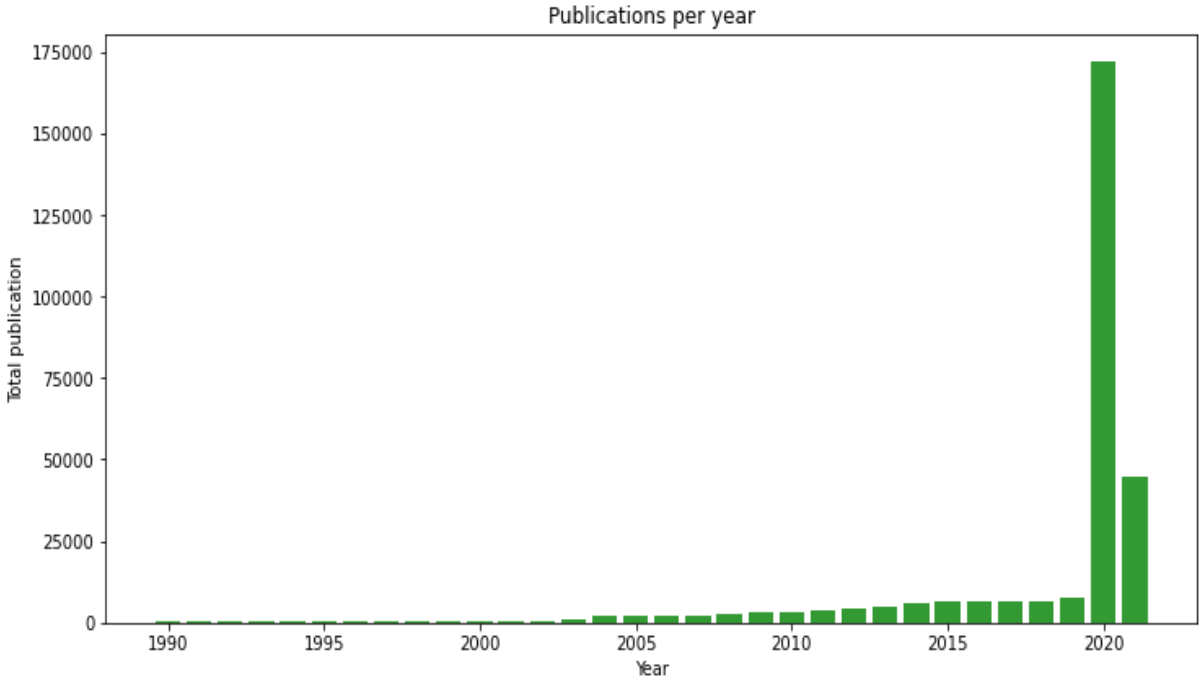
COVID-19 Analysis for the Transmission Modes and Symptoms

