

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

1. Overview
2. The Problem
3. The Solution
4. Data Source
5. Data Cleaning
6. Exploratory Data Analysis
7. Project Analysis Phase
8. Research Questions
9. Tools Used
10. ERD Diagram
11. ETL Pipeline Diagram
12. Machine Learning Model Results - XGBoost
13. XGBoostRegressor - Hyper-parameter Tuning with GridSearchCV
14. Dashboard Outline
15. Final Dashboard
16. Webapp
17. Limitations and Future Improvements

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.

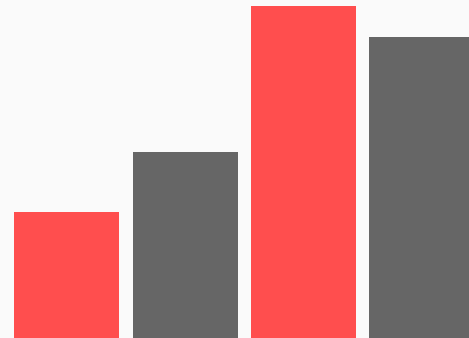


The problem

With the ever changing market it can be challenging for Airbnb hosts to determine the optimal rent prices for their properties.

Since price depends on numerous factors ranging from property type to amenities offered, as well as location customer reviews and ratings.

Hosts can often misinterpret the prices for their neighbourhoods and miss the opportunity to good profits for their listings.



A close-up photograph of a person's hand holding a stylus, poised to write on a tablet. The background is out of focus, showing bokeh light effects. The text 'The solution' is overlaid in white on the left side of the image.

The solution

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

Data Source

The file `clean_data_bourgh.csv`, contains data about Airbnb listings in Toronto, Canada. The dataset contains a total of 15171 records, where each row represents a unique listing and every column represents important data about the listing. The following is the description of some columns in this dataset.

1. **host_since**: The date that host listed their first Airbnb listing.
2. **host_response_rate**: How fast the host responded to customer inquiries.
3. **neighbourhood**: The Toronto neighbourhood of the listing.
4. **property_type**: Type of property (Entire home, private room, share room, hotel room)
5. **price**: Price of the listing property per day
6. **bourgh**: The Toronto bourgh where the of the listing property

Data Cleaning

Cleaned the original listings.csv file using Pandas library and created clean_data_borough.csv for project analysis

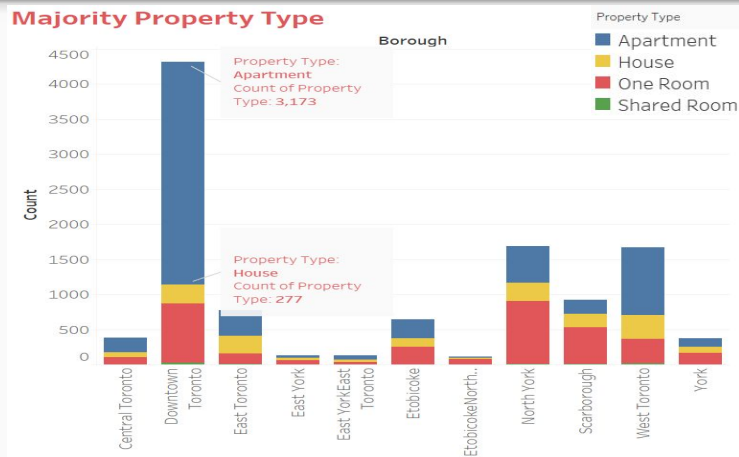
1. Dropped 37 columns that are not relevant for project analysis
2. Scraped Toronto Postal Codes from Wikipedia to determine the relation between postal codes, neighbourhood and boroughs
3. Added postal codes and borough data to original dataset to match each neighbourhood with a borough



