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:

Role by Frequency

Frequency chart (stacked bar) by Role: A, ADV, CONJ, DET, INTJ, N, nlpNounGroup, NUM, PN, PPOS, PRO, V

```
/*****************************************************************
* SAS Visual Text Analytics
* Sentiment Score Code
*
* Modify the following macro variables to match your needs.
*****************************************************************/

/* specifies CAS library information for the CAS table that you would like to score. You must modify the value to provide the name of the library that contains the table to be scored. */
%let input_caslib_name = "{input_caslib_name}";

/* specifies the CAS table you would like to score. You must modify the value to provide the name of the input table, such as "MyTable". Do not include an extension. */
%let input_table_name = "{input_cas_table_name}";

/* specifies the column in the CAS table that contains a unique document identifier. You must modify the value to provide the name of the document identifier column in the table. */
%let key_column = "{doc_id_column_name}";

/* specifies the column in the CAS table that contains the text data to score. You must modify the value to provide the name of the text column in the table. */
%let document_column = "{text_column_name}";

/* specifies the CAS library to write the score output tables. You must modify the value to provide the name of the library that will contain the output tables that the score code produces. */
%let output_caslib_name = "{output_caslib_name}";

/* specifies the sentiment output CAS table to produce */
%let output_sentiment_table_name = "out_sentiment";

/* specifies the matches output CAS table to produce */
%let output_matches_table_name = "out_sent_matches";

/* specifies the features output CAS table to produce */
%let output_features_table_name = "out_sent_features";

/* specifies the language of the associated SAS Visual Text Analytics project. This should be set automatically to the language you selected when you created your project. */
%let language = "ENGLISH";

/* specifies the hostname for the CAS server. This should be set automatically to the host for the associated SAS Visual Text Analytics project. */
%let cas_server_hostname = "sas-cas-server-default-client";

/* specifies the port for the CAS server. This should be set automatically to the host for the associated SAS Visual Text Analytics project. */
%let cas_server_port = 5570;

/* creates a session */
cas mysas1 host=&cas_server_hostname port=&cas_server_port;
libname mysas1 cas sessref=mysas1 datalimit=all;

/* calls the scoring action */
proc cas;
    session mysas1;
    loadactionset "sentimentAnalysis";

    action applySent;
        param
            table={caslib=&input_caslib_name, name=&input_table_name}
            docId=&key_column
            text=&document_column
            language=&language
            casOut={caslib=&output_caslib_name, name=&output_sentiment_table_name, replace=TRUE}
            matchOut={caslib=&output_caslib_name, name=&output_matches_table_name, replace=TRUE}
            featureOut={caslib=&output_caslib_name, name=&output_features_table_name, replace=TRUE}
        ;
    run;
quit;
```

Descriptive Statistics

Measure	Terms in a Sentence	Terms in a Document
Minimum	1	3
Maximum	560	816
Mean	21.8878	236.7472

Topics (11)

	Topic	Created by	Documents ↓
☐	+card, +credit card, +credit, +charge, +purchase	System	400
☐	+debt, +call, +say, n't, +number	System	392
☐	xx, xxxx, +letter, +escrow, +receive	System	379
☐	+modification, +loan, +loan modification, +mortgage, +home	System	348
☐	+debt, +identity, +letter, +theft, +information	System	328
☐	+check, +account, +check, +deposit, +money	System	317
☐	+report, +credit, +credit report, +report, +debt	System	307
☐	+late, +late fee, +payment, +fee, +day	System	289
☐	+interest, +rate, +interest rate, +balance, +loan	System	273
☐	+fee, +overdraft, +charge, +overdraft fee, +account	System	265

Terms

All (4588)   Matched   🔍 Filter

	Term	Role	Documents	Frequency ↓
☐	xxxx	PN	2148	26336
☐	not	ADV	1784	5336
☐	bank	PN	1381	3779
☐	▸ account	N	1102	3516
☐	▸ bank	N	1416	3505
☐	xx	PN	495	3002
☐	▸ payment	N	885	2820
☐	▸ credit	N	949	2549
☐	▸ loan	N	740	2336
☐	▸ tell	V	940	1984

