­

|  |
| --- |
|  |
| Parameter conclusions for external data and Fontys IT data  Fontys IT & NovoData |
| |  |  |  | | --- | --- | --- | | Burak Basal, Martin Frenken, Nam Le, Hugo Meur, Jara van de Mortel, Tim Oomen, Djinn Opten, Jorrit Overeem, Hai Tran | 6/23/21 | Data Driven Business Lab | |

**Parameter conclusions for external data and Fontys IT data**

Fontys IT & NovoData



Burak Basal, Martin Frenken, Nam Le, Hugo Meur, Jara van de Mortel, Tim Oomen, Djinn Opten, Jorrit Overeem, Hai Tran

**Project client:**

R. van der Vorst

June 23, 2021

Fontys University of applied sciences

**Project coaches:**

B. van Gennip & J. Cornelissen

Content

[1 Introduction 5](#_Toc75169364)

[2 Choices for data sources 5](#_Toc75169365)

[2.1 Personcount from WiFi data 5](#_Toc75169366)

[2.2 Parameters from Priva sensordata 6](#_Toc75169367)

[2.3 External sources 7](#_Toc75169368)

[2.3.1 Chosen parameters with API’s 7](#_Toc75169369)

[2.3.2 API’s that are not (yet) chosen 8](#_Toc75169370)

[3 Optimal values for parameters 10](#_Toc75169371)

[3.1 Light intensity 10](#_Toc75169372)

[3.2 CO2 values and overall air quality 12](#_Toc75169373)

[3.3 Temperature 13](#_Toc75169374)

[3.4 Plants 14](#_Toc75169375)

[4 Data to store 14](#_Toc75169376)

[4.1 Available and required data 14](#_Toc75169377)

[4.2 Update and Intervalplanning 15](#_Toc75169378)

[4.2.1 Data from the Priva sensors 15](#_Toc75169379)

[4.2.2 Data from other sources 15](#_Toc75169380)

[4.3 Structure 16](#_Toc75169381)

[4.3.1 Storage of data 16](#_Toc75169382)

[4.3.2 Analysing the data 17](#_Toc75169383)

[4.4 Estimated required storage 18](#_Toc75169384)

[4.5 Cost calculation for database 19](#_Toc75169385)

[4.5.1 Calculations based on current situation 20](#_Toc75169386)

[4.5.2 Prediction based on the expansion of the project 20](#_Toc75169387)

[4.6 Dashboard data choices 21](#_Toc75169388)

[4.7 Future (Im)possibilities with Azure 23](#_Toc75169389)

[5 Reliability and Influences of people on sensordata 27](#_Toc75169390)

[5.1 Light sensors 27](#_Toc75169391)

[5.1.1 Places of sensors 27](#_Toc75169392)

[5.1.2 Automated changes 27](#_Toc75169393)

[5.1.3 Visitors and artificial light 27](#_Toc75169394)

[5.2 CO2 sensors 28](#_Toc75169395)

[5.2.1 Automated changes 28](#_Toc75169396)

[5.2.2 Visitors and CO2 concentration 28](#_Toc75169397)

[5.3 Temperature sensors 29](#_Toc75169398)

[5.3.1 Automated changes 29](#_Toc75169399)

[5.3.2 Visitors and temperature 30](#_Toc75169400)

[6 Bibliography 31](#_Toc75169401)

# Introduction

This report follows up the project plan of NovoData for the client Fontys IT. Fontys IT has placed multiple sensors in their new building R10 in Eindhoven. They want to use these sensors for their mission – offering inspiring, challenging, and excellent higher education, and executing applied research that is meaningful for society. (Fontys, 2021) The task of NovoData is to create a database for this sensor data, and to add new external sources as well. As a result, Fontys IT will be able to see correlations/draw conclusions based on certain patterns. NovoData will start this project by collecting data for the 4th floor of R10. The coming chapters describe the parameter choices, findings, reliability, and data storing structure of NovoData for the Fontys IT project, with conclusions that are drawn between February and June 2021:

* Chapter 2 shows why certain parameters are chosen and which parameters are not (yet) included. This is useful for next project groups.
* Chapter 3 gives the optimal values for these parameters. Fontys IT can use this as a goal to achieve in R10.
* Chapter 4 describes how much data is required and available to store, with the associated costs. Both Fontys IT and next groups can use this information as a reference.
* Chapter 5 explains the influence of people on the values of the sensor data. Next groups can prove these influences with graphs for the long term.

# Choices for data sources

Data can be placed in three different categories, namely: Wi-Fi data from 12CU, Priva sensor data from Fontys, and externals sources. It is important what kind of data should be considered while making a new database. Section 2.1, 2.2. and 2.3 cover these categories respectively.

## Personcount from WiFi data

Fontys IT uses a software service from 12CU that converts Wi-Fi data to information about presence of visitors. Users can filter on buildings, floors, time frames and visitors (e.g., students from specific studies). With this information, NovoData can see the utilization rate of the 4th floor of R10 on a specific day. The software 12CU provides insights by giving snapshots of the occupancy every 10 minutes. If the number of people on the 4th floor is known, it is possible to compare this with sensor data to see if there are any correlations or influences. However, the data is not available per room but only per floor.

Figure 1 shows how the graph looks like when filtering on the 7th of April 2021, for the 4th floor of R10. No more than 18 persons have been detected at the same moment on that floor. The upper left corner shows two different datasets that could be imported as csv files. The file “personcounts” will give the same information that is showed in the graph.

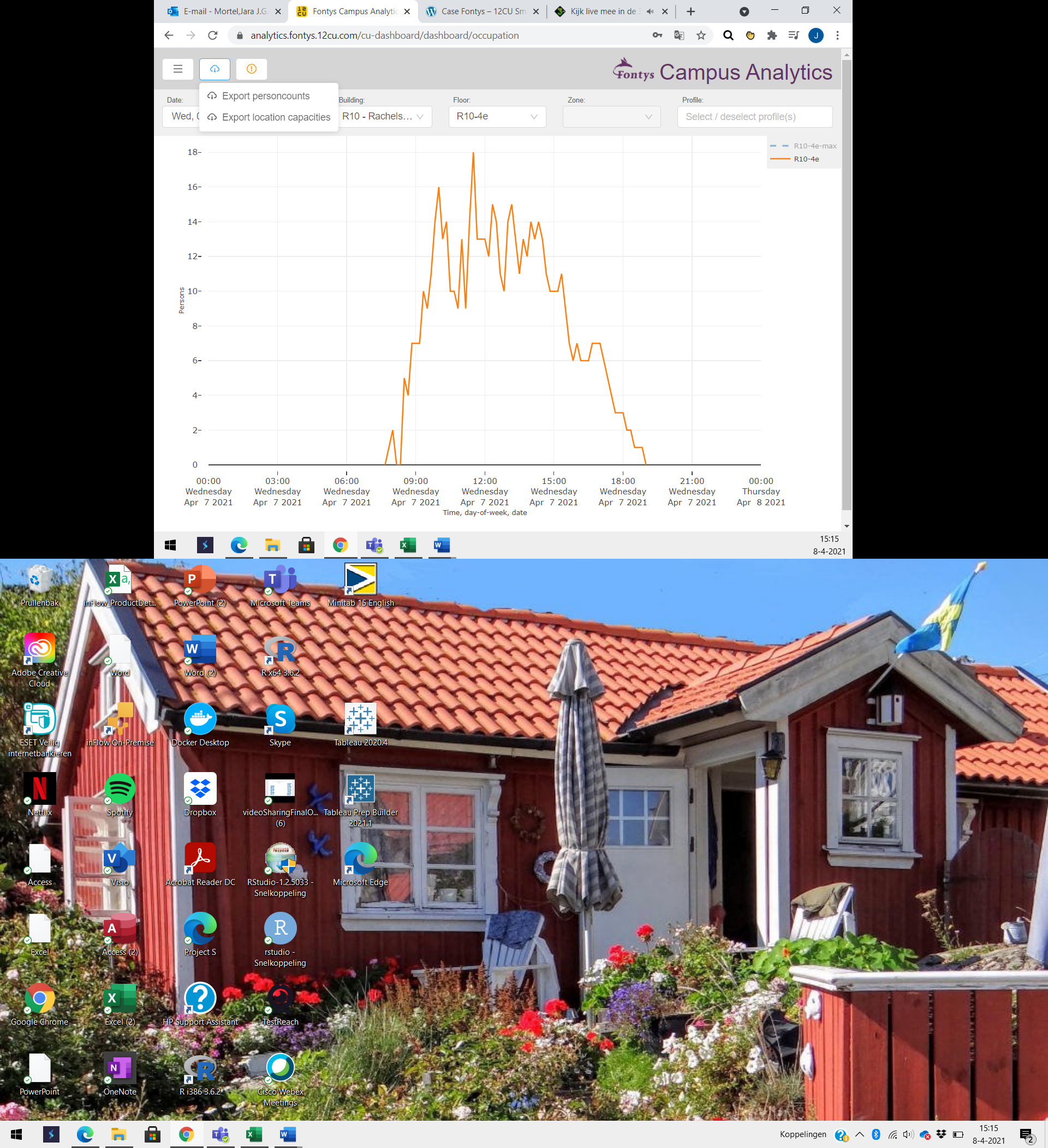


Figure 1: Available information of 12CU

## Parameters from Priva sensordata

Fontys placed several sensor types, which results in dozens of different individual Priva sensors in R10. NovoData only looks at the light, air quality and room temperature sensors for this project. Based on literatures studies, these parameters have relatively much influence on study success. Other sensors were inlet and outlet temperature, ventilation, and wind speed. The students cannot directly perceive these parameters, so they will not be considered. The required data and reasons are summed up in the table below.

