DATA SCIENCE

PROJECT REPORT

(Project Semester January-April 2025)

**Hospital Directory Analysis and Visualization Using Python**

Submitted by

Aryan Agarwal

Registration No. 12324865

B.Tech CSE, Section K23KM

Course Code: INT375

Under the Guidance of

Mr. Anchal Kaundal, UID: 29612 , Assistant Professor

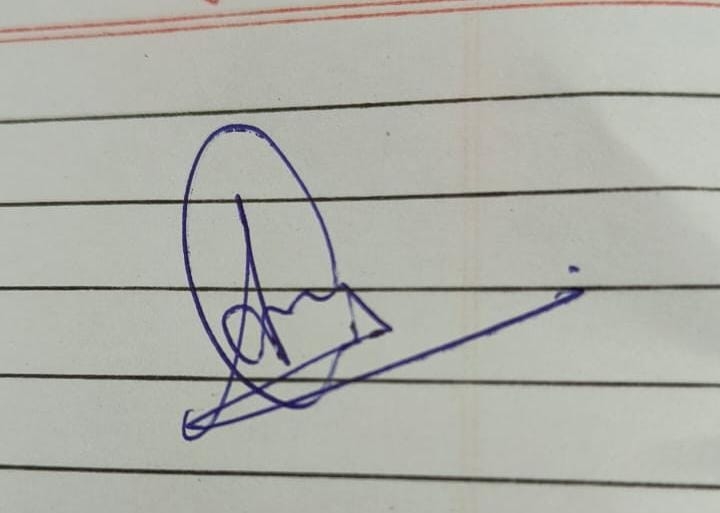
Discipline of CSE

**Lovely School of Computer Science & Engineering**

**Lovely Professional University, Phagwara**

# Declaration

I, Aryan Agarwal, student of B.Tech Computer Science and Engineering under CSE Discipline at Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 12/04/2025 Signature:   
Registration No. 12324865 Name: Aryan Agarwal

# Certificate

This is to certify that [Aryan Agarwal] bearing Registration no. [12324865] has completed [INT375] project titled, **“**Hospital Directory Analysis and Visualization Using Python**”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

**Signature and Name of the Supervisor: Mr. Anchal Kaundal**

**Designation of the Supervisor**

**School of Computer Science and Engineering**

Lovely Professional University

Phagwara, Punjab.

Date: 12/04/2025

# Acknowledgement

I express my sincere gratitude to my faculty guide, Mr. Anchal Kaundal, for his valuable support and timely guidance. I also thank the School of Computer Science and Engineering, LPU, for providing the resources necessary for this project.

# 1. Introduction

This project focuses on analyzing the hospital directory data of India to gain insights into the distribution of hospitals across states, categories, and bed availability. The objective is to visualize this data through Python and derive useful conclusions that can assist in policy-making, resource allocation, and healthcare analytics.

# 2. Source of dataset

The dataset used in this project was obtained from the open government health directories. It includes detailed information on hospitals across different Indian states, including hospital name, address, contact numbers, category, and number of beds.

# 3. EDA process

The Exploratory Data Analysis (EDA) was performed using:  
- Pandas: For data manipulation  
- Seaborn and Matplotlib: For data visualization  
- Steps:  
 - Data cleaning (handling missing values)  
 - Data transformation  
 - Grouping and aggregation  
 - Creating visual plots to derive insights

