IBM Capstone Presentation

Using Foursquare Location Data

Connor Fairbanks

Problem:

Using Foursquare location data, what US cities would be best suited for an app that recommended local businesses.

City Criteria:

- High population to allow for high user base
- High percentage of Foursquare venues in the Foursquare database to allow for high quality and robust recommendations.

Desired Result:

A popular app that users would enjoy and would eventually allow for advertising revenue to be earned due to the high amount of regular users.

Data Sources:

- Wikipedia page showing US cities ranked by population
- page included city longitude and latitude data

- Foursquare Location data accessed through search API

Data Manipulation tools used:

- Pandas library in python used to read in data from Wikipedia and clean the dataframes

How the data was cleaned:

- removed unwanted characters from titles and cells
- dropped unneeded cells
- split location cell into longitude and latitude columns
- extracted top 50 cities by population and made a new dataframe
- Used Foursquare API to extract number of venues per city and made new dataframe from that.
- Sorted final dataframe by number of foursquare venues

Data Manipulation tools used:

- Example of Wikipedia dataframe before and after cleaning:

Before:

	2018rank	City	State[c]	2018estimate	2010Census	Change	2016 land area	2016 land area.1	2016 population density	2016 population density.1	Location
0	1	New York[d]	New York	8398748	8175133	+2.74%	301.5 sq mi	780.9 km2	28,317/sq mi	10,933/km2	40°39′49″N 73°56′19″W / 40.6635°N 73.9387°W
1	2	Los Angeles	California	3990456	3792621	+5.22%	468.7 sq mi	1,213.9 km2	8,484/sq mi	3,276/km2	34°01′10″N 118°24′39″W / 34.0194°N 118.4108°W

After:

	rank	City	State	2018 Population	Land Area (square kms)	Population Density (per square km)	Longitude	Latitude
0	1	New York	New York	8398748	780.9	10933	-74.0060	40.7128
1	2	Los Angeles	California	3990456	1213.9	3276	-118.4108	34.0194

Results:

- Here is a portion of the desired results table showing the best cities to start providing the new location recommendation app using foursquare location data:

	City	State	Number of Venues on Foursquare
0	Baltimore	Maryland	122
1	Seattle	Washington	122
2	Nashville	Tennessee	122
3	Long Beach	California	121
4	Jacksonville	Florida	120
5	Kansas City	Missouri	120
6	Albuquerque	New Mexico	120
7	San Diego	California	120
8	Portland	Oregon	119
9	Arlington	Texas	118
10	New York	New York	117
11	Memphis	Tennessee	116
12	Sacramento	California	114
13	San Francisco	California	114
14	Tampa	Florida	114
15	Tulsa	Oklahoma	112
16	Tucson	Arizona	112
17	Fort Worth	Texas	111
18	Dallas	Texas	110
19	Colorado Springs	Colorado	105
20	Louisville	Kentucky	104
21	Los Angeles	California	104
22	Atlanta	Georgia	102
23	Philadelphia	Pennsylvania	101

Discussion:

- Many unexpected cities ranked high on the list of venues per city (e.g. Nashville Tennessee was tied for number one with 122 venues)
- Location data may not be accurate: When San Francisco and New York City coordinates from Wikipedia table were changed to their accurate coordiantes found on a google search, the number of returned venues on those cities changed dramatically. San Franciso changed from 4 to 114, and New York City changed from 88 to 117. This shows that the location data from Wikipedia table is slightly inaccurate.

- If I had more time, I would choose a new data source for longitude and latitude data of cities.

Conclusion:

This analysis showed which highly populated cities in the United States would achieve the best results from an app recommending locations using the foursquare location database