Date and clusters

To get idea about dates and sales on that on each cluster we plotted graph of Date vs Total_Sales for each cluster using matplotlib.pyplot.

cluster_info = {

0: ['France', 'Portugal', 'Spain', 'Ireland', 'United Kingdom', 'Iceland'],

1: ['Israel', 'Lebanon', 'United Arab Emirates', 'Saudi Arabia', 'Cyprus'],

2: ['Poland', 'Austria', 'Lithuania', 'Germany', 'Greece', 'Switzerland', 'Belgium', 'Netherlands', 'Italy', 'Denmark'],

3: ['Norway', 'Sweden', 'Finland'],

4: ['Brazil'],

5: ['South Africa'],

6: ['Canada', 'United States of America'],

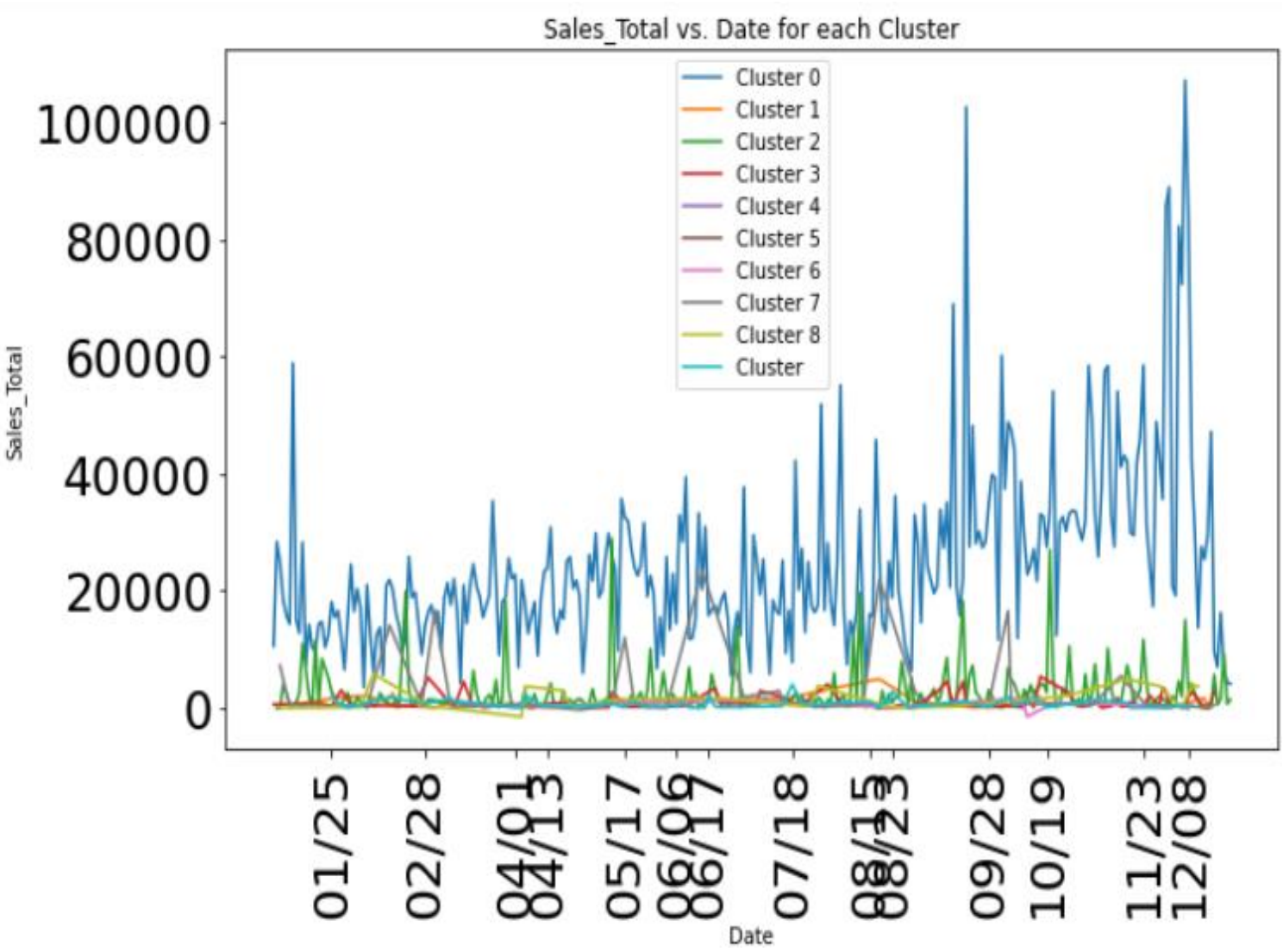
7: ['Australia'],

8: ['Japan']

