Course: CMPT 353 Project
Instructor: Greg Baker
Team Fuel and EVs: Parmveer Dayal, Edoardo Rinaldi, Kyle Mollard

Analyzing Fossil Fuel Consumption, Gas Prices, and Electric Vehicle Sales for a Sustainable Future

August 4, 2023
https://github.sfu.ca/kem16/353project

1. INTRODUCTION

The global energy landscape has witnessed significant shifts in recent years and, as the world faces challenges related to climate change and sustainable development, understanding the dynamics between fossil fuel availability, gas prices, and electric vehicles sales becomes crucial.
In our project, we analyzed data related to fossil fuel consumption to find the countries that consume the most oil and attempted to find whether there was a correlation between fuel consumption and EV sales. Moreover, as we progressed with our analysis, we also wanted to explore the possibility of a correlation between gas prices and EV sales, to better understand whether the recent increase in gas prices has had an influence on consumers' behavior towards EV's.

2. METHODS

2.1 Fuel Consumption

The fuel consumption data was downloaded from a Kaggle CSV file [1], containing fuel consumption data for each country, as well as the world as a whole. Fossil fuels consist of natural gas, crude oil and coal so, through the use of Pandas, we only considered the column containing crude oil data, since it is the resource that is refined and converted to gasoline [2]. Although we were unable to obtain the ratio of how much crude oil each country converts to gasoline, we based each country's crude oil breakdown to that of Canada's, which is about 65% of crude oil being converted into various fuels for transport [3].
Using Pandas, we grouped our data by country and summed the annual fuel consumption for each country. The data provided contained information from 1980-2019, but we only considered the timeline from 1992-2019. The reason for this adjustment is that in 1991 the Cold War ended leading to the breaking apart of the USSR, which significantly changed Russia's fuel consumption rates, as shown in Fig.1 [4][5].

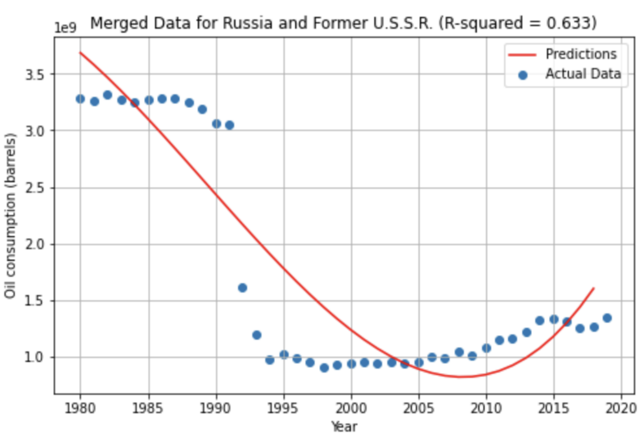

Figure 1 - USSR breaks apart into Russia

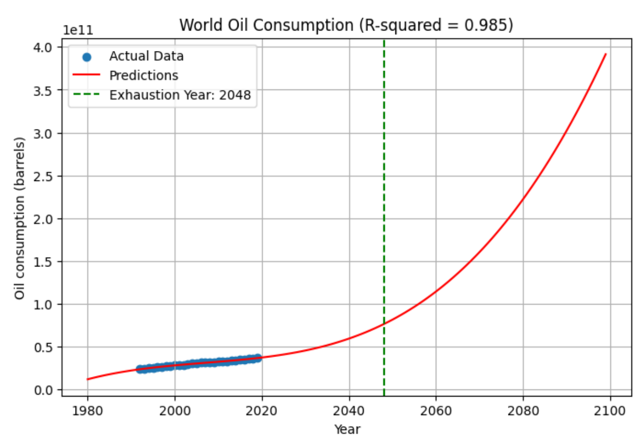

Figure 2 - World's Oil Consumption and Depletion

A second degree polynomial was used to fit each country's data, except for the US, which required a third-degree polynomial given the more prominent fluctuations in the data points.
Finally, to determine the range of years to make our future predictions, we analyzed the world's oil consumption, fitted it with a second degree polynomial and compared it to the current record of available crude oil reserves [6].

