In [34]:
```python
import pandas as pd
```

In [35]:
```python
sql_dataset = pd.read_csv(r'D:\Data Science and Machine Learning\March\7th_SQL_For_Data_analysts\dataset_1_202503072201.csv'
```

In [38]:
```python
sql_dataset.head()
```

Out[38]:

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

5 rows × 27 columns

In [39]:
```python
sql_dataset.iloc[:,1:3]
```

Out[39]:

	passanger	weather
0	Alone	Sunny
1	Friend(s)	Sunny
2	Friend(s)	Sunny
3	Friend(s)	Sunny
4	Friend(s)	Sunny
...	...	...
12679	Partner	Rainy
12680	Alone	Rainy
12681	Alone	Snowy
12682	Alone	Snowy
12683	Alone	Sunny

12684 rows × 2 columns

In [40]:
```python
sql_dataset.loc[:,'passanger':'weather']
```

Out[40]:

	passanger	weather
0	Alone	Sunny
1	Friend(s)	Sunny
2	Friend(s)	Sunny
3	Friend(s)	Sunny
4	Friend(s)	Sunny
...	...	...
12679	Partner	Rainy
12680	Alone	Rainy
12681	Alone	Snowy
12682	Alone	Snowy
12683	Alone	Sunny

12684 rows × 2 columns

In [41]: `sql_dataset[['passanger','weather']]`

Out[41]:

	passanger	weather
0	Alone	Sunny
1	Friend(s)	Sunny
2	Friend(s)	Sunny
3	Friend(s)	Sunny
4	Friend(s)	Sunny
...	...	...
12679	Partner	Rainy
12680	Alone	Rainy
12681	Alone	Snowy
12682	Alone	Snowy
12683	Alone	Sunny

12684 rows × 2 columns

In [42]: `sql_dataset.head(8)`

Out[42]:

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

8 rows × 27 columns

In [43]: `sql_dataset[:8]`

Out[43]:

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

8 rows × 27 columns

In [44]: `sql_dataset['passanger'].unique()`

Out[44]: `array(['Alone', 'Friend(s)', 'Kid(s)', 'Partner'], dtype=object)`

In [45]: `sql_dataset[sql_dataset['destination'] == 'Home']`

Out[45]:

	destination	passanger	weather	temperature	time	coupon	expiration	gender	age	maritalStatus	...	CarryAway	RestaurantLessThan20
13	Home	Alone	Sunny	55	6PM	Bar	1d	Female	21	Unmarried partner	...	NaN	4~8
14	Home	Alone	Sunny	55	6PM	Restaurant(20-50)	1d	Female	21	Unmarried partner	...	NaN	4~8
15	Home	Alone	Sunny	80	6PM	Coffee House	2h	Female	21	Unmarried partner	...	NaN	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

In [46]: `sql_dataset.query("destination == 'Home'")`

Out[46]:

	destination	passanger	weather	temperature	time	coupon	expiration	gender	age	maritalStatus	...	CarryAway	RestaurantLessThan20
13	Home	Alone	Sunny	55	6PM	Bar	1d	Female	21	Unmarried partner	...	NaN	4~8
14	Home	Alone	Sunny	55	6PM	Restaurant(20-50)	1d	Female	21	Unmarried partner	...	NaN	4~8
15	Home	Alone	Sunny	80	6PM	Coffee House	2h	Female	21	Unmarried partner	...	NaN	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

In [47]: `sql_dataset.loc[sql_dataset['destination'] == 'Home']`

Out[47]:

	destination	passanger	weather	temperature	time	coupon	expiration	gender	age	maritalStatus	...	CarryAway	RestaurantLessThan20
13	Home	Alone	Sunny	55	6PM	Bar	1d	Female	21	Unmarried partner	...	NaN	4~8
14	Home	Alone	Sunny	55	6PM	Restaurant(20-50)	1d	Female	21	Unmarried partner	...	NaN	4~8
15	Home	Alone	Sunny	80	6PM	Coffee House	2h	Female	21	Unmarried partner	...	NaN	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