# 4.1. State-wise Hospital Distribution

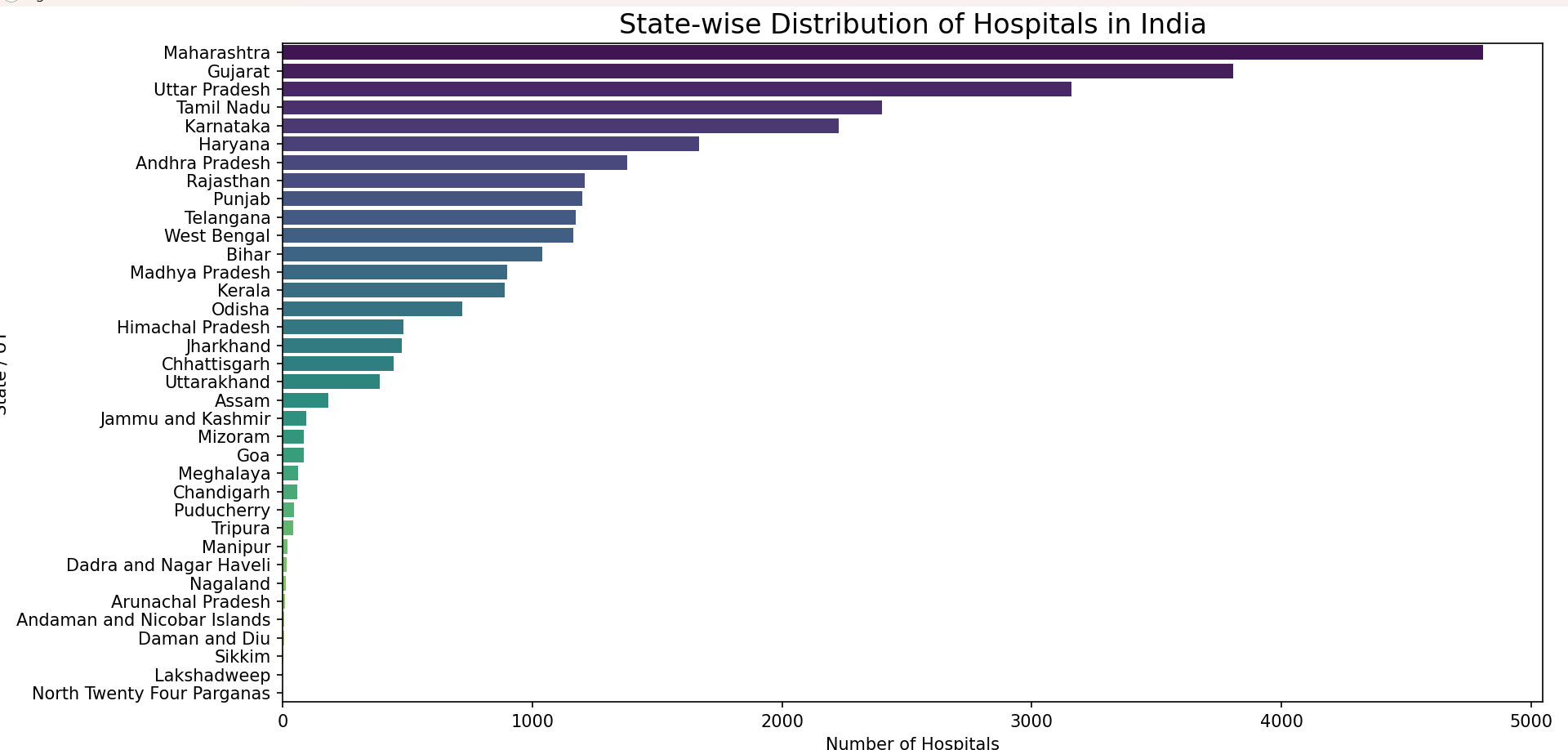
Introduction  
This analysis helps identify which states have the highest number of hospitals.

