

**PROJECT REPORT**

**07 May 2023**

**Gary O'Donoghue**

**FOR AVIATION**

A picture containing clipart

Description automatically generated

**CERTIFICATE**

**ANALYTICS**

**IN DATA**

Contents

[**Abstract** 3](#_Toc134543840)

[**Introduction** 3](#_Toc134543841)

[**Datasets** 4](#_Toc134543842)

[**Implementation Process** 5](#_Toc134543843)

[**Results** 13](#_Toc134543844)

[**Insights** 17](#_Toc134543845)

[**References** 17](#_Toc134543846)

GitHub URL - https://github.com/garyod/UCDPA\_Gary\_ODonoghue

# **Abstract**

The project aims to explore the relationship between population growth and air travel passenger numbers over the past 40 years using Python for data analysis. The project will begin by collecting and cleaning data on population growth and air travel passenger numbers for the entire world.

The data will be obtained from publicly available sources such as the United Nations (UN) and International Energy Agency (IEA).

After cleaning the data, statistical techniques will be applied to explore the relationship between population growth and air travel passenger numbers. This will involve using tools such as correlation analysis, linear regression, and data visualization libraries such as Matplotlib or Seaborn to analyse and present the data.

The project will also consider various factors that may affect the relationship between population growth and air travel passenger numbers, such as economic growth, technological advancements, and changes in travel behaviour. By examining these factors, the project will seek to provide a more nuanced understanding of the relationship between population growth and air travel passenger numbers over the past 40 years.

# **Introduction**

Over the past few decades, air travel has become increasingly accessible and affordable, leading to a surge in the number of people traveling by air. At the same time, the world's population has continued to grow, reaching 7.9 billion in 2021, according to the United Nations. As these two trends have unfolded, questions have arisen about the relationship between population growth and air travel passenger numbers.

Understanding the relationship between population growth and air travel is important for a variety of reasons. For policymakers, it is essential to understand how population growth and air travel are linked in order to plan for infrastructure development, manage resources, and promote sustainable travel. For the aviation industry, understanding the relationship between population growth and air travel is crucial for forecasting demand and planning routes.

This project aims to explore the relationship between population growth and air travel passenger numbers over the past 40 years using Python for data analysis. By examining this relationship, we hope to provide insights into the factors that influence air travel demand, and identify patterns and trends in air travel behaviour.

# **Datasets**

I have used two datasets in this project which I intend to analyse

1. World Population 1950 – 2100
2. Passenger Data

**World Population 1950 – 2100;**

This dataset contains world population and predicted world population numbers ranging between the years 1950 and 2100. The dataset was chosen due to it being a highly recognised intergovernmental organization that operates on a global scale.

Source - https://population.un.org/wpp/Download/Standard/CSV/

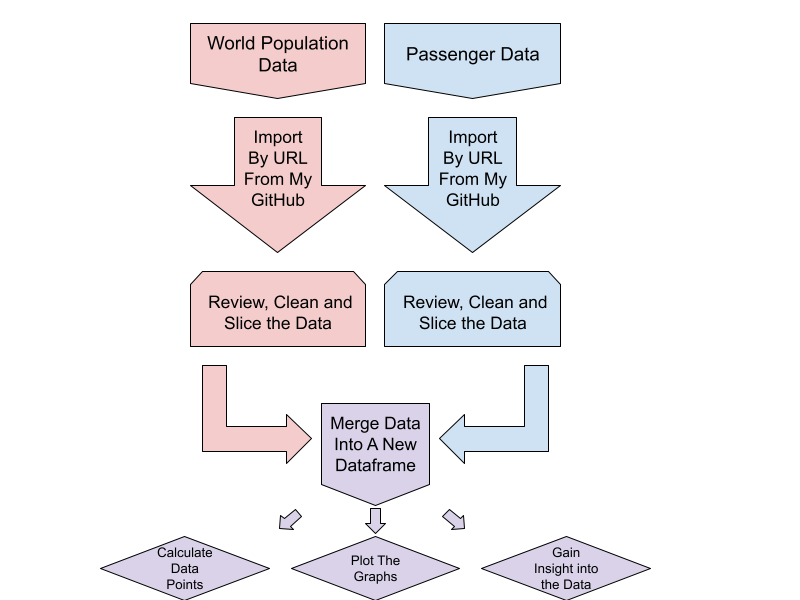
**Passenger Data;**

This dataset contains all aircraft passenger numbers ranging between the years 1980 and 2020. The dataset was chosen due to it being a recognised Aviation body with data openly available to the public.

