

Chipotle

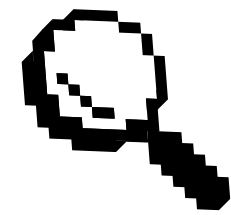
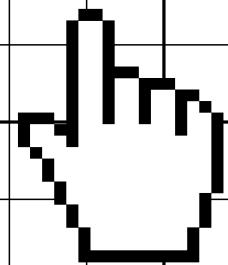
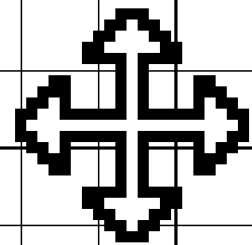
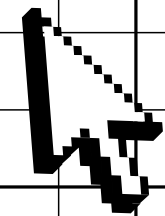
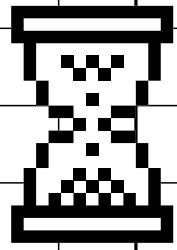
주문 데이터 분석

6조

서윤선, 노현정, 한슬기

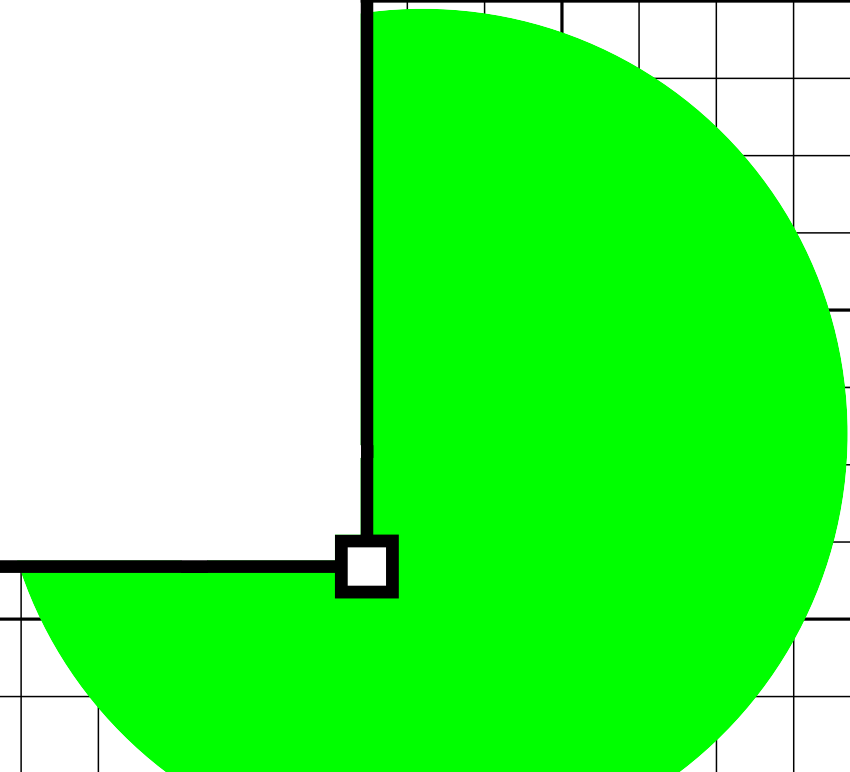
데이터

보내기



목차

- 1 데이터 읽기
- 2 데이터 기초 정보 탐색
- 3 인사이트 발견을 위한 탐색 & 시각화
- 4 전처리 & 통계
- 5 탐색적 데이터 분석 (질문하기)



데이터 분석 과정

1 기본정보 탐색

- 1-1. head(), info(), describe(), isnull()
- 1-2. 그 외 다양한 방법으로 기본 정보 탐색

2 주요 정보 탐색

- 2-1. 메뉴 갯수 및 메뉴별 주문갯수 확인
- 2-2. Top10메뉴 확인

3 데이터 전처리

- 3-1. 수 데이터 확인
- 3-2. 평균/표준편차 구하기

4 추가사항

- 4-1. Top10 메뉴별 가격 같이 출력
- 4-2. 질문 생각해 보기
- Q. 상위 소비자가 구매한 상품과 인기 메뉴의 순위도 일치하는가?

