Compare two cities for relocation and map neighbourhood for rental accommodation.

Lawrence D'silva

Aug 1st 2020

1. Introduction

Background

When relocating a user tries to find out certain parameter to know more about the city eg; Cost of living, healthcare facility, pollution and crime rate etc; The user also would look for rental accommodation close to his work area or if the user is looking for a job close to neighbourhood where there are job option.

In this scenario two cities will be compared based on parameter like cost of living(Market, Restaurant, Transport, Child Care), Pollution and Crime rate of Mumbai and Bangalore two metro cities of India.

For rental accommodation will be considering neighbourhood close to major job location Electronic City in Bangalore and Bandra Kurla complex in Mumbai, both these places are major source for jobs having close to 200+ companies.

For these neighbourhoods top 5 venue details will be displayed based on their clusters.

Problems

Mumbai and Bangalore are huge cities having their own pros and cons, The city is also split into various zones, to pin points which zone would be suitable add another complexity, but both the cities are tremendous job opportunity and lot a various accommodation option. Having a consolidated view of the two cities and well as providing rental details based on nearby area where there is huge job opportunity will help the user with their decision making and provide support.

Getting the data is another challenge as it is not readily available and there are no official site, had to browse through numerous sites including official government site to scrub the data.

2. Data Source and Cleansing

For Market, Restaurants, Health care, Transport, Child care, Pollution and Crime rates, retrieved the data from website numbeo.
In total 3 features and 50 rows
For properties in Bangalore and Mumbai scrubbed the data from Magicbricks.
In total 18 rows and 6 features.

The data were not in a downloadable format. Had to scrap the data using Beautiful soup module in python into a csv format.

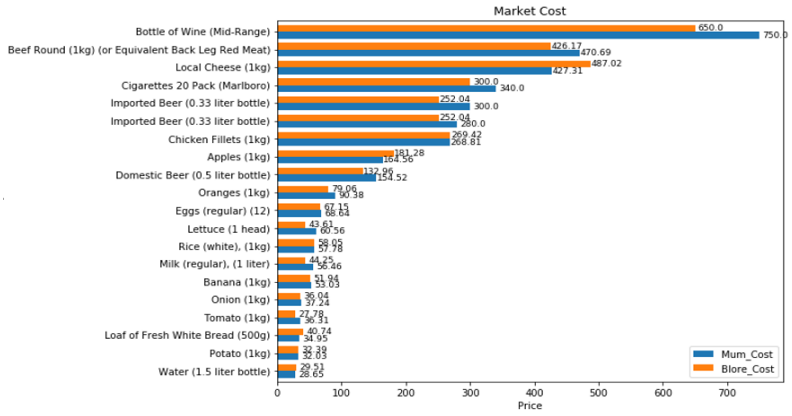
The data for cost of living were categorized based on Market, Restaurants, Transport, Utilities, Child Care. They had 50 rows and 3 features, for better visualization split this into individual data frames and multiple plots for each category.
Utilities category was dropped as there were only 3 rows and it did not provide much information to the user.

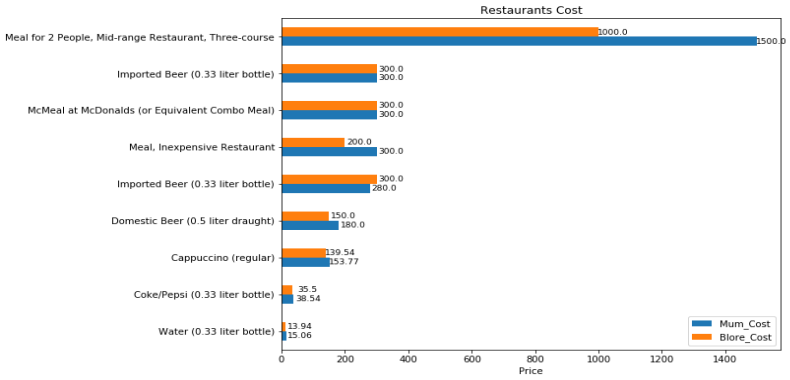
The data for neighborhood with various accommodation options were available without co-ordinates, used folium to get their correct co-ordinates and foursquare to map it.
Retrieved a various venue in those neighborhoods using foursquare and plotted them based on clusters.

