**[1조 : R파고]**

**프로젝트 결과 보고서**

프로젝트명: <R파고 쇼핑몰>

**1. 팀 소개**

1조 R파고는 R과 파이썬의 고수가 되어 알파고와 같은 AI를 개발하자는 의미를 가지고 있습니다..

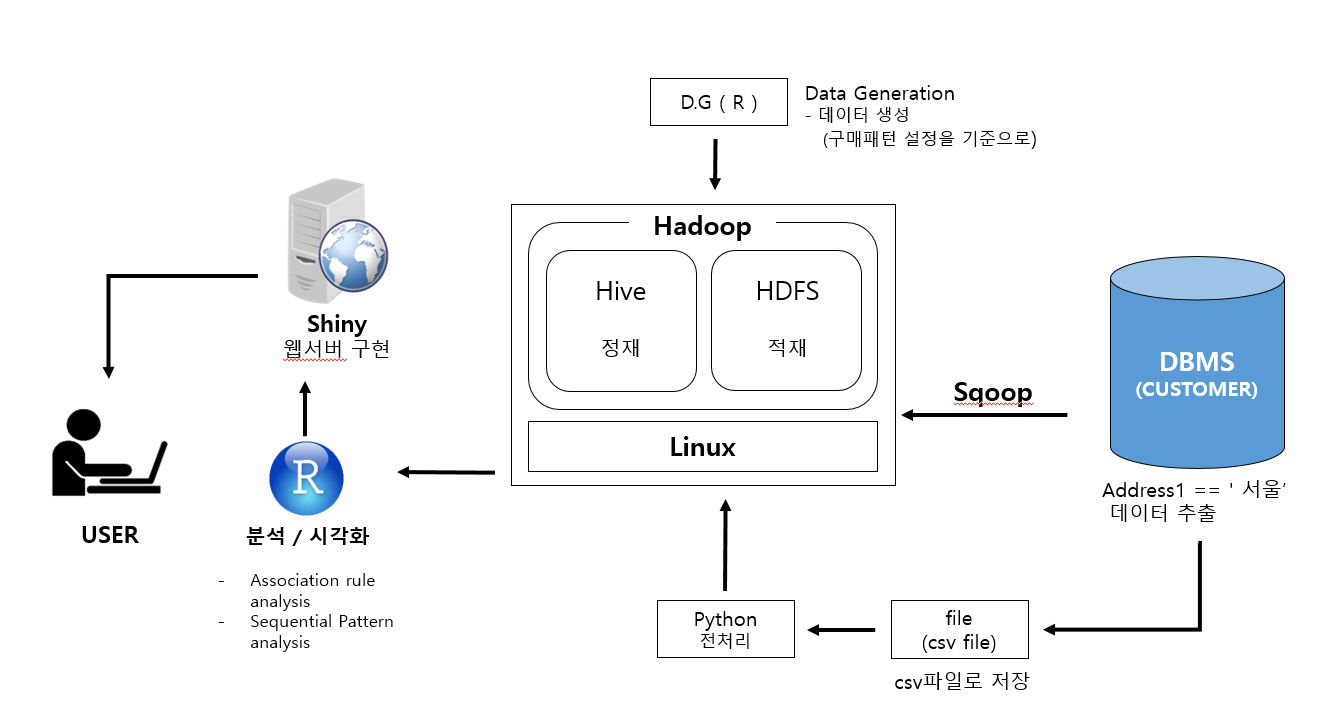
|  |  |
| --- | --- |
| 김진실© | 시스템 구축, 데이터 분석 |
| 김세훈 | 데이터분석, 웹서버 구축 |
| 김한얼 | 시스템 구축, 데이터 분석 |
| 마한올 | 데이터 분석, 발표 |
| 송치원 | 웹서버 구축, 데이터 분석 |

