

```
1 from google.colab import files
2 uploaded = files.upload()
```

파일 선택 chipotle.tsv

- **chipotle.tsv**(n/a) - 364975 bytes, last modified: 2021. 7. 2. - 100% done
Saving chipotle.tsv to chipotle (1).tsv

```
1 import pandas as pd
2 chipo = pd.read_csv("chipotle.tsv", encoding="cp949", sep="Wt ")
3 chipo.shape

(4622, 5)
```

1 chipo.head(10) #앞부터 10개 데이터 출력

	order_id	quantity	item_name	choice_description	item_price
0	1	1	Chips and Fresh Tomato Salsa	NaN	\$2.39
1	1	1	Izze	[Clementine]	\$3.39
2	1	1	Nantucket Nectar	[Apple]	\$3.39
3	1	1	Chips and Tomatillo-Green Chili Salsa	NaN	\$2.39
4	2	2	Chicken Bowl	[Tomatillo-Red Chili Salsa (Hot), [Black Beans...	\$16.98
5	3	1	Chicken Bowl	[Fresh Tomato Salsa (Mild), [Rice, Cheese, Sou...	\$10.98
6	3	1	Side of Chips	NaN	\$1.69
7	4	1	Steak Burrito	[Tomatillo Red Chili Salsa, [Fajita Vegetables...	\$11.75
8	4	1	Steak Soft Tacos	[Tomatillo Green Chili Salsa, [Pinto Beans, Ch...	\$9.25
9	5	1	Steak Burrito	[Fresh Tomato Salsa, [Rice, Black Beans, Pinto...	\$9.25

1 chipo.describe()

	order_id	quantity
count	4622.000000	4622.000000
mean	927.254868	1.075725

```
1 chipo['quantity'].describe() #기초 통계 구하는 함수 숫자 데이터만 가능 .describe()
```

```
count      4622.000000
mean        1.075725
std         0.410186
min         1.000000
25%         1.000000
50%         1.000000
75%         1.000000
max         15.000000
Name: quantity, dtype: float64
```

```
1 chipo['order_id'].unique()
```

```
array([ 1, 2, 3, ..., 1832, 1833, 1834])
```

```
1 len(chipo['order_id'].unique()) #order_id의 개수 중복을 제거함
```

```
1834
```

```
1 len(chipo['item_name'].unique()) #item_name의 개수 중복을 제거함
```

```
50
```

```
1 # 가장 많이 판매된 메뉴 top 10은?
```

```
2 chipo['item_name'].value_counts()[:10]
```

```
Chicken Bowl      726
Chicken Burrito   553
Chips and Guacamole 479
Steak Burrito     368
Canned Soft Drink 301
Chips             211
Steak Bowl        211
Bottled Water     162
Chicken Soft Tacos 115
Chips and Fresh Tomato Salsa 110
Name: item_name, dtype: int64
```

```
1 item_count=chipo['item_name'].value_counts()[:10]
```

```
2 for index,(val,cnt) in enumerate(item_count.iteritems(), start=1):
```

```
3     print("Top",index,":",val,cnt)
```

```
Top 1 : Chicken Bowl 726
Top 2 : Chicken Burrito 553
Top 3 : Chips and Guacamole 479
Top 4 : Steak Burrito 368
Top 5 : Canned Soft Drink 301
Top 6 : Chips 211
```

Top 7 : Steak Bowl 211
 Top 8 : Bottled Water 162
 Top 9 : Chicken Soft Tacos 115
 Top 10 : Chips and Fresh Tomato Salsa 110

```
1 i=0
2 for letter in ['A','B','C','D']:
3     print('{}번 : {}'.format(i,letter))
4     i+=1
```

0번 : A
 1번 : B
 2번 : C
 3번 : D

```
1 for letter in enumerate(['A','B','C','D'],start=1):
2     print(letter)
```

(1, 'A')
 (2, 'B')
 (3, 'C')
 (4, 'D')

메뉴별 판매량은 얼마나 될까?

```
1 chipo.groupby('item_name')['order_id'].count()
```

item_name	
6 Pack Soft Drink	54
Barbacoa Bowl	66
Barbacoa Burrito	91
Barbacoa Crispy Tacos	11
Barbacoa Salad Bowl	10
Barbacoa Soft Tacos	25
Bottled Water	162
Bowl	2
Burrito	6
Canned Soda	104
Canned Soft Drink	301
Carnitas Bowl	68
Carnitas Burrito	59
Carnitas Crispy Tacos	7
Carnitas Salad	1
Carnitas Salad Bowl	6
Carnitas Soft Tacos	40
Chicken Bowl	726
Chicken Burrito	553
Chicken Crispy Tacos	47
Chicken Salad	9
Chicken Salad Bowl	110
Chicken Soft Tacos	115
Chips	211
Chips and Fresh Tomato Salsa	110
Chips and Guacamole	479
Chips and Mild Fresh Tomato Salsa	1

