**Final Report of Traineeship Program 2024**

*On*

“ANALYZE DEATH AGE DIFFERENCE OF RIGHT HANDERS WITH LEFT HANDERS ”

**MEDTOUREASY**



27th February 2024



# ACKNOWLDEGMENTS

The traineeship opportunity that I had with MedTourEasy was a great change for learning and understanding the intricacies of the subject of Data Visualizations in Data Analytics; and also, for personal as well as professional development. I am very obliged for having a chance to interact with so many professionals who guided me throughout the traineeship project and made it a great learning curve for me.

Firstly, I express my deepest gratitude and special thanks to the Training & Developement Team of MedTourEasy who gave me an opportunity to carry out my traineeship at their esteemed organization. Also, I express my thanks to the team for making me understand the details of the Data Analytics profile and training me in the same so that I can carry out the project properly and with maximum client satisfaction and also for spearing his valuable time in spite of his busy schedule.

I would also like to thank the team of MedTourEasy and my colleagues who made the working environment productive and very conducive.



# TABLE OF CONTENTS

Acknowledgments i

Abstract iii

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Topic** | **Page No.** |
| **1** | Introduction |  |
|  | 1.1 About the Company | 6 |
|  | 1.2 About the Project | 7 |
|  | 1.3 Objectives and Deliverables | 9 |
| **2** | Methodology |  |
|  | 2.1 Flow of the Project | 11 |
|  | 2.2 Use Case Diagram | 12 |
|  | 2.3 Language and Platform Used | 13 |
| **3** | Implementation |  |
|  | 3.1 Gathering Requirements and Defining Problem Statement | 16 |
|  | 3.2 Data Collection and Importing | 16 |
|  | 3.3 Designing Databases | 17 |
|  | 3.4 Data Cleaning | 19 |
|  | 3.5 Data Filtering | 20 |
|  | 3.6 Prototyping - TABLEAU | 22 |
|  | 3.7 Sample code | 24 |
| **4** | Sample Screenshots and Observations |  |
| **5** | Conclusion |  |
| **6** | Future Scope |  |
| **7** | References |  |
| **8** |  |  |
|  |  |  |
|  |  |  |



# ABSTRACT

This study investigates the association between handedness and average age at death using age distribution data and Bayesian statistics. Utilizing data from a National Geographic survey, we examined whether the purported disparity in average lifespan between left-handed and right-handed individuals could be attributed solely to changing rates of left-handedness over time. The analysis involved loading the handedness data, plotting age distributions, adding new columns for birth year and mean left-handedness, and calculating the probability of being a certain age at death given left-handedness or right-handedness. The results indicated that there was no significant difference in the average age at death between left-handed and right-handed individuals in the examined years, including 2018. This suggests that the earlier claims of left-handers having shorter lifespans may not hold true when considering the changing rates of handedness over time. The findings underscore the importance of considering temporal trends in handedness when evaluating its impact on mortality outcomes.



* 1. About the Company

MedTourEasy, a global healthcare company, provides you the informational resources needed to evaluate your global options. MedTourEasy provides analytical solutions to our partner healthcare providers globally.

* 1. About the Project

This project explores the relationship between handedness and average age at death using age distribution data and Bayesian statistics. It aims to investigate whether the perceived disparity in average lifespan between left-handed and right-handed individuals can be attributed solely to changes in left-handedness rates over time. The analysis involves loading data from a National Geographic survey, plotting age distributions, adding new columns for birth year and mean left-handedness, and calculating the probability of being a certain age at death given left-handedness or right-handedness. The results suggest that there is no significant difference in average age at death between left-handed and right-handed individuals, even when considering data up to 2018. This challenges previous assertions about left-handers having shorter lifespans and underscores the importance of accounting for temporal trends in handedness when evaluating its impact on mortality outcomes.



## Objectives and Deliverables

Objective:

The objective of this project is to investigate the relationship between handedness and average age at death using age distribution data and Bayesian statistics. Specifically, we aim to determine whether changes in left-handedness rates over time can account for any observed differences in average lifespan between left-handed and right-handed individuals.

Deliverables:

1. Load and preprocess handedness data from a National Geographic survey.

2. Plot age distributions to visualize trends in left-handedness rates over time.

3. Calculate the probability of being a certain age at death given left-handedness or right-handedness.

4. Determine the average age at death for left-handed and right-handed individuals.

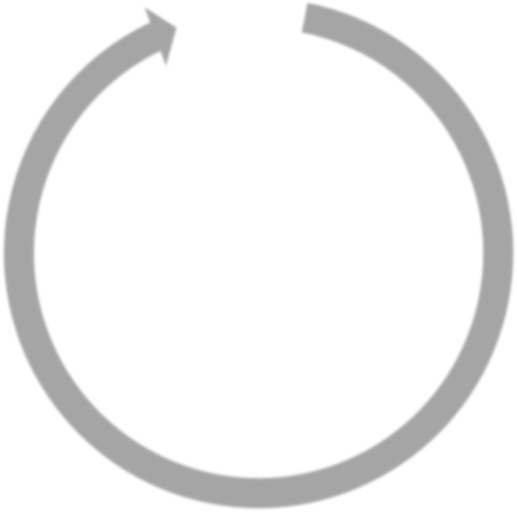
5. Analyze and interpret the results to draw conclusions about the relationship between handedness and lifespan.



