**COURSE: DATA MANAGEMENT AND BIG DATA (ALY 6110)**

**CRN: 71205**

**FINAL PROJECT : ENERGY CONSUMPTION IN NETHERLANDS**

SUBMITTED BY –

**Group: OMEGA**

Neeti Panjwani – [panjwani.n@husky.neu.edu](mailto:panjwani.n@husky.neu.edu)

Shivani Kolungade – [kolungade.s@husky.neu.edu](mailto:kolungade.s@husky.neu.edu)

Sarvesh Prabhu – [prabhu.sa@husky.neu.edu](mailto:prabhu.sa@husky.neu.edu)

**Summary:**

We as a team are working on a dataset that provides information on the energy consumption in Netherlands. The main intention is to understand the energy consumption and analyze the outcomes for the amount of energy produced by the Enexis energy organization. It is the decision that we are ultimately going to take to plan these at various sectors so that we can collect different kind of data and process this together to receive data that can be easily understood to save costs and reduce the consumption for a final goal. Not just that, we will also concentrate on the trends of smart meters and will determine what factors are responsible for the changes.

**The real-world problem:**

We as a team are working on a dataset that provides information on the energy consumption in Netherlands. The country being the significant processing and transportation hub, we would concentrate on how the 3 major network administrators work together to provide energy to nearly the entire country. The main intention is to understand the energy consumption and analyze the outcomes for the amount of energy produced by the Liander energy organization. It is the decision that we are ultimately going to take to plan these at various sectors so that we can collect different kind of data and process this together to receive data that can be easily understood to save costs and reduce the consumption for a final goal. Not just that, we will also concentrate on the trends of smart meters and will determine what factors are responsible for the changes.

**Business Questions/ Insights:**

* How are the smart meters spreading?
* What is the overall energy consumption by the Enexis over the years?
* What are the factors responsible for the changes in energy consumption?

ABOUT THE DATASET: Variables

* city: Name of the city
* zipcode: range of zipcode for the city
* num\_connections: Number of connections in the range of zipcodes
* perc\_of\_active\_connections: Percentage of active connections in the zipcode range
* type\_conn\_perc: percentage of presence of the principal type of connection in the zipcode range
* annual\_consume: Annual consume. Kwh for electricity, m3 for gas
* smartmeter\_perc: percentage of smartmeters in the zipcode ranges
* year: data is spread over 10 years from 2010 to 2019

Data Pre-processing:

* Dropped the columns with no significant data.
* Merged the files based on the year and the common columns(grouped).
* Concentrated on Enexis’s Electricity and Gas Consumption separately.

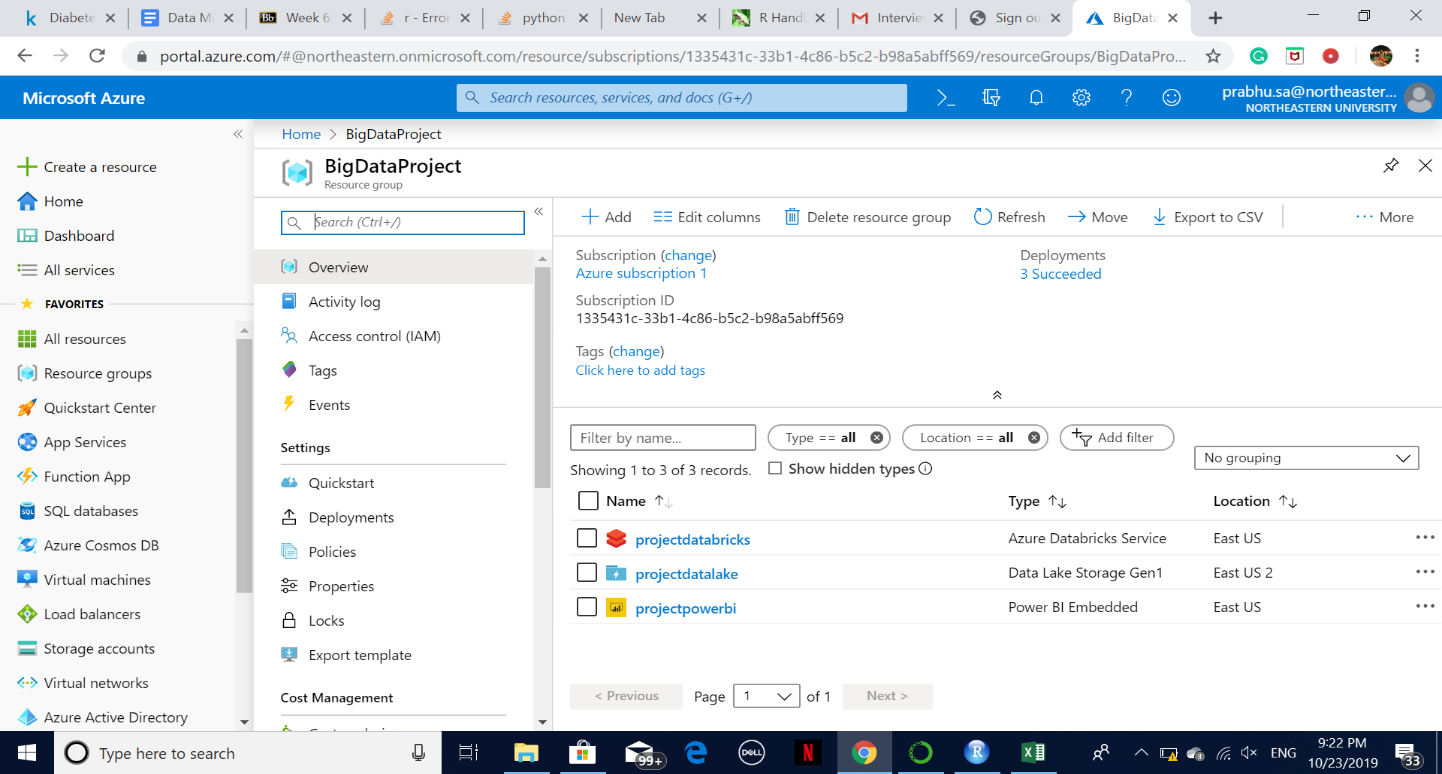
How will this benefit?

