Thomas LeBlanc

R and Python in SQL Server 2017

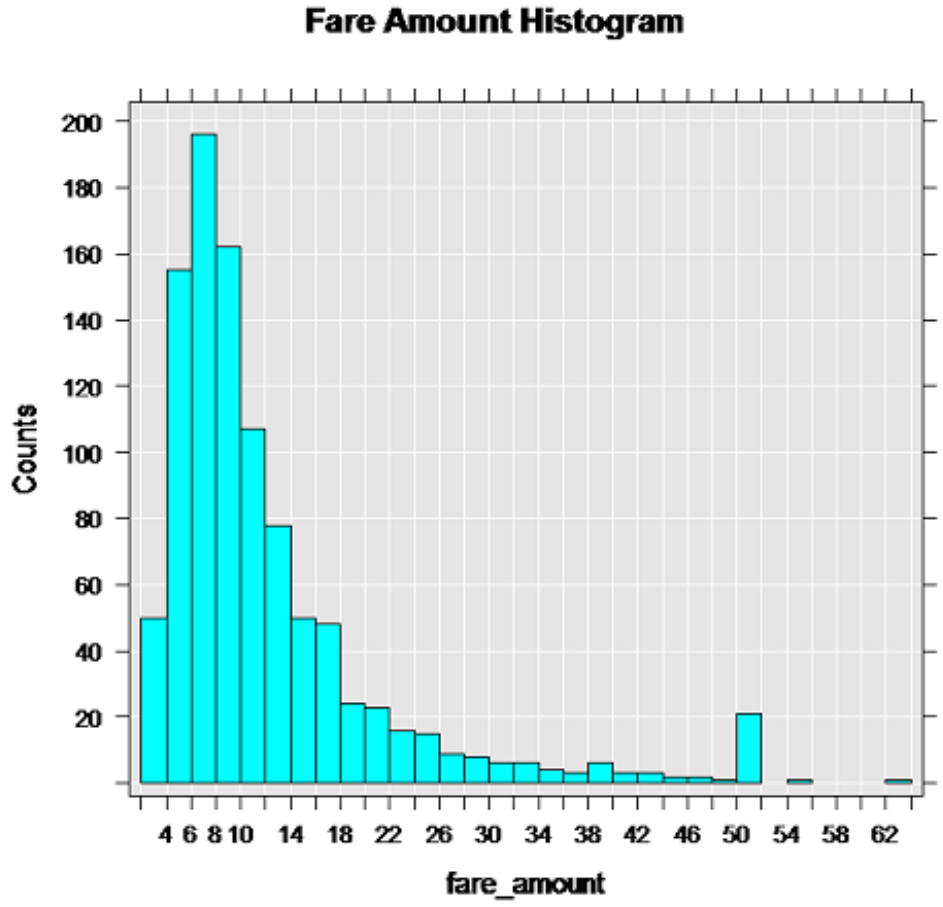
- **About Me…**
  - DW/BI Architect
  - Database Normalization “Nut”
    - Crazy about Dimensions
  - Retired Developer
  - TheSmilingDBA – twitter, blog & gmail
  - PASS volunteer - Chair of Excel BI VC

What are we talking about?

- What are these options
  - “Data Science” and “Machine Learning” options
- Installation (2016 R, 2017 R & Python)
- Configuration
- R Example
- Python Example
- Why?

What are these Options?

- Data Science
  - Analysis/Predictions
- Machine Learning
  - Training an algorithm
  - Predict from model
- Trends
  - Business Intelligence
  - Analytics
- Separate Process
  - Manage Memory
- Data is close to ML language



Installation

No Internet Access – Download separately

Installation (In-Database)