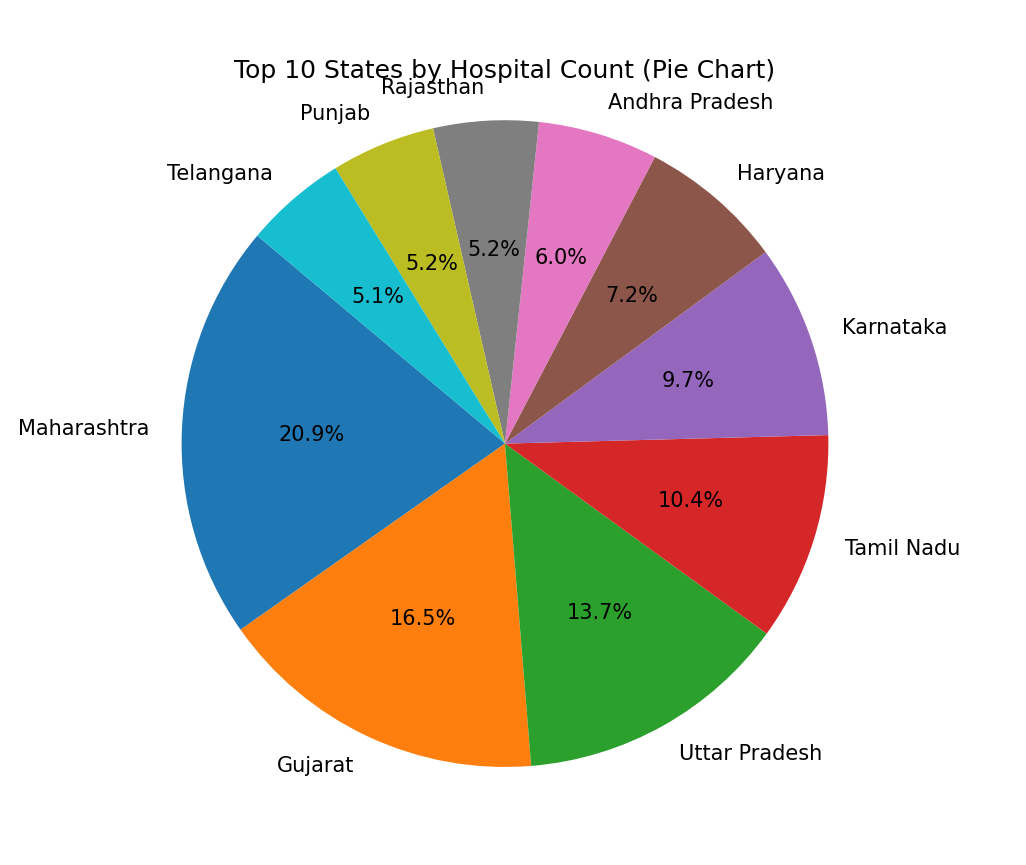
General Description  
We used value counts grouped by the State column to get hospital counts per state.

Functions and Formulas  
- value\_counts() for counting hospitals  
- barplot and pie for visual representation

Analysis Results  
States like Maharashtra, Uttar Pradesh, and Tamil Nadu have the highest number of hospitals.

Visualization  
- Horizontal bar plot of hospital counts by state  
- Pie chart for top 10 states





