DATA 512 – Human-Centred Data Science

Course Project – Part 4 – Final Report

Air Quality and the Economy – Cheyenne, WY

Arjun Sharma

Introduction:

Wildfires are one of the latest accompaniments of the disastrous consequences of climate change. One of the direct consequences of wildfires is an increase in the smoke particle density in the air, which worsens the Air Quality Index. The EPA is one of several entities that measure the Air Quality Index, and the availability of the AQI would further inform society at all levels about their level of air pollution. A worse AQI means less healthy air for the population to breathe in, leading to various economic downturns, and a general decline in the multitude of health-related metrics that define how healthy a population is. As such, a higher AQI is in general a hindrance to the economical and social development of a region. The goal of this analysis is to explore how the economy is impacted by the AQI and the larger industries in Cheyenne. In addition, some of these facets will also be projected into the future so as to inform the city council and the mayor about potential strategies for the future.

Cheyenne is the capital city of the state of Wyoming, USA. It has a population of about 65,000 people. Mining, logging, construction, transportation and goods trading are some of the largest industries that Cheyenne's economy thrives and to an extent, depends upon.

Cheyenne is a low-risk city with respect to wildfires according to FEMA [1], and while that may be the case currently, the future may be different as we will explore in this report later. Climate change continues to exacerbate the situation across the world, and Cheyenne may be no different. In fact, the potential worsening of the wildfire situation and the worsening of the AQI over the long-term, can impact the mining, construction, transportation industries – industries that Cheyenne almost relies upon [2, 3, 4].

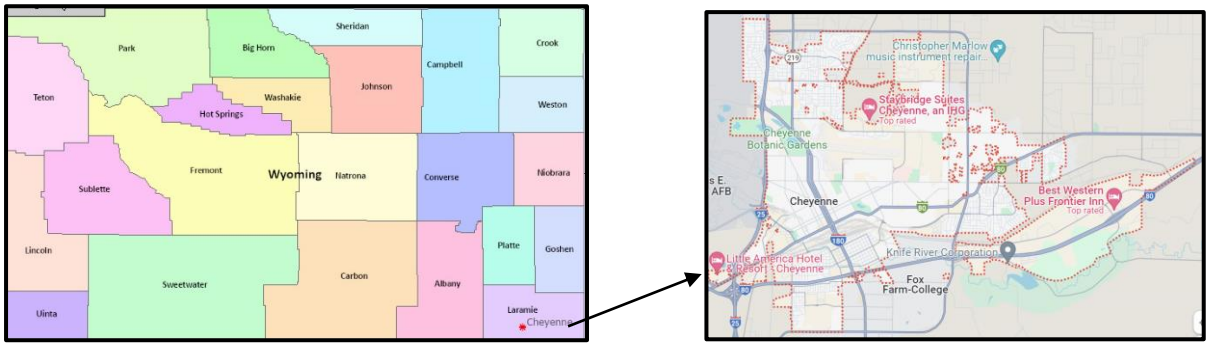

Figure 1: *Cheyenne, WY*

Quick Facts:
Population: 65,132
County: Laramie, WY
Area: 32.37 sq. mi.
Elevation: 6,086 ft (1,855 m)
Industries: Mining, Construction, Transportation, Trading, Military

Background:
With the knowledge of how wildfire tend to affect the environment and subsequently, the economy, I wanted to understand the trends of wildfire occurrence in and around Cheyenne, WY.

The wildfire data that may help answer this is sourced from the United States Geological Survey (USGS). It contains polygon data for wildfires in the US from sometime in the 1800s to 2020. For the purpose of the analysis, I restrict myself to 1963 onwards. The dataset also contains information on the fire type and the acreage burned during a fire. The data is also restricted to 1250 miles from the city of Cheyenne.