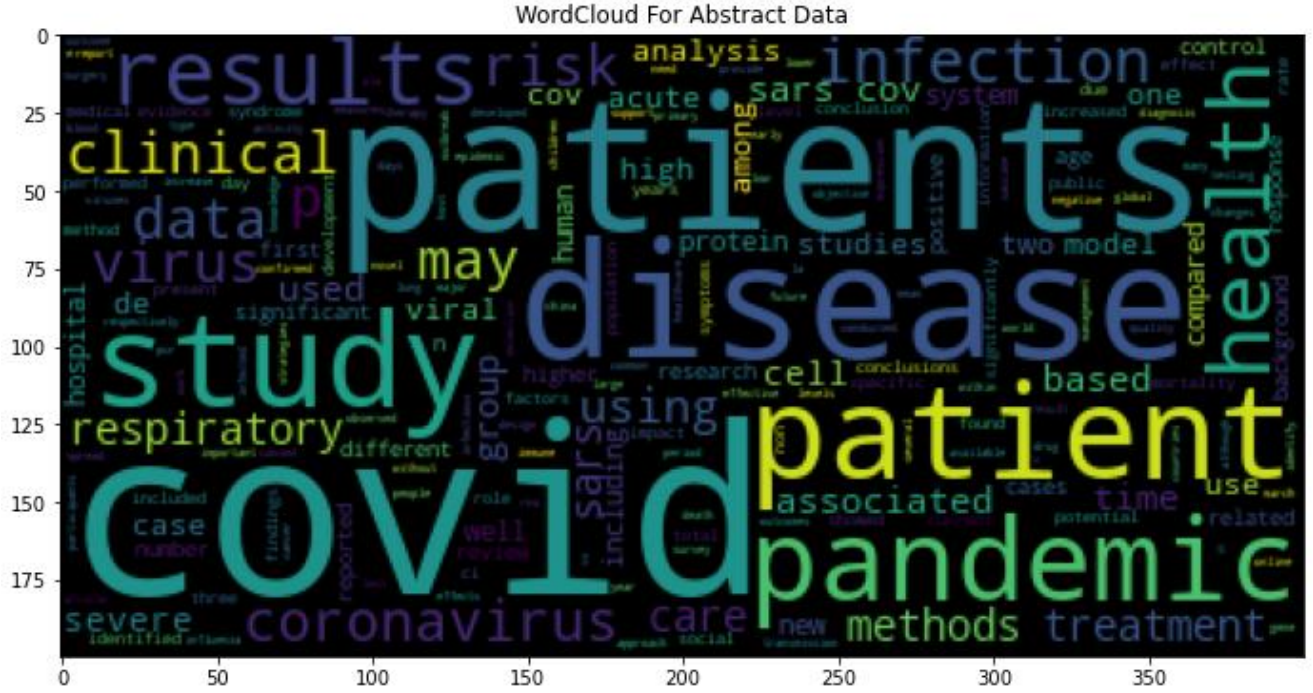
Submitted by,

Ruchit Shah

Student Number: 1005677830

Exploratory Data Analysis

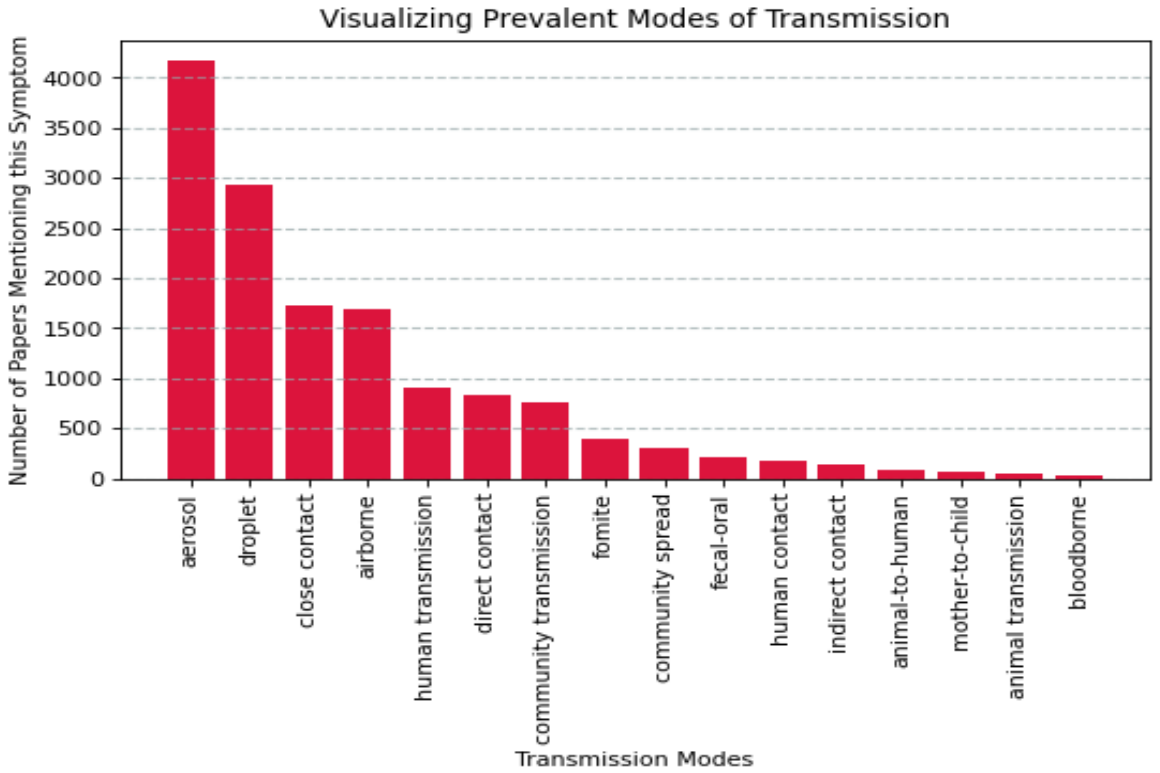




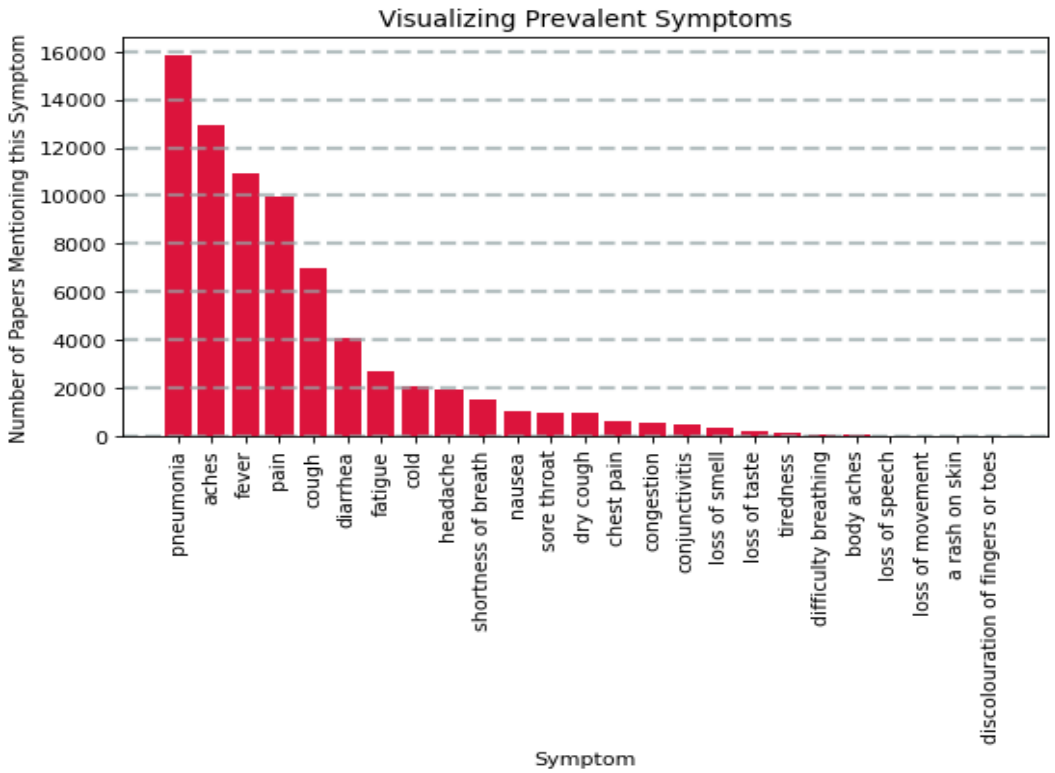
It is seen from the word cloud that most of the words described in the abstract is like patients, covid, disease, study, pandemic, health, clinical, results, methods, treatments, infection, risk, etc. These explains that the database is mainly based on the covid which describes about the study of infection and disease. It details the methods and treatment for the coronavirus, the infection rate and analysis for the COVID.

The publication year figure explains that most of the research papers are published in the year 2020 and 2021. This implies that the database is mainly related to the COVID-19. Though, it includes the study of the previous coronavirus like SARS, but it mainly explains the symptoms, transmission modes, risk, incubation rate, etc for the latest coronavirus, i.e., COVID-19.

