Factsheet

SamKnows One – Analytics

# Analytics

## Core analytics
Our core analytics functionality allows you to chart metrics in ways that make the data useful and easy to spot changes and make comparisons; before presenting or sharing that data within your organisation.

## Performance metrics
At SamKnows we've devised metrics based upon individual test results in order to give us the maximum amount of useful data from individual test runs. For example from our continuous UDP latency test we derive packet loss, jitter upstream, jitter downstream, latency, MOS and a series of other more advanced metrics. A full list of these metrics commonly available can be seen on our test methodology documentation.

In SamKnows One, these metrics are the foundation of what you're charting so you don't need to have intimate technical knowledge of the test's output data in order to use it.

We provide an interface for both raw and aggregated metric data.

## Line charts
Line charts aggregate data and plot one or more data series as lines on a graph. Time is plotted along the x-axis, and the metric value is plotted along the y-axis.

Each data series is plotted with a different colour. When plotting a time series line graph, users are also prompted to select how they wish to aggregate data by time. Users may aggregate by hour (or groups of hours), day, week and month. Users may also aggregate by "hour of day", which groups all test results from the same hour (across multiple days)

into a single point, thus producing a graph plotting the "typical day".
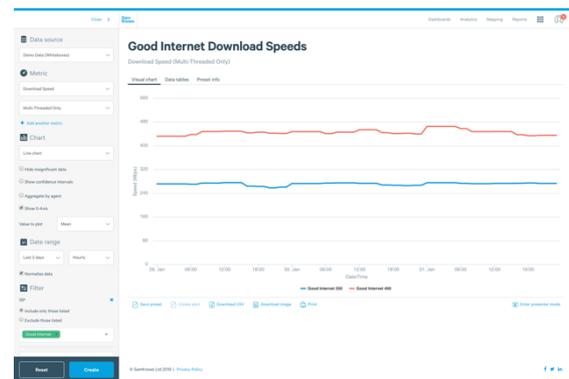


## Scatter plot
The scatter plots plot each individual measurement result on the graph. Time is plotted along the x-axis, and the metric is plotted along the y-axis. Each data series is plotted with a different colour.

Scatter plots should be used with caution, as the large amount of test results transmitted back to the client (graphs are rendered client side) may cause web browsers to struggle render the load. SamKnows generally suggests limiting raw data scatter plots to a few days.
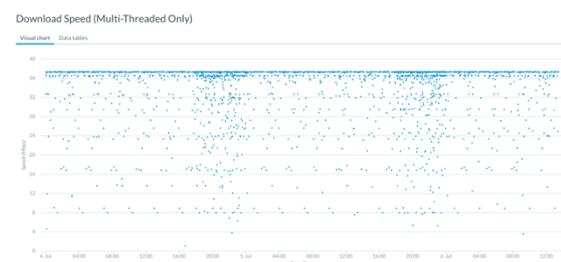


## Bar charts
Bar charts can be useful to provide an at a glance view of averages or other values when comparing different data series. The value being plotted is displayed on the y-axis and

the x-axis displaying the split of the data that each bar represents (for example product).



Download Speed (Multi-Threaded Only)
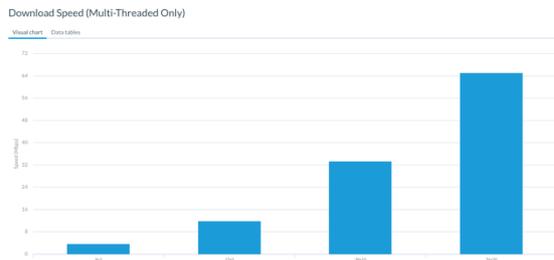
### Heatmaps

Heatmaps plot test results on a chart similar to the scatter plot, giving higher colour intensity to clusters of results. This may provide a more visually appealing alternative to the scatter plots. Time is plotted along the x-axis, and the metric is plotted along the y-axis. Only a single data series may be plotted on a heatmap.



