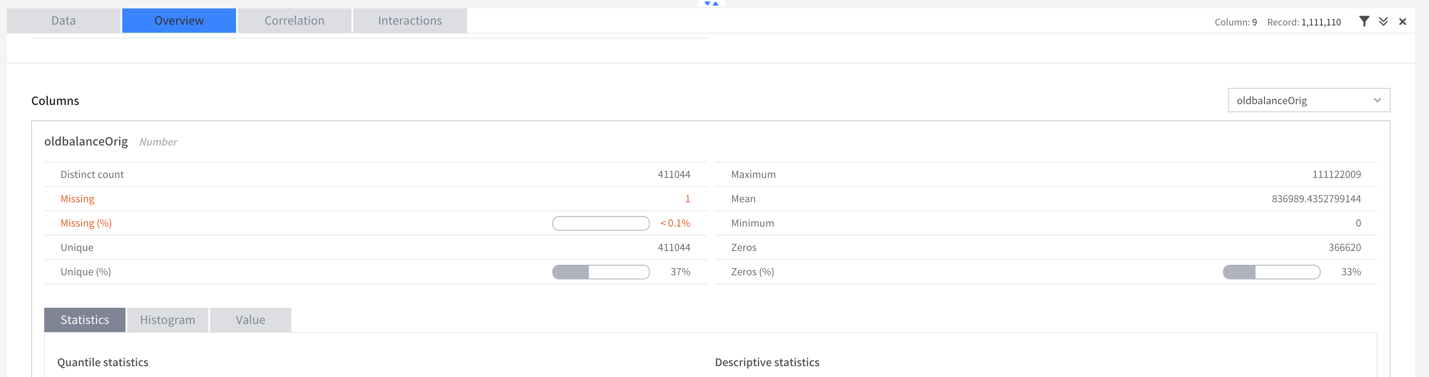
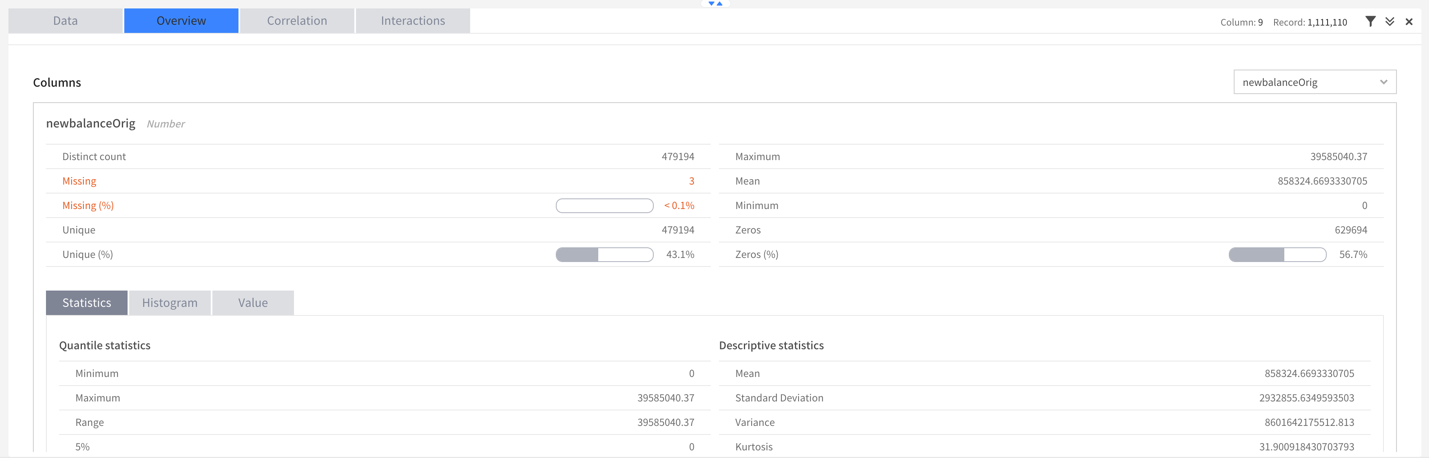
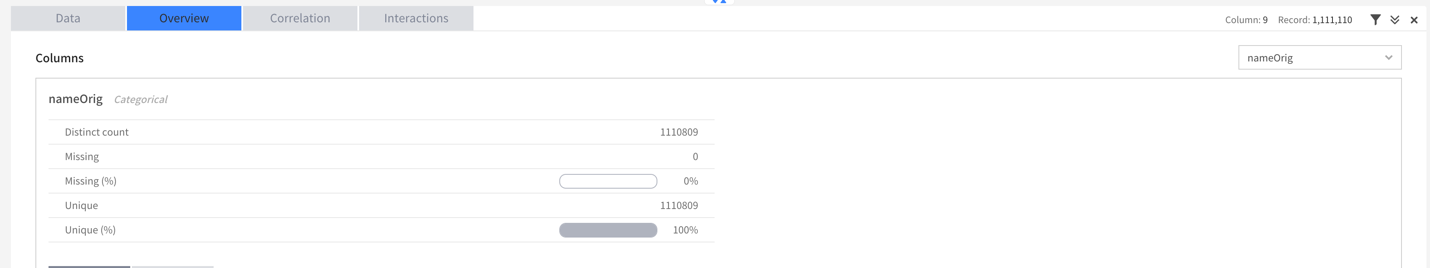
1. 결측치 처리
   1. oldbalanceOrg – missing row가 1개 존재하므로 결측치 처리 필요



