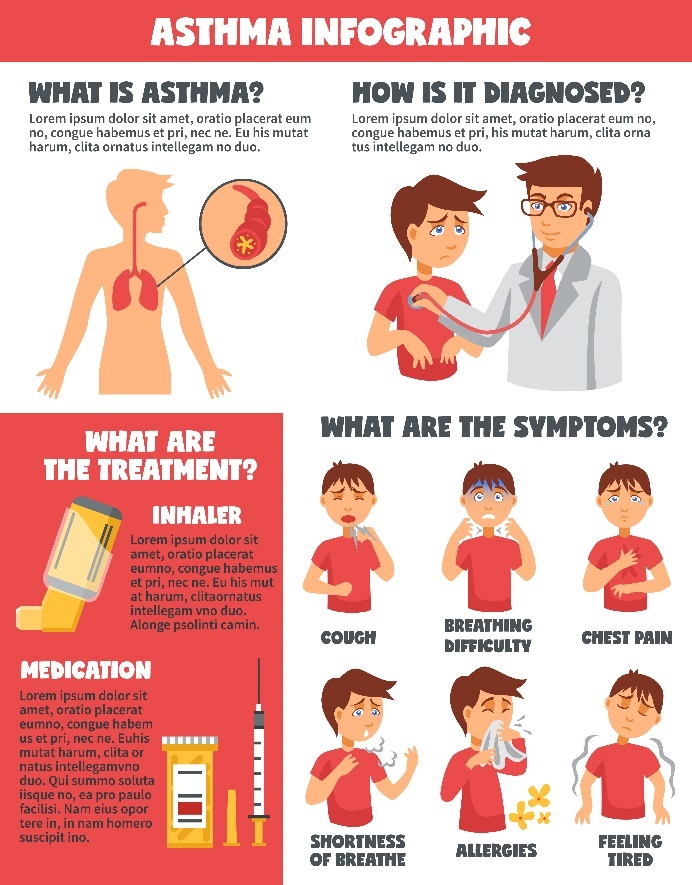
**WHAT IS ASTHMA**

Asthma is a chronic respiratory condition characterized by inflammation and narrowing of the airways, which makes breathing difficult. It can cause recurrent episodes of wheezing, breathlessness, chest tightness, and coughing, especially at night or early in the morning. Asthma symptoms can range from mild to severe and can vary over time.

Triggers for asthma attacks can include allergens (like pollen, dust mites, or animal dander), respiratory infections, air pollutants, cold air, physical exertion, and certain medications or food additives. Management typically involves avoiding triggers where possible and using medications like bronchodilators (to relax the airway muscles) and anti-inflammatory drugs (to reduce airway inflammation).

It's important for individuals with asthma to work closely with healthcare providers to develop a personalized asthma action plan to effectively manage their condition and minimize the impact on their daily life.



**Project description:**

**Asthma Disease Dataset**

**About Dataset:**

This dataset contains extensive health information for 2,392 patients diagnosed with Asthma Disease, uniquely identified with IDs ranging from 5034 to 7425. It includes demographic details, lifestyle factors, environmental and allergy factors, medical history, clinical measurements, symptoms, and a diagnosis indicator. This dataset is valuable for researchers and data scientists aiming to explore factors associated with Asthma, develop predictive models, and conduct statistical analyses.

Tables of the content:

* Patient Information
* Patient ID
* Demographic Details
* Lifestyle Factors
* Environmental and Allergy Factors
* Medical History
* Clinical Measurements
* Symptoms
* Diagnosis Information
* Confidential Information

**Patient Information**

**Patient ID**

* **Patient ID**: A unique identifier assigned to each patient (5034 to 7425).

**Demographic Details**

* **Age**: The age of the patients ranges from 5 to 80 years.
* **Gender**: Gender of the patients, where 0 represents Male and 1 represents Female.
* **Ethnicity**:

**What is ethnicity:**

Ethnicity refers to the social group that a person belongs to, characterized by a shared culture, language, ancestry, and often, geographic origin. Unlike race, which is often associated with physical characteristics, ethnicity is more about cultural identity. Ethnic groups share common cultural traits, such as traditions, customs, values, and sometimes religion or language.

