Lecture 1

Models:

most models are “supervised learning”

What are we going to cover:

kNN – nearest neighbors

logistic regression

regression

decision trees

naïve Bayesian classification

linear and quadratic discriminant

random forest

clustering methods

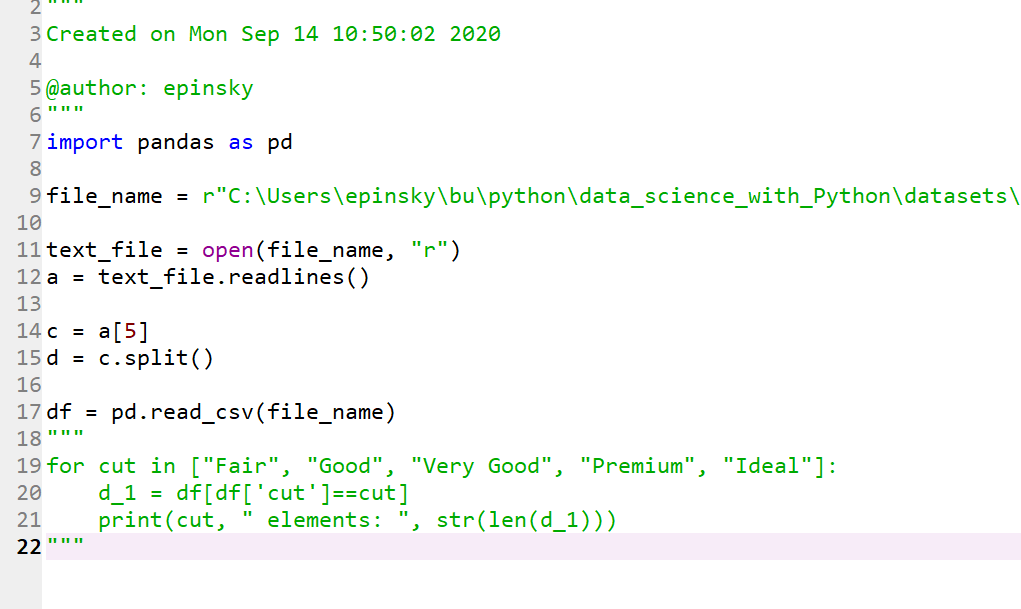
Programming in Python only !!!!

Assignment 1 (two parts)

1. Part 1: use the Python script “read\_and\_save data” to download daily data for a stock for 5 years

(choose a stock that starts with the same letter as first letter of your last name)

1. Part 2: analyze “diamonds” dataset



assume that you read a list of lines

use that list to answer questions

2 ways to compute averages fro fractions:

2/3 and ¼

One way (mu\_a)

mu\_a = ½ \* (2/3 + ¼) = ½ \* (8/12 + 3/12) = 11/24

Another way:

mu\_b = (2 + 1)/(3 + 4) = 3/7

compute:

(11/24 – 3/7)/(3/7)\*100 = 6.9%

for the hyperlink use:

<https://www.foxnews.com/lifestyle/flawless-diamond-100-carats-sothebys>

What do you submit (by Sept 21)

1. python script(s) for diamond set
2. word file with answers

You will put both in a “zipped” folder and name the folder as follows:

jim\_jones\_userid\_homework\_1

submit this zipped folder under assignment 1 tab in the blackboard

For no.3, are we supposed to show the parameters for the maximum price/carat and mimic price/carat?

YES

x[::-3]