Airbnb Toronto Price Prediction

Capstone Project Presentation

UofT SCS Data Analytics Boot Camp
Aug 10th, 2022

Contributors:
Carolina Semerano
Sukanya Ghosh
Yesha Tharwala

Link to Project Presentation

Contents

Tools Used

1. Github (Project link)
2. Database - PostgreSQL (Database details link)
3. Tableau (Dashboard link)
4. Python
5. Flask
6. HTML
7. Amazon Web Services RDS

Overview

Airbnb is an online marketplace for short term rentals. Airbnb allows people from all over the world to host their homes as someone's next stay.

Properties can range from houses, apartments to single and shared rooms and are priced per night or per stay.

The problem

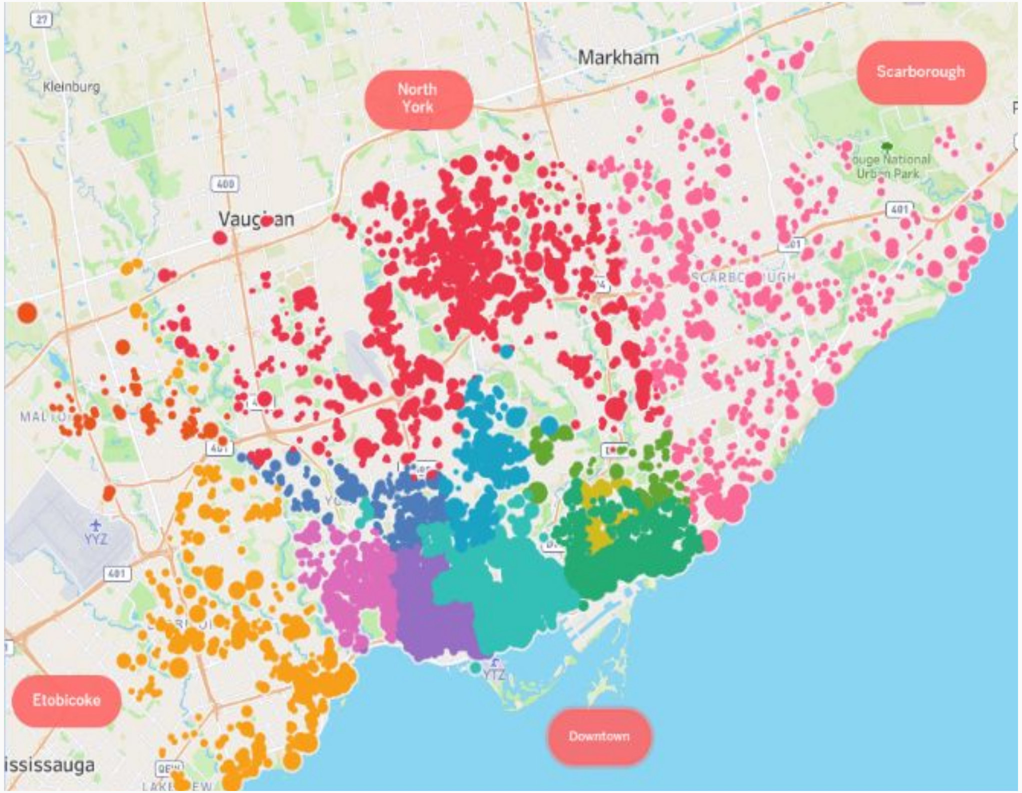
How do I determine the optimal rent price for my property?

The solution

Construct a data driven solution by using machine learning to predict rental prices for each property.

Exploratory Data Analysis

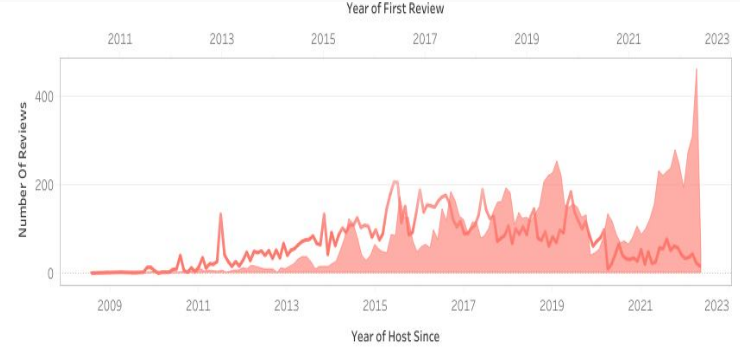
Airbnb listings distribution map



- There are approx. 15, 000 Airbnb listings in Toronto
- Exploratory Data Analysis on
    - Neighbourhoods
    - Boroughs
    - Reviews
    - Host Details
    - Amenities
    - Number of guests

