Comparison of Cities - Karachi vs Islamabad

Mirza Mawaz Khan

28th January, 2020

1. Introduction:

1.1. Background:

During this capstone project we would be looking at Real Estate industry's data of two major cities in Pakistan, Karachi and Islamabad, taken from Zameen.com. Karachi is the industrial hub of Pakistan with over 14.9 Million people of various groups, affiliations, backgrounds and ethnicities residing there. On the other hand Islamabad is the capital of Pakistan which has seen growth in Real Estate industry in past 1-2 years as compared to other cities.

Zameen.com is one of the largest online real estate portals in Pakistan. Founded in 2006, it has become one of the top five online portals in real estate industry in the world. Currently the influx of new listings on their portal, whether it be for property sales or property rental or project sales, has surpassed 350,000 listings per month.

However, with such large amount of data, Pakistan's real estate market has yet to fully adapt to digital transformation. There are various factors, both controllable and uncontrollable on part of the industry, still Zameen has provided a platform for data enthusiasts to generate insights and to provide guidelines based on those insights to all the stakeholders involved.

1.2 The Case:

Having large data influx without proper evaluation and insights is of no use. People in Pakistan make real estate investment decisions based on personal preferences, Geo-Political situation or previous history. However none of these factors take in to account the accessibility, growth and diverse nature of each geographical location, their cultural preferences and surroundings.

This study takes in to account the locality data for all the locations/neighborhoods under observation in dataset and gives a better picture of which type of property is best suited in which locality, taking in to both similarities and differences between each location.

1.3 Interest:

Being part of Pakistan's real estate industry, it seems fitting for a data enthusiast to contribute towards this industry, which is ripe with opportunities. On the other hand, with years of experience in sales, this data, shared openly can help make the conversations between real estate sales consultants and their potential buyers smoother and help the up and coming real estate salesperson gain better insights to progress their careers.

1.4 Stakeholders:

This report will help potential buyers and investors in making educated decisions regarding their purchases or investment. Stakeholders for this report will be buyers, investors and Real Estate marketing portals.

2. Data & Source:

2.1. Source:

The dataset is taken from an online Real Estate portal "Zameen.com", shared publicly on Kaggle, which lists various real estate listing in various cities, including Karachi and Islamabad.

2.2. Dataset Features:

The dataset has following features:

Features	Description	Data Type	Relevance
Property_id	Unique ID assigned to each listing on zameen.com	String	No
Location_id	Unique ID based on each location for each listing	String	No
Page_url	URL from where the listing is taken from	String	No
Property_type	The type of property listed	String	Yes
Price	Listed pricing of property	Float	Yes
Location	Neighborhood/Community of each listing in a city	String	Yes
City	Name of the city the listing is in	String	Yes
Province_name	Name of the province	String	No
Latitude	Coordinates of listing	Float	Yes
Longitude	Coordinates of listing	Float	Yes
Area	Size of the property listed	String	Yes
Purpose	Property either listed as “For Sale” or “For Rent”	String	Yes

However we would only be using the features which are relevant to our study, as categorized in above table. Raw data consisted of 179838 rows and 310 columns. Following is the header (first 5 rows) of the extracted dataset, based on relevance:

	property_type	price	location	city	latitude	longitude	area	purpose
0	Flat	10000000.0	G-10	Islamabad	33.679890	73.012640	4 Marla	For Sale
1	Flat	6900000.0	E-11	Islamabad	33.700993	72.971492	5.6 Marla	For Sale
2	House	16500000.0	G-15	Islamabad	33.631486	72.926559	8 Marla	For Sale
3	House	43500000.0	Bani Gala	Islamabad	33.707573	73.151199	2 Kanal	For Sale
4	House	7000000.0	DHA Defence	Islamabad	33.492591	73.301339	8 Marla	For Sale

2.3 Data Cleaning:

Although we have dropped irrelevant features/columns from our dataset, however we still have 179838 rows to analyze for inconsistencies before we can do our exploratory data analysis and apply clustering algorithms. There may be many irregularities when confronting raw data of any dataset. In our particular case, we encountered the following:

- NaN Values:
  - Out of the total of 179,838 entries in our dataset, 12,512 were dropped as they contained NaN values.
- Incorrect Data Type (dtype) in Data Frame:
  - From our selected features, we observe that feature “area” has type string and contains both the value and dimensions of the size of a listing. On top of that, the dimensions used in this datasets are two

different dimensions, which makes the comparison even more inconsistent. In order to rectify this, we corrected the data type and added a new feature “Size_in_Sq_Yards” which provides consistent dimension in Sq. Yards for each listing’s size.

- Inconsistent or Incorrect entries in Size and Price:
  - There could be a various reasons for wrong data entry when creating a listing, due to which a listing may have either an inconsistent price tag of just 0, 1 or 2 Rupees or incorrect size of listing. For example a Penthouse listed at 1 Sq. Yards size. All such entries were dropped.
- Irrelevant Information:
  - As our study focuses primarily on the sale side of real estate industry, taking in to account “For Rent” properties would not give correct insights, therefore we only consider “For Sale” entries.

After cleaning the data, normalization of Price variable and sorting, we have 62,337 entries and 9 features in our dataframe. Following is the header of dataframe:

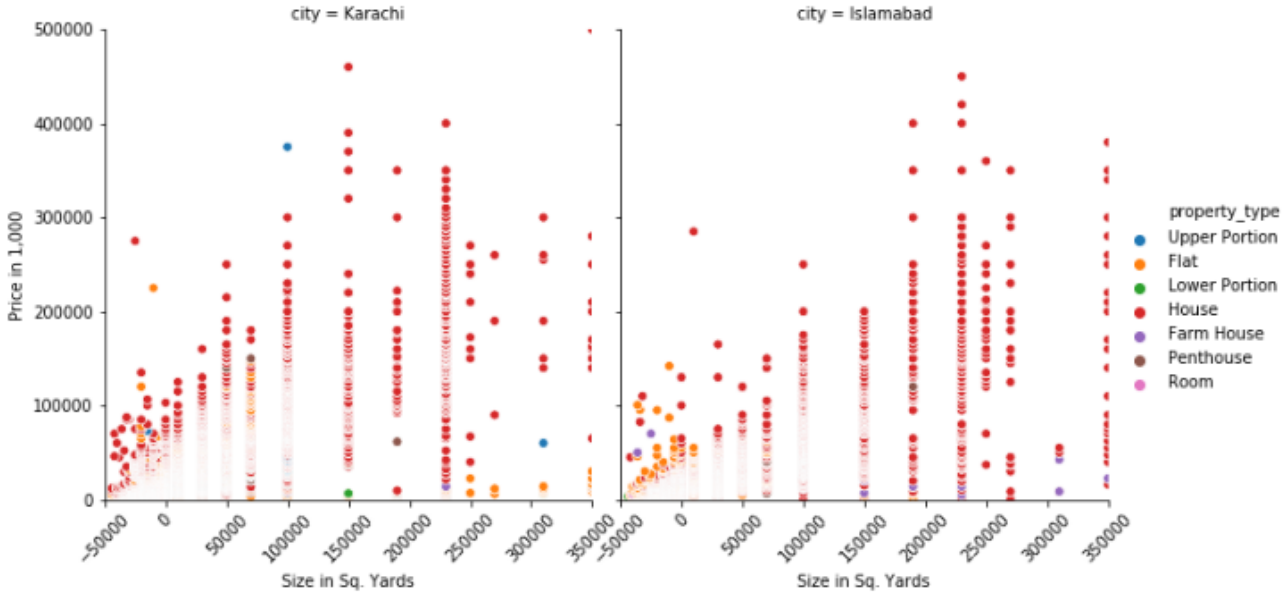
	property_type	price	location	city	latitude	longitude	purpose	Size_in_Sq_Yards	norm_p
0	Upper Portion	110000.0	DHA Defence	Karachi	24.794215	67.064610	For Sale	750.0	110.0
1	Upper Portion	130000.0	DHA Defence	Karachi	24.821639	67.071691	For Sale	750.0	130.0
2	Flat	135000.0	Clifton	Karachi	24.813927	67.012610	For Sale	240.0	135.0
3	Lower Portion	150000.0	DHA Defence	Karachi	24.810265	67.043552	For Sale	750.0	150.0
4	House	150000.0	G-13	Islamabad	33.650065	72.963681	For Sale	400.0	150.0

3. Exploratory Data Analysis:

3.1. Size vs Price:

For most of the buyers, the two key features in making a buying decision are size of property and its price. From a buyer’s perspective they both maintain a linear relationship. However in our dataset, we have other features as well which affect to a degree on the whole decision making process and for that we first analyze size and price to see how they fit when visualized.

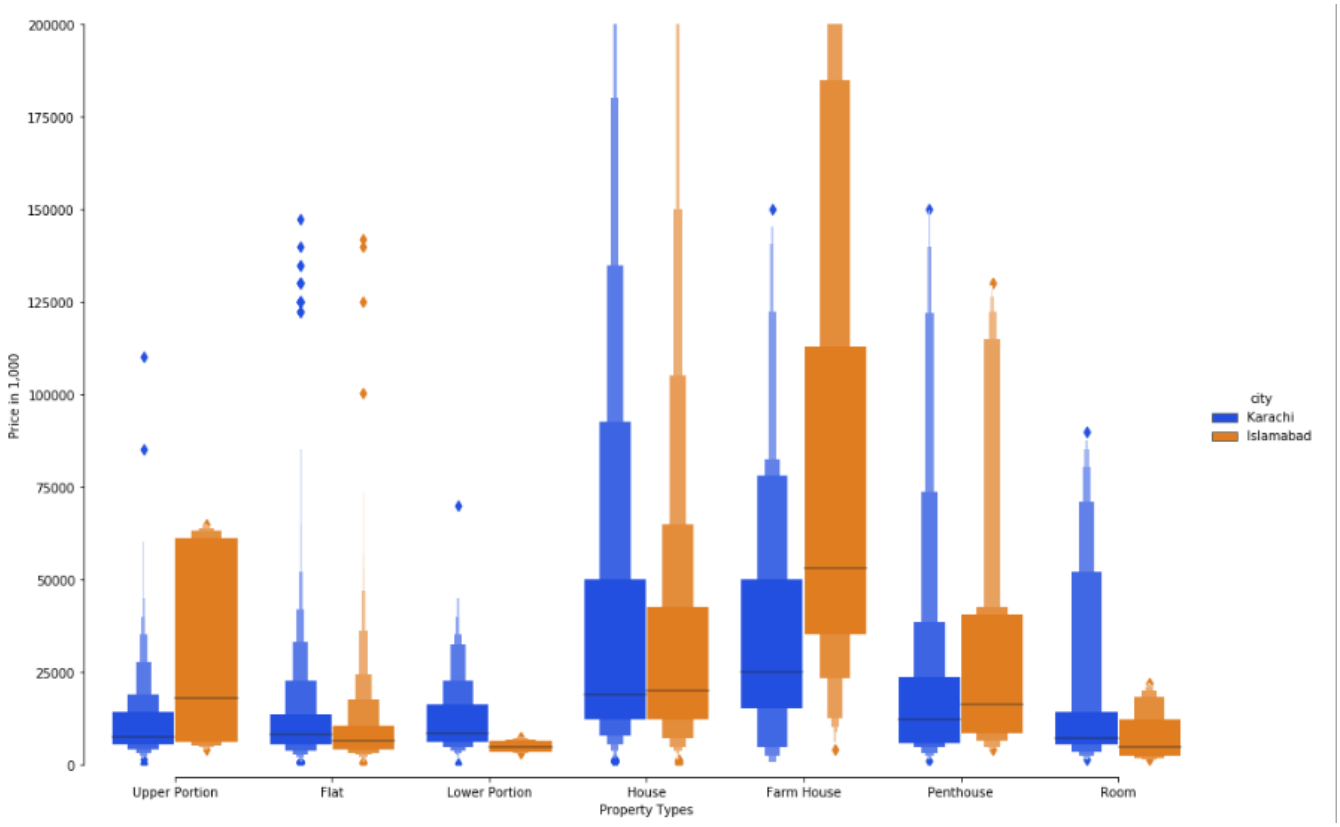
Following is the relational scatterplot of features Size and Price with Size being on the x-axis and Price on the y-axis. FacetGrid divides the two subplots based on City feature, whereas Markers are based on Property Type feature of dataset:



The scatterplot of Size and Price variables above do not show a clear picture of the landscape. We therefore opt for other visualization methods such as boxplots to segment our data and later, further classify based on other features separately such as Property Type, City and Type of Listing.

