Content-based Recommender System Utilizing User Data

(Applied Data Science Capstone)

Jay Mitchell

1/15/2020

Venue locations value

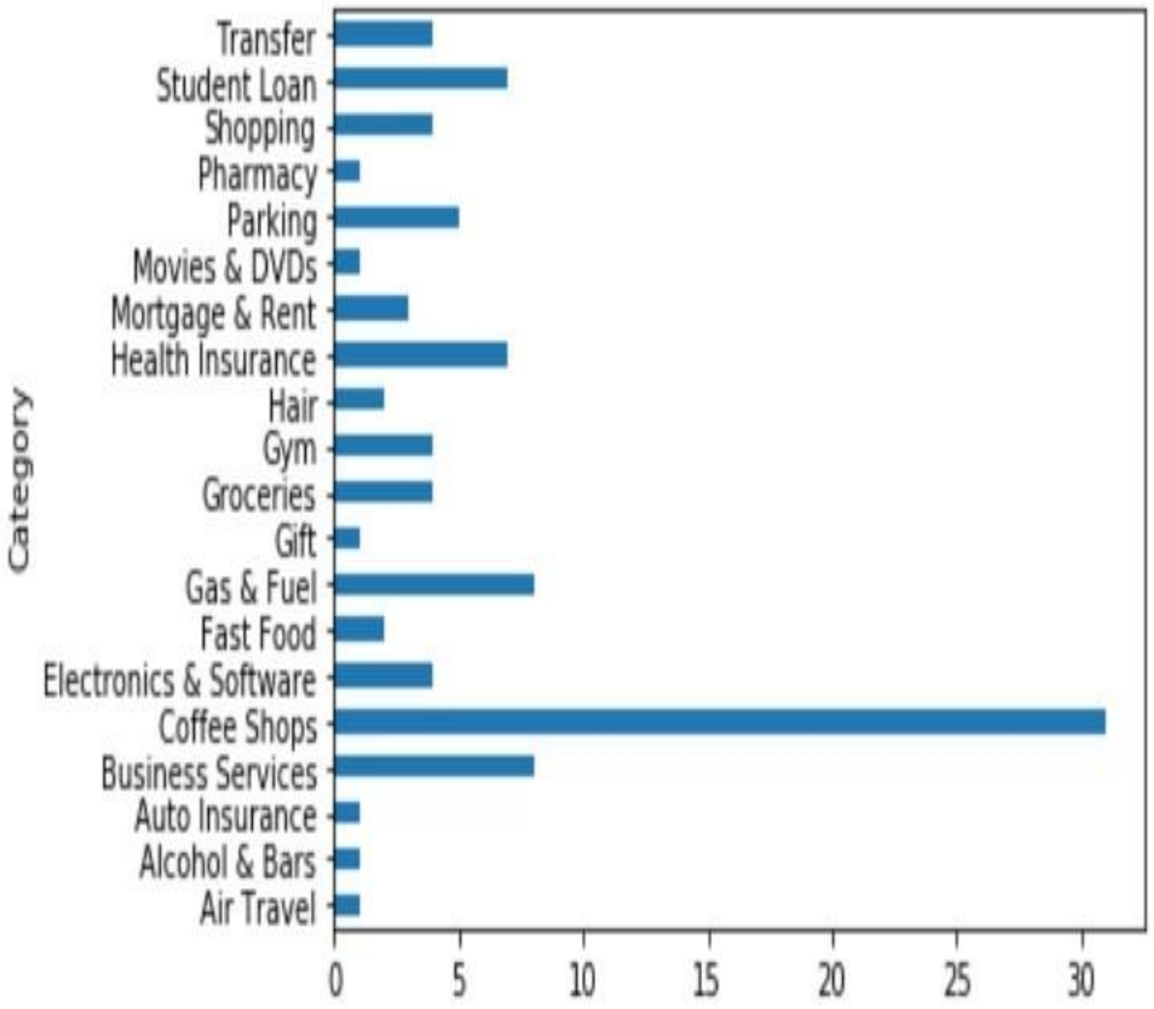
- Service providers
  - Capitalize on data (e.g., Foursquare, Mint)

