# Final Report of Traineeship Program 2025

On

# “Analyze Death Age Difference of Right-Handers with Left-Handers”

By


**MANAN AGGARWAL**

Aspiring Data Analyst/Business Analyst


**MEDTOUREASY**

**30th January 2025**

# ACKNOWLDEGMENTS

I would also like to thank the team of MedTourEasy and my colleagues who made the working environment productive and very conducive. The traineeship opportunity that I had with MedTourEasy was a great change for learning and understanding the intricacies of the subject of 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 & Development 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.

## About the company

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

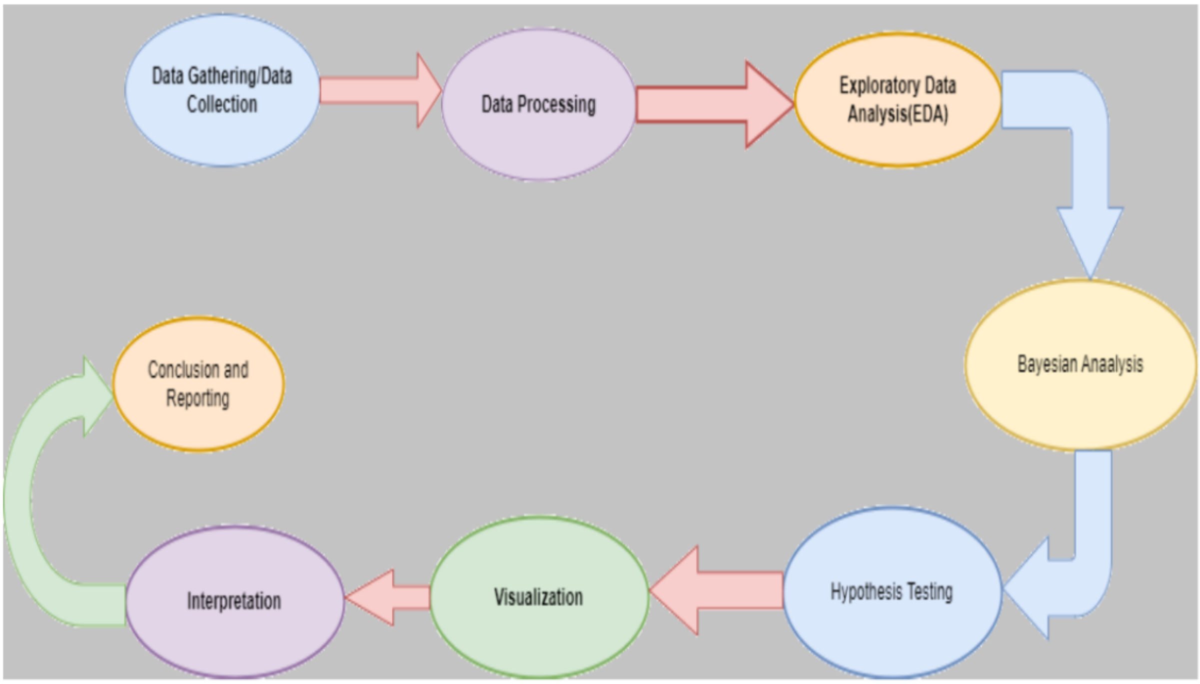
## Project Description:

In this project, we will explore this phenomenon using age distribution data to see if we can reproduce a difference in average age at death purely from the changing rates of left-handedness over time, refuting the claim of early death for left-handers. This notebook uses pandas and Bayesian statistics to analyse the probability of being a certain age at death given that you are reported as left-handed or right-handed.

