SAS® FORUM
UNITED KINGDOM 2015

SAS In-Memory Statistics and SAS Studio

Steve Ludlow

THE POWER TO KNOW.

What is SAS Studio?

What is IMSTAT?

Demonstration

New SAS Programming Environment

Available

Consistent

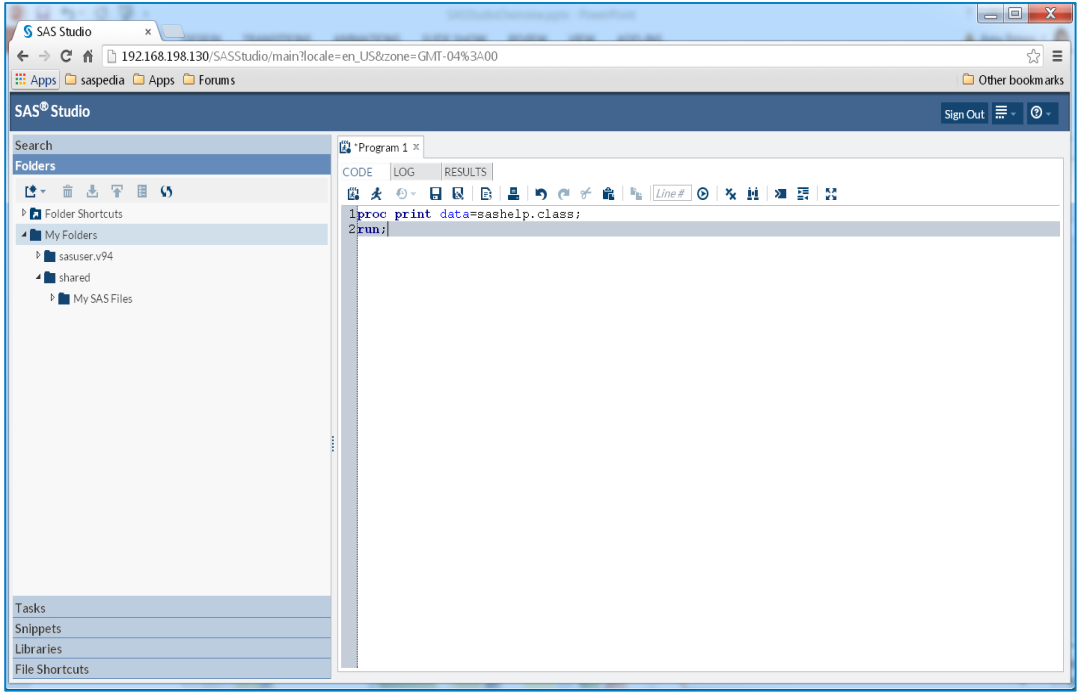
Assistive

Available through all modern web browsers

Available

Consistent

Assistive

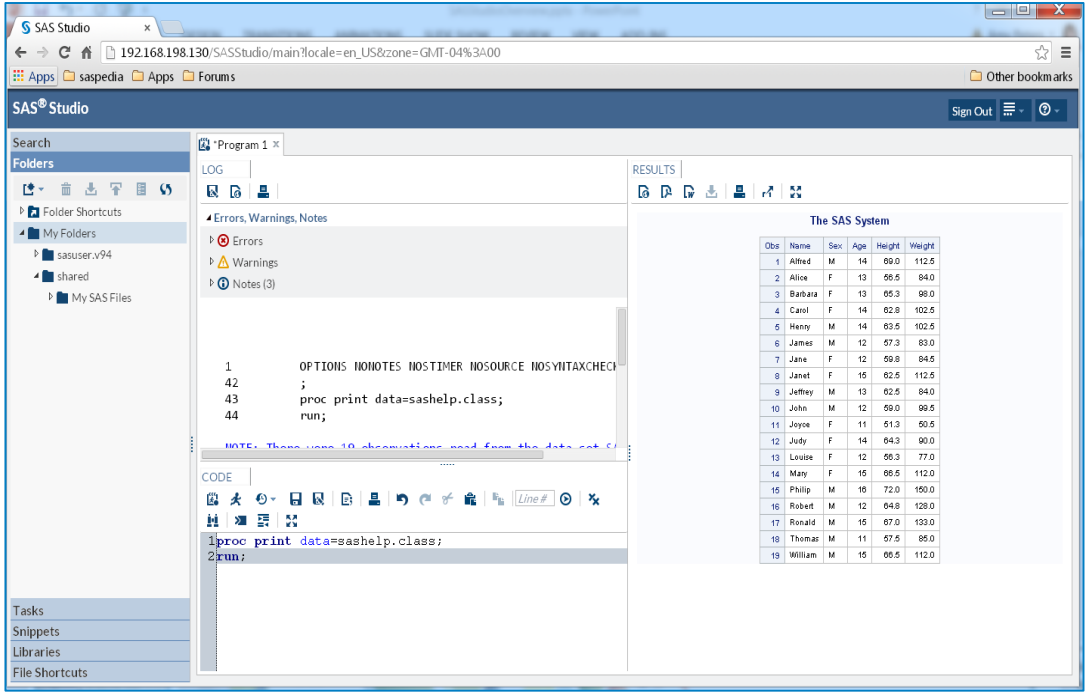


