Nibret Daba

INFM600 Information Environment – fall 2015

**Information Organization Assignment**

This document is a full-workflow of how to explore data. By analyzing through the data, the intention is to find out if issuing more liquor license has any relation with violence and vehicle accident in Howard County, MD.

**Step 1: Finding Data Sets**

* 1. **Data Sources**

Three datasets are identified from three different data sources. Here below is list of the data source.

1. Howard County Open Data Portal: (<https://opendata.howardcountymd.gov/>)
2. State of Maryland Open Data Portal: (<https://data.maryland.gov/>)
3. Comptroller of Maryland Online Database: (<https://interactive.marylandtaxes.com/webapps/licprt/user/ilu_QueryRetailer.asp>)
   1. **Data Set**

Howard County Data Portal, Howard County Maryland. (2015). *Howard County Police Department Call For Service: 2014*. Retrieved from <https://opendata.howardcountymd.gov/Public-Safety/Howard-County-Police-Department-Call-For-Service-2/qccx-65fg> on 2015, September 29.

This dataset contains records of calls that are made to request for police servicing. The data contain reason of the call, date, location, time, statistical reporting area and beat.

Maryland Government, Department of Information Technology. (2013). *2012 Vehicle Collisions Investigated by State Police*. Retrieved from <https://data.maryland.gov/Public-Safety/2012-Vehicle-Collisions-Investigated-by-State-Poli/pdvh-tf2u> on 2015, September 23.

This dataset contains records of vehicle collisions investigated by state police. The data has accident date and time, location, county, days of the week, road among other things.

Comptroller of Maryland. (2015). *Alcohol and Tobacco Tax, Alcoholic Beverage Retail Licenses* [Database Record]*.* Retrieved from Alcoholic Beverage Retail Licenses Database.<https://interactive.marylandtaxes.com/webapps/licprt/user/ilu_QueryRetailer.asp> on 2015, October 04.

This dataset contains records of alcohol and tobacco tax payers’ information that have active licenses. The result from the database query contains tax registration number, county, date, corporate and trade names of the businesses.

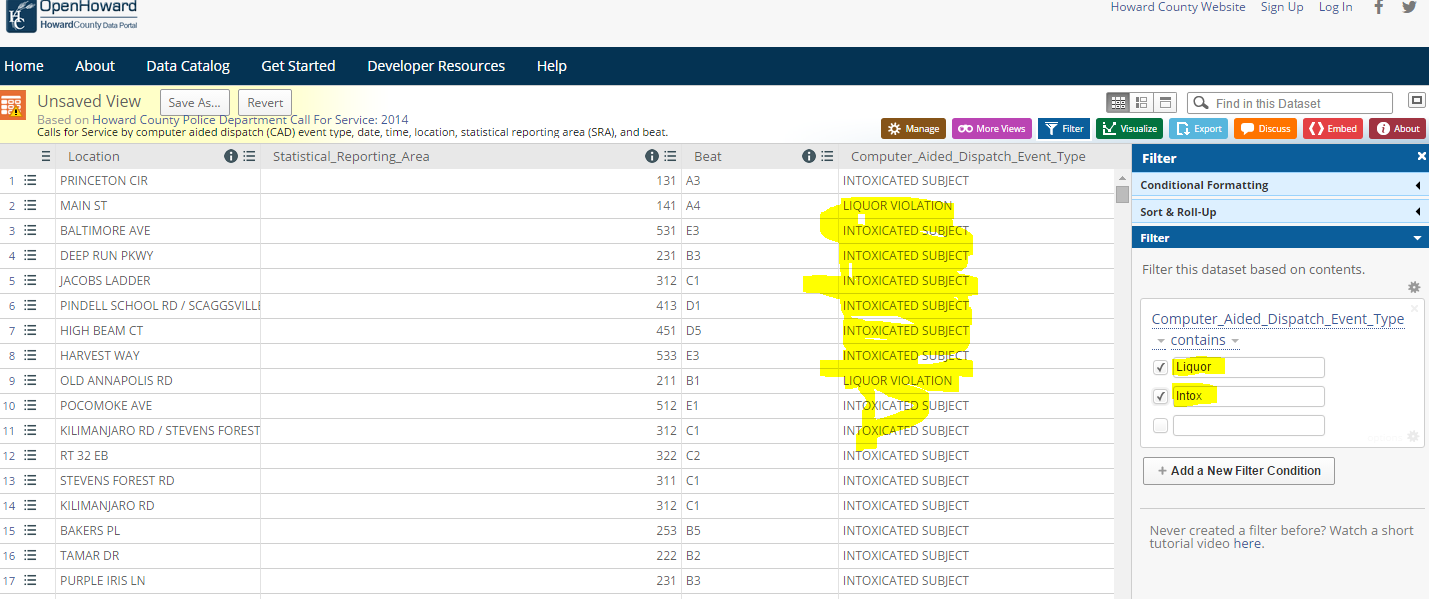
**Step 2: Formulating Research Question**

