**Power BI: How to Analyze FIFA 2014 Worldcup**

# Data Extraction

I use more than one data set for analysis, so let’s start with some basic data sets;

## Goals Scored by Teams

Number of goals scored by each team listed in this URL: <http://www.fifa.com/worldcup/archive/brazil2014/statistics/teams/goal-scored.html>

Steps to get this data set into Power BI is;

* Open Power BI Desktop
* Choose Get Data
* In Get Data Window, under Other, click on Web

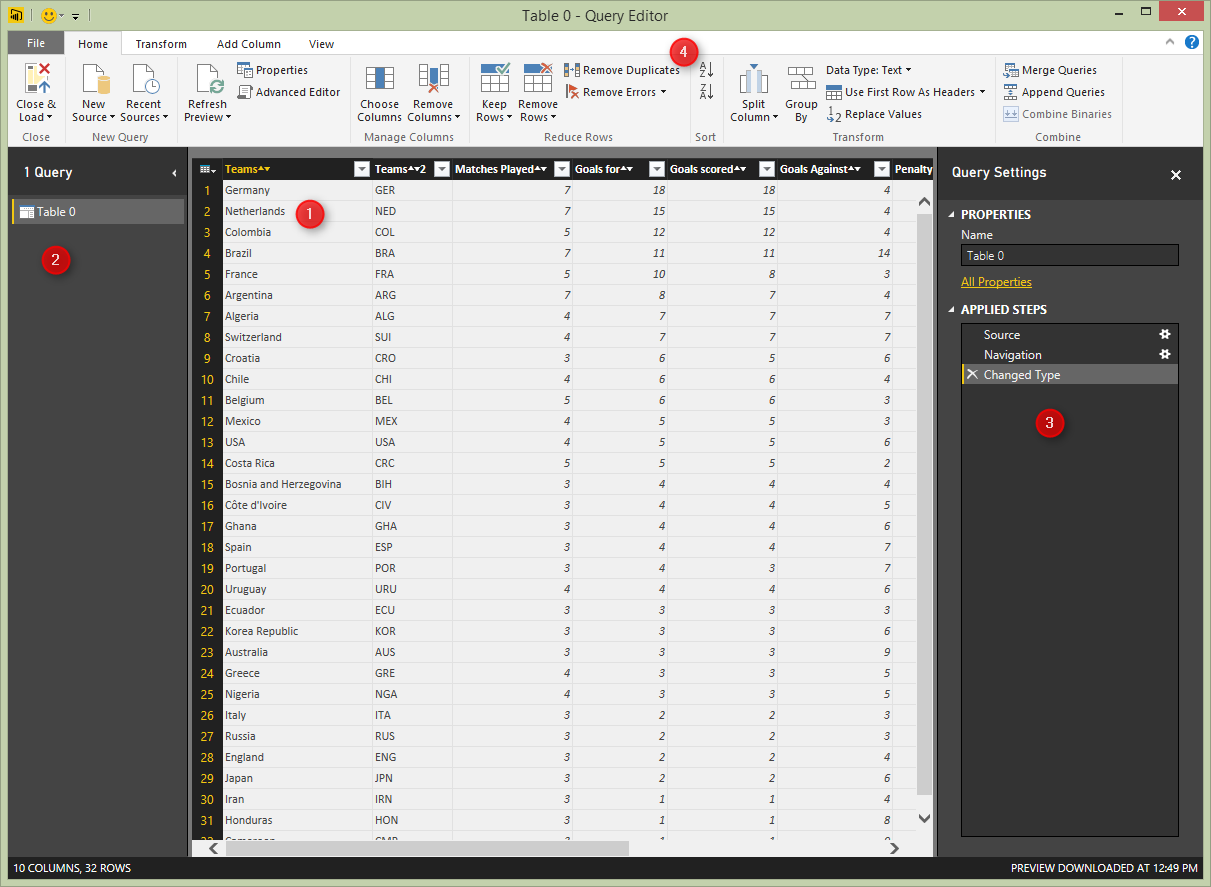
In “From Web” dialog box, Enter the URL and click OK; <http://www.fifa.com/worldcup/archive/brazil2014/statistics/teams/goal-scored.html>

a Navigator window will open as below, click on the table that contains data set of goal scorers and then click on Edit to open Query Editor window

The Query Editor window will appear with the data set loaded and displayed in the main pane.

