DIAML: Assignment 1

by Anthony Rizkallah

Table of contents

1 Python Packages

Numpy

Math

Pandas

Matplotlib

2 Question 1

2.1 Steps

Folding a sheet of paper of thickness d over n iterations to reach a specific height h, follows the logic that every iteration will double its thickness to *2d* of the previous thickness. So, 1mm becomes 2mm, 2mm become 4mm, 4mm become 8mm, and so on. It could be written as follows:

$$h = d \times 2_1 \times 2_2 \times 2_3 \ldots 2_n$$

where the subscript n is the number of the fold. Similarly, it can be written as such,

$$h = d \times 2^n$$

Solving for n requires mathematical manipulation such that n is brought down from the exponent. So, applying $\log_2$ to the equation on both sides takes care of that.

$$\log_2\left(\frac{h}{d}\right) = n$$

Now h and d are given as the target height and paper thickness, respectively, where $h = 8,848m$ and $d = 1mm = 1 \times 10^{-3}m$. Then, the number of folds $n = \log_2(8848 \times 10^3)$.

Note: I also provided a solution using code rather than just a mathematical one.

2.2 Results

It takes approx. 23 folds to reach $8,848m$. However, the result is approx. 23.077, and there is no such thing as a fraction of a fold, so it is a debate whether logically it takes 24 folds even though at the 24$^{\text{th}}$ it would have reached about 17,000m. Even though a graph was not required, I provided one to make it easier to visualize the effect.

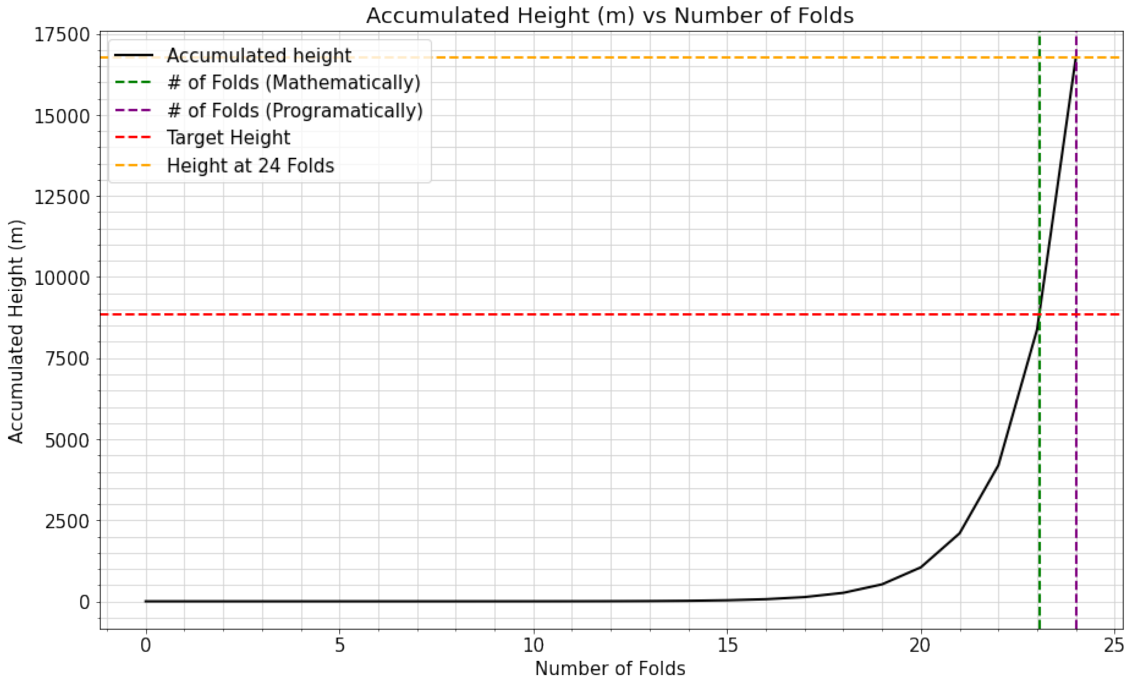

Figure 1. Accumulated Height (m) of a paper with thickness 1mm with respect to the number of folds

2.3 Insights

There are many conclusions that can be made from this exercise. One I find interesting is assuming I invest 0.001 Dollars (not practically feasible) in a business model that doubles my money every day, then after 23 days, I would have made 8,848 Dollars. For humans, who are better equipped to understand linear relations, it is hard to understand the compounding effect of doubling a very small number many times, and how fast it explodes to larger numbers after short period of time.

3 Question 2

3.1 Steps

The water reservoir is losing water at a rate of $V(t) = V_0\, e^{-at}$, where V_0 is the initial amount of water, a is the rate at which the reservoir is losing water, and t is the time. To find the time t at which the reservoir loses half of its water, I start by setting $V(t) = \frac{1}{2}V_0$. Then,

$$\begin{aligned}
\frac{1}{2}V_0 &= V_0\, e^{-at} \\
\frac{1}{2} &= e^{-at} \\
\ln\left(\frac{1}{2}\right) &= -at \\
\frac{\ln\left(\frac{1}{2}\right)}{-a} &= t
\end{aligned}$$

Then, the time it takes for the $t = -10 \times \ln\left(\frac{1}{2}\right)$

3.2 Results

It takes approximately 6.931 units of time to reach a half-full reservoir. This means that any time after 6.931 units of time, is less than half.

$$T > 6.931\,\text{units}\,(\text{time})$$

I also provided a graph showing the decreasing graph below.

