

IN PROGRESS

Out[14]: [Click here to toggle on/off the raw code.](#)

1. Introduction

Airbnb is easily one of the most successful start-ups in recent memory. Airbnb's business model has been a disruptive force in the short-term rental market and was valued at \$31 billion prior to the Corona Virus epidemic.

Out[15]: [Click here to toggle on/off the raw code.](#)

However, like many other businesses, COVID-19 had dire impacts on the company's business model. Due to the massive wave of lockdowns all across the world, the industry's valuation plummeted to nearly half of what it once was.

In this analysis, I will be exploring the Airbnb rental market in San Francisco from 7/3/2018 through 6/7/2020 to develop a better understanding about the impact COVID-19 is having specifically on SF market and what it means for the company moving forward in 2020.

1.1 Objectives

- Conduct an analysis of the impact COVID-19 has had on Airbnb's daily gross rental income in the San Francisco area since the lockdown began on 3/19/2020. Forecast 2020 Q2, Q3, and Q4 losses if lockdown conditions remain the same for the rest of the year.
- Create a per night rental price predictor for Airbnb properties in San Francisco. Extract out the most pertinent features that predict prices best. Create a guide for tri-kinne hosts real estate features and amenities that can be added to Airbnb rental to increase per night rental price.

1.2 Impact

Understanding the impact COVID-19 has had on daily gross rental income will allow us to forecast losses for the rest of the year. Creating a worst-case scenario of earnings for the rest of 2020 will allow company leadership to manage expectations for shareholders and table/prioritize certain projects over others for the foreseeable future.

Developing an accurate price predictor model will provide Airbnb hosts a more accurate representation of how much their property is worth on the rental platform. Doing so will increase the likelihood tourists and long-term renters will want to rent their property. Additionally, providing an accurate guide of features and amenities for first-time hosts to consider including with their rental will not only allow for more successful first-time experiences, but also increase the number of regular Airbnb hosts throughout the year.

2. Table of Contents

- Introduction
 - 1.1 Introduction
 - 1.2 Impact
- Table of Contents
- Data Overview
 - 3.1 About the Data
 - 3.2 Preview Data
 - 3.3 Data Overview - Metrics
- Impact of COVID 19 on Gross Rental Revenue
 - 4.1 Methodology
 - 4.2 Results
 - 4.3 Decision Recommendations
- Airbnb Rental Prices in San Francisco
 - 5.1 Methodology
 - 5.2 Base Model Decision Tree
 - 5.3 Tuned XGBRegressor Model
 - 5.4 Model Discussion and New Host Recommendations
- Conclusion
 - 6.1 Takeaways
 - 6.2 Future Research

3. Data Overview

3.1 About the Data

Airbnb Listings data was acquired from Inside Airbnb (no affiliation with Airbnb). Inside Airbnb is an independent, non-commercial set of tools and data that allows you to explore how Airbnb is really being used in cities around the world. By analyzing publicly available information about a city's Airbnb's listings, Inside Airbnb provides filters and key metrics so you can see how Airbnb is being used to compete with the residential housing market.

3.2 Preview Data

Cleaned Data Shape: (126258, 48)

| accommodancy | amenities | availability_30 | availability_365 | availability_60 | availability_90 | bathrooms | bed_type | bedrooms | beds | calculated |
|--------------|-----------|---------------------------------------------------------|------------------|-----------------|-----------------|-----------|----------|----------|------|------------|
| 0 | 3 | TV, Cable TV, Internet, Wi-Fi, Kitchen, Pets fr... | 0 | 77 | 0 | 1 | 1.0 | Real Bed | 1.0 | 2.0 |
| 1 | 2 | TV, Internet, Wi-Fi, Kitchen, Free street parking... | 30 | 365 | 0 | 0 | 4.0 | Real Bed | 1.0 | 1.0 |
| 2 | 5 | TV, Internet, Wi-Fi, Kitchen, Heating, Parking... | 30 | 90 | 60 | 90 | 1.5 | Real Bed | 2.0 | 2.0 |
| 3 | 6 | TV, Cable TV, Internet, Wi-Fi, Kitchen, Free parking... | 0 | 20 | 5 | 9 | 1.0 | Real Bed | 2.0 | 3.0 |
| 4 | 3 | TV, Cable TV, Internet, Wi-Fi, Paid parking off... | 4 | 131 | 19 | 42 | 1.0 | Real Bed | 1.0 | 2.0 |

