```
[3]: import pandas as pd
     import sqlite3
```

```
[6]: conn = sqlite3.connect('C:/Users/Amrutha Thalla/Downloads/SQLite Test.db')
     conn
```

```
[6]: <sqlite3.Connection at 0x259f6a88400>
```

```
[7]: query = "SELECT * FROM dataset_1"
     df = pd.read_sql(query, conn)
```

```
[10]: conn.close()
```

```
[13]: print(df.head())
```

```
          destination  passanger  weather  temperature   time  \
0   No Urgent Place        Alone    Sunny           55    2PM
1   No Urgent Place   Friend(s)    Sunny           80   10AM
2   No Urgent Place   Friend(s)    Sunny           80   10AM
3   No Urgent Place   Friend(s)    Sunny           80    2PM
4   No Urgent Place   Friend(s)    Sunny           80    2PM

                 coupon  expiration  gender  age      maritalStatus  ...  \
0       Restaurant(<20)          1d  Female   21  Unmarried partner  ...
1          Coffee House          2h  Female   21  Unmarried partner  ...
2   Carry out & Take away         2h  Female   21  Unmarried partner  ...
3          Coffee House          2h  Female   21  Unmarried partner  ...
4          Coffee House          1d  Female   21  Unmarried partner  ...

   CarryAway  RestaurantLessThan20  Restaurant20To50  toCoupon_GEQ5min  \
0                            4~8               1~3                 1
1                            4~8               1~3                 1
2                            4~8               1~3                 1
3                            4~8               1~3                 1
4                            4~8               1~3                 1

   toCoupon_GEQ15min  toCoupon_GEQ25min  direction_same  direction_opp  Y  \
0                  0                  0               0              1  1
1                  0                  0               0              1  0
2                  1                  0               0              1  1
3                  1                  0               0              1  0
4                  1                  0               0              1  0

   row_count
0          1
1          2
2          3
3          4
4          5

[5 rows x 27 columns]
```

```python
df.head(10)
```

	destination	passanger	weather	temperature	time	coupon	expiration	gender	age	maritalStatus	...	CarryAway	RestaurantLessThan20	Restaurant20T
0	No Urgent Place	Alone	Sunny	55	2PM	Restaurant(<20)	1d	Female	21	Unmarried partner	...		4~8	1
1	No Urgent Place	Friend(s)	Sunny	80	10AM	Coffee House	2h	Female	21	Unmarried partner	...		4~8	1
2	No Urgent Place	Friend(s)	Sunny	80	10AM	Carry out & Take away	2h	Female	21	Unmarried partner	...		4~8	1
3	No Urgent Place	Friend(s)	Sunny	80	2PM	Coffee House	2h	Female	21	Unmarried partner	...		4~8	1
4	No Urgent Place	Friend(s)	Sunny	80	2PM	Coffee House	1d	Female	21	Unmarried partner	...		4~8	1
5	No Urgent Place	Friend(s)	Sunny	80	6PM	Restaurant(<20)	2h	Female	21	Unmarried partner	...		4~8	1
6	No Urgent Place	Friend(s)	Sunny	55	2PM	Carry out & Take away	1d	Female	21	Unmarried partner	...		4~8	1
7	No Urgent Place	Kid(s)	Sunny	80	10AM	Restaurant(<20)	2h	Female	21	Unmarried partner	...		4~8	1
8	No Urgent Place	Kid(s)	Sunny	80	10AM	Carry out & Take away	2h	Female	21	Unmarried partner	...		4~8	1
9	No Urgent Place	Kid(s)	Sunny	80	10AM	Bar	1d	Female	21	Unmarried partner	...		4~8	1

```python
df['passanger'].unique()
```

```
array(['Alone', 'Friend(s)', 'Kid(s)', 'Partner'], dtype=object)
```

```python
df[df['destination']=='Home']
```

