INSIDE

This is a CAPSTONE PROJECT done for the DATA SCIENCE SPECIALIZATION. The presentation will present the work done under below headings

- **PROJECT OVERVIEW:**
- **DATA:**
- **METHODOLOGY:**
- **EXPLORATORY VIEW:**
- **DATA GATHERED:**
- **RESULTS AND CONCLUSION:**
- **ADDITIONAL NOTES:**

PROJECT OVERVIEW

SUBJECT: Find the best city in and around Cincinnati, OH to launch a new Asian Grocery Supermarket.

CLIENT AD BACKGROUND:
A business chain that has Asian grocery markets across various states within USA. They want to expand their business in mid-west and are contemplating Cincinnati neighborhood as their destination. We will help them in researching and analyzing various neighborhoods in the region and find the best location to launch their new store.

OBJECTIVE:
Research the main cities within the Greater Cincinnati Area of Ohio (OH) and present their potential to be a good fit for launching a new Asian Grocery Supermarket.
Explore the cities in terms of ***Restaurants*** and ***Schools*** to get an idea of Asian population in these cities.

DATA

We will be using the below datasets –

Cities: For cities:
https://en.wikipedia.org/wiki/Cincinnati_metropolitan_area

Locations: for locations data data of cities and other venues we will use geolocator

Restaurants: For restaurants analysis, we will use the FOURSQUARE database and APIs.
https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&query={}&ll={},{}&radius={}&limit={}

Schools and Schools Districts:

i) Fall Enrollment (Headcount) - October 2018 Public Districts and Buildings
http://education.ohio.gov/Topics/Data/Frequently-Requested-Data/Enrollment-Data

ii) Performance Index Score Rankings
http://education.ohio.gov/lists_and_rankings

METHODOLOGY

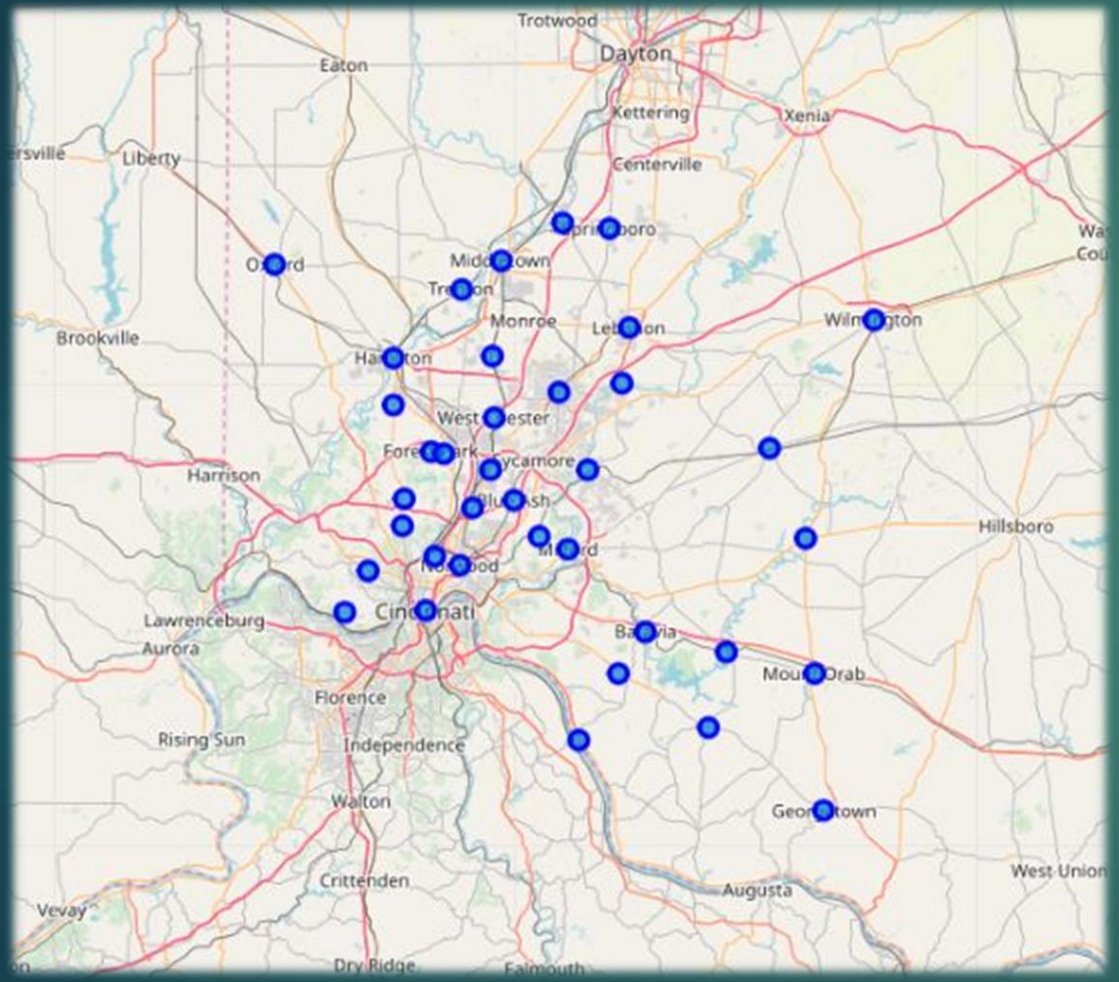
We used the below information about each city to represent its suitability for the new store

