

IBM Coursera Capstone Project

Rent Properties Analysis in Rotterdam, the Netherlands

Sivaporn Homvanish

21 January 2021

Introduction: Business Problem

Rotterdam is the 2nd largest city in the Netherlands. It is a major logistic and economic center and Europe's largest seaport. Moreover, it is always one of the main shipping industry centers which means that there are many companies located in this city¹. Consequently, employees and workers from different cities and countries move to Rotterdam. This creates high housing demand and supply also increases.

Rooms, houses, apartments in the Rotterdam are growing and very competitive. Hence, it is very difficult for new players who want to do their business in properties due to many offers are on the market.

In this study, we would like to analyze the rented properties in the Rotterdam. What are the main features that tenants are looking for? How to improve the properties and make it stands out from other competitors? Hence, the targeted stakeholders who will get benefit from this study will be housing companies, private rooms/apartment/house owners.

Data

According to the business problem, properties data in the Rotterdam and close-by venues around the properties area are required for the analysis.

First, the data from Kaggle, an online community of data scientists and machine learning practitioners², is used (<https://www.kaggle.com/juangesino/netherlands-rent-properties>). This data has real-estate rental market information in the Netherlands such as room size, rental fee, room capacity, etc. However, we will mainly focus on Rotterdam city. Other real estates in other cities in the Netherlands will be excluded from this study.

Moreover, we used Foursquare API to gather near-by places around the properties e.g., restaurants, tram/train stations, bars, supermarket, etc. This information will be used together with properties data to see if the places around the properties have impact on the tenet decision.

Data Exploration and Cleaning

The output below shows the raw data retrieved from Kaggle website. The properties data consists of 62 features and 46,722 properties in all cities in the Netherlands. This raw data has many duplicated features and this study mainly focuses on the rented properties in Rotterdam city because we would like to find

¹ <https://en.wikipedia.org/wiki/Rotterdam>

² <https://en.wikipedia.org/wiki/Kaggle>

main features of the room that attract potential tenants. Hence, the data is needed to be cleaned and filtered.

	_id	additionalCosts	additionalCostsDescription	additionalCostsRaw	areaRaw	areaSqm	city	coverImageUrl
0	{'Soid': '5d2b113a43cbfd7c77a998f4'}	50.0	NaN	\n â¬ 50\n ...	14 m2	14	Rotterdam	https://resources.kamernet.nl/image/913b4b03-5...
1	{'Soid': '5d2b113a43cbfd7c77a9991a'}	0.0	NaN	\n â¬ 0\n ...	30 m2	30	Amsterdam	https://resources.kamernet.nl/image/5e11d6b5-8...
2	{'Soid': '5d2b113a43cbfd7c77a99931'}	NaN	NaN	\n -\n ...	11 m2	11	Amsterdam	https://resources.kamernet.nl/image/74b93a27-a...
3	{'Soid': '5d2b113a43cbfd7c77a9994a'}	NaN	NaN	-	16 m2	16	Assen	https://resources.kamernet.nl/image/84e95365-6...
4	{'Soid': '5d2b113b43cbfd7c77a9997c'}	NaN	NaN	\n -\n ...	22 m2	22	Rotterdam	https://resources.kamernet.nl/Content/images/p...
5	{'Soid': '5d2b113b43cbfd7c77a99993'}	18.0	NaN	\n â¬ 18\n ...	14 m2	14	Groningen	https://resources.kamernet.nl/image/5756f247-7...
6	{'Soid': '5d2b113b43cbfd7c77a999ac'}	NaN	NaN	-	22 m2	22	Zeist	https://resources.kamernet.nl/image/1ca3d9b5-1...

Moreover, some features are categorization value and some machine learning algorithms are suitable with numerical data only. Therefore, we converted them to numerical value as follows.

