OFM3 — OFM3 TASK 3: ASSOCIATION RULES AND LIFT ANALYSIS

**DATA MINING II — D212**

**PRFA — OFM3**

COMPETENCIES

**4030.06.6** : **Pattern Prediction**

The graduate predicts patterns in data using association rules and lift analysis.

INTRODUCTION

In this task, you will act as an analyst and create a data mining report. In doing so, you must select one of the data dictionary and data set files to use for your report from the following link: [Data Sets and Associated Data Dictionaries](https://lrps.wgu.edu/provision/227080252).

You should also refer to the data dictionary file for your chosen dataset from the provided link. You will use Python or R to analyze the given data and create a data mining report in a word processor (e.g., Microsoft Word). Throughout the submission, you must visually represent each step of your work and the findings of your data analysis.

*Note: All algorithms and visual representations used need to be captured either in tables or as screenshots added into the submitted word document. A separate Microsoft Excel (.xls or .xlsx) document of the cleaned data should be submitted along with the written aspects of the data mining report.*

SCENARIO

**Scenario 1**

One of the most critical factors in customer relationship management that directly affects a company’s long-term profitability is understanding its customers. When a company can better understand its customer characteristics, it is better able to target products and marketing campaigns for customers, resulting in better profits for the company in the long term.

You are an analyst for a telecommunications company that wants to better understand the characteristics of its customers. You have been asked to perform a market basket analysis to analyze customer data to identify key associations of your customer purchases, ultimately allowing better business and strategic decision-making.

**Scenario 2**

One of the most critical factors in patient relationship management that directly affects a hospital’s long-term cost effectiveness is understanding its patients and the conditions leading to hospital admissions. When a hospital can better understand its patients’ characteristics, it is better able to target treatment to patients, resulting in more effective cost of care for the hospital in the long term.

You are an analyst for a hospital that wants to better understand the characteristics of its patients. You have been asked to perform a market basket analysis to analyze patient data to identify key associations of your patients, ultimately allowing better business and strategic decision-making for the hospital.

REQUIREMENTS

*Your submission must be your original work. No more than a combined total of 30% of the submission and no more than a 10% match to any one individual source can be directly quoted or closely paraphrased from sources, even if cited correctly. The originality report that is provided when you submit your task can be used as a guide.  
  
 You must use the rubric to direct the creation of your submission because it provides detailed criteria that will be used to evaluate your work. Each requirement below may be evaluated by more than one rubric aspect. The rubric aspect titles may contain hyperlinks to relevant portions of the course.*

*Tasks may****not****be submitted as cloud links, such as links to Google Docs, Google Slides, OneDrive, etc., unless specified in the task requirements. All other submissions must be file types that are uploaded and submitted as attachments (e.g., .docx, .pdf, .ppt).*

**Part I: Research Question**

A.  Describe the purpose of this data mining report by doing the following:

1.  Propose **one** question relevant to a real-world organizational situation that you will answer using market basket analysis.

2.  Define **one** goal of the data analysis. Ensure that your goal is reasonable within the scope of the scenario and is represented in the available data.

**Part II: Market Basket Justification**

B.  Explain the reasons for using market basket analysis by doing the following:

1.  Explain how market basket analyzes the selected dataset. Include expected outcomes.

2.  Provide **one** example of transactions in the dataset.

3.  Summarize **one** assumption of market basket analysis.

**Part III: Data Preparation and Analysis**

C.  Prepare and perform market basket analysis by doing the following:

1.  Transform the dataset to make it suitable for market basket analysis. Include a copy of the cleaned dataset.

2.  Execute the code used to generate association rules with the Apriori algorithm. Provide screenshots that demonstrate the error-free functionality of the code.

3.  Provide values for the support, lift, and confidence of the association rules table.

4.  Identify the top **three** rules generated by the Apriori algorithm. Include a screenshot of the top rules along with their summaries.

**Part IV: Data Summary and Implications**

D.  Summarize your data analysis by doing the following:

1.  Summarize the significance of support, lift, and confidence from the results of the analysis.

2.  Discuss the practical significance of the findings from the analysis.

3.  Recommend a course of action for the real-world organizational situation from part A1 based on your results from part D1.

**Part V: Attachments**

E.  Provide a Panopto video recording that includes a demonstration of the functionality of the code used for the analysis and a summary of the programming environment.

*Note: The audiovisual recording should feature you visibly presenting the material (i.e., not in voiceover or embedded video) and should simultaneously capture both you and your multimedia presentation.*

*Note: For instructions on how to access and use Panopto, use the "Panopto How-To Videos" web link provided below. To access Panopto's website, navigate to the web link titled "Panopto Access," and then choose to log in using the “WGU” option. If prompted, log in using your WGU student portal credentials, and then it will forward you to Panopto’s website.*

*To submit your recording, upload it to the Panopto drop box titled “Data Mining I – NVM2.” Once the recording has been uploaded and processed in Panopto's system, retrieve the URL of the recording from Panopto and copy and paste it into the Links option. Upload the remaining task requirements using the Attachments option.*

F.  Record *all* web sources used to acquire data or segments of third-party code to support the application. Ensure the web sources are reliable.

G.  Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased, or summarized.

H.  Demonstrate professional communication in the content and presentation of your submission.

**File Restrictions**

File name may contain only letters, numbers, spaces, and these symbols: ! - \_ . \* ' ( )  
File size limit: 200 MB  
File types allowed: doc, docx, rtf, xls, xlsx, ppt, pptx, odt, pdf, txt, qt, mov, mpg, avi, mp3, wav, mp4, wma, flv, asf, mpeg, wmv, m4v, svg, tif, tiff, jpeg, jpg, gif, png, zip, rar, tar, 7z

#Read and check the data set in Python using Pandas' read\_csv command

#Display Medical dataframe

#Evaluate the data structure to understand input data better

#Re-validate column data types, null and missing value

# Drop records with missing values and Review changes

# Replace empty values with 0

# Prepared data set copy:

# Convert the pandas dataset into list of lists format for use with Apriori algorithm