The proposed research question based on the dataset at hand would be “Does issuing more liquor license aggravate violence and vehicle accident in Howard County MD?”

**Step 3: Data Filtering**

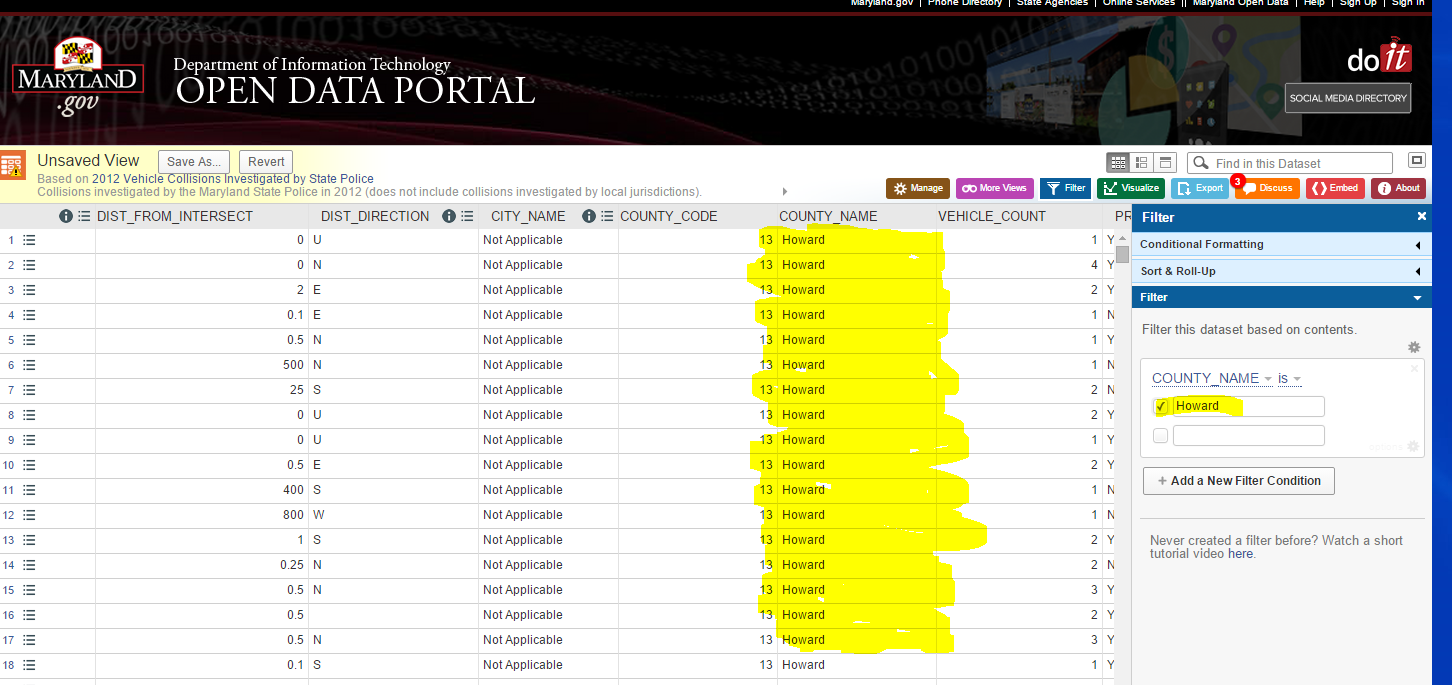
**3.1.** Dataset 1: Get records where Reason of call is **Liquor Violation and Intoxicated Subject.**