3.2. Property Types vs City vs Price:

Another visualization method that we opt is a box plot of features. In our boxplot feature Property Type is on x-axis, Price is on y-axis and we compare data from both Karachi and Islamabad. Karachi and Islamabad are plotted side by side for comparative.

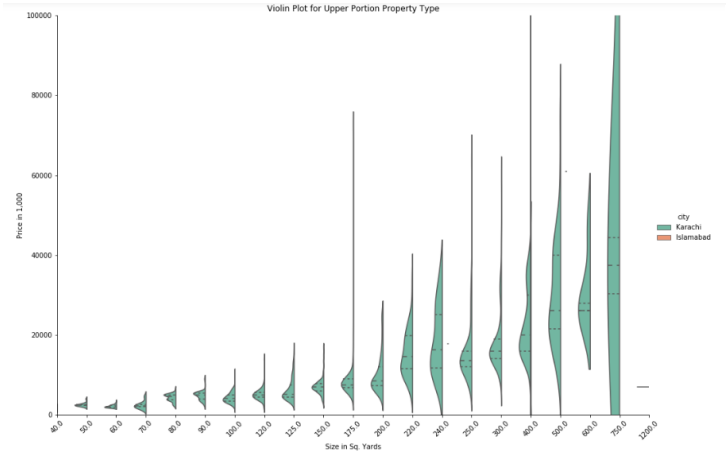


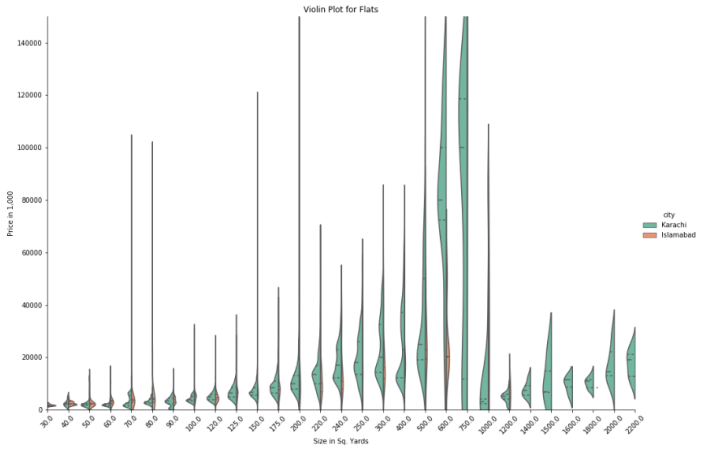
Above Boxplot shows a much refined picture of property prices in Karachi and Islamabad. This plot shows a comparison on bases of type of property listed in the two cities and is available for sale. From above plot we can observe the following:

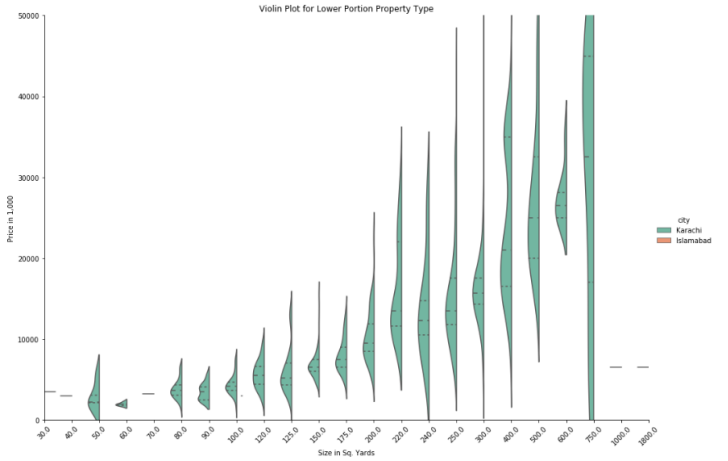
- Islamabad has higher property listing for "Farm House", "Penthouse" and "Upper Portion" categories
- Karachi has higher property listing for "House", "Lower Portion" and "Room Category"
- Plot does not differentiates between property size
- Plot does not show number of listings in each property type

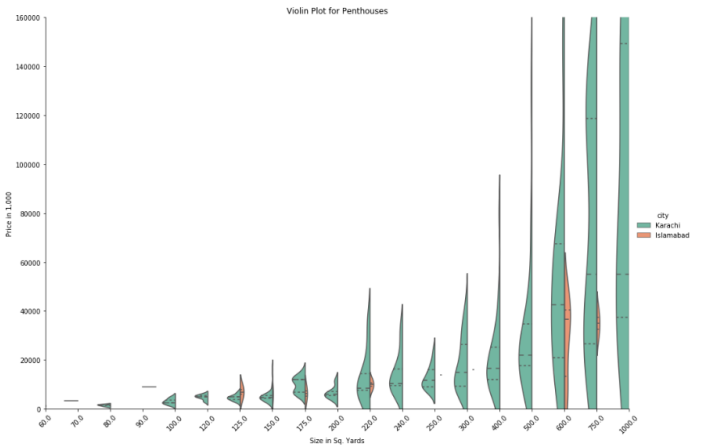
In order to answer the questions looming from above observations, we plot categorical pricing of each listing separately for all the property types.

