# Sunrise Futures 2017 Financial Data Modeling Competition Instructions

## Explanation of data

The data file you received is in plain text, in CSV (comma separated value) format. The data covers three instruments over a three month period and has data at one second intervals. The columns are as follows:

1. **Date** – The date as an integer starting at 0.
2. **Time** – The time as an integer.
3. **bid\_price\_inst1** – The best bid price of the first instrument.
4. **bid\_size\_inst1** – The best bid size of the first instrument.
5. **ask\_price\_inst1** – The best ask (AKA offer) price of the first instrument.
6. **ask\_size\_inst1** – The best ask (AKA offer) size of the first instrument.
7. **mid\_price\_inst1** – The average of the best bid and ask of the first instrument.
8. **net\_traded\_inst1** – The net sum of the buys (positive) and sells (negative) over the preceding one second of the first instrument.
9. **bid\_price\_inst2** – The best bid price of the second instrument.
10. **bid\_size\_inst2** – The best bid size of the second instrument.
11. **ask\_price\_inst2** – The best ask (AKA offer) price of the second instrument.
12. **ask\_size\_inst2** – The best ask (AKA offer) size of the second instrument.
13. **mid\_price\_inst2** – The average of the best bid and ask of the second instrument.
14. **net\_traded\_inst2** – The net sum of the buys (positive) and sells (negative) over the preceding one second of the second instrument.
15. **bid\_price\_inst3** – The best bid price of the third instrument.
16. **bid\_size\_inst3** – The best bid size of the third instrument.
17. **ask\_price\_inst3** – The best ask (AKA offer) price of the third instrument.
18. **ask\_size\_inst3** – The best ask (AKA offer) size of the third instrument.
19. **mid\_price\_inst3** – The average of the best bid and ask of the third instrument.
20. **net\_traded\_inst3** – The net sum of the buys (positive) and sells (negative) over the preceding one second of the third instrument.
21. **y\_value** – This is the y value for the linear regression (i.e., the value you are trying to model/predict).

## Goal

The objective is to use the fields in columns 1 through 20 to try and predict/explain the value in column 21. The measure of success will be the r-squared of a linear regression. You can manipulate or combine the data in any way of your choosing.

Note that r-squared is defined as [Coefficient Of Determination](https://en.wikipedia.org/wiki/Coefficient_of_determination) and is computed as follows:

R2=1− [ ∑i(yi−zi)2 / ∑i(yi−μ)2 ]

Where y is the actual value, μ is the mean of the actual values, and z is the predicted value.

Note that in the financial space even small r-squared values can be very meaningful, so don’t be discouraged by low values.

## Rules

Violating any of these rules will result in an immediate disqualification:

1. You cannot use column 21 (y-value) in your model at all.
2. You cannot use any data from a future row. In other words, to predict the y-value in line X you can use any data in the lines up to and including X, but no data from rows after X.

## Submission

In order to compete in the contest, you must submit the following:

1. A short paper/summary explaining what you did and what you found.
2. The code used to build your model. This code must be in one of the following:
   1. C/C++
   2. Java
   3. Python
   4. Perl

Your code will be used to verify your results and approach, but it will also be used to generate results over an out-of-sample data set, so please make sure that it is readable, runnable and understandable.

Email all of these items to [competition@sunrisefutures.com](mailto:competition@sunrisefutures.com) by midnight on Monday, April 17 in order to participate.

## Results/Questions

The winning teams will be selected based on the results and explanation of their findings as well as the out of sample performance. The prize will be presented on campus on Tuesday, April 25.

Please send any questions to [competition@sunrisefutures.com](mailto:competition@sunrisefutures.com).