VISUAL ANALYSIS OF AIR QUALITY DATA AT UPC

OLGA VILLAR CAIRÓ

Thesis supervisor
PERE PAU VÁZQUEZ ALCOCER (Department of Computer Science)

Degree
Bachelor's Degree in Data Science and Engineering

Bachelor's thesis

Facultat d'Informàtica de Barcelona (FIB)
Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona (ETSETB)
Facultat de Matemàtiques i Estadística (FME)

Universitat Politècnica de Catalunya (UPC) - BarcelonaTech

28/06/2024

First of all, I would like to express my deepest gratitude to my thesis supervisor, Professor Pere-Pau Vázquez, for his support and guidance during the development of this thesis. His dedication has been remarkable; I could rely on him at all times, and he has been a constant source of motivation.

I would also like to thank my family and friends, particularly my parents, Àngels and Jorge, for their unconditional love, support, and encouragement. My father has inspired me over the years with many stimulating conversations, and my mother has been an example of love, responsibility, and kindness. Furthermore, my two big sisters, Alba and Sònia, have always been there for me, serving as role models. To all my loved ones who have been an incredible source of motivation through my academic journey, thank you. This achievement would not have been possible without you.

Abstract

This project analyzes visually the air quality of several UPC classrooms in terms of temperature, CO_2 concentration and relative humidity, adding some insights about energy consumption too. Using mainly the Streamlit and Altair frameworks, various visualizations were carefully designed and integrated into a web application, highlighting key takeaways. The aim is to provide professionals with a tool that allows them to find patterns and assess classroom quality, enabling informed decisions for further improvement actions. Results indicate that the air quality is mostly outside the comfort zone recommended for workplaces, particularly in terms of temperature. The data reveals unhealthy CO_2 concentration levels, especially on the third floor of the Segarra building (ETSAB). These findings suggest the need for actions to improve classrooms air quality to ensure students well-being.

Aquest projecte analitza visualment la qualitat de l'aire de diverses aules de la UPC en termes de temperatura, concentració de CO_2 i humitat relativa, afegint també algunes perspectives sobre el consum energètic. Utilitzant principalment les plataformes Streamlit i Altair, es van dissenyar i integrar curosament diverses visualitzacions en una aplicació web, afegint-hi les conclusions principals. L'objectiu és proporcionar una eina pels professionals que els permeti trobar patrons i avaluar la qualitat de les aules, facilitant decisions informades per a dur a terme futures accions de millora. Els resultats indiquen que la qualitat de l'aire està majoritàriament fora de la zona de confort recomanada per a llocs de treball, especialment pel que fa a la temperatura. Les dades revelen nivells possiblement perjudicials de concentració de CO_2, especialment al tercer pis de l'edifici Segarra (ETSAB). Aquests descobriments suggereixen la necessitat de prendre mesures per millorar la qualitat de l'aire de les aules i així assegurar el benestar dels estudiants.

Este proyecto analiza visualmente la calidad del aire de varias aulas de la UPC en términos de temperatura, concentración de CO_2 y humedad relativa, añadiendo también algunas perspectivas sobre el consumo energético. Utilizando principalmente las plataformas Streamlit y Altair, se diseñaron e integraron cuidadosamente varias visualizaciones en una aplicación web, destacando las conclusiones clave. El objetivo es proporcionar a los profesionales una herramienta que les permita encontrar patrones y evaluar la calidad de las aulas, facilitando decisiones informadas para futuras acciones de mejora. Los resultados indican que la calidad del aire está mayoritariamente fuera de la zona de confort recomendada para lugares de trabajo, especialmente en cuanto a la temperatura. Los datos revelan niveles posiblemente perjudiciales de concentración de CO_2, especialmente en el tercer piso del edificio Segarra (ETSAB). Estos descubrimientos sugieren la necesidad de tomar medidas para mejorar la calidad del aire de las aulas y asegurar el bienestar de los estudiantes.

Keywords

visualization, chart, plot, temperature, CO_2 concentration, percentage of humidity / relative humidity, data, classroom, preprocessing, pattern, insight, exceedance, fall below, comfort, consumption, apparent temperature / heat index.

Contents

1. Introduction and state of the art

This bachelor's thesis is carried out at the Universitat Politècnica de Catalunya under the supervision of Pere-Pau Vázquez. The purpose of this project is to develop a visual analytics application that enables the comparison and analysis of air quality in some of the UPC classrooms, along with a brief evaluation of energy consumption.

This section aims to provide the background of the thesis, informing about the current analytics tool being used and highlighting the need for a specialized tool to analyze air quality. The subsequent sections will define the specific goals of the project and expose the problem, describing the expected scope of this research. Following this, the data and technologies used will be presented and explained. The process of developing the application, with all its visualizations, will be detailed. Finally, the thesis will conclude with a summary of the findings, a discussion of the ethical implications, and suggestions for future work. The appendix will contain all the visualizations.

1.1 Project Background

The Universitat Politècnica de Catalunya (UPC) has shown a growing interest in air quality, which is a crucial aspect for creating a favorable learning environment for students. Several studies have demonstrated the adverse effects of poor air quality on health and academic performance. Additionally, energy efficiency is crucial for sustainability in educational institutions and this topic has gained attention lately.

1.1.1 Significance of the project

Several projects at UPC reflect this interest on air quality and energy efficiency. UPC has a dedicated team of experts called *Servei d'Infraestructures* working everyday to take care of all the installations and develop plans to evolve into a sustainable university. One of their notable initiatives is the *QAire* project, which is somehow the basis of this thesis. It involved placing sensors in several indoor spaces at UPC to monitor air quality. Moreover, architecture professionals from ETSAB have also expressed interest on having a tool to analyze the air quality of their classrooms, relating it to architectural factors.

The data collected by the *QAire* project, along with energy consumption data, is displayed in an application called SIRENA, which stands for *Sistema d'Informació dels Recursos ENergètics i l'Aigua*. This application will be discussed in detail in Section 3, as it serves as data source for this project. SIRENA allows users to filter data by campus, building, classroom and date; it also provides the option to download the selected data in a CSV. See Figure 1 for an example screen of the app.

SIRENA is a powerful tool for an overview of the data, also enabling to choose the level of detail. However, its function is to display data rather than provide comparisons and facilitate its deep analysis. This limitation inspires the creation of a more specific tools that allows users to gain deeper insights by visualizing data in different ways. So, the point is to fully exploit data, to not only view it but also extract detailed insights, identify patterns and evaluate quantitative metrics of classroom air quality.

Thus, the necessity for this project emerges from the possibilities opened by SIRENA, as well as the interest of some experts in this area. This thesis aims to develop such a specialized tool to meet these needs and take conclusions that facilitate professionals the decision-making process on the improvement of UPC indoor spaces and energy consumption.

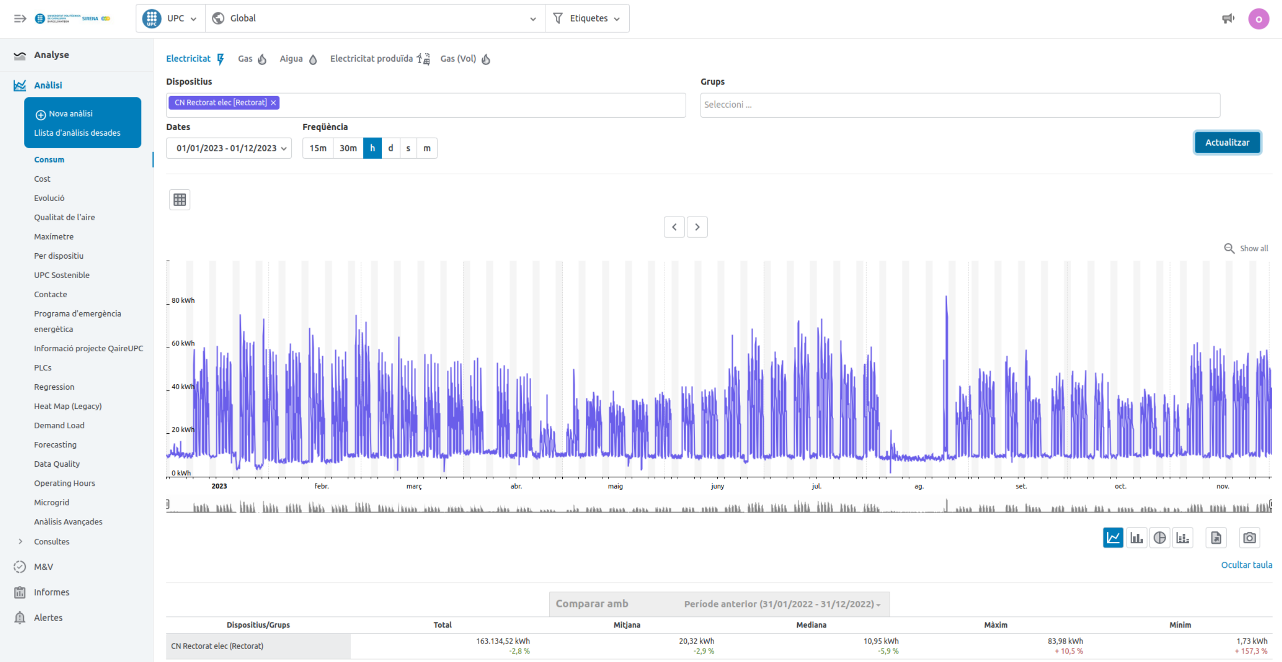

Figure 1: Screen of SIRENA showing Energy Consumption.

1.1.2 Basic concepts about Air Quality

Air quality can be measured according to three aspects: temperature, CO_2 concentration and relative humidity. All these variables have comfort values recommended for a working environment, which are shown in Table 1. The comfort zone has been deduced from different sources of information, as there are variable criteria. These values are crucial for the analysis of air quality, since the main point is to have a background to compare the actual values with the ideal ones. This way, air quality can be assessed.

Variable	Min Comfort	Max Comfort
Temperature	20°C	25°C
CO_2 **Concentration**	-	1000 ppm
Humidity	30%	60%

Table 1: Comfort Levels for Air Quality Variables

But, what does it happen when classrooms exceed these comfort thresholds? Air quality is not just about comfort, it's also about health. Firstly, poor air quality can have a significant impact on the health and well-being of teachers and students. Exposure to indoor air pollutants can cause a range of health problems, including headaches, eye irritation, respiratory problems, and allergic reactions. Long-term exposure to indoor air pollution has also been linked to more serious health issues such as lung cancer and heart disease.

Beyond air pollution, let's focus on the impact of the measured variables. Several investigations have shown that human beings can experiment headaches, sleepiness, poor concentration, loss of attention and increased heart rate when they are exposed to excessive CO_2 concentration. Regarding temperature, it clearly impacts the mood and productivity of people; for instance, warm environments difficult concentration and can also be a stress factor. Elevated humidity levels can lead to increased mould (fungus) growth, which can cause stuffiness and eye or skin irritation. Similarly, humid conditions are ideal for dust mites, which can trigger allergies and asthma in some individuals. Dry air is less harmful than humid, but it can cause employees to experience dry eyes or throats, and headaches. Furthermore, high humidity combined with high temperatures cause even higher apparent temperatures, making people feel hotter.

2. Project Goals

The initial goal of this project is to create a visualization tool capable of analyzing the air quality in some of the UPC classrooms, while also relating it to energy consumption. After consulting with several experts, the project was specifically defined to analyze these variables in the ETSAB classrooms.

Since there were no predefined requirements for the development of this project, the analysis subjects were open to numerous possibilities. That's why a list of interesting questions to answer was brainstormed after examining the available data. The chosen ones were the following:

1. Is there any relationship between the temperature of the campus and inside the classrooms? And with the energy consumption?
2. What are the patterns in energy consumption at ETSAB? Is it possible to identify any seasonal trends?
3. Is air quality of the classrooms appropriate for a working and healthy environment, in terms of temperature, concentration of CO_2, percentage of humidity and temperature? Which classrooms are the most problematic?
4. Which are the peak hours of CO_2 concentration in classrooms?
5. How is the impact of humidity on temperature? Does apparent temperature reach dangerous values?
6. How does air quality vary across the different floors of the building? Do windows affect the CO_2 concentration?

Therefore, the main objective of the thesis is to create graphics that provide answers to the previously mentioned questions. This will assist Sustainability and Architecture teams in reaching conclusions and implementing actions to improve air quality and reconsider energy consumption.

3. Data

In order to develop the project, a crucial step was to obtain and clean the available data. It's so important to have proper data to be able to answer the questions stated in Section 2 and get insights from the visualization. This section explains the data source, its content, and the preprocessing methods used.

3.1 Data Source

The primary data source was UPC's SIRENA tool. This tool was developed to monitor and assess the consumption of resources at UPC, including electricity, gas, water, photovoltaic production, and indoor air quality. The management of this tool is handled by UPC's sustainable community. No additional data sources were required beyond SIRENA.

Since SIRENA is a visualization tool, it provides a comprehensive overview of the data. Therefore, after taking a quick look at the available data displayed in SIRENA, all possible variables related to energy consumption and air quality were downloaded. To obtain the appropriate datasets, data was filtered specifically for ETSAB and from the year 2023. A sample frequency of one hour was used; although a 15-minute frequency was an option, it would have resulted in datasets that were too large and specific for the visualizations, which aim to analyze the entire year and identify global trends.

3.2 Data Collection

The datasets downloaded from SIRENA comprise four files: one for energy consumption (electricity) and three for air quality (temperature, humidity percentage, and CO_2 concentration). Data for gas and water consumption would have been useful, but SIRENA data for these variables is empty.

On the one hand, the energy dataset contains information about the energy meters placed in buildings A, B, and C, as well as the total consumption of ETSAB (all three buildings combined). These values are associated with their date and hour, and, as mentioned, were recorded every hour during 2023, resulting in 8760 samples for each variable.

On the other hand, the air quality data consists of three datasets, one for each variable mentioned before: temperature, humidity percentage, and CO_2 concentration; which were obtained through the sensors placed on the classrooms. For this project, data from classrooms in Building A (Segarra) was selected for analysis due to the building's size and number of floors, which can lead to more interesting conclusions. Another reason for this decision is that other buildings lack sensors in all classrooms, complicating their analysis. Thus, the three air quality datasets contain information for each of the 26 classrooms in building A, recorded hourly throughout 2023, also resulting in 8760 samples per dataset. The temperature dataset includes an additional variable for the outdoor temperature of Campus Sud, which is also of interest.

Before feeding the visualizations with the downloaded data, it is necessary to ensure the data is proper and clean.

3.3 Data Preprocessing

The first step in data preprocessing was to check the variables and understand their meaning to determine their usefulness or whether they should be deleted or renamed. For example, the air quality datasets initially included more classrooms than necessary, which were later filtered to keep only those of interest. Also, some variables had too complex names which were also changed to a more user-friendly ones.

Next, the existence of missing values was checked to ensure the data was clean. The data extracted from SIRENA had missing values that needed to be addressed. The solution applied was dual: if there were few missing values, forward imputation was performed to complete them, meaning that it filled NA/NaN values by propagating the last valid observation to next valid. If there were too many missing values, data quality was compromised, so these variables were deleted unless absolutely necessary.