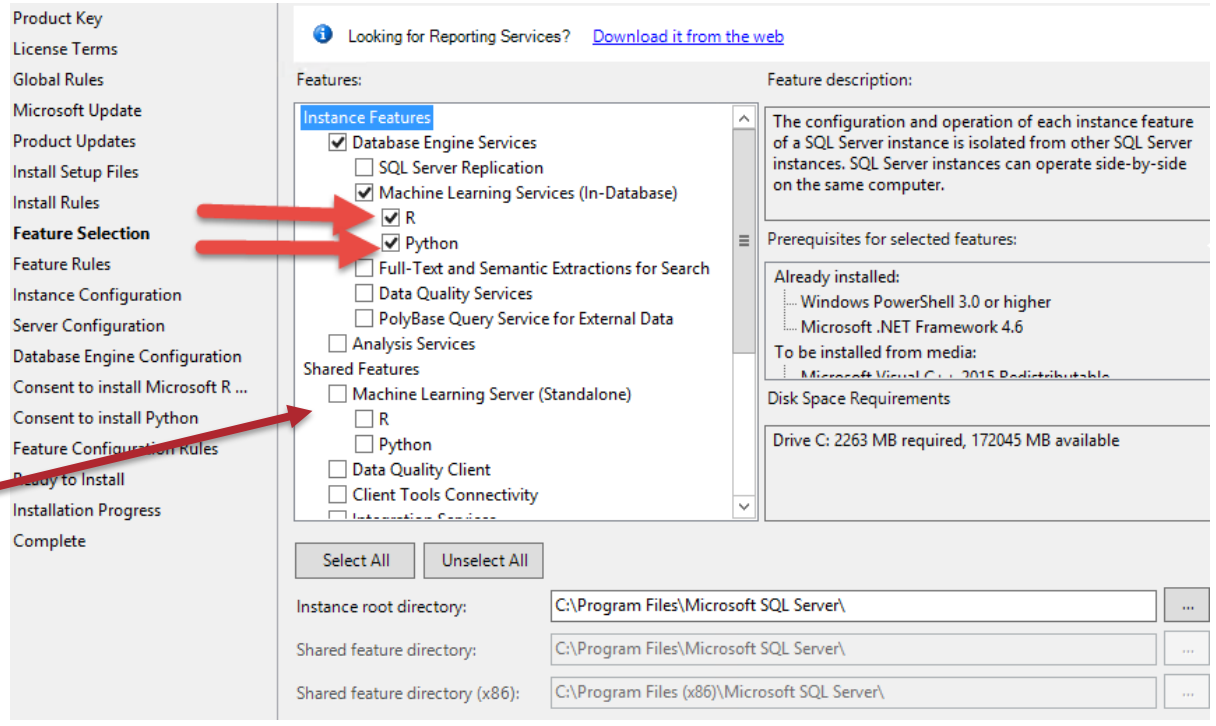
For an Instance

2 Options

R

Python

Standalone

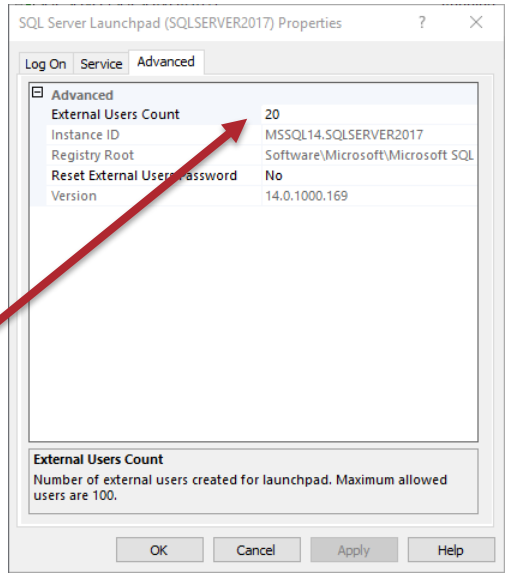


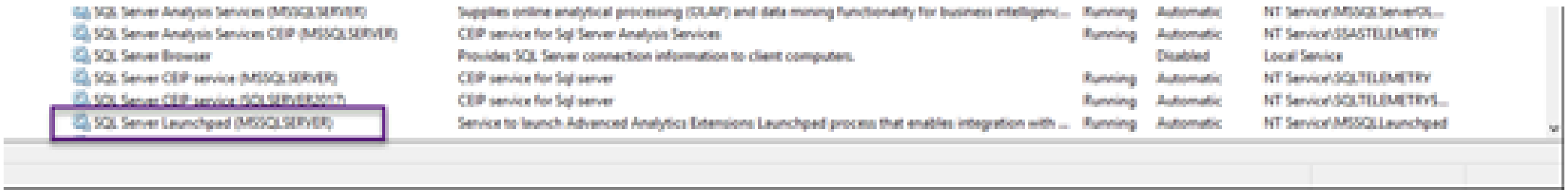
