

BUSINESS INTELLIGENCE FINAL PROJECT

TOPIC: LUNG CANCER – Factors to Facilitation



By:Shreeya Satav

INTRODUCTION

Lung cancer is a worldwide health problem that kills a large number of people each year. To improve patient results and mortality rates, it's important to find and treat cancer early on. The

lung cancer collection has a lot of medical information about patients, such as their traits, symptoms, and diagnoses. This information can be used to answer many questions about the disease. By looking at this data, experts and medical professionals can find out what makes people more likely to get lung cancer, which medicines work best for which patients, and how living choices affect how well patients do.

The lung cancer information can also be used to make models that predict how likely it is that a patient will get lung cancer. This lets screening and prevention efforts be more focused. This knowledge can also help make screening and diagnosing methods better, which can make it easier to find lung cancer early, when it is easiest to treat.

Overall, the lung cancer collection gives researchers and health care workers a great chance to learn more about this disease, find new ways to avoid and treat it, and, in the long run, make lung cancer less common around the world. By using these facts, we can improve how patients do and cut down on the number of people who die each year from lung cancer.

SUITABILITY OF DATA

Researchers and medical workers who want to learn more about lung cancer will find the lung cancer collection very useful. The dataset has a lot of different types of medical information, such as biographical data, symptoms, and diagnoses. This information can be used to answer a wide range of study questions about lung cancer.

The collection has both clinical and demographic data, which lets researchers look at how different risk factors, like age, gender, smoking status, and genetic history, affect the growth and spread of lung cancer. The information can also be used to study how living factors like food and exercise affect how well patients do and for which patients level patients certain generic treatment methods might or might not work.

SOURCE OF DATASET

The team has integrated two datasets for the visualization purposes,

1. <https://www.kaggle.com/datasets/mysarahmadbhat/lung-cancer?select=survey+lung+cancer.csv>
2. <https://www.kaggle.com/datasets/thedevastator/cancer-patients-and-air-pollution-a-new-link>

KEY AUDIENCE OF DATA VISUALIZATION

The team can make a simple but useful picture that shows the most important results from the dataset to show policy makers what personal and environmental factors may facilitate lung cancer. One idea would be to make a live screen that gives the person different ways to look at the information.

On the screen, there could be several pictures, such as a scatter plot that shows how causes and lung cancer are linked, and the likelihood for certain facilitating factors.

Overall, the goal of the image should be to help policy makers understand in a clear and easy way the most important things about lung cancer and how it affects public health. The team can show how important lung cancer research is and how much more help and money is needed in this area by showing this information in a way that is both visually appealing and interesting.

STORY TELLING BY DATA VISUALIZATION

The main questions:

1. What are the most common things that put people at risk for lung cancer?

I). **Descriptive:** What is the distribution of risk factors among the patient population?

II). **Predictive:** Can we predict which patients are at the highest risk for developing lung cancer based on their risk factors?

III). **Prescriptive:** How can we use this information to develop better screening and prevention strategies to reduce the incidence of lung cancer?

2. What factors influence patient outcomes?

I) **Descriptive:** What is the distribution of patient outcomes among the patient population, and what factors are most strongly associated with positive outcomes?

II) **Predictive:** Can we predict which patients are most likely to experience positive outcomes based on their medical history, treatment options, and other factors?

III) **Prescriptive:** How can we use this information to improve patient outcomes by developing better treatment plans, providing targeted support for patients, and addressing modifiable risk factors such as lifestyle factors or environmental exposures?

3. How to plan the most effective treatment?

I)**Descriptive:** What is the distribution of treatment options and outcomes among the patient population?