Each year's prediction was subtracted from the current available fossil fuels resources until they were completely exhausted. Although our analysis suggested complete depletion of resources by 2048, as shown in Fig. 2, many predictions of this nature have been largely underestimated, so we decided to extend our timeline to 2100. The limitations of this assumption will be further discussed in the "Limitations and Considerations" section at the end.

2.2 Fuel Cost

There are several datasets with annual gas prices available for the US, but it was challenging to find them for India, Japan, Russia and China. However, we were able to find statistics of gas prices for all the countries of interest [7], which allowed us to make a CSV file with the fuel costs.

The data points were very limited as they spanned from 2006-2022, except for China, which had only reports starting from 2009. Given the limited datasets and the different trends for each country, it was necessary to use different classifiers, as well adjusting the parameters to most accurately describe our data, as shown in the table below.

COUNTRY	Polynomial Features (degree)	kNN (neighbours)	Test size	Scores
India	-	3	0.3	0.338
Japan	5	-	0.4	0.633
Russia	5	-	0.3	0.914
China	5	-	0.5	0.423
US	-	3	0.2	0.830
World	-	3	0.2	0.756

Table I - Best Model Scores and Classifier Details

Usually, it is recommended to keep the size of the training data to 0.2-0.3, but the small samples from our dataset justified increasing the value to 0.4 or even 0.5 in the case of China, since it had even fewer data points than the other countries, as previously mentioned. Moreover, given the sudden fluctuations in prices, we limited the degree of our polynomial fit to 5 and the number of neighbours to 3 to avoid overfitting.

2.3 EV Sales

The EV sales data was downloaded from Kaggle and contained EV sales for every country and the world from 2010-2020. However, some countries such as Russia were missing data so it had to be merged from a different dataset, which is why its data starts from 2015. Similar to the Fuel Cost methods, the EV sales across the top five fuel consuming countries and the world were analyzed to identify trends in EV sales over time [8]. In order to do this, all countries selected were filtered and grouped by year to sum the sales for each year. Many tests such as polynomial regression, k-Nearest Neighbours, and Random Forest were conducted. A prediction model up to the year 2030 was also created for each test. From all the tests, polynomial regression performed the best to model and predict the relationship between the year and the total numbers of EVs sold. By fitting the countries data using polynomial features of degree 3 and either test sizes of 0.2 or 0.3, the majority of the models produced strong fits.

3. RESULTS AND DISCUSSION

3.1 Fuel Consumption

The R-squared values were very close to 1 for most countries, as shown in Fig.3 and Fig.4 below., indicating that the polynomial regression models effectively captured the oil consumption variations over time. However, both the US and Russia have the two lowest R-squared scores of 0.695 and 0.518, respectively. The lower scores can be attributed to the fact that their respective data points show much more fluctuations in comparison to the other three countries. For instance, we notice a spike in fuel consumption for the US from 2004-2007, since during those years the US was experiencing significant economic growth, low unemployment rates and an increase in car sales [9]. As for Russia, a likely reason for the low R-squared value, in addition to the fluctuations, is the initial outlier for 1992, immediately after the reformation of Russia from the USSR. In contrast to the majority of the world, Russia had more than enough oil, but kept prices low given its own needs and its growing infrastructure post Cold War, leading to the large consumption observed [5].

