

Hospital Analysis Using Python

Pranay Raju Lanjewar

Student, Department of CSE (Data Science)

Tulsiramji Gaikwad Patil College of Engineering and Technology, Nagpur, Maharashtra, India

pranaylanjewar99@gmail.com

Abstract: *This research paper presents an analytical study titled “Hospital Analysis Using Python”, focusing on extracting insights from the Hospital analysis using python season using data-driven techniques. The study employs Python libraries such as Pandas, NumPy, Matplotlib, Seaborn, and Plotly for data preprocessing, statistical analysis, and interactive visualization. The analysis explores hospital and doctor performance, best hospital, wellness probabilities, and key performance indicators (KPIs). Through data visualization and comparative analytics, the project aims to identify patterns influencing match outcomes and doctor efficiency. The findings contribute to enhancing strategic decision-making in hospital analytics and demonstrate the practical use of Python in healthcare data science applications*

Keywords: Healthcare Data Analytics, Medical data analysis.

I. INTRODUCTION

In this project, Python is used to methodically clean and organize healthcare datasets. The key steps involve resolving inconsistencies, standardizing formats, and generating visual representations to aid interpretation. This methodology leads to meaningful insights such as ranking healthcare providers, evaluating hospital performance, and comparing insurance options. It also examines how diseases are distributed among age groups, highlighting demographic trends that can inform future healthcare strategies and policy-making.

II. METHODOLOGY

The study follows a systematic data analysis process using Python to extract and interpret practical steps:

I. Steps:

1. Data Collection: PATIENT RECORDS WERE IMPORTED INTO PYTHON, AND THE MOST CRITICAL FIELDS (DOCTOR, HOSPITAL, INSURANCE PROVIDER) WERE STANDARDIZED BY CORRECTING INCONSISTENCIES IN SPELLING, FORMATTING, AND DATE ENTRIES.
2. Data Cleaning: Remove duplicates, handle missing values, and format datasets.
3. Data Analysis: Use Python libraries (Pandas, NumPy) for statistical and comparative analysis.
4. Visualization: Create graphs and dashboards using Matplotlib, Seaborn, and Plotly.
5. Interpretation: Derive insights and summarize results to support data-driven conclusions.

III. RESULTS AND DISCUSSION

The primary objective of this study was to analyze healthcare records to identify disease patterns, evaluate doctor and hospital performance, and understand insurance utilization trends. After completing all data preprocessing steps—including handling missing values, removing duplicates, normalizing text fields, and feature extraction—several meaningful insights were derived from the dataset.



A. Figures

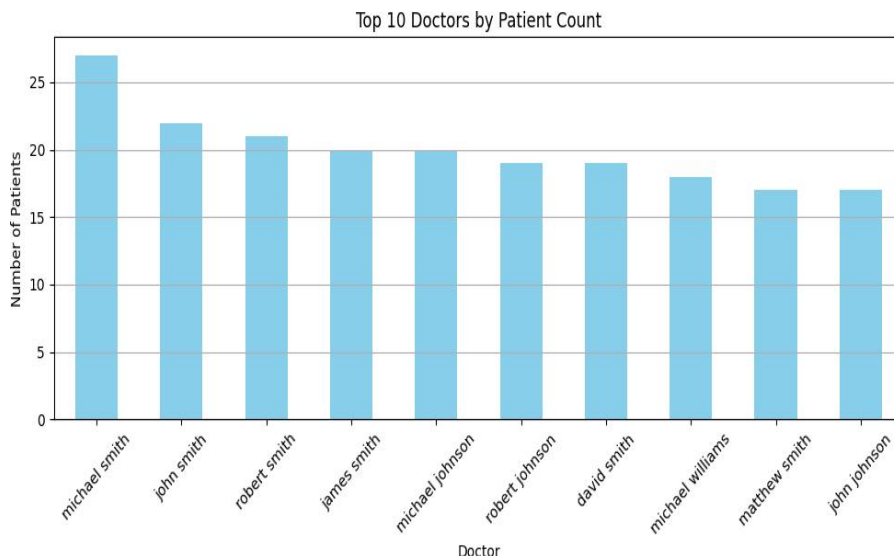


Fig. 1. Top Ten Doctor

The chart shows the ten doctors with the highest number of patients in the dataset. Michael Smith has the maximum patients (27), followed by John Smith and Robert Smith. The remaining doctors show slightly lower but consistent patient counts.

