



# Applied Capstone IBM Data Science

COMPARISON OF CITIES – KARACHI VS ISLAMABAD

# Introduction

## Background

Pakistan's real estate market is currently in need of digital transformation with new online portals emerging. One of the top leading online real estate portal in world Zameen was founded in 2006 in Pakistan and serves as forefront of majority of online data.

## Case

Using Zameen's publicly available dataset on Kaggle, we analyse real estate landscape in Pakistan

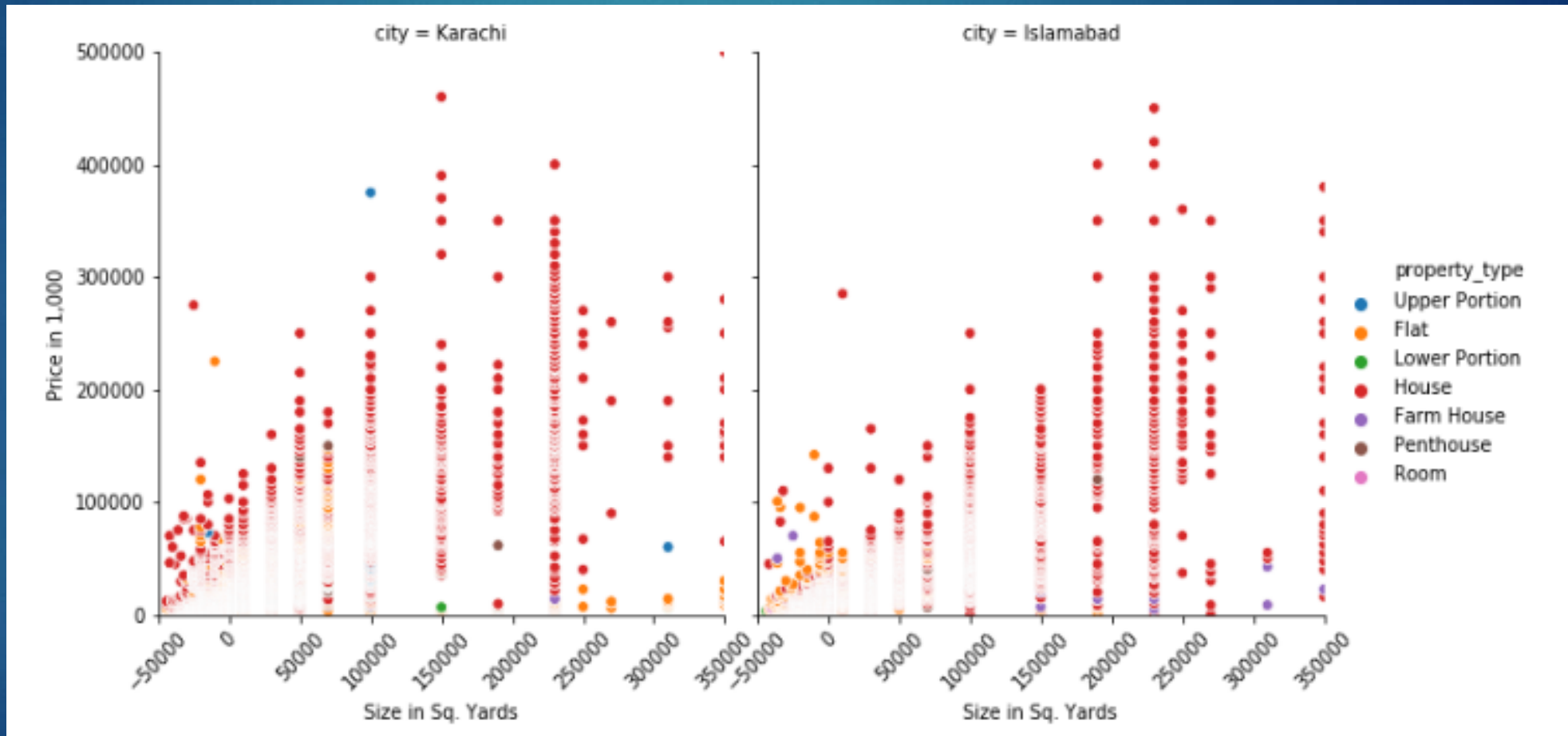
## Interest

Coming from a sales background and being in real estate industry, this study is also for self growth as well as to aid others.