**2. 프로젝트 주제**

쇼핑몰 판매자에게 효율적인 판매를 위해 데이터분석 토대의 컨설팅을 제공하며

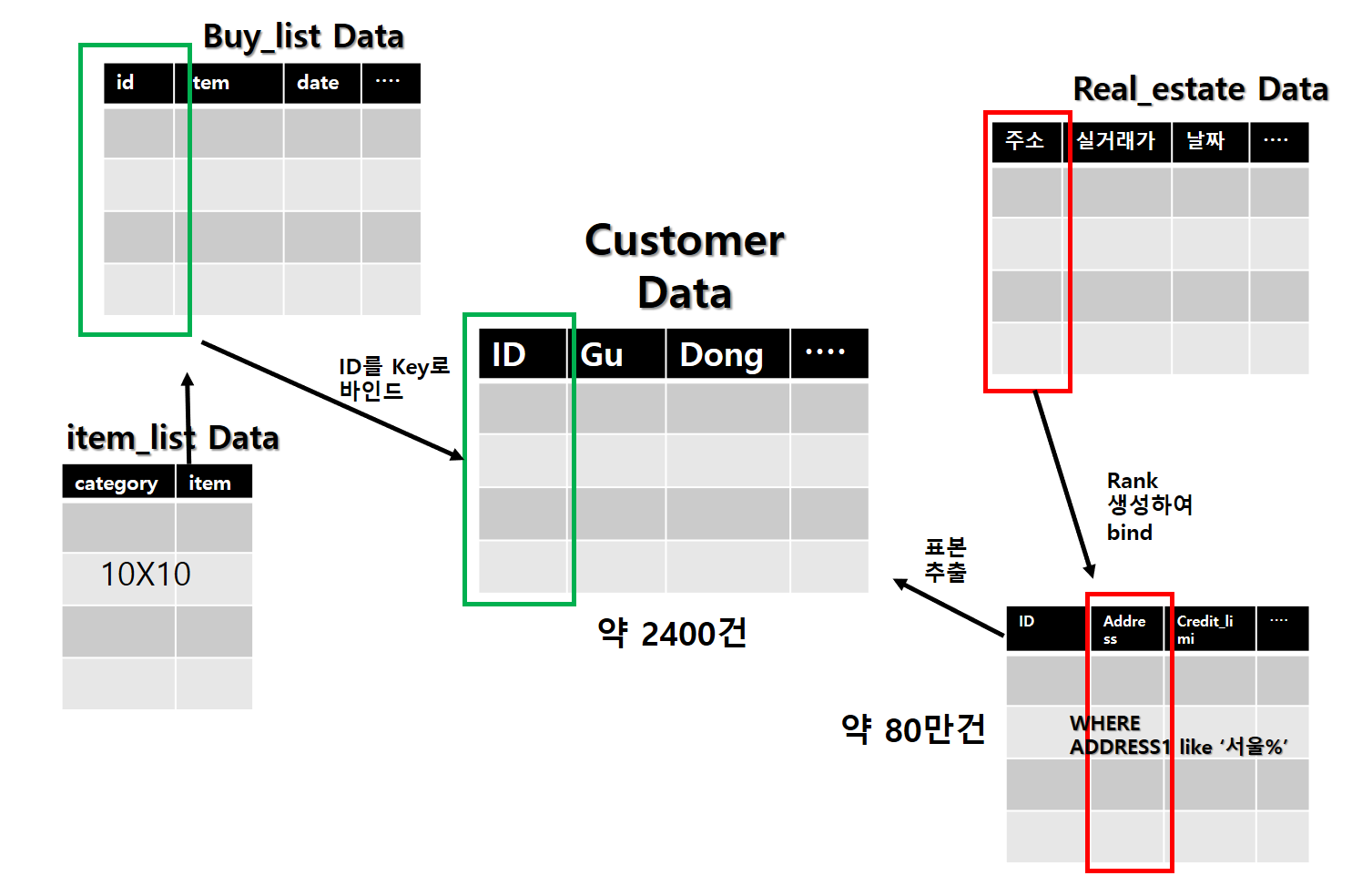
동시에 소비자의 익명성 보장으로 정보 유출의 위험을 줄이는 쇼핑몰 서비스 제공

**3. 시스템 아키텍쳐**



* sqoop으로 oracle DB를 Hadoop에 적재하여 Hive를 통하여 정제하고 Hive와 R을 연동하여 R에서 데이터 분석.

**4. 데이터 구조**



**Customer Data**

1. orcle DB에 customer 데이터 중 서울에 사는 고객들만 추출(전처리)
2. 국토교통부에서 **Real\_estate Data**로 고객들의 자산 수준 1~5등급으로 산정하여 **Customer Data**에 바인드(전처리)
3. 동별 3명의 고객을 표본으로 추출
4. 물품의 카테고리 한 개당 10개의 물품을 배정하여 총 100개의 **item\_list Data** 생성

4-1. 11번가 쇼핑몰의 카테고리와 물품 리스트를 참고하여 생성

4-2. 온라인 쇼핑몰의 물품 개수로 항목에 가중치를 부여하여 현실성 있는 구매리스트가 되도록 함.

