Project 6 cs112

Project 6 Learning objectives

- Gain additional experience implementing the Pandas library
- Gain additional experience implementing regression models

Project 6 Data file

- The data file is the AudUsd.csv file
  - This data is the forex currency month time frame data
    - There are 248 months of data in the file

- Screen shot of the first few lines of the file
  - Note that each line contains five pieces of data, delimited by the comma

```
Date,Open,High,Low,Close
12/1/1994,0.7688,0.78,0.7656,0.775
1/1/1995,0.775,0.7765,0.755,0.7563
2/1/1995,0.7564,0.7605,0.735,0.7387
3/1/1995,0.7388,0.7472,0.7212,0.7335
4/1/1995,0.7338,0.7462,0.7228,0.7275
5/1/1995,0.7274,0.744,0.7119,0.7177
6/1/1995,0.7179,0.7288,0.7072,0.7092
7/1/1995,0.7104,0.7409,0.707,0.7383
8/1/1995,0.7381,0.7591,0.7293,0.7509
9/1/1995,0.7503,0.7671,0.7435,0.7551
10/1/1995,0.7553,0.7705,0.7447,0.7605
```

Project 6 Problem to solve

- Read the data file
- Calculate the average price from the high and low
- Add this new data to an 'Average' column in the data frame
- Write the new data frame to file
- Read the newly written data frame
- Plot all the data and regression models: linear (n == 1) and cubic (n == 3)
  - Plot the index value on the x axis
  - Plot the average price value on the y axis
- Plot a subset of the data and the linear and cubic regression models for this subset
  - Subset start value: 20
 - Subset stop value: 80

Project 6 Modified data file

- The modified file has the added column for the Average
  - That is, the average of the high and the low prices

```
Date,Open,High,Low,Close,Average
12/1/1994,0.7688,0.7800,0.7656,0.7750,0.7728
1/1/1995,0.7750,0.7765,0.7550,0.7563,0.7657
2/1/1995,0.7564,0.7605,0.7350,0.7387,0.7477
3/1/1995,0.7388,0.7472,0.7212,0.7335,0.7342
4/1/1995,0.7338,0.7462,0.7228,0.7275,0.7345
5/1/1995,0.7274,0.7440,0.7119,0.7177,0.7280
6/1/1995,0.7179,0.7288,0.7072,0.7092,0.7180
7/1/1995,0.7104,0.7409,0.7070,0.7383,0.7240
8/1/1995,0.7381,0.7591,0.7293,0.7509,0.7442
9/1/1995,0.7503,0.7671,0.7435,0.7551,0.7553
10/1/1995,0.7553,0.7705,0.7447,0.7605,0.7576
```
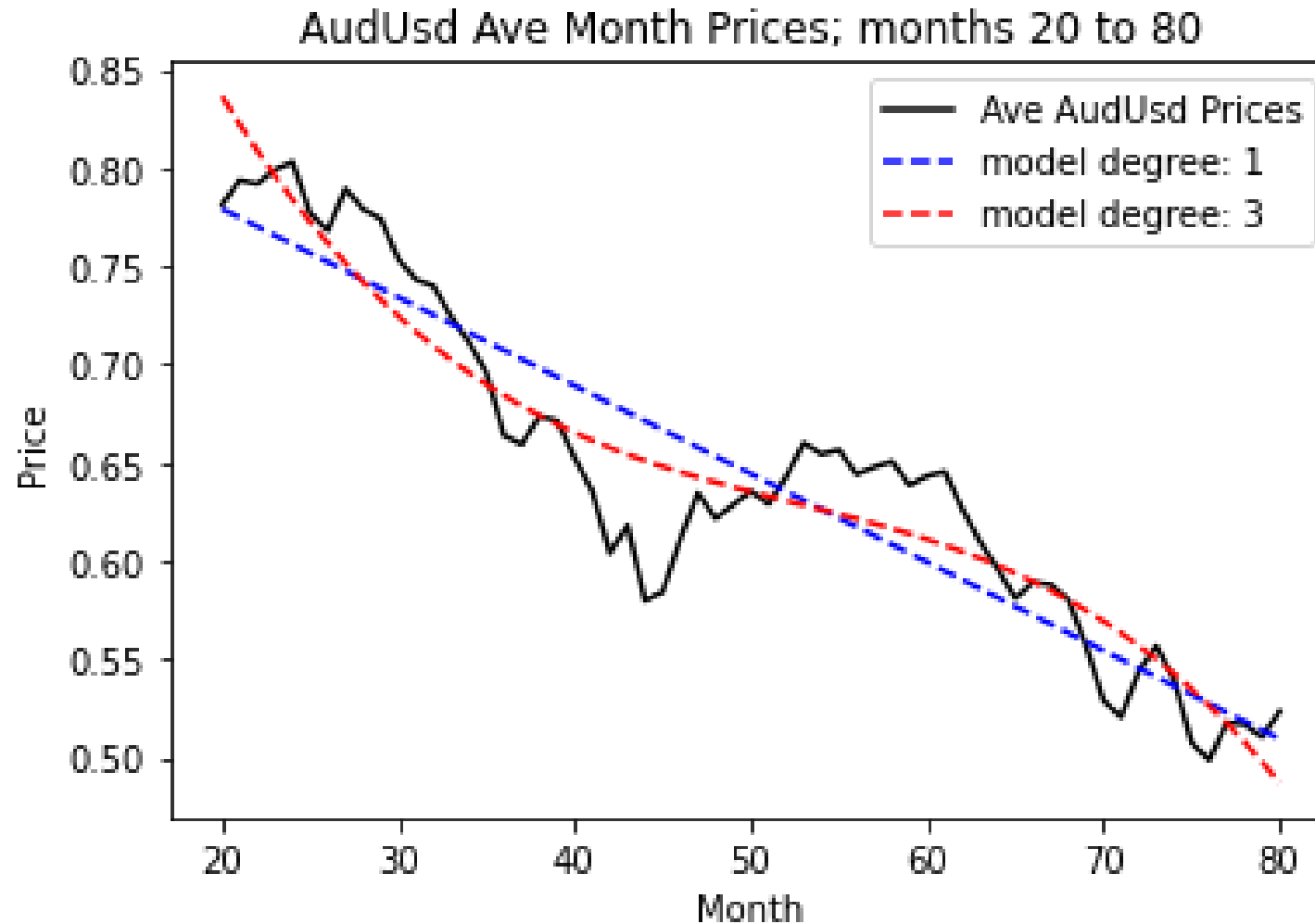
Project 6 Program output

