Introduction/Business Problem

I have been approached by a friend of means who is looking to explore the possibility of chartering a specialty running store in the Pittsburgh, PA city limits. My friend is an avid runner, but has never been to Pittsburgh. He is relocating to the area to live closer to his wife's family, and as such, will be starting over fresh in a new city with this business venture. He has reached out to me, as I have lived in the city for a number of years and am also an active runner in the city and am well connected with many local running groups.

I latched onto this idea because, I have long believed that the city of Pittsburgh is unique in that running is extremely very popular and the geography of the region offers some of the most unique and exciting running routes in the country, but there are very few, if any, running supply stores specifically catered towards runners located in the city.

After discussing this further with my friend, I have learned that he envisions a store located close enough to some of the popular running areas within the city so that local groups can meet at the store before and after runs. The owner also wants the store to be a hub for competitive runners in the area and would like it to be accessible enough to draw in the college and young adult runners within the area. My friend pictures a store that focuses on high quality customer service to runners of all levels and abilities and will offer custom shoe fittings and will showcase all the latest running apparel technologies and advancements. My friend has emphasized to me that location of the store is the most important factor and I should begin my analysis with this in mind. He would like the store to be as near as possible to the "trendy" neighborhoods where runners live and work while still being accessible to the municipal parks and other popular civic accommodations.

After I have presented him with the optimal location for the store, I should then begin an assessment of the competition that the store would face within this chosen location. These two parts shall form the backbone of the data analysis conducted for my friend.

This report is written in somewhat of a narrative style, relaying how my client would prefer the project to go. This is because this is typically how I've found these projects unfold. I present a little bit of analysis to the client or customer, and that person guides the project path direction from that point onward.

Data

From a data science perspective, this project will involve two parts. The first part shall be an analysis of location data using a database published by the Western Pennsylvania Regional Data Center. A GeoJson file of the parks and running trails within the city limits of Pittsburgh shall be analyzed in order to determine the optimal location for the running store as it relates to the criteria outlined by my friend who commissioned me for this study.

URL: http://pghgis-pittsburghpa.opendata.arcgis.com/datasets/e95593cb0a2d4ff194be9694b40614dc_0.geojson

This location analysis will be the primary that is used to locate the running apparel store. The output of this analysis shall be a proposed location (Latitude and Longitude) for the store so that it meets the criteria of my friend.

The second part of the analysis shall be an analysis of Foursquare API data to highlight key competitors or similar retail outlets located near the proposed store location (given the Latitude and Longitude of Part 1). This data analysis will focus on other popular stores of similar nature within a set distance from the store. The output of this analysis will be a list of a few "Top" competitors so that my friend can conduct an analysis on how to differentiate his store from other existing stores in the area and how to focus is marketing campaign to fill gaps that these stores are leaving in the community.

Data, where possible, is presented in a visual map form (if possible), and all other metrics from the data analysis shall be included in this report.

Methodology

Part 1

The first part of this data analysis consisted of analyzing the parks and green space areas immediately around a central location of the city in order to fully capture the entirety of the cityscape. This was done by creating a "Stamen Terrain" map focused around the central point within the city of Pittsburgh: Point State Park. The latitude and longitude of Point State Park was found using the Geolocator package, and a large Red circle was drawn at the center of this location showing the approximate radius that my friend was interested in searching for the location of his running store. This style of map made will make it extremely easy for my friend to visualize the green spaces within the city and be sure that his store location was near enough to some of these spaces. This was accomplished using the following commands.

```
# create map of Pittsburgh using latitude and longitude values
map_pitt = folium.Map(location=[latitude, longitude], zoom_start=13,tiles='Stamen Terrain')
map_pitt.add_child(folium.features.CircleMarker([latitude, longitude],radius=15, color='black',fill=True,fill_color='black',fill_opacity=0.6))

map_pitt.add_child(folium.features.CircleMarker([latitude, longitude],radius=520, color='red',fill=True,fill_color='red',fill_opacity=0.2))
```