We have been facilitated with data from the year 2009 to 2019 for both electricity and gas. There are 16 variables in the dataset such as city, consumption, zip code, year and so on with 2.5 million records in it. Analyzing such big data will be challenging as the big data or the kind of data being collected about the energy usage can help make the credit lines more efficient with the analytical tools that we will concentrate on like Hadoop that are now reaching the areas, where they couldn’t before. From our dataset, the concentration is more on the energy consumption which has the potential to provide a multitude of insights on energy efficiency, usage and more.

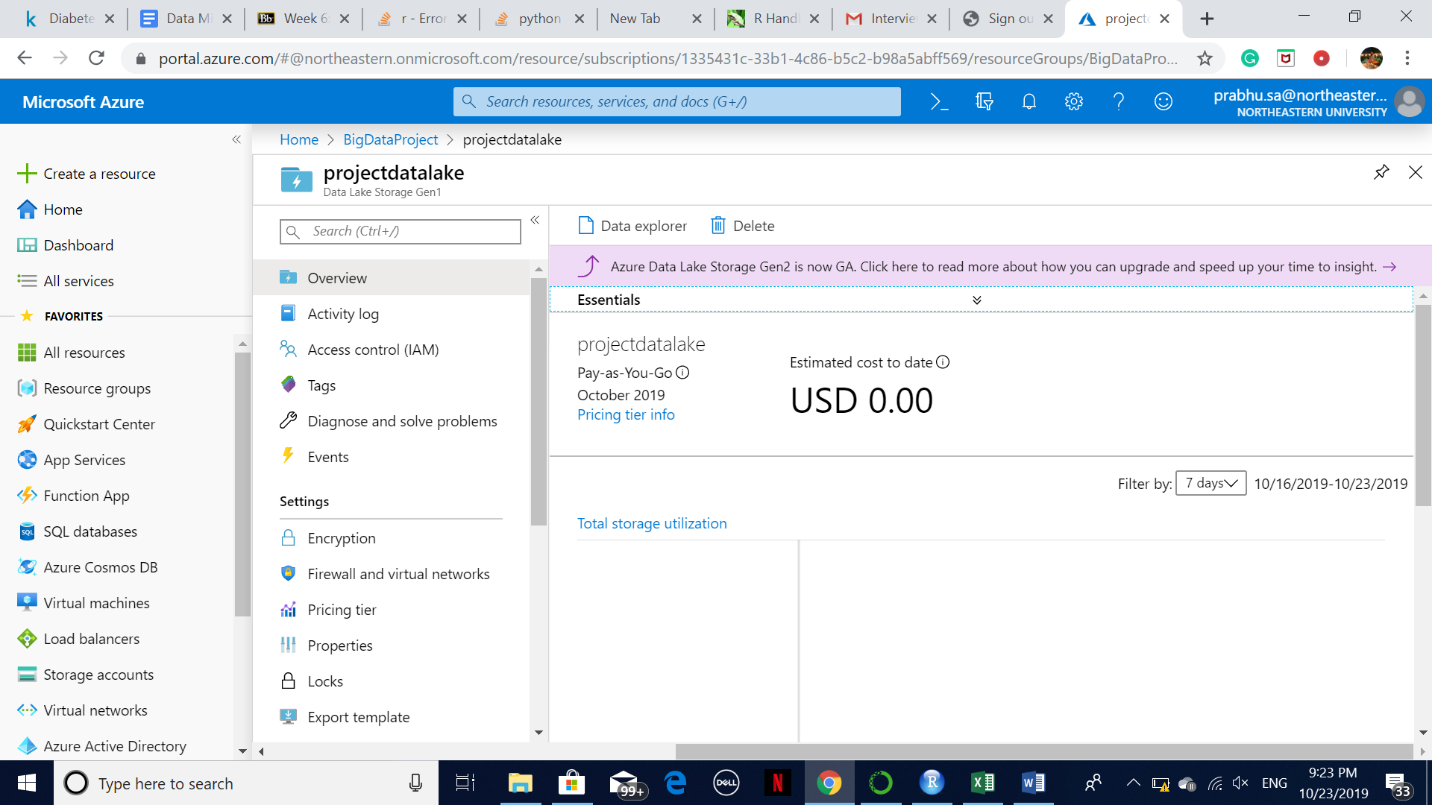
**Content:**

Steps followed:

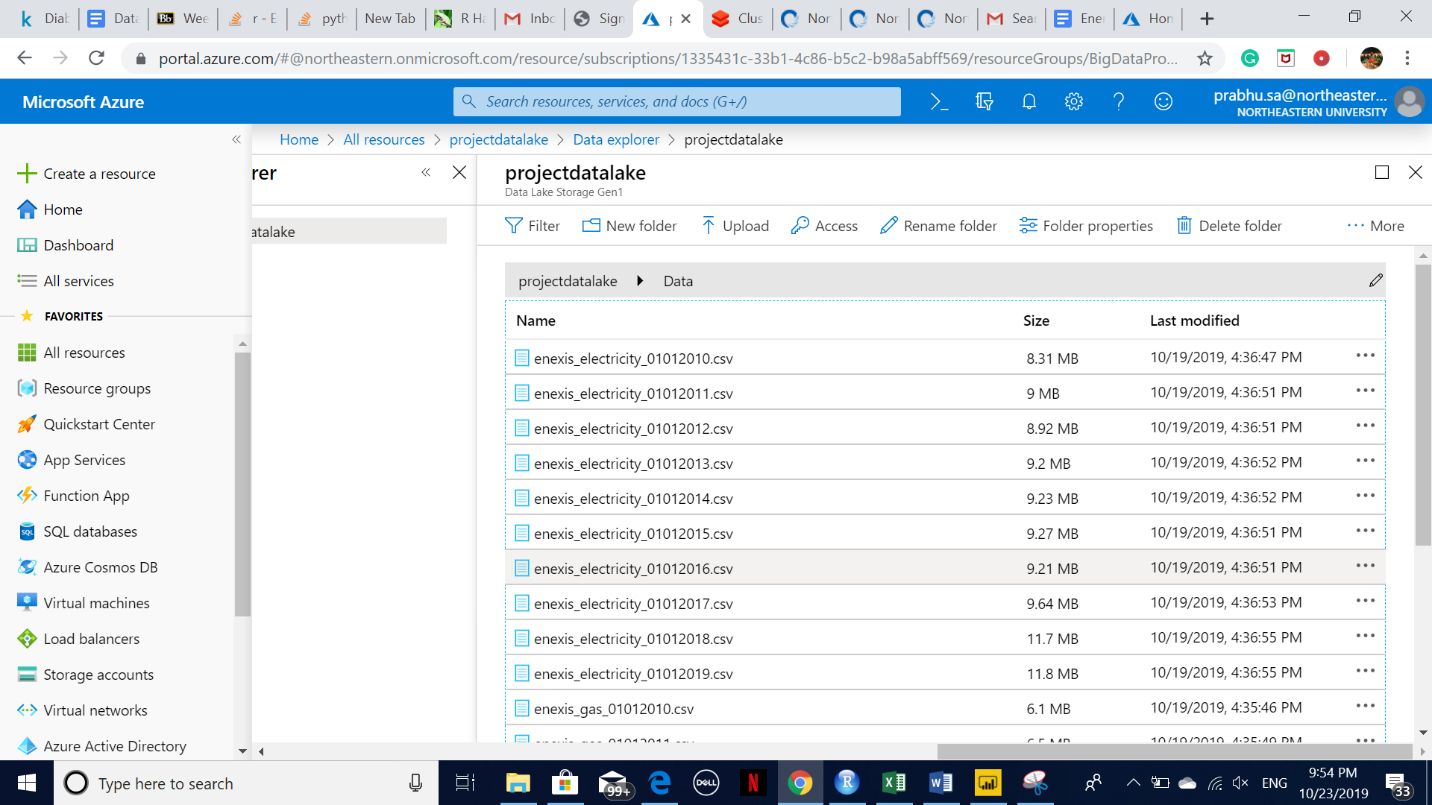
* Navigate to <https://portal.azure.com/#home> to login to Microsoft Azure portal.
* Click on Create Resources as shown above to create resources.

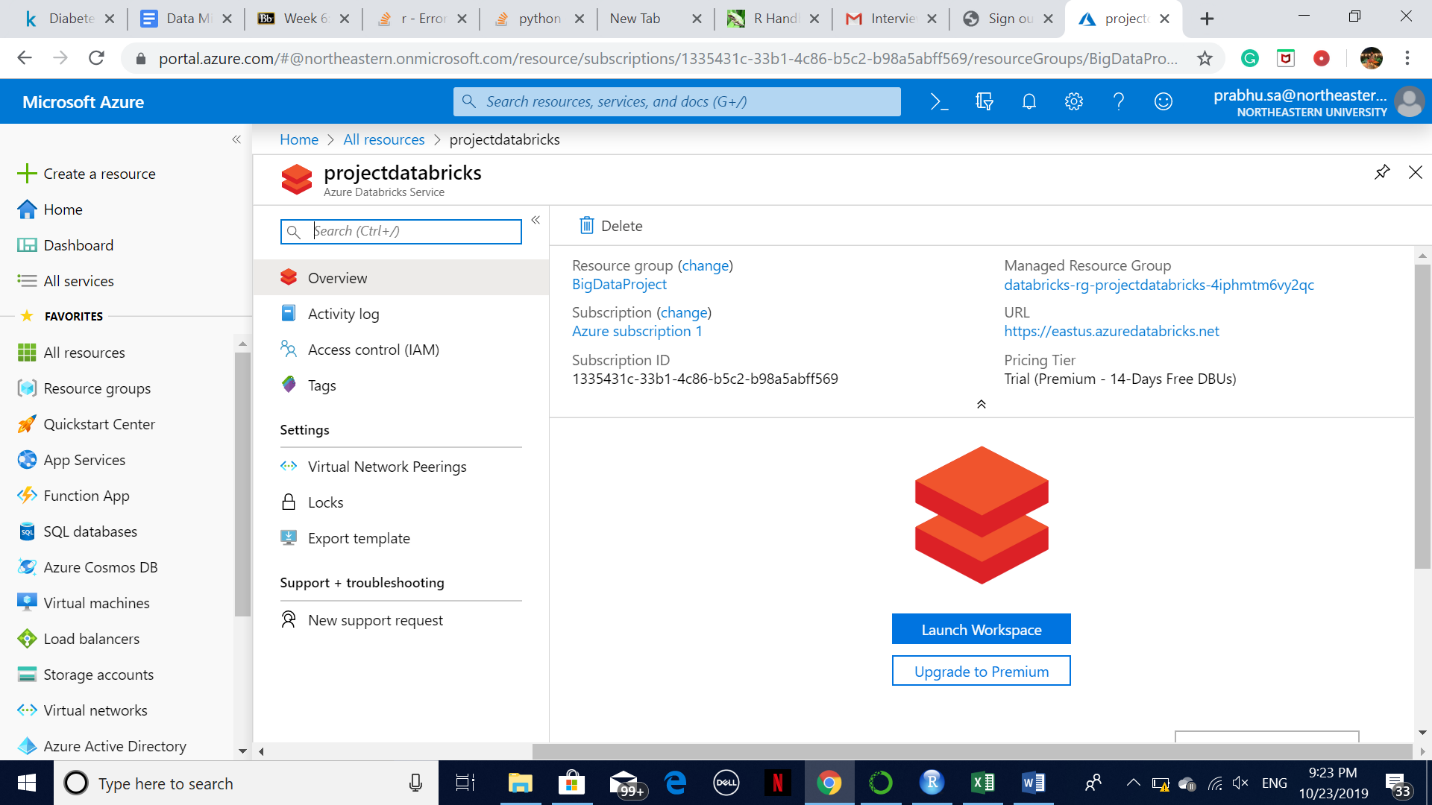


* Under Azure Marketplace tab, Click on Analytics to create New Datalake storage gen1.

