



Mahidol University
Wisdom of the Land



Lecture 15: Pandas

Lecturers:

Aj. Jidapa Kraisangka

Aj. Akara Supratak

Aj. Tipajin Thaipsisutikul



- โมดูลที่สำคัญของ Python
 - pandas
 - numpy





- ถูกสร้างขึ้นเพื่อประมวลผลข้อมูลขนาดใหญ่ (Big Data) ที่เป็นตาราง (Tabular Data)
 - [read_excel](#): Microsoft Excel
 - [read_csv](#): CSV file
 - [read_json](#): JSON file or JSON string
 - [read_sql](#): SQL query or database table
 - ...
- Pandas เก็บและประมวลผลข้อมูลตารางผ่าน [Series](#) หรือ [DataFrame](#)



เราสามารถทำการติดตั้ง pandas ด้วยคำสั่งด้านล่าง

```
pip install pandas
```



สร้างผ่าน dict

```
import pandas as pd
```

```
data = {  
    "calories": [420, 380, 390],  
    "duration": [50, 40, 45]  
}
```

```
df = pd.DataFrame(data)
```

```
print(df)
```



สร้างผ่าน dict

```
import pandas as pd
```

```
data = {  
    "calories": [420, 380, 390],  
    "duration": [50, 40, 45]  
}
```

ต้องมีขนาดเท่ากัน !!

```
df = pd.DataFrame(data)
```

```
print(df)
```



สร้างผ่าน dict

```
import pandas as pd
```

```
data = {
```

```
"calories": [420, 380, 390],
```

```
"duration": [50, 40, 45]
```

```
}
```

ชื่อคอลัมน์

ต้องมีขนาดเท่ากัน !!

```
df = pd.DataFrame(data)
```

```
print(df)
```



สร้างผ่าน dict

```
import pandas as pd
```

```
data = {
```

```
"calories": [420, 380, 390],
```

```
"duration": [50, 40, 45]
```

```
}
```

ชื่อคอลัมน์

ต้องมีขนาดเท่ากัน !!

```
df = pd.DataFrame(data)
```

```
print(df)
```

	calories	duration
0	420	50
1	380	40
2	390	45



อ่านข้อมูลจากไฟล์ Comma-separated Value (CSV)

```
import pandas as pd
```

```
df = pd.read_csv('data.csv')
```

```
print(df)
```

	Unnamed: 0	PassengerId	Survived	Pclass	...	Ticket	Fare	Cabin	Embarked
0	0	1	0	3	...	A/5 21171	7.2500	UNK	S
1	1	2	1	1	...	PC 17599	71.2833	C85	C
2	2	3	1	3	...	STON/O2. 3101282	7.9250	UNK	S
3	3	4	1	1	...	113803	53.1000	C123	S
4	4	5	0	3	...	373450	8.0500	UNK	S
5	5	6	0	3	...	330877	8.4583	UNK	Q
6	6	7	0	1	...	17463	51.8625	E46	S
7	7	8	0	3	...	349909	21.0750	UNK	S



แสดงข้อมูลใน DataFrame

```
import pandas as pd
```

```
df = pd.read_csv('data.csv')
```

```
print(df.head())
```

```
print(df.tail())
```

```
print(df.info())
```



แสดงข้อมูลใน DataFrame

```
import pandas as pd
```

```
df = pd.read_csv('data.csv')
```

```
print(df.head())
```

```
print(df.tail())
```

```
print(df.info())
```

```

    Unnamed: 0  PassengerId  Survived  Pclass  ...      Ticket      Fare  Cabin  Embarked
0            0           1         0       3  ...      A/5 21171    7.2500   UNK        S
1            1           2         1       1  ...      PC 17599   71.2833   C85        C
2            2           3         1       3  ...  STON/O2. 3101282   7.9250   UNK        S
3            3           4         1       1  ...      113803   53.1000  C123        S
4            4           5         0       3  ...      373450    8.0500   UNK        S