Exploration Data Analysis

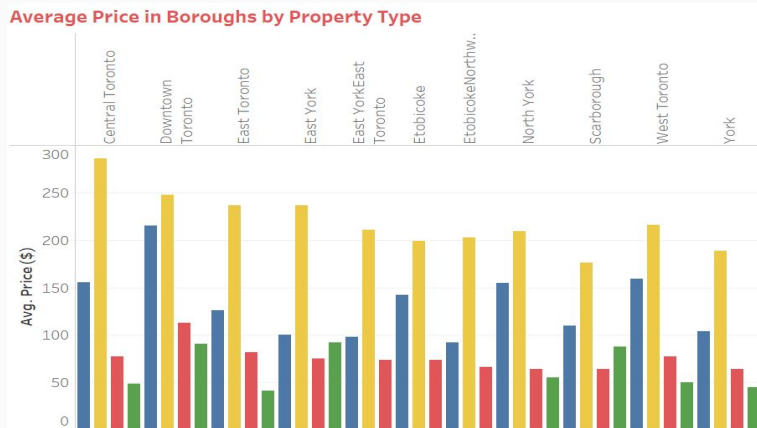
The data exploration phase of the project is conducted in Python and Tableau. We primarily analyzed the data based on four segments, host details, location, reviews and amenities. Within the host details segment our goal was to determine whether factors like being a superhost, having a verified identity and the number of listings the host has in the city has any impact on average prices. In the location segment we derived insights like the most and least expensive neighbourhoods in the city, the average prices in each Toronto borough, etc. The reviews segment we dived a little deeper into understanding the impact customer ratings and reviews have on prices and popularity. Lastly, we looked into popular amenities and whether or not they are offered by most Airbnb listings and the impact they have on average prices.

Exploration Data Analysis



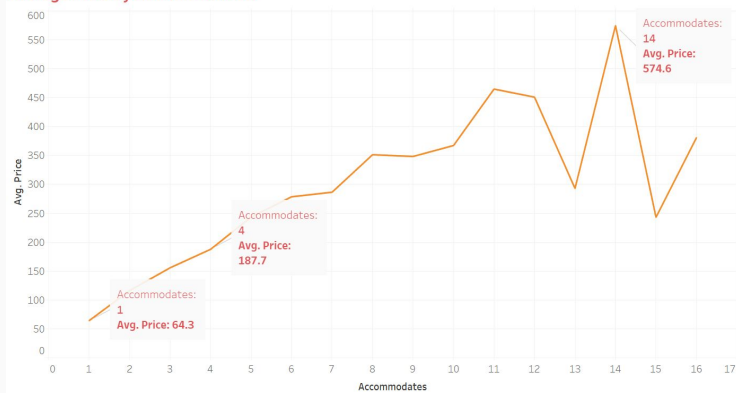
Observations:

- Majority property type - Apartment
- Most expensive property type - House
- Most expensive location - Downtown Toronto

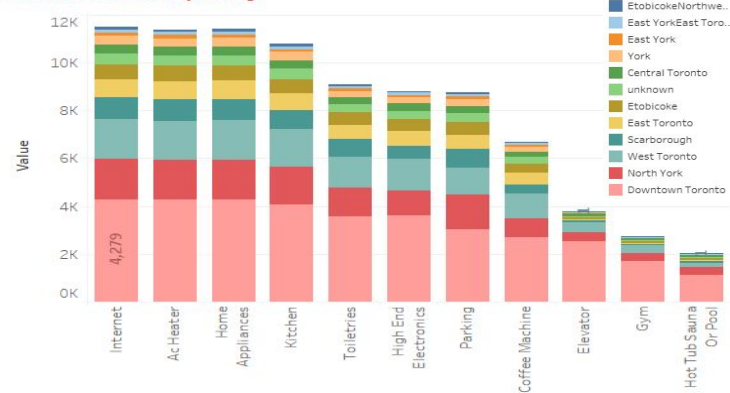


Exploration Data Analysis

Average Price by Accommodation



Number of Amenities by Borough



Observations:

- Highest price by accommodation - 14 guests (average price: CAD \$570)
- Most popular amenity - Internet
- Hosts charge higher prices for popular amenities

Project Analysis Phase

During the last segment the team was able to uncover some key insights on how different variables can affect average Airbnb prices in Toronto. The team analyzed the data by host, location, room type, property type, average customer reviews and rating, amenities offered, etc. On average we were able to see that there is a considerable difference in average prices based on the presence of each of these factors. Currently, the team is working towards analyzing the data through visualizations to derive any insights that can help us understand the impact different factors have on average price.

