

COMP4388

Weka Assignment

Dr Radi Jarrar

Maiia Kalajiawi (1170251)

Submitted onFeb 19, 2023

# 

# Question 1

Stochastic Gradient Descent (SGD) is an optimization algorithm commonly used for deep learning models, while Gradient Descent (GD) is a fundamental optimization algorithm used to minimize cost functions by iteratively adjusting parameters based on the gradient direction (Goodfellow et al., 2016). The primary difference between the two algorithms is that GD computes the gradient using the entire dataset, while SGD computes the gradient using a subset of the data, known as a mini-batch (Ruder, 2016). This allows SGD to be more efficient and faster than GD, especially for large datasets.

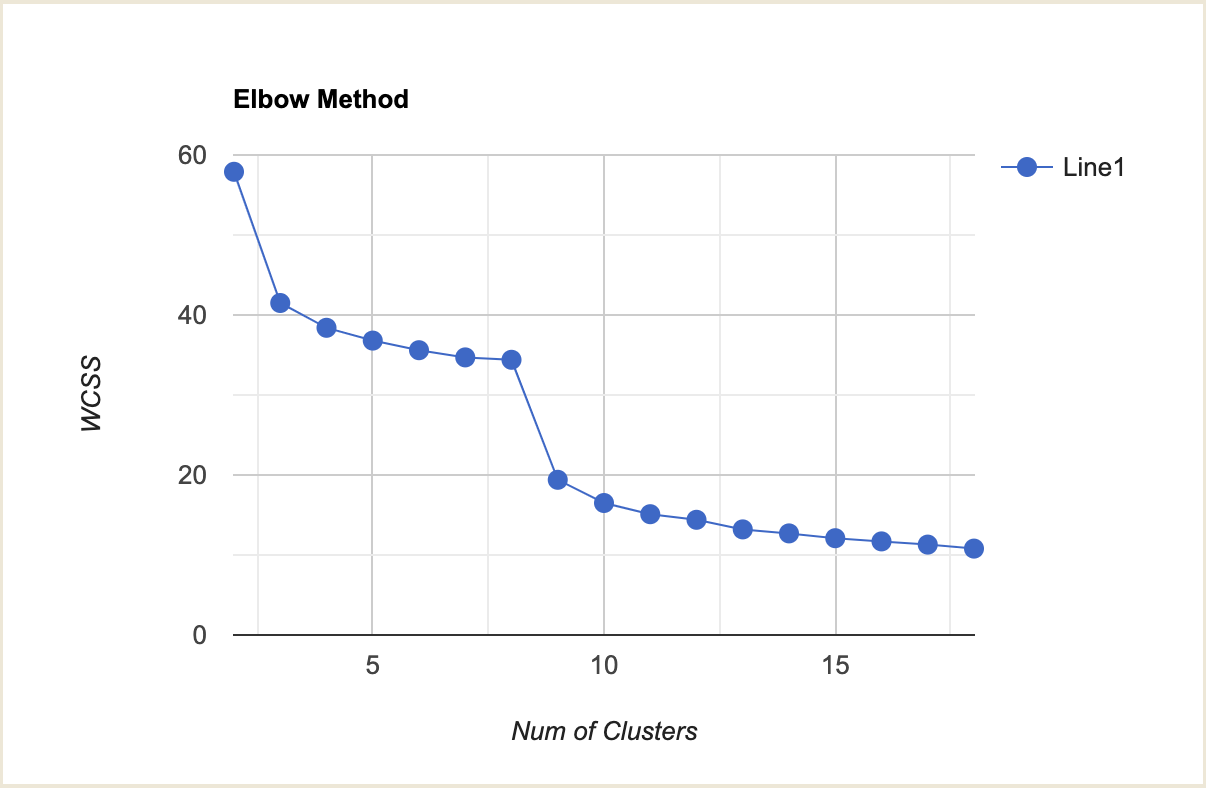
In GD, the parameters are updated after computing the gradient with respect to all training examples, while in SGD, the parameters are updated after computing the gradient with respect to a mini-batch of training examples (Goodfellow et al., 2016). This introduces randomness in the parameter updates, which can help SGD escape local minima and improve generalization. However, SGD can be less stable than GD since the updates are noisier and less precise than those in GD (Ruder, 2016). While SGD converges faster than GD, mini-batch gradient descent, which computes the gradient using a small batch of data, is a popular compromise between the two algorithms. It is less noisy than SGD and converges faster than GD (Goodfellow et al., 2016).