One interface for SAS Programming

Available

Consistent

Assistive

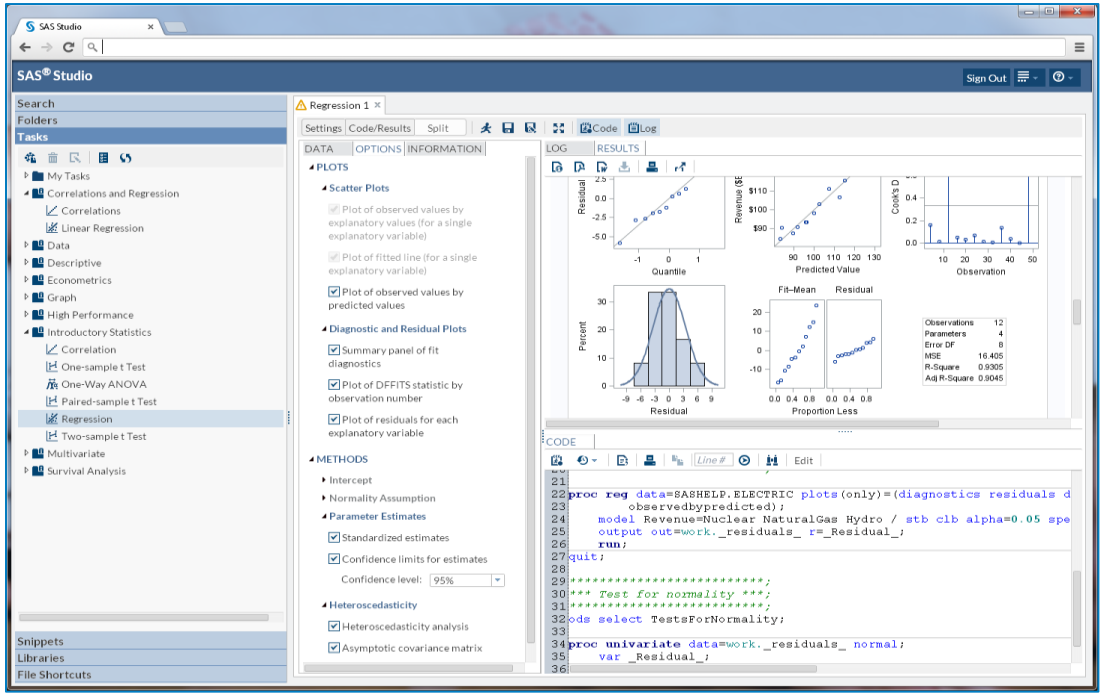


Increased Developer productivity

Available

Consistent

Assistive



Programming Interface

Programming interface

- Color-coded editor with auto-complete & pop-up syntax help
- Table viewer to sort or filter a table … and see/copy the generated SQL
- Log navigator
- Customise to create a “classic” DMS look