Source - <https://www.iea.org/data-and-statistics/charts/world-air-passenger-traffic-evolution-1980-2020>

# **Implementation Process**

**Project Flow Chart;**

****

**Process;**

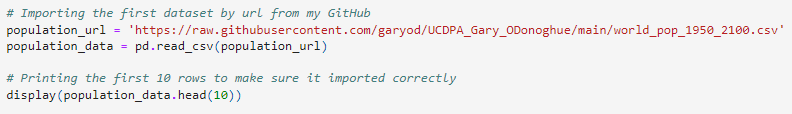
The project was implemented using Jupyter Notebook IDE and Python. I chose Jupyter Notebook because of how clean and easy it is to read and how blocks of code can be written like chapters in a book.  
On completion, the project was uploaded to my GitHub repo, for this course, in a file named main\_py.ipynb which is a Jupyter Notebook format.

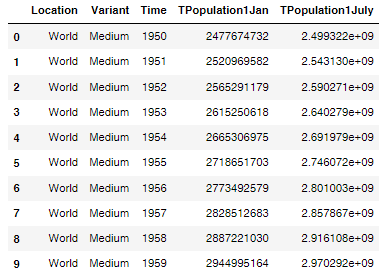
The following packages were used in the project;

1. Pandas as pd
   1. used to modify/manipulate the datasets
2. numpy as np
   1. For all calculations used in the datasets
3. Matplotlib.pyplot as plt
   1. Graphically represent data on charts

First step was to import the first Dataset. I did this through URL from my GitHub and imported into a pandas dataframe.

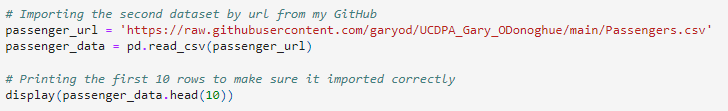
Then I read the first 10 rows to make sure it imported correctly;

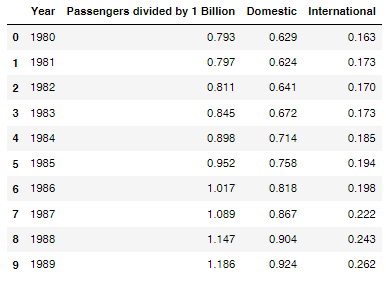




Next step is Importing the second dataset by URL from my GitHub.

Then I read the first 10 rows to make sure it imported correctly;





I choose to do it by URL again because I was having issues importing it locally due to my username on my machine. My username is Gary O’Donoghue and due to the fact it has an apostrophe (special character) it will not load from anywhere inside my profile, I thought it would be bad practice to store the file on the root of the C Drive so I opted to import both from GitHub. This can be advantageous though as the code can simply be downloaded and ran as is without the need of changing directories in the code.

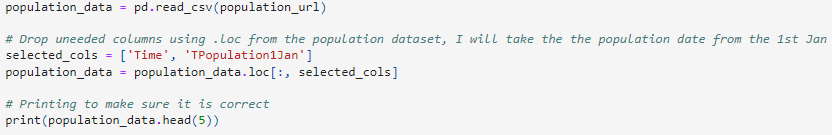
Now that both Datasets are imported fine and everything appears to be okay, I will now clean both datasets.

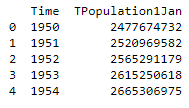
For the first Dataset, World Population Data, it appears as though there are a few columns which will be unnecessary for what I intend on doing.

On looking at the Dataset the following columns are all that are required; ‘Time’ and ‘TPopulation1Jan’.

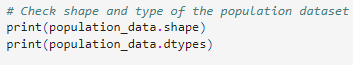
I will keep these columns using the .loc function.

