Competencies

**4159.1.1** : **Profiles Data**

The learner interprets a data dictionary to understand the data set.

**4159.1.2** : **Interprets Statistics and Visualization**

The learner interprets probability, descriptive and inferential statistics, and visualization.

**4159.1.3** : **Wrangles Data**

The learner wrangles data to ensure accuracy, format, and integrity relevant to the task being performed.

Introduction

Throughout your career in data analysis, you will be asked to identify patterns, trends, and correlations that offer valuable insights into a company’s business performance and customer preferences.

In this task, you will serve as a data analyst who has been asked to prepare and explore transactional data for an international retail store named Allias Megastore. This company wants to better understand its customers’ characteristics so it can more effectively target products and marketing campaigns toward its shoppers—thus resulting in improved profits for the company in the long term.

You will be given a dataset and an overview, including a data dictionary, and will be asked to perform market basket analysis to identify key associations of customer purchases—ultimately ensuring better business and strategic decision-making.

*Note: The IDE for this assessment is either Anaconda or RStudio, depending on which language you decide to use to complete the task. Please use the “WGU Virtual Lab Environment” web link below.*

Scenario

Refer to the scenario and information in the "Megastore Considerations and Data Dictionary" and "Megastore Dataset" supporting documents.

Requirements

Your submission must represent your original work and understanding of the course material. Most performance assessment submissions are automatically scanned through the WGU similarity checker. Students are strongly encouraged to wait for the similarity report to generate after uploading their work and then review it to ensure Academic Authenticity guidelines are met before submitting the file for evaluation. See [Understanding Similarity Reports](https://cm.wgu.edu/t5/Frequently-Asked-Questions/Understanding-Similarity-Reports/ta-p/252) for more information.    
  
**Grammarly Note:**   
Professional Communication will be automatically assessed through Grammarly for Education in most performance assessments before a student submits work for evaluation. Students are strongly encouraged to review the Grammarly for Education feedback prior to submitting work for evaluation, as the overall submission will not pass without this aspect passing. See [Use Grammarly for Education Effectively](https://cm.wgu.edu/t5/Academic-Coaching-Center/Use-Grammarly-for-Education-Effectively/ta-p/52276) for more information.    
  
**Microsoft Files Note:**  
Write your paper in Microsoft Word (.doc or .docx) unless another Microsoft product, or pdf, is specified in the task directions. Tasks may not be submitted as cloud links, such as links to Google Docs, Google Slides, OneDrive, etc.  All supporting documentation, such as screenshots and proof of experience, should be collected in a pdf file and submitted separately from the main file. For more information, please see [Computer System and Technology Requirements.](https://cm.wgu.edu/t5/WGU-Student-Policy-Handbook/Computer-System-and-Technology-Requirements/ta-p/78)    
 *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.*

**Part I: Research Question**

*Note: Your responses to the following task prompts must be provided in a document file. Unless otherwise specified, responses to PA requirements that are included in a Python or RStudio notebook will****not****be accepted.*

A.  Describe the purpose of your 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 provided scenario and is represented in the available data.

**Part II: Market Basket Justification**

*Note: Your responses to the following task prompts must be provided in a document file. Unless otherwise specified, responses to PA requirements that are included in a Python or RStudio notebook will****not****be accepted.*

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

1.  Explain how the market basket technique analyzes the provided dataset, including expected outcomes.

2.  Provide **one** example of a transaction in the dataset.

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

**Part III: Data Preparation and Analysis**

*Note: Your responses to the following task prompts must be provided in a document file. Unless otherwise specified, responses to PA requirements that are included in a Python or RStudio notebook will****not****be accepted.*

C.  Prepare the dataset for further analysis by doing the following:

1.  Wrangle (i.e., transform) data by doing the following:

a.  Select *x* number of categorical variables, choosing *at least* **two** ordinal variables and *at least* **two** nominal variables.

b.  Perform the appropriate encoding method (ordinal, label encoding, one-hot encoding) for *each* variable selected in part C1a.

c.  Transactionalize the data for market basket analysis.

d.  Explain and justify *each* step you took in parts C1a, C1b, and C1c.

2.  Include a copy of the cleaned dataset.

3.  Execute the code used to generate association rules with the Apriori algorithm. Provide a screenshot that demonstrates that the code is error-free.

4.  Provide values for the support, lift, and confidence of the association rules table. Include a screenshot of the values.

5.  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**

*Note: Your responses to the following task prompts must be provided in a document file. Unless otherwise specified, responses to PA requirements that are included in a Python or RStudio notebook will****not****be accepted.*

D.  Summarize your data analysis by doing the following:

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

2.  Explain the practical significance of your findings from the analysis.

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