

Recommending a Data Science Business in South Florida

Alfredo Castañeda

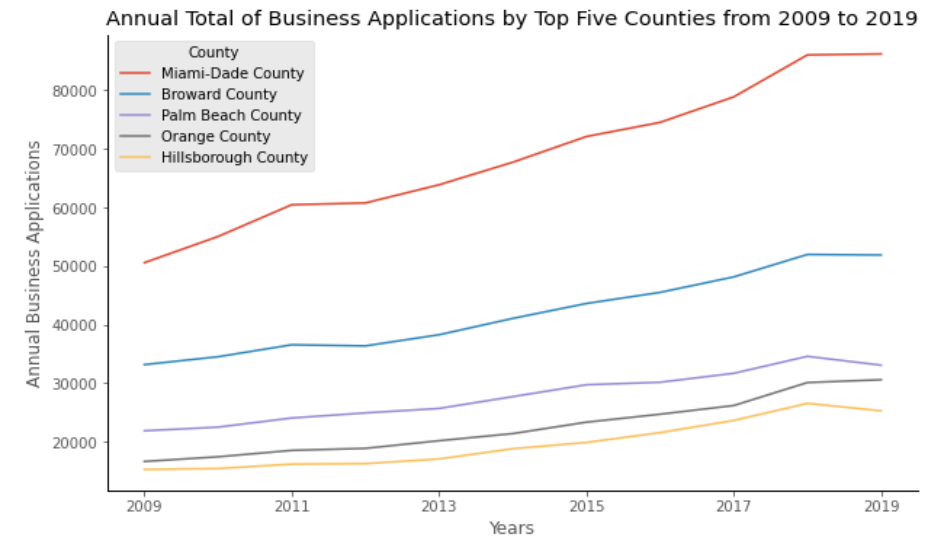
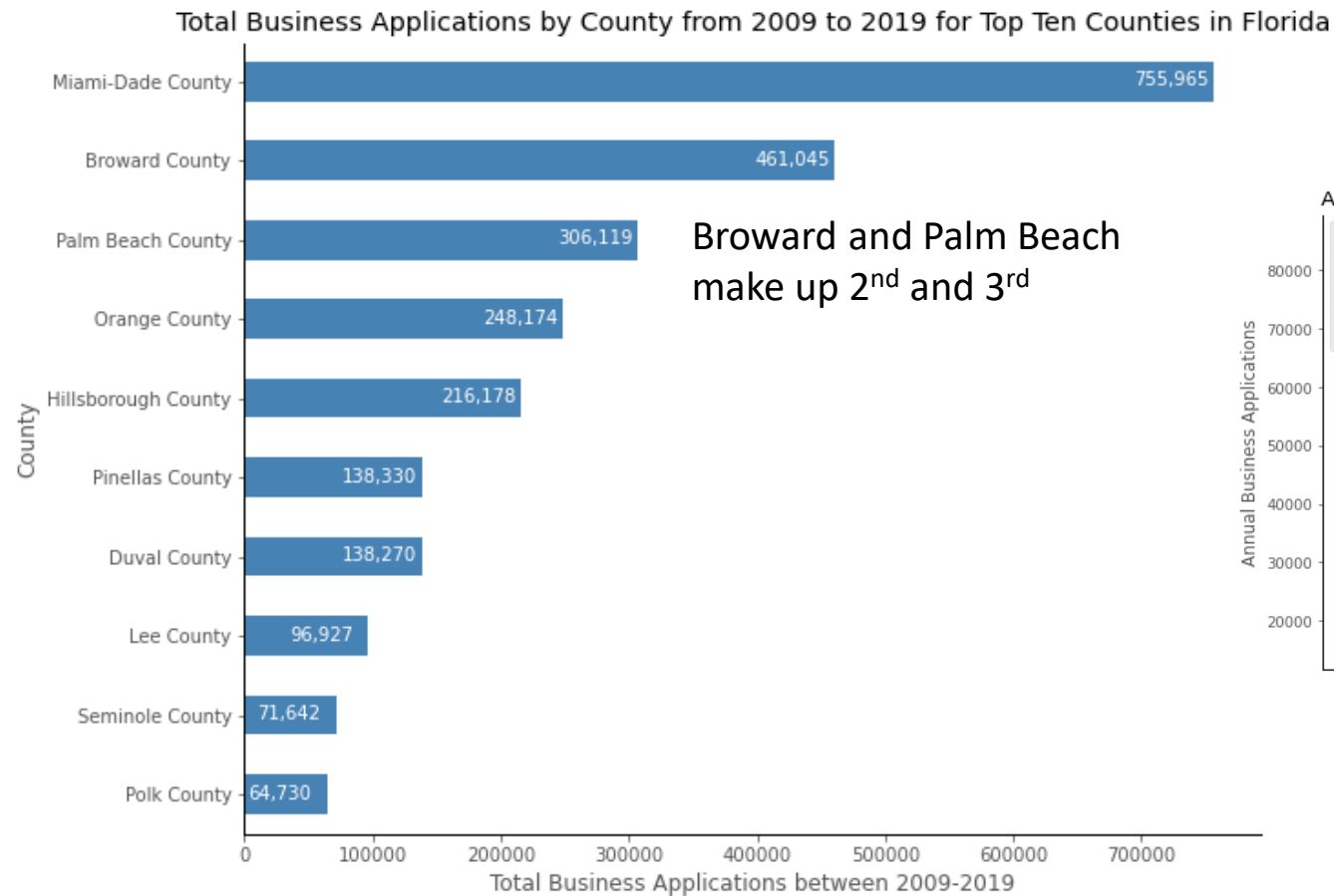
Why Florida?

