Lookalike report

Print(lookalike_df)

	CustomerID	Lookalikes
0	C0001	[(C0148, 1.0), (C0152, 1.0), (C0174, 1.0)]
1	C0002	[(C0159, 0.99), (C0134, 0.989), (C0133, 0.734)]
2	C0003	[(C0031, 1.0), (C0158, 1.0), (C0129, 0.988)]
3	C0004	[(C0012, 0.989), (C0001, 0.881), (C0148, 0.881)]
4	C0005	[(C0007, 1.0), (C0140, 0.991), (C0095, 0.746)]
5	C0006	[(C0025, 1.0), (C0076, 0.985), (C0187, 0.985)]
6	C0007	[(C0005, 1.0), (C0140, 0.991), (C0095, 0.746)]
7	C0008	[(C0109, 0.996), (C0059, 0.975), (C0098, 0.975)]
8	C0009	[(C0066, 1.0), (C0198, 0.991), (C0040, 0.743)]
9	C0010	[(C0061, 0.987), (C0132, 0.987), (C0077, 0.7)]
10	C0011	[(C0107, 1.0), (C0192, 0.985), (C0126, 0.985)]
11	C0012	[(C0004, 0.989), (C0001, 0.941), (C0148, 0.941)]
12	C0013	[(C0087, 1.0), (C0155, 1.0), (C0099, 0.989)]
13	C0014	[(C0060, 1.0), (C0089, 0.993), (C0151, 0.957)]
14	C0015	[(C0131, 0.99), (C0038, 0.891), (C0160, 0.891)]
15	C0016	[(C0183, 1.0), (C0029, 0.986), (C0094, 0.986)]
16	C0017	[(C0075, 0.993), (C0041, 0.991), (C0051, 0.991)]
17	C0018	[(C0157, 1.0), (C0035, 0.986), (C0047, 0.947)]
18	C0019	[(C0023, 1.0), (C0070, 0.941), (C0191, 0.667)]
19	C0020	[(C0130, 0.956), (C0120, 0.92), (C0198, 0.669)]
20	C0021	[(C0028, 1.0), (C0054, 1.0), (C0145, 1.0)]

- After plotting the histogram we come to conclusion that:

A wide spread suggests variability in the scores, which can be good if the variability aligns with meaningful customer distinctions

```
Mean similarity score: -0.005
Standard deviation of similarity scores: 0.370
```

#Mean being close to zero: Similarity scores are evenly distributed around zero, indicating that many pairs are dissimilar.

#High standard deviation: There's a large variance in similarity scores, meaning some customers are highly similar, while others are very dissimilar.