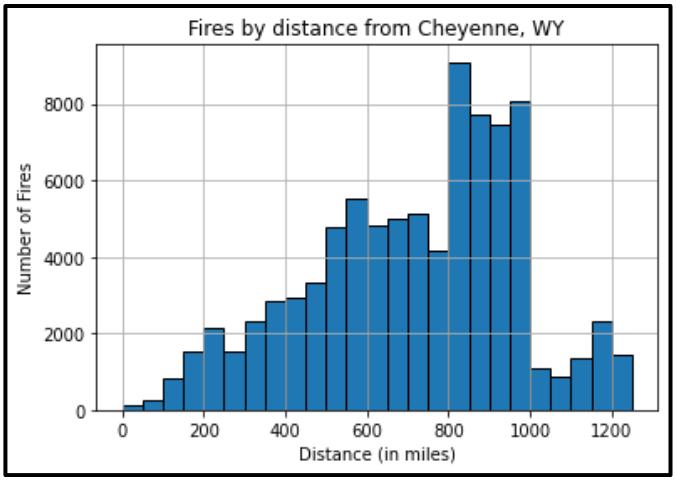

Figure 2: *The distance from Cheyenne for fires up to 1250 miles away from the city.*

Figure 2 presents the number of fires at a given distance from Cheyenne, WY. The histogram bins distance values at every 50 miles and presents the number of fires that lie in the said range. It is observed that there are very few fires (relatively) close to Cheyenne, and that number increases as the distance is increased, reaching a maximum at 800-850, but the next three largest bins following immediately after, upto 1000 miles. A sharp decline is observed for distances after that. That being said, there is a substantial number of fires even within 400 miles of Cheyenne, WY.

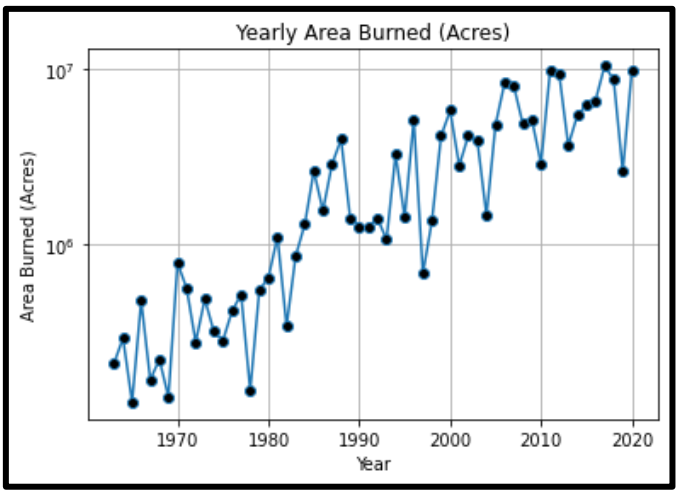

Figure 3: *Area Burned Yearly (up to 1250 miles away)*

Figure 3 presents the area burned because of fires in and around Cheyenne, WY up to a distance of 1250 miles from the city. The linear increase over time on a logarithmic scale is of utmost concern,

because the area burned in 2020 has increased by about ten-fold compared to just 30 years ago, in 1990.

Previous Results and Inferences:

In Part 1, I create my own benchmark to assess the smoke levels and to create an estimate of the smoke levels. The Smoke Estimate uses the data obtained from USGS as specified earlier. Mathematically, my smoke estimate is calculated as:

$$\text{Smoke\_Estimate} = \frac{GIS\_Hectares}{Distance * Distance} + (Fire\_Severity * Fire\_Severity)$$

The smoke estimate is a function that increases linearly with the area burned, and quadratically with the severity of the fire ("Wildfire" being most severe, and "Unknown: Likely Prescribed Fire" being least severe). The smoke estimate decreases quadratically with an increase in distance.

To assess the validity of my smoke estimates, I use air quality data from the United States Environmental Protection Agency (EPA). The EPA hosts an API called the Air Quality Service (AQS) which has air quality index (AQI) from air quality monitoring stations throughout the United States.

Comparing my Smoke Estimate with the EPA AQI, EPA AQI values for Cheyenne are only available from 1984, with the occasional absent value for a given year.

