



환자군 별 병원 물류 수요 예측 분석 (In Progress)

1. 개요

- 문제 정의:
 - 병원의 소모품은 재고 관리가 제대로 되지 않고, 수요 예측에 따라서 품목을 주문하는 것이 아닌, 간호사 개인의 경험에 따라서 주문하는 경향이 짙음
 - 병원의 환자 수요를 예측하고, 환자들의 방문 트렌드에 따라서 변동 자동 주문 실시
- 예상 결과물:
 - 처방전에 따른 환자 집단군.
 - 환자 집단군의 방문 빈도 시계열 (명/일), 및 시계열 예측 모델
 - 환자 집단군 방문량 예측에 따른 병원 소모품 수요 예측 모델
- 주요 역할:
 - 모델 개발

2. 데이터 소개

- OO, XX 병원들의 환자 처방 데이터 (대외비)

3. 데이터 클리닝

- 환자 ID, 방문 날짜, 성별, 주상병(Principal Disease), 부상병(Secondary Disease), 처방 구분, 처방명이 들어있는 데이터
 - 이를 모두 모아서 환자의 차트(Chart)를 구성.

예시

환자 ID: 1856

방문일(퇴원일): 20XX-XX-XX

입원/퇴원 여부: 외래

나이: XX

Sex: X

주상병(Primary Diagnoses): 만성 인두염 (J312)

부상병(Secondary Diagnoses): 만성 후두염, 만성 편도염 (J370, J350)

