



PASS
SUMMIT 2016

Data Preparation Is the Keystone

Reza Rad, Consultant, RADACAD





Please silence
cell phones

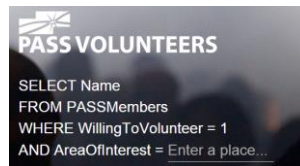
Explore Everything PASS Has to Offer



FREE ONLINE WEBINAR EVENTS



FREE 1-DAY LOCAL TRAINING EVENTS



VOLUNTEERING OPPORTUNITIES



**LOCAL USER GROUPS
AROUND THE WORLD**



**ONLINE SPECIAL INTEREST
USER GROUPS**



PASS COMMUNITY NEWSLETTER

PASS BLOG

WHITE PAPERS

SESSION RECORDINGS

FREE ONLINE RESOURCES



BUSINESS ANALYTICS TRAINING



BA INSIGHTS NEWSLETTER

Session Evaluations

Your feedback is
important and valuable.

3

ways to access

Submit by 5pm
Friday November 6th to
WIN prizes



Go to passSummit.com



Download the GuideBook App
and search: **PASS Summit 2016**



Follow the QR code link displayed
on session signage throughout the
conference venue and in the
program guide



Reza Rad

Consultant, RADACAD

DW/BI Consultant, Mentor, Trainer, Speaker
Microsoft Data Platform MVP
Author of SQL Server and BI books
Author of Power BI from Rookie to Rock Star book
Microsoft Certified Trainer
Microsoft Certified Professional
Co-Leader of NZ BI User Group

 /rezarad

 @rad_reza

 rezaradf

Agenda

Why Data Preparation?

