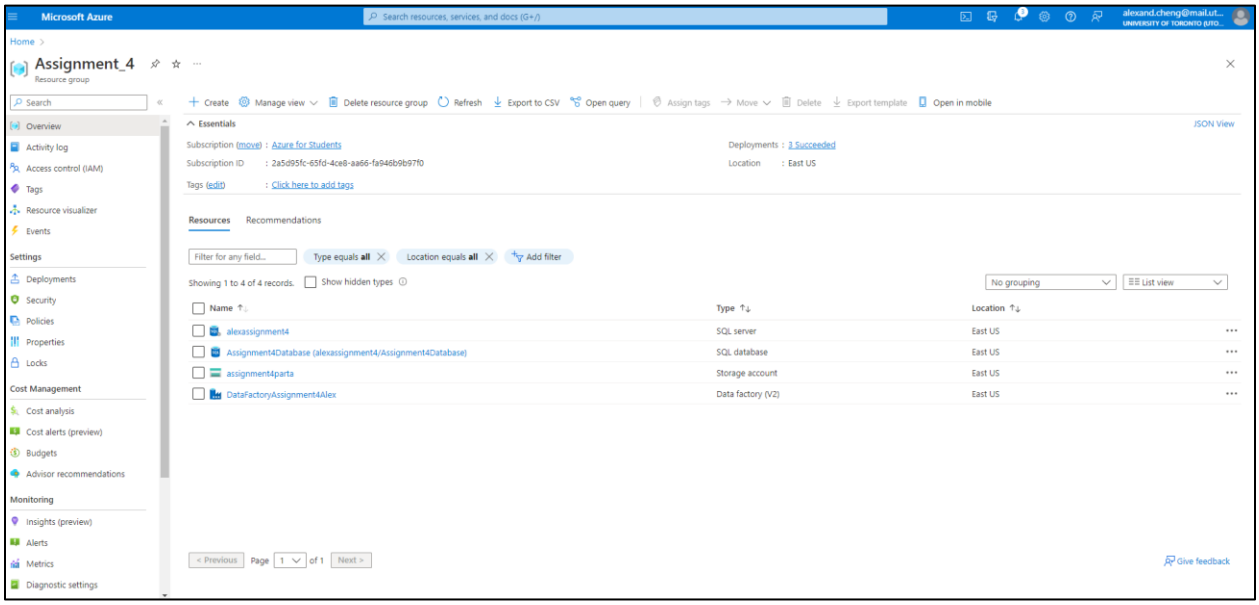
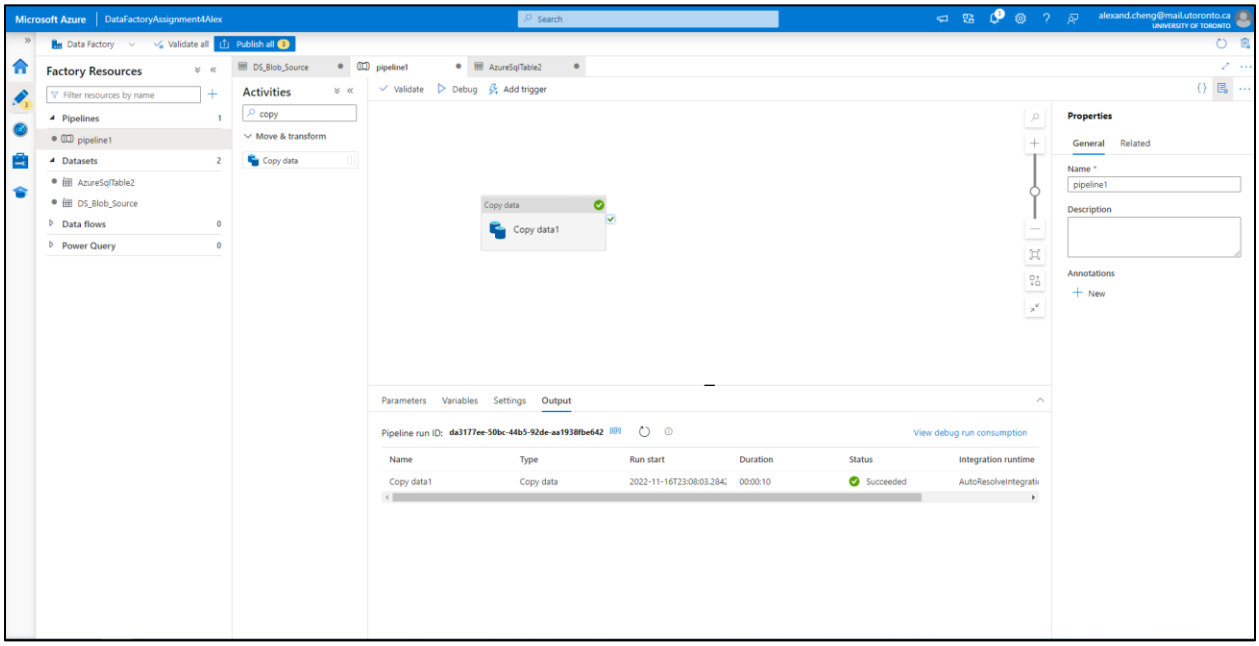
Alexander Cheng – 1001634298
Sunday, November 20th, 2022

