



A Project for the Course, Intro to Data Science, Mahidol University

CLUSTERING WINE

BASED ON THEIR CHEMICAL CONSTITUENTS



Presented by Jirapit Sripitak, Arthakorn Raksenawanich, Chitsanuwat Nitikunchainun

Agenda

Introduction to Wine	3
Dataset Preview	4
Objective and Expected Results	5
Exploratory Data Analysis (EDA)	7
Principal Component Analysis (PCA)	11
t-distributed stochastic neighbor embedding (t-SNE)	13
KMeans	15
Centroid-based clustering	17
Results and Interpretation	18
Conclusion	19

Introduction to Wine





Dataset Preview

The Attributes are:

- Alcohol
 - % of alcohol in wine
- Malic acid
 - Main acids found in grapes
- Ash
 - The residue left after the evaporation
- Alkalinity of ash
 - The sum of cations combined with the organic acids in the wine
- Magnesium
 - Stress-protectant factor in preventing cell death in yeast in wine
- Total phenols
 - Contributes to the astringency, bitterness, and body of the wine
- Flavanoids
 - Compounds responsible for the color, flavor, and mouthfeel of the wine
- Nonflavanoid phenols
 - Same as above but does not contain the flavone structure
- Proanthocyanins
 - The astringency and bitterness of wine
- Color intensity
 - Depth or darkness of the color of the wine
- Hue
 - Actual color of the wine
- OD280/OD315 of diluted wines
 - A measure of the absorbance ratio of light at two different wavelengths
- Proline
 - Amino acid that gives a bitter taste, slightly sweet or umami taste



Objectives

The objective of this script is to analyze a dataset of wines that contains numerous physical and chemical characteristics and cluster the wines according to how similar they are.

Objective and Expected Results

The script employs a KMeans clustering algorithm, Principal Component Analysis (PCA), and t-SNE for dimensionality reduction and visualization.

The performance of the clustering is evaluated using the silhouette score to ensure the quality of the resulting clusters.





Expected Results

Optimal clustering of wines based on their attributes, providing insights into the structure of the dataset and relationships between wine properties.

