CIDM 6355 Exam 2 Part 2 Submission

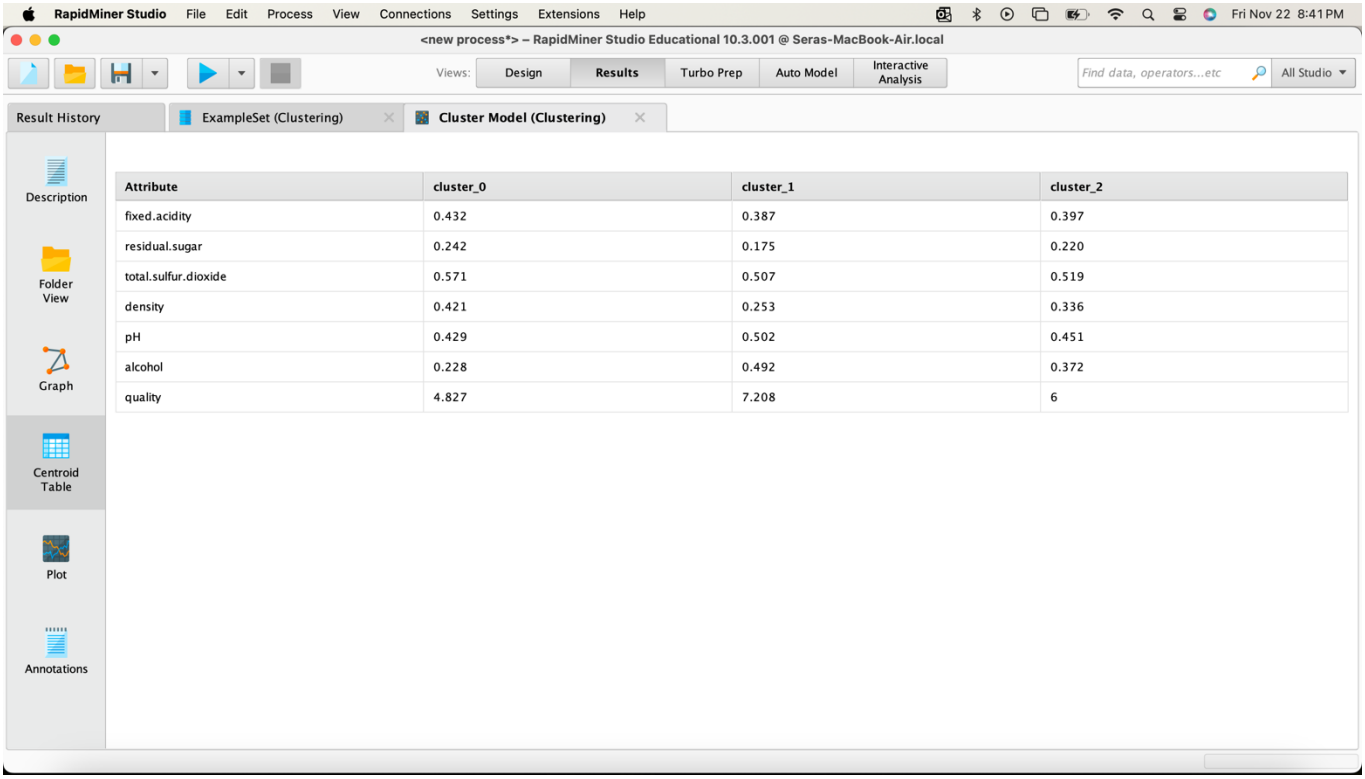
(50 points in total; due 11:59 pm CST, November 25, 2024)

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) Sera Hill

1. Step 2.3 Take a screenshot of the centroid table with date and time (Screenshot 1) and briefly describe each cluster. Your description must include each cluster's size and rating on each attribute (e.g., low density, medium pH, or high alcohol level). [10 points: 4 pts for your screenshot and 6 pts for your description with 2 pts for each cluster]