Research Questions

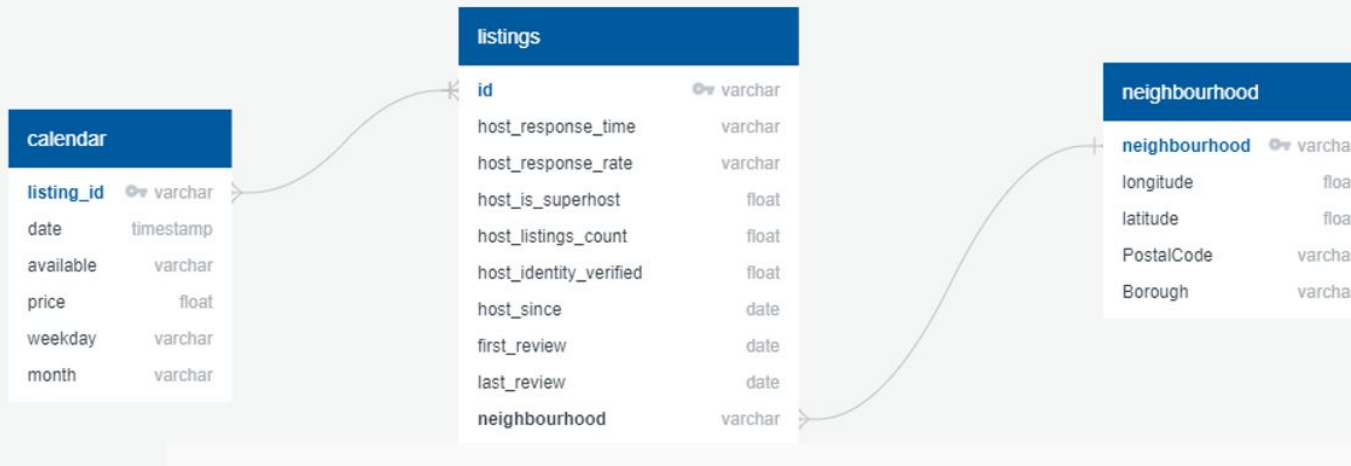
1. Relation between price and Room Type
2. Relation between price and Property Type
3. Top five most popular amenities
4. Top five most expensive locations
5. Relation between price and amenities
6. Relation between price and location
7. Relation between price and customer reviews and ratings
8. Popular properties by number of reviews
9. Which month has the most bookings
10. Highest number of listings by boroughs
11. Top five expensive and least expensive neighbourhoods and boroughs
12. Relation between price and host response time
13. Relation between price and host response rate