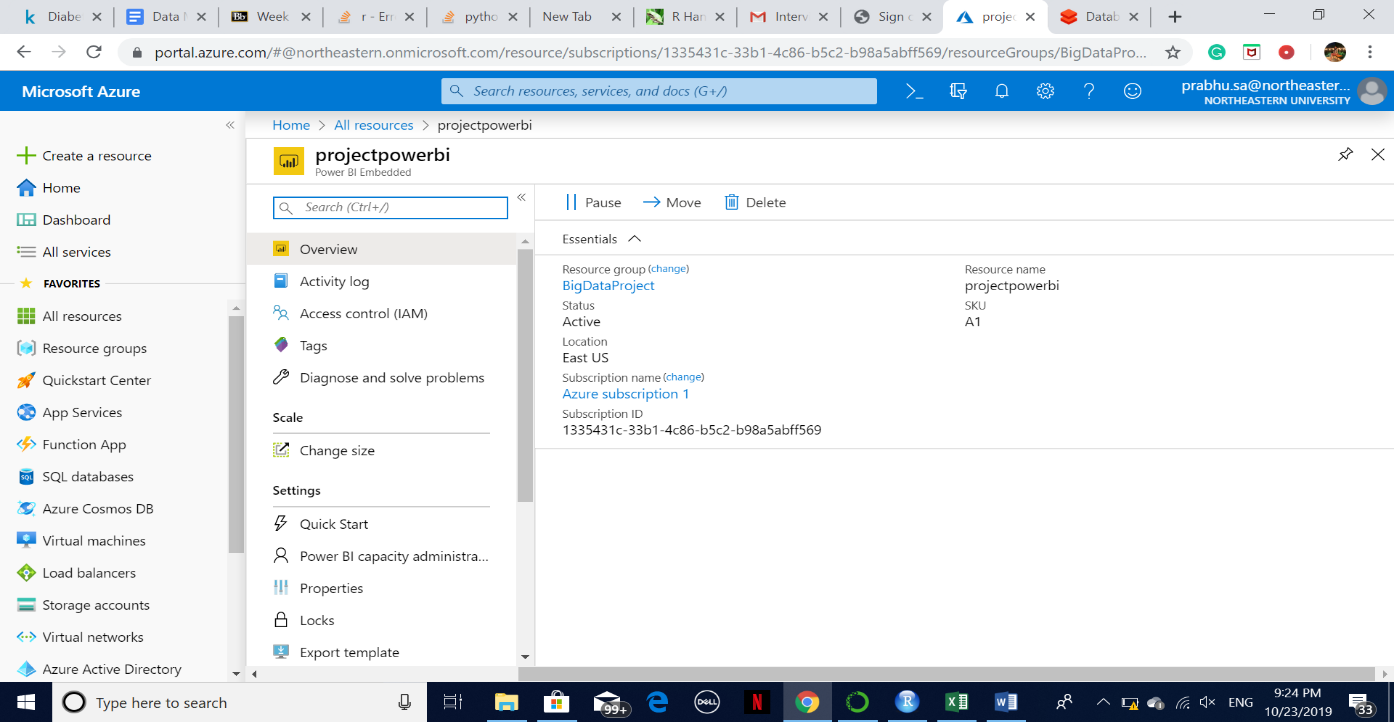


Again, click on Analytics tab under create resources to create Azure Databricks services which we will use as a platform for code operations.

