What I will focus on today:

* 20% - Describe: Good summary of the dataset, repository well laid-out and organised. Reasonable commits to the repository.

Create a git repository and make it available online for the lecturer to clone. The repository should contain all your work for this assessment. Within the repository, create a jupyter [6] notebook that uses descriptive statistics and plots to describe the Boston House Prices [1] dataset. This part is worth 20% of your overall mark.

1. Good summary of the dataset

Step 1: I reviewed the Boston Standard Metropolitan Statistical Area - Boston house prices dataset. <https://www.kaggle.com/c/boston-housing>

According to the description this dataset details the *Housing Values in Suburbs of Boston. Data description; The Boston data frame has 506 rows and 14 columns.*

We are provided with a ‘key’ which describes what the information in each column represents;

***crim****per capita crime rate by town.*

***zn****proportion of residential land zoned for lots over 25,000 sq.ft.*

***indus****proportion of non-retail business acres per town.*

***chas****Charles River dummy variable (= 1 if tract bounds river; 0 otherwise).*

***nox****nitrogen oxides concentration (parts per 10 million).*

***rm****average number of rooms per dwelling.*

***age*** *proportion of owner-occupied units built prior to 1940.*

***dis****weighted mean of distances to five Boston employment centres.*

***rad****index of accessibility to radial highways.*

***tax****full-value property-tax rate per $10,000.*

***ptratio****pupil-teacher ratio by town.*

***black****1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town.*

***lstat****lower status of the population (percent).*

***medv****median value of owner-occupied homes in $1000s.*

Source;

* Harrison, D. and Rubinfeld, D.L. (1978) Hedonic prices and the demand for clean air. J. Environ. Economics and Management 5, 81–102.
* Belsley D.A., Kuh, E. and Welsch, R.E. (1980) Regression Diagnostics. Identifying Influential Data and Sources of Collinearity. New York: Wiley.

I was immediately struck by one piece of information mentioned;

***black****1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town.*

This seemed very strange to me, and slightly offensive. I wondered if this would be explained a bit more when I became more familiar with the dataset.

When I click on ‘Data’ I expected to be presented with the dataset, but nothing appeared.

I revisited Moodle to see if Ian had mentioned that we need to create an account to access the dataset.

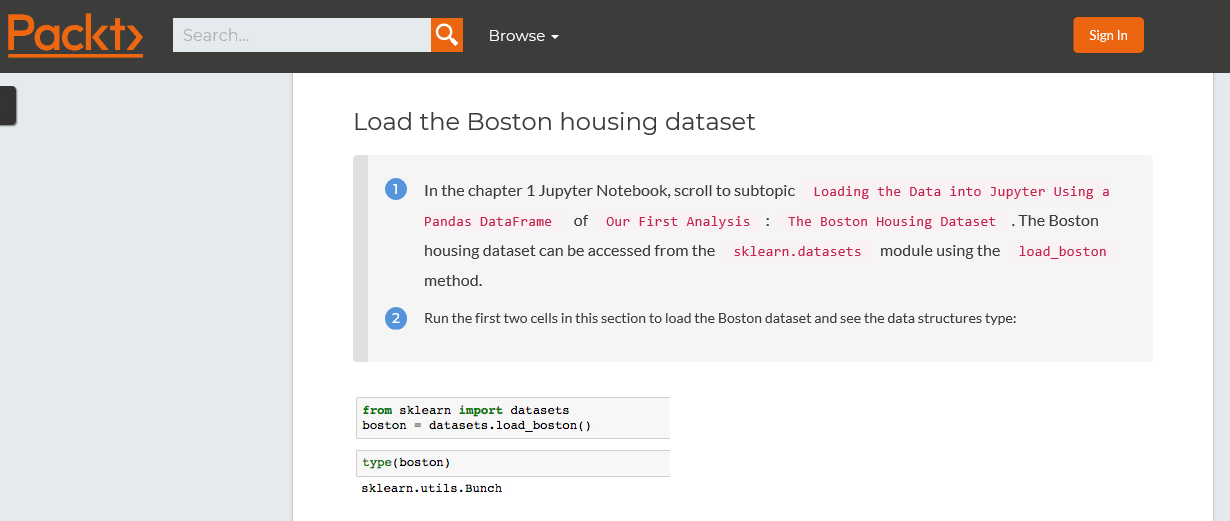
I noticed one comment from Ian in our Outlook group which is used for news, announcements and discussion. Ian’s comment was made in reply to a question posted by one of my fellow students;