|  |  |
| --- | --- |
| **Required data** | **Reason** |
| Light | The amount of daylight in a room has influence on study success. Students will be more focused and productive in lighter rooms. (Preto & Gomes, Lighting in the Workplace: Recommended, 2020) |
| CO2 | The amount of CO2 in an area can be an influence on the concentration levels of students. Students are more focused in rooms with a low CO2 concentration. (Wargocki, Wyon, & Fanger, PRODUCTIVITY IS AFFECTED BY THE AIR QUALITY IN OFFICES, 2000) |
| Room temperature | The temperature should not be too high or too low, because that will have a negative influence on the concentration of students. This is the reason we want to take room temperature in account. (Wargocki, Porras-Salazar, & Contreras-Espinoza, The relationship between classroom temperature and children’s performance in school, 2019) |

## External sources

The last category to add to the database is external data. This section consists of two parts: section 2.3.1 describes why certain APIs are used for the Fontys database. Section 2.3.2 describes why other APIs are not chosen, which can be useful for next project groups.

### Chosen parameters with API’s

The table below shows an overview of what data NovoData wants to add and why.

|  |  |  |
| --- | --- | --- |
| **Required data** | **URL for API’s** | **Reason** |
| Data about the weather: temperature, wind speed, humidity, UV index, cloudiness, and condition code | <https://dev.meteostat.net/> | NovoData wants to see if the temperature, wind speed and the condition code have influence on the number of students that are absent or present. When the sunlight (based on the condition code) is known, it is maybe possible to open/close the curtains automatically so that there is no bright sunlight in the student's faces. When time is left for NovoData, they can make a manual so that students and lecturers can open or close the curtains manually. The UV index and cloudiness can give an indication of the light intensity outdoors, which Novodata can compare with the light intensity in R10. Lastly, the humidity changes the perception of the temperature. (Palonen, Seppänen, & Jaakkola, 1993) |
| Air quality | <https://openweathermap.org/api/air-pollution> | When the air quality fluctuates much in Eindhoven, then it is maybe a possible come to some advice to open/close windows on certain days. For that, we need to make a trade-off between air quality outside and temperature/CO2 inside. |
| Data from NS, for public transport | <https://apiportal.ns.nl/> | Fontys IT can check the correlation between absence and delays in public transport, to check the influence of these delays on study success. |

***NS API Notice***

According to the NS API documentation the arrivals and departures à minimum of 10 arrivals/departures will be stored for the coming hour. For example, a station like Utrecht could have more then 10 arrivals the next hour, then all this data is stored as a single arrival / departure. Another station like Wijchen has 3 arrivals /departures the next hour. In this case trains are stored as duplicates to get to the minimum of 10.

### API’s that are not (yet) chosen

The table below shows an overview of what data NovoData did not add and why.

|  |  |  |
| --- | --- | --- |
| **Parameter** | **URL for API’s** | **Reason for leaving out the parameter** |
| Public transport: 9292 | <http://api.9292.nl/0.1/locations?lang=nl-NL&q=eindhoven> | A request is needed for this API. |
| Traffic transport: HERE | <https://developer.here.com/products/platform/data> | It is not clear if the outcomes are figures about traffic, of only (heat) maps. A login is needed, and it can only be used for non-commercial purposes. Next groups can however still decide to use this API, as Fontys has connections with HERE. The API is free but limited to 250.000 transactions per month. |
| Traffic transport: Google | <https://stackoverflow.com/questions/4600656/access-googles-traffic-data-through-a-web-service> | It is not clear if this API will really work. |
| Traffic transport: Google (2) | <https://developers.google.com/maps/documentation/roads/get-api-key> | It is not clear if this API is usable for traffic jams. It makes use of Google Cloud and costs 5 dollars per 1000 calls. |
| Traffic transport: OpenRouteService |  | This API gives routes from start until end, but there is no additional information about for example the travel time/speed. |
| Traffic transport: TomTom | <https://developer.tomtom.com/content/traffic-api-explorer#/Traffic%20Flow/get_traffic_map__versionNumber__tile_flow__type___zoom___x___y___format_> | This API looked like the best traffic API that NovoData could find, so they tried this one in Python. Some of the outcomes are the current driving speed and the needed time to drive over a certain road piece. This could maybe give an overview of the traffic jams and the correlation with student absence in R10. However, every API call gives only a road stroke of 1 kilometer. Therefore, Novodata wonders if this API is reliable and decided not to use it. Next groups can decide to do more research about the reliability of this API. |
| Weather data: KNMI | <https://pypi.org/project/knmi-py/> | It seems like this API is out of order. |
| Sun API: SunCalc | <https://www.suncalc.org/#/51.4359,5.4855,4/2021.03.02/20:15/1/3> | This API mentions the place of the sun for every day and every moment of the day, for every room. Novodata could check its reliability with other weather data (sunny/cloudy). This API has – because of limited time – not been implemented yet, but can maybe be useful for Fontys IT. Therefore, it is an idea for next groups to look at it. |
| API for CO2 | <https://docs.pymc.io/notebooks/GP-MaunaLoa.html> or <https://www.tinkerforge.com/en/doc/Software/Bricklets/CO2_Bricklet_Python.html> | Probably only data available for Mauna Loa. However, for the CO2 level there are probably no APIs needed, as Fontys IT can also use averages that are mentioned on internet sites. The CO2 level does not change per hour, but the concentration is rising over time because of the climate change. Next groups could maybe take this into account to compare it with the CO2 concentration in R10. If the CO2 concentration outside is higher than in R10, Fontys can choose to close the windows/ventilation system. |

# Optimal values for parameters

The Priva system makes it possible to automatically act when certain values are too high or too low. The Fontys users that have access to the overview of these Priva sensors can change the bandwidths. These bandwidths are triggers to change the values in the building automatically. To do this in a good way, it is important to know the optimal values for the different parameters, namely for CO2, light, and temperature.

## ­Light intensity

Good light has 21% influence on the performance of students. (Kennisrotonde, 2021) Light intensity or illuminance can be expressed in lux, which has a very broad scale. Before mentioning the optimal light intensity in classrooms, it is important to get a feeling for the number of lux in certain environments. Figure 2 shows some examples. (Torchspot, sd)

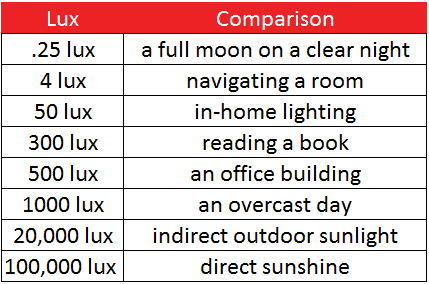


Figure 2: Examples of illuminances

As different rooms have different purposes, they also need different light intensities. Four sources have been consulted to get an idea of the optimal illuminance in a classroom, namely sources from Glamox, Fagerhult (2020), Obimex and KU Leuven (Glamox, sd) (Obimex, sd) (Fagerhult, 2020) and KU Leuven Technologiecampus Gent (2016). Three of them mention that at least 300 lux is needed in classrooms and that 500 to 750 lux is desirable, especially for reading and writing tasks. The education in R10 is based on applied research: the rooms are therefore not like old-fashioned classrooms, but they look more like office buildings because of the big spaces. For these types of rooms, 500 lux is the standard. Some of the sources mentioned before state that it is not desirable to increase the (artificial) light too much, as it does not improve the visual comfort, but it increases the chance of blindness. However, Wouter Boxhoorn stated during the Sports and Technology Seminar (2021) that healthy indoor light should be more than 500 lux. They say that less than 500 lux feels like darkness for the biological systems of humans, and they provide customers with light intensities between 500 and 1200 lux on the desktop. Natural light and higher light intensities have a positive influence on biological entrainment, sleep, well-being, and performance. (Preto & Gomes, Lighting in the Workplace: Recommended, 2020) It seems there is no clear maximum for natural light intensity, which would lead to the conclusion that studying nearby windows increases productivity. Boxhoorn (2021) states that the benefit of natural light is only present within a radius of about three meters from the windows.

To improve the benefits of the biological system, the light of the lamps should be adapted during the day. This focuses on the following parameters:

* Light intensity (illuminance)
* Colors: this can be for example blue, orange, or purple.
* Color temperature. Blue “cold” light has a relatively high value (up to 10.000 Kelvin), and orange “warm” light has a low value (about 1000 Kelvin).

This should correspond with the outside environment, so that people experience the same “natural environment” when they are inside.

Another important factor is the spread of light. Boxhoorn (2021) states that it is best that the light is spread homogeneously, so that every place gets the same light intensity.

Based on these facts, NovoData wants to draw a conclusion on the optimal illuminance in the Fontys R10 building. Fontys can choose to adjust the minimum or maximum desired illuminance for different rooms. NovoData concludes that light intensity should be at least 300 lux, but the optimal illuminance is somewhere between 500 and 1200 lux. Too much light can be annoying for work with laptops/computers.

To be able to adjust the illuminance to optimal values, actual data about light is needed. NovoData wants to measure this for the rooms on the 4th floor of R10. Figure 2 showed some examples of different light values, and the difference between indoor “artificial” light and indirect outdoor sunlight can be about 20.000 lux. It is therefore very important to know where the sensors in R10 are placed. More information about the places of the sensors can be found in section 5.1.

## CO2 values and overall air quality

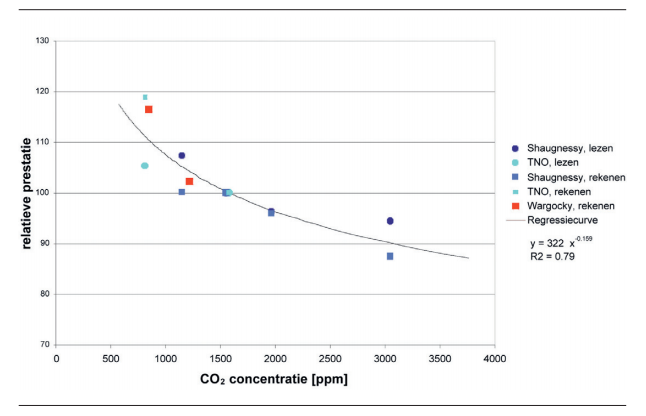
The air quality has 16% influence on the performance of students. (Kennisrotonde, 2021) In the R10 building Priva placed sensors to measure air quality in the classrooms on each floor. The air quality was therefore measured in CO2 (ppm) values. Many studies have been made to find out what the optimal CO2 level is to improve study success. All these studies on different schools let students take an exam on reading and calculation. These exams were taken in classrooms where the ppm concentration has different values and degrees of ventilation.  
  


Figure 3: Relation between relative performance and CO2 levels of concentration

Figure 3 shows the relations between the relative performance of students and the concentration of CO2 in ppm. When you follow the black line, you will see that the performance will drop when the concentration of CO2 gets higher. The black line also starts +/- 550 ppm. This is the optimum value with the best student performance. Ron Bertrams, who works at the housing department of Fontys, states that the air is “clean” up to 500 ppm, and automated changes are executed when the value is between 700 and 800 ppm. This could be an indication that the sensor indeed helps the R10 building to provide a better learning environment. We could use this measurement definitly as an indicator for study succes.

## Temperature

The last factor in the sensor data that has influence on study success, is the room temperature. Kennisrotonde (2021) mentions that the temperature has an influence of 12% on the study success. NovoData has considered four other sources to estimate the optimal temperature in classrooms from Aken (2017), Oebs (2018), Regus (2021) and GGD Rotterdam-Rijnmond (2021). The averages of the minimum and maximum temperature are then 20 and 23 °C. Different of these sources state that the optimum room temperature for focus and performance is 20 °C. However, 21 and 22 °C were the only values that were included in the bandwidths of all sources. NovoData therefore concludes that the room temperature should be between 20 and 22 °C, with 21 °C as an optimum.

## Plants

Plants belong to natural elements. Workers in offices with natural elements like plants (and sunlight) report that well-being and creativity increase with 15% and productivity with 6%. (Boxhoorn, 2021)

R10 also has plants, especially in relatively large rooms: room 4.60 counts 22 plants and room 4.70 counts 14 plants.

# Data to store

This chapter describes what data NovoData wants to use for the new database.

## Available and required data

The table below shows the available ( “Interval/Update from source”) and the desired data (“Desired interval/update”) from the sources used by NovoData. “Interval” explains the time in between two measurements, and “update” explains the time in between updating the new database. More information about these time intervals and updates will be given in the next section.

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter/source** | **Interval from source** | **Desired interval** | **Update from source** |
| Light in R10 | 8 minutes | 10 minutes | 24 hours |
| Temperature in R10 | 8 minutes | 10 minutes | 24 hours |
| CO2 in R10 | 8 minutes | 10 minutes | 24 hours |
| Meteostat API | Every whole hour | Every whole hour or every 10 minutes | 24 hours |
| UV index API | Every hour, at a random moment | Every whole hour or every 10 minutes | Every hour |
| Air quality API | Every whole hour | Every whole hour or every 10 minutes | Every hour |
| Wi-Fi data: occupancy R10 | 10 minutes | 10 minutes | 24 hours |
| NS API | Depends on train arrivals and departures | NovoData wants all available information | 24 hours |

## Update and Intervalplanning

In the previous chapters we talked about how and what data is going to be stored in our Azure database and the database for the PowerBI desktop. In this chapter we are going to talk about when and how often the data is retrieved from the external sources. Figure 1 shows the global architecture of the dataflow with the respective intervals.

### Data from the Priva sensors

The sensors are constantly measuring the parameters, but only snapshots for every 8 minutes are stored the Priva database. From here Fontys IT collects this data every 24 hours in their database. This is the data that we, as the project group, have access to.

Our Azure database is going to collect the data from the Fontys IT database daily at a set time of 9 am. We use a set time because the sensor data from Priva that is collected by Fontys IT can vary per sensor. With this timestamp we can make sure all the data from all the sensor is collected from the previous day.

### Data from other sources

Except for the Wi-Fi data that comes every 10 minutes, the data from other sources have no set intervals like the Fontys IT database had. These intervals can vary per source. For example: the Wi-Fi data is collected every day, but the weather data can have an hourly interval for the API. In this case it is also useful to collect the data from these sources daily.

According to the architecture, as shown in figure 4, this can be done at 9 AM each day. There is still a chance that we then overload the server with many requests form the server. We must find out what is the best option for the scheduled time.

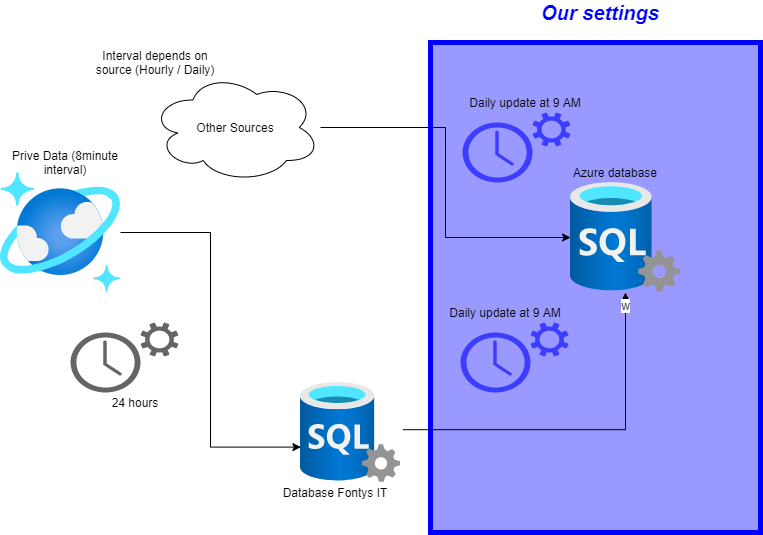


figure 4: Frequency planning architecture

We did not go with the architecture for the scheduling as shown in figure 4. The reason behind this is that some resources like Database Fontys IT are heavy load on the server. This source contains a lot of data. If we would schedule more sources on the same schedule time the server could get overloaded, and processes will be terminated. In the document *architecture serverless* you can find an overview of the scheduled time for each source.

