Motivation:-

It is generally considered that the more spending power of the people the more the products that are offered. This leads to the vague conclusion that higher the income (which in turn means more spending power of the people), more the number of restaurants (more products in which people can spend their money on). However if the world were to run based on such vague conclusions it would be a different ball game altogether. Hence in this project I embark on a mission to try and prove this hypothesis.

Brief description of data sources:-

Source 1 - All ZIP codes in Los Angeles
https://www.zip-codes.com/city/ca-los-angeles.asp

From the above link I scrape all the ZIP codes where the 'Type' column is 'Standard' and the 'Population' column has a value greater than 0 (Doing this gives us the list of all inhabited ZIP codes in Los Angeles).

Source 2 - Median, average and per capita incomes based on ZIP code
https://www.incomebyzipcode.com/california/90089 - An example link for the statistics of the 90089 ZIP code

From the above link corresponding to each of the ZIP codes scraped from source 1, we scarpe the median, average and per capita incomes.

Source 3 - Number of restaurants based on ZIP code using Documenu API
https://rapidapi.com/restaurantmenus/api/documenu/details - The description of the API can be found here.

Using the API, corresponding to each ZIP code we get the number of restaurants.

Analysis:-

After gathering our complete data from the sources described above, we consider every independent variable (ZIp codes, Mean incomes, Average incomes, Per capita incomes) along with the dependent variable (Number of

restaurants) and perform statistical T-Test to check if the dependent variable actually depends on the independent variables. To check this, we compare the p-value which is outputted by the model summary of statsmodel API . If the p-value is greater than 0.05 (The standard value for this threshold) we consider the Number of restaurants to be independent of the feature for which this test was performed. The data points along with the regression line is also plotted for all features individually and all features together. These graphs are saved too and have been displayed below.

