```
import pandas as pd

df=pd.read_csv(r"C:\Users\Admin\dataset_1_202508051858.csv")

df
```

```
             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
...                  ...        ...     ...          ...   ...
3111721             Home      Alone   Sunny           80   6PM
3111722             Home      Alone   Sunny           80   6PM
3111723             Home      Alone   Sunny           80   6PM
3111724             Home      Alone   Sunny           80  10PM
3111725             Home      Alone   Sunny           80  10PM

                       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  ...
...                        ...        ...     ...  ..                .
..  ...
3111721           Coffee House         1d  Female  26    Married
partner  ...
3111722        Restaurant(<20)         2h  Female  26    Married
partner  ...
3111723           Coffee House         2h  Female  26    Married
partner  ...
3111724        Restaurant(<20)         2h  Female  26    Married
partner  ...
3111725  Carry out & Take away         1d  Female  26    Married
partner  ...

         CarryAway RestaurantLessThan20 Restaurant20To50
toCoupon_GEQ5min  \
0              NaN                 4~8              1~3
1
1              NaN                 4~8              1~3
1
```

```
2               NaN                   4~8               1~3
1
3               NaN                   4~8               1~3
1
4               NaN                   4~8               1~3
1
...             ...                   ...               ...
...
3111721        4~8                   gt8               1~3
1
3111722        4~8                   gt8               1~3
1
3111723        4~8                   gt8               1~3
1
3111724        4~8                   gt8               1~3
1
3111725        4~8                   gt8               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
...                   ...               ...            ...            .
.. ..
3111721                 0                 0              0
1  1
3111722                 0                 0              1
0  1
3111723                 1                 1              0
1  1
3111724                 0                 0              1
0  1
3111725                 0                 0              1
0  1

        row_count
0               1
1               2
2               3
3               4
```

```
4               5
...           ...
3111721      4142
3111722      4143
3111723      4144
3111724      4145
3111725      4146

[3111726 rows x 27 columns]
```

```
df[['weather','temperature']]
```

```
        weather  temperature
0         Sunny           55
1         Sunny           80
2         Sunny           80
3         Sunny           80
4         Sunny           80
...         ...          ...
3111721   Sunny           80
3111722   Sunny           80
3111723   Sunny           80
3111724   Sunny           80
3111725   Sunny           80

[3111726 rows x 2 columns]
```

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

```
       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
5  No Urgent Place  Friend(s)   Sunny           80   6PM
6  No Urgent Place  Friend(s)   Sunny           55   2PM
7  No Urgent Place     Kid(s)   Sunny           80  10AM
8  No Urgent Place     Kid(s)   Sunny           80  10AM
9  No Urgent Place     Kid(s)   Sunny           80  10AM

                  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  ...
5        Restaurant(<20)         2h  Female  21  Unmarried
partner  ...
6  Carry out & Take away         1d  Female  21  Unmarried
partner  ...
7        Restaurant(<20)         2h  Female  21  Unmarried
partner  ...
8  Carry out & Take away         2h  Female  21  Unmarried
partner  ...
9                    Bar         1d  Female  21  Unmarried
partner  ...

   CarryAway RestaurantLessThan20 Restaurant20To50 toCoupon_GEQ5min  \
0        NaN                 4~8              1~3                 1
1        NaN                 4~8              1~3                 1
2        NaN                 4~8              1~3                 1
3        NaN                 4~8              1~3                 1
4        NaN                 4~8              1~3                 1
5        NaN                 4~8              1~3                 1
6        NaN                 4~8              1~3                 1
7        NaN                 4~8              1~3                 1
8        NaN                 4~8              1~3                 1
9        NaN                 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

5                 1                 0              0             1  1

6                 1                 0              0             1  1

7                 1                 0              0             1  1

8                 1                 0              0             1  1

9                 1                 0              0             1  0


  row_count
```

```
0          1
1          2
2          3
3          4
4          5
5          6
6          7
7          8
8          9
9         10

[10 rows x 27 columns]
```