## Structure

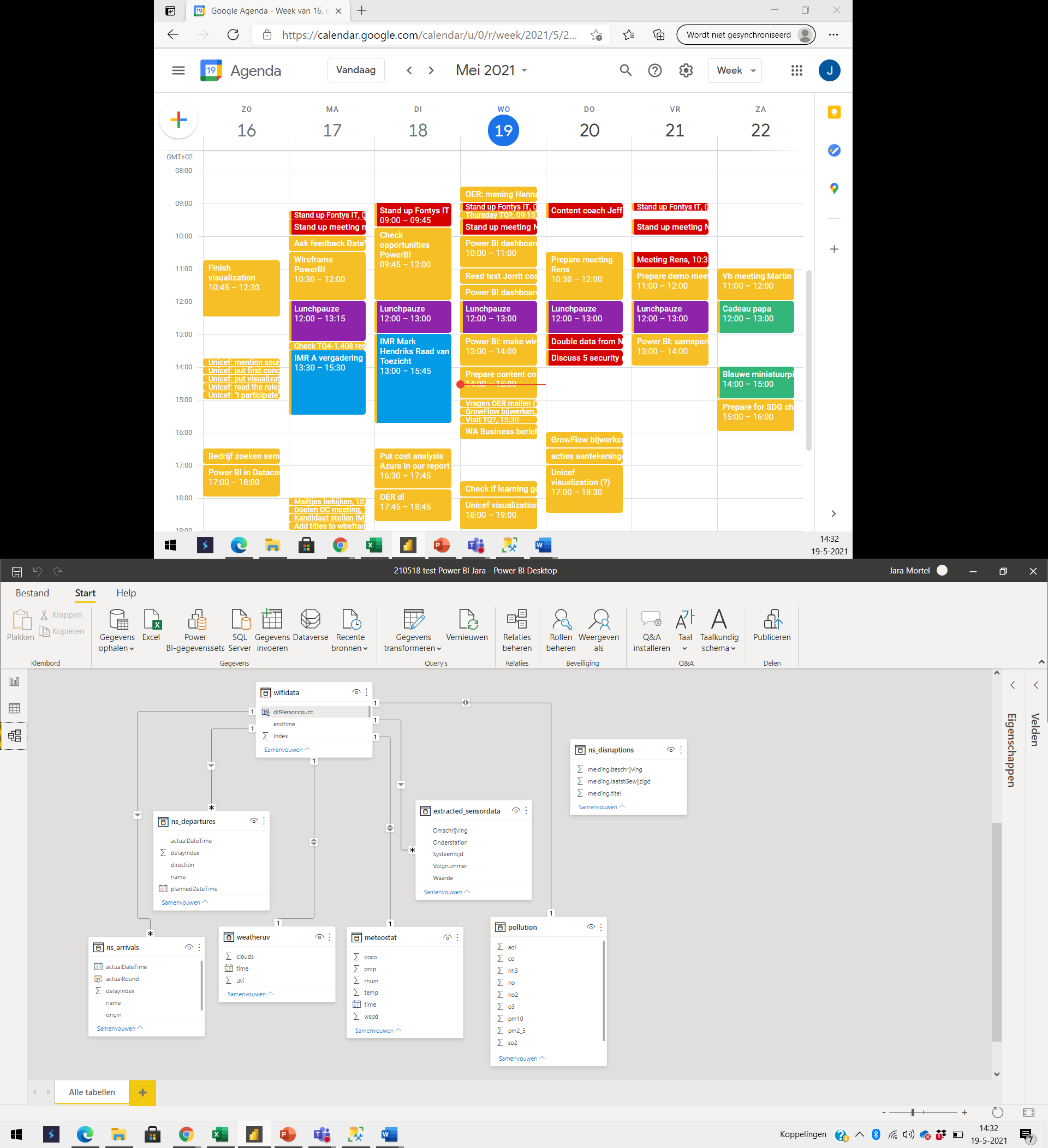
To create simplicity for the first deliverable we can go with 1 database for now. It is easier to maintain, and de data can be shared among different clients(as in departments, employees, software etc.). There are two possible solutions for further splitting up the data.

### Storage of data

The first being we create a table for each source, then use a dashboarding software like tableau for integration with the different data. The other solution is the create views that are application specific. So, for example a view for the scheduling department could have data which involves meteostat, NS data and WIFI data to have information about delays and their causes.

A major disadvantage of creating views is that they do not contain the actual data but a copy of the data that is queried for that specific view. This means that the data in the view needs to be updated every time there is new data available. And another disadvantage is that a view is not scalable like a table. You would be stuck with the data that is available in the view.

A better approach is to create tables for each source in the database and link the tables together on the datetime columns. These links, or key references as the called in SQL, can provide information on relations between two sources. For example, you can take values of CO2 from the sensor data source and the number of students of the Wi-Fi data source to see if there is a relation between number of students and the measured CO2 levels at the same timeframe. (See image below)



### Analysing the data

We want to look at the number of students, that we collected from the WIFI data and compare these numbers with other sources to see if there is a relation between these numbers. Each analysis task will be listed below. The graphs will show an hourly pattern and the user can filter on rooms and date. To find patterns on specific rooms or dates.

1. **Number of students and the measured values of CO2:** *These measures are important to improve performance. Therefore, we want to know if the amount students on the 4th floor have any influences on the CO2 levels. If so, what value must be reached to keep optimal performance in classrooms.*
2. **Room temperature:** *Simple graph that keeps track of the room temperature in different rooms on the fourth floor.*
3. **Number of students and the measured values of Lux (light):** *Light also improves performance. The correct amount of light can help students to read and focus. With this data we should see if the number of students is lower when there is less light available.*
4. **Weather conditions:** *Weather conditions can have influence on the number of students that are present on certain time. Analysis is based on the condition score. For example, with a score 1(snow, rainfall, storm etc.) there are maybe less students present then with clear weather (7 to 10). Apart from the score we can look at more factors like outside temperature.*
5. **Public transport (Nederlandse Spoorwegen):** *For now, we solely look at the arrivals, departures, and disruptions from the NS. To see a pattern in the flow of students.*

## Estimated required storage

NovoData made a database in Azure with an available storage of 2048 MB’s. This database is created in March 2020, and 40 MB’s of this capacity has been used two months later. As a result, 20 MB’s will be added every month. This is almost 0,98%, so Fontys IT can use the database with the current settings for at least 100 months. Table 1 shows the total overview based on the full months March and April.



NovoData had to do some assumptions for this calculation:

* As the same amount of data will be uploaded every day, NovoData concludes that the used storage space rises linearly.
* NovoData assumes that Fontys IT will not do more transactions in the future than already planned at this moment.

These calculations should be done regularly to make sure that the forecast is close to the reality. NovoData made an Excel file called “Export\_AzureCalculator”, in which Fontys IT can recalculate the forecasts by adding number of months the have been payed for. To keep track if history it is best the create a new tab for each month and change the number of months payed.

## Cost calculation for database

The costs are based on two services: The Azure SQL database (2GB) and the Azure Function. To calculate the cost yourself you can have a look at a small document we made to set up the calculator. Table 2 shows the cost of these resources.



Table 2: cost structure Azure database and Azure Function

For the Azure function there will be no costs before the 1.000.000 transactions of 400.000 GB limits. Above 1.000.000 transactions the cost will be 17 cents per million transections each month. The cost for GB per second is 0,0000014 cents after 400.000 GB.

Many apps have legal, compliance or other business purposes that require you to retain database backups for longer periods of time. By using the Long-Term Retention (LTR) feature, you can keep full backups for longer than 10 years. You can then restore each backup as a new database. For this database it means that you can store and restore to 5GB of data as back-up.

The current amount of data, when this document was written, is around 59MB since the beginning of March. That is 2,880859374% of the total 2048 MB (2GB) storage. This is equal to 0,960286458%per month of storage use. A full calculation of this estimate is shown in the table below. We can conclude from this cost analyses that Fontys IT does not exceed their maximum of 2GB soon (estimated around 8,5 years). But that is based on the current situation.

Based on the knowledge that Fontys will expand this project for the next 2 years, it is safer to say the amount of storage should be enough for around 5 years on the current monthly cost. More information about this reasoning can be found in section 4.5.2.

### Calculations based on current situation

The current setting would be good for the database for the next 100 months. After 100 months the capacity is at 97.66% of the available 2GB. Then after 100 months there will be 4.13 euros x 100 = 413 euros paid for the database.

Based on the current monthly data use and assuming they will not do more transactions than currently scheduled, NovoData comes to the following predicted cost for 100 months (time to full storage). The costs for the use of Azure function are none if Fontys has no more the 400.000 GB data use or does less than 1.000.000 transactions in a month.   


Table 3: cost forecast 100 months

### Prediction based on the expansion of the project

With the knowledge that Fontys IT wants to expand their resources and collect more data, we will expect that the database will reach full capacity sooner than the expected 100 months in the current situation.

Assuming a safe position of 5 years (60 months) and based on the current situation of only data form the 4th floor, it comes down to 247,78 total cost. To keep track of the cost and make a better prediction for the future, a new analysis should be done in case of adding more floors or annually,



Figure 5 Forecast on safe timeline of 60 months

## Dashboard data choices

All data is connected via their joint variable: date and time. However, some timestamps do not match because of small differences. The Wi-Fi data has intervals of 10 minutes and the sensor data has intervals of 8 minutes, so only 20% of the Wi-Fi datapoints can be matched with the sensor data. NovoData only used this data for the dashboards. However, next groups can choose to round some timestamps. The trade-off to make is quantity of matches vs. accuracy. Accuracy is more important for variables that fluctuate much because rounding can cause unreliable data. The table below shows again the parameter and time interval. It also shows the changes that can be made for matches, and the consequences for this action.