Tips for Data Preparation

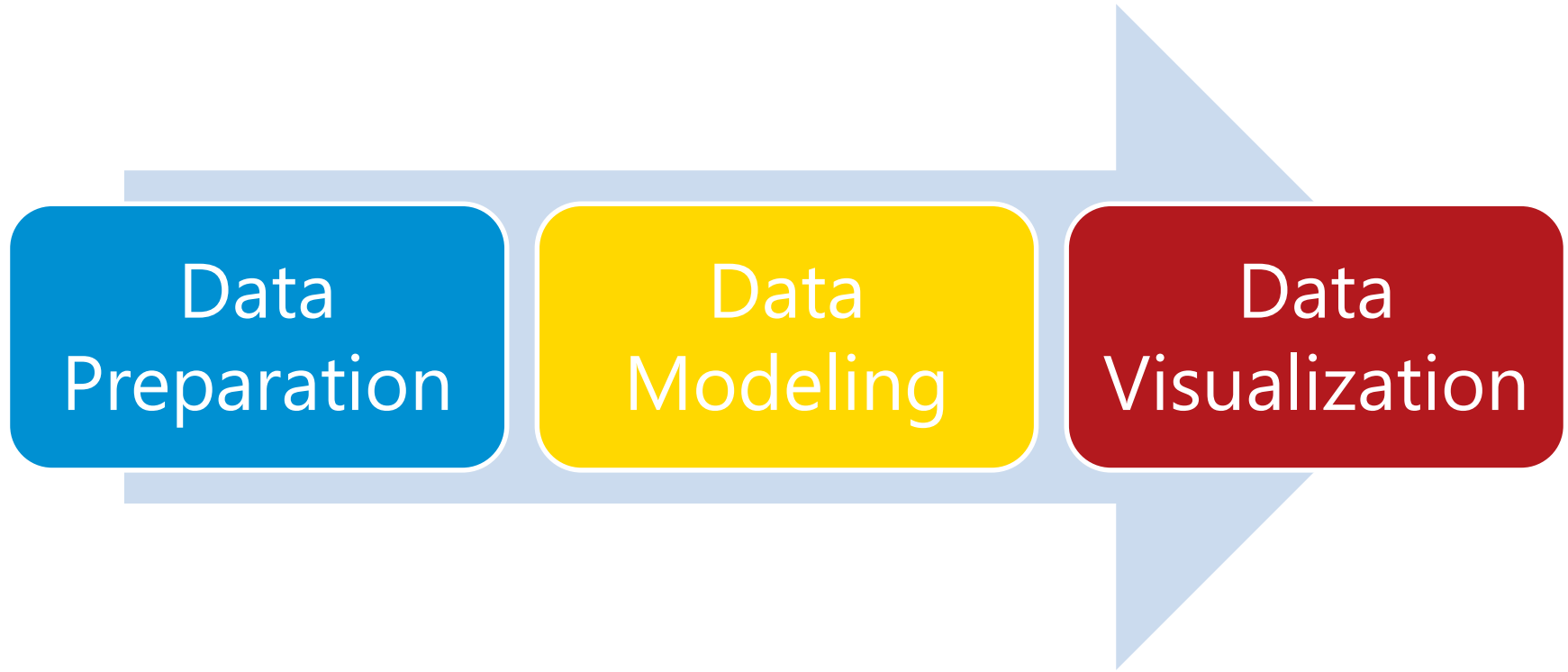
Introduction to Power Query

Power Query Formula Language: M

Top M Functionalities

Date Dimension with Power Query

Steps To Data Insight



Building Block; Keystone

Data Preparation is the very first step in Data Insight

Good Data Preparation => Useful Insight

Bad Data Preparation => Lots of Re-Work



Data Preparation

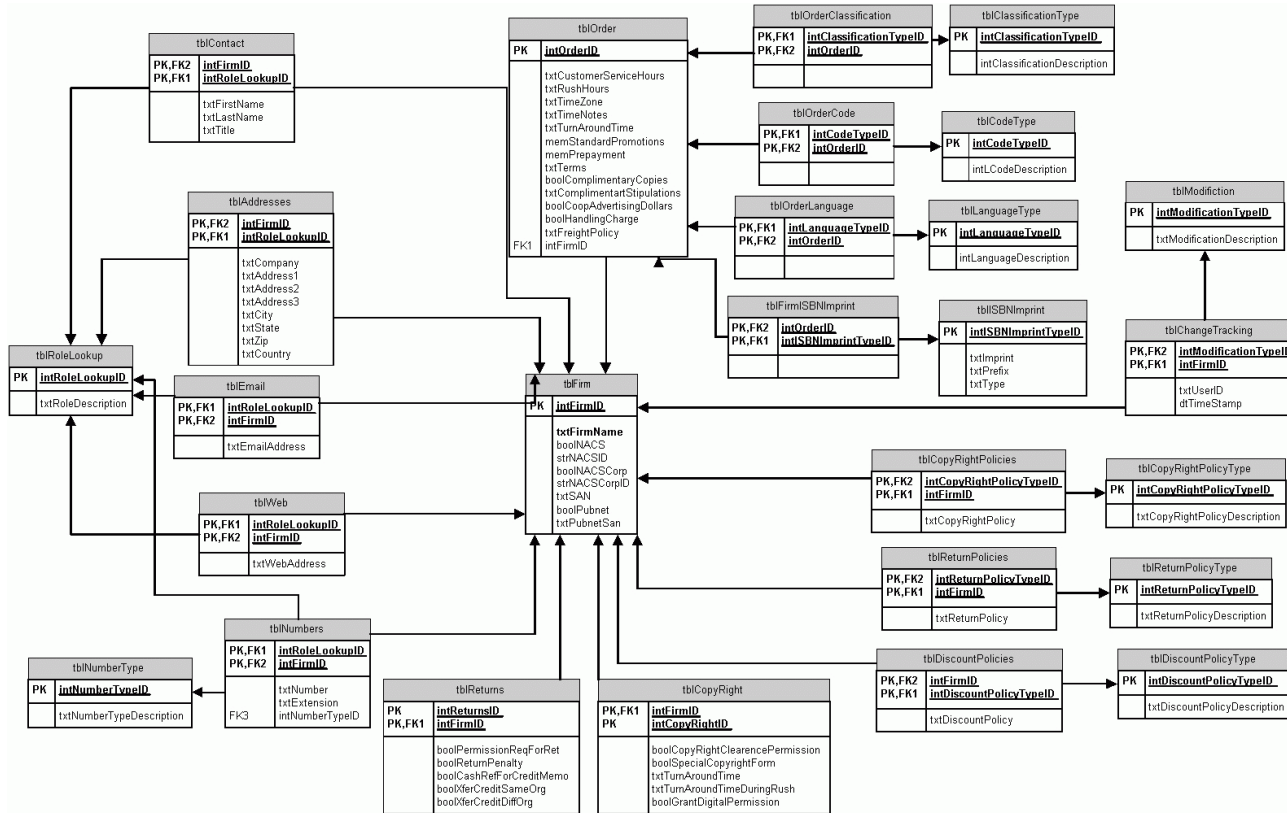
Getting Data from Multiple Sources

Data Cleansing

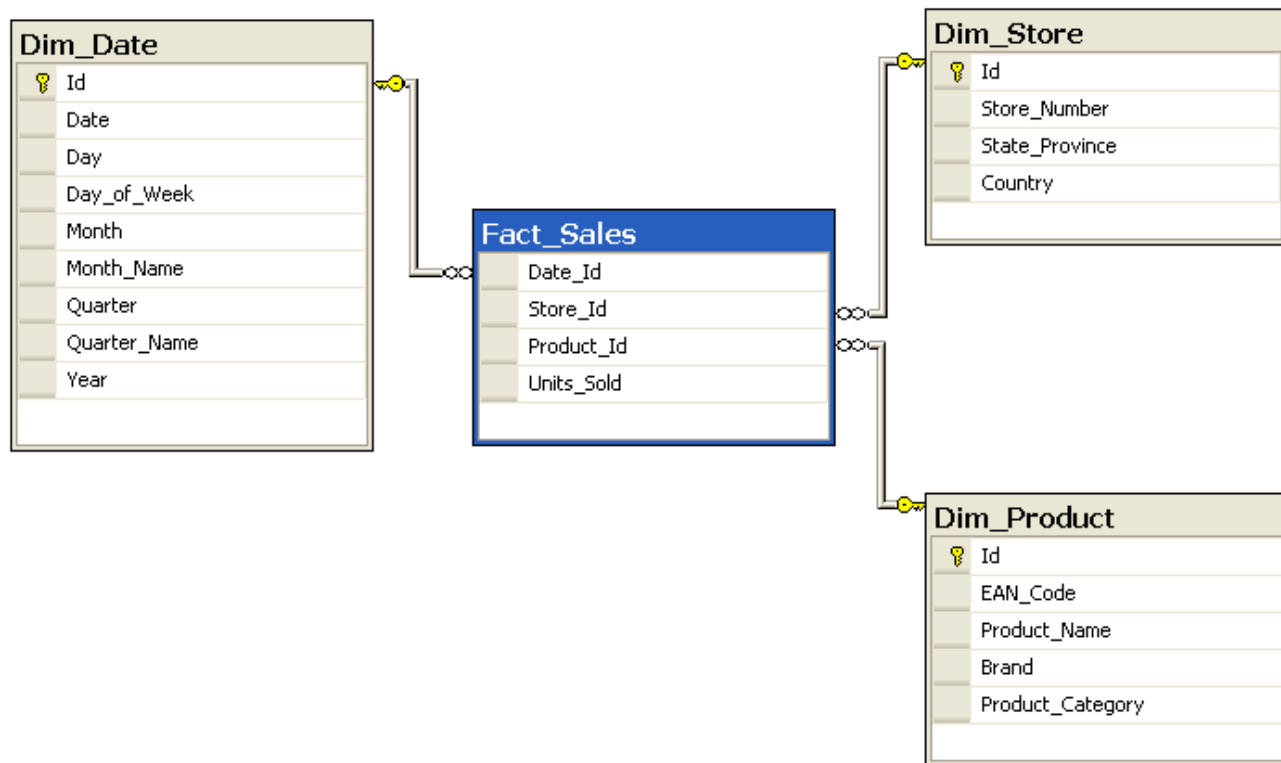
Data Transformation

Preparing Data for Modeling

Data Can Be Like This



Simpler; Better



Dimensional Modeling

Customer **Dimension**

Customer Name
Customer Age
Customer Geo
Customer
Contact Info
Customer Job
Title
Customer ID
CustomerKey

Sales **Fact** Table

Sales Amount
Quantity Sold
Profit

CustomerKey

ProductKey

OrderDateKey

Product Dimension

Product Name
Product Number
Product Brand
Category
Subcategory
ProductKey

Date Dimension

DateKey
Year
Quarter
Month
Day
Fiscal Columns

Tips to Consider #1

DO NOT add tables/files as is

Why?

