Predicting Rent prices in Toronto Canada

Salim Al-harthi

16 July 2021

1. Introduction

1.1 Background

People are renting houses and apartments to live in for years or months. It is important for people who rent a house or an apartment to get a fair price for the place they are renting.

By utilizing the information of the rent of different properties we can build a model that will accurately predict a fair price for rent according to different factors such as number of bathroom and bedrooms, number of different venues nearby and more.

1.2 problem

People who live in a city will have an idea on what a fair rent should be. However, for the people who are new to the city, they must research and find the fair price for rent, or they might end up paying an overpriced rent.

1.3 Interest

As mentioned, in 1.2 people who are moving to a new city will benefit from this solution since it will save them all the research time.

2. Data acquisition and cleaning

2.1 Data sources

Most of the data such as price, latitude, longitude, number of bedroom, and number of bathroom will be coming from Kaggel dataset here and the data related to the venues nearby the property will be coming from foursquare.com API.

2.2 Data cleaning

The data were downloaded from Kaggel contained the following columns:

Column name	Description	Data type	Example
Bedroom	How many bedrooms available	Integer	2
Bathroom	How many bathrooms available	float	2.0
Den	Whether den is available or not	Boolean	1
Address	Location	string	361 Front St W, Toronto, ON M5V 3R5, Canada
Lat	Latitude	float	43.643051
Long	Longitude	float	-79.391643
Price	Apartment Rental price per month in CAD	String	2450.0$

Table 1 Kaggel columns

Cleaning the data consist of removing ($) from **Price** column then convert it to Float.

The data from the API returned as Json object. And consist of several data. However, the important data points are:

Datapoint	Description	Data type	Example
Lat	Latitude	float	43.643051
Long	Longitude	float	-79.391643
Venue name	Venue name	String	Rathburn Plaz
categories	Venue categories	String	Salon / Barbershop
Categories icon	Icon url	URL	https://ss3.4sqi.net/img/categories_v2/shops/salon_barber_

Table 2 important data from API

For cleaning the API data, the **Categories icon** was reduced to main category for example.

The following **Categories icon:**

https://ss3.4sqi.net/img/categories_v2/shops/salon_barber_

‘https:’, ‘’, ‘ss3.4sqi.net’, ‘img’,’categories_v2’,’shops’,’salon_barber_’ it reduced to **shops.** The process output was 8 main categories:

Category	Description
food	Restaurants, coffee shops etc.
Shops	Supermarkets etc.
Building	Library office building and others
Art entertainment	Museums
Travel	Bus stops train and airports
Nightlife	Bars and clubs
Parks outdoors	Parks
Education	Schools and universities

Table 3 8 main categories

2.3 Feature selection

After data cleaning it is the time to select the features. For the features coming from Kaggel the dropped features were as following:

Feature	Reason to drop
Address	Latitude and Longitude provide the location of the property and Address are unique to each property.

Table 4 Dropped features

The remaining features:

Column name	Description	Data type
Bedroom	How many bedrooms available	Integer
Bathroom	How many bathrooms available	float
Den	Whether den is available or not	Boolean
Lat	Latitude	float
Long	Longitude	float
Price	Apartment Rental price per month in CAD	String

Table 5 Remaining features from Kaggel

For the features coming from the API. The features remained were the 8 main categories and their latitude and longitude. The features remained as the following:

Datapoint	Description	Data type

Lat	Latitude	float
Long	Longitude	float
categories	8 main categories	String

Table 6 Remaining features from API

Next is margining all features into one table. And the result table has the following features:

Column name	**Description**	**Data type**	**Example**
Bedroom	How many bedrooms available	Integer	2
Bathroom	How many bathrooms available	float	2.0
Den	Whether den is available or not	Boolean	1
Lat	Latitude	float	43.643051
Long	Longitude	float	-79.391643
Price	Apartment Rental price per month in CAD	Float	2450.0
food	Count of occurrence of the category near by the property	Integer	1
Shops	Count of occurrence of the category near by the property	Integer	2
Building	Count of occurrence of the category near by the property	Integer	9
Art entertainment	Count of occurrence of the category near by the property	Integer	1
Travel	Count of occurrence of the category near by the property	Integer	2
Nightlife	Count of occurrence of the category near by the property	Integer	0
Parks outdoors	Count of occurrence of the category near by the property	Integer	1
Education	Count of occurrence of the category near by the property	Integer	0

