DataGlacier Week 10 Deliverables
Project: Cross Selling Recommendations
Group Name: MacroStaff
Group Specialization: Data Analyst

Group Members:

1. Wasiq Ahmed:
   Email: wasiqahmed7@gmail.com

   Country: Turkey

   University: Bilkent University

Problem Description

A company called XYZ Credit Union currently has data about its customers and their purchased banking products. Since this data is highly clustered and not properly formatted, the company needs assistance in cleaning the data of potential invalid or repeated entries.

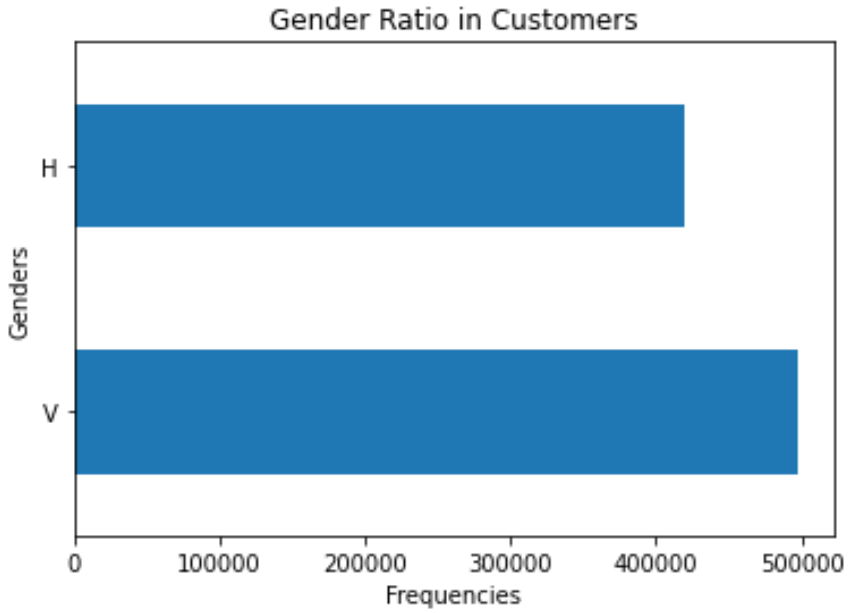
Github Repository

https://github.com/Wasiq147/DataGlacier-proejct.git

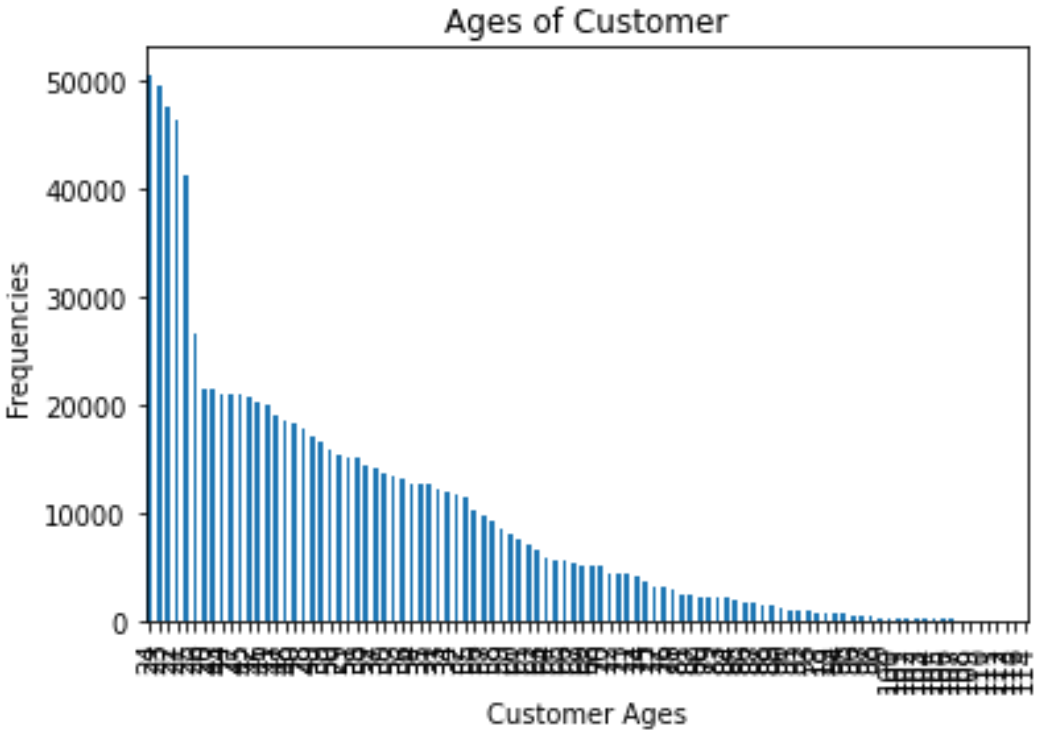
EDA Performed on Data

Once the data was cleaned using the different cleaning methods, we progressed to perform EDA techniques on the data and represent it using visual tools. Using python, we represented data in form of bar graphs to better observe potential trends or have a better analysis on the statistics.

