Campaign Management Template with SQL Server 2016 R Services – Introduction

In this template, we demonstrate how to develop and deploy end-to-end Campaign Management solutions with [SQL Server 2016 R Services](https://msdn.microsoft.com/en-us/library/mt674876.aspx).

We have used machine learning techniques like Random Forest & Gradient Boosting to model for Campaign Responses. We model the conversion rate for different Channel-day (of week)-time (of day) as a binary variable.

To run this solution yourself, use the [Solution How-To Guide](Solution%20How-To%20Guide.docx) to guide you through the setup and various paths you can follow.

The solution uses the following files:

**Copy of Input Datasets**

|  |  |
| --- | --- |
| File | Description |
| .\Data\campaign\_detail.csv | Campaign Metadata |
| .\Data\cm\_ad.csv | Analytical Dataset |
| .\Data\market\_touchdown.csv | Historical Campaign data including lead responses |
| .\Data\product.csv | Product Metadata |
| .\Data\lead\_demography.csv | Demographic data of the leads |

**Model Development in R**

|  |  |
| --- | --- |
| File | Description |
| .\R\input\_data.r | Simulates the 4 input datasets |
| .\R\feature\_engineering.r | Performs Feature Engg. On the input datasets |
| .\R\RF\_model\_train.r | Builds Random Forest Model |
| .\R\GBM\_model\_train.r | Builds Gradient Boosting Model |

**Model Development in SQL Server 2016 R Services**

|  |  |
| --- | --- |
| File | Description |
| .\SQLR\step0\_table\_structure\_input\_data.sql | SQL Script to create schema of the databases if the user wants to import the datasets instead of simulating them |
| .\SQLR\step1(a)\_campaign\_detail.sql | SQLR Script to create the campaign detail dataset |
| .\SQLR\step1(b)\_product.sql | SQLR Script to create Product dataset |
| .\SQLR\step1(c)\_lead\_demography.sql | SQLR Script to create Lead Demography dataset |
| .\SQLR\step1(d)\_market\_touchdown.sql | SQLR Script to create Market Touchdown dataset |
| .\SQLR\step2(a)\_preprocessing\_market\_touchdown.sql | Outliers in the market touchdown dataset are treated |
| .\SQLR\step2(b)\_preprocessing\_lead\_demography.sql | Missing values in the Lead demography table are treated |
| .\SQLR\step3\_feature\_engineering\_market\_touchdown.sql | Market touchdown dataset is aggregated and variables like #Emails, #Calls and #SMS are created |
| .\SQLR\step4\_ad\_creation.sql | SQLR Script to create Analytical Dataset and split it into Train and Test |
| .\SQLR\Step5(a)\_model\_train\_rf.sql | SQLR Script build Random Forest |
| .\SQLR\Step5(b)\_model\_train\_rf.sql | SQLR Script build Gradient Boosting Model |
| .\SQLR\step6\_models\_comparision.sql | SQLR Script to compute the model statistics of both the models |
| .\SQLR\step7\_scoring\_leads.sql | SQLR Script to select the champion model and score the Analytical dataset on the champion model |

**Automation with PowerShell**

|  |  |
| --- | --- |
| File | Description |
| .\SQLR\Analytical Dataset Creation.ps1 | Creates the Analytical/Scoring dataset |
| .\SQLR\Model Development.ps1 | Trains the Random Forest and Gradient Boosting Models |
| .\SQLR\Scoring.ps1 | Identifies the Champion Model and scores the Analytical/Scoring dataset |

In this template with SQL Server R Services, three versions of the implementation module have been showcased:

1. **Model Development in R IDE**. Run the code in R IDE (e.g., RStudio, R Tools for Visual Studio).
2. **Model Development in SQL**. Run the code in SQL Server using SQLR scripts
3. **Automation in PowerShell**. Run the PowerShell scripts which automates the Model Development and Scoring Process

The following is the directory structure for this template:

* **Data** This contains the copy of the simulated input data
* **R** This contains the R codes to simulate the input datasets, create the analytical datasets, train the models, identify champion model and score the analytical/scorings dataset
* **SQLR** This contains the SQLR codes to simulate the input datasets, create the analytical datasets, train the models, identify champion model and score the analytical/scorings dataset. It also contains PowerShell scripts automate the entire process.

To run this solution yourself, start with the [Solution How-To Guide](Solution%20How-To%20Guide.docx) which will guide you through the setup and various paths you can follow.