## Capstone Project - The Battle of Neighborhoods (Week 2) A full report consisting of all of the following components (15 marks): INTRODUCTION

Where you discuss the business problem and who would be interested in this project:  
In a city of my choice, Columbus, Ohio, if someone is looking to open a restaurant,  
where would you recommend that they open it?  
  
DISCUSSION OF THE BUSINESS PROBLEM:  
I am building a data analysis report for an entrepreneur who wants to open a restaurant here, in a Columbus suburb  
  
The business requirements being:

1. Opening a restaurant at a few of the wealthiest suburbs in and around Columbus, Ohio.
2. A residential suburb is preferred since it is a family restaurant.
3. The Food Company Inc. , though is very well funded, still needs to choose the perfect mix of audience location to start its first venture. If this is successful, they are planning to replicate the same in more suburbial locations.
4. Starting this undertaking on a relatively smaller budget since this is a pilot attempt  
   hence budgeting is quite a key factor.

WHO WOULD BE INTERESTED?  
This data analysis exploratory report is made for small business owners who are willing to try a pilot system of opening a family friendly restaurant offering healthier meals. Opening a restaurant is a considerable undertaking and, when done thoroughly and thoughtfully, can be a rewarding and profitable experience. By carefully following the steps outlined here, the business owner will be on his/her way to opening their doors for customers who desire their culinary services.

First move is very important, thereby choice of location is very important and hence I, as a data analyst have done my exploratory analysis in picking the best location that satisfies the business recommendations. At the end of my project presentation & discussions, the entrepreneur would be presented with picking one of the locations in either of these two suburbs-

**Dublin, Ohio or New Albany, Ohio**  
DATA where you describe the data that will be used to solve the problem and the source  
of the data.  
For picking the perfect location for the small business entrepreneur, the following data sources  
have been used in my analysis:  
I shortlisted two specific suburbs -Dublin and New Albany  
I collected the following data with respect to each of these suburbs

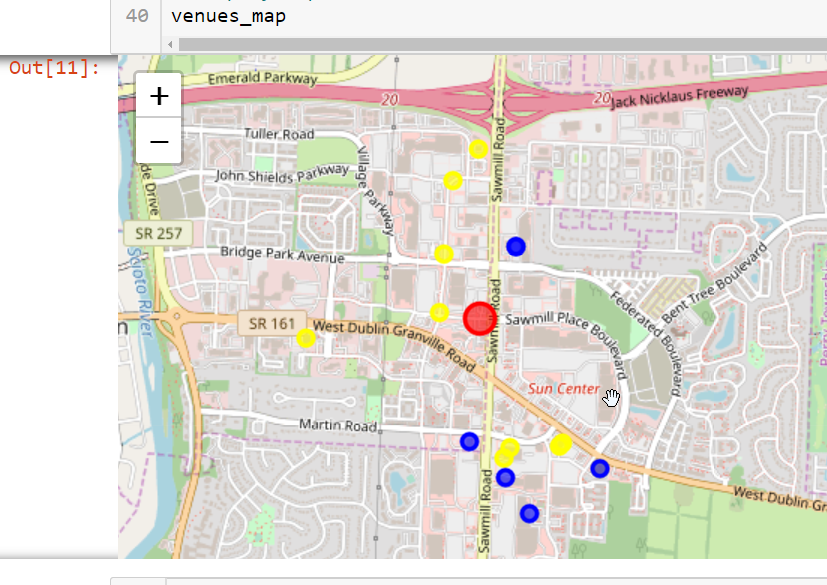
1. Demographics
2. Average Household Income
3. Availabilty of similar other restaurants in the vicinity
4. Average Retail Spending patterns in each of the two suburbs.
5. General eating and lifestyle options of the population in these two places  
     
   <http://worldpopulationreview.com/us-cities/dublin-oh-population/>  
   <http://worldpopulationreview.com/us-cities/new-albany-oh-population/>

Appropriate average household income for both these neighborhoods was obtained from the above mentioned link as well.  
  
<https://www.census.gov/quickfacts/fact/csv/newalbanycityohio,dublincityohio/PST045218>  
<http://zipatlas.com/us/zip-code-comparison/percentage-indian-population.htm>

With each of these links, we used FourSquare API to get to the URL, which inturn gave us only the desired values that can be used directly for data analysis.