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

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

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

```
        destination passanger weather  temperature  time  \
13             Home     Alone   Sunny           55   6PM
14             Home     Alone   Sunny           55   6PM
15             Home     Alone   Sunny           80   6PM
35             Home     Alone   Sunny           55   6PM
36             Home     Alone   Sunny           55   6PM
...             ...       ...     ...          ...   ...
3111721        Home     Alone   Sunny           80   6PM
3111722        Home     Alone   Sunny           80   6PM
3111723        Home     Alone   Sunny           80   6PM
3111724        Home     Alone   Sunny           80  10PM
3111725        Home     Alone   Sunny           80  10PM

                         coupon expiration  gender age
maritalStatus  ...  \
13                          Bar         1d  Female  21  Unmarried
partner  ...
14           Restaurant(20-50)         1d  Female  21  Unmarried
partner  ...
15                 Coffee House         2h  Female  21  Unmarried
partner  ...
35                          Bar         1d    Male  21
Single  ...
36           Restaurant(20-50)         1d    Male  21
Single  ...
...                         ...        ...     ...  ..                .
..  ...
3111721            Coffee House         1d  Female  26    Married
partner  ...
3111722          Restaurant(<20)         2h  Female  26    Married
partner  ...
```

```
3111723           Coffee House         2h  Female  26    Married
partner  ...
3111724       Restaurant(<20)         2h  Female  26    Married
partner  ...
3111725  Carry out & Take away         1d  Female  26    Married
partner  ...

         CarryAway RestaurantLessThan20 Restaurant20To50
toCoupon_GEQ5min  \
13             NaN                  4~8              1~3
1
14             NaN                  4~8              1~3
1
15             NaN                  4~8              1~3
1
35             4~8                  4~8            less1
1
36             4~8                  4~8            less1
1
...            ...                  ...              ...
...
3111721        4~8                  gt8              1~3
1
3111722        4~8                  gt8              1~3
1
3111723        4~8                  gt8              1~3
1
3111724        4~8                  gt8              1~3
1
3111725        4~8                  gt8              1~3
1

        toCoupon_GEQ15min toCoupon_GEQ25min direction_same
direction_opp  Y  \
13                      0                 0              1
0  1
14                      1                 0              0
1  1
15                      0                 0              0
1  0
35                      0                 0              1
0  1
36                      1                 0              0
1  0
...                   ...               ...            ...            .
.. ..
3111721                 0                 0              0
1  1
3111722                 0                 0              1
```

```
0  1
3111723                 1                 1              0
1  1
3111724                 0                 0              1
0  1
3111725                 0                 0              1
0  1

        row_count
13             14
14             15
15             16
35             36
36             37
...           ...
3111721      4142
3111722      4143
3111723      4144
3111724      4145
3111725      4146

[793646 rows x 27 columns]
```

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

```
             destination  passanger weather  temperature  time  \
2011253  No Urgent Place  Friend(s)   Rainy           55  10PM
1860992             Home    Partner   Sunny           55  10PM
1860990             Home      Alone   Rainy           55   6PM
517049   No Urgent Place      Alone   Snowy           30   2PM
517051   No Urgent Place    Partner   Sunny           55   2PM
...                  ...        ...     ...          ...   ...
1349885             Home      Alone   Sunny           80  10PM
1349883             Home      Alone   Sunny           80   6PM
2337157  No Urgent Place  Friend(s)   Sunny           80   2PM
1349702  No Urgent Place     Kid(s)   Sunny           80  10AM
0        No Urgent Place      Alone   Sunny           55   2PM

                   coupon expiration  gender     age      maritalStatus
...  \
2011253              Bar         2h    Male      46             Single
...
1860992              Bar         1d  Female      21    Married partner
...
1860990              Bar         1d  Female      21    Married partner
...
517049               Bar         1d    Male      41    Married partner
...
517051               Bar         1d    Male      41    Married partner
...
```

```
...                  ...        ...     ...     ...                ...
...
1349885  Restaurant(<20)         2h    Male      26             Single
...
1349883  Restaurant(<20)         2h    Male      26             Single
...
2337157  Restaurant(<20)         1d    Male      21             Single
...
1349702  Restaurant(<20)         1d  Female  50plus  Unmarried partner
...
0        Restaurant(<20)         1d  Female      21  Unmarried partner
...

         CarryAway RestaurantLessThan20 Restaurant20To50
toCoupon_GEQ5min  \
2011253        1~3                  4~8            less1
1
1860992        4~8                  1~3              1~3
1
1860990        4~8                  1~3              1~3
1
517049         gt8                  1~3              1~3
1
517051         gt8                  1~3              1~3
1
...            ...                  ...              ...
...
1349885        1~3                  1~3            less1
1
1349883        1~3                  1~3            less1
1
2337157        4~8                  1~3            less1
1
1349702        4~8                  1~3            less1
1
0              NaN                  4~8              1~3
1

        toCoupon_GEQ15min toCoupon_GEQ25min direction_same
direction_opp  Y  \
2011253                 1                 0              0
1  0
1860992                 1                 1              0
1  1
1860990                 1                 1              0
1  1
517049                  0                 0              0
1  1
517051                  0                 0              0
```

```
1  0
...                   ...                ...             ...           .
.. ..
1349885                 0                  0               1
0  1
1349883                 0                  0               1
0  1
2337157                 1                  0               0
1  1
1349702                 1                  0               0
1  0
0                       0                  0               0
1  1
```

```
        row_count
2011253      7182
1860992      9129
1860990      9127
517049       9690
517051       9692
...           ...
1349885      5382
1349883      5380
2337157      3302
1349702      5199
0               1
```

```
[3111726 rows x 27 columns]
```

