

quickly make a positive impact



**Jen Underwood**  
Founder & Principal Consultant  
Impact Analytix, LLC

813.435.5344  
jen@impactanalytix.com  
www.impactanalytix.com

© 2013 Impact Analytix, LLC

# Predictive Model Results Data Visualization with Tableau



# 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

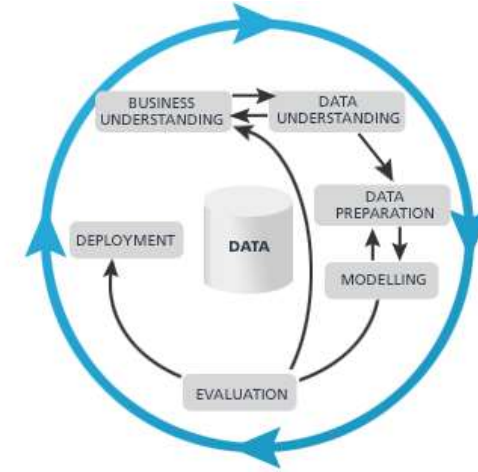


*"Decision making and the techniques and technologies to support and automate it will be the next competitive battleground for organizations. Those who are using business rules, data mining, analytics and optimization today are the shock troops of this next wave of business innovation."*

*- Tom Davenport, Competing on Analytics*

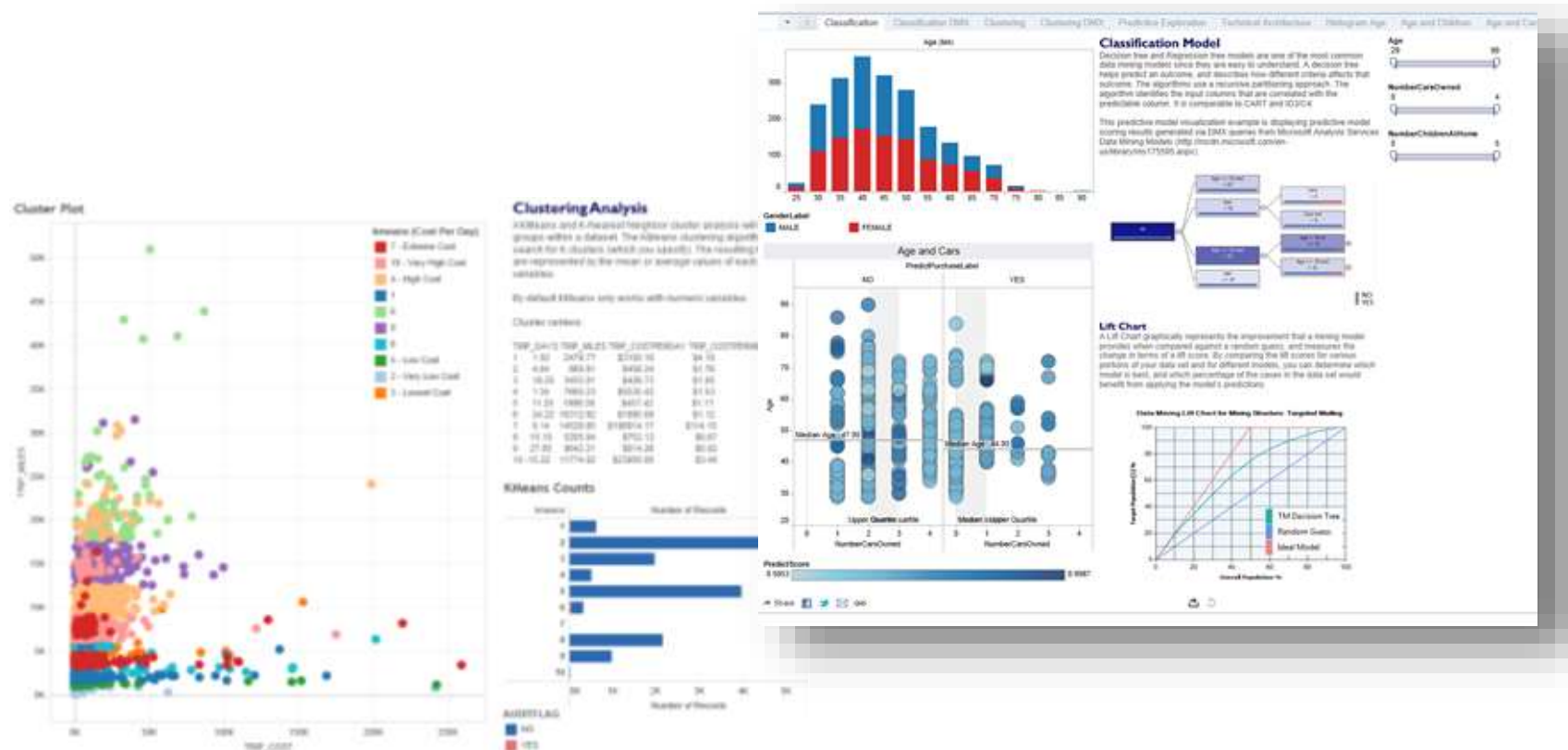
# Technical Solution Approach

- 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



# 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>



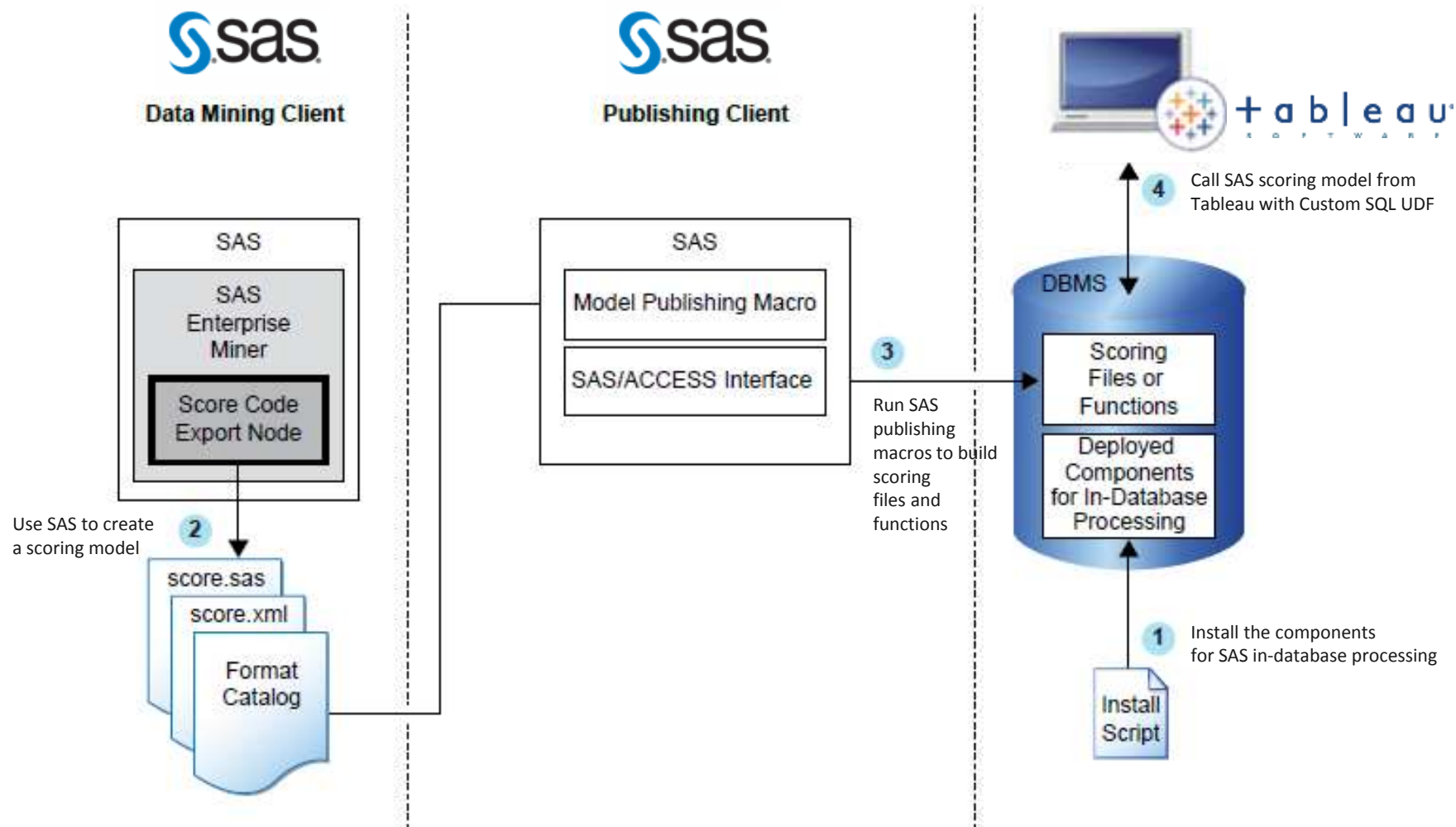
# Using SAS and SAS JMP with Tableau

# 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/indebug/65980/PDF/default/indebug.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