From performing the EDA on the data from ‘Test.csv’ we observed that in the customers to the XYZ Credit Union, we observe that customers that have listed their sex under the label ‘V’ are a total of 496,880, whereas the customers who have listed their sex under the label ‘H’ are a total of 419,111. This means that based on the data in the file ‘Test.csv’, the records show a ratio of customers with label ‘V’ to customers with a label ‘H’ as 1.19:1. While the ratio might not seem that far off, we need to take into account that the Credit Union deals with customers in large quantities, which means that even this relatively small difference reflects in up to several thousands of client records. A proper visualization of this data can be seen as follows as well:



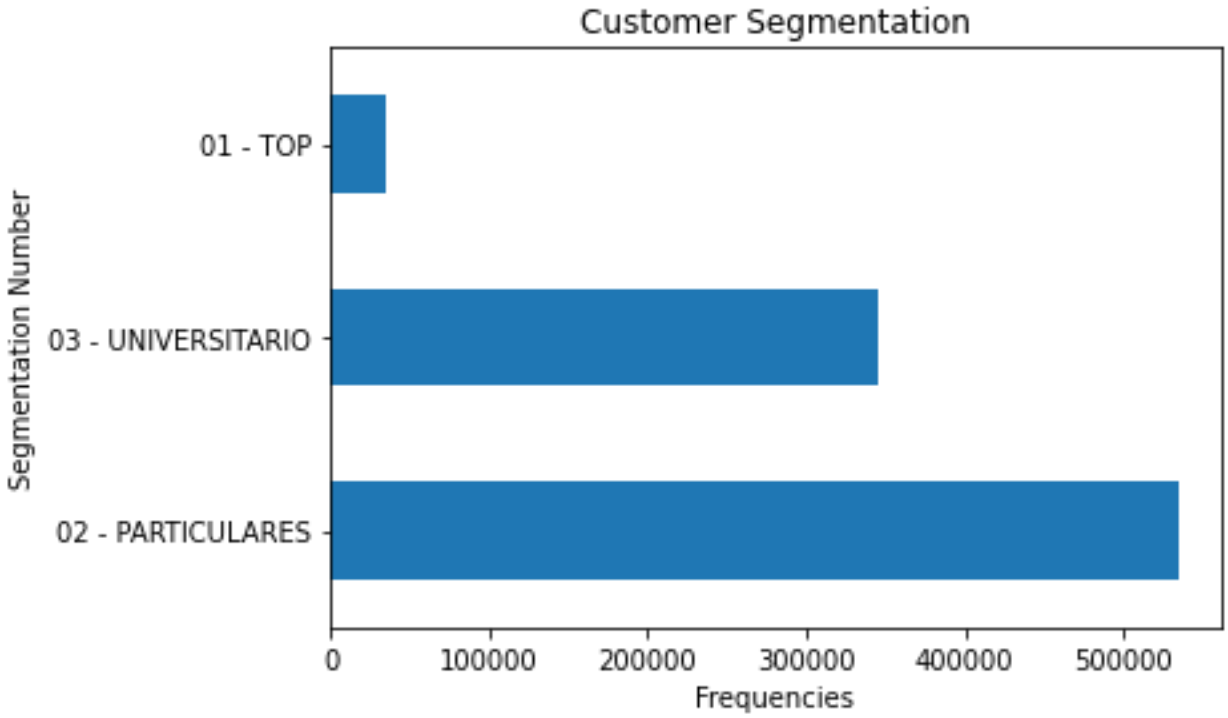
Additional to this analysis, another was performed to visualize and observe the age pattern of the current customers of the company. The reason for this particular analysis was to better understand the prime age group of the company, which would in turn help them with their future marketing strategies to sell banking products to their customers. The following graph was generated from the cleaned data in 'Test.csv':



Based on the generated bar graph, we can observe that in the age group of 18 – 116, the highest majority is ranges from the earliest legal age of 18 to the age of 27, after which we observe a steep decline in the

number of customers that are aged above 27. After this sharp decline, the rate of decline gradually slows down as the age continues to progress, up until the point that the entries of the higher ages are no longer visible on the graph. This shows that in the current listing of clients of XYZ Credit Union, the greatest majority is the age group of young professionals, which will help the company better understand the customer segment that they need to focus more on for future sales of their banking products.

Furthermore, another form of EDA was performed based on the data that was collected from 'Test.csv'. This section of data highlighted how the company had segmented its customer into different segments. These segments are to help XYZ Credit Union to better offer targeted services and products to their customers. Based on the information collected from 'Test.csv', the following graphic was generated in form of analysis:



From this infographic, it can be seen that the largest majority is the general collective of users who are not entitled to any specific scheme or unique benefits but are only seen as ordinary account holders. The second largest segment of customers consists of university students, who make up of the largest number of account holders for their age group, and as such represent a significant section. These students are usually entitled to specific programs and benefits when it comes to banking products, so properly identifying this section and targeting them with specific marketing strategies will benefit the company tremendously and allow them to properly sell a unique set of their banking products to this market sub-segment. The last remaining segment is the top clients of XYZ Credit Union. This segment usually consists of successful business leaders and investors, who also have a direct interest in the progression of the Credit Union itself. While these customers are barely even close to the other customer sub-segments, they are seen as the customer sub-segment that can have the largest impact on the bank and its progress plan. As such, these select numbered individuals can be ideal to sell the higher-end tier of banking products to, allowing the Credit Union to potentially generate higher yield in revenue from these customers, in comparison to customers in the other sub-sections.

Final Recommendations

The model proposed in this document is useful in not only cleaning the data to remove irrelevant data entries that might fabricate the actual data results, but rather the graphics generated at the end highlight the information listed in the data frame as well. As such, proposed code and model not only cleans the data, but also works actively in visualizing it as well.