

# 이커머스 데이터셋을 활용한 동적가격방법론 적용

발표자 : 허선우



본 프로젝트는 프로젝트의 문제 상황을 사전에 정의하지 않고  
상향식 방법론과 하향식 방법론을 혼합한 디자인 사고 방법론을 이용하여  
다음과 같은 5단계 분석 절차를 수행



## 데이터 및 데이터의 컬럼 구성은 다음과 같음

## 사용 데이터셋

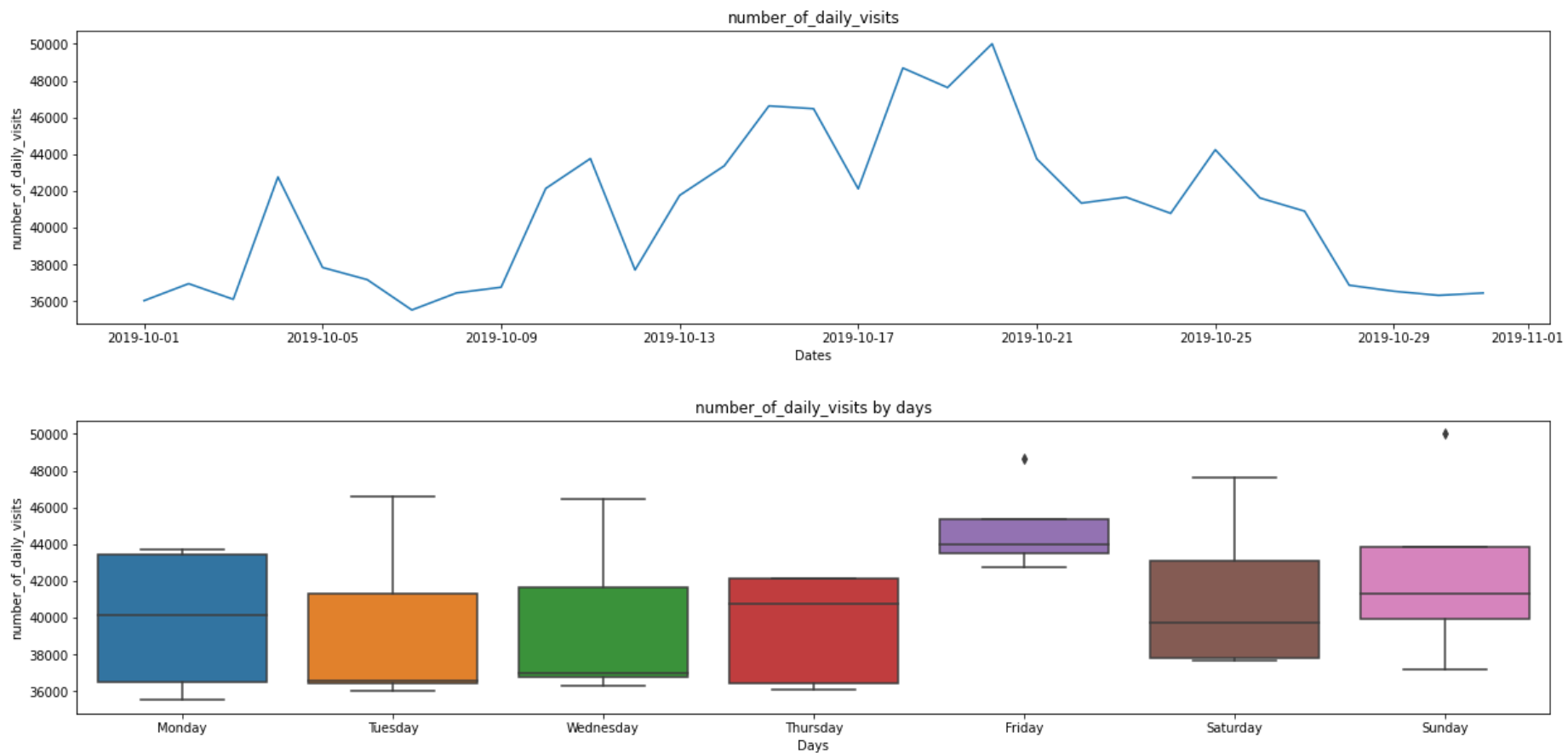
중동 이커머스 데이터셋, 2019년 10월 데이터, 스마트폰 카테고리의 삼성 스마트폰만 추출

event_time	이벤트가 발생한 시간 (UTC 기준)
event_type	이벤트 유형 (view, cart, remove_from_cart, purchase)
product_id	제품의 ID
category_id	제품의 카테고리 ID
category_code	제품의 카테고리 분류 (코드 이름)
brand	브랜드 이름
price	제품의 부동 소수점 가격
user_id	고정된 사용자 ID
user_session	일시적인 사용자 세션 ID. 각 사용자의 세션에 대해 동일. 사용종료 후 돌아올 때마다 변경