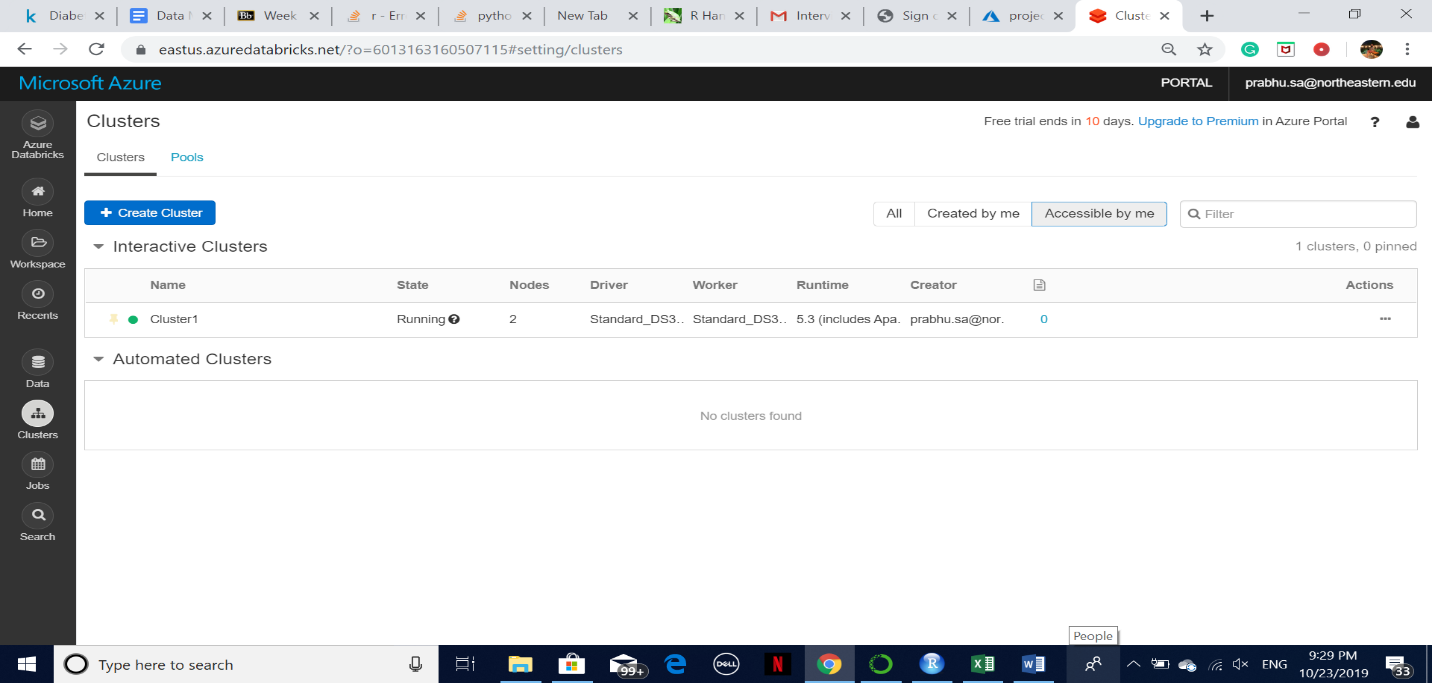


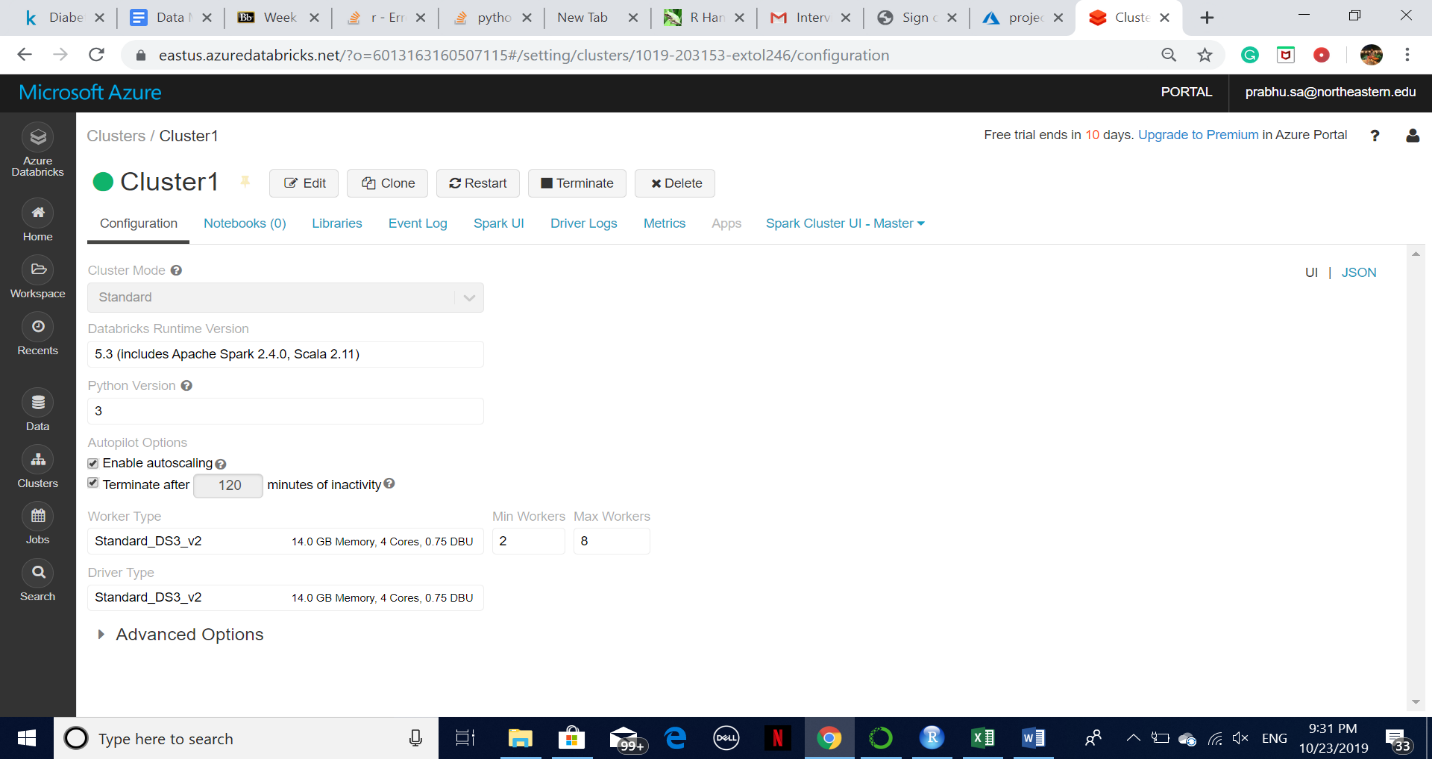


* Under analysis, go to power BI Embedded and fill all the required information like Resource name, Subscription, Resource group etc. and then click on create button at the bottom of the screen.
* Now, go to all resources and click on project data bricks which we created earlier

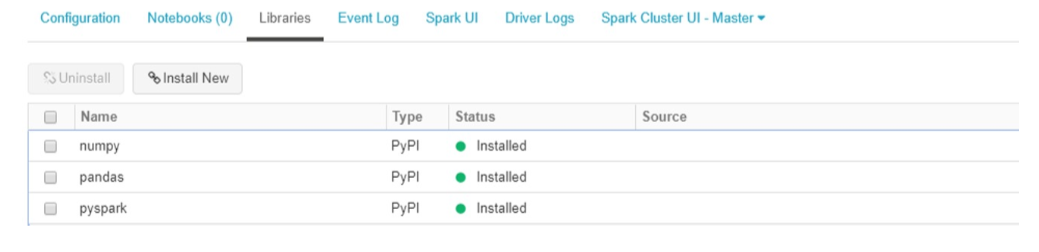


Fill all the required information given in above page and then click on Create Cluster to create new cluster.