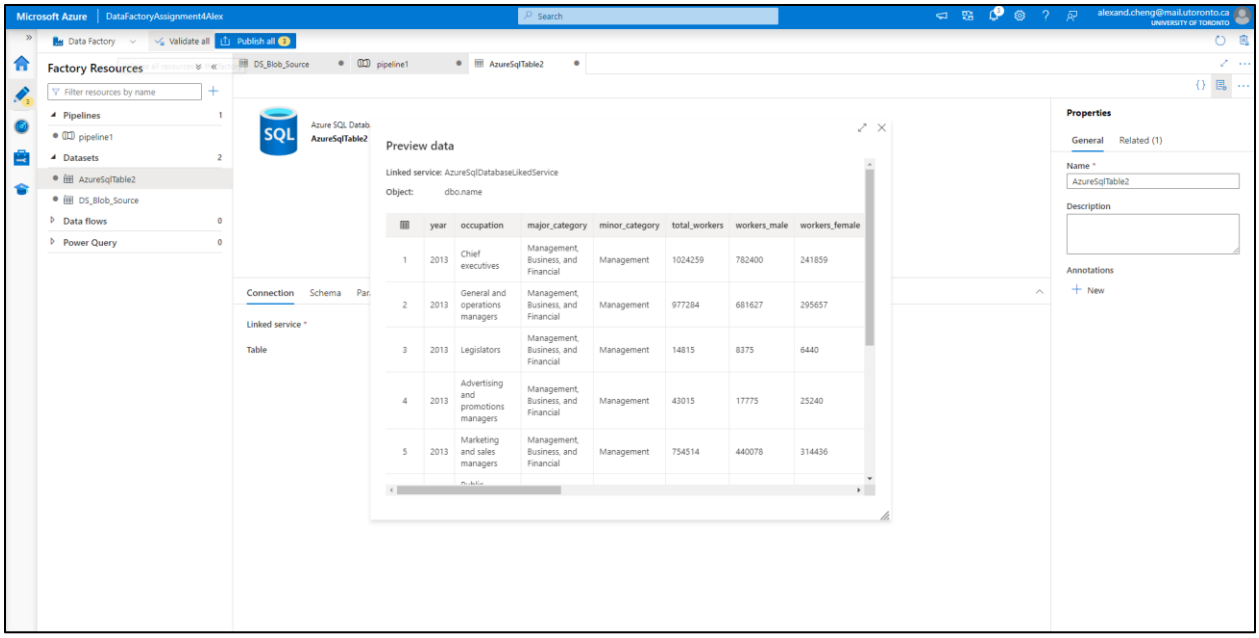
Part A:

1. Create a resource group in your Azure portal and deploy three resources. Azure Data Factory, Azure SQL DB and Blob storage account.



2. Now create a pipeline in Azure Data Factory and copy gender_jobs_data.csv file from the Blob storage account to Azure SQL DB. (First copy this file from your local machine to Blob Storage).