## 2. 문제 탐색

## 일별 방문 빈도

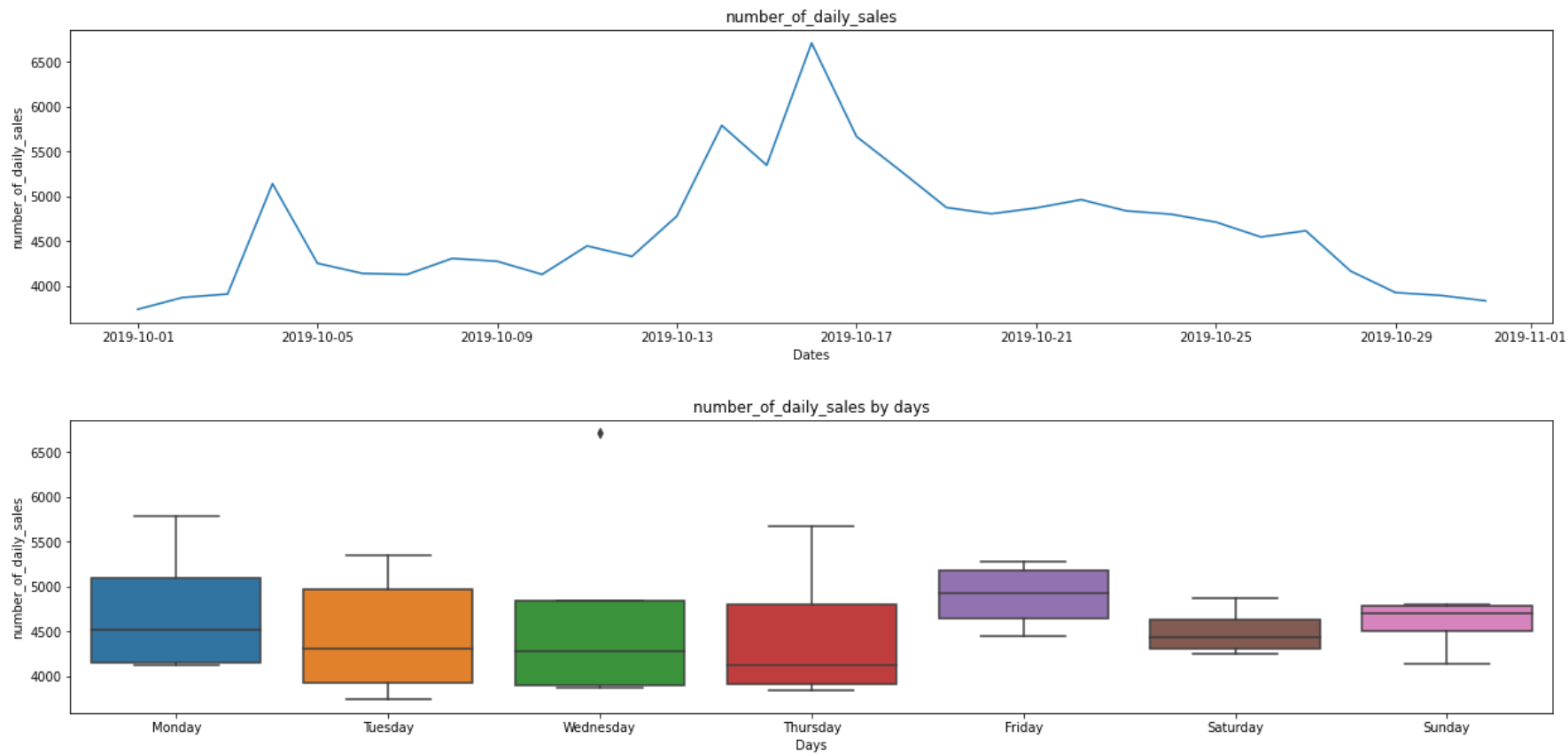
생성된 세션id 1개 당 방문 1회로 정의하고, 일별 방문 횟수를 시각화 하였음  
시각화 결과, 금요일의 방문 수가 다른 날에 비해 높은 것을 알 수 있었음



