

ASSIGNMENT DETAILS

Unit 8 Assignment: Categorization Analytics

Outcomes addressed in this activity:

Unit Outcomes:

- Analyze categorical situations in data analysis.
- Apply statistical methods to data sets in order to create categories.
- Interpret the results of categorical models.

Course Outcome:


IT527-4: Construct usable and effective data analytics models incorporating industry-recognized software and standard algorithms.

Purpose

This Assignment will enable you to practice categorization analytics in RStudio. Specifically, you will create and then interpret a k-Means model.

Assignment Instructions

The Unit 8 Assignment will give you an opportunity to practice some of the analytics skills you learned in your Reading this week, and also to reflect on that learning. To fulfill the Unit 8 Assignment, complete the following steps:

1. Download the [Heart Disease Risk](#)  comma separated values (CSV) file from Course Documents. Import it into a data frame in RStudio named Patients. In a Word document, use a screenshot to document successful completion of this step and label the screenshot appropriately.

Data Understanding:

- Age: Patient's age in years
- Marital_Status: 0 = Single, Never Married; 1 = Married; 2 = Divorced; 3 = Widowed
- Gender: 0=Female; 1=Male
- Weight_Category: 0=Normal; 1=Overweight; 2=Obese
- Cholesterol: Total cholesterol as measured by milligrams per deciliter of blood
- Stress_Management: 0=Did not attend classes; 1=Did attend classes
- Trait_Anxiety: A score from 0 (never anxious or stressed) to 100 (always anxious or stressed)

1. Create a k-means clustering model on the Patients data frame with four clusters. Show the size and centers for each of your four clusters in a centroid table. Place a screenshot into your Word document and label it. Write an explanation explaining how this size and center table would be interpreted by a data analyst.
2. Create a new data frame called PatientClusters and put the cluster number together with the Patient attributes into it. View this data frame in the RStudio data viewer and take a screenshot. Place it into your Word document and label it. Explain how a data analyst might use the data in this data frame.
3. In your Word document, classify clusters one through four as “Low Risk”, “Moderate Risk”, “High Risk”, or “Critical Risk”. Justify your classifications using the data in your k-means analysis.
4. Based on the data, list at least one thing that you would recommend to patients who fall into the “Critical Risk” category. Defend your recommendation with data from the analysis results.
5. Explain the relationship between Gender and the risk categories you assigned in Step 4 above. Use data from the k-means analysis to defend your explanation.

Assignment Requirements

Prepare your Assignment submission in Microsoft Word following standard APA formatting guidelines: Double spaced, Times New Roman 12-point font, and one-inch margins on all sides. Include a title page, table of contents, and references page. You do not need to write an abstract. Label all tables and figures. Cite sources appropriately both in the text of your writing (parenthetical citations) and on your references page (full APA citation format).

For more information on APA style formatting, go to Academic Writer under Academic Tools of this course.

Also review the university policy on plagiarism. If you have any questions, please contact your professor.

Directions for Submitting Your Assignment

Name your Assignment document according to this convention: Unit8-<yourname>.docx. Submit your completed Assignment to the Unit 8 Dropbox by the deadline.

Review the Unit Assignment Rubric under Course Resources, Grading Rubrics before beginning this activity.