```

```
[5 rows x 13 columns]
```

```

    Unnamed: 0  PassengerId  Survived  Pclass  ...      Ticket      Fare  Cabin  Embarked
886          886          887         0       2  ...      211536   13.00   UNK        S
887          887          888         1       1  ...      112053   30.00   B42        S
888          888          889         0       3  ...  W./C. 6607   23.45   UNK        S
889          889          890         1       1  ...      111369   30.00  C148        C
890          890          891         0       3  ...      370376    7.75   UNK        Q

```

```
[5 rows x 13 columns]
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 891 entries, 0 to 890
```

```
Data columns (total 13 columns):
```

```
Unnamed: 0      891 non-null int64
```

```
PassengerId     891 non-null int64
```

```
Survived        891 non-null int64
```

```
Pclass         891 non-null int64
```

```
Name           891 non-null object
```

```
Sex            891 non-null object
```

```
Age            891 non-null float64
```

```
SibSp          891 non-null int64
```

```
Parch          891 non-null int64
```

```
Ticket         891 non-null object
```

```
Fare           891 non-null float64
```

```
Cabin          891 non-null object
```

```
Embarked       891 non-null object
```

```
dtypes: float64(2), int64(6), object(5)
```

```
memory usage: 90.6+ KB
```

```
None
```



แสดงข้อมูลใน DataFrame

```
import pandas as pd
```

```
df = pd.read_csv('data.csv')
```

```
print(df.head())
```

```
print(df.tail())
```

```
print(df.info())
```

```
print(df.describe())
```

```

    Unnamed: 0  PassengerId  Survived  Pclass  ...      Ticket      Fare  Cabin  Embarked
0            0           1         0       3  ...      A/5 21171      7.2500   UNK       S
1            1           2         1       1  ...      PC 17599     71.2833   C85       C
2            2           3         1       3  ...  STON/O2. 3101282    7.9250   UNK       S
3            3           4         1       1  ...      113803     53.1000  C123       S
4            4           5         0       3  ...      373450      8.0500   UNK       S

```

```
[5 rows x 13 columns]
```

```

    Unnamed: 0  PassengerId  Survived  Pclass  ...      Ticket      Fare  Cabin  Embarked
886          886          887         0       2  ...      211536     13.00   UNK       S
887          887          888         1       1  ...      112053     30.00   B42       S
888          888          889         0       3  ...  W./C. 6607     23.45   UNK       S
889          889          890         1       1  ...      111369     30.00  C148       C
890          890          891         0       3  ...      370376      7.75   UNK       Q

```

```
[5 rows x 13 columns]
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 891 entries, 0 to 890
```

```
Data columns (total 13 columns):
```

```
Unnamed: 0      891 non-null int64
```

```
PassengerId     891 non-null int64
```

```
Survived        891 non-null int64
```

```
Pclass         891 non-null int64
```

```
Age            891 non-null float64
```

```
SibSp          891 non-null int64
```

```
Parch          891 non-null int64
```

```
Fare           891 non-null float64
```

```
dtype: object
```

```

count  Unnamed: 0  PassengerId  Survived  Pclass  Age      SibSp  Parch  Fare
mean    891.000000    891.000000    0.383838    2.308642    29.699118    0.523008    0.381594    32.204208
std     257.353842    257.353842    0.486592    0.836071    13.002015    1.102743    0.806057    49.693429
min       0.000000     1.000000    0.000000    0.000000     0.420000    0.000000    0.000000     0.000000
25%     222.500000    223.500000    0.000000    2.000000    22.000000    0.000000    0.000000     7.910400
50%     445.000000    446.000000    0.000000    3.000000    29.699118    0.000000    0.000000    14.454200
75%     667.500000    668.500000    1.000000    3.000000    35.000000    1.000000    0.000000    31.000000
max     890.000000    891.000000    1.000000    3.000000    80.000000    8.000000    6.000000    512.329200

