Inferior Goods:

The Effects of Fast Food Restaurant Density on Income

Ali Shalaby

Prof. Khazra

ECO225

April 16, 2022

1. Introduction

It is no question that food consumption and expenditure generates a noticeable portion of the United States' GDP. To be exact, consumer expenditure on food accounts for 6.4% of the US's GDP in 2016 and is expected to grow at a 2% rate each year (Gray, 2016). The proportion of food consumption that is allocated to fast food accounts for 36.6% (Fryar et al., 2018). Due to its low production cost and great taste, fast food is sold at a low price, further increasing its vast popularity. According to specific assumptions made in microeconomics, fast food is said to be an inferior good; meaning as an individual's income increases, the quantity of purchases that are made on fast food will decrease. This relationship is otherwise known as the income effect. In the case of inferior goods, the income effect is negative. Many factors can determine the natural behavior of a good, For example, an inferior good is usually cheap in price and low in quality. It is important to explore other causal factors that can act as predictors for income, allowing for future implications regarding more accurate forecasts on income per capita on a quarterly basis. Restaurant Density is the proportion of restaurants to population. This shows the concentration of restaurants in a given area. After some data cleaning, manipulation, and visualization, the data showed that restaurant density seemed to have a moderate correlation with disposable income per capita. A study in 2004 has shown that the obesity of individuals in low income areas was found to not be correlated to the number of restaurants near them (Burdette & Whitaker, 2004). Further implications from this study show that the data did not experience a significant relationship between fast food restaurant density for obsese individuals and income. Because fast food purchases have become more prevalent in the past decade, it is essential to analyze this relationship more specifically. Also, this study will focus more on the predictor itself, rather than its implicative health concerns.

Firstly, in order to study this relationship more specifically, burger restaurant density was found and an even stronger negative relationship with income per capita was observed. Thus economic implications seemed to have some responsibility for the correlation between fast food restaurant density(independent variable) and income per capita(outcome). Through API scraping from the Bureau of Economic Analysis, I was able to import data on takeouts per capita, non durable and durable consumption per capita. All used as predictors for disposable income per capita. After analyzing the separate correlation each IV had with income per capita, it was evident that three relationships were positively correlated, however, the relationships were either moderate or weak. This paper will showcase findings of restaurant density through the use of web-scraping, data visualization, mapping, machine learning and summary statistics. It will show the accuracy of this predictor with respect to income per capita, and conclude whether or not this variable can be added as an exogenous variable to the production identity function where income is equal to the summation of household consumption, investment, government expenditure, and net exports.

2. Data

The original dataset was imported from Kaggle and is a list of 10,000 fast food restaurants where each restaurant contained data for name of restaurant, address, category of fast food, city, state, source URL, latitude and longitude, postal code and company website. In order to explore the relationship between restaurant density and income per capita, some data cleaning had to be done.

2a. Income Data

It is evident that the outcome variable income per capita, is not present in this dataset. Therefore, I imported a dataset(x) from the Bureau of Economic Analysis that contains the per capita disposable income in 2019 grouped by state. This data set also contained the population and average disposable income for each state.

2b. Number of Restaurants

Once this dataset was imported, it is essential to find the number of restaurants grouped by each state and create a new data frame (y) that contains values of a given state in the United States and the number of restaurants that state contains. Data frames x and y were merged forming one dataset containing data for disposable income per capita, population, the number of restaurants and the population; where all variables are grouped by state.

2c. Restaurant Density

Finally, restaurant density can be calculated by means of dividing the number of restaurants by the population for each state. The 'restaurant density' variable can be added into the existing dataset. Moreover, to explore the behavior of fast food circulation throughout an economy, I found the restaurant density of all restaurants that served burgers.

