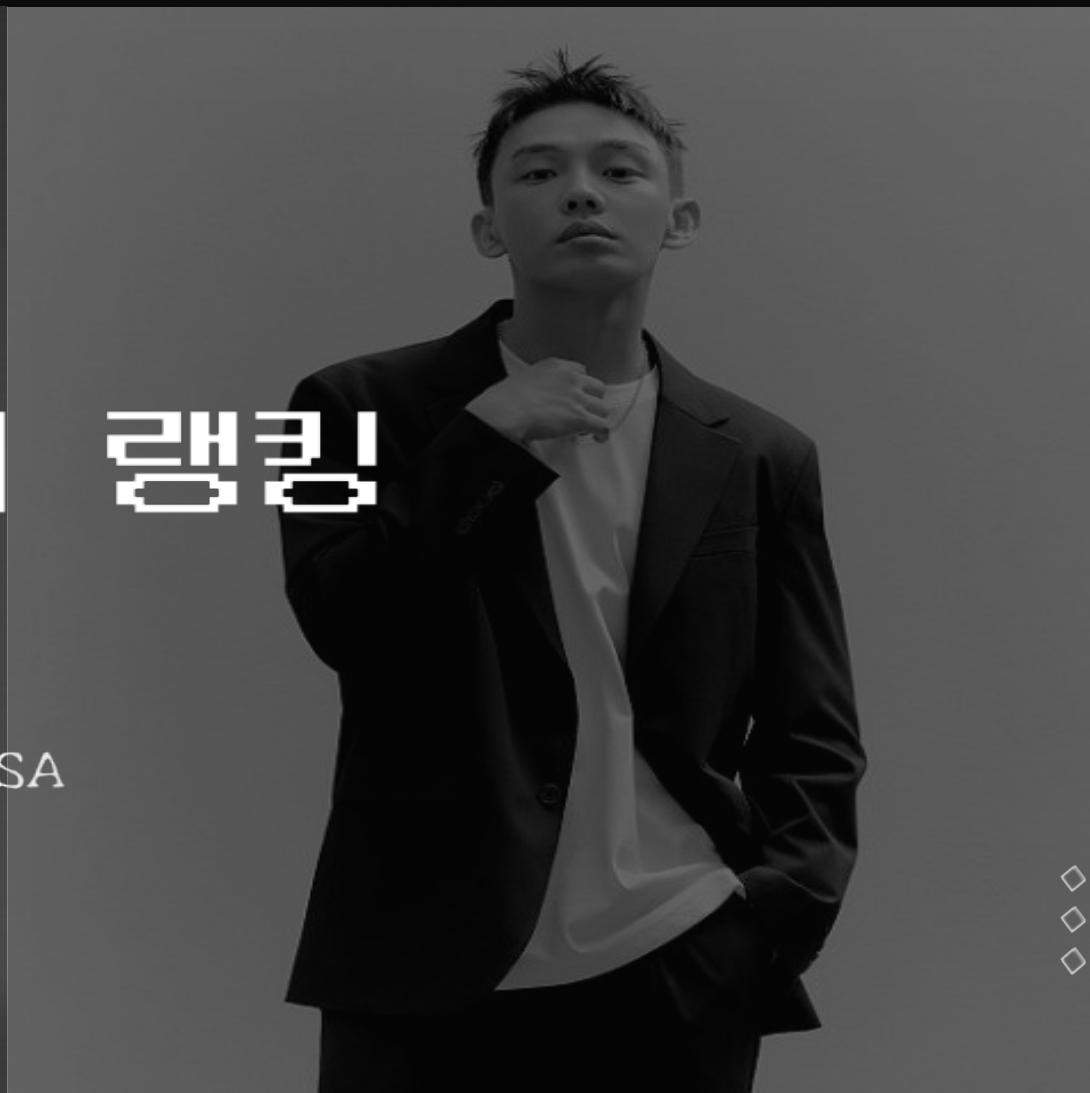


PORTFOLIO

Project No.1

