Module 3: Data Visualization

Demo II

edureka!

edureka!

© Brain4ce Education Solutions Pvt. Ltd.

Demo II

Domain - Retail

Focus – Visualize the sales data

Business challenge/requirement
BigMart is one of the biggest retailer of Europe and has operations across multiple countries. You are a data analyst in IT team of BigMart. Invoice and SKU wise Sales Data for Year 2011 is shared with you. You need to prepare meaningful charts to show case 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

edureka!

Data volume
- Approx 500K records – file BigMartSalesData.csv

Additional information
- NA

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

Approach to Solve
You have to use fundamentals of Matplotlib and plot the following 4 chart/graph

1. Plot Total Sales Per Month for Year 2011 and for the European Community. How was the total sales increased over months in Year 2011. Which month has lowest Sales?
2. Plot Total Sales Per Month for Year 2011 and for the European Community as Bar Chart. Is Bar Chart Better to visualize than Simple Plot?
3. Plot Pie Chart for Year 2011 Country Wise. Which Country contributes highest towards sales?

4. Plot Scatter Plot for the invoice amounts and see the concentration of amount. In which range most of the invoice amounts are concentrated

Enhancements for code

You can try these enhancements in code

1. In Pie Chart Play with Parameters shadow=True, startangle=90 and see how different the chart looks
2. In scatter plot change the color of Scatter Points

edureka!

Demo II Solution

1. Plot Total Sales Per Month for Year 2011 and for the European Community. How was the total sales increased over months in Year 2011. Which month has lowest Sales?

Solution

```
# Get Sales for the Year 2011 and for the country 'European Community'
import pandas as pd
import matplotlib.pyplot as plt

#  Read BigMartSales.csv as a dataframe called salesdata
sales_data = pd.read_csv('BigMartSalesData.csv')

# Check the basic data
print(sales_data.head())
print(sales_data.info())

print (" Getting Sales Data for Year 2011")
sales_2011_euro = sales_data[(sales_data['Year'] == 2011) &
(sales_data['Country'] == 'European Community' )]
print ("Getting Amount for Each Month")
sales_2011_month = sales_2011_euro.groupby('Month').count()
print(sales_2011_month)

# Simply Plot the Sales Data for 2011 , X Axis Month Number, Y -Axis Sales for
each month

sales_2011_euro.groupby('Month').count().plot()
plt.xlabel("Month Number")
plt.ylabel("Sales in Euro")
plt.title("Sales Per Month in Year 2011")
plt.show()

# Save the Plot Locally
plt.savefig("Year2011MonthWiseSales")
```

2. Plot Total Sales Per Month for Year 2011 and for the European Community as Bar Chart. Is Bar Chart Better to visualize than Simple Plot?

Solution

```
sales_2011_euro.groupby('Month').count().plot(kind="bar")
plt.xlabel("Month Number")
plt.ylabel("Sales in Euro")
plt.title("Sales Per Month in Year 2011")
plt.show()
# Enhancement Can you show the value of the bar
# Refer to https://matplotlib.org/gallery/api/barchart.html#sphx-glr-gallery-api-barchart-py
```

3. Plot Pie Chart for Year 2011 Country Wise. Which Country contributes highest towards sales?

Solution

```
sales_2011 = sales_data[sales_data['Year'] == 2011]
sales_country_wise = sales_2011.groupby('Country').sum()['Amount']
plt.title("Country Wise Contribution For 2011")
plt.pie(sales_country_wise.values,labels=sales_country_wise.index,autopct='%1.1f%%', )
plt.show()
# Enhancement -- Play With Parameters  shadow=True, startangle=90 etc in plt.pie and see how different the chart looks
```

4. Plot Scatter Plot for the invoice amounts and see the concentration of amount. In which range most of the invoice amounts are concentrated

Solution

```
sales_invoice_wise = sales_2011.groupby('InvoiceNo').sum()['Amount']
plt.scatter(sales_invoice_wise.values,sales_invoice_wise.values)
plt.grid(True)
plt.show()
# Enhancement -- Change the color of points use color=['red','green','blue']
```