August 12, 2018

Errata

All RapidMiner tutorial and solution processes, exercise solutions, screenshot files, presentation slides, and test questions have been updated to correspond to RapidMiner Studio 8.2. These materials are currently being updated to reflect changes resulting from the recent release of RapidMiner Studio 9.0. A main difference between RapidMiner Studio 8.2 and RapidMiner Studio 9.0 is the addition of the *Turbo Prep view*. The Turbo Prep view allows you to easily *Transform, Cleans, Generate, Pivot,* and *Merge* data. The Turbo Prep view is freely available for 30 days with the trial version of RapidMiner Studio 9.0 and is included as part of the educational version of the software. A brief introduction to the *Turbo Prep view* is given at the end of this document.

Visit http://krypton.mnsu.edu/~sa7379bt/ for download links to the latest versions of RapidMiner Studio and Weka Explorer. The site also offers links to the very latest RapidMiner Studio screenshot and Errata files.

Instructors, please email me at richard.roiger@mnsu.edu if you have any questions or concerns about the text or supplementary materials.

Chapter 5

Sections 5.2.1 through 5.2.4 assume that the *maximal depth* parameter for the *Decision Tree* operator is set at 5. This is clearly stated in the Chapter 5 screenshot file (*Screens_05).*

The *Cross Validation* operator replaces the *X-Validation* and *X-Prediction* operators. The *Cross Validation* operator input port label *exa* replaces *tra* and output port label *per* replaces *ave.* Any references to *X-Validation* or *X-Prediction* within the text should be replaced with the words *Cross Validation.*

The *Write Model* and *Read Model* operators have been replaced respectively with the *Store* and *Retrieve* operators. The *Store* operator differs from *Write Model* in that the created output file must be written to a location within the data repository. In a like manner, the *Retrieve* operator always imports from a location within the data repository.

Page 146 Section 5.1.2 4th line

Replace

eight templates

With

several templates

Page 167 bottom of page

Replace
All three bulleted items

With

- Create the process shown in Figure 5.30.
- Before you execute your process, click on *Store* and use the repository entry parameter to name your model and specify where it is to be stored.
- After your process executes, make sure the decision tree model has been written to the specified repository.

Page 182 right above Figure 5.49

Replace
min number of itemsets

With
Min items per itemset

Chapter 10

Page 306
Replace the paragraph beginning with --Lastly, this example used the X-*Validation* operator ….

With this paragraph:

Thus far our examples have limited the use of *Cross Validation* to situations where the output attribute is categorical. However, *Cross Validation* is also useful when the output attribute is numeric. To accommodate numeric output, we replace the *Performance (Classification)* operator with *Performance (Regression).* This operator offers several options for performance evaluation including *root mean squared error* and *absolute* error.

Page 314

Replace exercise 3 with this exercise.

3. Add the *Store* operator to the process in Figure 10.2. Run your process. Create a new process that retrieves the saved model and uses *Apply Model* and *Performance (Classification)* to apply your model to the XOR data.

Page 318

Replace
P(H|E) is the *conditional probability* that *H* is true given evidence *E*.

With
P(E|H) is the *conditional probability* that *E* is true given *H* is known to be true.

Chapter 13

Section 13.1

The exchange traded fund XIV is used to illustrate time series analysis. This fund was created to give traders an opportunity to bet against market volatility. February 2018 showed an unusual amount of market volatility so much so that XIV dropped from a February 2, 2018 price of $115 / share to a February 7, 2018 price of just over $6.00 / share. Shortly thereafter, Credit Suisse closed and eliminated the fund.

Section 13.1 is affected only to the extent that the student is unable to recreate the datasets used for illustrative purposes. This is not an issue as the XIV datasets are part of the student dataset library. Also, the Yahoo stock Web site has changed some data storage feature which in turn has currently disabled the proper functioning of the *Yahoo Historical Stock Data operator* (QuantX-1 Extension). The good news is that you can download historical stock data directly from:
https://finance.yahoo.com/
Simply type in the stock symbol of interest, locate "Historical Data", change the date range as desired, click *apply* then click on *download.* The historical data will be dumped to an Excel spreadsheet. The Excel spreadsheet can be loaded into RapidMiner Studio and used to replace the Yahoo Historical Stock Data operator shown in Figure 13.1.