** Unnamed cluster is for other countries

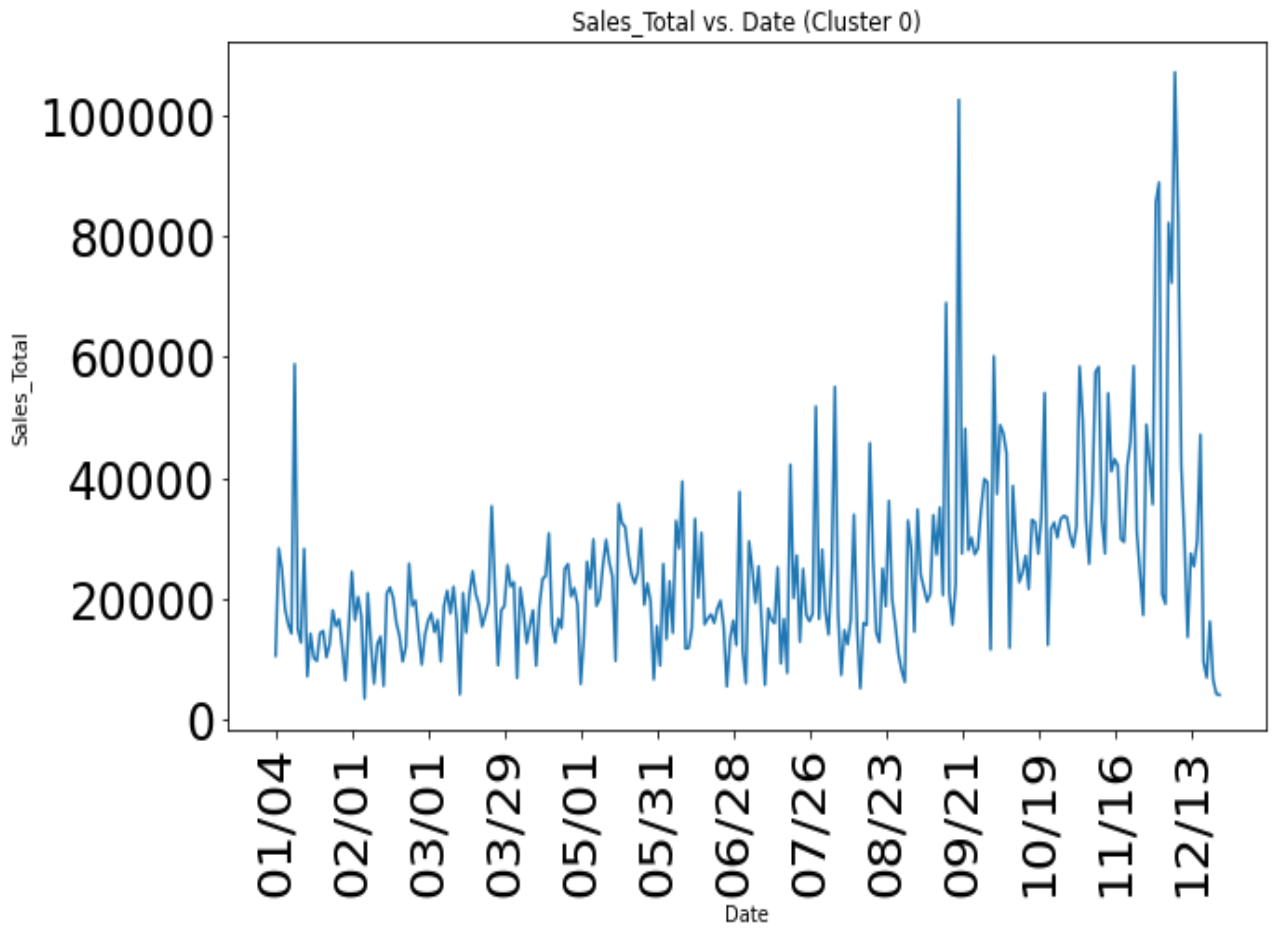
}

Plots-



This is all cluster in one plot.