## 2. 문제 탐색

## 일별 판매량

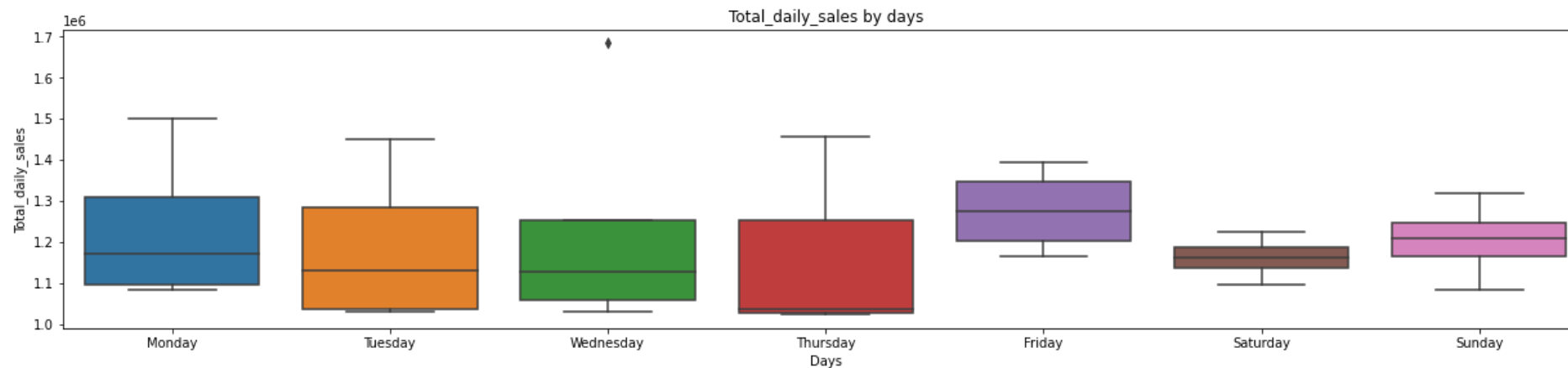
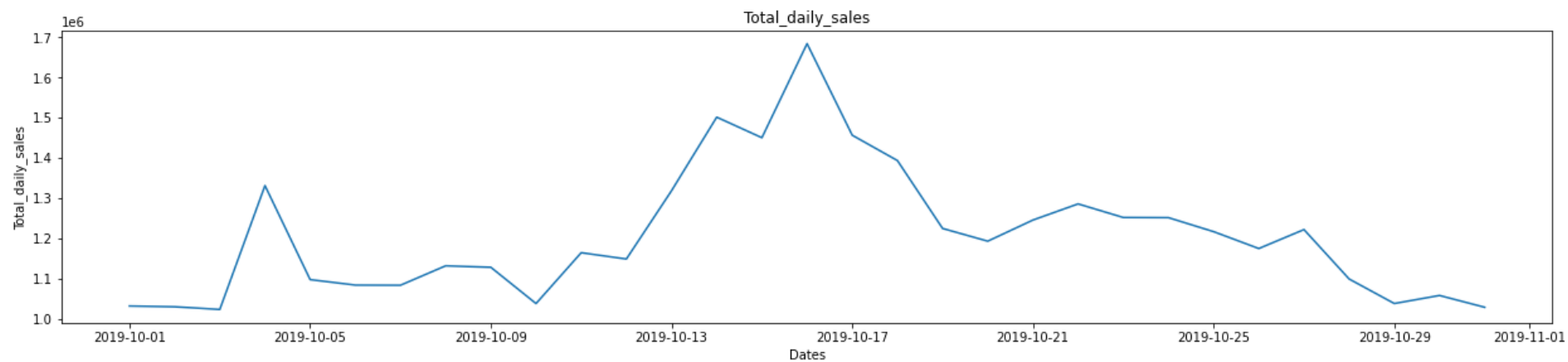
일별 판매 횟수를 시각화한 결과, 금요일의 판매량이 다른 날에 비해 높은 것을 확인함