1. **Customer Data**의 ID를 key로 고객의 구매개수가 정규분포를 이루도록 랜덤으로 구매 개수 배정하여 **Buy\_lista Data** 생성
2. **Buy\_list Data**에 2015년 1월부터 2018년 10월까지 월별로 row 생성 후 해당 월에 구매를 했으면 ‘o’ , 구매를 하지 않았으면 ‘x’ 를 배정

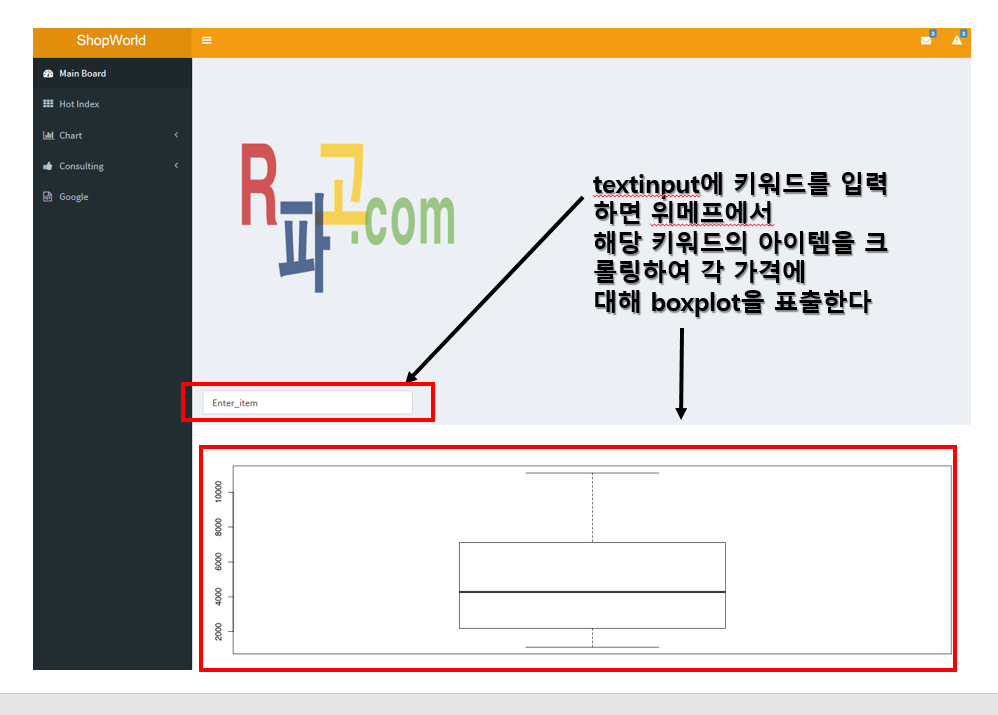
6-1. 지수함수를 사용하여 최근날짜일수록 ‘o’의 개수를 많이 배정

1. **item\_list**를 ‘o’의 개수만큼 list화 하여 **Buy\_list Data**에 물품 배정
2. **Buy\_list Data**를 **Customer Data**의 ID를 key로 바인드

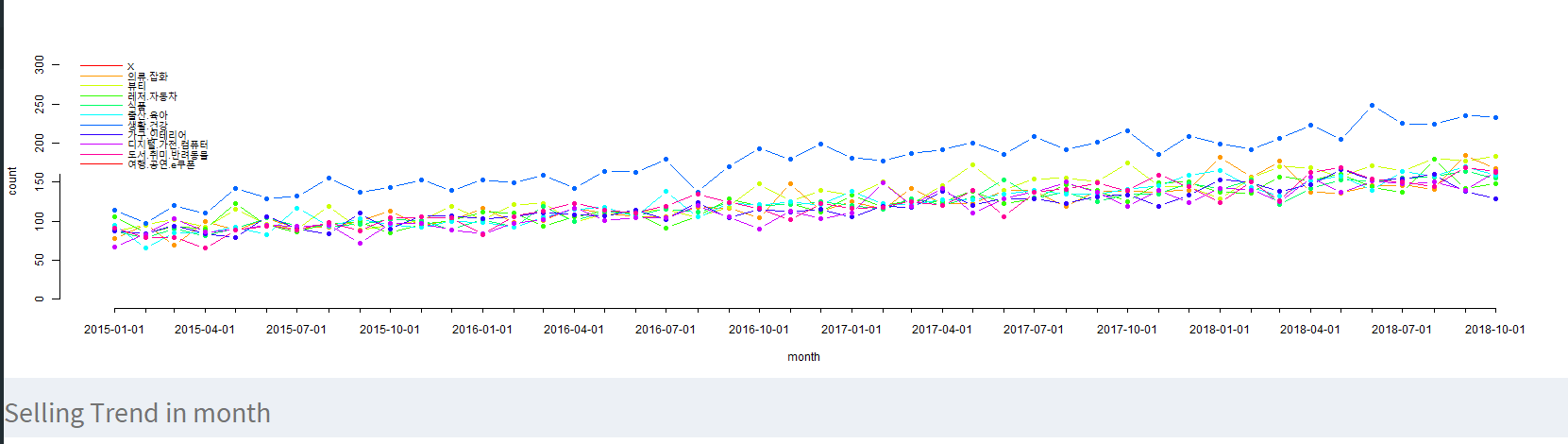
**->위와 같은 작업을 통하여 직관적으로 분석할 수 없는 Customer DataData 를 생성**