*You'll find the common datasets in many different configurations - part of the assignment is to source the dataset in a form conducive to what you need to do with it.*

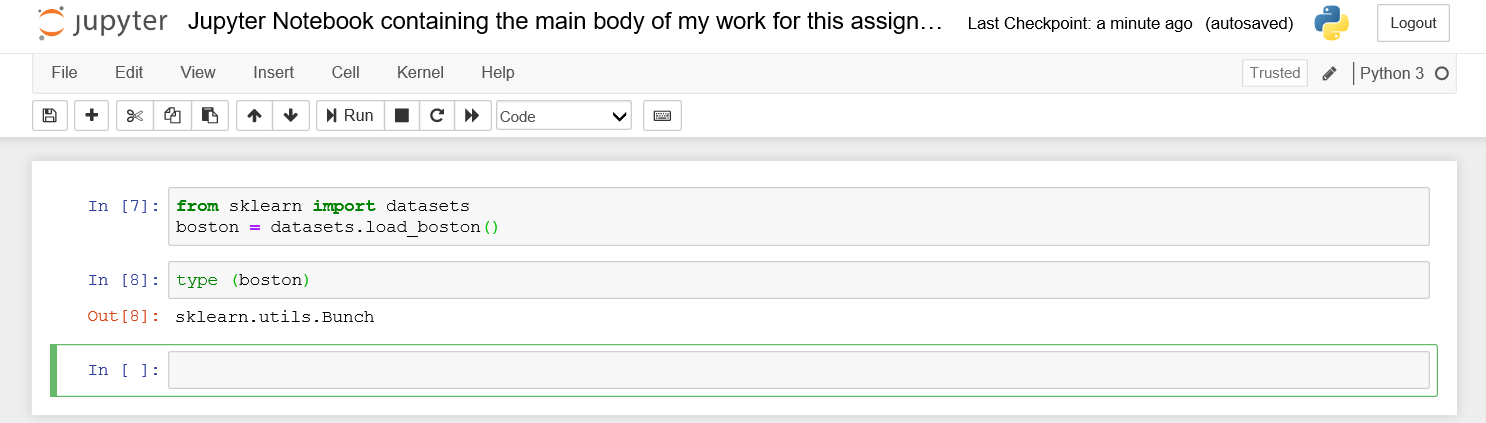
This could be important knowledge for me when completing this assignment.

I discovered the below website which seemed to explain how to load the dataset in Jupyter Notebook;

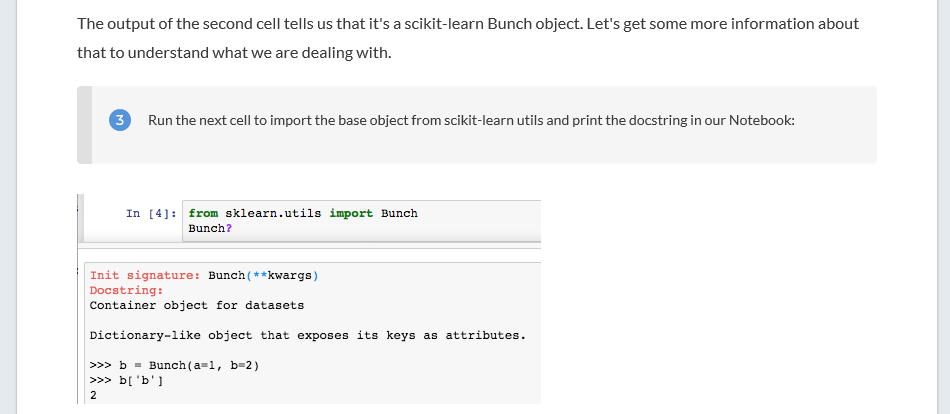
<https://subscription.packtpub.com/book/programming/9781789804744/1/ch01lvl1sec11/our-first-analysis-the-boston-housing-dataset>