# Question 2

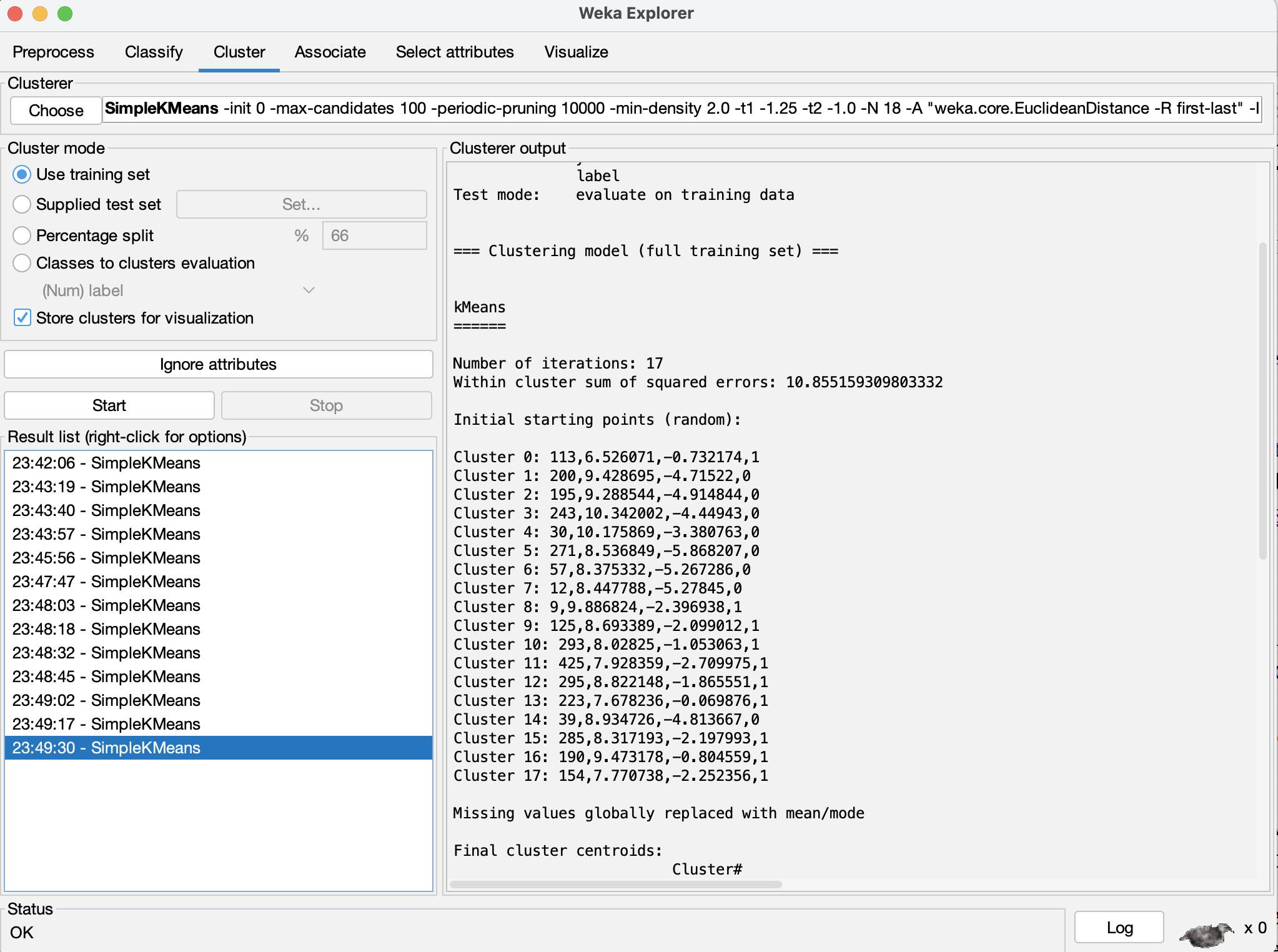
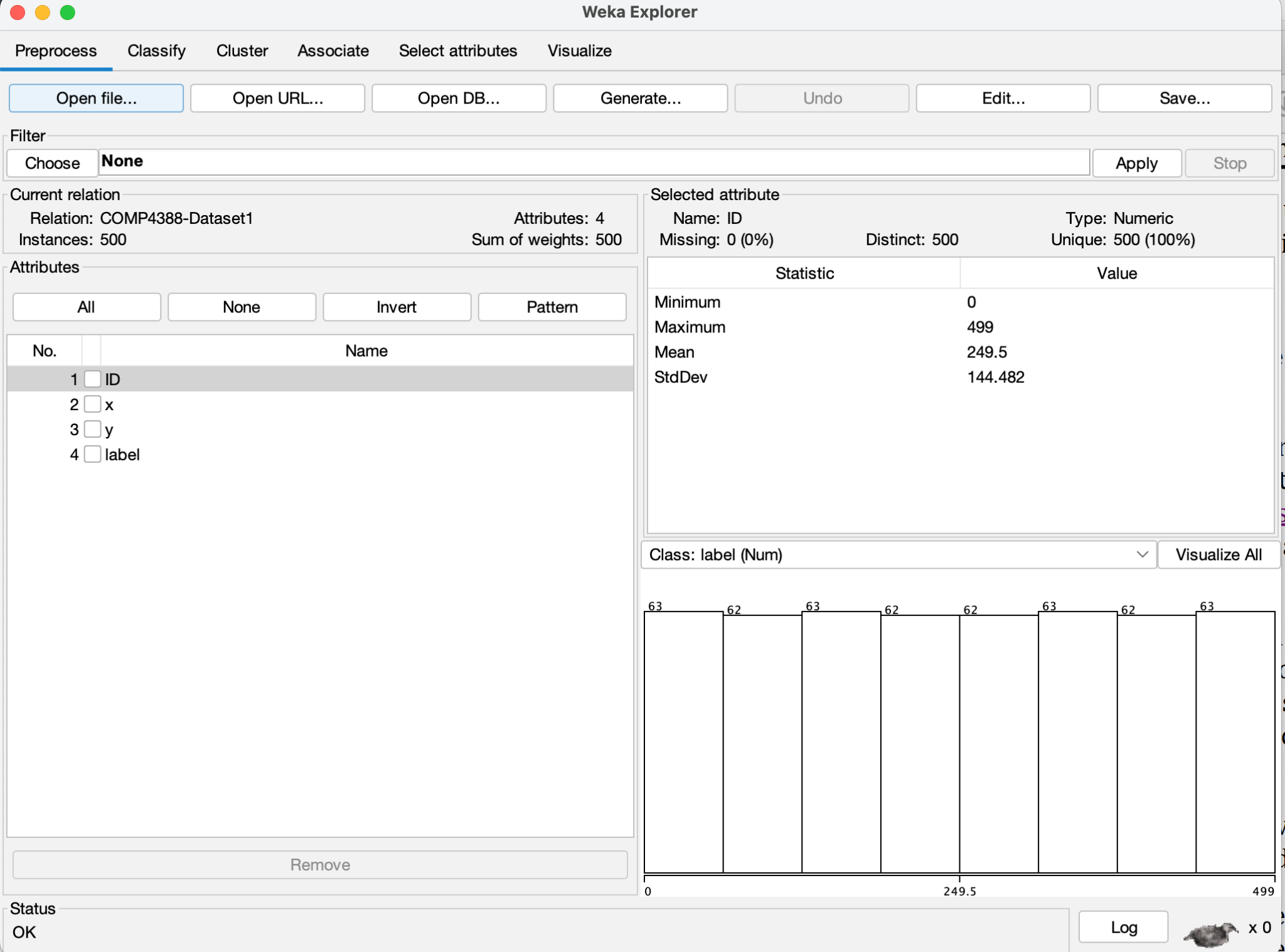
1. The two numeric variables x and y are likely to be the primary features used for clustering with the K-Means algorithm. These variables represent the coordinates of each data point in a two-dimensional space, and the algorithm will attempt to group together points that are close to each other in this space. The third numeric binary variable called label is not a feature but rather a label or class assigned to each data point. In other words, it represents the ground truth or known clustering of the data, which can be used to evaluate the accuracy of the K-Means algorithm's clustering results. However, it should not be included as a feature in the clustering process, as this would lead to biased results.
2. For this assignment, I tested the number of clusters (provided below) in the ranges of 2-18 and used the elbow method. I plotted the evaluation metric (WCSS) against the number of clusters and looked for an "elbow point" where the metric starts to level off. This elbow point indicates the number of clusters that provides the best trade-off between model complexity and clustering performance. Alternatively, one can use a method like the "gap statistic" that compares the observed WCSS to a reference distribution of WCSS values obtained from random data to determine the optimal number of clusters. Based on the graph 1 (shown below), I found that the best number of clusters is 10.

| **Number of Clusters** | **WCSS** |
| --- | --- |
| 2 | 57.9 |
| 3 | 41.5 |
| 4 | 38.4 |
| 5 | 36.8 |
| 6 | 35.6 |
| 7 | 34.7 |
| 8 | 34.4 |
| 9 | 19.4 |
| 10 | 16.5 |
| 11 | 15.1 |
| 12 | 14.4 |
| 13 | 13.2 |
| 14 | 12.7 |
| 15 | 12.1 |
| 16 | 11.7 |
| 17 | 11.3 |
| 18 | 10.8 |

| **Number of Clusters** | **WCSS** |
| --- | --- |
| 20 | 10.3 |
| 25 | 8.7 |
| 30 | 7.5 |



Graph 1 plots the evaluation metric against the number of clusters. The elbow point can be seen around 9-10 clusters.



