In [1]: import pandas as pd In [2]: df=pd.read_csv(r"C:\Users\TharunMahendra\OneDrive\Desktop\NIT\SQL\database extraction\SQLite ExtractedFile.csv") In [3]: **df** Out[3]: destination passanger weather temperature time coupon expiration gender age maritalStatus ... CarryAway RestaurantLessThan20 Restaurant20To50 to o No Urgent Unmarried Alone Sunny 55 2PM Restaurant(<20) 1d Female 21 NaN 4~8 1~3 Place partner No Urgent Unmarried Friend(s) 80 10AM Coffee House 2h Female NaN 1~3 Sunny partner No Urgent Carry out & 2 Friend(s) Sunny 80 10AM 2h Female 21 NaN 4~8 1~3 Place Take away partner Unmarried No Urgent Friend(s) 80 2PM Coffee House 2h Female 21 NaN 4~8 1~3 Sunny partner No Urgent Unmarried 4 Coffee House NaN 4~8 1~3 Friend(s) Sunny 80 2PM 1d Female 21 partner ... Carry out & 12679 Partner 55 6PM Male 26 Single ... 1~3 4~8 1~3 Take away Carry out & 12680 Work 55 7AM Single ... 1~3 4~8 1~3 Alone Rainy Take away 12681 Work Alone 30 7AM Coffee House 1d Male 26 Single ... 1~3 4~8 1~3 Bar 1~3 12682 Work Alone 30 1d Male 26 Single ... 1~3 4~8 Snowy 7AM Restaurant(20-12683 Work Alone Sunny 80 7AM 2h Male 26 Single ... 1~3 4~8 1~3

12684 rows × 27 columns

4

In [4]: df[['weather','temperature']]

> Sunny 12684 rows × 2 columns

Snowy

30

80

In [5]: df[0:10]

12682

12683

:		destination	passanger	weather	temperature	time	coupon	expiration	gender	age	maritalStatus	 CarryAway	Restaurant Less Than 20	Restaurant20To50	toCou
()	No Urgent Place	Alone	Sunny	55	2PM	Restaurant(<20)	1d	Female	21	Unmarried partner	 NaN	4~8	1~3	
	1	No Urgent Place	Friend(s)	Sunny	80	10AM	Coffee House	2h	Female	21	Unmarried partner	 NaN	4~8	1~3	
2	2	No Urgent Place	Friend(s)	Sunny	80	10AM	Carry out & Take away	2h	Female	21	Unmarried partner	 NaN	4~8	1~3	
3	3	No Urgent Place	Friend(s)	Sunny	80	2PM	Coffee House	2h	Female	21	Unmarried partner	 NaN	4~8	1~3	
4	4	No Urgent Place	Friend(s)	Sunny	80	2PM	Coffee House	1d	Female	21	Unmarried partner	 NaN	4~8	1~3	
	5	No Urgent Place	Friend(s)	Sunny	80	6PM	Restaurant(<20)	2h	Female	21	Unmarried partner	 NaN	4~8	1~3	
(6	No Urgent Place	Friend(s)	Sunny	55	2PM	Carry out & Take away	1d	Female	21	Unmarried partner	 NaN	4~8	1~3	
	7	No Urgent Place	Kid(s)	Sunny	80	10AM	Restaurant(<20)	2h	Female	21	Unmarried partner	 NaN	4~8	1~3	
8	3	No Urgent Place	Kid(s)	Sunny	80	10AM	Carry out & Take away	2h	Female	21	Unmarried partner	 NaN	4~8	1~3	
9	9	No Urgent Place	Kid(s)	Sunny	80	10AM	Bar	1d	Female	21	Unmarried partner	 NaN	4~8	1~3	
1	0 rc	ows × 27 col	umns												
	4														

In [6]: df.passanger.unique()

Out[6]: array(['Alone', 'Friend(s)', 'Kid(s)', 'Partner'], dtype=object)