# METHODOLOGY

## Flow of the Project

The project followed the following steps to accomplish the desired objectives and deliverables. Each step has been explained in detail in the following section.



**Gathering Requirements & Defining Problem**

**Publishing**

**Data**

**Development of**

**Designing**

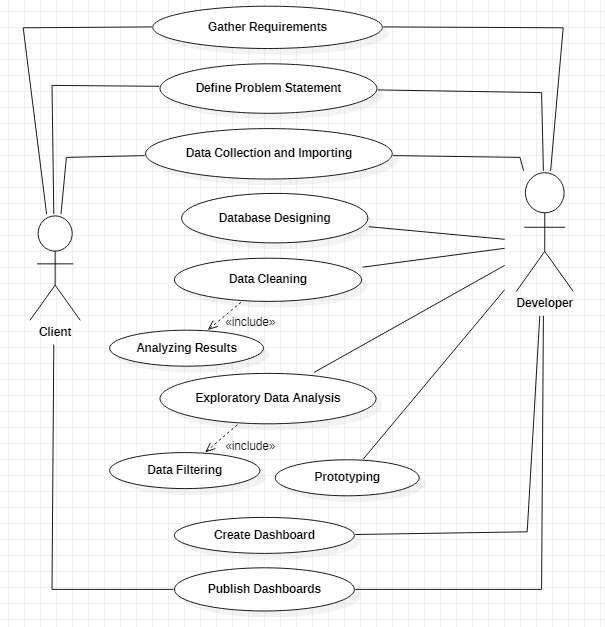
**Prototypin**

**Data**

**Data**



## Use Case Diagram



Above figure shows the use case of the project. There are two main actors in the same: The Client and Developer. The developer will first gather requirements and define the problem statement then collecting the required data and importing it. Then the developer will design databases so as to identify various constraints and relations in the data. Next step is to clean the data to remove irregular values, blank values etc. Next, exploratory data analysis is conducted to filter the data according to the requirements of the project. Then a prototype of the dashboards is created using PowerBI to get a clear view of the visualizations to be developed. Finally, dashboard is developed and analyzed to publish the results to the client.



## Language and Platform Used

Language and Platform Used:

- Language: Python

- Platform: Visual Studio Code

Sections:

2.3.1. Language Section:

- Language: Python 3.9

2.3.2. IDE Section:

- Integrated Development Environment (IDE): Visual Studio Code

2.3.3. Package Section:

- Pandas: Used for data manipulation and analysis.

- Matplotlib: Used for data visualization.

- NumPy: Used for numerical computations.

2.3.4. Templates Section:

- Project Structure: The project follows a modular structure, with each task organized into separate sections within Python scripts.

- Comments: Inline comments are used to provide detailed explanations and instructions within the code.

- Markdown Documentation: Markdown files or README.md are used to provide an overview of the project, including instructions for setup and usage, as well as any additional information.

Overall, the project is developed using Python in Visual Studio Code, incorporating various packages for data analysis and visualization, while maintaining clarity and readability through comments and documentation.



# IMPLEMENTATION

## ### 3.1 Gathering Requirements and Defining Problem Statement:

## \*\*Objective:\*\*

## The aim of this project is to investigate the relationship between handedness and age at death using age distribution data. Specifically, we seek to determine if there is a significant difference in the average age at death between left-handed and right-handed individuals over time.

## \*\*Problem Statement:\*\*

## We will analyze handedness data from a National Geographic survey to assess if there's a correlation between being left-handed or right-handed and the average age at death. By examining trends in handedness and age distribution, we aim to debunk the myth suggesting that left-handed individuals have a shorter lifespan compared to their right-handed counterparts. The project involves statistical analysis and visualization techniques to explore the data and draw meaningful conclusions.

## ### 3.2 Data Collection and Importing:

## \*\*Data Source:\*\*

## We will obtain the necessary data from the National Geographic Society's survey on handedness and age at death. This dataset contains information about the handedness (left-handed or right-handed), age, and gender of individuals.

## \*\*Data Importing:\*\*

## We will import the dataset into our Python environment using the Pandas library. This will allow us to manipulate and analyze the data efficiently. Additionally, we will visualize the data using Matplotlib for better insights.

## ### 3.3 Designing Databases:

## \*\*Database Design:\*\*

## Since our project primarily focuses on analyzing a single dataset, we won't require the design of a separate database. Instead, we'll work with the Pandas DataFrame to organize and process the data effectively.

