Technical Questions

Instructions

You are tasked with solving 3 technical problems including data extraction, manipulation and risk calculation and analysis.

We recommend to code Problems 1 and 2 in Jupyter notebook but you are free to use other languages too.

For problem 1, please consider the following aspects of your solutions:

- Maintainability - could someone easily modify/fix your code?
- Style - is the code readable and consistent with best coding practices?
- Reusability - is the code easy to adapt to solve similar problems?
- Documentation - are there meaningful comments in the code?
- Modularity - how easy would it be to integrate your script into a carious other processes?

Google is your friend - feel free to look things up, but please keep in mind that you will need to explain how you solved the problem and why you did it in this particular way.

Please read **note** carefully as they contain important information.

We will ask you talk us through your code during the next interview, please be prepared to cover the following points:

- Summary of problem - For audience who don't know this exercise
- How do you structure your solutions to solve the problem (code the solution)?
- What difficulties you encountered and how did you overcome it?
- If allowing more time, where you could/wanted do it better?

Problem 1: CSV Data

A fund of funds has allocation to five funds: A, B, C, D, and E. An automated process stores the total fund's allocation snapshot on a monthly basis to a csv file. These files are stored in **Exercises/Problem 1/Fund Allocations/**, with each report being stored in its own folder specifying the date on which it was created (i.e. report date).

The structure is as follows:

```
.../Problem 1/Fund Allocations/[report_date]/Fund_Allocations_[valuation_date].c
sv
```

E.g.

```
Fund Allocations/
├──20170102/
│      Fund_Allocations_20170101.csv
├──20170202/
│      Fund_Allocations_20170201.csv
├──20170302/
│      Fund_Allocations_20170301.csv
├──20170402/
│      Fund_Allocations_20170401.csv
├──20170502/
│      Fund_Allocations_20170501.csv
├──20170602/
│      Fund_Allocations_20170601.csv
├──20170702/
│      Fund_Allocations_20170701.csv
├──20170802/
│      Fund_Allocations_20170801.csv
├──20170903/
│      Fund_Allocations_20170901.csv
```

Each report file has the following structure (data can differ):

Fund	Allocation
Fund A	50
Fund B	75
Fund C	100
Fund D	150
Fund E	185

Business Problem: The team is often asked to provide time series information (e.g. Fund A allocation over time).

You are asked to write a Python script that combines all reports and saves them to a new csv file.

The final output should look like (data can differ):

Valuation Date	Fund	Allocation
2017-01-01	Fund A	50
2017-01-01	Fund B	75
2017-01-01	Fund C	100
...	...	...
2019-12-01	Fund C	700
2019-12-01	Fund D	400
2019-12-01	Fund E	185

Notes:

- It is not known how many files there are, we only know there are more than **35**, so the script should work for an arbitrary number of files
- Report file names always follow the same structure: Fund*Allocations*[ValuationDate].csv
- The valuation date for each report is specified in its filename in YYYYMMDD format
- ***your colleague may somehow mess up a few folder names***
- Only one file exists for each valuation date, so duplicate records are not an issue
- Often the *report_date* is simply *valuation_date + 1*, but that is not always the case
- You are free to use any Python Packages, but as a hint you should manage with OS and Pandas

In [2]:

```
# please write your solutions here
```

Problem 2: Estimate VaR95

In the above figure, we know: CVaR_92.5 = 15000 This means Conditional Value at Risk (Expected Shortfall) at 92.5% confidence level. CVaR_97.5 = 25000 This means Conditional Value at Risk (Expected Shortfall) at 97.5% confidence level.

Questions:

1. Calculate the purple area
2. Estimate VaR_95

In [ ]:

Problem 3: VaR calculation and Timeseries investigation

Consider the following funds' returns, and assume they follow normal distribution and their mean and standard deviations are: Mean, Annual Std: A: 1%, 10% B: 3%, 20% C: 1.5%, 15% D: 2%, 25% E: 1.2%, 13%

Assume the correlation between returns are: Fund A and C is 0.3 Fund B and D is 0.4 With no correlations between all others

Your colleague did a Monte carlo simulation of returns for these 5 funds and saved the results in Excel "Fund_Performance.xlsx" in "Problem 3" folder.

Questions - For all questions here, you can just do it in Excel or any application you like (Matlab or Python).

1. Could you please calculate Var_95, CVaR_92.5 and CVaR_97.5 for these 5 funds respectively from the ***simulated returns?
2. Could you please calculate *Parametric* Var_95, CVaR_92.5 and CVaR_97.5 for these 5 funds respectively?
3. Estimate VaR_95 using methods in Problem 2 and do a comparison to historical and parametric VaR_95.
4. Assume a portfolio composed of funds A to E equally weighted and rebalanced daily and please estimate VaR_95 and CVaR_95 of the portfolio.
5. Bonus point - *a. Check if your colleague did the Monte Carlo simulation of returns correctly from the results* b. How can you fix it? - Just describe what's the problem and how to fix it should be enough, not necessarily code it.

In [ ]: