

Team 2: Adam, Babatola, Iqra and Thinley

Thyroid Cancer Risk Analysis

Providing an in-depth analysis of dataset and uncovering key insights into which demographics are at high risk of having thyroid cancer.

Overview

Providing an in-depth analysis of dataset and uncovering key insights into which demographics are at risk of having thyroid cancer:

- **Hypothesis 1:** Age and gender influence thyroid cancer risk, with older individuals and females having a higher probability
 - ❑ **Validation:** Use box plots and regression analysis to explore how cancer risk varies across different age groups and gender distributions
- **Hypothesis 2:** Certain countries and ethnicities have a higher prevalence of thyroid cancer due to genetic and environmental factors
 - ❑ **Validation:** Conduct geospatial analysis and visualize the distribution of thyroid cancer cases across different regions.

Planning & Design

Ideation

Project Goal: Build interactive dashboards for data analysis

Medical Use Case: Improve decision-making through data insights of thyroid cancer risk based on demographics

Target Audience: Medical professionals, analysts, WHO and decision-makers

Design

User Stories: "As a Medical practitioner, we want to explore demographics of thyroid cancer risk dynamically."

Intuitive UI: Clean layouts, easy navigation

Accessibility: Readable colours, labeling

Interactivity: Clickable filters, zoomable charts and maps

Hypothesis: Which demographic has high risk of thyroid cancer

Technologies

Tools: Visual Studio Code, Jupyter Notebook, Power BI, PowerPoint

Wireframing: Balsamic Wireframes

Project Management: GitHub Projects, Google Meets

Version Control: GitHub for collaboration

Libraries & Frameworks: Python (Pandas, NumPy, Plotly, Seaborn, Ipywidgets, StatsModels), Power BI

Project Board

</

Assigned Tasks: 30

MoSCoW Prioritisation:

Must have: an interactive dashboard and proper documentation

Should have: simple and followable code

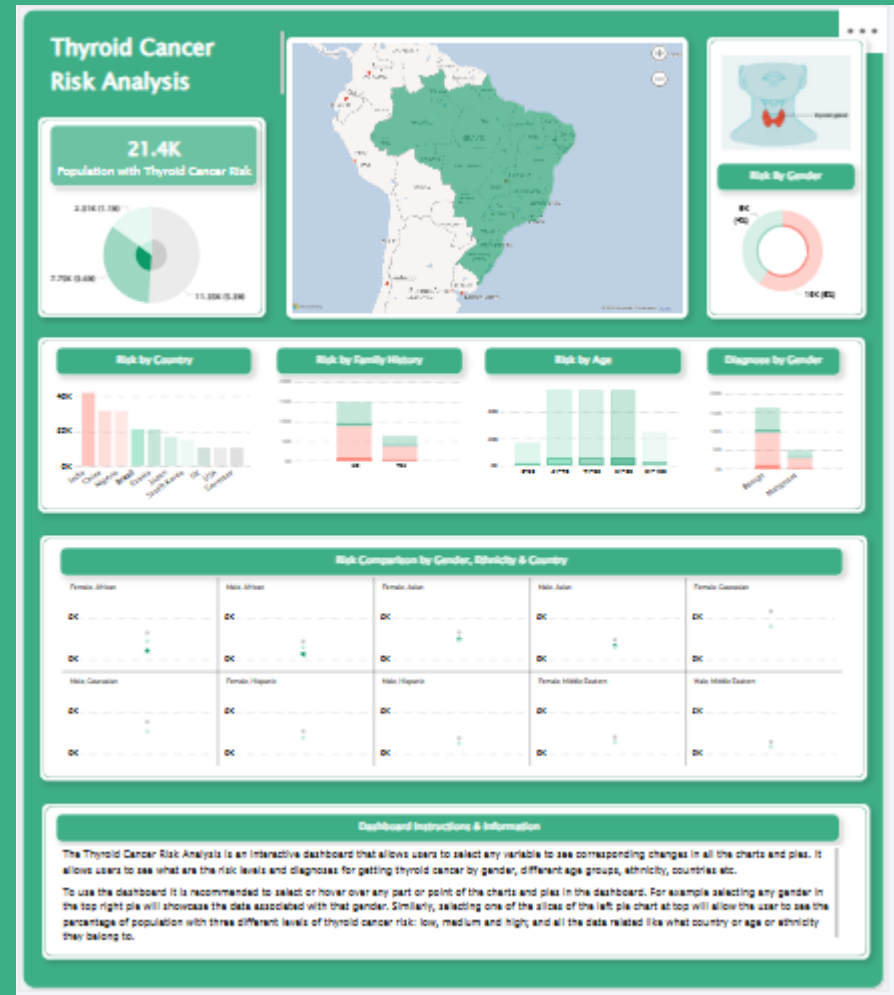
Could have: nicely stylized code and dashboard