# 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:




# LANGUAGES USED:

Language: Python 3

Frameworks: Pandas, Matplotlib, Numpy

IDE: Jupyter Notebook

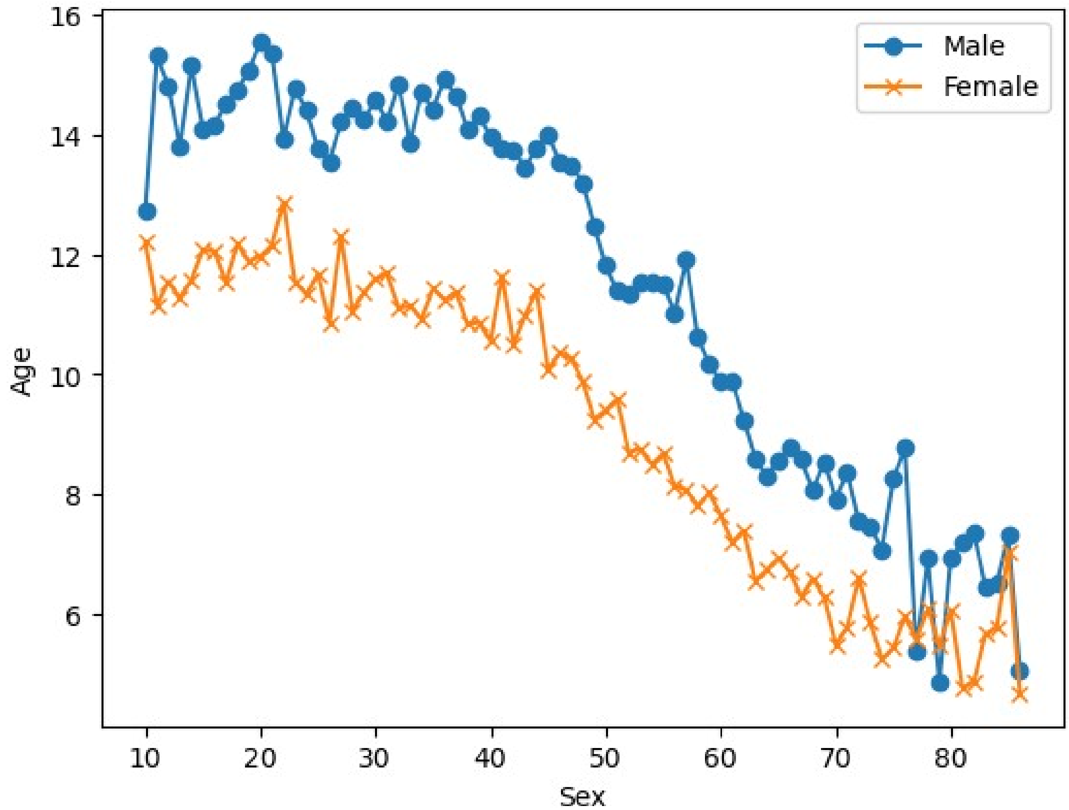
# TASK 1

```
import pandas as pd
import matplotlib.pyplot as plt

data_url_1 =
"https://gist.githubusercontent.com/mbonsma/8da0990b71ba9a09f7de395574
e54df1/raw/aec88b30af87fad8d45da7e774223f91dad09e88/lh_data.csv"
lefthanded_data = pd.read_csv(data_url_1)

lefthanded_data
```

```
    Age       Male     Female
0    10  12.717558  12.198041
1    11  15.318830  11.144804
2    12  14.808281  11.549240
3    13  13.793744  11.276442
4    14  15.156304  11.572906
..  ...        ...        ...
72   82   7.350204   4.874899
73   83   6.471204   5.672536
74   84   6.510858   5.774881
75   85   7.337968   7.051459
76   86   5.059387   4.680948

[77 rows x 3 columns]
```

