

A blurred photograph of a high-speed train, likely an AVE, moving through a station. The train is white with blue and grey accents. In the foreground, the dark silhouette of a person is visible on the left side of the frame.

Predicting High Speed Train Prices In Spain

By: Brian Jankowitz



Brian Jankowitz

Data Analyst

- Presented in front of audiences 300+ people
- Background in aviation, recruitment, and data analyzation

A photograph of a high-speed train, likely a Renfe AVE, traveling through a mountainous region. The train is blue and yellow, with a white stripe. It is moving on a set of tracks with overhead power lines. In the background, there are snow-capped mountains under a clear blue sky.

Agenda

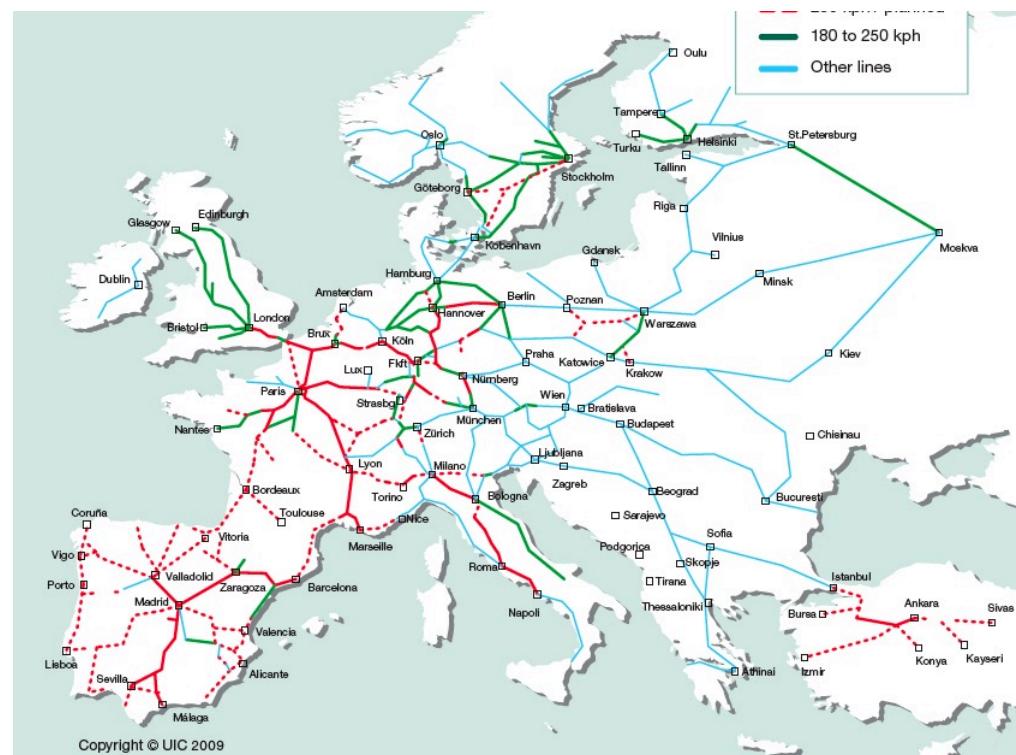
- Overview of High Speed Train Price in Spain
- Problem Statement
- The data
- Feature Engineering
- Models
- Model Selection

NEVER
GIVE UP
Because
GREAT THINGS
TAKE TIME

Keep moving
forward!

High Speed Train Overview

- 1,926 miles (3,100 km track)
- 193 mph (310km/h)
- AVE (Alta Velocidad Española) (Spanish High Speed)
 - Main train type for high speed train
- Barcelona
 - Get-away to other parts in Europe by high speed train going north
- Tickets sales are like airplane tickets
- It can be more convenient than flying





Problem Statement



- Buying a high speed train rail ticket can be just as tricky as buying an airplane ticket as prices fluctuate all the time based on multiple factors. The factors that make a high speed train ticket fluctuate can be because of the time, the supply, and the demand. Other factors that decide the cost of a ticket can be because of it's class and ticket type.
- Having prices vary so much, we are tasked with predicting the ticket prices of the high speed train this way the company can inform their customers when the best time to buy a train ticket is.