Transmission Modes



Symptoms



The words for the transmission modes and the symptoms are researched from the online data. Those words are searched in the abstract column of the data frame and summed up to get the total count of words used. These words are analysed later using the unsupervised machine learning methods.

Transmission Modes: The most common mode of transmission inferred from the plot is the aerosol and droplet. These two are the modes of transmission through air. While the transmission like animal to human, mother to child, bloodborne are least common modes. Therefore, the data verifies that the COVID spread mainly through air in ways of droplets, aerosols, and due to the close contact also.

Symptoms: The most common symptom is the pneumonia, aches, fever, and cough. These symptoms are related to the respiratory system which is the main cause of the COVID-19. These common symptoms mainly affect the lungs and leads to difficulty in breathing and adverse the situation.

ML and feature engineering methods

In this analysis, the unsupervised machine learning algorithm is implemented. The two **clustering** methods, Hierarchical clustering and K - Means clustering is implemented for the analysis of the transmission mode and symptoms.

Unsupervised Learning is a machine learning technique in which the users do not need to supervise the model. The model is developed to discover patterns and relations between the features. In supervised machine learning method, there is a target variable which is predicted by implementing the model. While, in unsupervised machine learning method, there is no target variable.

Hierarchical clustering: It is a type of algorithm which builds a hierarchy of clusters. It begins with all the data which is assigned to a cluster of their own. Here, two close cluster are going to be in the same cluster. This algorithm ends when there is only one cluster left.

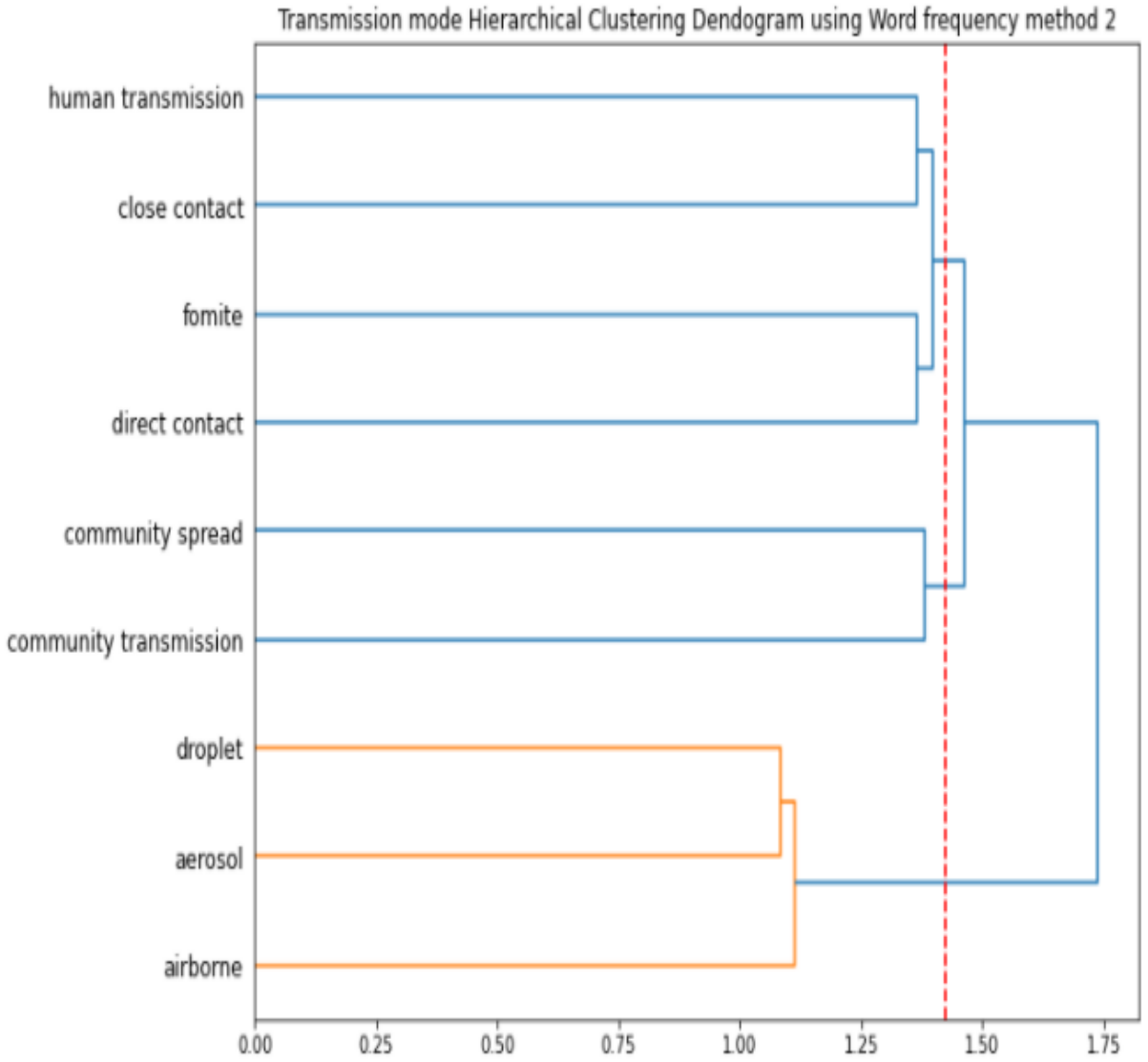
K-means clustering: It is an iterative clustering algorithm which helps you to find the highest value for every iteration. In this method, the number of clusters is initialized into k groups.