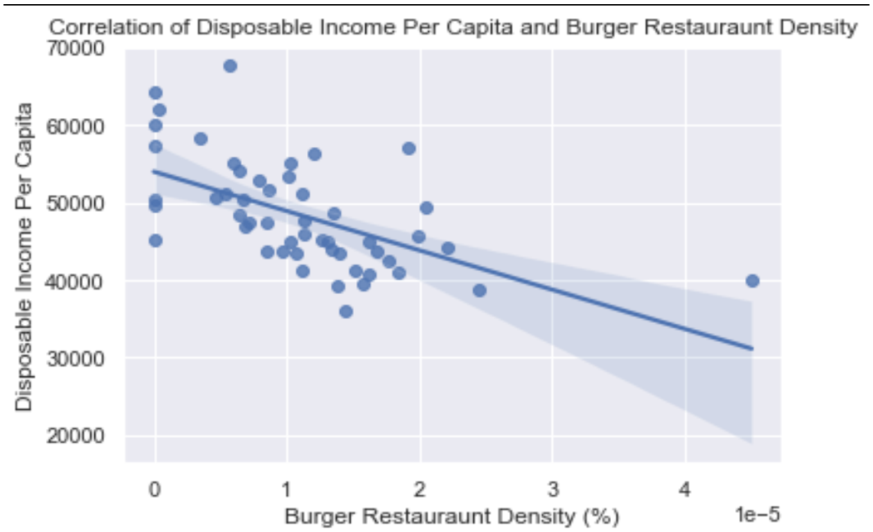
2d. Cafe Density

Cafe density was introduced to further explore the relationship and accuracy of burger restaurant density and income per capita. The negative relationship between burger restaurant density and income per capita means that as disposable income per capita increases in a given state, there will be a smaller density of burger restaurants within that state. If this hypothesis holds, because products served at cafes are not inferior goods, the correlation should be relatively weaker. If the relationship is not as negative or possibly even positive, there is more evidence that economic intuition is the main driver between such correlations.

