# A NEW BUSINESS IDEA

You want to open a new Ice Cream Shop in Chicago, but have a few questions in order to open this shop. You are hoping that data will help you solve them.

BUSINESS PROBLEM I

- You place a high importance on the safety of your customers, so you want to make sure you open up your ice cream shop in an area where you know your customers will be safe

- However, the crime that most often impacts a business, is theft. There have already been 27321 cases of theft in 2019 alone. In 2018, the highest crime was also theft, with a total 65200 cases

Number of theft cases in Chicago between 2015-2019
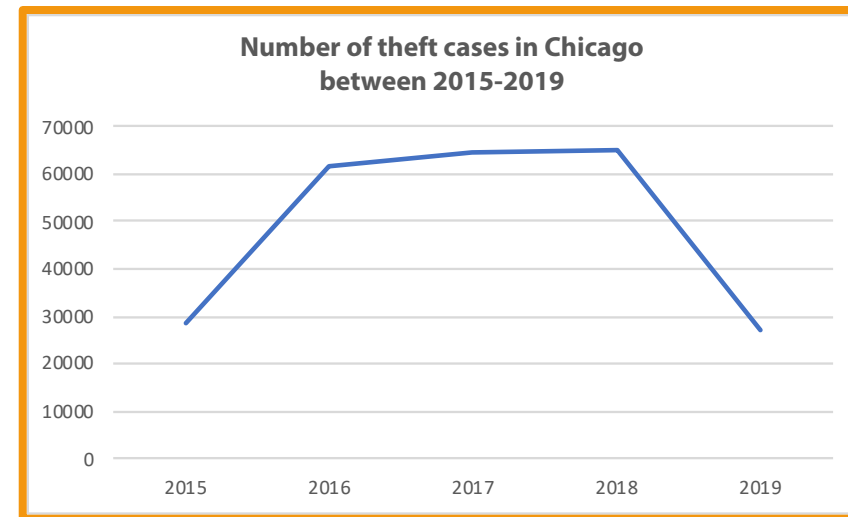
BUSINESS PROBLEM III

TARGET MARKET

- You know your target market will be families with young children. However, you do not know where you will target your marketing efforts

- According to Chicago Health Atlas, there were ±630,000 people in between the ages of 0 – 19 in Chicago in 2017. Maybe you can segment this target group into a smaller age group – what data would we need for that?

COMPETITORS

- Lastly, to determine of the five communities where you will locate your ice cream shop, the current level of supply in the community will be examined. In other words, how many ice cream shop competitors are there already in the chosen communities?

DATA

Six data sources were used.

For further information on each data source, please refer to the report:

- Governmental dataset called 'CHICAGO_PUBLIC_SCHOOLS'

- Geographical coordinates using Folium

- Foursquare data

- A governmental dataset called 'CENSURDATANEW'

- A governmental dataset called 'TOTALPOPULATION'

- A governmental dataset called 'CHILDOBESITY'

PREPARATION I

- Different datasets were imported into the IBM DB2 database in order to be extracted using the sql load extension. Multiple libraries also needed to be imported.

- Statistical tests were run in order to understand the data. For example, the following describes the statistics of the SAFETY_SCORE. We can see above that the minimum safety score is 1 and the max is 99. We can also see other statistics, like the mean score being 49.5 and that 75% of all safety scores are below 61.

```
count    513.000000
mean      49.504873
std       20.110837
min        1.000000
25%       35.000000
50%       48.000000
75%       61.000000
max       99.000000
Name: SAFETY_SCORE, dtype: float64
```

PREPARATION II

- Data needed to be cleaned in order to retrieve the data frame in the form we want.

- This included accessing only specific columns, sorting, setting and resetting indexes and creating new data frames.

- For example, as we wanted to focus on those with the highest safety score, scores are sorted in descending order in order to determine the top 5 safest communities.

	COMMUNITY_AREA_NAME	ZIP_Code	SAFETY_SCORE	Latitude	Longitude
24	FOREST GLEN	60646.000000	99.000000	41.999368	-87.762061
41	LOOP	60605.000000	92.000000	41.874419	-87.627755
46	MOUNT GREENWOOD	60655.000000	86.500000	41.692870	-87.706007
51	NORTH CENTER	60623.571429	85.166667	41.944746	-87.684155
38	LINCOLN PARK	60614.000000	81.833333	41.921793	-87.649618

ANALYSIS & RESULTS II

- In order to determine whether there are any geographical outliers, Folium will be used in order to plot the 5 communities with their coordinates onto the map.