Finally, once all datasets were clean, they were ready to be used in visualizations. It is worth mentioning that data needs to be transformed differently depending on each plot's needs, as the priority is to have suitable data for each visualization. So, further transformations and joins were performed but not detailed here. Additionally, data was grouped by different aggregation levels, and some necessary columns were created. Relevant changes will be explained in Section 5.1, with the development and design of the graphics.

3.3.1 Energy Data Preprocessing

In the energy data, all variables had 53 missing values (except building C, which had 54), which is not significant given the 8,760 samples. But, before completing the missing values through imputation, a quick exploration was performed, revealing an anomaly. From December 20 (2023) at 12 AM to December 22 (2023) at 8 AM, the values for all variables were missing. Furthermore, the first value after the gap (December 22 at 9 AM) was extremely high (about 20 times the usual values), and the last value before the gap was abnormally low. These two outliers would have distorted the imputation and caused a nonsensical peak in the graphics. Therefore, they were treated like the other missing values and imputed with the nearest previous value that wasn't missing.

3.3.2 Air Quality Data Preprocessing

Regarding air quality data, all three datasets required a similar preprocessing, as they had the same characteristics. Most of the building A classrooms presented a few missing values (maximum 9) so they were filled in the three datasets using the method explained before. However, classroom A-33 had two columns in each dataset to indicate the value of the variable. One of the duplicated columns was called *"Old"* and was completely empty (8760 missing values), so it was directly deleted in the three datasets, as it was not useful.

4. Technologies

In this section, the different used technologies will be briefly explained. It is worth mentioning that I already had some experience working with Pyhton, and Altair; and also some previous knowledge about Streamlit.

The technologies used in this Bachelor's Thesis have been divided in two main sections:

- Data: All the software that has been used to collect or generate the data for the visual analysis tool.
- Web Application: All the software that has been used to design the app, including its visualizations.

Figure 2: Logos of some of the used frameworks.

4.1 Data Technologies

First of all, the entire project was programmed in Python. For data preprocessing, different Jupyter notebooks were used and launched via Anaconda. The main package used for this step was Pandas, as it provides powerful data structures and data analysis tools, making it ideal for preprocessing large datasets. Pandas offers a wide range of functions for data manipulation, such as merging, reshaping, and filtering datasets, which are essential for cleaning and preparing data. Also, I was already familiar with it, which facilitates the whole task. Numpy was also used occasionally to complement Pandas.

4.2 Web Application Technologies

To build the application, I mostly chose libraries which I was already familiar with. To develop the application itself I used Streamlit and for the graphics creation I mainly used Altair.

4.2.1 Streamlit

Streamlit is an open-source Python framework for data scientists and AI/ML engineers designed for creating interactive and visually appealing web applications. It allows developers to quickly build and deploy data-driven web apps. Streamlit offers a variety of widgets for graphic interaction, enabling applications to update in real-time as users interact with the widgets.

One of the key advantages of using Streamlit to design the app is its ability to create dynamic interfaces due to its interactivity. Additionally, it supports various Python libraries (such as Numpy and Plotly) which is crucial for having access to the best available libraries when creating charts to find the most suitable one for each visualization. Another significant benefit of Streamlit is its caching mechanism and modularity, which enhances performance by preventing redundant computations and handling separately different components of the app.

4.2.2 Altair

Altair is a declarative statistical visualization library for Python, known for its effortless integration with Streamlit, making it highly readable and offering a multitude of functionalities for diverse chart types. Before this project, I had already acquired familiarity with Altair, so I knew how to use its interactive features and some of the multitude of functionalities.

Built on the Vega and Vega-Lite grammars, Altair simplifies the creation of high-quality data visualizations by enabling users to specify visual properties using an intuitive syntax. It integrates with Pandas DataFrames, facilitating data manipulation and analysis. Altair has interactive elements such as tooltips and filters, making it suitable for exploratory data analysis and research. Additionally, it supports statistical transformations and generates publishable outputs, enhancing its utility for various analytical tasks.

4.2.3 Plotly

Plotly was primarily utilized in cases where Altair fell short of meeting the requirements for certain charts, either due to limitations in functionality or specific chart needs that Altair couldn't address adequately. One of the notable advantages of Plotly charts is their user-friendly zoom in-out interactivity, as well as their visually appealing design.

Plotly stands out as a comprehensive, open-source visualization library specially designed for Python users, offering a multitude of chart types, including line plots, scatter plots, bar charts, histograms, maps, and even 3D plots. Its easy integration with popular data science libraries such as Pandas and NumPy makes it a preferred choice for data analysts and researchers. The interactive features of Plotly, such as hover information and zooming capabilities, greatly enhance data exploration and analysis.

5. Application Development

The main task of this project was the design and development of the app. The main objective of the application is to effectively transmit the answers to the questions stated in Section 2. Achieving this objective involves two crucial components: creating comprehensive visualizations and organizing them within the app, along with providing additional explanatory text to explain the results.

5.1 Visualization Design

The most important aspect of the project is the design of the visualizations, as they are crucial for effectively communicating the outcomes of the project. This section will provide a detailed explanation of the process carried out to develop each visualization, all of them were created using Altair except another library is specified. The goal is to identify and implement the most effective visual representations that accurately and clearly express the findings. However, not all graphics will be displayed in this section, for a complete set of visualizations refer to Appendix A.

5.1.1 Energy Consumption Visualizations

To begin, I focused on the energy consumption visualizations, more specifically on the first part of the question 2 stated in Section 2. The goal was to create an initial global view of all the available 2023 consumption data. Therefore, the first graphic is a line plot that shows the consumption of each ETSAB building as well as the total consumption of ETSAB, attached in Figure 3. To make the chart more comprehensive, user interaction was incorporated by adding custom date filtering for the start and end of a period. This feature allows users to select the period they wish to view using a pop-up calendar, enabling them to see more detailed data.

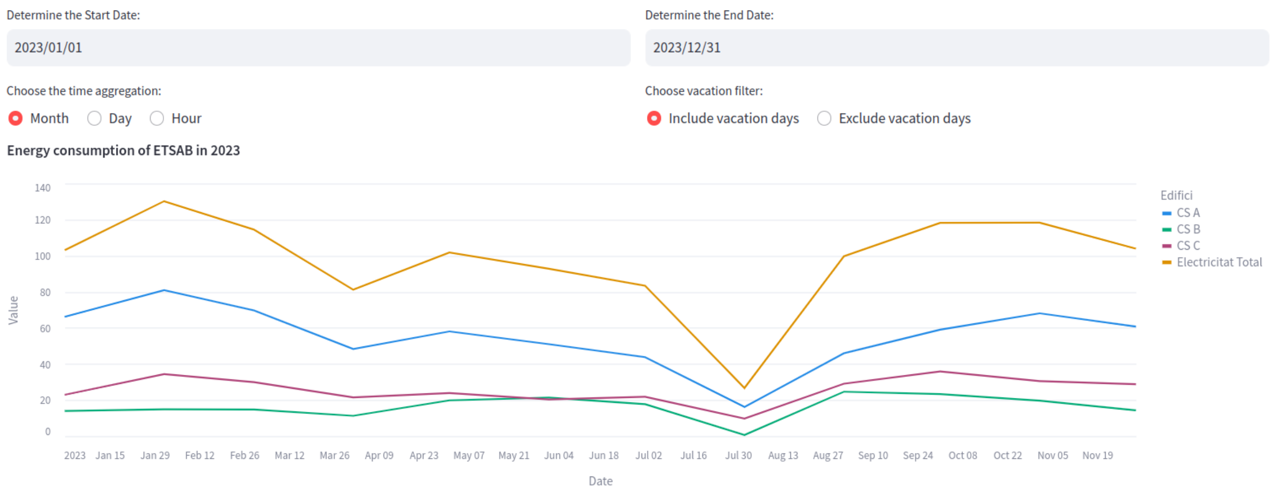

Figure 3: Energy Consumption Line Plot with its Interactive Options.

Additionally, users can choose the time aggregation of the data to observe either a more global trend or detailed consumption. The options are to view data on a monthly, daily or hourly basis. The aggregations are computed transparently for the user by automatically grouping the data according to the chosen aggregation and computing the mean. Finally, an option was included to exclude vacation days (including weekends), which significantly distort the consumption data, allowing users to observe global consumption trends without abnormal values. To get the vacation dates of the course, the academic calendars of courses 2022-2023 and 2023-2024 were checked. The exact dates are added in the Appendix B.

This visualization is accompanied by a map of ETSAB buildings to provide users with context regarding their locations. Additionally, a small pie chart represents the percentage contribution of each building to energy consumption with respect to the total. These two visualizations are shown in Figure 19 of Appendix A.

The next visualization also aims to address the question 2 raised in Section 2. To examine seasonal consumption patterns and variations, the idea involves comparing the hourly average consumption for each season using a line plot, facilitating the observation of daily consumption trends. To enhance the visualization's informativeness, this comparison is extended to include individual labor days, consequently encapsulating both seasonal differences and variations between labor days all in one plot.

Initially, the plot focused only on the energy consumption of Building A. However, to provide a more comprehensive view, an interactive button was added to display the consumption of each of the buildings separately, as well as the total consumption. This option revealed notable differences in consumption patterns among the buildings.

During the analysis, another challenge emerged. As most of the summer period coincides with vacation days when the campus is closed, the average consumption during this period is biased. To adress this issue, all vacation days were treated separately as "another season", and labeled accordingly to exclude them from each seasonal consumption, given that they are outliers. The final plot is shown in Figure 4:

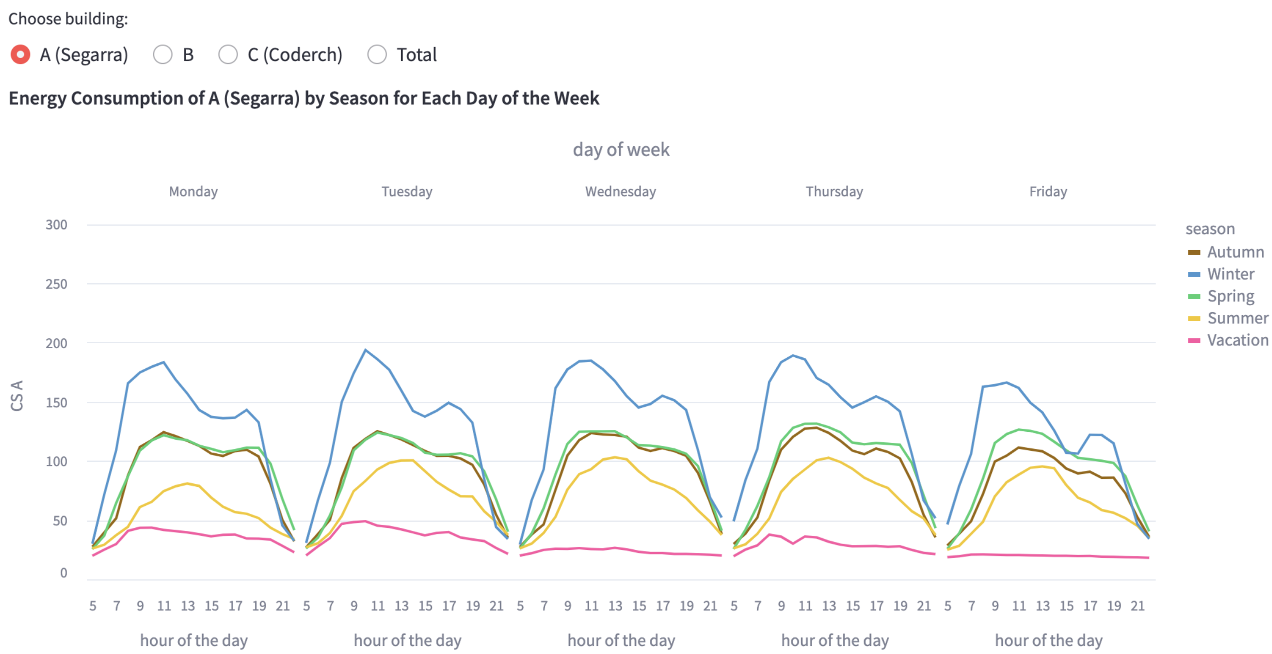

Figure 4: Seasonal Energy Consumption Line Plot.

The final energy consumption plot addresses the question 1. A combination of line and area charts is utilized for this purpose. The objective of the graphic is to illustrate the relationship between indoor and outdoor temperatures, their differences, and energy consumption, all within a single plot. Additionally, the plot shows the evolution of this relationship throughout the year 2023.

As the plot features two distinct variables, it uses two independent y-axes: one on the left for representing temperature and another on the right for energy consumption. Also, it uses different colors to help distinguish the variables. Regarding temperature, the plot displays two lines: the green line represents outdoor campus temperature, while the blue line represents the average temperature in classrooms of Building A, signifying indoor temperature. Furthermore, to highlight the difference between indoor and outdoor temperatures, the area between both lines is marked (with low opacity). This area enables users to quickly associate wider zones with high temperature difference and narrower areas indicating the opposite. The resulting chart is depicted in Figure 5.

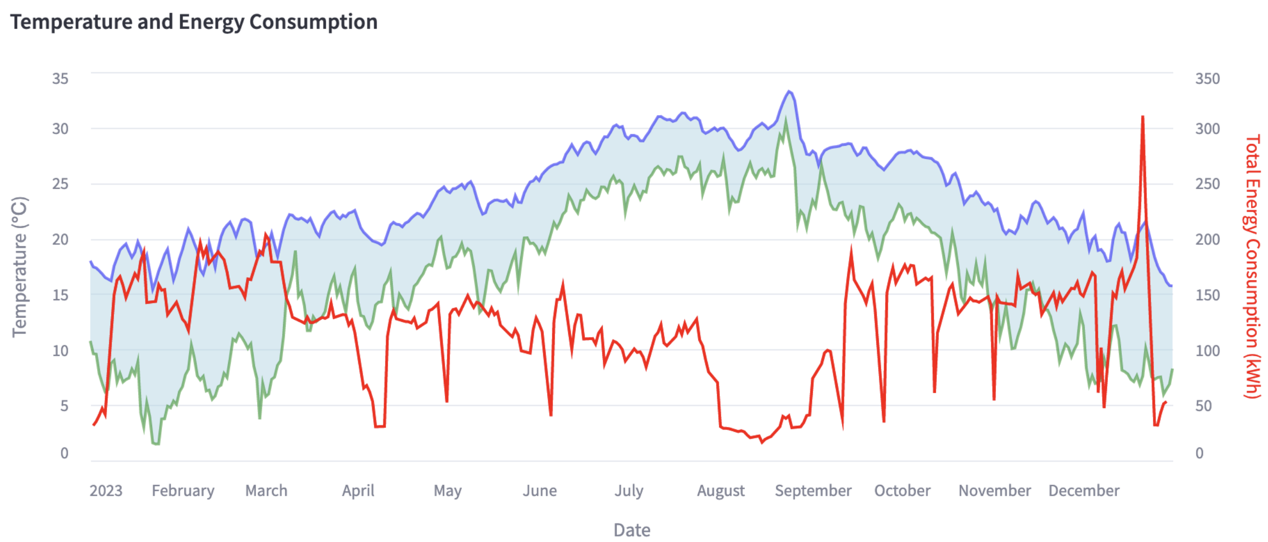

Figure 5: Relationship between Energy Consumption and Indoor-oudoor Temperatures.