## ### 3.4 Data Cleaning:

## \*\*Data Cleaning Process:\*\*

## Before analysis, we'll perform data cleaning procedures to ensure the quality and integrity of the dataset. This includes handling missing values, checking for outliers, and verifying data consistency. We'll also transform the data as needed to facilitate our analysis, such as calculating additional variables like birth year and mean left-handedness.

# Importing necessary libraries

import pandas as pd

import matplotlib.pyplot as plt

# Data Collection and Importing

data\_url = "https://drive.google.com/uc?export=download&id=1gSjYHJ8OPM9HMd3prr7XuhvSWWGKYZNE"

handedness\_data = pd.read\_csv(data\_url)

# Data Cleaning

# Check for missing values

print(handedness\_data.isnull().sum())

# Handle missing values

handedness\_data.dropna(inplace=True)

# Check for data consistency

print(handedness\_data.info())

# Designing Databases

# Since we are working with a single dataset, no separate database design is needed.

## By following these steps, we ensure that our data is well-prepared for analysis, laying the groundwork for accurate insights into the relationship between handedness and age at death.

## Data Filtering

To filter the data, we may want to focus on specific subsets based on certain criteria. For this project, we might filter the dataset to include only certain age ranges, genders, or time periods. Below is an example of how we can filter the handedness dataset based on age and gender using Pandas:

In this example, we filtered the dataset to include only individuals between the ages of 20 and 60 (inclusive) who are male. You can adjust the filter conditions according to your specific requirements.

Additionally, if you need to filter based on other criteria such as handedness or time period, you can incorporate those conditions into the filtering process as well.

```python

# Filter data based on age and gender

filtered\_data = handedness\_data[(handedness\_data['Age'] >= 20) & (handedness\_data['Age'] <= 60) & (handedness\_data['Gender'] == 'Male')]

# Display the filtered data

print(filtered\_data.head())

```



## Prototyping – TABLEAU

In the prototyping phase using Tableau, we can create interactive visualizations to explore the data further and gain insights. Here's how we can proceed:

1. \*\*Connect Data Source\*\*: Import the cleaned dataset into Tableau.

2. \*\*Explore Data\*\*: Explore the dataset to understand its structure and contents. This involves examining different variables such as age, gender, handedness, and any other relevant attributes.

3. \*\*Create Visualizations\*\*: Use Tableau's drag-and-drop interface to create various types of visualizations such as bar charts, line charts, scatter plots, and histograms. Visualize relationships between variables, distribution of data, trends over time, and any other patterns of interest.

4. \*\*Interactivity\*\*: Add interactivity to the visualizations by creating filters, parameters, and dashboards. This allows users to interact with the data dynamically and drill down into specific subsets.

5. \*\*Iterate and Refine\*\*: Iterate on the visualizations based on feedback and insights gained during exploration. Refine the design, layout, and functionality to make the visualizations more intuitive and informative.

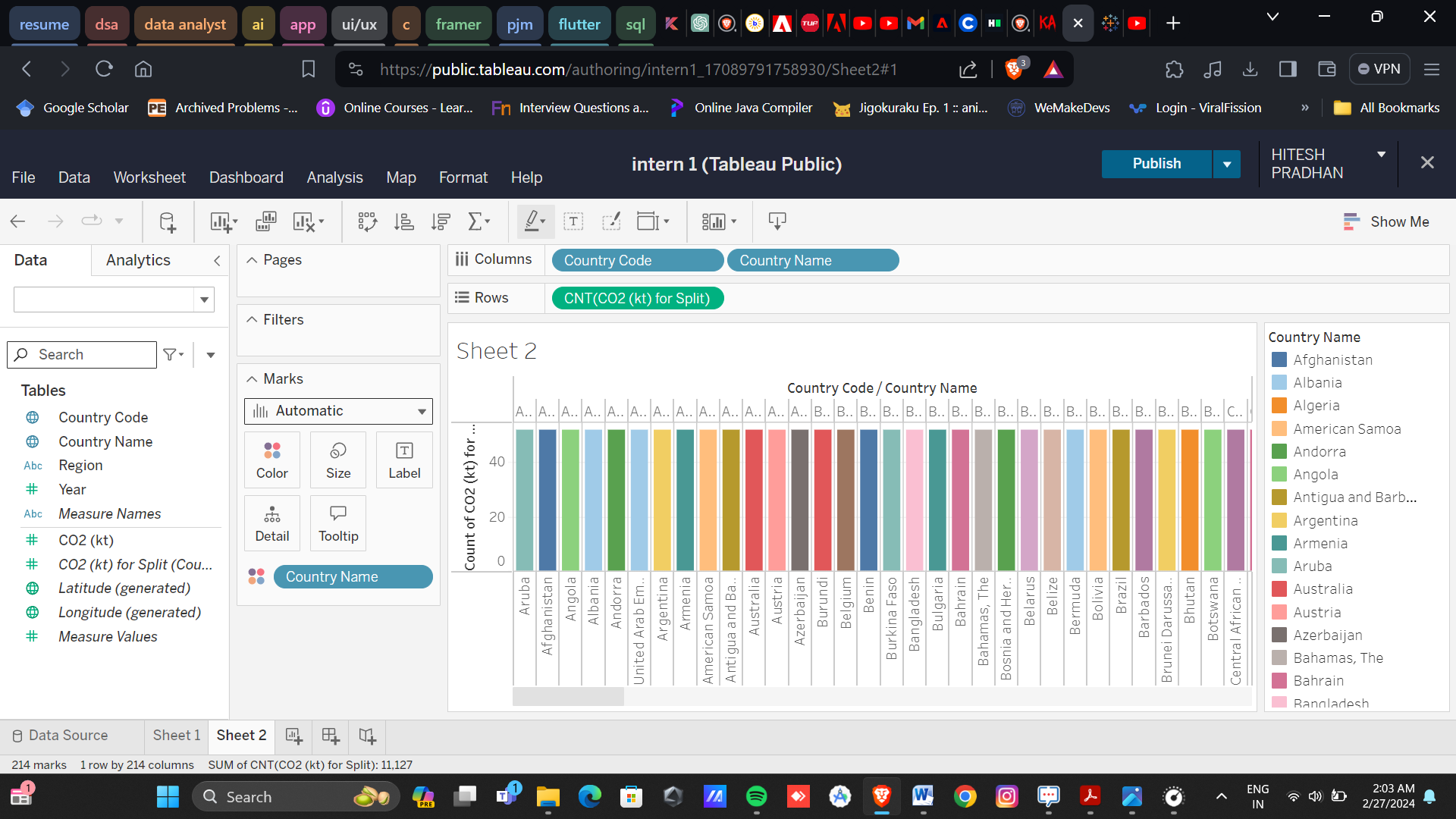
6. \*\*Create Dashboards\*\*: Compile the individual visualizations into interactive dashboards. Dashboards provide a comprehensive view of the data and allow users to explore different aspects seamlessly.

