Lookalike Model

Every customer exhibits a unique purchasing behavior. This can be attributed to various factors such as Region, Spending Capacity…

Here, we correlate the customers on the basis of the similarity they exhibit with each other by considering the aggregate metrics (grouped on Customer ID) such as Total Spending, Total Quantity and Unique Products.

We used Cosine Similarity to measure the similarity between the corresponding vectors, and chose the top 3 customers with the highest similarity score.

```
defaultdict(list,
            {'C0001': [('C0174', 0.9994009524324653),
              ('C0011', 0.9993929212174698),
              ('C0152', 0.9987651788245898)],
             'C0002': [('C0027', 0.9972484435911428),
              ('C0159', 0.9950099901581653),
              ('C0005', 0.9933284578274642)],
             'C0003': [('C0190', 0.9980730091178139),
              ('C0031', 0.9960594812464495),
              ('C0191', 0.9960222683025667)],
             'C0004': [('C0113', 0.998437721375097),
              ('C0104', 0.9965264907801978),
              ('C0102', 0.9964462978534467)],
             'C0005': [('C0159', 0.9996045162448027),
              ('C0007', 0.9986136741514055),
              ('C0002', 0.9933284578274642)],
             'C0006': [('C0048', 0.9978218308503074),
              ('C0187', 0.9977391266933187),
              ('C0137', 0.9963436081993936)],
             'C0007': [('C0159', 0.9990960415792354),
              ('C0005', 0.9986136741514055),
              ('C0002', 0.9931032440842279)],
             'C0008': [('C0098', 0.9914403319247044),
              ('C0049', 0.9913163451672555),
              ('C0194', 0.9867668486247554)],
...
              ('C0010', 0.9965956640371921),
              ('C0073', 0.9957715179681133)],
             'C0200': [('C0022', 0.9973708928483915),
              ('C0101', 0.9913195242728183),
              ('C0145', 0.9910652712119843)]})
```