Finding New Business Clients for Wi-Fi Services

Introduction/Business Problem

Many dining establishments such as restaurants and cafés now offer Wi-Fi as a way of attracting and keeping customers. For example, people who do not have a high-speed internet connection in their home often utilize Wi-Fi at such places; this can result in a steady customer base for the dining establishment (since patrons are expected to order food and beverages at such places). Wi-Fi availability is also useful for people who are traveling and don't have access to their usual home or office internet connection. Given a choice between having lunch at place that does have a Wi-Fi connection versus one that doesn't, a businessperson on the road would likely choose to eat at the café or restaurant that provides fast and free internet access.

It would make sense then for companies that **provide Wi-Fi to businesses** to try to identify those dining establishments which could benefit from the increased customer traffic that often results from Wi-Fi availability. However, finding those businesses is not simple: first, it involves an effort to identify **which types of businesses** would most benefit from making Wi-Fi available at their establishments. There are many types of dining establishments, of course, but not all of them may benefit from the addition of Wi-Fi – for example, customers may care a lot about whether or not Wi-Fi is available at a sandwich shop, but they might not care at all about whether or not it is available at a steakhouse.

One way to determine the importance of Wi-Fi for a given type of dining establishment (such as a coffee shop or donut shop) is to compare the Foursquare ratings of businesses of that type that do offer Wi-Fi versus the ratings of those which do not. For example, if it could be demonstrated that the average ratings of donut shops with Wi-Fi were significantly higher than those without Wi-Fi, this would lead to a business opportunity (for the internet provider) because venues with higher ratings are more likely to attract customers than those with lower ratings. If you can demonstrate to a business owner that the installation of Wi-Fi will lead to increased customer traffic, then that should help to sell internet services to that business. How would an internet provider obtain that information? Fortunately, the ratings for companies in many areas can be obtained electronically from Foursquare (as described in detail in the "Data" section).

Once it was determined which **types** of businesses would benefit from the addition of Wi-Fi, then the next step would involve identifying which **specific** businesses of that type do not yet have Wi-Fi (so that marketing efforts can be directly towards them). The Foursquare data also provides a way to easily obtain this information. First, a list of businesses of a specific type with a certain region can be obtained electronically by requesting it from the Foursquare website (using one of their application programming interfaces (i.e. APIs)). For example, if you wanted to obtain a list of all coffee shops within a 10,000 meter radius of the University of California, Los Angeles, you could do so by sending an electronic request with those parameters through a Foursquare API.

Once list of businesses were obtained, a salesperson at an internet provider could use it to identify potential candidates. One key step would be to narrow it down to only those businesses which do not currently offer Wi-Fi. Also, it would be essential to obtain contact information for all the businesses on the list. This is also available through Foursquare, so for each business on the list, you extract the contact information from the data received and use this contact information for direct mail campaigns,

cold calls or in-person sales calls. A diligent marketing campaign that utilized all of the information available from Foursquare could certainly result in new business for a Wi-Fi provider.

Data

As mentioned in the Introduction, the aim of this project is to gather data from Foursquare to determine which **types** of dining establishments will benefit the most by offering Wi-Fi (to their customers) as well as to determine which **specific** businesses should be targeted. For this project, I am going to focus on three types of dining establishments: coffee shops, sandwich places and donut shops. In addition, to keep the scope of this project manageable, I am going to focus on businesses in the greater Los Angeles area (though the conclusions drawn should be applicable to other major metropolitan regions).

To obtain this data, I relied upon the Foursquare database; specifically, I used a Python program that I wrote (called Gathering_Foursquare_Data_for_Analysis) to make calls to Foursquare using their various application programming interfaces (APIs); each call made to Foursquare contains the specifics of the information that I need (such as venue type and location); once the results are obtained from each call, the Python program parses the results and creates records (for each business) that can be used for my analysis.

