Data Science and Big Data Analysis

**Summary of the Dataset**:

Dataset Source: Buddy Move Dataset

URL: https://archive.ics.uci.edu/dataset/476/buddymove+data+set

Number of Objects: 249

Number of Attributes: 7

Attribute Types:

'Userid': Categorical

'Sports': Continous

'Religious': Continuous

'Nature': Continuous

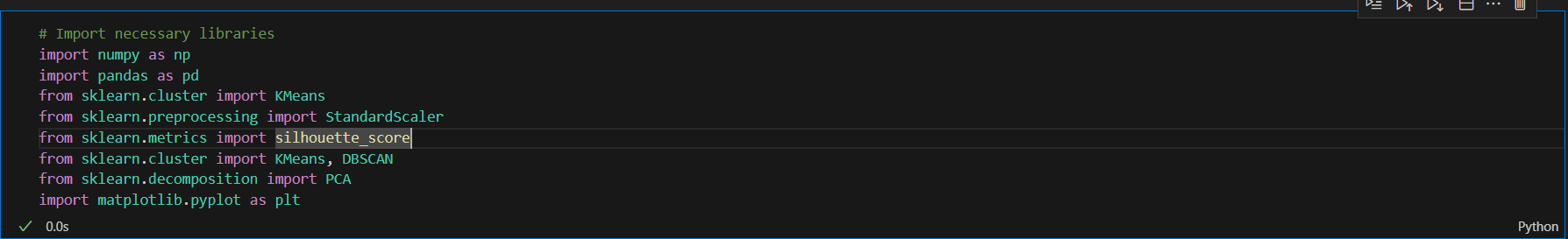
'Theatre': Continuous

'Shopping': Continuous

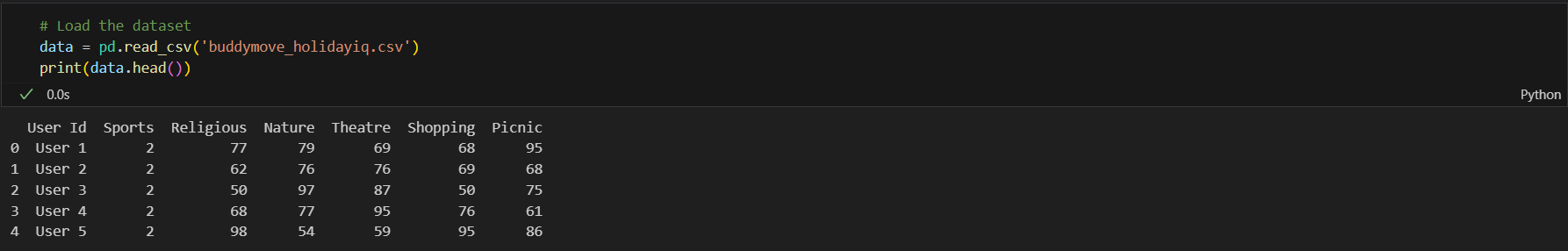
'Picnic': Continuous

**IMPLEMENTATION:**

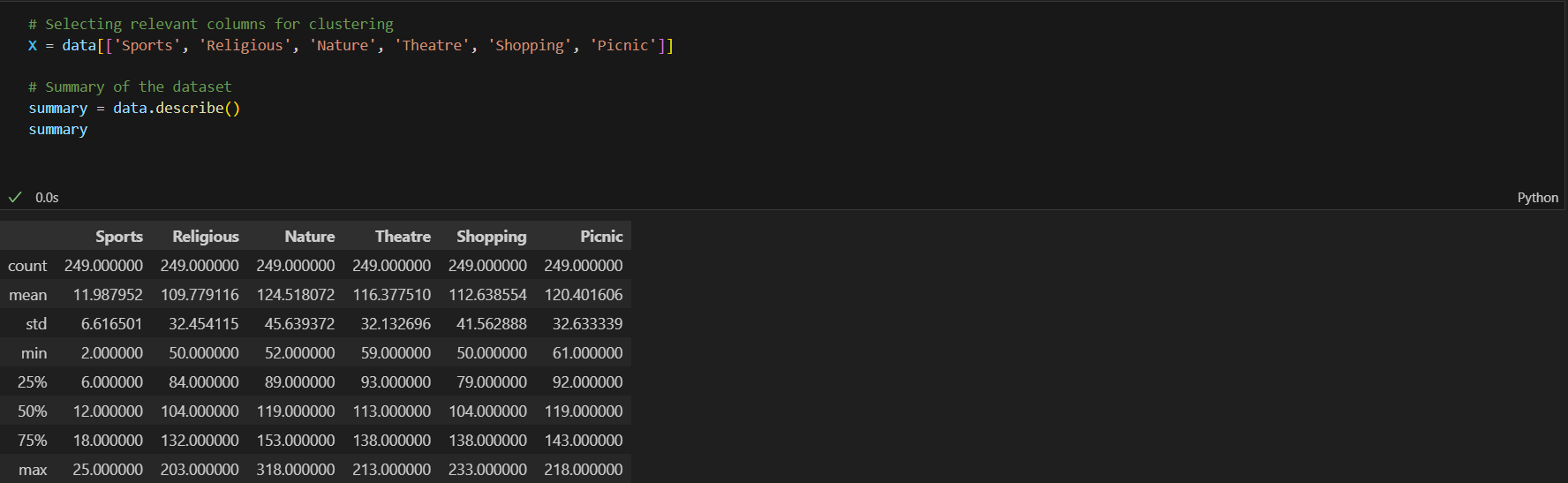
