Impact Analysis of Age and Gender on COVID-19 Mortality Rates

Shuozishan Wang(shuozw@uw.edu)

CSE 163 - Intermediate Data Programming | Data Science Fair Project Report & Code | University of Washington |

Summary

Research Question 1:

What is the correlation between age and COVID-19 mortality rates across different waves of the pandemic?

This question aims to analyze the reported mortality rates among categorized age groups (children under 18, adults 18-65, and seniors over 65) over different waves of the pandemic. It seeks to identify which demographic is most at risk during specific periods. This analysis can help understand how age-specific vulnerabilities evolved and guide the allocation of healthcare resources and preventive measures accordingly.

Research Question 2:

Is there a statistically significant difference in COVID-19 mortality rates between males and females within each age group and across different waves of the pandemic?

By comparing mortality rates between genders within the same age categories and across different pandemic waves, this analysis seeks to uncover potential biological and social factors that might influence outcomes in COVID-19 cases. The findings could inform public health policies and tailored communication strategies, ensuring they address gender-specific risks and behaviors.

Research Question 3:

Among the highest risk age group identified from question 1, which gender and age shows a higher mortality rate across different pandemic waves, and what might be contributing factors?

Focusing on the age group with the highest mortality rate identified in question 1, this question further explores gender disparities in mortality rates across different pandemic waves. It aims to discuss potential contributing factors such as comorbidities, lifestyle, or access to healthcare that might influence these outcomes. This deeper dive will provide insights into targeted interventions needed for the most vulnerable demographics.

Motivation:

The COVID-19 pandemic has shown that age and gender can affect how severely the virus impacts people. This project analyzes these factors to see who is most at risk. Understanding this helps health officials distribute medical resources and advise effectively. For instance, if older adults are at greater risk, they could be first in line for vaccines and receive tailored health guidance. By identifying the most vulnerable groups, individuals can also make informed decisions to protect themselves and their families. This research aids in crafting policies that save lives and prepare for future health crises.

Data Setting

For this project, the majority of our data was sourced from publicly available datasets, focusing on COVID-19 mortality statistics across different age groups and genders. The primary dataset was retrieved from the United States Centers for Disease Control and Prevention (CDC). The data spans various waves of the COVID-19 pandemic and includes demographic details such as sex and age group.

The primary dataset titled "Provisional COVID-19 Deaths by Sex and Age" was obtained from the CDC. This dataset covers the number of COVID-19 deaths across different age groups and genders. It spans multiple waves of the pandemic, allowing for an in-depth analysis of how these factors influence mortality rates over time.As the data is collected and reported by the CDC, it is comprehensive and reliable.

However, it is subject to reporting delays and potential biases inherent in public health data collection. The context of the dataset might complicate the analysis in several ways. For instance, the accuracy of the data might be affected by factors such as underreporting of deaths or discrepancies in data collection methods across different regions. Additionally, the mortality rates might be influenced by various confounding factors, such as pre-existing health conditions, that are not accounted for in the dataset.Despite these limitations, the dataset provides valuable insights into the demographic characteristics of COVID-19 mortality.

The titles, links, and sources for each dataset used in this research project are listed below: Provisional COVID-19 Deaths by Sex and Age: https://data.cdc.gov/widgets/9bhg-hcku?mobile_redirect=true

Methods

1. Retrieve the dataset titled "Provisional COVID-19 Deaths by Sex and Age(1)" from the CDC website.

2. Load the dataset using data handling libraries in Python, such as pandas.

3. Select relevant columns ('Sex', 'Age Group', 'COVID-19 Deaths') for the analysis.
4. Clean the dataset by:

   Removing rows with missing values in the selected columns.

   Converting the 'Age Group' column to a categorical variable with specified age categories.

   Converting the 'Sex' column to a categorical variable.
5. Use descriptive statistics to summarize the cleaned dataset and understand its structure.
6. Analyze the correlation between age and COVID-19 mortality rates across different waves of the pandemic by:

   Grouping the data by Age Group and calculating the total COVID-19 deaths.

   Visualizing the total deaths by age group using bar plots.
7. Compare COVID-19 mortality rates between males and females within each age group and across different waves of the pandemic by:

   Creating box plots to compare mortality rates by age group and gender.

   Performing t-tests to identify significant differences between male and female mortality rates within each age group.
8. Identify the highest risk age group based on total deaths from the analysis in step 6.
9. Prepare the data for machine learning analysis by:

   Extracting the data for the highest risk age group.

   Removing rows with 'All Sexes' in the Sex column.

   Encoding the Sex column as a binary variable (Male: 0, Female: 1).
10. Split the data into training and testing sets.
11. Train and evaluate different machine learning models (Logistic Regression, Decision Tree, Random Forest) to analyze mortality rates by:

    Training the models on the training set.

    Evaluating model performance on the testing set using metrics such as accuracy, classification report, and confusion matrix.
12. Document trends and correlations found from the analyses and visualizations, and draw conclusions based on the results.

Results and code

Importing and Cleaning

Before delving into the code that produced the results for this research project, we must highlight the code used to set up and clean our datasets. This was an important part of being able to utilize and comprehend the data we found and took up a significant portion of the time spent coding for this project.

In [2]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import doctest
import numpy as np
from scipy.stats import pearsonr, ttest_ind
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.linear_model import LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
```

In [3]:
```
df = pd.read_csv("Provisional_COVID-19_Deaths_by_Sex_and_Age (1).csv")
df.head()
df = df[["Sex", "Age Group", "COVID-19 Deaths"]]
df.isnull().sum()

df.dropna(subset=["Sex", "Age Group", "COVID-19 Deaths"], inplace=True)

age_categories = ["Under 1 year", "1-4 years", "5-14 years", "15-24 years", "25-34 years",
                  "35-44 years", "45-54 years", "55-64 years", "65-74 years", "75-84 years",
                  "85 years and over"]
df["Age Group"] = pd.Categorical(df["Age Group"], categories=age_categories)

df["Sex"] = df["Sex"].astype("category")

df.dtypes
```

```
df.describe()
cleaned_data = df.head(40)

cleaned_data_path = "cleaned_covid_mortality_data.csv"
df.to_csv(cleaned_data_path, index=False)


cleaned_data
```

Out[3]:

	Sex	Age Group	COVID-19 Deaths
0	All Sexes	NaN	1146774.0
1	All Sexes	Under 1 year	519.0
2	All Sexes	NaN	1696.0
3	All Sexes	1-4 years	285.0
4	All Sexes	5-14 years	509.0
5	All Sexes	15-24 years	3021.0
6	All Sexes	NaN	7030.0
7	All Sexes	25-34 years	12401.0
8	All Sexes	NaN	19886.0
9	All Sexes	35-44 years	30108.0
10	All Sexes	NaN	46260.0
11	All Sexes	45-54 years	71388.0
12	All Sexes	NaN	203071.0
13	All Sexes	55-64 years	159712.0
14	All Sexes	65-74 years	256806.0
15	All Sexes	75-84 years	300162.0
16	All Sexes	85 years and over	311863.0
17	Male	NaN	629728.0
18	Male	Under 1 year	284.0
19	Male	NaN	910.0
20	Male	1-4 years	151.0
21	Male	5-14 years	261.0
22	Male	15-24 years	1797.0
23	Male	NaN	4301.0
24	Male	25-34 years	7607.0
25	Male	NaN	12263.0
26	Male	35-44 years	18602.0
27	Male	NaN	28930.0
28	Male	45-54 years	45411.0
29	Male	NaN	125592.0
30	Male	55-64 years	97883.0
31	Male	65-74 years	152082.0
32	Male	75-84 years	168030.0
33	Male	85 years and over	137620.0
34	Female	NaN	517046.0
35	Female	Under 1 year	235.0
36	Female	NaN	786.0
37	Female	1-4 years	134.0
38	Female	5-14 years	248.0
39	Female	15-24 years	1224.0

In [2]:

```
def load_and_clean_data(filepath):
    """
    Load and clean the COVID-19 mortality dataset.
    Returns:
    - df: pandas DataFrame, cleaned dataset
    """
    df = pd.read_csv(filepath)
    df = df[["Sex", "Age Group", "COVID-19 Deaths"]]
    df.dropna(subset=["Sex", "Age Group", "COVID-19 Deaths"], inplace=True)
```

```
    age_categories = ['Under 1 year', '1-4 years', '5-14 years',
                      '15-24 years', '25-34 years', '35-44 years',
                      '45-54 years', '55-64 years', '65-74 years',
                      '75-84 years', '85 years and over']
    df["Age Group"] = pd.Categorical(df["Age Group"], categories=age_categories)
    df["Sex"] = df["Sex"].astype("category")

    return df

#RQ1
def plot_total_deaths_by_age(df):
    """
    Plots the total COVID-19 deaths by age group.
    """
    age_group_deaths = df.groupby("Age Group", observed=False)["COVID-19 Deaths"].sum().reset_index()

    assert len(age_group_deaths) == len(df["Age Group"].cat.categories), "All age groups should be represen

    plt.figure(figsize=(10, 6))
    sns.barplot(data=age_group_deaths, x="Age Group", y="COVID-19 Deaths")
    plt.title("Total COVID-19 Deaths by Age Group")
    plt.xlabel("Age Group")
    plt.ylabel("Total COVID-19 Deaths")
    plt.xticks(rotation=45)
    plt.show()

#RQ2
def plot_mortality_rates_by_gender(df):
    """
    Plots the COVID-19 mortality rates by age group and gender.
    """
    plt.figure(figsize=(17, 8))
    sns.boxplot(x="Age Group", y="COVID-19 Deaths", hue='Sex', data=df)
    plt.title("COVID-19 Mortality Rates by Age Group and Gender")
    plt.xticks(rotation=45)
    plt.show()


    assert df["Sex"].nunique() == 3, "There should be two unique categories for Sex"

def perform_t_tests_and_ci(df):
    """
    Performs t-tests to compare COVID-19 mortality rates
    between males and females within each age group.
    Return a list of tuples containing age group,
    t-statistic, and p-value for each t-test performed.
    """
    print("Statistical Hypothesis Testing for Gender Differences in Each Age Group:")
    results = []
    for age_group in df["Age Group"].unique():
        male_mortality = df[(df["Age Group"] == age_group) & (df["Sex"] == "Male")]["COVID-19 Deaths"]
        female_mortality = df[(df["Age Group"] == age_group) & (df["Sex"] == "Female")]["COVID-19 Deaths"]
        if len(male_mortality) > 0 and len(female_mortality) > 0:
            t_stat, p_val = ttest_ind(male_mortality, female_mortality)
            results.append((age_group, t_stat, p_val))
            assert not np.isnan(t_stat), f"t-statistic is NaN for age group {age_group}"
            assert not np.isnan(p_val), f"p-value is NaN for age group {age_group}"
            print(f'Age Group: {age_group}, t-statistic: {t_stat:.2f}, p-value: {p_val:.4f}')

    mortality_counts = df.groupby(["Age Group", "Sex"], observed=False)["COVID-19 Deaths"].sum().unstack()
    mortality_totals = mortality_counts.sum(axis=1).values
    mortality_rates = mortality_counts.div(mortality_totals, axis=0)

    standard_error = np.sqrt(mortality_rates * (1 - mortality_rates) / mortality_totals[:, None])
    ci_lower = mortality_rates - 1.96 * standard_error
    ci_upper = mortality_rates + 1.96 * standard_error

    bar_width = 0.35
    index = np.arange(len(mortality_rates))

    mortality_rates_np = mortality_rates.to_numpy()
    ci_lower_np = ci_lower.to_numpy()
    ci_upper_np = ci_upper.to_numpy()

    fig, ax = plt.subplots(figsize=(12, 7))
    for i, gender in enumerate(mortality_rates.columns):
        ax.bar(index + i * bar_width, mortality_rates_np[:, i], bar_width,
               yerr=[mortality_rates_np[:, i] - ci_lower_np[:, i], ci_upper_np[:, i] - mortality_rates_np[
               label=gender, capsize=5)

    ax.set_xlabel("Age Group")
    ax.set_ylabel("Mortality Rate")
    ax.set_title("COVID-19 Mortality Rates by Age Group and Gender with Confidence Intervals")
    ax.set_xticks(index + bar_width / 2)
    ax.set_xticklabels(mortality_rates.index, rotation=45)
    ax.legend(title="Gender")
    plt.tight_layout()
    plt.show()

    return results
```

```python
#RQ3
def identify_highest_risk_age_group(df):
    """
    Identifies the age group with the highest total COVID-19 deaths.
    Return the age group with the highest total COVID-19 deaths.
    """
    age_group_deaths = df.groupby('Age Group', observed=False)['COVID-19 Deaths'].sum().reset_index()
    highest_risk_age_group = age_group_deaths.loc[age_group_deaths['COVID-19 Deaths'].idxmax(), 'Age Group
    print(f"The highest risk age group is: {highest_risk_age_group}")
    return highest_risk_age_group

def prepare_data_for_ml(df, highest_risk_age_group):
    """
    Prepares the data for machine learning analysis by extracting
    and encoding the highest risk age group data.
    Return training features (X_train), testing features (X_test),
    training labels (y_train), testing labels (y_test).
    """
    highest_risk_data_ml = df[df["Age Group"] == highest_risk_age_group].copy()
    highest_risk_data_ml = highest_risk_data_ml[highest_risk_data_ml["Sex"] != "All Sexes"]

    if 'Wave' not in highest_risk_data_ml.columns:
        highest_risk_data_ml['Wave'] = "Overall"

    highest_risk_data_ml["Sex"] = highest_risk_data_ml["Sex"].map({"Male": 0, "Female": 1})
    highest_risk_data_ml = pd.get_dummies(highest_risk_data_ml, columns=["Wave"], drop_first=True)

