Course Project - Course Project Final

Shaun Pritchard

Rasmussen College

QMB4200

Terry Brookhouser

November 10, 2020

**Final Course Project**

This final Project I implement my skills as analytics Developer for a state government of Maryland, to predict police officer recruiting based on the historical crime data. My implementation as designated based of the given factors. Also, I created and changed database tables using SSMS in Azure so that both tables can be joined in a query. With both a column name that was identical and the same data type.

* SAS Scripts
* ETL job to load data into the database
* Number of crime incidents by City
* Apply scripting standard
* Test SAS script
* Optimizing SAS script used to read data from Cloud Database
* Applied Script standards

**Course Project - Real Time Analytics Case Study**

The initial purpose of the data is to analyze the specific requirements needed to hire more officers in given areas. We can use the data to show in specific locations crime rates, types, frequency, and locations of crimes. The data provides several key factors that will help to determine the number of police officers required to patrol specific areas. Initially, we would need more police officers in areas where a crime was higher in the variable of crime was more violent as opposed to nonviolent.

We would also need to discern which districts city agencies in sectors would need an increase of police officers to justify the rate and frequency of crimes in those specific areas. Initially want to hire more police officers in specific areas to lower the workload and pressure for officers. allowing them more sustainability-based on their initial duties which would increase their overall performance on each officer and raise their overall delivery of their performance of their duties.

The data can also be used from a historical perspective to determine if different seasons within the time frames of the year are more prone to criminal activity. How much more officers would be needed, and other areas are districts based on the historical crime data.

Essentially, some of the missing data might be required to further analyze the prediction of hiring more police officers. This would be to determine the categories of violent and non-violent crimes in the given areas, incident reports based on officers, number of officers employed per district, officer's performance, solved cases, unsolved cases, known reoccurring offenders, and potential ongoing crimes are given the variable dates and time. These factors would facilitate more sound information that could potentially resolve the issue of Manpower and hiring more officers for the given areas.

To justify reasoning in this additional data sets based on the factors provided. Knowing the categorical data label of violent or non-violent to facilitate in a higher frequency in Need for police officers in areas. For instance, if there are more violent crimes in one District as opposed to the other then we know that we need more police officers in that area.

Again, if there are more traffic violations are non-violent crime in other districts, we can assume that we need more officers of specific values such as beat cops or Highway Patrol officers, even officers who process internal data.

If we had the values on the performance of individual officers, then we can deduce whether or not those officers in a given area need more help than others.

If we understood the number of salt cases versus dog cases in the given areas, we could allocate more resources to solving crimes in those areas with more unsolved cases.

If we had a list of known reoccurring offenders, this would help us associate whether or not crimes in specific areas we need to yield more investigations there for needing more officers to conduct investigations such as Detectives.

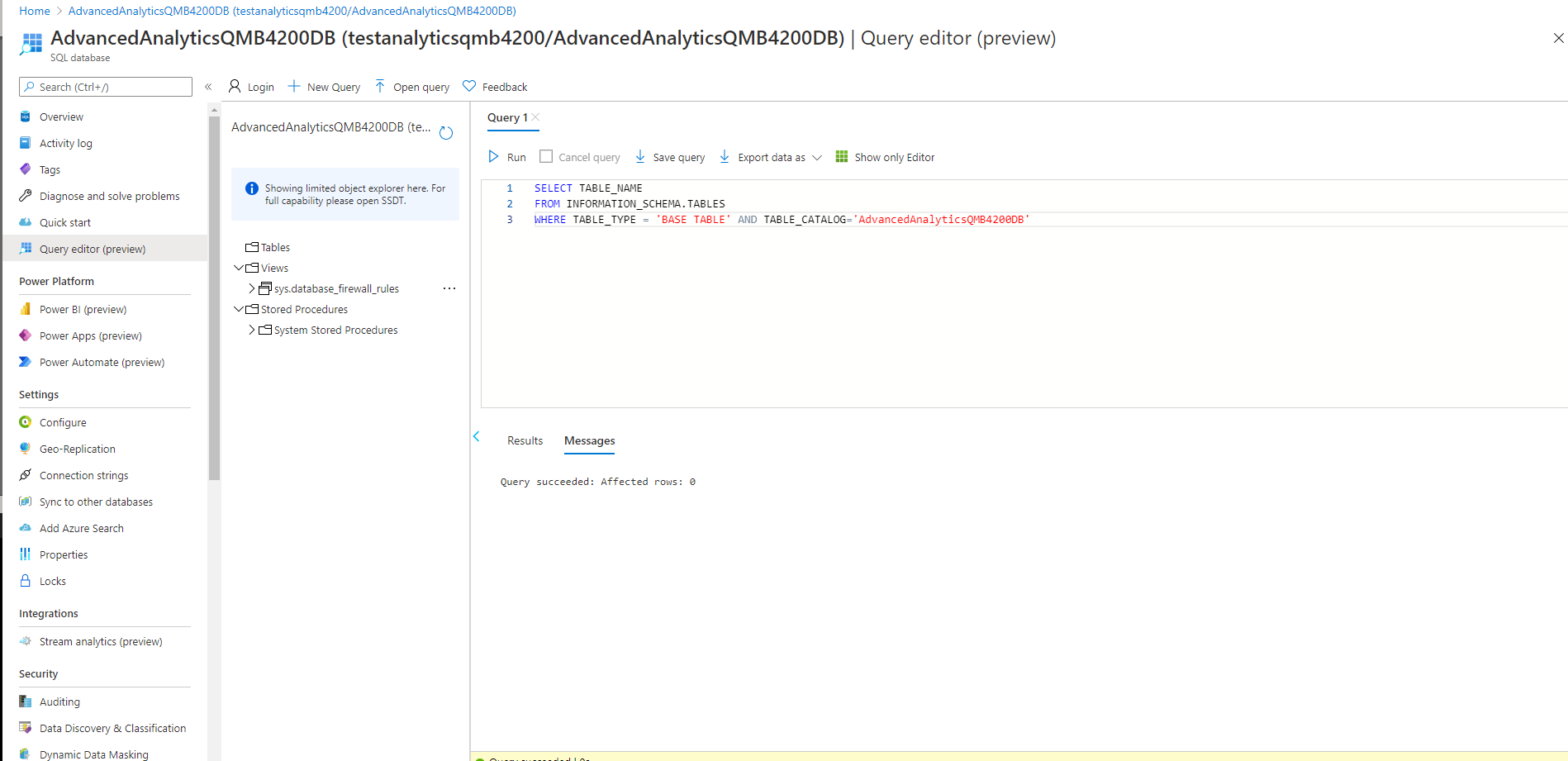
Also, knowing the potential ongoing crimes we can evaluate the need for more officers to facilitate handling these ongoing crimes in an ongoing manner. The data that we do have provides dates and times for specific crimes, but we do not have the data points that determine whether or not these crimes are ongoing violent or non-violent. So initially this categorical data would facilitate a need to associate the needed hiring of more officers.

