

```
In [1]: import pandas as pd
```

```
In [2]: df=pd.read_csv('g:/dataset/analysis/restaurant.csv')
```

```
In [3]: df
```

```
Out[3]:    total_bill  tip  gender  smoker  day  time  size
```

	total_bill	tip	gender	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4
...
239	29.03	5.92	Male	No	Sat	Dinner	3
240	27.18	2.00	Female	Yes	Sat	Dinner	2
241	22.67	2.00	Male	Yes	Sat	Dinner	2
242	17.82	1.75	Male	No	Sat	Dinner	2
243	18.78	3.00	Female	No	Thur	Dinner	2

244 rows × 7 columns

```
In [4]: df.total_bill.sum()
```

```
Out[4]: 4827.77
```

```
In [5]: df.describe()
```

```
Out[5]:    total_bill      tip      size
```

	total_bill	tip	size
count	244.000000	244.000000	244.000000
mean	19.785943	2.998279	2.569672
std	8.902412	1.383638	0.951100
min	3.070000	1.000000	1.000000
25%	13.347500	2.000000	2.000000
50%	17.795000	2.900000	2.000000
75%	24.127500	3.562500	3.000000
max	50.810000	10.000000	6.000000

```
In [16]: df.groupby(by='day')[['total_bill']].sum()
```

```
Out[16]: total_bill
```

day	
Fri	325.88
Sat	1778.40
Sun	1627.16
Thur	1096.33

```
In [17]: df.pivot_table(index='day',values='total_bill',aggfunc='sum')
```

```
Out[17]: total_bill
```

day	
Fri	325.88
Sat	1778.40
Sun	1627.16
Thur	1096.33

```
In [19]: df.pivot_table(index=['day','time'],values='total_bill',aggfunc='sum')
```

```
Out[19]: total_bill
```

day	time	
Fri	Dinner	235.96
	Lunch	89.92
Sat	Dinner	1778.40
Sun	Dinner	1627.16
Thur	Dinner	18.78
	Lunch	1077.55

```
In [20]: df.groupby(by=['day','time'])[['total_bill']].sum()
```

```
Out[20]: total_bill
```

day	time	
Fri	Dinner	235.96
	Lunch	89.92
Sat	Dinner	1778.40
Sun	Dinner	1627.16
Thur	Dinner	18.78
	Lunch	1077.55

```
In [22]: df.pivot_table(index=['day'],columns='time',values='total_bill',aggfunc='sum')
```

```
Out[22]: time Dinner Lunch
```

day		
Fri	235.96	89.92
Sat	1778.40	NaN
Sun	1627.16	NaN
Thur	18.78	1077.55

```
In [24]: df.pivot_table(index=['day'],columns='time',values='total_bill',aggfunc=['sum','mean','count'])
```

time	Dinner	Lunch	sum	mean	count	
			Dinner	Lunch	Dinner	Lunch
day						
Fri	235.96	89.92	19.663333	12.845714	12.0	7.0
Sat	1778.40	NaN	20.441379	NaN	87.0	NaN
Sun	1627.16	NaN	21.410000	NaN	76.0	NaN
Thur	18.78	1077.55	18.780000	17.664754	1.0	61.0

```
In [25]: resdf=df.pivot_table(index=['day'],columns='time',values='total_bill',aggfunc=['sum','mean','count'])
```

```
In [26]: resdf.to_excel('report.xlsx')
```

```
In [27]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 244 entries, 0 to 243
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   total_bill  244 non-null    float64
 1   tip          244 non-null    float64
 2   gender       244 non-null    object 
 3   smoker       244 non-null    object 
 4   day          244 non-null    object 
 5   time         244 non-null    object 
 6   size          244 non-null    int64  
dtypes: float64(2), int64(1), object(4)
memory usage: 13.5+ KB
```

```
In [28]: df[(df.time=='Lunch')&(df.day=='Sun')]
```

```
Out[28]: total_bill  tip  gender  smoker  day  time  size
```

```
In [31]: df[(df.time=='Dinner')&(df.gender=='Female')].shape
```

```
Out[31]: (52, 7)
```

```
In [33]: df[(df.time=='Dinner')&(df.gender=='Male')].shape
```

```
Out[33]: (124, 7)
```

```
In [35]: df[(df.time=='Dinner')].shape
```

```
Out[35]: (176, 7)
```

```
In [36]: df=pd.read_csv('g:/dataset/analysis/titanic.csv')
```

```
In [37]: df
```