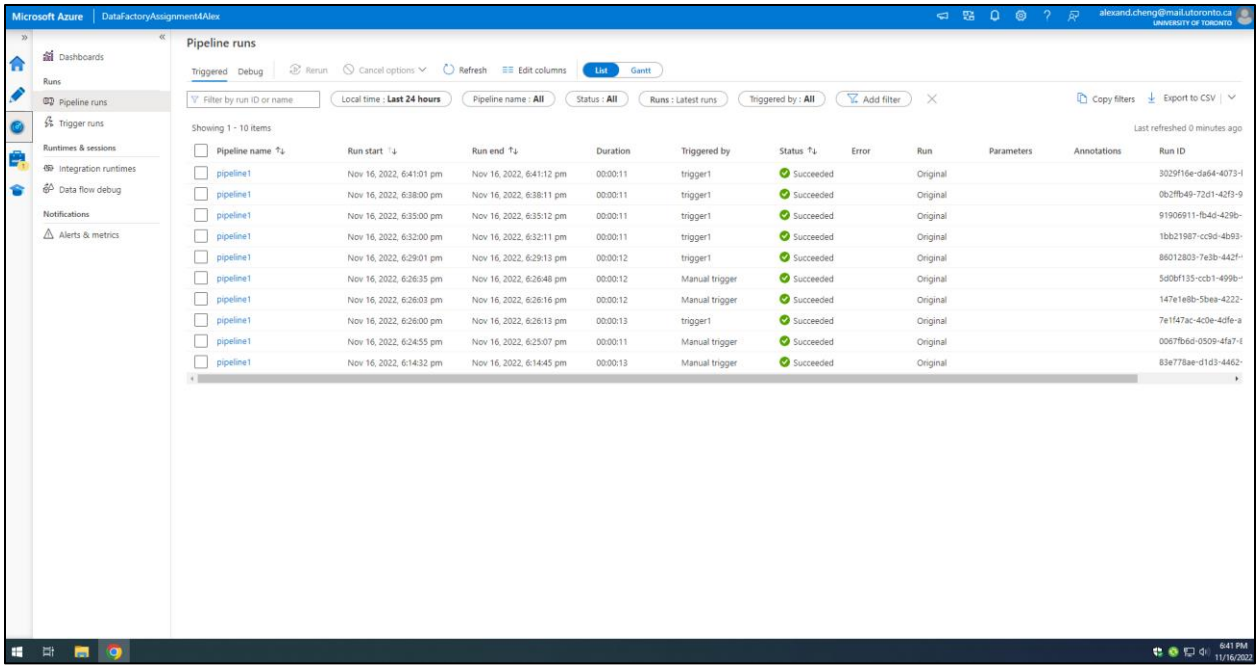


	year	occupation	major_category	minor_category	total_workers	workers_male	workers_female
1	2013	Chief executives	Management, Business, and Financial	Management	1024259	782400	241859
2	2013	General and operations managers	Management, Business, and Financial	Management	977284	681627	295657
3	2013	Legislators	Management, Business, and Financial	Management	14815	8375	6440
4	2013	Advertising and promotions managers	Management, Business, and Financial	Management	43015	17775	25240
5	2013	Marketing and sales managers	Management, Business, and Financial	Management	754514	440078	314436

3. Explain the different types of triggers available in ADF. Now create a schedule trigger and run your pipeline every 3 minutes. Show 5 successful runs.

Trigger Options: Used to schedule a Data pipeline run without any interventions.

- Schedule: Can run a Data pipeline according to a predetermined schedule with different scheduling intervals. Can choose the start and end dates, and on specific future calendar dates.
- Tumbling Window: Executes Data pipelines at a specified time slice or predetermined periodic time interval. The tumbling window sends the start and end times for each time window in the Database, returning all data between those periods.
- Storage Event-based: Triggers occurs in response to blob-related events such as generating or deleting a blob event in an Azure blob storage. These are also compatible with Azure Data Lakes.



4. A client needs to replicate objects from ADLS Gen 2 in Canada Central to ADLS Gen 2 in West Europe. Let's say they want to do this in a bi-directional way. How can you set this up? [Hint: This probably can be done using Azure Data Factory and Event Triggers. For eg; every time there is a new Blob on one side, it needs to be replicated on the other one]

The process will be described on how to implement this but will not be shown.

Cross regional replication provides data recovery in cases of failure, allows staggered times for updates which will minimize downtimes, and reduce the chances of regional disaster network outage. Azure storage account are used to deploy storage resources such as blob containers, file shares, tables, or queues.

We want Geo-redundancy Storage (GRS) or Geo-Zone-redundant storage (GZRS), replication between different regions. Can change the replication setting using the portal "Storage -> Data Management -> Redundancy -> Update settings to Geo-redundant storage and choose "West Europe"".

Need to perform a manual migration, which allows us to move a storage account to another region, although this may have downtime in which a conversion option is a in-place migration with no downtime. However, with the GRS or GZRS options, the data in the secondary region isn't available for read/write access.

Using Azure portal will need to export the storage account template (resources –> automation - > export template), modify the template with the new storage account name and location in the JSON file, move/deploy the storage account to create a new storage in the target location. Configure the new storage account, copy the data, there are many tools to copy data such as AzCopy. To copy the data, the

Alexander Cheng – 1001634298

ADLS Gen 2 Canada Central will be used as a source type and the ADLS Gen 2 West Europe will be used as a sink type.

Need to make sure that the storage in West Europe supports the desired replication settings (Also Geo-redundancy). Azure storage redundancy becomes more expensive when moving to geo-redundancy since it's a more sophisticated redundancy level.

Bi-directional sync is ideal for complex scenarios that involves many pipelines and dependencies. Need to create replication rules that determine the directory in the file system that will be replicated and the Zones that will be used in that replication. Without replication rules defined, each Zone's file system operates independently of the other. A tool such as WANdisco Fusion allows users control over how data is replicated between file systems and object stores.

Need to set up event triggers such that if there's a new blob in one Zone, it will replicate objects from ADLS Gen 2 Canada Central to ADLS Gen 2 in West Europe using the steps shown above. Another event trigger such that if there's a new blob in one Zone, will be used to replicate object from ADLS Gen 2 West Europe to ADLS Gen 2 in Canada Central using the steps shown above.

