Data Science Capstone

Introduction

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

This report centers around the last point. 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.

Interest

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.

The value added for business will be data analytics to aide with new venues around The City of Chicago. Considering the increasing foot traffic near transit stops.

Data

Data Source & Description

The Data was sourced from Github. The dataset includes coordinates (latitude and longitude) to entrances/exists of all 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. I will use the 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.

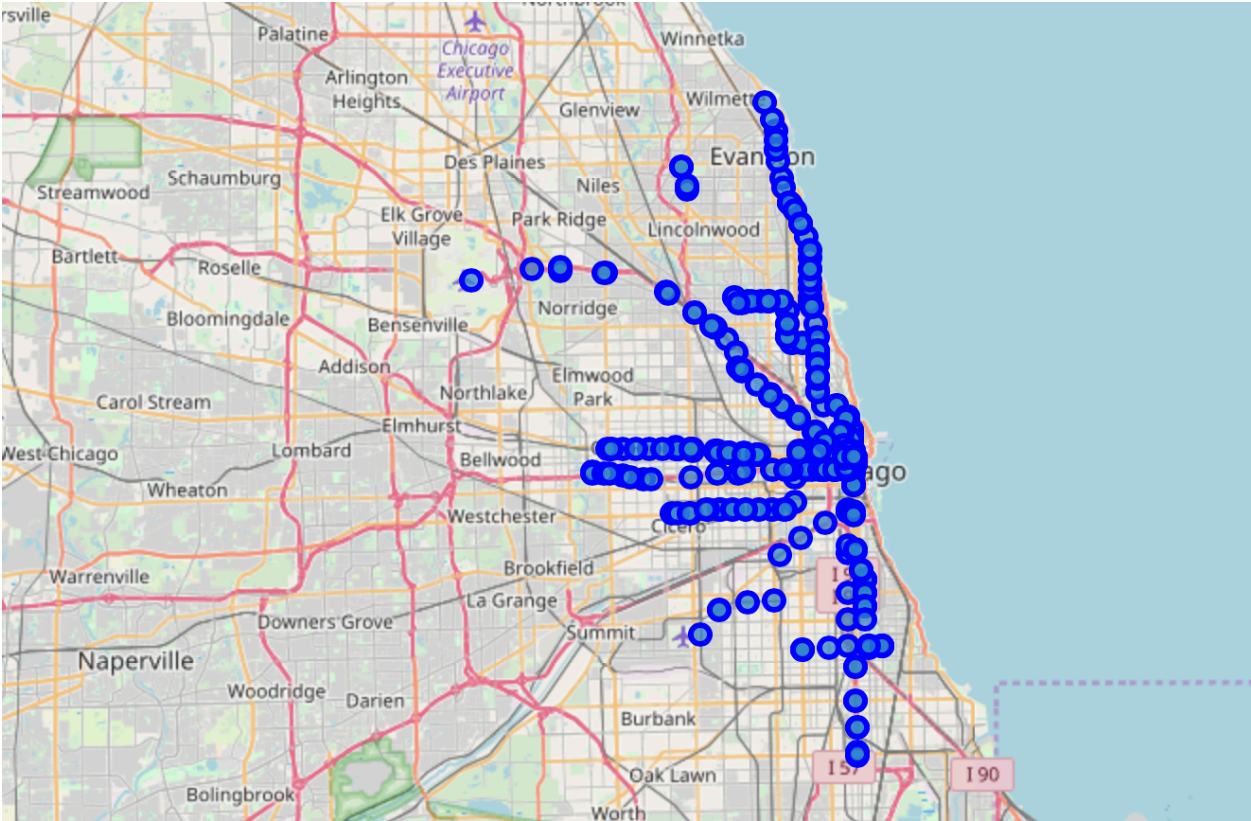|	Transit Stop	Transit Stop Latitude	Transit Stop Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	18th	41.857849	-87.669144	5 Rabanitos	41.857808	-87.670947	Mexican Restaurant
1	18th	41.857849	-87.669144	Carnitas Uruapan	41.857545	-87.669672	Mexican Restaurant
2	18th	41.857849	-87.669144	The Jibarito Stop	41.857789	-87.668101	Latin American Restaurant
3	18th	41.857849	-87.669144	Taqueria El Mezquite	41.857794	-87.670799	Mexican Restaurant
4	18th	41.857849	-87.669144	Paleteria El Sabor De Mexico	41.857776	-87.667811	Ice Cream Shop

Methodology & Analysis

Algorithm

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. Output behavioral segmentation data by grouping and creating profiles based on common venues.