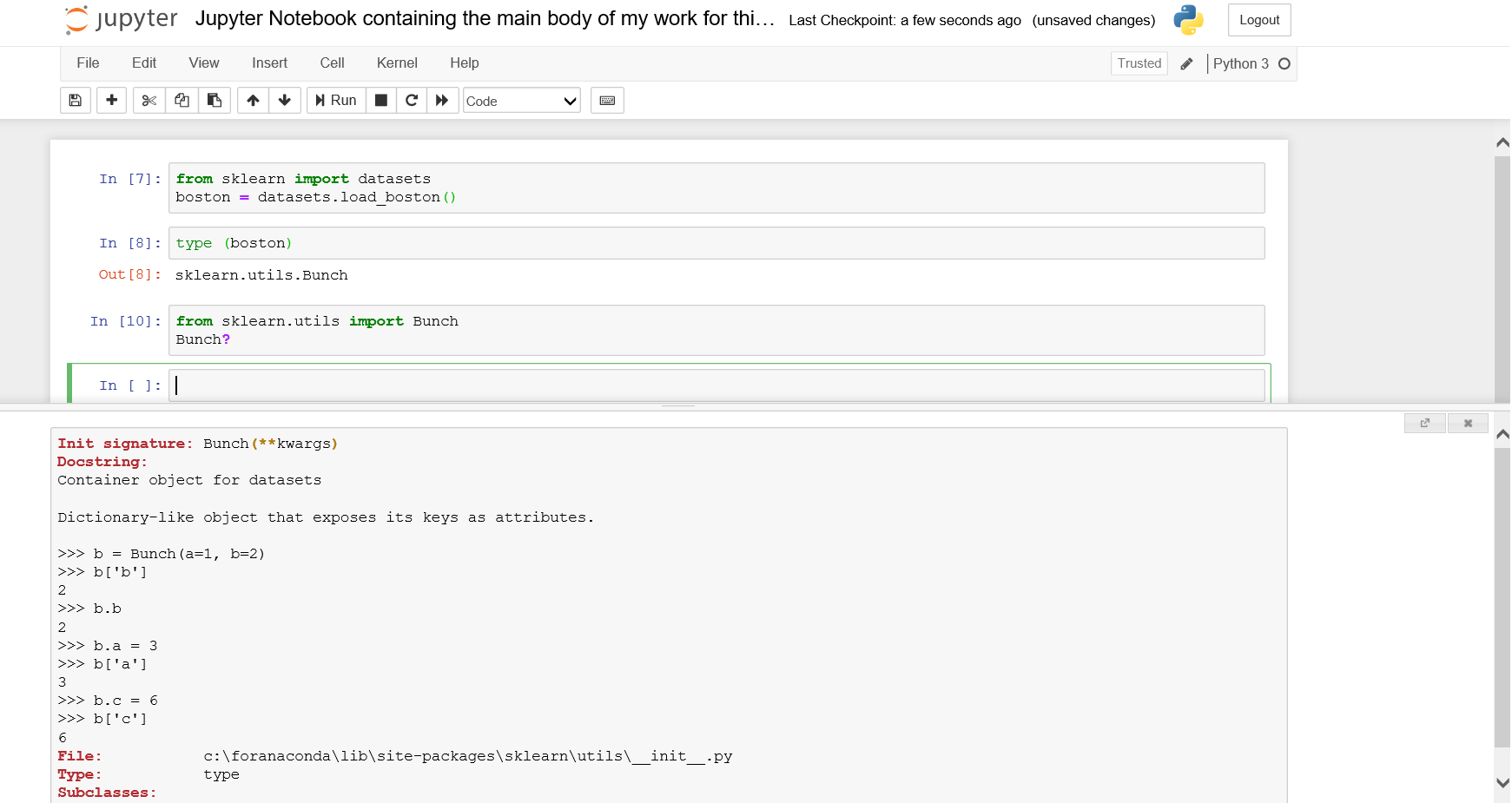
I followed their code as above in my Jupyter notebook and also got the same output;



I continued to follow the instructions from the website;

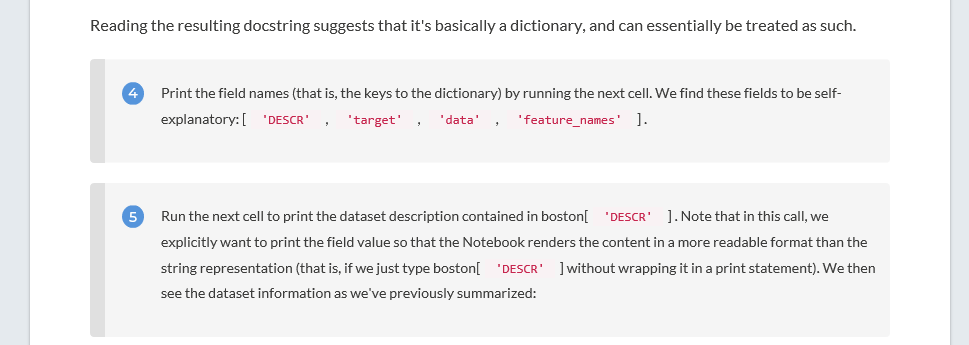


I encounter the same results as displayed on the website in my Jupyter Notebook;

