Sheth L.U.J College of Arts & Sir M.V. College of Science and Commerce
Data Science
PRACTICAL NO. 10

Aim: Data Visualization and Storytelling

a) Create meaningful visualizations using data visualization tools.
b) Combine multiple visualizations to tell a compelling data story.
c) Present the findings and insights in a clear and concise manner.

CODE:

- ***Importing libraries***

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

- ***Load Dataset***

```
df = pd.read_csv("healthcare-dataset-stroke-data.csv")
df.head()
```

- ***Dataset Description***

```
df.info()
df.describe(include="all")
```

- ***Handle Missing Values***

```
df['bmi'] = pd.to_numeric(df['bmi'], errors='coerce')
df['bmi'].fillna(df['bmi'].mean(), inplace=True)
df['smoking_status'].replace("Unknown", np.nan, inplace=True)
df['smoking_status'].fillna(df['smoking_status'].mode()[0], inplace=True)
df.isnull().sum()
```

- ***Bar Plot – Stroke Count by Gender***

```
plt.figure(figsize=(6,4))
sns.countplot(data=df, x='gender', hue='stroke')
plt.title("Stroke Count by Gender")
plt.show()
```

- ***Boxplot – Age vs Stroke***

```
plt.figure(figsize=(6,4))
sns.boxplot(data=df, x='stroke', y='age')
plt.title("Age Distribution for Stroke vs Non-Stroke Patients")
plt.show()
```

- ***Line Plot – Average Glucose Level by Age***

```
df_sorted = df.sort_values("age")
plt.figure(figsize=(8,4))
plt.plot(df_sorted['age'], df_sorted['avg_glucose_level'])
plt.title("Glucose Level Trend Across Age")
plt.xlabel("Age")
plt.ylabel("Avg Glucose Level")
plt.show()
```

- ***Heatmap – Correlation Between Numerical Variables***

```
plt.figure(figsize=(8,6))
numeric_df = df.select_dtypes(include=['int64','float64'])
sns.heatmap(numeric_df.corr(), annot=True, cmap="coolwarm")
plt.title("Correlation Heatmap")
plt.show()
```

- ***Violin Plot – BMI vs Work Type***

```
plt.figure(figsize=(8,5))
sns.violinplot(data=df, x='work_type', y='bmi')
plt.xticks(rotation=45)
plt.title("BMI Distribution Across Work Types")
plt.show()
```

- ***Pair Plot – Numerical Variables***

```
sns.pairplot(df[['age','avg_glucose_level','bmi','stroke']], hue='stroke')
plt.show()
```

- ***Scatter Plot – Glucose vs BMI***

```
plt.figure(figsize=(6,4))
sns.scatterplot(data=df, x='avg_glucose_level', y='bmi', hue='stroke')
plt.title("Glucose Level vs BMI")
plt.show()
```

- ***Distribution Plot – Age***

```
plt.figure(figsize=(6,4))
sns.histplot(df['age'], kde=True)
plt.title("Age Distribution")
plt.show()
```

- ***Pie Chart – Work Type Distribution***

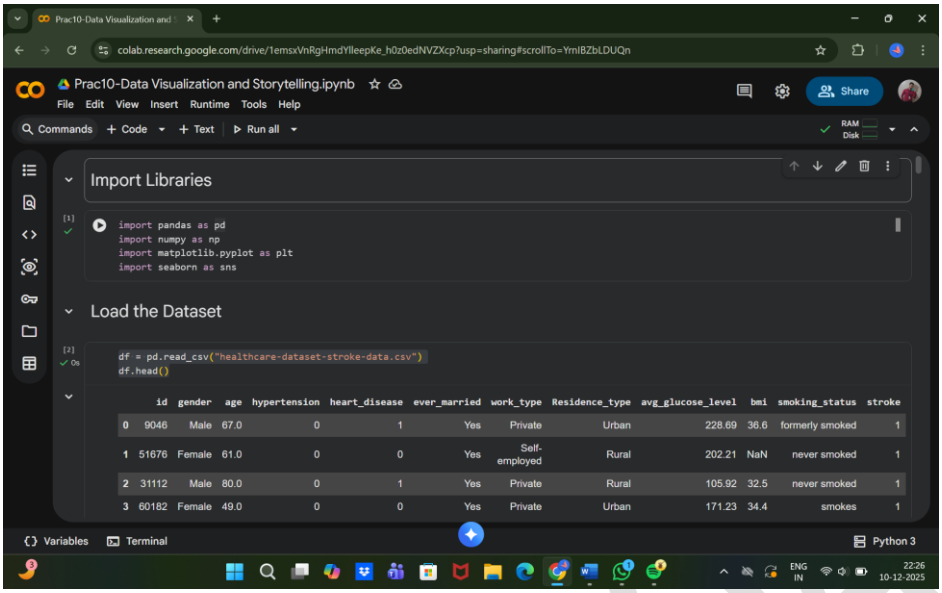
```
plt.figure(figsize=(6,6))
df['work_type'].value_counts().plot.pie(autopct='%1.1f%%')
plt.title("Work Type Distribution")
plt.show()
```