- One can clearly see in the map that there is 1 outlier, namely Mount Greenwood with zipcode 60655.

- This will be discarded from the options, due to its large distance from the other possible communities.
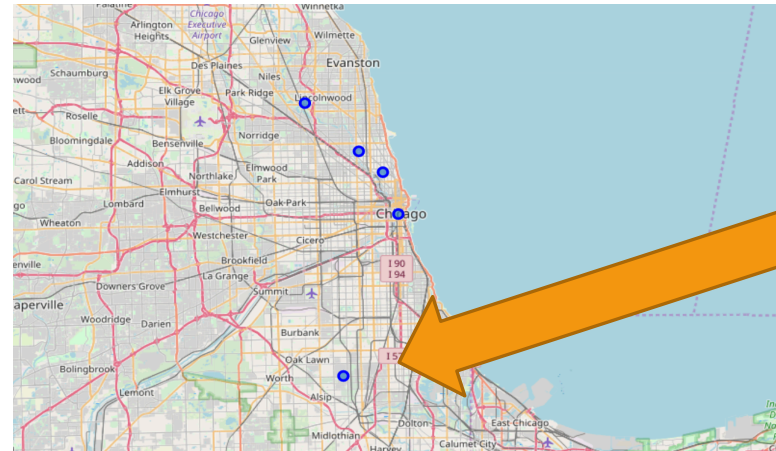
ANALYSIS & DISCUSSION II

- This analysis focused on whether there already is a supply of ice cream shops in each zipcode. The number of competitors in the surroundings will be examined.

- This step required creating a function that extracts the category of each venue, finds the common venues in each neighborhood and then converts this into a readable dataframe file.

- There is currently no common ice cream shop in 60623.

	Zipcode	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Comm Ven
0	60605	American Restaurant	Pub	Hostel	Indian Restaurant	Liquor Store	Music Ven
1	60614	Breakfast Spot	Yoga Studio	Pizza Place	French Restaurant	Japanese Restaurant	Cupcake Sh
2	60623	Gym / Fitness Center	Mediterranean Restaurant	Beer Garden	French Restaurant	Breakfast Spot	Wine Sh
3	60646	Sandwich Place	Ice Cream Shop	Salon / Barbershop	Italian Restaurant	Diner	Coffee Sh
4	60655	Mexican Restaurant	Gift Shop	Fast Food Restaurant	Donut Shop	Comic Shop	Convenience Sto

ANALYSIS & DISCUSSION III

- Further analysis into 60623 is needed to support the decision.

- The first analysis looks at PER_CAPITA_INCOME vs ZIPCODE.

- As there is a correlation between PER_CAPITA_INCOME vs the safety of a community, a high income should be correlated with a higher community safety.

- Thus, if the zipcode appears in the following graph, we are able to state that it is indeed in a safer community and is appropriate for further consideration.

Income per Community
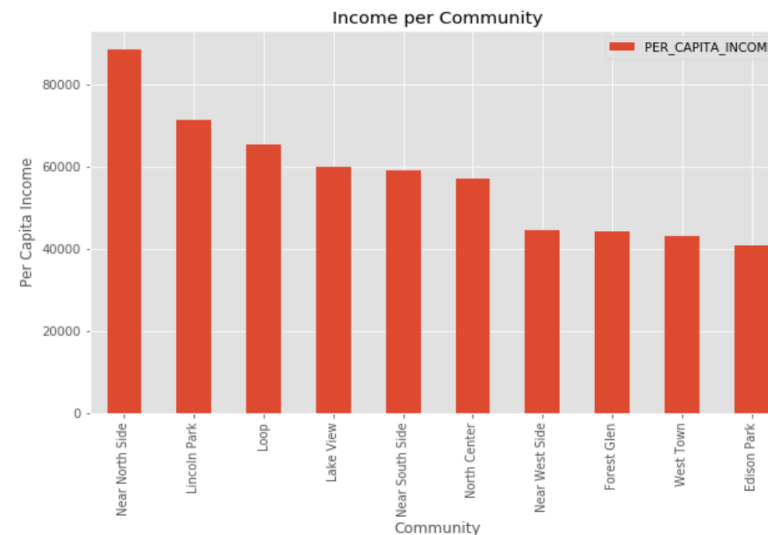
ANALYSIS & DISCUSSION IV

- In order to further support the choice of North Center, the relationship between child obesity and community (zipcode) will be analyzed.

