TREBALL DE FI DE GRAU

Grau en Estadística

Títol: Anàlisis i Visualitzacions del Benestar a les Aules de la Facultat de Matemàtiques i Estadística

Autor: Damari Alicia Paredes García

Director: Pere-Pau Vázquez Alcocer

Co-directora: Guadalupe Gómez Melis

Departament: Departament de ciències de la computació. Departament d'estadística i investigació operativa

Convocatòria: Setembre 2023

UNIVERSITAT DE BARCELONA

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

Resum

El punt de partida d'aquest treball és l'anàlisi estadístic de les dades recollides pel sistema *QaireUPC*. La base de dades inclou mesures de temperatura, humitat i nivells de diòxid de carboni (CO_2). Aquests paràmetres es consideren indicadors de la qualitat de l'aire i l'estudi es focalitza en deixar de manifest la situació de benestar d'estudiantat i professorat dintre de les aules de la Facultat de Matemàtiques i Estadística (FME) de la UPC. Aquest objectiu s'aconsegueix mitjançant mètodes d'anàlisi estadístic propis de la estadística descriptiva quantitativa. Simultàniament, al llarg del treball es fa una reflexió de la forta relació entre visualitzacions i anàlisis estadístic (sota un context de grans bases de dades) mitjançant la creació de visualitzacions amb la llibreria Altair de Python. Arribant a la conclusió de que actualment no totes les aules compleixen amb les condicions òptimes de benestar i reafirmant que els complements gràfics són una valuosa eina visual per comunicar la informació.

Paraules clau: Qualitat de l'aire, anàlisis descriptiu, visualitzacions, benestar, temperatura, humitat, CO_2, Altair.

Abstract

The starting point of this project is the statistical analysis of the data base obtained from the system *QaireUPC*. The data includes measurements of temperature, humidity, and levels of carbon dioxide (CO_2). These factors are considered indicators of air quality and the study tries to show the situation of wellbeing of the students and teachers in the classrooms of the Faculty of Mathematics and Statistics (FME) from the UPC. The goal is achieved by typical quantitative descriptive statistical methods. Simultaneously, there's a discussion throughout the paper about the strong relationship between the visual tools and the statistics field. In this instance, the objective is being tested with the creation of multiple visualizations using the Altair library from Python. Reaching the conclusions that many classrooms don't meet the optimal comfort levels and emphasizing the importance of the visualizations to communicate the information.

Key words: Quality air, descriptive analysis, visualizations, wellbeing, temperature, humidity, CO_2, Altair.

CLASSIFICACIÓ AMS

- *62-07 Data Analysis*
- *62-09 Graphical methods*
- *62H35 Image analysis*
- *62P12 Applications to environmental and related topics*
- *97K40 Descriptive statistics*

AGRAÏMENTS

Vull donar les gràcies primer de tot al meu tutor de TFG Pere-Pau, amb el suport del qual no hauria estat capaç de tirar endavant el projecte. També a totes les persones que m'han fet costat durant el temps de l'estudi. Per últim a tots els professors que han aportat el seu granet a la meva formació durant tota la carrera.

Índex de continguts

I. INTRODUCCIÓ

Aquest treball està realitzat com a Treball de Final de Grau (TFG) dintre del programa formatiu corresponent al Grau Interuniversitari d'Estadística de la Universitat de Barcelona (UB) i la Universitat Politècnica de Catalunya (UPC). Ha estat tutoritzat pel professor Pere-Pau Vázquez Alcocer del Departament de Ciències de la Computació i co-tutoritzat per la professora Guadalupe Gómez Melis del Departament d'Investigació Operativa.

1.1 Motivació

La motivació d'aquest treball ha estat en gran mesura gràcies al director del Treball de Final de Grau (TFG), Pere-Pau Vázquez. Va ser a les seves classes de Visualització de la Informació on va néixer l'interès de fer un treball que tingués un fort lligam amb la manera de com fem servir les diferents eines visuals per comunicar els diferents anàlisis estadístics.

Al principi es van proposar diversos temes però finalment el més destacat va ser l'anàlisi de la qualitat de l'aire en diferents aules de la Universitat Politècnica de Catalunya (UPC). De seguida el focus va passar a considerar el benestar d'alumnes i professors per mitjà d'aquest anàlisis. I a més, es tractava d'un projecte on hi havia la possibilitat de treballar amb les dades que s'estaven recollint en aquells instants a tota la universitat i la idea va ser prou atractiva com per servir de motivació d'aquest projecte.

1.2 Context

1.2.1 Qualitat de l'aire

Des de fa molts temps la problemàtica dels nivells de contaminació de l'aire és un tema constant relacionat amb el canvi climàtic i amb el benestar de la societat. Es pot trobar una gran quantitat d'estudis relacionats, com ara un article publicat al *National Geographic* al febrer de l'any 2019 on es diu que la *Environmental Protection Agency* (l'Agencia de Protecció Mediambiental), va enregistrar un 15% més de dies amb aire insalubre als Estats Units durant el 2018 i 2017 en comparació a la mitjana dels anys 2013 a 2016. A la pàgina web de *World Health Organization* trobem que l'any 2019 es va estimar que les condicions d'aire poc sa havien provocat 4.2 milions de morts prematures a tot el món.

S'ha de tenir present que una situació d'aire mal sa, tant a exteriors com interiors, no només és símptoma de contaminació ,i per tant de canvi climàtic, sinó que pot derivar en greus problemes de salut. Aquests poden ser derivats de llargues exposicions i poden acabar provocant migranyes, asma o altres problemes respiratoris, problemes cardíacs, càncer, etc. Com també poden ser derivats per exposicions a més curt termini amb resultats com ara mal de cap, mareig, irritacions oculars, etc. Tots aquests problemes no només estan provocat per contaminants com ara el diòxid de carboni (CO_2), sinó que també es veuen agreujats per altes temperatures i males condicions d'humitat de l'ambient.

En els últims anys s'han intentat implementar cada vegada més mesures per pal·liar aquesta situació, sobretot pel que fa a exteriors. Actualment a Barcelona hi ha diversos plans que s'estan executant per millorar la qualitat de l'aire, com ara afavorir el vehicles menys contaminats, crear zones de baixes emissions, fomentar el transport públic, entre altres mesures (https://ajuntament.barcelona.cat/qualitataire/es).

***Imatge* 1:** *Cartell indicador de la circulació en zones de baixes emissions*

1.2.2 Bases de dades

S'ha de tenir en compte que ens trobem en una època de revolució tecnològica i en la qual generem de manera constant una gran quantitat de dades. Aquest fet s'ha d'aprofitar per actuar davant d'aquest tipus de problemàtiques, ja sigui en temes de salut, temes socials o altres camps.

En l'actualitat existeixen moltes fonts d'informació i moltes eines que ens permeten tenir un accés pràcticament immediat a la informació en temps real. Però no es suficient amb acumular totes aquestes dades, s'ha de saber fer un bon ús de tot el que tenim a l'abast. S'ha de tenir

responsabilitat social i s'ha d'advocar per la reproductibilitat dels treballs, tenint en compte que no tots els mètodes són aplicables de la mateixa manera i en totes les situacions. S'han de saber aplicar els mètodes adequats per obtenir bons anàlisis.

***Imatge* 2.** *Aparell de mesura de la qualitat de l'aire a Barcelona*

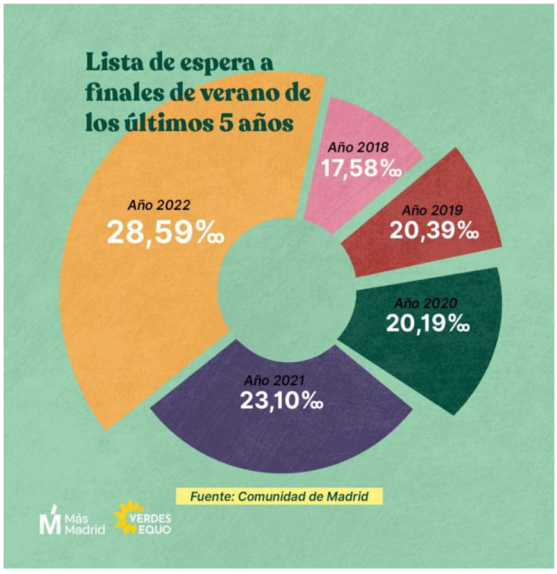
A la província de Barcelona, per exemple, es troben col·locats diversos dispositius que mesuren la qualitat de l'aire i a la pàgina web de la Generalitat de Catalunya (https://mediambient.gencat.cat/ca/05_ambits_dactuacio/atmosfera/qualitat_de_laire/vols-saber-que-respires/), podem trobar un mapa interactiu que ens permet saber quin és el nivell de qualitat de l'aire en les diferents localitzacions. Fent servir aquest tipus d'eines que enregistren dades al llarg de grans períodes i de manera pràcticament ininterrompuda obtenim grans bases de dades que serveixen per realitzar anàlisis i per conscienciar cada vegada més a les persones.

1.2.3 Visualització de dades

El mapa interactiu que s'ha comentat a l'apartat anterior és un exemple d'eina visual que podem trobar avui en dia i que ens permet tenir una idea ràpida de la situació estudiada.

