

BEST OF BOTH WORLDS: SAS AND OPEN SOURCE

SAS AS AN OPEN PLATFORM

OPEN ANALYTICS DAY
BUDAPEST BI FORUM, 26 OCT 2016

Tuba Islam
SAS Global Technology Practice

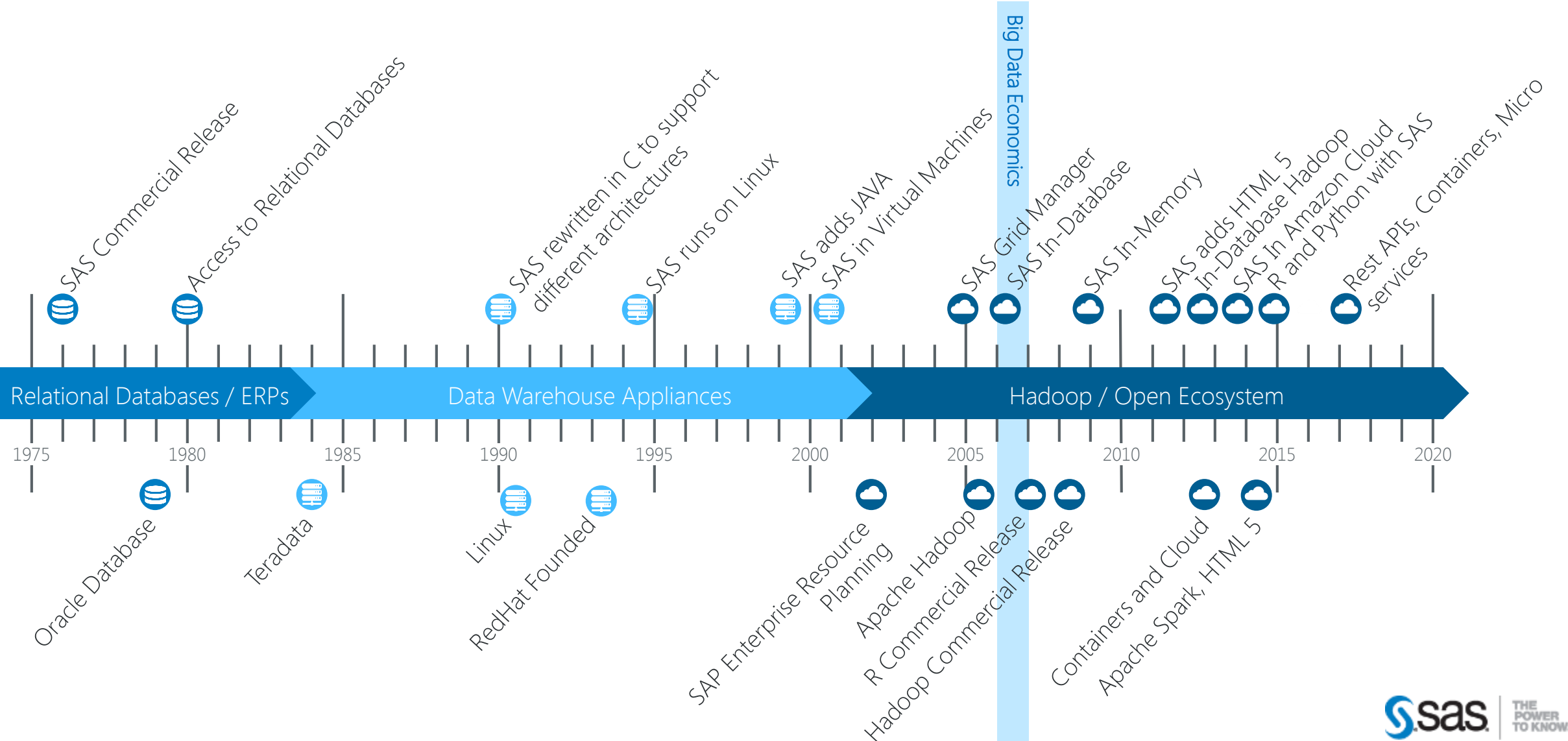


- SAS' openness
- How SAS embraces and extends open source
- Demonstration of open source and SAS integration