**Technique:** I used the “Filter” feature on the portal to accomplish this. Below is the screen shot showing the filters used and the result (Highlighted Yellow)



**3.2.** Dataset 2: Get records showing collision that happened only in Howard County, MD.

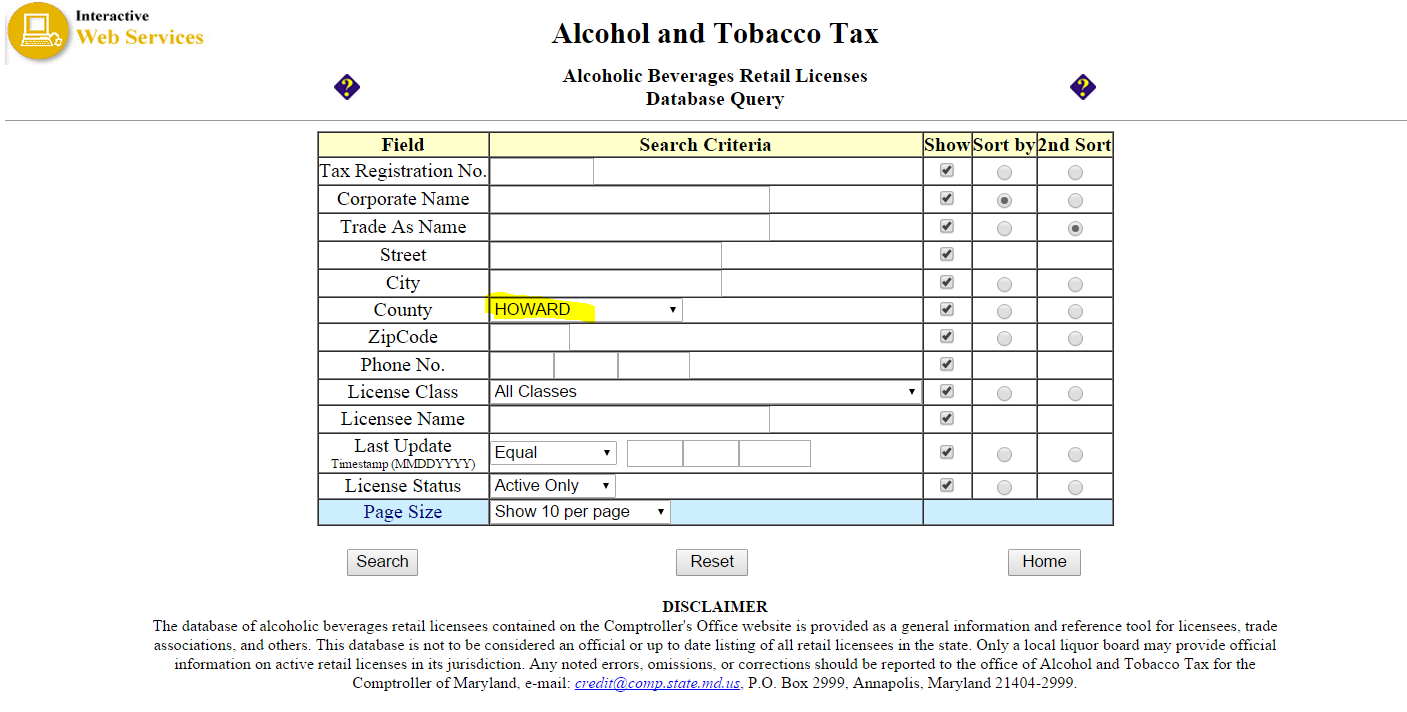
**Technique:** “Filter” feature of the portal is used. Below is the screen shot.