Table 7 All features

3. Exploratory Data Analysis

3.1 Relationship between features (Bathrooms, Bedrooms, Den, Latitude, and Longitude) and target (rent)

As shown in (Figure 1.1) the increase of number of bathrooms and bedrooms cause the minimum and maximum amount rent increases.

The properties with den has a higher minimum rent in compare to the properties without den as shown in (Figure 1).

As shown in (Figure 1) most of the properties are closer to (44, -80) latitude, and longitude.

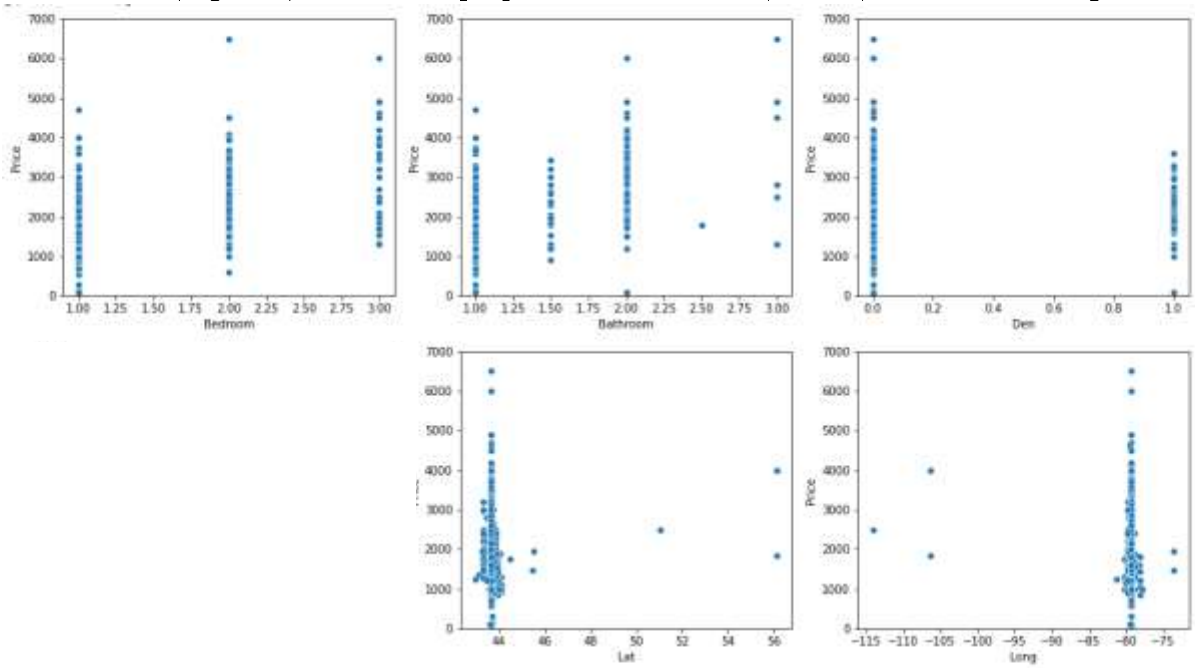

Figure 1 Relationship between features (Bathrooms, Bedrooms, Den, Latitude, and Longitude) and Price