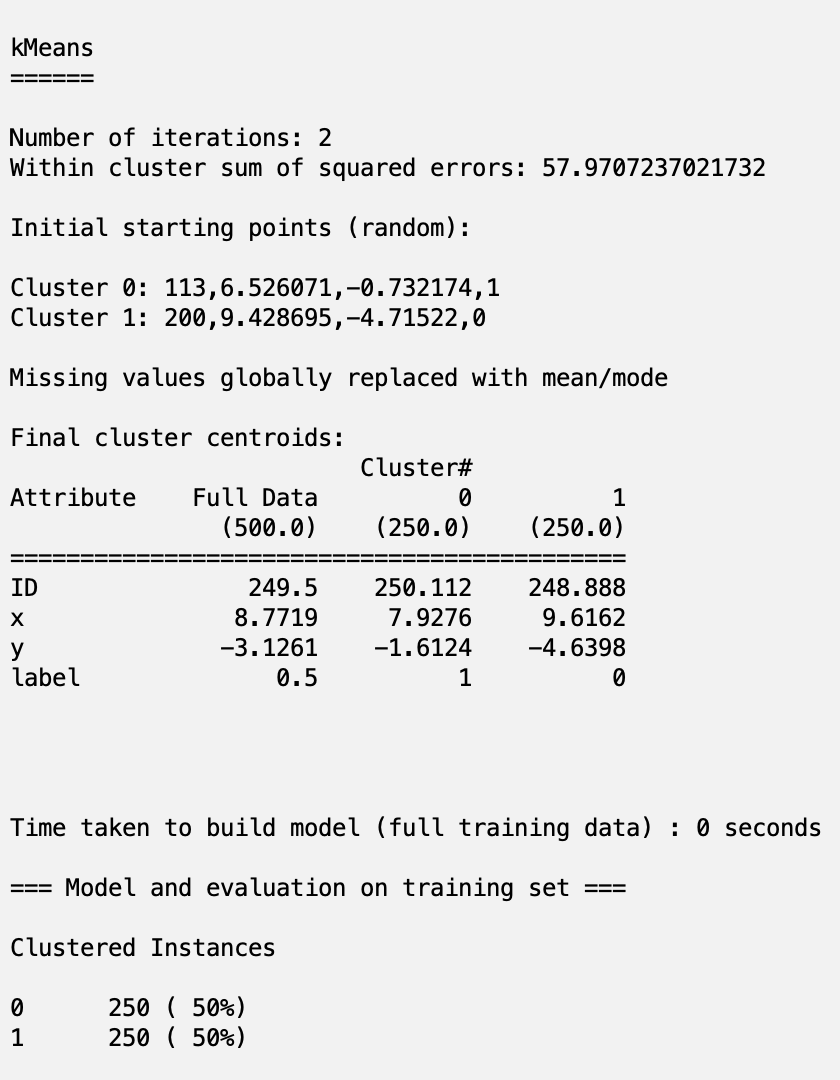


Figure 1 sets the number of clusters to 2. The WCSS result is very high and still needs further adjustment.

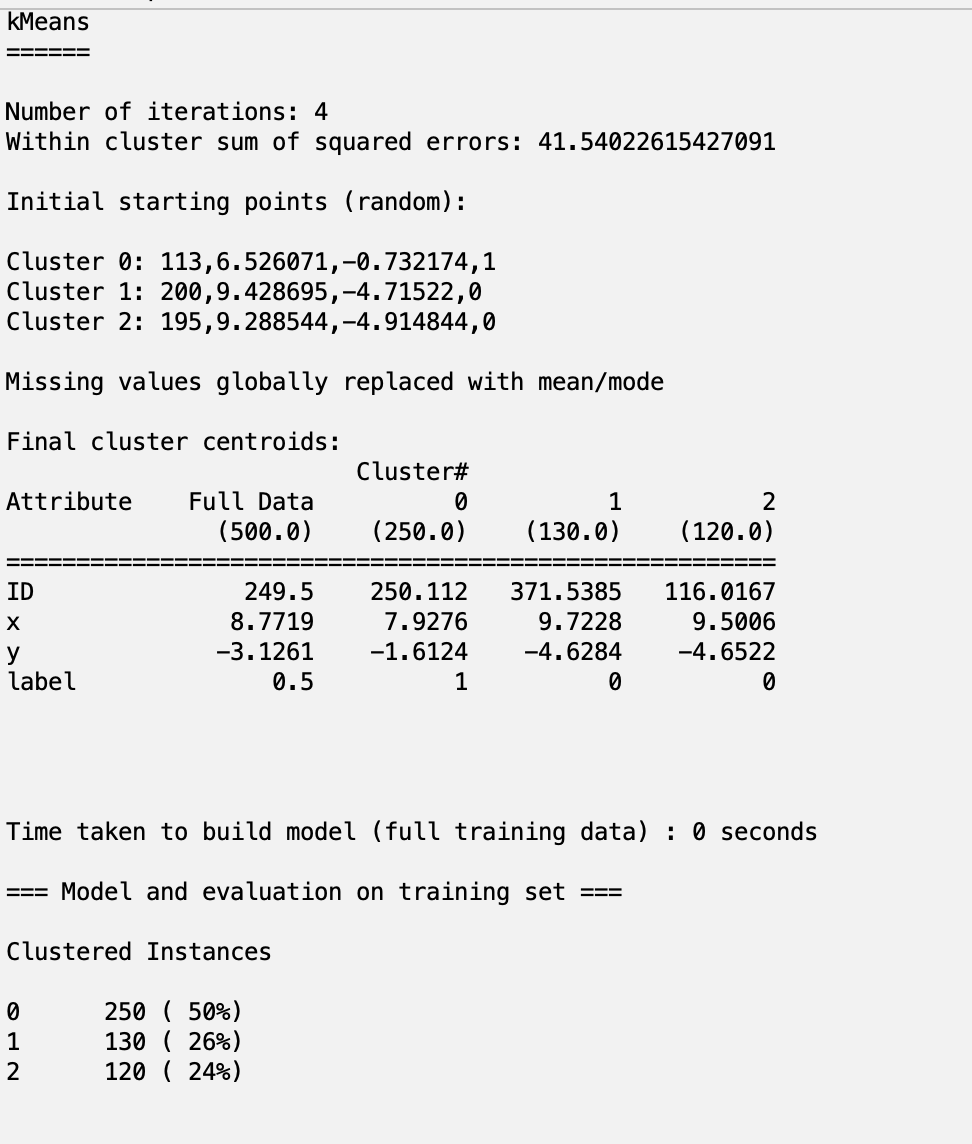


Figure 2 sets the number of clusters to 3. This time, the WCSS result has decreased by 28%. Might need further adjustment.