3.3 Data Overview - Metrics

Data Description

| | count | mean | std | min | 25% | 50% | 75% | max |
|----------------------------------------------|----------|---------------|--------------|--------------|---------------|---------------|---------------|---------------|
| accommodancy | 126258.0 | 3.217230e+01 | 1.947610e+00 | 1.000000 | 2.000000e+00 | 2.000000e+00 | 4.000000e+00 | 1.000000e+01 |
| availability_30 | 126258.0 | 9.534894e+00 | 1.048743e+01 | 0.000000 | 0.000000e+00 | 5.000000e+00 | 1.700000e+01 | 3.000000e+01 |
| availability_365 | 126258.0 | 1.408705e+02 | 1.266210e+02 | 0.000000 | 2.000000e+00 | 1.000000e+02 | 2.000000e+02 | 3.000000e+02 |
| availability_60 | 126258.0 | 2.260176e+01 | 2.308564e+01 | 0.000000 | 1.000000e+00 | 1.800000e+00 | 4.100000e+01 | 6.000000e+01 |
| availability_90 | 126258.0 | 3.027428e+01 | 3.108503e+01 | 0.000000 | 5.000000e+00 | 3.000000e+00 | 6.000000e+01 | 9.000000e+01 |
| bathrooms | 126258.0 | 1.341598e+00 | 0.540870e+01 | 0.000000 | 1.000000e+00 | 1.000000e+00 | 1.100000e+01 | 1.400000e+01 |
| bedrooms | 126258.0 | 1.354209e+00 | 0.912003e+01 | 0.000000 | 1.000000e+00 | 1.000000e+00 | 2.000000e+00 | 1.400000e+01 |
| beds | 126258.0 | 1.790309e+00 | 1.206100e+00 | 0.000000 | 1.000000e+00 | 1.000000e+00 | 2.000000e+00 | 1.000000e+01 |
| calculated_host_listings_count | 126258.0 | 1.164335e+00 | 3.761799e+01 | 0.000000 | 1.000000e+00 | 2.000000e+00 | 4.000000e+00 | 3.010000e+02 |
| calculated_host_listings_count_private_rooms | 105994.0 | 2.726702e+00 | 7.322385e+00 | 0.000000 | 0.000000e+00 | 0.000000e+00 | 2.000000e+00 | 8.700000e+00 |
| calculated_host_listings_count_shared_rooms | 105994.0 | 4.763628e+01 | 2.850416e+00 | 0.000000 | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 3.000000e+01 |
| cleaning_fee | 126258.0 | 0.963889e+01 | 7.891858e+01 | 0.000000 | 3.000000e+00 | 1.750000e+01 | 1.200000e+02 | 1.350000e+03 |
| cnnn_price | 126258.0 | 1.060889e+01 | 3.112158e+01 | 0.000000 | 0.000000e+00 | 5.000000e+00 | 2.000000e+01 | 3.000000e+02 |
| guests_included | 126258.0 | 1.787233e+00 | 1.407879e+00 | 1.000000 | 1.000000e+00 | 1.000000e+00 | 2.000000e+00 | 1.000000e+00 |
| host_has_profile_pic | 126258.0 | 0.964285e+01 | 7.532572e+02 | 0.000000 | 1.000000e+00 | 1.000000e+00 | 1.000000e+00 | 1.000000e+00 |
| host_id | 126258.0 | 4.915831e+07 | 6.886226e+07 | 28.000000 | 3.877879e+08 | 1.828818e+07 | 6.464862e+07 | 3.452837e+08 |
| host_identity_verified | 126258.0 | 0.709180e+01 | 4.948979e+01 | 0.000000 | 0.000000e+00 | 1.000000e+00 | 1.000000e+00 | 1.000000e+00 |