	PassengerId	Survived	Pclass	Name	gender	Age	SibSp	Parch	Ticket	Fare	Cabin
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	I
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th... Heikkinen, Miss. Laina	female	38.0	1	0	PC 17599	71.2833	C
2	3	1	3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	26.0	0	0	STON/O2. 3101282	7.9250	I
3	4	1	1	Allen, Mr. William Henry	male	35.0	1	0	113803	53.1000	C
4	5	0	3	Montvila, Rev. Juozas	male	35.0	0	0	373450	8.0500	I
...
886	887	0	2	Graham, Miss. Margaret Edith	male	27.0	0	0	211536	13.0000	I
887	888	1	1	Johnston, Miss. Catherine Helen "Carrie"	female	19.0	0	0	112053	30.0000	C
888	889	0	3	Behr, Mr. Karl Howell	female	Nan	1	2	W./C. 6607	23.4500	I
889	890	1	1	Dooley, Mr. Patrick	male	26.0	0	0	111369	30.0000	C
890	891	0	3		male	32.0	0	0	370376	7.7500	I

891 rows × 12 columns

```
In [38]: df.Survived.value_counts()
```

```
Out[38]: 0    549  
1    342  
Name: Survived, dtype: int64
```

```
In [39]: 549/891
```

```
Out[39]: 0.6161616161616161
```

```
In [40]: 342/891
```

```
Out[40]: 0.3838383838383838
```

```
In [41]: df.Pclass.value_counts()
```

```
Out[41]: 3    491  
1    216  
2    184  
Name: Pclass, dtype: int64
```

```
In [43]: df.pivot_table(index='Survived',columns='Pclass',values='Pclass',aggfunc='count')
```

```
-----  
ValueError                                     Traceback (most recent call last)  
Cell In[43], line 1  
----> 1 df.pivot_table(index='Survived',columns='Pclass',values='Pclass',aggfunc='count')  
  
File ~\anaconda3\Lib\site-packages\pandas\core\frame.py:8731, in DataFrame.pivot_table(self, values, index, columns, aggfunc, fill_value, margins, dropna, margins_name, observed, sort)  
    8714     @Substitution("")  
    8715     @Appender(_shared_docs["pivot_table"])  
    8716     def pivot_table(  
    (...)  
    8727         sort=True,  
    8728     ) -> DataFrame:  
    8729         from pandas.core.reshape.pivot import pivot_table  
-> 8731     return pivot_table(  
    8732         self,  
    8733         values=values,  
    8734         index=index,  
    8735         columns=columns,  
    8736         aggfunc=aggfunc,  
    8737         fill_value=fill_value,  
    8738         margins=margins,  
    8739         dropna=dropna,  
    8740         margins_name=margins_name,  
    8741         observed=observed,  
    8742         sort=sort,  
    8743     )  
  
File ~\anaconda3\Lib\site-packages\pandas\core\reshape\pivot.py:97, in pivot_table(data, values, index, columns, aggfunc, fill_value, margins, dropna, margins_name, observed, sort)  
    94     table = concat(pieces, keys=keys, axis=1)  
    95     return table._finalize_(data, method="pivot_table")  
---> 97 table = _internal_pivot_table(  
    98     data,  
    99     values,  
   100    index,  
   101    columns,  
   102    aggfunc,  
   103    fill_value,  
   104    margins,  
   105    dropna,  
