

HADOOP/HIVE

현대백화점 데이터 분석

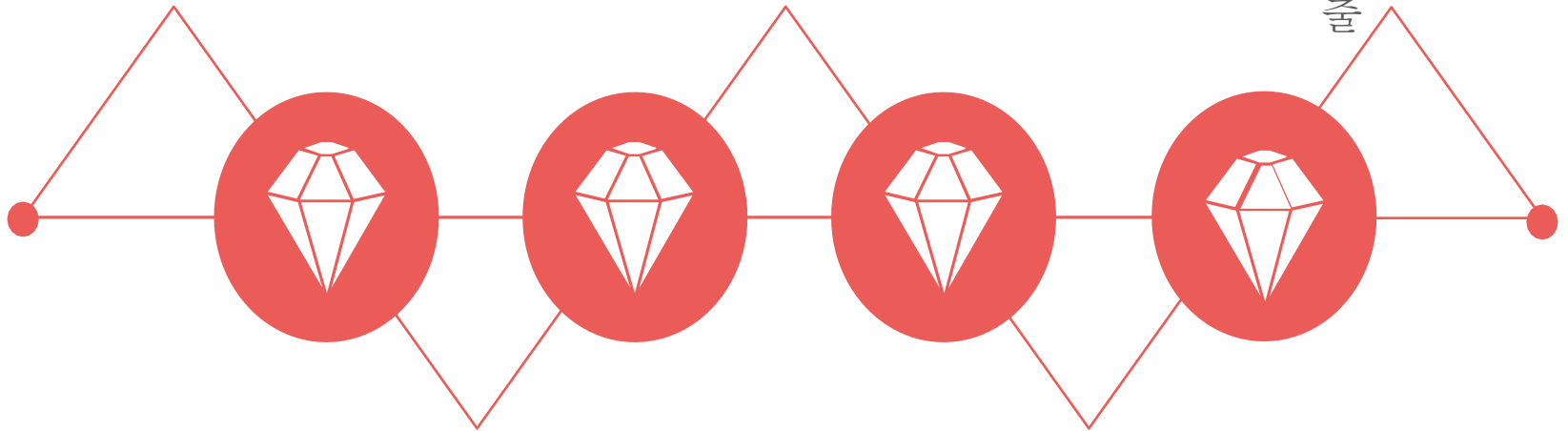
문제 정의

저장

결과 도출

수집 및 요약

처리



문제 정의



현대 백화점 명품과 일반 상품에 대한 구매 비교



현대 백화점 지점별 구매 시간 패턴 분석



현대 백화점 구매 금액이 높은 고객의 직업 유형

수집 및 요약



판매 파일



고객 파일



직업 파일



카드 파일



주소 파일

수집 및 요약



판매 파일

필드명	설명	도메인
sale_date	판매날짜	yyyymmdd (Customer Table 참조)
sale_time	판매 시간	hhmm
custid	고객아이디	1~50000
str_nm	지점명	신촌점, 본점, 무역점, 천호점...
goodcd	상품코드	
brd_nm	브랜드이름	식품, 슈퍼프리미아, 삼운, 엘지, 크리니크...
corner_nm	코너이름	주류,차류,화장잡화,욕실용품
pc_nm	상품군이름	일반식품,장신구,소품,식기...
part_nm	파트이름	공산품파트, 잡화파트, 가정용품....
team_nm	담당팀이름	식품팀, 잡화가용팀, 의류패션팀....
buyer_nm	상품분류	일반식품, 장신구, 조리욕실, 피혁B...
import_flg	수입여부	0 or 1
tot_amt	판매액	-101,000,000~14,790,000
dis_ant	할인액	-1479000~2301440
net_amt	실구매액	-7867900~9728000