SAS – MARKET PERCEPTION



EVOLUTION OF ANALYTICS MARKET AND SAS' OPEN TIMELINE



NEW OPEN ARCHITECTURE: SAS® VIYA™

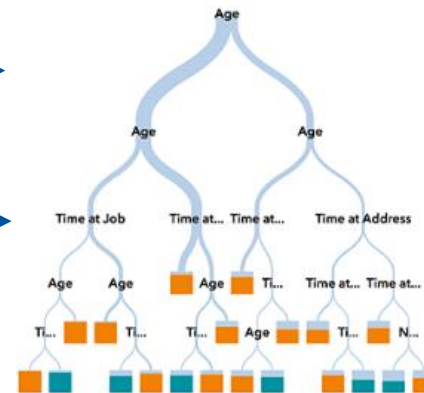
Visual Interfaces



Programming Interfaces



Future
Plans:



API Interfaces



BEING MORE OPEN: Share of Knowledge and Collaborative Work

The collage consists of four overlapping screenshots:

- Top Left:** A screenshot of the SAS Support Communities homepage. It features the SAS logo, navigation links (Products & Solutions, Industries, Support, Learn SAS, Partners, Community, About SAS), and a welcome message: "Welcome to SAS Support Communities". It encourages users to ask the community for help and join now.
- Top Right:** A screenshot of the SAS Blogs homepage. It has a blue header with "BLOG DIRECTORY", "SAS SOFTWARE", and "FOLLOW US". The main content area is titled "SAS Blogs" and "Connecting you to people, products & ideas from SAS". It features a "Featured Post" by Becky Graebe titled "'Data for good is in the DNA of this organization'".
- Bottom Left:** A screenshot of the GitHub page for SAS Software. It shows the organization's profile, repositories, and a list of people. Two repositories are highlighted: "sas-viya-programming" (Jupyter Notebook) and "unravl" (Java).
- Bottom Right:** A screenshot of the SAS Customer Support page. It features the SAS logo and tagline "THE POWER TO KNOW". It includes links to "support.sas.com", "Knowledge Base", "Support", "Training & Books", "Store", and "Sup". A section titled "SAS CUSTOMER SUPPORT /" lists various knowledge base topics like "Products & Solutions", "System Requirements", "Install Center", etc. There is also a promotional banner for a new SAS Support experience.

FREE SAS RESOURCES AVAILABLE ONLINE

FREE SOFTWARE, FREE TRAINING



SAS on Demand for Academics

42,956 total registered users

SAS University Edition

567,955 total downloads

SAS University Edition on AWS

2,638 total subscribers

8,621,212 total usage hours

- SAS' openness
- How SAS embraces and extends open source
- Demonstration of open source and SAS integration

OPEN DISCUSSIONS...



- Are all your models equal? *Or some are more equal?*
 - Spending **weeks** to build a fraud model vs a couple of **hours** for propensity.
 - Executing the churn detection **every day** vs a product launch **only once**.



- Analytical resources are scarce and they have diverse skills
 - How to enabling more people to develop many models easily and work collaboratively? How to allow data scientists to code in their language of choice?



- The speed of deployment brings value in return
 - How to move from development to production easily and efficiently with no risk of making errors? How to deal with massive deployments?

SAS as an Integrated Open Platform

- Joint development and deep integration with Hadoop
- Rapid prototyping and utilising all available resources
- Incorporating open source models into the analytical life cycle
- Providing governance and lineage of data and models
- Scalability and speed in deployment without any recoding



- SAS' opennness
- How SAS embraces and extends open source
- Demonstration of open source and SAS integration

