

## Pendleton Private Preview Update 24 Version: 0.1.1708.021x2

### New Features

- Column metrics view for data flows - recognizes type and shows data profile for each column
- Sampling
- Manage multiple samples
- Notifications for sample completed/canceled/aborted
- Random percentage sampling
- Remote sampling
- Append rows and append columns transforms
- Promote headers from first file or from every file
- Open remote Excel files

### Feedback and Fixes

- Editing a block below the sample block will not navigate to it anymore
- Fix bugs related to Join transform
- Text clustering works on columns with a long name
- Fix bugs in the data source wizard related to automatic suggestions and column offset width
- Edit dialog for inspectors and steps now shows the correct title
- Column widths are save correctly, even if some columns are removed
- Splitting JSON arrays produces columns with correct names
- Can click on panel labels to expand and collapse Inspectors, Dataflows and Steps panels
- Disallow commit for a transform if all columns are deselected
- Fix a path issue in Azure storage browser
- Wizard can detect and recognize more date formats
- Wizard will suggest Excel file type only if all files are Excel

## Pendleton Private Preview Update 23 Version: 0.1.1705.250x2

### New Features

- Supports multiple samples

### Feedback and Fixes

- (Windows only) Copy and Paste are reenabled in the menu
- Can now detect schema from Azure blob storage
- No more error notifications after accepting Handle missing values
- Notifies user when there isn't a header to be used when opening a dataflow
- Gives a dropdown menu of columns names for selecting a column a data type for a column
- Combine columns by example works with example

## Pendleton Private Preview Update 22 Version: 0.1.1705.150x2

### New Features

- Full-table profiling at a glance. To try it, click "(show profiling)" next to the dataflow name at the top of the window. This will show basic summary statistics for each column of the data.
- Error value tooltips now include the original value.
- Add Column By Example now supports numbers written in scientific notation.
- Clicking on the Help menu item with the version number will now copy that version to the clipboard.
- **For Top N sampling of remote plain-text/CSV files, we now only download the amount of data required for the sample, rather than always downloading the whole file.**
- Added Spark Job Parameters into run configuration

### Feedback and Fixes

- Fixed an issue that could prevent scrolling via mouse wheel.
- Fixed an issue that could cause a crash when loading a dataflow that refers to a nonexistent file.
- Fixed an issue where Excel files were not correctly detected
- Fixed an issue where the dialog titles were missing
- Data Profiling Inspector is no longer missing

## Pendleton Private Preview Update 21 Version: 0.1.1704.25032

### New Features

- Updated electron to version 1.6.2
- Option to have first column of DataFlow be the path of the source
- Replace Values transform now accepts 'None' as a replacement value
- It is now possible to create a fork from any of the visual filter transformations. Creating a fork will create a new DataFlow which contains the filtered out rows from the original source.

### Feedback and Fixes

- Improved dependency installation
- Improved error handling for ragged files
- Minor style changes to grid spacing
- Added a suffix to the default file name when saving a recovery file
- Searching for Reference Dataflows now filters to only show file with '.dprep' extension
- Fixed an issue where run configuration list was not updating after creating a new run configuration
- Fixed an issue where Python process could fail if username contained a space character
- Fixed an issue where current activity could be null
- Fixed an issue where dropdown menu could display wrong animation
- Fixed an issue surfaced when filter had an empty string

- Fixed minor issue with grid height when scrolling

## Pendleton Private Preview Update 20 Version: 0.1.1704.06032

### New Features

- It is now possible to handle delimited files with embedded new lines. You can find the option to do so in the Add Data wizard. Please be aware that turning this on will have a significant impact on performance when running in Spark.
- Added a new Expand JSON transform which can take JSON text in a column and expand it into multiple columns.
- We've made several improvements to how we handle multiple files in a single Dataflow. Among these, we will now detect column headers across all files.
- Added the ability to load any text file as a single column with one row per line in the file.
- Fixed-width files are now supported as a data source.

