Rental price vs. Neighborhoods

Coursera Data Science Capstone Project

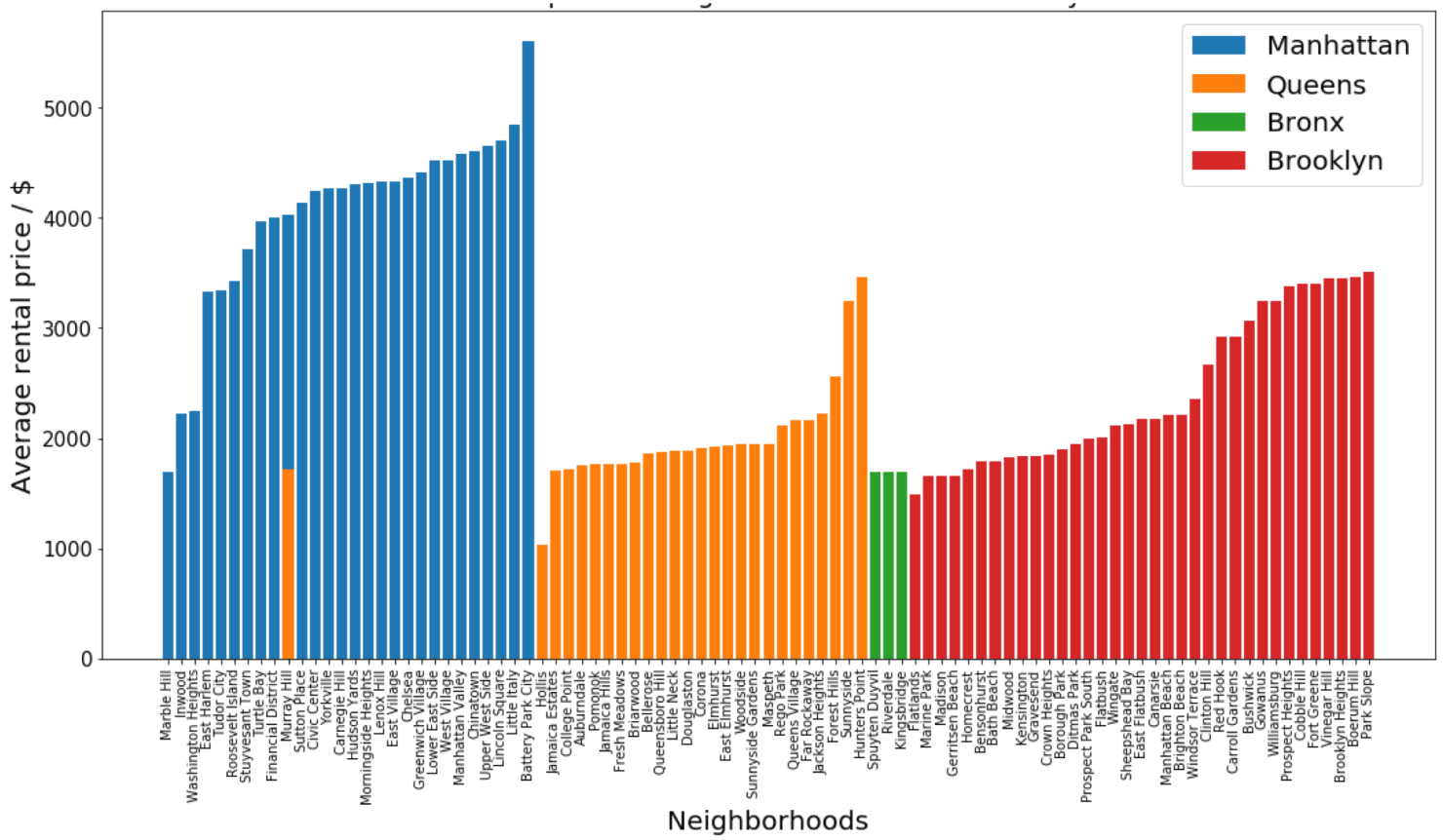
Weijing Dai

2019/04/11

Problem – New York City rental prices

- Renting a place in a city is always the first issue faced by employees, students and tourists.

- Finding a right place matching living requirement and financial budget is the common goal for this issue.



The reported rental prices vary considerably among the neighborhoods in New York City

- What factors cause this price variation.
- How those factors drive the price variation.
- Why those factors is essential in determining the price.

