

오픈 마켓 데이터 분석을 통한 제품 추천

무선청소기 모델별 비교 분석

■ 데이터 살펴보기

```
# 데이터 불러오기
import pandas as pd
data = pd.read_excel('data/danawa_data.xlsx')
data.head()
```

	카테고리	회사명	제품	가격	사용시간	흡입력
0	핸디/스틱청소기	샤오미	드리미 V10	173900	60.0	220.0
1	핸디/스틱청소기	원더스리빙	다이나킹 Z9	299000	65.0	220.0
2	핸디/스틱청소기	LG전자	코드제로 A9 A978	1005340	80.0	140.0
3	핸디/스틱청소기	델로라	V11 파워 300W	141000	70.0	220.0
4	핸디/스틱청소기	샤오미	드리미 V9	138800	60.0	200.0

```
# columns 데이터 확인
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 241 entries, 0 to 240
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype  
---  -
0   카테고리    241 non-null   object  
1   회사명      241 non-null   object  
2   제품        241 non-null   object  
3   가격        241 non-null   int64   
4   사용시간    218 non-null   float64  
5   흡입력      129 non-null   float64  
dtypes: float64(2), int64(1), object(3)
memory usage: 11.4+ KB
```

무선청소기 모델별 비교 분석

■ 데이터 살펴보기

```
# 흡입력 기준 정렬
top_list = data.sort_values(["흡입력"],
                             ascending = False)
top_list.head()
```

	카테고리	회사명	제품	가격	사용시간	흡입력
13	핸디/스틱청소기	DIBEA	F20 맥스	222990	50.0	250.0
127	핸디/스틱청소기	DIBEA	X30	259000	50.0	250.0
165	핸디/스틱청소기	DIBEA	TSX-25000A	244470	45.0	250.0
143	핸디/스틱청소기	DIBEA	F20 울트라 맥스	236550	60.0	250.0
152	핸디/스틱청소기	아이룸	RS1	178000	40.0	250.0

```
# 사용시간 기준 정렬
top_list = data.sort_values(["사용시간"],
                             ascending = False)
top_list.head()
```

	카테고리	회사명	제품	가격	사용시간	흡입력
111	핸디/스틱청소기	삼성전자	제트 VS20R9074S2	845990	120.0	200.0
5	핸디/스틱청소기	삼성전자	제트 VS20R9078S2	877880	120.0	200.0
153	핸디/스틱청소기	샤오미	이지에 YE-01	24740	120.0	NaN
16	핸디/스틱청소기	삼성전자	제트 VS20R9078S3	918120	120.0	200.0
76	핸디/스틱청소기	삼성전자	제트 VS20R9074S3	870910	120.0	200.0

무선청소기 모델별 비교 분석

■ 데이터 살펴보기

```
# 흡입력, 사용시간을 기준으로 정렬
```

```
top_list = data.sort_values(["사용시간", "흡입력"], ascending = False)
```

```
top_list.head()
```

	카테고리	회사명	제품	가격	사용시간	흡입력
5	핸디/스틱청소기	삼성전자	제트 VS20R9078S2	877880	120.0	200.0
16	핸디/스틱청소기	삼성전자	제트 VS20R9078S3	918120	120.0	200.0
76	핸디/스틱청소기	삼성전자	제트 VS20R9074S3	870910	120.0	200.0
109	핸디/스틱청소기	삼성전자	제트 VS20R9077Q3	931100	120.0	200.0
111	핸디/스틱청소기	삼성전자	제트 VS20R9074S2	845990	120.0	200.0

무선청소기 모델별 비교 분석

■ 데이터 살펴보기

```
# 평균값 정리
price_mean = data['가격'].mean()
suction_mean = data['흡입력'].mean()
use_time_mean = data['사용시간'].mean()
print("가격 평균값", price_mean)
print("흡입력 평균값", suction_mean)
print("사용시간 평균값", use_time_mean)
```

가격 평균값 296844.79253112036
흡입력 평균값 151.8294573643411
사용시간 평균값 43.38990825688074

```
# 가성비 좋은 제품 탐색
condition_data = data [
    (data['가격'] <= price_mean) &
    (data['흡입력'] >= suction_mean) &
    (data['사용시간'] >= use_time_mean)]
condition_data
```