### Feedback and Fixes

- Fixed some issues where Add Column by Example was failing for date and numeric fields.
- Fixed an issue where outlier detection could cause an app crash.
- Fixed an issue where exported Python code would fail if lines were skipped when importing data.
- Fixed an issue where modifying steps before the sample step could cause the app to crash.
- Fixed an issue where it was not possible to filter on an empty value.

## Pendleton Private Preview Update 19 Version: 0.1.1703.16012

### New save model

As previously announced, this release adopts an explicit-save model (like Word, etc.), instead of automatically saving to your .dprep package file whenever you make a change.

- Your changes will not be automatically saved to the .dprep file, unless you explicitly click File->Save (or Ctrl-S).
- If you are closing a package but have unsaved changes, you will be asked if you want to save.
- When you select Open Data Source, it will no longer create a .dprep file for you. The package will not be saved until you explicitly click Save or Save As. If you want to close it, you will be asked if you want to Save.
- If the app exits unexpectedly, the next time you open the package, you will be asked if you want to recover from the recovery file.

Note: Currently, closing the app window using the X on the top right corner will not prompt for saving changes. This will be fixed in an upcoming release.

Recovery files with extensions .dprep.tmp and .dprep.user.tmp will be created for every package that is modified. For every change made to the package, an auto-save will be made to the .tmp files. When an explicit save is done, the recovery files are deleted. If recovery files are found when attempting to open a package, you will have the option to open them and inspect the contents, at which point you can choose to save and overwrite the original package, or save it with a new package name.

When creating a new package by opening a data source, the default name is PackageX.dprep, but it is not actually created until you decide to save it, at which point you will be asked for a package name. If the app closes unexpectedly before it is saved, you can restore your progress by opening PackageX again, and choosing to load the recovery file.

## New Features

- Dataflows can now pull data directly from other dataflows, both within and between packages. Add Data Source and choose "Reference Dataflow" to try it out.

## Feedback and Fixes

- Improved error messages when Write to CSV fails.
- Fixed an issue that could cause dependency upgrades to fail.
- Fixed an issue where certain cell values in tooltips would be truncated.
- Fixed an issue where file encoding wasn't being detected properly.
- Fixed an issue where column headings would be misaligned with values after scrolling to the right.
- Fixed an issue where empty-string values would mislead type inference.
- Fixed an issue where the time series inspector wouldn't display any data containing errors.
- Fixed a crash while trying to apply type inference to a very large number of columns.
- Fixed several stability issues.

## Pendleton Private Preview Update 18 Version: 0.1.1702.28012

### New Features

- SQL Databases (MSSQL and SQL Azure) are now supported as a data source.
- We have added type inference to the Add Data Source wizard. After choosing your file, we will analyze the data and set the types of the different columns. You can always edit this within the wizard itself or after adding your data by editing the "Set Column Types" step.
- We can now load Parquet files as a data source.
- It is now possible to use Azure Blob and ADLS as destinations when writing out data.
- Added a new transformation to convert Unix timestamps to DateTime.
- It is now possible to include a screenshot with your feedback.
- More transforms now emit error values when dealing with bad data.

## Feedback and Fixes

- Fixed an issue where bringing up a context menu on null cells could crash the app.
- Fixed an issue where Spark and local execution were outputting different values for infinity.
- Fixed an issue where our loading indicator would leave artifacts behind.
- Fixed an issue where column widths would reset when making changes.
- Fixed an issue where some text boxes would swallow or repeat characters.
- Fixed an issue where Excel files were not auto-detected correctly.
- Fixed an issue where 0 was getting replaced with NaN in the numeric stats inspector.

## Pendleton Private Preview Update 17 Version: 0.1.1701.23012

### New Features

- All extension points that support custom Python code can now accept Python modules in addition to lambda expressions.
- There is a new "Handle Errors" transform that allows removing or keeping error values.
- Error values now show up in the Value Count inspector.