- 4th largest economy in the U.S. and 18th largest economy in the world (if Florida were a nation). (BEA & IMF).
- Ranked 4th in the nation for high-tech employment by CompTIA, Florida boasts nearly 237,000 high-tech workers.
- Number 3 for high-tech establishments (TechAmerica).

Source: Enterprise Florida website retrieved from <https://www.enterpriseflorida.com/why-florida/>

Where in Florida?

Palm Beach is most favored
due to less population and
less potential competition



Broward and Palm Beach
show ten-year consistent
growth (2009-2019).

Business Problem

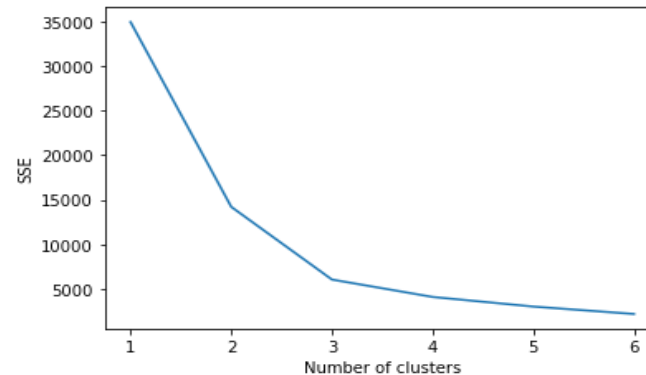
- Objective of the project is to analyze the businesses located within the cities of both Broward and Palm Beach counties.
- The focus will be on businesses within the four key industry areas: Information Technology, Life Sciences, Financial and Clean Tech.
- The benefit of this study is to produce a strategy tailored to service the four key industries from a location in either County.
- The target audience includes people interested in practicing data science or providing technology products to the selected industries across both counties.