Part B:

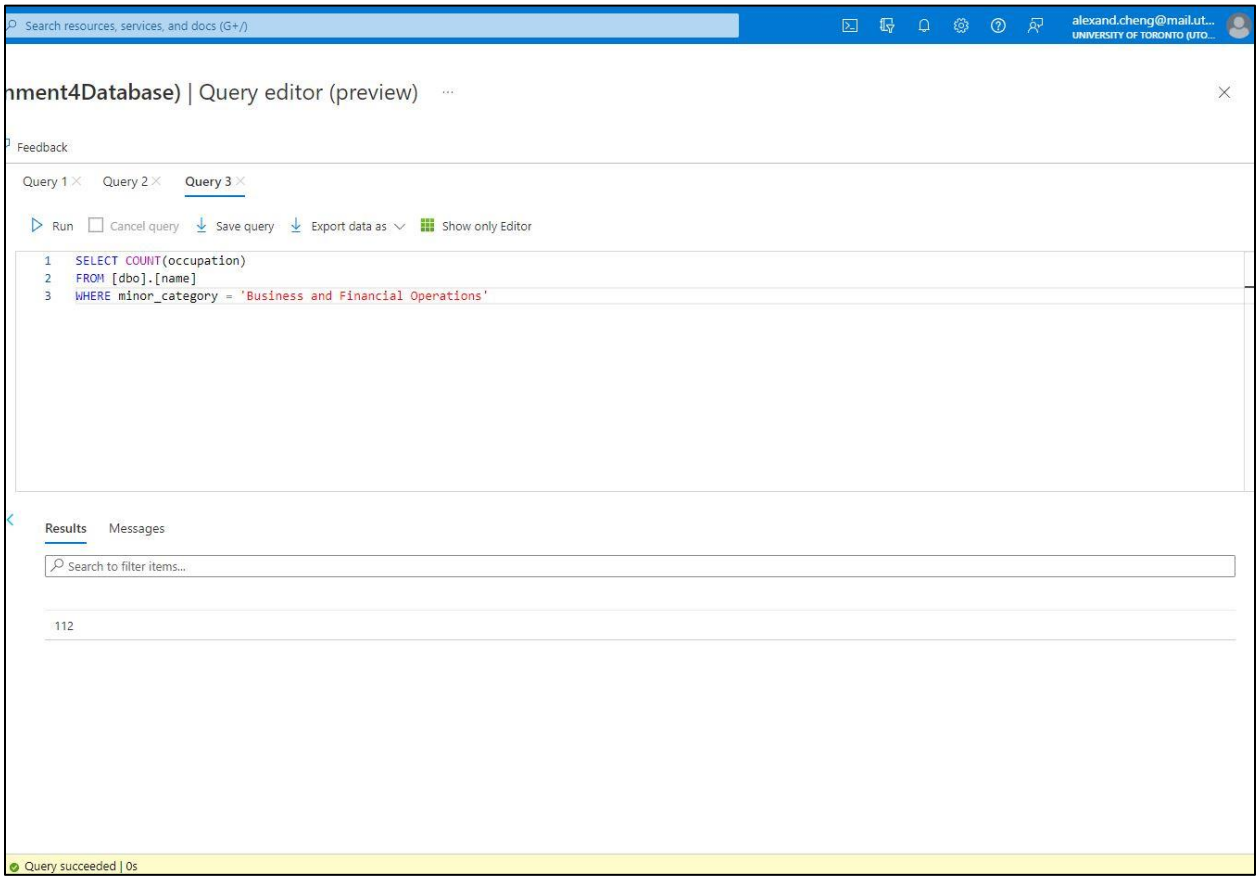
1. In the gender_jobs_data table - Filter all the OCCUPATIONS in MAJOR_CATEGORY of Computer, Engineering, and Science for the YEAR 2013
   - See Part B-1.csv for output