| host_is_superhost | 126258.0 | 0.110506e+01 | 4.948803e+01 | 0.000000 | 0.000000e+00 | 5.000000e+00 | 1.000000e+00 | 1.000000e+00 |
| host_response_rate | 126258.0 | 0.824807e+01 | 3.362359e+01 | 0.000000 | 9.700000e+01 | 1.000000e+00 | 1.000000e+00 | 1.000000e+00 |
| instant_bookable | 126258.0 | 1.700399e+07 | 1.148872e+07 | 958.000000 | 6.495015e+06 | 1.771786e+07 | 2.584328e+07 | 4.348820e+07 |
| is_location_exact | 126258.0 | 0.880397e+01 | 3.173658e+01 | 0.000000 | 0.000000e+00 | 1.000000e+00 | 1.000000e+00 | 1.000000e+00 |
| latitude | 126258.0 | 3.776330e+01 | 2.387956e+02 | 37.764120 | 3.776193e+01 | 3.776831e+01 | 3.776860e+01 | 3.793333e+01 |
| longitude | 126258.0 | -1.224327e+02 | 2.742094e+02 | -122.513065 | -1.224454e+02 | -1.224277e+02 | -1.224120e+02 | -1.223670e+02 |
| maximum_nights | 126258.0 | 1.747421e+04 | 6.048364e+06 | 1.000000 | 2.000000e+00 | 9.000000e+01 | 1.125000e+03 | 2.147748e+09 |
| minimum_nights | 126258.0 | 1.072055e+01 | 1.550510e+01 | 1.000000 | 2.000000e+00 | 3.000000e+00 | 3.000000e+01 | 9.990000e+02 |
| monthly_price | 126258.0 | 4.177345e+03 | 2.740558e+03 | 700.000000 | 2.000000e+00 | 3.000000e+00 | 4.999000e+03 | 2.500000e+04 |
| number_of_reviews | 126258.0 | 4.182232e+01 | 8.881759e+01 | 0.000000 | 0.000000e+00 | 2.000000e+00 | 6.000000e+01 | 8.210000e+02 |
| price | 126258.0 | 2.130139e+02 | 3.284834e+02 | 10.000000 | 9.000000e+01 | 1.490000e+02 | 2.300000e+02 | 1.000000e+04 |
| require_guest_phone_verification | 126258.0 | 0.468168e+02 | 2.273588e+01 | 0.000000 | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 1.000000e+00 |
| require_guest_profile_picture | 126258.0 | 4.589010e+02 | 0.282477e+01 | 0.000000 | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 1.000000e+00 |
| require_license | 126258.0 | 0.941800e+01 | 7.602401e+02 | 0.000000 | 1.000000e+00 | 1.000000e+00 | 1.000000e+00 | 1.000000e+00 |
| review_scores_rating | 126258.0 | 0.810777e+01 | 4.848869e+01 | 20.000000 | 8.000000e+01 | 9.000000e+01 | 9.000000e+01 | 1.000000e+02 |
| reviews_per_month | 126258.0 | 2.048854e+00 | 2.068254e+00 | 0.000000 | 4.200000e-01 | 1.290000e+00 | 3.500000e+00 | 4.101700e+01 |
| security_deposit | 126258.0 | 3.418327e+02 | 6.260580e+02 | 0.000000 | 0.000000e+00 | 1.500000e+02 | 5.000000e+02 | 5.000000e+03 |
| weekly_price | 20639.0 | 1.297647e+03 | 9.815052e+02 | 250.000000 | 7.500000e+02 | 1.000000e+03 | 1.100000e+03 | 9.999000e+03 |
| zipcode | 126258.0 | 9.411478e+04 | 1.131390e+01 | 86013.000000 | 9.410800e+04 | 9.411400e+04 | 9.412100e+04 | 9.495000e+04 |