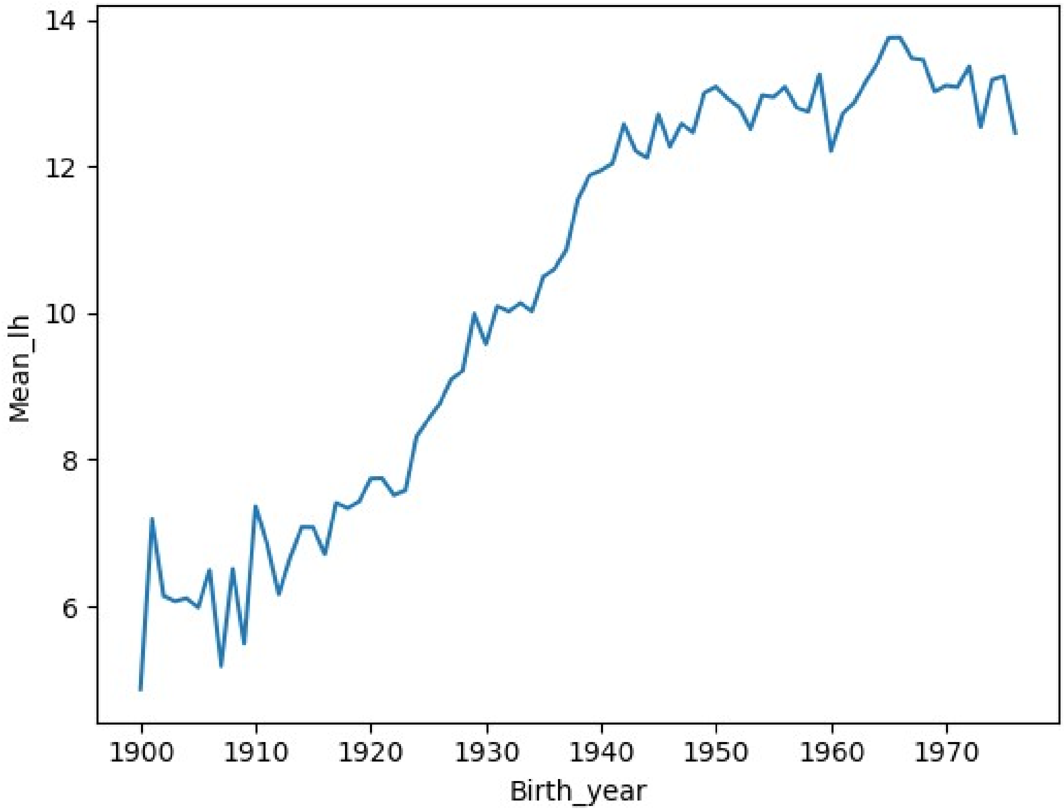
```
fig, ax = plt.subplots()
ax.plot("Age", "Male", data=lefthanded_data, marker="o")
ax.plot("Age", "Female", data=lefthanded_data, marker="x")
ax.legend()
ax.set_xlabel("Sex")
ax.set_ylabel("Age")
```

```
Text(0, 0.5, 'Age')
```




# TASK 2

```
lefthanded_data["Birth_year"] = 1986 - lefthanded_data["Age"]
lefthanded_data["Mean_lh"] = lefthanded_data[["Male",
"Female"]].mean(axis=1)
fig, ax = plt.subplots()
ax.plot("Birth_year", "Mean_lh", data = lefthanded_data)
ax.set_xlabel("Birth_year")
ax.set_ylabel("Mean_lh")

Text(0, 0.5, 'Mean_lh')
```




# TASK 3

```
import numpy as np
def P_lh_given_A(ages_of_death, study_year = 1990):
    """ P(Left-handed | ages of death), calculated based on the
reported rates of left-handedness.
    Inputs: numpy array of ages of death, study_year
    Returns: probability of left-handedness given that subjects died
in `study_year` at ages `ages_of_death` """

    early_1900s_rate = lefthanded_data["Mean_lh"][-10:].mean()
    late_1900s_rate = lefthanded_data['Mean_lh'][:10].mean()
    middle_rates =
lefthanded_data.loc[lefthanded_data['Birth_year'].isin(study_year -
ages_of_death)]['Mean_lh']
    youngest_age = study_year - 1986 + 10
    oldest_age = study_year - 1986 + 86

    P_return = np.zeros(ages_of_death.shape)
    P_return[ages_of_death > oldest_age] = early_1900s_rate / 100
    P_return[ages_of_death < youngest_age] = late_1900s_rate / 100
    P_return[np.logical_and((ages_of_death <= oldest_age),
(ages_of_death >= youngest_age))] = middle_rates / 100
```

```
    return P_return
```