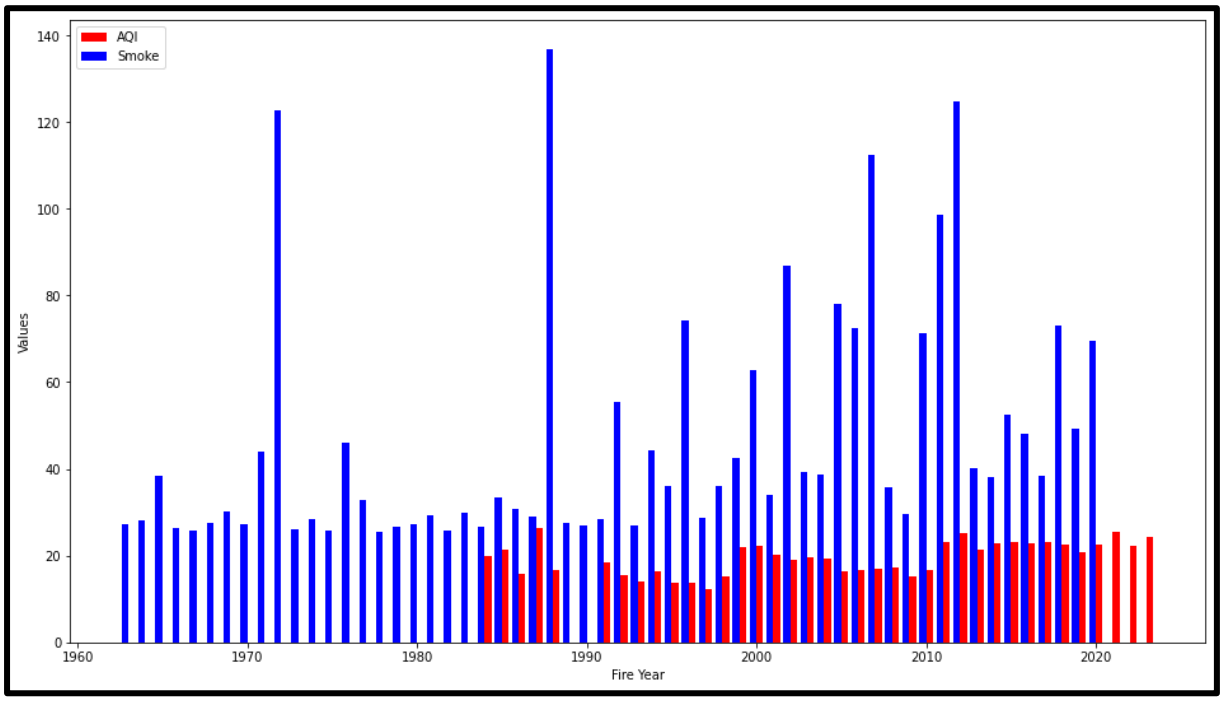

Figure 4: *EPA AQI and my smoke estimate over time*

Figure 4 presents the EPA AQI values from 1984, when compared with the smoke estimate values, I developed. The trend in their increase and decrease is quite similar, however, at some instances, my smoke estimate tends to overestimate the smoke levels in the air, and thus overestimates the smoke estimate.

Measuring levels of correlation may be of some help towards establishing a relationship between the 2 variables, and other important aspects that influence Cheyenne's economy.