3.3. Dataset 3: Get records showing alcohol and tobacco business for Howard County

**Technique:**  Querying the Database choosing “Howard” from the drop down list for County.

Screen shot:

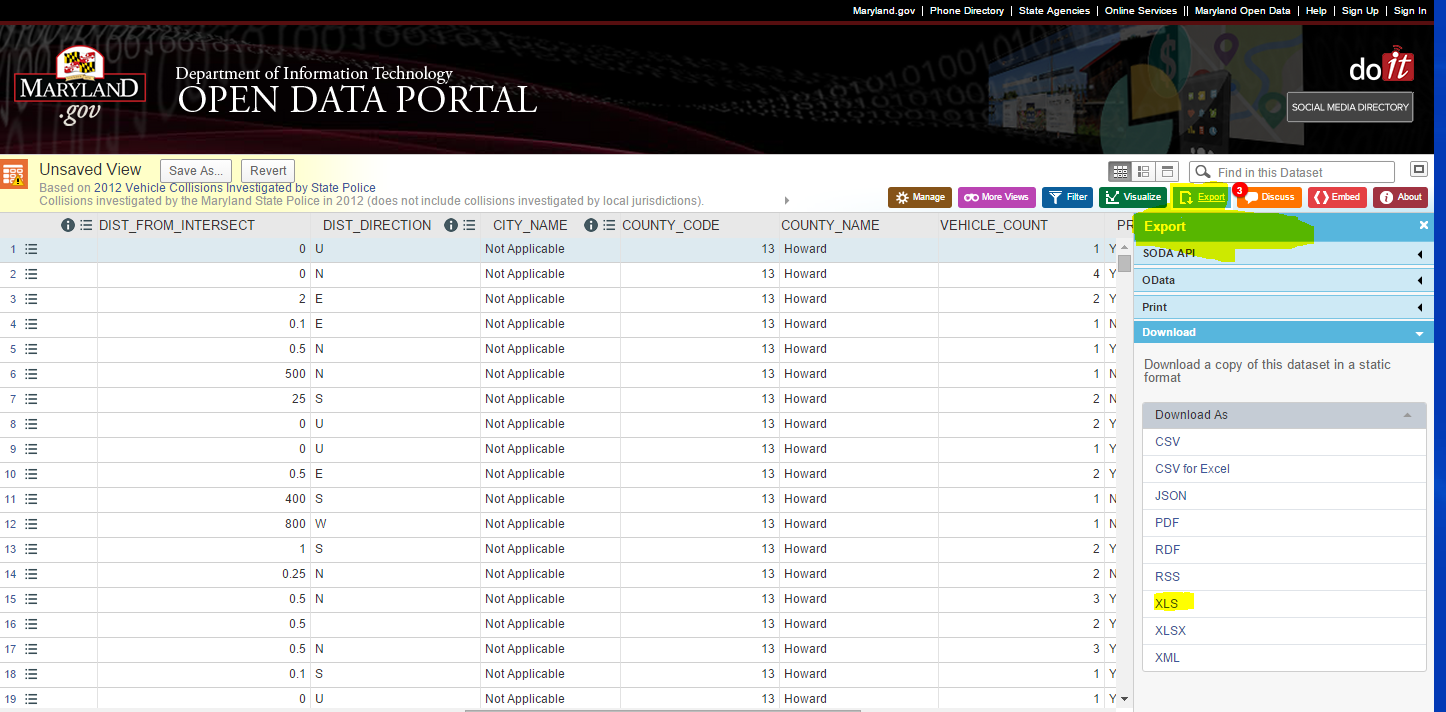


**Step 4: Exporting and Cleaning Data**

* 1. Exporting filtered data sets to Excel.

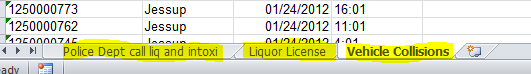
**Techniques:**

* + 1. For data set one and two, I used the **“Export”** feature of the data portal. Here below is the screen shot.



* + 1. For data set three, after I run the query against the database, I selected the whole record and pasted into excel

Here below is a screen shot where the three datasets exported into excel.



* 1. **Cleaning Data**

**Target:** Taking out extra spaces and irrelevant characters from the Location field of each dataset. This is because location is the common field between the three datasets and is a key field to merge the datasets.

