

Assignment -1

Name: Yeswanth Vallabhaneni

Student -ID: 16359049

Question 1:

Data Collection:

- Gathered the data provided in the table and converted this excel file into .csv file.

```
import pandas as pd

#load the dataset
load_data = pd.read_csv('rawData.csv')

[13] load_data.head(10)
```

	Height	Weight	Age	Grip strength	Frailty
0	65.8	112	30	30	N
1	71.5	136	19	31	N
2	69.4	153	45	29	N
3	68.2	142	22	28	Y
4	67.8	144	29	24	Y
5	68.7	123	50	26	N
6	69.8	141	51	22	Y
7	70.1	136	23	20	Y
8	67.9	112	17	19	N
9	66.8	120	39	31	N

Data Preprocessing:

- Check if there are any missing in the dataset.

```
Data Preprocessing

#Check for missing values
load_data.isnull().sum()
```

	0
Height	0
Weight	0
Age	0
Grip strength	0
Frailty	0

dtype: int64

- convert categorical variables into numerical.
- The 'Frailty' variable is converted from categorical (Y/N) to numerical (1/0).

```
#Convert 'Frailty' categorical variable to numerical
load_data['Frailty'] = load_data['Frailty'].replace({'Y': 1, 'N': 0})
load_data.head(10)
```

	Height	Weight	Age	Grip strength	Frailty
0	65.8	112	30	30	0
1	71.5	136	19	31	0
2	69.4	153	45	29	0
3	68.2	142	22	28	1
4	67.8	144	29	24	1
5	68.7	123	50	26	0
6	69.8	141	51	22	1
7	70.1	136	23	20	1
8	67.9	112	17	19	0
9	66.8	120	39	31	0

-The cleaned dataset is saved.

```
[17] load_data.to_csv('cleandata.csv', index=False)
```

Data Analysis:

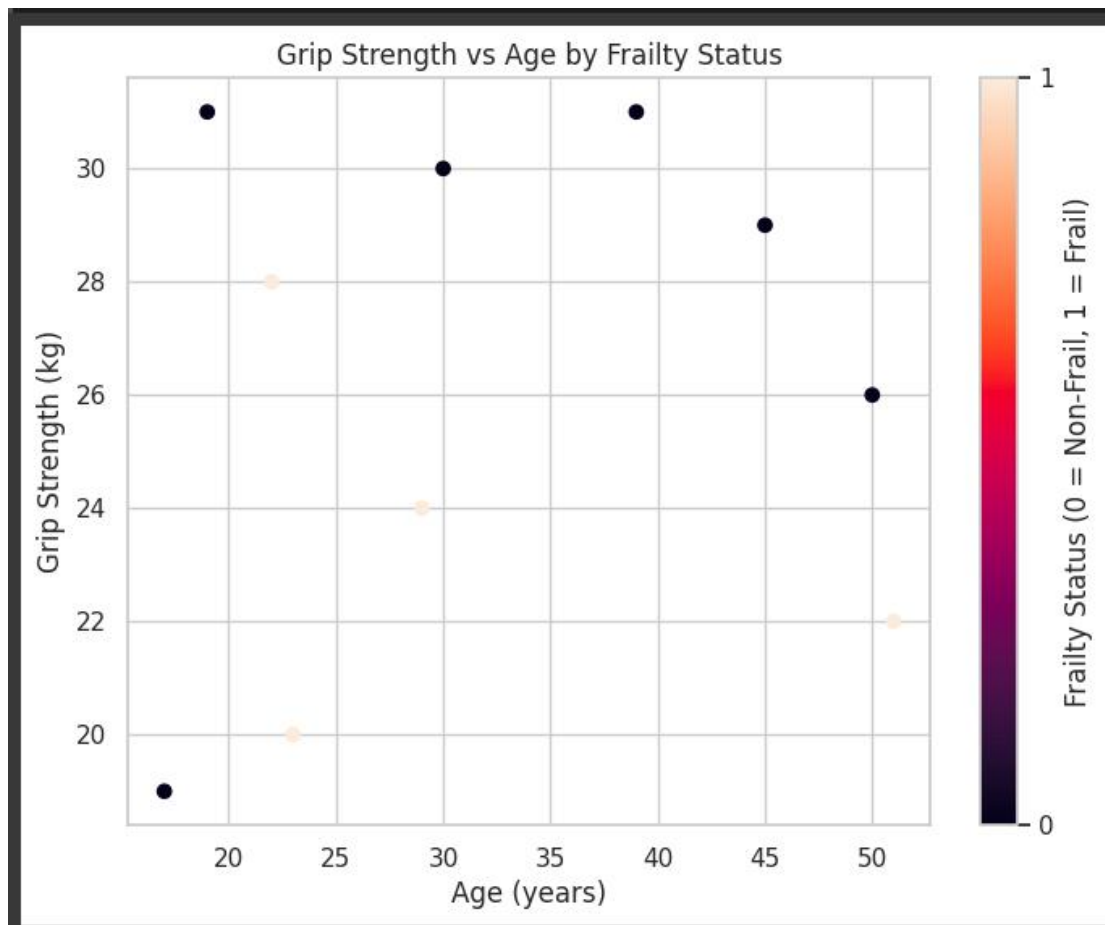
A T-Test is performed to compare the grip strength between frail and non-frail participants, helping to determine if there is a statistically significant difference in grip strength based on frailty status.

```
from scipy.stats import ttest_ind
# Separate Grip Strength based on Frailty status
frail_grip_strength = load_data[load_data['Frailty'] == 1]['Grip strength']