   106    margins_name,  
   107    observed,  
   108    sort,  
   109 )  
110 return table._finalize_(data, method="pivot_table")  
  
File ~\anaconda3\Lib\site-packages\pandas\core\reshape\pivot.py:166, in _internal_pivot_table(data, values, index, columns, aggfunc, fill_value, margins, dropna, margins_name, observed, sort)  
    163         pass  
    164     values = list(values)  
--> 166 grouped = data.groupby(keys, observed=observed, sort=sort)  
    167 msg = (  
    168     "pivot_table dropped a column because it failed to aggregate. This beh  
avior "  
    169     "is deprecated and will raise in a future version of pandas. Select on  
ly the "  
    170     "columns that can be aggregated."  
    171 )  
172 with rewrite_warning(
```

```

173     target_message="The default value of numeric_only",
174     target_category=FutureWarning,
175     new_message=msg,
176   ):

File ~\anaconda3\Lib\site-packages\pandas\core\frame.py:8402, in DataFrame.groupby
(self, by, axis, level, as_index, sort, group_keys, squeeze, observed, dropna)
    8399     raise TypeError("You have to supply one of 'by' and 'level'")
    8400 axis = self.get_axis_number(axis)
--> 8402 return DataFrameGroupBy(
    8403     obj=self,
    8404     keys=by,
    8405     axis=axis,
    8406     level=level,
    8407     as_index=as_index,
    8408     sort=sort,
    8409     group_keys=group_keys,
    8410     squeeze=squeeze,
    8411     observed=observed,
    8412     dropna=dropna,
    8413 )

File ~\anaconda3\Lib\site-packages\pandas\core\groupby\groupby.py:965, in GroupBy.__init__(self, obj, keys, axis, level, grouper, exclusions, selection, as_index, sort, group_keys, squeeze, observed, mutated, dropna)
    962 if grouper is None:
    963     from pandas.core.groupby.grouper import get_grouper
--> 965     grouper, exclusions, obj = get_grouper(
    966         obj,
    967         keys,
    968         axis=axis,
    969         level=level,
    970         sort=sort,
    971         observed=observed,
    972         mutated=self._mutated,
    973         dropna=self._dropna,
    974     )
    976 self._obj = obj
    977 self._axis = obj._get_axis_number(axis)

File ~\anaconda3\Lib\site-packages\pandas\core\groupby\grouper.py:883, in get_grouper(obj, key, axis, level, sort, observed, mutated, validate, dropna)
    879     in_axis, name, gpr = True, gpr, obj[gpr]
    880     if gpr.ndim != 1:
    881         # non-unique columns; raise here to get the name in the
    882         # exception message
--> 883         raise ValueError(f"Grouper for '{name}' not 1-dimensional")
    884     exclusions.add(name)
    885 elif obj._is_level_reference(gpr, axis=axis):

ValueError: Grouper for 'Pclass' not 1-dimensional

```

In [44]: df[(df.Survived==1)&(df.Pclass==1)].shape

Out[44]: (136, 12)

In [45]: df[(df.Survived==0)&(df.Pclass==1)].shape

Out[45]: (80, 12)

