웹에서 주문수 분석 테크닉

쇼핑몰 사이트 정보 분석

Python

- 데이터 불러오기
- 데이터 결합
- 매출데이터 조인
- 데이터 컬럼 만들기
- 통계량 파악/데이터 집계
- 매출 추이 시각화

쇼핑몰 사이트 정보 분석

Python

데이터 정보

No.	파일 이름	개요
1	customer_master.csv	고객 데이터, 이름, 성별 등
2	item_master.csv	취급하는 상품 데이터. 상품명, 가격 등
3–1	transaction_1.csv	구매내역 데이터
3–2	transaction_2.csv	3–1과 연결된 구매내역 분할 데이터
4–1	transaction_detail_1.csv	구매내역 상세 데이터
4–2	transaction_detail_2.csv	4–1과 연결된 분할 데이터

1. 구매내역, 구매내역 상세 데이터 세로로 연결하기

2. 구매내역 상세 데이터를 기준으로 고객정보, 제품정보를 가로로 연결

쇼핑몰 사이트 정보 분석

Python

구매내역 조인

```
transaction = pd.concat([transaction_1, transaction_2], ignore_index=True)
transaction.head()
```

```
print(len(transaction_1))
print(len(transaction_2))
print(len(transaction))
```

```
5000
1786
6786
```

쇼핑몰 사이트 정보 분석

Python

구매상세내역 조인

```
transaction_detail_2 = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/data/transaction_detail_2.csv')
transaction_detail=pd.concat([transaction_detail_1,transaction_detail_2], ignore_index=True)
transaction_detail.head()
```

쇼핑몰 사이트 정보 분석

Python

매출 데이터끼리 조인

```
join_data = pd.merge(transaction_detail, transaction[["transaction_id", "payment_date", "customer_id"]],
                     on="transaction_id", how="left")
join_data.head()
```

	detail_id	transaction_id	item_id	quantity	payment_date	customer_id
0	0	T0000000113	S005	1	2019-02-01 01:36:57	PL563502
1	1	T0000000114	S001	1	2019-02-01 01:37:23	HD678019
2	2	T0000000115	S003	1	2019-02-01 02:34:19	HD298120
3	3	T0000000116	S005	1	2019-02-01 02:47:23	IK452215
4	4	T0000000117	S002	2	2019-02-01 04:33:46	PL542865

쇼핑몰 사이트 정보 분석

Python

```
join_data = pd.merge(transaction_detail, transaction[["transaction_id", "payment_date", "customer_id"]],
                     on="transaction_id", how="left")
join_data.head()
```

	detail_id	transaction_id	item_id	quantity
0	0	T0000000113	S005	1
1	1	T0000000114	S001	1
2	2	T0000000115	S003	1

	transaction_id	price	payment_date	customer_id
0	T0000000113	210000	2019-02-01 01:36:57	PL563502
1	T0000000114	50000	2019-02-01 01:37:23	HD678019
2	T0000000115	120000	2019-02-01 02:34:19	HD298120
3	T0000000116	210000	2019-02-01 02:47:23	IK452215

쇼핑몰 사이트 정보 분석

Python

Customer_master 조인

```
join_data = pd.merge(join_data, customer_master, on="customer_id", how="left")
join_data.head()
```

item_master 조인 - 실습

	detail_id	transaction_id	item_id	quantity	payment_date	customer_id	customer_name	registration_date	email	gender	age	birth	pref	item_name	item_price
0	0	T0000000113	S005	1	2019-02-01 01:36:57	PL563502	김태경	2019-01-07 14:34	imoto_yoshimasa@example.com	M	30	1989-07-15	대전광역시	PC-E	210000
1	1	T0000000114	S001	1	2019-02-01 01:37:23	HD678019	김영웅	2019-01-27 18:00	mifune_rokurou@example.com	M	73	1945-11-29	서울특별시	PC-A	50000
2	2	T0000000115	S003	1	2019-02-01 02:34:19	HD298120	김강현	2019-01-11 8:16	yamane_kogan@example.com	M	42	1977-05-17	광주광역시	PC-C	120000
3	3	T0000000116	S005	1	2019-02-01 02:47:23	IK452215	김주한	2019-01-10 5:07	ikeda_natsumi@example.com	F	47	1972-03-17	인천광역시	PC-E	210000
4	4	T0000000117	S002	2	2019-02-01 04:33:46	PL542865	김영빈	2019-01-25 6:46	kurita_kenichi@example.com	M	74	1944-12-17	광주광역시	PC-B	85000

