Horvitz, Andrew
Capstone Project – IBM Data Science
10.4.2021

# IT Expansion

## 1. Introduction

### 1.1 Background :

Information technology managed service providers (MSPs) generally provided a set of "as a service" (XaaS) functions related to the maintenance, administration, and security of an organization's information systems and data. This generally includes support for an organization's users, some form of webhosting, and customized solutions for customers. MSPs enter in service-level agreements (SLAs) that define specifying performance information systems, definition, scope, and processes for termination of services.

A challenge for IT MSPs, like many other businesses, is to find new markets to serve and grow. One promising area is to provide IT managed services to small and medium sized businesses. The IT demands of small businesses are consuming enough that keeping up with them can become a full-time task. As such, many small businesses, rather than developing internal capabilities seek out MSPs that can take on the role of their "tech department". An IT MSP needs to actively seek out such businesses, and attempt to foster relationships.

From the point of view of the MSP there is potential long-term stability and viability by focusing on IT needs of small businesses. And these needs will only expand. However, to help bolster the chances of success geographic area is important consider in terms of a market to serve. While many functions can and are remotely executed an IT MSP also needs a physical footprint. Having an "onsite" presence, or at least accessibility, is critical to developing successful business relationships. Not only does equipment need to be attended to, having periodic in person touch points with clients helps foster fruitful business relationships.

In addition, different geographic markets may be better for structural reasons. For instance, if an IT MSP attempted to target businesses in affluent areas, the chances that the businesses they serve also serve the affluent is increased. While not immune, these communities tend to fair better during economic downturn or contraction and tend to outperform less affluent communities during periods of economic recovery or expansion.

## 1.2 Problem / Scenario:

A fictional company (the Company) provides services leveraging geolocation data to address the needs of organizations. An IT managed service provider specializing in serving small to medium size business (the Client) is looking to expand operations by establishing a satellite office. The Client approached the Company to provide recommendations for locations. The Client shared several requirements proposed locations including but not limited to:

1. Over 20 miles from current location, but not more than 45 miles of current location (located in Rockland County, NY 41.1489 N, 73.9830 W).
2. Mix of small to medium size businesses as well as highway, secondary, and tertiary routes and mass transit options within 5-mile radius. Minimize commuter chokepoints like bridges, tunnels, tolls, etc.
3. Off-street parking / lot.

## 1.3 Interest:

When a business grows it often does by expanding into a market it previously lacked access to. However, this can be a daunting task with many considerations to take into account. These involve the specific field the business operates in, the clients it serves, the needs of its employees, and prevailing market conditions. Leveraging geolocation data through analysis and presentation would be of value to stakeholders (i.e. other businesses) that find themselves in similar positions.

In the scenario outlined above an IT managed service provider (MSP) wants to expand operations. The requirements posed by the hypothetical company are based off of experiences I had working for a such a company several years ago.

1. Distance requirements: The Company would like to expand its service and coverage area, but be able to reasonably access from the home location. Although much management is done digitally and remotely equipment must be ordered, stocked, maintained, and serviced in the physical world (both onsite at Clients and in the Company's office). While somewhat arbitrary, the 55 mile (~88 km), is a modal highway speed limit in the region. Measured per hour, this means the satellite office would be roughly between a half hour and hours drive by auto given reasonable traffic conditions. Pushing towards the limit would expand potential service area. The minimum distance is sought to ensure the service area expands without too much coverage overlap.
2. Minimize commuter chokepoints between offices: Road congestion is a concern in the Company's current location. For this point in the proposed expansion, to better serve Clients and staff, locations that would exacerbate travel time should be avoided.
3. Access to transit: The Company would like easy connections between the new and existing location, easy access to reach clients and offices for those working on-site.
4. A mix of target businesses: Serving small to medium size businesses the Company would target the same in the location area. To best maximize access to onsite needs of clients the Company seeks to be physically close to a potential client base.
5. Off-street parking: Ease of access for employees, clients, and vendors.

## 2. Data acquisition and cleaning

### 2.1 Data sources

The scenario outlined asserts that serving businesses in affluent communities is most aligned with the interests of the Client. This means that information about affluent communities must be obtained. One way to operationalize affluence is by measuring per-capita income.