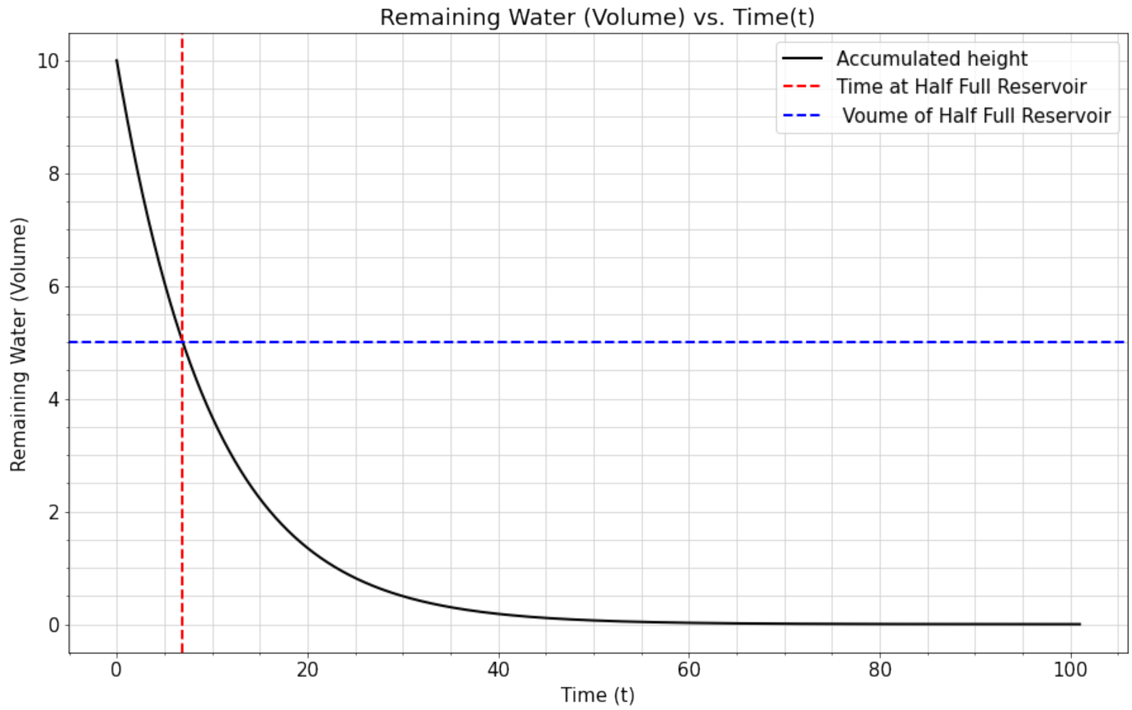

Figure 2. Remaining water in a reservoir that is decreasing as $V_0\, e^{-at}$ in volume with time.

3.3 Insights

I have mathemtical, observational, and physical insights. Mathematically, it makes sense that the water level is very high at first and later forms a plateau because e^{-x} has a limit that approaches 0 as x tends to infinity. Oberservationally, the rate at which the water is decreasing is fast at first, but then slows down– taking 6.931 seconds to reach half and then almost 54 seconds to reach 0. Physically, the pressure of the water in the tank at first is very high and then after water decreases, the pressure on the exit hole of the tank also decreases, resulting in slower decay.

4 Question 3

4.1 Steps

Depositing an initial amount of money– the principal P– in a bank account that offers an interest rate R, the amount will increase at this rate (assuming it is compounded once in a period). Then, it increases at a rate of $(1+R)$ for every compounding period T of the previous amount. So, the total amount earned at some time T can be written as follows:

$$A = P(1+R)_1 \times (1+R)_2 \times \cdots (1+R)_T$$

where the subscript T determines the time (minutes, days, years). Similarly,

$$A = P(1 + R)^T$$

However, if the interest rate R is the rate related to the time period T, but the rate is applied n times in that period, then the period T can be written as follows,

$$T = t \times n$$

Rewriting T in terms of t and n without adjusting the interest R is logically flawed because by definition R is directly related to T. If the compounding is done n times in the period T, then in a period $t < T$ it must be a fraction of R,

$$R = \frac{r}{n}$$

Finally,

$$A = P\left(1 + \frac{r}{n}\right)^{nt}$$

I could have gotten it directly from the internet, but it's always good practice to derive it.

4.2 Results

Compounding 100 dollars annually at an interest rate of 5% over 1, 2, 3, 4, and 5 years yields 105, 110, 116, 122, and 128 dollars respectively.

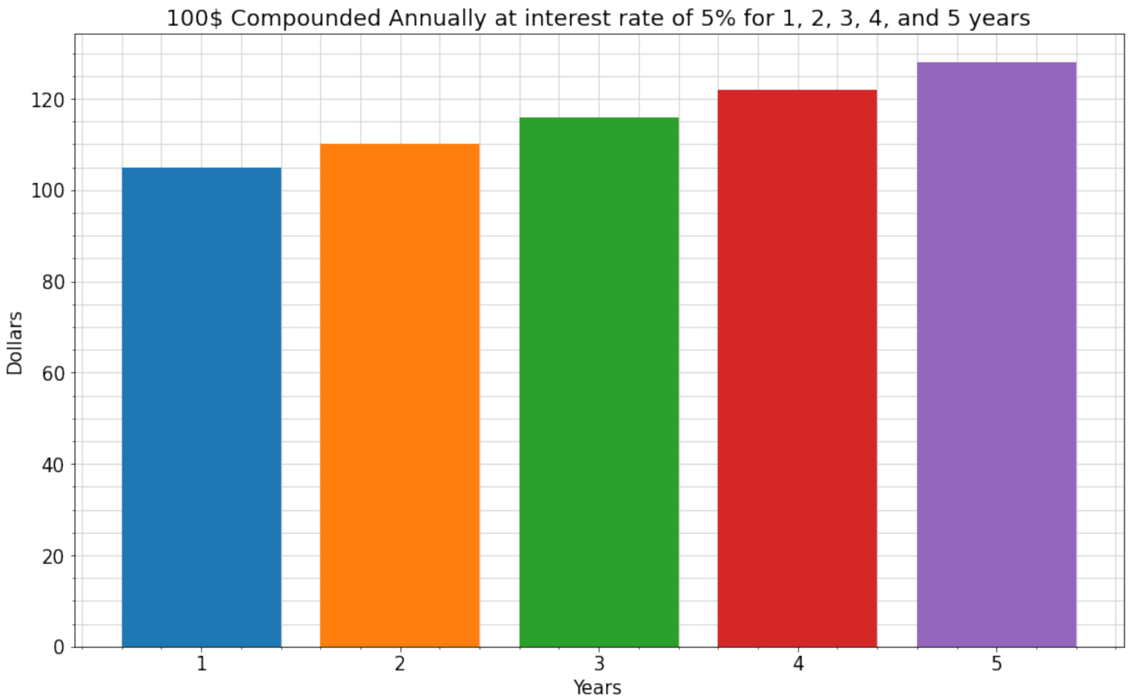

Figure 3. 100 dollars compounded annually at an interest rate of 5% for 1, 2, 3, 4, and 5 years

To see where this number will blow up, it would be interesting to do a bar graph tracking it over 100 years. Inflation and other economic factors will deminish that number in 100 years, but this is assuming that the principal is left to grow without any external assistance (adding more money).

