

Name	Varshil Shah
UID	2022301013
Branch	Final Year Btech Computer
Batch	ADV Batch D

**Dataset:** <https://www.kaggle.com/datasets/prasad22/healthcare-dataset>

## Theory -

Creating interactive dashboards with advanced and basic charts requires selecting the right tools and visualizations based on the data and the insights you want to present.

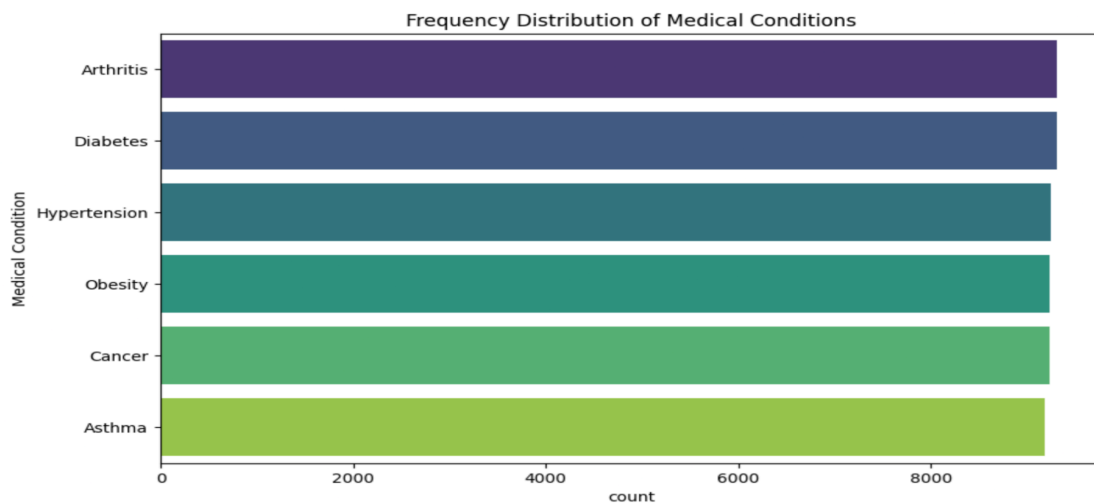
This dataset consists of 10,000 synthetic healthcare records, each simulating real-world patient information. It is structured with various attributes, including:

- **Patient Demographics:** Details like age, gender, and possibly geographic or socioeconomic indicators.
- **Medical Conditions:** Information on diagnoses such as heart disease, diabetes, or other health concerns.
- **Admission Details:** Information related to hospital visits, admission and discharge dates, and procedures performed.
- **Other Attributes:** Could include vital signs (e.g., cholesterol, blood pressure), lifestyle factors, and treatment outcomes.

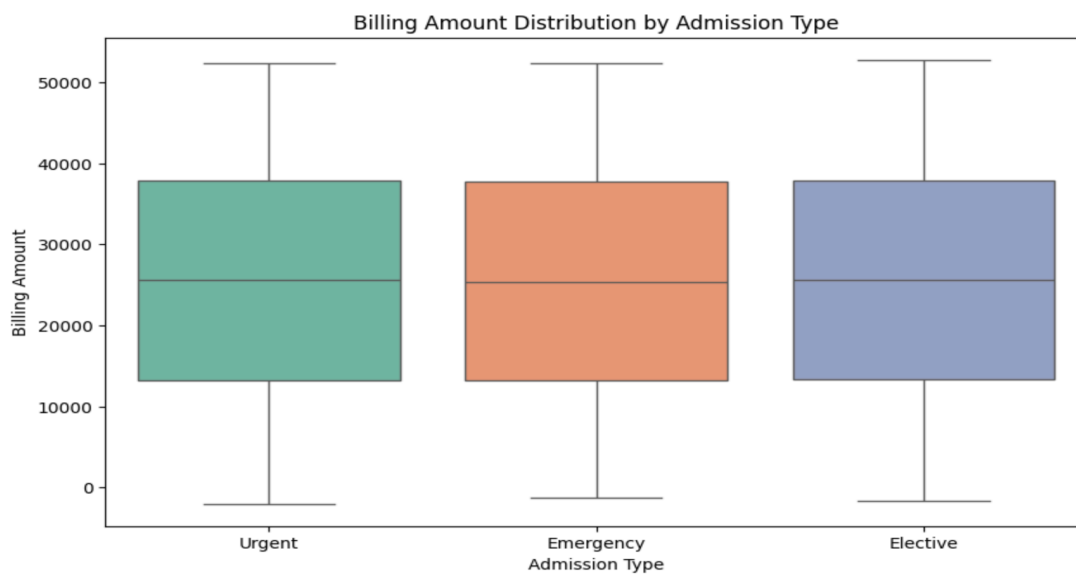
The dataset is entirely synthetic, meaning it does not involve real patients and is designed for educational and research purposes, making it safe to use for learning, testing models, and creating visualizations without privacy concerns.

