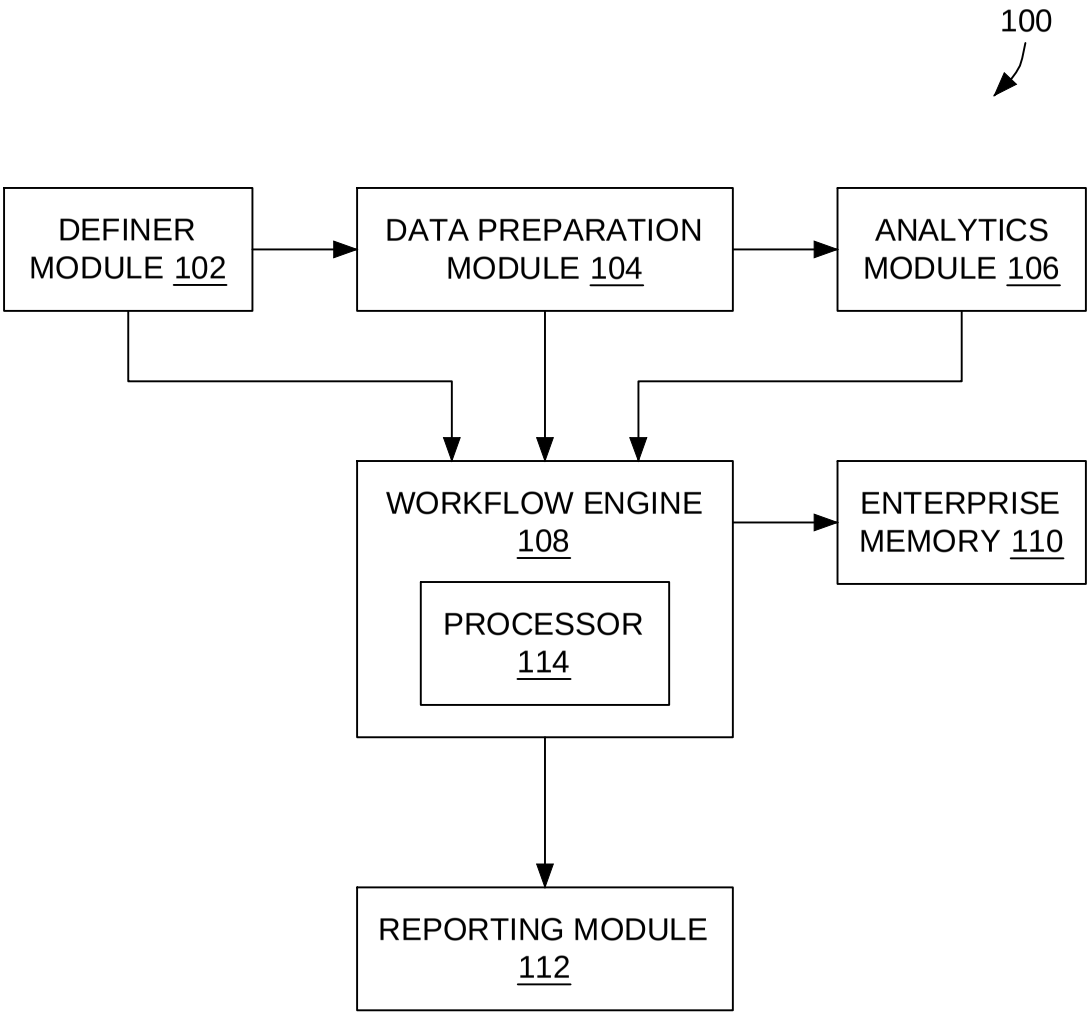

100
DEFINER MODULE 102
DATA PREPARATION MODULE 104
ANALYTICS MODULE 106
WORKFLOW ENGINE 108
PROCESSOR 114
ENTERPRISE MEMORY 110
REPORTING MODULE 112

FIG. 1

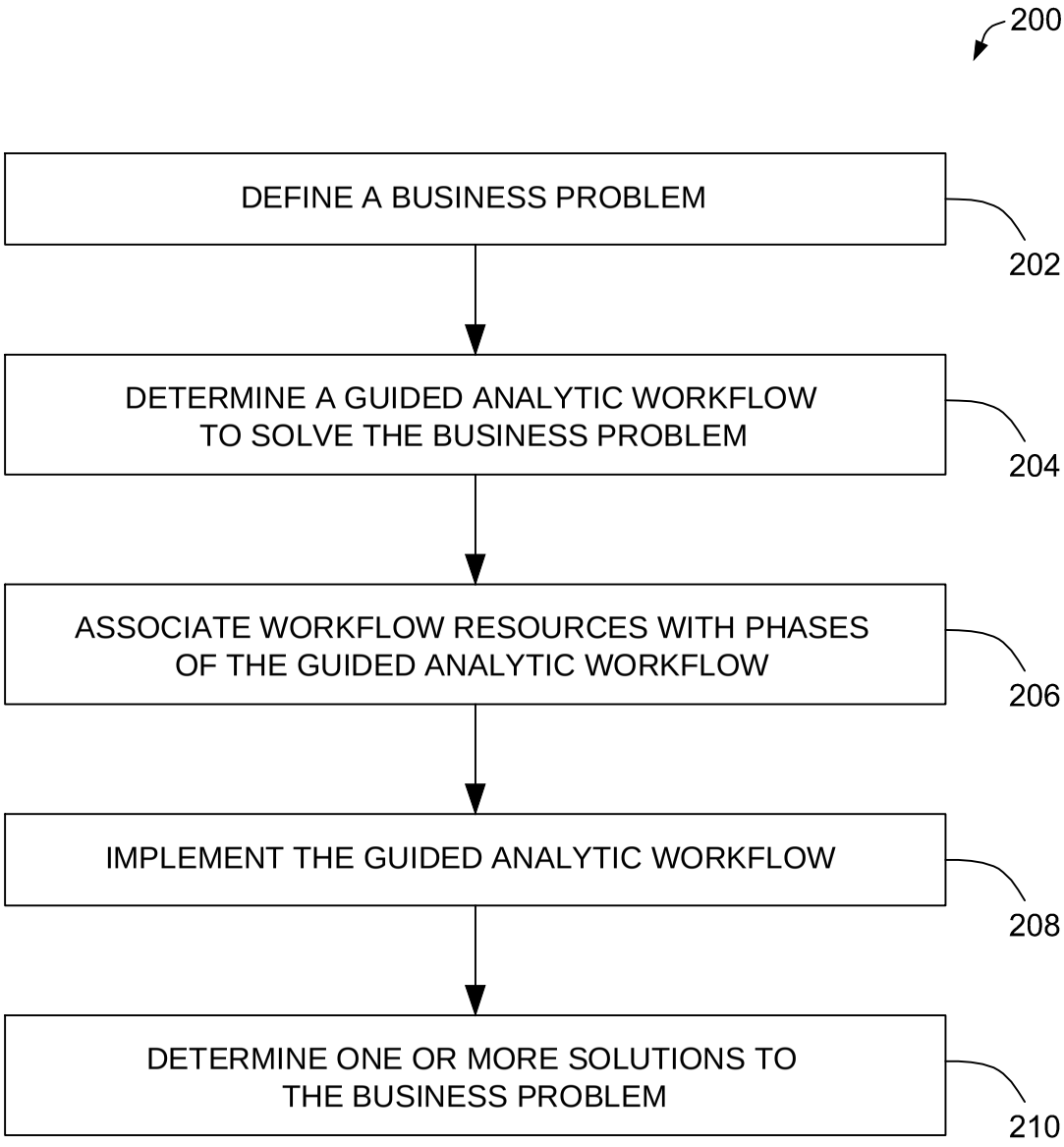

200
DEFINE A BUSINESS PROBLEM
202
DETERMINE A GUIDED ANALYTIC WORKFLOW TO SOLVE THE BUSINESS PROBLEM
204
ASSOCIATE WORKFLOW RESOURCES WITH PHASES OF THE GUIDED ANALYTIC WORKFLOW
206
IMPLEMENT THE GUIDED ANALYTIC WORKFLOW
208
DETERMINE ONE OR MORE SOLUTIONS TO THE BUSINESS PROBLEM
210

FIG. 2

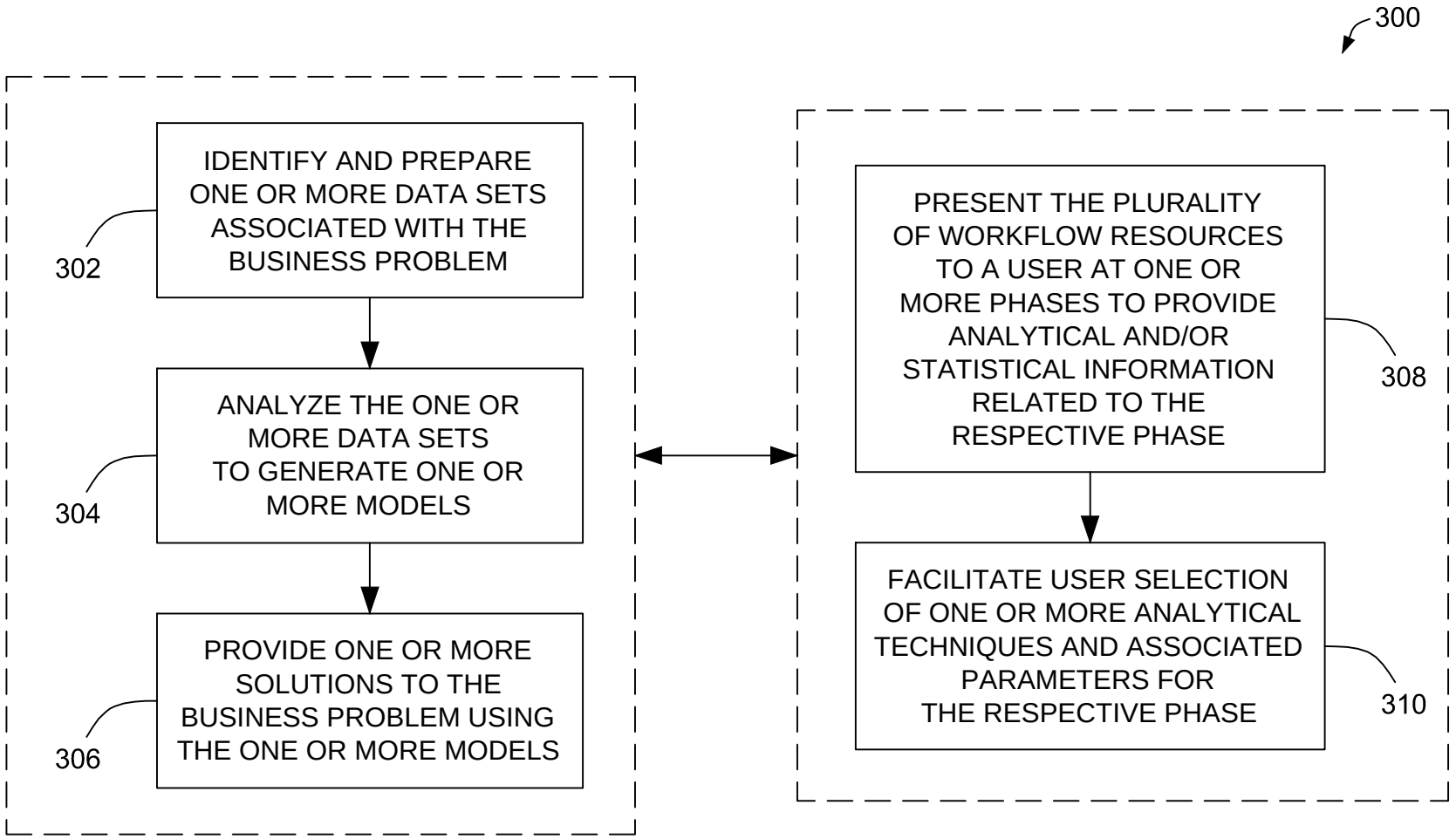

300
IDENTIFY AND PREPARE ONE OR MORE DATA SETS ASSOCIATED WITH THE BUSINESS PROBLEM
302
ANALYZE THE ONE OR MORE DATA SETS TO GENERATE ONE OR MORE MODELS
304
PROVIDE ONE OR MORE SOLUTIONS TO THE BUSINESS PROBLEM USING THE ONE OR MORE MODELS
306
PRESENT THE PLURALITY OF WORKFLOW RESOURCES TO A USER AT ONE OR MORE PHASES TO PROVIDE ANALYTICAL AND/OR STATISTICAL INFORMATION RELATED TO THE RESPECTIVE PHASE
308
FACILITATE USER SELECTION OF ONE OR MORE ANALYTICAL TECHNIQUES AND ASSOCIATED PARAMETERS FOR THE RESPECTIVE PHASE
310

FIG. 3

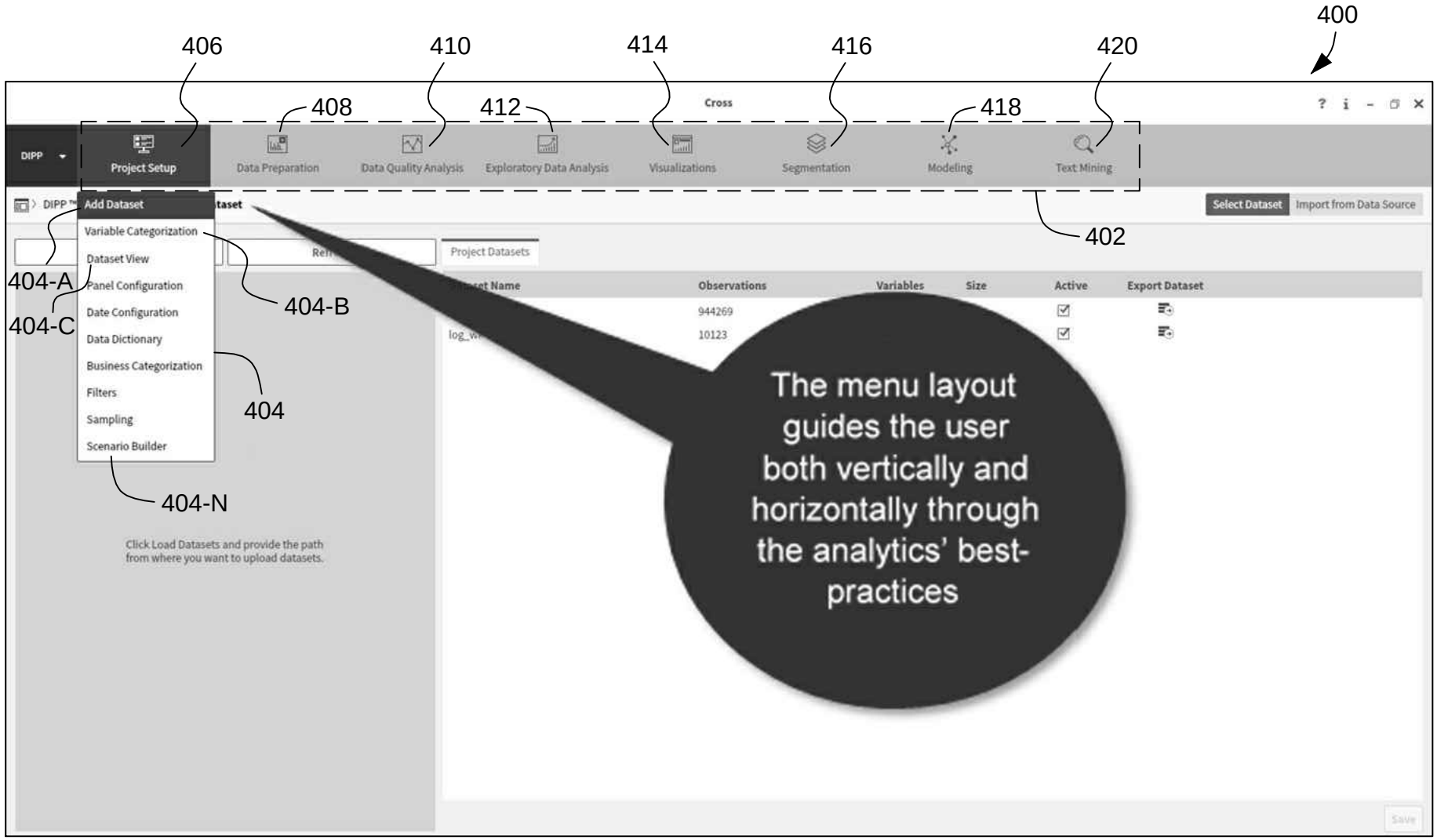

400
406
410
414
416
420
408
412
418
Cross
DIPP
Project Setup
Data Preparation
Data Quality Analysis
Exploratory Data Analysis
Visualizations
Segmentation
Modeling
Text Mining
Add Dataset
Variable Categorization
Dataset View
Panel Configuration
Date Configuration
Data Dictionary
Business Categorization
Filters
Sampling
Scenario Builder
404-A
404-B
404-C
404
404-N
402
Select Dataset
Import from Data Source
Project Datasets
Observations
Variables
Size
Active
Export Dataset
944269
10123
Click Load Datasets and provide the path from where you want to upload datasets.
The menu layout guides the user both vertically and horizontally through the analytics' best-practices

FIG. 4

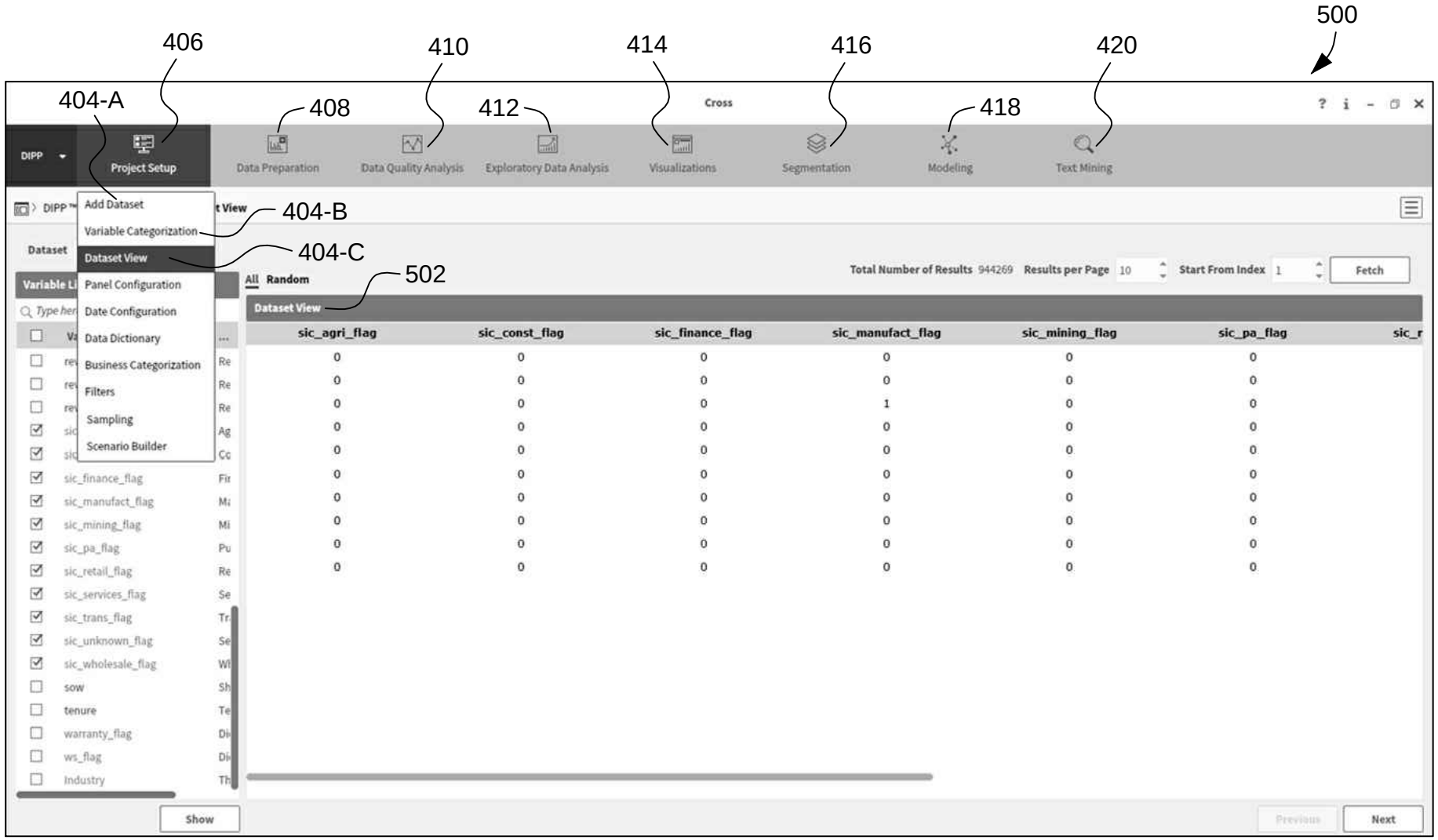

500
406
410
414
416
420
404-A
408
412
418
Cross
DIPP
Project Setup
Data Preparation
Data Quality Analysis
Exploratory Data Analysis
Visualizations
Segmentation
Modeling
Text Mining
Add Dataset
Variable Categorization
Dataset View
Panel Configuration
Date Configuration
Data Dictionary
Business Categorization
Filters
Sampling
Scenario Builder
404-B
404-C
502
Dataset
All
Random
Total Number of Results 944269
Results per Page 10
Start From Index 1
Fetch
Dataset View
sic_agri_flag
sic_const_flag
sic_finance_flag
sic_manufact_flag
sic_mining_flag
sic_pa_flag
sic_finance_flag
sic_manufact_flag
sic_mining_flag
sic_pa_flag
sic_retail_flag
sic_services_flag
sic_trans_flag
sic_unknown_flag
sic_wholesale_flag
sow
tenure
warranty_flag
ws_flag
Industry
Show
Previous
Next