A proposed solution to the scenario will begin by referencing communities reporting the highest per capita income in the United States. These areas are organized into to ZCATs (Zip Code Tabulation Area) which are designated by the US Census Bureau. This information will be scraped from the web (https://en.wikipedia.org/wiki/List_of_highest-income_ZIP_Code_Tabulation_Areas_in_the_United_States). Next, an independent file that cross lists ZCAT's with US Postal Service Zip Codes and geographic coordinates (https://raw.githubusercontent.com/adhorvitz/coursera_ibm_capstone/main/uszips_mod_zipgeoonly.csv) will read. After information from both sources is cleaned and concatenated it will be plotted on a map. At this point, potential locations will be checked against the Client's requirements.

After a location is preliminarily selected two additional pieces of information will be collected. First, an API call will be run to foursquare to get a sense of the number and geographic distribution of businesses / potential customers in the locale. Next, a search will be performed for potential competitors of the Client. Both potential customers and competitors will be mapped.

These maps along with (blank analysis to "measure" geographic density) will be presented to the Client for review.

## 2.2 Data cleaning and priming

Data was initially taken from two sources. A table containing ZCATs and per-capita income in the United States was scraped into a JSON file, reading into a table (table 1), and then converted to a dataframe. Taken in its entirety, the table listed five columns, "Rank", "Designation (name)", "ZCAT", "Population", and "Per Capita Income". In total, one hundred communities were ranked (table 1; N = 100).

A second table (table 2) sourced from census data records, provided "Zip Codes" along with corresponding geographic coordinates, "lat" (latitude) and "lng" longitude". The second table

was also rendered into a date frame containing. However, this second dataframe contained over 33,000 records (table 2; N = 33,121).

The two tables were merged into a single dataframe (merged_dataframe) with all of the necessary information required to plot. However, the merged_dataframe contained voluminous extraneous information given a target of 100 communities.

Merged_dataframe was then inspected, and a common values were found to exist in all extraneous records. For instance, in the "Rank" column in merged_dataframe "NaN" (Not a Number) was recorded for all records not taken from table_1. This meant by searching for, and eliminating any record with "NaN" the remaining records would the ones corresponding to the 100 communities having  a ranked value.

| | old_index | zip | lat | lng | rank | place | population | per_cap_inc |
|---|---|---|---|---|---|---|---|---|
| 0 | 473 | 2108 | 42.35767 | -71.06505 | 62.0 | Boston, Massachusetts | 3446.0 | 78771.0 |
| 1 | 474 | 2109 | 42.36459 | -71.05298 | 45.0 | Boston, Massachusetts | 3428.0 | 82689.0 |
| 2 | 475 | 2110 | 42.35826 | -71.05175 | 44.0 | Boston, Massachusetts | 1428.0 | 82736.0 |
| 3 | 524 | 2199 | 42.34747 | -71.08202 | 28.0 | Boston, Massachusetts | 1005.0 | 88974.0 |
| 4 | 579 | 2493 | 42.35894 | -71.30010 | 59.0 | Weston, Massachusetts | 11469.0 | 79640.0 |

*Figure 1 - df_income_zip_good.head()*

## 3. Methodology and Exploratory Analysis

With data now ready all the information necessary to begin plotting can take place. Folium was used to generate maps. The new dataframe iterated (df_zip_top_income) provides geocoordinates, names, populations, per capita income, and population for upper socioeconomic status by ZCAT. A map (map_usa) was then created with all 100 communities.

### **3.1** Visualizing Potential Locations – Client Specification 1

The first map plots potential locations for the prosed office based on high per capita income areas.

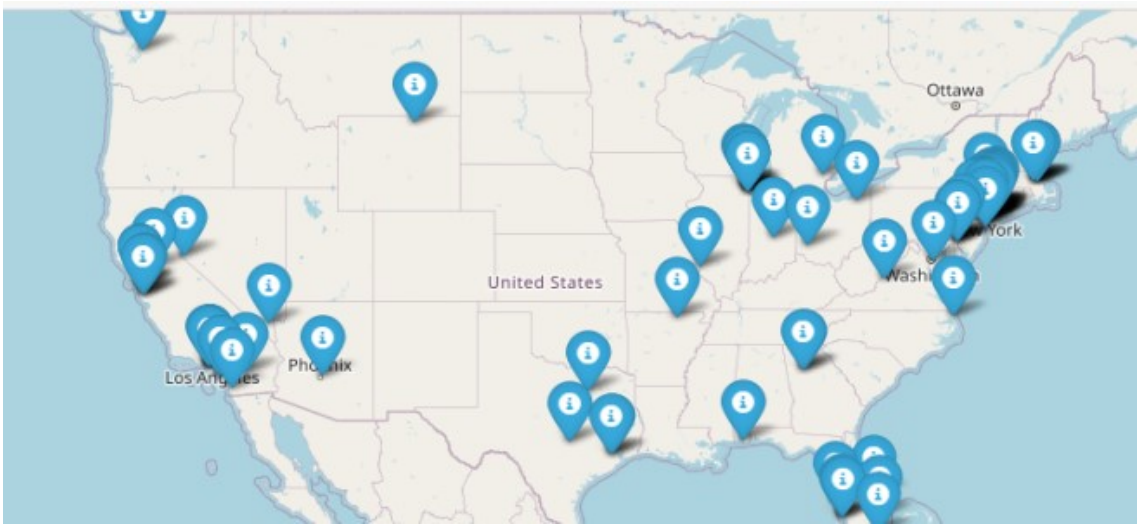


*Figure 2 - Map of 100 Top Per Capita Income Communities in the US*

### **3.2** Visualizing Potential Locations – Client Specification 2

The first map was overlayed with distance demarcations specified by the client. The inner circle and out circle represent a 20- and 45-mile radius from the current office location. With these specifications our initial list of 100 possible locations was effectively reduced to just under two dozen options (N = 24).

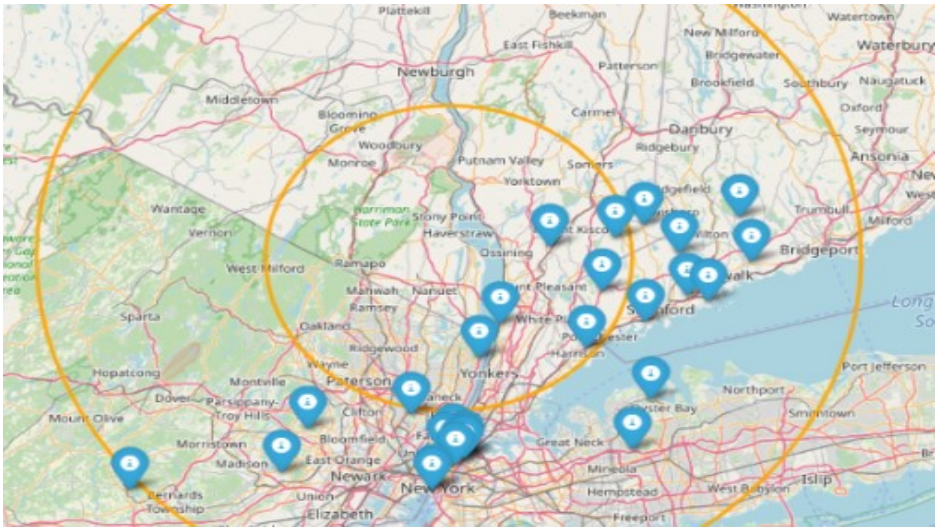


*Figure 3- Distance Boundaries from Existing Office*

## 3.3 Visualizing Potential Locations – Client Specification 3

Applying the next requirement to reduce or provide alternatives to traffic choke points (e.g. bridges, tunnels, tools) produced three suitable locations: 1) Essex Fells, NJ (Loc 1) 2) Short Hills, NJ, (Loc 2) and 3) Teterboro, NJ (Loc 3). At this juncture the next set of requirements were assessed. While avoiding choke points the Client requested that the location be within easy access to highways and mass transit. This was determined to be roughly five miles from each selected locations approximate geographic center.

