Air Traffic's Impact on Global Economy & Co2 Emissions Final Presentation

Executive Summary

Rising environmental concerns regarding air traffic emissions pose major roadblocks to the future of trade and travel in our global economy

Without efficient, reliable, sustainable air travel/trade, major impacts on the world economy are anticipated. Nations continue to impose CO_2 targets, yet emissions continue to rise

Using public data sources, our project identifies a need for global **gov't air**, **trade**, and **enviro regulating bodies**, **nationalized/privatized airlines**, to analyze current economic/ environmental relationships b/w air travel and trade

Data prep/manipulation thru Python/ Y-J trans tech. and LRM reveal 43% positive correlation b/w DVs & Ivs. Further in-depth analysis will provide regulators, private interests better decision-making insight

Business Goal(s)

Allowing stakeholders in-depth understanding of the links between **environmental** and **economic interests**

Evaluating recent/ long-term **trends**, isolating **outlier trends**, identifying **permanent barriers** to short-term and long-term profitability

Making detailed **short-term** and high-level **long-term projection analyses** for the global aviation industry/regulators

Data Analytics Goal(s)

Descriptive retrospective analysis linking **GDP growth** to **air traffic data** and **CO2 emissions**, examining causal links b/w variables

Making detailed **short-term** and high-level **long-term projection analyses** for the global aviation industry/regulators, focusing on int'l arrivals, CO2, GDP

Evaluating recent/ long-term **trends** of **GDP growth/decline** thru respective nation currency strength