**Course Project - SQL Server With Business Intelligence**

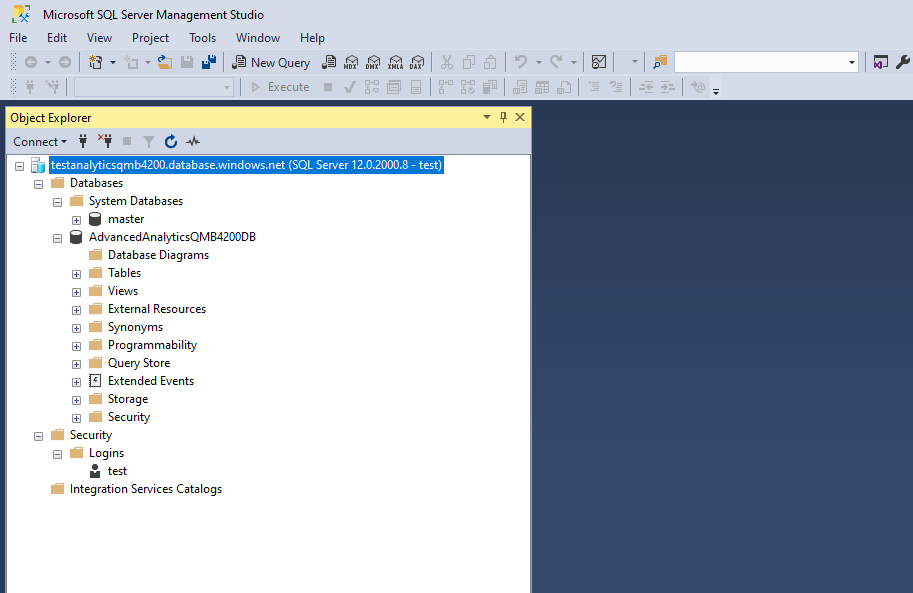
Here is the login to the created SQL server and Db.

*Note: I had to configure security system permissions to access it. Then I ran a query to test the database.*





SQL Server Studio login successful:



**Course Project - Write ETL Job and Migrate data into the Database**

I initially tried loading all the data columns. I ran into issues based off my character lengths allocation and other errors as posted below. I did successfully append 3 columns of data.

Here are the errors I ran into:

***TITLE: Package Validation Error ------------------------------ Package Validation Error ------------------------------ ADDITIONAL INFORMATION: Error at Data Flow Task [OLE DB Destination [600]]: The number of input columns for OLE DB Destination.Inputs[OLE DB Destination Input] cannot be zero. Error at Data Flow Task [SSIS.Pipeline]: "OLE DB Destination" failed validation and returned validation status "VS\_ISBROKEN". Error at Data Flow Task [SSIS.Pipeline]: One or more component failed validation. Error at Data Flow Task: There were errors during task validation. (Microsoft.DataTransformationServices.VsIntegration) ------------------------------ BUTTONS: OK ------------------------------***

Second error I had:

***Message: SSIS Warning Code DTS\_W\_MAXIMUMERRORCOUNTREACHED. The Execution method***

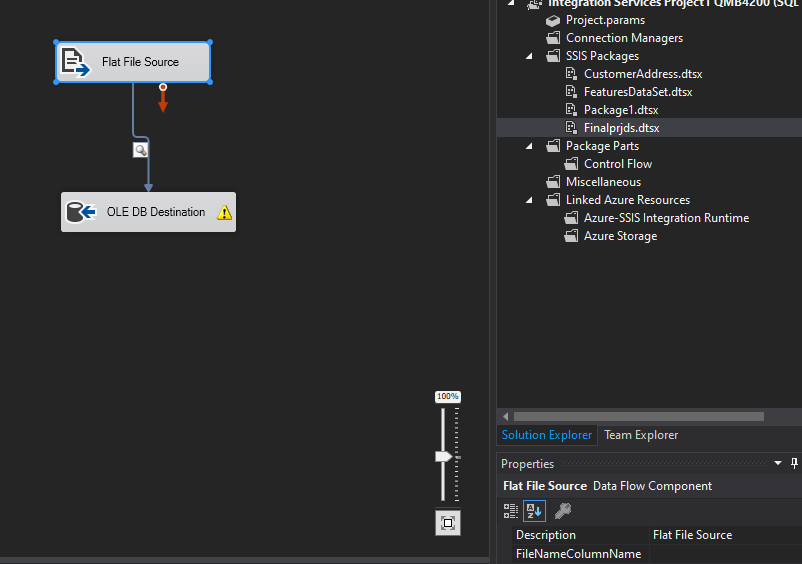
***succeeded, but the number of errors raised (1) reached the maximum allowed (1); resulting in failure.***

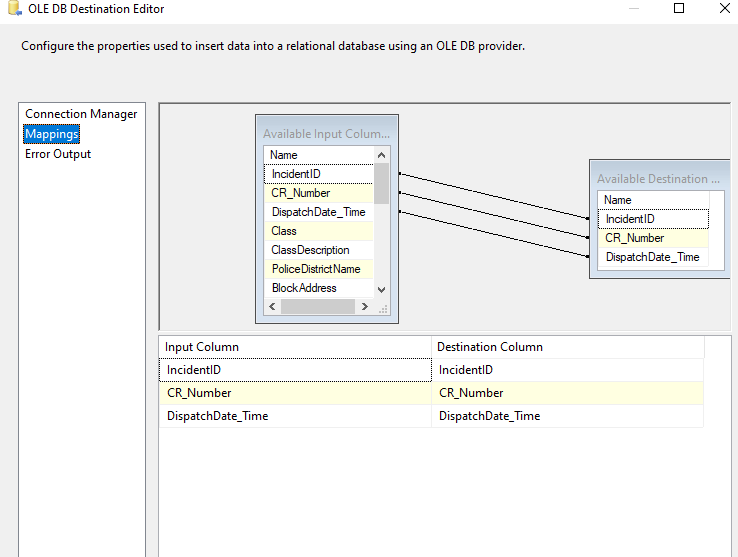
***This occurs when the number of errors reaches the number specified in MaximumErrorCount.***

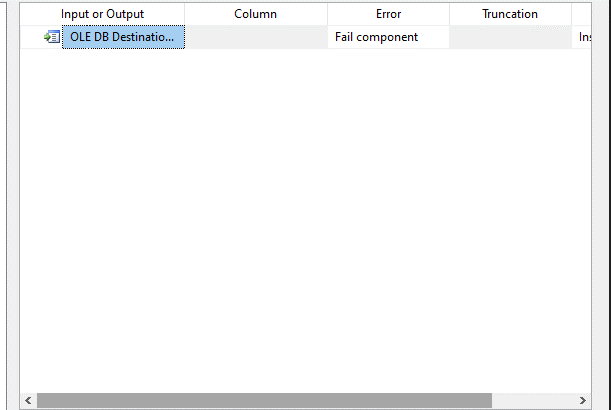
***Change the MaximumErrorCount or fix the errors.***

So, I restarted bot SQL studio and Visual studios and I got it to work. It seems every time I try to update my database tables, but I would refresh the object Explorer the old table was still there. I had to restart the entire program empty out the cache.