Methodology

- **Building the data set** - Extracted data from DB Hoovers Data based of 120 million business records. Based on select criteria.
- **Mapped Categories** – Using Key Industry Briefs mapped DB Hoover categories to Key Industries.
- **Preprocessing** – Added Geo coordinates with Geocoder, Cleaned data, removed missing data
- **Transforming** – Used dummy variables to match industry categories for modeling and analysis of clusters

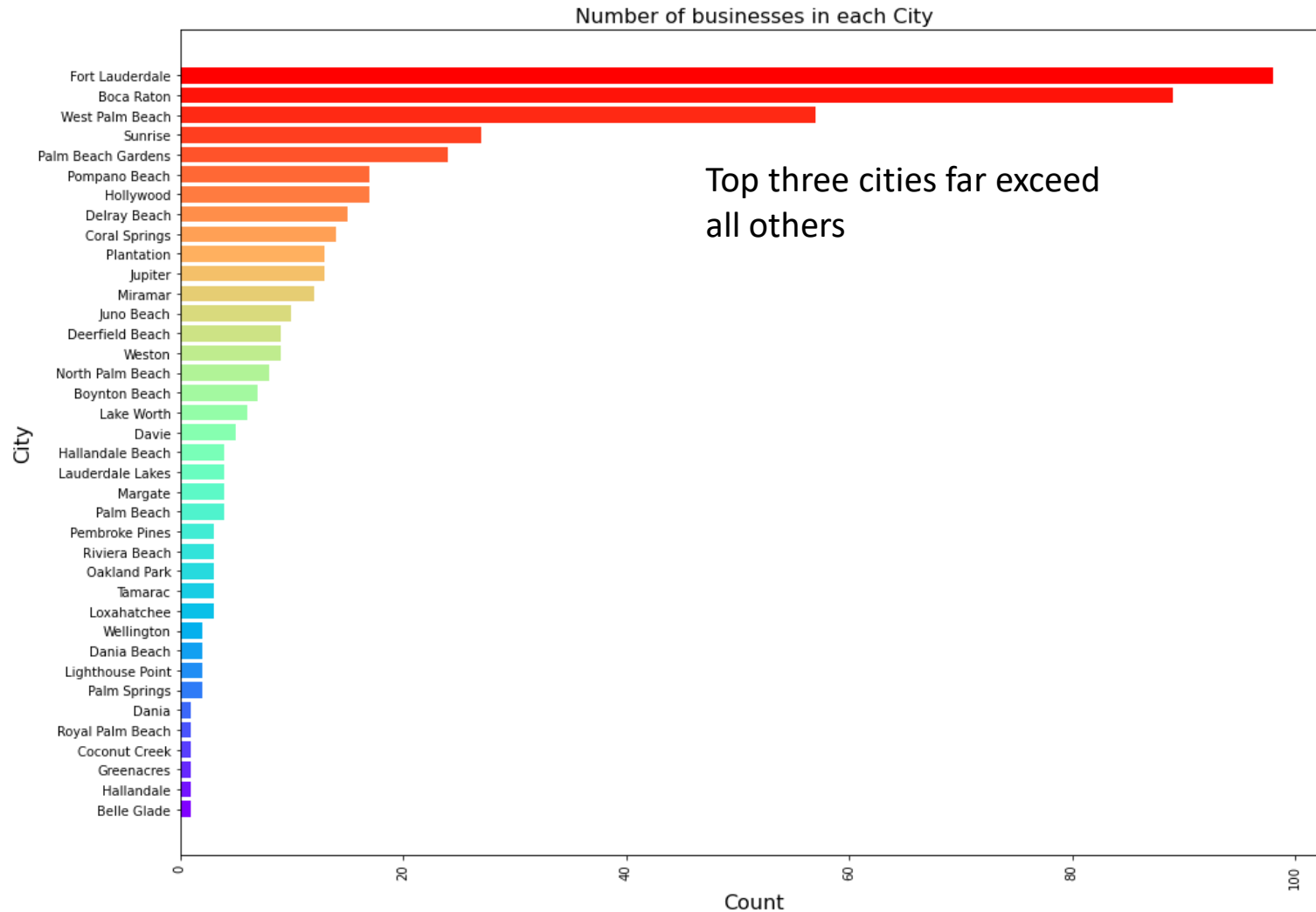
Methodology (cont.)

