**EM-624 Informatics for Engineering Management**

**EX-09 Analyzing News**

**Ashutosh Laxminarayan Gor**

**Interpretation:**

* Using the pandas, provided data was imported into a dataframe and all the unwanted blank spaces and na inputs were removed to avoid the errors during data processing.
* I created a column to contain the specific data captured from the dataframe of each news article.
* Using the string function covid related words were defined and used to capture covid related text from the data frame.
* Then I again created a new column to determine the each column contains covid-19 related article or not.
* Using the stopwords file provided along with the data file, I removed all the unwanted words that were not required for the data processing along with the words with length of less than 3 characters.
* Once again I created a new column to contain the cleaned text from previous process.
* Then I created a new list to contain all the words of all the news articles from the data file.
* Post that I created a new dataframe to hold only COVID-19 related news article for post processing.
* I then converted the new covid related dataframe into a list to hold all the articles.
* Similarly I created another data frame to contain all the non COVID-19 related nwews articles.
* Then creating a dictionary to determing the 20 most common used words in the data file along the frequency of there usage in the data file.
* Then printing the most common 20 words along with the frequency of its usage .
* Post which I calculated the percent of articles related to COVID-19 and printing all the COVID\_19 related news articles.
* Post that I used the vader sentiment library to calculated the sentiment behind all the non COVID-19 related news articles along with all the news articles provided in the data file.
* After the sentiment analysis, I used the wordcloud library to print the word clouds for words collected from all the news articles and another wordcloud with words collected from all the non COVID-19 related news articles from the provided data file.

**Expected Output:**

These are the 20 most common words used in all the news articles:

{'news': 440, 'covid': 371, 'said': 368, 'vaccine': 290, 'people': 286, 'google': 283, 'health': 275, 'cases': 273, 'coverage': 272, 'says': 167, 'government': 159, 'vaccination': 140, 'time': 136, 'just': 131, 'positive': 129, 'sheeran': 124, 'workers': 116, 'year': 115, 'need': 114, 'live': 112}

Percent of articles relate to COVID-19: 54.23% (417/769)

These are the COVID-19 related news articles:

0 Covid news - live: PM insists booster jabs wil...

1 Kolkata Municipal Corporation all set to reope...

2 Leaked memo hints 'immediate' roll-out of Covi...

4 Facing alien Isolation

5 Coronavirus live: UK travel rules relaxed as a...

...

760 I‚Äôm waiting to see my specialist, to find ou...

761 Assam Revised Covid-19 SOP, Imposes Night Curf...

762 No quarantine for fully vaccinated players at ...

763 Are flu vaccine side-effects worse this year? ...

764 Huge increase in imports checks, consumer conf...

Name: title, Length: 417, dtype: object

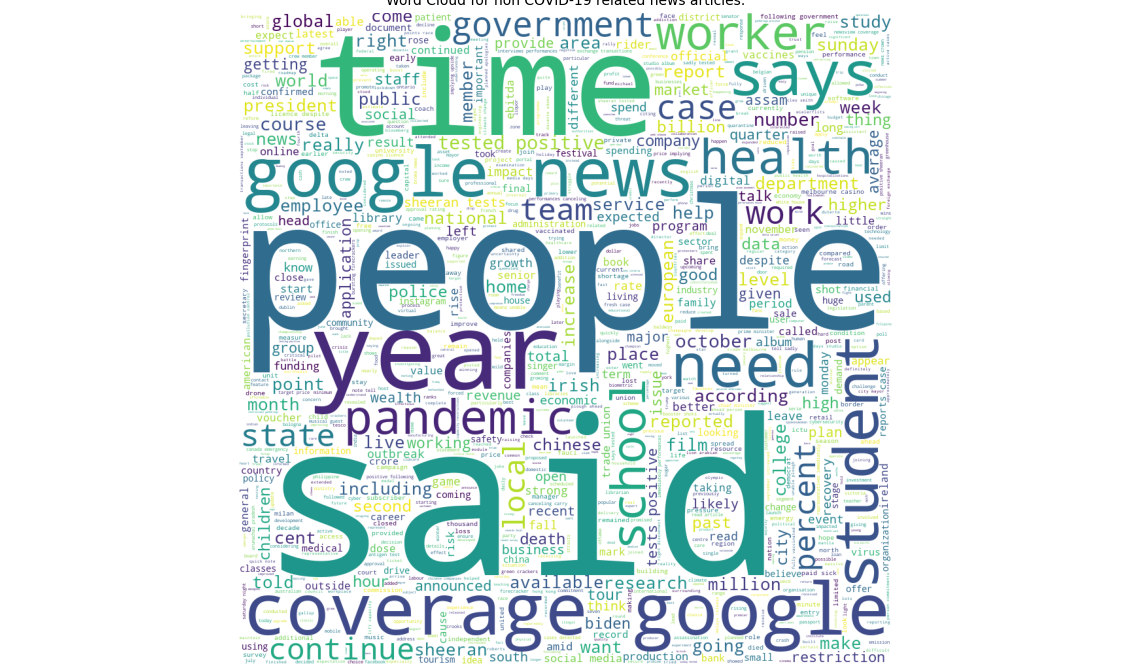
All the non COVID-19 related news articles are rated as: 17.7% Positive 8.1% Negative & 74.2% Neutral

All the news articles are rated as: 16.4% Positive 8.3% Negative & 75.3% Neutral

Text

Description automatically generatedWord Cloud for all the news articles:

Word Cloud for non COVID-19 related news articles:



Process finished with exit code 0