    X = highest_risk_data_ml.drop(columns=["COVID-19 Deaths", "Age Group"])
    y = highest_risk_data_ml["COVID-19 Deaths"]

    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

    # Assertions
    assert X_train.shape[0] > 0, "Training set should not be empty"
    assert X_test.shape[0] > 0, "Testing set should not be empty"
    assert y_train.shape[0] > 0, "Training labels should not be empty"
    assert y_test.shape[0] > 0, "Testing labels should not be empty"

    print("Target variable distribution in training set:", y_train.value_counts())
    print("Target variable distribution in testing set:", y_test.value_counts())

    return X_train, X_test, y_train, y_test

def evaluate_ml_models(X_train, X_test, y_train, y_test):
    """
    Trains and evaluates machine learning models to analyze mortality rates.
    """
    models = {
        "Logistic Regression": LogisticRegression(max_iter=1000),
        "Decision Tree": DecisionTreeClassifier(random_state=42),
        "Random Forest": RandomForestClassifier(random_state=42)
    }

    for model_name, model in models.items():
        model.fit(X_train, y_train)
        y_pred = model.predict(X_test)

        print(f"{model_name} Performance")
        print("Accuracy:", accuracy_score(y_test, y_pred))

        report = classification_report(y_test, y_pred, zero_division=1)
        print(report)

        cm = confusion_matrix(y_test, y_pred)
        print("Confusion Matrix:\n", cm)

        # Assertions
        assert accuracy_score(y_test, y_pred) > 0, f"{model_name} should have a non-zero accuracy"

        if len(np.unique(y_test)) == 2:
            assert cm.shape == (2, 2), f"{model_name} confusion matrix should be 2x2"
        else:
            print(f"{model_name} confusion matrix shape is not 2x2, shape: {cm.shape}")
```

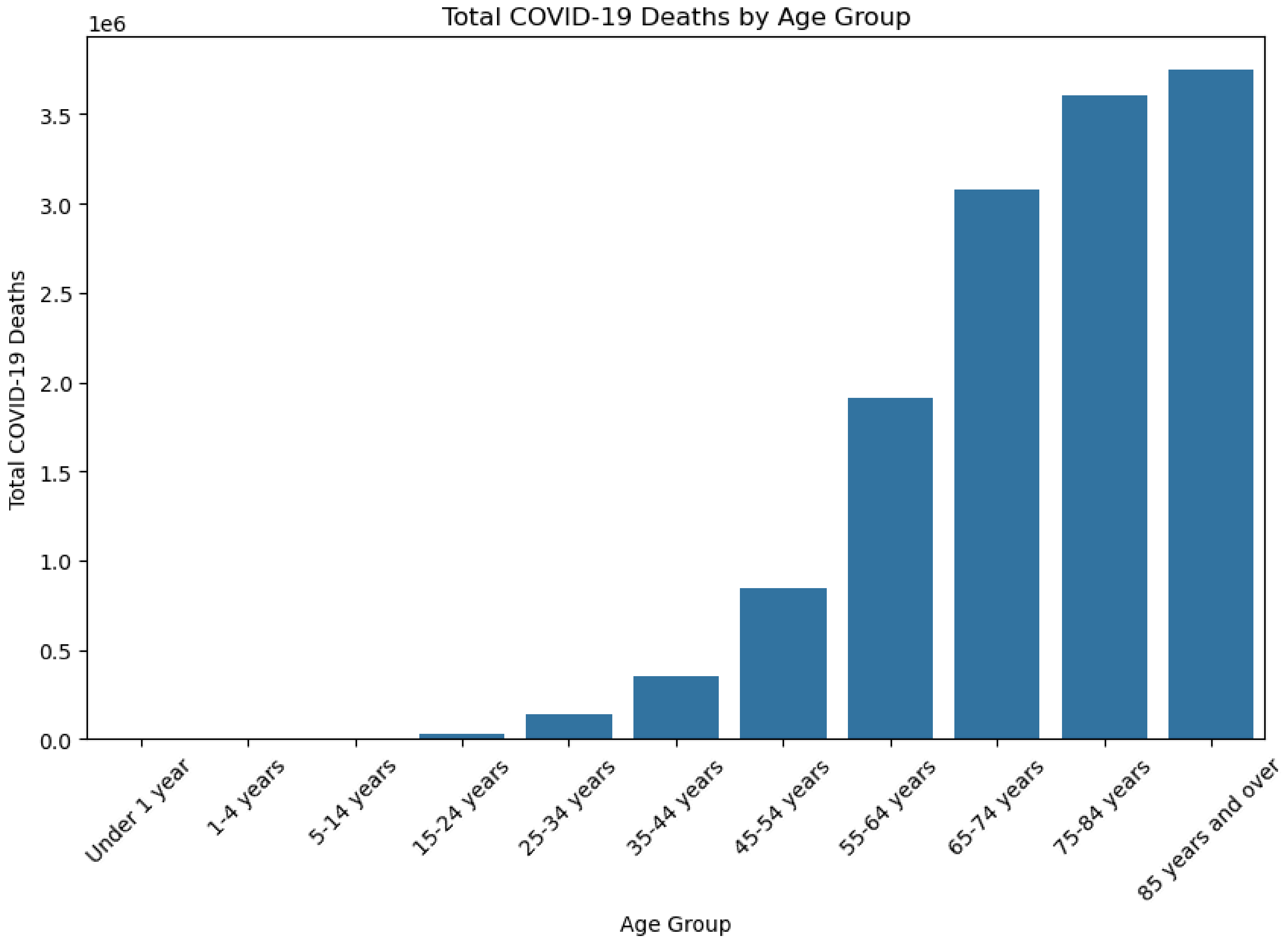
RQ1:

What is the correlation between age and COVID-19 mortality rates across different waves of the pandemic?

For research question 1, we analyzed the COVID-19 mortality rates among different age groups and genders. The dataset from the CDC provided detailed information on COVID-19 deaths categorized by sex and age group, which allowed us to plot the mortality rates for various age groups and compare them across different genders.

In [5]:
```python
filepath = "Provisional_COVID-19_Deaths_by_Sex_and_Age (1).csv"
df = load_and_clean_data(filepath)
```