## Advanced Charts

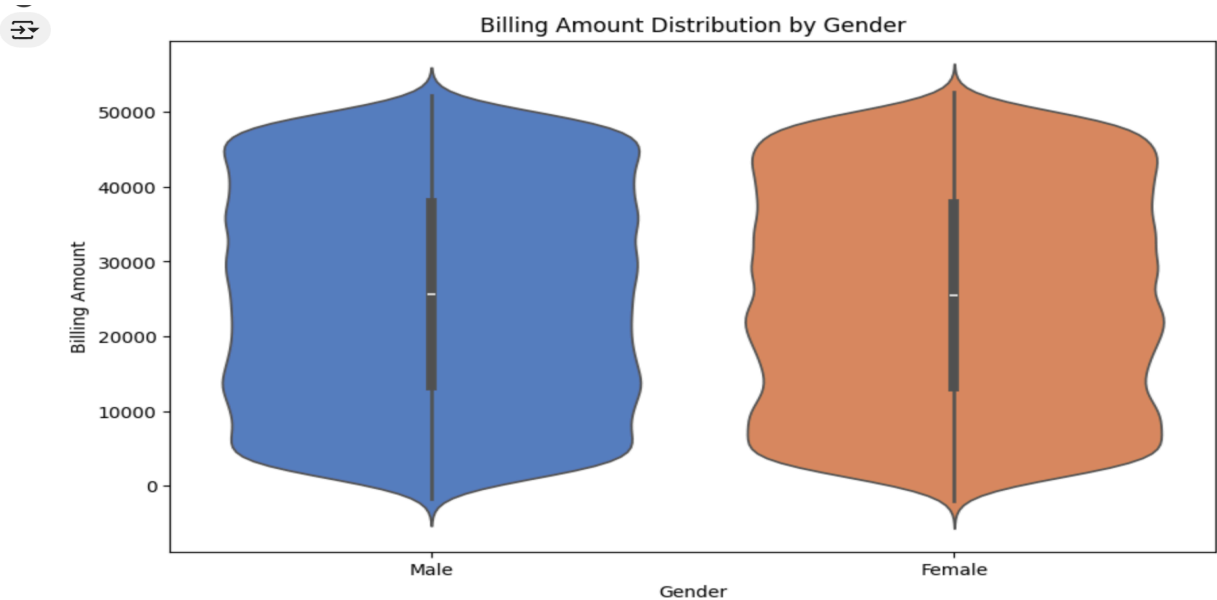
1. **Word Chart:** Displays the frequency of words from text data where larger words represent higher frequency. Useful for text-heavy datasets like patient feedback.