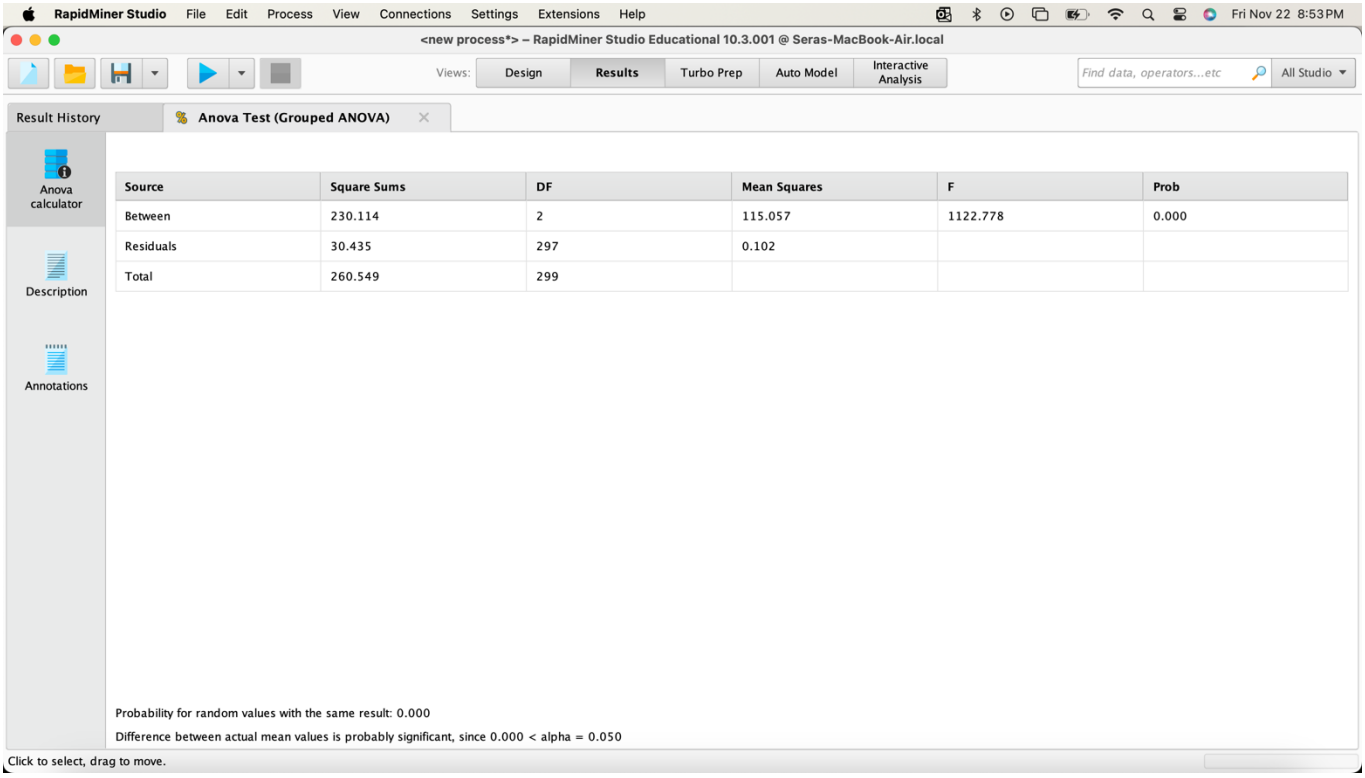
Attribute	cluster_0	cluster_1	cluster_2
fixed.acidity	0.432	0.387	0.397
residual.sugar	0.242	0.175	0.220
total.sulfur.dioxide	0.571	0.507	0.519
density	0.421	0.253	0.336
pH	0.429	0.502	0.451
alcohol	0.228	0.492	0.372
quality	4.827	7.208	6

The size of cluster_0 is 110 with low rating for residual sugar, and alcohol. A medium-low rating for fixed acidity, density, and pH. A medium rating for sulfur dioxide.

The size of cluster_1 is 53 with a low rating for, residual sugar, and density. A medium-low rating for fixed acidity, alcohol. A medium rating for sulfur dioxide, and pH.