Configuration

LaunchPad has to be running

Addition Users

Created Locally to run scripts





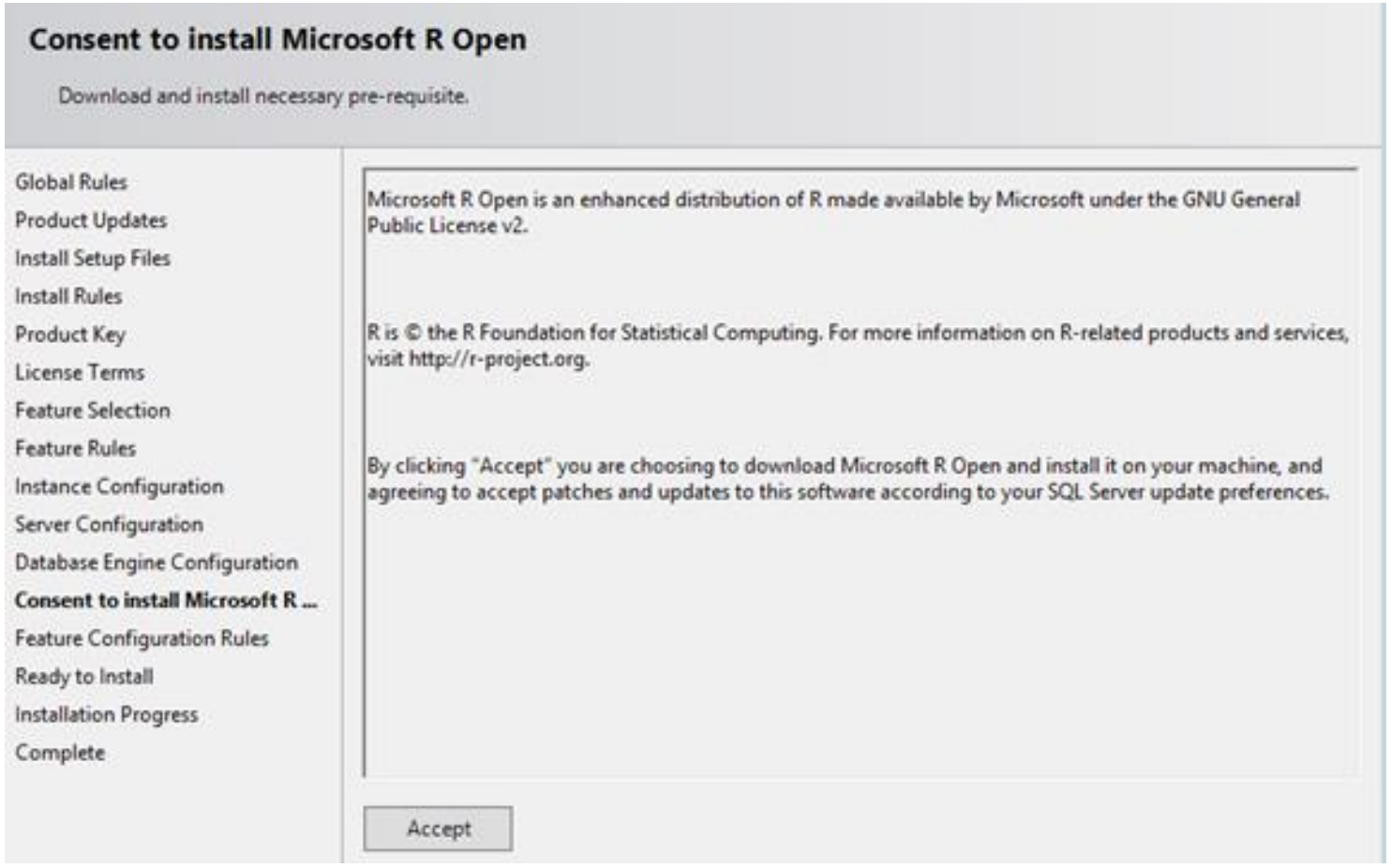
Configuration

Problems?