In [46]: df[(df.Survived==1)&(df.Pclass==2)].shape

```
Out[46]: (87, 12)
```

```
In [48]: df[(df.Survived==0)&(df.Pclass==2)].shape
```

```
Out[48]: (97, 12)
```

```
In [49]: df[(df.Survived==1)&(df.Pclass==3)].shape
```

```
Out[49]: (119, 12)
```

```
In [50]: df[(df.Survived==0)&(df.Pclass==3)].shape
```

```
Out[50]: (372, 12)
```

```
In [51]: df.gender.value_counts()
```

```
Out[51]: male      577  
female    314  
Name: gender, dtype: int64
```

```
In [52]: df[(df.Survived==1)&(df.gender=='male')].shape
```

```
Out[52]: (109, 12)
```

```
In [53]: df[(df.Survived==1)&(df.gender=='female')].shape
```

```
Out[53]: (233, 12)
```

```
In [54]: df[(df.Age>=0)&(df.Age<=10)].shape
```

```
Out[54]: (64, 12)
```

```
In [55]: df[(df.Age>=0)&(df.Age<=10)&(df.Survived==1)].shape
```

```
Out[55]: (38, 12)
```

```
In [57]: df[(df.Age>=60)].shape
```

```
Out[57]: (26, 12)
```

```
In [58]: df[(df.Age>=60)&(df.Survived==1)].shape
```

```
Out[58]: (7, 12)
```

```
In [59]: df[(df.Age>=60)&(df.Survived==1)&(df.gender=='female')].shape
```

```
Out[59]: (4, 12)
```

```
In [60]: df[(df.Age>=60)&(df.Survived==1)&(df.Pclass==1)].shape
```

```
Out[60]: (5, 12)
```

```
In [61]: df[(df.SibSp==0)&(df.Parch==0)].shape
```

```
Out[61]: (537, 12)
```

```
In [62]: df[(df.SibSp==0)&(df.Parch==0)&(df.Survived==1)].shape
```

```
Out[62]: (163, 12)
```

```
In [63]: df[(df.Pclass==1)&(df.Survived==1)&(df.gender=='male')].shape
```

```
Out[63]: (45, 12)
```

```
In [64]: df[(df.Pclass==1)&(df.gender=='male')].shape
```

```
Out[64]: (122, 12)
```

```
In [65]: df[(df.Pclass==1)&(df.Survived==1)&(df.gender=='female')].shape
```

```
Out[65]: (91, 12)
```

```
In [66]: df[(df.Pclass==1)&(df.gender=='female')].shape
```

```
Out[66]: (94, 12)
```

```
In [69]: df.Age.isnull().sum()
```

```
Out[69]: 177
```

```
In [70]: df.Survived.isnull().sum()
```

```
Out[70]: 0
```

```
In [71]: df.isnull().sum()
```

```
Out[71]: PassengerId      0  
Survived          0  
Pclass            0  
Name              0  
gender            0  
Age             177  
SibSp            0  
Parch            0  
Ticket           0  
Fare             0  
Cabin          687  
Embarked         2  
dtype: int64
```

```
In [73]: df[df.Age.isnull()]
```

Out[73]:

	PassengerId	Survived	Pclass	Name	gender	Age	SibSp	Parch	Ticket	Fare	C
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	N
17	18	1	2	Williams, Mr. Charles Eugene	male	NaN	0	0	244373	13.0000	N
19	20	1	3	Masselmani, Mrs. Fatima	female	NaN	0	0	2649	7.2250	N
26	27	0	3	Emir, Mr. Farred Chehab	male	NaN	0	0	2631	7.2250	N
28	29	1	3	O'Dwyer, Miss. Ellen "Nellie"	female	NaN	0	0	330959	7.8792	N
...
859	860	0	3	Razi, Mr. Raihed	male	NaN	0	0	2629	7.2292	N
863	864	0	3	Sage, Miss. Dorothy Edith "Dolly"	female	NaN	8	2	CA. 2343	69.5500	N
868	869	0	3	van Melkebeke, Mr. Philemon	male	NaN	0	0	345777	9.5000	N
878	879	0	3	Laleff, Mr. Kristo	male	NaN	0	0	349217	7.8958	N
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	N

177 rows × 12 columns

In [74]: `df[~(df.Age.isnull())]`

Out[74]:

	PassengerId	Survived	Pclass	Name	gender	Age	SibSp	Parch	Ticket	Fare	C
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	T
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1	0	PC 17599	71.2833	
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	T
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	T
...
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250	T
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	T
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	T

