



CP1_Project(AI12_2팀_황건하)

1. 서론

a. 프로젝트의 배경

i. 리셀테크 용어의 정의

리셀테크란 되판다는 뜻의 영어단어 ‘**resell**’ 과 ‘**재테크**’를 합친 말로 한정판 상품을 구입한 후, 더 높은 가격으로 되팔아 차익을 얻는 재테크 방식을 의미한다. 중고거래가 제품을 더 저렴하게 구입하게 이용된 반면 리셀테크는 이를 응용하여 수익까지 창출할 수 있게 되었다.

리셀테크를 위해 한정판 상품을 구매하는 사람들은 ‘리셀러’라고 불리며, 이 리셀러들은 백화점 앞에 이른 시간부터 줄을 서거나 한정판 상품에 대한 정보를 빠르게 입수해 희소성이 있는 상품들을 구입한다.

ii. MZ 세대는 왜 리셀테크에 열광하는지

리셀테크가 mz세대에게 이토록 인기있는 이유는 첫째로 **진입장벽이 낮다**는 것이다. 리셀테크에서는 주식처럼 경제의 흐름을 파악하거나 전문 지식을 필요로 하지도 않는다. 리셀테크 플랫폼에 접속만 할 수 있다면 누구든지 쉽게 리셀테크를 시작할 수 있는 것이다. 또한, **낮은 진입장벽에 비해 수익률은 매우 높다**는 점 역시 리셀테크가 인기있는 이유 중 하나이다. 희소 가치가 높고 수요가 많은 상품일수록 판매 가격은 점점 더 높아지기 때문에 단기간에 큰 수익을 낼

수 있다. 더불어 최소한 상품의 본래 가격 만큼은 보장받을 수 있기 때문에 손해 비용 역시 거의 발생하지 않아 다른 재테크들에 비해 훨씬 안정적이다.

마지막으로 바로 **mz세대의 소비 성향**이다. mz세대는 본인이 추구하는 방향과 맞는 제품이라면 과감하게 소비하는 ‘가치소비’ 성향을 띄고 있다. 이에 따라 자연스럽게 mz세대의 명품 구매 역시 증가하고 있으며 희귀하고 고급진 한정판 제품의 수요 역시 꾸준히 증가하고 있다. 또한 mz세대는 브랜드 자체에 관심이 많고 브랜드에 대한 친숙도 역시 높아 희소가치가 있는 제품을 빠르게 알아본다. 이러한 mz세대의 성향은 자연스럽게 리셀테크에 대한 관심으로 이어져 결국 mz세대의 대표 재테크 방법으로 급부상하고 있다.

b. 프로젝트 목적 설명

i. 목적 및 기대효과

재밌게 해 볼 수 있는 ‘신발 제테크’를 통해 **리셀 플랫폼 인식 확대**하고 **고객들이 보유한 신발의 잠재 가치를 쉽게 알 수 있게** 하여 **신규 고객을 확보 및 다방면으로의 리셀 시장 활성화**를 목표로 한다. 또한, 트렌드 분석을 통해 리셀 플랫폼 고객의 니즈(수요)를 파악하여 적합한 마케팅 전략을 수립하고자 한다.

2. 연구 및 개발 방법

a. 연구 및 개발에 필요한 데이터 셋 소개

i. 어떠한 데이터가 왜 필요한지

- 리셀 플랫폼의 신발 데이터
 - 상품명, 거래가, 소매가, 브랜드, 거래 날짜 등
 - 리셀 플랫폼의 신발 거래 정보를 활용해 고객이 가지고 있는 신발 브랜드 및 가격을 쉽게 예측하여 이용자에게 필요한 정보를 제공하기 위해

ii. 데이터의 수집 방법 혹은 출처

- 훈련, 테스트 데이터셋
 - 캐글에 공개되어 있는 리셀 플랫폼(StockX) 데이터 수집
- 트렌드 분석을 위한 데이터셋
 - 리셀 플랫폼(Kream)에서 동적 웹 크롤링을 사용하여 수집

iii. 데이터 설명

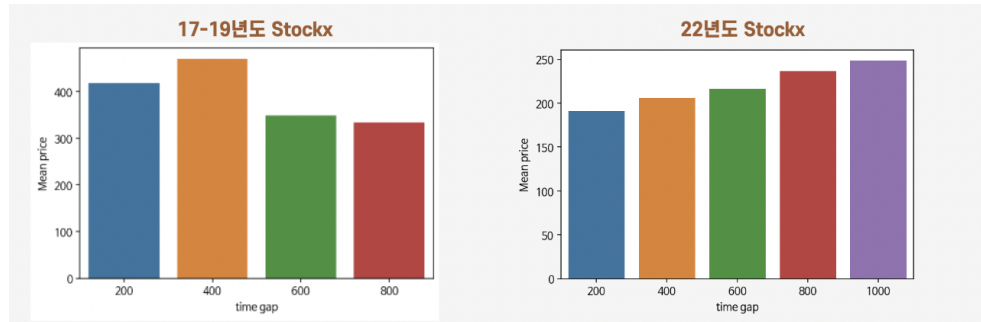
- train data
 - kaggle, StockX 2019 Data Contest에서 제공한 데이터, 10000개의 신발 데이터
(99956 rows × 8 columns)
 - columns : 주문일, 브랜드 이름, 신발 상품명, 판매가, 소매가, 출시일, 신발 사이즈, 구매자 지역
- test data
 - kaggle, StockX 최근 출시된 신발 데이터 (1000 rows × 17 columns)
- 트렌드 분석에 사용한 데이터
 - Kream 사이트 크롤링

iv. 전처리 및 시각화(EDA)

- 전처리
 - 출시일로 부터 시간이 지날 수록 판매가격이 낮아질거라고 추정하여, 주문일에서 출시일을 빼주는 `time gap` column 생성
 - 상품명에서 브랜드 이름을 추출하여 숫자형 데이터로 칼럼 생성
 - 비슷한 Sneaker Name(상품명)들 사이즈, 색상 차이로 가격이 비슷한 경우가 많아 이름을 하나로 통일
 - 모델링을 위해 Sneaker Name(신발 상품명)을 원핫인코딩
 - datetime 날짜 형태인 Order date를 모델링을 위해 정수 형태로 변경

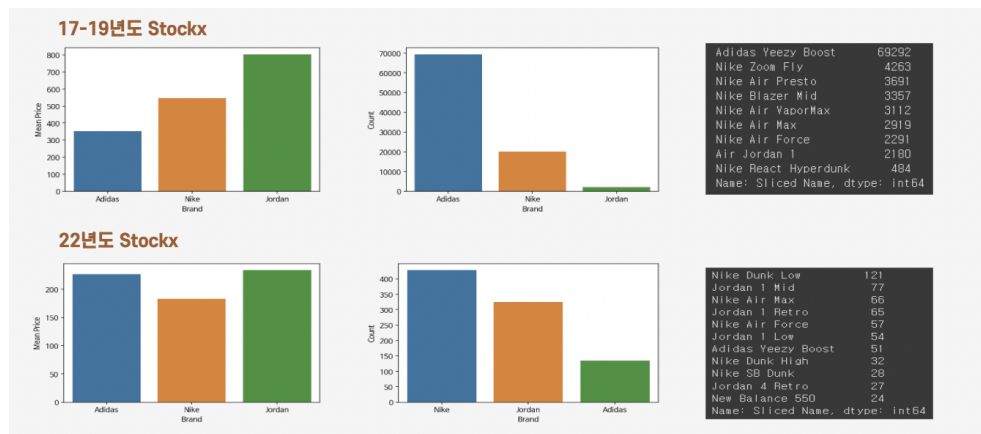
- 시각화(EDA)

[출시된지 200~1000일 사이에 판매된 제품의 평균가격 차트]



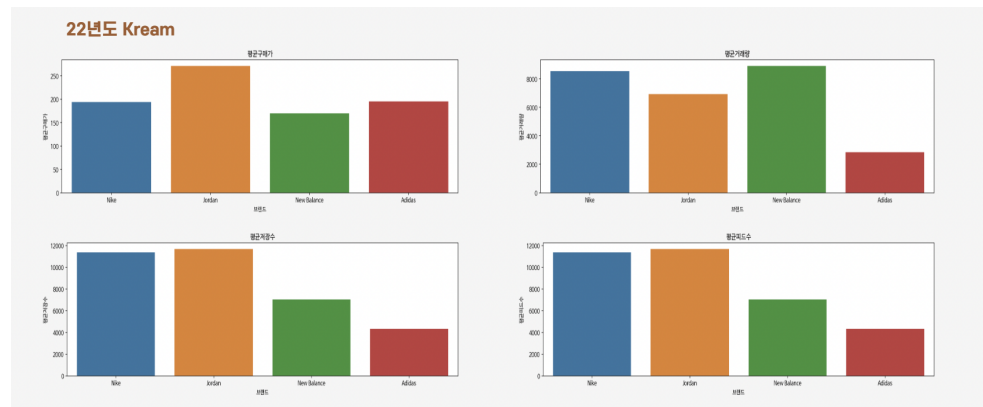
- 17-19년도의 Trend는 출시된지 얼마 지나지않았을 때에는 가격이 올랐다가,시간이 지나면서 점점 가격이 감소하는 경향을 보이며 오른쪽 차트의 최근 22년도에는 시간이 지날수록 평균가격이 점점 올라감을 알 수있다.
이를 보아 **과거**에는 출시된지 얼마 지나지않은 상품의 가치가 높게 평가되었고 반면에, **현재**는 시간이 지날수록 증가하는 희소성에 더 높은 가치를 부여하는 것으로 분석했다.

[브랜드(Adidas,Nike,Jordan)별 평균가격, 점유율(거래량) 차트]



- 17년도에는 평균가격이 조던이 가장 높고 아디다스가 가장 낮게 형성되었고 점유율은 아디다스가 월등히 높게 나왔다. 특히 아이다스의 이지부스트 제품군이 대부분을 차지하고있다.
22년도에는 나이키에 비해 아디다스와 조던이 조금 높은 가격대를 형성했고, 점유율은 과거 1등이었던 아디다스가 꼴지로 떨어졌고 나이키가 1위로 올라왔다.
이는 희소성있는 제품을 다루는 리셀 플랫폼의 특성상 고객들이 너도 나도 신고다니는 아디다스가 아닌 값이 상대적으로 저렴하고 대중들에게 희소성있는 브랜드로 니즈가 이동한 것을 보여준다.

[국내 리셀플랫폼 Kream의 브랜드별 평균구매가,거래량,저장수,피드수 차트]



- 다음으로 **국내 리셀시장**의 트렌드 파악을 위해 국내 리셀플랫폼 Kream에서 크롤링을 수행했습니다.
재밌는점은 우측에 있는 2개의 차트를 보시면 거래량이 높을 수록 인스타에 올린 게시글 수도 높은것이 당연하지만, 조던이 거래량에 비해 피드수가 상대적으로 높고 뉴발란스가 상대적으로 낮은 것을 알수 있습니다. 이는 **sns를 좋아하는 인플루언서들 에게는 조던과 나이키가 더 선호된다고** 유추했습니다.
또한, 왼쪽아래 차트에서 나이키와 조던이 평균 저장수가 높았고 이를 통해 **고객들이 나이키와 조던을 금액적인 측면때문에 선불리 구매하지 못하고 저장을 했을 가능성과 혹은 나이키,조던 브랜드를 이것저것 찾아보는 것을 좋아하고 저장해놓는 골수팬들이 나이키와 조던에 더 많다고** 유추해 보았습니다.

b. 연구 및 개발에 필요한 기술 스택 정리

i. 프로젝트에서 자신이 담당한 기술의 정의

● 셀레니움(selenium)

셀레니움은 웹 애플리케이션 테스트를 위한 프레임워크이다.

웹에 하는 명령을 코드화시켜서 작동시키고 다양한 브라우저 작동을 지원하며 크롤링에도 활용된다. 또한 현존하는 거의 모든 웹 브라우저를 다양한 언어를 통해 제어가 가능하다.

크롤링에서는 정적, 동적 페이지 크롤링으로도 접근이 불가능한 데이터에 접근할 때 유용하게 사용된다.

- 랜덤포레스트(RandomForest)

랜덤 포레스트는 다수의 결정 트리들을 학습하는 앙상블 방법이며 검출, 분류, 그리고 회귀 등 다양한 문제에 활용되고 있다. 의사결정 나무를 여러 개 사용하여 문제를 해결하는 방식으로 작동한다. Decision Tree는 나무 하나로 예측을 했다면 Random Forest는 여러 나무를 통해 종합된 결과를 바탕으로 예측한다. 학습 데이터의 부분집합을 랜덤하게 만들어 의사결정 나무를 구축하는 배깅(Bagging) 계열의 트리 모델이다.

- ii. 해당 기술(또는 연구방법)의 최근 장점과 단점

- 셀레니움(selenium)

- 셀레니움에서 제공하는 메서드는 실제 동작을 시뮬레이션해주어 마우스로 직접 클릭해야만 보이는 동적 페이지도 호출해준다. 셀레니움 실행 시 실시간으로 확인하면서 크롤링 및 자동화 동작을 확인할 수 있다.
- 컴퓨터 사양에 따라 느릴 수 있고, 브라우저를 직접 켜서 움직이므로 자원을 많이 쓰게 될 수 있다.

- 랜덤포레스트(RandomForest)

- 랜덤 포레스트는 다양성을 극대화 하여 예측력이 우수한 편이고 다수의 트리 예측 결과를 종합하여 의사결정을 진행하기 때문에 안정성도 높다.
- 과적합을 방지할 수 있고 결측치에 대해 강건하다.
- 데이터의 수가 많아지면 의사 결정나무에 비해 속도가 크게 떨어지고 결과에 대한 해석이 어려운 단점이 있다.

- iii. 기간 내 구현 가능한 범위 설정

1. Feature Engineering, EDA, Modeling(머신러닝)
2. 데이터 수집 및 크롤링
3. 데이터베이스 적재
4. 대시보드 생성
5. 웹 서비스 구현 및 배포

c. 개발 일정

구분	기간	활동	비고
사전 기획	6/24(금) ~ 6/28(화)	- 프로젝트 기획 및 주제선정 - 기획서 작성	- 아이디어 수집
데이터 수집	6/28(화) ~ 6/30(목)	- 데이터 크롤링 - 필요한 데이터 추가 수집	- 팀별 회의
데이터 전처리	6/30(목) ~ 7/3(일)	- 데이터 정제 및 EDA	- 팀별 회의
모델링	7/2(토) ~ 7/5(화)	- 베이스라인 모델 제작 - 모델 구현 및 보완	- 팀별 회의 - 발표자료 제작
서비스 구축	7/2(토) ~ 7/6(수)	- 웹서비스 구현 - 대시보드 제작	- 프로젝트 마무리
총 개발기간	6/24(금) ~ 7/6(수) (약 2주)	-	-

d. 협업 방식

- 팀원 구성 및 역할

AI12_김나연: 데이터 수집, 데이터 전처리/EDA 및 모델링, ppt 제작, 발표

AI12_송나현: 데이터 수집, 데이터 전처리/분석 및 모델링, ppt 제작, 발표

AI12_안원석: 데이터 수집, 웹 서비스/ 대시보드 구현 및 배포, 발표

AI12_황건하: 데이터 수집, 데이터 전처리/EDA 및 모델링, ppt 제작, 발표

- 협업 방식

- Discord

디스코드에서 자체 팀 채널을 만들어 사용하고 있으며 실시간으로 자료 및 의견을 공유하고 문제상황이 발생할 경우, 디스코드를 통해 소통

- Gather Town

게더타운에서 공식 미팅 시간에 팀원과 코치님 의견 공유

3. 결론(고찰)

a. CP1 프로젝트 전반에 대한 고찰 진행

i. 연구 및 개발 통해 알게된 점

분석을 통해 상품의 가격 변동이 어떻게 일어나는지 보면서 꽤나ダイナミック하게 가격이 변동되는 것을 알았고, 입찰 거래들도 활발하게 이루어짐을 알았다.

같은 세대의 고객들에게 인기 있는 플랫폼을 분석하고 신규 고객들은 더 쉽게, 기존 고객들은 더욱 활성화시키는 방향으로 리셀 시장에 진입할 수 있게 도와주는 웹 어플리케이션 서비스를 만듦으로써 힘들었던만큼 보람차고 성취감을 느꼈던 프로젝트였다.

처음으로 분석 및 개발의 팀 단위 프로젝트를 수행하며 업무, 역할 분담과 의견 조율, 아이디어 공유, 협업의 과정통해 개인이 혼자 하는 프로젝트보다 더 어렵고 힘들기도 했지만 긴 시간동안 의지하며 가까워졌고 프로젝트를 제출하고 다 같이 기뻐한 기억은 잊지 못할 경험이며, 성과 또한 혼자서 했으면 절대 이루지 못할 성과를 거두었다.

ii. 프로젝트 결과를 사회 혹은 비즈니스에 어떠한 방식으로 활용해 볼 수 있는지

본 프로젝트에서는 희소성이 높은 상품의 가격이 시간이 지남에 따라 증가하고 있음을 확인하였고, 세부적으로 어떤 브랜드의 상품이 많이 거래가 되고 높은 가격대를 형성하는지 알아봄을 통해 리셀플랫폼 트렌드를 파악할 수 있었다.

따라서, **고객의 브랜드 및 상품의 가격을 쉽게 예측**하여 **이용자에게 필요한 정보를 제공하고 본 프로젝트의 웹서비스를 통해 신규 고객을 확보**하는 마케팅 전략 수행에 활용할 수 있다.

iii. 연구 및 개발을 진행하며 수행하지 못했던 부분 및 아쉬웠던 부분

- db를 웹사이트와 연동시켜 대시보드 개선 필요
- 추후 Heroku 배포 재시도
- 모델 일반화와 성능 향상을 위해 타 브랜드 데이터 및 최근 2년간의 가격 변화 데이터수집 필요
- 데이터 수집의 어려움
 - 새로고침 되는 스크롤 웹사이트 크롤링 기술 부족 및 stockx 크롤링 접근 거부
 - 수집한 데이터들의 Feature 유사성 및 일치성 부족 ⇒ 모델 성능 하락
 - 절대적인 데이터 양 부족

iv. 향후 자신의 프로젝트 주제를 이어서 진행할 연구자 혹은 개발자를 위한 방향성 및 기술적 측면 제안

- 지속적인 크롤링 자동화를 통해 데이터베이스 최신화
- 고객 만족 향상을 위한 UI/UX 개선
- 정기적인 리셀 시장 트렌드 분석을 진행하여 변화하는 고객 니즈 충족

4. 참고 자료

- a. <https://www.mobiinside.co.kr/2021/10/26/mz-reselltech/>
[IT 트렌드 바로읽기] MZ세대는 왜 리셀테크에 열광하나
- b. <https://brunch.co.kr/@ddfe283f9d6b4b3/49>
- c. <https://www.koreaes.com/news/articleView.html?idxno=360518>
[한국 연예 스포츠 신문] [mz세대의 재테크] ②리셀테크

StockX 데이터 분석

Aa 이름	📅 날짜	👤 담당 자	▼ 상 태
<u>프로젝트 분야 및 주제 선정</u>	@2022년 6월 25일 → 2022년 6월 28일		
<u>데이터 크롤링</u>	@2022년 6월 27일 → 2022년 6월 30일		
<u>EDA 및 Feature Engineering, 베이스 모 델 제작</u>	@2022년 6월 29일 → 2022년 7월 3일		
<u>모델 성능 향상 및 웹페이지 제작</u>	@2022년 7월 4일 → 2022년 7월 5일		
<u>프로젝트 발표 자료 제작 및 영상 녹화</u>	@2022년 7월 6일		
<u>결과 공유, 문서작성 및 제출</u>	@2022년 7월 7일		
<u>제목 없음</u>	@2021년 11월 2일 → 2021년 11월 6일		
<u>분석 진행 총 기간(~8/12)</u>	@2020년 7월 2일 → 2020년 8월 12일		