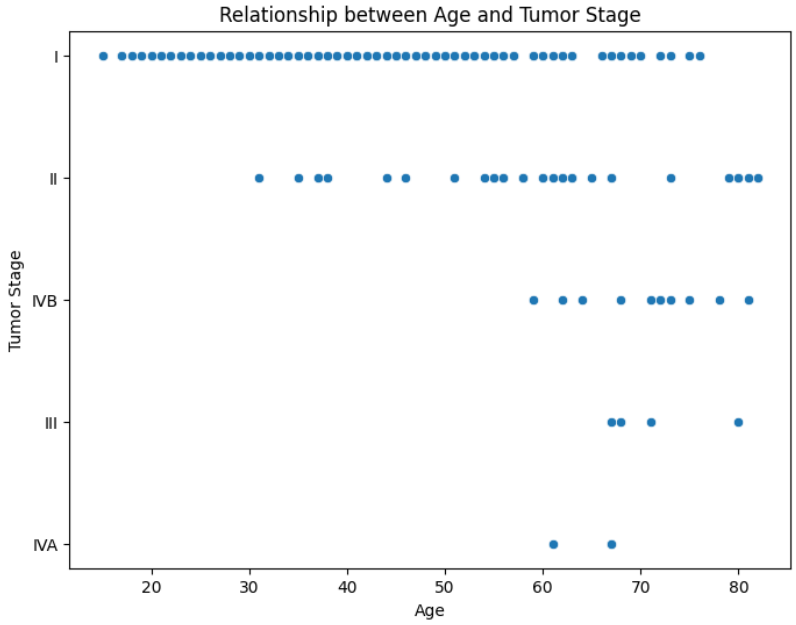
Statistics and Trends

Exploratory Data Analysis of Thyroid Dataset

Introduction:

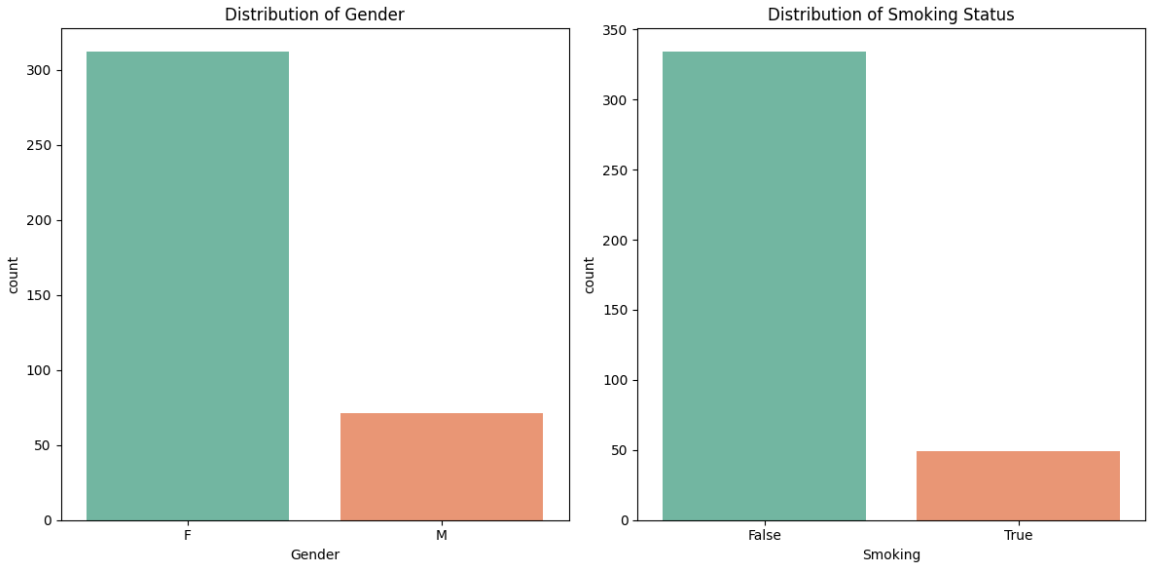
This report provides a comprehensive analysis of the thyroid disease dataset. The dataset includes various attributes related to patients and their conditions, allowing for an in-depth examination of potential patterns and correlations.

