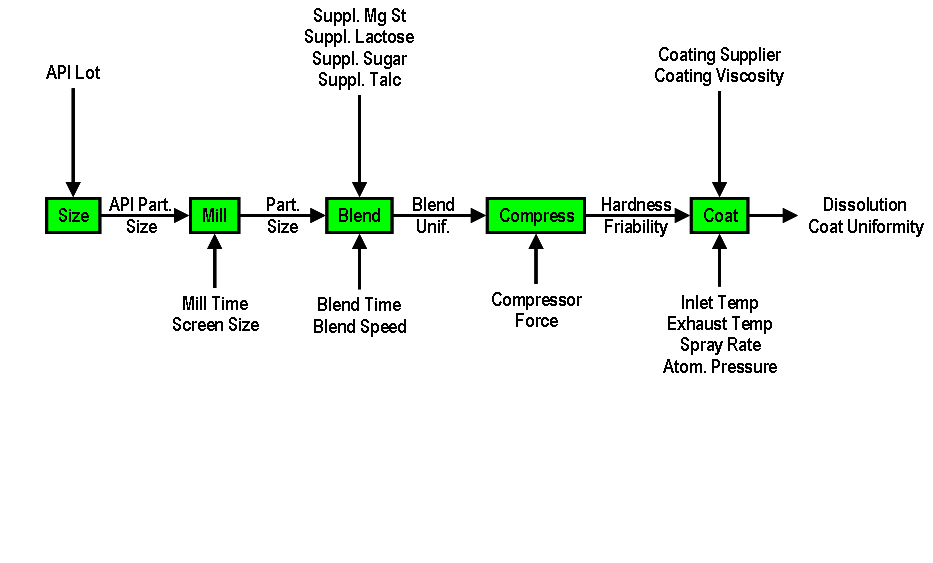
# Pharmaceutical Tablet production process



(specification limit Dissolution >70) :

