**CSE3020 – Data Visualization (ELA), Winter Semester 2021-2022**

**Lab Assignment IA7 – Slot L43-L44**

**By: Jonathan Rufus Samuel (20BCT0332) Dos: 21.03.2022**

---

# Lab Assignment – IA7 Hierarchical Data Visualizations - 2

## Note on Software used for following Visualizations: (Tableau)

Microsoft Excel is a spreadsheet developed by Microsoft for Windows, macOS, Android and iOS. It features calculation or computation capabilities, graphing tools, pivot tables, and a macro programming language called Visual Basic for Applications (VBA). Excel forms part of the Microsoft Office suite of software.

It includes:

- Easy to access from different sources.
- No need for any technical or programming knowledge, and Quick response for making a dashboard.
- In terms of connecting and sharing, it has various inbuilt advanced features such as: Collaboration and distribution, highly securable, Multiple data sources connection, Easy importation and exportation of the massive size of data.
- For easy accessibility and analysis, the data file can be downloaded locally on mobile or desktop, multilingual representation of data, real-time exploration of any dataset, etc.

## Note on Software used for following Visualizations: (Python)

Python is a High-Level Programming Language, commonly used for data analysis, AI and Machine Learning related fields.

It includes:

- Python is a **high-level programming language** that has English-like syntax. This makes it easier to read and understand the code.
- Python is really easy to **pick up** and **learn**, that is why a lot of people recommend Python to beginners. You need less lines of code to perform the same task as compared to other major languages like **C/C++** and **Java**.