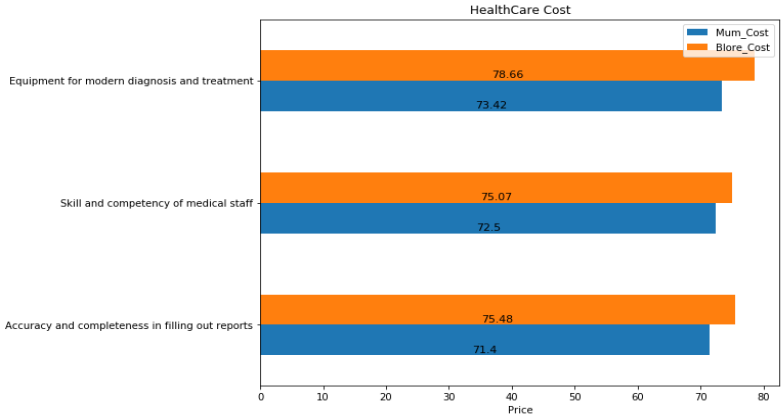
3. Exploratory Data Analysis

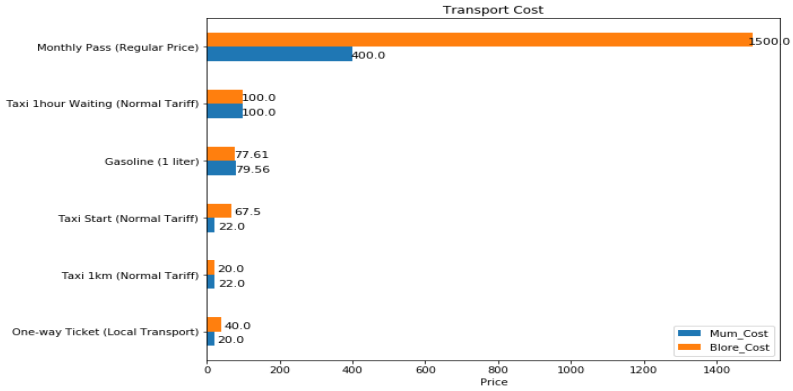
To get a better visualization of data and a single view to compare two cities, used horizontal bar plot. Split each category into separate dataframe.

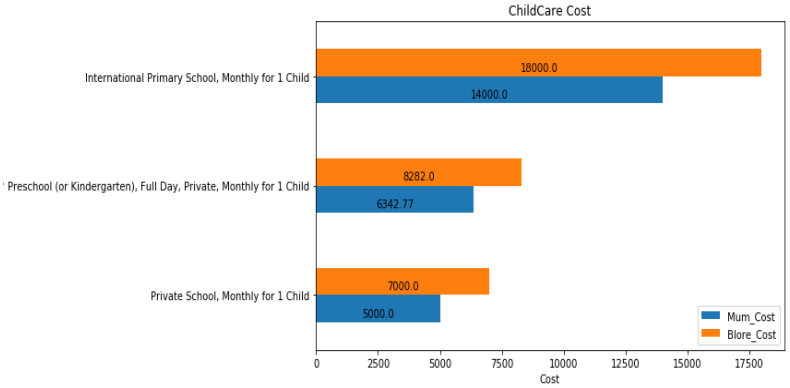
Cost of Living: Bangalore has an edge over Mumbai by a 1% based on cost differentiation for each item combining all the categories.



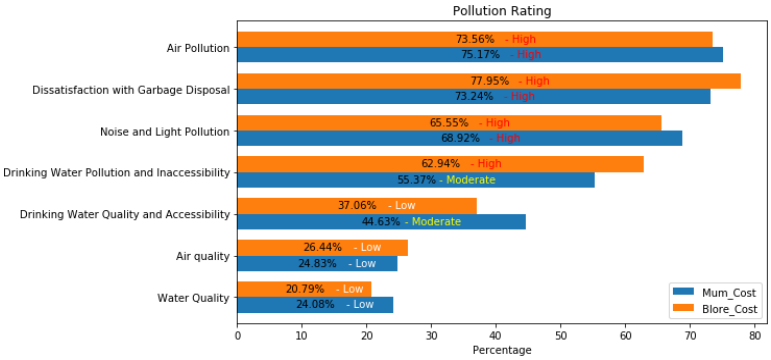


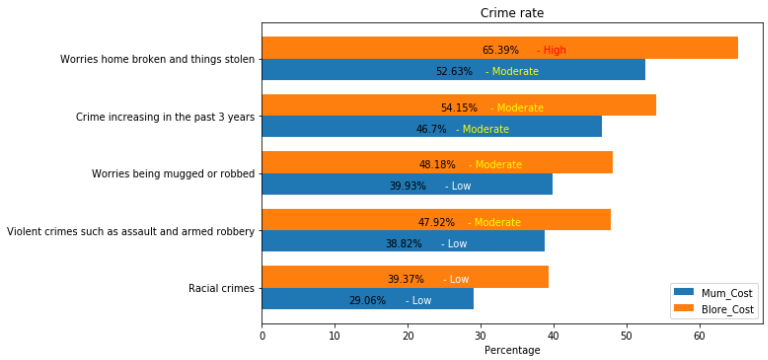






Pollution and Crime rate: On Crime and Pollution rate both the cities are on similar threshold with crime been comparatively low for Mumbai and pollution high in both the cities.