# Data Wrangling

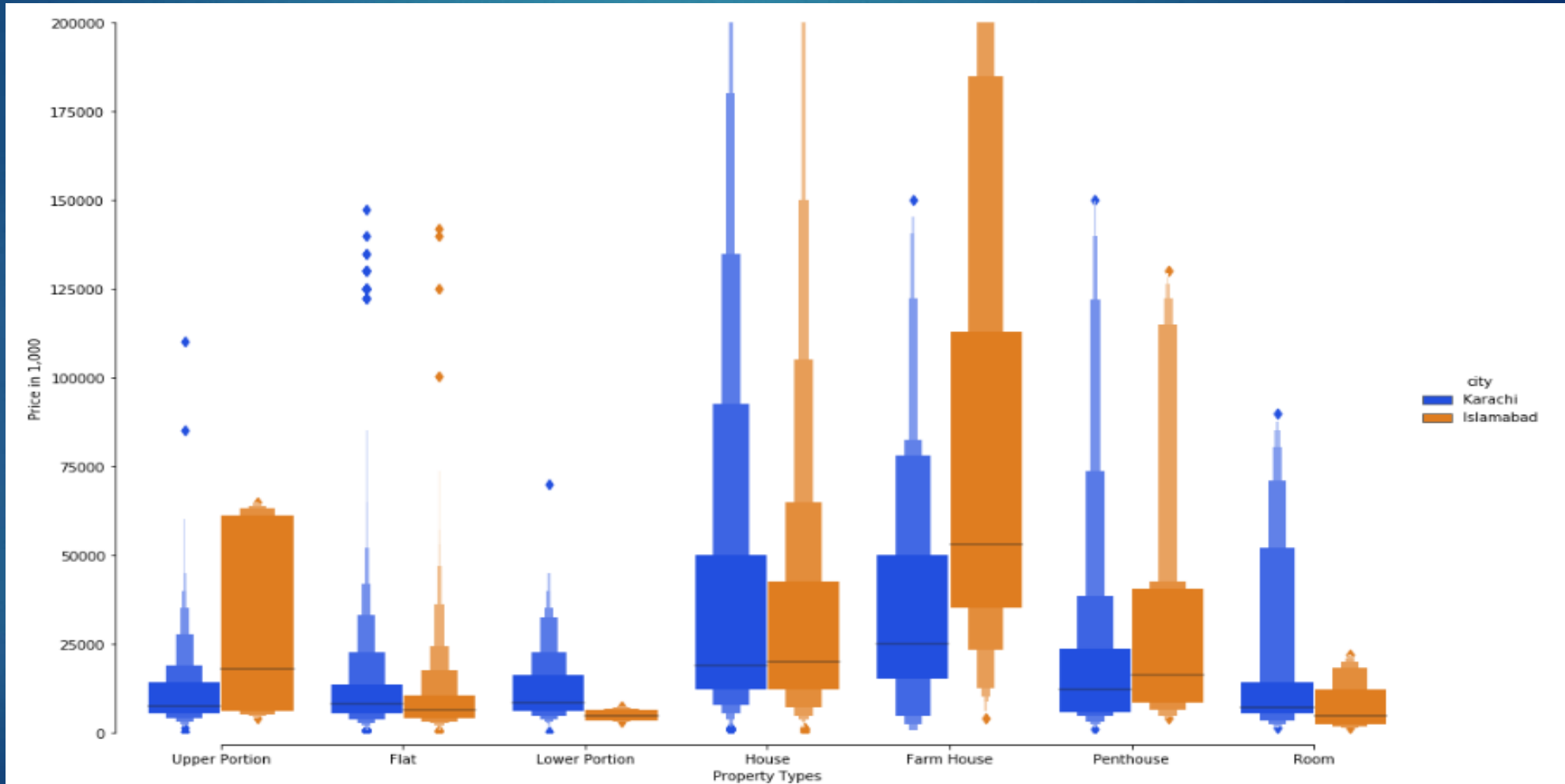
Data Source	Kaggle Open Datasets
Dataset	179838 entries, 310 columns
Relevant Features	Property Type, Price, Size, Location, City, Purpose, Longitude, Latitude
Methods	Dropping NaN values, Correcting data types, Correcting dimensional data, Removing inconsistent entries, Dropping