Advanced data visualization

Experiment-3

Name	Hardik Garg
UID	2021300036
Batch	Batch H
Department	COMPS A

Aim: Design Interactive Dashboards and Storytelling using Tableau / Power BI / R (Shiny) / Python(Streamlit/Flask) / D3.js to be performed on the dataset - Disease spread / Healthcare

Dataset: The dataset aims to explore the relationships between various factors and the diagnosis of asthma, potentially helping in understanding how different demographic groups are affected by this condition.

Key Features of the Dataset

1. **ID**: A unique identifier for each individual in the dataset.
2. **Age**: This column represents the age of each individual. Age can be a significant factor in asthma prevalence, as certain age groups may be more susceptible to asthma.
3. **Gender**: This feature is encoded as 0 for Male and 1 for Female. Gender differences can influence asthma prevalence and severity, often requiring gender-specific analysis.
4. **Ethnicity**: Ethnic background of each individual. Ethnicity is important as it might reflect genetic predispositions, cultural practices, or socioeconomic factors that could affect asthma risk.
5. **Diagnosis**: This column indicates whether the individual has been diagnosed with asthma or not. Typically, it's encoded as 0 for "No Asthma" and 1 for "Asthma". This variable is the primary outcome of interest.
6. **LungFunctionFEV1**: This feature likely represents the Forced Expiratory Volume in 1 second (FEV1), a critical measure of lung function. FEV1 is often used in asthma studies to assess the severity of airway obstruction.
7. **Environmental Exposures** (Potential): These may include factors such as smoking status, pollution levels, or exposure to allergens. Such data would help to identify environmental triggers that might exacerbate asthma symptoms.
8. **Family History** (Potential): Whether the individual has a family history of asthma or other respiratory conditions, which could indicate genetic susceptibility.
9. **Other Health Metrics** (Potential): There might be additional health metrics such as Body Mass Index (BMI), presence of other respiratory conditions, or comorbidities that could influence asthma risk or outcomes.

Dashboard

1. Age-related Lung Function Decline:

- The scatter plot titled "Lung Function vs Age" shows a general trend of decreasing lung function (FEV1) as age increases. This is expected as lung capacity naturally diminishes with age.
- The line graph "Average Lung Function (FEV1) by Age" further emphasizes this trend, demonstrating a clear downward slope in average FEV1 values across age groups.

2. Higher Asthma Prevalence in Younger Age Groups:

- The bar chart "Asthma Prevalence by Age" indicates that asthma is more prevalent in younger age groups, particularly between the ages of 10 and 30. This suggests that genetic factors and environmental exposures during childhood may play a significant role in asthma development.

3. Distinct Lung Function Patterns by Diagnosis:

- The box plot "Lung Function Distribution by Diagnosis" reveals that individuals with asthma generally have lower lung function compared to those without asthma. The median FEV1 values are significantly lower in the asthma group, indicating a more pronounced decline in lung function.

- The distribution of lung function within each group also differs. The asthma group shows a wider spread of values, suggesting greater variability in lung function among individuals with asthma.

4. Low Prevalence of Asthma Overall:

- The pie chart "Diagnosis Distribution" highlights that asthma is relatively uncommon in the overall population. The majority of individuals (94.85%) do not have asthma, while only 5.15% are diagnosed with the condition. This suggests that while asthma affects a significant number of people, it remains a minority diagnosis.

1. Lung Function vs Age:
- Shows a scatter plot of lung function vs age
- Data points are colored blue for no asthma and black for asthma
- Lung function values range from about 1.5 to 4 on the y-axis

2. Asthma Prevalence by Age:
- Bar chart showing asthma prevalence across different age groups
- Prevalence ranges from 0 to about 0.15 (15%)
- No clear trend visible in prevalence across age groups

3. Age Distribution:
- Histogram showing age distribution of patients
- Ages range from about 20 to 80 years old
- Most patients appear to be between 40-60 years old

4. Lung Function by Diagnosis:
- Box plots comparing lung function for asthma vs no asthma groups
- Median lung function appears similar between groups
- Asthma group shows slightly wider range of lung function values

5. Lung Function Trends by Age:
- Line graph showing average lung function (FEV1) by age for asthma and no asthma groups
- Both groups show fluctuations in lung function across ages
- No asthma group generally has higher lung function, especially in older ages

6. Diagnosis Distribution:
- Pie chart showing proportion of patients with and without asthma
- Vast majority (94.81%) do not have asthma
- Only a small slice (5.19%) represents patients with asthma

Conclusion:

Based on the data presented:

1. Asthma affects a small portion of the study population (5.19%).

2. Lung function varies widely across ages for both asthmatic and non-asthmatic individuals.
3. Non-asthmatic individuals generally maintain slightly higher lung function, especially at older ages.
4. The impact of asthma on lung function appears moderate rather than severe for most affected individuals.
5. Age alone does not strongly predict lung function, suggesting other factors are involved.

This study provides insights into the relationship between asthma, age, and lung function, highlighting the complex nature of respiratory health across different demographics.