Introduction

In this section the business problem at hand as well as the target audience for the research project “The Battle of Neighbourhoods” are presented.

Business problem

The problem I am going to solve with data from Foursquare and and other sources is which neighbourhood in Manhattan is most suitable for a Swedish pizza restaurant chain to establish. This (fictive) chain is called SwePizz and has an established brand in Sweden and it is time for them to expand internationally. They are confident that the brand will be strong in all areas of Manhattan and to choose their first entry in this city they want to find a neighbourhood with few pizza restaurants. I will achieve this by using data science methodology to find the Manhattan neighborhood that currently has the fewest pizza restaurants. The idea behind this is that the demand is assumed to be high in all areas since SwePizz's pizzas have historically captivated all categories of customers and therefore the focus is on finding an area with low supply in order to efficiently grow the customer base.

Target audience

This research will be valuable to the management team of SwePizz in order to make an effective entry into their next market. Entering the right area will help them gain traction and start building their brand in New York City, making it an important strategic decision for their long term international precense.

Data & Methodology

The data I will use for this project are:

- New York City neighborhood data from health.ny.gov
- Manhattan geodata from public.opendatasoft.com
- Venue categories and locations (latitude, longitude) from foursquare.com

The neighborhood data will show the names and postal codes for neighbourhoods in New York City, as well as the boroughs they are located in. Manhattan is the borough of focus for SwePizz. Two .csv files will be merged, one each from health.ny.gov and public.opendatasoft.com, to join postal code and neighborhood data with latitude and longitude data. The venue categories and locations data will show how many restaurants pizza in each Manhattan neighborhood by joining at approximate geographical coordinates (long/lat). Foursquare data will be retrieved to find nearby venues and then various instances of the pandas library will be used to refine the data to eventually create a table which displays the number of pizza restaurants in each Manhattan neighborhood.

Results, conclusion, and discussion

East Harlem is the Manhattan neighborhood with the least pizza restaurants and therefore the recommendation for SwePizz will be to establish there. In addition the low number of

pizza restaurants in East Harlem, the location is also known for its diverse food culture with influences from around the world (according to Wikipedia). SwePizz has a combined Swedish/Turkish heritage and therefore they should be able to contribute to the vibrant and interesting culture of East Harlem while getting a great start in their internationalization process. Further research could be done for SwePizz's management team, by providing a geographic visualization of the location of East Harlem as well as the locations of the existing pizza restaurants in the neighborhood. This could inform the decision making regarding exact location for establishment in East Harlem. Do SwePizz want to get close to the competition or do they want to keep an arms-length distance? To answer that question more analysis is needed, looking into things like center of attraction and flow of people in different areas.

To go through the results piece by piece, I will begin with what occurred first in the project, namely importing New York City neighborhood data. See the table below.

	Neighborhood	Borough	Postalcode
0	Central Bronx	Bronx	10453
1	Central Bronx	Bronx	10457
2	Central Bronx	Bronx	10460
3	Bronx Park and Fordham	Bronx	10458
4	Bronx Park and Fordham	Bronx	10467
...	...	...	...
173	South Shore	Staten Island	10312
174	Stapleton and St. George	Staten Island	10301
175	Stapleton and St. George	Staten Island	10304
176	Stapleton and St. George	Staten Island	10305
177	Mid-Island	Staten Island	10314

Then, I found data for Manhattan neighborhood geolocations, and imported those.

	Postalcode	Latitude	Longitude
0	10001	40.750742	-73.996530
1	10002	40.717040	-73.987000
2	10003	40.732509	-73.989350
3	10005	40.706019	-74.008580
4	10006	40.707904	-74.013420
...	...	...	...
161	10292	40.780751	-73.977182
162	10422	40.828279	-73.869454
163	11286	40.658825	-74.004495
164	11302	40.759450	-73.715016
165	11517	40.679845	-73.986415

Since this data was restricted to Manhattan and did not include the wider New York City, merging these to at postal code automatically cleaned the previous data set from the rest of the boroughs (Bronx, Brooklyn, Queens, and Staten Island). This resulted in the following table.