The first step in the process is to obtain lists of 1) coffee shops, 2) sandwich places and 3) donut shops in the Los Angeles area. For each type of venue, a call is made to Foursquare (using their API) to obtain a list of venues specified by type and location. Once the results of each call are received, the list of venues returned in the query is appended to a master list of venues stored in the Python program. (By using calls to Foursquare, I was able to obtain about 150 examples of each type of venue.)

The records received from Foursquare (in this type of query) contain many fields, but the program stores only those values which are relevant for my analysis, including:

- The ID number assigned to the venue by Foursquare
- The name of the venue
- The street address of the venue
- The city where the venue is located
- The state where the venue is located
- The ZIP Code where the state is located
- The latitude and longitude of the venue
- The phone number of the venue
- The type of the venue (i.e. “Coffee Shop”, “Sandwich Place”, “Donut Shop”)

For my analysis, I also need the current rating of the venue (a value on a sale of 0.0 to 10.0) as well as the value from a field which indicates whether or not the venue has Wi-Fi. However, these fields are not present on records received in the type of query described above, so in order to obtain that information, the program makes an additional call (using a different Foursquare API) to fetch detailed information for each venue in the list (including the values in the two aforementioned fields). When the results of each individual call are received, the program extracts the values the rating and “has wi-fi” fields and updates each record in my dataset accordingly.

(Because of limitations involved with obtaining data from Foursquare, I ran the "Gathering_Foursquare_Data_for_Analysis" the program multiple times, and in each run I obtained the data for a different type of business. The output for each run was stored in a CSV file labeled with the type of business (such as "Coffee_Shop.csv"). The analysis program (described in the next section) then read in the data from the individual files and compiled it into a single data set.

Methodology / Analysis

After obtaining the full set of detailed information for each of the venues, I then performed analysis on the venues using a second program (called Finding_New_Business_Clients_for_Wi-Fi_Services, also written in Python). In analysis program I carried out the following steps:

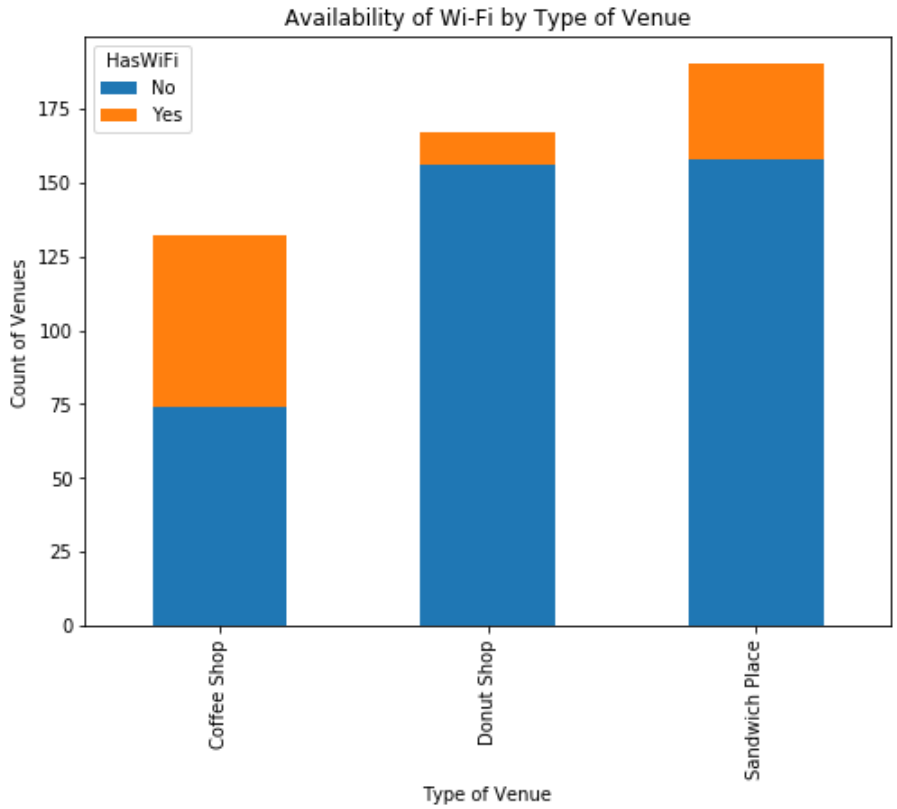
- I used the program to calculate and compare the rates of adoption of Wi-Fi across businesses of various types. The purpose of this was to show whether or not business of one type were more likely to offer Wi-Fi than other types. For example, I wanted to determine if coffee shops in general have greater availability of Wi-Fi than donut shops. (Knowing this information is useful in determining which types of venues to target – if it appears that donut shops are lagging behind other types of venues in terms of installing Wi-Fi, then that could mean that there are more opportunities for business among donut shops).
    - As part of this, I used a crosstab calculation that showed, for each type, how many businesses of that type offered Wi-Fi and how many did not
    - In addition, to illustrate the data, I generated a stacked bar chart that showed the rates of adoption of Wi-Fi for each type of business
    - Also, I performed a Chi-Square Test of Independence to determine whether any differences in the rates of Wi-Fi availability were statistically significant
- **Within each type of business**, I used the program to calculate the average Foursquare ratings of those business that offer Wi-Fi versus those which do not. The purpose of this was to demonstrate whether or not businesses (of a given type) that have Wi-Fi have significantly higher ratings that those which do not have Wi-Fi (because if it can be demonstrated that businesses of a certain type that offer Wi-Fi have higher ratings than businesses of that same type which do not, that could be used as a selling point).
    - For each type of business, the program calculated the average rating for those business that did offer Wi-Fi versus those which did not
    - In addition, to illustrate this data, individual box plots were generated (for each type of business) that show the average ratings for businesses with Wi-Fi versus those without Wi-Fi, so that a visual comparison can easily be made
    - Lastly, an Analysis of Variance (ANOVA) was performed to show whether any differences in ratings that are detected were statistically significant

Results

As mentioned above, the first step in the process was to examine rates of Wi-Fi availability across the three types of businesses. For each venue type, here are the counts of the venues that offer Wi-Fi versus those which do not:

```
Venue Type: Coffee Shop

Wi-Fi Yes: 58   (43.94%)
Wi-Fi No:  74   (56.06%)
Total:     132

Venue Type: Donut Shop

Wi-Fi Yes: 11   (6.59%)
Wi-Fi No:  156  (93.41%)
Total:     167

Venue Type: Sandwich Place

Wi-Fi Yes: 32   (16.84%)
Wi-Fi No:  158  (83.16%)
Total:     190
```

In addition, here is a stacked bar chart that shows the rates of Wi-Fi availability by venue:



Based on the counts above, we can see that coffee shops have the highest availability of Wi-Fi (at a rate of 43.94%) and donut shops have the lowest rate (at 6.59%). The rate of Wi-Fi availability at sandwich shops falls in between those two (at 16.84%, closer to donut shops than coffee shops). This is reinforced by the bar chart above, which shows the much higher rate of Wi-Fi availability at coffee shops and the relatively sparse availability of Wi-Fi at donut shops.

However, are those differences statistically significant? Is it possible that the differences are just due to chance, and that true rates of Wi-Fi availability are essentially the same across businesses? To determine whether or not the results were due to chance, a Chi-Square Test of Independence was performed. The results of that test are below:

```
Chi-Square value: 65.5217
p-value for Chi-Square test: 5.91751656923e-15
```