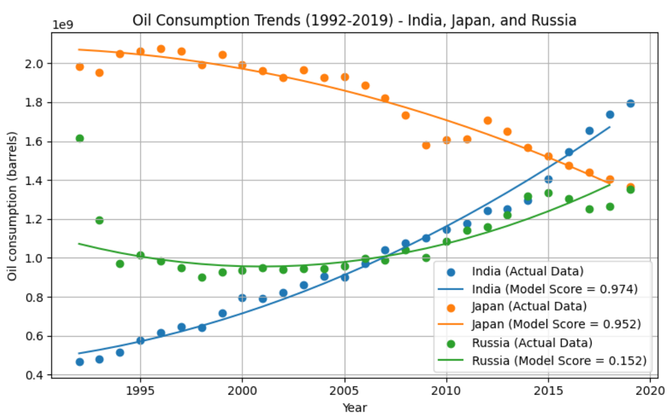

Figure 3 - Oil Consumption for India, Japan, Russia

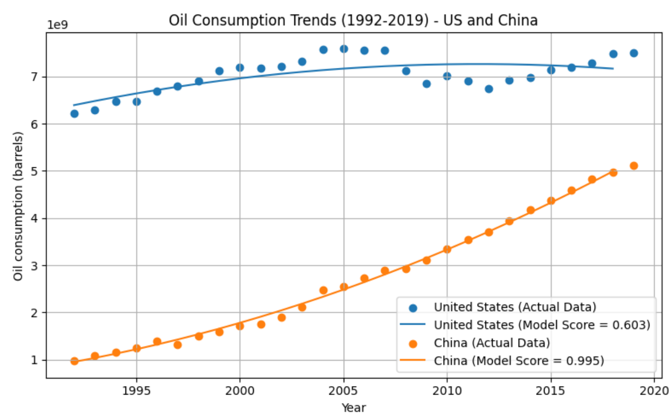

Figure 4 - Oil Consumption for US and China

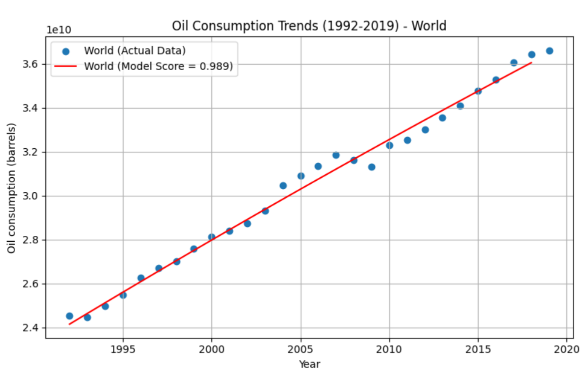

Figure 5 - Oil Consumption for World

From the analysis of oil consumption trends, we notice a steady increase in consumption for China and India, more fluctuations in the case of Russia and the US, while Japan is the only exception, showing a steady decrease. Given that Japan is a highly urbanized country with a large manufacturing industry, as well as having the third largest car market in the world, we were not expecting to see a decline in oil consumption. This is potentially attributed to the fuel efficiency of new cars, as well as an increase in popularity in hybrid and electric cars [10].

3.2 Gas Price

The prices of gasoline across the top five fuel consuming countries was analyzed by restricting the range between the years of 2006 and 2022, with the exception of China that only had gas prices reported from 2009. Unlike with the fuel consumption data, the gas prices data contained too few data points to make any sensible predictions in the long range (up until 2100), as we were originally planning. The trend for gas prices seems to always be on the rise, as shown in the figure below, which is expected given the limited supply of oil, with the exception of the years 2014-2016. During those years the world experienced a significant plunge in oil prices due to the US producing extremely large quantities of shale oil, which raised geopolitical concerns in foreign oil producing countries and resulted in a drop in oil prices [9][11][12].

With regards to classifier choices, Polynomial Features and K-Nearest Neighbors (kNN) yielded the best fit as discussed earlier, although kNN resulted in more choppy graphs given the limited data points and parametric nature of the classifier. We also tried to use MLPRegressor and Random Forest, but the fits were highly inaccurate due to the datasets containing very limited points, so we did not use them as a model.