- **Mining / Visualization** – Through data exploration explored data set and selected attributes for modeling.
- **Wrangling** – Manipulated data to aggregate information to facilitate clustering leveraging exploratory findings
- **Analysis** – Apply k-means clustering to partition the data into clusters and explore clusters.

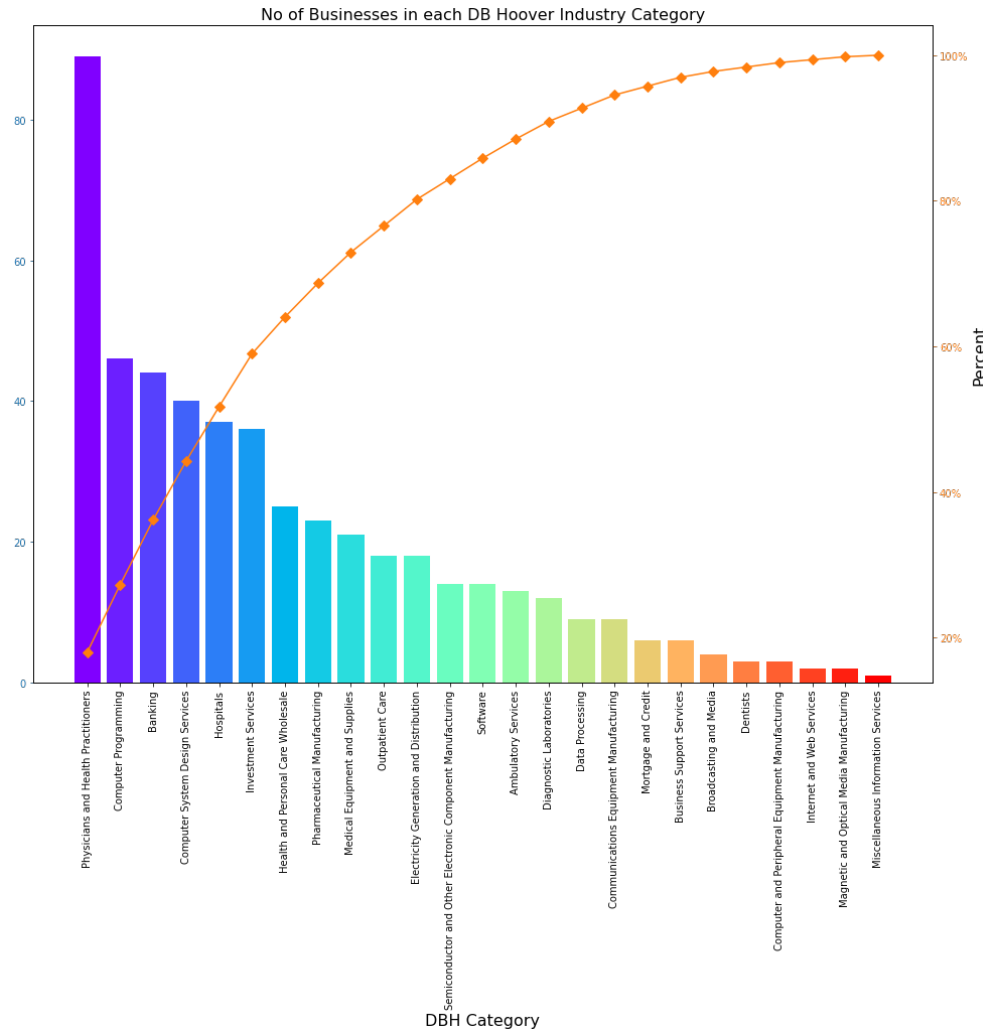


Optimal number of K-means clusters determined from elbow in graph.

A small group of cities outperform the rest



Physicians and Practioners are an exception



DBH Category	Count
Physicians and Health Practitioners	0.179798
Computer Programming	0.092929
Banking	0.088889
Computer System Design Services	0.080808
Hospitals	0.074747
Investment Services	0.072727
Health and Personal Care Wholesale	0.050505
Pharmaceutical Manufacturing	0.046465
Medical Equipment and Supplies	0.042424
Outpatient Care	0.036364


Physicians make
up 18% of total

All others are
more distributed.