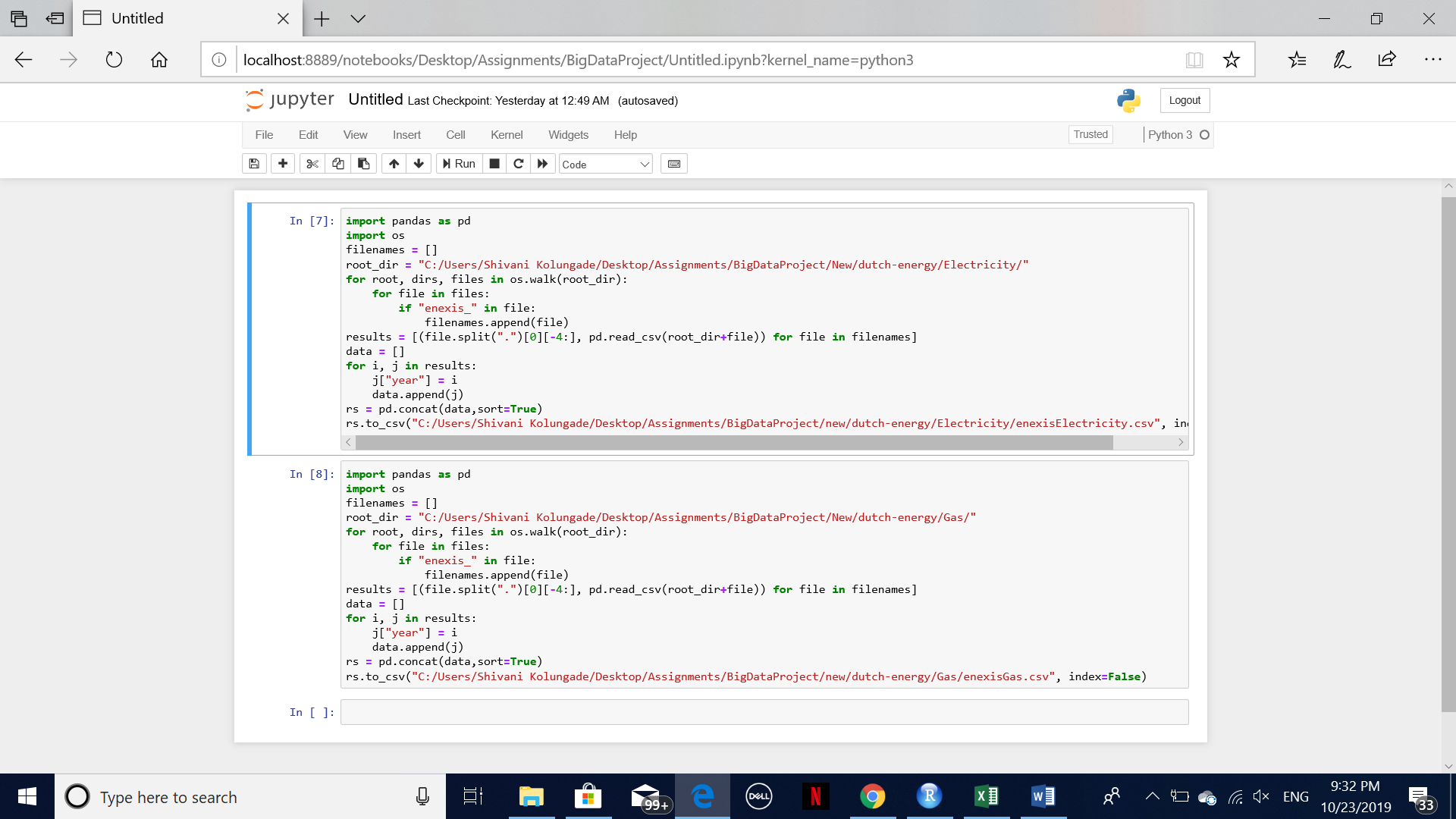




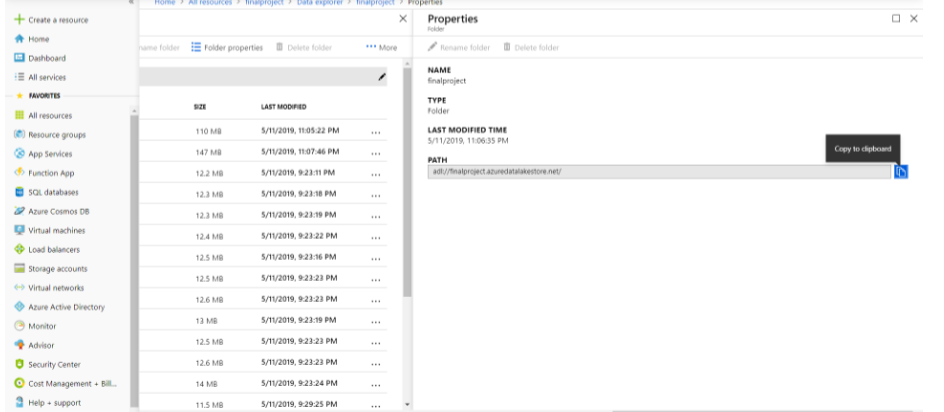
We have installed required libraries.



* Create new notebook by clicking on notebook and give it a name. Then click on Create button.

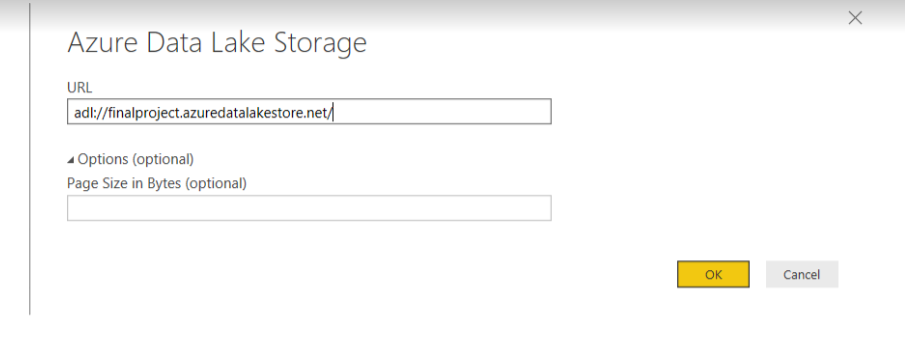


We have successfully done data preprocessing.