데이터 분석 보고서

기본정보

5개 항목 (order_id , quantity ,item_name ,choice_description, item_price) / 4622개 데이터

order ID: data['item_name'].unique() / total: 1834개 order

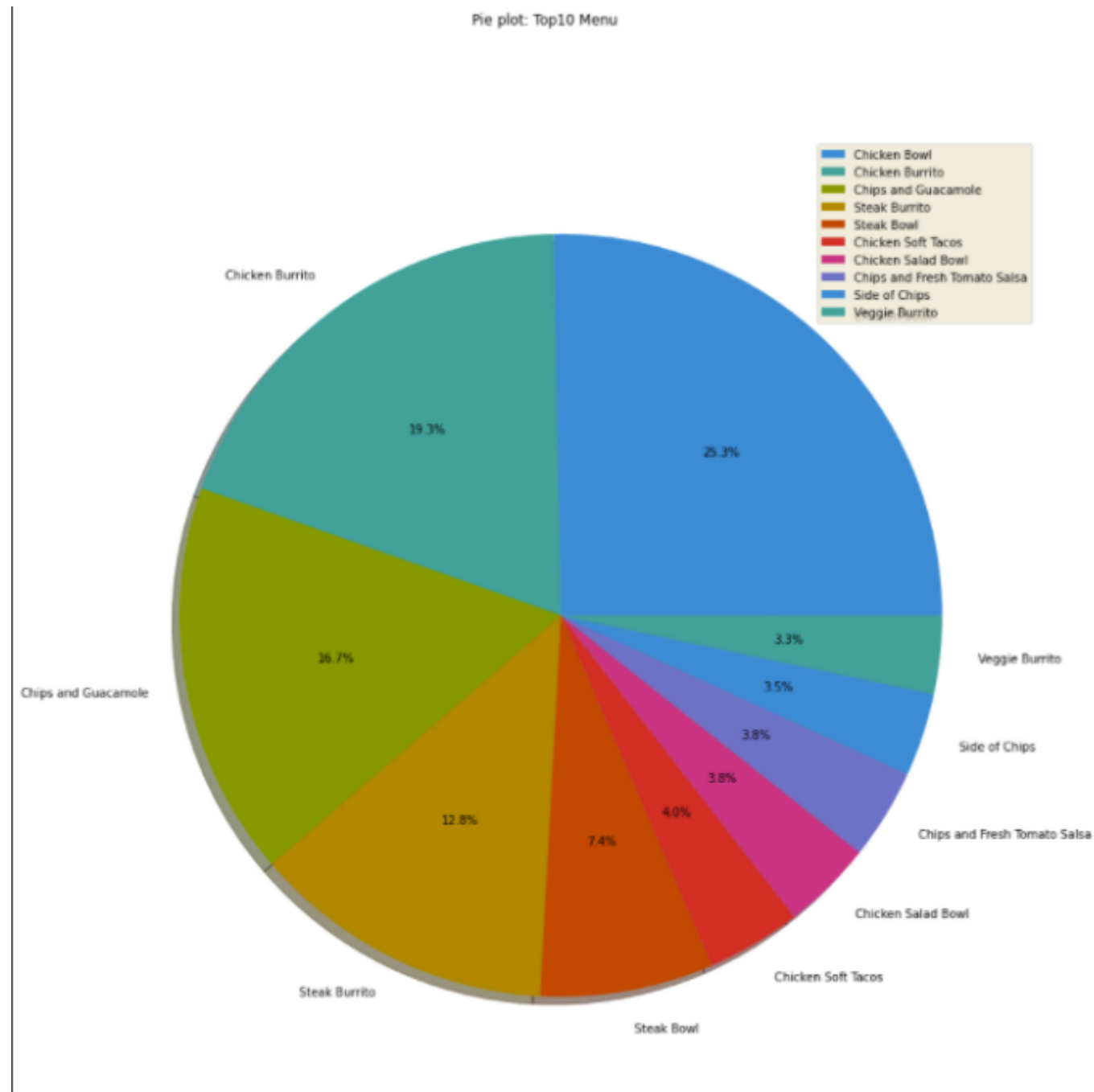
레스토랑 메뉴 갯수: data['item_name'].unique() / total: 50개