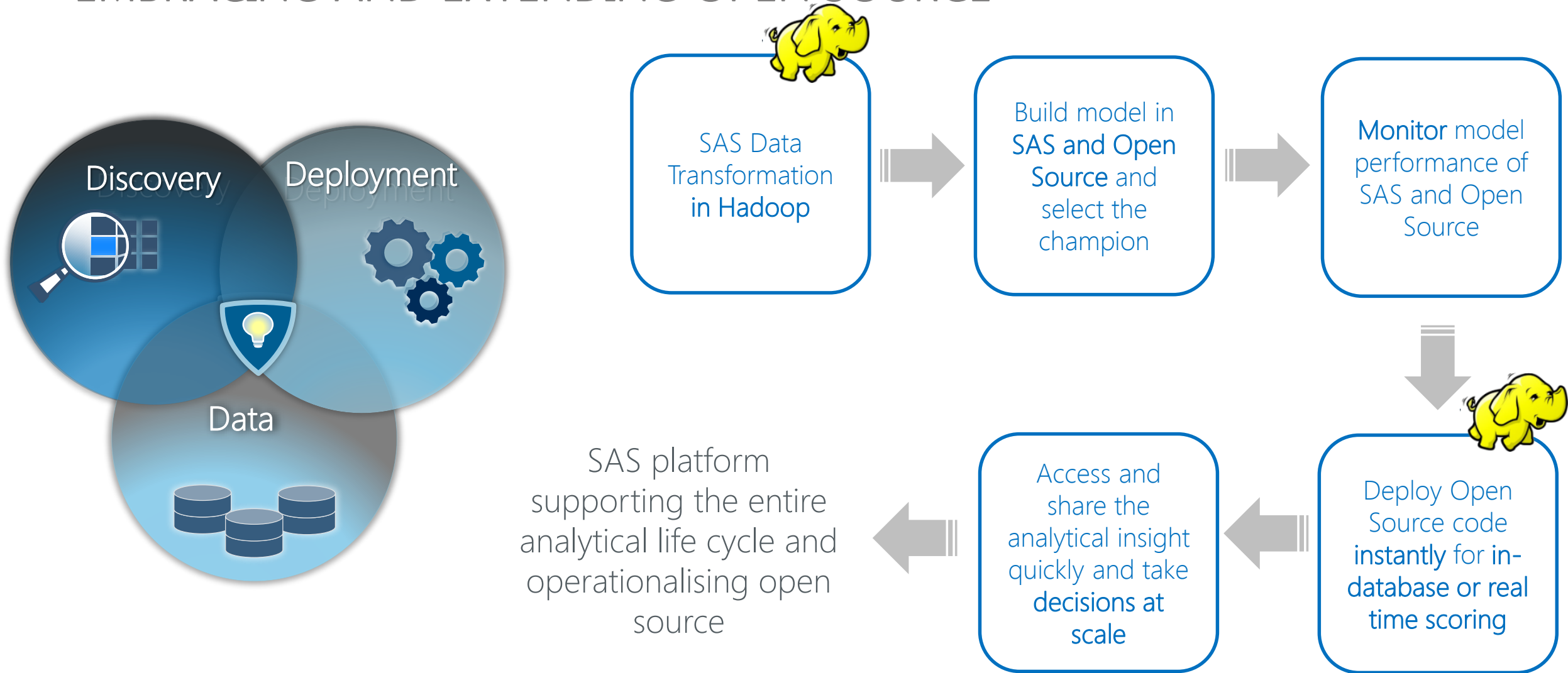


DEMO SCENARIO 1

OPERATIONALISING R MODELS WITH SAS

THE ANALYTICS LIFECYCLE – A Best Practice Open Journey

EMBRACING AND EXTENDING OPEN SOURCE



CENTRALISED MANAGEMENT ALL MODELS INCLUDING OPEN SOURCE

The screenshot displays the SAS Decision Manager web application. The left sidebar contains navigation links for My Tasks, Data, Business Rules, and Models. The Models section is expanded, showing a list of models. The main panel shows the 'HMEQ' model details, including tabs for Properties, Versions, Models, Variables, Scoring, Performance, Retrain, Reports, History, Attachments, and Comments. A table lists several models, with 'R Decision Tree' highlighted as the 'Champion' model. A callout box on the right lists three key features:

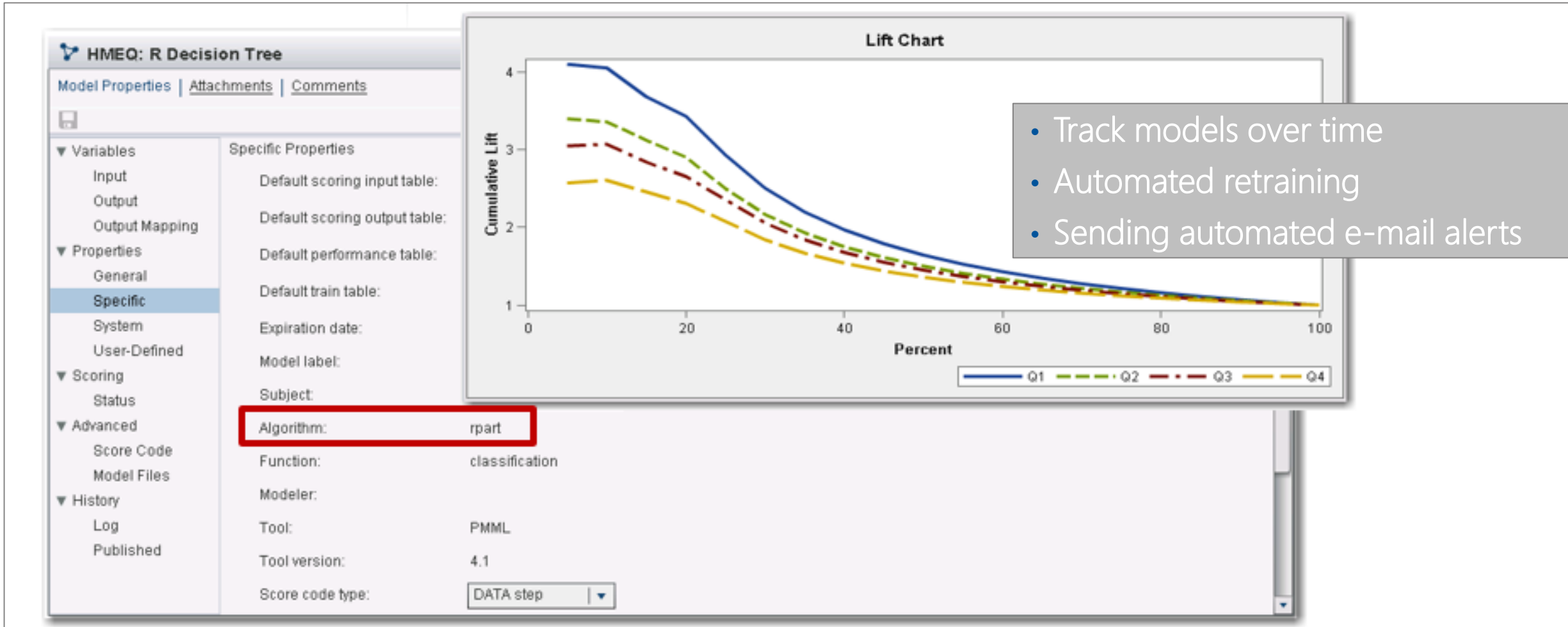
- Import models
- Manage all models in a single environment
- Declare the Champion

Name	Role	Version	Description	Model Type	Date Published	Owner
EM Regression		1.0		Classification	Oct 17, 2014 12:20 ...	sas
HPForest		1.0		Classification	Oct 17, 2014 01:52 ...	sas
Neural Net	Champion	1.0		Classification	Oct 17, 2014 12:37 ...	sas
R Decision Tree		1.0		Classification	Mar 2, 2015 05:07 PM	sasdemo
Regression	Challenger	1.0		Classification	Oct 17, 2014 12:38 ...	sas
STAT		1.0		Classification	Oct 17, 2014 12:40 ...	sas