- Enable R Accounts to Connect
- Install Additional Network Libraries
- Enable remote code execution
- Restart instance (not server)
- Windows Logs has some messages to help

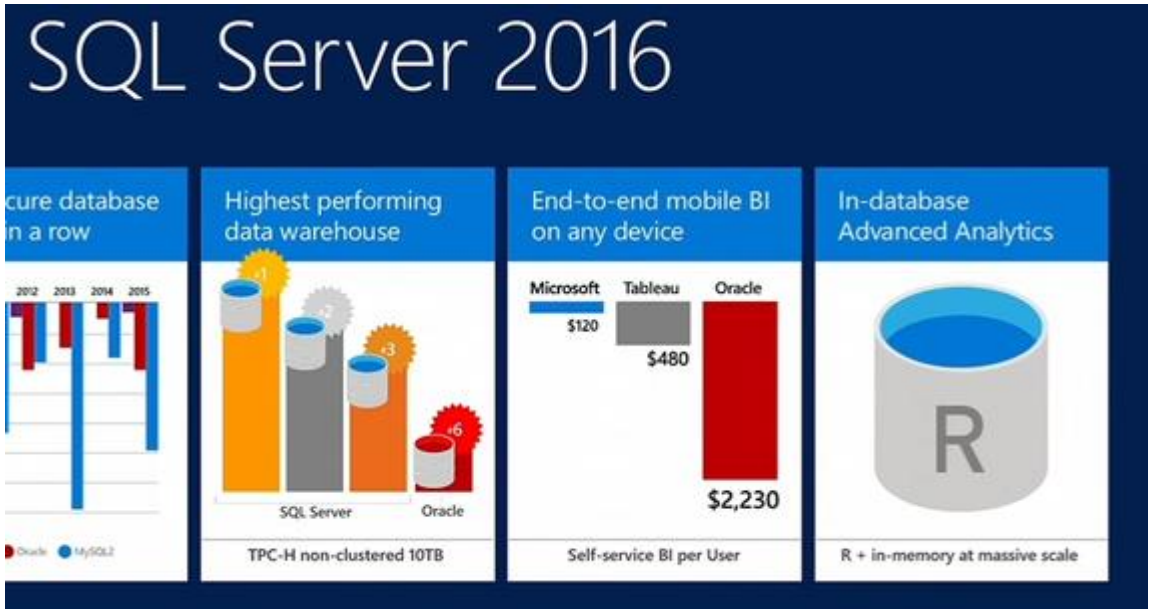
Search: R Services Installation and Upgrade FAQ

Demo - installation



R & Python

- Open Source
- Statistical Programming
  - Statistical Computing
  - Data Mining
  - Machine Learning
- Graphics
  - Data Visualizations