**Q) For the created hierarchical data or any other suitable dataset, perform the visualization using:**

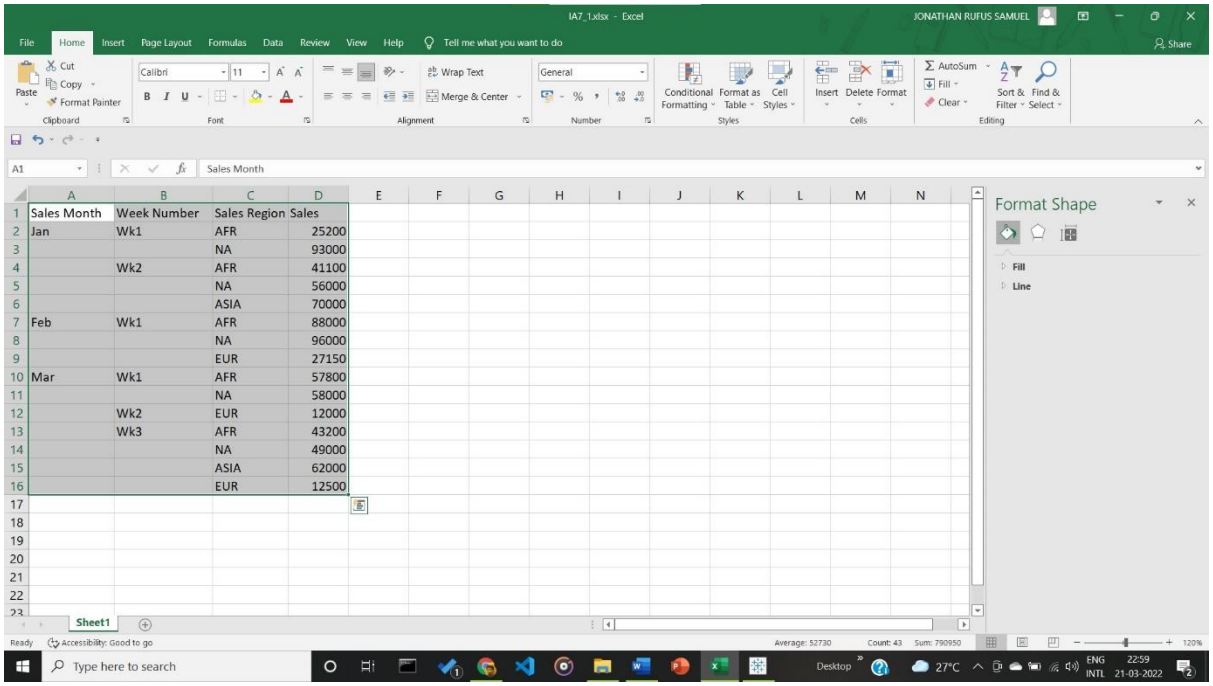
**a. Radial diagram**

**b. Sunburst diagram**
**c. Dendrogram**

**Answer**: Some Key points to note before visualization process:

# 1. Dataset Used:

The dataset is self-designed dataset, loosely based on the https://www.myexcelonline.com/blog/create-sunburst-chart-excel-2016/ dataset. A Hierarchical dataset of Sales in a week in different regions of the World.

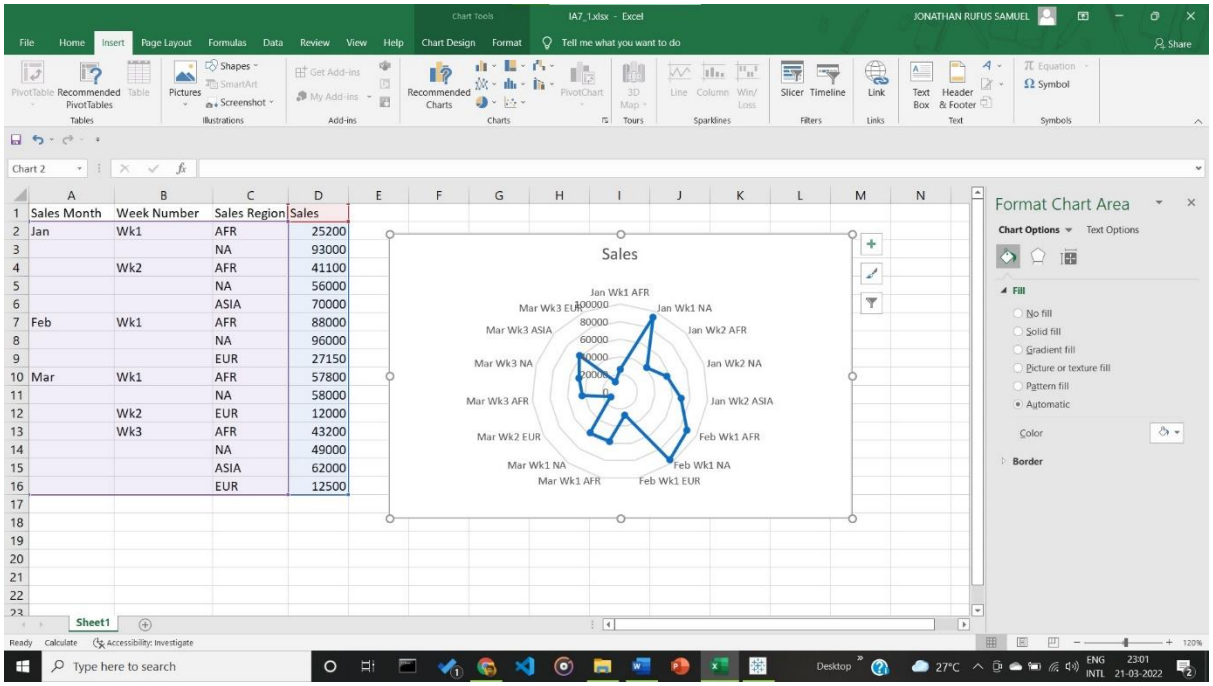



| Sales Month | Week Number | Sales Region | Sales |
|---|---|---|---|
| Jan | Wk1 | AFR | 25200 |
| | | NA | 93000 |
| | Wk2 | AFR | 41100 |
| | | NA | 56000 |
| | | ASIA | 70000 |
| Feb | Wk1 | AFR | 88000 |
| | | NA | 96000 |
| | | EUR | 27150 |
| Mar | Wk1 | AFR | 57800 |
| | | NA | 58000 |
| | Wk2 | EUR | 12000 |
| | Wk3 | AFR | 43200 |
| | | NA | 49000 |
| | | ASIA | 62000 |
| | | EUR | 12500 |

# 2. Visualization:

Visualization is Sales made in respective weeks in the months of January, February and March, in the regions of Africa, North America, Asia and Europe.

