IBM DATA SCIENCE SPECIALIZATION

APPLIED DATA SCIENCE CAPSTONE PROJECT

The Battle of Neighbourhoods: New York vs Toronto

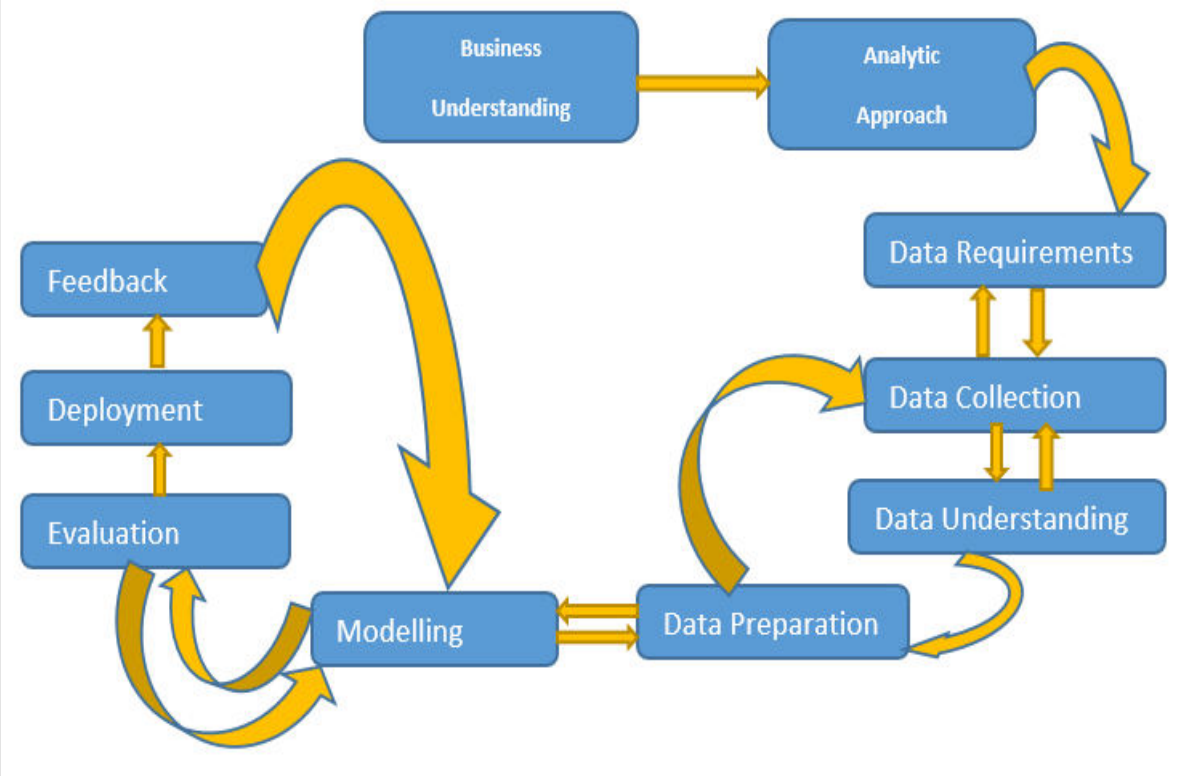
Musah Abdul-Mumuni

CONTENTS

- Introduction
- Resource Assessment – Potential Play types in Ghana(MAPS)
- Approach to funding request
- Possible sources of funding
- Way forward

INTRODUCTION & METHODOLOGY

- The decision to choose between North America´s two top financial capitals for the establishment of a popularly known fast food in Germany, Europe will be taken through the following Data Science Methodology



BUSINESS UNDERSTANDING

- A successful European based Doner (a very popular dish in Germany consisting of spiced lamb cooked on a spit and served in slices, typically with pitta bread) shop operator wishes to set up a similar shop in one of the top financial capitals in North America, New York or Toronto.
- To decide which city to setup the business in, Data would be collected online from sources such as Wikipedia with Foursquare API and with Python Jupyter the data is analysed to determine the existence of similar exotic type of businesses in the two Neighbourhoods

ANALYTICAL APPROACH

- Following the Data Science Methodology chart It starts with Business Understanding, where a definition of the goal of establishing a successful Doner Business in either New York or Toronto is defined. It is followed by Analytical Approach where the guidelines or patterns for achieving the goal is defined. In this case, Python Jupyter notebook is used together with Foursquare API for collecting Neighbourhood data online for the two cities to be analyzed.
- Data Analysis and Visualization part is made up of Data Requirement, collection, understanding and preparation stages are done iteratively. Questions are asked in the process and more data or similar better data are required, collected and prepared for analysis and create a model for solutions in setting can be found by manipulating the model. The model is then evaluated by checking other Neighbourhood venue within the model. Model evaluation goes hand-in-hand with model building as such, the modeling and evaluation stages are done iteratively. When satisfied of its performance it is deployed by establishing the business in the chosen city. Continuous feedback is fed into the model to achieve optimum functioning of the model.