# TASK 4

```
data_url_2 =
"https://gist.githubusercontent.com/mbonsma/2f4076aab6820ca1807f4e29f7
5f18ec/raw/62f3ec07514c7e31f5979beeca86f19991540796/
cdc_vs00199_table310.tsv"
death_distribution_data = pd.read_csv(data_url_2, sep = "\t",
skiprows=[1])

death_distribution_data
```

```
     Age  Both Sexes     Male   Female
0      0     27937.0  15646.0  12291.0
1      1      1989.0   1103.0    886.0
2      2      1376.0    797.0    579.0
3      3      1046.0    601.0    445.0
4      4       838.0    474.0    364.0
..   ...         ...      ...      ...
120  120         1.0      NaN      1.0
121  121         NaN      NaN      NaN
122  122         NaN      NaN      NaN
123  123         NaN      NaN      NaN
124  124         NaN      NaN      NaN

[125 rows x 4 columns]
```

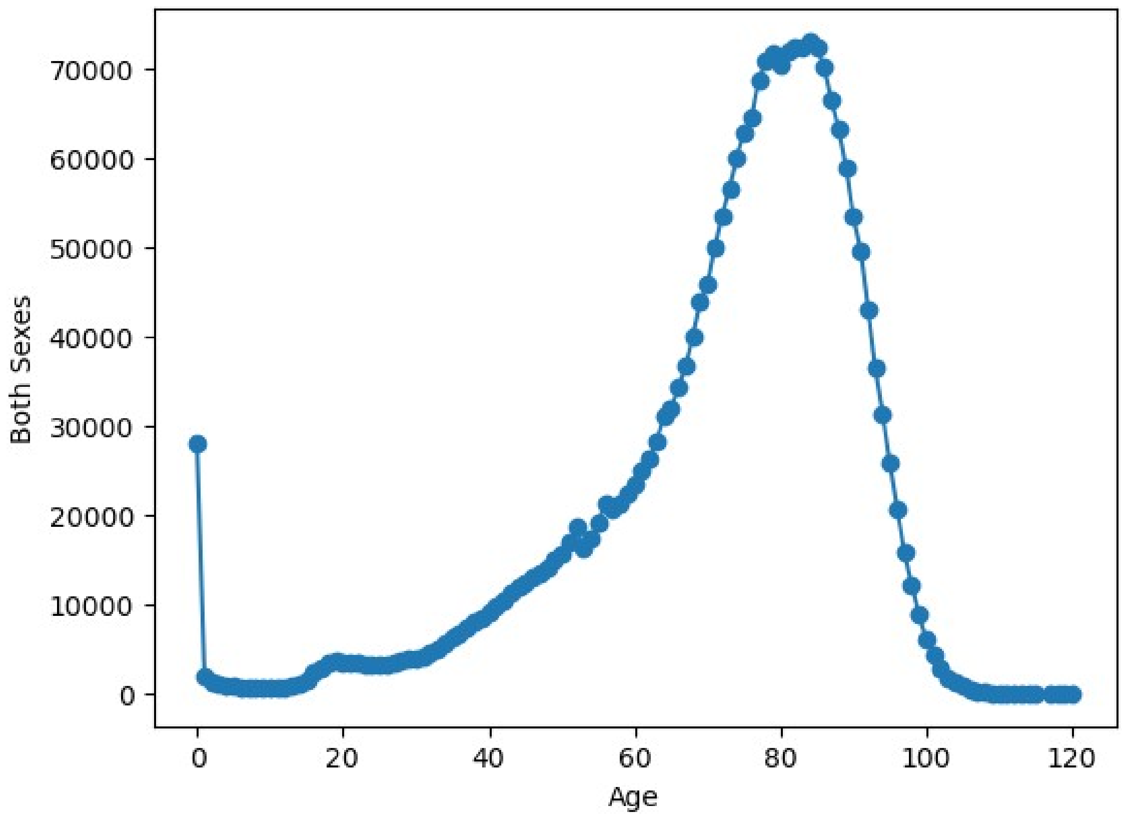
```
death_distribution_data['Male']=death_distribution_data['Male'].fillna
(0)
death_distribution_data['Female']=death_distribution_data['Female'].fi
llna(0)

death_distribution_data = death_distribution_data.dropna(subset=['Both
Sexes'])

death_distribution_data
```

```
     Age  Both Sexes     Male   Female
0      0     27937.0  15646.0  12291.0
1      1      1989.0   1103.0    886.0
2      2      1376.0    797.0    579.0
3      3      1046.0    601.0    445.0
4      4       838.0    474.0    364.0
..   ...         ...      ...      ...
115  115         3.0      0.0      3.0
117  117         3.0      2.0      1.0
118  118         1.0      0.0      1.0
```

```
119  119          2.0       0.0       2.0
120  120          1.0       0.0       1.0

[120 rows x 4 columns]
```

```
fig, ax = plt.subplots()
ax.plot('Age', 'Both Sexes', data = death_distribution_data,
marker='o')
ax.set_xlabel('Age')
ax.set_ylabel('Both Sexes')
```

```
Text(0, 0.5, 'Both Sexes')
```