In [48]: `sql_dataset.sort_values(by='coupon')`

Out[48]:

	destination	passanger	weather	temperature	time	coupon	expiration	gender	age	maritalStatus	...	CarryAway	RestaurantLessThan2
11702	Home	Partner	Sunny	30	10PM	Bar	2h	Female	50plus	Married partner	...	4~8	1~
9930	No Urgent Place	Alone	Snowy	30	2PM	Bar	1d	Female	21	Single	...	gt8	g
10632	Home	Alone	Rainy	55	6PM	Bar	1d	Male	21	Single	...	gt8	less
7997	No Urgent Place	Friend(s)	Rainy	55	10PM	Bar	2h	Male	26	Unmarried partner	...	4~8	nev
11166	Work	Alone	Snowy	30	7AM	Bar	1d	Female	41	Married partner	...	gt8	1~
...	...	...	...	...	...	...	...	...	...	...	...	...	
10476	Home	Alone	Sunny	80	6PM	Restaurant(<20)	1d	Female	31	Unmarried partner	...	1~3	1~
5447	Home	Alone	Sunny	80	10PM	Restaurant(<20)	2h	Female	50plus	Single	...	less1	less
10478	Home	Alone	Snowy	30	10PM	Restaurant(<20)	2h	Female	31	Unmarried partner	...	1~3	1~
5440	No Urgent Place	Alone	Sunny	80	2PM	Restaurant(<20)	2h	Female	50plus	Single	...	less1	less
0	No Urgent Place	Alone	Sunny	55	2PM	Restaurant(<20)	1d	Female	21	Unmarried partner	...	NaN	4~

12684 rows × 27 columns

In [49]: `sql_dataset[['passanger']].rename(columns={'passanger': 'Passenger'})`

Out[49]:

	Passenger
0	Alone
1	Friend(s)
2	Friend(s)
3	Friend(s)
4	Friend(s)
...	...
12679	Partner
12680	Alone
12681	Alone
12682	Alone
12683	Alone

12684 rows × 1 columns

In [50]: `sql_dataset.loc[:, ['passanger']].rename(columns={'passanger': 'Passenger'})`

Out[50]:

	Passenger
0	Alone
1	Friend(s)
2	Friend(s)
3	Friend(s)
4	Friend(s)
...	...
12679	Partner
12680	Alone
12681	Alone
12682	Alone
12683	Alone

12684 rows × 1 columns

In [51]: `sql_dataset.groupby('occupation').size()`

Out[51]:
```
occupation
Architecture & Engineering               175
Arts Design Entertainment Sports & Media 629
Building & Grounds Cleaning & Maintenance  44
Business & Financial                     544
Community & Social Services              241
Computer & Mathematical                 1408
Construction & Extraction                154
Education&Training&Library               943
Farming Fishing & Forestry                43
Food Preparation & Serving Related       298
Healthcare Practitioners & Technical     244
Healthcare Support                       242
Installation Maintenance & Repair        133
Legal                                    219
Life Physical Social Science             170
Management                               838
Office & Administrative Support          639
Personal Care & Service                  175
Production Occupations                   110
Protective Service                       175
Retired                                  495
Sales & Related                         1093
Student                                 1584
Transportation & Material Moving         218
Unemployed                              1870
dtype: int64
```

In [56]: `sql_dataset.groupby('weather')['temperature'].mean()`

Out[56]:
```
weather
Rainy     55.000000
Snowy     30.000000
Sunny     68.946271
Name: temperature, dtype: float64
```

In [55]: `sql_dataset.groupby('weather')['temperature'].mean().reset_index().rename(columns={'temperature': 'Average_Temperature'})`

Out[55]:

	weather	Average_Temperature
0	Rainy	55.000000
1	Snowy	30.000000
2	Sunny	68.946271

