Airbnb Analysis

Capstone Project Presentation

Link to Project Presentation

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.

Airbnb allows individuals to rent out their vacant properties to individuals looking for short term accommodation.

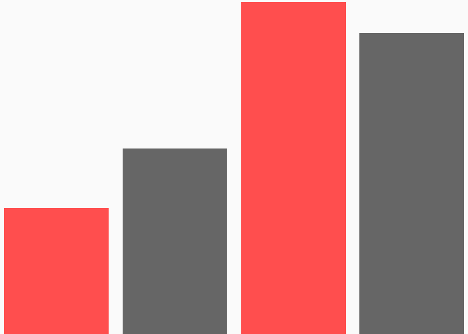
Properties can range from houses, apartments to single and shared rooms and are priced per night or per stay.

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.



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 8387 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

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.

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

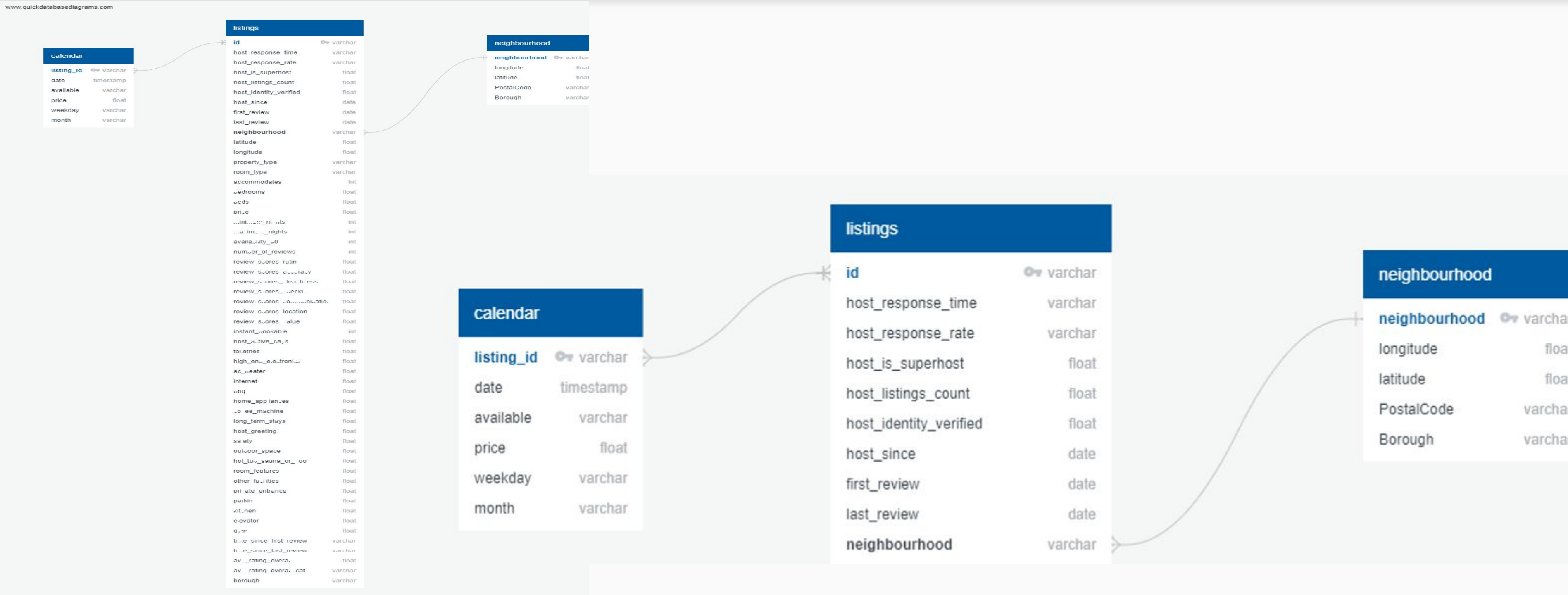
1. Relation between price and Room Type
2. Relation between price and Property Type
3. Top five most popular amenities
4. Top five most expensive locations
5. Relation between price and amenities
6. Relation between price and location
7. Relation between price and customer reviews and ratings
8. Popular 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

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 bourgh data to original dataset to match each neighbourhood with a bourgh

ERD Diagram



www.quickdatabasediagrams.com

calendar	
listing_id 🔑	varchar
date	timestamp
available	varchar
price	float
weekday	varchar
month	varchar

listings	
id 🔑	varchar
host_response_time	varchar
host_response_rate	varchar
host_is_superhost	float
host_listings_count	float
host_identity_verified	float
host_since	date
first_review	date
last_review	date
neighbourhood	varchar
latitude	float
longitude	float
property_type	varchar
room_type	varchar
accommodates	int
bedrooms	float
beds	float
price	float
minimum_nights	int
maximum_nights	int
availability_30	int
number_of_reviews	int
review_scores_rating	float
review_scores_accuracy	float
review_scores_cleanliness	float
review_scores_checkin	float
review_scores_communication	float
review_scores_location	float
review_scores_value	float
instant_bookable	int
host_active_days	float
toiletries	float
high_end_electronics	float
ac_heater	float
internet	float
bbq	float
home_appliances	float
coffee_machine	float
long_term_stays	float
host_greeting	float
safety	float
outdoor_space	float
hot_tub_sauna_or_pool	float
room_features	float
other_facilities	float
private_entrance	float
parking	float
kitchen	float
elevator	float
gym	float
time_since_first_review	varchar
time_since_last_review	varchar
avg_rating_overall	float
avg_rating_overall_cat	varchar
borough	varchar

