OFM3 — OFM3 TASK 3: ASSOCIATION RULES AND LIFT ANALYSIS

DATA MINING II — D212
PRFA — OFM3

TASK OVERVIEW | SUBMISSIONS | EVALUATION REPORT

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.

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 II – OFM3." 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

RUBRIC

A1:PROPOSAL OF QUESTION

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not propose 1 question answered using market basket analysis.	The submission proposes 1 question answered using market basket analysis that is not rele-	The submission proposes 1 question answered using market basket analysis that is relevant to a

	vant to a real-world organizational situation.	real-world organizational situation.

A2:DEFINED GOAL

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not define 1 goal for data analysis.	The submission defines 1 goal for data analysis, but the goal is not reasonable, is not within the scope of the scenario, or is not represented in the available data.	The submission defines 1 reasonable goal for data analysis that is within the scope of the scenario and is represented in the available data.

B1:EXPLANATION OF MARKET BASKET

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not explain how market basket analyzes the selected dataset.	The submission does not logically explain how market basket analyzes the selected dataset or includes inaccurate expected outcomes.	The submission logically explains how market basket analyzes the selected dataset and includes accurate expected outcomes.

B2:TRANSACTION EXAMPLE

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not include 1 example of transactions in the dataset.	The submission includes 1 inaccurate example of transactions in the dataset.	The submission includes 1 accurate example of transactions in the dataset.

B3:MARKET BASKET ASSUMPTION

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not summarize 1 assumption of market basket analysis.	The submission inadequately summarizes 1 assumption of market basket analysis.	The submission adequately summarizes 1 assumption of market basket analysis.

C1:TRANSFORMING THE DATASET

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not transform the dataset.	The submission transforms the dataset, but it is not suitable for market basket analysis. Or a copy of the cleaned dataset is not included.	The submission transforms the dataset to make it suitable for market basket analysis and includes a copy of the cleaned dataset.

C2:CODE EXECUTION

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not provide screenshots of the execution of the code used to generate association rules with the Apriori algorithm.	The submission provides screenshots, but 1 or more errors are evident during the execution of the code used to generate association rules with the Apriori algorithm.	The submission executes the code used to generate association rules with the Apriori algorithm and provides screenshots that demonstrate the error-free execution of the code.

C3:ASSOCIATION RULES TABLE

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not include values for the support, lift, or confidence of the association rules table.	The submission includes inaccurate values for the support, lift, or confidence of the association rules table.	The submission includes accurate values for the support, lift, and confidence of the association rules table.

C4:TOP THREE RULES

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not identify *any* rules generated by the Apriori algorithm or include a screenshot.	The submission includes a screenshot but inaccurately identifies 1 or more of the top 3 rules generated by the Apriori algorithm along with their summaries.	The submission includes a screenshot and accurately identifies the top 3 rules generated by the Apriori algorithm along with their summaries.

D1:SIGNIFICANCE OF SUPPORT, LIFT, AND CONFIDENCE SUMMARY

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not summarize the significance of sup-		The submission adequately summarizes the significance of sup-

port, lift, or confidence from the results of the analysis.	The submission inadequately summarizes the significance of support, lift, or confidence from the results of the analysis.	port, lift, and confidence from the results of the analysis.

D2:PRACTICAL SIGNIFICANCE OF FINDINGS

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not discuss the practical significance of the findings from the analysis.	The submission inadequately discusses the practical significance of the findings from the analysis.	The submission adequately discusses the practical significance of the findings from the analysis.

D3:COURSE OF ACTION

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not recommend a course of action for the real-world organizational situation from part A1.	The submission recommends a course of action for the real-world organizational situation from part A1, but the recommendation is inadequate or is not based on the results from part D1.	The submission adequately recommends a course of action for the real-world organizational situation from part A1 based on the results from part D1.

E:PANOPTO RECORDING

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not provide a Panopto video recording.	The submission provides a Panopto video recording, but it does not include a demonstration of the functionality of the code used for the analysis or a summary of the programming environment or *both*.	The submission provides 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.

F:WEB SOURCES

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not record web sources used to acquire	The submission records 1 or more unreliable web sources	The submission records *all* web sources used to acquire data or

data or segments of third-party code.	used to acquire data or segments of third-party code.	segments of third-party code, and the web sources are reliable.

G:SOURCES

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not include both in-text citations and a reference list for sources that are quoted, paraphrased, or summarized.	The submission includes in-text citations for sources that are quoted, paraphrased, or summarized and a reference list; however, the citations or reference list is incomplete or inaccurate.	The submission includes in-text citations for sources that are properly quoted, paraphrased, or summarized and a reference list that accurately identifies the author, date, title, and source location as available.

H:PROFESSIONAL COMMUNICATION

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
Content is unstructured, is disjointed, or contains pervasive errors in mechanics, usage, or grammar. Vocabulary or tone is unprofessional or distracts from the topic.	Content is poorly organized, is difficult to follow, or contains errors in mechanics, usage, or grammar that cause confusion. Terminology is misused or ineffective.	Content reflects attention to detail, is organized, and focuses on the main ideas as prescribed in the task or chosen by the candidate. Terminology is pertinent, is used correctly, and effectively conveys the intended meaning. Mechanics, usage, and grammar promote accurate interpretation and understanding.

WEB LINKS

Data Sets and Associated Data Dictionaries
If you have trouble with the link, copy and paste the link directly into your web browser.

Panopto Access
Sign in using the "WGU" option. If prompted, log in with your WGU student portal credentials, which should forward you to Panopto's website. If you have any problems accessing Panopto, please contact Assessment Services at assessmentservices@wgu.edu. It may take up to two business days to receive your WGU Panopto recording permissions once you have begun the course.

Panopto How-To Videos