IBM DATASCIENCE PROFESSIONAL CERTIFICATE
CAPSTONE PROJECT
FINAL REPORT
AUTHOR: DMYTRO SHPARYK

CONTENTS:

Introduction / Business Problem

In any big city you can see a lot of advertisement on the streets, like billboards, people giving flyers for advertising an event or product, etc. This advertisement is not designed to be changed during the day. What if we can implement a system where we keep track on the trending venues around a specific location during the day and change advertisement on digital billboards accordingly.

Here is a scenario:
Advertisement agency owns many advertising platforms (like digital billboards, digital ads in public transport, people with flyers, etc) in downtown of a big city (Toronto in our case). This agency has many clients with different products and services that are being advertised on these platforms.

During the day we collect information on the trending venues around specific locations (where digital ads platforms located) and then change ads depending on these trends. For example if we see that during winter morning coffee shops are trending, we can set ads platform to advertise our clients coffee shop, then good restaurant at lunch time and finally theatre premier in the evening. All this depends on trending venues for this time of a day and our observations.

Data

In this project I'll leverage the Foursquare location data to solve our problem. Foursquare allows us to see trending venues. Depending on the time when we run calls to Foursquare API, we might get different venues since the venues with the highest foot traffic are fetched live.

As an example I'm going to use two popular tourist attractions in NewYork and Toronto. Since both cities are at the same time zone, we will also be able to see if trends are different in these two cities at the same time of a day. To make it interesting, I'll pick similar places, tall buildings that people going at to watch nice view - Empire State Building in NewYork and CN Tower in Toronto.

In future, using the same technic, we can collect data from different parts of the day to improve ads for our agency.

Methodology

The main goal of this project was to see the difference in trending venues during different time of the day so we can compare them and decide on which ads are better to show in particular area. I used Nominatim from geopy library to convert interested address (in our case Empire State Building in NewYork and CN Tower in Toronto) into latitude and longitude values. Using Foursquare API I pulled information about venues that are trending at this exact time, meaning the places with the highest foot traffic around our interested places.

To be able to explore trends during different parts of the day, I decided to make checks if we have any trending venues at all, and if we have, save these venues to separate .cvs files, so that later we can just transfer it back to data frame and compare if needed. To track files easier, I added timestamp place name to files, so we have information about exact time and date of when information was collected. Also to visualize our observations and compare, I decided to use Folium library to show trending venues around interested location during different parts of the day.

Results

In the result of this project I was able to obtain everything I expected. We have files with timestamps from different time during the day. These files hold information about trending venues around required location in particular time of the day. In any moment we can compare/ process this files, compile historical data on trending venues to better target ads in future, etc.

Also visualization part works as expected. We can see different venues during the day around required location.

Parsing trending venues, saving to files with timestamps if any, showing it on the map if required, all this is defined as a functions so could be reused in any time.

Discussion

Some observations I want to discuss here. I did my calls to API twice a day with 2 hours difference: ~11am and ~1pm local time. I expected to see trending venues around both CN tower in Toronto and Empire State Building in NewYork. But both times Toronto didn't give me any trends. I suspect this happens because it's the middle of the usual working day and probably there is just much more trends in NewYork because of tourists (even this time of year) as opposite to Toronto.

The good part is that we got successful results for NewYork and trends are different between this 2 hours.

As you can see in notebook

- around 11am we had the following trending venues around Empire State:

2 train stations, convention centre, google office and art museum.

- around 1pm trending venues changed to:

we can still see 2 train stations, convention centre, google office and art museum in the list, but in addition we can see three restaurants as trending.

In our case with ads we can presume that changing our ads during lunch time to restaurants might be a good idea so as having some museum tour in our ads during morning hours.

Conclusion

This project shows how powerful collecting trends might be. With this technic we can collect data during long period of time and in future even train models to predict trending venues and correct them, which will allow us to make targeted ads depends on the time of the day more successful. Overall I'm happy with this project and there is many features we could think of that might be implemented in future using this project.