Would have: utilize the dataset to full potential

Project Backlog: Setting VS Code and Github collaboration

Features

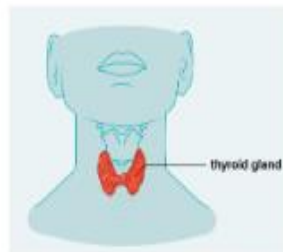
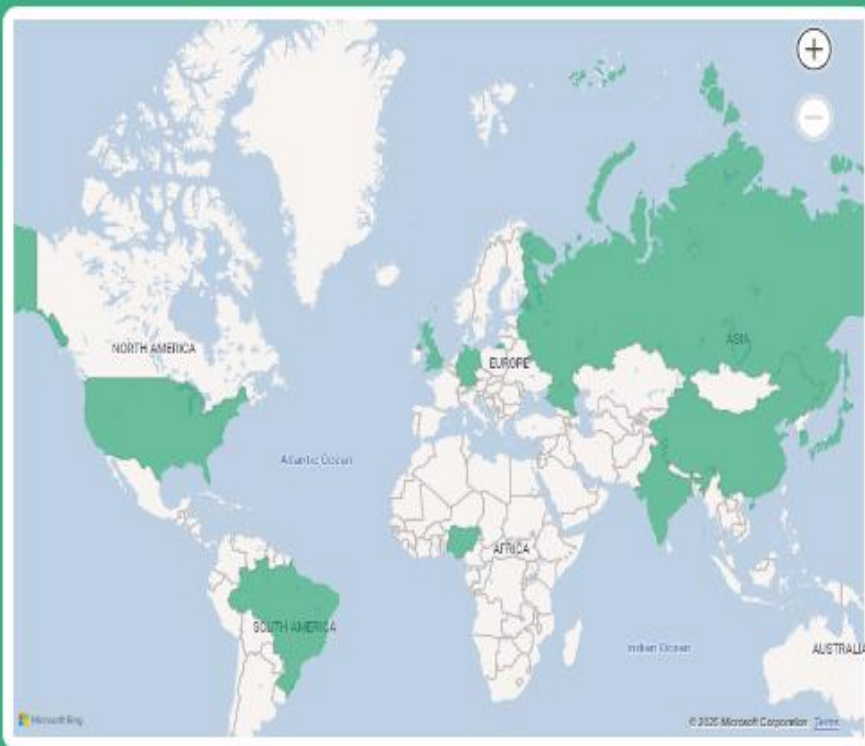
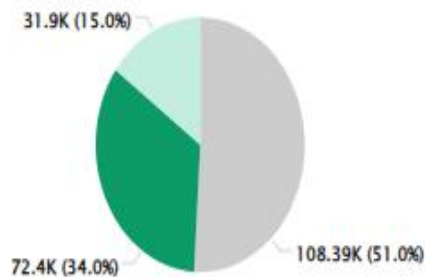
1. **Interactive:** all variables i.e. age, gender, country, ethnicity, diagnosis are interlinked so selecting one will show all the relevant data across different charts and map
2. At the bottom there are instructions on using the different features of the dashboard
3. Hovering over any chart will further explain the data specific e.g. regarding gender, amount of population at risk, country etc



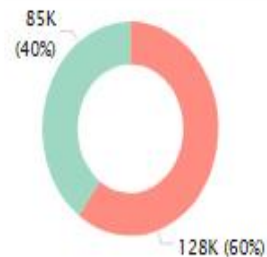
Thyroid Cancer Risk Analysis

212.7K

Population with Thyroid Cancer Risk



Risk By Gender



Risk by Country

Risk by Family History

Risk by Age

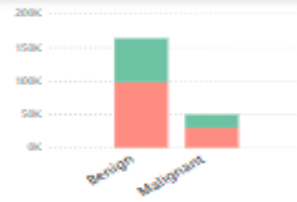
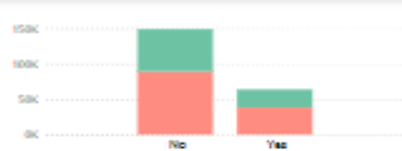
Diagnose by Gender

