

Relationship between United States Metropolitan GDP and Population Venue Preference

Fanqin Zhou 10/11/2019

1. Introduction

Money can't buy happiness, but would money change the definition of happiness? Wealth affects human behavior and preference, as learned from Psychology 101 in college. Studies shown that money alters how you value your time and effort. In fact, average GDP is a measurement of the wealthiness of a given city. Does wealth have anything to do with people's value? How about ethics? Does lack of wealth or GDP of a city effect addiction? Alcohol consumption?

2. Data acquisition and cleaning

From Wikipedia (https://en.wikipedia.org/wiki/List_of_cities_by_GDP). We obtained data for metropolitan GDP around the worl. In this study, we are only focusing on the United States.

	City/Metropolitan area	Country/Region	UNSDsub-region[1]	Official est.NominalGDP (\$BN)	Institution[2]2014 est.PPP-adjustedGDP (\$BN)	est.PPP-adjustedGDP (\$BN)	McKinsey[4]2010 est.NominalGDP (\$BN)	Other est.NominalGDP (\$BN)
0	Aachen-Liège-Maastricht	Germany Belgium Netherlands	Western Europe	NaN	99.7	NaN	NaN	NaN
1	Aberdeen	United Kingdom	Northern Europe	11.3 (2008)[5]	NaN	NaN	NaN	NaN
2	Abidjan	Ivory Coast	Africa	NaN	NaN	13	NaN	NaN
3	Abu Dhabi	United Arab Emirates	Western Asia	119[6]	178.3	NaN	67.1	NaN
4	Addis Ababa	Ethiopia	Africa	NaN	NaN	12	NaN	NaN

We will be focusing on GDP only. We will drop the irrelevant columns and clean up the data. Worcester does not have a GDP value in this table, thus we will ignore this city.

	City	Country	GDP
0	Akron	United States	37.30
1	Albany	United States	52.30
2	Albuquerque	United States	43.20
3	Allentown	United States	42.70
4	Atlanta	United States	385.50

We will be using foursquare location data to obtain geographic coordinates for each city:

The geographical coordinate of Akron are 41.083064, -81.518485.
The geographical coordinate of Albany are 42.6511674, -73.754968.
The geographical coordinate of Albuquerque are 35.0841034, -106.6509851.
The geographical coordinate of Allentown are 40.6022059, -75.4712794.
The geographical coordinate of Atlanta are 33.7490987, -84.3901849.
The geographical coordinate of Austin are 30.2711286, -97.7436995.
The geographical coordinate of Bakersfield are 35.3738712, -119.0194639.
The geographical coordinate of Baltimore are 39.2908816, -76.610759.
The geographical coordinate of Baton Rouge are 30.4459596, -91.18738.

	City	Country	GDP	Latitude	Longitude
0	Akron	United States	37.30	41.083064	-81.518485
1	Albany	United States	52.30	42.651167	-73.754968
2	Albuquerque	United States	43.20	35.084103	-106.650985
3	Allentown	United States	42.70	40.602206	-75.471279
4	Atlanta	United States	385.50	33.749099	-84.390185
5	Austin	United States	148.70	30.271129	-97.743700

Now, we have obtained our desired table for this study.

3. Popular Venue Generation

Using foursquare location data, we can get venue info around a certain latitude and longitude, with a radius of 500, we will find the top 30 most popular venues in a city:

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Akron	41.083064	-81.518485	Akron Art Museum	41.084153	-81.515494	Art Museum
1	Akron	41.083064	-81.518485	Akron Civic Theatre	41.080620	-81.520528	Performing Arts Venue
2	Akron	41.083064	-81.518485	Lock 3	41.079778	-81.520552	Music Venue
3	Akron	41.083064	-81.518485	The Lockview	41.079788	-81.519461	Bar
4	Akron	41.083064	-81.518485	Peanut Shoppe	41.079897	-81.519653	Food & Drink Shop

We analyze each city on their most popular venues:

	City	Accessories Store	Advertising Agency	Agriturismo	American Restaurant	Amphitheater	Antique Shop	Aquarium	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Musi
0	Akron	0.000000	0.000000	0.000000	0.066667	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.066667	0.033333
1	Albany	0.000000	0.000000	0.000000	0.066667	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.033333	0.000000
2	Albuquerque	0.000000	0.000000	0.000000	0.066667	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	Allentown	0.000000	0.000000	0.000000	0.066667	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.033333
4	Atlanta	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.033333	0.000000	0.033333	0.000000	0.000000	0.000000
5	Austin	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