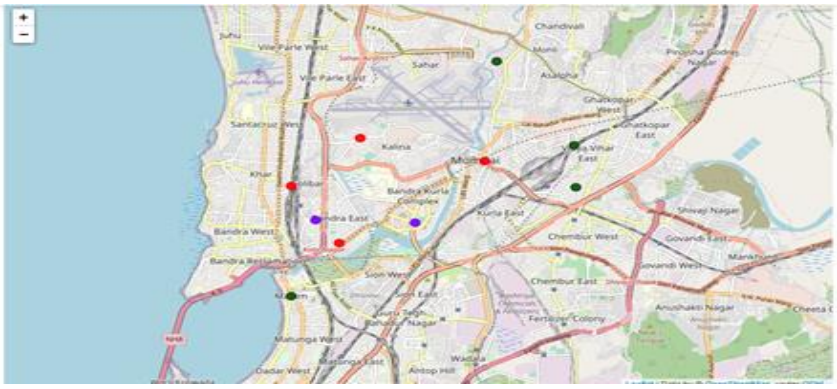




Rental accommodation for Mumbai:

The rental accommodation is near Bandra Kurla Complex within 6KM radius.

- Cluster 0 color red the range of property is 50667.0 - 65667.0
- Cluster 1 color purple the range of property is 85000.0 - 90000.0
- Cluster 2 color green the range of property is 32333.0 - 41000.0



Top 5 Venue details based on Neighbourhood along with their cluster info

	Neighborhood	Cluster	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Kalanagar	0	Restaurant	Indian Restaurant	Brazilian Restaurant	Sports Club	South American Restaurant
2	Khar east	0	Indian Restaurant	Bar	Lounge	Pub	Nightclub
4	King Circle	0	Mughlai Restaurant	Middle Eastern Restaurant	Flea Market	River	Indian Restaurant
8	Vakola	0	Chinese Restaurant	Seafood Restaurant	Bakery	Indian Restaurant	Thai Restaurant
0	BKC	1	Bar	Coffee Shop	Indian Restaurant	Chinese Restaurant	Thai Restaurant
3	Kherwadi	1	Indian Restaurant	Café	Diner	Pizza Place	Thai Restaurant
5	Mahim East	2	Indian Restaurant	Dessert Shop	Arcade	Asian Restaurant	Café
6	Sakinaka	2	Hotel	Hookah Bar	Coffee Shop	Thai Restaurant	Grocery Store
7	Tilak Nagar	2	Indian Restaurant	Fast Food Restaurant	Playground	Italian Restaurant	Chinese Restaurant
9	Vidyavihar	2	Fast Food Restaurant	Indian Restaurant	Bar	Cricket Ground	Juice Bar

Rental accommodation for Bangalore:

The rental accommodation is near Electronic city within 8KM radius.

- Cluster 0 color red the range of property is 19000.0 - 20000.0
- Cluster 1 color purple the range of property is 23667.0 - 23667.0
- Cluster 2 color red the range of property is 17333.0 - 18333.0

Top 5 Venue details based on Neighbourhood along with their cluster info

	Neighborhood	Cluster	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Begur Road	0	Tea Room	Asian Restaurant	Auto Garage	Mobility Store	Hotel
4	Maragondanahalli	0	Fruit & Vegetable Store	Restaurant	Women's Store	Comfort Food Restaurant	Food Court
5	Neeladri Nagar	1	Department Store	Indian Restaurant	Snack Place	Kerala Restaurant	Multicuisine Indian Restaurant
0	Ananth Nagar	2	Women's Store	Department Store	ATM	Sports Bar	Gaming Cafe
2	Doddathoguru	2	Asian Restaurant	Gym	Indian Restaurant	Food	Cosmetics Shop
3	Hosur road	2	Indian Restaurant	Chinese Restaurant	Fast Food Restaurant	Café	Lounge

4. Discussion

Mumbai and Bangalore are big cities with a high population density. Considered the closet neighbourhoods for rental accommodation based on accessibility to the prominent job locations. Rental are based on type of accommodation e.g.;
3BHK, 2BHK and 1BHK (BHK stands for Bedroom, living room and kitchen). Considered the average of accommodation for clustering.

I used the Kmeans algorithm as part of this clustering study. I set the optimum k value to 3. However, less than 10 neighbourhood coordinates were used. I performed data analysis through this information by adding the coordinates of neighbourhoods and home sales price averages as static data along with top 5 venue details for each neighbourhood.

I ended the study by visualizing the data and clustering information on the Bangalore and Mumbai map. In future studies
we can expand the details of the neighbourhoods for e.g.; closeness to education institute, open parks, accessibility to transport can be explored.

5. Conclusion

Purpose of this project was to identify which city would be good for relocation based on parameters like Cost of living (Market, Restaurants, Child care, Transport. Healthcare), Pollution and Crime. Compare each parameter for the location using horizontal bar.

This also consider rental accommodation and top 5 venue options within those neighborhoods close to Job cities. The rental rates are clustered based on average rental cost.

These clusters will act as a guide to the user for rental accommodation based on their budget.

There is room for improvement for instance parameter like accessibility to education institution, nearest transport systems, open parks can be considered when selecting the neighborhoods.

The items to compare within the parameters can be more refined.