D1

Hello, so today we are going to talk about some results we got from our project answering to our problematic which is "How do the buildings' characteristics impact their yearly energy consumption and can we cluster some different consumption profiles"

As a quick reminder, we will go trough some of the concepts I introduced during the last seminar:

D2

The first program I talked about was about the categorical and continuous energy consumption averages we can retrieve using the given properties of the buildings. The categorical features such intended uses or postal codes can be easily grouped because of their different names contrary as the continuous features, such as the gross area or the volume which have large ranges of possible values.

The algorithm consists of creating a list of characteristic categories or ranges. Then we browse this list and retrieve the average energy consumption of a certain number of buildings per category or range by merging the metadata and the consumption records using the property id.

At the same time, we are checking if we can actually retrieve the consumption, it happens that the records are partly missing or invalid.

Finally, we group these means by category or range to have a wider viewpoint of energy consumption.

And we can display these results with graphs to visualize them.

- List of characteristic categories

```
['511 Educational building buildings', '231 Kindergartens', '214 Health Centers']
```

- Average energy consumption by building

2	511 Educational building buildings	656301	NaN	26.44
3	511 Educational building buildings	677302	NaN	NaN
4	511 Educational building buildings	657302	NaN	NaN
5	231 Kindergartens	688202	NaN	NaN
6	231 Kindergartens	646101	32.77	13.63

- Average consumption group by category

|	intended_use	average_heat	average_electricity
0	231 Kindergartens	32.77	12.35
1	511 Educational building buildings	NaN	31.29
2	999 Buildings nec	30.52	12.41

D3

Here is an example of graphs we can have from this program showing the proportion of invalid and valid data by floor count and the heat (in orange) and electricity consumption (in blue) by floor count, which is a categorical feature. However, the data description provided didn't indicate the energy consumption unit measure, but we can guess from our findings it is probably in kW/h. The mix of these two graphs allows us to see which category is more or less relevant. For example, there is only one building with 0 floor and 2 buildings with 5 floors which have valid data, so we should be careful to not make too much assumption on these two types of buildings.
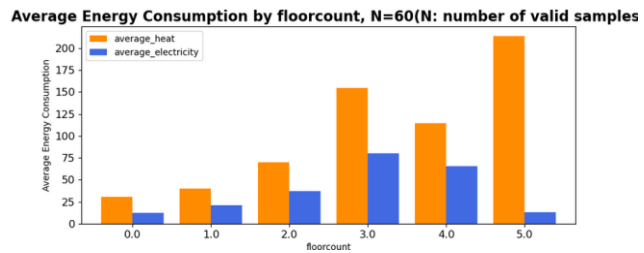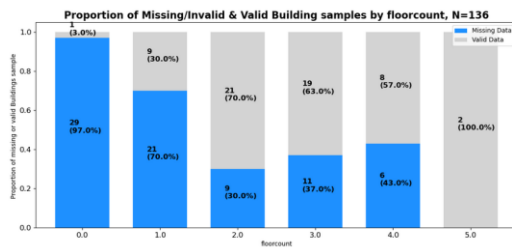
D4

And here is another example with continuous feature, which is the renovation year. The ranges are logarithmical with some continuous data (for example volume or gross area) in order to have to have a more relevant result. We can also display the proportion of invalid data the way it is displayed in the left graph. Moreover, since buildings renovated before 1976 don't have valid records, they are not displayed in the energy consumption graph on the right.
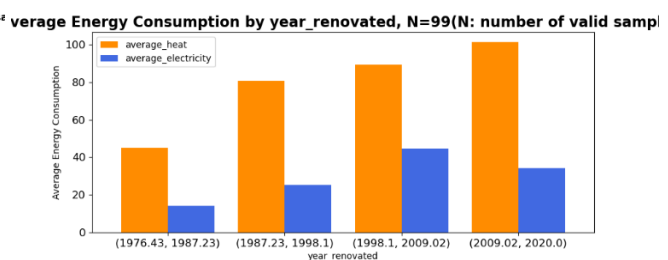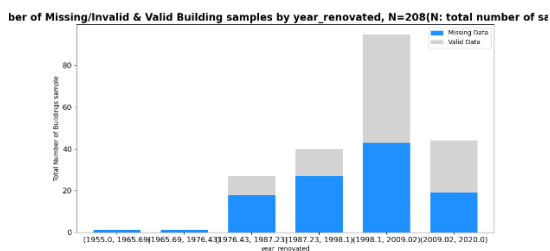

D5

So, these are some interesting graphs I collected using this program.

