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 analysis by doing the following:

1.  Summarize **one** research question that is relevant to a real-world organizational situation captured in the data set you have selected and that you will answer using logistic regression.

2.  Define the objectives or goals of the data analysis. Ensure that your objectives or goals are reasonable within the scope of the data dictionary and are represented in the available data.

**Part II: Method Justification**

B.  Describe logistic regression methods by doing the following:

1.  Summarize the assumptions of a logistic regression model.

2.  Describe the benefits of using the tool(s) you have chosen (i.e., Python, R, or both) in support of various phases of the analysis.

3.  Explain why logistic regression is an appropriate technique to analyze the research question summarized in Part I.

**Part III: Data Preparation**

C.  Summarize the data preparation process for logistic regression by doing the following:

1.  Describe your data preparation goals and the data manipulations that will be used to achieve the goals.

2.  Discuss the summary statistics, including the target variable and *all* predictor variables that you will need to gather from the data set to answer the research question.

3.  Explain the steps used to prepare the data for the analysis, including the annotated code.

4.  Generate univariate and bivariate visualizations of the distributions of variables in the cleaned data set. Include the target variable in your bivariate visualizations.

5.  Provide a copy of the prepared data set.

**Part IV: Model Comparison and Analysis**

D.  Compare an initial and a reduced logistic regression model by doing the following:

1.  Construct an initial logistic regression model from *all* predictors that were identified in Part C2

2.  Justify a statistically based variable selection procedure and a model evaluation metric to reduce the initial model in a way that aligns with the research question.

3.  Provide a reduced logistic regression model.

*Note: The output should include a screenshot of each model.*

E.  Analyze the data set using your reduced logistic regression model by doing the following:

1.  Explain your data analysis process by comparing the initial and reduced logistic regression models, including the following elements:

•  the logic of the variable selection technique

•  the model evaluation metric

2.  Provide the output and *any* calculations of the analysis you performed, including a confusion matrix.

*Note: The output should include the predictions from the refined model you used to perform the analysis.*

3.  Provide the code used to support the implementation of the logistic regression models.

**Part V: Data Summary and Implications**

F.  Summarize your findings and assumptions by doing the following:

1.  Discuss the results of your data analysis, including the following elements:

•  a regression equation for the reduced model

•  an interpretation of coefficients of the statistically significant variables of the model

•  the statistical and practical significance of the model

•  the limitations of the data analysis

2.  Recommend a course of action based on your results.

**Part VI: Demonstration**

G.  Provide a Panopto video recording that includes *all* of the following elements:

•  a demonstration of the functionality of the code used for the analysis

•  an identification of the version of the programming environment

•  a comparison of the **two** logistic regression models you used in your analysis

•  an interpretation of the coefficients

*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 “Logistic Regression Modeling – NBM2 | D208.” 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.*

H.  List the web sources used to acquire data or segments of third-party code to support the application. Ensure the web sources are reliable.

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

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