```
import pandas as pd
```

```
df = pd.read_csv("/content/income.csv")
df.head()
```

	age_group	income
0	18-30	472
1	18-30	657
2	31-50	662
3	18-30	263
4	31-50	445

```
df.age_group.unique()
```

```
array(['18-30', '31-50', '51-70'], dtype=object)
```

```
# count number of non null income value in each age group
df.groupby(df.age_group).count()
```

	income
age_group	
18-30	17
31-50	11
51-70	11

```
# minimum value of income in each age group
df.groupby(df.age_group).min()
```

	income
age_group	
18-30	155
31-50	203
51-70	54

```
# maximum value of income in each age group
df.groupby(df.age_group).max()
```

	income
age_group	
18-30	749
31-50	739
51-70	690

```
# mean of income in each age group
df.groupby(df.age_group).mean()
```

	income
age_group	
18-30	432.647059
31-50	423.272727
51-70	400.545455

```
# standard deviation of income in each age group
df.groupby(df.age_group).std()
```

	income
age_group	
18-30	207.238975
31-50	186.966356
51-70	206.593980

```
# describe gives count, mean, standard deviation, min, max, 25th percentile, 50th percentile and 75th percentile.
df.groupby(df.age_group).describe()
```

	income							
	count	mean	std	min	25%	50%	75%	max
age_group								
18-30	17.0	432.647059	207.238975	155.0	253.0	395.0	657.0	749.0
31-50	11.0	423.272727	186.966356	203.0	285.0	381.0	553.5	739.0
51-70	11.0	400.545455	206.593980	54.0	263.5	471.0	518.0	690.0

```
# Part 2
# Load iris data set
```

```
from sklearn import datasets
data = datasets.load_iris()
df = pd.DataFrame(data.data,columns=data.feature_names)
df['species'] = pd.Series(data.target)
df.head()
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	species
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0

```
df.species.unique()
```

```
array([0, 1, 2])
```

```
df.groupby(df.species)
```

```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x7ffbc0c43a90>
```

```
# use aggregation function like count() to get all quantitive variables
df.groupby(df.species).count()
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
species				
0	50	50	50	50
1	50	50	50	50
2	50	50	50	50

```
# max value of each quantitative variable according to categorical variable(species)
df.groupby(df.species).max()
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
species				
0	5.8	4.4	1.9	0.6
1	7.0	3.4	5.1	1.8
2	7.9	3.8	6.9	2.5

```
# min value of each quantitative variable according to categorical variable(species)
df.groupby(df.species).min()
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
species				
0	4.3	2.3	1.0	0.1
1	4.9	2.0	3.0	1.0
2	4.9	2.2	4.5	1.4

```
# mean of each quantitative variable according to categorical variable(species)
df.groupby(df.species).mean()
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
species				
0	5.006	3.428	1.462	0.246
1	5.936	2.770	4.260	1.326
2	6.588	2.974	5.552	2.026

```
# standard deviation of each quantitative variable according to categorical variable(species)
df.groupby(df.species).std()
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
species				
0	0.352490	0.379064	0.173664	0.105386

```
# for each categorical variable we get different summary statistics of sepal length column.
df.groupby(df.species)["sepal length (cm)"].describe()
```

	count	mean	std	min	25%	50%	75%	max
species								
0	50.0	5.006	0.352490	4.3	4.800	5.0	5.2	5.8
1	50.0	5.936	0.516171	4.9	5.600	5.9	6.3	7.0
2	50.0	6.588	0.635880	4.9	6.225	6.5	6.9	7.9

```
df.groupby(df.species)["sepal width (cm)"].describe()
```

	count	mean	std	min	25%	50%	75%	max
species								
0	50.0	3.428	0.379064	2.3	3.200	3.4	3.675	4.4
1	50.0	2.770	0.313798	2.0	2.525	2.8	3.000	3.4
2	50.0	2.974	0.322497	2.2	2.800	3.0	3.175	3.8

```
df.groupby(df.species)["petal length (cm)"].describe()
```

	count	mean	std	min	25%	50%	75%	max
species								
0	50.0	1.462	0.173664	1.0	1.4	1.50	1.575	1.9
1	50.0	4.260	0.469911	3.0	4.0	4.35	4.600	5.1
2	50.0	5.552	0.551895	4.5	5.1	5.55	5.875	6.9

```
df.groupby(df.species)["petal width (cm)"].describe()
```

	count	mean	std	min	25%	50%	75%	max
species								
0	50.0	0.246	0.105386	0.1	0.2	0.2	0.3	0.6
1	50.0	1.326	0.197753	1.0	1.2	1.3	1.5	1.8
2	50.0	2.026	0.274650	1.4	1.8	2.0	2.3	2.5