Page 417
2nd line

Replace
(0.731 vs. 0.816)
With
(0.645 vs. 0.807)

Appendix A

Page 453

Replace

LINK PENDING

With

https://www.crcpress.com/Data-Mining-A-Tutorial-Based-Primer-Second-Edition/Roiger/p/book/9781498763974

Community Data Sets – Watson Analytics

The links for the community data sets and trial version are given below:

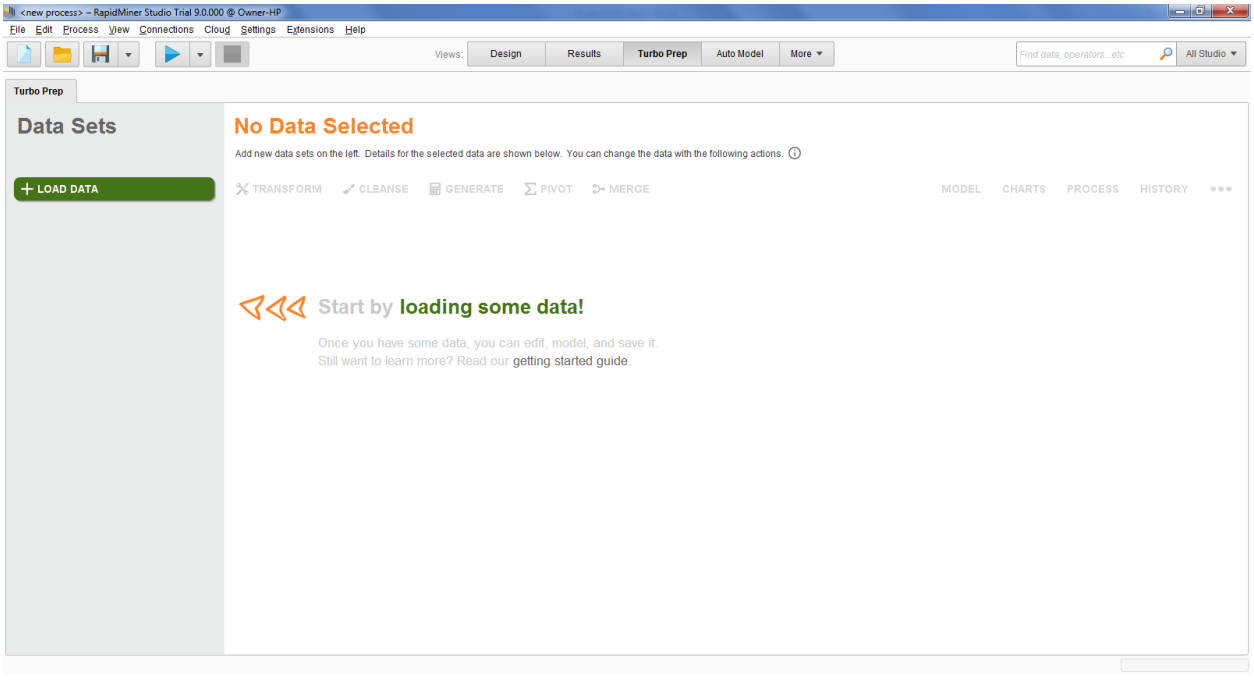
https://www.ibm.com/communities/analytics/watson-analytics-blog/guide-to-sample-datasets/