쇼핑몰 사이트 정보 분석

Python

필요한 데이터 컬럼 만들기

```
join_data["price"] = join_data["quantity"] * join_data["item_price"]
join_data[["quantity", "item_price","price"]].head()
```

	quantity	item_price	price
0	1	210000	210000
1	1	50000	50000
2	1	120000	120000
3	1	210000	210000
4	2	85000	170000

쇼핑몰 사이트 정보 분석

Python

데이터 검산

```
print(join_data["price"].sum())
print(transaction["price"].sum())
```

```
join_data["price"].sum() == transaction["price"].sum()
```

```
True
```

쇼핑몰 사이트 정보 분석

Python

결손치 파악

```
join_data.isnull().sum()
```

```
detail_id           0
transaction_id      0
item_id             0
quantity            0
payment_date        0
customer_id         0
customer_name       0
registration_date   0
email               0
gender              0
age                 0
birth               0
pref                0
item_name           0
item_price          0
price               0
dtype: int64
```

쇼핑몰 사이트 정보 분석

Python

기초통계량 파악하기

```
join_data.describe()
```

	detail_id	quantity	age	item_price	price
count	7144.000000	7144.000000	7144.000000	7144.000000	7144.000000
mean	3571.500000	1.199888	50.265677	121698.628219	135937.150056
std	2062.439494	0.513647	17.190314	64571.311830	68511.453297
min	0.000000	1.000000	20.000000	50000.000000	50000.000000
25%	1785.750000	1.000000	36.000000	50000.000000	85000.000000
50%	3571.500000	1.000000	50.000000	102500.000000	120000.000000
75%	5357.250000	1.000000	65.000000	187500.000000	210000.000000
max	7143.000000	4.000000	80.000000	210000.000000	420000.000000

쇼핑몰 사이트 정보 분석

Python

데이터 기간 파악

```
print(join_data["payment_date"].min())
print(join_data["payment_date"].max())
```

```
2019-02-01 01:36:57
2019-07-31 23:41:38
```

쇼핑몰 사이트 정보 분석

Python

월별 데이터 집계 - 연월 추출

```
join_data["payment_date"] = pd.to_datetime(join_data["payment_date"])
join_data["payment_month"] = join_data["payment_date"].dt.strftime("%Y%m")
join_data[["payment_date", "payment_month"]].head()
```

	payment_date	payment_month
0	2019-02-01 01:36:57	201902
1	2019-02-01 01:37:23	201902
2	2019-02-01 02:34:19	201902
3	2019-02-01 02:47:23	201902
4	2019-02-01 04:33:46	201902

쇼핑몰 사이트 정보 분석

Python

월 단위로 금액 집계

```
join_data.groupby("payment_month").sum()["price"]
```

```
payment_month
201902    160185000
201903    160370000
201904    160510000
201905    155420000
201906    164030000
201907    170620000
Name: price, dtype: int64
```

쇼핑몰 사이트 정보 분석

Python

Pivot_table 을 사용한 집계

```
pd.pivot_table(join_data, index='item_name', columns='payment_month', values=['price', 'quantity'], aggfunc='sum')
```

	price						quantity					
payment_month	201902	201903	201904	201905	201906	201907	201902	201903	201904	201905	201906	201907
item_name												
PC-A	24150000	26000000	25900000	24850000	26000000	25250000	483	520	518	497	520	505
PC-B	25245000	25500000	23460000	25330000	23970000	28220000	297	300	276	298	282	332
PC-C	19800000	19080000	21960000	20520000	21840000	19440000	165	159	183	171	182	162
PC-D	31140000	25740000	24300000	25920000	28800000	26100000	173	143	135	144	160	145
PC-E	59850000	64050000	64890000	58800000	63420000	71610000	285	305	309	280	302	341

행에 제품명, 열에 월, value 에는 집계하고 싶은 컬럼

쇼핑몰 사이트 정보 분석

Python

시각화

```
graph_data = pd.pivot_table(join_data, index='payment_month', columns='item_name', values='price', aggfunc='sum')
graph_data.head()
```

```
import matplotlib.pyplot as plt
%matplotlib inline
plt.plot(list(graph_data.index), graph_data["PC-A"], label='PC-A')
plt.plot(list(graph_data.index), graph_data["PC-B"], label='PC-B')
plt.plot(list(graph_data.index), graph_data["PC-C"], label='PC-C')
plt.plot(list(graph_data.index), graph_data["PC-D"], label='PC-D')
plt.plot(list(graph_data.index), graph_data["PC-E"], label='PC-E')
plt.legend()
```

```
<matplotlib.legend.Legend at 0x2617c72f160>
```