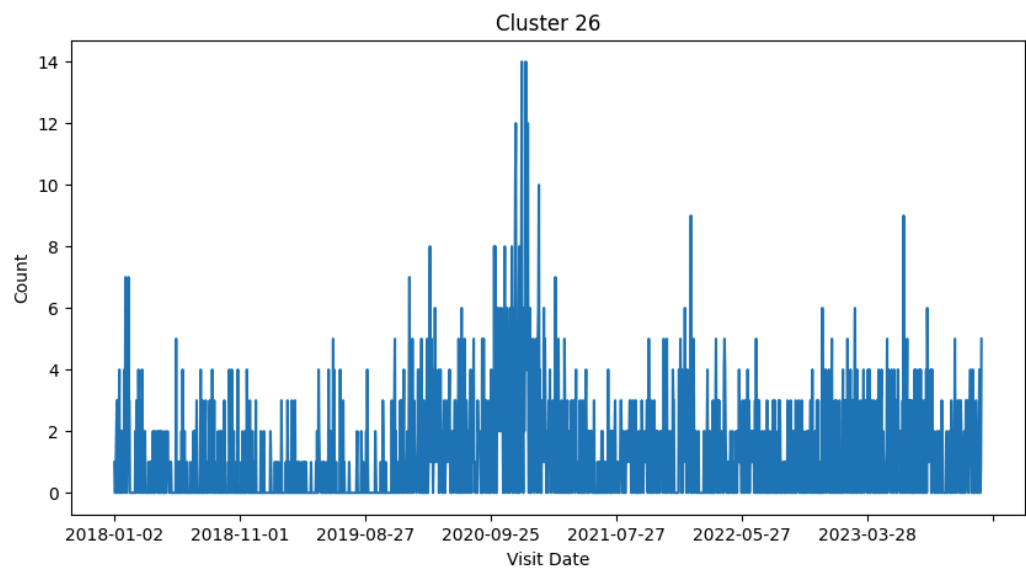
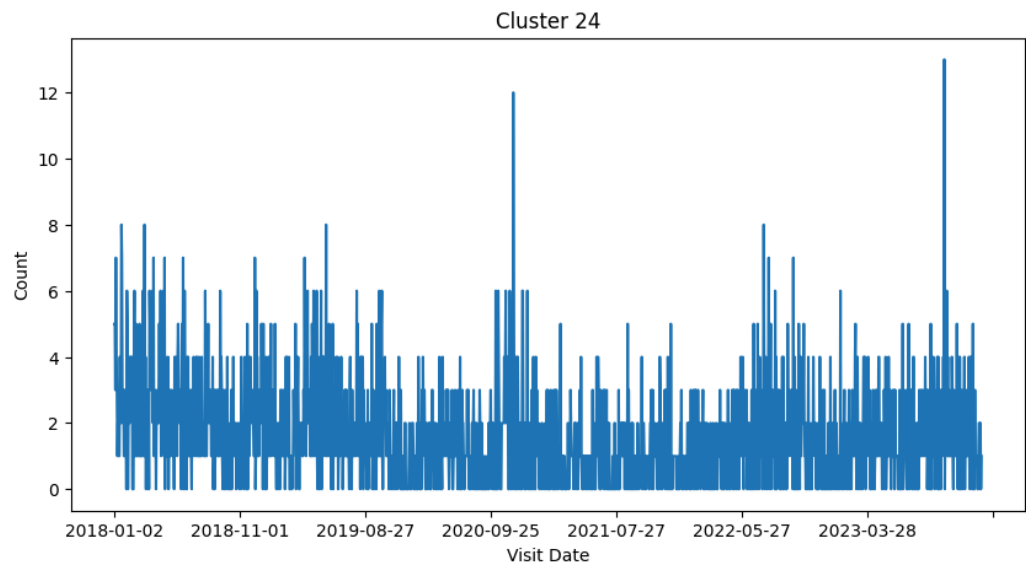
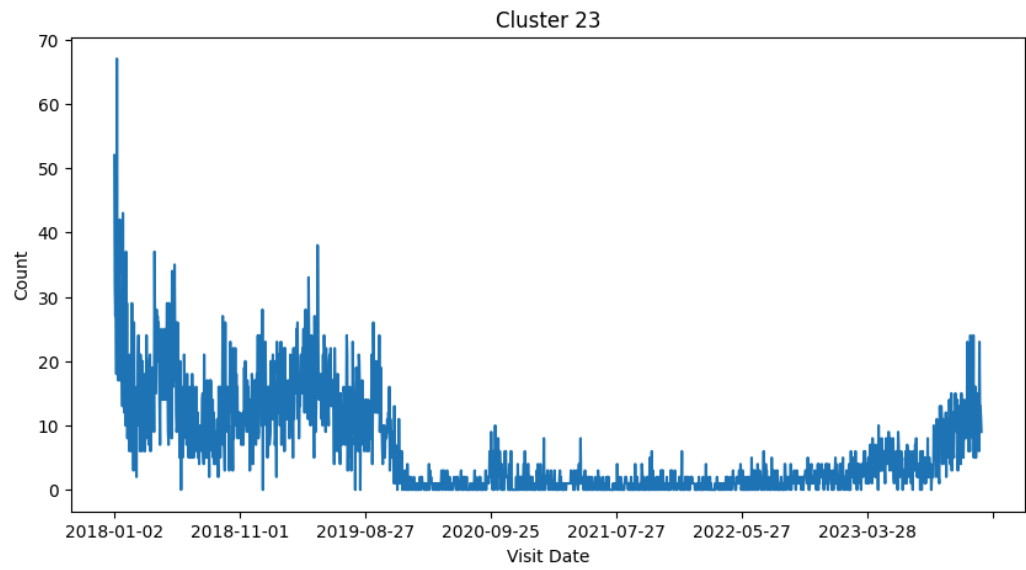
Department EN

- 내복약: 록솔정(암브록솔염산염), 세토펜정 325mg[삼아약품], 스티렌투엑스
- 내시경검사: 후두경검사
- 주사약제: 트리돌주50밀리그램(염산트라마돌) (주)유한양행

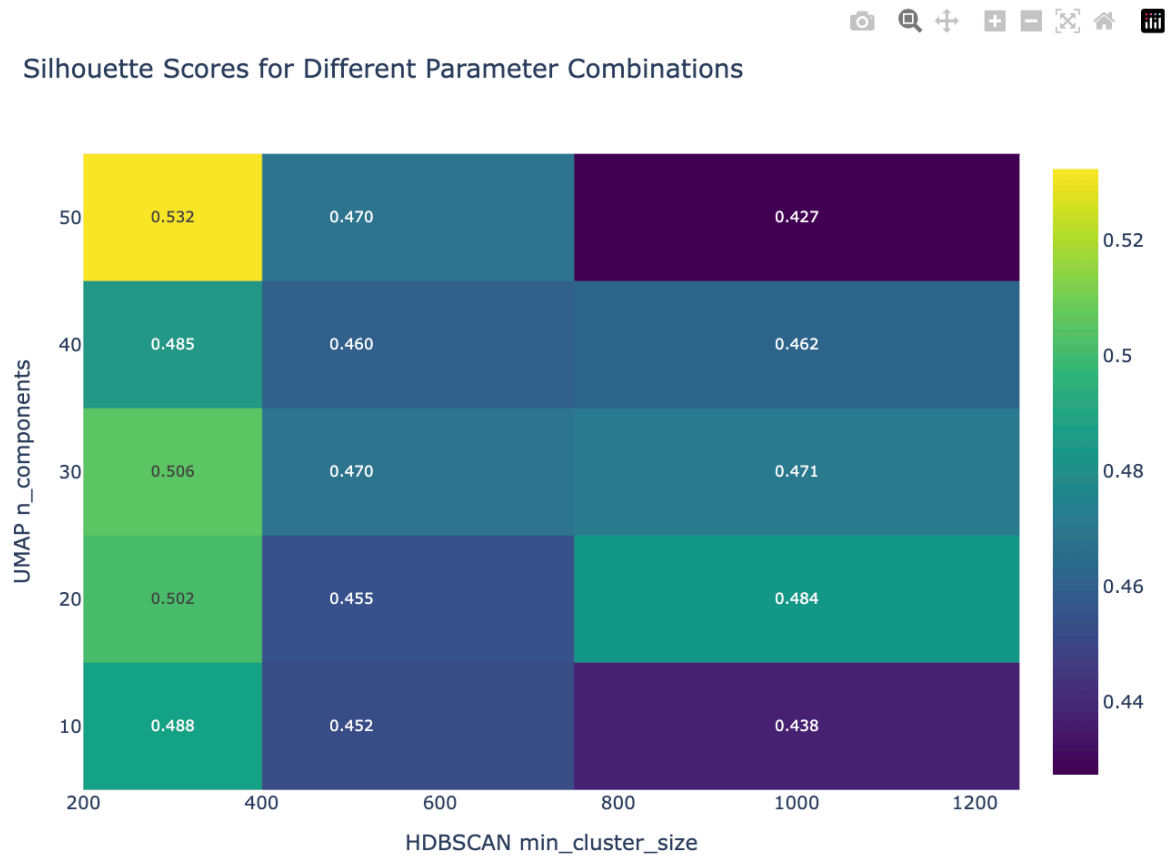
- OpenAI Embedding (text-embedding-ada-002)를 사용하여 임베딩 후 벡터 데이터베이스에 저장 ([Qdrant](#))
- 2018 ~ 2023년 데이터까지 클러스터링 모델 적합(fit)에 사용
 - 2024년 데이터는 fit된 모델을 이용하여 Clustering. (훈련자료와 분리)

4. 머신러닝 모델링 (In Progress)

- 환자군을 분류하기 위해 Clustering 모델 사용
 - 후보
 - HDBSCAN: DBSCAN의 업그레이드 버전
 - Hierarchical Clustering 과정을 함께 이용해서 과거 고정된 EPS를 기준으로 클러스터링을 하던 DBSCAN과는 다르게 군집의 밀도가 달라도 클러스터링 가능
 - Feature Selection: UMAP 사용
 - 1500여개의 Embedding을 묶기 위해서 UMAP 사용
 - PCA와는 다르게 Non-linear.