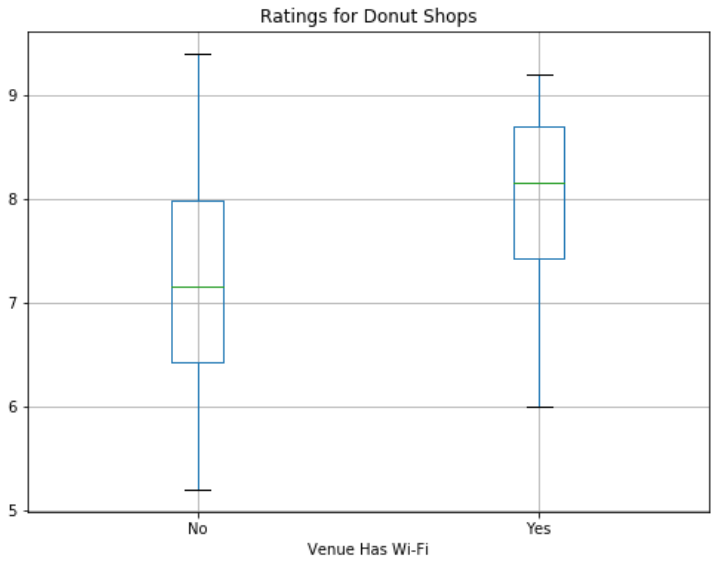
The Chi-Square value (65.5217) is high, indicating that the results (i.e. the differences in rates) are statistically significant. In addition, the very small p-value (5.91751656923e-15, which is far below the threshold of 0.05) tells us that the likelihood that the results were due to chance is extremely low. We can say with a great degree of confidence that the rates of availability of Wi-Fi clearly differ among the three types of businesses, with donut shops as those with the lowest rate of Wi-Fi availability.

The next step was to determine if, for each type of business, the average Foursquare rating for a business with Wi-Fi was higher than the average Foursquare rating for a business without Wi-Fi, and to what degree the differences were statistically significant.

Here are the average ratings for donut shops with and without Wi-Fi:

```
Wi-Fi - Yes: 7.9917
Wi-Fi - No:  7.2533
```

Also, here is boxplot that shows the distribution of ratings for donut shops with and without Wi-Fi side-by-side:



From the data (and the graph) above, we can see that donut shops with Wi-Fi have an average rating that is 0.7383 higher than those without. This means that a case could be made that Foursquare user rate donut shops with Wi-Fi nearly ¾ of a point higher than those without Wi-Fi. Many business owners would likely consider that to be a substantial increase – it could mean the difference between a rating of 7.4 and a rating of 8.1, for example. But is that value statistically significant? In order to determine that, an Analysis of Variance (ANOVA) test was performed on the data from donut shops. Below is the output generated by the analysis:

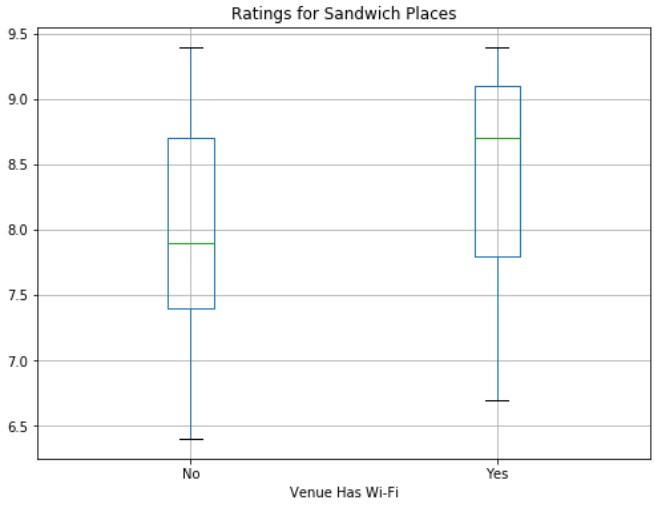
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Rating   R-squared:                       0.056
Model:                            OLS   Adj. R-squared:                  0.047
Method:                 Least Squares   F-statistic:                     5.987
Date:                Sun, 05 Aug 2018   Prob (F-statistic):             0.0162
Time:                        18:15:26   Log-Likelihood:                -141.86
No. Observations:                 102   AIC:                             287.7
Df Residuals:                     100   BIC:                             293.0
Df Model:                           1
Covariance Type:            nonrobust
================================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------
Intercept            7.2533      0.104     70.079      0.000       7.048       7.459
C(HasWiFi)[T.Yes]    0.7383      0.302      2.447      0.016       0.140       1.337
==============================================================================
Omnibus:                        3.715   Durbin-Watson:                   0.535
Prob(Omnibus):                  0.156   Jarque-Bera (JB):                2.137
Skew:                           0.075   Prob(JB):                        0.344
Kurtosis:                       2.307   Cond. No.                         3.15
==============================================================================
```

Looking at the output of the Analysis of Variance, we can see that the p-value for the coefficient that represents the increase in rating points for donut shops with Wi-Fi (i.e. "C(HasWiFi) [T.Yes]") is 0.016, which is far below the standard of 0.05 – this indicates that likelihood that the results were due to chance is very small (less than 2%). **This means we can say with great confidence that the average rating of donut shops with Wi-Fi is substantially higher than the average ratings of donut shops without Wi-Fi, and that difference is statistically significant.**

Next, let's look at results for sandwich places. Here are the average ratings for sandwich places with and without Wi-Fi:

```
Wi-Fi - Yes: 8.4571
Wi-Fi - No:  8.0171
```

Also, here is boxplot that shows the distribution of ratings for sandwich places with and without Wi-Fi side-by-side:



In this case, we can see that the average rating for sandwich places with Wi-Fi is 0.4401 points higher than those without Wi-Fi. This difference amounts to less than half a point, and is certainly not as dramatic a difference as it is for donut shops. Still, let's check to see if the difference in ratings for sandwich shops is statistically significant by performing an Analysis of Variance (the results of which are shown below).

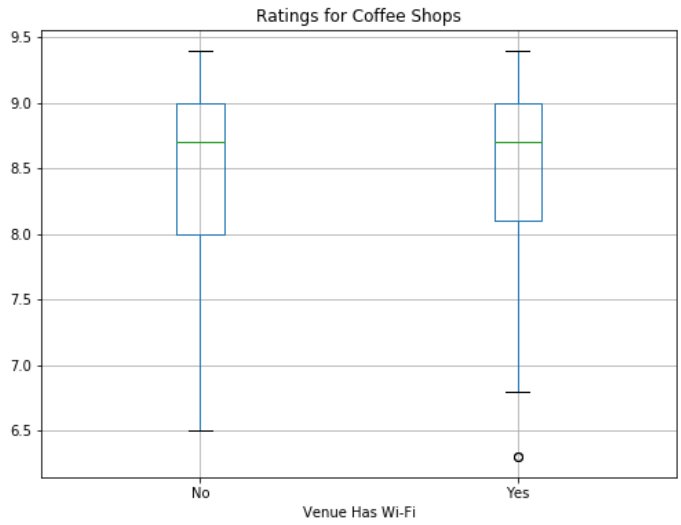
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Rating   R-squared:                       0.043
Model:                            OLS   Adj. R-squared:                  0.038
Method:                 Least Squares   F-statistic:                     8.832
Date:                Sun, 05 Aug 2018   Prob (F-statistic):            0.00333
Time:                        18:15:39   Log-Likelihood:                -235.78
No. Observations:                 199   AIC:                             475.6
Df Residuals:                     197   BIC:                             482.1
Df Model:                           1
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept        8.0171      0.062    129.097      0.000       7.895       8.140
C(HasWiFi)[T.Yes]  0.4401      0.148      2.972      0.003       0.148       0.732
==============================================================================
Omnibus:                       45.359   Durbin-Watson:                   1.566
Prob(Omnibus):                  0.000   Jarque-Bera (JB):                9.539
Skew:                          -0.075   Prob(JB):                      0.00849
Kurtosis:                       1.938   Cond. No.                         2.72
==============================================================================
```