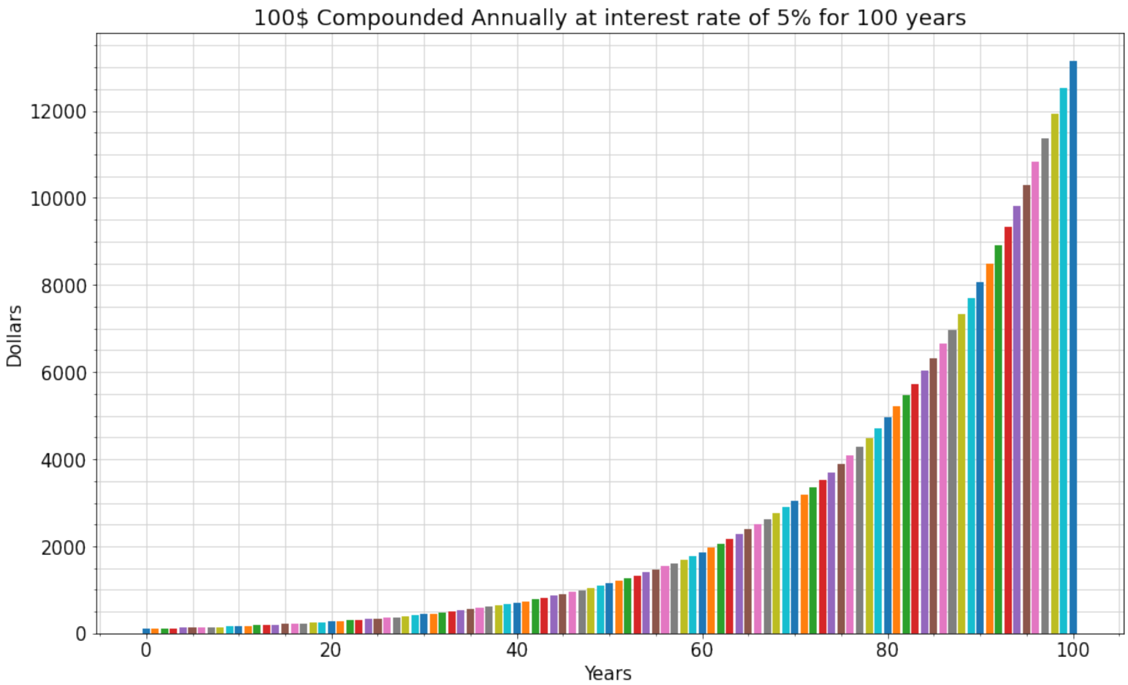

Figure 4. 100 dollars compounded annually at an interest rate of 5% for 100 years

4.3 Insights

The insight for this is similar to the one in Question 1, but drastically stretched because folding a paper has an "interest rate" of 100% instead of 5%. The idea is still the same– compounding starts slow then drastically blows up, and the factor determing how fast it will blow up is the rate at which it grows. However, no matter how large the rate is, it will always start slow compared to

how fast it will go at later times.

5 Question 4

5.1 Steps

For an amortizing loan with an interest r, finding the payment amount without directly using the amortization formula is a little bit complicated. So, I tried following a logical setup to better understand the formula, which led to a derivation. The main high-level idea is that the interest on the loan is being compounded monthly while the borrower is making payments which are first deducted from the interest, and then deducted from the principal. If the borrower does not make payments that are *at least* equal to the interest, the loan amount will increase. Moreover, the payments being made are such that if the interest is applied at time t, the payment towards that accumulation is done a period later, with a payment timestamp of $t-1$. So, assuming A_0 is the initial loan amount, A is the running amount of the loan, and t is the time at which interest is applied:

$$A(t) = A_0(1+r)^t$$

This is similar to the compounding formula, however, if the borrower is making a payment towards that loan, then the equation should look something like that:

$$A(t) = A_0(1+r)^t - P_{t-1};$$

noting the timestamp of $t-1$ refers to the payment being done towards a previously posted interest charge on the loan.

Again, P_{t-1} is an ambigious term and can be written in any way, but preferraby in terms of the interest amount for easier manipulation,

$$A(t) = A_0(1+r)^t - P(1+r)^{t-1}$$

Generally speaking, the formula makes sense, but it does not represent the *total* number of payments being made towards the principal. In other words, $t-1$ is the final payment to be made towards the loan, and this formula does not represent the accumulation. Hence,

$$A(t) = A_0(1+r)^t - \sum_{t=0}^{t-1} P(1+r)^t$$

is more accurate as it takes into account all the payments being made towards the loan, and the sum can be rewritten as a geometric series such that

$$\sum_{t=0}^{t-1} (1+r)^t = P\frac{(1+r)^t - 1}{(1+r) - 1}$$

To find the payment amount at time t, $A(t)$ must vanish to 0, so that the loan, with the accumulated interest, as well as payment contributions are complete. So,

$$\begin{aligned} P\frac{(1+r)^t - 1}{(1+r) - 1} &= A_0(1+r)^t \\ P &= A_0\frac{r(1+r)^t}{(1+r)^t - 1} \end{aligned}$$

which is exactly the amortization equation. I know you probably did not want a derivation, and at this point I will stop deriving things unless explicitly asked for.

5.2 Results

For a full amortization of the 20,000 dollar loan over 1, 2, and 3 years, the payments to the nearest dolalr are 1,777, 941, and 664 respectively.