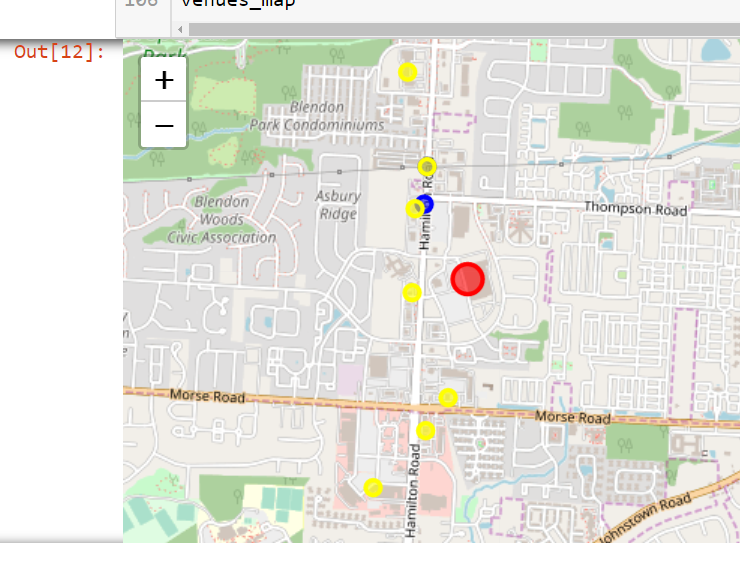
METHODOLOGY section which represents the main component of the report  
where you discuss and describe any exploratory data analysis that you did,  
any inferential statistical testing that you performed, if any,  
and what machine learnings were used and why.  
GET THE APPROPRIATE DATA FROM RELIABLE SOURCES,DO THE NECESSARY DATA WRANGLING TO MAKE THE DATA USABLE FOR MY ANALYSIS HERE and SHOW MY RESULTS ON AREA SPECIFIC MAPS  
Initially, I used the latitude & longitude of Dublin with Foursquare API calls to get the URL.  
From the URL, I converted those values to JSON. Then, these into a dataframe, that gave me 10 restaurants in all and 6 Indian restaurants in Dublin.

**DUBLIN MAP of Restaurants**

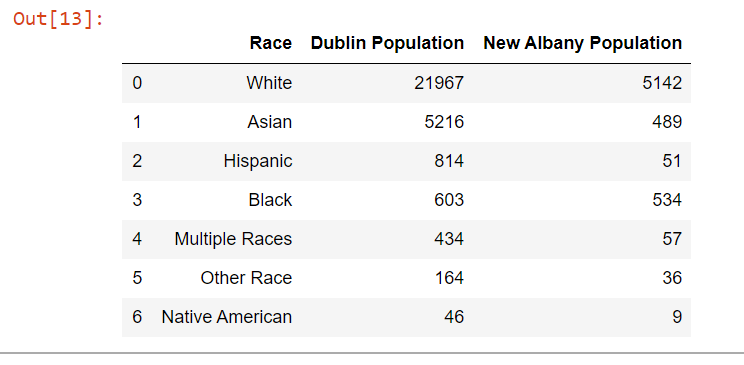
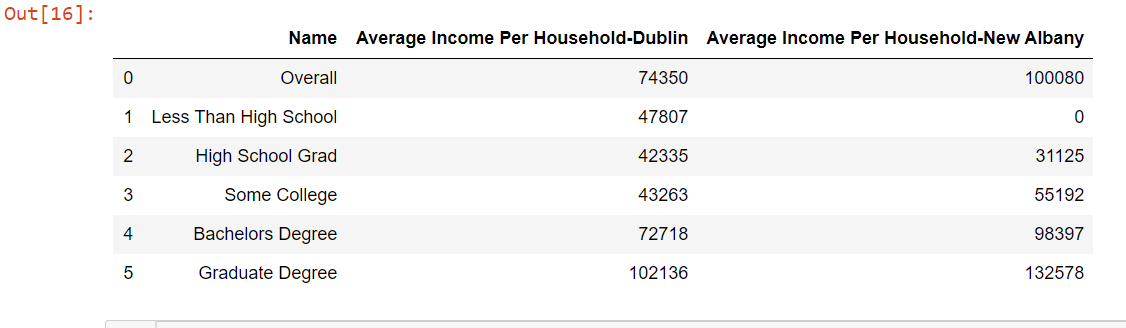