irrelevant data, Transformation and Normalization of data
Result	9 features and 62,337 useful entries retained in proper format

# Exploratory Data Analysis: Size vs Price

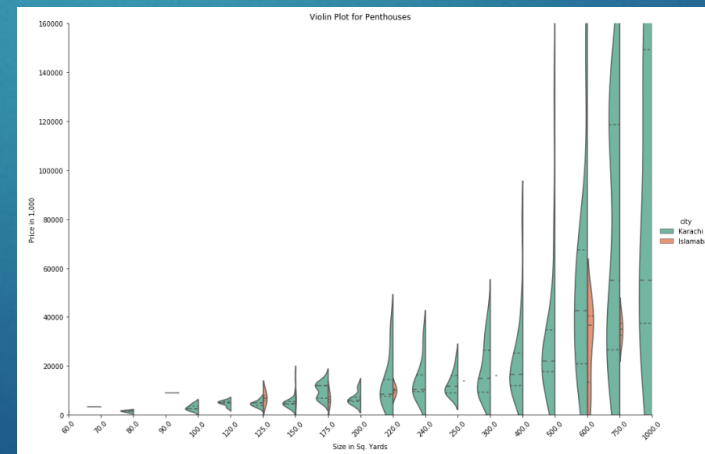
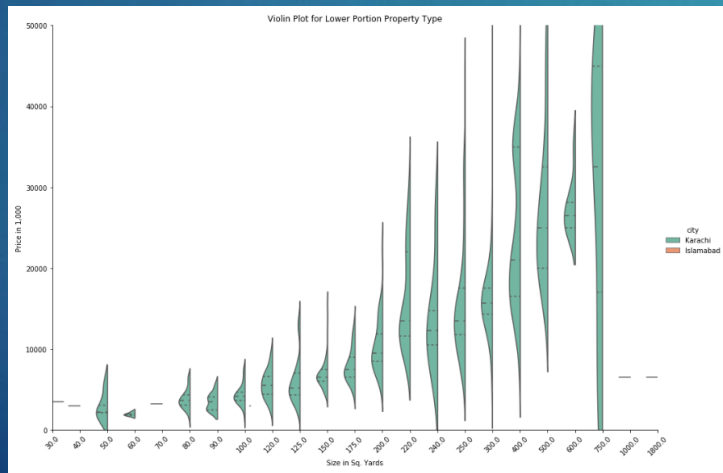
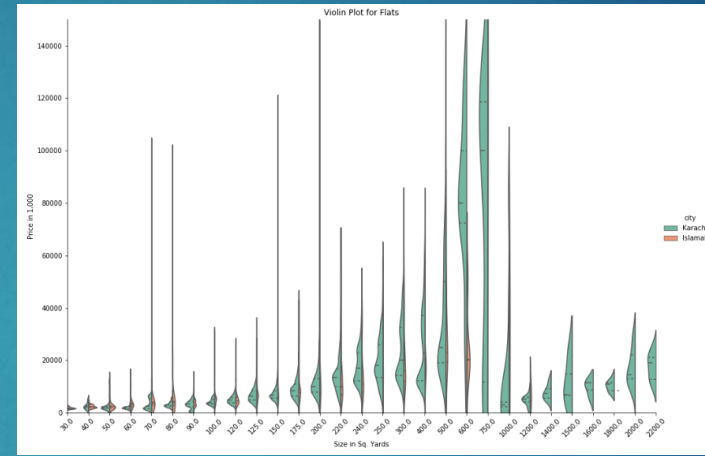
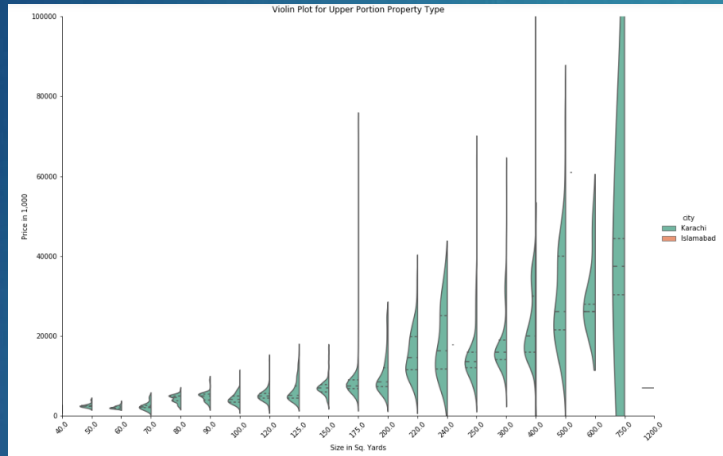