non_frail_grip_strength = load_data[load_data['Frailty'] == 0]['Grip strength']
grip_strength_ttest = ttest_ind(frail_grip_strength, non_frail_grip_strength, equal_var=False)
print("T-Test results for Grip Strength between Frail and Non-Frail Participants:")

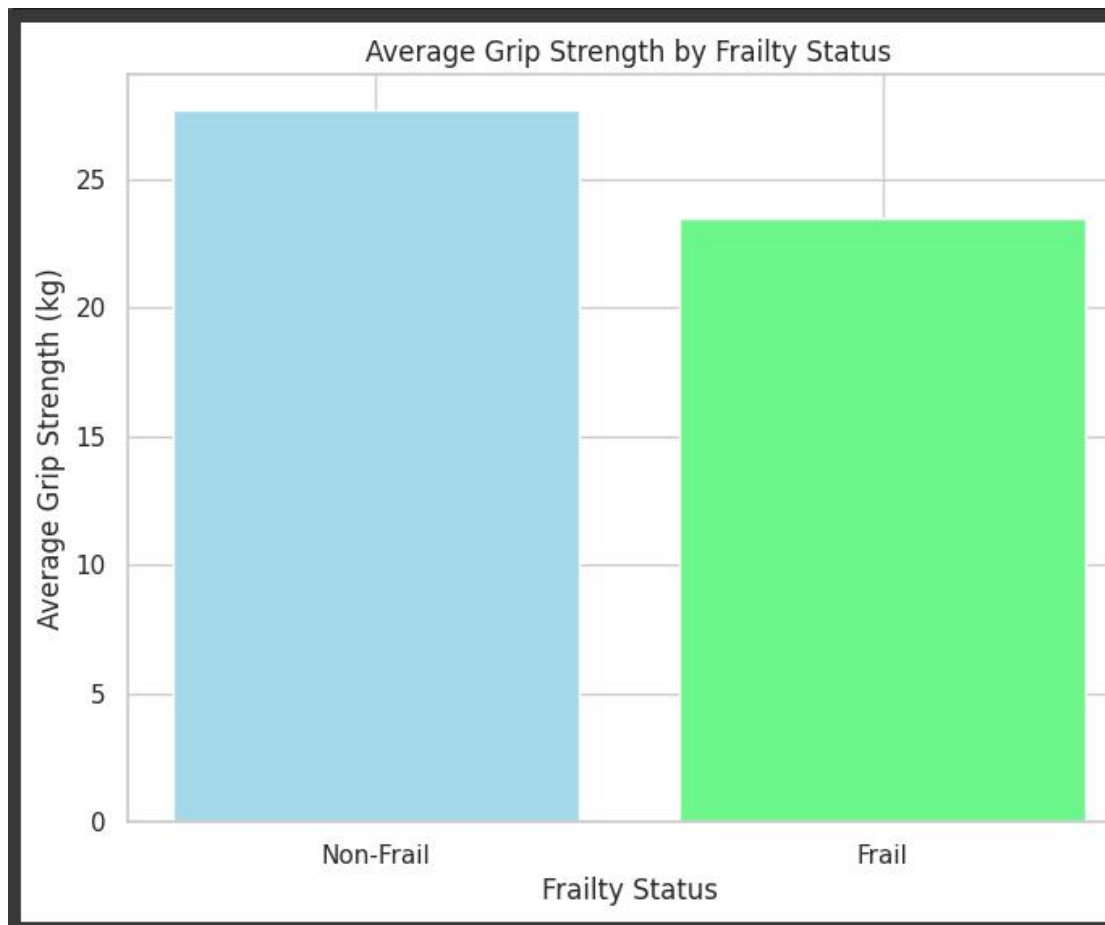
print("T-Statistic:", grip_strength_ttest.statistic)
print("P-Value:", grip_strength_ttest.pvalue)
```

