```
import pandas as pd
```

Python pandas library cheat sheet

import pandas library

```
import pandas as pd
```

read the csv file in pandas library

```
df=pd.read_csv("https://github.com/YBI-Foundation/Dataset/raw/main/Car%20Crashes.csv")
```

use .head function to show the 5 row of dataset

```
df.head()
```

	total	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses	abbrev
0	18.8	7.332	5.640	18.048	15.040	784.55	145.08	AL
1	18.1	7.421	4.525	16.290	17.014	1053.48	133.93	AK
2	18.6	6.510	5.208	15.624	17.856	899.47	110.35	AZ
3	22.4	4.032	5.824	21.056	21.280	827.34	142.39	AR
4	12.0	4.200	3.360	10.920	10.680	878.41	165.63	CA

✓ 0s completed at 1:35 PM

use tail to show last 5 row of dataset

```
df.tail()
```

	total	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses	abbrev
46	12.7	2.413	3.429	11.049	11.176	768.95	153.72	VA
47	10.6	4.452	3.498	8.692	9.116	890.03	111.62	WA
48	23.8	8.092	6.664	23.086	20.706	992.61	152.56	WV
49	13.8	4.968	4.554	5.382	11.592	670.31	106.62	WI
50	17.4	7.308	5.568	14.094	15.660	791.14	122.04	WY

to show the infor mation of data set it use the info function

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
Data columns (total 8 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   total           51 non-null     float64
 1   speeding        51 non-null     float64
 2   alcohol         51 non-null     float64
 3   not_distracted  51 non-null     float64
 4   no_previous     51 non-null     float64
 5   ins_premium     51 non-null     float64
 6   ins_losses      51 non-null     float64
```

```
 7   abbrev          51 non-null     object
dtypes: float64(7), object(1)
memory usage: 3.3+ KB
```

use decribe to describe the data set for count, mean,standered davision,max etc

```
df.describe()
```

	total	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses
count	51.000000	51.000000	51.000000	51.000000	51.000000	51.000000	51.000000
mean	15.790196	4.998196	4.886784	13.573176	14.004882	886.957647	134.493137
std	4.122002	2.017747	1.729133	4.508977	3.764672	178.296285	24.835922
min	5.900000	1.792000	1.593000	1.760000	5.900000	641.960000	82.750000
25%	12.750000	3.766500	3.894000	10.478000	11.348000	768.430000	114.645000
50%	15.600000	4.608000	4.554000	13.857000	13.775000	858.970000	136.050000
75%	18.500000	6.439000	5.604000	16.140000	16.755000	1007.945000	151.870000
max	23.900000	9.450000	10.038000	23.661000	21.280000	1301.520000	194.780000

to show the shap of the dataset use shape function

```
df.shape
```

```
(51, 8)
```

to show the column of the data set

```
df.columns
```

```
Index(['total', 'speeding', 'alcohol', 'not_distracted', 'no_previous',
       'ins_premium', 'ins_losses', 'abbrev'],
      dtype='object')
```

to show the particular columns

```
df.columns[[1,2]]
```

```
Index(['speeding', 'alcohol'], dtype='object')
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
Data columns (total 8 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   total           51 non-null     float64
 1   speeding        51 non-null     float64
 2   alcohol         51 non-null     float64
 3   not_distracted  51 non-null     float64
 4   no_previous     51 non-null     float64
 5   ins_premium     51 non-null     float64
 6   ins_losses      51 non-null     float64
 7   abbrev          51 non-null     object 
dtypes: float64(7), object(1)
memory usage: 3.3+ KB
```

	total	speeding	alcohol	not_distracted	no_previous	ins_premium	ins_losses
count	51.000000	51.000000	51.000000	51.000000	51.000000	51.000000	51.000000
mean	15.790196	4.998196	4.886784	13.573176	14.004882	886.957647	134.493137
std	4.122002	2.017747	1.729133	4.508977	3.764672	178.296285	24.835922
min	5.900000	1.792000	1.593000	1.760000	5.900000	641.960000	82.750000
25%	12.750000	3.766500	3.894000	10.478000	11.348000	768.430000	114.645000
50%	15.600000	4.608000	4.554000	13.857000	13.775000	858.970000	136.050000
75%	18.500000	6.439000	5.604000	16.140000	16.755000	1007.945000	151.870000
max	23.900000	9.450000	10.038000	23.661000	21.280000	1301.520000	194.780000