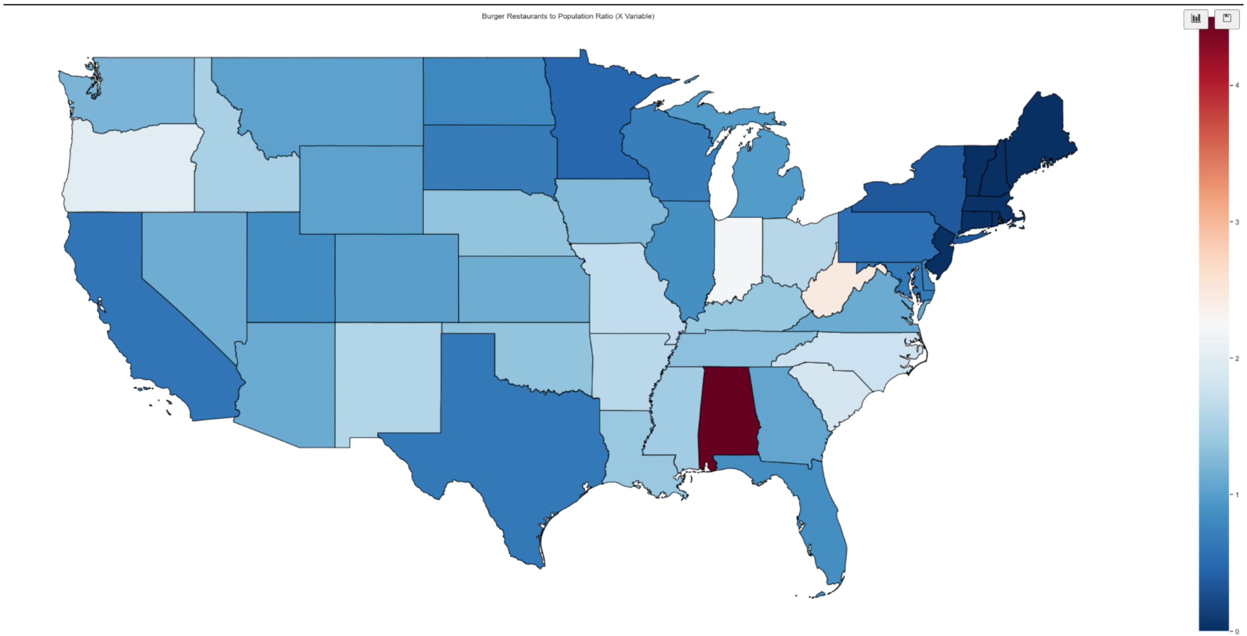
3. Summary Statistics

3a. Burger Restaurant density

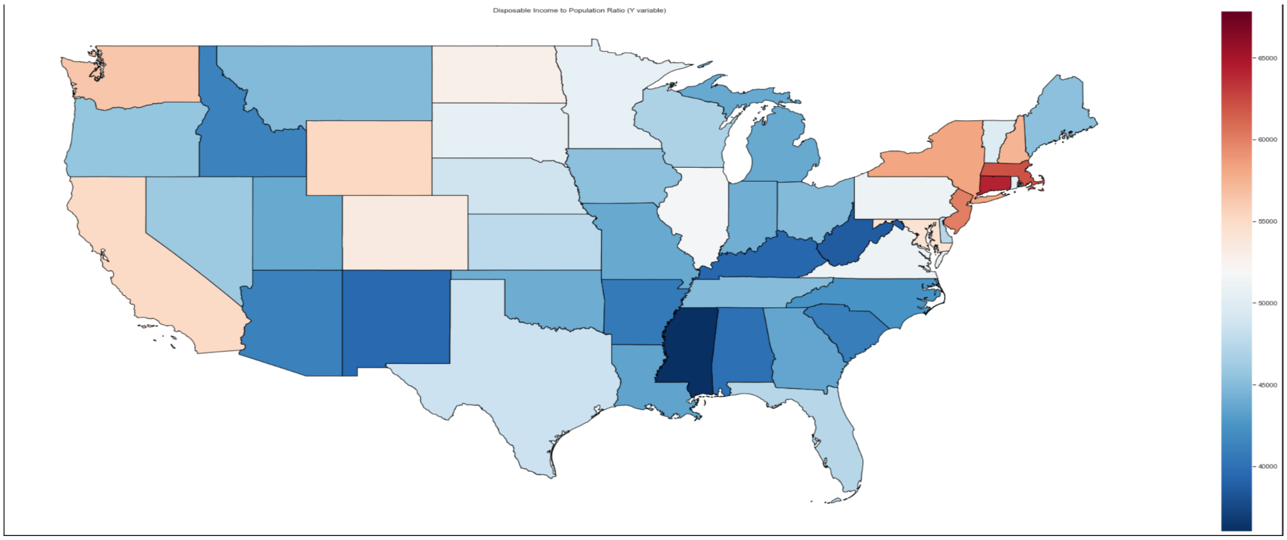
Although the original dataset imported from Kaggle represents 10,000 fast food restaurants across the United States, some restaurants did not necessarily serve fast food. For example, the dataset contained restaurants that served cafe goods, ethnic food, and healthy-marketed food. Therefore, to explore the true relationship of fast food restaurant density to income per capita, I filtered the data on all fast food restaurants that served burgers, a universally considered inferior good. After finding burger restaurant density, I generated a scatter plot showcasing the correlation between disposable income per capita and burger restaurant density.



It is evident that a moderate negative correlation is present between both variables. The correlation coefficient is -0.58. As predicted through economic intuition, this relationship would be negatively driven by the fact that fast food is an inferior good; thus, in high income states, it is more likely to have a smaller concentration of fast food restaurants. However, the data shows that the strength of this relationship is not strong, thus, more analysis must be done in order to confirm this IV’s accuracy on predicting income per capita. More visualizations must also be necessary in order to analyze this relationship. For instance, the maps generated below both represent the United States where the color is grouped by state. The first map showcases the burger restaurant density in each state.



