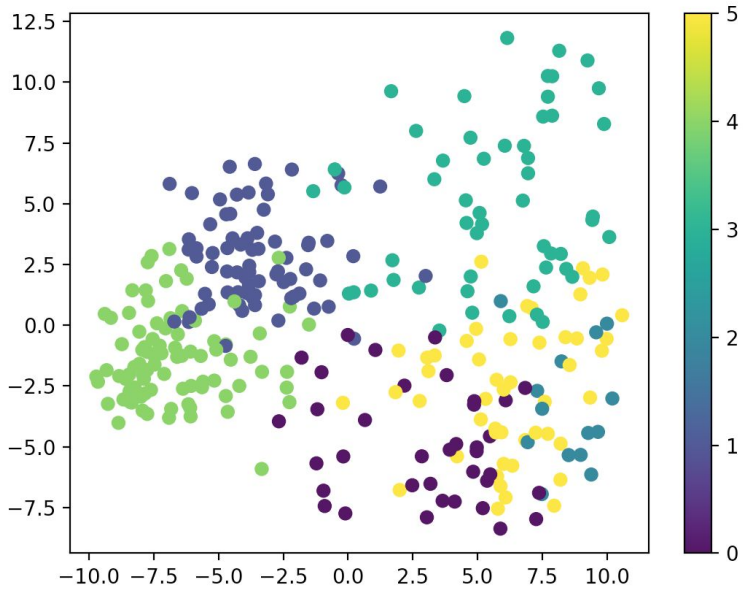
**Customer Analytics - Optional - Hw 3**

**Part (C) Visualization in Classical segmentation**




Though it is difficult to estimate what each cluster might indicate through the scatter-plot, we can still see the justification behind the clusters/segments. One thing that we must note is how each cluster seems to be in their own space. We can also say the cluster 4 and cluster 3 have less semblance than cluster 2 and cluster 3. Additionally, in our previous answer we had said that segment 5 is the best cluster for us to target. Seeing from this plot we can deduce that if we target segment 5 (yellow), we may also be able to target segment 3 and segment 0 without making any huge modifications.

**Part (D) Competion patterns:**

Looking at the data, I would say that brand A is our competitor. The way we have computed this is by using cosine similarity. We first computed the cosine similarity for each customer. Then we computed the preference of each customer for our brand and sorted it. Then using the sorted value we iterate through each as a baseline, and we deduced what is the percentage of customers who like our brand vs the other brands. At the beginning itself we can see that customers who are similar to our baseline, prefer brand A. Also, we can see in the distribution that most customers prefer brand A, but they are not similar to the people who prefer our brand. Therefore, we can say that brand A is our competitor and we need to try to convert their customers to ours.

```
brand preference
 defaultdict(<class 'int'>, {'P1': 1900, 'P3': 569, 'P2': 1842})
 ****************************************

               P1          P2          P3
count  311.000000  311.000000  311.000000
mean     0.309764    0.431333    0.258903
std      0.234041    0.309260    0.142624
min      0.007447    0.005543    0.006982
25%      0.105678    0.119950    0.152745
50%      0.232725    0.441702    0.256358
75%      0.542182    0.697961    0.354620
max      0.782914    0.980155    0.686380
Counter({'P2': 166, 'P1': 109, 'P3': 36})
```

**Part (E) Frequent occurrence of the segmentation like on Slide 35**

The primary underlying dimension of variability across people that is fairly universal is age.