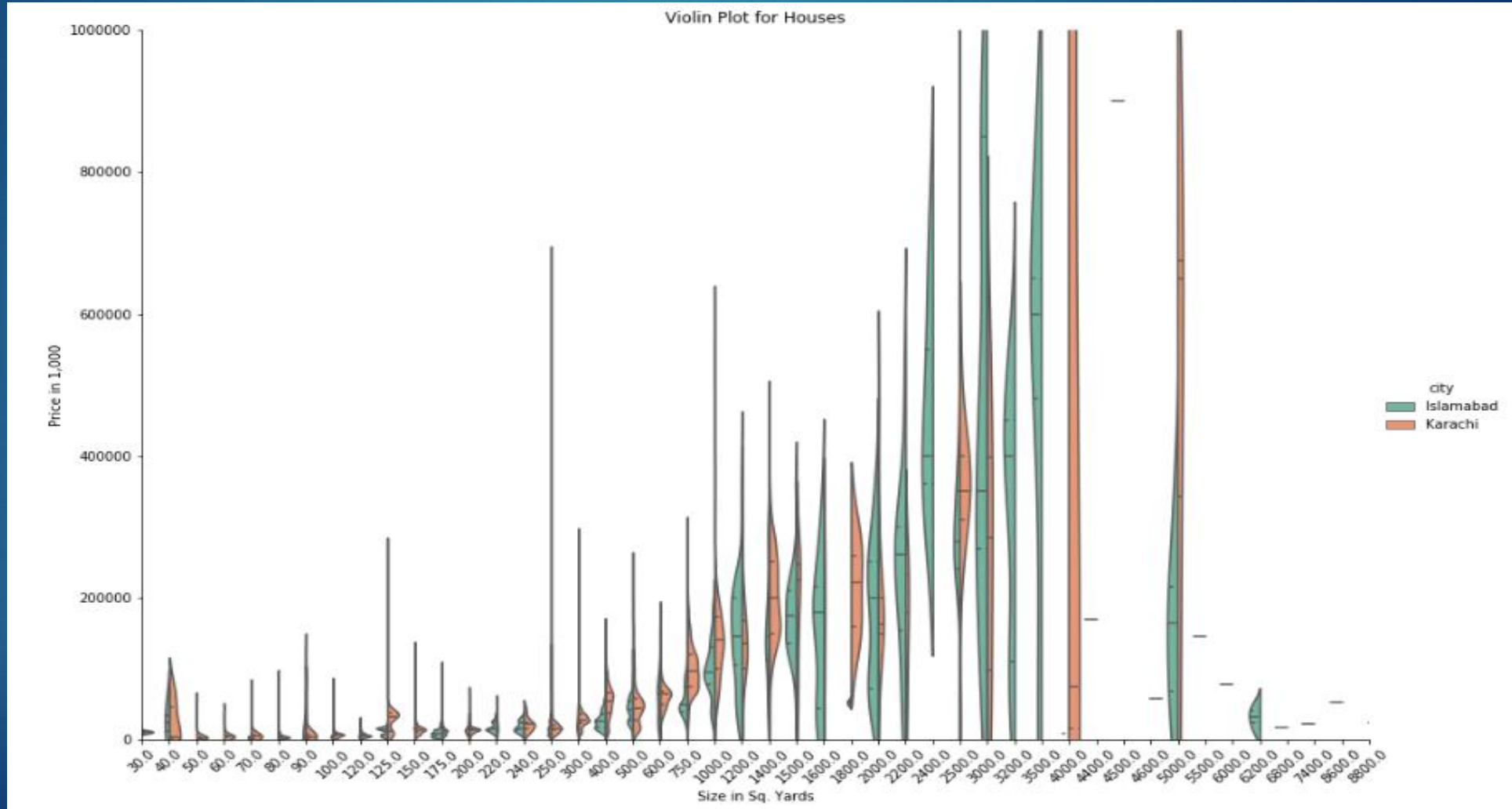
# Exploratory Data Analysis: Property Type vs Price



