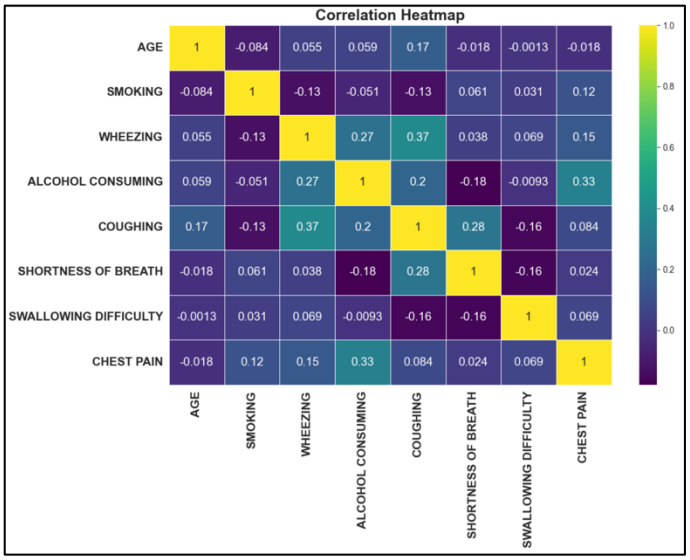
APPLIED DATA SCIENCE - 1

NAME : Dhanesh Kanakaraj
STUDENT ID : 23056970
GITHUB : https://github.com/kdhanesh619/Applied-Data-Science-Assignment-2.git
ASSIGNMENT : Statistics and Trends

PREDICTION OF LUNG CANCER

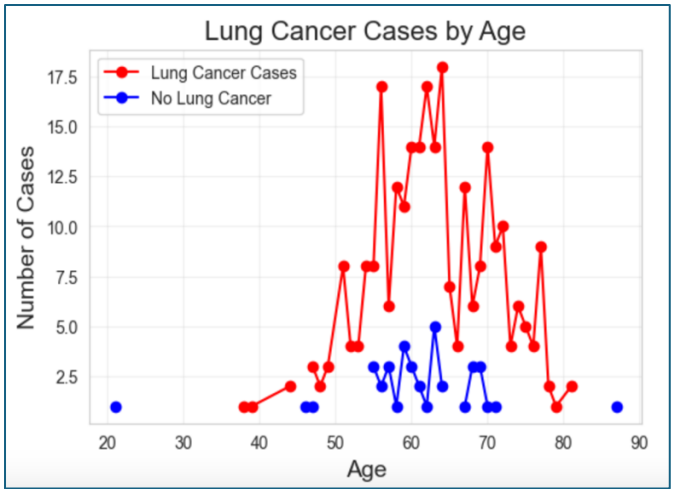
ABSTRACT

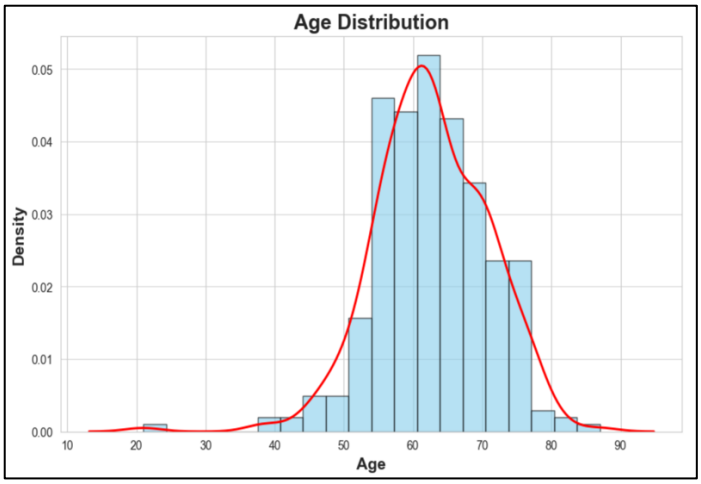
This data provides the report for the lung cancer patients according to the age category which was indicated based on some of the symptoms which can lead to cause lung cancer. The statistical visualizations help to get some insights with some patterns and their association with the Lung cancer. As a result, this data also helps to predict the patients whether they are diagnosed with the lung cancer to underline the importance of early detection for better health outcome.