Tools Used

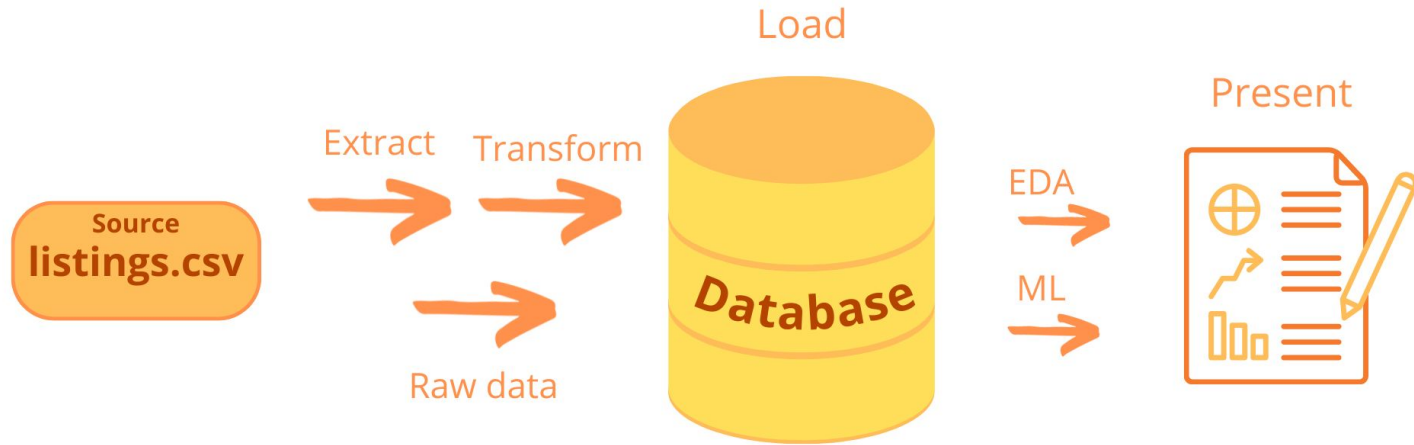
1. Github ([Project link](#))
2. Database - PostgreSQL ([Database details link](#)) and Amazon Web Services RDS
3. Tableau ([Dashboard link](#))
4. Python (Pandas, Matplotlib, Seaborn, Scikit-learn, Numpy)
5. Flask
6. HTML
7. CSS
8. Jupyter Notebook
9. Google Colab
10. XGBoost Regression Model (Scikit-learn)
11. GridSearch CV hyper-parameter tuning

ERD Diagram

www.quickdatabasediagrams.com



ETL Pipeline



Machine Learning Model Results - XGBoost

Model	RMSE	R^2	RMSE(%)	R^2 (%)
Linear Regression	79.39	0.550		
Support Vector Regression	100.08	0.284	-26%	-48%
GradientBoostingRegressor	69.52	0.655	12%	19%
Random Forest Regressor	69.17	0.658	13%	20%
Extra Trees Regressor	68.77	0.662	13%	20%
XGBoost Regressor	68.26	0.667	14%	21%
XGBoost Regressor + gridSearchCV	67.35	0.676	15%	23%

- Important features:
 - Neighbourhood/Borough
 - Property type
 - Accommodates
 - Bathroom
 - Minimum nights