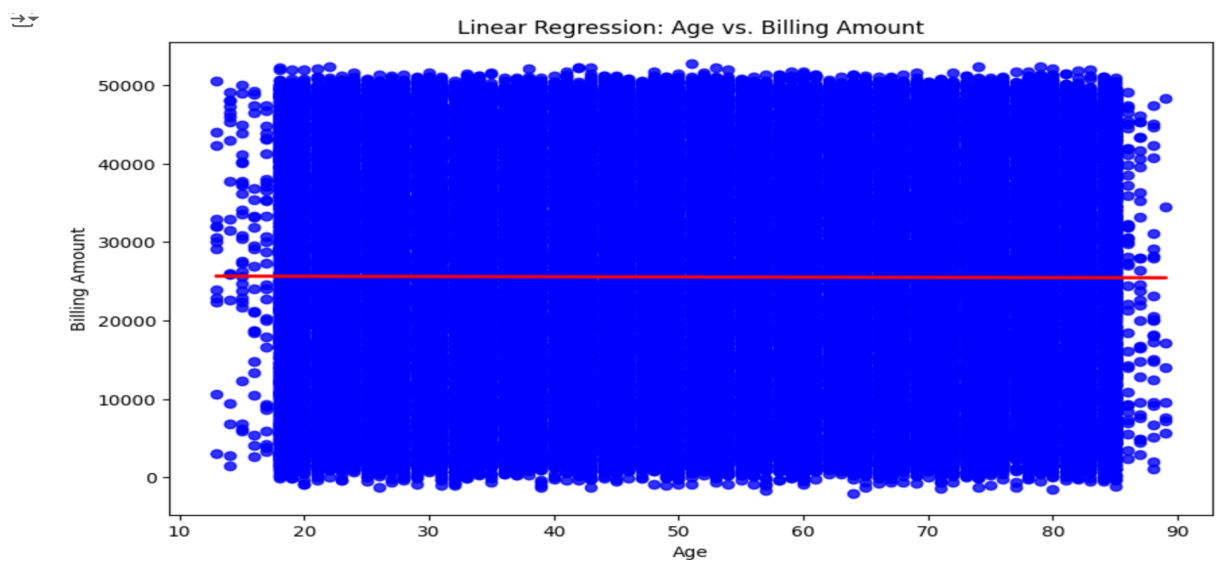
2. **Box and Whisker Plot:** Visualizes the distribution of data by showing the median, quartiles, and outliers. Ideal for comparing variables like cholesterol levels among different age groups.



3. **Violin Plot:** Combines a box plot with a density plot, showing the distribution of the data across different groups. It's useful for comparing the distribution of heart disease cases across genders.

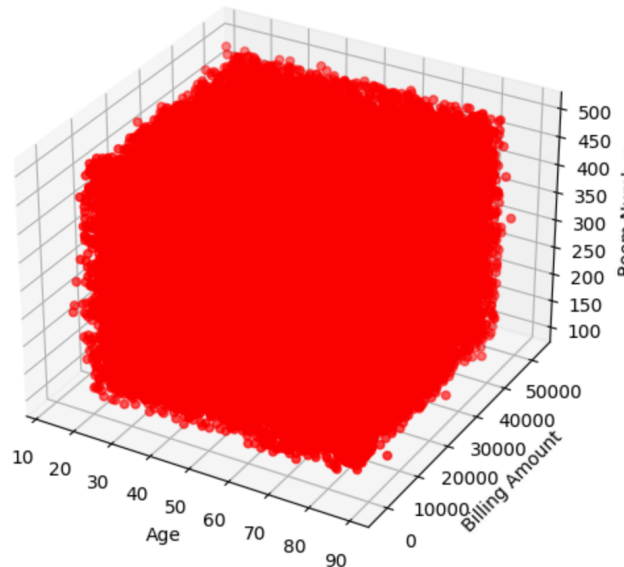


4. **Regression Plot (Linear and Nonlinear):** Shows the relationship between two continuous variables and includes a regression line (linear or nonlinear). For example, the relationship between age and cholesterol levels.