**SSIS Packages**



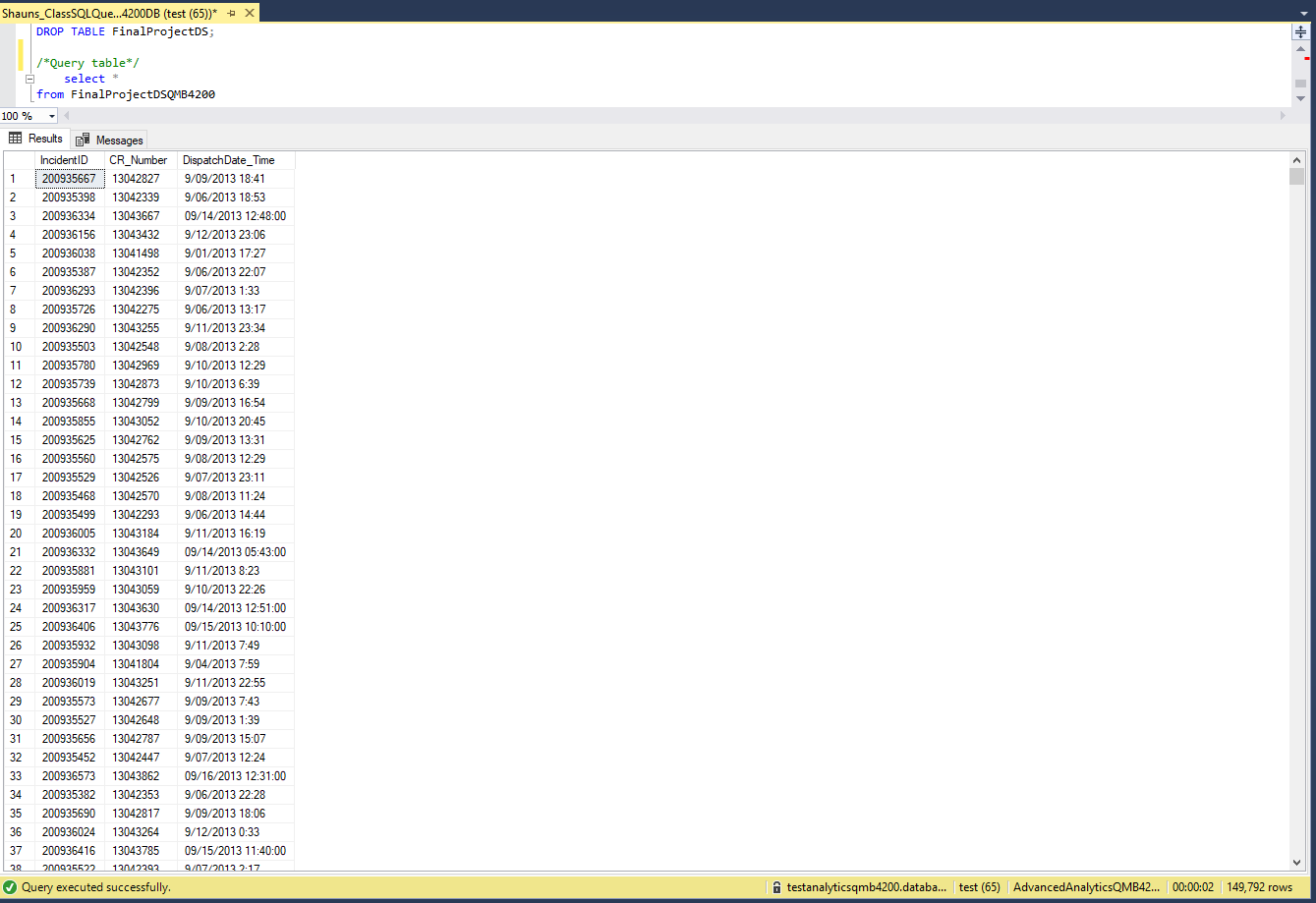


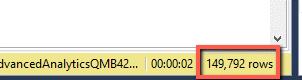


**SQL Table queries:**

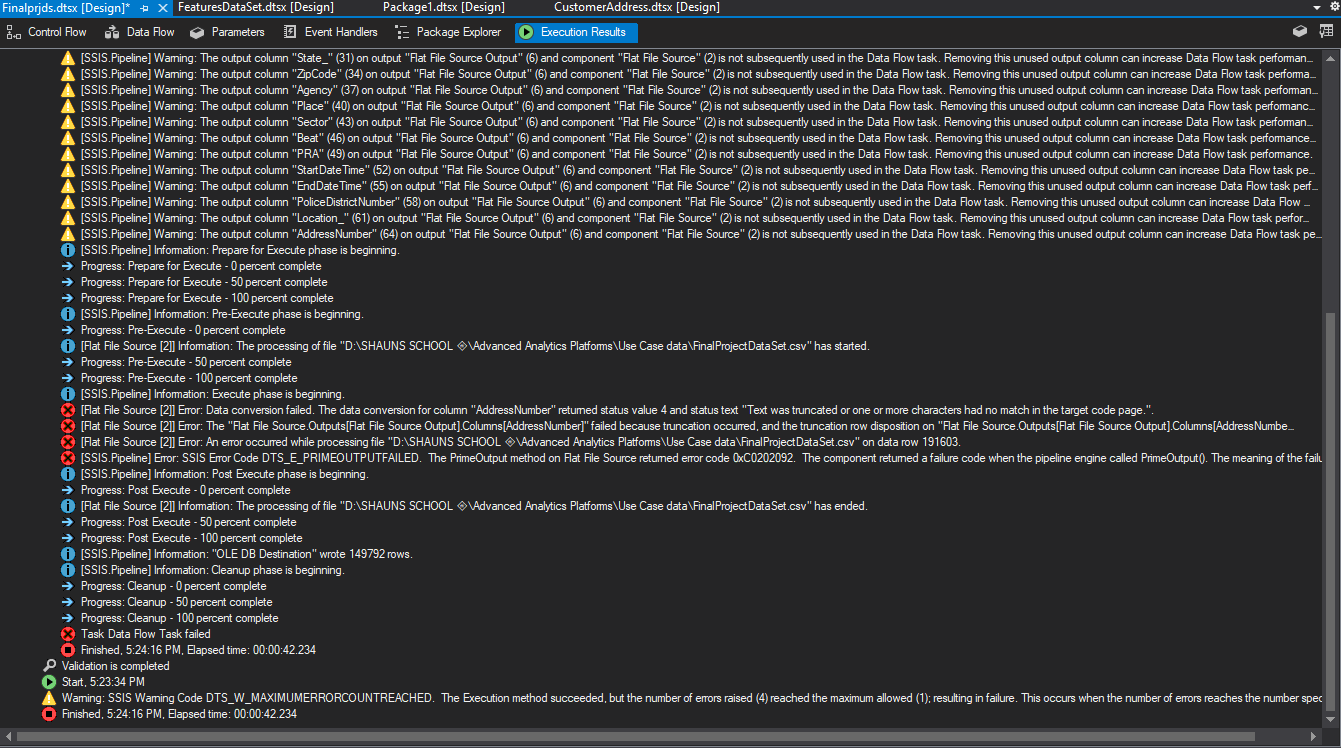


**Data results:**





**Execution results:**



Note:

I did however try to resolve some of the issues by looking up the error codes I was really stumped by the maximum error count reach. Then I can find is that the table name I was creating maybe it conflicted with what I was needing my SSIS package?

