**The Impact of County Resident Population, Median Household Income and Unemployment Rate on the Gross Domestic Product (GDP) Growth in Montgomery County-MD**

Fokwen Bechem, Emmanuel-293370

Colorado State University Global Campus

MIS 581-Data Analytics and Business Intelligence Capstone

Dr Ford

February 21, 2021

**Abstract**

*The Gross Domestic* Product (GDP) of Montgomery County is a measure of the total dollar value of goods and services produced by the county usually over a period of 1 year. This economic metric is the most-watched-for by the county Economic Council and government when making economic policy decisions. Over the last decade from 2009-2019, Montgomery County’s GDP has been steadily increasing with an average annual growth rate of about 3%.

This GDP growth has been driven in part by changes in certain macroeconomic parameters, amongst which are rising resident population, median household income and falling unemployment rates. The impact of these macroeconomic indicators on the county’s GDP growth trends will be the focus of this study. Using numeric datasets of these explanatory variables sourced from the Federal Reserve Bank public data repository for the 10-year period under review, Montgomery county’s GDP will be x-rayed with a view to forecasting future GDP numbers. The research questions will highlight if there is any potential relationship between GDP (response variable) and county population, unemployment rate and median household income (predictor variables). The multiple linear regression model built for this analysis used 90% of the data for training while the remaining 10% was used for testing the model and to predict future GDP for the county. The statistical results showed a strong positive correlation between GDP growth and rising county population and median income. There was also statistical evidence of an inverse relationship between GDP growth and falling unemployment. This study was limited in scope and data collection. More numeric data covering a longer period of time, in conjunction with qualitative data input, plus other important economic variables like taxation, inflation etc not factored into the model could yield deeper insights.

**The Impact of Resident Population, Median Household Income and Unemployment rate on the Gross Domestic Product (GDP) Growth in Montgomery County-Maryland**

**Introduction**

*The* Gross Domestic Product (GDP) of Montgomery county is one of the most important macro-economic indicators of the economic health and prospects of the county. Over the past 10 years (2009-2019), the GDP has seen an upward trend, propelled by a complex web of macro and micro economic factors. This GDP growth trend will be examined in this research study with the aid of numeric datasets of these explanatory variables. Effort will be made to determine if there exists a relationship between GDP growth (considered to be the dependent variable) and county population, median household income and the unemployment rate (assumed to be the independent variable). If this conjectured relationship is confirmed, then future GDP values for the county can be reasonably predicted by varying the predictor variables.

**Background**

*Montgomery County* is a local government area located in the state of Maryland between northern Virginia and Washington D.C. It’s one of the oldest counties in the nation established in 1776 at the State convention. According to [www.homesnacks.com](http://www.homesnacks.com), the county is home to about 1,043,530 residents making it the largest and most diverse county in the state. It’s median home price of $484,900 and median household income of $107,758 as at December 2020, puts the county among the top 20 wealthiest counties in the nation according to foxbusiness.com. The county government is headquartered in Rockville-its main city.

**Overview of Montgomery County Government**

*At the head* of the county government is the county executive elected for a 4-year term of office with the mandate to implement and enforce the county laws and run the county. Marc Elrich is the current county executive.

The legislature, made up of 9 council members, also get a 4-year mandate to draft legislation about land use and the budget. There’s also the judicial branch that comprises the court system, with the responsibility to settle disputes bordering on criminal and civil law.

*All 3 branches* of the government together constitute over 40 different departments ranging from Public safety, Health and Human Services, Public Schools, Public Libraries, Transportation, County Attorney’s office, Fire and Rescue etc, employing over 37000 employees providing services to over 1million county residents. With a total budget of about $20billion FY2021(operating-$5.8B, capital-$14B), the county government plays a central role in the local economy of the county as both a key investor, consumer and employer. Montgomery county is a strategic business hub that makes it an essential part of the economy of Maryland.

**Overview of the Study and Data Source**

County *Economic data* over a 10-year period from 2009-2019 will be used in this project. These vital economic measurements like the county population, county unemployment rate, median household income, house-price-index, etc have important implications for the county’s GDP growth and policymaking. The datasets of these important economic parameters used for this study are of the numeric and date data types in csv spreadsheet format. They were all sourced from the Federal Reserve Bank public repository-a reputable source, thus stamping the credibility of the datasets and the research results obtained therefrom. These numeric datasets are ideal to perform statistical and quantitative analysis which are needed to validate the research hypothesis statements put forward in this research, owing to the deductive nature of the research.

**Objectives**

*The goal* of this study is to attempt to explain GDP growth trends in Montgomery county as a function of county population, resident income, unemployment etc using data analytics techniques. Multiple Linear regression analysis will be used to verify if any relationship exists between the assumed independent economic variables (unemployment rate, county population etc) on the one hand, and the dependent variable (GDP) on the other hand. If correlations do exist, the strength of such relationships will be measured. The model developed will then be used to predict future GDP values for the county.

**Stating the Null and Alternative Hypothesis**

*In order to determine* possible relationships between the assumed independent economic variables and the supposed GDP dependent variable in Montgomery County over the last 10 years, it would be necessary to consider the below probing questions;

Q1-*Did the county population growth affect GDP trends over the last decade, and to what extent?*

Q2-*What was the impact of unemployment on GDP growth between 2009 and 2019 in Montgomery county*?

*Q3-Were high median household incomes for county residents associated with high GDP values over the last decade?*

The above mentioned simple, clear, unambiguous questions can only be explored further using hypothesis statements framed after these questions as follows;

HS1-

*Null Hypothesis (Ho*): The steady growth in the county’s population over the last decade had no significant effect on GDP growth over the same period.

*Alternative Hypothesis (Ha*): The steady growth in the county’s population had a positive effect on GDP growth in Montgomery county over the last decade.

HS2-

*Null Hypothesis (Ho):* The unemployment rate had no significant effect on GDP growth in Montgomery County over the last 10 years.

*Alternative Hypothesis (Ha):* The unemployment rate had a negative effect on the GDP growth rate in Montgomery County over the last 10 years.

HS3-

*Null Hypothesis (Ho):* Higher median household incomes for county residents had a negligible effect on the county’s GDP growth over the last decade.

*Alternative Hypothesis (Ha*): Higher median household incomes had a significant impact on GDP growth in Montgomery county over the last decade.

**Literature Review**

*Before making any assumptions or* analysis about possible correlations between Montgomery county’s population, unemployment rate, median household income etc and GDP growth trends over the 10-year period (2009-2019), it would be enlightening to review what some economic scholars and policymakers think about GDP and its drivers.

*Bettina (2015*), argued that the US government’s Quantitative Easing (QE) policy of buying bonds and keeping the interest low after the 2008 financial crisis boosted GDP growth which in tend brought down the unemployment to 6.1% in 2012 from a high of 10.6% at the peak of the crisis in 2009, even though jobless growth persisted. This affirms an inverse relationship between the rate of unemployment and economic growth (GDP) as affirmed by Okun (1962) p89-104, who alluded to this negative correlation on the grounds that employed people contribute to the national cake while unemployed people do not.

*However, Nguyen* *(2017*) , using panel data from 1997-2016 with a focus on South East Asia, showed that the unemployment rate had a positive effect on GDP growth for those developing countries whose data he reviewed, contrary to the popular view of an inverse relationship between unemployment rate and GDP growth. He added that export value and value-added manufacturing had the most significant effect on GDP growth for South East Asian Countries.