5.1.2 Air Quality Visualizations

Analyzing air quality requires focusing on two key aspects: assessing air quality based on the comfort values recommended for a productive working environment, and finding patterns on the building. Therefore, all visualizations in this subsection are focused in these specific aspects. Also, it's essential to note that when evaluating classroom air quality, only class hours on labor days should be considered to prevent contamination from non-representative data.

Given that air quality is measured using three variables (temperature, CO_2 concentration, and percentage of humidity) some visualizations will be replicated for each variable. To avoid repetition, these visualizations are only explained once in this section, but the complete set of visualizations is provided in Appendix A.

Temperature Graphics

The first visualization is intended to provide a comprehensive overview of the temperatures across all classes in Building A throughout 2023. To do so, a scatterplot was used to display all temperatures registered during the year for each class. However, a notable efficiency issue arose due to the large number of data points (8760), which could be solved by subsampling. But addressing this challenge required a careful approach, given the fact that it's very important to show outlier temperatures in this chart, and not remove them with the subsampling.

To mitigate the efficiency problem while preserving the data variation, a specific subsampling technique was employed. Initially, the mean temperature for each class at every time instant was computed. Then, for each time instant, the distance of each class's temperature from the mean was calculated and the points were sorted by this distance. After the sorting, the subset was composed by only two points near the mean and 20% of the most distant points (in order). This subsampling strategy ensures that the final plot is informative while improving performance. Furthermore, to enrich the visualization, comfort temperature zone for a working environment is highlighted in blue, and the points are colored to indicate whether they fell within the comfort range, exceed it, or fall below it. The resulting plot is shown in Figure 6:

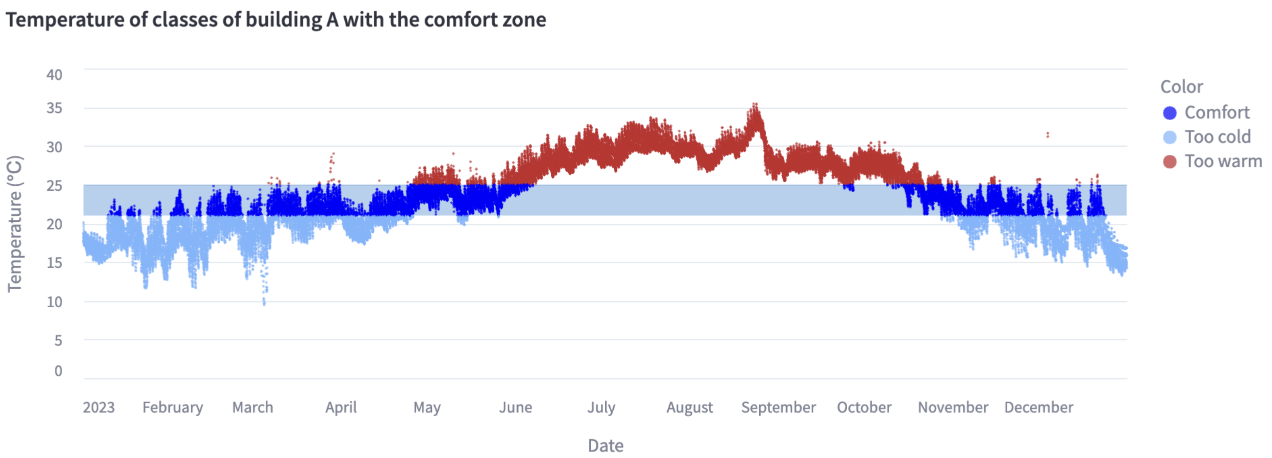

Figure 6: Classrooms Temperature Scatterplot with Comfort.

Then, the question 6 was investigated through the creation of three more graphics. The first one is a line plot that shows the temporal evolution of the average temperature across each floor of Building A through 2023. To enhance user interaction and achieve varying levels of detail, the visualization allows users to choose the time aggregation of the data and select a specific period. Given that all floors have similar temperatures, some lines overlap so an additional interactive feature was integrated to the plot. This feature enables users to select a subset of floors to highlight for more effective comparisons. This plot is displayed in the Appendix A, in Figure 20.

The second visualization in the floor-wise temperature analysis consists of a violin plot, which provides insights not only into the average temperatures but also the distribution of temperatures across different floors. In this plot, wider sections indicate higher density of data points, while narrower sections indicate lower density. It's also complemented by a box plot inside the violin areas, which illustrates the interquartile range and the median temperature of each floor. This visualization was generated through the Plotly library (for this this case, better than Altair), chosen for its simplicity and ability to create a visually appealing graphic. Furthermore, Plotly includes interactive toolboxes that display detailed values when passing the mouse over the data. See the plot in Figure 7.

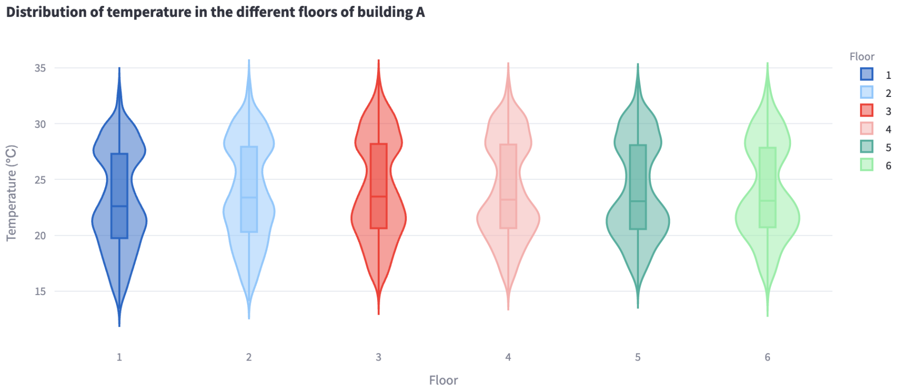

Figure 7: Floors Temperature Violin Plot.

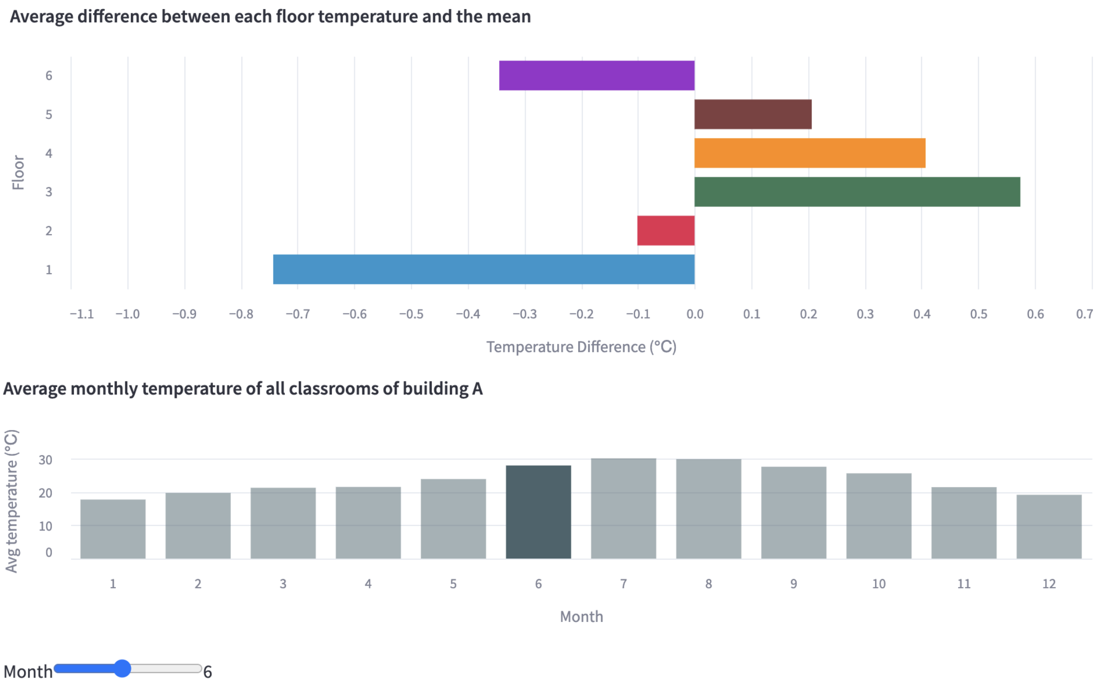

Figure 8: Average Difference between Floors Temperatures.

The plot shown in Figure 8 is the third one in floor-wise temperature analysis and it focuses on the evolution of difference between each floor's mean temperature and the overall building mean for each month. It allows for a more detailed comparison of temperature variations across floors than the first plot in this analysis. This visualization comprises two bar plots: the first and most significant displays the difference between the average monthly temperature per floor and building's average monthly temperature. This plot highlights whether a floor consistently remains colder or hotter than the average, and identifies the floors with the most deviation. A slider enables users to change the month being viewed. The second bar plot presents the monthly average temperature of the entire building, highlighting the month being currently viewed.

Last but not least, a visualization was designed to display the quality of each classroom in terms of temperature. To determine a quality metric from temperature data, the comfort zone was used. The number of samples exceeding the comfort temperature and those falling below it were counted. Following this, the percentage was computed, resulting in a metric that indicates the percentage of class hours each classroom spends above or below the comfort zone. To ensure a more accurate metric, only labor days and a time span from 8 AM to 9 PM were considered. Additionally, this metric was computed per month to provide more detailed insights, so the chart has the interactive option for choosing between global or monthly metrics. For the monthly metric, all samples within the month were included (even during vacation periods), and a slider appears to the screen to allow the user changing the month that's being visualized.

The final plot is a heatmap with the shape of the building, coloring each classroom based on the percentage of hours spent outside the comfort zone. To separately show instances of high and low temperatures, two heatmaps were created: one using a red scale for temperatures exceeding the comfort zone and another using a blue scale for temperatures below the comfort zone. A challenge arose with this visualization because the color scales needed to be consistent across months, but their values were really variable, making it difficult to distinguish differences between classes. The solution was to use a uniform scaleof 0% to 100% for the monthly plots to maintain consistency, but make the annual heatmap with its own scale to better appreciate class differences. The resulting charts are shown in Figure 9.

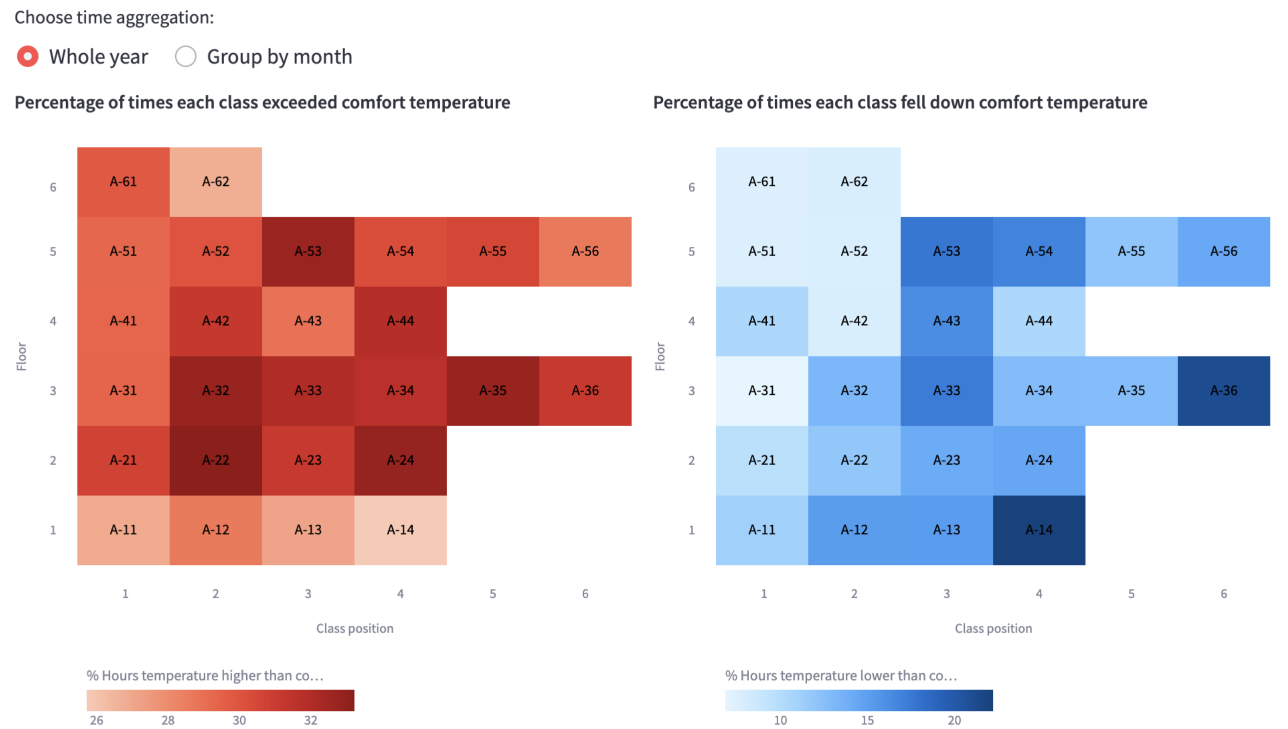

Figure 9: Classrooms Percentage of Hours Outside Comfort Temperatures Heatmap.

CO_2 Concentration Graphics

For CO_2 Concentration, some of the temperature plots were replicated. The first plot is the same as in the temperature tab, utilizing also the same subsampling method. The only difference is it shows CO_2 Concentration data and comfort values. The second visualization corresponds to the violin plot for floor-wise analysis already explained in the temperature section. Both visualizations are displayed in Appendix A, in Figure 21 and Figure 22.

The next chart is also only provided for the CO_2 Concentration section. It explores the relationship between this variable and the number of windows in each classroom. To count the number of windows of each classroom, the maps provided by architecture professors were used. In Table 2 the information about the number of windows of each classroom is given.

Number of Windows	Classrooms
4	A-31, A-32, A-33, A-34, A-35, A-36, A-41, A-51, A-52, A-53, A-54, A-55, A-56
5	A-61
6	A-11, A-12, A-13, A-14, A-21, A-22, A-23, A-24, A-43, A-62
7	A-42

Table 2: Number of Windows of each Classroom of Building Segarra

This analysis is performed with a box plot that illustrates the distribution of average CO_2 Concentration for classrooms grouped by the number of windows they have. The chart was created using Plotly, which offers visually appealing and easily generated box plots compared to Altair. The resulting plot is displayed in Appendix A, in Figure 23.

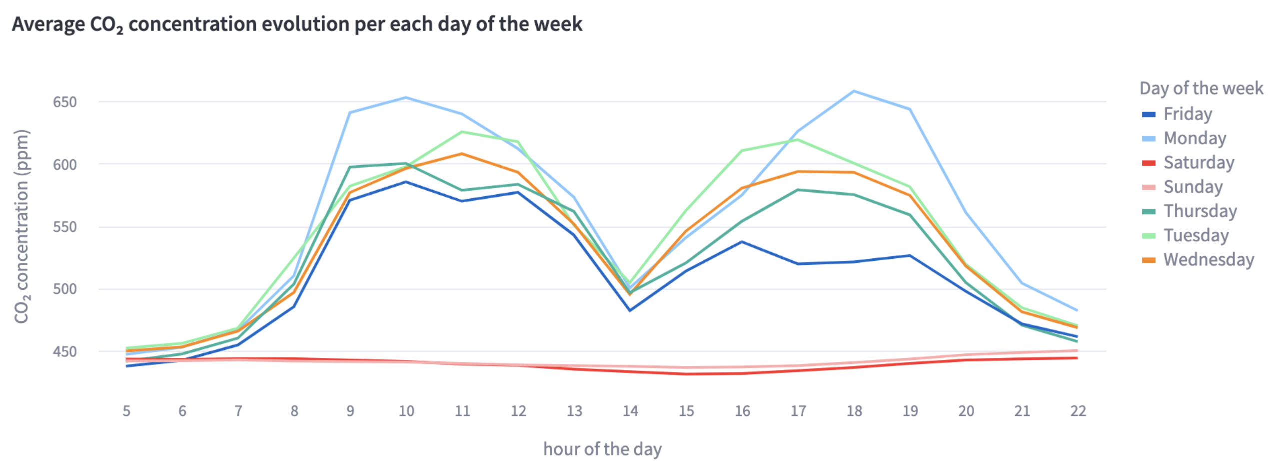

Figure 10: Average CO_2 Concentration Evolution for each Day.

The visualization shown in Figure 10 is specifically designed for the CO_2 Concentration data, illustrating its evolution throughout the day (from 5 AM to 10 PM) for each day of the week, including weekends. To ensure accurate representation, vacation days were excluded from this dataset to avoid distorting the normal values of CO_2 Concentration. This line plot shows various colored lines, each representing a different day of the week. The data was aggregated by computing the hourly average for each day of the week, providing a clear representation of the average CO_2 Concentration trend for each day.

The last plot of CO_2 Concentration analysis is the same as in the Temperature section: a heatmap indicating the percentage hours each classroom exceeds the comfort CO_2 Concentration threshold. The only difference is that only one heatmap is necessary, since the only case of analysis is when CO_2 Concentration is exceeded. The resulting plot is displayed in Appendix A, in Figure 24.

Percentage of Humidity Graphics

The first two humidity visualizations are repeated from the two previous variables. First, there is a subsampled scatterplot with the comfort zone specified, and secondly, a violin plot is shown for the floor-wise analysis. Both plots are displayed in Appendix B, in Figure 25 and Figure 26.

Then, a new section is created to show the impact of high humidity on high temperatures, called heat index (or apparent temperature). In this section there are two interesting visualizations. The first one's objective is to transmit the number of times that both temperature and humidity are high during labor days and class hours (from 8 AM to 9 PM). For this chart, a percentage of humidity higher than 40% and a temperature higher than 25°C are considered high. To show the frequency of this event, a bubble plot is used, which is a scatterplot that encodes the counts (occurrences) as the size of each point. So, the resulting chart consists on points of varying sizes being placed over the two axis, where the classes are represented on axis X and the months on the vertical axis. The resulting bubble plot is shown in figure 11.

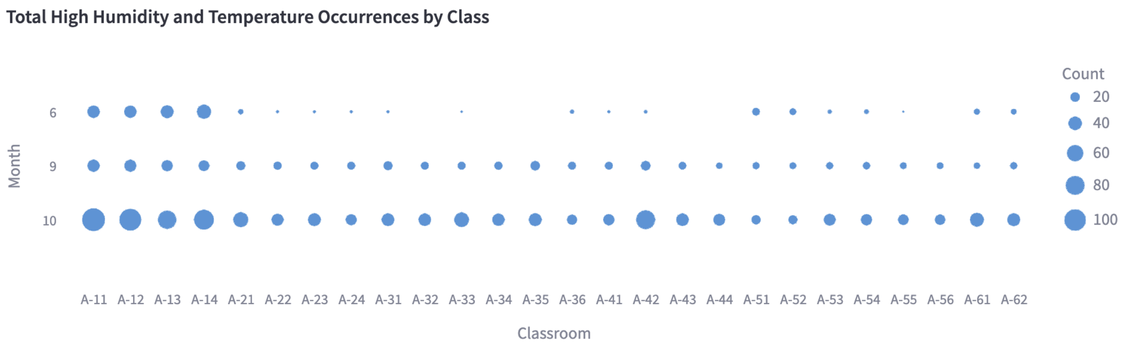

Figure 11: High Temperature and Humidity Occurrence Bubble Plot.

The second plot concerning apparent temperature aims to highlight the frequency and specific dates when certain classrooms reached caution-level sensations based on humidity and temperature. To achieve this, the heat index (apparent temperature) was calculated and categorized according to its corresponding sensation before plotting. It's worth noting that heat index only makes sense for temperatures above 26°and humidity percentages higher than 40%. The formula used to compute the heat index with these constraints is the following, with T for the temperature and H for the percentage of humidity:

$$HI = -8.78469476 + 1.61139411 \cdot T + 2.338548839 \cdot H - 0.14611605 \cdot T \cdot H - 0.012308094 \cdot T^2 - 0.016424828 \cdot H^2 + 0.002211732 \cdot T^2 \cdot H + 0.00072546 \cdot T \cdot H^2 - 0.000003582 \cdot T^2 \cdot H^2$$

Heat index can be classified into the categories of Table 3 according to its sensation.

Category	Temperature Range	Description
Caution	27 to 32°C	Possible fatigue from prolonged exposure or physical activity.
Extreme caution	33 to 40°C	Heat exhaustion, heat stroke, cramps. Likely with prolonged exposure or physical activity.
Danger	41 to 53°C	Heat stroke, heat exhaustion, cramps. Very likely with prolonged exposure or physical activity.
Extreme danger	54°C or more	Heat stroke, imminent heat exhaustion.

Table 3: Apparent Temperature Categories

A colored scatterplot was used to effectively transmit this information, displaying the apparent temperature for all class hours on labor days in 2023. Since the heat index can only be computed for temperatures above 26°C and humidity levels above 40%, the scatterplot exclusively includes data points meeting these conditions. An interactive tooltip was added to provide detailed information on humidity, temperature, and the classroom when passing the mouse over each point. The resulting plot is shown in Figure 12.

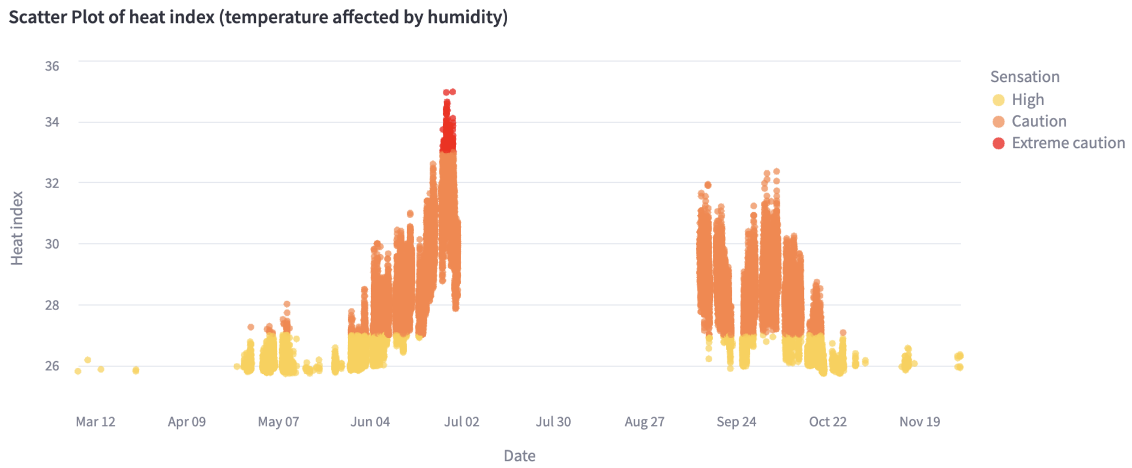

Figure 12: Scatterplot of Heat Index Sensation during Class Hours.

Finally, the last humidity chart is the same as the heatmap explained for temperature and CO_2 Concentration sections. It is shown in the Appendix A, in Figure 27.

5.1.3 Detailed Classroom Comparison Visualizations

This section serves as a comprehensive tool for comparing air quality across classrooms. While previous charts focused on analyzing individual air quality variables, they didn't facilitate direct comparisons between specific classes. Thus, the objective of this section is to provide a compact visualization capable of comparing selected classrooms across the three air quality variables: temperature, CO_2 Concentration and humidity.

The main visualization is a scatterplot that displays all the data points of the selected classrooms for the entire year of 2023, overlaid with the comfort zone specific to the chosen variable to display. Given the complexity of this visualization, it needs several interactive tools. Firstly, there's a Streamlit multiselector designed to choose classrooms, cleverly shaped to match the building structure, which is limited to the selection of five simultaneous classes to properly view the data (a message pops up in case it's not respected). Additionally, users can select which air quality variable they want to examine. Moreover, they have the option to view data either for the entire year or month-by-month (in this case, a slider is also provided), letting the user choose the level of detail of data.

A part from the interactive scatterplot, there are pie charts positioned below, each corresponding to a selected class. These pie charts display the percentage of data points falling inside, above, and below the comfort zone, with the class name as the title and the percentages indicated within the chart. All these pie charts are synchronized with the user's selections explained above. A small taste of this complex visualization is shown in Figure 13.

Figure 13: Screen Example of Detailed Classrooms Air Quality Visualization.

5.1.4 Air Quality Metrics Summary

In this final section, the aim is to sum up the information displayed on the previously presented heatmaps and delve deeper into the air quality evaluation of each class. The initial group of heatmaps is really similar to the ones from the other sections, with one heatmap per variable. The only difference is that now the focus shifts to highlighting the instances where data points fall outside the comfort zone, without differentiation between exceeding or falling below it. So, these three heatmaps depict the percentage of data points (hours) spent outside the comfort zone for each variable during labor days, for every class. A new feature is introduced, allowing users to select between global data, Quatrimester 1, or Quatrimester 2. This addition was suggested by architecture professionals who will utilize the app, as they expressed a need to differentiate between quarters due to variations in class schedules. The scale for each variable is uniform among global and quarter data, in order to facilitate comparisons between them. The chart is displayed in Appendix A, in Figure 28.

While this plot offers valuable insights, it also reveals an unresolved aspect: the importance of considering the magnitude of the variance among data points that fall outside the comfort zones. Hence, the second group of heatmaps is created, presenting a distinct and complementary metric. These graphics illustrate the normalized deviation from the comfort zone for each class across the three variables, displayed in three concatenated heatmaps. To compute this metric, the variance was calculated by measuring the distance between data points and the comfort zone of points that deviated from it. These distances were then averaged for each classroom and normalized by the comfort zone range of each variable. This process was repeated for global data, Quatrimester 1, and Quatrimester 2. The resulting heatmap is depicted in Figure 14.

Figure 14: Magnitude of each Class' Deviation from Comfort.

The usefulness of having both kinds of quality metrics is that it classes can be evaluated for both the number of occurrences being outside the comfort zone and the amount of deviation from the comfort zone. It helps detect possible classes that deviate a few times from the comfort zone but, when they do, they have really bad conditions. At the same time, it can help detect classes that deviate a lot of times form the comfort zone but the deviation is minimal, so they aren't a big problem.

Furthermore, as said before, it is required to normalise the sum of deviation magnitudes by the range of comfort zone of each variable, because the importance of the deviation is relative to the range. For instance, CO_2 Concentration has a healthy range from 0 to 1000 ppm, humidity from 30% to 60% and temperature from 20°C to 25°C. The difference between the ranges create the need of normalizing deviations in order to be able to compare this quality metric among the three air quality variables.

5.2 App Integration and Organization

Once all visualizations are designed, integrating them into the Streamlit framework and organizing them intuitively for users becomes essential. Also, given that some visualizations require a considerable execution time, it's necessary to plan their organization carefully. One effective strategy is to distribute charts across different tabs and pages, to avoid having large sections with a lot of graphics loading simultaneously. Despite this, it is worth mentioning that the goal of this project is not to focus on performance but to find a way to appropriately communicate information through the visualizations.

5.2.1 Performance Optimization

Regarding performance, it's necessary to gather information about the usage of Streamlit's tools that allow optimization. The first useful tool is @st.experimental_fragment, a decorator that turns any function into a "fragment" that can run independently of the wider page. When a user interacts with an input widget created by a fragment, Streamlit only reruns the specific fragment instead of the full script. This feature is particularly beneficial in scenarios with multiple slow-loading charts, each with its own controls, as happens in this application. It ensures that user interactions affect only the relevant chart, optimizing performance by minimizing unnecessary reloads of the entire page.

Another Streamlit's optimization tool is @st.cache_data, which is a decorator designed to cache functions that return data. However, the main priority for optimizing this project's performance lies in precomputing all necessary data transformations in preprocessing scripts and storing the result datasets in CSV files. This approach eliminates the need for the application to compute operations on the data dynamically. That's why this tool won't have a lot of impact on this app, but it is beneficial for functions that retrieve and reuse data for different charts, improving efficiency.

5.3 App Structure and Interpretive Guidance

The organization of the app is divided into 3 pages: HomePage, Energy Consumption and Air quality. The Air Quality page is divided into 5 tabs: Temperature, CO_2 concentration, Percentage of Humidity, Detailed classroom comparison and Quality metrics summary. The HomePage is designed to provide users with an overview of its content and organization, facilitating easy navigation and access to relevant sections.

An additional tool is designed to display each variable's statistics for various classes in a clear and interactive format. The tool is titled, in the CO_2 case, "Ranking: CO_2 Statistics for Each Class," and it's placed at the bottom of each variable's subsection. This tool uses Streamlit's *st.metric* to display the mean, minimum and maximum values of each variable for each class. It also has the interactive option to sort the classes by the mean variable value (both ascending and descending) or keep it in the building order. As there are 26 different classrooms, statistics are presented for each class using *st.expander*, which creates a compressible section for each class to save space on the app. This option is interesting as it provides a concise and visually appealing summary of each variable's main statistics for each class. Figure 15 shows part of this tool.

