**Thyroid Cancer Dataset**

Dataset : https://www.kaggle.com/datasets/abuchionwuegbusi/thyroid-cancer-recurrence-prediction/data

**Analysis and Interpretation:**

1. **Data Distribution and Imbalance:**
   * **The distribution of the target variable (Recurred) should be examined to identify any imbalance, as this can affect the performance of predictive models.**
   * **The demographic and clinical characteristics (e.g., age, gender, smoking history) should be analyzed to understand the profile of the patient population.**
2. **Correlation Analysis:**
   * **Investigate the correlation between features and the target variable to identify which features are most strongly associated with recurrence.**
   * **Analyze correlations among features to detect multicollinearity, which could influence model performance.**

**A black and white text

Description automatically generatedCorrelation:**

Note :

Conclusions from Correlation matrix :-

* Age has a negative correlation with Recurred -25.9897%.
* Gender has a positive correlation with Recurred 32.8189%.
* Smoking has a positive correlation with Recurred 33.2343%.
* Hx Radiotherapy has a positive correlation with Recurred 17.4487%.
* Physical Examination has a slight positive correlation with Recurred 14.8016%.
* Adenopathy has a slight positive correlation with Recurred 5.30636%.
* Pathology has a negative correlation with Recurred -3.27172%.
* Focality has a positive correlation with Recurred 38.3776%.
* Risk has a positive correlation with Recurred 73.3376%.
* T has a negative correlation with Recurred -55.6201%.
* N has a negative correlation with Recurred -63.2323%.
* M has a negative correlation with Recurred -35.456%.
* Stage has a negative correlation with Recurred -49.4137%.
* Response has a negative correlation with Recurred -78.9057%.

SO ,

A screen shot of a computer program

Description automatically generated

A bar chart with purple rectangular bars

Description automatically generatedSome Visualization to Describe the columns :

Overview to Relation between Risk Column and Stage Column

Overview to Relation between Recurred Column and Risk Column

Overview of Smoking Column

A graph of bar chart

Description automatically generatedA graph with purple bars

Description automatically generated

A bar chart with purple rectangles

Description automatically generated

Overview of T (Tumor Stage) Column

Overview to Relation between Focality Column and Gender Column

Overview to Relation between Gender Column and Smoking Column

A bar chart with different bars

Description automatically generatedA graph with purple bars

Description automatically generated