This is represented by color as the range plot shows the degree of burger restaurant density by color. When analyzing the plot, it is evident that there is high burger restaurant density in Alabama meaning there are many burger restaurants within a closer radius in this state. Intuitively, Alabama has a relatively higher demand in fast food than other states as fast food chains are concentrated throughout the state. The second map showcases the disposable income per capita grouped by state.

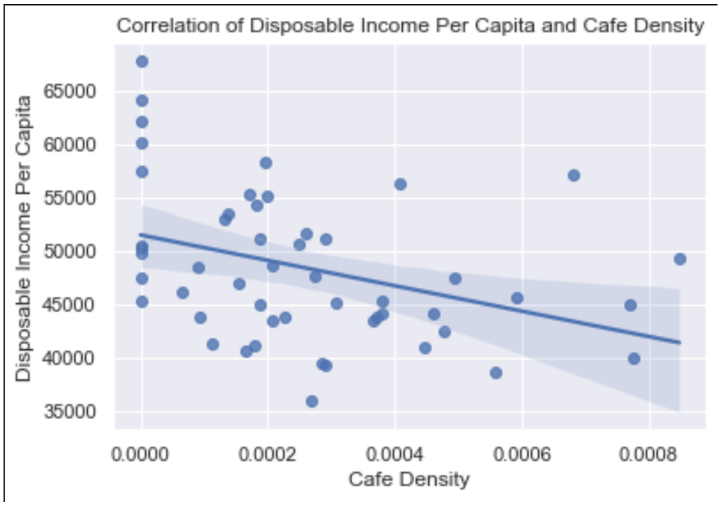


This is represented by the color as the range plot shows the degree of income per capita by color. In this map, there are more outliers. For example, in the northeast region, states such as New York, Rhode Island and Connecticut all have relatively high income per capita. This corresponds to very low burger restaurant density. Another example of outliers are Washington, California, and Wyoming, all having relatively high income per capita. This also corresponds to the first map as those states are represented as relatively darker colors, meaning relatively lower burger restaurant density.

These are just a few examples, however, in general, according to both maps, it is evident that a state with high income per capita, represented as a brighter color, will have a lower burger restaurant density in the other map represented as a darker color.

3b. Cafe Density

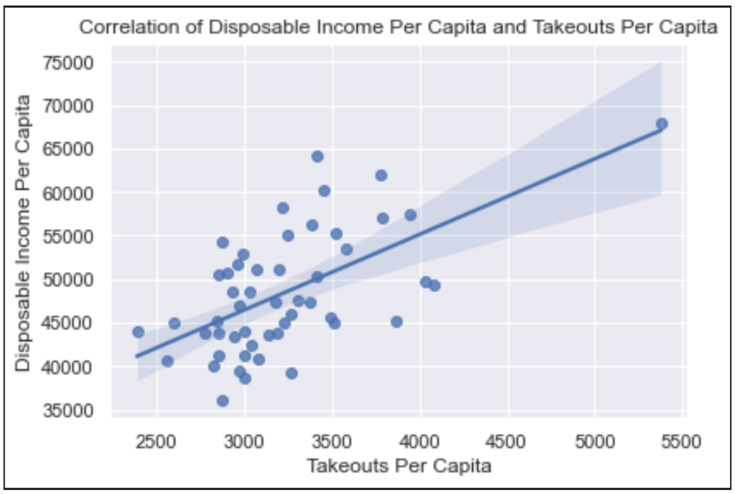
As stated previously, this variable is simply used as a control group in order to analyze the link between economic intuition and data correlation. The graph below represents the correlation between cafe restaurant density and disposable income per capita.