- Column 'furnish' ==> Unfurnished(0), Furnished(1), Uncarpeted(2)
- Column 'internet' ==> No(0), Yes(1), Unknown(2)

Below is the final properties data after finishing data transformation.

	id	areaSqm	furnish	internet	matchCapacity	rent	latitude	longitude	postalCode
0	5d2b113b43cbfd7c77a999c4	17	1	1	1	857	51.921123	4.505328	3063ER
1	5d2b113c43cbfd7c77a99a0d	25	1	1	4	600	51.893195	4.516478	3074WL
2	5d2b115d43cbfd7c77a9a588	29	1	2	1	530	51.905459	4.453957	3024TX
3	5d2b11b343cbfd7c77a9c6e4	22	0	1	1	350	51.890358	4.475083	3083GA
4	5d2b11b443cbfd7c77a9c72e	13	1	1	1	699	51.924588	4.486010	3011KM

We also consider nearby venues around properties area. Foursquare API comes to play with the condition that the distance between venues and properties must not be more than 500 meters and limit for 30 venues. The result shows below.

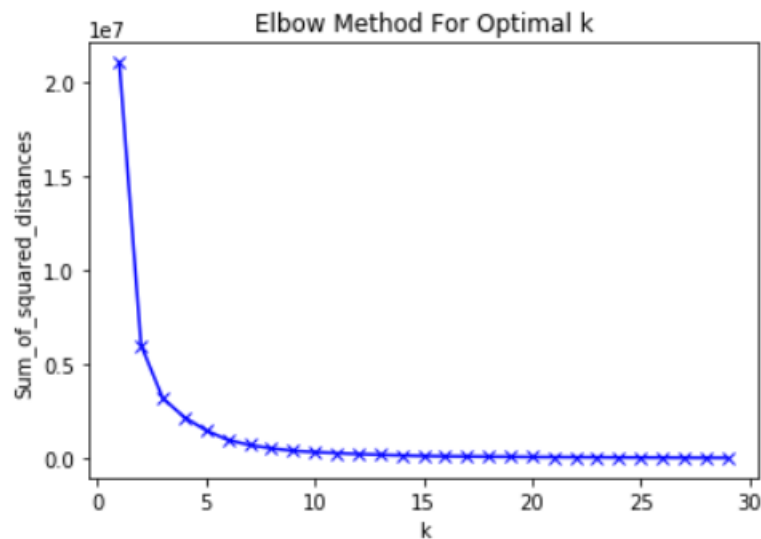
	id	id Latitude	id Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	5d2b113b43cbfd7c77a999c4	51.921123	4.505328	The Student Hotel Rotterdam	51.921319	4.505519	Hotel
1	5d2b113b43cbfd7c77a999c4	51.921123	4.505328	Aloha	51.919768	4.500730	Restaurant
2	5d2b113b43cbfd7c77a999c4	51.921123	4.505328	De Hemel op Aarde	51.922789	4.500983	Tapas Restaurant
3	5d2b113b43cbfd7c77a999c4	51.921123	4.505328	Bodymentors	51.921896	4.501067	Gym
4	5d2b113b43cbfd7c77a999c4	51.921123	4.505328	Tropicana	51.919768	4.501717	Pool

	id	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	5d2b113b43cbfd7c77a999c4	Ice Cream Shop	Bar	Deli / Bodega	Drugstore	Pool	Fried Chicken Joint	Gym	Supermarket	Greek Restaurant	Burger Joint
1	5d2b113c43cbfd7c77a99a0d	Tram Station	Soccer Stadium	Sandwich Place	Museum	Middle Eastern Restaurant	Market	Fast Food Restaurant	Train Station	Furniture / Home Store	Farmers Market
2	5d2b115d43cbfd7c77a9a588	Theater	Restaurant	Hotel	Café	Music Venue	Supermarket	Outdoor Sculpture	Coffee Shop	Bistro	Boat or Ferry
3	5d2b11b343cbfd7c77a9c6e4	Supermarket	Tram Station	Pool Hall	Flea Market	Burger Joint	Cosmetics Shop	Resort	Friterie	Fried Chicken Joint	French Restaurant
4	5d2b11b443cbfd7c77a9c72e	Italian Restaurant	Organic Grocery	Asian Restaurant	Bar	Hawaiian Restaurant	Bakery	Seafood Restaurant	Smoke Shop	Sandwich Place	Food & Drink Shop

Methodology

To find similar properties in Rotterdam, we used K-means clustering to determine and segment them into the same cluster. We used the properties data together with common venues nearby from Foursquare API as an input of the clustering.