The ethnicity of the patients, coded as follows:

* + Hispanic or Latino
  + Asian (e.g., Chinese, Indian, Filipino)
  + Black or African American
  + White (e.g., European ancestry)
  + Native American or Alaska Native
  + Pacific Islander (e.g., Hawaiian, Samoan)
* **Education Level**: The education level of the patients, coded as follows:

0: None

1: High School

2: Bachelor's

3: Higher

**Lifestyle Factors**

* **BMI**: Body Mass Index of the patients, ranging from 15 to 40.
* **Smoking**: Smoking status, where 0 indicates No and 1 indicates Yes.
* **Physical Activity**: Weekly physical activity in hours, ranging from 0 to 10.
* **Diet Quality**: Diet quality score, ranging from 0 to 10.
* **Sleep Quality**: Sleep quality score, ranging from 4 to 10.

**Environmental and Allergy Factors**

* **Pollution Exposure**: Exposure to pollution, score from 0 to 10.

Air pollution is the presence of one or more contaminants in the atmosphere, such as dust, fumes, gas, mist, odour, smoke or vapor, in quantities and duration that can be injurious to human health. The main pathway of exposure from air pollution is through the respiratory tract.

* **Pollen Exposure**: Exposure to pollen, score from 0 to 10.

Pollen is an airborne allergen that can affect our health. Pollen grains are tiny “seeds” dispersed from flowering plants, trees, grass, and weeds. The amount and type of pollen in the air depends on the season and geographic region. Though pollen counts are typically higher during the warmer seasons, some plants pollinate year-round.

* **Dust Exposure**: Exposure to dust, score from 0 to 10.

Dusts are tiny solid particles scattered or suspended in the air. The particles are "inorganic" or "organic," depending on the source of the dust. Inorganic dusts can come from grinding metals or minerals such as rock or soil. Examples of inorganic dusts are silica, asbestos, and coal.

* **Pet Allergy**: Pet allergy status, where 0 indicates No and 1 indicates Yes.

Pet allergy is an allergic reaction to proteins found in an animal's skin cells, saliva or urine. Signs of pet allergy include those common to hay fever, such as sneezing and runny nose. Some people may also experience signs of asthma, such as wheezing and difficulty breathing.

**Medical History:**

* **Family History Asthma**: Family history of asthma, where 0 indicates No and 1

indicates Yes.

* **History Of Allergies**: History of allergies, where 0 indicates No and 1 indicates Yes.
* **Eczema**:
  + - Eczema is a skin condition that causes dry and itchy patches of skin. It’s a common condition that isn’t contagious. Symptoms of eczema can flare up if you contact an irritant or an allergen. There are treatments available to help you manage symptoms, but there isn’t a cure.
    - Presence of eczema, where 0 indicates No and 1 indicates Yes.
* **Hay Fever**:
  + - Hay fever, also called allergic rhinitis, causes cold-like symptoms. These may include a runny nose, itchy eyes, congestion, sneezing and sinus pressure. But unlike a cold, hay fever isn't caused by a virus. Hay fever is caused by an allergic response to a harmless outdoor or indoor substance the body identifies as harmful (allergen).

Presence of hay fever, where 0 indicates No and 1 indicates Yes.

**Gastroesophageal Reflux**:

* **Gastroesophageal reflux disease** (**GERD**) a condition in which acidic gastric fluid flows backward into the oesophagus, resulting in heartburn: “frequent heartburn may be an indication of gastroesophageal reflux disease.

Presence of gastroesophageal reflux, where 0 indicates No and 1 indicates Yes.

**Clinical Measurements**

* **Lung Function FEV1**: Forced Expiratory Volume in 1 second (FEV1), ranging from 1.0 to 4.0 Liters.

FEV1 helps measure the progression of lung conditions such as chronic obstructive pulmonary disease (COPD) or asthma. FEV stands for forced expiratory volume, which is the air you exhale in 1 second. A low FEV1 suggests a breathing obstruction.

* **Lung Function FVC**:

forced vital capacity (FVC) is the amount of air that can be forcibly exhaled from your lungs after taking the deepest breath possible. It's measured by spirometry, which is a common breathing test to check lung function. Forced Vital Capacity (FVC), ranging from 1.5 to 6.0 Liters.