In [59]: `sql_dataset.groupby('weather')['temperature'].count()`

Out[59]:
```
weather
Rainy     1210
Snowy     1405
Sunny    10069
Name: temperature, dtype: int64
```

In [60]: `sql_dataset.groupby('weather')['temperature'].count().reset_index().rename(columns = {'temperature':'Count of Temparatures i`

Out[60]:

	weather	Count of Temparatures in Season
0	Rainy	1210
1	Snowy	1405
2	Sunny	10069

In [61]: `sql_dataset.groupby('weather')['temperature'].sum()`

Out[61]:
```
weather
Rainy     66550
Snowy     42150
Sunny    694220
Name: temperature, dtype: int64
```

In [62]: `sql_dataset.groupby('weather')['temperature'].sum().reset_index().rename(columns = {'temperature':'Count of Temparatures in`

Out[62]:

	weather	Count of Temparatures in Season
0	Rainy	66550
1	Snowy	42150
2	Sunny	694220

In [65]:
```python
sql_dataset.groupby('weather')['temperature'].min().reset_index().rename(columns = {'temperature':'Count of Temparatures in
```

Out[65]:

	weather	Count of Temparatures in Season
0	Rainy	55
1	Snowy	30
2	Sunny	30

In [66]:
```python
sql_dataset.groupby('weather')['temperature'].max().reset_index().rename(columns = {'temperature':'Count of Temparatures in
```

Out[66]:

	weather	Count of Temparatures in Season
0	Rainy	55
1	Snowy	30
2	Sunny	80

In [71]:
```python
sql_dataset['age'] = pd.to_numeric(sql_dataset['age'], errors='coerce')
sql_dataset.groupby('weather')['age'].mean().reset_index().rename(columns={'age': 'Average_age'})
```

Out[71]:

	weather	Average_age
0	Rainy	29.728025
1	Snowy	30.162186
2	Sunny	29.869465

In [77]:
```python
temp_df = sql_dataset.groupby('occupation')['age'].mean().reset_index().rename(columns={'age':'age'})
temp_df[temp_df['age'] > 36]
```

Out[77]:

	occupation	age
18	Production Occupations	37.250000
20	Retired	37.538462

In [78]:
```python
sql_dataset.groupby('occupation')['age'].mean().reset_index().query('age > 36').rename(columns={'age': 'Average_Age'})
```

Out[78]:

	occupation	Average_Age
18	Production Occupations	37.250000
20	Retired	37.538462

In [79]:
```python
union_dataset = pd.read_csv(r'D:\Data Science and Machine Learning\March\7th_SQL_For_Data_analysts\table_to_union_2025030817
```

In [84]:
```python
tables_union = pd.concat([sql_dataset,union_dataset], ignore_index=True).drop_duplicates()
tables_union
```

Out[84]:

	destination	passanger	weather	temperature	time	coupon	expiration	gender	age	maritalStatus	...	CarryAway	RestaurantLessThan20
0	No Urgent Place	Alone	Sunny	55	2PM	Restaurant(<20)	1d	Female	21.0	Unmarried partner	...	NaN	4~8
1	No Urgent Place	Friend(s)	Sunny	80	10AM	Coffee House	2h	Female	21.0	Unmarried partner	...	NaN	4~8
2	No Urgent Place	Friend(s)	Sunny	80	10AM	Carry out & Take away	2h	Female	21.0	Unmarried partner	...	NaN	4~8
3	No Urgent Place	Friend(s)	Sunny	80	2PM	Coffee House	2h	Female	21.0	Unmarried partner	...	NaN	4~8
4	No Urgent Place	Friend(s)	Sunny	80	2PM	Coffee House	1d	Female	21.0	Unmarried partner	...	NaN	4~8
...	...	...	...	...	...	...	...	...	...	...	...	...	...
12680	Work	Alone	Rainy	55	7AM	Carry out & Take away	1d	Male	26.0	Single	...	1~3	4~8
12681	Work	Alone	Snowy	30	7AM	Coffee House	1d	Male	26.0	Single	...	1~3	4~8
12682	Work	Alone	Snowy	30	7AM	Bar	1d	Male	26.0	Single	...	1~3	4~8
12683	Work	Alone	Sunny	80	7AM	Restaurant(20-50)	2h	Male	26.0	Single	...	1~3	4~8
12684	UNION	UNION	UNION	55	2PM	Restaurant(<20)	1d	Female	21.0	Unmarried partner	...	NaN	4~8