The data

OUTWARD JOURNEY

MADRID (TODAS) TO SEVILLA-SANTA JUSTA ↗

Date: 10 Diciembre 2015

Departure	Arrival	Duration	Train	Price From	Class	Fare	Options
06.20	09.16	2 h. 56 min.	AV City	26,70 € ►	Turista	Promo	<input checked="" type="checkbox"/>
07.00	09.30	2 h. 30 min.	AVE	52,85 € ►	Turista	Promo	<input checked="" type="checkbox"/>
07.30	09.50	2 h. 20 min.	AVE	85,20 € ►	Preferente	Promo +	<input type="checkbox"/>

Class Fare Seat

Preferente Promo + 85,20 € Selección de Asiento

Request Quiet Coach

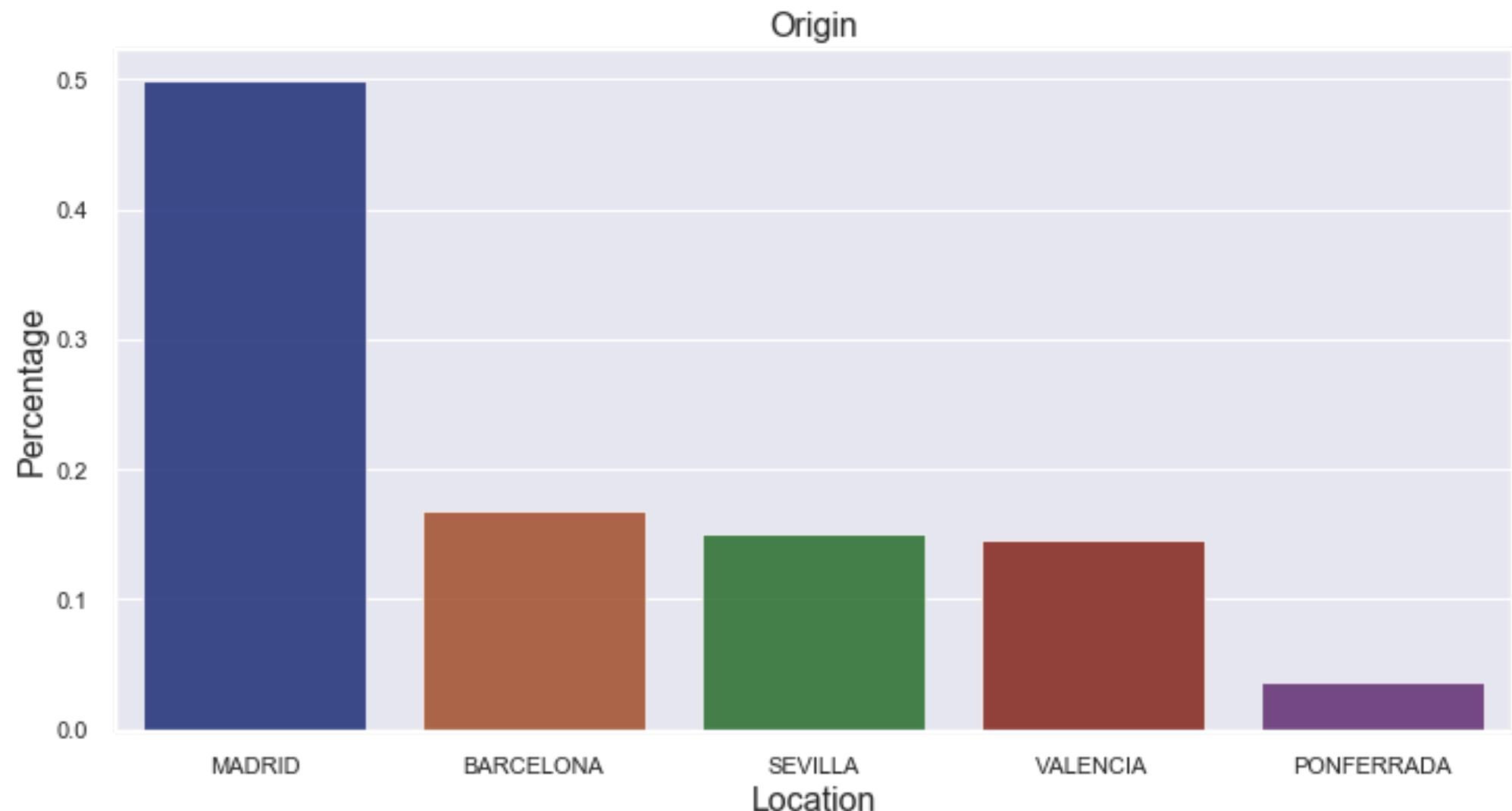
Conditions