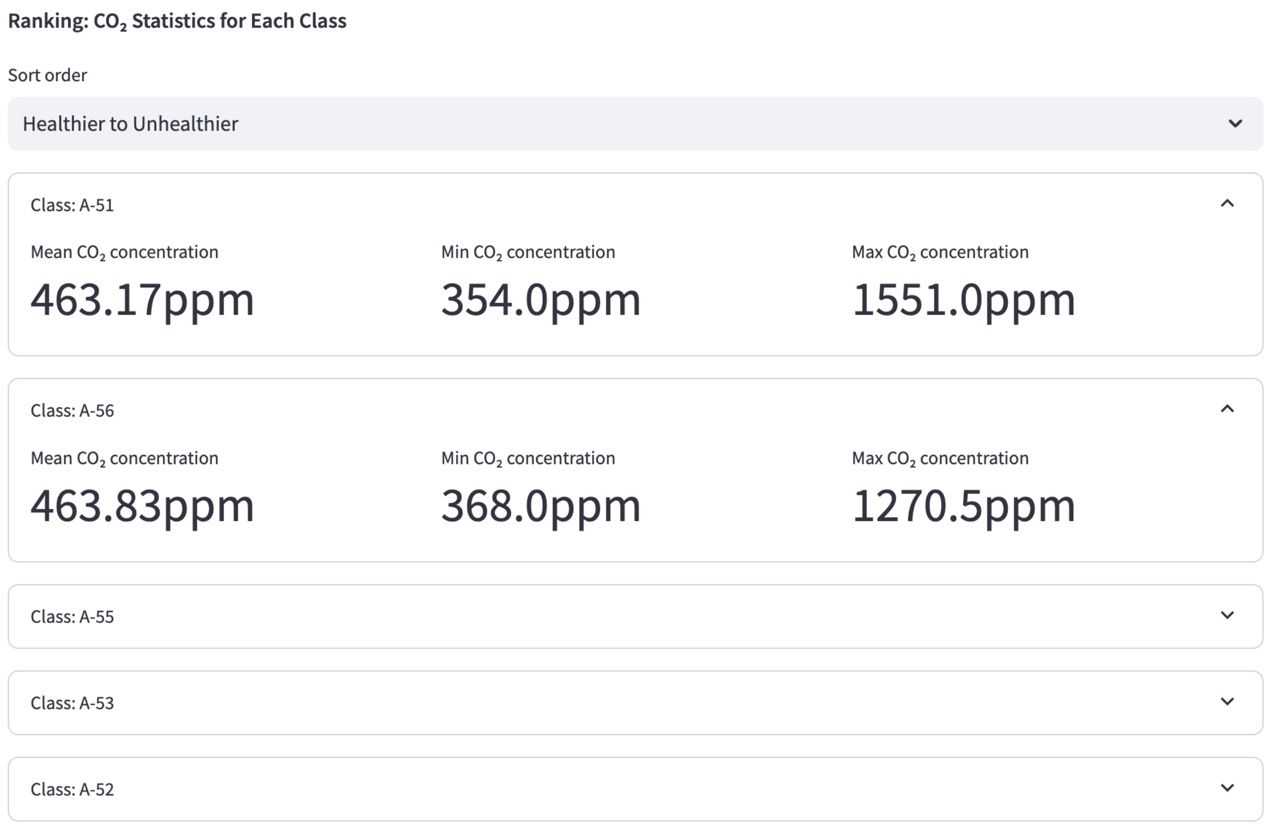

Figure 15: Class Ranking based on Average $CO2$ Concentration.

To enhance the usability of the visualizations, descriptive text was incorporated in the whole application and, more importantly, before each chart, providing informative messages to help users understand the data presented. Additionally, a concise summary of key insights and takeaways was appended after each chart, offering users helpful hints for interpreting the visualizations effectively. An example is shown on Figure 16.

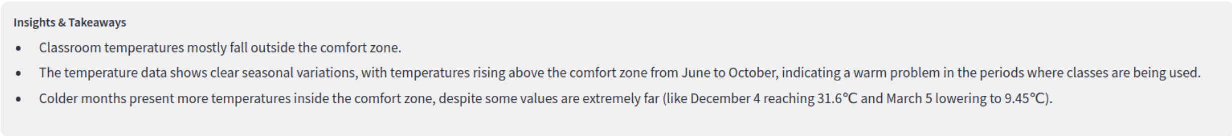

Figure 16: Insights & Takeaways section Example for a Temperature Chart.

Lastly, an "Additional Information" section has been included at the bottom of some sections (when needed), containing pertinent details such as the exact dates of vacation days that were considered for each section, providing users more context. An example of this section for the CO_2 tab is shown in Figure 17.

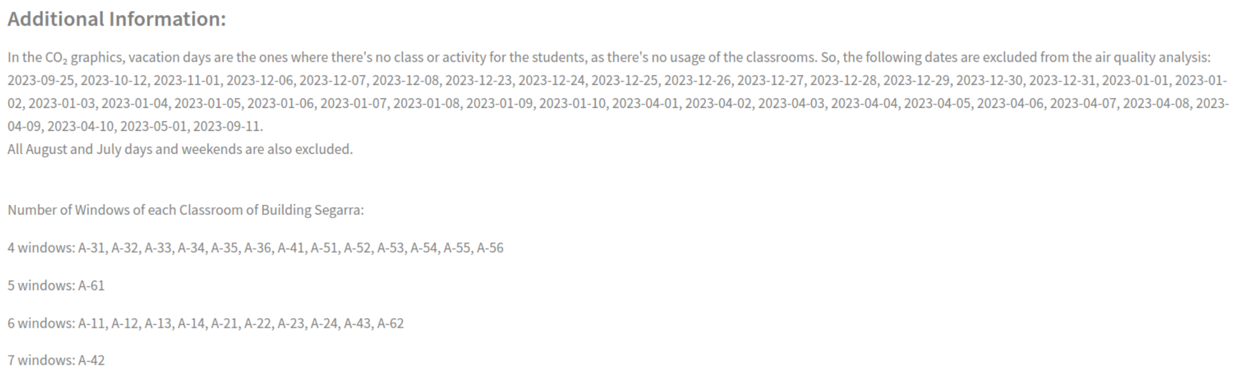

Figure 17: Additional Information section Example

6. Results

In this section, the findings derived from all the previous visualizations will be discussed. The discussion will be guided by the questions stated in section 2 (Project Goals), which should be answered. The aim is to draw meaningful conclusions and insights that highlight the utility of all the created visualizations.

6.1 Energy Consumption Analysis

Regarding energy consumption, the first important aspect to describe the general pattern of 2023 across the three buildings, shown in Figure 3 and Figure 19. The data reveals that Building A consumes 56.88% of the total ETSAB energy consumption, which is surprising considering its electricity usage is only for illumination and not heating. However, this higher consumption is understandable considering that Building A has 7 floors (the 7th floor is not shown in the plots as it doesn't have classrooms), while the other buildings only have a ground floor and a basement.

A clear recurring weekly pattern is evident, with energy consumption peaking during weekdays and dropping over weekends, emphasizing the influence of university activity on energy usage. Most weeks show a noticeable rise in consumption at the beginning of the week, which gradually decreases by Friday. Others weeks, in spite, show two weekly peaks: one at the beginning and another around the middle-end of the week.

Analyzing the evolution of consumption throughout the year, both Buildings A and C present similar energy consumption patterns, while Building B's pattern is more flat. From January to February, there is a notable increase in energy consumption, particularly during the first week of February. As spring arrives (March to May), energy consumption shows a moderate decline, which continues until August, when consumption drops significantly (even more than weekends) due to campus closure. At the beginning of September, consumption starts to increase until September 14th, where a small peak is observed due to the start of classes. Then, the consumption remains constant until the end of the year, where there's a strange peak on December 22nd likely related to the missing values and abnormalities detected around this date, as previously explained in the preprocessing section.

To delve into the details of seasonal energy consumption, we examine the plot of average seasonal energy consumption per building (Figure 4). Buildings A and C display similar daily energy consumption patterns across different seasons and days of the week, although Building A has notably higher consumption. Both buildings show two clear peaks in energy use: one in the morning (around 11 AM) and another in the late afternoon (around 5 PM), reflecting typical university activity hours. However, these peaks are much more pronounced in Building A compared to Building C. Building B, on the other hand, shows the lowest overall consumption and very different patterns for each season, for example, during summer, there's a peak in consumption around 1 PM every day of the week. Remarkably, Building B is the only one where summer exceeds the other seasons consumption. In contrast, Building A experiences the highest consumption during winter (due to the absence of natural illumination), while spring and autumn exhibit almost the same consumption patterns. In Building C, winter and autumn have the highest consumption levels, with summer exhibiting the lowest. Across all buildings, there is a significant reduction in energy consumption during vacation periods, with energy use dropping to minimal levels and remaining practically flat.

The final aspect of energy consumption to analyze is its relationship with the difference between outdoor and indoor temperatures. In the last energy plot (Figure 5), it can be seen that greater temperature differences (wider areas) correlate with higher total ETSAB energy consumption. Furthermore, during spring and summer, as the temperatures raise and their difference narrows, the consumption also decreases. So, this suggests that energy consumption is inversely proportional to indoor-outdoor temperatures difference. However, this relationship may also be influenced by other factors such as increased natural illumination hours or reduced campus activity during the summer months. The problem for this analysis is that only temperatures for classes of building A have been available, while this building's energy consumption is not allocated to heating. That's why it's difficult to directly relate temperature to energy consumption patterns.

6.2 Air Quality Analysis

In this subsection, the results from all the air quality visualizations for each variable will be explained. This detailed analysis will cover each variable individually, providing a comprehensive understanding of the air quality data from Building A of ETSAB.

6.2.1 Temperature Analysis

The primary focus of analyzing air quality data is to ensure that classroom conditions meet the recommended comfort standards for optimal working environments. Examining temperatures, it can be observed in figure 6 that, from November to March, certain classrooms fall below the comfort range, while temperatures rise above comfort levels from June to October. Notably, from June to mid-October, nearly all classrooms always exceed the comfort range, with some reaching temperatures as high as 30°C. Although this is not problematic during summer when classes are not in session, in the other months it is. Cold temperatures show less consecutive deviation over time compared to warm temperatures, having more samples inside the comfort range than warmer months. Some interesting outlier temperatures happen on December 4th, when class A-24 reaches 31 degrees, and a drop to almost 9°C happens in A-22 on March 5th. Additionally, sometimes there are classrooms colder than the comfort at the same time that others are warmer, which also shouldn't happen.

Let's delve into the floor-wise temperature analysis. In all the visualizations provided (Figures 20, 7 and 8), a pattern can be appreciated: the first floor consistently registers lower temperatures compared to the global mean, with an average difference of up to -1°C. On the contrary, the fourth floor consistently exhibits temperatures higher than the mean. Floors 5 and 3 tend to overpass the mean temperature more frequently than falling below it, while floor 2 typically records temperatures below the mean. However, the range of variation in average temperatures across floors is modest (from -1.1 to 0.65°C). The most significant positive and negative deviations occur in December, when the first floor negatively deviates the most and the sixth floor positively does the same.

To conclude, the temperature quality of each classroom will be analyzed. Figure 9 reveals a clear trend: floor 1 is the least likely to exceed the temperature comfort zone. However, it is worth mentioning that all classes exceed the 25°C threshold for 26% to 32% of class hours, and exceedances are spread, with nearly all classrooms affected except for those on floor 1 and A-62, which show fewer exceedance. On the contrary, cold temperatures (below 20°C) are reached in 6% to 22% of class hours, indicating that some classes are less problematic. Classrooms A-14 and A-36 show the most significant issues with temperatures falling below 20°C. Moreover, another pattern is observed: classrooms with higher numbers (located further north) experience cold temperatures more frequently than those numbered 1 and 2.

Monthly data analysis shows that from June to September, all classrooms exceed 25°C almost 100% of the time. The nearest months to this hot period in time are October and May. In October, the percentage of exceedances drops to 60%, which is higher than in May. In contrast, in May this percentage is maximum 37%, this month is the one with the most variability among classrooms in comfort zone exceedance (ranging from 4% to 37% of class hours). Regarding cold temperatures, December, January, and February are the most problematic months. Something curious happens in November, as all classrooms fall below 20°C for a maximum of 20% of the class hours, except for A-14 and A-24, which are down comfort 35% and 43% of the times respectively. This anomaly may indicate a cold problem in these particular classrooms, as they duplicate the percentage of hours down comfort of the other classrooms. It may be relevant the fact that both classes have a wall on in north facade. Finally, in March and April, low-floor classrooms exhibit the highest occurrences of cold temperatures.

6.2.2 CO_2 Concentration Analysis

Regarding CO_2 Concentration relative to the comfort zone, as depicted in Figure 21, student activity is extremely illustrated through CO_2 Concentration levels. During weekends or periods without classes, concentrations consistently remain within the comfort and healthy range. On the contrary, on weekdays, CO_2 Concentration tends to increase, sometimes exceeding the healthy CO_2 Concentration threshold, reaching values as high as 2200 ppm. Notably, during colder months such as November or February, CO_2 Concentration reaches higher peaks than in warmer months like September or May, due to reduced ventilation as people are less inclined to open windows when it's cold.

From the floor-wise analysis, it can be deduced that each floor's usage patterns (such as schedules or number of students) contribute to CO_2 Concentrations. The third floor stands out as having the highest CO_2 Concentration levels, indicating a significant concern. Floors 1 and 2 exhibit similar distributions of CO_2 Concentrations, like floors 5 and 6, which both maintain healthier levels. However, all floors exceed the recommended threshold of 1000 ppm for a healthy working environment.

The previous analysis can be linked to the study involving the number of windows. Figure 23 clearly illustrates that classrooms with four windows tend to exhibit higher CO_2 Concentrations. Surprisingly, classrooms with six windows show higher CO_2 Concentrations than those with five windows. This discrepancy may be attributed to the scheduling of sessions in these classrooms or other variables beyond the scope of the project's visualizations. The previous observation of floor 3 being the unhealthiest correlates directly with the fact that all classrooms on this floor have four windows (classrooms' windows data is shown in Table 2). It is important to note that all classrooms in Building A have windows with a surface area of $3.6m^2$, that's why the only factor needed to take into account is the number of windows. Moreover, as the windows are all equally separated, the number of windows correlates directly with the size of the classroom. This correlation complicates drawing direct conclusions based only on the number of windows.

Let's examine the evolution of CO_2 Concentrations throughout each day. Figure 10 illustrates that on weekdays, there are two peaks in CO_2 Concentration: one in the morning and another in the late afternoon. Mondays and Tuesdays present the highest CO_2 Concentrations, while Fridays tend to have the lowest levels among weekdays. During weekends, CO_2 Concentrations remain consistently flat at a healthy level, similar to outdoor air.

Finally, the quality of each classroom in terms of CO_2 Concentration relative to the comfort zone will be analysed. Figure 24 shows a clear trend indicating that the first three floors are the most problematic, frequently exceeding CO_2 Concentration comfort levels. It is known that these floors are the most used of the building. Classrooms A-13 and A-32 are the most affected, being outside the comfort zone more than

5% of the class hours. On a monthly basis, February and November show the most exceedance. The fifth floor consistently shows tiny exceedances, similar to the 4th and 6th, though the latter two floors exceed the comfort threshold a bit more. Floor 3 is the most problematic, particularly in November. As expected, vacation months show no exceedance.

6.2.3 Humidity Analysis

Firstly, comparing the comfort zone for humidity percentage with the actual measured values allows for an assessment of air quality. Figure 27 indicates that most samples fall within the comfort zone, unlike the temperature data. Humidity levels fluctuate significantly without a pronounced seasonal trend. However, a general pattern can be observed: humidity values tend to exceed the comfort zone in summer and fall lower than it in winter. An outlier occurs on October 23rd, with a humidity level reaching 80%.