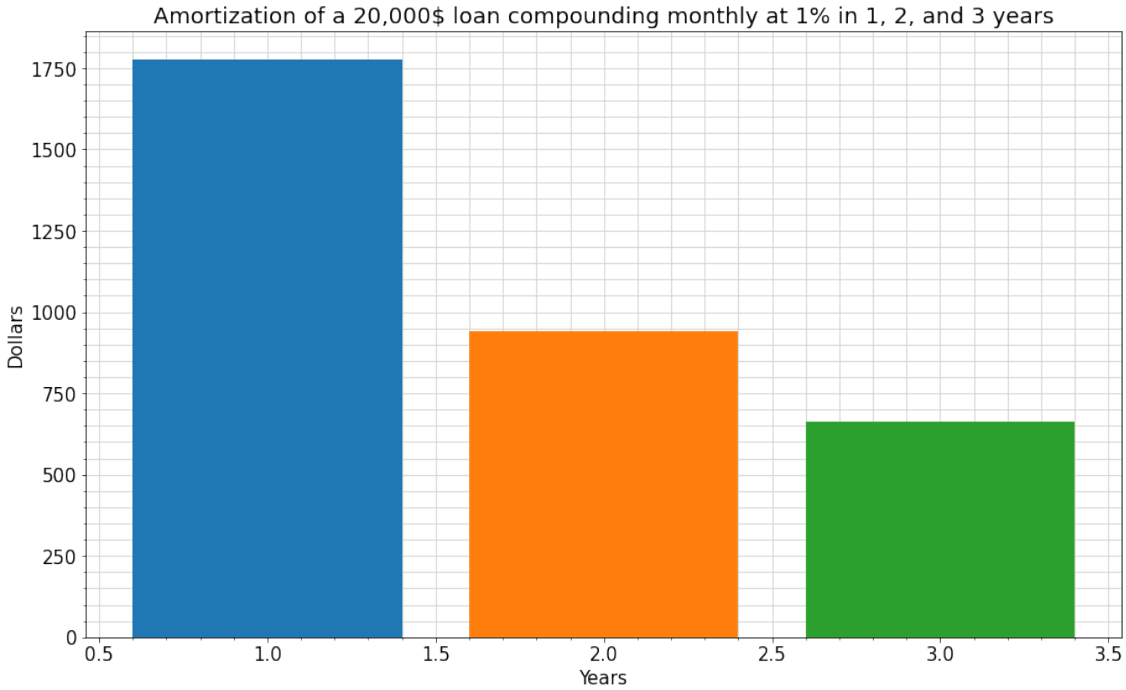

Figure 5. Amortization of a 20,000 dollar loan compounding monthly at 1% in 1, 2, and 3 years

5.3 Insights

What is intersting is the amount of interest accumulated for each of the three cases– 1, 2, and 3 years. Interest Paid in 1 year: 1,324 dollars, Interest Paid in 2 years: 2,584 dollars, Interest Paid in 3 years: 3,904 dollars

I realized that the payment towards the principal gets smaller with respect to the payment amount, and the interest starts taking a bigger portion of the monthly payment as the loan term gets longer. The following graph shows this relationship.

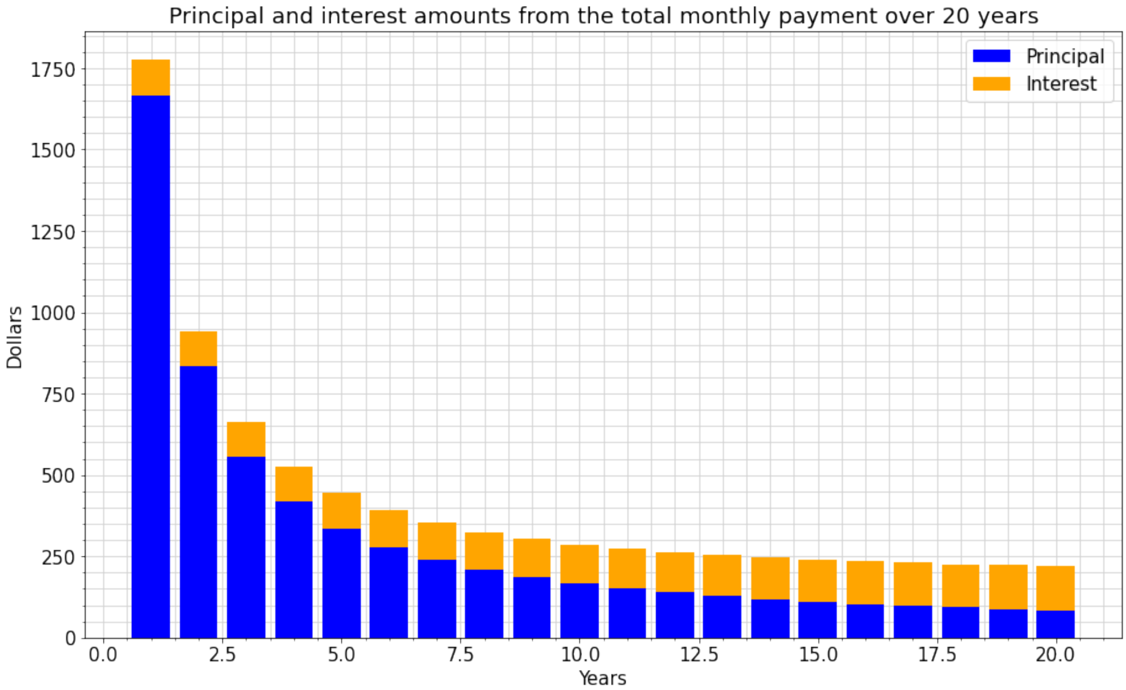

Figure 6. Principal and interest amounts from the total monthly payment over 20 years