neighbourhood	
neighbourhood 🔑	varchar
longitude	float
latitude	float
PostalCode	varchar
Borough	varchar

calendar	
listing_id 🔑	varchar
date	timestamp
available	varchar
price	float
weekday	varchar
month	varchar

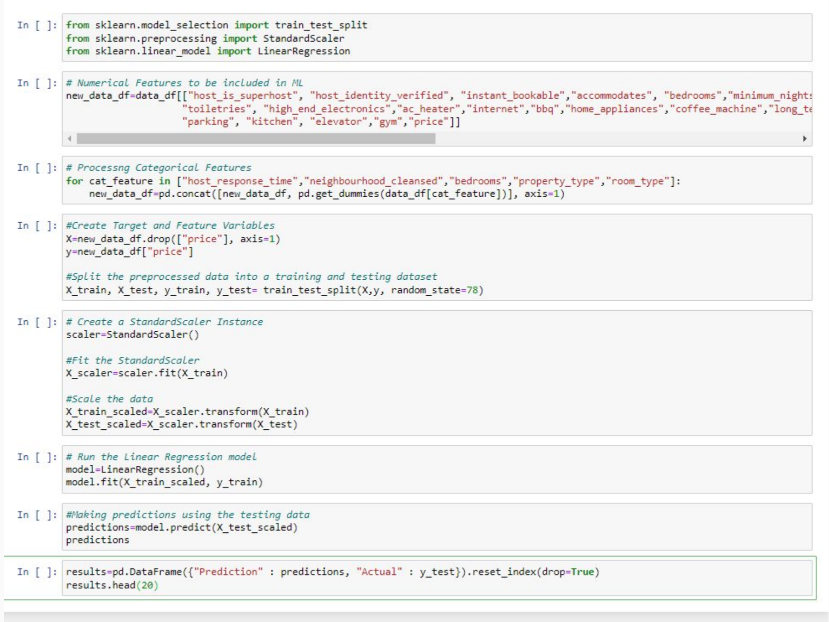
listings	
id 🔑	varchar
host_response_time	varchar
host_response_rate	varchar
host_is_superhost	float
host_listings_count	float
host_identity_verified	float
host_since	date
first_review	date
last_review	date
neighbourhood	varchar

neighbourhood	
neighbourhood 🔑	varcha
longitude	floa
latitude	floa
PostalCode	varcha
Borough	varcha

Machine Learning Model

Goal is to implement and compare results of these regression models to see which model helps us accurately predict prices

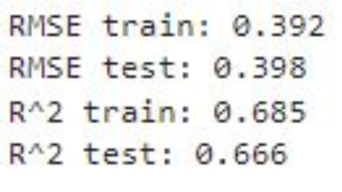
1. Linear Regression
2. Support Vector Regression
3. HistGradientBoostingRegressor
4. RandomForestRegressor
5. ExtraTreesRegressor
6. XGBoostRegressor



```
In [ ]: from sklearn.model_selection import train_test_split
        from sklearn.preprocessing import StandardScaler
        from sklearn.linear_model import LinearRegression

In [ ]: # Numerical Features to be included in ML
        new_data_df=data_df[["host_is_superhost", "host_identity_verified", "instant_bookable","accommodates", "bedrooms","minimum_night
                            "toiletries", "high_end_electronics","ac_heater","internet","bbq","home_appliances","coffee_machine","long_te
                            "parking", "kitchen", "elevator","gym","price"]]

In [ ]: # Processng Categorical Features
        for cat_feature in ["host_response_time","neighbourhood_cleansed","bedrooms","property_type","room_type"]:
            new_data_df=pd.concat([new_data_df, pd.get_dummies(data_df[cat_feature])], axis=1)

In [ ]: #Create Target and Feature Variables
        X=new_data_df.drop(["price"], axis=1)
        y=new_data_df["price"]

        #Split the preprocessed data into a training and testing dataset
        X_train, X_test, y_train, y_test= train_test_split(X,y, random_state=78)

In [ ]: # Create a StandardScaler Instance
        scaler=StandardScaler()

        #Fit the StandardScaler
        X_scaler=scaler.fit(X_train)

        #Scale the data
        X_train_scaled=X_scaler.transform(X_train)
        X_test_scaled=X_scaler.transform(X_test)

In [ ]: # Run the Linear Regression model
        model=LinearRegression()
        model.fit(X_train_scaled, y_train)

In [ ]: #Making predictions using the testing data
        predictions=model.predict(X_test_scaled)
        predictions

In [ ]: results=pd.DataFrame({"Prediction" : predictions, "Actual" : y_test}).reset_index(drop=True)
        results.head(20)
```