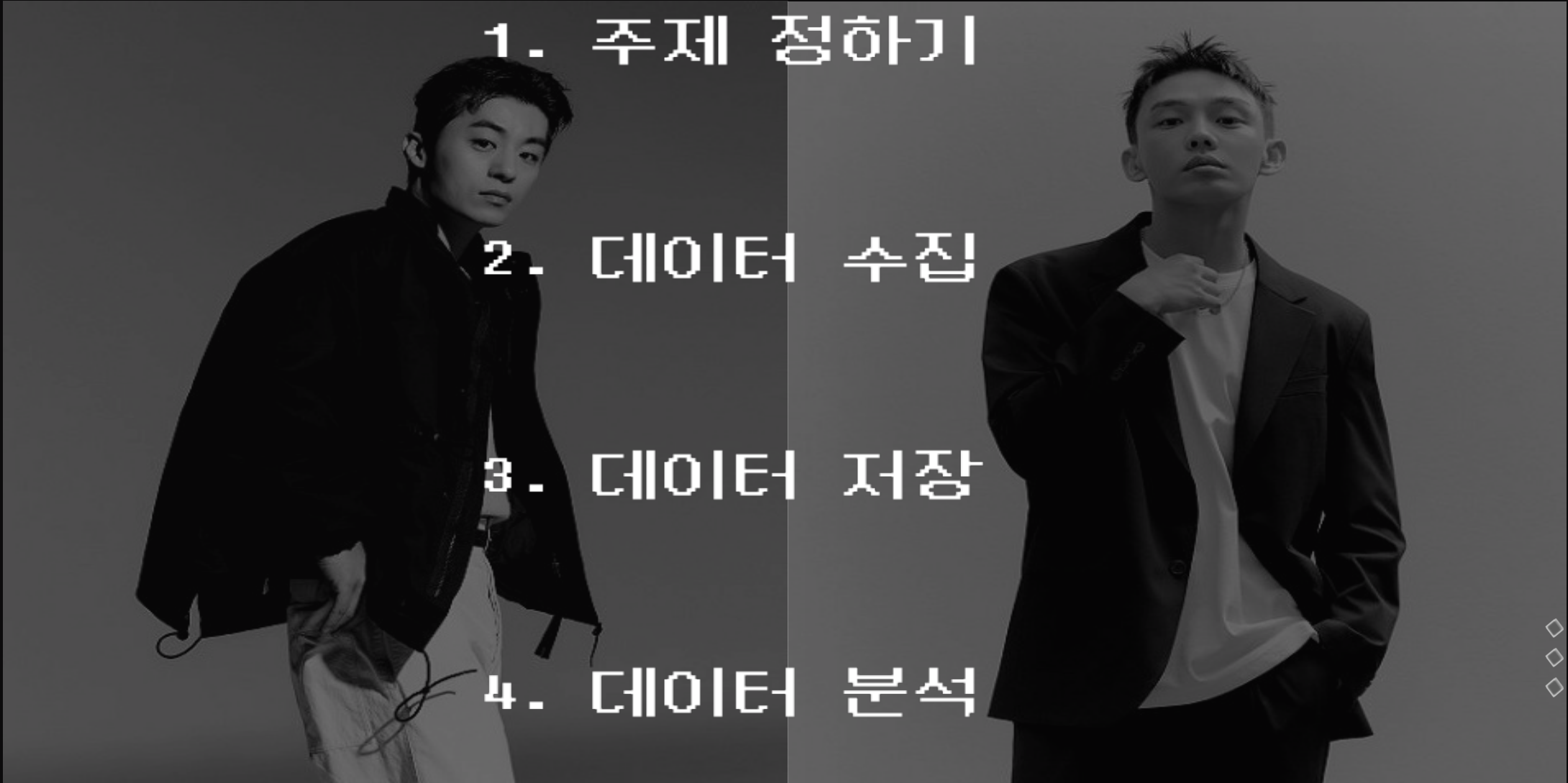
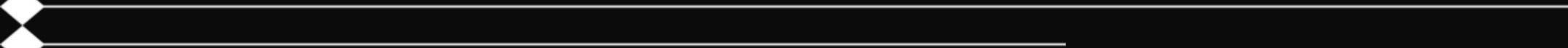