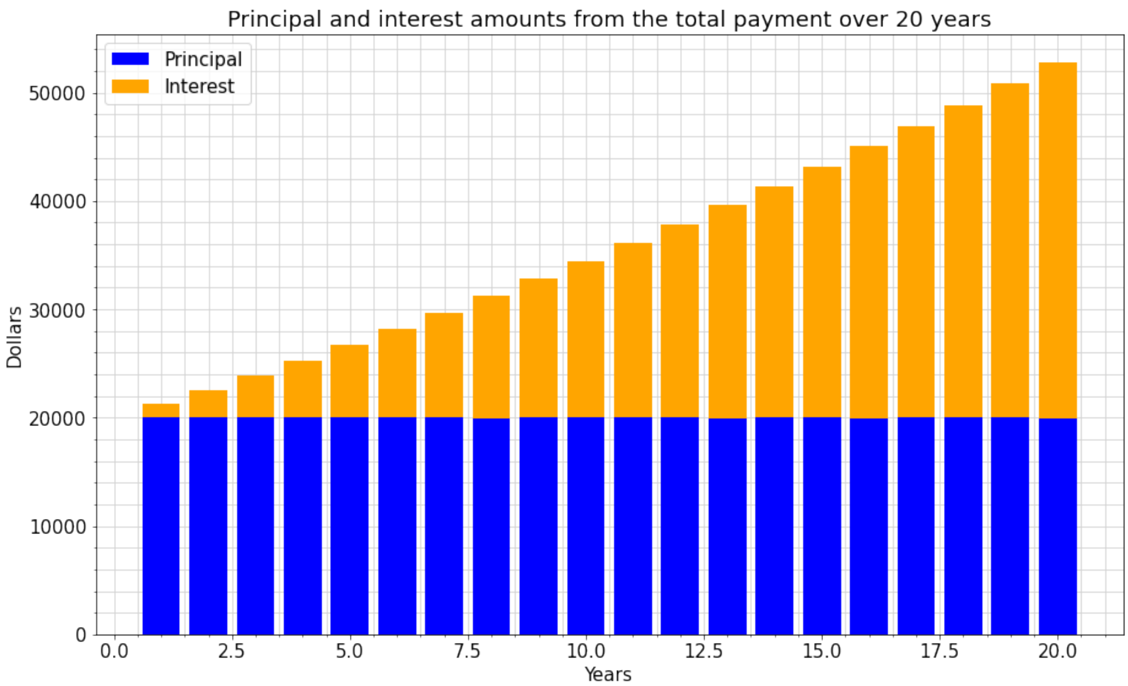

Figure 7. Principal and interest amounts from the total payment over 20 years.

6 Question 5

6.1 Steps

First, I won't take into account the initial investment A_0, but rather the number of customers $C(t)$ that is supposed to grow at a rate r each day. Instead of using the compounding formula, I will approximate $C = C_0(1+r)^t$ as $C = C_0 e^{rt}$, since it is the proper approximation and neater in code.

Next, to find how many days it takes, I can first setup the general equation of the business:

$$D(t) = p \times C(t)$$

where $D(t)$ is the total amount of profit, p is the profit per customer and $C(t)$ is the number of customers. Then, using the previous form for $C(t)$, initially,

$$D(t) = p \times C_0 e^{rt}$$

However, this formula reflects the profits earned at a snapshot of time t, but every day new customers come in and are added to the pool of customers that bring in more business profit, getting us closer to the 100,000 dollars initial investment. So,

$$\begin{aligned} D(t_0 = 0) &= 0 + pC_0 \\ D(t_1 = 1) &= D(0) + pC_0\, e^{r} \\ D(t_2 = 2) &= D(1) + pC_0\, e^{2r} \end{aligned}$$

The emerging pattern is that of a sum, which can be rewritten as

$$D(t) = \sum_{t=0}^{T} pC_0 e^{rt}$$

Using code, I looped over that some by setting $D(t)$ as 100,000 to find the amount of days it takes to accumulate 100,000 dollars.

6.2 Results

It takes 70 days to break even my initial investment of 100,000 dollars.

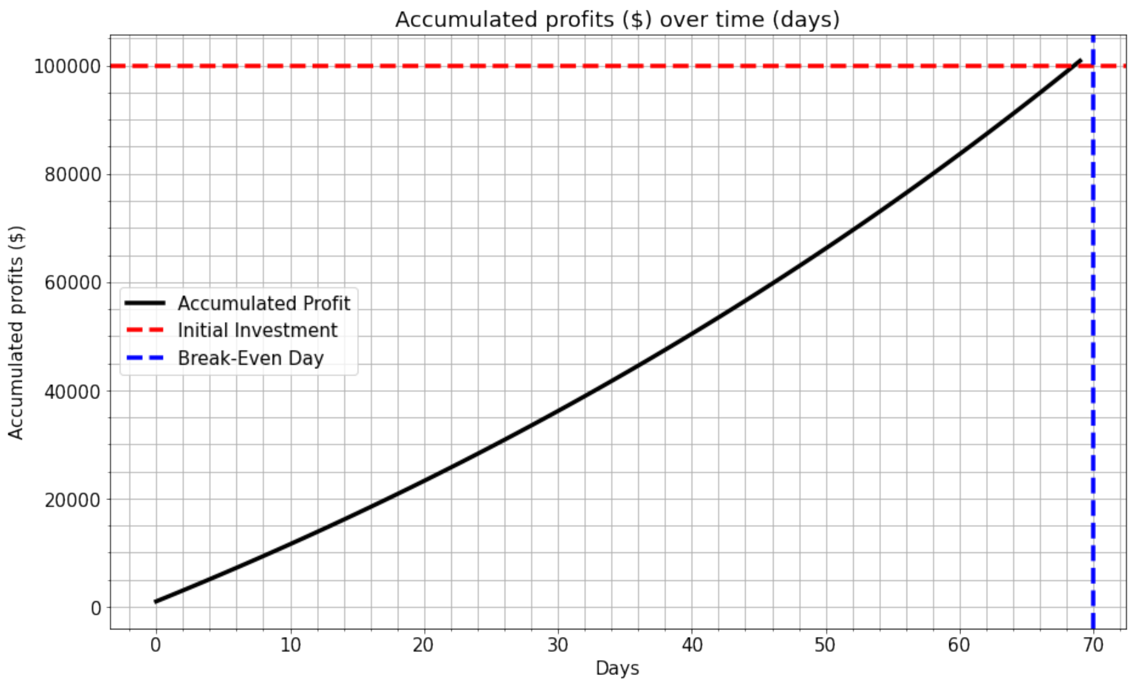

Figure 8. Accumulated profits from a business that is constantly growing its customer base at a rate of 1%, with each customer bringing in 10 dollars of profit. With an initial investment of 100,000 dollars, it takes 70 days to breakeven starting with 100 customers on day 1.

6.3 Insights

The relationship is linear between the accumulated profits and the time it takes to acquire enough customers to break even.

Investing in a succesful business model has two simple axioms: profits and a decent market size. Increasing the number of customers by small percentage daily can bring in a lot of profit quicker than one thinks. The difficult part is setting up a business that is a necessity so that all my customers come in on a daily basis. However, this is a simple business model, and the previous graph can act as a plan of action, meaning, if I don't meet my customer acquisition goals, I could have an alternative plan of action to acquire those customers. On the other hand, I can increase the profit margins if I see that customer acquisition and conversion is not working, and my current customers can handle a slightly more expensive service.

7 Question 6

7.1 Steps

The data retrieved has 5 columns– Date, Cases, Death, Diff, Noofdays. The *Diff* column represents the difference between Date_1 and Date_2 showcasing the missing dates between the two. The *Noofdays* column shows the span from the start date to the current date (depending on the row). The *Cases* and *Death* columns are the number of infected people with Ebola and the number of people who died from Ebola, respectively. I dropped the *Diff* and *Noofdays* columns because I am going to fill in the missing points using linear interpolation.

The linear interpolation equation is as follows,

$$y = y_1 + (x - x_1)\frac{y_2 - y_1}{x_2 - x_1}$$

The *matplotlib* library in Python can do that directly. Then, to find the at what point the cases and deaths exceed 100, 500, 1,000, 2,000, and 5,000, it is the first value after the target values, such that value $\geqslant$ target.

7.2 Results

The dates where the *cases* exceeded 100, 500, 1,000, 2,000, and 5,000 are 2014-03-27, 2014-06-13, 2014-07-15, 2014-08-12, and 2014-09-12 respectively.

The dates where the *deaths* exceeded 100, 500, 1,000, 2,000, and 5,000 are 2014-04-06, 2014-06-13, 2014-07-15, 2014-08-12, and 2014-09-12 respectively.

The table below shows the previous data in Y/M/D format.

Threshold	Date when Cases Exceeded Threshold	Date when Deaths Exceeded Threshold
100	2014-03-27	2014-04-06
500	2014-06-13	2014-07-05
1,000	2014-07-15	2014-08-09
2,000	2014-08-12	2014-09-03
5,000	2014-09-12	2014-10-24

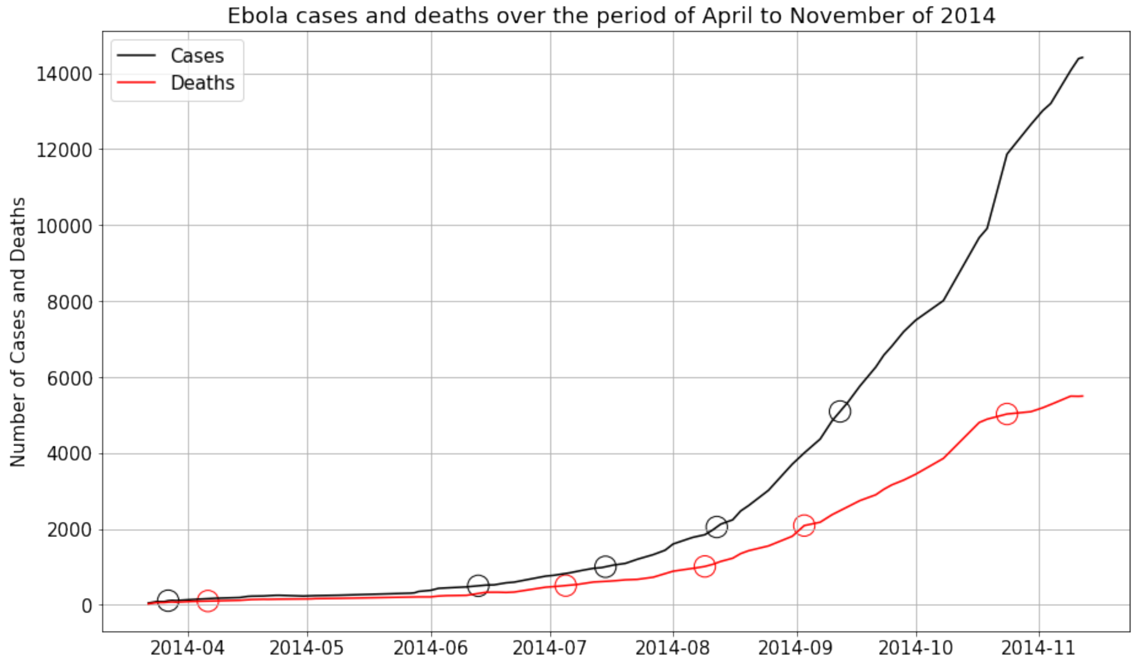

Figure 9. Ebola cases and deaths over the period of April to November of 2014. The circles represented in balck and red (Cases and Deaths, respectively) show when the 100, 500, 1,000, 2,000, and 5,000 thresholds have been exceeded respectively.

7.3 Insights

At the beginning, the number of cases and deaths were slowly increasing, and they later grow at a faster rate. This is showcased either graphically– very small slope and then a drastic change in direction for both the cases and deaths; logically– the distance between the 100, 500, 1,000, 2,000, and 5,000 points gets smaller, representing a faster spread of the virus. The number of cases remains in an upward direction, but the number of deaths seems to slow down in November of 2014 as the graph tends to plateau. This could be due to more understanding of the virus or death-preventative measures taken by healthcare professionals.

8 Question 7

8.1 Steps

The growth rate is dependent on having a value at time t and a value at time $t-1$ or $t+1$ (depends on what time t signifies). To find the rate at which the number of cases and deaths is evolving, I can use the formula

$$r(t) = \frac{p(t)}{p(t-1)} - 1$$

where $p(t)$ is the current number, $p(t-1)$ is the previous number, and subtracting 1 because the mother formula is

$$r(t) = \frac{p(t) - p(t-1)}{p(t-1)} = \frac{p(t)}{p(t-1)} - \frac{p(t-1)}{p(t-1)}$$

The *Matplotlib* pacakge in Python has can perform this without me having to apply this formula to every data point.

8.2 Results

The growth rate for cases is 2.51% and 2.33% for the deaths.

	Cases	Deaths
Rate	2.51%	2.33%

8.3 Insights

The rate of the deaths is less than the rate of cases, which is almost a good sign if the rate of deaths was not so high. This is also observed in the graph from Question 7, as the number of cases was larger than the number of deaths. However, the rates are very close to each other and this could be more alarming if the rate of deaths becomes equal or greater than the rate of cases. Healthcare professionals could be interested in those two numbers to track whether their efforts to contain the virus are working, and whether their efforts for attenuation are working.