But I was aware that there was a lot of inaccurate data in the final data project there is chunks of missing data there was Data with symbolic characters, more than 50 characters, and several other things I noticed.

1. Are we supposed to uncheck all the available external columns from The Source assistant that we are not using before we connect it to the source?
2. In order to Output columns of the table that we built correctly how do we handle Error output on the destination?
3. What is the maximum error count?

**Course Project - Perform Analysis using SAS Procedures**

In this lesson I used several scripts to calculate and run statistics:

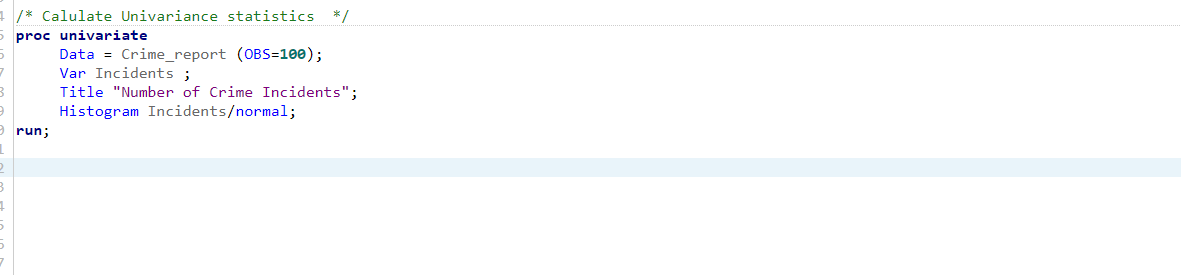
* Step 1: Create a SAS script and read the data from the SQL Table(FinalProjectDataSet.csv)
* Step 2: Use SAS proc methods and calculate number of crime incidents per zip code, mean score, Std Dev
* Step 3: Determine the police officer count based on the following the logic.

For this I used 2 scripts because I was not sure if I was supposed to use the Zip code data and calculate the frequency of duplicates where crimes occurred to apply the statistics and calculate the police officer count. So, I followed the example in class. I also imported the data set file and ran statistics directly to accumulate the lesson parameters.