714 rows × 12 columns



```
In [80]: a=df[df.Age.notnull()]
```

```
In [81]: a.Survived.value_counts()
```

```
Out[81]: 0    424
1    290
Name: Survived, dtype: int64
```

```
In [82]: a.Pclass.value_counts()
```

```
Out[82]: 3    355
          1    186
          2    173
          Name: Pclass, dtype: int64
```

```
In [83]: a[(a.Survived==1)&(a.Pclass==1)].shape
```

```
Out[83]: (122, 12)
```

```
In [84]: a[(a.Survived==0)&(a.Pclass==1)].shape
```

```
Out[84]: (64, 12)
```

```
In [85]: a[(a.Survived==1)&(a.Pclass==2)].shape
```

```
Out[85]: (83, 12)
```

```
In [86]: a[(a.Survived==0)&(a.Pclass==2)].shape
```

```
Out[86]: (90, 12)
```

```
In [87]: a[(a.Survived==1)&(a.Pclass==3)].shape
```

```
Out[87]: (85, 12)
```

```
In [88]: a[(a.Survived==0)&(a.Pclass==3)].shape
```

```
Out[88]: (270, 12)
```

```
In [89]: a.gender.value_counts()
```

```
Out[89]: male      453
          female    261
          Name: gender, dtype: int64
```

```
In [90]: a[(a.Survived==1)&(a.gender=='male')].shape
```

```
Out[90]: (93, 12)
```

```
In [91]: a[(a.Survived==1)&(a.gender=='female')].shape
```

```
Out[91]: (197, 12)
```

```
In [92]: a[(a.Age>=0)&(a.Age<=10)].shape
```

```
Out[92]: (64, 12)
```

```
In [95]: a[(a.SibSp==0)&(a.Parch==0)].shape
```

```
Out[95]: (404, 12)
```

```
In [96]: a[(a.SibSp==0)&(a.Parch==0)&(a.Survived==1)].shape
```

```
Out[96]: (130, 12)
```

```
In [97]: a[(a.Pclass==1)&(a.Survived==1)&(a.gender=='male')].shape
```

```
Out[97]: (40, 12)
```

```
In [98]: a[(a.Pclass==1)&(a.gender=='male')].shape
```

```
Out[98]: (101, 12)
```

```
In [99]: a[(a.Pclass==2)&(a.Survived==1)&(a.gender=='male')].shape
```

```
Out[99]: (15, 12)
```

```
In [101... a[(a.Pclass==2)&(a.gender=='male')].shape
```

```
Out[101]: (99, 12)
```

```
In [102... a[(a.Pclass==3)&(a.gender=='male')].shape
```

```
Out[102]: (253, 12)
```

```
In [103... a[(a.Pclass==3)&(a.Survived==1)&(a.gender=='male')].shape
```

```
Out[103]: (38, 12)
```

```
In [104... a[(a.Pclass==1)&(a.Survived==1)&(a.gender=='female')].shape
```

```
Out[104]: (82, 12)
```

```
In [105... a[(a.Pclass==2)&(a.Survived==1)&(a.gender=='female')].shape
```

```
Out[105]: (68, 12)
```

```
In [106... a[(a.Pclass==3)&(a.Survived==1)&(a.gender=='female')].shape
```

```
Out[106]: (47, 12)
```

```
In [107... a[(a.Pclass==3)&(a.gender=='female')].shape
```

```
Out[107]: (102, 12)
```

```
In [108... a[(a.Pclass==2)&(a.gender=='female')].shape
```

```
Out[108]: (74, 12)
```

```
In [109... a[(a.Pclass==1)&(a.gender=='female')].shape
```

```
Out[109]: (85, 12)
```

```
In [113... a[(a.Survived==1)&(a.gender=='male')].count()
```

```
Out[113]: PassengerId      93  
Survived        93  
Pclass          93  
Name            93  
gender          93  
Age             93  
SibSp           93  
Parch           93  
Ticket          93  
Fare            93  
Cabin           41  
Embarked        93  
dtype: int64
```

```
In [114...]: a[(a.Survived==1)&(a.gender=='female')].count()
```

```
Out[114]: PassengerId      197  
Survived        197  
Pclass          197  
Name            197  
gender          197  
Age             197  
SibSp           197  
Parch           197  
Ticket          197  
Fare            197  
Cabin           84  
Embarked        195  
dtype: int64
```

```
In [ ]:
```