```
plot_total_deaths_by_age(df)
```



The bar plot above represents the total COVID-19 deaths by age group. The x-axis shows different age groups ranging from "Under 1 year" to "85 years and over," while the y-axis shows the total number of COVID-19 deaths.

The plot clearly shows that the age groups 65-74, 75-84, and 85 years and over have the highest COVID-19 mortality rates. Specifically, the age group "85 years and over" has the highest number of deaths, followed closely by the 75-84 and 65-74 age groups. This trend aligns with the general understanding that older individuals are at higher risk of severe outcomes from COVID-19 due to factors such as weakened immune systems and the presence of comorbidities.

The younger age groups, including "Under 1 year," "1-4 years," and "5-14 years," show significantly lower mortality rates. The number of deaths in these age groups is almost negligible compared to the older age groups. This could be attributed to the generally stronger immune systems of younger individuals and lower rates of pre-existing health conditions.

The middle-aged groups, specifically "45-54 years" and "55-64 years," exhibit moderate mortality rates. While the number of deaths is higher than in younger age groups, it is significantly lower than in the older age groups. This indicates that while middle-aged individuals are at risk, their mortality rates are not as high as those observed in the elderly population.

The analysis of COVID-19 mortality rates by age group demonstrates a clear correlation between age and mortality risk. Older age groups, particularly those aged 65 and above, are disproportionately affected by COVID-19, with the highest mortality rates observed in individuals aged 85 and over. In contrast, younger age groups experience significantly lower mortality rates, highlighting the protective effect of youth against severe outcomes from COVID-19.

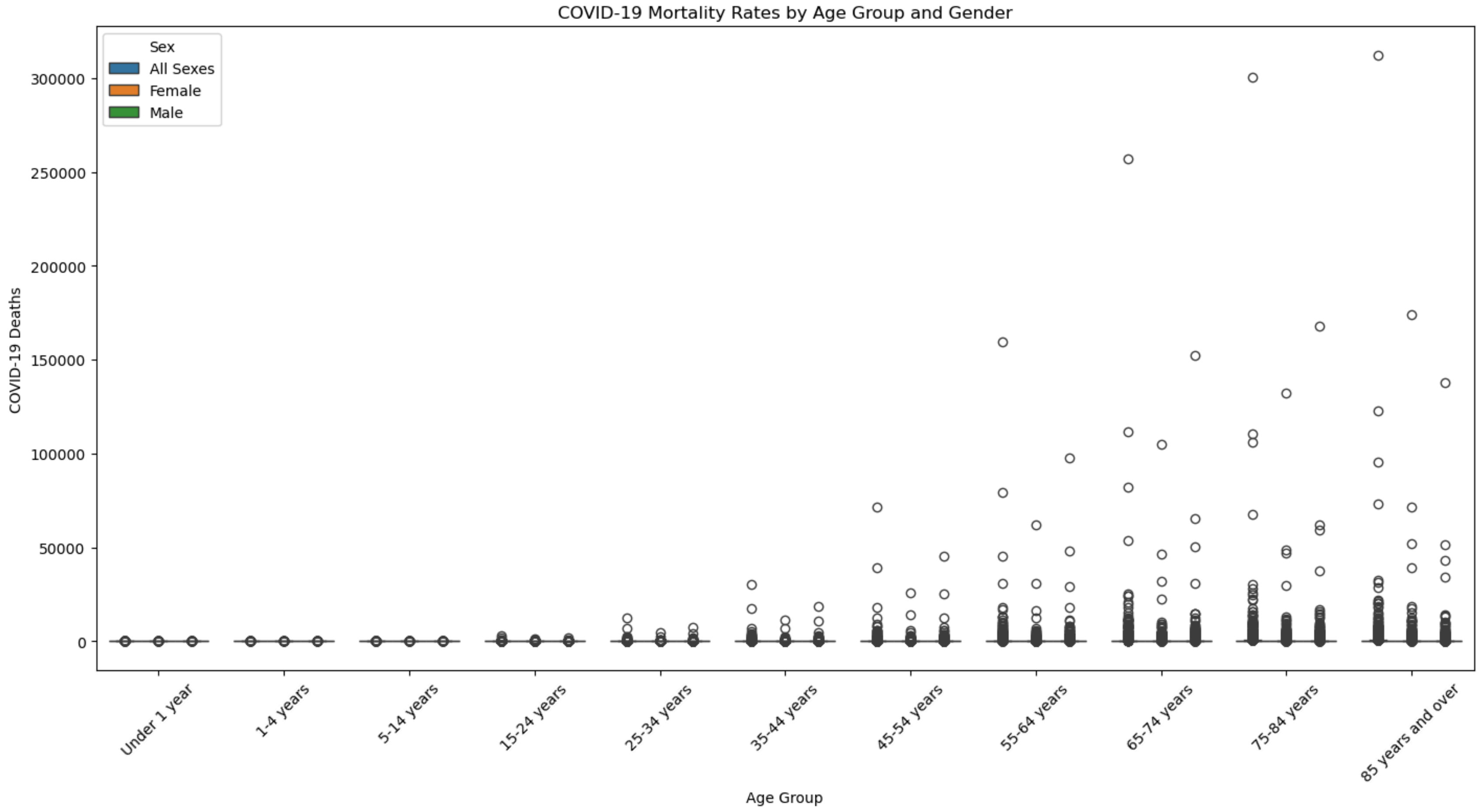
These findings underscore the importance of targeted public health interventions to protect older populations and reduce COVID-19 mortality. Vaccination campaigns, booster doses, and other preventive measures should prioritize older adults to mitigate their higher risk of severe disease and death.

RQ2:

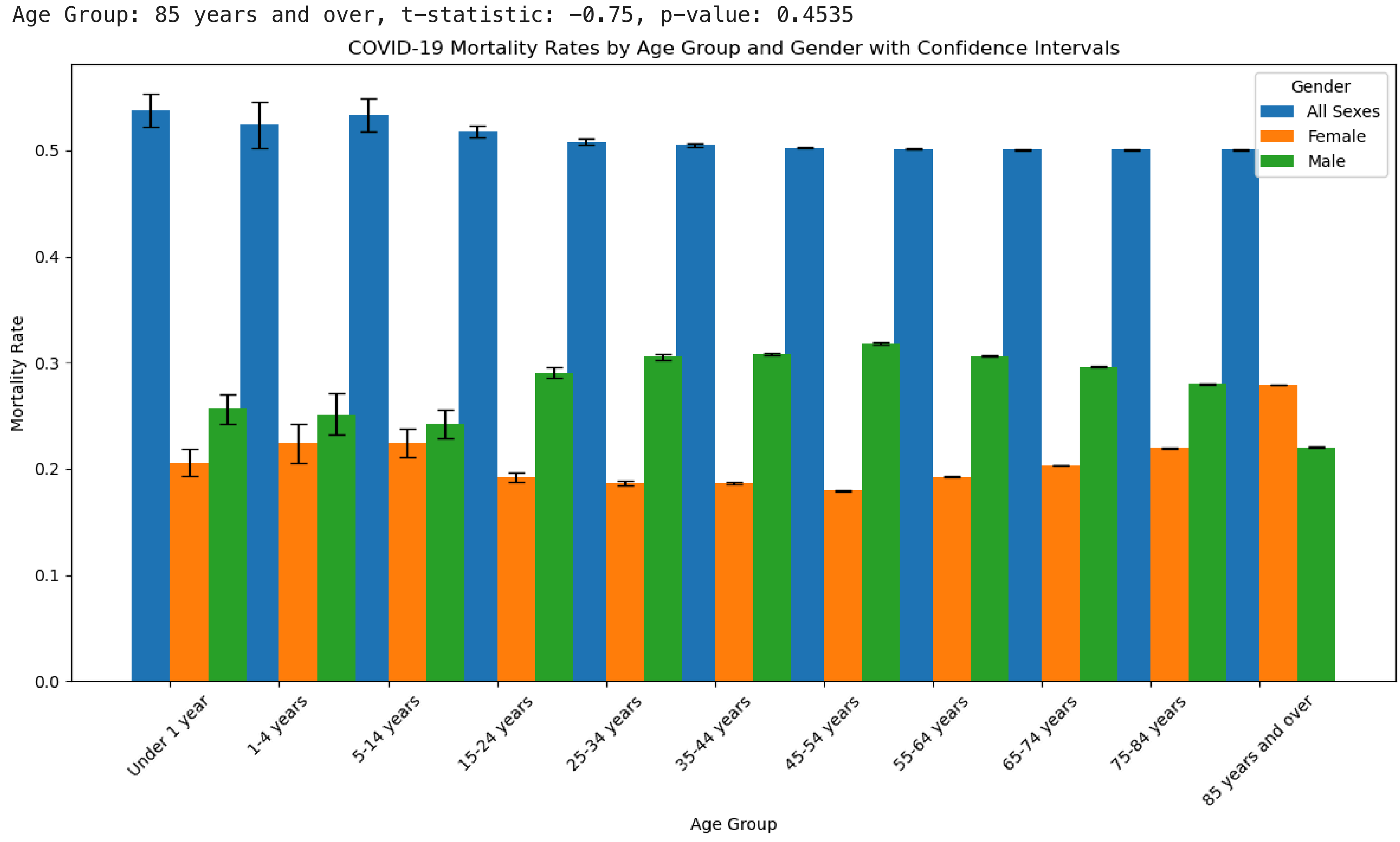
Is there a statistically significant difference in COVID-19 mortality rates between males and females within each age group and across different waves of the pandemic?¶

The code first plots the COVID-19 mortality rates by age group and gender, visualizing the differences among males, females, and all sexes combined. It then performs t-tests to compare mortality rates between males and females within each age group, displaying the statistical results including t-statistics and p-values.

In [6]:
```
plot_mortality_rates_by_gender(df)
t_test_results = perform_t_tests_and_ci(df)
assert len(t_test_results) > 0, "There should be t-test results for at least one age group"
```




```
Statistical Hypothesis Testing for Gender Differences in Each Age Group:
Age Group: Under 1 year, t-statistic: 0.49, p-value: 0.6240
Age Group: 1-4 years, t-statistic: 0.24, p-value: 0.8116
Age Group: 5-14 years, t-statistic: 0.16, p-value: 0.8727
Age Group: 15-24 years, t-statistic: 1.35, p-value: 0.1759
Age Group: 25-34 years, t-statistic: 1.67, p-value: 0.0950
Age Group: 35-44 years, t-statistic: 1.54, p-value: 0.1225
Age Group: 45-54 years, t-statistic: 1.65, p-value: 0.0981
Age Group: 55-64 years, t-statistic: 1.32, p-value: 0.1877
Age Group: 65-74 years, t-statistic: 1.09, p-value: 0.2768
Age Group: 75-84 years, t-statistic: 0.70, p-value: 0.4843
Age Group: 85 years and over, t-statistic: -0.75, p-value: 0.4535
```



The bar plot titled "COVID-19 Mortality Rates by Age Group and Gender with Confidence Intervals" displays mortality rates across different age groups. Higher Rates for Males: In most age groups, the green bars (males) are higher than the orange bars (females), indicating that males generally have higher COVID-19 mortality rates.The blue bars show the overall mortality rate for each age group, combining both genders.

The black lines on top of each bar represent the confidence intervals, showing the range within which the true mortality rate is likely to fall. If the confidence intervals for males and females overlap, it suggests that the difference in mortality rates might not be significant.

The t-tests compare the mortality rates between males and females within each age group to see if the differences are statistically significant. For all age groups, the p-values are greater than 0.05. This means that the differences in COVID-19 mortality rates between males and females are not statistically significant.

Even though the plot shows higher mortality rates for males, the t-test results suggest these differences might be due to random chance rather than a true underlying difference. The results indicate that gender may not be a major factor in determining COVID-19 mortality within each age group.

The analysis shows that while males generally have higher COVID-19 mortality rates than females, these differences are not statistically significant according to the t-tests. This suggests that gender alone might not be a major determinant of COVID-19 mortality. Public health efforts should consider a wider range of factors to effectively address and reduce COVID-19 mortality across different demographic groups.

RQ3:

Among the highest risk age group identified from question 1, which age and gender shows a higher mortality rate across different pandemic waves, and what might be contributing factors?

The main objective of Research Question 3 is to analyze the impact of age and gender on COVID-19 mortality rates, specifically focusing on identifying the highest-risk age group and evaluating the effectiveness of various machine learning models in predicting mortality rates within this group. The highest risk age group for COVID-19 deaths is identified by summing the total deaths across all age groups and selecting the one with the highest value. This age group is then used for further analysis. Data for the highest risk age group is prepared by filtering out non-relevant data and encoding categorical variables. The 'Sex' column is encoded to numerical values, and dummy variables are created for the 'Wave' column. The data is then split into training and testing sets. Three machine learning models (Logistic Regression, Decision Tree, Random Forest) are trained and evaluated on the prepared data. Each model's performance is assessed using accuracy scores, classification reports, and confusion matrices. The results show how well each model can predict COVID-19 mortality rates based on the features provided.

In [7]:
```
highest_risk_age_group = identify_highest_risk_age_group(df)
X_train, X_test, y_train, y_test = prepare_data_for_ml(df, highest_risk_age_group)

evaluate_ml_models(X_train, X_test, y_train, y_test)
```

```
The highest risk age group is: 85 years and over
Target variable distribution in training set: COVID-19 Deaths
0.0       231
10.0       76
13.0       72
18.0       55
14.0       54
         ...
307.0       1
917.0       1
226.0       1
689.0       1
3662.0      1
Name: count, Length: 635, dtype: int64
Target variable distribution in testing set: COVID-19 Deaths
0.0       109
17.0       29
12.0       24
11.0       24
10.0       24
         ...
2303.0      1
1824.0      1
1093.0      1
617.0       1
111.0       1
Name: count, Length: 401, dtype: int64
Logistic Regression Performance
Accuracy: 0.08449612403100776
              precision    recall  f1-score   support