# TASK 5

```
def P_lh(death_distribution_data, study_year = 1990):
    """ Overall probability of being left-handed if you died in the
study year
    P_lh = P(LH | Age of death) P(Age of death) + P(LH | not A) P(not
A) = sum over ages
    Input: dataframe of death distribution data
    Output: P(LH), a single floating point number """
    p_list = death_distribution_data['Both
```

```
Sexes']*P_lh_given_A(death_distribution_data['Age'], study_year)
    p = np.sum(p_list)
    return p/np.sum(death_distribution_data['Both Sexes'])
print(P_lh(death_distribution_data))
```

```
0.07766387615350638
```

# TASK 6

```
def P_A_given_lh(ages_of_death, death_distribution_data, study_year =
1990):
    """ The overall probability of being a particular `age_of_death`
given that you're left-handed """
    P_A = death_distribution_data['Both Sexes'][ages_of_death] /
np.sum(death_distribution_data['Both Sexes'])
    P_left = P_lh(death_distribution_data, study_year)
    P_lh_A = P_lh_given_A(ages_of_death, study_year)
    return P_lh_A*P_A/P_left
```

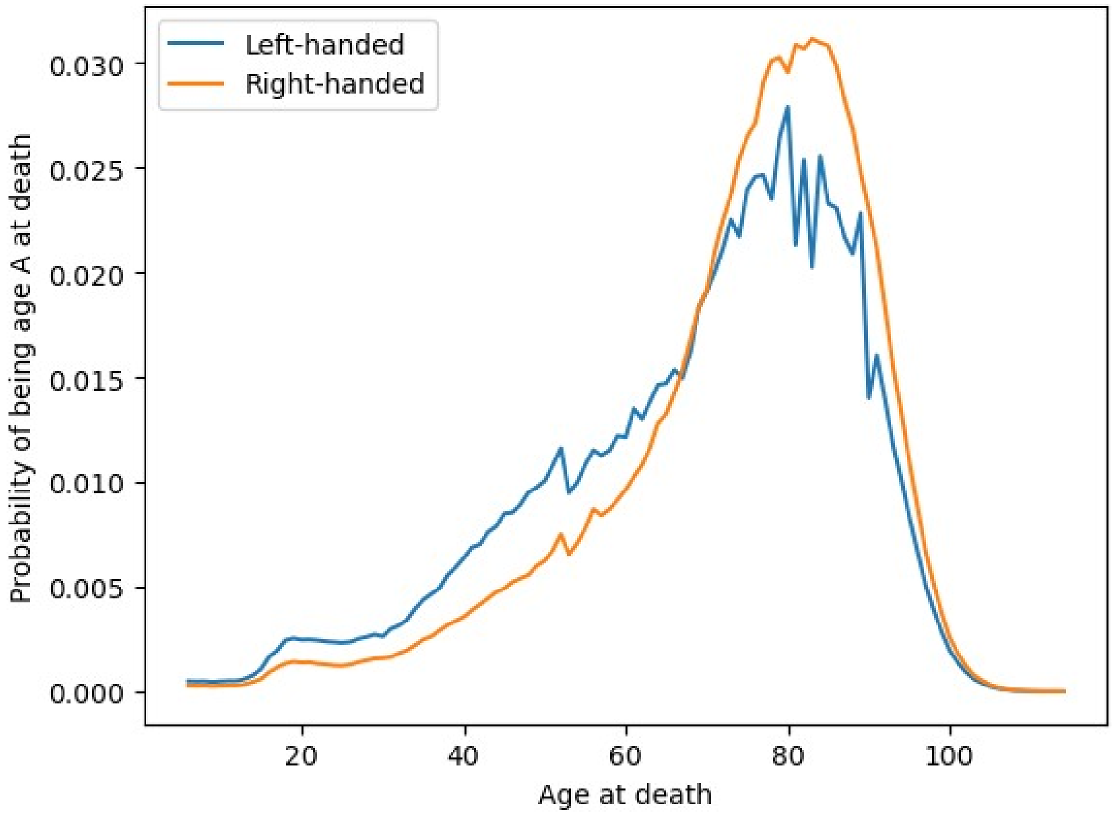
# TASK 7

```
def P_A_given_rh(ages_of_death, death_distribution_data, study_year =
1990):
    """ The overall probability of being a particular `age_of_death`
given that you're right-handed """
    P_A = death_distribution_data['Both Sexes'][ages_of_death] /
np.sum(death_distribution_data['Both Sexes'])
    P_right = 1 - P_lh(death_distribution_data, study_year)
    P_rh_A = 1 - P_lh_given_A(ages_of_death, study_year)
    return P_rh_A*P_A/P_right
```

