

엠브레인 패널 분석 경진대회 공모전

1 팀 소개 : Sweet Spot

골프에서 공이 Sweet Spot에 맞게 되면 거리와 방향성이 정확해져 임팩트가 가장 커진다. 주어진 데이터를 활용하여 가장 효율적인 방법으로 유의미한 결과를 도출해내고자 하는 것이 팀 목적과 부합한다고 생각해 팀이름으로 정하였다.

2 Project

2.1 Project 1: 숨은 타겟을 찾아라

2.1.1 문제 설정 및 목적

마크로밀 엠브레인의 경우 많은 회사들을 고객으로 두고 있으며, 회사에서는 마케팅을 통해 수익을 극대화하고 싶어 것이다. 마크로밀 엠브레인에서 기업 제품의 마케팅 이익을 극대화할 수 있도록 숨은 타겟을 찾아 준다면 좋을 것이다. 주어진 데이터를 활용하여 가장 효과적인 마케팅 타겟을 선택, 컨설팅하도록 sweet spot을 찾아주는 알고리즘을 만드는 것이 본 프로젝트의 목적이다.

2.1.2 데이터 전처리, 모델링 계획

먼저, 패널 프로파일 정보를 이용해 사람들의 특성 변수를 선별한다. 패널 프로파일 정보는 설문지 응답을 통해 얻은 데이터로, 가구정보, 의료, 패션, 헬스, 교육 등 다양한 변수를 이용할 수 있다. 위 변수들을 이용하여 비슷한 성향을 가진 사람들을 클러스터링 한 뒤, 가구소비재나 내구재를 구매한 경험 정보를 종속변수로 이용하여 실제로 구매하지는 않았지만 비슷한 성향을 가진 고객들에게 추천할 수 있다.

참고논문

- DeepWalk: Online Learning of Social Representations(KDD, 2014)
- LINE: Large-scale Information Network Embedding(WWW, 2015)
- Node2vec: Scalable Feature Learning for Networks(KDD, 2016)

2.1.3 기대 효과

마크로밀 엠브레인은 향후 기업의 마케팅 분석 프로젝트에서 더욱 효율적으로 숨은 고객층을 찾을 수 있을 것으로 판단된다. 또한, BPTO(Brand Price Trade-Off, 경쟁사와 자사 제품의 가격변화에 따른 시장 점유율의 변화를 예측할 수 있는 가격분석 모형)에도 도움이 될 것이다.

2.2 Project 2: 미세먼지와 생활패턴

2.2.1 문제 설정 및 목적

최근 미세먼지로 인해 야외 활동에 많은 제한이 있으며, 사람들의 소비 패턴도 많은 영향을 주고 있다. 첫째, 외식문화에 대한 영향이다. 사람들은 미세먼지가 심한 날에 배달음식을 선호할 것이고, 이를 고객들의 결제정보를 통하여 확인할 수 있다. 둘째, 고객들의 쇼핑 행태에서 오프라인 매장을 통한 쇼핑보다는 온라인 쇼핑몰의 결제 횟수가 증가할 것으로 예측된다.

미세먼지로 인한 소비패턴의 차이를 알아내고, 기업들의 영업 패턴에 적용하여 이익 창출에 기여하는 것이 본 프로젝트의 목적이다.

2.2.2 데이터 전처리, 모델링 계획

미세먼지와 관련된 기상청 데이터를 지역/일자별로 수집하고, 고객들의 결제액 데이터를 이용해 소비패턴을 파악할 것이다. 다음 비슷한 행동양식을 갖는 고객들을 클러스터링 하고, 이를 바탕으로 각 클러스터 별로 미세먼지 수준에 따른 고객들의 소비패턴에 차이가 있는지 분산분석을 통해 검증해볼 것이다.

2.2.3 기대 효과

미세먼지로 인하여 고객 소비패턴의 유의미한 차이가 발생하는 것을 확인할 수 있다면, 현재 미세먼지의 효과가 단순한 국민들의 건강 문제뿐만 아닌 국가 경제에도 영향을 줄 수 있다는 사실을 확인할 수 있다. 이를 바탕으로 국가 정책 수립에 있어 긍정적 영향을 줄 수 있을 것이다. 또한 미세 먼지 수준에 따른 고객들의 소비 패턴을 오프라인 매장의 영업 패턴에 적용하여 수익성을 극대화시킬 수 있을 것으로 예상된다.

2.3 Project 3: 음식 배달 수요 예측

2.3.1 문제 설정 및 목적

소상공인이나 소규모 프랜차이즈 등 외식 산업의 발달로 기존 배달서비스를 제공하지 않았던 음식점 입장에서는 판매 영역의 확대와 함께 매출에 따라 음식점 소속의 배달기사를 고용하는 것보다 대행업

체를 이용하는 것이 필요 경비를 줄이고 배달 직원의 사고 책임에서도 자유로워 인기를 얻고 있다. 배달을 위해 배달원을 적정하게 고용하여 사업을 운영하는 것이 수익 향상의 중요한 요소이다. 주어진 데이터로 지역/위치별 적정 배달 인원을 제안하는 것이 본 프로젝트의 목적이다.

2.3.2 데이터 전처리, 모델링 계획

'[PAYMENT] Summary Category.csv' 데이터의 업종 코드 열에서 코드 501번인 배달이 77719건으로, 2번째로 많은 데이터임을 확인할 수 있었다. 또한 Payment 데이터에서 'Area Code', 'Area Name', 'LATITUDE', 'LONGITUDE' 변수로 결제된 장소에 대한 정보를 알 수 있다. 이를 기반으로 다른 변수들과 외부 데이터의 다양한 변수들을 활용하여, 지역/위치 별 예상 주문 요청 건 수를 구하는 모델을 만들 계획이다.

2.3.3 기대 효과

예상 배달 요청 건 수보다 배달이 많은 경우, 대기 배달원의 수가 부족하여 업무에 차질이 발생할 수 있다. 반면 예상 배달 요청 건 수 보다 실제 배달 건 수가 현저히 적은 경우, 과도한 인건비로 인하여 수익률이 떨어질 수 있다. 따라서 적정 배달 건 수를 예상할 수 있는 모델을 만들 수 있다면 수익률을 향상시킬 수 있을 것이다.

3 결론

상기 프로젝트들 중, EDA를 통하여 데이터에 가장 적합한 주제를 선택하여 진행할 계획입니다.