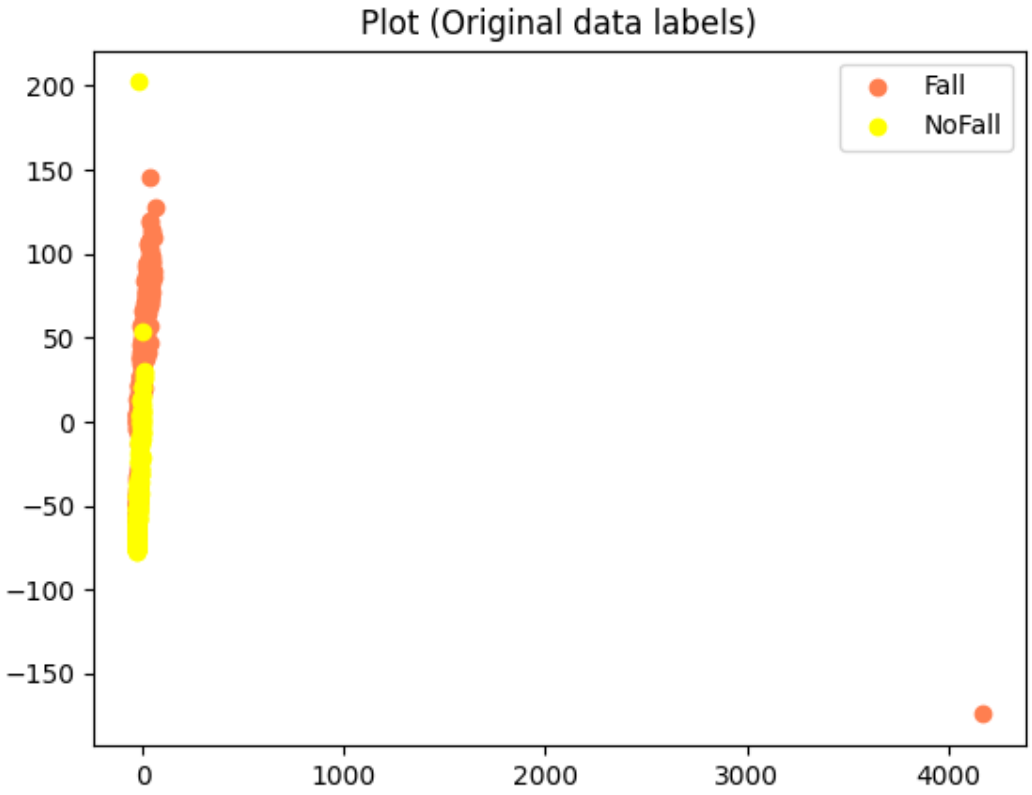
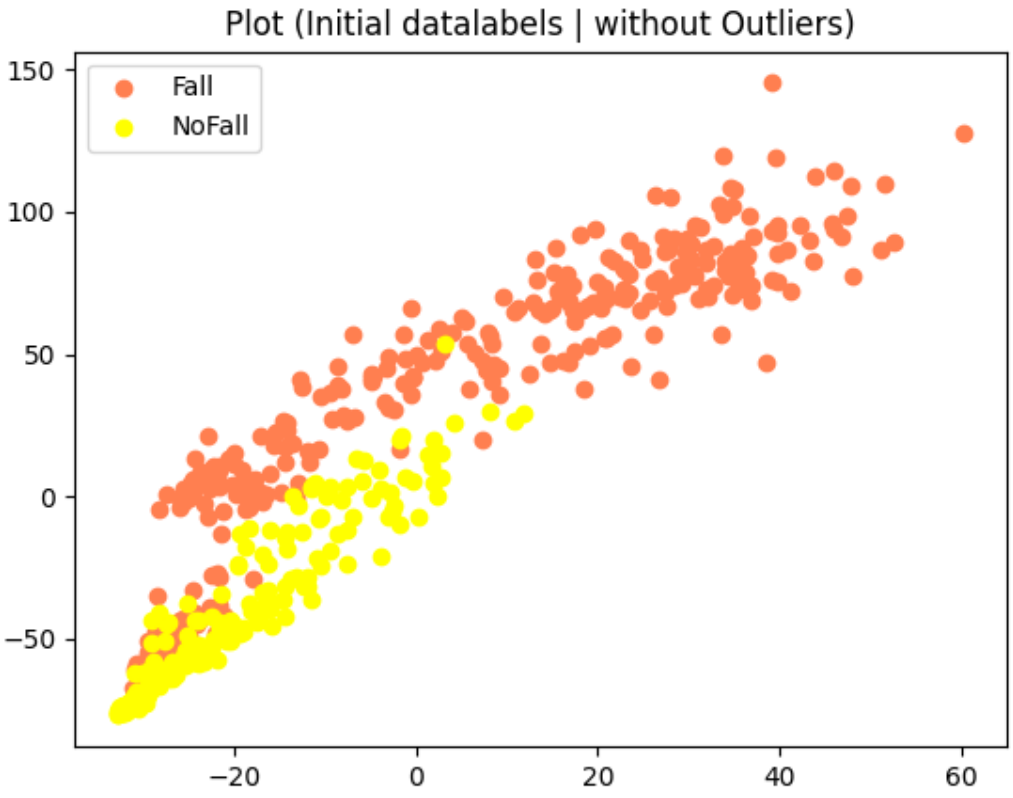
Muhammad Arham Khan
21701747
GE461
Homework 4