Download Speed (Multi-Threaded Only)
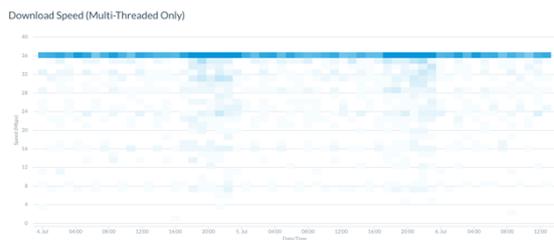
### Summary tables

The table chart type aggregates all results from the metric chosen and prints out selected data (such as averages, sample counts, percentiles and standard deviations), broken out by how you select to split the data (for example by product).

When a user uses the table chart, the selected metrics and values to display (such as average, sample count etc.) are displayed in the columns, whilst the desired splits (such as package) are show in the rows.

Download Speed (Multi-Threaded Only), Upload Speed (Multi-Threaded Only) split by Package / Product / Tier

| Package | Download Speed (Multi-Threaded Only) | | | | Upload Speed (Multi-Threaded Only) | | | |
|---|---|---|---|---|---|---|---|---|
| | 10th Percentile | 20th Percentile | 80th Percentile | Sample Count | 10th Percentile | 20th Percentile | 80th Percentile | Sample Count |
| Acme 17 | 3.74 Mbps | 5.88 Mbps | 16.78 Mbps | 22159 | 0.60 Mbps | 0.67 Mbps | 1.04 Mbps | 20490 |
| Acme 38 | 23.33 Mbps | 30.41 Mbps | 37.35 Mbps | 32667 | 4.09 Mbps | 6.18 Mbps | 9.33 Mbps | 32053 |
| Acme 76 | 46.45 Mbps | 56.56 Mbps | 74.67 Mbps | 20094 | 9.46 Mbps | 15.53 Mbps | 18.68 Mbps | 18521 |

### Aggregation: mean average

Many of our chart types allow you to select how you wish to aggregate data, and also

different values to return [functions] such as averages and percentiles. SamKnows One can either aggregate by all test results equally weighted, or take the average of each agent and then aggregate on top of that giving all measurement agents equal weighting, no matter how many test results they've reported. Our most standard aggregation is plotting the mean (average) of all the results.

SamKnows One allows users to aggregate by many standard time-series periods such as by hour (or groups of hours), day, week and month. Users may also aggregate by other things such as "hour of day", which groups all test results from the same hour (across multiple days) into a single point, thus producing a graph plotting the "typical day".

The following aggregations are available that aggregate by time in chronological order:

- hourly
- two_hourly
- four_hourly
- six_hourly
- daily
- weekly
- monthly
- quarterly

The following aggregations are available that use an element of time (e.g. the date or the hour of the day)

- hour_of_day
- day_of_week
- day_of_month

The following other aggregations are available:

### Presenter mode and exports

SamKnows One is designed to be able to help you explore and analyse the data you need, and then pull it out for presentations or reports as may be necessary.

Therefore users can export almost all our charts as high-resolution images of the chart image for sharing in documents or emails.

Both our line and scatter plot charts also allow for exporting of CSVs of data being displayed. There is also an option to print charts that will generate an appropriate chart image and take you to your computer's print dialog.

We also have a presenter view where any chart that is plotted may be shown full screen, without any window titles or menus, which makes it well suited for presentations or display on a screen. To enter presenter mode, click the "Presenter mode" button at the bottom right of all charts. To exit, press Esc on your keyboard.

## Data normalisation

Users can optionally enable a feature called "normalisation" when using time series line charts. This is useful when plotting results by hour, but the test schedule does not run test results every hour across every agent. Without normalisation, users will often see a saw-toothing effect, with some hours having measurement agents with much better results than other hours.

Normalisation forces SamKnows One to look at the underlying test schedule for the metric in question, and to normalise results across the blocks of hours that the tests should run over. This results in an even number of measurement agents and results being represented for all hours of the day, thus removing the spikes in the charts and making it easier to see overall trends.

## Prefiltering

SamKnows One allows users to enable or disable prefiltering. Prefiltering is a mechanism in order to try and remove data that is unhelpful and likely to be anomalous.

There are a series of custom SamKnows defined prefiltering rules per metric that aim to remove results we believe to be impossible or incredibly unlikely results, for example a recorded latency of less than 0.5ms would likely not be a valid measurement agent on a home network.

We also allow clients to add specific CPE and Whitebox measurement agents to be prefiltered from results. This is useful for abusive or lab test agents.