*According to Dragos* (2019), government fiscal policy played a role in GDP growth and in attracting Foreign Direct Investment (FDI) in Central and Eastern European Countries using data from 2005-2015.

*Meanwhile Sandra* (2012), opined that Latvia a country in Eastern Europe saw a surge in GDP growth per annum of 12% after joining the EU in 2004, though periods of recession followed. This GDP growth in part was due to the large viable population of the EU market that provided strong demand for Latvian exports.

**Research Methods**

*The* *Deductive* Research approach would be used in conducting this research project, in light of the research objectives, time constraints and the numeric nature of the datasets. The Deductive research methodology adopted, makes it suitable to run statistical algorithms to test the hypothesis statements listed above. Following the deductive methodology, research findings will flow from generalization to specific.

**Research Methodology**

*The* *quantitative research* methodology would be followed, considering that the deductive research approach has been adopted. In this regard, descriptive statistics will be used to perform exploratory data analysis on the numeric explanatory variables prior to making inferences or conclusions from the results of the hypothesis. There will be no qualitative input (surveys, questionnaires, opinions) from Montgomery county officials. The statistical evidence supporting the research findings will come from the data analysis phase of this study.

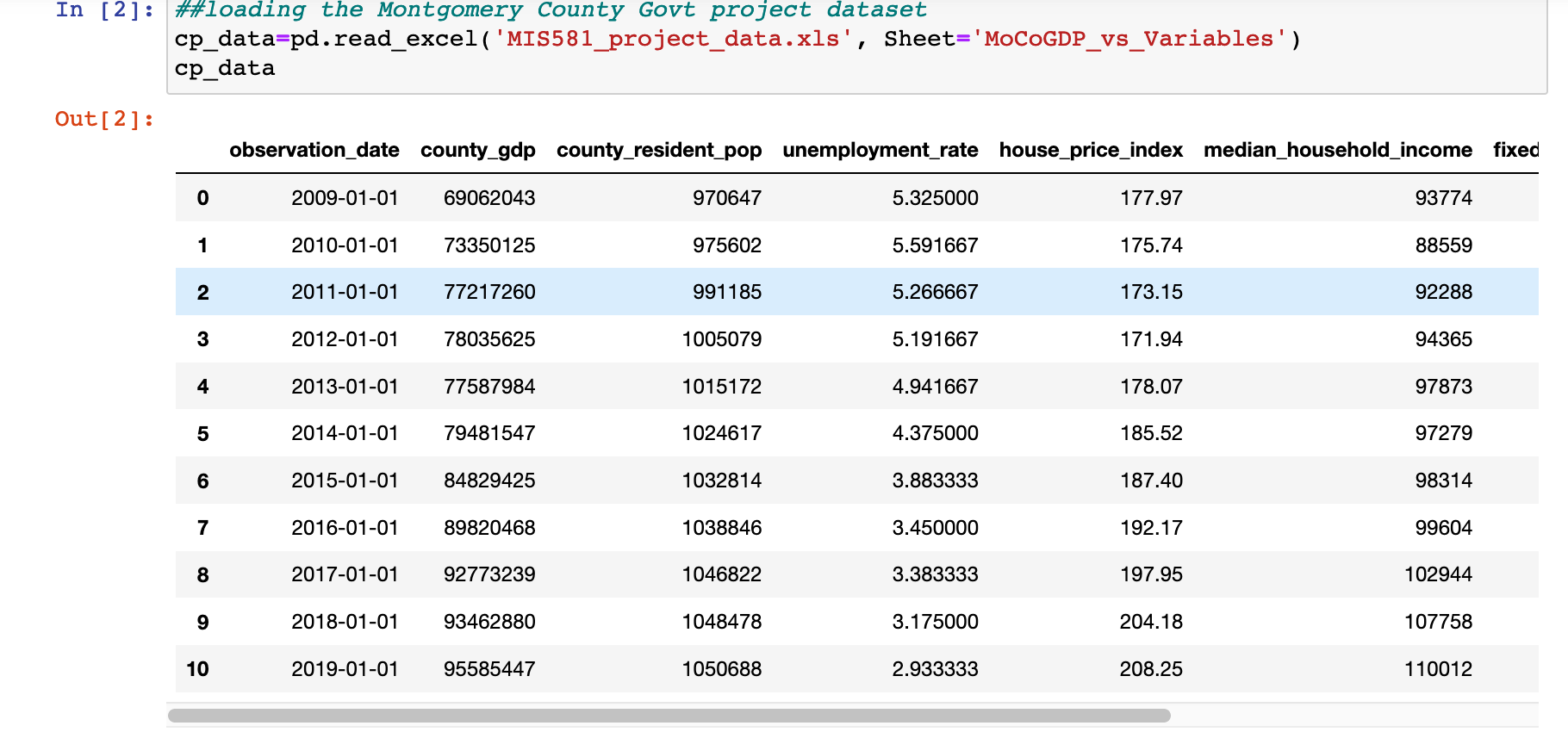
**Data Analysis and Presentation**

*In order to* better understand the data that would be used in validating the hypothesis stated above, it would be helpful to explore the data using descriptive summary statistics. Next, using python libraries, the multiple linear regression model will be built to explore the correlations between the GDP outcome variable and the other predictor variables. These relationships will be visualized using scatter plots. Because the dataset is small, 90% of the data will be used to train the model, while the remainder will be used to test or validate the model. Correlation coefficients and the p-value will be used to test the hypothesis at 5% level of significance, in order to make predictions about future GDP values. Finally, the accuracy of the model will be evaluated to judge its reliability.

**Exploring the nature and correlations in the datasets.**

Figure 1 below highlights the Montgomery county GDP capstone dataset after being successfully loaded in python jupyter notebook. The dataset is a small dataset named MIS\_581\_project\_data.xls with 11 observations(rows)and 8 variables(columns). The 4 main dependent variables that will be used to predict the response variable-GDP are county\_pop, unemployment\_rate,median\_household\_income and mean\_real\_wages.

*Fig1- MIS\_581\_project\_data.xls dataset uploaded on jupyter notebook*



The dataset above shows that the county’s GDP has been steadily increasing over the last decade, just like the county population, house\_price\_index and median\_household\_income. Within the same period, the unemployment rate has been on the decline reaching its lowest rate of 2.93% in 2019, while the fixed\_mortgage\_rate has been fluctuating in the lows. Hopefully, the test results of the hypothesis will provide some insights into these trends. The below line charts highlight the trends and evolution of the above economic parameters over time for Montgomery county.

