Final Term – Proposal

I am interested in thyroid cancer data because, despite being considered one of the less aggressive cancers and often categorized as "benign" in many cases, it is becoming increasingly prevalent, yet it doesn't receive the same level of attention as other cancers like skin or breast cancer. This lack of awareness surrounding thyroid cancer can lead to underdiagnosis, delayed detection, and a general lack of understanding about its risks and potential long-term impact on health.

By working with thyroid cancer data, I hope to uncover valuable insights that can contribute to identifying risk factors, enhancing prevention strategies, and ultimately improving patient outcomes. This is especially important in a time where data-driven approaches are increasingly shaping the future of healthcare. I believe that through careful analysis and a deeper understanding of thyroid cancer, we can help to bring more awareness to the disease and make a meaningful difference in how it is diagnosed and treated.

List of Features:

- Age: numerical
- Gender: categorical
- Country: categorical
- Ethnicity: categorical
- Family_History: categorical
- Radiation_Exposure: categorical
- Iodine_Decficiency: categorical
- Smoking: categorical
- Obesity: categorical
- Diabetes: categorical
- TSH_Level: numerical
- T3_Level: numerical
- T4_Level: numerical
- Nodule_Size: numerical
- Thyroid_Cancer_Risk: categorical
- Diagnosis: categorical

There is no need for feature engineering for EDA processes as there is no missing value.

Plots for numerical features: Boxplot, Heatmap

Plots for categorical features: Histogram, Bar Chart, Pie Chart, Tree Map, Sunburst