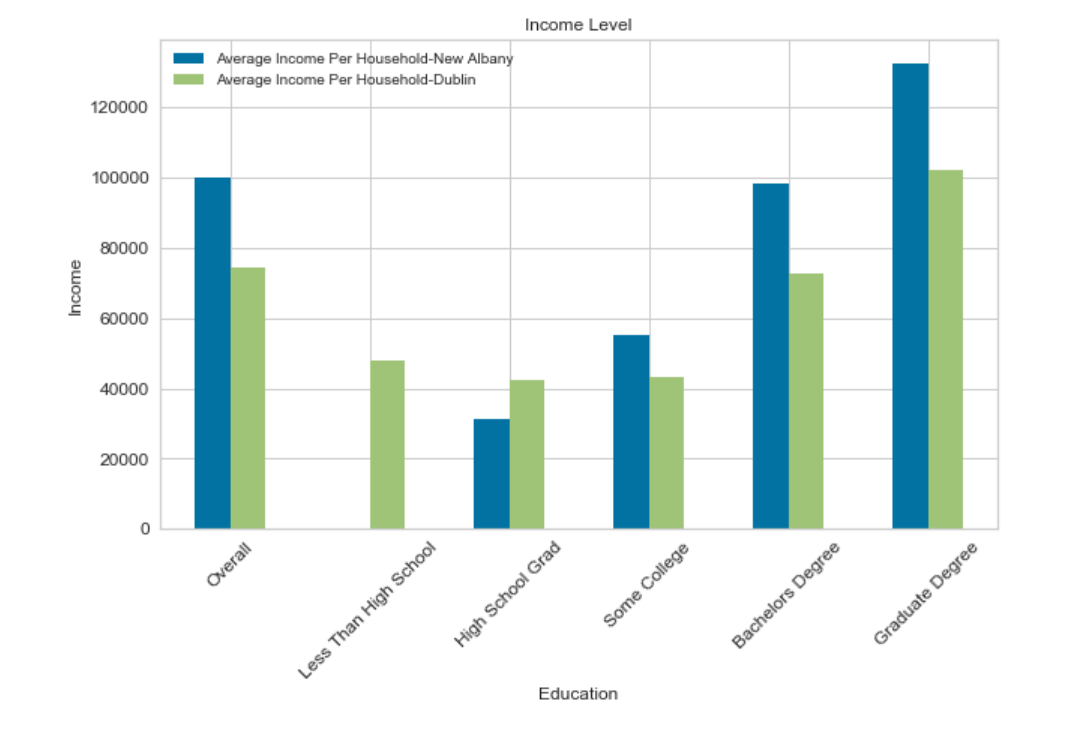
Then, I put these on a map -the 'new' one marked in red, other ones in 'Yellow’, Indian ones are in 'Blue'.  
Similarly the identical methodology for New Albany(NA) that leaves us with only 1 Indian restaurant in New Albany with 7 other surrounding restaurants.