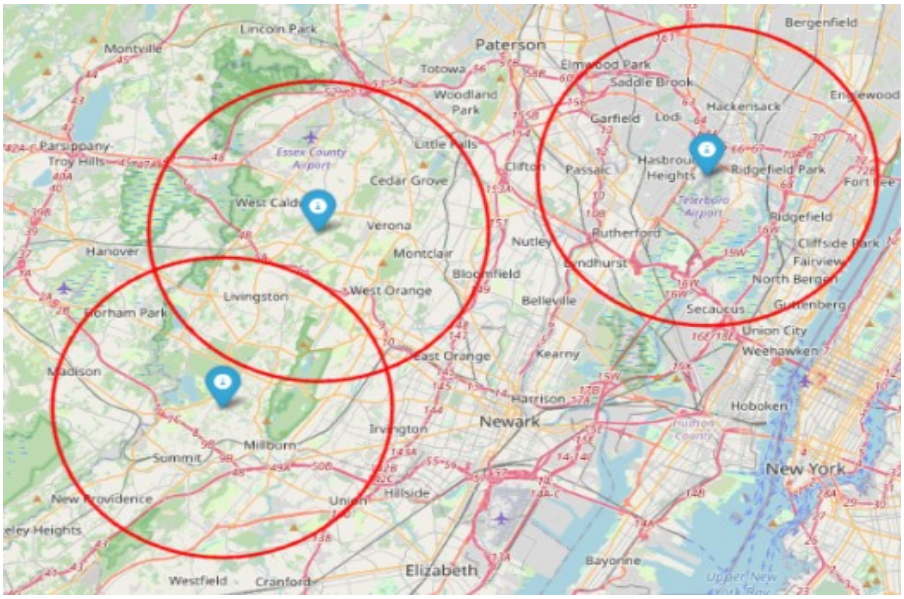


*Figure 4- Final Locations Relative to One Another*

- Location 1 is located within five miles of access to two interstate highways (I80 and I280), a major state limited access parkway (The Garden State Parkway), major state roads, nearly a dozen county numbered roads, and is served by three lines of NJ Transit (commuter rail) as well as a regional airport.
- Location 2 is located within five miles of access to one interstate highway (I78) a major state limited access parkway (The Garden State Parkway), major state roads, nearly a dozen county numbered roads, and is served by two lines of NJ Transit (commuter rail).
- Location 3 is located within five miles of access to two interstate highways (I95 and I80), several major state roads, nearly a dozen county numbered roads, and is served by four lines of NJ Transit (commuter rail) as well as a regional airport.

| | index | old_index | zip | lat | lng | rank | place | population | per_cap_inc |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 12 | 1991 | 7021 | 40.82686 | -74.27973 | 67.0 | Essex Fells, New Jersey | 2151.0 | 77787.0 |
| 1 | 13 | 2039 | 7078 | 40.74207 | -74.33378 | 21.0 | Short Hills, New Jersey | 12849.0 | 92940.0 |
| 2 | 14 | 2135 | 7608 | 40.85368 | -74.06129 | 92.0 | Teterboro, New Jersey | 18.0 | 72613.0 |

*Figure 5- Dataframe of Final Possible Locations*

### **3.4** Visualizing Potential Locations – Client Specification 4

Using Foursquare potential clients were identified in each location. They are all marked in red.