텍스트를 써주세요

데이터 결측값: data.isnull().sum() / choice_description 항목에서 1246개

인기메뉴 Top10

TOP10메뉴에서 의미없는 값을 제외한 인기메뉴 10개의 항목을 추출 및 시각화 함.



```
#물과 음료 drop
del new_Top10['Canned Soft Drink']
del new_Top10['Canned Soda']
del new_Top10['Bottled Water']
del new_Top10['Chips']
new_Top10
```

```
{'Chicken Bowl': 726,
'Chicken Burrito': 553,
'Chicken Salad Bowl': 110,
'Chicken Soft Tacos': 115,
'Chips and Fresh Tomato Salsa': 110,
'Chips and Guacamole': 479,
'Side of Chips': 101,
'Steak Bowl': 211,
'Steak Burrito': 368,
'Veggie Burrito': 95}
```

```
new_df_top10 = pd.DataFrame(new_Top10, index = ['cnt'] )
top10_plot = new_df_top10.T
top10_plot
```

	cnt
Chicken Bowl	726
Chicken Burrito	553
Chips and Guacamole	479
Steak Burrito	368
Steak Bowl	211
Chicken Soft Tacos	115
Chicken Salad Bowl	110
Chips and Fresh Tomato Salsa	110
Side of Chips	101
Veggie Burrito	95

인기메뉴TOP10

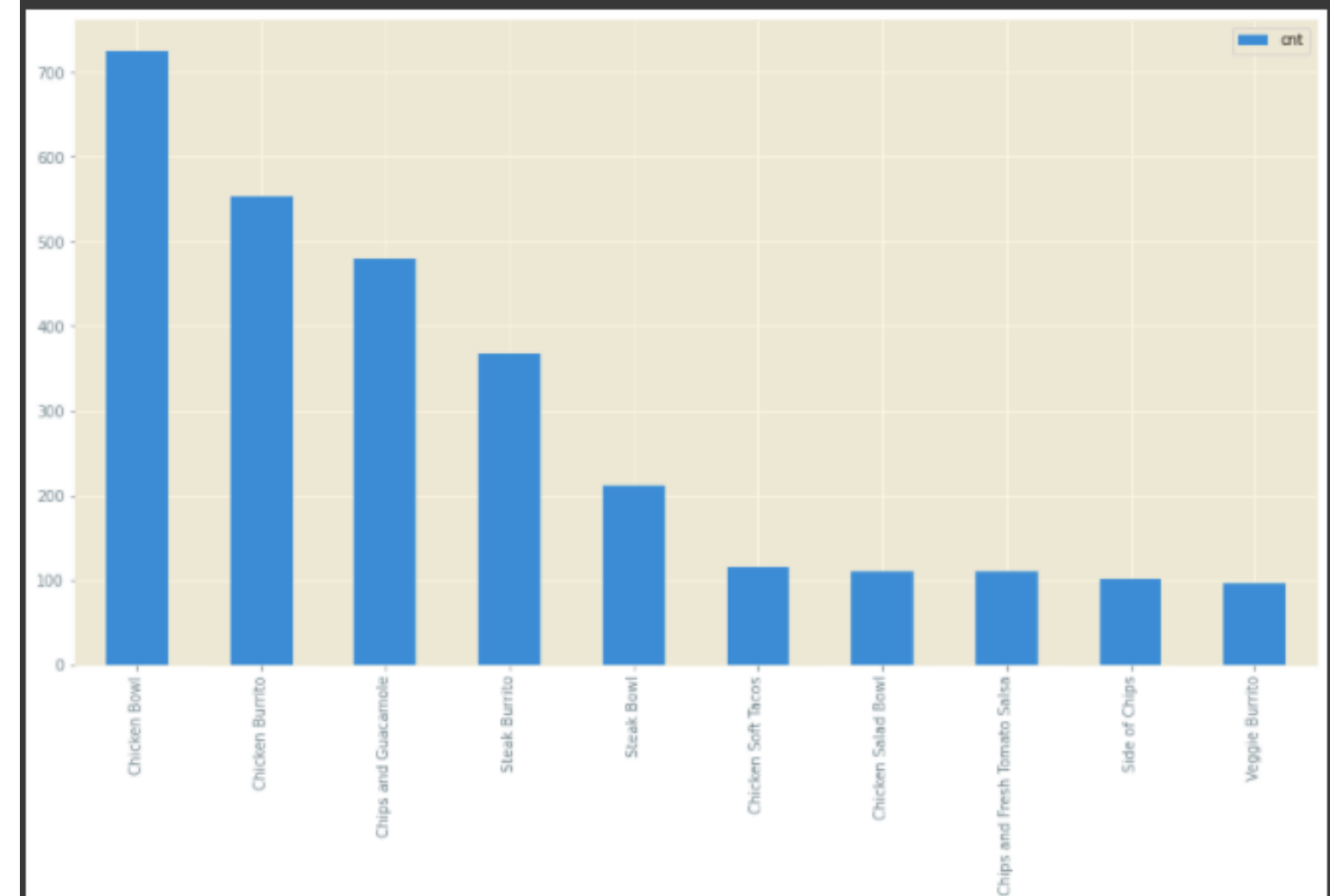
Bar Chart로 시각화

인사이트(Insight):