Part a:

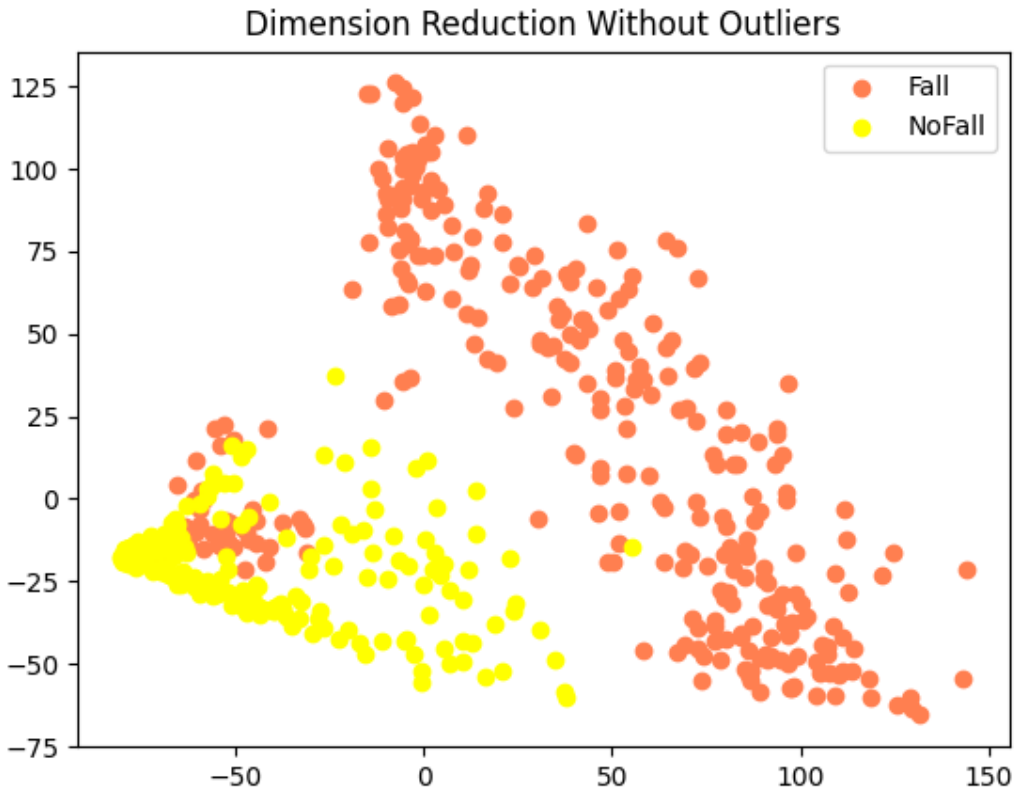


There are two outliers found in this data. One is at the end of x-axis, the other is at the end of the y-axis. Apart from these, the data shows a reasonable separability so we need to clear out these outliers for the PCA.

