

Data Sources and Description

The data used for the analysis was sourced through www.simplemaps.com. This website provided updated data (Jan 2021) on latitude and longitude for over 108,000 cities and towns in the US for all 50 states. The geodata was used in conjunction with the folium library to visualize the geographic locations of the counties. The data was read into the IDE via a CSV file.

Another website used was <https://data.census.gov/cedsci/>. This is the official website of the US Census Bureau which provided the population, population density and average income per household for the counties in Florida. The data was used to analyze each county using their attributes to identify the most suitable county for a startup. The data was read into IDE via a CSV file.

Foursquare is a social media mobile application used to discover various types of businesses in a geographic location. The application provided the trending venues by category for the cluster analysis model.

All of the datasets used in this project were checked for null/NAN values and duplicates. Columns with data not relevant to the scope of this project were removed (State Id, time zone, zip code, ranking, military, etc) The Foursquare category data was grouped by county and one hot encoded to convert the categorical data. Table 3.1 is a data dictionary for all of the datasets used in this project. Table 4.1 is the table created from www.simplemaps.com and <https://data.census.gov/cedsci/>

Table 3.1 – Data Dictionary

Term	Data Type	Description
County_name	Object	Name of the county
lat	Float64	Latitude of the object
lng	Float64	Longitude of the object
average income	Int64	Average income for each household in the county
population	Int64	Population of each county
Population density	Float64	Population density of each county
County	Object	Name of the county
County Latitude	Float64	Latitude of the county
County Longitude	Float64	Longitude of the county
Venue	String	Name of the venue
Venue latitude	Float64	Latitude of the venue
Venue Longitude	Float64	Longitude of the venue
Venue Category	String	Classification of the venue
Most Common Venue	String	10 most trending venues
Cluster Labels	Int64	The number of cluster in each county

Table 4.1 – County Data

	county_name	lat	lng	average income	population	population density
0	Broward	26.1412	-80.1464	59547	1951866	2271.75
1	Collier	29.9807	-81.8108	69653	140250	1143.53
2	Duval	29.4499	-83.2818	55832	1227744	1183.75
3	Hillsborough	27.5908	-81.5081	48452	3855386	1814.40
4	Lee	28.9681	-81.6482	57832	1528309	807.05

Foursquare Data

Foursquare is a local app which provides recommendations given a user's current location and interests. The Foursquare application was used to locate and identify trending venues in the 10 counties (Table 4.2) Foursquare provided the latitude and longitude, venue name and category for the top 196 trending venues for each state. The venues were grouped by counties and the top ten venues were listed for each county. Table (4.3)

Table 4.2 - Foursquare Results by County

	County	County Latitude	County Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Broward	26.1412	-80.1464	LeatherWerks	26.141577	-80.140398	Men's Store
1	Broward	26.1412	-80.1464	Wine Watch	26.136892	-80.140659	Wine Shop
2	Broward	26.1412	-80.1464	Laser Wolf	26.134695	-80.141203	Bar
3	Broward	26.1412	-80.1464	Hot Dog Heaven	26.136929	-80.143318	Hot Dog Joint
4	Broward	26.1412	-80.1464	Radio Active Records	26.134800	-80.137900	Record Shop

```
: florida_venues['Venue Category'].value_counts()
```

```
: Fast Food Restaurant      11
: Sandwich Place            9
: Grocery Store             8
: American Restaurant       7
: Pharmacy                  6
: Gas Station               6
: Hotel                     5
: Gay Bar                   5
: Coffee Shop               5
: Discount Store            5
: Mexican Restaurant        4
: Donut Shop                4
: Park                      4
: Pizza Place               4
: Convenience Store         4
: Seafood Restaurant        4
: Thrift / Vintage Store    4
```

Table 4.3 County Top 10 Venue by County

	County	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Broward	Gay Bar	Coffee Shop	Brewery	Gym	Sandwich Place	Fast Food Restaurant	New American Restaurant	Park	Thrift / Vintage Store	Clothing Store
1	Collier	Post Office	Wings Joint	Fast Food Restaurant	Cosmetics Shop	Cuban Restaurant	Dentist's Office	Diner	Discount Store	Dog Run	Donut Shop
2	Duval	Beach	Wings Joint	Cuban Restaurant	Dentist's Office	Diner	Discount Store	Dog Run	Donut Shop	Farm	Fast Food Restaurant
3	Hillsborough	Fast Food Restaurant	Pizza Place	Donut Shop	Sandwich Place	Chinese Restaurant	Discount Store	Pharmacy	American Restaurant	Grocery Store	Hotel
4	Lee	Convenience Store	Basketball Court	Gas Station	Farm	Lake	French Restaurant	Cuban Restaurant	Dentist's Office	Diner	Discount Store
5	Miami Dade	American Restaurant	Seafood Restaurant	Business Service	Italian Restaurant	Grocery Store	Pharmacy	Donut Shop	Gas Station	Home Service	Breakfast Spot
6	Monroe	Hotel	Grocery Store	Pet Service	Dog Run	Park	Café	French Restaurant	Steakhouse	Italian Restaurant	Jewelry Store