12685 rows × 27 columns

In [85]:
```python
tables_union = pd.concat([sql_dataset,union_dataset], ignore_index=True).drop_duplicates()
tables_union['destination'].unique()
```

Out[85]: array(['No Urgent Place', 'Home', 'Work', 'UNION'], dtype=object)

```
In [86]: join_dataset = pd.read_csv(r'D:\Data Science and Machine Learning\March\7th_SQL_For_Data_analysts\table_to_join_202503081735
         join_dataset
```

Out[86]:

	time	part_of_day
0	2PM	Afternoon
1	10AM	Morning
2	6PM	Evening
3	7AM	Morning
4	10PM	Night

```
In [90]: merged_dataset = pd.merge(sql_dataset,join_dataset,how='left',left_on='time',right_on='time')
         merged_dataset[['destination','time','part_of_day']].rename(columns={'destination':'Destination','time':'Dataset_Time','part
```

Out[90]:

	Destination	Dataset_Time	PartOfDay
0	No Urgent Place	2PM	Afternoon
1	No Urgent Place	10AM	Morning
2	No Urgent Place	10AM	Morning
3	No Urgent Place	2PM	Afternoon
4	No Urgent Place	2PM	Afternoon
...	...	...	...
12679	Home	6PM	Evening
12680	Work	7AM	Morning
12681	Work	7AM	Morning
12682	Work	7AM	Morning
12683	Work	7AM	Morning

12684 rows × 3 columns

```
In [94]: sql_dataset[sql_dataset['passanger'] == 'Alone'][['destination','passanger']].rename(columns = {'passanger':'Passenger','des
```

Out[94]:

	Destination	Passenger
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 [96]: `sql_dataset[sql_dataset['weather'].str.startswith('Sun')]`

Out[96]:

	destination	passanger	weather	temperature	time	coupon	expiration	gender	age	maritalStatus	...	CarryAway	RestaurantLessThan20
0	No Urgent Place	Alone	Sunny	55	2PM	Restaurant(<20)	1d	Female	21.0	Unmarried partner	...	NaN	4~8
1	No Urgent Place	Friend(s)	Sunny	80	10AM	Coffee House	2h	Female	21.0	Unmarried partner	...	NaN	4~8
2	No Urgent Place	Friend(s)	Sunny	80	10AM	Carry out & Take away	2h	Female	21.0	Unmarried partner	...	NaN	4~8
3	No Urgent Place	Friend(s)	Sunny	80	2PM	Coffee House	2h	Female	21.0	Unmarried partner	...	NaN	4~8
4	No Urgent Place	Friend(s)	Sunny	80	2PM	Coffee House	1d	Female	21.0	Unmarried partner	...	NaN	4~8
...	...	...	...	...	...	...	...	...	...	...	...	...	...
12673	Home	Alone	Sunny	30	6PM	Carry out & Take away	1d	Male	26.0	Single	...	1~3	4~8
12676	Home	Alone	Sunny	80	6PM	Restaurant(20-50)	1d	Male	26.0	Single	...	1~3	4~8
12677	Home	Partner	Sunny	30	6PM	Restaurant(<20)	1d	Male	26.0	Single	...	1~3	4~8
12678	Home	Partner	Sunny	30	10PM	Restaurant(<20)	2h	Male	26.0	Single	...	1~3	4~8
12683	Work	Alone	Sunny	80	7AM	Restaurant(20-50)	2h	Male	26.0	Single	...	1~3	4~8