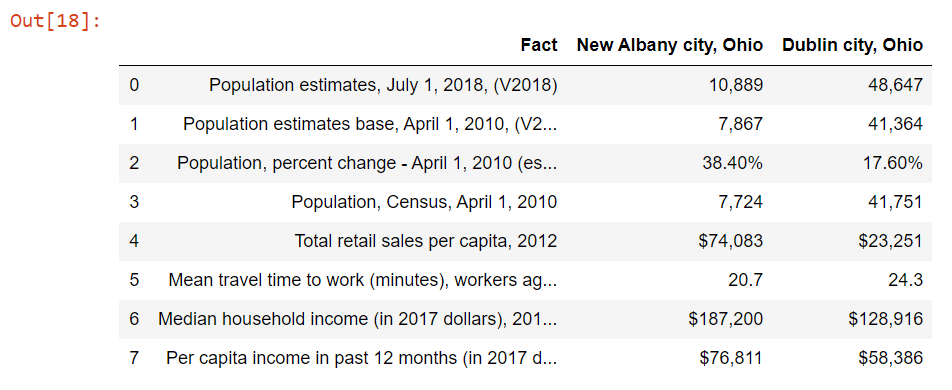
**NEW ALBANY MAP of Restaurants**

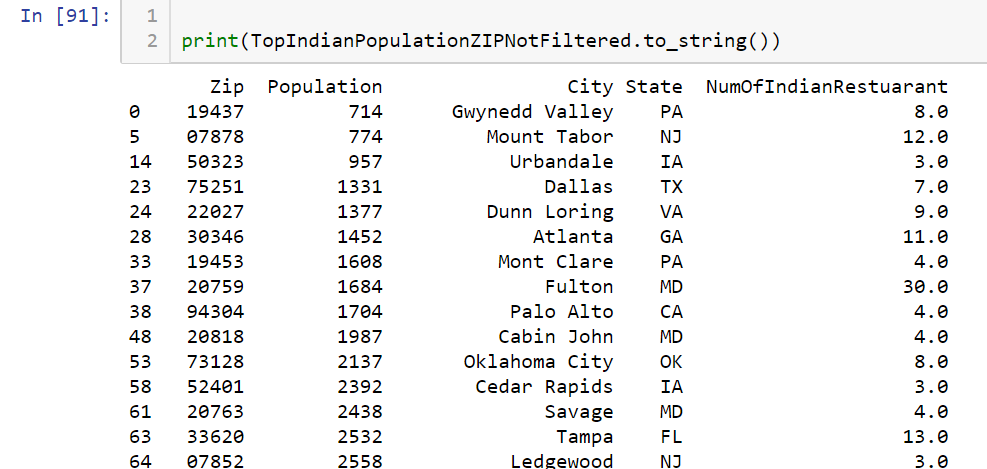
  
Next, the population comparision on both these suburbs. Got the Json data from the website  
<http://worldpopulationreview.com/us-cities/dublin-oh-population/>  
url = '<https://raw.githubusercontent.com/seethaparamesh/github-example/master/Dublin%20Demo%20data.json>'

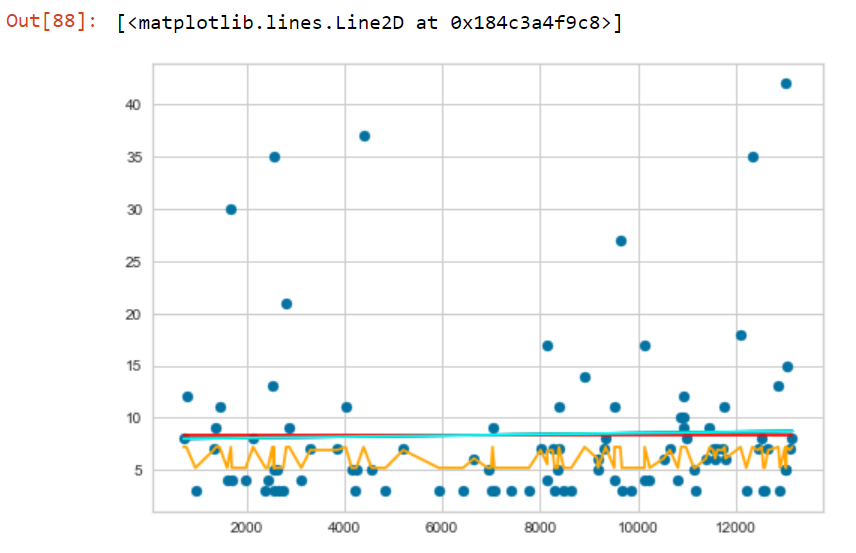
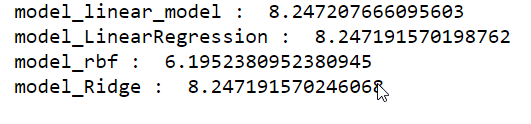
<http://worldpopulationreview.com/us-cities/new-albany-oh-population/>  
url = '<https://raw.githubusercontent.com/seethaparamesh/github-example/master/New%20Albany%20Demo.json>'

Converted these json into single dataframe each, one for NA & one for Dublin and then merged the 2 df.  
  