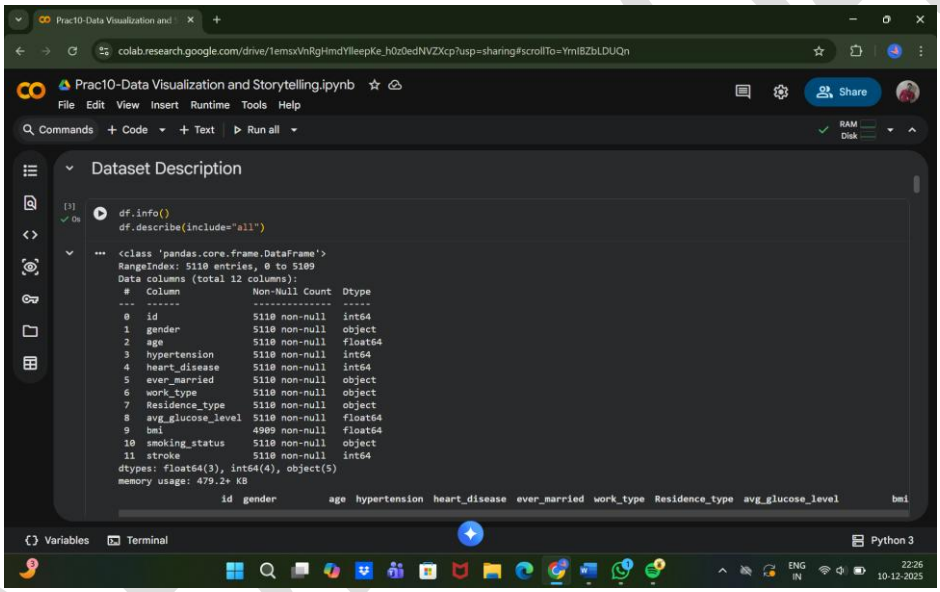
- ***Histogram – Glucose Level***

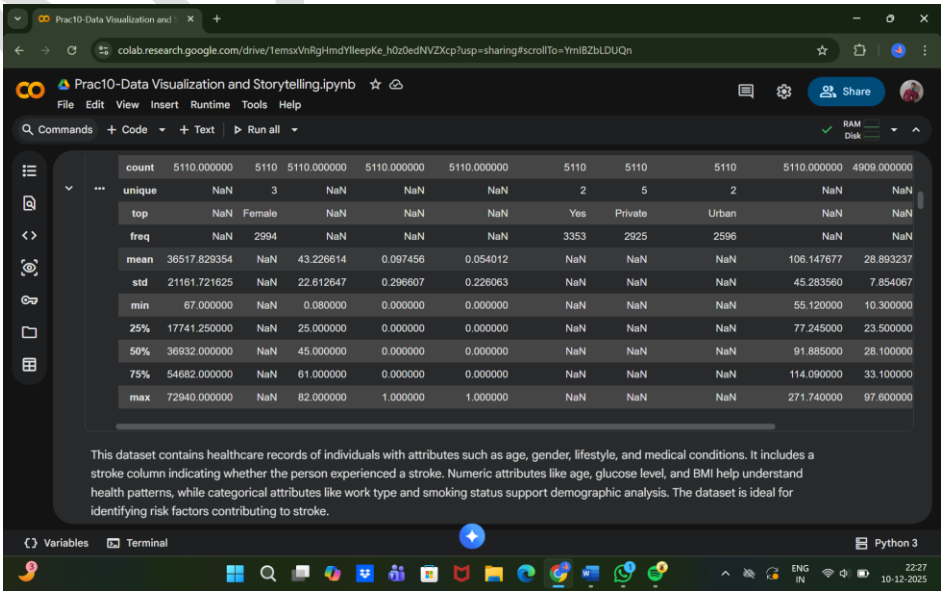
```
plt.figure(figsize=(6,4))
plt.hist(df['avg_glucose_level'], bins=20)
plt.title("Histogram of Glucose Levels")
plt.xlabel("Glucose Level")
plt.ylabel("Frequency")
plt.show()
```

Output:

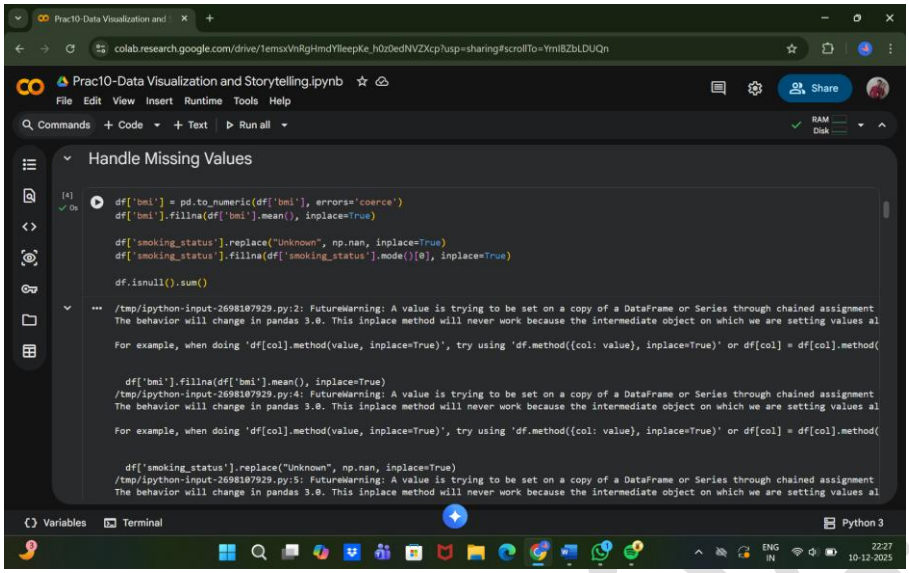
Sheth L.U.J College of Arts & Sir M.V. College of Science and Commerce
Data Science
PRACTICAL NO. 10