Data Wrangling

Data set collected from DB
Hoovers Data base

Name	City	Miles to Ft Laud	Miles to Palm Beach	DB Hoovers Industry	Industry	Ownership Type	Parent Country_Region
Kemet Corporation	Fort Lauderdale	0.04	41.34	Semiconductor and Other Electronic Component Manufacturing	InfoTech	Private	Taiwan Region
Ri/Bbnm Acquisition Corp.	Fort Lauderdale	0.16	41.53	Investment Services	Financial	Private	United States
Advanced Recovery Systems, LLC	Fort Lauderdale	0.22	41.40	Outpatient Care	LifeScience	Private	United States
Templeton International, Inc.	Fort Lauderdale	0.24	41.48	Investment Services	Financial	Private	United States
Templeton/Franklin Investment Services, Inc	Fort Lauderdale	0.24	41.48	Investment Services	Financial	Private	United States



	City	Count	MedD_PB
0	Belle Glade	1	38.13
1	Boca Raton	89	22.61
2	Boynton Beach	7	14.73
3	Coconut Creek	1	29.20

Data frame extracted for
aggregation of City counts and
median distance to PB location

Data frame extracted for seeding
of geo codes for 38 cities

	City	Businesses	Latitude	Longitude
0	Belle Glade	1	26.684510	-80.667558
1	Boca Raton	89	26.368306	-80.128932
2	Boynton Beach	7	26.531787	-80.090547



Added Geo Codes

Data Wrangling (cont.)

Data frames combined for analysis with k-means algorithm

	City	Clean Tech	Financial	InfoTech	LifeScience
0	Belle Glade	0	0	0	1
1	Boca Raton	0	16	34	39
2	Boynton Beach	0	1	0	6
3	Coconut Creek	0	0	0	1
4	Coral Springs	0	3	4	7



Data was aggregated by city

	City	Clean Tech	Financial	InfoTech	LifeScience
0	Fort Lauderdale	0	0	1	0
1	Fort Lauderdale	0	1	0	0
2	Fort Lauderdale	0	0	0	1

Data frame extracted for dummy variables for key industries

	City	Count	MedD_PB	Clean Tech	Financial	InfoTech	LifeScience
0	Belle Glade	1	38.13	0	0	0	1
1	Boca Raton	89	22.61	0	16	34	39
2	Boynton Beach	7	14.73	0	1	0	6
3	Coconut Creek	1	29.20	0	0	0	1
4	Coral Springs	14	33.27	0	3	4	7