Tables can be joined together to create more **flatten** and simpler structure => **Better Modeling**

DO Create flatten structure for tables (specially dimensions)

Tips to Consider #2

DO NOT flatten your FACT tables!

Why?

Fact tables are largest entities in your model. With flattening them you are making them even larger! =>
Memory Management & Performance Consideration

DO Instead of flattening fact tables, create relation to Dimension Tables

Tips to Consider #3

DO NOT leave naming as is

Why?

Name your tables and columns for the end user => **Better End User Experience, Better Q&A**

DO Name tables and columns as you design table in restaurant

Tips to Consider #4

DO NOT leave data types as is

Why?

Proper Data Types makes modelling easier => **Better Modeling**

DO Set Proper Data Types based on the data in each field

Tips to Consider #5

DO NOT load the whole data set if you don't require it

Why?

Filtering part of the data before loading it into memory is cost and performance effective => **Better Performance**

DO Filter Part of the data that is not required.

Data Preparation Tool

For Data Analyst: Power Query

Power Query for Power BI

Power Query for Excel



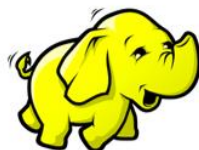
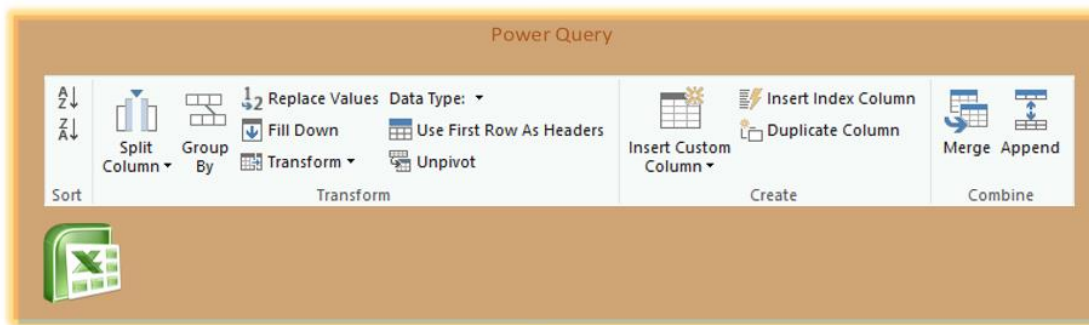
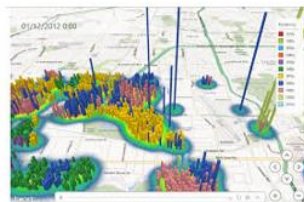
Introduction to Power Query



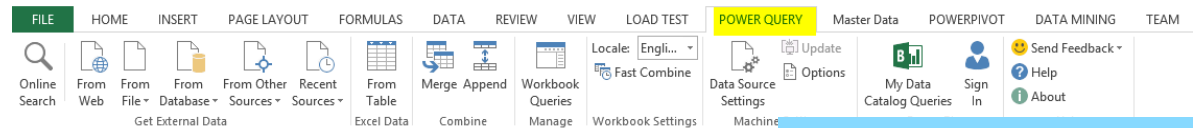
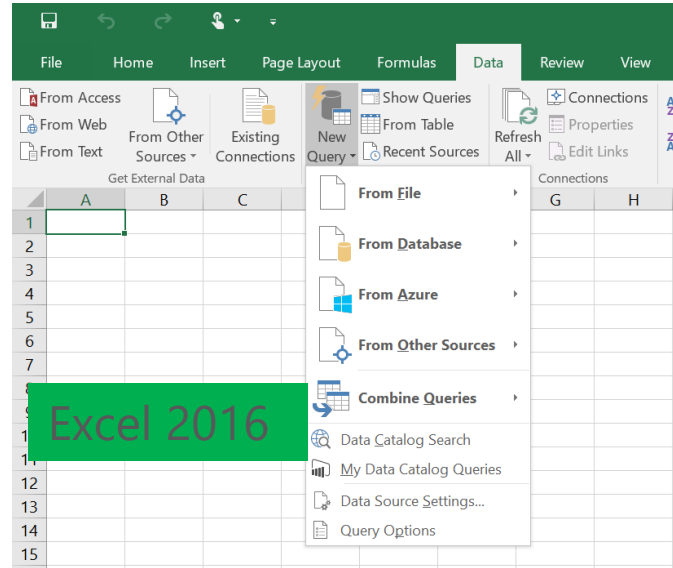
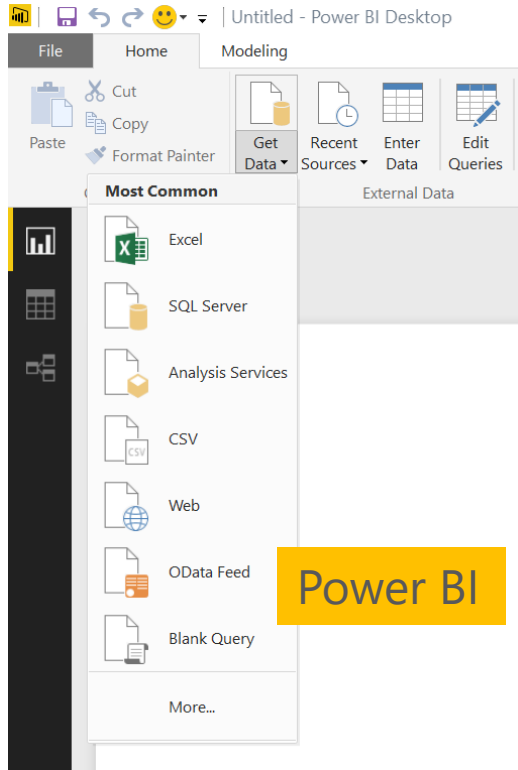
@Rad_Reza