When analyzing the relationship between cafe density and disposable income per capita, the relationship is very weak. Data shows that the correlation coefficient is -0.27. Intuitively, this makes sense. Because cafe goods are usually expensive, they are not considered to be inferior goods. However, they are not luxury goods either, thus implying this very weak negative relationship with disposable income per capita. The estimation of the strength of this relationship is represented by the line. It is evident that the line is almost horizontal, meaning the slope of the linear approximation is close to zero, ultimately representing the very weak relationship between both variables.

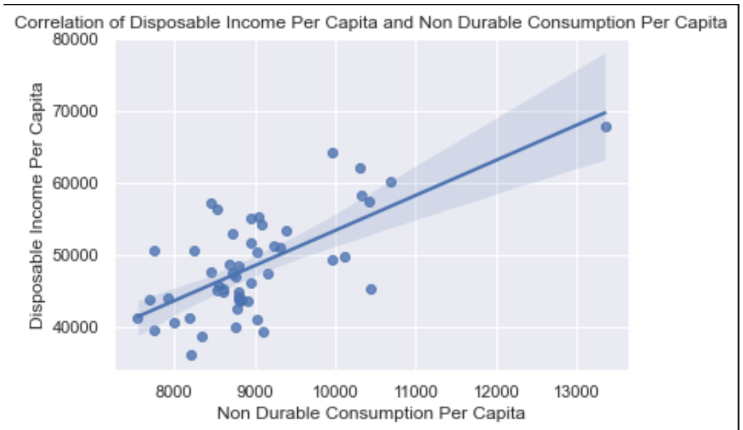
3c. External API Data

In order for regression and ML purposes, a model must be created in order to predict income per capita and burger restaurant density is just one independent variable. In order to explore in more depth this relationship, other factors must be considered. I scraped from the Bureau of Economic Analysis using their provided API which included three datasets of interest that may be properly coupled with burger restaurant density. Note that each dataset is grouped by state. (1) represents the amount of takeouts per capita in dollars in a given year, (2) represents the amount of non durable consumption in dollars in a given year, and (3) represents durable consumption in a given year. Merging these three datasets to my original dataset will enhance my research as integrating these three independent variables with burger restaurant density and creating a model around predicting income per capita may interact strongly with each other, resulting in more accurate predictions. These variables' relationship with income per capita are backed up by economic intuition and it is viable to analyze if this matches the data. Firstly, takeouts per capita could be a good indicator of representing inferior goods. The plot below represents the correlation between takeouts per capita and disposable income per capita.



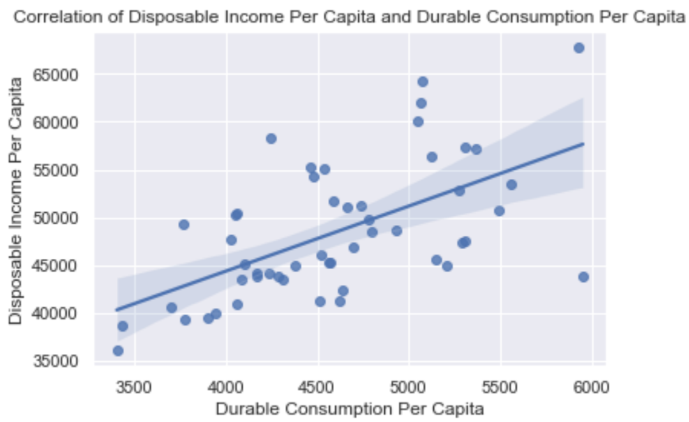
There appears to be a moderate positive correlation between disposable income per capita and take outs per capita. Technically, this means that as average disposable income per capita increases in a given state, then the average takeouts per capita will also increase. This contradicts economic intuition where there would be a negative relationship between both variables; however, take outs could range from fast food to expensive food. This may explain the positive correlation as takeouts per capita is a very broad category when using it as a predictor for disposable income per capita. The data suggests that the correlation coefficient between both variables is 0.6, suggesting that the individual relationship is positively moderate.