XGBoostRegressor- Hyper-parameter Tuning with GridSearchCV

Final Model After Hyper-parameter Tuning

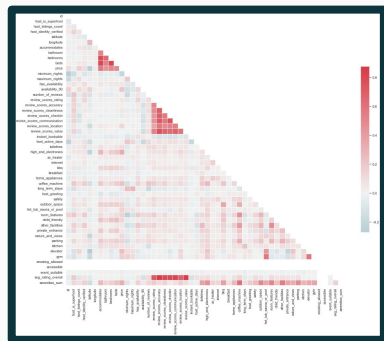
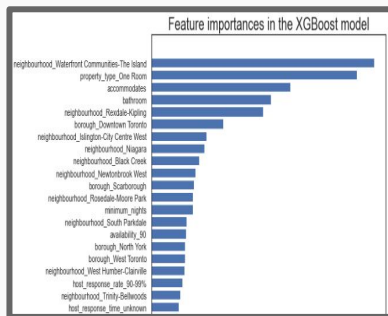
```
## Test optimum values for param_grid
param_grid = {
    'n_estimators': [100,250,500,1000],
    'eta':[0.05,0.1,0.2]
}

grid_search(param_grid,best_param)

## 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

Story Outline

Page 1



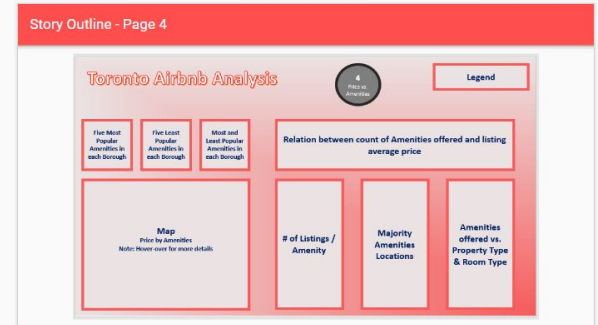
Page 2



Page 3



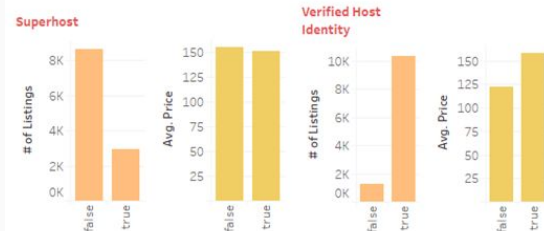
Page 4



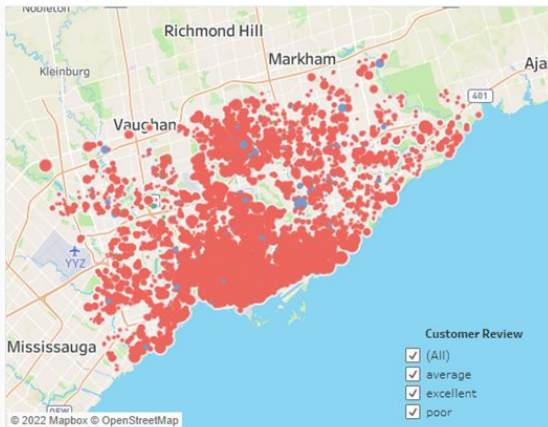
Toronto Airbnb Analysis

Relation between Price, Customer Ratings and Amenities

In this section, we see the relationship between average Airbnb prices and other factors such as customer reviews and ratings, host details amenities offered, etc. We determine things like the impact an excellent vs. poor rating has on number of listing and their prices



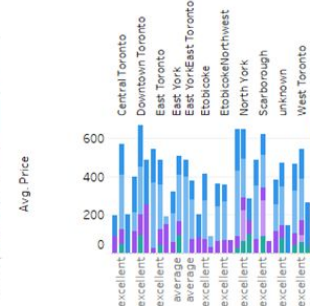
Average Ratings by Borough



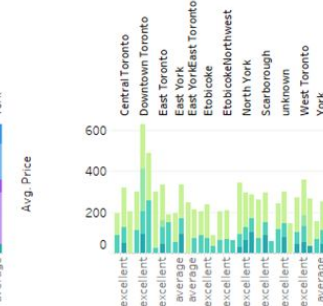
Avg Prices for listings since First Review



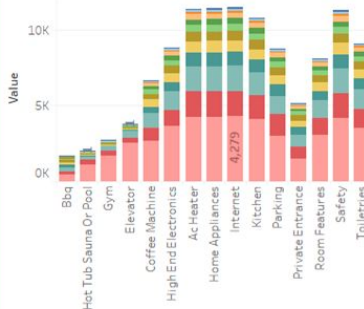
Avg Price in Boroughs by Customer Ratings



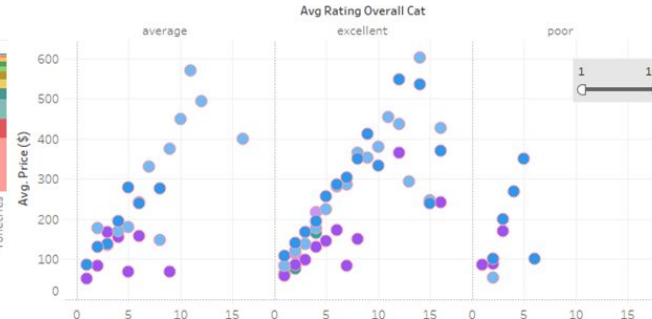
Avg Price in Boroughs by Customer Ratings



Number of Amenities by Borough



Avg Price based on Accommodations by Customer Ratings



Borough	Room Type	Customer Rating	Filter by Property Type
<input checked="" type="checkbox"/> EtobicokeNorth... <input type="checkbox"/> East YorkEast... <input type="checkbox"/> East York	<input type="checkbox"/> Entire home/.. <input type="checkbox"/> Hotel room <input type="checkbox"/> Private room	<input type="checkbox"/> average <input type="checkbox"/> excellent <input type="checkbox"/> poor	<input checked="" type="checkbox"/> (All) <input type="checkbox"/> Apartment <input type="checkbox"/> House