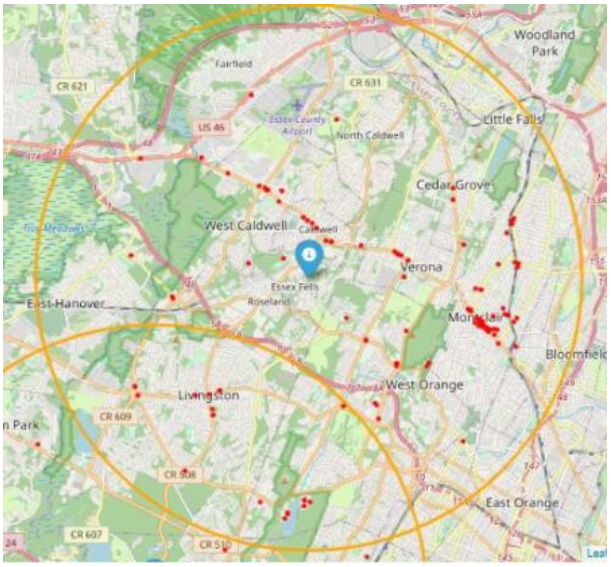


*Figure 6 - Essex Fells & Potential Clients*

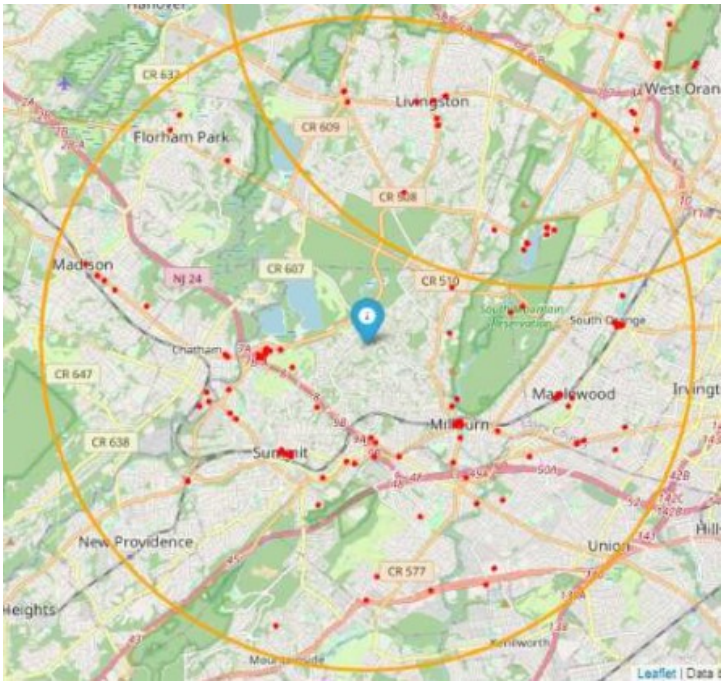


*Figure 7 - Short Hills & Potential Clients*

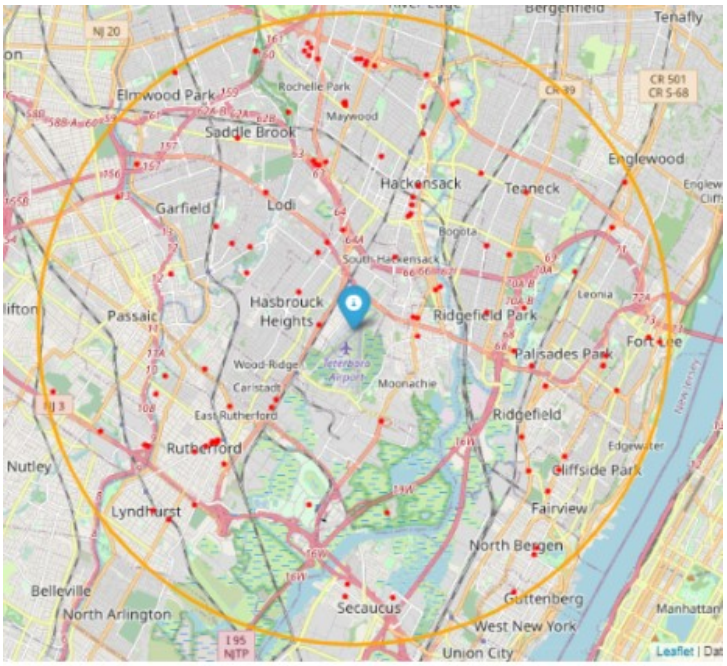


*Figure 8 - Teterboro & Potential Clients*

### **3.5** Visualizing Potential Locations – Client Specification 5 – parking

All areas offer commercial locations with off street parking. Local relators best suited to find individual properties after town is selected.

### **3.6** Visualizing Potential Locations – Client Specification – bonus – possible competitors

To better help our clients this report also sought to find possible competitors in the area. Data from Google Maps was organized into a csv file, and then mapped. Potential competitors are marked in black.

