

DATA SCIENCE CAPSTONE

GARY MOYO

Background & Discussion

- With increasing awareness of the climate change movement on the rise. The push to reduce CO₂ emissions is louder than before. On the business end, the opportunities to gain capital rise as well.
- Many industries will argue that a sudden requirement to adapt will have great bearings on the economy.
- Bloomberg reports that the global economy must invest \$1.8 trillion by 2030 to prepare for the effects of global warming. Bloomberg also reports that fighting climate change will aide economic growth. A large part of this argument being the skyrocketing costs spent on natural disasters in the last 30 years.
- On September 20, 2019 the German government announced an agreement to set a price on carbon emissions in a bid to meet a 2030 climate target of cutting greenhouse gases by 55% on 1990 levels.
 - *The plan will boost initiatives on “going green”.*
 - *VAT (sales tax) on rail tickets is set to fall from 19% to 7% on 1 January 2020 and operator Deutsche Bahn said it would waive any price increase. A further €1bn of annual investment is planned until 2030 to modernize and expand the rail network to cope with an expected increase in passengers.*

Problem & Interest

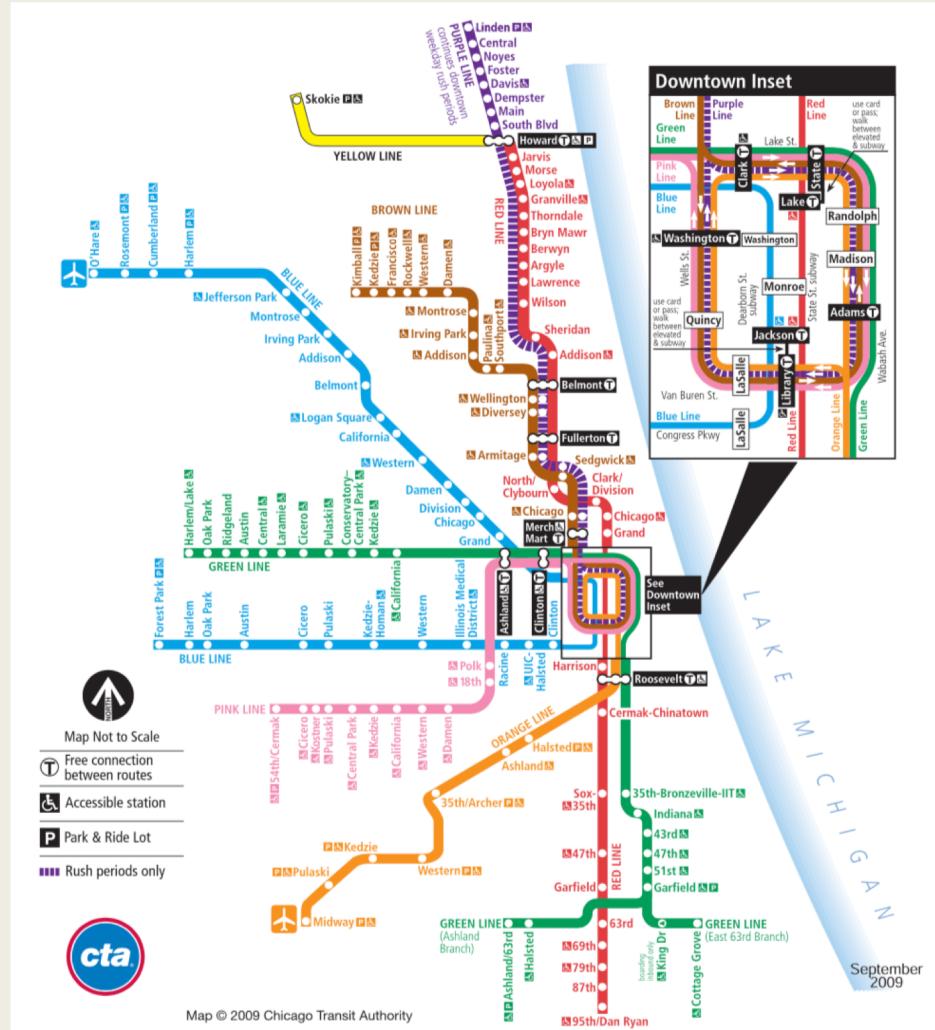
- As current Chicago resident, I believe it's only a matter of time until this coalition becomes Global and we are all required participate by law.
- The problem will be the sudden increase in gas prices due to carbon taxes, ceilings on industry CO₂ emissions and the transferring burden to consumers.
- The goal of this report is to provide current and incoming businesses with most common venue data focused around the Chicago Transit Authority (CTA) train service.