FIG. 5

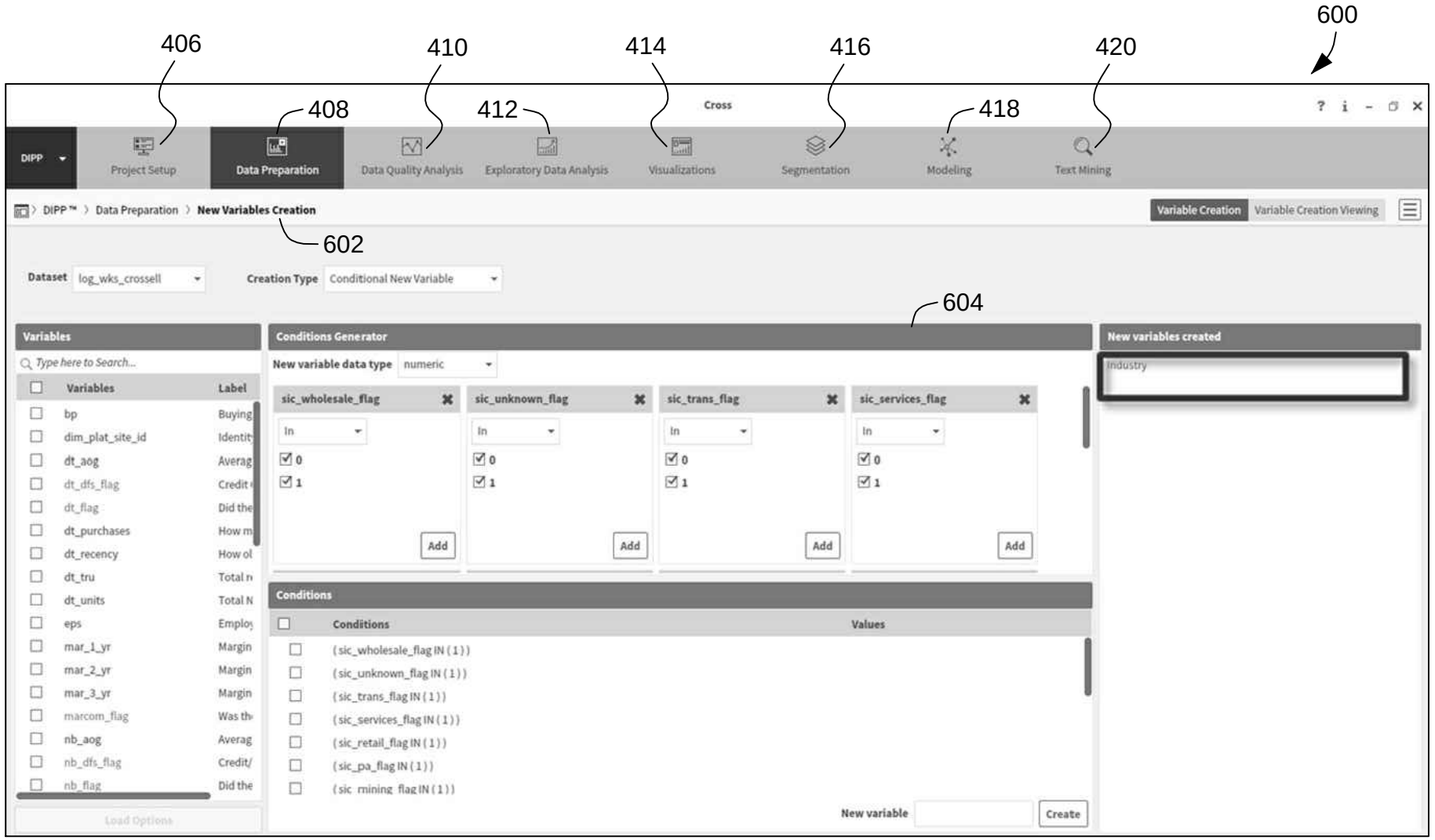

600
406
410
414
416
420
408
412
418
Cross
DIPP
Project Setup
Data Preparation
Data Quality Analysis
Exploratory Data Analysis
Visualizations
Segmentation
Modeling
Text Mining
DIPP™ 〉 Data Preparation 〉 New Variables Creation
602
Variable Creation
Variable Creation Viewing
Dataset log_wks_crossell
Creation Type Conditional New Variable
604
Variables
Type here to Search...
Variables
Label
bp
dim_plat_site_id
dt_aog
dt_dfs_flag
dt_flag
dt_purchases
dt_recency
dt_tru
dt_units
eps
mar_1_yr
mar_2_yr
mar_3_yr
marcom_flag
nb_aog
nb_dfs_flag
nb_flag
Load Options
Conditions Generator
New variable data type numeric
sic_wholesale_flag
sic_unknown_flag
sic_trans_flag
sic_services_flag
In
0
1
Add
Conditions
Conditions
Values
(sic_wholesale_flag IN (1))
(sic_unknown_flag IN (1))
(sic_trans_flag IN (1))
(sic_services_flag IN (1))
(sic_retail_flag IN (1))
(sic_pa_flag IN (1))
(sic_mining_flag IN (1))
New variable
Create
New variables created
Industry

FIG. 6

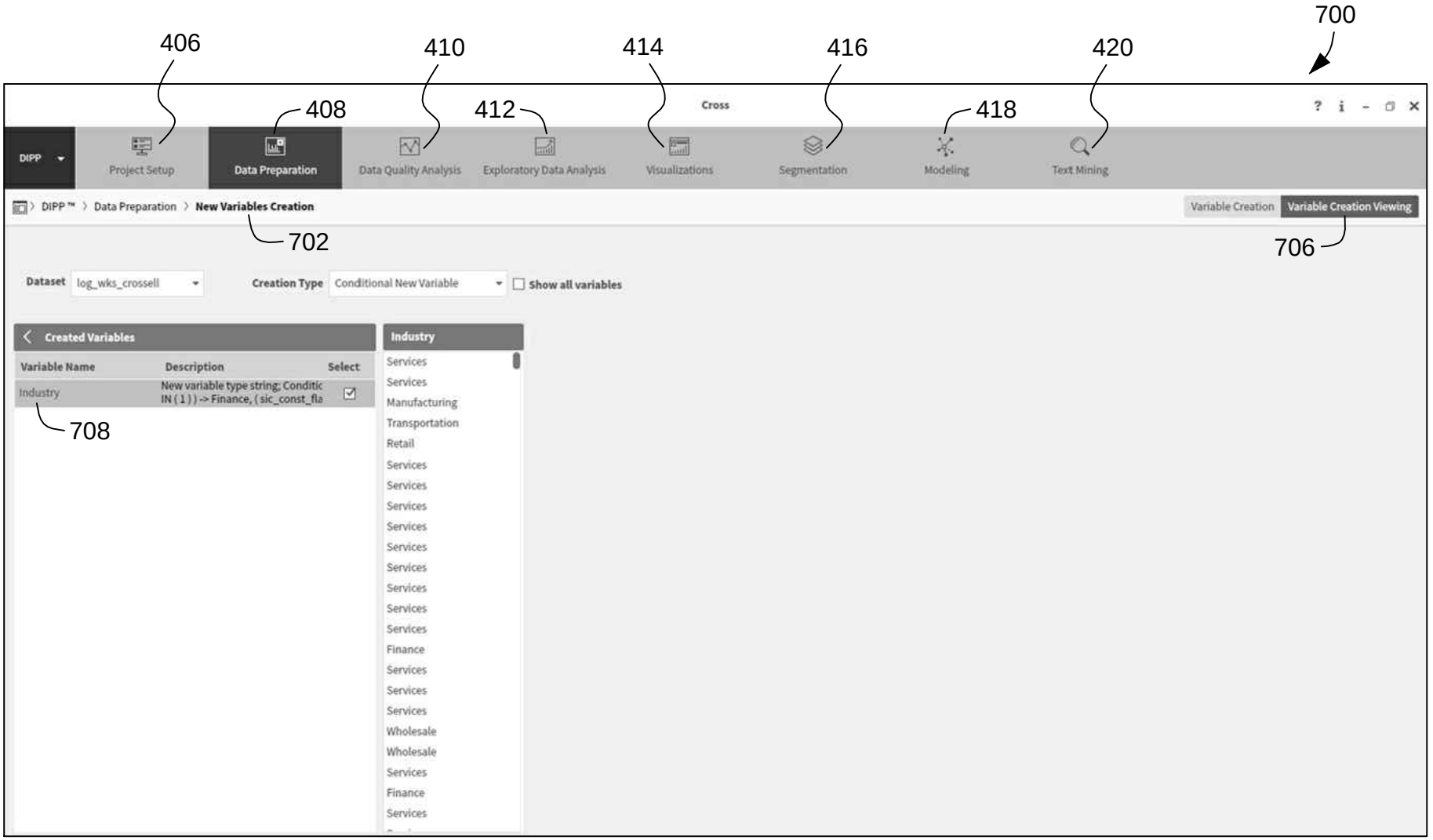

700
406
410
414
416
420
408
412
418
Cross
DIPP
Project Setup
Data Preparation
Data Quality Analysis
Exploratory Data Analysis
Visualizations
Segmentation
Modeling
Text Mining
DIPP™ > Data Preparation > New Variables Creation
Variable Creation
Variable Creation Viewing
702
706
Dataset
log_wks_crossell
Creation Type
Conditional New Variable
Show all variables
Created Variables
Variable Name
Description
Select
Industry
New variable type string; Conditic
IN (1) -> Finance, (sic_const_fla
708
Industry
Services
Services
Manufacturing
Transportation
Retail
Services
Services
Services
Services
Services
Services
Services
Services
Services
Services
Finance
Services
Services
Services
Wholesale
Wholesale
Services
Finance
Services

FIG. 7

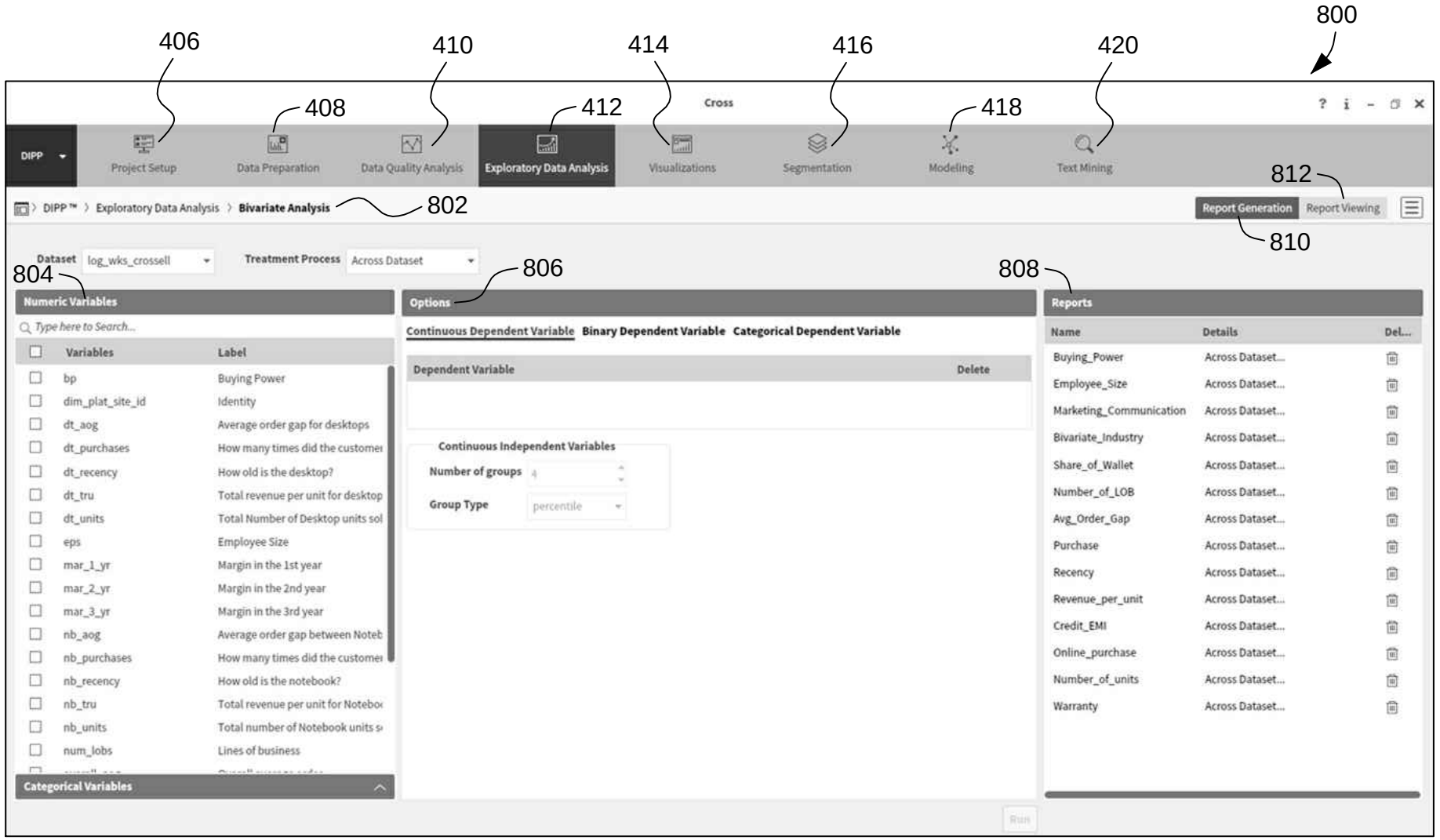

800
406
410
414
416
420
408
412
418
Cross
DIPP
Project Setup
Data Preparation
Data Quality Analysis
Exploratory Data Analysis
Visualizations
Segmentation
Modeling
Text Mining
812
DIPP ™ 〉 Exploratory Data Analysis 〉 Bivariate Analysis
802
Report Generation
Report Viewing
810
Dataset log_wks_crossell
Treatment Process Across Dataset
804
806
808
Numeric Variables
Type here to Search...
Variables
Label
bp Buying Power
dim_plat_site_id Identity
dt_aog Average order gap for desktops
dt_purchases How many times did the customer
dt_recency How old is the desktop?
dt_tru Total revenue per unit for desktop
dt_units Total Number of Desktop units sol
eps Employee Size
mar_1_yr Margin in the 1st year
mar_2_yr Margin in the 2nd year
mar_3_yr Margin in the 3rd year
nb_aog Average order gap between Noteb
nb_purchases How many times did the customer
nb_recency How old is the notebook?
nb_tru Total revenue per unit for Noteboo
nb_units Total number of Notebook units s
num_lobs Lines of business
Categorical Variables
Options
Continuous Dependent Variable
Binary Dependent Variable
Categorical Dependent Variable
Dependent Variable
Delete
Continuous Independent Variables
Number of groups 4
Group Type percentile
Run
Reports
Name
Details
Del...
Buying_Power Across Dataset...
Employee_Size Across Dataset...
Marketing_Communication Across Dataset...
Bivariate_Industry Across Dataset...
Share_of_Wallet Across Dataset...
Number_of_LOB Across Dataset...
Avg_Order_Gap Across Dataset...
Purchase Across Dataset...
Recency Across Dataset...
Revenue_per_unit Across Dataset...
Credit_EMI Across Dataset...
Online_purchase Across Dataset...
Number_of_units Across Dataset...
Warranty Across Dataset...

