

BA222 Problem Set 2

Due: Thursday October 12, 2022 at 12PM (Noon)
No Late PSets Accepted

Fall 2023

Format – Please Read

This problem set has 2 parts. In the first part, you will complete tasks related to basic Python programming. In the second part, you will complete tasks related to the use of Pandas.

Please submit one single Jupyter .ipynb notebook containing both parts, entitled “(Your Name) Problem Set 2.ipynb”

Coding Standards

- All cells of your notebook should run without error to receive full credit.
- Code is expected to be commented
- All functions should have docstrings
- Make sure each cell is labeled with a comment so we know which question it’s part of
- Use informative variable names (e.g. if you had a list of cities, name it ListofCities rather than just x or myList)
- Please space your operators: write

`3 + 2`

instead of

`3+2`

1 Part 1: Python Basics

Question 1: Compound Interest Calculator

When you invest money with a fixed rate of return, the money will grow exponentially over time.

If the amount invested (the principal) is K , and the money compounds every year with year interest rate r , then the total amount after T years is:

$$K * (1 + r)^T$$

- a) **Write a function** called “Compound” that takes 3 user inputs: K , the principal; r , the yearly interest rate, and T , the number of years of compounding.

The function should return the dollar value of the money, (including the principal), that will be had at the end of T years.

The function should return a floating point number, rounded to 2 decimal places.

- Example:

```
>>> Compound(1000, 0.05, 10)
1628.89
```

- Example:

```
>>> Compound(100, 0.04, 20)
219.11
```

- b) **Answer the following question** with your calculator: which is worth more? \$5,000 compounding for 10 years at a 6% interest rate, or \$2,500 compounding for 25 years at a 6% interest rate?

Question 2: Checking First and Last

Write a function called FirstLastSame that takes a single input, a list of numbers, and tests whether the first and last values in the list are the same.

The function should return either True or False

For example:

```
>>> FirstLastSame([1, 0, 1])
True
```

```
>>> FirstLast([10, 9, 8, 7, 6])
False
```

2 Part 2: Numpy and Pandas

Question 3 – Numpy Simulation and Drop-Shipping Business

Drop shipping is a business model where a person sells an item, online, without keeping any inventory. These businesses leverage major retail platforms, like Amazon and Etsy, to sell items that the business owner sources from other companies.

Consider the formula: $Profit = Revenue - Cost$

In drop shipping, the profit is simple: for every unit sold, the company makes revenue $P * Q$, where price is P and Q is quantity sold. The company incurs costs $C * Q + F$, where C is the per-unit cost, and F is any fixed costs.

The per-unit costs C have 2 features: the cost of buying the good from the manufacturer, and cost of shipping the good. The fixed costs are the cost of any advertising and setup to get the drop-shipping company started.

$$C * Q + F = UnitCost * Q + Shipping * Q + F$$

Therefore, in this space, the profit is given by:

$$Profit = Price * Q - (Unit Cost + Shipping) * Q - F$$

In this example, the firm faces the following distributions at a monthly level:

- The price they receive for the good they are selling is normally distributed around a mean of 20, with a standard deviation of 2. (It is not a function of quantity)
- Quantity sold is a discrete uniformly distributed integer, between 100 and 200 units.
- The firm pays a fixed cost of \$500 per month for advertising and running the business.
- The firm pays \$5 flat for shipping each product.
- The supplier's unit cost is normally distributed, around a mean of \$3, with a standard deviation of \$0.50.

Follow these steps to simulate data on 1,000 months of the business's profit.

- a) Import numpy.
- b) Simulate 1,000 observations of prices from the normal distribution described in the example above.

The syntax is:

```
np.random.normal(mean, stdev, shape)
```

where “shape” is how many observations you want.

Save it as a variable name of your choosing.

- c) Simulate 1,000 observations of quantity sold, using values from the example above. The function you want is called `randint`; it is in the Lecture 7 notes. Save it as a variable name of your choosing.
- d) Simulate 1,000 observations of unit cost, from the normal distribution described in the example above. Save it as a variable name of your choosing.
- e) Using your saved variables, and the numbers from the example above, create 1,000 simulated values for the firm’s profits.

Don’t forget to include flat-rate shipping costs, and monthly advertising

Save it as a variable name of your choosing.

- f) What are the mean, median, max, min, and standard deviation of your month-level simulated profits?
- g) Import Matplotlib.

Produce a histogram of these values using matplotlib. What is the shape of this distribution, roughly? Include a screenshot in your submission.

Question 4 – Real Data Analysis in Pandas

- a) Import pandas

Load in the file `CongressTerms.csv` using the Pandas `read_csv` function. It is available on the Questrom blackboard homework site. Use the `head()` function to see the top of your dataset

- b) How many observations are in this dataset?
How many Congress vs Senate observations? (Use the “chamber” column).
- c) What is the mean age of senators in the 107th session of congress?
(You can tell the 107th session using the “congress” column)
- d) Who is the oldest senator in this data? (Hint: check which rows have age equal to the max age).
Does this seem like a reasonable value? Google the Senator and see if the result you found is correct.
- e) In the 110th Congress, what was the most common first name of House members?
How many times does it appear?