## Feedback and Fixes

- A link to our User Voice has been added to the Send Feedback dialog.
- Fixed a bug where changing the size of a sample could result in an error.
- Convert to Boolean now treats 1 as true and 0 as false.

## Pendleton Private Preview Update 16 Version: 0.1.1701.06012

### New Features

- Error values now show up in the data quality. We are in the process of updating transformations to distinguish between error result and null result.
- Added a new Outlier Detection process that runs automatically after adding a new dataflow. We're working on ways to surface the outliers detected.
- We've added the ability to use custom Python code to load data into a new dataflow. In the Add Data dialog, choose "Custom" as the file type to use this.
- Inspectors now have a minimize button.

## Feedback and Fixes

- We've made significant performance improvements throughout the app.
- Some fixes to the Derive Column UX.
- Fixed a bug where certain cell values were being truncated in tooltips.
- Layout fixes to the Add Data dialog.

- Fixed a bug where Distinct Rows was always returning a single row when no columns were selected.
- Fixed a bug where quickly deleting multiple inspectors could crash the app.
- Fixed several minor UX and layout issues.

## Pendleton Private Preview Update 15 Version: 01.1612.15032

### New Features

- We've made significant changes to the structure of our exported Python scripts. Please try this out and let us know what you think.
- Loading a data source has been split into three separate steps. As part of this, we've also made significant improvements to our sampling performance.
- It is now possible to refresh the sample you're working on. Just right click on the sample step and choose "Refresh".
- We have a new column tooltip.

### Feedback and Fixes

- Performance improvements to grid rendering.
- Performance improvements for all transformations.
- We no longer refresh inspectors when transformations that don't affect them are added.
- Fixed a bug where random samples were causing an increase in working set.
- We now remember what was the active dataflow when opening a package.
- Fixed several bugs with inspector filters.
- Fixed a random crash that could occur when a step had an error.
- Several fixes to inspectors' rendering.

## Pendleton Private Preview Update 14 Version 01.1612.01042

### New Features

- We've added auto-complete support to filter transformations
- Auto-completion support for Derive Column by Example and Split Column by Example
- Automatic detection of file types and properties, including encoding
- We've added a pattern inspector that shows the different patterns found in a string column
- We've added context menus for cells which provide quick access to some transformations
- It is now possible to run your wrangling script locally independent of your interactive session. You can find this under File -> Run -> New Run Configuration
- We've added support for renaming columns right from the column header. Just click on the column name when the column is selected.
- Azure Data Lake Store is now supported as a data source for both local and Spark runs.
- New Keep Columns transformation.

- We've added a new Merge Columns by Example transformation that allows you to merge several columns in a new column.
- The Inspectors and Dataflows pane now start collapsed and will expand automatically.

### Feedback and Fixes

- Fixed a bug where opening a data set where a column name looks like a date could crash Pendleton
- Fixed filter transformations to work correctly with null values
- Fixed some issues with directory selection in the Azure browser
- Fixed some issues where step status would not match the data in the grid
- Separators have been added to context menus
- Fixed a bug where Trim Whitespace was not dealing with null values correctly
- Fixed a bug where our data grid was being created twice, leading to worsened performance

## Pendleton Private Preview Update 12 Version 01.1610.31012

### New Features

- New UX for Split Column by Example which allows you to specify delimiters and several other settings. .

### Feedback and Fixes

- Fixed a bug where the Time Series inspector would display nothing.
- Fixed a bug where summarize could corrupt the grid after committing or cancelling it.
- Pane labels now rotate when the panes are collapsed.
- The block actions dropdown could show up off-screen.

## Pendleton Private Preview Update 11 Version: 01.1610.25011

### New Features

- There's a new Filter column transform that requires no Python code.
- We have a new File Picker and have made some changes to our add data UI.
- It is possible to specify a glob pattern as the input in the new file picker.
- It is now possible to open Excel and JSON files from Azure. Please keep in mind that sampling is not currently supported for Excel and sampling for JSON files is line-based. So don't try to open a 500 GB single-line JSON file.
- The Read CSV and Read TSV data sources have been merged into a single Read Delimited source.