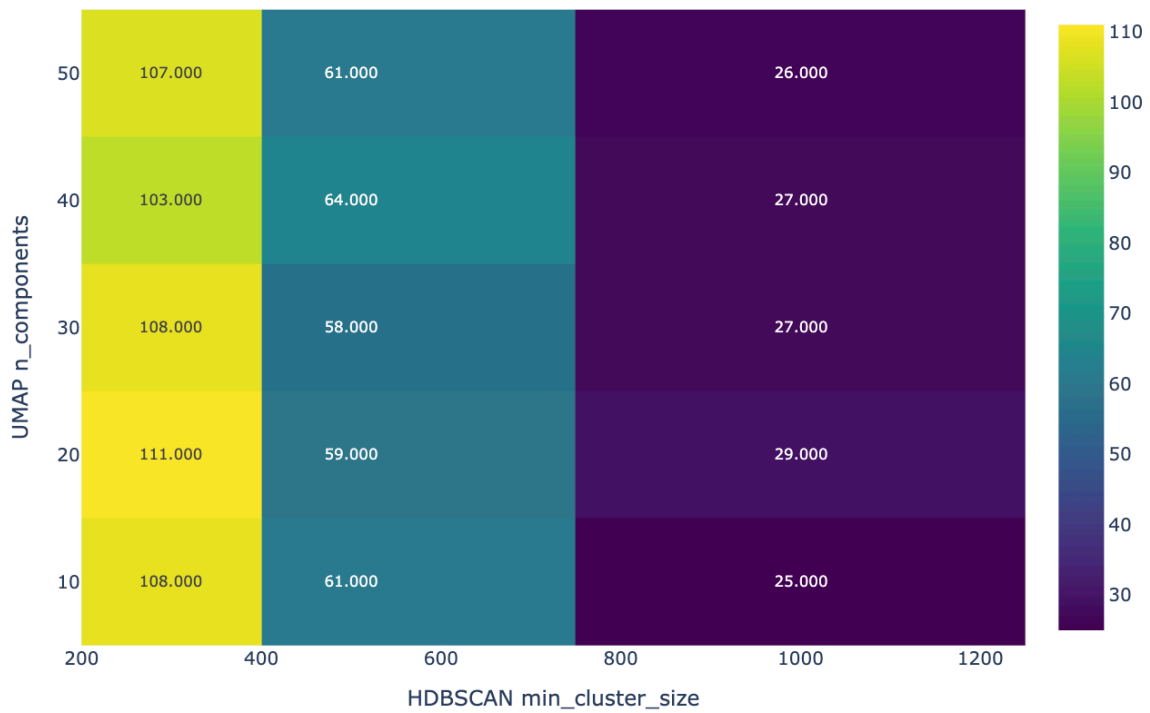


- 실루엣 점수로 판단한 HDBSCAN & UMAP 최적 하이퍼파라미터 찾기

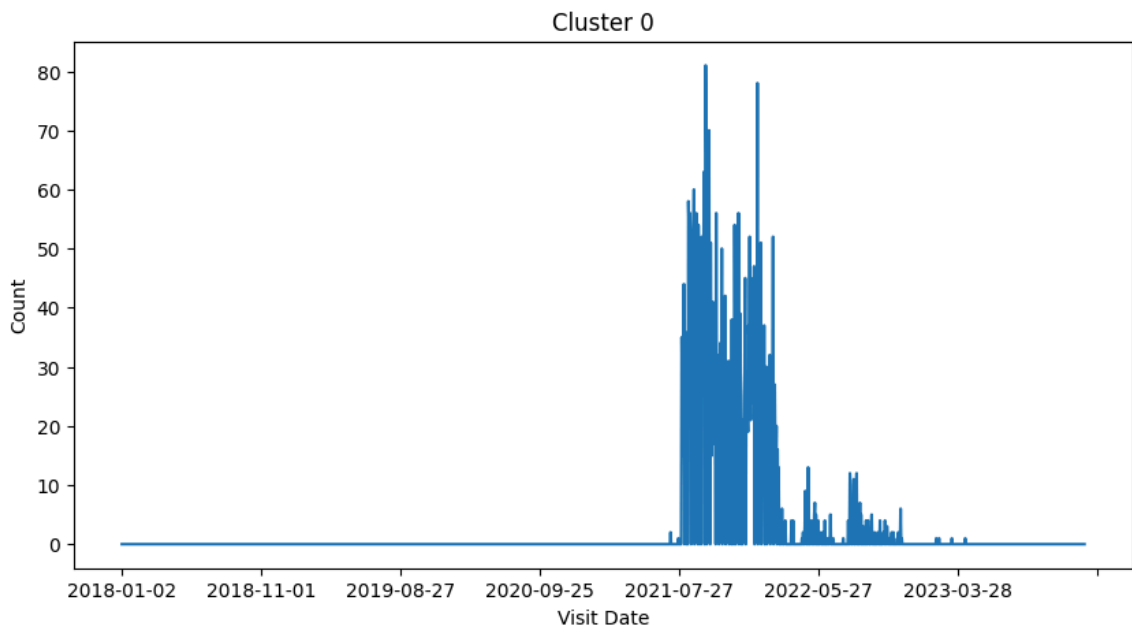