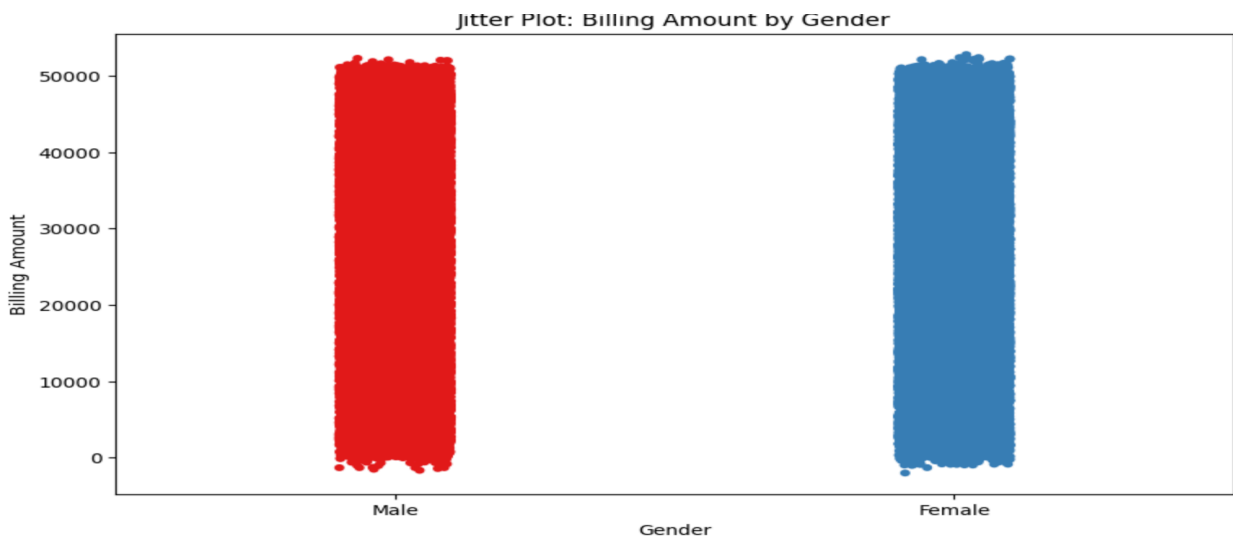


5. **3D Chart:** Adds a third dimension to traditional charts, which can help visualize multivariable relationships. It might show age, cholesterol, and heart disease risk simultaneously.

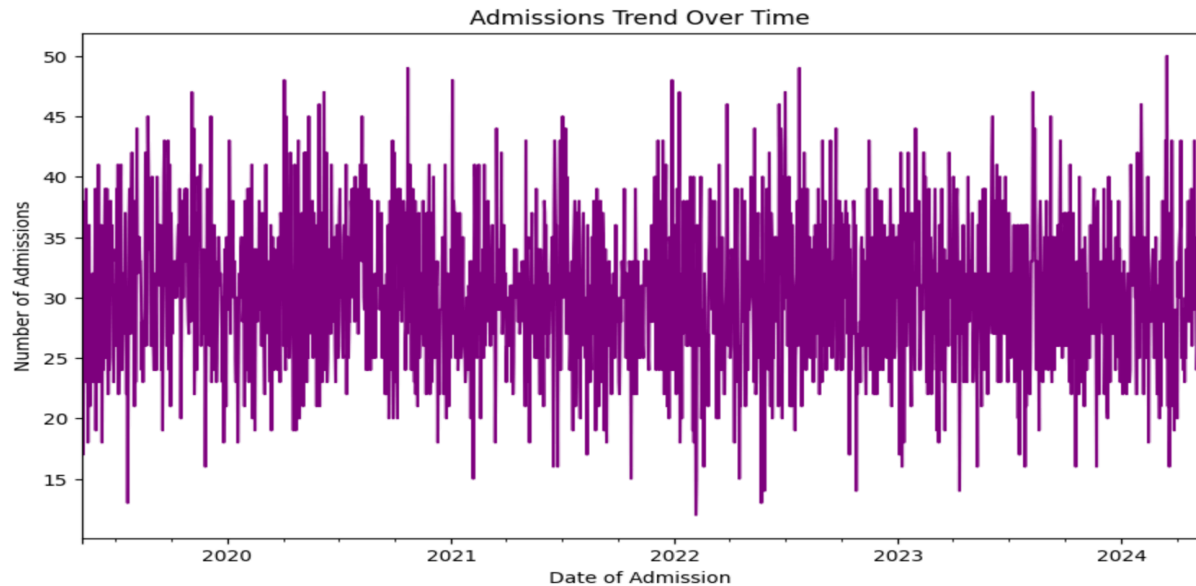
3D Scatter Plot: Age, Billing Amount, and Room Number