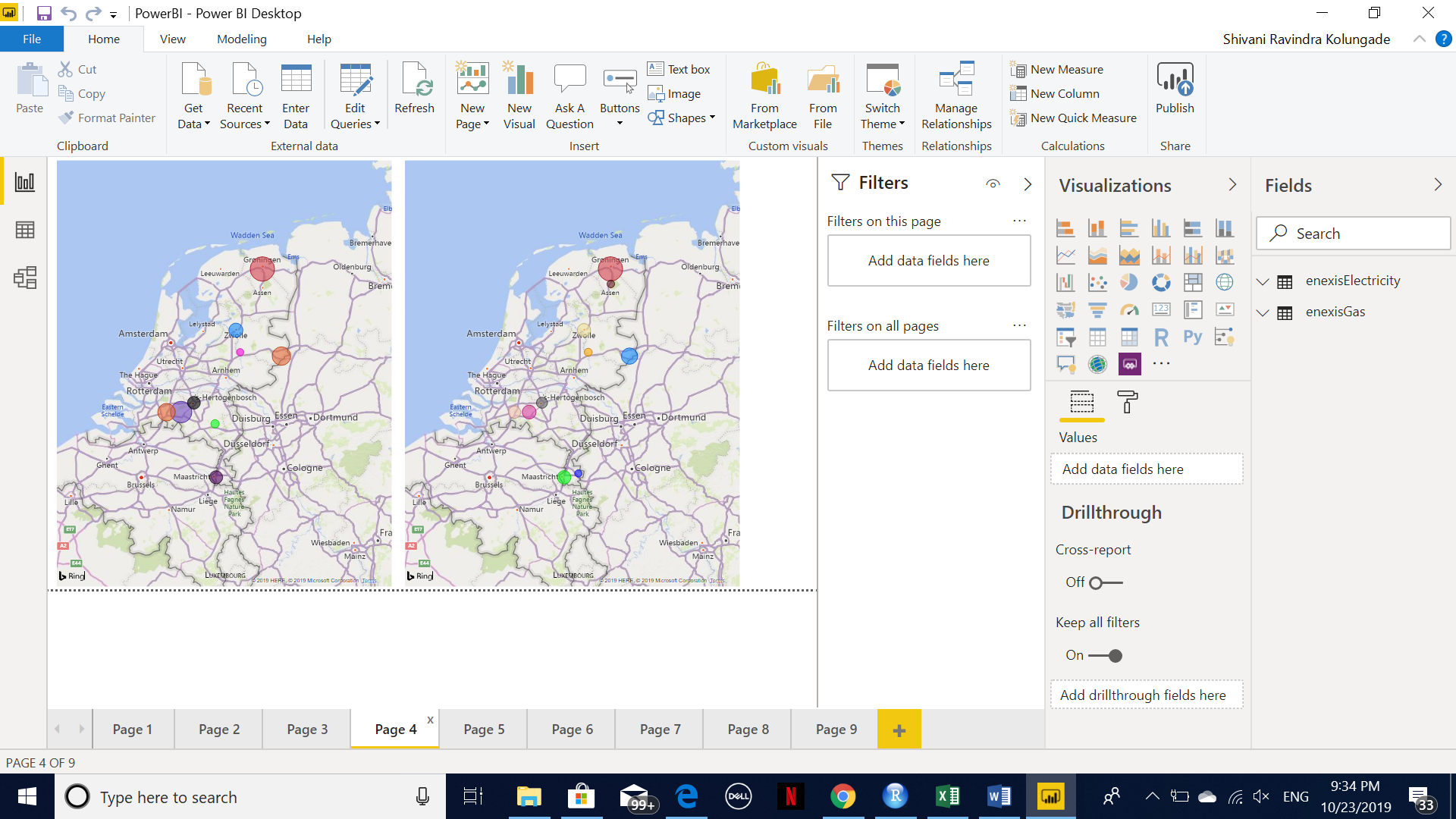


From properties section we copied the path that we will use to build a connection.

In the data lake, we are mounting the data in the storage which is shown below:

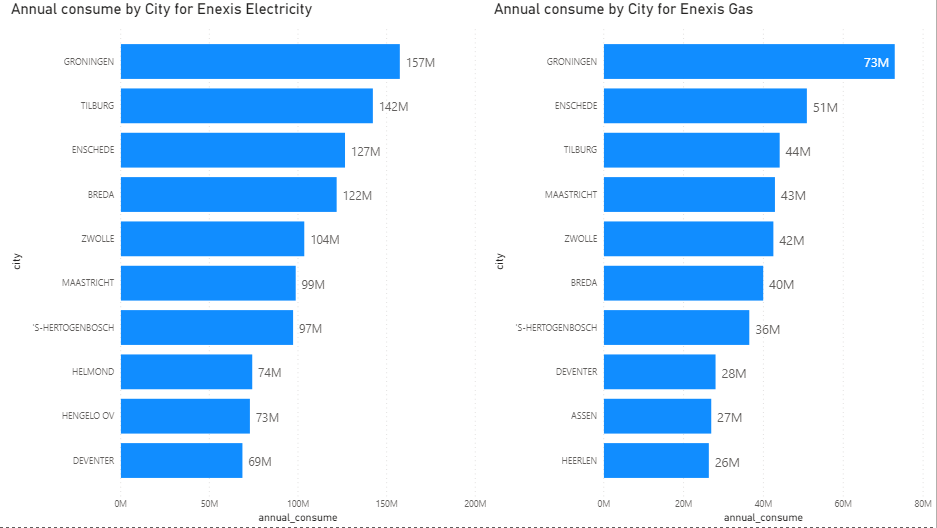


* We copied the path as shown above and clicked on Ok.
* Then we started working on visualization.