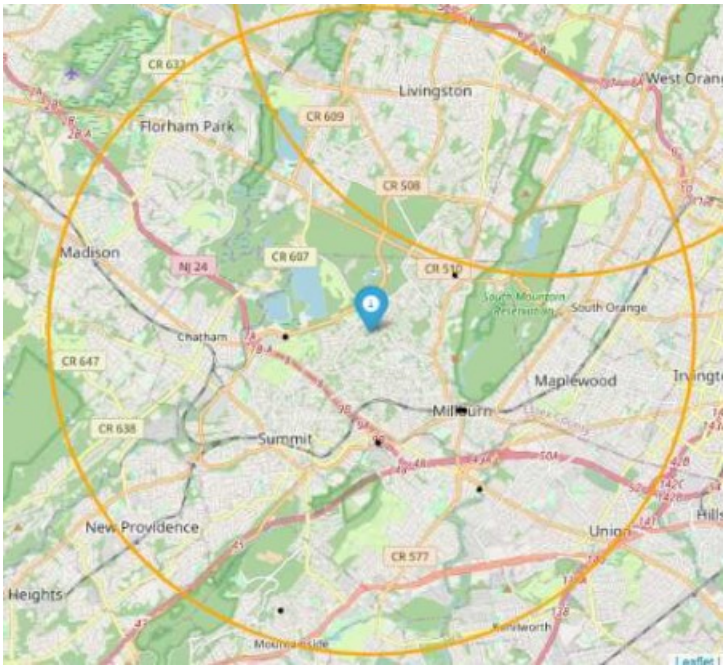


*Figure 9 - Short Hills Potential Competitors*

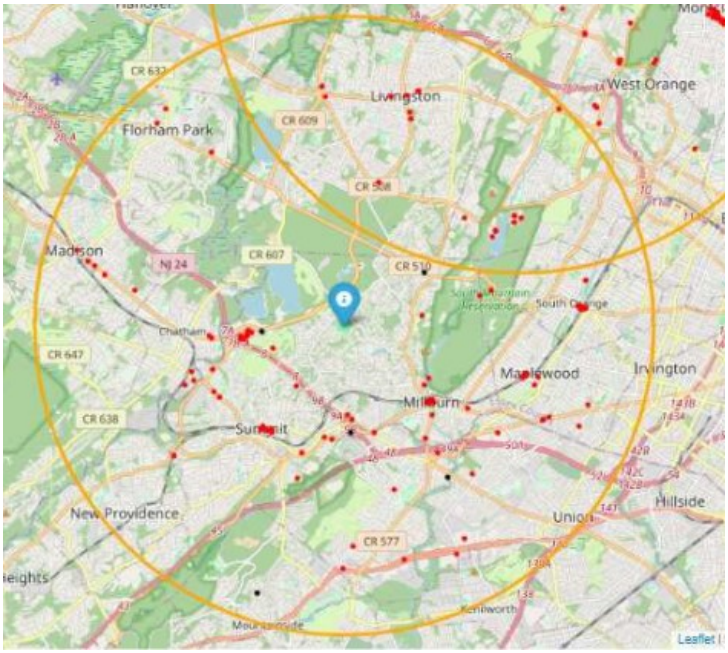


*Figure 10 - Short Hills – Potential Competitors & Potential Clients*

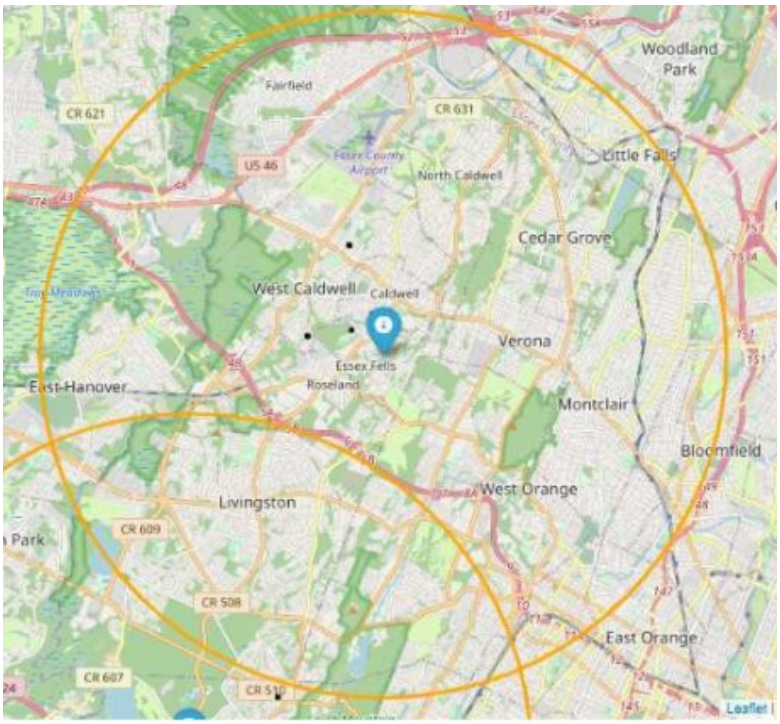


*Figure 11 - Essex Fells Potential Competitors*

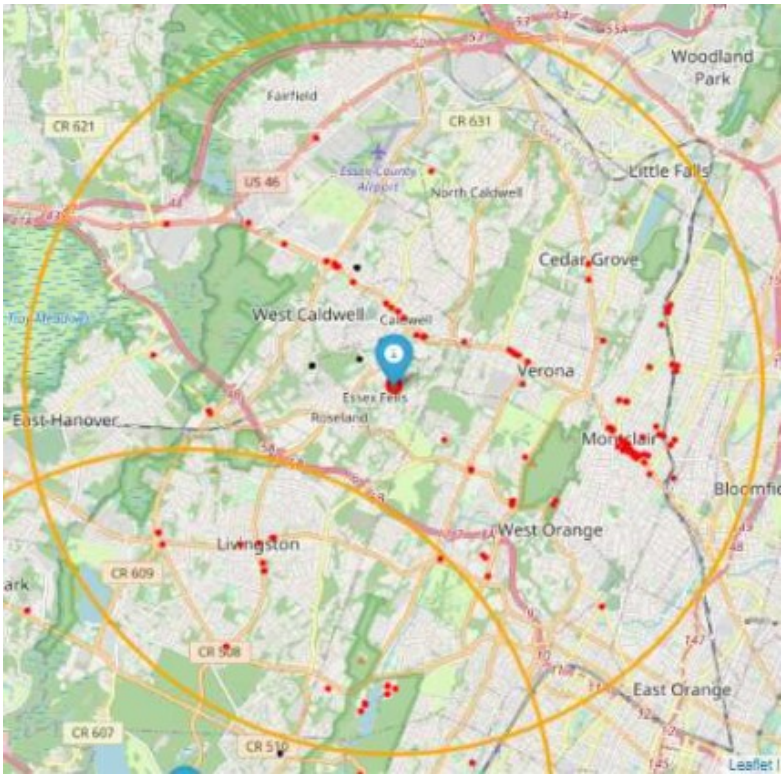


*Figure 12 - Essex Fells – Potential Competitors & Potential Clients*

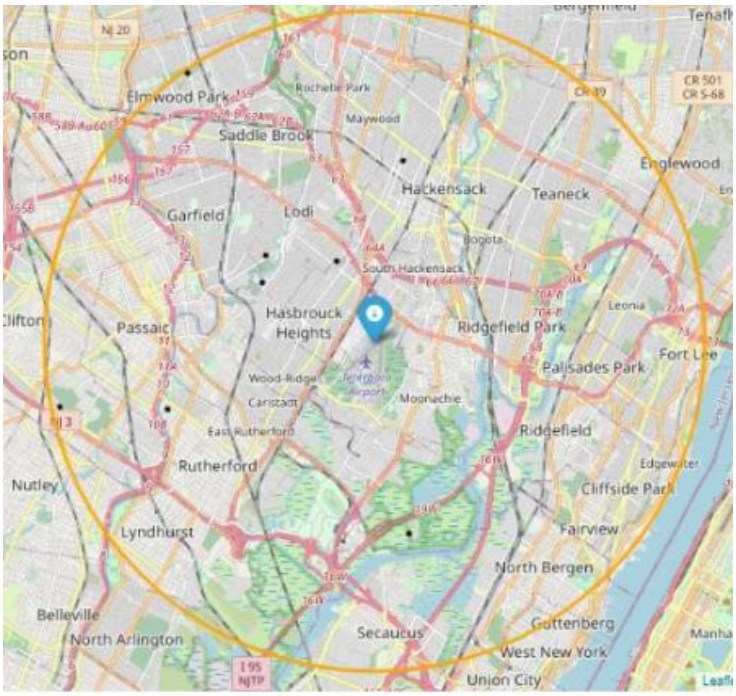


*Figure 13 - Teterboro Potential Competitors*

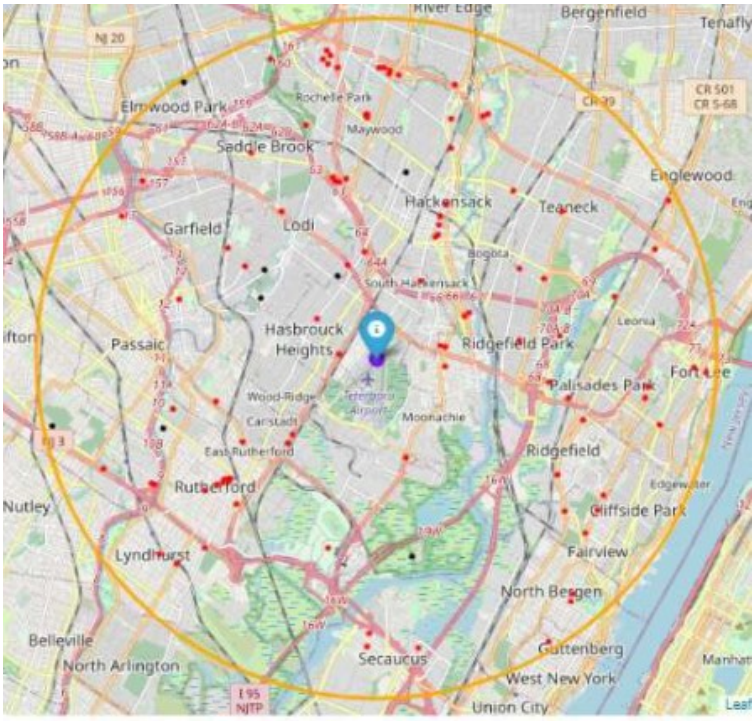


*Figure 14 - Teterboro – Potential Competitors & Potential Clients*

## 4. Recommendations / Discussion

After reviewing the visual data, **the Company recommends to the Client to consider** not an area, but rather **a corridor that exists between Essex Fells and Short Hills**. Both locations (Essex Fells and Short Hills) contain a large number of potential clients, access to major transits options, and fall within the specified distance as per the Client's requirements. Since there is overlap, the interstitial area between the two communities might be ideal given the Client's parameters.