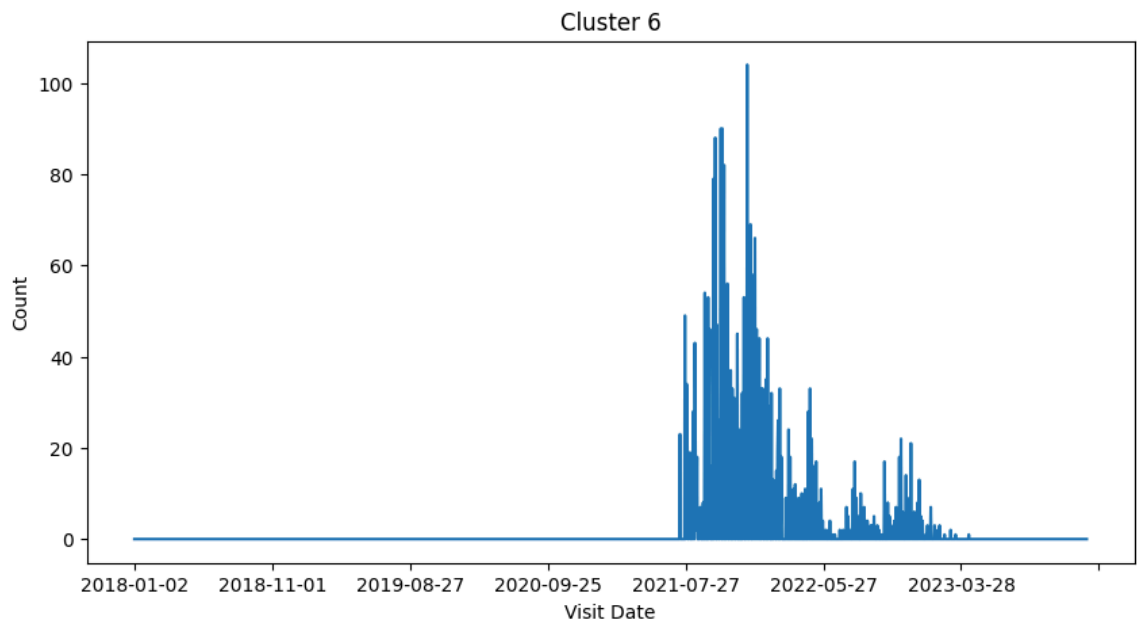
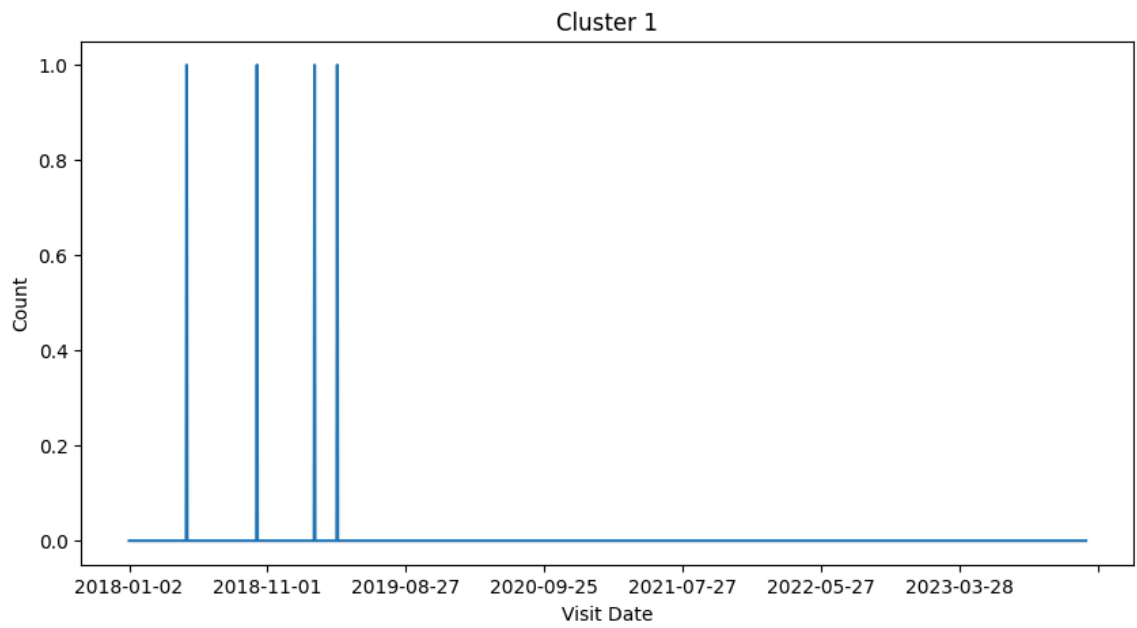
- HDBSCAN & UMAP 하이퍼파라미터 별 클러스터 개수