Looking at the output of the Analysis of Variance for sandwich place we can see that the p-value for the coefficient that represents the increase in rating points for sandwich places with Wi-Fi (i.e. "C(HasWiFi)[T.Yes]") is 0.003, which is also far below the standard of 0.05. This very low p-value indicates that the likelihood that the results were due to chance is extremely small (less than 1%). In this case also then we can say **with great confidence that the average rating of sandwich with Wi-Fi is higher than the average ratings of sandwich place without Wi-Fi, and that difference is statistically significant.**

Let's now look at our last category of businesses, coffee shops. Here are the average ratings for coffee shops with and without Wi-Fi:

```
Wi-Fi - Yes: 8.5052
Wi-Fi - No:  8.3875
```

Also, here is boxplot that shows the distribution of ratings for coffee shops with and without Wi-Fi side-by-side:



Looking at this data, we can see that the difference in average ratings between coffee shops with Wi-Fi and those without seems paltry, as coffee shops with Wi-Fi have only a 0.1177 point advantage over those that do not have Wi-Fi. This certainly does not seem like much of a selling point (at least for coffee shops), but for the sake of completeness, let's see if that difference is statistically significant. Below is the output from an Analysis of Variance performed on the coffee shop ratings:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Rating   R-squared:                       0.006
Model:                            OLS   Adj. R-squared:                 -0.002
Method:                 Least Squares   F-statistic:                    0.7750
Date:                Sun, 05 Aug 2018   Prob (F-statistic):              0.380
Time:                        18:15:45   Log-Likelihood:                -147.37
No. Observations:                 130   AIC:                             298.7
Df Residuals:                     128   BIC:                             304.5
Df Model:                           1
Covariance Type:            nonrobust
================================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------------------
Intercept                8.3875      0.089     93.941      0.000       8.211       8.564
C(HasWiFi)[T.Yes]        0.1177      0.134      0.880      0.380      -0.147       0.382
==============================================================================
Omnibus:                       17.971   Durbin-Watson:                   0.731
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               21.277
Skew:                          -0.981   Prob(JB):                     2.40e-05
Kurtosis:                       3.278   Cond. No.                         2.51
==============================================================================
```

In this case, the p-value of the coefficient that represents the difference in ratings between coffee shops with Wi-Fi and those without is 0.380, far above the 0.05 threshold. Based on this, we cannot reject the idea that the 0.1177 difference was due to chance. **This effectively means there is no demonstrable difference in the average rating of coffee shops with Wi-Fi versus those without.**