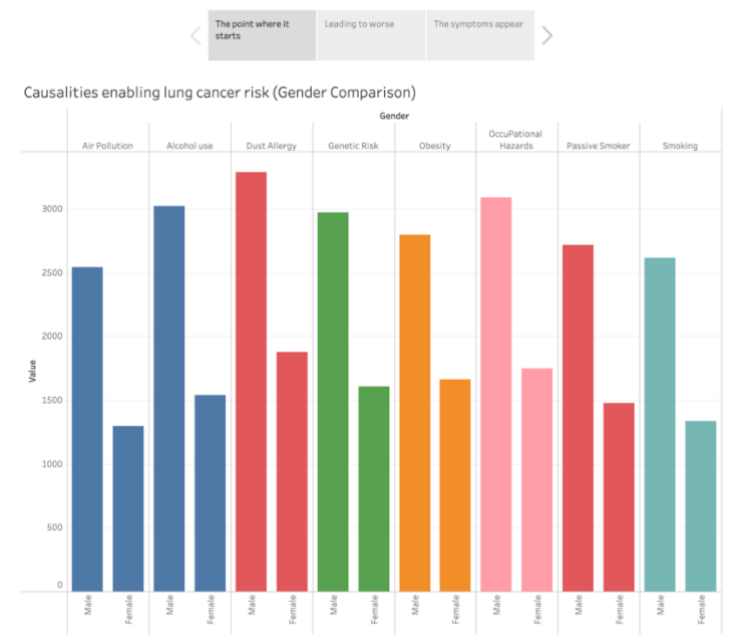
- **II)Predictive:** Can we predict which treatment options are likely to be most effective for a given patient based on their personal health standpoint?

III)Prescriptive: How can we use this information to develop more personalized treatment plans and improve patient outcomes?

Questions Story 1 : What are the most common things that put people at risk for lung cancer?

- **Descriptive:** What is the distribution of risk factors among the patient population?

Story 1



The visualization is a bar chart that shows the causalities enabling lung cancer risk (Gender Comparison) in terms of major influencing factors such as Air Pollution, Alcohol Use, Dust Allergy, Genetic Risk Obesity, Occupational Hazard, Smoking.

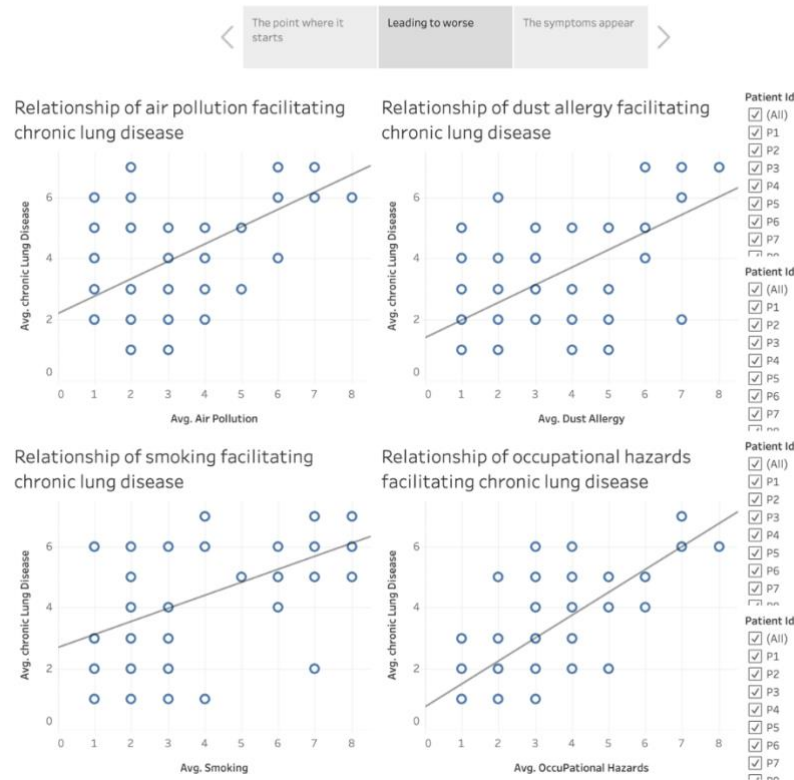
The y-axis represents the number of male and female affected by Lung Cancer and the x-axis represents the causes.

The x-axis is divided into different segments listing the causes that can lead to Lung Cancer. The bars are colored differently based on the causes i.e. each cause has a different colored bar.

The height of the bars represents the frequency of Lung Cancer in each category. The bar chart suggests that Males are at higher risk of developing Lung Cancer than Females and dust allergy, occupational hazard, genetic risks appear to be the primary reason for Lung Cancer in both genders.

- **Predictive:** Can we predict which patients are at the highest risk for developing lung cancer based on their risk factors?

Story 1



There are four scatter plots consisting of regression analysis of the correlating factors which shows relationships between the causes and lung disease.

All scatter plots indicate a positive upward relationship between them and lung disease with higher levels of causes associated with higher levels of lung disease.