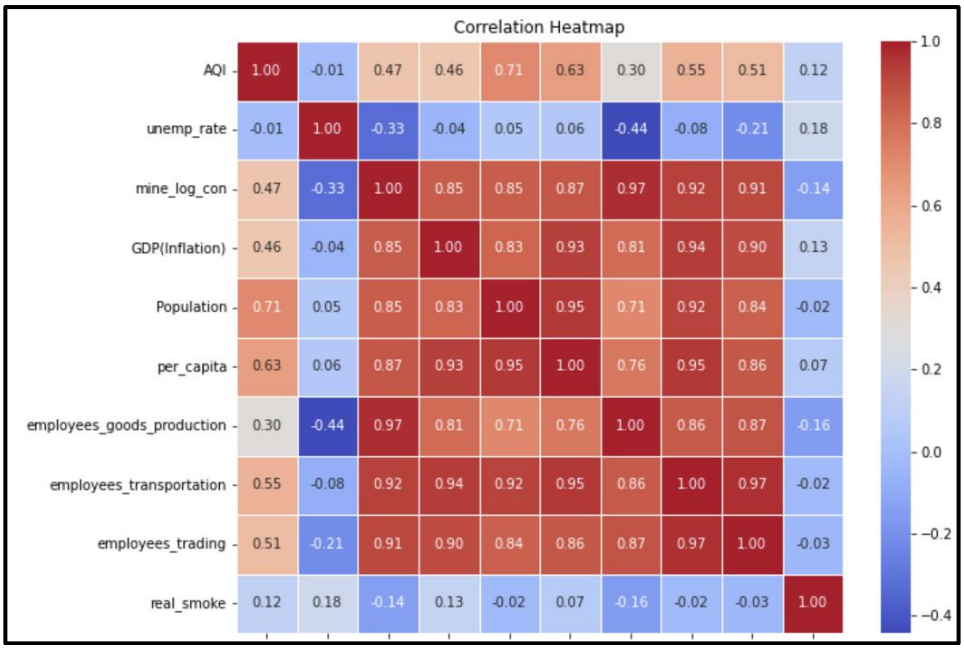

Figure 5: *Correlation heatmap between several important facets of Cheyenne's economy and environment*

The EPA AQI has strong linear relationships with:

1. Levels of employment in mining, logging, construction
2. The GDP when adjusted for inflation
3. Population
4. GDP Per Capita, adjusted for inflation
5. Employment in transportation
6. Employment in goods trading

Conversely, my smoke estimate, named 'real_smoke' has a very weak relationship with the AQI as well as the other aspects that intuitively and (now shown) practically influence the AQI. As a next step, I used ordinary least squares to predict the AQI as a function of the smoke estimate and the year. The results of which are below:

Dep. Variable:	AQI	**R-squared:**	0.433
Model:	OLS	**Adj. R-squared:**	0.370
Method:	Least Squares	**F-statistic:**	6.870
Date:	Mon, 11 Dec 2023	**Prob (F-statistic):**	0.00607
Time:	21:42:34	**Log-Likelihood:**	-46.971
No. Observations:	21	**AIC:**	99.94
Df Residuals:	18	**BIC:**	103.1
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-645.2062	180.591	-3.573	0.002	-1024.614	-265.798
Year	0.3309	0.090	3.688	0.002	0.142	0.519
real_smoke	0.0058	0.018	0.316	0.756	-0.033	0.044

Omnibus:	1.511	**Durbin-Watson:**	0.856
Prob(Omnibus):	0.470	**Jarque-Bera (JB):**	0.994
Skew:	-0.196	**Prob(JB):**	0.608
Kurtosis:	2.009	**Cond. No.**	6.80e+05

Figure 6: *AQI as a function of year and real_smoke, based on OLS*

Unfortunately, this was the best metric I was able to develop based on modification of the formula, the different terms included in calculating the smoke estimate, weightage given to each aspect, etc., For further analysis, I will use the EPA AQI as opposed to my own smoke estimate for a more informed analysis, where there is much higher assurance of me having made the correct statistical assumptions and that I have the right data at this time.

Exploratory Projections:

Based on the context established on the factors that support Cheyenne's economy and how they may be potentially affected by wildfires and worsening air quality, a few exploratory projections of some of these variables is warranted and is of interest so as to help the city council and mayor lead the city with a more informed mindset. These models are based on the SARIMAX algorithm for time-series forecasting. Any breaks in the line indicate the absence of data for the said year(s).

