Hormonal Behaviour and Thyroid Detection Using EDA

Tanishq Varpe — SOE23201030051
November 15, 2025

1 Introduction

This mini project focuses on understanding patterns related to thyroid disorders using Exploratory Data Analysis (EDA). Thyroid disorders are influenced by factors such as age, gender, hormonal imbalance, and medical history. By analyzing the cleaned thyroid dataset, we aim to identify relationships among clinical factors and how they impact thyroid function.

2 Objective

To apply EDA techniques and visualize relationships among medical parameters that influence thyroid disorders. The aim is to interpret hidden patterns that help understand factors contributing to normal vs abnormal thyroid conditions.

3 Dataset Description

The dataset used is **Thyroid-Cleaned.csv**, which contains medical information related to thyroid patients. This dataset was cleaned before analysis by removing missing values, fixing inconsistent data, and selecting relevant columns.

Key Columns Used

- **age** - Age of the patient.
- **sex** - Gender of the patient.
- **tsh** - Thyroid Stimulating Hormone level.
- **t3** - Triiodothyronine hormone level.
- **tt4** - Total Thyroxine level.
- **t4u** - Thyroxine Uptake.
- **class** - Thyroid diagnosis category (e.g., Normal, Hypothyroid, Hyperthyroid).

4 Steps Followed

1. Import Libraries

Loaded R libraries such as `ggplot2`, `dplyr`, and `tidyr` for data cleaning and visualization.

2. Load Dataset

Used `read.csv()` to load the cleaned thyroid data.

3. Inspect and Clean Data

Performed the following:

- Removed missing values using `na.omit()`.
- Eliminated duplicate records using `distinct()`.
- Converted character columns into factors.

4. Descriptive Statistics

Calculated summary statistics using `summary()` to understand the distribution of age, hormone levels, and diagnostic class.

5. Correlation Analysis

Computed pairwise correlations for numeric variables:

- TSH vs T3
- TSH vs TT4
- T3 vs T4U

This helped identify strong and weak relationships among hormone levels.

6. Advanced Data Visualization

Created multiple advanced plots to explore deeper medical relationships:

- **Age vs TSH scatter plot** to observe how thyroid function varies with age.
- **Hormone Interaction Plot (T3 vs TT4)** to analyze hormone behavior.
- **Class-wise Mean TSH Bar Plot** to compare hormone levels across thyroid classes.
- **Gender vs Disorder Bar Plot** to check if thyroid disorders are more common in females.
- **TSH vs T3 Scatter Plot by Class** to see separation between normal and abnormal thyroid cases.

7. Interpretation of Visual Findings

Each plot was interpreted to understand medical behavior of thyroid hormones and diagnostic patterns.

5 Results and Observations

- **Age Influence:** Older individuals showed higher TSH levels, indicating increased risk of hypothyroidism with age.
- **Gender Patterns:** Females had a significantly higher number of thyroid disorder cases compared to males.
- **Hormonal Relationships:**
  - High TSH was associated with low T3 and T4, a typical indicator of hypothyroidism.
  - Hyperthyroid patients showed very low TSH and elevated T3/T4 values.
- **Visualization Findings:**
  - Scatter plots confirmed clear separation between normal and abnormal thyroid groups.
  - Strong correlations were seen between T3 and TT4 suggesting linked hormone secretion.

6 Conclusion

This mini project successfully demonstrates how EDA can reveal meaningful medical insights from thyroid data. By analyzing hormone interactions, age patterns, and diagnostic categories, we observed:

- Age and gender significantly influence thyroid disorders.
- TSH is a strong indicator of thyroid health and shows clear separation between normal and abnormal classes.
- Hormone relationships confirm medical expectations of thyroid functionality.

Overall, this analysis highlights the importance of EDA in understanding health datasets and identifying medical patterns that support clinical decision-making.