Overall, these scatter plots highlight the different factors that may contribute to the development of lung disease; for example, air pollution and occupational hazards have significant impact on facilitating chronic lung disease. It also suggests that reducing exposure to these risk factors may be effective in terms of preventing or reducing the risk of lung disease. Policy makers who work with the environment, labor or health industry will benefit from such insights.

- **Prescriptive:** How can we use this information to develop better screening and prevention strategies to reduce the incidence of lung cancer?



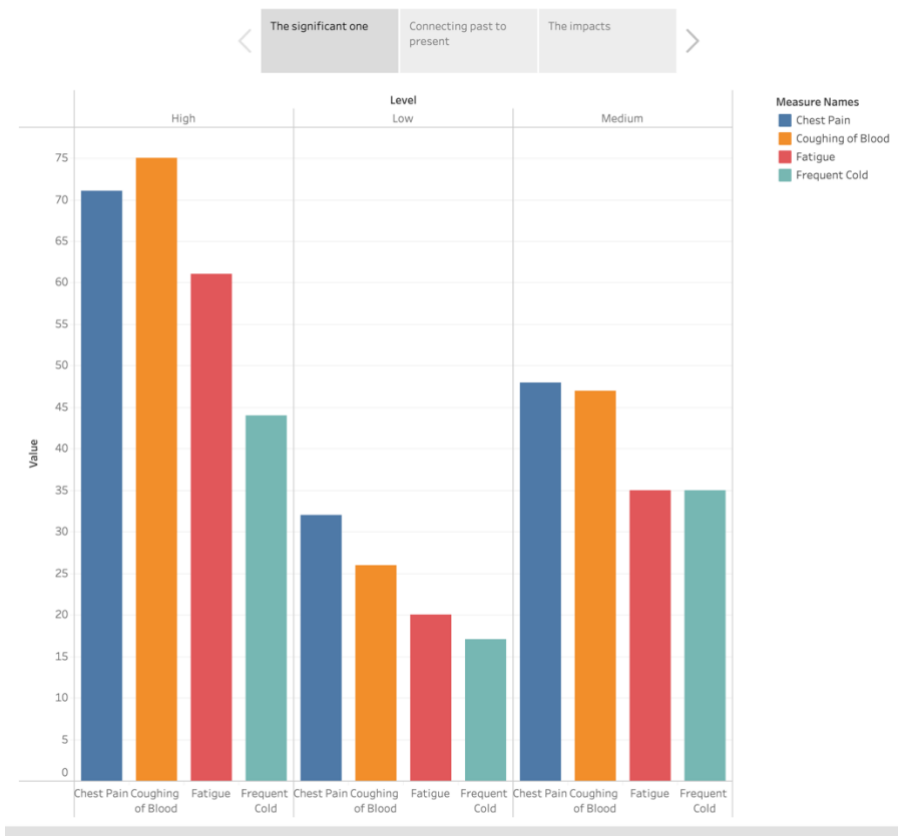
The bar charts represent risk intensity with different symptoms which are further divided into three segments. The three segments represent Coughing of blood, weight loss and fatigue respectively.

The chart provides individualized patient information, which patient is at what risk level and which of these intensity symptoms are highly affiliated with them. For example, a certain patient with a high level of risk can have all three of these symptoms. But for each individual the intensity of a symptom is different. For example, patient 213 is on the high-risk index. He has all three of the symptoms, but his loss of weight is more significant than other symptoms. Through this visual one can understand who needs more care in terms of what aspect.

Questions Story 2 : What factors influence patient outcomes?

- **Descriptive** : What is the distribution of patient outcomes among the patient population, and what factors are most strongly associated with negative outcomes?(patient level risk)

Story 2



This bar chart is the accumulation of all the patients datasets that we have it shows that the patient with high risk level of being diagnosed with lung cancer shows greater severity in coughing of blood and lesser intensity in terms of frequent cold.