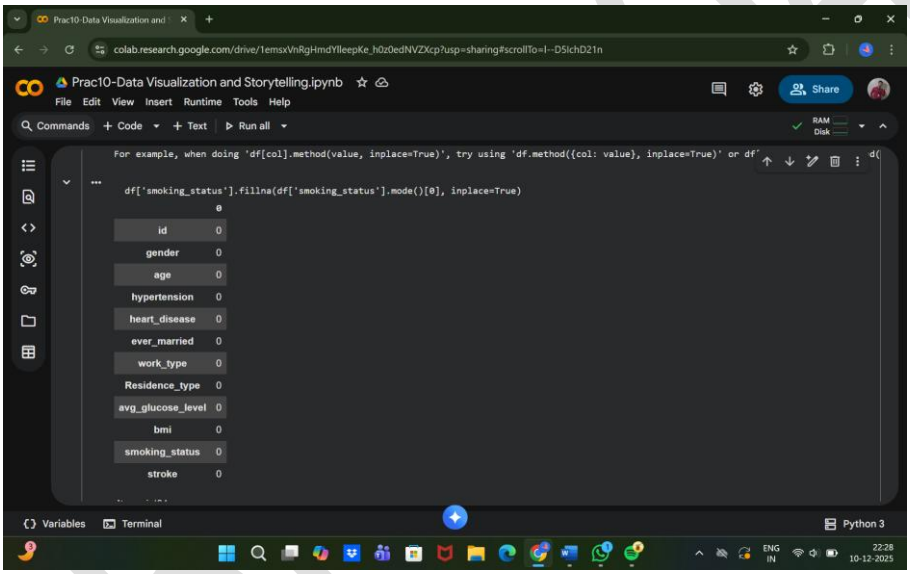


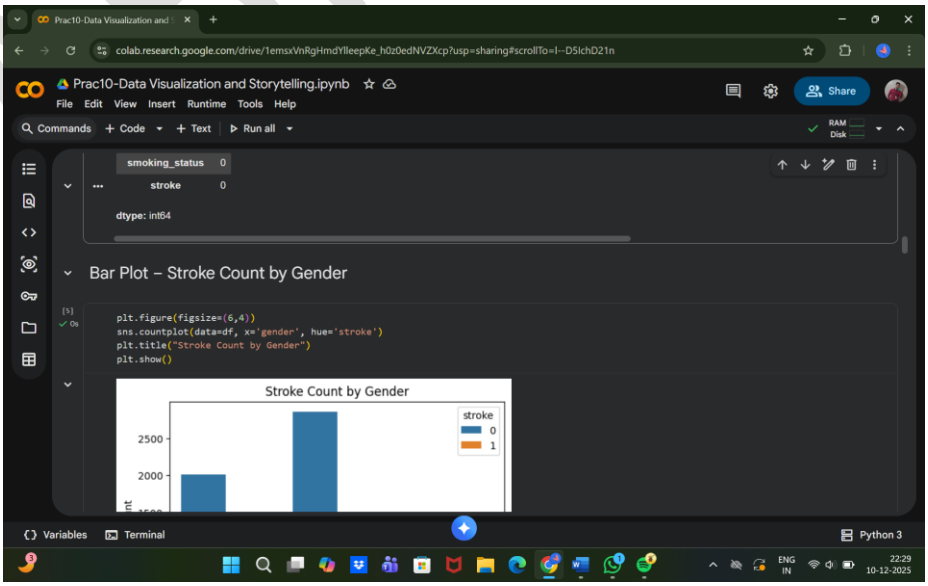




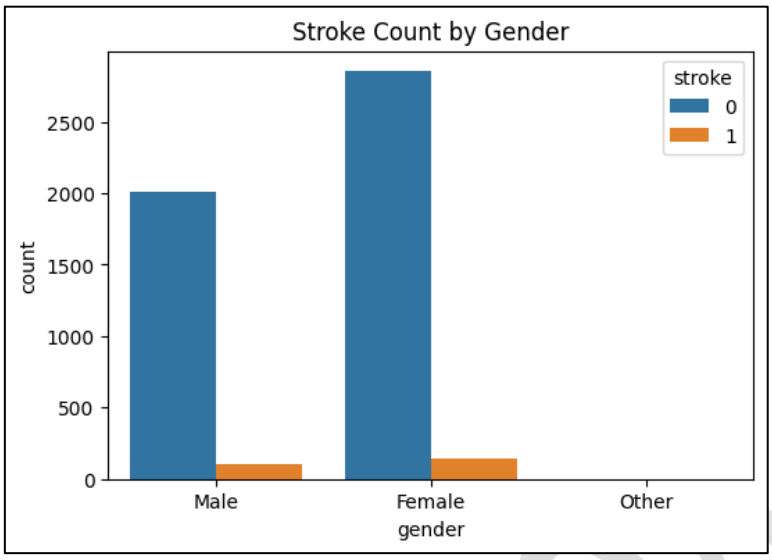
Sheth L.U.J College of Arts & Sir M.V. College of Science and Commerce
Data Science
PRACTICAL NO. 10