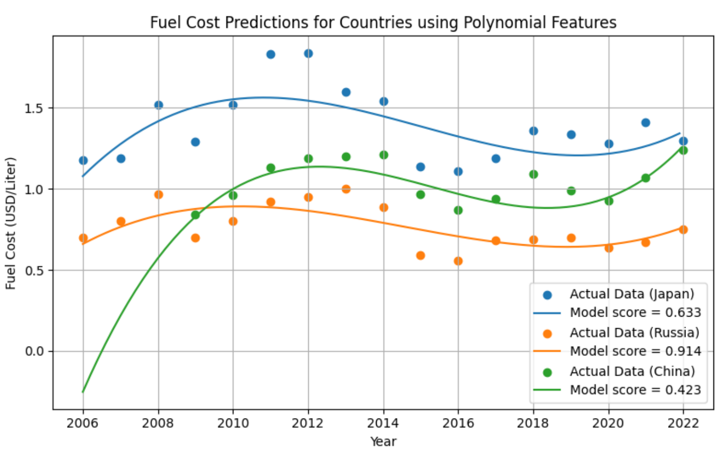

Figure 6 - Gas Price Predictions with Polynomial Feature

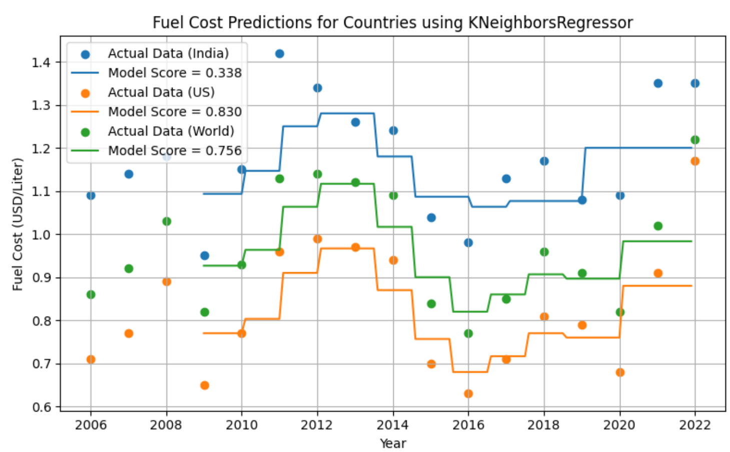

Figure 7 - Gas Prices Predictions with K-Neighbors Regressor

3.3 Electric Vehicle (EV) Sales

Polynomial regression was also used to predict EV sales across 2020-2030 to offer insights into potential EV sales trends. Model evaluation was performed using the R-squared metric on a validation set, with higher R-squared values indicating better model fit and accuracy. The figures below show the polynomial regression test for each country and the world. Every country's data shows EV sales increasing by a significant amount since 2020. The US sold the most EVs, with over 1,000,000 sold in just 2022 alone. Although Russia has sold by far the least amount of cars, their sales from 2020 to 2021 increased by over 300% [13]. Also, based on this model, every country's sales are expected to continue growing exponentially until at least 2030, which is consistent with the data and R-squared test.

Despite the success found in the majority of models, it is important to consider the issues found with the small and negative R-squared values in Japan and India, respectively [14]. R-squared values close to 0 imply a weak fit, indicating the model may not fully capture the variability in the EV sales data for these countries. For example, in Japan's case, the dip in sales during 2018-2020, seems to be caused by Japan's reluctance to continue investing in EV technology and infrastructure [15], unlike the other countries in this study. Secondly, many Japanese citizens continue to prefer Japanese cars and refuse to drive cars manufactured abroad. Also, hybrid cars in Japan are incredibly popular, especially the Toyota Prius [16]. On the other hand, India's EV sales jumped by 387% from 2020 to 2021 and by another 400% the following year. As shown in the plot and R-squared value, these two years act as outliers, making it difficult to find a strong model. However, we expect this increase in sales to continue as 2023 is projected to easily surpass 2022's sales [10]. Outside of these two countries, the remaining countries and the world provide strong R-squared values and evidence that EV sales will continue to increase yearly.