Getting to the business requirement of opening a restaurant in 'one of the wealthiest neighborhoods’, I went ahead and created a dataframe from the average household income data json that i got from the following websites.  
  
url = '<https://raw.githubusercontent.com/seethaparamesh/github-example/master/Dublin%20Income.json>'  
  
  
  
url='<https://raw.githubusercontent.com/seethaparamesh/github-example/master/New%20Albany%20Income.json>'  
  


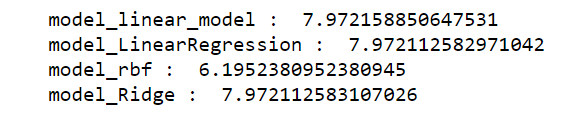
Comparison of the average income between the 2 suburbs in 1 dataframe  
  
Bar Chart depicting the average household income comparison in both the suburbs  
  
  
 Validate that analysis of data got from  
 [worldpopulationreview.com/us-cities/new-albany-oh-population/](http://worldpopulationreview.com/us-cities/new-albany-oh-population/)

matches a secondary source.  
# US Census data got from  
<https://www.census.gov/quickfacts/fact/csv/newalbanycityohio,dublincityohio/PST045218>  
 Snapshot of the population growth  


Based on Census data-  
Population of New Albany is growing at 38.4% in the past 8 years as compared to only 17.6% for Dublin.  
Average medium income from Census data also match the previous analysis that New Albany is more affluent for high earners.  
Per Capita of retail sales also show that Consumer Expenditure is significantly higher in New Albany as compared to Dublin.  
  
Thus ends the initial analysis of the two suburbs that the business owner would be interested in.  
  
FUNDAMENTAL DISTRIBUTION OF INDIAN RESTAURANTS ACROSS INDIAN POPULATION IN THE US  
  
Next, let’s do the fundamental distribution of Indian restaurants in a list of Cities all across the US.  
  
#<http://zipatlas.com/us/zip-code-comparison/percentage-indian-population.htm>  
  


Performed Data Wrangling and Data Cleanup  
  
  
Now, we do the similar analysis of all the cities in the above mentioned df , iterating through all zipcodes from the above df.  
At each zipcode, I am calculating simultaneously the number of Indian restaurants.  
  
At this, Foursquare provides a maximum of only 50 venues per zipcode.Hence any zipcode that has a 'num' value of 50 and 0 are considered outliers(this data would pretty much be irrelevant to us) and have been excluded from the data analysis.  
After these outliers have been dealt with, I have made a scatter plot to show the distribution of  
the number of Indian restaurants against the total Indian population in each of the zipcodes of cities  
that we have considered in the table above.  
This data would be used as the training data and I have used the following supervised learning models:  
Lasso  
Linear Regression  
ElasticNetCV  
SVRrbf  
Ridge  
   
   
 At the end of this data modelling & prediction, the SVRrbf turns out to be the best prediction model for this training data.  
The final point is to use these models on our test data under consideration- New Albany Vs. Dublin.  
Applying all these prediction models on both the location gives us the conclusion that  
New Albany has better prospects to be successful than Dublin for opening a new Indian restaurant.  
 Predation based on Dublin Population  


 Predation based on New Albany Population



RESULTS  
From all the analysis above, in between the two suburbs in question, opening a restaurant  
in New Albany has better chances of survival than in Dublin that can be seen from the  
factors considered above.  
  
DISCUSSION section where you discuss any observations you noted and any recommendations  
you can make based on the results.  
My recommendation would be to start off the restaurant in New Albany as compared to Dublin  
because of various major factors like the ratio of people to restaurants, the lack of many similar cuisines In New Albany, the average spending pattern/power of the New Albany Residents  
and relatively growing population since New Albany is only an up and coming  
village compared to the already established city of Dublin.  
  
CONCLUSION section where you conclude the report.  
The report thus concludes that for an entrepreneur with the above mentioned business  
requirements, New Albany, Ohio would be a very good option to start off his new  
restaurant on the basis of the exploratory data analysis, and statistical inferences.