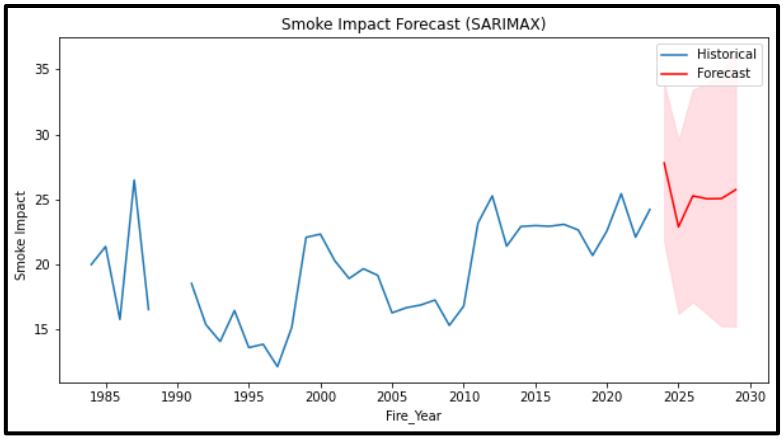

Figure 7: *EPA AQI projections for the next 5 years*

Looking at a projection for the EPA's AQI over the next 5 years in figure 7, it appears as though the current situation with the economic activities hurting Cheyenne's air quality will continue to be the case. More land burned over time adds to this. My projection implies that the Air quality will only continue to worsen.

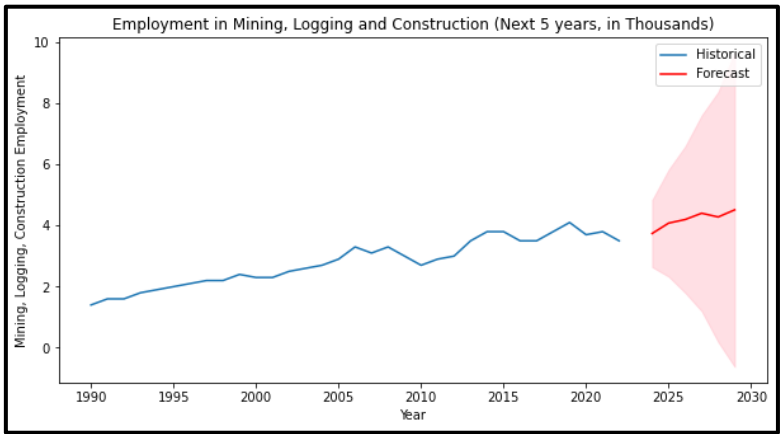

Figure 8: *Employment in Mining, Logging, Construction projected for the next 5 years*

Figure 8 represents the forecast for the level of employment in the city over the next 5 years. The number of employees in this sector has been steadily increasing since 1990 and that is expected to continue even for the foreseeable future.

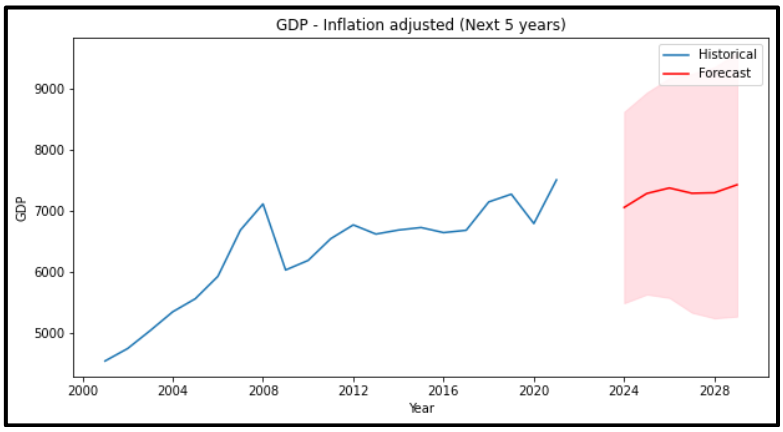

Figure 9: *Inflation Adjusted GDP projection for the next 5 years*

Figure 9 represents the GDP forecast for the next 5 years into the future. While the GDP is expected to increase steadily over time, the increase does not keep up with the environmental consequences that I described earlier. It appears that Cheyenne's economy may fall behind if we do not adapt to the world's resolve for sustainable development.