**a. Visualization using Radial Diagram for Given Hierarchical Dataset Visualization**

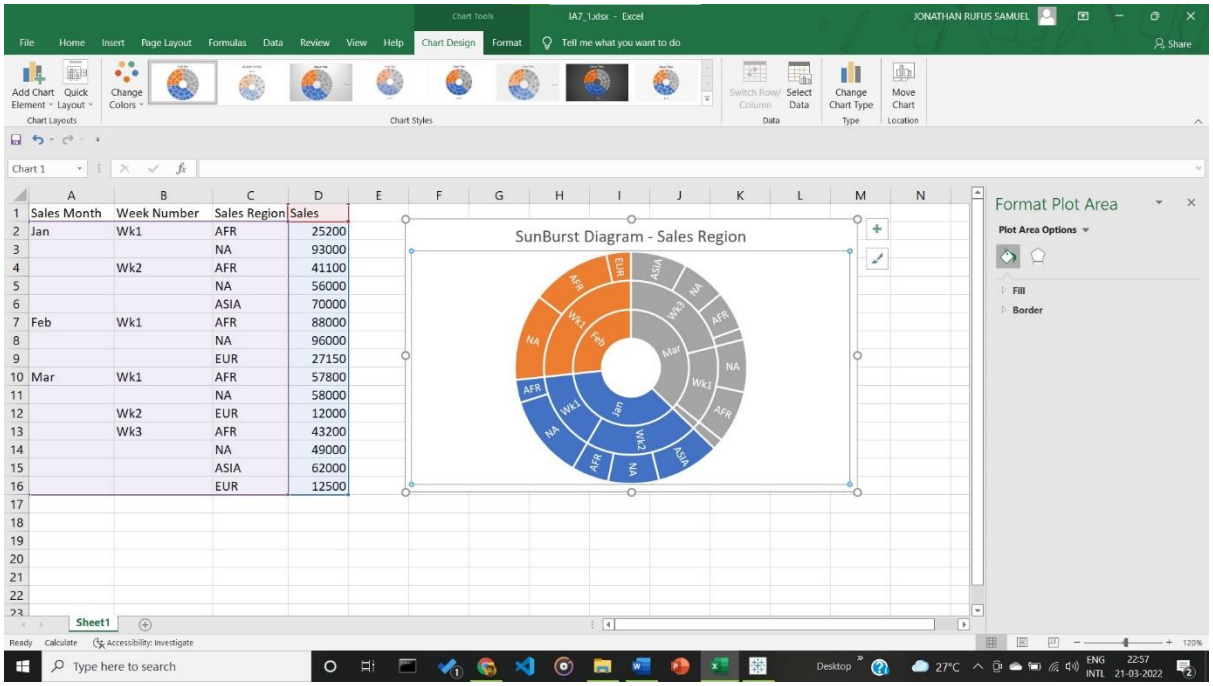



| | A | B | C | D |
|---|---|---|---|---|
| 1 | Sales Month | Week Number | Sales Region | Sales |
| 2 | Jan | Wk1 | AFR | 25200 |
| 3 | | | NA | 93000 |
| 4 | | Wk2 | AFR | 41100 |
| 5 | | | NA | 56000 |
| 6 | | | ASIA | 70000 |
| 7 | Feb | Wk1 | AFR | 88000 |
| 8 | | | NA | 96000 |
| 9 | | | EUR | 27150 |
| 10 | Mar | Wk1 | AFR | 57800 |
| 11 | | | NA | 58000 |
| 12 | | Wk2 | EUR | 12000 |
| 13 | | Wk3 | AFR | 43200 |
| 14 | | | NA | 49000 |
| 15 | | | ASIA | 62000 |
| 16 | | | EUR | 12500 |