FIG. 8

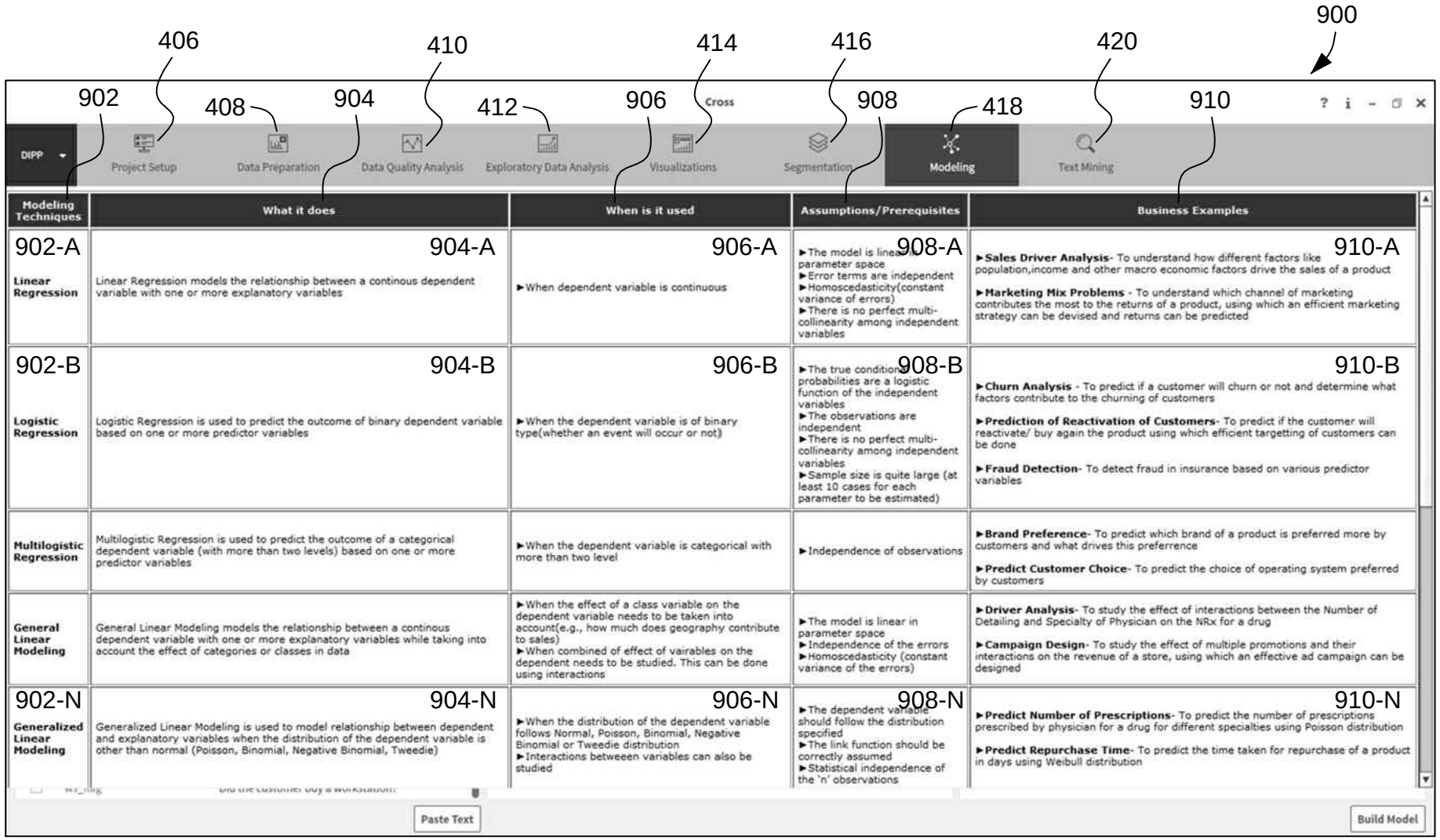

900
406
410
414
416
420
902
408
904
412
906
Cross
908
418
910
DIPP
Project Setup
Data Preparation
Data Quality Analysis
Exploratory Data Analysis
Visualizations
Segmentation
Modeling
Text Mining
Modeling Techniques
What it does
When is it used
Assumptions/Prerequisites
Business Examples
902-A
Linear Regression
Linear Regression models the relationship between a continous dependent variable with one or more explanatory variables
904-A
►When dependent variable is continuous
906-A
►The model is linear in parameter space
►Error terms are independent
►Homoscedasticity(constant variance of errors)
►There is no perfect multi-collinearity among independent variables
908-A
►Sales Driver Analysis- To understand how different factors like population,income and other macro economic factors drive the sales of a product
►Marketing Mix Problems - To understand which channel of marketing contributes the most to the returns of a product, using which an efficient marketing strategy can be devised and returns can be predicted
910-A
902-B
Logistic Regression
Logistic Regression is used to predict the outcome of binary dependent variable based on one or more predictor variables
904-B
►When the dependent variable is of binary type(whether an event will occur or not)
906-B
►The true conditional probabilities are a logistic function of the independent variables
►The observations are independent
►There is no perfect multi-collinearity among independent variables
►Sample size is quite large (at least 10 cases for each parameter to be estimated)
908-B
►Churn Analysis - To predict if a customer will churn or not and determine what factors contribute to the churning of customers
►Prediction of Reactivation of Customers- To predict if the customer will reactivate/ buy again the product using which efficient targetting of customers can be done
►Fraud Detection- To detect fraud in insurance based on various predictor variables
910-B
Multilogistic Regression
Multilogistic Regression is used to predict the outcome of a categorical dependent variable (with more than two levels) based on one or more predictor variables
►When the dependent variable is categorical with more than two level
►Independence of observations
►Brand Preference- To predict which brand of a product is preferred more by customers and what drives this preferrence
►Predict Customer Choice- To predict the choice of operating system preferred by customers
General Linear Modeling
General Linear Modeling models the relationship between a continous dependent variable with one or more explanatory variables while taking into account the effect of categories or classes in data
►When the effect of a class variable on the dependent variable needs to be taken into account(e.g., how much does geography contribute to sales)
►When combined of effect of vairables on the dependent needs to be studied. This can be done using interactions
►The model is linear in parameter space
►Independence of the errors
►Homoscedasticity (constant variance of the errors)
►Driver Analysis- To study the effect of interactions between the Number of Detailing and Specialty of Physician on the NRx for a drug
►Campaign Design- To study the effect of multiple promotions and their interactions on the revenue of a store, using which an effective ad campaign can be designed
902-N
Generalized Linear Modeling
Generalized Linear Modeling is used to model relationship between dependent and explanatory variables when the distribution of the dependent variable is other than normal (Poisson, Binomial, Negative Binomial, Tweedie)
904-N
►When the distribution of the dependent variable follows Normal, Poisson, Binomial, Negative Binomial or Tweedie distribution
►Interactions betweeen variables can also be studied
906-N
►The dependent variable should follow the distribution specified
►The link function should be correctly assumed
►Statistical independence of the 'n' observations
908-N
►Predict Number of Prescriptions- To predict the number of prescriptions prescribed by physician for a drug for different specialties using Poisson distribution
►Predict Repurchase Time- To predict the time taken for repurchase of a product in days using Weibull distribution
910-N
Paste Text
Build Model

FIG. 9

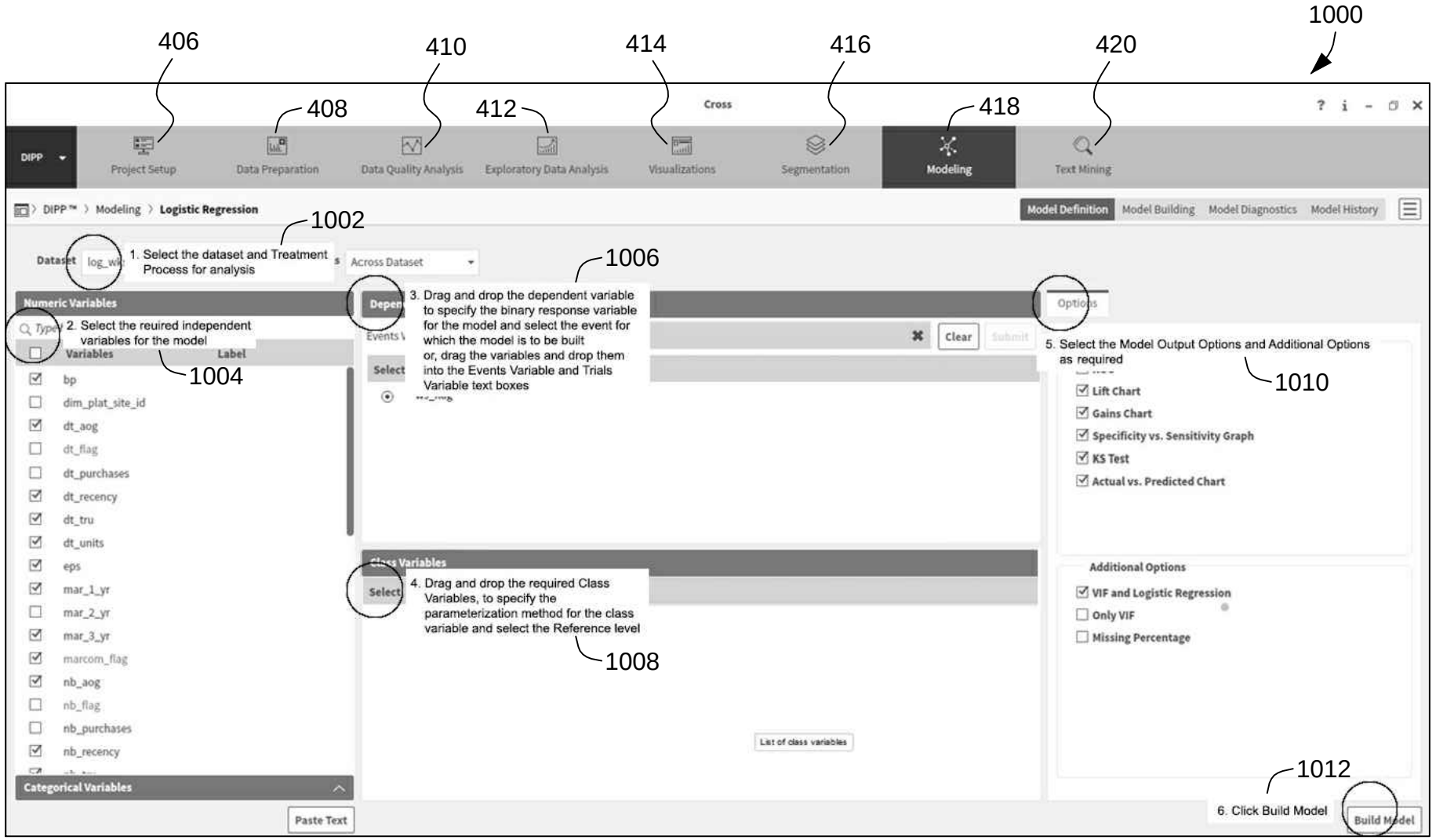

1000
406
410
414
416
420
408
412
418
Cross
DIPP
Project Setup
Data Preparation
Data Quality Analysis
Exploratory Data Analysis
Visualizations
Segmentation
Modeling
Text Mining
DIPP™ > Modeling > Logistic Regression
Model Definition
Model Building
Model Diagnostics
Model History
1002
Dataset
1. Select the dataset and Treatment Process for analysis
Across Dataset
1006
Numeric Variables
2. Select the reuired independent variables for the model
Variables
Label
1004
bp
dim_plat_site_id
dt_aog
dt_flag
dt_purchases
dt_recency
dt_tru
dt_units
eps
mar_1_yr
mar_2_yr
mar_3_yr
marcom_flag
nb_aog
nb_flag
nb_purchases
nb_recency
Categorical Variables
3. Drag and drop the dependent variable to specify the binary response variable for the model and select the event for which the model is to be built or, drag the variables and drop them into the Events Variable and Trials Variable text boxes
Clear
Submit
Class Variables
4. Drag and drop the required Class Variables, to specify the parameterization method for the class variable and select the Reference level
1008
List of class variables
Paste Text
Options
5. Select the Model Output Options and Additional Options as required
1010
Lift Chart
Gains Chart
Specificity vs. Sensitivity Graph
KS Test
Actual vs. Predicted Chart
Additional Options
VIF and Logistic Regression
Only VIF
Missing Percentage
1012
6. Click Build Model
Build Model

FIG. 10

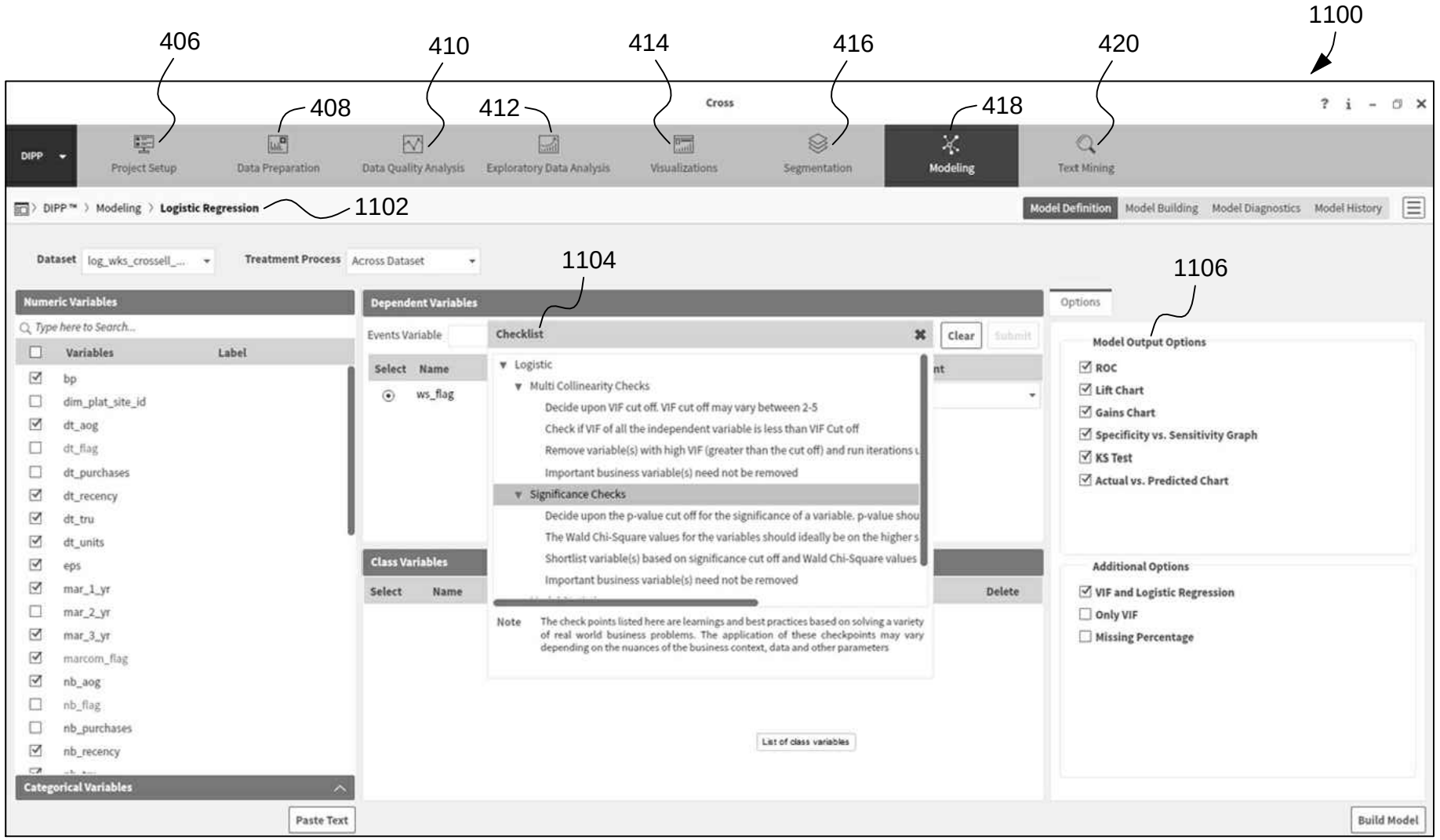

1100
406
410
414
416
420
408
412
418
Cross
DIPP
Project Setup
Data Preparation
Data Quality Analysis
Exploratory Data Analysis
Visualizations
Segmentation
Modeling
Text Mining
DIPP™ 〉 Modeling 〉 Logistic Regression
1102
Model Definition
Model Building
Model Diagnostics
Model History
Dataset
log_wks_crossell_...
Treatment Process
Across Dataset
1104
1106
Numeric Variables
Type here to Search...
Variables
Label
bp
dim_plat_site_id
dt_aog
dt_flag
dt_purchases
dt_recency
dt_tru
dt_units
eps
mar_1_yr
mar_2_yr
mar_3_yr
marcom_flag
nb_aog
nb_flag
nb_purchases
nb_recency
Categorical Variables
Paste Text
Dependent Variables
Events Variable
Checklist
Clear
Submit
Select
Name
ws_flag
Logistic
Multi Collinearity Checks
Decide upon VIF cut off. VIF cut off may vary between 2-5
Check if VIF of all the independent variable is less than VIF Cut off
Remove variable(s) with high VIF (greater than the cut off) and run iterations
Important business variable(s) need not be removed
Significance Checks
Decide upon the p-value cut off for the significance of a variable. p-value shou
The Wald Chi-Square values for the variables should ideally be on the higher s
Shortlist variable(s) based on significance cut off and Wald Chi-Square values
Important business variable(s) need not be removed
Class Variables
Select
Name
Delete
Note
The check points listed here are learnings and best practices based on solving a variety of real world business problems. The application of these checkpoints may vary depending on the nuances of the business context, data and other parameters
List of class variables
Options
Model Output Options
ROC
Lift Chart
Gains Chart
Specificity vs. Sensitivity Graph
KS Test
Actual vs. Predicted Chart
Additional Options
VIF and Logistic Regression
Only VIF
Missing Percentage
Build Model