https://www.ibm.com/analytics/watson-analytics/us-en/

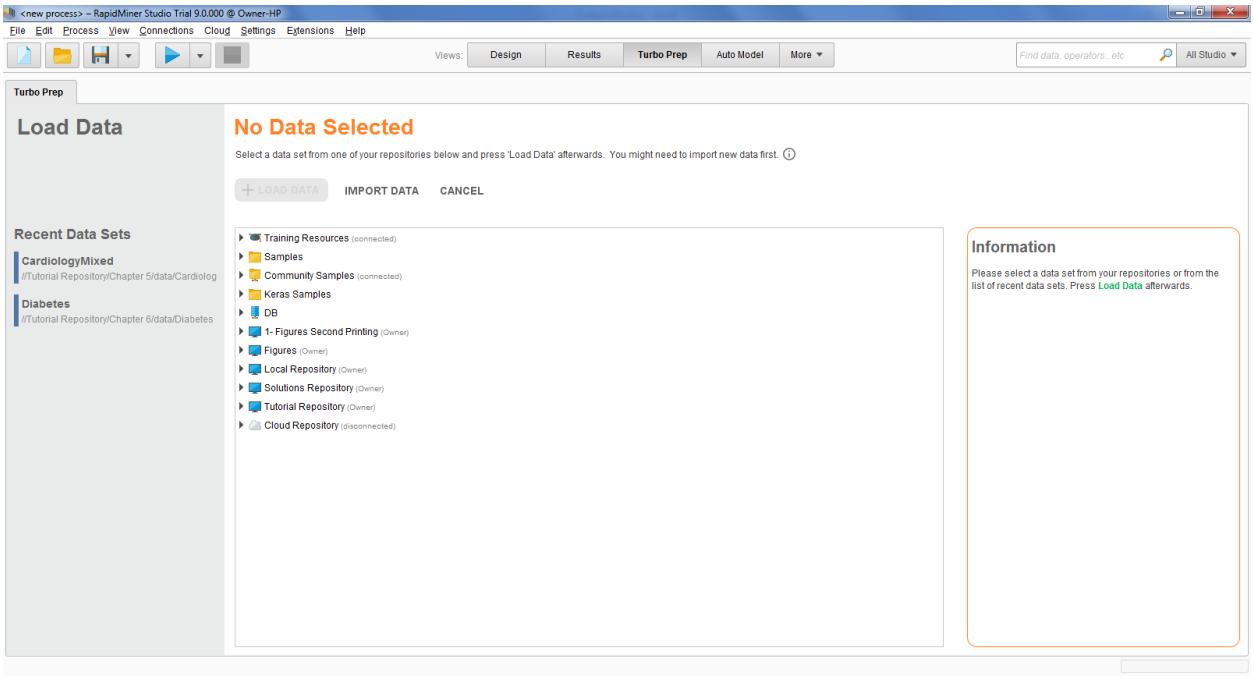
RapidMiner's Turbo Prep View

Here is a simple example of one way to use the Turbo Prep View.

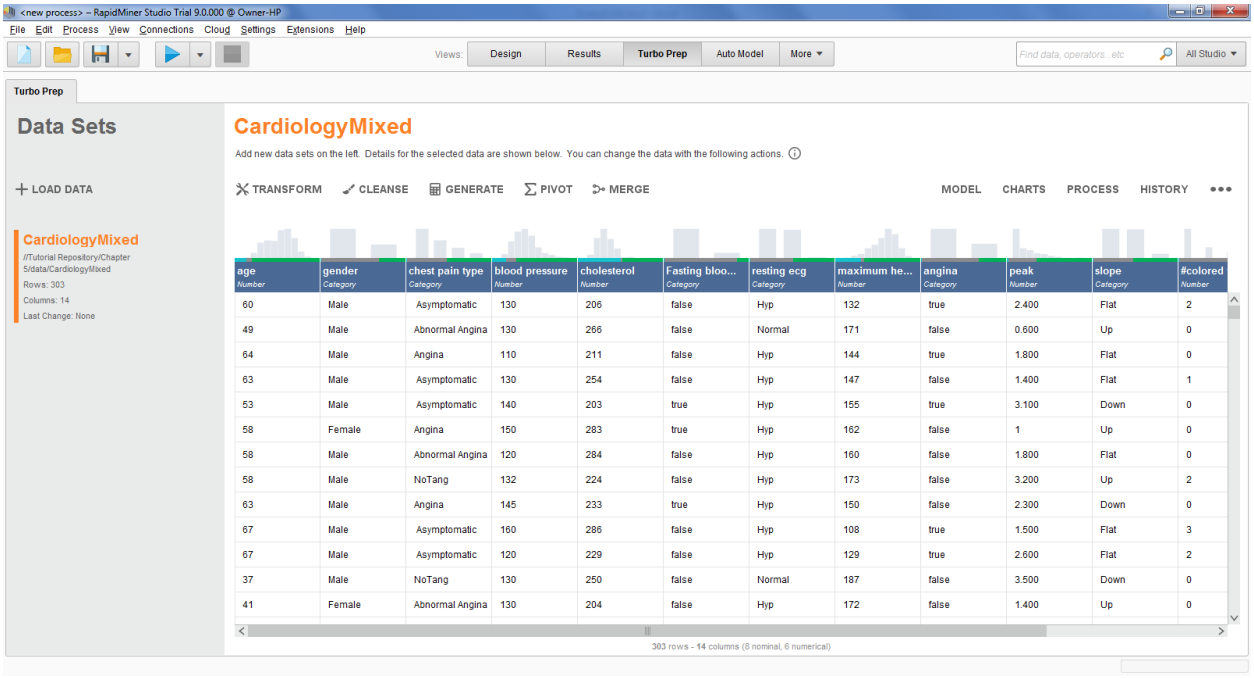
From the *Views* menu click on *Turbo Prep*. Your screen will appear as below:



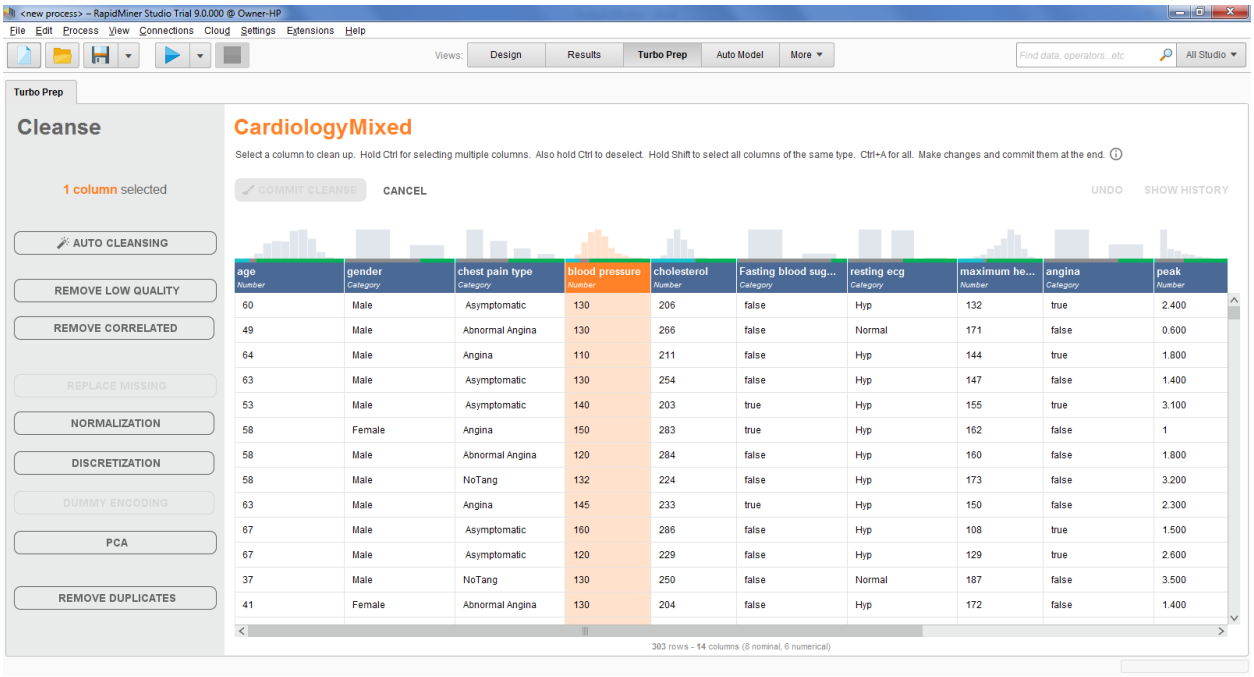
Click on +Load Data to see a screen similar to the following:



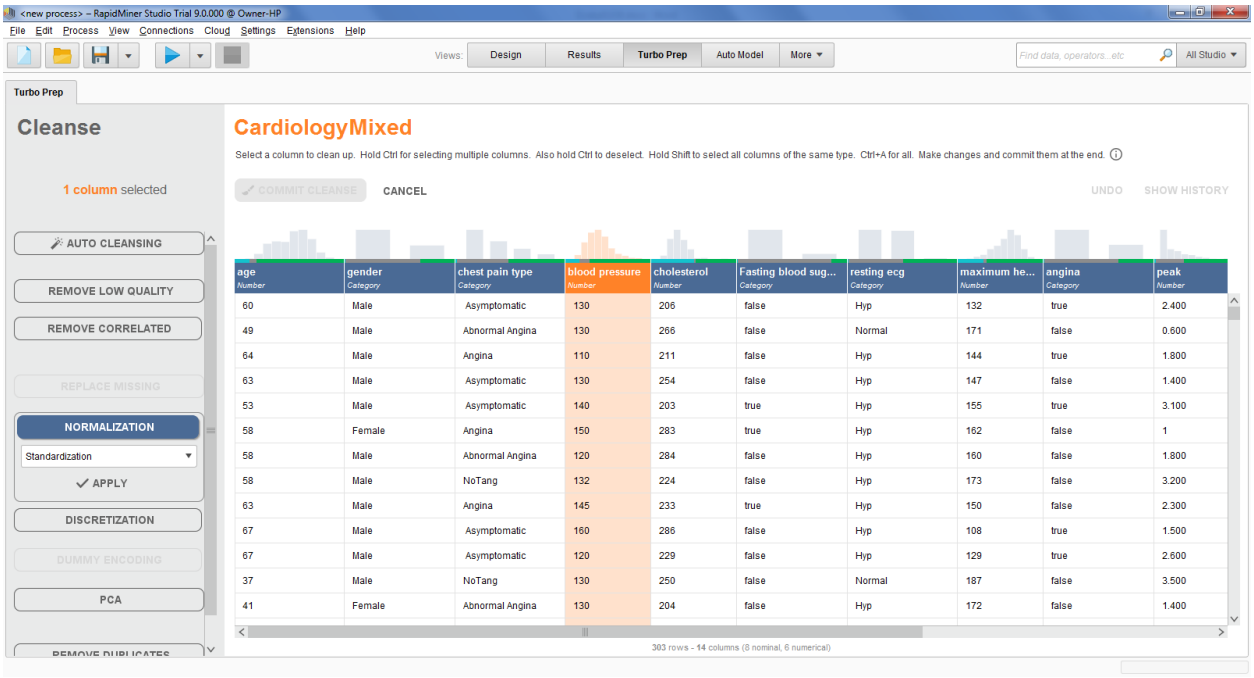
Use your mouse to load the CardiologyMixed data set. This data set is found in the *Chapter 5 Data* subfolder within the *Tutorial Respository*. Your screen will appear as below:



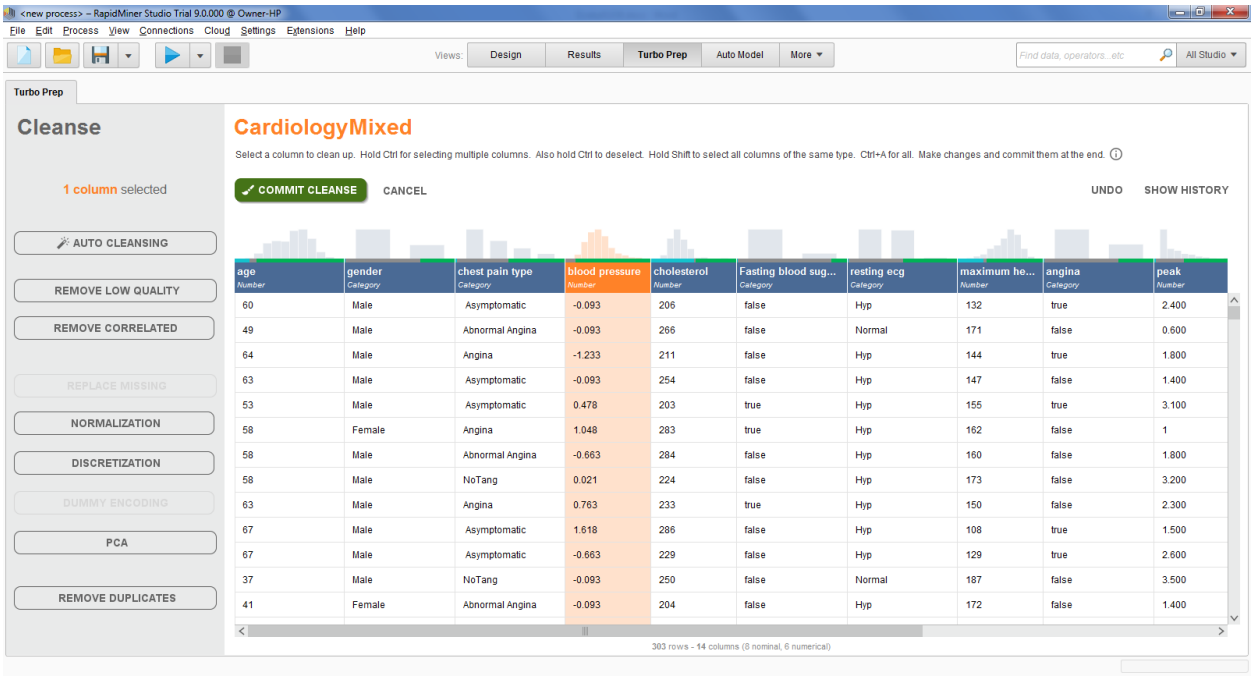
Highlight the “blood pressure” column with a click of your mouse. Next, click “Cleanse”. Your screen will appear as below:



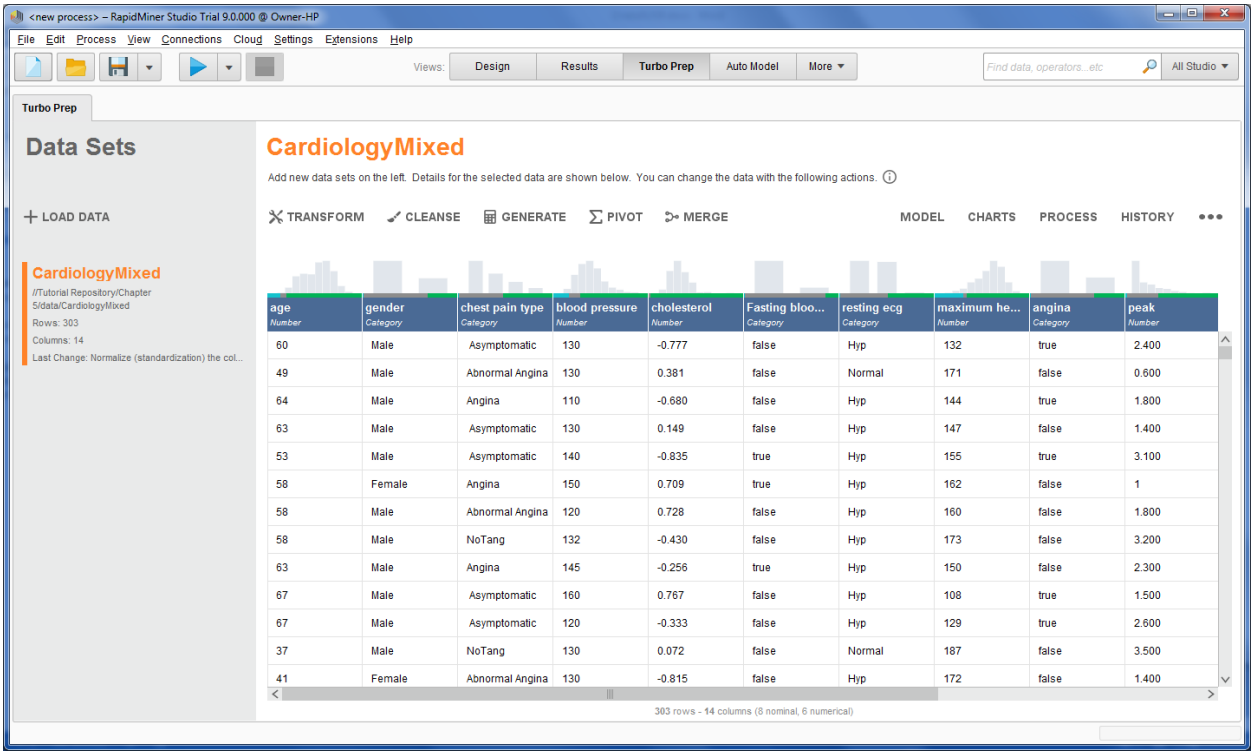
Click *Normalization* and set the normalization type to *standardization* to see the following:



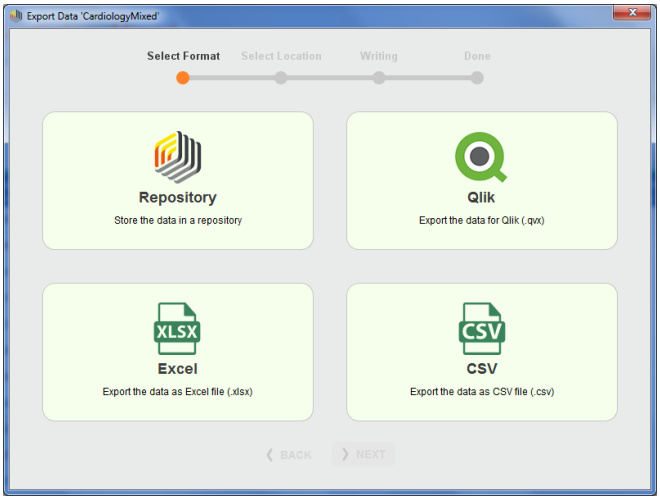
Click Apply to see the screen below:



To create a new data set with the specified change click on *commit cleanse* to see the screen below:



Click on '**. . .**' seen in the upper right of your screen. Choose *Export* to see the following:



Click on the action we wish to take. A click on *Repository* then on *next* allows you to store the modified data set in the RapidMiner repository of your choice. As you will learn, data manipulation is a breeze with Turbo Prep!