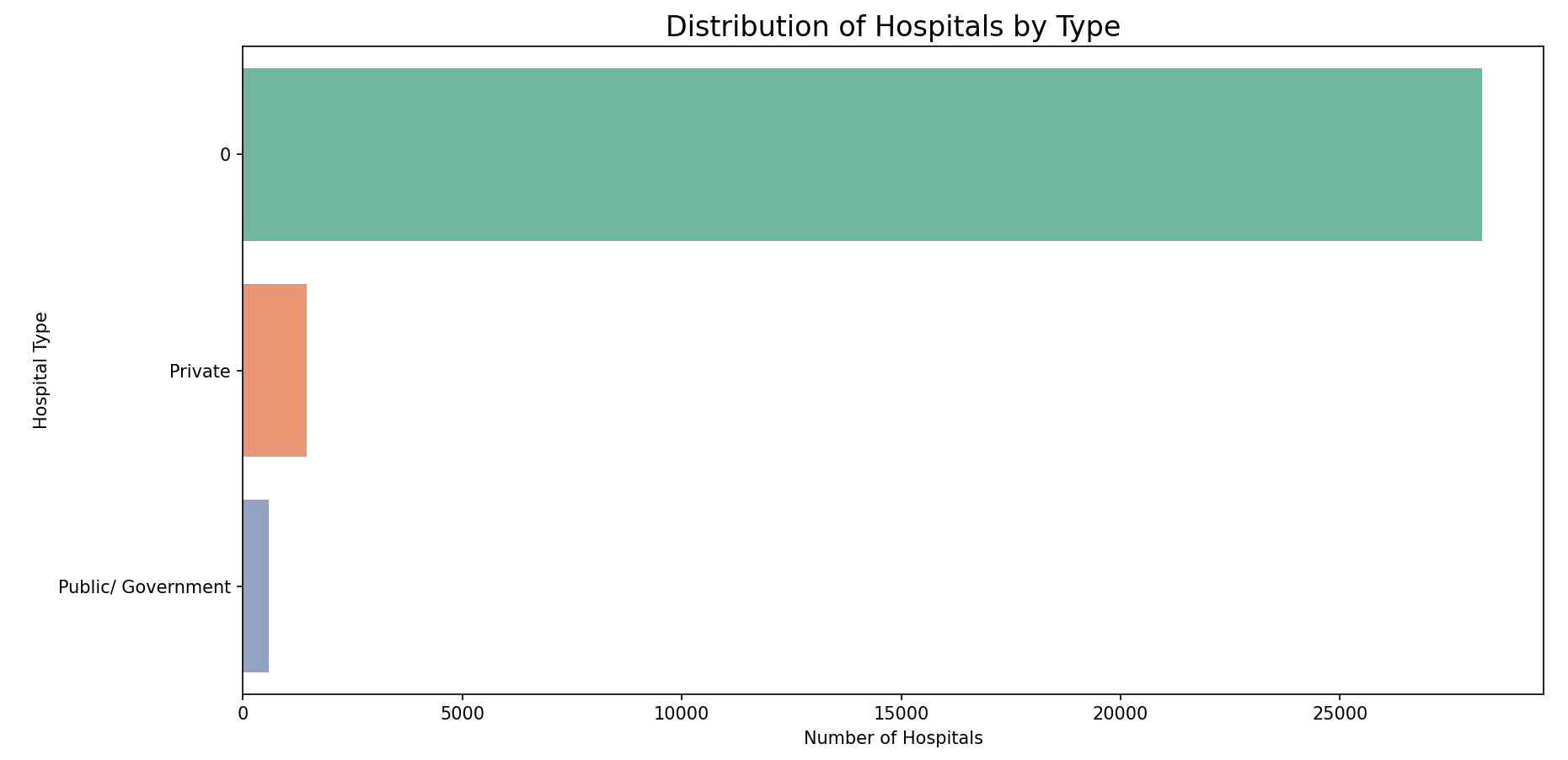
# 4.2. Hospital Type Distribution

Introduction  
Shows the distribution of hospitals by their categories (Govt, Pvt, etc.)

Functions and Formulas  
- value\_counts() on Hospital\_Category  
- groupby + unstack() for heatmap

Results  
Government hospitals are predominant.  
A heatmap visualizes the distribution across states.