Regarding the differences between floors, Figure 26 shows that humidity levels fluctuate more widely on the lower floors, except for the sixth floor, which reaches slightly higher values than the fourth and fifth floors. Given that the humidity comfort zone ranges from 30% to 60%, the plot confirms that most values fall within this range.

Now, let's analyze the impact of humidity on temperature. First, we examine the frequency of high humidity and high temperature occurring simultaneously during class hours, in Figure 11. The three months where this issue is most prevalent are June, September, and October, in increasing order of severity. Additionally, there is a noticeable trend that classrooms on the first floor experience a higher incidence of both elevated temperature and humidity.

Delving deeper into this topic, apparent temperature can be analyzed to better understand students' real sensations. Knowing that Figure 12 shows data only from class hours on weekdays, excluding July and August, reveals that caution and extreme caution levels are reached during these conditions, meaning that students and teachers suffer them. Extreme caution, defined as apparent temperatures starting at 33°C, is only reached at the end of June, a period when regular classes are not in session. However, this time is still used for activities such as thesis presentations and exam reevaluations. The biggest concern is the caution level, corresponding to apparent temperatures between 27°C and 32°C, which is reached during June, September and the first half of October. These high temperatures are excessive for a working environment, which ideally should not exceed 25°C.

Information gathered from various sources tells that caution heat index can cause potential fatigue and discomfort from physical activity, while extreme caution levels can lead to more severe effects such as heat cramps, heat exhaustion, or heat stroke due to prolonged exposure or physical exertion. The good point is that danger and extreme danger levels of apparent temperature (the highest categories) are not reached in ETSAB's Building A classrooms. However, even reaching the caution level indicates a need for concern and corrective action.

The final subject of analysis is each classroom's air quality, focusing now on humidity. The analysis compares the total percentage of hours each classroom falls below or exceeds the comfort zone during class hours on weekdays. Figure 27 reveals that floor 1 has the most issues with high humidity, as noted before. Another trend is high humidity is more likely in the classrooms numbered 1 on each floor, likely due to their proximity to the bathrooms. Low humidity levels, defined as below 30%, occur more sporadically through the building, making the pattern less clear. Low humidity is more common than high humidity. The most problematic classrooms for low humidity are also those numbered 1 on each floor, however, now the 5th floor is more affected.

Monthly analysis shows that high humidity levels occur from June to October, with August and October being the most problematic months. Low humidity levels are prevalent from November to March, with January and February being the most problematic. The most affected classrooms by low humidity are A-31, A-61 and A-52, falling below comfort more than 35% of the hours. Regarding high humidity, A-11 is clearly the most problematic, along with all classrooms on floor 1. Class A-42 also shows surprisingly high humidity levels.

6.2.4 Global Air Quality of ETSAB Building A Classrooms

This last subsection provides a comprehensive comparison of the air quality in ETSAB Building A classrooms, focusing on all variables together and comparing data between quarters. The results indicate that temperature is the variable most frequently out of the comfort zone, with all classes experiencing deviations for 20% to 60% of the class hours. Moreover, temperature deviations are widely distributed, with a slight trend showing more frequent deviations in the northern (higher numbered) classrooms. However, CO_2 Concentration issues are more prominent on the lower three floors, while humidity deviations are dispersed but tend to be higher in classrooms numbered as 1, as previously said, and best in the middle classrooms of the building. Given that each variable shows different patterns, there's no evidence in terms of overall air quality. For instance: A-36 has the worst occurrence of temperature deviations but is medium on CO_2 Concentration levels and one of the best in humidity. Similarly, most classrooms have a mix of issues, there's no evident best or worst class.

Comparing data by course and each quarter reveals that the first quarter has the highest percentage of class hours outside the comfort zone for all variables, indicating the worst air quality. Humidity shows a significant difference. In terms of CO_2 Concentration and temperature, the first quarter shows a more global trend of deviations, while the second quarter shows more localized problems. Some interesting aspects are that classroom A-61 is problematic in the first quarter but not in the second, potentially related to scheduling differences. On the contrary, A-32 is problematic in the second quarter but not in the first.

The last and one of the most important analysis to perform is the one of variable deviation magnitude. By examining the quality metrics based on deviation from comfort, it is evident that the third floor has the worst quality classrooms in terms of temperature and CO_2 Concentration. Classroom A-36 deviates the most in temperature during the first quarter and is also one of the worst for CO_2 Concentration in the same period. A-31 has the highest CO_2 Concentration deviations from comfort in the first quarter. In the second quarter, A-33 is the most deviated for CO_2, while the others remain low deviated.

Comparing the number of deviations with the magnitude of deviations allows taking additional conclusions. For temperature, deviation occurrences are dispersed throughout the building, but classrooms A-33 and A-36 show the highest deviation magnitudes, specifically A-36 in the first quarter and A-33 in the second. Regarding CO_2, A-13 has the most frequent deviations in the first quarter, but A-31 has the highest magnitude of deviations. In the second quarter, A-32 is both the most frequently deviated and has the largest deviations, indicating a significant problem from February to June. For humidity, the occurrence and magnitude of deviations follow the same distribution throughout the building, with more pronounced deviations on the 5th and 6th floors.

7. Conclusions

7.1 Achieved Objectives

To finalize, the initial questions will be revisited, and their answers summarized and provided here. Only the most important results will be stated.

1. **Is there any relationship between the temperature of the campus and inside the classrooms? And with the energy consumption?**

   As expected, it has been observed that greater indoor-outdoor temperature differences lead to higher energy consumption. Additionally, during spring and summer, when temperatures are higher, energy consumption tends to decrease.

2. **What are the patterns in energy consumption at ETSAB? Is it possible to identify any seasonal trends?**

   Seasonal patterns vary significantly depending on the building. Building A exhibits the clearest common pattern among all seasons, with two energy peaks: one in the morning and another in the late afternoon. Among all the buildings, autumn consistently is one of the periods with the highest consumption. Surprisingly, Building B has summer as one of the seasons with higher consumption.

3. **Is air quality of the classrooms appropriate for a working and healthy environment, in terms of temperature, concentration of CO_2, percentage of humidity and temperature? Which classrooms are the most problematic?**

   Air quality of Building Segarra (A) classrooms is not always in the comfort levels it should be. Temperature is the most frequently deviated variable from comfort, with all classes being out of comfort for 30% to 50% of the class hours during 2023. Humidity is the second most frequently worst variable, and CO_2 concentration the best, being out of comfort maximum 6% of the hours. Air quality problems are quite variable depending on the time of the year and the variable, but some of the worst classes are A-36, A-33 and A-13.

4. **Which are the peak hours of CO_2 concentration in classrooms?**

   On all weekdays, CO_2 peaks typically occur around 11 AM and 5 PM. However, on Mondays, the peaks shift slightly to 10 AM and 6 PM. On Fridays, the peaks are less pronounced compared to other days.

5. **How is the impact of humidity on temperature? Does apparent temperature reach dangerous values?**

   High temperature values combined with high humidity can alter the perceived temperature in humans, making warm conditions more uncomfortable. In Building A classrooms, this combination of high humidity and temperature occurs in June, September, and October. Among these months, October is the worst, with the main problem on the first floor. During class hours in these months, the apparent temperature reaches a level labeled as caution. At the end of June, the sensation reaches extreme caution with an apparent temperature higher than 33°C.

6. **How does air quality vary across the different floors of the building? Do windows affect the CO_2 concentration?**

   Regarding humidity, Floor 1 has the biggest problem due to its wide range. This floor also experiences the lowest temperatures, while Floors 3 and 4 have higher temperatures. Floor 3 is the worst in terms of CO_2 concentration, coinciding with the fact that the classrooms on this floor are smaller and have only four windows. Number of windows clearly affects air quality.

7.2 Ethical Implications

The primary motivation behind this project lies in the university's commitment to the well-being of its students and professors, as well as the responsible consumption of energy. By developing this tool, my goal is to make a positive ethical impact in these areas by providing professionals access to crucial information that facilitates decision-making.

The first step to promote energy efficiency is understanding and assessing the actual situation, analysing it to then take action. So, this app serves as a tool for professionals to conduct this analysis. By gaining insights into energy consumption patterns and efficiency opportunities, decision-makers can identify areas for improvement and implement strategies that reduce environmental impact.

Furthermore, ensuring air quality in the university environment not only a practical necessity but also a matter of ethical importance. After COVID-19 pandemic, this topic started being a bigger concern. Enhancing air quality is essential for supporting the health and well-being of students and workers.

With this project, I aspire to contribute to the university's sustainability goals and ethical principles. In terms of Objectives for Sustainable Development (ODS), this project contributes to three of them: quality education (Goal 4), responsible production and consumption (Goal 12), and climate action (Goal 13). Through informed decision-making, UPC can create an environment that supports the academic mission prioritizing the well-being of all people involved.

7.3 Future Work

After developing this project, several opportunities and ideas have emerged. Firstly, this project focuses on the data from ETSAB in 2023, with air quality analysis limited to Building A. Therefore, an important next step could involve extending the analysis to include other buildings and areas of the campus, as well as data from other years. A potential project could involve expanding the developed app to accommodate different dates and campus zones.

Furthermore, this tool could be customized to meet the specific needs of environmental and architecture professionals, among others. Lots of possible visualizations can be designed, so, obtaining professional feedback on the insights gained could lead to the development of additional visualizations or improvements to the existing ones.

Another area of focus could be optimizing the performance of the app. While this project primarily concentrated on developing visualizations that effectively communicate findings and address key questions, improving the speed and efficiency of the app could enhance user experience.

Additionally, the utilization of machine learning methods could further improve the app's capabilities. By incorporating classroom schedules, it may be possible to predict more specific air quality patterns and identify periods and courses during which air quality tends to be at its worst.

References

[1] Servei d'Infraestructures UPC 2023. *Plànols de Distribució ETSAB 2023*

[2] Snowflake Inc n.d. *API reference (Streamlit)* Retrieved March 23, 2024, from https://docs.streamlit.io/develop/api-reference

[3] Vega-Altair Developers n.d. *API reference (Altair)* n.d. Retrieved March 3, 2024, from https://altair-viz.github.io/user_guide/api.html

[4] Plotly n.d. *Plotly Open Source Graphing Library for Python* n.d. Retrieved April 22, 2024, from https://plotly.com/python/

[5] NumFOCUS Inc n.d. *API reference (Pandas)*. Retrieved March 3, 2024, from https://pandas.pydata.org/docs/reference/index.html#api

[6] Jesús Soto 2020, *Índice de sensación térmica (Heat Index)*. Clima.bio. Retrieved June 6, 2024, from https://clima.bio/indice-de-sensacion-termica-heat-index/

[7] Servei d'Infraestructures UPC 2021, *Informe SIRENA 2020: Avaluació del consum d'energia i aigua de la UPC* . Retrieved March 15, 2024, from https://sostenible.upc.edu/ca/reptes/imatges/informe-sirena-2020.pdf

[8] UPC Sostenible n.d., *Campus Saludable: Disseny de l'estratègia. Quins reptes tenim a la UPC?*. Retrieved March 15, 2024, from https://sostenible.upc.edu/ca/reptes/imatges/campus-saludable-sessio-2-_on-som_.pdf/view

[9] Praegresus n.d, *Encycolorpedia*. Retrieved April 12, 2024, from https://encycolorpedia.com/named

[10] *Color Schemes* n.d. Retrieved April 12, 2024, from https://vega.github.io/vega/docs/schemes/

[11] *GeeksforGeeks* n.d. Retrieved March 15, 2024, from https://www.geeksforgeeks.org/

[12] Altair Engineering, Inc. n.d. *Altair Community* n.d. Retrieved April 12, 2024, from https://community.altair.com/community/en

[13] DEXMA n.d, *SIRENA*. Retrieved March 2, 2024, from https://upcsirena.app.dexma.com/analysis/consumption/dis

[14] Overleaf n.d, *Documentation*. Retrieved May 26, 2024, from https://es.overleaf.com/learn

[15] Apache Software Foundation (2017), *Echarts Examples*. Retrieved March 19, 2024, from https://echarts.apache.org/examples/en/index.html

[16] CO_2 Meter 2023. *Carbon Dioxide Levels Chart*. Retrieved May 5, 2024, from https://www.co2meter.com/en-es/blogs/news/carbon-dioxide-indoor-levels-chart.

[17] Disruptive Technologies 2022. *Carbon Dioxide (CO2) and Employee Productivity: All You Need To Know*. Retrieved May 5, 2024, from https://www.disruptive-technologies.com/explore/carbon-dioxide-co2-employee-productivity.

[18] The Severn Group 2018. *The Right Office Humidity for Your Building*. Retrieved May 5, 2024, from https://www.theseverngroup.com/office-humidity/.

[19] Quirónprevención 2018. *Temperatura en la oficina ¿cuál es la ideal?*. Retrieved May 5, 2024, from https://www.quironprevencion.com/blogs/es/prevenidos/temperatura-oficina-ideal.

[20] Marie Bannister 2020. *Humidity: the surprising factor that can impact employee wellbeing*. Retrieved May 11, 2024, from https://www.airthings.com/business/resources/humidity-office

[21] *Comunitat UPC Sostenible* n.d. Retrieved June 7, 2024, from https://sostenible.upc.edu/ca.

[22] United Nations: Department of Economic and Social Affairs n.d. *THE 17 GOALS*. Retrieved June 15, 2024, from https://sdgs.un.org/goals

[23] Stack Exchange Inc. n.d. *StackOverflow* Retrieved March 3, 2024, from https://stackoverflow.com/

[24] OpenAI (2022). ChatGPT. Retrieved March 3, 2024, from https://chat.openai.com/

[25] *Posit Community*. Retrieved April 17, 2024, from https://forum.posit.co/

A. Appendix: Visualizations

This appendix contains all the visualizations generated during the project placed in the app. While the main report includes only the most important visualizations to answer the key questions, this annex provides a complete set of charts and graphs to offer a comprehensive view of the data analysis performed, as well as the HomePage.