Fig2 -*chart below showing unemployment gradually falling over time*

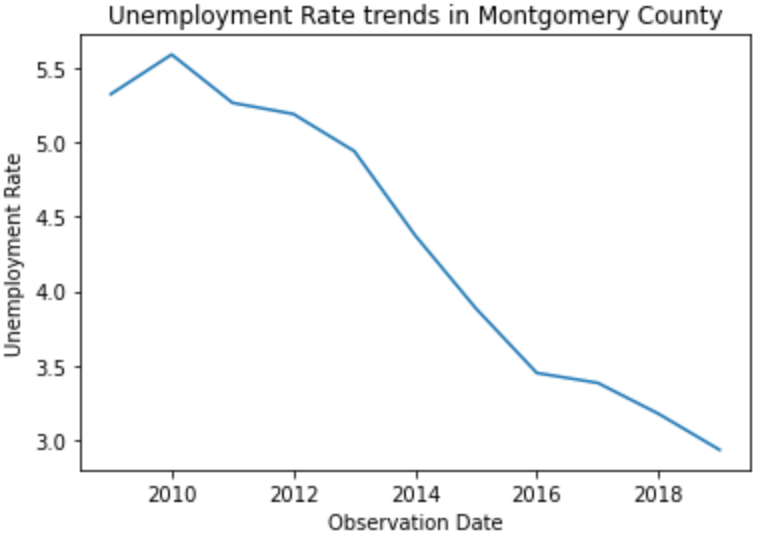


Fig3- *Chart below shows the median\_household income has been steadily rising*

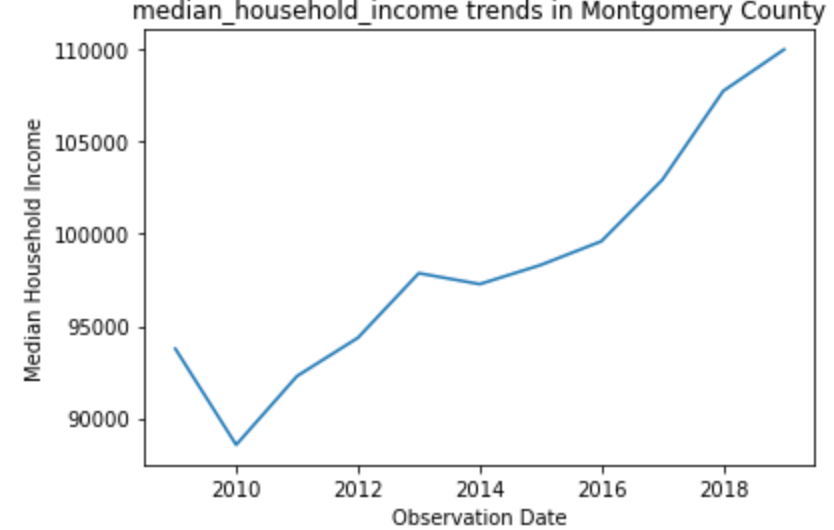


Fig4- *Chart shows the county population has been rising over the last decade*

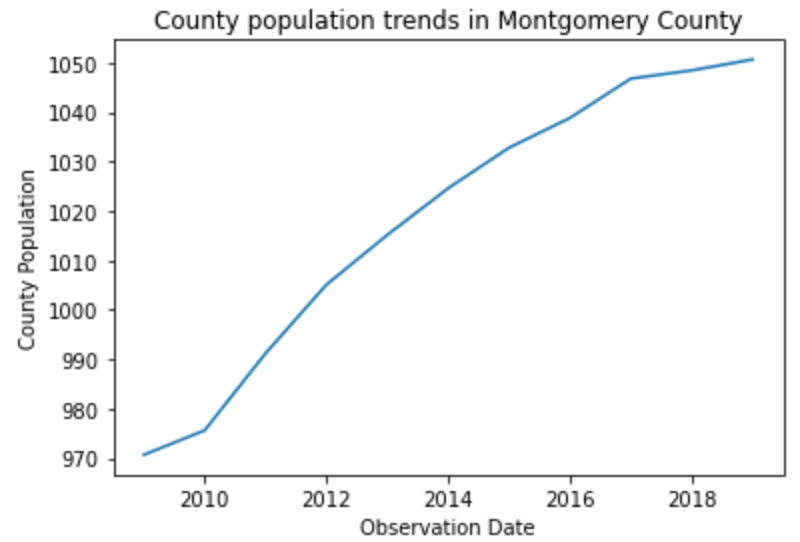
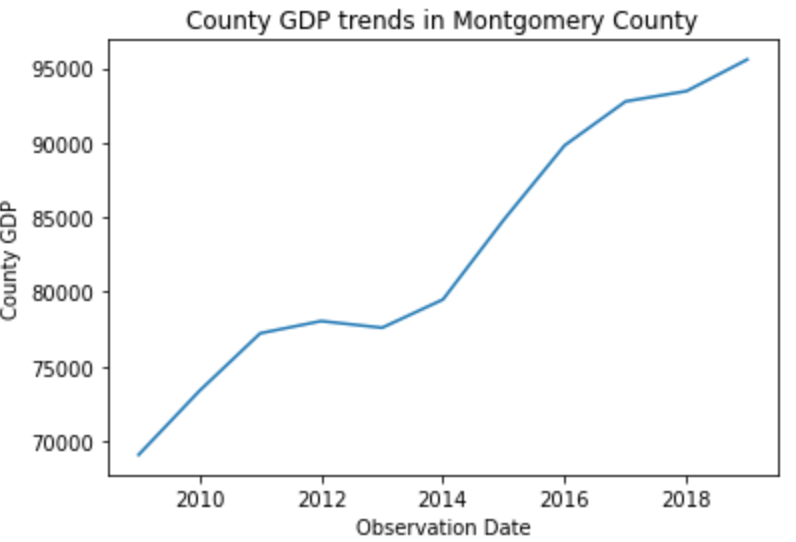
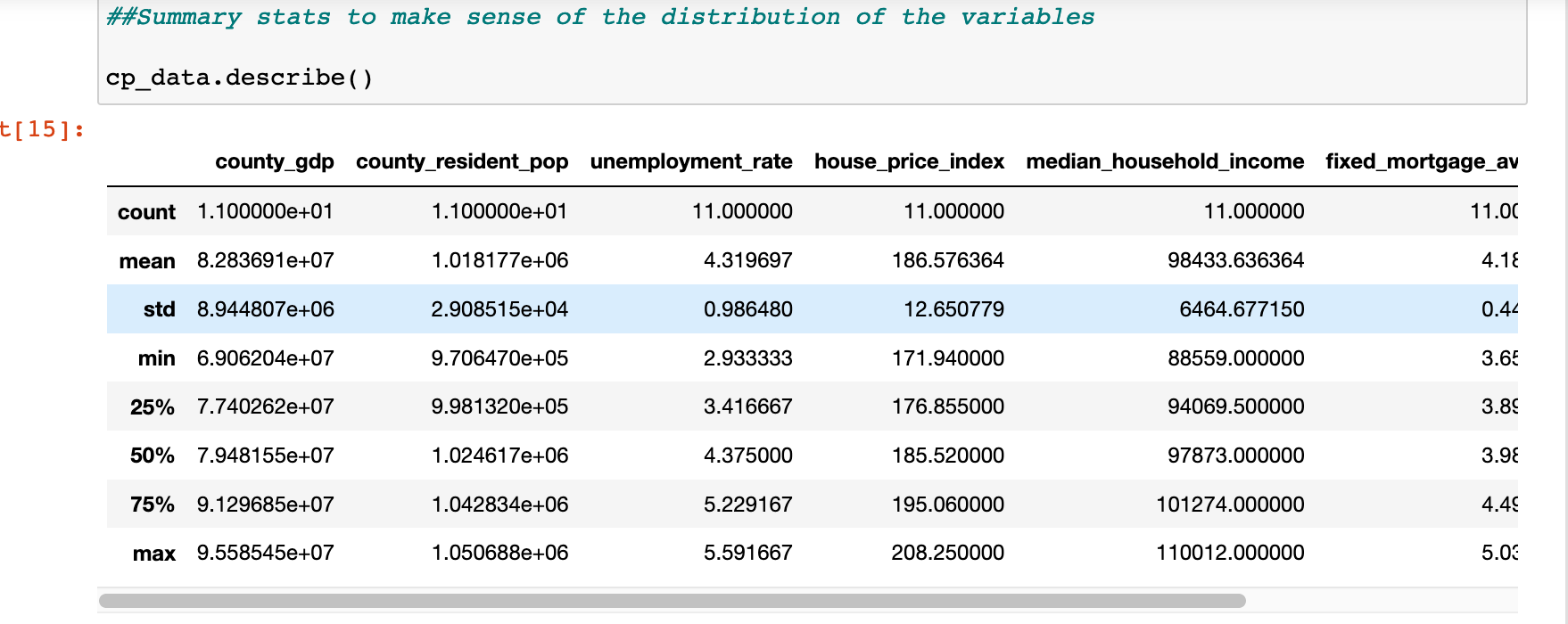