```

```
None
```



การเลือกคอลัมน์ใน DataFrame

```
print(df.columns.values)           # .values เป็นการเปลี่ยนข้อมูลเป็น NumPy
print(df['PassengerId'])           # pd.Series
print(df['PassengerId'].values)    # np.ndarray
```

```
['Unnamed: 0' 'PassengerId' 'Survived' 'Pclass' 'Name' 'Sex' 'Age' 'SibSp'
 'Parch' 'Ticket' 'Fare' 'Cabin' 'Embarked']
0      1
1      2
2      3
3      4
4      5
...
886    887
887    888
888    889
889    890
890    891
Name: PassengerId, Length: 891, dtype: int64
[ 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18
 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36
 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54
 ...
 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864
 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882
 883 884 885 886 887 888 889 890 891]
```



```
import numpy as np
import pandas as pd

df = pd.read_csv('data.csv')
print(df.head(10) ['PassengerId'].values)    # np.ndarray
print(df.iloc[:10] ['PassengerId'])          # pd.Series

# NumPy กับ Pandas ถูกออกแบบมาให้ทำงานร่วมกันได้
print(np.sum(df.head(10) ['PassengerId']))
print(np.sum(df.iloc[:10] ['PassengerId'].values))

# สามารถใช้กับ function อื่นนอกจาก np.sum ได้เหมือนกัน
```



```
import numpy as np
import pandas as pd

df = pd.read_csv('data.csv')
print(df.head(10)['PassengerId'].values) # np.ndarray
print(df.iloc[:10]['PassengerId'])      # pd.Series

# NumPy กับ Pandas ถูกออกแบบมาให้ทำงานร่วมกันได้
print(np.sum(df.head(10)['PassengerId']))
print(np.sum(df.iloc[:10]['PassengerId'].values))

# สามารถใช้กับ function อื่นนอกจาก np.sum ได้เหมือนกัน
```

ใช้ระบุ index หรือทำ slicing เสมือนเป็น 2D array ได้ เช่น
.iloc[4,5]
.iloc[:,3:6]



```
import numpy as np
import pandas as pd

df = pd.read_csv('data.csv')
print(df.head(10))
print(df.iloc[0:10])

# NumPy กับ Pandas
print(np.sum(df['PassengerId']))
print(np.sum(df['PassengerId']))

# สามารถใช้กับ function
```

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

```
Name: PassengerId, dtype: int64
55
55
```




แสดงเฉพาะข้อมูลคอลัมน์ PassengerId กับ Pclass

```
import pandas as pd

df = pd.read_csv('data.csv')
small_df = df.head(10) # df.iloc[:10]
print(small_df[['PassengerId', 'Pclass']])
print(small_df[['PassengerId', 'Pclass']].values)
```

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

[[ 1  3]
 [ 2  1]
 [ 3  3]
 [ 4  1]
 [ 5  3]
 [ 6  3]
 [ 7  1]
 [ 8  3]
 [ 9  3]
[10  2]]
```



แสดงเฉพาะผู้โดยสารที่มี Pclass==1

```
import pandas as pd

df = pd.read_csv('data.csv')
first_class_df = df[df['Pclass']==1]
print(first_class_df)
```

	Unnamed: 0	PassengerId	Survived	Pclass	...	Ticket	Fare	Cabin	Embarked
1	1	2	1	1	...	PC 17599	71.2833	C85	C
3	3	4	1	1	...	113803	53.1000	C123	S
6	6	7	0	1	...	17463	51.8625	E46	S
11	11	12	1	1	...	113783	26.5500	C103	S
23	23	24	1	1	...	113788	35.5000	A6	S
27	27	28	0	1	...	19950	263.0000	C23 C25 C27	S
30	30	31	0	1	...	PC 17601	27.7208	UNK	C
...									



แสดงเฉพาะผู้โดยสารที่มี Sex เป็น 'male'

```
import pandas as pd
```

```
df = pd.read_csv('data.csv')
```



หาค่าเฉลี่ยอายุของผู้โดยสารที่มี Sex เป็น 'male'

```
import pandas as pd
```

```
df = pd.read_csv('data.csv')
```



หาค่าเฉลี่ยอายุของผู้โดยสารผู้ชายและผู้หญิง



เลือกข้อมูลผู้โดยสารที่มีชื่อบางส่วนเป็น James