Alexander Cheng – 1001634298

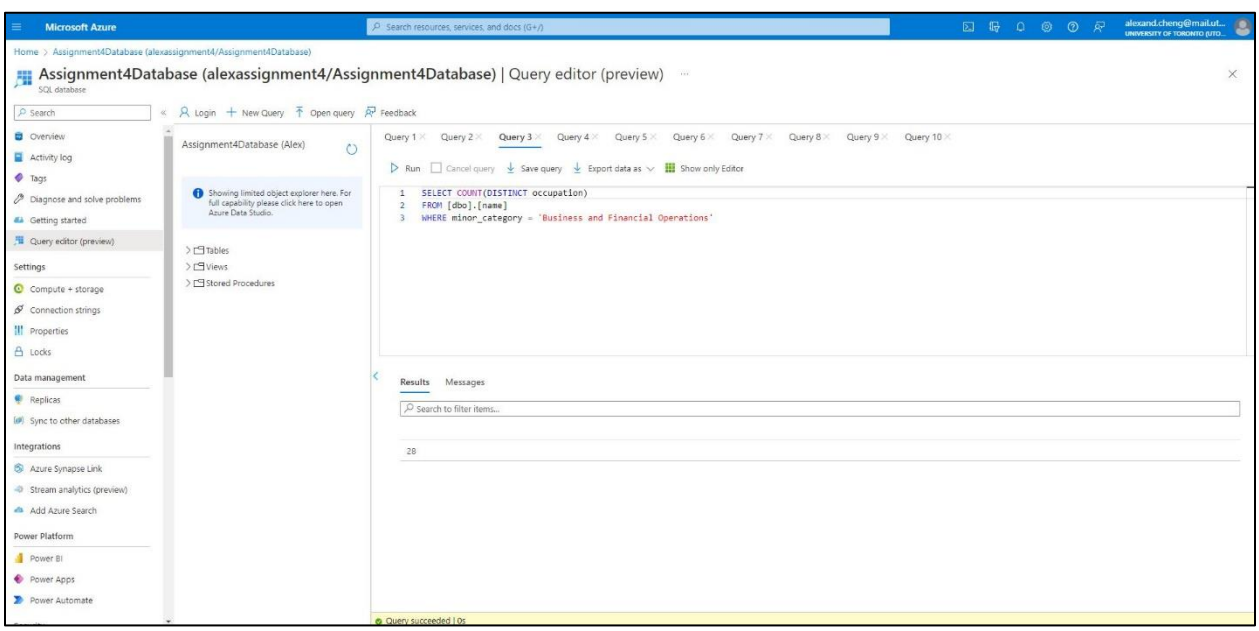
occupation
Web developers
Computer support specialists
Database administrators
Network and computer systems administrators
Computer network architects
Computer , all other
Actuaries
Mathematicians
Operations research analysts
Statisticians
Miscellaneous mathematical science
Architects, except naval
Surveyors, cartographers, and photogrammetrists
Aerospace engineers
Agricultural engineers
Biomedical engineers
Chemical engineers
Civil engineers
Computer hardware engineers
Electrical and electronics engineers
Environmental engineers
Economists
Computer and information research scientists
Computer systems analysts
Information security analysts
Computer programmers
Software developers, applications and systems software
Industrial engineers, including health and safety
Marine engineers and naval architects
Materials engineers
Mechanical engineers
Mining and geological engineers, including mining safety engineers
Nuclear engineers
Petroleum engineers
Engineers, all other
Drafters
Engineering technicians, except drafters

Computer systems analysts
Information security analysts
Computer programmers
Software developers, applications and systems software
Industrial engineers, including health and safety
Marine engineers and naval architects
Materials engineers
Mechanical engineers
Mining and geological engineers, including mining safety engineers
Nuclear engineers
Petroleum engineers
Engineers, all other
Drafters
Engineering technicians, except drafters
Surveying and mapping technicians
Agricultural and food scientists
Biological scientists
Conservation scientists and foresters
Medical scientists
Life scientists, all other
Astronomers and physicists
Atmospheric and space scientists
Chemists and materials scientists
Environmental scientists and geoscientists
Physical scientists, all other
Survey researchers
Psychologists
Urban and regional planners
Miscellaneous social scientists and related workers, including sociologists
Agricultural and food science technicians
Biological technicians
Chemical technicians
Geological and petroleum technicians
Nuclear technicians
Social science research assistants
Miscellaneous life, physical, and social science technicians

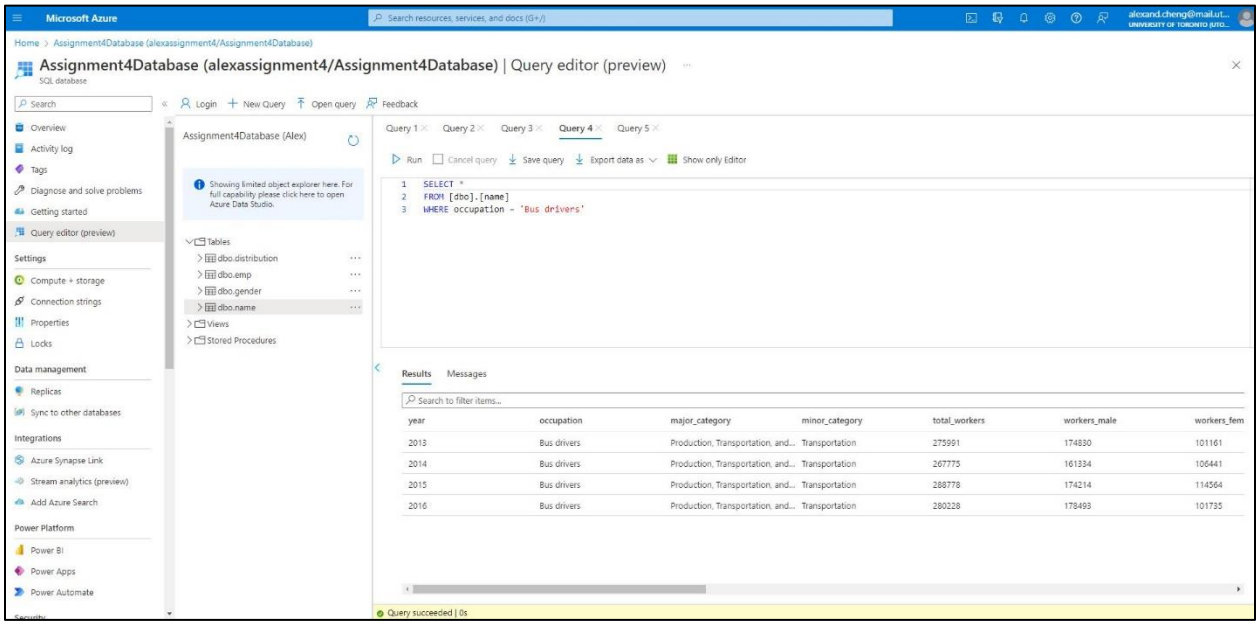
2. In the gender_jobs_data table - How many OCCUPATIONS exist in the MINOR_CATEGORY of Business and Financial Operations overall?