Discussion (Observations / Recommendations)

To summarize the results:

- We can demonstrate (with a 98% degree of confidence) that the average Foursquare rating of donut shops that offer Wi-Fi is approximately ¾ of a point higher than donut shops that do not offer Wi-Fi. We can also show that the rate of availability of Wi-Fi at donut shops is significantly lower than either of the other two types of venues.
- We can demonstrate (with a 99% degree of confidence) that the average Foursquare rating of sandwich places that offer Wi-Fi is approximately 0.4 points higher than sandwich places that do not offer Wi-Fi. We can also show that the availability of Wi-Fi at sandwich shops is somewhat greater than that of donut shops but far less than that of coffee shops.
- We can demonstrate that there is no statistically significant difference in the Foursquare ratings of coffee shops with Wi-Fi versus coffee shops without Wi-Fi

From these results, we can see that the largest business opportunity (in terms of selling Wi-Fi to business clients) is to target donut shops. It can be demonstrated that donut shops with Wi-Fi have an average Foursquare rating that is ¾ of a point higher than donut shops that do not, and a case can be made (to donut shop owners) that a higher rating leads to increased visits – given the choice between a donut shop with a rating of 7.4 and one with a rating of 8.1, the average Foursquare user is much more likely to choose one with the higher rating. In addition, Foursquare users (who are obviously tech-savvy) are more likely to appreciate free Wi-Fi and are more likely to give favorable ratings to business that offer it (which again can lead to additional business). Also, including the boxplot chart (that shows the difference in ratings) as part of a sales presentation would help reinforce that message. Lastly, since donut shops have lowest rates of adoption of Wi-Fi (i.e. less than 7%), they represent an untapped market, much more so than the other two types of venues.

Is it worth it to attempt to sell Wi-Fi to the other two types of venues? In the case of sandwich places, the message would not be as powerful, since the results showed that sandwich places with Wi-Fi only had about a 0.4 increase in ratings versus those that did not. A sales pitch using that value would not likely to be effective. Still, we can see that there is still a substantial market for Wi-Fi among sandwich places (since less than 17% of such venues offer it), so it may be worth pursuing those businesses, but only after any and all campaigns aimed at donut shops were completed.

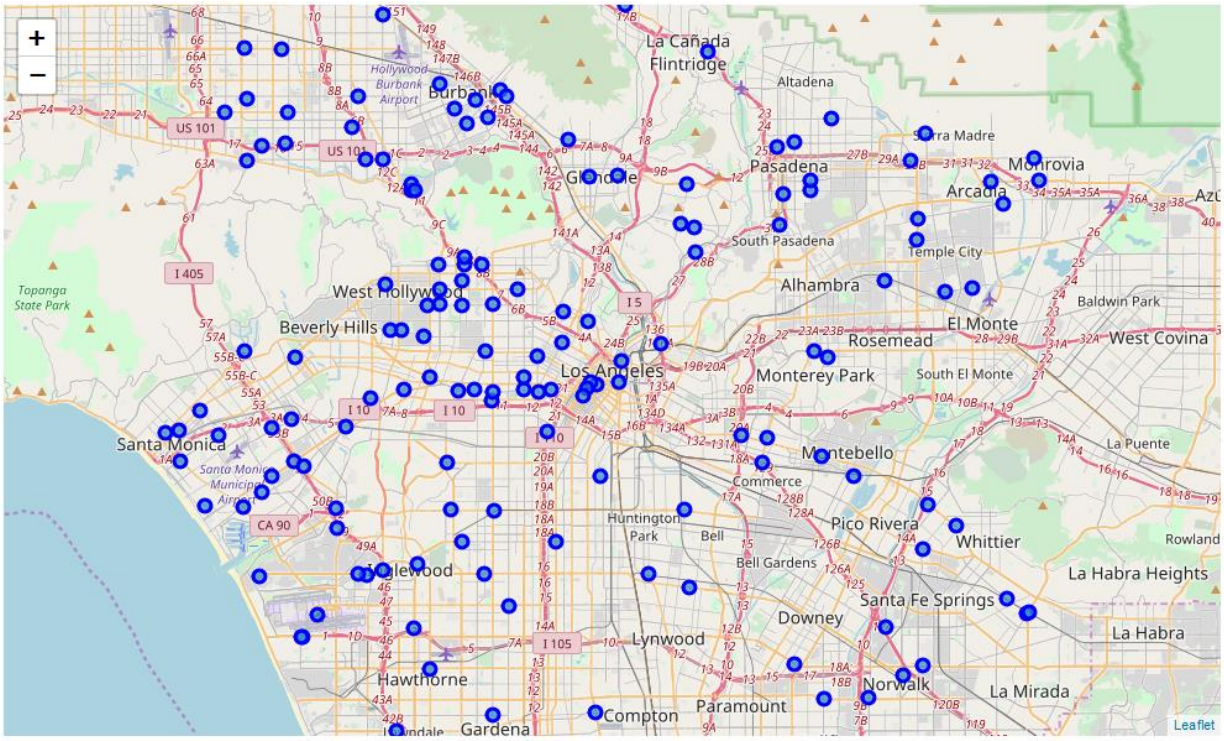
For coffee shops, the results from this study show that those venues deserve the least amount of attention, since it cannot be demonstrated that they gain anything (in terms of ratings) by offering Wi-Fi, so a sales pitch based on an potential increase in ratings would not be feasible in this case. This market is only worth pursing once all other markets had been exhausted (and some other approach would be required in the sales pitch – one that was not based a potential increase in ratings).