수집 및 요약



고객 파일

필드명	설명	도메인
custid	고객 아이디	1~50000
sex	구매 시간	0: 유효값
birth	생일	yyyymmdd
card_flg1	카드유형	0000~9999 (Card Table 참조)
mrg_flg	결혼기념일	yyyymmdd
h_type1	주거형태코드	A:아파트, B:빌딩, H:병원, N:단독주택, V:빌라, X: 기타, 7:기타, 8:기타
h_type2	주거현황코드	0:기타, 1:본인소유, 2: 배우자소유, 3: 부모소유, 4: 전세, 5:기타, 6: 기타, 7: 기타, 8: 기타
hobby	취미코드	0001~ 0017 : 등산,스포츠,여행,낚시,관람,컴퓨터, 요리,서예,미술,공예,음악,꽃꽂이,수집,독서,작문,바둑,기타
job_stype	직업코드	00000~99999 (Job Table 참조)
ent_date	카드가입일	yyyymmdd
home_zip1	주택우편번호1	000~999 (Address Table 참조)
home_zip2	주택 우편번호2	000~999 (Address Table 참조)
work_zip1	직장우편번호1	000~999 (Address Table 참조)
work_zip2	직장우편번호2	000~999 (Address Table 참조)
cus_stype	고객등급	1-12 : 초우량1,2,3,4등급 / 우량1,2,3,4등급 / 고정 1,2등급 / 일반 1,2등급

수집 및 요약



주소 파일

필드명	설명	도메인
post_num	우편번호	000000~999999
serial_num	일련번호	000~999
si_do	시도	강원도, 경상북도, 충청도...
si_gun_gu	시군구	남구, 포항시, 칠곡군, 청송군.....
eup_meon_d o	읍면동	진도면, 청송읍, 파천면, 북산읍...
li	리	송도, 호도, 조도, 노도, 양도.....

수집 및 요약



직업 파일

필드명	설명	도메인
job_stype	직업코드	0000~9999
job_nm	직업군	현대백화점, 판촉사원, 무역,은행, 제조업
job_nm_gr	직업군 그룹	그룹사, 직원, 개인사업,공공기관,개인사업...



카드 파일

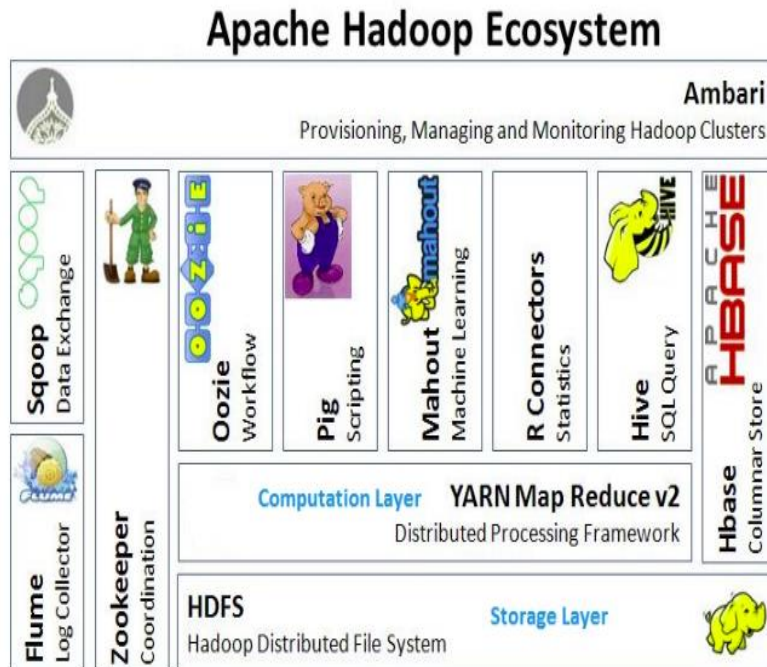
필드명	설명	도메인
card_flg1	카드코드	0000~9999
card_sflg_nm	카드유형	회장단 및 당사 중역,임대업주,퇴직자,Gold