같은 데이터 라도 Pie Chart는 비율을 쉽게 확인 할 수 있고, Bar Chart는 수량으로 파악되어 메뉴간 판매량 차이가 더 쉽게 파악함.

CHECK!

```
# 의미없는 항목을 뺀 인기 메뉴 Bar chart로 시각화
top10_plot.plot(kind = 'bar', figsize=(15,8))
plt.show()
```



1

데이터 변환

object 자료형을 수자료형으로

```
[92] #데이터 타입 변경 방법: 1
data['item_price'] = data['item_price'].str.replace('$','').astype(float)

[ ] # 데이터 타입 변경 방법2. (참고자료:http://hleecaster.com/pandas-data-cleaning-2/)
data["item_price"] = data["item_price"].replace('[\$',]', '', regex=True)
data["item_price"] = pd.to_numeric(data["item_price"])
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4622 entries, 0 to 4621
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   order_id        4622 non-null   int64
1   quantity        4622 non-null   int64
2   item_name       4622 non-null   object
3   choice_description 3376 non-null   object
4   item_price      4622 non-null   float64
5   menu_no         4622 non-null   int64
dtypes: float64(1), int64(3), object(2)
memory usage: 216.8+ KB
```

2

평균 구하기

```
print('레스토랑 메뉴 갯수: ', len(menu), '개')
print('테이블당 주문 수량 평균: ', round(data_mean['quantity'],0), '개')
print('테이블당 지불 금액 평균: ', round(data_mean['item_price'],0), '달러')
print('가격이 가장 비싼 메뉴: ',df_item_price['item_price'].max(), '달러', '메뉴이름:', hpm['item_name'].unique())
print('\n')
print('인기 메뉴 TOP10: ')
top10_menu

레스토랑 메뉴 갯수: 50 개
테이블당 주문 수량 평균: 3.0 개
테이블당 지불 금액 평균: 19.0 달러
가격이 가장 비싼 메뉴: 9.39 달러 메뉴이름: ['Steak Salad Bowl' 'Barbacoa Salad Bowl' 'Carnitas Salad Bowl']
```

데이터 뽐내기

1. Heatmap

2. Crosstab

데이터의 상관관계를 다양한 시각화 방법을
통하여 알아봄.