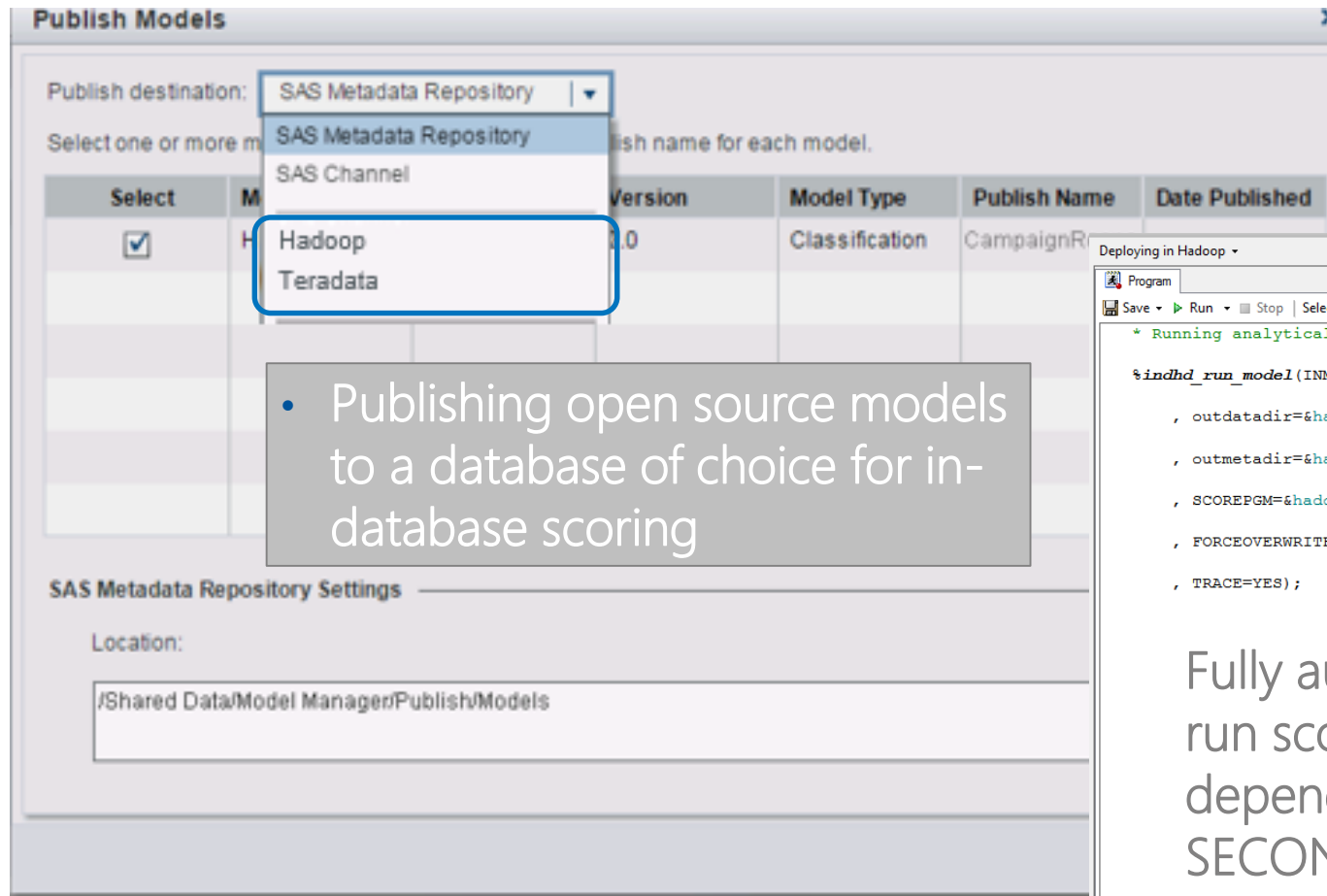
Alerts: 0 Total, 0 New

User: SAS Installer ID

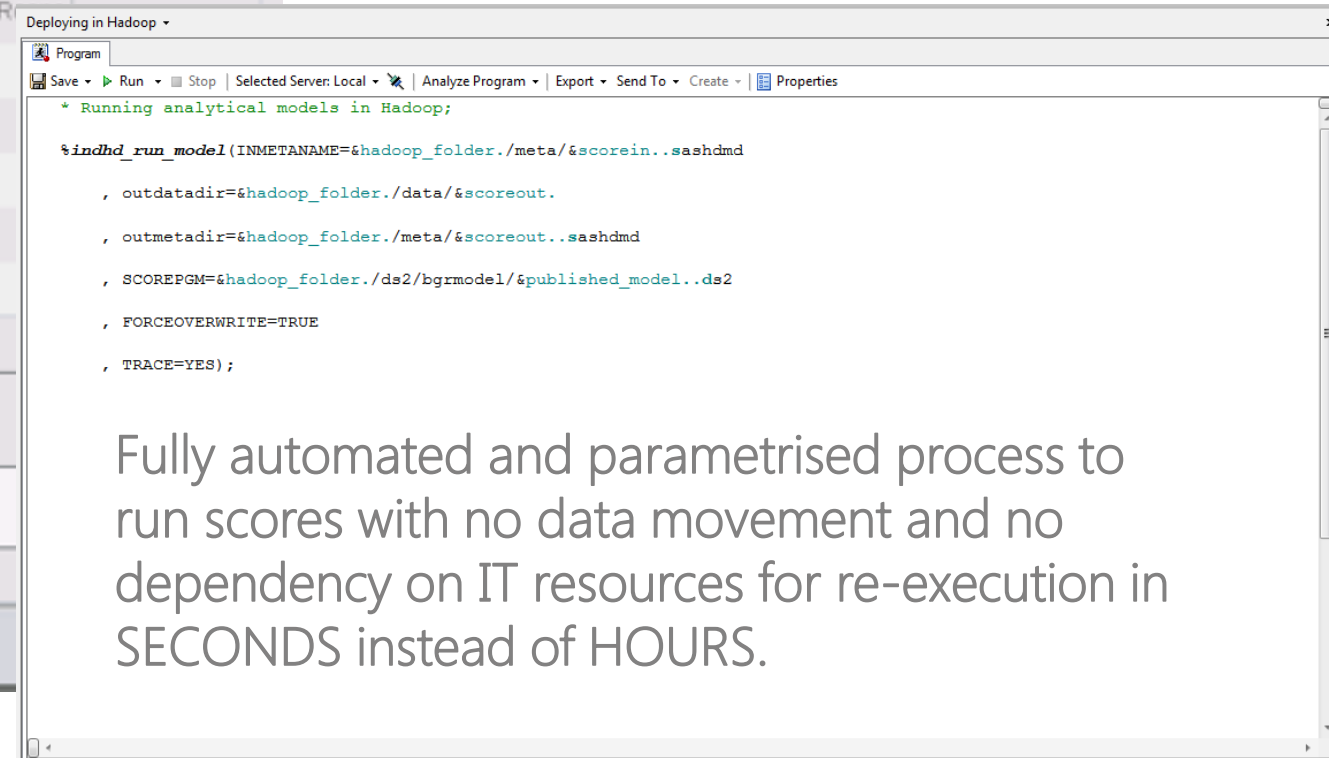
PERFORMANCE TRACKING AND AUTOMATED RETRAINING



PUBLISHING IN-DATABASE OR DEPLOYING IN REAL-TIME



- Publishing open source models to a database of choice for in-database scoring



Fully automated and parametrised process to run scores with no data movement and no dependency on IT resources for re-execution in SECONDS instead of HOURS.



DEMO SCENARIO 2

USING SAS AND PYTHON TOGETHER IN JUPYTER NOTEBOOK



Python is an interpreted, object-oriented, high-level programming language with modular libraries that support data extraction and processing.

The Jupyter Notebook is a web application that allows you to create and share documents that contain live code, equations, visualizations and explanatory text.

The notebook has support for over 40 languages including SAS and Python.

```

In [1]: libname c_data './bkbddm';
        filename s_hpc './bkbddm/zcta_hpcplus_score.sas';

        ***Load macro for varlist;
        %include './bkbddm/zcta_prerun.sas';

Lastly executed on Fri, Oct 30 2015 at 12:52 PM in 0 s

```

```

Out[1]:
5. ods listing close; ods htmlout options(bitmap_mode='inline') device=png; ods graphics on / output=png;
NOTE: Writing HTML Body File: ST0007
6. %*"""/;
7. libname c_data './bkbddm';
NOTE: LIBREF C_DATA was successfully assigned as follows:
Engine: V8
Physical Name: /home/jadawn/bkbddm
Filename s_hpc './bkbddm/zcta_hpcplus_score.sas';
8.
9.
10. ***Load macro for varlist;
11. %include './bkbddm/zcta_prerun.sas'/*"""/;
NOTE: Data file C:\DATA\ZCTA_SAS2.DAT is in a format that is native to another host, or the file encoding does not match the