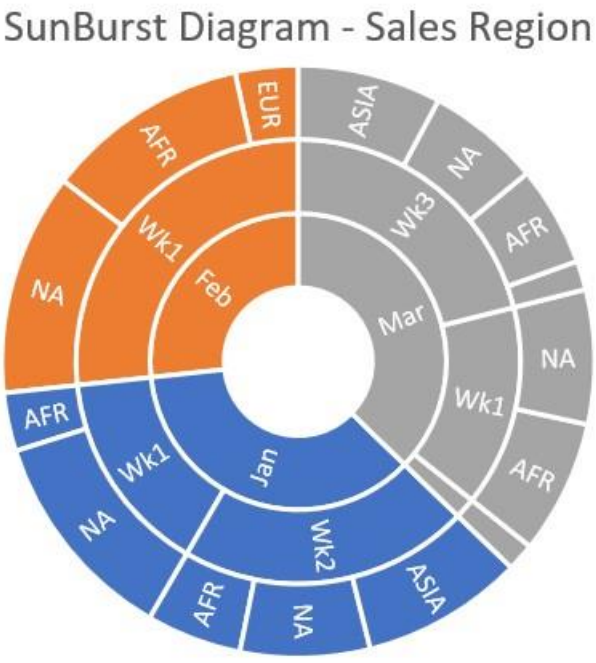

**Inference:**

Radial Diagram is displayed as required, for all the weeks from the months Jan, Feb and March. It can be inferred that least sales were made in the 3rd week of March and the Max sales were made in the 1st week of February.

**b. Visualization using Sunburst Diagram for Given Hierarchical Dataset Visualization**




| | A | B | C | D |
|---|---|---|---|---|
| 1 | Sales Month | Week Number | Sales Region | Sales |
| 2 | Jan | Wk1 | AFR | 25200 |
| 3 | | | NA | 93000 |
| 4 | | Wk2 | AFR | 41100 |
| 5 | | | NA | 56000 |
| 6 | | | ASIA | 70000 |
| 7 | Feb | Wk1 | AFR | 88000 |
| 8 | | | NA | 96000 |
| 9 | | | EUR | 27150 |
| 10 | Mar | Wk1 | AFR | 57800 |
| 11 | | | NA | 58000 |
| 12 | | Wk2 | EUR | 12000 |
| 13 | | Wk3 | AFR | 43200 |
| 14 | | | NA | 49000 |
| 15 | | | ASIA | 62000 |
| 16 | | | EUR | 12500 |




**Inference:**

Sunburst Diagram is displayed as required, for the months Jan, Feb and March. It can be inferred that minimum sales were made in the month of February from the Sunburst Diagram shown above.

**c. Visualization using Dendrogram for Given Hierarchical Dataset Visualization**

**Code:**

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

data = pd.read_csv('IA7\IA7.csv')
print(data.head(15))

from sklearn.preprocessing import normalize
data_scaled = normalize(data)
data_scaled = pd.DataFrame(data_scaled, columns=data.columns)
print(data_scaled.head())

import scipy.cluster.hierarchy as shc
plt.figure(figsize=(10, 7))
plt.title("Dendrograms")
dend = shc.dendrogram(shc.linkage(data_scaled, method='ward'))