# Technical Solution Architecture





# SAS In-Database Processing

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

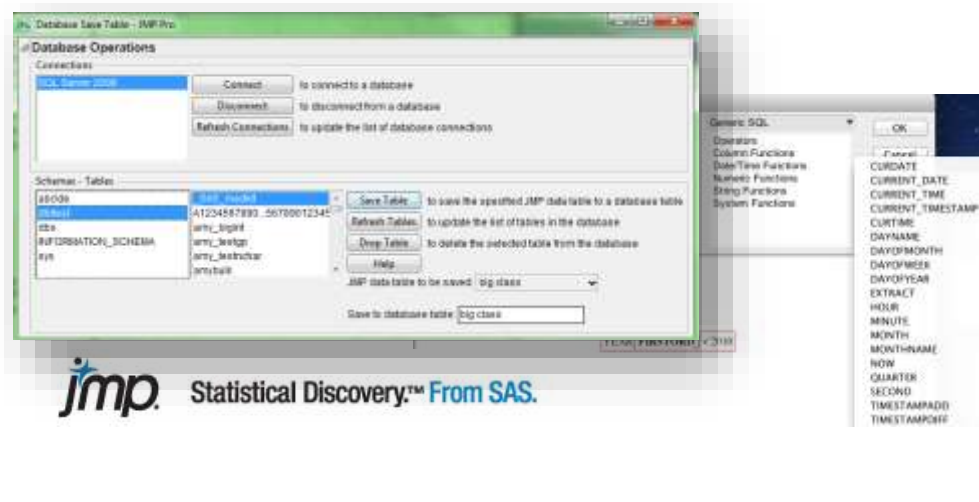
In-Database Feature	Software Required	Database Supported
Format publishing and the SAS_PUT( ) function	<ul style="list-style-type: none"><li>•Base SAS</li><li>•SAS/ACCESS Interface to the DBMS</li></ul>	Aster, DB2 under UNIX, Greenplum, Netezza, Oracle and Teradata
Scoring Models	<ul style="list-style-type: none"><li>•Base SAS</li><li>•SAS/ACCESS Interface to the DBMS</li><li>•SAS Scoring Accelerator</li><li>•SAS Model Manager (optional)</li></ul>	Aster, DB2 under UNIX, Greenplum, Netezza, Oracle and Teradata
Base SAS procedures: FREQ, RANK, REPORT, SORT SUMMARY/MEANS, TABULATE	<ul style="list-style-type: none"><li>•Base SAS</li><li>•SAS/ACCESS Interface to the DBMS</li></ul>	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	<ul style="list-style-type: none"><li>•Base SAS (for CORR)</li><li>•SAS/ACCESS Interface to Teradata</li><li>•SAS/STAT (for CANCORR, FACTOR, PRINCOMP, REG, SCORE, VARCLUS)</li><li>•SAS/ETS (for TIMESERIES)</li><li>•SAS Enterprise Miner (for DMDB, DMINE, DMREG)</li><li>•SAS Analytics Accelerator</li></ul>	Teradata

Source: <http://support.sas.com/documentation/cdl/en/indebug/65980/PDF/default/indebug.pdf>

# 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](http://jmp.com/about/events/summit2012/resources/Paper_Brian_Corcoran.pdf)

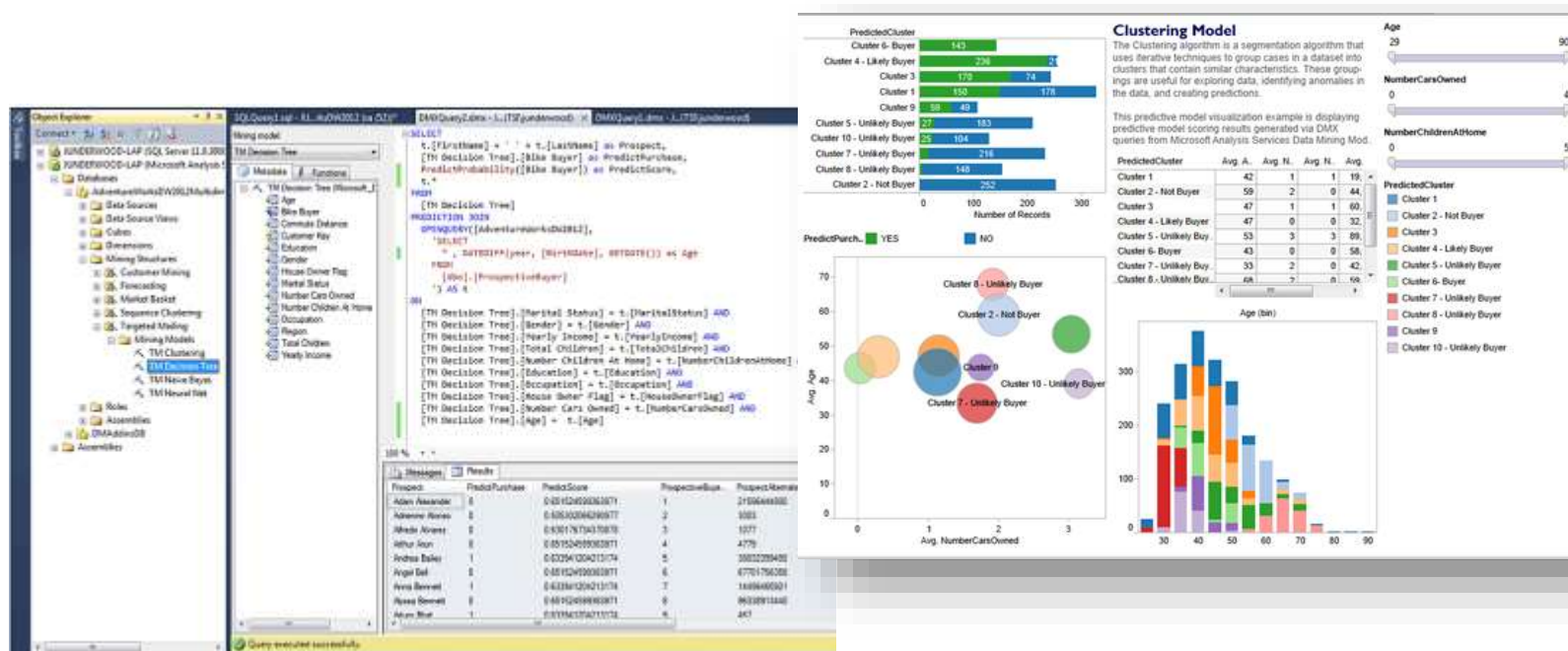
Tableau can directly connect to Excel, text, files and a wide variety of databases, issue SQL statements, call database specific UDFs in Custom SQL or Raw SQL functions, pass Tableau measure values or Tableau parameters for dynamic result data visualization