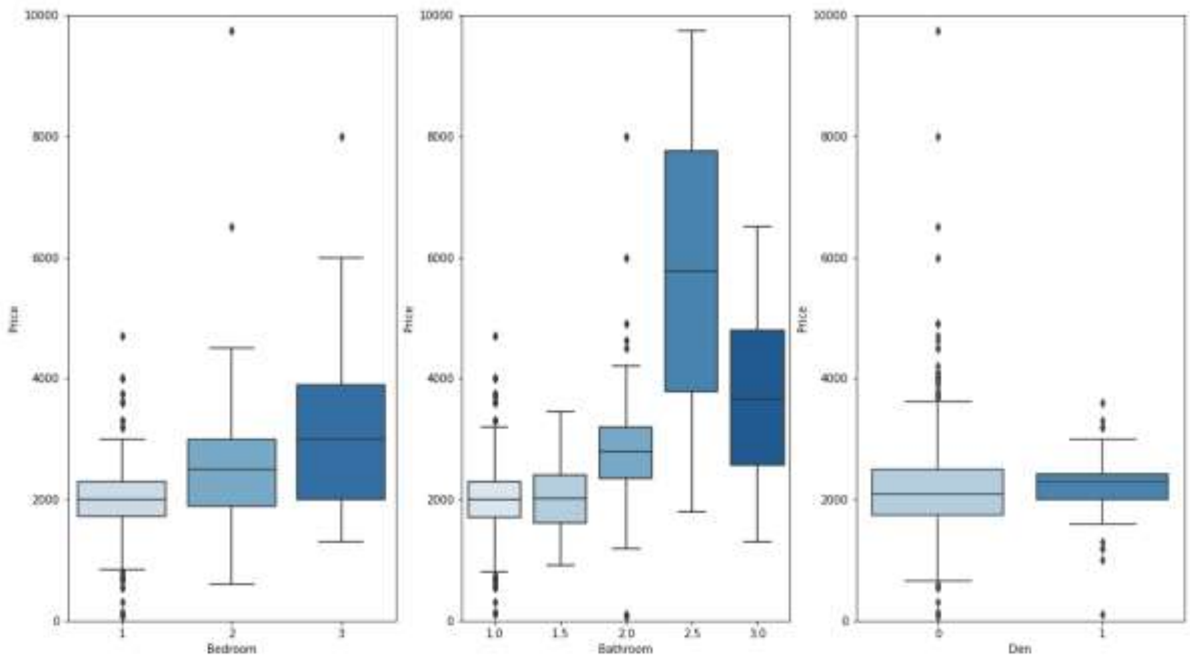

Figure 1.1 box plot features (Bathrooms, Bedrooms, and Den) and Price

3.2 Relationship between features (Latitude, and Longitude) and target (rent)

As shown in (Figure 2) most of the property from latitude, and longitude (43.65,-79.4) the lower the maximum rent price is.