First, on these 3 graphs, I noticed that the size of a building linearly impacts the general energy consumption as we can see on these three graphs: the more building's gross area, volume or floor count is, the more energy it consumes. But surprisingly, we can that the electricity consumption average decreases as the floor count exceed 4. Unfortunately, this is probably due to the lack of recordings for buildings having more than 4 floors.
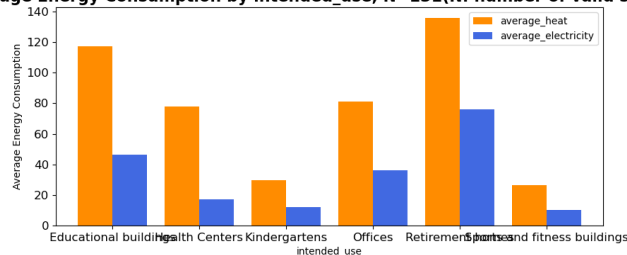
And again, the data description didn't provide the unit measure of volume and gross area but we can guess they are cubic and square meters.
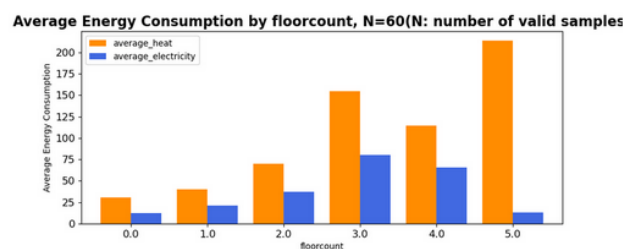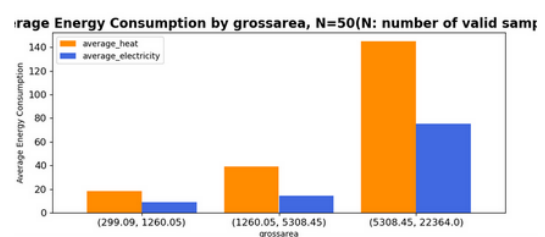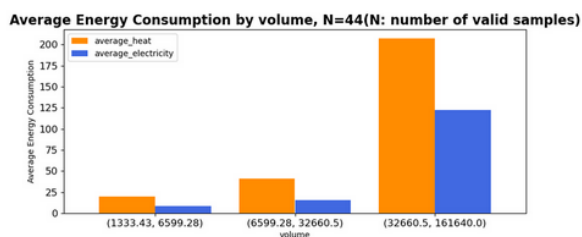
D6

I also realized that some intended uses have more impact, for example, here Retirement homes and educational buildings clearly show a difference with the other buildings.
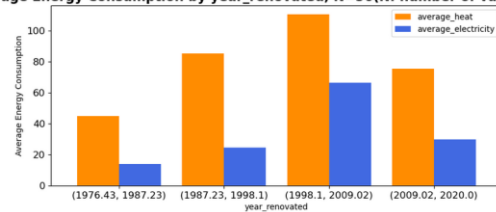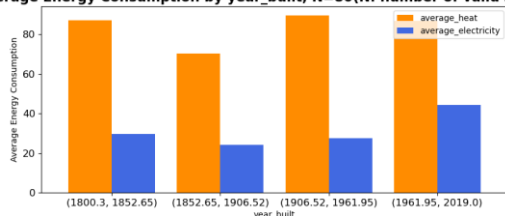

D7

And finally, although the year of construction doesn't seem to have a clear impact on buildings' energy consumption, the year of renovation shows some interesting results: buildings which were renovated between 1998 and 2009 are more consumptive than the buildings renovated before and after this time window.

###

Then, to have a better idea of how these different buildings energy consumption change over time, we can plot the average heat or electricity for each characteristic's category or range by time window. For example: _____

###

D8

So, with these findings, we can try to focus on some particular buildings' properties relationship towards the energy consumption.