# 아우터 랭킹

MUSINSA

김건영, 안은선, 이재영F, 전한성





1. 주제 정하기

2. 데이터 수집

3. 데이터 저장

4. 데이터 분석

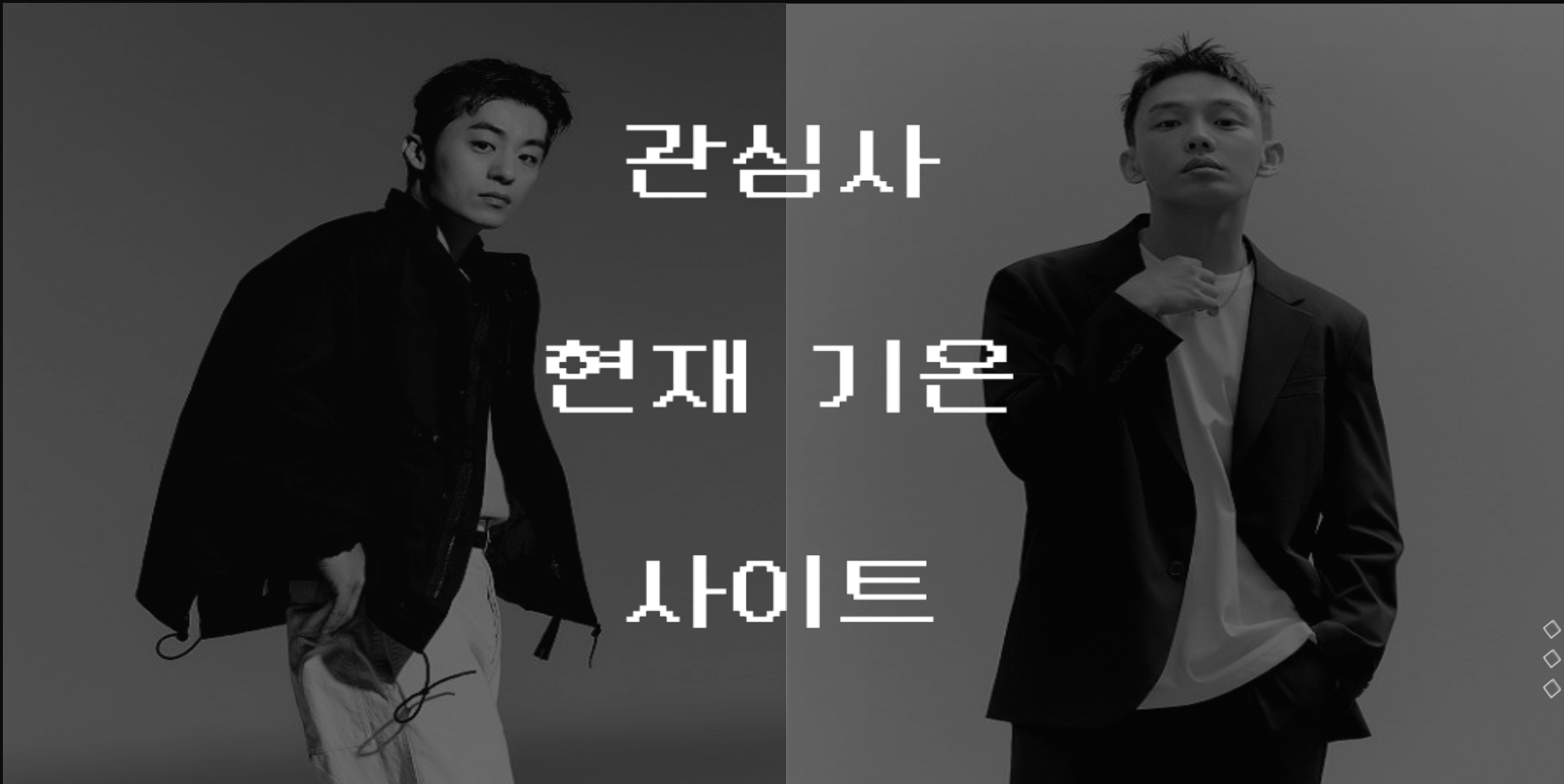




# 1

## 주제 정하기





## Part 1 >> 질문 만들기



결정

겨울을 앞둔 시기인 만큼, 사람들의 관심이 많을 아우터 카테고리의 랭킹을 분석하기로

- 할인을 많이 하는 브랜드
- 할인 쿠폰을 증정하는 브랜드
- 가격대별 제품 순위
- 리뷰수 top10 제품
- 누적 판매량별 순위



# 2 데이터 수집



## Part 2 >> 웹 스크래핑

### 무신사 스토어의 아우터 카테고리 월간 랭킹의 1페이지를 스크래핑 (1~90위)

→ 수집할 데이터 : 제품의 순위, 브랜드명, 제품명, 리뷰수, 가격대, 쿠폰 유무, 누적 판매량

```
import requests
from bs4 import BeautifulSoup

url = "https://www.musinsa.com/ranking/best?period=month&age=ALL&mainCategory=002&subCategory=&leafCategory=&price=&golf=false&kids=false&newProduct=false"

user_agent = "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/107.0.0.0 Safari/537.36"

headers = {"User-Agent" : user_agent}
response = requests.get(url, headers)
response.raise_for_status()

soup = BeautifulSoup(response.text, 'lxml')
```

## Part 2 >> 웹 스크래핑

### 무신사 스토어의 아우터 카테고리 월간 랭킹의 1페이지를 스크래핑 (1~90위)

→ 수집할 데이터 : 제품의 순위, 브랜드명, 제품명, 리뷰수, 가격대, 쿠폰 유무, 누적 판매량

```
# 누적판매량 추가
ii = 0
driver = webdriver.Chrome()
while ii < 90 :

    driver.get("https://www.musinsa.com/ranking/best?period=month&age=ALL&mainCategory=002&subCategory=002")
    items = driver.find_elements_by_css_selector("img.lazyload.lazy")
    items[ii].click()
    time.sleep(1)
    sum_sale = driver.find_element_by_id("sales_1y_qty")
    sst = sum_sale.text

    if '만' in sst :
        if '.' in sst :
            sst = sst.replace('.', '')
            sst += '000'