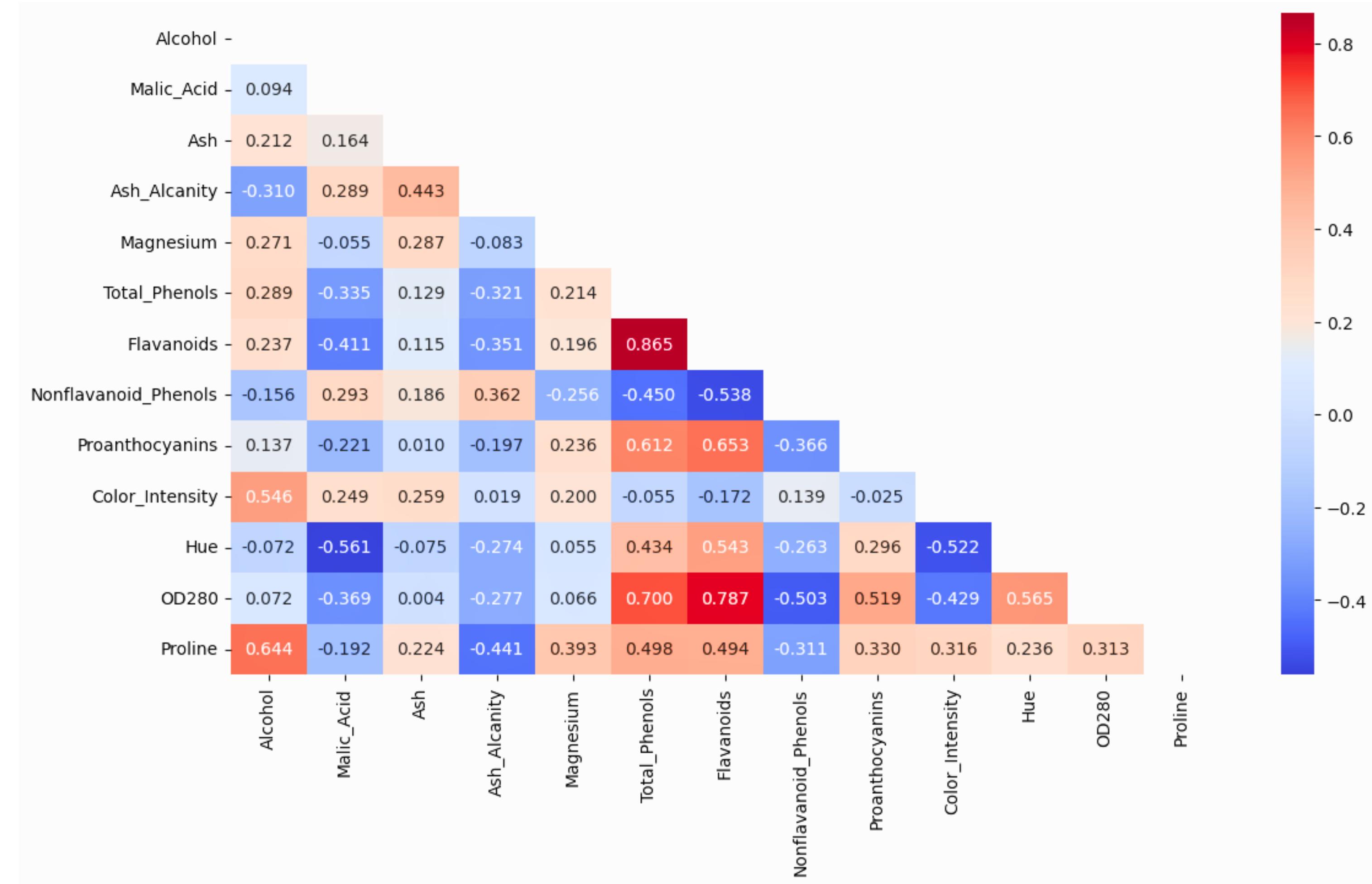
Objective and Expected Results

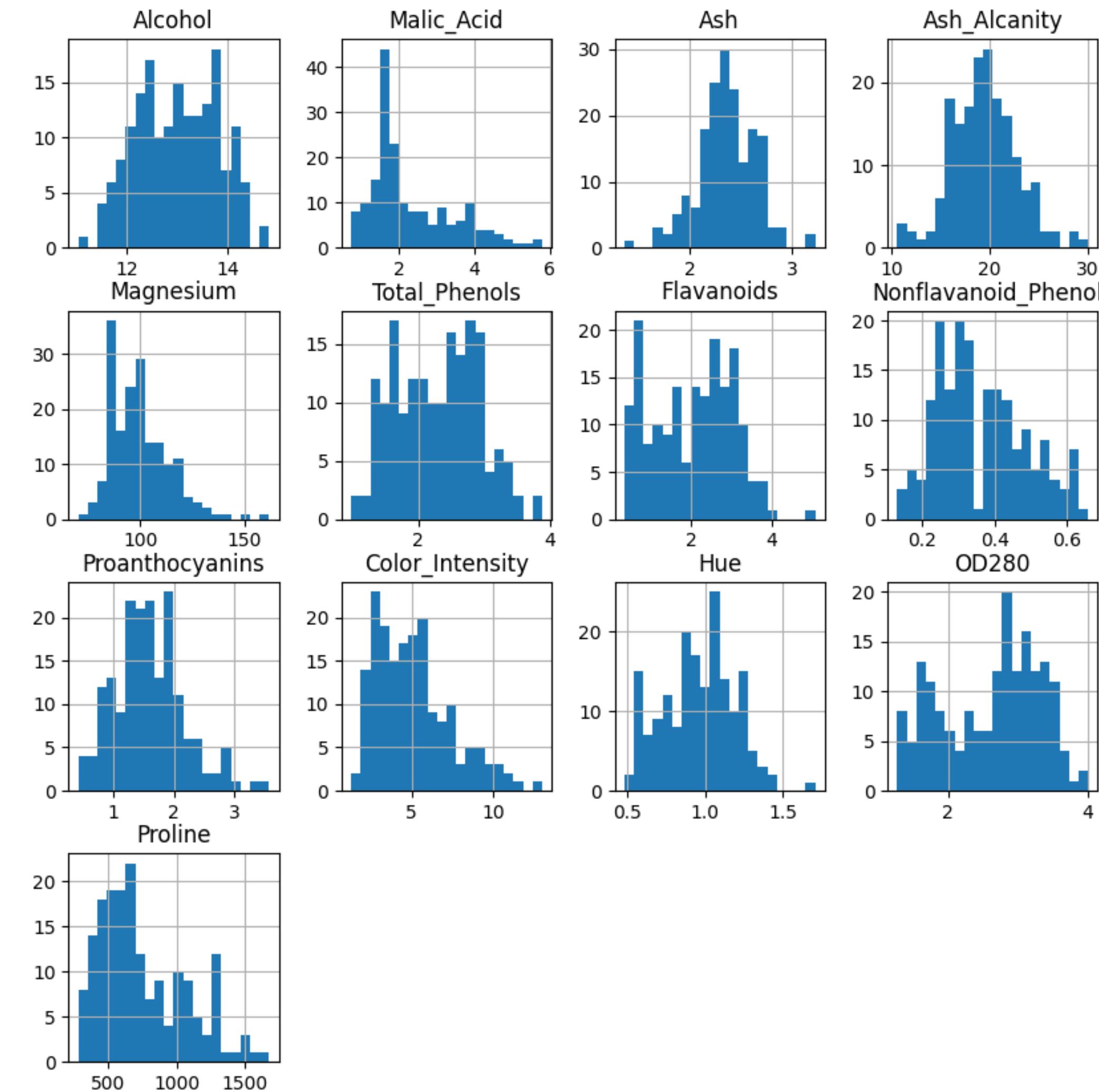
Using PCA and t-SNE to reduce the dataset's dimensions and visualize it, we can more easily analyze the clustering results and identify patterns.