저장

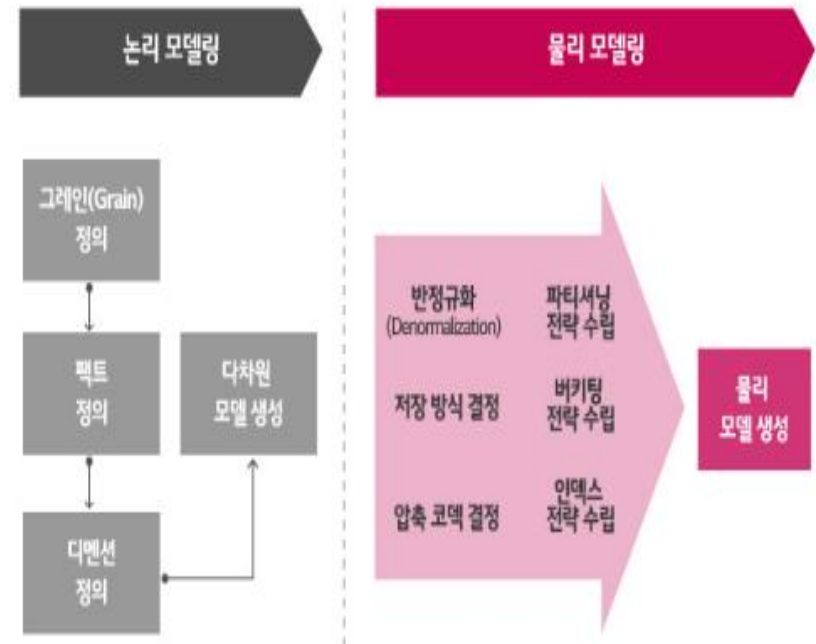
01. 저장 환경 구성

빅데이터 분석을 위한 환경 구성



02. 데이터 이관 및 적재

데이터웨어하우스 모델링 및 적재





저장 데이터 이관 및 적재

자료형 결정



스키마 생성



적재

```
sale_date    int
sale_time    int
custid       int
str_nm       string
goodcd       string
brd_nm       string
corner_nm    string
pc_nm        string
part_nm      string
team_nm      string
buyer_nm     string
import_flg   string
tot_amt      int
dis_amt      int
net_amt      int
inst_mon     tinyint
```

```
create table if not exists
hds_sales_info
(
  sale_date int,
  sale_time int,
  custid int,
  str_nm string,
  goodcd string,
  brd_nm string,
  corner_nm string,
  buyer_nm string,
  import_flg string,
  tot_amt int,
  dis_amt int,
  net_amt int,
  inst_mon tinyint
) row format delimited
fields terminated by ','
stored as textfile;
```

```
load data
local inpath
"hds_sale_info.csv"
overwrite into table
hds_sales_info;
```

처리

명품과 일반 상품에 대한 구매 비교
SQL 작성 및 결과 테이블 출력



판매 테이블

```
create table if not exists output
(
  custid int,
  kk int,
  corner_nm string,
  tot_amt int
)
row format delimited
fields terminated by ','
stored as textfile;
```

```
create table if not exists output2
(
  custid int,
  kk int,
  tot_amt int
)
row format delimited
fields terminated by ','
stored as textfile;
```

##명품 구매고객

```
insert into table output
select custid, count(custid) as
kk, sum(tot_amt), corner_nm from
hds_sales_info_1 where
corner_nm='Luxury'
group by custid, corner_nm order by
kk desc limit 10;
```

##구매고객

```
insert into table output2
select custid, count(custid) as
kk, sum(tot_amt) from
hds_sales_info_1 where
corner_nm<>'Luxury'
group by custid order by kk desc limit
10;
```

처리 지점별 구매 시간 패턴 분석 SQL 작성 및 결과 테이블 출력



판매 테이블

```
create table str_sales (  
  sale_data int,  
  sale_time int,  
  str_nm string,  
  tot_amt int  
)  
row format delimited  
fields terminated by ','  
stored as textfile;
```

```
create table output (  
  sale_data int,  
  sale_time int,  
  str_nm string,  
  tot_amt int  
)  
row format delimited  
fields terminated by ','  
stored as textfile;
```

```
insert into table str_sales  
select sale_date,sale_time-  
substr(sale_time,2),str_nm, sum(tot_amt)  
tot_amt  
from hds_sales_info_1  
group by sale_date,sale_time, str_nm;
```

```
insert into table hds_sales_info from select  
* from hds_sales_info_1;
```

```
create table str_sales (  
  sale_time int,  
  str_nm string,  
  tot_amt int
```

```
insert into table str_sales  
select sale_date,sale_time-  
substr(sale_time,2),str_nm, sum(tot_amt)  
tot_amt  
from hds_sales_info_1  
group by sale_date,sale_time, str_nm;
```