**5. 서비스 화면**

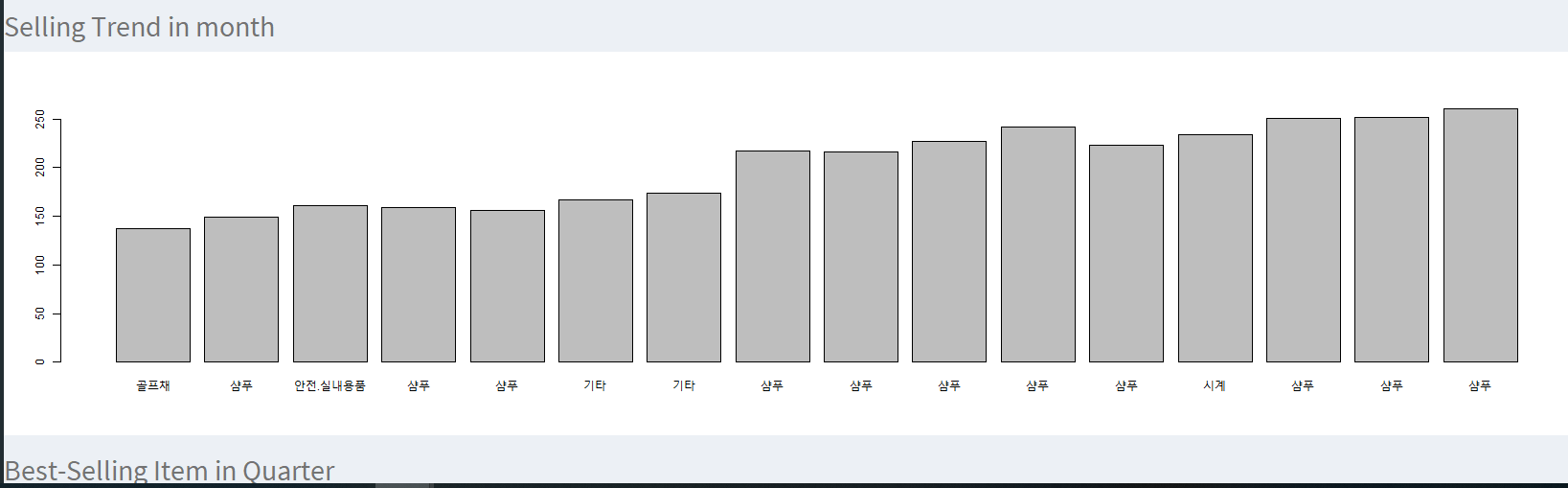
**<1.Main Board>**



**<2.Hot Index>**

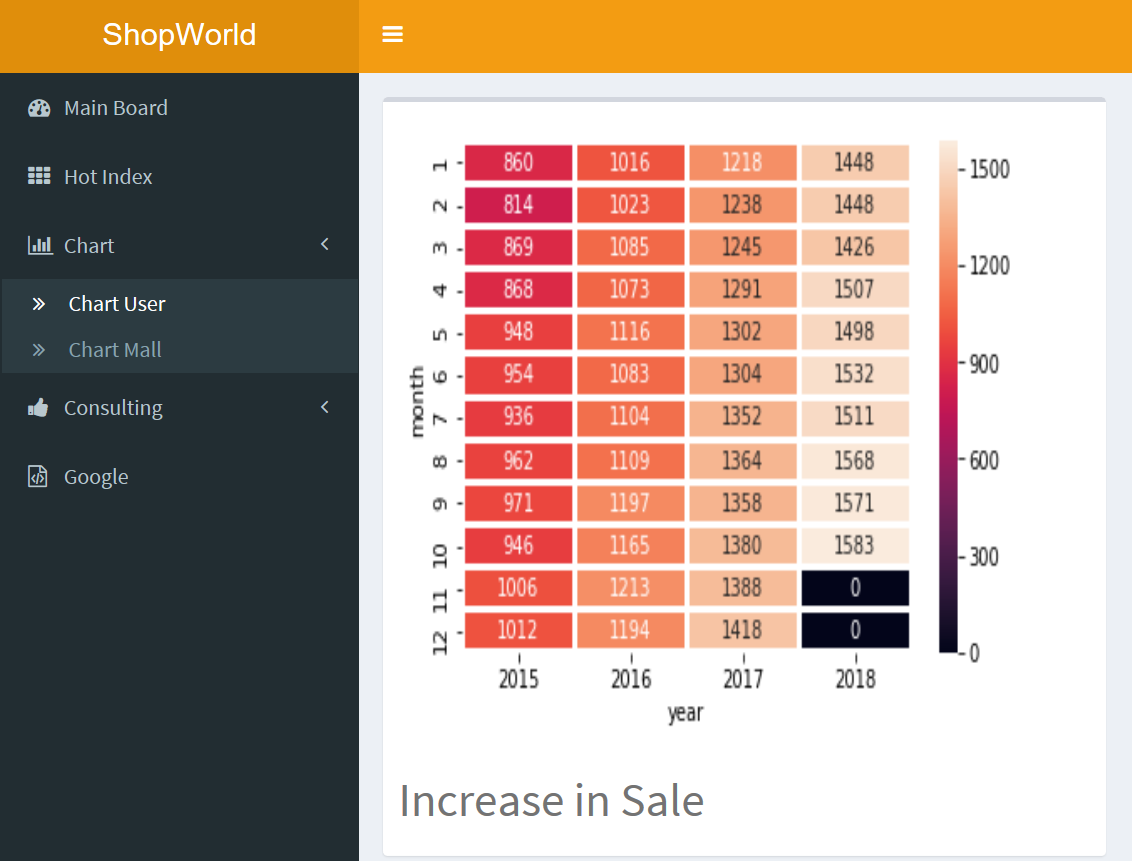


**:**쇼핑몰 운영기간 중 각 품목별 판매량을 추세선으로 표현