DATA

- Data sources for Toronto and New York
- https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- https://cocl.us/new_york_dataset
- Initial data of Scarborough and Queens selected to represent both Toronto and New York respectively

	Borough	Neighborhood	Latitude	Longitude
0	Scarborough	Rouge,Malvern	43.806686	-79.194353
1	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497
2	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711
3	Scarborough	Woburn	43.770992	-79.216917
4	Scarborough	Cedarbrae	43.773136	-79.239476

	Borough	Neighborhood	Latitude	Longitude
0	Queens	Astoria	40.768509	-73.915654
1	Queens	Woodside	40.746349	-73.901842
2	Queens	Jackson Heights	40.751981	-73.882821
3	Queens	Elmhurst	40.744049	-73.881656
4	Queens	Howard Beach	40.654225	-73.838138

DATA ANALYSIS

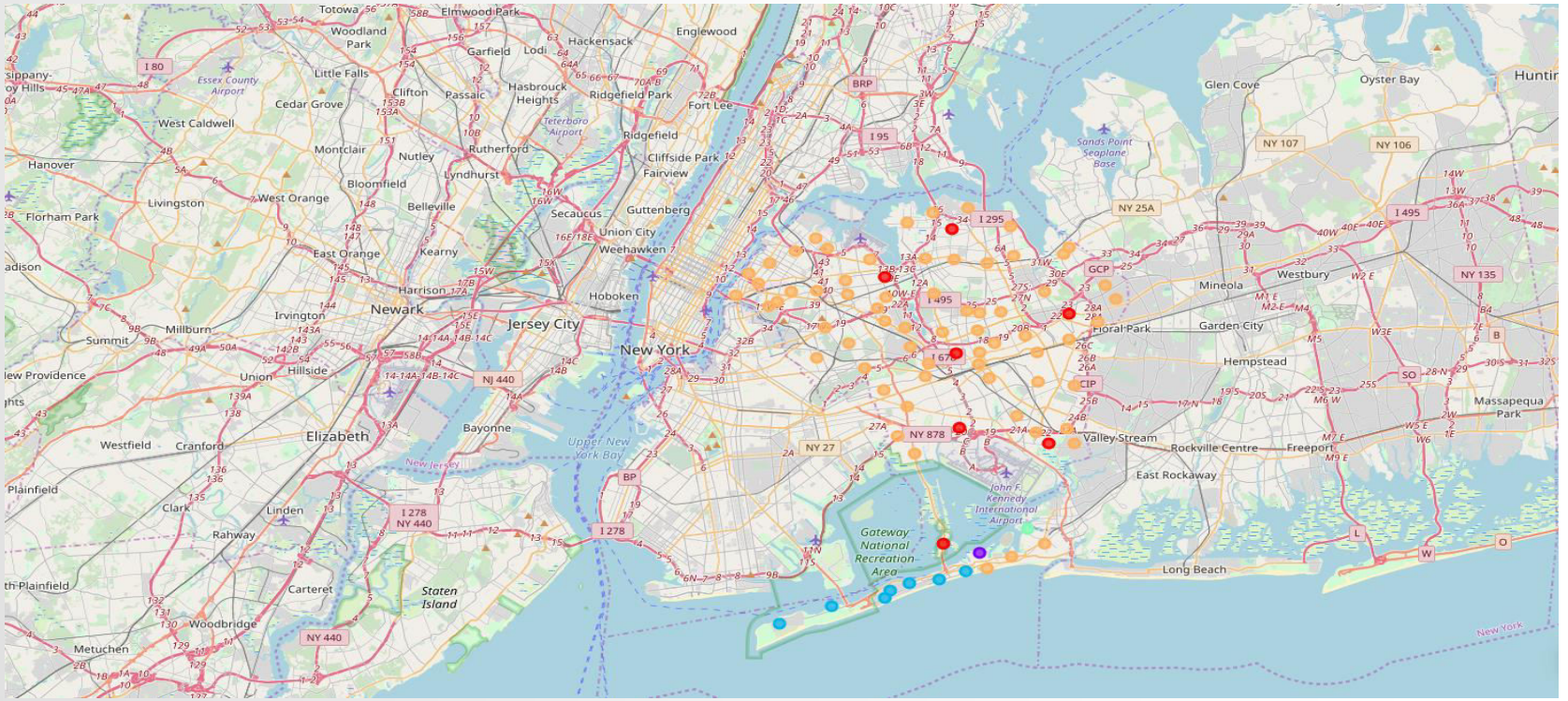
- A dataframe showing venue data around Astoria in New York (left) and Rouge, Malvern (right) in Toronto with their corresponding coordinates

	name	categories	lat	lng
0	Favela Grill	Brazilian Restaurant	40.767348	-73.917897
1	Titan Foods Inc.	Gourmet Shop	40.769198	-73.919253
2	CrossFit Queens	Gym	40.769404	-73.918977
3	Sitan Muay Thai	Martial Arts Dojo	40.766108	-73.913224
4	Al-sham Sweets and Pastries	Middle Eastern Restaurant	40.768077	-73.911561

	name	categories	lat	lng
0	Images Salon & Spa	Spa	43.802283	-79.198565
1	Staples Morningside	Paper / Office Supplies Store	43.800285	-79.196607
2	Caribbean Wave	Caribbean Restaurant	43.798558	-79.195777
3	Wendy's	Fast Food Restaurant	43.802008	-79.198080
4	Wendy's	Fast Food Restaurant	43.807448	-79.199056