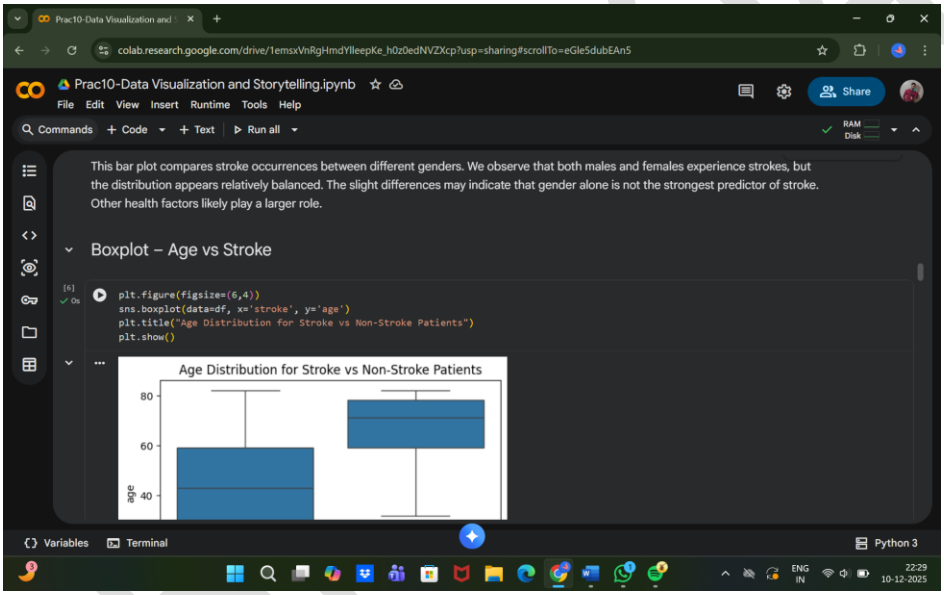


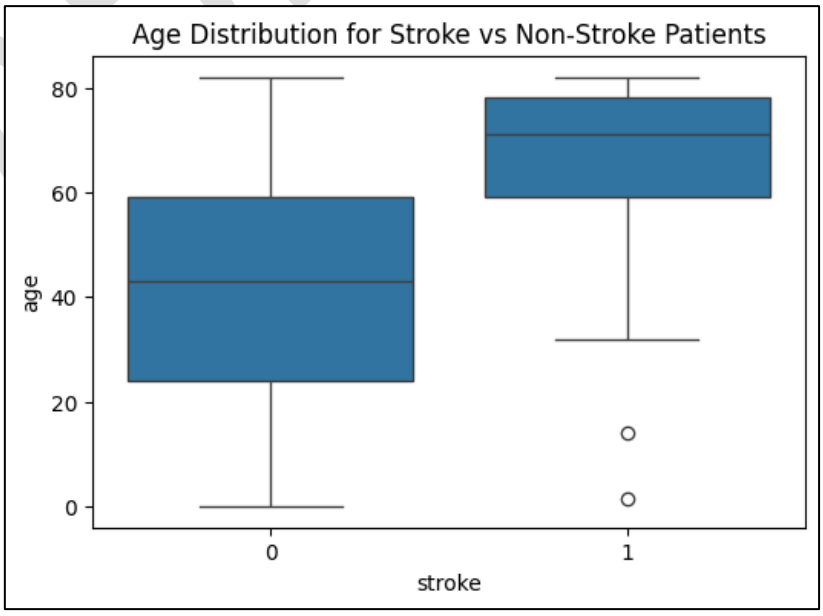




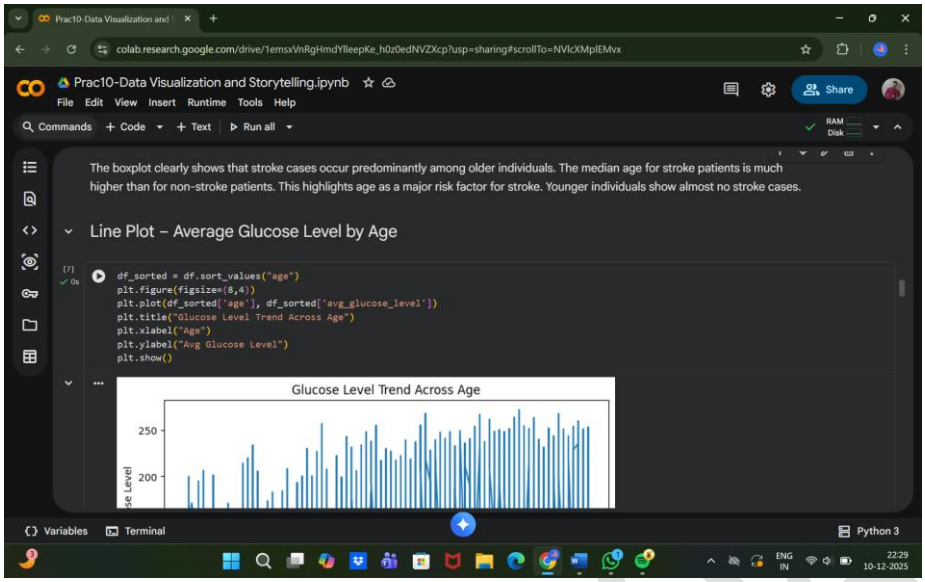
Sheth L.U.J College of Arts & Sir M.V. College of Science and Commerce
Data Science
PRACTICAL NO. 10

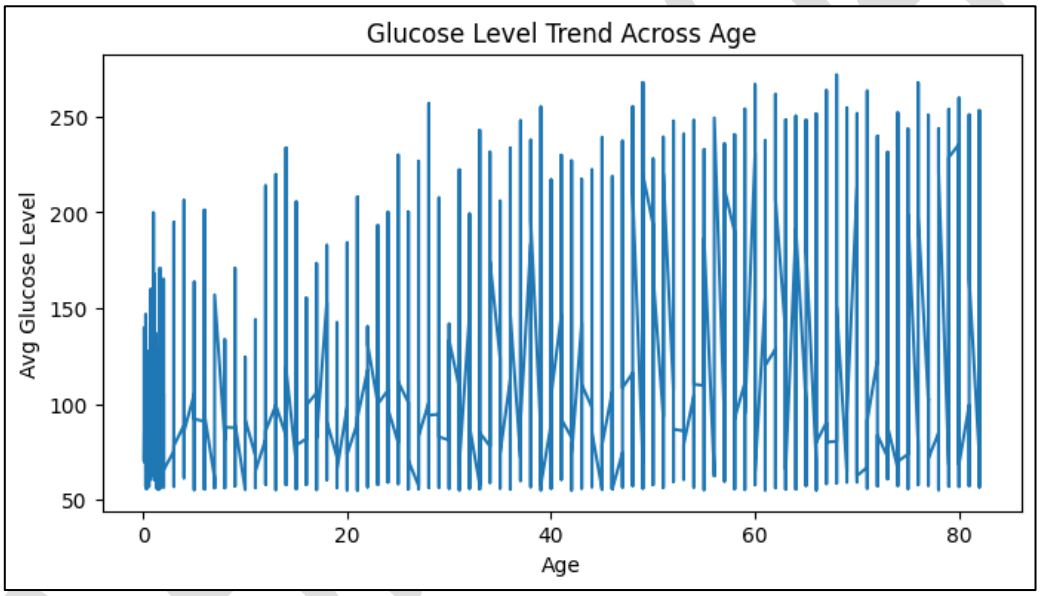


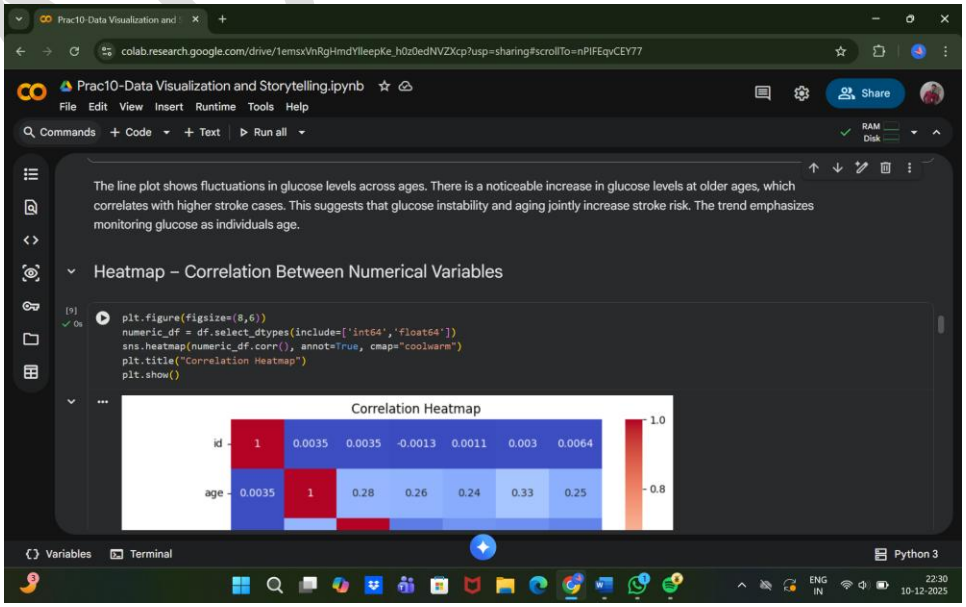




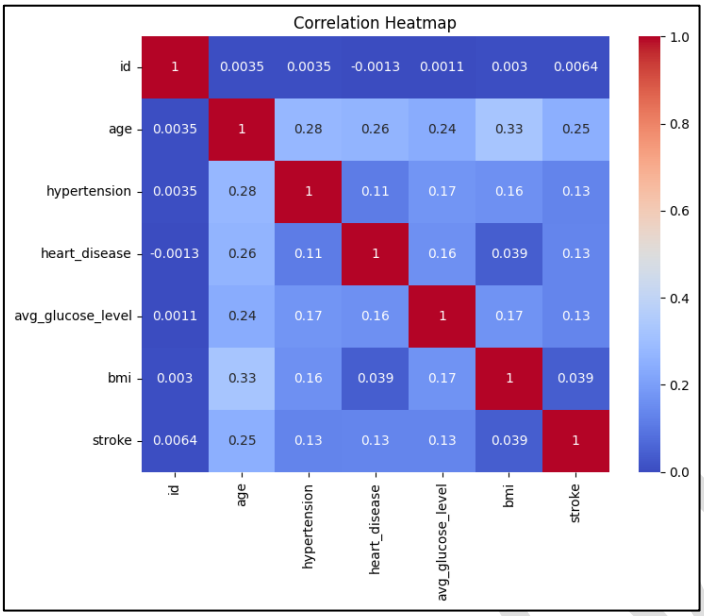
Sheth L.U.J College of Arts & Sir M.V. College of Science and Commerce
Data Science
PRACTICAL NO. 10