And print the first 5 rows to confirm;

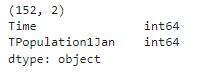




Now I will check the shape and type of the dataset

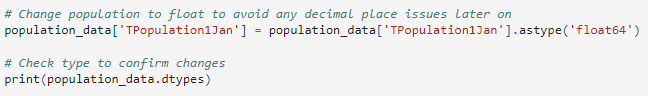


Which Returns;



I plan to divide the population column by a billion later to make any visuals a lot more legible. Prior to doing so I will change the column type to float64 to avoid any issues we may encounter later with decimal points (I encountered an issue whereby I lost the decimal places and all my values were rounded to the nearest whole number which caused inaccurate visuals on the graph)

I will also print the type of the dataset again to confirm everything changed.

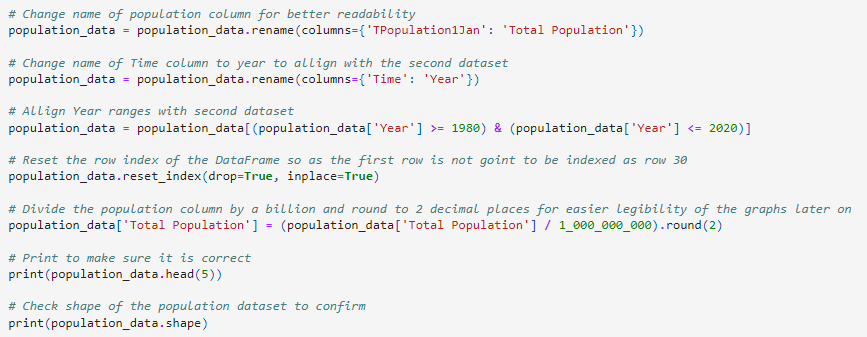


Which Returns;

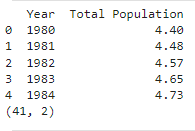


Now I will clean the dataset;

* change the column names for better legibility
* filter only for the years 1980 – 2020,
* reset the index as dropping the first 30 rows will cause the index to start from 30
* divide the population column by 1 Billion for better legibility in the graphs later on and also round it to 2 decimal places.
* print the first 5 rows to make sure the column names changed, the index changed, it starts at year 1980, the population is now divided by 1 Billion and rounded to 2 decimal places.
* print shape to confirm we have the correct number of rows upon dropping several rows



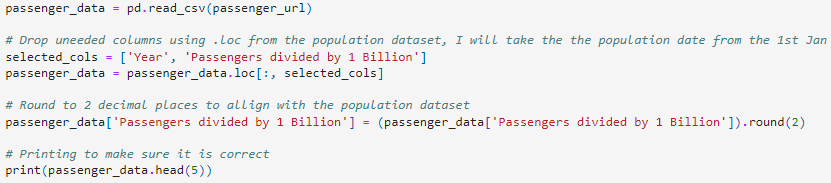
Which Returns;



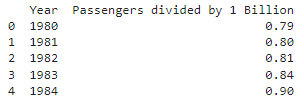
That is the first dataset cleaned, sliced and ready for later on, now I will start cleaning and slicing the second dataset.

Load the dataset and drop unnecessary columns using the .loc function

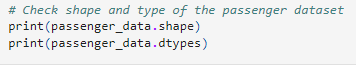
Also round the “Passengers divided by 1 Billion” column and also printing to make sure it is correct



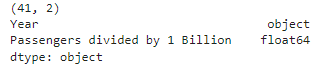
Which Returns;



Now I will check the shape and type of the dataset and see if what I have done aligned both datasets.



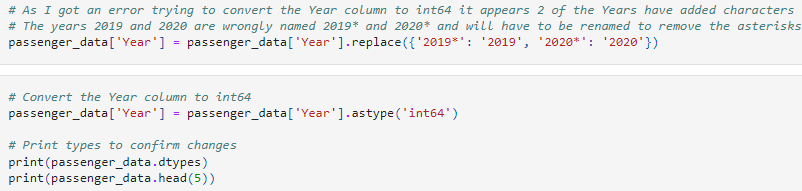
Which Returns;



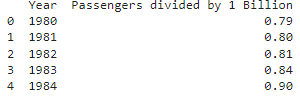
I will have to convert the ‘Year’ column to align and merge with the first dataset and print to confirm changes

Further in the project while I was trying to merge both datasets I ran into an issue whereby a couple of the ‘Years’ in the passenger\_data datasetwere oincorrectly labelled, both 2019 & 2020 had asterisks at the end of their names which caused errors on merging.