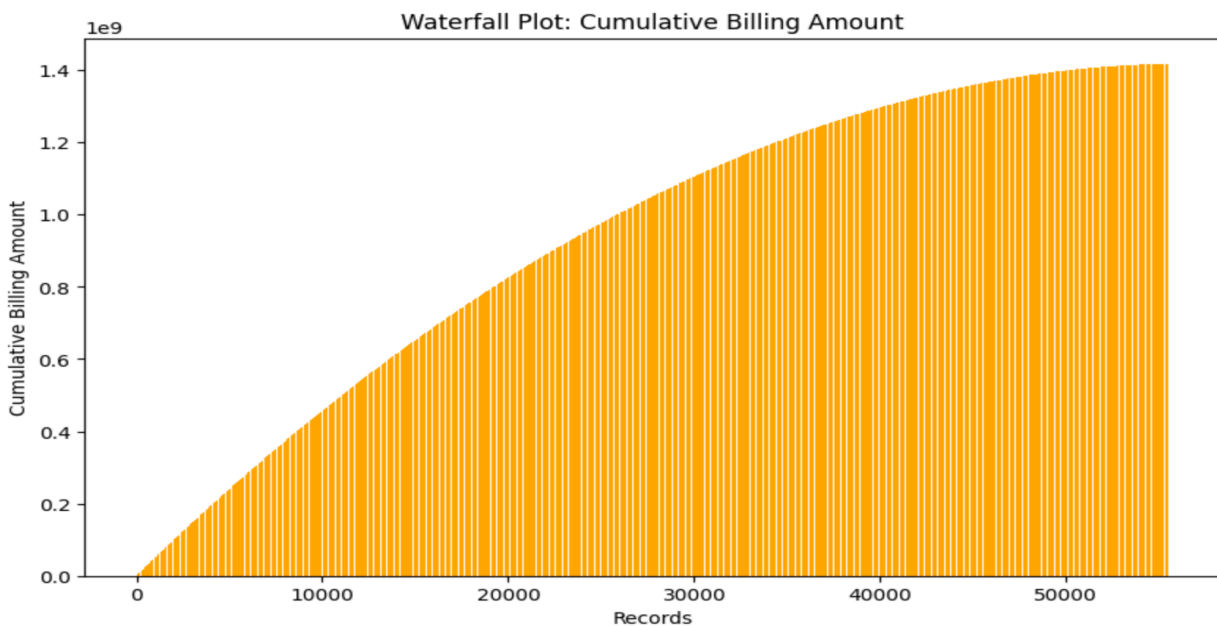
6. **Jitter Plot:** Adds random noise to points in a scatter plot to avoid overplotting when data points overlap. Useful when visualizing large datasets with many overlapping points, such as age vs. heart disease.



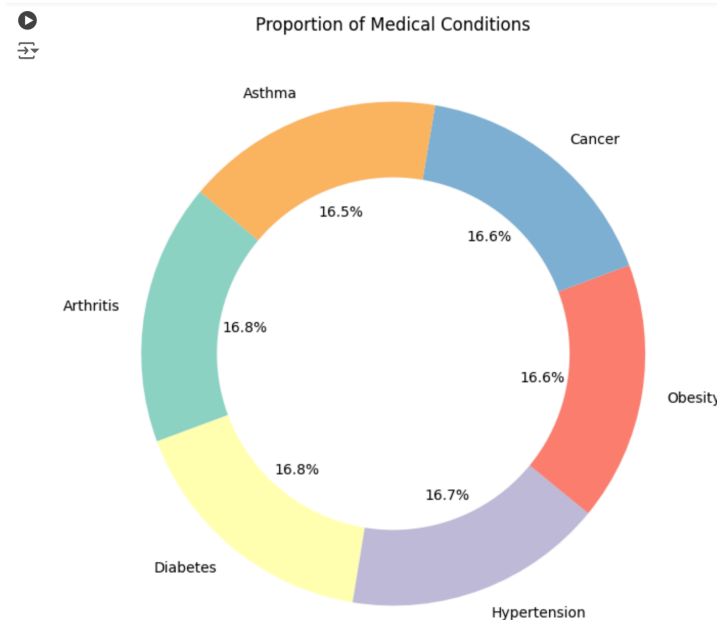
7. **Line Chart:** Displays trends over time by connecting data points with a continuous line. It can show trends like heart disease cases over different months or years.



8. **Waterfall Chart:** Illustrates sequential changes in a measure, such as the accumulation of heart disease risk factors.



9. **Donut Chart:** A variation of a pie chart that shows part-to-whole relationships but with a blank center. It can show proportions, like the percentage of different types of heart disease



## Conclusion -

In conclusion, the synthetic healthcare dataset with 10,000 records provides a valuable resource for analyzing patient demographics, medical conditions, and healthcare trends. By using various advanced and basic visualization techniques such as bar charts, scatter plots, regression analysis, and more, key insights can be drawn about disease spread, risk factors, and patient outcomes. The synthetic nature of the data allows for exploration and experimentation in a safe, non-commercial environment, making it ideal for educational purposes and gaining a deeper understanding of healthcare data analysis.