

# Data Visualization

## Assignment

Name :- D. Jaya Surya

VTU No :- VTU21708

Subject :- Data Visualization

Slot No :- S<sub>2</sub>h5

Department :- CSE(AI ML)

Course code :- 10212CS214

## Healthcare - Patient Data Analysis

1. Explain data abstraction and identify data types

(Categorical: gender, Continuous: age, cost).

A. Data abstraction

→ Data abstraction is the process of simplifying complex raw hospital data into a structured and meaningful format suitable for analysis.

→ In the healthcare context, patient data may include hundreds of attributes such as personal details, medical history, diagnosis & hospital.

By abstracting, we select only the attributes necessary for our analysis:-

→ Patient - ID

→ Gender

→ Age

→ Treatment - Cost

→ Diagnosis

→ Doctor - ID

→ Referral - ID

→ City

→ This simplification helps in performing analytical operations efficiently & reduces redundancy.

2. Apply EDA (Histogram for age, boxplot for treatment cost).

A. EDA helps understand data patterns, detect outliers and summarize characteristics using visual tools such as histograms and box plots.

a) Histogram for Age

```
import Pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv("Patient_data.csv")
plt.hist(df["Age"], bins=10, color='skyblue',
          edgecolor='black')
plt.title("Distribution of Patient Ages")
plt.xlabel("Age")
plt.ylabel("Number of Patients")
plt.show()
```

2. ~~Box~~ Boxplot for Treatment Cost.

```
plt.boxplot(df["Treatment-cost"])
plt.title("Boxplot of Treatment Cost")
plt.ylabel("Cost (₹)")
plt.show()
```

3. Use Network Visualization for doctor-Patient referrals & text Visualization for diagnosis notes.

A. Doctor - Patient Referral Network

```
import network as nx
```

```
import matplotlib.pyplot as plt
```

```
G = nx.from_pandas_edgelist(df, 'Doctor_ID', 'Referral_ID')
```

```
plt.figure(figsize=(7,7))
```

```
nx.draw(G, with_labels=True, node_color='lightgreen',  
        node_size=800, edge_color='gray')
```

```
plt.title("Doctor - Patient Referral Network")
```

```
plt.show()
```

b) Text - Visualization for Diagnosis Notes.

```
from wordcloud import WordCloud
```

```
text = " ".join(df["Diagnosis"].astype(str))
```

```
wc = WordCloud(width=800, height=400, background-  
               color='white')
```

```
plt.imshow(wc, interpolation='bilinear')
```

```
plt.axis('off')
```

```
plt.title("Word Cloud of Diagnosis Notes")
```

```
plt.show()
```



4. Map patient address on a City - level heatmap.

A. City-level heatmap

```
import folium
import pandas as pd

df = pd.DataFrame({
    'City': ['Chennai', 'Madurai', 'Gimbatore', 'Salem'],
    'Count': [50, 30, 40, 20],
    'Latitude': [13.0827, 9.9252, 11.0168, 11.6643],
    'Longitude': [80.2707, 78.2198, 76.9558, 78.1460]
})

m = folium.Map(location = [11.0, 78.0], zoom_start=7)

for _, row in df.iterrows():
    folium.CircleMarker(
        location = [row['Latitude'], row['Longitude']],
        radius = row['Count']/2,
        color = 'red',
        fill = True,
        fill_opacity = 0.5,
        popup = f'{row["City"]}: {row["Count"]} Patients'
    ).add_to(m)

m.save("Patient-heatmap.html")
```

5. Create a hospital Performance dashboard.

A. Key Components.

Metric	Purpose
Total Patients	Shows patient Volume
Average Treatment Cost	Tracks hospital Pricing trends Helps in resource planning
Most Common Disease	Indicates demand distribution
City with Most Patients	Shows doctor Collaboration Patterns
Referral Network	

import plotly.express as px

fig = px.bar(df, x='city', y='Treatment\_Cost',  
color='City',  
title='Average Treatment Cost by City')

fig.show()

→ Decision-makers can track Patient inflow, Cost Variations & disease Patterns.

→ The System Supports better planning & Improves healthcare efficiency.