




Investor Profiling

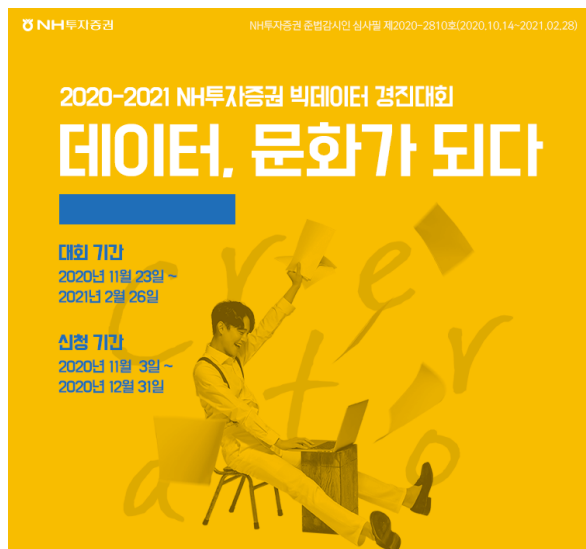
문구영 성유지 이영신





L.Point 공모전이 올해 열리지 않아,
새로운 공모전을 찾게 되었습니다 πππ

NH 투자증권 빅데이터 분석대회



- League 2: Y&Z 세대 투자자 프로파일링
- 1980년대 후반부터 현재까지의 출생연도를 가진 사람들의 투자 행동 분석
- 주어진 데이터로 다양한 인사이트를 끌어내야
- 시각화, 데이터 마이닝

저희 조의 계획

2020 11 NOV

일 SUN	월 MON	화 TUE	수 WED	목 THU	금 FRI	토 SAT
1	2	3	4	5	6	7 85
8	9	10	11	12	13	14
유사 주제의 데이터 분석 대회 수상작 코드 구현						
15 음 10.1	16	17	18	19	20	21
금융 분야에서 활용되는 빅데이터 기법 공부						
22 소설	23	24	25	26	27	28
주어진 데이터 EDA						
29 음 10.15	30					

유사 주제의 데이터 분석 대회 수상작 코드 구현

CUS_INFO.CSV(10,000건)
ACT_INFO.CSV(23,959건)
IEM_INFO.CSV(5,065건)
TRD_KR.CSV(3,312,664건)
TRD_OSS.CSV(29,301건)
data_schema_vf: 데이터 명세

대회 데이터 셋 11/23에 발표!
Customer, action, trade, item 등으로 추정

-> 유사한 데이터셋으로 미리 연습

1. KCB 금융스타일 시각화 경진대회

1. credit_card_data.csv : 전국민 카드 및 대출 이용통계 데이터

연령대/거주지/성별 구분에 따른 카드 및 대출 이용통계 데이터로, 구분 값 별 카드이용금액 비교, 대출금액 및 대출 특이사항 비교 등이 가능하다.

2. jeju_financial_life_data.csv : 지역 거주지(제주도) 금융라이프 데이터

제주도의 우편번호 단위 지역민의 금융생활 통계 데이터
우편번호 단위 별, 거주지 블록 별 소득 비교, 채무 현황, 다양한 컬럼을 활용하여 인사이트를 발굴할 수 있다.

2. 신용카드 거래 데이터 시각화

train.csv: 한국의 약 2,000개 상점의 신용카드 거래 내역 데이터

데이터 탐색

- 결측치 제거 및 유니크 값
- 시간에 따른 탐색
- Amount의 심화 탐색
- installments(할부) 심화 탐색

클러스터링

- 일별 거래의 변화율
- 시간별 거래의 변화율
- Prophet을 통한
클러스터 별 트렌드 분석

변수중요도

- Feature Engineering
- Correlation
- Tree Importance
- Permutation Importance
- Shap value

금융 분야에서 활용되는 빅데이터 기법 공부

Y&Z세대 투자 데이터이므로 투자 관련 지표에 대한 공부와
실제 시장에서 어떠한 기법이 쓰이는지 공부할 필요성이 있다고 느낌!

-> 자문을 통해 기법 선정 후 공부하기로 계획

감사합니다! 화이팅 😊