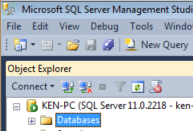
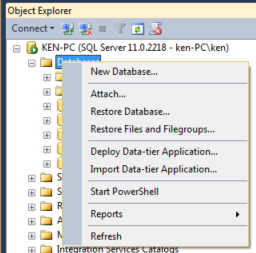
For Assignment 2 in Big Data, you will ultimately use Tableau to visualize a large dataset. The real-world problem that you need to overcome lies in the fact that the data is not in the nice, clean form of an Excel spreadsheet already formatted and cleansed to make visualization simple for you. For A2, your data resides in a data warehouse called AdventureworksDW2012 (I’ll just call it DW from here on out). This tutorial document will walk you through the process of adding the DW to your SQL Server on your lab machine, creating a query that pulls some data from this DW by creating a database object called a “View”. A view is basically a stored query that can then be easily invoked/called for by another program or system. For this assignment, that other program will be Tableau!

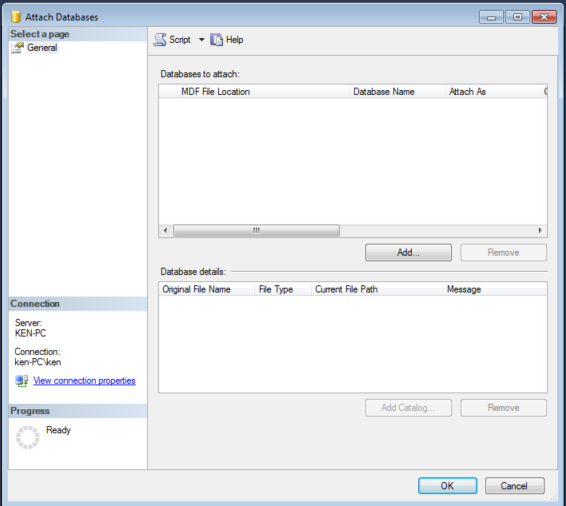
1. Open up SQL Server Management Studio. When asked to log into it, it will ask you for a server name. This will be RBB0113XX where RBB0113 is the room number. If you are in a different room number you will have a different prefix. Look on the top of your monitor to determine what the XX is in that string. If your machine already has a DB in it called AdventureWorksDW1012 in it then please proceed to step 2. If not then you must add the DW to your SQL Server on your lab machine by performing the following.
   1. Open up SQL Server and expand out the Databases by clicking on the + sign in front of it.



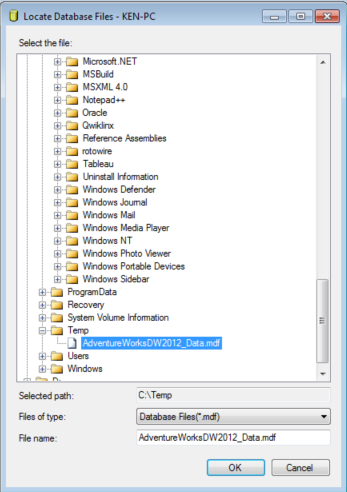
* 1. If you see the AdventureworksDW2012 in the list, skip to Step 2. If not proceed to the next step.
  2. Right click on Databases and then select “Attach”.



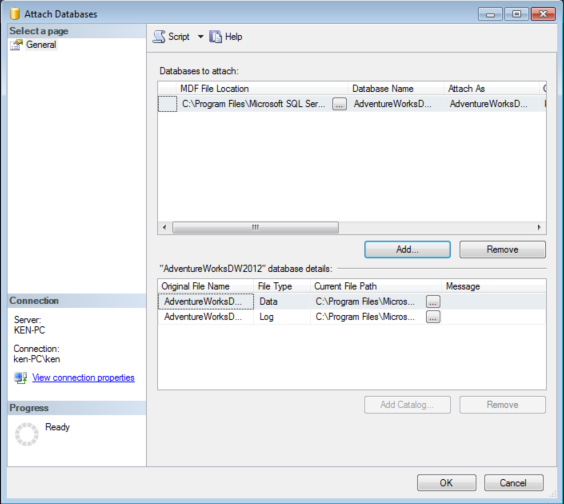
* 1. An Attach Databases dialog box opens up. Click on the “Add” Button.



