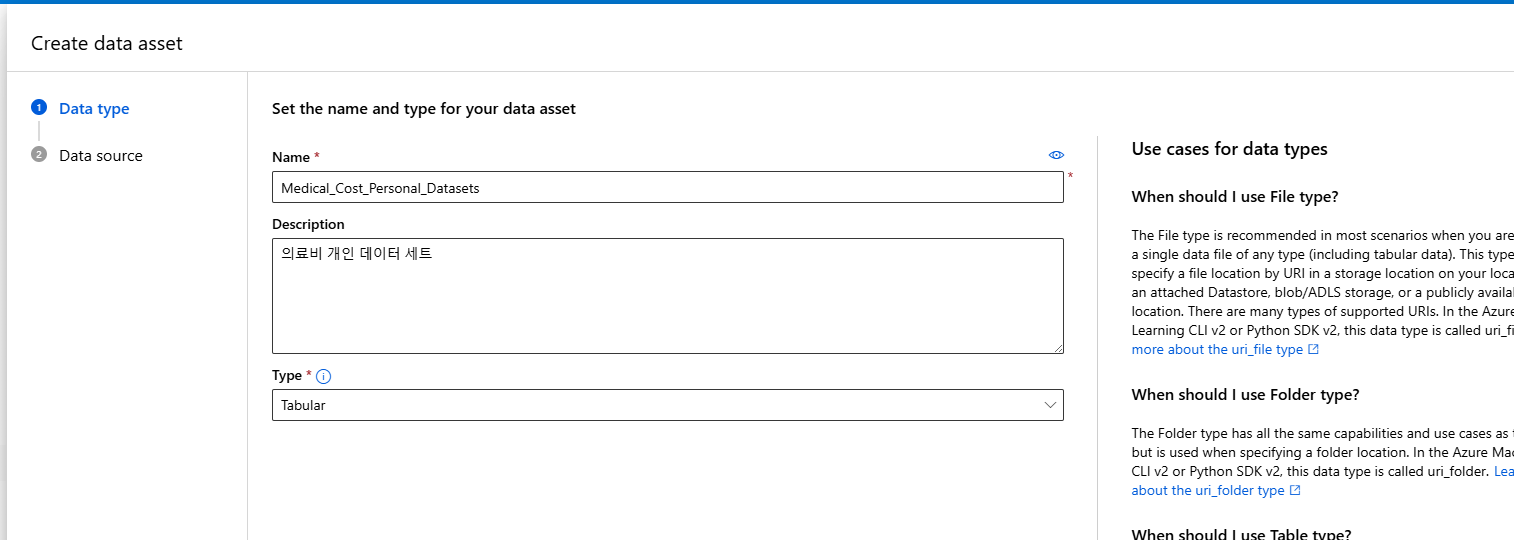
# <오전>

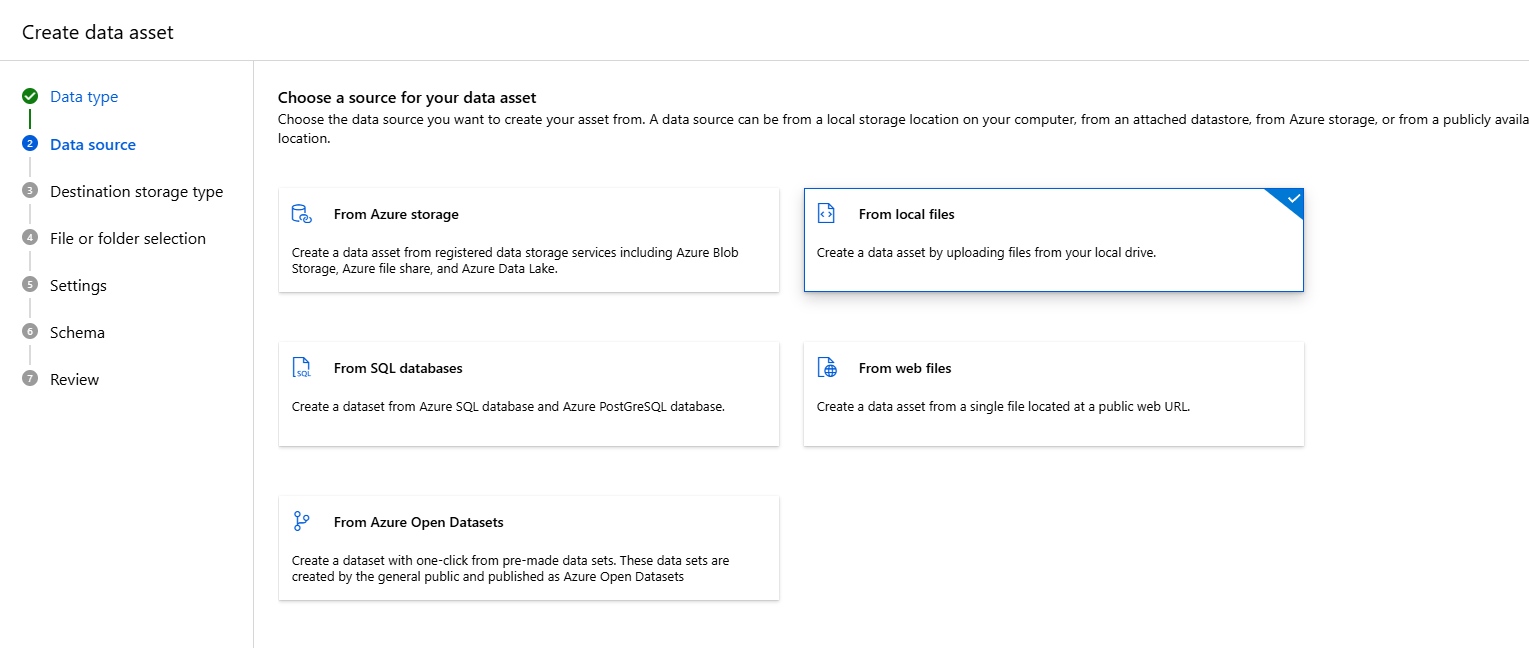
## [환경]

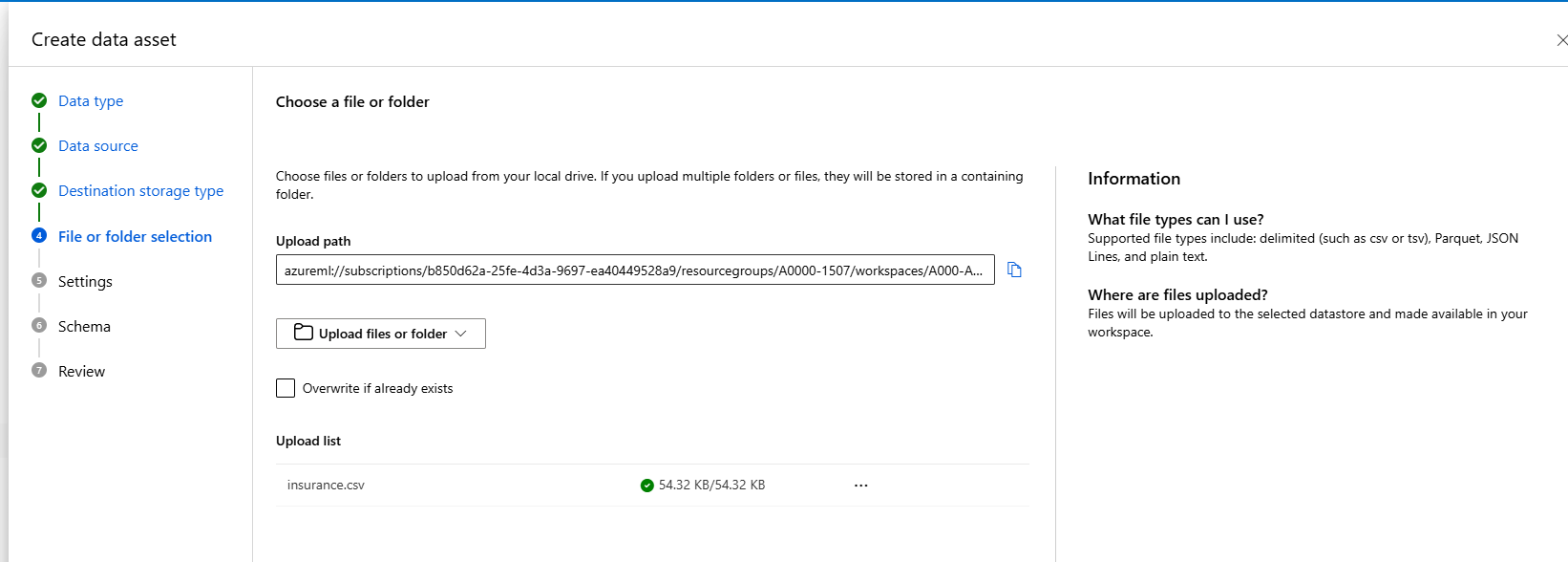
* Azure
* Kaggle(캐글)에서 csv 파일 다운로드

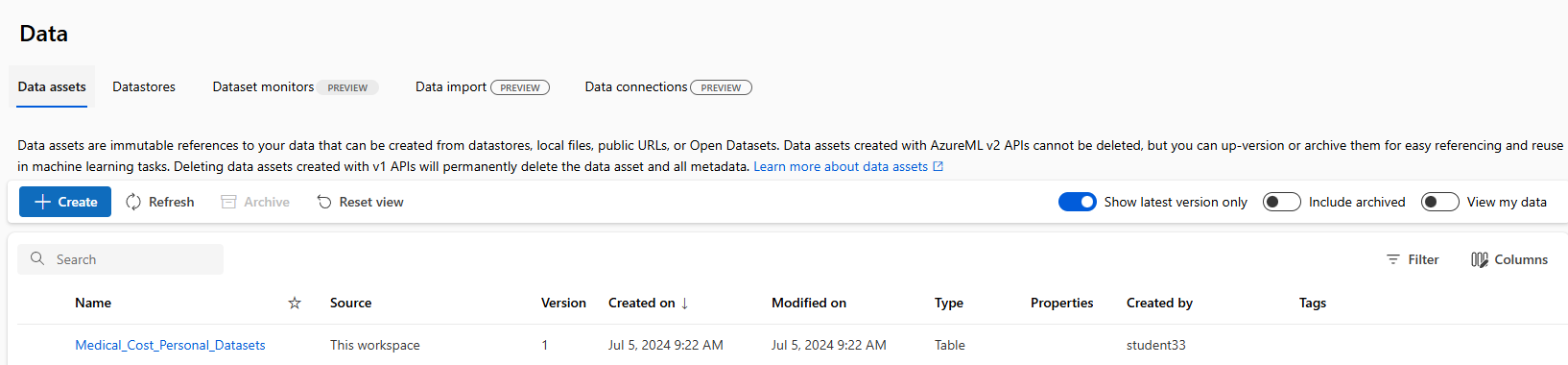
## [개인 실습 - Medical\_Cost\_Personal\_Datasets]

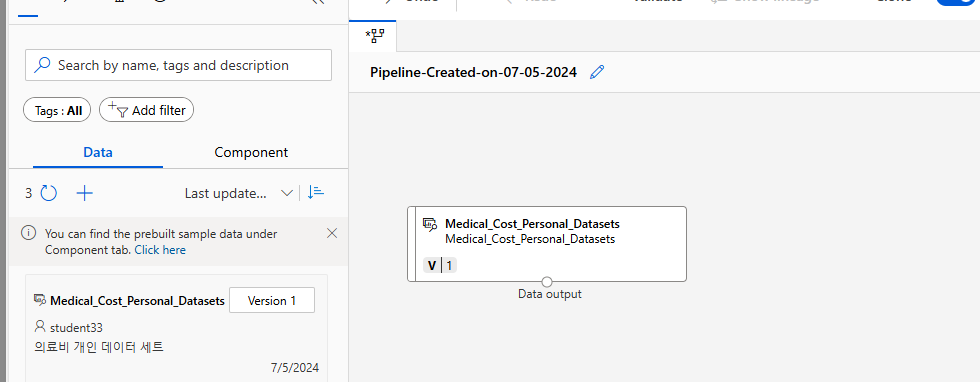
Kaggle에서 데이터 수집 (<https://www.kaggle.com/datasets/mirichoi0218/insurance>)