## 2. 문제 탐색

## 일별 매출액

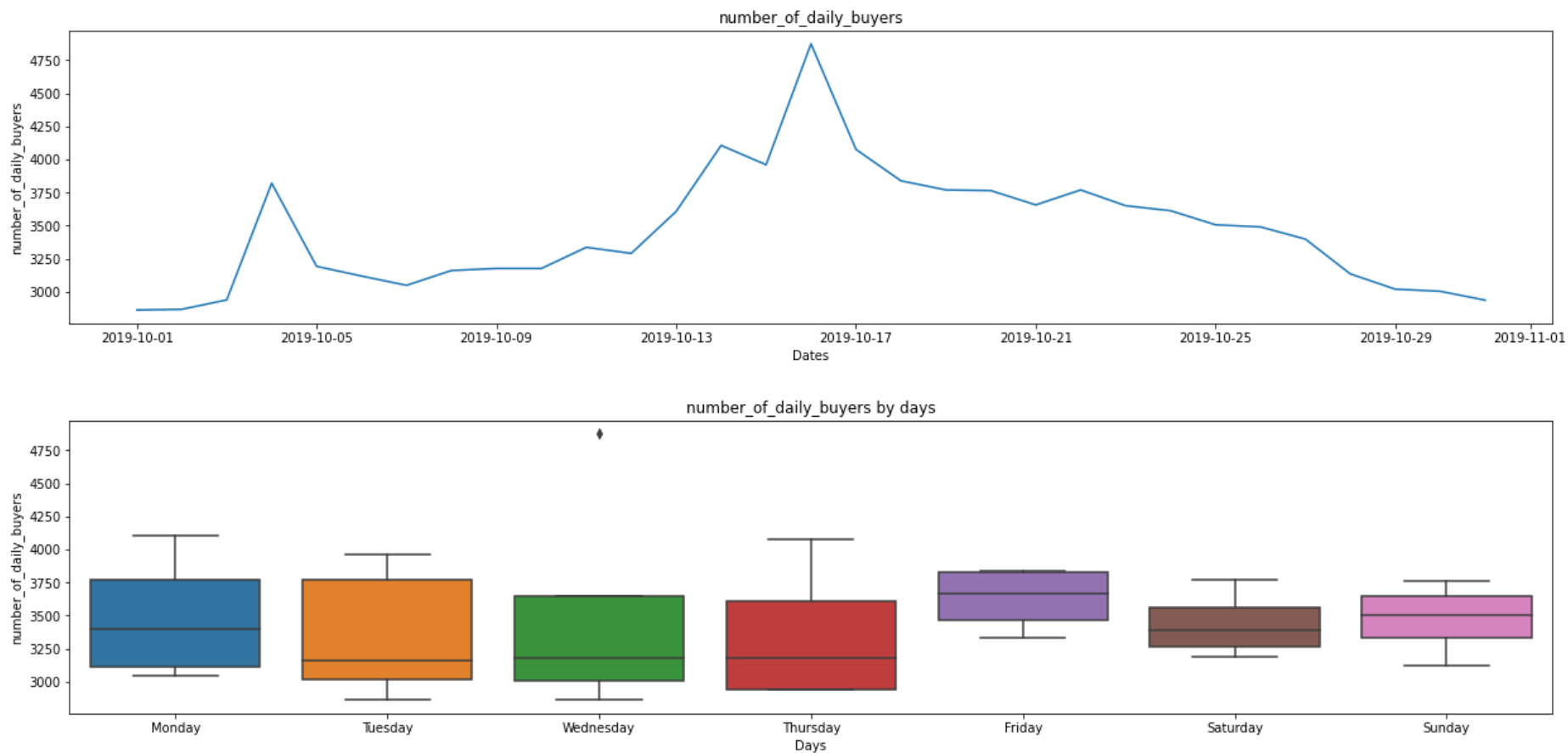
일별 매출액을 시각화한 결과, 금요일의 매출액이 다른 날에 비해 높은 것을 확인함