* **Import necessary libraries**



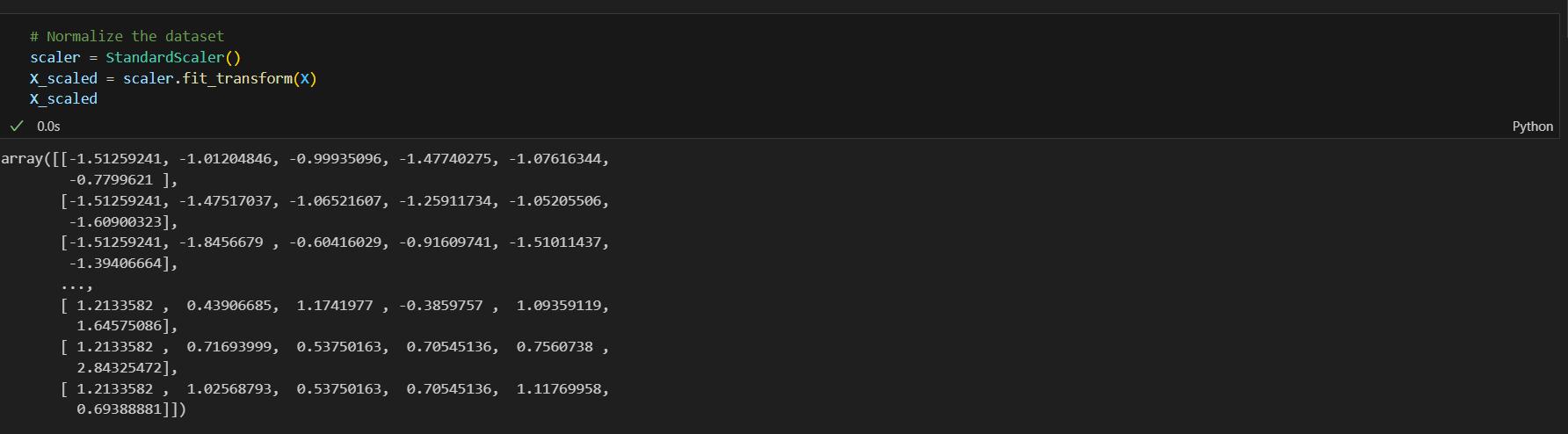
* **Load the dataset**



* **Preprocess the dataset**

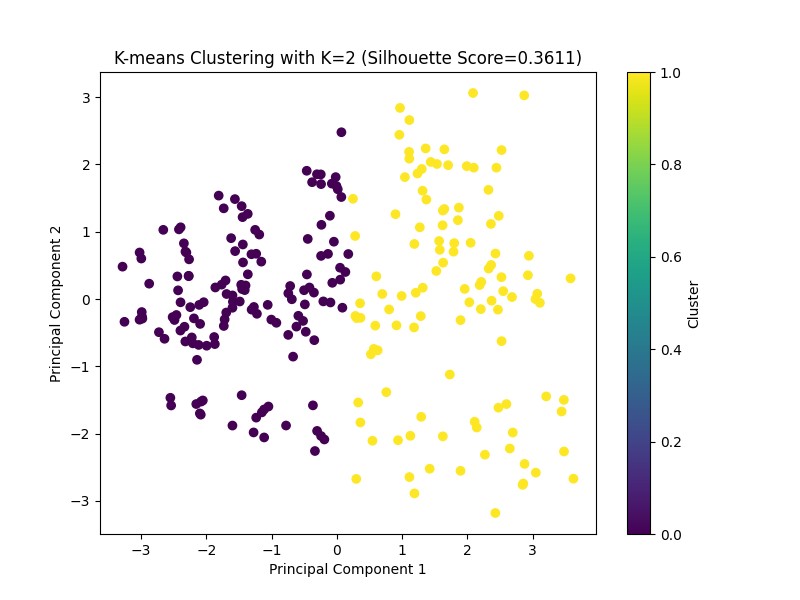


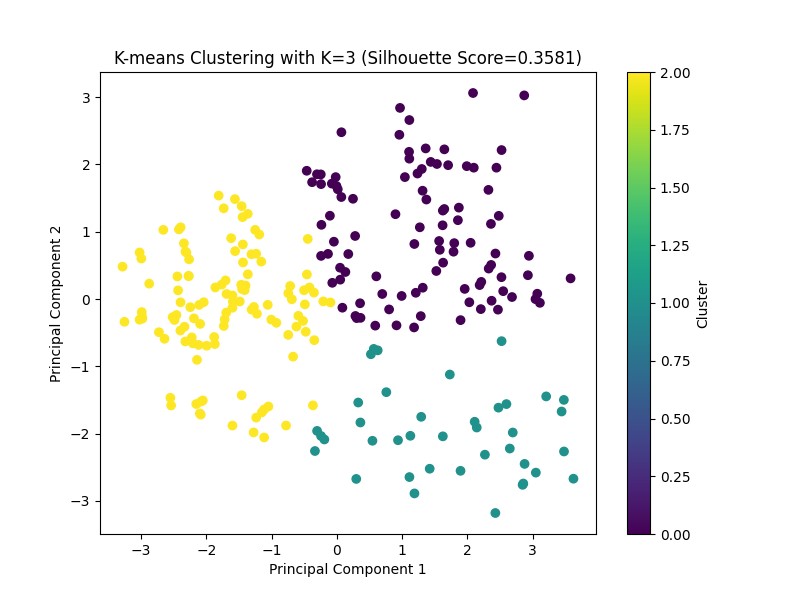
* **Normalize the dataset**