Segons la definició de T. Munzner "Computer-based visualization systems provide **visual** representations **of datasets** designed to **help people** carry out tasks more **effectively**". És a dir, la visualització de les dades ha de ser una eina visual de representació de dades que ajudi a les persones a realitzar tasques de forma més efectiva. Per tant, s'ha de ser conscient del que es vol transmetre i de qui serà el receptor d'aquesta informació. Ja que la manera com es visualitzen els resultats és molt important per tal de transmetre la informació correcte.

***Imatge* 3:** *Exemples de classe. Visualitzacions inexactes o amb una clares tendències.*



Lamentablement no tothom en fa un bon ús i podem veure a diferents fonts d'informació com es té poca cura o es manipulen les eines visuals per tal de modificar la percepció dels usuaris menys coneixedors de la matèria.

De la importància de les visualitzacions es parlarà en més detall dintre de marc del projecte.

1.2.4 Pandèmia Covid-19

Finalment, hi ha un altre esdeveniment important que cal situar pel context d'aquest projecte, la pandèmia de Covid-19.

A finals de l'any 2019 i principis del 2020 es va declarar una crisi sanitària a nivell internacional davant d'una malaltia provocada per un virus (*SARS-COV-2*) de transmissió aèria, la necessitat de controlar de manera efectiva els sistemes de ventilació dels espais tancats per tal de controlar l'expansió de la que va acabar sent una pandèmia, es va veure incrementada.

A arrel d'aquesta situació es va potenciar, mantenint-se també un cop passada la crisis, la investigació de cara a l'estat de la qualitat de l'aire als interiors dels edificis. I és a partir d'aquest succés que neix l'eina **QaireUPC** que es va presentar per realitzar aquest treball i que recull mesures de nivells de CO2, temperatura i humitat en diferents instal·lacions de la UPC. Tots tres són indicadors de la qualitat de l'aire i agreujants de malalties en nivells inadequats.

1.3 Objectius

Un cop es va definir quina era la temàtica principal d'aquest treball es van plantejar diferents objectius.

L'objectiu principal d'aquest treball és fer un tractament estadístic de les dades recollides pel sistema *QaireUPC* corresponents a les aules de la Facultat de Matemàtiques i Estadística (FME). L'anàlisi tindrà un caràcter principal de tipus descriptiu i a continuació es pretenen crear una sèrie de visualitzacions relacionades amb les dades.

Concretament s'esperen obtenir resultats que puguin ser il·lustratius sobre el benestar dels estudiants a les aules de la facultat. Ja que la qualitat de l'aire ens permet saber si els professors i els estudiants pateixen mals de cap, estat de somnolència o si baixa la productivitat tant d'ensenyament del professor com d'aprenentatge dels alumnes.

Amb les eines visuals es vol aconseguir donar una visió clara de les condicions generals de la qualitat de l'aire a la universitat. Amb els objectius de ser eficients i fàcils d'interpretar.

També es vol comprovar si hi ha algun tipus comportament anòmal que sigui destacable i d'interès tant per a la comunitat de la facultat com per a les persones encarregades de gestionar la plataforma de SIRENA.

Finalment, i de manera transversal, també es contempla la possible relació entre la base de dades estudiada i l'àmbit de sostenibilitat.

1.4 Estructura del treball

El cos principal d'aquest treball es dividirà en les dos idees principals explicades als objectius. Una primera part més relacionada amb l'anàlisi estadístic i una segons part més centrada en les visualitzacions.

Pel que fa a la estructura general el projecte estarà dividit en els apartats següents:

- **Introducció:** Motivacions, context, objectius i estructura del projecte.

- **Marc del projecte:** Marc dintre del qual s'ha realitzat aquest treball. Context de la procedència de les dades. I importància de les visualitzacions a l'estadística.

- **Metodologia:** Metodologies aplicades i descripció detallada del procés en l'elaboració d'aquest treball. Començant per la fase d'obtenció de les dades, l'establiment d'objectius i de valors límits dels paràmetre, la descripció dels processos estadístics contemplats i finalment la descripció dels recursos informàtics utilitzats.

- **Anàlisis estadístic descriptiu:** Capítol on es posarà de manifest els mètodes estadístics explicats a la metodologia en el context de la base de dades escollida. Concretament els propis de l'anàlisi descriptiva .

- **Visualitzacions:** Capítol on es presentaran les diferents visualitzacions obtingudes al llarg de tot el treball, juntament amb l'anàlisi corresponent.

- **Conclusions:** Conclusions del treball i valoracions finals del projecte.

- **Treball futur:** Possibles direccions per continuar aquest treball.

II. MARC DEL PROJECTE

En aquest apartat es parlarà sobre el marc on es realitza aquest treball, ja que les dades han estat recollides gràcies a una eina que forma part d'un programa de la UPC que vol promoure la utilització de dades de la mateixa universitat per realitzar diferents estudis. I d'aquesta manera crear espais de millora del campus com entitat.

2.1 Projecte Sirena

L'any 2006, sota un context de canvi climàtic, crisi energètica i crisi econòmica. I amb informes que destacaven la necessitat de dirigir els esforços cap a energies més sostenibles (*Stern Review on the Economics of Climate Change*[1]) i augments en demanda energètica, entre altres. Tant a nivell local, com internacional, es van adoptar diversos compromisos polítics i socials. Cal destacar que els resultats mostraven que a la Unió Europea prop del 40% de l'energia final es produïa en edificacions.

Davant la prioritat d'actuació per part d'aquest àmbit, la UPC va plantejar la necessitat de posar en marxa un sistema d'informació energètica per part de la UPC (*Pla UPC Sostenible* i Declaració de Sostenibilitat i Pla d'Estalvi Energètic 2010-2014). D'aquí neix el **projecte SIRENA**[2] (***S**istema d'**I**nformació dels **R**ecursos **EN**ergètics i l'**A**igua de la UPC*), una eina que permet fer el seguiment i avaluar l'evolució dels consums de subministraments de la UPC (electricitat, gas, aigua) així com la producció fotovoltaica i la qualitat de l'aire dels espais interiors de la UPC.

Després d'uns primers informes, algunes proves pilot i 13 plans d'Optimització energètica durant els primers anys. Finalment, es publiquen un seguit d'indicadors i d'informes regulars (Informe SIRENA 2016, Informe SIRENA 2018, Informe SIRENA 2019, Informe SIRENA 2020, Informe SIRENA 2022) sobre els canvis que s'observen als diversos edificis de la comunitat UPC.

[1] Stern, N. H. (2006). *The Economics of Climate Change : The Stern Review*. Cambridge Univ. Press.

[2] La pàgina web compta amb un accés obert per a tothom, al qual es pot accedir amb el següent enllaç: https://upcsirena.app.dexma.com/dashboard/widgets.htm. Però per fer aquest treball es va haver de demanar un accés amb compte d'usuari.

Imatge 4: *Portada de l'últim informe SIRENA realitzat fis a moment (2022)*

És dintre d'aquest projecte que es situa la recollida de la base de dades que es fa servir en aquest treball. Més concretament dintre del *Projecte QaireUPC* que forma part del Projecte SIRENA.

2.2 Projecte QaireUPC

Davant la pandèmia de la COVID-19 (desembre 2019) es va posar de manifest la importància de la renovació de l'aire a les aules per reduir la propagació del virus. Com part de les accions de la universitat per fer front a aquesta crisi sanitària es va impulsar el projecte **QaireUPC** (https://sostenible.upc.edu/ca/ca/qaireupc).

Des del mes d'octubre de 2020 la UPC va iniciar un campanya de presa de dades per monitoritzar l'evolució dels paràmetres següents:

- Concentració de CO2 (ppm)
- Temperatura (ºC)
- Humitat relativa (%)

L'objectiu primer de la recollida i l'estudi d'aquestes dades va ser, poder crear plans d'actuació adequats per reduir els contagis. Un cop l'impacte de la propagació del virus es va veure reduït i la majoria de la població va estar vacunada o va començar a estar immunitzada per causes naturals, aquest tipus de recollida de dades no va deixar de ser necessària per avaluar altres aspectes importants.

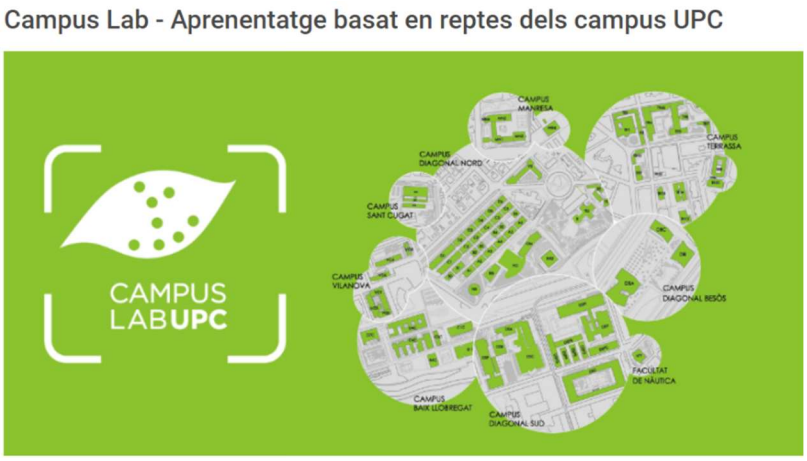
Imatge 5: *Aparells situats a les classes*

A la seva pàgina web es pot trobar un llistat de les diferents edificacions i les diferents àrees on es troben col·locats els dispositius. Així com petits informes de diferents campus de la UPC.