```
import pandas as pd
import numpy as np

df = pd.read_csv('data.csv')
james_df = df[df['Name'].str.contains('James')]
print(james_df[['PassengerId', 'Name', 'Pclass', 'Embarked']])
```



เลือกข้อมูลผู้โดยสารที่มีชื่อบางส่วนเป็น James

```
import pandas as pd
import numpy as np

df = pd.read_csv('data.csv')
james_df = df[df['Name'].str.contains('James')]
print(james_df[['PassengerId', 'Name', 'Pclass', 'Embarked']])
```

	PassengerId	Name	Pclass	Embarked
5	6	Moran, Mr. James	3	Q
67	68	Crease, Mr. Ernest James	3	S
134	135	Sobey, Mr. Samuel James Hayden	2	S
150	151	Bateman, Rev. Robert James	2	S
161	162	Watt, Mrs. James (Elizabeth "Bessie" Inglis Mi...	2	S
174	175	Smith, Mr. James Clinch	1	C
194	195	Brown, Mrs. James Joseph (Margaret Tobin)	1	C
221	222	Bracken, Mr. James H	2	S
250	251	Reed, Mr. James George	3	S
299	300	Baxter, Mrs. James (Helene DeLaudeniére Chaput)	1	C
416	417	Drew, Mrs. James Vivian (Lulu Thorne Christian)	2	S
428	429	Flynn, Mr. James	3	Q
468	469	Scanlan, Mr. James	3	Q
511	512	Webber, Mr. James	3	S
512	513	McGough, Mr. James Robert	1	S
525	526	Farrell, Mr. James	3	Q
551	552	Sharp, Mr. Percival James R	2	S
582	583	Downton, Mr. William James	2	S
592	593	Elsbury, Mr. William James	3	S
696	697	Kelly, Mr. James	3	S
740	741	Hawksford, Mr. Walter James	1	S
811	812	Lester, Mr. James	3	S
812	813	Slemen, Mr. Richard James	2	S
883	884	Banfield, Mr. Frederick James	2	S



เลือกข้อมูลผู้โดยสารที่มีชื่อบางส่วนเป็น James

มีชื่อ James เป็นส่วนหนึ่งก็ได้

```
import pandas as pd
import numpy as np

df = pd.read_csv('data.csv')
james_df = df[df['Name'].str.contains('James')]
print(james_df[['PassengerId', 'Name', 'Pclass', 'Embarked']])
```

PassengerId	Name	Pclass	Embarked
5	Moran, Mr. James	3	Q
67	Crease, Mr. Ernest James	3	S
134	Sobey, Mr. Samuel James Hayden	2	S
150	Bateman, Rev. Robert James	2	S
161	Watt, Mrs. James (Elizabeth "Bessie" Inglis Mi...	2	S
174	Smith, Mr. James Clinch	1	C
194	Brown, Mrs. James Joseph (Margaret Tobin)	1	C
221	Bracken, Mr. James H	2	S
250	Reed, Mr. James George	3	S
299	Baxter, Mrs. James (Helene DeLaudeniére Chaput)	1	C
416	Drew, Mrs. James Vivian (Lulu Thorne Christian)	2	S
428	Flynn, Mr. James	3	Q
468	Scanlan, Mr. James	3	Q
511	Webber, Mr. James	3	S
512	McGough, Mr. James Robert	1	S
525	Farrell, Mr. James	3	Q
551	Sharp, Mr. Percival James F	2	S
582	Downton, Mr. William James	2	S
592	Elsbury, Mr. William James	3	S
696	Kelly, Mr. James	3	S
740	Hawksford, Mr. Walter James	1	S
811	Lester, Mr. James	3	S
812	Slemen, Mr. Richard James	2	S
883	Banfield, Mr. Frederick James	2	S



เลือกข้อมูลผู้โดยสารที่ขึ้นเรือจากท่า 'Q' หรือ 'S'

