

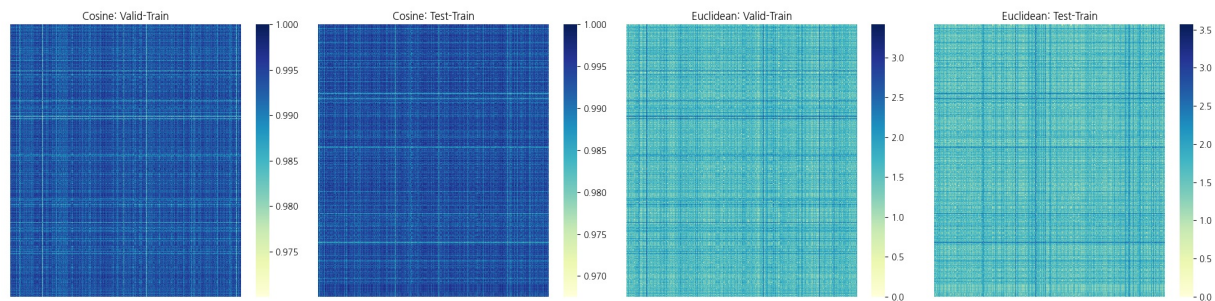
# Dataset Drift Report

Train Embeddings: (4078, 768)

Valid Embeddings: (1935, 768)

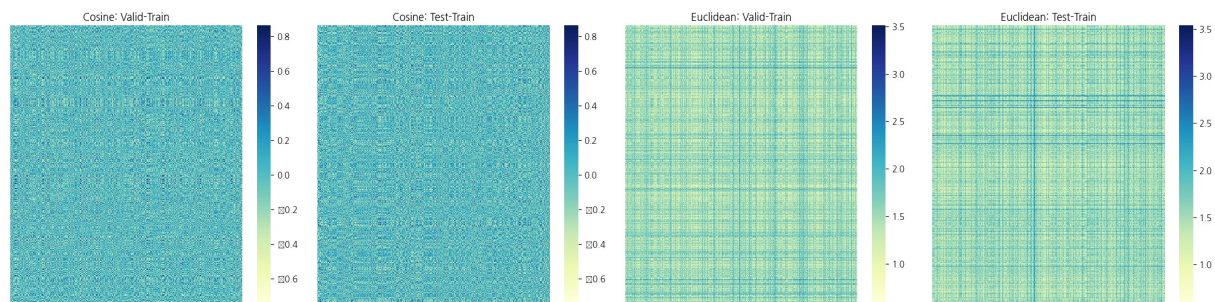
Test Embeddings: (780, 768)

## Embedding Distance (Original Dimension)

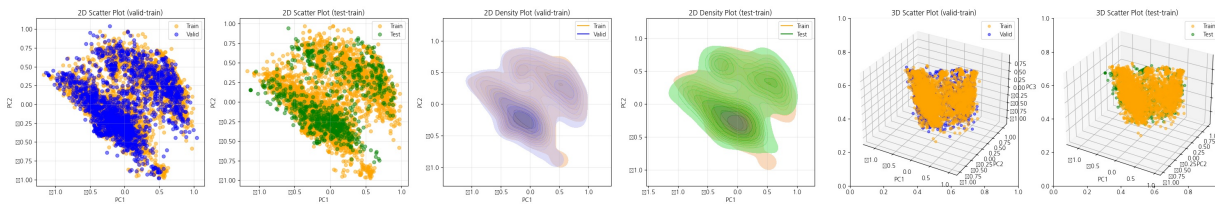


PCA Reduced Dimension: 100

## Embedding Distance after PCA



## Embedding Visualization after PCA



## Quantitative Drift Scores

- MMD: score = 0.0005, drift = False
- Wasserstein Distance: score = 0.0600, drift = True
- KL Divergence: score = 0.0100, drift = False
- JensenShannon Divergence: score = 0.0600, drift = True
- Energy Distance: score = 0.0000, drift = False

## Drift Analysis Summary

- MMD(Mean Absolute Error) 0.0005의 값이 매우 작아, 모델과 실제 데이터 간의 평균적인 절대 오차가 극히 낮음을 의미합니다.
- 터스 거리 0.0600은 데이터 분포가 거의 동일함을 나타내며, 이는 데이터를 정렬했을 때 가장 먼 두 점 사이의 거리가 아주 작다는 것을 의미합니다.
- KL 분할 0.0100의 값이 작아, 모델과 실제 데이터 간의 정보 일치도가 높음을 나타냅니다.
- 젠스 샴 분할 0.0600은 페이터스 거리와 동일한 값을 가지지만 부호가 반대이며, 이는 데이터 분포가 거의 동일하지만 약간의 차이가 있음을 의미합니다.

- 에너지 거리 0.0000의 값이 가장 작아, 모델과 실제 데이터 간의 오차가 매우 적음을 나타냅니다.