Snippets & Tasks

Snippets

SAS® Studio

Sign Out

Search
Folders
Tasks
Snippets

- Favorites
- Code Snippets
  - My Snippets
  - Data
    - Import CSV File
    - Import XLS File
    - DS2 Package
    - DS2 Code
    - DS2 Thread
    - Generate CSV File
    - Generate PowerPoint Slide
    - Generate XML File
    - Simulate Linear Regression Data
    - Simulate One-Way ANOVA Data
  - Descriptive
  - Graph
  - IML
  - Macro

Generate CSV File

CODE | LOG | RESULTS

```
/* Stream a CSV representation of SASHELP.CARS directly to the user's browser. */

proc export data=sashelp.cars
            outfile=_dataout
            dbms=csv replace;
run;

%let _DATAOUT_MIME_TYPE=text/csv;
%let _DATAOUT_NAME=cars.csv;
```

Tasks

SAS® Studio

Sign Out

Search
Folders
Tasks

- My Tasks
- Data
- Graph
- Statistics
  - Summary Statistics
  - Distribution Analysis
  - One-Way Frequencies
  - Correlations
  - Table Analysis
  - One-sample t Test
  - Paired-sample t Test
  - Two-sample t Test
  - One-Way ANOVA
  - Nonparametric One-Way A
  - Linear Regression