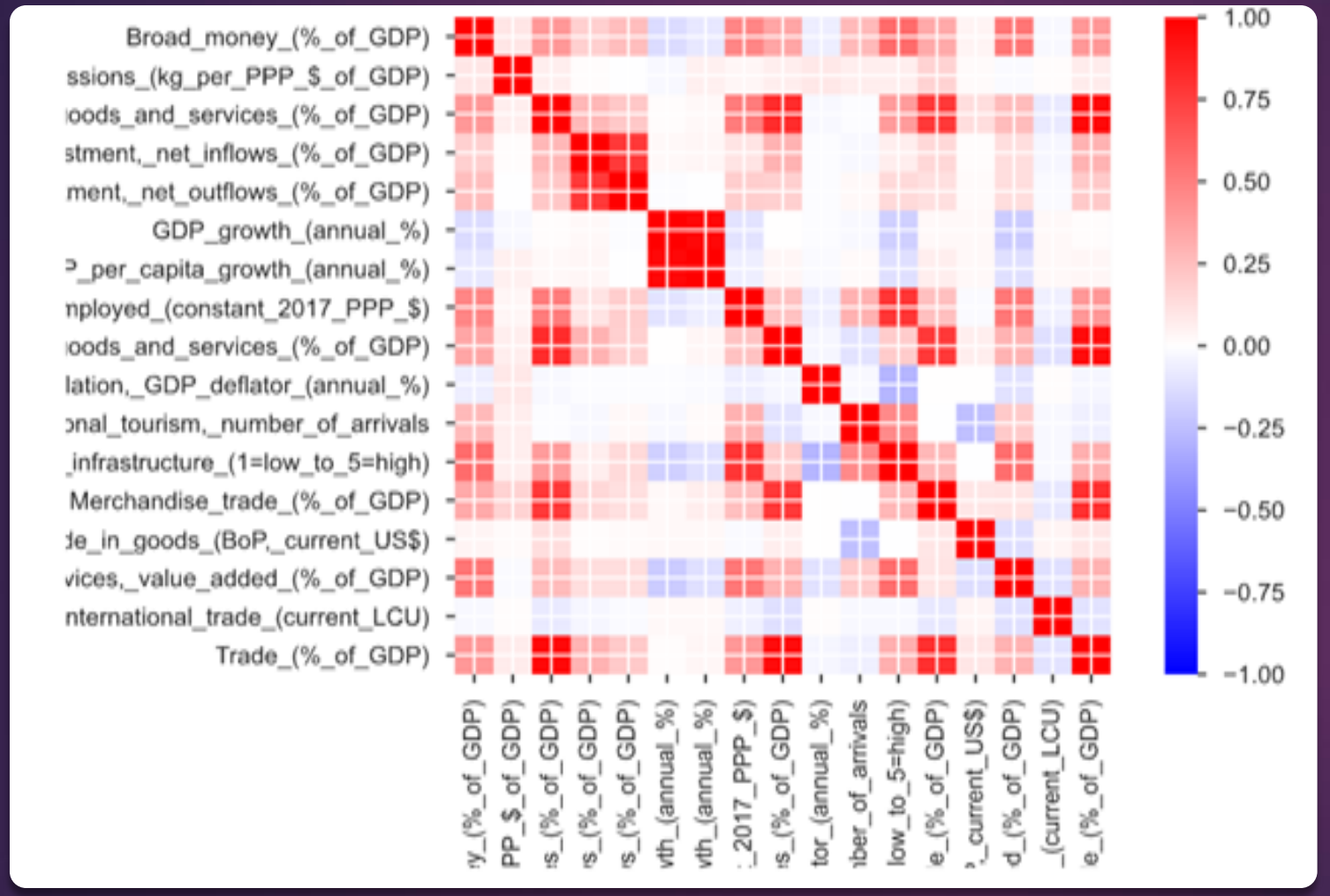
Data Description and Preparation

Data Sources

- Data was gathered from various sources and required **processing**, **cleaning**, **review**, outlier removal, distribution identification, ID'd **missing**, **irrelevant**, **duplicate values**, eliminated **non-uniform data** through **Python** using **profiling**
- **4989 records** in total comprised **17 columns** and **16 DVs**
- All data before 1995 was removed, and all relevant was combined data into **single format** and **regression analysis** was performed through **Python**
- Also dropped country name/code/series code, then wrote data into csv, then dropped columns w/ high missing values and were replaced null values w/ 0s
- **Processing** for **correlation(s)** between **key variables** of reference, **Tableau** and **PowerBI** was used for **included visualization(s)**
- Data set was split 80% train and 20% test

Preparation Details

- Using **classification analysis pollution data** was examined, classifying contributors separately defined as: **road transport** and **air transport**, if available: **rail transport**
- **EDA** performed through **Python** showed unbalanced dataset
- Air passenger dataset underwent **classification analysis**, ID'd as passenger, cargo, domestic & international travels
- Used **outlier detection/regression analysis** when predicting future trend(s) **of Co2 output** related to **domestic/ int'l travel** activity
- Ran scatterplot to id dist., split into x/y Vs, created PL for OMLC
- Dev'd LRM, YJTT deployed to optimize results



Pearson R Correlation Heatmap

Data Analysis Solution

Linear Regression Model

- R-Sq: Evaluated extend to which variance of single variables related to 2nd V variance. If R2 model is 1-100%, model is = .5, half of observed output causally linked to model input
- Adj R-Sq: Adjusted sq for number of model predictors. Adj R-sq increases at a higher rate of improvement than expected by chance.
- Decrease is observed when predictor improves model by less than expected. Adj R-sq is reliably positive, and always lower than R-sq
- RMSE: RMSE observe sqrt of mean of all error, used to check common error metric for all LRM

Linear Regression Outcomes

- The coefficients of the variables were in a range of between -9.61 to 9.29. which states that there is less correlation between the variables
- The R2 and the Adjusted R2 had the same result of 18%. Which states that relationship between the dependent and independent variable is very low

Original Model

Method: Linear regression

Intercept: 8281199.96

	Coefficients
Broad money (% of GDP)	2.065909e+06
CO2 emissions (kg per PPP $ of GDP)	8.264646e+05
Exports of goods and services (% of GDP)	6.803247e+14
GDP growth (annual %)	-9.612838e+06
Foreign direct investment, net inflows (% of GDP)	-1.062560e+06
Foreign direct investment, net outflows (% of GDP)	2.850820e+05
GDP per capita growth (annual %)	1.028319e+07
Imports of goods and services (% of GDP)	6.470432e+14
Merchandise trade (% of GDP)	3.208575e+03
Inflation, GDP deflator (annual %)	-1.580671e+05
Trade (% of GDP)	-1.282993e+15
Services, value added (% of GDP)	9.296560e+05
GDP per person employed (constant 2017 PPP $)	6.652097e+06
Net trade in goods (BoP, current US$)	-5.338564e+06

R2: 0.18
Adj_R2: 0.18
Mean Absolute Error: 10256488.91
Mean Squared Error: 466004413230199.19
Root Mean Squared Error: 21587135.36

