TORONTO RENTAL PRICES

PROBLEM STATEMENT

- Rental prices in Toronto are among the highest in the country
- The market is competitive and decisions need to be made quickly
- The city is large enough that it can be difficult to estimate an appropriate price for any given property that considers not only the property, but also the neighborhood

DATA

- 2018 Toronto Rental Data

	Bedroom	Bathroom	Den	Address	Lat	Long	Price
0	2	2.0	0	3985 Grand Park Drive, 3985 Grand Park Dr, Mis...	43.581639	-79.648193	$2,450.00
1	1	1.0	1	361 Front St W, Toronto, ON M5V 3R5, Canada	43.643051	-79.391643	$2,150.00
2	1	1.0	0	89 McGill Street, Toronto, ON, M5B 0B1	43.660605	-79.378635	$1,950.00

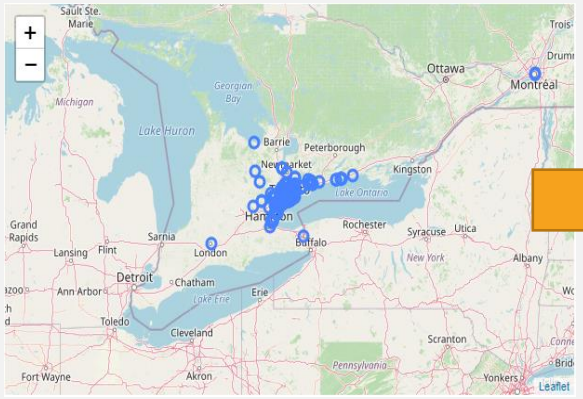
- Foursquare Location Data

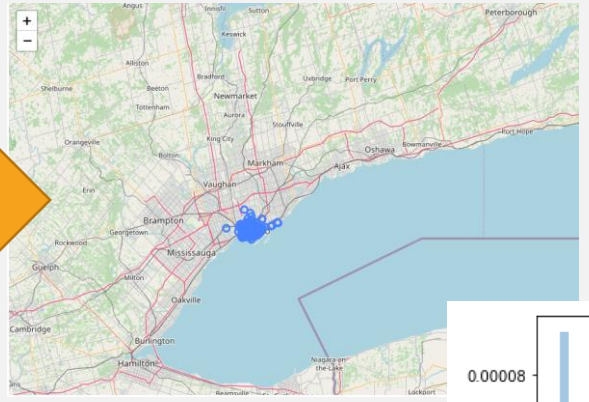
	Cluster	Cluster Latitude	Cluster Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	0	43.648388	-79.391511	Pai	43.647923	-79.388579	Thai Restaurant
1	0	43.648388	-79.391511	Byblos Toronto	43.647615	-79.388381	Mediterranean Restaurant
2	0	43.648388	-79.391511	YYoga	43.649725	-79.391983	Yoga Studio

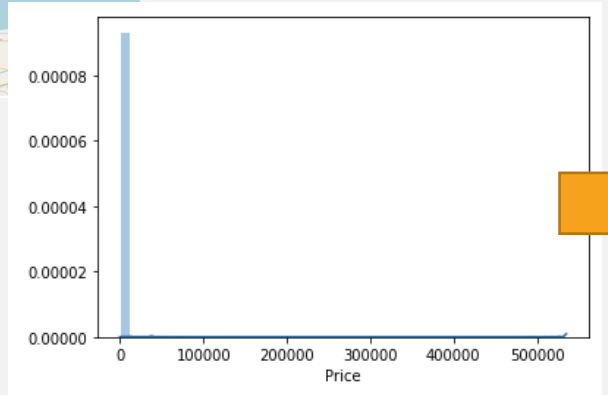
- Geopy distances (to determine distance of rental from city centre)