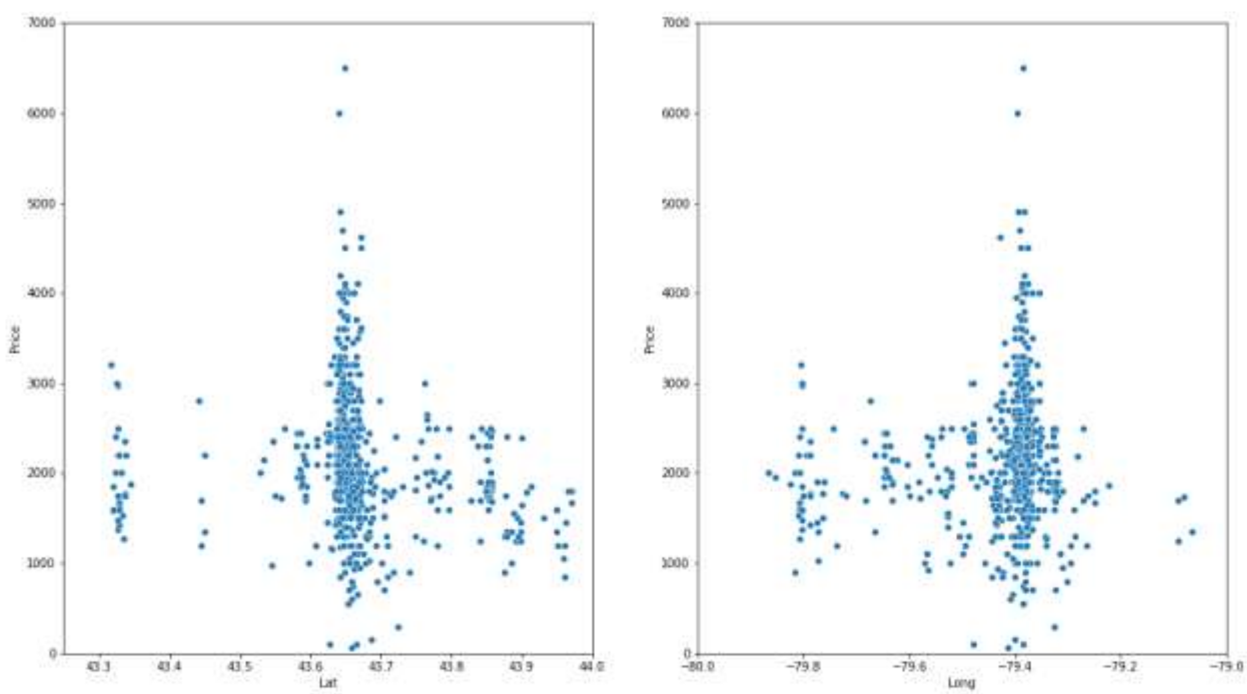

Figure 2 Relationship between features (Latitude, and Longitude) and Price

3.3 Relationship between 8 main categories and target (rent)

As shown in (Figure 3) the in crease in number of travel, parks_outdoors, and arts_entertainment, cause the minimum and maximum amount rent increases.

As shown in (figure 3.1) the change in number of occurrences changes the range of the price.