The next independent variable is non durable consumption per capita. This variable could also closely resemble inferior goods as non durable goods are goods that are consumed for every day use. The plot below represents the correlation between non durable consumption per capita and disposable income per capita.



Data shows that the correlation coefficient with respect to income per capita is 0.67. Visually, there appears to be a moderate positive correlation. This means that as average disposable income per capita increases in a given state, the average consumption of goods that are non durable will also increase per household. Non durable consumption can range from goods that are inferior to luxurious; Also an important consideration is many goods that are basic necessities such as toothbrushes, cleaning supplies, food, fall under this category. Therefore it makes sense that the more a given household makes, the more they will consume in such necessities. The correlation may not be strong, possibly due to the permanent income hypothesis; where households tend to smooth their consumption over time based on their marginal propensity to consume. This means households may not consume as much as they can and instead will save a portion of their income in order to smooth consumption over their lifetime.

Finally, the last IV is durable consumption per capita. Durable goods are goods that don't have to be purchased often and last for an extended amount of time. The plot below represents the correlation between disposable consumption per capita and disposable income per capita.



The data shows that the correlation coefficient with respect to income per capita is 0.59. Visually, there appears to be a moderate positive correlation between disposable income per capita and durable consumption per capita. Theoretically, this means that as average disposable income per capita increases in a given state, the average consumption of goods that are durable will increase. Durable goods are goods that last for a longer period of time. Many appreciating assets fall under this category with the exception of vehicle purchases. It would seem to make more sense if there was a stronger correlation than non durable consumption as many durable goods tend to cost more. However, due to the PIH and depending on the household's marginal propensity to consume, it may explain the strength of this correlation as many households tend to reinvest savings.

4. Results

Although visualizations and statistical summaries are useful in understanding a bivariate relationship, it is essential to further explore the accuracy and interactions of such independent variables with respect to the outcome.

4a. OLS regression

In order to measure the accuracy of any linear regression, it is essential to analyze the p value, f-stat and the R squared value. Analyzing each of these measures is viable to conclude the accuracy of the models. If an IV has a corresponding p value that is greater than 0.05 (rejection rule), that IV does not significantly predict the outcome. An f-statistic is a value between 0 and 1 and compares the joint effect of all variables together. A large f-stat means something in the model is significant and corresponds to a lower p-value. The probability of the f-statistic is the p-value of the entire model. Therefore, if this value is lower than alpha = 0.05, then the model is statistically significant. Finally, r squared shows how well the model fits the correlation on each point. Below, is the OLS regression results of income per capita, given predictors, takeouts per capita, burger restaurant density, durable and non durable consumption per capita.