In [7]: df[df['destination']=='Home']

Uses											. , ,		Restaurant20To50
Home	Alone	Sunny	55	6PM	Bar	1d	Female	21	Unmarried partner		NaN	4~8	1~3
Home	Alone	Sunny	55	6PM	Restaurant(20- 50)	1d	Female	21	Unmarried partner		NaN	4~8	1~3
Home	Alone	Sunny	80	6PM	Coffee House	2h	Female	21	Unmarried partner		NaN	4~8	1~3
Home	Alone	Sunny	55	6PM	Bar	1d	Male	21	Single		4~8	4~8	less1
Home	Alone	Sunny	55	6PM	Restaurant(20- 50)	1d	Male	21	Single		4~8	4~8	less1
•••					***								
Home	Alone	Snowy	30	10PM	Coffee House	2h	Male	26	Single		1~3	4~8	1~3
Home	Alone	Sunny	80	6PM	Restaurant(20- 50)	1d	Male	26	Single		1~3	4~8	1~3
Home	Partner	Sunny	30	6PM	Restaurant(<20)	1d	Male	26	Single		1~3	4~8	1~3
Home	Partner	Sunny	30	10PM	Restaurant(<20)	2h	Male	26	Single		1~3	4~8	1~3
Home	Partner	Rainy	55	6PM	Carry out & Take away	1d	Male	26	Single		1~3	4~8	1~3
	Home Home Home Home Home Home	Home Alone Home Alone Home Alone Home Alone Home Partner Home Partner	Home Alone Sunny Home Alone Sunny Home Alone Sunny Home Alone Snowy Home Alone Sunny Home Partner Sunny Home Partner Sunny Home Partner Rainy	Home Alone Sunny 80 Home Alone Sunny 55 Home Alone Sunny 55 Home Alone Snowy 30 Home Alone Sunny 80 Home Partner Sunny 30 Home Partner Sunny 30 Home Partner Sunny 30 Home Partner Rainy 55	Home Alone Sunny 80 6PM Home Alone Sunny 55 6PM Home Alone Sunny 55 6PM Home Alone Snowy 30 10PM Home Alone Sunny 80 6PM Home Partner Sunny 30 6PM Home Partner Sunny 30 10PM Home Partner Sunny 30 10PM Home Partner Rainy 55 6PM	Home	Home	Home Alone Sunny 80 6PM Coffee House 2h Female Home Alone Sunny 55 6PM Bar 1d Male Home Alone Sunny 55 6PM Restaurant(20-50) 1d Male Home Alone Snowy 30 10PM Coffee House 2h Male Home Alone Sunny 80 6PM Restaurant(20-50) 1d Male Home Partner Sunny 30 6PM Restaurant(<20)	Home Alone Sunny 80 6PM Coffee House 2h Female 21 Home Alone Sunny 55 6PM Bar 1d Male 21 Home Alone Sunny 55 6PM Restaurant(20-50) 1d Male 21	Home	Home Alone Sunny 80 6PM Coffee House 2h Female 21 Unmarried partner	Home Alone Sunny 80 6PM Coffee House 2h Female 21 Unmarried partner NaN	Home Alone Sunny 80 6PM Coffee House 2h Female 21 Unmarried partner NaN 4~8 Home Alone Sunny 55 6PM Bar 1d Male 21 Single 4~8 4~8 Home Alone Sunny 55 6PM Restaurant(20-50) 1d Male 21 Single 4~8 4~8

In [8]: df.sort_values('coupon')

