Capstone Project: Predicting the Best Possible Location for a Bilingual Child Care Venue in Chicago, IL.

1. Introduction: Business Problem

Studies have shown that children under the age of 7 possess an innate ease of learning a secondary language; when children learn a secondary language early in their lives, they use the same part of their brains that is used to acquire their native language. In addition, early secondary language acquisition has proven to provide many cognitive and social benefits that will last a lifetime.

Based on these studies, Company ABC has developed a very successful business model, which consists in offering bilingual (English and Spanish) daycare services for children aged between 0 and 7 years. The company has already opened several profitable bilingual daycares in many cities across the United States and because almost 29% of Chicago's population is "Hispanic or Latino", the company has made the decision to open up a new daycare in Chicago as it seeks to expand its business into the city.

Company ABC is interested in knowing what would be the best location (neighborhood and zip code) for opening their new bilingual child daycare in the City of Chicago. As a result, the company has ordered the present study, which will help its business expansion team better understand the different residential neighborhoods within the city of Chicago, and ultimately select the best possible location for their new bilingual child daycare.

2. Data

The following datasets were used to assess the feasibility of opening a new bilingual child daycare in each one of the different zip codes within the city of Chicago, with the ultimate purpose of selecting the right location for the business:

2.1 The zip codes dataset: Company ABC target markets are located in Cook and DuPage counties, which are the most densely populated counties in the Chicago greater metropolitan area. In order to create a data frame that contained all of the zip codes available in these two counties, two zip codes lists, one for each county, were obtained and then merged into a single pandas data frame. The two zip code lists were obtained from the following websites: https://www.zipcodestogo.com/Cook/IL/ and https://www.zipcodestogo.com/Dupage/IL/.

These lists were then read into two pandas data frames by using the pandas.read_html() method. The two data frames were then merged to produce one single data frame named **zip_df**. The dimension of the resulting data frame was 276 rows by 2 columns. The first column named "Zip Code" contains all 276 zip codes, while the second column named "City" contains the names of the cities in each zip code.

```
print(zip_df.shape)
zip_df.head()
```

```
(276, 2)
```

	Zip Code	City
0	60004	Arlington Heights
1	60005	Arlington Heights
2	60006	Arlington Heights
3	60007	Elk Grove Village
4	60008	Rolling Meadows

2.2 **The 2018 income tax dataset** from the United States Internal Revenue Service (IRS) was obtained from https://www.irs.gov/pub/irs-soi/18zpallnoagi.csv. This data set provides detailed household income information, as well as some useful demographic information, such as number of households and number of claimed dependants by zip code.

The dataset was downloaded as a .csv file and read into a pandas data frame by using the pandas.read_csv() method. The data frame was then joined with the previously created **zip_df** data frame in order to return only the matching zip codes. The resulting data frame was named **IRS_df** and returned a total of 197 zip codes and five columns. The first two columns show the zip codes and city names of the 197 matching zip codes, while the remaining three columns show the number of households, the income per household and the dependants per household by zip code. The IRS_df dataset was used as follows:

i. The information regarding the number of dependents per household was considered very valuable. This data yielded very useful insights that were used to determine where most of the families with child dependents actually live. These results were compared against the number of child care venues in each zip code to analyze the ratios of child care venues to total number of dependants, which helped support the final recommendation.

ii. Another important metric is household income. Child care is not cheap, and families who can afford it need to make a reasonable amount of income. The average income per household per zip code was analyzed to determine whether families in specific zip codes were be able to afford bilingual child care.

```
print(IRS_df.shape)
IRS_df.head()
```

```
(197, 5)
```

	Zip Code	City	Number of households	Income per household	Dependants per household
0	60004	Arlington Heights	26490	112728.65	0.587769
1	60005	Arlington Heights	15580	103211.36	0.535302
3	60007	Elk Grove Village	18480	75952.98	0.483225
4	60008	Rolling Meadows	11670	69124.76	0.608398
6	60016	Des Plaines	31820	60761.57	0.520113

2.3 **The coordinates dataset:** The information regarding the latitude and longitude coordinates of every zip code was pulled from https://www2.census.gov/geo/docs/maps-data/data/gazetteer/2019_Gazetteer/2019_Gaz_zcta_national.zip as a zip file which contained a .csv file inside. The dataset was read into a pandas data frame by using the pandas.read_csv() method. The original .csv file, besides commas, had one additional type of separator, "\t" to separate the values. For this reason, I had to specify sep='\t' as one of the parameters of the method used to read the file into a pandas dataframe.