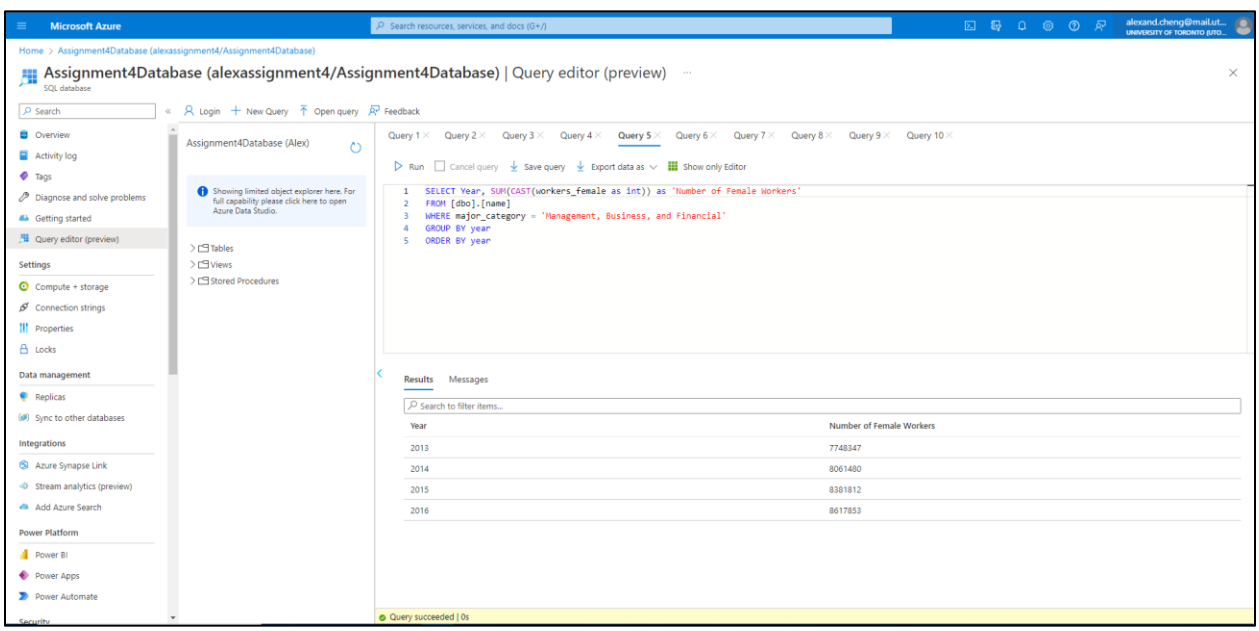
If asking for unique occupations:



3. In the gender_jobs_data table - Get all relevant information for bus drivers across all years
   - See Part B-1.csv for output