Fig5- *Chart shows the county GDP has been steadily increasing over the last decade*



Next, are summary statistics that describe the variables of the datasets. These stats highlight the nature of the distribution, spread, skewness, patterns and trends in these economic variables over the last decade as seen below;

Fig6- *Summary Descriptive stats of the Montgomery county dataset*



The summary stats above show that the county recorded its highest GDP of over USD95 billion and median income of $110k in 2019 with an average GDP and median income of USD82 billion and $98k respectively. Meanwhile, the highest rate of unemployment of 5.59% was seen in 2010 with an average unemployment rate of 4.3% over the 10-year period. The county population has also been on the rise, reaching the million mark in 2012. The housing market had its best year in 2019 as seen in the highest house-price-index and mortgage rates.

With the trends and patterns in the data understood and visualized, the multiple linear regression model can now be built on the dataset.

**Building and Running the Multiple Linear Regression Model**

*To build a robust* multiple linear regression model on the MIS\_581\_project.xls dataset shown above, 90% of the data was used for training the model, while the remaining 10% was used to test and validate the model. This statistical algorithm was built using python machine learning libraries as shown in Fig7 below;

Fig7-

*the multiple linear regression being trained and fitted on the dataset*

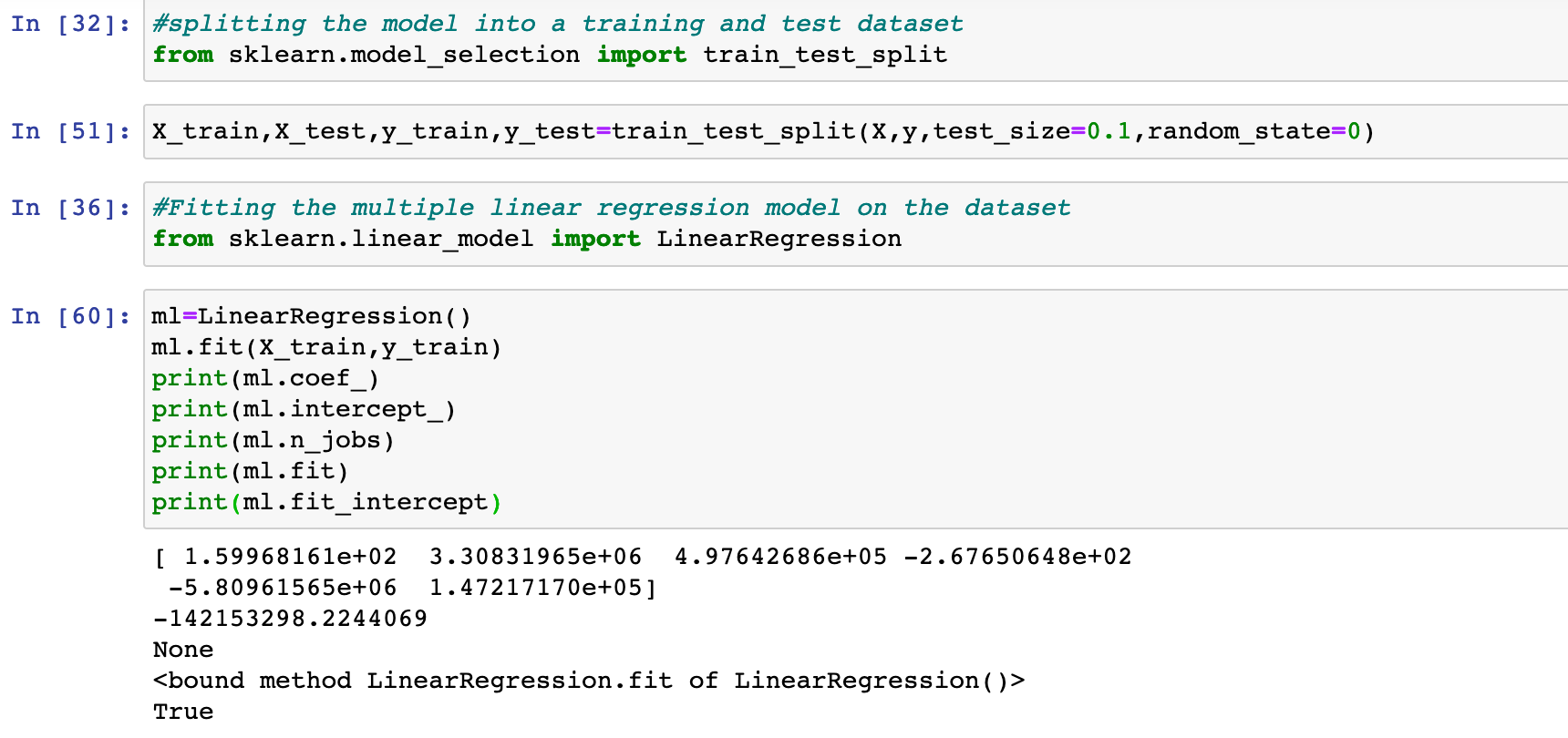


Fig7 above shows the successful results of the multiple linear regression model using GDP as the outcome variable(Y) and the independent economic variables of interest-county\_pop, unemployment\_rate,median\_household\_income,mean\_real\_wages(X). The regression output also shows the coefficients and intercepts that will be useful in predicting future GDP. Based on the regression coefficients, it would appear that, the unemployment rate is the strongest negative predictor of future GDP growth, while the county\_pop, median\_household\_income are the strongest positive predictors of county GDP.

**Calculating the Correlation Coefficients**

*The coefficients* produced in the results above together with the correlation coefficients help determine the relationships and strength thereof between the outcome variable GDP and the predictor variables. Values range between -1 and 1 with negative values indicating inverse relationship while positive values indicate direct relationship. The correlation coefficients show that unemployment rate has a strong negative correlation with county\_gdp, with a correlation coefficient of -0.96. County\_resident population and median\_household income with correlation coefficients of 0.94 and 0.896 respectively, appear to have a strong positive correlation with county\_gdp meaning the move in the same direction. Mean real wages, or the effect of inflation appear to have a moderate positive relationship with GDP as seen by its coefficient of 0.77. The below scatter plots and heat map illustrate the relationship between the GDP and the predictor variables as revealed by the correlation coefficients.

