```
In [1]: %run test.py
```

《Python 程序设计》作业二：数据分析和可视化

班级：计算机 232 姓名：马凌峰 学号：202307020122

(1) 导入数据并查看前10行数据

成功读取 housing.csv 文件

前10行数据：

数据集前10行

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	MEDV
0	0.006320	18.000000	2.310000	0	0.538000	6.575000	65.200000	4.090000	1	296.000000	15.300000	396.900000	4.980000	24.000000
1	0.027310	0.000000	7.070000	0	0.469000	6.421000	78.900000	4.967100	2	242.000000	17.800000	396.900000	9.140000	21.600000
2	0.027290	0.000000	7.070000	0	0.469000	7.185000	61.100000	4.967100	2	242.000000	17.800000	392.830000	4.030000	34.700000
3	0.032370	0.000000	2.180000	0	0.458000	6.998000	45.800000	6.062200	3	222.000000	18.700000	394.630000	2.940000	33.400000
4	0.069050	0.000000	2.180000	0	0.458000	7.147000	54.200000	6.062200	3	222.000000	18.700000	396.900000	5.330000	36.200000
5	0.029850	0.000000	2.180000	0	0.458000	6.430000	58.700000	6.062200	3	222.000000	18.700000	394.120000	5.210000	28.700000
6	0.088290	12.500000	7.870000	0	0.524000	6.012000	66.600000	5.560500	5	311.000000	15.200000	395.600000	12.430000	22.900000
7	0.144550	12.500000	7.870000	0	0.524000	6.172000	96.100000	5.950500	5	311.000000	15.200000	396.900000	19.150000	27.100000
8	0.211240	12.500000	7.870000	0	0.524000	5.631000	100.000000	6.082100	5	311.000000	15.200000	386.630000	29.930000	16.500000
9	0.170040	12.500000	7.870000	0	0.524000	6.004000	85.900000	6.592100	5	311.000000	15.200000	386.710000	17.100000	18.900000

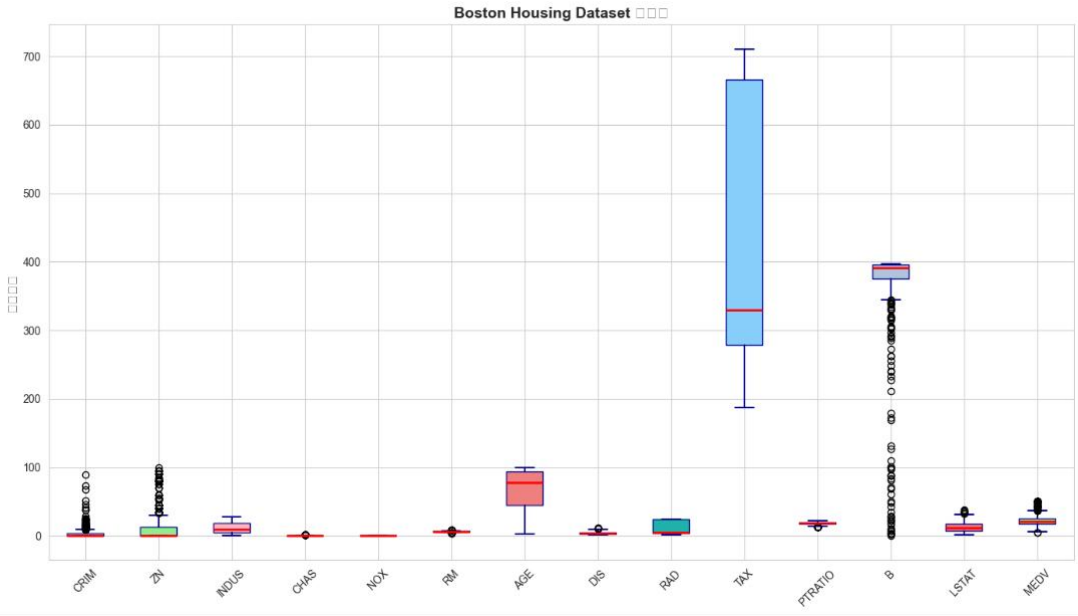
(2) 数据的描述统计

描述性统计：

数据集描述性统计

	count	mean	std	min	25%	50%	75%	max
CRIM	506	3.613524	8.601545	0.006320	0.082045	0.256510	3.677083	88.976200
ZN	506	11.363636	23.322453	0.000000	0.000000	0.000000	12.500000	100.000000
INDUS	506	11.136779	6.860353	0.460000	5.190000	9.690000	18.100000	27.740000
CHAS	506	0.069170	0.253994	0.000000	0.000000	0.000000	0.000000	1.000000
NOX	506	0.554695	0.115878	0.385000	0.449000	0.538000	0.624000	0.871000
RM	506	6.284634	0.702617	3.561000	5.885500	6.208500	6.623500	8.780000
AGE	506	68.574901	28.148861	2.900000	45.025000	77.500000	94.075000	100.000000
DIS	506	3.795043	2.105710	1.129600	2.100175	3.207450	5.188425	12.126500
RAD	506	9.549407	8.707259	1.000000	4.000000	5.000000	24.000000	24.000000
TAX	506	408.237154	168.537116	187.000000	279.000000	330.000000	666.000000	711.000000
PTRATIO	506	18.455534	2.164946	12.600000	17.400000	19.050000	20.200000	22.000000
B	506	356.674032	91.294864	0.320000	375.377500	391.440000	396.225000	396.900000
LSTAT	506	12.653063	7.141062	1.730000	6.950000	11.360000	16.955000	37.970000
MEDV	506	22.532806	9.197104	5.000000	17.025000	21.200000	25.000000	50.000000

(3) 绘制数据的箱形图



(4) 数据的10%和90%分位数

所有特征的10%和90%分位数：

特征分位数

	10%分位数	90%分位数
CRIM	0.038195	10.753000
ZN	0.000000	42.500000
INDUS	2.910000	19.580000
CHAS	0.000000	0.000000
NOX	0.427000	0.713000
RM	5.593500	7.151500
AGE	26.950000	98.800000
DIS	1.628300	6.816600
RAD	3.000000	24.000000
TAX	233.000000	666.000000
PTRATIO	14.750000	20.900000
B	290.270000	396.900000
LSTAT	4.680000	23.035000
MEDV	12.750000	34.800000

目标值(MEDV)的10%和90%分位数：

目标变量分位数

分位数	MEDV值
10%	12.7500
90%	34.8000