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

```
df
```

```
             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
...                  ...        ...     ...          ...   ...
3111721             Home      Alone   Sunny           80   6PM
3111722             Home      Alone   Sunny           80   6PM
3111723             Home      Alone   Sunny           80   6PM
3111724             Home      Alone   Sunny           80  10PM
3111725             Home      Alone   Sunny           80  10PM

                        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  ...
...                        ...        ...     ...  ..                .
..  ...
3111721           Coffee House         1d  Female  26    Married
partner  ...
3111722        Restaurant(<20)         2h  Female  26    Married
partner  ...
3111723           Coffee House         2h  Female  26    Married
partner  ...
3111724        Restaurant(<20)         2h  Female  26    Married
partner  ...
3111725  Carry out & Take away         1d  Female  26    Married
partner  ...

         CarryAway RestaurantLessThan20 Restaurant20To50
toCoupon_GEQ5min  \
0              NaN                  4~8              1~3
1
1              NaN                  4~8              1~3
1
2              NaN                  4~8              1~3
1
3              NaN                  4~8              1~3
1
4              NaN                  4~8              1~3
1
...            ...                  ...              ...
...
3111721        4~8                  gt8              1~3
1
3111722        4~8                  gt8              1~3
1
3111723        4~8                  gt8              1~3
1
3111724        4~8                  gt8              1~3
1
3111725        4~8                  gt8              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
...                   ...                ...             ...             .
.. ..
3111721                 0                  0               0
1  1
3111722                 0                  0               1
0  1
3111723                 1                  1               0
1  1
3111724                 0                  0               1
0  1
3111725                 0                  0               1
0  1

        row_count
0               1
1               2
2               3
3               4
4               5
...           ...
3111721      4142
3111722      4143
3111723      4144
3111724      4145
3111725      4146

[3111726 rows x 27 columns]
```

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

```
                                    occupation   Count
0                  Architecture & Engineering   42984
1    Arts Design Entertainment Sports & Media  154387
2   Building & Grounds Cleaning & Maintenance   10780
3                        Business & Financial  133418
4                 Community & Social Services   59089
5                     Computer & Mathematical  345439
6                   Construction & Extraction   37796
7                  Education&Training&Library  231343
8                  Farming Fishing & Forestry   10535
9          Food Preparation & Serving Related   73076
```

```
10       Healthcare Practitioners & Technical   59863
11                         Healthcare Support   59356
12            Installation Maintenance & Repair   32630
13                                       Legal   53742
14                  Life Physical Social Science   41776
15                                  Management  205708
16              Office & Administrative Support  156731
17                     Personal Care & Service   42918
18                      Production Occupations   26950
19                          Protective Service   42979
20                                     Retired  121473
21                              Sales & Related  268179
22                                     Student  388562
23              Transportation & Material Moving   53476
24                                  Unemployed  458536
```

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

```
  weather   avg_temp
0   Rainy  55.000000
1   Snowy  30.000000
2   Sunny  68.947152
```

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

```
  weather  Count_temp
0   Rainy      296450
1   Snowy      344225
2   Sunny     2471051
```

```
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
```

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

```
  weather   sum_temp
0   Rainy   16304750
1   Snowy   10326750
2   Sunny  170371930
```

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

```
  weather  min_temp
0   Rainy        55
1   Snowy        30
2   Sunny        30
```