처리

구매 금액이 높은 고객의 직업
SQL 작성 및 결과 테이블 출력



판매 테이블 고객 테이블



직업 테이블

```
create table sales_customers(job_stype  
string,custid int,)
```

```
insert into table sales_customers  
select c.custid,job_stype,tot_amt
```

```
from hds_sales_info_1 s
```

```
join hds_customers c
```

```
on s.custid=c.custid;
```

```
create table output(
```

```
job_stype string,
```

```
job_nm string,
```

```
job_nm_gr string,
```

```
custid int,
```

```
tot_amt int
```

```
)row format delimited
```

```
fields terminated by ','
```

```
stored as textfile;
```

```
insert into table output
```

```
select j.job_stype, job_nm, job_nm_gr,
```

```
custid, tot_amt
```

```
from sales_customers_tot t
```

```
join hds_jobs j on t.job_stype=j.job_stype;
```

분석

명품과 일반 상품 구매 비교

구매 총액과 횟수

결과 파일

custid	구매횟수	구매총액	luxury		custid	구매횟수	구매총액	
25950	49	13791500	WN	구매횟수 평균	33995	671	38270284	구매횟수 평균
42800	48	17985600	WN	2.4795287	32067	589	35251614	34.362989
49427	36	14052800	WN		43915	511	62516419	
47992	33	15216000	WN	구매총액 평균	40230	501	13365905	구매총액 평균
42131	31	5693500	WN	₩871,954	26295	491	92868830	₩2,992,329
38565	29	13254000	WN		47509	479	14495775	
42214	29	2769000	WN		24321	474	17229001	
42013	28	9085400	WN		32778	451	38941417	
39912	28	3409300	WN		37639	444	77126175	
33453	27	8880000	WN		48163	436	19809518	
48902	24	7134500	WN		16613	436	9678437	
13573	23	6410800	WN		42339	435	20325401	
47291	23	14928300	WN		34717	432	62455805	
4670	23	3818000	WN		26728	417	28412163	
34952	23	458000	WN		49448	417	4682925	
40926	21	8880500	WN		36574	417	29952687	
35339	21	4925600	WN		31665	416	40056099	
37198	20	8927400	WN		49831	413	27001548	
14002	19	3454000	WN		34172	408	12916140	