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter/source** | **Interval from source** | **Possible time customization** | **Consequences** |
| Light in R10 | 8 minutes | Round timestamps to 5 (or 10) minutes. | Now 50% of the datapoints can be compared to the Wi-Fi data, instead of only 20%. However, rounding data can cause unreliable data, as some data initially belonged to a time of 1 or 2 minutes earlier or later! Especially for light this is a big disadvantage, because the illuminance can immediately change due to lamps, sunscreens, and clouds. That is why it is better to leave the dataset like this without rounding, or to switch to 5 minutes instead of 10 minutes. However, the last option would make it possible to use all light datapoints. |
| Temperature in R10 | 8 minutes | Rounded times to 10 minutes. | Now 100% of the datapoints can be compared to the Wi-Fi data, instead of only 20%. However, rounding data can cause unreliable data, as some data initially belonged to a time of 1 to 5 minutes earlier or later. The temperature does not fluctuate much, so rounding 10 minutes could be possible. More research for this must be done by next groups. |
| CO2 in R10 | 8 minutes | Rounded times to 5 minutes. | Now 50% of the datapoints can be compared to the Wi-Fi data, instead of only 20%. However, rounding data can cause unreliable data, as some data initially belonged to a time of 1 or 2 minutes earlier or later. That is why NovoData thinks it is better to round at 5 minutes instead of 10 minutes. That last option would make it possible to use all CO2 datapoints. |
| Meteostat API | Every whole hour | No changes |  |
| UV index API | Every hour, at a random moment | Round to whole hours | This has already been done by NovoData. The datapoints deviate up to 6 minutes from the whole hours. As the UV index does not fluctuate very much, this change has minor negative consequences. |
| Pollution API | Every whole hour | No changes |  |
| Air quality sensor data | 8 minutes | No changes | NovoData did not change anything in the timestamps to compare more datapoints for air quality. However, they made a new air quality index, to be able to compare this with the air quality outside. The disadvantage of this new measurement is the fact that inside and outside air quality indexes use different variables. |
| Wi-Fi data: occupancy R10 | 10 minutes | No changes |  |
| NS data | Depends on trains | Round train arrivals/ departures to 10 minutes | This has already been done by NovoData for the actual arrivals and departures. The idea behind this, is to be able to compare all train routes to other datasets. The disadvantage of losing train connections is bigger than the inaccuracy of a few minutes by rounding timestamps. However, if next groups decide to display the arrival and departure times, it is recommended to use the real (not rounded) timestamps. |

## Future (Im)possibilities with Azure

This section is about the possibilities in Azure. A lot of functions will be discussed later in this section, but Azure has one disadvantage: the updates of the database depend on the Azure Function, so it is not possible to switch to another host. Therefore, next groups cannot search for cheaper options to reduce costs for Fontys IT.

Azure has different options that next groups can explore:

* Machine Learning
* Azure Form Recognizer
* Azure Data Exploring

**Azure Form Recognizer**

*“Azure Form Recognizer applies advanced machine learning to accurately extract text, key-value pairs, tables, and structures from documents. With just a few samples you can tailor Azure Form Recognizer to understand your documents, both on-premises and in the cloud. Turn documents into usable data at a fraction of the time and cost, so you can focus more time acting on the information rather than compiling it. “* (Microsoft Form Recognizer, 2021).

This service could be useful for other groups in future projects, because Fontys IT can learn from form data(online survey data, documents) to predict study success in this way. However, for the next group this is not interesting, because this a step too far and you need to really think about the usefulness of this kind of analytics and take in account the GDPR restrictions.

This service is rather expensive in use. The first 500 pages are free of charge. Then the price will be up to €42,165 per 1.000 pages and another €8,433 per 1000 pages to recognize data from receipts and business cards (Microsoft FR Pricing, 2021)

**Machine Learning with Azure**

Machine learning is giving knowledge to a computer or model by providing data, more important data patterns, so that the model can be trained to predict future outcomes. Fontys, IT wants to apply this to their data to find patterns that can help Fontys to predict the number of students at school, schedule maintenances etc. (Microsoft Azure ML Service, 2021)

Machine Learning Service is expensive and has a steep learning curve, so learning how it works could be very time consuming. We can also suggest making use of Python and / or R to apply machine learning on the data. Their libraries do the same work and are free.

The costs for the West-European region can be between 35 and 11.200 euro’s a month, depending on the storage and processing capacity. The pricing is based on pay per use. (Microsoft\_MLCalc, 2021)

Figure 5 gives an example of what machine learning can do for companies: it can for example predict when a certain car will have issues. Users can check the details by clicking on the issue. (Microsoft Mechanics, 2019)

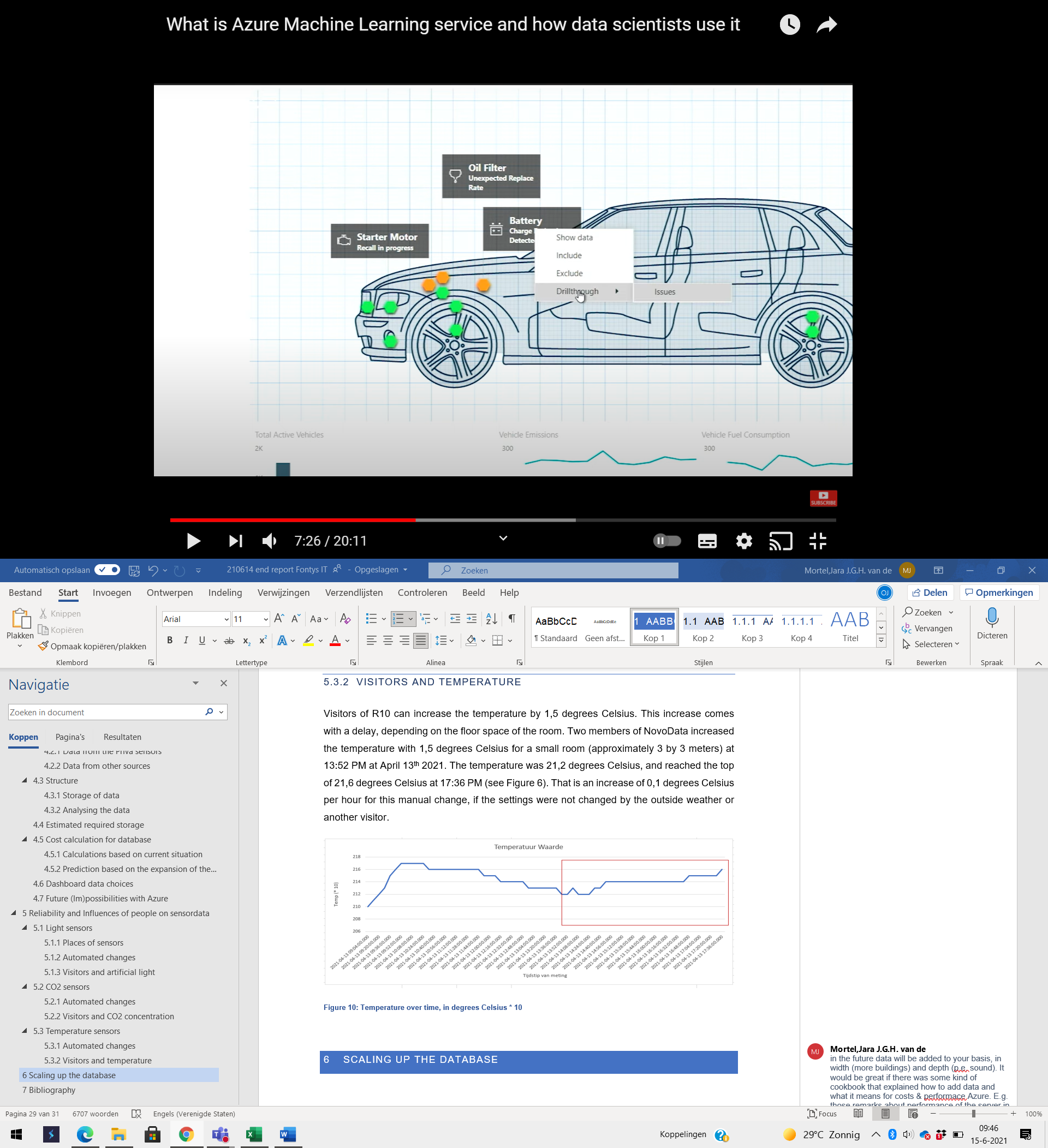


Figure 5: Predictive maintenance with Machine Learning

**Azure Data Explorer**

Azure Data Explorer makes real-time analysis possible to structure data and identify patterns, trends, and deviations in the datasets. Detailed information can be obtained via self-made queries. It helps users to create dashboards and to make sense of the data. The advantage of Azure Data Explorer is that non-data scientists can easily use it without the need of Jupyter Notebook or other data science related tools. (Microsoft Form Recognizer, 2021) An example of possible outcomes is shown in Figure 5. (Azure Data Explorer, 2021) The dashboard can give real time data, which is an advantage compared to Power BI: for that tool, you must refresh the data manually. However, Azure Data Explorer is not cheap: if you pay by use, it is €0,093/core per hour. This is €67,72/core for one month. (Microsoft Azure, 2021)