## Discover the Tablet data

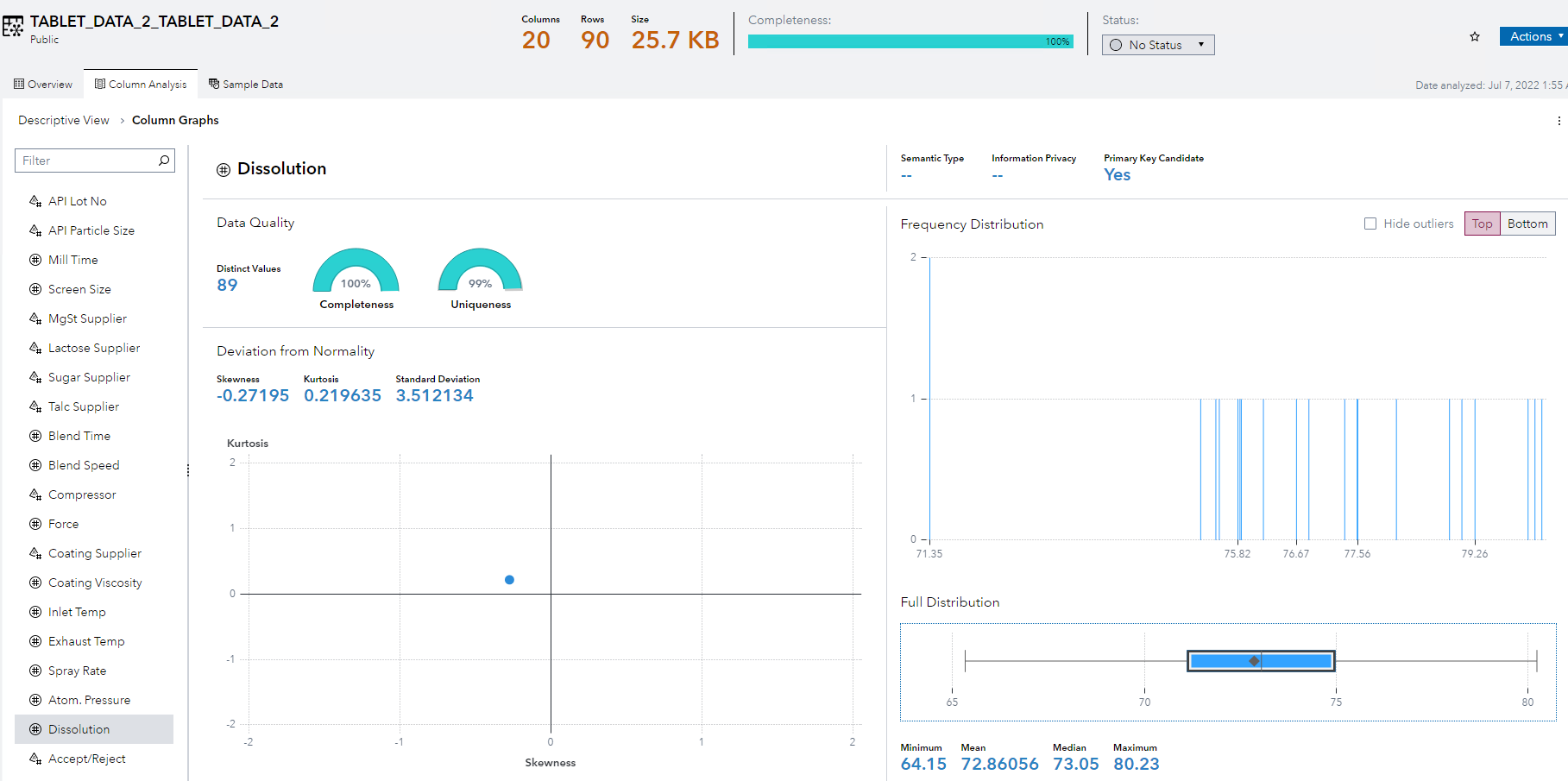
Download the Case Study from

<https://github.com/paulvanmol/PolymerCase.git>

and download/copy it to your d:\workshop\polymercase folder.

1. Open Manage Data: <http://server.demo.sas.com/SASDataExplorer/>
2. Select the import tab: local files: d:\workshop\Tablet\_data\_2.xlsx
3. Import the Tablet\_Data\_2 in the public cas library
4. In Discover Data, select the Discovery Agents
5. Select public and Action Run Now to discover all Public data
6. Select Overview and search for the tablet data
7. Select columns and look at Distribution, histograms and boxplots.
8. What does the Accept/reject tell you: how many tablet Lot No are rejected
9. Check Dissolution: boxplot what is the minimum value

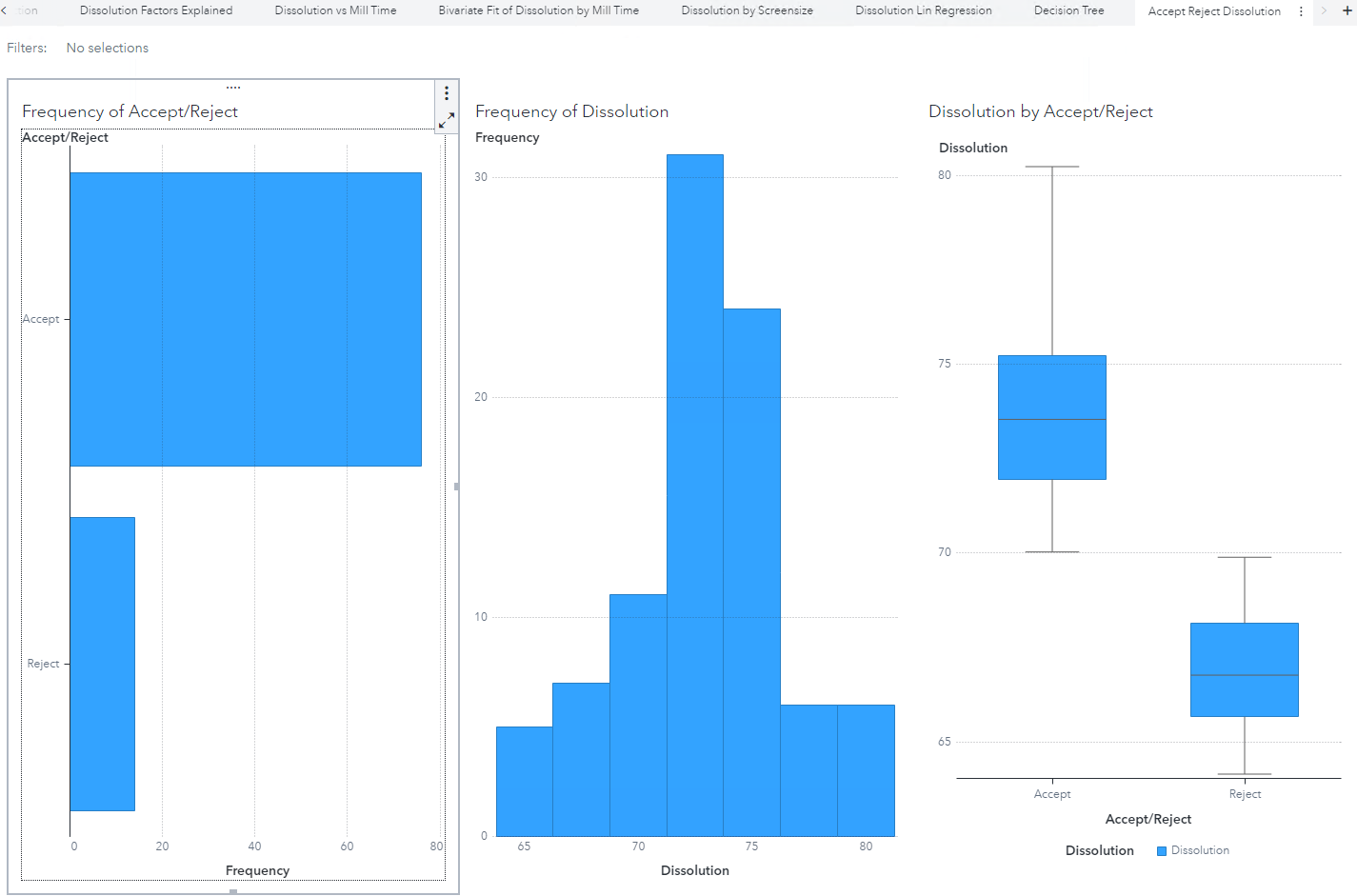
Answer:



## Explore the Tablet Data:

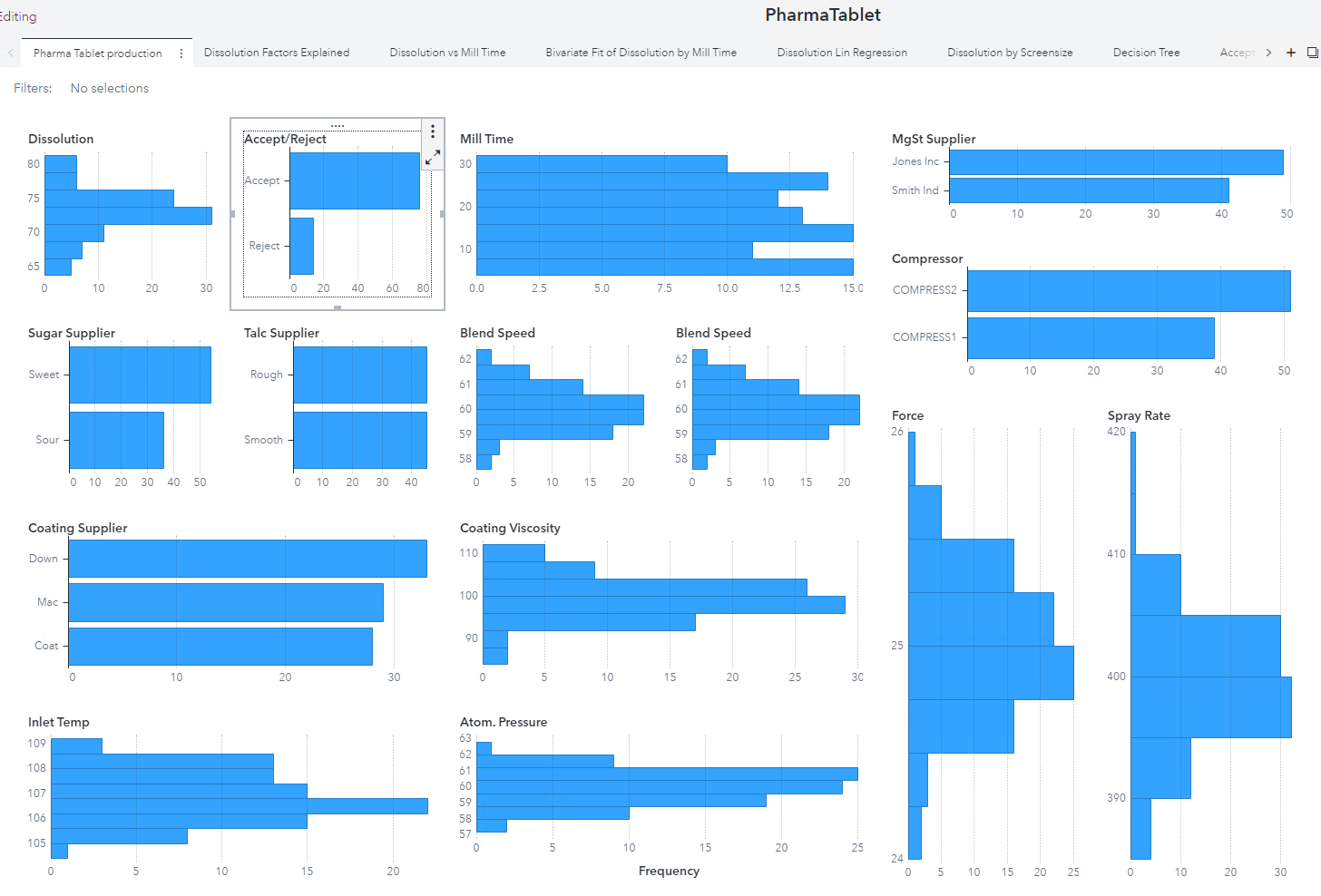
1. Create Histograms & Barcharts in SAS Visual Analytics
   1. Histograms for numeric fields
   2. Bar Charts for categorical fields
   3. X Axis label remove
   4. Graph no automatic title
   5. Save Template histogram to objects as Histogram\_notitle
   6. Save Template barchart to objects as barchart\_notitle
   7. Add multiple histograms and barcharts for the different data items to one page
   8. Add Automatic Interactions between all objects
2. Look at the Data Item Accept/Reject and the value of Dissolutions