```
df.shape
```

```
(51, 8)
```

```
df.columns
```

```
Index(['total', 'speeding', 'alcohol', 'not_distracted', 'no_previous',
       'ins_premium', 'ins_losses', 'abbrev'],
      dtype='object')
```

data manuplation

```
import pandas as pd
```

```
df=pd.read_csv('https://github.com/YBI-Foundation/Dataset/raw/main/Titanic.csv')
```

```
df.describe()
```

	pclass	survived	age	sibsp	parch	fare	body
count	1309.000000	1309.000000	1046.000000	1309.000000	1309.000000	1308.000000	121.000000
mean	2.294882	0.381971	29.881138	0.498854	0.385027	33.295479	160.809917
std	0.837836	0.486055	14.413493	1.041658	0.865560	51.758668	97.696922
min	1.000000	0.000000	0.170000	0.000000	0.000000	0.000000	1.000000
25%	2.000000	0.000000	21.000000	0.000000	0.000000	7.895800	72.000000
50%	3.000000	0.000000	28.000000	0.000000	0.000000	14.454200	155.000000
75%	3.000000	1.000000	39.000000	1.000000	0.000000	31.275000	256.000000
max	3.000000	1.000000	80.000000	8.000000	9.000000	512.329200	328.000000

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1309 entries, 0 to 1308
Data columns (total 14 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   pclass     1309 non-null   int64
 1   survived   1309 non-null   int64
 2   name       1309 non-null   object
 3   sex        1309 non-null   object
 4   age        1046 non-null   float64
 5   sibsp      1309 non-null   int64
 6   parch      1309 non-null   int64
 7   ticket     1309 non-null   object
 8   fare       1308 non-null   float64
 9   cabin      295 non-null    object
 10  embarked   1307 non-null   object
 11  boat       486 non-null    object
 12  body       121 non-null    float64
 13  home.dest  745 non-null    object
```

```
 13  home.dest  745 non-null    object
dtypes: float64(3), int64(4), object(7)
memory usage: 143.3+ KB
```

```
df.head(9)
```

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
0	1	1	Allen, Miss. Elisabeth Walton	female	29.00	0	0	24160	211.3375	B5	S	2	NaN	St Louis, MO
1	1	1	Allison, Master. Hudson Trevor	male	0.92	1	2	113781	151.5500	C22 C26	S	11	NaN	Montreal, PQ / Chesterville, ON
2	1	0	Allison, Miss. Helen Loraine	female	2.00	1	2	113781	151.5500	C22 C26	S	NaN	NaN	Montreal, PQ / Chesterville, ON
			Allison, Mr.											Montreal, PQ /

```
df.describe(include= 'all')
```

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	eml
count	1309.000000	1309.000000	1309	1309	1046.000000	1309.000000	1309.000000	1309	1308.000000	295	
unique	NaN	NaN	1307	2	NaN	NaN	NaN	929	NaN	186	
top	NaN	NaN	Connolly, Miss. Kate	male	NaN	NaN	NaN	CA. 2343	NaN	C23 C25 C27	
freq	NaN	NaN	2	843	NaN	NaN	NaN	11	NaN	6	
mean	2.294882	0.381971	NaN	NaN	29.881138	0.498854	0.385027	NaN	33.295479	NaN	

std	0.837836	0.486055	NaN	NaN	14.413493	1.041658	0.865560	NaN	51.758668	NaN
min	1.000000	0.000000	NaN	NaN	0.170000	0.000000	0.000000	NaN	0.000000	NaN
25%	2.000000	0.000000	NaN	NaN	21.000000	0.000000	0.000000	NaN	7.895800	NaN
50%	3.000000	0.000000	NaN	NaN	28.000000	0.000000	0.000000	NaN	14.454200	NaN
75%	3.000000	1.000000	NaN	NaN	39.000000	1.000000	0.000000	NaN	31.275000	NaN

to find the nth largest age of the data set

```
df['age'].nlargest(3)
```

```
14      80.0
61      76.0
1235    74.0
Name: age, dtype: float64
```

to find the nth smallest age of the data set

```
df['age'].nsmallest(5)
```

```
763     0.17
747     0.33
1240    0.42
427     0.67
657     0.75
Name: age, dtype: float64
```

find max age of data set

Double-click (or enter) to edit

```
df['age'].max()
```

```
80.0
```

```
df['age'].min()
```

```
0.17
```

change data type

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1309 entries, 0 to 1308
Data columns (total 14 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   pclass     1309 non-null   int64  
 1   survived   1309 non-null   int64  
 2   name       1309 non-null   object 
 3   sex        1309 non-null   object 
 4   age        1046 non-null   float64
 5   sibsp      1309 non-null   int64  
 6   parch      1309 non-null   int64  
 7   ticket     1309 non-null   object 
 8   fare       1308 non-null   float64
 9   cabin      295 non-null    object 
 10  embarked   1307 non-null   object 
 11  boat       486 non-null    object
```