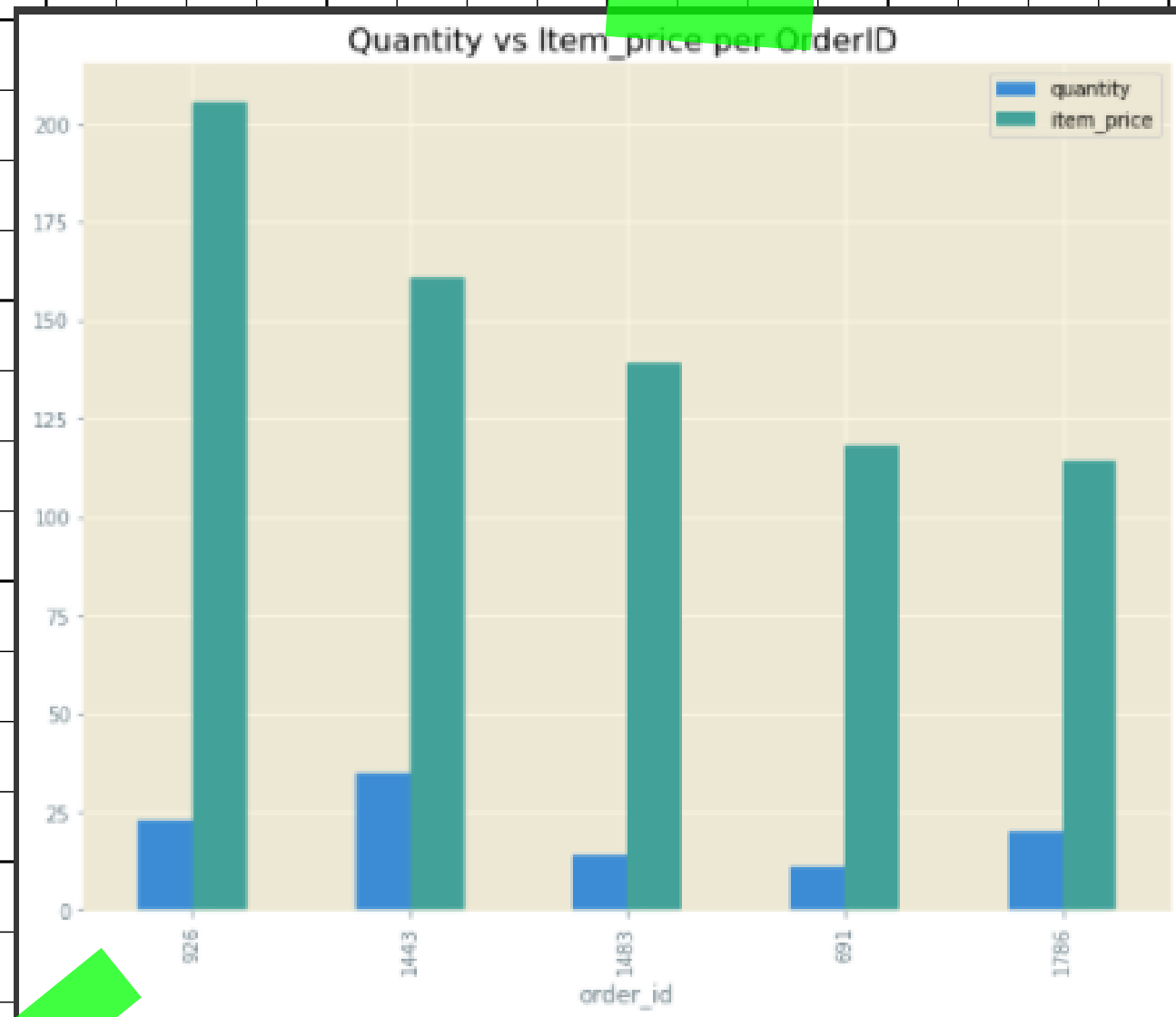
CHECK!



피어슨 상관관계 분석

```
1 # 주문량별 메뉴 분석표  
2 pd.crosstab(data['quantity'], data['menu_no'], margins=True).style.background_gradient(cmap='PuBu')
```

menu_no	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	All				
quantity																																																							
1	693	521	455	352	264	202	193	139	110	105	97	84	99	94	91	83	66	66	58	54	53	46	44	41	40	34	29	27	25	25	21	20	17	18	18	10	10	9	6	6	6	6	6	4	2	2	1	1	1	1	4355				
2	31	28	22	14	27	8	17	12	5	3	13	19	0	0	0	2	1	0	1	1	1	2	3	2	0	1	2	2	2	0	1	0	1	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	224			
3	2	2	1	2	7	1	1	6	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	28			
4	0	2	1	0	3	0	0	2	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10				
5	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1				
7	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1				
8	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1				
10	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1				
15	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1				
All	726	553	479	368	301	211	211	162	115	110	110	104	101	95	91	85	68	66	59	55	54	48	47	43	40	35	31	29	27	25	22	20	20	18	18	11	10	9	7	7	6	6	6	4	2	2	2	1	1	1	4622				



```
[107] # 상위 소비자 TOP5
grouped_data = data[['quantity', 'item_price']].groupby(data['order_id']).sum()
a = grouped_data.sort_values('item_price', ascending=False)
order_top5 = a.iloc[:5, :]
order_top5.reset_index()
```

	order_id	quantity	item_price
0	926	23	205.25
1	1443	35	160.74
2	1483	14	139.00
3	691	11	118.25
4	1786	20	114.30

