OFM4 — OFM4 TASK 3: ASSOCIATION RULES AND LIFT ANALYSIS

DATA MINING II — D212

PRFA — OFM4

Task Overview   Submissions   Evaluation Report

COMPETENCIES

4030.6.6 : **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. You must select one of the data dictionary and data set files to use for your report from the following web link: "Data Sets and Associated Data Dictionaries."

You should also refer to the data dictionary file for your chosen data set from the above 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 the customers. When a company can better understand its customers' 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 enabling 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 the patients and the conditions leading to hospital admissions. When a hospital understands its patients' characteristics, it is better able to target treatment to patients, resulting in a 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 enabling 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 similarity 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 your 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 your goal is reasonable within the scope of the selected 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 data set. Include expected outcomes.
   2. Provide **one** example of transactions in the data set.
   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 data set to make it suitable for market basket analysis. Include a copy of the cleaned data set.
   2. Execute the code used to generate association rules with the Apriori algorithm. Provide screenshots that demonstrate that the code is error free.
   3. Provide values for the support, lift, and confidence of the association rules table.
   4. Explain the top **three** relevant rules generated by the Apriori algorithm. Include a screenshot of the top three relevant rules.

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

Part V: Attachments

E. Provide a Panopto video recording that includes the presenter and a vocalized demonstration showing all code used, the code being executed, and the results of all code used in the task.

1. Include the presenter and a vocalized demonstration describing the programs used to complete this task in the Panopto video recording.

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 – OFM4" 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 you 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, csv, 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.	The submission proposes 1 question that cannot be answered using market basket analysis. Or the question is not relevant to a real-world organizational situation.	The submission proposes 1 question that can be answered using market basket analysis. The question is relevant to a real-world organizational situation.

A2:DEFINED GOAL

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not define 1 goal of the data analysis.	The submission defines 1 goal of the data analysis, but the goal is not reasonable, is not within the scope of the selected scenario, or is not represented in the available data.	The submission defines 1 reasonable goal of the data analysis that is within the scope of the selected 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 data set.	The submission does not logically explain how market basket analyzes the selected data set or does not include expected outcomes.	The submission logically explains how market basket analyzes the selected data set and includes expected outcomes.

B2:TRANSACTION EXAMPLE

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

B3:MARKET BASKET ASSUMPTION

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not sum-		The submission accurately sum-

marize 1 assumption of market basket analysis.

The submission inaccurately summarizes 1 assumption of market basket analysis.

marizes 1 assumption of market basket analysis.

C1:TRANSFORMING THE DATA SET

NOT EVIDENT

The submission does not transform the data set.

APPROACHING COMPETENCE

The submission transforms the data set, but it is not suitable for market basket analysis. Or a copy of the cleaned data set is not included.

COMPETENT

The submission transforms the data set to make it suitable for market basket analysis and includes a copy of the cleaned data set.

C2:CODE EXECUTION

NOT EVIDENT

The submission does not provide screenshots of the execution of the code used to generate association rules with the Apriori algorithm.

APPROACHING COMPETENCE

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.

COMPETENT

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

The submission does not include values for the support, lift, or confidence of the association rules table.

APPROACHING COMPETENCE

The submission includes inappropriate values for the support, lift, or confidence of the association rules table.

COMPETENT

The submission includes appropriate values for the support, lift, and confidence of the association rules table.

C4:TOP THREE RULES

NOT EVIDENT

The submission does not explain *any* rules generated by the Apriori algorithm or does not include a screenshot.

APPROACHING COMPETENCE

The submission includes a screenshot but inaccurately explains 1 or more of the top 3 relevant rules generated by the

COMPETENT

The submission includes a screenshot of the top 3 relevant rules and accurately explains the top 3 relevant rules generated by the Apriori algorithm.

Apriori algorithm.

D1:SIGNIFICANCE OF SUPPORT, LIFT, AND CONFIDENCE SUMMARY

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not summarize the significance of support, lift, or confidence from the results of the analysis.	The submission inaccurately summarizes the significance of support, lift, or confidence from the results of the analysis.	The submission accurately summarizes the significance of support, 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 inaccurately discusses the practical significance of the findings from the analysis.	The submission accurately 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 inaccurate or is not based on the results from part D1.	The submission accurately recommends a course of action for the real-world organizational situation from part A1 based on the results from part D1.

E:PANOPTO VIDEO OF CODE

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
A Panopto video recording of the code used is not provided, or the link provided for the video is not functional.	A Panopto video recording is provided, but a full demonstration of the code used, the code being executed, or the results of the code used in the task is not provided, or the video does not capture both the presenter	A Panopto video recording is provided that includes a full demonstration of the code used, the code being executed, and the results of the code used in the task. For the duration of the presentation, the video captures both the presenter and

and the vocalized demonstration.

the vocalized demonstration.

E1:PANOPTO VIDEO OF PROGRAMS

NOT EVIDENT

A Panopto video recording of the programs used is not provided.

APPROACHING COMPETENCE

A Panopto video recording is provided, but a complete description of the programs used to complete the task is not provided, or the video does not capture both the presenter and the vocalized presentation describing the programs used to complete the task.

COMPETENT

A Panopto video recording is provided that includes a complete description of the programs used to complete the task. For the duration of the presentation, the video captures both the presenter and the vocalized presentation describing the programs used to complete the task.

F:SOURCES FOR THIRD-PARTY CODE

NOT EVIDENT

The submission does not record web sources used to acquire data or segments of third-party code.

APPROACHING COMPETENCE

The submission records 1 or more unreliable web sources used to acquire data or segments of third-party code.

COMPETENT

The submission records *all* web sources used to acquire data or segments of third-party code, and the web sources are reliable.

G:SOURCES

NOT EVIDENT

The submission does not include both in-text citations and a reference list for sources that are quoted, paraphrased, or summarized.

APPROACHING COMPETENCE

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.

COMPETENT

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

This submission includes professional communication er-

APPROACHING COMPETENCE

This submission includes pro-

COMPETENT

This submission demonstrates correct use of spelling, grammar,

rors related to spelling, grammar, punctuation, and sentence fluency. For best results, please focus on the specific Correctness errors identified by Grammarly for Education to help guide your revisions. If you need additional assistance preparing your submission, please contact your Instructor.

fessional communication errors related to spelling, grammar, punctuation, and/or sentence fluency. For best results, please focus on the specific Correctness errors identified by Grammarly for Education to help guide your revisions.

punctuation, and sentence fluency. You have demonstrated quality professional communication skills in this submission.

WEB LINKS

Data Sets and Associated Data Dictionaries
If you have trouble with the link, copy and paste the URL 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