FIG. 11

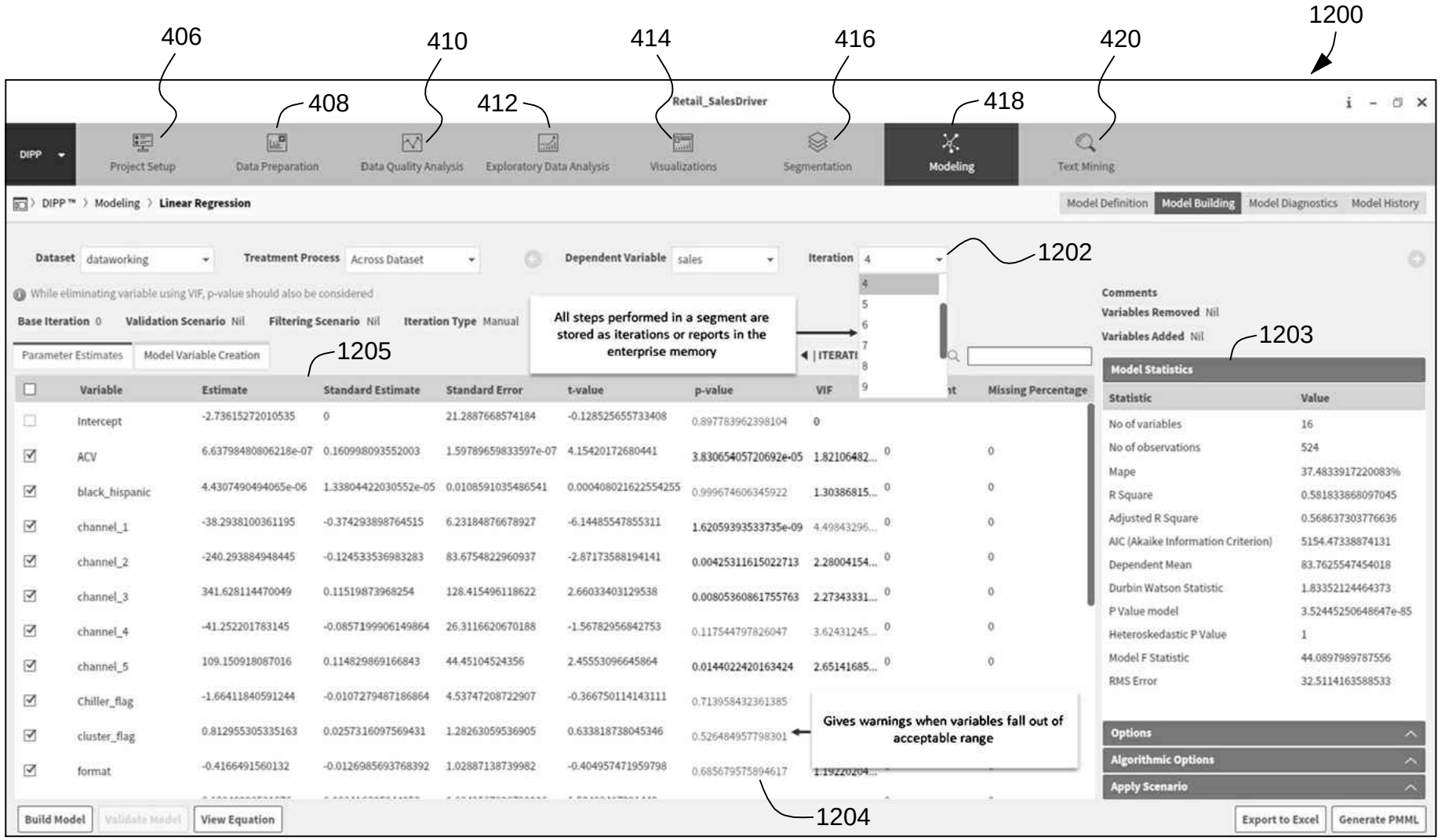

1200
406
410
414
416
420
408
412
418
Retail_SalesDriver
DIPP
Project Setup
Data Preparation
Data Quality Analysis
Exploratory Data Analysis
Visualizations
Segmentation
Modeling
Text Mining
DIPP™ > Modeling > Linear Regression
Model Definition
Model Building
Model Diagnostics
Model History
Dataset dataworking
Treatment Process Across Dataset
Dependent Variable sales
Iteration 4
1202
While eliminating variable using VIF, p-value should also be considered
Base Iteration 0
Validation Scenario Nil
Filtering Scenario Nil
Iteration Type Manual
All steps performed in a segment are stored as iterations or reports in the enterprise memory
Comments
Variables Removed Nil
Variables Added Nil
1203
Parameter Estimates
Model Variable Creation
1205
Variable
Estimate
Standard Estimate
Standard Error
t-value
p-value
VIF
Missing Percentage
Intercept -2.73615272010535 0 21.2887668574184 -0.128525655733408 0.897783962398104 0
ACV 6.63798480806218e-07 0.160998093552003 1.59789659833597e-07 4.15420172680441 3.83065405720692e-05 1.82106482... 0 0
black_hispanic 4.4307490494065e-06 1.33804422030552e-05 0.0108591035486541 0.000408021622554255 0.999674606345922 1.30386815... 0 0
channel_1 -38.2938100361195 -0.374293898764515 6.23184876678927 -6.14485547855311 1.62059393533735e-09 4.49843296... 0 0
channel_2 -240.293884948445 -0.124533536983283 83.6754822960937 -2.87173588194141 0.00425311615022713 2.28004154... 0 0
channel_3 341.628114470049 0.11519873968254 128.415496118622 2.66033403129538 0.00805360861755763 2.27343331... 0 0
channel_4 -41.252201783145 -0.0857199906149864 26.3116620670188 -1.56782956842753 0.117544797826047 3.62431245... 0 0
channel_5 109.150918087016 0.114829869166843 44.45104524356 2.45553096645864 0.0144022420163424 2.65141685... 0 0
Chiller_flag -1.66411840591244 -0.0107279487186864 4.53747208722907 -0.366750114143111 0.713958432361385
cluster_flag 0.812955305335163 0.0257316097569431 1.28263059536905 0.633818738045346 0.526484957798301
format -0.4166491560132 -0.0126985693768392 1.02887138739982 -0.404957471959798 0.685679575894617
Gives warnings when variables fall out of acceptable range
1204
Model Statistics
Statistic
Value
No of variables 16
No of observations 524
Mape 37.48339172200 83%
R Square 0.581833868097045
Adjusted R Square 0.568637303776636
AIC (Akaike Information Criterion) 5154.47338874131
Dependent Mean 83.7625547454018
Durbin Watson Statistic 1.83352124464373
P Value model 3.52445250648647e-85
Heteroskedastic P Value 1
Model F Statistic 44.0897989787556
RMS Error 32.5114163588533
Options
Algorithmic Options
Apply Scenario
Build Model
Validate Model
View Equation
Export to Excel
Generate PMML

FIG. 12

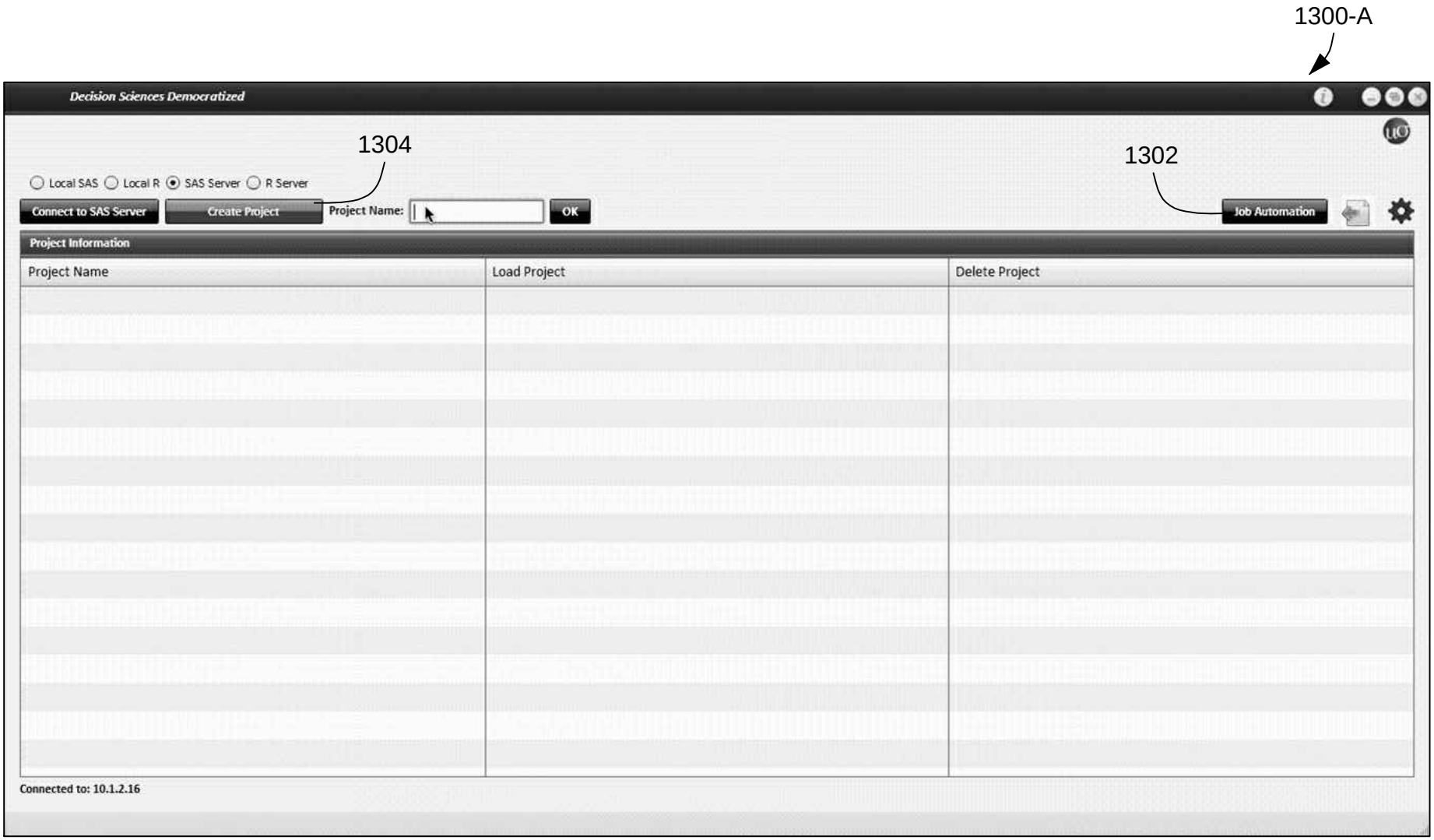

1300-A
Decision Sciences Democratized
1304
1302
Local SAS
Local R
SAS Server
R Server
Connect to SAS Server
Create Project
Project Name:
OK
Job Automation
Project Information
Project Name
Load Project
Delete Project
Connected to: 10.1.2.16

FIG. 13-A

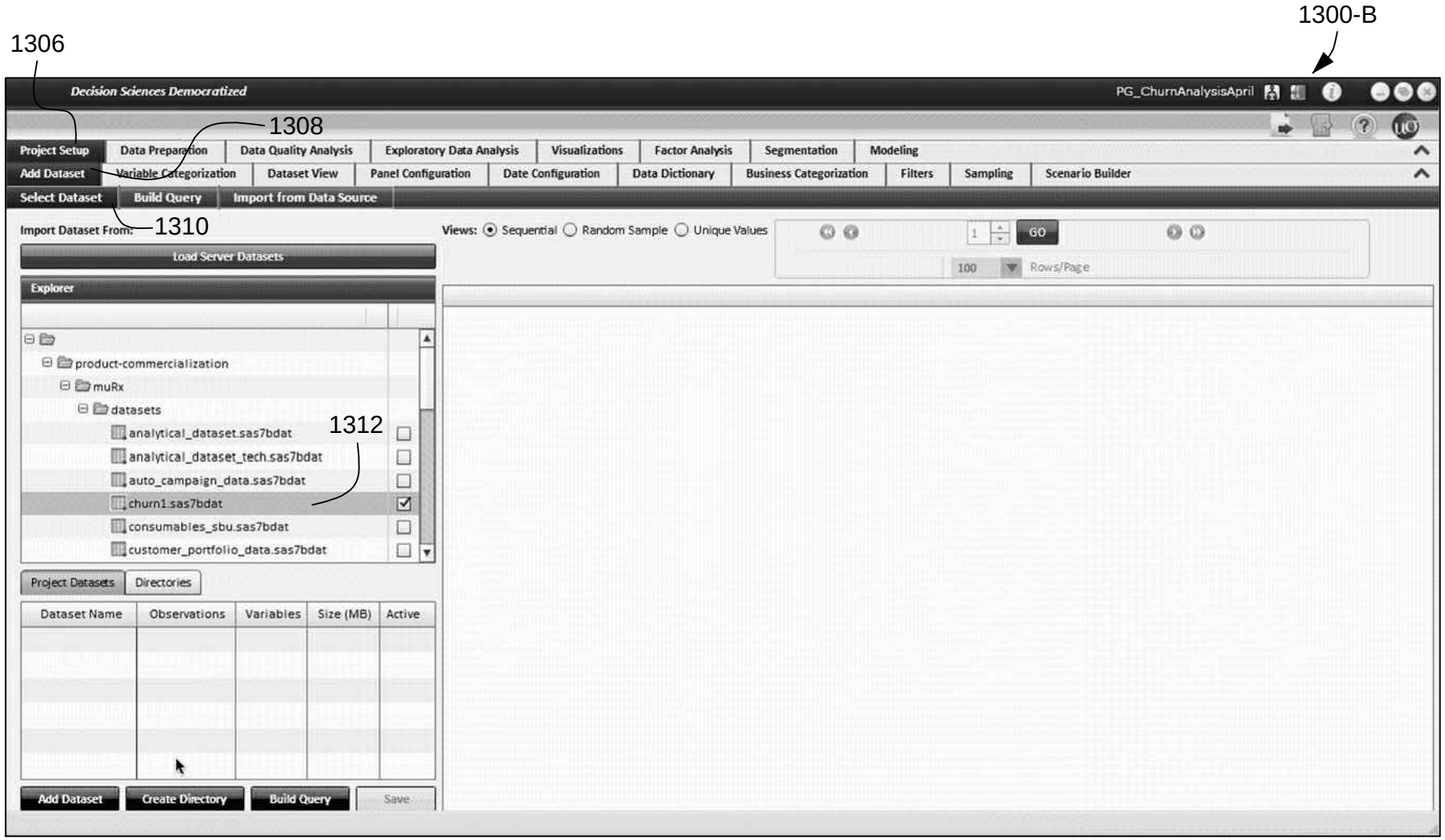

1300-B
1306
1308
1310
1312
Decision Sciences Democratized
PG_ChurnAnalysisApril
Project Setup
Data Preparation
Data Quality Analysis
Exploratory Data Analysis
Visualizations
Factor Analysis
Segmentation
Modeling
Add Dataset
Variable Categorization
Dataset View
Panel Configuration
Date Configuration
Data Dictionary
Business Categorization
Filters
Sampling
Scenario Builder
Select Dataset
Build Query
Import from Data Source
Import Dataset From:
Load Server Datasets
Views: Sequential Random Sample Unique Values
GO
100
Rows/Page
Explorer
product-commercialization
muRx
datasets
analytical_dataset.sas7bdat
analytical_dataset_tech.sas7bdat
auto_campaign_data.sas7bdat
churn1.sas7bdat
consumables_sbu.sas7bdat
customer_portfolio_data.sas7bdat
Project Datasets
Directories
Dataset Name
Observations
Variables
Size (MB)
Active
Add Dataset
Create Directory
Build Query
Save

FIG. 13-B

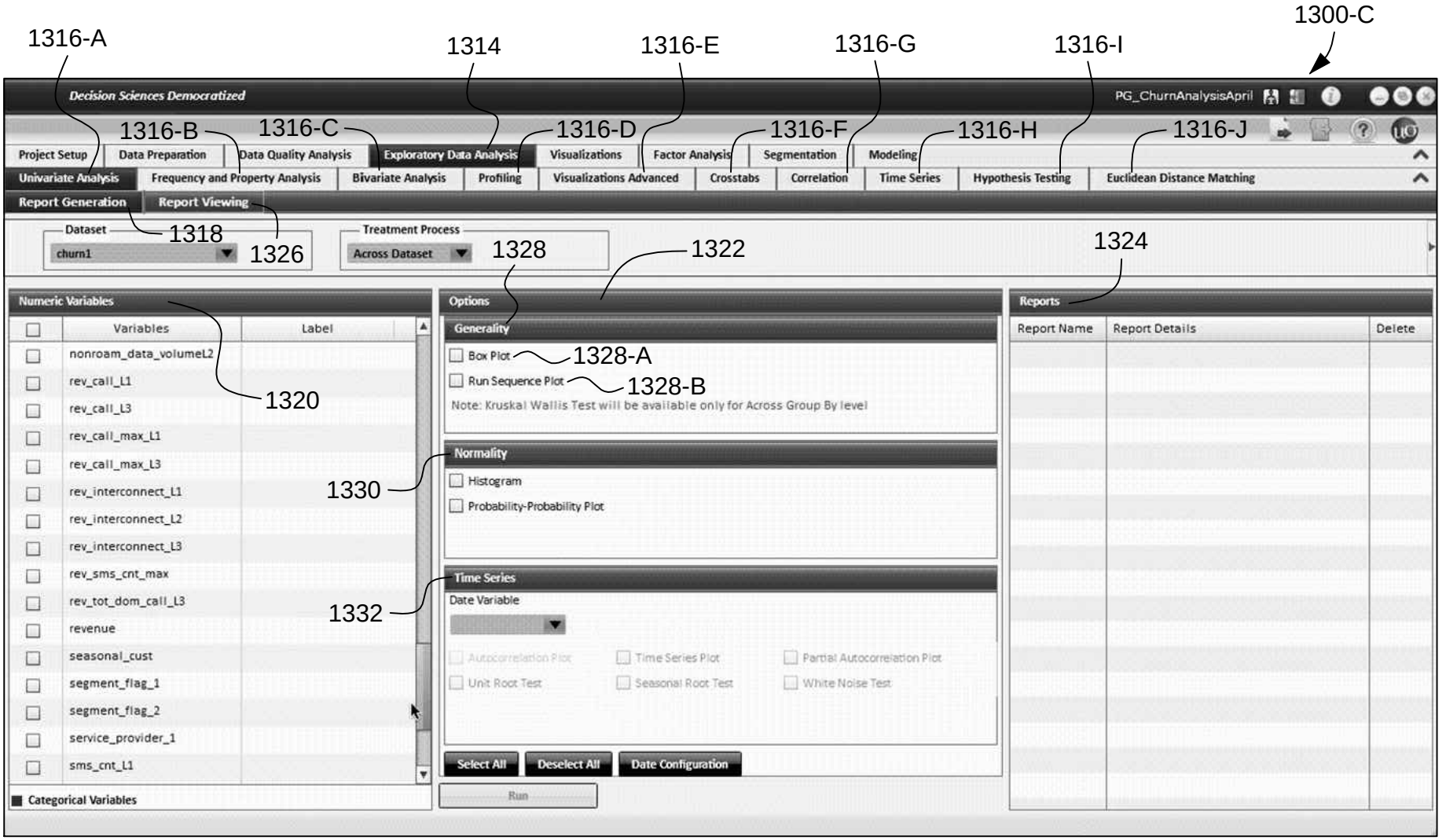

1300-C
1316-A
1314
1316-E
1316-G
1316-I
Decision Sciences Democratized
PG_ChurnAnalysisApril
1316-B
1316-C
1316-D
1316-F
1316-H
1316-J
Project Setup
Data Preparation
Data Quality Analysis
Exploratory Data Analysis
Visualizations
Factor Analysis
Segmentation
Modeling
Univariate Analysis
Frequency and Property Analysis
Bivariate Analysis
Profiling
Visualizations Advanced
Crosstabs
Correlation
Time Series
Hypothesis Testing
Euclidean Distance Matching
Report Generation
Report Viewing
1318
1326
Dataset
churn1
Treatment Process
Across Dataset
1328
1322
1324
Numeric Variables
Variables
Label
nonroam_data_volumeL2
rev_call_L1
rev_call_L3
rev_call_max_L1
rev_call_max_L3
rev_interconnect_L1
rev_interconnect_L2
rev_interconnect_L3
rev_sms_cnt_max
rev_tot_dom_call_L3
revenue
seasonal_cust
segment_flag_1
segment_flag_2
service_provider_1
sms_cnt_L1
1320
Categorical Variables
Options
Generality
Box Plot
1328-A
Run Sequence Plot
1328-B
Note: Kruskal Wallis Test will be available only for Across Group By level
Normality
1330
Histogram
Probability-Probability Plot
Time Series
1332
Date Variable
Autocorrelation Plot
Time Series Plot
Partial Autocorrelation Plot
Unit Root Test
Seasonal Root Test
White Noise Test
Select All
Deselect All
Date Configuration
Run
Reports
Report Name
Report Details
Delete