The link of the app is: https://ppkbzcssgnvlursgqtlcfp.streamlit.app/

A.1 HomePage

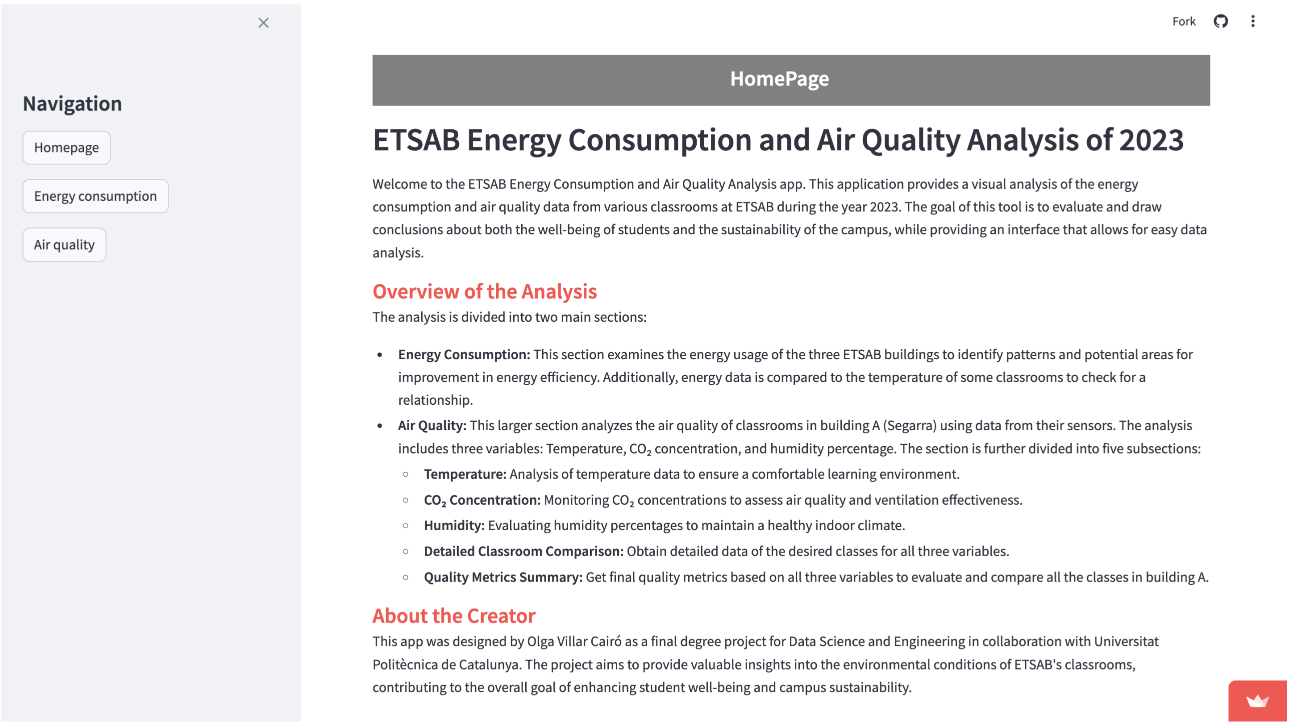

Figure 18: App HomePage.

A.2 Energy Consumption Visualizations

The first visualization in this section of the app is shown in Figure 3.

The other energy charts are shown in Figure 4 and Figure 5.

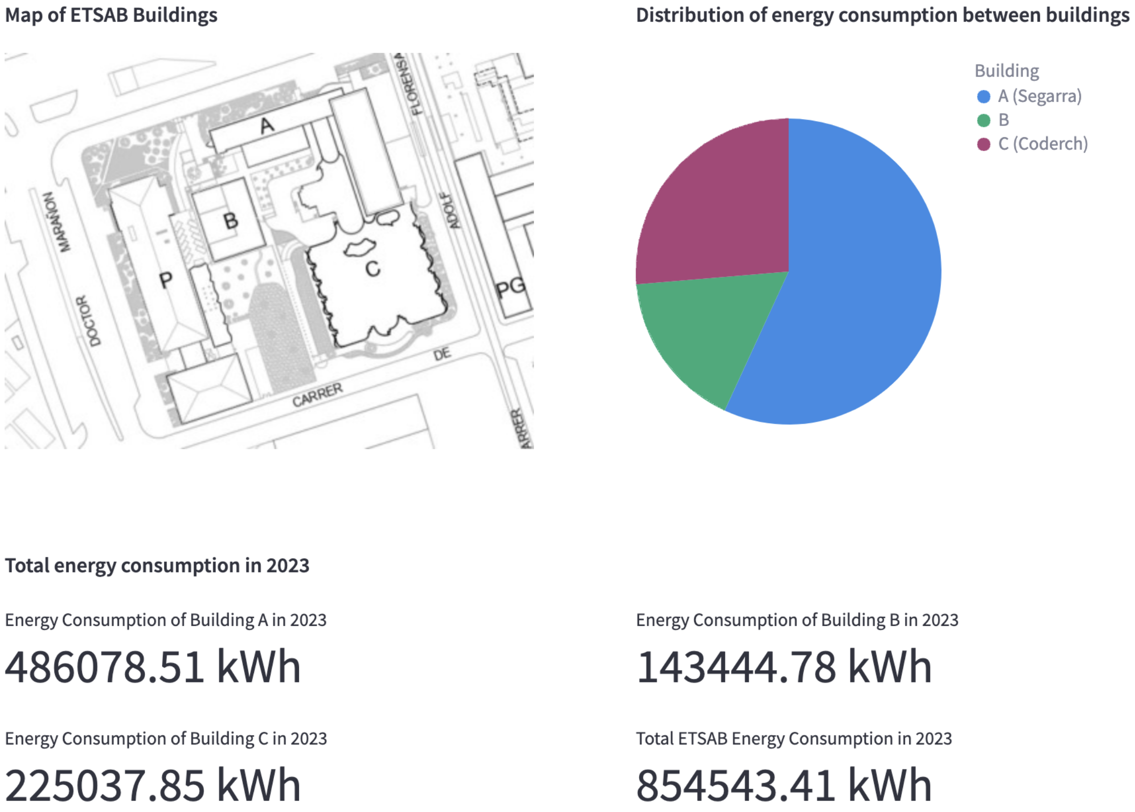

Figure 19: Map of ETSAB Buildings and Pie Chart of Energy Consumption.

A.3 Air Quality Visualizations

A.3.1 Temperature Visualizations

The fist temperature visualization is shown in Figure 6.

The other charts in the app for temperature are shown in Figure 7, Figure 8 and Figure 9.

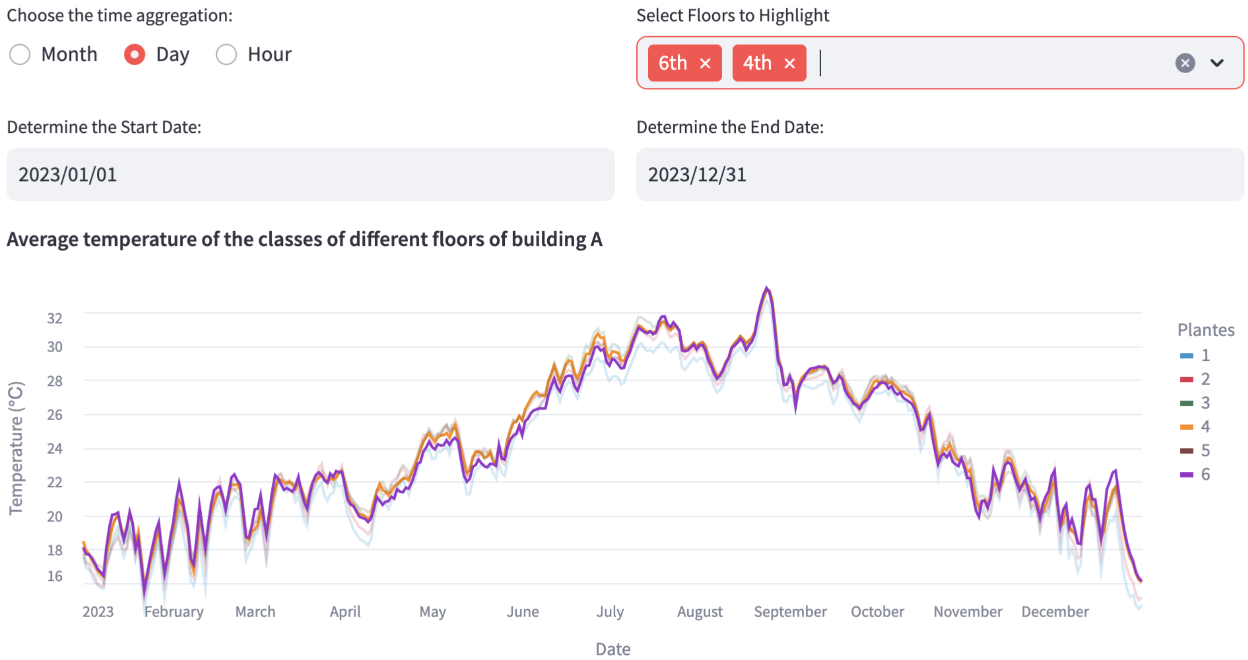

Figure 20: Average Floor Temperature.

A.3.2 CO_2 Concentration Visualizations

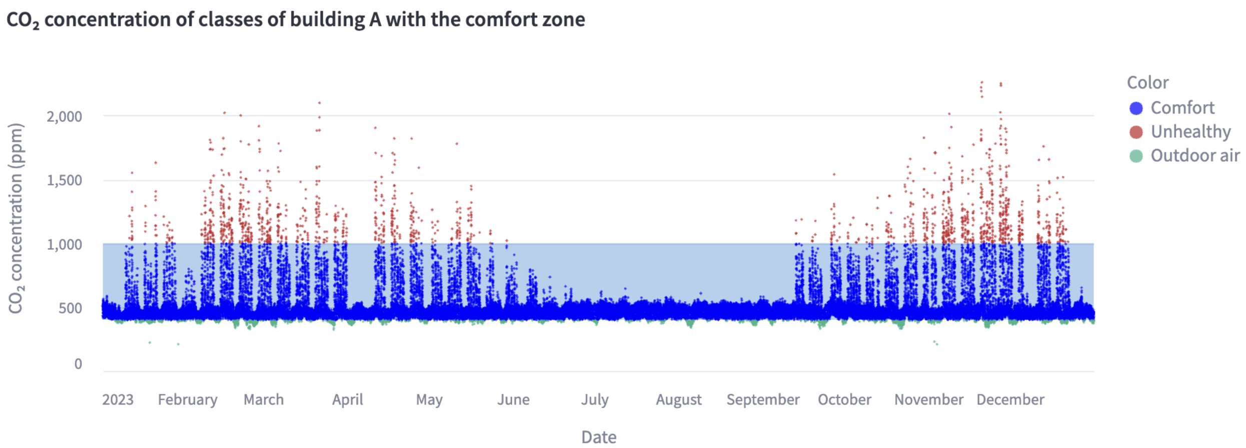

Figure 21: CO_2 Concentration and Comfort Scatterplot.

Another visualization for this section of the app is shown in Figure 10.

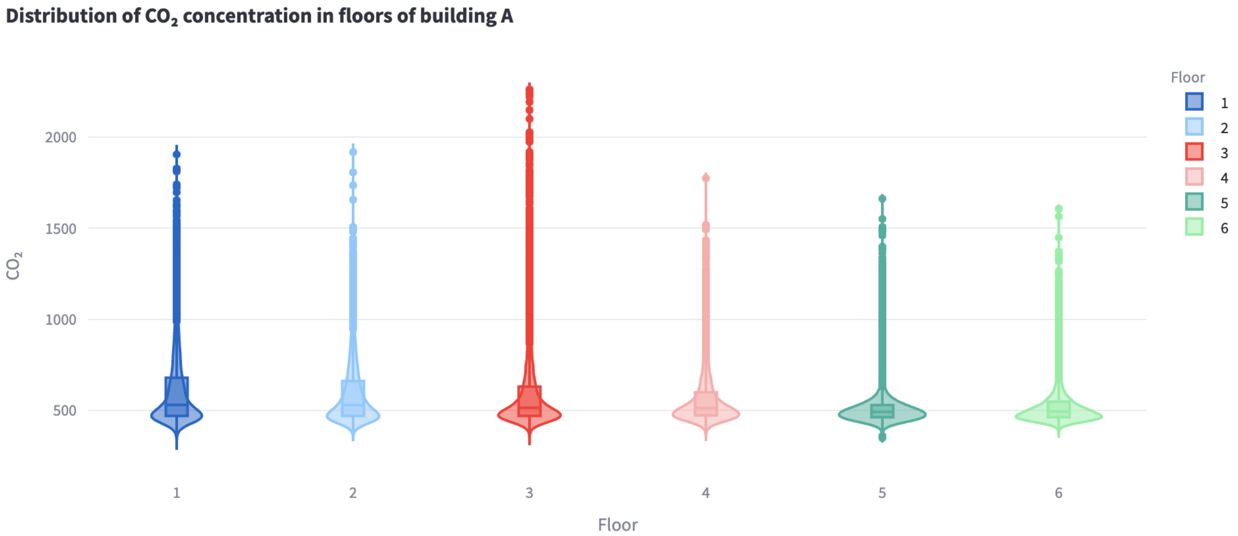

Figure 22: CO_2 Concentration Violin Plot for Floors.

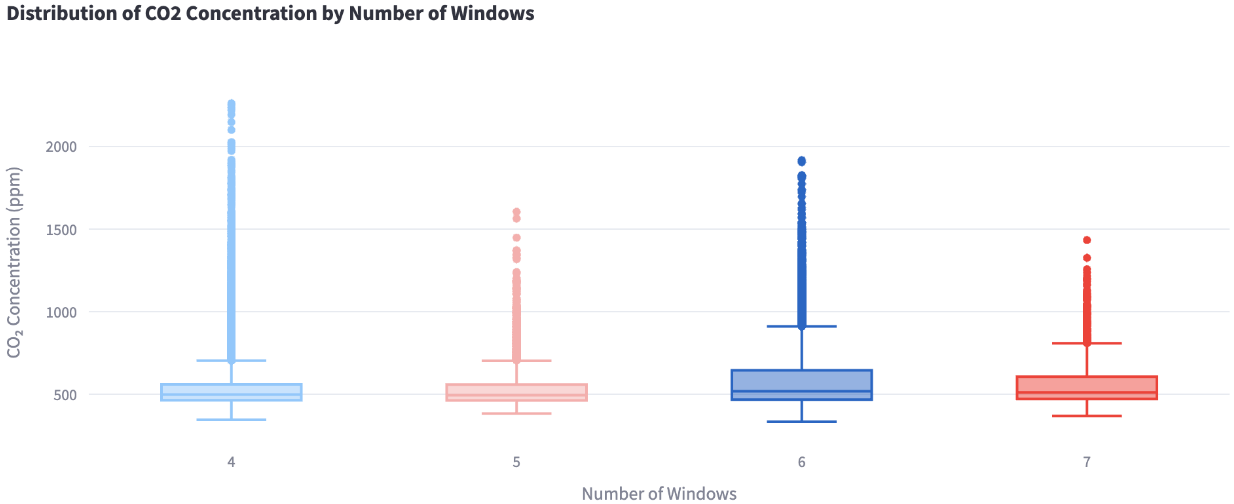

Figure 23: CO_2 Concentration Distribution for Number of Windows.

