

DATA VISUALIZATION

Course code:- 10212C8214

Siva

V. Shiva Siri
VTU 21754
B.TECH, CSE
Task-1
Slot-S2

Healthcare - Patient Data Analytics

Explain

Data abstraction and identify data types

Data Abstraction and Data Types

a) Meaning of Data Abstraction

- Data abstraction refers to simplifying complex real-world data into a structured and analyzable form.
- In healthcare, patient records are abstracted into a dataset containing only relevant information such as age, gender, cost, diagnosis, etc.
- It helps in data organization, analysis, and decision-making.

b. Common Healthcare Data Attributes

Attribute	Description	Data Type	Example
Patient-ID	Unique identifier for each patient	Nominal	P101
Gender	Patient's gender	Categorical	Male/Female
Age	Patient's age	Continuous (Numerical)	45
Diagnosis	Disease name	Categorical (Text)	Diabetes
Treatment-Cost	Cost of treatment	Continuous (Numerical)	₹25000
Doctor-ID	Treating doctor's ID	Nominal	D202
Referral-ID	Doctor referred to	Nominal	D305
City	Patient's location	Categorical	Hyderabad

c) Data type Summary :-

- Categorical: Qualitative data such as gender, city, diagnosis.
- Continuous: Quantitative data such as age, treatment cost.
- Nominal: Labels or names without numerical meaning (patient-ID, Doctor-ID)

Q. Apply EDA (histogram for age, boxplot for treatment cost)

a) Purpose:

- To explore and summarize key characteristics of data
- Detect patterns, trends and outliers
- Helps prepare data for deeper modeling

b) Visualization 1 - Histogram for Age

Tool: python

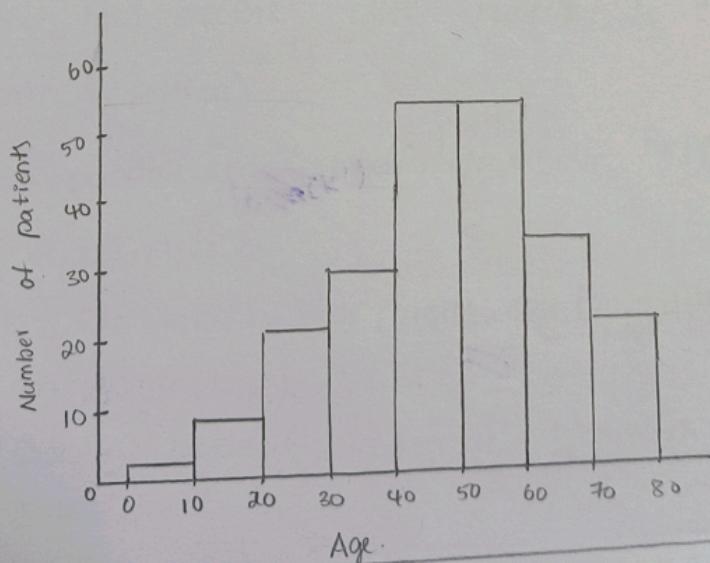
```
plt.hist(df['Age'], bins=10, edgecolor='black')
```

```
plt.title('Distribution of patient Age')
```

```
plt.xlabel('Age')
```

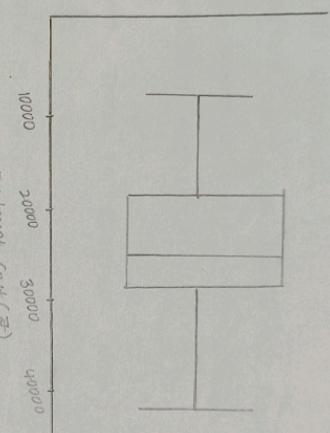
```
plt.ylabel('Number of patients')
```

```
plt.show()
```



c) Visualization 2 - Box plot for Treatment Cost

```
Tool :- Seaborn  
sns.boxplot(x=df['Treatment_Cost'])  
plt.title('Boxplot of Treatment Cost')  
plt.xlabel('Treatment Cost')  
plt.show()
```



3. Use network visualization for doctor-patient referrals and text visualization for diagnosis notes

a) Doctor - Patient Referral Network

Purpose:-

To visualize how doctors refer patients to other specialists

Tool:- NetworkX (Python)

```
G=nx.from_pandas_edgelist(df,'Doctor_ID','Referral_To',create_=True)  
nx.draw(G,with_labels=True,node_color='lightblue',node_size=1500,arrowstyle='->')  
plt.title('Doctor - patient Referral Network')  
plt.show()
```

④ Text visualization - Diagnosis Notes

a) purpose:

To find common diseases or symptoms in patient diagnosis text.

b) tool: word cloud

```
from wordcloud import WordCloud  
text = " ".join(df['Diagnosis'].T)  
wordcloud = WordCloud(width=800, height=400, background_color='white').generate(text)
```

```
plt.imshow(wordcloud, interpolation='bilinear')  
plt.title('Common terms in Diagnosis Notes')  
plt.show()
```

4. Geographical Mapping - City-level Heatmap

a) purpose:

- To visualize where most patients are located geographically.
- Help plan medical resources and outreach programs.

b) tool: folium (python)

```
import folium  
from folium.plugins import HeatMap  
m = folium.Map(location=[17.3850, -8.4867], zoom_start=6)  
HeatMap(df['Latitude'], 'Longitude').add_to(m)  
m.save('patient-heatmap.html')
```

c) Interpretation :-

- Red areas = High patient concentration
- Blue areas = Low patient concentration
- Useful for locating high - demand zones and planning hospital branches or mobile units.

5. Create a hospital performance dashboard

a) purpose:-

- To monitor and evaluate hospital performance visually.
- Combines multiple metrics into one interactive view.
- Tools: Tableau, PowerBI, or Plotly Dash.

b) Key components:-

Metric	Visualization	Insight
Number of patients per Department	Bar chart	Identifies busiest departments
Average Treatment Cost of Disease	Box plot / Bar chart	Cost control & Efficiency
Patient Age Distribution	Map	Regional patient density
Doctor Referral Network	Network Graph	Doctor collaboration pattern
Top performing Doctors	Table or Rank Chart	Distinguish best performers

↳ Example - Boxchart

```
import plotly.express as px  
fig = px.box(df, x='Department', y='Patient_Count', title='Patients  
per Department')  
fig.show()
```