Answer



1. Explore interactions between low Dissolution (below 70) and other data items

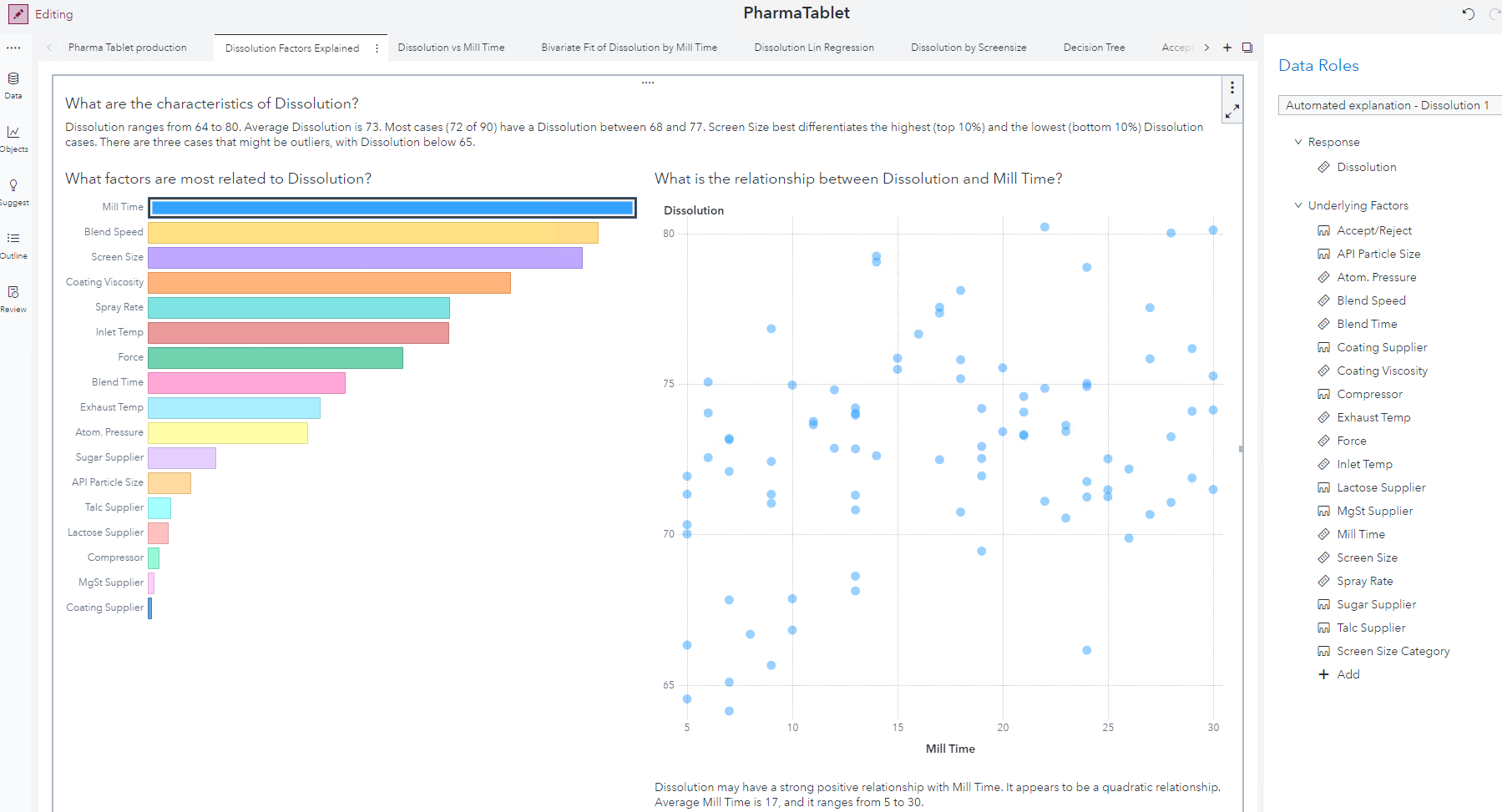
Answer:



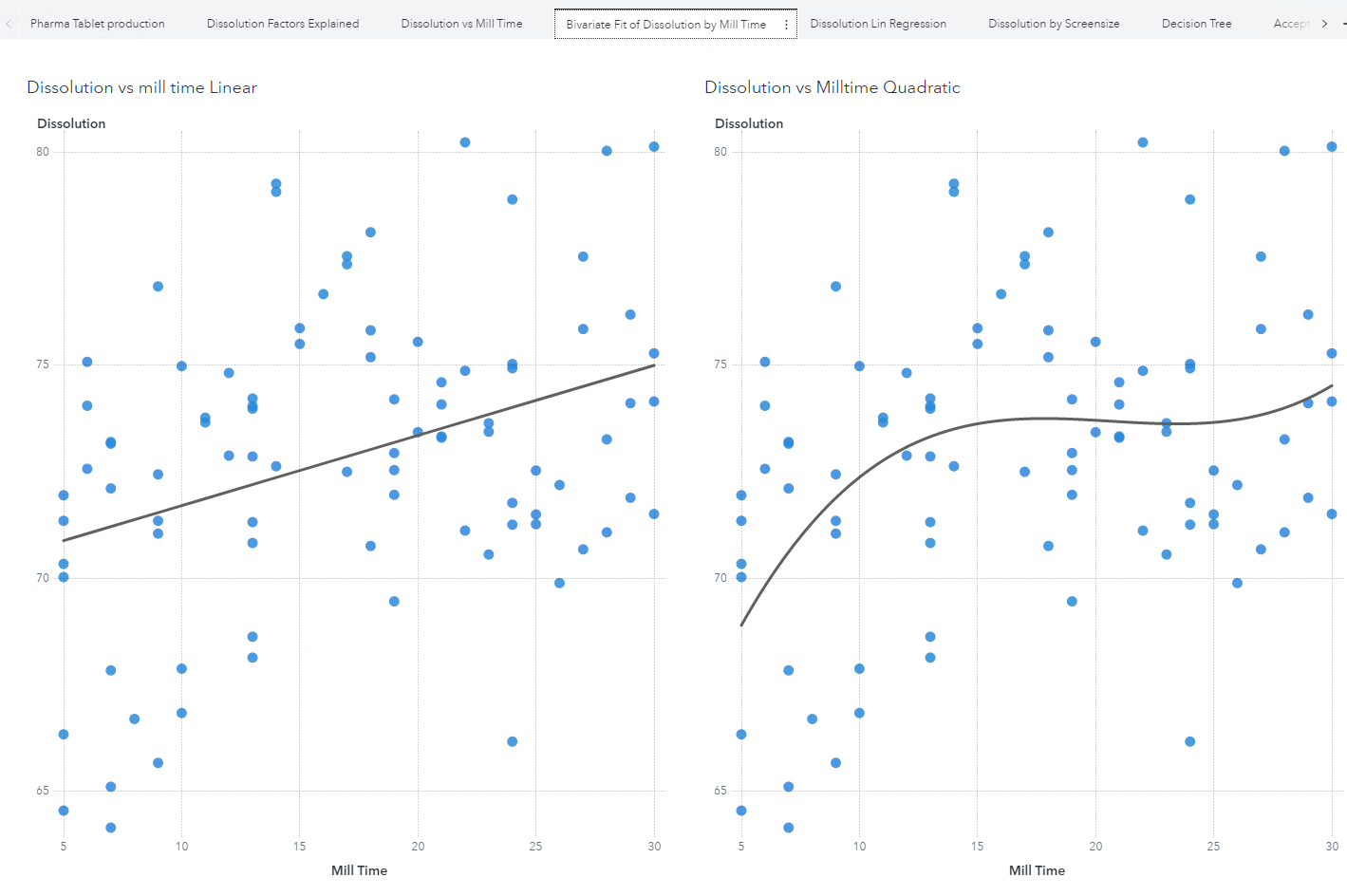
## Explain Dissolution: What are the most important factors?

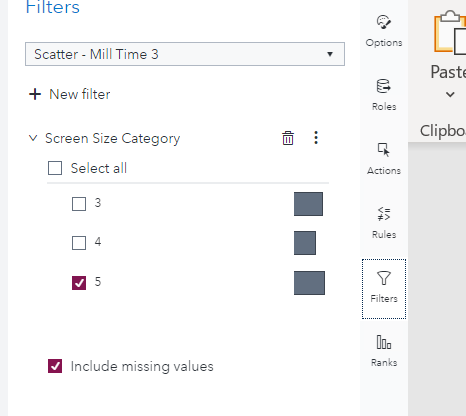
1. Automatic Explanation of Dissolution: what are the most important factors that have influence on Dissolution?
   1. Answer: Milltime, Screen size, …?

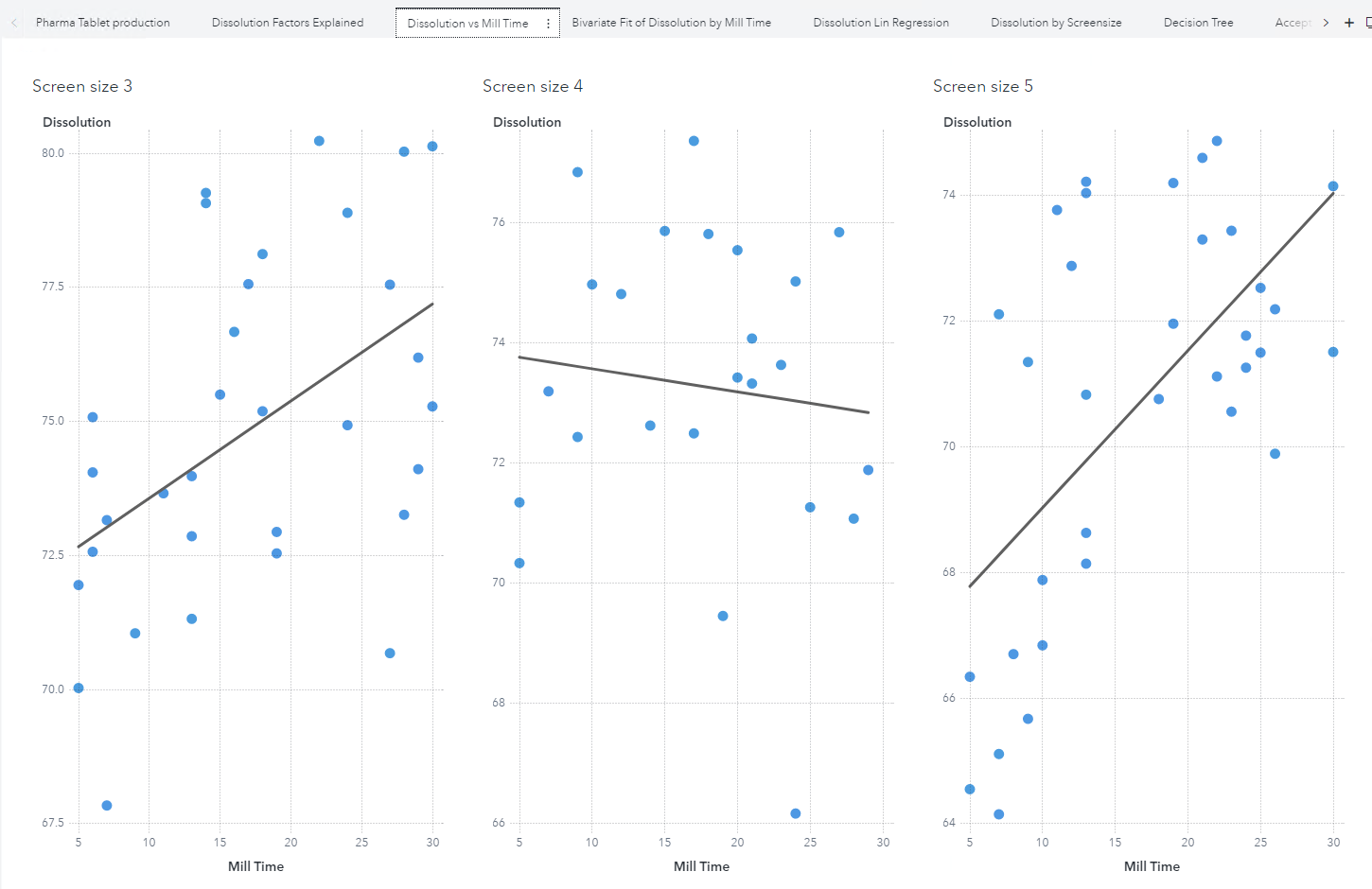
Answer:



1. Scatter plot of Dissolution and Milltime
   1. Add a fitline (Quadratic or Cubic or Best fit)



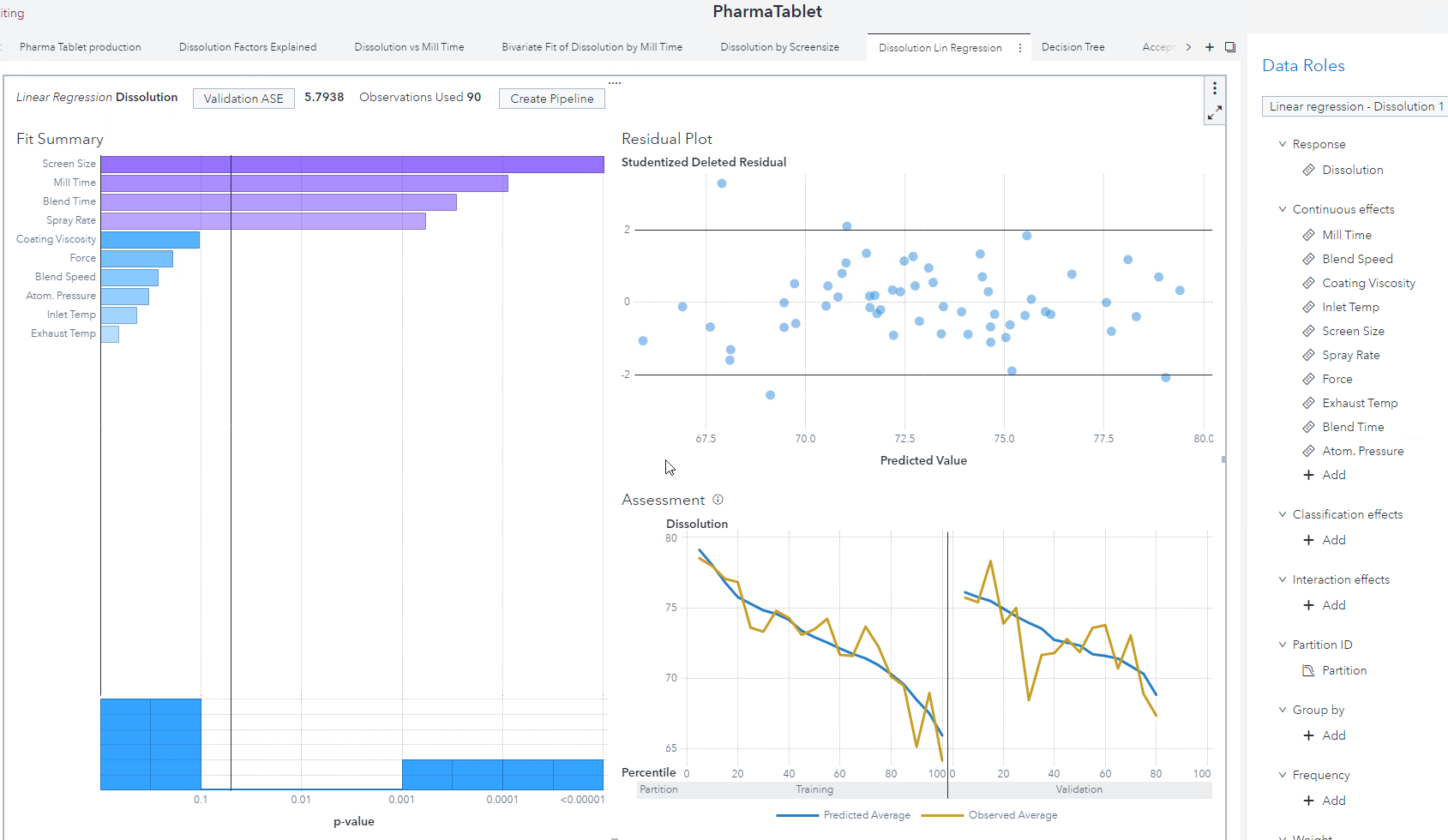
1. Add a page with scatterplots of Mill Time and Dissolution and use the filters to check the differences by Screen size:
2. 