In a sign of what's at stake, a report included a finding showing that annual costs for natural disasters topped the 30-year average of \$140 billion in seven of the past 10 years -- and that the number of extreme weather events since 1980 has tripled. *Photographer: Luke Sharrett/Bloomberg*

Data

- CTA train line (Red, Blue, Brown, Green, Orange, Pink, Purple, Yellow) stops.
- Data downloaded will be cleaned and simplified so that each point includes the stops latitude/longitude coordinates and the respective train line.
- Foursquare API to explore the most common venues near all train lines. This data will help us group the transit stops into clusters and analyze each.



Methodology

- We will use the k-means clustering algorithm. K-means clustering is a simple but popular unsupervised machine learning algorithm.
- Why? K-means clustering works exceptional well with numerical data.
- Behavioral segmentation data by grouping and creating profiles based on common venues.
- It will compare and find certain similarities between each cluster.
- Give business insights into Chicago transit stop areas.

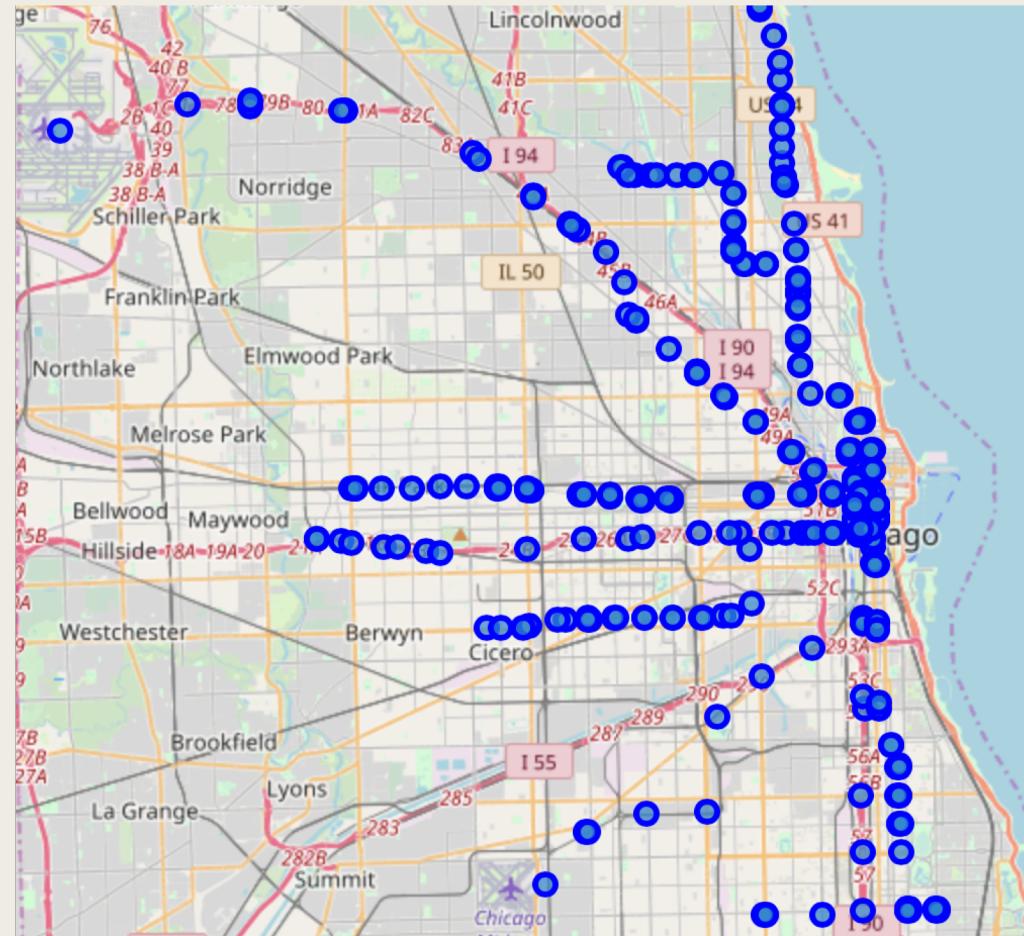
Analysis

- Input vectors will be CTA transit stop coordinates and the Foursquare most common venues dataset.
- Geopy library used to get the latitude and longitude values of Chicago.
- Folium used for data visualization.
- Explored each Transit Stop in dataframe.
- Explored nearby venues in a 500 meter radius.