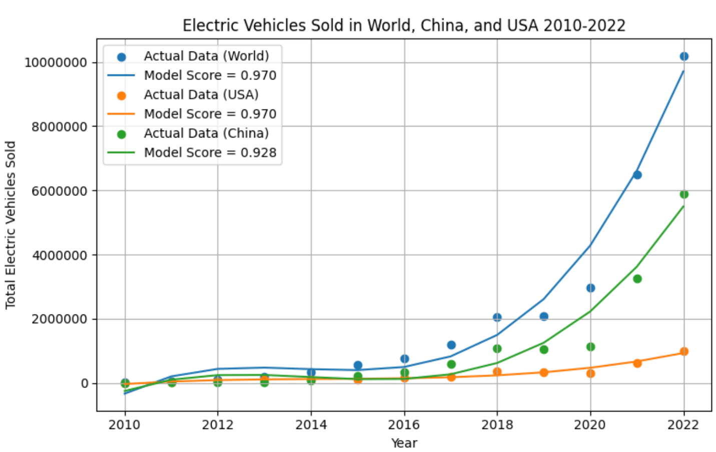

Figure 8 - EVs Sold in World, China and USA

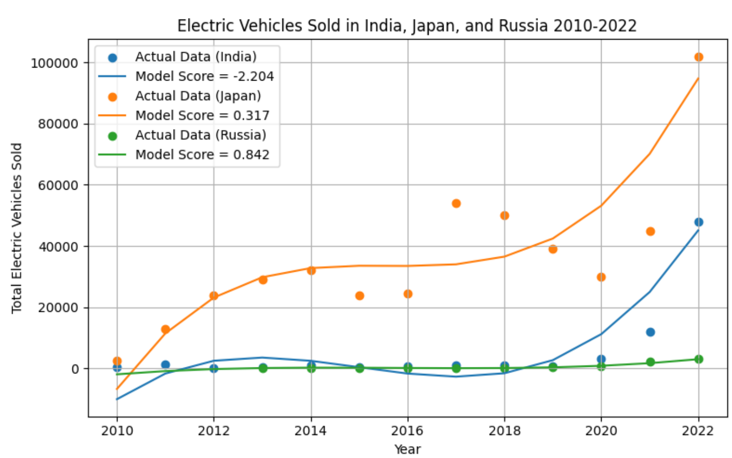

Figure 9 - EVs Sold in India, Japan, and Russia

3.4 Correlation Investigation

To investigate correlations between gas prices, fuel consumption, and electric vehicle (EV) sales we used the corrcoeff function from Numpy. At the country level, the results showed minimal to no significant correlations; however, the correlation coefficient for World Fuel Consumption and EV Sales was calculated to be 0.953, indicating a strong positive relationship, as the Fig.9 and Fig.10 below show. Conversely, the correlation between World Gas Price and EV Sales yielded a coefficient of 0.308, suggesting a relatively weak or insignificant relationship. The correlation values seem counterintuitive, as we would expect to see a negative correlation between fuel consumption and car sales and a more strong positive correlation between gas prices and EV sales.

In the past decade, however, gasoline cars have become more affordable and transportation networks have been evolving world wide, while EV's have been steadily gaining traction amongst consumers, hence both fuel consumption and EV sales have been increasing.

On the other hand, since EV sales have been showing a much more predictable trend, whereas gas prices are always prone to sudden fluctuations, it is reasonable to see a relatively low coefficient relationship between the two variables. Also, even though EV sales continue to increase at a rapid pace, there are still significantly more gasoline cars being sold in comparison, which is shown by the fuel consumption continuing to increase.

