**CIDM 6355 Exam 2 Part 2 Submission**

**Due 11:59 PM CT Nov 12, 2023 (Total 70 points)**

Requirements: This exam is open book, open slides, and open notes. However, this is an individual exam, so you are not allowed to collaborate nor discuss with anyone else before the due time of the exam. Any question about the exam should be addressed to the instructor. You are required to follow the instruction to complete all the questions and deliverables. You are not allowed to share your RM processes, R scripts, screenshots, or answers with other students or parties; otherwise, such a behavior will be reported to the university authority. In addition, it is your responsibility to make your answers meet the required format; otherwise, you might lose points because of wrong format.

Please read, understand, and comply with these requirements in this exam by typing your name as below.

**Name (First Last) Jordan Unfred**

1. Step 2.3 Take a screenshot of the centroid table with date and time (Screenshot 1). [5 points].

A black text on a white background

Description automatically generatedA screenshot of a computer

Description automatically generated

1. Step 2.6 Take a screenshot of the ANOVA table with date and time (Screenshot 2) and briefly describe your conclusion. Your conclusion must be based on both Steps 2.4 and 2.5. [9 points: 5 pts for your screenshot and 4 pts for your description]

Cluster\_1: 123

Cluster\_2: 110

Cluster\_0: 67

A screenshot of a computer

Description automatically generated

The analysis tells us that there are clear and important differences in the quality ratings of the wines in the three clusters. The way we grouped the wines into clusters has a big impact on how they're rated, meaning that the clusters are not similar when it comes to quality. To figure out which clusters are different from each other, we might need to do more detailed tests which can help us understand which specific groups of wines have very different quality ratings compared to others.

1. Step 3.3 Take a screenshot of your output (cluster size and centroids) with date and time (Screenshot 3) [5 points]

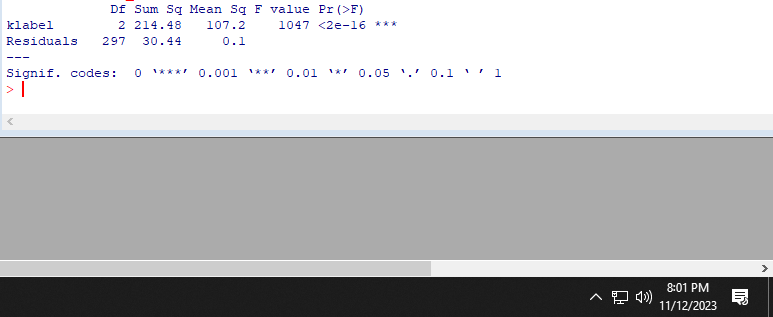
A white background with black text

Description automatically generated

A screenshot of a computer

Description automatically generated

1. Step 3.6 Take a screenshot of the ANOVA table with date and time (Screenshot 4) and briefly describe your conclusion based on the ANOVA table. [9 points: 5 pts for your screenshot and 4 pts for your description]



1. Step 4.3 Take a screenshot of the bar chart with date and time (Screenshot 5) and briefly describe your conclusion. Your conclusion must include each cluster’s size and their average quality ratings. [9 points: 5 pts for your screenshot and 4 pts for your description]

A black text on a white background

Description automatically generated

A screenshot of a computer

Description automatically generated

Averages

Cluster\_1: 0.55

Cluster\_0: 0.49  
Clsuter\_2: 0.72

In our hierarchical clustering analysis performed using RapidMiner, we applied the algorithm to the dataset, resulting in the formation of three distinct clusters. These clusters have been denoted as Cluster 0, Cluster 1, and Cluster 2, with the following respective sizes:

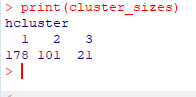
Cluster 0: Comprising 147 items

Cluster 1: Consisting of 83 items

Cluster 2: Encompassing 70 items

The cluster sizes indicate varying quantities of items within each group, with Cluster 0 being the largest and Cluster 2 being the smallest. These findings suggest that our hierarchical clustering model has effectively grouped the data into three clusters, primarily based on the considered attributes.

1. Step 5.5 Take a screenshot of the bar chart with date and time (Screenshot 6) and briefly describe your conclusion. Your conclusion must include each cluster’s size (Step 5.3) and their average quality ratings. [9 points: 5 pts for your screenshot and 4 pts for your description]



A screenshot of a computer

Description automatically generated

In our hierarchical clustering analysis, we applied the algorithm to the dataset resulting in the formation of three clusters. These clusters have been labeled as Cluster 1, Cluster 2, and Cluster 3, and their respective sizes are as follows:

* Cluster 1: 178 items
* Cluster 2: 101 items
* Cluster 3: 21 items

The cluster sizes reveal varying levels of items in each group, with Cluster 1 being the largest and Cluster 3 the smallest. These results suggest that our hierarchical clustering model has grouped the data into three clusters based on the attributes under consideration.

1. Step 6.2 Please compute each pair’s match rate and attach the corresponding screenshot below. You must show how your match rate is computed; your screenshot (e.g., a PivotTable) must show how clusters from each model matched with each other. Your screenshots do not have to show date and time. [24 pts: 2 pts. for each match rate and 2 pts for each screenshot]

Attention: To calculate the match rate, please illustrate how it is computed, as demonstrated in the following example; otherwise, 1 point will be deducted. You should provide a screenshot that clearly displays how each pair of clusters is matched, similar to the example shown; otherwise, 1 point will be deducted.

|  |  |  |
| --- | --- | --- |
| Model Pair | Match Rate | A screenshot to support your match rate |
| Models 1 & 2 |  |  |
| Models 1 & 3 |  |  |
| Models 1 & 4 |  |  |
| Models 2 & 3 |  |  |
| Models 2 & 4 |  |  |
| Models 3 & 4 |  |  |