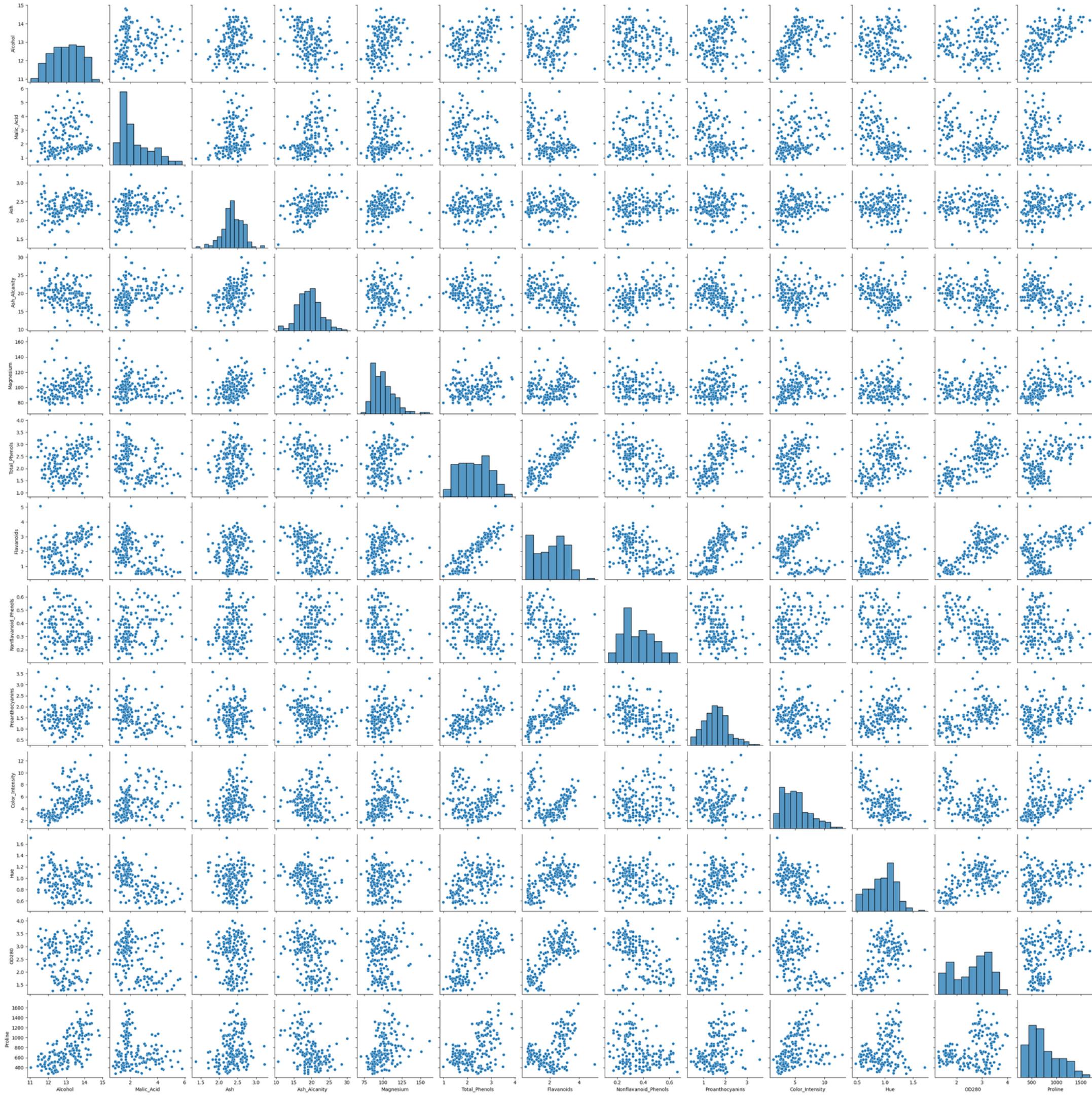
Evaluation of clustering quality using the silhouette score and providing a basis for further analysis or potential improvements.