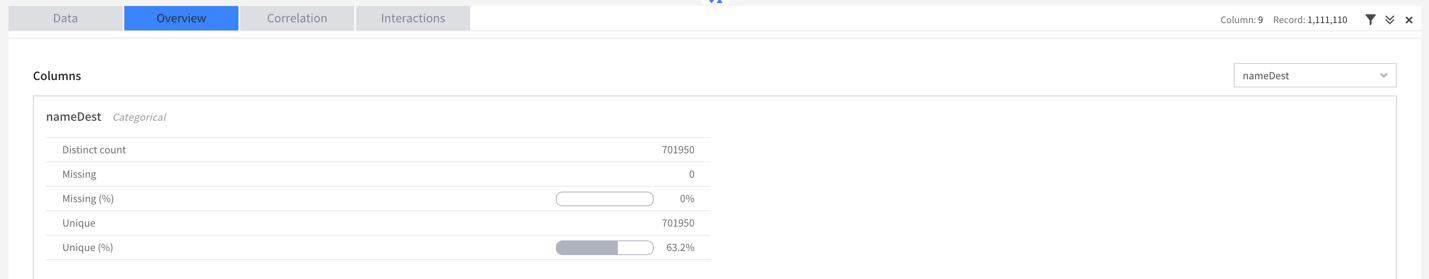
* 1. newbalanceOrig – missing row 가 3개 존재하므로 결측치 처리 필요



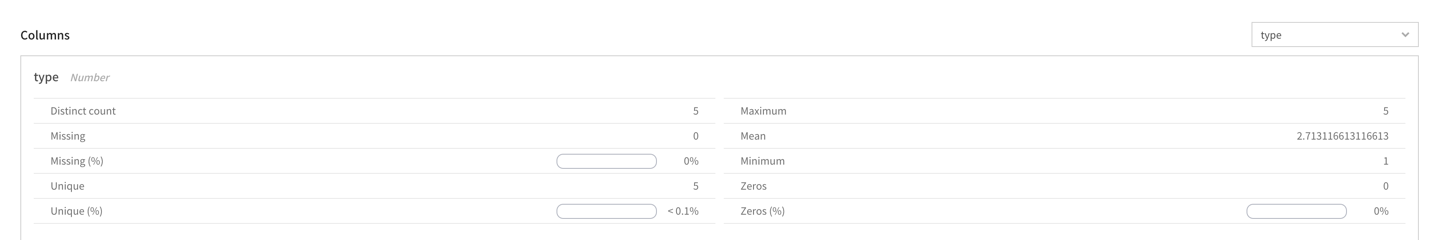
1. 불필요한 데이터 삭제 – 컬럼
   1. nameOrig – Unique가 50% 넘으므로 불필요함 (100%)



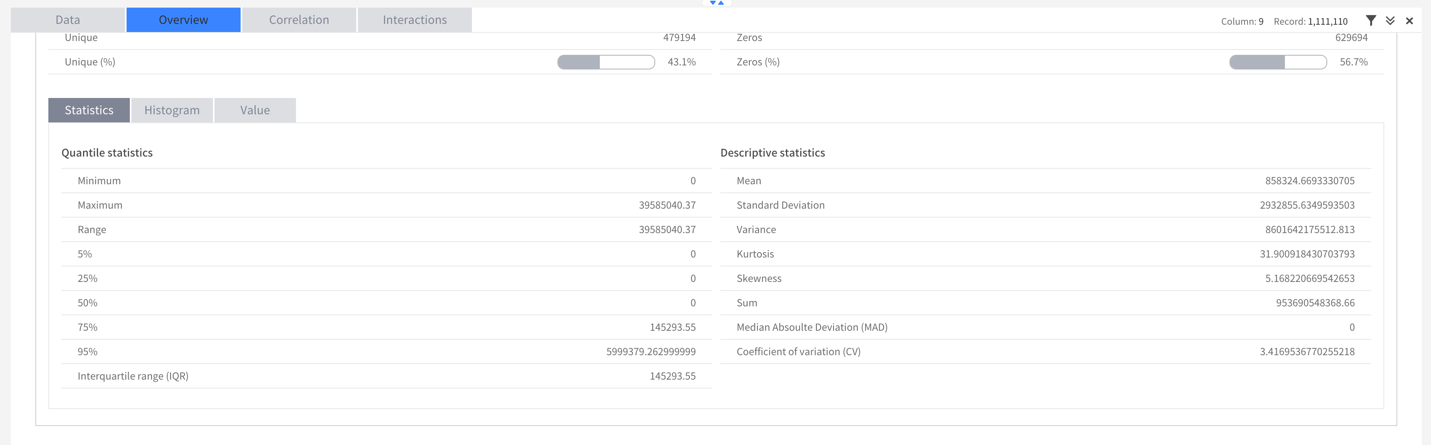
* 1. nameDest – Unique가 50% 넘으므로 불필요함 ( 63.2%)



1. 불필요한 데이터 삭제 – 로우 (Optional)
2. 가변수화
   1. Type 컬럼 – distinct count가 5이고 unique가 0.1% 미만으로 범주형 데이터로 판단되어 가변수화가 필요함



1. 아웃라이어 처리
   1. newbalanceOrig – 조건이 평균+4\*표준편차보다 큰 row 삭제인데, newbalanceOrig의 max값이 이에 미치지 못함, 근거 부족함



nowage@gmail.com