2.3 Campus Lab

Cal destacar que el programa SIRENA forma part d'una sèrie de projectes promoguts per la Comunitat UPC Sostenible. Aquest conjunt de projectes busquen millorar les infraestructures i processos de la UPC amb l'objectiu d'un compris social entre la universitat i la universitat com a ens. El programa en el seu conjunt és diu **CampusLAB**[3] i tal com es descriu a la seva pàgina web, "*es tracta d'un programa inspirat en els living lab que vincula l'aprenentatge de l'estudiantat, el coneixement del PDI i l'expertesa tècnica del PAS*".

Imatge* 6: Programa CampusLab*



Des del 2009 s'han realitzat diferents Treballs de Final d'Estudis (TFE) relacionats amb aquest programa i donat que el TFG descrit en aquest treball es basava en les dades recollides per una de les eines del CampusLab, gràcies a aquest programa se'm va donar l' oportunitat de poder participar en aquest programa de forma activa, no només amb la realització d'aquest treball, sinó que també vaig poder participar en una taula rodona de les **Jornades UPC de compromís social i comunitari** d'aquest any[4], celebrada a l'Escola Tècnica Superior d'Arquitectura del Vallès (ETSAV), el passat 12 de juliol.

[3] https://sostenible.upc.edu/ca/campus-lab

[4] En aquest enllaç trobareu informació sobre el desenvolupament de la jornada que va tenir lloc el 12 de juliol de 2023. https://canviaelmon.upc.edu/ca/esdeveniments/et-proposem/jornades-upc-de-compromis-social-i-comunitari

Malgrat aquest programa ha estat en funcionament des de fa un temps, sota altres noms i amb programes pilots previs, és tracta d'una eina que pocs estudiants coneixen avui en dia. Per aquest motiu en aquesta jornada es va donar molta importància en la manera d'arribar a l'estudiantat. Com part d'aquest programa vaig poder aportar la meva experiència sobre com havia decidit treballar amb aquestes dades.

2.4 Importància de les visualitzacions

Per tancar el marc dintre del qual s'ha creat aquest projecte, s'ha de parlar sobre la importància de les visualitzacions en el camp de l'estadística.

La creació de visualitzacions és indispensable per als estadístics, ja que serveix per crear un lligam entre les dades sense processar i els resultats significatius. Les eines visuals ens permeten transformar tota una sèrie de càlculs i desenvolupaments estadístics de caire més teòric, així com passar de dades més complexes, a formes visuals més pròximes i fàcils de pair. Ja sigui per als més experts en la matèria com per aquells usuaris menys coneixedors. Com eines visuals tenen dos funcionalitats clarament diferenciades.

Per una banda, serveixen als estadístics de guia per continuar les seves recerques o per prendre decisions. Per exemple, en els anàlisis de tipus exploratori un diagrama de barres et pot servir per intuir el tipus de distribució que seguiran les dades o, en el cas d'anàlisis més complexos (com ACM-Anàlisi de Components Múltiples o PCA-Anàlisi de Components Principals), et permeten descobrir patrons o relacions entre variables que permeten crear o confirmar hipòtesis.

D'altra banda també tenen la funcionalitat de servir com eina de comunicació. Al tenir un format visual acostuma a ser més accessible per a la major part del públic i permet destacar els elements més importants. Les notícies a la televisió o a la premsa, per exemple, acostumen a anar acompanyades de gràfics per poder arribar a un públic més ample. I als articles científics de tipus estadístic no els pot faltar un gràfic per destacar els punts més importants o interessants.

Cal dir, què a més, no són pocs els elements que s'han de tenir en compte alhora de la seva elaboració. El títol explicarà de forma clara de que van les dades? Els colors o formes destacaran de forma adequada els punts importants a comunicar? Hi haurà massa informació en un gràfic de forma que acabarà sent confús? Els usuaris patiran algun tipus d'afectació visual i com a conseqüències es perdrà informació? El llenguatge dintre de la visualització és correcte a nivell social? Els eixos estaran representats en l'escala correcta? Totes aquestes

qüestions i altres han d'estar presents quan es vol crear una visualització, sobretot si es vol ser curós i es pretén evitar una mala interpretació.

En resum, les visualitzacions en el camp de l'estadística milloren l'accessibilitat, la interpretabilitat i l'impacte causat pels estudis estadístics.

III. METODOLOGIA

En aquest capítol es presenta la metodologia seguida en la realització del treball. Per una banda es detallaran les metodologies que tenen a veure amb l'obtenció de la base de dades i l'elaboració dels anàlisis estadístics. I per una altra la descripció dels recursos informàtics que s'han fet servir per crear les diferents visualitzacions.

S'ha de considerar que en tot moment el procés d'anàlisi va ser fer-se en paral·lel, és adir, tant la part analítica descriptiva com gràfica es feien a la vegada.

3.1 Obtenció de la base de dades

3.1.1 Funcionalitat plataforma SIRENA

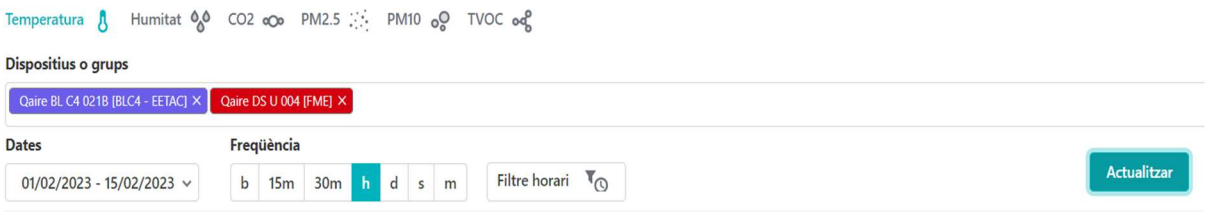
Les dades aquí tractades corresponen als valors mesurats gràcies als sistemes de recollida de dades del projecte *QaireUPC* en diferents aules de l'edifici de la **Facultat de Matemàtiques i Estadística** (FME, Campus Sud). Es va decidir treballar només amb les dades d'aquesta facultat degut a la proximitat amb l'estudiant (entorn habitual d'estudis), i per raons de volum de dades.

Es pretenia tenir una base de dades el més completa possible. Per aquest motiu es van escollir totes les aules de la FME on hi havia aparells que enregistraven mesures i es va considerar el període més llarg pel qual hi havia dades disponibles, en aquest cas, corresponent al període entre **gener de 2021 i febrer de 2023**. S'ha tenir en compte que aquest treball es pretenia entregar en la convocatòria de juny, però malgrat l'entrega final s'ha realitzat en la convocatòria de setembre, es va decidir no ampliar més el rang de dades.

En el moment que es van descarregar les dades de la eina QaireUPC la pàgina encara estava en procés de millora, de fet ho continua estant en el moment de l'entrega d'aquest treball, i no totes les funcionalitats estaven disponibles. Tot i tenir un accés obert per a tothom a la pàgina web del projecte SIRENA, per obtenir les dades d'aquest treball es va haver de demanar una acreditació per fer servir l'aplicació informàtica que en el moment de l'entrega d'aquest treball ja està més actualitzada. Les identificacions les van proporcionar des dels serveis TIC de la UPC.

Un cop es van poden consultar totes les dades, l'aplicació et permetia seleccionar les aules i el tipus de paràmetre o indicador que es vol veure. També et permetia descarregar les dades que haguessis seleccionat (en format **excel**).

Imatge 7: *Panell de selecció de l'aplicació de SIRENA*



Com es pot observar en la imatge del panell de selecció, les aules s'han de seleccionar d'una en una, i d'aquesta manera es van afegint al grup de dispositius que es vol veure. Pel que fa al període de dades, en el moment d'aquest treball, només et permetia seleccionar un rang no superior a un any. Així doncs, per obtenir les dades de tot el període mencionat, es van haver de seleccionar 3 rangs diferents, un per cada any des de 2021 fins al 2023.

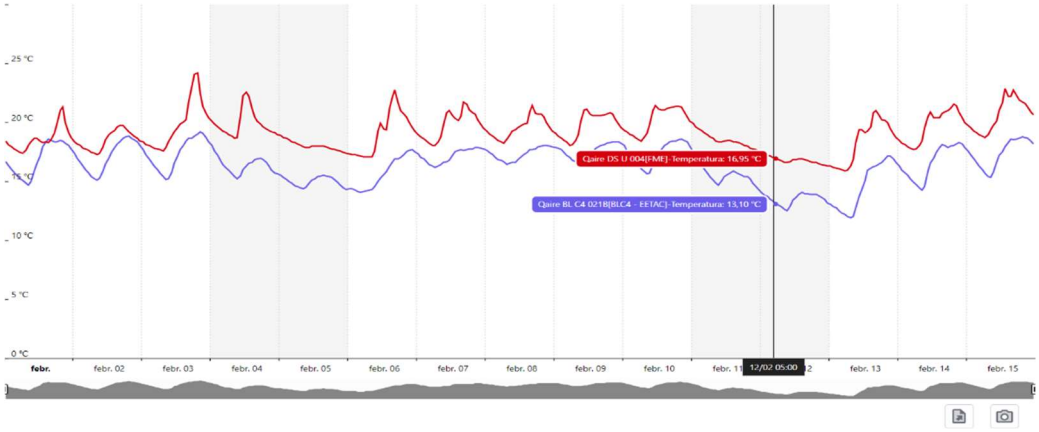
Amb l'objectiu de tenir la base de dades més completa es va decidir agafar la granularitat[5] més petita possible, que en aquest cas era de 15 min. I seleccionar tots els dies en que hi hagués dades, ja que també es permetia la selecció per hores i dies.