```
import pandas as pd
import numpy as np

df = pd.read_csv('data.csv')
embark_df = df[df['Embarked'].isin(['Q','S'])]
print(embark_df[['PassengerId', 'Name', 'Pclass', 'Embarked']])
```



เลือกข้อมูลผู้โดยสารที่ขึ้นเรือจากท่า 'Q' หรือ 'S'

```
import pandas as pd
```

```
import numpy as np
```

```
df = pd.read_csv('data.csv')
```

```
embark_df = df[df['Embarked'].isin(['Q', 'S'])]
```

```
print(embark_df[['PassengerId', 'Name', 'Pclass', 'Embarked']])
```

ต้องเป็น list หรือ numpy array ของค่า



เลือกข้อมูลผู้โดยสารที่ขึ้นเรือจากท่า 'Q' หรือ 'S'

```
import pandas as pd  
import numpy as np
```

```
df = pd.read_csv('data.csv')  
embark_df = df[df['Embarked']  
print(embark_df[['PassengerId',
```

```
PassengerId      Name  Pclass Embarked  
0                1  Braund, Mr. Owen Harris    3         S  
2                3  Heikkinen, Miss. Laina    3         S  
3                4 Futrelle, Mrs. Jacques Heath (Lily May Peel)    1         S  
4                5  Allen, Mr. William Henry    3         S  
5                6  Moran, Mr. James    3         Q  
..            ...            ...            ...  
885            886  Rice, Mrs. William (Margaret Norton)    3         Q  
886            887  Montvila, Rev. Juozas    2         S  
887            888  Graham, Miss. Margaret Edith    1         S  
888            889  Johnston, Miss. Catherine Helen "Carrie"    3         S  
890            891  Dooley, Mr. Patrick    3         Q  
  
[721 rows x 4 columns]
```



เลือกข้อมูลผู้โดยสารที่ขึ้นเรือจากท่า 'Q' หรือ 'S'

```
import pandas as pd  
import numpy as np
```

มีค่าเท่ากับ 'Q' หรือ 'S' เท่านั้น

```
df = pd.read_csv('data.csv')  
embark_df = df[df['Embarked']  
print(embark_df[['PassengerId',
```

	PassengerId	Name	Pclass	Embarked
0	1	Braund, Mr. Owen Harris	3	S
2	3	Heikkinen, Miss. Laina	3	S
3	4	Futrelle, Mrs. Jacques Heath (Lily May Peel)	1	S
4	5	Allen, Mr. William Henry	3	S
5	6	Moran, Mr. James	3	Q
..
885	886	Rice, Mrs. William (Margaret Norton)	3	Q
886	887	Montvila, Rev. Juozas	2	S
887	888	Graham, Miss. Margaret Edith	1	S
888	889	Johnston, Miss. Catherine Helen "Carrie"	3	S
890	891	Dooley, Mr. Patrick	3	Q

[721 rows x 4 columns]



ไม่ควรใช้ `isin` ถ้าเราจะอนุญาตให้เหมือนกันแค่บางส่วนของ string (substring)

```
import pandas as pd
```

```
import numpy as np
```

```
df = pd.read_csv('data.csv')
```

```
james_df = df[df['Name'].isin(['James'])]
```

```
print(james_df [['PassengerId', 'Name', 'Pclass', 'Embarked']])
```



ไม่ควรใช้ `isin` ถ้าเราจะอนุญาตให้เหมือนกันแค่บางส่วนของ string (substring)

```
import pandas as pd
import numpy as np

df = pd.read_csv('data.csv')
james_df = df[df['Name'].isin(['James'])]
print(james_df[['PassengerId', 'Name', 'Pclass', 'Embarked']])
```

ไม่มีผู้โดยสารที่มีชื่อตรงกับ 'James'

```
Empty DataFrame
Columns: [PassengerId, Name, Pclass, Embarked]
Index: []
```



หาค่าเฉลี่ยค่าโดยสาร (Fare) ของผู้โดยสารที่มีชื่อบางส่วนเป็น William