Linear Regression Outcomes

- The linear model was optimized with grid search function
- Results of the R2, Adjusted R2, MAE and the RMSE outputs are similar to non-optimized model

```
Optimized Model

Model Name: LinearRegression()

Best Parameters: {'clf__copy_X': True, 'clf__fit_intercept': True}

Intercept: 8281199.96

Model coefficients:
Broad money (% of GDP) 2065909.18
CO2 emissions (kg per PPP $ of GDP) 826464.65
Exports of goods and services (% of GDP) 680324724601488.2
GDP growth (annual %) -9612837.51
Foreign direct investment, net inflows (% of GDP) -1062560.19
Foreign direct investment, net outflows (% of GDP) 285082.02
GDP per capita growth (annual %) 10283194.08
Imports of goods and services (% of GDP) 647043245062714.8
Merchandise trade (% of GDP) 3208.58
Inflation, GDP deflator (annual %) -158067.07
Trade (% of GDP) -1282993138094050.2
Services, value added (% of GDP) 929656.01
GDP per person employed (constant 2017 PPP $) 6652096.93
Net trade in goods (BoP, current US$) -5338564.47

R2: 0.18
Adj_R2: 0.18
Mean Absolute Error: 10256488.91
Mean Squared Error: 466004413230199.19
Root Mean Squared Error: 21587135.36
```

Linear Regression Outcomes

Using YJ modeling we transformed numeric variables to achieve normal distribution

- The coefficients of the variables were between -9.61 to 9.29 which is same as previous model results
- The R2 & Adjusted R2 is same which is at 43% which is 25% higher than optimized liner model results
- The RMSE value is at 2.13

Transformed Model - Yeo-Johnson

Method: Linear regression

Intercept: -0.00

	Coefficients
Broad money (% of GDP)	2.065909e+06
CO2 emissions (kg per PPP $ of GDP)	8.264646e+05
Exports of goods and services (% of GDP)	6.803247e+14
GDP growth (annual %)	-9.612838e+06
Foreign direct investment, net inflows (% of GDP)	-1.062560e+06
Foreign direct investment, net outflows (% of GDP)	2.850820e+05
GDP per capita growth (annual %)	1.028319e+07
Imports of goods and services (% of GDP)	6.470432e+14
Merchandise trade (% of GDP)	3.208575e+03
Inflation, GDP deflator (annual %)	-1.580671e+05
Trade (% of GDP)	-1.282993e+15
Services, value added (% of GDP)	9.296560e+05
GDP per person employed (constant 2017 PPP $)	6.652097e+06
Net trade in goods (BoP, current US$)	-5.338564e+06

R2: 0.43
Adj_R2: 0.43
Root Mean Squared Error: 2.13

Polynomial Regression Experiment

We decided to implement an experimental polynomial regression model to compare its performance with the linear models.

- Optimal degree was 2
- Had the best performance metrics
- Given additional time we would further explore polynomial models to see if they perform better than linear models

```
Polynominal Regression Model

Polynomial Degree: 2

Intercept: 174358127.61


R2: 0.48
Adj R2: 0.48
Root Mean Squared Error: 17239378.50
```

Conclusions

Advantages

- Our intended outcome was detailed 4-year retrospective analytical model
- Allowing more precise marketing, regulating, manufacturing, trade-related decisions to be made, intended to be used by:
  - Gov't regulators
  - Aviation industry bodies
  - Industry-based manufacturing

Limitations

- Anticipated and experienced limitations include:
  - Difficulty of optimal model selection
  - Scarcity of public info and rigid project lifespan
  - Can only provide predictive modeling based on data collected at specific date(s)
  - As mentioned in previous ppt, diverse public datasets impacted accuracy
  - While positive correlation of early models exists, right variable combination would require more time and effort than is currently available
  - For future studies will need to engage in revised collection, may be able to devise a strategy to update regularly

Conclusions

Operational Recommendations

- After transformation accuracy increased, but models did not meet required standards
- Possibility of inaccurate early models confirmed, in future, with a longer lifespan, we intend to magnify scope of project, and we have identified need to focus on polynomial or other regression modeling
- For next time it is recommended:
  - Outlier detection models should be implemented early and often for correction purposes, i.e. Tukey
  - Further model exploration to be carried out to identify suitable model(s), with a focus on PNRM, narrow down from there
  - Remove LRM focus, focus primarily on PNRM, while exploring other models, and identify optimal combination of IVs through this lens