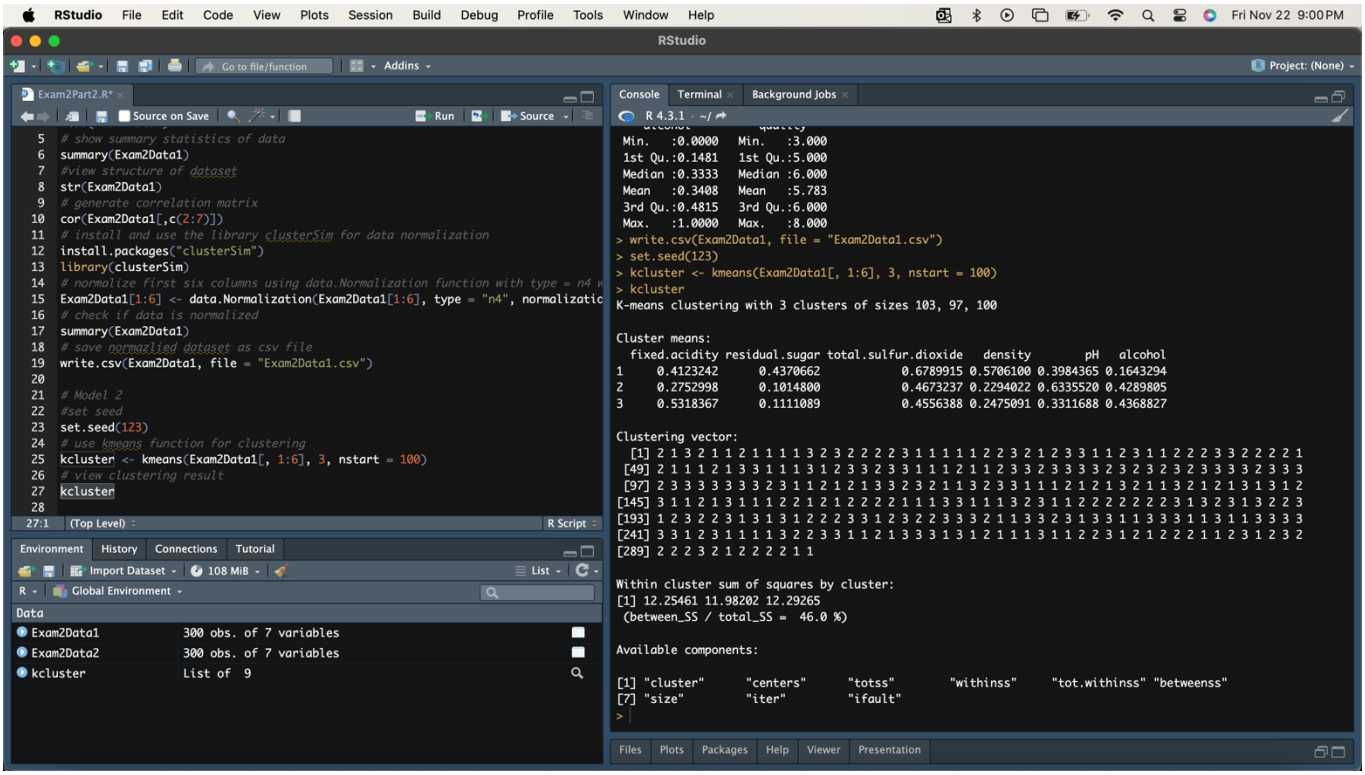
The size of cluster_2 is 137 with a low rating for residual sugar, and density. A medium-low rating for fixed acidity, pH, and alcohol. A medium rating for sulfur dioxide.

2. Step 2.6 Take a screenshot of the ANOVA table with date and time (Screenshot 2) and describe your conclusion. Be sure to include whether the null hypothesis is rejected at the 0.05 level and explain the implications of this result. Your conclusion must be based on both Steps 2.4 and 2.5. [8 points: 4 pts for your screenshot and 4 pts for your description]