```
 12  body       121 non-null    float64
 13  home.dest  745 non-null    object
dtypes: float64(3), int64(4), object(7)
memory usage: 143.3+ KB
```

```
df['pclass']=df['pclass'].astype(object)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1309 entries, 0 to 1308
Data columns (total 14 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   pclass     1309 non-null   object
 1   survived   1309 non-null   int64
 2   name       1309 non-null   object
 3   sex        1309 non-null   object
 4   age        1046 non-null   float64
 5   sibsp      1309 non-null   int64
 6   parch      1309 non-null   int64
 7   ticket     1309 non-null   object
 8   fare       1308 non-null   float64
 9   cabin      295 non-null    object
 10  embarked   1307 non-null   object
 11  boat       486 non-null    object
 12  body       121 non-null    float64
 13  home.dest  745 non-null    object
dtypes: float64(3), int64(3), object(8)
memory usage: 143.3+ KB
```

```
df['survived']=df['survived'].astype(float)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1309 entries, 0 to 1308
Data columns (total 14 columns):
```

```
Data columns (total 14 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   pclass     1309 non-null   object
 1   survived   1309 non-null   float64
 2   name       1309 non-null   object
 3   sex        1309 non-null   object
 4   age        1046 non-null   float64
 5   sibsp      1309 non-null   int64
 6   parch      1309 non-null   int64
 7   ticket     1309 non-null   object
 8   fare       1308 non-null   float64
 9   cabin      295 non-null    object
 10  embarked   1307 non-null   object
 11  boat       486 non-null    object
 12  body       121 non-null    float64
 13  home.dest  745 non-null    object
dtypes: float64(4), int64(2), object(8)
memory usage: 143.3+ KB
```

find the corelation of the data set

```
df.corr()
```

	survived	age	sibsp	parch	fare	body
survived	1.000000	-0.055512	-0.027825	0.082660	0.244265	NaN
age	-0.055512	1.000000	-0.243699	-0.150917	0.178740	0.058809
sibsp	-0.027825	-0.243699	1.000000	0.373587	0.160238	-0.099961
parch	0.082660	-0.150917	0.373587	1.000000	0.221539	0.051099
fare	0.244265	0.178740	0.160238	0.221539	1.000000	-0.043110
body	NaN	0.058809	-0.099961	0.051099	-0.043110	1.000000

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

```
array(['S', 'C', nan, 'Q'], dtype=object)
```

```
df['embarked'].value_counts()
```

```
S    914
C    270
Q    123
Name: embarked, dtype: int64
```

```
df['sex'].value_counts()
```

```
male      843
female    466
Name: sex, dtype: int64
```

```
df.replace({'sex':{'male':0,'female':1}},inplace=True)
```

```
df.head()
```

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.dest
0	1	1.0	Allen, Miss. Elisabeth Walton	1	29.00	0	0	24160	211.3375	B5	S	2	NaN	St Louis, MO
1	1	1.0	Allison, Master. Hudson Trevor	0	0.92	1	2	113781	151.5500	C22 C26	S	11	NaN	Montreal, PQ / Chesterville, ON
2	1	0.0	Allison, Miss. Helen	1	2.00	1	2	113781	151.5500	C22	S	NaN	NaN	Montreal, PQ /

Loraine C26 Chesterville, ON

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1309 entries, 0 to 1308
Data columns (total 14 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   pclass     1309 non-null   object 
 1   survived   1309 non-null   float64
 2   name       1309 non-null   object 
 3   sex        1309 non-null   object 
 4   age        1046 non-null   float64
 5   sibsp      1309 non-null   int64  
 6   parch      1309 non-null   int64  
 7   ticket     1309 non-null   object 
 8   fare       1308 non-null   float64
 9   cabin      295 non-null    object 
 10  embarked   1307 non-null   object 
 11  boat       486 non-null    object 
 12  body       121 non-null    float64
 13  home.dest  745 non-null    object 
dtypes: float64(4), int64(2), object(8)
memory usage: 143.3+ KB
```

```
df.describe()
```