Visualization  
- Bar plot of categories  
- Heatmap of state-wise categories



# 

# 4.3. Data Cleaning & Missing Value Handling

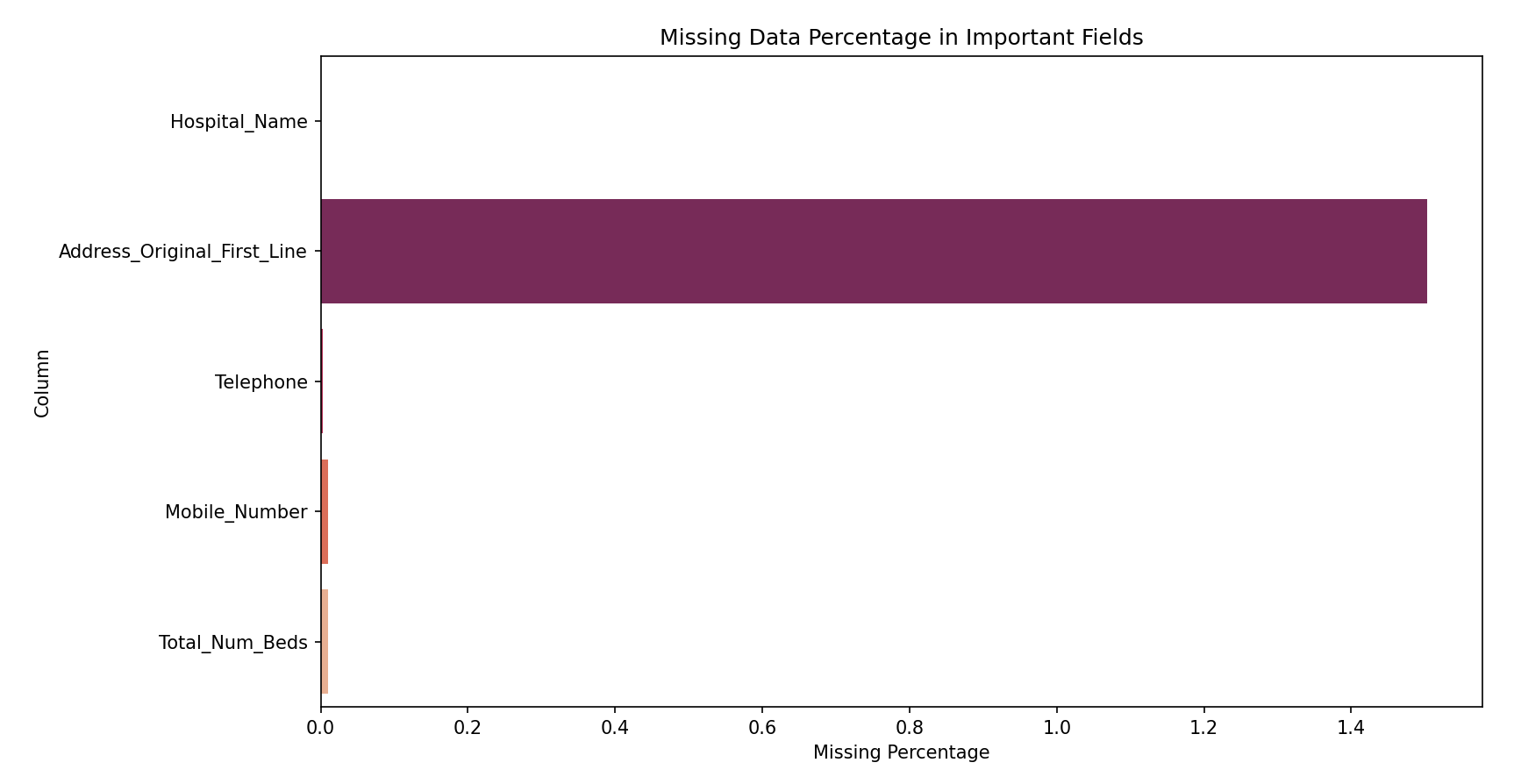
Introduction  
Addresses missing values in critical fields.

Steps Taken  
- Analyzed % of missing values in key columns  
- Replaced missing values where possible (e.g., contact numbers)

Functions and Formulas  
- isnull().mean()  
- fillna() for imputing values  
- dropna() where data was unusable

Results  
Data was cleaned and prepared for accurate analysis.