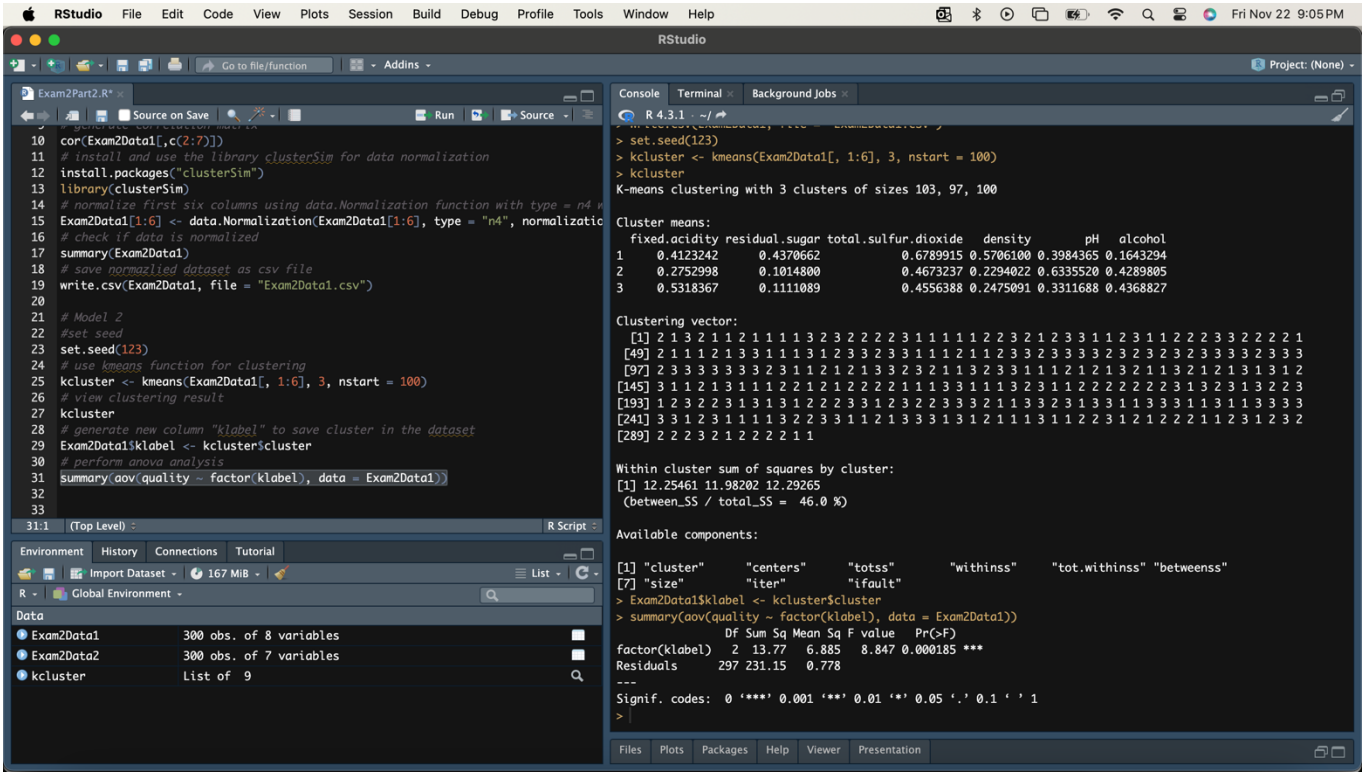


The null hypothesis is rejected at the 0.05 level. This means there is a significant difference between the average quality values between the clusters.

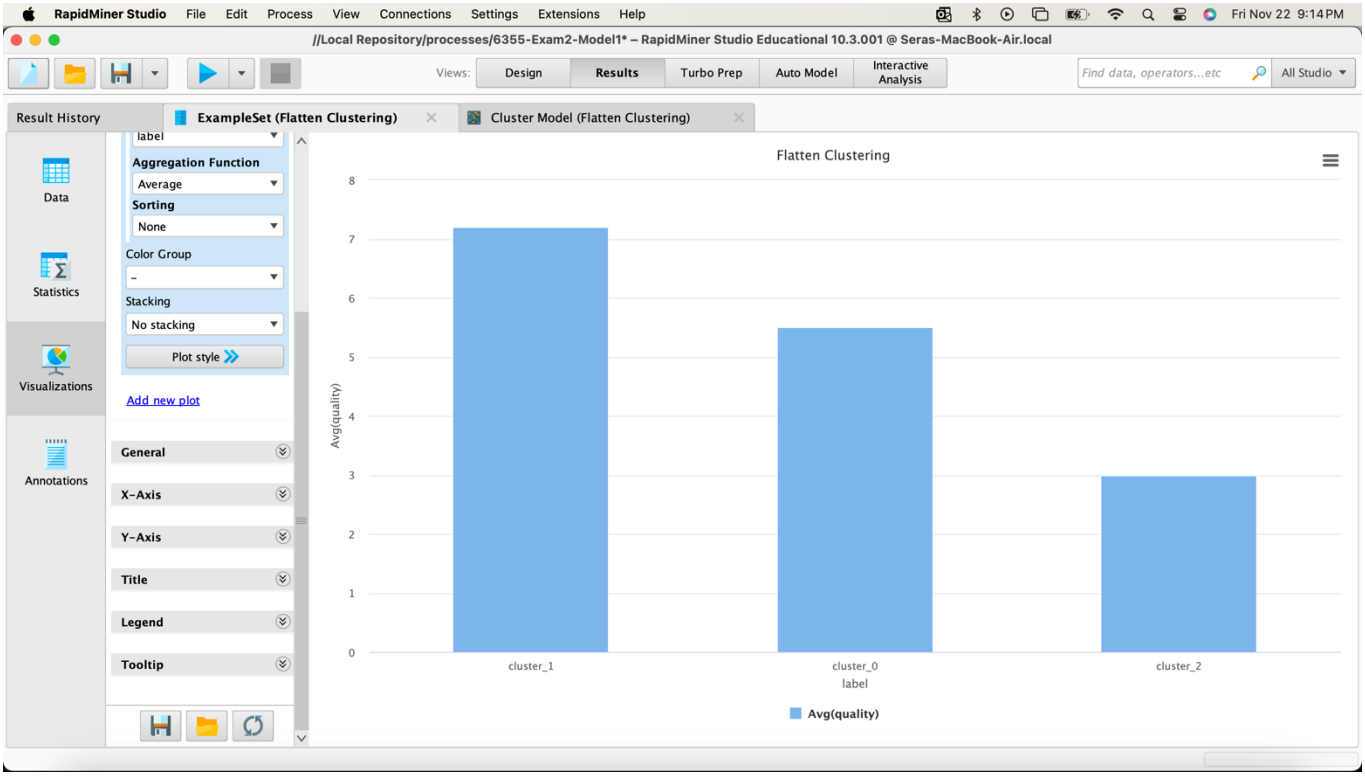
3. Step 3.3 Take a screenshot of your output (cluster size and centroids) with date and time (Screenshot 3) [4 points]