### Feedback and Fixes

- Fixed an issue where OS X builds were unable to download dependencies.
- The Time Series inspector would fail to render if its arguments were modified.
- Deleting an inspector while the grid was loading could lead to a crash.
- The Sankey diagram in the Join dialog is now correctly updated as the data changes.

## Pendleton Private Preview Update 10 Version: 01.1610.07022

### New Features

- The Write To CSV dialog now shows a file picker to choose the output path.
- We've reworked some of the UX for Derive Column by Example. We're still working on additional features and changes for this transform and we expect to bring them online soon.

### Feedback and Fixes

- Hovering over a recent package in the start screen now shows its full path in the tooltip.
- We've made numerous fixes to our installation process for OS X.
- The numeric stats inspector was failing when the column contained NaNs.
- Dialogs now have cancel buttons.
- Join could cause a crash when one of the join sources failed to execute.

## Pendleton Private Preview Update 9 Version: 01.1610.02012

### New Features

- Added some animations to make it more obvious when a new step gets added to a data flow.
- We've added the ability to show non-printable characters in the grid.
- We've added support for specifying a delimiter and a fill strategy for Split Column by Example.
- We've added a Replace Values transform.

### Feedback and Fixes

- Update notifications will now bring up the notifications pop-up instead of just lighting up the notification bell.
- Several fixes and improvements to Azure Blob browsing.
- Different icons are now used for each table in the join wizard.
- The Activate button in the EULA page is now greyed out until the EUAL has been agreed to.



- It is no longer possible to resize some of the panes until they break layout.
- Opening an Azure Blob no longer results in random single-letter data flow names.
- Several fixes for the OS X version of the app. If you've previously been unable to run the app on OS X, please try again.

## Pendleton Private Preview Update 8 Version: 01.1609.25012

### New Features

- It is now possible to browse for and open Azure Blobs from within Pendleton. We're actively working on making this experience better and are **looking forward to suggestions and feedback**. Please give it a try!
- We now show more descriptive names for some of the steps in the history pane.
- Mapping values from one dictionary to another in the 'Add Column by Example' transform. For example, now we can map '0' and '1' to 'Male' and 'Female' respectively. You need to give as many examples as there are distinct values in the source domain.
- Add Column by Example now supports examples that perform division by 1000 e.g. 2788 -> 2.7B

### Feedback and Fixes

- Convert to date now allows you to specify a static date to use when converting time values.
- Changed a few labels in the UI to be more consistent.
- The Sankey diagram in the join wizard now displays a warning when there's not enough space.
- We've fixed an issue that could cause the app to become completely unresponsive after switching dataflows.
- Character inputs in some dialogs were getting truncated.
- Fixed several installation issues when installing to a path with a space.
- Fixed some cases where Split by Example was splitting too much.
- Add Column by Example will no longer fill columns with a constant value when it can't learn a better program.

## Pendleton Private Preview Update 7 Version: 01.1609.19012

### New Features

- Inspectors now show a loading indicator when they're doing work instead of just "Data Not Available". "Data Not Available" will still come up if the inspector is unable to calculate data.
- Inspectors now execute independently of the core script, leading to speed improvements when applying transformations.

- We've added a new interactive 'Missing Values' transform that allows for replacing missing values with a specific value or just removing them.
- It is now possible to sample JSON files.
- We now support opening a directory of files in addition to a single file. The current implementation requires that all files have the same schema. We're actively working on this feature and could use any feedback or suggestions you might have.
- We now remember the size and collapsed state of the dataflows pane and the inspectors well.

### Feedback and Fixes

- Fixed some type inference issues when loading Excel files.
- Changes to some labels and text in the UI to make it more consistent.
- Fixed an issue where menus could become disabled and the UI would be unresponsive to all actions.
- Fixed an issue where switching between data flows could lead to a bad UI state.

