|  |  |
| --- | --- |
|  |  |
| Data Analysis of Global Aviation Emissions |  |
|  | A picture containing clipart  Description automatically generated |
|  | 30/11/2022Data Analytics in Aviation *Sean Clarke*  *UCD Professional Academy* |

Table of Contents

Abstract3

Introduction3

Dataset3

Implementation Process4

Results9

Insights……………………………………………………………………………………………………………………………………………………………11

GITHUB - <https://github.com/clarkesean/UCDPA_Sean_Clarke>

Abstract

This report will seek to analyse the global emissions of the aviation sector, domestic and international. Aviation accounts [for approx. 2.5%](https://ourworldindata.org/co2-emissions-from-aviation) of global CO2 emissions annually (Ritchie, n.d.). The report will apply data analytics and data visualisation to compare flight emissions by country per capita and highlight potential trends or outliers.

The python programming language will be used to process, clean and visualise the data sets. This report will use two main data sources; ‘CO₂ emissions from domestic air travel, 2018’ and ‘CO₂ emissions from international air travel, 2018’. Both data sets rank nations aviation emissions per capita. Domestic aviation represents flights which depart and arrive within the same country while International aviation emissions are allocated to the country of departure for each flight.

Using data analytics this report will examine the societal and economic factors contributing to the results found.

Introduction

These data sources will be used to draw insights regarding inequality of emissions around the world, as we will see the variation in aviation emissions per capita between nations is significant. We will also analyse the primary factors behind these variations, from a nations GDP to tourism and commerce. As a sector, Aviation has a relatively small contribution to global emissions, however, this figure is heavily affected by the fact that many nations have little to no contribution to these emissions.

In order to analyse this data I will utilise the lessons from the UCD coursework which practised data analytics using various python tools. Python has a library of add-on features and libraries such as matplotlib and seaborn to aid with visualising this data.

This project will do the following:

1. Present and analyse the data sets relating to CO₂ emissions from air travel, both International and domestic.
2. Use data cleaning techniques in Python to ensure accurate and representative results.
3. Use data visualisation in Python to create various charts which will allow insights to be drawn from the data.
4. Present results and insights that can be concluded from the analysis.

Dataset

The primary datasets used in this project are ‘CO₂ emissions from domestic air travel, 2018’ and ‘CO₂ emissions from international air travel, 2018’. Both datasets were downloaded as CSV. files from ‘Our World in Data’, <https://ourworldindata.org>, who publish the “research and data to make progress against the world's largest problems”. Both datasets measure emissions per capita based on census population figures. As such several third world nations are omitted from these figures as no information is available.

Implementation Process

This project was implemented using the Jupyter Notebook IDE and python. Jupyter notebook is an open-source IDE that is a web-based interactive computational environment. The created Jupyter Notebook file was then exported to GitHub in a newly created repository for the purpose of this assignment.

The program is contained in one file “main.py”.

The project reads all local files used in this project. These datasets are then prepared, cleaned and modified. The called functions used in this project are as follows;

1. Pandas as pd
   1. used to efficiently manipulate the datasets
2. Matplotlib.pyplot as plt
   1. Graphically represent data
3. numpy as np
   1. For numerical calculations

The first step is reading my data in Jupyter Notebook and printing the first 10 rows to ensure the data had loaded correctly.

Below I will load the first dataset, a CSV file, saved localy, ‘*per\_capita\_co2\_domestic\_aviation.csv'*.

* *data\_domestic = pd.read\_csv('/Users/Sean/Desktop/per\_capita\_co2\_domestic\_aviation.csv')*
* *display(data\_domestic.head(10))*

This has worked as intended and I can see the correct dataset.

I’m going to load my second dataset and work on these two simultaneously as the same process will be required for both.

- *data\_international = pd.read\_csv('/Users/Sean/Desktop/per\_capita\_co2\_international\_aviation.csv')*

*-* *display(data\_international.head(10))*

I’m keeping a simple naming convention for my dataframes here – data\_domestic and data\_international

Now I want to make sure both datasets have the same number of rows, as they should have the same list of countries. I will also check the data types.

* *print(data\_international.shape)*
* *print(data\_international.dtypes)*
* *print(data\_domestic.shape)*
* *print(data\_domestic.dtypes)*

Both return (106, 4) which is as expected, so we can proceed.

My columns at this point are *Entity - object*

*Code - object*

*Year - int64*

*Per\_capita\_domestic/international\_aviation\_CO2 - float64*

Now I want to remove the column for ‘Year’ from both dataframes as the year for every entry is the same.

I will use the .loc function for this.