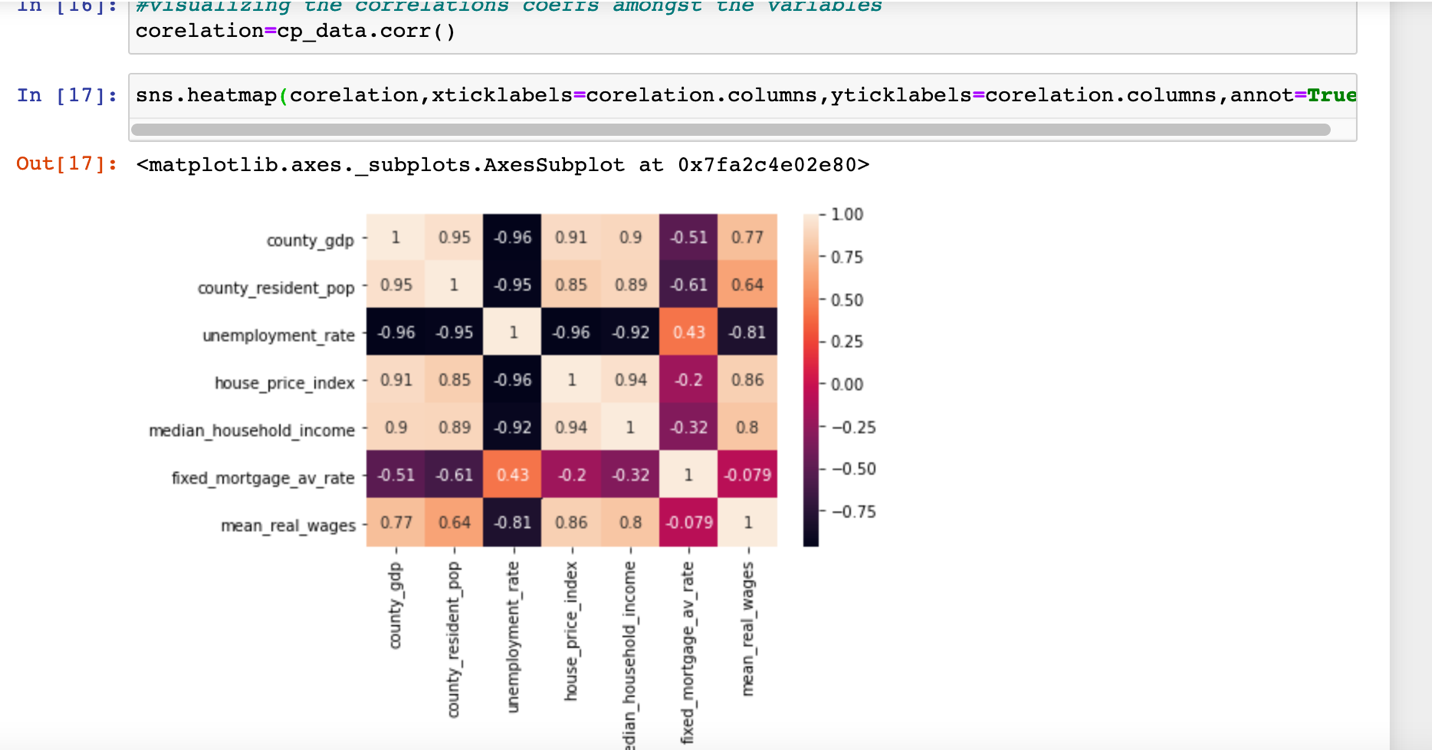
Fig9 *Heatmap showing the correlation coefficients between variables in the linear model*

Fig10- *scatte*r *plot showing strong direct relationship btw county\_pop and GDP*

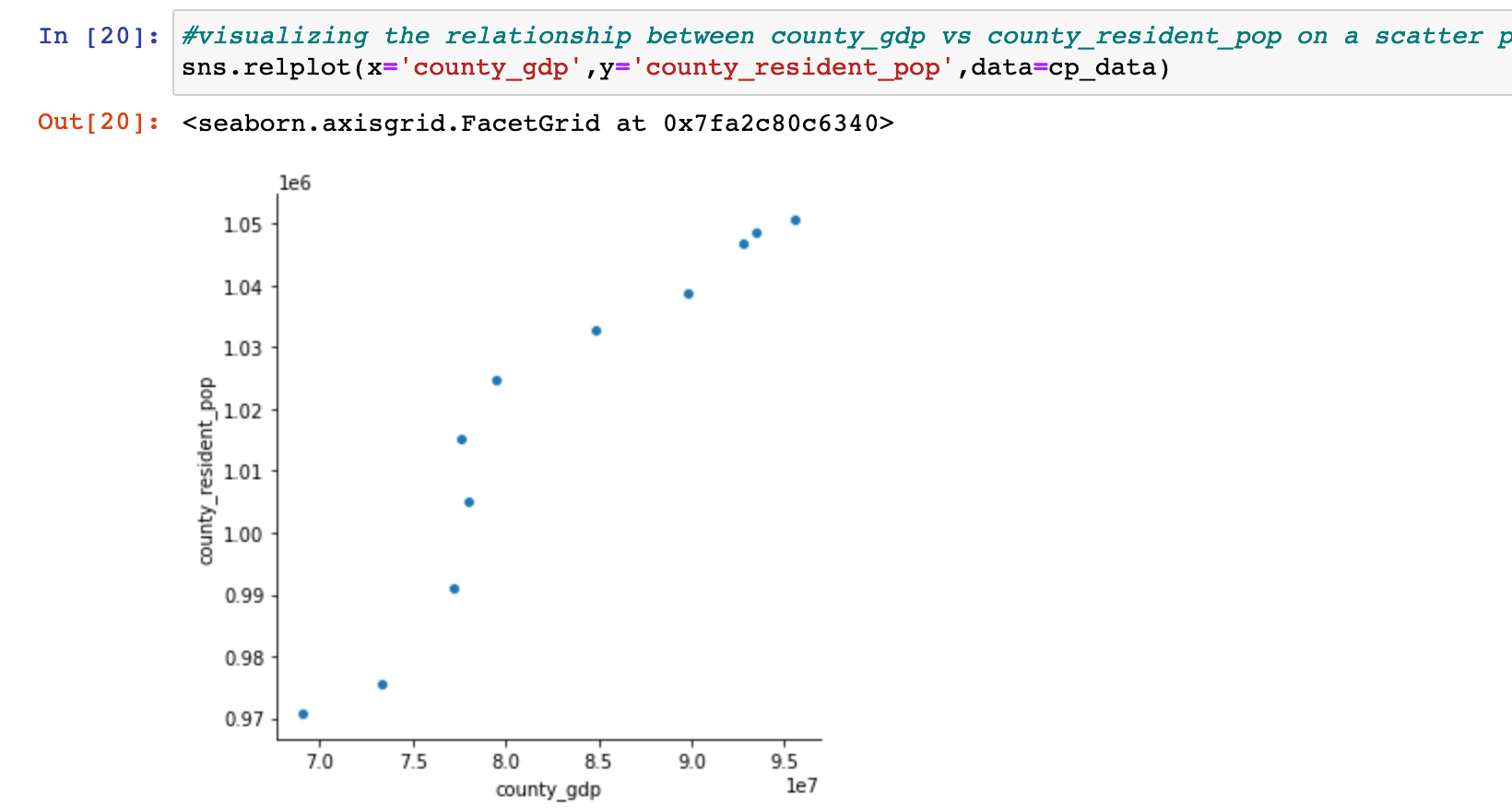


Fig11- *scatter plot showing strong inverse relationship between unemployment and GDP*