**Technique:** Excel **Trim and LEFT** functions is used.

**Result:** The three data sets have common column with the same structure of data to join on.

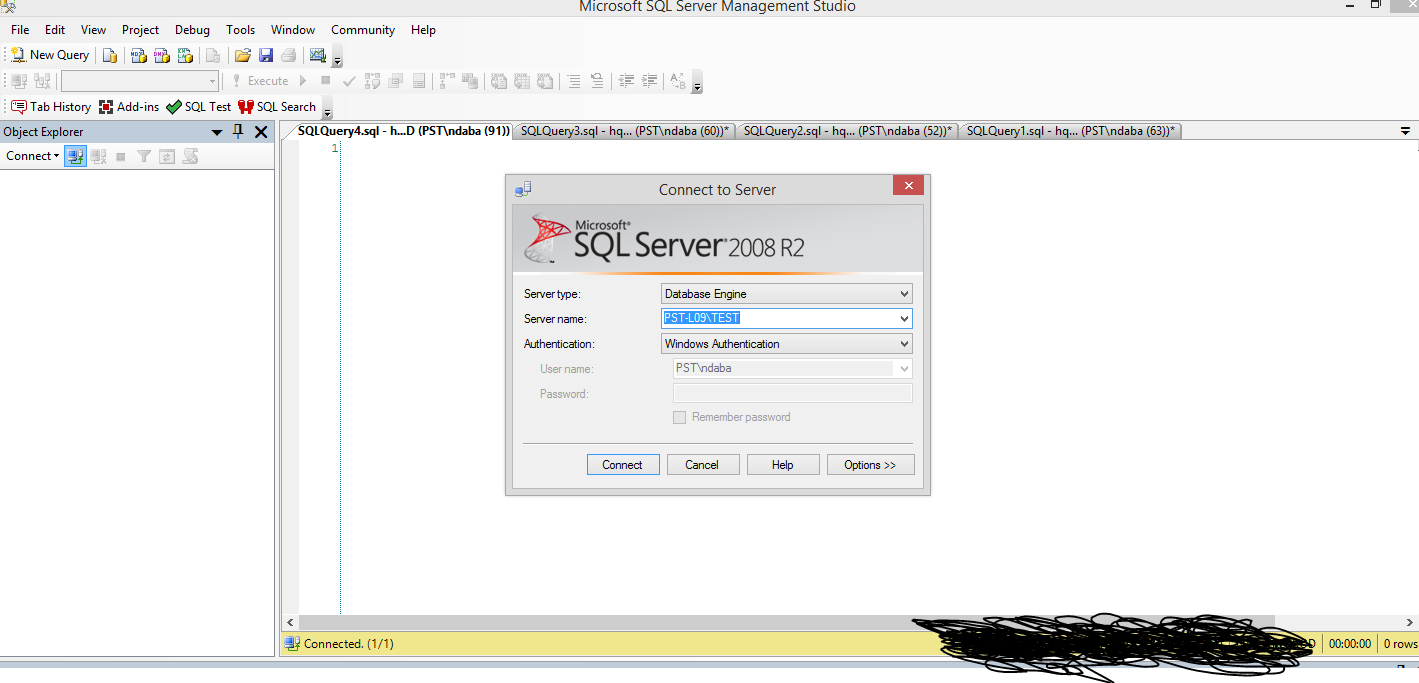
**Step 5: Merging Data**

The merging between the three data sets is accomplished using SQL script. For that, the datasets are imported to SQL Server into three separate tables. The following steps and screen shots illustrate how the datasets get into SQL Server.