Out[8]:		destination	passanger	weather	temperature	time	coupon	expiration	gender	age	maritalStatus	 CarryAway	RestaurantLessThan20	Restaurant20To50
	11702	Home	Partner	Sunny	30	10PM	Bar	2h	Female	50plus	Married partner	 4~8	1~3	less1
	9930	No Urgent Place	Alone	Snowy	30	2PM	Bar	1d	Female	21	Single	 gt8	gt8	4~8
	10632	Home	Alone	Rainy	55	6PM	Bar	1d	Male	21	Single	 gt8	less1	less1
	7997	No Urgent Place	Friend(s)	Rainy	55	10PM	Bar	2h	Male	26	Unmarried partner	 4~8	never	1~3
	11166	Work	Alone	Snowy	30	7AM	Bar	1d	Female	41	Married partner	 gt8	1~3	less1
										•••		 		
	10476	Home	Alone	Sunny	80	6PM	Restaurant(<20)	1d	Female	31	Unmarried partner	 1~3	1~3	less1
	5447	Home	Alone	Sunny	80	10PM	Restaurant(<20)	2h	Female	50plus	Single	 less1	less1	never
	10478	Home	Alone	Snowy	30	10PM	Restaurant(<20)	2h	Female	31	Unmarried partner	 1~3	1~3	less1
	5440	No Urgent Place	Alone	Sunny	80	2PM	Restaurant(<20)	2h	Female	50plus	Single	 less1	less1	never
	0	No Urgent Place	Alone	Sunny	55	2PM	Restaurant(<20)	1d	Female	21	Unmarried partner	 NaN	4~8	1~3
	12684 rc	ws × 27 colu	mns											
	4													•
In [9]:	df.rena	ame(columns=	{'destinat	ion':'Des	tination'},ir	place=	True)							
In [10]:	df.head	1()												

Out[10]:		Destination	passanger	weather	temperature	time	coupon	expiration	gender	age	maritalStatus	 CarryAway	RestaurantLessThan20	Restaurant20To50	toCou
	0	No Urgent Place	Alone	Sunny	55	2PM	Restaurant(<20)	1d	Female	21	Unmarried partner	 NaN	4~8	1~3	
	1	No Urgent Place	Friend(s)	Sunny	80	10AM	Coffee House	2h	Female	21	Unmarried partner	 NaN	4~8	1~3	
	2	No Urgent Place	Friend(s)	Sunny	80	10AM	Carry out & Take away	2h	Female	21	Unmarried partner	 NaN	4~8	1~3	
	3	No Urgent Place	Friend(s)	Sunny	80	2PM	Coffee House	2h	Female	21	Unmarried partner	 NaN	4~8	1~3	
	4	No Urgent Place	Friend(s)	Sunny	80	2PM	Coffee House	1d	Female	21	Unmarried partner	 NaN	4~8	1~3	
	5 rc	ws × 27 colu	mns												
	4														▶
In [11]:	df.	groupby('oc	cupation')	.size().t	o_frame('cour	nt').re	set_index()								

```
0
                          Architecture & Engineering
                                                  175
        1 Arts Design Entertainment Sports & Media
                                                   629
          2 Building & Grounds Cleaning & Maintenance
                                                   44
         3
                               Business & Financial 544
          4
                        Community & Social Services 241
                      Computer & Mathematical 1408
          5
          6
                           Construction & Extraction 154
          7
                         Education&Training&Library 943
                          Farming Fishing & Forestry
                                                   43
          8
                   Food Preparation & Serving Related 298
         10
                   Healthcare Practitioners & Technical
                                                  244
         11
                                Healthcare Support 242
         12
                     Installation Maintenance & Repair
                                                  133
                      Legal 219
         13
         14
                          Life Physical Social Science
                                                  170
                                    Management 838
         15
         16
                       Office & Administrative Support
         17
                           Personal Care & Service 175
         18
                             Production Occupations
                                                  110
                                  Protective Service 175
         19
         20
                                         Retired 495
         21
                                   Sales & Related 1093
         22
                                         Student 1584
         23
                     Transportation & Material Moving 218
         24
                                     Unemployed 1870
In [12]: df.groupby('weather')['temperature'].mean().to_frame('avg_temp').reset_index()
Out[12]: weather avg_temp
             Rainy 55.000000
        1 Snowy 30.000000
         2 Sunny 68.946271
In [13]: df.groupby('weather').size().to_frame('count_temp').reset_index()
Out[13]:
          weather count_temp
        0
            Rainy
                          1210
        1 Snowy
                         1405
                         10069
             Sunny
In [14]: df.groupby('weather')['temperature'].nunique().to_frame('distinct_temp').reset_index()
Out[14]: weather distinct_temp
             Rainy
        1 Snowy 1
         2 Sunny
In [15]: df.groupby('weather')['temperature'].sum().to_frame('sum_temp').reset_index()
Out[15]: weather sum_temp
         1 Snowy
                     42150
                      694220
             Sunny
In [16]: df.groupby('weather')['temperature'].min().to_frame('min_temp').reset_index()
Out[16]: weather min_temp
              Rainy
        1 Snowy
                         30
```