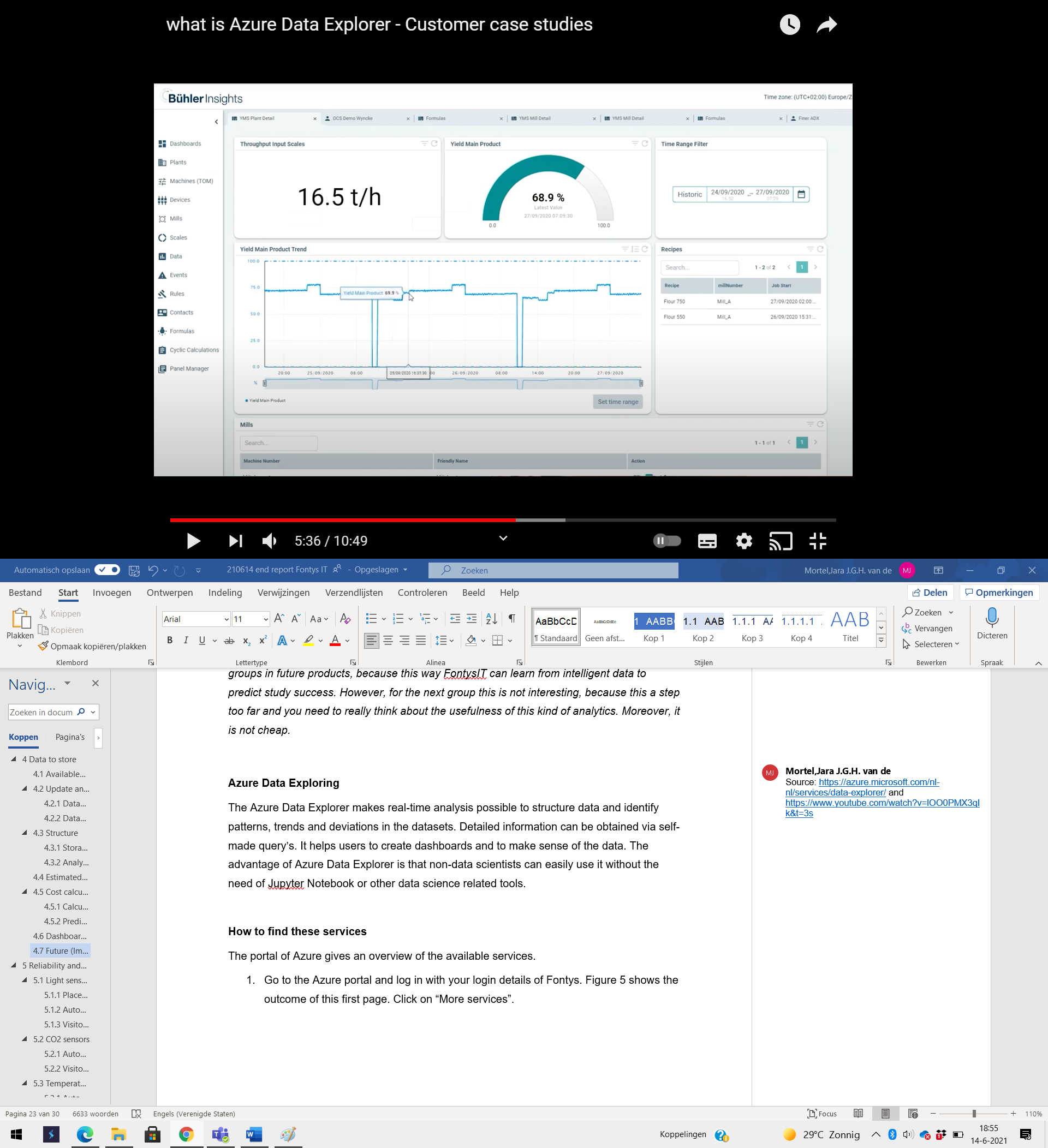


Figure 6: Dashboard of Azure Data Explorer

**How to find these services**

The portal of Azure gives an overview of the available services.

1. Go to the Azure portal and log in with your login details of Fontys. Figure 6 shows the outcome of this first page. Click on “More services”.

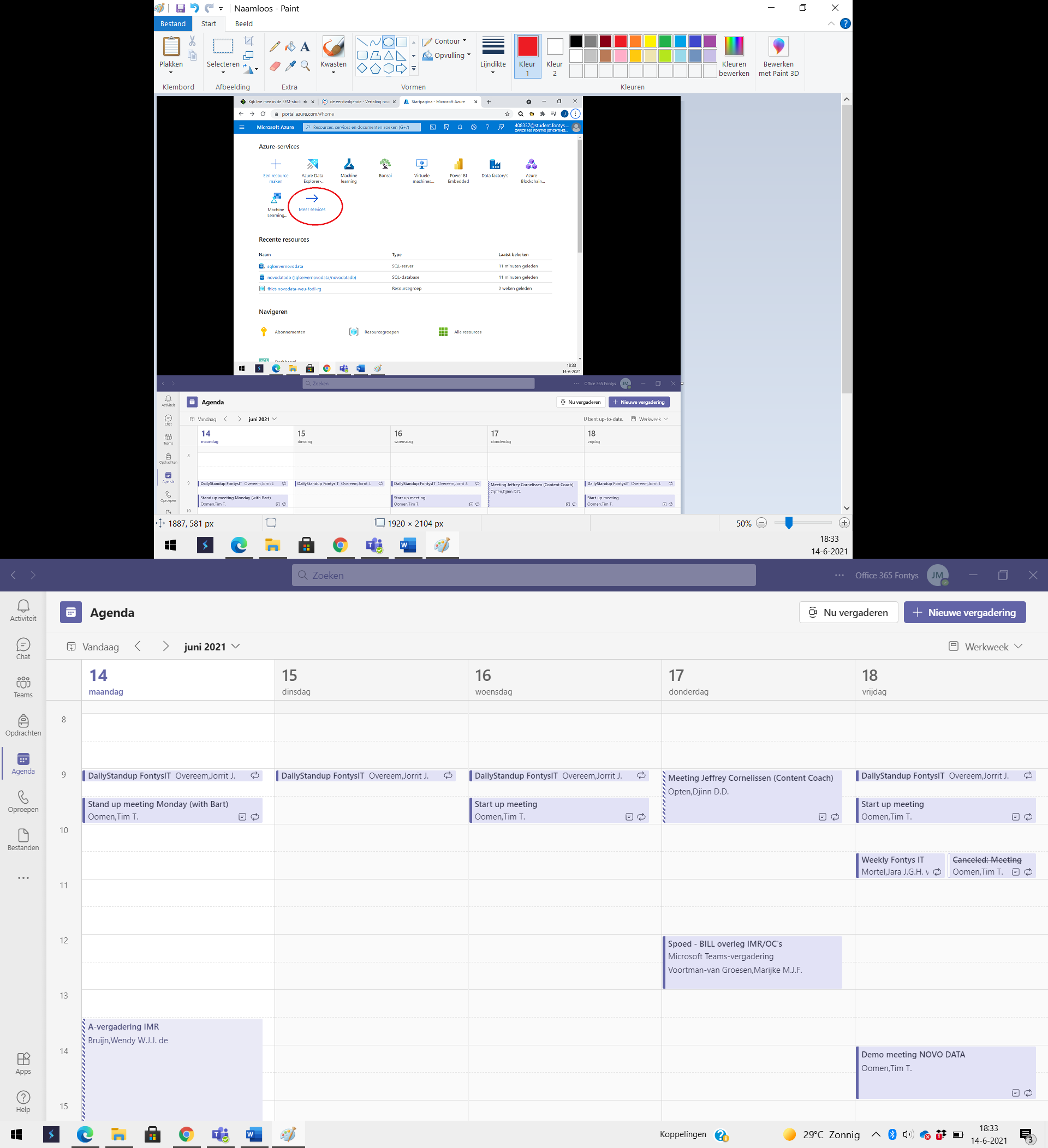


Figure 7: Azure home portal

1. All categories of services are shown on the left side (see Figure 6). If you click on a category, the portal will show detailed services for that category. Some services are visible in multiple categories.