4. Impact of COVID 19 on Gross Rental Revenue

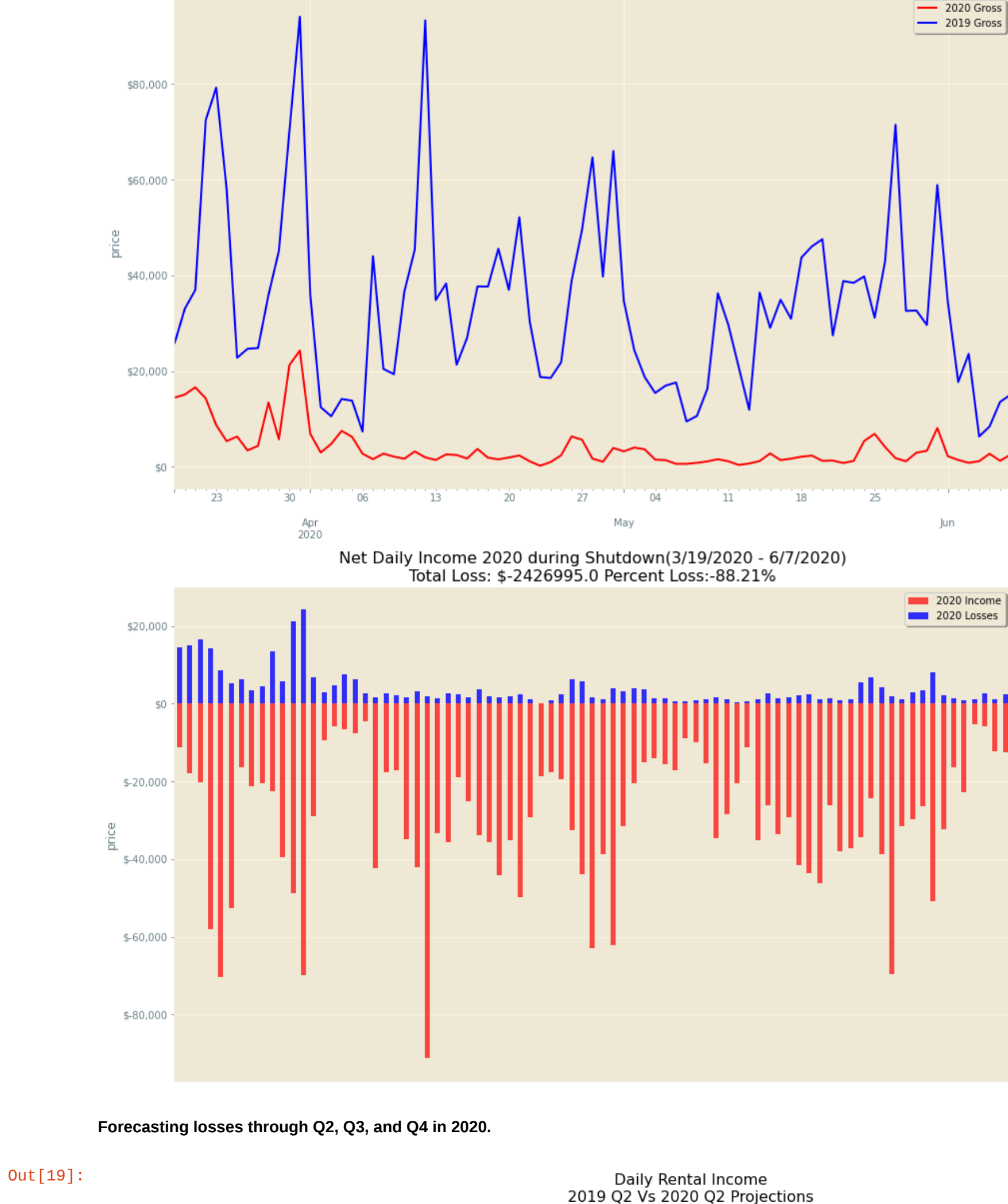
4.1 Methodology

Due to the relatively new impact of COVID-19 shutdown, there is not enough data to accurately model an accurate time series forecast. To circumvent this, I calculated the average daily percentage loss based on the data we presently have concerning the lockdown (March 19/2020 - 6/7/2020) and then extrapolated out that loss against our previous 2019 data for Q2, Q3, and Q4 of 2020. It is safe to assume the 2020 would have resulted in (at least) not slightly greater) revenue than 2019, so this extrapolation will serve as a worst-case scenario.