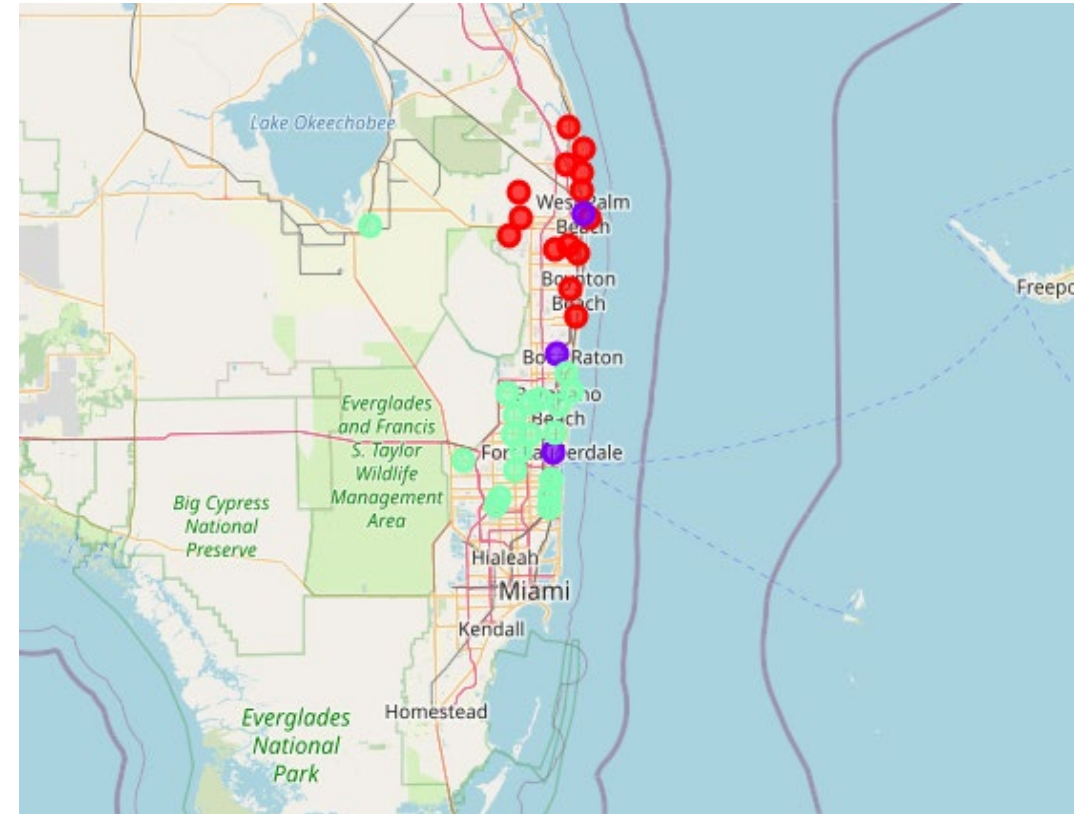
Final Data frame seeded with Cluster labels and merged with Geocode city data

	City	Cluster Labels	Businesses	LifeScience	InfoTech	Financial	Clean Tech	MedD_PB	Latitude	Longitude
0	Belle Glade	2	1	1	0	0	0	38.13	26.684510	-80.667558
1	Boca Raton	1	89	39	34	16	0	22.61	26.368306	-80.128932
2	Boynton Beach	0	7	6	0	1	0	14.73	26.531787	-80.090547
3	Coconut Creek	2	1	1	0	0	0	29.20	26.251748	-80.178935
4	Coral Springs	2	14	7	4	3	0	33.27	26.271192	-80.270604

Results / Discussion

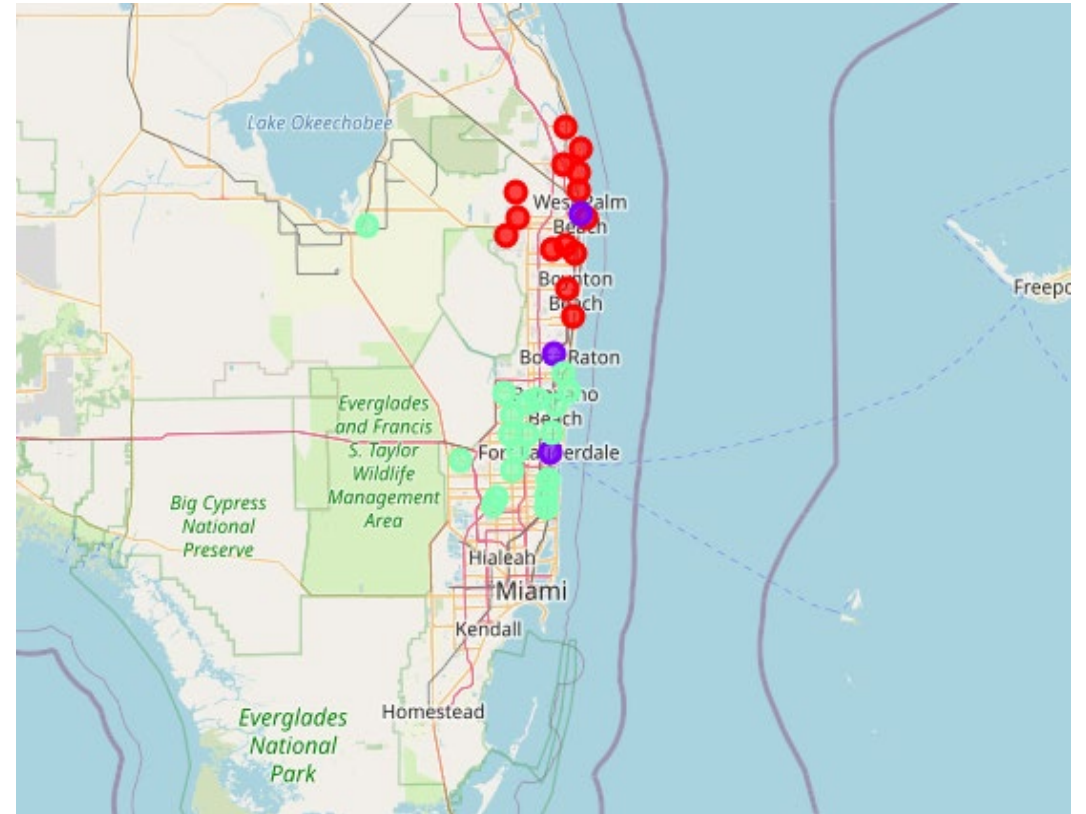
- Colors Red, Green and Purple represent the three clusters North, South and High Density.