Snippets
Libraries
File Shortcuts

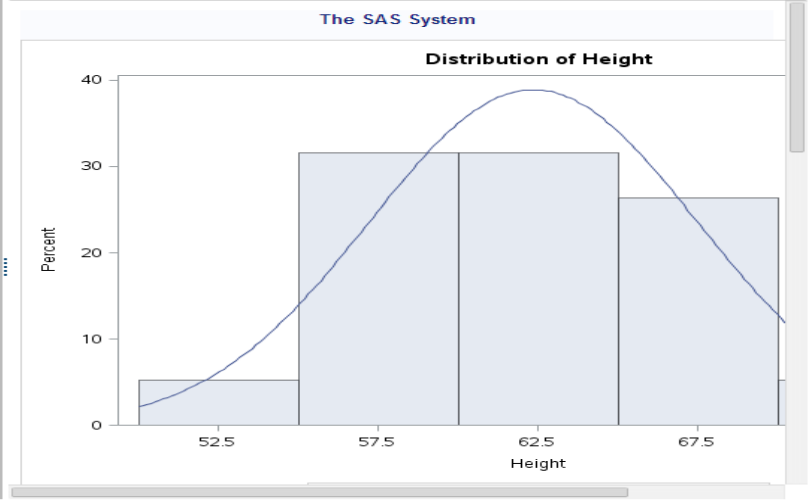
Distribution Analysis 1

Settings | Code/Results | Split | Code | Log

DATA | OPTIONS | INFORMATION

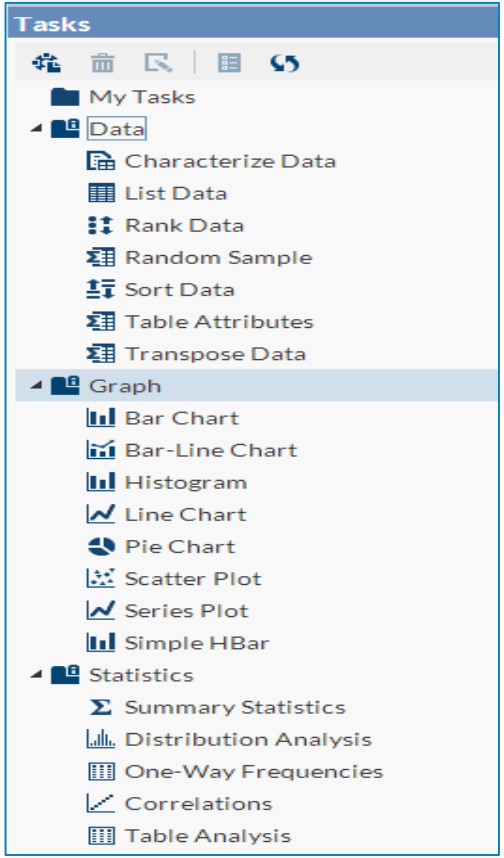
- EXPLORING DATA
  - Histogram
    - Add normal curve
    - Add kernel density estimate
    - Add inset statistics
  - Inset Statistics
- CHECKING FOR NORMALITY
  - Goodness-of-fit tests
  - Histogram with normal curve
    - Add inset statistics
  - Normal probability plot
    - Add inset statistics
  - Normal quantile-quantile plot
    - Add inset statistics
  - Inset Statistics
- FITTING DISTRIBUTIONS

CODE | LOG | RESULTS



Tasks Authoring

- SAS Studio allows for a flexible task framework to create tasks.



Delivery

Delivery - How you get SAS Studio

- Single-User. With Base SAS running on Microsoft Windows or with the SAS University Edition, you access SAS Studio through your local web browser on the same machine as SAS is installed.

- Multi-User. If you have SAS installed on a server, SAS Studio provides basic multi-user access to submit SAS code to that server. You can access SAS Studio from your browser (on any machine) by pointing to the server where SAS is running. You must have credentials to log into the SAS server machine. The content (files, data) you are allowed to access depends on the authority granted to you on the SAS server being accessed.

- Enterprise. If you have licensed Integration Technologies, you have full enterprise access to SAS Studio, which includes security & access control provided via SAS Metadata Server.

Delivery - Supported browsers

- SAS Studio is a pure HTML5 application that requires no browser plug-ins and supports all the major web browsers.

Microsoft Windows

- Microsoft Internet Explorer 9 or later
- Mozilla Firefox 21 or later
- Google Chrome 27 or later

Linux

- Mozilla Firefox 21 or later
- Google Chrome 27 or later

Mobile (iOS) = Beta

Apple Macintosh OS X

- Apple Safari 6.0 or later
- Mozilla Firefox 21 or later
- Google Chrome 27 or later

Computer is SUKSEL ▸ Win7x64 (C:) ▸ Users ▸ suksel ▸ .sasstudiohost_3.2

File Edit View Tools Help

Organize ▾ Open ▾ Share with ▾ Burn New folder

Favorites
- Desktop
- Downloads
- Recent Places
- Google Drive
- Dropbox
- My Cubby

Name	Date modified	Type	Size
access.lock	26/01/2015 15:38	LOCK File	0 KB
config.properties	18/02/2015 08:57	PROPERTIES File	1 KB
currentUrl.txt	18/02/2015 08:35	Text Document	1 KB
server.lock	26/01/2015 15:38	LOCK File	0 KB
serverlog.txt	18/02/2015 08:35	Text Document	7 KB
uuid.txt	18/02/2015 08:35	Text Document	1 KB

config.properties - Notepad

File Edit Format View Help

```
browserpath=C:\\Program Files (x86)\\Google\\Chrome\\Application\\chrome.exe
```

Delivery - How you get SAS Studio

Visit the SAS Studio product page (http://support.sas.com/software/products/sasstudio/index.html) for:

- Documentation
- Videos
- Papers

sas THE POWER TO KNOW.