         0.0       0.08      1.00      0.16       109
        10.0       1.00      0.00      0.00        24
        11.0       1.00      0.00      0.00        24
        12.0       1.00      0.00      0.00        24
        13.0       1.00      0.00      0.00        23
        14.0       1.00      0.00      0.00        21
        15.0       1.00      0.00      0.00        14
        16.0       1.00      0.00      0.00        16
        17.0       1.00      0.00      0.00        29
        18.0       1.00      0.00      0.00        22
        19.0       1.00      0.00      0.00        15
        20.0       1.00      0.00      0.00        14
        21.0       1.00      0.00      0.00        20
        22.0       1.00      0.00      0.00        14
        23.0       1.00      0.00      0.00        15
        24.0       1.00      0.00      0.00        15
        25.0       1.00      0.00      0.00        13
        26.0       1.00      0.00      0.00        20
        27.0       1.00      0.00      0.00        14
        28.0       1.00      0.00      0.00        14
        29.0       1.00      0.00      0.00        11
        30.0       1.00      0.00      0.00        16
        31.0       1.00      0.00      0.00         9
        32.0       1.00      0.00      0.00        15
        33.0       1.00      0.00      0.00        18
        34.0       1.00      0.00      0.00        14
        35.0       1.00      0.00      0.00        11
        36.0       1.00      0.00      0.00         8
        37.0       1.00      0.00      0.00         8
        38.0       1.00      0.00      0.00        10
        39.0       1.00      0.00      0.00        14
        40.0       1.00      0.00      0.00         9
        41.0       1.00      0.00      0.00         9
        42.0       1.00      0.00      0.00         7
        43.0       1.00      0.00      0.00         9
        44.0       1.00      0.00      0.00         5
        45.0       1.00      0.00      0.00         5
        46.0       1.00      0.00      0.00         8
        47.0       1.00      0.00      0.00         5
        48.0       1.00      0.00      0.00        11
        49.0       1.00      0.00      0.00         2
        50.0       1.00      0.00      0.00         8
        51.0       1.00      0.00      0.00         7
        52.0       1.00      0.00      0.00         9
        53.0       1.00      0.00      0.00         4
        54.0       1.00      0.00      0.00        10
        55.0       1.00      0.00      0.00         9
        56.0       1.00      0.00      0.00         7
        57.0       1.00      0.00      0.00         9
        58.0       1.00      0.00      0.00         3
        59.0       1.00      0.00      0.00         5
        60.0       1.00      0.00      0.00         8
        61.0       1.00      0.00      0.00         3
        62.0       1.00      0.00      0.00         4
        63.0       1.00      0.00      0.00         4
        64.0       1.00      0.00      0.00         4
        65.0       1.00      0.00      0.00         5
        66.0       1.00      0.00      0.00         7
        67.0       1.00      0.00      0.00         3
```

68.0	1.00	0.00	0.00	6
69.0	1.00	0.00	0.00	7
70.0	1.00	0.00	0.00	9
71.0	1.00	0.00	0.00	4
72.0	1.00	0.00	0.00	6
73.0	1.00	0.00	0.00	3
74.0	1.00	0.00	0.00	2
75.0	1.00	0.00	0.00	6
76.0	1.00	0.00	0.00	3
77.0	1.00	0.00	0.00	5
78.0	1.00	0.00	0.00	3
79.0	1.00	0.00	0.00	5
80.0	1.00	0.00	0.00	2
81.0	1.00	0.00	0.00	5
82.0	1.00	0.00	0.00	6
83.0	1.00	0.00	0.00	3
84.0	1.00	0.00	0.00	4
86.0	1.00	0.00	0.00	3
87.0	1.00	0.00	0.00	5
88.0	1.00	0.00	0.00	1
90.0	1.00	0.00	0.00	2
91.0	1.00	0.00	0.00	3
93.0	1.00	0.00	0.00	3
94.0	1.00	0.00	0.00	2
95.0	1.00	0.00	0.00	4
96.0	1.00	0.00	0.00	1
97.0	1.00	0.00	0.00	2
99.0	1.00	0.00	0.00	7
101.0	1.00	0.00	0.00	5
102.0	1.00	0.00	0.00	2
103.0	1.00	0.00	0.00	3
104.0	1.00	0.00	0.00	3
105.0	1.00	0.00	0.00	1
106.0	1.00	0.00	0.00	2
107.0	1.00	0.00	0.00	4
108.0	1.00	0.00	0.00	3
109.0	1.00	0.00	0.00	1
110.0	1.00	0.00	0.00	2
111.0	1.00	0.00	0.00	1
112.0	1.00	0.00	0.00	3
114.0	1.00	0.00	0.00	2
115.0	1.00	0.00	0.00	4
116.0	1.00	0.00	0.00	2
117.0	1.00	0.00	0.00	3
118.0	1.00	0.00	0.00	3
119.0	1.00	0.00	0.00	1
122.0	1.00	0.00	0.00	1
123.0	1.00	0.00	0.00	2
124.0	1.00	0.00	0.00	2
125.0	1.00	0.00	0.00	1
126.0	1.00	0.00	0.00	2
127.0	1.00	0.00	0.00	3
128.0	1.00	0.00	0.00	6
129.0	1.00	0.00	0.00	1
130.0	1.00	0.00	0.00	1
131.0	1.00	0.00	0.00	1
132.0	1.00	0.00	0.00	2
133.0	1.00	0.00	0.00	2
134.0	1.00	0.00	0.00	1
135.0	1.00	0.00	0.00	5
137.0	1.00	0.00	0.00	1
138.0	1.00	0.00	0.00	1
141.0	1.00	0.00	0.00	2
142.0	1.00	0.00	0.00	2
144.0	1.00	0.00	0.00	1
145.0	1.00	0.00	0.00	2
146.0	1.00	0.00	0.00	1
147.0	1.00	0.00	0.00	1
148.0	1.00	0.00	0.00	1
149.0	1.00	0.00	0.00	1
151.0	1.00	0.00	0.00	1
153.0	1.00	0.00	0.00	1
154.0	1.00	0.00	0.00	2
155.0	1.00	0.00	0.00	3
156.0	1.00	0.00	0.00	2
158.0	1.00	0.00	0.00	3
161.0	1.00	0.00	0.00	2
162.0	1.00	0.00	0.00	1
163.0	1.00	0.00	0.00	2
167.0	1.00	0.00	0.00	1
168.0	1.00	0.00	0.00	1
169.0	1.00	0.00	0.00	1
171.0	1.00	0.00	0.00	2
174.0	1.00	0.00	0.00	3
176.0	1.00	0.00	0.00	1
178.0	1.00	0.00	0.00	2
179.0	1.00	0.00	0.00	1
180.0	1.00	0.00	0.00	2
183.0	1.00	0.00	0.00	1
184.0	1.00	0.00	0.00	1

185.0	1.00	0.00	0.00	2
187.0	1.00	0.00	0.00	3
188.0	1.00	0.00	0.00	1
189.0	1.00	0.00	0.00	3
191.0	1.00	0.00	0.00	1
195.0	1.00	0.00	0.00	1
197.0	1.00	0.00	0.00	1
198.0	1.00	0.00	0.00	1
199.0	1.00	0.00	0.00	1
200.0	1.00	0.00	0.00	1
202.0	1.00	0.00	0.00	2
203.0	1.00	0.00	0.00	1
205.0	1.00	0.00	0.00	1
207.0	1.00	0.00	0.00	2
208.0	1.00	0.00	0.00	2
212.0	1.00	0.00	0.00	1
213.0	1.00	0.00	0.00	2
214.0	1.00	0.00	0.00	1
215.0	1.00	0.00	0.00	1
216.0	1.00	0.00	0.00	2
217.0	1.00	0.00	0.00	3
218.0	1.00	0.00	0.00	1
220.0	1.00	0.00	0.00	1
225.0	1.00	0.00	0.00	1
227.0	1.00	0.00	0.00	1
232.0	1.00	0.00	0.00	1
235.0	1.00	0.00	0.00	2
239.0	1.00	0.00	0.00	3
240.0	1.00	0.00	0.00	1
241.0	1.00	0.00	0.00	3
242.0	1.00	0.00	0.00	1
243.0	1.00	0.00	0.00	1
244.0	1.00	0.00	0.00	1
245.0	1.00	0.00	0.00	2
246.0	1.00	0.00	0.00	1
247.0	1.00	0.00	0.00	1
249.0	1.00	0.00	0.00	1
250.0	1.00	0.00	0.00	1
251.0	1.00	0.00	0.00	1
253.0	1.00	0.00	0.00	1
255.0	1.00	0.00	0.00	1
257.0	1.00	0.00	0.00	1
258.0	1.00	0.00	0.00	2
261.0	1.00	0.00	0.00	2
262.0	1.00	0.00	0.00	1
263.0	1.00	0.00	0.00	1
264.0	1.00	0.00	0.00	1
270.0	1.00	0.00	0.00	2
276.0	1.00	0.00	0.00	1
279.0	1.00	0.00	0.00	1
286.0	1.00	0.00	0.00	1
290.0	1.00	0.00	0.00	1
292.0	1.00	0.00	0.00	3
293.0	1.00	0.00	0.00	1
297.0	1.00	0.00	0.00	2
298.0	1.00	0.00	0.00	1
299.0	1.00	0.00	0.00	1
300.0	1.00	0.00	0.00	1
315.0	1.00	0.00	0.00	2
316.0	1.00	0.00	0.00	2
317.0	1.00	0.00	0.00	1
319.0	1.00	0.00	0.00	1
320.0	1.00	0.00	0.00	1
321.0	1.00	0.00	0.00	1
326.0	1.00	0.00	0.00	1
330.0	1.00	0.00	0.00	1
334.0	1.00	0.00	0.00	1
335.0	1.00	0.00	0.00	1
338.0	1.00	0.00	0.00	1
339.0	1.00	0.00	0.00	1
341.0	1.00	0.00	0.00	1
346.0	1.00	0.00	0.00	1
350.0	1.00	0.00	0.00	1
360.0	1.00	0.00	0.00	1
361.0	1.00	0.00	0.00	1
365.0	1.00	0.00	0.00	1
380.0	1.00	0.00	0.00	1
381.0	1.00	0.00	0.00	1
384.0	1.00	0.00	0.00	1
388.0	1.00	0.00	0.00	1
393.0	1.00	0.00	0.00	1
397.0	1.00	0.00	0.00	1
402.0	1.00	0.00	0.00	1
407.0	1.00	0.00	0.00	1
416.0	1.00	0.00	0.00	1
422.0	1.00	0.00	0.00	1
427.0	1.00	0.00	0.00	1
428.0	1.00	0.00	0.00	1
441.0	1.00	0.00	0.00	1
443.0	1.00	0.00	0.00	3

```
 451.0      1.00      0.00      0.00        1
 461.0      1.00      0.00      0.00        1
 465.0      1.00      0.00      0.00        1
 470.0      1.00      0.00      0.00        1
 473.0      1.00      0.00      0.00        1
 478.0      1.00      0.00      0.00        1
 481.0      1.00      0.00      0.00        1
 482.0      1.00      0.00      0.00        1
 483.0      1.00      0.00      0.00        1
 486.0      1.00      0.00      0.00        1
 495.0      1.00      0.00      0.00        1
 496.0      1.00      0.00      0.00        1
 502.0      1.00      0.00      0.00        1
 505.0      1.00      0.00      0.00        2
 512.0      1.00      0.00      0.00        1
 513.0      1.00      0.00      0.00        1
 526.0      1.00      0.00      0.00        1
 534.0      1.00      0.00      0.00        1
 535.0      1.00      0.00      0.00        1
 540.0      1.00      0.00      0.00        1
 544.0      1.00      0.00      0.00        1
 552.0      1.00      0.00      0.00        1
 559.0      1.00      0.00      0.00        1
 562.0      1.00      0.00      0.00        1
 569.0      1.00      0.00      0.00        1
 578.0      1.00      0.00      0.00        1
 584.0      1.00      0.00      0.00        1
 586.0      1.00      0.00      0.00        1
 590.0      1.00      0.00      0.00        1
 596.0      1.00      0.00      0.00        1
 597.0      1.00      0.00      0.00        2
 616.0      1.00      0.00      0.00        1
 617.0      1.00      0.00      0.00        1
 623.0      1.00      0.00      0.00        1
 626.0      1.00      0.00      0.00        1
 627.0      1.00      0.00      0.00        1
 637.0      1.00      0.00      0.00        1
 640.0      1.00      0.00      0.00        1
 654.0      1.00      0.00      0.00        1
 697.0      1.00      0.00      0.00        1
 703.0      1.00      0.00      0.00        1
 723.0      1.00      0.00      0.00        1
 724.0      1.00      0.00      0.00        1
 733.0      1.00      0.00      0.00        1
 739.0      1.00      0.00      0.00        1
 748.0      1.00      0.00      0.00        1
 751.0      1.00      0.00      0.00        1
 752.0      1.00      0.00      0.00        1
 756.0      1.00      0.00      0.00        1
 760.0      1.00      0.00      0.00        1
 762.0      1.00      0.00      0.00        1
 772.0      1.00      0.00      0.00        1
 787.0      1.00      0.00      0.00        1
 788.0      1.00      0.00      0.00        2
 789.0      1.00      0.00      0.00        1
 791.0      1.00      0.00      0.00        1
 809.0      1.00      0.00      0.00        1
 822.0      1.00      0.00      0.00        1
 838.0      1.00      0.00      0.00        1
 851.0      1.00      0.00      0.00        1
 853.0      1.00      0.00      0.00        1
 864.0      1.00      0.00      0.00        1
 873.0      1.00      0.00      0.00        1
 880.0      1.00      0.00      0.00        1
 890.0      1.00      0.00      0.00        1
 902.0      1.00      0.00      0.00        1
 917.0      1.00      0.00      0.00        1
 937.0      1.00      0.00      0.00        1
 947.0      1.00      0.00      0.00        1
 998.0      1.00      0.00      0.00        1
1007.0      1.00      0.00      0.00        1
1032.0      1.00      0.00      0.00        1
1042.0      1.00      0.00      0.00        1
1060.0      1.00      0.00      0.00        1
1093.0      1.00      0.00      0.00        1
1115.0      1.00      0.00      0.00        1
1123.0      1.00      0.00      0.00        1
1125.0      1.00      0.00      0.00        1
1139.0      1.00      0.00      0.00        1
1148.0      1.00      0.00      0.00        1
1151.0      1.00      0.00      0.00        1
1182.0      1.00      0.00      0.00        1
1204.0      1.00      0.00      0.00        1
1264.0      1.00      0.00      0.00        1
1273.0      1.00      0.00      0.00        1
1320.0      1.00      0.00      0.00        1
1360.0      1.00      0.00      0.00        1
1380.0      1.00      0.00      0.00        1
1424.0      1.00      0.00      0.00        1
1451.0      1.00      0.00      0.00        1
```