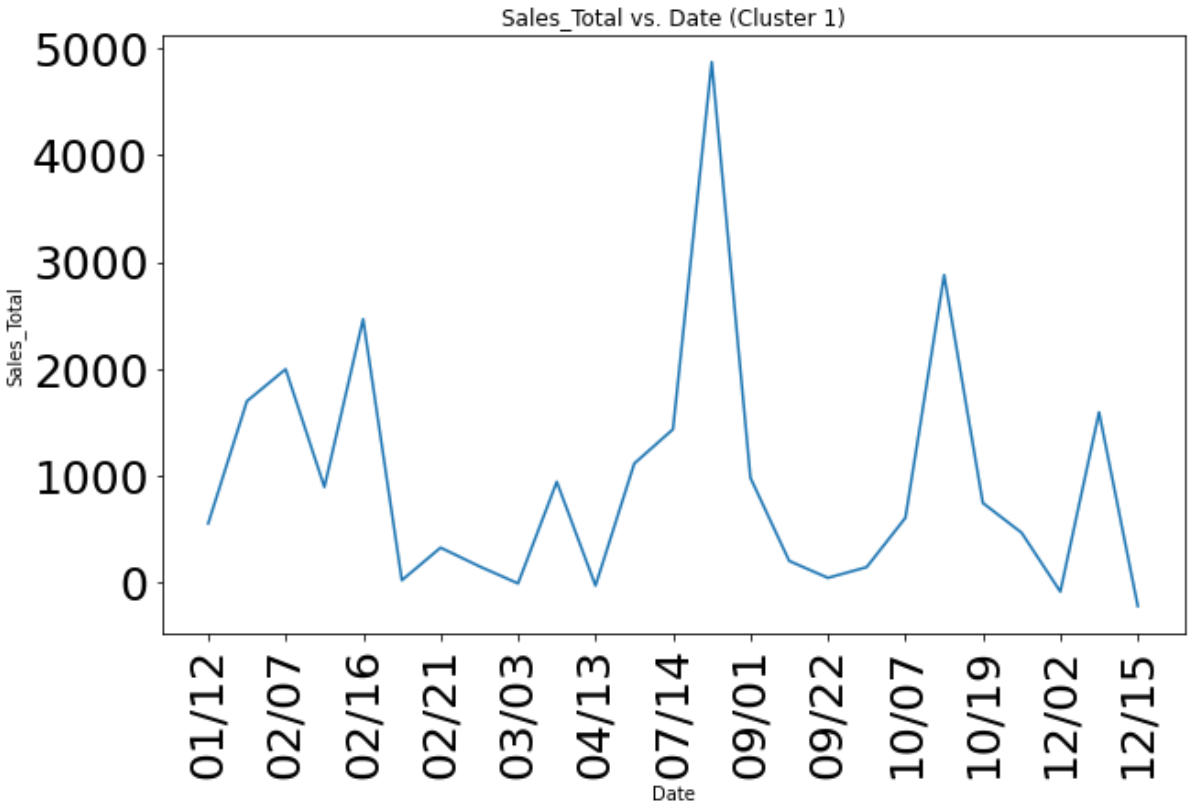
Cluster 0 is dominating.