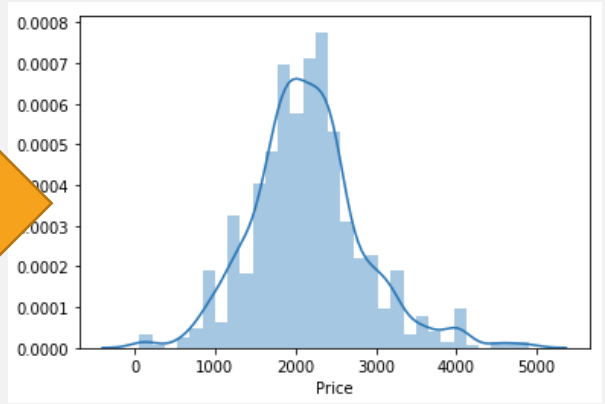
METHODOLOGY – DATA PREPARATION

- Remove outliers and include only Toronto data based on postal codes



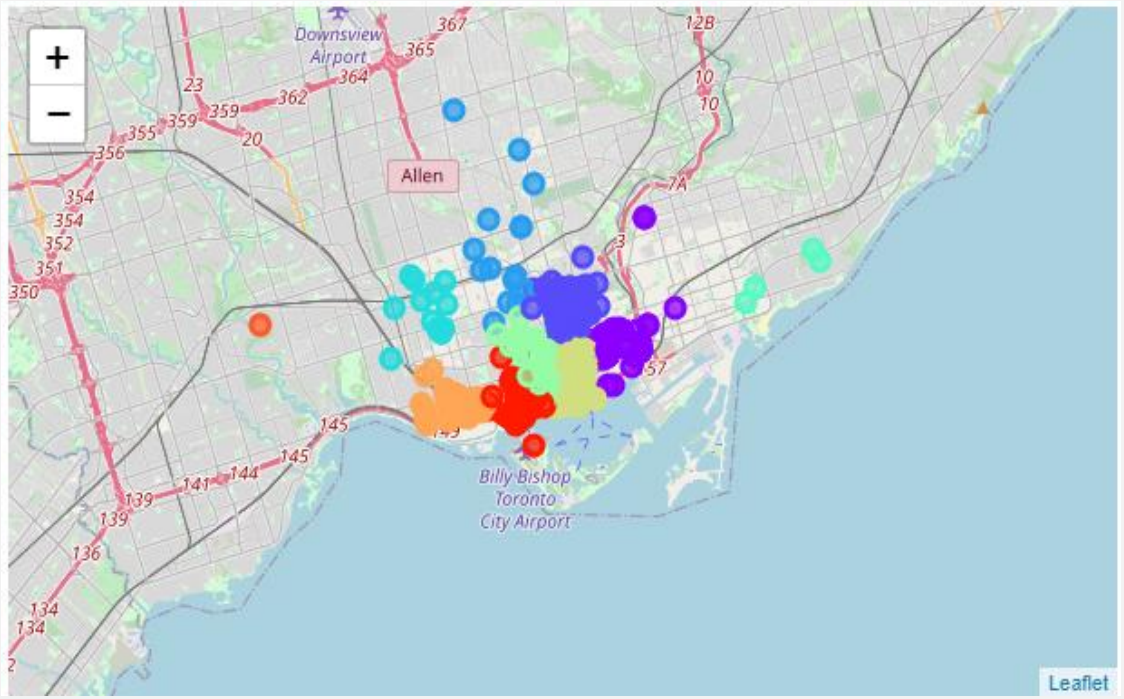






METHODOLOGY - CLUSTERING

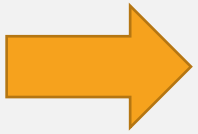
- Create 10 clusters based on location to collect Foursquare data using kmeans



METHODOLOGY – FOURSQUARE DATA

- Assume that the higher the number of restaurants among the top 100 venues in a neighborhood, the more desirable and more expensive the area
- Extract restaurants from 194 venue categories based on keywords:

```
'Comic Shop', 'Hostel', 'Food Truck', 'Gastropub', 'Beer Bar',
'Ramen Restaurant', 'Event Space', 'Shopping Mall',
'Japanese Restaurant', 'Cupcake Shop', 'Music Store',
'South American Restaurant', 'Middle Eastern Restaurant',
'Cocktail Bar', 'Comedy Club', 'Beer Store', 'Indian Restaurant',
'Wine Bar', 'Grocery Store', 'Salon / Barbershop', 'Whisky Bar',
'Vietnamese Restaurant', 'Tapas Restaurant', 'Bookstore',
'Ethiopian Restaurant', 'Korean Restaurant',
'Other Great Outdoors', 'Taco Place', 'Boutique', 'Breakfast Spot',
'Candy Store', 'Snack Place', 'Soup Place', 'Asian Restaurant',
'Farmers Market', 'Pool', 'Historic Site', 'Liquor Store',
'Chocolate Shop', 'Tech Startup', 'Deli / Bodega',
'Sushi Restaurant', 'Botanical Garden', 'Farm', 'Hotel Bar',
'Gift Shop', 'Pet Store', 'Athletics & Sports', 'Creperie',
```