plt.figure(figsize=(10, 7))
plt.title("Dendrograms")
dend = shc.dendrogram(shc.linkage(data_scaled, method='ward'))
plt.axhline(y=6, color='r', linestyle='--')
plt.show()
```

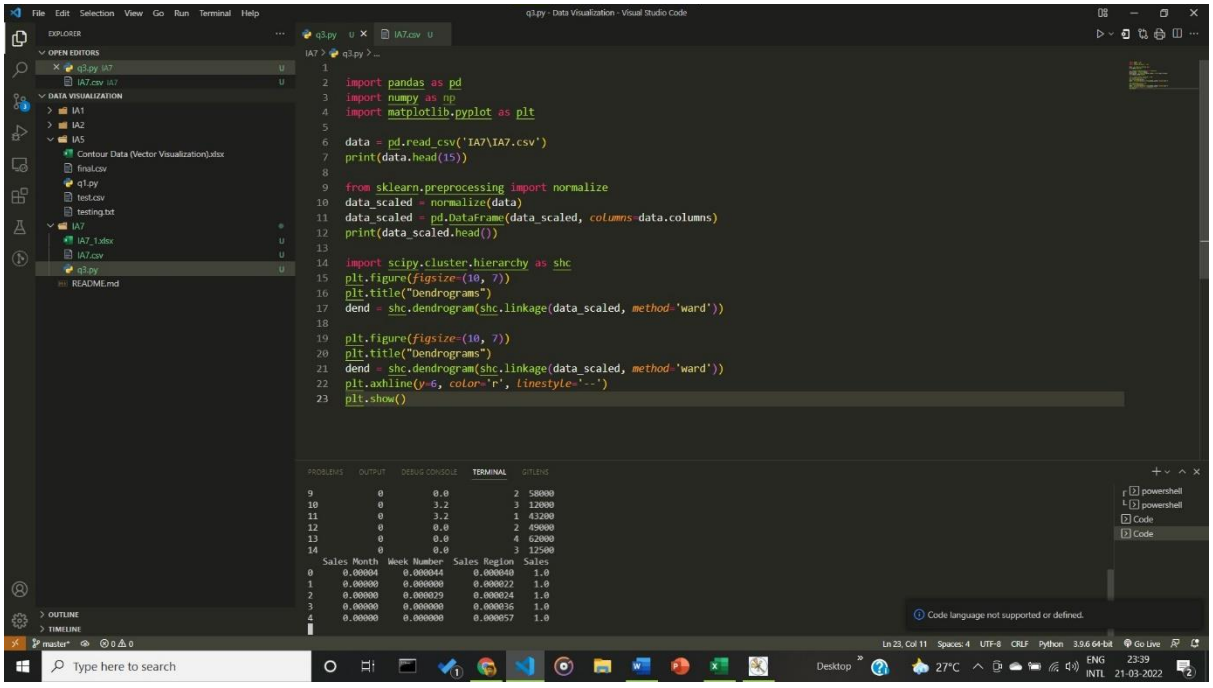
**Output:**

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Try the new cross-platform PowerShell https://aka.ms/pscore6

PS D:\CompSci - Learn\Python\Data Visualization> python -u "d:\CompSci -
Learn\Python\Data Visualization\IA7\q3.py"
   Sales Month  Week Number  Sales Region  Sales
0      1      1.1          1  25200
1      0      0.0          2  93000
2      0      1.2          1  41100
3      0      0.0          2  56000
4      0      0.0          4  70000
```

```
5          2       2.1          1  88000
6          0       0.0          2  96000
7          0       0.0          3  27150
8          3       3.1          1  57800
9          0       0.0          2  58000
10         0       3.2          3  12000
11         0       3.2          1  43200
12         0       0.0          2  49000
13         0       0.0          4  62000
14         0       0.0          3  12500
   Sales Month  Week Number  Sales Region  Sales
0      0.00004     0.000044      0.000040    1.0
1      0.00000     0.000000      0.000022    1.0
2      0.00000     0.000029      0.000024    1.0
3      0.00000     0.000000      0.000036    1.0
4      0.00000     0.000000      0.000057    1.0
PS D:\CompSci - Learn\Python\Data Visualization>
```