        else :
            sst += '0000'

        sst = sst.replace('만 개 이상', '')

    elif '천' in sst :
        if '.' in sst :
            sst = sst.replace('.', '')
            sst += '00'
        else :
            sst += '000'
        sst = sst.replace('천 개 이상', '')

    else :
        sst = sst.replace('개 이상', '')
```

```
# 브랜드
brands = soup.select('form#goodsRankForm div.list-box.box div.article_info p.item_title a')

brd = ""+brands[ii].text+""

# 제품명
names = soup.select('form#goodsRankForm div.list-box.box div.article_info p.list_info a')

nm = ""+(names[ii]['title'])+""

# 다 넣은 곳에서 쿠폰 있는거 없는거 찾기
all_items = soup.select('div.right_area.page-contents-best form#goodsRankForm div.list-box.box .li_box')

a = all_items[ii].find('p', {'class' : 'mu-icon mu-icon-coupon'})
if a == None :
    coup = "x"
else :
    coup = "o"

# 리뷰 수
points = soup.select('div.right_area.page-contents-best form#goodsRankForm div.list-box.box div.article_info p.point span.count')

pp = points[ii].text.replace(',','')

# 가격대
prices = soup.select('div.right_area.page-contents-best form#goodsRankForm div.list-box.box div.article_info p.price')

price11 = prices[ii].text.split()

if len(price11) == 2:
    pri1 = price11[0][0:-1].replace(',','')
    pri2 = price11[1][0:-1].replace(',','')
else:
    pri1 = price11[0][0:-1].replace(',','')
    pri2 = 0

print('INSERT INTO musinsa VALUES ('+ii+1+', '+brd+', '+nm+', '+pri1+', '+pri2+', '+pp+', '+coup+', '+sst+',');')

ii += 1
```