FIG. 13-C

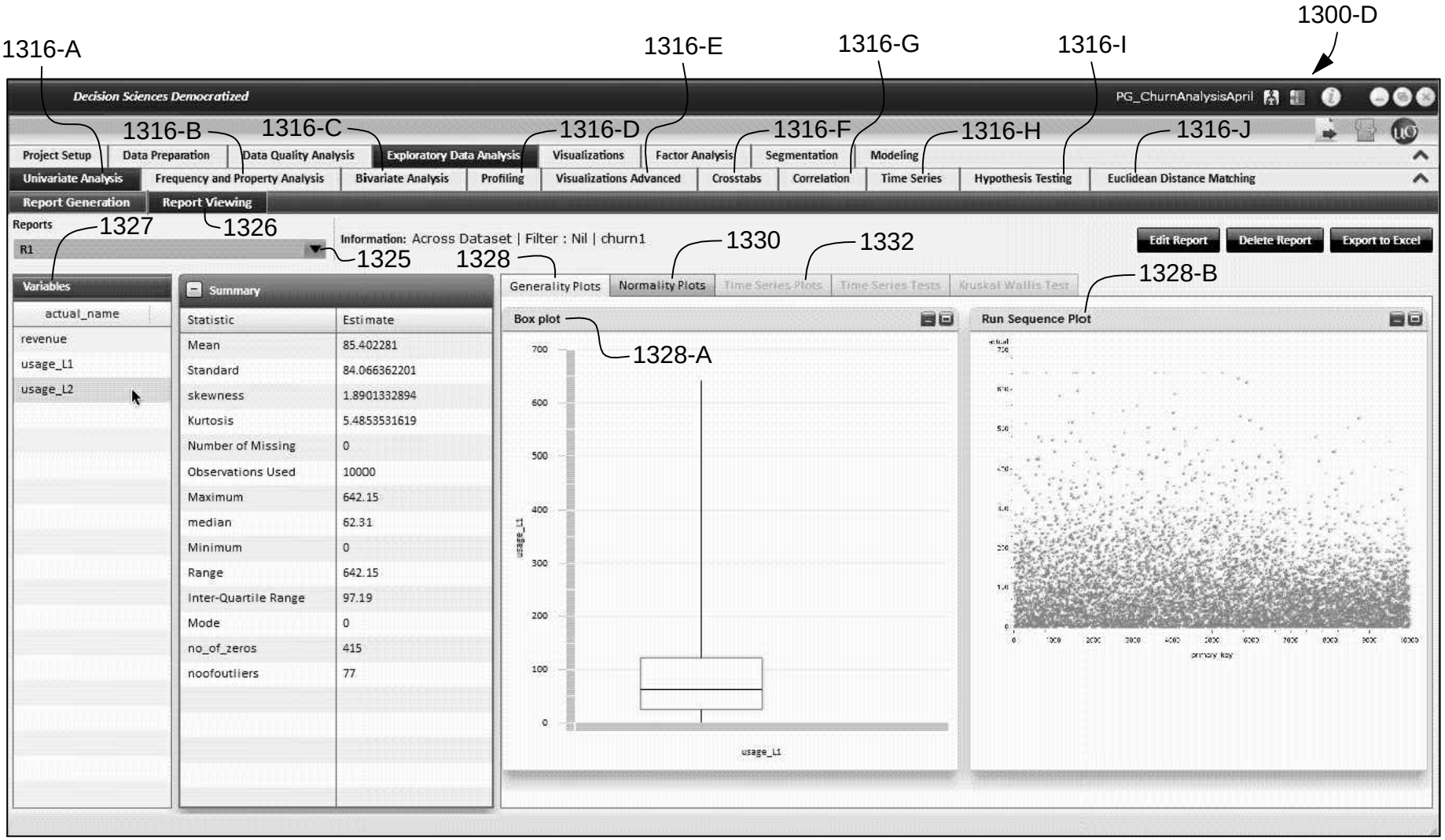

1300-D
1316-A
1316-B
1316-C
1316-D
1316-E
1316-F
1316-G
1316-H
1316-I
1316-J
Decision Sciences Democratized
PG_ChurnAnalysisApril
Project Setup
Data Preparation
Data Quality Analysis
Exploratory Data Analysis
Visualizations
Factor Analysis
Segmentation
Modeling
Univariate Analysis
Frequency and Property Analysis
Bivariate Analysis
Profiling
Visualizations Advanced
Crosstabs
Correlation
Time Series
Hypothesis Testing
Euclidean Distance Matching
Report Generation
Report Viewing
Reports
1327
1326
R1
1325
Information: Across Dataset | Filter : Nil | churn1
1330
1332
1328
Edit Report
Delete Report
Export to Excel
Variables
actual_name
revenue
usage_L1
usage_L2
Summary
Statistic
Estimate
Mean 85.402281
Standard 84.066362201
skewness 1.8901332894
Kurtosis 5.4853531619
Number of Missing 0
Observations Used 10000
Maximum 642.15
median 62.31
Minimum 0
Range 642.15
Inter-Quartile Range 97.19
Mode 0
no_of_zeros 415
noofoutliers 77
Generality Plots
Normality Plots
Time Series Plots
Time Series Tests
Kruskal Wallis Test
1328-B
Box plot
1328-A
Run Sequence Plot
usage_L1
primary key

FIG. 13-D


1300-E
1334
1336
Click to view the project work flow
Project Workflow
Project Setup
Data Preparation
EDA
Add dataset
Univariate Analysis

FIG. 13-E

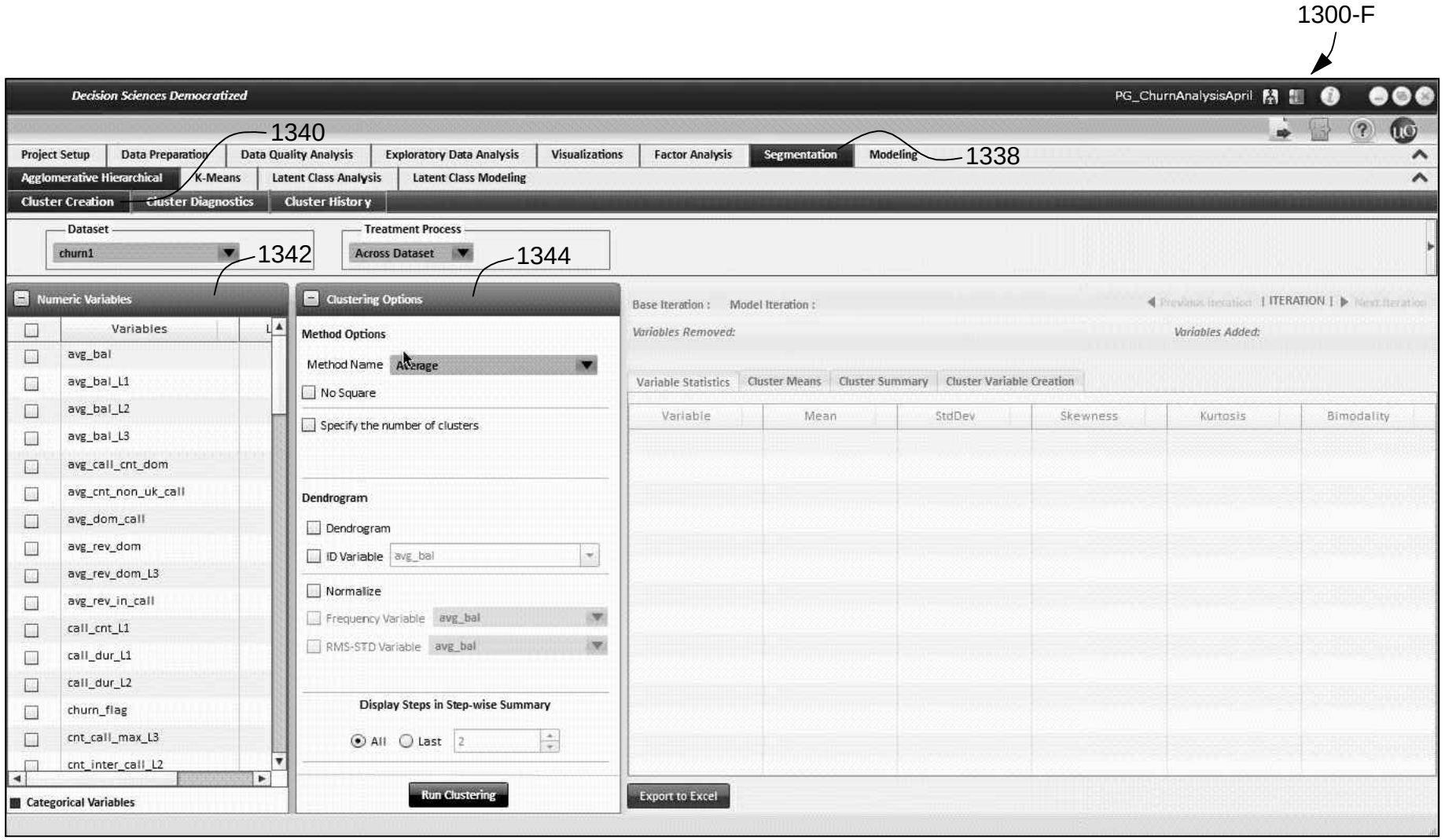

1300-F
Decision Sciences Democratized
PG_ChurnAnalysisApril
1340
1338
Project Setup
Data Preparation
Data Quality Analysis
Exploratory Data Analysis
Visualizations
Factor Analysis
Segmentation
Modeling
Agglomerative Hierarchical
K-Means
Latent Class Analysis
Latent Class Modeling
Cluster Creation
Cluster Diagnostics
Cluster History
Dataset
churn1
1342
Treatment Process
Across Dataset
1344
Numeric Variables
Variables
avg_bal
avg_bal_L1
avg_bal_L2
avg_bal_L3
avg_call_cnt_dom
avg_cnt_non_uk_call
avg_dom_call
avg_rev_dom
avg_rev_dom_L3
avg_rev_in_call
call_cnt_L1
call_dur_L1
call_dur_L2
churn_flag
cnt_call_max_L3
cnt_inter_call_L2
Categorical Variables
Clustering Options
Method Options
Method Name
Average
No Square
Specify the number of clusters
Dendrogram
Dendrogram
ID Variable
avg_bal
Normalize
Frequency Variable
avg_bal
RMS-STD Variable
avg_bal
Display Steps in Step-wise Summary
All
Last
2
Run Clustering
Base Iteration :
Model Iteration :
Previous Iteration
ITERATION
Next Iteration
Variables Removed:
Variables Added:
Variable Statistics
Cluster Means
Cluster Summary
Cluster Variable Creation
Variable
Mean
StdDev
Skewness
Kurtosis
Bimodality
Export to Excel

FIG. 13-F

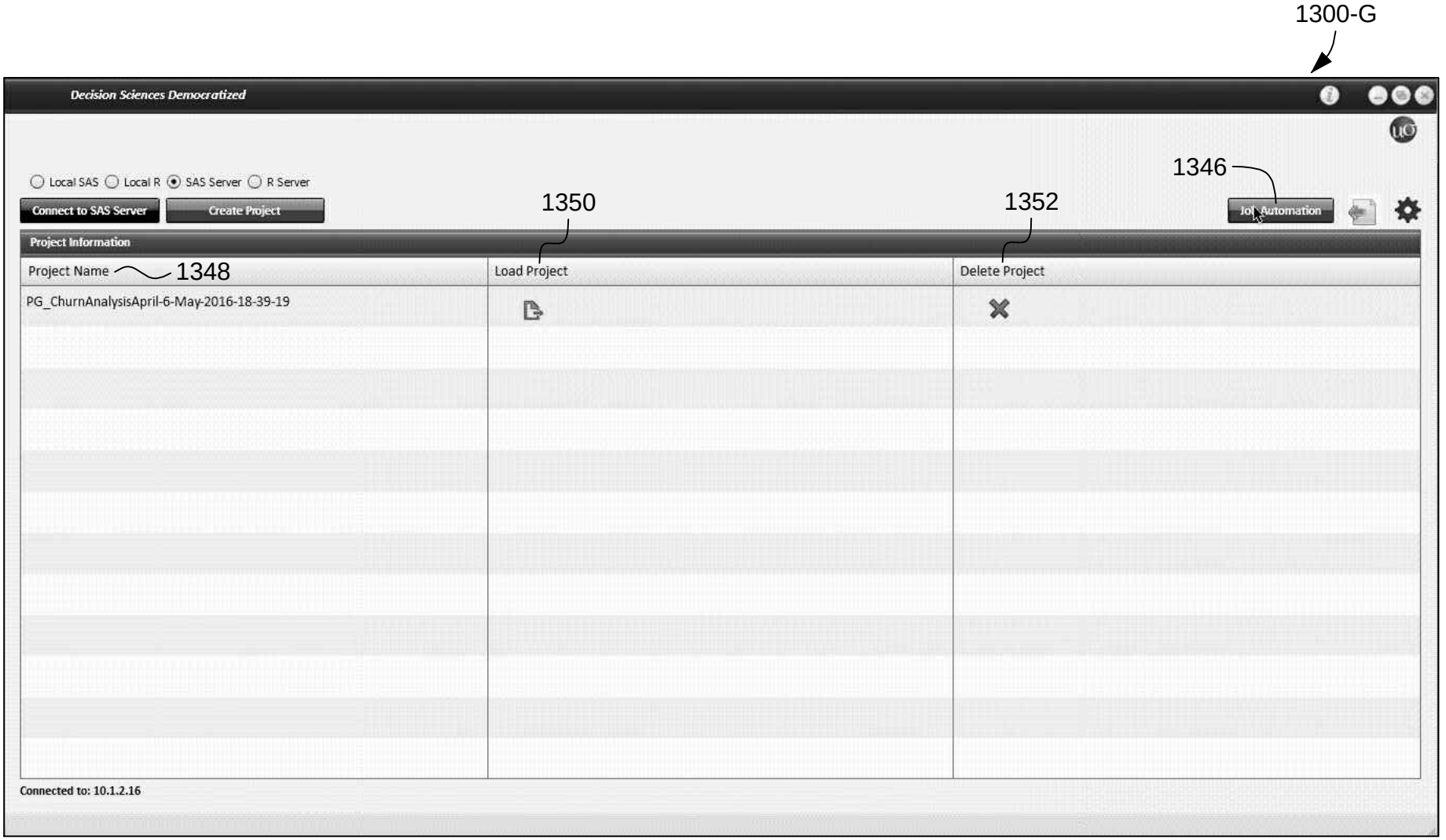

1300-G
Decision Sciences Democratized
Local SAS
Local R
SAS Server
R Server
Connect to SAS Server
Create Project
1346
Job Automation
1350
1352
Project Information
Project Name
1348
Load Project
Delete Project
PG_ChurnAnalysisApril-6-May-2016-18-39-19
Connected to: 10.1.2.16

FIG. 13-G

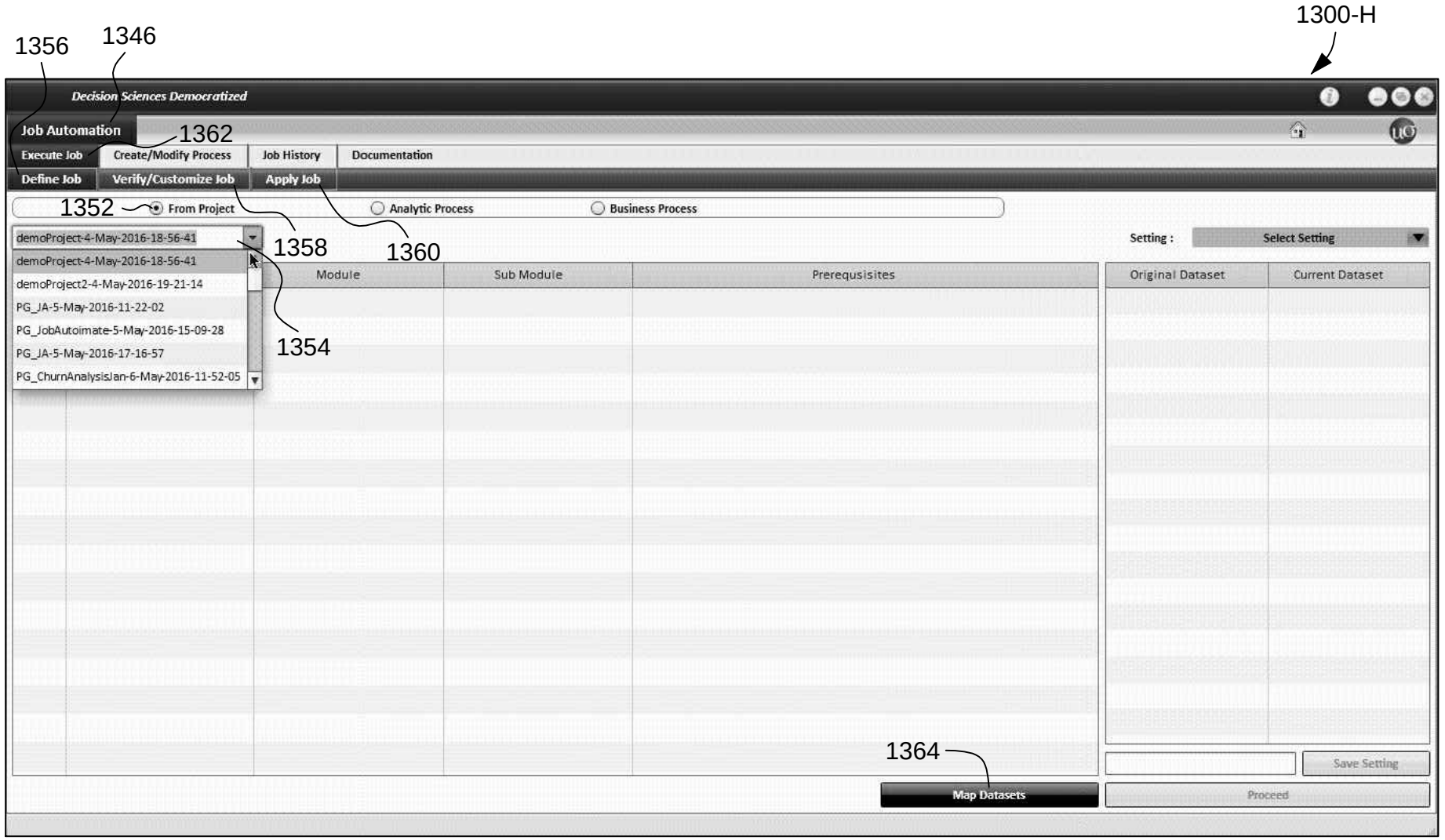

1300-H
1356
1346
1362
Decision Sciences Democratized
Job Automation
Execute Job
Create/Modify Process
Job History
Documentation
Define Job
Verify/Customize Job
Apply Job
1352
From Project
Analytic Process
Business Process
demoProject-4-May-2016-18-56-41
1358
1360
Setting :
Select Setting
demoProject-4-May-2016-18-56-41
demoProject2-4-May-2016-19-21-14
PG_JA-5-May-2016-11-22-02
PG_JobAutoimate-5-May-2016-15-09-28
PG_JA-5-May-2016-17-16-57
PG_ChurnAnalysisJan-6-May-2016-11-52-05
1354
Module
Sub Module
Prerequsisites
Original Dataset
Current Dataset
1364
Map Datasets
Save Setting
Proceed