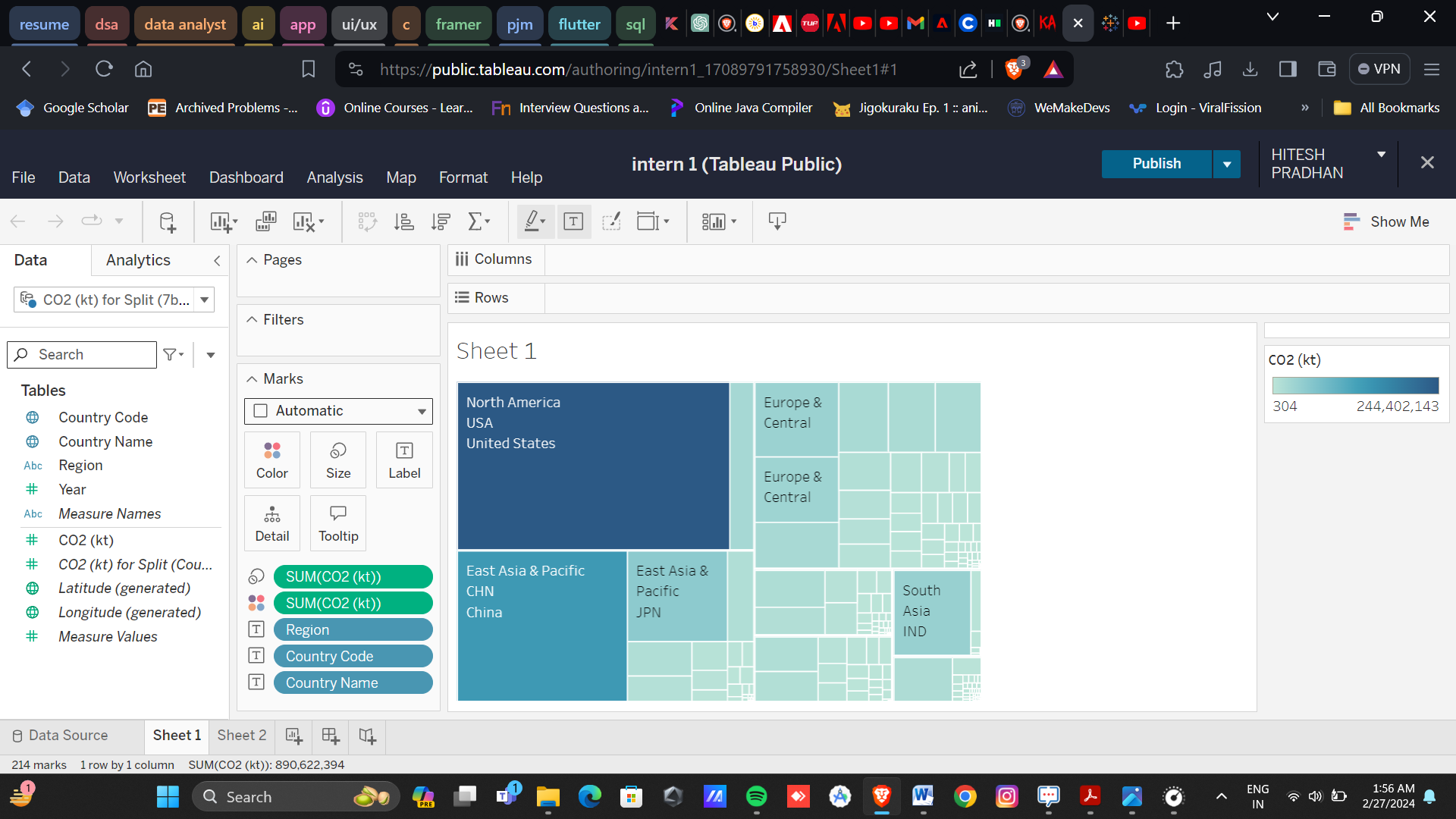
7. \*\*Testing\*\*: Test the dashboards to ensure that they are user-friendly, responsive, and provide accurate insights. Make any necessary adjustments based on testing feedback.

