

CARNEGIE MELLON UNIVERSITY
DATA, INFERENCE & APPLIED MACHINE LEARNING (CDA 410)
ASSIGNMENT 1

INSTRUCTIONS

- Submissions should be made via canvas.
- **Single** Python/MATLAB code file(.ipynb or .m) **[Do not Submit checkpoints for .ipynb]**. In addition, each line of code should be documented by text. This demonstrates that the code is unique and owned by the student.
- Assignment report(.pdf) with full evidence that the assignment was completed by the student and demonstrate a full understanding of each step in the process including textual descriptions of each result (statistics, table, graph etc) represents and insights that can be gained.
- Indicate the libraries you have used in your code at the beginning of the report (After the title page).
- Using ChatGPT for any assignment is not allowed as it could lead to being flagged for plagiarism.
- Data files (as given).

Submission process:

1. Put source code **file and data files** in a single folder
2. Name of the folder should be the same as your andrew ID
3. **Zip this folder and attach the zipped file to the submission page**
4. After attaching zipped file, click on "Add Another File" from assignment submission page and **attach your report**
5. Submit your assignment

N.B. This new process will allow us to compile your reports in **Turnitin** to check for plagiarism.

Specific reasons for a submission being classified as incomplete include:

- Failure to correctly name your folder with your Andrew ID, report, and code file with andrewID_DIAMLE_assignment1. For example, mesharry_DIAMLE_assignment1, mesharry_DIAMLE_assignment2 and mesharry_DIAMLE_assignment3.
- A missing report describing the steps, results, and insights
- A missing dataset required for running the code
- A missing code file such as .ipynb or .m file
- An error in the file path needed to run the code

The student is responsible for checking that their submission is complete. Students will lose 10% as for late submission even if the submission is repaired during the 24 hours after the deadline has passed, and receive 0 for the assignment if it is not repaired.

The submission deadline is Eastern Time (ET) **Monday 04 September 2024 11:59 PM**

No.	Question	Format	Value
1	A piece of paper is 1mm thick. Assuming you can fold it as many times as you want, how many folds would it take to exceed the height of Mount Everest at 8,848 m?	Number	10%
2	The volume of water in a reservoir decreases at an exponential rate, following $v(t) = v(0)\exp(-at)$ with $a=0.1$. How much time, t , does it take for the volume to decrease to less than one half of its initial volume, $v(0)$?	Number	10%
3	If you deposit \$100 in a bank account that offers an annualized interest rate of 5% (compounded annually), how much money will you have (round to the \$) after one, two, three, four and five years?	Five Numbers	10%
4	Suppose you want to buy a car worth \$20,000. A financial institution can provide a loan with a monthly interest rate of 1%. What is the monthly payment to pay off the debt in one, two and three years (rounded to the nearest \$)?	Three Numbers	10%
5	You are about to set up a new business and will invest \$100,000. On day one you expect to have 100 customers and the number of customers will grow at a rate of 1% per day. If each customer provides profits of \$10, how many days will it take to repay your initial investment based on cumulated profits? Plot cumulated profits per day, show initial investment and mark breakeven day.	Graph Number of days	10%
6	Using data from http://bit.ly/1JJyf29 and linear interpolation, estimate the dates when the number of cases and deaths due to Ebola exceeded 100, 500, 1000, 2000 and 5000. Graph the cases and deaths (observations and interpolations) and mark the dates when thresholds were exceeded with a circle.	Graph Ten dates	10%
7	Using data from 2014, downloaded in the previous question, what is the average growth rate per day, as a percentage, in the number of Ebola cases and deaths?	Two Numbers	10%
8	Using the same data, plot the number of deaths versus the number of cases and estimate the average ratio of Ebola deaths to cases.	Graph Ratio	10%
9	Obtain daily prices for two ETFs called 'SPY' and 'TLT' which track the S&P500 index and long-term Treasury Bond. Select the adjusted closing prices. Plot the two time series during 01/01/2014 – 08/31/2015 and make them comparable by starting from prices of \$100 on the first day in 01/01/2014 – 08/31/2015.	Graph	10%
10	For the ETFs on the previous question, calculate daily returns, $r(t) = p(t)/p(t-1)-1$, for each trading day in the same time period as above. Calculate the average, min and max daily return for each of the two ETFs during the time period and express these as percentages.	Six numbers	10%