# Doctor Data Analysis

**Data Collection:**

As part of Data Collection, I have chosen a website named “www.drdata.in” which has list of many doctors, with their details. Then, I checked the required tags to get the information from. Later, I did the web scraping for the required information using BeautifulSoup and saved all the data in csv file at last.

A screenshot of a computer

Description automatically generated

**Data Cleaning:**

After scraping the relevant data from the websites, I have loaded the csv file into python using pandas, got an overview of how the data looks like and identified obvious issues that needed to be addressed. To handle unrelated data I have removed ‘Details’ column as it doesn’t contain any data. Also replaced blank values and null values such as ‘NA’, ‘null’, ‘ ’ with nan and dropped the nan. Removed all the duplicates and made sure the data has only distinct values drop\_duplicates()  
A screenshot of a computer

Description automatically generated

**Model Training:**

For model training, I have installed Private GPT to local repo and trained based on the cleaned csv file we had. A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

**Findings:**

* According to the model’s results we have came to know that the doctors who have been mostly specialized is in Allopathic Family Physician and the lowest being Gastro Physician.
* Number of Doctors who have graduated with a degree of BDS are 3.
* Top 5 states with respect to producing highest number of doctors are Madhya Pradesh, Uttar Pradesh, Gujarat, Bihar, Rajasthan.

**Model Performance:**

Overall Model Performance was decent. The model was able to return the answers with decent speed and accuracy. The ability to generate responses that are not only accurate but also contextually appropriate and tailored to the user’s needs or questions further boosts the model’s effectiveness. As the data was properly cleaned and accurate the model was able to read the file properly and gave the responses without any coherence and error rate at the accuracy. The model was able to adapt to new data and learn from feedback which significantly enhanced the performance over time.

Ultimately, the performance of private GPT model is judged by the continuous monitoring, testing and refining the model and it is necessary to maintain and improve the model’s performance.