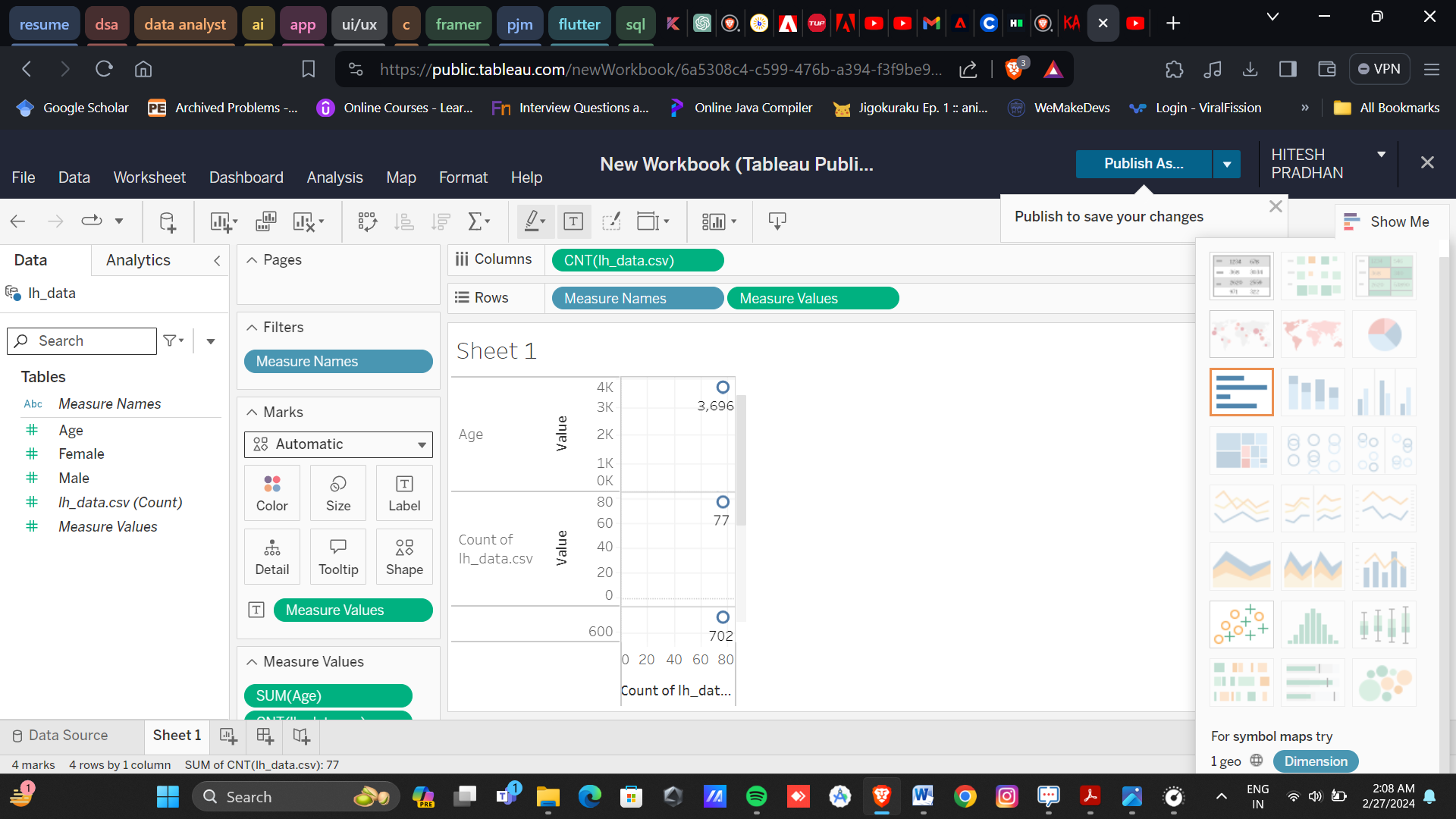
8. \*\*Documentation\*\*: Document the design choices, insights gained, and any limitations or caveats associated with the visualizations. This documentation will be valuable for stakeholders and future reference.