3.1.2 Qualitat de l'aire a la plataforma SIRENA

Cal destacar que a la mateixa plataforma hi ha una gran varietat de contingut que s'obté dels diferents aparells instal·lats, mesures com el consum d'electricitat, aigua o gas, estan ben recollides (en la seva majoria aquelles relacionades amb l'electricitat) i representades al llarg de l'aplicació. Però pel que fa a la qualitat de l'aire encara queda feina a fer.

[5] Nivell de detall de la información en una base de dades.

Imatge 8: *Gràfic de línies disponible per a la qualitat d'aire*



La plataforma, encara en procés de millora, no permetia escollir les mides de les partícules recollides de CO_2 (PM_{25} o PM_{10}). I també ofereix menys opcions de visualitzacions de les que ofereixen a altres apartats, com el consum energètic. El gràfic disponible per observar els diferents indicadors de la qualitat de l'aire era un gràfic de línies amb el qual es pot interactuar i veure els valors del que s'ha seleccionat. Al llarg d'aquest treball s'intentarà millor aquest tipus de gràfic, de manera que la selecció de diferents aspectes de les dades pugui ser més eficaç.

3.1.3 Base de dades per a les visualitzacions

Com s'ha comentat prèviament, al descarregar les dades es van obtenir tres arxius diferents d'Excel. Per poder fer el tractament de les dades de forma més còmoda es va decidir treballar cada paràmetre en arxius diferents de tipus R-markdown per fer els anàlisis estadístics de tractament de *NA*, per obtenir les diferents taules dels estadístics principals i per qualsevol manipulació prèvia a la realització dels gràfics, com la creació de variables noves.

Per poder fer les visualitzacions a través de google colab, fent servir Python i la llibreria Altair, es necessitava que la base de dades estigués el més neta possible. Per aquest motiu es van crear tres bases de dades, una per a cada paràmetre estudiat, més adaptada per treballar les visualitzacions.

3.2 Definició del objectius

Lo primer que s'havia de fer, una vegada s'havia decidit la temàtica, era definir una sèrie d'objectius per saber cap a on aniria el projecte.

El procés va ser més llarg de l'esperat, ja que malgrat hi havia una idea del que es volia aconseguir en general, la base de dades era més gran que la majoria amb les quals s'havia treballat fins el moment i contenia una gran quantitat de valors faltants o *missings*. També s'havia de decidir que era el que realment volíem aconseguir amb les visualitzacions.

La definició de com havien de ser les visualitzacions havia d'estar fortament lligada amb els objectius. Què volia saber l'usuari? Voldria fer comparacions? Voldria saber molt o poc detall? Totes aquestes qüestions i algunes més havien de ser considerades per tal de crear bones eines visuals.

Es va establir que lo més important era destacar si les mesures estaven dintre o fora dels llindars adequats de cada indicador per tal de saber si hi havia benestar dintre de les aules. La millor manera de reflectir aquesta situació és mitjançant comparacions però en comptes de decidir des del principi quins gràfics eren els més indicats per això, es va realitzar tot una sèrie de proves al llarg del temps dedicat a aquest treball. Sempre fent reunions amb el tutor per tal d'aclarir si el camí seguit era correcte o si calia reconsiderar algun plantejament.

Per aquest motiu en aquest treball no es trobarà una idea de visualització final, sinó més aviat un seguit de gràfics de prova i error.

3.3 Paràmetres i qualitat de l'aire

Donat que la finalitat del treball era poder definir el benestar dels estudiants a les aules, s'han d'establir els valors dintre dels quals les mesures són òptimes.

3.3.1 Temperatura i Humitat

Pel que fa a la temperatura i segons els informes del IDAE (*Instituto para la Diversificación y Ahorro de Energía*), tenint en compte la norma **UNE EN ISO 8996** i seguint el **Real Decreto 178/2021, de 23 de marzo** (on es fa una revisió de l'original de l'any 2007) del *Reglamento de Instalaciones Térmicas en los Edificios* (RITE).

S'estableix que sota les condicions especifiques d'activitat metabòlica sedentària de 1,2 met[6] i amb grau de vestimenta entre 0,5 clo[7] a l'estiu i 1 clo a l'hivern i un PPD (percentatge de

[6] Met: unitat de mesura de l'energia que desprèn el cos humà i que correspon al metabolisme d'una persona sana, asseguda i sense treballar.

[7] Clo: unitat que mesura el nivell d'aïllament tèrmic de la vestimenta. Un clo = 0 correspon a un home nu y a partir d'una determinada fórmula es calcula la resta de possibles composicions.

persones insatisfetes) < 10%, assumint una velocitat d'aire baix (< 0.1 m/s), les condicions de temperatura i humitat relativa òptimes als interiors serien:

Taula 1: *Valors òptims de benestar a l'interior*

ESTACIÓ	***TEMPERATURA OPERATIVA ° C***	***HUMITAT RELATIVA %***
ESTIU	23 - 25	45 - 60
HIVERN	21 - 23	40 - 50

Com en aquest cas els participants de l'estudi són els estudiants que es troben asseguts i no fan un fort exercici físic, a part de prendre apunts, en la major part del temps que assisteixen a classe, es considerarà que el met d'un estudiant es troba entre 1.2 i 1.3 (el met corresponent a un treball moderat com el que es pot realitzar en oficines). Per tant, es considera que els valors òptims de la taula són els adequats per aplicar a aquest treball i no s'han de modificar.

Com ja s'havia mencionat anteriorment, les condicions per molt per sota o molt per sobre d'aquests valors poden tenir diverses conseqüències. Pel que fa als efectes immediats, les males condicions de temperatura provoquen incomoditat (tant si es té molt fred com molta calor) que poden derivar en poca atenció per part de l'alumnat i professorat (sensació de cansament o somnolència). Les males condicions d'humitat també poden ser un risc per a la salut dintre de les aules, un ambient completament sec pot derivar en irritació o sequedat les mucoses i membranes de la gola. A més, poden provocar sensació d'irritació i sequedat en ulls i pell.

3.3.2 Nivell de CO_2

Pel que fa als nivells òptims de CO_2, farem servir les indicacions que se'ns proporciona a la pàgina web de la QaireUPC dels valors de IDA[8].

[8] IDA: *Calidad del Aire Interior*

Taula 2: IDA, *valors de classificació de qualitat de l'aire segons RITE*

Qualitat aire	*Tipus d'espais*	*Nivell de referència IDA (en ppm)*	*Mitjana exterior (en ppm)*	*Nivell màxim* (en ppm)*
ÒPTIMA. IDA 1	Hospitals, clíniques, laboratoris i guarderies.	350	400	750
BONA. IDA 2	**Oficines**, residències, **sales de lectura**, museus, sales de tribunals, **aules d'ensenyament i assimilables** i piscines.	**500**	**400**	**900**
MITJANA. IDA 3	Edificis comercials, cines, teatres, **sales d'actes**, restaurants, cafeteries, bars, **gimnasos**, locals per a l'esport (llevat de piscines) i **sales d'ordinadors**	800	400	1.200
BAIXA. IDA 4		1.200	400	1.600

Se'ns indica que per als nivells de CO_2 s'accepta un marge de ±100 ppm, degut a les variacions en la concentració del CO_2 exterior i la tolerància de les sondes de medició.

Pel que fa a aquest treball ens interessaren sobretot els nivells de BONA. IDA2, que corresponen a les aules. Tot i que també es té a la base de dades sales d'ordinadors i sales d'actes.

3.4 Mètodes estadístics

En aquest apartat es detallaran els diversos processos estadístics realitzats.

3.4.1 Pre-processament

Pel que fa a la part estadística el primer que calia fer era un pre-processament de les dades, per identificar i/o millorar la qualitat de la base de dades.

Es van haver de modificar els fitxers d'Excel originals de les bases de dades per tal que la lectura fos més còmode. També es van haver d'ajuntar en una mateixa base de dades els tres documents d'Excel que es van obtenir al descarregar les dades (un per cada any 2021, 2022 i 2023).

Donat que els atributs de la base de dades eren les aules on s'havien recollit les mesures i els registres corresponien a cada mesura pressa (una mesura per unitat de temps). No han estat necessàries gaires recodificacions o codificacions de les variables com les que s'havien vist prèviament a les classes. Un exemple el trobem en les enquesta, en que cal codificar els atributs, per exemple, Home = H i Dona = D o Home = 0 i Dona = 1, o correccions tipogràfiques.

Però en aquest cas s'havia de considerar si es volien crear noves variables auxiliars per poder fer les comparacions després. Amb aquest objectiu és van crear variables secundàries que feien referència al moment en que s'havien recollit les dades (dia de la setmana, mes, etc. veure *Annex 1*).

Dintre del pre-processament de les dades també calia revisar el nombre de *missings* de la base de dades. Però donada la importància que aquest tractament ha tingut en aquest treball se'n parlarà en un apartat separat.

3.4.2 Tractament dels valors missings (NA)

Per veure quants missings (NA) hi havia a la base de dades lo primer que es va fer va ser uns gràfics (amb la llibreria *DataExplorer*) fent servir R. D'aquest gràfics (els quals veurem en la part de l'anàlisi estadístic) es va observar que hi havia una quantitat molt gran de valors faltants.