R & Python

Services has to be installed

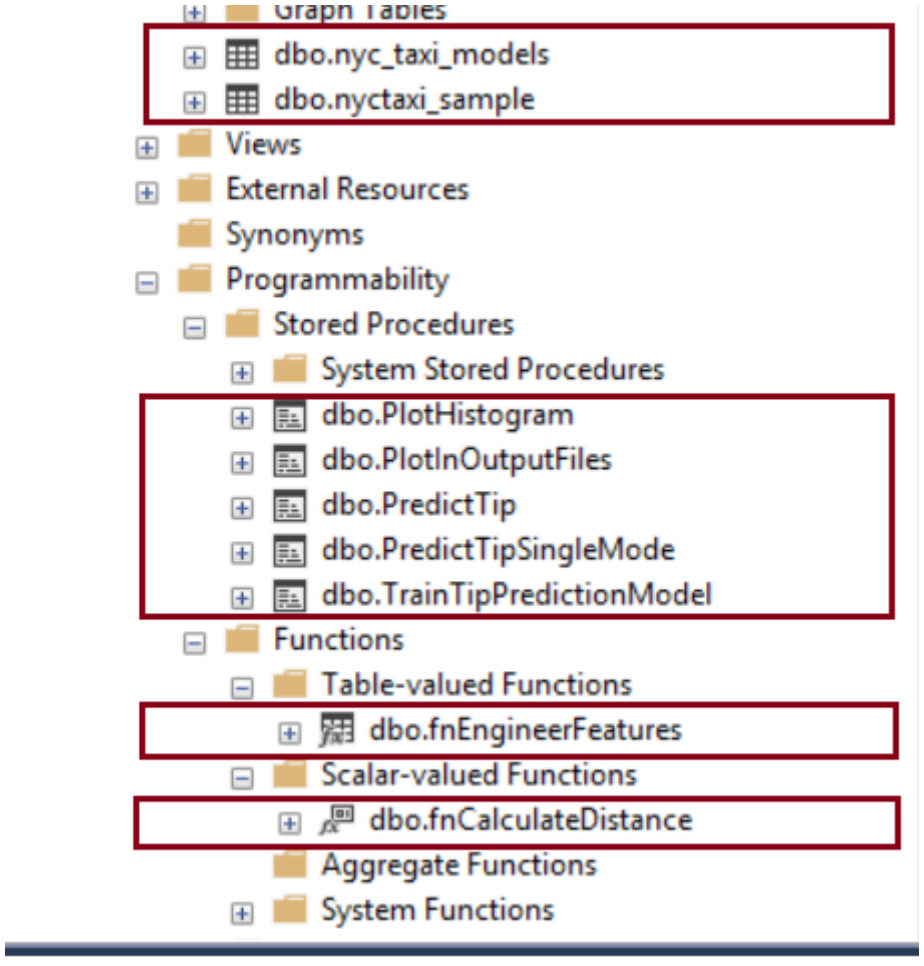
- Integrated with SQL Server

Use SSMS

- Mix with T-SQL
- Data and language together

Free libraries and packages

- FREE!!!



Configure

SP_CONFIGURE ‘external scripts enabled’:
GO
RECONFIGURE;
GO

Restart Instance

Make sure Launchpad is running

R Script syntax

```SQL
EXEC sp_execute_external_script
  @language =N'R',
  @script=N'OutputDataSet<-InputDataSet',
  @input_data_1 =N'SELECT 1 AS hello'
  WITH RESULT SETS (([Hello World] int));
GO
```

@Language (R or Python)

@script (R Script)

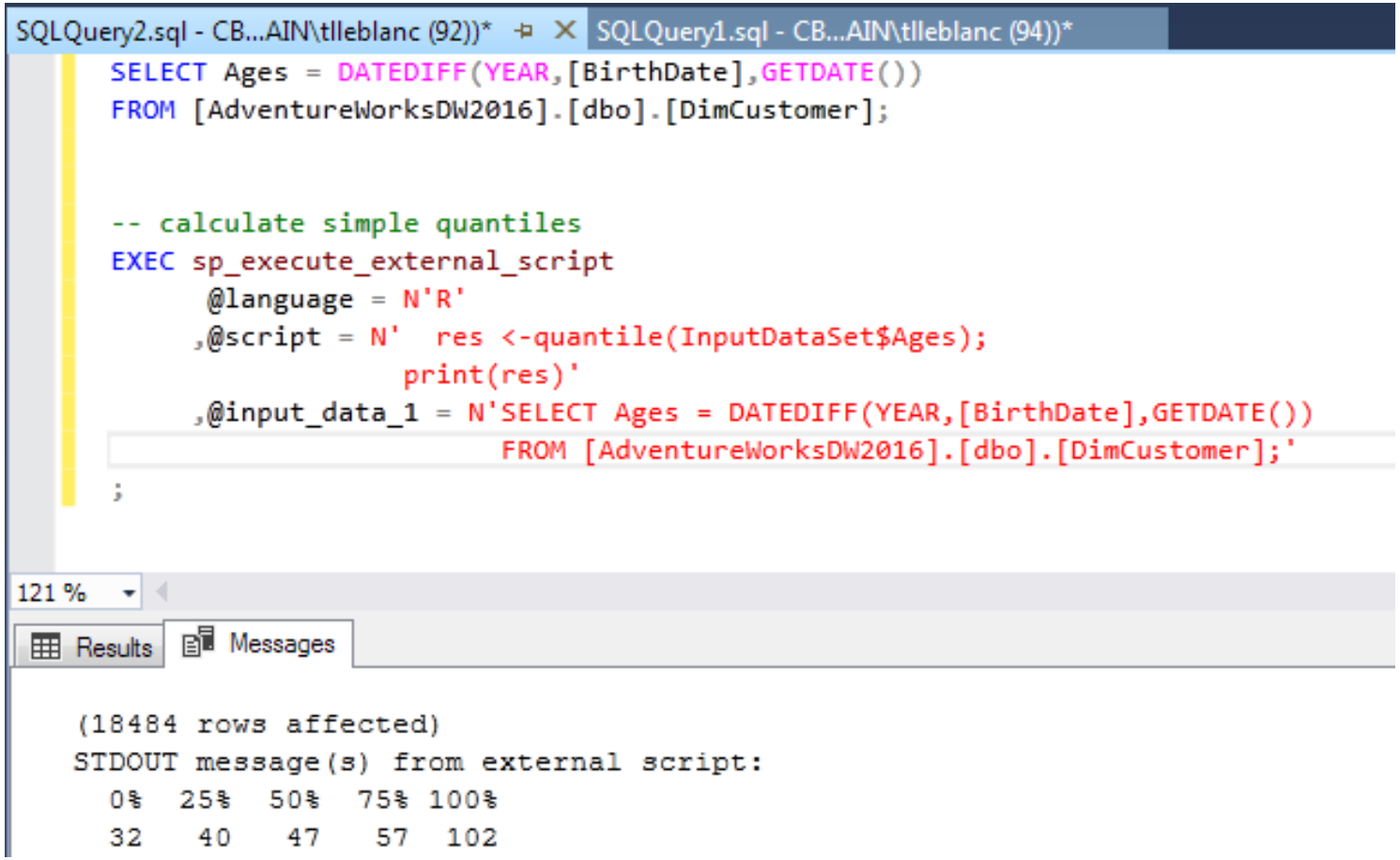
@input_data_1 – data imputed by query, passed to R