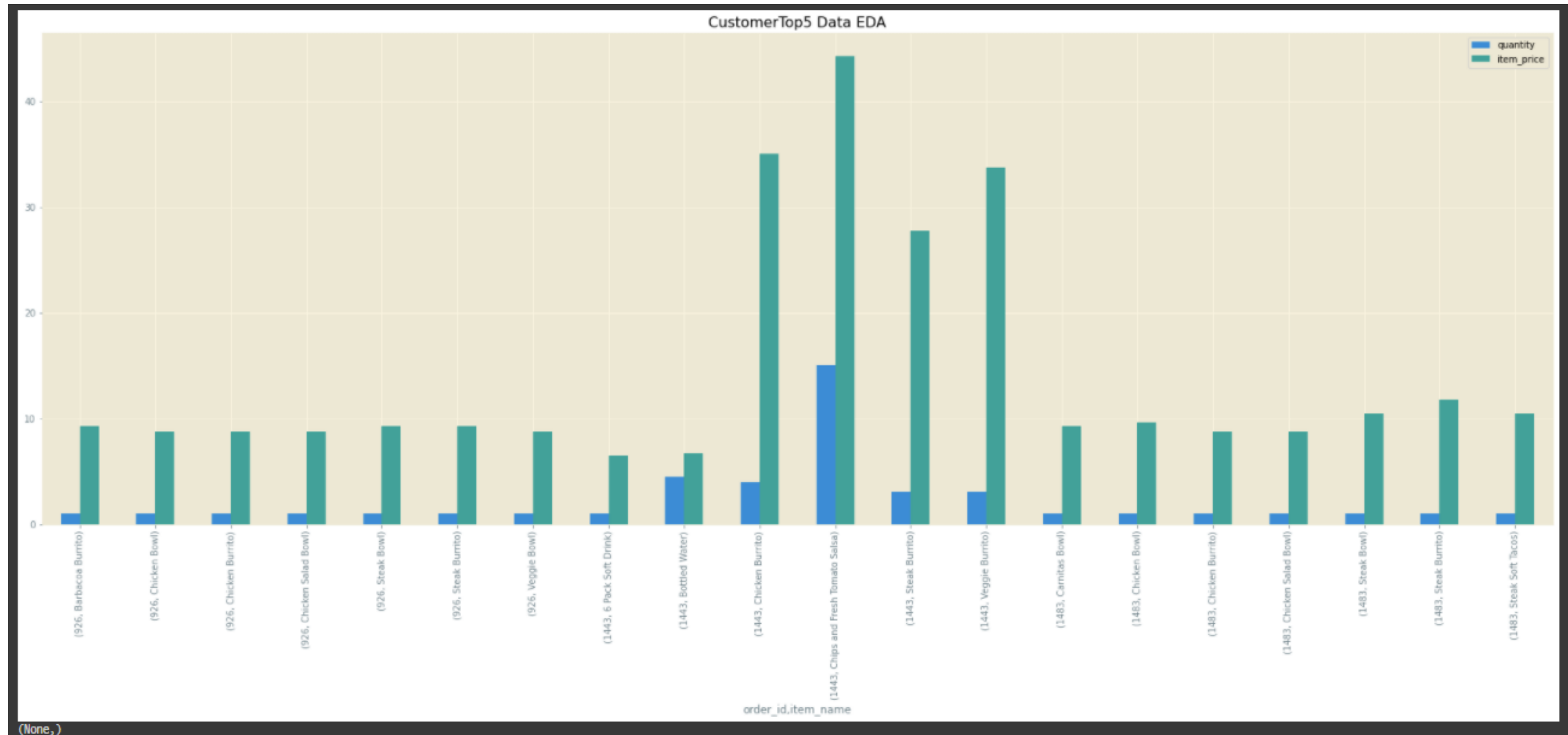
추가질문

상위소비자 TOP5 데이터 분석

가설 :

인기메뉴와 상위소비자를 비교분석하여
주요 타겟 및 시장 내 포지션을 설정 할 수 있는
의미있는 데이터를 도출할 수 있다.

상위 소비자 Top3



1

상위소비자 Top3

참고자료: <https://hayden-archive.tistory.com/286>
 aa = df_order_top.groupby(by=['order_id', 'item_name'])
 aa.sum()