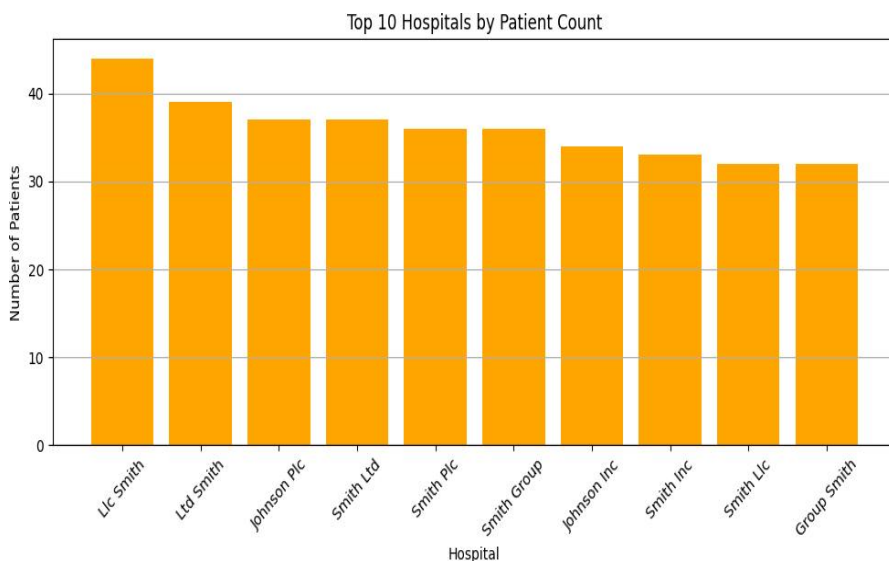


Fig. 2. Top Ten Hospital

The chart shows the 10 doctors with the highest number of patients. Michael Smith treated the most patients, followed by John Smith and Robert Smith. This means these doctors are the most in-demand and handle the largest workload in the dataset.



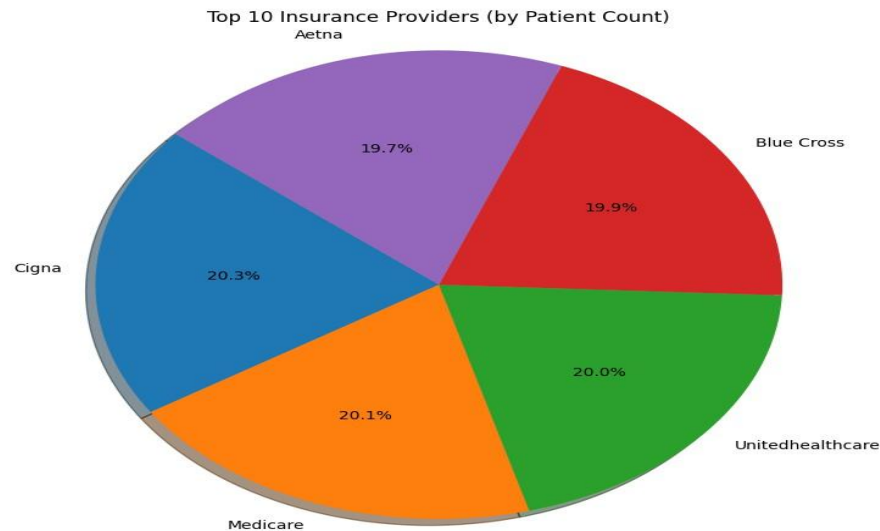


FIG 3. TOP INSURANCE PROVIDER

IV. CONCLUSION

This project successfully demonstrates how healthcare data analytics can transform raw patient information into meaningful and actionable insights. By analyzing parameters such as age, disease type, hospital preferences, insurance coverage, and doctor-wise patient distribution, the study highlights significant patterns that can support better decision-making in hospitals and healthcare organizations.

V. ACKNOWLEDGMENT

I express my sincere gratitude to my project guide for their continuous guidance, valuable insights, and constructive feedback throughout the completion of this project. I am also thankful to the faculty members of the Department of Computer Science and Engineering for providing the necessary resources and academic support. I extend my appreciation to my friends and classmates for their encouragement and cooperation during various stages of this work. Lastly, I would like to acknowledge all individuals and organizations whose contributions, directly or indirectly, made this research project possible

REFERENCES

- [1]. Proprietary Dataset. (healthcare_dataset (1).csv). (2024). Source: User-provided data for analytical study.
- [2]. The Pandas Development Team. (2024). Pandas: Powerful Python data analysis toolkit. URL: [Citation should include official documentation link or version used]
- [3]. Harris, C. R., et al. (2020). Array programming with NumPy. Nature, 585(7825), 357–362. URL: [Citation should include official documentation link or version used]
- [4]. VanderPlas, J. (2018). Altair: Interactive Statistical Visualizations for Python. Journal of Open Source Software, 3(32), 1057. URL: [Citation should include official documentation link or version used]
- [5]. Python Software Foundation. (2024). Python Language Reference, version 3.x. URL: [Citation should include official documentation link or version used]