Documents

All (2413)   Matched   Search

Consumer_complaint_narrative	Sentiment
I live paycheck to paycheck. I check my bank account balance daily if not XXXX a day. The bank I use, Bank B offers online and text banking so it makes it very convenient. The problem is they do not match up. Some of the dates that transaction posts change. For example on XXXX night, the XXXX, I checked my balance through text banking. I had ($380.00) in my account. The next morning, XXXX the XXXX, when I got to work at XXXX I went online and checked my balance and transactions. It still said I had a balance of ($380.00). At this point I realized I still had XXXX checks outstanding and I needed to transfer money from my savings. I transferred ($1200.00)….	🙂
I have been attempting to rectify the situation with Bank A by calling various executives including the Assistant General Counsel to no avail. After further review of my account I located ($490.00) worth of bank fees on XXXX account alone within the last month. All of these were based off of Bank A allowing electronic payments to post and then charging my account a bank fee. The most recent fee was for a ($7.00) charge that they returned to a vendor that last month they choose to pay as a courtesy '' when the transaction was for well over ($100.00) and the account was overdrawn for a lot more than it was when they returned the ($7.00) transaction. When I asked…	🙁
I have a checking account at Bank D. When I wrote a check of ($1000.00) to XXXX XXXX for paying my rent (for my wife, XXXX XXXX 's rent in XXXX) on XX/XX/XXXX, somebody named XXXX XXXX, instead of XXXX XXXX, cashed the check to her own Bank D account by wiping out the recipient 's name (XXXX XXXX) and put her name on the check. Bank D did not look carefully for the payee authenticity and paid the money to that person (XXXX XXXX). I filed a clam to Bank D for ($1000.00) on XX/XX/XXXX. Bank D Representative in that office told me that the check was deposited in a Bank D account, i would be easy for them to find the theft since they…	🙁

Number of Documents Per Topic
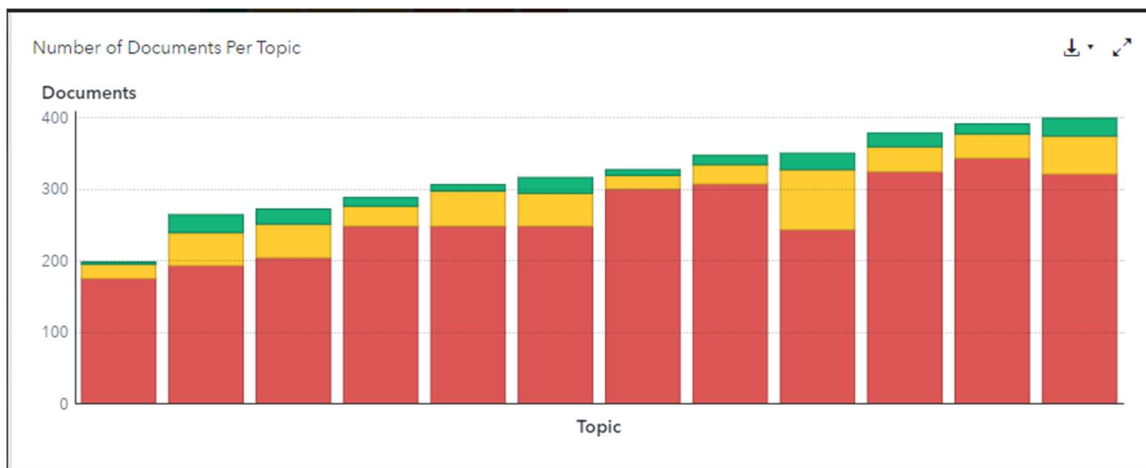
Conclusion:

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