Visualization  
- Bar chart showing % of missing data per column



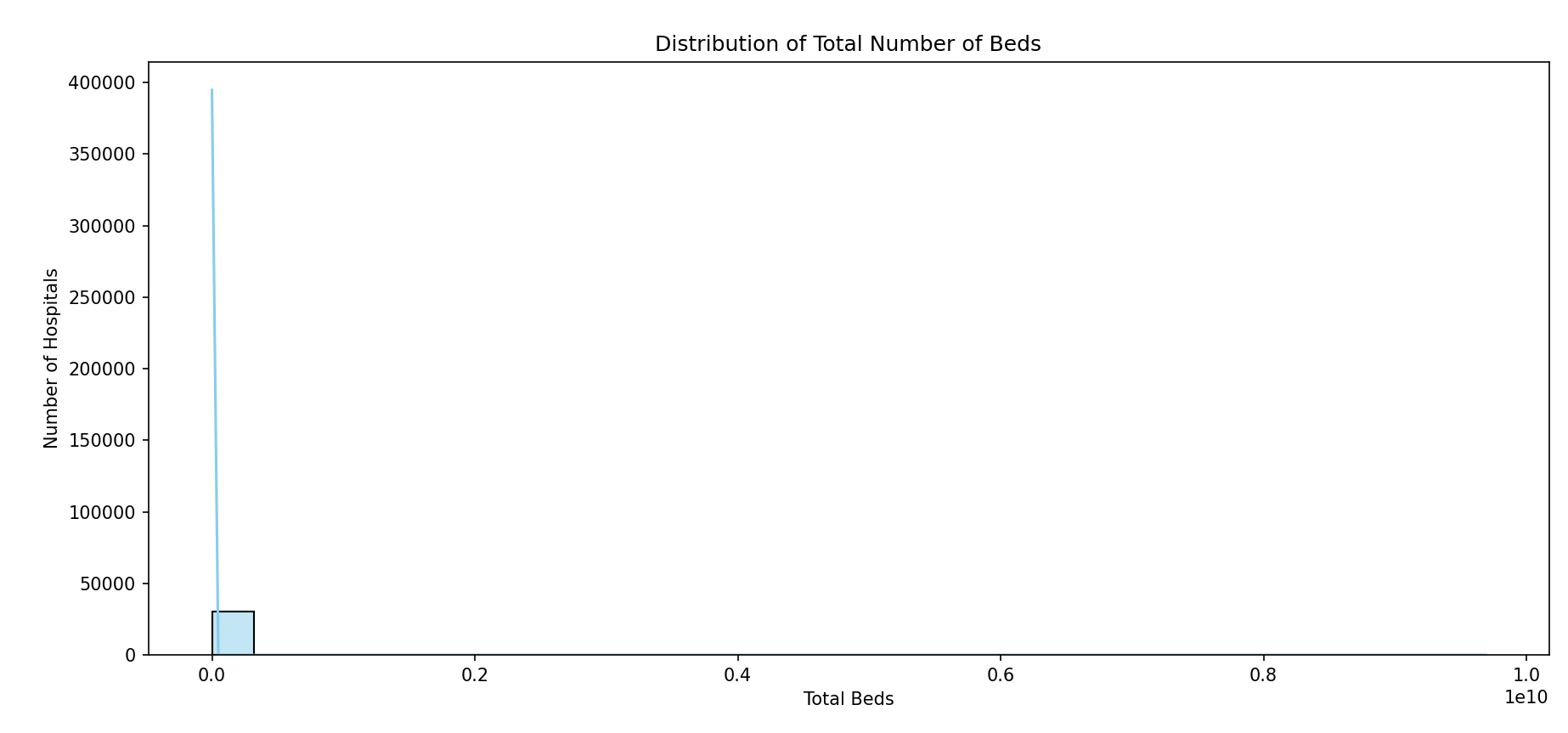
# 4.4. Summary Statistics of Hospital Beds

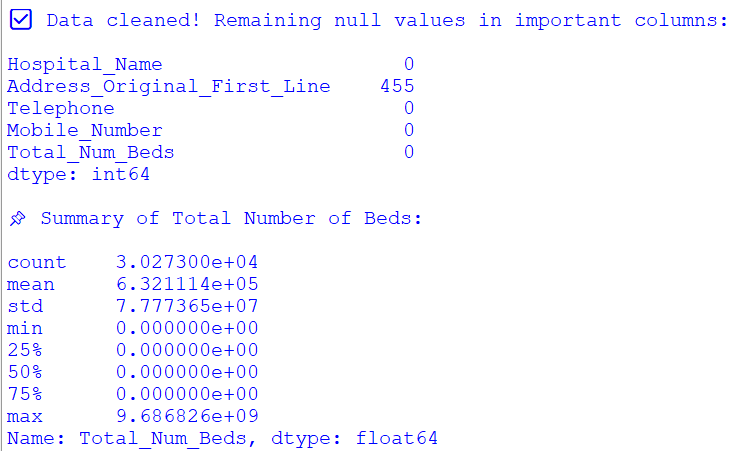
Introduction  
Analyzes hospital capacity via number of beds.

Methods  
- Descriptive statistics (describe())  
- Histogram of total bed distribution  
- Top 10 states by total beds

Results  
Wide variation exists; few hospitals have high bed counts, most are smaller.

Visualization  
- Histogram of bed counts  
- Bar chart of top 10 states by total beds

