Activity 1

1.

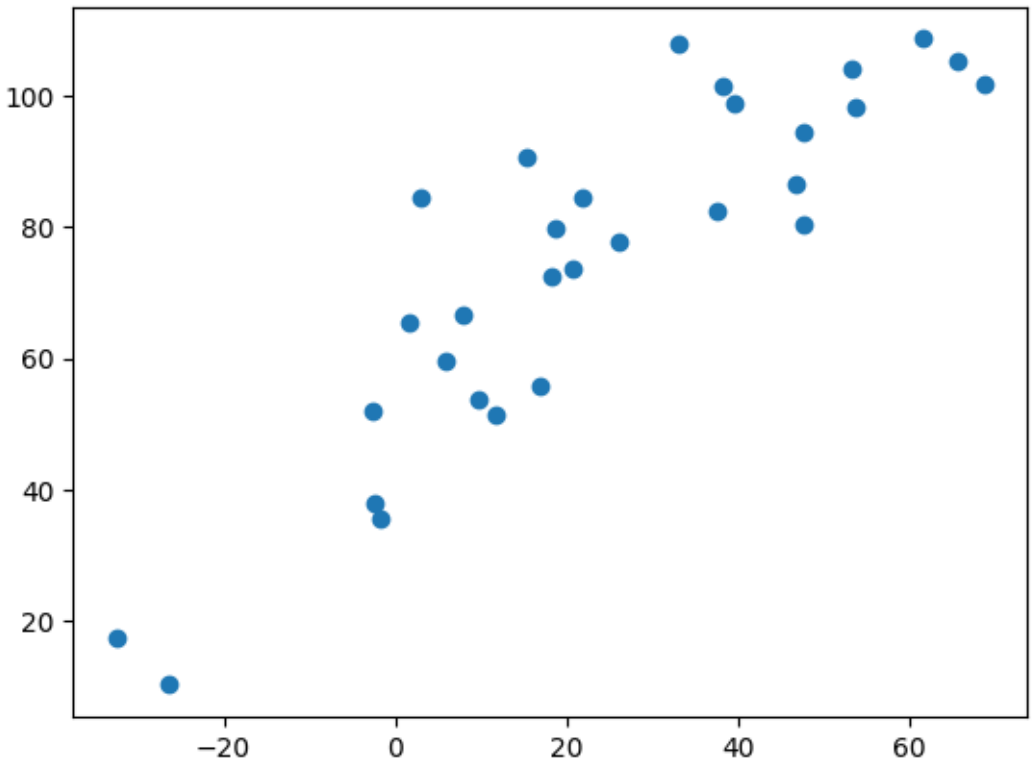
The exercise centered on analyzing data through covariance and Pearson correlation to evaluate relationships between variables. While covariance revealed how two variables change together, its interpretation was limited by differences in scale. In contrast, Pearson's correlation offered a standardized measure ranging from -1 to +1, allowing for clearer assessment of both the strength and direction of relationships. Applying these methods in Python enhanced our ability to interpret real-world datasets and determine the appropriate use of each statistical measure.

data1: mean=23.502 stdv=25.228

data2: mean=74.650 stdv=26.041

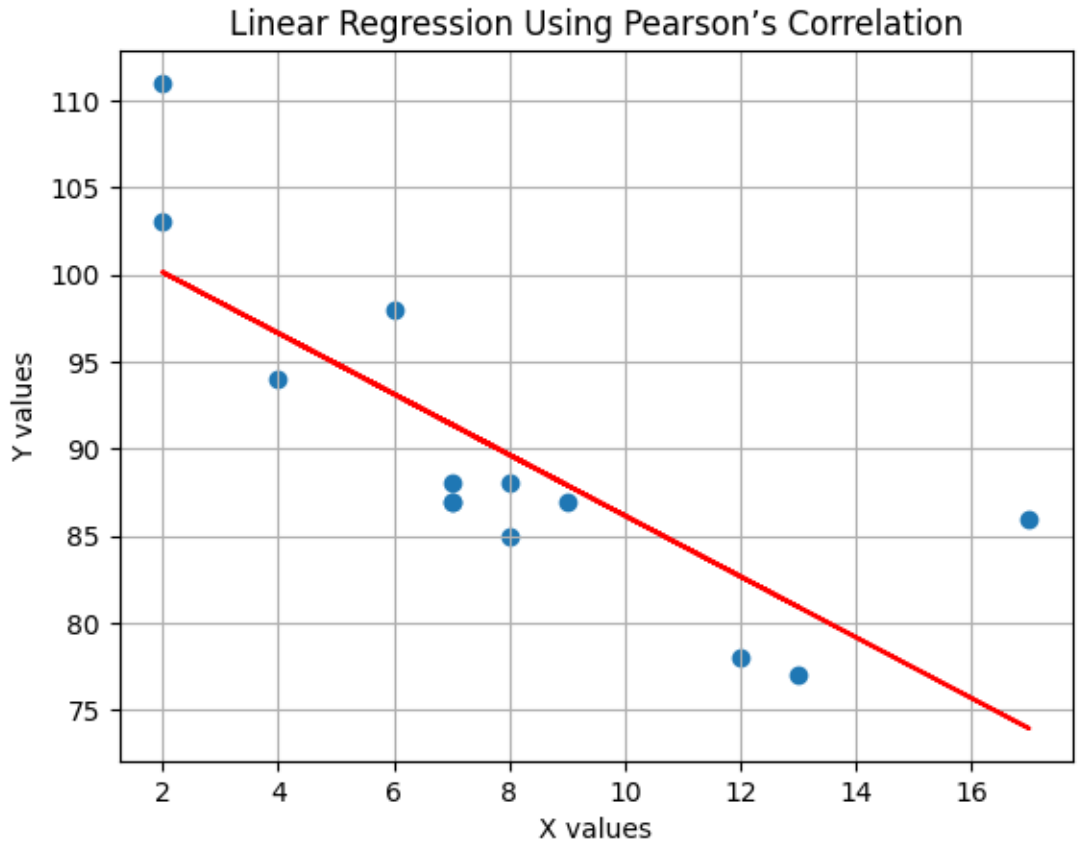
Covariance: 599.774

Pearsons correlation: 0.883



2.

During this task, various Google Colab programs were run to analyze the difficulties often encountered in unprocessed data and the techniques employed to manage them. The visual below represents the output of a linear regression model developed using Pearson's correlation metric.



Correlation: -0.784

3.

The goal of this task was to explore polynomial regression by fitting a third-degree polynomial to a non-linear dataset. We used NumPy to calculate the best-fit curve and matplotlib to visualise the results. The plotted curve followed the shape of the data much more closely than a linear model would, highlighting how polynomial regression can handle more complex trends. This activity emphasised the importance of choosing the right model for the right data, especially when visual analysis shows clear curvature that linear regression can't capture.