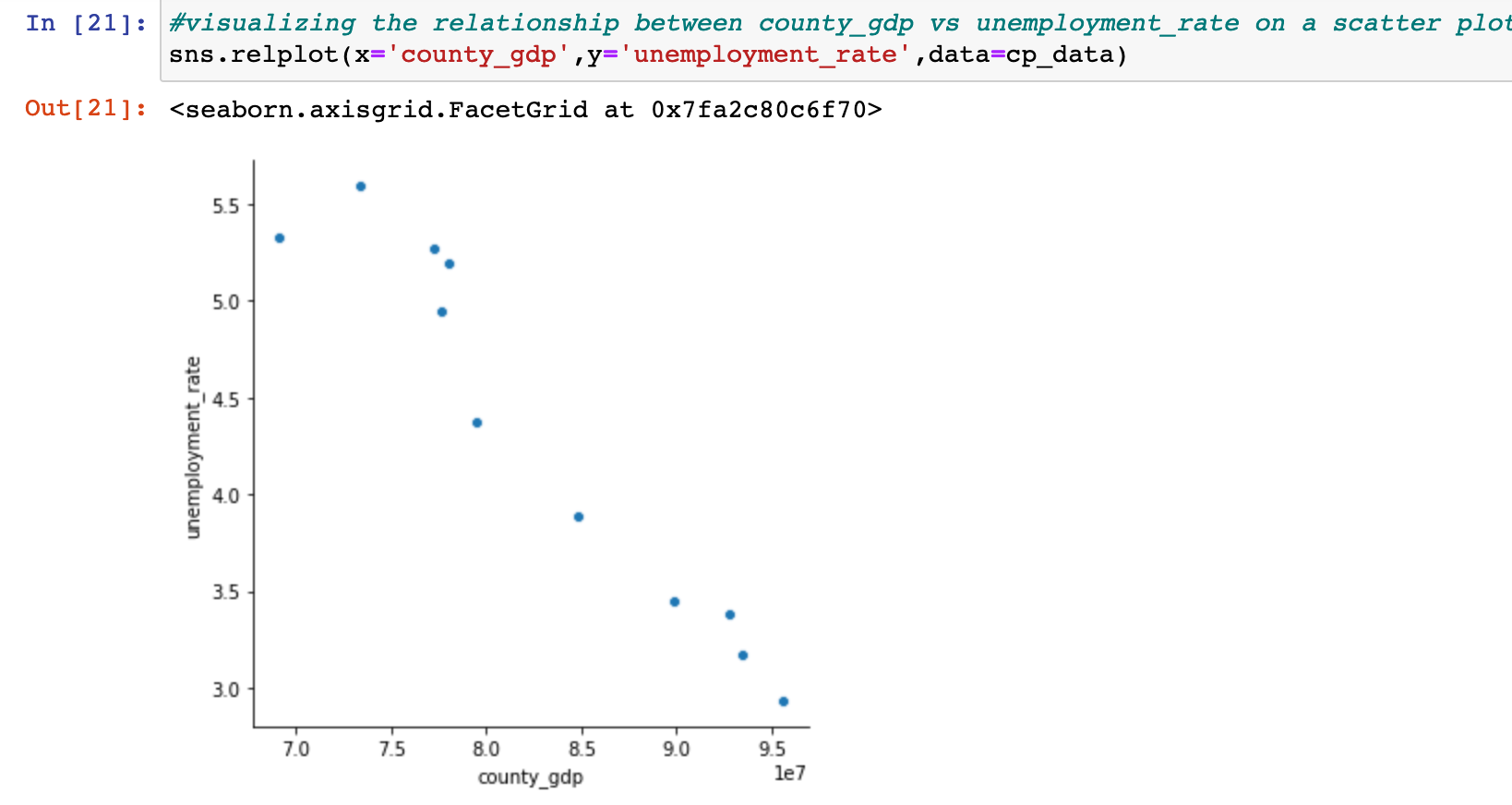
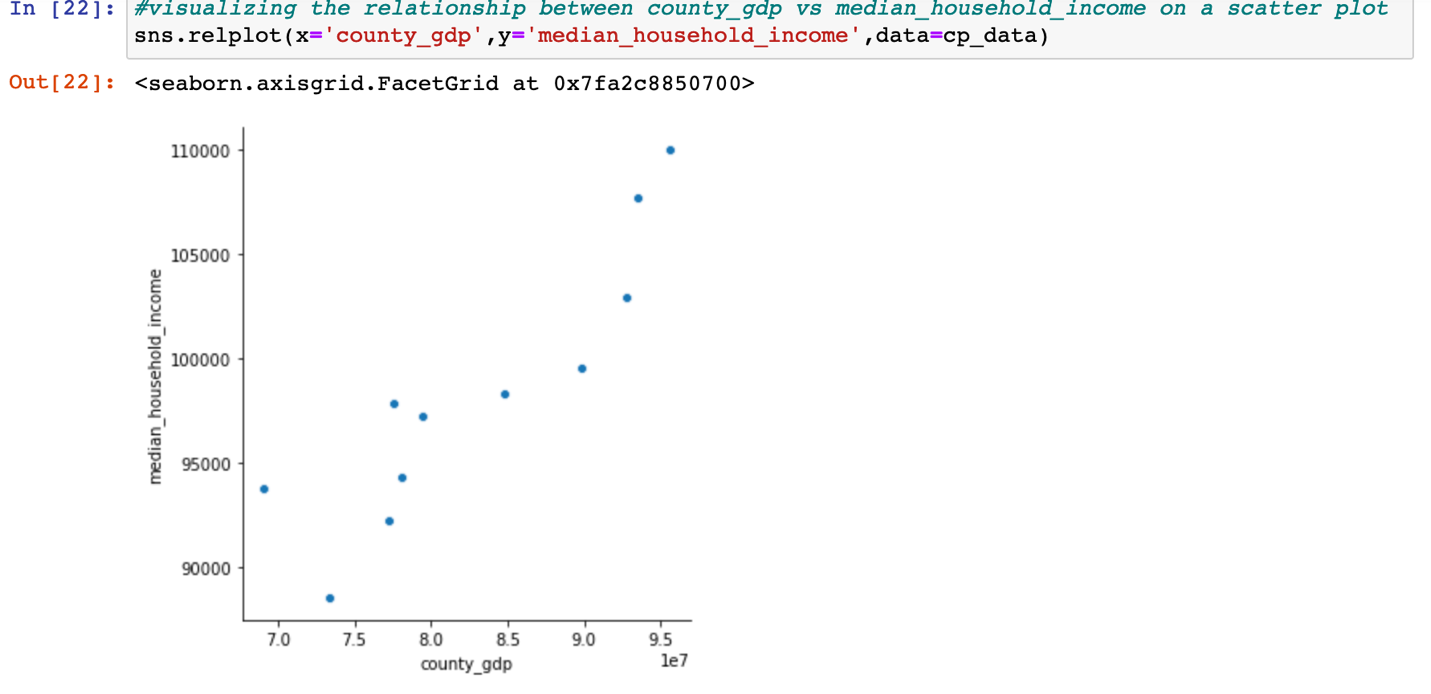