3

데이터 저장



## Part 3 >> 테이블 설계 • 저장

# 4. musinsa 구문 만들기

SQL = """

```
create table musinsa (  
    ranks int not null,  
    brand varchar(20) not null,  
    title varchar(150) not null,  
    ori_price varchar(20),  
    dis_price varchar(20),  
    review_count int,  
    coupon varchar(20),  
    sale_amt int(10),  
    primary key(ranks)
```

);

"""

Pymysql 모듈을 이용해서 데이터 저장

ranks : 순위 (pk)

brand : 브랜드명

title : 제품명

ori\_price : 원가

dis\_price : 할인가

review\_count : 리뷰 수

coupon : 쿠폰

sale\_amt : 누적 판매량

# 4 데이터 분석



## Part 4 >> 데이터 분석

# 랭킹 중 할인을 가장 많이 하는 브랜드  
top10

라퍼지스토어	10
커버넛	7
스파오	4
쿠어	4
파르티멘토	4
코드그래피	4
예일	4
무신사 스탠다드	4
드로우핏	3
도프제이슨	3

→ 가장 할인을 많이 하는 브랜드는 라퍼지스토어로,  
90위 랭킹 중 10개의 제품이나 세일 중이다.

## Part 4 >> 데이터 분석

### # 랭킹 중 할인 쿠폰을 증정하는 브랜드

쿠어

→ 90위 랭킹 중,

쿠폰을 증정하는 제품의 브랜드는 총 7개가 나왔다.

라퍼지스토어

예일

노이어

무신사 스탠다드

케이투

파르티멘토

## Part 4 >> 데이터 분석

# 가격대별 아웃터 순위 > 10만원 이하

→ 10만원 이하의 제품은 총 19개가 나왔다.

# 10만원 이상, 20만원 이하

→ 10만원 이상, 20만원 이하의 제품은 총 43개가 나왔다.

# 20만원 이상

→ 20만원 이상의 제품은 총 28개가 나왔다.

결과가 양이 너무 많아서 생략! 자세한 코드는 [git.hub](#)의 `mysql` 파일 참조

## Part 4 >> 데이터 분석

# 가격대별 아우터 순위 > 30만원 이상 가격순 top10

24	노이어	캐시미어 더블 브레스티드 코트 (블랙)	425000	297000	55	o	1000
52	메종미네드	BLACK DOUBLE LONG GOOSE DOWN	389000	289000	1845	x	2700
59	케이투	씬에어(THINAIR)ECO LITE_Black	350000	252000	1260	o	1500
40	노스페이스	NJ1DN50J 화이트라벨 노벨티 눅시 다운 자켓	348000	0	529	x	2000
3	노스페이스	NJ1DN75A 남성 1996 에코 눅시 자켓	339000	0	277	x	2200
84	노스페이스	NJ1DN75B 남성 1996 에코 눅시 자켓	339000	0	85	x	500
61	도프제이슨	오버핏 램스킨 싱글 자켓	339000	198000	7321	x	8800
45	쿠어	MTR 오버핏 더블 롱 코트 (블랙)	334000	300600	6580	o	2900
5	쿠어	MTR 발마칸 코트 (블랙)	329000	296100	2732	o	4700
26	쿠어	MTR 발마칸 코트 (멜란지토프)	329000	296100	2235	o	3100

→ 90위 랭킹 중, 원가가 가장 비싼 제품은  
노이어 브랜드의 42만 5000원짜리 코트로 나왔다!