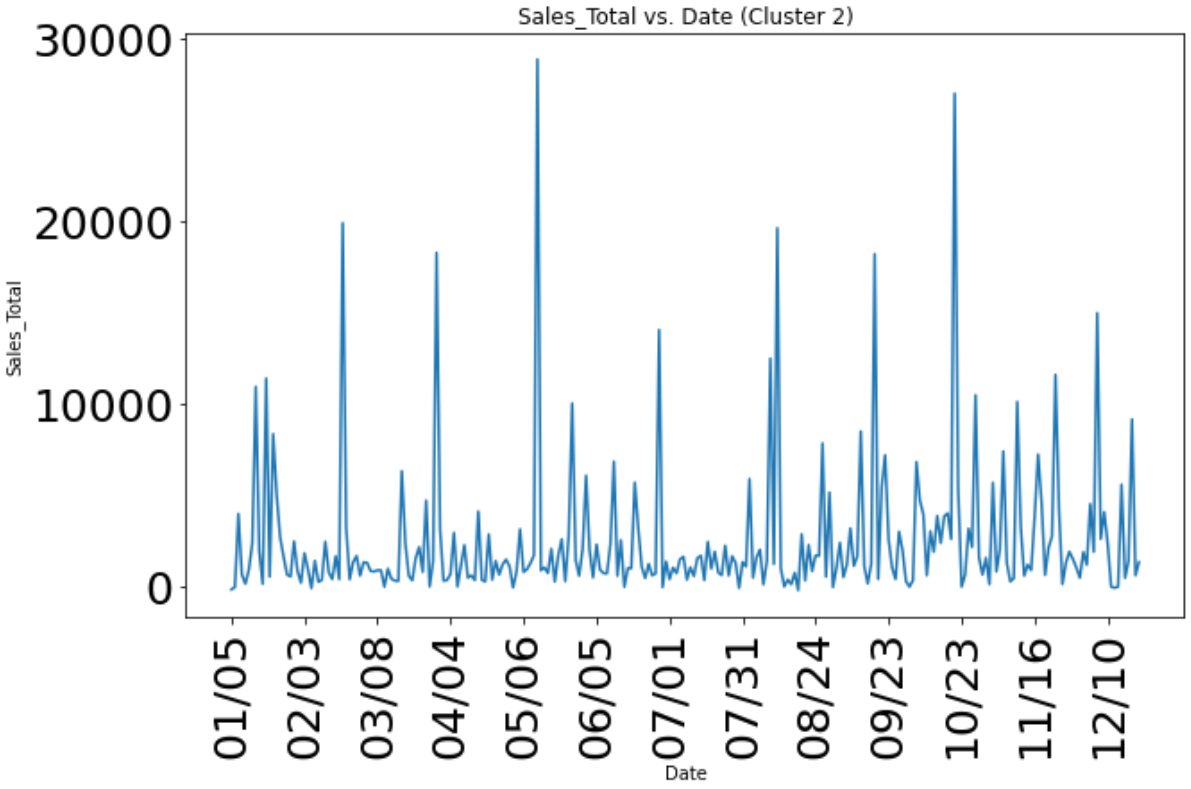


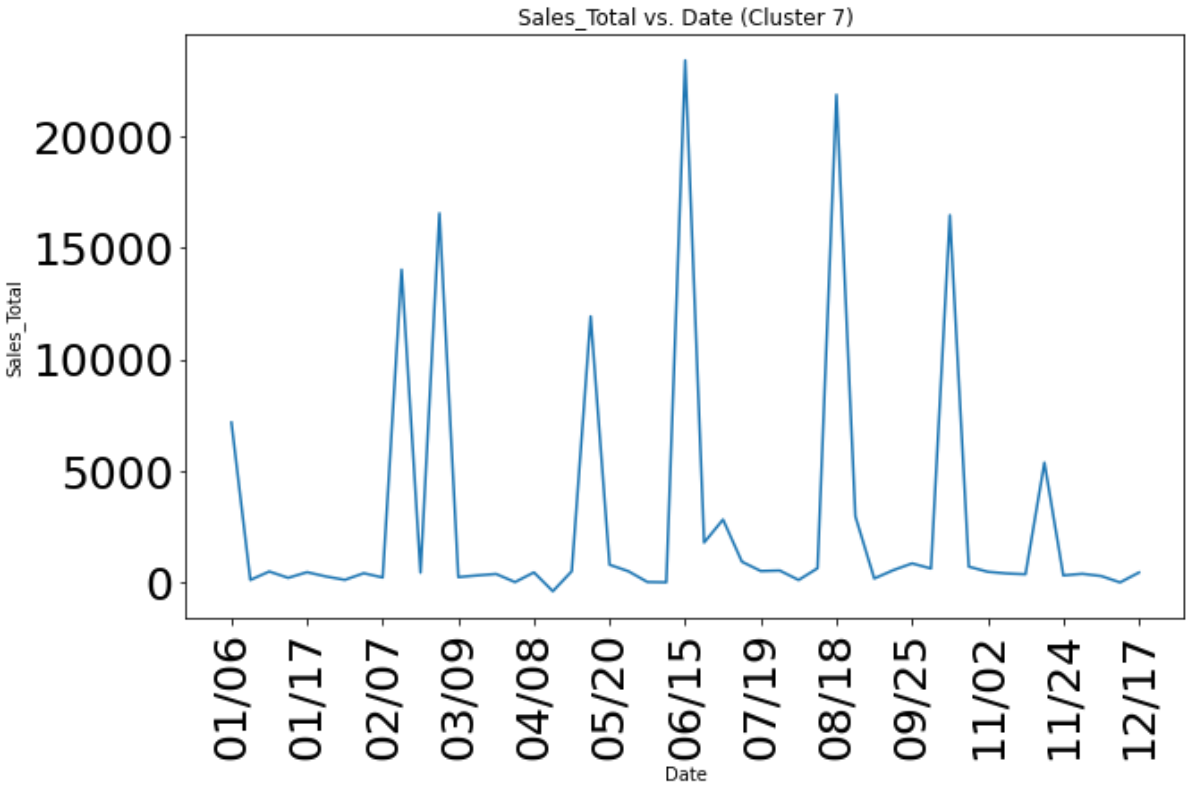
This is plot only for cluster 0 (UK and Neighbouring European countries)