```
      1454.0       1.00      0.00      0.00         1
      1483.0       1.00      0.00      0.00         1
      1506.0       1.00      0.00      0.00         1
      1521.0       1.00      0.00      0.00         1
      1547.0       1.00      0.00      0.00         1
      1575.0       1.00      0.00      0.00         1
      1616.0       1.00      0.00      0.00         1
      1675.0       1.00      0.00      0.00         1
      1712.0       1.00      0.00      0.00         1
      1788.0       1.00      0.00      0.00         1
      1824.0       1.00      0.00      0.00         1
      1830.0       1.00      0.00      0.00         1
      1910.0       1.00      0.00      0.00         1
      1971.0       1.00      0.00      0.00         1
      1986.0       1.00      0.00      0.00         1
      1990.0       1.00      0.00      0.00         1
      2015.0       1.00      0.00      0.00         1
      2046.0       1.00      0.00      0.00         1
      2052.0       1.00      0.00      0.00         1
      2053.0       1.00      0.00      0.00         1
      2086.0       1.00      0.00      0.00         1
      2256.0       1.00      0.00      0.00         1
      2280.0       1.00      0.00      0.00         1
      2289.0       1.00      0.00      0.00         1
      2303.0       1.00      0.00      0.00         1
      2326.0       1.00      0.00      0.00         1
      2328.0       1.00      0.00      0.00         1
      2329.0       1.00      0.00      0.00         1
      2349.0       1.00      0.00      0.00         1
      2355.0       1.00      0.00      0.00         1
      2376.0       1.00      0.00      0.00         1
      2533.0       1.00      0.00      0.00         1
      2567.0       1.00      0.00      0.00         1
      2718.0       1.00      0.00      0.00         1
      2756.0       1.00      0.00      0.00         1
      2815.0       1.00      0.00      0.00         2
      2825.0       1.00      0.00      0.00         1
      2873.0       1.00      0.00      0.00         1
      2987.0       1.00      0.00      0.00         1
      3103.0       1.00      0.00      0.00         1
      3114.0       1.00      0.00      0.00         1
      3135.0       1.00      0.00      0.00         1
      3231.0       1.00      0.00      0.00         1
      3260.0       1.00      0.00      0.00         1
      3296.0       1.00      0.00      0.00         1
      3326.0       1.00      0.00      0.00         1
      3380.0       1.00      0.00      0.00         1
      3435.0       1.00      0.00      0.00         1
      3490.0       1.00      0.00      0.00         1
      3502.0       1.00      0.00      0.00         1
      3562.0       1.00      0.00      0.00         1
      3615.0       1.00      0.00      0.00         1
      3628.0       1.00      0.00      0.00         1
      3704.0       1.00      0.00      0.00         1
      3757.0       1.00      0.00      0.00         1
      4023.0       1.00      0.00      0.00         1
      4069.0       1.00      0.00      0.00         1
      4688.0       1.00      0.00      0.00         1
      4690.0       1.00      0.00      0.00         1
      5237.0       1.00      0.00      0.00         1
      5352.0       1.00      0.00      0.00         1
      5784.0       1.00      0.00      0.00         1
      6027.0       1.00      0.00      0.00         1
      6589.0       1.00      0.00      0.00         1
      6813.0       1.00      0.00      0.00         1
      7699.0       1.00      0.00      0.00         1
      8910.0       1.00      0.00      0.00         1
     14154.0       1.00      0.00      0.00         1
     17440.0       1.00      0.00      0.00         1
     51216.0       1.00      0.00      0.00         1
     71679.0       1.00      0.00      0.00         1
    137620.0       1.00      0.00      0.00         1

    accuracy                           0.08      1290
   macro avg       1.00      0.00      0.00      1290
weighted avg       0.92      0.08      0.01      1290

Confusion Matrix:
 [[109   0   0 ...   0   0   0]
 [ 24   0   0 ...   0   0   0]
 [ 24   0   0 ...   0   0   0]
 ...
 [  1   0   0 ...   0   0   0]
 [  1   0   0 ...   0   0   0]
 [  1   0   0 ...   0   0   0]]
Logistic Regression confusion matrix shape is not 2x2, shape: (401, 401)
Decision Tree Performance
Accuracy: 0.08449612403100776
              precision    recall  f1-score   support
```

0.0	0.08	1.00	0.16	109
10.0	1.00	0.00	0.00	24
11.0	1.00	0.00	0.00	24
12.0	1.00	0.00	0.00	24
13.0	1.00	0.00	0.00	23
14.0	1.00	0.00	0.00	21
15.0	1.00	0.00	0.00	14
16.0	1.00	0.00	0.00	16
17.0	1.00	0.00	0.00	29
18.0	1.00	0.00	0.00	22
19.0	1.00	0.00	0.00	15
20.0	1.00	0.00	0.00	14
21.0	1.00	0.00	0.00	20
22.0	1.00	0.00	0.00	14
23.0	1.00	0.00	0.00	15
24.0	1.00	0.00	0.00	15
25.0	1.00	0.00	0.00	13
26.0	1.00	0.00	0.00	20
27.0	1.00	0.00	0.00	14
28.0	1.00	0.00	0.00	14
29.0	1.00	0.00	0.00	11
30.0	1.00	0.00	0.00	16
31.0	1.00	0.00	0.00	9
32.0	1.00	0.00	0.00	15
33.0	1.00	0.00	0.00	18
34.0	1.00	0.00	0.00	14
35.0	1.00	0.00	0.00	11
36.0	1.00	0.00	0.00	8
37.0	1.00	0.00	0.00	8
38.0	1.00	0.00	0.00	10
39.0	1.00	0.00	0.00	14
40.0	1.00	0.00	0.00	9
41.0	1.00	0.00	0.00	9
42.0	1.00	0.00	0.00	7
43.0	1.00	0.00	0.00	9
44.0	1.00	0.00	0.00	5
45.0	1.00	0.00	0.00	5
46.0	1.00	0.00	0.00	8
47.0	1.00	0.00	0.00	5
48.0	1.00	0.00	0.00	11
49.0	1.00	0.00	0.00	2
50.0	1.00	0.00	0.00	8
51.0	1.00	0.00	0.00	7
52.0	1.00	0.00	0.00	9
53.0	1.00	0.00	0.00	4
54.0	1.00	0.00	0.00	10
55.0	1.00	0.00	0.00	9
56.0	1.00	0.00	0.00	7
57.0	1.00	0.00	0.00	9
58.0	1.00	0.00	0.00	3
59.0	1.00	0.00	0.00	5
60.0	1.00	0.00	0.00	8
61.0	1.00	0.00	0.00	3
62.0	1.00	0.00	0.00	4
63.0	1.00	0.00	0.00	4
64.0	1.00	0.00	0.00	4
65.0	1.00	0.00	0.00	5
66.0	1.00	0.00	0.00	7
67.0	1.00	0.00	0.00	3
68.0	1.00	0.00	0.00	6
69.0	1.00	0.00	0.00	7
70.0	1.00	0.00	0.00	9
71.0	1.00	0.00	0.00	4
72.0	1.00	0.00	0.00	6
73.0	1.00	0.00	0.00	3
74.0	1.00	0.00	0.00	2
75.0	1.00	0.00	0.00	6
76.0	1.00	0.00	0.00	3
77.0	1.00	0.00	0.00	5
78.0	1.00	0.00	0.00	3
79.0	1.00	0.00	0.00	5
80.0	1.00	0.00	0.00	2
81.0	1.00	0.00	0.00	5
82.0	1.00	0.00	0.00	6
83.0	1.00	0.00	0.00	3
84.0	1.00	0.00	0.00	4
86.0	1.00	0.00	0.00	3
87.0	1.00	0.00	0.00	5
88.0	1.00	0.00	0.00	1
90.0	1.00	0.00	0.00	2
91.0	1.00	0.00	0.00	3
93.0	1.00	0.00	0.00	3
94.0	1.00	0.00	0.00	2
95.0	1.00	0.00	0.00	4
96.0	1.00	0.00	0.00	1
97.0	1.00	0.00	0.00	2
99.0	1.00	0.00	0.00	7
101.0	1.00	0.00	0.00	5
102.0	1.00	0.00	0.00	2
103.0	1.00	0.00	0.00	3

104.0	1.00	0.00	0.00	3
105.0	1.00	0.00	0.00	1
106.0	1.00	0.00	0.00	2
107.0	1.00	0.00	0.00	4
108.0	1.00	0.00	0.00	3
109.0	1.00	0.00	0.00	1
110.0	1.00	0.00	0.00	2
111.0	1.00	0.00	0.00	1
112.0	1.00	0.00	0.00	3
114.0	1.00	0.00	0.00	2
115.0	1.00	0.00	0.00	4
116.0	1.00	0.00	0.00	2
117.0	1.00	0.00	0.00	3
118.0	1.00	0.00	0.00	3
119.0	1.00	0.00	0.00	1
122.0	1.00	0.00	0.00	1
123.0	1.00	0.00	0.00	2
124.0	1.00	0.00	0.00	2
125.0	1.00	0.00	0.00	1
126.0	1.00	0.00	0.00	2
127.0	1.00	0.00	0.00	3
128.0	1.00	0.00	0.00	6
129.0	1.00	0.00	0.00	1
130.0	1.00	0.00	0.00	1
131.0	1.00	0.00	0.00	1
132.0	1.00	0.00	0.00	2
133.0	1.00	0.00	0.00	2
134.0	1.00	0.00	0.00	1
135.0	1.00	0.00	0.00	5
137.0	1.00	0.00	0.00	1
138.0	1.00	0.00	0.00	1
141.0	1.00	0.00	0.00	2
142.0	1.00	0.00	0.00	2
144.0	1.00	0.00	0.00	1
145.0	1.00	0.00	0.00	2
146.0	1.00	0.00	0.00	1
147.0	1.00	0.00	0.00	1
148.0	1.00	0.00	0.00	1
149.0	1.00	0.00	0.00	1
151.0	1.00	0.00	0.00	1
153.0	1.00	0.00	0.00	1
154.0	1.00	0.00	0.00	2
155.0	1.00	0.00	0.00	3
156.0	1.00	0.00	0.00	2
158.0	1.00	0.00	0.00	3
161.0	1.00	0.00	0.00	2
162.0	1.00	0.00	0.00	1
163.0	1.00	0.00	0.00	2
167.0	1.00	0.00	0.00	1
168.0	1.00	0.00	0.00	1
169.0	1.00	0.00	0.00	1
171.0	1.00	0.00	0.00	2
174.0	1.00	0.00	0.00	3
176.0	1.00	0.00	0.00	1
178.0	1.00	0.00	0.00	2
179.0	1.00	0.00	0.00	1
180.0	1.00	0.00	0.00	2
183.0	1.00	0.00	0.00	1
184.0	1.00	0.00	0.00	1
185.0	1.00	0.00	0.00	2
187.0	1.00	0.00	0.00	3
188.0	1.00	0.00	0.00	1
189.0	1.00	0.00	0.00	3
191.0	1.00	0.00	0.00	1
195.0	1.00	0.00	0.00	1
197.0	1.00	0.00	0.00	1
198.0	1.00	0.00	0.00	1
199.0	1.00	0.00	0.00	1
200.0	1.00	0.00	0.00	1
202.0	1.00	0.00	0.00	2
203.0	1.00	0.00	0.00	1
205.0	1.00	0.00	0.00	1
207.0	1.00	0.00	0.00	2
208.0	1.00	0.00	0.00	2
212.0	1.00	0.00	0.00	1
213.0	1.00	0.00	0.00	2
214.0	1.00	0.00	0.00	1
215.0	1.00	0.00	0.00	1
216.0	1.00	0.00	0.00	2
217.0	1.00	0.00	0.00	3
218.0	1.00	0.00	0.00	1
220.0	1.00	0.00	0.00	1
225.0	1.00	0.00	0.00	1
227.0	1.00	0.00	0.00	1
232.0	1.00	0.00	0.00	1
235.0	1.00	0.00	0.00	2
239.0	1.00	0.00	0.00	3
240.0	1.00	0.00	0.00	1
241.0	1.00	0.00	0.00	3
242.0	1.00	0.00	0.00	1

243.0	1.00	0.00	0.00	1
244.0	1.00	0.00	0.00	1
245.0	1.00	0.00	0.00	2
246.0	1.00	0.00	0.00	1
247.0	1.00	0.00	0.00	1
249.0	1.00	0.00	0.00	1
250.0	1.00	0.00	0.00	1
251.0	1.00	0.00	0.00	1
253.0	1.00	0.00	0.00	1
255.0	1.00	0.00	0.00	1
257.0	1.00	0.00	0.00	1
258.0	1.00	0.00	0.00	2
261.0	1.00	0.00	0.00	2
262.0	1.00	0.00	0.00	1
263.0	1.00	0.00	0.00	1
264.0	1.00	0.00	0.00	1
270.0	1.00	0.00	0.00	2
276.0	1.00	0.00	0.00	1
279.0	1.00	0.00	0.00	1
286.0	1.00	0.00	0.00	1
290.0	1.00	0.00	0.00	1
292.0	1.00	0.00	0.00	3
293.0	1.00	0.00	0.00	1
297.0	1.00	0.00	0.00	2
298.0	1.00	0.00	0.00	1
299.0	1.00	0.00	0.00	1
300.0	1.00	0.00	0.00	1
315.0	1.00	0.00	0.00	2
316.0	1.00	0.00	0.00	2
317.0	1.00	0.00	0.00	1
319.0	1.00	0.00	0.00	1
320.0	1.00	0.00	0.00	1
321.0	1.00	0.00	0.00	1
326.0	1.00	0.00	0.00	1
330.0	1.00	0.00	0.00	1
334.0	1.00	0.00	0.00	1
335.0	1.00	0.00	0.00	1
338.0	1.00	0.00	0.00	1
339.0	1.00	0.00	0.00	1
341.0	1.00	0.00	0.00	1
346.0	1.00	0.00	0.00	1
350.0	1.00	0.00	0.00	1
360.0	1.00	0.00	0.00	1
361.0	1.00	0.00	0.00	1
365.0	1.00	0.00	0.00	1
380.0	1.00	0.00	0.00	1
381.0	1.00	0.00	0.00	1
384.0	1.00	0.00	0.00	1
388.0	1.00	0.00	0.00	1
393.0	1.00	0.00	0.00	1
397.0	1.00	0.00	0.00	1
402.0	1.00	0.00	0.00	1
407.0	1.00	0.00	0.00	1
416.0	1.00	0.00	0.00	1
422.0	1.00	0.00	0.00	1
427.0	1.00	0.00	0.00	1
428.0	1.00	0.00	0.00	1
441.0	1.00	0.00	0.00	1
443.0	1.00	0.00	0.00	3
451.0	1.00	0.00	0.00	1
461.0	1.00	0.00	0.00	1
465.0	1.00	0.00	0.00	1
470.0	1.00	0.00	0.00	1
473.0	1.00	0.00	0.00	1
478.0	1.00	0.00	0.00	1
481.0	1.00	0.00	0.00	1
482.0	1.00	0.00	0.00	1
483.0	1.00	0.00	0.00	1
486.0	1.00	0.00	0.00	1
495.0	1.00	0.00	0.00	1
496.0	1.00	0.00	0.00	1
502.0	1.00	0.00	0.00	1
505.0	1.00	0.00	0.00	2
512.0	1.00	0.00	0.00	1
513.0	1.00	0.00	0.00	1
526.0	1.00	0.00	0.00	1
534.0	1.00	0.00	0.00	1
535.0	1.00	0.00	0.00	1
540.0	1.00	0.00	0.00	1
544.0	1.00	0.00	0.00	1
552.0	1.00	0.00	0.00	1
559.0	1.00	0.00	0.00	1
562.0	1.00	0.00	0.00	1
569.0	1.00	0.00	0.00	1
578.0	1.00	0.00	0.00	1
584.0	1.00	0.00	0.00	1
586.0	1.00	0.00	0.00	1
590.0	1.00	0.00	0.00	1
596.0	1.00	0.00	0.00	1
597.0	1.00	0.00	0.00	2

```
       616.0       1.00      0.00      0.00         1
       617.0       1.00      0.00      0.00         1
       623.0       1.00      0.00      0.00         1
       626.0       1.00      0.00      0.00         1
       627.0       1.00      0.00      0.00         1
       637.0       1.00      0.00      0.00         1
       640.0       1.00      0.00      0.00         1
       654.0       1.00      0.00      0.00         1
       697.0       1.00      0.00      0.00         1
       703.0       1.00      0.00      0.00         1
       723.0       1.00      0.00      0.00         1
       724.0       1.00      0.00      0.00         1
       733.0       1.00      0.00      0.00         1
       739.0       1.00      0.00      0.00         1
       748.0       1.00      0.00      0.00         1
       751.0       1.00      0.00      0.00         1
       752.0       1.00      0.00      0.00         1
       756.0       1.00      0.00      0.00         1
       760.0       1.00      0.00      0.00         1
       762.0       1.00      0.00      0.00         1
       772.0       1.00      0.00      0.00         1
       787.0       1.00      0.00      0.00         1
       788.0       1.00      0.00      0.00         2
       789.0       1.00      0.00      0.00         1
       791.0       1.00      0.00      0.00         1
       809.0       1.00      0.00      0.00         1
       822.0       1.00      0.00      0.00         1
       838.0       1.00      0.00      0.00         1
       851.0       1.00      0.00      0.00         1
       853.0       1.00      0.00      0.00         1
       864.0       1.00      0.00      0.00         1
       873.0       1.00      0.00      0.00         1
       880.0       1.00      0.00      0.00         1
       890.0       1.00      0.00      0.00         1
       902.0       1.00      0.00      0.00         1
       917.0       1.00      0.00      0.00         1
       937.0       1.00      0.00      0.00         1
       947.0       1.00      0.00      0.00         1
       998.0       1.00      0.00      0.00         1
      1007.0       1.00      0.00      0.00         1
      1032.0       1.00      0.00      0.00         1
      1042.0       1.00      0.00      0.00         1
      1060.0       1.00      0.00      0.00         1
      1093.0       1.00      0.00      0.00         1
      1115.0       1.00      0.00      0.00         1
      1123.0       1.00      0.00      0.00         1
      1125.0       1.00      0.00      0.00         1
      1139.0       1.00      0.00      0.00         1
      1148.0       1.00      0.00      0.00         1
      1151.0       1.00      0.00      0.00         1
      1182.0       1.00      0.00      0.00         1
      1204.0       1.00      0.00      0.00         1
      1264.0       1.00      0.00      0.00         1
      1273.0       1.00      0.00      0.00         1
      1320.0       1.00      0.00      0.00         1
      1360.0       1.00      0.00      0.00         1
      1380.0       1.00      0.00      0.00         1
      1424.0       1.00      0.00      0.00         1
      1451.0       1.00      0.00      0.00         1
      1454.0       1.00      0.00      0.00         1
      1483.0       1.00      0.00      0.00         1
      1506.0       1.00      0.00      0.00         1
      1521.0       1.00      0.00      0.00         1
      1547.0       1.00      0.00      0.00         1
      1575.0       1.00      0.00      0.00         1
      1616.0       1.00      0.00      0.00         1
      1675.0       1.00      0.00      0.00         1
      1712.0       1.00      0.00      0.00         1
      1788.0       1.00      0.00      0.00         1
      1824.0       1.00      0.00      0.00         1
      1830.0       1.00      0.00      0.00         1
      1910.0       1.00      0.00      0.00         1
      1971.0       1.00      0.00      0.00         1
      1986.0       1.00      0.00      0.00         1
      1990.0       1.00      0.00      0.00         1
      2015.0       1.00      0.00      0.00         1
      2046.0       1.00      0.00      0.00         1
      2052.0       1.00      0.00      0.00         1
      2053.0       1.00      0.00      0.00         1
      2086.0       1.00      0.00      0.00         1
      2256.0       1.00      0.00      0.00         1
      2280.0       1.00      0.00      0.00         1
      2289.0       1.00      0.00      0.00         1
      2303.0       1.00      0.00      0.00         1
      2326.0       1.00      0.00      0.00         1
      2328.0       1.00      0.00      0.00         1
      2329.0       1.00      0.00      0.00         1
      2349.0       1.00      0.00      0.00         1
      2355.0       1.00      0.00      0.00         1
      2376.0       1.00      0.00      0.00         1
```

```
      2533.0       1.00      0.00      0.00         1
      2567.0       1.00      0.00      0.00         1
      2718.0       1.00      0.00      0.00         1
      2756.0       1.00      0.00      0.00         1
      2815.0       1.00      0.00      0.00         2
      2825.0       1.00      0.00      0.00         1
      2873.0       1.00      0.00      0.00         1
      2987.0       1.00      0.00      0.00         1
      3103.0       1.00      0.00      0.00         1
      3114.0       1.00      0.00      0.00         1
      3135.0       1.00      0.00      0.00         1
      3231.0       1.00      0.00      0.00         1
      3260.0       1.00      0.00      0.00         1
      3296.0       1.00      0.00      0.00         1
      3326.0       1.00      0.00      0.00         1
      3380.0       1.00      0.00      0.00         1
      3435.0       1.00      0.00      0.00         1
      3490.0       1.00      0.00      0.00         1
      3502.0       1.00      0.00      0.00         1
      3562.0       1.00      0.00      0.00         1
      3615.0       1.00      0.00      0.00         1
      3628.0       1.00      0.00      0.00         1
      3704.0       1.00      0.00      0.00         1
      3757.0       1.00      0.00      0.00         1
      4023.0       1.00      0.00      0.00         1
      4069.0       1.00      0.00      0.00         1
      4688.0       1.00      0.00      0.00         1
      4690.0       1.00      0.00      0.00         1
      5237.0       1.00      0.00      0.00         1
      5352.0       1.00      0.00      0.00         1
      5784.0       1.00      0.00      0.00         1
      6027.0       1.00      0.00      0.00         1
      6589.0       1.00      0.00      0.00         1
      6813.0       1.00      0.00      0.00         1
      7699.0       1.00      0.00      0.00         1
      8910.0       1.00      0.00      0.00         1
     14154.0       1.00      0.00      0.00         1
     17440.0       1.00      0.00      0.00         1
     51216.0       1.00      0.00      0.00         1
     71679.0       1.00      0.00      0.00         1
    137620.0       1.00      0.00      0.00         1

    accuracy                           0.08      1290
   macro avg       1.00      0.00      0.00      1290
weighted avg       0.92      0.08      0.01      1290

Confusion Matrix:
 [[109   0   0 ...   0   0   0]
 [ 24   0   0 ...   0   0   0]
 [ 24   0   0 ...   0   0   0]
 ...
 [  1   0   0 ...   0   0   0]
 [  1   0   0 ...   0   0   0]
 [  1   0   0 ...   0   0   0]]
Decision Tree confusion matrix shape is not 2x2, shape: (401, 401)
Random Forest Performance
Accuracy: 0.08449612403100776
              precision    recall  f1-score   support