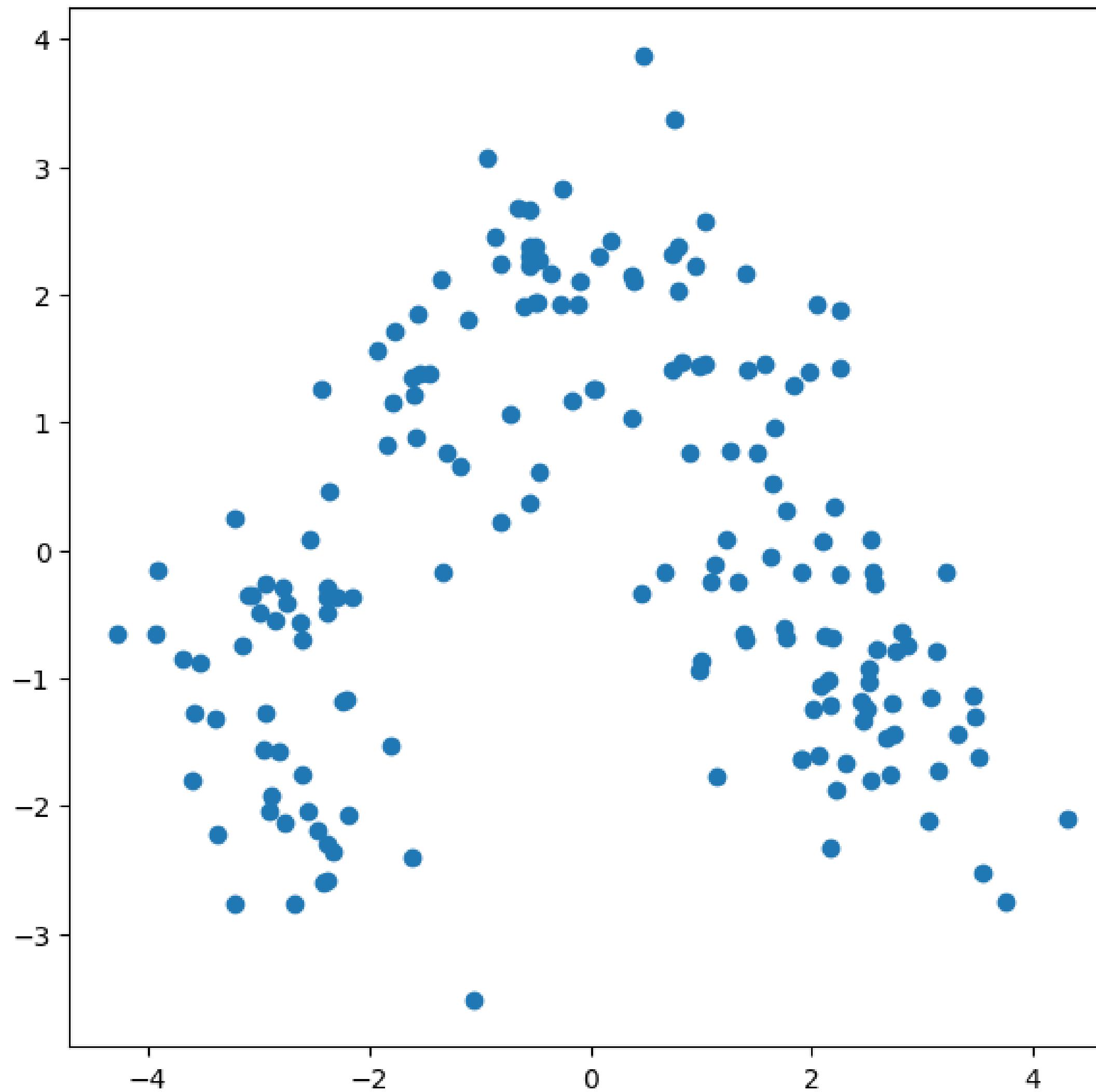
EXPLORATORY DATA ANALYSIS (EDA)