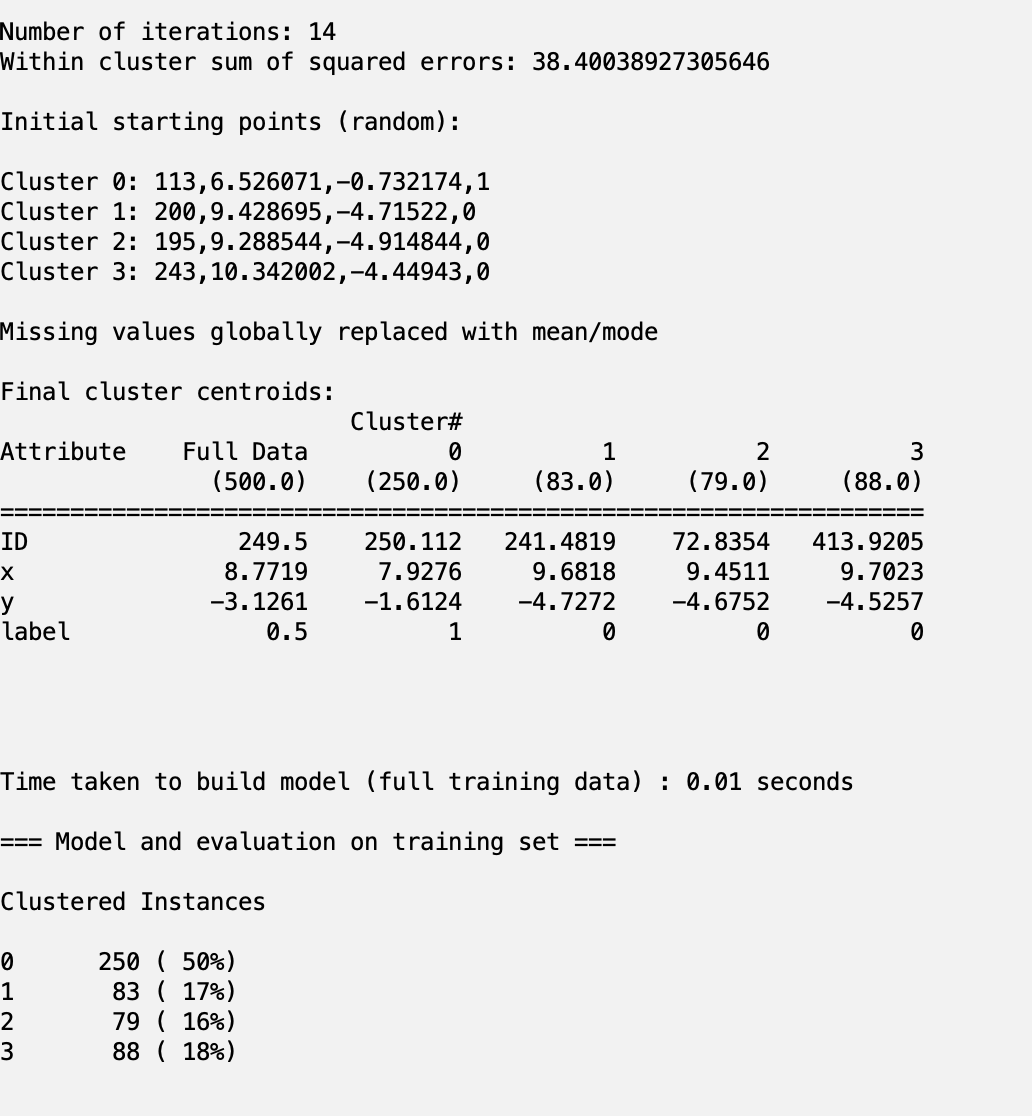


Figure 3 sets the number of clusters to 4. This time, the WCSS result has decreased again by 11 degrees. Might need further adjustment.

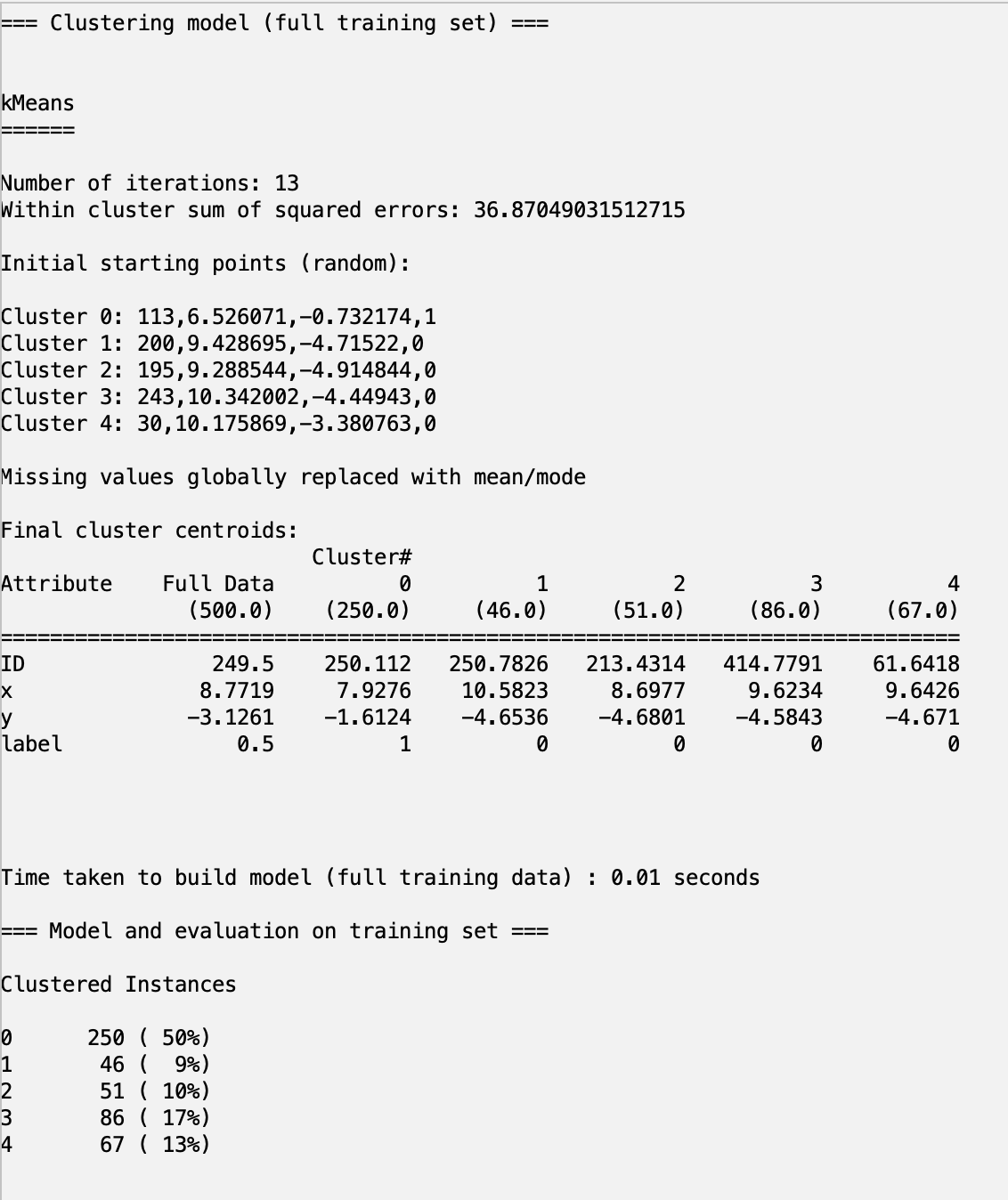


Figure 4 sets the number of clusters to 5. This time, the WCSS result has decreased again by 11 degrees. Might need further adjustment.

