

Programing in python.

1. **Database system**
 - a. Done together in working session.
2. **Data transformation, visualization, and summary statistics**
 - a. Packages: *Pandas*, *NumPy*, *SciPy*, *Statsmodels*, *Matplotlib*
 - i. 25-minut YouTube tutorial [HERE](#)
3. **Vectorized operations**
 - a. Packages: *Numpy*
 - i. 8-minute YouTube tutorial [HERE](#)
4. **Multithreading**
 - a. Packages: *Concurrent Futures*, *ThreadPool Executor*
 - i. 10-minute YouTube tutorial [HERE](#)
5. **Producing plots**
 - a. Packages: *Matplotlib*
 - i. Examples [HERE](#)
6. **Building indices**
 - a. Value weight
 - b. Equal weight

Testing Your Abilities:

1. **Retrieve data from database.**
2. **Use *Pandas* & *Numpy* to manipulate the returns data.**
 - a. Calculate some summary statistics (average return, variance, standard deviation, etc.)
3. **Produce visualizations using *Matplotlib*.**
 - a. Create a histogram of returns.
 - b. Plotting prices
 - i. Plot 2 stocks on one chart, where the y axis are independent.
 - ii. Build an equal weighted index using the returns of 2 stocks.
 - c. Use rolling windows to calculate the average historical return and variance with a 12 month look-back.
 - i. Plot this value over time using a line chart. Be sure to include legends and annotations where needed. (just ask *ChatGPT*).
4. **Perform a regression analysis between 2 sets of returns.**
 - a. One where you regress R_t^i to R_t^j
 - i. This representation the direct cross-sectional regression that is commonly used for exposure factors like market beta.
 - b. One where you regress R_{t-1}^i to R_t^j
 - i. In this one, we are using R_{t-1}^i to predict the future price of R_t^j . This is what we will use for estimating expected returns in factor modeling.
 - c. Interpret the statistical significance of your findings.
 - i. What is the regression equation for these two variables?
 - ii. Note your conclusion – is the y-intercept different from zero? Is the regression coefficient β different from zero? Is the error term ϵ constant or not? What is the R^2 ?
 - iii. Create a scatter plot of 2 sets of returns from the regression analysis (R_{t-1}^i to R_t^j) and superimpose the regression line using *Matplotlib*. Tutorial [HERE](#).
5. **Use *Concurrent Futures* to speed up the operations that would normally be done in a loop.**
 - a. This is for conceptual purposes only, the transformations you decide to do here are up to you. The Objective is to apply multithreading to any function. This will become very useful later.