10069 rows × 27 columns

In [97]: `sql_dataset[sql_dataset['weather'].str.contains('^Sun', case=True, na=False)]`

Out[97]:

	destination	passanger	weather	temperature	time	coupon	expiration	gender	age	maritalStatus	...	CarryAway	RestaurantLessThan20
0	No Urgent Place	Alone	Sunny	55	2PM	Restaurant(<20)	1d	Female	21.0	Unmarried partner	...	NaN	4~8
1	No Urgent Place	Friend(s)	Sunny	80	10AM	Coffee House	2h	Female	21.0	Unmarried partner	...	NaN	4~8
2	No Urgent Place	Friend(s)	Sunny	80	10AM	Carry out & Take away	2h	Female	21.0	Unmarried partner	...	NaN	4~8
3	No Urgent Place	Friend(s)	Sunny	80	2PM	Coffee House	2h	Female	21.0	Unmarried partner	...	NaN	4~8
4	No Urgent Place	Friend(s)	Sunny	80	2PM	Coffee House	1d	Female	21.0	Unmarried partner	...	NaN	4~8
...	...	...	...	...	...	...	...	...	...	...	...	...	...
12673	Home	Alone	Sunny	30	6PM	Carry out & Take away	1d	Male	26.0	Single	...	1~3	4~8
12676	Home	Alone	Sunny	80	6PM	Restaurant(20-50)	1d	Male	26.0	Single	...	1~3	4~8
12677	Home	Partner	Sunny	30	6PM	Restaurant(<20)	1d	Male	26.0	Single	...	1~3	4~8
12678	Home	Partner	Sunny	30	10PM	Restaurant(<20)	2h	Male	26.0	Single	...	1~3	4~8
12683	Work	Alone	Sunny	80	7AM	Restaurant(20-50)	2h	Male	26.0	Single	...	1~3	4~8

10069 rows × 27 columns

In [104]: `sql_dataset[(sql_dataset['temperature'] >= 30) & (sql_dataset['temperature'] < 75)]['temperature'].unique()`

Out[104]: `array([55, 30], dtype=int64)`

In [107]: `sql_dataset[(sql_dataset['occupation'] == 'Sales & Related') | (sql_dataset['occupation'] == 'Management')]['occupation']`

Out[107]:
```
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
Name: occupation, Length: 1931, dtype: object
```

```
In [109]: sql_dataset[sql_dataset['occupation'].isin(['Sales & Related','Management'])]['occupation']
```

```
Out[109]: 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
          Name: occupation, Length: 1931, dtype: object
```

```
In [116]: sql_dataset.groupby('weather')['temperature'].mean()
```

```
Out[116]: weather
          Rainy    55.000000
          Snowy    30.000000
          Sunny    68.946271
          Name: temperature, dtype: float64
```

```
In [118]: temp_set = sql_dataset
          temp_set['Average_Temperature'] = temp_set.groupby('weather')['temperature'].transform('mean')
          temp_set[['destination','weather','Average_Temperature']].sort_values(by = 'Average_Temperature')
```

Out[118]:

	destination	weather	Average_Temperature
12682	Work	Snowy	30.000000
10758	No Urgent Place	Snowy	30.000000
7369	Work	Snowy	30.000000
7368	Work	Snowy	30.000000
10765	Home	Snowy	30.000000
...	...	...	...
4231	Home	Sunny	68.946271
4232	Work	Sunny	68.946271
4233	Work	Sunny	68.946271
4235	Work	Sunny	68.946271
12683	Work	Sunny	68.946271