For example, I started to focus on the most recorded intended uses because the more data we have, the more reliable it is. And so, the 5 most recorded intended uses are educational buildings, Kindergartens, Health Centers, Sports buildings and Retirement homes.
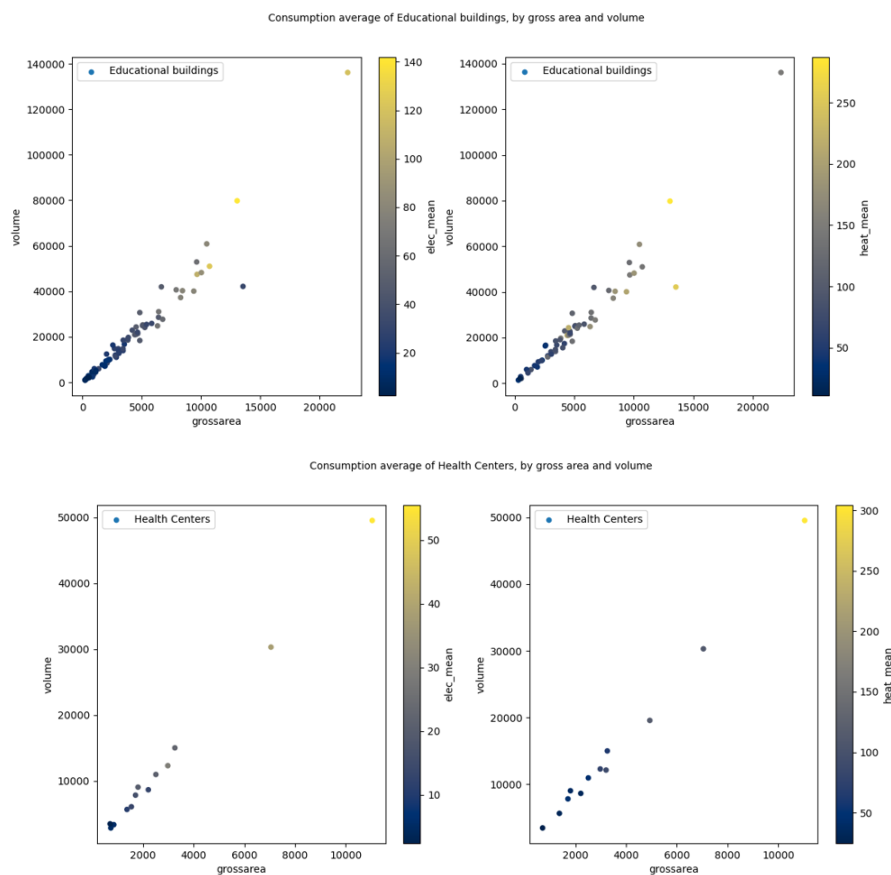
If we look at the size of these intended uses using the volume attribute, we have this graph, showing that educational buildings, Retirement homes and Health centers are the biggest building by average.