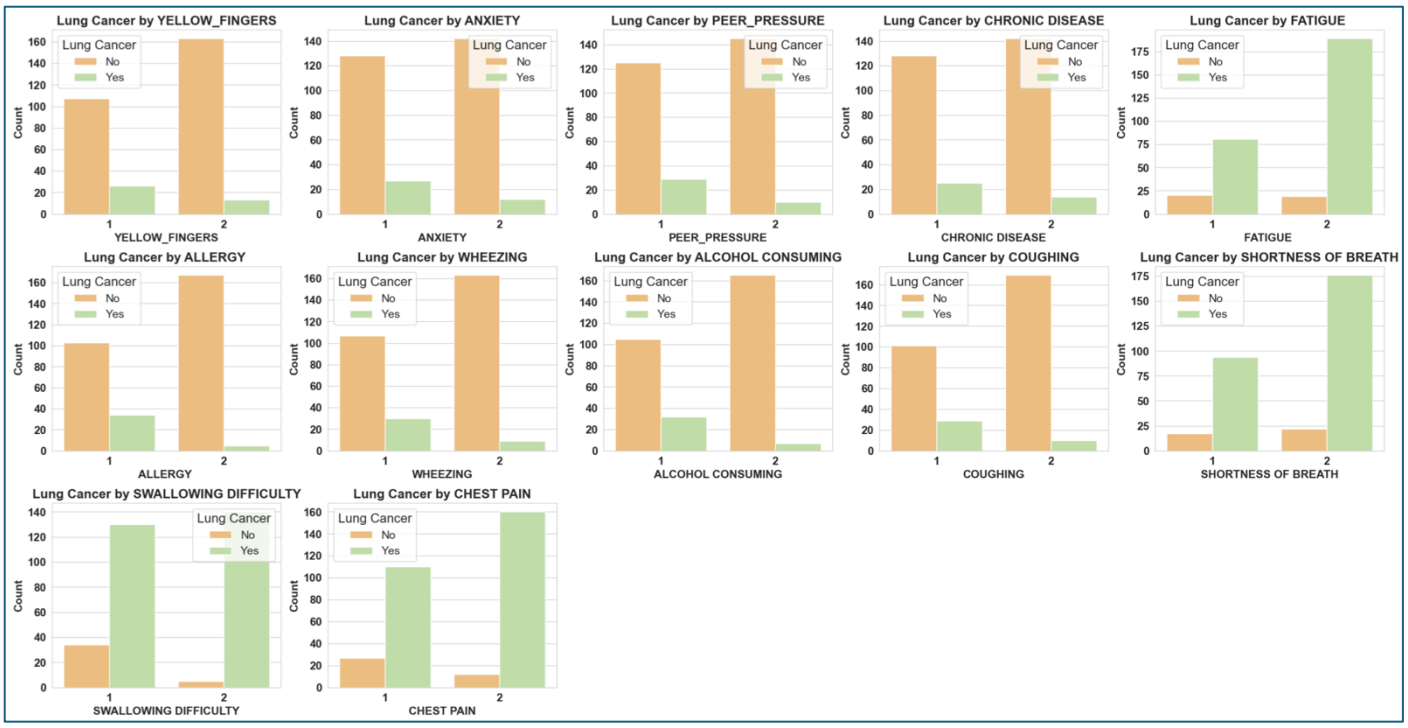
The heatmap provided above shows the correlation for major symptoms of lung cancer. Wheezing and coughing are linked 0.37, which says that they mostly co-occur, while the alcohol consumption and chest pain also shows some correlation 0.33. Coughing and shortness of breath values are correlated moderately with 0.28, showing some common respiratory distressing patterns. Smoking is weaker but it links with chest pain 0.12. Overall, age and swallowing difficulty shows minimal correlations with other symptoms, indicating less relevance in this specific content.

The line plot shows the relationship between age and the number of lung cancer cases. It displays how lung cancer diagnoses which are marked in red varies across different age groups compared to those without the condition which are marked in blue. The markers and a clear grid improves readability, making it easier to observe age-related trends in lung cancer possibilities. This visualization provides insight like whether lung cancer prevalence increases with age and helps to identify age groups that may require more focused healthcare attention.





The histogram shows the age distribution within the dataset, showing a concentration of individuals between 50 and 70 years old. The high frequency occurs around the 60-65 age range, indicating that this statistic is most represented. There are fewer individuals in both younger who are below 40 and older who are above 80 age groups, suggesting a skewness towards middle-aged and older peoples. This distribution may indicate that lung cancer data in this study primarily focuses on those who are aged 50-70, helping to detect lung cancer early and improve health outcomes.



The bar charts are used to compare the presence of lung cancer with various symptoms. For symptoms like Yellow Fingers, Anxiety, Peer Pressure, and Chronic Disease, a higher proportion of individuals without lung cancer is observed, which are indicated by the taller blue bars. At the same time, symptoms such as Fatigue, Wheezing, Shortness of Breath, Swallowing Difficulty, and Chest Pain shows a more balanced or even higher count among those with lung cancer, compared with the taller orange bars. This suggests that certain symptoms are more closely linked with lung cancer, highlighting the key areas for focusing on medical assessment in lung cancer risk evaluation.

CONCLUSION

The analysis of lung cancer data reveals significant associations between various symptoms and the presence of lung cancer. The statistical visualizations, including heat maps, line plots, histograms, and bar plots, provide a better understanding of how the factors like fatigue, wheezing, shortness of breath, and chest pain are more formative in individuals diagnosed with lung cancer. Additionally, age distribution highlights that middle-aged and older adults are more frequently affected, showing the need for age-based screening and preventive measures. As a result, these insights implements the critical role of early detection and symptom awareness in managing and potentially reducing lung cancer risks. This report shows a data-driven approach to better inform healthcare strategies and to improve the outcomes for those who are at risk.