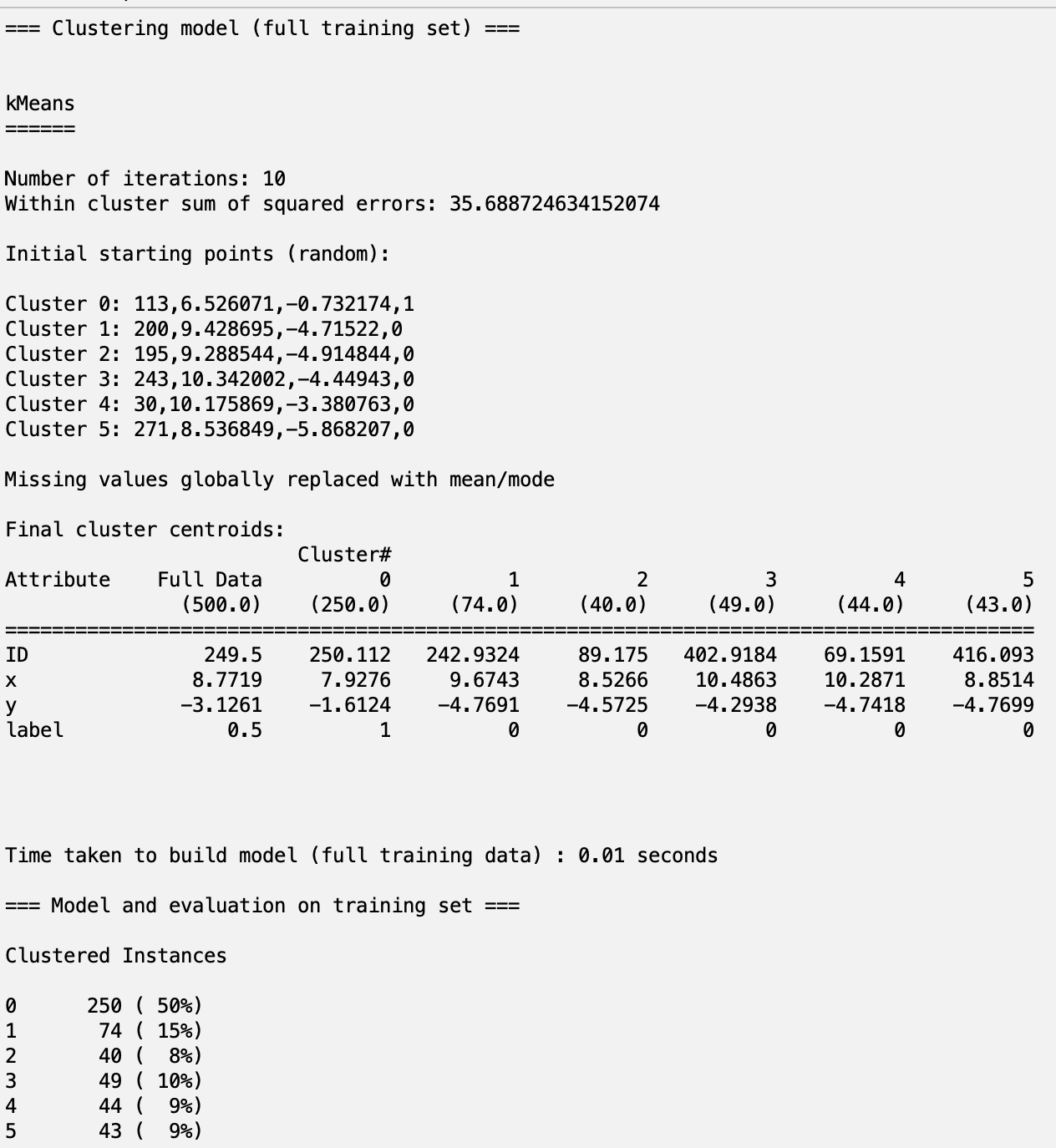


Figure 5 sets the number of clusters to 6. This time, the WCSS result has decreased again by 11 degrees. Might need further adjustment.

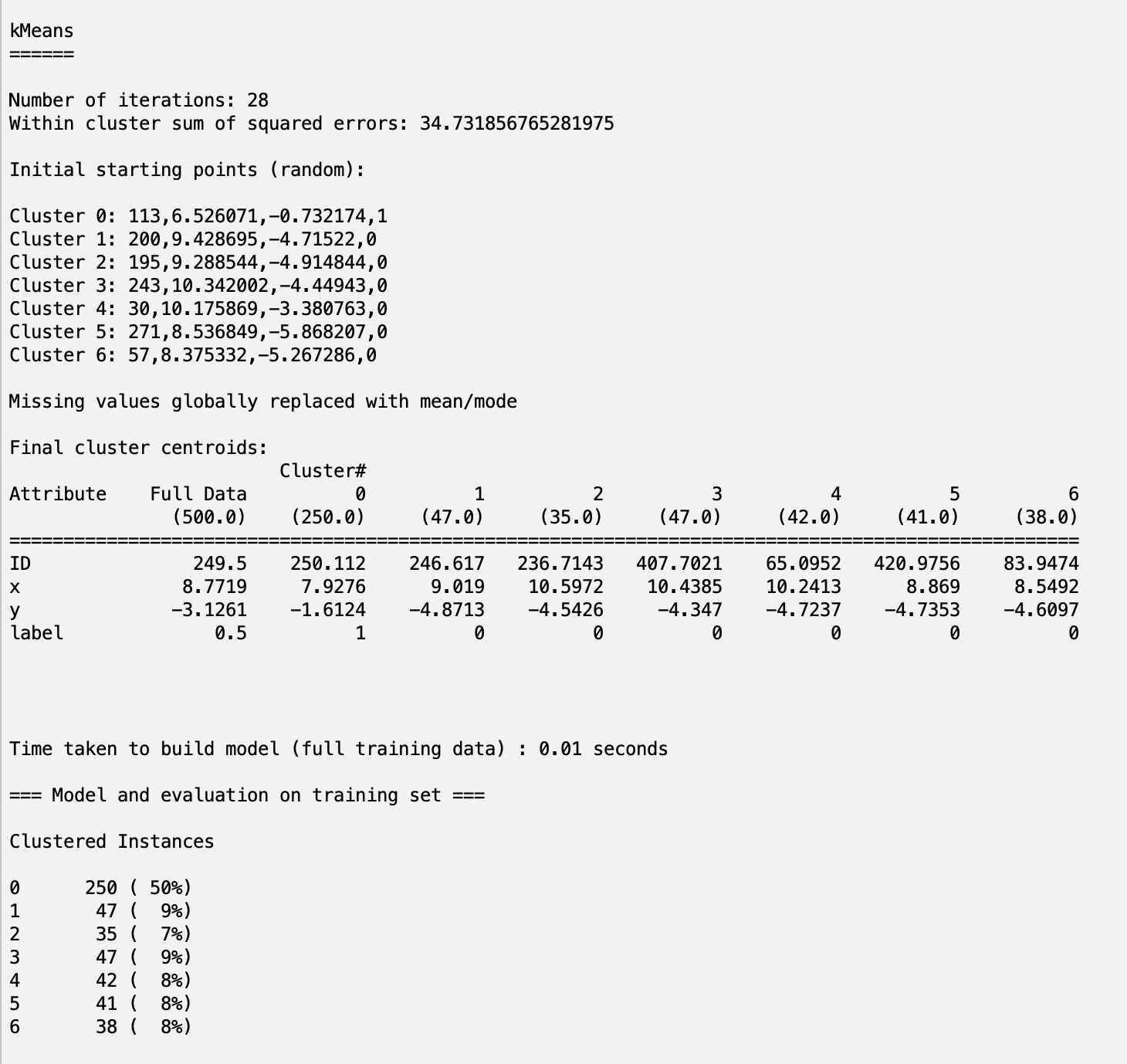


Figure 6 sets the number of clusters to 7. This time, the WCSS result has decreased again by 11 degrees. Might need further adjustment.