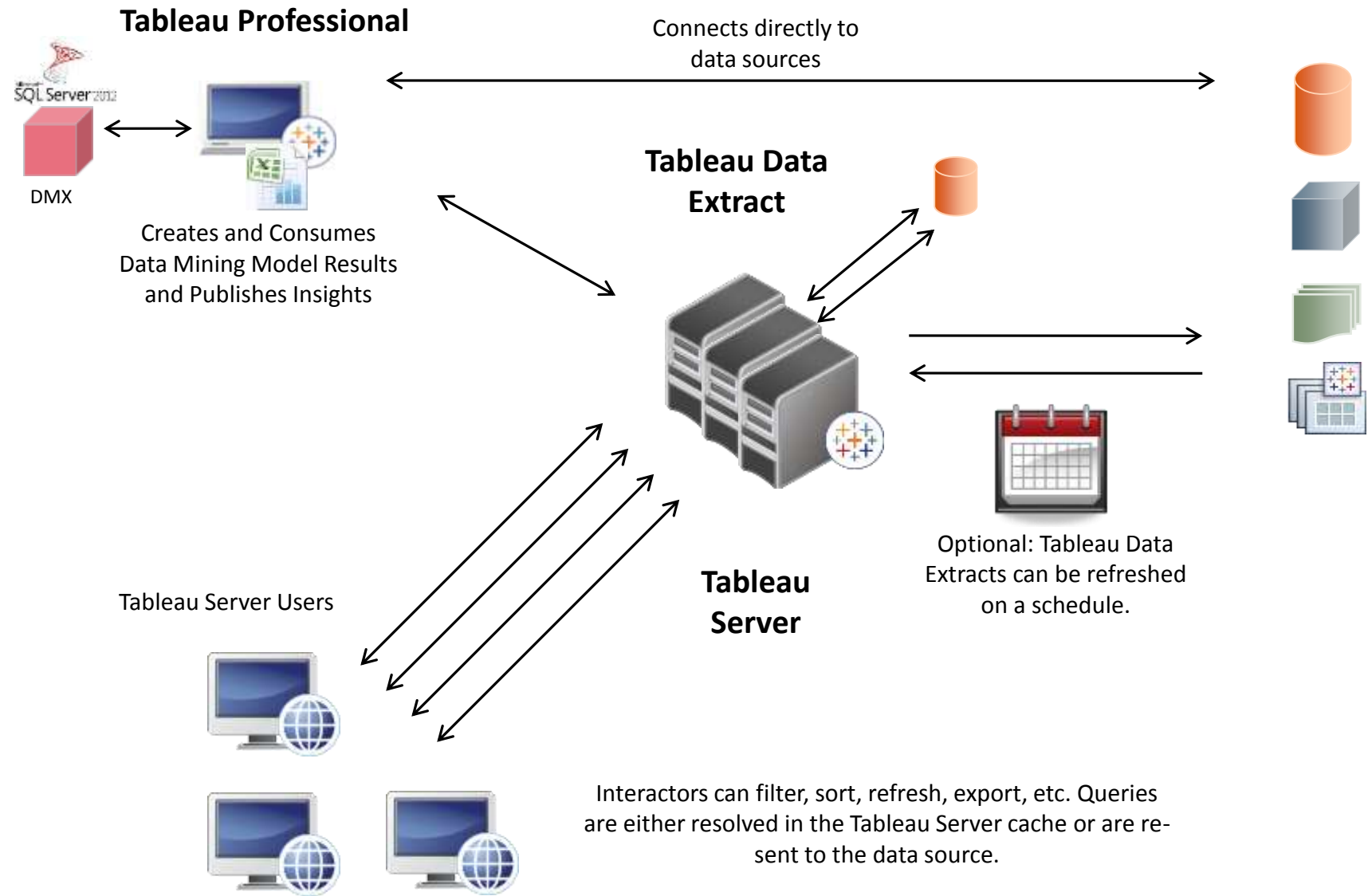
# Using Microsoft SQL Server Analysis Services Prediction Queries DMX with Tableau

# 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



# Using R or Revolution Analytics R with Tableau

# **Using Microsoft SQL Server Analysis Services Prediction Queries DMX with Tableau**



# 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

# Use R directly with Tableau Custom SQL

The screenshot displays the Tableau interface with a PostgreSQL connection window open. The connection window includes fields for server name, database, username, password, and a custom SQL query. The main Tableau window shows a data table with columns 'name', 'age', and 'salary'.

**PostgreSQL Connection Window:**

- Step 1: Enter a server name: `app-08.starschema.biz` Port: `5432`
- Step 2: Enter a database on the server: `rtest`
- Step 3: Enter information to log on to the database:  
Username: `tfoldi`  
Password: `*****`
- Step 4: Establish the connection: `Connect`
- Step 5: Select a table or view from the database:  
☐ Single Table ☐ Multiple Tables ☒ Custom SQL  
`SELECT *  
FROM get_emps(<Parameters.Salary Multiplier>)  
-- get_emps is an R function`
- Step 6: Give the connection a name for use in Tableau: `Custom SQL (rtest)`

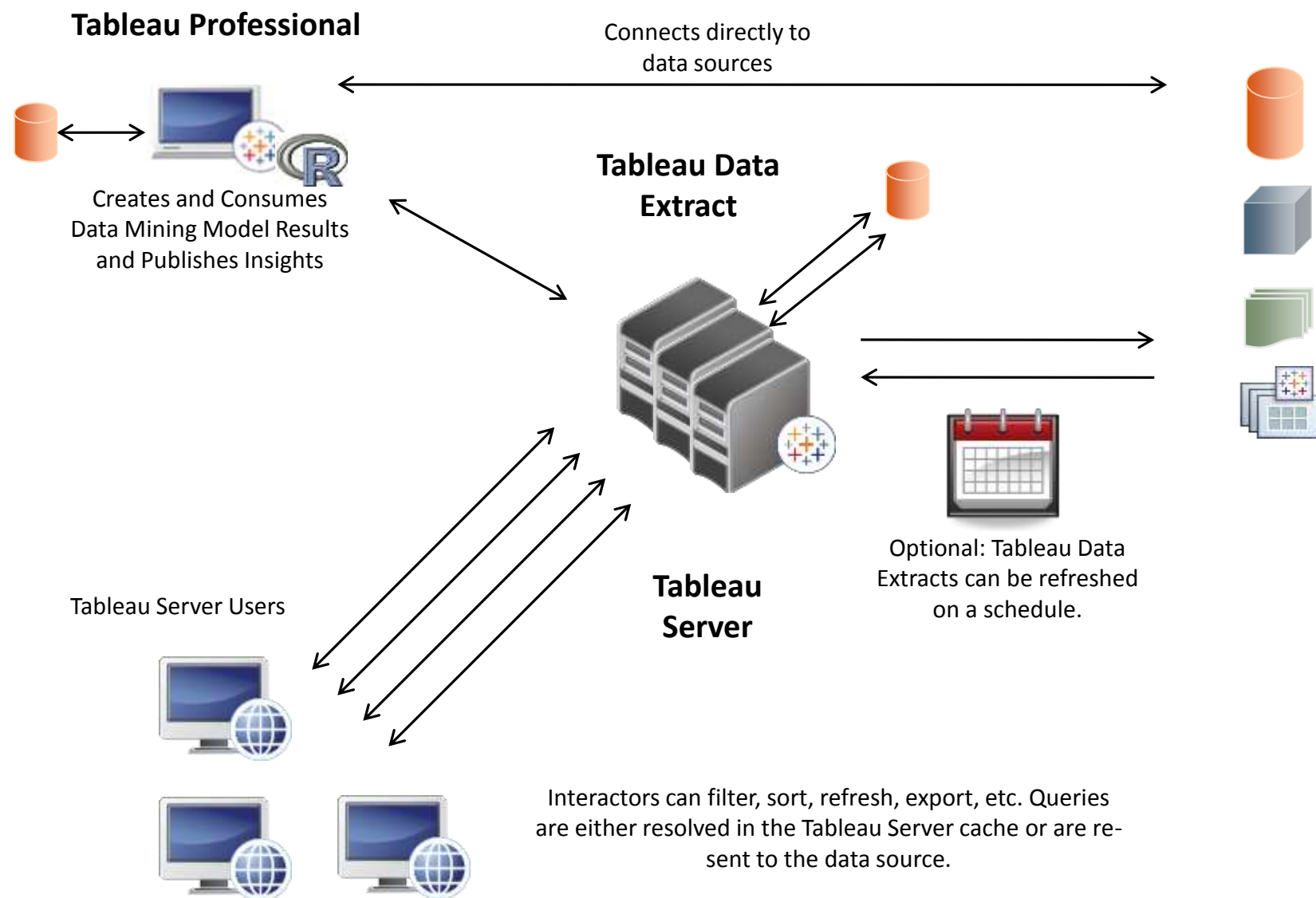
**Tableau Main Window:**

- Tableau - R\_integration
- File Data Worksheet Dashboard Analysis Map Format Server Window Help
- Data: Custom SQL (rtest), emp (rtest)
- Dimensions: name, Measure Names
- Columns: Measure Names
- Rows: name
- Table Data:

name	age	salary
Jim	25	120,000
Joe	41	250,000
Jon	35	50,000
- Measures: SUM(age), SUM(salary)
- Parameters: Salary Multiplier
- Status: 6 marks, 3 rows by 2 columns, SUM of Measure Values: 420,101

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

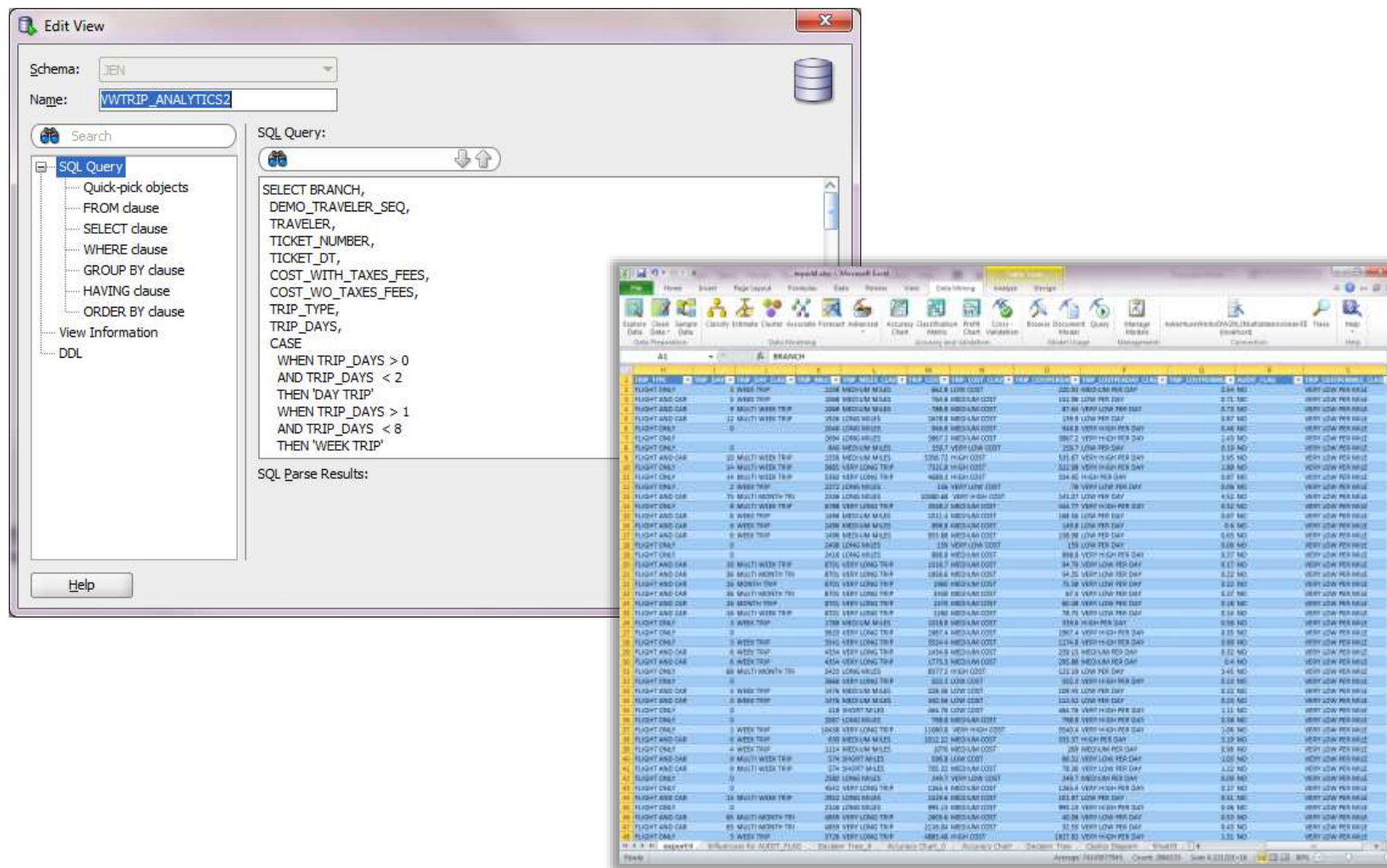
# Technical Solution Architecture



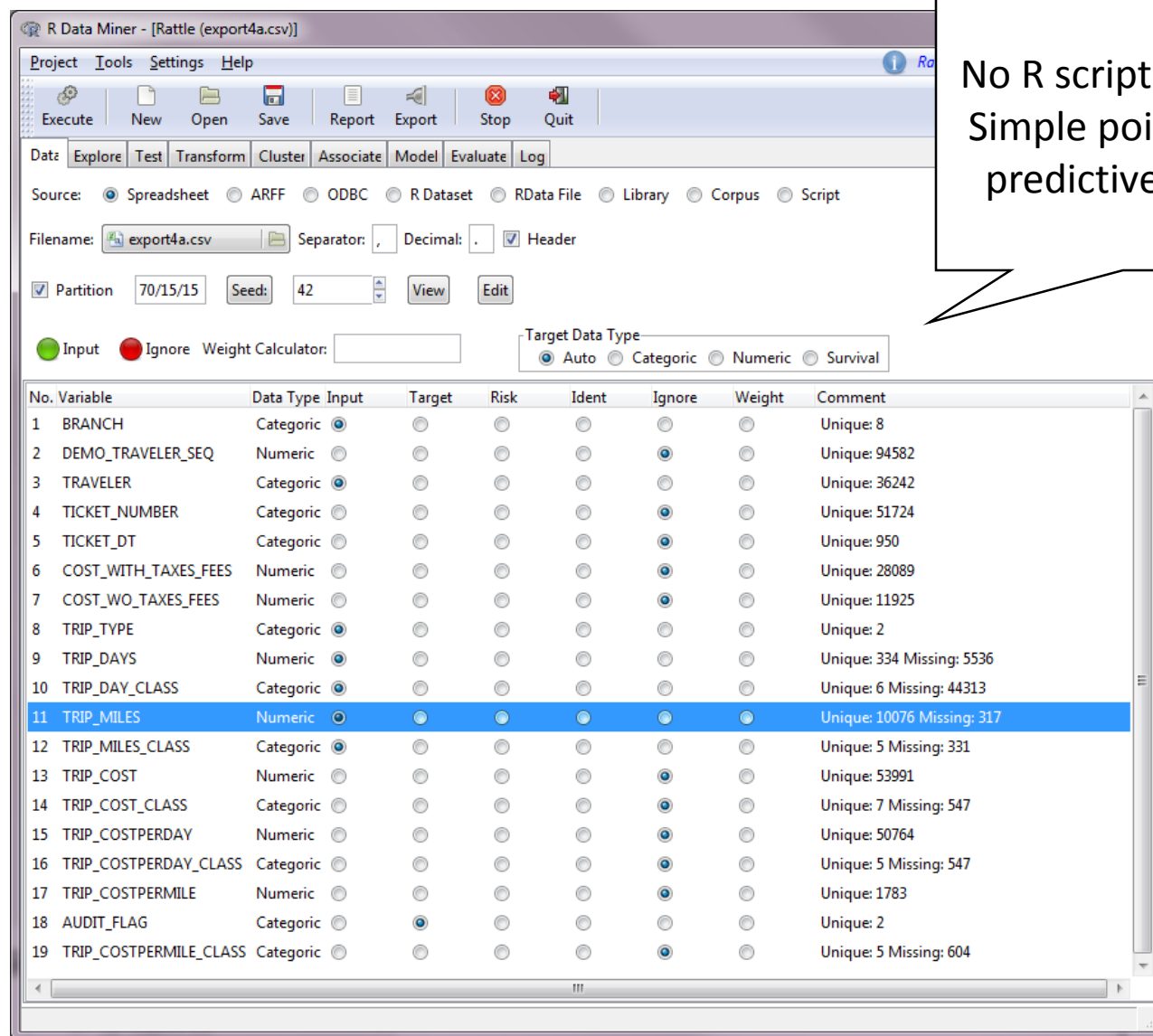
# Using R with Tableau

## A Step-by-Step Example

# Preparing Data

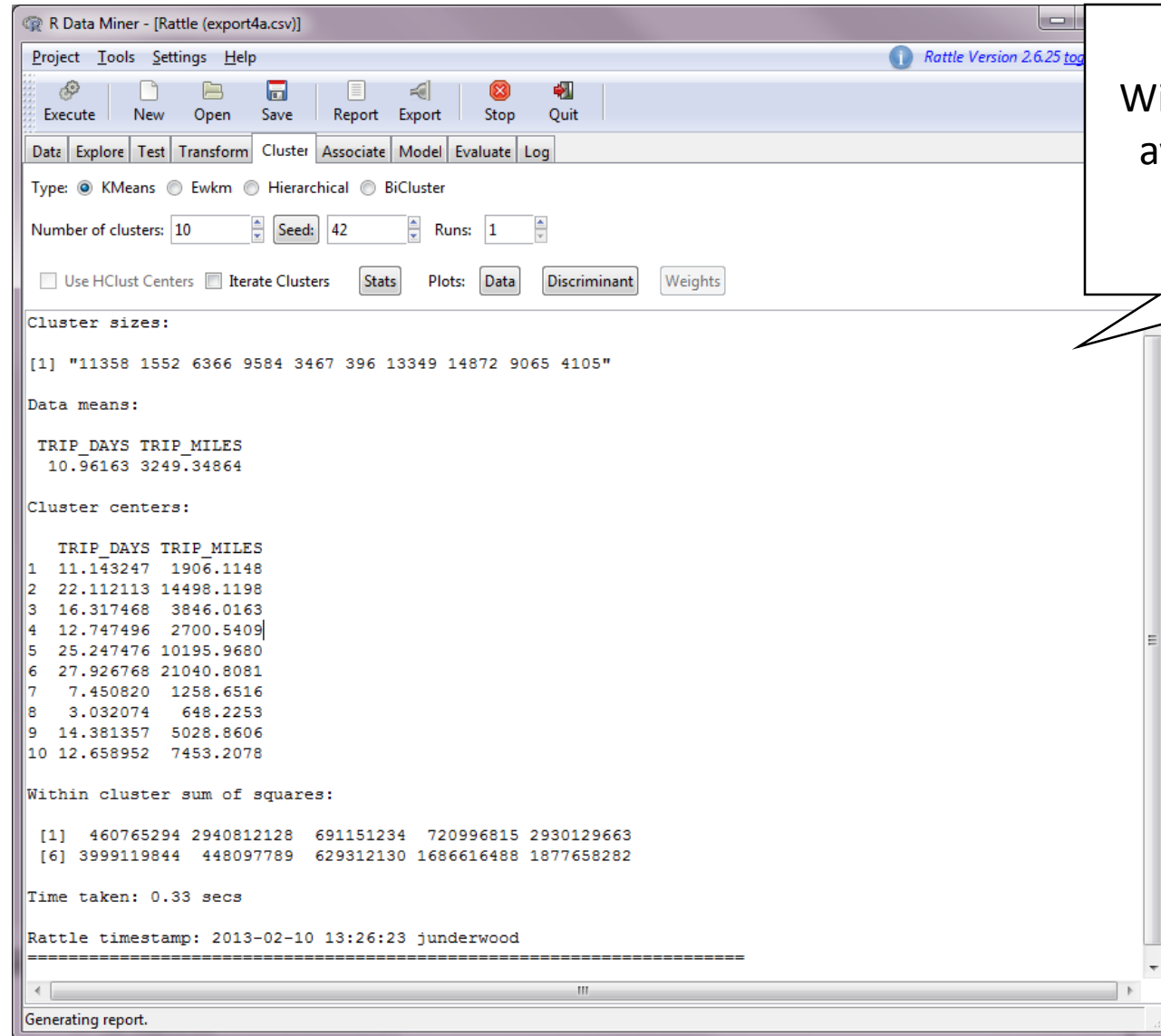


# Loading data in R



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

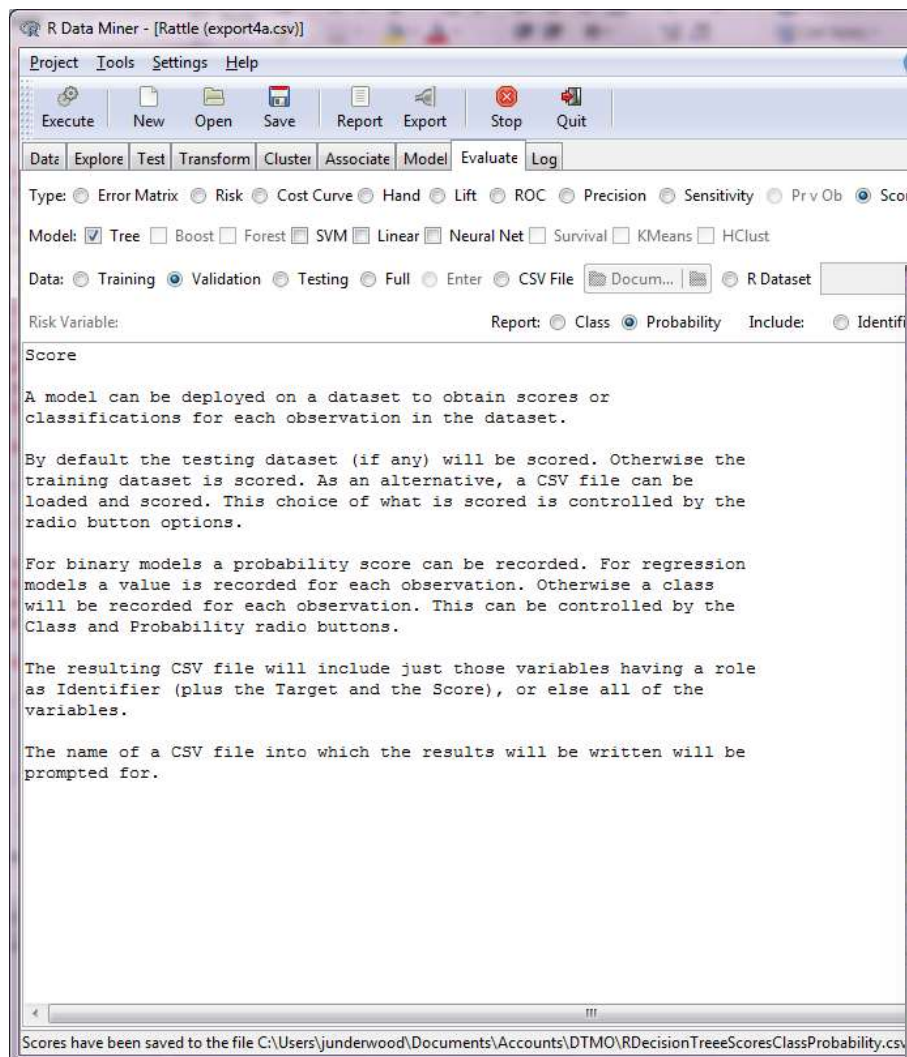
# Running Data Mining Models in R



Wide variety of freely available predictive algorithms.



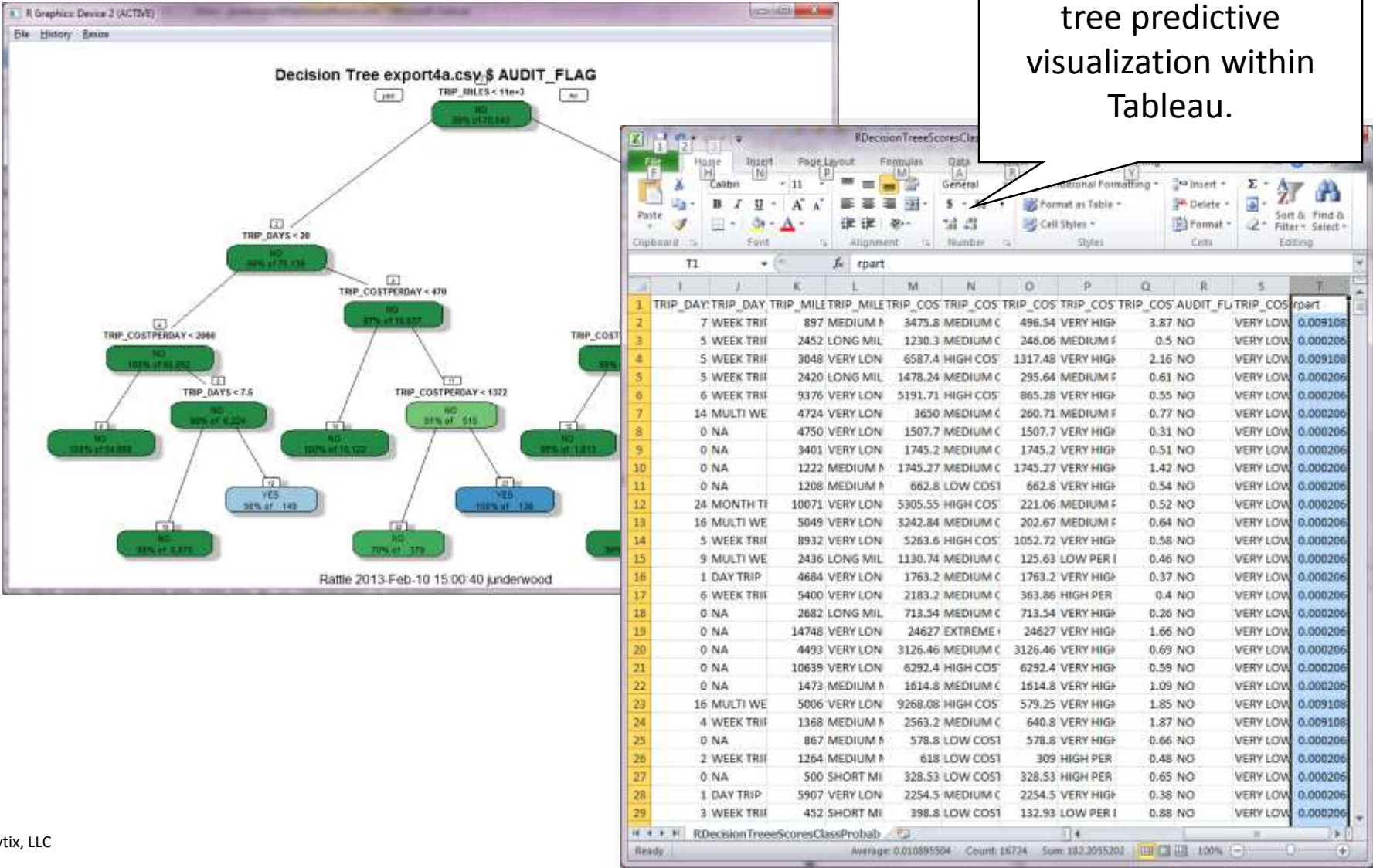
# Generating Model Prediction Scores



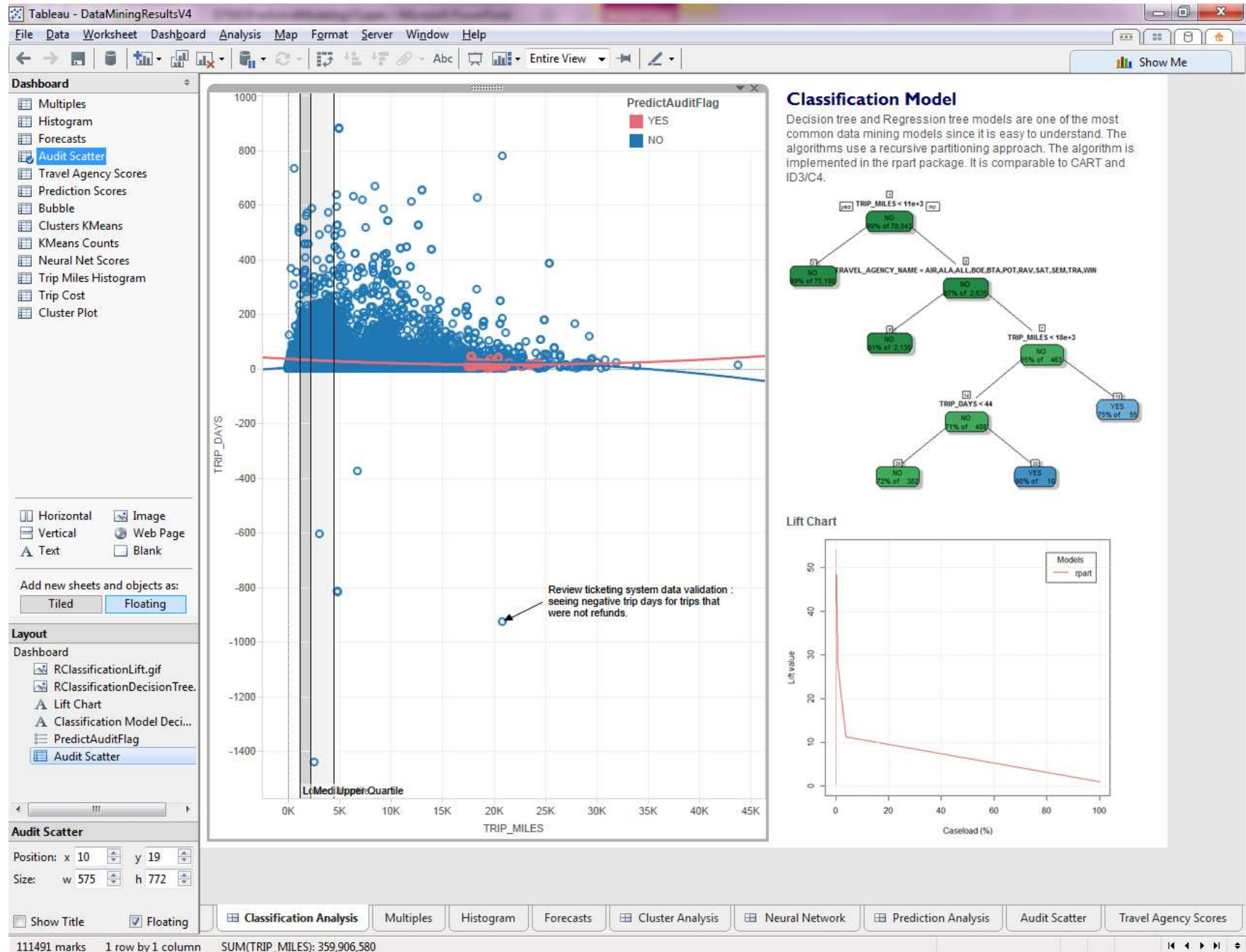
Export predictive models and scores to Excel or back to a database for Tableau.

RDecisionTreeScoresClassProbability																					
File Home Insert Page Layout Formulas Data Review Developer																					
Clipboard Font Paragraph Styles Cells																					
T1 rpart																					
	TRIP	DAY	TRIP	DAY	TRIP	MILE	TRIP	MILE	TRIP	COS	TRIP	COS	TRIP	COS	TRIP	COS	AUDIT	FU	TRIP	COS	rpart
1	7	WEEK	TRIP		897	MEDIUM		3475.8	MEDIUM		496.54	VERY HIGH		3.87	NO		VERY LOW			0.009108	
2	5	WEEK	TRIP		2452	LONG MIL		1230.3	MEDIUM		246.06	MEDIUM F		0.5	NO		VERY LOW			0.000206	
3	5	WEEK	TRIP		3048	VERY LON		6587.4	HIGH COS		1317.48	VERY HIGH		2.16	NO		VERY LOW			0.009108	
4	5	WEEK	TRIP		2420	LONG MIL		1478.24	MEDIUM		295.64	MEDIUM F		0.61	NO		VERY LOW			0.000206	
5	6	WEEK	TRIP		9376	VERY LON		5191.71	HIGH COS		865.28	VERY HIGH		0.55	NO		VERY LOW			0.000206	
6	14	MULTI WE			4724	VERY LON		3650	MEDIUM		260.71	MEDIUM F		0.77	NO		VERY LOW			0.000206	
7	0	NA			4750	VERY LON		1507.7	MEDIUM		1507.7	VERY HIGH		0.31	NO		VERY LOW			0.000206	
8	0	NA			3401	VERY LON		1745.2	MEDIUM		1745.2	VERY HIGH		0.51	NO		VERY LOW			0.000206	
9	0	NA			1222	MEDIUM M		1745.27	MEDIUM		1745.27	VERY HIGH		1.42	NO		VERY LOW			0.000206	
10	0	NA			1208	MEDIUM M		662.8	LOW COST		662.8	VERY HIGH		0.54	NO		VERY LOW			0.000206	
11	24	MONTH TI			10071	VERY LON		5305.55	HIGH COS		221.06	MEDIUM F		0.52	NO		VERY LOW			0.000206	
12	16	MULTI WE			5049	VERY LON		3242.84	MEDIUM		202.67	MEDIUM F		0.64	NO		VERY LOW			0.000206	
13	5	WEEK	TRIP		8932	VERY LON		5263.6	HIGH COS		1052.72	VERY HIGH		0.58	NO		VERY LOW			0.000206	
14	9	MULTI WE			2436	LONG MIL		1130.74	MEDIUM		125.63	LOW PER		0.46	NO		VERY LOW			0.000206	
15	1	DAY	TRIP		4684	VERY LON		1763.2	MEDIUM		1763.2	VERY HIGH		0.37	NO		VERY LOW			0.000206	
16	6	WEEK	TRIP		5400	VERY LON		2183.2	MEDIUM		363.86	HIGH PER		0.4	NO		VERY LOW			0.000206	
17	0	NA			2682	LONG MIL		713.54	MEDIUM		713.54	VERY HIGH		0.26	NO		VERY LOW			0.000206	
18	0	NA			14748	VERY LON		24627	EXTREME		24627	VERY HIGH		1.66	NO		VERY LOW			0.000206	
19	0	NA			4493	VERY LON		3126.46	MEDIUM		3126.46	VERY HIGH		0.69	NO		VERY LOW			0.000206	
20	0	NA			10639	VERY LON		6292.4	HIGH COS		6292.4	VERY HIGH		0.59	NO		VERY LOW			0.000206	
21	0	NA			1473	MEDIUM M		1614.8	MEDIUM		1614.8	VERY HIGH		1.09	NO		VERY LOW			0.000206	
22	16	MULTI WE			5006	VERY LON		9268.08	HIGH COS		579.25	VERY HIGH		1.85	NO		VERY LOW			0.009108	
23	4	WEEK	TRIP		1368	MEDIUM M		2563.2	MEDIUM		640.8	VERY HIGH		1.87	NO		VERY LOW			0.009108	
24	0	NA			867	MEDIUM M		578.8	LOW COST		578.8	VERY HIGH		0.66	NO		VERY LOW			0.000206	
25	2	WEEK	TRIP		1264	MEDIUM M		618	LOW COST		309	HIGH PER		0.48	NO		VERY LOW			0.000206	
26	0	NA			500	SHORT MIL		328.53	LOW COST		328.53	HIGH PER		0.65	NO		VERY LOW			0.000206	
27	1	DAY	TRIP		5907	VERY LON		2254.5	MEDIUM		2254.5	VERY HIGH		0.38	NO		VERY LOW			0.000206	
28	3	WEEK	TRIP		452	SHORT MIL		398.8	LOW COST		132.93	LOW PER		0.88	NO		VERY LOW			0.000206	
RDecisionTreeScoresClassProbability																					
Average: 0.001435506 Count: 30 Size: 0.04157674																					