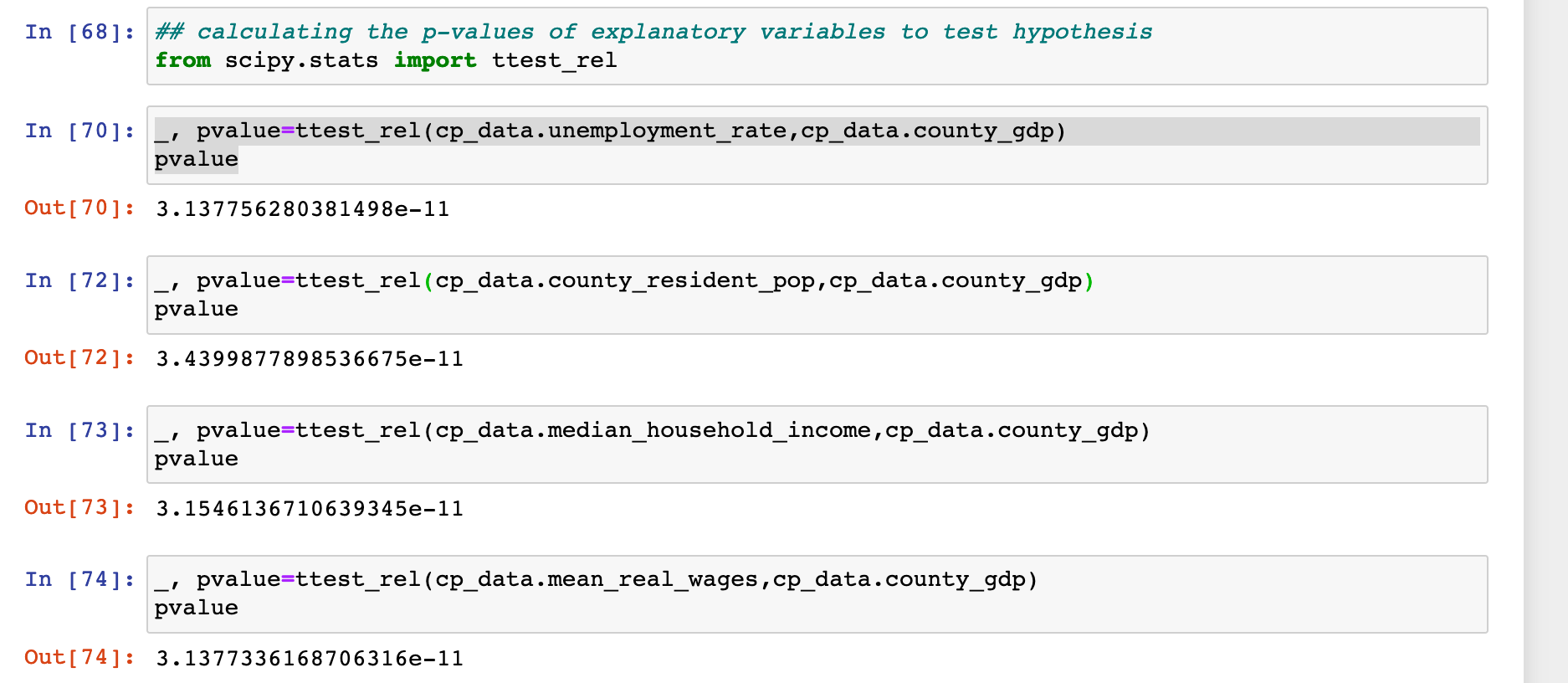
Fig 12-*scatter plot showing strong positive effect between median\_income and GDP*



**Validating the Research Hypothesis**

*Having confirmed* the relationships between GDP and the other economic explanatory variables and the strength of such relationships, the next phase is to either accept or reject the null hypothesis based on the statistical evidence. To that end, the p-values will be compared with Alpha at 5% level of significance or 95% confidence interval. If the p-value is less than, Alpha (5%), then sufficient statistical evidence exists to reject the null hypothesis in favor of the alternative hypothesis and vice versa. Figure 12 below shows the p-values for the explanatory variables under review;

Fig12- *the results of the p-values of the predictor variables*



-From figure 12 above, the p-value 3.439987789853667e-11 of the county\_pop variable is far lower than 0.05(the level of significance or alpha).

Decision1-

*Reject the Null Hypothesis in favor of the Alternative Hypothesis.* This simply means that sufficient statistical evidence exists to suggest that the steady growth in the county’s population over the last decade had a significant positive effect on GDP growth. In effect, there is sufficient evidence to confirm a strong direct relationship between population growth and GDP growth in Montgomery county over the last decade. This could be because the more people the county has, the more purchasing power the county has. This means more people are likely to be in the work force who will earn income, pay taxes, set up businesses and spend money. Increase in consumer demand will push businesses to take loans and expand further. This surge in business investment will raise interest rates and GDP. All these activities boost the GDP of the county.

-Moving to the unemployment variable, the p-value of 3.1377566280381498e-11 is far lower than 0.05 the alpha value or level of significance.

Decision2-

*Reject the Null Hypothesis in favor of the Alternative Hypothesis*. This means that sufficient statistical evidence exists to suggest that the unemployment rate in the county adversely affected GDP growth over the last 10 years. In other words, there is statistical evidence to support a strong inverse relationship between unemployment rate and GDP growth rate over the past t10 years in Montgomery county. This could be so, because the more people are unemployed in the county, the purchasing power drops, the county receives less in taxes and instead pays more in unemployment benefits which reduces the GDP of the county. Because people are not earning income while unemployed, they spend less, economic activity slows down, and businesses scale back operations and lay off more workers as result leading to an endless cycle of economic decline. The COVID-19 pandemic slow down and the financial crisis of 2007-2010 are noteworthy examples.

-Finally, the p-value of the median\_household\_income of 3.154613671063934e-11 is far lower than the 5% level of significance.

Decision3-

*Reject the Null hypothesis in favor of the alternative hypothesis*. This means that there is sufficient statistical evidence to suggest strong positive relationship between median household income and GDP in the county over the last decade. The reason could be because, when people earn more money, they spend more and the government collects more in taxes. Businesses also flourish because of high purchasing power which will likely incentivize business expansion which again leads to an increase in investment which increases the GDP of the county.

It’s worth mentioning that, the economy of the county is complex and inter-dependent not just on internal county dynamics alone. Stock market activity, Federal Reserve monetary policy decisions, federal government fiscal policy, external shocks from international stock markets, natural disasters and geopolitical events can significantly affect the GDP of Montgomery county without directly affecting the explanatory variables mentioned in this study.