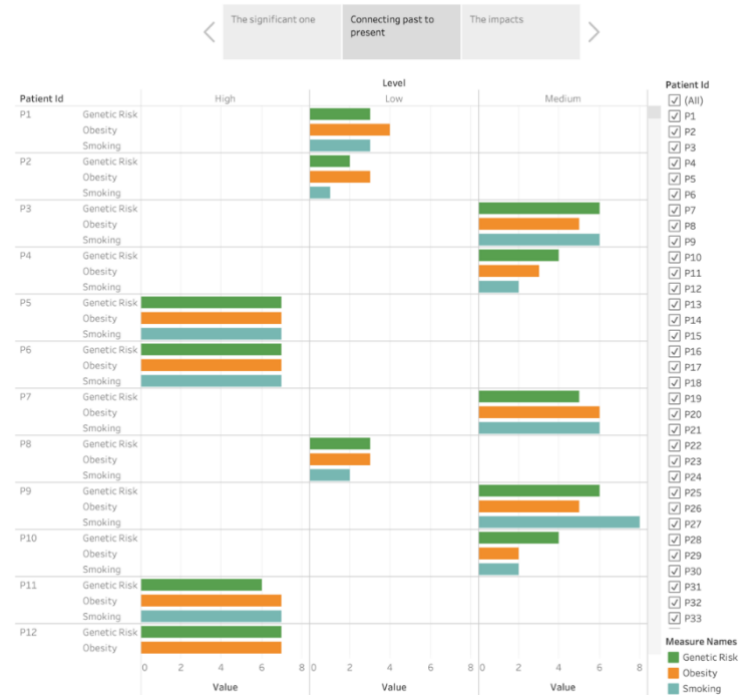
On the other hand, if you are at lower level of being diagnosed with lung cancer they have higher chest pain but coughing of blood has reduced significantly.

And those who are at medium risk level of being diagnosed with lung cancer shows almost similar severity of chest pain and coughing of blood.

Therefore, higher the level of being diagnosed with lung cancer , the intensity of the symptom also goes higher.

- **Predictive:** Can we predict which patients are most likely to experience negative outcome?

Story 2



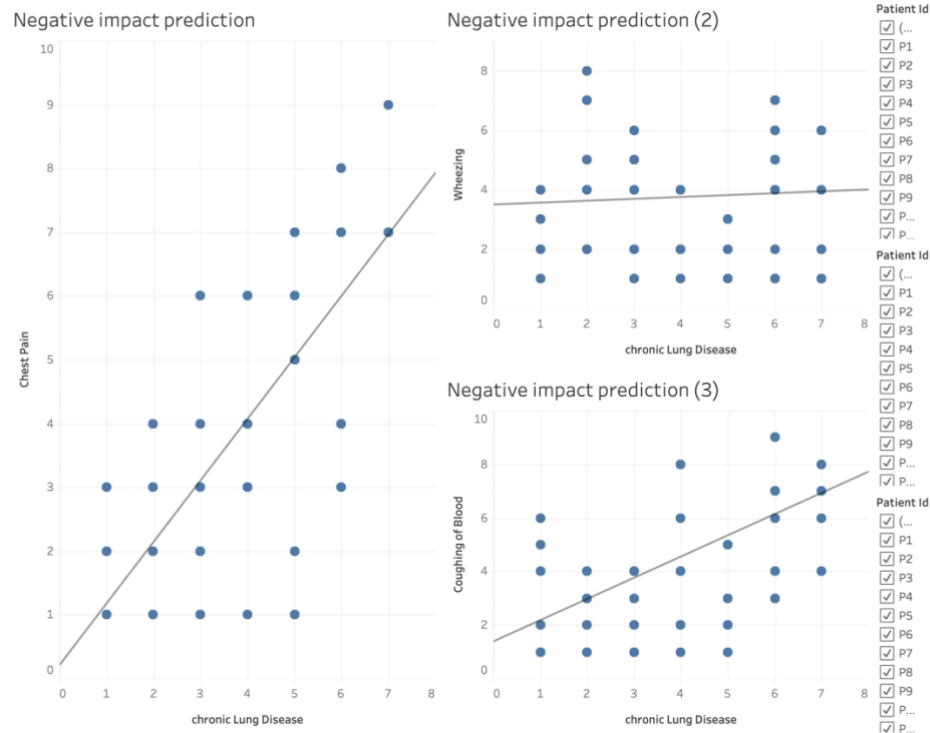
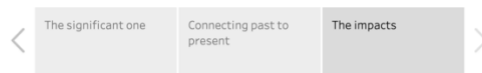
Genetic Risk is something that comes with the family. Obesity is something that comes from the lifestyle you have and smoking is a personal habit.

We already know what the intensity level of lung cancer the patients are in. For example, if a patient has medium level intensity of diagnosed with lung cancer, however if any one of the causes like genetic risk, obesity or smoking are high then the likelihood of the patient getting lung cancer is more. Therefore, these factors can accelerate the patient to go to the higher level of lung cancer.