This dataset contained all the latitude and longitude coordinates of all the zip codes in the United States, as well as the areas in square miles of each zip code. The dataset was processed and cleaned to only include the zip codes of interest and then merged with the **IRS_df** dataframe to create the **Chicago_df** dataframe.

```
print(Chicago_df.shape)
Chicago_df.head()
```

(197, 8)

	Zip Code	City	Number of households	Income per household	Dependants per household	Area (SQMI)	Latitude	Longitude
0	60004	Arlington Heights	26490	112728.65	0.587769	11.083	42.112780	-87.979542
1	60005	Arlington Heights	15580	103211.36	0.535302	6.559	42.064490	-87.985462
2	60007	Elk Grove Village	18480	75952.98	0.483225	14.104	42.007517	-87.992860
3	60008	Rolling Meadows	11670	69124.76	0.608398	5.175	42.069786	-88.016221
4	60016	Des Plaines	31820	60761.57	0.520113	10.622	42.049573	-87.895003

2.4 The venues dataset was obtained through a loop in python that pulled multiple .json files from Foursquare and then processed and combined them into a single pandas data frame named **Chicago_venues**. The .json files provided location, name and venue category information about venues that are located within a specified radius from the latitude and longitude coordinates of each zip code. The venues dataset was used as follows:

i. The venues dataset was first used to extract the venue categories and group them together by zip code. In order to do this, one-hot encoding was performed on the venue categories, resulting in a total of 444 venue categories.

ii. Since a good part of the company's target customers are Hispanics and Latin Americans, this dataset was used to predict which neighborhoods in Chicago have the highest number of Hispanic and Latino communities. The correlations between certain venue categories and the existence of bilingual childcare venues were calculated by using simple and multiple regression models. This analysis yielded some interesting insights into where the company's target customers are located.

```
print(Chicago_venues.shape)
Chicago_venues.head(10)
```

(15965, 7)

	Zip Code	Zip Code Latitude	Zip Code Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	60004	42.11278	-87.979542	Trader Joe's	42.113167	-87.982543	Grocery Store
1	60004	42.11278	-87.979542	Binny's Beverage Depot	42.113807	-87.984634	Liquor Store
2	60004	42.11278	-87.979542	Chipotle Mexican Grill	42.112781	-87.978320	Mexican Restaurant
3	60004	42.11278	-87.979542	Sun Shui	42.115282	-87.985345	Chinese Restaurant
4	60004	42.11278	-87.979542	Panera Bread	42.112463	-87.981708	Bakery
5	60004	42.11278	-87.979542	LA Fitness	42.107957	-87.978448	Gym
6	60004	42.11278	-87.979542	Olive Garden	42.109389	-87.974479	Italian Restaurant
7	60004	42.11278	-87.979542	Gail's Carriage Place	42.113484	-87.979423	Breakfast Spot
8	60004	42.11278	-87.979542	Smash Burger	42.114233	-87.983625	Burger Joint
9	60004	42.11278	-87.979542	Ulta Beauty - Curbside Pickup Only	42.112753	-87.981959	Cosmetics Shop

2.5 The childcare venues dataset: Since not enough information regarding childcare venues could be obtained through Foursquare, I had to find an additional dataset containing relevant information regarding the locations of all childcare venues currently available in Chicago. Luckily, I was able to find a dataset from the State of Illinois which contained all childcare venues available in the state of Illinois. This data was obtained from the following website: https://sunshine.dcfs.illinois.gov/Content/Licensing/Daycare/ProviderLookup.aspx.

Since a direct link to the .csv file was inexistent in the above-mentioned website, the dataset was downloaded straight from the website and then uploaded into my Github repository as a .csv file. The file was then pulled and read into pandas data frame by using the pandas.read_csv() method. I filtered the dataframe to only include the zip codes available in the **IRS_df** data frame. I also applied a filter on the license status to make sure that only childcare venues with active licenses were included in the dataframe. All of the childcare venues were then grouped by zip code, and since there were 11 zip codes without any childcare venues, the data frame resulted in 186 rows, one for each zip code. The resulting data frame was called **daycare_df**.

The first column in the resulting dataframe contains all 186 zip codes. The second column shows the total number of childcare venues by zip code, while the third column shows the number of bilingual childcare venues by zip code. The fourth column shows the total available children capacity of all childcare venues by zip code, while the fifth column shows the available children capacity of the bilingual (English-Spanish) childcare venues by zip code.

The information in the **daycare_df** data frame was used to evaluate how saturated or competitive the markets were in each one of the zip codes. The more saturated or competitive the market is, the less desirable it is. Specifically, the dataset was used to calculate the ratio of daycare venues per dependants in each zip code. The purpose of this analysis was to avoid selecting zip codes where the market was already overly-saturated with other childcare venues.

```
daycare_df.head()
```

```
There are currently 1619 active child daycare venues in Cook and DuPage counties, of which 94 are bilingual (English-Spanish)
The total child capacity is 130008 children and total bilingual capacity is 9576 children.
The dimensions of daycare_df are: (186, 5)
```

	Zip	TotalCount	BilingualCount	TotalCapacity	BilingualCapacity
0	60004	14	0	1475	0
1	60005	5	0	372	0
2	60007	2	0	160	0
3	60008	1	0	90	0
4	60016	11	0	864	0

3. Methodology:

The main purpose of the analysis is to help Company ABC find the best possible zip code in the city of Chicago to start up a new bilingual child care center. In order to do this I have designed the following multi-step methodology process:

3.1 Simple linear regression models

The first step in the methodology process was to use multiple different simple linear regression models (one for each venue category) to find how individual venue categories were correlated with the existence of bilingual childcare venues in each one of the zip codes.