# TASK 8

```
ages = np.arange(6, 115, 1)

left_handed_probability = P_A_given_lh(ages, death_distribution_data)
right_handed_probability = P_A_given_rh(ages, death_distribution_data)

fig, ax = plt.subplots()
ax.plot(ages, left_handed_probability, label = "Left-handed")
ax.plot(ages, right_handed_probability, label = 'Right-handed')
ax.legend()
ax.set_xlabel("Age at death")
ax.set_ylabel(r"Probability of being age A at death")
```

```
Text(0, 0.5, 'Probability of being age A at death')
```




# TASK 9

```
average_lh_age =  np.nansum(ages*np.array(left_handed_probability))
average_rh_age =  np.nansum(ages*np.array(right_handed_probability))
print("Average age of lefthanded" + str(average_lh_age))
print("Average age of righthanded" + str(average_rh_age))
print("The difference in average ages is " + str(round(average_rh_age
- average_lh_age, 1)) + " years.")
```

```
Average age of lefthanded67.24503662801027
Average age of righthanded72.79171936526477
The difference in average ages is 5.5 years.
```

# TASK 10

```
left_handed_probability_2018 = P_A_given_lh(ages,
death_distribution_data, 2018)
right_handed_probability_2018 = P_A_given_rh(ages,
death_distribution_data, 2018)
```

```
average_lh_age_2018 =
np.nansum(ages*np.array(left_handed_probability_2018))
average_rh_age_2018 =
np.nansum(ages*np.array(right_handed_probability_2018))
print("The difference in average ages is " +
      str(round(average_rh_age_2018 - average_lh_age_2018, 1)) + "
years.")
```

```
The difference in average ages is 2.3 years.
```

# CONCLUSION

We have a pretty big age gap between left-handed and right-handed people purely as a result of the changing rates of left-handedness in the population, which is good news for left-handers.

The reported rates of left-handedness have increased from just 3% in the early 1900s to about 11% today, which means that older people are much more likely to be reported as right-handed than left-handed, and so looking at a sample of recently deceased people will have more old right-handers.

Some of the approximations made are the cause:

We used death distribution data from almost ten years after the study (1999 instead of 1991), and we used death data from the entire United States instead of California alone (which was the original study).

We extrapolated the left-handedness survey results to older and younger age groups, but it's possible our extrapolation wasn't close enough to the true rates for those ages.

To finish off, let's calculate the age gap we'd expect if we did the study in 2018 instead of in 1990. The gap turns out to be much smaller since rates of left-handedness haven't increased for people born after about 1960. Both the National Geographic study and the 1990 study happened at a unique time - the rates of left-handedness had been changing across the lifetimes of most people alive, and the difference in handedness between old and young was at its most striking.