- Plot all the data

AudUsd Ave Month Prices; months 0 to 248
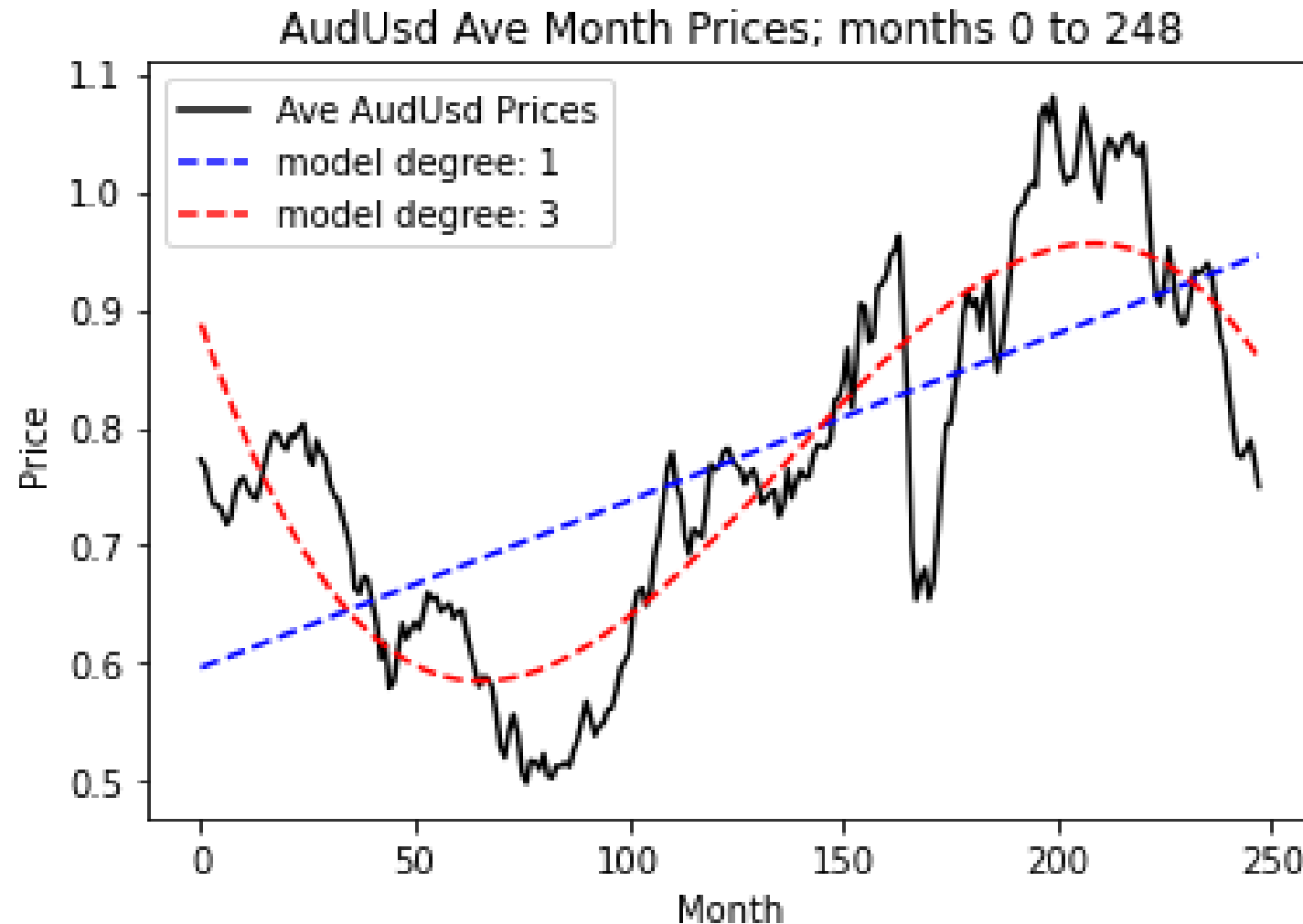
Project 6 Program output