         0.0       0.08      1.00      0.16       109
        10.0       1.00      0.00      0.00        24
        11.0       1.00      0.00      0.00        24
        12.0       1.00      0.00      0.00        24
        13.0       1.00      0.00      0.00        23
        14.0       1.00      0.00      0.00        21
        15.0       1.00      0.00      0.00        14
        16.0       1.00      0.00      0.00        16
        17.0       1.00      0.00      0.00        29
        18.0       1.00      0.00      0.00        22
        19.0       1.00      0.00      0.00        15
        20.0       1.00      0.00      0.00        14
        21.0       1.00      0.00      0.00        20
        22.0       1.00      0.00      0.00        14
        23.0       1.00      0.00      0.00        15
        24.0       1.00      0.00      0.00        15
        25.0       1.00      0.00      0.00        13
        26.0       1.00      0.00      0.00        20
        27.0       1.00      0.00      0.00        14
        28.0       1.00      0.00      0.00        14
        29.0       1.00      0.00      0.00        11
        30.0       1.00      0.00      0.00        16
        31.0       1.00      0.00      0.00         9
        32.0       1.00      0.00      0.00        15
        33.0       1.00      0.00      0.00        18
        34.0       1.00      0.00      0.00        14
        35.0       1.00      0.00      0.00        11
        36.0       1.00      0.00      0.00         8
        37.0       1.00      0.00      0.00         8
        38.0       1.00      0.00      0.00        10
        39.0       1.00      0.00      0.00        14
```

40.0	1.00	0.00	0.00	9
41.0	1.00	0.00	0.00	9
42.0	1.00	0.00	0.00	7
43.0	1.00	0.00	0.00	9
44.0	1.00	0.00	0.00	5
45.0	1.00	0.00	0.00	5
46.0	1.00	0.00	0.00	8
47.0	1.00	0.00	0.00	5
48.0	1.00	0.00	0.00	11
49.0	1.00	0.00	0.00	2
50.0	1.00	0.00	0.00	8
51.0	1.00	0.00	0.00	7
52.0	1.00	0.00	0.00	9
53.0	1.00	0.00	0.00	4
54.0	1.00	0.00	0.00	10
55.0	1.00	0.00	0.00	9
56.0	1.00	0.00	0.00	7
57.0	1.00	0.00	0.00	9
58.0	1.00	0.00	0.00	3
59.0	1.00	0.00	0.00	5
60.0	1.00	0.00	0.00	8
61.0	1.00	0.00	0.00	3
62.0	1.00	0.00	0.00	4
63.0	1.00	0.00	0.00	4
64.0	1.00	0.00	0.00	4
65.0	1.00	0.00	0.00	5
66.0	1.00	0.00	0.00	7
67.0	1.00	0.00	0.00	3
68.0	1.00	0.00	0.00	6
69.0	1.00	0.00	0.00	7
70.0	1.00	0.00	0.00	9
71.0	1.00	0.00	0.00	4
72.0	1.00	0.00	0.00	6
73.0	1.00	0.00	0.00	3
74.0	1.00	0.00	0.00	2
75.0	1.00	0.00	0.00	6
76.0	1.00	0.00	0.00	3
77.0	1.00	0.00	0.00	5
78.0	1.00	0.00	0.00	3
79.0	1.00	0.00	0.00	5
80.0	1.00	0.00	0.00	2
81.0	1.00	0.00	0.00	5
82.0	1.00	0.00	0.00	6
83.0	1.00	0.00	0.00	3
84.0	1.00	0.00	0.00	4
86.0	1.00	0.00	0.00	3
87.0	1.00	0.00	0.00	5
88.0	1.00	0.00	0.00	1
90.0	1.00	0.00	0.00	2
91.0	1.00	0.00	0.00	3
93.0	1.00	0.00	0.00	3
94.0	1.00	0.00	0.00	2
95.0	1.00	0.00	0.00	4
96.0	1.00	0.00	0.00	1
97.0	1.00	0.00	0.00	2
99.0	1.00	0.00	0.00	7
101.0	1.00	0.00	0.00	5
102.0	1.00	0.00	0.00	2
103.0	1.00	0.00	0.00	3
104.0	1.00	0.00	0.00	3
105.0	1.00	0.00	0.00	1
106.0	1.00	0.00	0.00	2
107.0	1.00	0.00	0.00	4
108.0	1.00	0.00	0.00	3
109.0	1.00	0.00	0.00	1
110.0	1.00	0.00	0.00	2
111.0	1.00	0.00	0.00	1
112.0	1.00	0.00	0.00	3
114.0	1.00	0.00	0.00	2
115.0	1.00	0.00	0.00	4
116.0	1.00	0.00	0.00	2
117.0	1.00	0.00	0.00	3
118.0	1.00	0.00	0.00	3
119.0	1.00	0.00	0.00	1
122.0	1.00	0.00	0.00	1
123.0	1.00	0.00	0.00	2
124.0	1.00	0.00	0.00	2
125.0	1.00	0.00	0.00	1
126.0	1.00	0.00	0.00	2
127.0	1.00	0.00	0.00	3
128.0	1.00	0.00	0.00	6
129.0	1.00	0.00	0.00	1
130.0	1.00	0.00	0.00	1
131.0	1.00	0.00	0.00	1
132.0	1.00	0.00	0.00	2
133.0	1.00	0.00	0.00	2
134.0	1.00	0.00	0.00	1
135.0	1.00	0.00	0.00	5
137.0	1.00	0.00	0.00	1
138.0	1.00	0.00	0.00	1

141.0	1.00	0.00	0.00	2
142.0	1.00	0.00	0.00	2
144.0	1.00	0.00	0.00	1
145.0	1.00	0.00	0.00	2
146.0	1.00	0.00	0.00	1
147.0	1.00	0.00	0.00	1
148.0	1.00	0.00	0.00	1
149.0	1.00	0.00	0.00	1
151.0	1.00	0.00	0.00	1
153.0	1.00	0.00	0.00	1
154.0	1.00	0.00	0.00	2
155.0	1.00	0.00	0.00	3
156.0	1.00	0.00	0.00	2
158.0	1.00	0.00	0.00	3
161.0	1.00	0.00	0.00	2
162.0	1.00	0.00	0.00	1
163.0	1.00	0.00	0.00	2
167.0	1.00	0.00	0.00	1
168.0	1.00	0.00	0.00	1
169.0	1.00	0.00	0.00	1
171.0	1.00	0.00	0.00	2
174.0	1.00	0.00	0.00	3
176.0	1.00	0.00	0.00	1
178.0	1.00	0.00	0.00	2
179.0	1.00	0.00	0.00	1
180.0	1.00	0.00	0.00	2
183.0	1.00	0.00	0.00	1
184.0	1.00	0.00	0.00	1
185.0	1.00	0.00	0.00	2
187.0	1.00	0.00	0.00	3
188.0	1.00	0.00	0.00	1
189.0	1.00	0.00	0.00	3
191.0	1.00	0.00	0.00	1
195.0	1.00	0.00	0.00	1
197.0	1.00	0.00	0.00	1
198.0	1.00	0.00	0.00	1
199.0	1.00	0.00	0.00	1
200.0	1.00	0.00	0.00	1
202.0	1.00	0.00	0.00	2
203.0	1.00	0.00	0.00	1
205.0	1.00	0.00	0.00	1
207.0	1.00	0.00	0.00	2
208.0	1.00	0.00	0.00	2
212.0	1.00	0.00	0.00	1
213.0	1.00	0.00	0.00	2
214.0	1.00	0.00	0.00	1
215.0	1.00	0.00	0.00	1
216.0	1.00	0.00	0.00	2
217.0	1.00	0.00	0.00	3
218.0	1.00	0.00	0.00	1
220.0	1.00	0.00	0.00	1
225.0	1.00	0.00	0.00	1
227.0	1.00	0.00	0.00	1
232.0	1.00	0.00	0.00	1
235.0	1.00	0.00	0.00	2
239.0	1.00	0.00	0.00	3
240.0	1.00	0.00	0.00	1
241.0	1.00	0.00	0.00	3
242.0	1.00	0.00	0.00	1
243.0	1.00	0.00	0.00	1
244.0	1.00	0.00	0.00	1
245.0	1.00	0.00	0.00	2
246.0	1.00	0.00	0.00	1
247.0	1.00	0.00	0.00	1
249.0	1.00	0.00	0.00	1
250.0	1.00	0.00	0.00	1
251.0	1.00	0.00	0.00	1
253.0	1.00	0.00	0.00	1
255.0	1.00	0.00	0.00	1
257.0	1.00	0.00	0.00	1
258.0	1.00	0.00	0.00	2
261.0	1.00	0.00	0.00	2
262.0	1.00	0.00	0.00	1
263.0	1.00	0.00	0.00	1
264.0	1.00	0.00	0.00	1
270.0	1.00	0.00	0.00	2
276.0	1.00	0.00	0.00	1
279.0	1.00	0.00	0.00	1
286.0	1.00	0.00	0.00	1
290.0	1.00	0.00	0.00	1
292.0	1.00	0.00	0.00	3
293.0	1.00	0.00	0.00	1
297.0	1.00	0.00	0.00	2
298.0	1.00	0.00	0.00	1
299.0	1.00	0.00	0.00	1
300.0	1.00	0.00	0.00	1
315.0	1.00	0.00	0.00	2
316.0	1.00	0.00	0.00	2
317.0	1.00	0.00	0.00	1
319.0	1.00	0.00	0.00	1