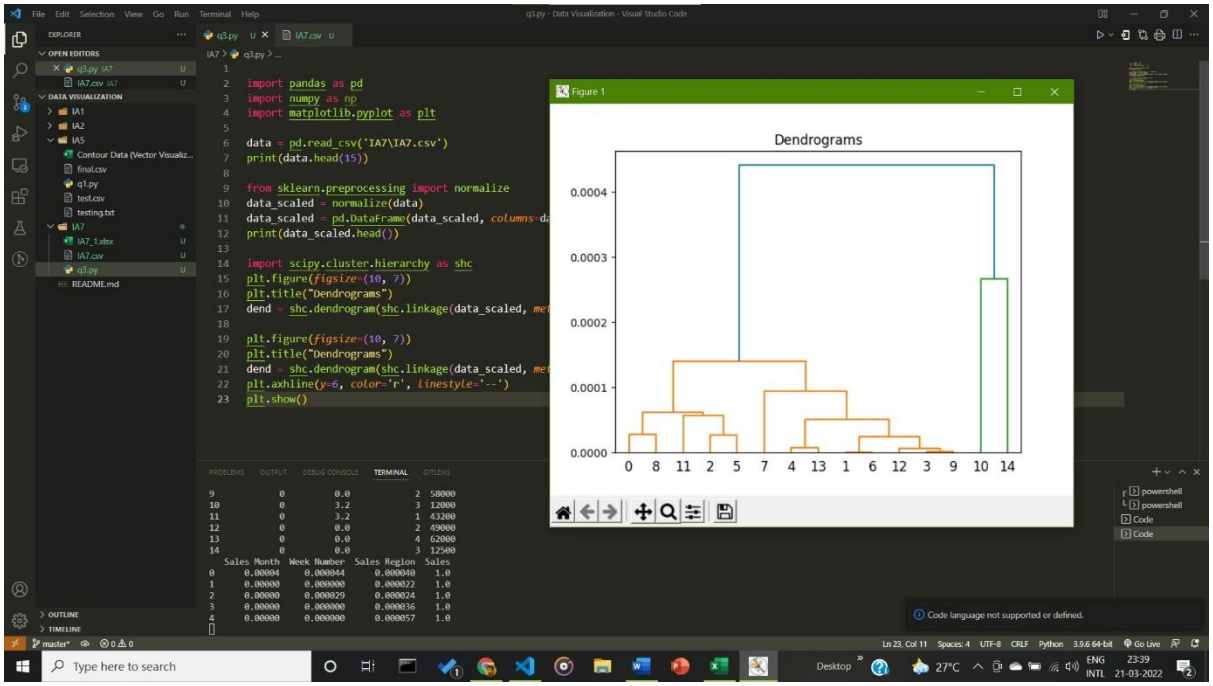


Figure 1
Dendrograms
0.0004
0.0003
0.0002
0.0001
0.0000
0 8 11 2 5 7 4 13 1 6 12 3 9 10 14