* *data\_international = pd.read\_csv('/Users/Sean/Desktop/per\_capita\_co2\_international\_aviation.csv')*
* *selected\_cols = ['Entity', 'Code', 'Per\_capita\_international\_aviation\_CO2']*
* *df\_updated\_international = data\_international.loc[:, selected\_cols]*
* *print(df\_updated\_international)*

Now the same process will be repeated for my other dataset.

* *data\_domestic = pd.read\_csv('/Users/Sean/Desktop/per\_capita\_co2\_domestic\_aviation.csv')*
* *selected\_cols = ['Entity', 'Code', 'Per\_capita\_domestic\_aviation\_CO2']*
* *df\_updated\_domestic = data\_domestic.loc[:, selected\_cols]*
* *print(df\_updated\_domestic)*

The remaining columns are all required. The ‘code’ column may be useful when creating visualisations.

Now I want to try sorted one of my dataframes and since both are made up of the same list of countries.

If we sort in descending order by emissions we can get an idea of what our visualisations should look like for each dataframe.

* *df\_updated\_domestic.sort\_values(by=['Per\_capita\_domestic\_aviation\_CO2'], inplace=True, ascending=False)*
* *print(df\_updated\_domestic)*
* Entity Code Per\_capita\_domestic\_aviation\_CO2
* 101 United\_States USA 385.520397
* 4 Australia AUS 267.167242
* 69 Norway NOR 209.230182
* 67 New\_Zealand NZL 174.188537
* 16 Canada CAN 168.267380

Now we can see the 5 largest values for the domestic dataset.

Before I create my visualisations I want to merge both dataframes into one. This will be straightforward since they share common values (Entity and Code).

I will call this dataframe df\_merged.

* *df\_merged = pd.merge(df\_updated\_domestic, df\_updated\_international, on = "Entity", how = "outer")*
* *display(df\_merged.head(10))*

Now both columns are merged but I have an unnecessary column. Code\_x and Code\_y contain the same vales – three letter Country code.

I will remove ‘Code\_Y’.

* *selected\_cols = ['Entity', 'Code\_x', 'Per\_capita\_domestic\_aviation\_CO2', 'Per\_capita\_international\_aviation\_CO2']*
* *df\_merged = df\_merged.loc[:, selected\_cols]*
* *display(df\_merged.head(10))*

Now my data is ready for visualisations

To get a quick visualisation of a section of the dataframe I will create lists and use plt.plot

This graph will show the 10 highest values for domestic emmisions.

- *Country = [ 'USA', 'AUS', 'NOR', 'NZL', 'CAN', 'JPN', 'ISL', 'CHL', 'FRA', 'SAU' ]*

*- Domestic\_c02 = [ 385.520397, 267.167242, 209.230182, 174.188537, 168.267380, 73.958502, 71.323861, 70.458354, 69.666608, 65.243270 ]*

*- plt.plot(Country, Domestic\_c02)*

*- plt.title('Domestic C02 by nation')*

*-plt.xlabel('Country')*

*- plt.ylabel('Domestic c02 in KG')*

*- plt.show()*

Now I will try a similiar visualisation for the same nations international.

First I will need to sort by descending order for this column.

*- df\_merged.sort\_values(by=['Per\_capita\_international\_aviation\_CO2'], inplace=True, ascending=False)*

*- display(df\_merged.head(10))*

*- Country = [ 'ISL', 'QAT', 'ARE', 'SGP', 'MLT', 'NZL', 'MUS', 'IRL', 'CHE', 'AUS' ]*

*- International\_c02 = [ 3505.599607, 2472.695468, 2195.111739, 1741.036404, 991.639201, 640.278593, 599.817907, 574.126181, 513.284541, 495.948006 ]*

*- plt.plot(Country, International\_c02)*

*- plt.title('International C02 by nation')*

*- plt.xlabel('Country')*

*- plt.ylabel('c02 in KG')*

*- plt.show()*

Now I will look at the countries with the highest domestic emissions and compare their international emissions

import numpy as np

import matplotlib.pyplot as plt

X = ['USA','AUS','NOR','NZL','CAN']

domestic = [385.520397,267.167242,209.230182,174.188537, 168.26738]

international = [170.601004,495.948006,298.030303,640.278593,294.512443]

X\_axis = np.arange(len(X))

plt.bar(X\_axis - 0.2, domestic, 0.4, label = 'Domestic')

plt.bar(X\_axis + 0.2, international, 0.4, label = 'International')

plt.xticks(X\_axis, X)

plt.xlabel("Country")

plt.ylabel("Emissions in KG")