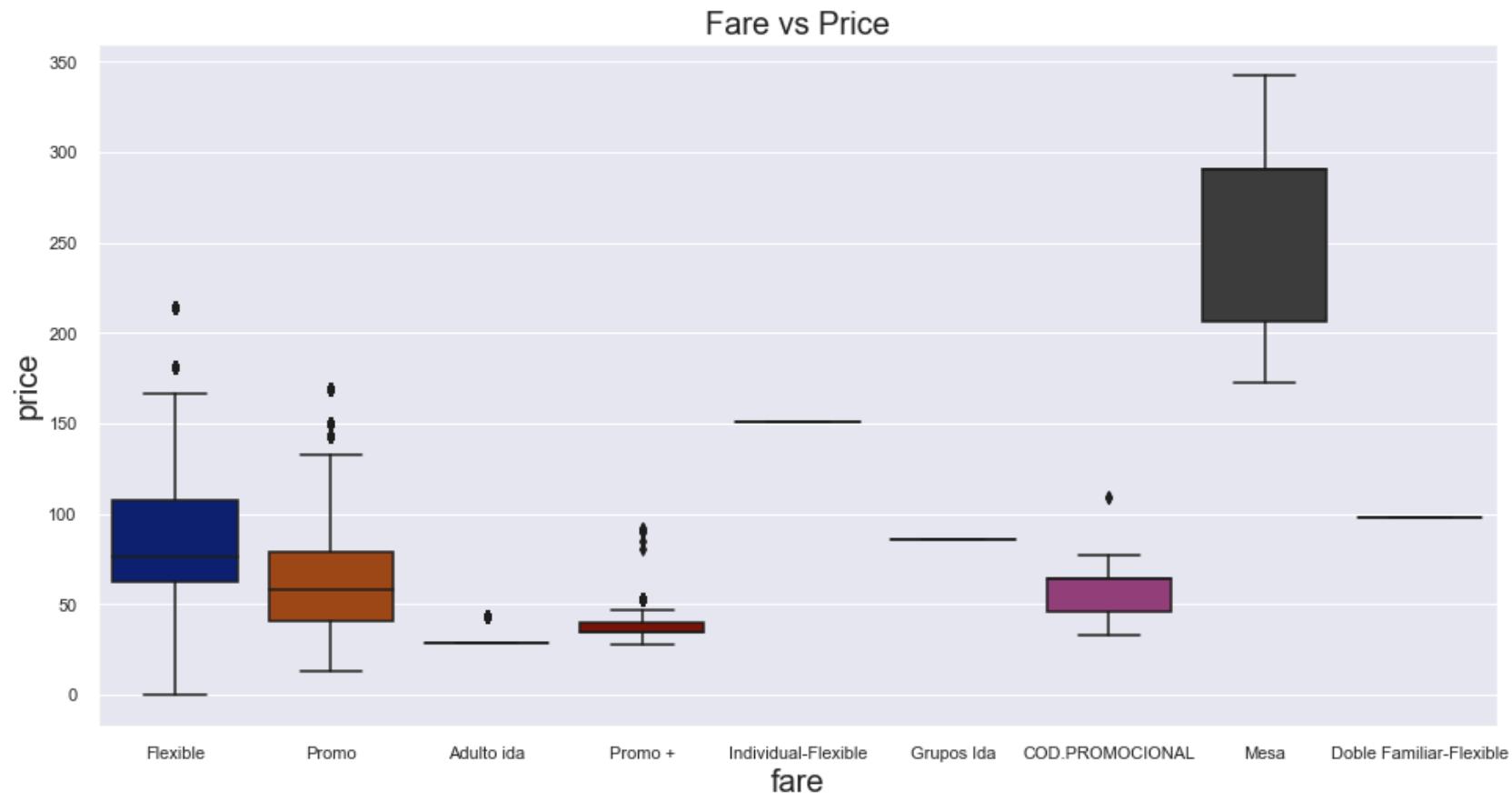
Departure Popularity

Bigger circle- more departure
Smaller circle- less departures

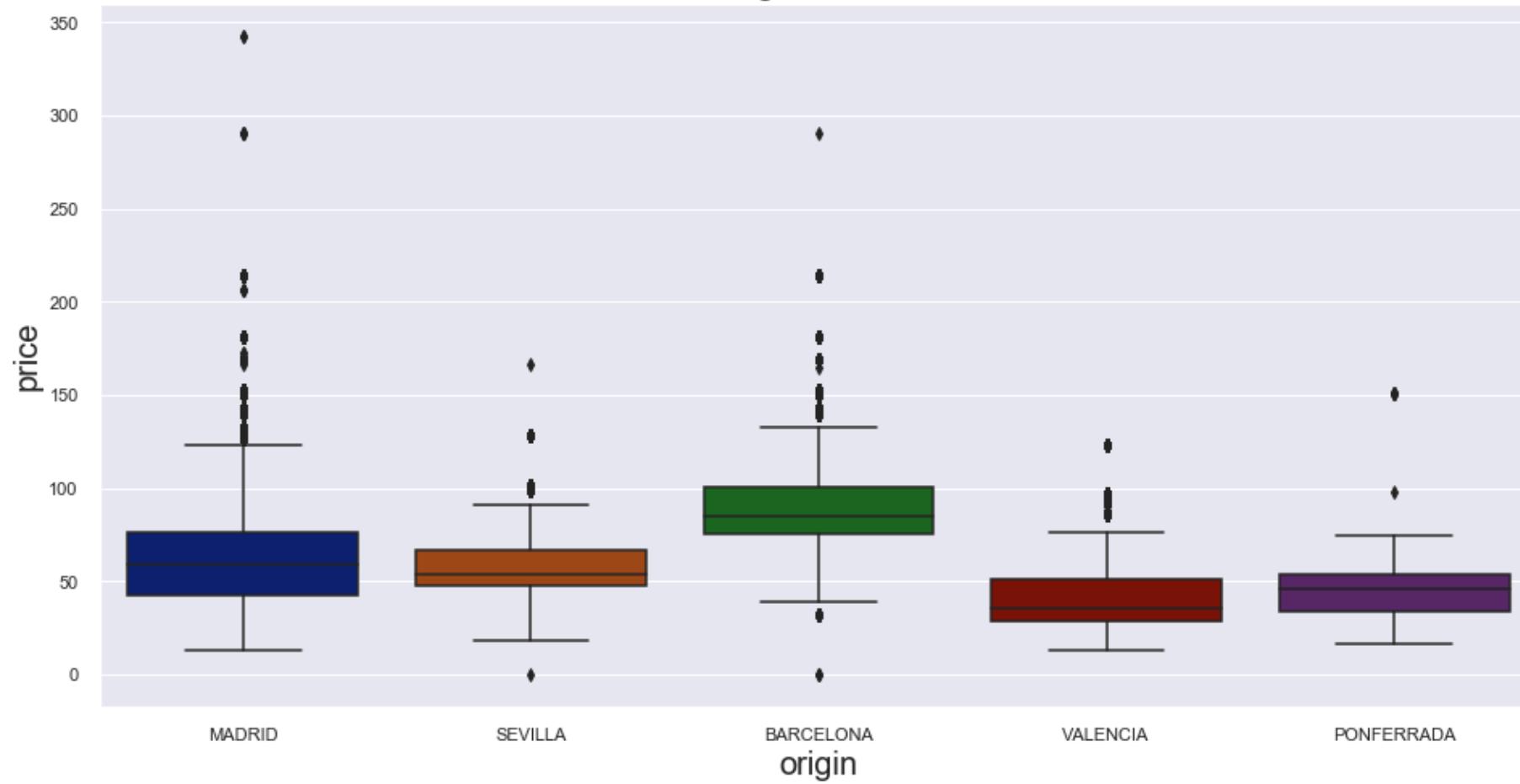


