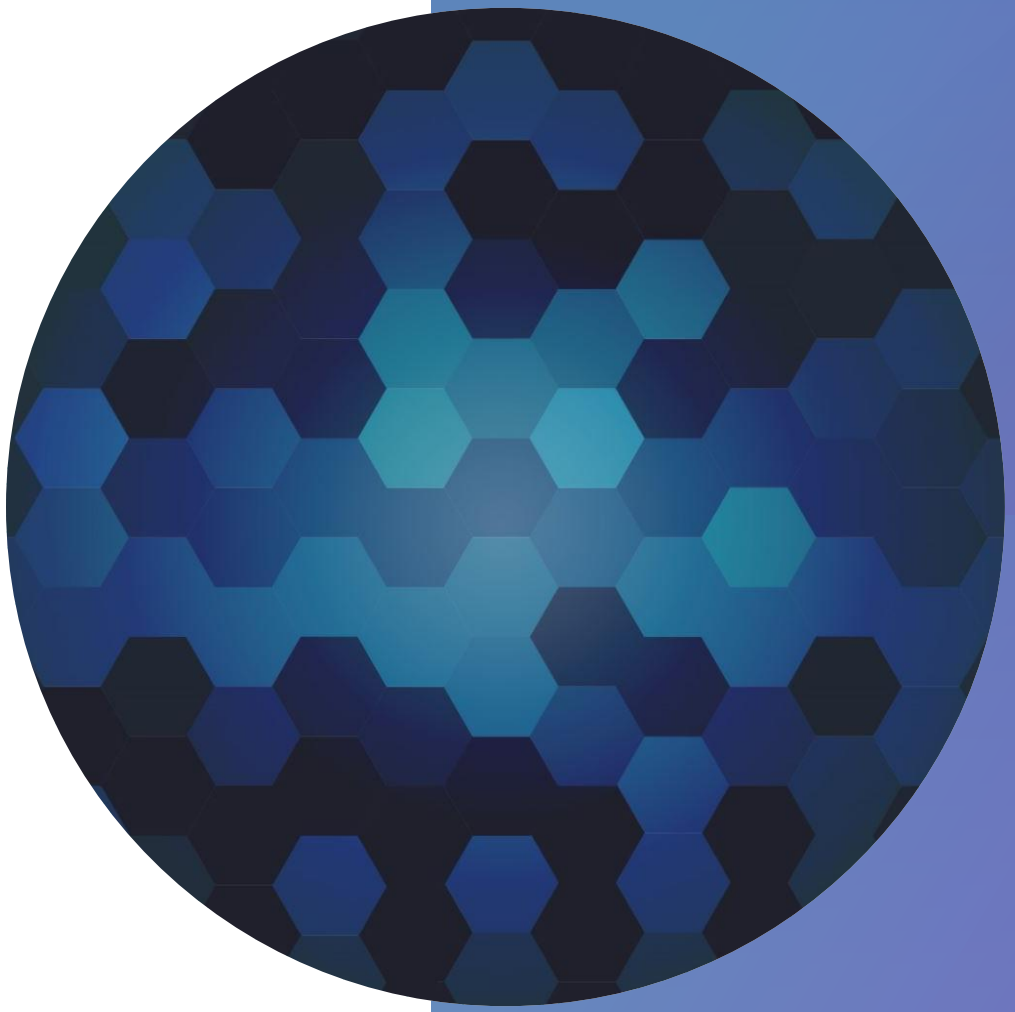
CAPSTONE PROJECT BATTLE OF NEIGHBORHOODS



CHOOSING LOCATION FOR IT TRAINING BUSINESS

SUMMARY

- INTRODUCTION: WHAT PROJECT ADDRESSES AND WHY. THE PROBLEM AND WHO IT WILL HELP
- DATA: DESCRIPTION AND SOURCE OF DATA.
- METHODOLOGY: HOW I AM ADDRESSING THE PROBLEM.
- RESULTS: DISCUSS THE ANALYSIS RESULTS.
- OBSERVATIONS NOTED AND ANY RECOMMENDATIONS YOU CAN MAKE BASED ON THE RESULTS.
- CONCLUSION

INTRODUCTION

- A TECH ENTREPRENEUR IS LOOKING TO START AN IT TRAINING BUSINESS IN NIGERIA AND NEEDS TO DECIDE BETWEEN TWO LOCATIONS; OYO STATE OR RIVERS STATE
- THIS PROJECT WOULD HELP ANYONE WHO IS LOOKING FOR HOW TO CHOOSE THE BEST LOCATION TO START AN IT TECHNICAL TRAINING SCHOOL.