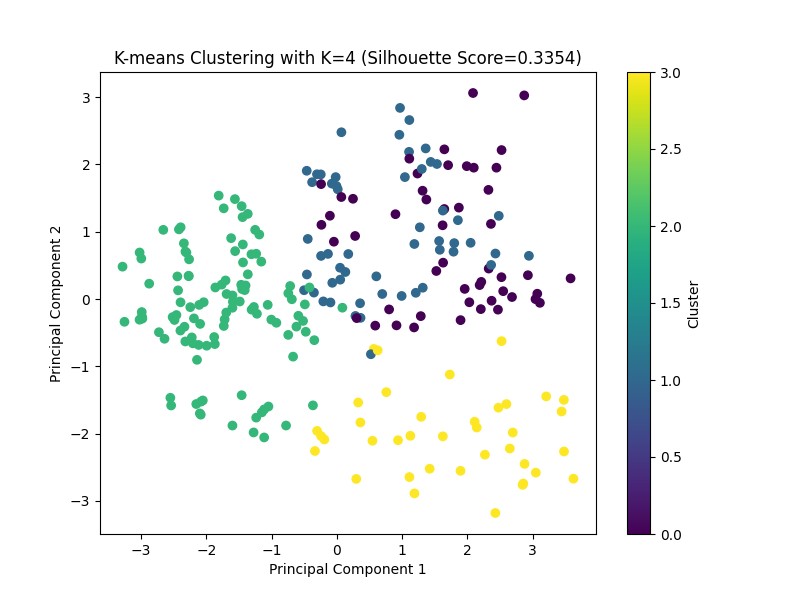


* **Perform K-means clustering**

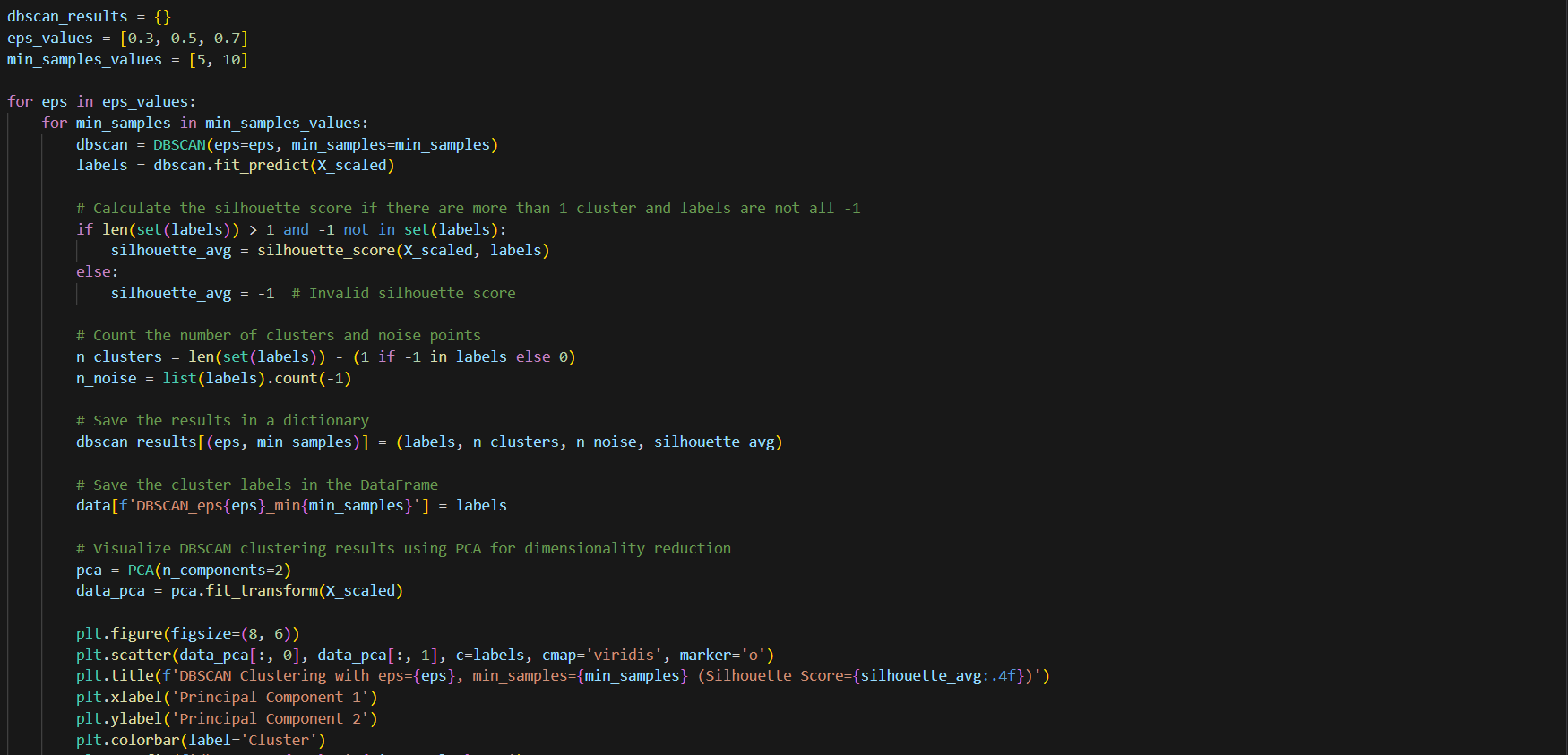


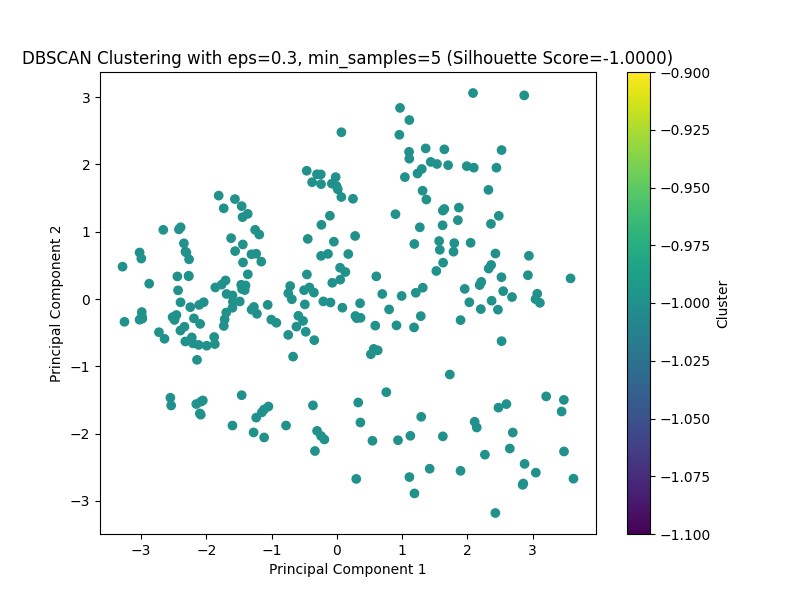


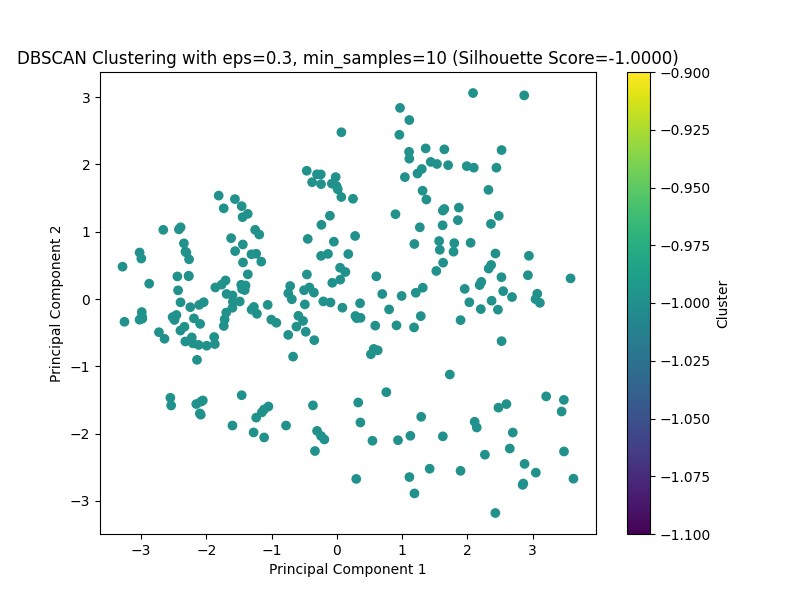


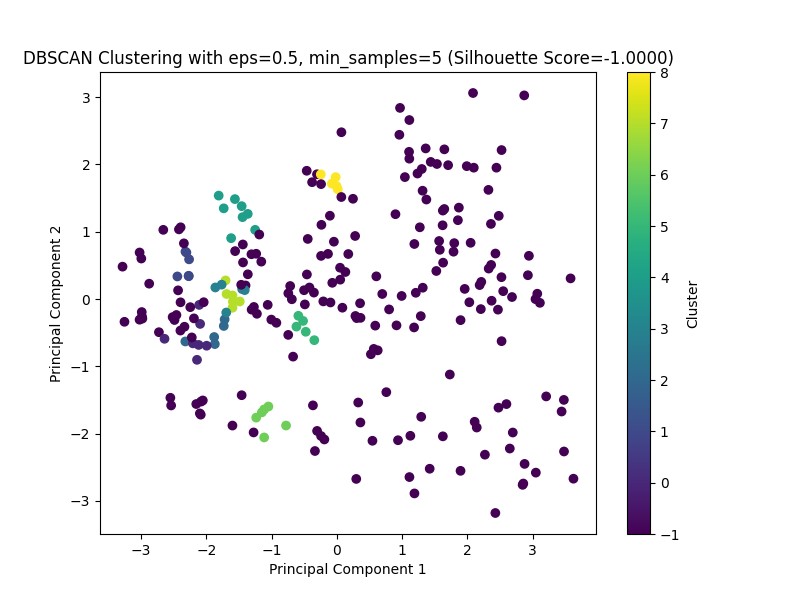


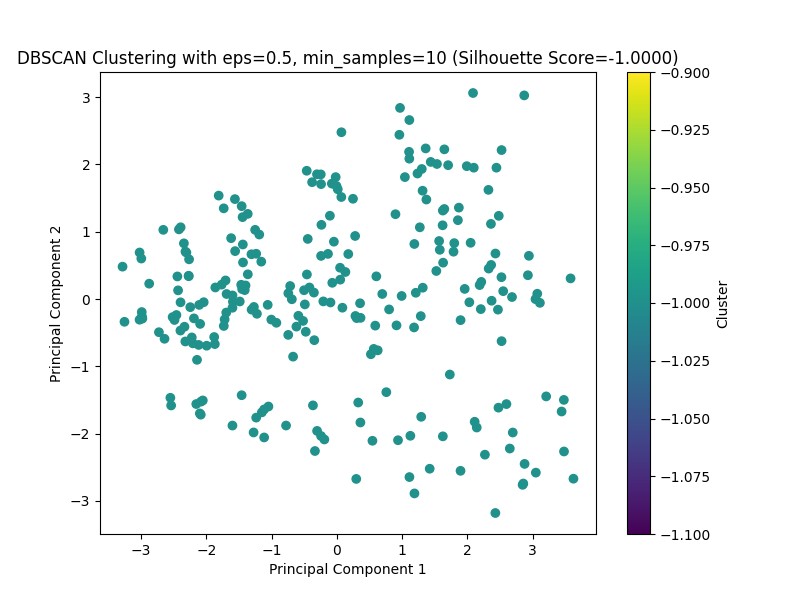