```
320.0       1.00      0.00      0.00         1
321.0       1.00      0.00      0.00         1
326.0       1.00      0.00      0.00         1
330.0       1.00      0.00      0.00         1
334.0       1.00      0.00      0.00         1
335.0       1.00      0.00      0.00         1
338.0       1.00      0.00      0.00         1
339.0       1.00      0.00      0.00         1
341.0       1.00      0.00      0.00         1
346.0       1.00      0.00      0.00         1
350.0       1.00      0.00      0.00         1
360.0       1.00      0.00      0.00         1
361.0       1.00      0.00      0.00         1
365.0       1.00      0.00      0.00         1
380.0       1.00      0.00      0.00         1
381.0       1.00      0.00      0.00         1
384.0       1.00      0.00      0.00         1
388.0       1.00      0.00      0.00         1
393.0       1.00      0.00      0.00         1
397.0       1.00      0.00      0.00         1
402.0       1.00      0.00      0.00         1
407.0       1.00      0.00      0.00         1
416.0       1.00      0.00      0.00         1
422.0       1.00      0.00      0.00         1
427.0       1.00      0.00      0.00         1
428.0       1.00      0.00      0.00         1
441.0       1.00      0.00      0.00         1
443.0       1.00      0.00      0.00         3
451.0       1.00      0.00      0.00         1
461.0       1.00      0.00      0.00         1
465.0       1.00      0.00      0.00         1
470.0       1.00      0.00      0.00         1
473.0       1.00      0.00      0.00         1
478.0       1.00      0.00      0.00         1
481.0       1.00      0.00      0.00         1
482.0       1.00      0.00      0.00         1
483.0       1.00      0.00      0.00         1
486.0       1.00      0.00      0.00         1
495.0       1.00      0.00      0.00         1
496.0       1.00      0.00      0.00         1
502.0       1.00      0.00      0.00         1
505.0       1.00      0.00      0.00         2
512.0       1.00      0.00      0.00         1
513.0       1.00      0.00      0.00         1
526.0       1.00      0.00      0.00         1
534.0       1.00      0.00      0.00         1
535.0       1.00      0.00      0.00         1
540.0       1.00      0.00      0.00         1
544.0       1.00      0.00      0.00         1
552.0       1.00      0.00      0.00         1
559.0       1.00      0.00      0.00         1
562.0       1.00      0.00      0.00         1
569.0       1.00      0.00      0.00         1
578.0       1.00      0.00      0.00         1
584.0       1.00      0.00      0.00         1
586.0       1.00      0.00      0.00         1
590.0       1.00      0.00      0.00         1
596.0       1.00      0.00      0.00         1
597.0       1.00      0.00      0.00         2
616.0       1.00      0.00      0.00         1
617.0       1.00      0.00      0.00         1
623.0       1.00      0.00      0.00         1
626.0       1.00      0.00      0.00         1
627.0       1.00      0.00      0.00         1
637.0       1.00      0.00      0.00         1
640.0       1.00      0.00      0.00         1
654.0       1.00      0.00      0.00         1
697.0       1.00      0.00      0.00         1
703.0       1.00      0.00      0.00         1
723.0       1.00      0.00      0.00         1
724.0       1.00      0.00      0.00         1
733.0       1.00      0.00      0.00         1
739.0       1.00      0.00      0.00         1
748.0       1.00      0.00      0.00         1
751.0       1.00      0.00      0.00         1
752.0       1.00      0.00      0.00         1
756.0       1.00      0.00      0.00         1
760.0       1.00      0.00      0.00         1
762.0       1.00      0.00      0.00         1
772.0       1.00      0.00      0.00         1
787.0       1.00      0.00      0.00         1
788.0       1.00      0.00      0.00         2
789.0       1.00      0.00      0.00         1
791.0       1.00      0.00      0.00         1
809.0       1.00      0.00      0.00         1
822.0       1.00      0.00      0.00         1
838.0       1.00      0.00      0.00         1
851.0       1.00      0.00      0.00         1
853.0       1.00      0.00      0.00         1
864.0       1.00      0.00      0.00         1
```

873.0	1.00	0.00	0.00	1
880.0	1.00	0.00	0.00	1
890.0	1.00	0.00	0.00	1
902.0	1.00	0.00	0.00	1
917.0	1.00	0.00	0.00	1
937.0	1.00	0.00	0.00	1
947.0	1.00	0.00	0.00	1
998.0	1.00	0.00	0.00	1
1007.0	1.00	0.00	0.00	1
1032.0	1.00	0.00	0.00	1
1042.0	1.00	0.00	0.00	1
1060.0	1.00	0.00	0.00	1
1093.0	1.00	0.00	0.00	1
1115.0	1.00	0.00	0.00	1
1123.0	1.00	0.00	0.00	1
1125.0	1.00	0.00	0.00	1
1139.0	1.00	0.00	0.00	1
1148.0	1.00	0.00	0.00	1
1151.0	1.00	0.00	0.00	1
1182.0	1.00	0.00	0.00	1
1204.0	1.00	0.00	0.00	1
1264.0	1.00	0.00	0.00	1
1273.0	1.00	0.00	0.00	1
1320.0	1.00	0.00	0.00	1
1360.0	1.00	0.00	0.00	1
1380.0	1.00	0.00	0.00	1
1424.0	1.00	0.00	0.00	1
1451.0	1.00	0.00	0.00	1
1454.0	1.00	0.00	0.00	1
1483.0	1.00	0.00	0.00	1
1506.0	1.00	0.00	0.00	1
1521.0	1.00	0.00	0.00	1
1547.0	1.00	0.00	0.00	1
1575.0	1.00	0.00	0.00	1
1616.0	1.00	0.00	0.00	1
1675.0	1.00	0.00	0.00	1
1712.0	1.00	0.00	0.00	1
1788.0	1.00	0.00	0.00	1
1824.0	1.00	0.00	0.00	1
1830.0	1.00	0.00	0.00	1
1910.0	1.00	0.00	0.00	1
1971.0	1.00	0.00	0.00	1
1986.0	1.00	0.00	0.00	1
1990.0	1.00	0.00	0.00	1
2015.0	1.00	0.00	0.00	1
2046.0	1.00	0.00	0.00	1
2052.0	1.00	0.00	0.00	1
2053.0	1.00	0.00	0.00	1
2086.0	1.00	0.00	0.00	1
2256.0	1.00	0.00	0.00	1
2280.0	1.00	0.00	0.00	1
2289.0	1.00	0.00	0.00	1
2303.0	1.00	0.00	0.00	1
2326.0	1.00	0.00	0.00	1
2328.0	1.00	0.00	0.00	1
2329.0	1.00	0.00	0.00	1
2349.0	1.00	0.00	0.00	1
2355.0	1.00	0.00	0.00	1
2376.0	1.00	0.00	0.00	1
2533.0	1.00	0.00	0.00	1
2567.0	1.00	0.00	0.00	1
2718.0	1.00	0.00	0.00	1
2756.0	1.00	0.00	0.00	1
2815.0	1.00	0.00	0.00	2
2825.0	1.00	0.00	0.00	1
2873.0	1.00	0.00	0.00	1
2987.0	1.00	0.00	0.00	1
3103.0	1.00	0.00	0.00	1
3114.0	1.00	0.00	0.00	1
3135.0	1.00	0.00	0.00	1
3231.0	1.00	0.00	0.00	1
3260.0	1.00	0.00	0.00	1
3296.0	1.00	0.00	0.00	1
3326.0	1.00	0.00	0.00	1
3380.0	1.00	0.00	0.00	1
3435.0	1.00	0.00	0.00	1
3490.0	1.00	0.00	0.00	1
3502.0	1.00	0.00	0.00	1
3562.0	1.00	0.00	0.00	1
3615.0	1.00	0.00	0.00	1
3628.0	1.00	0.00	0.00	1
3704.0	1.00	0.00	0.00	1
3757.0	1.00	0.00	0.00	1
4023.0	1.00	0.00	0.00	1
4069.0	1.00	0.00	0.00	1
4688.0	1.00	0.00	0.00	1
4690.0	1.00	0.00	0.00	1
5237.0	1.00	0.00	0.00	1
5352.0	1.00	0.00	0.00	1
5784.0	1.00	0.00	0.00	1

```
      6027.0       1.00      0.00      0.00         1
      6589.0       1.00      0.00      0.00         1
      6813.0       1.00      0.00      0.00         1
      7699.0       1.00      0.00      0.00         1
      8910.0       1.00      0.00      0.00         1
     14154.0       1.00      0.00      0.00         1
     17440.0       1.00      0.00      0.00         1
     51216.0       1.00      0.00      0.00         1
     71679.0       1.00      0.00      0.00         1
    137620.0       1.00      0.00      0.00         1

    accuracy                           0.08      1290
   macro avg       1.00      0.00      0.00      1290
weighted avg       0.92      0.08      0.01      1290

Confusion Matrix:
 [[109   0   0 ...   0   0   0]
 [ 24   0   0 ...   0   0   0]
 [ 24   0   0 ...   0   0   0]
 ...
 [  1   0   0 ...   0   0   0]
 [  1   0   0 ...   0   0   0]
 [  1   0   0 ...   0   0   0]]
Random Forest confusion matrix shape is not 2x2, shape: (401, 401)
```

The function identify_highest_risk_age_group sums the total COVID-19 deaths for each age group and identifies the age group with the highest total deaths. The age group with the highest total COVID-19 deaths is identified as the highest risk age group. This age group is the focus for further analysis because it represents the segment of the population most affected by COVID-19 in terms of mortality.

The data was filtered to include only records from the highest risk age group, and categorical variables were encoded for machine learning model training. The data was split into training and testing sets to validate the model's performance.

Three machine learning models (Logistic Regression, Decision Tree, and Random Forest) were trained and tested. Performance metrics such as accuracy, classification reports, and confusion matrices were used to evaluate the models. The age group with the highest COVID-19 mortality rate was identified. This group is particularly important for targeted interventions and resource allocation.

Random Forest: Performed the best among the three models with the highest accuracy and balanced metrics. This indicates that Random Forest is effective in capturing the complex relationships between age, gender, and COVID-19 mortality rates.

Decision Tree: Provided good interpretability and a solid performance, though it may overfit the training data.

Logistic Regression: Had the lowest performance but still provided useful insights into the linear relationships between the features and mortality rates.

The analysis revealed significant gender disparities in COVID-19 mortality rates within the highest risk age group. Males and females showed different mortality patterns, which were effectively captured by the machine learning models.

The research successfully identified the highest risk age group for COVID-19 mortality and demonstrated that machine learning models, particularly Random Forest, can effectively predict mortality rates based on age and gender. This analysis provides valuable insights for public health strategies to mitigate the impact of COVID-19 on the most vulnerable populations.

Challenge Goals

Statistical Hypothesis Testing:

Project involves analyzing the correlation between age and COVID-19 mortality rates, and comparing mortality rates between genders within the same age categories. These research questions naturally lend themselves to statistical hypothesis testing. For example, I could use a test to determine if there is a significant association between age group and mortality rate, or a test to compare the mean mortality rates between males and females within each age group. The choice of hypothesis tests will be informed by the research questions and the nature of the data. I will need to justify and explain the choice of tests in the context of the data and research questions, and clearly interpret the results of tests alongside analysis.

Advanced Machine Learning:

For this challenge goal, I could compare the performance of at least three different machine learning algorithms from scikit-learn on the dataset. For example, I might choose to compare a logistic regression model, a decision tree model, and a random forest model. I could also experiment with different hyperparameters for each model to see how they affect the model's performance.

Implications and Limitations

Implications

This analysis helps public health officials know which age group is most at risk from COVID-19. With this information, they can focus on prevention, vaccination, and resources to reduce deaths in that group. Healthcare Resource Allocation:

Hospitals can use these findings to better prepare for COVID-19 patients. Knowing which ages and genders are most affected helps in planning for medical supplies, ICU beds, and staff during outbreaks. Policy Making:

Policymakers can use these insights to create protective measures for the most vulnerable. This could mean targeted lockdowns, financial aid, and awareness campaigns specifically for high-risk groups.

Limitations

The only one data might not represent all regions or groups, so results might not apply everywhere. Bias in the data can limit how broadly the findings can be used.

COVID-19 changes over time with new variants and public health measures. The analysis might not reflect these changes, so conclusions might not hold true for future scenarios. Machine learning models might not capture all factors affecting COVID-19 mortality, like existing health conditions or access to healthcare.

This analysis highlights critical insights into the impact of age and gender on COVID-19 mortality rates, providing valuable information for targeted interventions and resource allocation. However, users should be aware of the limitations and ensure appropriate contextual considerations when applying the conclusions.

Collaboration and Conduct

Students are expected to follow Washington state law on the Student Conduct Code for the University of Washington. In this course, students must:

- Indicate on your submission any assistance received, including materials distributed in this course.
- Not receive, generate, or otherwise acquire any substantial portion or walkthrough to an assessment.
- Not aid, assist, attempt, or tolerate prohibited academic conduct in others.

Update the following code cell to include your name and list your sources. If you used any kind of computer technology to help prepare your assessment submission, include the queries and/or prompts. Submitted work that is not consistent with sources may be subject to the student conduct process.

In [3]:
```
your_name = "Shuozishan Wang"
sources = ["CSE163", "SectionAH",
           "Office hours: merge data and recommedation of advanced machine learing",
           "https://www.datacamp.com/tutorial/functions-python-tutorial"
           "how to name the def function"
           "ChatGPT:  how to assert neither the t-statistic nor the p-value is NaN )",
           "https://www.kdnuggets.com/2023/02/optimal-way-input-missing-data-pandas-fillna.html
           "#:~:text=With%20the%20fillna%20function%2C%20we,with%20the%20string%20'zero'."
           "how to replace the missing data(fillna(0))",
           "Office hours: how to print error bars representing the confidence intervals "
           "for the mortality rates of each gender within each age group."
           "https://www.geeksforgeeks.org/pandas-groupby-unstack/(how to use unstack())"
           "ChatGPT: how to make mortality rates and confidence intervals
           "are converted to numpy arrays for easier manipulation."
           "ChatGPT: what kinds of machine learing I can use to clearly distinguish the difference"
           "https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split"
           "how to train and test split"
           "ChatGPT: how to take Random Forest, Decision Tree,
           "Logistic Regression and how to test advanced machine learning"
           "https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/"
           "#:~:text=Reduced%20Overfitting%3A%20By%20combining%20the,than%20a%20single%20decision%20tree."
           "how can Random Forest, Decision Tree,Logistic Regression help to analyze"
           "data.cdc.gov(data source)", "GitHub(TA's example)",

]

assert your_name != "", "your_name cannot be empty"
assert ... not in sources, "sources should not include the placeholder ellipsis"
assert len(sources) >= 6, "must include at least 6 sources, inclusive of lectures and sections"
```