9 Question 8

9.1 Steps

To find the ratio of deaths to cases of Ebola, I can just divide the number of deaths at time t by the number of cases at time t. I did that and added it to a separate column so that finding the mean becomes easy.

9.2 Results

The ratio of deaths to cases of Ebola is about 0.56.

The graph below shows the Deaths vs. Cases plot for Ebola cases.

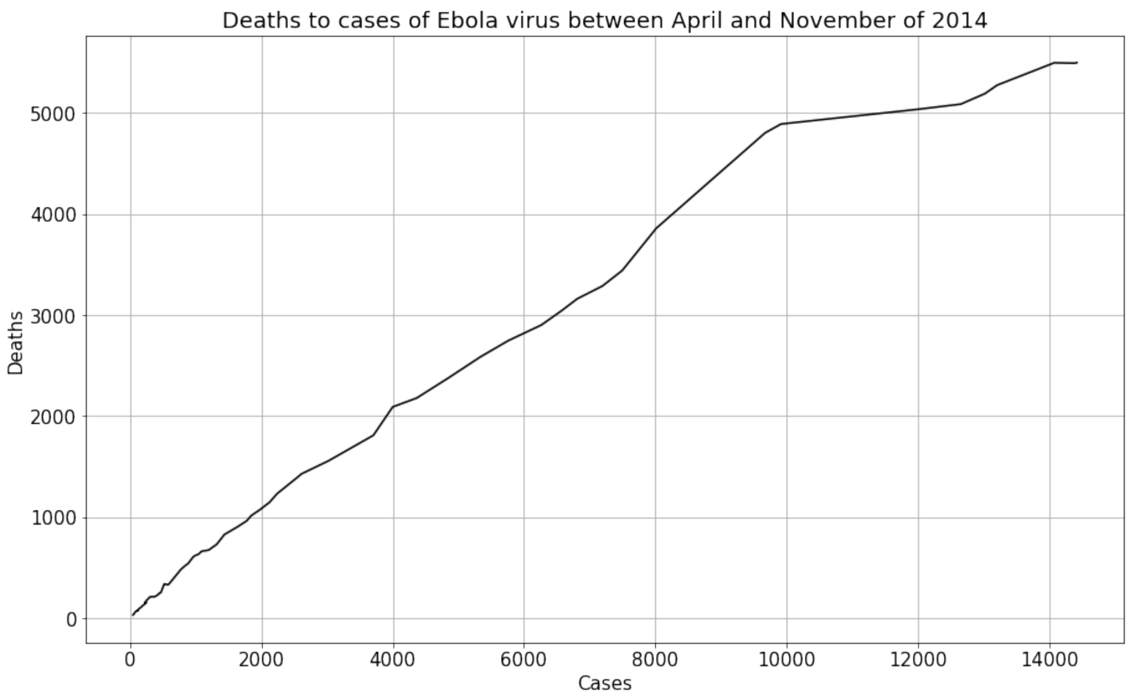

Figure 10. Deaths to cases of Ebola virus between April and November of 2014.

9.3 Insights

A ratio of 0.56 means that for every 100 cases of Ebola, 56 people die and 44 people survive. The graph shows that relationship with some outliers. At around 14,000 cases, the graph is tending to a plateau, but there is not enough data to prove this. So, I graphed the ratio of deaths to cases over time to see if the rate of deaths is decreasing.

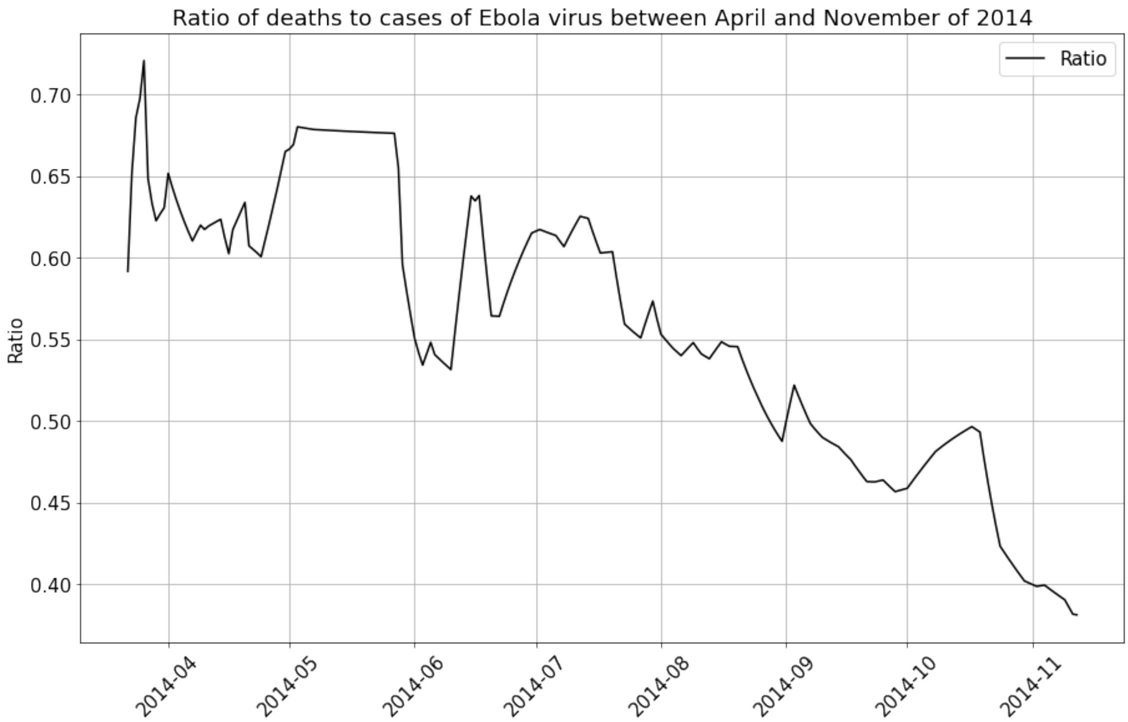

Figure 11. Ratio of deaths to cases of Ebola virus between April and November of 2014