구매 총액 평균 = 명품 < 일반 상품

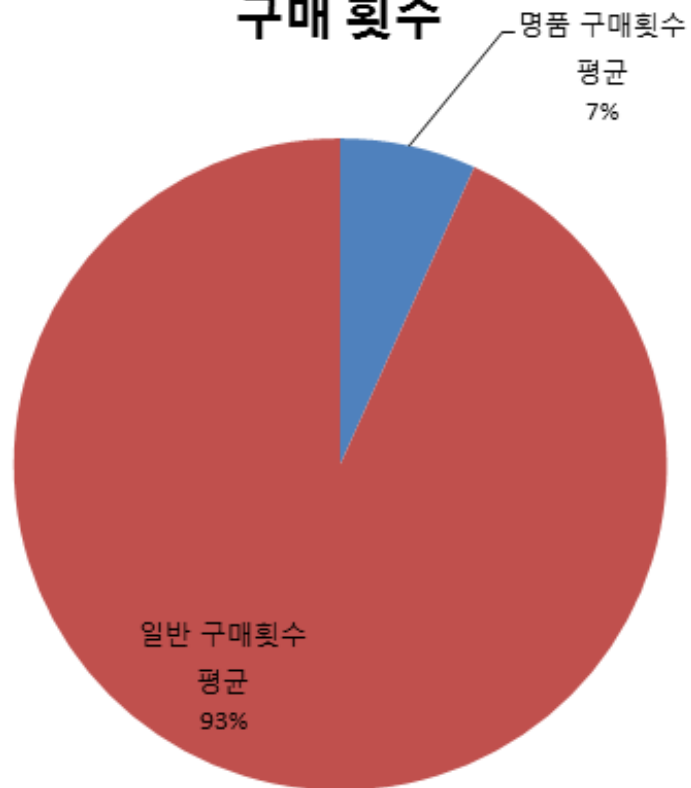
구매 횟수 평균 = 명품 < 일반 상품

분석

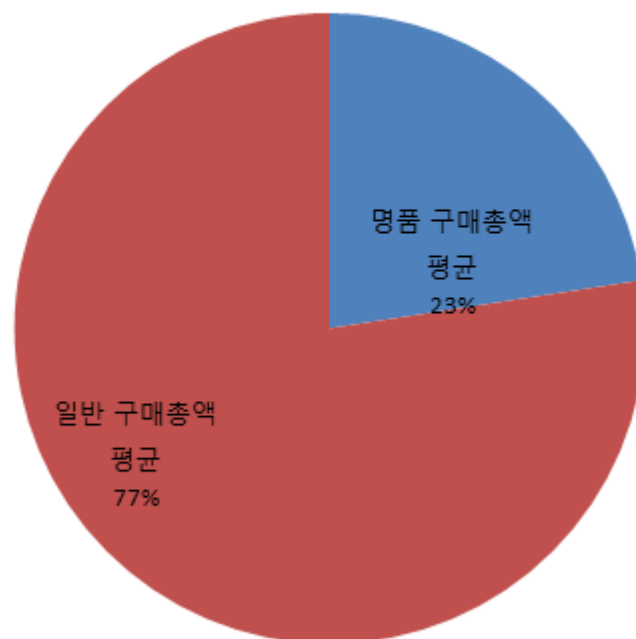
명품과 일반 상품 구매 비교

그래프

구매 횟수



구매 총액



요약

명품과 일반 상품 구매 비교

명품 매출액 평균 = 전체 매출 평균의 1/4

적절한 고객 마케팅이 필요하다.

명품을 구매한 사람의 특징을 추출하여 비슷한 고객에게 타겟 홍보

일반 고객중 구매액이 높은 고객에게 타겟 홍보



분석

지점별 주요 구매 시간 패턴 분석 방문고객수와 판매 금액

결과 파일

최대

최소



신촌

시간 방문고객수 판매 금액

9	272	27699.67
10	6479	420039.65
11	27074	2040133.39
12	34931	2638350.94
13	39417	3166433.42
14	43847	3709587.60
15	49942	4167514.77
16	53039	4426207.69
17	52356	4228513.43
18	53822	4221129.08
19	48901	3929939.49
20	3766	366537.96



압구정(본점)

시간 방문문고객수 판매 금액

8	2	-37.000