These factors will help us analyze whether a person is more likely to get lung cancer.

- **Prescriptive:** How can we use this information to improve patient outcomes by developing better treatment plans, providing targeted support for patients, and addressing modifiable risk factors such as lifestyle factors or environmental exposures?

Story 2



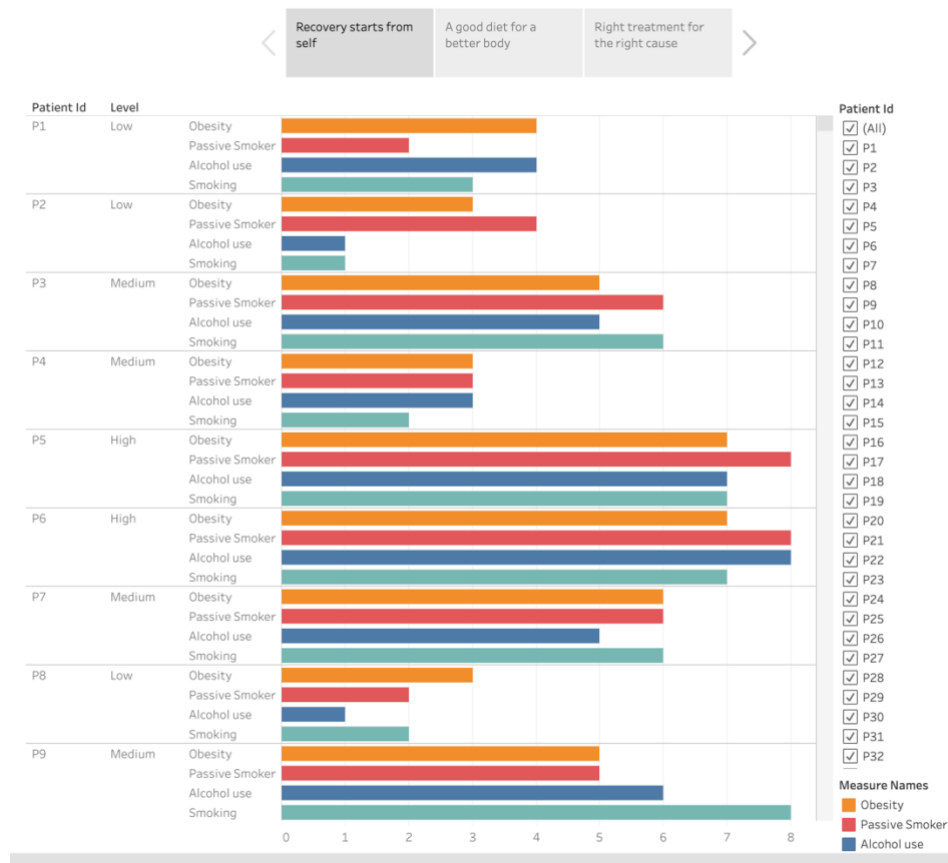
Here we are doing negative impact prediction in distribution model. As you can see in the distribution model there is an upward trend which shows that chronic lung disease increases chest pain increases.

Similarly, for negative impact prediction two as we can observe as the chronic lung disease is lower wheezing is also lower and for the negative impact prediction the chronic lung disease is higher because of which coughing of blood in the patient is higher.

Questions Story 3: How to plan the most effective treatment?

- **Descriptive:** What is the distribution of treatment options and outcomes among the patient population?

Story 3



The bar chart shows different aspects based on individual lifestyles and habits which are obesity, passive smoker, alcohol use and smoking that will lead to lung cancer

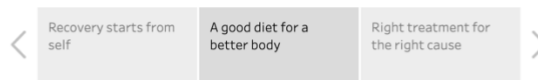
Based on the bar chart the doctors can use this data to personalize treatment for everyone. For example a passive smoker can be affected by lung cancer and in order to prevent it doctors can advice to change the surroundings.

Also, a person with obesity can be affected by lung cancer and to prevent it doctors can provide that person with a balanced diet.