- density of Indian/Chinese/other Asian restaurants within 5 miles of the cities
- population of the cities
- number of Pre/Elem/Med/High schools within 5 miles of those cities
- grade/rating of the main school district serving each city
- Total enrollment in the school district this current year and possibly # of Asian students enrolled

We gathered all the above information individually for each city for each city. We then did some initial exploratory analysis, did some cleansing, consolidation, some transformations and combined all the usable data for each city in a data frame with one row for each city.

We then assigned score to each city for each of the criteria and took the final score for each city. Then Final Index Score was used to rank the cities for its suitability. To further assist the decision making we also visually represented the final score of the cities on the map along with their location.

EXPLORATORY ANALYSIS



[6]:
```
gcincities[['Latitude','Longitude']] = df[['Latitude','Longitude']]
gcincities.head()
```

[6]:

	cityname	population	LEAId	Latitude	Longitude
0	Cincinnati, Ohio	298,843	43752	39.101454	-84.512460
1	Hamilton, Butler County, OH	62,447	44107	39.399507	-84.561343
2	Liberty Township, Butler County, Ohio,37	259"	46110	39.401558	-84.409678
3	Middletown, Butler County, Ohio	48,694	44404	39.515058	-84.398276
4	Fairfield, Butler, Ohio	42,510	46102	39.345895	-84.560503

We started with exploring **37 Cities**, they are plotted on the map for initial view

GATHERED DATA -RESTAURANTS

```
[45]: gcincities_venues_data.groupby('Venue Category2').count()
```

[45]:

Venue Category2	City	City Latitude	City Longitude	Venue Name	Venue Latitude	Venue Longitude
Chinese Restaurant	485	485	485	485	485	485
Indian Restaurant	217	217	217	217	217	217
Other Asian Restaurant	47	47	47	47	47	47

We were able to find **907 Restaurants** around those cities under our categories.

```
[53]: gcincities_venues_data1.groupby('City').count()
```

[53]:

City	City Latitude	City Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Amelia, Clermont County, Ohio	7	7	7	7	7	7
Blue Ash, Hamilton County, Ohio	39	39	39	39	39	39
Cincinnati, Ohio	57	57	57	57	57	57
College Hill, Hamilton County, Ohio	27	27	27	27	27	27
Delhi Township, Hamilton County, Ohio	23	23	23	23	23	23
Fairfield, Butler, Ohio	14	14	14	14	14	14
Forest Park, Hamilton County, Ohio	28	28	28	28	28	28
Franklin, Warren County, Ohio	7	7	7	7	7	7
Hamilton, Butler County, OH	10	10	10	10	10	10
Indian Hill, Ohio	30	30	30	30	30	30
Liberty Township, Butler County, Ohio,37	21	21	21	21	21	21
Loveland, Hamilton County, Ohio	25	25	25	25	25	25

GATHERED DATA - SCHOOLS

```
[73]: gcincities_schools_data.groupby('Venue Category2').count()
```

[73]:

Venue Category2	City	City Latitude	City Longitude	Venue Name	Venue Latitude	Venue Longitude	Venue Category
Elementary School	483	483	483	483	483	483	483
High School	365	365	365	365	365	365	365
Middle School	120	120	120	120	120	120	120
Preschool	449	449	449	449	449	449	449

```
[81]: schools_dist.head()
```

[81]:

	LEAIRN	LEA Name	SD_PI	SD_Grade	ENRL_PRE	ENRL_ELEM	ENRL_MID	ENRL_HIGH	STUDENT_ASIAN
0	442	Manchester Local	83.898	D	20	384	202	250	10
1	43489	Akron City	67.03	D	553	9816	4534	6373	1807
2	43497	Alliance City	73.179	D	240	1201	648	954	10
3	43505	Ashland City	90.805	C	134	1433	751	872	23
4	43513	Ashtabula Area City	69.595	D	174	1521	779	945	10

We looked at
1407 Schools and
evaluated
21 School Districts
in and around the cities

GATHERED DATA - CONSOLIDATED

```
[92]: gcincities_grouped_add2.head()
```

[92]:

City	REST_Chinese	SCH_Elementary	SCH_High	REST_Indian	SCH_Middle	REST_Asian	SCH_Pre	CITY_Population	SD_Name	SD_PI	SD_Grade	ENRL_PRE	ENRL_ELEM	ENRL_MID	ENRL_HIGH	SD_STUDENT_ASIAN
Amelia, Clermont County, Ohio	5	7	3	2	3	0	9	4,801	West Clermont Local	87.406	C	217	3799	1934	2301	167
Blue Ash, Hamilton County, Ohio	20	23	22	14	6	5	27	12,114	Sycamore Community City	103.694	B	91	2501	1240	1640	890
Cincinnati, Ohio	37	34	34	17	6	3	23	298,843	Cincinnati City	72.477	D	1597	16158	7566	9454	498
College Hill, Hamilton County, Ohio	18	25	18	9	5	0	13	14,133	North College Hill City	59.359	F	19	761	437	393	10
Delhi Township, Hamilton County, Ohio	17	14	14	5	4	1	13	29,510	Oak Hills Local	96.997	B	143	3364	1894	2232	90

We consolidated all the data in a single frame with one row for each city, along with all the features/information that we are using to evaluate the cities

GATHERED DATA – INDEX / RANKED

```
[100]: final_data_rep.loc[:,'FINAL_IND'] = final_data_rep.loc[:,'REST_IND'] + final_data_rep.loc[:,'SCH_IND'] + final_data_rep.loc[:,'AST_IND'] + final_data_rep.loc[:,'ENR_IND'] + final_data_rep.loc[:,'SD_IND']
       final_data_rep.sort_values('FINAL_IND',ascending=False)
```

[100]:

	City	CITY_Population	SD_Grade	SD_STUDENT_ASIAN	REST_Tot	SCH_Tot	ENRL_Tot	REST_IND	SCH_IND	SD_IND	AST_IND	ENR_IND	FINAL_IND
12	Mason, Warren County, Ohio	30,712	B	2820	35	58	10396	6	4	3	4	4	21
2	Cincinnati, Ohio	298,843	D	498	57	97	34775	8	4	1	3	4	20
1	Blue Ash, Hamilton County, Ohio	12,114	B	890	39	78	5472	6	4	3	4	3	20
26	Westwood, Hamilton County, Ohio	29,950	D	498	37	52	34775	6	4	1	3	4	18
10	Liberty Township, Butler County, Ohio,37	259"	B	1188	21	40	16262	4	3	3	4	4	18
22	Springdale, Hamilton County, Ohio	11,223	D	209	37	63	5955	6	4	1	3	3	17
19	Sharonville, Hamilton County, Ohio	13,560	D	209	48	71	5955	6	4	1	3	3	17
9	Indian Hill, Ohio	5,785	A	215	30	58	2030	4	4	4	3	1	16
5	Fairfield, Butler, Ohio	42,510	C	561	14	44	9887	4	3	1	4	3	15
23	St. Bernard, Hamilton County, Ohio	4,368	D	10	60	105	905	8	4	1	0	1	14
16	Norwood, Hamilton County, Ohio	19,207	C	10	60	100	1906	8	4	1	0	1	14
6	Forest Park, Hamilton County, Ohio	18,720	D	241	28	66	3758	4	4	1	3	1	13
4	Delhi Township, Hamilton County, Ohio	29,510	B	90	23	45	7633	4	3	3	0	3	13
14	Milford, Clermont County, Ohio	6,709	B	104	20	33	6614	4	2	3	1	3	13
25	West Chester, Butler County, Ohio	64000	C	10	47	62	1089	6	4	1	0	1	12
11	Loveland, Hamilton County, Ohio	12,081	B	103	25	43	4592	4	3	3	1	1	12
18	Reading, Hamilton County, Ohio	10,385	C	17	49	85	1535	6	4	1	0	1	12
21	Springboro, Warren County, Ohio	17,409	B	182	5	30	6153	2	2	3	1	3	11
20	South Lebanon, Warren County, Ohio	4,115	B	226	9	28	4784	2	2	3	3	1	11
3	College Hill, Hamilton County, Ohio	14,133	F	10	27	61	1610	4	4	1	0	1	10