**Symptoms**

* **Wheezing**:

Wheezing is the shrill whistle or coarse rattle you hear when your airway is partially blocked. It might be blocked because of an allergic reaction, a cold, bronchitis or allergies. Wheezing is also a symptom of asthma, pneumonia, heart failure and more. It could go away on its own, or it could be a sign of a serious condition.

Presence of wheezing, where 0 indicates No and 1 indicates Yes.

* **Shortness Of Breath**:

Breathing is an automatic action that you perform without thinking, but it’s not a reflex you should take for granted. If you’ve ever lost your breath, you know it can be a frightening event. The minute you have to think about breathing, you intuitively know that something is off. Shortness of breath can affect anyone, including healthy people. It’s a subjective feeling, so its presentation differs from person to person.

Presence of shortness of breath, where 0 indicates No and 1 indicates Yes.

* **Chest Tightness**:

Chest tightness is a sensation of pressure, discomfort, or constriction in the chest area that can cause shortness of breath.

Presence of chest tightness, where 0 indicates No and 1 indicates Yes.

* **Coughing**:

Coughing is a standard way of clearing your throat. When your airways become clogged with mucus or foreign particles such as smoke or dust, a cough serves as a reflexive reaction that helps clear the particles and make breathing easier.

Presence of coughing, where 0 indicates No and 1 indicates Yes.

* **Nighttime Symptoms**:

Nighttime or nocturnal asthma is a type of asthma where symptoms get worse at night. It is a very common type of asthma. About 75 percent of people with asthma get woken up by symptoms at least once a week. Nearly half of people with asthma experience symptoms every night

Presence of nighttime symptoms, where 0 indicates No and 1 indicates Yes.

* **Exercise Induced**:

Exercise-induced asthma (exercise-induced bronchoconstriction) happens when your airways get smaller during physical activity, triggering asthma symptoms. This might include coughing, wheezing or shortness of breath. Warming up, inhalers and other medications can prevent asthma episodes and open your airways when you do have an episode.

Presence of symptoms induced by exercise, where 0 indicates No and 1 indicates Yes.

**Diagnosis Information**

* **Diagnosis**:

Your doctor will perform a physical exam to rule out other possible conditions, such as a respiratory infection or chronic obstructive pulmonary disease (COPD). Your doctor will also ask you questions about your signs and symptoms and about any other health problems.

Diagnosis status for Asthma, where 0 indicates No and 1 indicates Yes.

**Tests to measure lung function:**

You may be given lung function tests to determine how much air moves in and out as you breathe. These tests may include:

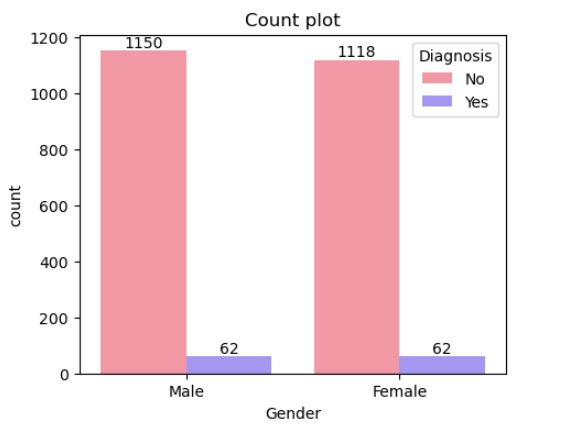
* + Spirometry
  + Peak flow

**Confidential Information**

* **Doctor In Charge**: This column contains confidential information about the doctor in charge, with "Dr Confide" as the value for all patients.