Our input vectors will be the CTA transit stop coordinates and the Foursquare most common venues dataset. This will compare and find certain similarities between each cluster. It will give business insights into Chicago transit stop areas.

Exploratory Data Analysis

The geopy library was used to get the latitude and longitude values of Chicago. Folium was used for data visualization. The dataframe has 109 transit stops.


I looked at the nearby venues in a 500meter radius. Analyzed each Transit Stop and Top 5 most common venues. Explored the amount of venues for each transit stop. Then ran k-means to cluster transit stops into 5 clusters. Created dataframe to store clusters and top 5 venues for each transit stop.

	Transit Stop	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	18th	Mexican Restaurant	Pizza Place	Ice Cream Shop	Convenience Store	Coffee Shop
1	35th-Bronzeville-IIT	Sandwich Place	Bus Stop	Train Station	Shipping Store	Snack Place
2	35th/Archer	Bakery	Thrift / Vintage Store	Supermarket	Diner	Gas Station
3	43rd	Boutique	Harbor / Marina	Sports Bar	Fast Food Restaurant	Smoke Shop
4	47th	American Restaurant	Art Gallery	Plaza	Jazz Club	Breakfast Spot

Results

Observations & Discussion

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.
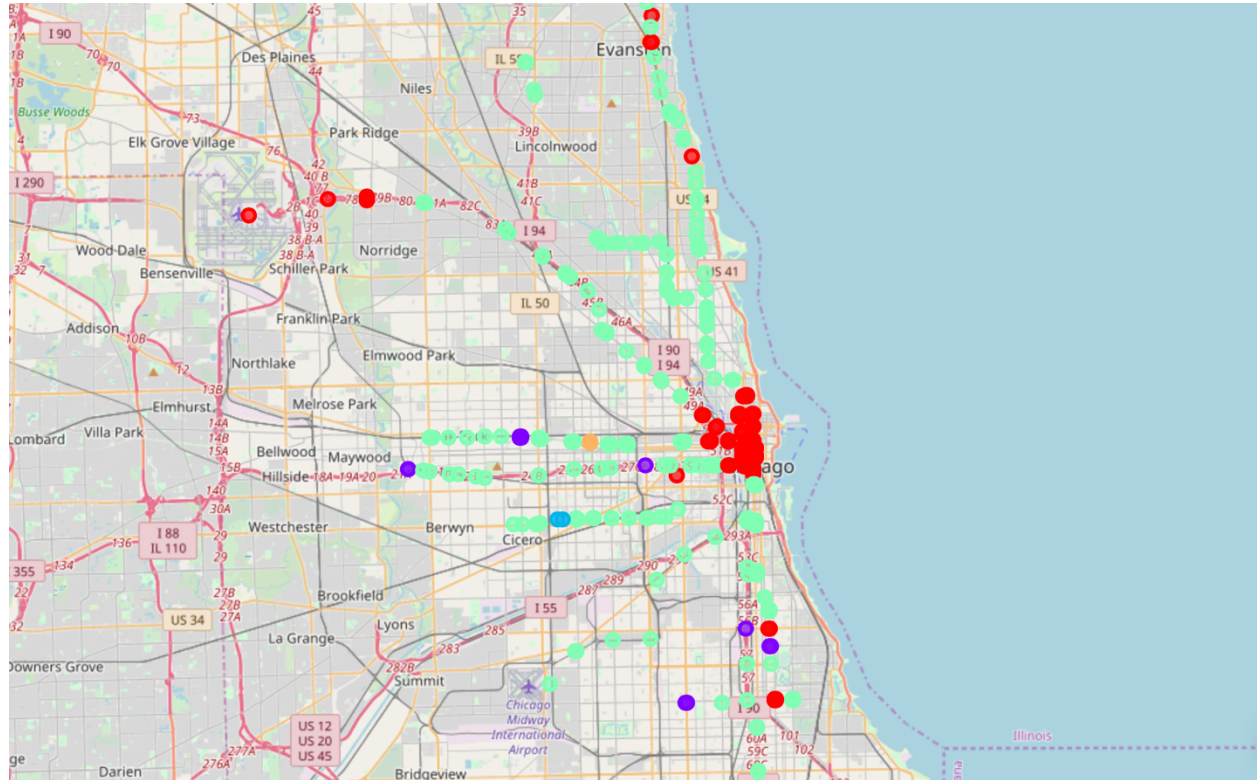
Recommendations

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

Works Cited

https://www.bloomberg.com/news/articles/2019-08-19/fighting-climate-change-will-help-
economic-growth-study-finds

https://www.bloomberg.com/news/articles/2019-09-09/the-massive-cost-of-not-adapting-to-
climate-change