## Pendleton Private Preview Update 6 Version: 01.1609.12032

### New Features

- We've made some changes to our UI. You will now find data flows on a pane on the left side of the screen instead of in tabs at the top.
- It is now possible to edit auto-split transforms to specify whether to keep delimiters as columns or not.

### Feedback and Fixes

- More fixes and reliability improvements to PROSE fill.
- Fixed a memory leak with the map inspector that could cause the app to crash.
- Fixed handling for some dates in Excel files.

## Pendleton Private Preview Update 5 Version: 01.1609.01022

### New Features

- Support for reading Excel files has been added.
- There is a new option to export your wrangling script as a Jupyter notebook. You'll find it under the File menu.
- Column name restrictions have been significantly reduced. Columns can now have any name as long as it's unique within the data flow.

- It is now possible to show whitespace within cell values. This can be enabled from the View menu.
- Support for TSV files has been added.

### Feedback and Fixes

- Fixed an issue where custom transforms could cause the app to crash.
- The map inspector would display "Data Not Available" when the data provided fell outside acceptable longitude and latitude. It now displays a more helpful error.
- Several fixes to Add Column by Example and Auto-split.
- Submit buttons were displaying the text "submitLabel" as their content.
- Closing a dialog by hitting Esc would leave all menus disabled.
- The dialog for rename column now pre-populates the field for the new name with the current name.
- It was possible to delete blocks while in Summarize, leading to unexpected behavior.
- The app showed up as Electron.exe in task manager instead of Pendleton.exe

## Pendleton Private Preview Update 4 Version: 01.1608.29032

### New Features

- It is now possible to edit the name of an "Add Column by Example" column.
- A "Convert to boolean" transform has been added.
- Random sampling for CSV files has been added.

### Feedback and Fixes

- Python processes were being restarted unnecessarily while the app was running.
- Context menus now respect multiple selection.
- You can now double click on the file type selectors in the add data source menu.
- The separator property for CSV files accepted as string as input and then failed. It should be a character.
- Editing dialogs for transformations that accepted multiple columns as arguments (rename, for example) could result in transformation failures.
- We've sped up our boot time quite a bit.
- Fixed an issue where custom Python code could cause the app to crash.

## Pendleton Private Preview Update 3 Version: 01.1608.18012

### New Features

- Responding to one of our most requested features, we have added context menus to data flow tabs and grid columns. Right clicking on a column will pop up a context menu that allows you to apply transformations or add inspectors. Right clicking on a data flow tab will bring up options to delete or rename it.

- There's a new "Trim Whitespace" transform that removes leading and/or trailing whitespace from the values in a column.
- We now show an indicator at the end of an activity when it is used as the source of a join. You can use this to navigate to the resulting activity.
- We've added a new inspector to the app: a map inspector. This inspector lets you plot a series of coordinates on a map and is great for visualizing outliers in location data.

## Feedback and Fixes

- We've added improved error messages in different places: the Write To CSV transform, join errors, and JSON errors, among others.
- Loading JSON files sometimes resulted in a crash.
- Some cleanup and bug fixes in dialog layouts.
- The steps in the history sometimes moved when bringing up their dropdown menu.
- In summarize, the count of a datetime column was returning an unexpected date instead of an actual count.
- When multiple inspectors were added at once (by selecting multiple columns), they were getting added in reverse order.
- Cancelling out of summarize would leave the wrong step selected.
- Tooltips in some inspectors were getting stuck when going in and out of summarize mode.
- The numeric stats inspector was not scrollable and was getting truncated.
- The count of a datetime column in summarize was returning a datetime value instead of an actual count.
- Dropdown menus for steps were causing steps to move and shift.
- Fixed a few issues in join that could cause either the preview or the resulting grid to not be rendered.
- The numeric stats inspector was shown as an option for non-numeric columns. It should be disabled.
- The file type icons in the add data source dialog can now be double clicked.