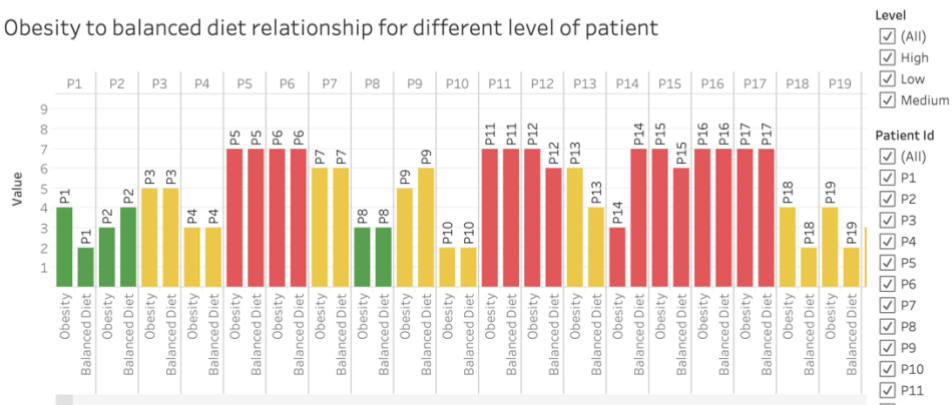
The chart gives a general idea of how different measures are considered when determining the individualized treatment levels for patients.

- **Predictive:** Can we predict which treatment options are likely to be most effective for a given patient based on their personal health standpoint?

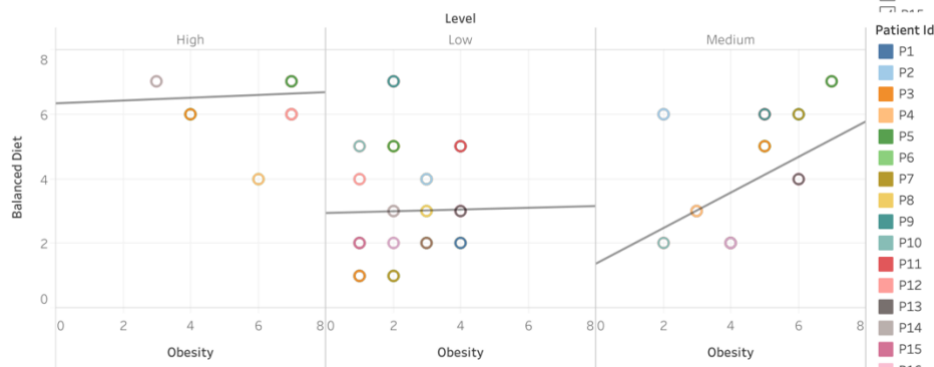
Story 3



Obesity to balanced diet relationship for different level of patient



high risk patients and balanced diet importance



Above visualization represents certain treatment method (Balanced Diet for the patient). A person with high obesity needs a more balanced and strict diet.

This bar chart shows the intensity of all the patients with obesity and how important balanced diet is for the patient.

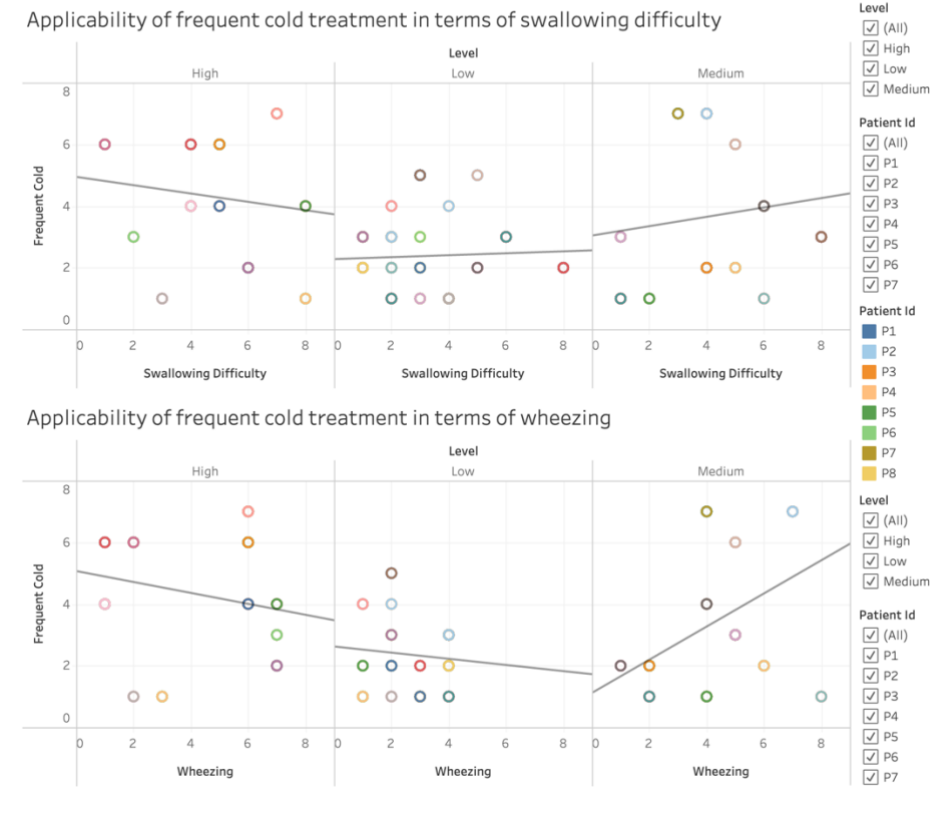
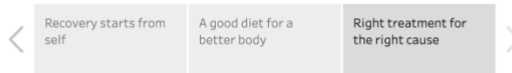
As observed from the distribution model, if the intensity of obesity is high only having balanced diet will not be helpful and patients will be more prone to be affected by lung cancer. In this high level of intensity other than the balanced diet the patients should consider different treatment methods.