Additionally, the five colleges with a presence within city limits were added as points of interest. These locations were queried using the Geolocator package and the addresses for each location were used to look up the latitude and longitude of each university. The picture below is a snapshot of the code showing how this was done.

```
#This impoorts the location of the Univerity of Pittsburgh
address_upitt = '4200 Fifth Ave, Pittsburgh, PA'

geolocator_upitt = Nominatim(user_agent="foursquare_agent")
location_upitt = geolocator_upitt.geocode(address_upitt)
latitude_upitt = location_upitt.latitude
longitude_upitt = location_upitt.longitude
#print(latitude_upitt, longitude_upitt)
print('The geograpical coordinate of the University of Pittsburgh are {}, {}.'.format(latitude_upitt, longitude_upitt))
```

However, this city mapping was not the only tool employed to determine the potential location for the park. My friend was interested to see so much green space in the map viewer, but was curious exactly what the biggest parks were in the city of Pittsburgh. In this way, he tasked me to come up with a list and location of the five biggest parks and where they were located in proximity to the area that he was targeting for his store. For this, I found a JSON file of all the parks in the city of Pittsburgh, as mentioned in the data section. This file was interrogated in the notebook, as evidenced by the commands and code in the notebook. The five largest parks were found from this large GeoJSON data set and written to a

table along with their square footage of each park. The commands to do this are lengthy and as such, are not reproduced in this report, but they can be found in the notebook.

Part 2

When I showed the totality of results from Part 1 to my friend, he was extremely excited. These two parks: Schenley and Frick park are both extremely close to my friend's top two schools of choice: University of Pittsburgh and Carnegie Mellon University. My friend said that he would begin to look at empty real estate in this area, right between the two largest parks of Schenley and Frick parks as a potential landing spot for his running store concept. He was extremely excited about the possibility of collegiate running programs partnering with his store and leading runs in both parks with local athletes from such a store location. He also was very excited about the prospect being able to host regular running groups as they could depart right from the storefront and run in either (or both) of Pittsburgh's largest parks.