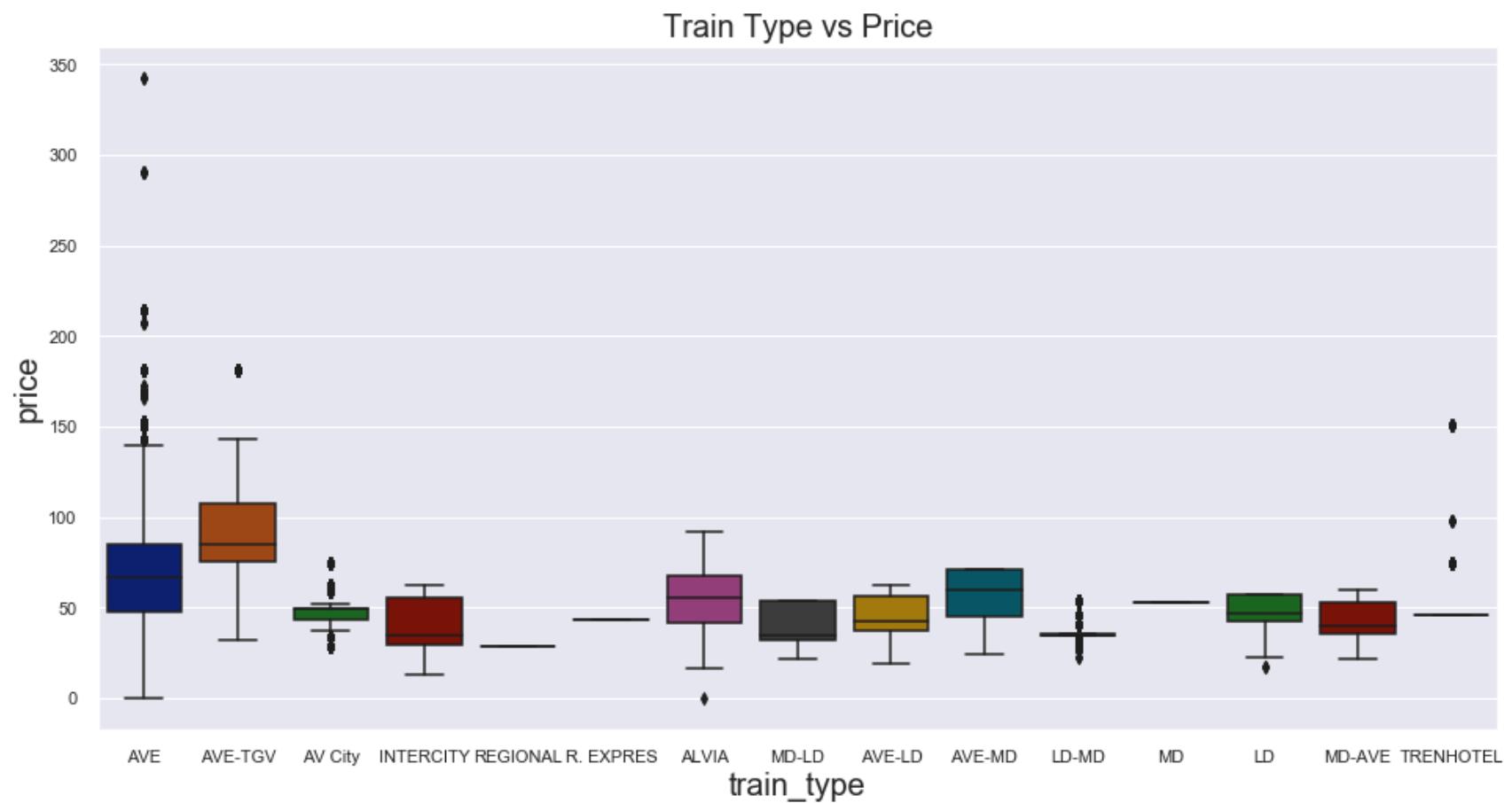


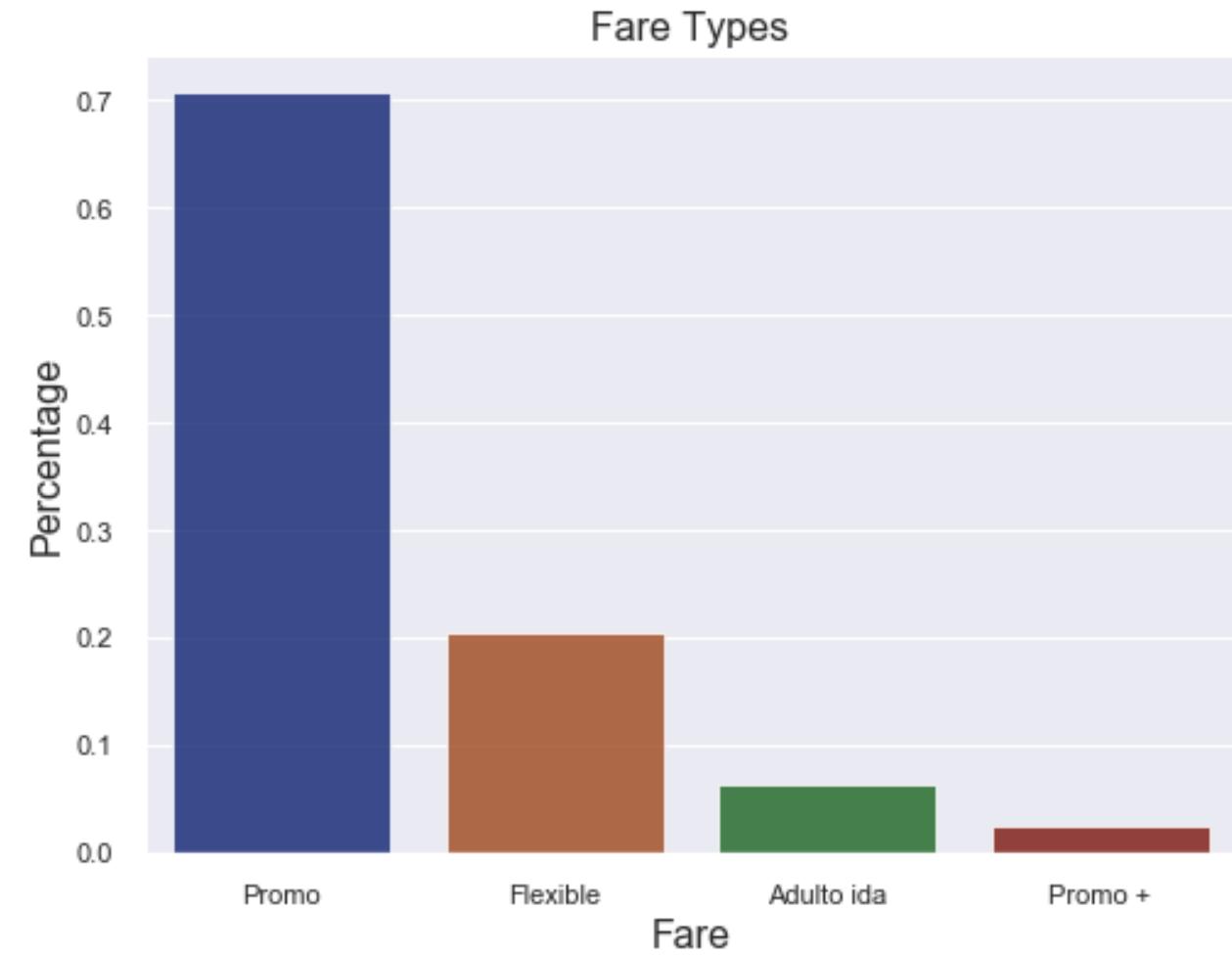
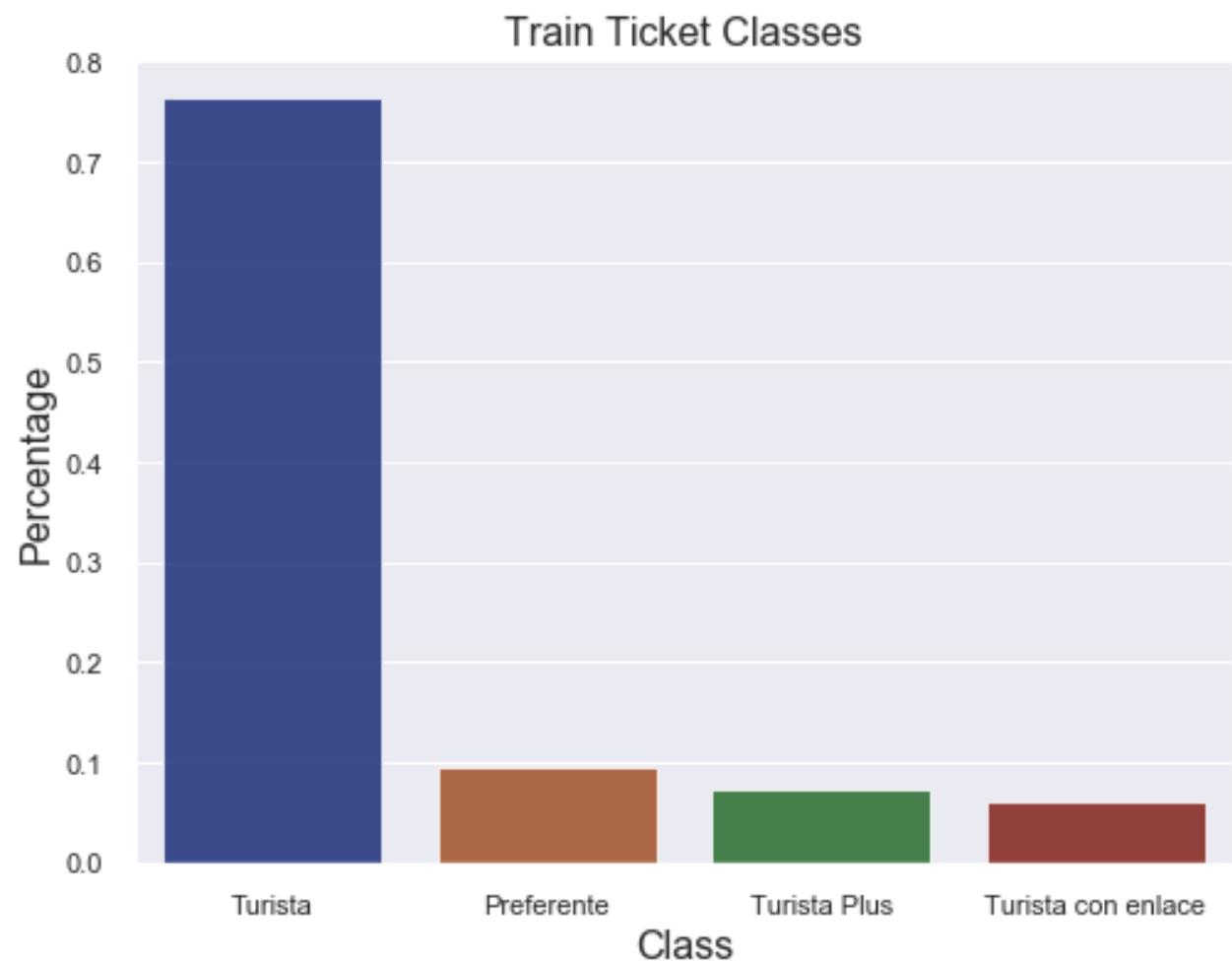