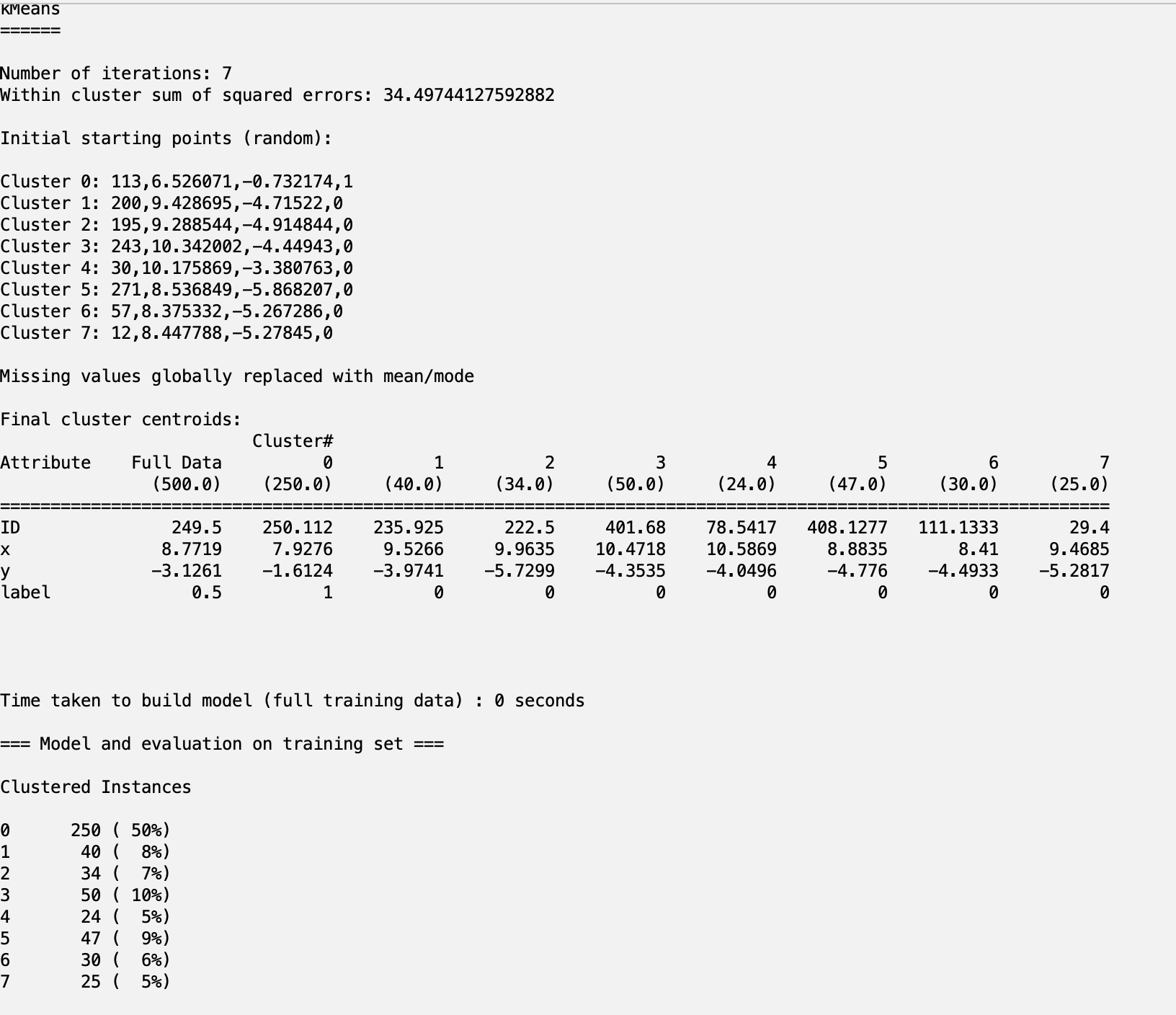


Figure 7 sets the number of clusters to 8. This time, the WCSS result has decreased again by 11 degrees. Might need further adjustment.