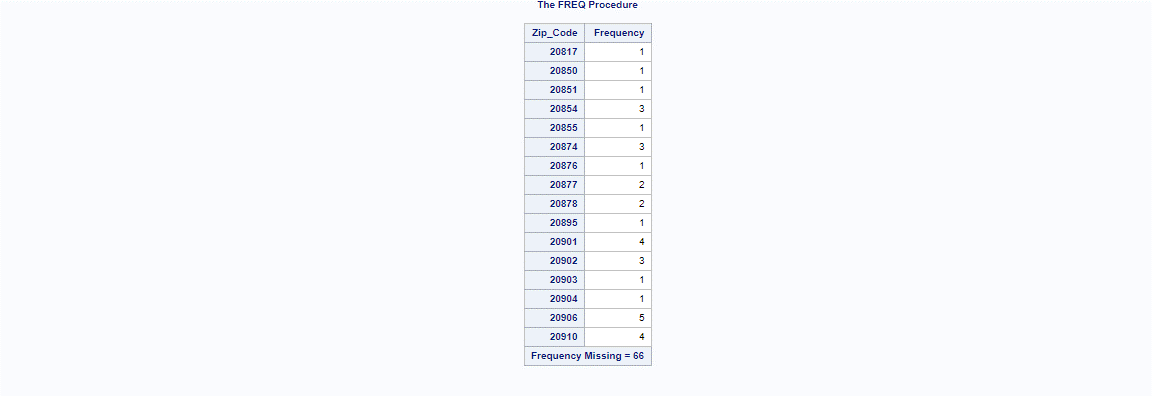
**Script 1** – My custom Script

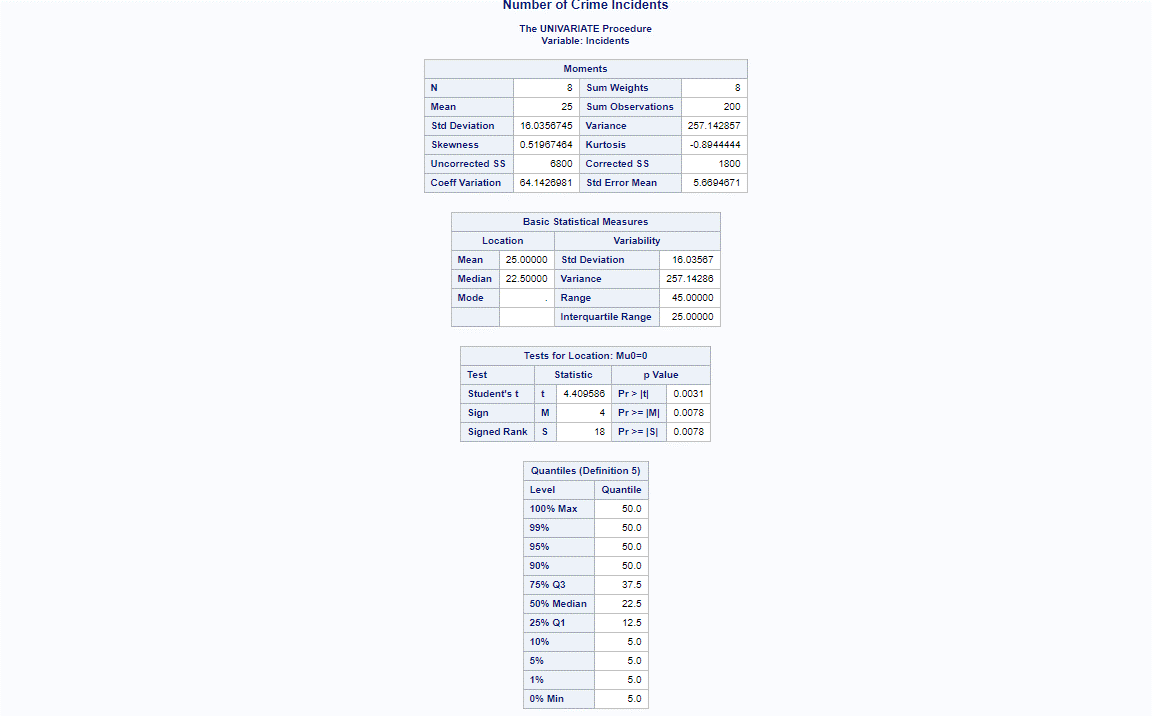
**Script 2** - Custom scripts based off in class example