Feature Engineering Methods:

Method 1: Word Count – In this method, the total number of words in the entire database is summed up and used for the clustering method. The distance matrix is calculated using the scipy library.

Method 2: Word frequency normalization – In this method, the ratio of words occurring in the abstract to the total number of words in that particular abstract is stored in a matrix. Then, this matrix is used to obtain the linkage matrix and implemented on clustering algorithm.

Transmission Mode Analysis



- The Hierarchical clustering with the feature engineering method 2 gives the superior results for the transmission mode analysis. The detailed analysis of the cluster is mentioned below.

- **Cluster 1**: Human transmission, close contact, fomite, and direct contact. This cluster has all the modes which are spread through the human and nearby interactions. It can be inferred that these types of modes should be prevented altogether. Therefore, the **social distancing** is the key to prevent the transmission.

- **Cluster 2**: Community spread, and community transmission. This cluster relates to the modes related to the spread of community. The community gathering and get-to-gather is one of the primary cause of COVID transmission. Therefore, the **lockdown** and **ban on gathering** is one of the main implementation to prevent the spread of COVID.

- **Cluster 3**: Droplets, Aerosol, and Airborne. This cluster combines all the modes through air. The sneezing or coughing spread the virus through droplets and aerosol which proves to be the mode of COVID transmission. Therefore, the use of **masks** and **sanitization** helps to prevent this cluster transmission.

COVID Symptoms Analysis

- Hierarchical method gives the 5 clusters which are shown in the dendrogram. For the K – means clustering also, the 5-cluster input is provided. The clusters for the K-means clustering method gives more sensible insights for the COVID symptoms analysis. The clusters are explained as below.

- **Cluster 1**: Pneumonia, fever, cough, fatigue, shortness of breath, sore throat, dry cough. In this case, it makes more sense than the hierarchical clustering result. The symptoms mentioned in this clustered are much more related to the cough and breathing condition which have infection in the lungs like pneumonia. Therefore, the **ventilation** and other medication related to breathing problem may be more effective.

- **Cluster 2**: Aches, headache. These two symptoms generally occurs together.

- **Cluster 3**: Pain, chest pain. These two symptoms also generally occurs together.

- **Cluster 4**: Diarrhea, nausea. These two symptoms are related to the issues of stomach and makes sense that this affect the digestion system. Therefore, medications related to the stomach issues may be provided to relief COVID symptoms.

- **Cluster 5**: Cold, Congestion, conjunctivitis, loss of smell. These are common issues related to cold situation and sinus issues. It affects the smell as due to cold, the sense of smelling get affected drastically.