També es va fer un càlcul, per mitja de taules de freqüències, de la quantitat de valors *missings* de la base de dades. El resultat va ser sorprenent ja que els valors eren molt elevats.

Després d'observar els resultats i els comportaments de les dades i de concloure el motiu d'aquest succés (es veurà amb detall a l'anàlisi), s'arriba a la conclusió de que el millor és treballar amb una granularitat més gran. Amb aquest procediment el problema dels NA es veu raonablement reduït.

Fer el tractament d'aquest tipus de dades és necessari i molt important per obtenir un bon anàlisi posterior. Una quantitat molt elevada de NA pot derivar en un anàlisis poc rigorós però

igualment ho serà si s'eliminen les variables o atributs que els contenen, sense pensar en les conseqüències d'estar eliminant també informació necessària de la base de dades.

3.4.3 Taules

En aquest treball les taules més importants són les que descriuran les característiques inicials de les aules a estudiar. Aquestes taules que contindran els estimadors més comuns i ens donaran una idea general de la situació.

Per entendre que es volem dir definirem primer el concepte d'estadístic i estimador:

Definició 1 (Estadístic). *Sigui $X_1,\dots,X_n$ una mostra aleatòria simple de X. Suposant que X pren valors en $X \subseteq \mathbb{R}$. Qualsevol funció.*

$$\begin{aligned} T: \quad & \mathcal{X}^n \quad \to \quad \mathbb{R} \\ & (x_1,\dots,x_n) \mapsto T(x_1,\dots,x_n) \end{aligned}$$

Aplicada a $(X_1,\dots,X_n)$ *és un estadístic:*

$$T_n = T(X_1,\dots,X_n)$$

Definició 2 (Estimador). Quan un estadístic T és utilitzat amb el propòsit d'estimar un paràmetre θ direm que T és un estimador de θ

En el cas d'aquest treball s'entendrà que els valors representats a les taules són estadístics i no estimadors, ja que no s'estimaran els paràmetres poblacionals, sinó que es farà l'anàlisi únicament dels valors observats i es consideraran com mesures descriptives. A continuació es defineixen els estadístics observats que prendrem com mesures de resum:

- *Mínim*: És el valor més petit de entre totes les observacions registrades.

$$Mínim = \text{mín}\,(x_1,\dots,x_n)$$

- *Màxim*: És el valor més gran de entre totes les observacions registrades.

$$Màxim = \text{màx}\,(x_1,\dots,x_n)$$

- *Quantils*: Els quantils són mesures de posició estadística de tipus no central que divideixen una distribució de dades en parts iguals. Els que s'utilitzaran en aquestes taules seran els corresponents als quartils Q_1 i Q_3 .

- Q_1 corresponent al percentil 0.25, és el valor que separa el 25% inferior de les dades del 75% superior.

- Q_3 corresponent al percentil 0.75, és el valor que separa el 75% inferior de les dades del 25% superior.

- *Mediana*: És una mesura de posició central, corresponent al percentil 0.50 i al segon quartil (Q_2). En aquest cas és el valor que separa el 50% de les dades inferiors del 50% superior.

- *Mitjana*: És una mesura de tendència central i es calcula de la següent forma:

$$\bar{x} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)$$

- *Desviació típica*: És una mesura de dispersió o variació de les dades. I és l'arrel quadrada de la variància. La seva fórmula és:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

3.4.4 Outliers o dades atípiques

Els valors atípics o *outliers* són dades que es troben allunyades de la resta de dades d'un mateix grup. Es podria considerar que no segueixen el "patró" de la resta de valors i s'ha d'anar amb compte amb l'anàlisi estadístic que es fa si hi ha moltes dades d'aquest tipus, ja que podrien portar a conclusions errònies.

El tractament que s'acostuma a fer d'aquests valors depèn del tipus d'estudi que es vulgui realitzar. En cas de que es volgués fer algun anàlisi on s'hagués de fer alguna predicció i/o es volgués realitzar algun tipus de regressió, els *outliers* serien susceptibles d'interferir sobre el comportament dels models estadístics estudiats.

Una forma de saber si les dades contenen *outliers* o valors atípics, és fent servir eines visuals com els *boxplots*. Donat que aquest treball pretén centrar-se també en l'àmbit de la visualització de dades es farà servir aquesta eina per a la seva detecció.

S'ha de recordar que per tal de considerar-se un valor atípic el valor ha de complir una de les dues condicions següents:

$$q < Q_1 - 1.5 * IQR \text{ ó } q > Q_3 + 1.5 * IQR$$

on IQR és el Rang InterQuartílic, Q_1 el primer quartil i Q_3 el tercer. En aquest cas estaríem parlat d'un valor atípic dèbil, però es consideraria fort si en comptes de 1.5 es considerés el valor 3, és a dir, 3 vegades la diferència entre Q_1 i Q_3.

En aquest treball no s'ha realitzat cap model lineal però sí es farà un estudi dels *outliers* a nivell descriptiu.

3.4.5 Normalitat

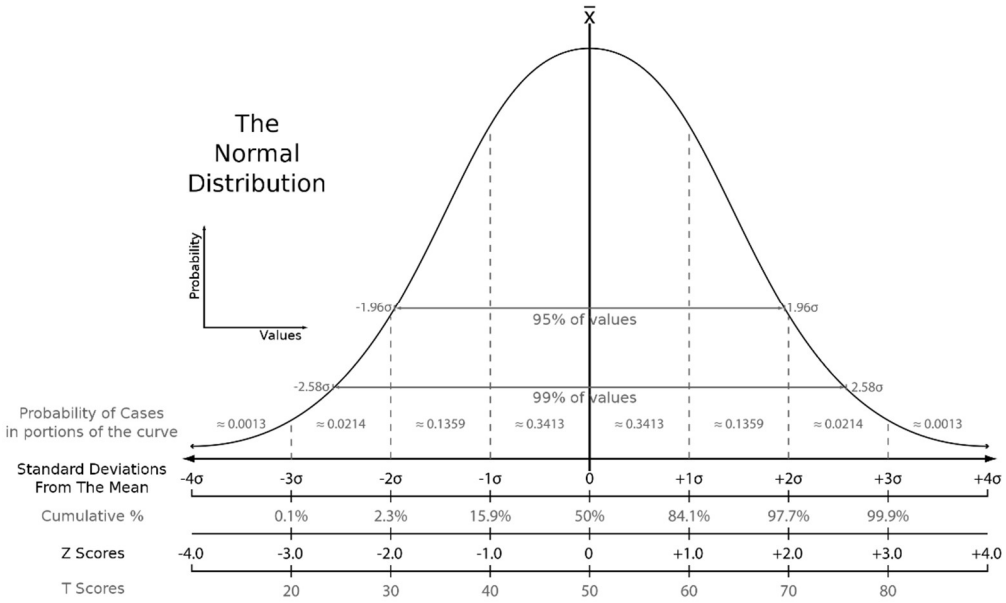
Conèixer la distribució de les dades es un factor important si es volen fer anàlisis, sobretot al modelitzar, ja que acostuma a ser una condició necessària i/o suficient per diferents proves. Tot i així, hi ha moltes eines estadístiques que ens permeten treballar amb dades no normals, com ara els mètodes no paramètrics. En aquest treball farem una comprovació sobre si les dades segueixen aquest tipus de distribució però no serà una condició necessària per al seu desenvolupament.

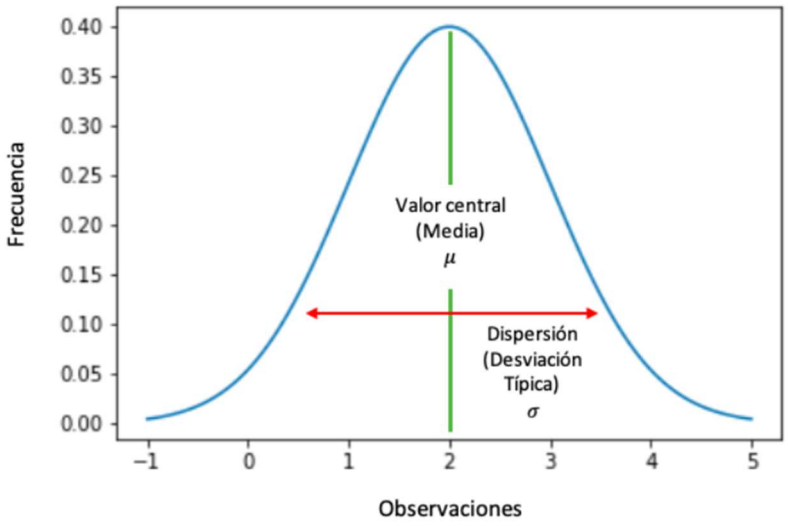
La distribució normal és una distribució de probabilitat continua, amb tendència central i de tipus simètric. Els paràmetres característics de la distribució normal són la mitjana i la variància. La funció de densitat de la normal és:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

On $\mu \in \mathbb{R}$ és el paràmetre de la mitjana (com estimador). I $\sigma^2 \in \mathbb{R} > 0$ és la variància (sent σla desviació típica).

***Imatge* 9:** *Exemple gràfic de la forma d'una distribució normal i els elements associats*





La distribució normal s'expressa com $X \sim N(\mu, \sigma)$. I un cas habitualment utilitzat és el de la distribució normal estàndard, on $\mu = 0$ *i* $\sigma = 1$, $X \sim N(0,1)$.

