

Data Science Challenge

WELCOME

Thank you for participating! We're very excited to see your results. To ensure that your submission is valid, we ask that you please:

- Submit your solutions by Thursday, May 23, 2019 3:56 PM GMT. You have 7 days, 0 hours, and 0 minutes remaining (Please let us know if you need more time!)
- Submit only your own individual work
- Submit your answers on this page
- Show all your work; we recommend a well-documented Jupyter or R notebook (you can upload any notebooks and code files at the bottom of this page)

If you encounter any difficulties (starting the Challenge, loading data into memory, etc.), please reach out to us at DataScienceChallenge@twosigma.com. We're happy to help!

GETTING STARTED

This Challenge consists of two parts:

- A dataset of stock market returns
- A dataset of spending by employees of the State of Oklahoma

You will be asked to answer several quantitative and qualitative questions. Answering these questions will require understanding and manipulating the data, as well as being familiar with standard statistical techniques.

[Download Stock Returns Dataset](#)[Download Oklahoma Spending Dataset](#)

If you have any questions or need anything at all, please email us at DataScienceChallenge@twosigma.com.

A few helpful tips:

- Dealing with messy data is part of the Challenge
- We recommend that you take the Challenge using tools with which you are most comfortable. If you have not previously done data analysis on your own computer, both [Jupyter Notebook](#) and [R Studio](#) are free for personal use. Setting up a Python environment can sometimes be tricky; we recommend [Anaconda](#) for people who are new to the Python ecosystem

SECTION 1: STOCK MARKET DATA

Momentum and Mean Reversion

Momentum and **mean reversion** are common trading strategies. For the purposes of this Challenge, an asset that exhibits momentum is defined as an asset whose price returns are more likely to go up (down) on day t if the return went up (down) on day $t-1$. In other words, the stock exhibits a positive autocorrelation. Mean reversion is the opposite of momentum. Stocks are more likely to go up (down) on day t if that stock went down (up) on day $t-1$.

You are provided below with a simulated dataset of a series of stock returns. These returns have been generated with a predetermined average momentum during one period and a predetermined average mean reversion in another period.

Please use this dataset to answer the questions below. In order to do so, you will need to clean the dataset. It comes with a number of flaws commonly seen in datasets we receive. In the interest of time, please feel free to drop any records that are problematic.

Questions:

1. In what month did the returns shift from exhibiting mean reversion to exhibiting momentum, or from exhibiting momentum to exhibiting mean reversion? Before this month, stocks exhibit one pattern (on average), and after this month stock returns exhibit another (on average).

Please submit your answer in the format *mm-yyyy* (ex: 01-2018).

2. During the time period when these stock returns had the momentum property, what was the average momentum? Please note this is a single number: the average across all stock returns in the time period.

3. During the time period when these stock returns exhibited mean reversion, what was the average mean reversion? Please note this is a single number: the average across all stock returns in the time period.

4. Why might the market shift from momentum being dominant to mean reversion being dominant (or the other way around)?

5. Why might trading momentum or mean reversion succeed as a strategy? Why might it fail?

SECTION 2: OKLAHOMA STATE SPENDING

This dataset comprises individual credit card transactions by employees of the State of Oklahoma. Theh dataset is a slightly altered version of the raw data found [here](#). Please use this dataset to answer the questions below. Answering this questions requires cleaning and manipulating the dataset. Please document any methods you use to deal with messy data.

Questions:

1. What is the total amount of spending captured in this dataset?

2. How much was spent at WW GRAINGER?

3. How much was spent at WM SUPERCENTER?

4. What is the standard deviation of the total monthly spending in the dataset?

5. Please describe the process you would follow to build a model on this dataset to make predictions about the stock market. Please note this is a hypothetical only - there is no need to build an actual model.

6. What biases might this dataset have if you tried to use it to model equities?

7. (Optional) Do you have any other observations about this dataset?

SECTION 3: FEEDBACK

The following questions are not used for assessment purposes.

1. (Optional) How long did it take you to complete this Challenge?

2. (Optional) Any ways we can improve this Challenge? We welcome your feedback!

3. (Optional) At Two Sigma, we find patterns, meaning and relationships in the world's data to create value for our investors. We are constantly on the lookout for new and interesting datasets. If you know of a dataset that might be of interest to Two Sigma, please describe it here (with any relevant links).


Save

Please be sure to save your work before leaving this site.

FILE UPLOAD

Please upload workbook(s), code, and any other files relevant to your work. To submit multiple files, please package them in a .zip file. Only the final file uploaded will be considered.

No file currently uploaded.

 Upload file

Submit Challenge

Please contact DataScienceChallenge@twosigma.com for support.

Property of Two Sigma. Proprietary and Confidential - Not for Redistribution. [Privacy Policy](#)