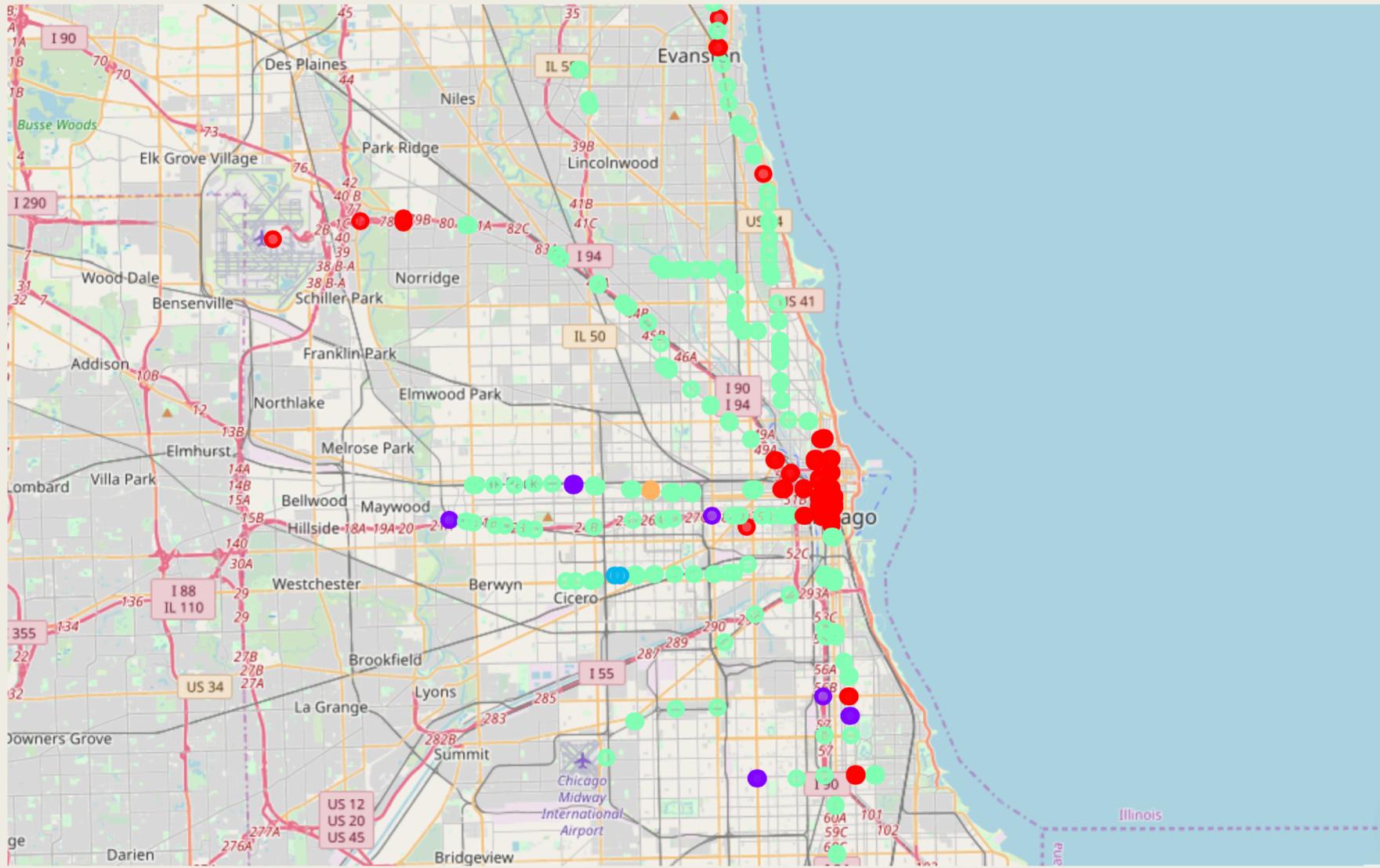
	Transit Stop	Latitude	Longitude
0	18th	41.857849	-87.669144
1	35th/Archer	41.829274	-87.680632
2	95th-Dan Ryan	41.722729	-87.624410
3	Adams/Wabash	41.879715	-87.625997
4	Addison	41.946604	-87.718406

Analysis cont'd

- Analyzed each Transit Stop and Top 5 most common venues.
- Ran k-means to cluster transit stops into 5 clusters.
- Created dataframe to store clusters and top 5 venues for each transit stop.
- Examined each cluster to determine the distinguishing characteristic.



Results & Recommendations



Results & Recommendations cont'd

- 365 unique categories.
- 35 venues returned by Foursquare.
- Restaurants and Hotels are the most common venues for transit stops closest to downtown.
- General entertainment, parks and grocery stores are much more common as you get further away from downtown
- Prospective business owners can use this data to differentiate.
- Strategically place services such as bikes, scooters or other transit sharing business.

Conclusion

- Multiple use cases for prospective business.
 - Creating profiles based on activity monitoring.
 - Defining personas based on interests