	survived	age	sibsp	parch	fare	body
count	1309.000000	1046.000000	1309.000000	1309.000000	1308.000000	121.000000
mean	0.381971	29.881138	0.498854	0.385027	33.295479	160.809917
std	0.486055	14.413493	1.041658	0.865560	51.758668	97.696922
min	0.000000	0.170000	0.000000	0.000000	0.000000	1.000000
25%	0.000000	21.000000	0.000000	0.000000	7.895800	72.000000

50%	0.000000	28.000000	0.000000	0.000000	14.454200	155.000000
75%	1.000000	39.000000	1.000000	0.000000	31.275000	256.000000
max	1.000000	80.000000	8.000000	9.000000	512.329200	328.000000

```
df.loc[3,['age','fare']]
```

```
age       30.0
fare    151.55
Name: 3, dtype: object
```

```
import pandas as pd
```

```
df=pd.read_csv('https://github.com/YBI-Foundation/Dataset/raw/main/Titanic.csv')
```

```
df.iloc[3,4]
```

```
30.0
```

```
df[(df['age']>70)]
```

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.d
9	1	0	Artagaveytia, Mr. Ramon	male	71.0	0	0	PC 17609	49.5042	NaN	C	NaN	22.0	Montevid Urug
14	1	1	Barkworth, Mr. Algernon Henry Wilson	male	80.0	0	0	27042	30.0000	A23	S	B	NaN	Hes Yc
61	1	1	Cavendish, Mrs. Tyrell William (Julia	female	76.0	1	0	19877	78.8500	C46	S	6	NaN	Little C Hall, St

			Florence...											
135	1	0	Goldschmidt, Mr. George	male	71.0	0	0	PC	34.6542	A5	C	NaN	NaN	New Y

```
df.loc[10:25,['pclass','fare','age']]
```

	pclass	fare	age
10	1	227.5250	47.0
11	1	227.5250	18.0
12	1	69.3000	24.0
13	1	78.8500	26.0
14	1	30.0000	80.0
15	1	25.9250	NaN
16	1	247.5208	24.0
17	1	247.5208	50.0
18	1	76.2917	32.0
19	1	75.2417	36.0
20	1	52.5542	37.0
21	1	52.5542	47.0
22	1	30.0000	26.0
23	1	227.5250	42.0
24	1	221.7792	29.0
25	1	26.0000	25.0

```
df.iloc[10:25,[0,8,4]]
```

	pclass	fare	age
10	1	227.5250	47.0
11	1	227.5250	18.0
12	1	69.3000	24.0
13	1	78.8500	26.0
14	1	30.0000	80.0
15	1	25.9250	NaN
16	1	247.5208	24.0
17	1	247.5208	50.0
18	1	76.2917	32.0
19	1	75.2417	36.0
20	1	52.5542	37.0
21	1	52.5542	47.0
22	1	30.0000	26.0
23	1	227.5250	42.0
24	1	221.7792	29.0

```
df.loc[(df['age']>=35),'pclass':'age']
```

	pclass	survived	name	sex	age
5	1	1	Anderson, Mr. Harry	male	48.0
6	1	1	Andrews, Miss. Kornelia Theodosia	female	63.0
7	1	0	Andrews, Mr. Thomas Jr	male	39.0
8	1	1	Appleton, Mrs. Edward Dale (Charlotte Lamson)	female	53.0

9	1	0	Artagaveytia, Mr. Ramon	male	71.0
...	...	...	...	...	...
1286	3	1	Whabee, Mrs. George Joseph (Shawneene Abi-Saab)	female	38.0
1287	3	0	Widegren, Mr. Carl/Charles Peter	male	51.0
1290	3	1	Wilkes, Mrs. James (Ellen Needs)	female	47.0
1298	3	0	Wittevrongel, Mr. Camille	male	36.0
1301	3	0	Youseff, Mr. Gerious	male	45.5

