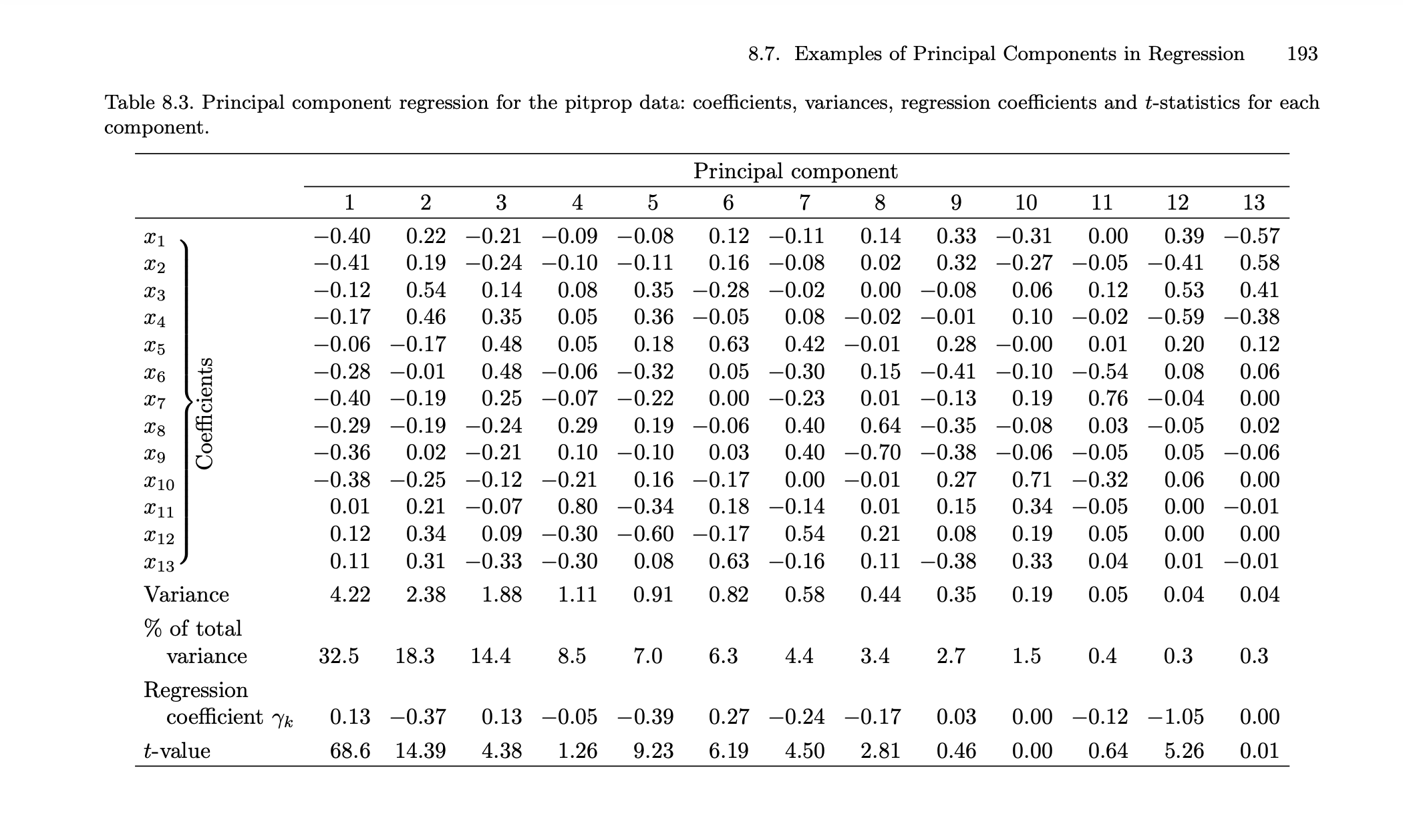
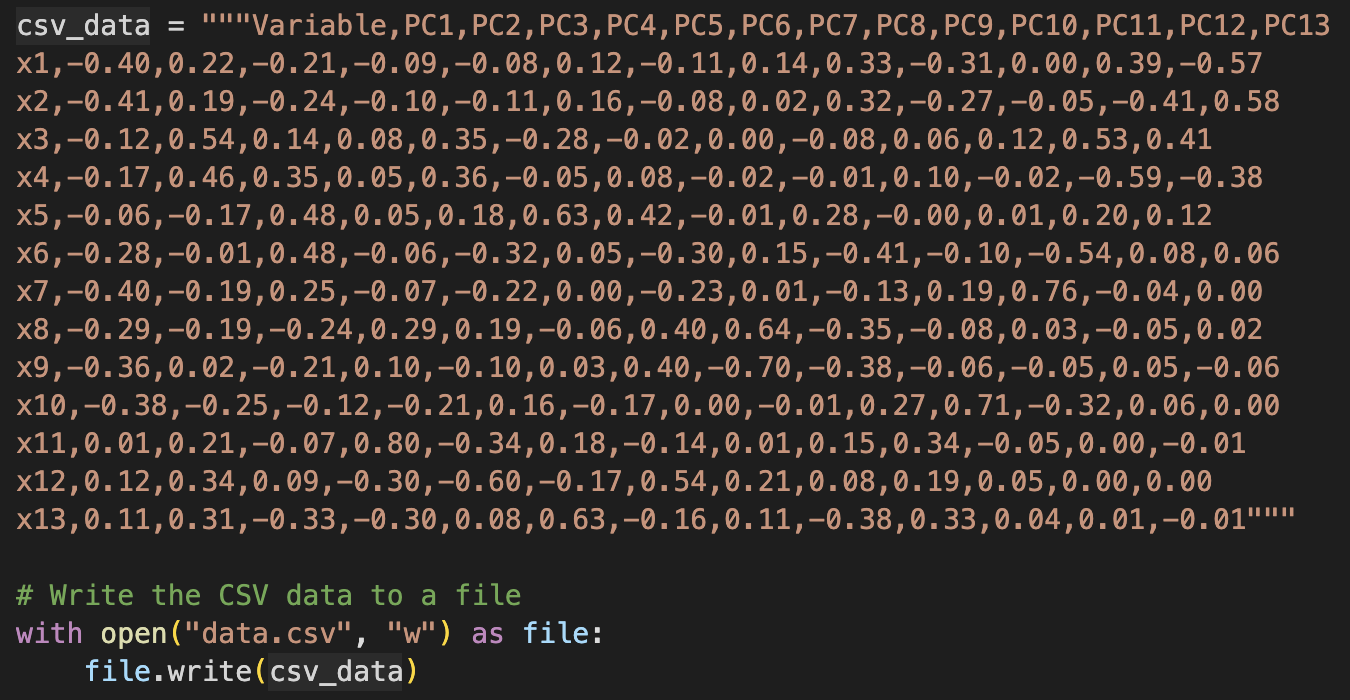
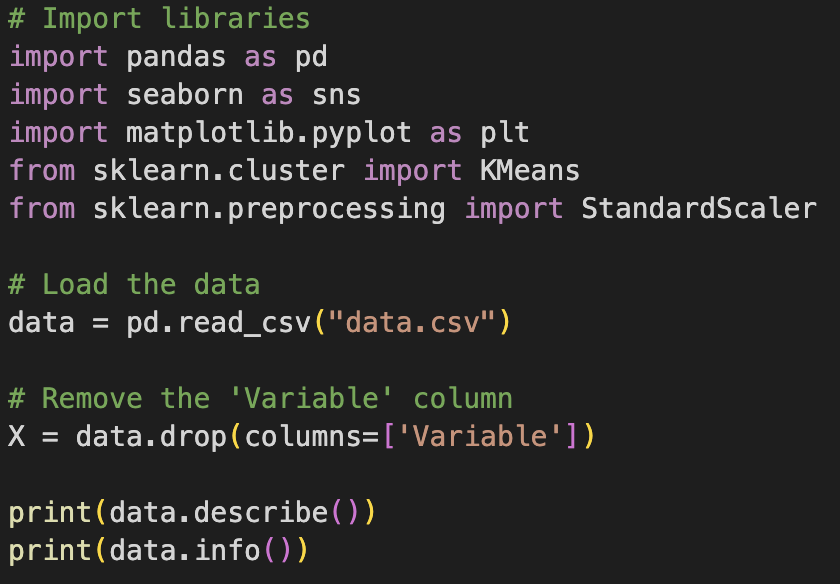
**Clustering analysis - K-mean**