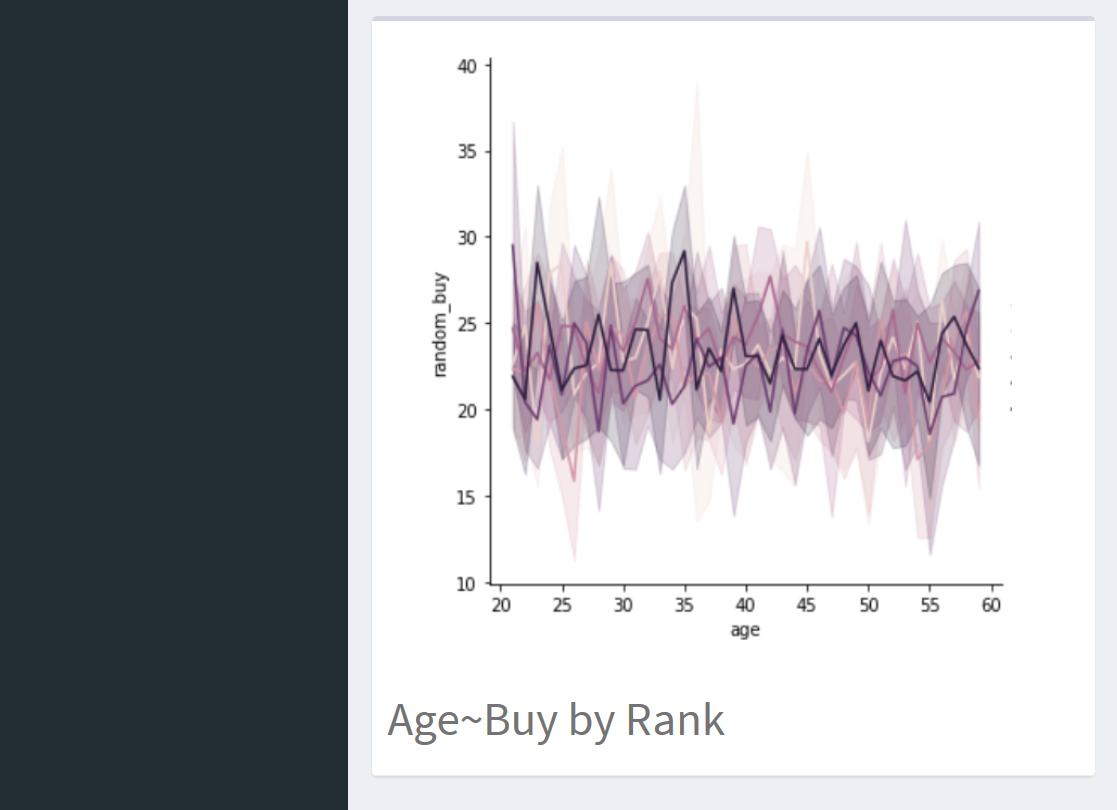


:각 월별 최고 판매 아이템의 이름과 그 판매량을 분석

**<3. Chart -Chart User>**



:연도별 월 판매량을 분석하여 시각화 – 옅은색일수록 많은 판매량



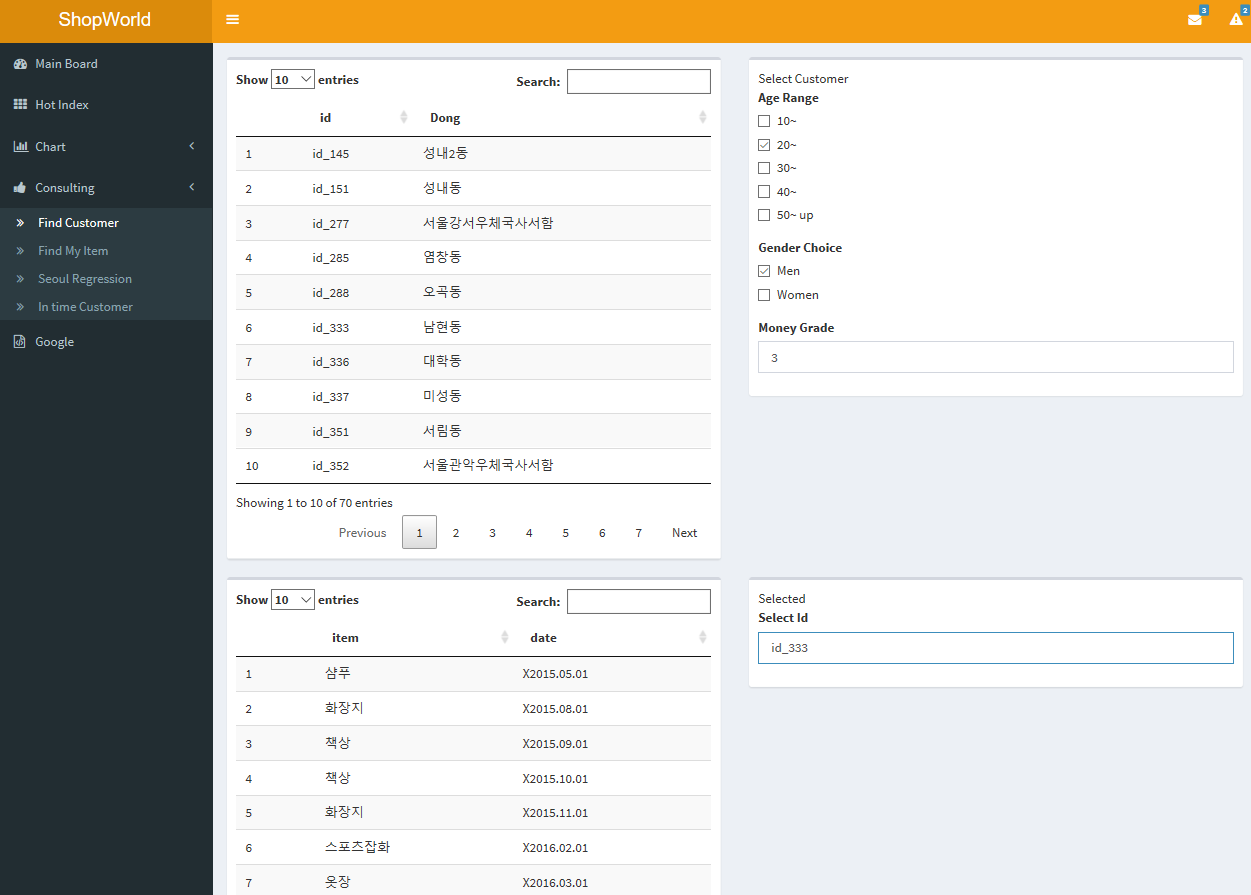
:연령대별 구매 개수와 자산지수 분석하여 시각화

**<3.Chart – Chart Mall>**



:R파고 쇼핑몰의 시간에 따른 소비자 구매량 추세선과 순이익 그래프

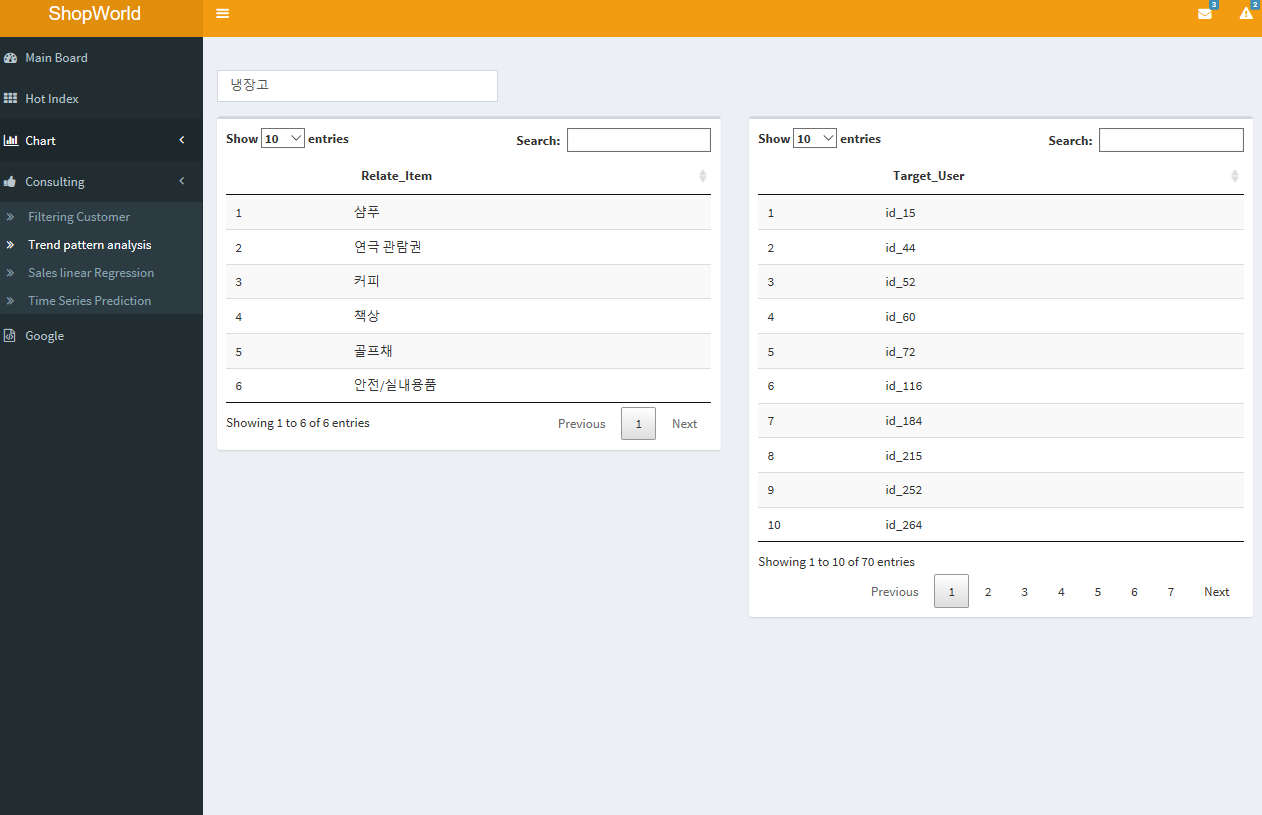
**<Consulting –Find Customer>**