Providing software solutions since 1976

support.sas.com | Knowledge Base | Support | Training & Books | Happenings | Store | Support Communities

KNOWLEDGE BASE / DOCUMENTATION

KNOWLEDGE BASE

- Products & Solutions
- System Requirements
- Install Center
- Third-Party Software Reference
- Documentation
  - What's New in SAS
  - Product Index A-Z
  - SAS Analytical Products 13.2
  - SAS Analytical Products 13.1
  - SAS 9.4
  - SAS Analytical Products 12.1
  - SAS 9.3
  - SAS 9.2
  - Earlier SAS Releases
- Papers
- Samples & SAS Notes
- Focus Areas

SAS® Studio

3.3 | 3.2 | 3.1

- Accessibility Features of SAS Studio 3.3 HTML
- Getting Started with Programming in SAS Studio 3.3 PDF | HTML
- Getting Started with Programming in SAS Studio 3.3 – non-English PDF files

  Select a language
- SAS Studio 3.3: Administrator's Guide PDF | HTML
- SAS Studio 3.3: Developer's Guide to Writing Custom Tasks PDF | HTML
- SAS Studio 3.3: System Requirements PDF
- SAS Studio 3.3: User's Guide PDF | HTML
- SAS Studio 3.3: User's Guide – non-English PDF files

  Select a language

Videos

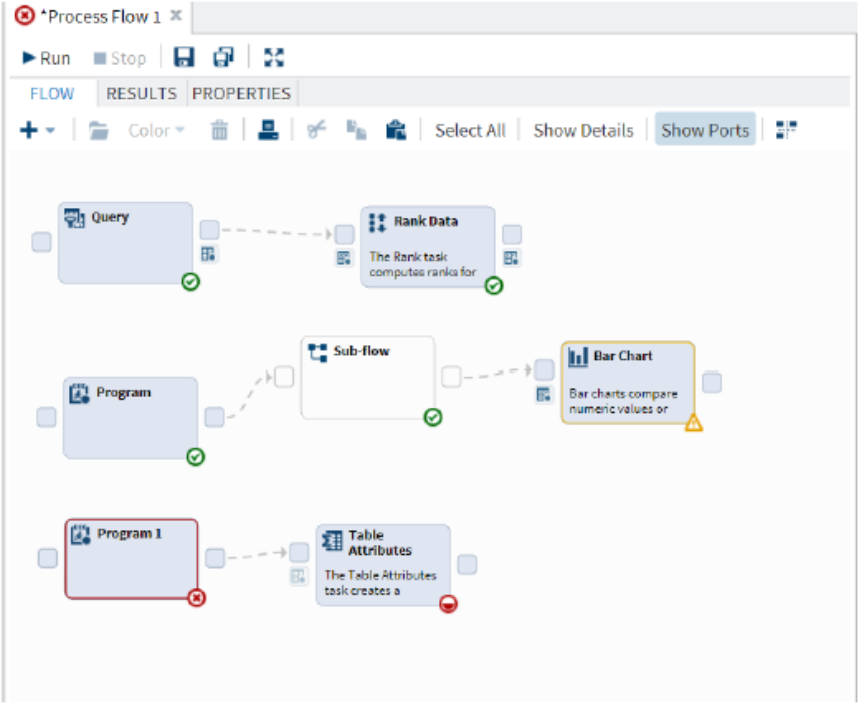
- SAS Studio: Video Library

  View these free videos to learn the basics of **SAS Studio**.

Programming References

- SAS 9.4 Programmer's Bookshelf

SAS Studio 3.3 – 15W08 – 17Feb2015



- Already on the Support Site
- SAS Studio 3.3 Users Guide
- SAS Studio 3.3 Administrators Guide
- SAS Studio 3.3 Developers Guide

SAS® FORUM
UNITED KINGDOM 2015

SAS® In-Memory Statistics

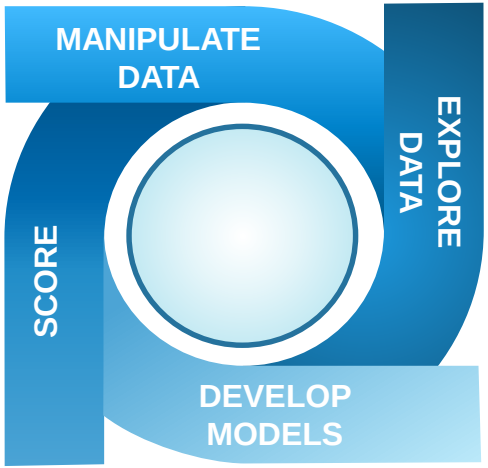
THE POWER TO KNOW.