* **Perform DBSCAN clustering**

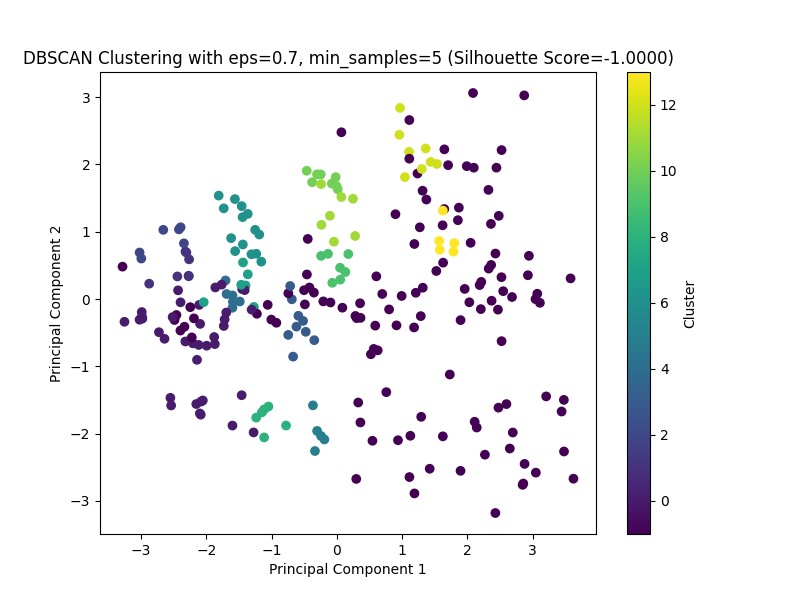


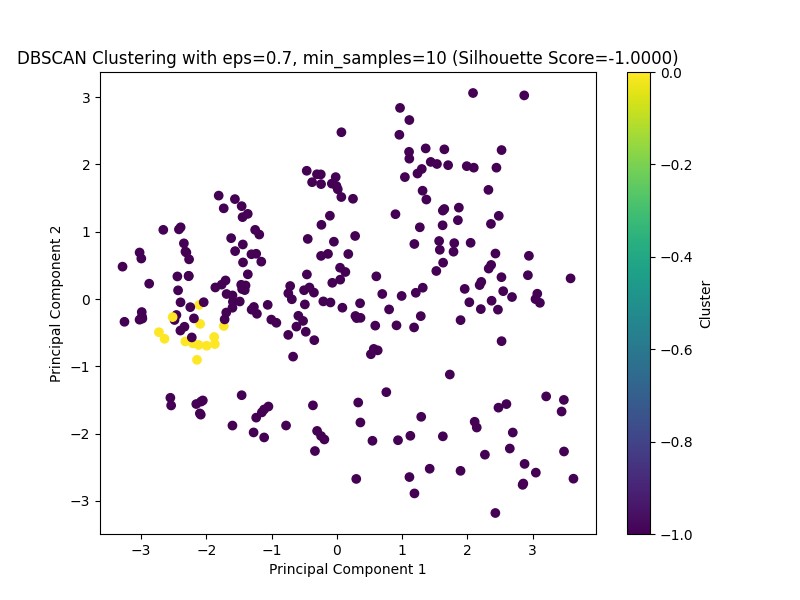




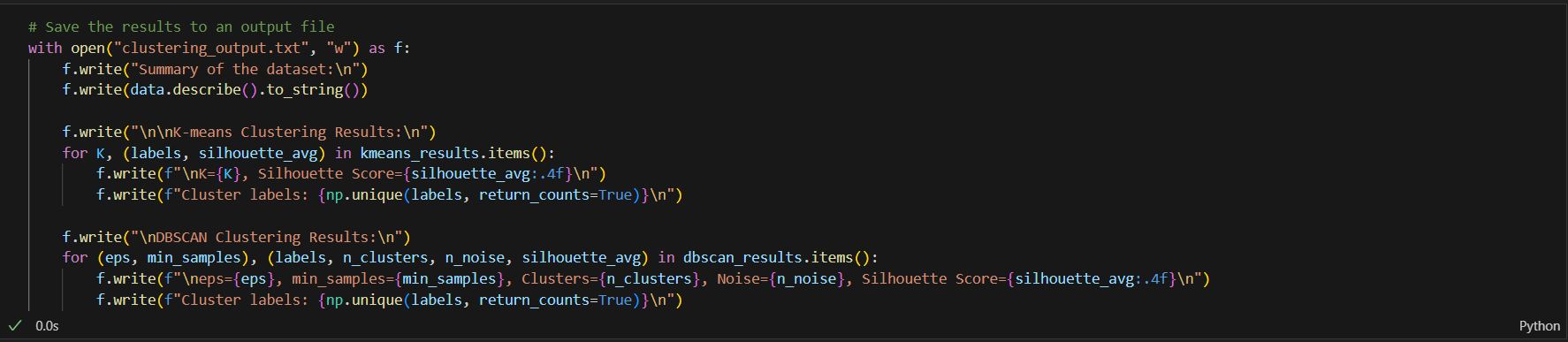








* **Save the results to an output file**



* **Conclusion**:

In this assignment, we explored clustering techniques, specifically K-means and DBSCAN, applied to the BuddyMove dataset. The dataset, consisting of 249 observations and 7 attributes related to user reviews across different categories, was first preprocessed by handling missing values and normalizing numerical attributes. This preprocessing step ensured that the data was suitable for clustering analysis. Through this exploration, we aimed to identify patterns in user review behavior and classify users based on their review activities in different categories.