## Part 4 >> 데이터 분석

### # 리뷰 수 많은 아우터와 브랜드 순위(top 10)

					review		
6	토피	2WAY 스웨트 후드 집업 (MELANGE GREY)	45000	37000	52900	x	149000
71	무신사 스탠다드	베이식 블레이저 [블랙]	79900	67890	46214	x	21000
37	무신사 스탠다드	릴렉스드 베이식 블레이저 [블랙]	81900	73690	27521	x	44000
82	무신사 스탠다드	오버사이즈 블레이저 [블랙]	83900	71290	24392	x	20000
15	무신사 스탠다드	캐시미어 블렌드 오버사이즈 싱글 코트 [블랙]	159900	143890	19098	x	12000
7	라퍼지스토어	[리뉴얼 ver.][3M 신솔레이트]오리지널 M-1965 피쉬테일 파카_Original Khaki	138000	89900	17402	o	37000
49	라퍼지스토어	[매트블랙] 버핑레더 미니멀 포켓 블루종_Matt Black	94000	65900	12626	o	30000
20	도프제이슨	오버핏 비건레더 싱글 자켓	129000	82550	12423	x	23000
67	파르티멘토	[비건 레더] 필드 자켓 블랙	111700	72600	12414	x	19000
2	스파오	베이직 푸퍼_SPJPC4TC11	69900	62910	11741	x	41000

→ 아직 한겨울이 오지 않아서 패딩보다는 후드집업, 블레이저, 자켓등의 제품 리뷰가 많은 것을 알 수 있다.

데이터를 수집한 사이트가 무신사 스토어인 만큼, 브랜드도 무신사 스탠다드의 비중이 높은 것을 확인할 수 있다.



## Part 4 >> 데이터 분석

### # 누적 판매량 top10

6	토피	2WAY 스웨트 후드 집업 (MELANGE GREY)	149000
37	무신사 스탠다드	릴렉스드 베이식 블레이저 [블랙]	44000
2	스파오	베이직 푸퍼_SPJPC4TC11	41000
7	라퍼지스토어	[리뉴얼 ver.][3M 신슬레이트]오리지널 M-1965 피쉬테일 파카_Original Khaki	37000
9	스파오	베이직 퍼플리스 집업_SPFZC4TU01	34000
49	라퍼지스토어	[매트블랙] 버핑레더 미니멀 포켓 블루종_Matt Black	30000
75	스파오	리버서블 푸퍼_SPJDC4VC14	24000
20	도프제이슨	오버핏 비건레더 싱글 자켓	23000
10	라퍼지스토어	버핑레더 오버핏 블루종_Black	22000
11	스파오	파스텔 푸퍼_SPJPC4TG01	21000

→ 가장 많이 팔린 제품은 토피 브랜드의 스웨트 후드 집업으로, 누적 판매량이 14.9만개 이상이다!

## Part 4 >> 데이터 분석

### # 5만원에서 20만원 사이 제품의 누적 판매량 top10

37	무신사 스탠다드	릴렉스드 베이식 블레이저 [블랙]	44000
2	스파오	베이직 푸퍼_SPJPC4TC11	41000
7	라퍼지스토어	[리뉴얼 ver.][3M 신슬레이트]오리지널 M-1965 피쉬테일 파카_Original Khaki	37000
49	라퍼지스토어	[매트블랙] 버핑레더 미니멀 포켓 블루종_Matt Black	30000
75	스파오	리버서블 푸퍼_SPJDC4VC14	24000
20	도프제이슨	오버핏 비건레더 싱글 자켓	23000
10	라퍼지스토어	버핑레더 오버핏 블루종_Black	22000
11	스파오	파스텔 푸퍼_SPJPC4TG01	21000
71	무신사 스탠다드	베이식 블레이저 [블랙]	21000
82	무신사 스탠다드	오버사이즈 블레이저 [블랙]	20000

→ 앞서 리뷰수 순위로 보았던 것과 비슷한 결과가 나왔다. 리뷰수가 누적 판매량에 비례하다는 것을 알 수 있다.

## Part 4 >> 데이터 분석

### # 총정리

- 최근 한달간은 후드집업, 블레이저, 자켓 등의 수요가 더 많다는 것을 알 수 있다.  
더 많은 소비를 위해 할인과 쿠폰을 증정하는 브랜드도 많았고,  
월간 랭킹 안의 제품들은 주로 10만원~20만원대 사이에 많이 분포되어있음을 알 수 있었다.





PORTFOLIO

#Project No.1

Thank you for watching

