



 CAKD5 광민아엘사를부탁해조

1차 프로젝트를 위한 데이터 분석 보고서

김기현 문성윤 박광민 최종원

CONTENTS

1. 유통업계 현황 분석 - 외부요인
2. 과제 도출
3. 데이터 분석
4. 데이터 재분류
5. 앞으로의 계획

01. 유통업계 현황 분석 - 외부요인

2015년 유통 업계 키워드는 'H.A.R.D.'..롯데마트 유통업계 이슈·소비 트렌드 분석

- Hesitate to Buy: 경기(景氣)침체 장기화로 대형 세일 잇따라
- Anytime, Anywhere: 유통업체들의 배송 경쟁 (온·오프라인)
- Renew Everything: 오프라인 매장 강점 어필 위한 변신 시도
- Desire to Safe: 메르스, 육가공육 이슈로 고객 확보 경쟁 치열

2015년 6~8월 메르스 영향으로 고객 급감

2015년 10~12월 정부 코리아 블랙프라이데이
민간주도 쇼핑대전 'K-세일데이' 행사

년	월	총매출액(천단위)	월매출평균
2014	1	28249482	25533
2014	2	22817825	23783
2014	3	26627456	23116
2014	4	26412674	23460
2014	5	28456014	23815
2014	6	25677279	21928
2014	7	24578377	19997
2014	8	24884479	20498
2014	9	25486858	22743
2014	10	30603048	25491
2014	11	33925358	29518
2014	12	31882990	25537
2015	1	27754412	22057
2015	2	27685970	25099
2015	3	29014094	23066
2015	4	29084253	22991
2015	5	30894338	23518
2015	6	24856083	19477
2015	7	26064429	19593
2015	8	25333692	19597
2015	9	28138430	23997
2015	10	31990296	28106
2015	11	34021501	29880
2015	12	32579818	27728

02. 과제 도출

고객 유지를 위해 기존구매고객 대상 구매 감소고객 예측 모델

고객 유지가 회사 성장에 중요한 이유



경제성

기존 고객을 유지하는 것보다 새로운 고객을 확보하는 비용이 5 ~ 25 배 더 비쌉니다.

ROI

고객 보유율을 5 % 증가 시키면 회사 수익이 25-95 % 증가할 수 있습니다

충성도

유지 된 고객은 더 자주 구매하고 새로운 고객보다 더 많이 지출합니다. 그들은 제품이나 서비스의 가치를 배웠고 계속해서 다시 돌아 왔습니다.

소개 & 추천

만족스럽고 충성도가 높은 고객은 회사를 칭찬하고 친구와 가족에게 추천 할 가능성이 높습니다. 새로운 고객을 무료로 데려옵니다

03. 데이터 분석

구매상품(PURPROD) 테이블 분석

1. 구매일자에서 '월'을 빼서 따로 속성 추가

2. 월별로 총매출액, 월매출평균 확인

-> 2014년, 2015년 둘 다 6~8월 매출이 낮고 10월 ~ 12월 매출이 높은 것으로 확인할 수 있었음

```
-- 월 추가
ALTER TABLE PURPROD ADD MONTH NUMBER;
UPDATE PURPROD SET MONTH=substr(구매일자,5,2);
-- 월별로 총매출액, 월매출평균 살펴보기
select year 년, month 월, round(sum(구매금액)/1000) "총매출액(천단위)", round(avg(구매금액)) "월매출평균"
from purprod group by year, month order by year, month;
```

	년	월	총매출액(천단위)	월매출평균
1	2014	1	28249482	25533
2	2014	2	22817825	23783
3	2014	3	26627456	23116
4	2014	4	26412674	23460
5	2014	5	28456014	23815
6	2014	6	25677279	21928
7	2014	7	24578377	19997
8	2014	8	24884479	20498
9	2014	9	25486858	22743
10	2014	10	30603048	25491
11	2014	11	33925358	29518
12	2014	12	31882990	25537
13	2015	1	27754412	22057
14	2015	2	27685970	25099
15	2015	3	29014094	23066
16	2015	4	29084253	22991
17	2015	5	30894338	23518
18	2015	6	24856083	19477
19	2015	7	26064429	19593
20	2015	8	25333692	19597
21	2015	9	28138430	23997
22	2015	10	31990296	28106
23	2015	11	34021501	29880
24	2015	12	32579818	27728

03. 데이터 분석

구매상품(PURPROD) 테이블 분석

성별 별로 총매출액, 월매출평균 확인

-> 구매금액과 구매횟수는 여성이 압도적으로 높으나
평균금액으로 봤을 땐 남성이 더 높은 것을 알 수 있었음

-- 여성과 남성의 구매금액, 평균금액, 구매 횟수를 비교

```
SELECT YEAR, MONTH, 성별, ROUND(SUM(구매금액)/1000) 구매금액, ROUND(AVG(구매금액)) 평균금액,  
count(구매금액) 구매횟수  
FROM PURPROD P, CUSTDEMO C  
WHERE P.고객번호=C.고객번호  
GROUP BY YEAR, MONTH, 성별  
ORDER BY YEAR, MONTH, 성별;
```