**Model prediction and Evaluation**

The objective of building this data analytics predictive model is to;

1. *Not only to explain and evaluate GDP changes over the years using the other independent metrics, but also to;*
2. *Help forecast the outcome variable (GDP) in future years with these explanatory economic variables derived from different datasets that were not used in building or* training the model. Since 90% of the dataset was used to train the model, the remaining 10% of the dataset will now be used to validate the model by predicting future GDP values. In this way, the accuracy or reliability of the prediction model can be measured. *R2 is a measure of the model accuracy*. The higher the value of r2, the more accurate and reliable the model is for prediction. The below screenshots show the r2 value for the Montgomery County GDP model and predicted GDP values using the test dataset;

Fig13- *r2 value of model accuracy and predicted values of GDP*

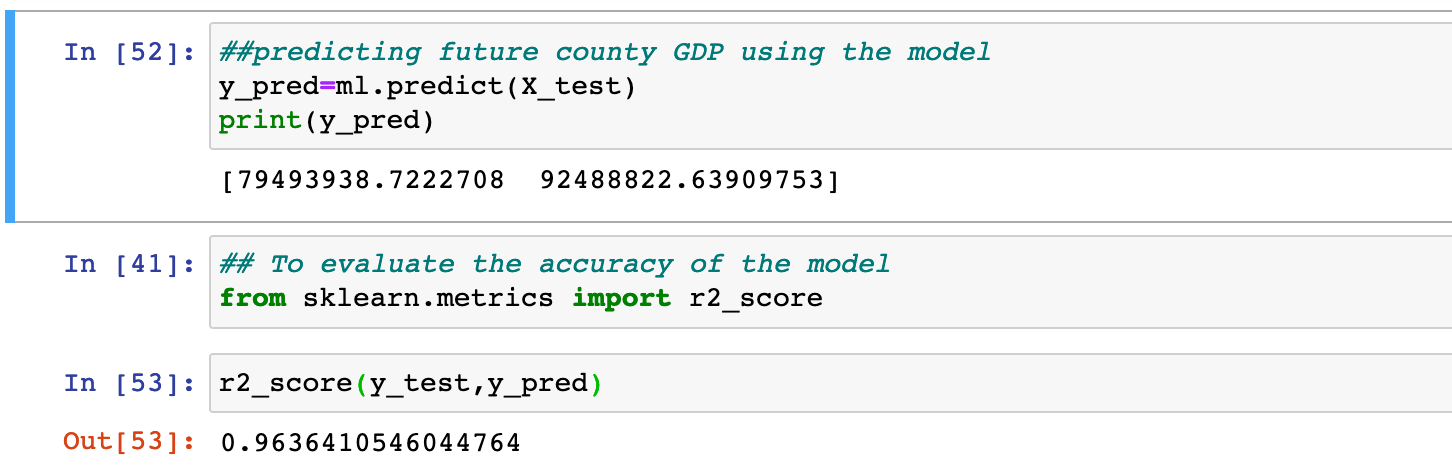


Figure 13 above shows the linear model built to predict GDP of Montgomery county using the county\_pop, unemployment rate and median household income is about 96% accurate. This is a pretty good model. The model also predicts the county GDP for the next 2 years after 2019 to be about $79.5B and 92.B for 2020 and 2021 respectively. It’s worth mentioning that, a more dependable model will factor in more explanatory variables and will use a far larger dataset for model training since the bigger the dataset, the more capable and accurate the machine learning prediction model.

**Limitations of the Study**

*This capstone* research is in no way exhaustive. It uses a small public data spanning only a 10-year period. Machine learning prediction models are more accurate and reliable with a larger training dataset. Time and circumstances change the economic landscape which could affect study outcomes. Because economics deals with human behavior which is complex and erratic, figures on a spreadsheet alone cannot fully explain economic events like GDP. Moreover, only a select economic variables were used as explanatory variables. Furthermore, other macro-economic factors like taxation, level of inflation, foreign exchange rate, Balance of Payments at the state or federal level, the level of technology, government policy, external market shocks, geopolitical events and other qualitative factors that cannot be quantified or modeled, can and do influence the GDP and economic growth of Montgomery county.

**Ethical Considerations**

*Data privacy*, security and ethics evoke strong feelings in this age of digitization and mass surveillance by governments. The datasets used for this study are aggregate datasets obtained from a public source. Montgomery county as a government institution has privileged access to millions of records of county residents. This data could be anything from biometric, demographic, financial, personal etc that is voluntarily given during surveys by residents or surrendered to fulfill obligations to comply with tax laws, census, government audits of businesses etc. These individual datasets then cumulate to form the aggregate datasets used in this study.

The county government has the duty to exercise due care and professionalism to make sure data in its possession is only used for the lawful purpose of conducting government business. In this regard therefore, Ferguson et al (2019) argues that unauthorized sharing, selling or use of the data for any other purpose will be an unethical breach of trust that violates the privacy of its residents.

**Findings/Recommendation**

*Based on the* results of this study, the GDP growth in Montgomery county over the last decade has been shown to be positively driven by the steady increase in population and median household income. Conversely, high unemployment rates negatively affected GDP growth in the county within the same period. Therefore, policies that encourage immigration to the county like tax cuts, deregulation (that makes it easy to set up businesses) will be good for the economic health and future prospects of the county. This will then create new businesses that will hire people and hence keep the unemployment rate down. The reduced unemployment rate will then boost GDP growth as has been demonstrated. However though, rising costs of housing in the county is pushing people out to cheaper counties. Raising the minimum wage also boosts the median income and spending power of county residents which tend to boost the GDP.

**Summary/Conclusion**

*The impact* of economic variables like county population, unemployment rate, median household income etc have been used to analyze GDP growth trends in Montgomery county over the last decade. Python statistical algorithms were used to build a multiple linear regression model on the dataset that predicted GDP growth in upcoming years. The deductive research method was adopted to validate hypothetical correlations between the GDP and key economic variables with the training dataset. Ethical considerations in sourcing and storing the datasets were considered. Future researchers could expand on the datasets and factor in qualitative data for more insights. It is hoped that the study findings and recommendations could be useful to Montgomery County officials in crafting future plans and policies.

**References**

### 

Bettina (2015). Should we increase Economic Growth or Boost Employment*? -* *Public Finance Quarterly (0031-496X). 2015, Vol. 60 Issue 2, p249-269. 21p.*

Dragos (2019). The Competition and factors influencing the Gross Domestic Product and *Foreign Direct Investments of CEE Countries-Economic Research-Ekonomska Istrazivanja.*

*Dec2019, Vol. 32 Issue 1, p876-893. 18p.*

### Ferguson (2019). Journal of Learning Analytics, v6 n3 p25-30 2019*. (EJ1237573)-*[*Ethical Challenges for Learning* ***Analytics***](https://eds-a-ebscohost-com.csuglobal.idm.oclc.org/eds/viewarticle/render?data=dGJyMPPp44rp2%2fdV0%2bnjisfk5Ie45PFJsKuxTbSk63nn5Kx94um%2bSq2prUquqK84sLCyS7iqrzjOw6SM8Nfsi9%2fZ8oHt5Od8u6%2bwTrKnrk21qqSL59q7TuKtt0%2fk2%2bNFtdzheauqtn3jo7Z94terUOHb4UivprF5r6i3WPHb8Yvn5eyF5eiySK6spITf5OVV4%2bSkfOCz44rn2aSM3927Wcyc34a7u8hJsKm1TbWppH7t6Ot58rPkjeri8n326gAA&vid=0&sid=6d797fee-7fca-48ee-8eda-8cec0103a129@sessionmgr4006)

Nguyen (2017). Impacts of Macroeconomic Indicators on Economic Growth in Southeast Asia: *A* Panel Data Analysis*- Journal of Economics and Business Research, Vol XXIII, Iss 2, Pp 111-128 (2017)*

Okun (1962). ‘‘Potential GNP: Its Measurement and Significance*.”*

*In: Proceedings of the Business and Economic Statistics Section of the American*

*Statistical Association. Alexandria, VA: American Statistical Association, 1962,*

*pp. 89-104.*

*Sandra (2012).* Strategic Choices for the Economic Development of Latvia as an EU member*-*state-*European Integration Studies. 2012, Issue 6, p134-139. 6p.*

**Retrieved From**

[www.foxbusiness.com](http://www.foxbusiness.com)

[www.homesnacks.com](http://www.homesnacks.com)