Finally we also allow you to filter out likely insignificant results on line charts, qualifying results would be the most recent 2 or 3 aggregate results if they have very low sample/agent counts compared to other data points and are significantly different from all other points (are the minimum or maximum point). This helps filter out results from incomplete recent hours/days where the result may be skewed.

## Filtering of data by date

You can filter down a specific date range or use one of our default ranges for example 'Last 7 days' which will update when used in presets.

## Filtering of data by time

You can also filter down to a specific time on a date, instead of just being able to filter to whole days

Filtering of data by custom metadata
Filters allow test results to be included or excluded in the report based upon one or more filter criteria. Users can use "include" or "exclude" filters, and can choose one or more values from a list. Multiple filters can be used simultaneously.

Users can filter by any of the common metadata fields that SamKnows associates with measurement agents and results as well as custom metadata fields provided by a client such as exchange, CMTS, geography or firmware version. Common splits/filters provided by SamKnows for fixed broadband for example would include measurement target, product, and ISP. Further there may be filters specific to a metric such as thread count, IP version or target. Finally you can also filter down to view analytics results of single specific measurement agents.

## Splitting of data

Splits define how results will be split up on the resulting chart. Each combination of splits effectively creates a new data series on the resulting chart. This might, in the case of line series charts, For example, a user may choose to split by US State, and would therefore see lines on the chart for California, New York, Texas, and so on. Combining filters and splits together allows users to report on and compare any combination of results that they wish.

## Saved chart presets

Users can save chart configurations as named 'presets' that are either visible just to themselves or to everyone in their organisation. Presets allow a quick and easy way to go back to previously created charts. They also form the basis for alerts and you use presets to build dashboard pages. These presets can also be updated and deleted at a later data. They are accessible from the main analytics screen in SamKnows One where they are prominently featured and simply clicking on one will load that chart configuration and render the chart immediately.
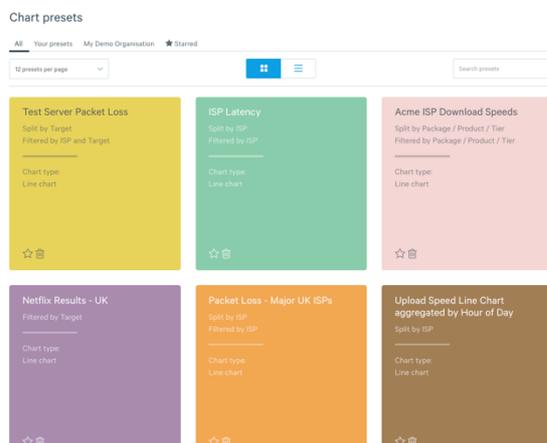


## Sharing of presets with organisation

Presets can be shared within an organisation or, the default, to only be accessible to the person who created them.

## Advanced analytics

Whilst core analytics gives you insights into your data, advanced analytics takes this to the next level providing more statistical methods of analysing your data so you can delve into

diagnosing issues, provide insight for analysts and probe SamKnows One to answer specific questions for your marketing teams.

## Plot multiple datasets on a chart

In addition to being able to split data as normal, you can also add to an existing graph entire new data series or graphs to overlay complete with their own sets of filters, splits. This allows you to compare metrics (for example latency versus download), different more complex data sets, overlay different chart types, or even compare data from different sources (such as mobile latency on Wi-Fi versus web based test latency).

## Contextual environmental information

Contextual environmental information will show you some pieces of environment data collected at the time of the test running when hovering on points on raw charts (e.g. Raw Map or Scatter Plot) and in-depth environmental information when doing graph inspection.
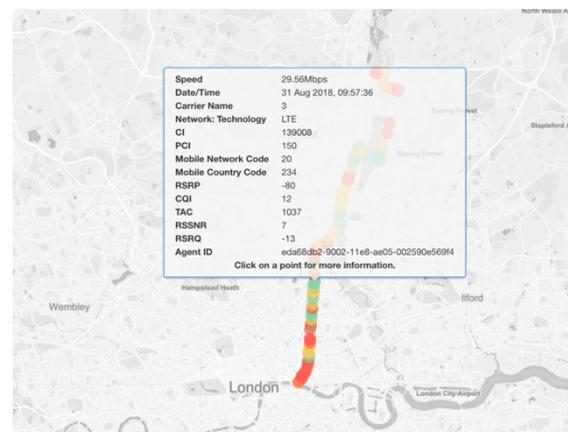