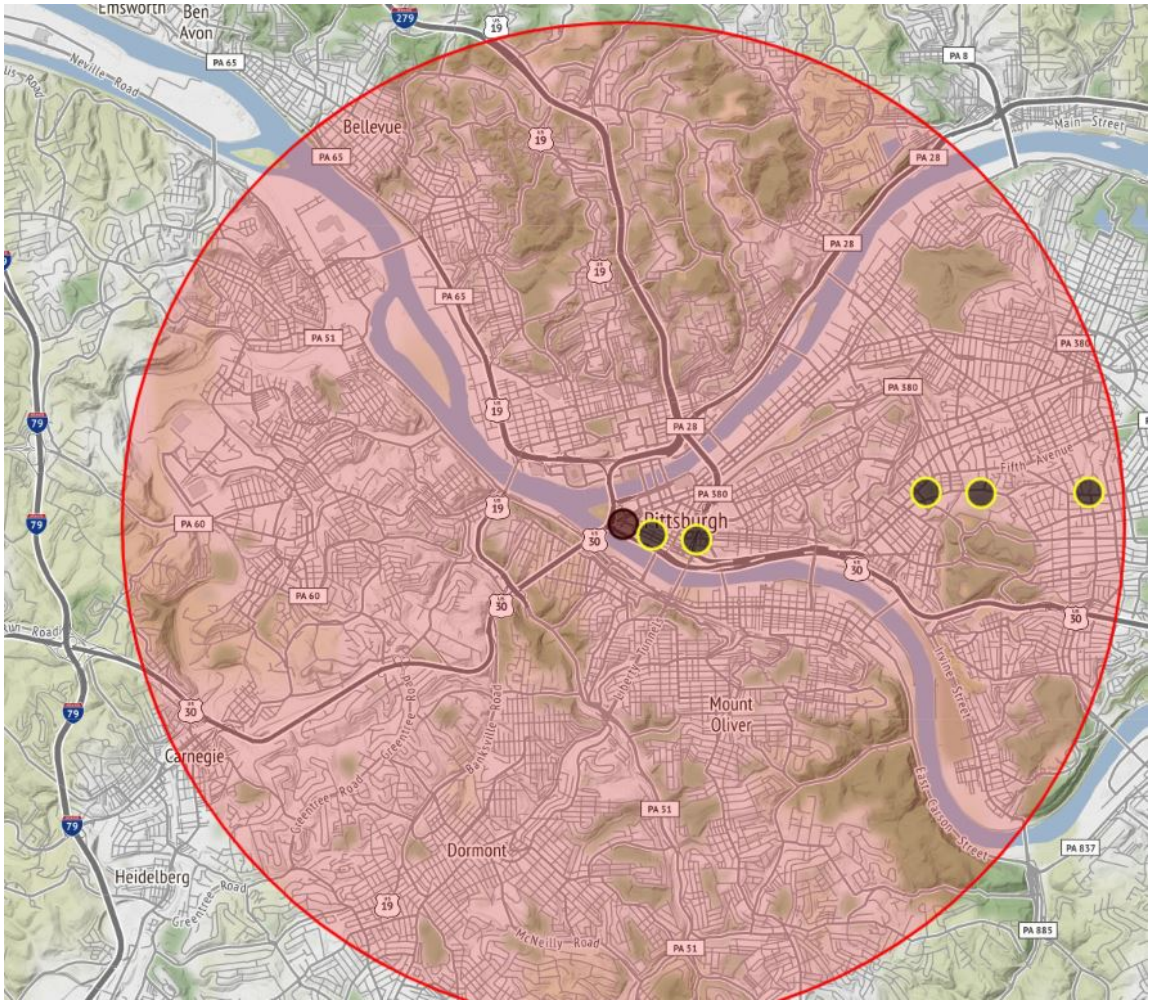
The second part of this data analysis focuses on the output of Part 1 and the input from my friend to this previous assessment. Based on the output of the Part 1 analysis, as described in the Results section, my friend got back to me with a location of an empty storefront that he found interesting in the Squirrel Hill Neighborhood of Pittsburgh. He thought that this commercial area would be an absolutely perfect location for his running store. However, he wasn't sure what the competition looked like in the area. He tasked me with trying to locate as many potential competitor stores within the area as I could. I set about doing this with a Foursquare API data query to find out what kinds of retailers and venues were in the area and whether any of them might represent significant competition for my friend's running store. Using the hypothetical store location that he gave me, I submitted a Foursquare API that returned 48 venues within a "reasonable" radius of 3 kilometers. I chose 3 kilometers for the search radius in order to capture as many venues as possible within a reasonable driving distance from the people who live and work in the running store's hypothetical chosen area.

I then sorted these venues by venue type so that he could easily see the types of venues that were in the area and which ones might be a direct competitor of his new store. I provided him this data in a table format because he wanted to go down the list carefully and research each potential venue that might be a competitor. One interesting, unforeseen result that came from this search was that many of the venues returned from the Foursquare API were actually in the "trail" category and represented popular running trails and spots near the potential running store location. My friend was very excited to have learned this.

Results

Part 1

All of this information was plotted to the previously created “Folium” map points of interest for my friend were added to the map in order to serve as a visual marker. The resulting map that was created is shown below.



This folium map was the primary visual aide that my friend used to narrow down the location of his running store. He appreciated that he was able to see on the map exactly where each school was located, using the pop-up box. He got excited when he realized how close Carnegie Mellon University and the University of Pittsburgh were from each other, as evidenced by my Map. This really got him thinking that this was the area he’d like to target, but he wanted a bit more information on the parks and running areas in the city before he really made an assessment of the area in which he wanted to have his store located.

Based on the methodology section, I set about to find the 5 largest parks by square footage area were found from the publicly available GeoJSON file. As discussed in the Discussion section, the top two parks

by size (Frick and Schenley Parks) were very close to each other which drove the initial narrowing of the search for a venue.