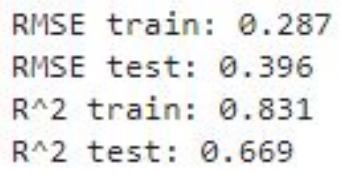
Model Results

Linear Regression Model



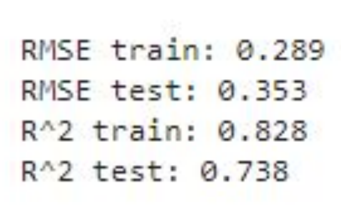
```
RMSE train: 0.392
RMSE test: 0.398
R^2 train: 0.685
R^2 test: 0.666
```

Support Vector Regression Model



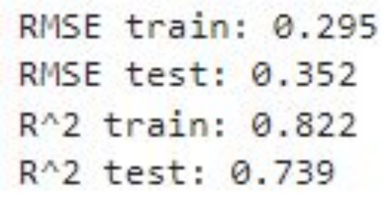
```
RMSE train: 0.287
RMSE test: 0.396
R^2 train: 0.831
R^2 test: 0.669
```

XGBoost Regressor



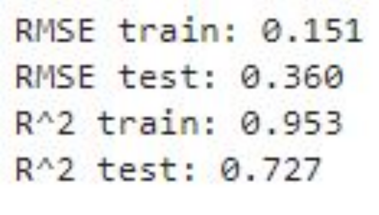
```
RMSE train: 0.289
RMSE test: 0.353
R^2 train: 0.828
R^2 test: 0.738
```

HistGradientBoosting Regressor



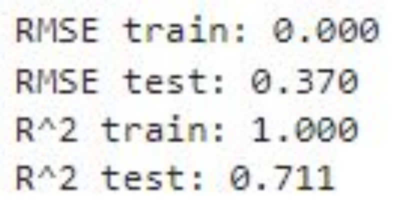
```
RMSE train: 0.295
RMSE test: 0.352
R^2 train: 0.822
R^2 test: 0.739
```

RandomForest Regressor



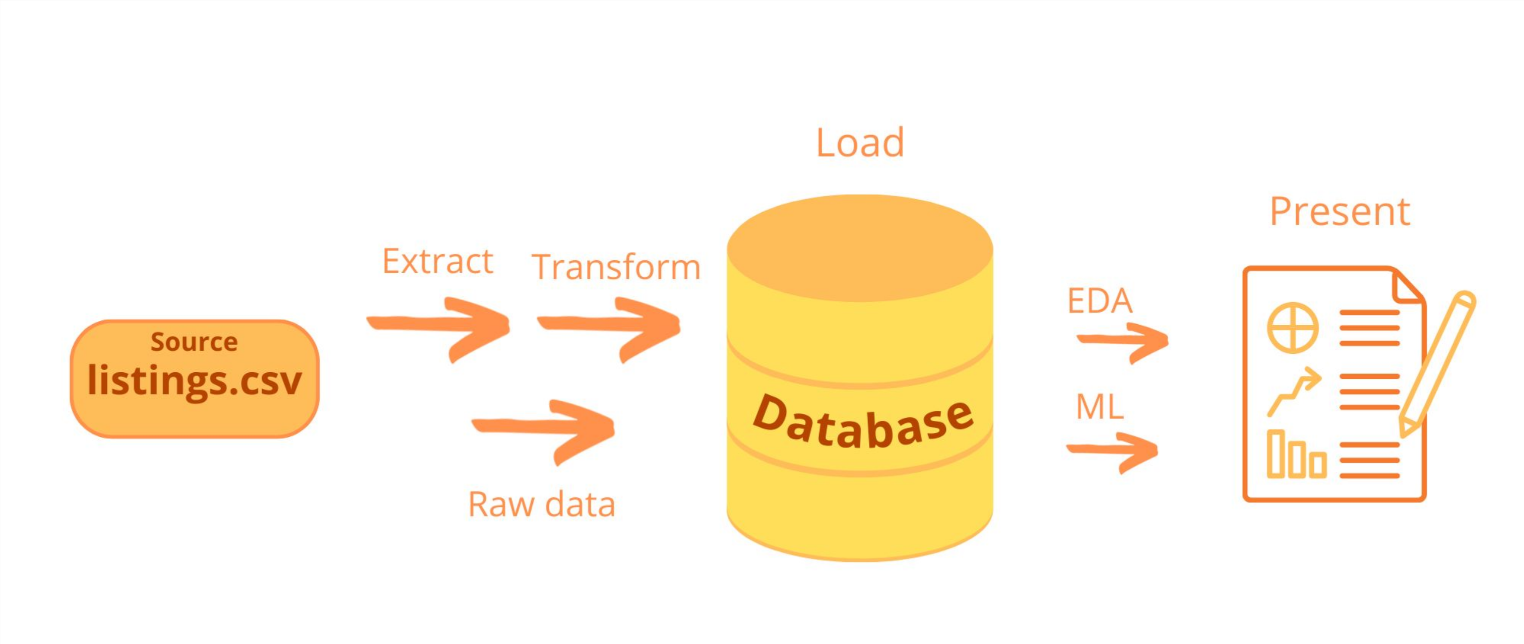
```
RMSE train: 0.151
RMSE test: 0.360
R^2 train: 0.953
R^2 test: 0.727
```