MODELLING AND EVALUATION

- Map showing 5 venue clusters of Queens Borough in New York



FIRST CLUSTER DATA OF NEW YORK SHOWING VARIOUS EXOTIC RESTAURANTS AMONG THE TOP MOST COMMON VENUES

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Astoria	Bar	Middle Eastern Restaurant	Greek Restaurant	Hookah Bar	Seafood Restaurant	Mediterranean Restaurant	Bakery	Ice Cream Shop	Salon / Barbershop	Chinese Restaurant
1	Woodside	Grocery Store	Thai Restaurant	Bakery	Filipino Restaurant	Latin American Restaurant	American Restaurant	Donut Shop	Pizza Place	Pub	Bar
2	Jackson Heights	Latin American Restaurant	Peruvian Restaurant	Bakery	South American Restaurant	Mobile Phone Shop	Thai Restaurant	Mexican Restaurant	Diner	Spanish Restaurant	Clothing Store
3	Elmhurst	Thai Restaurant	Mexican Restaurant	Chinese Restaurant	South American Restaurant	Bubble Tea Shop	Vietnamese Restaurant	Bar	Sushi Restaurant	Malay Restaurant	Snack Place
4	Howard Beach	Italian Restaurant	Ice Cream Shop	Fast Food Restaurant	Sandwich Place	Bagel Shop	Pharmacy	Deli / Bodega	Jewelry Store	Gym	Breakfast Spot
5	Corona	Bakery	Mexican Restaurant	Ice Cream Shop	Pizza Place	Convenience Store	Deli / Bodega	Park	Restaurant	Donut Shop	Sandwich Place
6	Forest Hills	Gym / Fitness Center	Gym	Yoga Studio	Pizza Place	Convenience Store	Park	Thai Restaurant	Pharmacy	Video Game Store	Optical Shop
7	Kew Gardens	Chinese Restaurant	Donut Shop	Pet Store	Pharmacy	Cosmetics Shop	Bank	Pizza Place	Bar	Park	Indian Restaurant
8	Richmond Hill	Pizza Place	Bank	Latin American Restaurant	Lounge	Caribbean Restaurant	Bakery	Deli / Bodega	Pet Service	Diner	Discount Store
9	Flushing	Hotpot Restaurant	Chinese Restaurant	Korean Restaurant	Bubble Tea Shop	Construction & Landscaping	Asian Restaurant	Gym / Fitness Center	Bakery	Gym	Karaoke Bar
10	Long Island City	Coffee Shop	Hotel	Pizza Place	Bar	Café	Deli / Bodega	Mexican Restaurant	Gym / Fitness Center	Donut Shop	Chinese Restaurant
11	Sunnyside	Pizza Place	Italian Restaurant	South American Restaurant	Grocery Store	Bakery	Hotel	Discount Store	Chinese Restaurant	Coffee Shop	Taco Place
12	East Elmhurst	Donut Shop	Ice Cream Shop	Coffee Shop	Bus Station	Supermarket	Gas Station	Lake	Latin American Restaurant	Rental Car Location	Hotel Bar
13	Maspeth	Pizza Place	Diner	Grocery Store	Mobile Phone Shop	Bank	Ice Cream Shop	Bakery	Sandwich Place	Sushi Restaurant	Lounge
14	Ridgewood	Mexican Restaurant	Bank	Bakery	Pizza Place	Mobile Phone Shop	Greek Restaurant	Grocery Store	Restaurant	Korean Restaurant	Sushi Restaurant
15	Glendale	Arts & Crafts Store	Pizza Place	Brewery	Food & Drink Shop	Food	Flower Shop	Fish Market	Fish & Chip Shop	Filipino Restaurant	Electronics Store

DEPLOYMENT AND FEEDBACK

- When the model has been manipulated iteratively enough and it is found to meet the standard, the Doner shop is establish in the chosen area. Feed back as to the relevance of the model is continuously fed in it. This allows for optimization of the model to be used in the future by people of similar interest in establishing a similar business in a similar situation.

CONCLUSION

- The descriptive model created shows the location of exotic restaurants in North America`s two financial capitals New York and Toronto and the top most common venues.
- From the clusters we can see that exotic restaurants, such as Chinese, Thai, and Greek make up the top most visited places. It is also clear that despite Toronto having its share of exotic restaurants, New York with its large population and many Neighbourhoods offers more exotic restaurants.
- The decision to choose to establish a Doner business in New York over Toronto can be established. However, Toronto has the advantage of offering less competition but a population that seems very interested in exotic restaurants.

FUTURE DIRECTIONS

A more thorough analysis could be made by comparing income, spending on food especially exotic, fast food, etc.

Also comparison of North York (Toronto) with Queens (New York) or Manhattan in New York with North York can be made.

The distance between the venues can also be computed to find a place with the highest number of customers or visits to make the location of business establishment optimum.