```
import pandas as pd
```

```
df = pd.read_csv('data.csv')
```



นับจำนวนของค่าในคอลัมน์ : `value_counts()`

```
import pandas as pd
```

```
df = pd.read_csv('data.csv')
```

```
print(df['Embarked'].value_counts()) # pd.Series
```

```
print(df['Embarked'].value_counts().to_dict()) # dict
```

```
S      644
C      168
Q       77
UNK       2
Name: Embarked, dtype: int64
{'S': 644, 'C': 168, 'Q': 77, 'UNK': 2}
```




หาค่าที่ไม่ซ้ำกันในแต่ละคอลัมน์ : `unique()`

```
import pandas as pd

df = pd.read_csv('data.csv')
print(df['Pclass'].unique())      # numpy array
print(df['Embarked'].unique())    # numpy array

# ระวัง: ไม่ควรใช้กับตัวเลขจำนวนจริง (ที่จะมีค่าแตกต่างกันเยอะมาก) เช่น ค่าโดยสาร
# print(df['Fare'].unique())
```

```
[3 1 2]
['S' 'C' 'Q' 'UNK']
```



เรียงลำดับโดยดูจากค่าในคอลัมน์ต่าง ๆ : `sort_values()`

```
import pandas as pd

df = pd.read_csv('data.csv')
print(df.sort_values(by=['Pclass'])) # ค่าตั้งต้นมาตรฐาน ascending=True (เรียงจากน้อยไปมาก)
print(df.sort_values(by=['Pclass'], ascending=False)) # เรียงจากมากไปน้อย
```



เรียงลำดับโดยดูจากค่าในคอลัมน์ต่าง ๆ : `sort_values()`

```
import pandas as pd

df = pd.read_csv('data.csv')
print(df.sort_values(by=['Pclass'])) # ค่าตั้งต้นมาตรฐาน ascending=True (เรียงจากน้อยไปมาก)
print(df.sort_values(by=['Pclass'], ascending=False)) # เรียงจากมากไปน้อย

# เรียงด้วย Pclass จากมากไปน้อยตามด้วย Fare จากมากไปน้อย
print(df.sort_values(by=['Pclass', 'Fare'], ascending=False)[['Pclass', 'Fare']])

# เรียงด้วย Pclass จากน้อยไปมากตามด้วย Fare จากมากไปน้อย
print(df.sort_values(by=['Pclass', 'Fare'], ascending=[True, False])[['Pclass', 'Fare']])
```



เราสามารถรวมเงื่อนไขต่าง ๆ ได้โดยใช้

- `&` : and
- `|` : or
- `~` : not



แสดงผู้โดยสารผู้ชายที่ซื้อตั๋วโดยสารถูกกว่า 15.25

```
import pandas as pd

df = pd.read_csv('data.csv')
print(df[(df['Sex']=='male') & (df['Fare']<15.25)])
```



แสดงผู้โดยสารผู้ชายที่ซื้อตั๋วโดยสารถูกกว่า 15.25

```
import pandas as pd
```

ต้องมีวงเล็บ () ครอบเงื่อนไขต่าง ๆ เวลาที่ถูกเชื่อมด้วย &, | หรือ ~ ด้วย