	카테고리	회사명	제품	가격	사용시간	흡입력
0	핸디/스틱청소기	샤오미	드리미 V10	173900	60.0	220.0
3	핸디/스틱청소기	델로라	V11 파워 300W	141000	70.0	220.0
4	핸디/스틱청소기	샤오미	드리미 V9	138800	60.0	200.0
13	핸디/스틱청소기	DIBEA	F20 맥스	222990	50.0	250.0
18	핸디/스틱청소기	DIBEA	M500 쿼텀	248640	50.0	250.0
42	핸디/스틱청소기	DIBEA	F20 프로	161970	50.0	220.0
73	핸디/스틱청소기	JDL	tech 타이폰 DV-889DC-X	137160	50.0	200.0

무선청소기 모델별 비교 분석

■ 데이터 시각화

```
# 라이브러리 임포트 및 한글 글꼴 설정
from matplotlib import font_manager, rc
import platform
font_path = ''
if platform.system() == 'Windows':
    font_path = 'c:/Windows/Fonts/malgun.ttf'
    font_name = font_manager.FontProperties(fname = font_path).get_name()
    rc('font', family = font_name)
elif platform.system() == 'Darwin':
    font_path = '/Users/$USER/Library/Fonts/AppleGothic.ttf'
    rc('font', family = 'AppleGothic')
else:
    print('Check your OS system')

%matplotlib inline
```

무선청소기 모델별 비교 분석

■ 데이터 시각화

```
# 시각화를 위해 null data 정리 하기위해 데이터 확인
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 241 entries, 0 to 240
Data columns (total 6 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   카테고리    241 non-null    object
 1   회사명      241 non-null    object
 2   제품        241 non-null    object
 3   가격        241 non-null    int64
 4   사용시간    218 non-null    float64
 5   흡입력      129 non-null    float64
dtypes: float64(2), int64(1), object(3)
memory usage: 11.4+ KB
```

```
# null data 확인
data.isnull()
```

	카테고리	회사명	제품	가격	사용시간	흡입력
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
...
236	False	False	False	False	False	False
237	False	False	False	False	True	True

```
data['흡입력'].isnull()
```

```
0      False
1      False
2      False
3      False
4      False
...
236     False
237       True
238       True
239     False
240     False
Name: 흡입력, Length: 241, dtype: bool
```

무선청소기 모델별 비교 분석

■ 데이터 시각화

```
# 결측값 없애기
chart_data = data.dropna(axis = 0)
chart_data.head()
```

	카테고리	회사명	제품	가격	사용시간	흡입력
0	핸디/스틱청소기	샤오미	드리미 V10	173900	60.0	220.0
1	핸디/스틱청소기	원더스리빙	다이나킹 Z9	299000	65.0	220.0
2	핸디/스틱청소기	LG전자	코드제로 A9 A978	1005340	80.0	140.0
3	핸디/스틱청소기	델로라	V11 파워 300W	141000	70.0	220.0
4	핸디/스틱청소기	샤오미	드리미 V9	138800	60.0	200.0

```
# null data 확인
chart_data['흡입력'].isnull()
```

```
0      False
1      False
2      False
3      False
4      False
...
231     False
235     False
236     False
239     False
240     False
Name: 흡입력, Length: 123, dtype: bool
```

```
chart_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 123 entries, 0 to 240
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype
---  -
0   카테고리    123 non-null   object
1   회사명      123 non-null   object
2   제품        123 non-null   object
3   가격        123 non-null   int64
4   사용시간    123 non-null   float64
5   흡입력      123 non-null   float64
dtypes: float64(2), int64(1), object(3)
memory usage: 6.7+ KB
```


무선청소기 모델별 비교 분석

■ 데이터 시각화

```
# 흡입력, 사용시간의 최댓값/최솟값 정리
suction_max = chart_data['흡입력'].max()
suction_mean = chart_data['흡입력'].mean()
use_time_max = chart_data['사용시간'].max()
use_time_mean = chart_data['사용시간'].mean()
```

```
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# 청소기 성능 시각화
```

```
plt.figure(figsize=(20, 10))
```