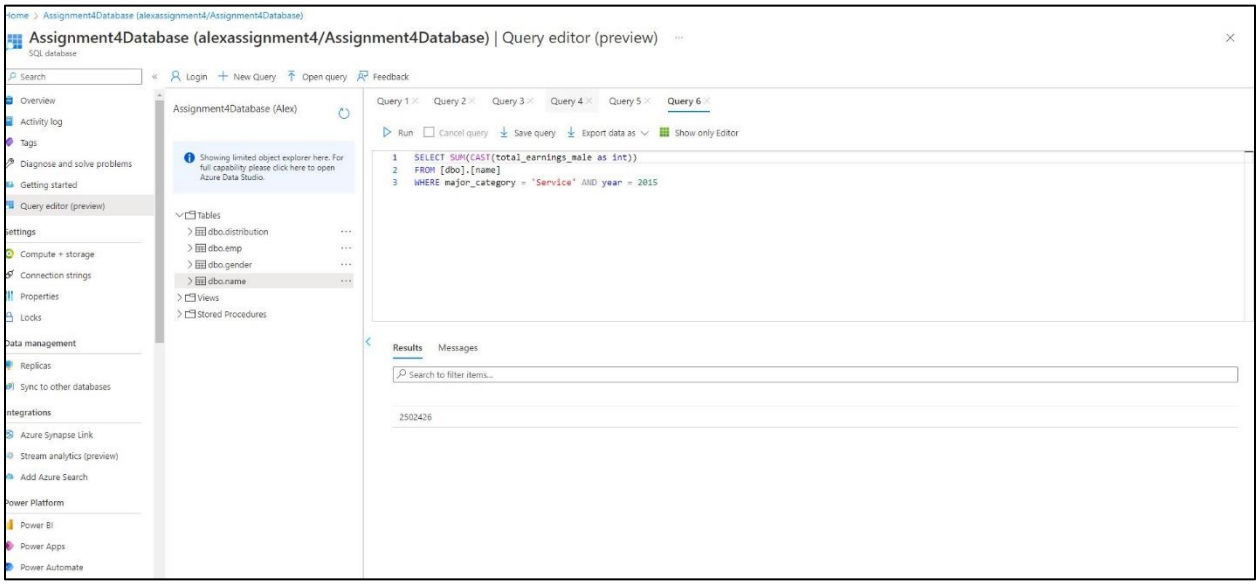


4. In the gender_jobs_data table - Summarize the total number of WORKERS_FEMALE in the MAJOR_CATEGORY of Management, Business, and Financial by each year