<http://www.radacad.com>



Comes with Different Shapes and Sizes



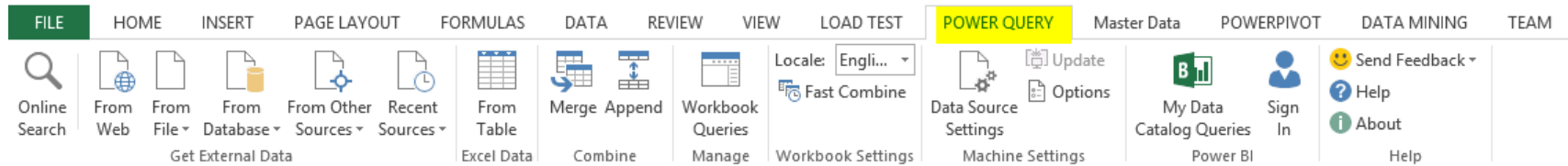
Excel 2013, 2010

How to Get it?

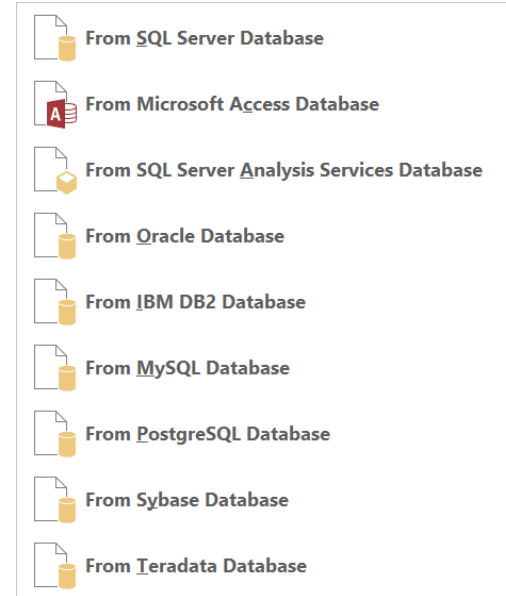
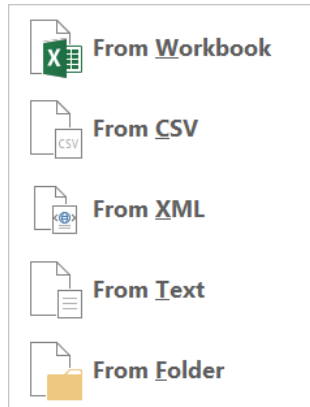
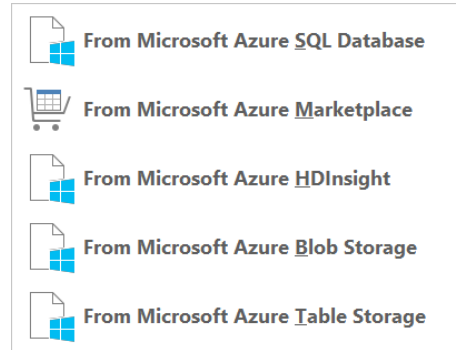
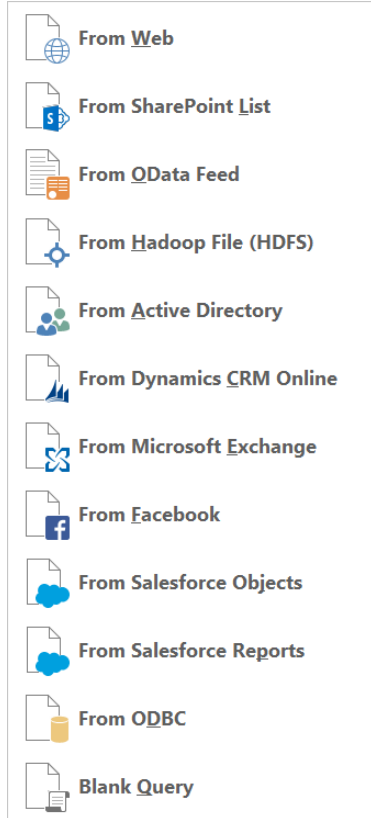
Power BI Desktop

Built-in for Excel 2016

Free Add-in download for Excel 2013, and 2010



Wide Range of Data Sources Supported



Data Transformations on GUI

Title	Year	IMDb Rating
The Shawshank Redemption		9.2
The Godfather		9.2
The Godfather: Part II		9
The Dark Knight		8.9
Pulp Fiction		8.9
The Good, the Bad and the Ugly		8.9
Schindler's List		8.9
12 Angry Men		8.9
The Lord of the Rings: The Return of the King		8.9
Fight Club		8.8
The Lord of the Rings: The Fellowship of the Ring		8.8
Star Wars: Episode V - The Empire Strikes Back		
Inception		
One Flew Over the Cuckoo's Nest		
Forrest Gump		
Goodfellas		
The Lord of the Rings: The Two Towers		
Star Wars: Episode IV - A New Hope		
The Matrix		
Seven Samurai		
City of God		

Remove

Remove Other Columns

Use First Row As Headers

Duplicate Column

Split Column

Remove Duplicates

Remove Errors

Replace Values...

Fill Down

Change Type

Transform

Insert Custom Column...

Insert Index Column

Group By...

Unpivot

Move

Rename...