- Customers
  - Find amenities similar to those normally used

- Venues
  - New patrons may increase sales

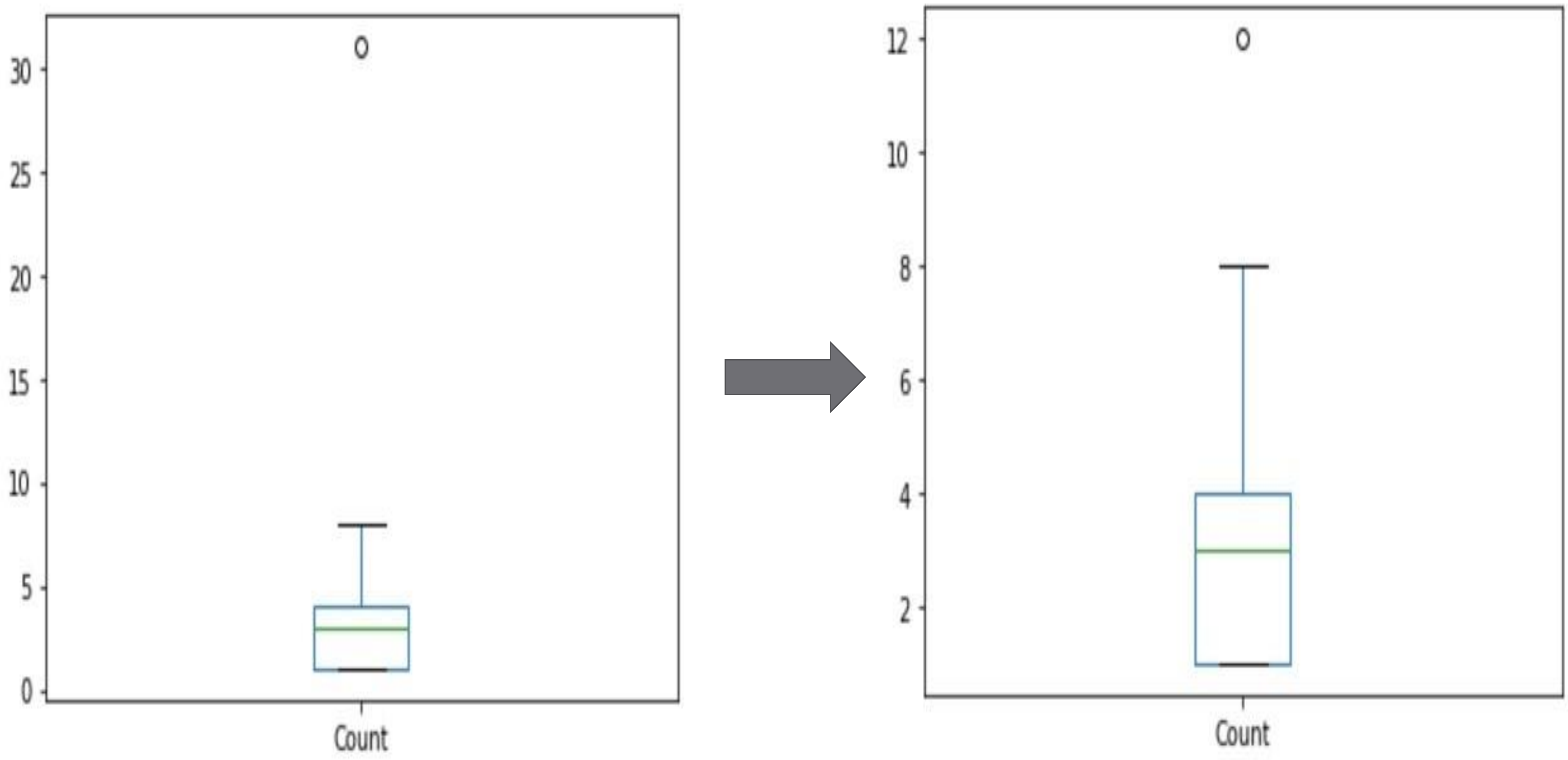
Data acquisition and cleaning

- Mint
  - Text file downloaded from website
  - Select columns and rows of interest to generate categories
  - Compare categories by statistical means (generate NormScore)



- Foursquare
  - Use Mint categories as search terms
  - Locate nearby venues using location
  - Find ratings for each venue

Outliers



Initial Results - Mint

- Simple feature scaling
  - Positive values
  - 0 to 1
  - Multiplied by 10 to equal venue rating weight (0 to 10 scale)
  - NormScore

	Category	Count	NormScore
0	Alcohol & Bars	1	0.833333
1	Coffee Shops	12	10.000000
2	Electronics & Software	4	3.333333
3	Fast Food	2	1.666667
4	Gas & Fuel	8	6.666667
5	Gift	1	0.833333
6	Groceries	4	3.333333
7	Gym	4	3.333333
8	Hair	2	1.666667
9	Movies & DVDs	1	0.833333
10	Pharmacy	1	0.833333
11	Shopping	4	3.333333

