**Project Report**

**Location Based Revenue Prediction**

**Contents:**

1. **Introduction**
2. **Data Required**
3. **Methodology**
4. **Discussion and Future Scope**
5. **Conclusion**

# **Introduction:**

In recent times, for any restaurant to gain successful opening along high revenues, it is imperative to choose a location that fits the culinary tastes of their neighbourhood. Similarly, large restaurants chains like McDonalds, Dominos, KFC etc. looks for various parameters before choosing the location for their next franchise. In this context, the data-driven decision making can be a boon to these restaurant chain’s management. It is possible to identify the next location of the franchise with high potential revenue based on the revenue of the franchises as existing location.

In this project, I propose a solution which predicts the revenues of the pizza restaurants based on the features of the locality in which the restaurant is set to open.

# **Data Required:**

For this project, I will use to Foursquare API to get the location-based data and explore the various venues available nearby the restaurants. For the purpose of proof of concept (POC), I will manually populate revenues of different venues to train the model. These revenues will be randomly generated and may differ considerably if the project is implemented on a real-world dataset with actual revenue figures.

I used the Datafiniti dataset from Kaggle containing information about the pizza restaurants in USA such as address, location data, menus, average price of pizza etc. I used foursquare location API to retrieve nearby venues around a restaurant.

Link to the dataset: <https://www.kaggle.com/datafiniti/pizza-restaurants-and-the-pizza-they-sell>

I will populate a dataframe consisting of various venue categories around a restaurant along with its randomly generated revenue. I will then use regression models to train and make predications based on the above data.

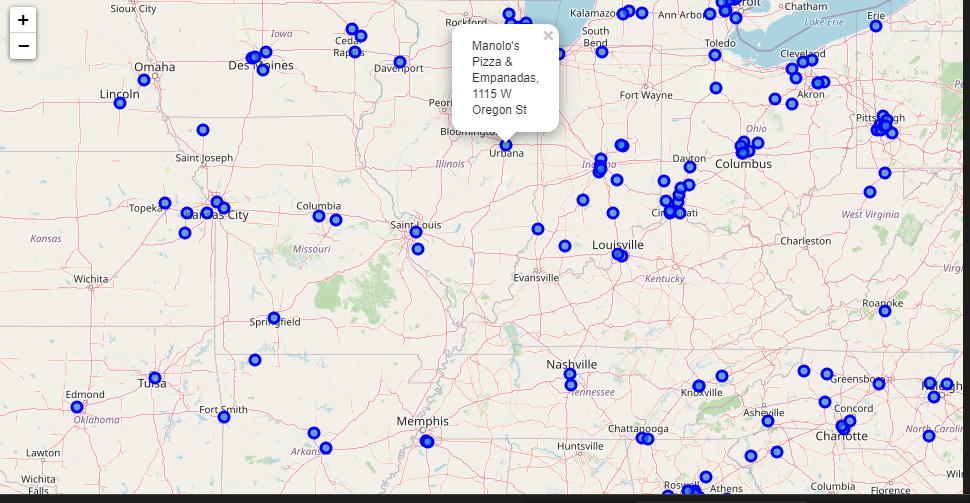
# **Methodology:**

1. **Data Normalization:**

The dataframe created should be normalize so that it can be fitted in regression models. This is to ensure that the gradients are not subsided in the regression algorithm. There were a total of 300+ venue categories and we applied one-hot encoding to the venue category columns. This gave me the dataframe of size (m,352) of which majority of the values are dominantly zero. Then I grouped the rows by Restaurant names and find the mean to normalize the data.

1. **Data Visualisation:**

I used the folium data visualization library to plot venues on a map using their lat-long coordinates.



1. **Regression Model:**

For this project, I used XgBoost Algorithm. XGBoost is a scalable and accurate implementation of gradient boosting machines and it has proven to push the limits of computing power for boosted trees algorithms as it was built and developed for the sole purpose of model performance and computational speed. Specifically, it was engineered to exploit every bit of memory and hardware resources for tree boosting algorithms.

Configuration:

alpha=10, base\_score=0.5, booster='gbtree', colsample\_bylevel=1,

colsample\_bynode=1, colsample\_bytree=0.3, gamma=0, gpu\_id=-1,

importance\_type='gain', interaction\_constraints='',

learning\_rate=0.1, max\_delta\_step=0, max\_depth=10,

min\_child\_weight=1, missing=nan, monotone\_constraints='()',

n\_estimators=100, n\_jobs=0, num\_parallel\_tree=1,

objective='reg:linear', random\_state=0, reg\_alpha=10, reg\_lambda=1,

scale\_pos\_weight=1, subsample=1, tree\_method='exact',

validate\_parameters=1, verbosity=None

# **Discussion and Future Scope:**

This project can be considered as a POC for a real life business implementation which can be useful to accurately predict the revenue of a restaurant. Additionally, it can be designed to suggest the locations for the restaurants to maximize the profit. The accuracy and predictions of the model can be improved by data enrichments using the population and demographics of the location. Moreover, revenue also depends on the variety of pizzas served, their ratings in the popular sites like Zomato etc. These improvements can guide further development of this model.

# **Conclusion:**

This was the first end to end ML project developed by me. There is a lot of room for improvement and your feedback is really appreciated and will help me motivate to continue my venture into the field of Machine Learning.