



2020 IC-PBL 사례연구

온라인 쇼핑 고객 데이터 분석

인문사회연구소지원사업 2차년 성과물



제 출 문

본 보고서를
「IC-PBL 사례연구: 온라인 쇼핑 고객 데이터 분석」
최종 결과물로 제출합니다.

2020년 11월

연구책임자: 신윤식 (에이모)

※ 본 보고서는 2019년 대한민국 교육부와 한국연구재단의 지원을
받아 수행된 연구임(NRF-2019S1A5C2A04083153)

〈목차〉

I. 개요	4
1.1. 배경	4
1.2. 거시환경	5
II. 데이터 분석	8
2.1. 분석대상 및 요구사항	8
2.2. 분석환경 및 환경구성	9
2.3. 분석 프레임워크	13
2.4. 문제정의	14
2.5. 데이터 전처리	14
2.6. EDA(Exploratory Data Analysis, 탐색적 데이터 분석)	17
2.7. 분석모델 적용	24
2.8. 결론도출	32
III. 정리	35
참고자료	36

〈표 차례〉

〈표 I-1〉 2020년 7월 e-commerce 상위 10개 웹사이트 순방문자 및 도달률 ...	6
〈표 II-1〉 데이터셋 명세	8
〈표 II-2〉 Colab, 드라이브 연동 코드	11
〈그림 II-8〉 K-means Clustering 알고리즘 실행과정	29
〈표 II-2〉 프로모션 적합 고객의 정량적 조건	34

〈그림 차례〉

〈그림 I-1〉 북미 온라인 소매유통 거래액 추이(2000~2018)	5
〈그림 I-2〉 국내 온라인 전용 쇼핑몰 거래액 추이(2017.1Q ~ 2020.2Q)	6
〈그림 I-3〉 세계 리테일 분석 시장가치(2016-2022E)	7
〈그림 II-1〉 구글 로그인 후 메인페이지	9
〈그림 II-2〉 구글 로그인 후 메인페이지	9
〈그림 II-3〉 구글 드라이브에 파일 복사	10
〈그림 II-4〉 Colab 파일 생성	10
〈그림 II-5〉 구글 Colab 접속화면	10
〈그림 II-6〉 Colab, 드라이브 연동을 위한 코드 입력	11
〈그림 II-7〉 Colab, 드라이브 연동 계정 로그인 및 인증코드 복사	11
〈그림 II-7〉 Colab, 드라이브 연동 인증코드 입력 및 완료 확인	12
〈그림 II-5〉 데이터 분석 Framework	13
〈그림 II-6〉 분석대상, 목표, 방법 구체화 과정	14
〈그림 II-7〉 Colab에서 드라이브 데이터 파일 찾기	15
〈그림 II-9〉 RFM 결과 시각화	33
〈그림 II-10〉 충성고객 군집 기술통계량	33
〈그림 II-11〉 RFM 기술통계량	34
〈그림 III-1〉 전체 분석 진행과정	35