* 1. Importing from Excel to SQL Server

**Techniques:**

* + 1. Connecting to SQL Server

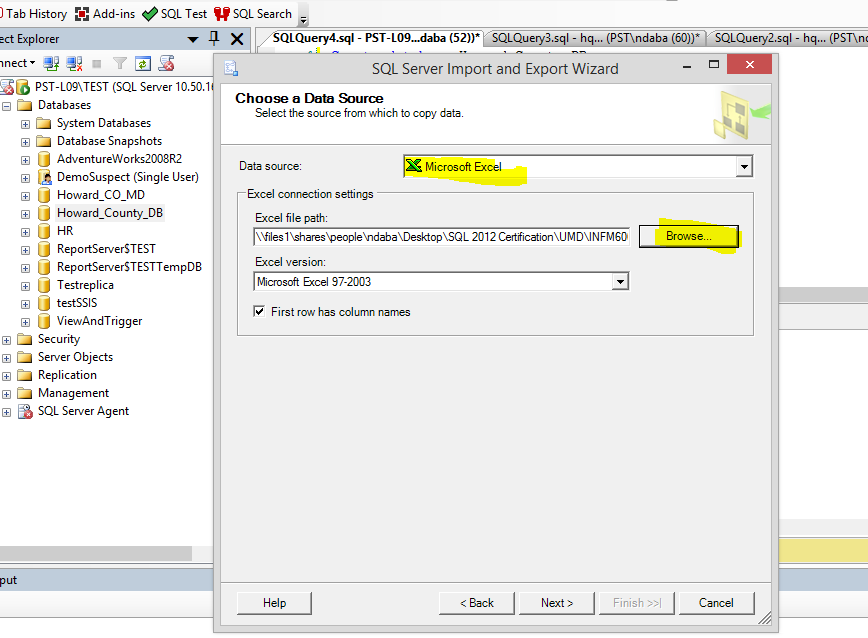


* + 1. Creating a database and importing the excel files

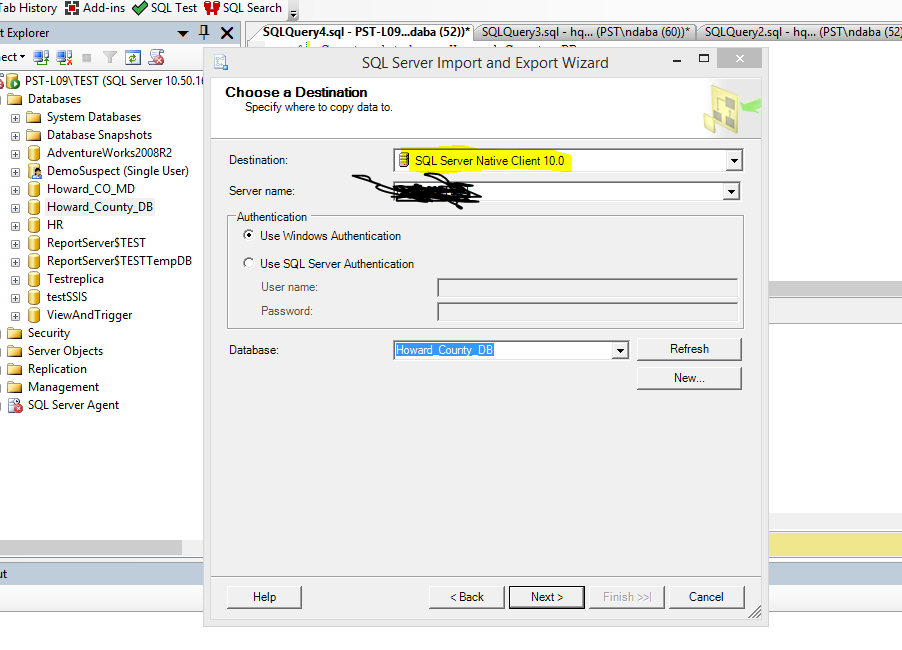
Database Creating Script:

Create database [Howard\_County\_DB]