Customer Review	Property Type	Avg. Price	Filter by Room Type
<input checked="" type="checkbox"/> (All) <input type="checkbox"/> average <input type="checkbox"/> excellent	<input type="checkbox"/> Apartment <input type="checkbox"/> House <input type="checkbox"/> One Room	<input type="checkbox"/> 14.0 <input type="checkbox"/> 200.0 <input type="checkbox"/> 400.0	<input checked="" type="checkbox"/> (All) <input type="checkbox"/> Entire home/apt <input type="checkbox"/> Hotel room

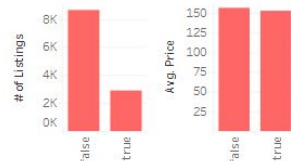
Final Dashboard

Toronto Airbnb Analysis

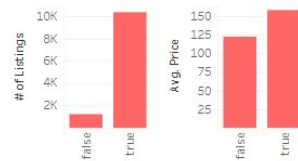
Relation between Price and Host Details

In this section, we see the relationship between average Airbnb prices and some host details. What impact does being a superhost and having verified identity mean when it comes to the number of listing and their prices

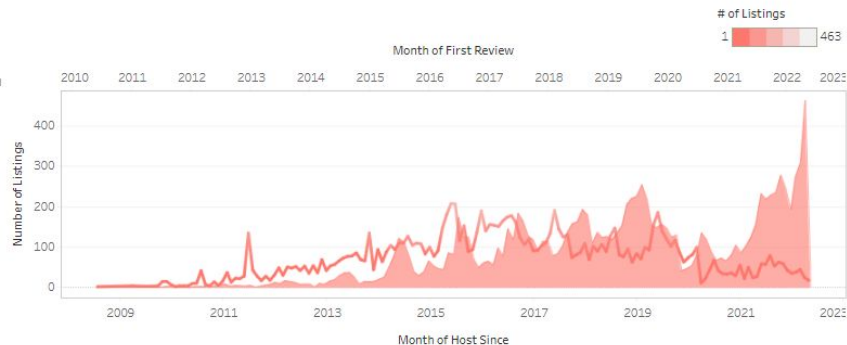
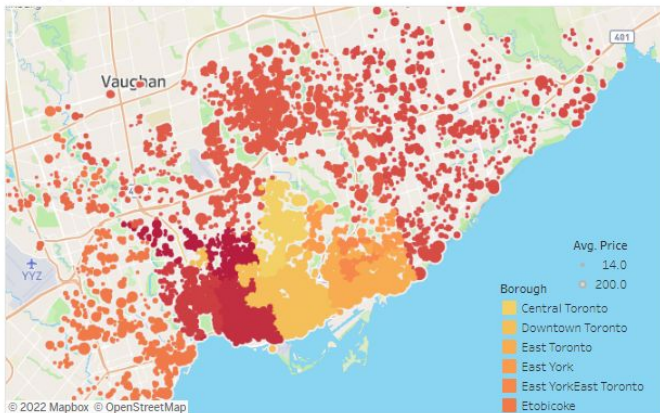
Superhost



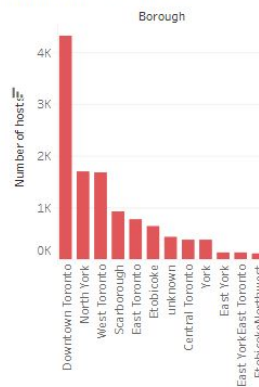
Verified Host Identity



Average Prices by Borough



Majority Host Location



Majority Host Location & Property Type

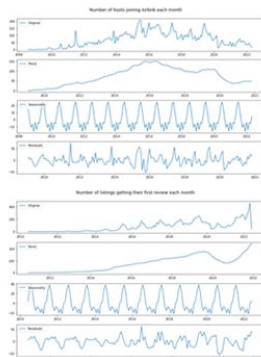


Final Dashboard

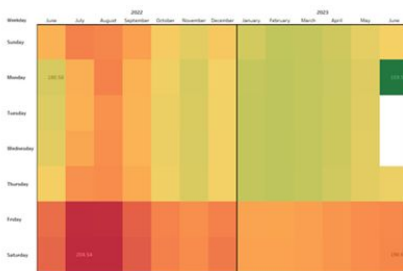
Toronto Airbnb Analysis

Relation between Price and Other factors

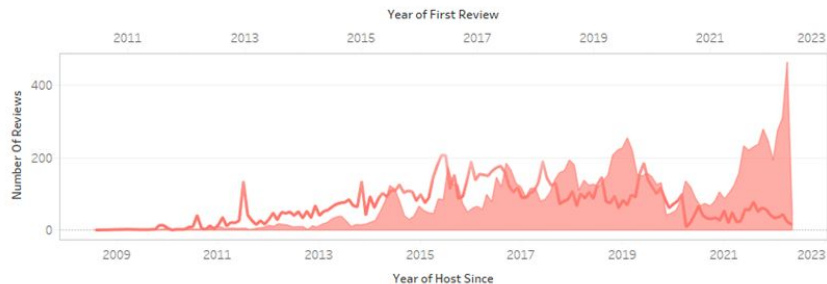
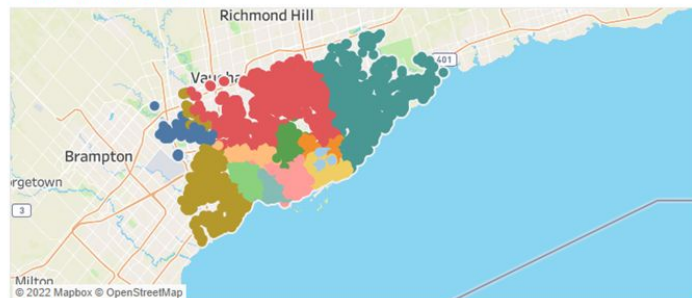