However, for people at medium risk balanced diet will be more beneficial treat obesity and avoid being affected by lung cancer.

Overall, these graphs indicate that consuming a balanced diet is essential for sustaining good health and avoiding disease. As part of a healthy lifestyle, encouraging individuals to consume a balanced diet is highlighted by the visualizations.

- **Prescriptive:** How can we use this information to develop more personalized treatment plans and improve patient outcomes?

Story 3



This is the descriptive model of how can we use this information to plan right treatment for the causes.

For example, a person with high intensity of lung cancer having swallowing difficulty, frequent cold remedy will not be enough for them and they should consider other factors responsible. On the contrary, for a person with medium intensity of lung cancer, frequent cold remedy will be effective.

Similarly, a person with high intensity of lung cancer having wheezing problem, frequent cold remedy will not be enough for them and considering other factors will be helpful. However, for a person with medium intensity of lung cancer, frequent cold remedy will be beneficial.

CONCLUSION

Lung cancer is a highly complex disease that necessitates an all-encompassing strategy to enhance patient outcomes. Our data analysis suggests that multiple risk factors, including environmental exposures, lifestyle factors, and underlying medical conditions, contribute to the development of lung cancer. Identifying and addressing modifiable risk factors, such as quitting smoking, reducing exposure to air pollution and other environmental pollutants, and promoting healthy lifestyle choices, can aid in reducing the incidence of lung cancer.

In addition, the analysis reveals that patient outcomes are influenced by a variety of factors, including medical history, treatment options, and social determinants of health. Personalized treatment regimens that take these variables into consideration can enhance patient outcomes. For patients to receive the best possible care and support, effective communication and collaboration among healthcare providers, patients, and their families are essential.

To achieve these objectives, it is essential to develop improved screening and prevention strategies, such as assessing high-risk individuals and instituting interventions supported by scientific evidence to reduce modifiable risk factors. It is also crucial to foster collaboration between healthcare providers, researchers, and policymakers in order to identify and resolve knowledge deficits about lung cancer and to develop effective interventions to enhance patient outcomes.

WHAT HAVE WE LEARNT FROM DATA VISUALIZATION CONCEPTS?

We have gained several valuable insights about lung cancer and its effects on patients thanks to these data visualization concepts.

Several risk factors, including smoking, exposure to environmental contaminants, and genetics, are associated with a higher incidence of lung cancer.

Second, we have investigated the effect of various interventions on patient outcomes and identified factors significantly associated with positive outcomes, such as early detection and individualized treatment plans.

Finally, we have seen how data visualization can facilitate the communication of complex medical information in a plain and understandable manner, making it simpler for patients, healthcare providers, and other stakeholders to comprehend the factors that influence patient outcomes and collaborate to improve them. We can develop more effective lung cancer prevention, diagnosis, and treatment strategies by utilizing data visualization techniques and technologies.

HOW HAS THE DATA VISUALIZATION PROCESS HELPED YOU FIND THE ANSWERS TO THE QUESTIONS YOU DEVELOPED?

The visualization of data assisted us in recognizing patterns, trends, and connections within the data. By visualizing the data, we were able to identify critical insights, including the distribution of risk factors and outcomes, the efficacy of various treatments, and the relationship between variables.