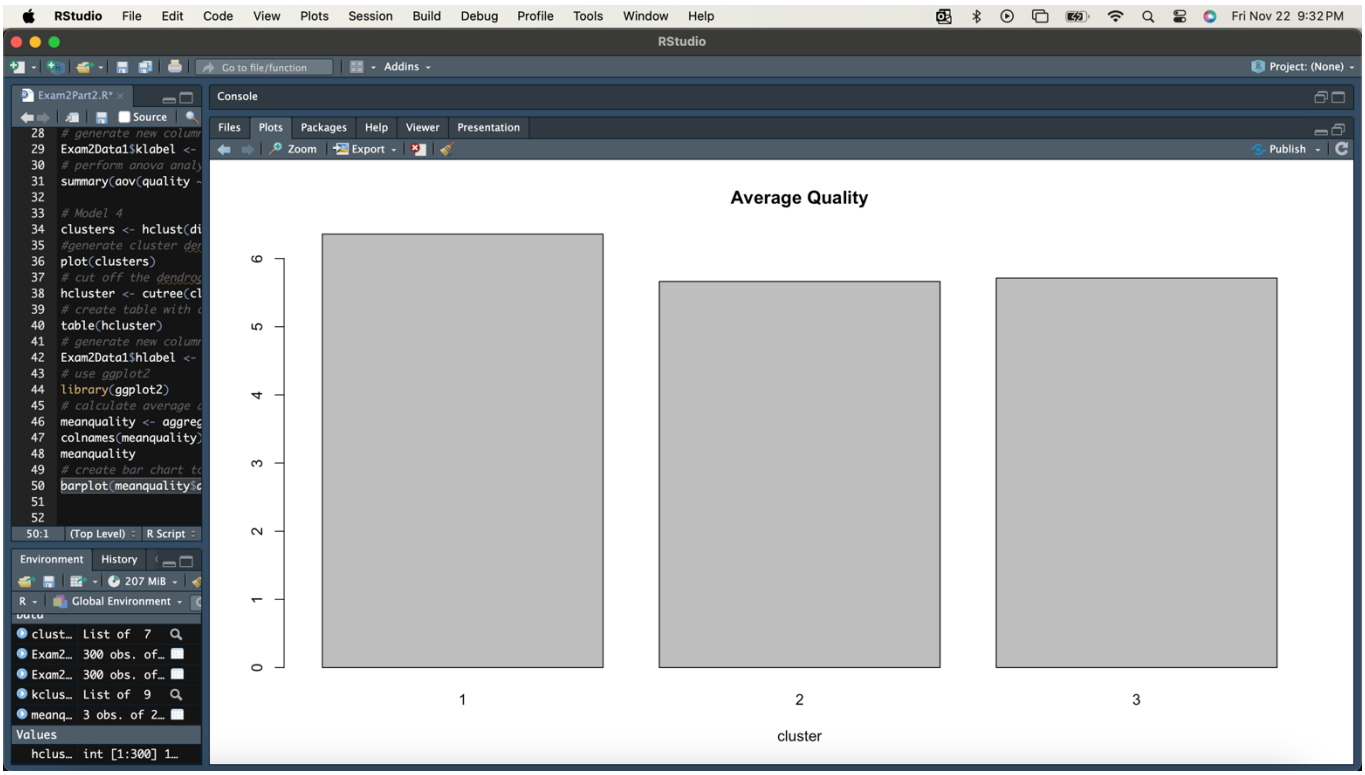
4. Step 3.6 Take a screenshot of the ANOVA table with date and time (Screenshot 4). [4 points]