SAS In-Memory Statistics

Provides a single *interactive* programming environment to perform

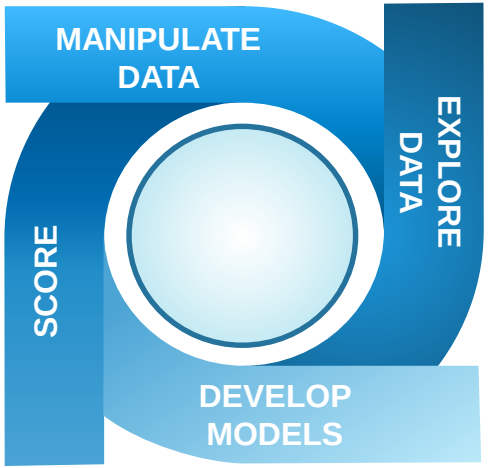
- analytical data preparation
- variable transformations
- exploratory analysis
- statistical modeling and machine learning
- integrated modeling comparison and scoring

- Takes advantage of distributed in-memory computing optimized for analytical workloads



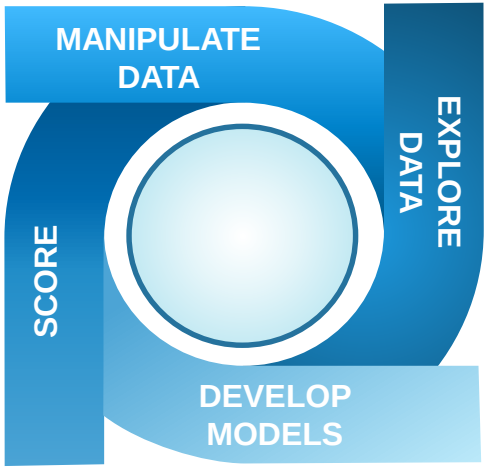
SAS In-Memory Statistics

- All required components for successful data science (analytics lifecycle) at your fingertips in one integrated programming environment
- Supports the experimental, iterative nature of analytical model development
- Faster time to insight - nimble and agile
- Analyze data sets of any size and any structure
- Explore and compare different modeling approaches to build the optimal champion model faster
- Better collaboration by empowering multiple users to work on the same data concurrently
- Maintain even distribution of workload among the nodes in a distributed computing environment



SAS In-Memory Statistics

- Utilize compute resources more efficiently
  - Load data once for entire exploration and modeling process
  - Multiple User Concurrency
- Improve data governance by eliminating data duplication
- Utilize distributed computing environment for high-performance data science
- An in-memory architecture eliminates costly data movement
- Easier maintenance with web-based clients



Capabilities of SAS® In-memory Statistics

Data Manipulation
- Aggregate
- Compute
- Update
- Append
- Set
- Schema
- DeleteRows
- DropTables
- PurgeTempTables

Data Exploration
- Boxplot
- Corr
- Crosstab
- Distinct
- Fetch
- Frequency
- Histogram
- KDE
- MDSummary
- Percentile
- Summary
- TopK

Predictive Modeling
- Decision Tree
- Forecast
- Gen Linear Model
- Linear Regression
- Logistic Regression
- Random Forests
- Neural Networks

Descriptive Modeling
- Association
- Path Analysis
- Clustering (k-means)
- Clustering (DBSCAN)