Appendices

Dataset info

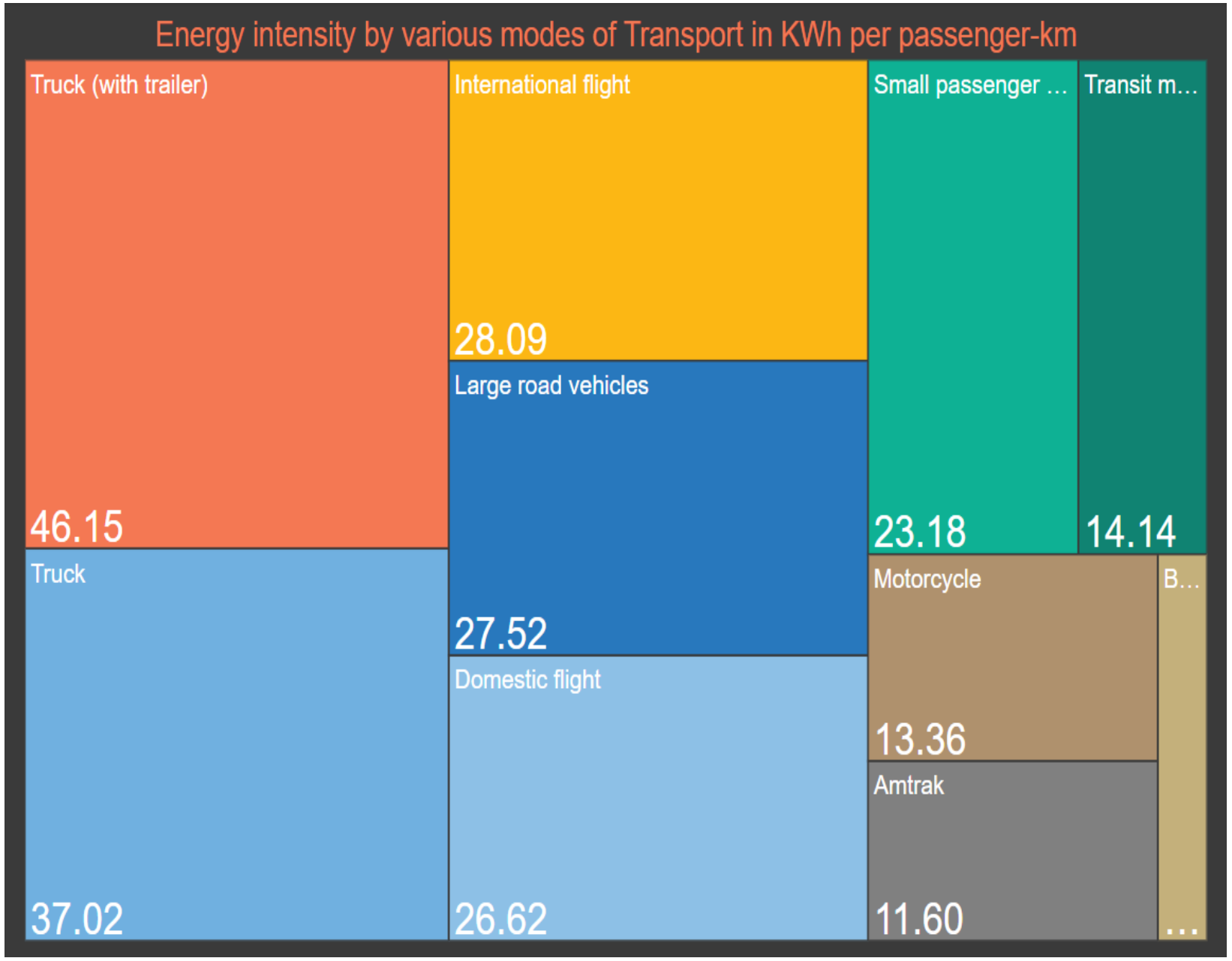
Number of variables	17
Number of observations	4989
Missing cells	13188 (15.5%)
Duplicate rows	6 (0.1%)
Total size in memory	662.7 KiB