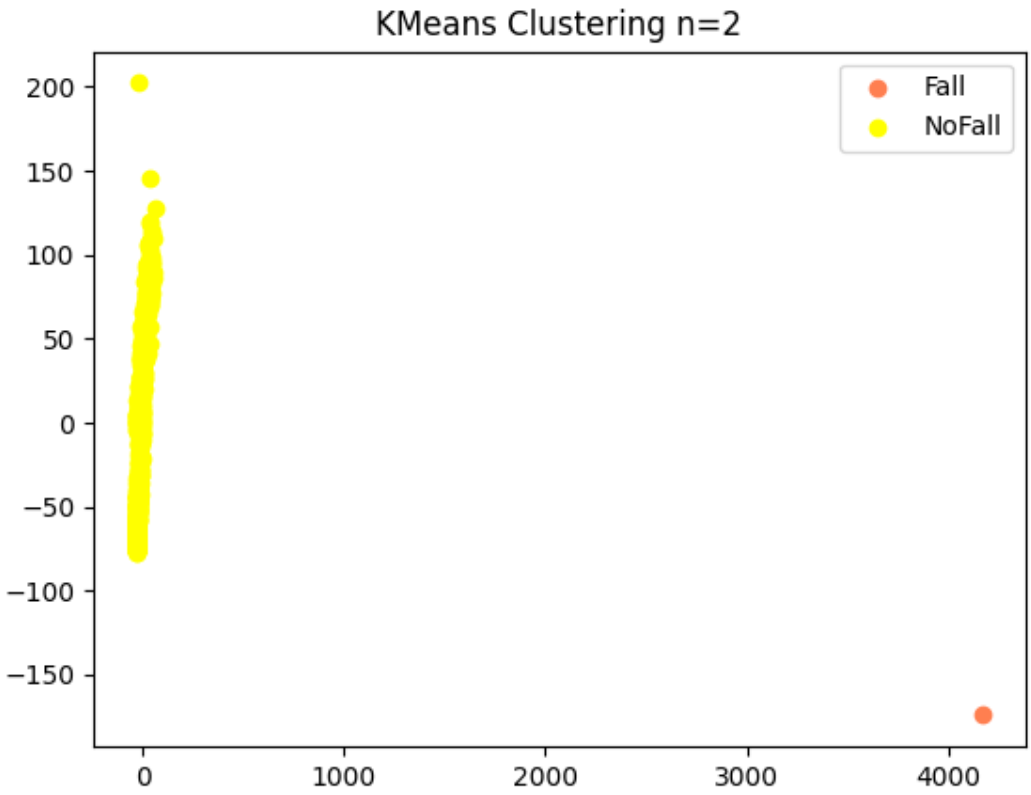
The first PCA explains 75.3% of the variance while the second one explains a mere 8.5%.