There is a general trend of decline in the ratio of deaths to cases over time, and it is pointed out in the graph in Question 7 as the number of deaths

10 Question 9

10.1 Steps

To normalize a set of values, I first have to find the scaling ratio, which in this case it is dependent on the first value being 100. After doing that I would have found a ratio that I would scale everything to. It takes the form of

$$\text{Scaled Number} = \frac{\text{Target Number}}{\text{Original Number}} \times \text{Original Number}$$

I would apply this to all values in the dataset to get my desired scale. For example, [150, 200, 250] scaled to start with 100 becomes:

$$\begin{aligned}
\text{Scaled Number}_1 &= \frac{100}{150}150 = 100; \left(\text{Scaling Factor is } \frac{100}{150}\right) \\
\text{Scaled Number}_2 &= \frac{100}{150}200 = 133.33 \\
\text{Scaled Number}_3 &= \frac{100}{150}250 = 166.67
\end{aligned}$$

Then, if the original numbers are required at anytime in the analysis, multiplying by the *inverse* of the scaling factor gives them back.

10.2 Results

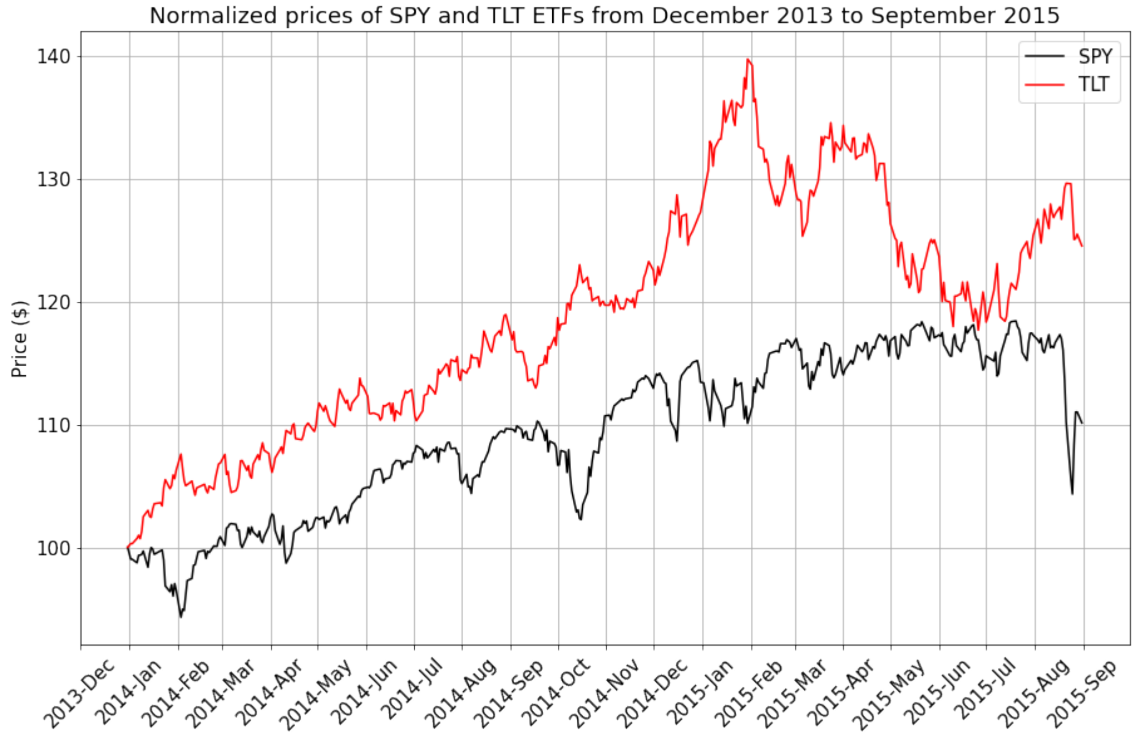

Figure 12. Normalized prices of SPY and TLT ETFs from December 2013 to September 2015

10.3 Insights

The two ETFs are inversely related– mirroring each other almost all the time. This is beneficial because knowledge about one of them can give knowledge about the other. This comes in handy when one of them has more parameters for a more accurate prediction.

To my knowledge, treasury bonds have an inversely related relationship with stocks because when the economy is tumultuous or actually in bad shape, bonds offer a safer investment. So, the fact that they are inveresly related actually makes sense

11 Question 10

11.1 Steps

11.1.1 Mean Daily Return

(Similar to question 7)

The growth rate is dependent on having a value at time t and a value at time $t-1$ or $t+1$ (depends on what time t signifies). To find the rate at which the price is evolving, I can use the formula

$$r(t) = \frac{p(t)}{p(t-1)} - 1$$

where $p(t)$ is the current number, $p(t-1)$ is the previous number, and subtracting 1 because the mother formula is

$$r(t) = \frac{p(t) - p(t-1)}{p(t-1)} = \frac{p(t)}{p(t-1)} - \frac{p(t-1)}{p(t-1)}$$

The *Matplotlib* pacakge in Python has can perform this without me having to apply this formula to every data point.

11.1.2 Minimum and Maximum Daily Returns

After finding all the daily rates, the minimum and maximum daily returns would be at some time t_{min} and t_{max} (referring to the time at which there is a minium and maximum value for the rate), there will be some minimum and maximum rate for that time.

11.2 Results

Type	SPY	TLT
Mean Daily Return	0.03%	0.06%
Minimum Daily Return	-4.21%	-2.43%
Maximum Daily Return	3.84%	2.65%

11.3 Insights

First, the TLT ETF has a higher mean daily return than the SPY. The minimum daily return of the SPY is larger (more negative) than that of TLT. The maximum daily return of SPY is greater than that of the TLT.

However, the SPY is riskier than the TLT because the difference between the minimum and maximum daily returns is about 8.05% whereas that of the TLT is about 5.08%. Combining this with the fact that TLT has an average daily return of 0.06% vs 0.03% of the SPY, the TLT seems to be the better option.

As I mentioned in Question 9 Insights– treasury bonds are safer than stocks, this is a very appealing investment; this is a sweet spot for an investor because combining the safety of Treasury Bonds and the higher average rate of return, the TLT is much more appealing.