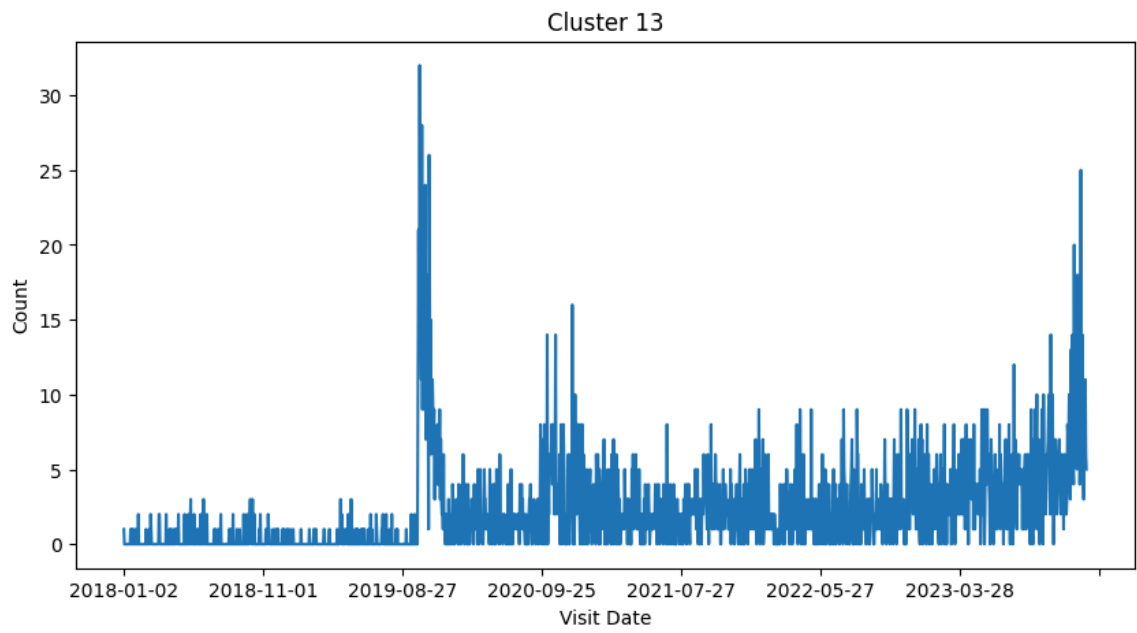
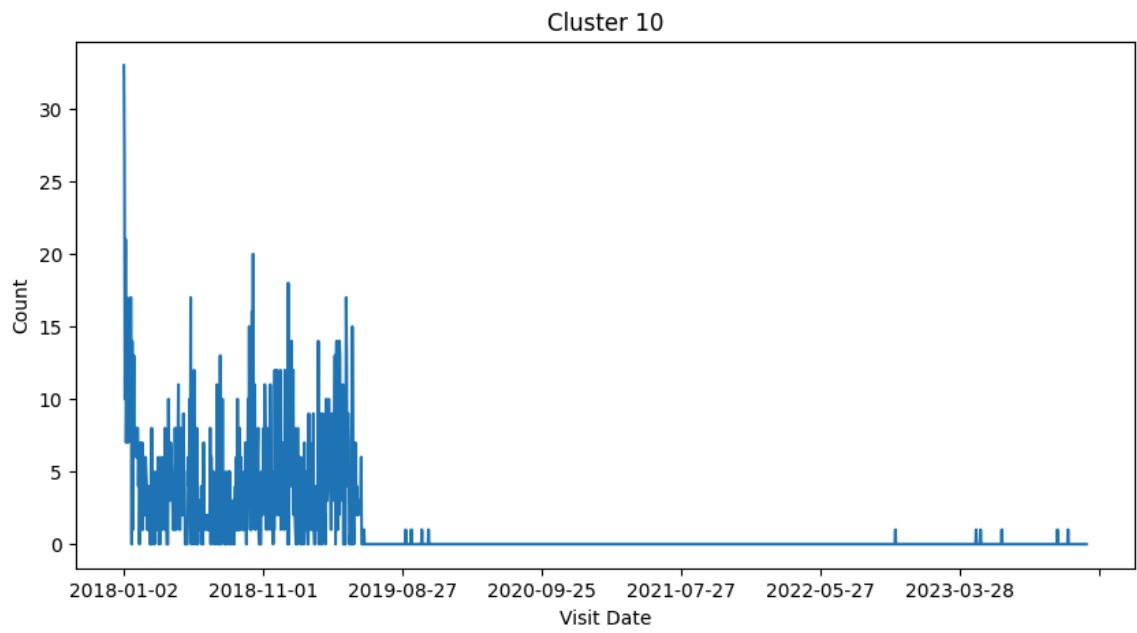
Silhouette Scores for Different Parameter Combinations