Drill Down

Add as New Query

lowercase

UPPERCASE


Capitalize Each Word

Trim

Clean

JSON

XML




DEMO: IMDB Movies Data Mashup



@Rad_Reza



<http://www.radacad.com>



Power Query Formula Language: M



@Rad_Reza



<http://www.radacad.com>

What is M?

Everything that Happens on GUI works with a Code Behind
Code Behind is a Functional Language: M

```
let
    TableA = #table({"CustomerId", "TranDate", "TranCount"},
        {
            {1, DateTime.FromText("2014-01-01 01:00:00.000"), 10},
            {1, DateTime.FromText("2014-01-01 02:00:00.000"), 5},
            {1, DateTime.FromText("2014-01-03 01:00:00.000"), 5},
            {1, DateTime.FromText("2014-01-04 02:00:00.000"), 80}
        },

    TableB = #table({"CustomerId", "TranDate", "TranCount"},
        {
            {1, DateTime.FromText("2014-01-01 02:00:00.000"), 10},
            {1, DateTime.FromText("2014-01-01 03:00:00.000"), 5},
            {1, DateTime.FromText("2014-01-02 01:00:00.000"), 20},
            {1, DateTime.FromText("2014-01-02 03:00:00.000"), 15},
            {2, DateTime.FromText("2014-01-01 01:00:00.000"), 5},
            {2, DateTime.FromText("2014-01-01 02:00:00.000"), 80}
        },
    TableATransformed=Table.Sort(
        Table.AddColumn(TableA, "Date", each Date.From([TranDate]))
        , {"CustomerId", "TranDate"}
    )
in
    Table.Group(TableATransformed, {"CustomerId", "Date"}, {"Total", each List.Last([TranCount])})
```

More about Formula Language

M is much more powerful than Power Query GUI;

Not all functionalities of M implemented through GUI.

If you want to be Professional In Power Query; You should be Expert in M

M is a functional language

M Syntax

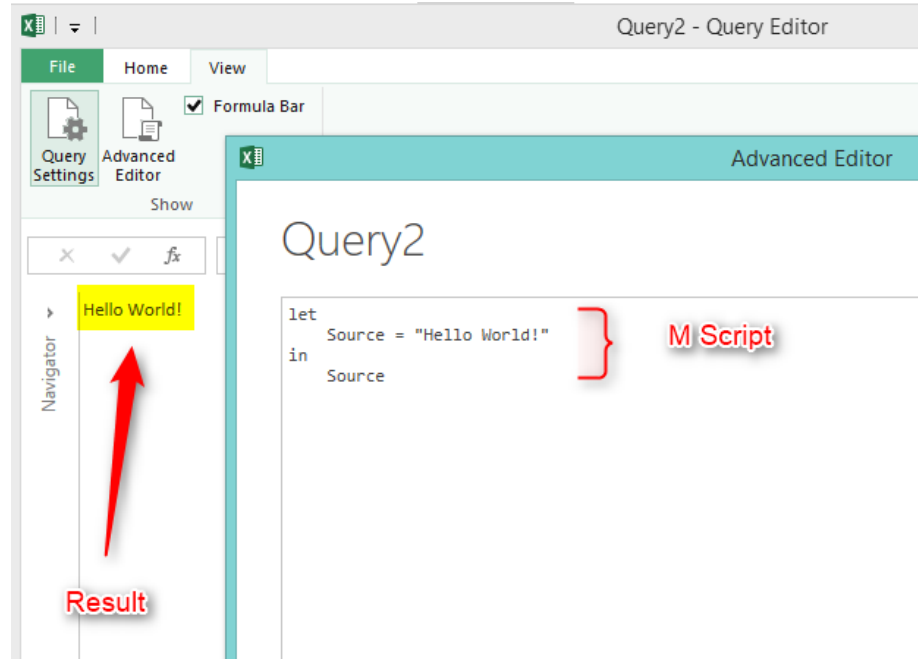
let

x=1

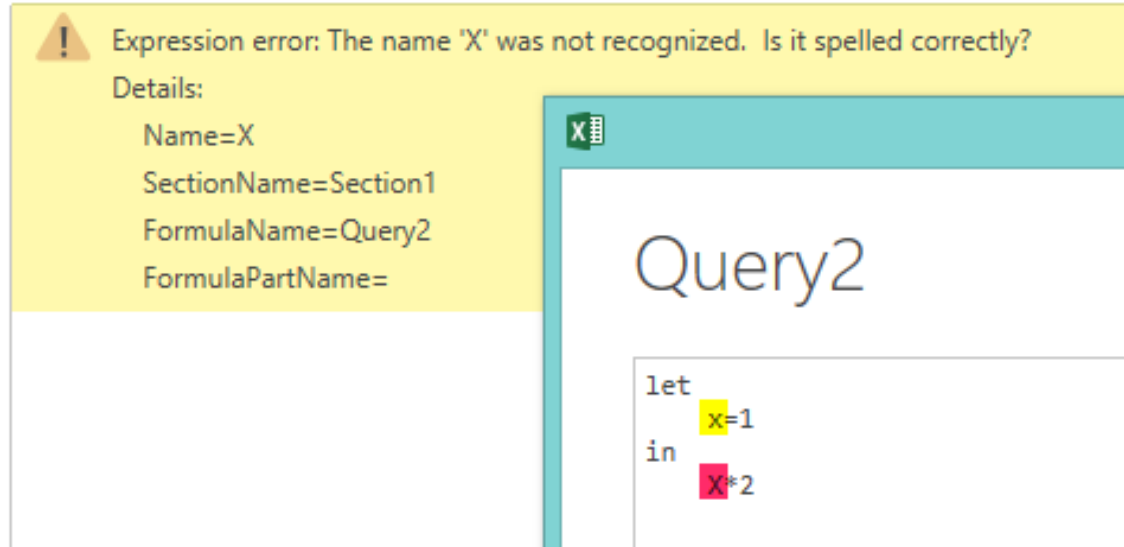
in

x

Structure of M



M is Case Sensitive



The image shows a screenshot of a Power BI error message and a query editor snippet. The error message is in a yellow box with a warning icon and text: "Expression error: The name 'X' was not recognized. Is it spelled correctly?". Below the error message, the details are listed: "Name=X", "SectionName=Section1", "FormulaName=Query2", and "FormulaPartName=". To the right of the error message is a snippet of a query editor window titled "Query2". The query text is: "let", "x=1", "in", "X*2". The "x=1" is highlighted in yellow and the "X*2" is highlighted in red.