Dataset Sample

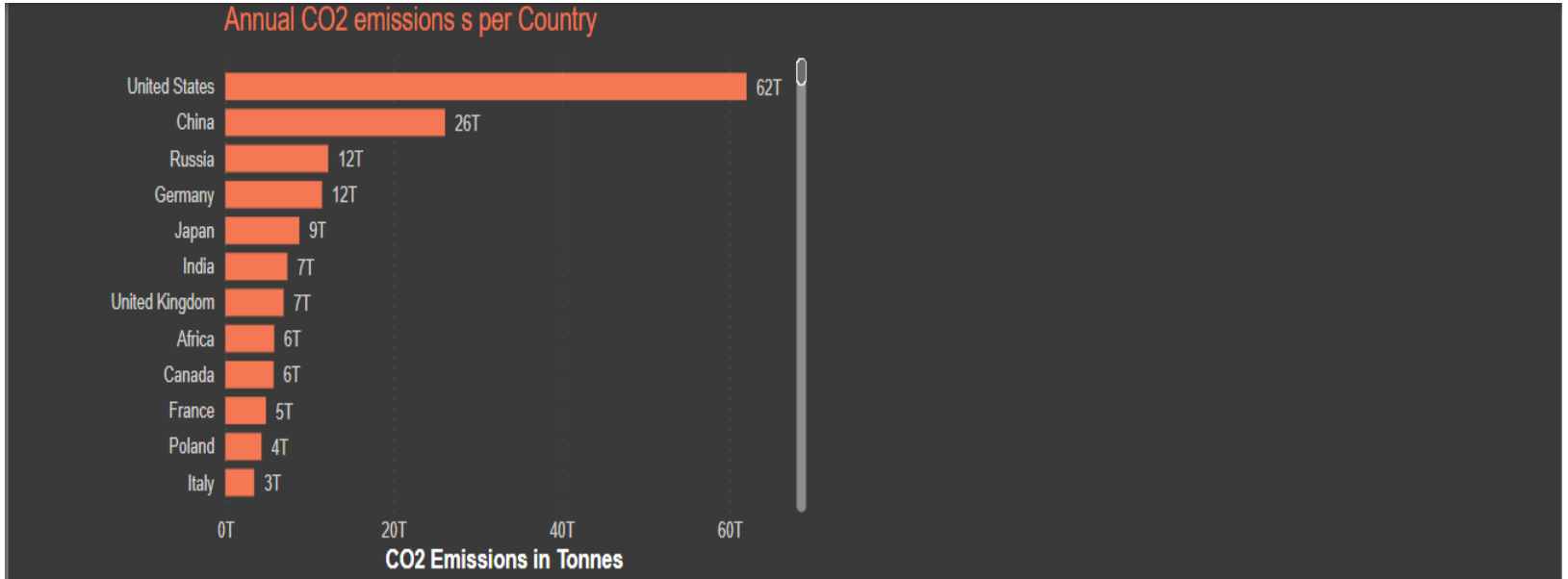
	Broad money (% of GDP)	CO2 emissions (kg per PPP $ of GDP)	Exports of goods and services (% of GDP)	GDP growth (annual %)	Foreign direct investment, net inflows (% of GDP)	Foreign direct investment, net outflows (% of GDP)	GDP per capita growth (annual %)	Imports of goods and services (% of GDP)	Merchandise trade (% of GDP)	Inflation, GDP deflator (annual %)	Trade (% of GDP)	Services, value added (% of GDP)	GDP per person employed (constant 2017 PPP $
10	0.000000	0.000000	99.500000	-0.408163	0.000000	0.000000	-0.195793	123.600000	176.000000	-1.365583	223.200000	0.000000	0.00000
11	0.000000	0.000000	100.405680	-4.098361	0.000000	0.000000	-3.371501	129.411765	203.853955	2.813675	229.817444	0.000000	0.00000
12	0.000000	0.000000	97.490347	1.851852	0.000000	0.000000	3.146660	120.656371	152.702703	3.160612	218.146718	0.000000	0.00000
13	0.000000	0.000000	114.464286	-2.657343	0.000000	0.000000	-1.172397	135.357143	161.607143	11.059335	249.821429	0.000000	0.00000
14	0.000000	0.000000	78.518519	-4.166667	0.000000	0.000000	-2.788561	82.814815	143.111111	25.776398	161.333333	0.000000	0.00000