FIG. 13-H

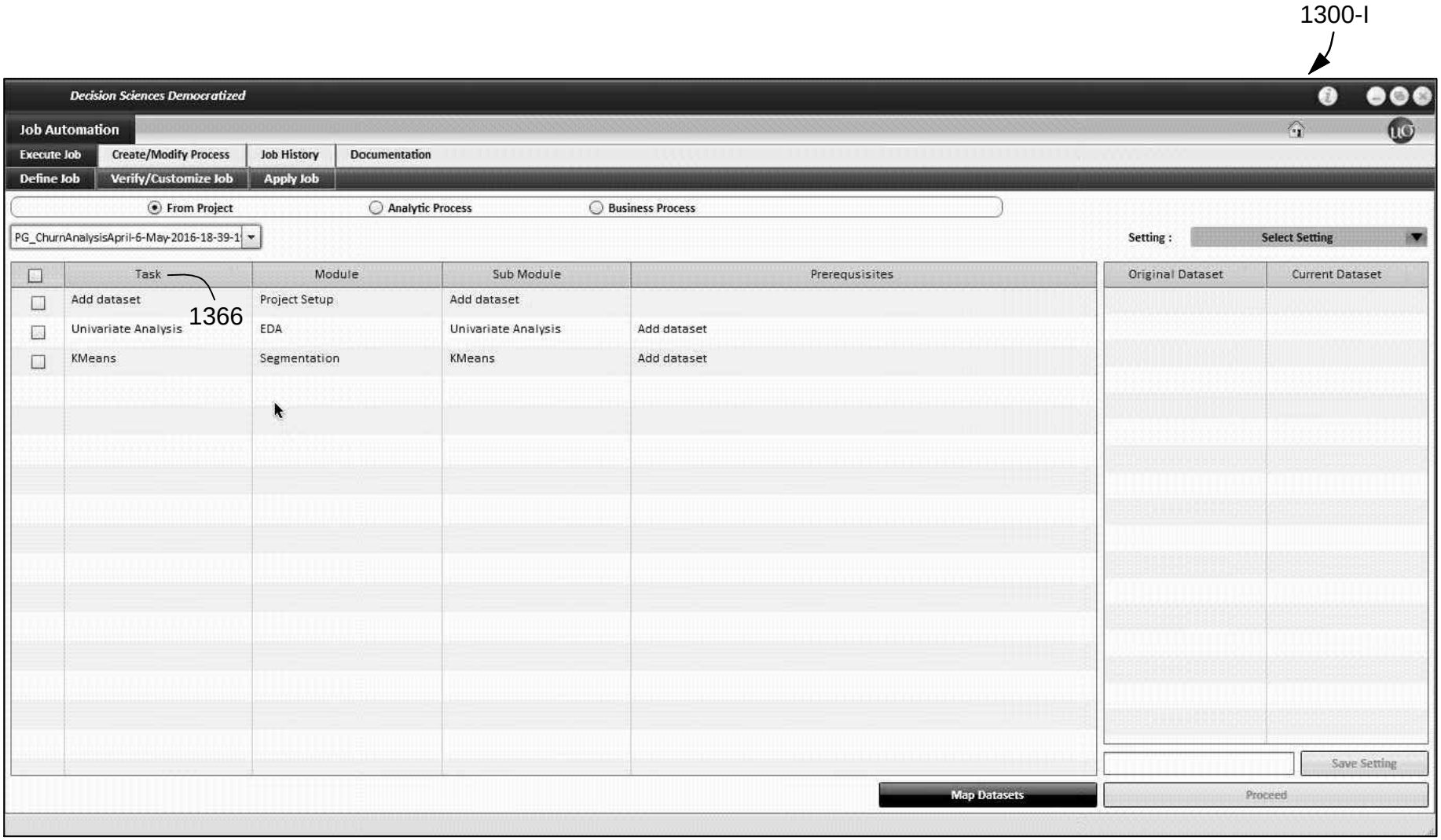

1300-I
Decision Sciences Democratized
Job Automation
Execute Job
Create/Modify Process
Job History
Documentation
Define Job
Verify/Customize Job
Apply Job
From Project
Analytic Process
Business Process
PG_ChurnAnalysisApril-6-May-2016-18-39-1
Setting :
Select Setting
Task
Module
Sub Module
Prerequisites
Original Dataset
Current Dataset
1366
Add dataset
Project Setup
Add dataset
Univariate Analysis
EDA
Univariate Analysis
Add dataset
KMeans
Segmentation
KMeans
Add dataset
Save Setting
Map Datasets
Proceed

FIG. 13-I

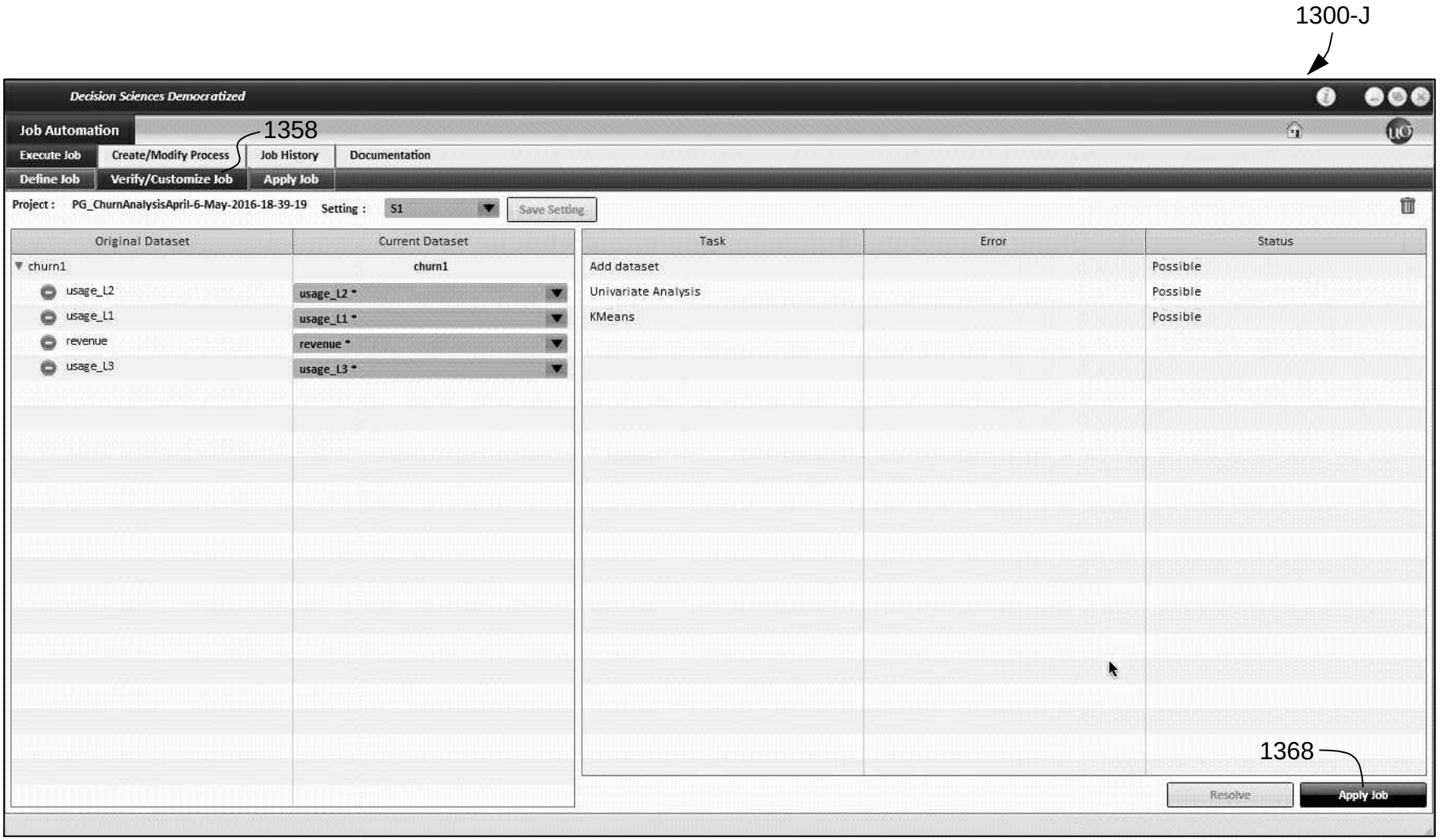

1300-J
Decision Sciences Democratized
Job Automation
1358
Execute Job
Create/Modify Process
Job History
Documentation
Define Job
Verify/Customize Job
Apply Job
Project : PG_ChurnAnalysisApril-6-May-2016-18-39-19
Setting : S1
Save Setting
Original Dataset
Current Dataset
churn1
churn1
usage_L2
usage_L2 *
usage_L1
usage_L1 *
revenue
revenue *
usage_L3
usage_L3 *
Task
Error
Status
Add dataset
Possible
Univariate Analysis
Possible
KMeans
Possible
1368
Resolve
Apply Job

FIG. 13-J

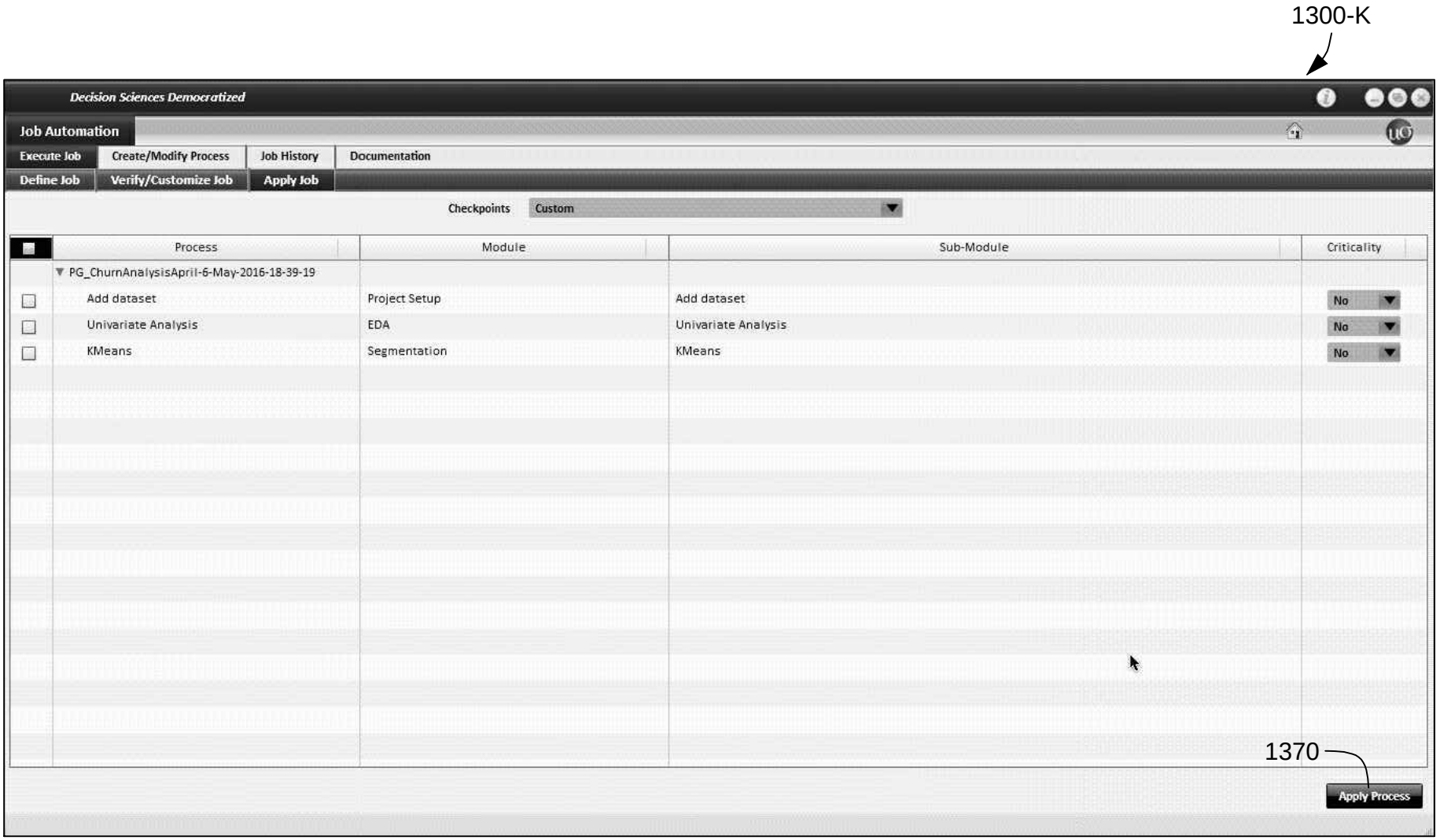

1300-K
Decision Sciences Democratized
Job Automation
Execute Job
Create/Modify Process
Job History
Documentation
Define Job
Verify/Customize Job
Apply Job
Checkpoints
Custom
Process
Module
Sub-Module
Criticality
PG_ChurnAnalysisApril-6-May-2016-18-39-19
Add dataset
Project Setup
Add dataset
No
Univariate Analysis
EDA
Univariate Analysis
No
KMeans
Segmentation
KMeans
No
1370
Apply Process

FIG. 13-K

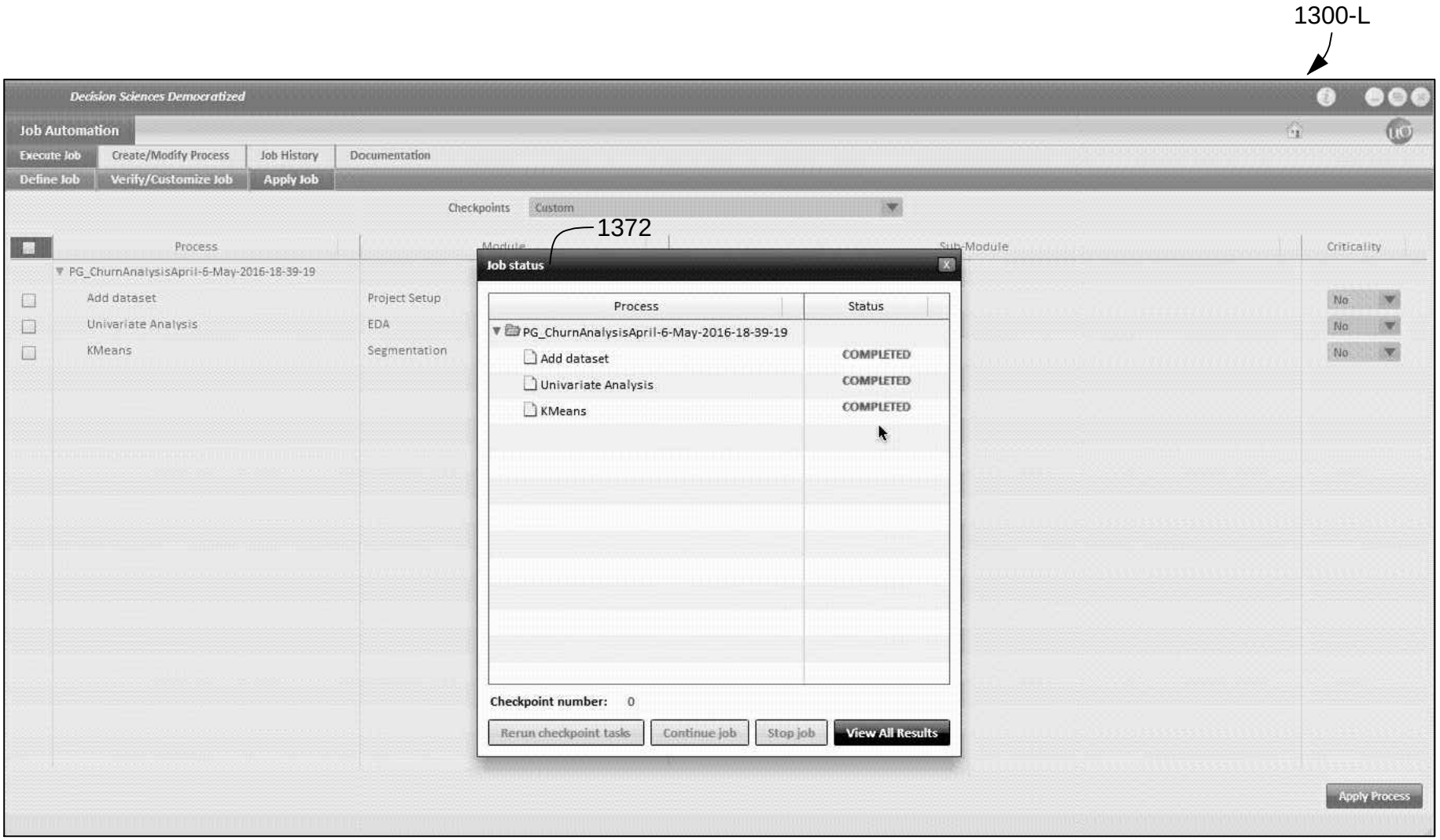

1300-L
Decision Sciences Democratized
Job Automation
Execute Job
Create/Modify Process
Job History
Documentation
Define Job
Verify/Customize Job
Apply Job
Checkpoints
Custom
1372
Process
Module
Sub-Module
Criticality
PG_ChurnAnalysisApril-6-May-2016-18-39-19
Add dataset
Project Setup
Univariate Analysis
EDA
KMeans
Segmentation
No
No
No
Job status
Process
Status
PG_ChurnAnalysisApril-6-May-2016-18-39-19
Add dataset
COMPLETED
Univariate Analysis
COMPLETED
KMeans
COMPLETED
Checkpoint number: 0
Rerun checkpoint tasks
Continue job
Stop job
View All Results
Apply Process

FIG. 13-L

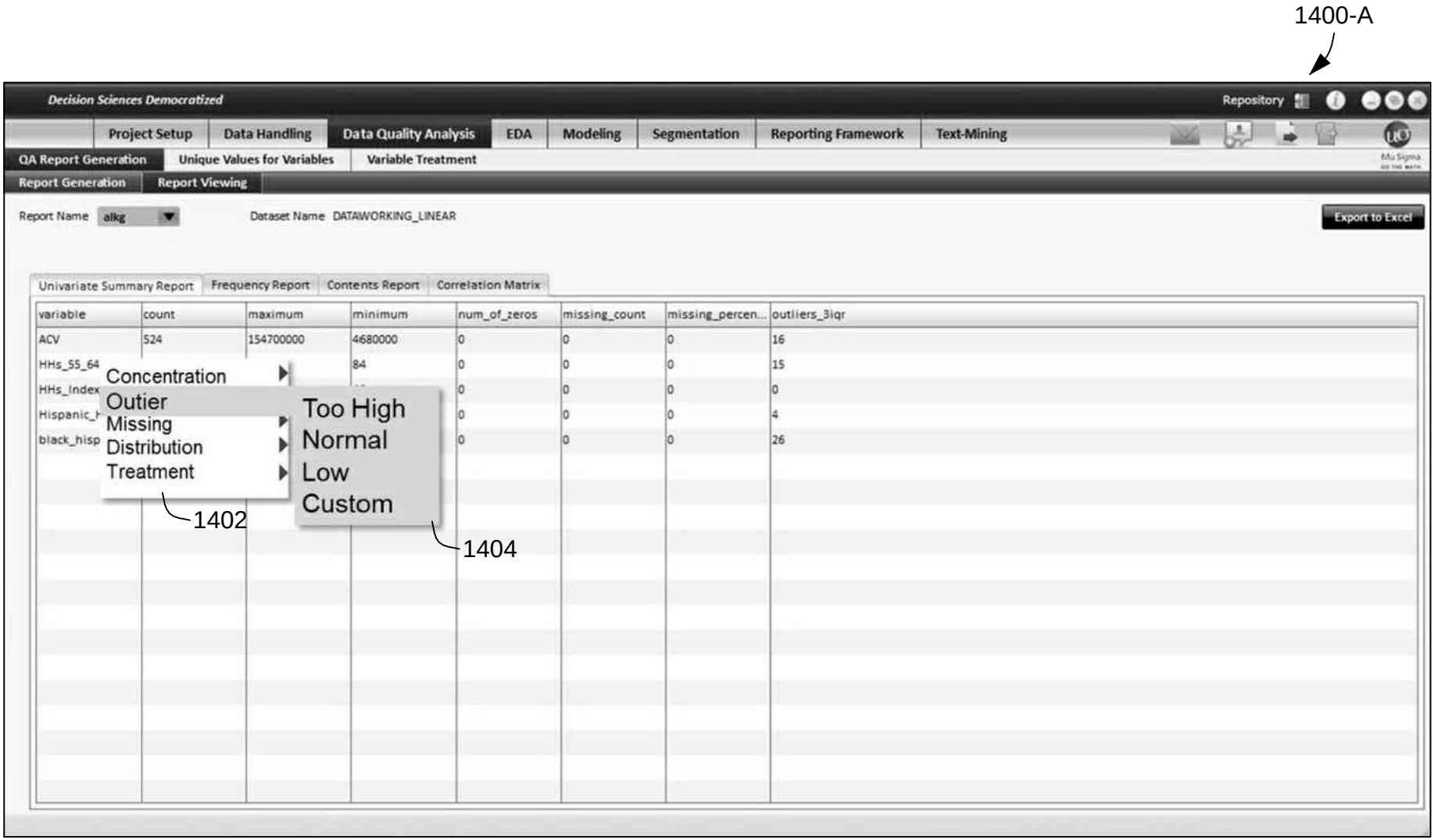

1400-A
Decision Sciences Democratized
Repository
Project Setup
Data Handling
Data Quality Analysis
EDA
Modeling
Segmentation
Reporting Framework
Text-Mining
QA Report Generation
Unique Values for Variables
Variable Treatment
Report Generation
Report Viewing
Report Name
alkg
Dataset Name DATAWORKING_LINEAR
Export to Excel
Univariate Summary Report
Frequency Report
Contents Report
Correlation Matrix
variable
count
maximum
minimum
num_of_zeros
missing_count
missing_percen...
outliers_3iqr
ACV
524
154700000
4680000
0
0
0
16
HHs_55_64
84
0
0
0
15
HHs_Index
0
0
0
0
Hispanic_H
0
0
0
4
black_hisp
0
0
0
26
Concentration
Outier
Missing
Distribution
Treatment
Too High
Normal
Low
Custom
1402
1404

FIG. 14-A


1400-B
Decision Sciences Democratized
Repository
Project Setup
Data Handling
Data Quality Analysis
EDA
Modeling
Segmentation
Reporting Framework
Text-Mining
Linear Regression
Logistic Regression
General Linear Model
Generalized Linear Model
Mixed
Multinomial Logisitic Regression
ARIMAX
Auto Regression
Factor Analysis
Checklist
Model Definition
Model Building
Model Diagnostics
Model History
Filter : sdaef
Compress Dataset
Dataset
DATAWORKING_LINEAR
Treatment Process
Across Dataset
Dependent Variable
ACV
Iteration
1
1406
1412-A
1412-B
1408
Properties
Behavior
Relations
Custom
1412-C
1412-D
1406
1410
P Mean close to max
P Normally distributed
B High correlation with sales
B Varies across storeformat
1412-E
1412-F
High Impact
Low impact
Comments:
Variables Removed:
Variables Added:
Base Iteration:1
Validation Scenario: nil
Filtering Scenario
Previous Iteration
ITERATION
Next Iteration
Parameter Estimates
Model Variable Creation
Select...
Variable
VIF
Missin...
Missing ...
Intercept
black_hispanic
HHs_55_64
HHs_Index_Income_75K_9999K
Hispanic_HHs_Index
ACV
Chiller_flag
P164_Demand_Index
P26_Demand_Index
Total_Selling_Area
channel_1
channel_2
channel_3
Model Stats
Model Statistics
Statistic
Value
No of variables 5
No of observations 524
Mape 9.923757%
R Square 0.96160472
Adjusted R Square 0.9613
AIC (Akaike Information Criter 15905.5782
Dependent Mean 29695573
Durbin Watson Statistic 1.979
P Value model 0
Heteroskedastic P Value 1
Options
Algorithmic Options
Validation Scenario
Outlier Scenario
Run Manual Regression
Generate Model Report
Validate Model

FIG. 14-B

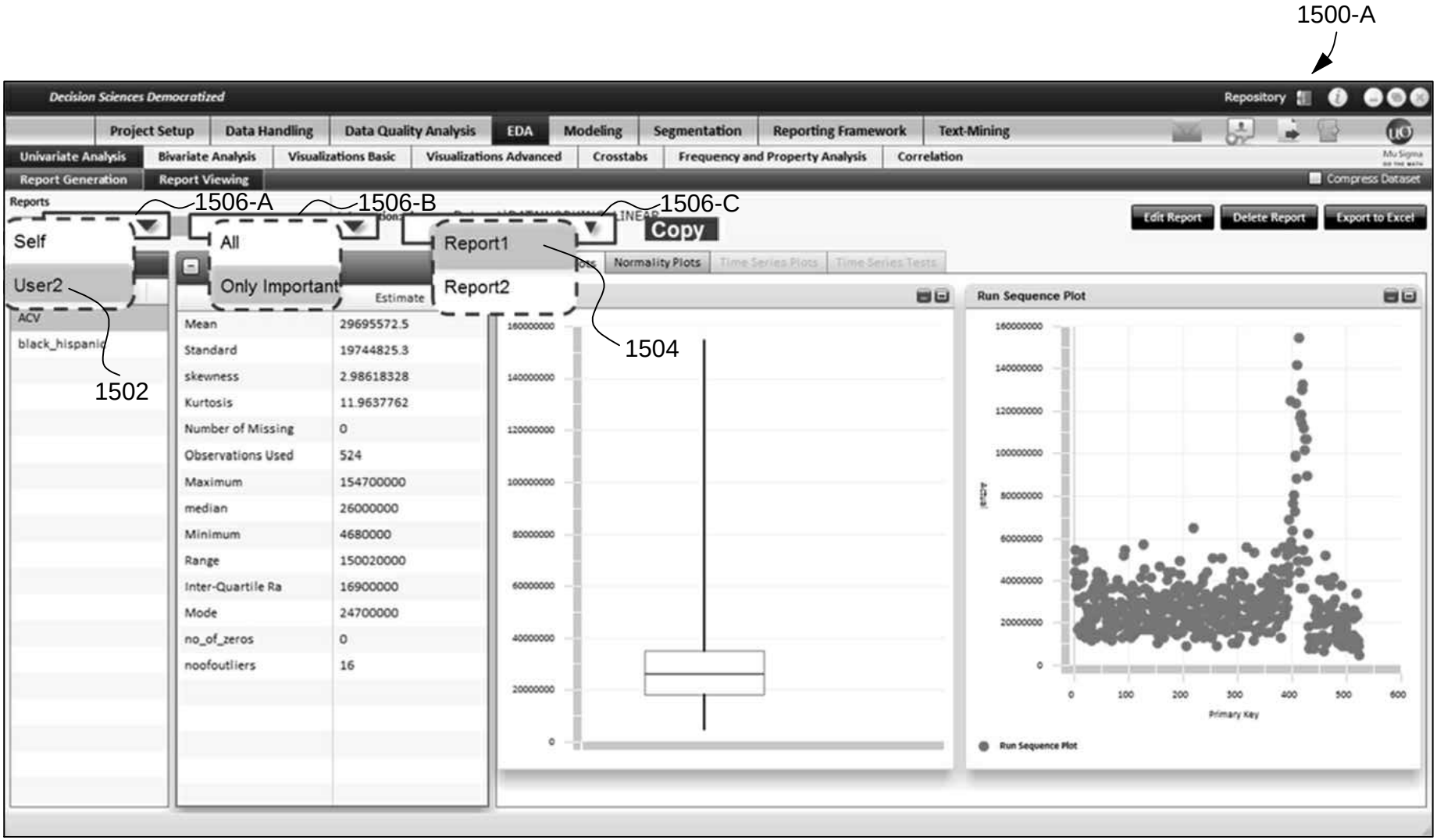

1500-A
Decision Sciences Democratized
Repository
Project Setup
Data Handling
Data Quality Analysis
EDA
Modeling
Segmentation
Reporting Framework
Text-Mining
Univariate Analysis
Bivariate Analysis
Visualizations Basic
Visualizations Advanced
Crosstabs
Frequency and Property Analysis
Correlation
Report Generation
Report Viewing
Compress Dataset
Reports
1506-A
1506-B
1506-C
Self
User2
All
Only Important
Report1
Report2
Copy
Edit Report
Delete Report
Export to Excel
Normality Plots
Time Series Plots
Time Series Tests
Estimate
ACV
black_hispanic
1502
1504
Mean 29695572.5
Standard 19744825.3
skewness 2.98618328
Kurtosis 11.9637762
Number of Missing 0
Observations Used 524
Maximum 154700000
median 26000000
Minimum 4680000
Range 150020000
Inter-Quartile Ra 16900000
Mode 24700000
no_of_zeros 0
noofoutliers 16
Run Sequence Plot
Primary Key
Run Sequence Plot

FIG. 15-A

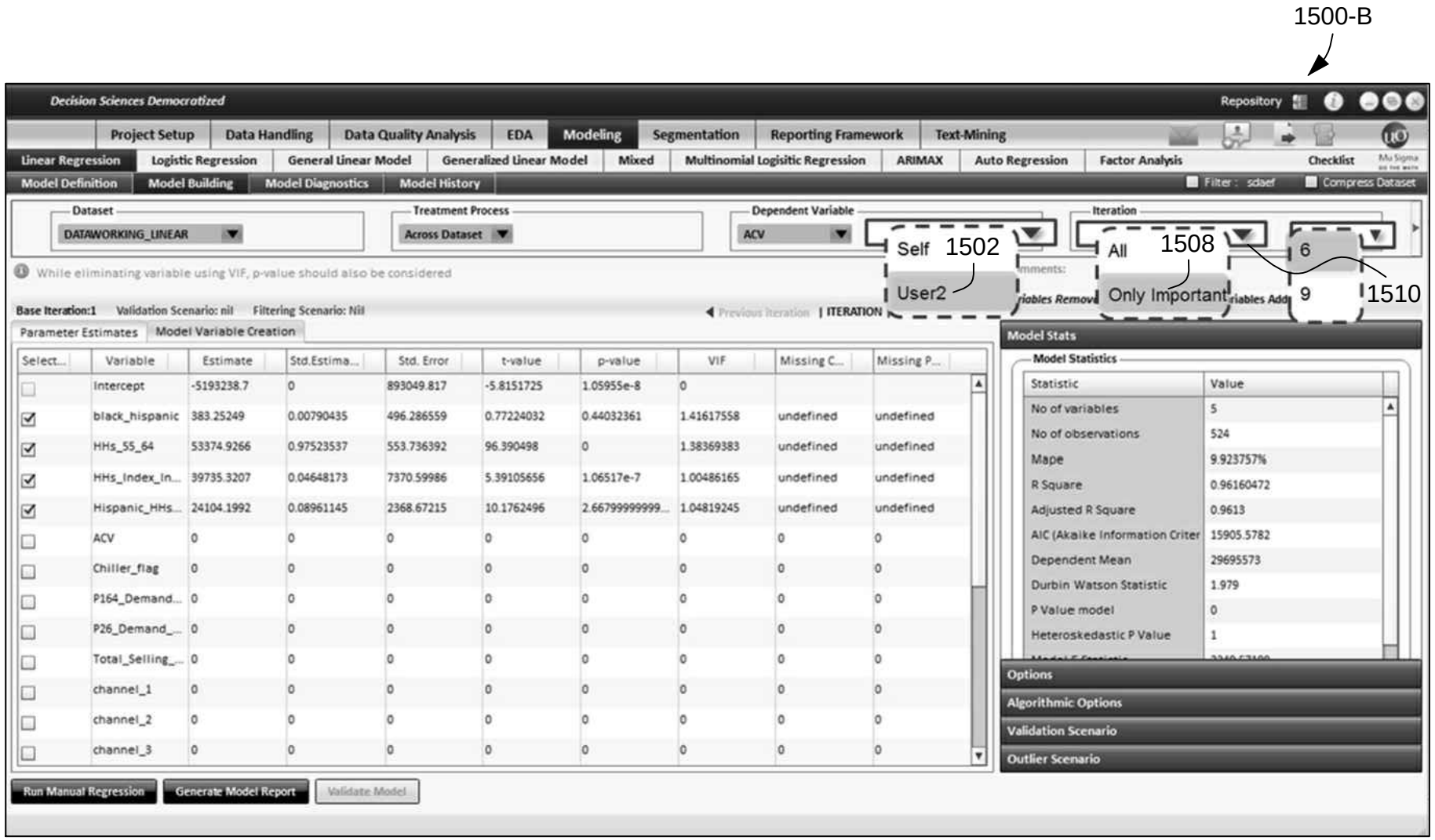
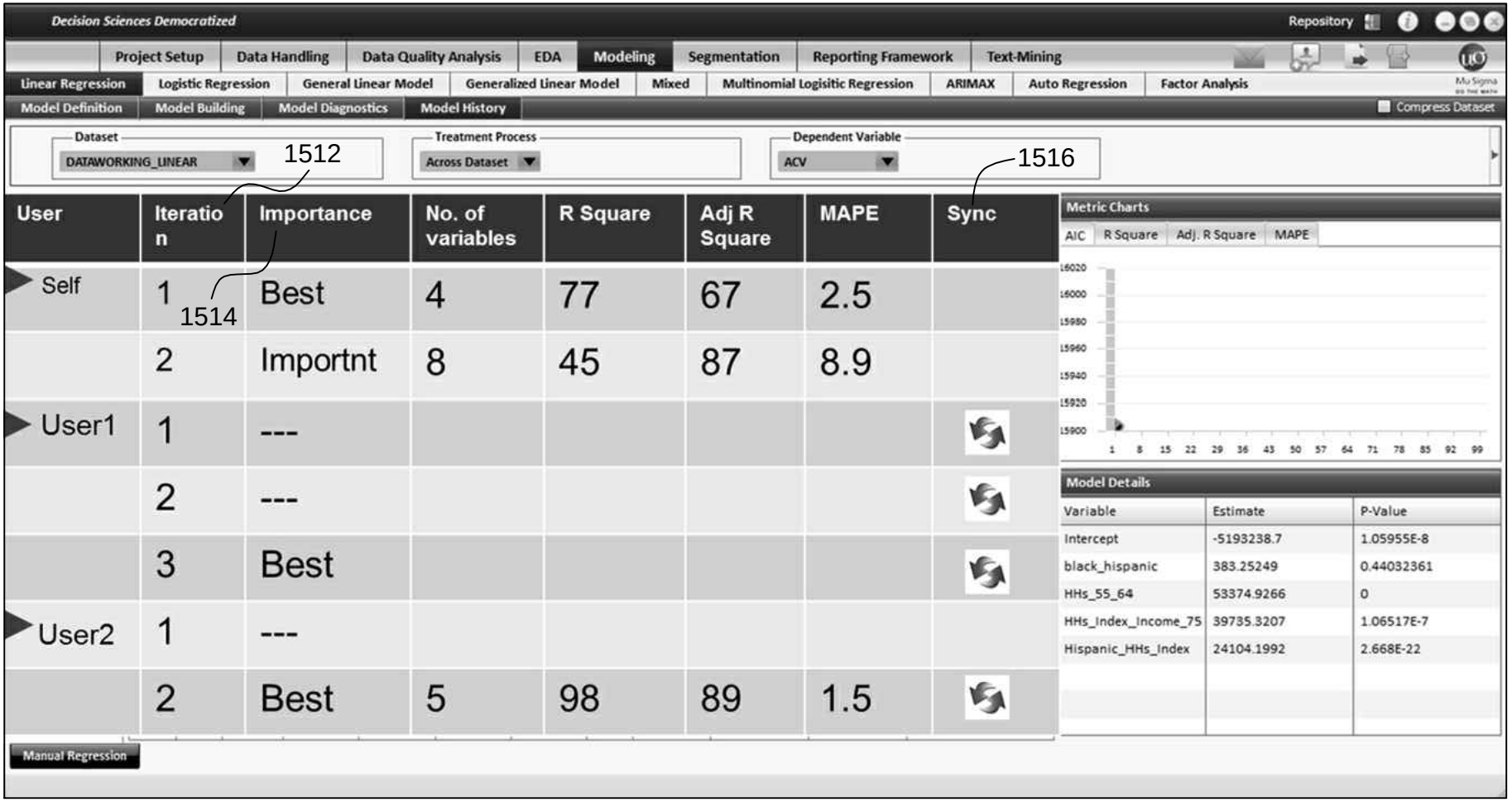

1500-B
Decision Sciences Democratized
Repository
Project Setup
Data Handling
Data Quality Analysis
EDA
Modeling
Segmentation
Reporting Framework
Text-Mining
Linear Regression
Logistic Regression
General Linear Model
Generalized Linear Model
Mixed
Multinomial Logisitic Regression
ARIMAX
Auto Regression
Factor Analysis
Checklist
Model Definition
Model Building
Model Diagnostics
Model History
Filter : sdaef
Compress Dataset
Dataset
DATAWORKING_LINEAR
Treatment Process
Across Dataset
Dependent Variable
ACV
Iteration
Self
1502
User2
All
1508
Only Important
6
9
1510
While eliminating variable using VIF, p-value should also be considered
Base Iteration:1
Validation Scenario: nil
Filtering Scenario: Nil
Previous Iteration
ITERATION
Parameter Estimates
Model Variable Creation
Select...
Variable
Estimate
Std.Estima...
Std. Error
t-value
p-value
VIF
Missing C...
Missing P...
Intercept -5193238.7 0 893049.817 -5.8151725 1.05955e-8 0
black_hispanic 383.25249 0.00790435 496.286559 0.77224032 0.44032361 1.41617558 undefined undefined
HHs_55_64 53374.9266 0.97523537 553.736392 96.390498 0 1.38369383 undefined undefined
HHs_Index_In... 39735.3207 0.04648173 7370.59986 5.39105656 1.06517e-7 1.00486165 undefined undefined
Hispanic_HHs... 24104.1992 0.08961145 2368.67215 10.1762496 2.66799999999... 1.04819245 undefined undefined
ACV 0 0 0 0 0 0 0 0
Chiller_flag 0 0 0 0 0 0 0 0
P164_Demand... 0 0 0 0 0 0 0 0
P26_Demand_... 0 0 0 0 0 0 0 0
Total_Selling_... 0 0 0 0 0 0 0 0
channel_1 0 0 0 0 0 0 0 0
channel_2 0 0 0 0 0 0 0 0
channel_3 0 0 0 0 0 0 0 0
Model Stats
Model Statistics
Statistic
Value
No of variables 5
No of observations 524
Mape 9.923757%
R Square 0.96160472
Adjusted R Square 0.9613
AIC (Akaike Information Criter 15905.5782
Dependent Mean 29695573
Durbin Watson Statistic 1.979
P Value model 0
Heteroskedastic P Value 1
Options
Algorithmic Options
Validation Scenario
Outlier Scenario
Run Manual Regression
Generate Model Report
Validate Model

FIG. 15-B


1500-C
Decision Sciences Democratized
Repository
Project Setup
Data Handling
Data Quality Analysis
EDA
Modeling
Segmentation
Reporting Framework
Text-Mining
Linear Regression
Logistic Regression
General Linear Model
Generalized Linear Model
Mixed
Multinomial Logisitic Regression
ARIMAX
Auto Regression
Factor Analysis
Model Definition
Model Building
Model Diagnostics
Model History
Compress Dataset
Dataset
DATAWORKING_LINEAR
Treatment Process
Across Dataset
Dependent Variable
ACV
1512
1516
1514
User
Iteration
Importance
No. of variables
R Square
Adj R Square
MAPE
Sync
Self
1
Best
4
77
67
2.5
2
Importnt
8
45
87
8.9
User1
1

2

3
Best
User2
1

2
Best
5
98
89
1.5
Manual Regression
Metric Charts
AIC
R Square
Adj. R Square
MAPE
16020
16000
15980
15960
15940
15920
15900
1
8
15
22
29
36
43
50
57
64
71
78
85
92
99
Model Details
Variable
Estimate
P-Value
Intercept
-5193238.7
1.05955E-8
black_hispanic
383.25249
0.44032361
HHs_55_64
53374.9266
0
HHs_Index_Income_75
39735.3207
1.06517E-7
Hispanic_HHs_Index
24104.1992
2.668E-22

FIG. 15-C

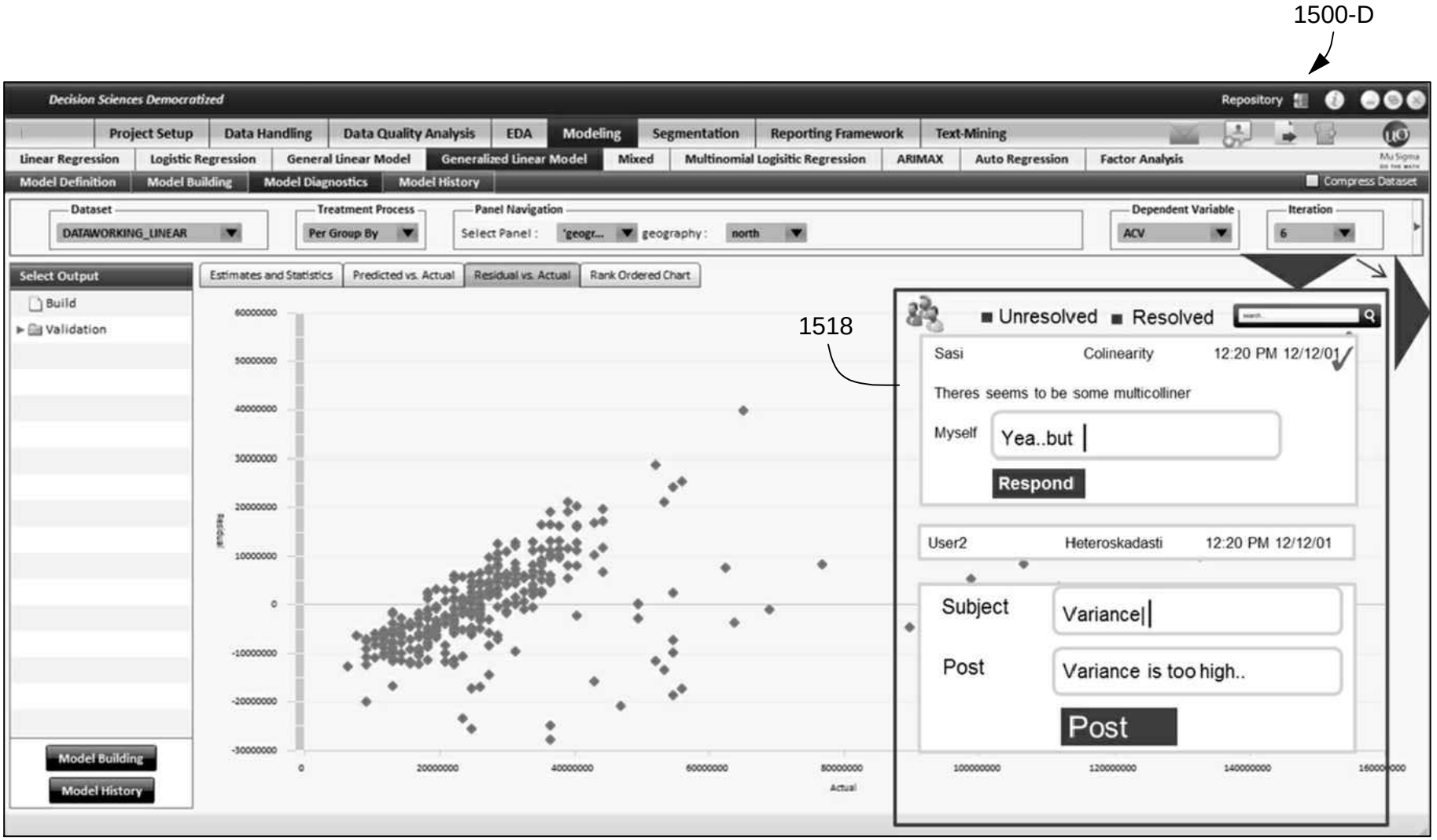

1500-D
Decision Sciences Democratized
Repository
Project Setup
Data Handling
Data Quality Analysis
EDA
Modeling
Segmentation
Reporting Framework
Text-Mining
Linear Regression
Logistic Regression
General Linear Model
Generalized Linear Model
Mixed
Multinomial Logisitic Regression
ARIMAX
Auto Regression
Factor Analysis
Model Definition
Model Building
Model Diagnostics
Model History
Compress Dataset
Dataset
DATAWORKING_LINEAR
Treatment Process
Per Group By
Panel Navigation
Select Panel :
'geogr...
geography :
north
Dependent Variable
ACV
Iteration
6
Select Output
Build
Validation
Model Building
Model History
Estimates and Statistics
Predicted vs. Actual
Residual vs. Actual
Rank Ordered Chart
1518
Unresolved
Resolved
Sasi
Colinearity
12:20 PM 12/12/01
Theres seems to be some multicolliner
Myself
Yea..but
Respond
User2
Heteroskadasti
12:20 PM 12/12/01
Subject
Variance
Post
Variance is too high..
Post
Residual
Actual

FIG. 15-D

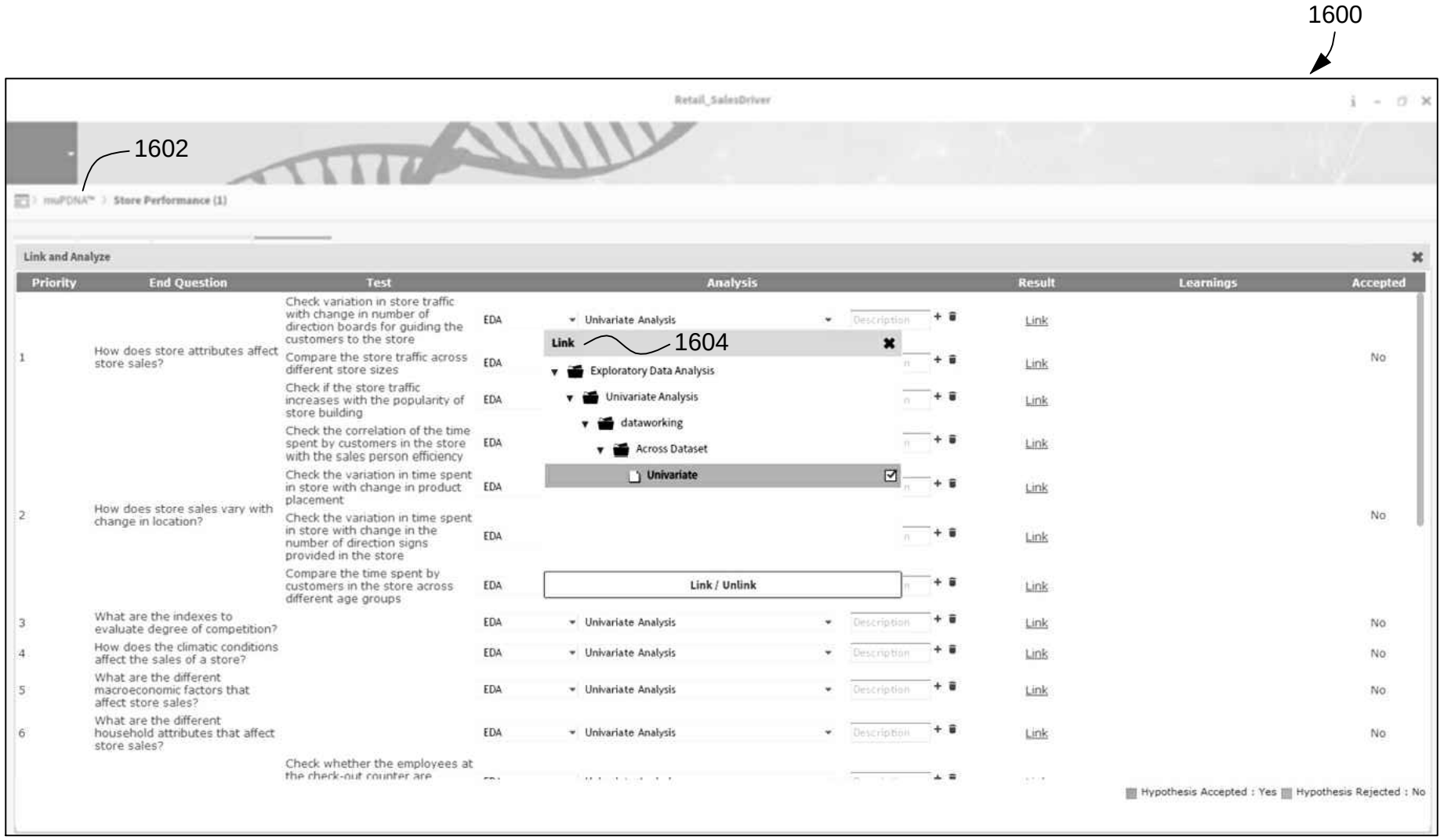

1600
Retail_SalesDriver
1602
myPDNA™ > Store Performance (1)
Link and Analyze
Priority
End Question
Test
Analysis
Result
Learnings
Accepted
1
How does store attributes affect store sales?
Check variation in store traffic with change in number of direction boards for guiding the customers to the store
Compare the store traffic across different store sizes
Check if the store traffic increases with the popularity of store building
Check the correlation of the time spent by customers in the store with the sales person efficiency
Check the variation in time spent in store with change in product placement
2
How does store sales vary with change in location?
Check the variation in time spent in store with change in the number of direction signs provided in the store
Compare the time spent by customers in the store across different age groups
3
What are the indexes to evaluate degree of competition?
4
How does the climatic conditions affect the sales of a store?
5
What are the different macroeconomic factors that affect store sales?
6
What are the different household attributes that affect store sales?
Check whether the employees at the check-out counter are
EDA
Univariate Analysis
Description
Link
No
1604
Exploratory Data Analysis
dataworking
Across Dataset
Univariate
Link / Unlink
Hypothesis Accepted : Yes
Hypothesis Rejected : No

FIG. 16

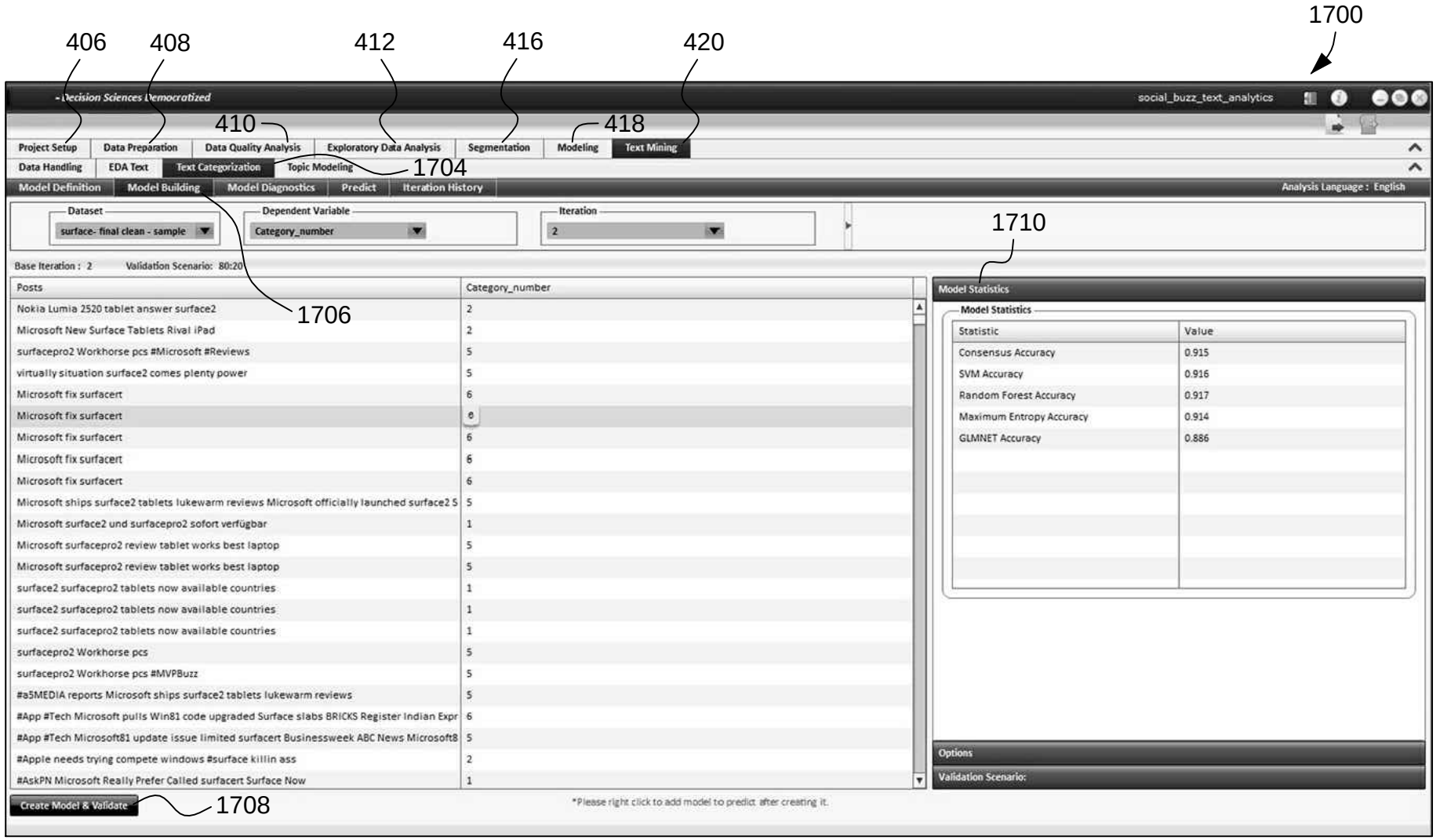

1700
406
408
412
416
420
410
418
1704
1706
1708
1710
- Decision Sciences Democratized
social_buzz_text_analytics
Project Setup
Data Preparation
Data Quality Analysis
Exploratory Data Analysis
Segmentation
Modeling
Text Mining
Data Handling
EDA Text
Text Categorization
Topic Modeling
Model Definition
Model Building
Model Diagnostics
Predict
Iteration History
Analysis Language : English
Dataset
surface- final clean - sample
Dependent Variable
Category_number
Iteration
2
Base Iteration : 2
Validation Scenario: 80:20
Posts
Category_number
Nokia Lumia 2520 tablet answer surface2 2
Microsoft New Surface Tablets Rival iPad 2
surfacepro2 Workhorse pcs #Microsoft #Reviews 5
virtually situation surface2 comes plenty power 5
Microsoft fix surfacert 6
Microsoft fix surfacert 6
Microsoft fix surfacert 6
Microsoft fix surfacert 6
Microsoft fix surfacert 6
Microsoft ships surface2 tablets lukewarm reviews Microsoft officially launched surface2 S 5
Microsoft surface2 und surfacepro2 sofort verfügbar 1
Microsoft surfacepro2 review tablet works best laptop 5
Microsoft surfacepro2 review tablet works best laptop 5
surface2 surfacepro2 tablets now available countries 1
surface2 surfacepro2 tablets now available countries 1
surface2 surfacepro2 tablets now available countries 1
surfacepro2 Workhorse pcs 5
surfacepro2 Workhorse pcs #MVPBuzz 5
#a5MEDIA reports Microsoft ships surface2 tablets lukewarm reviews 5
#App #Tech Microsoft pulls Win81 code upgraded Surface slabs BRICKS Register Indian Expr 6
#App #Tech Microsoft81 update issue limited surfacert Businessweek ABC News Microsoft8 5
#Apple needs trying compete windows #surface killin ass 2
#AskPN Microsoft Really Prefer Called surfacert Surface Now 1
Model Statistics
Model Statistics
Statistic
Value
Consensus Accuracy 0.915
SVM Accuracy 0.916
Random Forest Accuracy 0.917
Maximum Entropy Accuracy 0.914
GLMNET Accuracy 0.886
Options
Validation Scenario:
Create Model & Validate
*Please right click to add model to predict after creating it.

FIG. 17

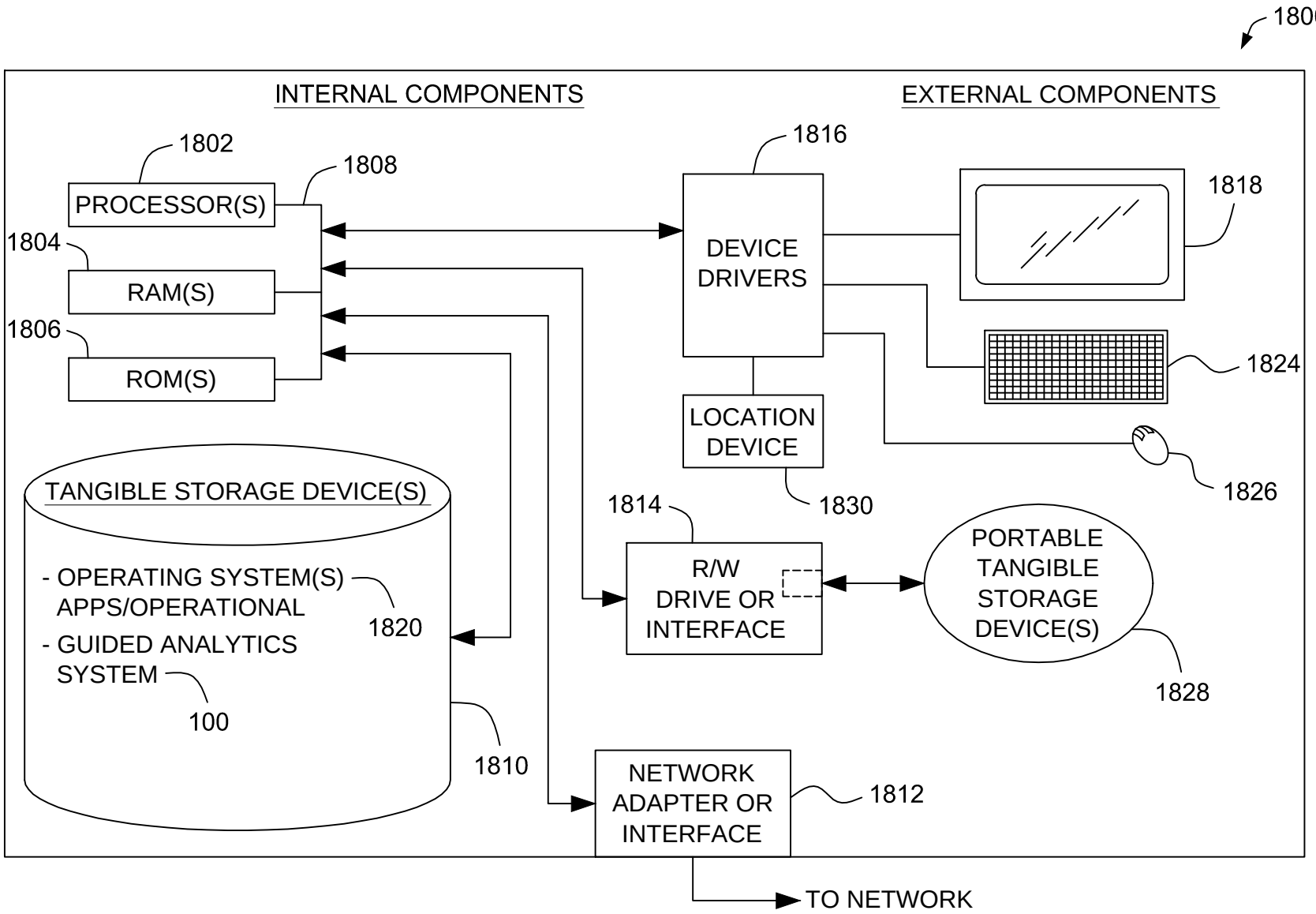

1800
INTERNAL COMPONENTS
EXTERNAL COMPONENTS
1802
1808
PROCESSOR(S)
1804
RAM(S)
1806
ROM(S)
1816
DEVICE DRIVERS
1818
1824
LOCATION DEVICE
1826
1830
TANGIBLE STORAGE DEVICE(S)
- OPERATING SYSTEM(S) APPS/OPERATIONAL
1820
- GUIDED ANALYTICS SYSTEM
100
1810
1814
R/W DRIVE OR INTERFACE
PORTABLE TANGIBLE STORAGE DEVICE(S)
1828
NETWORK ADAPTER OR INTERFACE
1812
TO NETWORK

FIG. 18