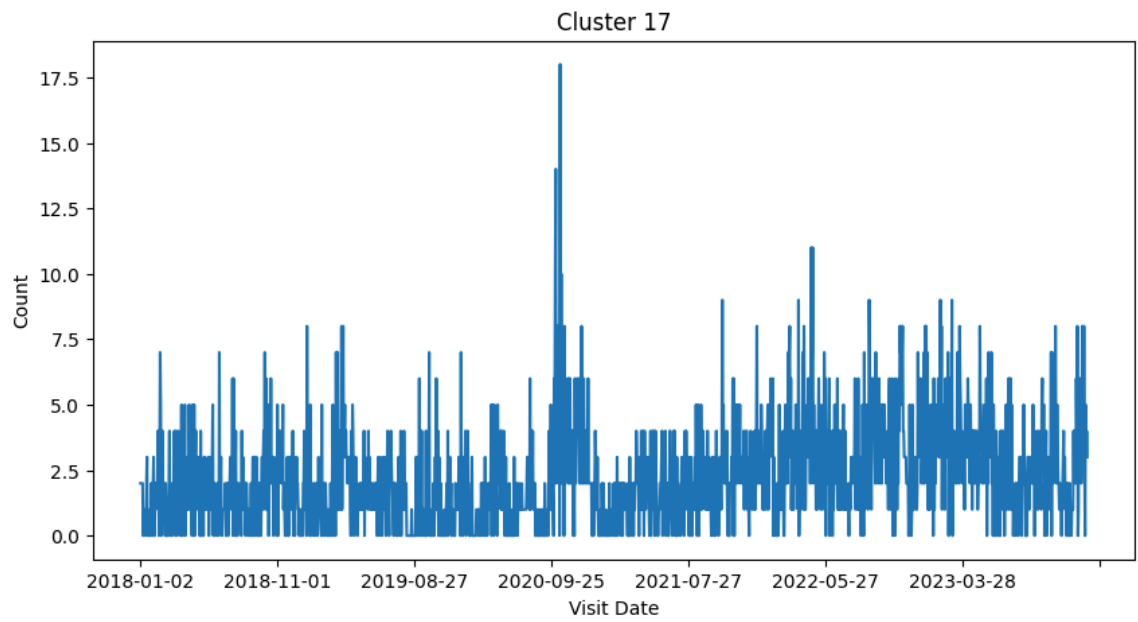


▼ 20개의 클러스터를 유지하면서 Silhouette Score 가 높은 UMAP (20) HDBSCAN (1000) 버전









5. 결론