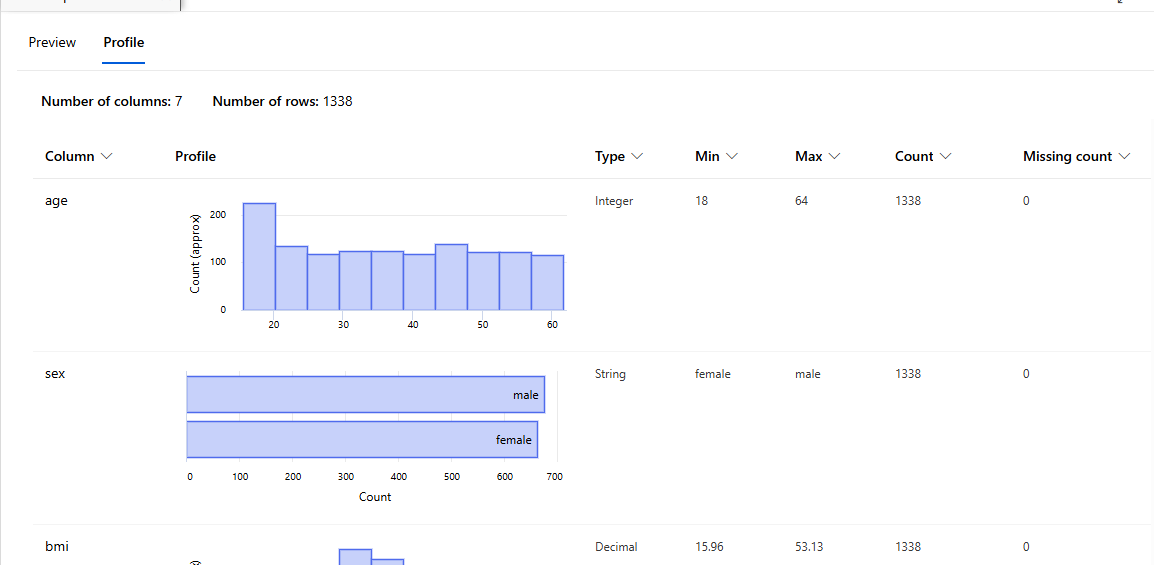


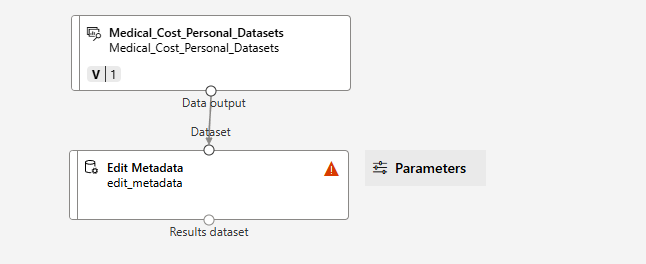


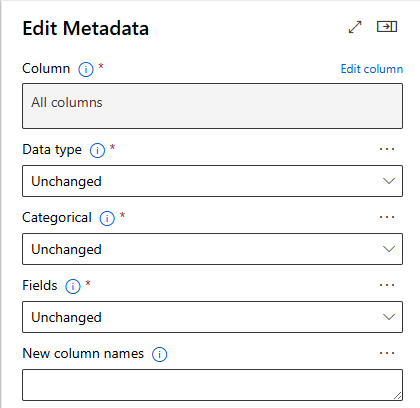


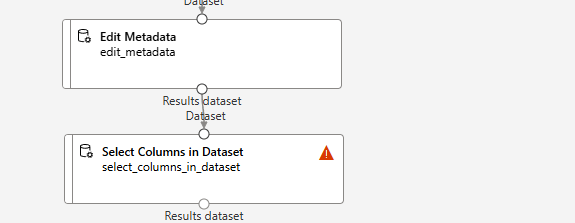




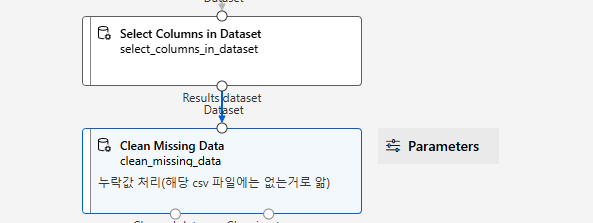


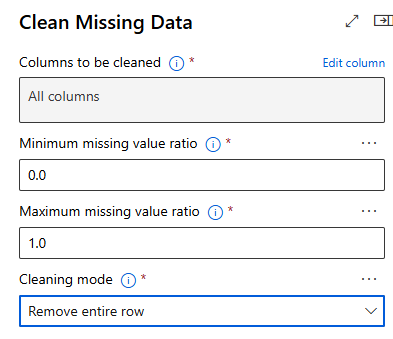


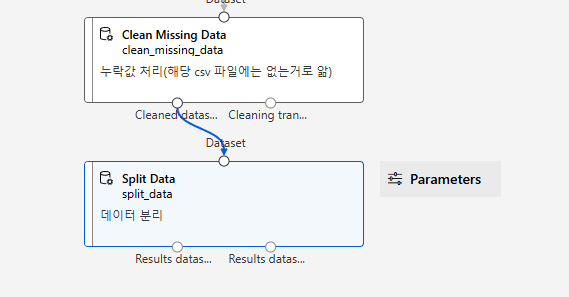


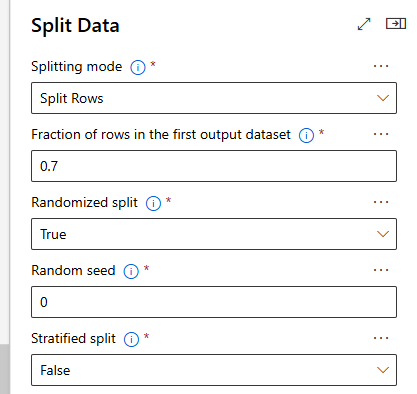


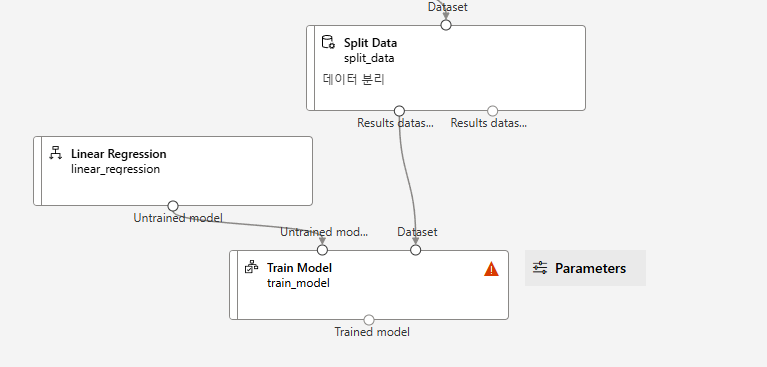


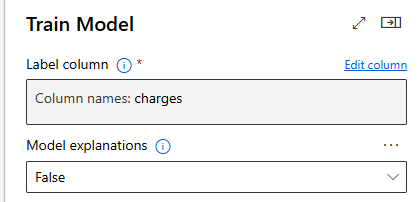


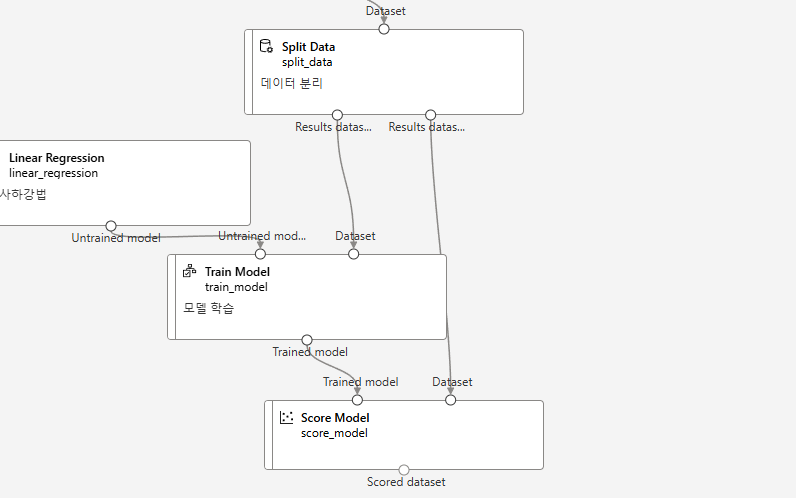


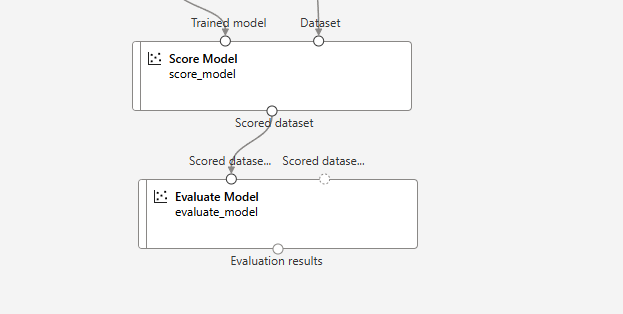




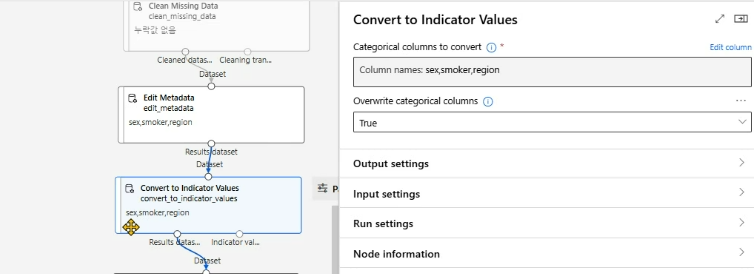






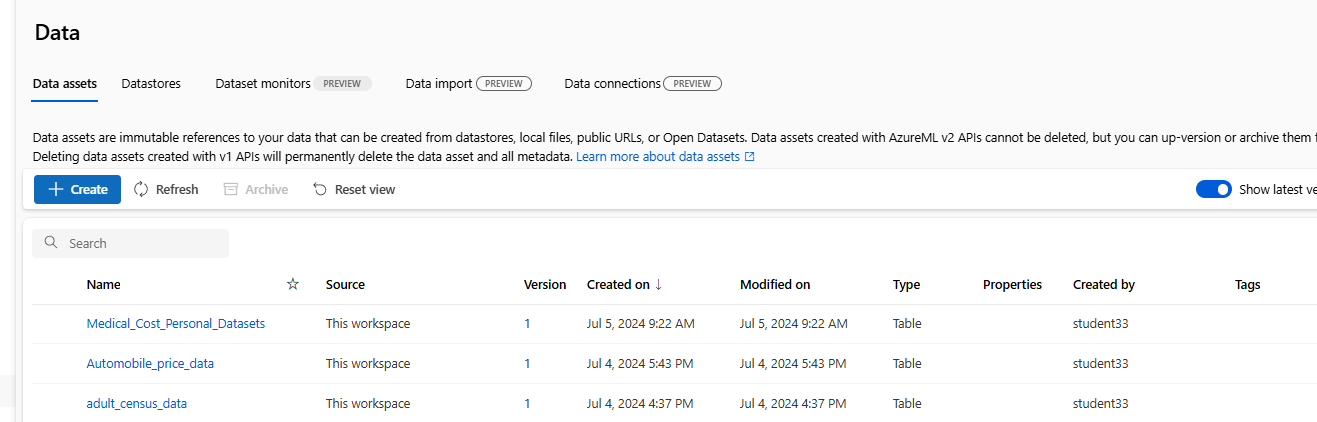


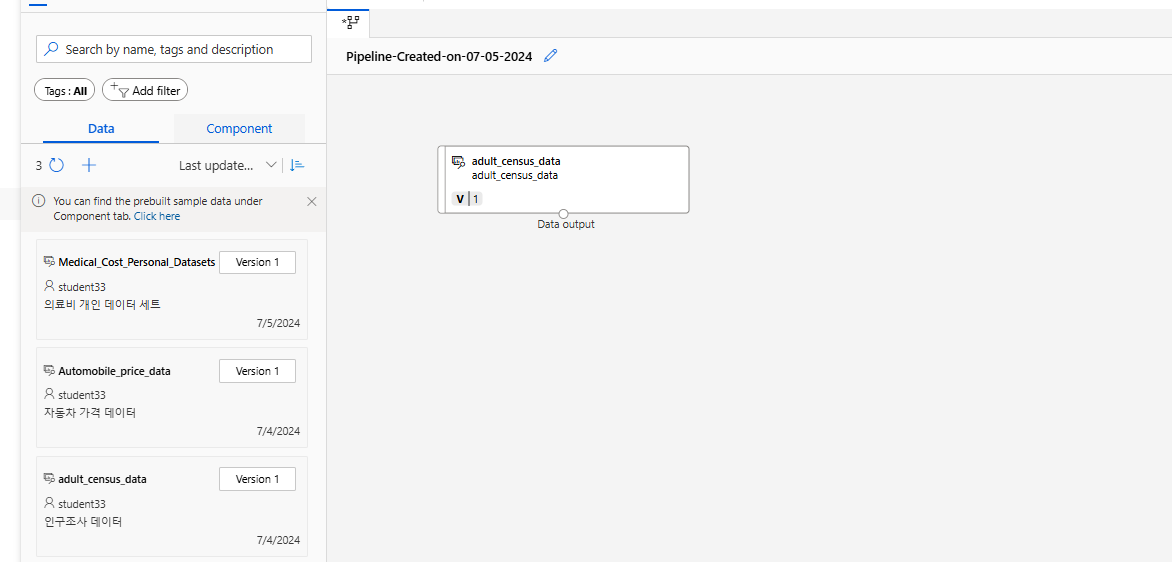


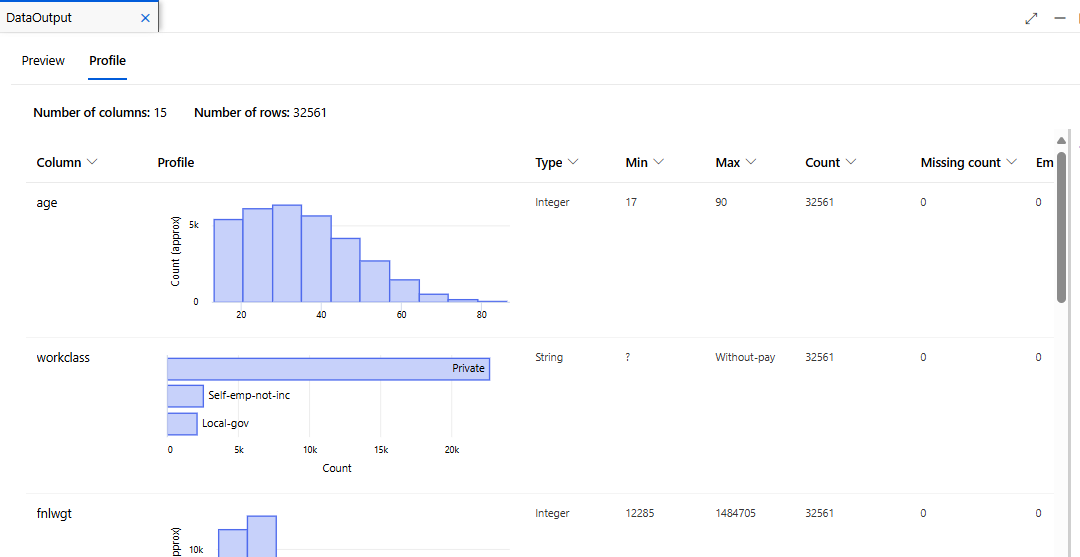


## [실습 - 분류MLD\_개인수입예측\_카이제곱FS]

Data는 기존의 adult\_census\_data 사용

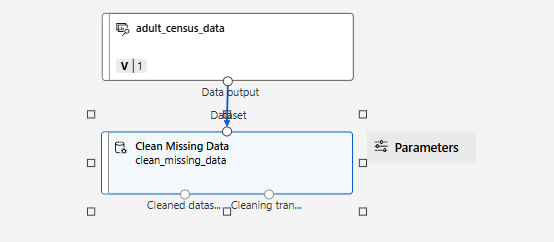


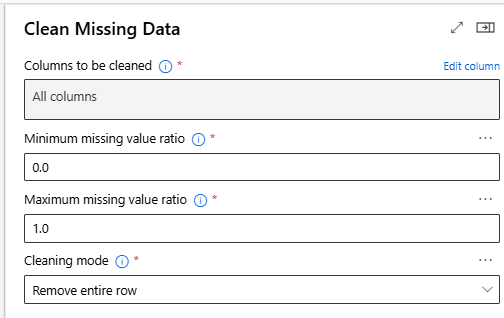




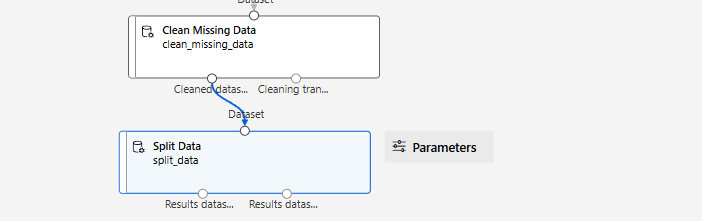
### 데이터 준비

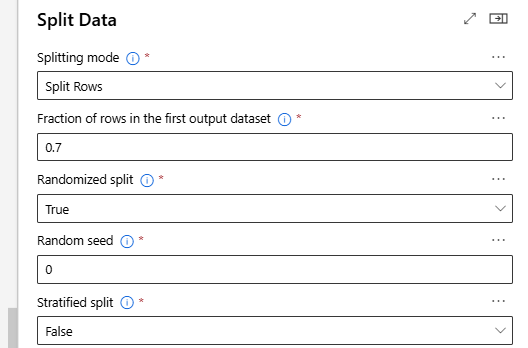
#### 누락값 처리



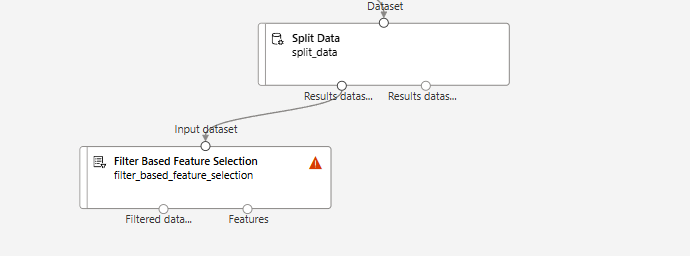


#### 데이터 분리

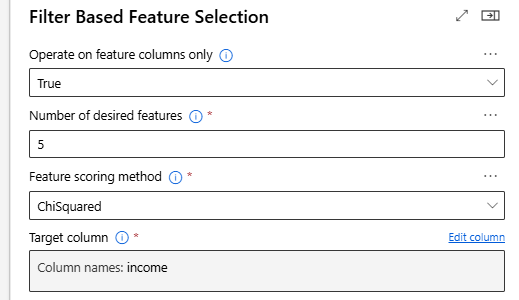




#### 특성 선택 - Filter based Feature Selection

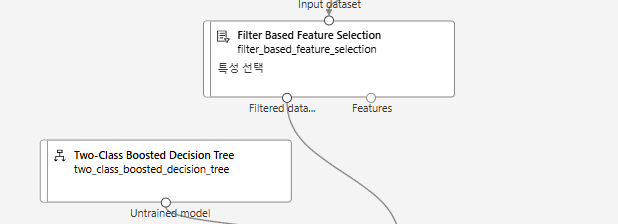


선택할 컬럼 5개, 특성 선택 방법 ChiSquard

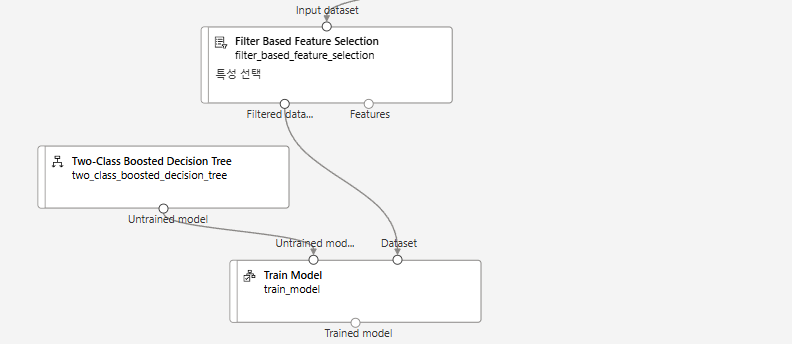


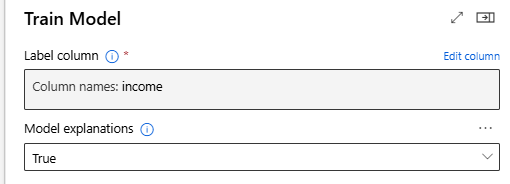
### 모델링/평가

#### 모델링 알고리즘 선택

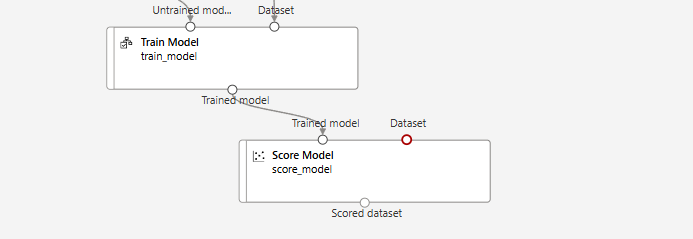


#### 모델 학습(훈련)

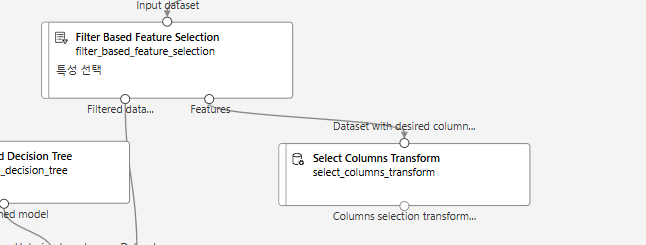


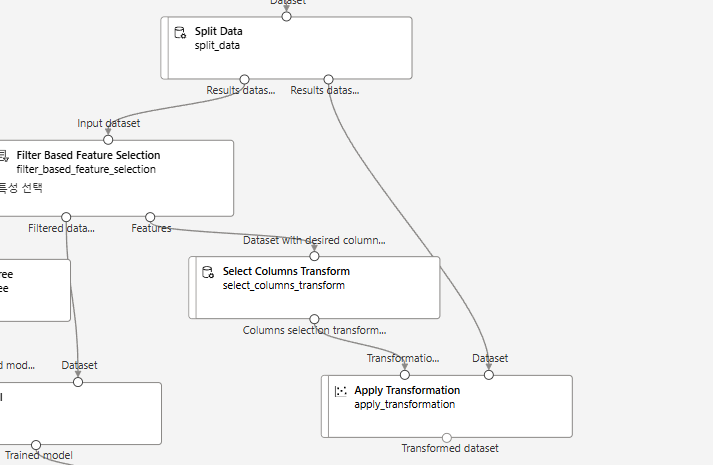


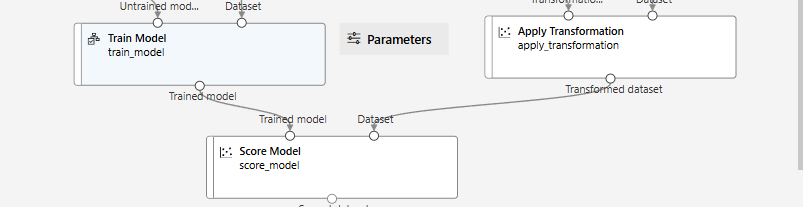
#### 모델 테스트



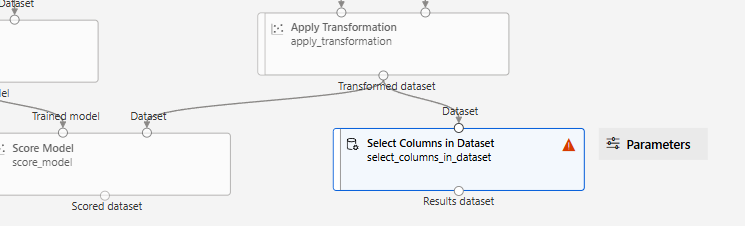
학습 데이터와 컬럼 수를 맞추기 위해 다른 작업을 이어나간다.