3.3. Individual Visualization: Property Type vs Price

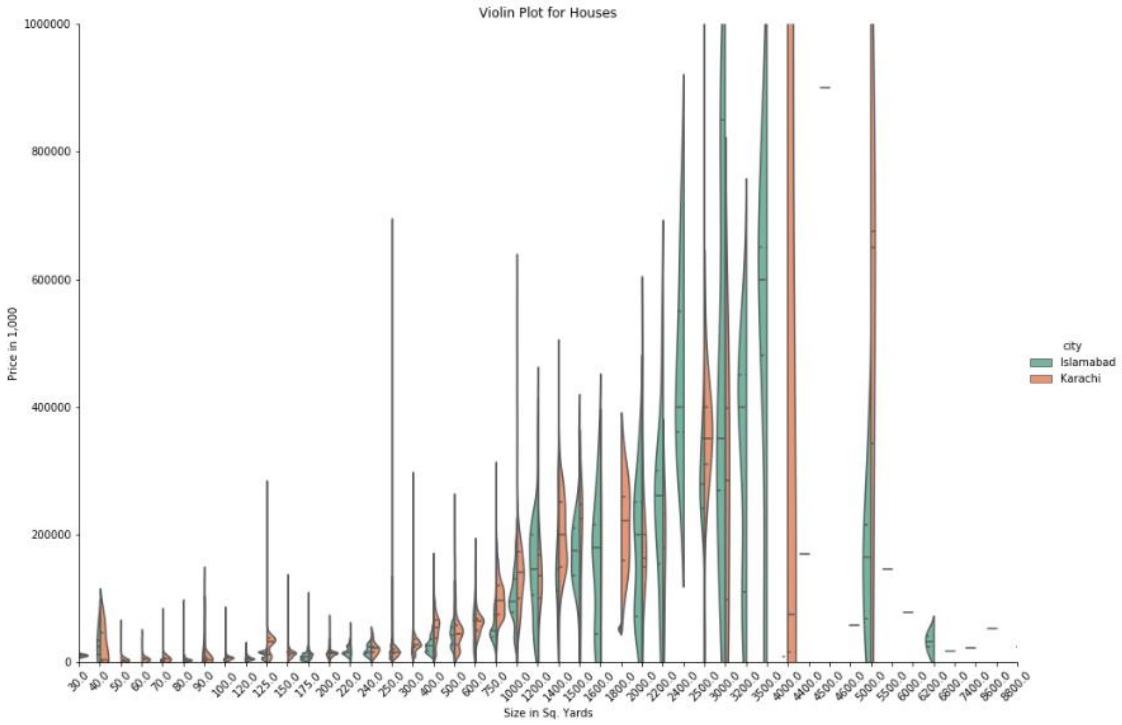




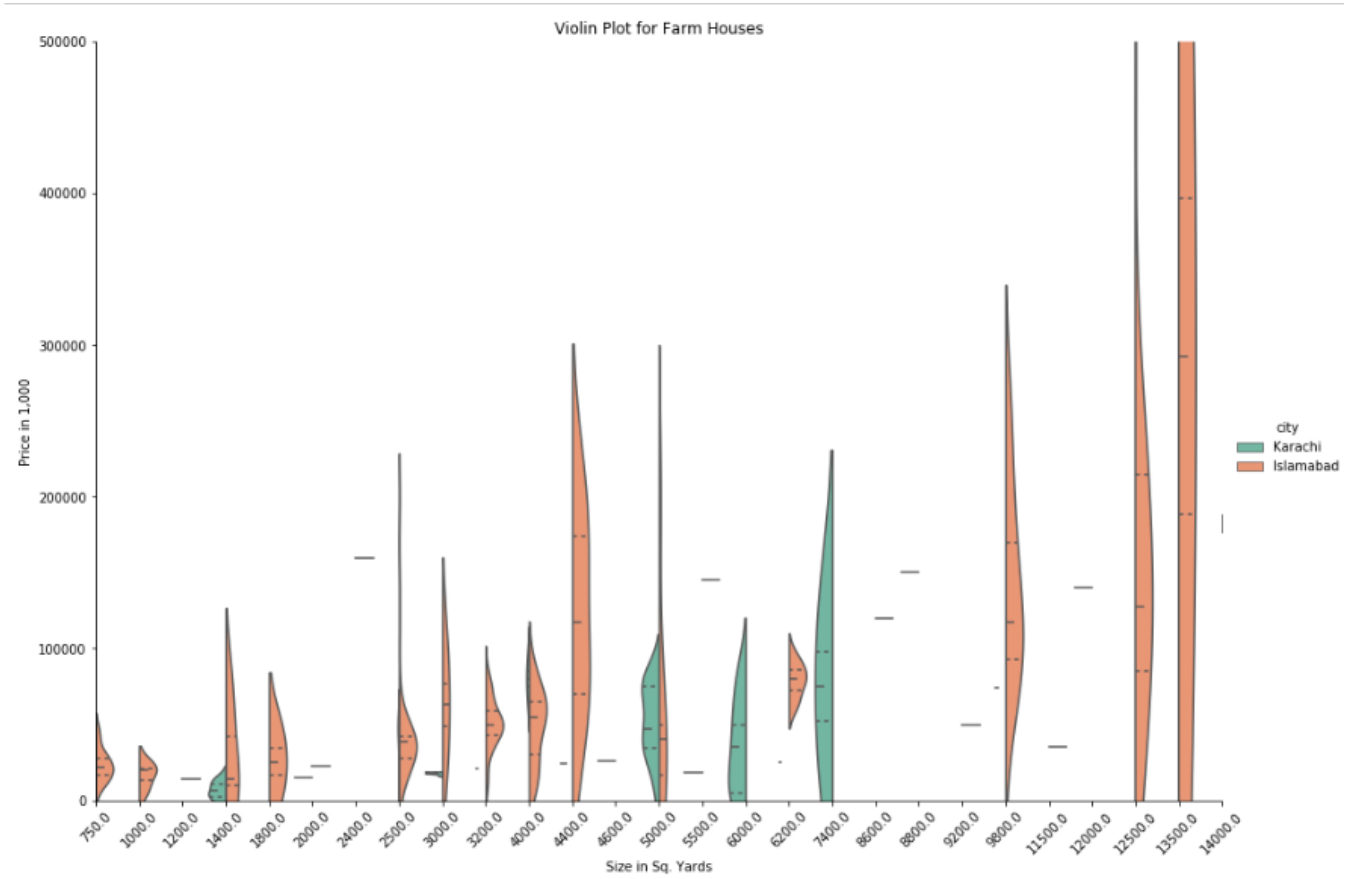




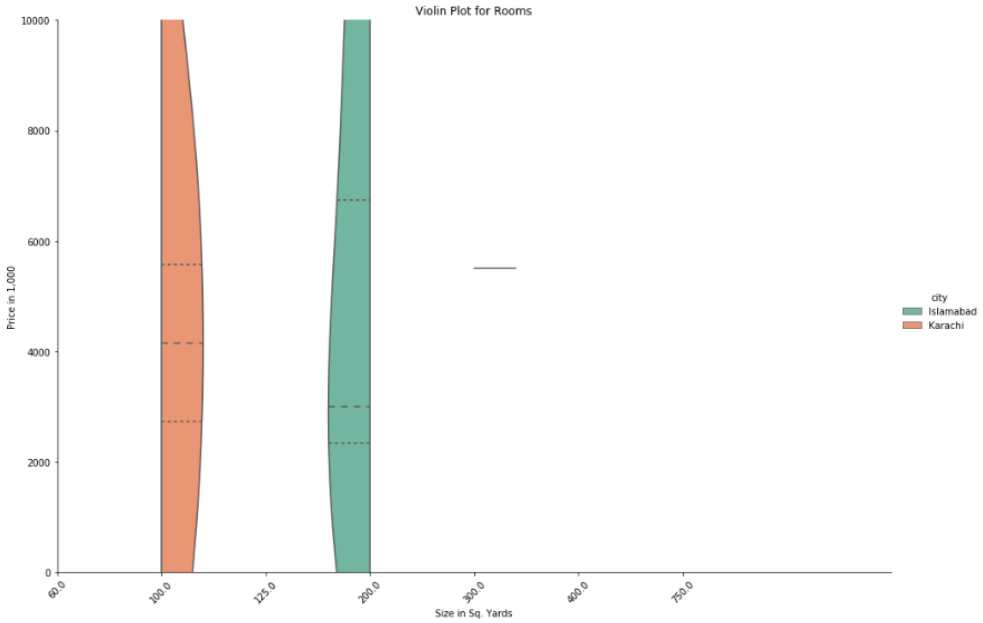
For “Upper Portions”, “Lower Portions”, “Flats” and “Penthouses” Karachi city (in green) has a diverse range of options. Although some property listing are inconsistent with observed linear relationship between price and size, specifically for “Flats” in plot above, however these constitute to a small sample.



The above plot for “Houses” show an interesting distribution between the two cities. For houses up to 600 sq. yards Karachi (in orange) has more listings and are relatively high priced than Islamabad (in green). 600 sq. yards and above both cities go head to head with Islamabad being significantly high priced than Karachi in larger sizes.



As can be observed from above plot, in Farm Houses Islamabad (in orange) has more listing in varying sizes as compared to Karachi. In the plot below, there is not a lot to take as individual rooms listed as a property are rare in both cities, However they do have a few listed in one or two categories each, as shown below.



We will now further explore each of the city separately, using Foursquare API. For each of the listings in their respective categories, Foursquare API will give us the neighboring data. This data will be used for clustering and segmentation of each location in both cities. Based on similarities and differences between each clustered data for both cities and the visualizations above, we can make better decisions in terms of pricing and locality.

4. Segmenting & Clustering using Foursquare API and KMeans Clustering:

4.1. Gathering Data from Foursquare API:

In each of our listings we have following features:

- Location: Name of location where the listing is.
- Longitude & Latitude: Coordinates for particular location of listing.

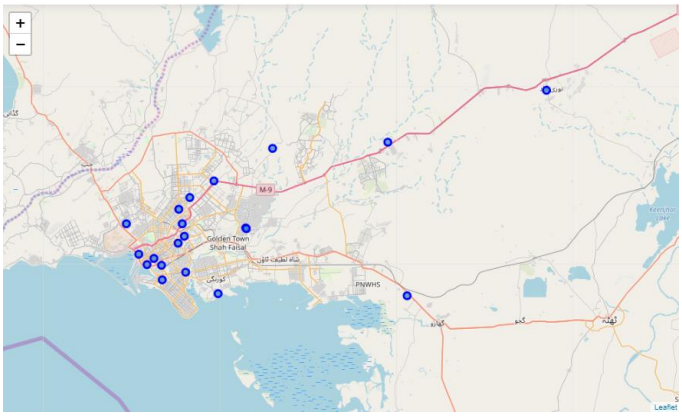
We can utilize these 3 features from our dataframe and use as input parameters for sending request to Foursquare API and Folium library, which in return will give us all the popular venues in a specified radius from each of the location in our dataset. These venues can then be plotted using Folium library to give us an interactive map of the city with datapoints marked on the map.

Our goal is to classify each location based on its top 10 most common popular venues nearby. From our dataset we gather that we have following locations for both Karachi (on left) and Islamabad (on right):

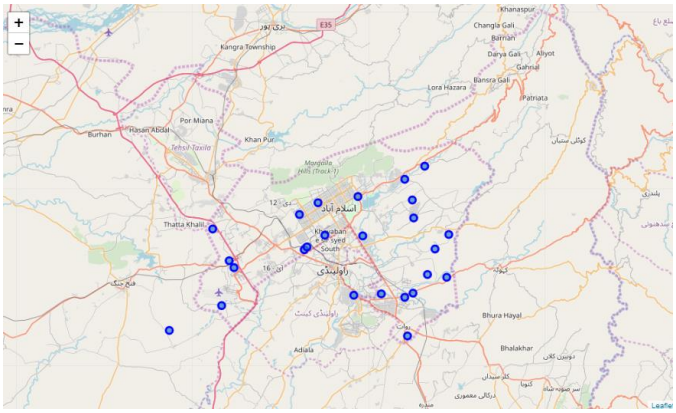
	location	longitude	latitude
0	Altaf Hussain Road	66.961912	24.919284
1	Dhabeji	67.521236	24.789435
2	Erum Villas	67.076658	24.896713
3	Gadap Road	67.253365	25.053956
4	Gobal Town	67.088202	24.966257
5	Goth Ibrahim Haidri	67.144605	24.792572
6	Gulshan-e-Ghazian	67.135509	24.996614
7	Gulshan-e-Jami	67.199470	24.911917
8	Hoshang Road	67.031274	24.843565
9	Humaira Town	67.200003	24.910078
10	Jamaluddin Afghani Road	67.064471	24.884039
11	Karachi Golf City	67.481782	25.066076
12	Mauripur Road	66.986321	24.863855
13	Nooriabad	67.798475	25.158484
14	Old Clifton	67.032915	24.817883
15	Old Queens Road	67.002655	24.845359
16	Peoples Colony	67.065446	24.945035
17	Royal Defence Tower	67.079122	24.831493
18	Shahra-e-Jahangir	67.073334	24.918598
19	Shahra-e-Liaquat	67.016451	24.856768

