**Part I: Research Question**

A.  Describe **one** question or decision that you will address using the data set you chose. The summarized question or decision must be relevant to a realistic organizational need or situation.

* Identify the which service(s) has the highest churn rate

B.  Describe the variables in the data set and indicate the specific type of data being described. Use examples from the data set that support your claims.

**Part II: Data-Cleaning Plan**

C.  Explain the plan for cleaning the data by doing the following:

1.  Propose a plan that includes the relevant techniques and specific steps needed to identify anomalies in the data set.

2.  Justify your approach for assessing the quality of the data, include:

•  characteristics of the data being assessed,

•  the approach used to assess the quality.

3.  Justify your selected programming language and any libraries and packages that will support the data-cleaning process.

4.  Provide the code you will use to identify the anomalies in the data.

**Part III: Data Cleaning**

D.  Summarize the data-cleaning process by doing the following:

1.  Describe the findings, including all anomalies, from the implementation of the data-cleaning plan from part C.

2.  Justify your methods for mitigating each type of discovered anomaly in the data set.

3.  Summarize the outcome from the implementation of each data-cleaning step.

4.  Provide the code used to mitigate anomalies.

5.  Provide a copy of the cleaned data set.

6.  Summarize the limitations of the data-cleaning process.

7.  Discuss how the limitations in part D6 affect the analysis of the question or decision from part A.

E.  Apply principal component analysis (PCA) to identify the significant features of the data set by doing the following:

1.  List the principal components in the data set.

2.  Describe how you identified the principal components of the data set.

3.  Describe how the organization can benefit from the results of the PCA

**Part IV. Supporting Documents**

F.  Provide a Panopto recording that demonstrates the warning- and error-free functionality of the code used to support the discovery of anomalies and the data cleaning process and summarizes the programming environment.

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 Cleaning – NUM2 \ D206” 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.

G.  Reference the web sources used to acquire segments of third-party code to support the application. Be sure the web sources are reliable.

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

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