# Classification Decision Tree

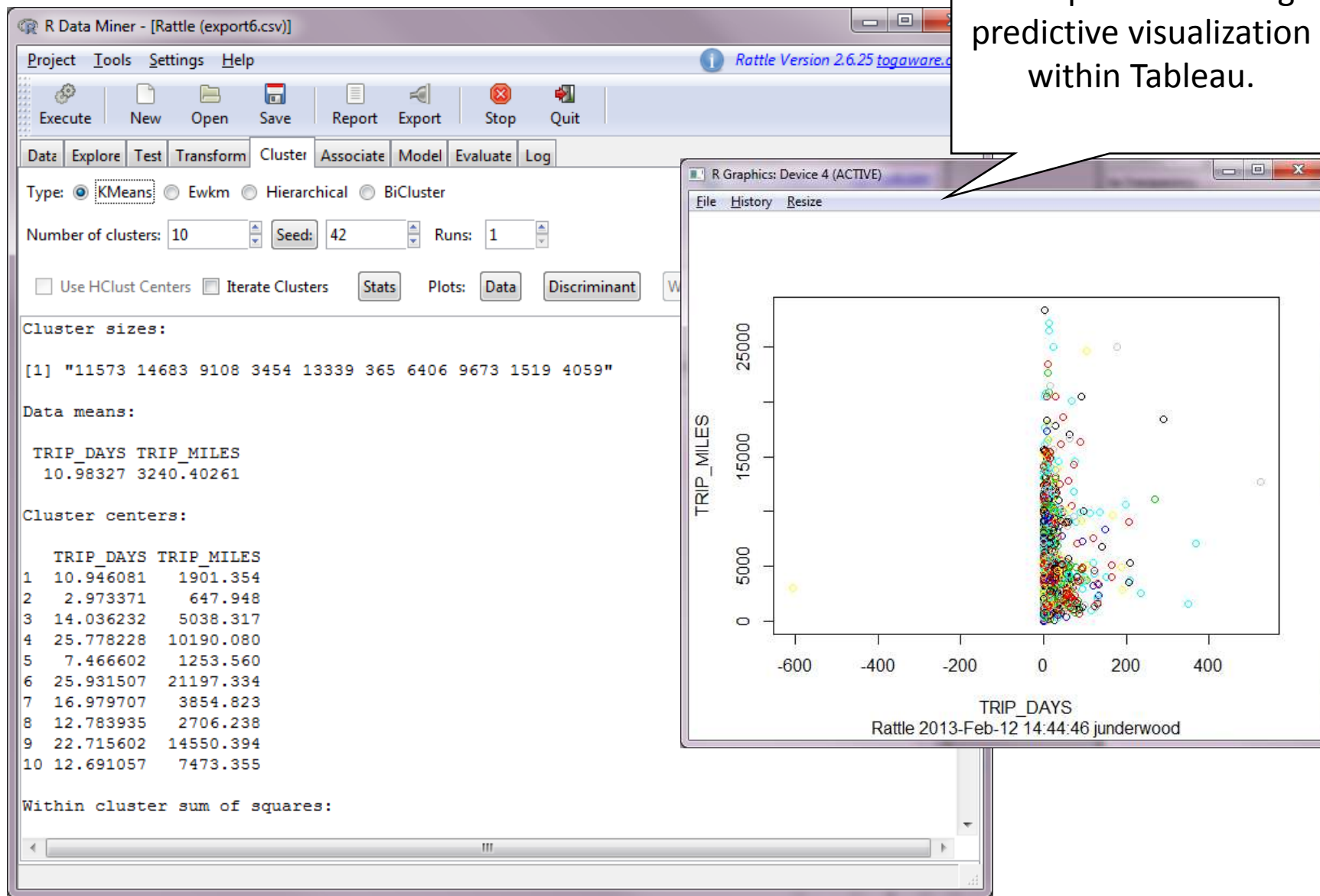






# Clustering

Example R clustering  
predictive visualization  
within Tableau.







# Neural Net with R

```
R Data Miner - [Rattle (export4a.csv)]
Project Tools Settings Help
Execute New Open Save Report Export Stop Quit
Data Explore Test Transform Cluster Associate Model Evaluate Log
Type: Tree Forest Boost SVM Linear Neural Net Survival All
Target: AUDIT_FLAG
Hidden Layer Nodes: 10

Summary of the Neural Net model (built using nnet):
A 29-10-1 network with 340 weights.
Inputs: BRANCHDEPARTMENT OF THE AIR FORCE, BRANCHDEPARTMENT OF THE ARMY, BRANCHJOINT STAFF/COMMAND, BRANCHOTHER DOD COMPONENTS, BRANCHUNITED STATES MAR GUARD, TRIP TYPEFLIGHT ONLY, TRIP_DAYS, TRIP_DAY_CLASSLONG TRIP, TRIP_DAY_CLASS TRIP_DAY_CLASSMULTI MONTH TRIP, TRIP_DAY_CLASSMULTI WEEK TRIP, TRIP_DAY_CLASS TRIP_MILES_CLASSMEDIUM MILES, TRIP_MILES_CLASSSHORT MILES, TRIP_MILES_CLASSVERY SHORT MILES, TRIP_COSTPERDAY, TRIP_COSTPERDAY_CLASSLOW PER TRIP_COSTPERDAY_CLASSMEDIUM PER DAY, TRIP_COSTPERDAY_CLASSVERY HIGH PER DAY, TRIP_COSTPERMILE, TRIP_COSTPERMILE_CLASSLOW PER MILE, TRIP_COSTPERMILE_CLASSVERY HIGH PER DAY, TRIP_COSTPERMILE_CLASSVERY LOW PER MI
Output: as.factor(AUDIT_FLAG).
Sum of Squares Residuals: 689.0000.

Neural Network build options: skip-layer connections; entropy fitting.

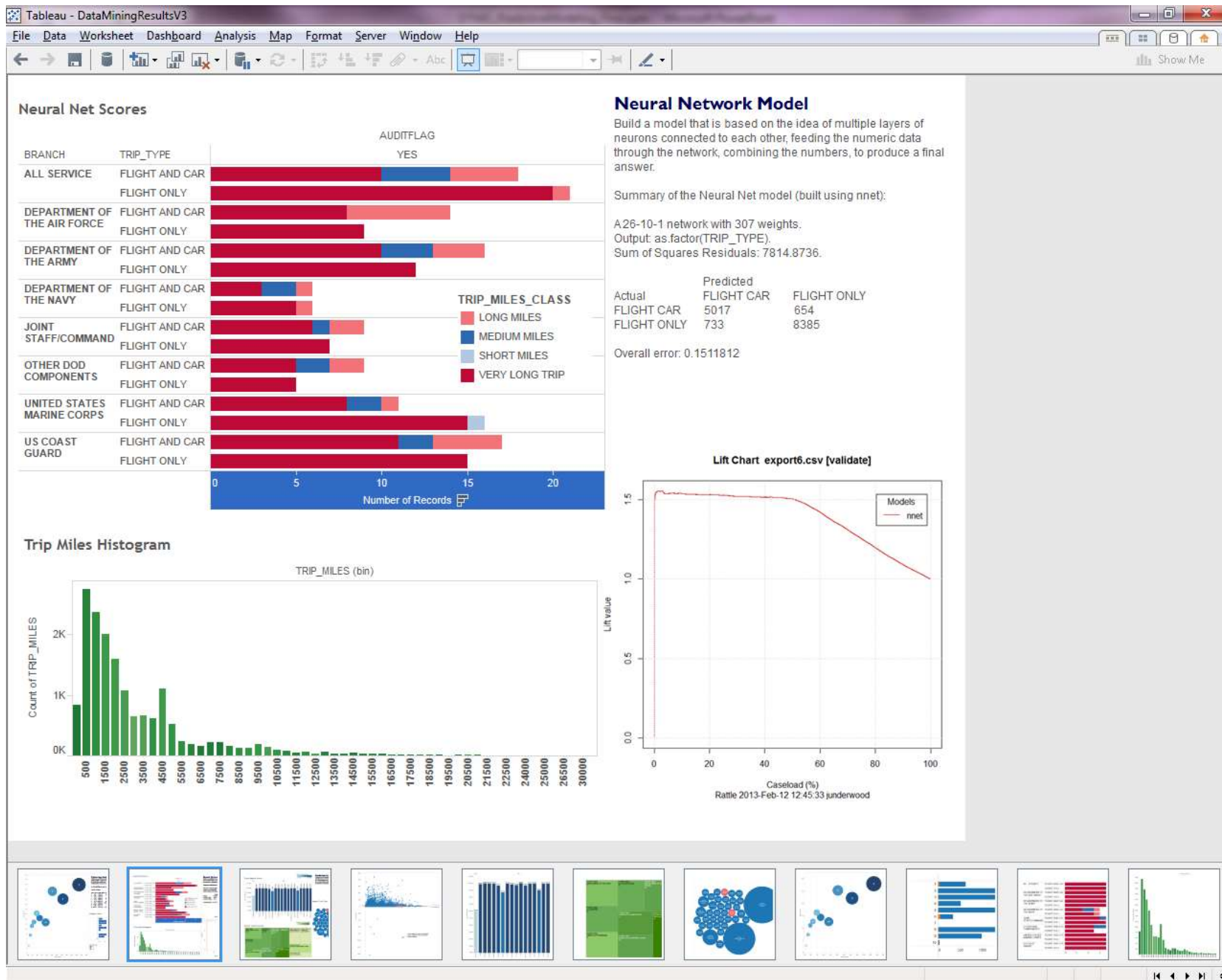
In the following table:
b represents the bias associated with a node
h1 represents hidden layer node 1
i1 represents input node 1 (i.e., input variable 1)
o represents the output node

Weights for node h1:
b->h1 i1->h1 i2->h1 i3->h1 i4->h1 i5->h1 i6->h1
-0.66 0.23 0.29 -0.31 -0.68 -0.36 0.27
i7->h1 i8->h1 i9->h1 i10->h1 i11->h1 i12->h1 i13->h1
0.23 -0.31 -0.18 0.31 -0.02 0.29 -0.50
i14->h1 i15->h1 i16->h1 i17->h1 i18->h1 i19->h1 i20->h1
0.39 0.28 -0.16 -0.55 -0.52 0.25 -0.63
i21->h1 i22->h1 i23->h1 i24->h1 i25->h1 i26->h1 i27->h1
-0.15 -0.03 -0.20 0.30 -0.16 -0.04 0.49
i28->h1 i29->h1
0.56 0.44

The Neural Net model has been built. Time taken: 1.78 secs
```