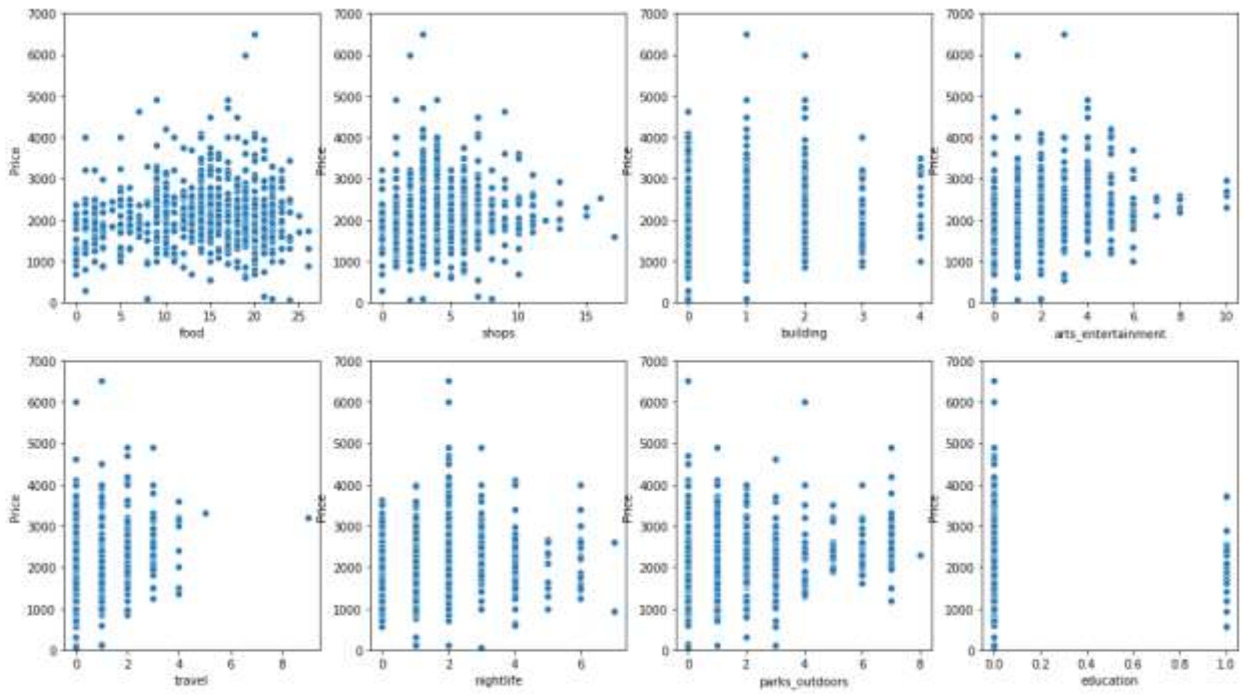

Figure 3 Relationship between 8 main categories and Price

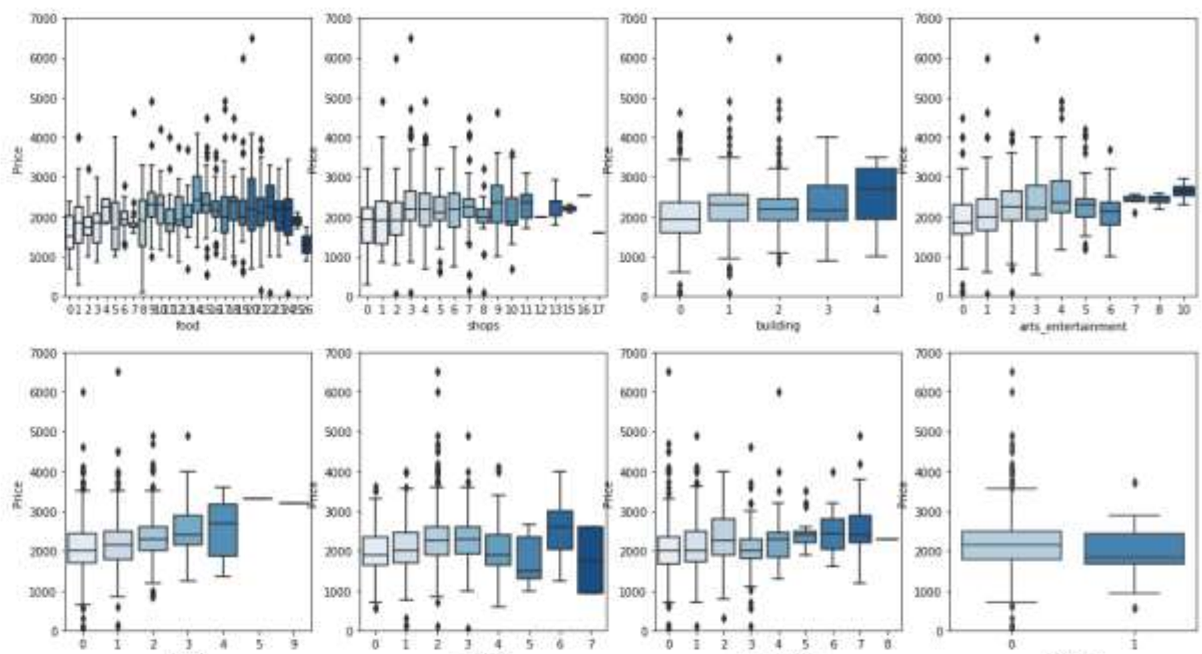

Figure 4.1 box plot Relationship between features (Latitude, and Longitude) and Price

3.4 Distribution of properties

As shown in (Figure 4) most of the properties are closer to the middle of Toronto with some on the edges of the city.

Figure 4 Distribution of properties

4. Predictive Modeling

To Predict the rent of a property a regression model is the best fit. In this case 3 regression models have been used Gradient Boosting Regressor, Random Forest Regressor, and KNN Regressor

4.1 Solution to the problems

The problem was to predict the rent for different properties. Therefore, the model must be able to predict a value that close to the real value hence using R^2.