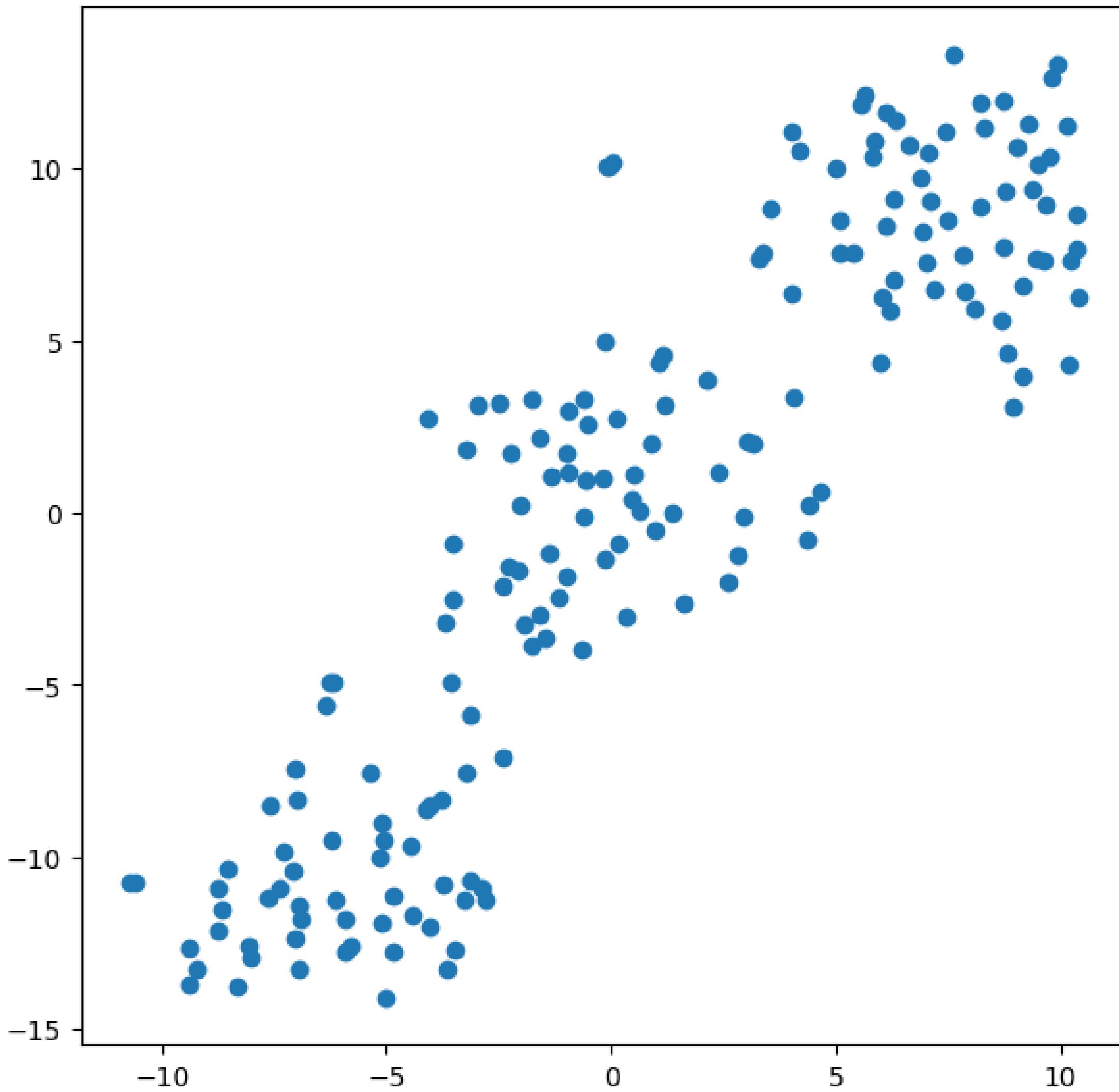




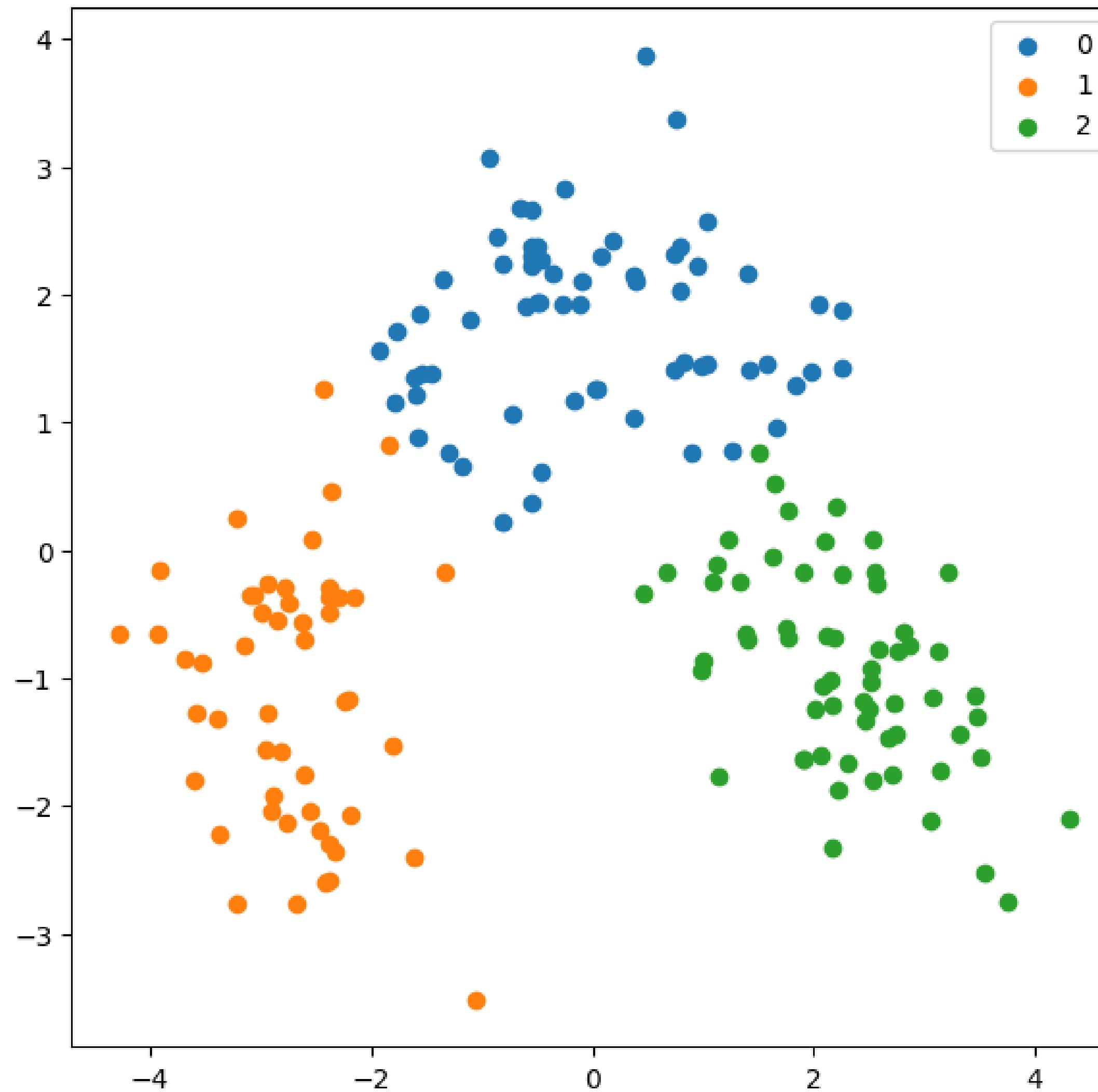
PRINCIPAL COMPONENT ANALYSIS (PCA)