OLS Regression Results

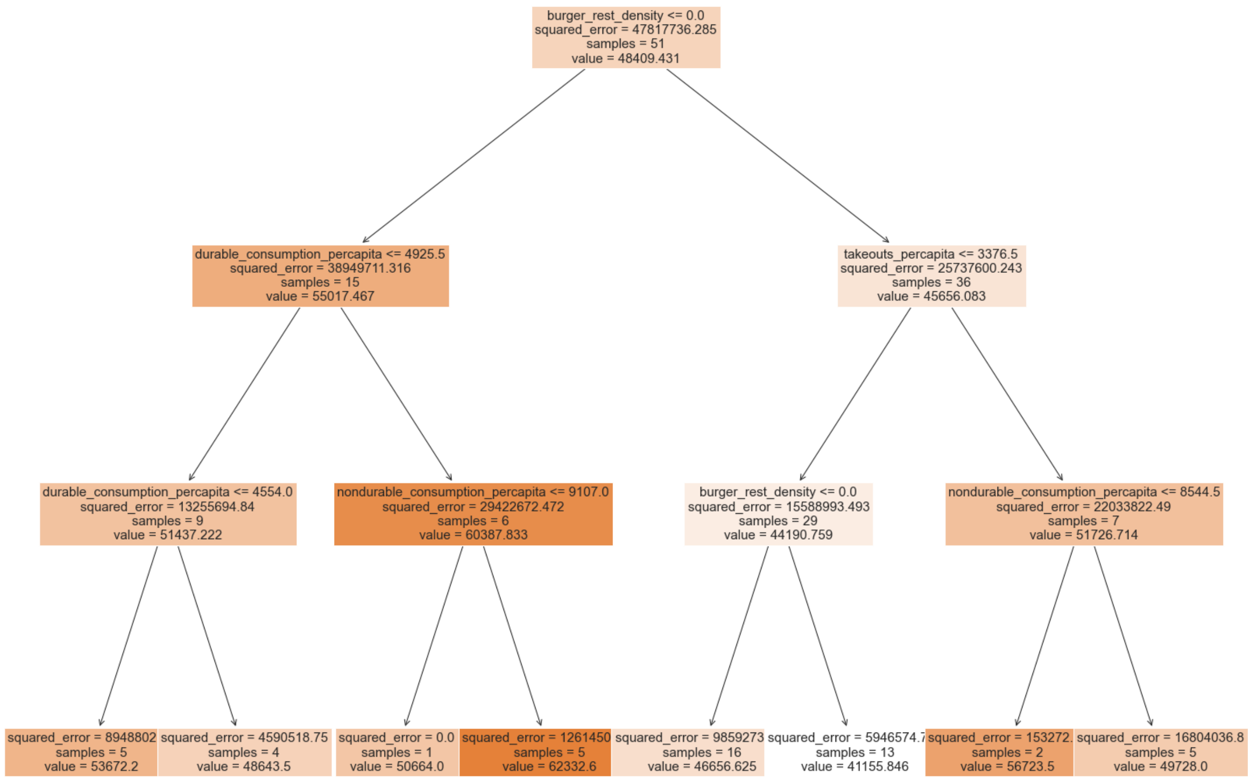
Dep. Variable:	Per capita disposable personal income (dollars) 2/	R-squared:	0.649
Model:	OLS	Adj. R-squared:	0.618
Method:	Least Squares	F-statistic:	21.24
Date:	Fri, 15 Apr 2022	Prob (F-statistic):	5.63e-10
Time:	23:12:11	Log-Likelihood:	-496.60
No. Observations:	51	AIC:	1003.
Df Residuals:	46	BIC:	1013.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	3171.5018	8437.767	0.376	0.709	-1.38e+04	2.02e+04
takeouts_percapita	0.2271	2.312	0.098	0.922	-4.427	4.881
burger_rest_density	-2.189e+08	9.1e+07	-2.406	0.020	-4.02e+08	-3.58e+07
nondurable_consumption_percapita	3.2816	1.154	2.845	0.007	0.959	5.604
durable_consumption_percapita	3.7924	1.205	3.148	0.003	1.367	6.218

It is evident that the R squared value is 0.65. This means most of the variation in the data can be explained by this regression model. Furthermore, the probability of the f-statistic is less than the rejection rule of 0.05 meaning the model itself is statistically significant. This also means that the joint effect of all independent variables included make predicting income per capita more likely at a given data point. Finally, the p values of three out of the four variables are statistically significant, however, takeouts per capita is not statistically significant. In general, integrating the burger restaurant density to the external data made the model more accurate in predicting the average income per capita in a given state.

4b. Regression Tree

It is suitable to use techniques under the umbrella of machine learning to further explore the model generated by OLS regression. This model includes the independent variables from the BEA data and burger restaurant density. In this case a regression tree will be utilized in order for more precise, clear and user-friendly predictions based on the inputs of variables. The output below is a regression tree using input variables, burger restaurant density, takeouts per capita, durable and nondurable consumption.