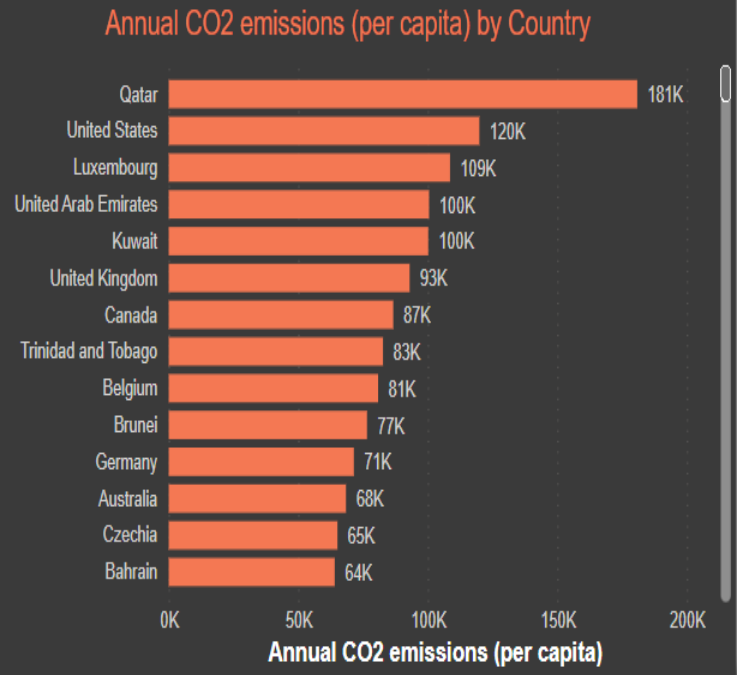
Null Values

	percent_missing
Logistics performance index: Quality of trade and transport-related infrastructure (1=low to 5=high)	81.639607
Taxes on international trade (current LCU)	45.259571
Broad money (% of GDP)	20.665464
Foreign direct investment, net outflows (% of GDP)	14.852676
CO2 emissions (kg per PPP $ of GDP)	14.752455
Net trade in goods (BoP, current US$)	13.589898
GDP per person employed (constant 2017 PPP $)	11.385047
International tourism, number of arrivals	11.124474
Imports of goods and services (% of GDP)	10.142313
Trade (% of GDP)	10.142313
Exports of goods and services (% of GDP)	10.142313
Services, value added (% of GDP)	8.558829
Foreign direct investment, net inflows (% of GDP)	4.810583
Merchandise trade (% of GDP)	2.926438
GDP per capita growth (annual %)	1.463219
Inflation, GDP deflator (annual %)	1.443175
GDP growth (annual %)	1.443175