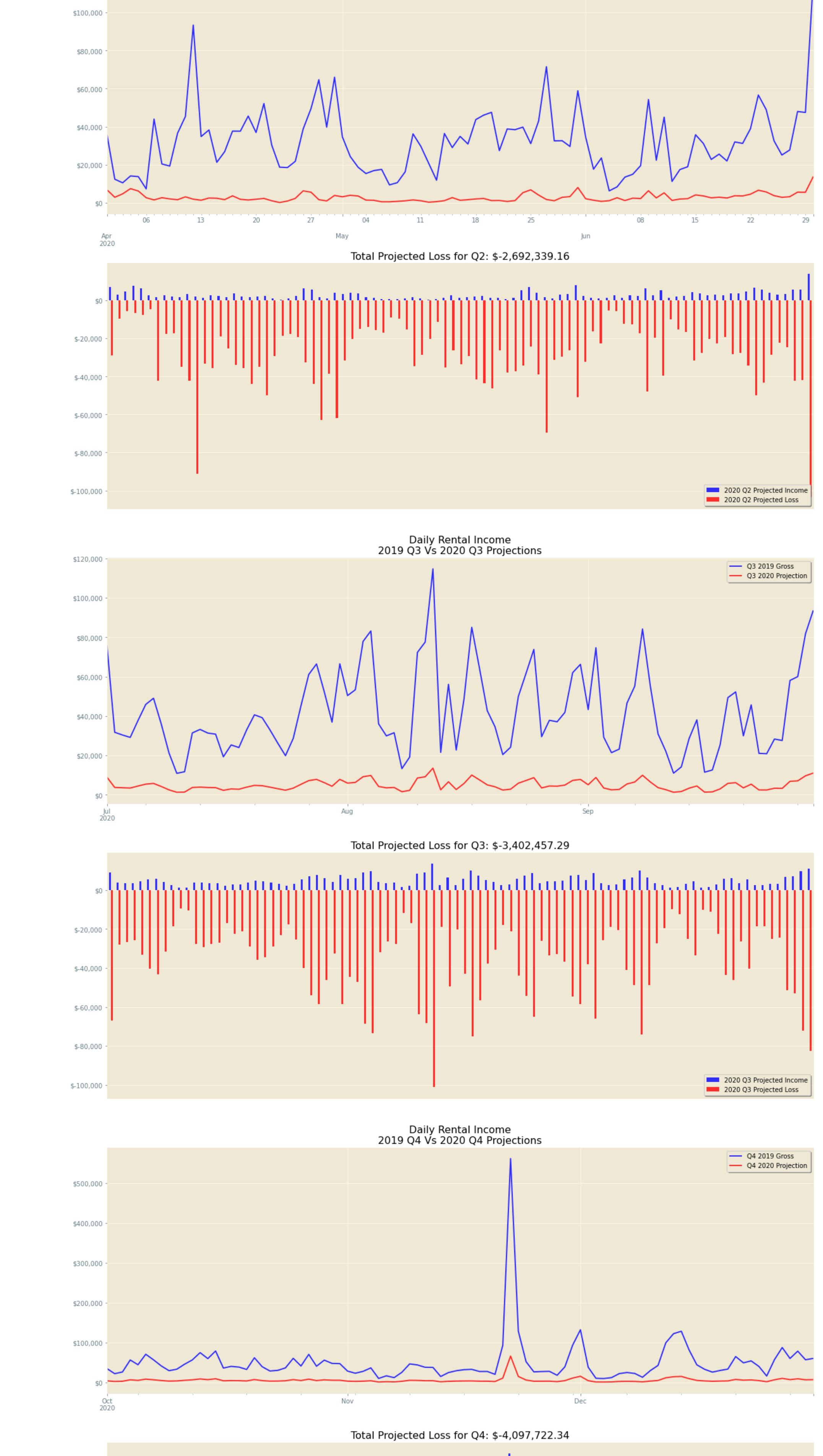
4.2 Results

Losses Due to Initial Shutdown

The loss due to COVID-19 shutdown for the data we have concerning 2020 is below, data ranges from 3/19/2020 - 6/7/2020



Forecasting Losses through Q2, Q3, and Q4 in 2020.



4.3 Decision Recommendations

San Francisco Daily Gross Income Analysis Results:

- Q2 Projected Loss: \$2,692,339.16
- Q3 Projected Loss: \$3,402,457.29
- Q4 Projected Loss: \$4,097,722.34
- Projected Net 2020 Losses: \$10,192,518.79

Looking at the results, things do not look great should the lockdown remain in place as currently constructed for the rest of 2020 in San Francisco. Due to concerns from COVID 19, it would be a wise decision to only offer listings where the full-property is rented to a single-person/family. No single rooms should be permitted. Airbnb does already require cleaning to be done after each rental is completed (which is included in the price), so consumers will feel more at ease should we make this mandatory.