- “Restaurant”
- “Joint”
- “Place”
- “Coffee Shop”
- “Breakfast Spot”

METHODOLOGY – MODEL DEVELOPMENT

- Split data into training and testing sets
- Using training set to perform multiple linear regression of increasing order polynomials
- Choose model based on highest R^2 value

X1	X2	X3	X4	X5	Y
# of Bedrooms	# of Bathrooms	Existence of Den	# of Restaurants	Distance to Centre	Price

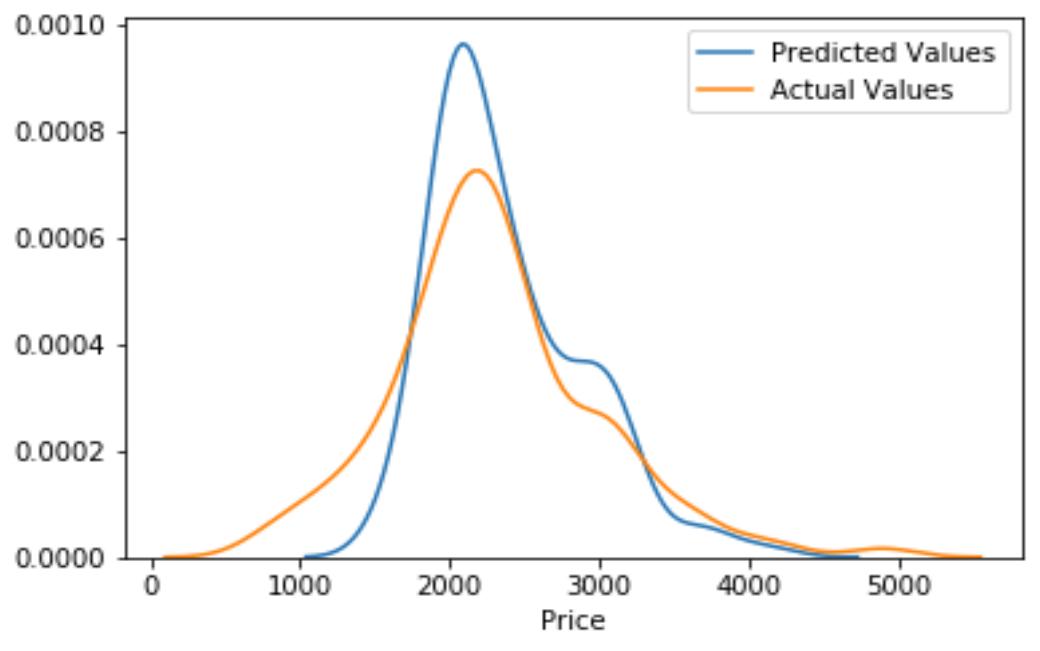
RESULTS

- Best model fit was achieved with a first order polynomial and an R^2 value of 0.364

Price = 548(# of bedrooms) + 492(# of bathrooms) + 367(existence of den) – 13(# of restaurants) – 71(distance to centre) + 1497

RESULTS

- Testing set distribution of actual versus predicted values



DISCUSSION & CONCLUSIONS

- Overall, model fit was not very strong with a maximum R^2 value of 0.364
- The impact of restaurants in the area was opposite of expected (i.e. more restaurants decreased the price)
  - There may be a better way to use Foursquare location data to evaluate the impact of nearby amenities on price
- The availability of more data may help to refine the model, but it appears that other inputs would be required to explain more of the variability