	destination	passanger	weather	temperature	time	coupon	expiration	gender	age	maritalStatus	...	CarryAway	RestaurantLessThan20	Restaurant
13	Home	Alone	Sunny	55	6PM	Bar	1d	Female	21	Unmarried partner	...		4~8	
14	Home	Alone	Sunny	55	6PM	Restaurant(20-50)	1d	Female	21	Unmarried partner	...		4~8	
15	Home	Alone	Sunny	80	6PM	Coffee House	2h	Female	21	Unmarried partner	...		4~8	
35	Home	Alone	Sunny	55	6PM	Bar	1d	Male	21	Single	...	4~8	4~8	
36	Home	Alone	Sunny	55	6PM	Restaurant(20-50)	1d	Male	21	Single	...	4~8	4~8	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
12675	Home	Alone	Snowy	30	10PM	Coffee House	2h	Male	26	Single	...	1~3	4~8	
12676	Home	Alone	Sunny	80	6PM	Restaurant(20-50)	1d	Male	26	Single	...	1~3	4~8	
12677	Home	Partner	Sunny	30	6PM	Restaurant(<20)	1d	Male	26	Single	...	1~3	4~8	
12678	Home	Partner	Sunny	30	10PM	Restaurant(<20)	2h	Male	26	Single	...	1~3	4~8	
12679	Home	Partner	Rainy	55	6PM	Carry out & Take away	1d	Male	26	Single	...	1~3	4~8	

3237 rows × 27 columns

```
df.sort_values('coupon')
```

	destination	passanger	weather	temperature	time	coupon	expiration	gender	age	maritalStatus	...	CarryAway	RestaurantLessThan20	Restau
11702	Home	Partner	Sunny	30	10PM	Bar	2h	Female	50plus	Married partner	...	4~8	1~3	
9930	No Urgent Place	Alone	Snowy	30	2PM	Bar	1d	Female	21	Single	...	gt8	gt8	
10632	Home	Alone	Rainy	55	6PM	Bar	1d	Male	21	Single	...	gt8	less1	
7997	No Urgent Place	Friend(s)	Rainy	55	10PM	Bar	2h	Male	26	Unmarried partner	...	4~8	never	
11166	Work	Alone	Snowy	30	7AM	Bar	1d	Female	41	Married partner	...	gt8	1~3	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
10476	Home	Alone	Sunny	80	6PM	Restaurant(<20)	1d	Female	31	Unmarried partner	...	1~3	1~3	
5447	Home	Alone	Sunny	80	10PM	Restaurant(<20)	2h	Female	50plus	Single	...	less1	less1	
10478	Home	Alone	Snowy	30	10PM	Restaurant(<20)	2h	Female	31	Unmarried partner	...	1~3	1~3	
5440	No Urgent Place	Alone	Sunny	80	2PM	Restaurant(<20)	2h	Female	50plus	Single	...	less1	less1	
0	No Urgent Place	Alone	Sunny	55	2PM	Restaurant(<20)	1d	Female	21	Unmarried partner	...		4~8	

```
[20]: df.rename(columns={'destination':'Destination'},inplace=True)

[22]: df.columns

[22]: Index(['Destination', 'passanger', 'weather', 'temperature', 'time', 'coupon',
             'expiration', 'gender', 'age', 'maritalStatus', 'has_children',
             'education', 'occupation', 'income', 'car', 'Bar', 'CoffeeHouse',
             'CarryAway', 'RestaurantLessThan20', 'Restaurant20To50',
             'toCoupon_GEQ5min', 'toCoupon_GEQ15min', 'toCoupon_GEQ25min',
             'direction_same', 'direction_opp', 'Y', 'row_count'],
            dtype='object')
```

```
df.groupby('occupation').size().to_frame('Count').reset_index()
```

	occupation	Count
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
5	Computer & Mathematical	1408
6	Construction & Extraction	154
7	Education&Training&Library	943
8	Farming Fishing & Forestry	43
9	Food Preparation & Serving Related	298
10	Healthcare Practitioners & Technical	244
11	Healthcare Support	242
12	Installation Maintenance & Repair	133
13	Legal	219
14	Life Physical Social Science	170
15	Management	838
16	Office & Administrative Support	639
17	Personal Care & Service	175
18	Production Occupations	110
19	Protective Service	175
20	Retired	495
21	Sales & Related	1093
22	Student	1584
23	Transportation & Material Moving	218
24	Unemployed	1870

```python
[24]: df.groupby('weather')['temperature'].mean().to_frame('avg_temp').reset_index()
```

