

Comparison of SQL and Python (Pandas)

```
In [1]: import warnings
warnings.filterwarnings("ignore")
```

```
In [2]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [3]: SQL = pd.read_csv(r"C:\Users\JANHAVI\Desktop\SQL Project\dataset_1.csv")
```

```
In [4]: SQL
```

Out[4]:

	destination	passanger	weather	temperature	time	coupon	expiration	gender
0	No Urgent Place	Alone	Sunny	55	2PM	Restaurant(<20)	1d	Female
1	No Urgent Place	Friend(s)	Sunny	80	10AM	Coffee House	2h	Female
2	No Urgent Place	Friend(s)	Sunny	80	10AM	Carry out & Take away	2h	Female
3	No Urgent Place	Friend(s)	Sunny	80	2PM	Coffee House	2h	Female
4	No Urgent Place	Friend(s)	Sunny	80	2PM	Coffee House	1d	Female
...
12679	Home	Partner	Rainy	55	6PM	Carry out & Take away	1d	Male
12680	Work	Alone	Rainy	55	7AM	Carry out & Take away	1d	Male
12681	Work	Alone	Snowy	30	7AM	Coffee House	1d	Male
12682	Work	Alone	Snowy	30	7AM	Bar	1d	Male
12683	Work	Alone	Sunny	80	7AM	Restaurant(20-50)	2h	Male

12684 rows × 27 columns

```
In [5]: SQL.head()
```

Out[5]:

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

5 rows × 27 columns

In [6]: `SQL.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12684 entries, 0 to 12683
Data columns (total 27 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   destination                          12684 non-null  object
1   passanger                           12684 non-null  object
2   weather                             12684 non-null  object
3   temperature                         12684 non-null  int64
4   time                                12684 non-null  object
5   coupon                              12684 non-null  object
6   expiration                           12684 non-null  object
7   gender                              12684 non-null  object
8   age                                 12684 non-null  object
9   maritalStatus                       12684 non-null  object
10  has_children                        12684 non-null  int64
11  education                           12684 non-null  object
12  occupation                           12684 non-null  object
13  income                              12684 non-null  object
14  car                                 108 non-null    object
15  Bar                                 12577 non-null  object
16  CoffeeHouse                         12467 non-null  object
17  CarryAway                           12533 non-null  object
18  RestaurantLessThan20                12554 non-null  object
19  Restaurant20To50                    12495 non-null  object
20  toCoupon_GEQ5min                    12684 non-null  int64
21  toCoupon_GEQ15min                   12684 non-null  int64
22  toCoupon_GEQ25min                   12684 non-null  int64
23  direction_same                      12684 non-null  int64
24  direction_opp                       12684 non-null  int64
25  Y                                    12684 non-null  int64
26  row_count                           12684 non-null  int64
dtypes: int64(9), object(18)
memory usage: 2.6+ MB
```

In [7]: `SQL.describe()`

Out[7]:

	temperature	has_children	toCoupon_GEQ5min	toCoupon_GEQ15min	toCoupon_GEQ25min
count	12684.000000	12684.000000	12684.0	12684.000000	12684.000000
mean	63.301798	0.414144	1.0	0.561495	0.119126
std	19.154486	0.492593	0.0	0.496224	0.323950
min	30.000000	0.000000	1.0	0.000000	0.000000
25%	55.000000	0.000000	1.0	0.000000	0.000000
50%	80.000000	0.000000	1.0	1.000000	0.000000
75%	80.000000	1.000000	1.0	1.000000	0.000000
max	80.000000	1.000000	1.0	1.000000	1.000000

In [8]: `SQL.shape`

Out[8]: (12684, 27)

In [9]: `SQL[['weather', 'temperature']]`

Out[9]:

	weather	temperature
0	Sunny	55
1	Sunny	80
2	Sunny	80
3	Sunny	80
4	Sunny	80
...
12679	Rainy	55
12680	Rainy	55
12681	Snowy	30
12682	Snowy	30
12683	Sunny	80

12684 rows × 2 columns

In [10]: `SQL['passanger'].unique()`

Out[10]: array(['Alone', 'Friend(s)', 'Kid(s)', 'Partner'], dtype=object)

In [11]: `SQL[SQL['destination']=='Home']`

Out[11]:

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

3237 rows × 27 columns

In [12]: `SQL.sort_values('coupon')`

Out[12]:

	destination	passanger	weather	temperature	time	coupon	expiration	gender
11702	Home	Partner	Sunny	30	10PM	Bar	2h	Female
9930	No Urgent Place	Alone	Snowy	30	2PM	Bar	1d	Female
10632	Home	Alone	Rainy	55	6PM	Bar	1d	Male
7997	No Urgent Place	Friend(s)	Rainy	55	10PM	Bar	2h	Male
11166	Work	Alone	Snowy	30	7AM	Bar	1d	Female
...
10476	Home	Alone	Sunny	80	6PM	Restaurant(<20)	1d	Female
5447	Home	Alone	Sunny	80	10PM	Restaurant(<20)	2h	Female
10478	Home	Alone	Snowy	30	10PM	Restaurant(<20)	2h	Female
5440	No Urgent Place	Alone	Sunny	80	2PM	Restaurant(<20)	2h	Female
0	No Urgent Place	Alone	Sunny	55	2PM	Restaurant(<20)	1d	Female

12684 rows × 27 columns

In [13]: `SQL.rename(columns={'destination': 'Destination'}, inplace=True)`In [14]: `SQL`

Out[14]:

	Destination	passanger	weather	temperature	time	coupon	expiration	gender
0	No Urgent Place	Alone	Sunny	55	2PM	Restaurant(<20)	1d	Female
1	No Urgent Place	Friend(s)	Sunny	80	10AM	Coffee House	2h	Female
2	No Urgent Place	Friend(s)	Sunny	80	10AM	Carry out & Take away	2h	Female
3	No Urgent Place	Friend(s)	Sunny	80	2PM	Coffee House	2h	Female
4	No Urgent Place	Friend(s)	Sunny	80	2PM	Coffee House	1d	Female
...
12679	Home	Partner	Rainy	55	6PM	Carry out & Take away	1d	Male
12680	Work	Alone	Rainy	55	7AM	Carry out & Take away	1d	Male
12681	Work	Alone	Snowy	30	7AM	Coffee House	1d	Male
12682	Work	Alone	Snowy	30	7AM	Bar	1d	Male
12683	Work	Alone	Sunny	80	7AM	Restaurant(20-50)	2h	Male

12684 rows × 27 columns

In [15]: `SQL.groupby('occupation').size().to_frame('Count').reset_index()`

Out[15]:

	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

In [16]: `SQL.groupby('weather')['temperature'].mean().to_frame('avg_temp').reset_index()`

Out[16]:

	weather	avg_temp
0	Rainy	55.000000
1	Snowy	30.000000
2	Sunny	68.946271

In [17]: `SQL.groupby('weather')['temperature'].size().to_frame('Count_temp').reset_index()`

Out[17]:

	weather	Count_temp
0	Rainy	1210
1	Snowy	1405
2	Sunny	10069

In [18]: `SQL.groupby('weather')['temperature'].nunique().to_frame('count_distinct_temp').reset_index()`

Out[18]:

	weather	count_distinct_temp
0	Rainy	1
1	Snowy	1
2	Sunny	3

In [19]: `SQL.groupby('weather')['temperature'].sum().to_frame('sum_temp').reset_index()`

Out[19]:

	weather	sum_temp
0	Rainy	66550
1	Snowy	42150
2	Sunny	694220

In [20]: `SQL.groupby('weather')['temperature'].min().to_frame('min_temp').reset_index()`

Out[20]:

	weather	min_temp
0	Rainy	55
1	Snowy	30
2	Sunny	30

In [21]: `SQL.groupby('weather')['temperature'].max().to_frame('max_temp').reset_index()`

Out[21]:

	weather	max_temp
0	Rainy	55
1	Snowy	30
2	Sunny	80

In [22]: `SQL.groupby('occupation').filter(lambda x: x['occupation'].iloc[0] == 'Student').groupby('occupation').size()`

Out[22]:

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

In [23]: `SQL1 = pd.read_csv(r"C:\Users\JANHAVI\Desktop\SQL Project\table_to_join_20250806135", engine='python')`

Out[23]:

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

In [26]: `pd.concat([SQL, SQL1])['Destination'].drop_duplicates()`

Out[26]:

0	No Urgent Place
13	Home
16	Work
0	NaN

Name: Destination, dtype: object

In [27]: `pd.concat([SQL, SQL1])['Destination'].drop_duplicates().reset_index(drop=True)`

Out[27]:

0	No Urgent Place
1	Home
2	Work
3	NaN

Name: Destination, dtype: object

In [28]: `SQL2 = pd.read_csv(r"C:\Users\JANHAVI\Desktop\SQL Project\table_to_union_2025080613\SQL2")`

Out[28]:

	destination	passanger	weather	temperature	time	coupon	expiration	gender	age
0	UNION	UNION	UNION	55	2PM	Restaurant(<20)	1d	Female	21

1 rows × 27 columns

In [31]: `pd.merge(SQL, SQL1[['time', 'part_of_day']], on='time', how='inner')[['Destination']]`

Out[31]:

	Destination	time	part_of_day
0	No Urgent Place	2PM	Afternoon
1	No Urgent Place	2PM	Afternoon
2	No Urgent Place	2PM	Afternoon
3	No Urgent Place	2PM	Afternoon
4	No Urgent Place	2PM	Afternoon
...
12679	No Urgent Place	10PM	Night
12680	No Urgent Place	10PM	Night
12681	Home	10PM	Night
12682	Home	10PM	Night
12683	Home	10PM	Night

12684 rows × 3 columns

In [33]: SQL[SQL['passanger'] == 'Alone'][['Destination', 'passanger']]

Out[33]:

	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 [34]: SQL[SQL['weather'].str.startswith('Sun')]

Out[34]:

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

10069 rows × 27 columns



In [36]: SQL[(SQL['temperature'] >= 29) & (SQL['temperature'] <= 75)][['temperature']].unique()

Out[36]: array([55, 30], dtype=int64)

In [37]: SQL[SQL['occupation'].isin(['Sales & Related', 'Management'])][['occupation']]

Out[37]:

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