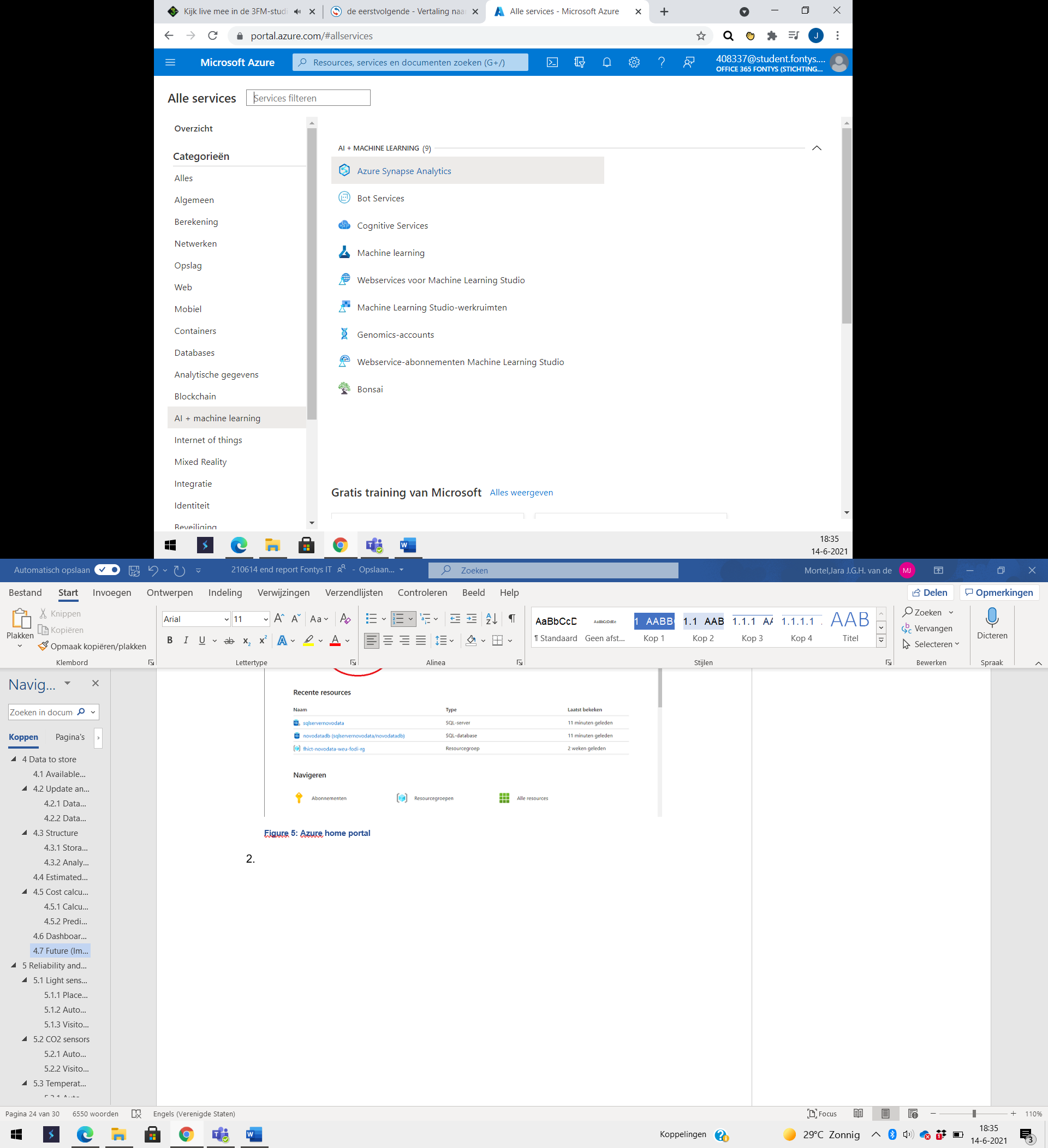


Figure 8: Azure services per category

# Reliability and Influences of people on sensordata

To check the reliability and flexibility of the measurements, it is important to know the influences of automated settings and visitors. Visitors can change settings manually, or (unintentionally) by sudden changes in the number of visitors in a certain room. The impact of automated settings and persons on the 4th floor will be explained for each parameter separately. The conclusions are based on findings on April 13th, 2021, during a visit to R10.

## Light sensors

This section is separated into three parts, that explain the influence of the sensor place, automated changes, and visitors on the light measurements.

### Places of sensors

As mentioned in section 3.1, the distance between the windows and light sensors have influence on the measurements. If the sensors measure a certain value, it does not mean that the whole room has that exact value of illuminance. NovoData noticed that the light sensors are two to three meters away from the windows, and 15 to 40 cm from the lamps. This can give a distorted picture of the outcomes for the illuminance.

### Automated changes

The level for light indoors is set on 500 lux. The light goes on when people enter the room, which is based on motion sensors. If there is too much light based on measurements outdoors, then the sunscreens go down. This is managed centrally, but it is possible to change it manually per room.

### Visitors and artificial light

Visitors cannot put on/off the light manually for all rooms. In that case, sensors detect people and put on the light automatically. After some time (estimated to be 30 minutes) the lights go out when the last visitor has left the room. Figure 5 shows the average illuminance in different cases, measured in three different rooms. The data is only based on nineteen measurements, but it looks like the sunscreens have more influence on the illuminance than the lights. However, it should be noted that these measurements have been done on a sunny day. How much natural daylight fills the room depends on the position of the sun and the position of the room in the building. This could also have an influence on how much lux is measured at a given time.

The right side of the table shows the illuminance for lights on and off with the sunscreen down. There is no significant difference, so the expectation is that the lights have no function when the sun is shining outside. More research should be done to confirm this statement, but it could save Fontys some money if this is really the case.



Figure 9: measured illuminance per situation

Section 3.1 also mentioned that the color and color warmth has influence on the light perception. The color warmth in R10 cannot be changed, unlike the light intensity. The lamps have white or slightly blue colors with an estimated value of 5000 Kelvin.

## CO2 sensors

This section is separated into two parts, that explain the influence of the automated changes and visitors on the CO2 measurements.

### Automated changes

As mentioned before, the air is “clean” up to 500 ppm, and their Priva sensors have a maximum value of 700-800 ppm before the ventilation kicks in. As mentioned by Bron (2020), this is based on the presence of visitors. This means that the number of people has a strong influence on the CO2 concentration. Section 5.2.2 will explain this with the use of an example.

Because of the ventilation, it is not necessary to open windows to reduce the CO2 concentration.

### Visitors and CO2 concentration

The number of visitors has influence on the CO2 concentration in the air. Three members of NovoData experimented with this, by staying in a classroom for a longer time and checking the CO2 concentration for that specific room and time afterwards. Before entering the room, the CO2 concentration was about 445 ppm. After staying there for 20 minutes, the CO2 concentration reached 472 ppm. Figure 5 shows the influence of the visitors in a graph. There is a clear correlation between the presence of the visitors and the increased CO2 concentration.

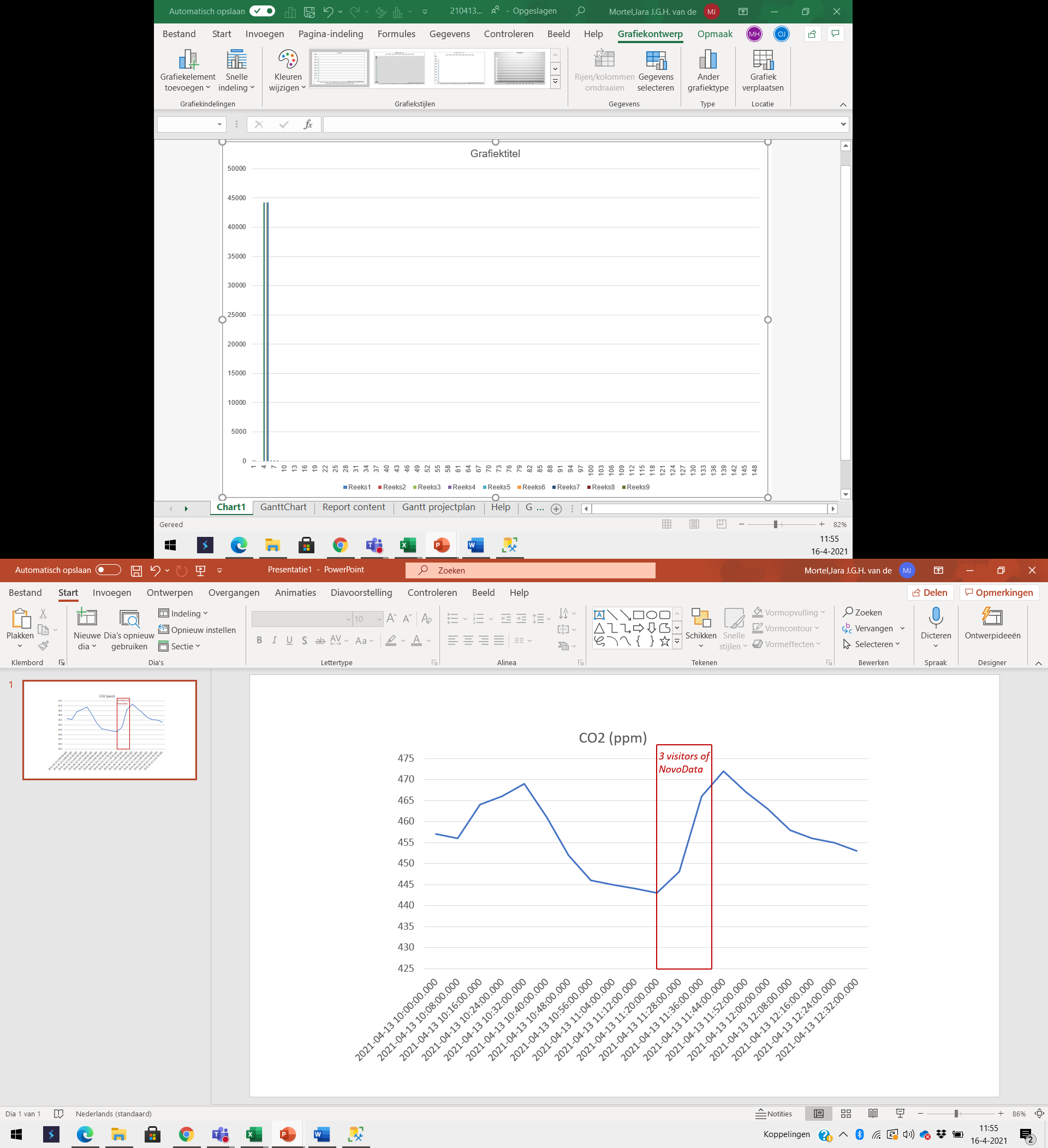


Figure 10: CO2 over time, in parts per million

## Temperature sensors

This section is separated into two parts, that explain the influence of the automated changes and visitors on the temperature measurements.

