

5. Startup Data analysis (패션 쇼핑몰 추천 스타트업)

○ Data

- `order.head()` → (867, 5)
 - 주문이 일어난 로그데이터
 - `timestamp`(주문시각), `user_id`(주문을 한 유저ID), `goods_id`(상품의 id), `shop_id`(쇼핑몰의 id), `price`(상품의 가격)
- `user.head()` → (10000, 3)
 - `user_id`, `os`, `age`
- `shop.head()` → (200, 4)
 - `shop_id`, `name`(고객이름), `category`(쇼핑몰 분류), `age`(타겟 연령), `style`(쇼핑몰 스타일 태그)
- `log.head()` → (105815, 6)
 - `timestamp`, `event_origin`(이벤트 발생한 앱 위치), `event_name`(발생한 이벤트 명), `event_goods_id`(이벤트 발생한 상품 고유 식별자), `event_shop_id` 등
 - ※ 컬럼별 상세명세는 주피터 파일을 참고 바랍니다

○ Environment

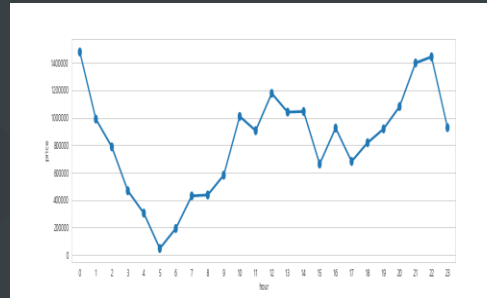
- Python 3.73 with `jupyter==1.0.0`
- pandas, numpy, matplotlib

➤ Perpose

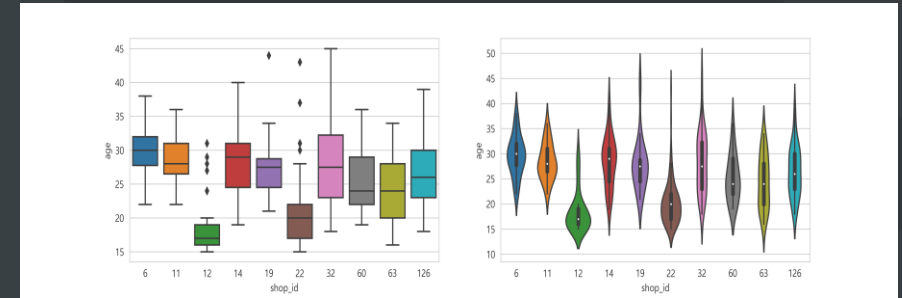
DB 접근(SQL) → 고객 행동데이터 데이터 분석 및 시각화 → 로그 데이터 분석

- 데이터베이스에서 SQL을 통해 고객 정보, 거래 정보, 상품 정보, 그리고 고객의 행동 정보를 로드하여, 데이터 시각화를 통해 데이터를 이해하고 분석하는데 중점을 둠
- 로그 데이터분석의 핵심 수치(page duration, session, 체류 시간)를 구하고 이를 위한 전처리를 진행 후 매출개선을 위한 인사이트를 얻는데 목적을 두고 데이터 분석을 진행

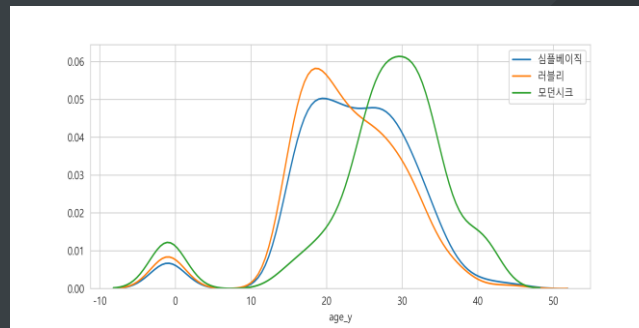
➤ Visualization



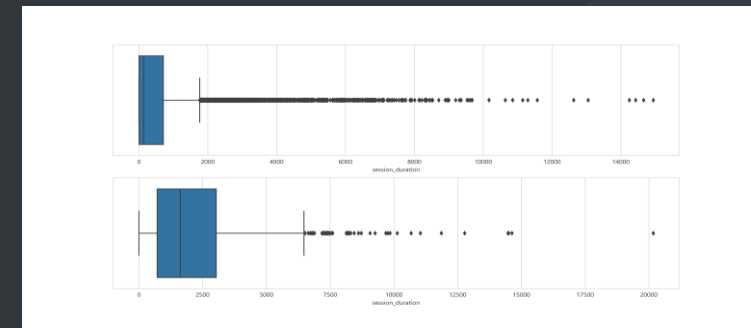
【시간별 Price 구간화(binning)】



【매출 Top 10 쇼핑몰 구매자들의 연령대를 쇼핑몰별로 시각화】



【매출 Top 3 스타일의 구매 연령대 분포 시각화】



【구매/비구매 session별 평균 체류시간 시각화】