In this section, we see the relationship between average Airbnb prices and some other factors like seasonality. What impact does a certain month, or day of the week have on the number of listing and their prices.



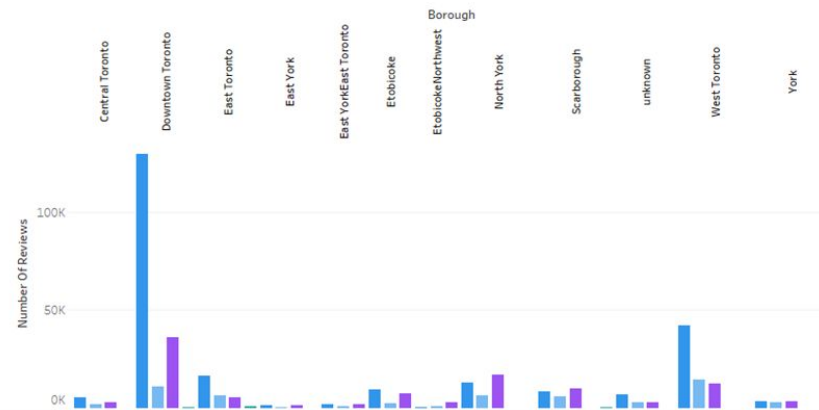
Average Prices by Month and Weekday



Number of Listings by Borough

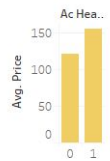
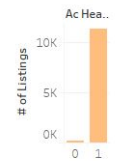


Majority Property Type

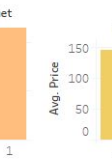
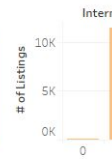


Price vs. Amenities

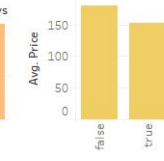
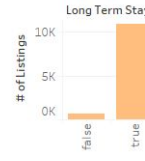
Average Price vs. AC Heater



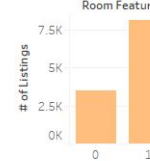
Average Price vs. Internet



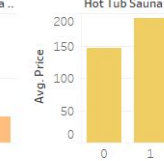
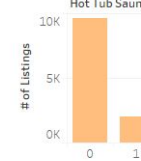
Average Price vs. Long Term Stays



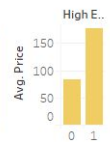
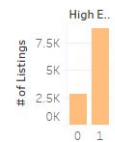
Average Price vs. Room Features



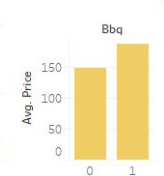
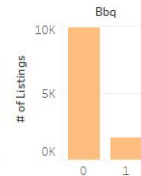
Average Price vs. Hot Tub/Sauna/Pool



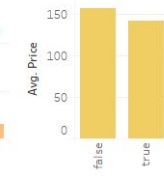
Average Price vs. High end electronics



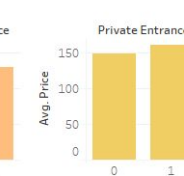
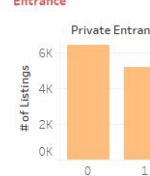
Average Price vs. BBQ



Average Price vs. Host Greetings



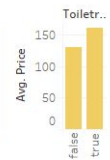
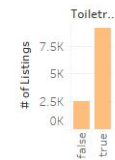
Average Price vs. Private Entrance



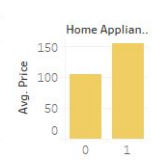
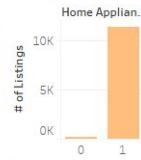
Average Price vs. Other Facilities



Average Price vs. Toiletries



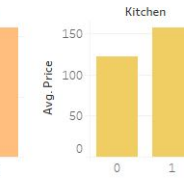
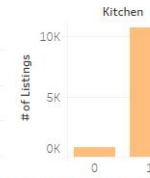
Average Price vs. Home Appliances