session encoding. Cross Environment Data Access will be used, which might require additional CPU resources and might reduce
performance.
NOTE: PROCEDURE SQL used (Total process time):
      real time          0.08 seconds
      cpu time           0.01 seconds

101001 001w 101001 002w 101001 003w 101001 004w 101001 005w 101001 006w 101001 007w 101001 008w 101001 009w 101001 010w 101001 011w
101001 012w 101001 013w 101001 014w 101001 015w 101001 016w 101001 017w 101001 018w 101001 019w 101001 020w 101001 021w

```

Initial Exploration and Standardization of Variables

- Use the [CONTENTS](#) procedure to see the names, types, labels of variables
- Use the [STDIZE](#) procedure to create a new dataset name zcta_base in the WORK directory based on the zcta_base dataset found in our C_DATA library. Do the same for the ZCTA_new dataset

```

In [3]: proc contents data=c_data.zcta_base; run;
        proc stdize data=c_data.zcta_base out=zcta_base;
          var b1;
        run;
        proc stdize data=c_data.zcta_new out=zcta_new;
          var b1;
        run;

Lastly executed on Fri, Oct 30 2015 at

```

<https://github.com/ipython/ipython/wiki/IPython-kernels-for-other-languages>

IPython kernels for other languages

Bo edited this page 21 hours ago · 124 revisions

IPython/Jupyter kernels:

The Kernel Zero, is of course [IPython](#), which you can get through [ipykernel](#), and still comes (for now) as a dependency of [jupyter](#). The IPython kernel can be thought as a reference implementation, here are other available kernels:

Name	Jupyter/IPython Version	Language(s) Version	3rd party dependencies	Example Notebooks
sas_kernel	Jupyter 4.0	python >= 3.3	SAS 9.4 or higher	
IPyKernel	Jupyter 4.0	python 2.7, >= 3.3	pyzmq	
Julia		julia >= 0.3		
IHaskell		ghc >= 7.6		
IRuby		ruby >= 2.1		
IJavascript		nodejs >= 0.10		