To do this, I first needed to perform one-hot encoding on the venue categories of the Chicago_venues dataset. Once I had all of the venue categories as columns in a dataframe, I proceeded to group the dataframe by zip code and applied a **sum()** function to the venue categories to return a count for each one of the venue categories by zip code. This resulted in a dataframe which displayed the number of venues in each category by zip code. I called this dataframe **Chicago_grouped** and it looked like this:

```
print(Chicago_grouped.shape)
Chicago_grouped.head()
```

```
(197, 445)
```

	Zip Code	ATM	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Amphitheater	Animal Shelter	Antique Shop	Aquarium
0	60004	0	0	0	0	0	0	0	0	0	3	0	0	0	0
1	60005	0	0	0	0	0	0	0	0	0	3	0	0	0	0
2	60007	0	0	0	0	0	0	0	0	0	2	0	0	0	0
3	60008	0	0	0	0	0	0	0	0	0	1	0	0	0	0
4	60016	0	0	0	0	0	0	0	0	0	1	0	0	0	0

Next, I merged the childcare venue information from the previously explained **daycare_df** dataframe with the **Chicago_grouped** dataframe by adding two additional columns to the **Chicago_grouped** dataframe: the first additional column displayed the total childcare venue count by zip code, while the second column displayed the bilingual childcare venue count by zip code. I called the resulting dataframe **Chicago_daycare**, and it looked like this:

```
print(Chicago_daycare.shape)
Chicago_daycare.head()
```

```
(197, 447)
```

ehouse	Warehouse Store	Water Park	Waterfall	Waterfront	Weight Loss Center	Whisky Bar	Wine Bar	Wine Shop	Winery	Wings Joint	Women's Store	Yoga Studio	Zoo	Zoo Exhibit	TotalCount	BilingualCount
0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	14.0	0.0
0	0	0	0	0	0	0	1	0	0	0	0	2	0	0	5.0	0.0
0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	2.0	0.0
0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1.0	0.0
0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	11.0	0.0

After creating the **Chicago_daycare** dataframe I was finally able to define simple linear regression models between each venue category and the number of bilingual child care venues available in each zip code in order to find out the venue categories that were more strongly correlated with bilingual child care venues.

In order to do this, I used sklearn.linear_model.LinearRegression() to define the simple linear regression model. I then looped through each individual venue category in the **Chicago_daycare** dataframe and fitted the model multiple times in each iteration with the venue category count of each individual zip code as the independent variable and the bilingual childcare count as the dependent variable. I then extracted the R2 score of every model fit and put it into a dataframe. I then sorted the dataframe in descending order to find the top 20 venue categories that were more strongly correlated with bilingual childcare venues. I called the resulting dataframe **R2_bilingual**, which looked like this:

```
R2_bilingual
```

	VenueCategory	r2BilingualCount
0	Mexican Restaurant	0.207892
1	Taco Place	0.216447
2	Rock Club	0.206948
3	Latin American Restaurant	0.202632
4	Cocktail Bar	0.167925
5	Dessert Shop	0.160914
6	Heliport	0.146841
7	Speakeasy	0.123189
8	Ukrainian Restaurant	0.095090
9	Café	0.094379
10	Casino	0.088312
11	Brewery	0.088113
12	Caribbean Restaurant	0.082222
13	Art Gallery	0.075530
14	Cuban Restaurant	0.074918
15	Food & Drink Shop	0.074176
16	Argentinian Restaurant	0.072530
17	Stadium	0.072233
18	Pie Shop	0.060757
19	Street Art	0.059227

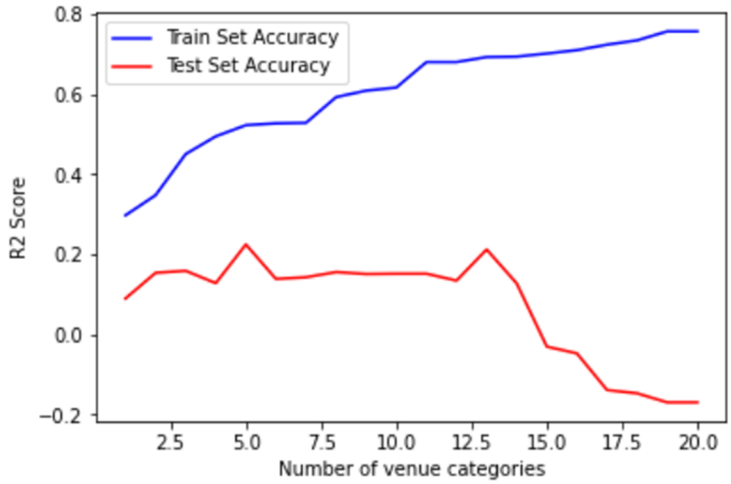
Individual R2 scores show how strongly correlated each venue category is with bilingual childcare venues.

3.2 Multiple linear regression model

A multiple linear regression machine learning algorithm was built to calculate the correlation between a set of venue categories, which were the independent variables, and the existence of a bilingual child care venue, which was the dependent variable.