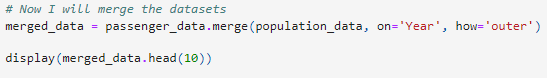
A simple fix was to rename the ‘Years’ correctly using .replace



Which Returns;



Now I will Merge both datasets and call the resulting dataset ‘merged\_data’ and display the result



Which Returns;



All appears *correct.*

# **Results**

Now it is time to plot the data.

First I started by creating a figure and axis

Then set the x axis to be the ‘Year’ column

Then set the y axis to be in the range of 1 – 10

Then plot the Passengers divided by 1 Billion column against the Year column

Now plot the data of the ‘Total Population’ column against the ‘Year’ column.

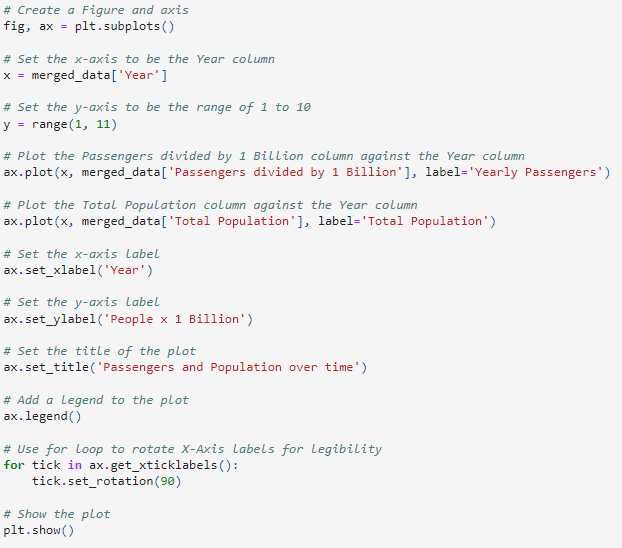
Setting the x-axis & y-axis labels

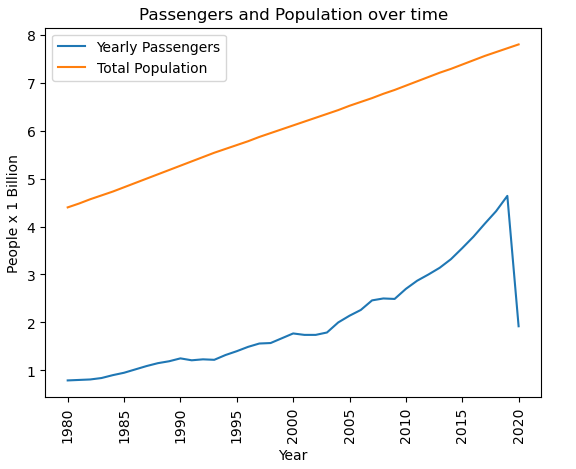
Giving it a name and setting the Title of the plot

Add a legend to differentiate between population and passengers

For better legibility of the ‘Year’ numbers on the x axis I will use a for loop to rotate 90 degrees

Lastly was to plot the graph time to plot the graph





I will also create a bar graph with the same data to see if any valuable insight can be taken away from that also.

First I started by creating a figure and axis and then group the data by 5 Year intervals and calculate the mean over those 5 years

Then set the x-axis to be the year intervals, offset by 1 year for aesthetics of the bars

Then set the width of the bars (2.00 seemed to be the best fit)

Now I will offset the position of the bars so as they are not on top of each but rather beside each other. This will greatly improve the readability of the graph.