1. Relational Graph: Relationship between Age and Tumor Stage



Most patients are in Tumor Stage I, regardless of age. Higher tumor stages (II, III, IVA, and IVB) are more common in older patients, indicating that age may be a risk factor for advanced tumor stages.

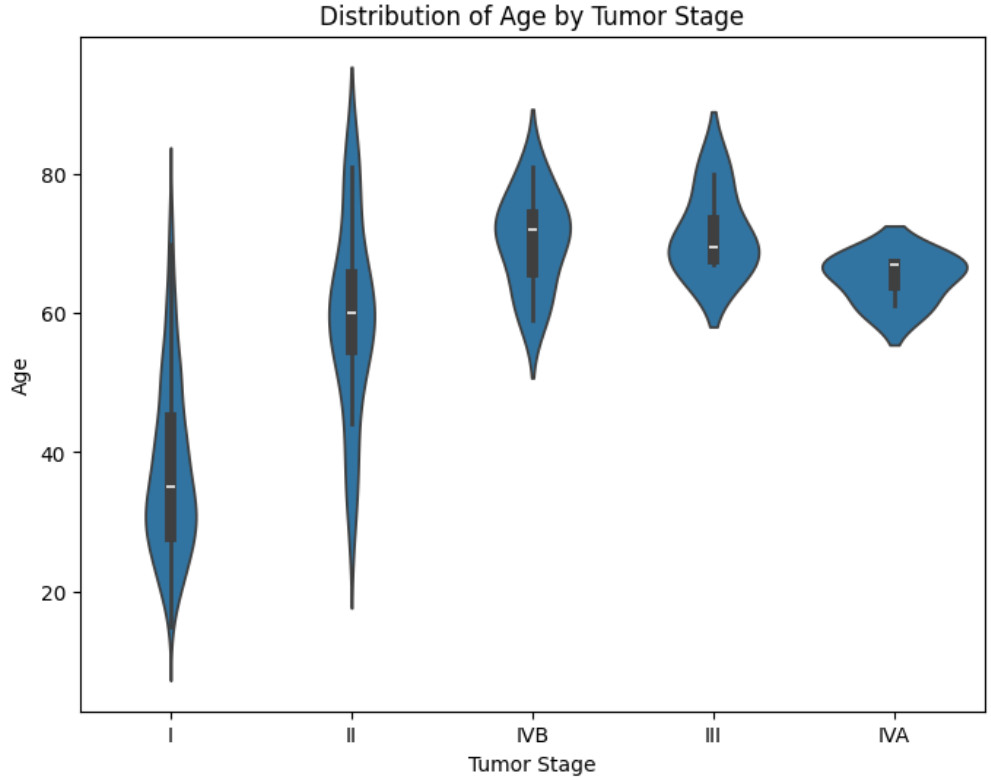
2. Histogram: Distribution of Fuel Types



Gender Distribution: A significant gender disparity exists, with more female patients (312) compared to male patients (71). Smoking Status: Most patients are non-smokers (342), indicating that smoking is not prevalent in this dataset.

3. Statistical Graph: Distribution of Age by Tumor Stage

Tumor Stage I: Includes a wide age range, from young adults to the elderly. Tumor Stage II: Higher concentration of patients in their 50s and 60s. Tumor Stage IVB: Generally older patients, peaking around 60 to 70 years. Tumor Stages III and IVA: Tend to occur in older patients, but smaller sample sizes make it difficult to draw definitive conclusions.



Conclusion:

The analysis reveals that most thyroid tumor cases are diagnosed at Stage I across various age groups, indicating early detection. Additionally, the dataset highlights a significant disparity in gender distribution, with a higher prevalence of cases among females and non-smokers, suggesting potential gender-based and lifestyle-related factors in thyroid disease incidence.

Name: Yuvaraj Jammi

Student ID: 23023543

Code: https://github.com/Yuva3543/ADS1-statistics-and-trends

DataSet: https://www.kaggle.com/datasets/jainaru/thyroid-disease-data