	YEAR	MONTH	성별	구매금액	평균금액	구매횟수
1	2014	1	F	22540425	24772	909898
2	2014	1	M	5709057	29058	196473
3	2014	2	F	18103313	23032	785996
4	2014	2	M	4714512	27187	173408
5	2014	3	F	21330543	22625	942787
6	2014	3	M	5296912	25331	209107
7	2014	4	F	20885505	22665	921472
8	2014	4	M	5527169	27041	204400
9	2014	5	F	23023676	23617	974870
10	2014	5	M	5432338	24691	220009
11	2014	6	F	20575191	21502	956877
12	2014	6	M	5102089	23832	214085
13	2014	7	F	19724393	19601	1006291
14	2014	7	M	4853984	21783	222831
15	2014	8	F	19782659	19910	993604
16	2014	8	M	5101820	23151	220370
17	2014	9	F	20342000	22206	916047
18	2014	9	M	5144859	25145	204611
19	2014	10	F	24733580	25184	982118
20	2014	10	M	5869468	26870	218439
21	2014	11	F	26969797	28738	938462
22	2014	11	M	6955561	32990	210838
23	2014	12	F	25385775	24831	1022358
24	2014	12	M	6497216	28728	226163
25	2015	1	F	22180071	21529	1030247
26	2015	1	M	5574341	24443	228052
27	2015	2	F	22218469	24623	902345
28	2015	2	M	5467501	27241	200708
29	2015	3	F	23469108	22906	1024582
30	2015	3	M	5544987	23769	233289

03. 데이터 분석

구매상품(PURPROD) 테이블 분석

지역별 나이에 따른 고객분포와 소비금액

```
-- 지역별 나이에 따른 고객분포와 소비금액
select 거주지역, 성별, 연령대, count(c.고객번호) 총고객수, sum(구매금액) 총구매액
from custdemo c, purprod p where c.고객번호=p.고객번호
group by 거주지역, 성별, 연령대 order by 거주지역;
```

	거주지역	성별	연령대	총고객수	총구매액
1	010	F	20세~24세	668	4933616
2	010	F	25세~29세	4291	116270057
3	010	F	30세~34세	15053	607458974
4	010	F	35세~39세	20453	452177767
5	010	F	40세~44세	43918	824473840
6	010	F	45세~49세	46595	774781046
7	010	F	50세~54세	24492	863995727
8	010	F	55세~59세	21700	739233189
9	010	F	60세이상	17087	624007229
10	010	M	25세~29세	4159	34562052
11	010	M	30세~34세	3491	115246855
12	010	M	35세~39세	4147	167469864
13	010	M	40세~44세	6142	345988735
14	010	M	45세~49세	11132	150787211
15	010	M	55세~59세	5533	28920097
16	010	M	60세이상	3258	80811612
17	013	F	20세~24세	1824	26385751
18	013	F	25세~29세	5154	156271461
19	013	F	30세~34세	7572	392224290
20	013	F	35세~39세	30772	1006282995
21	013	F	40세~44세	47191	1240286258
22	013	F	45세~49세	62626	1861065777
23	013	F	50세~54세	42669	1379281204
24	013	F	55세~59세	19476	705102433
25	013	F	60세이상	16637	725036911
26	013	M	19세이하	175	2009580
27	013	M	25세~29세	4735	39600459
28	013	M	30세~34세	2154	7062038

03. 데이터 분석

멤버십여부(MEMBERSHIP) 테이블 분석

1. 전체 고객 수 19383명,
총 멤버십 발급 수 7456건, 멤버십 가입자 수 6498명
-> 전체 고객 중 약 34%의 고객이 멤버십에 가입함

2. 멤버십을 2건 이상 중복 발급 받은 고객은 총 897명

3. 고객별로 어떤 멤버십을 발급 받았는지 확인

```
-- 전체 고객 수 19383명
select count(distinct 고객번호) 전체고객수 from purprod;
-- 총 멤버십 발급수 7456건
select count(고객번호) 총멤버십발급수 from membership;
-- 멤버십 가입자수 6498명
select count(distinct 고객번호) 멤버십가입자수 from membership;
-- 중복되는 가입자수
SELECT COUNT(COUNT(고객번호)) "멤버십 중복 가입자수" FROM MEMBERSHIP GROUP BY 고객번호 HAVING COUNT(고객번호) > 1;
-- 멤버십 몇 개 발급 받았는지
-- 2건의 멤버십 발급받은 837명, 3건 59명, 4건 1명 -> 897명
SELECT COUNT(COUNT(고객번호)) FROM MEMBERSHIP GROUP BY 고객번호 HAVING COUNT(고객번호) = 3;
-- 고객별로 어떤 멤버십을 쓰는지 확인 가능
SELECT 고객번호, 멤버십명 FROM MEMBERSHIP WHERE 고객번호 = ANY(SELECT 고객번호
FROM MEMBERSHIP GROUP BY 고객번호 HAVING COUNT(고객번호) > 1);
```