Something to consider for Airbnb as well is moving into the long-term rental market. The most profitable/popular listings on the platform belong to long-term rental companies who provide fully furnished dwellings in San Francisco. A potential strategy to ensure secured revenue would be to work with local apartments in SF and use our platform/advertising as a means to attempt to book empty dwellings. Airbnb thus could then capture either a commission or a percentage of the monthly rent in long-term deals.

5. Airbnb Rental Prices in San Francisco

5.1 Methodology

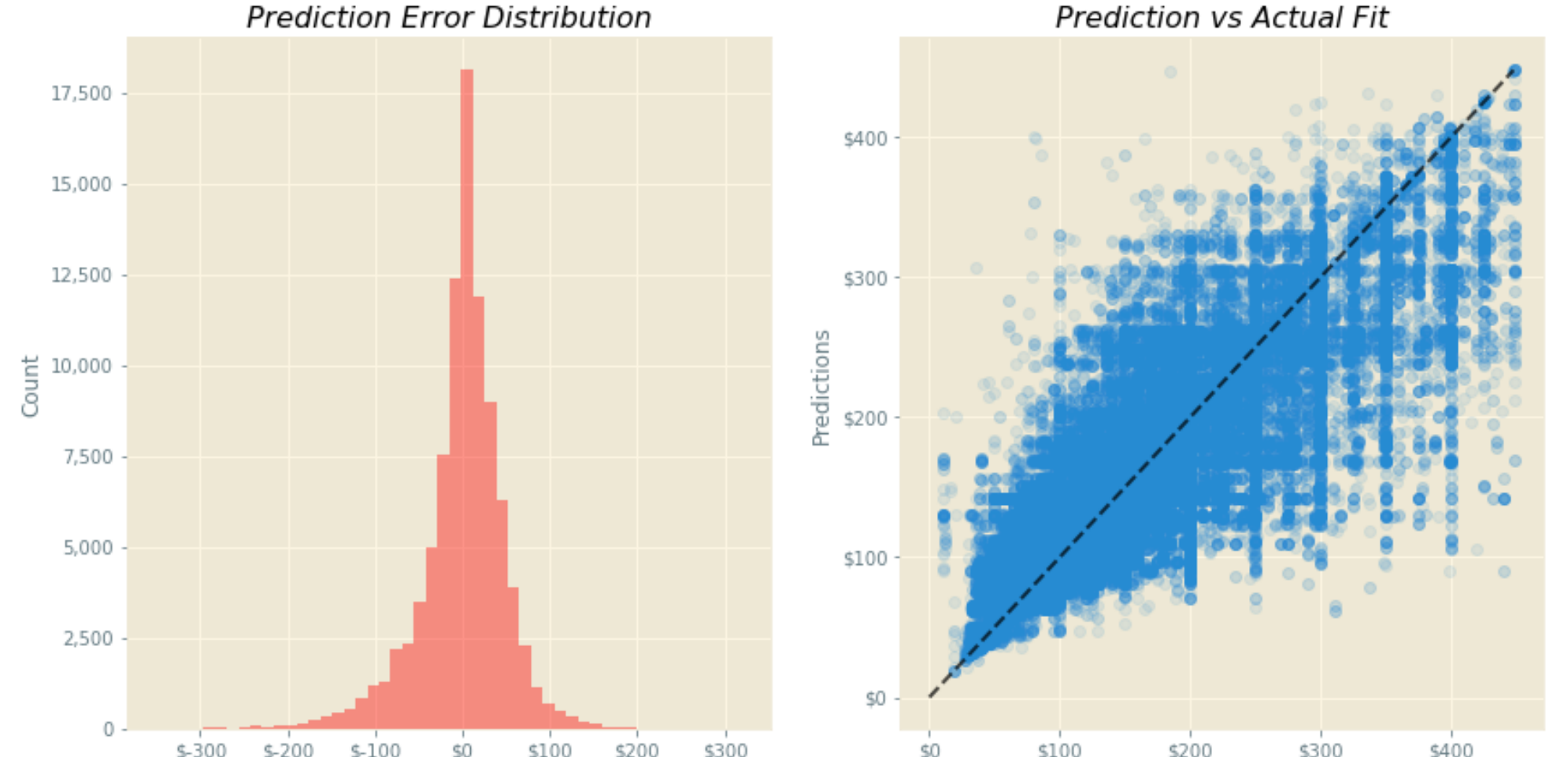
After aggregating and cleaning Airbnb listings data in San Francisco, created a preprocessing pipeline that scaled numeric and encoded categorical variables. Created a base model decision tree as initial rental price predictor to serve as a baseline for testing out ensemble and boosting methods. After testing a few more complex models, settled on optimizing an XGBRegressor model using GridSearchCV and additionally extracted the most important features to the models' performance to serve as a reference to new Airbnb hosts.