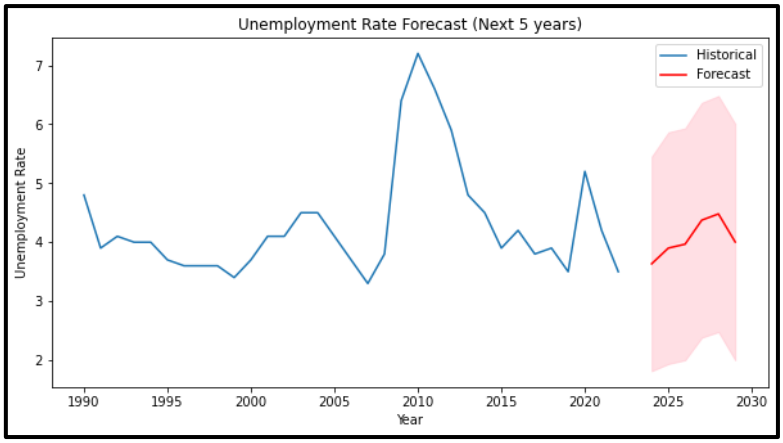

Figure 10: *Unemployment forecast for the next 5 years.*

Figure 10 represents the level of unemployment (%), the peak between 2008 and 2012 can be attributed to the great crash in 2008, it appears Cheyenne took a substantial hit and took substantial time to recover as well. To add on to my point about the GDP, the unemployment rate projections convey a similar message. Over the next 5 years, unemployment is expected to increase slightly. As the world moves to more sustainable means to grow their economy, it is time we look at more environmentally-friendly alternatives to make a living.

From these exploratory projections, it appears that the city of Cheyenne's economy is losing its insofar successful capitalization of industries like mining, logging, trading, construction, etc., as the world begins to transition to more sustainable methods of growing their economies. It is possible Cheyenne's economy may worsen if measures to introduce a shift in their economic activities is not put in place soon.

Methodology:

The goal of this project was to assess the impact of different economic activities on the AQI of Cheyenne, WY and their impact to Cheyenne's economy. One way to measure how change in AQI may affect the economy and vice-versa. I will use the Ordinary Least Squares Model to predict the per-capita income as a response to the set of predictors being:

1. Levels of employment in mining, logging, construction
2. Population
3. Employment in transportation
4. Employment in goods production
5. Unemployment Rate
6. Employees in goods trading
7. AQI

.

Save for the AQI data which I obtained from the EPA; I obtained data for the other predictors from the Federal Reserve Bank of St. Louis (FRED).

I have chosen the per-capita income as a metric to compare economic development, this would represent the average person's economic success or failure over time. In order for the average to prevent being skewed, the unemployment rate is taken into consideration as well, as part of the model.

The model only takes data from 2001 to 2021 into account (data over 21 years). This is due to inconsistencies in the time periods for which data is available for different aspects. This is the only timeframe for which all columns have consistent non-null data.

To determine the relationship between the GDP per-capita and the other predictors, the mathematical relationship is modelled as:

$$gdp\_per\_capita = \beta_0 + \beta_1 (emp\_min\_log\_con) + \beta_2 (population) + \beta_3 (emp\_transport) + \beta_4 (emp\_production) + \beta_5 (unemployment\_rate) + \beta_6 (emp\_goods\_trade) + \beta_7 (AQI) \quad - (1)$$

The OLS model is fitted, as described above, and the results I obtained are below:

OLS Regression Results

Dep. Variable:	per_capita	**R-squared:**	0.983
Model:	OLS	**Adj. R-squared:**	0.972
Method:	Least Squares	**F-statistic:**	87.58
Date:	Mon, 11 Dec 2023	**Prob (F-statistic):**	1.83e-09
Time:	22:17:22	**Log-Likelihood:**	-174.44
No. Observations:	21	**AIC:**	366.9
Df Residuals:	12	**BIC:**	376.3
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-3.987e+06	9.41e+05	-4.235	0.001	-6.04e+06	-1.94e+06
Year	2042.5155	480.817	4.248	0.001	994.906	3090.125
unemp_rate	1906.1049	578.934	3.292	0.006	644.716	3167.493
mine_log_con	-1630.1290	6770.795	-0.241	0.814	-1.64e+04	1.31e+04
Population	-1457.7793	457.047	-3.190	0.008	-2453.600	-461.959
employees_goods_production	5204.0679	6773.214	0.768	0.457	-9553.497	2e+04
employees_transportation	1613.1466	4469.290	0.361	0.724	-8124.600	1.14e+04
employees_trading	1416.0878	2459.866	0.576	0.575	-3943.500	6775.676
AQI	627.9932	205.862	3.051	0.010	179.458	1076.529