### Automated changes

Fontys made a basic setting for temperature when people enter a certain room. The inlet temperature will be changed when the temperature is too high or too low.

### Visitors and temperature

Visitors of R10 can increase the temperature by 1,5 degrees Celsius. This increase comes with a delay, depending on the floor space of the room. Two members of NovoData increased the temperature with 1,5 degrees Celsius for a small room (approximately 3 by 3 meters) at 13:52 PM on April 13th, 2021. The temperature was 21,2 degrees Celsius and reached the top of 21,6 degrees Celsius at 17:36 PM (see Figure 6). That is an increase of 0,1 degrees Celsius per hour for this manual change, if the settings were not changed by the outside weather or another visitor.

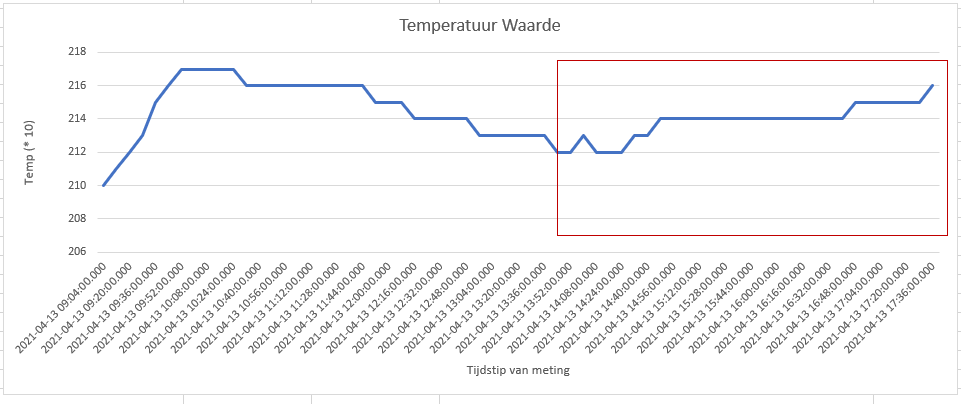


Figure 11: Temperature over time, in degrees Celsius \* 10

# Bibliography

Aken, J. v. (2017, oktober 9). *Nieuws*. Retrieved from AOb: https://www.aob.nl/nieuws/juridisch-advies-koud-in-de-klas/

Azure Data Explorer. (2021, March 18). *what is Azure Data Explorer - Customer case studies*. Retrieved from YouTube: https://www.youtube.com/watch?v=lOO0PMX3qIk&t=3s

Boxhoorn, W. (2021, Januari 27). Sports en Technology Seminar. *Sports en Technology Seminar*.

Bron. (2020, juli 15). *Paradepaardje R10 eerste Fontys-pand met label A++++.* Retrieved from Bron: https://bron.fontys.nl/PARADEPAARDJE-R10-EERSTE-FONTYS-PAND-MET-LABEL-A/

CMD METHODS. (2020, 09 24). *CMD METHODS PACK*. Retrieved from cmdmethods.nl: www.cmdmethods.nl

Fagerhult. (2020). *KLASLOKALEN Een inclusieve leeromgeving*. Retrieved from Fagerhult: https://www.fagerhult.com/nl/kennis-centrum/light-guides/onderwijs-en-leeromgeving/klaslokalen/

GGD Rotterdam-Rijnmond. (2021). *Te heet op school*. Retrieved from GGD Rotterdam-Rijnmond: https://www.ggdrotterdamrijnmond.nl/professionals/te-heet-op-school/

Glamox. (n.d.). *Klaslokalen*. Retrieved from Glamox: https://glamox.com/nl/solutions/test-small-class-room#:~:text=De%20minimale%20verlichtingssterkte%20in%20het,hinderlijke%20verblinding%20zijn%20ook%20hoog.

Kennisrotonde. (2021). *Vragen en antwoorden*. Retrieved from Kennisrotonde: https://www.kennisrotonde.nl/vraag-en-antwoord/onderwijs-in-de-buitenlucht#:~:text=Te%20hoge%20temperaturen%20gaan%20gepaard,belangrijk%20voor%20het%20cognitief%20presteren

KU Leuven Technologiecampus Gent. (2016). *Licht in scholen - van beginner tot expert.* Leuven: W. Ryckaert.

Microsoft Azure. (2021). *Prijzen voor Azure Data Explorer*. Retrieved from Microsoft Azure: https://azure.microsoft.com/nl-nl/pricing/details/data-explorer/

Microsoft Azure ML Service. (2021). *Azure Machine Learning*. Retrieved from Azure Machine Learning: https://azure.microsoft.com/nl-nl/services/machine-learning/

Microsoft Form Recognizer. (2021). *Azure Form Recognizer*. Retrieved from Azure Form Recognizer: https://azure.microsoft.com/nl-nl/services/form-recognizer/

Microsoft FR Pricing. (2021). *Prijzen voor Azure Form Recognizer*. Retrieved from Prijzen voor Azure Form Recognizer: https://azure.microsoft.com/nl-nl/pricing/details/form-recognizer/

Microsoft Mechanics. (2019, October 28). *What is Azure Machine Learning service and how data scientists use it*. Retrieved from YouTube: https://www.youtube.com/watch?v=X7GR4ANn45s&t=446s

Microsoft\_MLCalc. (2021). *Prijscalculator*. Retrieved from Prijscalculator: https://azure.microsoft.com/nl-nl/pricing/calculator/

Obimex. (n.d.). *Het menselijk oog*. Retrieved from Obimex: https://www.obimex.nl/pdf/MenselijkOog.pdf

Oebs. (2018, december 21). *De ideale setting om te studeren*. Retrieved from Oebs: https://oebs.be/de-ideale-setting-om-te-studeren/#:~:text=Op%20basis%20van%20het%20onderzoek,21%20en%20de%2023%20graden.

Palonen, J., Seppänen, O., & Jaakkola, J. J. (1993). *The Effects of Air Temperature and Relative*. Retrieved from https://www.aivc.org/sites/default/files/airbase\_7553.pdf

Preto, S., & Gomes, C. C. (2020, November 9). *Lighting in the Workplace: Recommended*. Retrieved from https://www.researchgate.net/profile/Sandra-Preto/publication/325960961\_Lighting\_in\_the\_Workplace\_Recommended\_Illuminance\_lux\_at\_Workplace\_Environs/links/5b5b1667458515c4b24b6c67/Lighting-in-the-Workplace-Recommended-Illuminance-lux-at-Workplace-Environs.

Preto, S., & Gomes, C. (n.d.). *Lighting in the Workplace: Recommended Illuminance (lux) at Workplace Environs*. Retrieved from https://www.researchgate.net/profile/Sandra-Preto/publication/325960961\_Lighting\_in\_the\_Workplace\_Recommended\_Illuminance\_lux\_at\_Workplace\_Environs/links/5b5b1667458515c4b24b6c67/Lighting-in-the-Workplace-Recommended-Illuminance-lux-at-Workplace-Environs.

Regus. (2021). *Te warm of te koud? De kantoortemperatuur is belangrijker dan u denkt*. Retrieved from Regus: https://www.regus.nl/work-netherlands/nl-nl/hot-cold-office-temperature-matters-think/#:~:text=Uit%20onderzoek(1)%20is%20gebleken,bevorderen%20de%20ontwikkeling%20van%20ziekten.

Torchspot. (n.d.). *Lumens, Lux and Candela Explained*. Retrieved from Torchspot: https://www.torchspot.com/lumens-lux-and-candela/

Wargocki, P., Porras-Salazar, J. A., & Contreras-Espinoza, S. (2019, Juni 15). *The relationship between classroom temperature and children’s performance in school*. Retrieved from ScienceDirect: https://www.sciencedirect.com/science/article/abs/pii/S0360132319302987

Wargocki, P., Wyon, D. P., & Fanger, P. O. (2000, Januari). *PRODUCTIVITY IS AFFECTED BY THE AIR QUALITY IN OFFICES*. Retrieved from http://www.seedengr.com/Productivity%20is%20Affected%20by%20the%20air%20quality%20in%20Offices.pdf