12684 rows × 3 columns

```
In [119]: temp_set = sql_dataset
          temp_set['Average_Temperature'] = temp_set.groupby('weather')['temperature'].transform('mean')
          temp_set[['destination','weather','Average_Temperature']].sort_values(by = 'index')
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
<ipython-input-119-0b4d918e7e4d> in <module>
      1 temp_set = sql_dataset
      2 temp_set['Average_Temperature'] = temp_set.groupby('weather')['temperature'].transform('mean')
----> 3 temp_set[['destination','weather','Average_Temperature']].sort_values(by = 'index')

~\Anaconda3\lib\site-packages\pandas\core\frame.py in sort_values(self, by, axis, ascending, inplace, kind, na_position)
   4991
   4992             by = by[0]
-> 4993             k = self._get_label_or_level_values(by, axis=axis)
   4994
   4995             if isinstance(ascending, (tuple, list)):

~\Anaconda3\lib\site-packages\pandas\core\generic.py in _get_label_or_level_values(self, key, axis)
   1772             values = self.axes[axis].get_level_values(key)._values
   1773         else:
-> 1774             raise KeyError(key)
   1775
   1776         # Check for duplicates

KeyError: 'index'
```

```
In [129]: temp_set = sql_dataset.copy()
          temp_set.dtypes
          temp_set['destination'].astype('str')
          temp_set['destination'].fillna('Unknown', inplace=True)
          temp_set['Row_Number'] = sql_dataset.groupby('weather')['destination'].rank(method='first').astype(int)
          temp_set[['destination','weather','Row_Number']]
```

```
          ---------------------------------------------------------------------
          TypeError                                 Traceback (most recent call last)
          <ipython-input-129-f0888511c1f3> in <module>
                3 temp_set['destination'].astype('str')
                4 temp_set['destination'].fillna('Unknown', inplace=True)
          ----> 5 temp_set['Row_Number'] = sql_dataset.groupby('weather')['destination'].rank(method='first').astype(int)
                6 temp_set[['destination','weather','Row_Number']]

          ~\Anaconda3\lib\site-packages\pandas\core\groupby\groupby.py in rank(self, method, ascending, na_option, pct, axis)
             2124             na_option=na_option,
             2125             pct=pct,
          -> 2126             axis=axis,
             2127         )
             2128

          ~\Anaconda3\lib\site-packages\pandas\core\groupby\groupby.py in _cython_transform(self, how, numeric_only, **kwargs)
              855
              856             try:
          --> 857                 result, names = self.grouper.transform(obj.values, how, **kwargs)
              858             except NotImplementedError:
              859                 continue

          ~\Anaconda3\lib\site-packages\pandas\core\groupby\ops.py in transform(self, values, how, axis, **kwargs)
              598
              599     def transform(self, values, how, axis=0, **kwargs):
          --> 600         return self._cython_operation("transform", values, how, axis, **kwargs)
              601
              602     def _aggregate(

          ~\Anaconda3\lib\site-packages\pandas\core\groupby\ops.py in _cython_operation(self, kind, values, how, axis, min_count, **kwargs)
              558             # TODO: min_count
              559             result = self._transform(
          --> 560                 result, values, labels, func, is_numeric, is_datetimelike, **kwargs
              561             )
              562

          ~\Anaconda3\lib\site-packages\pandas\core\groupby\ops.py in _transform(self, result, values, comp_ids, transform_func, is_numeric, is_datetimelike, **kwargs)
              655             )
              656         else:
          --> 657             transform_func(result, values, comp_ids, ngroups, is_datetimelike, **kwargs)
              658
              659         return result

          ~\Anaconda3\lib\site-packages\pandas\core\groupby\ops.py in wrapper(*args, **kwargs)
              435
              436             def wrapper(*args, **kwargs):
          --> 437                 return f(afunc, *args, **kwargs)
              438
              439             # need to curry our sub-function

          ~\Anaconda3\lib\site-packages\pandas\core\groupby\ops.py in <lambda>(func, a, b, c, d, e, **kwargs)
              387                     kwargs.get("ascending", True),
              388                     kwargs.get("pct", False),
          --> 389                     kwargs.get("na_option", "keep"),
              390                 ),
              391             },

          TypeError: 'NoneType' object is not callable
```

```
In [ ]:
```

```
In [ ]:
```