```
df = pd.read_csv('data.csv')
```

```
print(df[(df['Sex']=='male') & (df['Fare']<15.25)])
```

ผู้โดยสารผู้ชาย

ค่าตั๋วถูกกว่า 15.25

	Unnamed: 0	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	0	1	0	3	Braund, Mr. Owen Harris	male	22.000000	1	0	A/5 2117	7.2500	UNK	S
4	4	5	0	3	Allen, Mr. William Henry	male	35.000000	0	0	37345	8.0500	UNK	S
5	5	6	0	3	Moran, Mr. James	male	29.699118	0	0	33087	8.4583	UNK	Q
12	12	13	0	3	Saunders, Mr. William Henry	male	20.000000	0	0	A/5. 215	8.0500	UNK	S
17	17	18	1	2	Williams, Mr. Charles Eugene	male	29.699118	0	0	24437	13.0000	UNK	S
...
881	881	882	0	3	Markun, Mr. Johann	male	33.000000	0	0	34925	7.8958	UNK	S
883	883	884	0	2	Banfield, Mr. Frederick James	male	28.000000	0	0	C.A./SOTON 3406	10.5000	UNK	S
884	884	885	0	3	Sutehall, Mr. Henry Jr	male	25.000000	0	0	SOTON/OQ 39207	7.0500	UNK	S
886	886	887	0	2	Montvila, Rev. Juozas	male	27.000000	0	0	21153	13.0000	UNK	S
890	890	891	0	3	Dooley, Mr. Patrick	male	32.000000	0	0	37037	7.7500	UNK	Q

[349 rows x 13 columns]



แสดงผู้โดยสารผู้ชายที่ซื้อตั๋วโดยสารถูกกว่า 15.25

```
import pandas as pd

df = pd.read_csv('data.csv')
sel_df = df[(df['Sex']=='male') & (df['Fare']<15.25)]

# เรียงตามราคาตั๋ว
sel_df = sel_df.sort_values(by=['Fare'], ascending=False)

# แสดงผลเฉพาะคอลัมน์ 'Name', 'Sex', 'Fare'
print(sel_df[['Name', 'Sex', 'Fare']])
```



แสดงผู้โดยสารผู้ชายที่ซื้อตั๋วโดยสารถูกกว่า 15.25

```
import pandas as pd

df = pd.read_csv('data.csv')
sel_df = df[(df['Sex']=='male') & (df['Fare']<15.25)]

# เรียงตามราคาตั๋ว
sel_df = sel_df.sort_values(by=['Fare'], ascending=False)

# แสดงผลเฉพาะคอลัมน์ 'Name', 'Sex', 'Fare'
print(sel_df[['Name', 'Sex', 'Fare']])
```

```
      Name      Sex      Fare
65      Moubarek, Master. Gerios  male  15.2458
709  Moubarek, Master. Halim Gonios ("William George")  male  15.2458
497  Shellard, Mr. Frederick William  male  15.1000
181  Pernot, Mr. Rene  male  15.0500
135  Richard, Mr. Emile  male  15.0458
..      ...      ...
263  Harrison, Mr. William  male  0.0000
271  Tornquist, Mr. William Henry  male  0.0000
277  Parkes, Mr. Francis "Frank"  male  0.0000
302  Johnson, Mr. William Cahoon Jr  male  0.0000
732  Knight, Mr. Robert J  male  0.0000
```

[349 rows x 3 columns]



จำนวนผู้โดยสารผู้ชายและผู้หญิงที่ไม่ได้ขึ้นเรือจากท่า 'S'

```
import pandas as pd

df = pd.read_csv('data.csv')
sel_df = df[~df['Embarked'].isin(['S'])]
```



จำนวนผู้โดยสารผู้ชายและผู้หญิงที่ไม่ได้ขึ้นเรือจากท่า 'S'

```
import pandas as pd

df = pd.read_csv('data.csv')
sel_df = df[~df['Embarked'].isin(['S'])]
print(sel_df)
print(sel_df['Sex'].value counts())
```

Unnamed: 0	PassengerId	Survived	Pclass	Name	Sex	...	SibSp	Parch	Ticket	Fare	Cabin
Embarked											
1	1	2	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	...	1	0	PC 17599	71.2833	C85
5	5	6	0	Moran, Mr. James	male	...	0	0	330877	8.4583	UNK
9	9	10	1	Nasser, Mrs. Nicholas (Adele Achem)	female	...	1	0	237736	30.0708	UNK
16	16	17	0	Rice, Master. Eugene	male	...	4	1	382652	29.1250	UNK
19	19	20	1	Masselmani, Mrs. Fatima	female	...	0	0	2649	7.2250	UNK
...
875	875	876	1	Najib, Miss. Adele Kiamie "Jane"	female	...	0	0	2667	7.2250	UNK
879	879	880	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	...	0	1	11767	83.1583	C50