Finally, we scored each city for each criteria (REST_IND, SCH_IND etc…) and took the final Index score (FINAL_IND). Used the final score to rank the cities, the higher the city the more potential for the new store.

RESULTS

So we evaluated **37 Cities.**
907 Restaurants.
1407 Schools
21 School Districts

We were able to filter our 10 cities based on initial exploration, and finally represent 27 cities with all the detailed information and their final score.

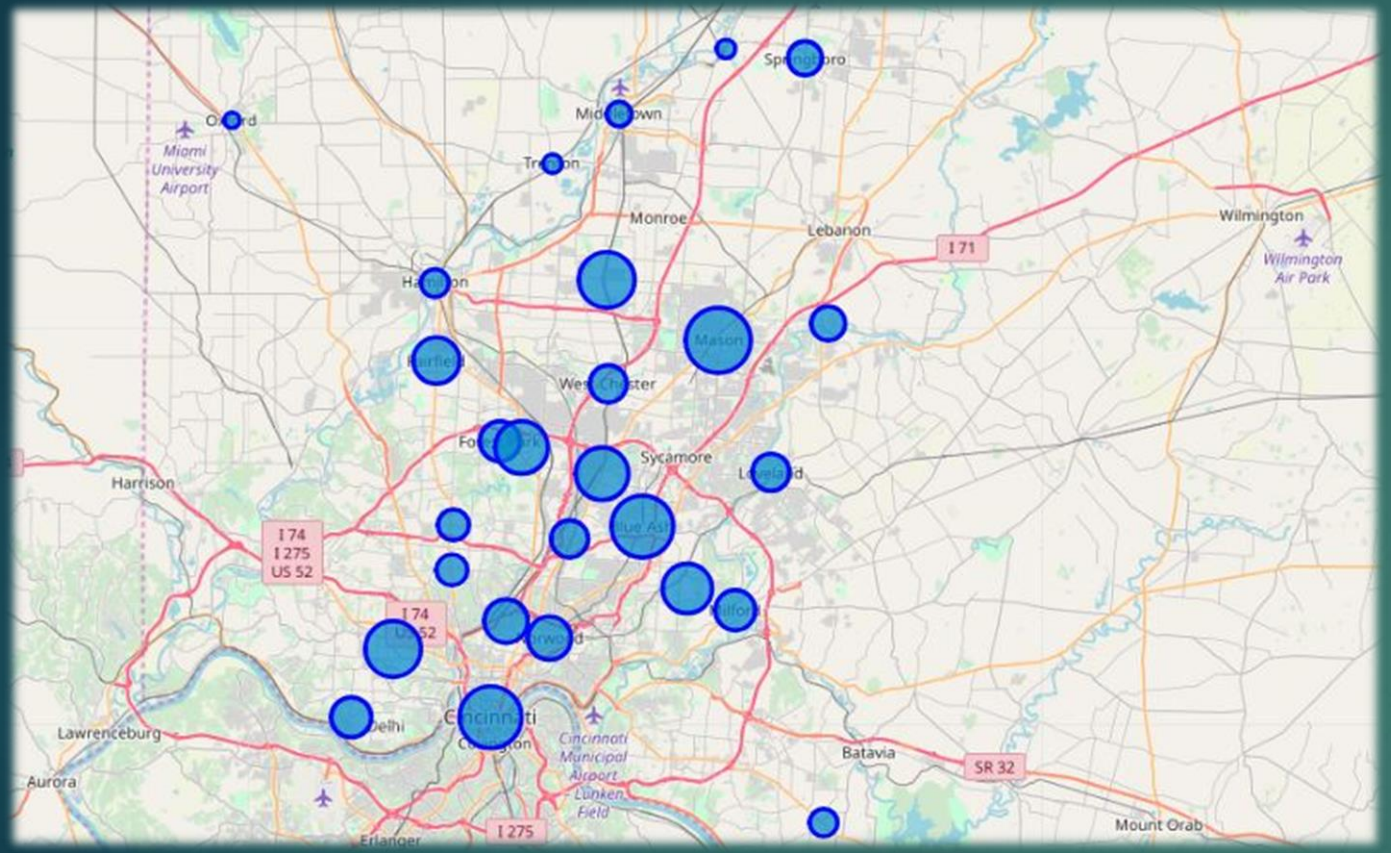
	City	CITY_Population	SD_Grade	SD_STUDENT_ASIAN	REST_Tot	SCH_Tot	ENRL_Tot	REST_IND	SCH_IND	SD_IND	AST_IND	ENR_IND	FINAL_IND
12	Mason, Warren County, Ohio	30,712	B	2820	35	58	10396	6	4	3	4	4	21
2	Cincinnati, Ohio	298,843	D	498	57	97	34775	8	4	1	3	4	20
1	Blue Ash, Hamilton County, Ohio	12,114	B	890	39	78	5472	6	4	3	4	3	20
26	Westwood, Hamilton County, Ohio	29,950	D	498	37	52	34775	6	4	1	3	4	18
10	Liberty Township, Butler County, Ohio,37	259"	B	1188	21	40	16262	4	3	3	4	4	18
22	Springdale, Hamilton County, Ohio	11,223	D	209	37	63	5955	6	4	1	3	3	17
19	Sharonville, Hamilton County, Ohio	13,560	D	209	48	71	5955	6	4	1	3	3	17