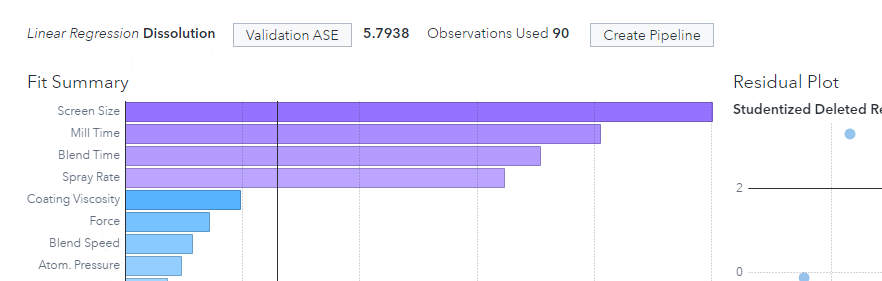
1. Boxplot of Dissolution by Screen Size (ideally an ANOVA in SAS Studio)
   1. Make duplicate of Screen Size as Category data item
   2. Add Boxplot of Dissolution by Screensize



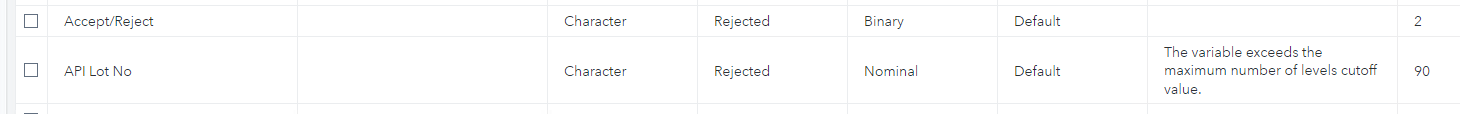
1. Linear Regression of Dissolution and all other Data items
   1. Create a partition ID variable (80 % training and 20 % Validation, stratified by Accept/reject)
   2. Target: Dissolution
   3. All other variables: except API Lot No and Accept Reject



1. Decision Tree of Dissolution ?
   1. Compare models
2. Create Pipeline Model from Linear Regression
   1. Create Pipeline>Add to New Project



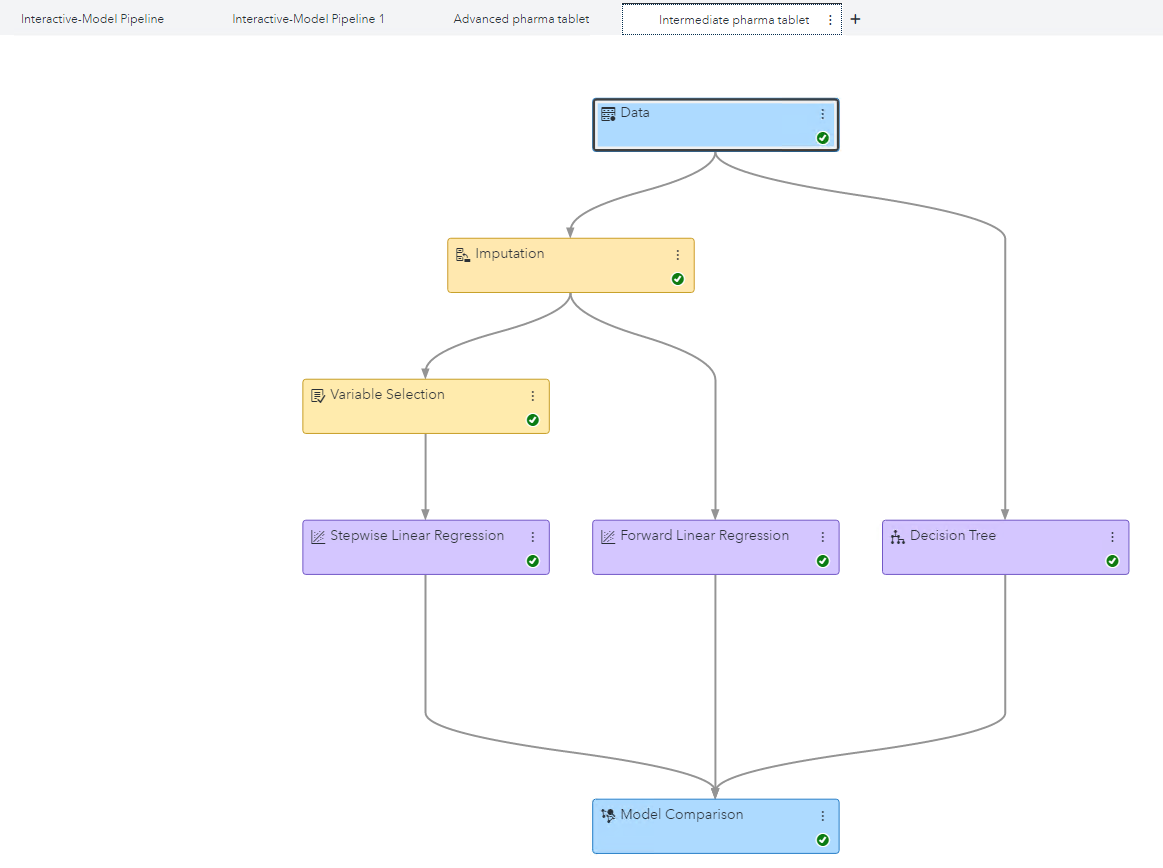
* 1. Reject the Accept/Reject and API Lot No as it identifies the batches (change from input to ID or rejected)