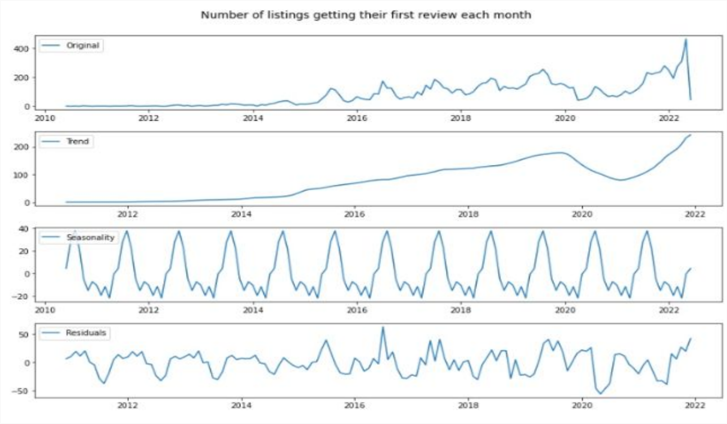
Exploratory Data Analysis

Number of hosts joining vs. first review

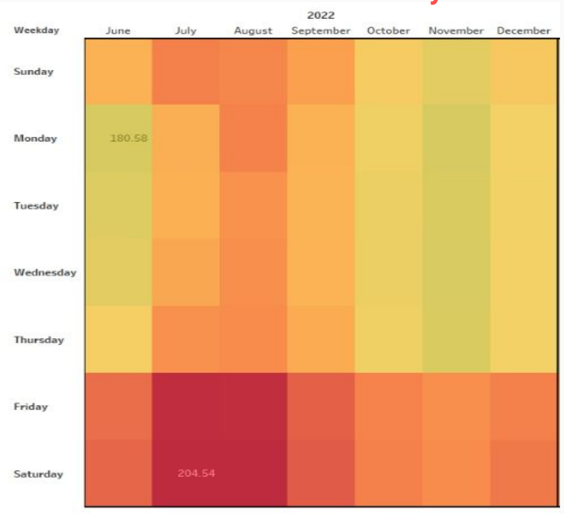


Yearly change in Average Price

Seasonality



Month and Weekday

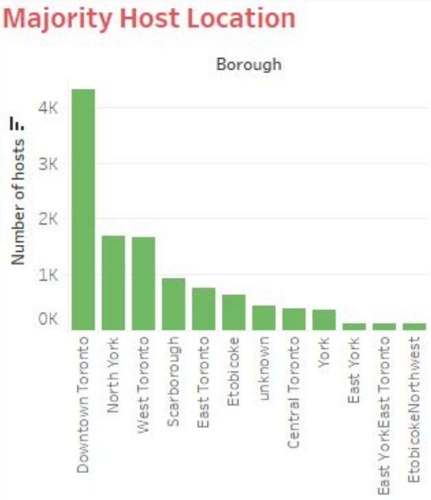


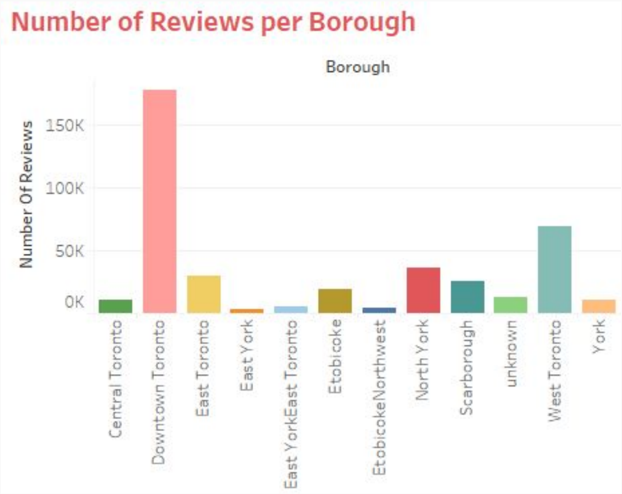
Observations:

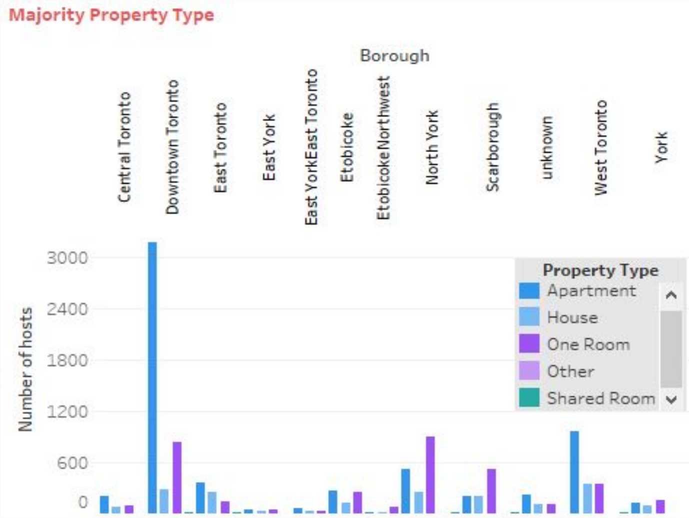
- Airbnb prices have gone up by 75% in 2022 compared to 2010
- Summer is the most popular season for Airbnb business
- Average prices are highest in August, and during the week the highest prices are on Saturdays

Exploratory Data Analysis

➢ Downtown Toronto is the most popular location among hosts as well as guests

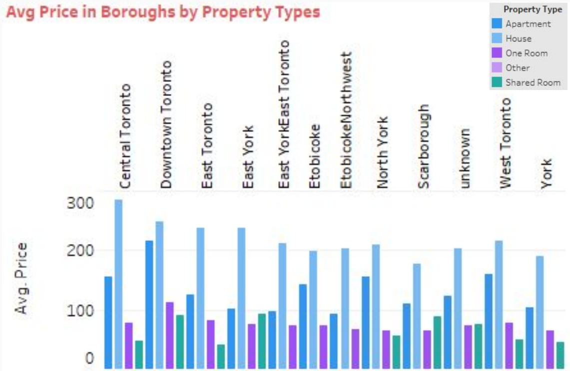


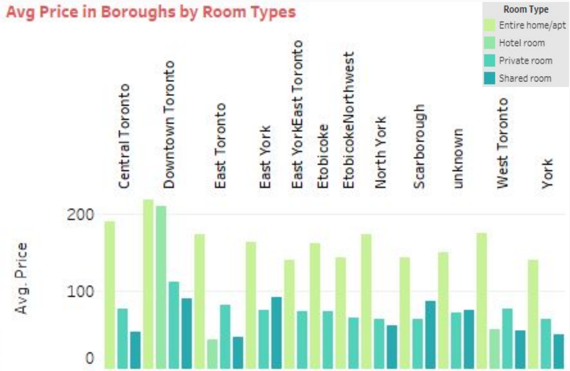


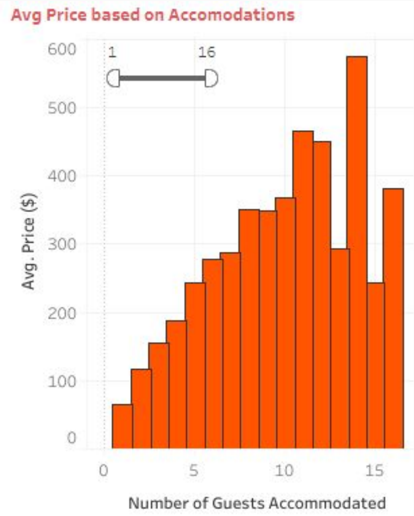


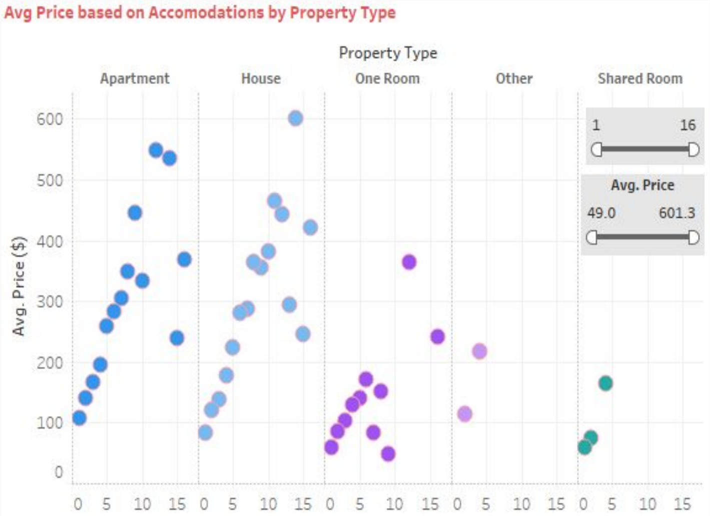
➢ More than 50% of the Airbnb listings in Toronto are Apartments