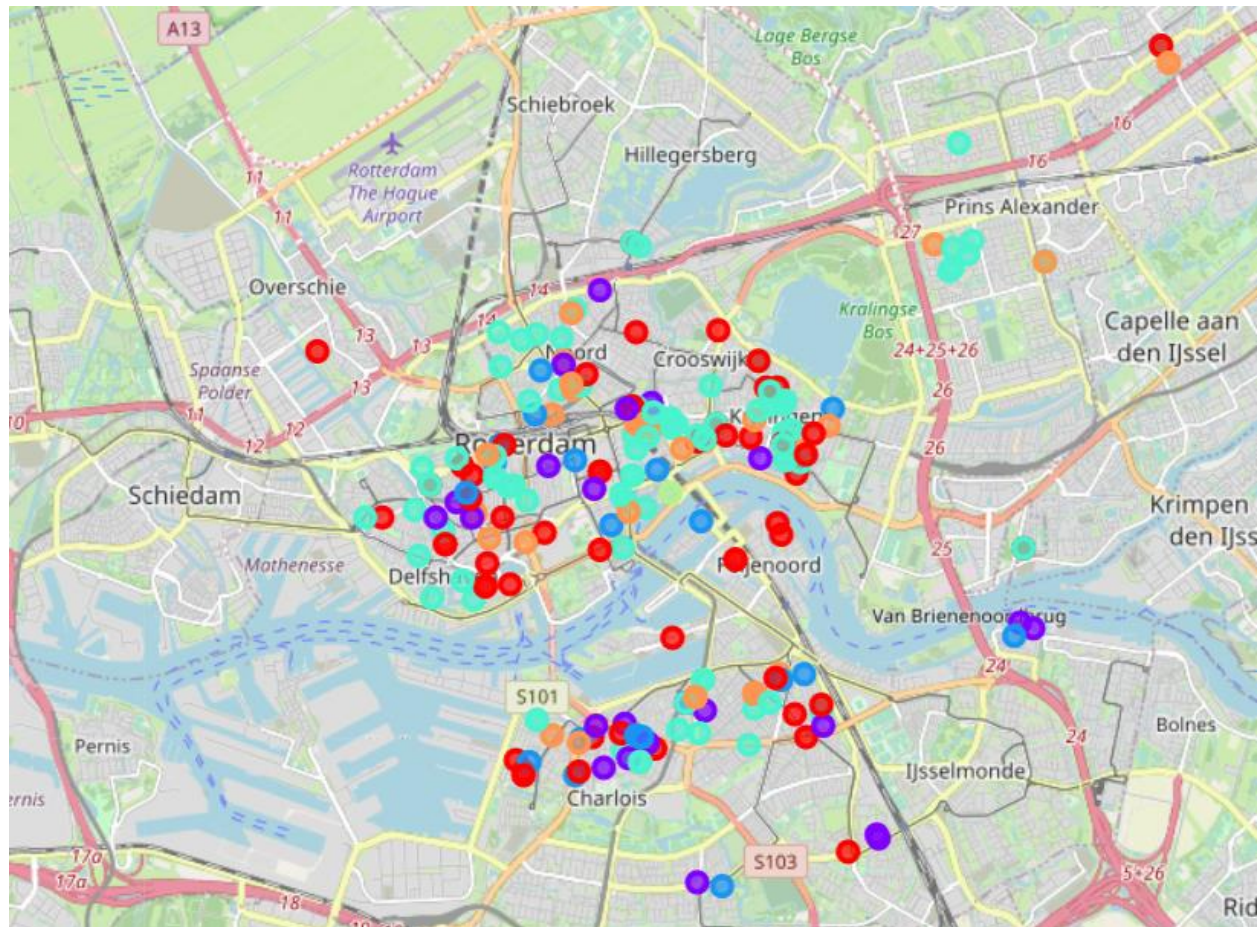
However, we first need to define the number of clusters (K) by using elbow curve to identify the most suitable number for this study.