WITH RESULTS SETS – schema of returned data

R is case sensitive

R Script

Demo



Rules

Variable follow rules for SQL identifiers

Input and Output parameters must be in order

- @input_data_1 and @output_data_1
- Before _name parameters

Only one input data set

Only one output data set

WITH RESULTS SET has to specify the schmea

Developing R Code

Use R Studio or R Tools for VS

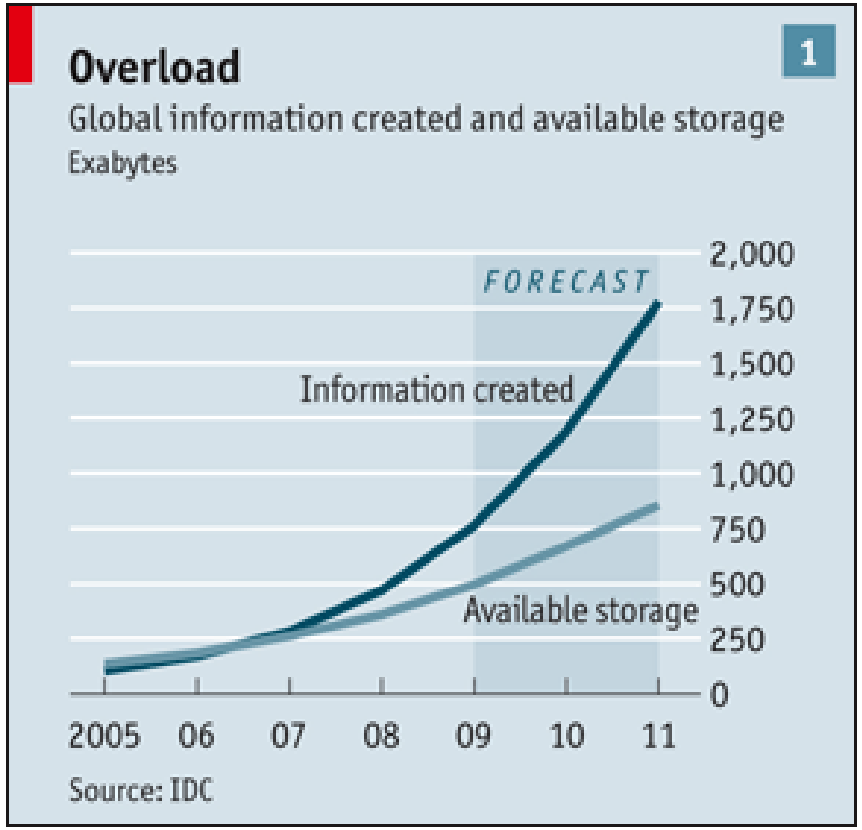
Not great messages from SPs

Learn R First

- use small data set

Move to SP in SQL Server

- use large data sets



Python syntax

```SQL
execute sp_execute_external_script
@language = N'Python',
@script = N'
a = 1
b = 2
c = a/b
s = pandas.Series(c, index =["simple math example 1", "simple math example 2"])
print(s)
'
```

@Language (R or Python)
@script (Python Script)
@input_data_1 – data returned by query
WITH RESULTS SETS – schema of returned data
Python is case sensitive, indention & syntax sensitive ☺