Initial Results - Foursqure

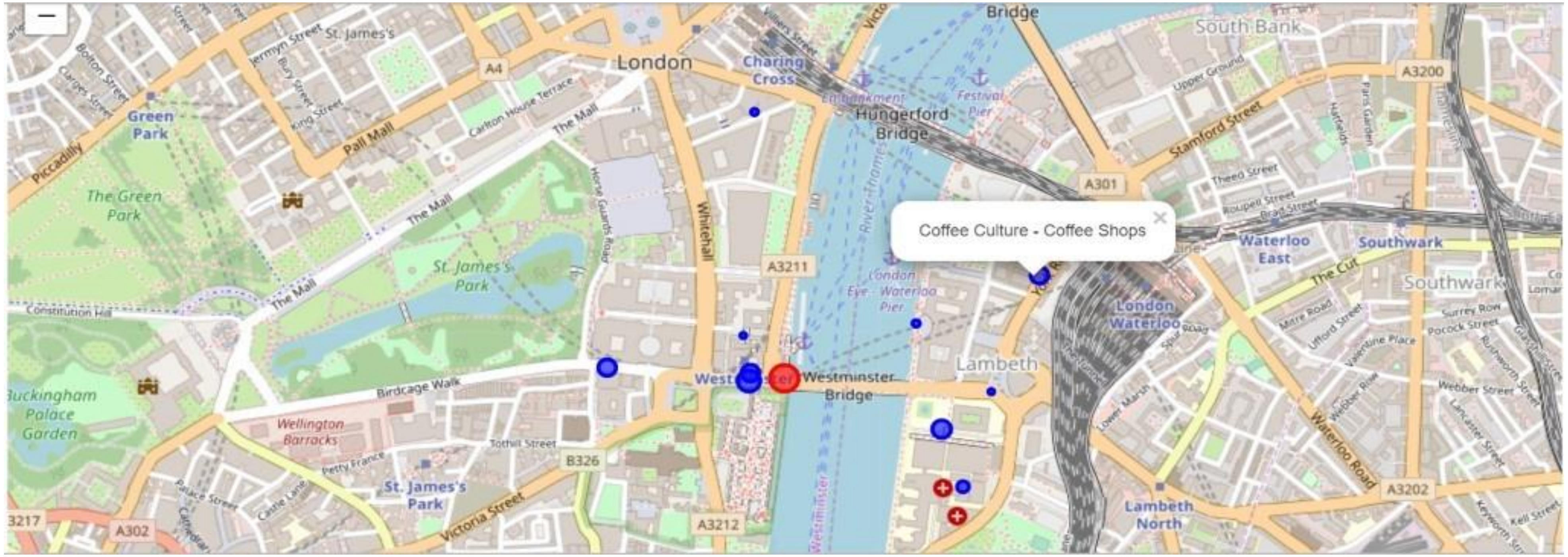
	categories	hasPerk	id	location.address	location.cc	location.city	location.country	location.crossStreet	location.distance
0	[{'id': '4bf58dd8d48988d123941735', 'name': 'W...	0.0	5d45ce66aea54e0008ec02dc	Unit 35, Old School Pl	GB	Croydon	United Kingdom	NaN	484.0
1	[{'id': '4bf58dd8d48988d1f9941735', 'name': 'F...	0.0	5d24d9d8dbde1100259c8731	13 Essex Street	GB	London	United Kingdom	NaN	652.0
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	[{'id': '4bf58dd8d48988d1e0931735', 'name': 'C...	0.0	5aaffc87918713b5d1dfa36	Westminster Tube Station	GB	London	United Kingdom	NaN	74.0
6	[{'id': '4bf58dd8d48988d1e0931735', 'name': 'C...	0.0	4cac1b4d3bfebdcb80e6db78	Portcullis House	GB	London	United Kingdom	NaN	71.0
7	[{'id': '4bf58dd8d48988d1e0931735', 'name': 'C...	0.0	4bbfc85bf964a520609c2ce3	St. Thomas Hospital	GB	London	United Kingdom	Lambeth Palace Road	348.0
8	[{'id': '4bf58dd8d48988d143941735', 'name': 'B...	0.0	4b546e28f964a520fba27e3	49 York Rd.	GB	Waterloo	United Kingdom	NaN	584.0
9	[{'id': '4bf58dd8d48988d1e0931735', 'name': 'C...	0.0	4fb60120e4b0a5e32711bb58	One Great George St	GB	London	United Kingdom	NaN	373.0
10	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
11	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
12	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
13	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
14	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
15	[{'id': '4bf58dd8d48988d118951735', 'name': 'G...	0.0	4debf42d45dd3093a8b36453	244-246 Westminster Bridge Rd	GB	London	United Kingdom	NaN	466.0

Final Results - Combined

- Venue rating and NormScore generated Weightscore
- Top 10 venues
- Venue Name, Latitude, and Longitude for map