345 rows × 5 columns

```
df.loc[ (df['age']>=35)&(df['sex']=='female' )]
```

	pclass	survived	name	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	home.d
6	1	1	Andrews, Miss. Kornelia Theodosia	female	63.0	1	0	13502	77.9583	D7	S	10	NaN	Huds
8	1	1	Appleton, Mrs. Edward Dale (Charlotte Lamson)	female	53.0	2	0	11769	51.4792	C101	S	D	NaN	Bays Quee
17	1	1	Baxter, Mrs. James (Helene DeLaudeniere Chaput)	female	50.0	0	1	PC 17558	247.5208	B58 B60	C	6	NaN	Montr
21	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	1	11751	52.5542	D35	S	5	NaN	New Y

Bidois, Miss PC

first import the file

```
import pandas as pd
```

```
tips=pd.read_csv('https://github.com/YBI-Foundation/Dataset/raw/main/Tips%20Payment%20Data.csv')
```

display first five row of the data set

```
tips.head()
```

	Total Bill	Tip	Gender	Smoker	Day	Time	Size	Bill Per Person	Payer Name	CC Number	Payment ID
0	16.99	1.01	Female	No	Sun	Dinner	2	8.49	Christy Cunningham	3560325168603410	Sun2959
1	10.34	1.66	Male	No	Sun	Dinner	3	3.45	Douglas Tucker	4478071379779230	Sun4608
2	21.01	3.50	Male	No	Sun	Dinner	3	7.00	Travis Walters	6011812112971320	Sun4458
3	23.68	3.31	Male	No	Sun	Dinner	2	11.84	Nathaniel Harris	4676137647685990	Sun5260

calculate the persentage of tip to total bill

```
tips['Tip']/tips['Total Bill']*100
```

```
0       5.944673
1      16.054159
```

```
2      16.658734
3      13.978041
4      14.680765
         ...
239    20.392697
240     7.358352
241     8.822232
242     9.820426
243    15.974441
Length: 244, dtype: float64
```

creat a new column of the persentage tip

```
tip_persentage=tips['Tip']/tips['Total Bill']*100
```

```
tip_persentage
```

```
0       5.944673
1      16.054159
2      16.658734
3      13.978041
4      14.680765
         ...
239    20.392697
240     7.358352
241     8.822232
242     9.820426
243    15.974441
Length: 244, dtype: float64
```

insert persentage tip in existing tips datafram

round upto one decemal place the tip persentage column value

```
tips['tip_persentage']=tips['Tip']/tips['Total Bill']*100
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: PerformanceWarning: DataFrame is highly fragmented.
  """Entry point for launching an IPython kernel.
```

```
tips.head()
```

	Total Bill	Tip	Gender	Smoker	Day	Time	Size	Bill Per Person	Payer Name	CC Number	...	12.413108242303872	7.9
0	16.99	1.01	Female	No	Sun	Dinner	2	8.49	Christy Cunningham	3560325168603410	...	12.413108	
1	10.34	1.66	Male	No	Sun	Dinner	3	3.45	Douglas Tucker	4478071379779230	...	12.413108	
2	21.01	3.50	Male	No	Sun	Dinner	3	7.00	Travis Walters	6011812112971320	...	12.413108	
3	23.68	3.31	Male	No	Sun	Dinner	2	11.84	Nathaniel Harris	4676137647685990	...	12.413108	
4	24.59	3.61	Female	No	Sun	Dinner	4	6.15	Tonya Carter	4832732618637220	...	12.413108	

5 rows × 253 columns

Double-click (or enter) to edit

round upto one decemal place the tip persentage column

```
tips['tip persentage']=tips['tip persentage'].round(1)
```

```
tips.head()
```

	Total Bill	Tip	Gender	Smoker	Day	Time	Size	Bill Per Person	Payer Name	CC Number	...	12.413108242303872	7.9
0	16.99	1.01	Female	No	Sun	Dinner	2	8.49	Christy Cunningham	3560325168603410	...	12.413108	
1	10.34	1.66	Male	No	Sun	Dinner	3	3.45	Douglas Tucker	4478071379779230	...	12.413108	
2	21.01	3.50	Male	No	Sun	Dinner	3	7.00	Travis Walters	6011812112971320	...	12.413108	
3	23.68	3.31	Male	No	Sun	Dinner	2	11.84	Nathaniel Harris	4676137647685990	...	12.413108	
4	24.59	3.61	Female	No	Sun	Dinner	4	6.15	Tonya Carter	4832732618637220	...	12.413108	

