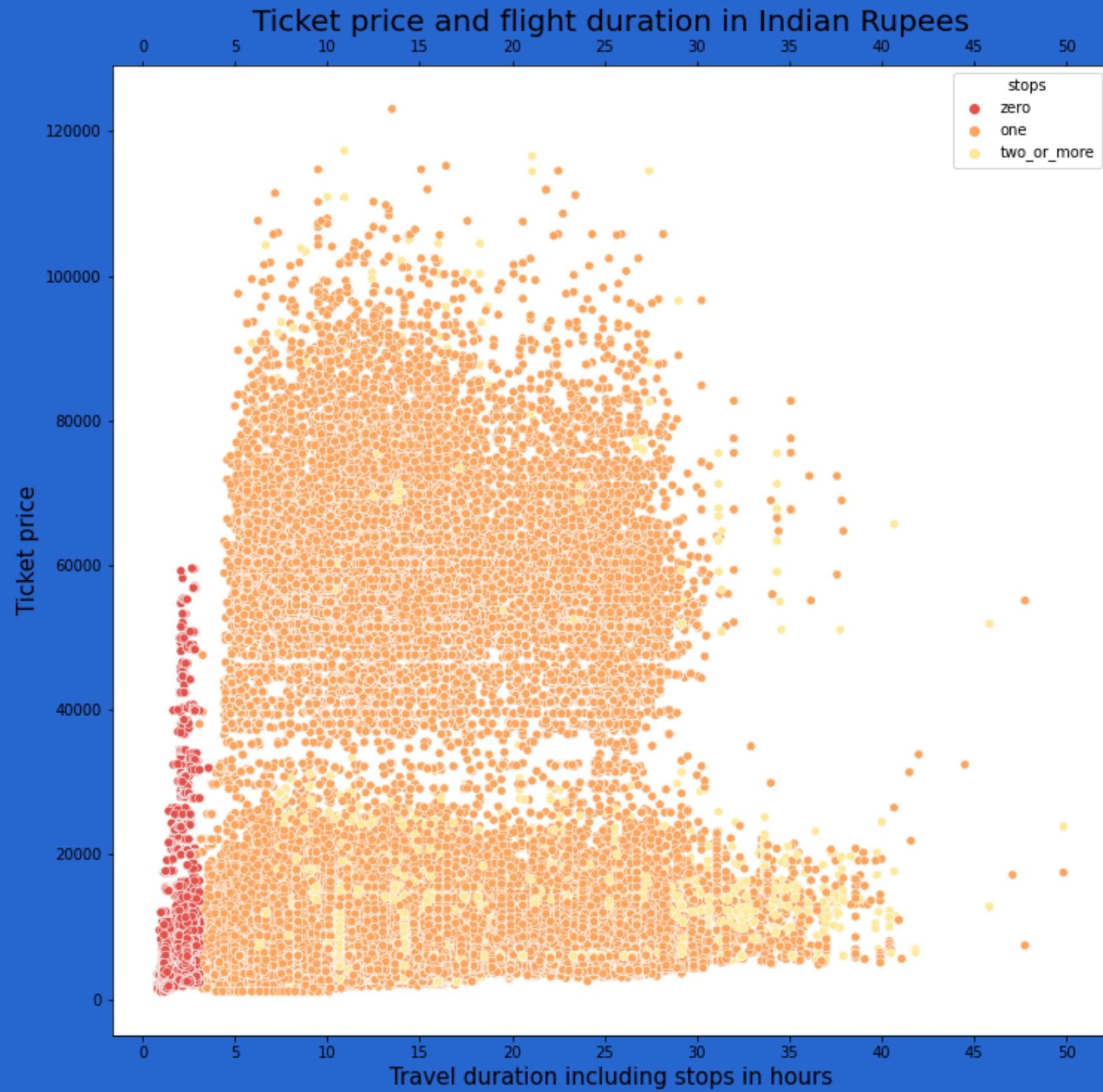
₹1

\$0.013
0.13Kr

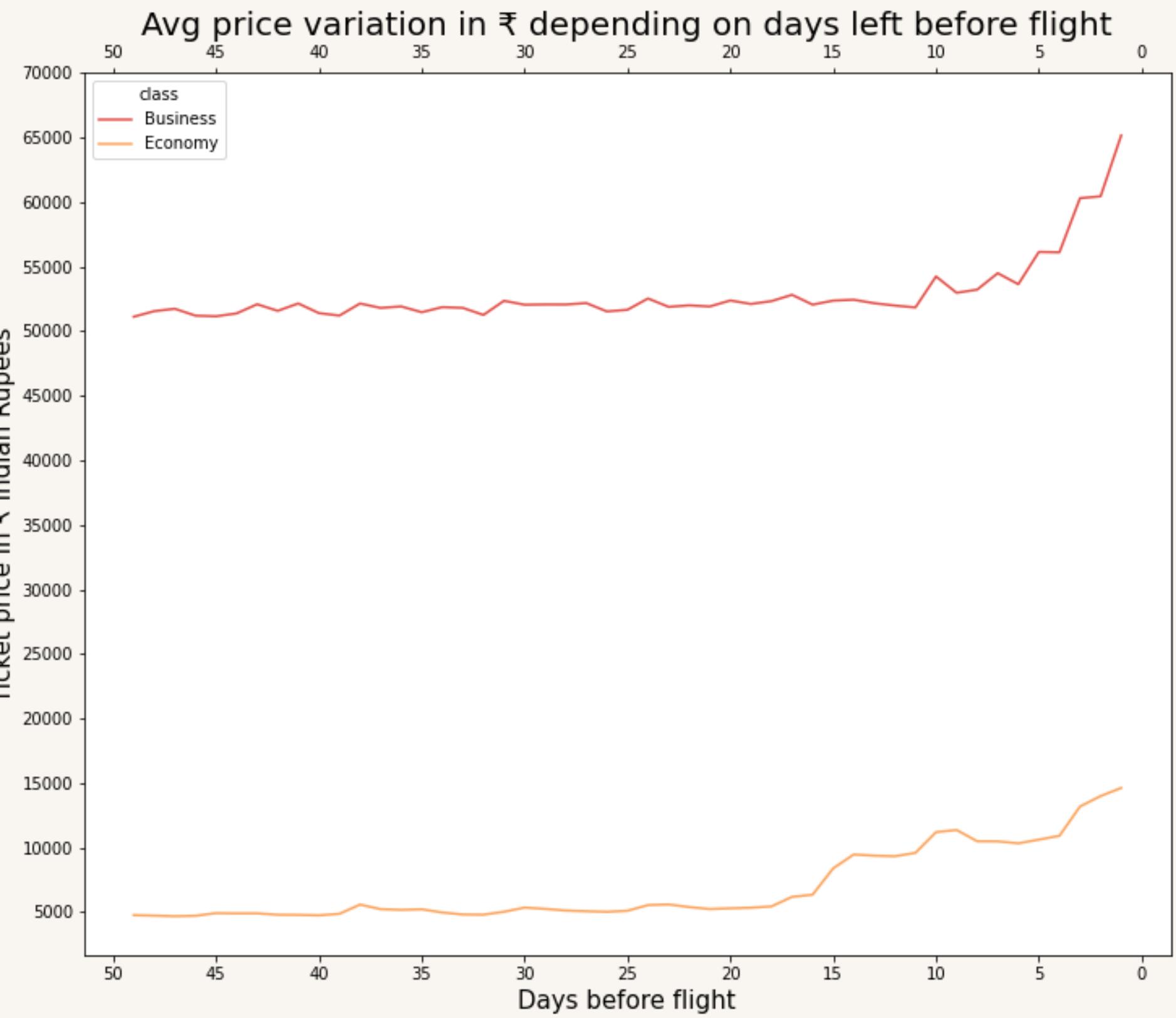
Class Business and Economy



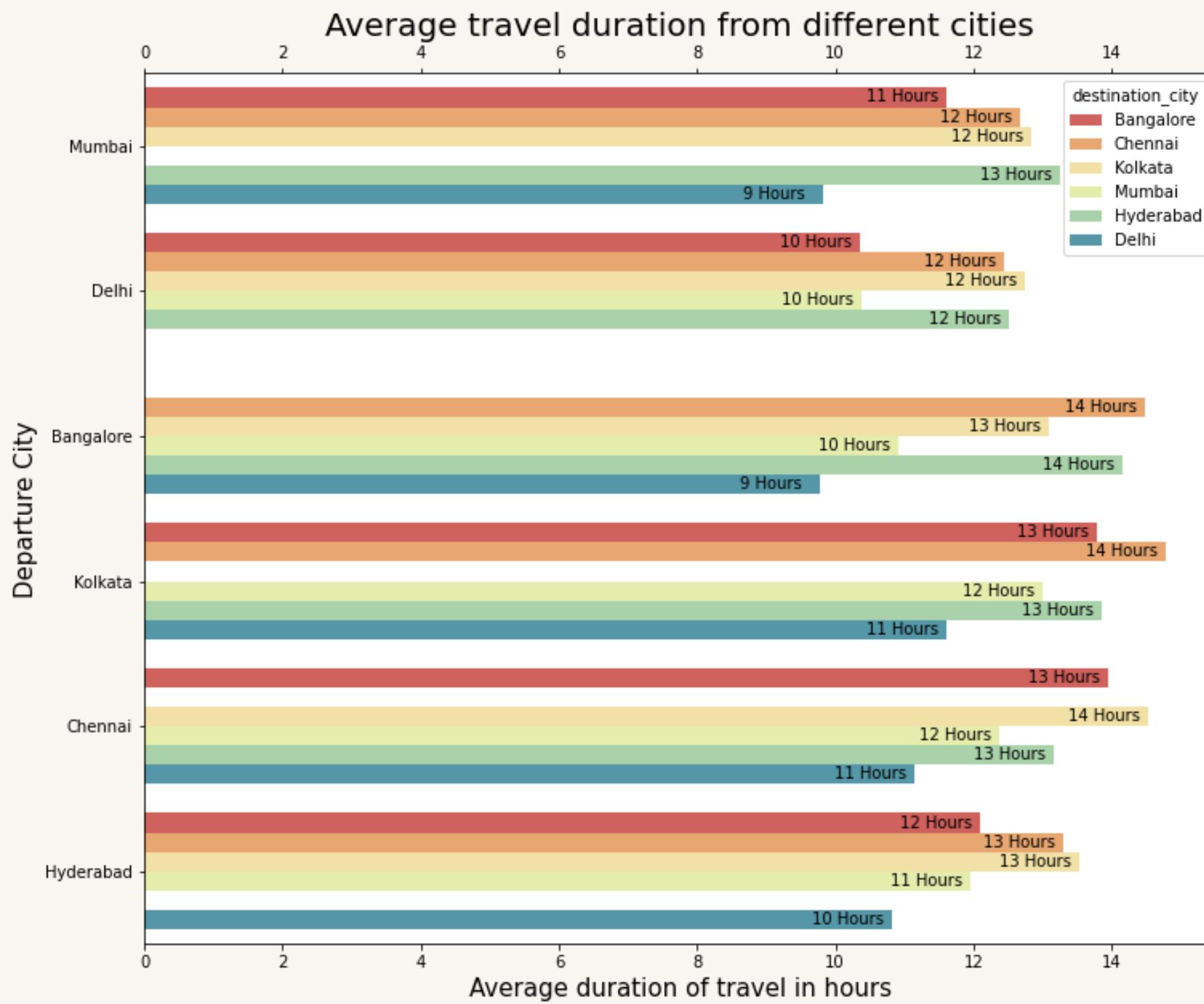
Number of layovers