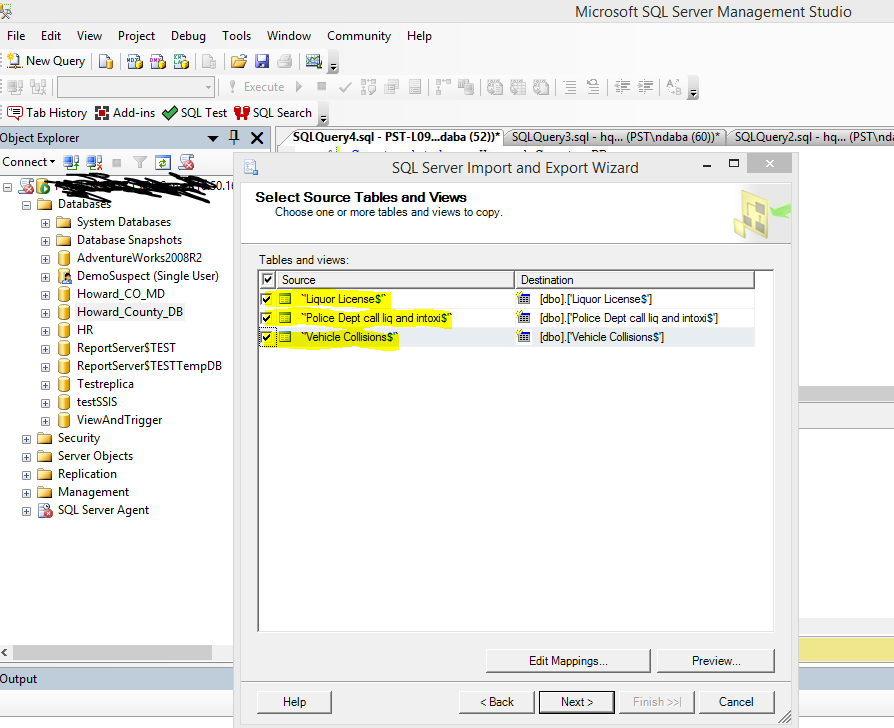
* + 1. Importing the datasets to the newly created database
* Right click on to the database ([Howard\_County\_DB]>>>Task>>>Import Data
* Pass the welcome page
* Choose Microsoft Excel as a data source
* Browse to the excel file from 3.1



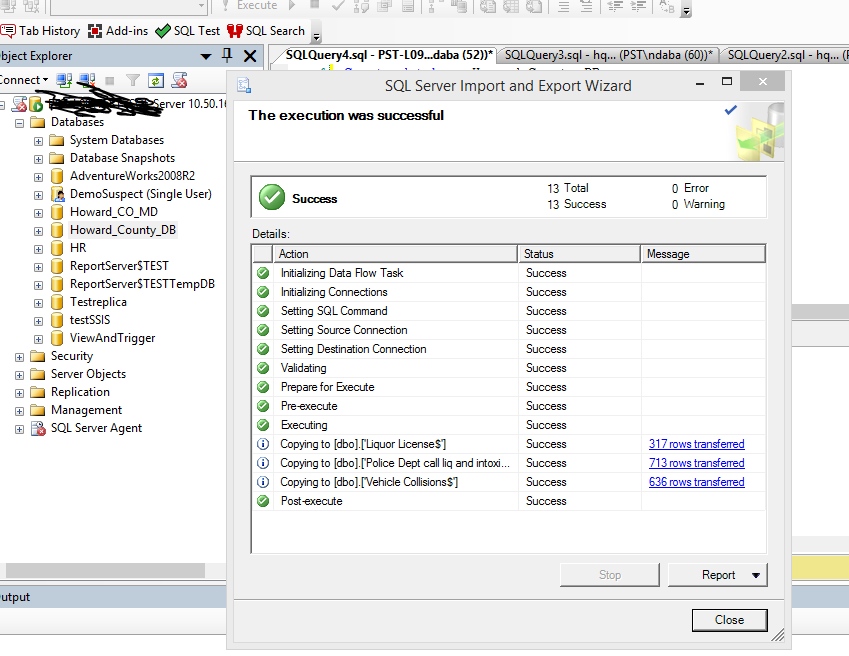
* Choose SQL Server as a destination

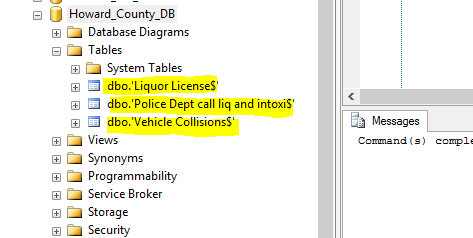


* Choose the three data sets to import



* Successful Import to SQL Server



* The three data sets are imported to SQL Server tables
* 
  1. Write a Query to Join the three tables

use Howard\_County\_DB

Select \*

from dbo.['Liquor License$'] as a

Join dbo.['Police Dept call$'] as b

on a.[StreetAddr1] = b.[Location]

join dbo.['Vehicle Collisions$'] as c

on a.[StreetAddr1] = c.[INTERSECT\_ROAD]

**Documentation:**

1. Write the README file
2. Creating Github account
3. Upload the following documents to a Github Repository:
   1. README file
   2. The Processing Document
   3. The final dataset
   4. The original three datasets