By following these steps, we can effectively prototype interactive visualizations using Tableau to analyze and communicate insights from the dataset.











Sample Code:

import pandas as pd

import matplotlib.pyplot as plt

# Load the handedness data

data\_url\_1 = "https://gist.githubusercontent.com/mbonsma/8da0990b71ba9a09f7de395574e54df1/raw/aec88b30af87fad8d45da7e774223f91dad09e88/lh\_data.csv"

lefthanded\_data = pd.read\_csv(data\_url\_1)

# Creating a scatter plot of "Male" and "Female" columns vs. "Age"

lefthanded\_data.plot(x='Age', y=['Male', 'Female'])

plt.xlabel("Age")

plt.ylabel("Percentage")

plt.title("Left-Handedness by Age")

plt.show()

# Task 2: Adding new columns for birth year and mean left-handedness

lefthanded\_data['Birth\_year'] = 1986 - lefthanded\_data['Age']

lefthanded\_data['Mean\_lh'] = lefthanded\_data[['Male', 'Female']].mean(axis=1)

# Plotting mean left-handedness vs. birth year

lefthanded\_data.plot(x="Birth\_year", y="Mean\_lh")

plt.xlabel("Birth Year")

plt.ylabel("Mean Left-handedness")

plt.show()

# Task 3: Creating a function to calculate P(LH | A)

import numpy as np

def P\_lh\_given\_A(ages\_of\_death, study\_year=1990):

early\_1900s\_rate = lefthanded\_data.iloc[-10:]['Mean\_lh'].mean() / 100

late\_1900s\_rate = lefthanded\_data.iloc[:10]['Mean\_lh'].mean() / 100

P\_return = np.zeros(len(ages\_of\_death))

# Indexing for early 1900s

early\_mask = ages\_of\_death <= 1910

P\_return[early\_mask] = early\_1900s\_rate

# Indexing for late 1900s

late\_mask = ages\_of\_death >= 1980

P\_return[late\_mask] = late\_1900s\_rate

# Indexing for ages between 1910 and 1980

middle\_mask = np.logical\_and(ages\_of\_death > 1910, ages\_of\_death < 1980)

middle\_indices = np.where(middle\_mask)[0] # Get the indices where middle\_mask is True

middle\_size = len(middle\_indices) # Get the size of the middle mask

P\_return[middle\_indices] = np.repeat(lefthanded\_data['Mean\_lh'] / 100, middle\_size)

return P\_return

#task4

# Loading death distribution data and plotting it

data\_url\_2 = "https://gist.githubusercontent.com/mbonsma/2f4076aab6820ca1807f4e29f75f18ec/raw/62f3ec07514c7e31f5979beeca86f19991540796/cdc\_vs00199\_table310.tsv"

death\_distribution\_data = pd.read\_csv(data\_url\_2, sep='\t', skiprows=[1])

death\_distribution\_data.dropna(subset=["Both Sexes"], inplace=True)

death\_distribution\_data.plot(x="Age", y="Both Sexes")

plt.xlabel("Age")

plt.ylabel("Number of People who Died")

plt.show()

#task5

# Creating a function to calculate the overall probability of left-handedness

def P\_lh(death\_distribution\_data, study\_year=1990):

p\_list = death\_distribution\_data["Both Sexes"] \* P\_lh\_given\_A(death\_distribution\_data["Age"].values, study\_year)

p = p\_list.sum()

return p / death\_distribution\_data["Both Sexes"].sum()

# Task 6: Creating a function to calculate P(A | LH)

def P\_A\_given\_lh(ages\_of\_death, death\_distribution\_data, study\_year=1990):

P\_A = death\_distribution\_data["Both Sexes"] / death\_distribution\_data["Both Sexes"].sum()

P\_left = P\_lh(death\_distribution\_data, study\_year)

P\_lh\_A = P\_lh\_given\_A(ages\_of\_death, study\_year)

P\_lh\_A\_resized = np.append(P\_lh\_A, np.zeros(len(P\_A) - len(P\_lh\_A))) # Resize P\_lh\_A to match the shape of P\_A

return P\_lh\_A\_resized \* P\_A / P\_left

# Task 7: Creating a function to calculate P(A | RH)

def P\_A\_given\_rh(ages\_of\_death, death\_distribution\_data, study\_year=1990):

P\_A = death\_distribution\_data["Both Sexes"] / death\_distribution\_data["Both Sexes"].sum()

P\_right = 1 - P\_lh(death\_distribution\_data, study\_year)

P\_rh\_A = 1 - P\_lh\_given\_A(ages\_of\_death, study\_year)