The CONTENTS Procedure

Data Set Name
Member Type
Engine
Created
Last Modified

Using the ABC method

In this call we use the following between 2 and 18 clusters

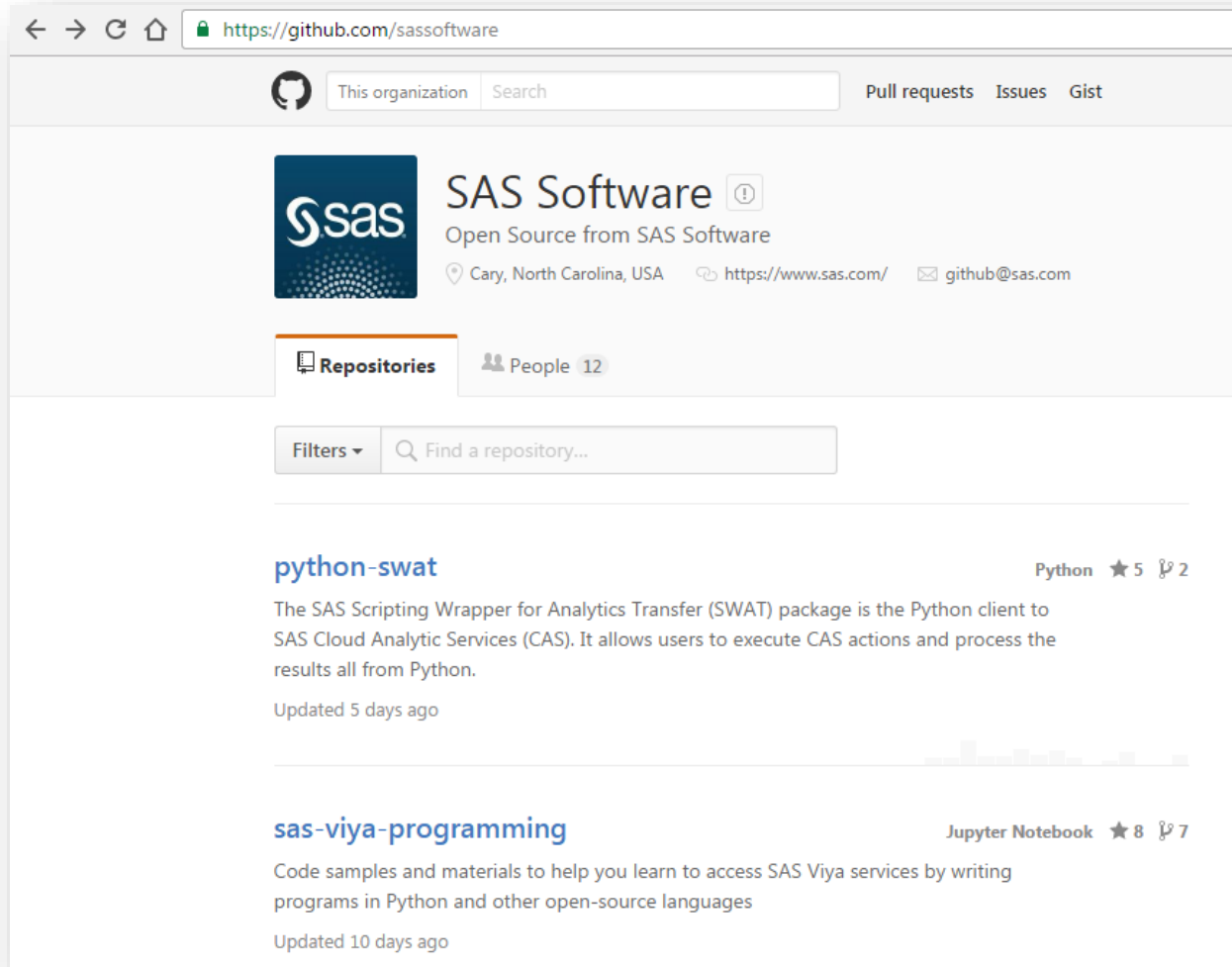
```

In [5]: ods select none;
        proc hpcplus;
          data=zcta_base;
          maxclusters= 18;
          maxiters= 100;
          seed= 9878;
          NOC= ABC(B= 1 min;
            *code files_hpc;
          id zcta;
          input b1 /level=;
          ods output ABCSta;
        run;
        proc sgplot;
          data= ABC;
          scatter x= k y= C;
          xaxis grid integer;
          yaxis label= 'ABC';
        run;

Lastly executed on Fri, Oct 30 2015 at