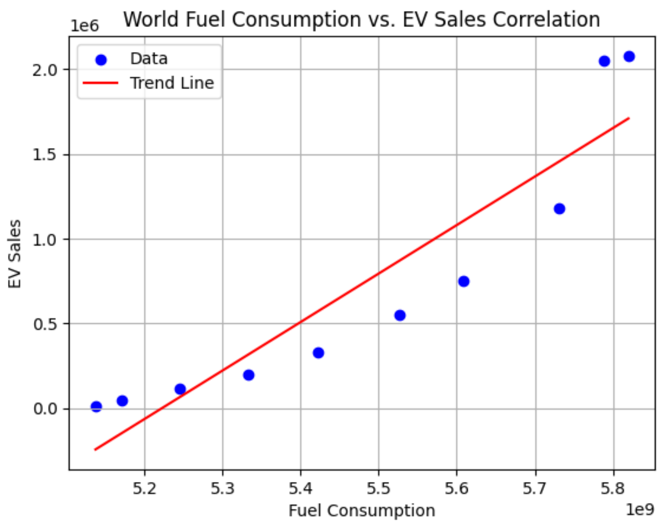

Figure 9 - EV Sales vs. Fuel Consumption

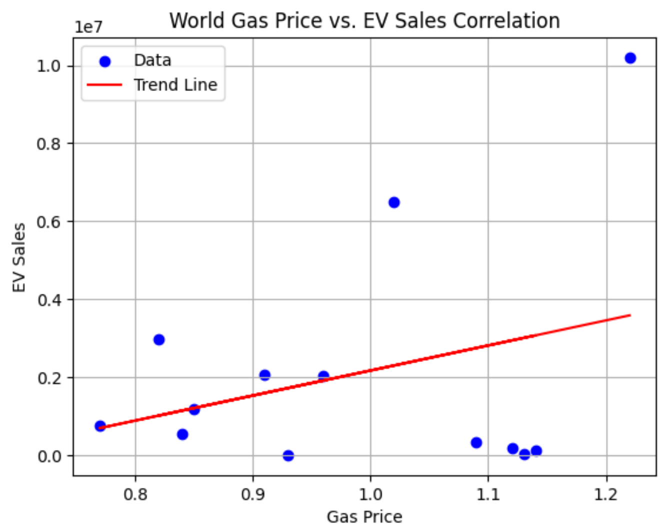

Figure 10 - EV Sales vs. Gas Price

4. CONCLUSION

Analysis of the fuel consumption data showed that the top five oil-consuming countries in the world, with the exception of Japan, are predicted to have increasing consumption rates until 2100, should the fossil fuels reserves be sufficient until then. This is expected as every country's consumption rates continue to increase and the model will also predict an increase, since it does not account for certain variables such as the recent increase in EV sales, how much oil is left, or the economic status of the people living in a specific country.

After investigating EV sales related to each of the five countries, we found a negligible correlation between sales and fuel consumption. However, there was a much stronger positive correlation of 0.953 between electric car sales and fuel consumption on a global scale, suggesting a potential relationship. On the other hand, the correlation between EV sales and gas prices had a value of 0.308, indicating a weaker relationship.

Although the correlation findings neglect many variables that should be accounted for, such as economic growth, market fluctuations and geopolitical factors specific to each country, we cannot conclude that a rise in EV sales resulted in a drop in fuel consumption. This question will be better answered in 10 or more years, since we expect to see not only more EVs but also other gasoline powered engines switching to batteries. As of today, there are simply not enough EVs on the road compared to gasoline cars, in order to significantly impact fuel consumption. Lastly, the low correlation coefficient between EV sales and gas prices may suggest a weak relationship between the two variables on a global scale, but in order to draw more significant conclusions, the data should be analyzed and modeled over a longer time span.

5. LIMITATIONS

The analysis of fuel consumption, gas price, and EV sales was influenced by various factors, with constraints due to data availability, especially for countries in Asia and Eastern Europe. One of our assumptions, for instance, was that Canada's crude oil to gasoline conversion (65% of crude oil is refined in various forms of gasoline) applied to all the countries in the analysis, but to draw a more accurate conclusion the exact conversion rate should be known for each country. Also, there have been many predictions on when fossil fuel resources would be exhausted, which have all been proven wrong with the discovery of new drilling sites and implementation of more advanced technologies, so our prediction that fossil fuels will be depleted by 2048 is a rough estimate. Moreover, both the fuel consumption and gas prices data had relatively few data points. To build a better model and make better predictions, it would be advisable to obtain the consumption rates and gas prices on a monthly basis over a 20 year span, which would allow our classifiers to perform significantly better. Also, this project did not not look into why we are seeing a big rise in EVs over the past couple years. Two potential reasons for this spike are that EVs are much more accessible today, with more large car manufacturers producing EVs, and many countries funding projects to limit carbon footprint. Since EVs have just been gaining traction in recent years, to build a better model of EV sales and make more accurate future predictions, it may be necessary to track the sales trends over a longer period of time, especially within the next 5-10 years, as car manufacturers continue to release more EV models at potentially lower prices.

6. PROJECT EXPERIENCE SUMMARIES

Parmveer Dayal

- Organized, cleaned, and transformed data using the Numpy and Pandas
- Applied machine learning techniques through Python's SciKit learn library to analyze the data and make predictions
- Analyzed the fuel consumption, gas price, and EV sales data to find a correlation between them
- Processed data into presentable results in Python

Edoardo Rinaldi

- Researched, cleaned and prepared data using Pandas and Numpy
- Critically analyzed and fitted the data through machine learning techniques learnt in class (SciKit learn library)
- Created models, as well as made predictions and plots related to the datasets
- Worked with other team members to explain quantitatively and qualitatively our findings in a technical report

Kyle Mollard

- Worked collaboratively in a team of three members, utilizing GitHub and Google Docs to create a comprehensive data analysis report.
- Evaluated, cleaned, and transformed data using Python's Pandas and Numpy libraries in preparation for model fitting
- Investigated outliers thoroughly to contextualize and understand the data's domain
- Reviewed a variety of machine learning methods using the SciKit-learn library

