

The Signals V2 “Cosmic” dataset is officially released. Download it [here](#).

[

cosmic_art

1600×1600 509 KB

](https://forum.numer.ai/uploads/default/original/2X/3/38f52531475822a1b382f7565dc90f798702ac91.jpeg)

Similar to the [V5 Atlas](#) dataset for the Numerai tournament, the Cosmic release focuses on universe expansion - including even more stocks than the Atlas dataset. We also substantially improved our numerai_ticker and the country feature column.

In the Atlas forum post, we showed v5 models perform much better than v4 models solely due to the larger universe. The case is similar for Signals V2 data. Here are the diagnostics for V1 example validation predictions scored against the V1 dataset:

[

v1_diagnostics

1494×1482 181 KB

](https://forum.numer.ai/uploads/default/original/2X/9/96cca2bc9e69b45331fcbd66b7f20e915d08d7f6.png)

And here are the diagnostics for V2 example predictions scored against the V2 dataset:

[

v2_diagnostics

1494×1486 182 KB

](https://forum.numer.ai/uploads/default/original/2X/5/5d41d1a8281ec857e65a82a183d8b0623de31729.png)

You can see that there is an increase in the mean and Sharpe for 3 of the primary metrics. No change to the model or training parameters, just a larger universe. Note: while churn of our example model does increase, we believe this is negligible and we know that churn is relatively easy to control.

Below, you can see that the ticker universe has greatly expanded from around 5000 stocks in recent years to well over 6000. This is at least a 20% increase for all eras in the last decade:

[

cosmic_universe

984×584 68.7 KB

](https://forum.numer.ai/uploads/default/original/2X/2/26bec36596b7a9de03cc3272ef55663b1ff9d473.png)

We've added several new countries to the dataset including Chile, China, Colombia, Egypt, India, Morocco, Peru, Qatar, Russia, and UAE. We also fixed several consistency issues with our ticker and country columns. Countries now have consistent codes throughout the entire dataset, and notations such as shareclass are either made consistent or removed altogether if unnecessary.

Signals V2 data will be used for Signals submissions starting Dec 3, 2024. Our records indicate this would not significantly affect the community since we do not require full coverage of the Signals universe.

Signals V2 data will be used for Signals scores in rounds starting on or after Jan 1, 2025.

Furthermore, with this release we are officially deprecating Signals V0, V1, and the ticker map files. These files have been de-listed from the website, but we will allow users to continue downloading these files via the API until Jan 1, 2025. On this date, these files will be removed and will no longer be available for download. They will not be updated each round and the historical files will be deleted. We are unable to provide alternatives.