Predictive Model Results
Data Visualization with Tableau

Jen Underwood
Founder & Principal Consultant
Impact Analytix, LLC

813.435.5344
jen@impactanalytix.com
www.impactanalytix.com

About the Speaker

- Impact Analytix, LLC is a boutique BI and predictive analytics consulting firm that values projects that truly make a difference
- Jen Underwood, Founder & Principal Consultant

- ~20 years of business intelligence industry experience
- Former Global Microsoft BI and Analytics Technical Product Manager and seasoned BI implementer
- Passionate technology evangelist and volunteer, TDWI, PASS, SharePoint Conference, and Microsoft TechEd
- Bachelor of Business Administration degree
  University of Wisconsin Milwaukee

  Post Graduate Certificate Computer Science - Data Mining
  University of California, San Diego

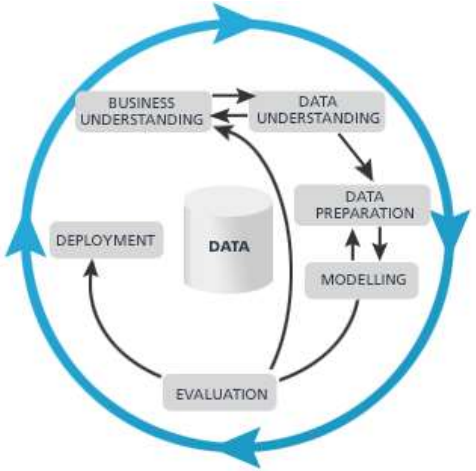
impact analytix

"Decision making and the techniques and technologies to support and automate it will be the next competitive battleground for organizations. Those who are using business rules, data mining, analytics and optimization today are the shock troops of this next wave of business innovation."

- Tom Davenport, Competing on Analytics

impact analytix

Technical Solution Approach

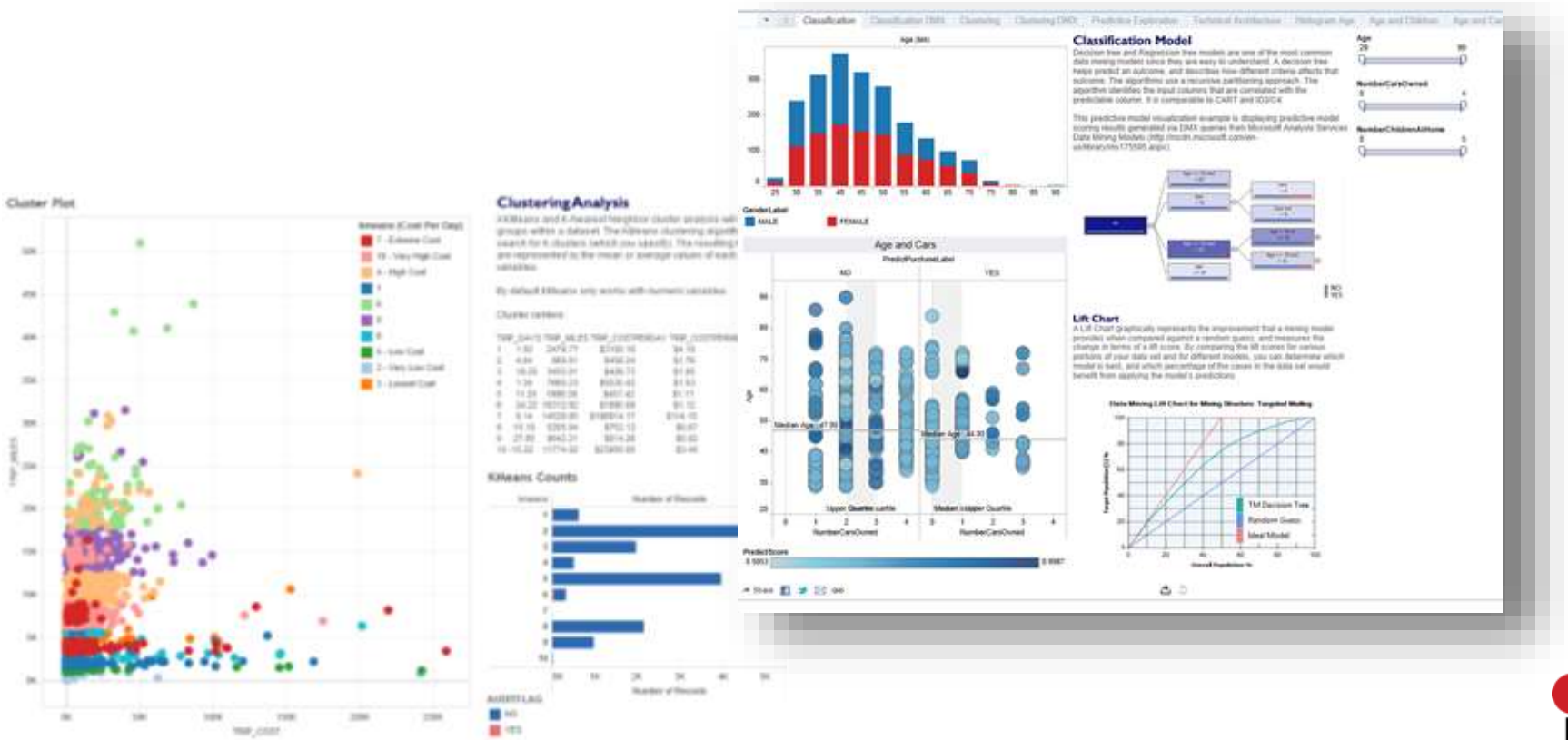


- Tableau supplements existing predictive applications
- Follow the CRISP-DM Process
- Database views or queries to prepare data
  - Flatten and prepare raw data sets, add bins/classes, and transforms in database or data mining application
- Use data mining application to create mining models
  - SAS, JMP, SPSS, R, Revolution Analytics R, Weka, Rapid Miner, Oracle Data Miner Microsoft Analysis Services, Teradata, or other programs
  - Note some databases do have embedded data mining capabilities
- Query data mining results for data visualization in Tableau
  - Export scored results back and forth via text files for use in Tableau or directly connect to databases with Tableau Custom SQL
  - Note that Tableau can call native database UDFs in Custom SQL or Raw SQL functions, pass Tableau measure values or Tableau parameters for truly dynamic, predictive result set data visualization

impact analytix

Predictive Modeling Data Visualization

- Sample predictive scored model data visualizations of Clustering and Classification Models
- A hands-on, live, online demo is posted to Tableau Public at http://public.tableausoftware.com/views/PredictiveDataVisualizationwithSSASDataMining/Classification



impact analytix

Using SAS and SAS JMP with Tableau

impact analytix

SAS and Tableau

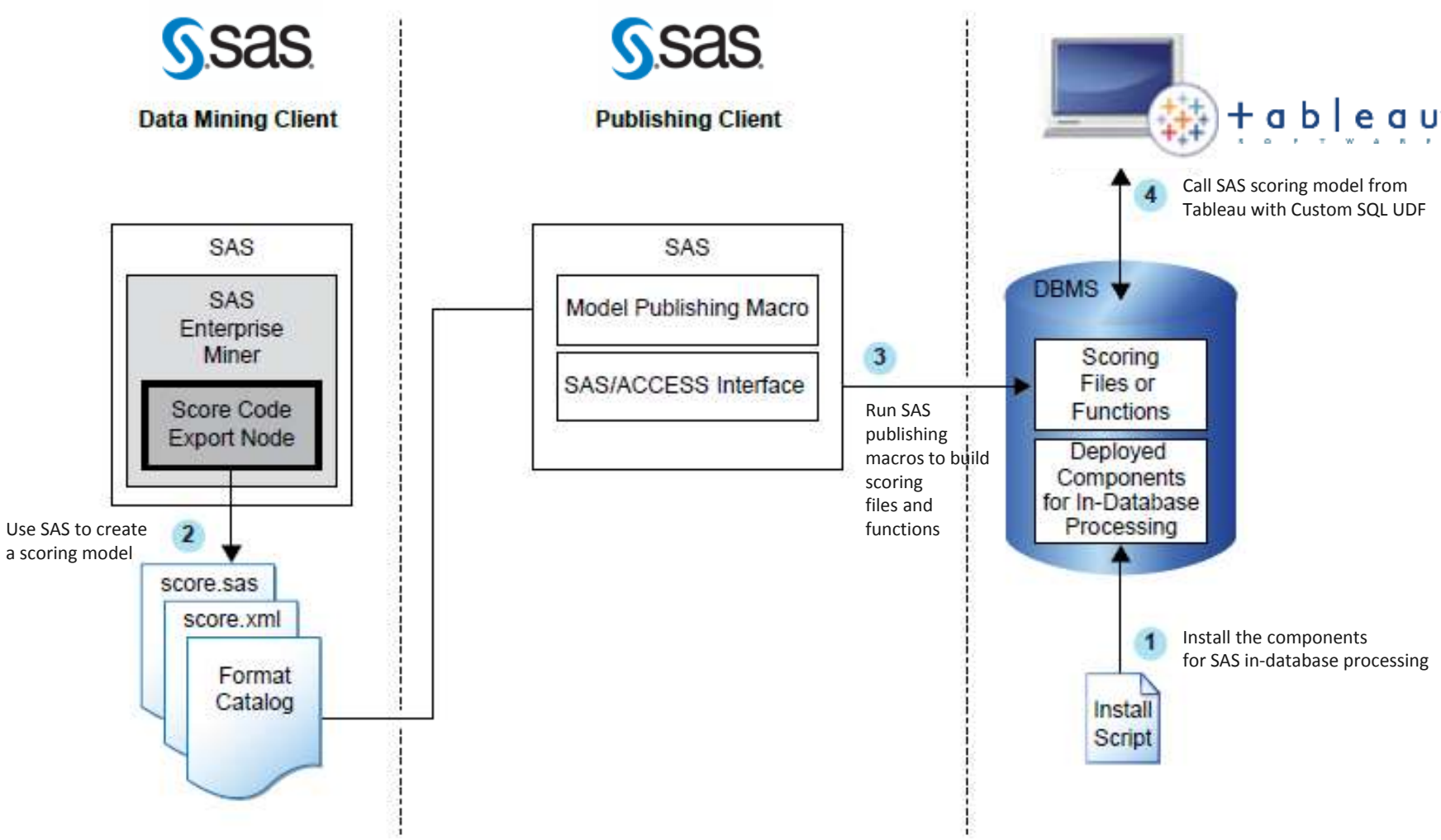
- SAS has several options for exporting data for visualization within Tableau
  - SAS export to text or database; database is optimal for exceptionally large datasets
  - SAS ODBC in Tableau http://kb.tableausoftware.com/articles/knowledgebase/tableau-and-sas-odbc
  - SAS/ACCESS for Relational Databases and SAS In-Database Processing
    - Note SAS/ACCESS engine issues SQL SELECT * to the target database
    - SAS In-Database Processing and the SAS Scoring Accelerator is more efficient in very large dataset scenarios since it does not require the transfer of data
  - Refer to SAS/ACCESS for Relational Databases Reference http://support.sas.com/documentation/cdl/en/acreldb/65247/PDF/default/acreldb.pdf
  - Refer to SAS In-Database Products User's Guide http://support.sas.com/documentation/cdl/en/indbug/65980/PDF/default/indbug.pdf

- SAS In-Database Processing has Base SAS, SAS_PUT(), and SAS Scoring within databases such as Teradata, Oracle, Aster, Netezza, Greenplum, DB2, and Hadoop via an SAS Embedded Process
  - For example, in Teradata Aster SAS_SCORE can be implemented as UDF and Tableau can call database UDFs in Custom SQL or Raw SQL functions, pass Tableau measure values or Tableau parameters for dynamic result data visualization

- SAS Embedded Process
  - SAS server process that runs within a supported database to read and write data
  - Integrates SAS solutions, SAS analytic processes, and third-party database systems

impact analytix

Technical Solution Architecture



Adapted from Source: http://support.sas.com/documentation/cdl/en/indbug/65980/PDF/default/indbug.pdf

SAS In-Database Processing

- SAS products needed for various In-Database Processing features as of SAS v9.3:

In-Database Feature	Software Required	Database Supported
Format publishing and the SAS_PUT( ) function	•Base SAS •SAS/ACCESS Interface to the DBMS	Aster, DB2 under UNIX, Greenplum, Netezza, Oracle and Teradata
Scoring Models	•Base SAS •SAS/ACCESS Interface to the DBMS •SAS Scoring Accelerator •SAS Model Manager (optional)	Aster, DB2 under UNIX, Greenplum, Netezza, Oracle and Teradata
Base SAS procedures: FREQ, RANK, REPORT, SORT SUMMARY/MEANS, TABULATE	•Base SAS •SAS/ACCESS Interface to the DBMS	Aster, DB2 under UNIX and PC Hosts, Greenplum, Oracle, Netezza and Teradata
SAS/STAT procedures: CORR, CANCORR, DMDB, DMINE, DMREG, FACTOR, PRINCOMP, REG, SCORE, TIMESERIES, VARCLUS	•Base SAS (for CORR) •SAS/ACCESS Interface to Teradata •SAS/STAT (for CANCORR, FACTOR,PRINCOMP, REG, SCORE, VARCLUS) •SAS/ETS (for TIMESERIES) •SAS Enterprise Miner (for DMDB, DMINE, DMREG) •SAS Analytics Accelerator	Teradata

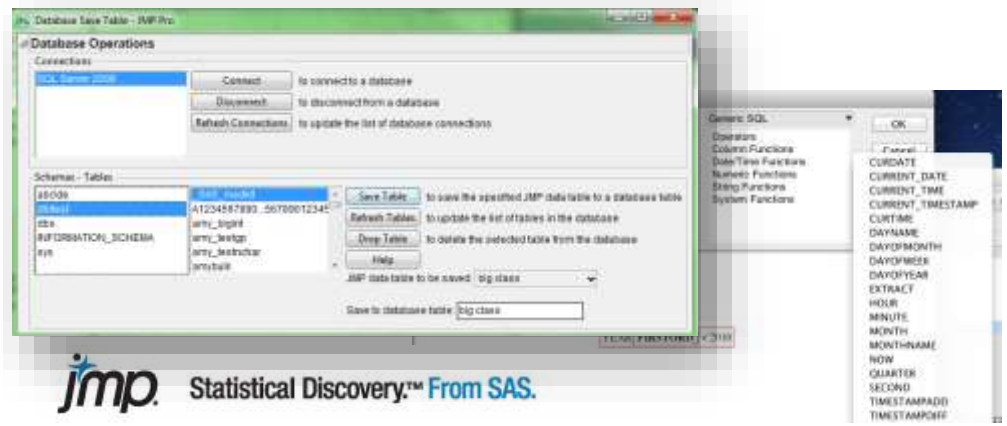
Source: http://support.sas.com/documentation/cdl/en/indbug/65980/PDF/default/indbug.pdf

impact analytix

SAS JMP and Tableau

- JMP provides several interfaces for importing and exporting data
  - SAS connection, R interface, Excel add-in, Text files, JMP scripting language JSL and ODBC
  - ODBC SQL CREATE TABLE command to save a JMP table to the target database
    (Note: this requires appropriate user create table permissions within target database)
  - Examples of databases currently used with JMP are Oracle, SQL Server, Teradata IBM DB2, Microsoft Access and MySQL
  - Excellent technical white paper by Brian Corcoran, SAS Institute
    http://jmp.com/about/events/summit2012/resources/Paper_Brian_Corcoran.pdf

Tableau can directly connect to Excel, text, files and a wide variety of databases, issue SQL statements, call database specific UDFs in Custom SQL or Raw SQL functions, pass Tableau measure values or Tableau parameters for dynamic result data visualization

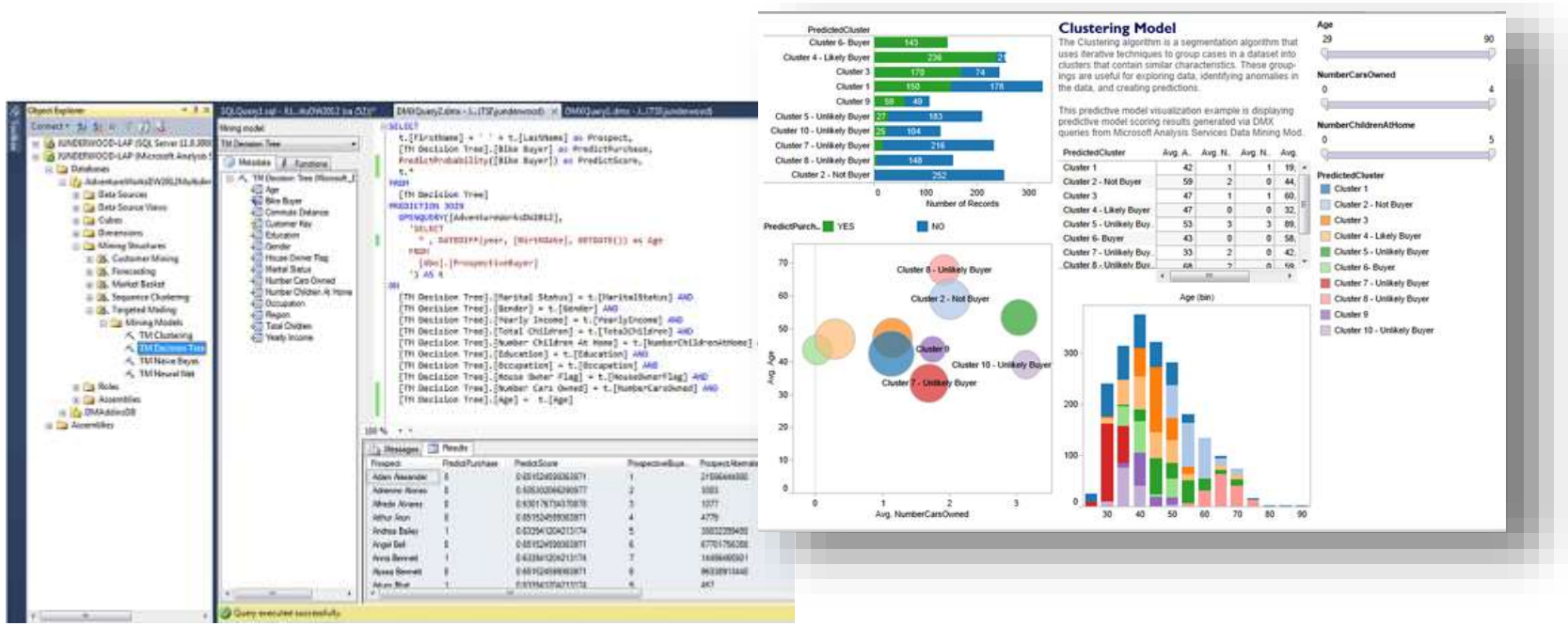


Using Microsoft SQL Server Analysis Services Prediction Queries DMX with Tableau

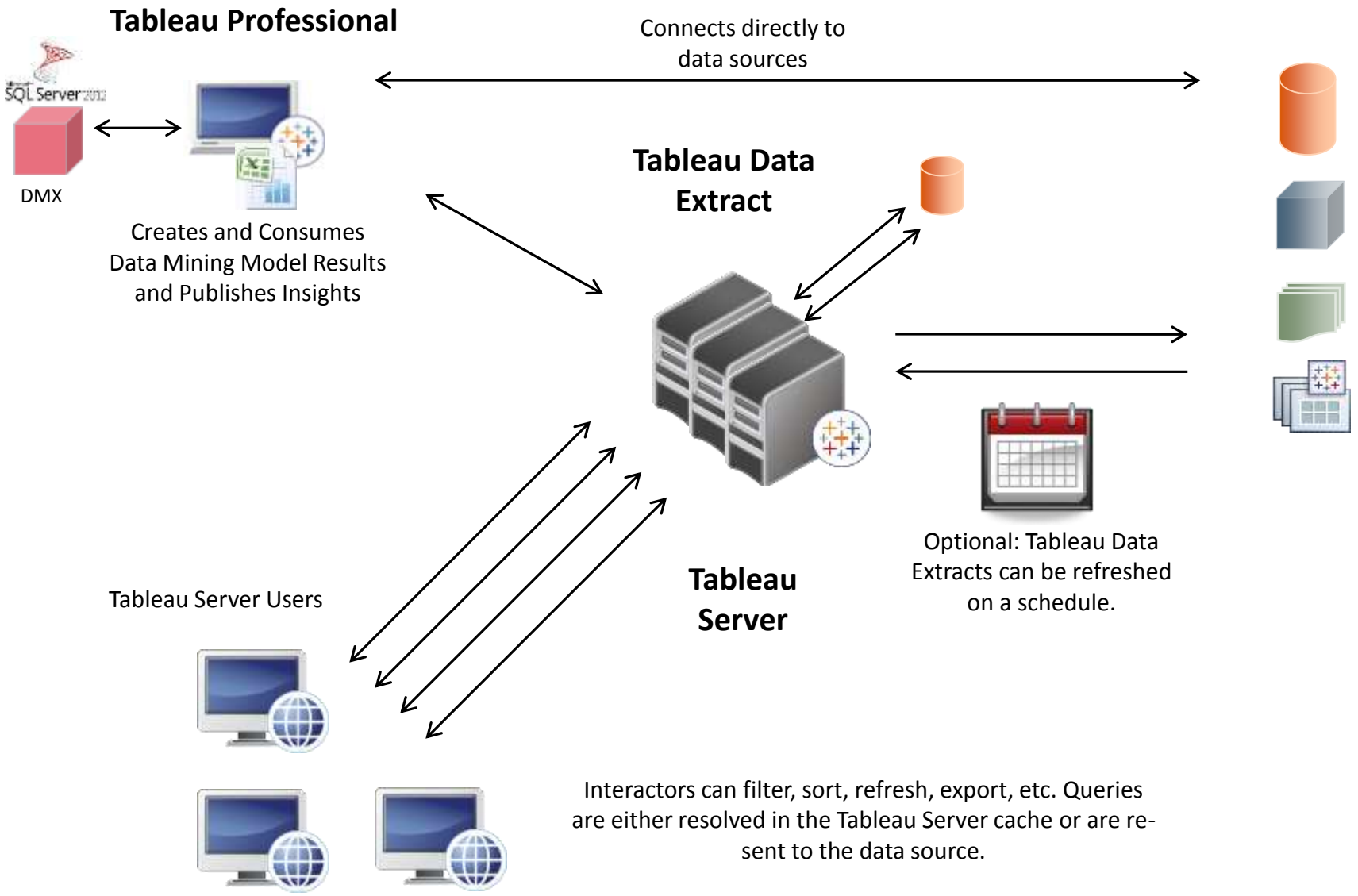
impact analytix

Analysis Services Prediction Queries (DMX)

- Analysis Services Data Mining Model Prediction Queries (DMX) can be visualized within Tableau to explore scored mining model datasets
  - Prediction DMX queried data can be saved to a text file or a database table for usage within Tableau just like any other data source type
  - See a live Tableau Public online demo with detailed solution description at http://public.tableausoftware.com/views/PredictiveDataVisualizationwithSSASDataMining/Classification
  - Refer to Analysis Service Data Mining http://msdn.microsoft.com/en-us/library/ms175595.aspx and Prediction Queries http://msdn.microsoft.com/en-us/library/hh213169.aspx



Technical Solution Architecture



Interactors can filter, sort, refresh, export, etc. Queries are either resolved in the Tableau Server cache or are re-sent to the data source.

Using R or Revolution Analytics R with Tableau

impact analytix

Using Microsoft SQL Server Analysis Services Prediction Queries DMX with Tableau

impact analytix

R, Revolution Analytics R or PL/R

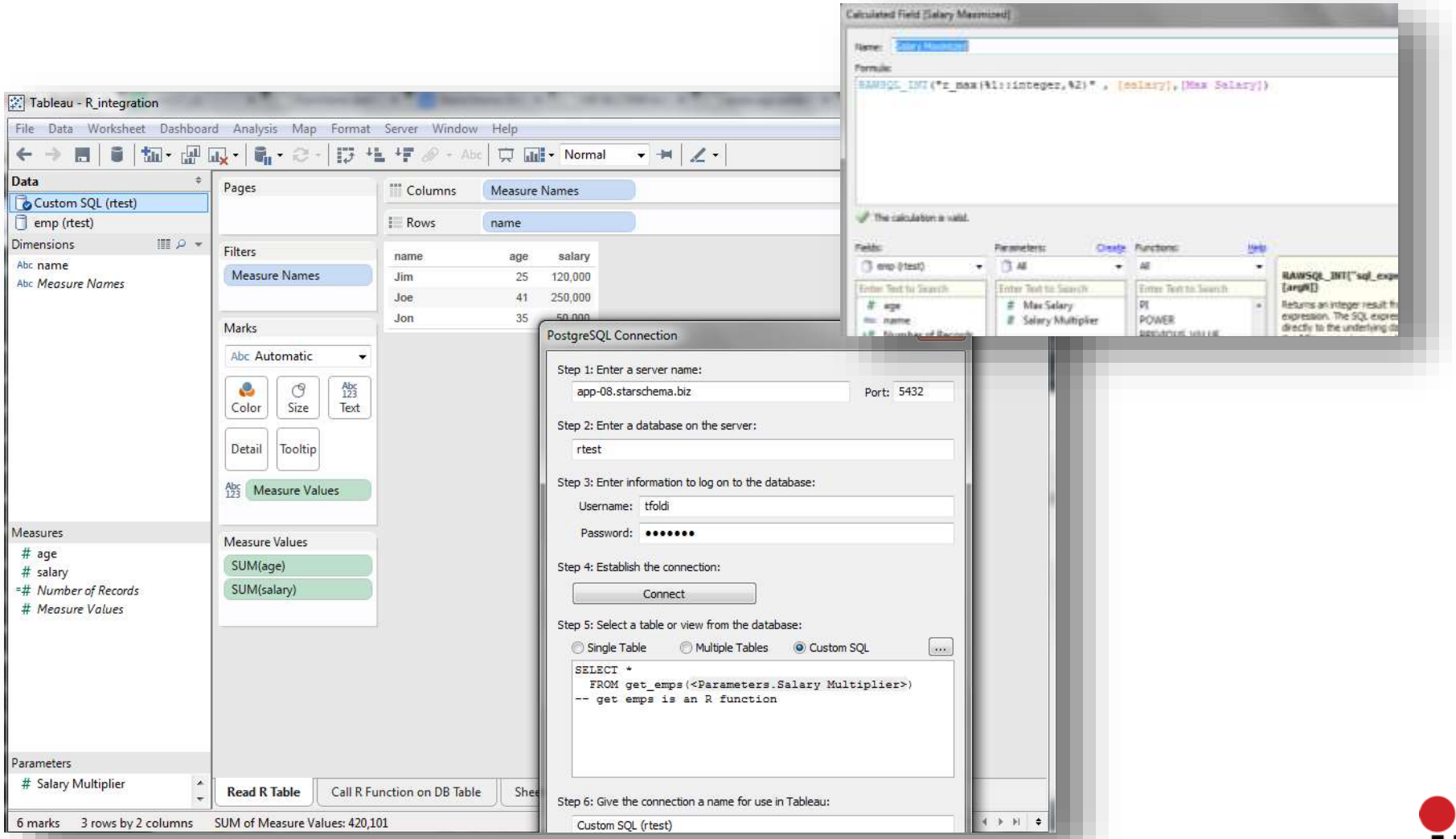
- Call R directly with Tableau Custom SQL
  - Call R functions from Tableau with databases that support PL/R functions
  - Refer to http://community.tableausoftware.com/message/206708#206708
  - Note R package “foreign v0.8+” can read and write data stored by Minitab, S, SAS, SPSS, Stata, and Systat http://cran.r-project.org/web/packages/foreign/index.html

- Use R, RODBC and Rattle
  - Load data with .csv or use RODBC
  - Explore data set, identify predictive influencers, evaluate various data mining models, further transform and experiment with variables
  - Run data set Score Report to generate scored output for use in Tableau

- Note that Revolution Analytics R version is more robust and highly scalable for scenarios that exceed open source R limitations

impact analytix

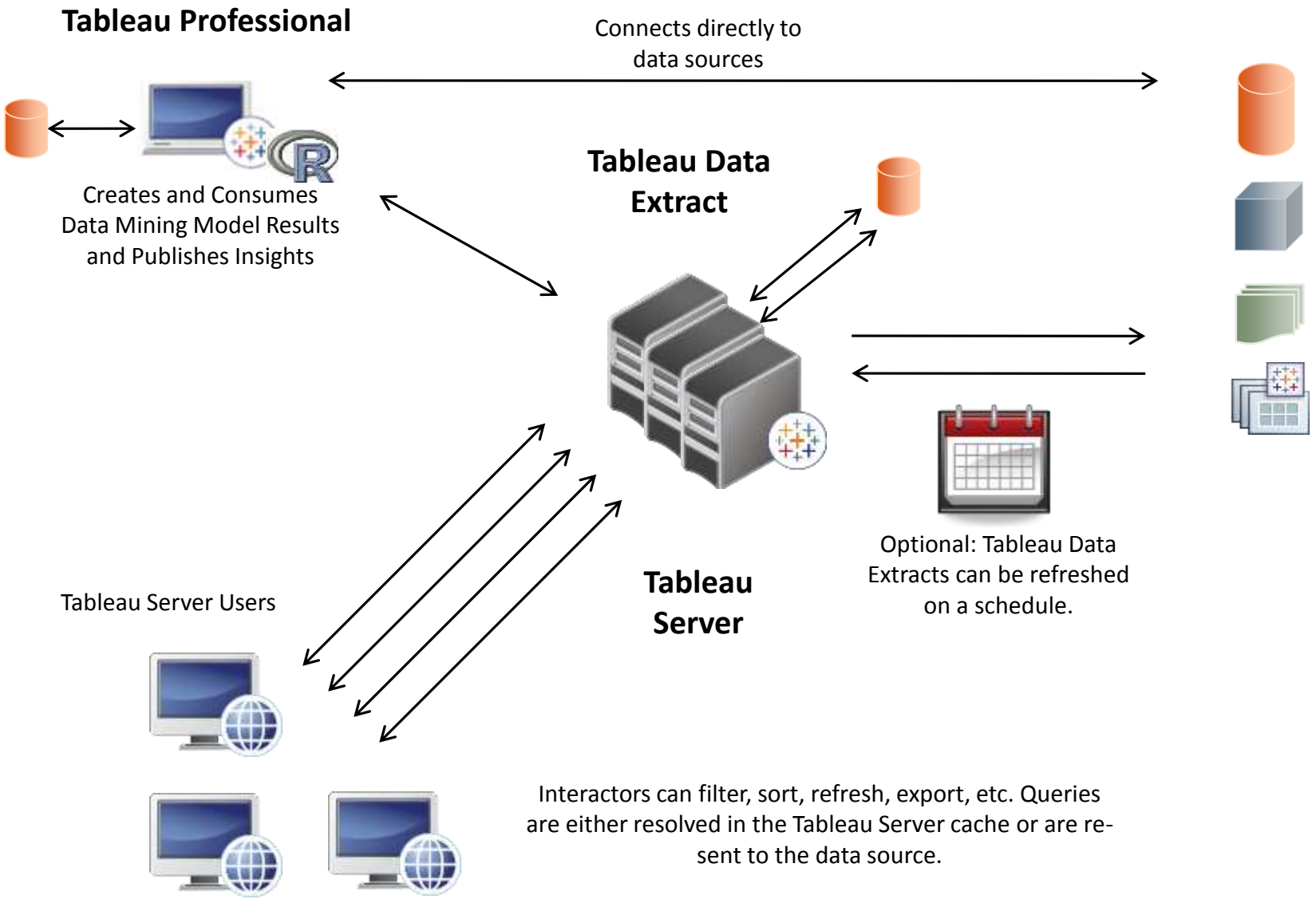
Use R directly with Tableau Custom SQL



Source: http://community.tableausoftware.com/message/206708#206708

impact analytix

Technical Solution Architecture



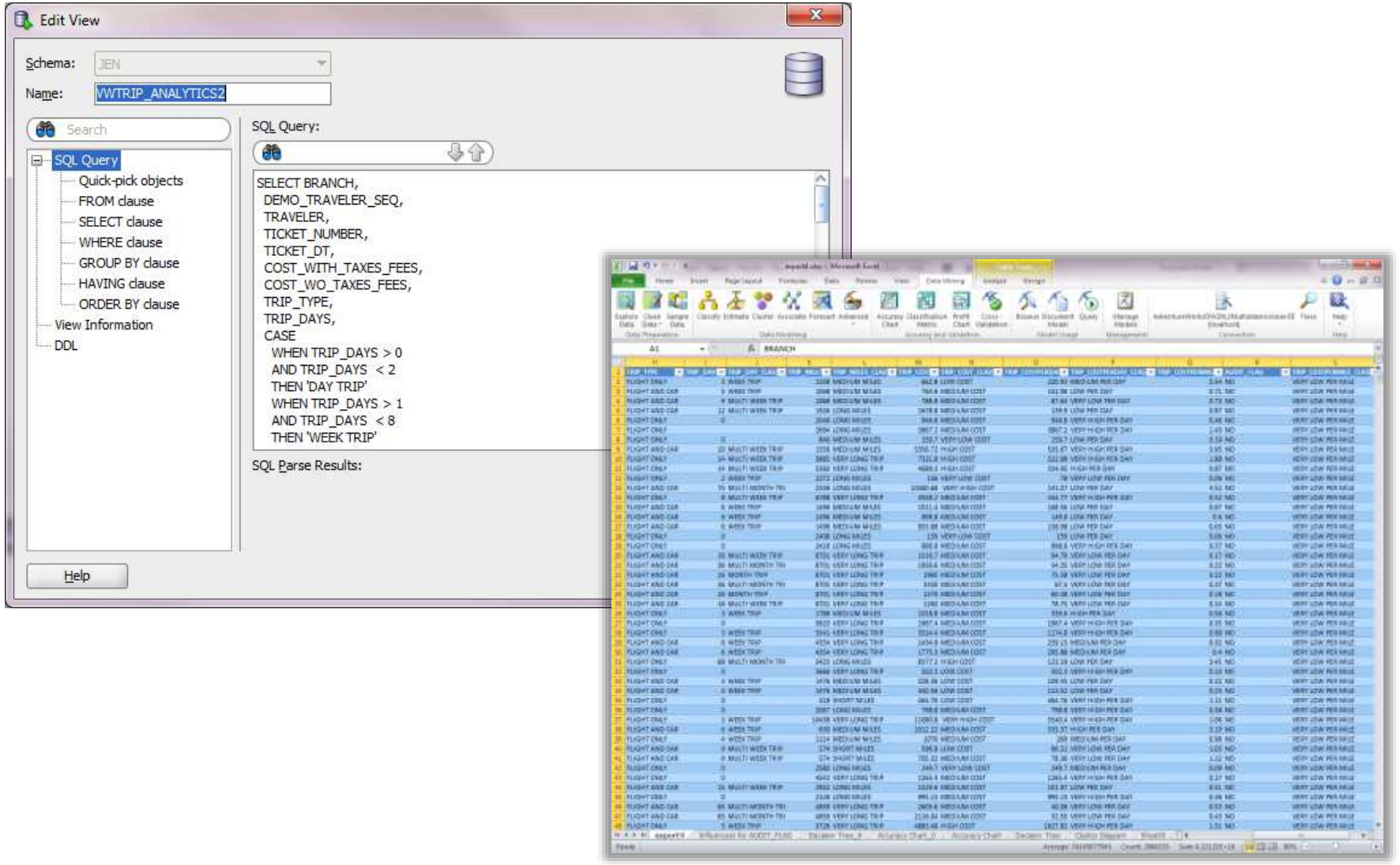
Interactors can filter, sort, refresh, export, etc. Queries are either resolved in the Tableau Server cache or are re-sent to the data source.

impact analytix

Using R with Tableau
A Step-by-Step Example

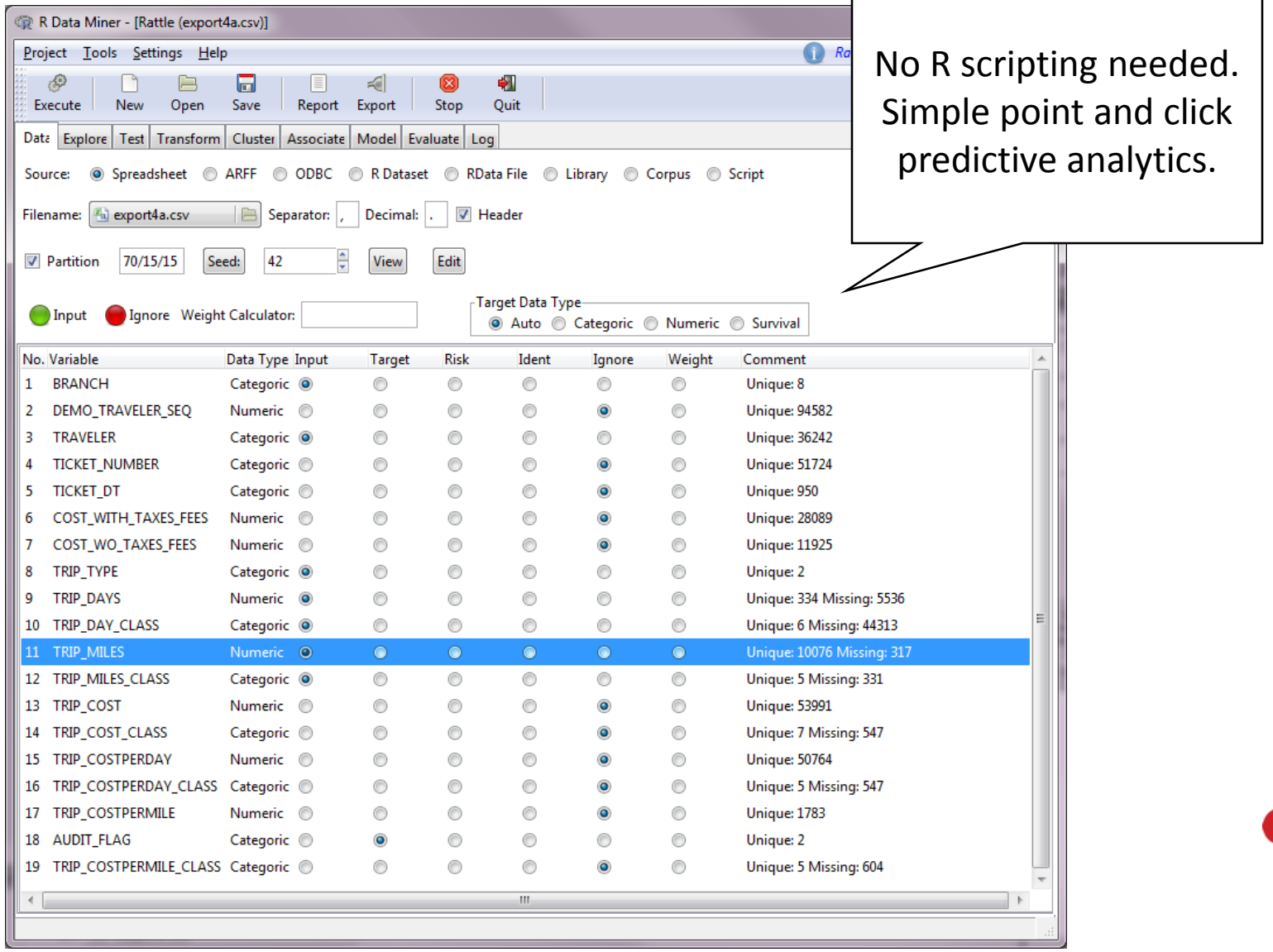
impact analytix

Preparing Data



impact analytix

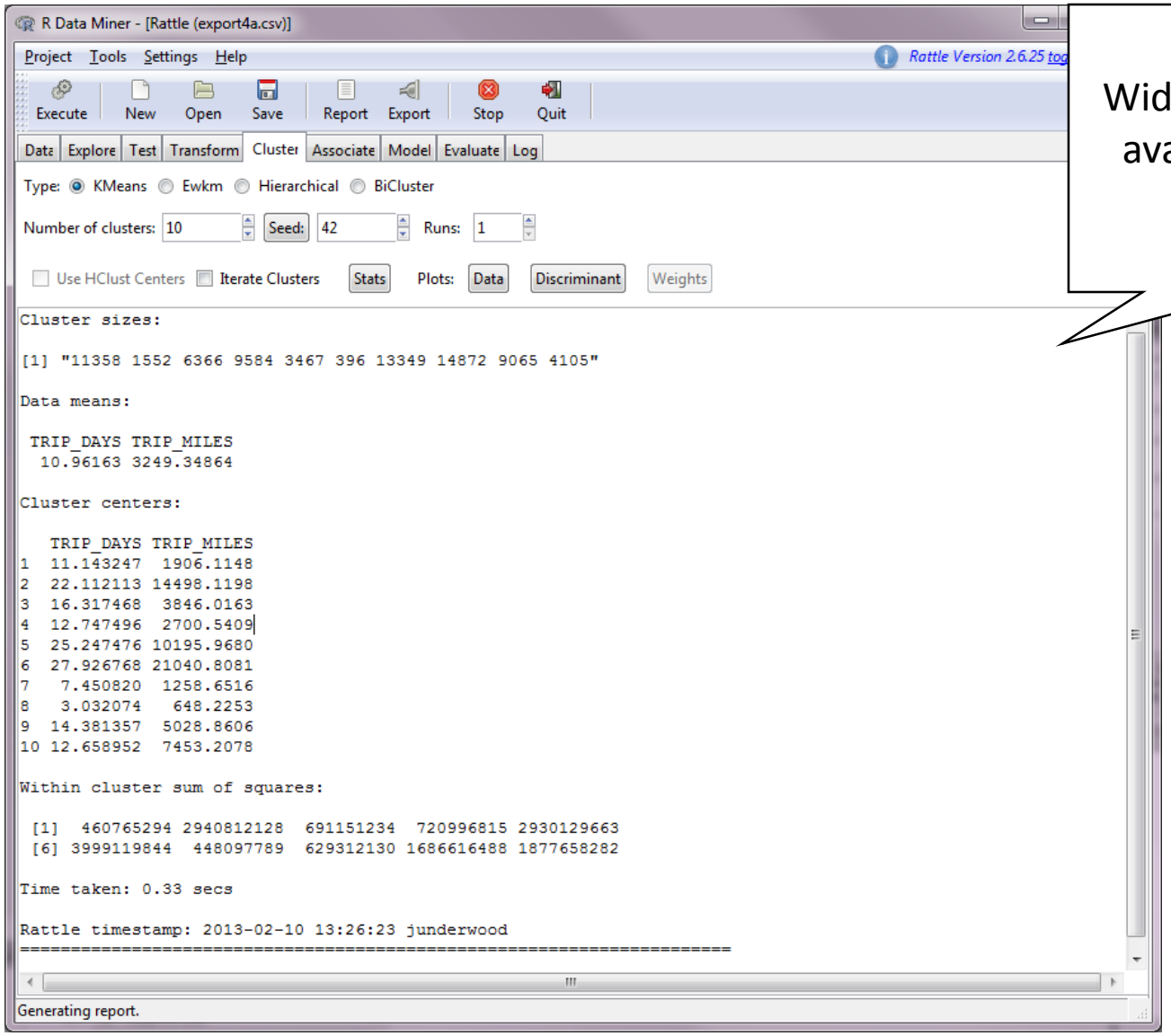
Loading data in R



No R scripting needed. Simple point and click predictive analytics.

impact analytix

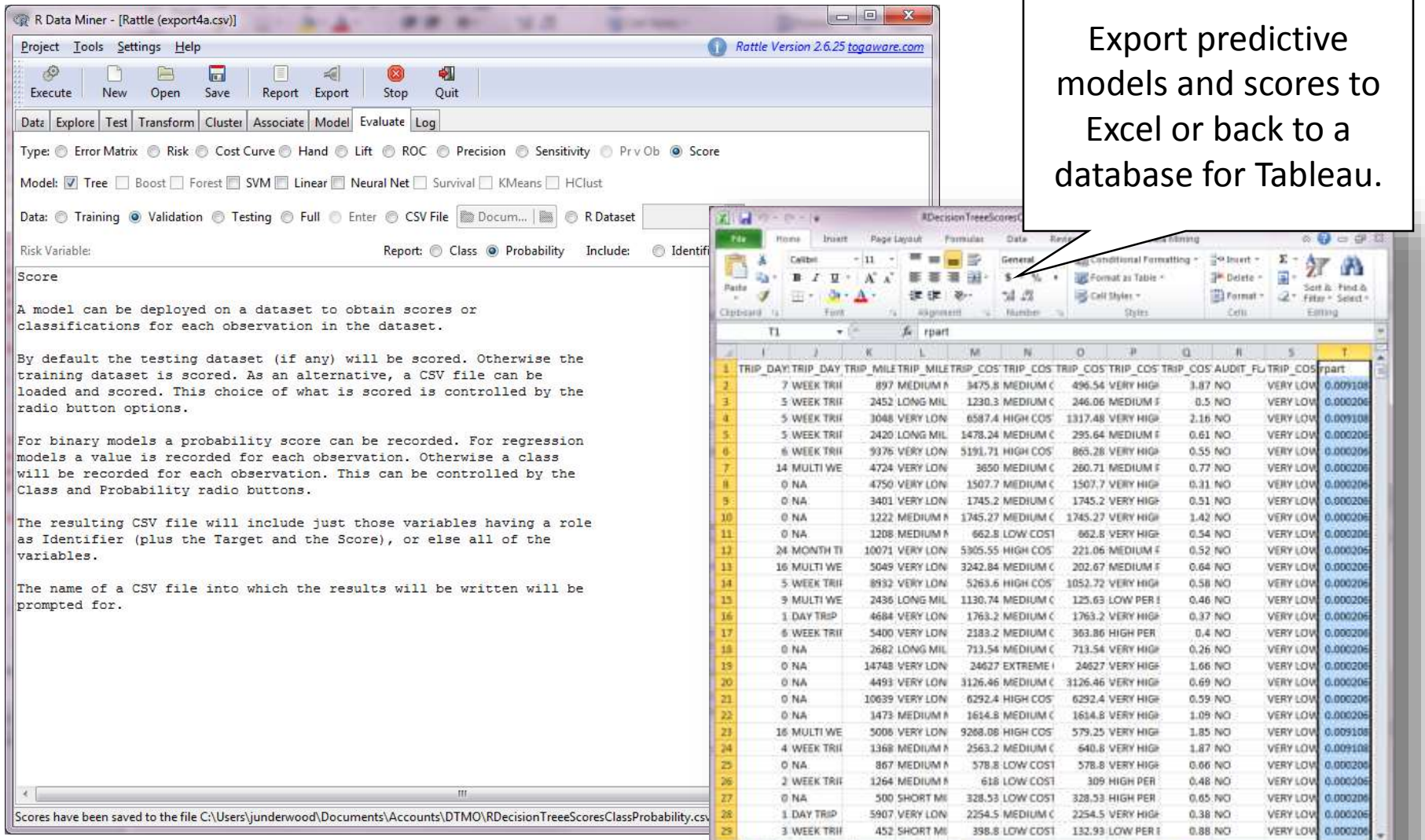
Running Data Mining Models in R



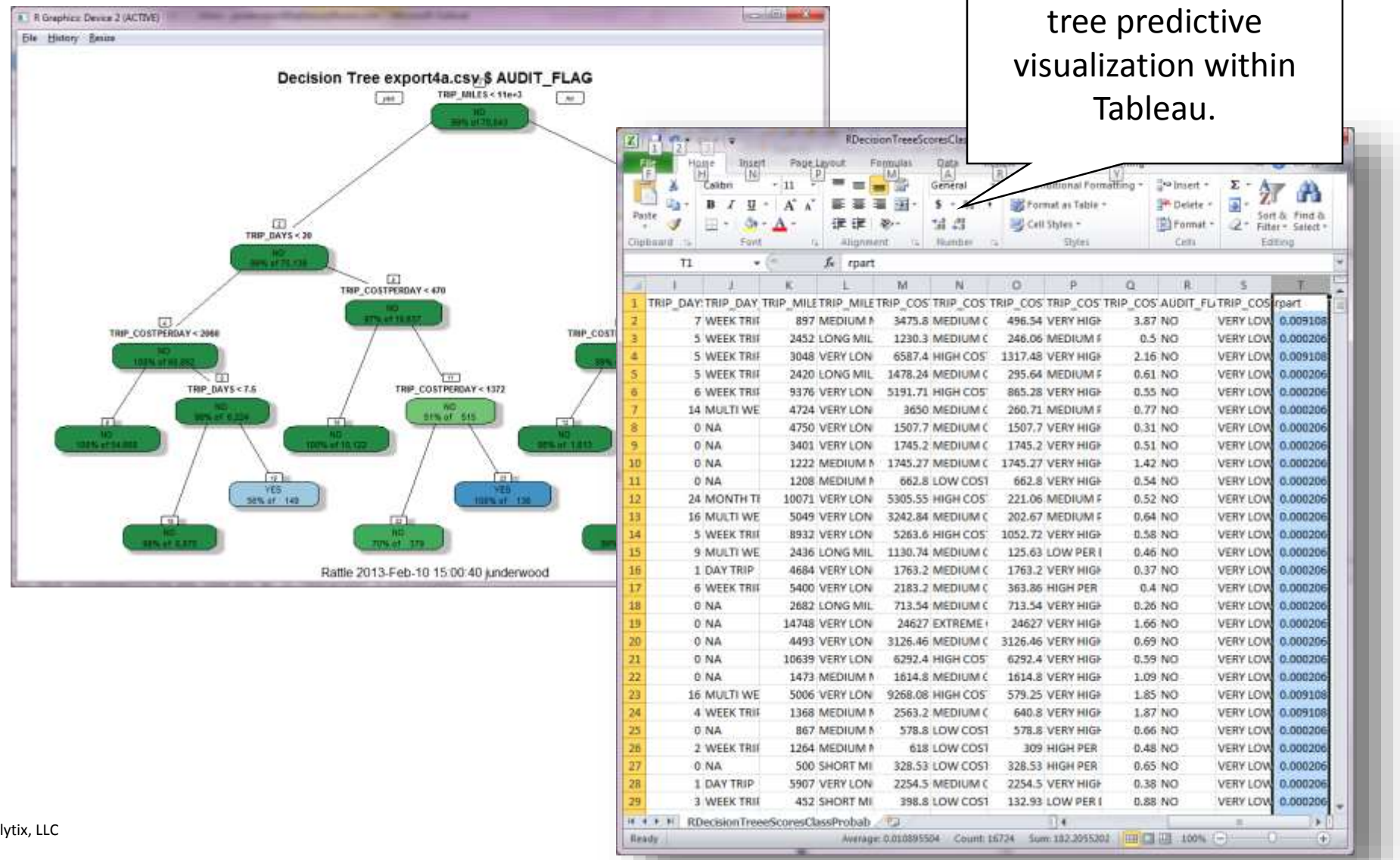
Wide variety of freely available predictive algorithms.

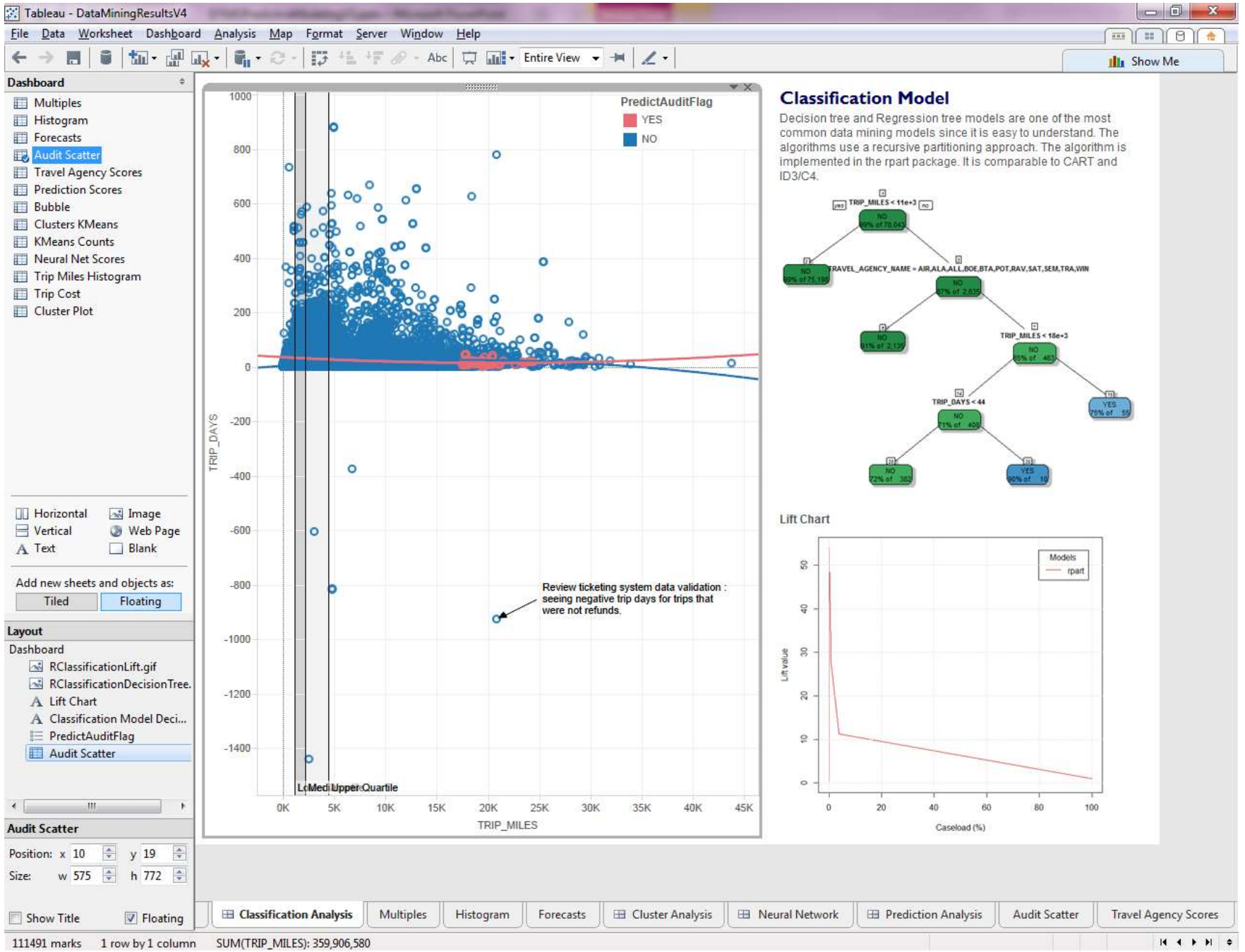
impact analytix

Generating Model Prediction Scores



Classification Decision Tree

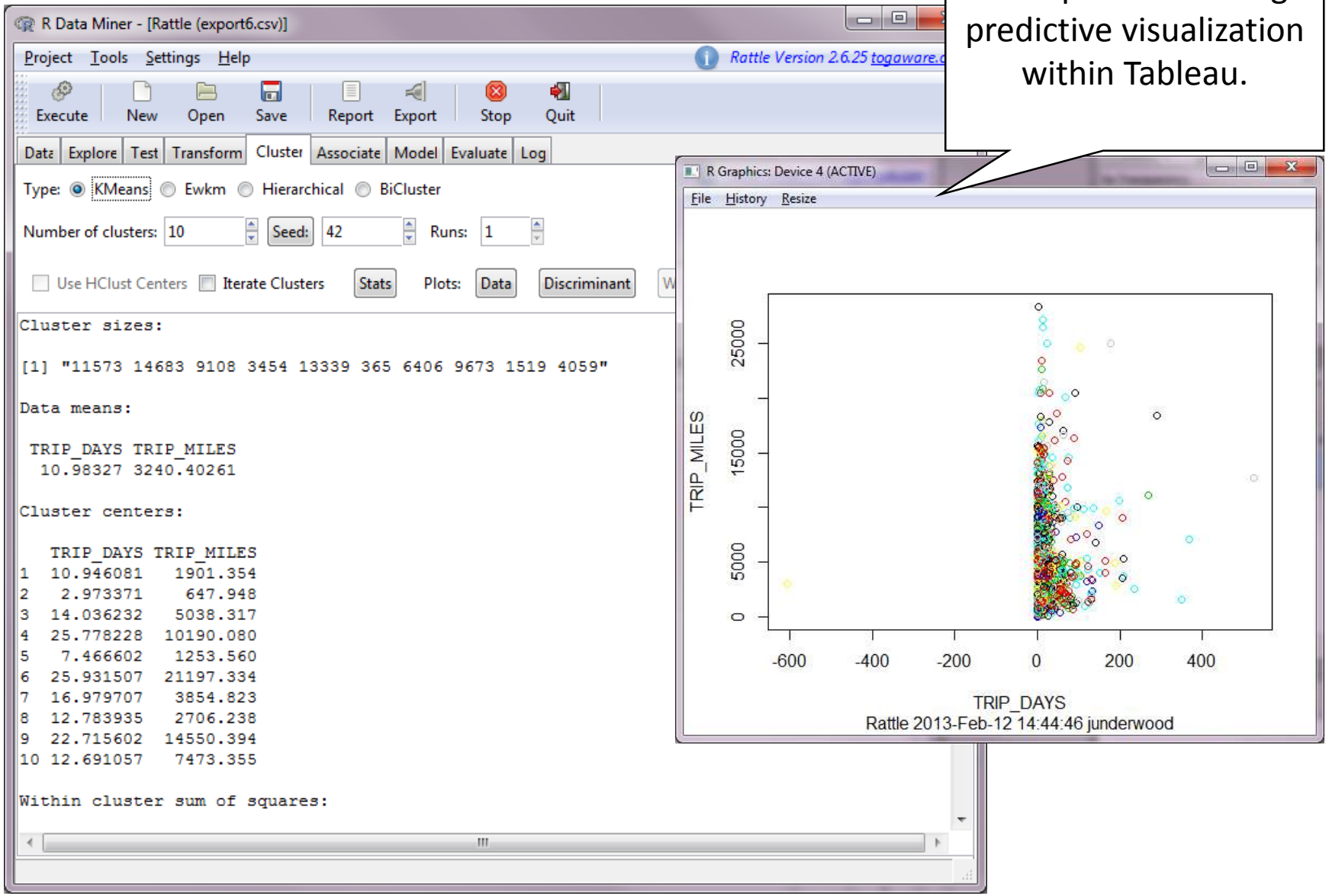




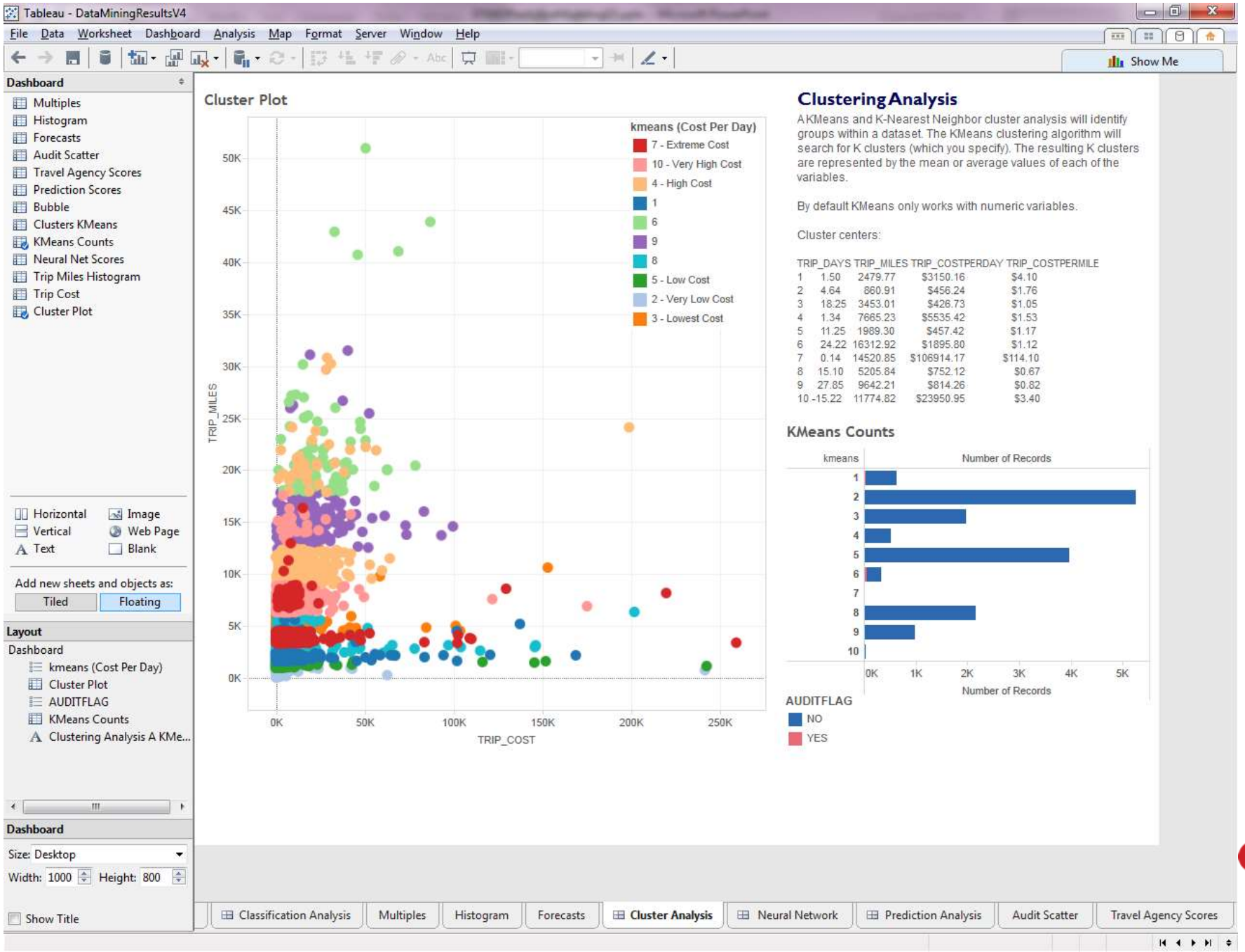
impact analytix

Clustering

Example R clustering predictive visualization within Tableau.

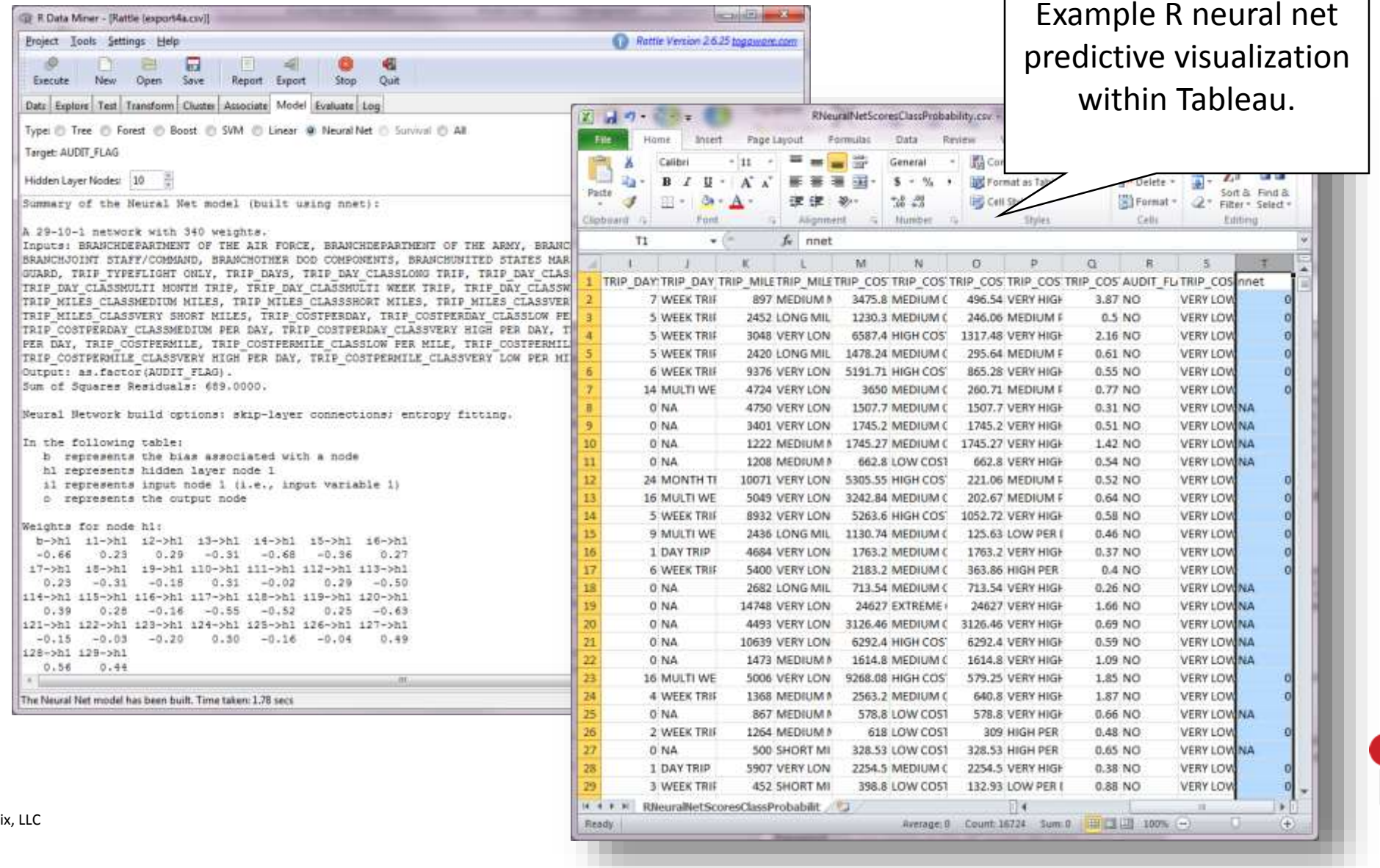


impact analytix



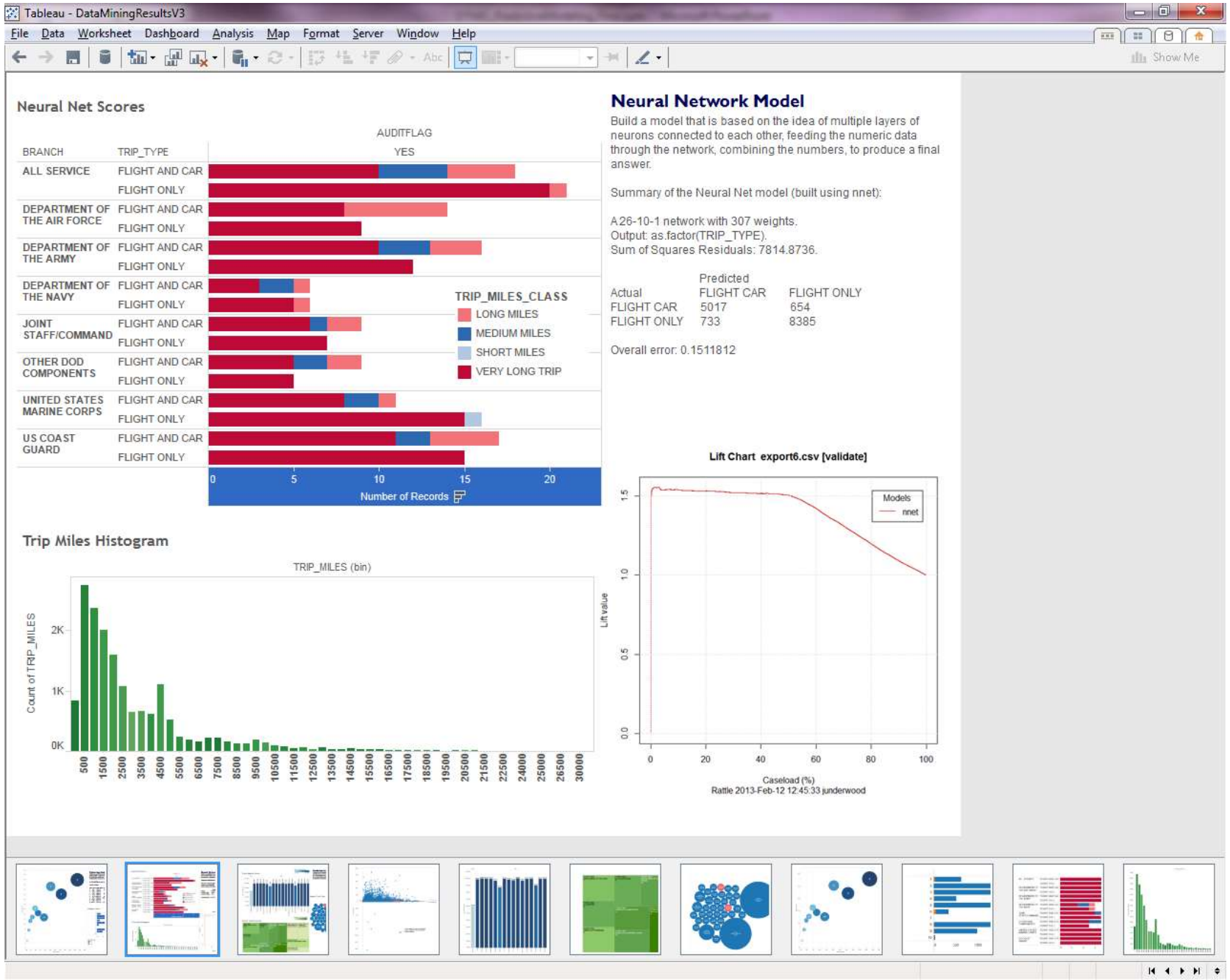
impact analytix

Neural Net with R

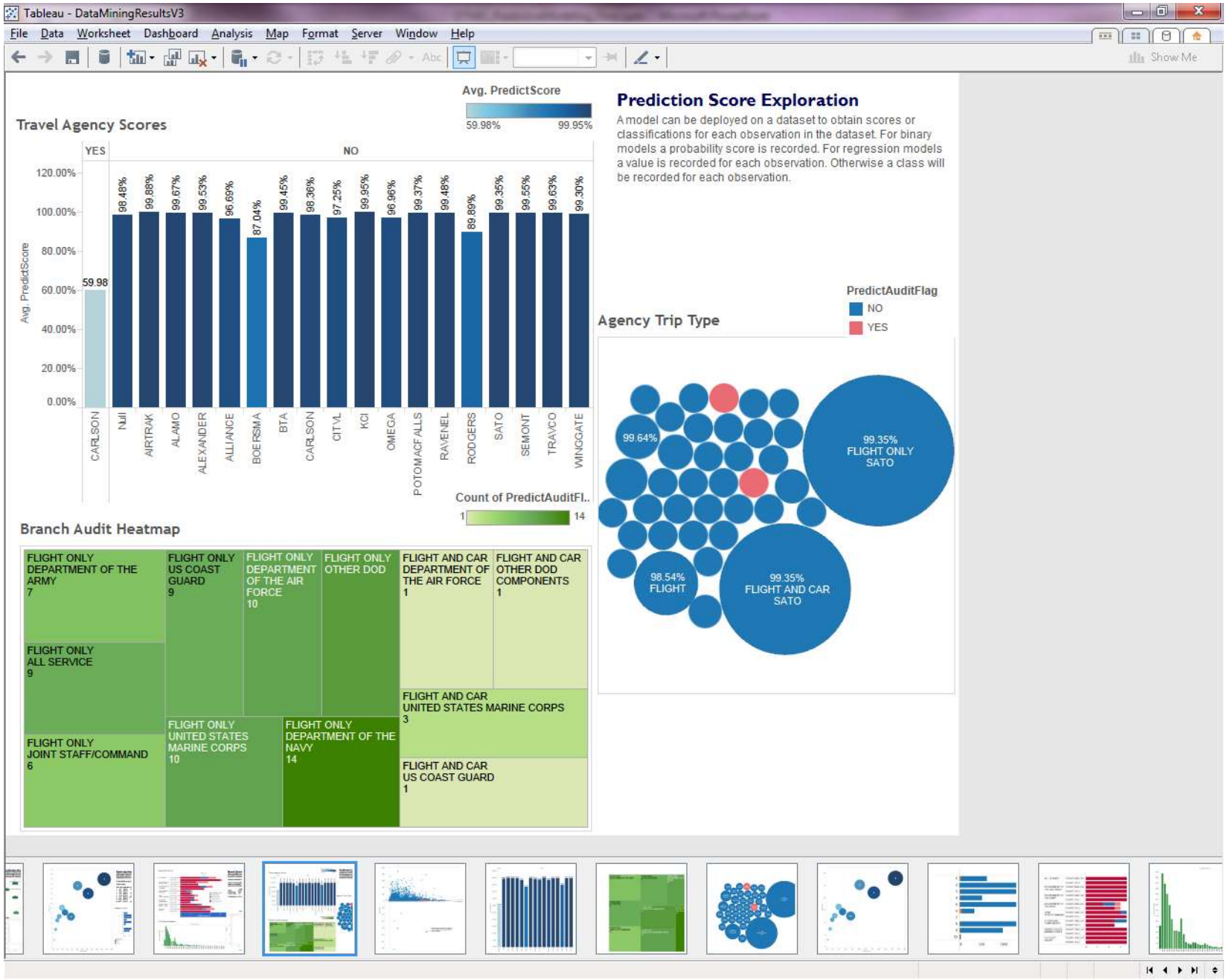


Example R neural net predictive visualization within Tableau.

impact analytix



impact analytix



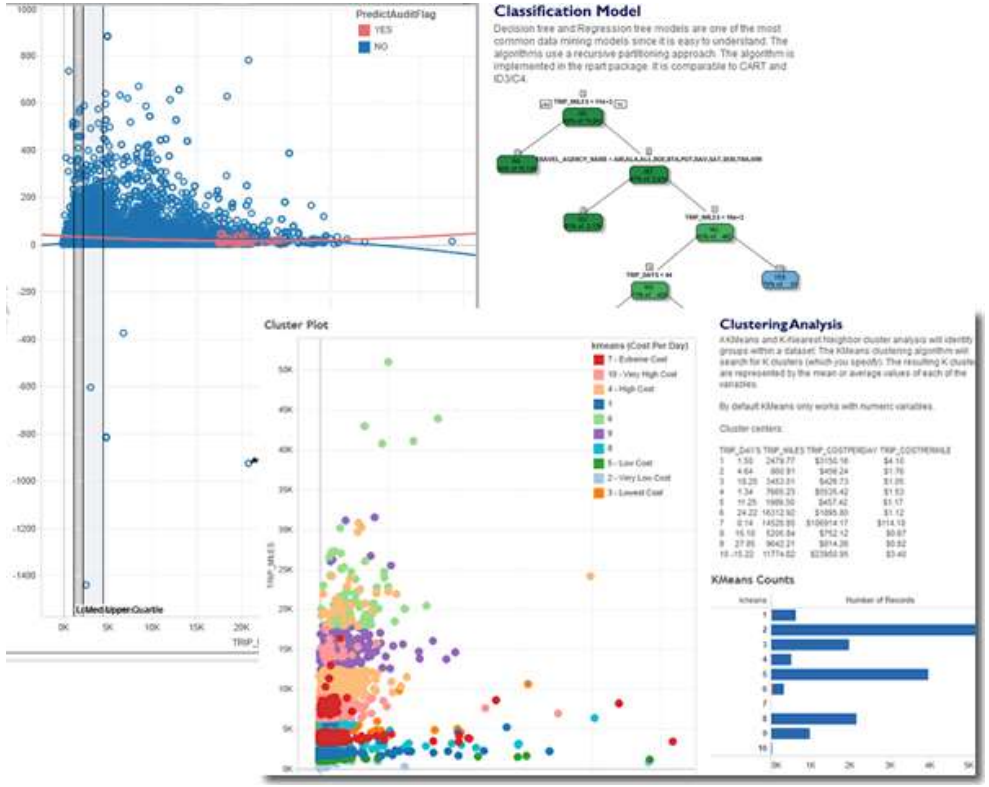
impact analytix

Additional Resources

- http://www.impactanalytix.com/presentations.htm
- http://www.slideshare.net/idigdata/predictive-model-data-visualization

KDD-Nuggets http://www.kdnuggets.com

- RapidMiner http://rapid-i.com
- R Statistical Computing http://www.r-project.org
- Revolution Analytics http://ww.revolutionanalytics.com
- Microsoft http://www.microsoft.com
- SAP http://www.sap.com
- Oracle http://www.oracle.com
- IBM http://www-01.ibm.com
- Teradata http://www.teradata.com
- Tableau http://www.tableausoftware.com
- Spotfire http://spotfire.tibco.com
- SAS http://www.sas.com
- IBM SPSS http://www-01.ibm.com/software/analytics/spss
- Mahout https://cwiki.apache.org/confluence/display/MAHOUT/Algorithms
- Weka Open Source Data Mining http://www.cs.waikato.ac.nz/ml/weka



http://public.tableausoftware.com/views/PredictiveDataVisualizationwithSSASDataMining/Classification?:embed=y

impact analytix