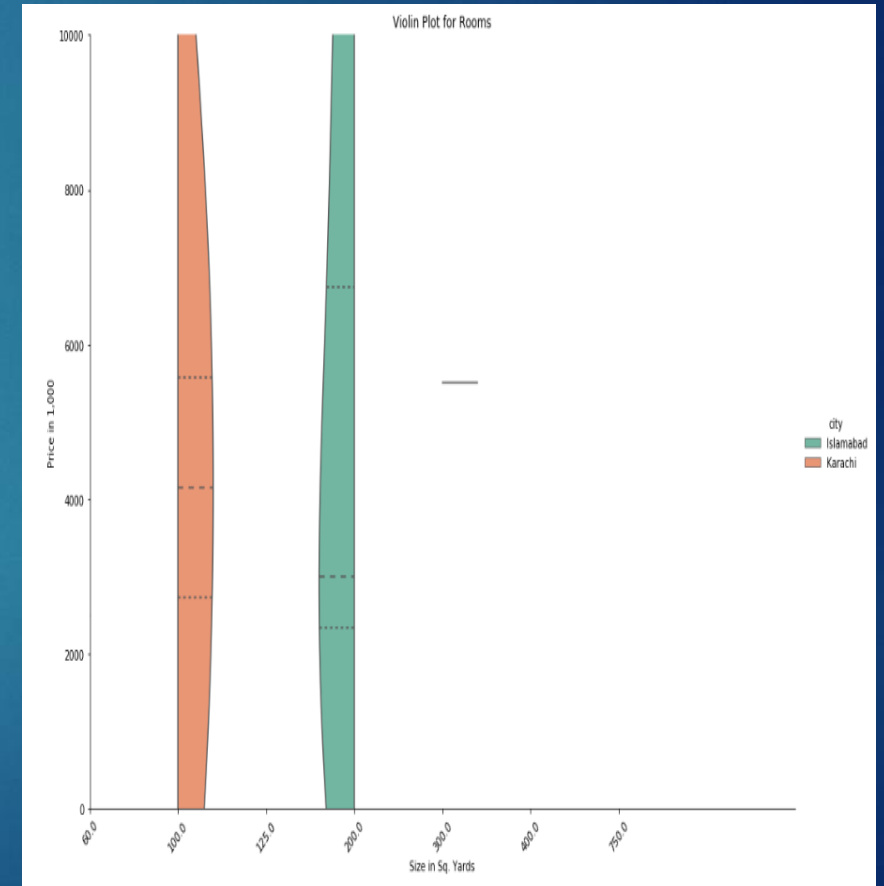
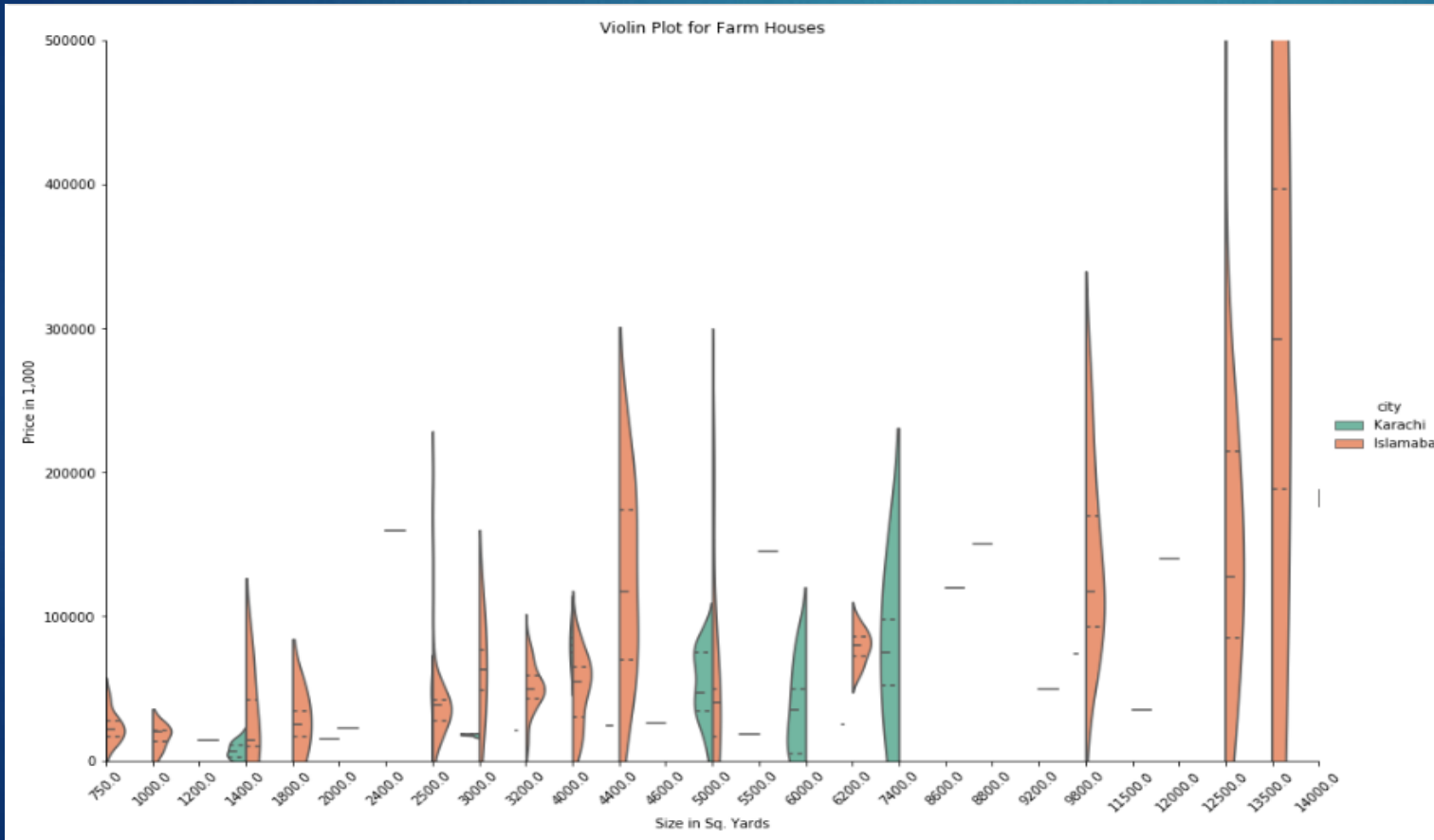
# Exploratory Data Analysis: Size vs Price – Filtered Property Type



# Exploratory Data Analysis: Size vs Price – Filtered Property Type

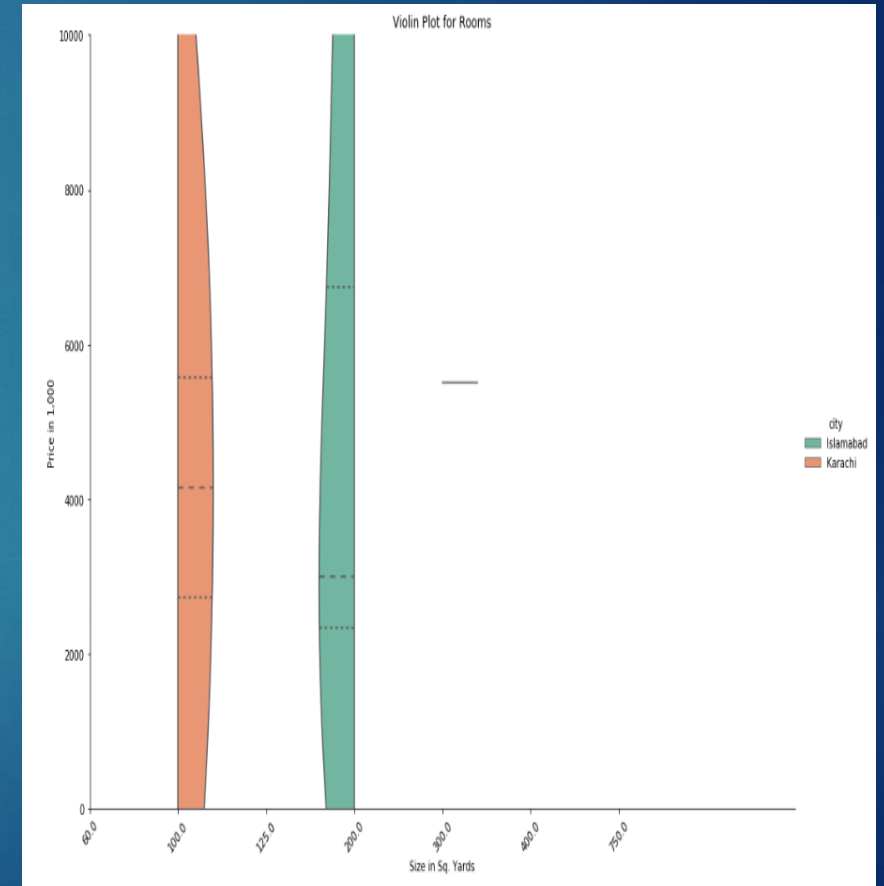
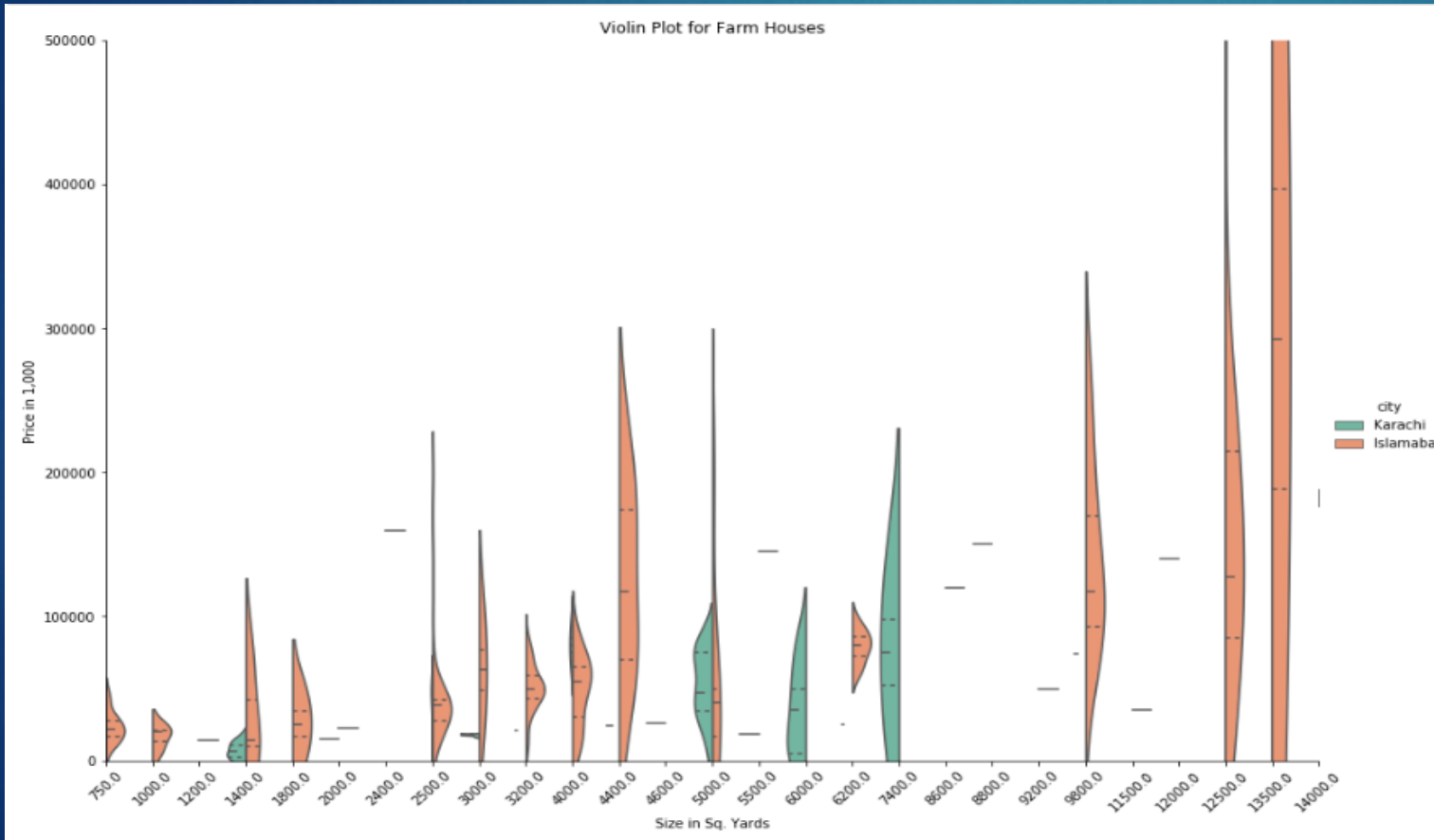