Chips and Roasted Chili Corn Salsa	22
Chips and Roasted Chili-Corn Salsa	18
Chips and Tomatillo Green Chili Salsa	43
Chips and Tomatillo Red Chili Salsa	48
Chips and Tomatillo-Green Chili Salsa	31
Chips and Tomatillo-Red Chili Salsa	20
Crispy Tacos	2
Izze	20
Nantucket Nectar	27
Salad	2
Side of Chips	101
Steak Bowl	211
Steak Burrito	368
Steak Crispy Tacos	35
Steak Salad	4
Steak Salad Bowl	29
Steak Soft Tacos	55
Veggie Bowl	85
Veggie Burrito	95
Veggie Crispy Tacos	1
Veggie Salad	6
Veggie Salad Bowl	18
Veggie Soft Tacos	7
Name: order_id, dtype: int64	

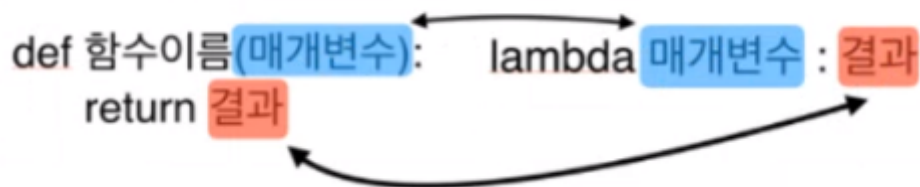
```
1 chipo.groupby('item_name')['quantity'].sum()
```

item_name	
6 Pack Soft Drink	55
Barbacoa Bowl	66
Barbacoa Burrito	91
Barbacoa Crispy Tacos	12
Barbacoa Salad Bowl	10
Barbacoa Soft Tacos	25
Bottled Water	211
Bowl	4
Burrito	6
Canned Soda	126
Canned Soft Drink	351
Carnitas Bowl	71
Carnitas Burrito	60
Carnitas Crispy Tacos	8
Carnitas Salad	1
Carnitas Salad Bowl	6
Carnitas Soft Tacos	40
Chicken Bowl	761
Chicken Burrito	591
Chicken Crispy Tacos	50
Chicken Salad	9
Chicken Salad Bowl	123
Chicken Soft Tacos	120
Chips	230
Chips and Fresh Tomato Salsa	130
Chips and Guacamole	506
Chips and Mild Fresh Tomato Salsa	1
Chips and Roasted Chili Corn Salsa	23
Chips and Roasted Chili-Corn Salsa	18
Chips and Tomatillo Green Chili Salsa	45
Chips and Tomatillo Red Chili Salsa	50
Chips and Tomatillo-Green Chili Salsa	33

Chips and Tomatillo-Red Chili Salsa	25
Crispy Tacos	2
Izze	20
Nantucket Nectar	29
Salad	2
Side of Chips	110
Steak Bowl	221
Steak Burrito	386
Steak Crispy Tacos	36
Steak Salad	4
Steak Salad Bowl	31
Steak Soft Tacos	56
Veggie Bowl	87
Veggie Burrito	97
Veggie Crispy Tacos	1
Veggie Salad	6
Veggie Salad Bowl	18
Veggie Soft Tacos	8

Name: quantity, dtype: int64

lambda 인자리스트: 표현식



```
1 chipo['item_price']=chipo['item_price'].apply(lambda x: float(x[1:]))
2 chipo['item_price']
```

0	2.39
1	3.39
2	3.39
3	2.39
4	16.98
...	
4617	11.75
4618	11.75
4619	11.25
4620	8.75
4621	8.75

Name: item_price, Length: 4622, dtype: float64

[퀴즈1] 주문량 평균 계산 금액 출력하기

```
1 chipo.groupby('order_id')['item_price'].sum().mean()

18.81142857142869
```

[퀴즈2] 한 주문당 20달러 이상 지불한 주문 번호(id)출력하기

```
1 order_price = chipo.groupby('order_id').sum()
```

```
1 order_price = omtpo.groupby('order_id').sum()
```

```
2 order_price
```

	quantity	item_price
order_id		
1	4	11.56
2	2	16.98
3	2	12.67
4	2	21.00
5	2	13.70
...
1830	2	23.00
1831	3	12.90
1832	2	13.20
1833	2	23.50
1834	3	28.75

1834 rows × 2 columns

```
1 order_price[order_price.item_price>=20]
```

	quantity	item_price
order_id		
4	2	21.00
18	4	24.90
20	4	36.20
21	3	23.36
26	2	20.50
...
1827	5	32.95
1829	3	24.25
1830	2	23.00
1833	2	23.50
1834	3	28.75

589 rows × 2 columns

[퀴즈3] 'Veggie Salad Bowl'이 몇 번 주문되었는지 구하기

```
1 chipo_salad = chipo[chipo['item_name']=='Veggie Salad Bowl']
2 len(chipo_salad)
```

18

```
1 order_count = chipo.groupby('order_id').sum()
2 order_count
```

	quantity	item_price
order_id		
1	4	11.56
2	2	16.98
3	2	12.67
4	2	21.00
5	2	13.70
...
1830	2	23.00
1831	3	12.90
1832	2	13.20
1833	2	23.50
1834	3	28.75

1834 rows × 2 columns

```
1 result = order_count[order_count.quantity >= 2]
```

```
1 result.count()
```

```
↳ quantity    1778
   item_price  1778
   dtype: int64
```

✓ 0초 오전 10:02에 완료됨