ExtraTrees Regressor



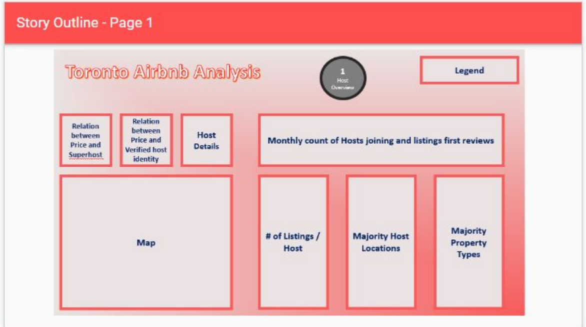
```
RMSE train: 0.000
RMSE test: 0.370
R^2 train: 1.000
R^2 test: 0.711
```

ETL Pipeline



Story Outline

Page 1



Story Outline - Page 1

Toronto Airbnb Analysis

1 Host Overview

Legend

Relation between Price and Superhost

Relation between Price and Verified host identity

Host Details

Monthly count of Hosts joining and listings first reviews

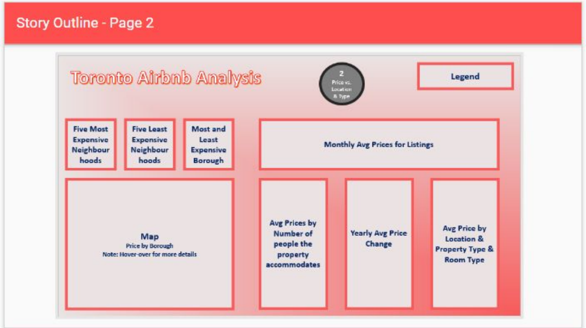
Map

# of Listings / Host

Majority Host Locations

Majority Property Types

Page 2



Story Outline - Page 2

Toronto Airbnb Analysis

2 Price vs. Location & Type

Legend

Five Most Expensive Neighbour hoods

Five Least Expensive Neighbour hoods

Most and Least Expensive Borough

Monthly Avg Prices for Listings

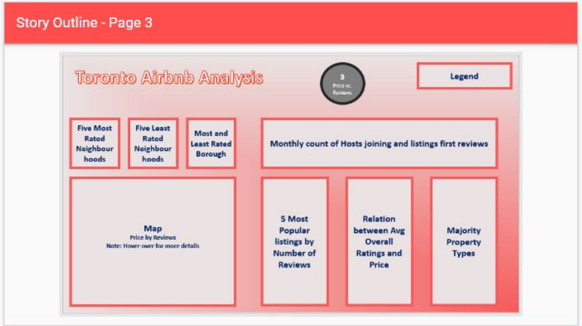
Map
Price by Borough
Note: Hover-over for more details

Avg Prices by Number of people the property accommodates

Yearly Avg Price Change

Avg Price by Location & Property Type & Room Type

Page 3



Story Outline - Page 3

Toronto Airbnb Analysis

3 Price vs. Reviews

Legend

Five Most Rated Neighbour hoods

Five Least Rated Neighbour hoods

Most and Least Rated Borough

Monthly count of Hosts joining and listings first reviews

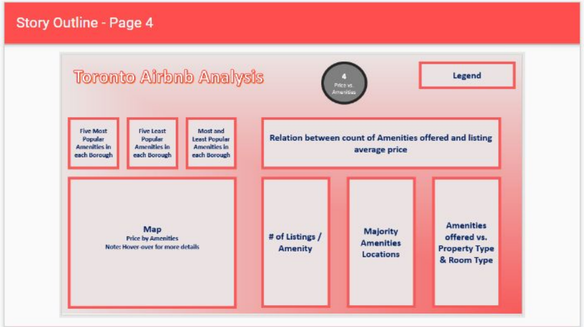
Map
Price by Reviews
Note: Hover-over for more details