# Exploratory Data Analysis: Size vs Price – Filtered Property Type





# Exploratory Data Analysis: Size vs Price – Filtered Property Type



# Segmentation & Clustering:

## Venues Fetched from Foursquare API

### Dataframe Header

	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

	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

# Segmentation & Clustering:

## Venues Fetched from Foursquare API

Location (Karachi)	Number of Venues
Erum Villas	21
Gobal Town	2
Goth Ibrahim Haidri	2
Gulshan-e-Ghazian	3
Gulshan-e-Jami	3
Hoshang Road	53
Humaira Town	3
Jamaluddin Afghani Road	63
Mauripur Road	4
Old Clifton	70
Old Queens Road	12
Peoples Colony	23
Royal Defence Tower	14
Shahra-e-Jahangir	16
Shahra-e-Liaquat	42

Location (Islamabad)	Number of Venues
12th Avenue	5
Agro Farming Scheme	1
Airline Avenue	1
Aiza Garden	5
Al Qaim Town	4
Bokra Road	4
Club Road	26
F-9	26
Jagiot Road	3
Karakoram Enclave 1	20
Korang Road	6
Lawyers Society	8
PTV Colony	8
Sihala	2
Sihala Valley	1

# Segmentation & Clustering: Venue Categories Ranked (Karachi)

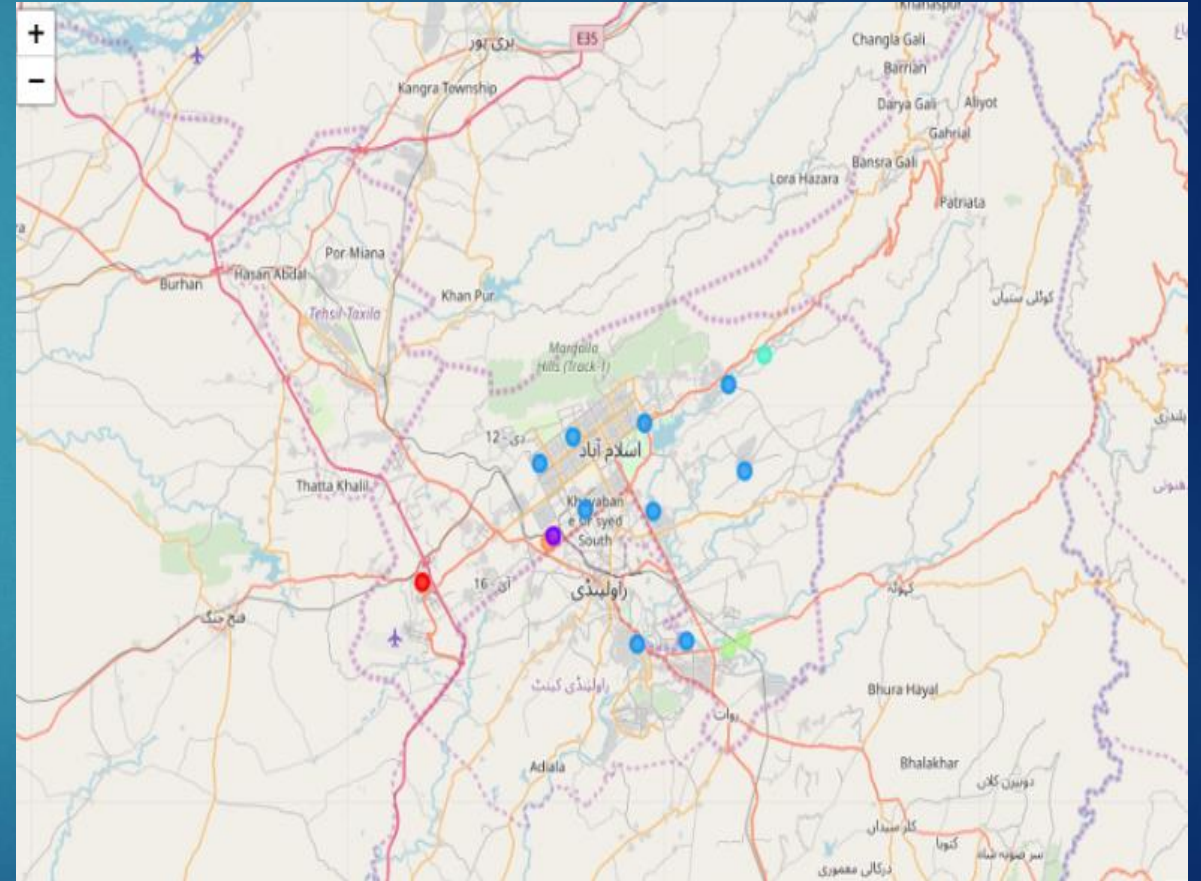
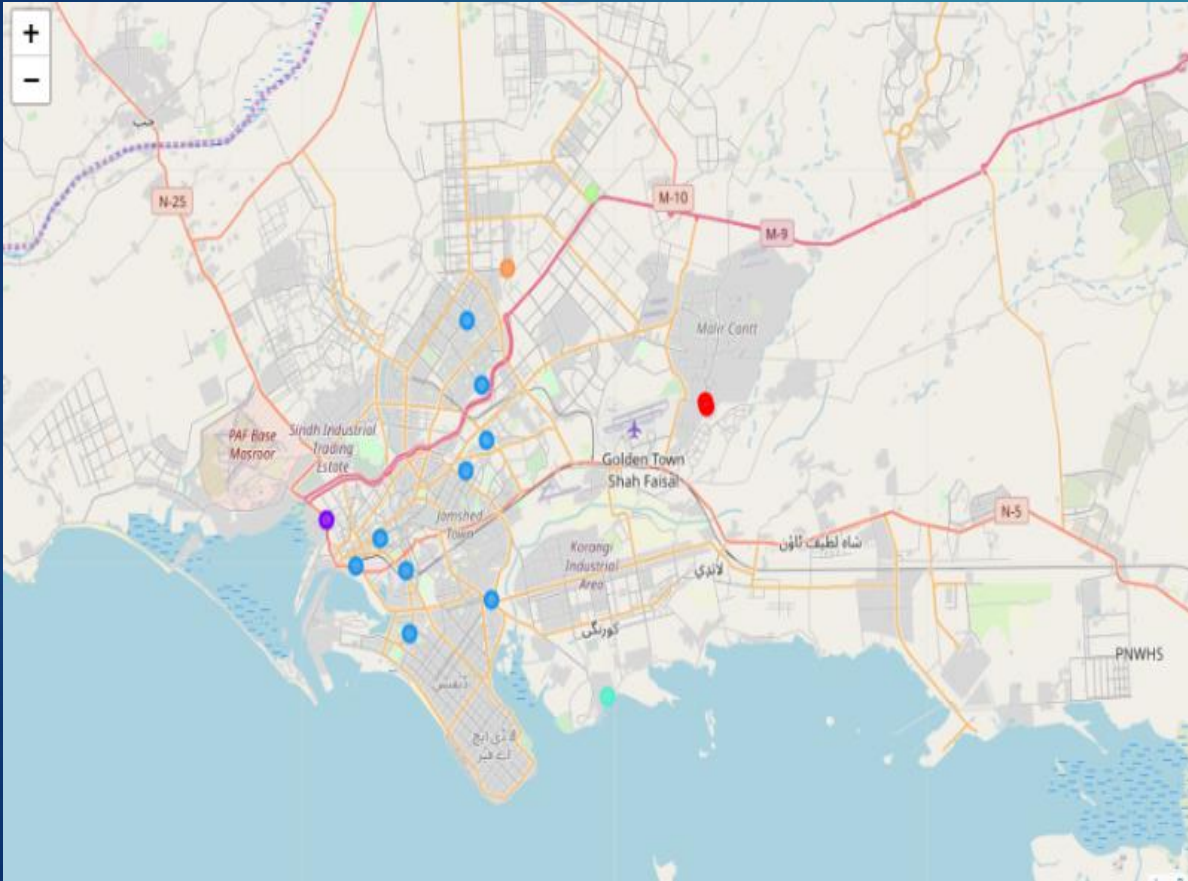
	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Erum Villas	Pharmacy	Fast Food Restaurant	Burger Joint	Pakistani Restaurant	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

# Segmentation & Clustering: Venue Categories Ranked (Islamabad)

	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	Jaglot 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



# Segmentation & Clustering: KMeans Clustering



# Segmentation & Clustering:

## KMeans Clustering Observations

---

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.

# Segmentation & Clustering:

## KMeans Clustering Observations

---

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.