전체고객수	
1	19383
총멤버십발급수	
1	7456
멤버십가입자수	
1	6498
멤버십 중복 가입자수	
1	897

	고객번호	멤버십명
1	03225	다둥이
2	03225	하이마트
3	03326	다둥이
4	03326	하이마트
5	03354	롭스
6	03354	하이마트
7	03503	다둥이
8	03503	하이마트
9	03638	다둥이
10	03638	하이마트
11	03644	다둥이
12	03644	하이마트
13	03687	다둥이
14	03687	하이마트
15	03710	롭스
16	03710	하이마트
17	03719	다둥이
18	03719	하이마트
19	03737	다둥이
20	03737	하이마트
21	03799	다둥이
22	03799	하이마트
23	04096	다둥이

03. 데이터 분석

경쟁사이용(COMPET) 테이블 분석

제휴사별 고객의 경쟁사 방문횟수

-> B 제휴사는 '마트'로 경쟁사 방문 횟수가 가장 많음

```
--제휴사별 고객의 경쟁사 방문횟수
```

```
select 제휴사, count(경쟁사) 경쟁사방문횟수 from compet group by 제휴사;
```

	제휴사	경쟁사방문횟수
1	B	13526
2	C	5176
3	A	6385
4	D	3072

채널이용(CHANNEL) 테이블 분석

제휴사별 채널 누적이용횟수(3개월기준)

-> B 제휴사는 '마트'로 제휴사 중 온라인몰, 모바일앱 누적이용 횟수가 가장 많음

```
-- 제휴사별 채널 누적이용횟수
```

```
select 제휴사, count(*) 누적이용횟수 from channel group by 제휴사;
```

제휴사	누적이용횟수
C MOBILE/APP	46
B ONLINEMALL	1328
D MOBILE/APP	227
B MOBILE/APP	5170
A MOBILE/APP	1435
C ONLINEMALL	618

대분류

01 가공식품

02 신선식품

03 일상용품

05 의약품/의료기기

06 교육/문화용품

07 디지털/가전

08 가구/인테리어

09 의류

10 전문스포츠/레저

11 패션잡화

99 기타상품

04. 데이터 재분류

가공식품, 신선식품, 일상용품, 의약품/의료기기, 교육/문화용품, 디지털/가전,
가구/인테리어, 의류, 전문스포츠/레저, 패션잡화, 기타상품 + 유아동품, 명품, 외식


유통상품지식뱅크 PRODUCT BANK KOREA

“

[유통상품지식뱅크]는 국내 시장에서 유통되는 바코드 부착 상품에 대한 가장 정확하고 상세한 데이터를 제공합니다. 3,000여종의 상품카테고리 기준으로, 상품당 50~700여개의 데이터 항목을 표준화된 체계로 수집/관리하고 있는 국내 최고의 고품질 상품데이터를 활용해 보시기 바랍니다.

”

가공식품	신선식품	일상용품	의약품/의료기기	교육/문화용품	디지털/가전
					
356,808	13,179	490,031	48,086	62,358	64,346

가구/인테리어	의류	전문스포츠/레저	패션잡화	기타상품
				
18,507	75,316	5,902	3,202	526

1	가공식품
2	신선식품
3	일상용품
4	의약품/의료기기
5	교육/문화용품
6	디지털/가전
7	가구/인테리어
8	의류
9	전문스포츠/레저
10	패션잡화
11	명품
12	유아동
13	외식
14	기타

04. 데이터 재분류

소비자 구매행동 특성에 따른 편의품, 선매품, 전문품 분류

구분	편의품	선매품	전문품
구매 전 계획정도	거의 없다.	있다.	상당히 있다.
가격	저가	중, 고가	고가
브랜드 충성도	거의 없다.	있다.	특정상표 선호
고객 쇼핑 노력	최소한	보통	최대한
제품회전율	빠름	느림	느림
카테고리	가공식품 신선식품 일상용품 의약품/의료기기 교육/문화용품 외식 기타	디지털/가전 가구/인테리어 의류 전문스포츠/레저 패션잡화 유아동품	명품

05. 앞으로의 계획

EDA (Exploratory Data Analysis) 탐색적 데이터 분석



머신러닝 모델 결정