! Expression error: The name 'X' was not recognized. Is it spelled correctly?

Details:

- Name=X
- SectionName=Section1
- FormulaName=Query2
- FormulaPartName=

Query2

```
let
    x=1
in
    X*2
```

Three Base Structures in M

	CustomerId	TranDate	TranCount
1	1	1/01/2014 1:00:00 a.m.	10
2	1	1/01/2014 2:00:00 a.m.	5
3	1	3/01/2014 1:00:00 a.m.	5
4	1	4/01/2014 2:00:00 a.m.	80

Table

CustomerId	1
TranDate	1/01/2014 2:00:00 a.m.
TranCount	5

Record

	List
1	This
2	is
3	a
4	list

List



Top M Functionalities

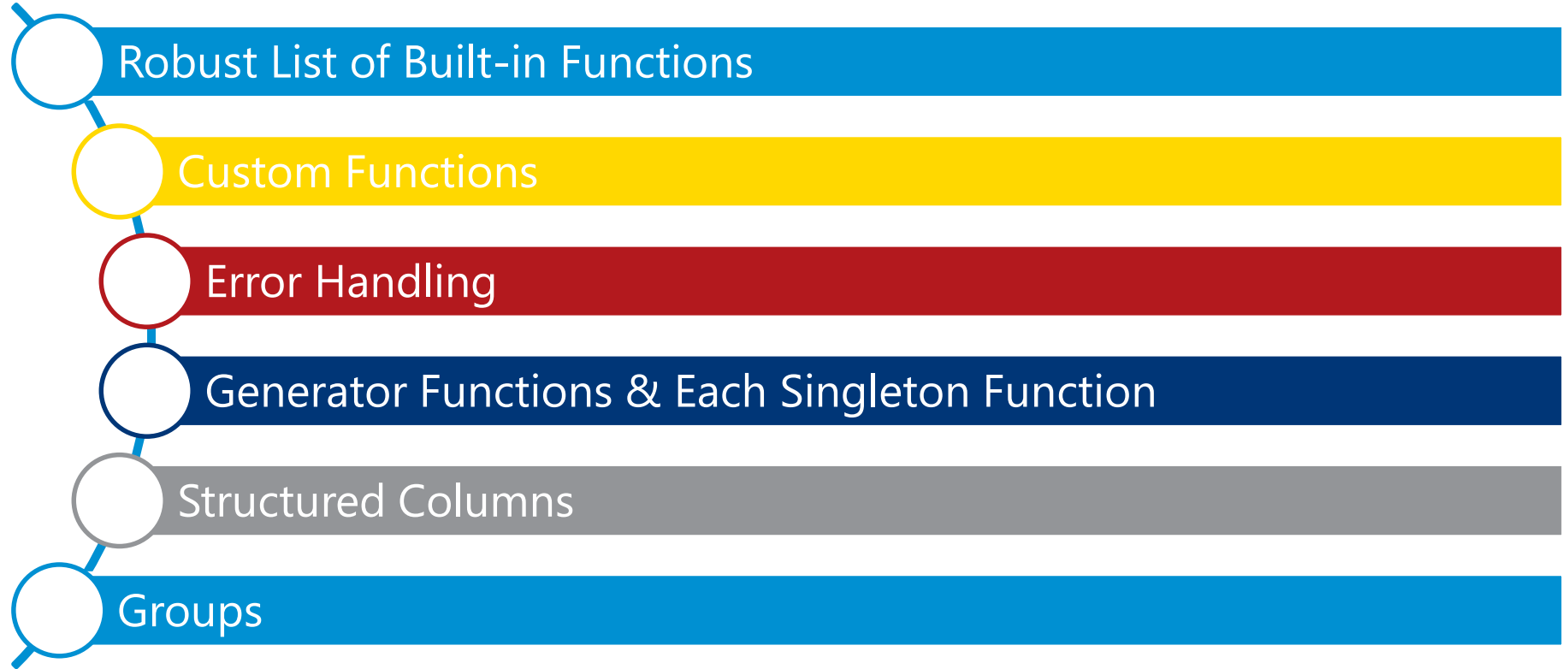


@Rad_Reza



<http://www.radacad.com>

Top M Functionalities



Robust Built-in Functions

Table Functions

Date Functions

#shared keyword

✓	fx	= #shared
	Date.Year	Function
	Date.Month	Function
	Date.Day	Function
	Date.AddDays	Function
	Date.AddWeeks	Function
	Date.AddMonths	Function
	Date.AddQuarters	Function
	Date.AddYears	Function
	Date.IsInPreviousDay	Function
	Date.IsInCurrentDay	Function
	Date.IsInNextDay	Function
	Date.IsInPreviousWeek	Function