Omnibus:	0.690	**Durbin-Watson:**	2.379
Prob(Omnibus):	0.708	**Jarque-Bera (JB):**	0.265
Skew:	-0.275	**Prob(JB):**	0.876
Kurtosis:	2.963	**Cond. No.**	6.70e+06

Figure 11: *gdp_per_capita as a response predicted against predictors in (1)*

Figure 11 consists of a statistical summary of the OLS model. It is observed that the per_capita_income has a statistically significant association with the AQI, if we were to control year, unemployment rate and the population. It indicates that even a city at relatively low risk like Cheyenne is prone to having its economy impacted by wildfires that may impact the AQI. This is even more so of an impact, considering that the per_capita_income does not have a statistically significant relationship with the number of employees in mining, construction, production, transportation, and trading – some industries that are said to be pillars of Cheyenne's economy.

To check if my model satisfies linearity assumptions, I plotted the residuals vs. fitted values.

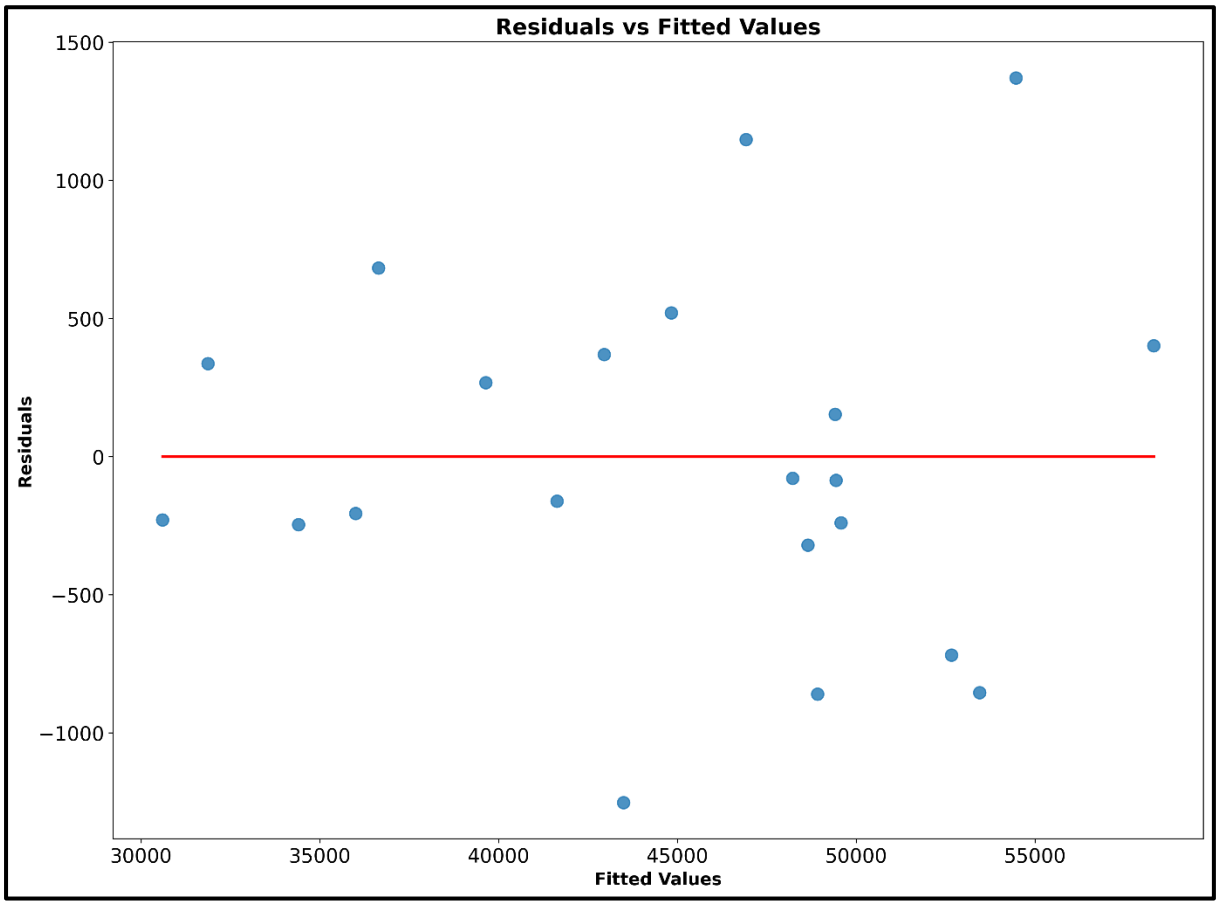

Figure 12: *Residuals vs. Fitted Values Graph*

Figure 12 depicts the residuals vs. fitted values graph for the GDP per-capita. Owing to the flatline of the line, I infer that the linearity assumption is upheld, with the exception of some anomalous values.

Inference:

Over the course of this project, in the case of Cheyenne, I infer that:

- AQI affects the per capita income
- The industries Cheyenne Depends on, will not yield as much revenue in the future; the city council is urged to shift their attention towards other industries that are likely to last in the future
- Focus on renewable sources of energy. Wind energy is a good place to start. Wyoming is home to 2 of the top 5 windiest cities in the US [5] (Cheyenne being one of them), thus there is tremendous potential with harnessing wind energy.

In addition, despite how Cheyenne is at low risk at this time, this can change in the near future, as the area burned due to fires continues to increase, and the number of wildfires continues to increase as well.

Potential Improvements:

1. Availability of data for more years would have ensured a more complete picture. We currently have 20 years of data, which tells us some concerning stories, but we could be more assured if we had more data to work with
2. Weather-related data was absent, weather-data varies seasonally, and the season variation can also be tied to the likelihood of a fire breaking out.

Conclusion:

Wildfires are another manifestation of climate change, which among other factors of life, will negatively impact the economy. In the case of Cheyenne, WY, policymakers would do well by emphasizing on a different set of industries that they can pivot to, based on their strengths. The city council would also be benefitted by running awareness campaigns on the same concern. Ultimately, it comes down to citizens to be individually motivated for their collective efforts to make an appreciable difference.

REFERENCES

[1] Federal Emergency Management Agency (FEMA), "National Wildfire Risk Map", https://hazards.fema.gov/nri/map

[2] Wildfires Signal Miners Need To Worry About Climate Change, The Financial Post, https://financialpost.com/commodities/mining/wildfires-signal-miners-need-to-worry-about-climate-change

[3] How wildfires affect labor income, IZA Newsroom, https://newsroom.iza.org/en/archive/research/how-wildfires-affect-labor-income/

[4] The Effects of Wildfires on Logistics and the Supply Chain, More Than Shipping, https://www.morethanshipping.com/the-effects-of-wildfires-on-logistics-and-the-supply-chain/#:~:text=In%20extreme%20cases%2C%20wildfires%20can,making%20transportation%20even%20more%20challenging.

[5] Cheyenne Fifth Windiest City in the United States, Wyoming News, https://www.wyomingnews.com/news/cheyenne-fifth-windiest-city-in-united-states/article_b333d0d8-a1c9-5d58-b963-e25b29664eb4.html

Data Sources:

1. USGS, "Combined wildland fire datasets for the United States and certain territories, 1800s - Present (combined wildland fire polygons)", December 8, 2021, https://www.sciencebase.gov/catalog/item/61aa537dd34eb622f699df81

2. EPA, "Air Quality System (AQS) API" https://aqs.epa.gov/aqsweb/documents/data_api.html.

3. Federal Reserve Bank of St. Louis, data available at: https://fred.stlouisfed.org/series/ATNHPIUS45500Q