Example R neural net predictive visualization within Tableau.

RNeuralNetScoresClassProbability.csv												
File Home Insert Page Layout Formulas Data Review												
Clipboard Font Alignment Number Styles Cells Editing												
Calibri 11 General Format as Table Cell Styles Delete Sort & Find & Filter Select												
T1 nnet												
	I	J	K	L	M	N	O	P	Q	R	S	T
1	TRIP_DAY	TRIP_DAY	TRIP_MILE	TRIP_MILE	TRIP_COST	TRIP_COST	TRIP_COST	TRIP_COST	TRIP_COST	AUDIT_FL	TRIP_COST	nnet
2	7	WEEK TRIP	897	MEDIUM	3475.8	MEDIUM	496.54	VERY HIGH	3.87	NO	VERY LOW	0
3	5	WEEK TRIP	2452	LONG MIL	1230.3	MEDIUM	246.06	MEDIUM	0.5	NO	VERY LOW	0
4	5	WEEK TRIP	3048	VERY LON	6587.4	HIGH COS	1317.48	VERY HIGH	2.16	NO	VERY LOW	0
5	5	WEEK TRIP	2420	LONG MIL	1478.24	MEDIUM	295.64	MEDIUM	0.61	NO	VERY LOW	0
6	6	WEEK TRIP	9376	VERY LON	5191.71	HIGH COS	865.28	VERY HIGH	0.55	NO	VERY LOW	0
7	14	MULTI WE	4724	VERY LON	3650	MEDIUM	260.71	MEDIUM	0.77	NO	VERY LOW	0
8	0	NA	4750	VERY LON	1507.7	MEDIUM	1507.7	VERY HIGH	0.31	NO	VERY LOW	NA
9	0	NA	3401	VERY LON	1745.2	MEDIUM	1745.2	VERY HIGH	0.51	NO	VERY LOW	NA
10	0	NA	1222	MEDIUM	1745.27	MEDIUM	1745.27	VERY HIGH	1.42	NO	VERY LOW	NA
11	0	NA	1208	MEDIUM	662.8	LOW COS	662.8	VERY HIGH	0.54	NO	VERY LOW	NA
12	24	MONTH TI	10071	VERY LON	5305.55	HIGH COS	221.06	MEDIUM	0.52	NO	VERY LOW	0
13	16	MULTI WE	5049	VERY LON	3242.84	MEDIUM	202.67	MEDIUM	0.64	NO	VERY LOW	0
14	5	WEEK TRIP	8932	VERY LON	5263.6	HIGH COS	1052.72	VERY HIGH	0.58	NO	VERY LOW	0
15	9	MULTI WE	2436	LONG MIL	1130.74	MEDIUM	125.63	LOW PER	0.46	NO	VERY LOW	0
16	1	DAY TRIP	4684	VERY LON	1763.2	MEDIUM	1763.2	VERY HIGH	0.37	NO	VERY LOW	0
17	6	WEEK TRIP	5400	VERY LON	2183.2	MEDIUM	363.86	HIGH PER	0.4	NO	VERY LOW	0
18	0	NA	2682	LONG MIL	713.54	MEDIUM	713.54	VERY HIGH	0.26	NO	VERY LOW	NA
19	0	NA	14748	VERY LON	24627	EXTREME	24627	VERY HIGH	1.66	NO	VERY LOW	NA
20	0	NA	4493	VERY LON	3126.46	MEDIUM	3126.46	VERY HIGH	0.69	NO	VERY LOW	NA
21	0	NA	10639	VERY LON	6292.4	HIGH COS	6292.4	VERY HIGH	0.59	NO	VERY LOW	NA
22	0	NA	1473	MEDIUM	1614.8	MEDIUM	1614.8	VERY HIGH	1.09	NO	VERY LOW	NA
23	16	MULTI WE	5006	VERY LON	9268.08	HIGH COS	579.25	VERY HIGH	1.85	NO	VERY LOW	0
24	4	WEEK TRIP	1368	MEDIUM	2563.2	MEDIUM	640.8	VERY HIGH	1.87	NO	VERY LOW	0
25	0	NA	867	MEDIUM	578.8	LOW COS	578.8	VERY HIGH	0.66	NO	VERY LOW	NA
26	2	WEEK TRIP	1264	MEDIUM	618	LOW COS	309	HIGH PER	0.48	NO	VERY LOW	0
27	0	NA	500	SHORT MI	328.53	LOW COS	328.53	HIGH PER	0.65	NO	VERY LOW	NA
28	1	DAY TRIP	5907	VERY LON	2254.5	MEDIUM	2254.5	VERY HIGH	0.38	NO	VERY LOW	0
29	3	WEEK TRIP	452	SHORT MI	398.8	LOW COS	132.93	LOW PER	0.88	NO	VERY LOW	0
RNeuralNetScoresClassProbability												
Average: 0 Count: 16724 Sum: 0 100%												

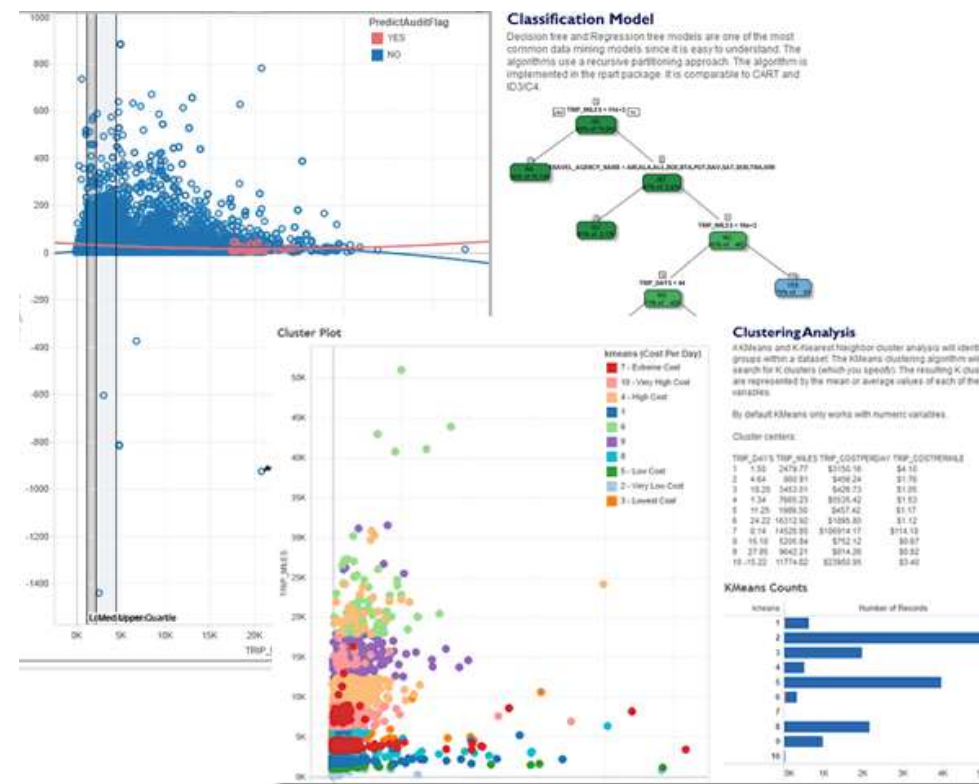






# Additional Resources

- <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://www.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/PredictiveDataVisualization  
withSSASDataMining/Classification?:embed=y](http://public.tableausoftware.com/views/PredictiveDataVisualization%20withSSASDataMining/Classification?:embed=y)