T-Test results for Grip Strength between Frail and Non-Frail Participants:
T-Statistic: -1.634999993460006
P-Value: 0.1415730416628566

- Created a Scatter Plot of Grip Strength vs Age by Frailty



Bar Plot of Average Grip Strength by Frailty



Question 2:

Data collection:

```
# Load the dataset
load_data = pd.read_csv('StudentsPerformance.csv')
load_data.head(10)
```

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	72	72	74
1	female	group C	some college	standard	completed	69	90	88
2	female	group B	master's degree	standard	none	90	95	93
3	male	group A	associate's degree	free/reduced	none	47	57	44
4	male	group C	some college	standard	none	76	78	75
5	female	group B	associate's degree	standard	none	71	83	78
6	female	group B	some college	standard	completed	88	95	92
7	male	group B	some college	free/reduced	none	40	43	39
8	male	group D	high school	free/reduced	completed	64	64	67
9	female	group B	high school	free/reduced	none	38	60	50

Next steps: [Generate code with load_data](#) [View recommended plots](#) [New interactive sheet](#)

Data Preprocessing:

```
✓ 0s Data Preprocessing

#Check for missing values
print(load_data.isnull().sum())

gender          0
race/ethnicity  0
parental level of education  0
lunch           0
test preparation course  0
math score      0
reading score   0
writing score   0
dtype: int64

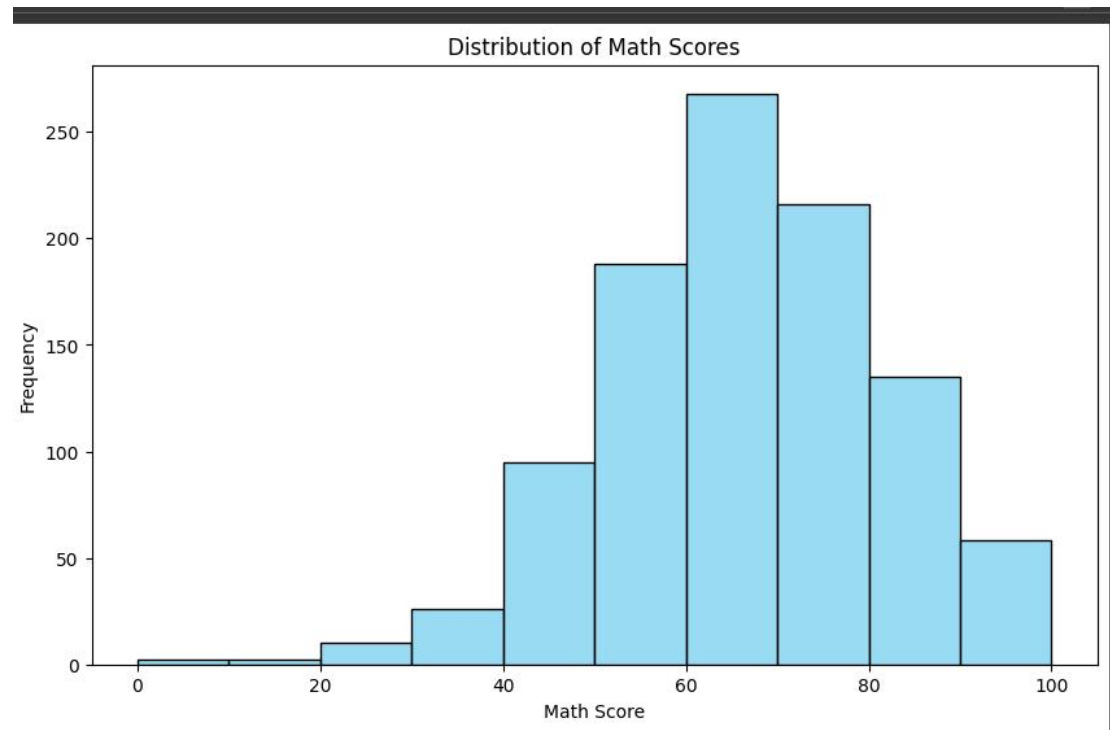
Data Visualization

Histogram Representation of Distribution of Math Scores

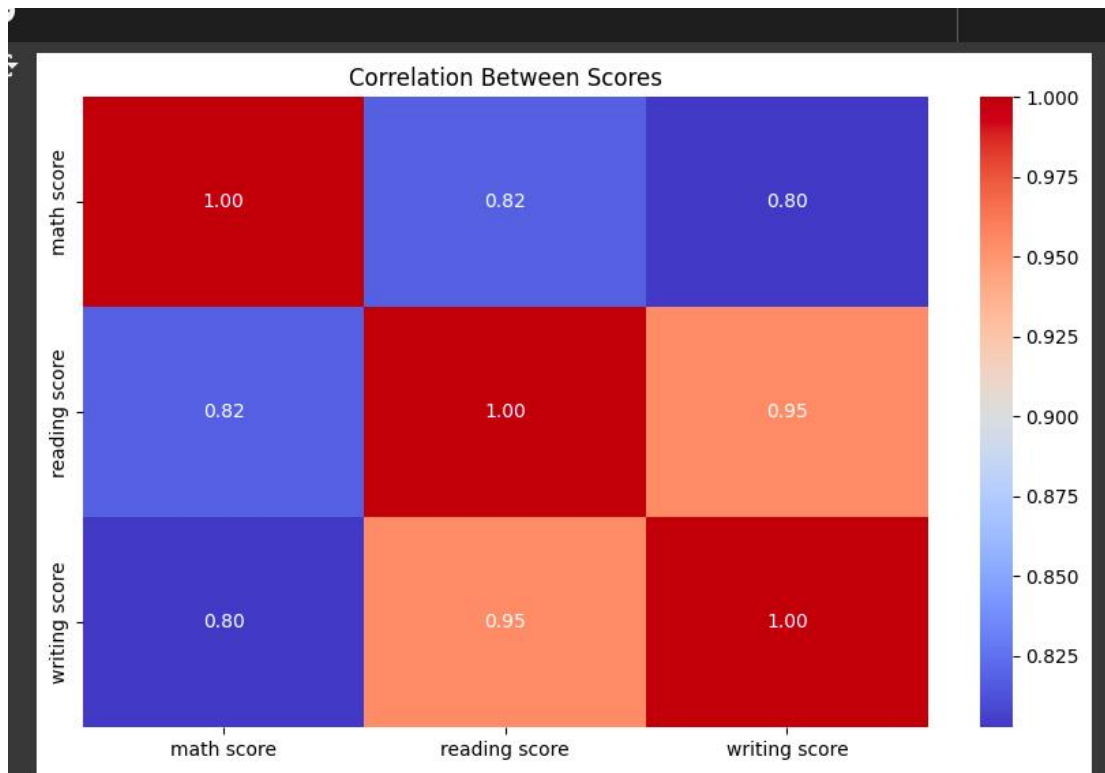
[38] # Visualization 1: Distribution of Math Scores (Histogram)
```

