Campaign Management

Execution with R Scripts

For the purposes of a quick demo, we can use a small dataset. To create a smaller dataset follow the steps in [**Data Setup**](Data%20Setup.docx)**.docx**.

Make sure you have set up your SQL Server and ODBC connection between SQL and PowerBI by following the instructions in [**START HERE.docx**](START%20HERE.docx). Then proceed with the steps below to run the solution template using the R files.

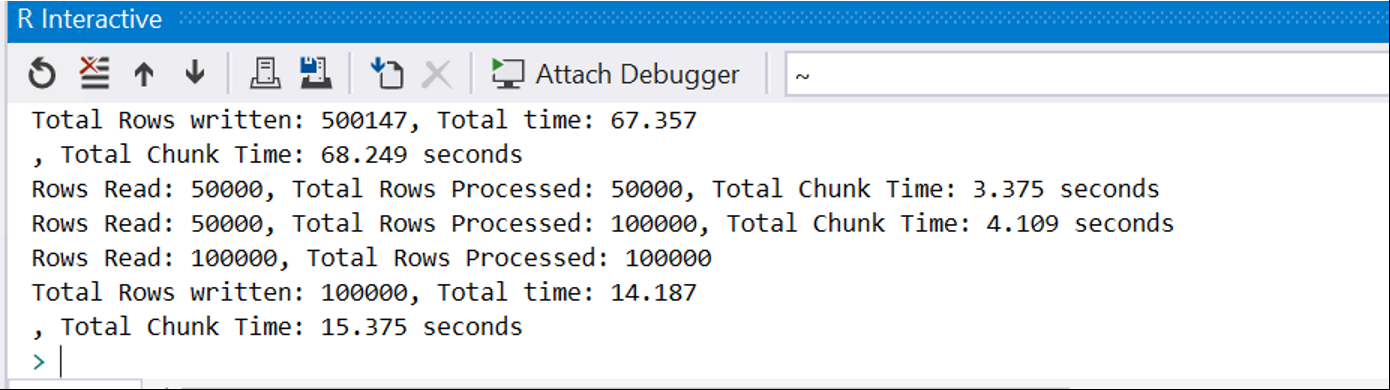
Running these scripts will walk through the initial R code used to create this solution – dataset creation, modeling, and scoring as described in the [data scientist page](https://github.com/Microsoft/SQL-Server-R-Services-Samples/tree/master/CampaignManagement/CampaignManagement/Resources/data-scientist.md) on GitHub.

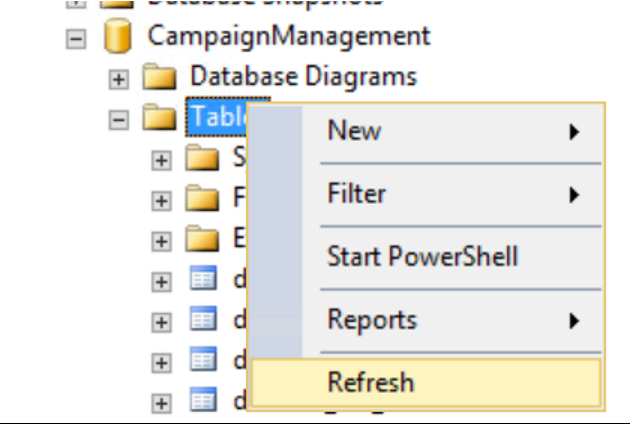
The code you see here was later incorporated into the .sql files. The PowerShell scripts then invoke the .sql files to fully automate the solution.

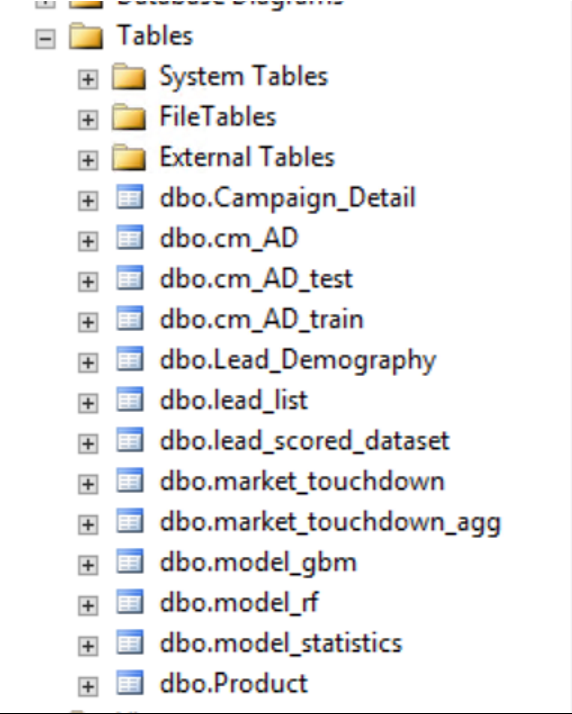
Solution with R Server

1. *Open the four files in the R directory in the R IDE of your choice.*
2. *Replace the connection string at the top of each file with details of your login and database name in each of the four files. For example:*

 *(Note: You can use “.” for the server name as shown here if using a local SQL Server (on the same machine as your code).*

1. *The scripts perform the following actions:*
   1. **Step1\_input\_data.R**: Simulates the 4 input datasets
   2. **Step2\_data\_preprocessing.R**: Performs preprocessing steps like outlier treatment and missing value treatment on the input datasets
   3. **Step3\_feature\_engineering\_AD\_creation.R**: Performs Feature Engineering and creates the Analytical Dataset. Feature Engineering consists of creating new variables in the market touchdown dataset by aggregating the data in multiple levels. The table is aggregated at a lead level, so variables like channel which will have more than one value for each user are pivoted and aggregated to from variables like SMS count, Email count, Call Count, Last Communication Channel, Second Last Communication Channel etc.
   4. **Step4\_model\_rf\_gbm.R**: Builds the Random Forest & Gradient Boosting models, identifies the champion model and scores the Analytical dataset
2. *Run each script in order. Note some may take some time to finish. You’ll know they are done when you put cursor in the Console area (labeled “R Interactive” in RTVS) and it is no longer spinning. Also when done you’ll see the command prompt “>” ready for the next interactive command.*
3. *After each step completes, feel free to go back to SSMS and look at the contents of the database. You’ll need to right click on Database and “Refresh” to see the most recent set of results.*



1. *When you have finished with all four scripts, you should see the following tables in SSMS in your Database:*
2. *Now proceed to Visualizing Results with PowerBI by following the instructions in* [*Visualize Results*](Visualize%20Results.docx)*.docx.*