Data Analysis

1. **Titanic Data Analysis Project**
   The sinking of the Titanic is one of the most infamous shipwrecks in history. We analyze the train dataset and answer 5 questions.
   1) Who were the passengers on the Titanic? (Age, Gender, Class…etc.
      → check the gender, then separate the gender by class
      Moreover, we take the person under 10 years old as children, then separate them to female, male and children
      → to visualize the data using 'sns.factorplot' and 'sns.FacetGrid' to plot multiple kedplots on one plot
   2) What deck were the passengers on and how does that relate to their class?
      → drop missing value
      make a list for the Cabin level then using 'sns.factorplot' to visualize them
   3) Where did the passengers come from?
      → count the value 'Embarked' categorized by 'C, Q, S', then separate them by class and visualize it
   4) Who was alone and who was with family?
      → separate them by adding the 'SibSp' and 'Parch', and test that if them are larger than 0. then separate them by class and visualize it
   5) What factors helped someone survive the sinking?
      → from Class Factor, Age factor and Deck factor to analysis it

2. **Stock Market Analysis Project**
   In this project we will be looking at data from the "Apple, Google, Microsoft and Amazon" stock market over the past year. We will use pandas to get stock information, visualize different aspects of it, and finally we will look at a few ways of analyzing the risk of a stock, based on its previous performance history. We will also be predicting future stock prices through a Monte Carlo method.
   We analysis the data from aspects:
   1) Basic analysis of stock information
      → closing price, total volume of stock, moving average for the stock
      In this section we'll handle requesting stock information with pandas and analyze basic attributes of a stock.
   2) Daily return Analysis
      We're now going to analyze the risk of the stock. In order to do so we'll need to take a closer look at the daily changes of the stock, and not just its absolute value.
      → Analyze the individual stock daily return
      Analyze the returns of all the stocks in our list → build a Data Frame with all the ['Close'] columns

Compare the daily percentage return and closing price of two stocks to check how correlated using ‘sns.jointplot’ or ‘sns.pairplot’ (visual analysis of all the comparisons)

Plot the actual numerical values for the correlations between the stock’s daily return values and closing price

3) Risk Analysis
   There are many ways we can quantify risk, one of the most basic ways using the information we've gathered on daily percentage returns is by comparing the expected return with the standard deviation of the daily returns.
   → Apple is the best stock (strong expected return with lower risk)

4) Value at Risk
   a) Bootstrap Method
      For this method we will calculate the empirical quantiles from a histogram of daily returns.
   b) Monte Carlo Method
      Using the Monte Carlo to run many trials with random market conditions, then we'll calculate portfolio losses for each trial. After this, we'll use the aggregation of all these simulations to establish how risky the stock is.