40K

200K

150K

200K



Risk Comparison by Gender, Ethnicity & Country

Female, African



Male, African



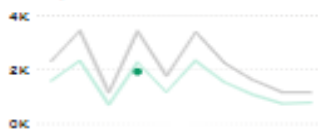
Female, Asian



Male, Asian



Female, Caucasian



Male, Caucasian



Female, Hispanic



Male, Hispanic



Female, Middle Eastern



Male, Middle Eastern

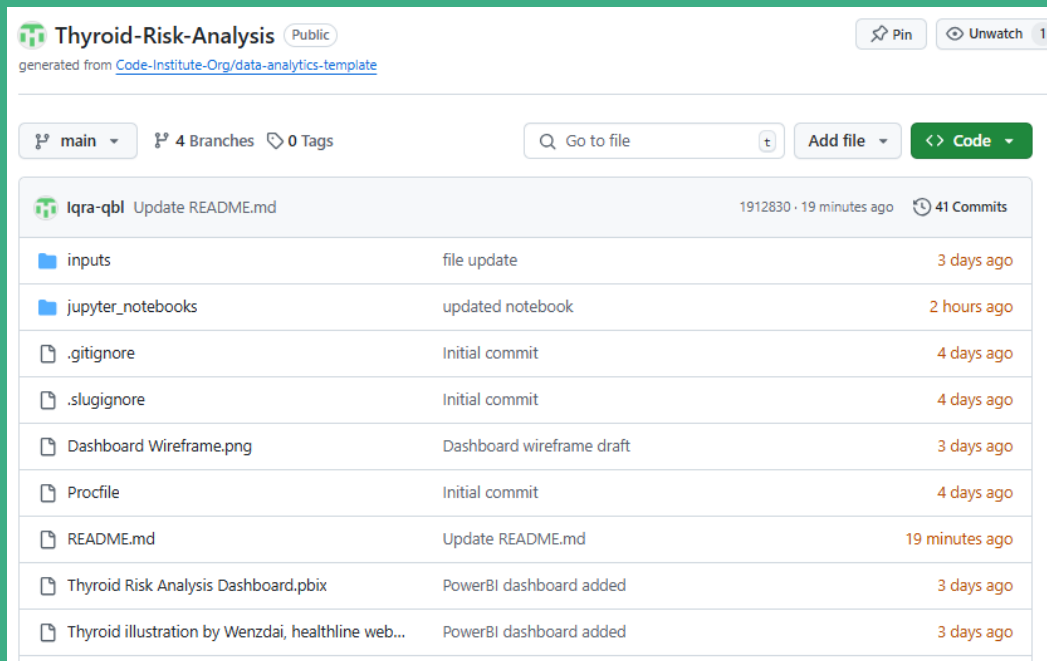


Dashboard Instructions & Information

The Thyroid Cancer Risk Analysis is an interactive dashboard that allows users to select any variable to see corresponding changes in all the charts and pies. It allows users to see what are the risk levels and diagnoses for getting thyroid cancer by gender, different age groups, ethnicity, countries etc.

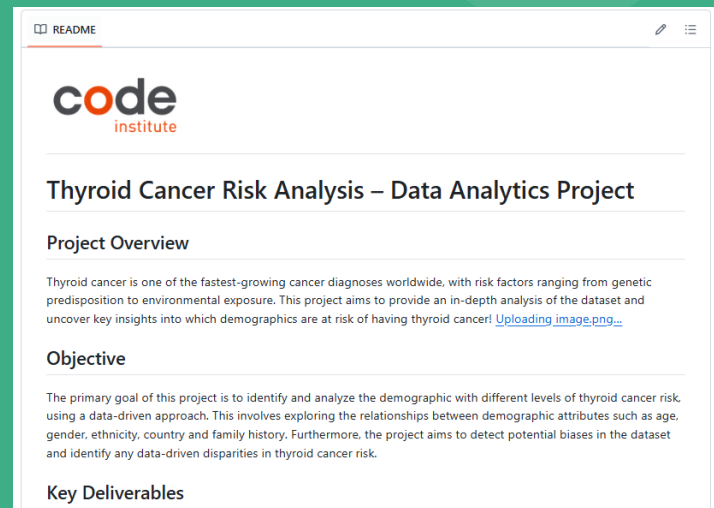
To use the dashboard it is recommended to select or hover over any part or point of the charts and pies in the dashboard. For example selecting any gender in the top right pie will showcase the data associated with that gender. Similarly, selecting one of the slices of the left pie chart at top will allow the user to see the percentage of population with three different levels of thyroid cancer risk: low, medium and high; and all the data related like what country or age or ethnicity they belong to.