**Script 1 – Code:**

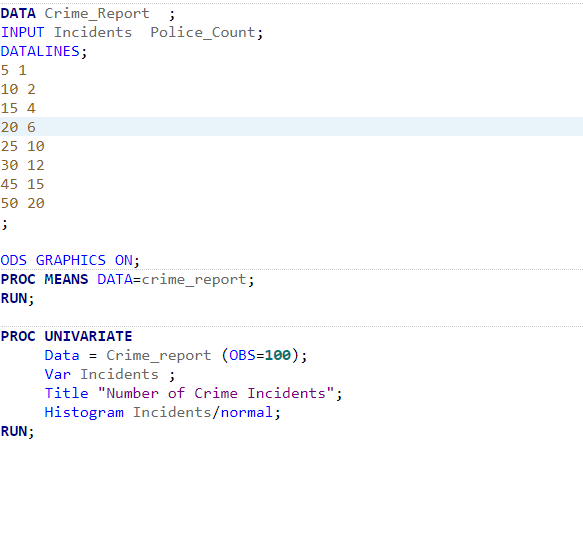
 

**Script 1 – Output:**



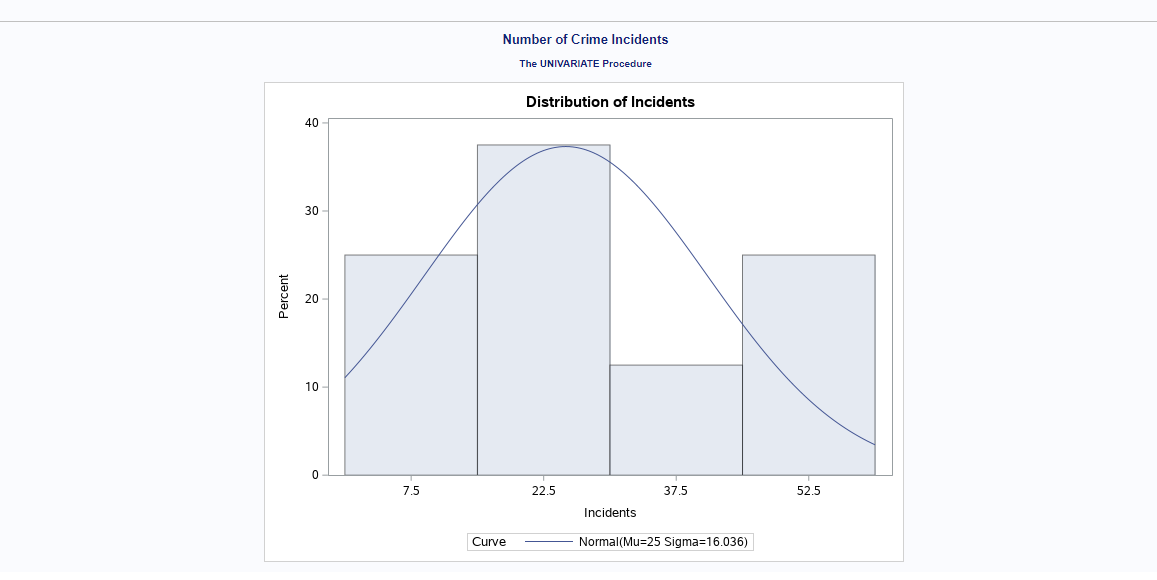


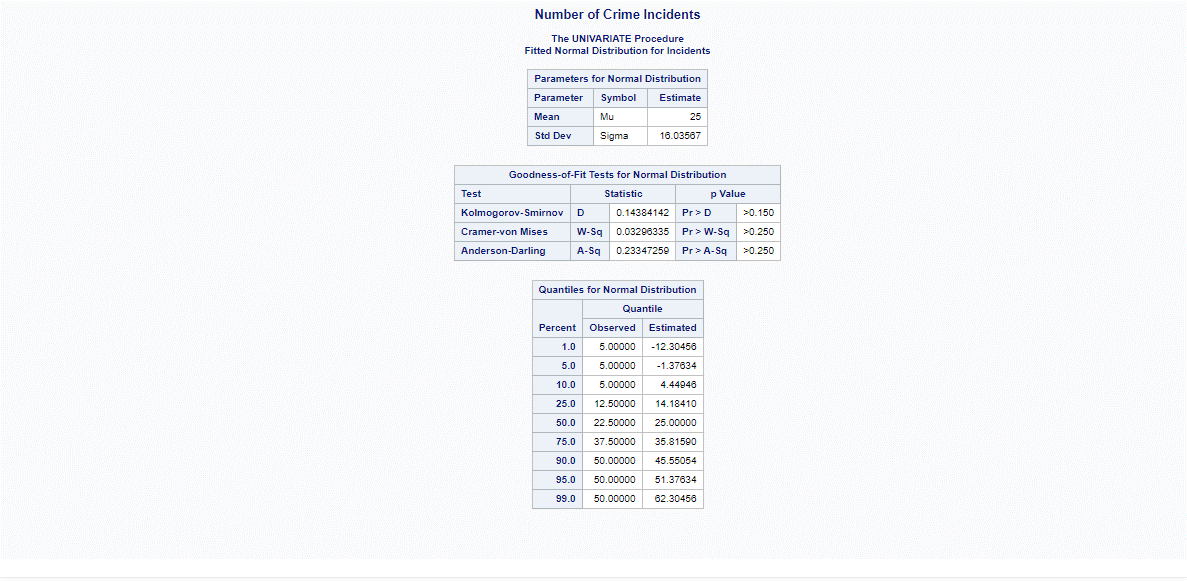
**Script 2 – Code:**

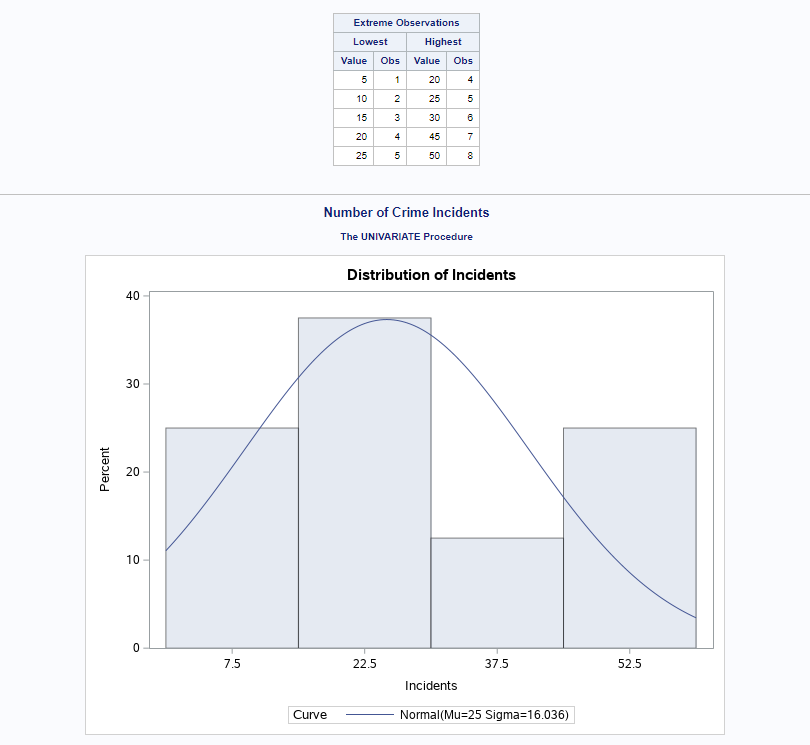


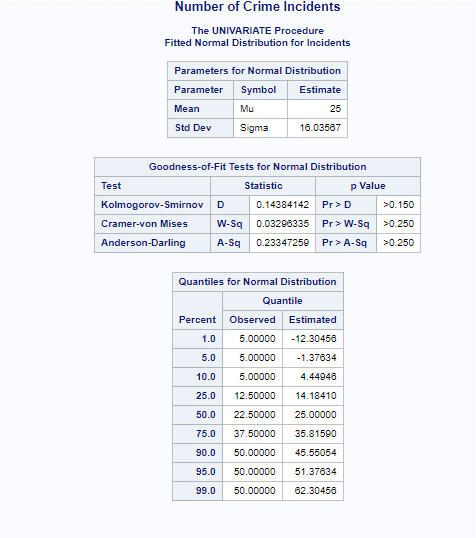
**Script 2 – Output:**







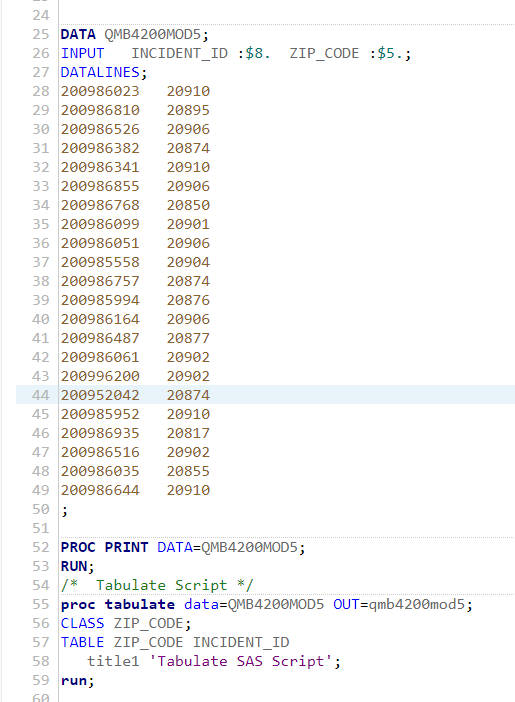




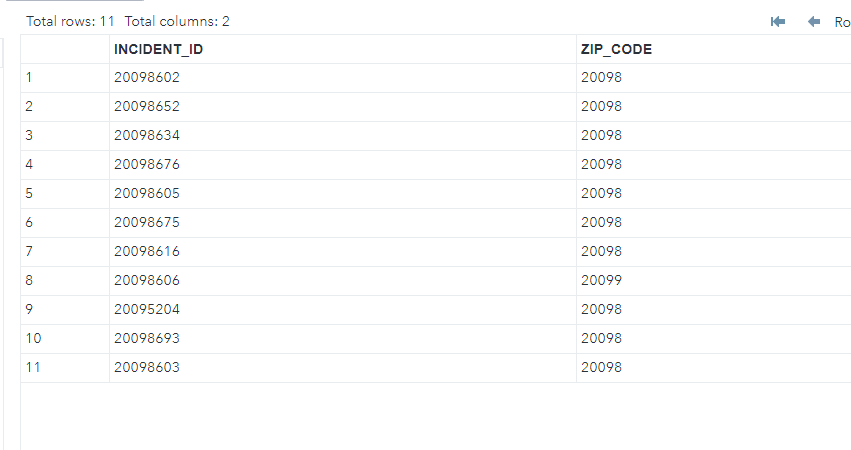
**Course Project - SAS Custom Reporting**

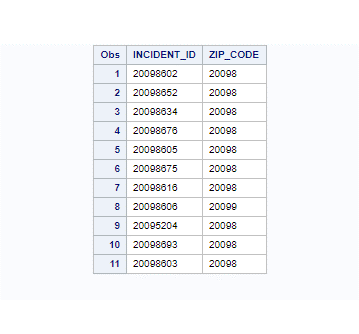
Custom SAS scripts using PROC TABULATE and PROC REPORT for QMB4200 Module 5.

**PROC TABULATE Code:**

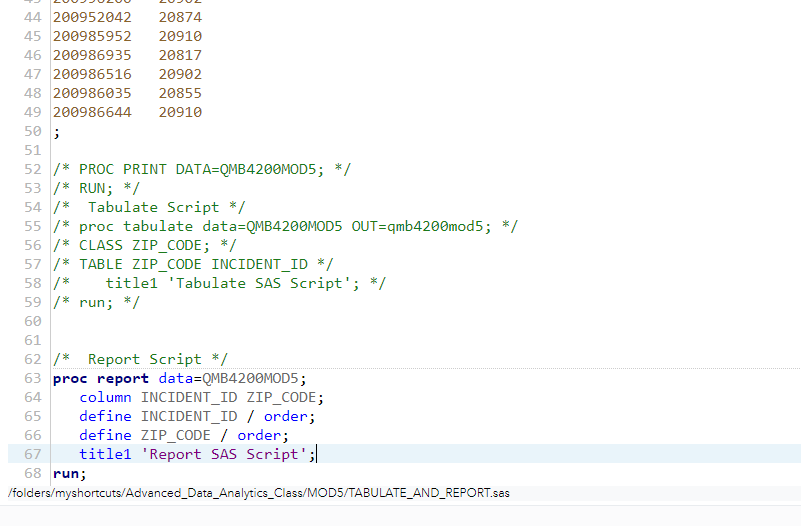