Table Functions

Table.AddColumn

Table.RemoveColumns

Table.ReorderColumns

Table.SelectColumns

Table.ReplaceValue

Table.Sort

Table.AddIndexColumn

Table.FillDown

Custom Functions

Re-use your code

Reduce redundancy

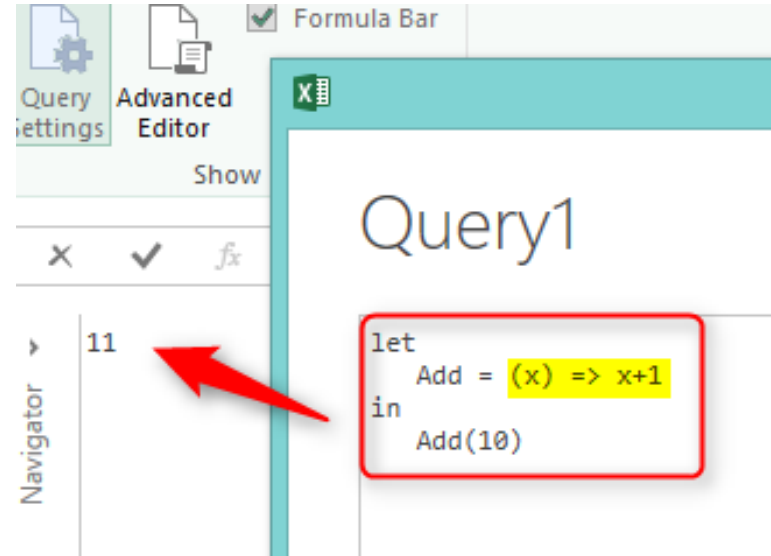
Increase consistency

Lambda Expression

$(x) \Rightarrow x + 1$

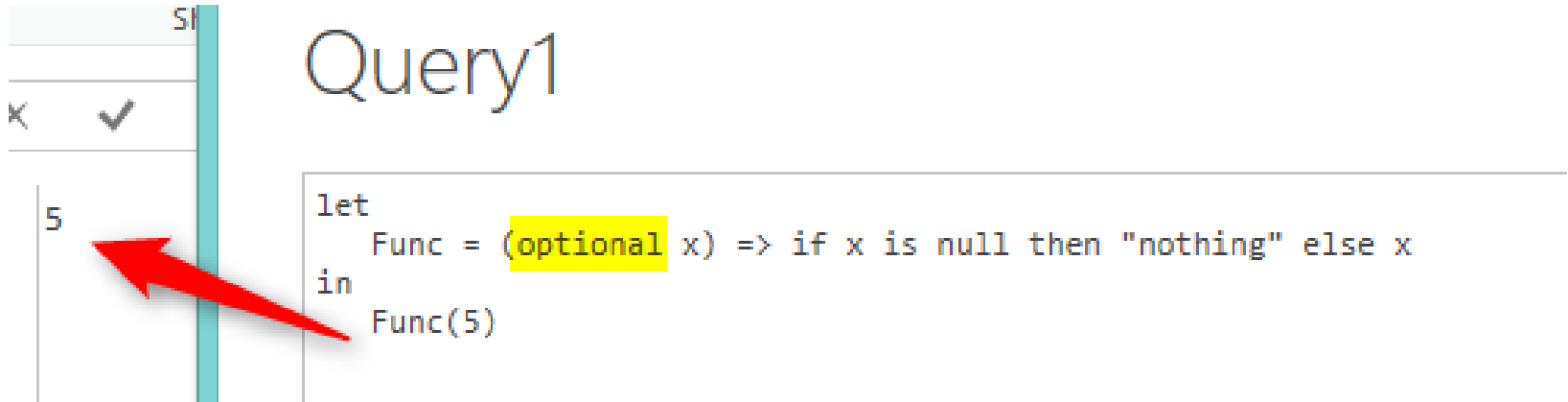
Input Parameters \Rightarrow Function Body

$(x, y) \Rightarrow x + y$



Parameters

Parameters (optional, required)



The image shows a screenshot of a Power BI query editor. On the left, a table with one column and one row containing the value 5 is visible. A red arrow points from the value 5 in the table to the word 'optional' in the DAX formula on the right. The DAX formula is as follows:

```
let  
    Func = (optional x) => if x is null then "nothing" else x  
in  
    Func(5)
```

The word 'optional' is highlighted in yellow in the original image.

Multi-Line Functions

Cascade let/in clause

```
let  
  Func = (x) =>  
    let  
      <body>  
    in  
      <return value>  
in  
  Func(5)
```

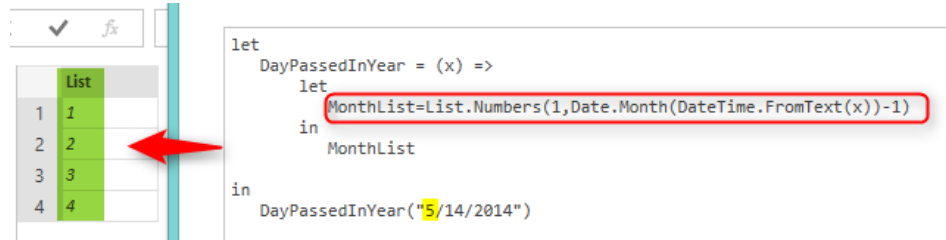
Body lines can be separated with comma

Generators

List can be generated

List.Dates(<start date>,<end date>)

Generates List of Dates from <start date> to <end date>



The image shows a spreadsheet on the left and a code editor on the right. The spreadsheet has a column labeled 'List' with values 1, 2, 3, and 4. A red arrow points from the code editor to the spreadsheet. The code editor shows a function definition for DayPassedInYear, which uses List.Dates to generate a list of dates.

```
let  
  DayPassedInYear = (x) =>  
    let  
      MonthList = List.Dates(1, Date.Month(DateTime.FromText(x))-1)  
    in  
      MonthList  
in  
  DayPassedInYear("5/14/2014")
```