Create the bars for the 'Passengers divided by 1 Billion' & 'Total Population' columns

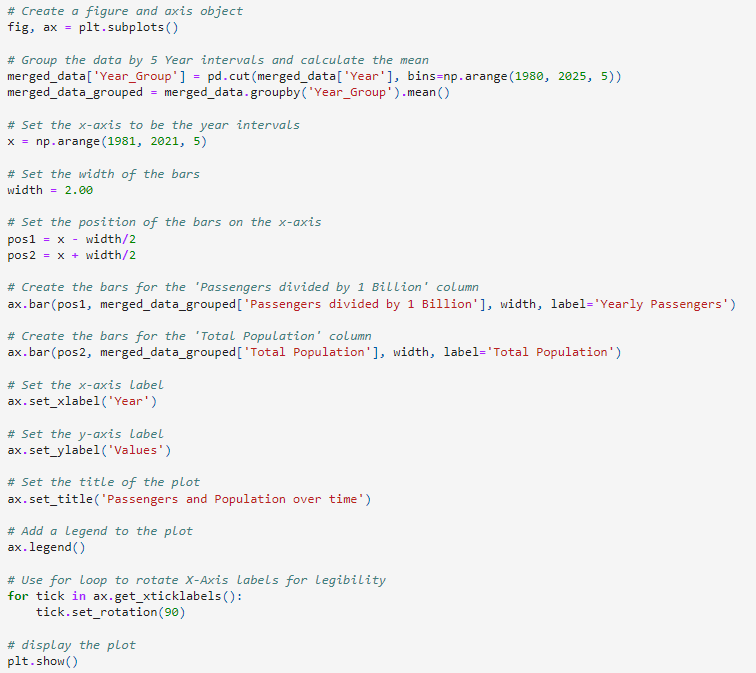
Set the x-axis & y-axis labels

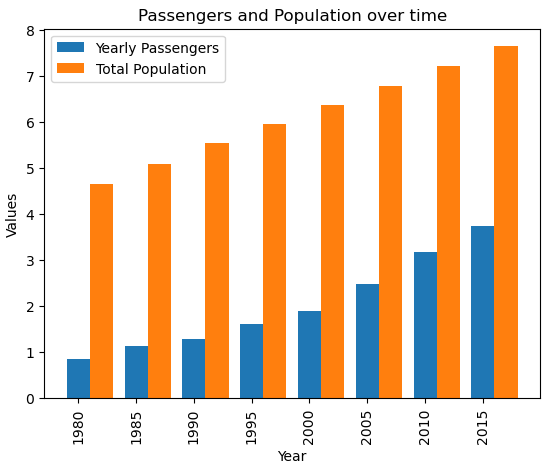
Set the title of the plot

Add a legend to the plot

For better legibility of the ‘Year’ numbers on the x axis I will use a for loop to rotate 90 degrees

Displaying the plot





# **Insights**

1. As can be expected both World Population and average yearly aircraft passengers have increased over the past 40 years. The significant rise in population has led to an increase in the demand for air travel. As this rise in world population continues, more people are likely to travel by air to stay connected.
2. As we can see on both graphs, the growth rate of passengers has significantly outpaced the growth rate of world population, suggesting air travel is becoming more accessible and more affordable for people all over the world. This is reflected in the bar chart whereby we can see passenger numbers were over 50% of the world’s population by 2020.
3. What we can also see from this data is the impact of major global events on air travel and population. For example, the events of 9/11 led to a significant drop in air travel for a period of time and how global events seem to have zero impact on population growth.
4. As the world population continues to grow, it is likely that demand for air travel will also continue to grow, and it will be important to balance this demand with the need to reduce the environmental impact of air travel.
5. The relationship between world population and aircraft passenger numbers is not strictly linear, as other factors such as economic growth, political stability, and airline industry competition also play a role as evident from both charts suggesting population growth is not the biggest factor in airline passenger growth

# **References**

**World Population 1950 – 2100;**

This dataset contains world population and predicted world population numbers ranging between the years 1950 and 2100. The dataset was chosen due to it being a highly recognised intergovernmental organization that operates on a global scale.

Source - https://population.un.org/wpp/Download/Standard/CSV/

**Passenger Data;**

This dataset contains all aircraft passenger numbers ranging between the years 1980 and 2020. The dataset was chosen due to it being a recognised Aviation body with data openly available to the public.

Source - <https://www.iea.org/data-and-statistics/charts/world-air-passenger-traffic-evolution-1980-2020>