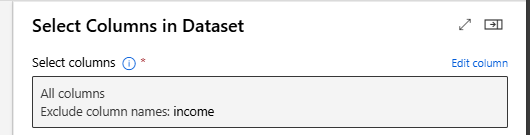






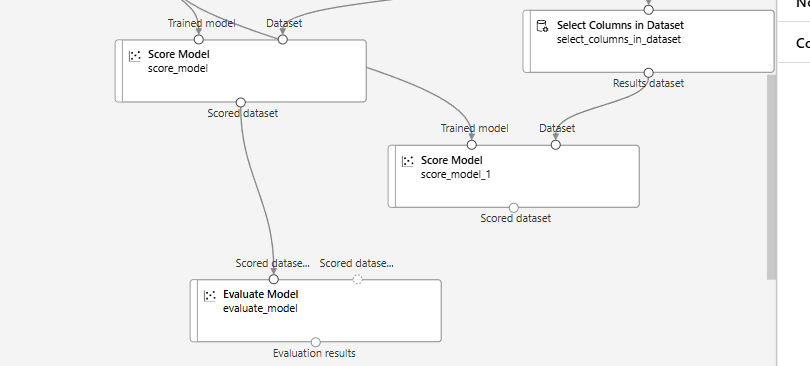
유추를 위한 Score Model

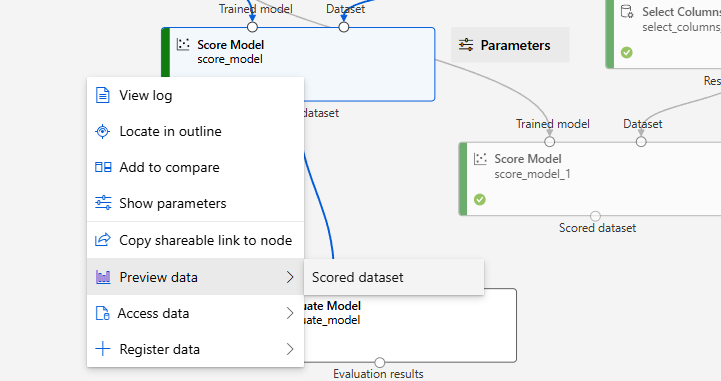


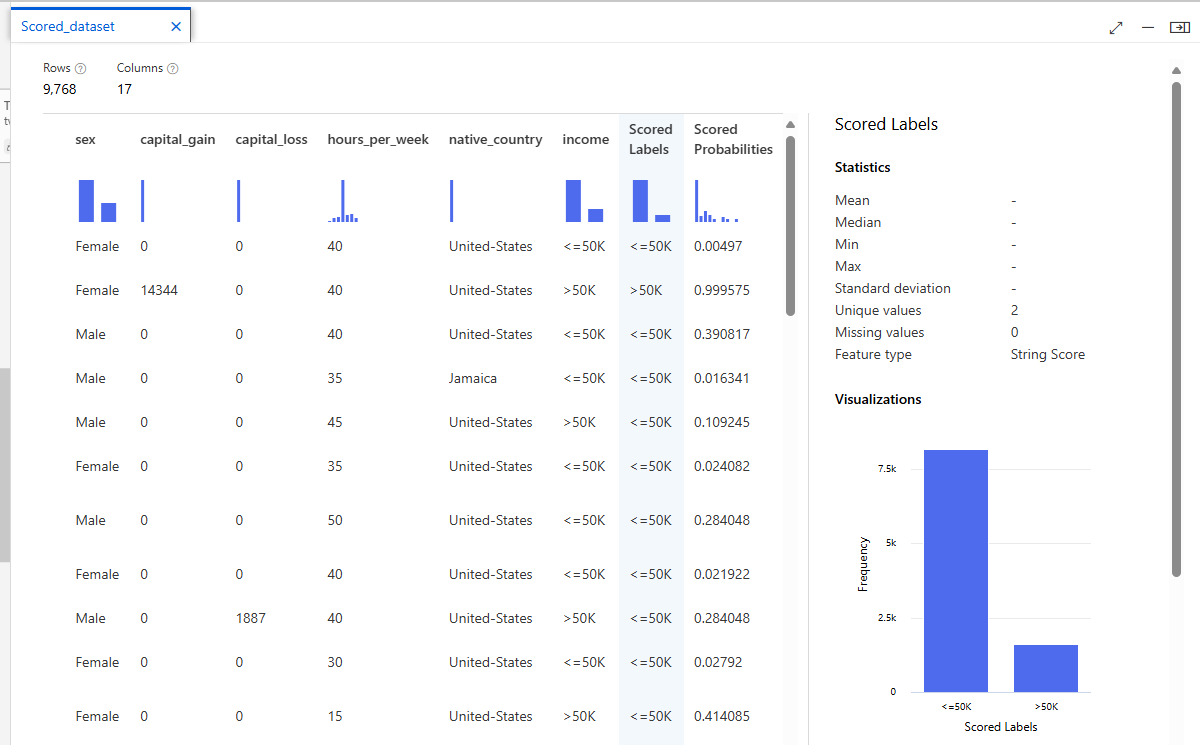




#### 모델 평가









## [실습 - 분류MLD\_신용위험예측\_SVM]

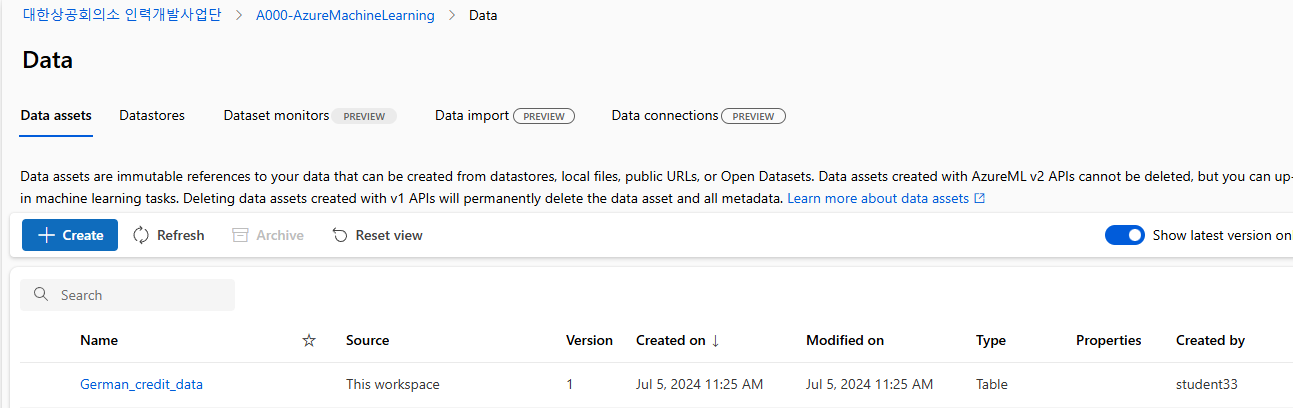
### 실습 준비

#### 데이터 주십

https://archive.ics.uci.edu/dataset/144/statlog+german+credit+data 에서 다운로드

이후 머신러닝 리소스 만들기 생략

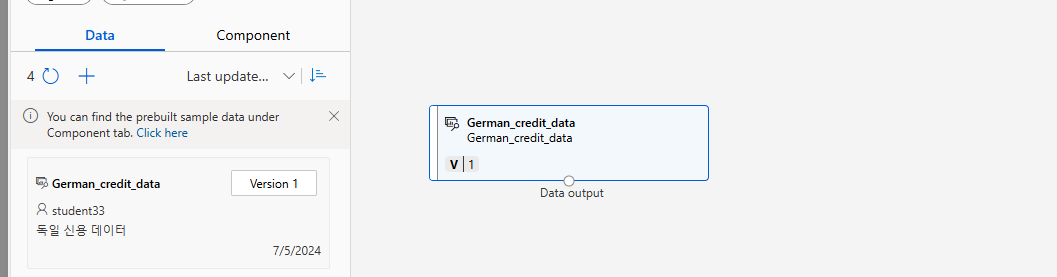
#### 데이터 세트 등록



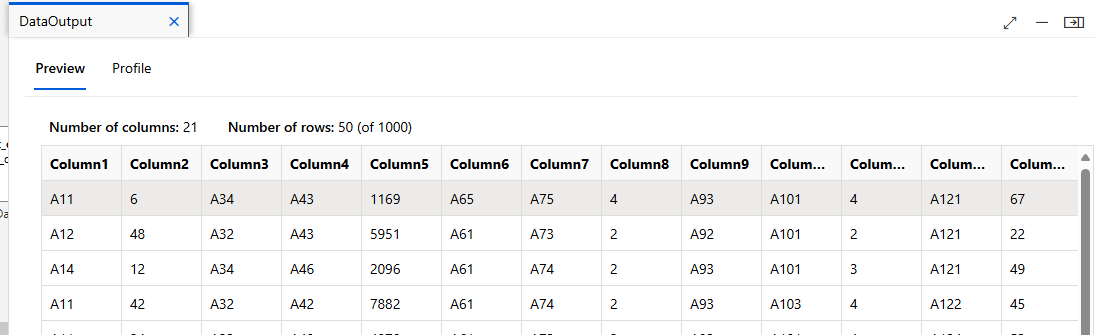
이후 컴퓨트 생략

### 데이터 수집/이해

#### 데이터 세트 가져오기



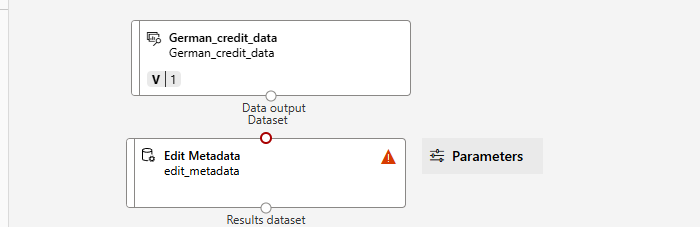
#### 데이터 이해

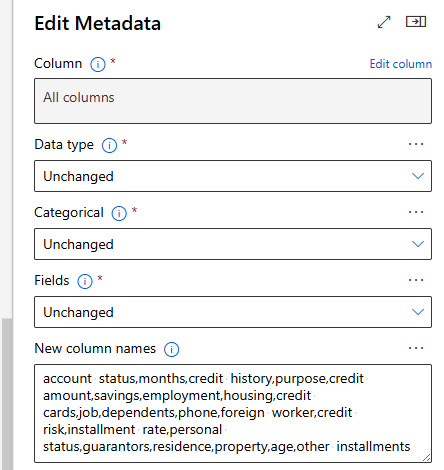




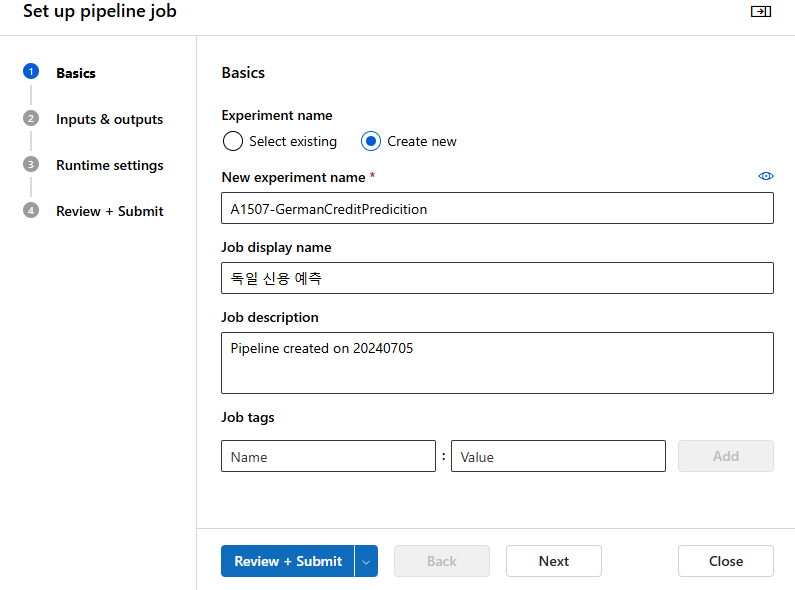
### 데이터 준비

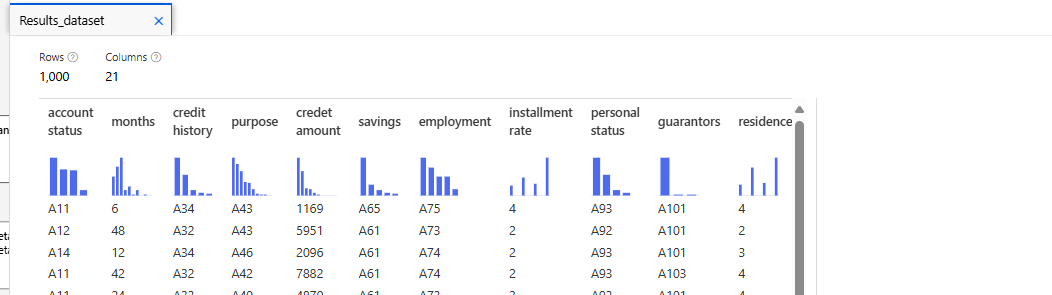
#### 메타데이터 변환 – 컬럼명 부여



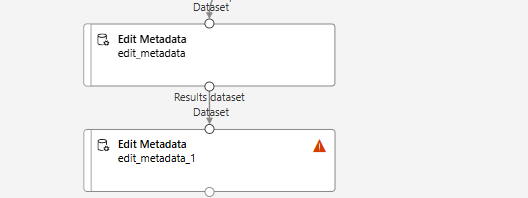


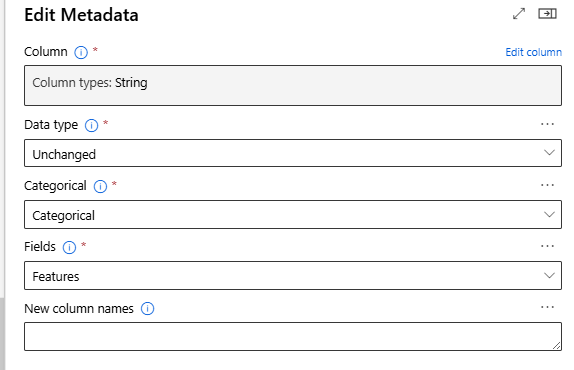
#### 중간점검

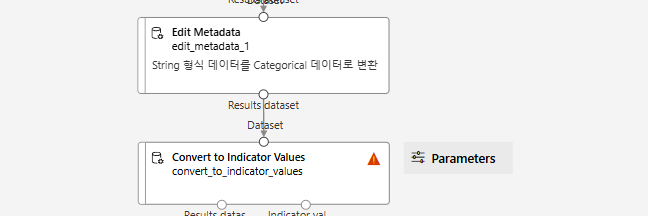


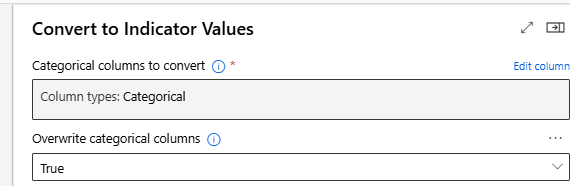


#### 데이터 변환 – String -> Category -> Indicator value

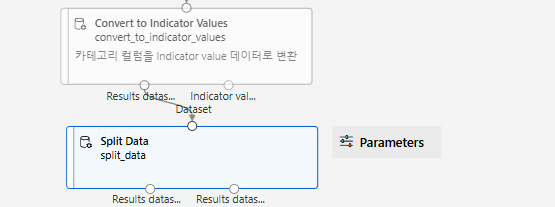


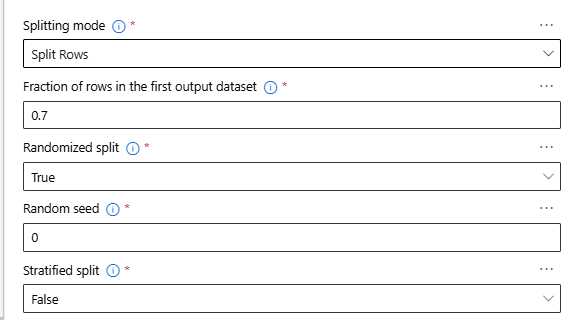




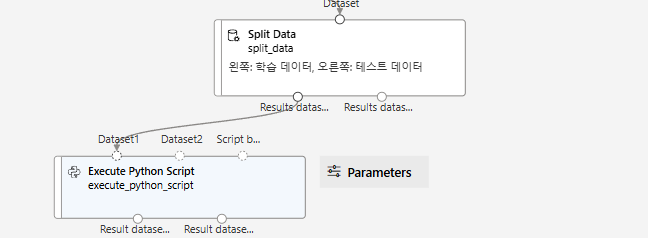


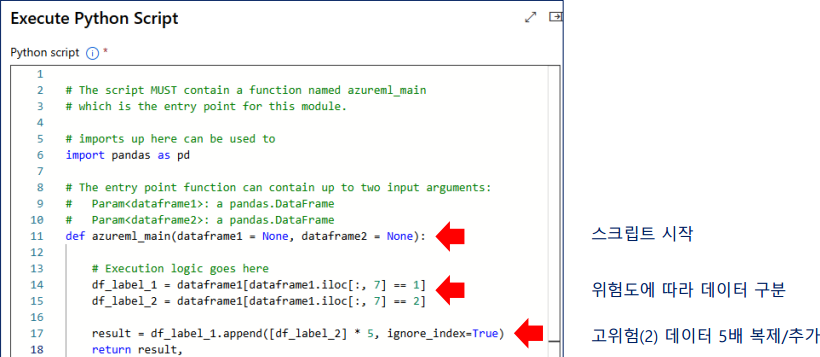