7. REFERENCES

[1]*Fuel Production and Consumption(1980-2021)*. (n.d.). https://www.kaggle.com/datasets/shawkatsujon/worldwide-fuel-production-and-consumption

[2]Ritchie, H., Roser, M., & Rosado, P. (2022). *Fossil Fuels*. Our World in Data. https://ourworldindata.org/fossil-fuels

[3]CAPP. (2019). *What Is Oil Used For | Uses for Oil and Petroleum Products*. CAPP. https://www.capp.ca/oil/uses-for-oil/

[4]The Editors of Encyclopedia Britannica. (2018). collapse of the Soviet Union | Causes, Facts, Events, & Effects. In *Encyclopædia Britannica*. https://www.britannica.com/event/the-collapse-of-the-Soviet-Union

[5] Uchitelle, L. (1992, July 12). THE WORLD; Russians Line Up for Gas as Refineries Sit On Cheap Oil. *The New York Times*. https://www.nytimes.com/1992/07/12/weekinreview/the-world-russians-line-up-for-gas-as-refineries-sit-on-cheap-oil.html

[6]WorldOmeter. (2016). *World Oil Statistics - Worldometer*. Www.worldometers.info. https://www.worldometers.info/oil/

[7]*End-Use Prices Data Explorer – Data Tools*. (n.d.). IEA. https://www.iea.org/data-and-statistics/data-tools/end-use-prices-data-explorer?tab=Yearly+prices

[8]*Historic sales of electric vehicles*. (n.d.). Www.kaggle.com. Retrieved August 5, 2023, from https://www.kaggle.com/datasets/edsonmarin/historic-sales-of-electric-vehicles

[9]Stocker, M., Baffes, J., & Vorisek, D. (2018, January 18). *What triggered the oil price plunge of 2014-2016 and why it failed to deliver an economic impetus in eight charts*. World Bank Blogs. https://blogs.worldbank.org/developmenttalk/what-triggered-oil-price-plunge-2014-2016-and-why-it-failed-deliver-economic-impetus-eight-charts

[10]U.S Energy Information Administration. (2020, November 2). *International - U.S. Energy Information Administration (EIA)*. www.eia.gov. https://www.eia.gov/international/analysis/country/JPN

[11]*With the Benefit of Hindsight: The Impact of the 2014-2016 Oil Drop Collapse.* World Bank - Global Economic Prospects, 2018 https://thedocs.worldbank.org/en/doc/910311512412250749-0050022017/original/GlobalEconomicProspectsJan2018TopicalIssueoilpricecollapse.pdf

[12]*A Crude Shock : Explaining the Impact of the 2014-16 Oil Price Decline Across Exporters*. (2014). IMF. https://www.imf.org/en/Publications/WP/Issues/2017/07/18/A-Crude-Shock-Explaining-the-Impact-of-the-2014-16-Oil-Price-Decline-Across-Exporters-44966

[13] “Electric vehicles (EV) car parc in Russia from 2013 to January 2023, by mileage”, *Statistica*, 28 Feb. 2023, https://www-statista-com.proxy.lib.sfu.ca/statistics/1082308/russia-ev-car-parc-by-mileage/

[14]www.ETAuto.com. (n.d.). *EV sales in India increases 174% YoY to 12,47,120 units in FY23 - ET Auto*. ETAuto.com. Retrieved August 5, 2023, from https://auto.economictimes.indiatimes.com/news/industry/ev-sales-in-india-increases-174-yoy-to-1247120-units-in-fy23/100156256

[15]“Dooley, B., & Ueno, H. (2021, March 9). Why Japan Is Holding Back as the World Rushes Toward Electric Cars. *The New York Times*. https://www.nytimes.com/2021/03/09/business/electric-cars-japan.html

[16]Bekker, H. (2023, January 11). *2022 (Full Year) Japan: Best-Selling Car Brands and Market Analysis*. Car Sales Statistics. https://www.best-selling-cars.com/japan/2022-full-year-japan-best-selling-car-brands-and-market-analysis/

[17]Hamilton, J. (2009). *Causes and Consequences of the Oil Shock of 2007-08*. https://www.brookings.edu/wp-content/uploads/2016/07/2009a_bpea_hamilton-1.pdf