

# India TB Dataset: Workshop Instructions

## Introduction

Welcome! For today's workshop, you'll create a short Quarto-based report exploring the India TB pathways dataset, incorporating data visualization with Plotly Express and tables with iTables. This will give you practice with report generation in Quarto.

## Steps

- 1) Download the dataset from this link, unzip it (extract all on Windows), then copy the .csv file into a data folder in your graph\_courses\_python repository:

[https://github.com/the-graph-courses/pbb\\_q4\\_2024\\_materials/raw/refs/heads/main/week\\_04\\_workshop/data/india\\_tb\\_pathways\\_and\\_costs.zip](https://github.com/the-graph-courses/pbb_q4_2024_materials/raw/refs/heads/main/week_04_workshop/data/india_tb_pathways_and_costs.zip)

(You can find out more about the dataset [here](#) if you're interested.)

- 2) Open your graph\_courses\_python folder in VS Code. Create a new Quarto file (.qmd) called week\_04\_workshop.qmd using the side panel.
- 3) Copy this YAML header into your file:

```
---  
title: "India TB Pathways Analysis"  
format: html  
author: "YOUR NAME HERE"  
---
```

(You may need to update this YAML later to change the execution options.)

- 4) Import the required packages (pandas, plotly.express, itables). (You should have installed these previously. But install them if you need to. For plotly.express, you install `plotly` not `plotly.express`.)
- 5) Under an appropriate section heading, add the following dataset description about the India TB Pathways dataset:

*The India TB Pathways dataset is based on a study conducted in Vellore district, Tamil Nadu, India. It examines the pre-diagnostic costs and health-seeking behavior of 880 new adult pulmonary tuberculosis patients registered at Primary Health Centres (PHCs).*

- 6) Load the data using `pd.read_csv()`
- 7) Use the `show()` function from `itables` to display the first few rows of your dataset
- 8) Download the following image of the Tamil Nadu map and add it to an `images` folder in your repository:

<https://imgur.com/VdhR2fr>

- 9) Embed that image in your report using the appropriate Quarto syntax. Include a relevant description.

Here is an example of how to embed an image:

```
![Map of Tamil Nadu, India. Vellore district in the north-east is where the study was conducted.](ima
```

- 10) Under an appropriate section heading, create a bivariate plot using Plotly Express that shows a relationship between two variables of your choice in the dataset. Add a brief interpretation of what your plot reveals.
- 11) The most important part: Render your report to an HTML format. Make sure it renders successfully. Your code should not be visible in the rendered output. (You will submit only your `.qmd` file, but from that file, we will be able to tell if your report renders successfully or not.)

## OPTIONAL CHALLENGE: Highlight Key Numbers with Inline Code

Use inline Python code to enhance your narrative. You might choose to highlight, for example, the maximum and minimum values of your numerical variables, or the most common values of your categorical variables.

Reference the Quarto documentation for guidance: <https://quarto.org/docs/computations/inline-code.html>

Here's a quick example using the `tips` dataset to show how to extract useful information:

```
tips = px.data.tips()
max_tip = int(tips.tip.max()) # int needed to print cleanly
common_day = tips.day.mode()[0]
```

You can then use these variables in your narrative like this:

The max tip amount was `{python} max_tip`.`

The most common day in the `tips` dataset was `{python} common_day`.`

The values there will then be dynamically inserted into your narrative. This allows you to highlight key numbers in your report without having to manually update the text every time you re-run your code or your dataset changes.

## Submission

To submit your work, please upload your .qmd file to the submission page. Make sure it renders successfully before submitting!