Out[24]:

	Park Name	Size Area	Latitude	Longitude
0	Frick Park	4.236350e+06	[[[-79.8976574229661, 40.4395292781351], [-79....	[[[-79.8976574229661, 40.4395292781351], [-79....
1	Schenley Park	3.035710e+06	[[[-79.9459692465441, 40.434687524324], [-79.9...	[[[-79.9459692465441, 40.434687524324], [-79.9...
2	Highland Park	2.646461e+06	[[[-79.9137561554273, 40.4860365267797], [-79....	[[[-79.9137561554273, 40.4860365267797], [-79....
3	Riverview Park	1.811034e+06	[[[-80.0130276883929, 40.4832761374082], [-80....	[[[-80.0130276883929, 40.4832761374082], [-80....
4	Emerald View Regional Park	7.727741e+05	[[[-80.0314879138697, 40.4399943174112], [-80....	[[[-80.0314879138697, 40.4399943174112], [-80....

This was then used to drive the results and methods of Part 2.

Part 2

The venues of interest, as chosen by my friend, is shown below in a RED dot in a folium map created below. This is the approximate location of where the owner and I believe the store should be located, after seeing the results of the Part 1 analysis. The store should be located in this commercial district

Based on this location, a Foursquare API search was run in order to determine if there were any similar running stores or places that might be direct competitors of the store. Based on this analysis and a search radius of 3 kilometers, 100 total venues were found in the area surrounding the hypothetical chosen location. The absolute top 5 closest venues are shown in the image below.

Out[59]:

	name	categories	lat	lng
0	The Manor	Indie Movie Theater	40.437247	-79.922673
1	Waffallonia	Breakfast Spot	40.437772	-79.922844
2	Hidden Harbor	Cocktail Bar	40.437673	-79.919409
3	Gluuteny Bakery	Bakery	40.435304	-79.922782
4	Taiwanese Bistro Cafe 33	Taiwanese Restaurant	40.437643	-79.919236

A few Gym venues and some Trails and parks were returned which were the most similar venue types to the running store concept. However, these venues would most likely not be competitors of the store, but would rather be welcome complements to the store.

Discussion

From the totality of the results and methodology mentioned above, there are a few things that can be noted from the data analysis. I believe that it is important to note that the parameters, boundaries, and results of this analysis were sharply shaped by the initial set of criteria given at the outset of this exercise. Given a different set of initial conditions, it's entirely possible that someone could arrive at an entirely different "optimal" location, and so it's important to note that the results of this analysis can only optimize a solution for the set of initial conditions.

In light of this however, I stand by the following recommendations based on the totality of the analysis detailed in Part 1 and Part 2 of this document. The location chosen meets the following initial criteria by being close to large running parks, near local universities, and not immediately adjacent any direct, known competitors.

However, there are three additional suggestions that I propose based on the results of this data analysis in order to assist my friend in his journey to open a running store in the Pittsburgh metro area.

1) Continue to look for real estate in the commercial are I've outlined: This will help determine if the chosen commercial district is economically feasible for a running store. This analysis did not examine commercial real estate prices, and so this type of further assessment would be extremely beneficial.

2) Conduct further research on the "potential competitors" list that I've provided: I believe this data analysis shows that there are not many stores in the chosen area that would be a direct competitor to the running store, I still think there is great benefit to be gained by going down the list of venues one by one and doing further "in person" research to really see what kinds of products and experiences the stores are offering to customers.

3) Conduct a Survey: I believe that some type of targeted survey to the store's potential demographic will generate a lot of additional, valuable information that can then be analyzed in a further assessment of the feasibility of opening a store. A survey will really be able to tell my friend a lot about the current purchasing and exercise habits of potential customers in the store.

All of these recommendations will be an opportunity to generate more data, that I believe could be a great "jumping off point" for more analysis.

Conclusion

This concludes the report on a potential location for a new running store in Pittsburgh, PA. The store location that was selected for this analysis was driven by data while still leaving room for human interpretation and preferences.

That being said, I believe that the analysis presented in this document shows a great spot for a running store that meets the following criteria:

1) Proximity to large running parks
2) Proximity to local universities and young people
3) Not located near any other competitors