## Confidence intervals

Line graphs can also be enabled to display zones for confidence intervals, and allow for inspection of specific points to view scatter plots of the test results contributing towards an aggregated point.
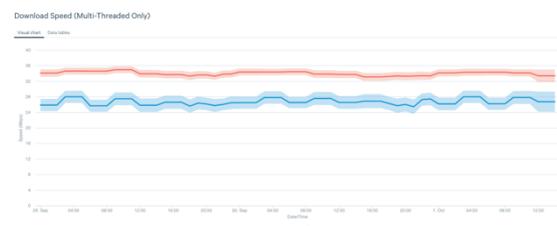
## Agent or test based aggregation NEW

When aggregating by test (the default), it will perform a simple aggregation over all of the tests run in each bucket (e.g. hour) however when aggregating by agent, it will first, for each bucket, aggregate all of the results for each agent before then performing an aggregation over those agent results. This helps to prevent an agent that reports more or less often than others from skewing averages and is commonly required for using data in an official capacity such as regulation or marketing.

## Aggregation

Percentiles:
In addition to aggregating by the mean average value, you can also plot percentiles with Advanced Analytics.

Minimum and maximum:
You can plot the minimum and maximum values of a dataset with Advanced Analytics.

Standard deviation:
You can plot the population standard deviation of a dataset with Advanced Analytics.

Interquartile range:
You can also plot the magnitude of the interquartile range, demonstrating the spread of values that 25% versus 75% of measurement agents or tests achieved.

## Plot sample and agent count

You can plot sample and agent counts, allowing you to monitor and visualise the number of reporting measurement agents or test results. This can be a great way to see faults and also add context to information to ensure you have an adequate number of sample/reporting measurement agents.

## Graph inspection

Graph inspection allows you to click on points in order to see more information. For aggregated points it will load a data table of the raw data behind it and a scatter plot. If it's a raw/unaggregated data point it will display more information such as a link to view the

agent in agent administration, environmental information etc.
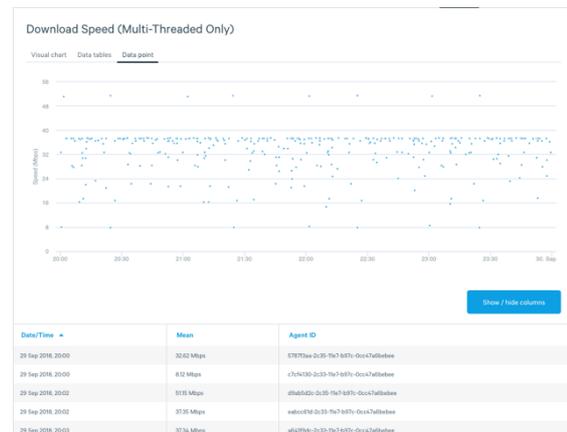


## Range filters

Range filters allow you to not just filter to specific quantitative data but also continuous data as well. This means you can filter by things such as the metric value itself (to exclude outliers for example) or filter down to only show aggregate data points that have a set minimum sample or agent count.

## Box plot

Box plots charts can be useful in order to see the range of results in which most of the test results lie between as they show the median, the upper and lower quartiles, and the minimum and maximum of each data series. Box plots are displayed vertically with multiple box plots being placed next to each other horizontally. The y-axis displays the metric being plotted and the x-axis displays the split of the data that each box plot represents (for example technology or ISP).



## CDF plot

CDF (Cumulative Distribution Function) plots show the cumulative distribution of measurement results. The shape of this curve allows to engineers to identify patterns in the

data (e.g. step changes indicate clustering of results around certain values) and also to immediately recognise whether performance is normally distributed amongst users or not.

Data series are represented as curved lines on the CDF plot. The value on the y-axis indicates what proportion of users achieved at the given value on the x-axis or greater. It is also possible to do a tail distribution where the y-axis indicates the proportion of users who achieved the given value on the x-axis or lower where this may be more appropriate (for example latency).

Download Speed (Multi-Threaded Only)