	location	longitude	latitude
0	12th Avenue	72.996002	33.624151
1	Agro Farming Scheme	73.236661	33.762376
2	Ahmed Town	72.831454	33.531531
3	Airline Avenue	72.855749	33.594532
4	Aiza Garden	73.150396	33.550364
5	Al Qaim Town	73.112588	33.646824
6	Alhamra Avenue	73.257129	33.624947
7	Atomic Energy Employee Society	73.201904	33.480422
8	Bokra Road	73.000788	33.628883
9	C-19	72.813134	33.658424
10	Chirah	73.284686	33.648973
11	Club Road	73.103435	33.711489
12	F-9	73.022861	33.701493
13	Islamabad View Valley	72.726960	33.489352
14	Jagiot Road	73.214595	33.676393
15	Karakoram Enclave 1	72.985775	33.681755
16	Korang Road	73.037486	33.647328
17	Lawyers Society	73.095081	33.548369
18	Malot	73.210916	33.706436
19	OPF Valley	73.241898	33.582412
20	PAF Tarnol	72.846068	33.605559
21	PTV Colony	73.196239	33.740507
22	Sehala Farm House	73.280159	33.577969
23	Sihala	73.196883	33.545091
24	Sihala Valley	73.212615	33.551548

Using Folium library, we now plot these locations on the map for better visualizing our data, before we call Foursquare API, first for Karachi (on left) and then for Islamabad (on right) below.



(Karachi)



(Islamabad)

Next, we use this location data and request Foursquare API for venues nearby. Following are the first 5 rows of data gathered for both cities, stored in two separate dataframes:

	Location	Location Latitude	Location Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Erum Villas	24.896713	67.076658	Aga Khan Sports & Rehabilitation Center	24.892157	67.079419	Gym / Fitness Center
1	Erum Villas	24.896713	67.076658	Shazz Supermarket	24.897260	67.079086	Supermarket
2	Erum Villas	24.896713	67.076658	Burger Inc	24.897225	67.078910	Burger Joint
3	Erum Villas	24.896713	67.076658	Dunkin'	24.904756	67.078900	Donut Shop
4	Erum Villas	24.896713	67.076658	McDonald's	24.891543	67.081012	Fast Food Restaurant
5	Erum Villas	24.896713	67.076658	Karachi Arts Council Auditorium	24.895419	67.062449	Theater

Total 331 venues were returned for locations in Karachi city

	Location	Location Latitude	Location Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	12th Avenue	33.624151	72.996002	Long Munir Legal service	33.630000	72.998503	Lawyer
1	12th Avenue	33.624151	72.996002	Kainaat Travels	33.614996	72.998674	Bus Station
2	12th Avenue	33.624151	72.996002	Ghosia	33.614955	72.993000	Bus Station
3	12th Avenue	33.624151	72.996002	Bilal Travels	33.613581	72.996500	Bus Station
4	12th Avenue	33.624151	72.996002	Lahori Tikka Point	33.614151	73.005882	BBQ Joint
5	Agro Farming Scheme	33.762376	73.236661	Chattar Park	33.769464	73.224375	Park

Total 120 venues were returned for locations in Islamabad city

Foursquare API returns with valuable information such as names of individual venues, their coordinates and their category. We can utilize “Venue Category” feature in table above in our KMeans clusters. Number of venues per location is also given below:

Location (Karachi)	Number of Venues	Location (Islamabad)	Number of Venues
Erum Villas	21	12th Avenue	5
Gobal Town	2	Agro Farming Scheme	1
Goth Ibrahim Haidri	2	Airline Avenue	1
Gulshan-e-Ghazian	3	Aiza Garden	5
Gulshan-e-Jami	3	Al Qaim Town	4
Hoshang Road	53	Bokra Road	4
Humaira Town	3	Club Road	26
Jamaluddin Afghani Road	63	F-9	26
Mauripur Road	4	Jagiot Road	3

Old Clifton	70	Karakoram Enclave 1	20
Old Queens Road	12	Korang Road	6
Peoples Colony	23	Lawyers Society	8
Royal Defence Tower	14	PTV Colony	8
Shahra-e-Jahangir	16	Sihala	2
Shahra-e-Liaquat	42	Sihala Valley	1

From the data above, we can observe that many locations were automatically dropped from consideration when there was no active data available on Foursquare API. This limitation however can be improved upon as more users are actively participating on that platform. Furthermore number of location fetched for Karachi is significantly higher than those for Islamabad, mainly due to the fact that Karachi is comparatively more densely populated than Islamabad and thus has more likelihood of user activity, reviews and posting of venues on Foursquare in a given region. Moving forward, we can now use this data and apply KMeans clustering.

4.2. KMeans Clustering | Separately for Karachi & Islamabad:

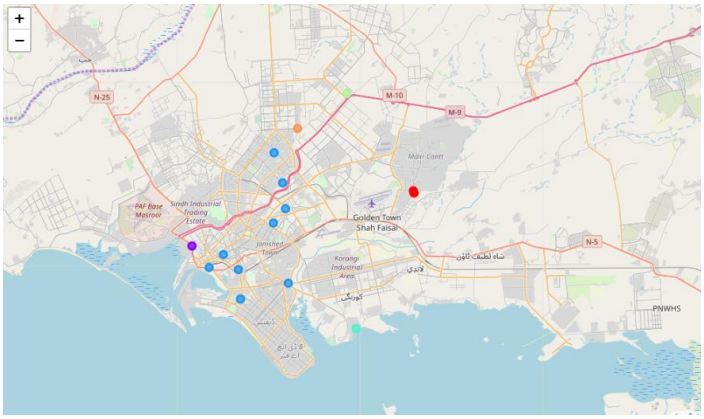
First step in applying KMeans to our categorical data, we need to identify the more recurring and highly rated venue categories and sort them in order. Following is the listed data of Karachi city with each venue category listed for all locations:

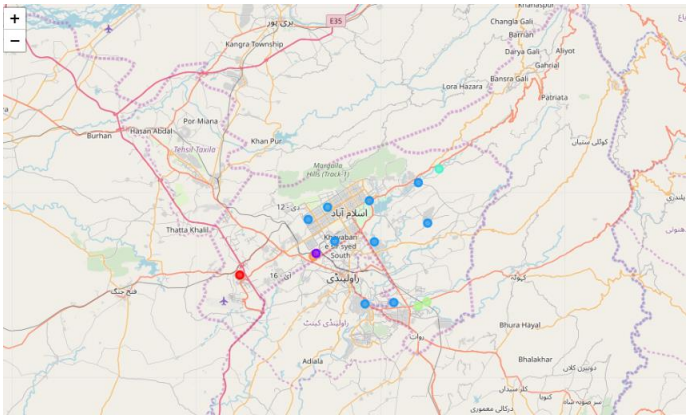
	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Erum Villas	Pharmacy	Fast Food Restaurant	Burger Joint	Pakistani Restaurant	Night Market	Department Store	Cricket Ground	Dessert Shop	Gym / Fitness Center	Pizza Place
1	Gobal Town	Breakfast Spot	Pharmacy	Historic Site	Fish & Chips Shop	Department Store	Dessert Shop	Diner	Donut Shop	Electronics Store	English Restaurant
2	Goth Ibrahim Haidri	Pier	Sports Club	Food & Drink Shop	Department Store	Dessert Shop	Diner	Donut Shop	Electronics Store	English Restaurant	Fast Food Restaurant
3	Gulshan-e-Ghazian	Restaurant	BBQ Joint	Train Station	Food & Drink Shop	Dessert Shop	Diner	Donut Shop	Electronics Store	English Restaurant	Fast Food Restaurant
4	Gulshan-e-Jami	Hotel	Café	Pizza Place	Athletics & Sports	Food & Drink Shop	Dessert Shop	Diner	Donut Shop	Electronics Store	English Restaurant
5	Hoshang Road	Hotel	Asian Restaurant	Fast Food Restaurant	Café	Performing Arts Venue	Japanese Restaurant	Italian Restaurant	Gym	Shopping Mall	Social Club
6	Humaira Town	Hotel	Café	Cricket Ground	Athletics & Sports	Food Court	Diner	Donut Shop	Electronics Store	English Restaurant	Fast Food Restaurant
7	Jamaluddin Afghani Road	Fast Food Restaurant	Pizza Place	Burger Joint	Department Store	BBQ Joint	Ice Cream Shop	Pakistani Restaurant	Chinese Restaurant	Dessert Shop	Market
8	Mauripur Road	Soccer Field	Asian Restaurant	Coffee Shop	Restaurant	Train Station	Dessert Shop	Diner	Donut Shop	Electronics Store	English Restaurant
9	Old Clifton	Dessert Shop	Café	Fast Food Restaurant	Burger Joint	Clothing Store	BBQ Joint	Bakery	Coffee Shop	Ice Cream Shop	Pizza Place
10	Old Queens Road	Fast Food Restaurant	Hotel	Seafood Restaurant	BBQ Joint	Italian Restaurant	Pizza Place	Market	Restaurant	Chinese Restaurant	Café
11	Peoples Colony	Fast Food Restaurant	Department Store	Ice Cream Shop	Bakery	Burger Joint	Bus Station	Cricket Ground	Pakistani Restaurant	Electronics Store	Pizza Place
12	Royal Defence Tower	Bakery	BBQ Joint	Convenience Store	Diner	Tea Room	Pizza Place	Coffee Shop	Juice Bar	Market	Department Store
13	Shahra-e-Jahangir	Bakery	Pizza Place	Indian Restaurant	Snack Place	Ice Cream Shop	Movie Theater	Restaurant	Pakistani Restaurant	Park	BBQ Joint
14	Shahra-e-Liaquat	Fast Food Restaurant	Market	Café	Ice Cream Shop	Shopping Mall	Clothing Store	Hotel	Asian Restaurant	Pakistani Restaurant	Beach

Following is the listed data of Islamabad city with each venue category listed for all locations:

	Location	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	12th Avenue	Bus Station	BBQ Joint	Lawyer	Zoo	Fast Food Restaurant	Hotel	Home Service	Gym / Fitness Center	Gym	Golf Course
1	Agro Farming Scheme	Park	Zoo	Hotel	Home Service	Gym / Fitness Center	Gym	Golf Course	Garden	Food Truck	Fish & Chips Shop
2	Airline Avenue	Toll Plaza	Zoo	Falafel Restaurant	Hotel	Home Service	Gym / Fitness Center	Gym	Golf Course	Garden	Food Truck
3	Aiza Garden	Pharmacy	BBQ Joint	Café	Donut Shop	Ice Cream Shop	Fast Food Restaurant	Hotel	Home Service	Gym / Fitness Center	Gym
4	Al Qaim Town	Asian Restaurant	Hotel	Lake	Restaurant	Farm	Home Service	Gym / Fitness Center	Gym	Golf Course	Garden
5	Bokra Road	Lawyer	Home Service	Mobile Phone Shop	Zoo	Ice Cream Shop	Hotel	Gym / Fitness Center	Gym	Golf Course	Garden
6	Club Road	Hotel	Gym	Golf Course	Café	Restaurant	Coffee Shop	Film Studio	Lounge	Middle Eastern Restaurant	Movie Theater
7	F-9	Park	Pizza Place	Café	Fast Food Restaurant	Bakery	Tea Room	Pakistani Restaurant	Italian Restaurant	Falafel Restaurant	Donut Shop
8	Jagiot Road	Zoo	Café	Asian Restaurant	Tibetan Restaurant	Falafel Restaurant	Home Service	Gym / Fitness Center	Gym	Golf Course	Garden
9	Karakoram Enclave 1	Café	Ice Cream Shop	Bakery	Coffee Shop	Pakistani Restaurant	Wings Joint	Gym	Diner	Pharmacy	Gym / Fitness Center
10	Korang Road	Indian Restaurant	Bakery	Big Box Store	Café	Department Store	Market	Film Studio	Hotel	Home Service	Gym / Fitness Center
11	Lawyers Society	Asian Restaurant	Playground	Italian Restaurant	Garden	Food Truck	Pharmacy	Ice Cream Shop	Supermarket	Bus Station	Film Studio
12	PTV Colony	Fast Food Restaurant	Pakistani Restaurant	Tibetan Restaurant	IT Services	Pharmacy	Juice Bar	Farm	Department Store	Fish & Chips Shop	Film Studio
13	Sihala	Light Rail Station	Pharmacy	Zoo	Farm	Hotel	Home Service	Gym / Fitness Center	Gym	Golf Course	Garden
14	Sihala Valley	Light Rail Station	Zoo	Farm	Hotel	Home Service	Gym / Fitness Center	Gym	Golf Course	Garden	Food Truck