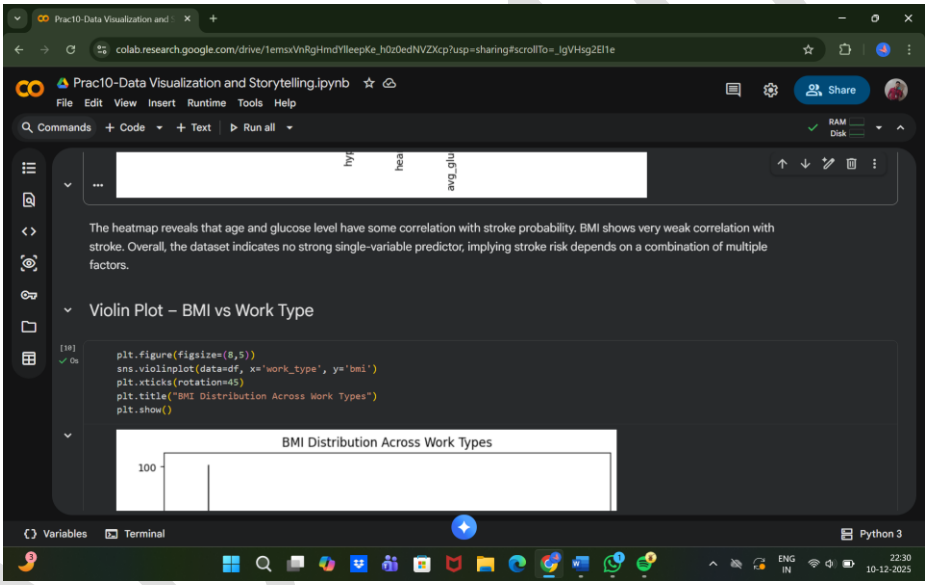


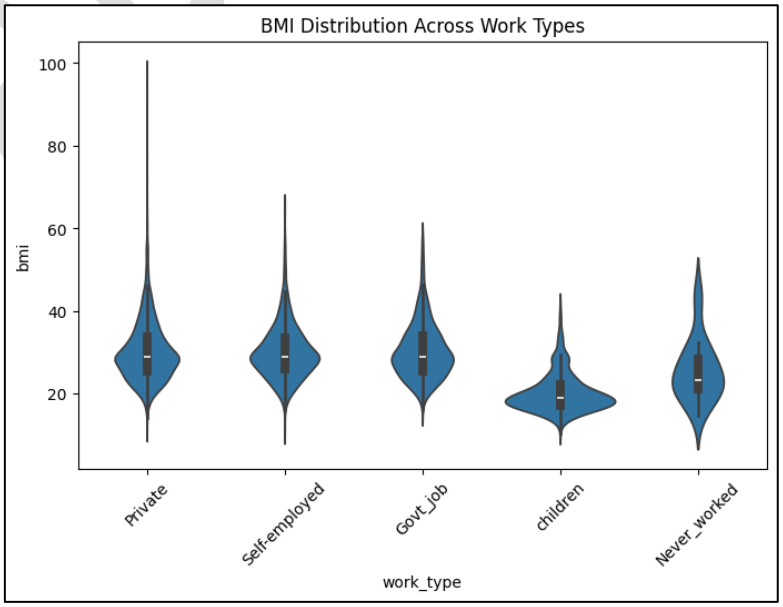




Sheth L.U.J College of Arts & Sir M.V. College of Science and Commerce
Data Science
PRACTICAL NO. 10