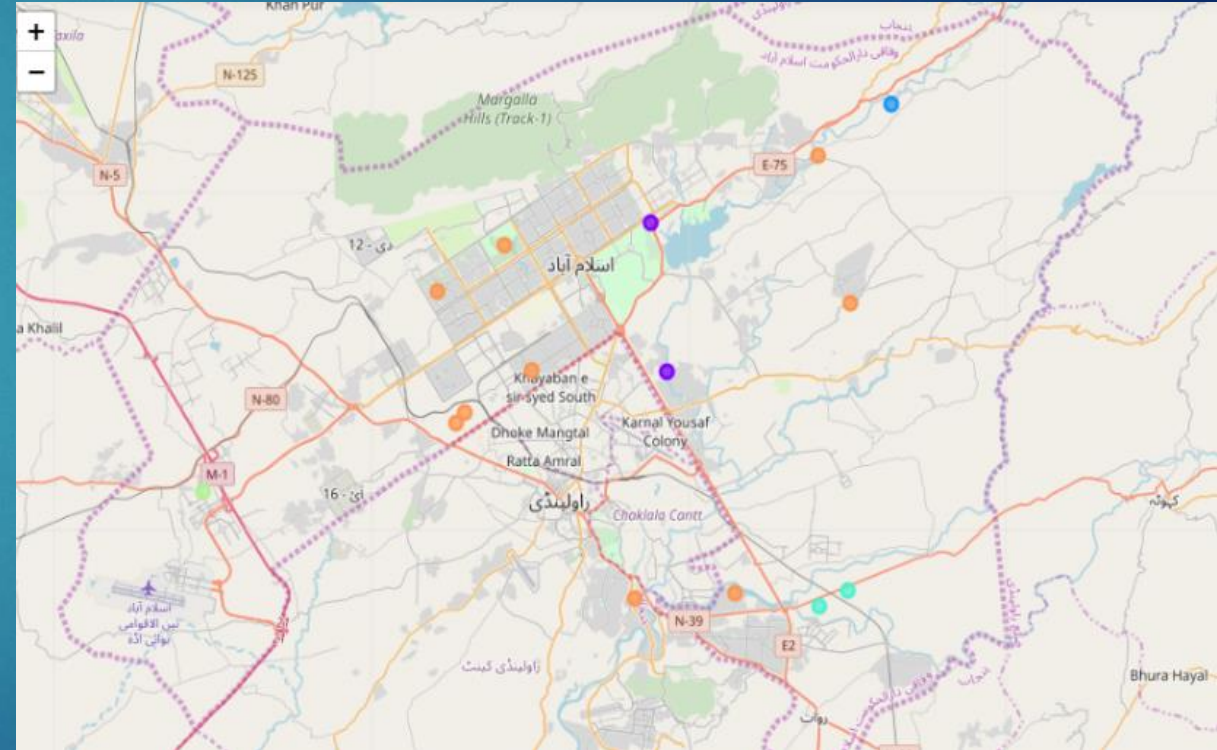
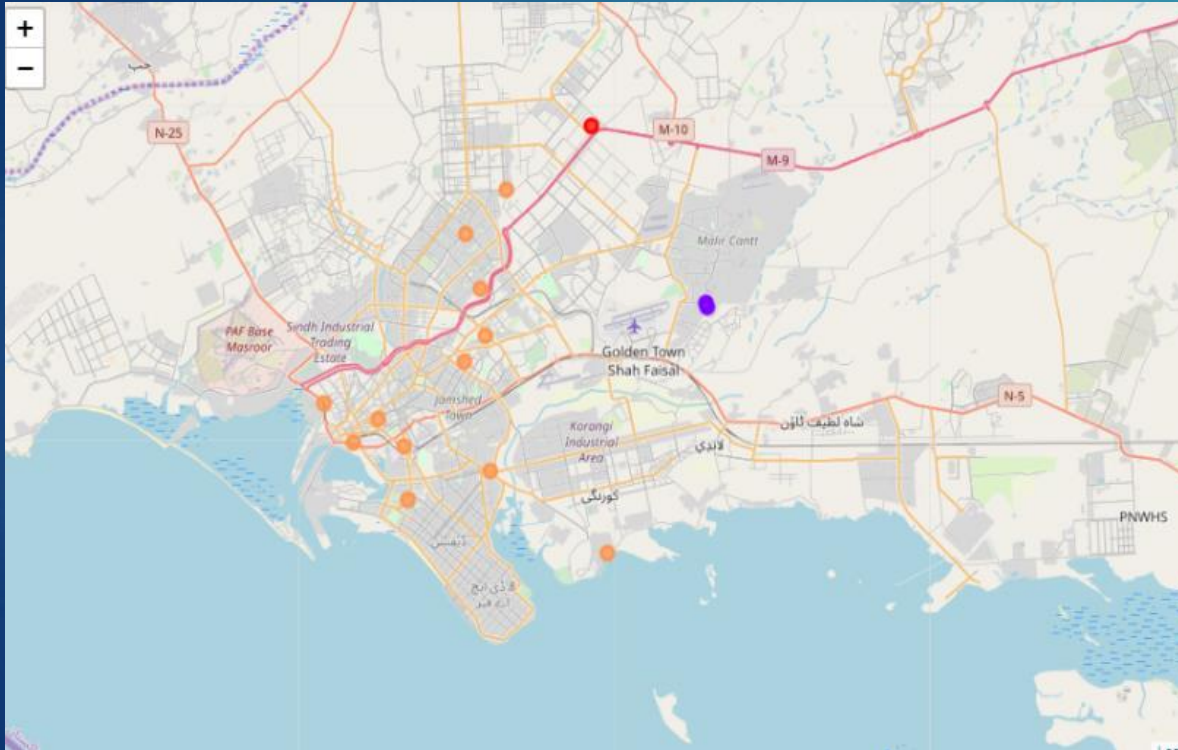
---

Islamabad's clusters #3 and #4 are also similar, however differentiated by frequencies of common venues.



# Segmentation & Clustering:

## KMeans Clustering Inter-City



# Segmentation & Clustering: KMeans Clustering Inter-City Observations

---

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.



# Conclusion

---

Clarified which type of property is readily available and at comparatively lower price in a city in our exploratory data analysis.

---

Explored other variables as well and how they affect the pricing overall.

---

Leveraged Foursquare API's data to better understand each locality and what it has to offer on top of its price tag.

---

Segmented each locality based on similarities and differences between the venues and we observed where and how the two cities meet and differentiate.

# Further Study

---

## 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.

---



# Thank You!

MIRZA MAWAZ KHAN

[WWW.GITHUB.COM/MAWAZ/](https://www.github.com/mawaz/)