While Teterboro did fulfil the requirements specified by the Client there were more potential competitors identified in the area. Moreover, given its proximity to New York City (within 6.5 miles / 10.5 KM) from the George Washington Bridge, one of the busiest bridge for vehicular traffic in the world, being located further west would prove beneficial.

An IT managed service provider (the Client) is seeking to expand operations by establishing a second location. The Client had several requirements concerning distance from current location, traffic and transit options, as well as a mix of small to medium sized business that would create a pool of potential clients in the local area.

## 5. Conclusions

After considering the Client requirements the Company drew on existing social scientific knowledge which suggests affluent communities are generally better insulated from economic downturns, and generally have better infrastructure and services relative to less affluent communities. Moreover, while affluent residents may not work or have business in the communities, they live they are likely be patrons of local businesses. Being in a position to provide business services to business in affluent communities creates a potential opportunity, while mitigating certain risks for the Client.

Collecting data on per capita income based on geographic location produced a list of 100 localities spread across the United States. Factoring in Client requirements, three potential locations were selected. Next, potential clients were identified and plotted using data from Foursquare. As a a complimentary feature in analysis, potential competitors were mapped for each location.

After reviewing the visual data a recommendation was made to the Client to consider an interstitial location within the areas Essex Fells and Short Hills area.

At this point the recommendation is brough back to the client for feedback. Further geospatial analysis of the final location selected by the Client can be performed.