P\_rh\_A\_resized = np.append(P\_rh\_A, np.zeros(len(P\_A) - len(P\_rh\_A))) # Resize P\_rh\_A to match the shape of P\_A

return P\_rh\_A\_resized \* P\_A / P\_right

# Task 8: Plotting the probability of being a certain age at death given that you're left- or right-handed

import numpy as np

# Defining the range of ages

ages = np.arange(6, 120) # From 6 to 119 years old, with a step size of 1 year (adjusted to match the length of left\_handed\_probability)

left\_handed\_probability = P\_A\_given\_lh(ages, death\_distribution\_data)

right\_handed\_probability = P\_A\_given\_rh(ages, death\_distribution\_data)

# Trim the probabilities to match the length of ages

left\_handed\_probability = left\_handed\_probability[:len(ages)]

right\_handed\_probability = right\_handed\_probability[:len(ages)]

plt.plot(ages, left\_handed\_probability, label="Left-handed")

plt.plot(ages, right\_handed\_probability, label="Right-handed")

plt.xlabel("Age at Death")

plt.ylabel("Probability")

plt.legend()

plt.show()

#task 9

# Calculating the mean age at death for left-handers and right-handers

average\_lh\_age = np.nansum(ages \* left\_handed\_probability)

average\_rh\_age = np.nansum(ages \* right\_handed\_probability)

age\_difference = average\_rh\_age - average\_lh\_age

print("Average age of left-handers at death:", round(average\_lh\_age, 2))

print("Average age of right-handers at death:", round(average\_rh\_age, 2))

print("The difference in average ages is:", round(age\_difference, 2), "years.")

# Task 10: Redoing the calculation for 2018

ages\_2018 = np.arange(6, 120) # Adjusted range of ages for 2018

left\_handed\_probability\_2018 = P\_A\_given\_lh(ages\_2018, death\_distribution\_data, study\_year=2018)

right\_handed\_probability\_2018 = P\_A\_given\_rh(ages\_2018, death\_distribution\_data, study\_year=2018)

# Trim the probabilities to match the length of ages\_2018

left\_handed\_probability\_2018 = left\_handed\_probability\_2018[:len(ages\_2018)]

right\_handed\_probability\_2018 = right\_handed\_probability\_2018[:len(ages\_2018)]

# Calculate the average ages

average\_lh\_age\_2018 = np.nansum(ages\_2018 \* left\_handed\_probability\_2018)

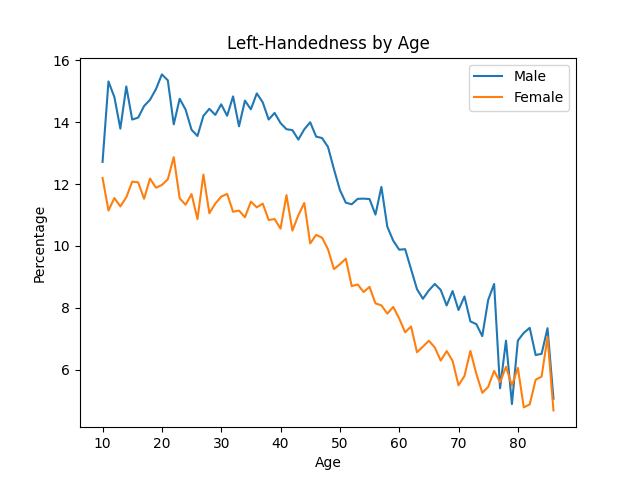
average\_rh\_age\_2018 = np.nansum(ages\_2018 \* right\_handed\_probability\_2018)