5 rows × 253 columns

Drop column payer number

```
tips=tips.drop(['Payer Name'],axis=1)
```

```
tips.head()
```

	Total Bill	Tip	Gender	Smoker	Day	Time	Size	Bill Per Person	CC Number	Payment ID	...	12.413108242303872	7.9365
0	16.99	1.01	Female	No	Sun	Dinner	2	8.49	3560325168603410	Sun2959		12.413108	

1	10.34	1.66	Male	No	Sun	Dinner	3	3.45	4478071379779230	Sun4608	...	12.413108
2	21.01	3.50	Male	No	Sun	Dinner	3	7.00	6011812112971320	Sun4458	...	12.413108
3	23.68	3.31	Male	No	Sun	Dinner	2	11.84	4676137647685990	Sun5260	...	12.413108
4	24.59	3.61	Female	No	Sun	Dinner	4	6.15	4832732618637220	Sun2251	...	12.413108

5 rows × 252 columns

index tips dataframe as per payment id

```
tips.set_index('Payment ID')
```

Payment ID	Total Bill	Tip	Gender	Smoker	Day	Time	Size	Bill Per Person	CC Number	5.9446733372572105	...	12.413108
Sun2959	16.99	1.01	Female	No	Sun	Dinner	2	8.49	3560325168603410	5.944673	...	
Sun4608	10.34	1.66	Male	No	Sun	Dinner	3	3.45	4478071379779230	5.944673	...	
Sun4458	21.01	3.50	Male	No	Sun	Dinner	3	7.00	6011812112971320	5.944673	...	
Sun5260	23.68	3.31	Male	No	Sun	Dinner	2	11.84	4676137647685990	5.944673	...	
Sun2251	24.59	3.61	Female	No	Sun	Dinner	4	6.15	4832732618637220	5.944673	...	
...	...	...	...	...	...	...	...	...	...	...	...	
Sat2657	29.03	5.92	Male	No	Sat	Dinner	3	9.68	5296068606052840	5.944673	...	
Sat1766	27.18	2.00	Female	Yes	Sat	Dinner	2	13.59	3506806155565400	5.944673	...	

Sat3880	22.67	2.00	Male	Yes	Sat	Dinner	2	11.34	6011891618747190	5.944673	...
Sat17	17.82	1.75	Male	No	Sat	Dinner	2	8.91	4375220550950	5.944673	...
Thur672	18.78	3.00	Female	No	Thur	Dinner	2	9.39	3511451626698130	5.944673	...

244 rows × 251 columns

change the tips dataframe as per payment ID

```
tips=tips.set_index('Payment ID')
```

```
tips.head()
```

Payment ID	Total Bill	Tip	Gender	Smoker	Day	Time	Size	Bill Per Person	CC Number	5.9446733372572105	...	12.413108
Sun2959	16.99	1.01	Female	No	Sun	Dinner	2	8.49	3560325168603410	5.944673	...	
Sun4608	10.34	1.66	Male	No	Sun	Dinner	3	3.45	4478071379779230	5.944673	...	
Sun4458	21.01	3.50	Male	No	Sun	Dinner	3	7.00	6011812112971320	5.944673	...	
Sun5260	23.68	3.31	Male	No	Sun	Dinner	2	11.84	4676137647685990	5.944673	...	
Sun2251	24.59	3.61	Female	No	Sun	Dinner	4	6.15	4832732618637220	5.944673	...	

5 rows × 251 columns

reset index of tips dataframe to row Index

reset index of tips dataframe to row index

```
tips=tips.reset_index()
```

```
tips.head()
```

	Payment ID	Total Bill	Tip	Gender	Smoker	Day	Time	Size	Bill Per Person	CC Number	...	12.413108242303872	7.9365
0	Sun2959	16.99	1.01	Female	No	Sun	Dinner	2	8.49	3560325168603410	...	12.413108	
1	Sun4608	10.34	1.66	Male	No	Sun	Dinner	3	3.45	4478071379779230	...	12.413108	
2	Sun4458	21.01	3.50	Male	No	Sun	Dinner	3	7.00	6011812112971320	...	12.413108	
3	Sun5260	23.68	3.31	Male	No	Sun	Dinner	2	11.84	4676137647685990	...	12.413108	
4	Sun2251	24.59	3.61	Female	No	Sun	Dinner	4	6.15	4832732618637220	...	12.413108	

5 rows × 252 columns