[24]:

	weather	avg_temp
0	Rainy	55.000000
1	Snowy	30.000000
2	Sunny	68.946271

```python
df.groupby('weather')['temperature'].size().to_frame('Count_temp').reset_index()
```

	weather	Count_temp
0	Rainy	1210
1	Snowy	1405
2	Sunny	10069

```python
df.groupby('weather')['temperature'].nunique().to_frame('count_distinct_temp').reset_index()
```

	weather	count_distinct_temp
0	Rainy	1
1	Snowy	1
2	Sunny	3

```python
df.groupby('weather')['temperature'].sum().to_frame('sum_temp').reset_index()
```

	weather	sum_temp
0	Rainy	66550
1	Snowy	42150
2	Sunny	694220

```python
df.groupby('weather')['temperature'].min().to_frame('min_temp').reset_index()
```

	weather	min_temp
0	Rainy	55
1	Snowy	30
2	Sunny	30

```python
df.groupby('weather')['temperature'].max().to_frame('max_temp').reset_index()
```

	weather	max_temp
0	Rainy	55
1	Snowy	30
2	Sunny	80

```python
df.groupby('occupation').filter(lambda x: x['occupation'].iloc[0] ==
'Student').groupby('occupation').size()
```

```
occupation
Student    1584
dtype: int64
```

```
df[df['passanger'] == 'Alone'][['Destination', 'passanger']]
```

	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

```
df[df['weather'].str.startswith('Sun')]
```

	Destination	passanger	weather	temperature	time	coupon	expiration	gender	age	maritalStatus	...	CarryAway	RestaurantLessThan20	Restaura
0	No Urgent Place	Alone	Sunny	55	2PM	Restaurant(<20)	1d	Female	21	Unmarried partner	...		4~8	
1	No Urgent Place	Friend(s)	Sunny	80	10AM	Coffee House	2h	Female	21	Unmarried partner	...		4~8	
2	No Urgent Place	Friend(s)	Sunny	80	10AM	Carry out & Take away	2h	Female	21	Unmarried partner	...		4~8	
3	No Urgent Place	Friend(s)	Sunny	80	2PM	Coffee House	2h	Female	21	Unmarried partner	...		4~8	
4	No Urgent Place	Friend(s)	Sunny	80	2PM	Coffee House	1d	Female	21	Unmarried partner	...		4~8	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
12673	Home	Alone	Sunny	30	6PM	Carry out & Take away	1d	Male	26	Single	...	1~3	4~8	
12676	Home	Alone	Sunny	80	6PM	Restaurant(20-50)	1d	Male	26	Single	...	1~3	4~8	
12677	Home	Partner	Sunny	30	6PM	Restaurant(<20)	1d	Male	26	Single	...	1~3	4~8	
12678	Home	Partner	Sunny	30	10PM	Restaurant(<20)	2h	Male	26	Single	...	1~3	4~8	
12683	Work	Alone	Sunny	80	7AM	Restaurant(20-50)	2h	Male	26	Single	...	1~3	4~8	

10069 rows × 27 columns

```
[38]: df[(df['temperature'] >= 29) & (df['temperature'] <= 75)]['temperature'].unique()

[38]: array([55, 30])

[ ]:
```

```
df[df['occupation'].isin(['Sales & Related', 'Management'])][['occupation']]
```

	occupation
193	Sales & Related
194	Sales & Related
195	Sales & Related
196	Sales & Related
197	Sales & Related
...	...
12679	Sales & Related
12680	Sales & Related
12681	Sales & Related
12682	Sales & Related
12683	Sales & Related

1931 rows × 1 columns