Reprocessing data

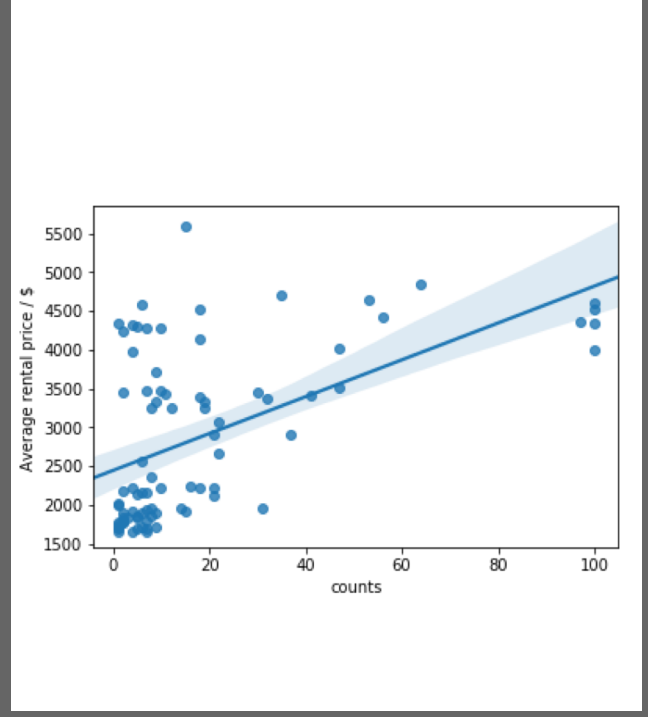
	Borough	Neighborhood	Average rental price / $	Latitude	Longitude
0	Manhattan	Marble Hill	1694	40.876551	-73.910660
1	Manhattan	Inwood	2225	40.867684	-73.921210
2	Manhattan	Washington Heights	2243	40.851903	-73.936900
3	Manhattan	Randalls and Wards Islands	2336	0.000000	0.000000
4	Manhattan	East Harlem	3334	40.792249	-73.944182

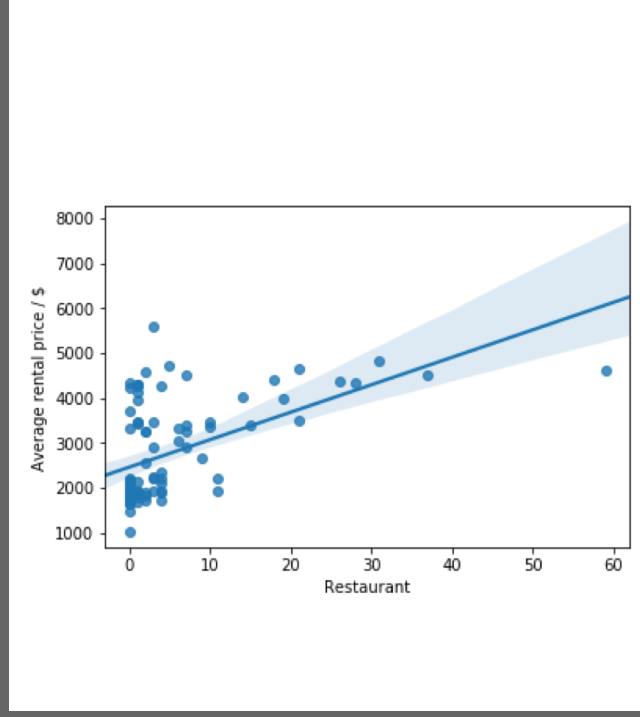
	Borough	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Manhattan	Marble Hill	40.876551	-73.91066	Metro North - Marble Hill Train Station	40.874698	-73.912052	Train Station
1	Manhattan	Marble Hill	40.876551	-73.91066	Marble Hill Youthmarket	40.874519	-73.910394	Farmers Market
2	Manhattan	Marble Hill	40.876551	-73.91066	marble hill pharmacy	40.875050	-73.909195	Pharmacy
3	Manhattan	Marble Hill	40.876551	-73.91066	MTA Subway - 225th St/Marble Hill (1)	40.874486	-73.909589	Metro Station
4	Manhattan	Marble Hill	40.876551	-73.91066	Marble Hill Playground	40.877765	-73.907994	Playground

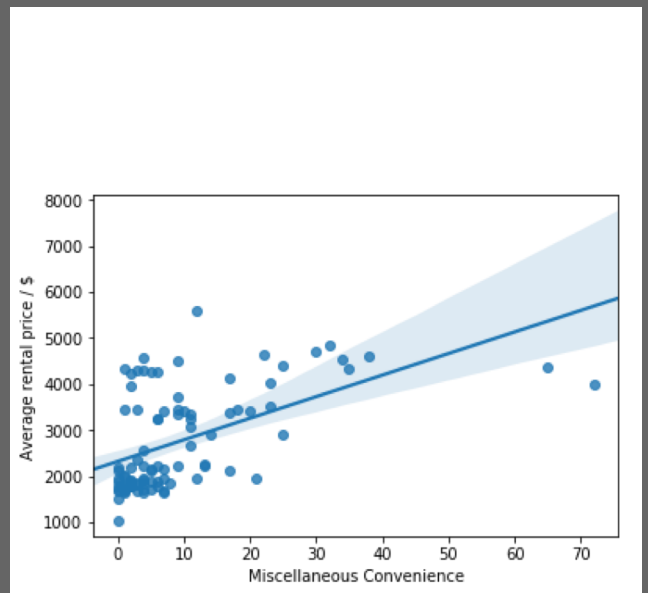
	Borough	Neighborhood	Average rental price / $	Latitude	Longitude	Restaurant	Entertainment	Miscellaneous Convenience
0	Manhattan	Marble Hill	1694	40.876551	-73.910660	0	0	7
1	Manhattan	Inwood	2225	40.867684	-73.921210	4	1	13
2	Manhattan	Washington Heights	2243	40.851903	-73.936900	3	0	13
3	Manhattan	East Harlem	3334	40.792249	-73.944182	6	2	11
4	Manhattan	Tudor City	3338	40.746917	-73.971219	0	0	9