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat	lng	neighborhood	postalCode	state	id	WeightScore	Category
2	Costa Coffee	[{'id': '4bf58dd8d48988d1e0931735', 'name': 'C...	Westminster Tube Station	GB	London	United Kingdom	NaN	74.0	[Westminster Tube Station, London, Greater Lon...	[{'label': 'display', 'lat': 51.50093364186963...	51.500934	-0.124805	NaN	S W1A	Greater London	NaN	79.000000	Coffee Shops
4	AMT Coffee	[{'id': '4bf58dd8d48988d1e0931735', 'name': 'C...	St. Thomas Hospital	GB	London	United Kingdom	Lambeth Palace Road	348.0	[St. Thomas Hospital (Lambeth Palace Road), Lo...	[{'label': 'display', 'lat': 51.49997320903263...	51.499973	-0.118974	NaN	SE1 7EH	Greater London	NaN	66.000000	Coffee Shops
3	Despatch Box Coffee Shop	[{'id': '4bf58dd8d48988d1e0931735', 'name': 'C...	Portcullis House	GB	London	United Kingdom	NaN	71.0	[Portcullis House, London, Greater London, SW1...	[{'label': 'display', 'lat': 51.50111406437572...	51.501114	-0.124743	NaN	SW1 A 2	Greater London	NaN	62.408193	Coffee Shops
6	Costa Coffee	[{'id': '4bf58dd8d48988d1e0931735', 'name': 'C...	One Great George St	GB	London	United Kingdom	NaN	373.0	[One Great George St, London, Greater London, ...	[{'label': 'display', 'lat': 51.50121966342386...	51.501220	-0.129109	NaN	NaN	Greater London	NaN	62.408193	Coffee Shops
5	Coffee Culture	[{'id': '4bf58dd8d48988d143941735', 'name': 'B...	49 York Rd.	GB	Waterloo	United Kingdom	NaN	584.0	[49 York Rd., Waterloo, Greater London, SE1 7N...	[{'label': 'display', 'lat': 51.50304865891913...	51.503049	-0.115980	NaN	SE1 7NJ	Greater London	NaN	62.000000	Coffee Shops
12	Gassiot House	[{'id': '4bf58dd8d48988d196941735', 'name': 'H...	Lambeth Palace Road	GB	London	United Kingdom	NaN	444.0	[Lambeth Palace Road, London, Greater London, ...	[{'label': 'display', 'lat': 51.498833, 'lng':...	51.498833	-0.118327	Lambeth	SE1 7EW	Greater London	NaN	41.605462	Gas & Fuel
19	The Gym & Club at County Hall	[{'id': '4bf58dd8d48988d176941735', 'name': 'G...	County Hall	GB	London	United Kingdom	Belvedere Road	302.0	[County Hall (Belvedere Road), London, Greater...	[{'label': 'display', 'lat': 51.50207893534426...	51.502079	-0.119739	NaN	SE1 7PB	Greater London	NaN	25.333333	Gym
17	ESPA Life Gym	[{'id': '4bf58dd8d48988d176941735', 'name': 'G...	10 Whitehall Pl	GB	London	United Kingdom	NaN	597.0	[10 Whitehall Pl, London, Greater London, Unit...	[{'label': 'display', 'lat': 51.50631843415083...	51.506318	-0.124648	NaN	NaN	Greater London	NaN	24.666667	Gym
21	Hotel Gym	[{'id': '4bf58dd8d48988d176941735', 'name': 'G...	NaN	GB	NaN	United Kingdom	NaN	435.0	[United Kingdom]	[{'label': 'display', 'lat': 51.50074, 'lng':...	51.500740	-0.117470	NaN	NaN	NaN	NaN	20.802731	Gym
20	Westminster Gym	[{'id': '4bf58dd8d48988d176941735', 'name': 'G...	Derby Gate, 1 Canon Row	GB	City of Westminster	United Kingdom	NaN	130.0	[Derby Gate, 1 Canon Row, City of Westminster,...	[{'label': 'display', 'lat': 51.50185929300493...	51.501859	-0.124989	NaN	SW1A 2JN	Greater London	NaN	20.802731	Gym

Output



- Folium mapping tool
- Current location (Large red dot)
- Venues (Blue dots with the radius proportional to WeightScore)
- Interactable (venue name and category)

Discussion

- Effective at generating local venues
- Data
  - Utilize other data sources based on the particular user
  - Update data from time to time to ensure relevancy
  - More generalized categories may be used
  - k means clustering algorithm to determine similar categories
- Processing
  - Final recommendation skewed
  - Counts may not be the best way to determine a user profile
  - Reduce the impact of outliers
  - Binning Profit
- Profit
  - Popup label hyperlink to incentive
  - Include income to data providers and app developer

Conclusion

- Recommendation engine profiled a user
- Historical data
- Determine nearby venues
- Venues rating
- Interactive Folium map

Code: https://github.com/jmitchell4390/Coursera_Capstone/blob/master/Capstone_Project.ipynb

Map: https://nbviewer.jupyter.org/github/jmitchell4390/Coursera_Capstone/blob/master/Capstone_Project.ipynb

Data: https://github.com/jmitchell4390/Coursera_Capstone/blob/master/transactions.txt