BA

Experiment 8

Name:Ayush Jain

SAP ID: 60004200132

Div: B

Batch: B3

Aim: To perform Text Analysis in SAS

Theory:

Text analysis is the process of using computer systems to read and understand human-written text for business insights. Text analysis software can independently classify, sort, and extract information from text to identify patterns, relationships, sentiments, and other actionable knowledge.

SAS is a reliable tool for analysing text data source.

Analyzing a text data source in SAS (Statistical Analysis System) involves several steps. Here's a brief overview of the process:

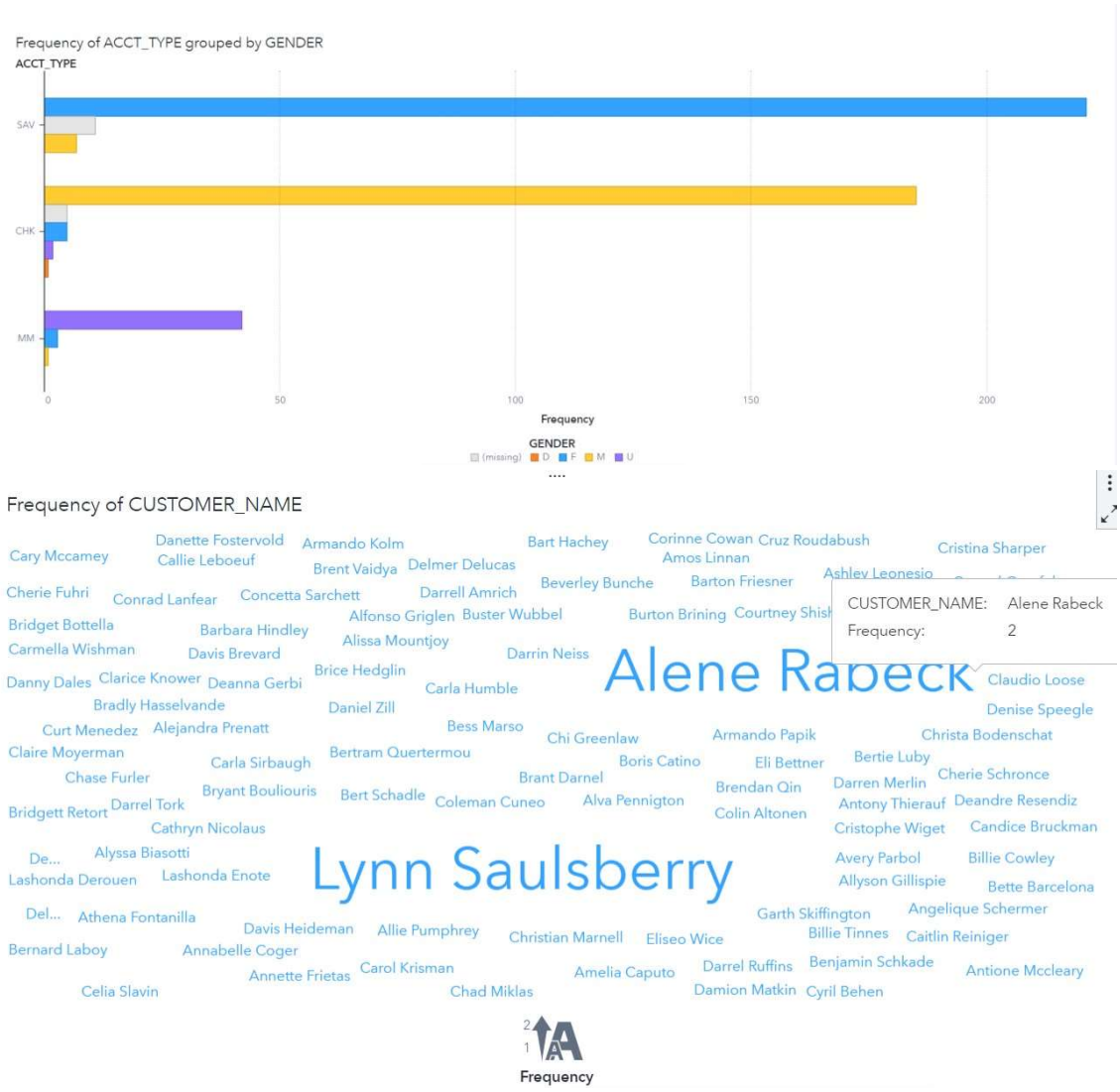
Import the text data: The first step is to import the text data into SAS. You can use the nfile statement to read the text data from a file, or use the import procedure to import data from a database.

Data cleaning: Text data often contains noise or unwanted characters, which need to be removed. You can use SAS functions like compress and substr to remove unwanted characters.

Text preprocessing: Text data needs to be preprocessed before it can be analyzed. This involves converting text to lowercase, removing stopwords (common words like "the" and "and"), stemming (reducing words to their root form), and tokenizing (breaking text into individual words or phrases). Text analysis: Once the text data has been preprocessed, you can perform various analyses. For example, you can count the frequency of individual words, calculate the sentiment of the text (positive, negative or neutral), or perform topic modeling to identify the main themes in the text.

Visualization: Finally, you can use SAS to create visualizations of your text data analysis. SAS provides various tools for creating charts and graphs that help to summarize and communicate your findings.

Output:



Conclusion:

Thus, SAS provides a comprehensive set of tools for analyzing text data, including data cleaning, preprocessing, analysis and visualization.