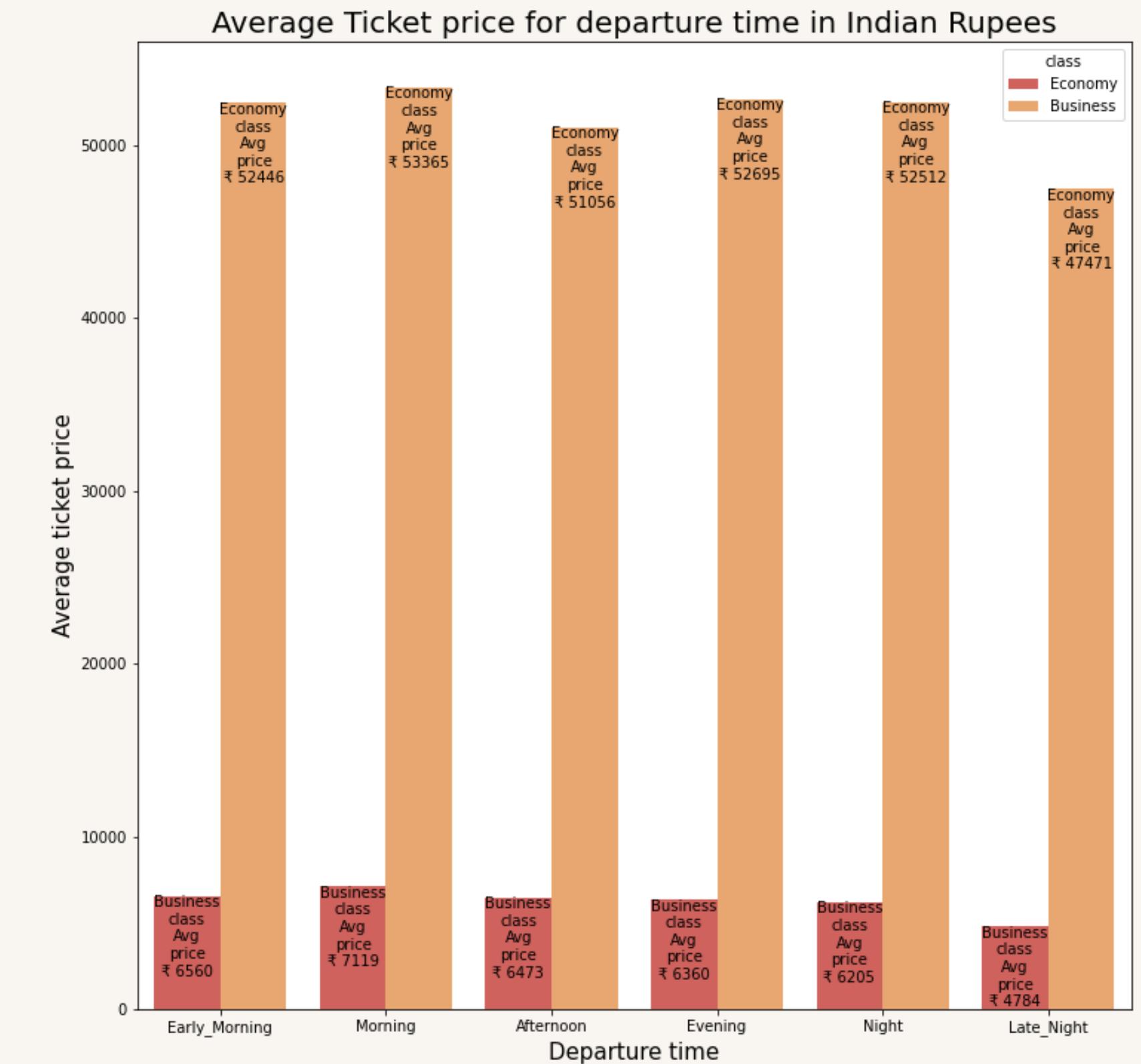
How days left until departure affects pricing



Average duration of flights



Price depending on departure time of day



**Predicting the ticket price
with a R2 baseline of 91%**

The R Squared of my model

With the RandomForestRegressor()

i got a R2 test score of:

98.1%



What can this be used for?

- Adjusting pricing of tickets to be competitive,
 - Finding holes in the market
- Planning when to buy tickets to save money
- Setting up new airlines

Challenges

Running the models with this big of a dataset and different hyperparameters took a long time.

Highlights

My plotting skills have evolved