It can be used as **Loop** structure

Error Handling

Control of Execution when error happens

Sending proper error messages

Preventing failure with messages for troubleshooting

Structured Column



Date Dimension with Power Query



@Rad_Reza



<http://www.radacad.com>

Date Dimension

	Year	Month	Day	FullDateAl...	DateKey	DateFullName	Fiscal Year	Fiscal Quarter	Calendar Quar...	IsWeekDay	DayOfWeek	Month Name	Day of Week Name
1	2013	6	15	6/15/2013	20130615	15 June 2013	2013		4	2	0	6 June	Saturday
2	2013	6	16	6/16/2013	20130616	16 June 2013	2013		4	2	0	0 June	Sunday
3	2013	6	17	6/17/2013	20130617	17 June 2013	2013		4	2	1	1 June	Monday
4	2013	6	18	6/18/2013	20130618	18 June 2013	2013		4	2	1	2 June	Tuesday
5	2013	6	19	6/19/2013	20130619	19 June 2013	2013		4	2	1	3 June	Wednesday
6	2013	6	20	6/20/2013	20130620	20 June 2013	2013		4	2	1	4 June	Thursday
7	2013	6	21	6/21/2013	20130621	21 June 2013	2013		4	2	1	5 June	Friday
8	2013	6	22	6/22/2013	20130622	22 June 2013	2013		4	2	0	6 June	Saturday
9	2013	6	23	6/23/2013	20130623	23 June 2013	2013		4	2	0	0 June	Sunday
10	2013	6	24	6/24/2013	20130624	24 June 2013	2013		4	2	1	1 June	Monday
11	2013	6	25	6/25/2013	20130625	25 June 2013	2013		4	2	1	2 June	Tuesday
12	2013	6	26	6/26/2013	20130626	26 June 2013	2013		4	2	1	3 June	Wednesday
13	2013	6	27	6/27/2013	20130627	27 June 2013	2013		4	2	1	4 June	Thursday
14	2013	6	28	6/28/2013	20130628	28 June 2013	2013		4	2	1	5 June	Friday
15	2013	6	29	6/29/2013	20130629	29 June 2013	2013		4	2	0	6 June	Saturday
16	2013	6	30	6/30/2013	20130630	30 June 2013	2013		4	2	0	0 June	Sunday
17	2013	7	1	7/1/2013	20130701	01 July 2013	2014		1	3	1	1 July	Monday
18	2013	7	2	7/2/2013	20130702	02 July 2013	2014		1	3	1	2 July	Tuesday
19	2013	7	3	7/3/2013	20130703	03 July 2013	2014		1	3	1	3 July	Wednesday
20	2013	7	4	7/4/2013	20130704	04 July 2013	2014		1	3	1	4 July	Thursday
21	2013	7	5	7/5/2013	20130705	05 July 2013	2014		1	3	1	5 July	Friday
22	2013	7	6	7/6/2013	20130706	06 July 2013	2014		1	3	0	6 July	Saturday

Features

Configurable via table in Excel Sheet

Fiscal Columns

Public Holidays Fetched Live

Year	Month	Day	F...	D...	DateFullName	F...	F...	C...	I...	D...	Month Name	Day of Week Name	HolidayDescription	IsPublicHoliday
2014	4	14	4/14/2014	14 April 2014	2014	4	2	1	1	April	Monday		null	0
2014	4	15	4/15/2014	15 April 2014	2014	4	2	1	2	April	Tuesday		null	0
2014	4	16	4/16/2014	16 April 2014	2014	4	2	1	3	April	Wednesday		null	0
2014	4	17	4/17/2014	17 April 2014	2014	4	2	1	4	April	Thursday		null	0
2014	4	18	4/18/2014	18 April 2014	2014	4	2	1	5	April	Friday	Good Friday		1
2014	4	19	4/19/2014	19 April 2014	2014	4	2	0	6	April	Saturday		null	0
2014	4	20	4/20/2014	20 April 2014	2014	4	2	0	0	April	Sunday		null	0
2014	4	21	4/21/2014	21 April 2014	2014	4	2	1	1	April	Monday	Easter Monday		1
2014	4	22	4/22/2014	22 April 2014	2014	4	2	1	2	April	Tuesday		null	0
2014	4	23	4/23/2014	23 April 2014	2014	4	2	1	3	April	Wednesday		null	0
2014	4	24	4/24/2014	24 April 2014	2014	4	2	1	4	April	Thursday		null	0

How it works

Generators; to build the base structure

EACH; to apply transformations on each item in list

Date/Text Functions

OData

Table Functions

Date Functions

Public Holidays

<https://www.opm.gov/policy-data-oversight/snow-dismissal-procedures/federal-holidays/#url=Overview>

FEDERAL HOLIDAYS

Overview 2020 2019 2018 2017 2016 2015 2014 2013 2012 2011

2020

2020 Holiday Schedule

Date	Holiday
Wednesday, January 1	New Year's Day
Monday, January 20	Birthday of Martin Luther King, Jr.
Monday, February 17*	Washington's Birthday
Monday, May 25	Memorial Day
Friday, July 3**	Independence Day
Monday, September 7	Labor Day
Monday, October 12	Columbus Day
Wednesday, November 11	Veterans Day
Thursday, November 26	Thanksgiving Day
Friday, December 25	Christmas Day

*This holiday is designated as "Washington's Birthday" in section 6103(a) of title 5 of the United States Code, which is the law that specifies holidays for Federal employees. Though other institutions such as state and local governments and private businesses may use other names, it is our policy to always refer to holidays by the names designated in the law.

**July 4, 2020 (the legal public holiday for Independence Day), falls on a Saturday. For most Federal employees, Friday, July 3, will be treated as a holiday for pay and leave purposes. (See 5 U.S.C. 6103(b).)



Date Dimension

DEMO



@Rad_Reza



<http://www.radacad.com>

Grouping

Empower it with M

Summary

- Data Preparation
- Introduction to Power Query
- Power Query Formula Language: M
- Custom Function, Generators, Built-in Functions, Error Handling, Structured Columns, Grouping
- Date Dimension

References to Study More

Power BI from Rookie to Rock Star book: FREE

<http://www.radacad.com/online-book-power-bi-from-rookie-to-rockstar>

Power Query Formula Categories online Help:

<http://office.microsoft.com/en-001/excel-help/power-query-formula-categories-HA104122363.aspx>

Power BI Self-Paced Training Course/Videos:

<http://www.learn.radacad.com>



@Rad_Reza



<http://www.radacad.com>

Session Evaluations

Your feedback is
important and valuable.

3

ways to access

Submit by 5pm
Friday November 6th to
WIN prizes



Go to passSummit.com



Download the GuideBook App
and search: **PASS Summit 2016**



Follow the QR code link displayed
on session signage throughout the
conference venue and in the
program guide



Thank You

Learn more from

Reza Rad

reza@radacad.com or follow @Rad_Reza