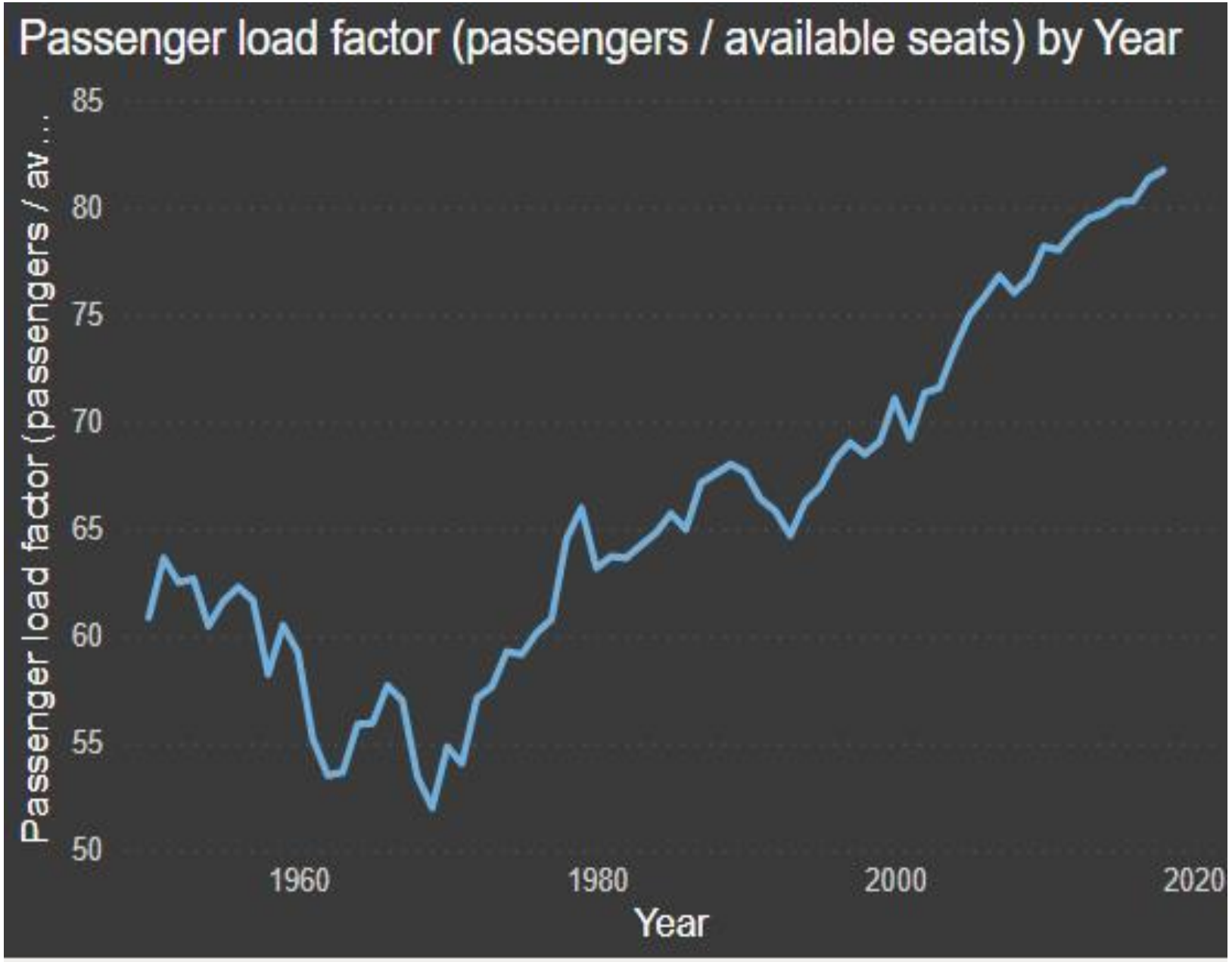


Visualization 1

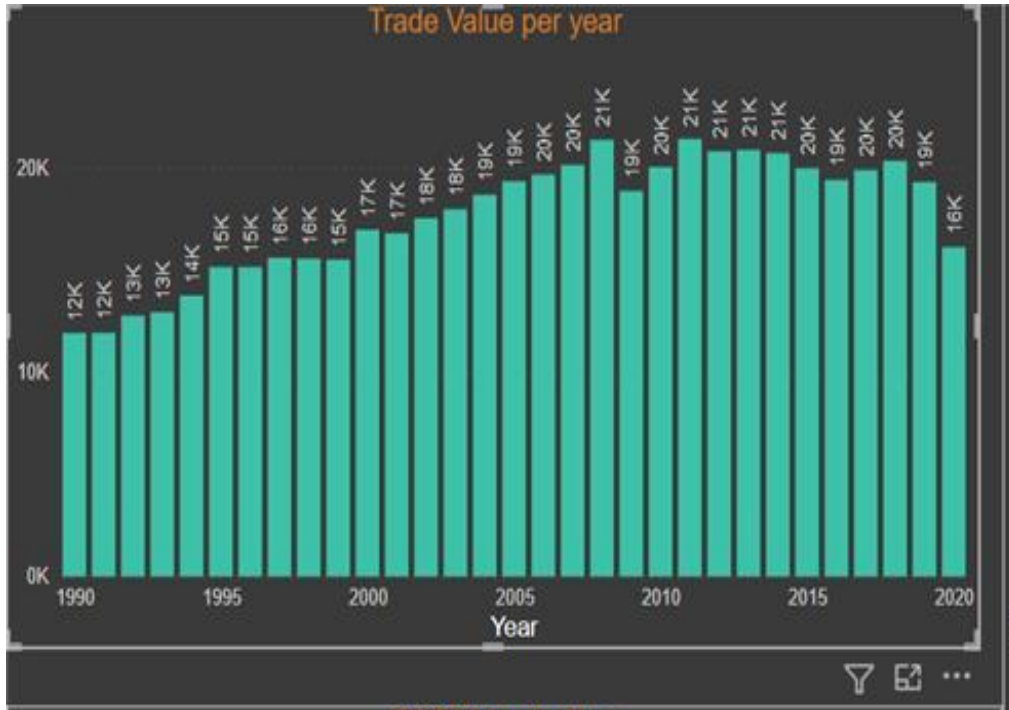


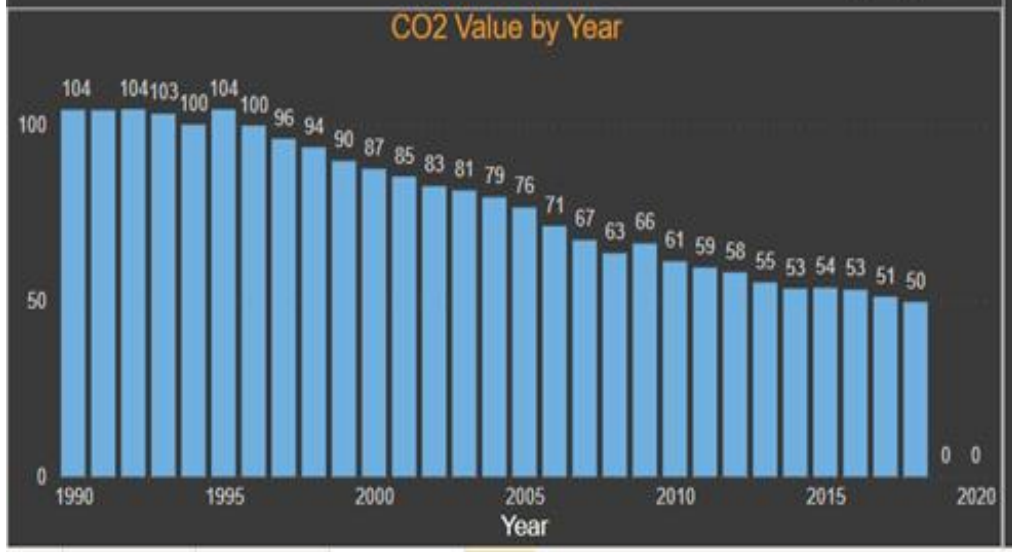


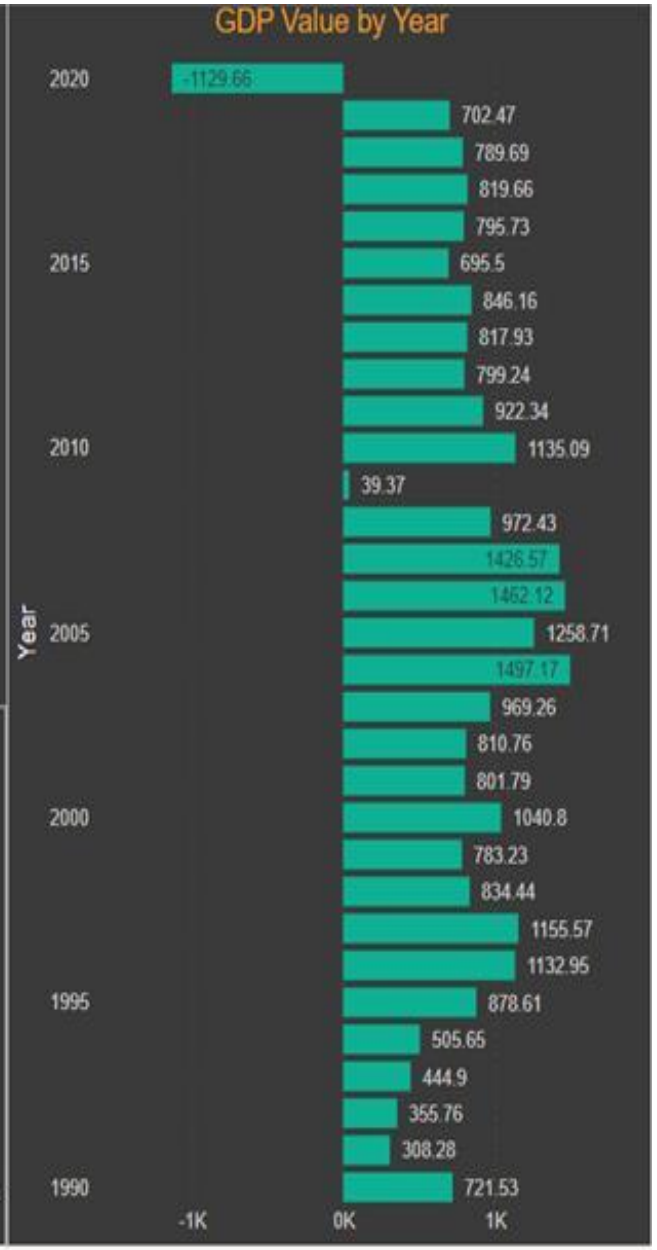
Visualization 2



Visualization 3







Visualization 4

References

- PWT 10.0. (2021). Retrieved 6 February 2022, from https://www.rug.nl/ggdc/productivity/pwt/
- Air Traffic Passenger Data. (2022). Retrieved 6 February 2022, from https://www.kaggle.com/yamqwe/air-traffic-passenger-datae
- COVID-19 Air Traffic Dashboard . (2022). Retrieved 6 February 2022, from https://www.icao.int/sustainability/Pages/COVID-19-Air-Traffic-Dashboard.aspx
- Ritchie, H., & Roser, M. (2020). Energy. *Our World In Data*. Retrieved from https://ourworldindata.org/transport
- World Bank Open Data | Data. (2022). Retrieved 6 February 2022, from https://data.worldbank.org/?most_recent_year_desc=false