Cluster Labels	Total	LifeScience	InfoTech	Financial	CleanTech
North Cluster	99	51	16	19	13
High Density Cluster	244	105	89	45	5
South Cluster	152	85	45	22	0



Results / Discussion (cont.)

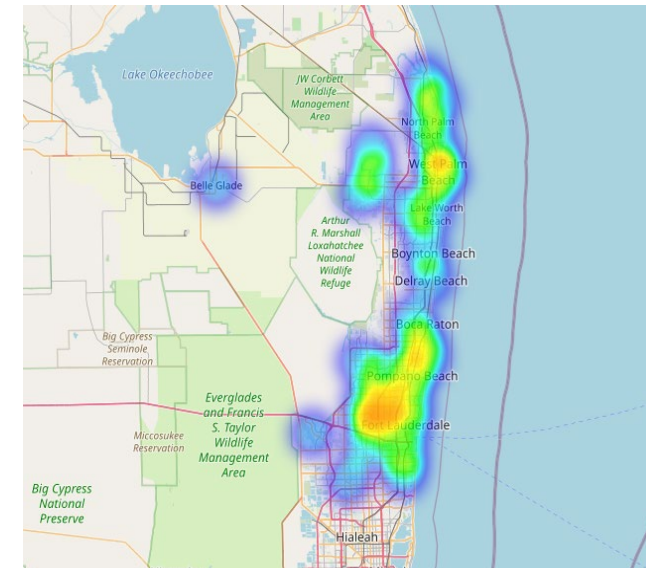
- High Density Cluster is made of three cities yet accounts for largest share in total of business.
- Life Science and Info Tech are most represented Industries. Clean tech is the smallest.
- Life Science is skewed due to the large number of Physicians.
- Palm Beach has two of high-density cities.



Conclusions and Potential Applications

- This study identified 495 businesses with a potential for data science services through potential contracting or consulting.
- The results of this study can inform a prospective entrepreneur to target 38 cities / 495 businesses based on location and/or industry.
- This data can further be analyzed if one desires to target a specific industry or select city.

Cluster Labels	Total	LifeScience	InfoTech	Financial	CleanTech
North Cluster	99	51	16	19	13
High Density Cluster	244	105	89	45	5
South Cluster	152	85	45	22	0



References

- Source US Census Bureau, 2020 November 24, Business Formation Statistics, Retrieved from <https://www.census.gov/econ/bfs/index.html>
- Source Wikipedia, 2020 November 20, Miami metropolitan area, Retrieved from https://en.wikipedia.org/wiki/Miami_metropolitan_area
- Source Enterprise Florida, 2020 November 11, Industries, Retrieved from <https://www.enterpriseflorida.com/industries/>