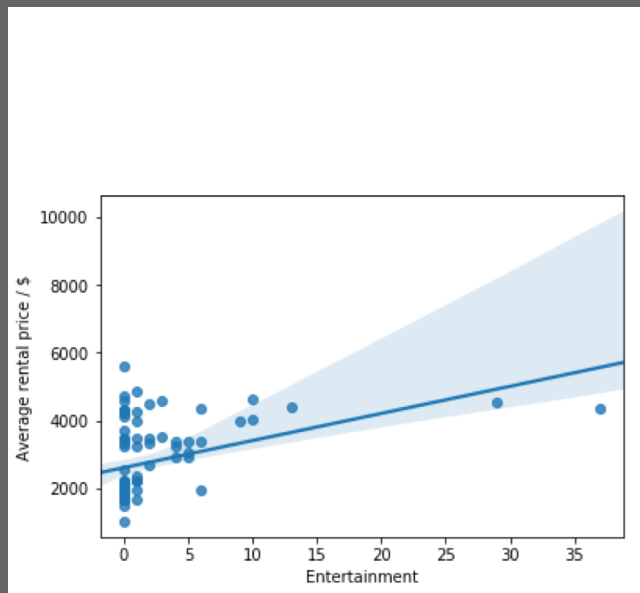
Simple linear regression

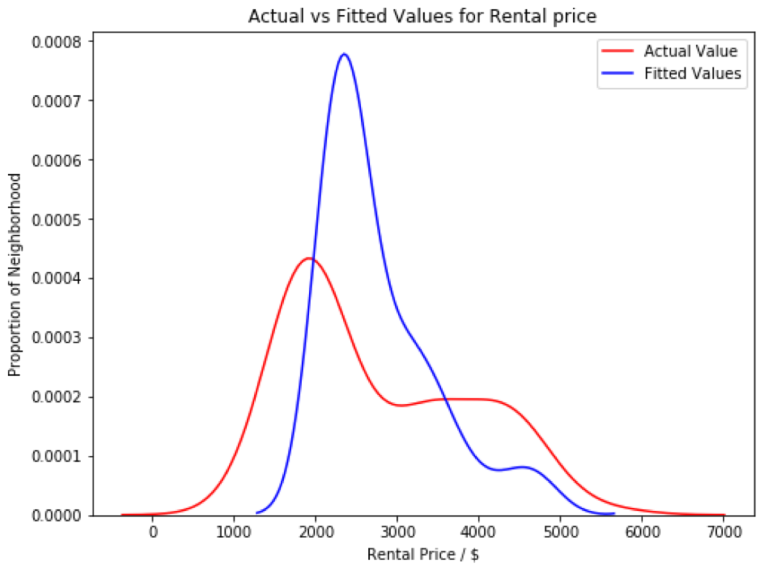
Table 1 The statistical coefficients			
	R-sqaured	Pearson coefficient	p-value
Price vs. Counts	0.291	0.540	1.418e-7
Price vs. Restaurant	0.275	0.524	8.248e-8
Price vs. Entertainment	0.154	0.393	0.0001
Price vs. Miscellaneous Convenience	0.290	0.538	3.170e-9