#### 데이터 분리

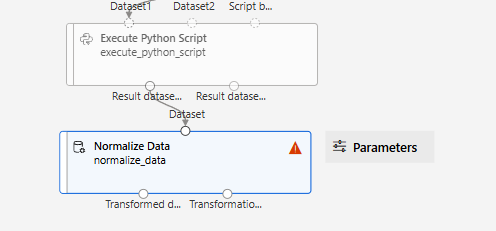


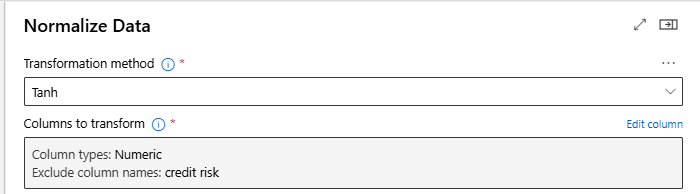


#### 데이터 변환 – 위험도 고려 및 정규화

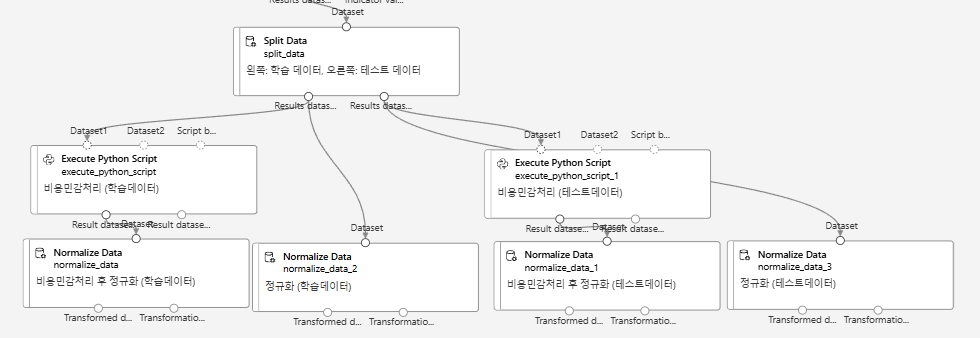






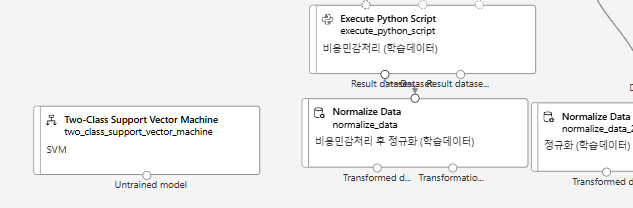


테스트 데이터도 똑같이 반복 + 파이썬 비용민감처리를 안 한 정보도 확인

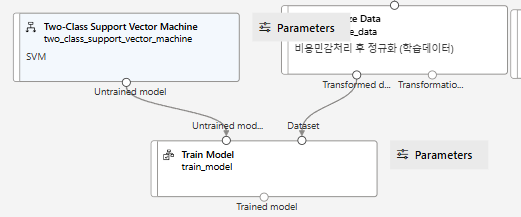


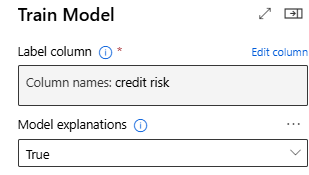
### 모델링/평가

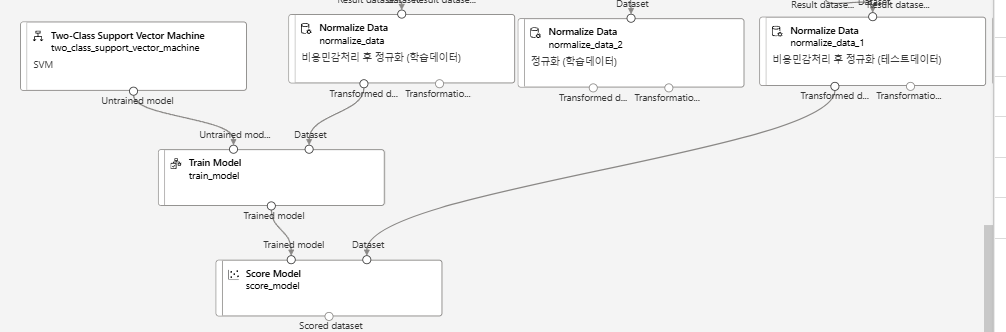
#### 모델링 알고리즘 선택

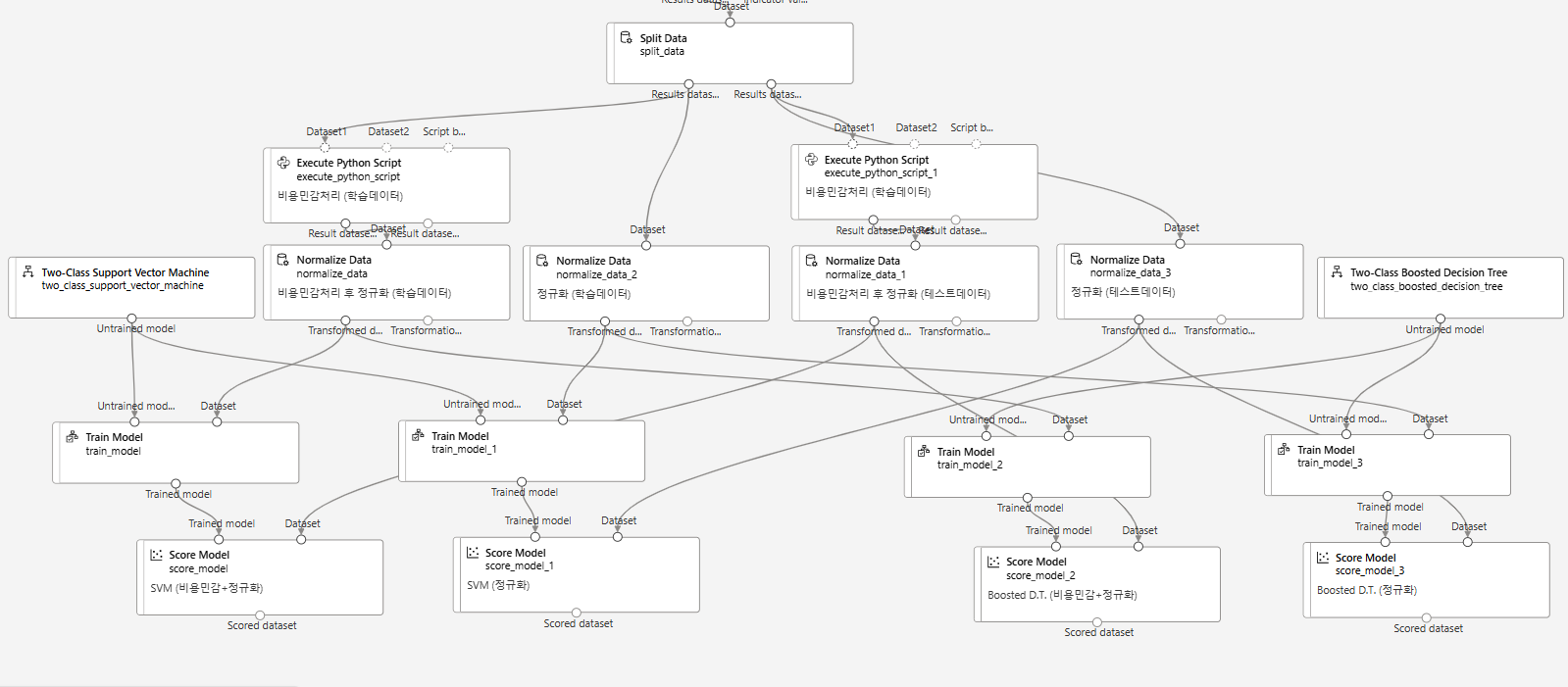


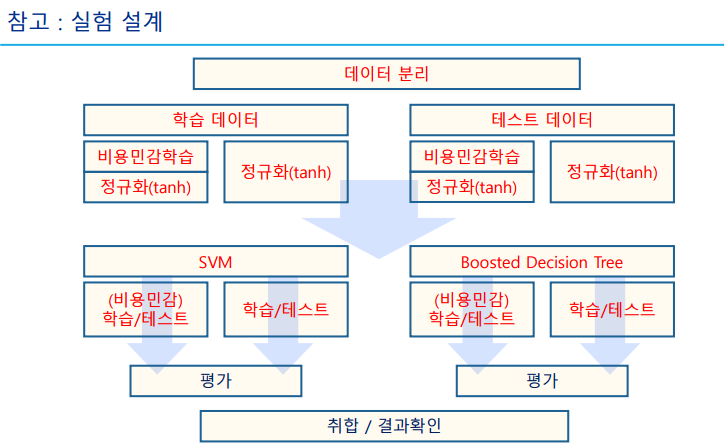
#### 모델 학습 및 테스트

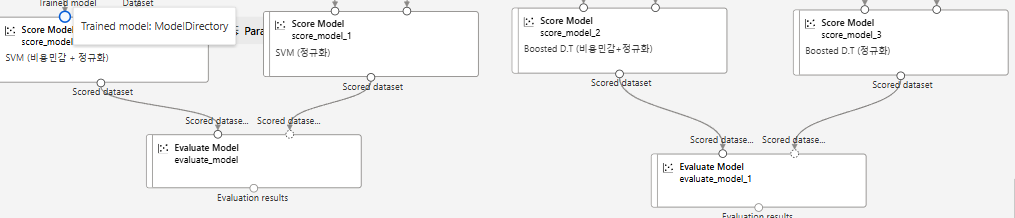






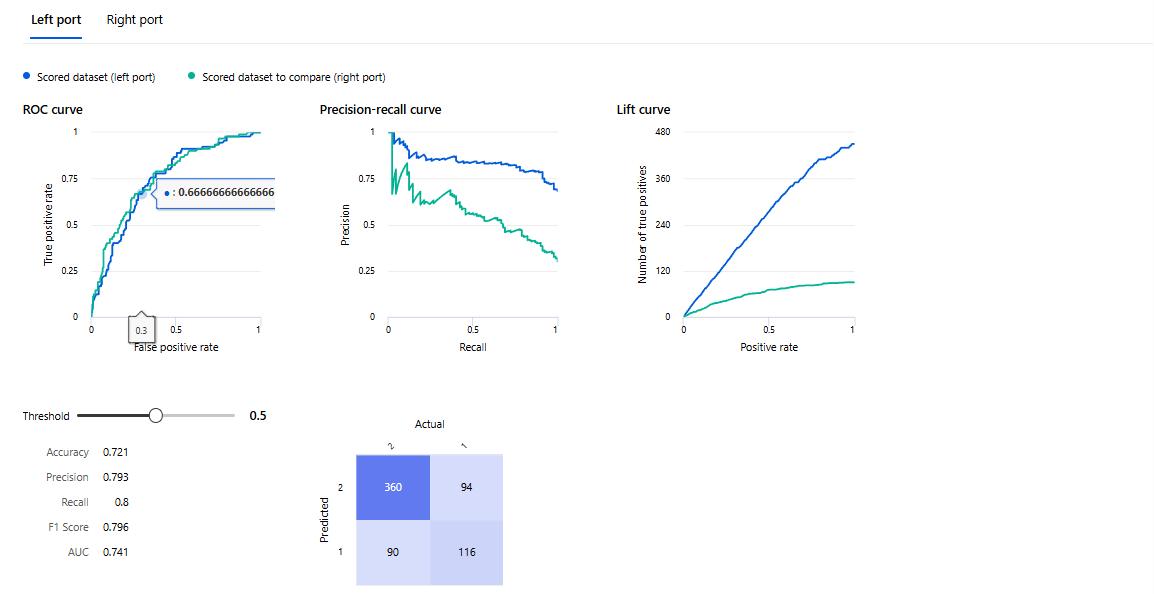


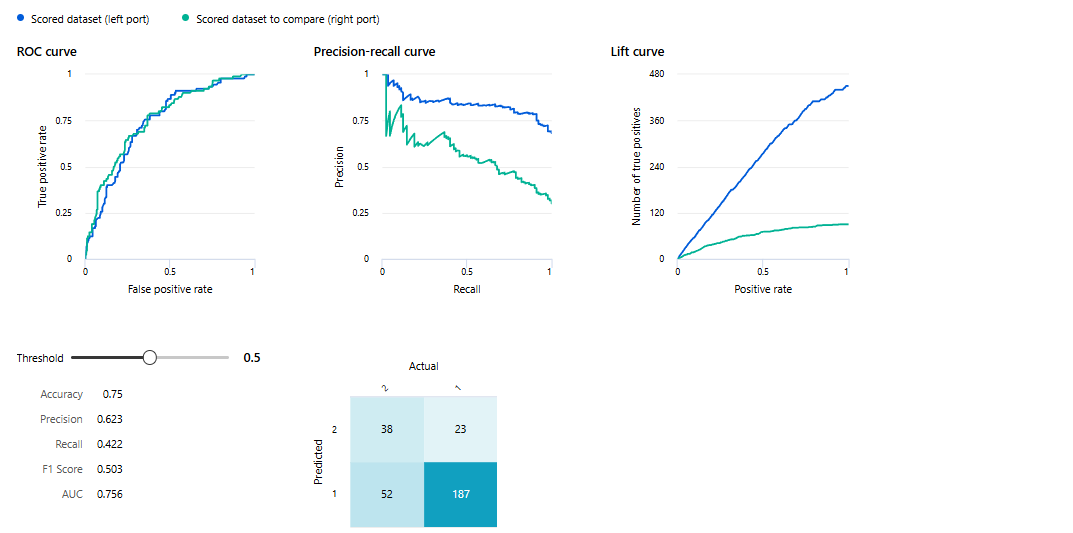




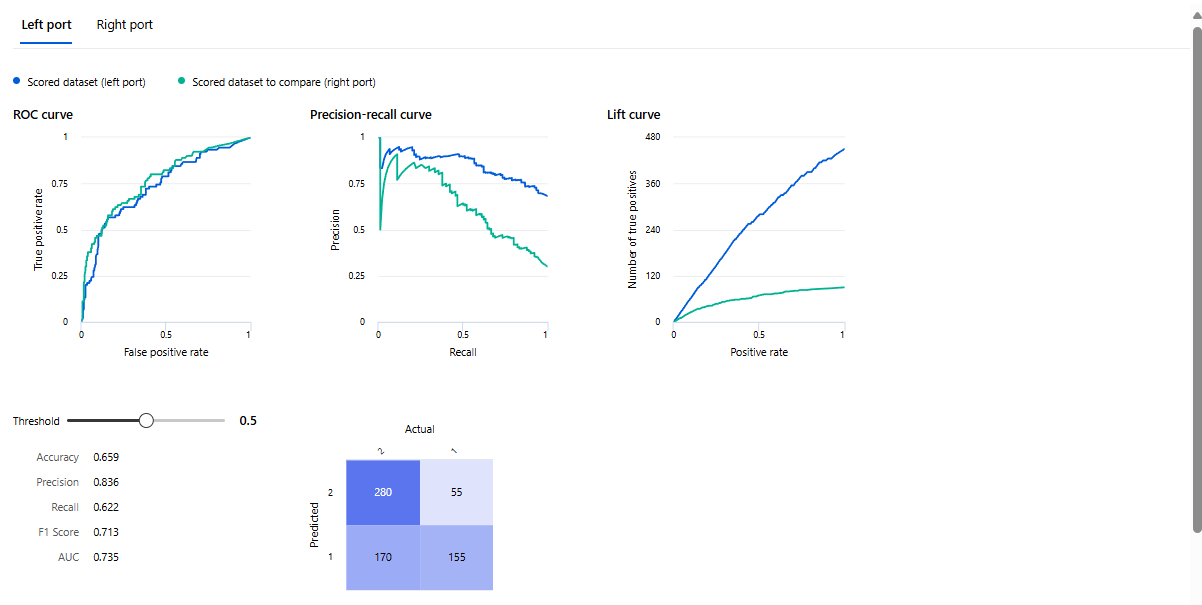
#### 중간 점검

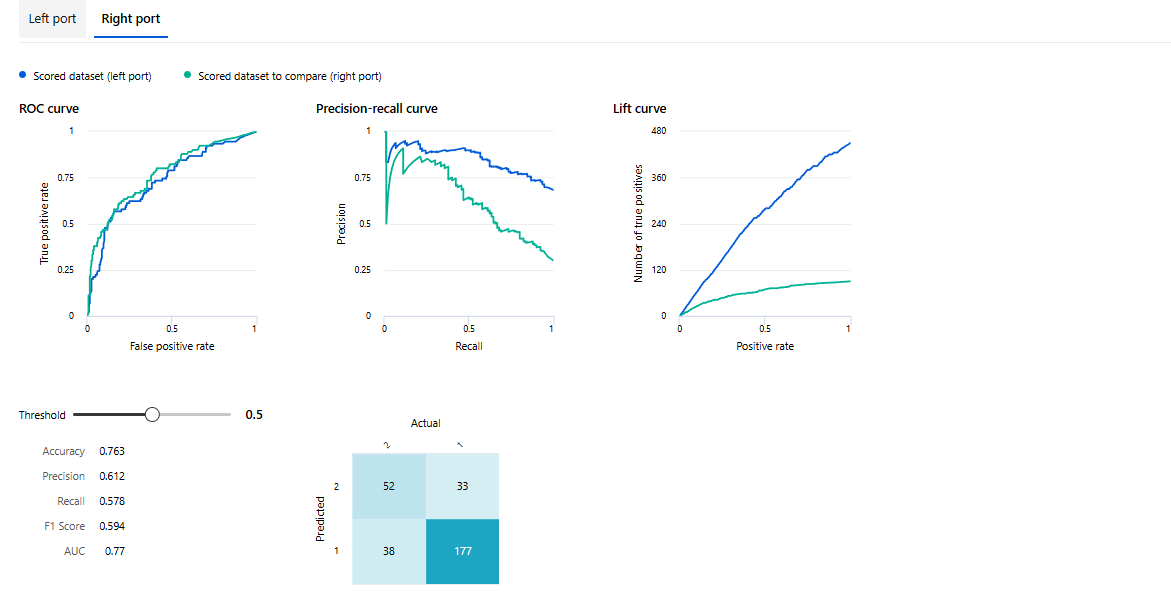
학습 데이터



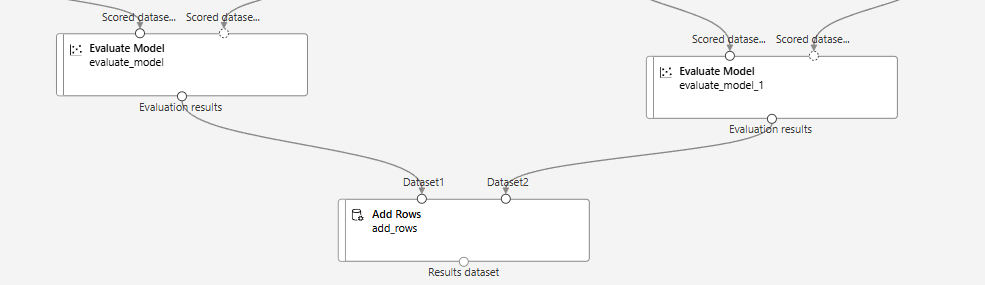


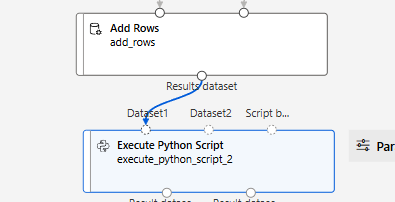
테스트 데이터



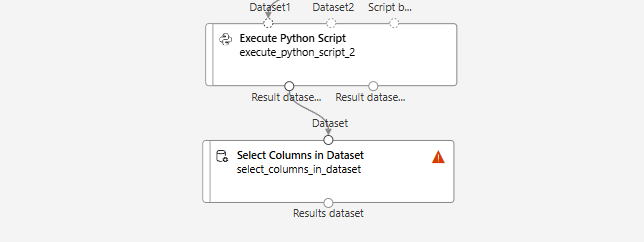


#### 모델 평가(취합)

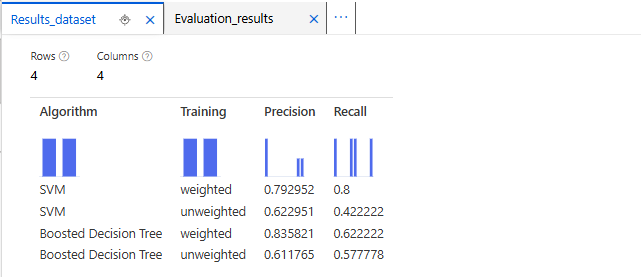












# <오후>

## [실습 - 분류MLD\_대출자격예측\_로지스틱회귀\_층화추출]

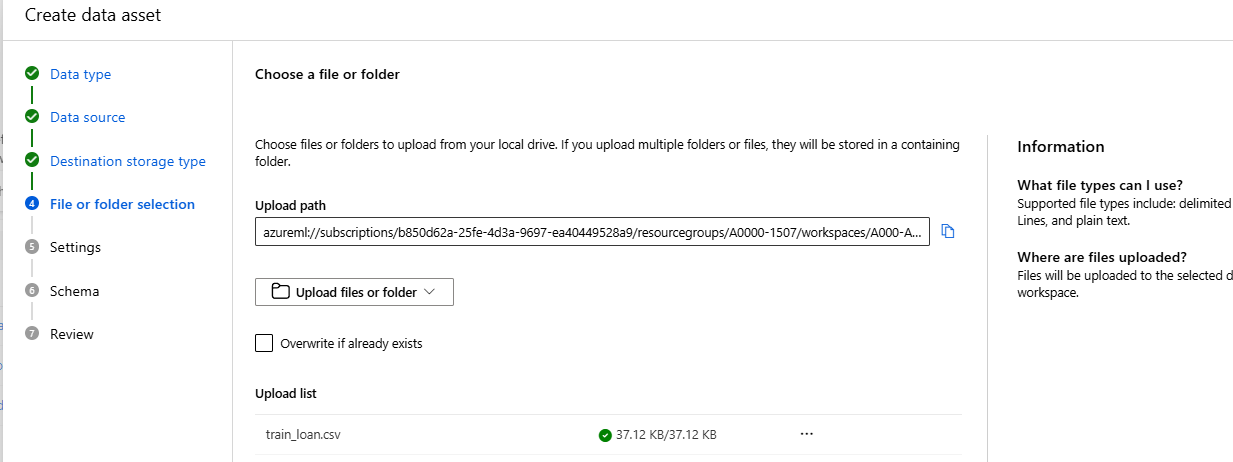
Kaggle에서 데이터 수집

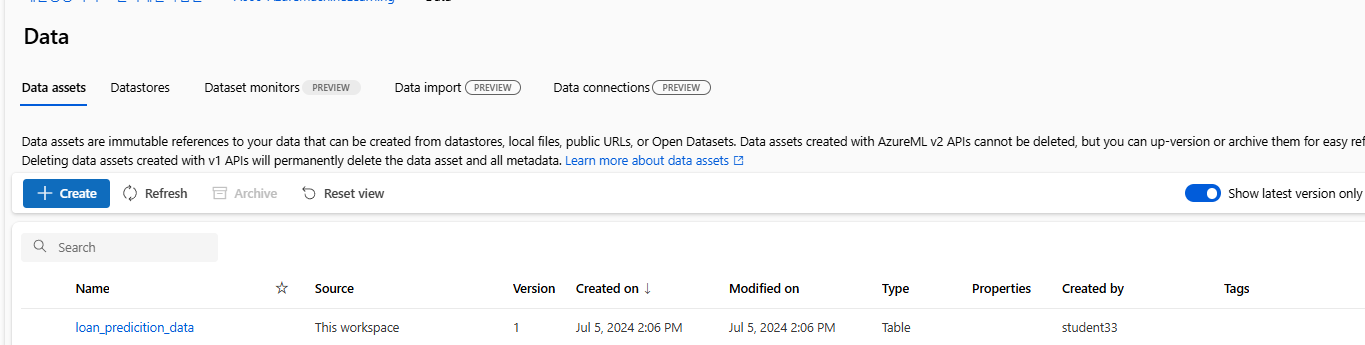
(<https://www.kaggle.com/datasets/altruistdelhite04/loan-prediction-problem-dataset>)

