

Module 7: Data Visualization Using Matplotlib and Seaborn

Assignment

edureka!

edureka!

© Brain4ce Education Solutions Pvt. Ltd.

Assignment

1. You are given a dataset, which is present in the LMS, containing the number of hurricanes occurring in the United States along the coast of the Atlantic. Load the data from the dataset into your program and plot a Bar Graph of the data, taking the Year as the X-axis and the number of hurricanes occurring as the Y-axis.
2. The given dataset records the data of city temperatures over the years 2014 and 2015. Plot the histogram of the temperatures over this period for the cities of San Francisco and Moscow.
3. Plot a pie-chart of the number of models released by every manufacturer, recorded in the provided data. Also, mention the name of the manufacturer with the largest releases.
4. Use sample-salesv2.csv file to carry out the following:

Phase 1 -Read the data

Phase 2 –Describe the data

Describe the data on the **unit price**.

Phase 3 –Filter the data

Create a new DataFrame having columns as **name**, **net_price**, **date**, and group all the records according to **name**.

Phase 4 –Plotting graph

Plot the graph after calculating the total sales by each customer. The customer name should be on the X-axis and total sales on the Y-axis.

5. Let the X-axis and Y-axis data points be as the following:

X = [1,2,3,4]

y = [20, 21, 20.5, 20.8]

5.1: Draw a simple plot

5.2: Configure the line and markers in a simple plot

5.3: Configure the axes

5.4: Give the title to the graph & labels of the X-axis and Y-axis.

5.5: Give the error bar if y_error = [0.12, 0.13, 0.2, 0.1]

5.6: Define width, height as figsize=(4,5) DPI and adjust plot dpi=100

5.7: Give a font size of 14

5.8: Draw a scatter graph of any 50 random values across X and Y axes

5.9: Create a DataFrame from the following data:

'first_name': ['Jason', 'Molly', 'Tina', 'Jake', 'Amy'],

'last_name': ['Miller', 'Jacobson', 'Ali', 'Milner', 'Cooze'],

'female': [0, 1, 1, 0, 1],

'age': [42, 52, 36, 24, 73],

'preTestScore': [4, 24, 31, 2, 3],

'postTestScore': [25, 94, 57, 62, 70]

Draw a Scatterplot of preTestScore and postTestScore, with the size of each point determined by age.

5.10: Draw a Scatterplot from the data in question 9 of preTestScore and postTestScore with the size = 300 and the color determined by sex

Case Study:

Domain – Retail

Focus – Visualize the sales data

Business challenge/requirement

BigMart is one of the biggest retailers in Europe and has operations across multiple countries. You are a data analyst in the IT team of BigMart. Invoice and SKU wise sales data for the Year 2011 is shared with you. You need to prepare meaningful charts to showcase the various sales trends for 2011 to top management.

Key issues

Data should be displayed pictorially to capture the attention of top management.

Considerations

NONE

Data volume

Approx. 500K records – file BigMartSalesData.csv

Additional information

NA

Business benefits

This exercise is an annual exercise and BigMart makes an important investment decision based on trends.

Approach to Solve

You must use the fundamentals of Matplotlib covered in module 5 and plot the following four charts/graph:

1. Plot total sales per month for the Year 2011. How have the total sales increased over the months in the Year 2011? Which month has the lowest sales?
2. Plot a Pie Chart for the Year 2011 Country Wise. Which Country contributes highest towards sales?
3. Plot Scatter Plot for the invoice amounts and see the concentration of amount. In which range have most of the invoice amounts concentrated?

Enhancements for code

You can try the following enhancements in code:

1. In Pie Chart, shadow=true, startangle=90 and see how different the chart looks
2. In the Scatter Plot, change the color of Scatter Points

edureka!