Query Editor is the Power Query editor window, which gives you ability to transform data and then load it into the model that Power BI can use. There are different panes in Power Query Editor window here;

1. Main pane that shows a preview of the data set after applying transformation listed in the Query Setting pane.
2. Query pane, which shows all queries used
3. Query setting pane; here you can see steps applied on the data set, you can delete them or change them as you want.
4. Transformation Pane; all transformation options listed in top menu here, you can choose simply to work with any of them.



This data set is almost ready to use, the only transformation I want to make here is to remove extra two characters at the end of each column name that came from icons used in the FIFA website. so rename every column with right click on that and choosing “rename column” option from the menu.

Also rename the Query name to TeamGoals in the right hand side pane in Query Setting section. Here is the result set after above changes;

## Passes by Teams

The second data set is passed by teams which we get from this url: <http://www.fifa.com/worldcup/archive/brazil2014/statistics/teams/passes.html>

* Without closing Query Editor window, on the Menu (under Home, New Query) click on New Source, and choose from Web. In the From Web paste URL above and press OK.
* In Navigator window choose the table that contains passes by teams data set, and click OK.
* Now in the same Query Editor window you will see Table 0 loaded with it’s data set in the main pane
* For this data set again we have to change column names and remove two extra characters from each column name. Rename Team2 column name to Teams Abrv.
* There are some percentage columns in this data set, which considered as text column (because of % character in the value), we want to remove the character and change the data type to be Decimal number.
* Right click on Passes Completed Percentage column (this is renamed column version of Passes Completed 2), and choose Replace Values
* Now change the data type of this column to Decimal number
* Do the same for Crosses Completed Percentage column (renamed from Crosses Completed 2)
* Rename the query to TeamPasses. The result should looks like below screenshot

## Teams Defending

I would also like to add another data set to this analysis which is Teams Defending result set. URL: <http://www.fifa.com/worldcup/archive/brazil2014/statistics/teams/defending.html>

* Get Data again, from New Source, From Web, and enter URL above.
* After loading the data set in Query Editor apply below changes (as illustrated above)
* Rename columns and remove extra two characters
* Change the data type of percentage columns to Decimal number
* Rename the query to TeamDefending

# Data Modelling

We can add much more data sets from FIFA official website. but for this example let’s keep this simple and stick to three data sets above. In this section we would model relationship between these data sets, and model it to be used in reporting section.

in Query Editor window click on Close & Load icon in the Home tab, under Close section

**This is what ETL (Extraction, Transformation and Loading)**

## Relationship Tab

In the Power BI Desktop from left pane click on Relationships tab and you will see the relationship created itself by the similarity of column name between queries.

You can also change the relationship if you want. current relationship is based on Teams column as the key column.

## Data Tab

In the data tab you can view the data set, and set formatting of the columns or add new measures (DAX measures) if you want.  Screenshot below shows how to access data tab from the left pane. You can also see the Modeling menu that gives you ability to create and edit measures, as well as format them in appropriate way. you can also access managing relationships from this menu.

Let’s set formatting of percentage columns appropriate;

* in the Fields pane in right hand side click on TeamPasses.
* In the data set click on Passes Completed Percentage, and from Formatting section in Modeling menu choose percentage.
* Do this for all percentage columns in TeamPasses and TeamDefending

Now our model is ready enough to be used in reporting section

# Visualization

Now let’s make some visualization with the Report tab based on the data in model we’ve built so far.

## Goals Scored Column Chart

From the Fields section under TeamGoals, drag and drop teams into the main pane, then drag and drop Goals Scored on it.

## Map

Double click on the map from visualization section. Then from TeamPasses Drag and drop Teams into the Location section of the map setting. Bring Passes Completed into the values section. Let’s change the color, and color saturate it by number of matches each team played.

## Gauge

Now let’s create a gauge for goals scored. set the value as Goals Scored. Minimum value as Min of Goals Scored, Maximum value as Max of Goals scored, and the Target as Average of Goals Scored.

## Clustered Column and Line Chart

as the last visualization item in this report I would like to show a column chart combined with line chart. the column values shows Completed Clearance from TeamDefendings, and the line value shows number of Saves.

# Publish

Now let’s publish this report into Power BI website.