5 Most Popular listings by Number of Reviews

Relation between Avg Overall Ratings and Price

Majority Property Types

Page 4



Story Outline - Page 4

Toronto Airbnb Analysis

4 Price vs. Amenities

Legend

Five Most Popular Amenities in each Borough

Five Least Popular Amenities in each Borough

Most and Least Popular Amenities in each Borough

Relation between count of Amenities offered and listing average price

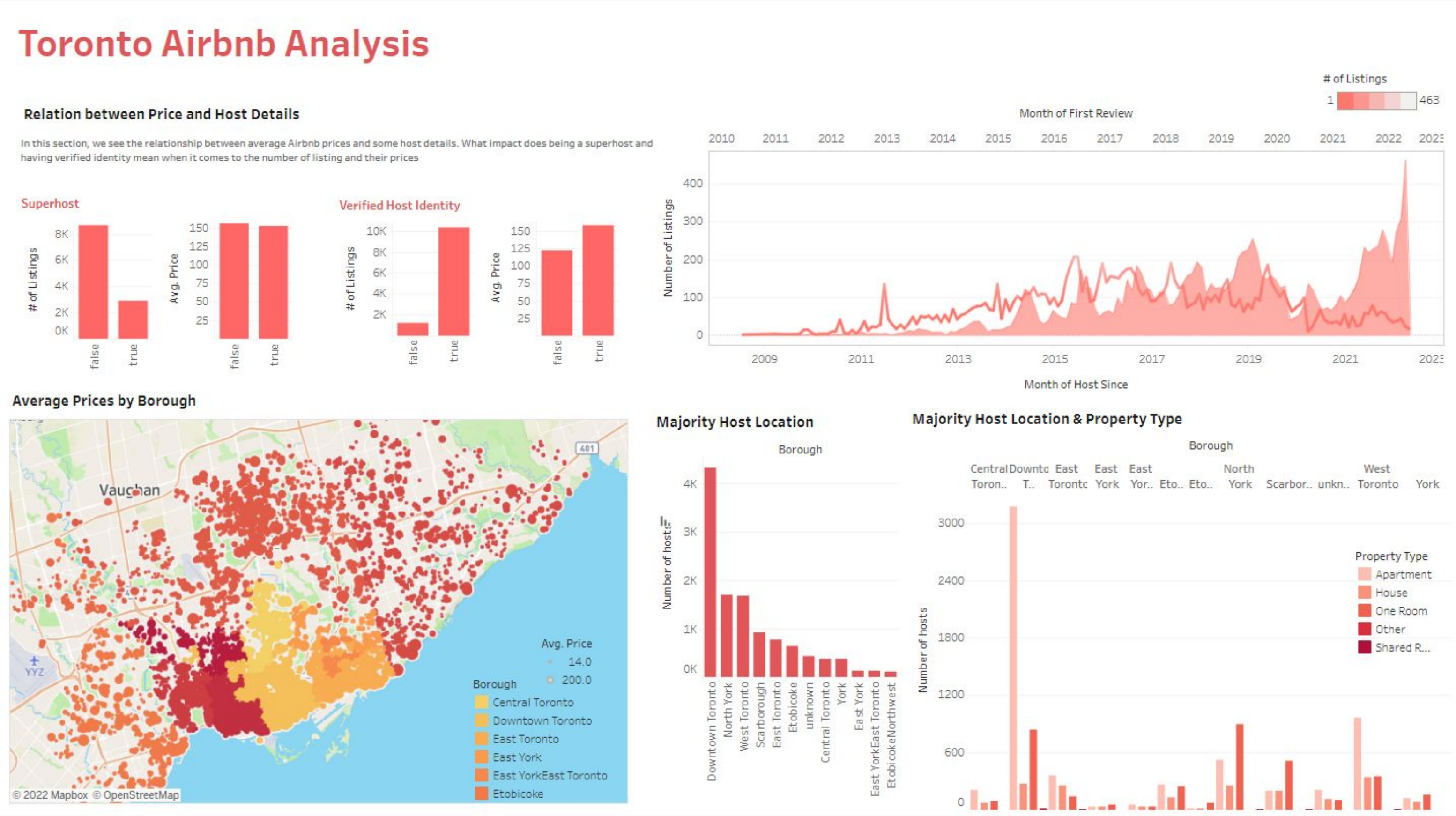
Map
Price by Amenities
Note: Hover-over for more details