From this, we generate the top 10 venues for the city:

----Akron----

	venue	freq
0	Art Gallery	0.07
1	American Restaurant	0.07
2	Music Venue	0.07
3	Italian Restaurant	0.07
4	Coffee Shop	0.07
5	Performing Arts Venue	0.07
6	Fast Food Restaurant	0.03
7	Boutique	0.03
8	Food & Drink Shop	0.03
9	Speakeasy	0.03

----Allentown----

	venue	freq
0	Brewery	0.07
1	Farmers Market	0.07
2	American Restaurant	0.07
3	New American Restaurant	0.07
4	Gastropub	0.03
5	Theater	0.03
6	Mexican Restaurant	0.03
7	Park	0.03
8	Middle Eastern Restaurant	0.03
9	Clothing Store	0.03

----Albany----

	venue	freq
0	Café	0.17
1	Pub	0.13
2	Theater	0.07
3	Hotel	0.07
4	American Restaurant	0.07
5	Park	0.07
6	Plaza	0.03
7	Nature Preserve	0.03
8	English Restaurant	0.03
9	Sushi Restaurant	0.03

----Atlanta----

	venue	freq
0	Beer Bar	0.03
1	Korean Restaurant	0.03
2	Latin American Restaurant	0.03
3	Museum	0.03
4	Seafood Restaurant	0.03
5	Outdoor Sculpture	0.03
6	Greek Restaurant	0.03
7	Farmers Market	0.03
8	General Entertainment	0.03
9	Mediterranean Restaurant	0.03

----Albuquerque----

	venue	freq
0	Hotel	0.07
1	Bar	0.07
2	American Restaurant	0.07
3	Mexican Restaurant	0.07
4	Brewery	0.07
5	Café	0.07
6	Coffee Shop	0.07
7	Zoo	0.03
8	Brazilian Restaurant	0.03
9	Burger Joint	0.03

----Austin----

	venue	freq
0	Hotel	0.13
1	Movie Theater	0.07
2	Seafood Restaurant	0.07
3	Coffee Shop	0.07
4	Cocktail Bar	0.07
5	Capitol Building	0.07
6	Food Court	0.03
7	Grocery Store	0.03
8	Lounge	0.03
9	Nightclub	0.03

3.1 Categorization

Notice we really have a lot of too detailed venue type, such as: “fast food restaurant”, “American Restaurant”, “New American restaurant”, etc. It is necessary to group them into more inclusive categories.

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Akron	Performing Arts Venue	American Restaurant	Art Gallery	Music Venue	Italian Restaurant
1	Albany	Café	Pub	Park	Theater	American Restaurant
2	Albuquerque	Bar	Coffee Shop	Hotel	American Restaurant	Brewery
3	Allentown	Farmers Market	American Restaurant	Brewery	New American Restaurant	Gym / Fitness Center
4	Atlanta	Park	Music Venue	Poke Place	Basketball Stadium	Beer Bar
5	Austin	Hotel	Capitol Building	Movie Theater	Coffee Shop	Cocktail Bar
6	Bakersfield	Mexican Restaurant	Coffee Shop	Thai Restaurant	Bar	Ice Cream Shop

In this study, we are mostly interested in the GDP effect on 3 big categories: “Monument/Landmark”, “Food & Beverage” and “Entertainment”.

```
df4[col[i]].replace('Burger Joint', "Restaurant", inplace=True)
df4[col[i]].replace('Deli / Bodega', "Restaurant", inplace=True)
df4[col[i]].replace('Mexican Restaurant', "Restaurant", inplace=True)
df4[col[i]].replace('Japanese Restaurant', "Restaurant", inplace=True)
df4[col[i]].replace('American Restaurant', "Restaurant", inplace=True)
df4[col[i]].replace('French Restaurant', "Restaurant", inplace=True)
```