4.2 Performances of different models

The models that have been used to predict the rent value had performed closely. The average in the accuracy between the models was 84% with KNN Regressor on top as shown in the table8

Model	Gradient Boosting Regressor	, Random Forest Regressor	KNN Regressor
Accuracy (R^2)	88.3%	77%	84.0%

Table 8 Models accuracy

In figure5 show the predicted value and real value for each property. Figure 5 confirms that the models are predicting good values not randomly selecting a value

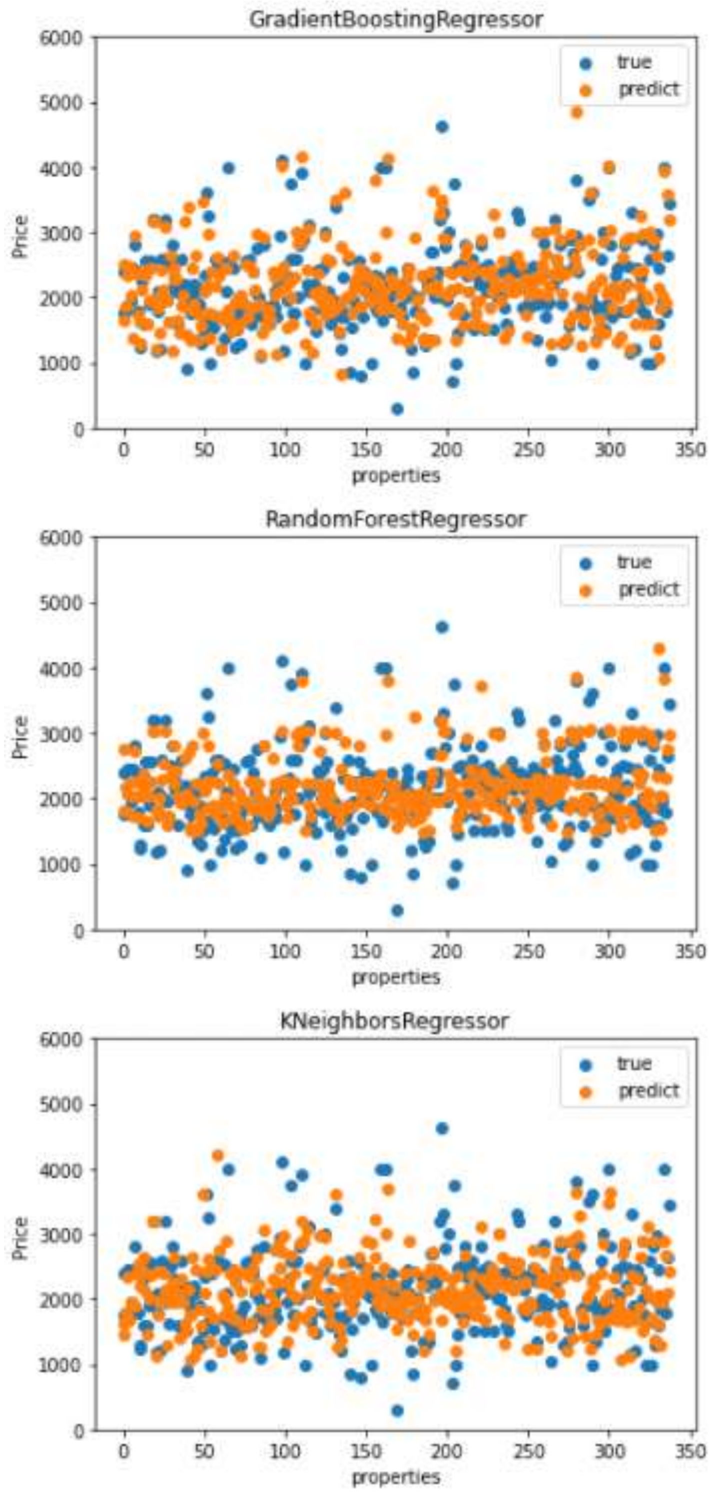

Figure 5 Prediction vs real price

5. Conclusions

In conclusion in this study, I attempted to predict the rent of some properties in Toronto. I have performed analyses on the features from Kaggel dataset as well as foursquare.com API. I used three different models Gradient Boosting Regressor, Random Forest Regressor, and KNN Regressor to predict the rent of the properties and found that KNN Regressor performed the best in term of accuracy (R^2).

6. Future directions

For the future of this project, I'm planning to build a simple website that will take the latitude, longitude, number of bathroom bedroom and if there is a den and output a predicted rent price.

In addition, this project must be updated regularly because rent prices are affected by economical changes.