- In this dataset, there are 77 communities.

- The following data frame presents the 10 communities with the lowest child obesity score.

- North Center (60623) is in the top 10 with lowest obesity rates.

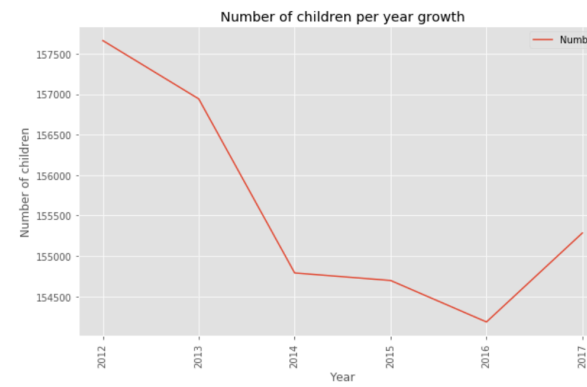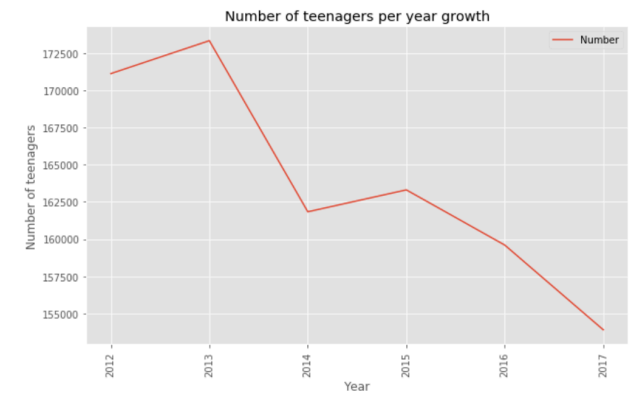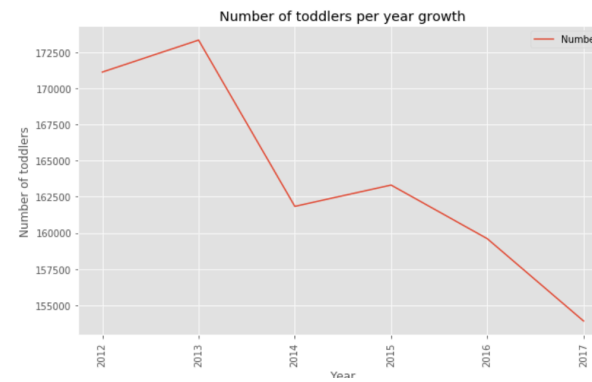|	Indicator	Year	Geography	Community	WeightPercent
66	Child obesity	2012-2013	Community Area	Lincoln Park	11.5
3	Child obesity	2012-2013	Community Area	Forest Glen	13.2
55	Child obesity	2012-2013	Community Area	Lakeview	13.3
33	Child obesity	2012-2013	Community Area	Lincoln Square	14.1
76	Child obesity	2012-2013	Community Area	Edison Park	14.4
1	Child obesity	2012-2013	Community Area	Norwood Park	14.4
27	Child obesity	2012-2013	Community Area	Armour Square	14.8
73	Child obesity	2012-2013	Community Area	O'Hare	15.4
25	Child obesity	2012-2013	Community Area	Loop	16.0
44	Child obesity	2012-2013	Community Area	North Center	17.4

ANALYSIS & DISCUSSION V

- Another business decision to be determined was how to segment the target market into smaller groups that would be easier to target with marketing efforts.

- We would want to analyze the growth of the number of children, per segment, per year.

- The categories are:

Toddlers	5-9
Children	10-14
Teenagers	15-19

ANALYSIS & DISCUSSION V

- The resulting plotted graphs are:

Number of toddlers per year growth

Number of teenagers per year growth

Number of children per year growth

RESULTS I

- Firstly, it was determined that the communities with the highest safety score are:

Community Name	Zipcode	Score
Forest Glen	60646	99
Loop	60605	92
Mount Greenwood	60655	87
North Center	60623	85
Lincoln Park	60614	82