# of Listings / Amenity

Majority Amenities Locations

Amenities offered vs. Property Type & Room Type

Dashboard 1 - Page 1



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

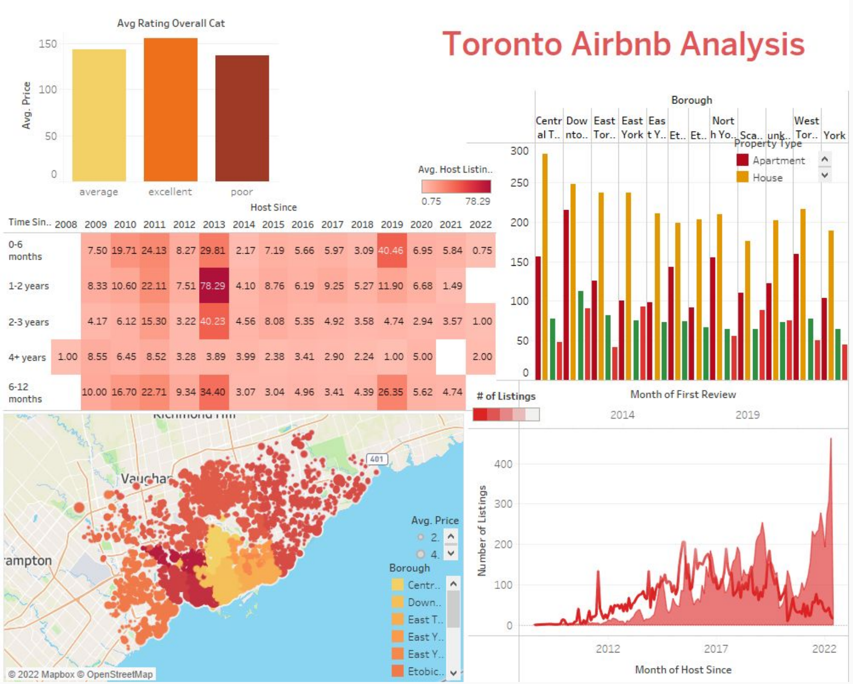
Average Prices by Borough

Majority Host Location

Majority Host Location & Property Type

Dashboard Drafts

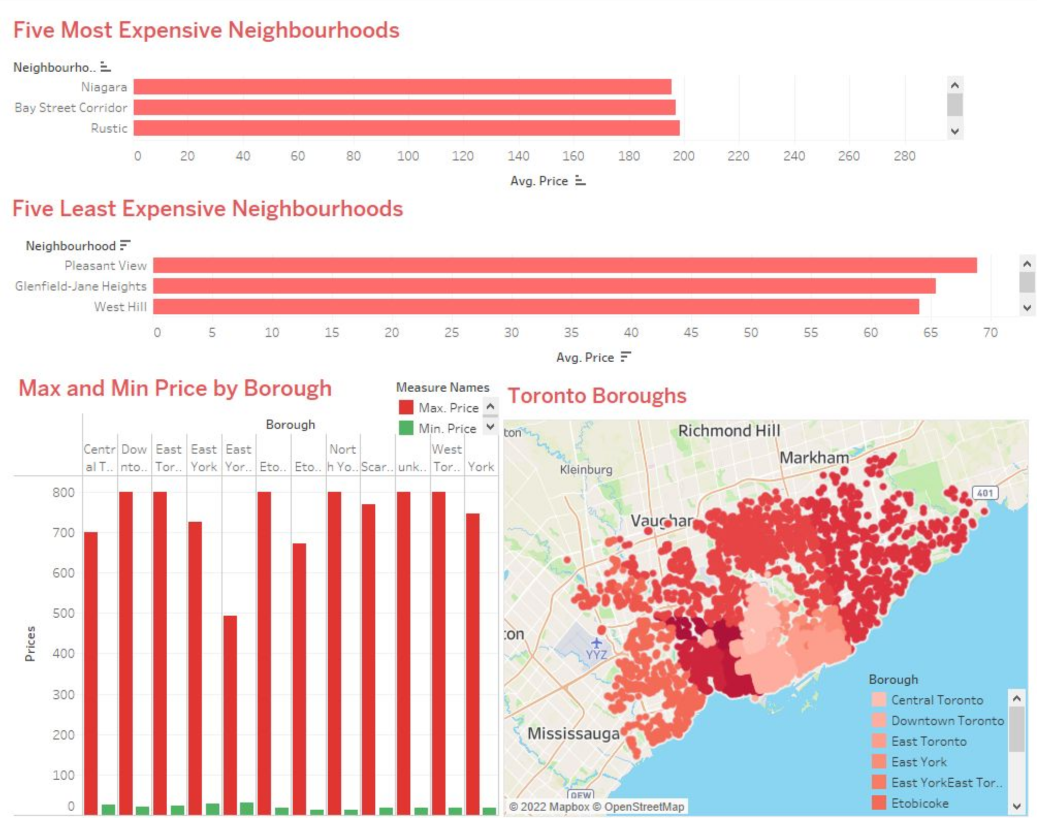
Price vs. Reviews Analysis



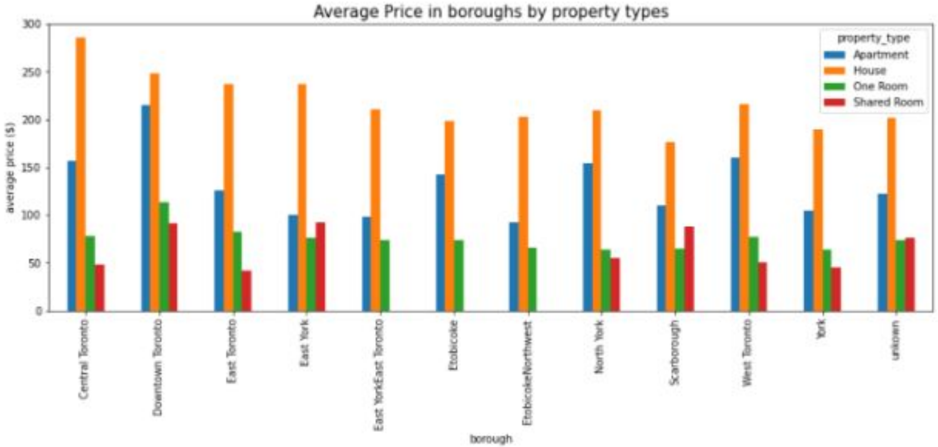
Host Since

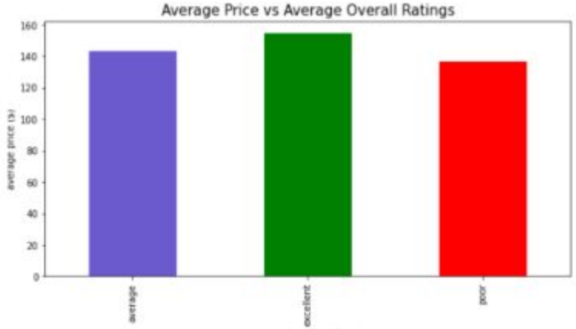
Time Sin..	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
0-6 months		7.50	19.71	24.13	8.27	29.81	2.17	7.19	5.66	5.97	3.09	40.46	6.95	5.84	0.75
1-2 years		8.33	10.60	22.11	7.51	78.29	4.10	8.76	6.19	9.25	5.27	11.90	6.68	1.49	
2-3 years		4.17	6.12	15.30	3.22	40.23	4.56	8.08	5.35	4.92	3.58	4.74	2.94	3.57	1.00
4+ years	1.00	8.55	6.45	8.52	3.28	3.89	3.99	2.38	3.41	2.90	2.24	1.00	5.00		2.00
6-12 months		10.00	16.70	22.71	9.34	34.40	3.07	3.04	4.96	3.41	4.39	26.35	5.62	4.74	

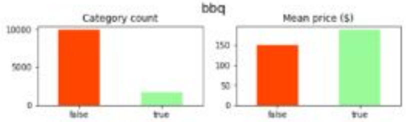
Price vs. Neighbourhood Analysis

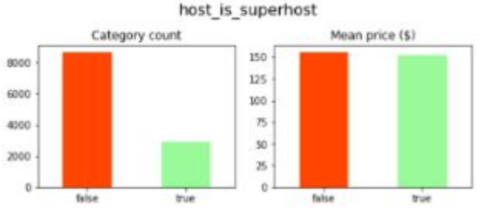


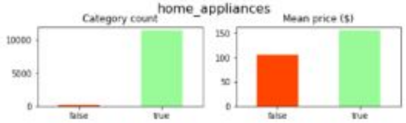
Other Visualizations

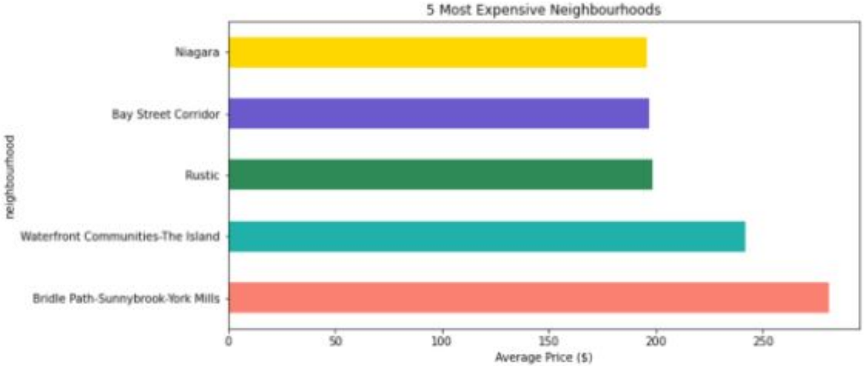


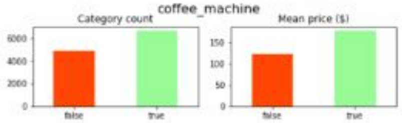


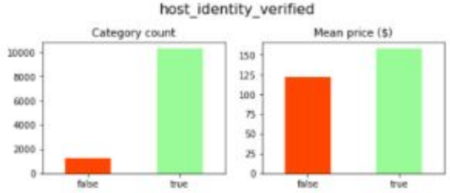


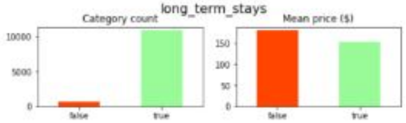


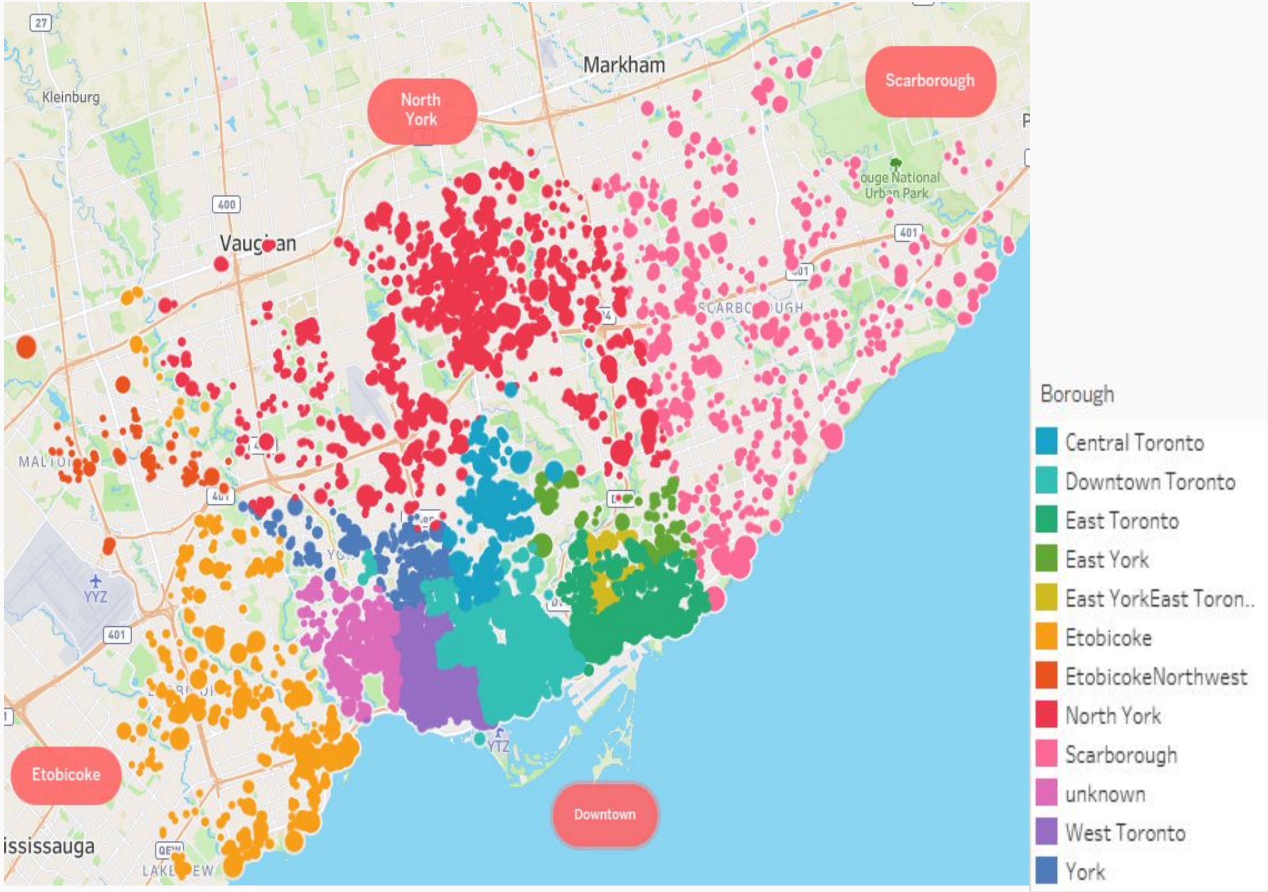


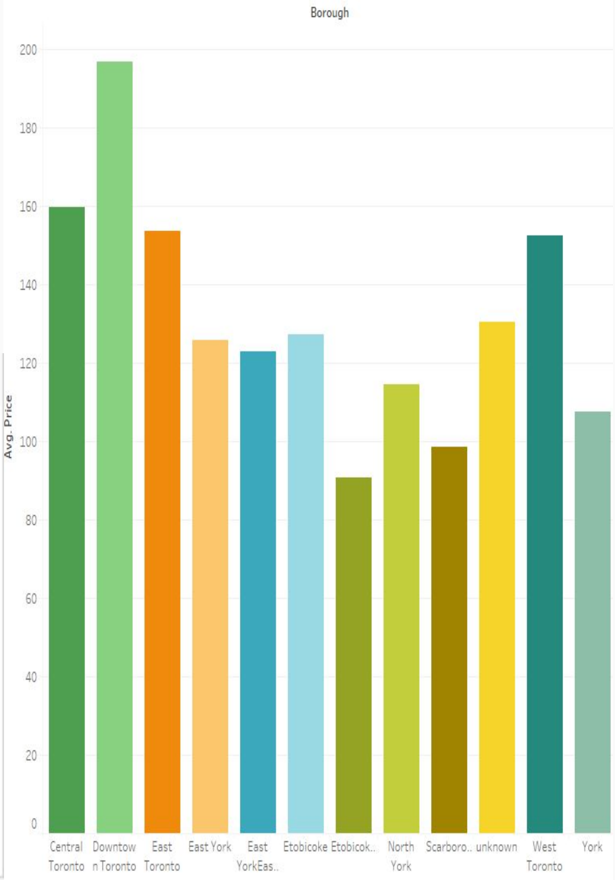









Markham
Kleinburg
North York
Scarborough
27
400
Vaughan
401
ouge National Urban Park
SCARBOROUGH
MALTON
YYZ
YTZ
Etobicoke
Downtown
issisaauga
QEW
LAKEVIEW
Borough
Central Toronto
Downtown Toronto
East Toronto
East York
East YorkEast Toron..
Etobicoke
EtobicokeNorthwest
North York
Scarborough
unknown
West Toronto
York


Borough
Avg. Price
200
180
160
140
120
100
80
60
40
20
0
Central Toronto
Downtown Toronto
East Toronto
East York
East YorkEas..
Etobicoke
Etobicok..
North York
Scarboro..
unknown
West Toronto
York