Out[11]:

occupation count

```
In [17]: df.groupby('weather')['temperature'].max().to_frame('max_temp').reset_index()
Out[17]: weather max_temp
            Rainy
                         55
        1 Snowy 30
        2 Sunny
                         80
In [18]: df.groupby('occupation').filter(lambda x:x['occupation'].iloc[0]=='Student').groupby('occupation').size()
Out[18]: occupation
Student 1584
dtype: int64
In [19]: df[df['passanger'] == 'Alone'][['Destination', 'passanger']]
Out[19]: Destination passanger
           0 No Urgent Place
                                Alone
        13 Home Alone
           14
                      Home
                                Alone
        15
                    Home
                               Alone
           16
                      Work
                               Alone
         12676
                      Home
                                Alone
        12680
                      Work
                               Alone
         12681
                       Work
                                Alone
         12682
                       Work
                              Alone
         12683
                       Work
                                Alone
        7305 rows × 2 columns
In [20]: df[df['weather'].str.startswith('Sun')]
```

ut[20]:		Destination	passanger	weather	temperature	time	coupon	expiration	gender	age	maritalStatus	c	arryAway	RestaurantLessThan20	Restaurant20To50
	0	No Urgent Place	Alone	Sunny	55	2PM	Restaurant(<20)	1d	Female	21	Unmarried partner		NaN	4~8	1~3
	1	No Urgent Place	Friend(s)	Sunny	80	10AM	Coffee House	2h	Female	21	Unmarried partner		NaN	4~8	1~3
	2	No Urgent Place	Friend(s)	Sunny	80	10AM	Carry out & Take away	2h	Female	21	Unmarried partner		NaN	4~8	1~3
	3	No Urgent Place	Friend(s)	Sunny	80	2PM	Coffee House	2h	Female	21	Unmarried partner		NaN	4~8	1~3
	4	No Urgent Place	Friend(s)	Sunny	80	2PM	Coffee House	1d	Female	21	Unmarried partner		NaN	4~8	1~3
	12673	Home	Alone	Sunny	30	6PM	Carry out & Take away	1d	Male	26	Single		1~3	4~8	1~3
	12676	Home	Alone	Sunny	80	6PM	Restaurant(20- 50)	1d	Male	26	Single		1~3	4~8	1~3
	12677	Home	Partner	Sunny	30	6PM	Restaurant(<20)	1d	Male	26	Single		1~3	4~8	1~3
	12678	Home	Partner	Sunny	30	10PM	Restaurant(<20)	2h	Male	26	Single		1~3	4~8	1~3
	12683	Work	Alone	Sunny	80	7AM	Restaurant(20- 50)	2h	Male	26	Single		1~3	4~8	1~3
	10069 rc	ws × 27 colu	mns												
	4		_	_				_	_	-					
[21]:	df[(df['temperatur	e'] >= 29)	& (df['t	emperature']	<= 75)]['temperature'].unique()							
:[21]:	annav([55, 30], dt	-vno-int64)												

In [22]: df[df['occupation'].isin(['Sales & Related', 'Management'])][['occupation']]











