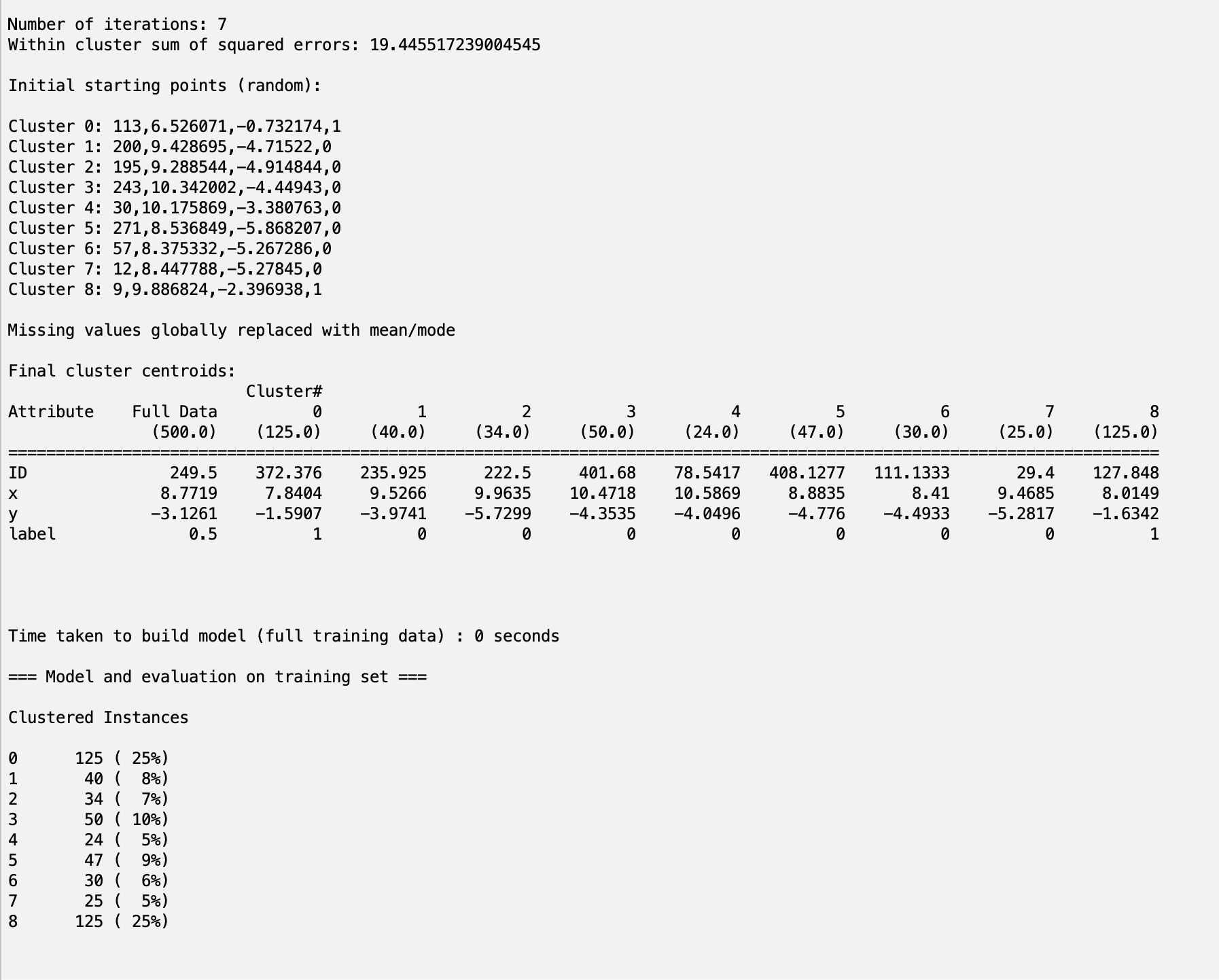
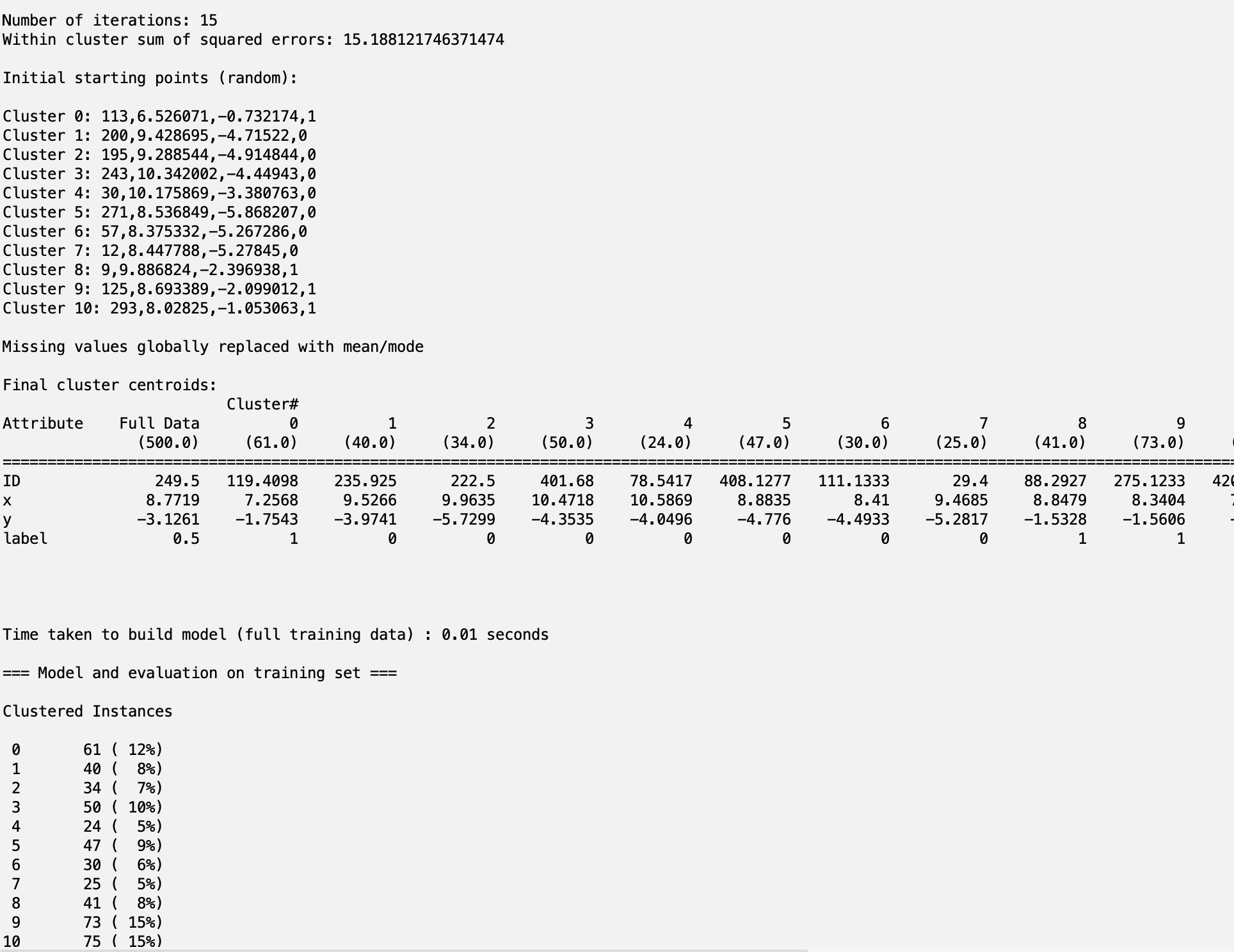
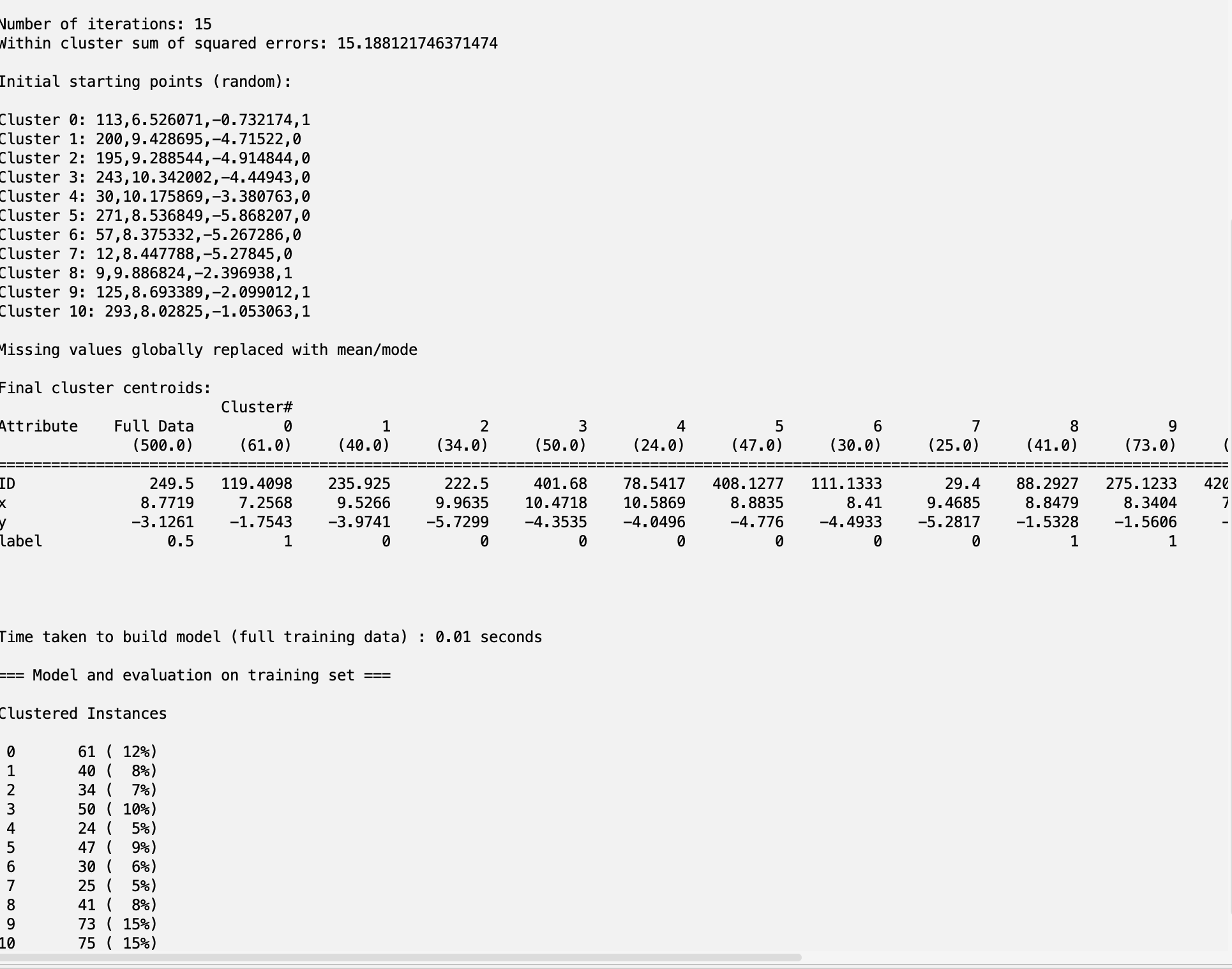
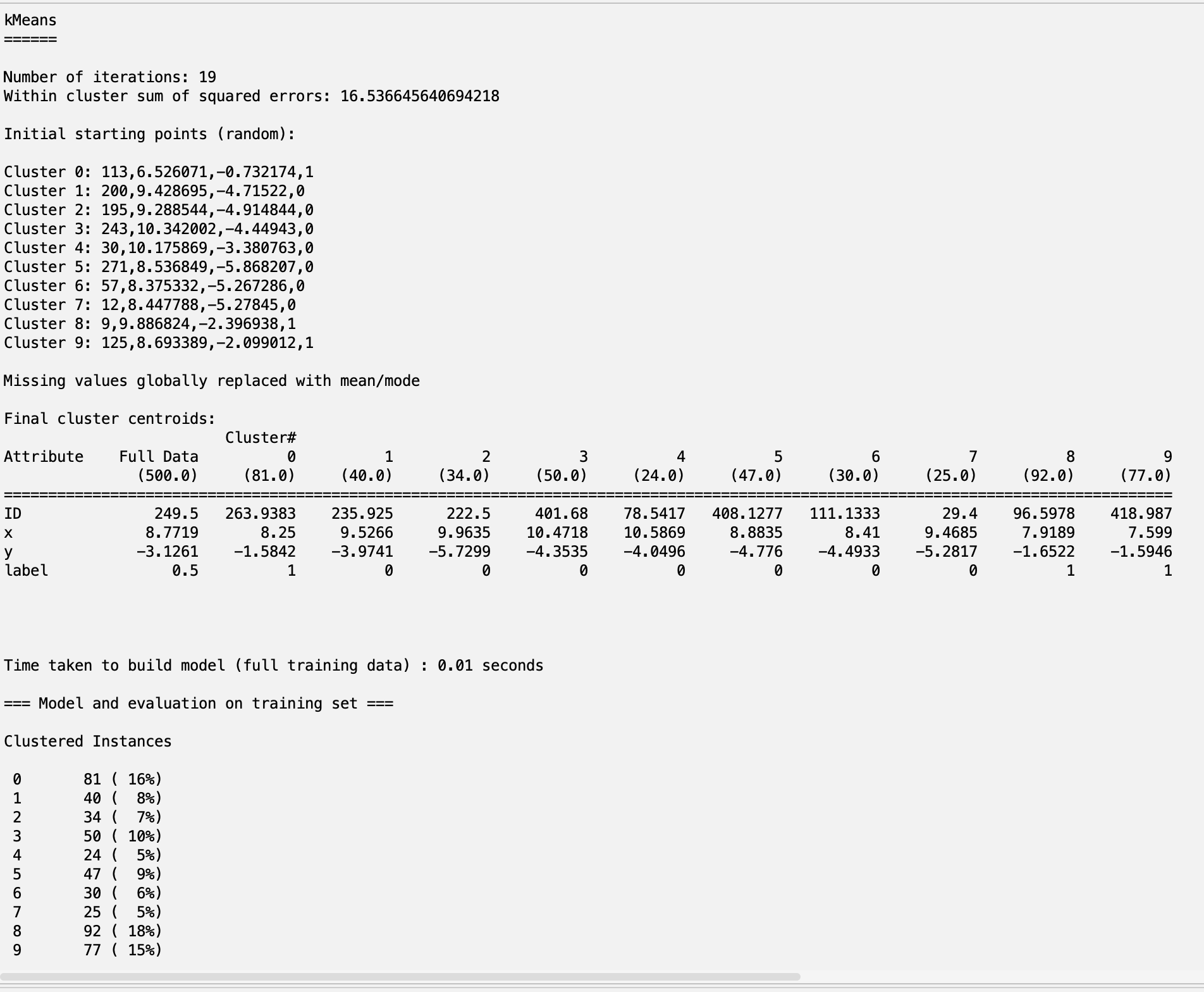


Figure 8 sets the number of clusters to 9. This time, the WCSS result has decreased again by 11 degrees. Might need further adjustment.



1. Here are some of the common evaluation metrics that can be used to test the accuracy of clustering:
   1. Adjusted Rand Index (ARI): measures the similarity between the true labels and the predicted clusters, adjusted for chance agreement. ARI ranges from -1 to 1, where a value of 1 indicates perfect agreement between the two, and a value of 0 indicates no agreement beyond chance.
   2. F1 score: measures the harmonic mean of precision and recall of the predicted clusters with respect to the true labels. F1 score ranges from 0 to 1, where a value of 1 indicates perfect precision and recall.
   3. Silhouette score: measures the quality of clustering by computing the average silhouette coefficient for each data point, which reflects how similar it is to its own cluster compared to other clusters. Silhouette score ranges from -1 to 1, where a value close to 1 indicates good clustering and a value close to -1 indicates poor clustering.
   4. Within-Cluster Sum of Squares (WCSS): measures the total squared distance of each data point to its nearest cluster center, which reflects the compactness of each cluster. A lower WCSS indicates better clustering. This one was used in the Weka tool.

# **References**

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
2. Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.