**PROC TABULATE Output:**



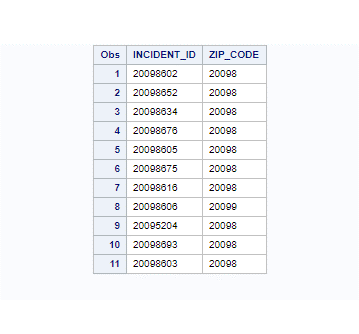


**PROC REPORT Code:**

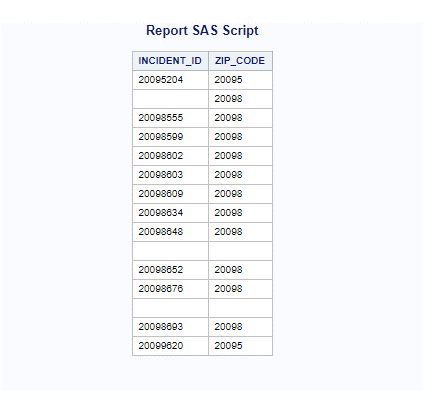


**PROC REPORT Output:**

Here is the report with my original data used in PROC TABULATE



Here is the data report after I duplicated some values.



**Course Project - Alter the existing Azure database tables using SSMS**

In this final section I have designed several normalized databases in Azure SQL, **INSTRUCTORS** table and **COURSES** table. Using the **InstructorID** entry as the primary key and foreign key within the **COURSES** table. I have designed the tables with normal form where the primary key of 1 table composes exactly 1 column.