	Postalcode	Latitude	Longitude	Neighborhood	Borough
0	10001	40.750742	-73.99653	Chelsea and Clinton	Manhattan
1	10002	40.717040	-73.98700	Lower East Side	Manhattan
2	10003	40.732509	-73.98935	Lower East Side	Manhattan
3	10005	40.706019	-74.00858	Lower Manhattan	Manhattan
4	10006	40.707904	-74.01342	Lower Manhattan	Manhattan
5	10007	40.714754	-74.00721	Lower Manhattan	Manhattan
6	10009	40.727093	-73.97864	Lower East Side	Manhattan
7	10010	40.739022	-73.98205	Gramercy Park and Murray Hill	Manhattan
8	10011	40.741012	-74.00012	Chelsea and Clinton	Manhattan
9	10012	40.725960	-73.99834	Greenwich Village and Soho	Manhattan
10	10013	40.720666	-74.00526	Greenwich Village and Soho	Manhattan
11	10014	40.734718	-74.00532	Greenwich Village and Soho	Manhattan
12	10016	40.746180	-73.97759	Gramercy Park and Murray Hill	Manhattan
13	10017	40.752159	-73.97231	Gramercy Park and Murray Hill	Manhattan
14	10018	40.755101	-73.99337	Chelsea and Clinton	Manhattan
15	10019	40.765714	-73.98560	Chelsea and Clinton	Manhattan
16	10020	40.758667	-73.98024	Chelsea and Clinton	Manhattan
17	10021	40.768420	-73.96045	Upper East Side	Manhattan
18	10022	40.759015	-73.96732	Gramercy Park and Murray Hill	Manhattan
19	10023	40.776099	-73.98285	Upper West Side	Manhattan
20	10024	40.786387	-73.97709	Upper West Side	Manhattan
21	10025	40.798502	-73.96811	Upper West Side	Manhattan
22	10026	40.802853	-73.95471	Central Harlem	Manhattan
23	10027	40.812665	-73.95499	Central Harlem	Manhattan
24	10028	40.776777	-73.95410	Upper East Side	Manhattan
25	10029	40.791586	-73.94575	East Harlem	Manhattan
26	10030	40.818151	-73.94351	Central Harlem	Manhattan
27	10031	40.826201	-73.94879	Inwood and Washington Heights	Manhattan
28	10032	40.840686	-73.94154	Inwood and Washington Heights	Manhattan
29	10033	40.848764	-73.93496	Inwood and Washington Heights	Manhattan
30	10034	40.867653	-73.92000	Inwood and Washington Heights	Manhattan
31	10035	40.802395	-73.93359	East Harlem	Manhattan
32	10036	40.759511	-73.99019	Chelsea and Clinton	Manhattan
33	10037	40.813385	-73.93616	Central Harlem	Manhattan
34	10038	40.709677	-74.00365	Lower Manhattan	Manhattan
35	10039	40.826181	-73.93710	Central Harlem	Manhattan
36	10040	40.858704	-73.92853	Inwood and Washington Heights	Manhattan
37	10044	40.762174	-73.94917	Upper East Side	Manhattan
38	10128	40.781894	-73.95039	Upper East Side	Manhattan
39	10280	40.707467	-74.01780	Lower Manhattan	Manhattan

Having this data in place, I just needed to relate it to the venue data via Foursquare. Doing this, I found nearby venues based on approximate geodata. I also got category data, telling

me whether a venue was a “Pizza Place” or not.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Chelsea and Clinton	40.750742	-73.99653	New York Pizza Suprema	40.750124	-73.994992	Pizza Place
1	Chelsea and Clinton	40.750742	-73.99653	You Should Be Dancing.....! / Club 412	40.750306	-73.994743	Dance Studio
2	Chelsea and Clinton	40.750742	-73.99653	Music Choice	40.752632	-73.994585	Music Venue
3	Chelsea and Clinton	40.750742	-73.99653	Madison Square Garden	40.750752	-73.993542	Basketball Stadium
4	Chelsea and Clinton	40.750742	-73.99653	Bluestone Lane	40.752068	-73.998848	Coffee Shop
...	...	...	...	...	...	...	...
3253	Lower Manhattan	40.707467	-74.01780	Gym @ The Terrace Club	40.708995	-74.018358	Gym
3254	Lower Manhattan	40.707467	-74.01780	Stone Garden	40.705965	-74.018846	Outdoor Sculpture
3255	Lower Manhattan	40.707467	-74.01780	Swedish Midsummer At Wagner Park	40.705408	-74.018726	Park
3256	Lower Manhattan	40.707467	-74.01780	Eyes by Louise Bourgeios	40.704828	-74.018796	Outdoor Sculpture
3257	Lower Manhattan	40.707467	-74.01780	Pier A	40.704221	-74.018379	Pier

The pandas library then came to great use to manipulate the data in order to answer my initial question “which Manhattan neighborhood has the fewest pizza restaurants?”. Pandas allowed me to filter away all venue categories except “Pizza Place”.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Chelsea and Clinton	40.750742	-73.99653	New York Pizza Suprema	40.750124	-73.994992	Pizza Place
108	Lower East Side	40.717040	-73.98700	Champion Pizza - Ludlow	40.719190	-73.988850	Pizza Place
176	Lower East Side	40.717040	-73.98700	Sauce Pizzeria	40.720368	-73.988830	Pizza Place
190	Lower East Side	40.717040	-73.98700	Scarr's Pizza	40.715335	-73.991649	Pizza Place
195	Lower East Side	40.732509	-73.98935	Joe's Pizza	40.733234	-73.987672	Pizza Place
...	...	...	...	...	...	...	...
3098	Upper East Side	40.781894	-73.95039	Marinara Pizza Upper East	40.782538	-73.953359	Pizza Place
3109	Upper East Side	40.781894	-73.95039	Nick's Restaurant & Pizzeria	40.782923	-73.948014	Pizza Place
3172	Upper East Side	40.781894	-73.95039	Luigi's Pizzeria	40.778222	-73.948426	Pizza Place
3186	Upper East Side	40.781894	-73.95039	Domino's Pizza	40.782746	-73.945352	Pizza Place
3196	Lower Manhattan	40.707467	-74.01780	Inatteso Pizzabar Casano	40.705886	-74.016661	Pizza Place

Finally I reached the conclusion that East Harlem had the fewest pizza restaurants and therefore East Harlem would be my recommendation for SwePizz’s for international establishment.

Neighborhood	Venue
Central Harlem	5
Chelsea and Clinton	9
East Harlem	3
Gramercy Park and Murray Hill	8
Greenwich Village and Soho	4
Inwood and Washington Heights	16
Lower East Side	9
Lower Manhattan	12
Upper East Side	12
Upper West Side	6

Then small adjustments with pandas to rename “Venue” into “Pizza Restaurants”

Neighborhood	Pizza Restaurants
Central Harlem	5
Chelsea and Clinton	9
East Harlem	3
Gramercy Park and Murray Hill	8
Greenwich Village and Soho	4
Inwood and Washington Heights	16
Lower East Side	9
Lower Manhattan	12
Upper East Side	12
Upper West Side	6

Can also see that Greenwich Village and Soho only has one more pizza restaurant that East Harlem, making that a potentially attractive next area to enter.