Text Analytics
- Parsing
- SVD
- Topic generation
- Document projection

Recommendation Systems
- Association
- Clustering
- kNN
- SVD
- Ensemble

Evaluation, Deployment
- Assess
  - Misclassification matrix
  - Lift, ROC, Concordance
- Score
- Training / Validation

Utilities
- Where
- GroupBy
- TableInfo, ColumnInfo, ServerInfo
- Partition, Balance
- Store, Replay, Free
- Table, Promote

HDFS I/O
- Sasiola
- Sashdat
- Anyfile Reader

SAS® Studio - SAS Programmer

Sign Out

Search

Folders

- Folder Shortcuts
- My Documents
- Desktop
- My Folders
- C:\
  - $Recycle.Bin
  - demodata
  - Documents and Settings
  - Intel
  - MSOCache
  - PerfLogs
  - Program Files
  - Program Files (x86)
  - ProgramData
  - Public
    - IMSTAT
      - Knowledge
        - data
          - data_analysis_usin
          - demo_short_with_
          - demo_short_with_
          - demo_short_with_
          - SAS® LASR™.pptn
          - Thumbs.db
    - VSTAT
  - Recovery
  - SASv9
  - Support
  - System Volume Information
  - temp
  - Users
  - V9setup
  - Windows

Tasks

Snippets

Libraries

File Shortcuts

*demo_short_with_ODS_10032013_v3_smpUSE.sas

CODE | LOG

```
proc imstat ;
  /*--- M E R G E ---*/
  table lasr.cardatap;
  schema carinfo (modelkey=modelkey / prefix=info) / mode=table;
run;
```

```
41 table lasr.carinfo;
42 fetch / from=1 to=5 format;
43 run;
44
45 /*--- E X P L O R E ---*/
46 table lasr.cardatap;
47 distinct _all_ / save=tab1;
48 run;
49 * the name list in LIST1 is automatically created;
50 * no need to type in variable names one by one;
51 store tab1(where="NDistinct le 5", 1)=list1;
52 run;
53 frequency &list1;
54 crosstab wheelTypeID*wheelType;
55 corr mmr:;
56 ods output frequency=work.freq1;
57 run;
58 table lasr.cardatap(tempnames=(avgOdo));
59 summary avgOdo / tempnames=avgOdo tempexpress="avgOdo = vehOdo / v
60     groupby=badbuy partition;
61 run;
62 quit;
63 ods graphics / height=4in width=12in;
64
65 proc sgpanel data=freq1;
66     panelby column / novarname columns=8 rows=1 uniscale=row;
67     vbar formattedvalue / group=column response=frequency datalabe
68 run;
69
70 proc imstat ;
71     /*--- M E R G E ---*/
72     table lasr.cardatap;
73     schema carinfo (modelkey=modelkey / prefix=info) / mode=table;
74 run;
75
76 /*--- C L E A N - U P ---*/
77 table lasr.&_templast_;
78 promote cardata2;
79 where info_nationality ne 'NULL';
80 run;
81 table lasr.cardata2;
82 where;
83 compute delta1 "delta1 = vehbcost - mmracquisitionretailaveragepri
84 compute delta2 "delta2 = vehbcost - mmrcurrentretailaverageprice
```

RESULTS

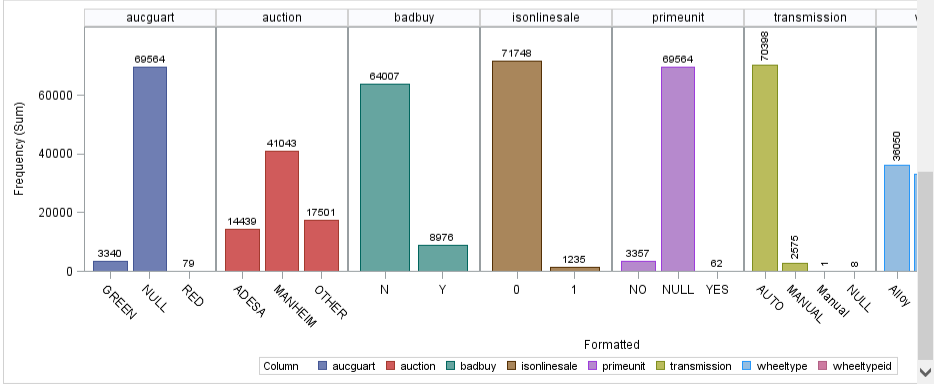
1	36050	0	0	0
2	0	33004	0	0
3	0	0	0	755
NULL	0	0	3169	0