T-DISTRIBUTED STOCHASTIC NEIGHBOR EMBEDDING (T-SNE)



KMEANS



CENTROID-BASED CLUSTERING

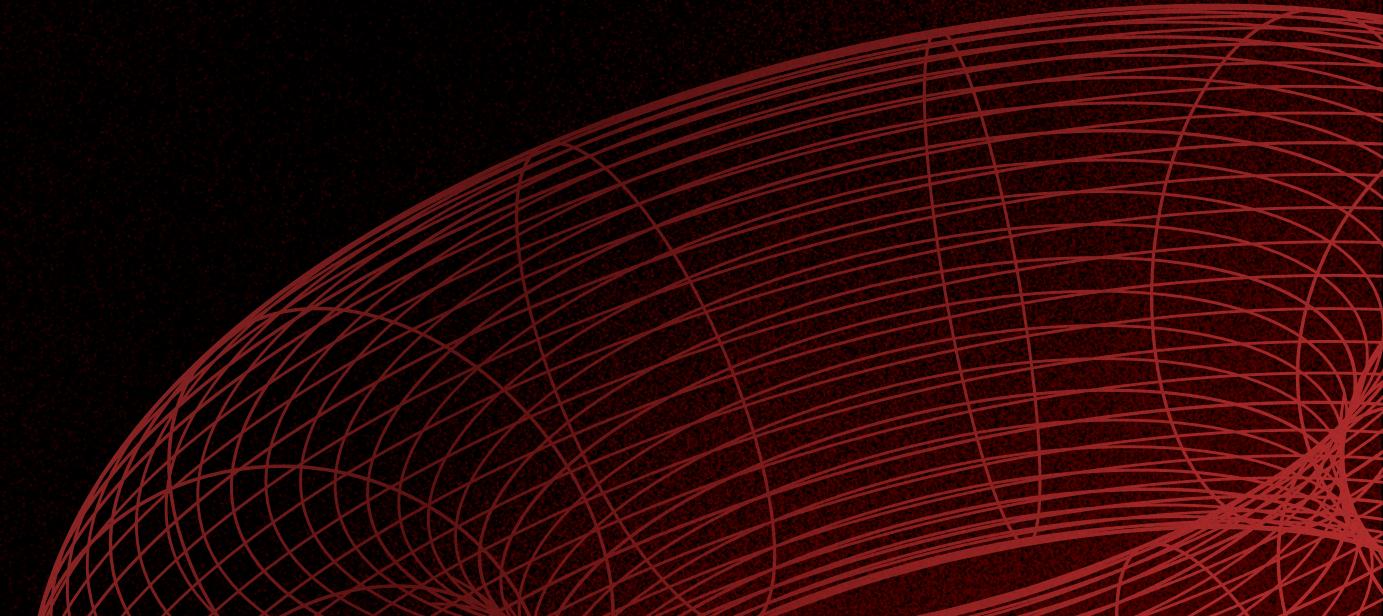
Results and Interpretation

From these centroids, we can make the following observations

-
- CLUSTER 0** has wines with relatively lower alcohol content, total phenols, flavonoids, and proline levels. It has an average malic acid content, ash alkalinity, and magnesium levels. The color intensity is relatively low compared to other clusters.
-
- CLUSTER 1** has wines with higher malic acid content, ash alcalinity, and magnesium levels. It has a relatively lower total phenols, flavanoids, hue, and OD280. The color intensity is the highest among the clusters.
-
- CLUSTER 2** has wines with the highest alcohol content, total phenols, flavanoids, hue, OD280, and proline levels. It has the lowest ash alcalinity and nonflavanoid phenols. The color intensity is in the middle compared to the other clusters.

CONCLUSION

-
- CLUSTER 0** represents wines with lower alcohol content and relatively balanced attributes, leaning towards lower total phenols, flavanoids, and proline levels.
-
- CLUSTER 1** represents wines with a high malic acid content, high color intensity, and lower levels of total phenols and flavanoids.
-
- CLUSTER 2** represents high-quality wines with high alcohol content, high levels of total phenols, flavanoids, hue, OD280, and proline, and low ash alkalinity and nonflavanoid phenols.





A Project for the Course, Intro to Data Science, Mahidol University

THANK YOU FOR YOUR ATTENTION!



Presented by Jirapit Sripitak, Arthakorn Raksenawanich, Chitsanuwat Nitikunchainun