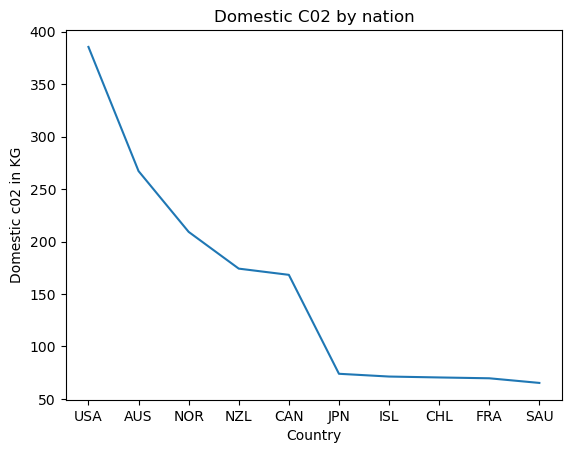
plt.title("Domestic Vs International Emissions")

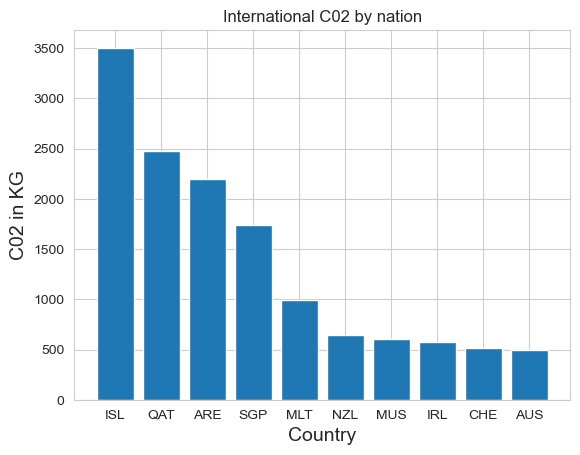
plt.legend()

plt.show()

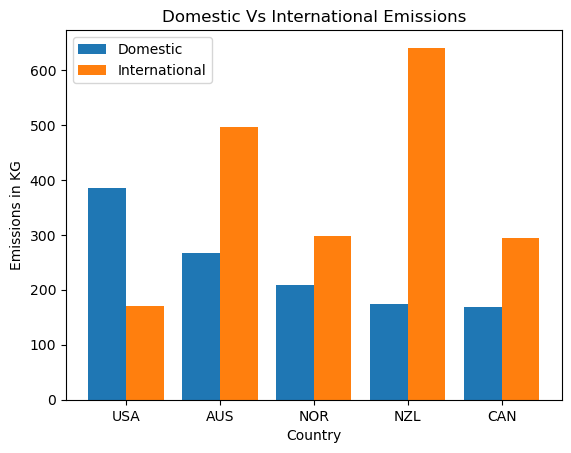
Results

Graph 1

Graph 2



Graph 3



Insights

*Graph 1*

* If we examine Graph 1 we can make some observations regarding domestic emissions per capita. Firstly, we would assume that wealthier nations would be likely to have higher emissions because people can generally afford to fly. This is supported by the data as we can the 5 largest values are wealthy first world countries ; USA, Australia, Norway, New Zealand and Canada.
* Secondly, we would assume that larger countries may have more internal flights – and this is supported by the data in Graph 1 as we see can a [correlation](https://ourworldindata.org/grapher/per-capita-co2-domestic-aviation-vs-land-area) between land area and domestic flight emissions; in small countries people are more likely to travel by car or other means. However there are some outliers here, such as New Zealand. This is likely due to New Zealand’s relatively cheap domestic flights and its geographical isolation leading to more domestic travel.

*Graph 2*

* Some of the largest emitters for international emissions per person in 2018 were Iceland, Qatar, United Arab Emirates, Singapore and Malta. Once again we see an overrepresentation of wealthy countries here. As we can also see, four of the five largest values are from geographically small nations, which would likely result in increased international travel, the inverse of what we noticed in our last insight.
* It would also seem that geographical isolation contributes to these international emission figures. Supporting this is the Icelandic figure of 3’505 Kilograms, which is the largest figure per capita by a significant margin. Further support of this is the fact that 6 of the 10 highest emitters per capita are island nations.

*Graph 3*

* As we can see in Graph 3 there is no clear corelation between the highest domestic emitters and their international emissions. One notable detail here is the USA, which is one of the only examples of a higher domestic emission per capita than international. This suggests US citizens often holiday and travel to other parts of the US.

References

# Our World in Data, *Climate change and flying: what share of global CO2 emissions come from aviation?*

<https://ourworldindata.org/co2-emissions-from-aviation>