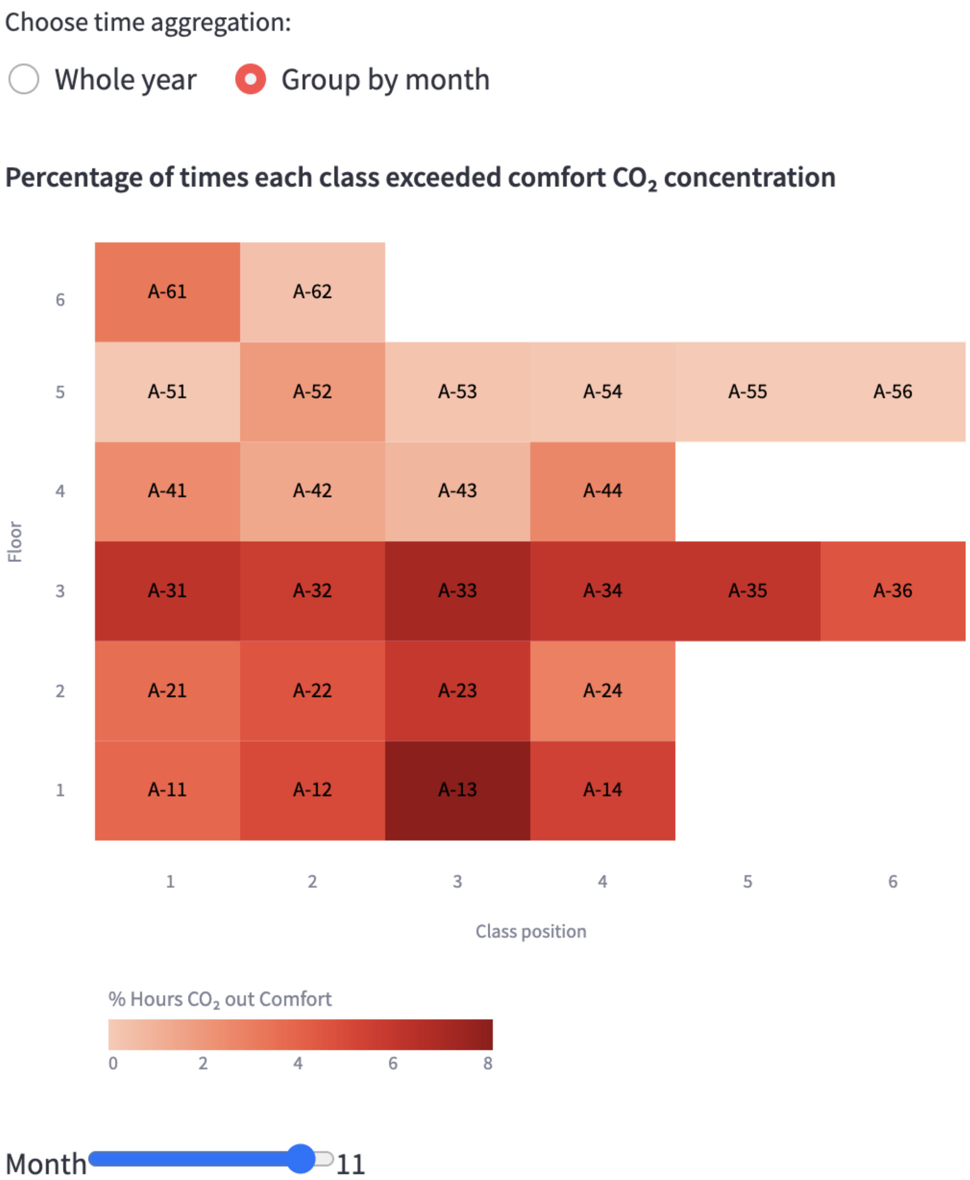

Figure 24: Percentage of Hours with CO_2 Concentration Higher than Comfort.

A.3.3 Humidity Visualizations

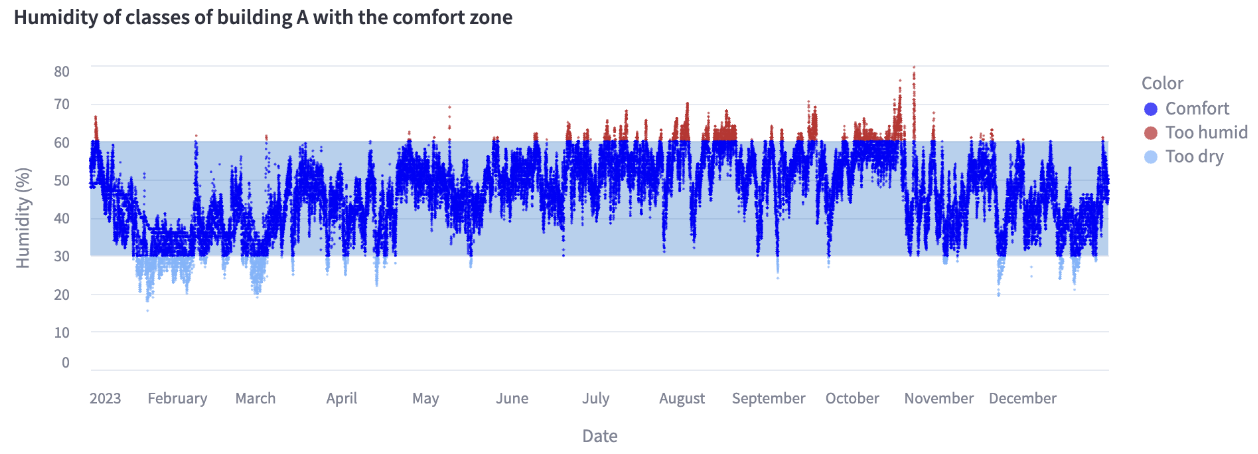

Figure 25: Percentage of Humidity and Comfort Scatterplot.

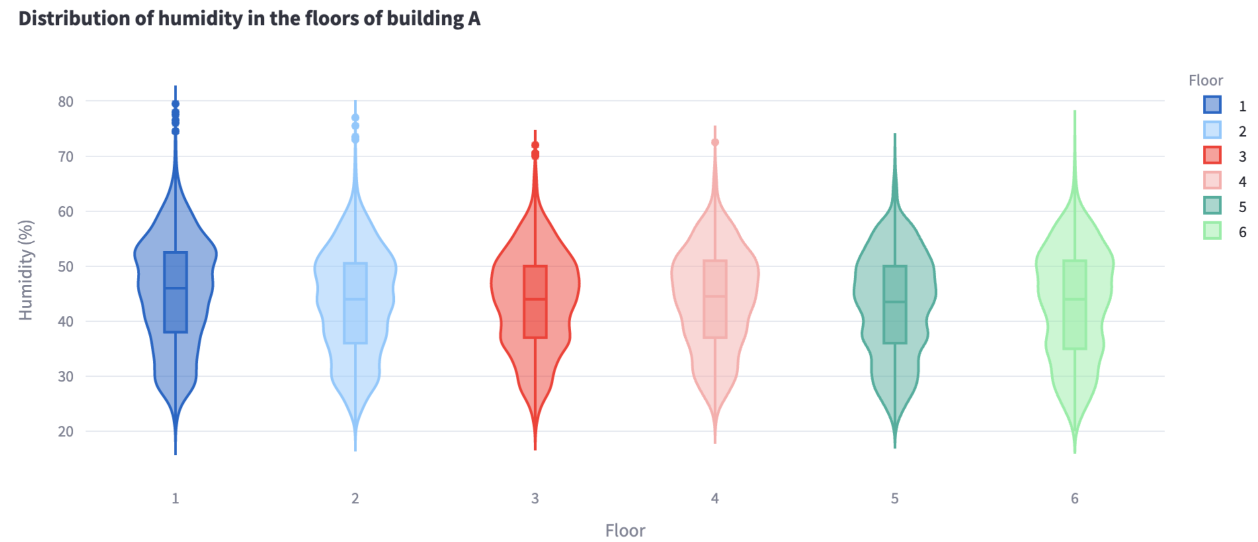

Figure 26: Percentage of Humidity Violin Plot for Floors.

The other two plots of this section of the app are shown in Figure 11 and 12.

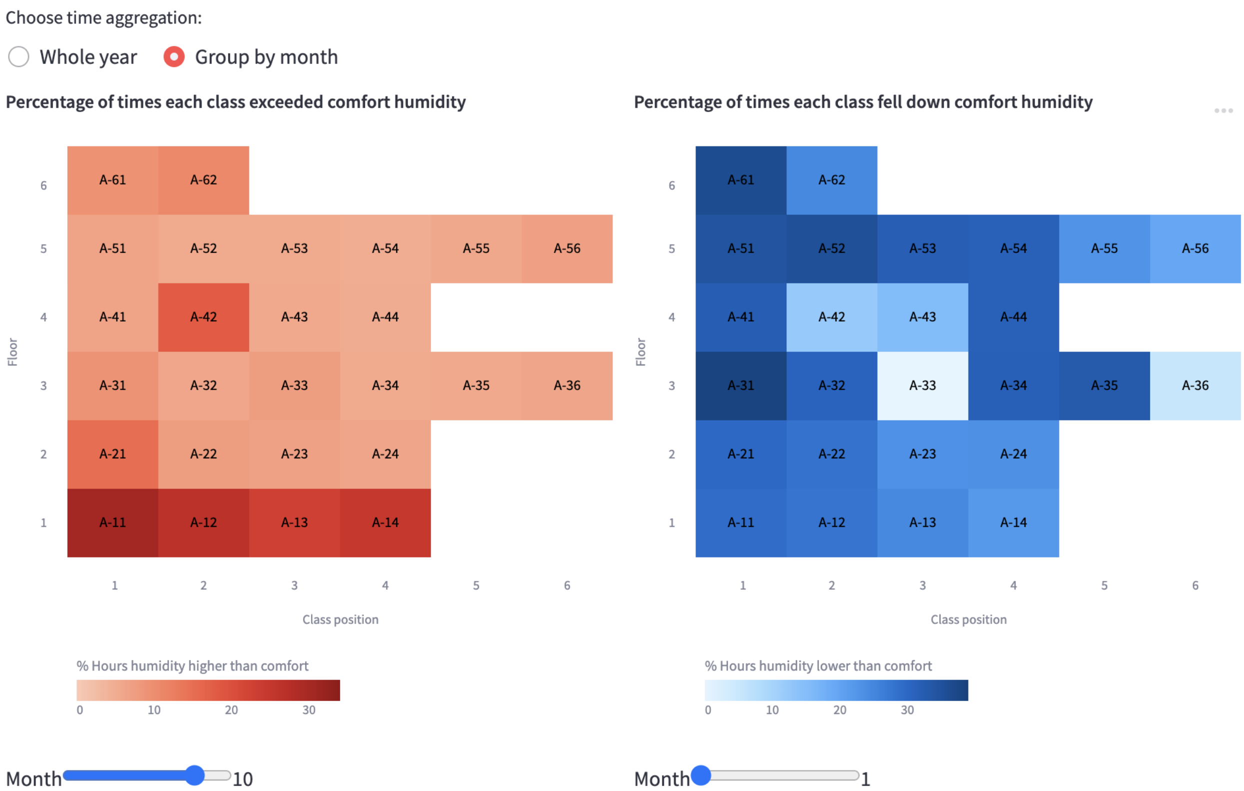

Figure 27: Percentage of Hours with Percentage of Humidity Outside Comfort.

A.3.4 Detailed Classroom Comparison

This subsection visualizations are shown in Figure 13.

A.3.5 Quality Metrics Summary

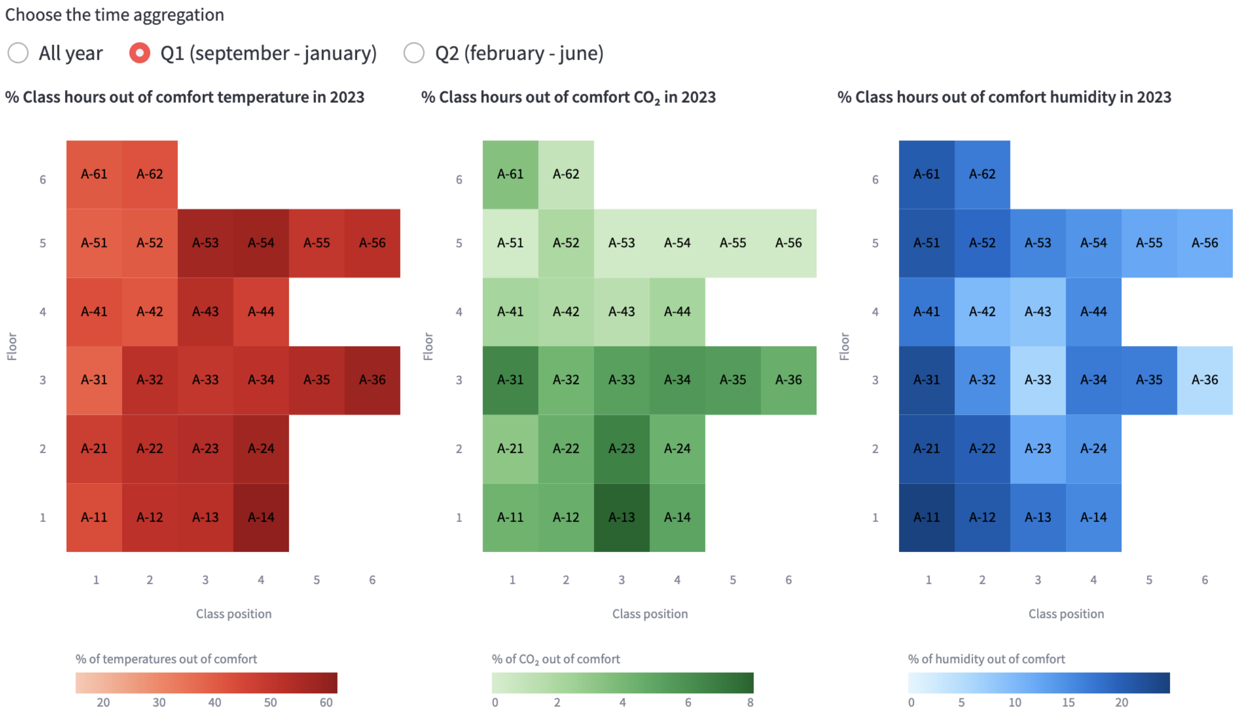

Figure 28: Percentage of Class Hours being Out of Comfort for each Variable.

The last chart is shown in Figure 14.

B. Vacation Days of 2023

In the energy graphics, the vacation days are the ones where the campus is completely closed, as there's no consumption. So, the following dates are excluded from the energy consumption analysis: 2023-09-25, 2023-10-12, 2023-11-01, 2023-12-06, 2023-12-07, 2023-12-08, 2023-12-23, 2023-12-24, 2023-12-25, 2023-12-26, 2023-12-27, 2023-12-28, 2023-12-29, 2023-12-30, 2023-12-31, 2023-01-01, 2023-01-02, 2023-01-03, 2023-01-04, 2023-01-05, 2023-01-06, 2023-01-07, 2023-01-08, 2023-01-09, 2023-01-10, 2023-04-01, 2023-04-02, 2023-04-03, 2023-04-04, 2023-04-05, 2023-04-06, 2023-04-07, 2023-04-08, 2023-04-09, 2023-04-10, 2023-05-01, 2023-09-11. All August days and weekends are also excluded.

In the temperature graphics, the vacation days are the ones where there's no class or activity for the students, as there's no usage of the classrooms. So, the same dates as in energy graphics are excluded from the air quality data, adding July, which is also excluded.