Next, we ran KMeans algorithm with K = 6, for both Karachi and Islamabad separately and plotted the resulting clusters on map using Folium library (left = Karachi, right = Islamabad):





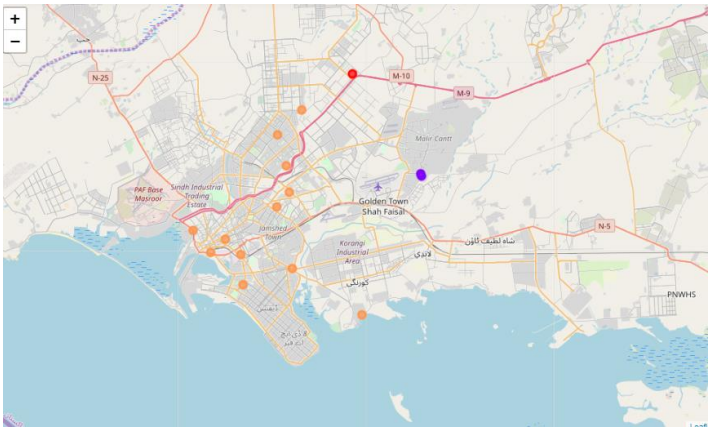
We further tabulate the clusters individually to observe similarities and differentiation between each cluster for both cities. From the clustered data, for both Karachi and Islamabad, we observe the following:

- Most common venues in majority of localities in Karachi are Fast Food Chains and Restaurants.
- Islamabad's cluster labelled #2 is more similar to Karachi in terms of Fast Food Chains than other clusters.
- Islamabad's clusters are diverse in terms of venues as they have Pharmacies, Hotels, IT shops and Fitness Centres in all of the clusters.
- Karachi's cluster #1 and #3 are differentiated only by top 2 venues in their respective localities.
- Islamabad's clusters #3 and #4 are also similar, however differentiated by frequencies of common venues.

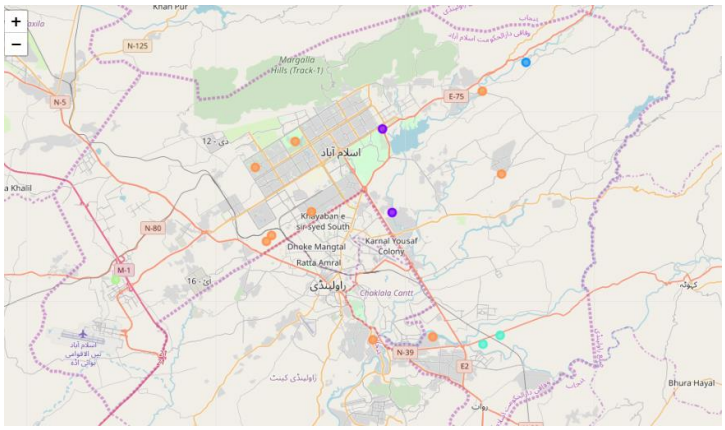
Above clustering gives a more improved perspective of the kind of cities these two are, however this information is still not enough to make a call on purchasing property. Customers and investors come from varying backgrounds and with that their taste, preferences and buying power varies as well. In order to entertain such diversity, we need to first compare all the localities together so that differences and similarities between both cities can be better visualized. In our observations from plotting each property type separately, we gather that there are significant difference between the two cities based on property types they offer and clustering can give us a different perspective altogether.

4.3. Inter-City Clustering Using KMeans:

Following similar step in previous section, we implement KMeans but this time combine the data of both cities and run KMeans with K = 6, plot map using Folium library, as followings:



(Karachi)



(Islamabad)

We then tabulate our data and observe the following:

- Data in clusters 0, 2, 3, 4 are completely distinct in terms of city-wise classification. Clusters 2, 3 and 4 complete comprise of locations from Islamabad and Cluster 0 is completely from Karachi.
- Clusters 1 and 5 show even distribution of localities from both cities. These clusters are differentiated based on vicinity to Hotels as Cluster 1 has top most venue category as Hotel, while other categories somewhat overlap, differentiating only in frequency.

5. Conclusion:

To conclude, we looked at the data from two different perspectives. We first clarified which type of property is readily available and at comparatively lower price in a city in our exploratory data analysis. We not only looked at relationship between Price and Size, we also explored other variables as well and how they affect the pricing overall. We then leveraged Foursquare API's data to better understand each locality and what it has to offer on top of its price tag. We segmented each locality based on similarities and differences between the venues and we observed where and how the two cities meet and differentiate.

6. Further Study:

Although the study itself contains all the steps and parameters to be self-contained, however following are the recommendations, based on above observations, which can prove useful in taking this study forward:

- Descriptive Analysis: Now that we understand each property type and each location, we can filter out property types based on recommended locations only and then prepare an analysis of which property listings are worth investing in.
- Comparative Algorithm: In this study KMeans clustering algorithm was used with K=6, however we can use other algorithm as well such as Agglomerative clustering and compare its results with KMeans and refine our study further based on the best possible algorithm.
- User Inputs: As our data fetched from Foursquare API did not return promising results, as expected for some localities, which resulted in those localities being dropped from comparison, we can look for other API or approached to better facilitate this issue and in turn give better input to our algorithm to run clustering on. This will vastly improve our understanding of data and help us make better decisions.