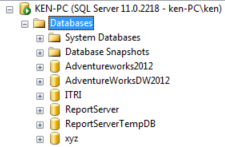
* 1. Your lab machine is set up to store all of the database files used by SQL Server in the C:\Program Files\Microsoft SQL Server\MSSQL11.MSSQLSERVER\MSSQL\DATA directory and we have inserted a complete version of this DW into this directory for your convenience (It’s all about you today). Your machine probably already defaults to this path but maybe not!. Click on the AdventureWorksDW2012\_Data.mdf (master data file) and then click on OK.



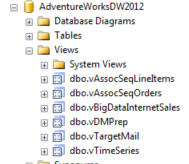
* 1. This will take you back to the Attach Databases dialog box where you will click on the OK button.



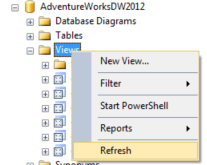
The AdventureWorksDW2012 database should now be listed in your list of databases under the Databases folder!



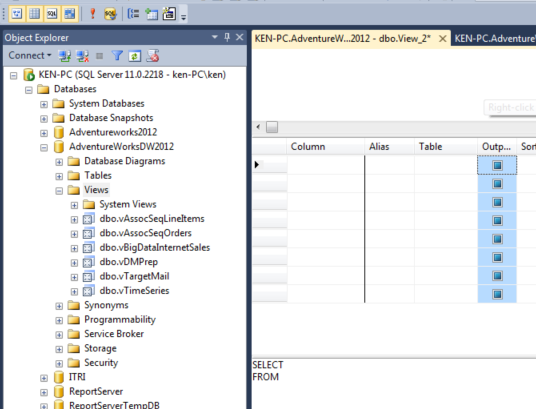
1. Create a View in the DW.
   1. Expand out the AdventuresWorksDW2012 folder and then expand out the Views folder



* 1. Add a view by right-clicking on the Views and then “New View”



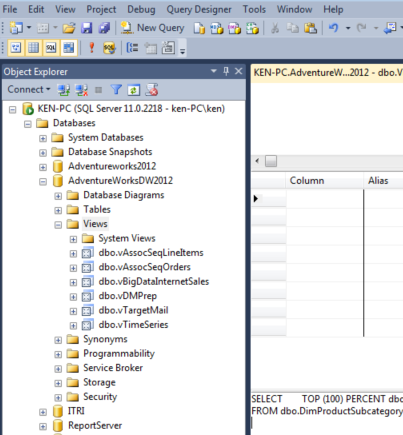
You can either spend hours creating your own view (not recommended unless you are adventurous!) or just continue on to the next step.

* 1. Close the Add Table Dialog box that pops up by default.
  2. In the view designer, highlight the “SELECT FROM” text in the 3rd pane and delete it.
  3. 
  4. In this pane, copy and paste the following query…

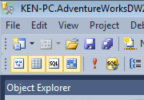
SELECT TOP (100) PERCENT dbo.DimCustomer.CustomerKey, dbo.DimCustomer.FirstName, dbo.DimCustomer.LastName, dbo.DimCustomer.BirthDate, dbo.DimCustomer.MaritalStatus, dbo.DimCustomer.Gender, dbo.DimCustomer.YearlyIncome, dbo.DimCustomer.TotalChildren, dbo.DimCustomer.NumberChildrenAtHome, dbo.DimCustomer.EnglishEducation, dbo.DimCustomer.EnglishOccupation, dbo.DimCustomer.HouseOwnerFlag, dbo.DimCustomer.NumberCarsOwned, dbo.DimCustomer.AddressLine1, dbo.DimCustomer.CommuteDistance, dbo.DimCustomer.DateFirstPurchase, dbo.DimGeography.City, dbo.DimGeography.StateProvinceName, dbo.DimGeography.CountryRegionCode, dbo.DimGeography.PostalCode, dbo.DimSalesTerritory.SalesTerritoryRegion, dbo.DimSalesTerritory.SalesTerritoryCountry, dbo.DimSalesTerritory.SalesTerritoryGroup, dbo.DimProduct.EnglishProductName, dbo.DimProduct.StandardCost, dbo.DimProduct.ListPrice, dbo.DimProduct.ProductLine, dbo.DimProductSubcategory.EnglishProductSubcategoryName, dbo.DimProductCategory.EnglishProductCategoryName, dbo.FactInternetSales.SalesOrderNumber, dbo.FactInternetSales.SalesOrderLineNumber, dbo.FactInternetSales.SalesAmount, dbo.FactInternetSales.OrderDate, dbo.FactInternetSales.ShipDate, dbo.FactInternetSales.Freight, dbo.FactInternetSales.OrderQuantity