### 실습 준비

#### 데이터세트 등록

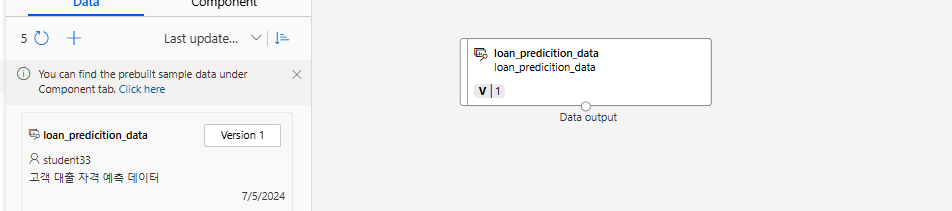




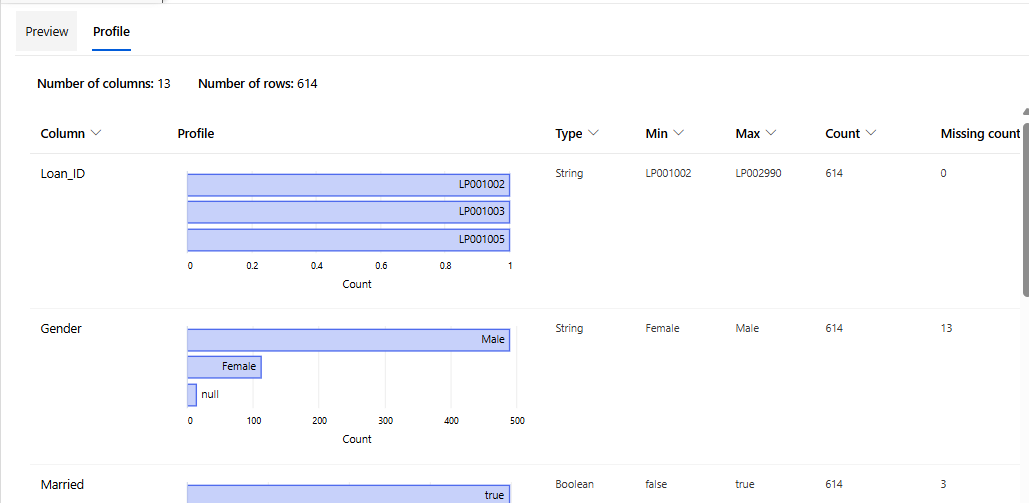


### 데이터 수집/이해

#### 디자이너 시작 및 데이터 세트 가져오기

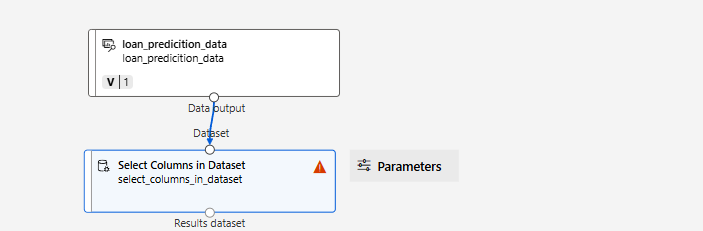


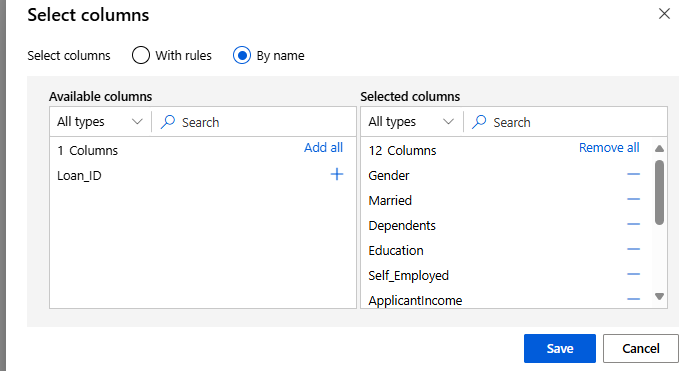
#### 데이터 이해



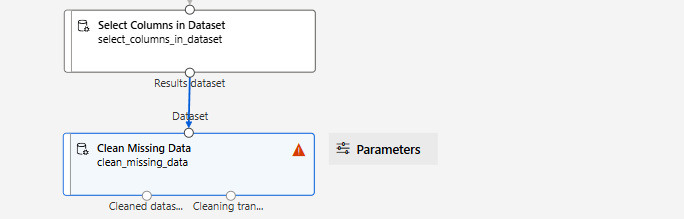
### 데이터 준비

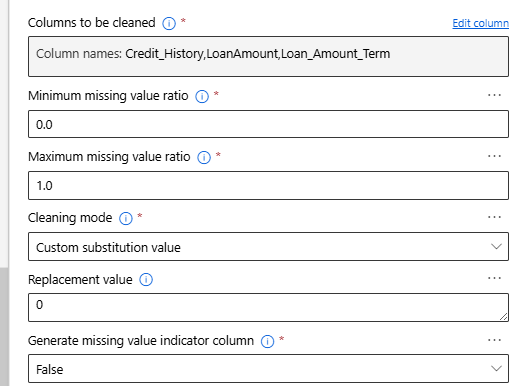
#### 특성 선택 – 불필요한 컬럼 제외

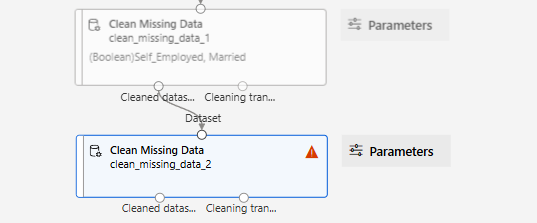


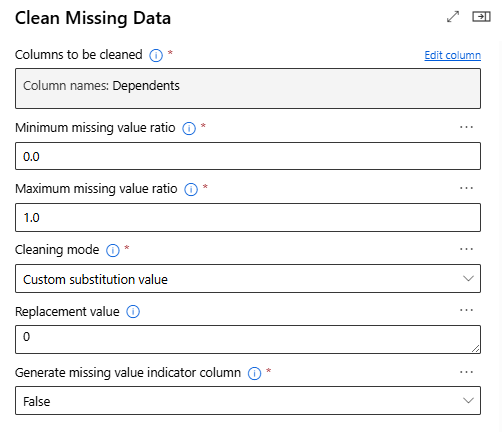


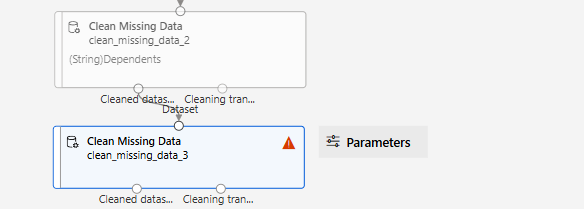
#### 누락값 처리

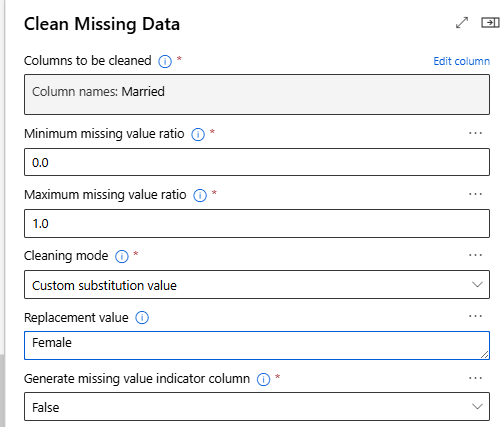






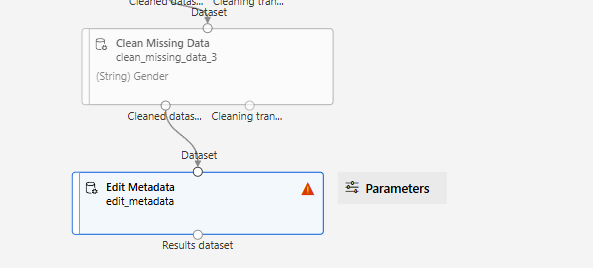


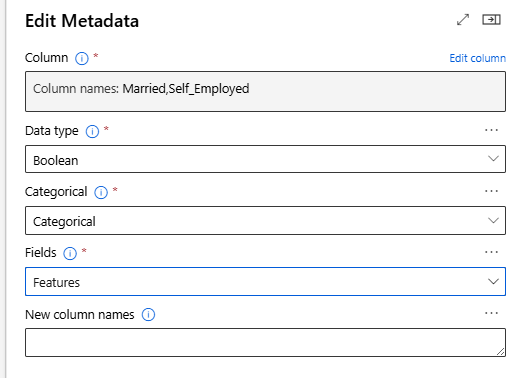




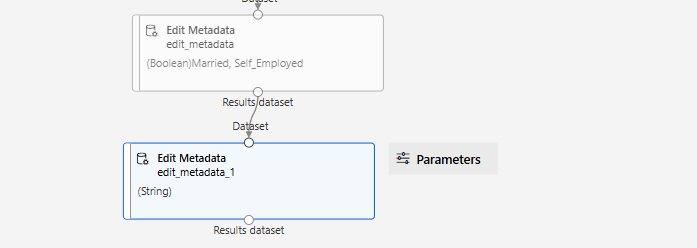
#### 데이터 변환 : String → Category → Indicator value

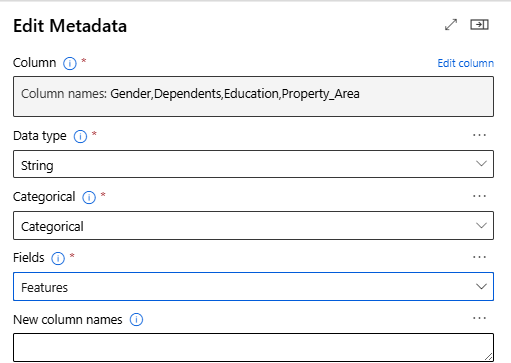
Boolean→Category : Married, Self\_Employed



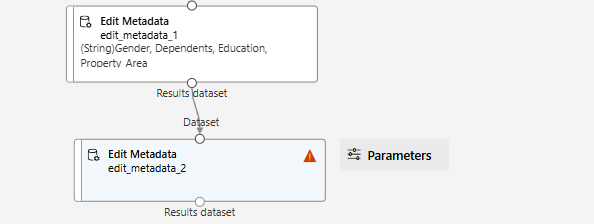


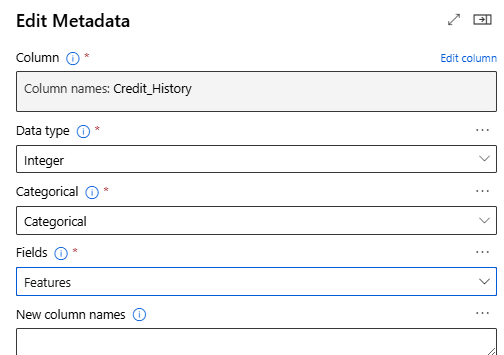
String→Category : Gender, Dependents, Education, Property\_Area