5. Step 4.3 Take a screenshot of the bar chart with date and time (Screenshot 5). [4 points]



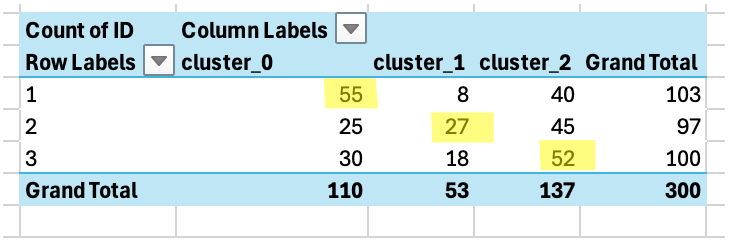
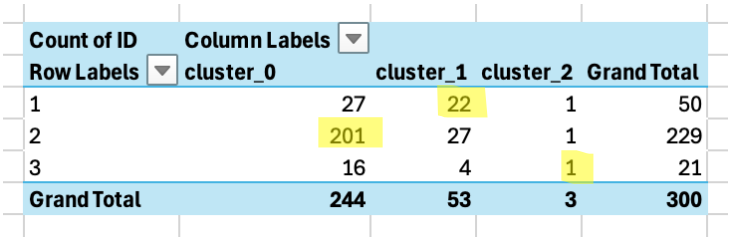
6. Step 5.5 Take a screenshot of the bar chart with date and time (Screenshot 6) and briefly describe your conclusion. Your conclusion must include each cluster's size (Step 5.3) and their average quality ratings. [8 points: 4 pts for your screenshot and 4 pts for your description]



The size of cluster 1 is 50. It has the highest average quality rating compared to the other clusters. The size of cluster 2 is 229. It has the lowest average quality rating.

The size of cluster 3 is 21. It has a mid-level average quality rating, but not by a lot compared to cluster 2.

7. Step 6.2 Please compute the match rate of each pair of models 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. [8 pts: 2 pts. for each match rate and 2 pts for each screenshot].
   In order to calculate the match rate, please explain how it is computed, using the following example provided in Exam 2 Part 2 Instruction as a reference. Failure to do so will result in a zero point on the match rate you provide. Additionally, you must provide a screenshot clearly showcasing how each pair of clusters is matched, similar to the example provided. Failure to provide this will also result in a zero point on each the screenshot you provide.

Model Pair	Match Rate (4 pts)	A screenshot to support your match rate (4 pts)
Models 1 & 2	(55+27+52)/300 = 45% I followed the first method outlined in the appendix in the instructions.	
Models 3 & 4	(201+22+1)/300 = 75% I followed the section method outline in the appendix.	

Count of ID	Column Labels			
Row Labels	cluster_0	cluster_1	cluster_2	Grand Total
1	55	8	40	103
2	25	27	45	97
3	30	18	52	100
Grand Total	**110**	**53**	**137**	**300**

Count of ID	Column Labels			
Row Labels	cluster_0	cluster_1	cluster_2	Grand Total
1	27	22	1	50
2	201	27	1	229
3	16	4	1	21
Grand Total	**244**	**53**	**3**	**300**

8. Compare the results of ANOVA analyses from Steps 2.6, 3.6, 4.4, and 5.6, and then discuss whether they align or differ [4 pts].

   Each ANOVA table from steps 2.6, 3.6, 4.4, and 5.6 show that the null hypothesis can be rejected because the p-value is lower than our alpha of 0.05. Each analysis shows that there is a significant difference between the average quality value between the clusters.