Exploratory Data Analysis





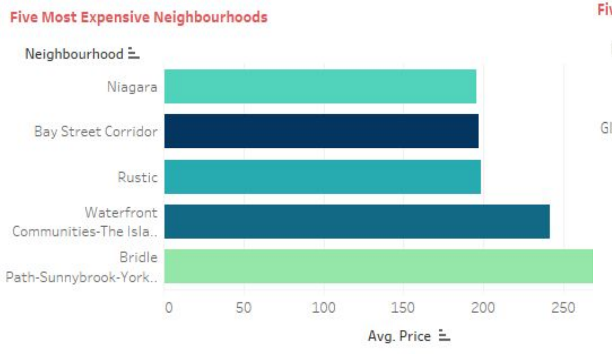


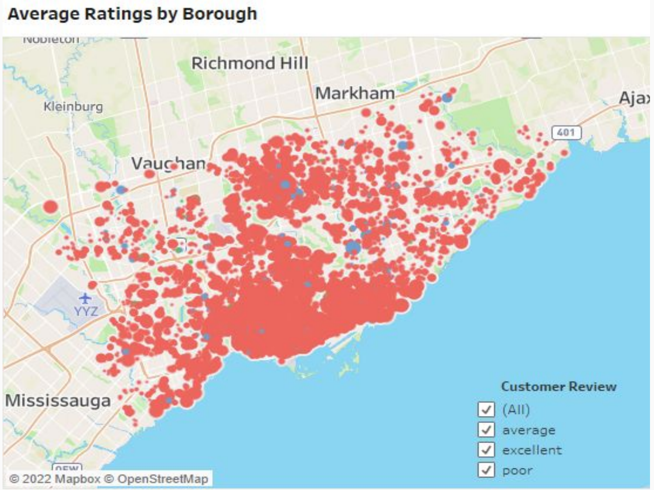


Observations:

- Renting an entire home or apartment is the most expensive in all boroughs
- In Downtown Toronto, both Apartments and entire homes can cost more than CAD $200 per night
- Average Price by Room Type
    - Entire home - CAD $180
    - Private room - CAD $80
    - Shared room - CAD $70
    - Hotel room - CAD $60
- Listings that accommodate 14 people are the most expensive at approx. CAD $580
- Single guests can rent for less than CAD $100 per night

Exploratory Data Analysis

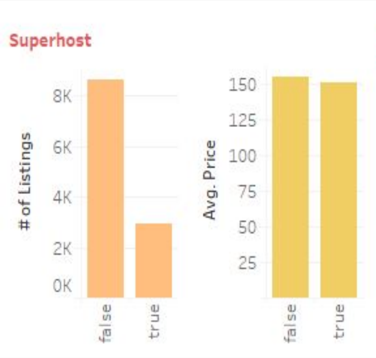


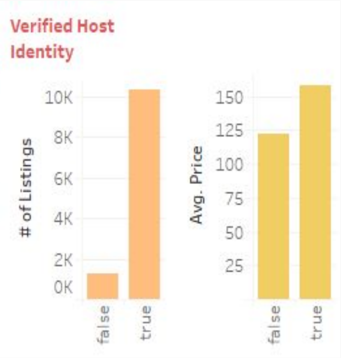


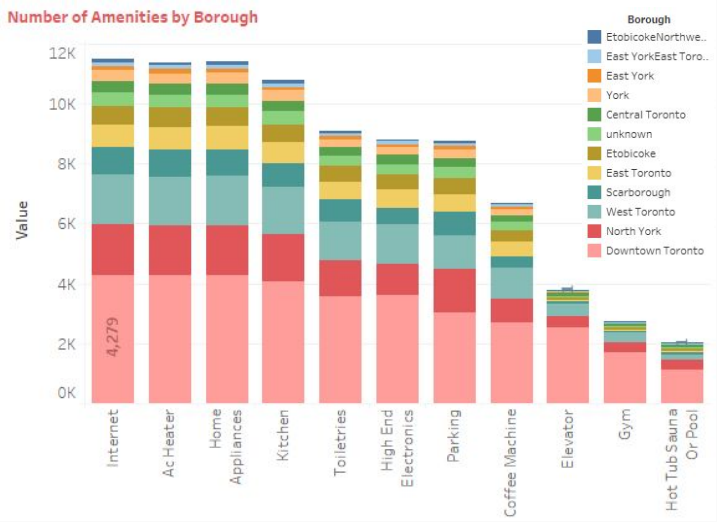
Observations:

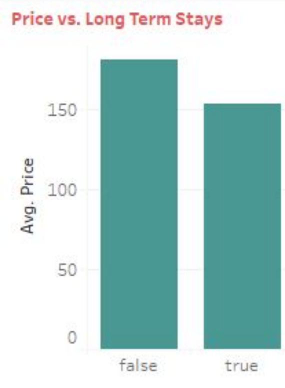
- The five most expensive neighbourhoods
  - Bridle Path-Sunnybrook-Yorkmills
  - Waterfront Communities-The Island
  - Rustic
  - Bay Street Corridor
  - Niagara
- The five least expensive neighbourhoods
  - Black Creek
  - Eglinton East
  - West Hill
  - Glenfield-Jane Heights
  - Pleasant View
- Most of the properties have “excellent” ratings
- Price benefit of 8-14% over properties with “poor” or “average” ratings

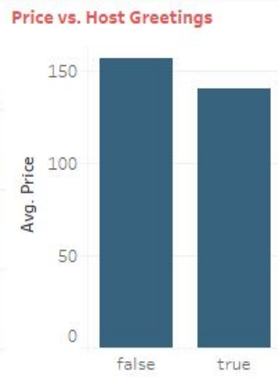
Exploratory Data Analysis











Observations:

- Around 25% of hosts are superhosts. No significant impact on average price
- Around 90% hosts have verified ID. Average prices approx. 25% higher
- 29% of hosts allow instant booking. No significant impact on average price
- **Popular Amenities** - Internet, AC Heater, Home Appliances, Kitchen, Toiletries, Electronics, Parking, Coffee Machines, etc.
- Higher prices are charged for amenities listed above along with some other amenities like gym, elevator, pool, etc.
- Long term stays and host greetings have lower average prices

Machine Learning Model - XGBoost

Linear Regression Model

```
RMSE train: 80.427
RMSE test: 79.391
R^2 train: 0.555
R^2 test: 0.550
```

Support Vector Regression Model

```
RMSE train: 101.333
RMSE test: 100.078
R^2 train: 0.294
R^2 test: 0.284
```

XGBoost Regressor

```
RMSE train: 46.833
RMSE test: 68.264
R^2 train: 0.849
R^2 test: 0.667
```

HistGradientBoosting Regressor

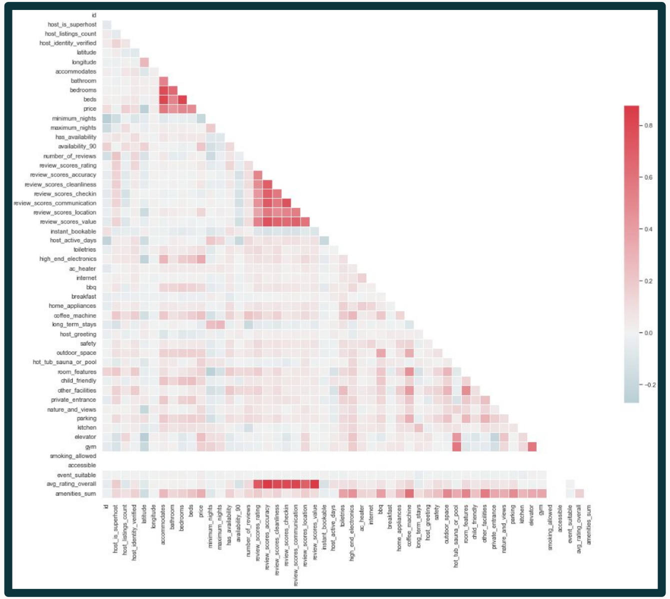
```
RMSE train: 67.088
RMSE test: 69.522
R^2 train: 0.690
R^2 test: 0.655
```

RandomForest Regressor

```
RMSE train: 26.395
RMSE test: 69.168
R^2 train: 0.952
R^2 test: 0.658
```

ExtraTrees Regressor

```
RMSE train: 0.076
RMSE test: 68.757
R^2 train: 1.000
R^2 test: 0.662
```