Data Visualization

1.Histogram Representation of Distribution of Math Scores

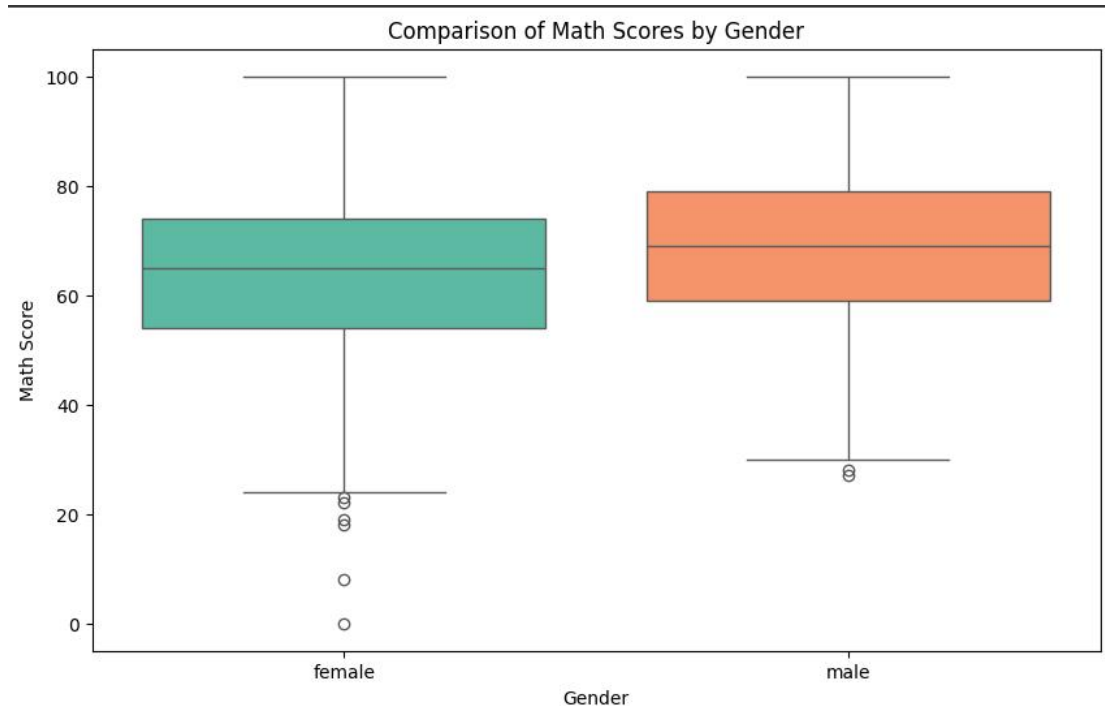


2. Correlation between scores



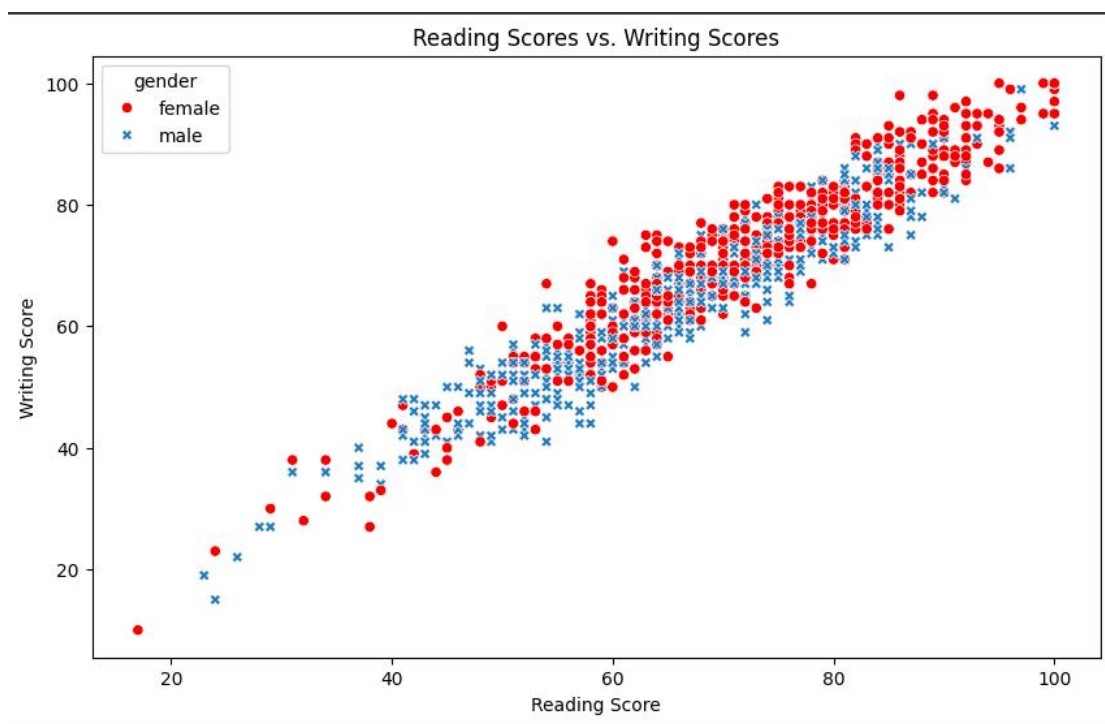
3. Box Plot of Math Scores by Gender

- This box plot visualizes the distribution of math scores among different genders, highlighting central tendency and variability.
- The box plot displays the median, quartiles, and potential outliers of math scores for each gender.



4. Scatter Plot of Reading Scores vs. Writing Scores

This scatter plot visualizes the relationship between reading scores and writing scores, differentiated by gender.



5. Line Plot of Average Scores Over Different Education Levels

This line plot visualizes the average scores in math, reading, and writing based on different parental education levels.