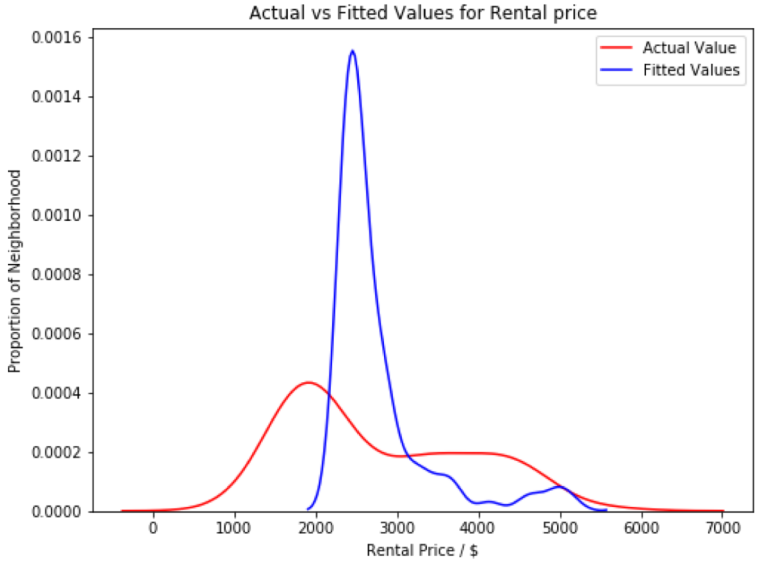










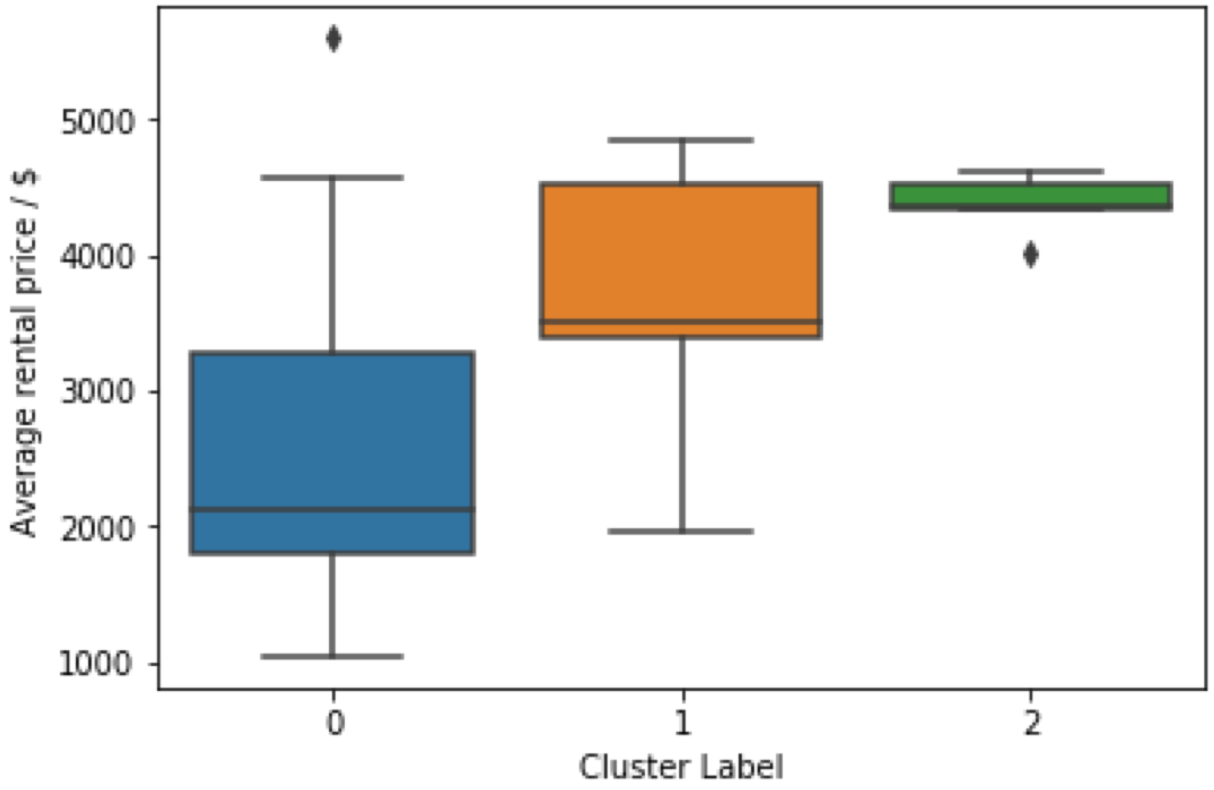


Multiple variables fitting

- The fitting methods discussed above indicate that the three selected variables are positively correlated to the average rental prices to some degree. However, there is no function capable to describe the correlation accurately.

Clustering vs. Price

Ideally, the overlaps of the price ranges between different clusters should be minimised if the clusters greatly differ from each other with respect to the average rental price.



In spite of the overlap of the price ranges is tremendous, the median price of each cluster splits significantly

Conclusion

- It proves that the increasing convenience by having more Restaurant and Miscellaneous Convenience types of venues can elevate the rental price of a neighborhoods.
- Wealthier people can afford higher rental price, suggesting that they possibly have more money to spend on other activities, such as dining outside, joining recreations and visiting different venues.
- this project also suggests the importance of the structure of data used for analysis. Better recommendation can be achieved by classifying the venues into less detailed but distinct categories.