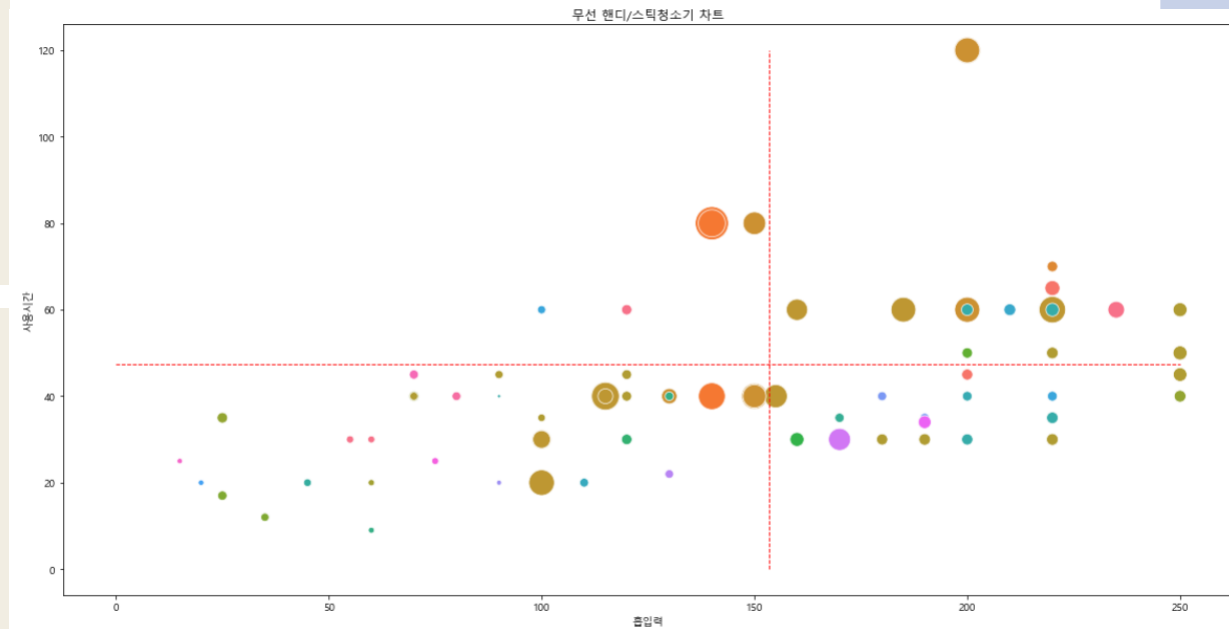
```
plt.title("무선 핸디/스틱청소기 차트")
```

```
sns.scatterplot(x = '흡입력', y = '사용시간', size = '가격', hue = chart_data['회사명'],
                data = chart_data, sizes = (10, 1000), legend = False)
```

```
plt.plot([0, suction_max], [use_time_mean, use_time_mean], 'r--', lw = 1 )
```

```
plt.plot([suction_mean, suction_mean], [0, use_time_max], 'r--', lw = 1 )
```

```
plt.show()
```