The first problem that I encountered prior to building the multiple linear regression model, was to figure out the optimal number of independent variables that I needed to include in the model. There were a total of 444 venue categories, and I needed to figure out which and how many of these categories to include in the multiple linear regression model. Too many venue categories would result in an over-fitted model, while too few would result in an under-fitted model.

To find out the optimal number of venue categories, I designed a loop that tested the multiple linear regression model with the top 20 venue categories that had the strongest correlations with the bilingual childcare venues in each zip code and returned the R2 accuracies of the model with each iteration. The first iteration tested the model with the single one most correlated venue category, the second iteration tested the model with the top two venue categories, the third with the top three, and so on. I executed the loop with both a train and a test data set to see how the in-sample and out-of-sample accuracies of the model compared against each other. With this loop I was able to ultimately generate the following line chart which compared the accuracies of the model when using different number of venue categories:



By looking at the chart above, I was able to determine the optimal number of venue categories and used these categories to fit the multiple linear regression model accordingly. The resulting multiple linear regression model was then used to predict the number of bilingual child care venues in each zip code based on the nearby venue categories. I calculated the results for each zip code and put them into a pandas dataframe that I named **Predicted_locations** and looked like this:

```
print(Predicted_locations.shape)
Predicted_locations.head(10)
```

```
(197, 2)
```

	ZipCode	BilingualPredicted
0	60608	7.7[illegible]815
1	60622	7.000000
2	60617	6.000000
3	60647	5.000000
4	60623	3.731458
5	60616	3.210185
6	60804	3.091467
7	60612	3.000000
8	60632	2.505285
9	60641	2.448508

Number of predicted bilingual childcare venues by zip code.

The higher the magnitude of the prediction, the more likely that bilingual childcare venues would exist in the specified zip code based on the nearby venue categories.

3.3 Analysis of important market and demographic factors

There are some very important demographic and market factors that were considered and analyzed before selecting the best possible location for the new bilingual childcare venue. The zip code locations were analyzed based on the following four main factors:

- **Income per household:** This metric was used to analyze the childcare affordability of the households.
- **Total number of dependants per zip code:** This metric was used to analyze the size of the markets.
- **Bilingual childcare children capacity per dependant:** This metric was used to analyze the saturation of the markets in each zip code location for bilingual childcare venues.

- **Total childcare children capacity per dependant:** This metric was used to analyze the saturation of the markets in each zip code location for any type of childcare venues.

Prior to analyzing each one of these factors, I put them all together into a pandas dataframe by combining the **Predicted_locations** dataframe with demographic information extracted from the **IRS_df** and **daycare_df** dataframes. I named the resulting dataframe **Predicted_bilingual**, and it looked like this:

```
print(Predicted_bilingual.shape)
Predicted_bilingual.head(10)
```

```
(197, 6)
```

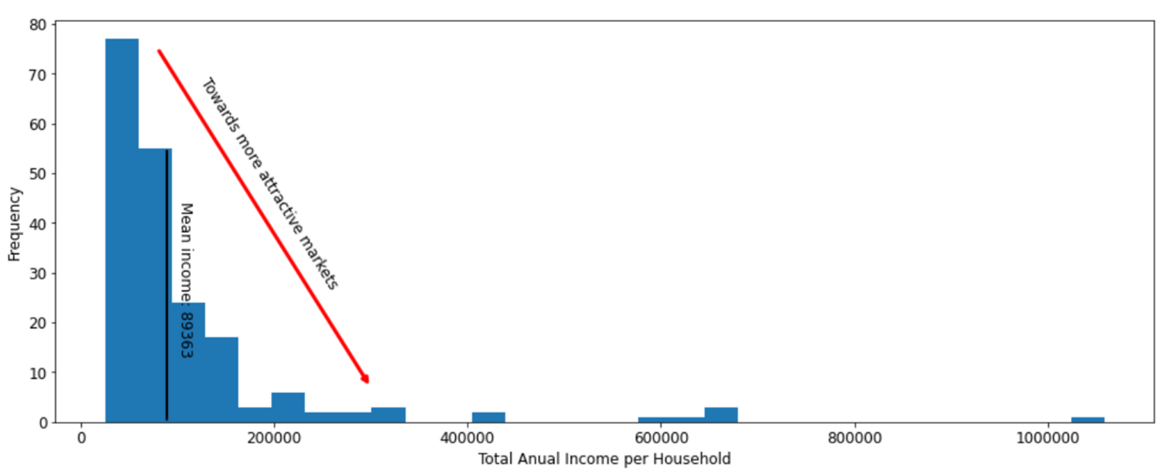
	ZipCode	Income per household	TotalDependants	BilingualPredicted	BilCap/Dependant	TotalCap/Dependant
0	60608	47111.52	18960.0	7.789815	0.068196	0.116192
1	60622	115882.63	8630.0	7.000000	0.087717	0.171147
2	60617	40541.92	26370.0	6.000000	0.022677	0.070648
3	60647	82541.76	18070.0	5.000000	0.026619	0.136525
4	60623	30856.84	34060.0	3.731458	0.024486	0.041574
5	60616	68185.80	10230.0	3.210185	0.018084	0.223460
6	60804	35927.35	34060.0	3.091467	0.009278	0.032149
7	60612	56555.27	9340.0	3.000000	0.019272	0.097109
8	60632	36378.62	36080.0	2.505285	0.012140	0.026414
9	60641	52903.96	21470.0	2.448508	0.000699	0.030461