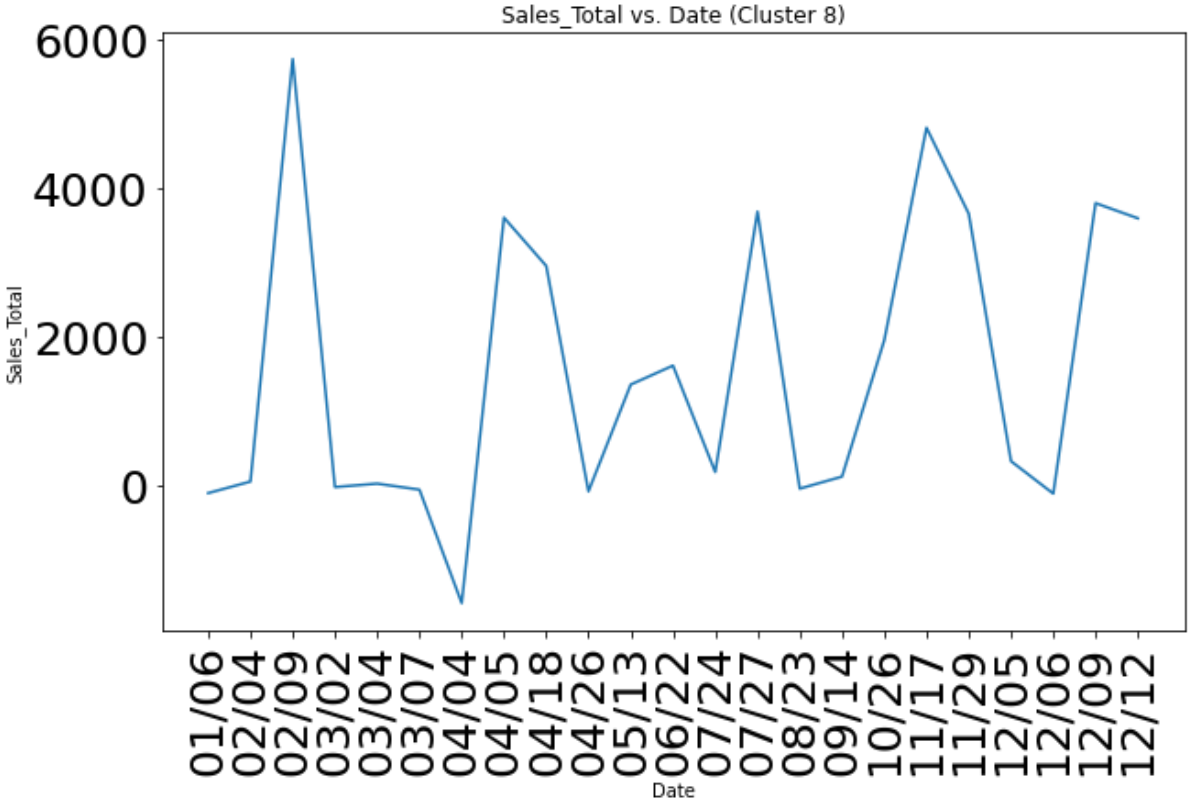
We can see maximum sales are around 10th December, 25th November and 20th September. January to September sales are around 35000 to 60000.