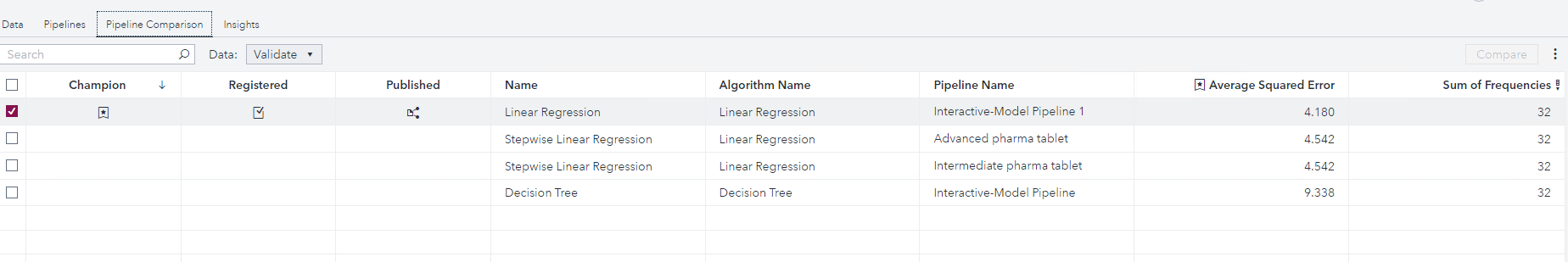
1. Experiment with Intermediate Pipeline template for Interval variable

Dissolution is the Target variable which is interval not Category

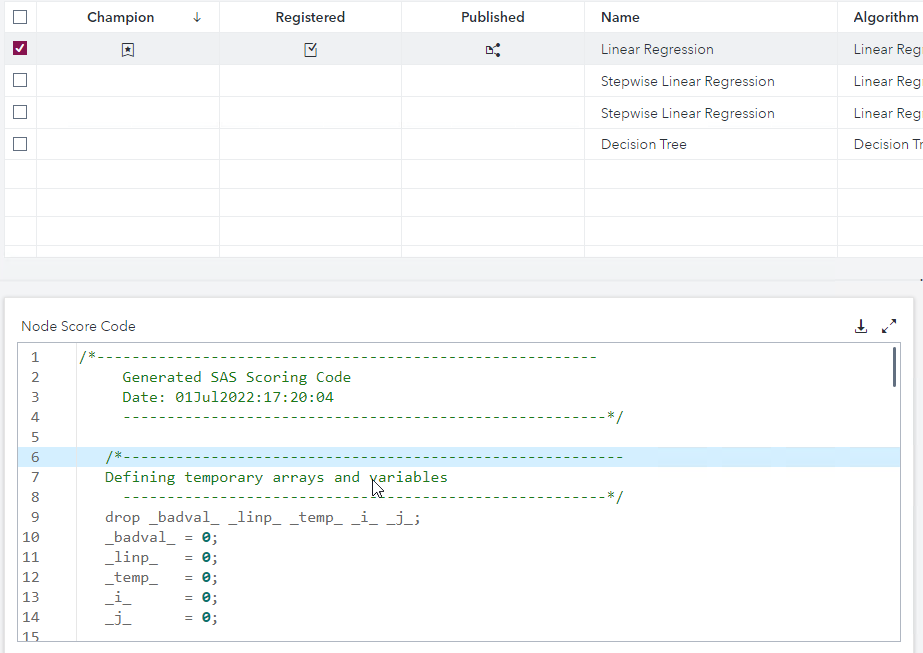
The Intermediate pipeline template applies some best practices like missing value imputation, variable selection and then creates several other models:



1. Compare models: the best model is still the Linear Regression



1. Download the Score code of the Linear Regression Model:



1. Optimize the Dissolution to avoid production with less dan 70?
   1. What are the optimal production parameters?
   2. Use Automated Prediction?
   3. Use Score code? How to optimize parameters with OR?