Python syntax

Output

- return as data frame
- even single value

Data Frame – Multiple columns

Column – List like object (Series)

Cell of data frame – called by index

Uses pandas library

Use tables to store data or model and results

Python Script

Demo

SQLQuery2.sql - CB...AIN\tlleblanc (92))* | SQLQuery1.sql - CB...AIN\tlleblanc (94))*

```
SELECT Ages = DATEDIFF(YEAR,[BirthDate],GETDATE())
FROM [AdventureWorksDW2016].[dbo].[DimCustomer];


-- calculate simple quantiles
EXEC sp_execute_external_script
      @language = N'R'
     ,@script = N'  res <-quantile(InputDataSet$Ages);
                  print(res)'
     ,@input_data_1 = N'SELECT Ages = DATEDIFF(YEAR,[BirthDate],GETDATE())
                        FROM [AdventureWorksDW2016].[dbo].[DimCustomer];'
;
```

121 %

Results | Messages

```
(18484 rows affected)
STDOUT message(s) from external script:
  0%  25%  50%  75% 100%
  32   40   47   57  102
```

Developing Python Code

Use VS Code

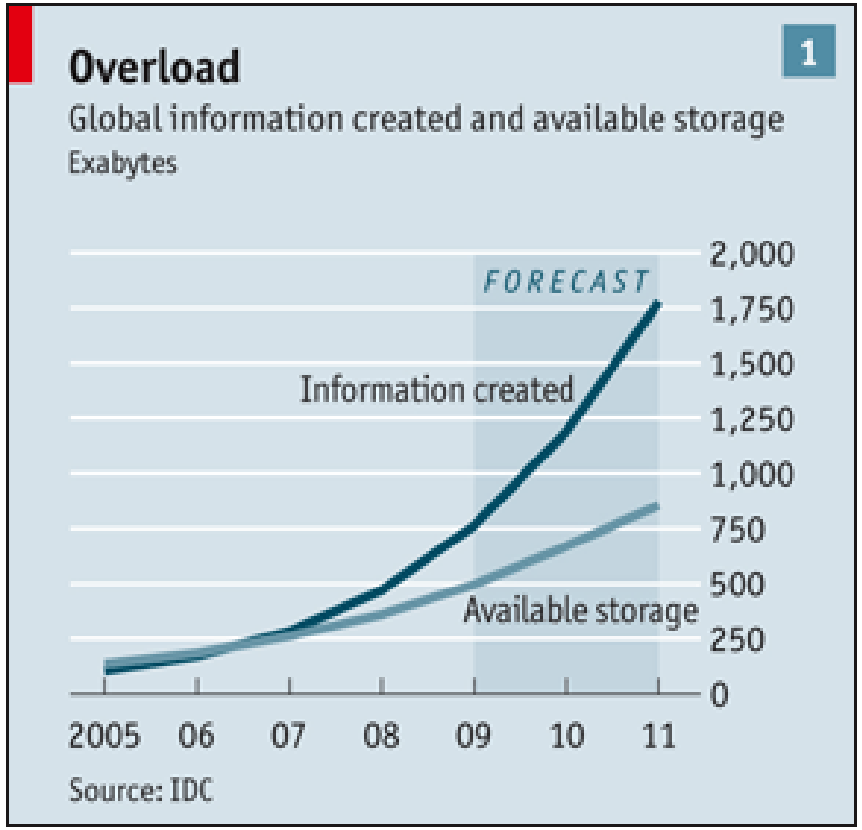
Not great messages from SPs

Learn Python First

- use small data set

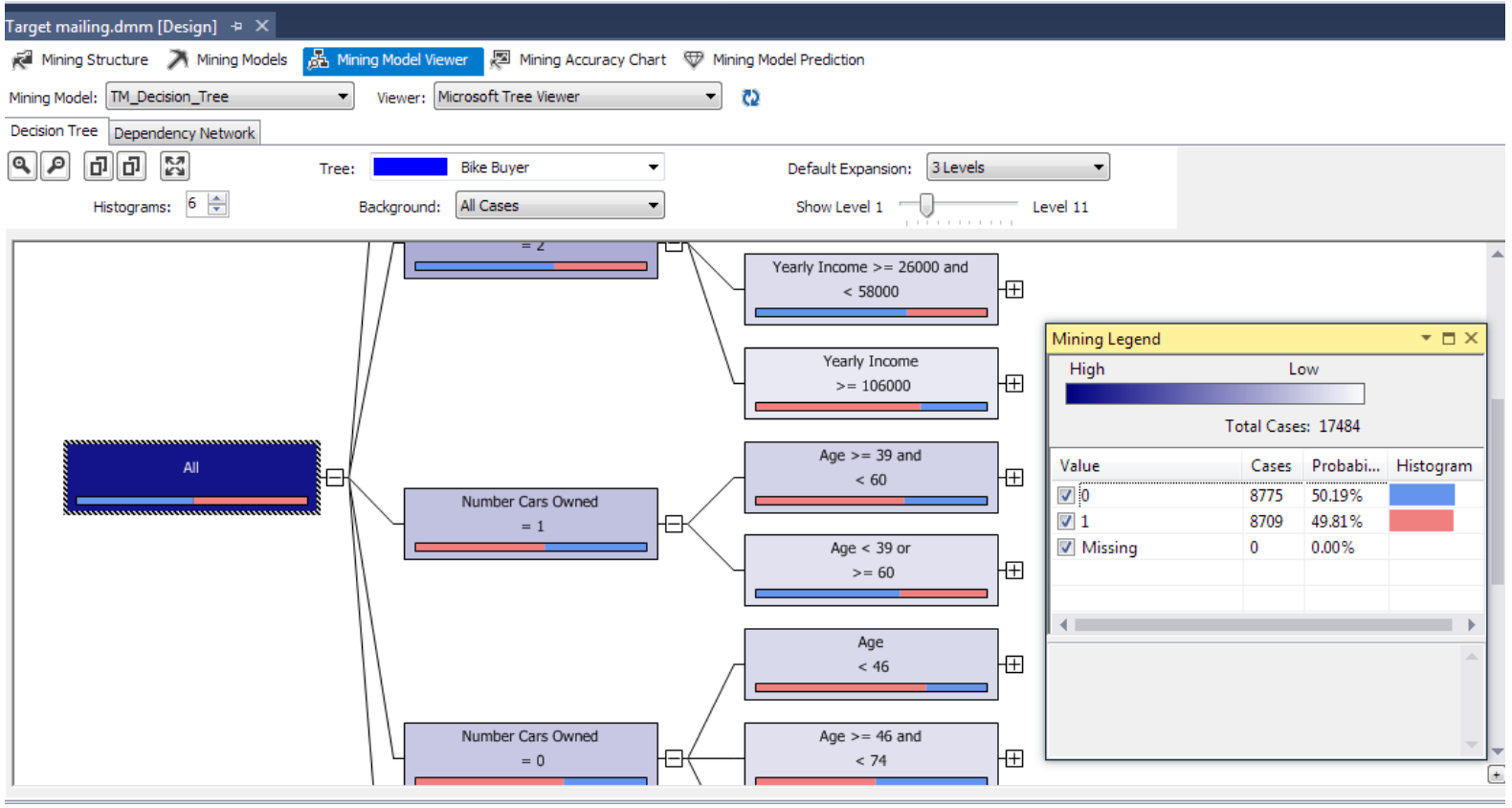
Move to SP in SQL Server

- use large data sets



Why?

Let's Look at Data Mining in SSAS



Where?

2 Sites

Search SQL Server 2017 R (Python) Tutorial

Quickstart: "Hello world" R script in SQL Server

07/14/2018 • 3 minutes to read • Contributors all

APPLIES TO: SQL Server (Windows only) Azure SQL Database Azure SQL Data Warehouse Parallel Data Warehouse

SQL Server includes R language feature support for in-database analytics on resident SQL Server data. You can use open-source R functions, third-party packages, and built-in Microsoft R packages for predictive analytics at scale.

In this quickstart, you learn key concepts by running a "Hello World" R script inT-SQL, with an introduction to the **sp_execute_external_script** system stored procedure. R script execution is through stored procedures. You can either use the sp_execute_external_script stored procedure and pass R script in as an input parameter as demonstrated in this quickstart, or wrap R script in a custom stored procedure.

Prerequisites

Thank You

Please visit Sponsors!!!

Thomas LeBlanc

Twitter @TheSmilingDBA
TheSmilingDBA@blogspot.com
TheSmilingDBA@gmail.com