Per conversations with Company ABC, the company said that they wanted to focus on the following markets:

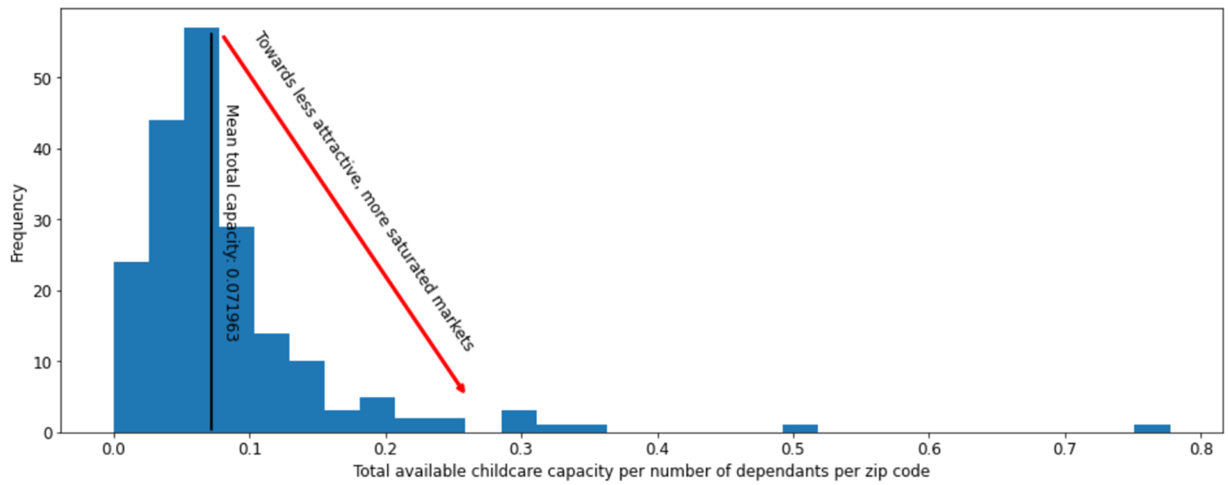
- Zip codes where annual income per household is greater than 75% of the mean annual income.
- Zip codes where total available childcare capacity per dependant is less than or equal to 150% of the mean.
- Since the bilingual daycare market in Chicago is generally underserved, the threshhold for the available bilingual childcare capacity per dependant should be less than or equal to 0.02.

In order to analyze and filter the data based on the criteria and thresholds discussed with Company ABC, I proceeded to calculate the mean income per household of the dataset, as well as the mean childcare capacities per dependant. With this information, I was able to create the following histograms to better visualize the distribution of the aggregated data and better understand the market as a whole:

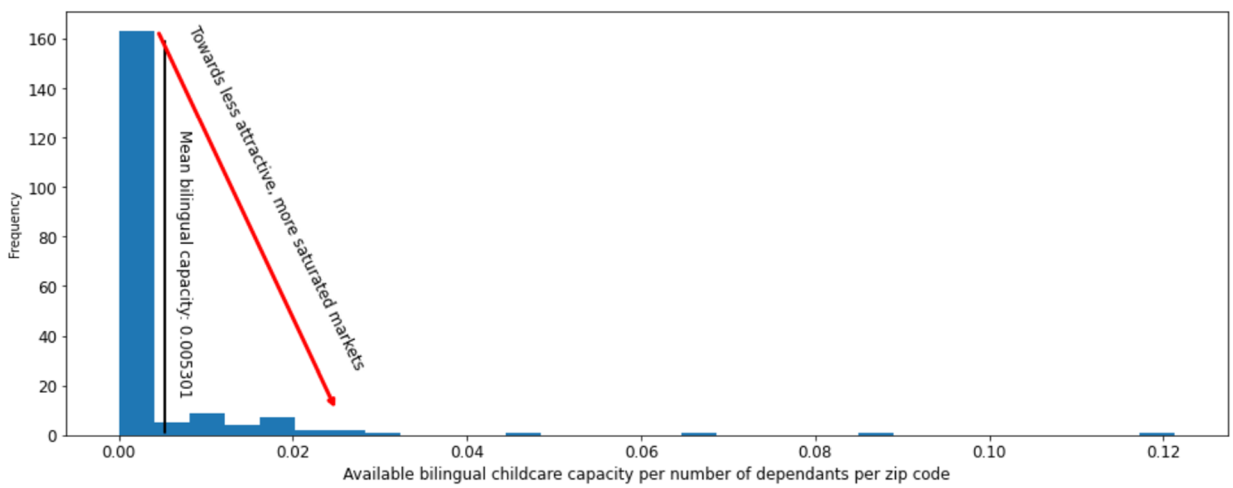
Annual Income per household:



Total childcare capacity per dependant per zip code:



Bilingual childcare capacity per dependant per zip code:



After analyzing and understanding the markets, I used the thresholds given by Company ABC to narrow down the data and come up with a list of zip codes that met the conditions specified by the company. I then sorted this list on descending order, based on the bilingual predicted locations obtained with the multiple linear regression model described in the previous section. I put this list in a pandas dataframe, named it **Predicted_final**, and it looked like this:

```
print(Predicted_final.shape)
Predicted_final.head(10)
```

```
(72, 6)
```

	ZipCode	Income per household	TotalDependants	BilingualPredicted	BilCap/Dependant	TotalCap/Dependant
0	60618	87771.90	23740.0	1.485818	0.018787	0.103623
1	60007	75952.98	8930.0	0.934508	0.000000	0.017917
2	60076	84219.31	11100.0	0.785158	0.000000	0.075225
3	60515	131704.18	8420.0	0.779750	0.000000	0.063777
4	60077	70701.72	8190.0	0.775296	0.000000	0.096947
5	60148	78409.80	14000.0	0.699080	0.000000	0.055929
6	60068	151425.14	12100.0	0.604227	0.000000	0.051818
7	60655	87303.12	8680.0	0.600217	0.000000	0.007834
8	60606	580587.41	540.0	0.598937	0.000000	0.000000
9	60189	134025.58	8990.0	0.563421	0.000000	0.029922

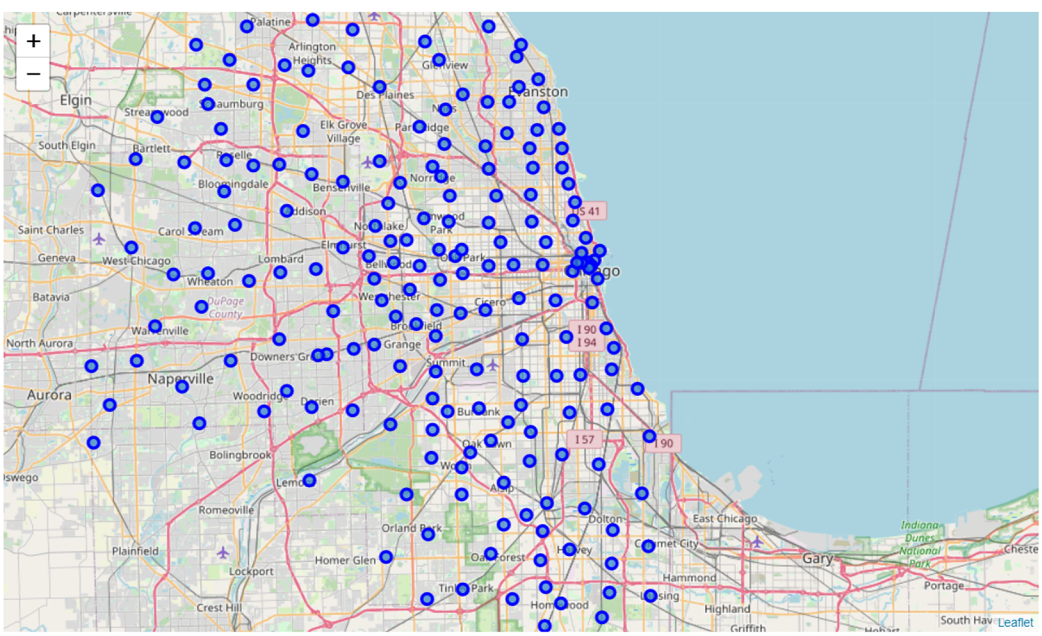
In the dataframe above, the zip codes at the top of the table represent the best possible locations based on the multiple linear regression model, while still meeting the criteria discussed with Company ABC. The top 3 zip codes in this table were selected and individually analyzed to come up with a final recommendation.

4. Results:

I will split this section into different parts in order to go over the results that were obtained in each one of the steps described in the Data and Methodology sections of this report.

4.1 Dataset results

As described in the Data section of this report, the data used in the analysis came from multiple sources. The data was pulled, processed, transformed and put into multiple dataframes that were then used in the analysis. The first significant results that were obtained when pulling the data, were the zip code lists and the latitude and longitude coordinates of each zip code. The map below shows a visualization of the results that were obtained after the initial processing of the geographical data with each blue point representing a zip code:



After merging the data from the zip code lists with the data from the IRS, the total number of zip codes was reduced from 276 to 197 because some zip codes did not have any reported household income. I excluded these zip codes because I was only interested in zip codes with actual households.

The venues dataset obtained from Foursquare resulted in a total of 15,965 venues and 444 venue categories spread out throughout the 197 zip codes. Only one childcare venue was returned from this dataset.