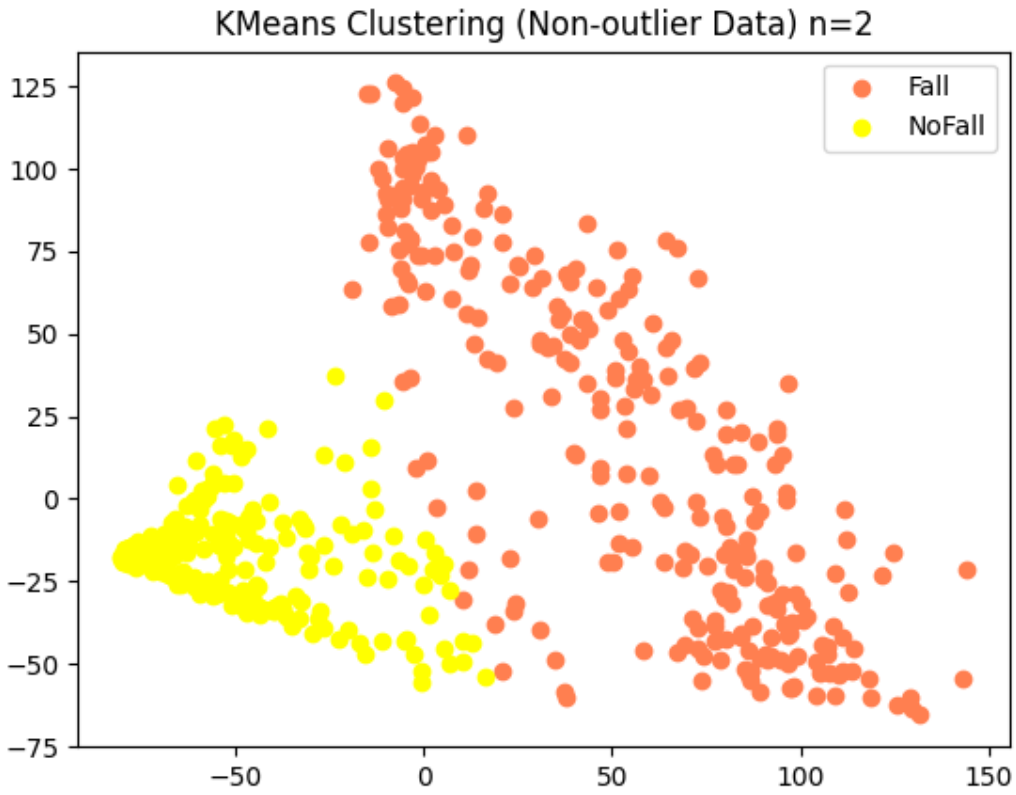
This plot shows the PCA with outliers present and it shows its unseucessful as demonstrated by the left region