XGBoostRegressor - Hyper-parameter Tuning with GridSearchCV

Final Model After Hyper-parameter Tuning

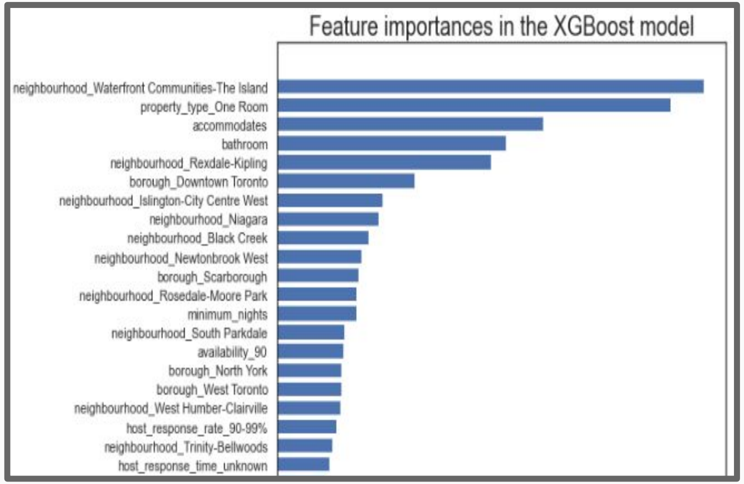
```python
# # Test optimum values for param_grid
param_grid = {
    'n_estimators': [100,250,500,1000],
    'eta':[0.05,0.1,0.2]
}

grid_search(param_grid,best_param)
```

```python
# # Run the XGBoost Regression model with optimized parameters
model=xgb.XGBRegressor(random_state=0, verbosity=1,**best_param)
model.fit(X_train_scaled, y_train)
training_pred=model.predict(X_train_scaled)
predictions=model.predict(X_test_scaled)
r2_score(y_test,predictions)

rmse_training=np.sqrt(mean_squared_error(y_train,training_pred))
rmse_model=np.sqrt(mean_squared_error(y_test, predictions))
print('RMSE train: %.3f' % rmse_training)
print('RMSE test: %.3f' % rmse_model)
print('R^2 train: %.3f' % (r2_score(y_train,training_pred )))
print('R^2 test: %.3f' % (r2_score(y_test, predictions)))
```

```
RMSE train: 37.916
RMSE test: 67.563
R^2 train: 0.901
R^2 test: 0.672
```



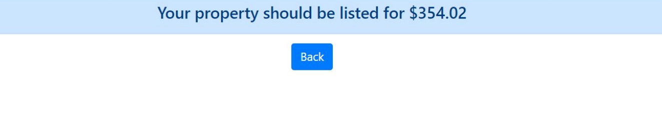
Results:

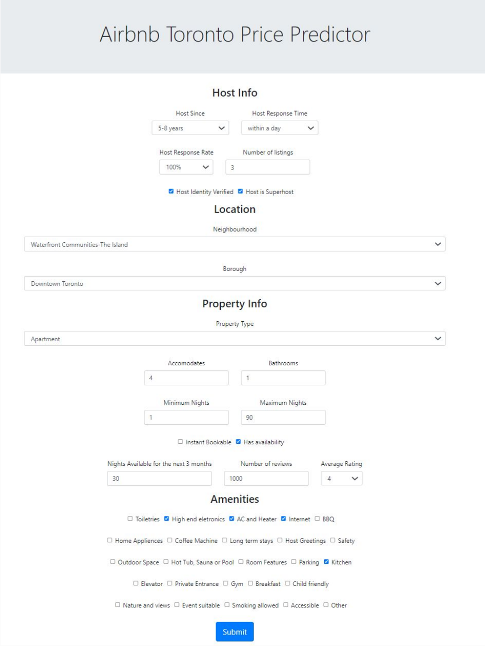
- RMSE test score: 67.35%
- R-Squared score: 67.6%

- Important features:
    - Neighbourhood
    - Property type
    - Accommodates
    - Bathroom
    - Borough
    - Minimum nights
    - Host response rate

Webpage

We created an exported machine learning model on webapp with Flask and HTML. This webapp can be used by potential Airbnb hosts to predict optimal prices for their properties.





Limitations and Future Improvements

- Limited data volume
- Lack of data on important factors like bookings, cancellation policy, security deposit, etc.
- No data regarding points of interests, restaurants or cafes around the property
- Have not performed sentiment analysis on customer reviews and ratings