9	188	9583.288
10	8506	777022.732
11	37497	4148976.529
12	45118	4882382.357
13	51574	5897955.863
14	59197	7460232.204
15	66289	8411546.621
16	65912	8255402.728
17	63098	7524689.537
18	59797	6800143.336
19	36463	3779378.692
20	2825	315314.606
21	41	-6091.100
22	3	-301.500



천호

시간 방문고객수 판매 금액

9	68	5788.637
10	5621	429798.587
11	23180	1869159.532
12	30395	2344228.348
13	34026	2824031.742
14	38791	3462689.143
15	44263	3907955.842
16	46398	4139088.134
17	45576	3891074.789
18	44098	3696847.657
19	23434	1900210.719
20	1988	181447.985
21	9	-4169.760



삼성

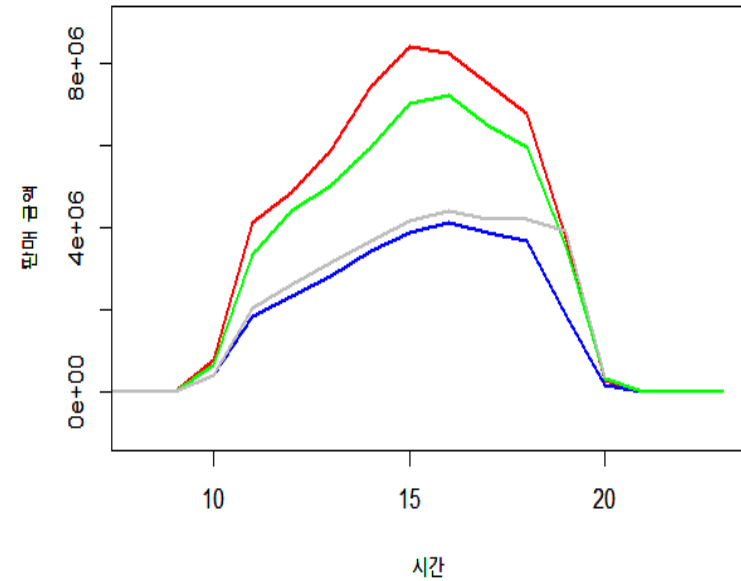
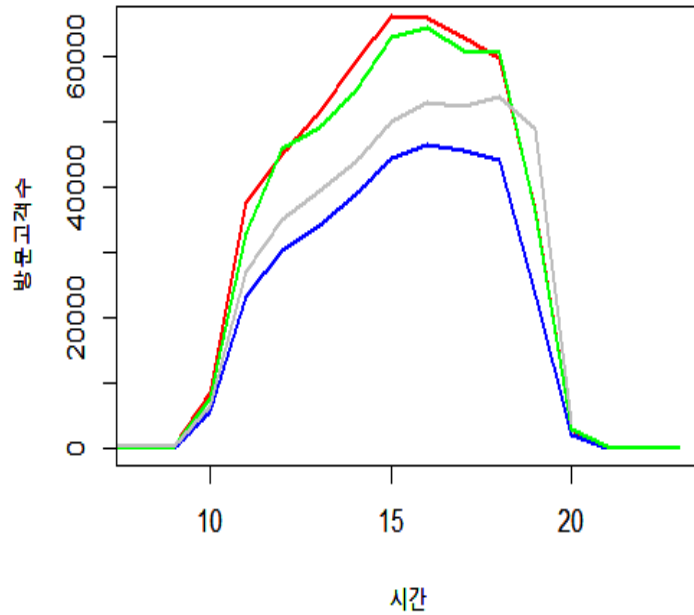
시간 방문고객수 판매 금액