El mètode que farem servir per la seva comprovació serà de tipus visual majoritàriament, mitjançant *QQPlots* i també amb alguns *Violin Plots*. Però també s'executaran altres proves analítiques. La primera serà el test de Shapiro Whilk i la segona el test de Anderson-Darling.

- **Shapiro Wilk:** es tracta d'una prova per contrastar la normalitat de les dades amb les hipòtesis:

H_0: La mostra prové d'una distribució normal
H_1: La mostra no prové d'una distribució normal

És tracta d'una de les proves més habituals per comprovar la normalitat. Si la hipòtesis nul·la es rebutja, tenim evidències per pensar que la mostra no prové d'una distribució normal. L'estadístic de la prova és $W = \frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$, on $x_{(i)}$és el valor que ocupa la posició i (un cop ordenada la mostra de menor a major), $\bar{x}$ és la mitjana mostral i $a_i = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^2}$ on $m_1, \dots, m_n$ representen els valors mitjos de l'estadístic ordenat i V la matriu de covariàncies del mateix estadístic ordenat. W oscil·la entre 0 i 1 i la hipòtesis nul·la es rebutjarà si W és massa petit.

- **Anderson-Darling**: es tracta d'una prova de bondat d'ajust derivada del test de Kolmógorov-Smirnov que serveix per contrastar si la mostra prové d'una determinada distribució, la diferencia principal és que aquest test dona més pes que el test de Kolmógorov-Smirnov als valors extrems. Les hipòtesis són:

  H_0: La mostra s'ajusta a una distribució específica.
  H_1: La mostra no s'ajusta a la distribució específica.

  En aquest cas es farà servir per comprovar si les dades s'ajusten a una distribució normal. Per una distribució qualsevol la fórmula de l'estadístic de prova prové d'una funció de distribució acumulativa F tal que $A^2 = -n - S, on \quad S = \sum_{i=1}^{n} \frac{(2_i - 1)}{n} [lnF(Y_i) + \ln(1 - F(Y_{n+1-i})]$. Per cada distribució probabilística la funció variarà. En el cas de la normal $F(Y_i) = \Phi([x_i - \bar{x}]/s)$.

3.5 Recursos informàtics

A continuació es detallen breument els recursos informàtics que s'han fet servir en l'elaboració d'aquest treball.

3.5.1 RStudio

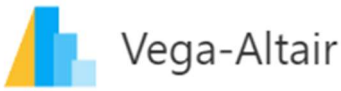
***Imatge* 10:** *Logo de R-Studio*

L'eina principal amb la qual s'ha realitzat la primera part del treball és *RStudio*. Un entorn integrat pel llenguatge de programació R, que serveix com una gran eina en l'anàlisi estadístic. És un programari lliure i actualment és una de les eines informàtiques bàsiques al grau d'estadística. Per aquest treball s'ha fet servir concretament *R-markdown*, que permet integrar tant codi com funcions d'escriptura similar a làtex.

3.5.1 Vega-Altair, Pandas i Python

Les eines principals amb les que s'han treballat les visualitzacions han estat *Vega-Altair* i *Pandas*, les dues són llibreries per *Python*. Aquest és un altre llenguatge de programació que no està contemplat dintre del programari del grau d'estadística, però que s'ha fet servir a les classes de Visualització de la informació juntament amb les llibreries esmentades.

***Imatge* 11:** *Logo de la llibreria Altair*



Pandas és una llibreria especialitzada en la manipulació i l'anàlisi de dades. I Vega-Altair és una llibreria específica per a la creació de visualitzacions.

3.5.2 Google Colab

***Imatge* 12:** *Logo de la eina Google Colab*

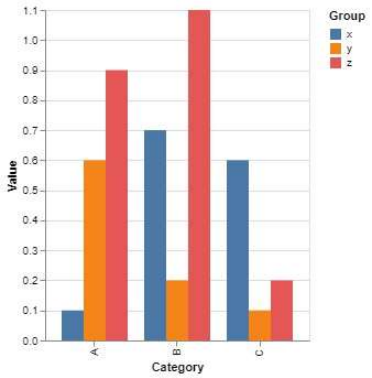
Per poder executar el codi de Python s'ha fet servir la eina Google Colab o "Colaboratory", ja que permet l'execució del codi des del navegador. A més, en quant a aspecte i funcionalitat funciona de manera molt similar a com treballa R-markdown.

3.6Tipus de Visualitzacions

En els següents subapartats es parlarà breument d'algunes de les característiques dels gràfics amb els que s'ha treballat i també si s'acostumen a fer servir en estadística i amb quins objectius. No tots els gràfics aquí descrits han acabat sent els considerats més adients per a les dades d'aquest projecte i alguns altres es detallen perquè formen part d'alguna visualització i no com a elements individuals.

3.6.1 Bar charts

Imatge* 13*: *Bar Chart amb categories*

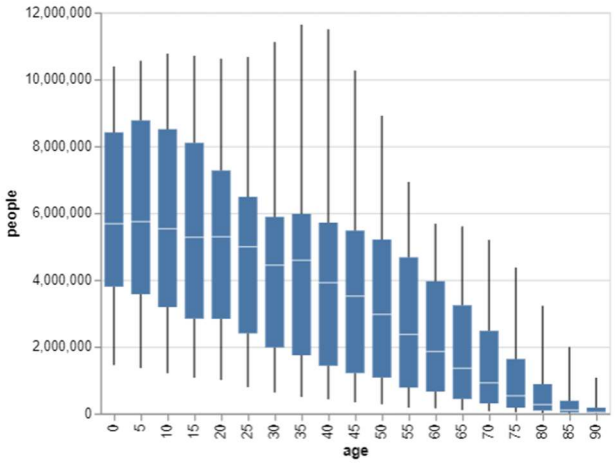


Els *bar charts* o diagrames de barres són gràfics que mostren habitualment la freqüència d'aparició de les observacions o registres. Es pot representar per categories i en estadística serveixen tant per tenir una idea de les freqüències dels valors absoluts, com per intuir quina distribució seguiran les dades. També ens permetrien saber si les dades estan balancejades dintre dels diferents grups.

3.6.2 Boxplot

Els *boxplots* o diagrames de caixes és un tipus de gràfic format, com el propi nom indica, per "caixes" on es representen els estadístics mitjana, quartil Q_1, quartil Q_3 , màxim i mínim. S'acostumen a fer servir per inspeccionar la distribució, la simetria i la dispersió de les dades. També és una eina utilitzada per trobar valors atípics o *outliers*.

***Imatge* 14:** *Boxplot*

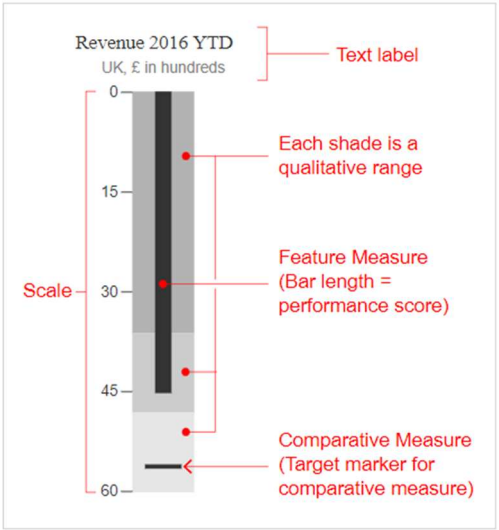


La línia horitzontal blanca mostra la mitjana de les dades, la línia negra vertical representa el valor mínim i el valor màxim corresponent mentre que la forma de la caixa està formada per la diferència entre els quantils 1 i 3 (l'anomenat Rang InterQuartílic- IQR), i dona una idea de la distribució i dispersió de les dades. En cas de trobar *outliers* aquests es representarien fora dels límits de la línia que negra en forma de punts.

3.6.3 Bullet graph

El gràfic *bullet* és una variació dels gràfics de barres creat originalment per Stephen Few, inspirat en la forma dels termòmetres, permet aportar més informació en un únic context visual. Serveix principalment per fer comparacions entre diferents variables o categories. I no s'acostuma a utilitzar en estadística com element individual.

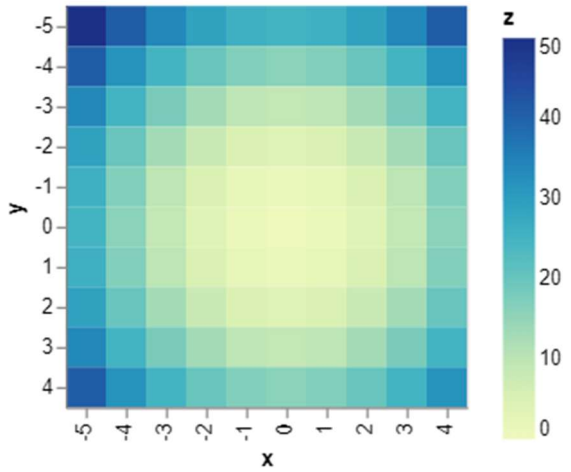
***Imatge* 15:** *Bullet graph i els elements que el formen*



3.6.1 Heatmaps

Els *heatmaps* són gràfics que permeten mostrar comportaments de bases de dades de forma eficient. Mesuren la magnitud d'un fenomen per mitja de colors en dues dimensions. La variació dels colors pot dependre del to o de la intensitat. Habitualment també els podem trobar per representar correlacions entre variables i de fet són eficaços per mostrar relacions entre variables de qualsevol tipus. La forma més freqüent als estudis es la d'un quadrat però realment el podem trobar representat sobre qualsevol superfície (mapes).