5.2 Base Model Decision Tree

Out[20]:

Base Tree Regressor Model Error Evaluation

Root Mean Squared Error:48.86



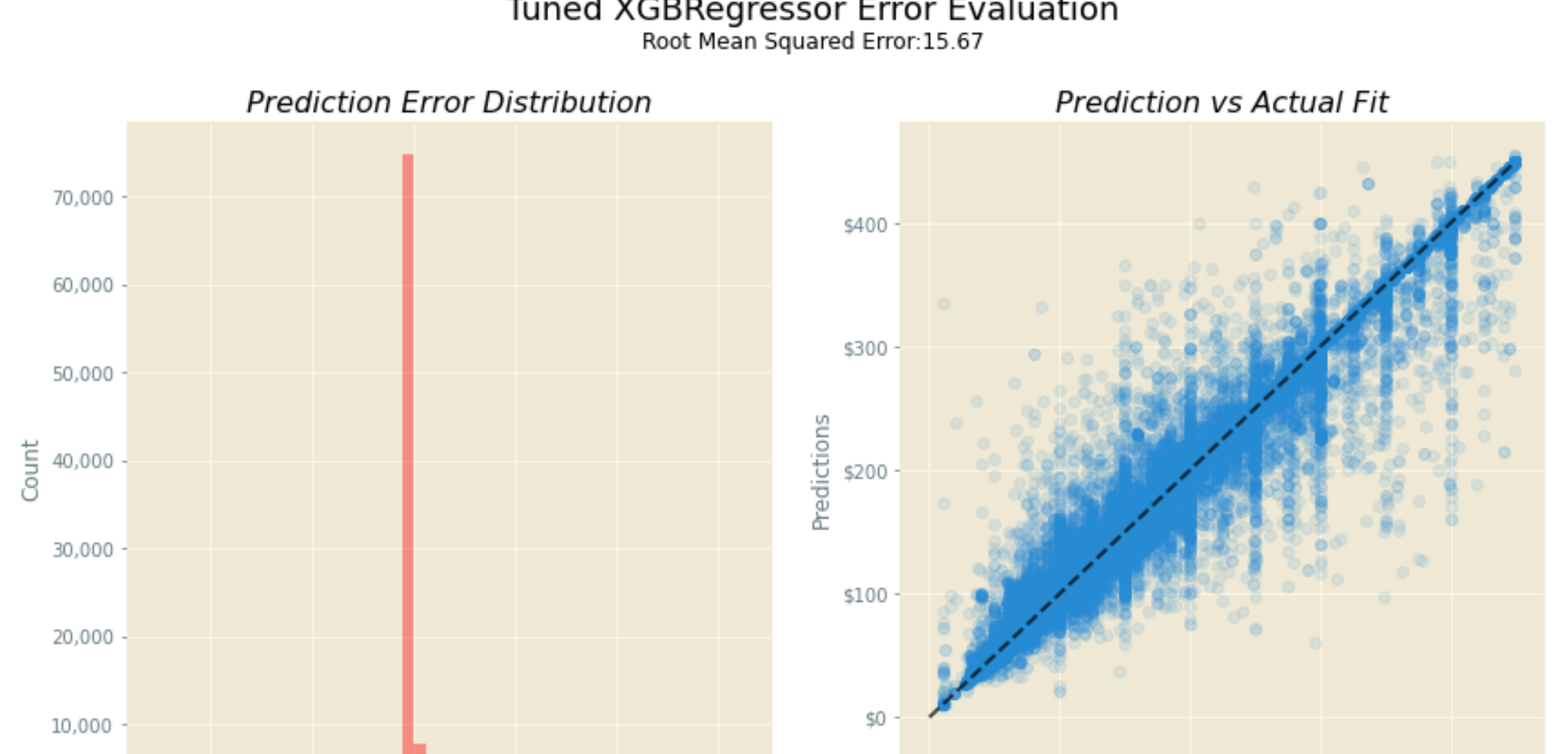
5.3 Tuned XGBRegressor Price Predictor

Tuned XGBRegressor performance on Training data

Out[21]:

Tuned XGBRegressor Error Evaluation

Root Mean Squared Error:15.67

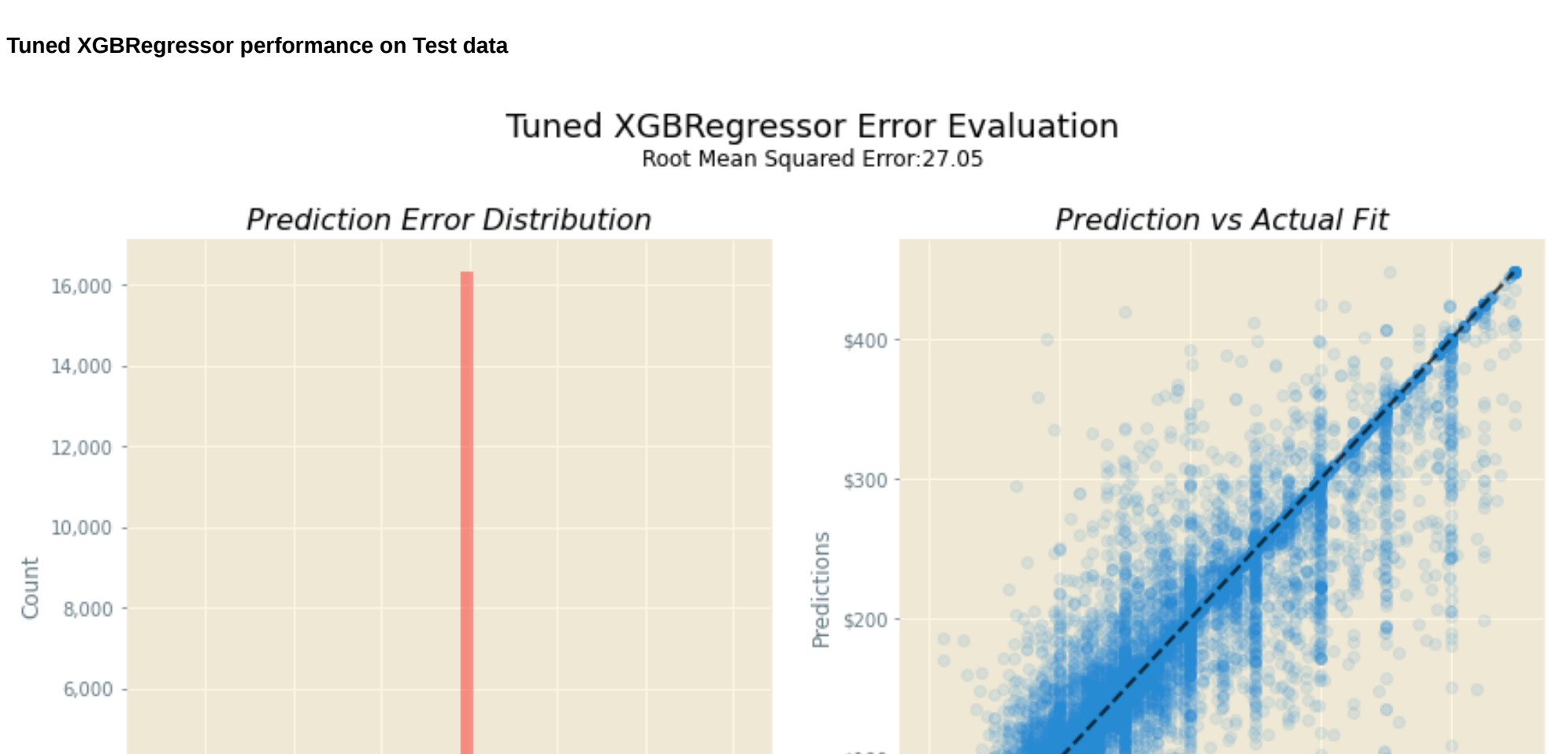


Tuned XGBRegressor performance on Test data

Out[22]:

Tuned XGBRegressor Error Evaluation

Root Mean Squared Error:27.05



Feature Importance of tuned XGBRegressor model

Out[23]:

