심화과정에서 미니 프로젝트

파이썬을 통한 실무과정

주제에 대한 선정

데이터에 대한 분석, 데이터의 이해

학습을 잘하고 있는지

가설을 잘세워서 하는지

발표1 : 비오는 날 사람들의 생활습관분석

김경만, 김혜란, 이현지, 신진영

1. 주제개요

비오는 날씨에 생각확장

1. 데이터 수집 및 전처리  
   변수이름 정리 - 강수량, 대중교통, 교통사고, 따릉이 이용자 수  
   데이터 수집 - 2020-06 ~ 2020-08 서울지역

강수량 데이터 - Nan -> 0으로 처리

6월 1일 - 8월 31일 (92개)

버스데이터의 이상치 확인 - 이상치 정리

네이버 데이터 조인(검색횟수) - 음식, 계절가전, 기타

감정분석???

1. 시각화

부침개, 계절가전 상관관계 높음

따릉이가 비 오는 것에 이용자 수 감소

찌개, 부침개 의미 큼

교통사고 의미 없음

1. 인사이트 도출 및 반성

인사이트 : 비오는 날 따릉이를 수리할 수 있음??

비오는 날 찌개류, 부침개류, 제습기에 대한 수요 대비 재고 보충

이륜차는 항상 안전??

반성점 : 데이터 수가 적음, 일별 데이터 구하기 힘듬, 원하는 데이터 구한점 아쉬움

개별평가

=> Y, X에 값에 대한 의미 파악X

=> 6월 1일 - 8월 31일 뉴스 크롤링해서 단어 파악하면 의미있는 단어찾을 수 있음

평가

1. 단발적인 데이터를 조인해서 별도의 빅데이터 분석이 없어서 인사이트가 없었다.
2. 파이썬을 통해서 프레임을 좋았으나, 어떤 가설을 세우고, 데이터를 고민을 할지..

발표2 : 야! 너도 부자될 수 있어!

박재영 임세정 박다은

1. 주제선정

주제선정배경 : 소비자행동 -> 주가

목표 :

1. 데이터 탐색

2020.01 - 2020.06 신용카드 데이터, 주가데이터, **뉴스키워드 사이트**

데이터 전처리 : 카드기록이 30일이하, 데이터부족 판단, 삭제

데이터 형변환 : 파생변수 code

1. 시각화

업종 선택 -> 기업 선정 -> 주가 비교

업종 선택 : 교통 및 숙박 => 렌터카, 여객선, 철도, 택시는 꾸준히 상승

뉴스키워드 검색 : ex) 택시 키워드 검색 : 카카오모빌리티

주가비교 - 수혜주 ETF vs KOSPI

업종선택 : 가전/가구

뉴스키워드 :

1. 결과 및 결론

보완점 : 데이터&도메인 지식의 부족  
 주가는 매출 총액에 의해 결정x, 이자율, 환율, 원자재 가격, 미국증시

평가

1. 주식은 이자율, 환율, 원자재 가격에 대한 영향이 많은데 고려하지 않은점
2. 시각화 측면은 분석하려고 모습 보임
3. 데이터 다루는

발표3 :

평가

1. 데이터 출처, 미국을 선택한 이유?
2. 뻔한 분석 내용…
3. 머신러닝 쪽으로 가면 괜찮을듯

발표4 :야! 너도 집 구할 수 있어

* 주택과 다른 데이터의 상관관계에 대하여

아파트 변수, 캐글 출처

전국아파트 총거래량, 평균전용면적(5년)

전국 아파트 평균 매매가

아파트 평당가격 시계열 그래프

시계열 데이터 추세 분해 - 일정한 패턴

시계열 데이터 이동변경

기준 금리와 아파트 거래량의 관계 - 금리에 따라서 따라가는 모습

환율과 아파트평당가격 관계 -

이자율& 아파트평당가격 상관분석

-지역별 발레학원의 갯수와 집값의 상관관계

-> 상관관계수 0.46

-업종별 인구수에 관하여(feat. 집값)  
 생산가능 인구 수

업종별 인구 수

집값과 업종의 상관관계 - 강남구

=> 목차 없이 진행된 상황 아쉬움

평가

1. 부동산, 다중영향, 거시미시 ㅡ 상관분석만 할 뿐 데이터 분석이 아니다
2. 팀 프로젝트로서 같이 하면 좋았을 거 같았음

발표5 :The figure of Korean contents in Netflix

이동호, 박진영, 이현진, 김지영

넷플릭스에 있는 한국콘텐츠

데이터셋 : 케글자료

가장 많은 작품에 출연한 배우 - 이준혁, 엄효섭 / 단순 빈도는 조연들

컨텐츠를 고르는 이유는 배우가 아니다

콘텐츠 개수 - 미국, 인도, 한국, 일본, 프랑스

콘텐츠 국가

콘텐츠 영향력 - 평점 합

넷플릭스가 투자를 많이 하는 이유? 한국의 콘텐츠 영향력이 높아서

=> 분석 값 수치… 없어서 아쉬움, 상관계수 없음

=> 넷플릭스를 보는 국가별 인구 수

평가

1. 비정형데이터를 한건지, 감성적 데이터를 했는지? DB데이터가 적음
2. 가설 세우는 것은 좋았으나 데이터 분석 보다는 통계적으로 분석함
3. 비정형데이터를 활용하면 좋을거 같음

발표6 : 프로듀스 청하 프로젝트 청하의 다음 신곡은 언제?

MNH entertainment 이태윤, 김예윤, 이승연, 장우재

2018-19 인기 많은 청하...

청하, 코로나 확진 이후

미국, 영국 극찬 이지만 한국시장 떨어짐

데이터 : 3년간 멜론 차트 분석

1) 크롤링

주차별 순위, 곡 정보

2) 최근 3년간 전체

월요일 출시, 금요일 출시, 빌보드 차트 집계기간 금요일 효율,

12월 연말 시상식 이전인 10월 11월 곡 출시 많음

3) 청하

차트 순위 변동

4) 주요 경쟁 가수

방탄소년단, 블랙핑크, 태연, 세븐틴

사고의 확장 - 장르 선호도와 경제 지표

결론

=> 통계적인 자료, 데이터 분석이 없었다.

=> 갑자기 결론… 목차가 없어서 아쉬움..

=> 코드 보여주는 것은 캡쳐로 했으면 좋았을 것 같다.

평가

1. 분석이 낮고, 전체적으로 데이터… 멜론만 했던 점이 아쉬움
2. 청하가 나오는 TV프로그램, 검색 연관관계