Simple Linear Regression for median_incomes

Simple Linear Regression for per_capita_incomes
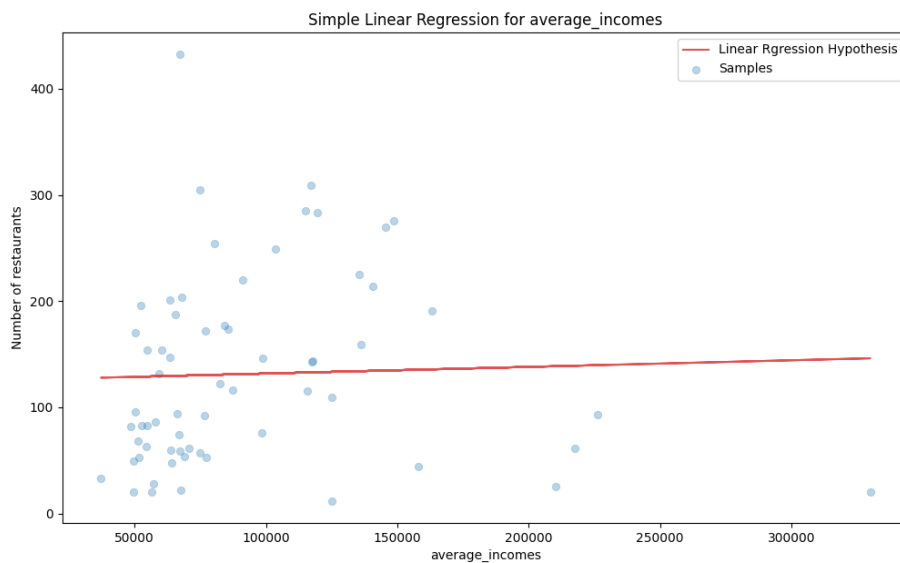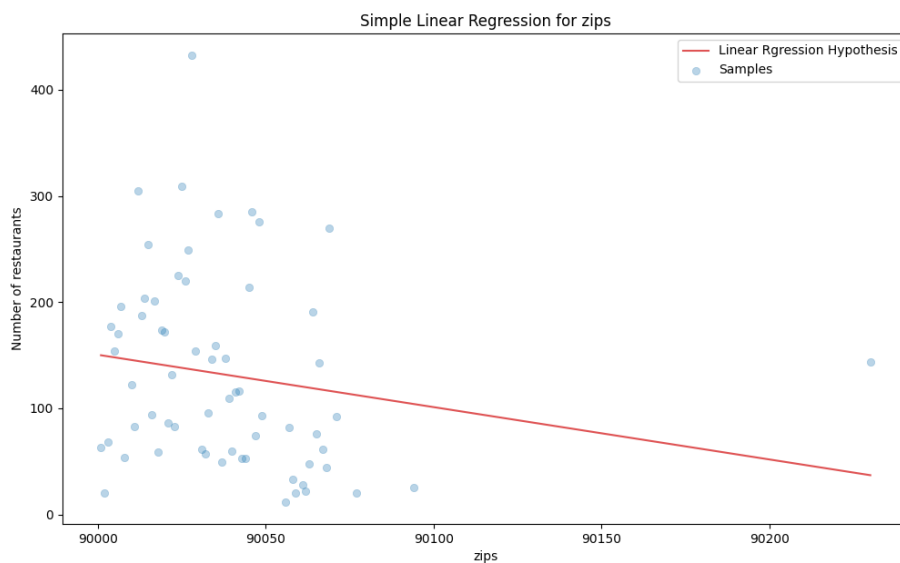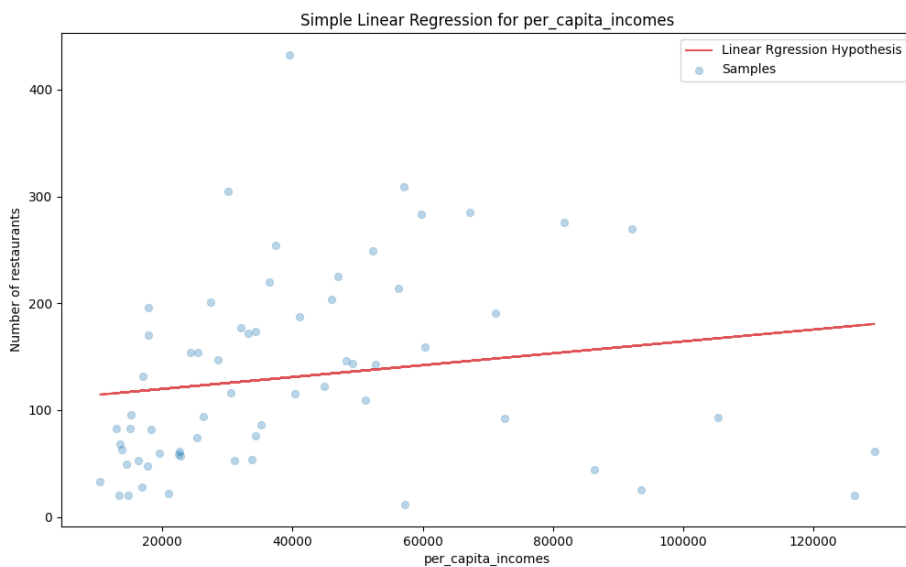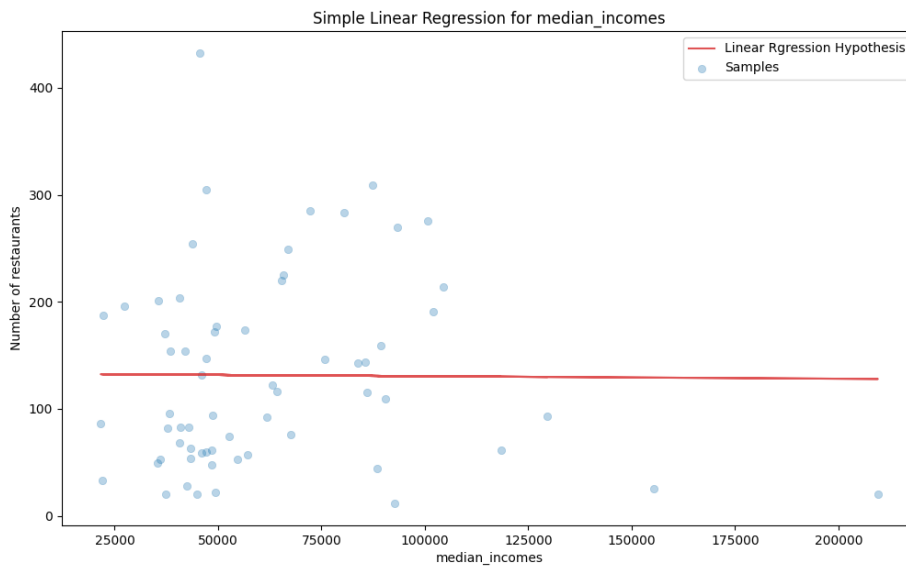
<u>Conclusions:-</u>

Since the p-values for all individual variables is greater than 0.05, we can conclude that the number of restaurants does not depend on either ZIP Code, Median income, Average income or per capita income. Hence we do not make a predictive model as none of the features are useful in predicting the number of restaurants.