```

SAS on GitHub



SAS kernel for Jupyter Notebook:

https://github.com/sassoftware/sas_kernel

SAS Cell Magic for Python:

<https://github.com/sassoftware/saspy>

SAS Viya Open API for Python:

<https://github.com/sassoftware/python-swat>

jupyter Analysing Customer Reviews using SAS Text Mining in Python Last Checkpoint: a minute ago (unsaved changes)

File Edit View Insert Cell Kernel Help

Python [Root] ○










 Markdown ▼
 
 CellToolbar
 



Voice of the Customer: Processing online airline reviews

Loading SAS kernel and SAS Cell Magic for Jupyter Notebook: <https://github.com/sassoftware/saspy>

```
In [ ]: %load_ext saspy.sas_magic
```

Display available cell magics - %%SAS is one of them.

```
In [ ]: %lsmagic
```

Define a SAS library

```
In [ ]: %%SAS
libname mylib "/home/sasinst/data";
```

Import review data from web using python (for demo purpose only)

```
In [ ]: import requests
dls = "http://www.airlinequality.com/review-pages/latest-airline-reviews/all_reviews.xls"
resp = requests.get(dls)
output = open('airline_reviews.xls', 'wb')
```


A POTENTIAL BUSINESS APPLICATION FOR IMAGE PROCESSING

Severity analysis of car damage



Biggest challenge for image processing (as always) : Quality of the data

- Condition of the day (dark, bright)
- Resolution of the image (blurred, zoomed in, out)
- Angle of the capture (alignment, referencing)

Python Scripts | SAS Viya Image Process |

localhost8888/notebooks/Python%20Scripts/SAS%20Viya%20Image%20Processing%20Demo%20for%20Car%20Damage.ipynb

jupyter SAS Viya Image Processing Demo for Car Damage Last Checkpoint: 20 minutes ago (unsaved changes)

File Edit View Insert Cell Kernel Help Python [Root]

Apply noise reduction and binarization


```
In [35]: r = s.image.processImages(casout={'name': 'resultingImages', 'replace': True},
    imageTable={'name': 'inputTable_resized'},
    imageFunctions=[
        {'options': {'functiontype': 'CONVERT_COLOR'}}, #change color space

        {'options': {'functiontype': 'BILATERAL_FILTER', #noise reduction
                     'diameter': 13, 'sigmacolor': 30, 'sigmaspace': 30}},

        {'options': {'functiontype': 'THRESHOLD', #image binarization
                     'type': 'OTSU', 'value': 125}}
    ])
print(r)
outTable = s.CASTable('resultingImages')
imageShow(s, outTable, 0, rimage)
```

NOTE: Table INPUTTABLE_RESIZED contains compressed images.
NOTE: 4 out of 4 images were processed successfully and saved as compressed images to the Cloud Analytic Services table resulti
rgImages.
+ Elapsed: 0.408s, user: 1.44s, sys: 0.056s, mem: 195mb

Calling SAS Viya image processing actions from python



CUSTOMER SUCCESS STORIES

MUNICH RE INSURANCE

“With these new technologies (Hadoop and SAS Text Analytics), we are considerably enhancing our global ability to combine our customers’ data with our own findings and expert knowledge.”

Chief Executive Officer

A MAJOR INTERNATIONAL BANK

“SAS has outpaced any vendor in its speed and breadth... in moving with us to open source, they have turned on a dime on delivering for us. SAS is data-agnostic, making analyzing data coming in from anywhere easy. Management can now visualize the results of analytics on our customer data in seconds.”

Chief Data Officer

NORTH AMERICAN TELCO SERVICE PROVIDER

“The proven capability, security and the continuity of SAS is its strength within our organization. Open source has its place, and will continue to work in tandem with our SAS implementation.”

Principal Analyst, Business Systems

SAS – MARKET PERCEPTION



BEST OF BOTH WORLDS: SAS AND OPEN SOURCE

SAS AS AN OPEN PLATFORM

✓ Quick Access to Intelligence

- Enabling more users
- Operationalising analytics

✓ Improved Productivity

- Freedom of choice of language
- Utilising all resources
- Collaborative environment

✓ Governance and Lineage

- Managing an enterprise ecosystem

THANK YOU FOR LISTENING

Tuba Islam

tuba.islam@sas.com

Global Technology Practice