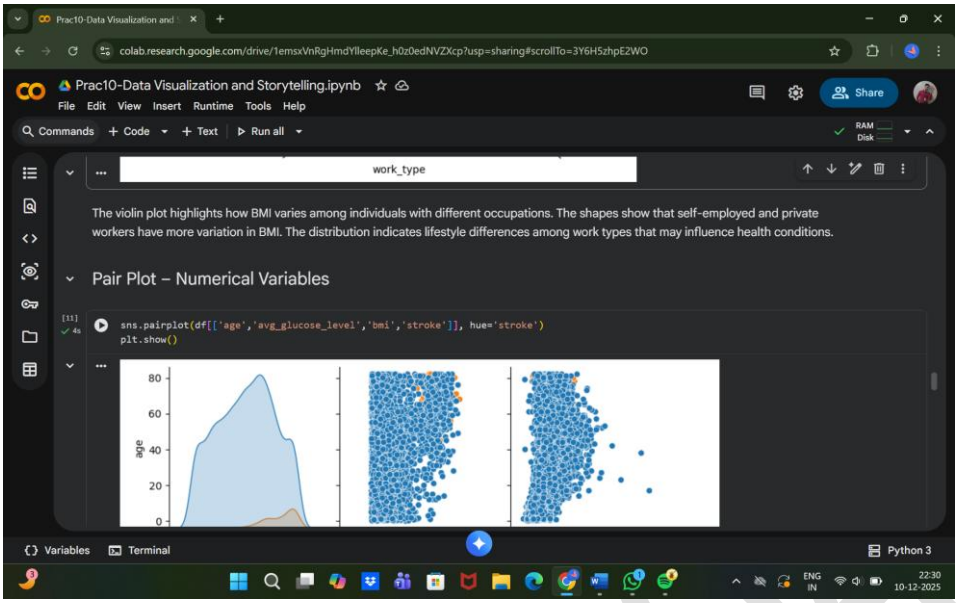


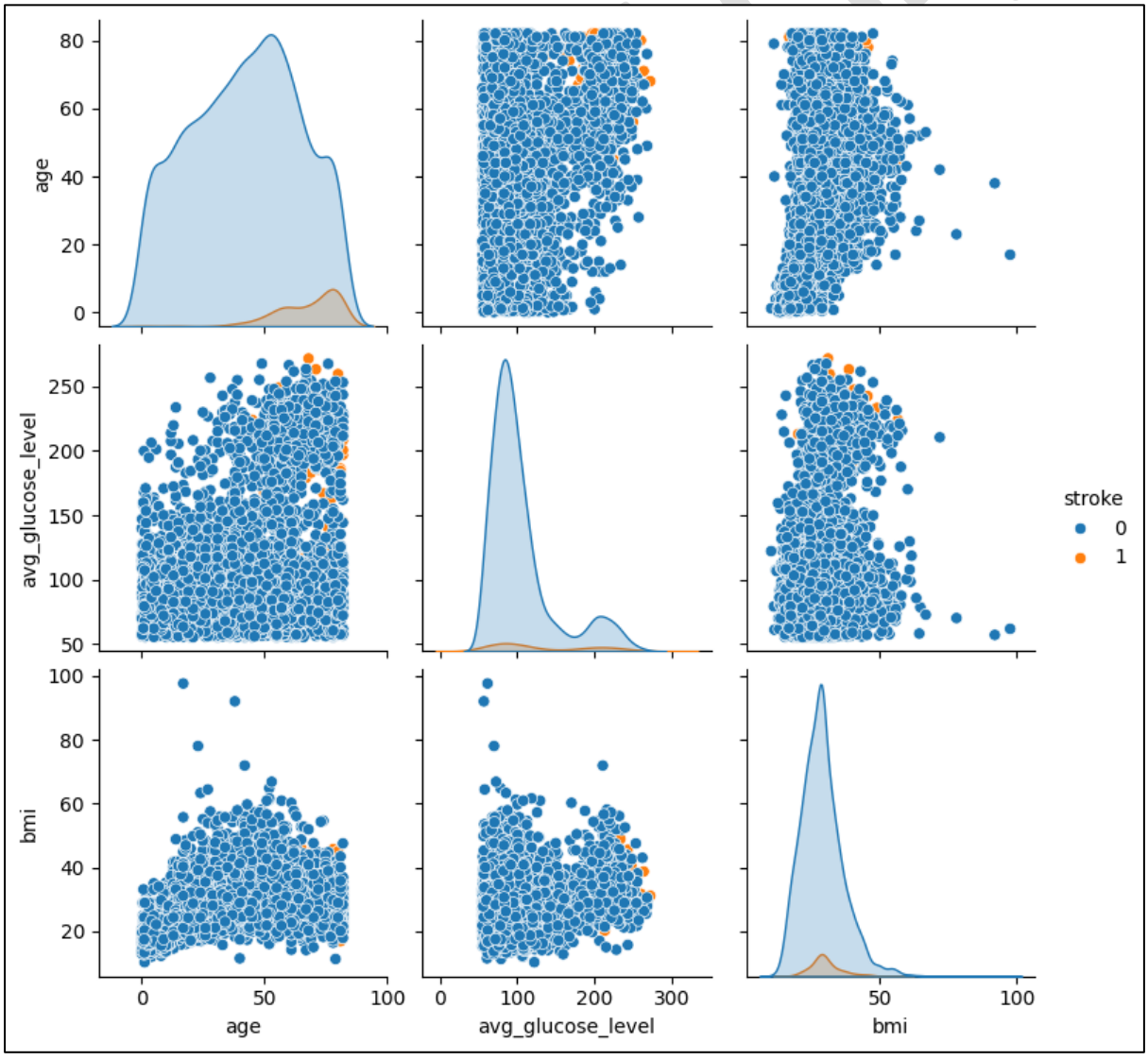


Sheth L.U.J College of Arts & Sir M.V. College of Science and Commerce

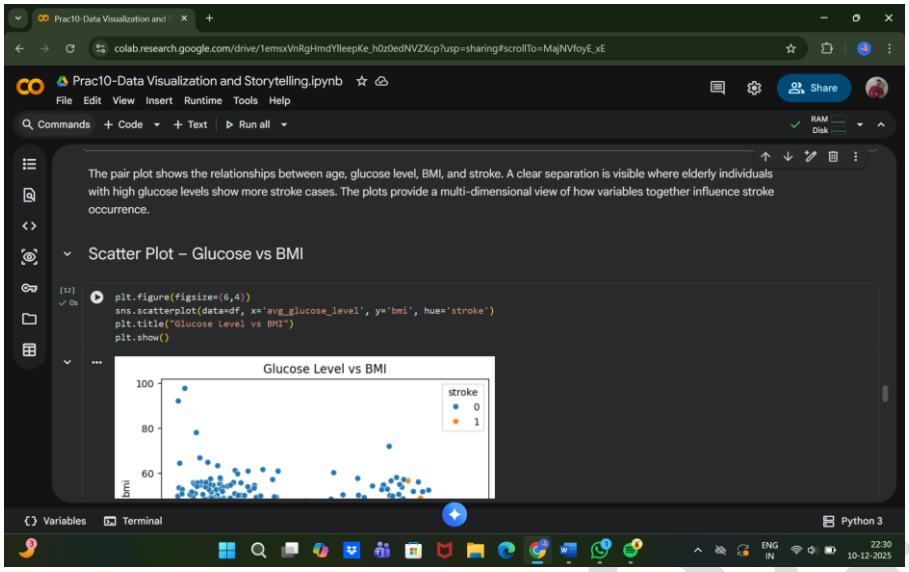
Data Science