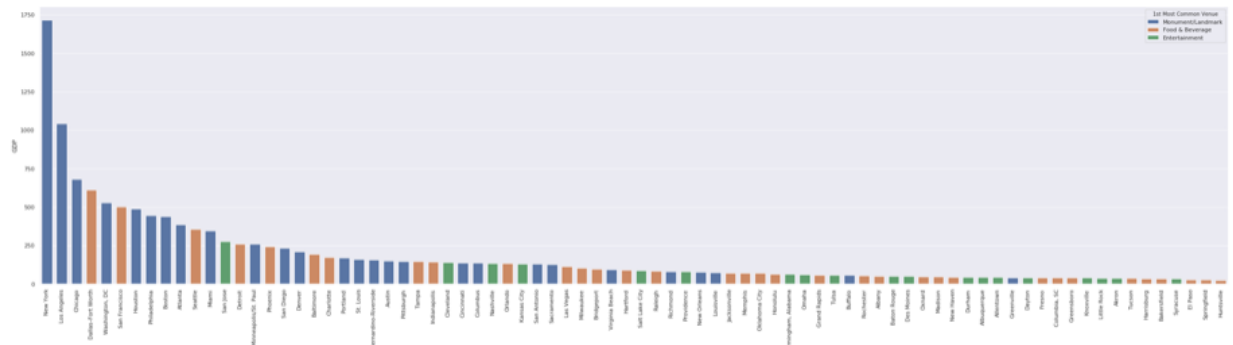
After transformation, our table looks like this:

	City	GDP	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	New York	1717.71	Monument/Landmark	Monument/Landmark	Monument/Landmark	Monument/Landmark	Health/Sport
1	Los Angeles	1043.73	Monument/Landmark	Food & Beverage	Entertainment	Food & Beverage	Monument/Landmark
2	Chicago	679.60	Monument/Landmark	Entertainment	Food & Beverage	Monument/Landmark	Monument/Landmark
3	Dallas–Fort Worth	613.00	Food & Beverage	Entertainment	Entertainment	Food & Beverage	Food & Beverage
4	Washington, DC	529.90	Monument/Landmark	Entertainment	Monument/Landmark	Food & Beverage	Monument/Landmark
5	San Francisco	500.70	Food & Beverage	Health/Sport	Entertainment	Food & Beverage	Entertainment
6	Houston	490.00	Monument/Landmark	Entertainment	Entertainment	Entertainment	Food & Beverage
7	Philadelphia	444.90	Monument/Landmark	Food & Beverage	Monument/Landmark	Food & Beverage	Food & Beverage
8	Boston	438.60	Monument/Landmark	Monument/Landmark	Food & Beverage	Food & Beverage	Entertainment
9	Atlanta	385.50	Monument/Landmark	Entertainment	Food & Beverage	Health/Sport	Entertainment

4. Finding

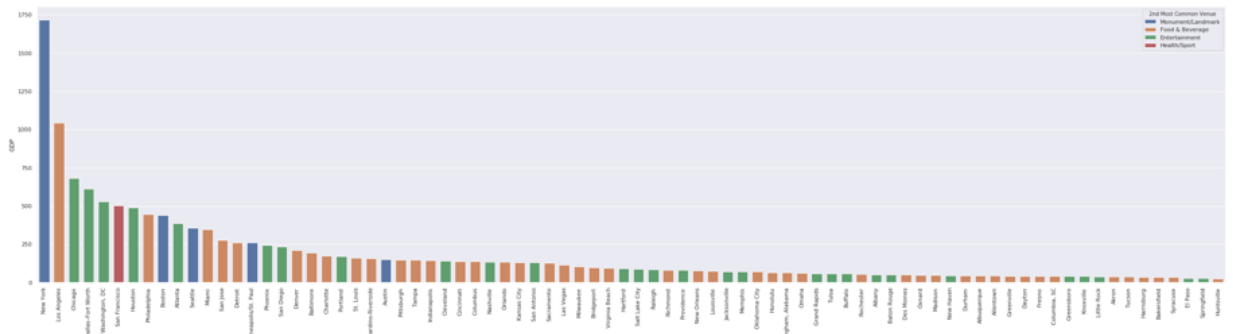
By looking at cities from higher GDP to lower GDP, we see that "Food & Beverage" spread evenly throughout the GPD span, "Monument/Landmark" seems to be most popular among high GPD cities,

"Entertainment" seems to be most popular among lower GDP cities.



In graph, blue indicates "Monument/landmark", orange denotes "Food & Beverage", green denotes "Entertainment".

let's see if the same holds true for 2nd popular venues



In graph, blue indicates "Monument/landmark", orange denotes "Food & Beverage", green denotes "Entertainment".

while Monument/Landmark lost popularity, same rule still holds true. However, Entertainment seems to be distributed evenly amongst cities.

5. Conclusion

From data, we can conclude that "Monument/Landmark" is the most popular among high GDP cities while "Entertainment" is the most popular among lower GDP cities. While this does not indicate any casual relationship, we can see that higher GDP does have an effect on people's preference. It is also interesting to see that GDP does not influence "Food & Beverage" a lot.