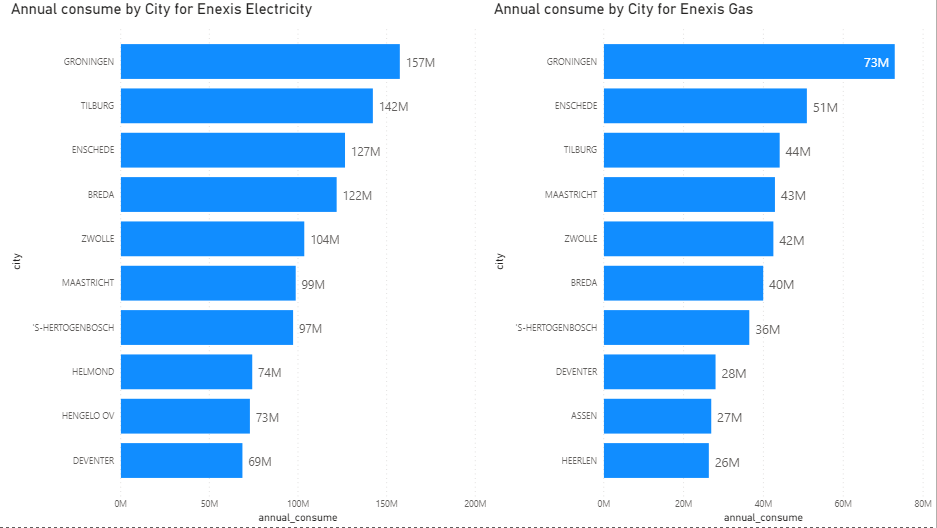
**Visualization: (PowerBI)**

The total Energy consumption of Electricity in Netherlands was

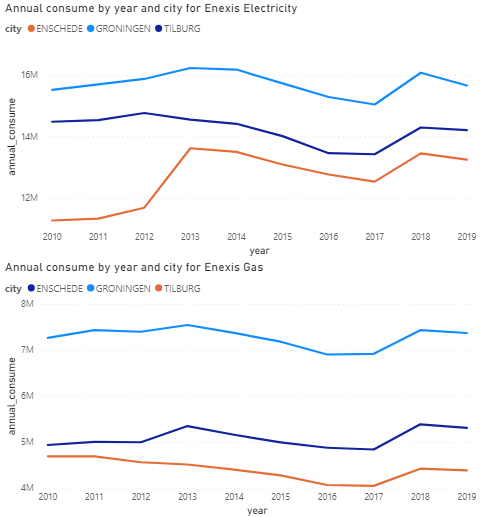


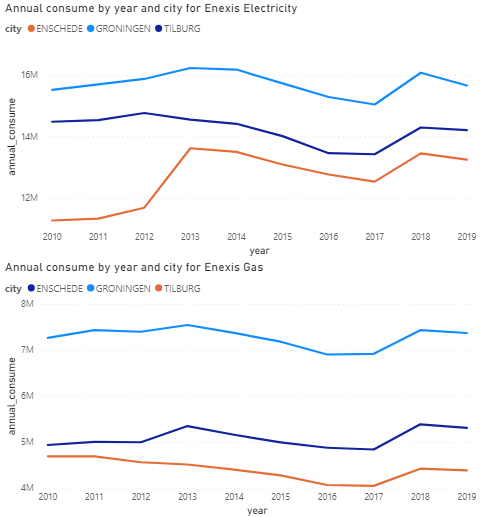
After filtering the top 10 cities we see that the highest consumption was in Groningen with about 157M KWh annual consumption.

While the total Energy consumption of Gas in Netherlands was 73M cubic meter and again for the city of Groningen.

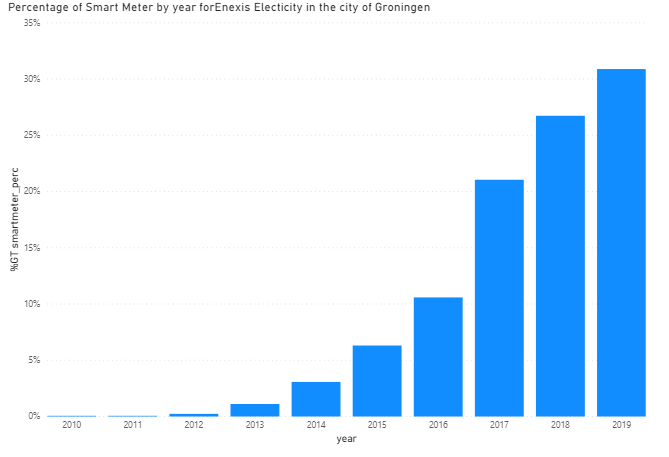


To understand the energy consumption trends in the top 3 cities, we have differentiated this with a line graph shown below.

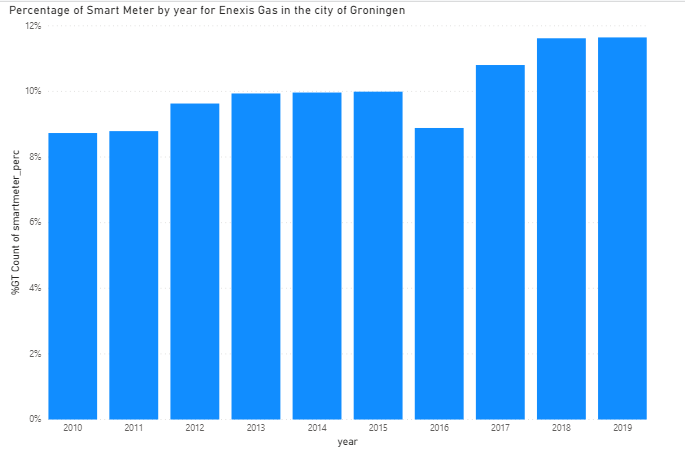




The trend lines above show that there was a linear change in the year 2017. This was due to the increase in promotion values of the usage of LEDs and smart meters. The situation had turned old school and many were fluctuating with the decision as to what was more effective.



The Enexis administrator was well versed with the gradual increase in the energy consumption of Electricity in Groningen, where as there was no much change in the Gas consumption. (shown below)



**Conclusions:**

* The total electricity consumption has been roughly constant over the years except for their major city **Groningen**.
* While the number of connections growing linearly, so the consumption per connection in cities seemingly decreased, because of using more energy efficient products such as LEDs and awareness of saving energy everywhere.
* Its observed that number of smart meters have increased tremendously over the years.
* So, we can conclude that, if Enexis will deploy more smart meters in Groningen followed with other cities in the Netherlands, it will help them increasing their revenue.
* Smart meters will give them slot data that will help them monitoring the consumption patterns of energy for their customers and generate bills accordingly. With the help of these smart meters, every user can have control over energy consumption and save environment.