This clustering analysis assignment using this data below for finding K-means using python library including pandas, seaborn, matplotlib and sklearn  


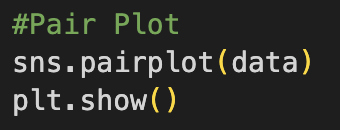
On the first step we’re preparing data for google colab by writing a script to write a csv file in to colab’s storage for avoiding data losses when runtime get reset



After that we start to import and analyze data before performing data calculation



**Data calculations**

* **Pair Plot**  
    
  - visualizes pairwise relationships between all numerical columns in the dataset.  
  - Helps to identify patterns, trends, and correlations that may influence clustering results.
* **K-means Clustering**A computer screen shot of a program code

  Description automatically generated1. Data Scaling  
  - Standardization ensures all variables have the same scale (mean = 0, standard deviation = 1).  
  - K-Means is sensitive to the scale of data; therefore, scaling is a necessary preprocessing step.  
  2. K-Means Clustering for Different k values  
  - Visualize each k values to a graph
* **Elbow Method to Determine Optimal k**A black screen with text on it

  Description automatically generated- The Elbow Method helps identify the optimal number of clusters.  
  - Inertia is plotted against k values:  
   - Initially, inertia decreases rapidly as k increases.  
   - At a certain point (the “elbow”), the decrease slows down, indicating the optimal number of clusters.  
  - The “elbow” point is the value of k where adding more clusters does not significantly reduce inertia.

**Results**

Elbow Curve

A graph of a number of clusters

Description automatically generated

From this elbow curve result as “At a certain point, the decrease slows down, indicating the optimal number of clusters”. We can see that at k = 2 is where the decrease slow down. So, we choose 2 as an optimal k value

A diagram of a number of red and yellow dots

Description automatically generated

**Graph of K = 2 K-means clustering result**