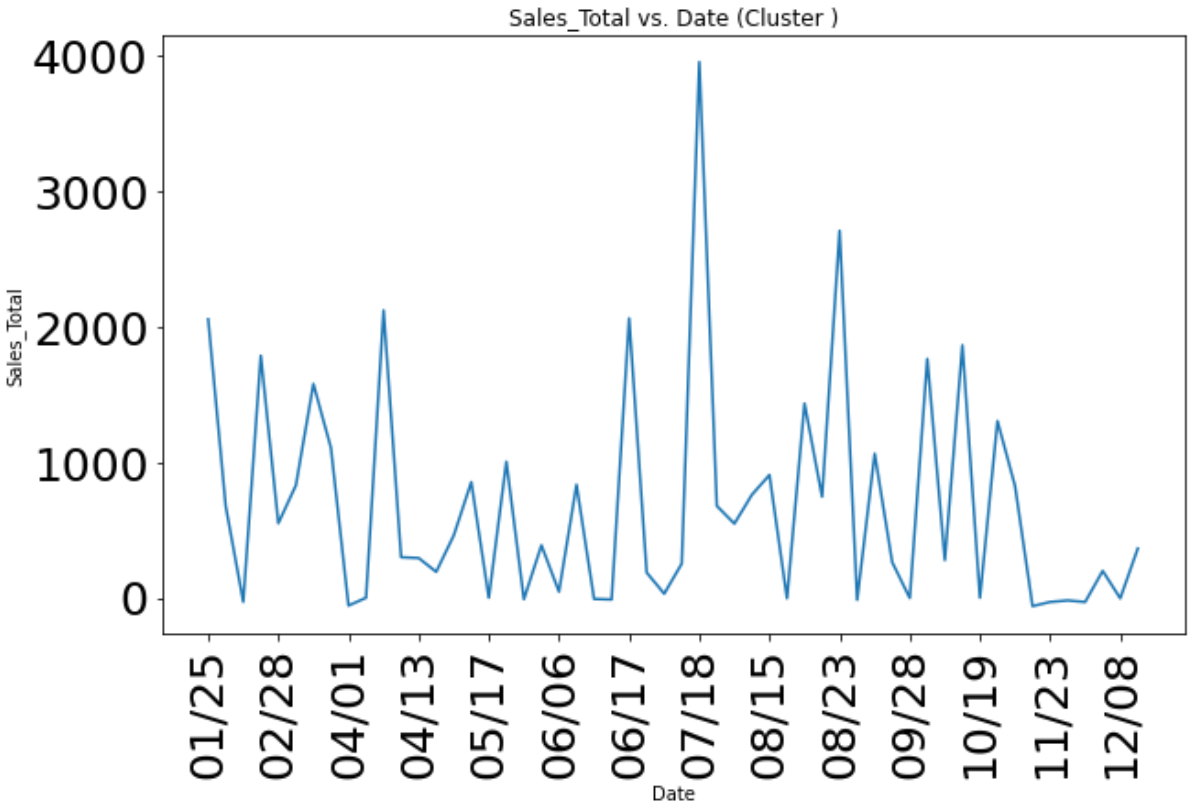
We plotted for every cluster individually as follows: --


Sales_Total vs. Date (Cluster 1)
Sales_Total
5000
4000
3000
2000
1000
0
01/12
02/07
02/16
02/21
03/03
04/13
07/14
09/01
09/22
10/07
10/19
12/02
12/15
Date


Sales_Total vs. Date (Cluster 2)
Sales_Total
30000
20000
10000
0
01/05
02/03
03/08
04/04
05/06
06/05
07/01
07/31
08/24
09/23
10/23
11/16
12/10
Date


Sales_Total vs. Date (Cluster 7)
Sales_Total
20000
15000
10000
5000
0
01/06
01/17
02/07
03/09
04/08
05/20
06/15
07/19
08/18
09/25
11/02
11/24
12/17
Date


Sales_Total vs. Date (Cluster 8)
Sales_Total
6000
4000
2000
0
01/06
02/04
02/09
03/02
03/04
03/07
04/04
04/05
04/18
04/26
05/13
06/22
07/24
07/27
08/23
09/14
10/26
11/17
11/29
12/05
12/06
12/09
12/12
Date


Sales_Total vs. Date (Cluster)
Sales_Total
4000
3000
2000
1000
0
01/25
02/28
04/01
04/13
05/17
06/06
06/17
07/18
08/15
08/23
09/28
10/19
11/23
12/08
Date