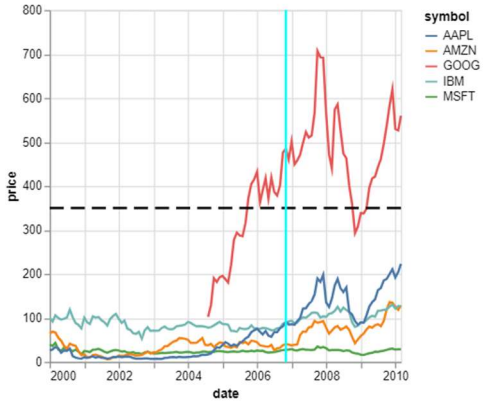
Imatge* 16*: *Heatmap*



3.6.2 Line charts

Els *line charts* són una de les eines més utilitzades per representar dades, ja que permet veure tendències i comportaments al llarg del temps. La seva representació es basa en la unió d'una sèrie de punts (si son temporals de forma ordenada) units per línies. Són les eines bàsiques dels estudis de sèries temporals.

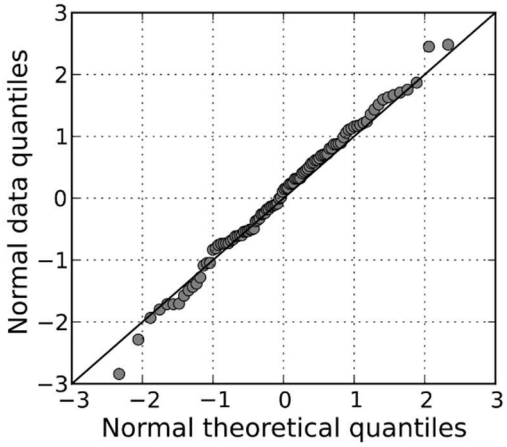
Imatge* 17*: *Line chart*



3.6.3 QQPlot

Els *QQplot* són gràfics específics per fer anàlisis de distribucions de probabilitat de les mostres o poblacions estudiades. Són utilitzats molt especialment per comprovar la normalitat de les dades. La representació es fa per mitjà de punts que representen dels quantils de la distribució calculats a l'eix X.

Imatge* 18*: *QQPlot*



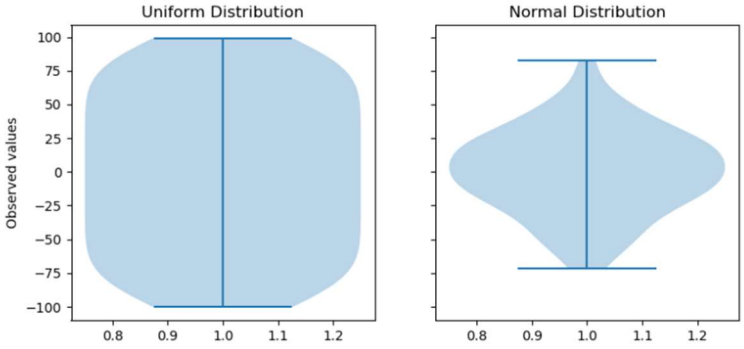
En el cas de la normalitat, si la figura que formen els punts s'aproxima a una recta, podrien dir que la distribució de les dades és normal. S'acostumen a representar amb una línia diagonal que permeti distingir si efectivament les dades s'ajusten a una recta.
Aquesta recta es pot representar de diferents maneres. Una d'elles és fer servir els quantils Q_1 i Q_3 (0.25 i 0.75, respectivament) per tal de traçar la recta a la qual s'haurien d'ajustar les dades.

Una altre forma de crear aquest línia és per mitjà de la recta de regressió de les dades, en aquest cas s'estima la recta a partir d'una estimació lineal per tal de mostrar la relació entre dues variables. Tot i que aquest mètode es fa servir, hi ha qui no el considera el mètode més adequat, ja que no es tracta de punts aleatoris representats on es procedeix a fer un ajust, sinó que els valors del QQPlot han estat ordenats i per tant manipulats. Malgrat això, la interpretació que se'n fa seria la mateixa.

3.6.4 Violin Plot

Els *violin plots* són una combinació dels *boxplot* juntament amb gràfics de densitat, de manera que serveixen per veure la distribució de les dades, com a forma alternativa als gràfics de densitats.

Imatge* 19*: Violin plots



IV. ANÀLISIS ESTADÍSTIC DESCRIPTIU

En aquest apartat es comentaran els anàlisis realitzats juntament amb les taules numèriques corresponents. Tots els processos aplicats a les diferents variables han estat similars en la seva execució, per aquest motiu l'únic que variaria en els apartats separats per variables, si fos el cas, serien els resultats explícits. A més, donat que la base de dades més completa és la referent a l'any 2022 (els motius s'expliquen a l'apartat de *4.2 Tractament de missings*), alguns dels anàlisis s'aplicaran només a aquesta variable.

4.1 Característiques de les variables

A continuació es defineixen en detall les característiques de l'entorn de les variables, juntament amb comentaris i indicacions sobre la seva manipulació per a l'estudi realitzat.

4.1.1 Indicacions al llegir les dades

En un primer moment l'arxiu d'Excel que es descarrega del sistema de SIRENA conté informació referent al període de les dades i valors referents per aules com la mitjana, la mediana, el màxim i el mínim. En aquest moment ja s'observava que no a totes les hores hi havia valors registrats.

Imatge 20: Aspecte de les dades al moment de la descarrega dels registres de Temperatures

Account	UPC	**From**	01/01/2022
Customer	UPCnet	**To**	31/12/2022
		Qaire DS U 004[FME]-Temperatura[ºC]	**Qaire DS U 005[FME]-Temperatura[ºC]**
	AVERAGE	23,38	23,71
	MEDIAN	23,40	23,90
	MAX	31,70	32,10
	MIN	13,60	14,30
Date	**Time**	**Qaire DS U 004[FME]-Temperatura[ºC]**	**Qaire DS U 005[FME]-Temperatura[ºC]**
01/01/2022	00:00	17,30	17,80
01/01/2022	00:15		
01/01/2022	00:30	17,30	17,80
01/01/2022	00:45		
01/01/2022	01:00	17,20	17,80
01/01/2022	01:15		

Aquest fet va fer sospitar que s'haurien de modificar les base de dades, de fet, al llegir els documents amb R el sistema caracteritzava algunes de les aules com variables *logi* (de tipus lògic, referent a valors de TRUE o FALSE). La causa d'aquest fenomen era la gran quantitat de *missings* d'algunes de les aules, ja que no a totes els passaven el mateix. És va haver de modificar la lectura especificant el tipus de variable que eren les aules (numèriques).

Lo primer que es va pensar va ser que potser hi havia un error en la descarrega o que la granularitat no arribava a ser realment de 15 minuts. És va plantejar eliminar les files

corresponents als valors faltants, però després d'una observació detinguda es va arribar a la conclusió de que no tots els registres coincidien en la falta de dades. És a dir, hi havia aules que prenien registres a les 00:15 però d'altres que les prenien a les 00:45, amb intervals de mitja hora, però sense coincidir.

Imatge 21: *Imatge de com la falta de dades era no coincident*

17:30	26,80	
17:45		27,40
18:00	26,90	
18:15		27,30

4.1.2 Propietats de les variables i bases de dades

- **Període d'estudi inicial:** de maig de 2021 a febrer de 2023 (un total de 23 mesos). Aquest període va ser definit agafat el rang més ampli de dades disponibles al sistema.
- **Granularitat inicial:** mesures preses cada 15 minuts.
- **Atributs:** Aules.
- **Registres:** Mesures preses en dies i hores determinades.

Les aules que s'han considerat en aquest treball són totes aquelles que comptaven amb registres en el període designat.

Taula 3: *Taula que mostra les aules de l'estudi*

Aules	Tipus	Planta	Edificis
23	3	3	1
DSU Aula S01	Aula	Planta -1	FME
DSU Aula S02	Aula	Planta -1	FME
DSU Aula S03	Aula	Planta -1	FME
DSU Aula S04	Aula	Planta -1	FME
DSU Aula S05	Aula	Planta -1	FME
DSU Aula 001	Aula	Planta 0	FME
DSU Aula 002	Aula	Planta 0	FME
DSU Aula 003	Aula	Planta 0	FME
DS U 004	Aula	Planta 0	FME
DS U 005	Aula	Planta 0	FME
DSU Aula 7	Aula	Planta 0	FME
DSU PC1	Sala d'Ordinadors	Planta 0	FME
DSU PC2	Sala d'Ordinadors	Planta 0	FME
DSU PC3	Sala d'Ordinadors	Planta 0	FME
DSU 101	Aula	Planta 1	FME
DSU Aula 102	Aula	Planta 1	FME
DSU Aula 103	Aula	Planta 1	FME
DSU Biblioteca	Otro	Planta 1	FME
DSU Sala d'actes	Otro	Planta -1	FME
DSU Sala de juntes	Otro	Planta 0	FME
DSU Sala estudis	Otro	Planta 1	FME
DSU Sala estudis CFIS	Otro	Planta 1	FME
DSU Seminari	Otro	Planta 0	FME

Posteriorment per realitzar les visualitzacions a les aules se'ls posarà una identificació sobre la plata a la qual pertanyen.

Al principi també es van considerar dades de consum però es van descartar per focalitzar en els grups de dades més relacionats amb la qualitat de l'aire. És a dir, les mesures de temperatura, humitat i CO_2.