**PLOTS:**

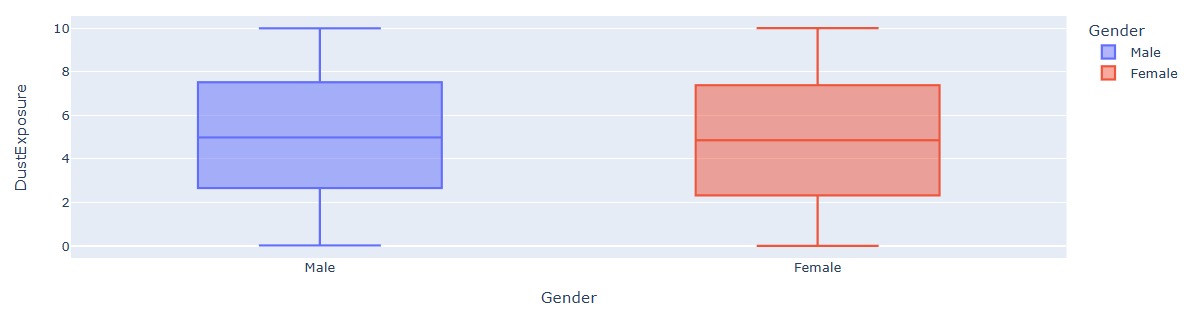
**COUNT PLOT**



Insight

* The count plot provided shows the distribution of asthma diagnosis among males and female.
* The plot compares the counts of males and female who were diagnosed with asthma and those who were not.
* The categories are split into "No" and "Yes “No indicating no asthma diagnosis, Yes indicate Yes asthma diagnosis
* There are more males (1150) than females (1118) The number of diagnosed cases is the same for both males and female 62.

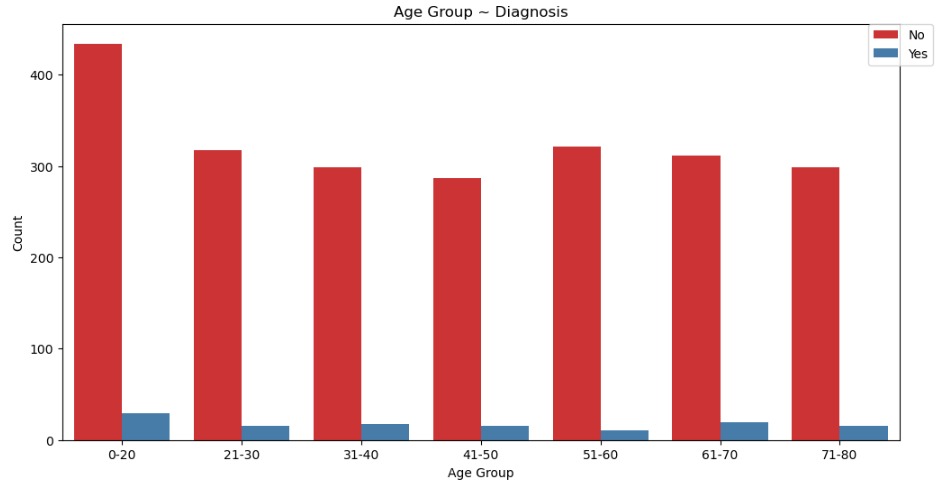
**BOX PLOT**



**I**nsight

* The box plot shows the distribution of Gender and Physical activity
* Male -Physical activity
* Median dust exposure is around 4.98.
* The interquartile range (IQR) is from approximately 2.655 to 7.515.
* Minimum dust exposure is 0.02.
* Maximum dust exposure is 9.99.
* Males have a wider range of dust exposure compared to females.
* The spread of dust exposure values is larger in males

