DATA GLACIER VIRTUAL INTERNSHIP

CROSS SELLING RECOMMENDATION-GROUP PROJECT

WEEK 9: DELIVERABLES

GROUP NAME: HEGY

Team members:

Name: B. Harika
Email: harikabreddy444@gmail.com
Country: India
College/ Company: Data Glacier
Specialization: Data Analyst

Name: Yusuf Yuhan
Email: yusufyuhan98.yy@gmail.com
Country: Srilanka
College/ Company: The Open University of Srilanka
Specialization: Data Analyst

Name: Ebaghae Imhanlahimi
Email: imhanlahimiw@gmail.com
Country: America
College/ Company: Data Glacier
Specialization: Data Analyst

Name: Gladys Kalas
Email: gladys@kalas.me
Country: USA
College/ Company: Data Glacier
Specialization: Data Analyst

Peers Comments on Code:

Gladys Kalas comments on team members code:

@HarikaReddyB

- A clean data set with translated columns names from Spanish to English.
  data free from duplications.
- Detected and treated the null values using Median method, using dropna function to remove
  few column values that were missing important information, and Fillna methods.
- Checked for the outliers detection using the Standard Deviation method.
  Treated the outliers using .clip method.

@Ebaghae

- Cleaned data changing column names from Spanish to English interpretation.
- Deleted null values of categorical and demographic columns with sensitive information.
- Filled some columns null values with the mode of the column.
- Replaced the Spanish values in columns (Gender, Customer type, Customer relation, Employee index, Residence index, Foreigner index, Deceased index, etc.) with their English interpretations.
- Replaced categorical columns values (Activity index and segmentation) with their equivalent column values interpretations. Changed data types of columns as required to save memory.
- Checked and treated for outliers in the data.

@mlyusufyuhan

- Obtained a clean data set featuring checking of null values, imputing null values with forward fill , backward fill and mode accordingly.
- Changing data types. Detecting the outliers in the data with IQR and upper and lower caps ,by capping using lambda function

B. Harika Comments on team members code:

@mlyusufyuhan

- Yusuf has worked on cleaning the code, Translated the columns name, and imputing the missing values using Mode, Forwardfill, and Backwardfill methods.
- Worked on Detecting the outliers using the IQR method and Treating them using the Median.

@onlygladys

- Gladys worked on translating the columns and assigned the columns variables.
- Checked on each column for unique values, and treated the null and missing values using dropna( ), fillna( ), and Mean methods.
- Detection of outliers using Visualization, and IQR methods, and treating the outliers using the Capping method.

@Ebaghae

- Ebaghae worked on treating the Null values using Mode, and Fillna( ), Dropna( ) for the columns which have sensitive details.
- Replaced the values of categorical columns with their respective interpretations.
- Converted the data types to save memory, and Detection and Treatment of outliers using the percentile method.

Ebaghae Imhanlahimi comments on team members code:

@HarikaReddyB

- Your notebook is well organized and well detailed.
- I commend the way you broke down your codes making it easy to understand.
- I noticed that when detecting and treating outliers, you clipped some important categorical columns outliers (Age and Gross income). I don't know if this is the best approach to treating their outliers, as some sensitive data which is relevant for analysis may have been lost.
- Overall, your data looks well transformed and cleaned based on your work so far.

@onlygladys

- Your work is neatly organized.
- You properly checked for errors and resolved them.
- I however noticed that when treating your outliers, you trimmed and capped some sensitive columns (Age, and Gross income). These values that were trimmed and capped may be relevant for analysis.
- overall, your notebook is very good.

@mlyusufyuhan

- Your work is neatly organized.
- You detected problems in the data and treated them.
- Your work does not show any code that replaced the values in specific columns with their appropriate interpretation. This interpretation would have been relevant for analysis.
- Overall your notebook is neat and good.

Yusuf Yuhan Comments on team members code:

@onlygladys:

- Great job on translating and assigning variables to the columns! That must have taken some time.
- It's good to see that you checked for unique values in each column and treated the missing values using various methods. That will help ensure the data is clean.
- Nice work using visualization and the IQR method to detect outliers, and the capping method to treat them. That's a robust approach to handling outliers.

@Ebaghae

- It's impressive that you were able to handle sensitive details in the data by using appropriate methods to treat null values.
- Good job replacing categorical column values with their respective interpretations - that will make the data more interpretable.
- It's always helpful to save memory when working with large datasets, so it's great to see that you converted the data types. And using the percentile method to detect and treat outliers is a solid approach.

@HarikaReddyB

- Great job on the data preprocessing tasks, @harika!
- I noticed that you took charge of changing data types, which is essential for optimizing memory usage.
- You also did a great job in detecting and treating outliers, which can significantly impact the accuracy of the analysis.
- I appreciated how you were able to replace and change values in columns, especially for categorical variables, which is crucial in making sense of the data.
- treating null values is a crucial step, and I'm glad you were able to handle it well. Great work!

GitHub Repo Link:

https://github.com/HarikaReddyB/Cross_selling_recommendation---Group_Project/tree/main/Week%209