:첫번째 프레임에서는 선택한 범위의 고객들의 인원수와 주소를 보여준다

두번째 프레임에서는 선택한 id의 고객이 구매한 품목과 날짜를 분석해준다.

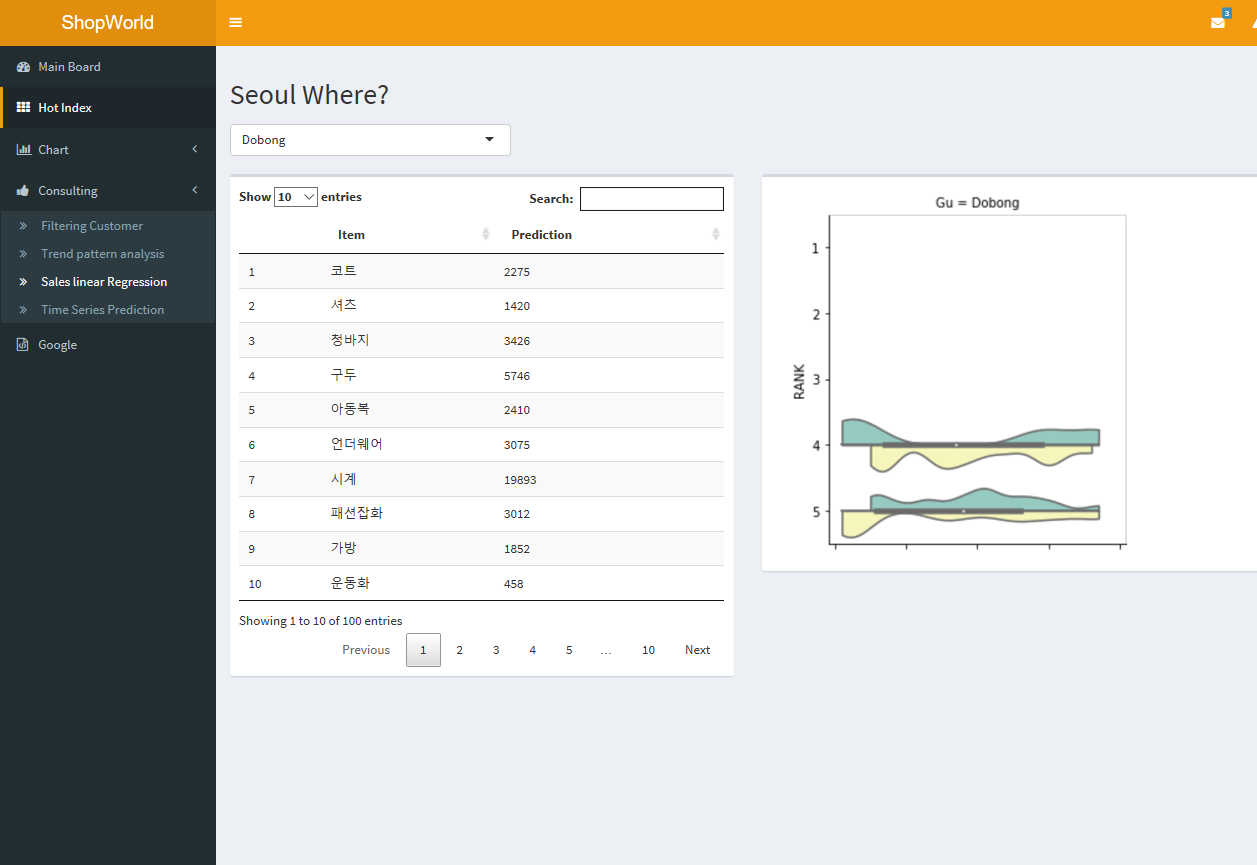
**<4.Consulting – Trend pattern analysis>**



:Input창에 물품을 입력하면 첫번째 프레임에는 추세패턴 분석을 통해 구매 유도력이 높은 상위 6개 제품을 제시.

두번째 프레임에는 첫번째 프레임에 있는 물품들을 구매했음에도 input창에 입력한 물품을 구매하지 않은 잠재 고객으로 고객 ID를 보여준다.

**<4.Counsulting- Sales linear Regression>**

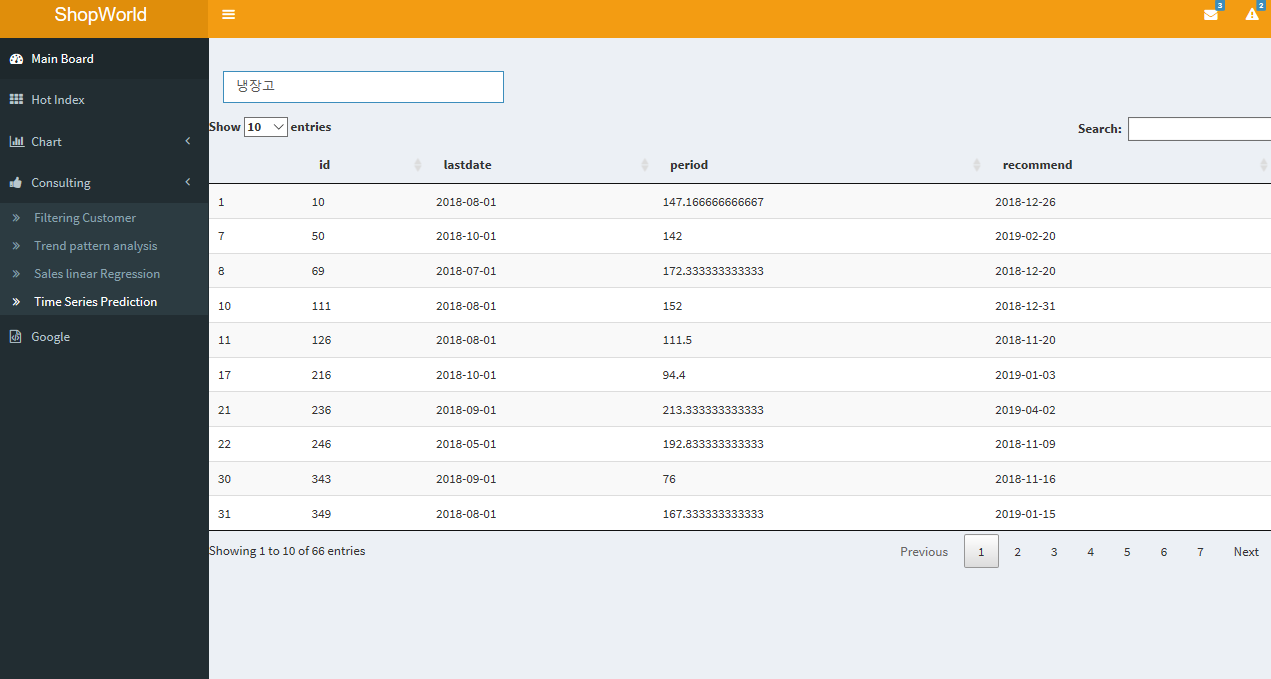


: 데이터 선택창에서 구의 이름을 선택하면

첫번째 프레임에서는3년치 아이템별 판매량을 구별로 평균내어 그것을 다시 서울 구별 전체인구에 곱해, 2018년11월 구별 아이템 수요량을 예측

두번째 프레임의 그 구의 연령대별 분포도와 자산 순위를 그래프로 분석한 시각화 자료를 제공

**<4.Consulting – Time Series Prediction>**



:input창에 물품을 입력하면 그 물품을 3번이상 구매한 고객들을 추출하여 구매한 기간을 분석하여 다음 구매 날짜를 예측해준다.

**7. 개발**