## 2. 문제 탐색

## 일별 구매자 수

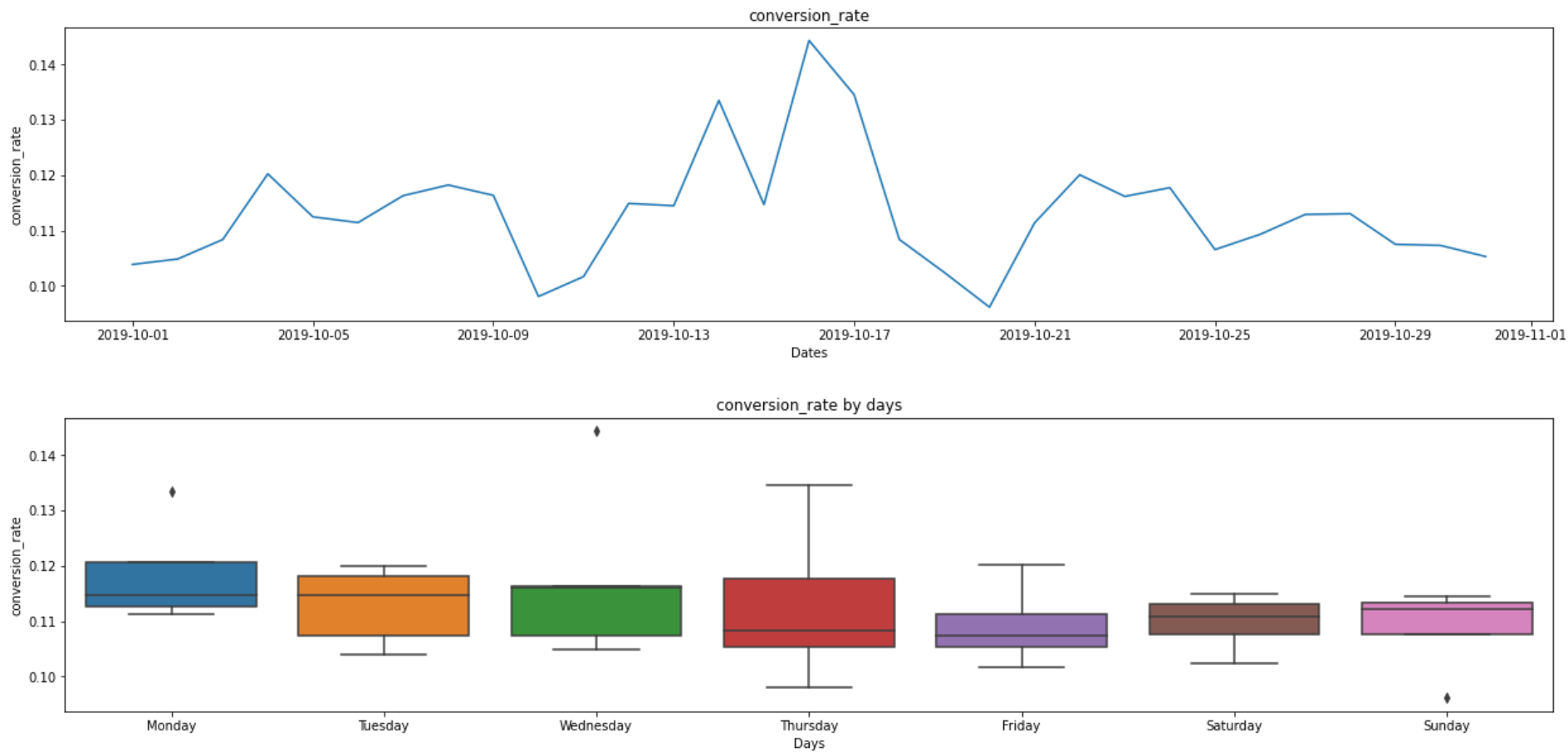
일별 구매자 수를 시각화한 결과, 금요일의 구매자 수가 가장 높은 것을 확인할 수 있었음