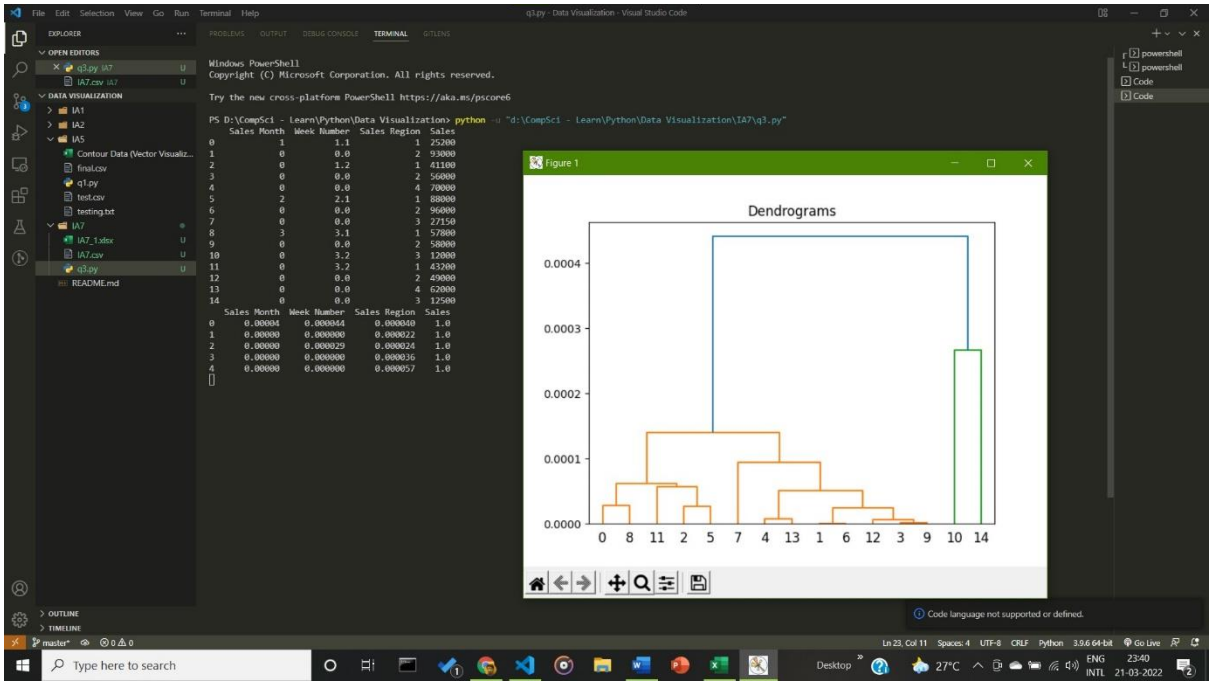
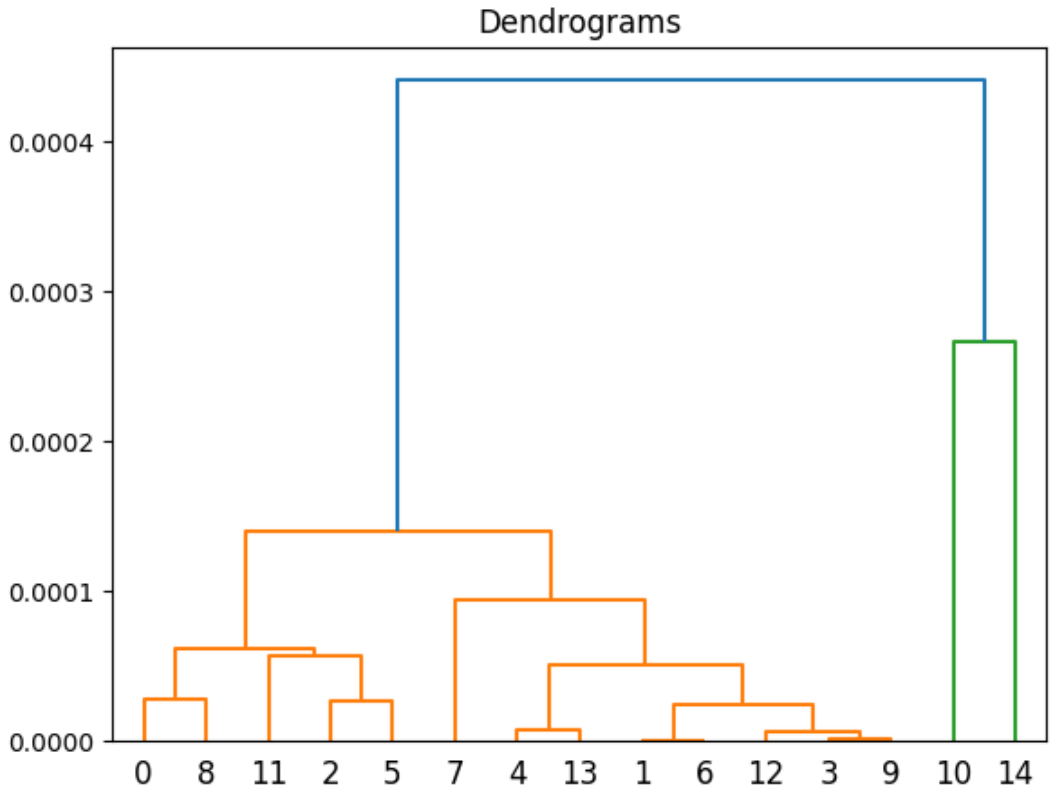


Figure 1
Dendrograms
0.0004
0.0003
0.0002
0.0001
0.0000
0 8 11 2 5 7 4 13 1 6 12 3 9 10 14




**Inference:**

The Dendrogram is displayed as required. The x axis sows the Weeks leading from January to March, with the Y axis depicting the sales made. Max sales were made between weeks 5 to 13.

---