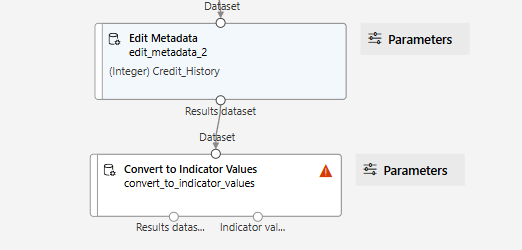


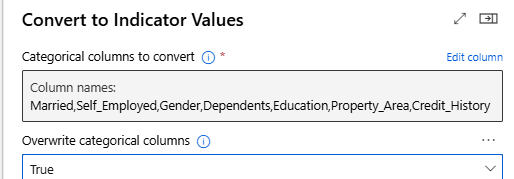
Integer→Category : Credit\_History



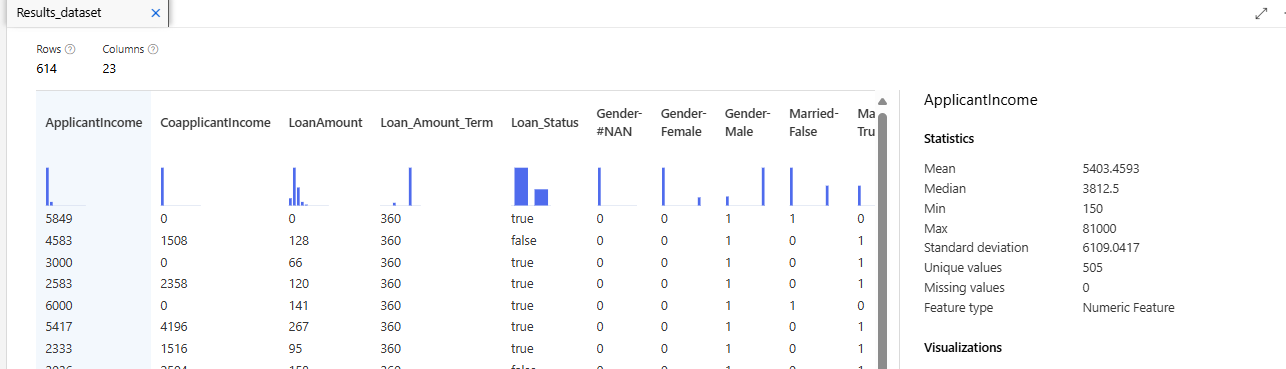


Category→ Indicator value

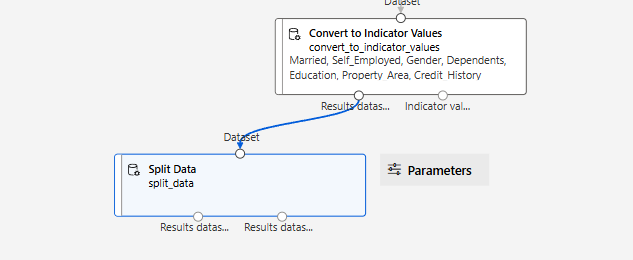


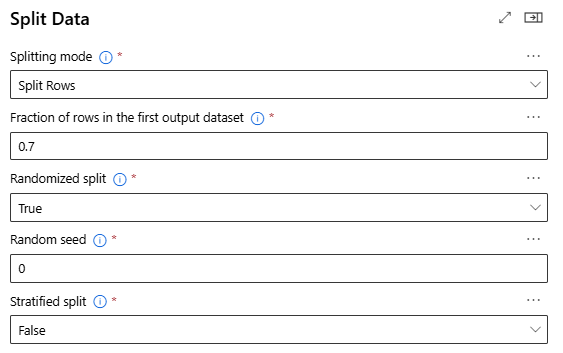


#### 중간 점검



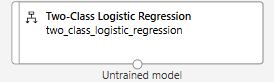
#### 데이터 분리



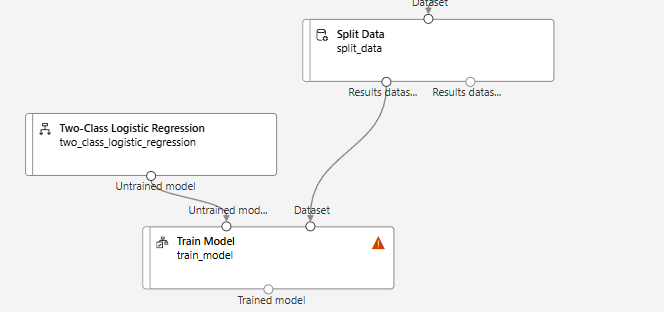


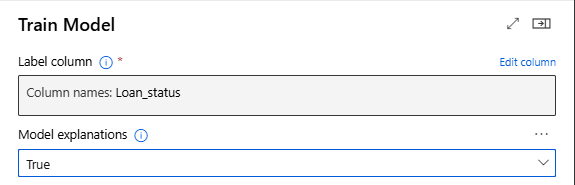
### 모델링/평가

#### 모델링 알고리즘 선택

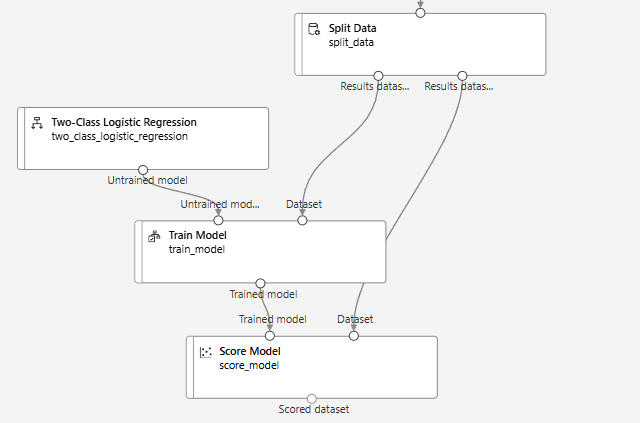


#### 모델 학습(훈련)

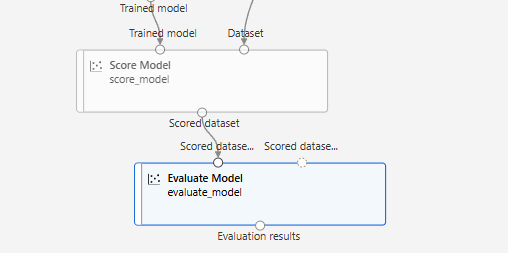


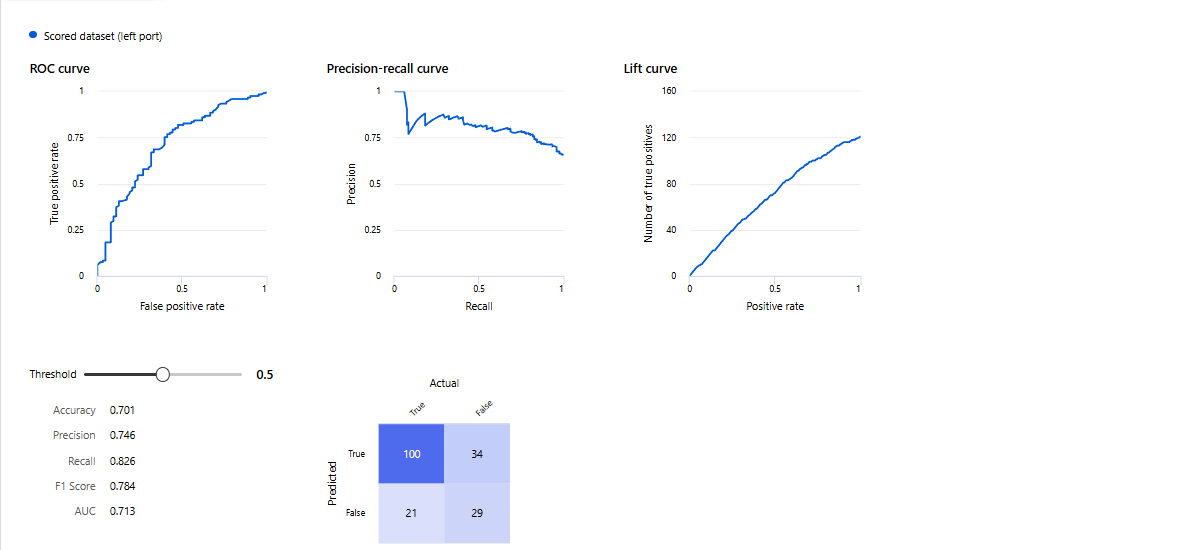


#### 모델 테스트



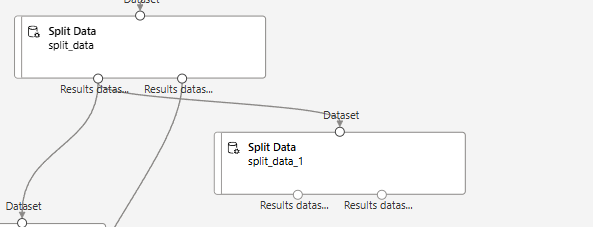
#### 모델 평가

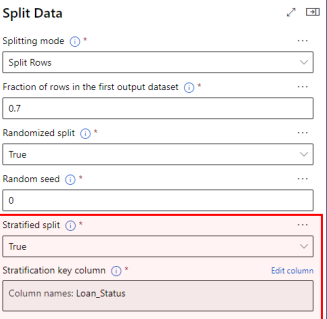




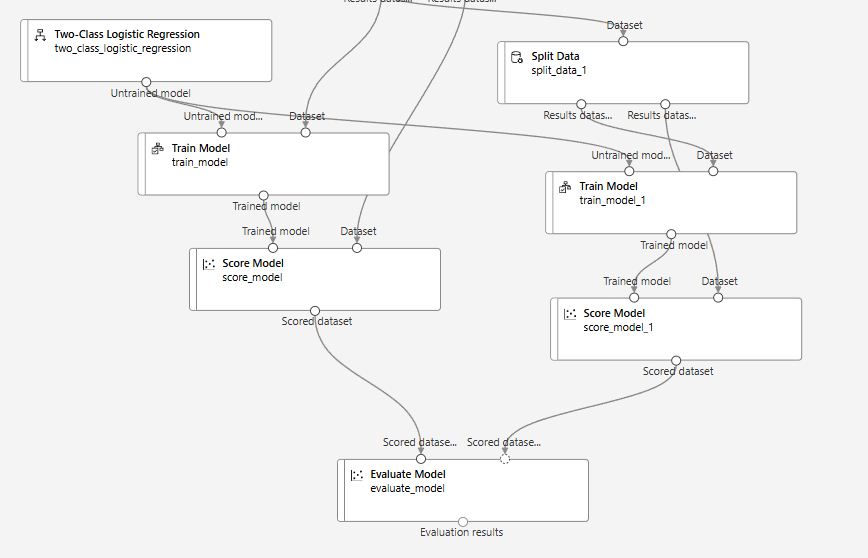
### 추가 실습 – 층화 추출

#### 학습 데이터와 테스트 데이터로 분리 – 층화 추출

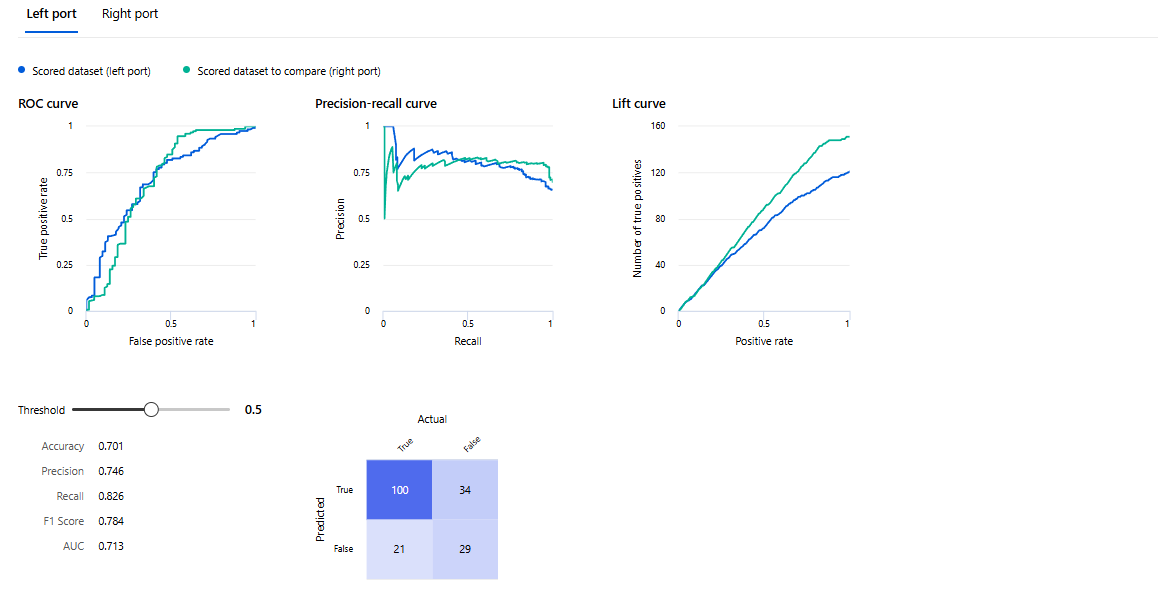




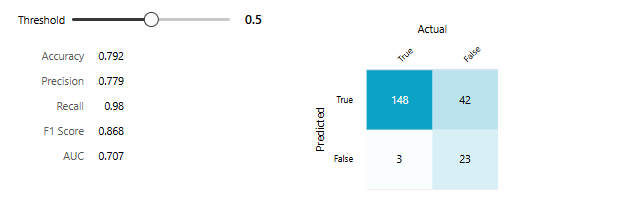
후 전과 같이 구성



#### 모델 평가 결과 – 층화 추출

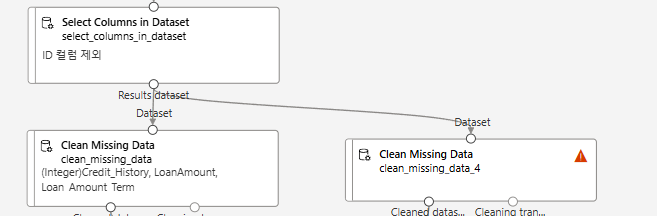


얘가 층화 추출

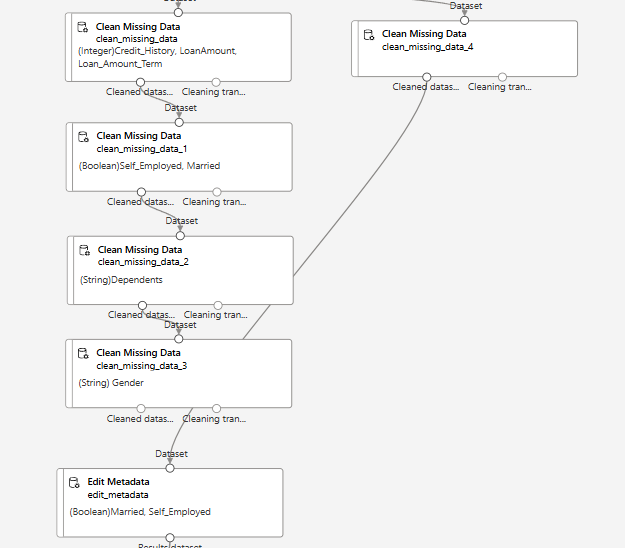


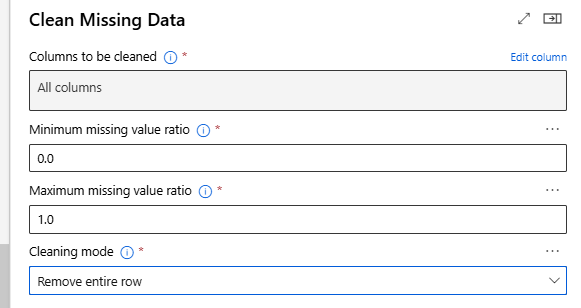
### 추가 실습 – 누락값 컬럼 제거

#### 누락값 처리 – 데이터 샘플 제거

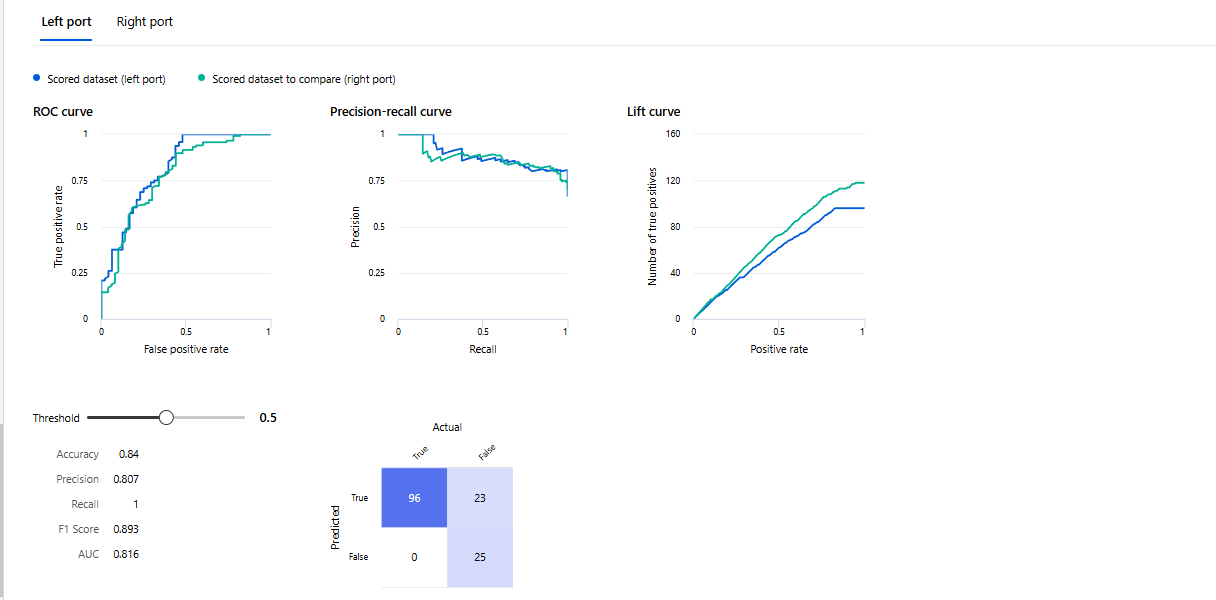


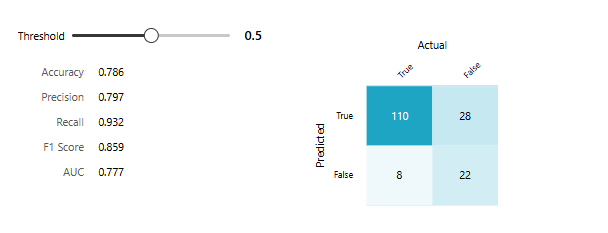
Clean Missing Data 생성 후 선만 다시 이어주고 저장





#### 모델 평과 결과 – 누락값 컬럼 제거





## [실습 - 분류MLD\_CRM고객이탈예측\_BoostedDT\_SMOTE]

(Lab : 오버샘플링을 이용한 CRM 고객 이탈 예측 모델 구현)

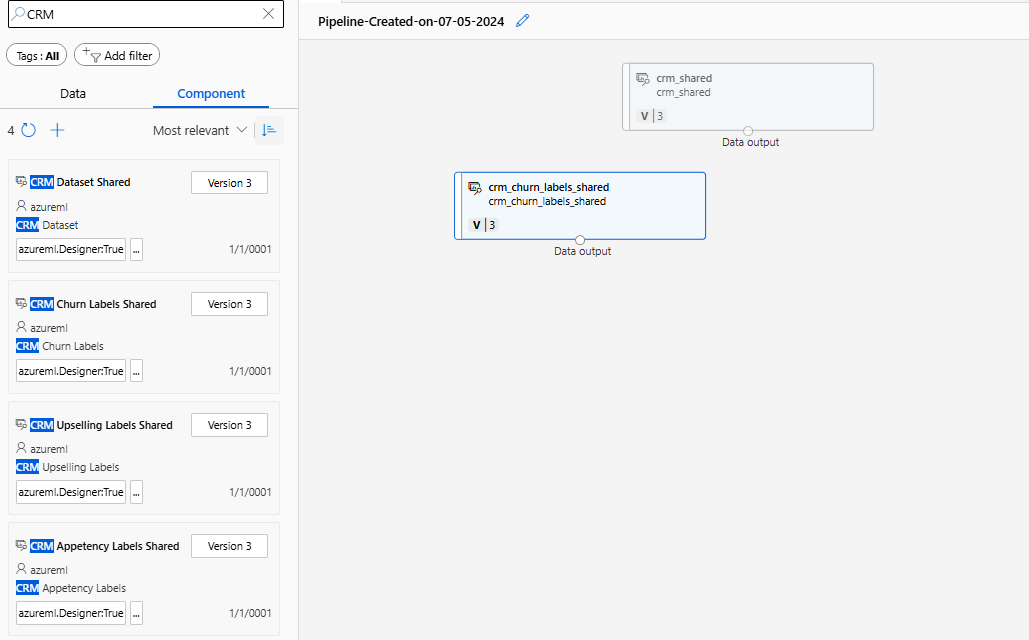
### 실습 준비

#### 데이터 수집

애저에 등록되어 있는 샘플 데이터 사용

### 데이터 수집/이해

#### 디자이너 생성 및 데이터 세트 가져오기

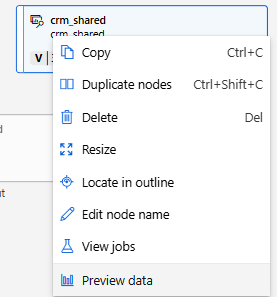


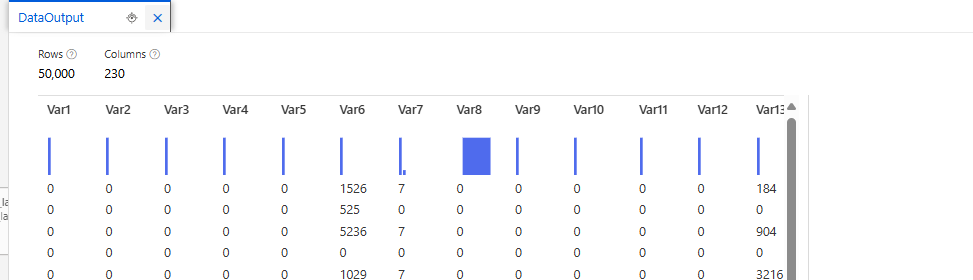
• CRM Data Shared : 특성 데이터 (230개 컬럼)

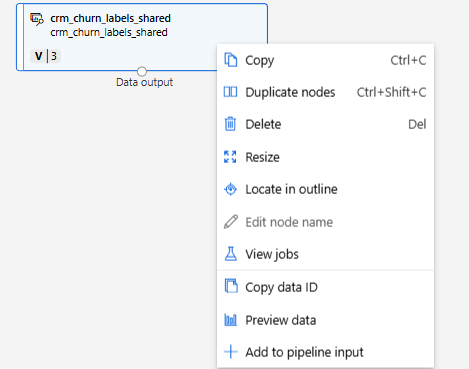
• CRM Churn Labels Shared :이탈 고객 레이블

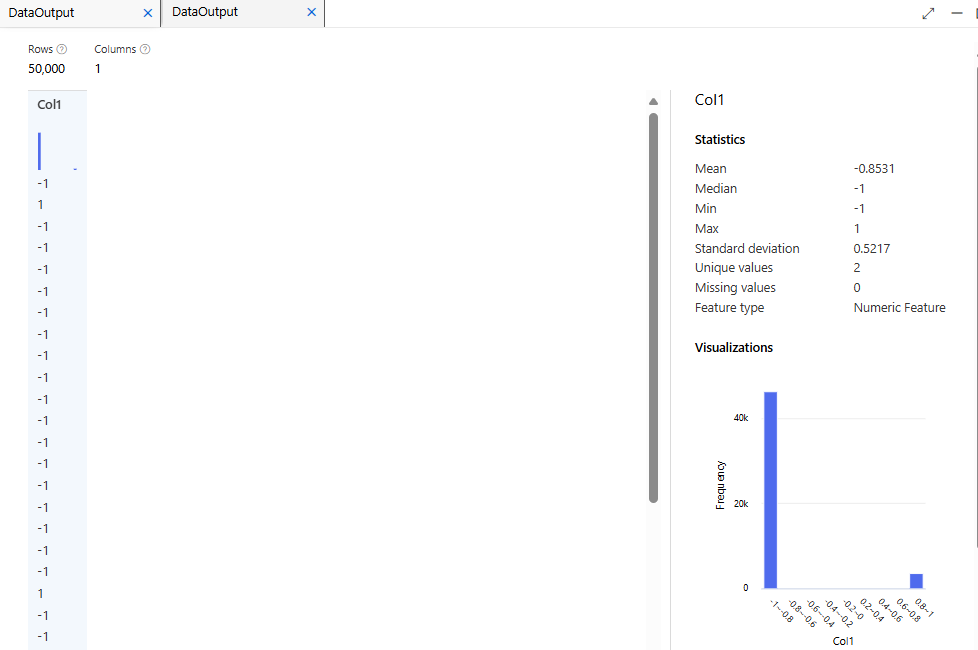
• CRM Upselling Labels Shared : 새로운 서비스 잠재 고객 레이블

• CRM Appentency Labels Shared : 마케팅에 호의적인 고객 레이블







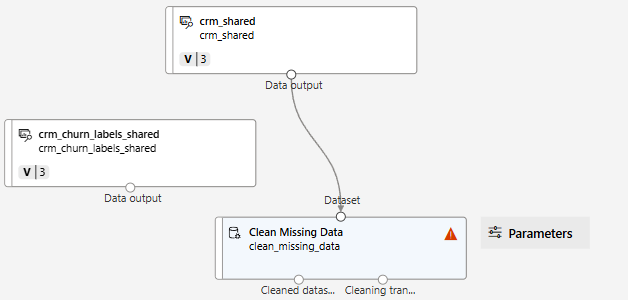


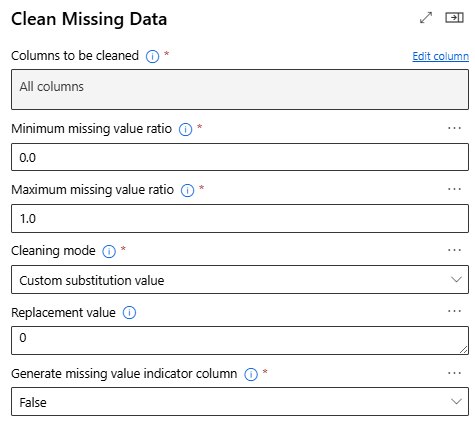
### 데이터 준비

#### 특성 선택

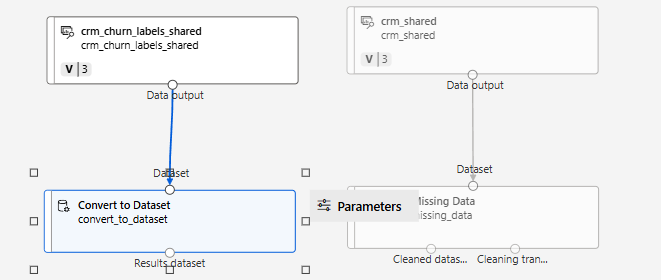
할 게 없다.

#### 누락값 처리



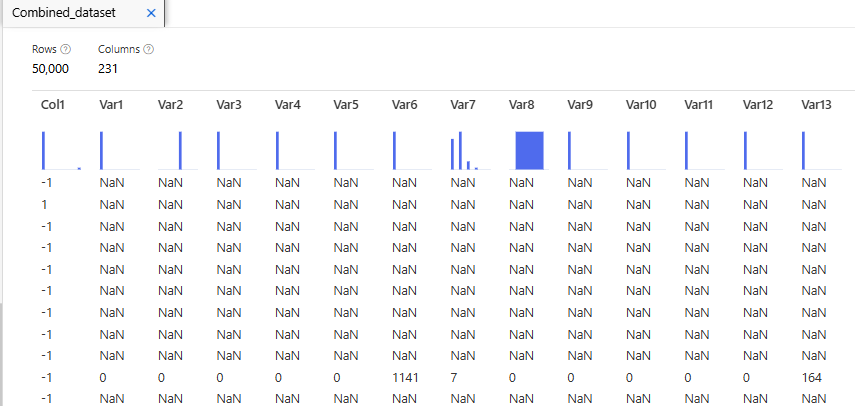


#### 데이터 변환 – 열 병합

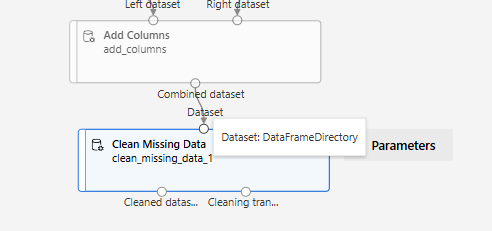




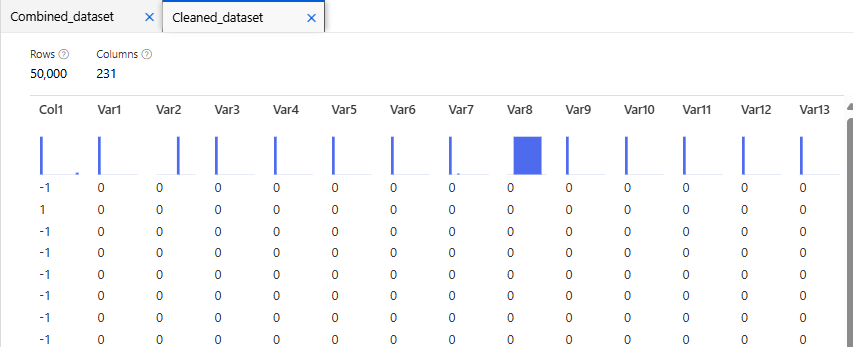
#### 중간 점검



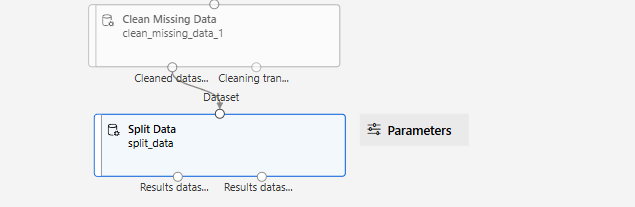
#### 누락값 처리(NaN)