## 2. 문제 탐색

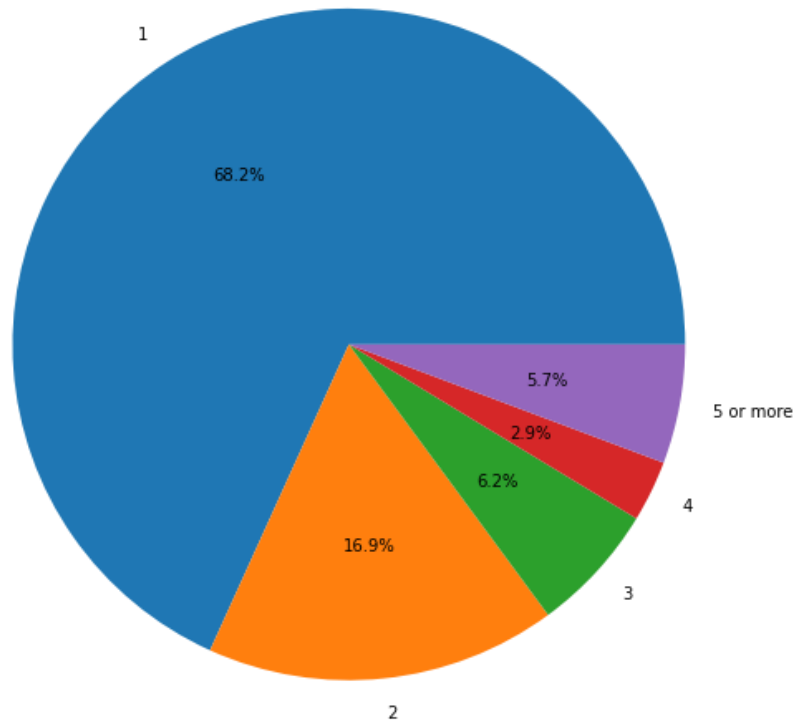
## 일별 전환율

한편, “판매량 / 방문 횟수 = 전환율”로 정의 하고  
전환율을 시각화한 결과, 평균적으로 금요일의 전환율이 가장 낮은 것으로 나타남





10월의 구매 고객별 구매횟수를 시각화 한 결과, 2회 이상 구매 고객은 전체의 31.8%에 불과함  
스마트폰의 구매 주기가 일반적으로 2년 내외임을 고려할 때, 다음의 결과는 타당성이 있어 보임



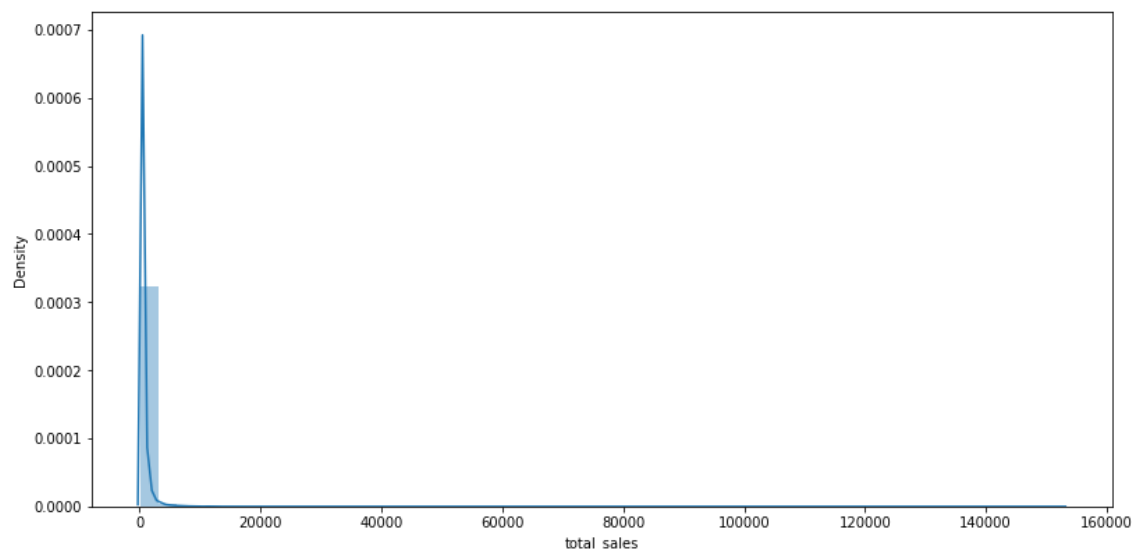
1. 10월 삼성 스마트폰 구매 고객은 총 76,524명

2. 고객들의 구매 횟수 분포:

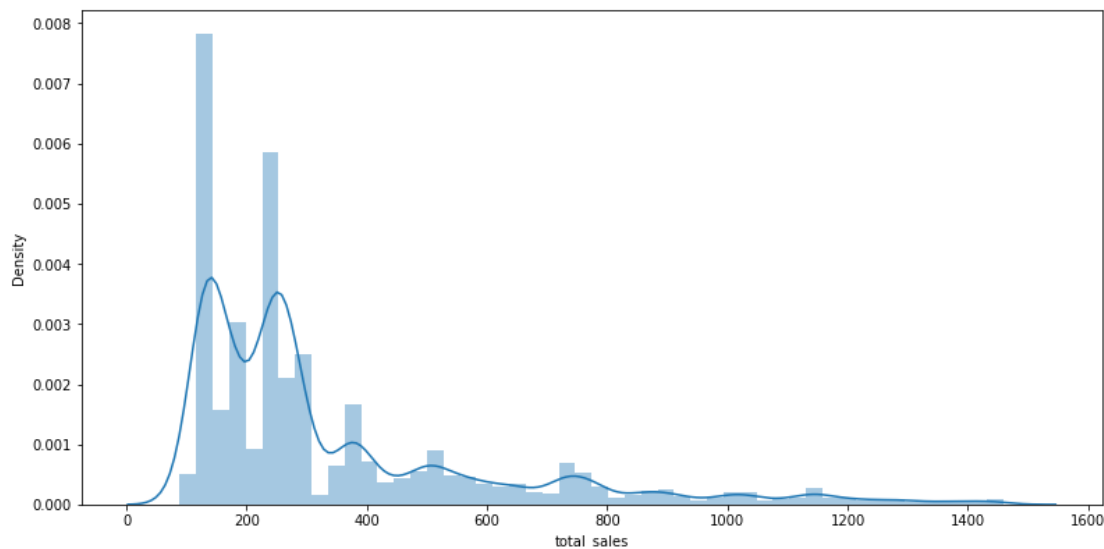
- 68.2%의 고객이 1회 구매
- 16.9%의 고객이 2회 구매
- 6.2%의 고객이 3회 구매
- 2.9%의 고객이 4회 구매
- 5.7%의 고객이 5회 이상 구매

매출 상위 5%를 제외한 고객의 최대 구매 금액은 1,459.3달러로, 전체 고객의 최대 구매 금액인 152,800.1달러보다 훨씬 낮음  
이는 소수의 고객이 매우 높은 매출을 올리고 있음을 보여줌

### 전체 고객의 매출 분포



### 매출 상위 5%를 제외한 고객의 매출 분포



### 3. 문제 정의

### 연구 가설 수립

본 과제에서는 프로젝트의 목적을 매출증대로 설정하고 예상 해결 방안으로 Dynamic Pricing을 제안하고자 함  
이에 따라, Dynamic Pricing의 적용의 타당성을 확인하기 위해  
가격 변동과 방문빈도 변화, 장바구니 이용 빈도 변화, 판매량 변화와의 상관관계를 검증하였음

프로젝트 목적	매출 증대
예상 해결방안	Dynamic Pricing 활용
문제정의	1. 어떻게 Dynamic Pricing을 적용할 것인가?
	2. Dynamic Pricing은 과연 효과가 있을 것인가? (가설 검증을 통한 필요성 제시)
연구 가설 수립	1. 가격 변동은 방문 빈도 변화와 음(-)의 상관관계를 갖는다
	2. 가격 변동은 장바구니 이용 빈도 변화와 음(-)의 상관관계를 갖는다
	3. 가격 변동은 판매량 변화와 음(-)의 상관관계를 갖는다

## 4. 가설 검정

세 가지 가설 모두, 가격 변동과 음의 상관관계를 가지며, 유의확률 0.05 미만이므로, 검증 결과에 대한 타당성을 갖는다고 볼 수 있음  
결과적으로 가격이 하락하면, 방문빈도, 장바구니 이용빈도, 판매량은 상승한다고 볼 수 있음  
따라서, Dynamic Pricing의 효과가 클 것으로 예상됨

### 1. 가격 변동은 방문 빈도 변화와 음(-)의 상관관계를 갖는다

Pearson Correlation:  $-0.21$  (p-value: 0.000)

### 2. 가격 변동은 장바구니 이용 빈도 변화와 음(-)의 상관관계를 갖는다

Pearson Correlation:  $-0.18$  (p-value: 0.023)

### 3. 가격 변동은 판매량 변화와 음(-)의 상관관계를 갖는다

Pearson Correlation:  $-0.22$  (p-value: 0.000)