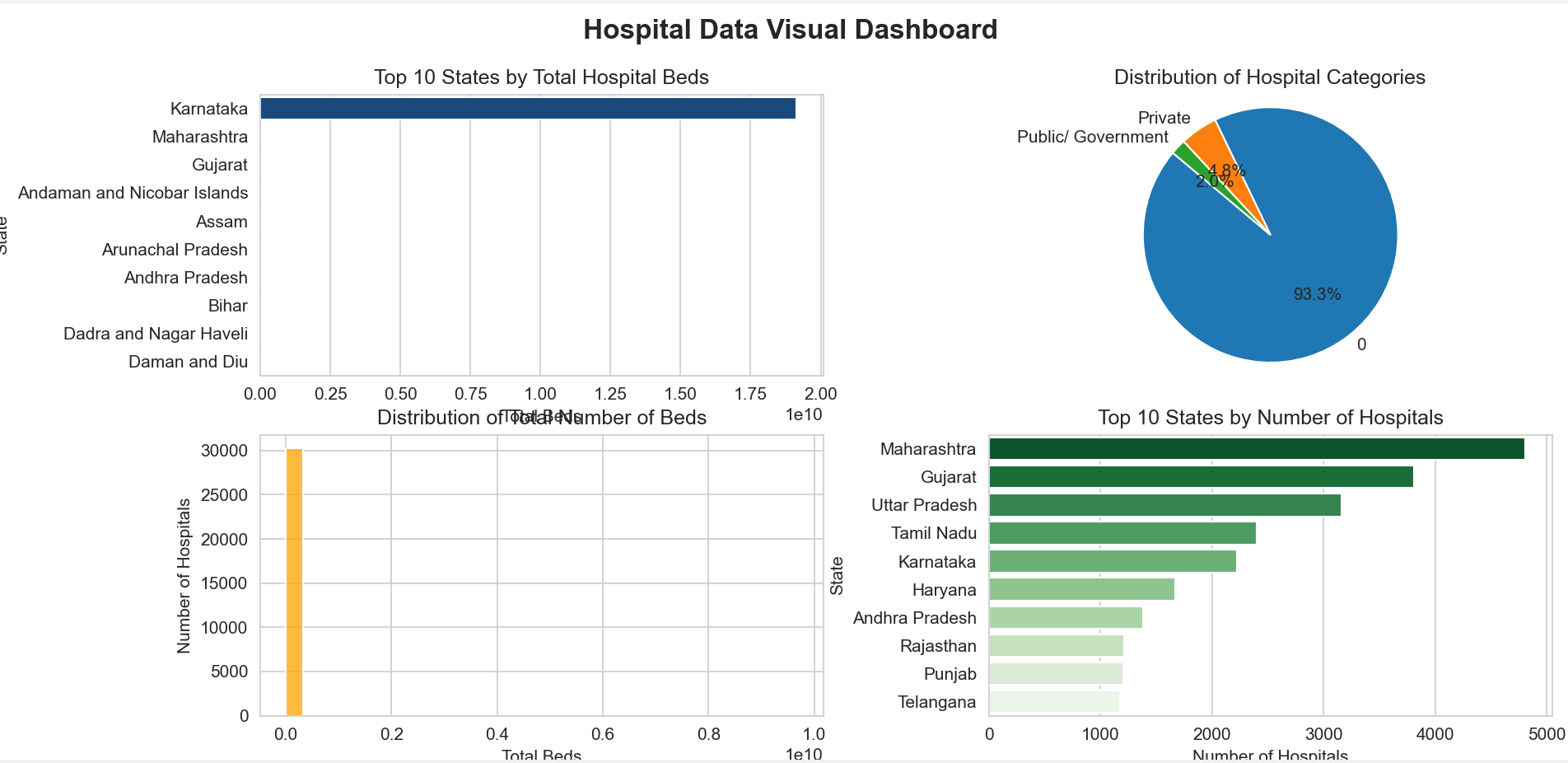




# 4.5. Dashboard - Visual Summary

Description  
Created a 2x2 dashboard plot showcasing the summary of major insights.

Charts Included  
- Top 10 States by Beds  
- Pie chart of Hospital Categories  
- Histogram of Beds  
- States by Number of Hospitals



# 5. Conclusion

The project successfully explored hospital data across India using Python. It highlighted disparities in healthcare infrastructure, revealed missing data issues, and visualized key trends. These insights can support better decision-making in healthcare planning.

# 6. Future scope

- Build interactive dashboards using Power BI or Tableau  
- Incorporate more datasets like patient inflow, mortality, and specialties  
- Perform geospatial analysis using geopandas  
- Deploy the analysis on web applications using Flask/Streamlit

# 7. References

1. Government of India Hospital Directory Dataset, data.gov.in  
2. Matplotlib Documentation, https://matplotlib.org  
3. Seaborn Documentation, https://seaborn.pydata.org  
4. IEEE Reference Style Guide, IEEE.org