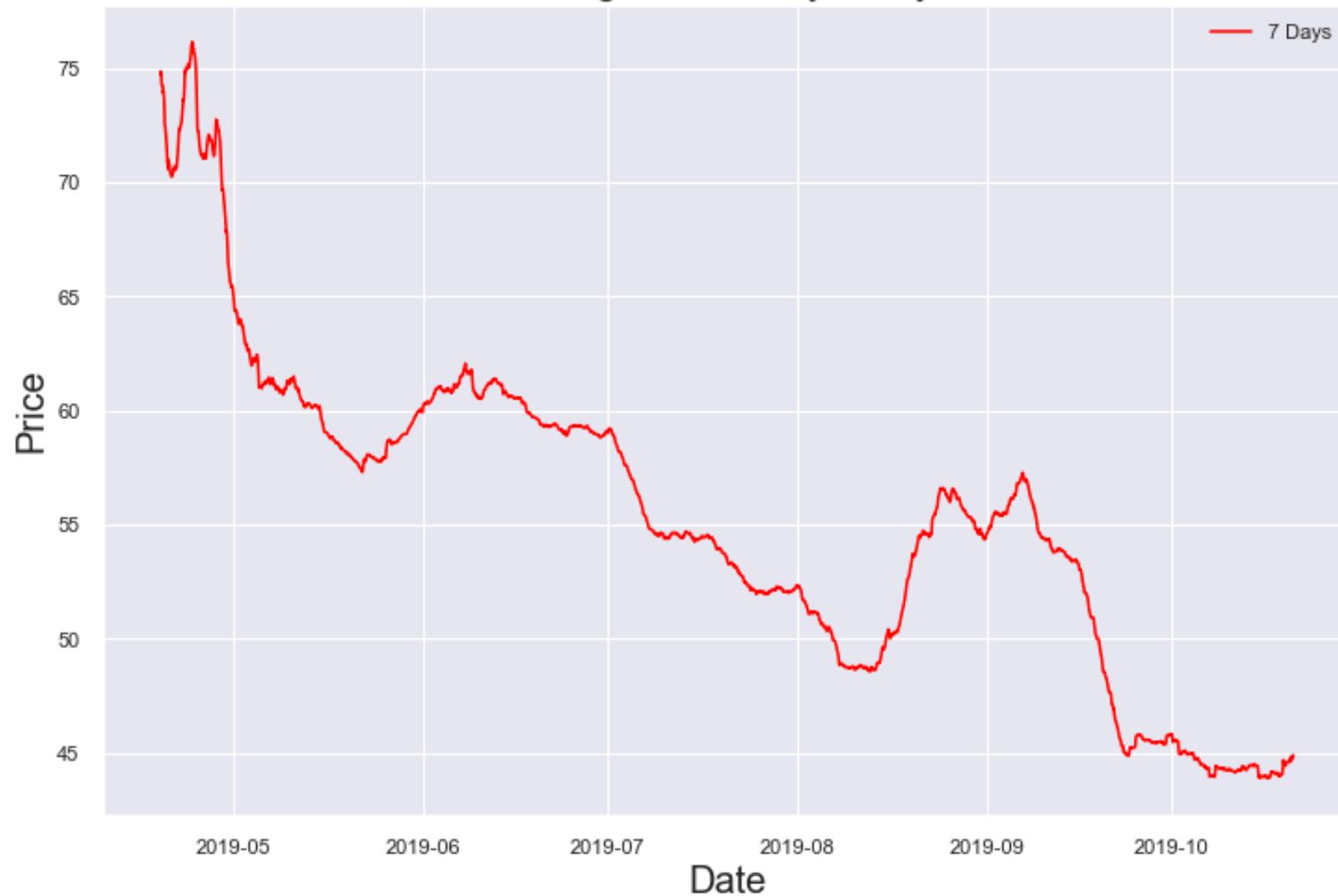
Origin vs Price







Rolling Mean Every 7 Days



Feature Engineering

- Features Added:
 - Start Month
 - Start Day
 - Start hour
 - End Month
 - End Day
 - End Hour
- Why?
 - Prices more expensive in summer
 - Prices vary by time

Models

Model	Train	Test	R^2	Absolute MSE	Time (Seconds)
Linear Regression	85.3%	.85	-.863	7.05	3.78
Decision Trees	95.4%	95%	.95	3.07	16.7
Bagging	95.4%	95%	.95	3.12	41.1
Random Forest	95.4%	95%	.95	3.05	415

Baseline Score

Train: 5.3%

Test: 5.3%

Model Selection

- Decision Trees
- Run Time
- R^2 Score





Conclusion

- We can predict the correct prices 95% of the time
- Travel company can provide this information to their customers
- Customers will be informed when the best time to buy a train ticket is





Recomendatoins

- Data should be analyzed that contains prices for when a person buy a ticket one week in advancne versus one month to predict how much in advance a customer should buy a ticket.