Pairwise Correlations for Table WORK.CARDATAP

Column	Row	mmracquisitionauctionaveragepric	mmracquisitionauctioncleanprice	mmracquisitionretailaveragepriceprice	mmracquisitonretailcleanprice
mmracquisitionauctionaveragepric	1	1.0000	0.9903	0.9104	0.9102
mmracquisitionauctioncleanprice	2	0.9903	1.0000	0.9027	0.9182
mmracquisitionretailaveragepriceprice	3	0.9104	0.9027	1.0000	0.9902
mmracquisitonretailcleanprice	4	0.9102	0.9182	0.9902	1.0000
mmrcurrentauctionaveragepriceprice	5	0.9378	0.9236	0.8516	0.8469
mmrcurrentauctioncleanprice	6	0.9325	0.9321	0.8494	0.8557
mmrcurrentretailaveragepriceprice	7	0.8718	0.8607	0.9131	0.9032
mmrcurrentretailcleanprice	8	0.8710	0.8712	0.9061	0.9068

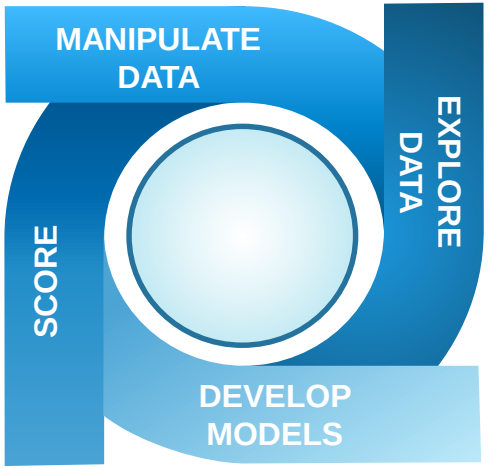
Summary Statistics for Table WORK.CARDATAP

auction	badbuy	Column	Min	Max	N	Sum	Mean	Std Dev.	Std Error	Coefficient of Variation	Number Missing
ADESA	N	avgOdo	2074.14	59767	12246	182129149	14873	5032.88	45.4799	33.8400	0
ADESA	Y	avgOdo	2961.14	38340	2193	29867437.5	13619	4298.17	91.7834	31.5591	0
MANHEIM	N	avgOdo	766.86	83729	36328	547241384	15064	5443.37	28.5593	36.1352	0
MANHEIM	Y	avgOdo	1870.13	47410	4715	62358270.7	13226	4292.05	62.5064	32.4528	0
OTHER	N	avgOdo	2118.89	48523	15433	256772573	16638	5807.60	46.7489	34.9059	0
OTHER	Y	avgOdo	1206.25	42628	2068	29776623.4	14399	4865.38	106.99	33.7903	0



Demonstration

SAS In-Memory Statistics



- **Faster** time to insight
- **Interactive** analytics
- **Seamless integration** of
  - *Data Access and Preparation*
  - *Data Exploration*
  - *Model Development*
  - *Model Scoring*
- Proven and **state-of-the-art** analytical and machine-learning algorithms
- **Scalable**
- **Concurrent** access to data loaded in-memory in a **multi-user** environment

SAS® FORUM
UNITED KINGDOM 2015

Thank you

sas THE POWER TO KNOW.

www.SAS.com