DATA DESCRIPTION AND SOURCE

- IT WAS VERY CHALLENGING GATHERING DATA ON NIGERIA
- I RECRUITED THE HELP OF A FRIEND IN NIGERIA- JOHN AJAYI. HE LOCATED DATA OF NIGERIA'S 2016 CENSUS – HTTPS://DATA.WORLD/OCHA-NIGERIA/A7C3DE5E-FF27-4746-99CD-05F2AD9B1066
- OYO DATA

State regions	Age 4 to 5 Pop	Age 6 to 13 Pop	Age 14 to 17 Pop	Age 18 to 22 Pop	Age 23+ Pop	Total Pop	Latitude	Longitude
Oyo North	77,552	558,224	358,344	335,667	1,329,787	2,659,574	8.525903	3.616589
Oyo South	172,518	265,887	345,668	624,567	1,000,654	2,567,254	7.374447	3.271742
Oyo Central	200,584	156,716	338,547	613,453	1,257,954	2,409,294	7.968394	3.571628

- RIVERS DATA

State regions	Age 4 to 5 Pop	Age 6 to 13 Pop	Age 14 to 17 Pop	Age 18 to 22 Pop	Age 23+ Population	Tota Pop	Latitude	Longitude
Rivers East	215,760	225,884	335,486	785,774	1,228,554	2,791,458	4.889828	7.107192
Rivers South East	205,901	265,524	300,258	424,125	602,655	1,798,463	4.645197	7.441428
Rivers West	356,254	385,263	412,549	658,416	679,009	2,491,491	4.831936	6.592625

DATA DESCRIPTION AND SOURCE

- AS THE PROJECT PROGRESSED, I REALIZED THAT FOURSQUARE DOES NOT HAVE DATA ON MY LOCATIONS OF CHOICE. USING FOURSQUARE IS REQUIRED FOR THIS PROJECT
- I SWITCHED TO NEW YORK STATE AND ONTARIO PROVINCE.
- NEW YORK'S POPULATION FROM HTTPS://WWW.NEWYORK-DEMOGRAPHICS.COM/CITIES_BY_POPULATION AND ONTARIO'S POPULATION DATA FROM HTTPS://WWW.CITYPOPULATION.DE/EN/CANADA/CITIES/ONTARIO/
- THEN HIGH SCHOOL DATA ON BOTH LOCATIONS FROM FOURSQUARE

METHODOLOGY

- K-MEANS CLUSTERING AND GROUPBY WERE USED TO ANALYZE THE DATA.

- BASED ON THE TABLE OF DATA GATHERED, A CITY WITH THE HIGHEST POPULATION WAS SELECTED FROM EACH STATE. THEN THE DATA OF SCHOOLS IN THOSE CITIES WERE GATHERED USING FOURSQUARE

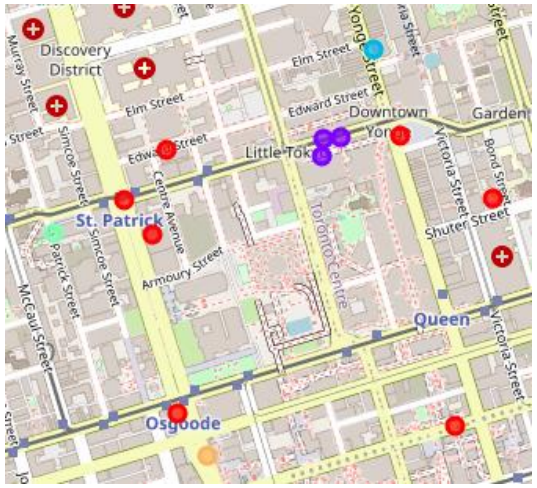
METHODOLOGY

- GROUPBY AND K-MEANS OF TORONTO

- GROUPBY AND K-MEANS OF NEW YORK CITY

```
Torontodataframe_filtered.groupby('postalCode').count()
```

postalCode	name	categories	city	lat	lng
L3R 0M1	1	1	1	1	1
M2J 1Y3	1	1	1	1	1
M4P 1A9	1	1	1	1	1
M5B 2B9	1	1	1	1	1
M5B 2K1	1	1	1	1	1
M5G 1G6	1	1	1	1	1
M5G 2C3	1	1	1	1	1
M5G 2C5	2	2	2	2	2
M5H 3E5	1	1	1	1	1
M5M 3G5	1	1	1	1	1
M5T 2X7	1	1	1	1	1
N1H 7V2	1	1	1	1	1




```
NYdataframe_filtered.groupby('postalCode').count()
```

postalCode	name	categories	city	lat	lng	state
10007	10	10	10	10	10	10
10013	8	8	8	8	8	8
10038	15	15	15	15	15	15
10271	1	1	1	1	1	1
10279	1	1	1	1	1	1
10452	1	1	1	1	1	1
11201	1	1	1	1	1	1

RESULT & CONCLUSION

- FOR CANADA, IN THE ANALYSIS DONE, WE CAN SEE THAT THE ZIP CODE M5G 2C5 HAS MORE HIGH SCHOOLS THAN THE OTHER ZIP CODES

- FOR NEW YORK, IN THE ANALYSIS DONE, WE CAN SEE THAT THE ZIP CODE 10038 HAS MORE HIGH SCHOOLS THAN THE OTHER ZIP CODES. AND FOLLOWING CLOSE TO IT IS 10007.

RESULT & CONCLUSION

- BASED ON THE COMBINATION OF HIGHEST POPULATION AND K-MEAN, I OBSERVE THAT ZIP CODE 10038 IN NEW YORK HAS MORE CLUSTER OF HIGH SCHOOLS AMONG THE SELECTED CITIES
- I WOULD HIGHLY RECOMMEND STARTING AN IT TRAINING BUSINESS IN NEW YORK CITY
- IF I WAS ABLE TO GATHER DATA FROM FOURSQUARE FOR THE LOCATIONS I HAD ORIGINALLY CHOSEN FOR THE PROJECT, THIS IS THE SAME PROCESS I WOULD HAVE FOLLOWED TO ARRIVE AT A LOCATION RECOMMENDATION.