

Big Data Lab 06: Data Visualization in Spyder

DengFeng Qiao Yi He MingYi Wang

1 Introduction

This lab involves data visualization using Microsoft Power BI and Spyder. We explore basic data handling and visualization techniques using Python libraries.

2 Tasks and Code Samples

2.1 1. Data Loading

Load the Stroke dataset using `pandas`:

```
import pandas as pd
data = pd.read_csv("Stroke_dataset.csv")
```

2.2 2. Displaying Data

Show the first five records of the dataset:

```
print(data.head(5))
```

2.3 3. Histograms

Generate histograms for all numerical fields:

```
data.hist()
```

Create a histogram with customized bins for `avg_glucose_level`:

```
data.hist("avg_glucose_level", bins=20)
```

2.4 4. Multiple Variable Histogram

Create a subset of data and plot multiple histograms:

```
data1 = data[['avg_glucose_level', 'bmi']]
data1.plot.hist()
```

2.5 5. Bar Chart

Visualize Residence Type with a bar chart:

```
data["Residence_type"].value_counts().plot(kind="bar")
```

2.6 6. Scatter Plot

Generate a scatter plot to compare **age** and **bmi**:

```
data.plot.scatter(x="age", y="bmi")
```

2.7 7. Pie Chart

Plot the distribution of **work_type** with a pie chart:

```
ax = data["work_type"].value_counts().plot(kind="pie", title="Wo  
ax.xaxis.set_visible(False)  
ax.yaxis.set_visible(False)
```

2.8 8. Spotify Dataset

Load and plot the Spotify dataset:

```
import pandas as pd  
data = pd.read_csv("spotify.csv")  
data.plot(x="Date", kind="line", figsize=(14,6), xlabel="", ylab
```