Dynamic Pricing을 위해, XGBoost와 수리 최적화 모델을 사용하고자 하며, 적용 순서는 다음과 같음

1. XGBoost로 판매량을 예측 모델 구축 (시계열 변수와 다음 주차 판매 예상가격을 독립 변수로 이용)
2. 아래 목적식과 제약식에 따라, 수리 모델링을 진행하여 최적해(적정가격)를 산출

- 변수 정의

$$P_i = \{i | i = \text{Price of product id}(i)\}$$

$$Q_i = \{i | i = \text{Quantity of product id}(i)\}$$

$$P_{pred_i} = \{i | i = \text{Predicted price of product id}(i)\}$$

$$Q_{pred_i} = \{i | i = \text{Predicted quantity of product id}(i)\}$$

- 목적식과 제약식

$$\text{Max} : \sum_{i=1}^n P_{pred_i} * Q_{pred_i}$$

$$\text{Constraints}_1 : 0.95P_i < P_{pred_i}$$

$$\text{Constraints}_2 : P_{pred_i} < 1.1P_i$$

## 6. 효과 검증

## XGBoost 성능지표 확인

XGBoost를 이용하여 미래 판매가격과 판매수량을 예측한 결과 MAPE 기준 0.02, 0.46으로 나타남  
기 학습된 XGBoost 모델을 바탕으로 최적화를 진행하여 Dynamic Pricing을 수행할 수 있었음

### 실제 판매결과

	product_id	price	sales
0	1000978	301.140000	5
1	1002098	370.640000	1
2	1002100	370.640000	3
3	1002101	370.640000	16
4	1002103	414.530000	2
...	...	...	...
104	1005211	193.954200	50
105	1005212	194.060361	83
106	1005215	1544.284211	19
107	1005216	1544.152500	4
108	1005225	193.833529	17

### XGBoost 예측결과

price_prediction	sales_prediction
334.414642	6.249163
361.696716	3.643797
369.174194	6.249163
374.696167	13.189700
463.568512	30.598507
...	...
194.886139	27.261774
195.040405	29.114676
1509.487793	11.183790
1509.828857	7.024426
188.986221	12.521174

### 예측결과 최적화

optimized_price	optimized_sales
331.254000	6
407.704000	3
407.704000	6
407.704000	13
383.603006	49
...	...
213.349620	27
213.466398	29
1698.712632	11
1698.567750	7
213.216882	12

### XGBoost 성능지표

#### 판매가격 예측모델

MAE: 11.91  
MSE: 399.23  
RMSE: 19.98  
MAPE: 0.02

#### 판매수량 예측모델

MAE: 29.89  
MSE: 5275.53  
RMSE: 72.63  
MAPE: 0.46

제약식을 수정하여 민감도 분석을 수행한 결과, 가격 범위가 넓어질 수록 예상 매출 증가율이 높아지는 것을 확인 할 수 있었음  
 한편, 가격 상승에 대한 예상치 못한 부작용과 가격하한이 낮아지더라도 예상 매출 증가율이 높아지는 것을 고려할 때,  
 가격 하한을 낮추어 Dynamic Pricing을 수행하는 것을 제안함

$$Constraints_1 : 0.95P_i < P_{pred_i}$$

$$Constraints_2 : P_{pred_i} < 1.05P_i$$

매출(실제) : 4221609.1  
 매출(예측) : 4206942.5  
 매출(최적화): 4439896.54  
 예상 매출 증가율: 0.06

$$Constraints_1 : 0.90P_i < P_{pred_i}$$

$$Constraints_2 : P_{pred_i} < 1.05P_i$$

매출(실제) : 4221609.1  
 매출(예측) : 4206942.5  
 매출(최적화): 4444382.43  
 예상 매출 증가율: 0.06

$$Constraints_1 : 0.9P_i < P_{pred_i}$$

$$Constraints_2 : P_{pred_i} < 1.1P_i$$

매출(실제) : 4221609.1  
 매출(예측) : 4206942.5  
 매출(최적화): 4700026.15  
 예상 매출 증가율: 0.12

$$Constraints_1 : 0.8P_i < P_{pred_i}$$

$$Constraints_2 : P_{pred_i} < 1.1P_i$$

매출(실제) : 4221609.1  
 매출(예측) : 4206942.5  
 매출(최적화): 4787299.11  
 예상 매출 증가율: 0.14

**감사합니다**