- After geographically analyzing the location of all communities, you immediately notice that the community with zipcode 60655 is an outlier from the others. Therefore, we will remove this zipcode from the possibilities.
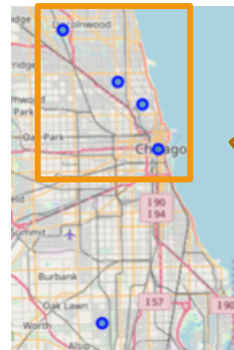
RESULTS II

- Moreover, by looking at the most common venues of the remaining 4 communities, it was found that:

> 60605's 6th most common venue is an ice cream shop
> 60614's 7th most common venue is an ice cream shop
> **60623 does not have an ice cream shop**
> 60646's 2nd most common venue is an ice cream shop

- There is sufficient supply in the zipcodes 60605, 60614 and 60646. However, it would be interesting to take a further look into 60623, as there is currently no common ice cream shop.

- When looking at the income_per_capita, each of the four communities is part of the top 6 highest income_per_capita communities:

> 60605: 3rd highest per capita income
> 60614: 2nd highest per capita income
> **60623: 6th highest per capita income**
> 60646: 8th highest per capita income

- We can safely assume that the community chosen, North Center with zipcode 60623, fulfills the safety criterion and thus is an appropriate choice.

RESULTS III

- Each of the four communities is in the top 10 lowest child obesity rate communities, of the total 77 communities:

  60605: 9th lowest in child obesity rate
 60614: 1st lowest in child obesity rate
  60623: 10th lowest in child obesity rate
  60646: 2nd lowest in child obesity rate

- North Center is in the top 10 with lowest obesity rates. This data can both show us a threat, or opportunity with opening an ice cream shop in North Center.

- It may be a threat, as the community is very health-conscious with its resulting low child obesity score, meaning that parents do not often take their children to an "unhealthy place".

- On the other hand, it can be also seen as an opportunity, as we are now aware of their health-conscious behavior and thus can include "healthier" options.

RESULTS IV

- Lastly, the target analysis graph demonstrated a general downward trend in the number of toddlers (5-9) and teenagers (15-19) between 2012 and 2017.

- On the other hand, although there was an initial downward trend of number of children over the years, it is slowly increasing since 2016.

- Therefore, the business conclusion we can make from this data is to focus on families with children between 10-14 years old in order to maximize our customer reach.

Number of children per year growth
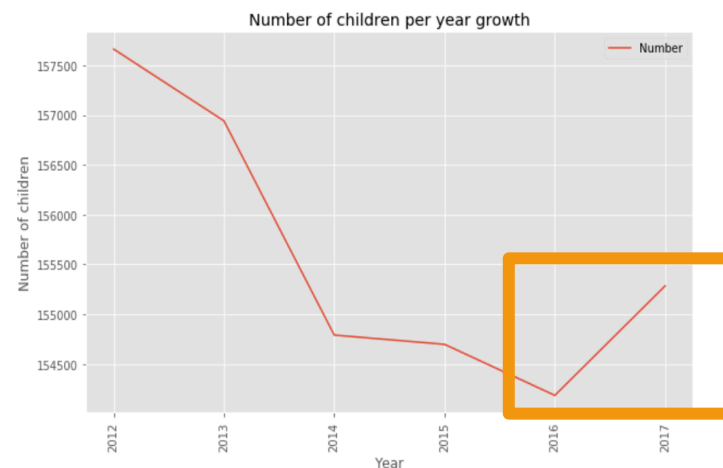
CONCLUSION

- Safety is an important criteria considering the target audience of families with children. The top 5 communities were chosen to be 60646, 60614, 60605, 60623 and 60655 with the highest safety scores.

- Geographical outliers, i.e. any community which is extremely far from the rest, was not considered. Zipcode 60655 was excluded.

- There are many competitors (i.e. sufficient supply) in the zipcodes 60605, 60614 and 60646. However, there is currently no common ice cream shop that families go to in 60623 (North Center).

- There is a general downward trend in number of toddlers (5-9) and teenagers (15-19), but a slow upward trend of children (10-14). Therefore, the targe audience is refined to focus on families with children between 10-14 years old in order to maximize customer reach.

- Deeper analysis into zipcode 60623 showed that they are one of the top 6 communities in Chicago with the highest income per capita and thus, this confirms our criterion of being a safe zipcode.

- Deeper analysis into zipcode 60623 (North Center) showed that it is amongst the top 10 'healthiest' communities due to its low child obesity score. This can present both a threat in terms of preferences of target market, but it can also be an opportunity in terms of changing your product portfolio to include healthier options.