- Plot only months 20 to 80

AudUsd Ave Month Prices; months 20 to 80

The plotData( ) function docstring

- Writing this function is one of the program requirements
- The docstring seen here specifies the requirements that must be implemented by this function

```
In [47]: help(plotData)
Help on function plotData in module __main__:

plotData(pDegrees, start, stop, dFrm, strCol)
    Assumes:
        pDegrees is a list of ints of regression model degrees
        start is an int which is the beginning of the subset to plot
        stop is an int which is the end of the subset to plot
        dFrm is the DataFrame
        strCol is a string which specifies the column to plot on y axis
    Function plots the experimental data and the models specified by pDegrees
        The index values are on the x axis
        The data, and predicted values, in the column specified by strCol are on the y axis
```

To begin solving Project 6 do the following:

- Create a new folder named ***yourLastName***-project6 folder
- Copy the utilitiesModule.py file from Lab 20 and paste in this folder
  - You will need the formatRegressionGraph( ) function in that module
- Copy the pandaUtilities.py file from Lab 20 and paste this inside this folder
  - This file does not need to be modified
- From the Canvas Project 6 Assignment area, download the AudUsd.csv file and save this in the ***yourLastName***-project6 folder
- Start a new Python file named project6.py
  - Save this inside the ***yourLastName***-project6 folder
- Verify that your name is in the header of all .py files
- After understanding the problem that needs to be solved, identify the pseudocode that you will need to follow to solve this
  - This will help you avoid "spaghetti code."

Program requirements:

- The program generates correct output
  - The output must be similar to what is seen in the sample output in earlier slides
- The program implements a plotData( ) function
  - The docstring displayed in an earlier slide specifies the requirements of this function

- The program meets all the criteria in the grading rubric
  - The grading rubric is available in the Project 6 assignment area of Canvas

How to submit your program for grading:

- The "header" of all files submitted must contain your name
- The **folder** that contains the four files must be named *yourLastName*-project6
  - For example, if Mary Smith is submitting this assignment, the folder should be named:

    Smith-project6

- Verify that the folder contains the following four files:
  - AudUsd.csv
  - pandaUtilities.py
  - utilitiesModule.py
  - project6.py
  - Note that there will probably be a fifth file, AudUsd-ave.csv
    - This is the file written by your program
    - You may leave this file in the folder that is being submitted; it will be overwritten when we run your program
- Zip the *yourLastName*-project6 folder
- Submit the *yourLastName*-project6.**zip** folder to the Project 6 assignment area of Canvas
- The criteria that will be used to grade your program can be found in the Project 6 assignment grading rubric