From the elbow curve, we could clearly see that when the number of clusters is above 6, the squared distance values barely change. Hence, we select $k = 6$ as an appropriate number for k-means clustering.

Then, we apply K-mean clustering with the data to cluster properties having similar features. The result is visualized on map by using latitudes and longitudes. Different clusters are represented by different colors of dot.

	id	areaSqm	furnish	internet	matchCapacity	rent	latitude	longitude	postalCode	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	5d2b113b43cbfd7c77a999c4	17	1	1	1	857	51.921123	4.505328	3063ER	1	Ice Cream Shop	Bar	Deli / Bodega
1	5d2b113c43cbfd7c77a99a0d	25	1	1	4	600	51.893195	4.516478	3074WL	3	Tram Station	Soccer Stadium	Sandwich Place
2	5d2b115d43cbfd7c77a9a588	29	1	2	1	530	51.905459	4.453957	3024TX	3	Theater	Restaurant	Hotel
3	5d2b11b343cbfd7c77a9c6e4	22	0	1	1	350	51.890358	4.475083	3083GA	0	Supermarket	Tram Station	Pool Hall
4	5d2b11b443cbfd7c77a9c72e	13	1	1	1	699	51.924588	4.486010	3011KM	5	Italian Restaurant	Organic Grocery	Asian Restaurant



Results and Discussion

Before analyzing the result, we ranked the clusters by the number properties. It could be seen that most of the properties fall into cluster 3, 0, 5, 1, 2 and 4, respectively. Then, we examine all clusters in detail to see the common features within a cluster and the difference across clusters.

Based on the result of the model, the properties in Rotterdam can be classified into 6 groups ordered by number of properties which are.

1. **Cluster 3** (90 properties):

Furnished room for 1 person with internet, low rental price (450-600 euros/month) and have restaurants, drug store, supermarkets, and coffee shop nearby.

2. **Cluster 0** (56 properties):

Furnished/unfurnished room for 1 person with very low rental price (250-450 euros/month) and have restaurants, drug store, supermarkets, and coffee shop nearby.

3. **Cluster 5** (36 properties):

Furnished room for 1-2 persons with internet, medium rental price (650-800 euros/month) and have restaurants, supermarkets, and coffee shops nearby.

4. **Cluster 1** (25 properties):

Unfurnished room for 1-2 persons (>50 m²) with medium to high rental price and have restaurants, supermarkets, drug store and coffee shops nearby.

5. **Cluster 2** (21 properties):

Furnished/unfurnished room for 4 persons (>45 m²) with high rental price (>1,100 euros/month) and have restaurants and public transportation nearby.

6. **Cluster 4** (2 properties):

Very high rental price (1,950 euros/month) and big size (>120m²)

Conclusion

All 6 clusters were summarized and analyzed based on the prominent similar information within the same cluster.

Most of rented properties are for 1 person with low rental price (450-600 euros/month) following by very low rental price (250-450 euros/month). For 1-2 persons rooms, the price is in medium to high range which is mixed between furnished and unfurnished. Lastly, the least popular properties have very high rental price and large living area.

However, we could see that the features having big influence are the rental price, room capacity and size, respectively. Meanwhile, the venues nearby are similar across clusters. Most commons are restaurants, supermarket, drug store.

As a result, housing companies, private rooms/apartment/house owners can use this information for their properties improvement, price adjustment, advertising, and suitable locations with nice venues.