The **daycare_df** dataframe only returned existing childcare venues in 186 zip codes out of the 197 zip codes in the list, which means that 11 zip codes did not have any active childcare venues. The dataframe returned a total of 1619 active childcare venues, 94 of which were English-Spanish bilingual. The total childcare capacity for all types of childcare venues was 130,008 children, 9,576 of which were for bilingual types of venues.

4.2 Simple linear regression models results

The simple linear regression models that were used to find the individual correlations between each one of the venue categories and the bilingual childcare venues returned the following top 20 venue categories along with their R2 scores. The higher R2 score, the more accurate is the simple linear regression model that was used to find the correlation between that specific venue category and bilingual childcare venues. In other words, the higher the R2 score, the stronger the correlation:

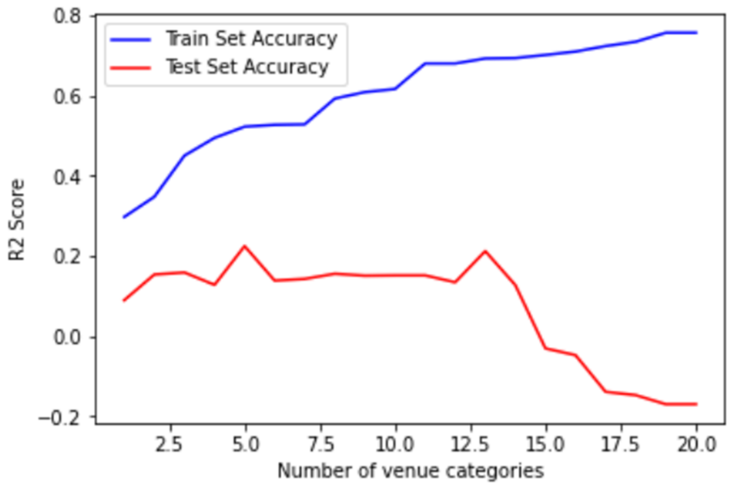
Place	Venue Category	R2 score
1	Mexican Restaurant	0.297892
2	Taco Place	0.216447
3	Rock Club	0.206948
4	Latin American Restaurant	0.202632
5	Cocktail Bar	0.167925
6	Dessert Shop	0.160914
7	Heliport	0.146841
8	Speakeasy	0.123189
9	Ukrainian Restaurant	0.09509
10	Café	0.094379

Place	Venue Category	R2 score
11	Casino	0.088312
12	Brewery	0.088113
13	Caribbean Restaurant	0.082222
14	Art Gallery	0.07553
15	Cuban Restaurant	0.074918
16	Food & Drink Shop	0.074176
17	Argentinian Restaurant	0.07253
18	Stadium	0.072233
19	Pie Shop	0.060757
20	Street Art	0.059227

Note that among the top five venue categories that are most strongly correlated with bilingual child care venues are Latin American, Taco and Mexican restaurants. With these results one can predict that a good part of the Hispanic and Latino population could be living in the areas where these types of venues are located. This assumption will be used further in the analysis to better predict the best possible location for the English-Spanish bilingual childcare.

4.3 Multiple linear regression model results

As discussed in section 3.2 of this report, a multiple regression model was built in order to accurately predict the number of bilingual childcare venues in each zip code based on nearby venues. In order to optimize the number of independent variables that were used in the model, I created a loop that tested the in-sample and out-of-sample accuracy the model with the top 20 venue categories found in the simple linear regression models described above. The chart below was produced by the loop and I used it to determine that 13 was the optimal number of categories for the multiple linear regression model before the out-of-sample accuracy started to drop.