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

```
  weather  max_temp
0   Rainy        55
1   Snowy        30
2   Sunny        80
```

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

```
occupation
Student    388562
dtype: int64
```

```
pd.concat([df,df1])['destination'].drop_duplicates()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call
last)
Cell In[26], line 1
----> 1 pd.concat([df,df1])['destination'].drop_duplicates()

NameError: name 'df1' is not defined
```

```
pd.merge(df, df2[['time', 'part_of_day']], on='time', how='inner')
[['destination', 'time', 'part_of_day']]
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call
last)
Cell In[27], line 1
----> 1 pd.merge(df, df2[['time', 'part_of_day']], on='time',
how='inner')[['destination', 'time', 'part_of_day']]

NameError: name 'df2' is not defined
```

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

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call
last)
Cell In[28], line 1
----> 1 df[df['passanger'] == 'Alone'][['destination', 'passanger']]
```

```
File ~\anaconda3\Lib\site-packages\pandas\core\frame.py:4108, in
DataFrame.__getitem__(self, key)
   4106     if is_iterator(key):
   4107         key = list(key)
-> 4108     indexer = self.columns._get_indexer_strict(key, "columns")
[1]
   4110 # take() does not accept boolean indexers
   4111 if getattr(indexer, "dtype", None) == bool:

File ~\anaconda3\Lib\site-packages\pandas\core\indexes\base.py:6200,
in Index._get_indexer_strict(self, key, axis_name)
   6197 else:
   6198     keyarr, indexer, new_indexer =
self._reindex_non_unique(keyarr)
-> 6200 self._raise_if_missing(keyarr, indexer, axis_name)
   6202 keyarr = self.take(indexer)
   6203 if isinstance(key, Index):
   6204     # GH 42790 - Preserve name from an Index

File ~\anaconda3\Lib\site-packages\pandas\core\indexes\base.py:6252,
in Index._raise_if_missing(self, key, indexer, axis_name)
   6249     raise KeyError(f"None of [{key}] are in the
[{axis_name}]")
   6251 not_found = list(ensure_index(key)[missing_mask.nonzero()
[0]].unique())
-> 6252 raise KeyError(f"{not_found} not in index")

KeyError: "['destination'] not in index"
```

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

```
             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
...                  ...        ...     ...          ...   ...
3111721             Home      Alone   Sunny           80   6PM
3111722             Home      Alone   Sunny           80   6PM
3111723             Home      Alone   Sunny           80   6PM
3111724             Home      Alone   Sunny           80  10PM
3111725             Home      Alone   Sunny           80  10PM

                        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  ...
...                         ...        ...     ...  ..                 .
..  ...
3111721          Coffee House          1d  Female  26    Married
partner  ...
3111722       Restaurant(<20)          2h  Female  26    Married
partner  ...
3111723          Coffee House          2h  Female  26    Married
partner  ...
3111724       Restaurant(<20)          2h  Female  26    Married
partner  ...
3111725  Carry out & Take away         1d  Female  26    Married
partner  ...

         CarryAway RestaurantLessThan20 Restaurant20To50
toCoupon_GEQ5min  \
0              NaN                  4~8              1~3
1
1              NaN                  4~8              1~3
1
2              NaN                  4~8              1~3
1
3              NaN                  4~8              1~3
1
4              NaN                  4~8              1~3
1
...            ...                  ...              ...
...
3111721        4~8                  gt8              1~3
1
3111722        4~8                  gt8              1~3
1
3111723        4~8                  gt8              1~3
1
3111724        4~8                  gt8              1~3
1
3111725        4~8                  gt8              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
...                    ...                 ...              ...              .
.. ..
3111721                  0                   0                0
1  1
3111722                  0                   0                1
0  1
3111723                  1                   1                0
1  1
3111724                  0                   0                1
0  1
3111725                  0                   0                1
0  1
```

	row_count
0	1
1	2
2	3
3	4
4	5
...	...
3111721	4142
3111722	4143
3111723	4144
3111724	4145
3111725	4146

[2471051 rows x 27 columns]

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

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

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

	occupation
193	Sales & Related
194	Sales & Related
195	Sales & Related
196	Sales & Related
197	Sales & Related

```
...                    ...
3111639  Sales & Related
3111640  Sales & Related
3111641  Sales & Related
3111642  Sales & Related
3111643  Sales & Related

[473887 rows x 1 columns]
```