무선청소기 모델별 비교 분석

■ 데이터 시각화

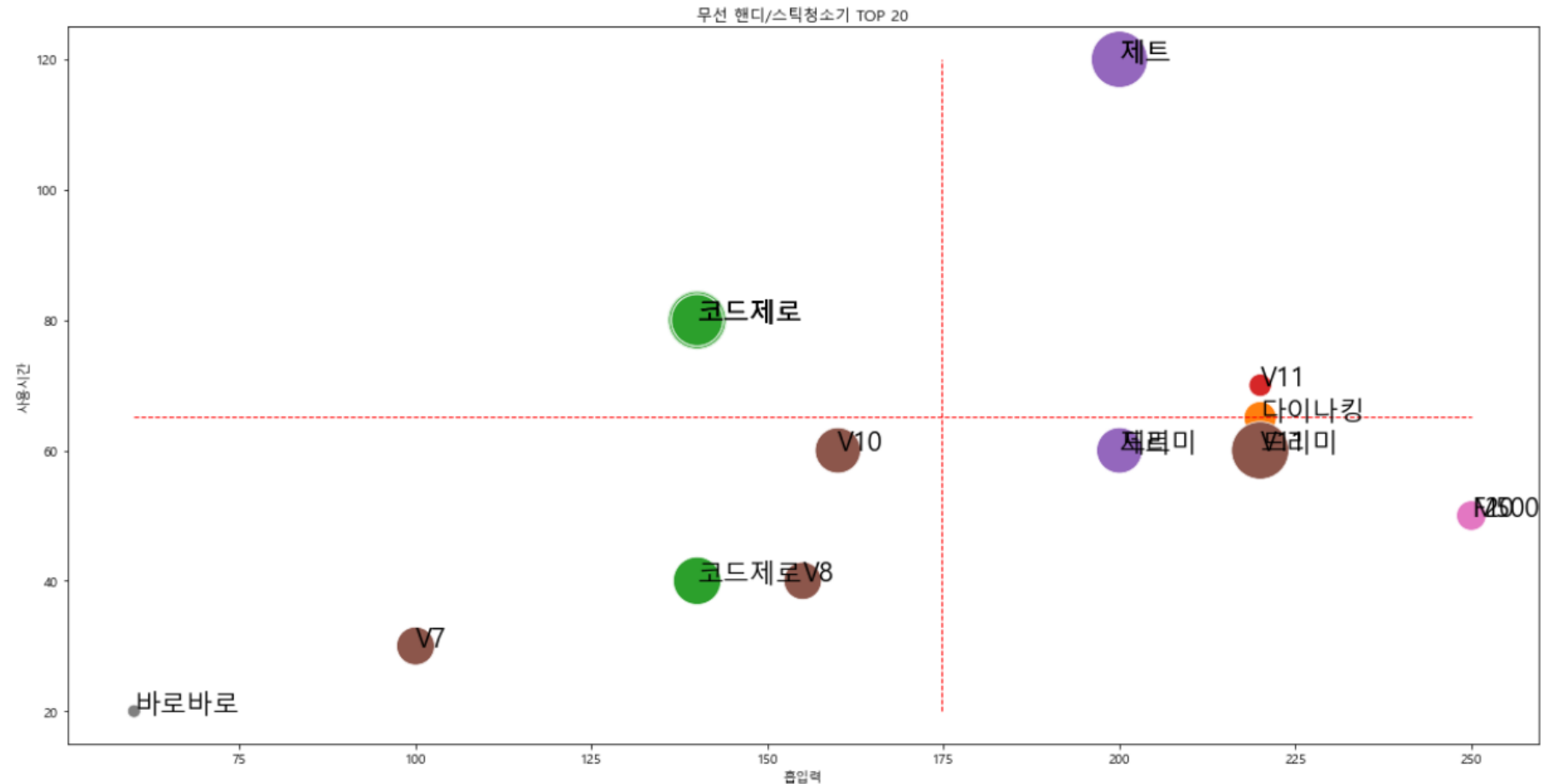
```
# 인기 제품 20개 선택
chart_data_selected = chart_data[:20]
len(chart_data_selected)

# 흡입력, 사용시간의 최댓값, 최솟값 구하기
suction_max = chart_data_selected['흡입력'].max()
suction_mean = chart_data_selected['흡입력'].mean()
use_time_max = chart_data_selected['사용시간'].max()
use_time_mean = chart_data_selected['사용시간'].mean()
plt.figure(figsize=(20, 10))
plt.title("무선 핸디/스틱청소기 TOP 20")
sns.scatterplot(x = '흡입력', y = '사용시간', size = '가격',
                hue = chart_data_selected['회사명'],
                data = chart_data_selected, sizes = (100, 2000),
                legend = False)

plt.plot([60, suction_max], [use_time_mean, use_time_mean], 'r--', lw = 1 )
plt.plot([suction_mean, suction_mean], [20, use_time_max], 'r--', lw = 1 )
for index, row in chart_data_selected.iterrows():
    x = row['흡입력']
    y = row['사용시간']
    s = row['제품'].split(' ')[0]
    plt.text(x, y, s, size=20)
plt.show()
```

무선청소기 모델별 비교 분석

■ 데이터 시각화



감사합니다