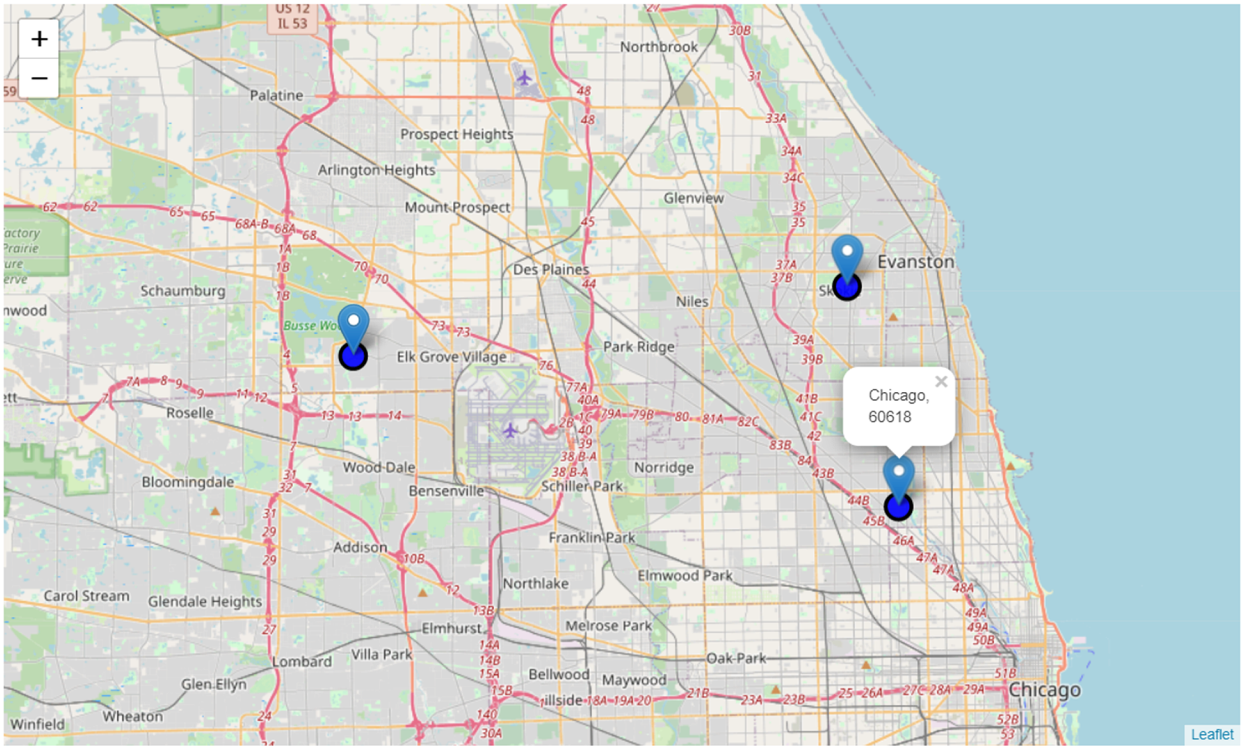


Categories in the places 1 through 13 in the tables above were fitted into the model. I then used the model to predict the number of bilingual childcares in each one of the 197 zip codes. The model yielded an in-sample R2 accuracy of 0.6931, while its out-of-sample R2 accuracy resulted much lower at 0.2126.

4.4. Final Results

After applying the thresholds given by Company ABC, discussed in section 3.3 of this report, I was able to narrow down the data from 197 zip codes to just 72. I then sorted the data to show the zip codes with the highest bilingual childcare prediction on the top of the table and proceeded to separately analyze each one of the top 3 zip codes in the table. I found out that the best zip code for starting up a new bilingual childcare location was zip code 60618, Chicago.

The following map shows the location of the top three zip codes. Zip code 60618 is the one that is closer to Chicago city center:



5. Discussion

The resulting top three zip codes were analyzed separately and their characteristics and demographics were compared against each other to ultimately determine which of the three zip codes offers the best location for Company's ABC bilingual child daycare. The following comparisons offer some interesting insights into the sizes of their Latin American populations and how their total childcare capacities and bilingual childcare capacities compare to the number of claimed dependants in each zip code:

60618: There are a total of 30 childcare locations in 60618, of which 4 are bilingual. By looking at the data, we can conclude that zip code 60618 has by far the largest population of the three, with more than 44k households and almost 24k claimed dependants. The total childcare capacity in this zip code is 2460 children, which represents roughly about 10% of the claimed dependant population in this zip code. The bilingual childcare capacity is 446. Even though supply might seem high for bilingual childcares in this zip code, it is not really that high considering that more than 40% of the population in this zip code are Hispanics and Latin Americans according to https://www.illinois-demographics.com/60618-demographics. 60618 is also the 2nd most populated zip code in the state of Illinois. At 87,771 USD, zip code 60618 has the highest average annual income per household of the three.

60007: There are just 2 childcare locations in this zip code, none of which are bilingual. Zip code 60007 (Elk Grove Village) is a residential area with a relatively high number of households and very low number of childcare venues. The total number of claimed dependants in this zip code is almost 9k, which makes the childcare capacity of just 160 look very low. The childcare market in this zip code is likely to be underserved, which could make it a great opportunity for starting a new childcare center in this zip code. However, only 11.8 % of people living in this zip code are Hispanics or Latinos according to https://www.illinois-demographics.com/60007-demographics. Zip code 60007 has the lowest average annual income of the three at almost 76k USD per household.

60076: Zip code 60076 has the fewest number of households of the three, but not the fewest number of claimed dependants. This zip code has a little more than 11k claimed dependants and a total childcare capacity of 835 children. There are no bilingual childcare venues in this zip code, which is consistent with its low population of Hispanics and Latin Americans, which according to https://www.illinois-demographics.com/60618-demographics, accounts for just 10.6% of its population, the lowest of the three. The average annual income per household in Zip code 60076 is slightly more than 84k USD.

6. Conclusion:

Since more than 40% of the population in zip code 60618 are Hispanics and Latin Americans, and since there are almost 24 thousand claimed dependants in 60618, zip code 60618 is my recommended location for starting up a new bilingual child care venue. Even though competition in zip code 60618 could be tougher than in 60007 and 60076, the market in 60618 is substantially larger than in the other two zip codes, especially for a bilingual (English-Spanish) childcare.

In addition, zip code 60618 is located closer to the city center of Chicago than zip codes 60007 and 60076. Moreover, 60618 is located right in between Evanston and Chicago, which makes 60618 a good strategic location that would allow Company ABC to offer its bilingual childcare services to residents of Evanston who commute to Chicago and vice versa.