8	1	-296.00
9	105	16171.74
10	7455	669917.78
11	32932	3357857.52
12	45898	4433690.04
13	49056	5019461.24
14	54615	5979471.75
15	62871	7026071.58
16	64303	7225645.95
17	60917	6510254.64
18	60708	5981809.34
19	36179	3550274.78
20	2989	323115.71
21	178	10281.74
22	4	256.30
23	2	65.00

분석

지점별 주요 구매 시간 패턴 분석 방문고객수와 판매 금액 그래프

빨강 : 압구정
연두색 : 삼성
회색 : 신촌
파란색 : 천호

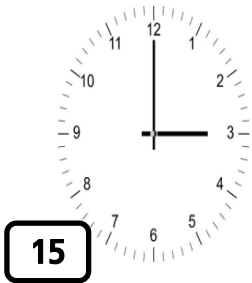




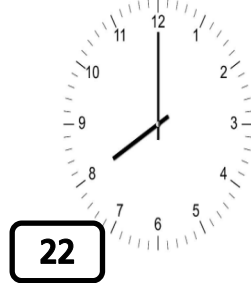
분석 지점별 주요 구매 시간 패턴 분석 최대 최소 방문 시간

압구정(본점)

방문 최대 66289

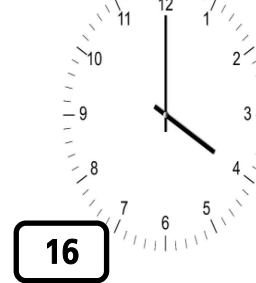


방문 최소 3

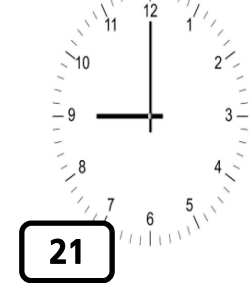


천호

방문 최대 46398

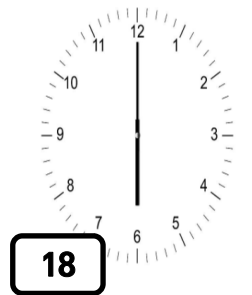


방문 최소 9

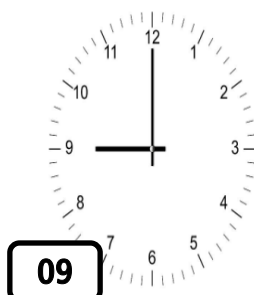


신촌

방문 최대 53822

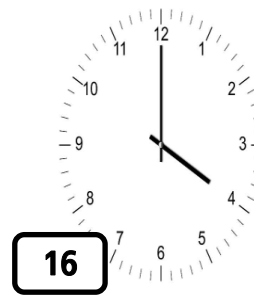


방문 최소 272

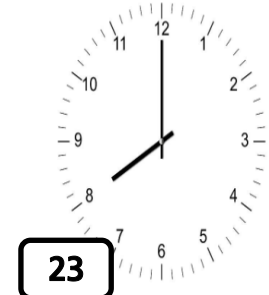


삼성

방문 최대 64303



방문 최소 2

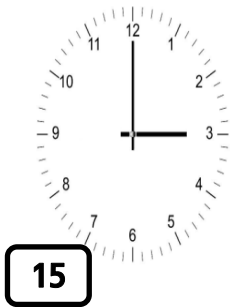




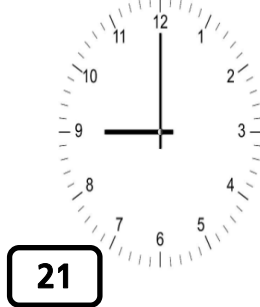
분석 지점별 주요 구매 시간 패턴 분석 최대 최소 판매 시간

압구정(본점)

판매 최대

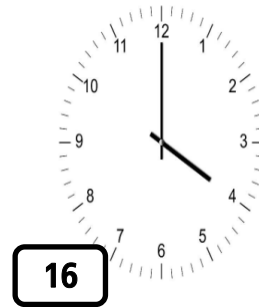


판매 최소

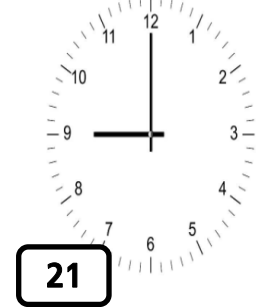


천호

판매 최대

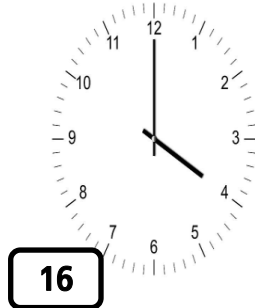


판매 최소

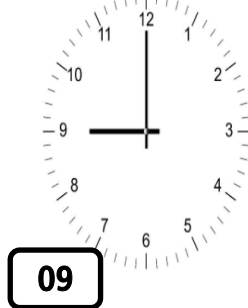


신촌

판매 최대



판매 최소

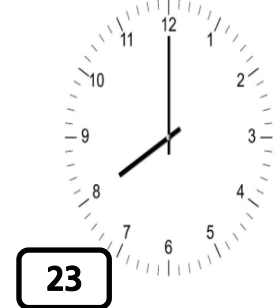


삼성

판매 최대



판매 최소



요약

언제 운영하는 것이 효율적일까?



10시 30분~

19시

20시

21시

22시

23시



신촌

48901

3766



압구정(본점)

36463

2825

41

3



천호

23434

1988

9



삼성

36179

2989

178


4

2

요약

방문객이 많은 지점의 순서?  압구정(본점) > 삼성점 > 신촌점 > 천호점

판매 금액이 많은 지점의 순서?  압구정(본점) > 삼성점 > 신촌점 > 천호점

방문객이 많은 시간대는?  오후 3 ~ 6

판매 금액이 높은 시간대는?  오후 3 ~ 4

방문객이 낮은 시간대는?  오전 9 , 오후 9~11

판매 금액이 낮은 시간대는?  오전 9 , 오후 9~11

효율적인 시간에 운영하는 지점은?  신촌점



분석 및 요약

구매 금액이 높은 고객의 직업
유형

결과 파일

	A	B	C	D	E
1	job_stype	job_nm	job_nm_gr	custid	tot_amt
2	0	monitor	employee	19443	66000
7	3	monitor	employee	16612	8465809
8	3	monitor	employee	28982	5516312
9	3	monitor	employee	31982	475662
10	3	monitor	employee	32607	2723163
11	3	monitor	employee	34585	506000
12	3	monitor	employee	35383	2739600
13	3	monitor	employee	39089	6823034
14	3	monitor	employee	39238	3296000
15	3	monitor	employee	39743	1794249
16	3	monitor	employee	39837	8586284
17	3	monitor	employee	12691	256104
18	3	monitor	employee	22882	929031
19	3	monitor	employee	16021	5800200
20	20	Hyundai C	employee	25563	8879500
21	20	Hyundai C	employee	31637	9498318
22	20	Hyundai C	employee	3260	1220882

결론

가장 구매 금액이
높은 사람의 직업군?



증권회사의 투자자

가장 구매 금액이
낮은 사람의 직업군?



공공기업 건설업자