

Performing Arts Theater Hotspots around the United States

Introduction/Business Problem

In this report, we will look at Performing Arts Theaters around the United States, to identify communities that have a thriving performing arts theater community, and locations that have a relatively deficient presence. The generated report will be of interest to those considering relocating to areas with thriving communities, such as actors, managers, designers, patrons or investors, or for those looking to establish theater in areas that are currently underrepresented.

For example, cities with either a high, or a higher than predicted, number of theaters will be considered hotspots, and those with less will be considered “growth opportunities.” Additional insights may be gathered while analyzing the data.

Data Sources

We will get data for all large cities in the United States, getting data from:

1) https://en.wikipedia.org/wiki/List_of_United_States_cities_by_population

Lists cities with populations over 100,000, growth rate (percent growth from 2010 to 2018), and population density (from 2016) available from the same source. The data looks like this:

R a n k	City	State	2018 estima te	2010 Cens us	Cha nge	2016 land area		2016 population density		Locat ion
1	New York	New York	8,398,748	8,175,133	+2.74%	301.5 sq mi	780.9 k m ²	28,317/s q mi	10,933/ km ²	40.6635° N 73.9387° W
2	Los Angeles	California	3,990,456	3,792,621	+5.22%	468.7 sq mi	1,213.9 km ²	8,484/sq mi	3,276/k m ²	34.0194° N 118.4108° W
3	Chicago	Illinois	2,705,994	2,695,598	+0.39%	227.3 sq mi	588.7 k m ²	11,900/s q mi	4,600/k m ²	41.8376° N 87.6818° W

2) Foursquare (<https://foursquare.com/>) will be used to identify the number of performing arts theaters within 10km of each city center. From foursquare, we will use, under “categories”, the performing arts id, {'id': '4bf58dd8d48988d1e5931735'}, and then filter to include {'name': 'Theater'} but not other names, such as {'name': 'Music Venue'} or {'name': 'Concert Hall'} which show up under the same id.

The raw (json) data returned from foursquare looks like this:

```
{'meta':
  {'code': 200, 'requestId': '5e8e6135618f43001bebae82'},
'response': {'venues': [
  {'id': '4a0b05a7f964a520bb741fe3',
  'name': 'The Walter Kerr Theatre',
  'location': {'address': '219 W 48th St',
  'crossStreet': 'btwn Broadway & 8th Ave',
  'lat': 40.76061320906541, 'lng': -73.98572244899813,
  'labeledLatLngs': [{'label': 'display',
    'lat': 40.76061320906541,
    'lng': -73.98572244899813},
    {'label': '?', 'lat': 40.760562, 'lng': -73.985672}],
  'distance': 327,
  'postalCode': '10036',
  'cc': 'US',
  'city': 'New York',
  'state': 'NY',
  'country': 'United States',
  'formattedAddress': ['219 W 48th St (btwn Broadway & 8th Ave)',
    'New York, NY 10036',
    'United States']},
  'categories': [{'id': '4bf58dd8d48988d137941735',
    'name': 'Theater',
    'pluralName': 'Theaters',
    'shortName': 'Theater',
    'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/arts_entertainm
ent/performingarts_theater_',
    'suffix': '.png'},
    'primary': True}],
  'referralId': 'v-1586389276',
  'hasPerk': False}]}
```

3) <http://www.univsearch.com/family-search.php?family=50>
lists the colleges and universities in the United States with Visual and Performing Arts Schools and Department, with their city and state. An example is here:

COLLEGE/ UNIVERSITY NAME	CITY	STATE
A1 BUSINESS AND TECHNICAL COLLEGE	CAGUAS	PR
ABILENE CHRISTIAN UNIVERSITY	ABILENE	TX
ACADEMIA SERRANT INC	PONCE	PR
ACADEMY EDUCATION CENTER INC	MINNEAPOLIS	MN
ACADEMY OF ART COLLEGE	SAN FRANCISCO	CA

Each type of data will be appropriately manipulated to generate a normal or uniform histogram, and then normalized to have zero mean and unit standard deviation. A linear regression model will be used to predict the number of performing arts theaters based on the city population, the number of local colleges, and other available demographic data.