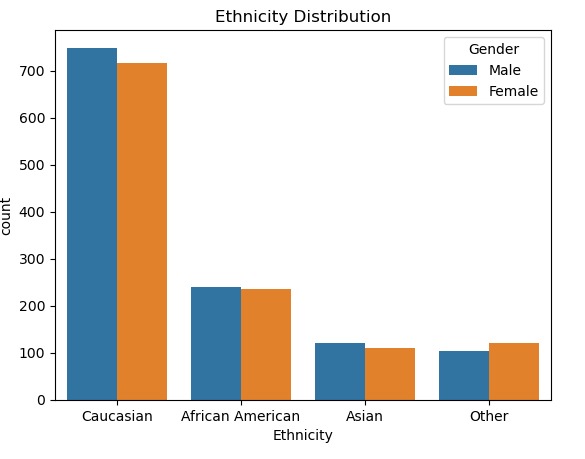
**BARPLOT**



Insight,

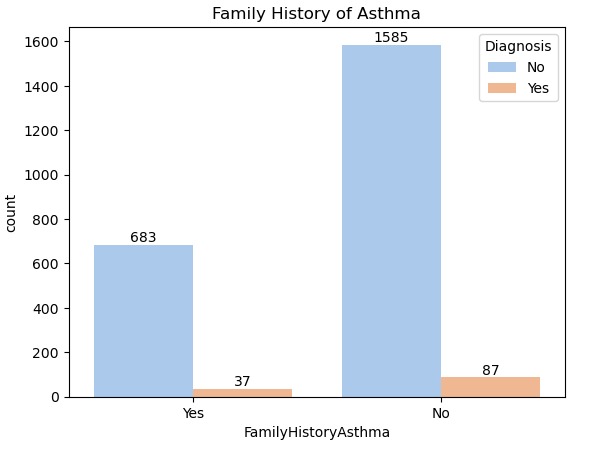
* The highest count of diagnoses as "No" is 0-20 age group.
* Next the highest age group is (51-60) The highest count of diagnoses as "yes" age group also 0-20.
* next highest count of diagnoses as "yes" age group is 61-70

**COUNT PLOT**



Insight

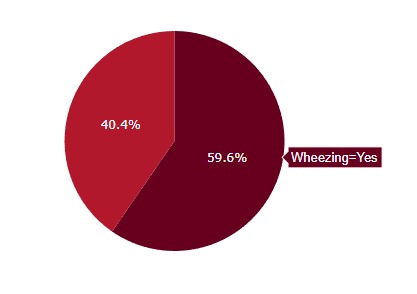
* The plot categorizes individuals into four ethnic groups: Caucasian, African American, Asian, and Other.
* Caucasian: The largest group, with approximately 700 males and 600 females.
* African American: The second largest group, with around 250 males and slightly fewer females.
* Asian: The third largest group, with close to 100 males and a similar number of females.
* Other: The smallest group, with less than 100 individuals of each gender.
* Overall, the plot shows that the Caucasian group has the highest count, followed by African American, Asian, and then Other.



Insight

* There is no link between family history of asthma Vs diagnosis because both categories of diagnosis ratio of similar
* Yes radio value 5.14, No radio value =5.203

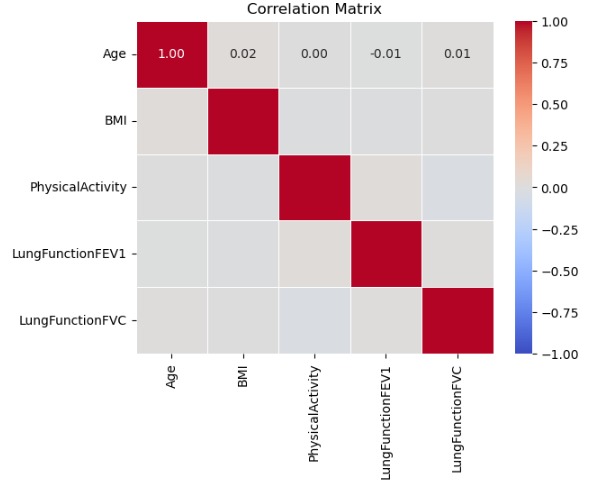
**PIE CHART**



Insight

* From Asthma Disease Dataset, we have totally 2,392 people's information.
* 59.6% of the people in the dataset have wheezing.
* 40.4% of the people in the dataset do not have wheezing.

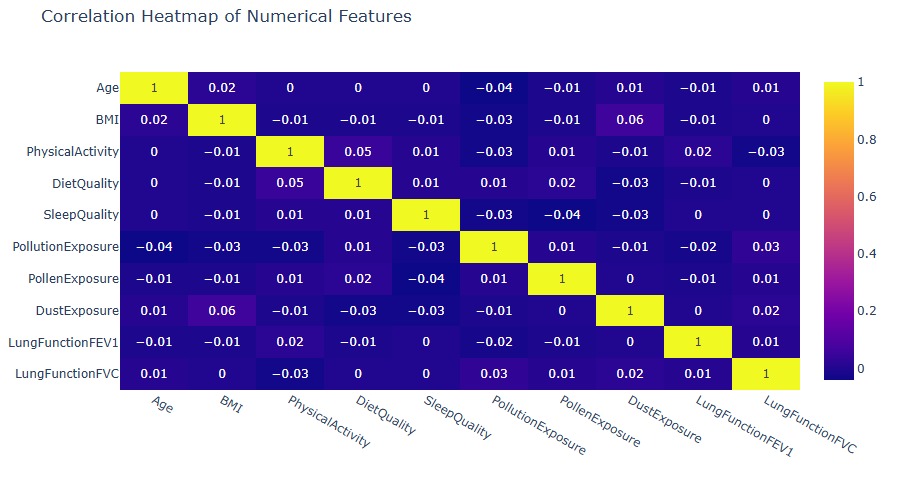
**HEATMAP**



Insight

* 0.01-Very weak negative correlation (almost no relationship)
* 0,01-Very weak positive correlation (almost no relationship)
* 0-no relation
* 0.02-Very weak positive correlation (no relationship)
* Most of the variables show very weak correlations with each other, with correlation coefficients close to zero.
* This indicates that there is little to no linear relationship between these variables