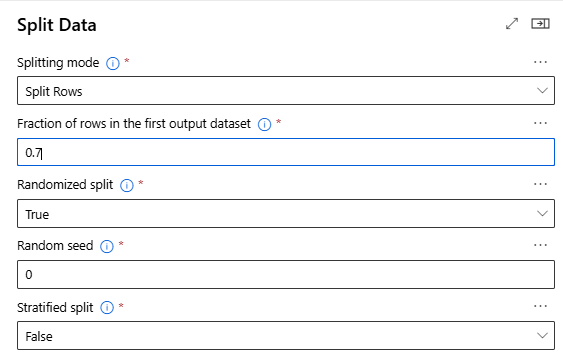


#### 중간 점검 비교 실행

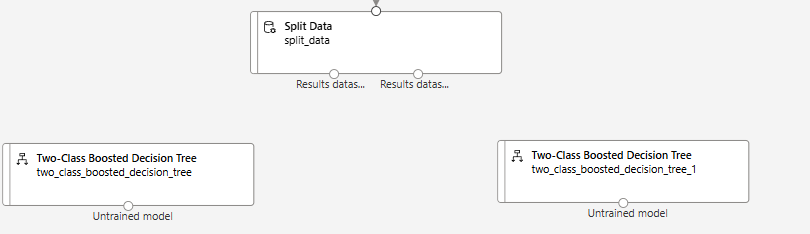


#### 데이터 분리

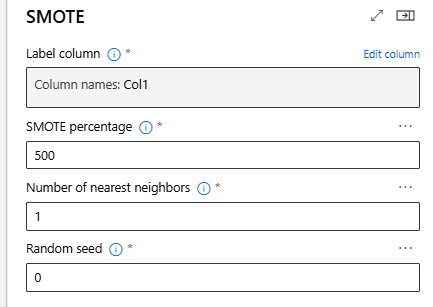




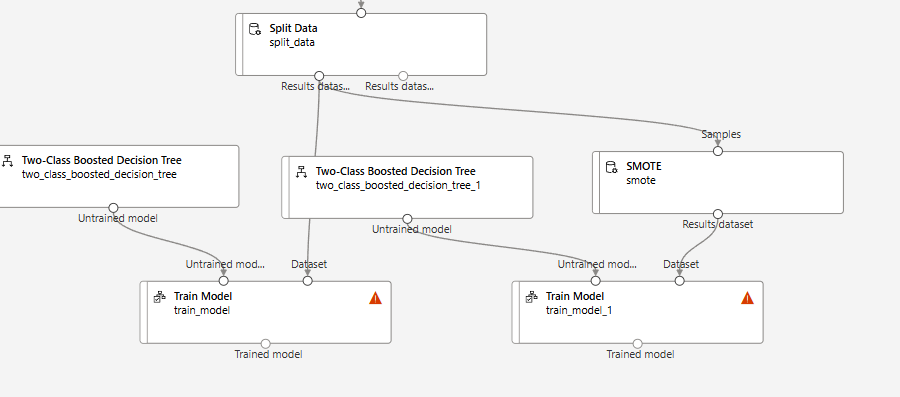
### 모델링/평가





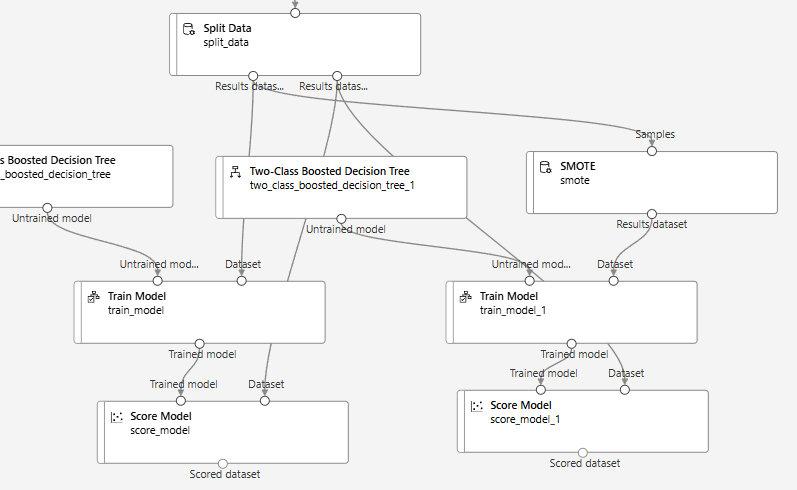


#### 모델 학습(훈련)

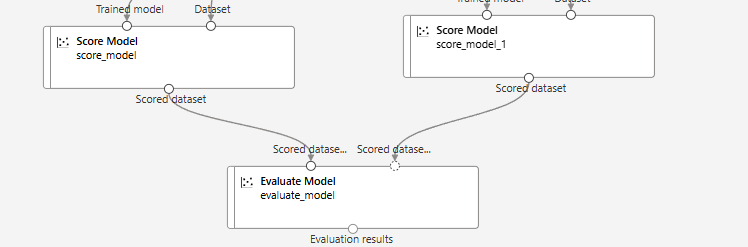


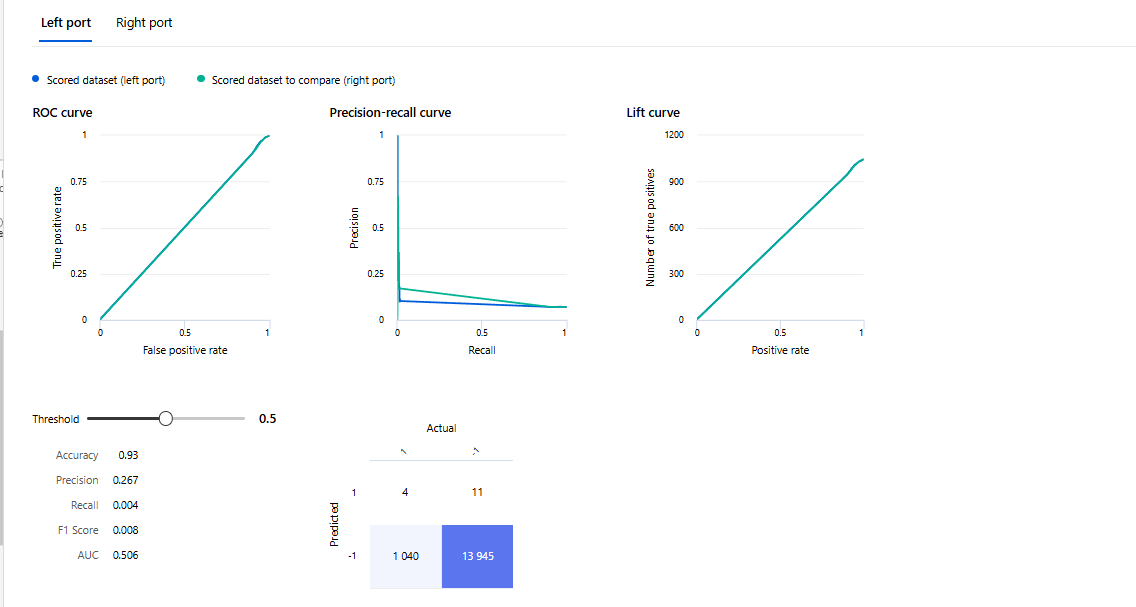


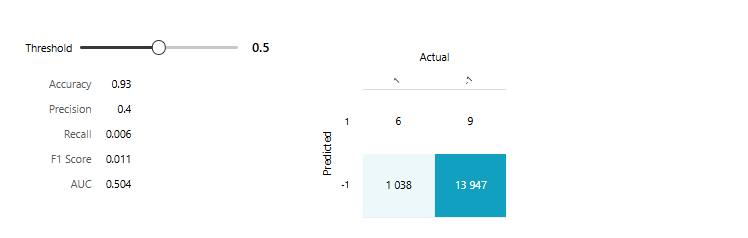
#### 모델 테스트



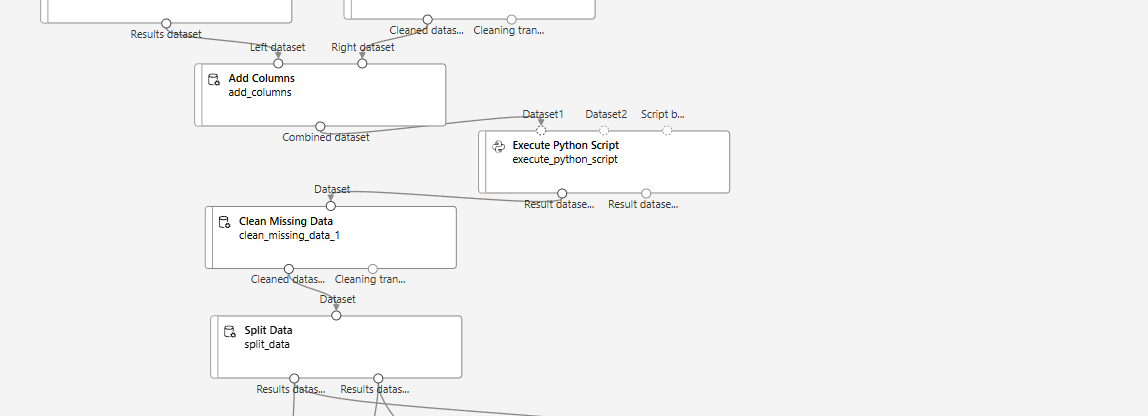
#### 모델 평가

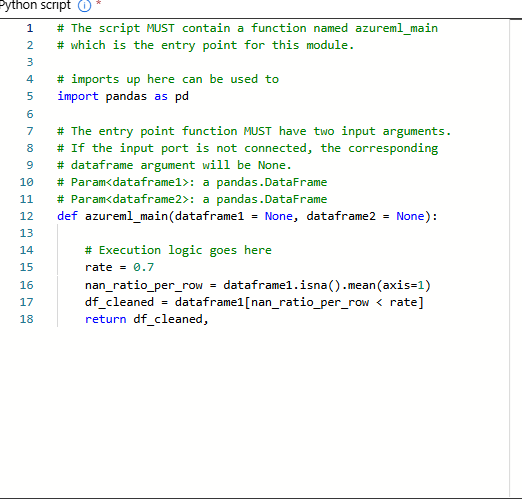


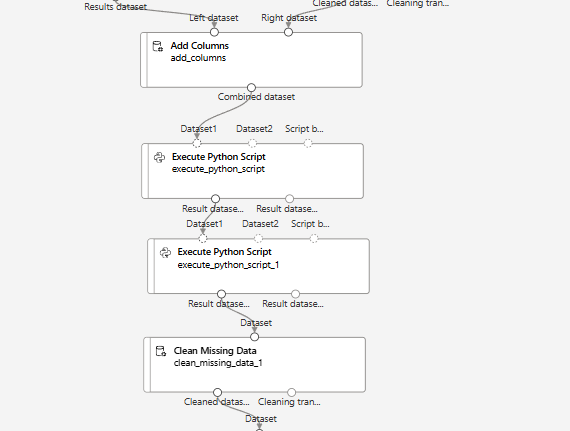


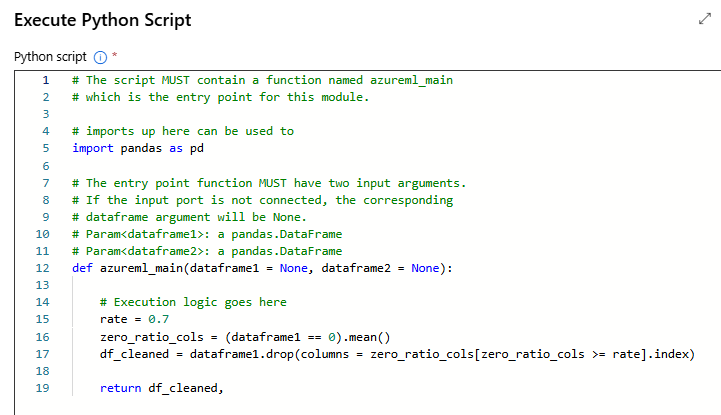


### 추가 실습 – 누락값 많은 데이터행/컬럼 제거









## [실습 - 개인수입예측\_교차검증]

### 실습 준비

#### 데이터 수집

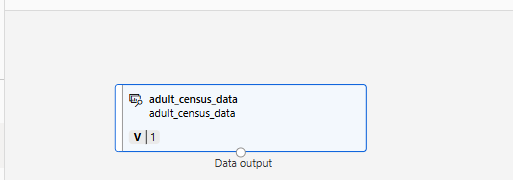
kaggle에서 adult census income 검색 후 사용 (<https://www.kaggle.com/datasets/uciml/adult-census-income>)

#### 애저 접속 -> 리소스 생성 -> 데이터 세트 등록 -> 컴퓨트 대상 설정

생략

### 데이터 수집/이해

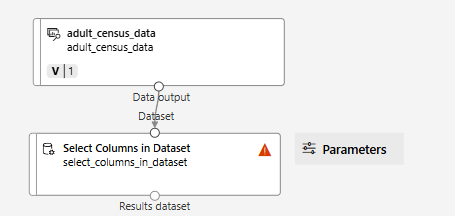
#### 디자이너 시작 및 데이터 세트 가져오기

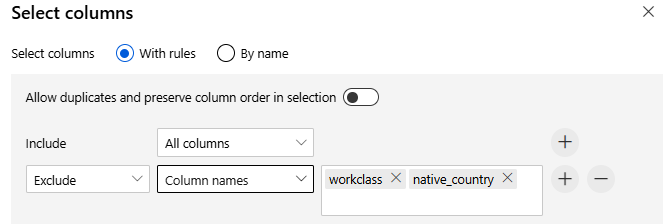


#### 데이터 이해

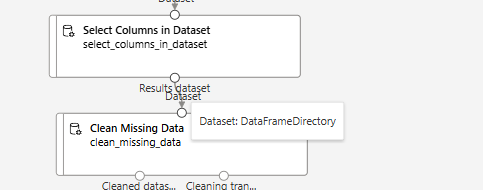


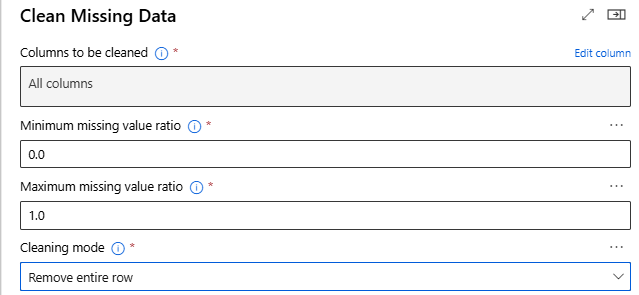
### 데이터 준비





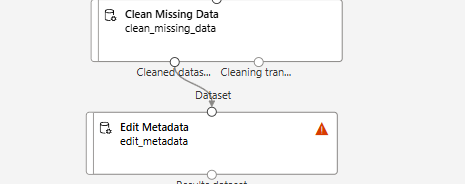
#### 누락값 처리 – 데이터 샘플 제거

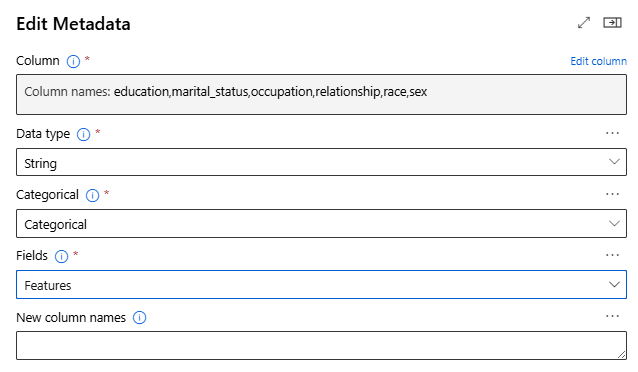




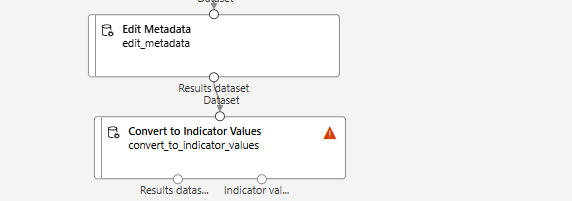
#### 데이터 변환 : String → Category → Indicator value

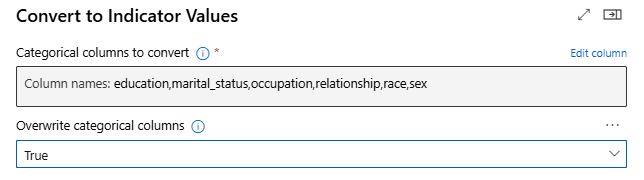
String → Category



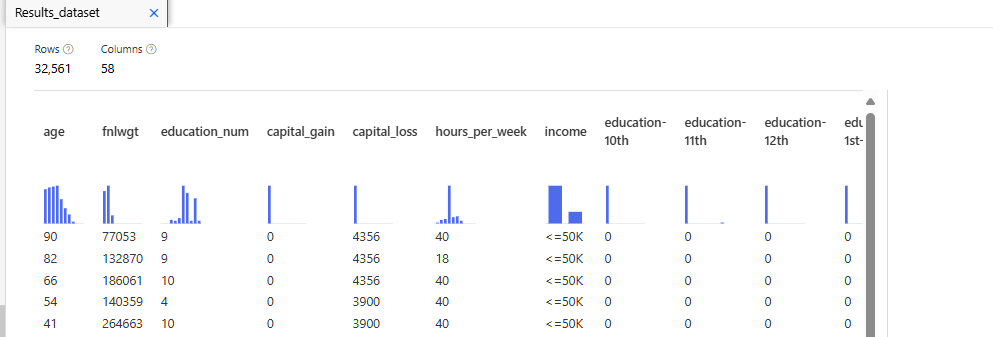


Category→ Indicator value

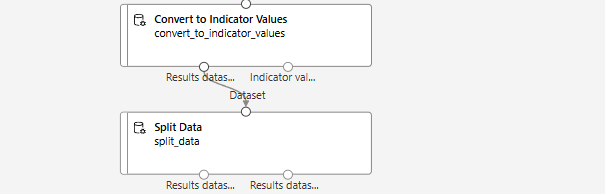


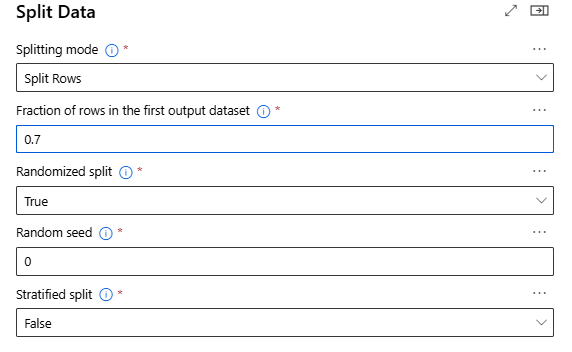


#### 중간 점검



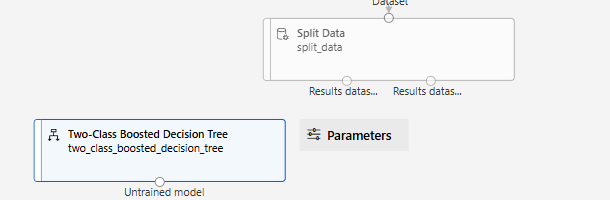
#### 데이터 분리



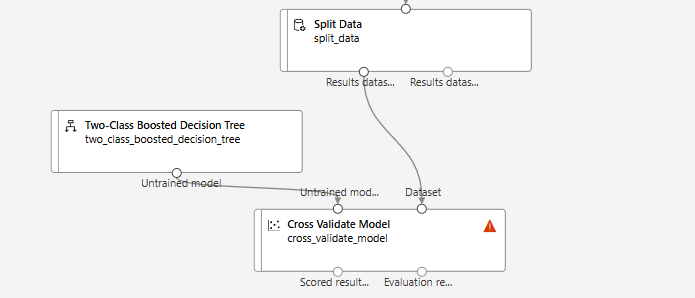


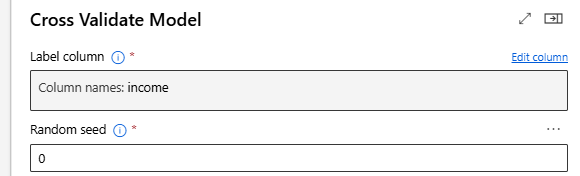
### 모델링/평가

#### 모델링 알고리즘 선택

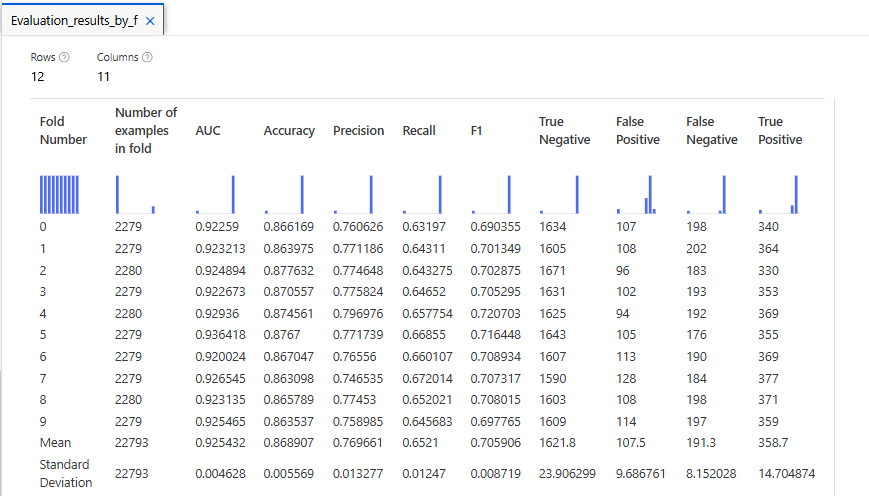


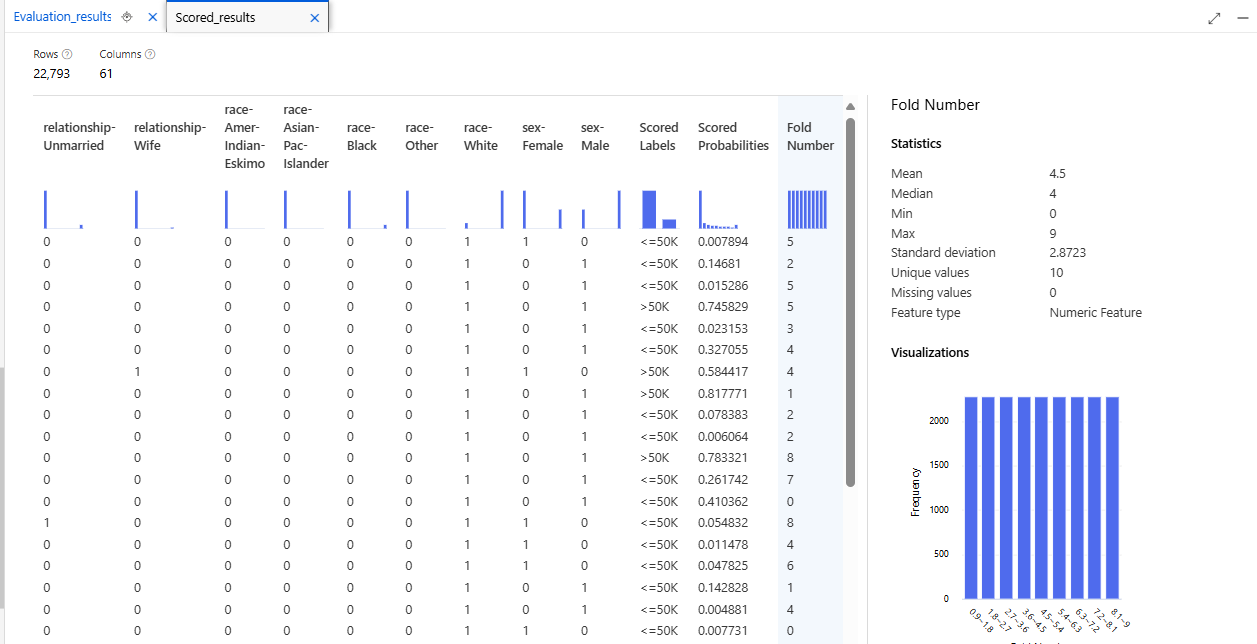
#### 교차 검증





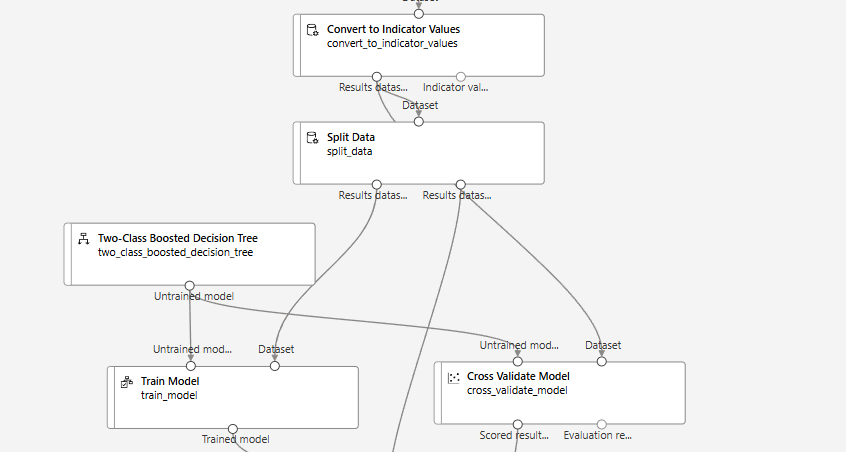
#### 교차 검증 – 폴드 별 결과 확인

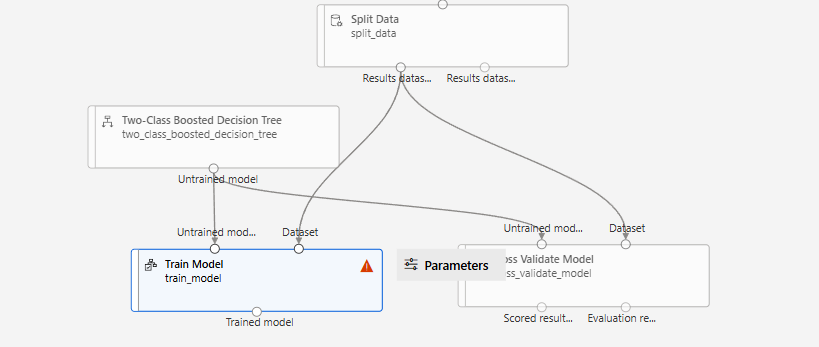


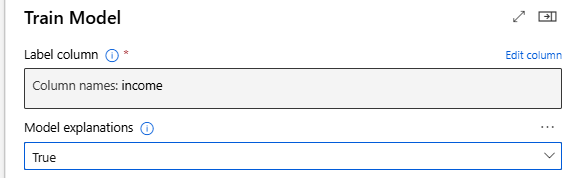


#### 모델 학습(훈련)

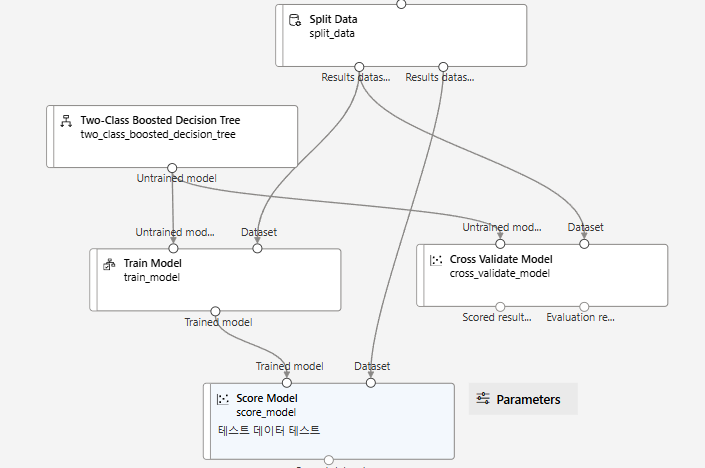
70% 학습 데이터와 전체 데이터 교차 검증을 위해 변경



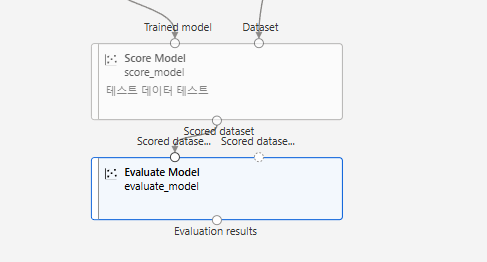




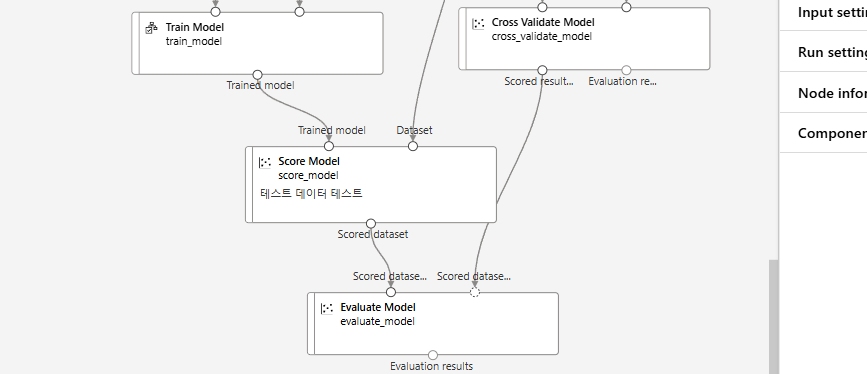
#### 모델 테스트

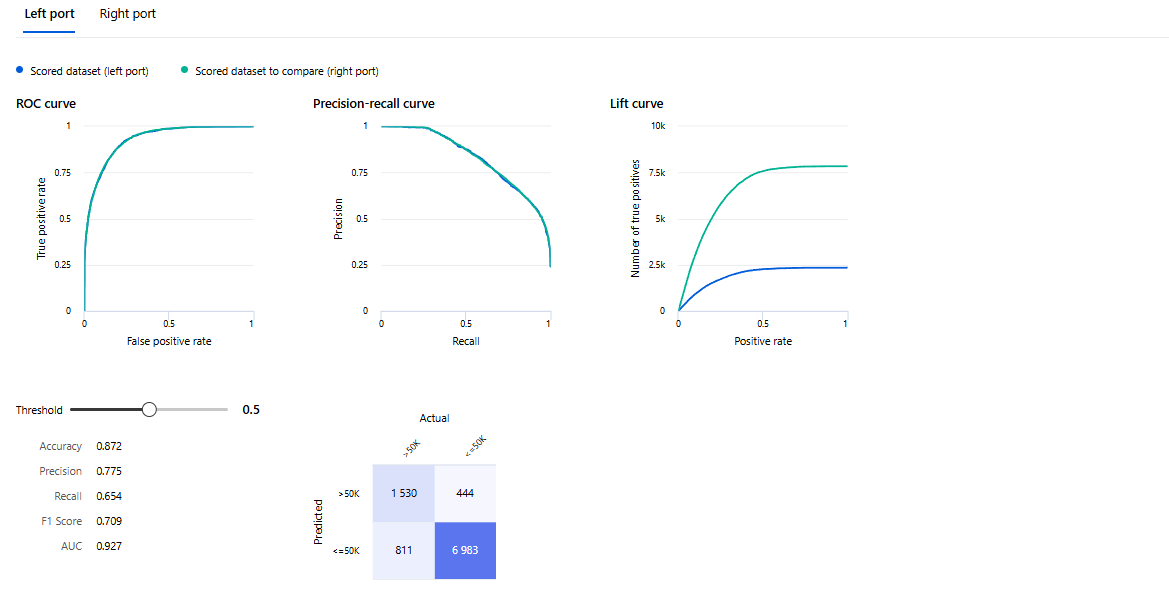


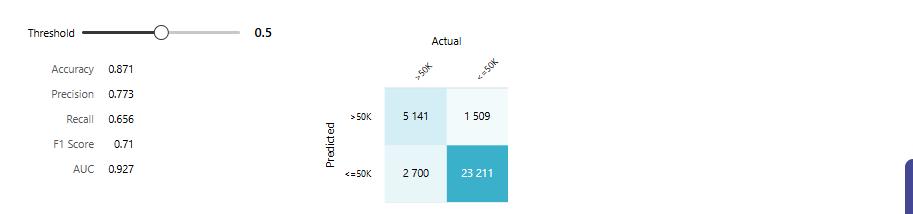
#### 모델 평가



결과 비교를 위해 Cross Validata Model의 Scored result를 Evaluate Model의 두 번째 입력으로 연결







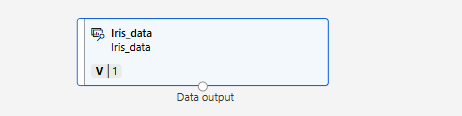
## [실습 - 분류MLD\_붓꽃군집화\_K-means\_로지스틱회귀]

### 실습 준비

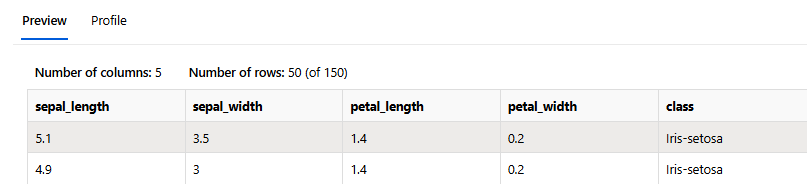
<https://archive.ics.uci.edu/dataset/53/iris> 에서 파일 다운로드

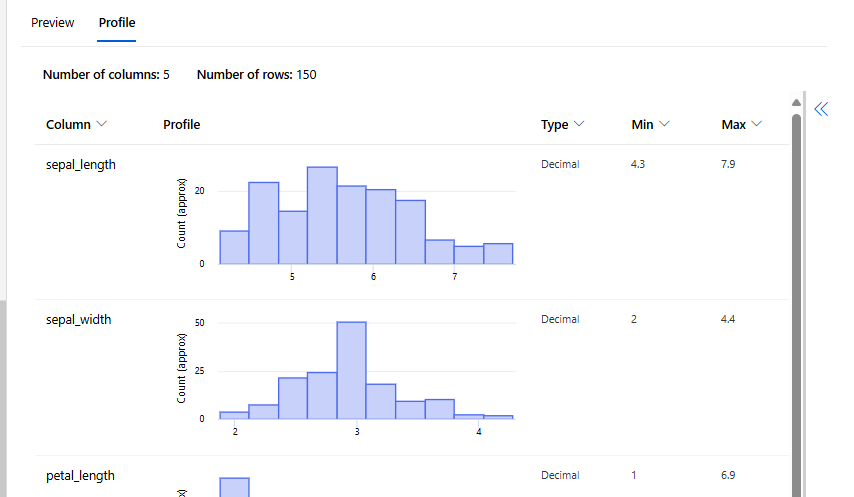
### 데이터 수집/이해

#### 디자니너 시작 및 파이프라인 생성



#### 데이터 이해



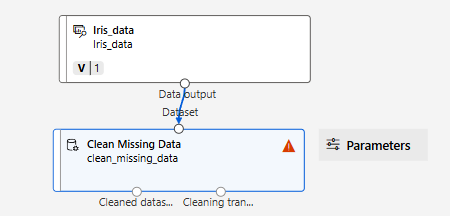


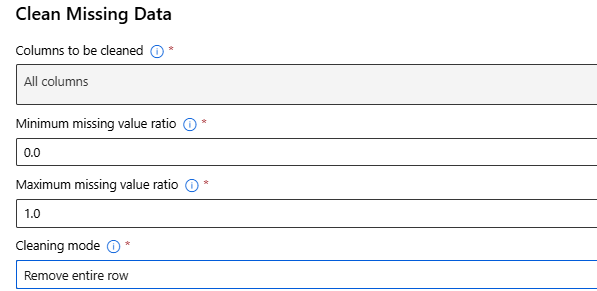
### 데이터 준비

#### 특성 선택

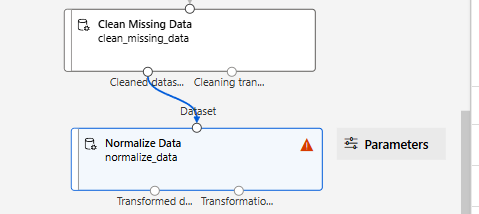
생략

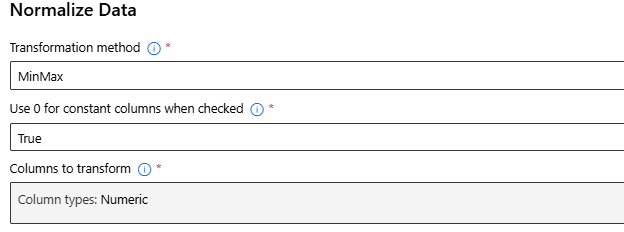
#### 누락값 처리



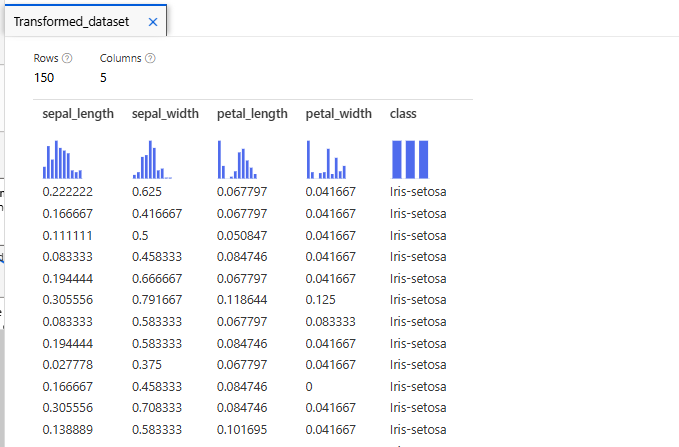


#### 데이터 변환 – 정규화

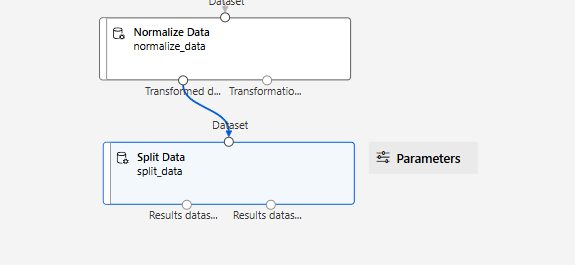


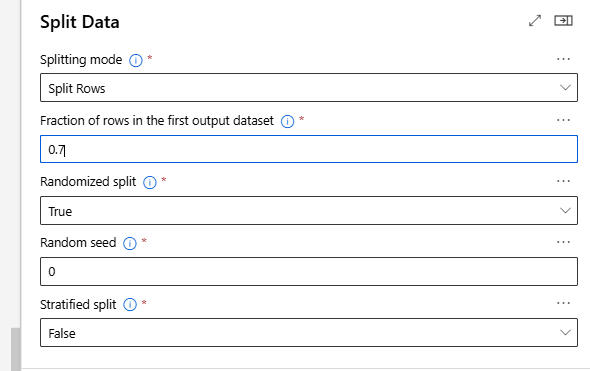


#### 중간 점검



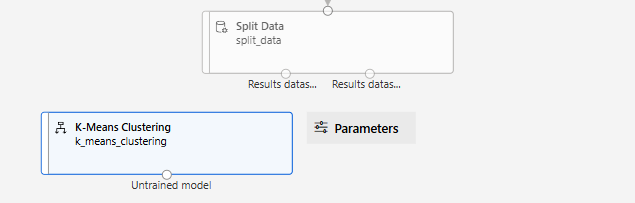
#### 데이터 분리

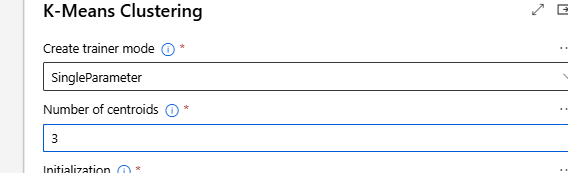


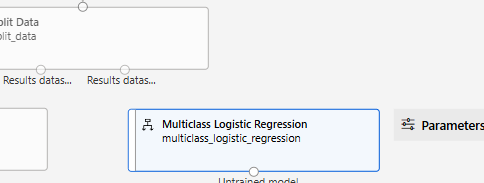


### 모델링/평가

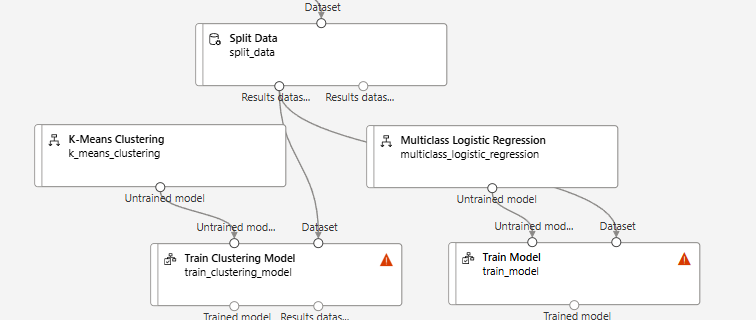
#### 알고리즘 선택

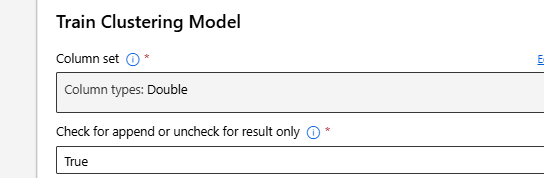




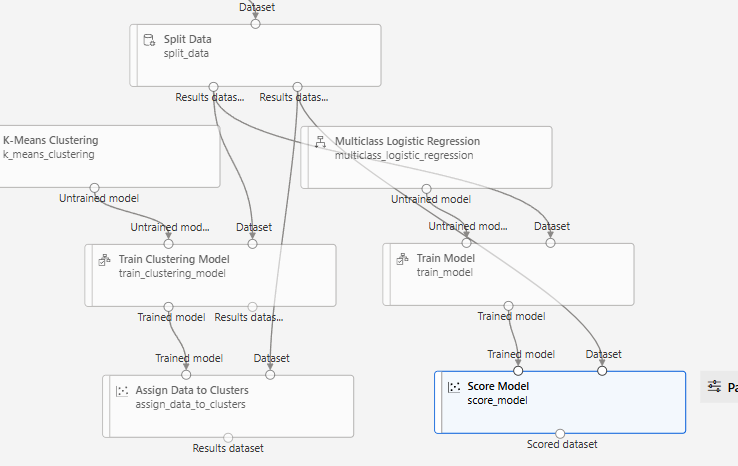


#### 모델 학습(훈련)





#### 모델 테스트



#### 모델 평가 (군집화에 대해서만 평가)