Given such independent variables, each cell tries to minimize the error of prediction. The last cell outputs the prediction given a certain path based on the condition of a sample of independent variables. Each branch splits the data in two groups. The first branch splits the data into two internal nodes: durable consumption and take outs per capita. Income per capita is predicted to be $55,017.47 if burger restaurant density is less than zero or $45,656 if burger restaurant density is greater than zero. If the first internal node, durable consumption per capita, is less than or equal to $4925.5, then income per capita is predicted at a value at $51,437.22. Otherwise, income per capita is $60,357.53. If the second internal node, takeouts per capita, is less than or equal to $3376.50, income per capita is predicted to be $44,190.76. Otherwise, income per capita is $51,726.71. Now we reach the second set of internal nodes, containing four nodes. The first two nodes are split by durable and non durable consumption per capita. If durable consumption per capita is less than or equal to $4554, then income per capita is predicted to be $53,672.2. Otherwise, income per capita is $45,643.50 If non-durable consumption per capita is less than $9107 then income per capita is $50,664. Otherwise income per capita is $62,332.6. In general, there is an increase in income per capita if non durable, durable, and take outs per capita increase. This was displayed when visualizing the correlation of each independent variable with respect to income per capita. The error of prediction or the Mean Squared Error (MSE) is extremely high at each node. The MSE is determined by the squared difference between the actual Y value and the predicted Y value. Therefore the metrics of the outcome variable is what drives the magnitude of the MSE. For example, this experiment's outcome variable is income per capita where each of its values are in the tens of thousands. Therefore, it is likely that the difference between the actual result and its predicted value will still be in the thousands; thus, the square of this difference will be in the millions.This is why the nodes in the regression tree all have extremely high errors of prediction. The MSE is a measure of accuracy but has its limitations in interpretation. Other factors must be considered in order to determine the accuracy and strength of a model such as the p-value, f-statistic, etc.

5. Conclusion

Analysts must always understand that correlation does not equal causation. Even if there were a 1 to 1 correlation between both variables, that does not mean that the independent variable is the causal factor for the outcome. Furthermore, there are many explanations that can hinder the strength of a correlation. Economically, factors in this case such as the personal income hypothesis, a household's marginal propensity to consume, and the state of the economy all contribute to the non-traditional patterns analyzed from the data. Going back to the moderate correlation between burger restaurant density and disposable income per capita, I cannot infer that this variable is a causal factor in the change in income. I can further explore this relationship by capturing more data, and use more specific data tools to get more accurate results; however, no matter how much predictive analysis is done, a prediction cannot be guaranteed. ML tools utilized such as OLS regression and the regression tree exposed the accuracy of this relationship. Therefore, it is evident that burger restaurant density is not a causal determinant to income per capita. This is evident due to a moderately low correlation coefficient, high MSEs, and moderate R squared values. However, fast food data and its relationship to national income can be further explored due to its backing by microeconomic theory.

One limitation experienced in this research was the data was aggregated throughout the entire project as values were grouped by state. Given the original dataset and when merging the BEA data on city instead of state, many data points were lost as there were only a few common cities between both data frames. Therefore, A more detailed dataset given could have proven for more accurate results. There was less data to work with, thus, statistical significance and analytical trends have risks of being misleading due to lack of data. Although the Central Limit Theorem states that a sample of $n = 30$ is enough for the data to converge to the parameter of interest, the data was aggregated at a very high rate in order to prevent data loss.

Works Cited

Alex Gray, S. W. (2016, December 7). *This map shows how much each country spends on food*. World Economic Forum. Retrieved April 16, 2022, from https://www.weforum.org/agenda/2016/12/this-map-shows-how-much-each-country-spends-on-food/

Burdette, H. L., & Whitaker, R. C. (2004). Neighborhood playgrounds, fast food restaurants, and crime: relationships to overweight in low-income preschool children. *Preventive medicine*, *38*(1), 57-63.

Fryar CD, Hughes JP, Herrick KA, Ahluwalia, N. Fast food consumption among adults in the United States, 2013–2016. NCHS Data Brief, no 322. Hyattsville, MD: National Center for Health Statistics. 2018