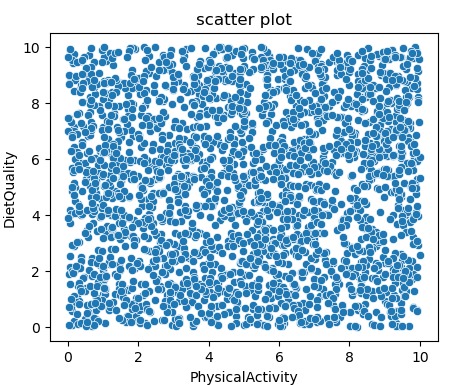
**HEATMAP**



Insight

* There are no correlation between these column.
* based on our data there are no correlation.

**SCATTERPLOT**



Insight

* The scatter plot shared visualizes the relationship between two variables: "Physical Activity" (x-axis) and "Diet Quality" (y-axis).
* There is no clear linear trend observable in the scatter plot.
* The scatter plot indicates that there is a wide range of "Physical Activity" and "Diet Quality" among the individuals with no strong correlation between the two variables.

**Statistical Analysis:**

**2 Sample Independent T-Test**

A two-sample independent t-test compares the means of two independent groups to determine if there is a statistically significant difference between them. It assumes that the samples are randomly selected, the populations are normally distributed, and the variances are equal. The test calculates the t-statistic, which measures the difference between the group means relative to the variability of the samples. A p-value is then derived from the t-statistic to assess the significance of the observed difference. If the p-value is below a chosen significance level (e.g., 0.05), the null hypothesis of no difference is rejected.

* t-statistic: 0.15667650344092288
* p-value: 0.8755130390832333
* If the p-value < alpha reject the null hypothesis.
* p-value > alpha, fail to reject the null hypothesis.

**Hypotheses**

* **Null Hypothesis (H0):** The null hypothesis states that there is no significant difference in the mean age between the male and female groups. (population mean of male is equal to female)
* **Alternative Hypothesis (HA):** The alternative hypothesis states that there is a significant difference in the mean age between the male and female groups. (population mean of male is different from female)

Since the p-value is much greater than 0.05, we fail to reject the null hypothesis. we have enough evidence to support null hypothesis. population mean of male is equal to female.

**Chi-square t-test**

The chi-square test assesses whether there is a significant association between categorical variables. It compares the observed frequencies in each category to the expected frequencies, which are calculated under the assumption that the variables are independent. The test statistic is calculated by summing the squared differences between observed and expected frequencies, divided by the expected frequencies. A large chi-square value indicates a greater discrepancy between observed and expected frequencies, suggesting an association. The significance is determined by comparing the test statistic to a chi-square distribution with appropriate degrees of freedom.

The chi-square test was conducted to determine if there is a significant association between two categorical variables (e.g., "Family history of asthma" and "diagnosis ").

* Chi-Square Statistic: 1.240214194954464
* p-value: 0.871434267986058
* Degrees of Freedom (dof): 4

**Hypotheses**

* **Null Hypothesis:** There is no link between the two categorical variables
* **Alternative Hypothesis:** There is an link between the two categorical variables
* Significance Level alpha: 0.05.
* If the p-value < alpha reject the null hypothesis.
* p-value > alpha, fail to reject the null hypothesis.
* p-value: 0.871434267986058

Since the p-value is much greater than 0.05, we fail to reject the null hypothesis. we have enough evidence to support null hypothesis, where there is no significant link between "Family History Asthma" and "Diagnosis".

**CONCLUSION OF THE PROJECT:**

In conclusion, asthma is a chronic respiratory disease that requires ongoing management to prevent symptoms and exacerbations. With the proper treatment and lifestyle modifications, people with asthma can lead active and healthy lives.