FROM dbo.DimProductSubcategory INNER JOIN dbo.DimProduct ON dbo.DimProductSubcategory.ProductSubcategoryKey = dbo.DimProduct.ProductSubcategoryKey INNER JOIN dbo.DimProductCategory ON dbo.DimProductSubcategory.ProductCategoryKey = dbo.DimProductCategory.ProductCategoryKey INNER JOIN dbo.FactInternetSales ON dbo.DimProduct.ProductKey = dbo.FactInternetSales.ProductKey INNER JOIN dbo.DimCustomer ON dbo.FactInternetSales.CustomerKey = dbo.DimCustomer.CustomerKey INNER JOIN dbo.DimGeography ON dbo.DimCustomer.GeographyKey = dbo.DimGeography.GeographyKey INNER JOIN dbo.DimSalesTerritory ON dbo.FactInternetSales.SalesTerritoryKey = dbo.DimSalesTerritory.SalesTerritoryKey AND dbo.DimGeography.SalesTerritoryKey = dbo.DimSalesTerritory.SalesTerritoryKey

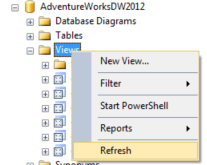
* 1. Execute the view by hitting the F5 key or by clicking on the red exclamation point.



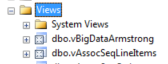
* 1. The view should execute and render the results of the query in the bottom pane of the view design page.
  2. Click on the disk icon on the toolbar and then save the view as “vBigDataLastName” (the view at the start of the name stands for view)



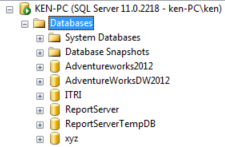
* 1. Right click on the Views folder and hit refresh



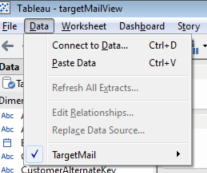
* 1. Your view should now be located as an item in this folder.



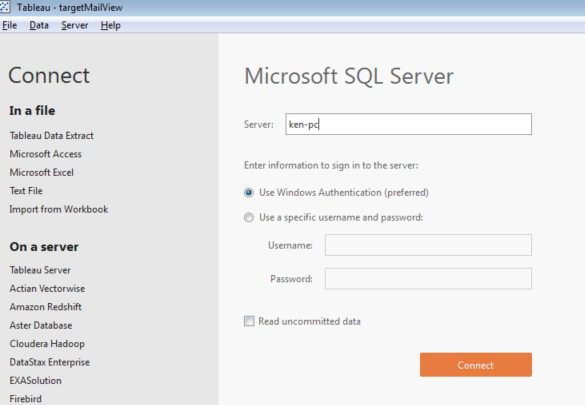
1. Create Data source in Tableau that points to this new view of your data in your DW.
   1. While you are still looking at your SQL Server databases, note the name of the SQL Server name on your machine. On my machine, it is named “KEN-PC” in the following location (yours is probably named RBB0113XX where RBB0113 is the room number where the X is your machine’s number:

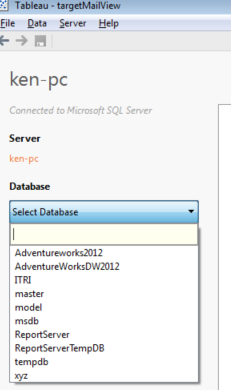


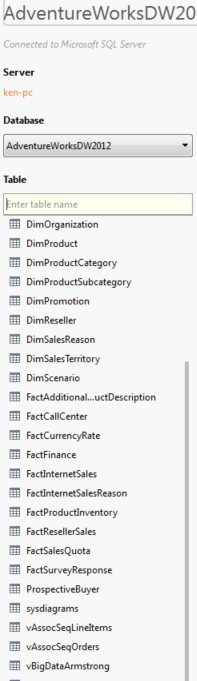
* 1. Open up Tableau and click on “Data” and then “Connect to Data”



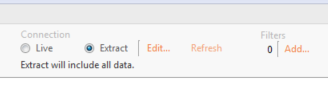
* 1. Scroll down to the “On a Server” section and then select Microsoft SQL Server
  2. The following dialog box pops up. Enter in the name of your SQL Server name that you found in step a) above: In the classroom, this is probably RBB0113XX where RBB0113 is the room number. Then click on “Connect”



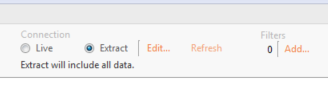
* 1. Select your DW from the “Select Database” dropdown list.
  2. Find your view in the list of tables. You named it something like: vBigDataLastName



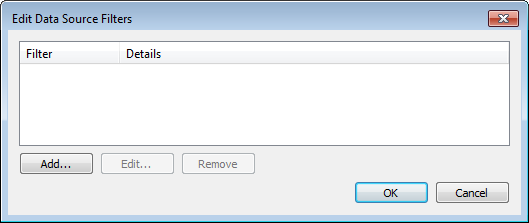
* 1. Drag this view into the top pane of the page that says “Drag Tables Here”
  2. In order to embed this data into your Tableau workbook, you will need to set it up so that Tableau extracts the data from your query and then embeds it into your workbook so that you can work on it at home on a machine that DOESN’T have SQL Server on it! To do this click on the Extract radio button on the upper right corner of Tableau.



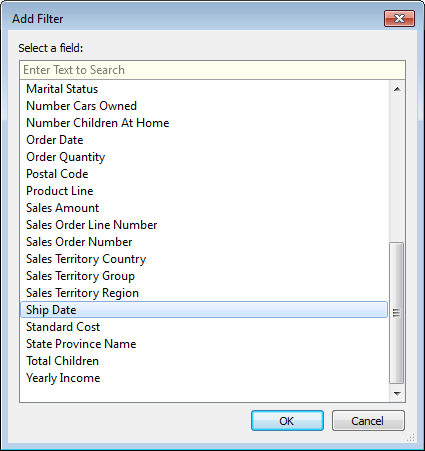
* 1. Your view has just a few sales in the last 2 months of data. If you include them, in most of your visualizations, it will look like your sales went into the tank for July and August of the last year so we need to exclude them. To do this, we can apply a filter at this data source extraction level so that our data source does not even have them. By filtering the datasource itself, you won’t have to do it on EVERY chart that you make! Click on “Add” underneath the “Filters” area of Tableau.



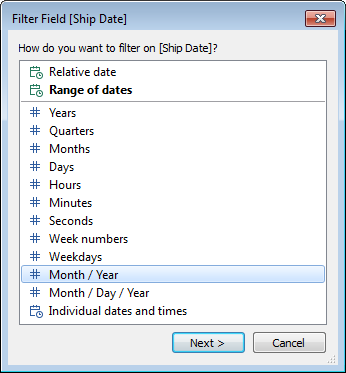
* 1. This opens up the Edit Data Source Filters dialog box. Add a filter by clicking on the “Add” button.