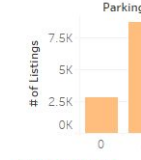
Average Price vs. and Safety



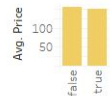
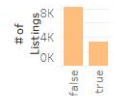
Average Price vs. Kitchen



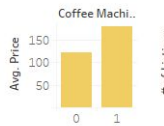
Average Price vs. Parking



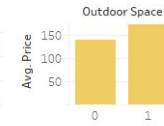
Instant Booking Privileges



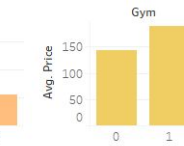
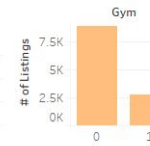
Average Price vs. Coffee Machine



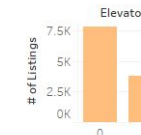
Average Price vs. Outdoor Space



Average Price vs. Gym



Average Price vs. Elevator



Webapp

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.

Your property should be listed for \$354.02

Back

Airbnb Toronto Price Predictor

Host Info

Host Since	Host Response Time
5-8 years	within a day

Host Response Rate	Number of listings
100%	3

☒ Host Identity Verified ☒ Host is Superhost

Location

Neighbourhood

Waterfront Communities-The Island

Borough

Downtown Toronto

Property Info

Property Type

Apartment

Accommodates

4

Bathrooms

1

Minimum Nights

1

Maximum Nights

90

☐ Instant Bookable ☒ Has availability

Nights Available for the next 3 months

30

Number of reviews

1000

Average Rating

4

Amenities

- ☐ Toiletries ☒ High end electronics ☒ AC and Heater ☒ Internet ☐ BBQ
- ☐ Home Appliances ☐ Coffee Machine ☐ Long term stays ☐ Host Greetings ☐ Safety
- ☐ Outdoor Space ☐ Hot Tub, Sauna or Pool ☐ Room Features ☐ Parking ☒ Kitchen
- ☐ Elevator ☐ Private Entrance ☐ Gym ☐ Breakfast ☐ Child friendly
- ☐ Nature and views ☐ Event suitable ☐ Smoking allowed ☐ Accessible ☐ Other

Submit

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