Tools for Selling Wi-Fi to Donut Shops

Since the recommendation of this study is to target donut shops as potential Wi-Fi customers, let's leverage the information gathered in order to create some useful data for a sales campaign. To this end, the output of the program also generates a list of local donut shops (i.e. local to the Los Angeles area) which do not have Wi-Fi. A salesperson can use this list for either a direct mail campaign, cold calling (via phone) or in-person visits. Here is an example of that list:

	A	B	C	D	E	F	G	H	I
1	**Name**	**Address**	**City**	**State**	**ZIP Code**	**Latitude**	**Longitude**	**Phone Number**	**Rating**
83	Yum Yum Donuts	8139 W Manchester Ave	Playa del Rey	CA	90293	33.95854168	-118.4388375	(310) 823-0045	7.4
84	Randy's Donuts & Chinese Food	210 N Market St	Inglewood	CA	90301	33.964636	-118.350975	(310) 673-1515	
85	Randy's Donuts	805 W Manchester Blvd	Inglewood	CA	90301	33.96176315	-118.3703202	(310) 645-4607	8.7
86	Golden Donuts and Ice Cream	11051 Hawthorne Blvd	Lennox	CA	90304	33.93461329	-118.3530747		
87	Dunkin' Donuts	1132 Wilshire Blvd	Santa Monica	CA	90401	34.025021	-118.490702	(310) 576-9200	7.6
88	Santa Monica Donuts	2822 Santa Monica Blvd	Santa Monica	CA	90404	34.03518084	-118.4717006		7.3
89	DK's Donuts and Bakery	1614 Santa Monica Blvd	Santa Monica	CA	90404	34.02578275	-118.4834093	(310) 829-2512	8.7
90	Yum Yum Donuts	2628 Pico Blvd	Santa Monica	CA	90405	34.02345959	-118.4617761	(310) 593-0119	6.1
91	Donut King	1912 Lincoln Blvd	Santa Monica	CA	90405	34.01183011	-118.4829363		7.4
92	Yum Yum Donuts		Torrance	CA	90504	33.8737615	-118.3436071		
93	Us Donuts & Yogurt	11719 Whittier Blvd	Whittier	CA	90601	33.98220513	-118.0515363	(562) 695-8867	
94	USA Donuts		Whittier	CA	90603	33.941918	-118.012085		
95	USA Donuts and Croissants	10255 Colima Rd	Whittier	CA	90603	33.941941	-118.011002	(562) 941-0422	

Also, the program generated a map shows the locations of each of the businesses (on the list). This map can be used by a salesperson to plan in-person visits to the various donut shops that he or she wants to target:



Conclusion

In this study, we have assessed potential markets for new business customers of Wi-Fi services. Based on data obtained from Foursquare, we have determined that the greatest potential for new business clients is donut shops, because their current rate of Wi-Fi adoption is low and it can be demonstrated that they can potentially get a significant gain in business from the increased ratings that often accompany donut shops that offer Wi-Fi. Thus, the greatest amount of energy and effort should be aimed at securing donut shops as new Wi-Fi customers. It was also demonstrated that sandwich shops could also be a potential market, but one which should only be pursued once all efforts to sell Wi-Fi to donut shops had been exhausted. Lastly, we have concluded that efforts to obtain coffee shops as business clients should be the lowest priority, because the market is smaller and coffee shops have the least to gain by adopting Wi-Fi.