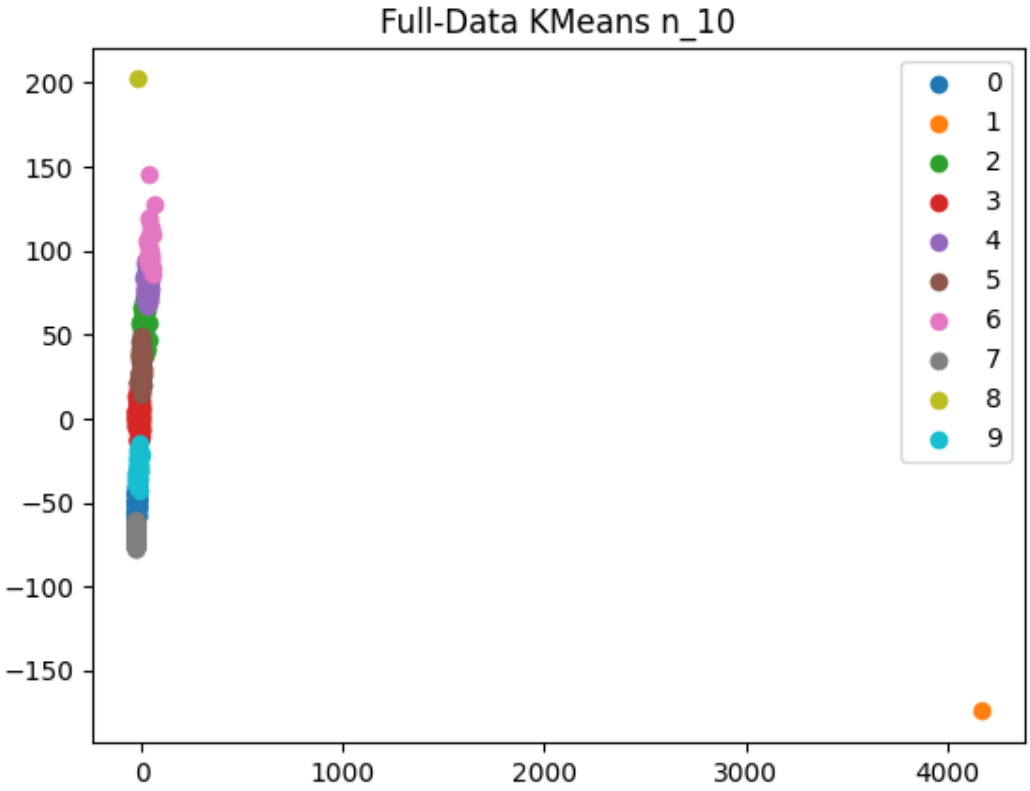


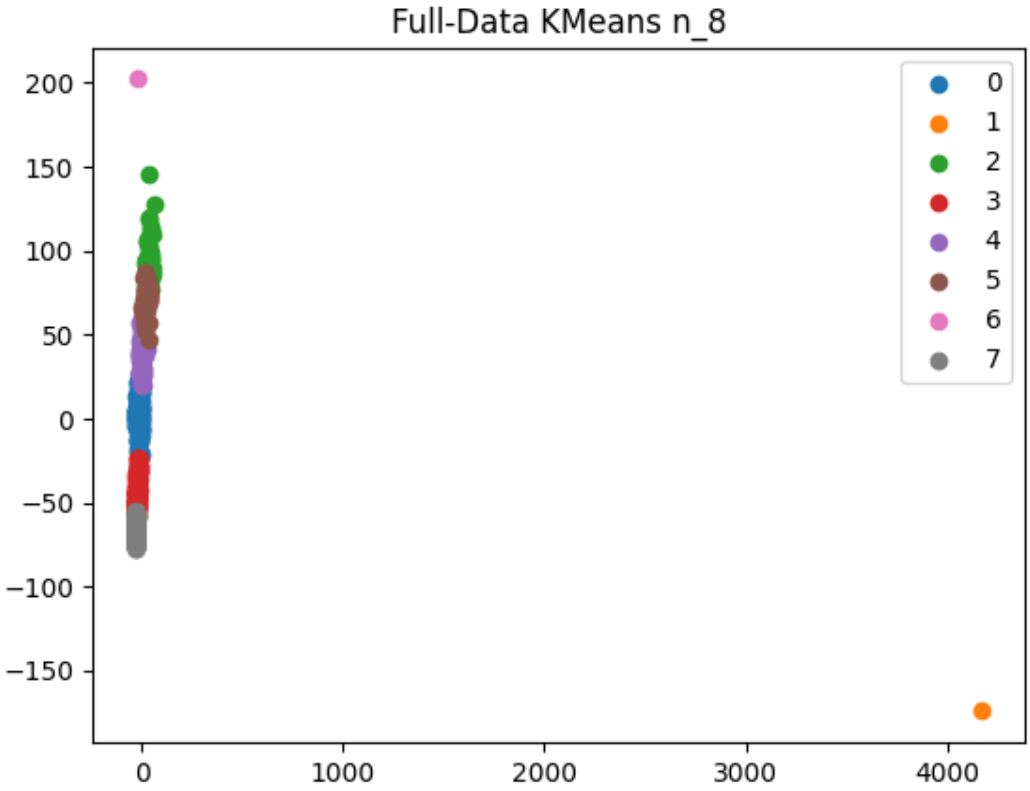
After removing the outliers rpesenting in the first image, this is the PCA outcome. This pattern presents a better outcome and represents a better idea of the unsupervised learining.



As demonstrated by the graph, the K gains are slightly unsuccessful when outliers are very different than other values. The ornage color is Fall but the results are not good as demonstrated by the graph.


KMeans Clustering (Non-outlier Data) n=2
Fall
NoFall
125
100
75
50
25
0
−25
−50
−75
−50
0
50
100
150


Full-Data KMeans n_10
200
150
100
50
0
−50
−100
−150
0
1000
2000
3000
4000
0
1
2
3
4
5
6
7
8
9


Full-Data KMeans n_8
200
150
100
50
0
−50
−100
−150
0
1000
2000
3000
4000
0
1
2
3
4
5
6
7