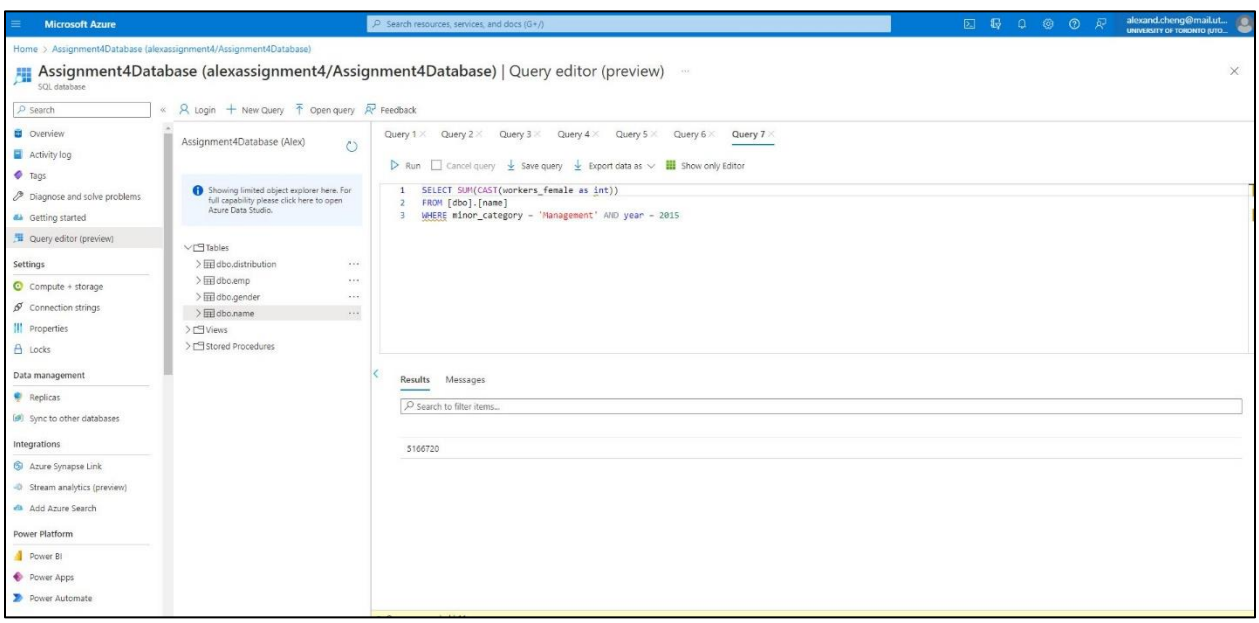


Alexander Cheng – 1001634298

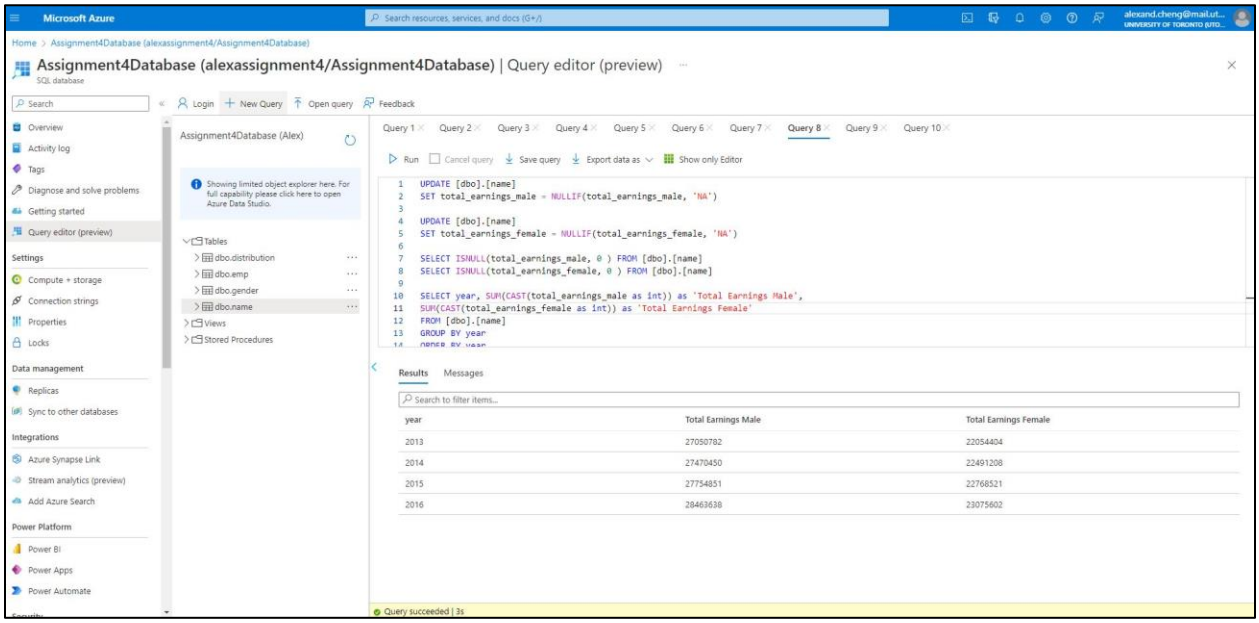
5. In the gender_jobs_data table - What were the total earnings of male (TOTAL_EARNINGS_MALE) employees in the Service MAJOR_CATEGORY for the year 2015?



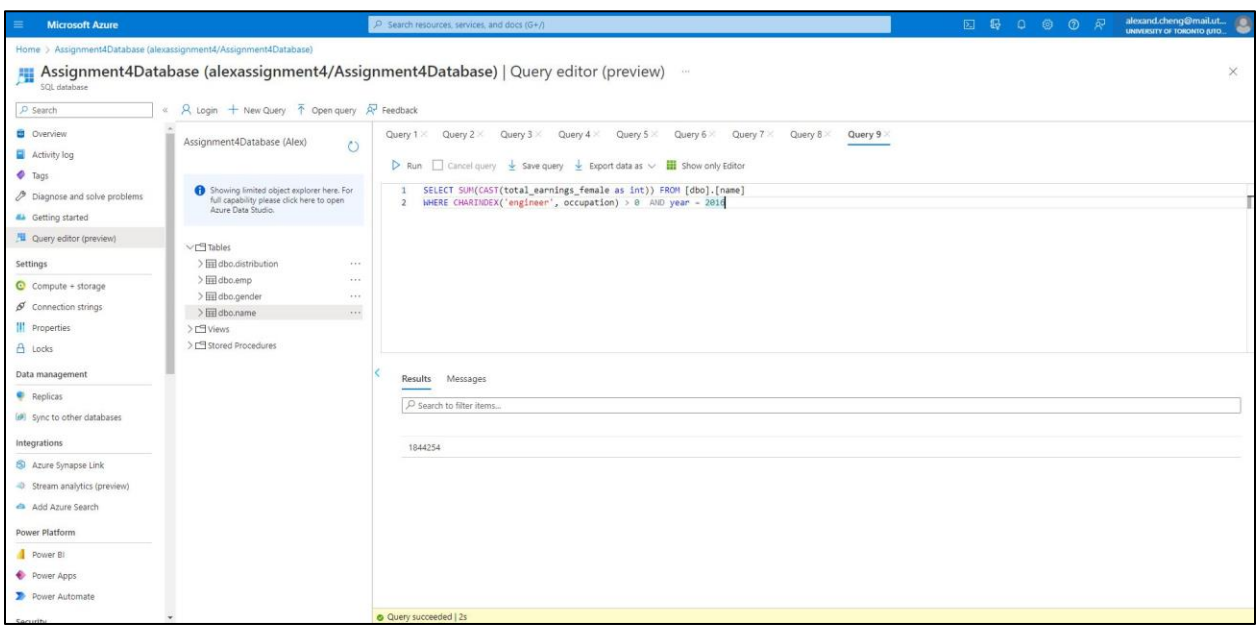
6. In the gender_jobs_data table - How many female workers were in management roles in the year 2015?



7. In the gender_jobs_data table - Compare the TOTAL_EARNINGS_MALE and TOTAL_EARNINGS_FEMALE earnings irrespective of occupation by each year
   - Filter out NA values as 0's



8. In the gender_jobs_data table - How much money (TOTAL_EARNINGS_FEMALE) did female workers make as engineers in 2016?
   - Included every occupation that included the word 'Engineer', and summed up the total earnings



Alexander Cheng – 1001634298

9. What is the total number of full-time and part-time female workers versus male workers year over year?

– Divide by 100 since the numbers given are in percentages