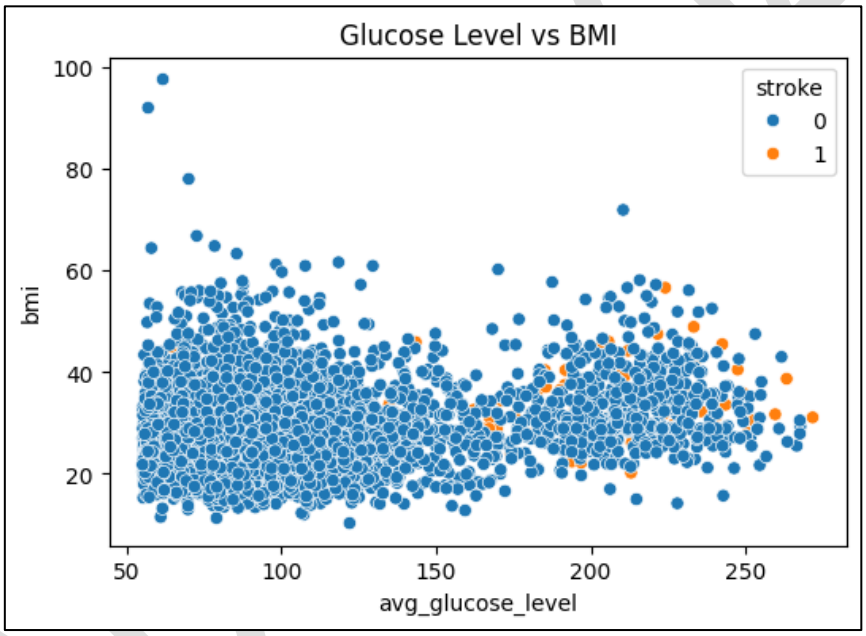
PRACTICAL NO. 10

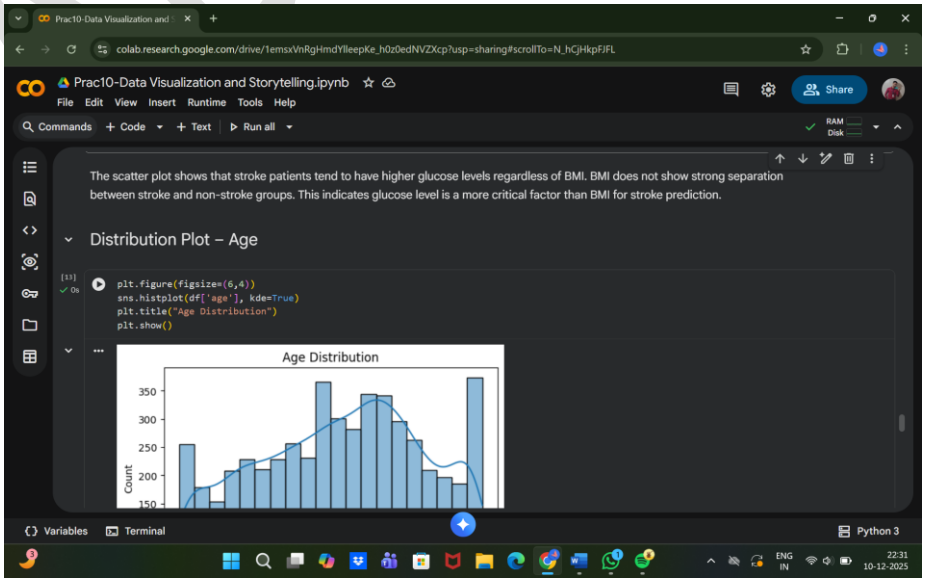




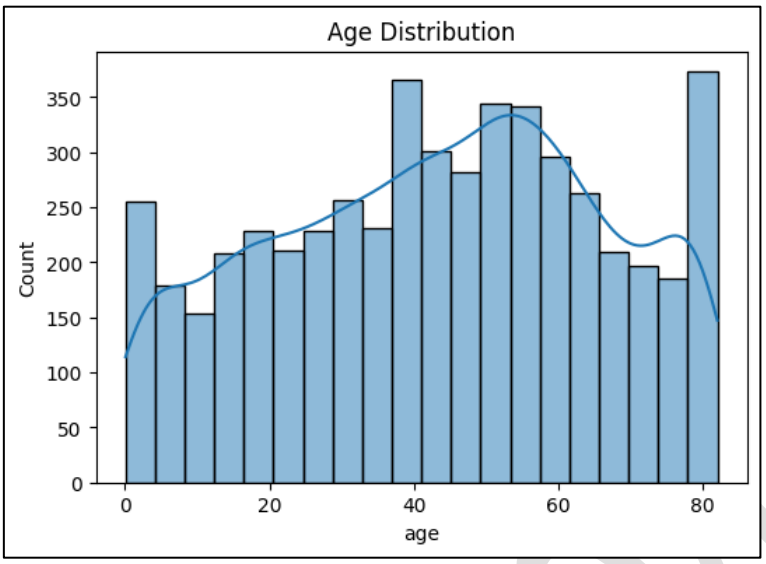
Sheth L.U.J College of Arts & Sir M.V. College of Science and Commerce
Data Science
PRACTICAL NO. 10