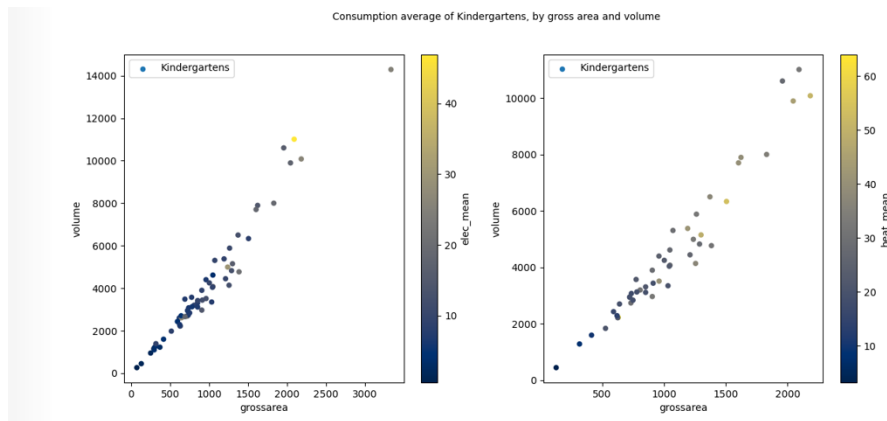Building average volume according to their intended_use
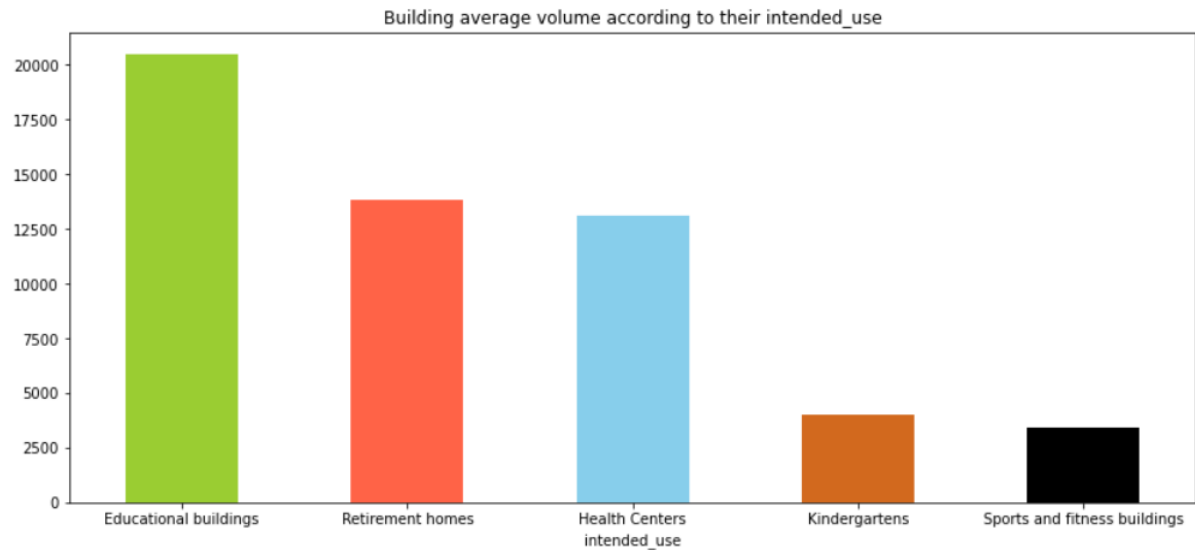
The educational high average can easily be explained by the large campuses in the region such as Linnanmaa and Kontinkangas. However, since health centers use a lot of energy consumptive infrastructures (scanners, heating for all bedrooms or bathrooms), we can wonder if their energy consumption will exceed the educational buildings in some cases.

This why I created two types of graphs by using some parts of my first feature extraction program to explore this question.

First graphs:

Consumption average of Educational buildings, by gross area and volume

Consumption average of Health Centers, by gross area and volume

Consumption average of Kindergartens, by gross area and volume

On this first set of graphs, we are plotting building's volume in function of their gross area, on top of having a shade of colors indicating the electricity average consumption of buildings on the left and the heat average consumption on the right.

We can first see that there is a clear correlation between buildings' gross area and volume. And most of the time, the size of the buildings linearly impacts their energy consumption but there are some surprises.

For example, the heat consumption of the educational buildings and kindergartens can be important as the building is small and vice-versa. This can be seen by the shade change not being very smooth. That means the heat consumption of these buildings doesn't always depend on their size or at least not as much as in the other cases.

Second graphs:

Heat Consumption By volume and Group By Intended Use

Electricity Consumption By volume and Group By Intended Use
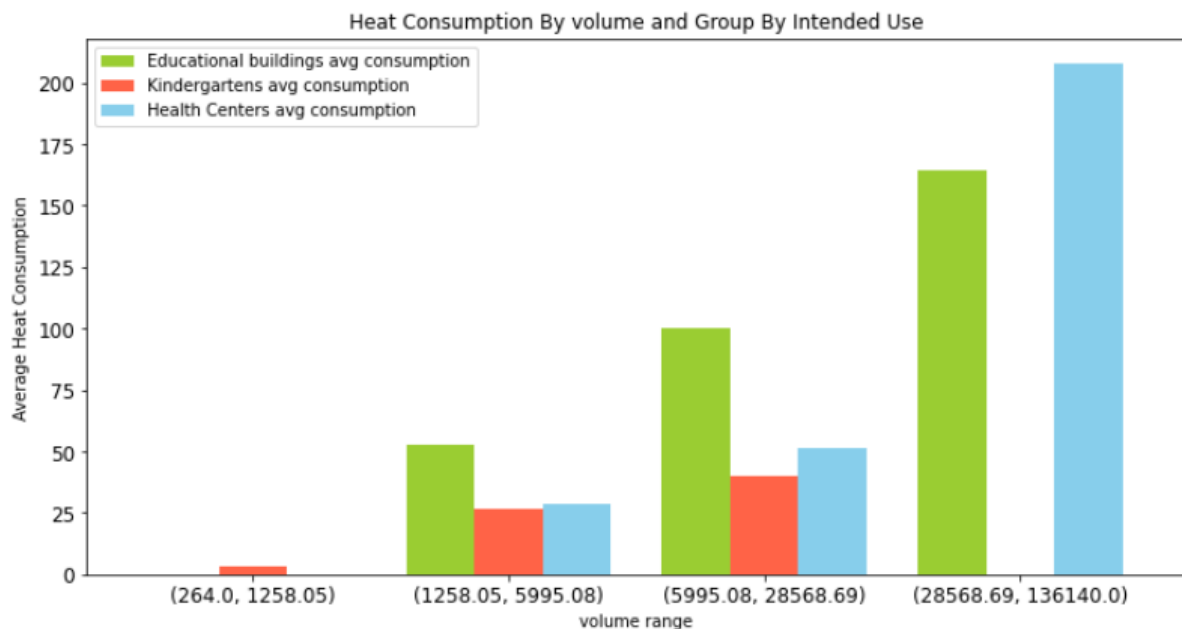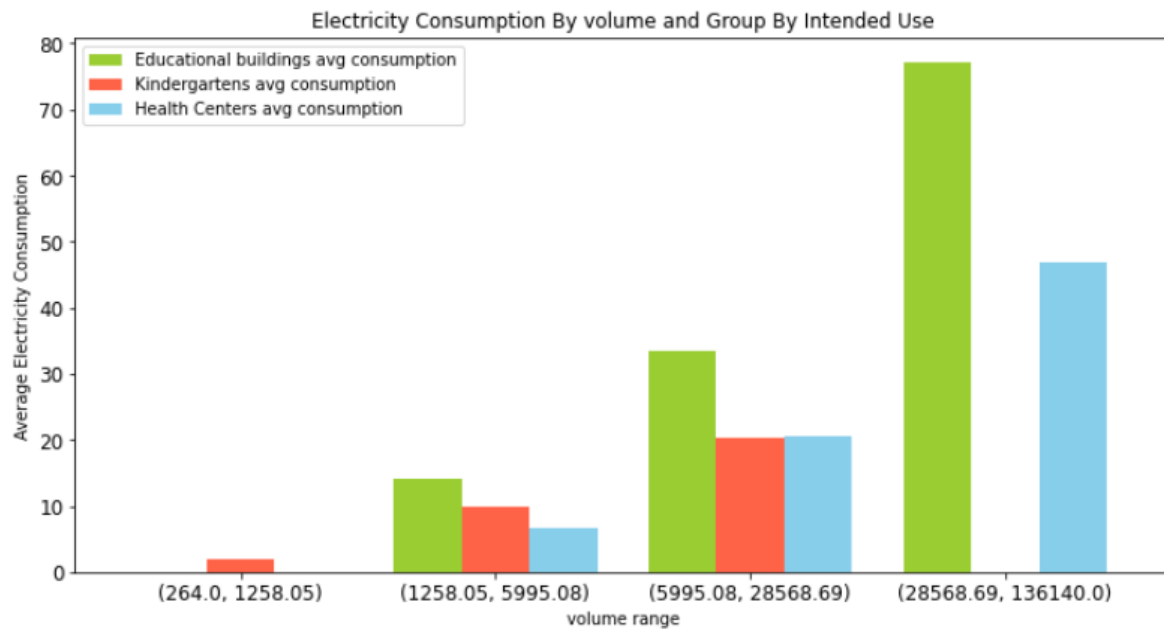
On this second set of graphs, as you can see, I am using the continuous program by using the volume as the x axis and the energy consumption average as the y axis. And finally, instead of grouping the data by the energy consumption type, I choose heat (or electricity) and I group the data by the 3 most recorded intended uses (Educational buildings, Retirement homes and Health centers). Unfortunately, the graphs are not complete because of some missing and invalid data so we like some bars.

Anyway, here, we can see that educational buildings are the most consumptive type of buildings in most cases. However, for very voluminous buildings (so in the last volume range), the health centers' heat consumption actually exceeds the educational buildings' heat consumption. That means that large health centers such as hospitals are consuming more heat than large educational buildings such as universities but less electricity.