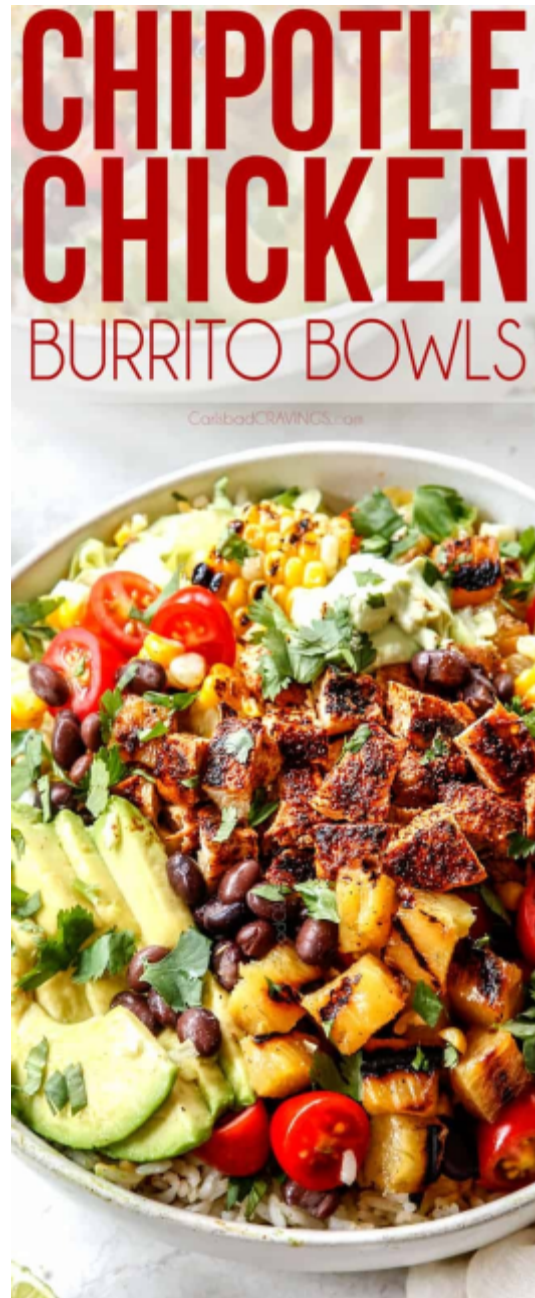
order_id	item_name	quantity	item_price
926	Barbacoa Burrito	1	9.25
	Chicken Bowl	9	78.75
	Chicken Burrito	4	35.00
	Chicken Salad Bowl	1	8.75
	Steak Bowl	5	46.25
	Steak Burrito	2	18.50
	Veggie Bowl	1	8.75
1443	6 Pack Soft Drink	1	6.49
	Bottled Water	9	13.50
	Chicken Burrito	4	35.00
	Chips and Fresh Tomato Salsa	15	44.25
	Steak Burrito	3	27.75
	Veggie Burrito	3	33.75
	Carnitas Bowl	1	9.25
1483	Chicken Bowl	3	28.75
	Chicken Burrito	1	8.75

2

인기메뉴 Top10

	cnt
Chicken Bowl	726
Chicken Burrito	553
Chips and Guacamole	479
Steak Burrito	368
Steak Bowl	211
Chicken Soft Tacos	115
Chicken Salad Bowl	110
Chips and Fresh Tomato Salsa	110
Side of Chips	101

Chipotle 선호도 1위 == 너도나도 'Chichken Burrito & Bowl '



결론:

상위소비자 TOP5 데이터 분석 결과
상위 소비자가 선호하는 메뉴와 인기메뉴가 상당부분 동
이한 것으로 분석되었다. 따라서 메뉴 리뉴얼 , 이벤트등
의 사업 프로젝트 기획시 이런 점을 충분히 고려해 볼 수
있을 것으로 생각된다.

또한, 추가 매출 지표에 대한 분석을 통하여
보다 정확한 시장분석이 가능하다.

😊 노현정 코멘트: 팀원분들과 수업 외에 추가적인 메소드, 시각화 툴, 자료에서 인사이트를 얻는 법을 공유할 수 있어서 많은 도움이 되었습니다~

🦊 서윤선 코멘트:

EDA 어렵지만, 고민한 만큼 재미있는 분야일 것 같습니다. 부족한 건 항상 채워지지 않는듯 싶어서 막막했지만, 어려운 문제일수록 팀원들과 더 많이 얘기하고 같이 할 수 있었던 것 같아 좋았습니다.

👉 한슬기 코멘트: 너무너무 어렵고.. 노트북이 터질 것 같지만.. 오랜만에 뇌용량을 잔뜩 쓴 기분이라 마치 힘든 운동을 마친 기분..? 열심히 코딩 운동 하겠습니다.

감사합니다!

THANK YOU!