Cities like Mason, Cincinnati, Blue Ash, Westwood, Liberty Township scored high and ranked up.

8	Hamilton, Butler County, OH	62,447	D	56	10	35	10067	2	2	1	0	4	9
13	Middletown, Butler County, Ohio	48,694	D	15	6	24	6338	2	2	1	0	3	8
24	Trenton, Butler County, Ohio	11,869	C	20	6	22	3739	2	2	1	0	1	6
7	Franklin, Warren County, Ohio	27,294	C	18	7	30	2777	2	2	1	0	1	6
17	Oxford, Butler County, Ohio	21,943	C	65	5	8	2997	2	1	1	0	1	5

While cities like Hamilton, Oxford, Trenton, Franklin scored very low and ranked at the bottom

RESULTS AND CONCLUSION



We plotted the final list of cities on map with their circle size according to their final index.

It is easier to see that the cities like Mason, OH and Blue Ash, OH are most potential candidates for the new market because of their higher circle and also proximity to other cities that has good Asian population.

Our client finally could use this information along with few other off line checks (like competition, their reviews, real state rates etc.) to decide that they would open their new Asian Super Market in **Mason, OH.**

We helped out client choose the best fit city to launch the new Asian Grocery Supermarket. We achieved this by ***collecting, cleansing, preparing, analyzing, summarizing, and presenting*** the data in the best way so they can visualize the information in easiest possible way.

ADDITIONAL NOTES:

The work done here shows, in a rather simple way, how data science can be used to make such important decisions.

This work can be further extended/modified to include more features and criteria like *Real State prices*, *# of Competitors* etc. to make it more exhaustive. Further the work can be extended using various other data science and machine learning technologies like *Classification, Clustering* to service various other clients and their specific asks. For example:

i) An individual customer who wants to relocate from some other city of US to greater Cincinnati can use us to know which city is best to rent/buy a home for the family. We can analyze his original city and find a city with similar characteristics from the list that we gathered within Cincinnati area using clustering techniques.

ii) Any other new business ventures that wants to come to Cincinnati can use this information for demographic analysis of cities. like to host Korean Movies in theaters, open Italian Restaurant etc.

And yes, this entire thing can be easily replicated for any other city, metropolitan area across the globe with little tweaking!

THANK YOU!!