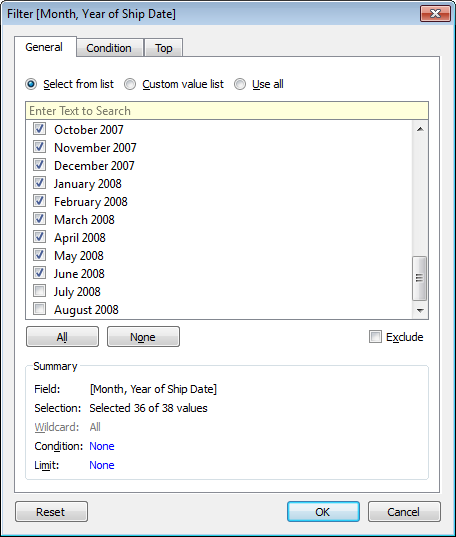
* 1. Select “Ship Date” from the list of fields:



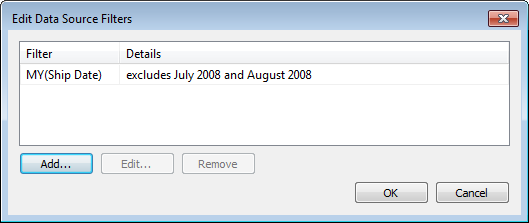
* 1. Select “Month/Year” from the Filter Field dialog box and then click “Next”



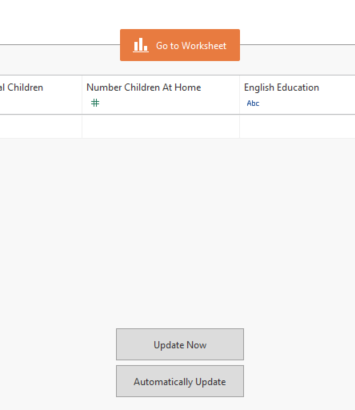
* 1. And then select all of the months except for July 2008 & August 2008 and then click on the OK button.



* 1. Click on OK on the Edit Data Source Filters dialog box.

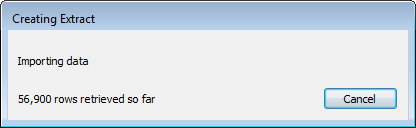


* 1. You can click on the Update Now button to see if your data has made it to your view.

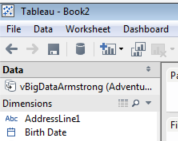


* 1. When you are happy, click on the Go to Worksheet button on the above image.

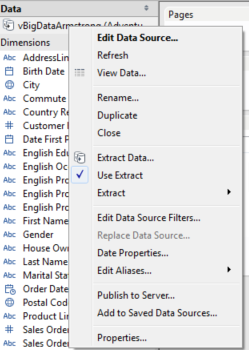
1. You will now be asked to save your Tableau data extract (tde) file. Save it where you will remember it!
2. You will see the following box pop up as it gradually copies all of data from your View into your Tableau workbook!



1. Your Datasource now listed in Tableau… Verify that your data made the trip.
   1. Right click on the Data source in Tableau

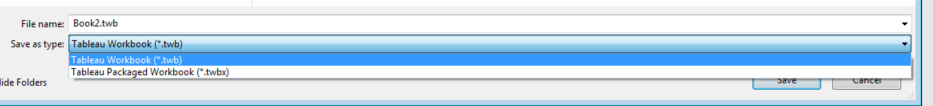


* 1. Select “View Data”



* 1. By default, the data viewer only shows the first 10,000 rows. Change this 10,000 to 100,000 and see how many rows that you have. Take a couple of minutes to look over your columns of data and see what each row looks like and what it is capable of telling you. You can also have Tableau “describe” your data Dimensions and Measures to show you what your data looks like.

1. **Finally, save your file by clicking on “File” and then “Save”. Give your file a name of A2-LastName FirstName and MAKE SURE to change the file type to Tableau Packaged Workbook (twbx) as opposed to the default file type of Tableau Workbook (twb).**



**If you don’t save it with this file type, the data will not be embedded into your file, you won’t be able to work on it anywhere else and WE WON’T be able to GRADE IT!**

1. Whew… Good Job! You are now ready to use Tableau to answer all of the questions/problems contained in Assignment 2. Please proceed to that assignment now.