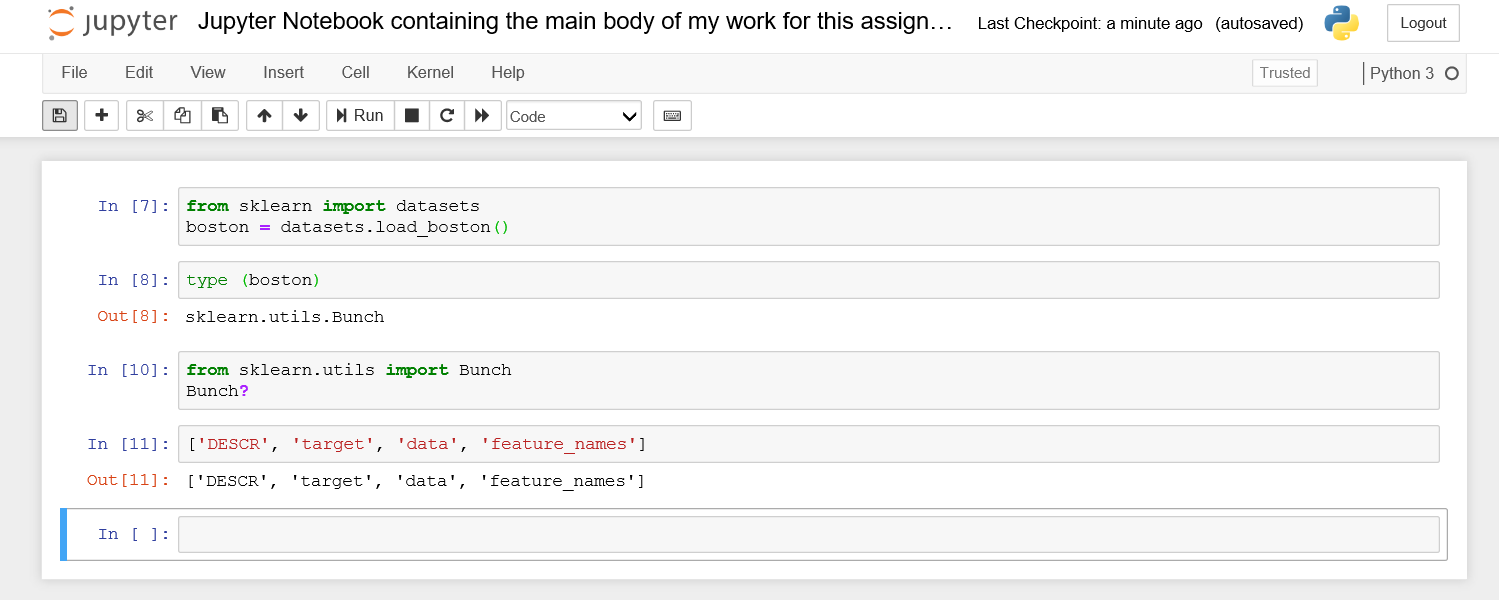


The website explains that ‘Reading the resulting docstring suggests that it's basically a dictionary, and can essentially be treated as such.’ (<https://subscription.packtpub.com/book/programming/9781789804744/1/ch01lvl1sec11/our-first-analysis-the-boston-housing-dataset>)

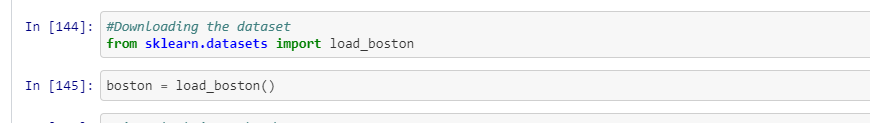
The website guides me on to the next steps;



I stopped at this point as I became confused;



From M:



**My Jupyter Notebook should use descriptive statistics and plots to describe the Boston House Prices dataset.**

[~~https://www.ritchieng.com/machine-learning-project-boston-home-prices/~~](https://www.ritchieng.com/machine-learning-project-boston-home-prices/)