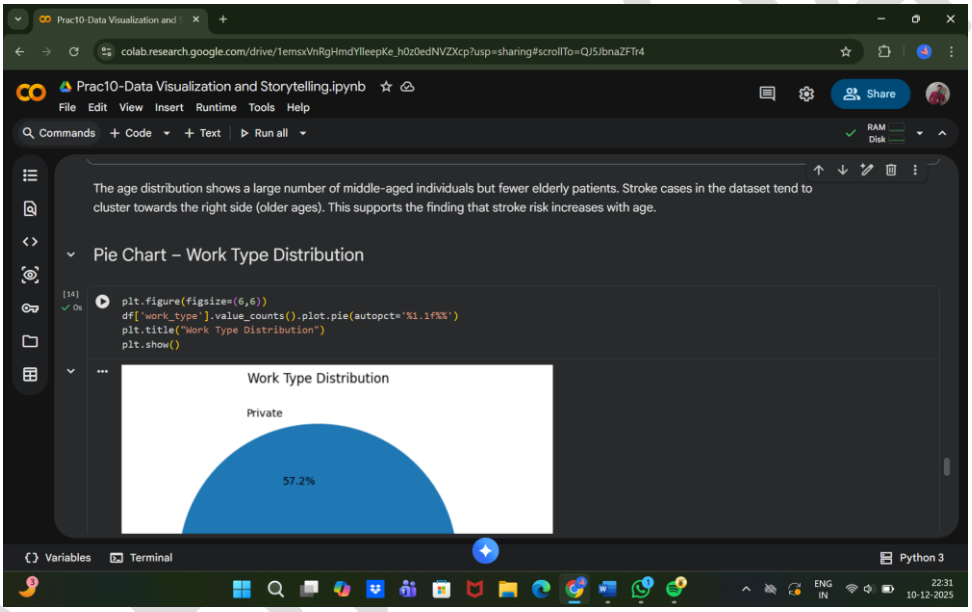


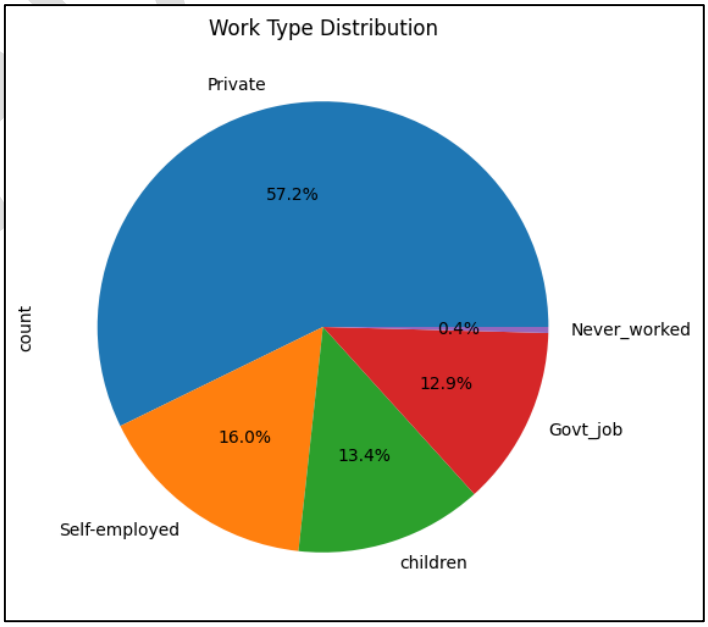


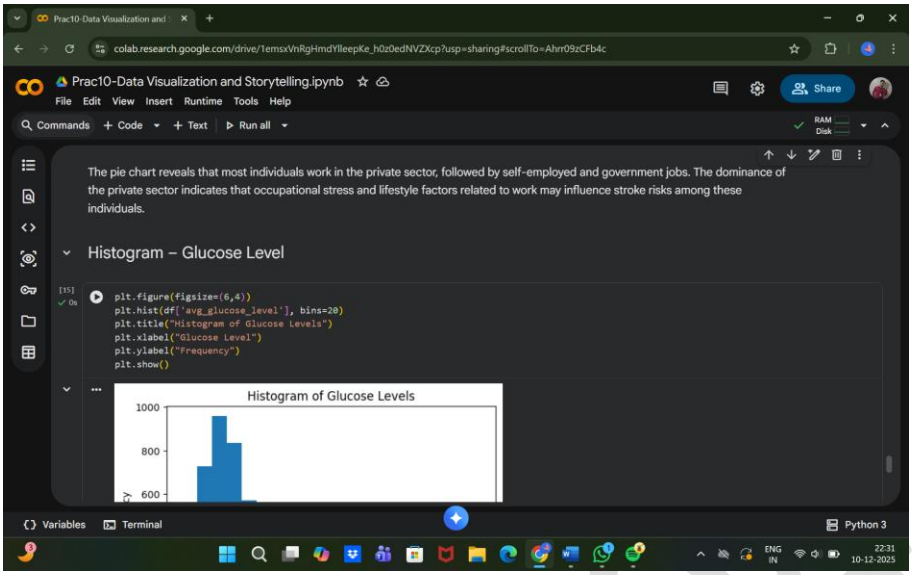


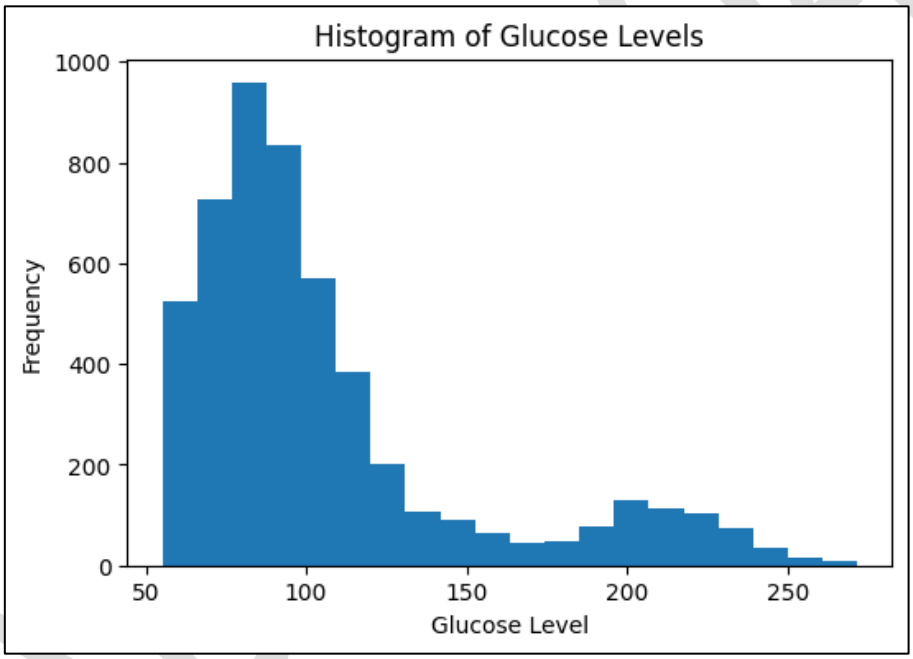
Sheth L.U.J College of Arts & Sir M.V. College of Science and Commerce
Data Science
PRACTICAL NO. 10

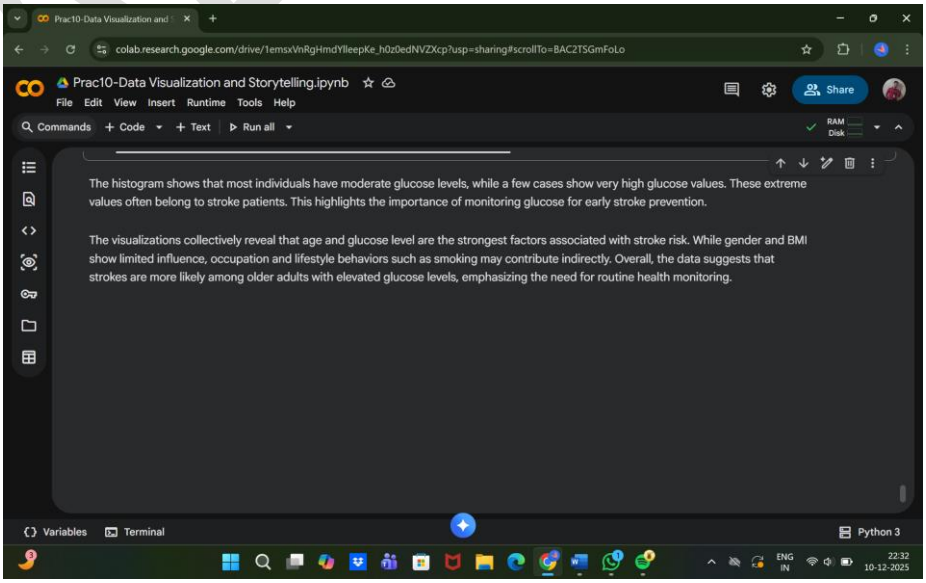







Prac10-Data Visualization and Storytelling.ipynb
File Edit View Insert Runtime Tools Help
Share
The pie chart reveals that most individuals work in the private sector, followed by self-employed and government jobs. The dominance of the private sector indicates that occupational stress and lifestyle factors related to work may influence stroke risks among these individuals.
Histogram – Glucose Level
plt.figure(figsize=(6,4))
plt.hist(df['avg_glucose_level'], bins=20)
plt.title("Histogram of Glucose Levels")
plt.xlabel("Glucose Level")
plt.ylabel("Frequency")
plt.show()
Histogram of Glucose Levels
1000
800
600
Variables
Terminal
Python 3


Histogram of Glucose Levels
Frequency
1000
800
600
400
200
0
50
100
150
200
250
Glucose Level


Prac10-Data Visualization and Storytelling.ipynb
File Edit View Insert Runtime Tools Help
Share
The histogram shows that most individuals have moderate glucose levels, while a few cases show very high glucose values. These extreme values often belong to stroke patients. This highlights the importance of monitoring glucose for early stroke prevention.
The visualizations collectively reveal that age and glucose level are the strongest factors associated with stroke risk. While gender and BMI show limited influence, occupation and lifestyle behaviors such as smoking may contribute indirectly. Overall, the data suggests that strokes are more likely among older adults with elevated glucose levels, emphasizing the need for routine health monitoring.
Variables
Terminal
Python 3