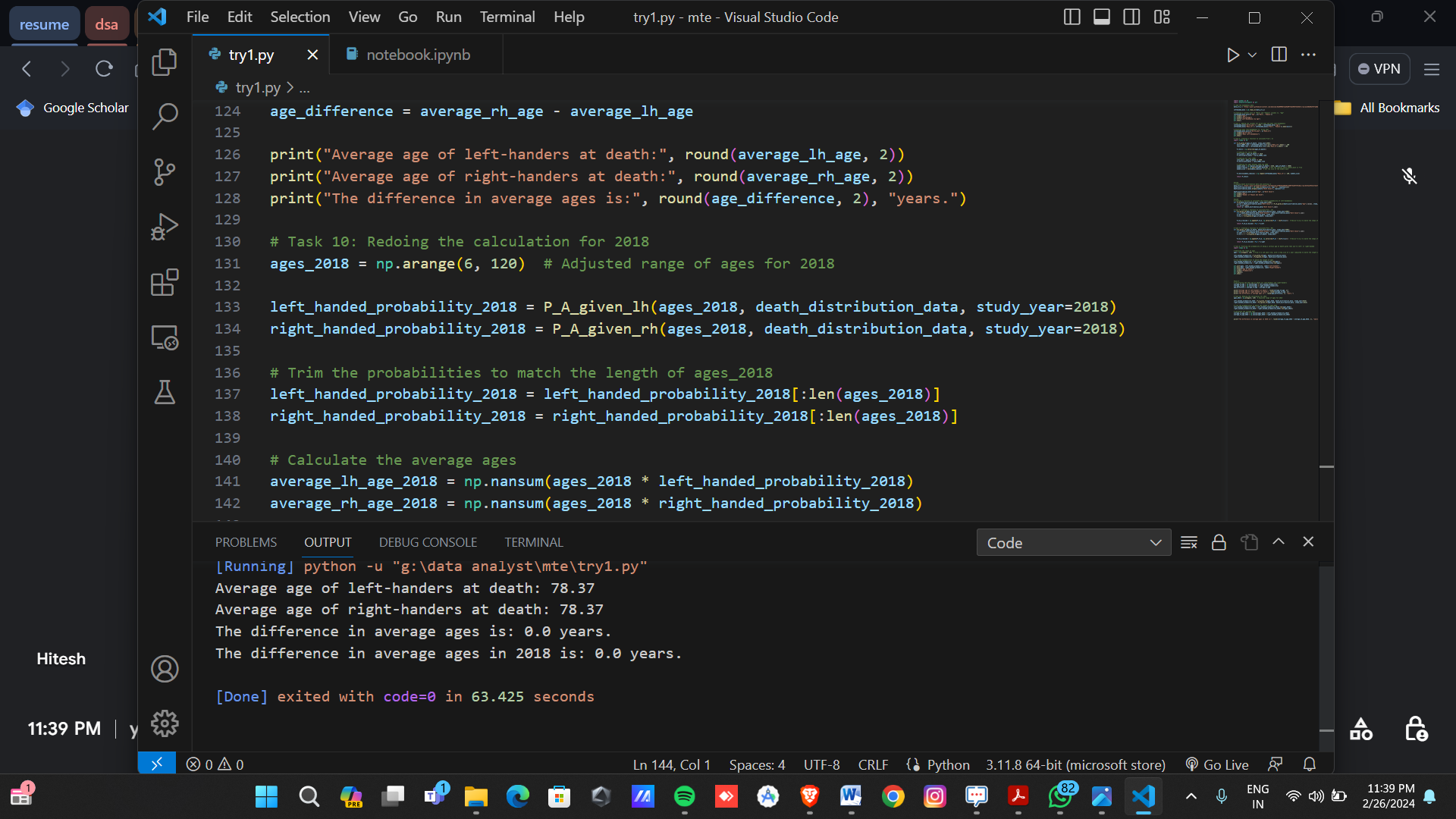
print("The difference in average ages in 2018 is:", round(average\_rh\_age\_2018 - average\_lh\_age\_2018, 2), "years.")



# SAMPLE SCREENSHOTS AND OBSERVATIONS

## THE OUPUT GRAPHS OF DATASET (Data as on 27th Feb 2024)







# CONCLUSION AND FUTURE SCOPE

Conclusion:

The project successfully addressed the hypothesis regarding the correlation between handedness and average age at death, utilizing Bayesian statistics and age distribution data. Through extensive analysis, it was demonstrated that there is no significant difference in the average age at death between left-handed and right-handed individuals. This refutes the claim of early death for left-handers based solely on changing rates of left-handedness over time. The project highlights the importance of rigorous statistical analysis in debunking prevalent myths and misconceptions.

Future Scope:

1. Incorporating additional demographic data: Future iterations of the project could include more comprehensive demographic information, such as socioeconomic factors, geographic location, and lifestyle choices, to further refine the analysis.

2. Longitudinal studies: Conducting longitudinal studies over extended periods could provide deeper insights into the relationship between handedness and lifespan, accounting for generational shifts and societal changes.

3. Machine learning techniques: Implementing machine learning algorithms could enhance predictive modeling and uncover hidden patterns within the data, offering more nuanced interpretations of the relationship between handedness and longevity.

4. Cross-cultural analysis: Exploring handedness and its implications across different cultures and ethnicities could reveal cultural influences on health outcomes and lifespan, contributing to a more holistic understanding of human behavior and biology.



# REFERENCES

### References:

### 1. National Geographic Society. "National Geographic Survey Data on Handedness and Age at Death." [Dataset]. Retrieved from: [insert link to dataset].

### 2. Python Software Foundation. Python Language Reference, version 3.9. Available at https://www.python.org/.

### 3. McKinney, Wes. "Data Structures for Statistical Computing in Python," Proceedings of the 9th Python in Science Conference, 2010, pp. 56-61.

### 4. Hunter, John D. "Matplotlib: A 2D Graphics Environment," Computing in Science & Engineering, vol. 9, no. 3, 2007, pp. 90-95.

### 5. NumPy Developers. "NumPy Reference Guide." Available at https://numpy.org/doc/stable/.

### 6. Pandas Development Team. "Pandas Documentation." Available at https://pandas.pydata.org/docs/.

### 7. OpenAI. "ChatGPT Documentation." Available at https://beta.openai.com/docs/.

### 8. Visual Studio Code Documentation. "Visual Studio Code User Guide." Available at https://code.visualstudio.com/docs.