Documentation, Testing & Version Control



The screenshot shows the GitHub interface for the 'Thyroid-Risk-Analysis' repository. At the top, it indicates the repository is 'Public' and was generated from a template. Below this, there are buttons for 'Pin' and 'Unwatch'. The main section displays the repository's structure, including a list of files and folders. The files listed are: 'inputs' (file update, 3 days ago), 'jupyter_notebooks' (updated notebook, 2 hours ago), '.gitignore' (Initial commit, 4 days ago), '.slugignore' (Initial commit, 4 days ago), 'Dashboard Wireframe.png' (Dashboard wireframe draft, 3 days ago), 'Procfile' (Initial commit, 4 days ago), 'README.md' (Update README.md, 19 minutes ago), 'Thyroid Risk Analysis Dashboard.pbix' (PowerBI dashboard added, 3 days ago), and 'Thyroid illustration by Wenzdai, healthline web...' (PowerBI dashboard added, 3 days ago).

File/Folder	Commit Message	Time Ago
inputs	file update	3 days ago
jupyter_notebooks	updated notebook	2 hours ago
.gitignore	Initial commit	4 days ago
.slugignore	Initial commit	4 days ago
Dashboard Wireframe.png	Dashboard wireframe draft	3 days ago
Procfile	Initial commit	4 days ago
README.md	Update README.md	19 minutes ago
Thyroid Risk Analysis Dashboard.pbix	PowerBI dashboard added	3 days ago
Thyroid illustration by Wenzdai, healthline web...	PowerBI dashboard added	3 days ago



The screenshot shows the README file for the 'Thyroid Cancer Risk Analysis' project. The README is titled 'Thyroid Cancer Risk Analysis – Data Analytics Project' and includes a 'Project Overview' section. The overview states that thyroid cancer is one of the fastest-growing cancer diagnoses worldwide, with risk factors ranging from genetic predisposition to environmental exposure. The project aims to provide an in-depth analysis of the dataset and uncover key insights into which demographics are at risk of having thyroid cancer. A link to 'Uploading image.png...' is provided. The README also includes an 'Objective' section, which states that the primary goal of the project is to identify and analyze the demographic with different levels of thyroid cancer risk using a data-driven approach. This involves exploring the relationships between demographic attributes such as age, gender, ethnicity, country and family history. Furthermore, the project aims to detect potential biases in the dataset and identify any data-driven disparities in thyroid cancer risk. Finally, the README includes a 'Key Deliverables' section.

Thyroid Cancer Risk Analysis – Data Analytics Project

Project Overview

Thyroid cancer is one of the fastest-growing cancer diagnoses worldwide, with risk factors ranging from genetic predisposition to environmental exposure. This project aims to provide an in-depth analysis of the dataset and uncover key insights into which demographics are at risk of having thyroid cancer! [Uploading image.png...](#)

Objective

The primary goal of this project is to identify and analyze the demographic with different levels of thyroid cancer risk using a data-driven approach. This involves exploring the relationships between demographic attributes such as age, gender, ethnicity, country and family history. Furthermore, the project aims to detect potential biases in the dataset and identify any data-driven disparities in thyroid cancer risk.

Key Deliverables

GitHub Repository: <https://github.com/lqra-qbl/Thyroid-Risk-Analysis>

Insights & Finding

Key Data Insights

- **Caucasians** are the most represented group, with 63,669 entries, while the **Middle Eastern** individuals are the least represented, with only 21,335 entries, **Asians** with 53,261 entries, **Africans** with 42,414 entries, and **Hispanics** with 32,012 entries, showing an uneven distribution
- Females are more likely to have thyroid cancer especially Asian females
- Chi-square test p-value = 0.5102 ($p > 0.05$), which means no statistically significant difference in the diagnosis proportions across the genders

- Chi-squared test p-value for age groups: 0.5586, analysis does not show any bias in thyroid cancer diagnosis based on age groups within this dataset
- Most countries show a similar risk distribution: **Low risk (~53–54%)** with most common category, **Medium risk (~35–36%) follows**, High risk (~10–11%) is the least frequent in most countries. This suggests that thyroid cancer risk is generally low in most global regions
- India Shows a Unique Risk Pattern High-risk individuals make up 32.86% of the population (significantly higher than other countries)
- Japan (10.06%) and South Korea (10.50%) have the lowest high-risk proportions



Collaboration & Outcomes

Outcomes

Are you happy with the final product?
Yes

What do you hope to achieve in the next development cycle?

Fully utilize the dataset to its full potential

What would you do differently if you could start again?

We identified the bias but didn't take any action on it, we left the dataset as it was because of time constraints and no experience with handling this situation.

Development Problems

Problems that arose during development?:

Git collaboration

In group conflicts and resolutions?
No

Did you find any of the behaviour related content useful? Teamwork, problem solving etc?

Yes, mindful collaboration and problem- solving, work division

Interactivity: Overall good

Summary

Overall group dynamic:
Good, friendly, professional

Overall satisfaction: 9/10 (-1 for git)

What we learned: github collaboration

Our experiences: We had many issues with git commands but we had good troubleshooting and good mindset for the project



Q&A