Taula 4: *Variables principals recollides per QaireUPC*

Variables	Unitats	Categoria
Temperatura	ºC	Principal
Humitat	%	Principal
C02	ppm	Principal

Donat que les mesures havien estat preses durant períodes de màxim un any i amb gran freqüència, es va decidir crear noves variables per tal de veure si al segmentar per aquestes noves variables hi havia algun patró diferenciat.

Taula 5: *Variables secundàries derivades de l'estudi*

Variables	Unitats
Setmanes	Dies
Caps de setmana	Dies
Festius	Dies
Quatrimestres	Dies(mesos)

Malgrat aquestes van ser les variables que es van considerar inicialment, al final algunes es van descartar i es van crear altres, de forma que:

- **Setmanes** es van agafar de forma completa, integrant els caps de setmana. I anat de dilluns a diumenge.
- Els dies **festius** com a tal finalment no es van estudiar per falta de temps.
- Es va crear una variable **Tipus** que dividia els dies segons "setmana" i "cap de setmana".
- **Data** de recollida de les dades es va separar per poder treballar millor. Creant les variables **Hora**, **Mes** i **Dia.**
- La variable **Quadrimestre** es categoritzar a partir de la data i segons els mesos.
    - Q1: Primer quadrimestre de setembre a desembre.
    - Q2: Segon quadrimestres de febrer a maig.
    - Exàmens: Època general d'exàmens al gener i juny.
    - Estiu: Mesos amb menys estudiants juliol i agost.

Després de considerar el perquè de la gran quantitat de valors *missings* (tema que es discutirà en l'apartat 4.2), es va acabar concloent que lo millor era crear diferents bases de dades corresponents als anys. On la base de dades més completa la trobarem associada a l'any 2022 i per tant serà aquesta base de dades la que farem servir majoritàriament per l'anàlisi.

La base de dades completa, considerant els tres indicadors i abans de ser tractada, estava formada per **1.523.355** registres.

A les taules següents podem observar el nombre de *missings* per aula de cada variable i també el nombre de valors vàlids.

Taula 6: *Nombre de missings per aula i variable*

	Temperatura	Humitat	CO2	Total missings
Aules	NA1	NA2	NA3	NAT
Qaire DS U 004	28694	28694	28696	86084
Qaire DS U 005	28591	28591	28593	85775
Qaire DS U 101	28458	28458	28460	85376
Qaire DSU Aula 001	36853	36853	36855	110561
Qaire DSU Aula 002	37217	37217	37219	111653
Qaire DSU Aula 003	36845	36845	36847	110537
Qaire DSU Aula 102	37290	37290	37292	111872
Qaire DSU Aula 103	37321	37321	37323	111965
Qaire DSU Aula 7	37263	37263	37265	111791
Qaire DSU Aula S02	37441	37441	37443	112325
Qaire DSU Aula S03	37411	37411	37413	112235
Qaire DSU Aula S04	37369	37369	37371	112109
Qaire DSU Aula S05	37196	37196	37198	111590
Qaire DSU Biblioteca	37583	37583	37585	112751
Qaire DSU PC1	37361	37361	37363	112085
Qaire DSU PC2	37447	37447	37449	112343
Qaire DSU PC3	37526	37526	37528	112580
Qaire DS U S01	28278	28278	28280	84836
Qaire DSU Sala d'actes	37433	37433	37435	112301
Qaire DSU Sala de juntes	37412	37412	37414	112238
Qaire DSU Sala estudis	37433	37433	37435	112301
Qaire DSU Sala estudis CFIS	37536	37536	37538	112610
Qaire DSU Seminari	37405	37405	37407	112217
Totals	**823363**	**823363**	**823409**	**2470135**

Com es pot observar el nombre de valors faltants de les variables sembla ser gairebé el mateix, excepte per algunes desenes de valors cosa que podria indicar que podria haver passat alguna una falla del sistema de recollida. Tenint en compte que el total de valors faltants és de 2.470.135, el **percentatge de *missings* d'aquesta base de dades és del 61.85%**, més del 50% de la base da dades, cosa que suggereix una actuació immediata contra aquest problema. Les tres aules que recullen menys valors NA són la Aula 004, l'aula 005 i la 101, les dues primeres de la planta 1 i l'altre de la planta 0.

En la *taula 7*, podem veure observar el nombre de registres vàlids de la base de dades. En aquest cas no hi ha cap diferència entre el nombre de valors de les diferents variables, per tant, es descartaria que el sistema hagués deixat de funcionar (en cas que hagués fallat només en la recollida d'alguna variable i no de forma completa, ja que en aquest cas no ho podríem saber) i que l'origen de les dades que falten és un altre. Les aules amb més registres serien la S01 i la 101, seguides de l'aula 004 i l'aula 005. La diferència de valors en aquest cas queda explicada pel fet que els dispositius no es van col·locar a totes les aules a la vegada. La seva adaptació va ser progressiva.

Taula 7: *Nombre total de valors vàlids per aula i variable*

	Temperatura	Humitat	CO2	Total vàlid
Aules	N1	N2	N3	NT
Qaire DS U 004	29182	29182	29182	87546
Qaire DS U 005	29285	29285	29285	87855
Qaire DS U 101	29418	29418	29418	88254
Qaire DSU Aula 001	21023	21023	21023	63069
Qaire DSU Aula 002	20659	20659	20659	61977
Qaire DSU Aula 003	21031	21031	21031	63093
Qaire DSU Aula 102	20586	20586	20586	61758
Qaire DSU Aula 103	20555	20555	20555	61665
Qaire DSU Aula 7	20613	20613	20613	61839
Qaire DSU Aula S02	20435	20435	20435	61305
Qaire DSU Aula S03	20465	20465	20465	61395
Qaire DSU Aula S04	20507	20507	20507	61521
Qaire DSU Aula S05	20680	20680	20680	62040
Qaire DSU Biblioteca	20293	20293	20293	60879
Qaire DSU PC1	20515	20515	20515	61545
Qaire DSU PC2	20429	20429	20429	61287
Qaire DSU PC3	20350	20350	20350	61050
Qaire DS U S01	29598	29598	29598	88794
Qaire DSU Sala d'actes	20443	20443	20443	61329
Qaire DSU Sala de juntes	20464	20464	20464	61392
Qaire DSU Sala estudis	20443	20443	20443	61329
Qaire DSU Sala estudis CFIS	20340	20340	20340	61020
Qaire DSU Seminari	20471	20471	20471	61413
Totals	**507785**	**507785**	**507785**	**1523355**

Una vegada es va fer el tractament dels *missings* la quantitat total de registres era de 813.606, el 53.4% de les dades originals.

4.2Tractament dels valors *missings*

L'aspecte de la base de dades respecte els *missings* va ser sobtant des del principi. A la part de les visualitzacions es poden veure els gràfics on queda visible quina era la situació, a part del que hem comentat amb les taules.

En aquest treball es van optar ver veure el comportament de les dades, observant les hores a les quals hi havia registres i després d'un procés una senzill però llarg d'observació, es va arribar a la conclusió que les dades no es prenien totes a la mateixa hora perquè hi havia una component aleatòria, probablement derivada del funcionament dels aparells. Així, hi havia aules que seguien un patró de començar a recollir dades a les 00:15h i després cada mitja hora, i es mantenia d'aquesta manera durant dos dies (l'anomenarem període d'estabilitat), però el tercer dia començava a recollir dades a les 23:00h i així cada mitja hora durant els pròxims dos dies, al cap de tercer podia tornar a canviar el moment de recollida de les dades. Tenint com a conseqüència que dintre de diversos dies el nombre mesures preses per hores varies.

Aquest tipus de comportament era diferent per cada aula, algunes podien arribar a tenir períodes estables de fins i tot 7 dies, però sense seguir un patró clar, per exemple no importava la planta on estigués l'aula. Malgrat la plataforma ens prometia un granularitat cada 15 minuts, la realitat era que les dades es prenien cada 30 minuts i quedaven quarts d'hora sense informació.

Per poder solucionar aquest fet es va decidir recalcular les dades en base a la hora. Per cada aula es va fer servir la funció *aggregate* de R per fer la mitja dels valors de cada hora del dia i el mes concrets, tal com es mostra al codi de la imatge. Amb aquest procediment es va assegurar que els valors recalculats no fossin gaire diferents als originals, ja que dintre d'una mateixa hora les mesures no diferien molt, però entre diferents hores ja hi havia canvis més grans.

Imatge 22: *Forma de recalcular el valors*

```
P_1_AulaS01_h21 <- aggregate( `Qaire DS U S01[FME]-Humitat[%]` ~ h + d + dia + mes,
                              data = DSUAulaS01_h21, FUN = mean)
```

És va fer servir la mitjana ja que semblava la opció que deixa els valors menys modificats i perquè en qüestió d'estimadors és tracta d'un dels més robust. També es podria haver escollit algun altre estadístic com ara el mínim o el màxim. Aquest mètode va permetre mantenir la base de dades el més similar possible a la original, tot i que els nombre de valors a estudiar es va reduir a l pràcticament la meitat.

El problema dels *missing* estava pràcticament solucionat, excepte pel fet de que a no tots els anys hi havia el mateix nombre de dades. Per tant, intentar unificar tota la base de dades completa tornava a crear una quantitat important de *missings*. Per aquesta raó es va decidir mantenir els *dataframes* per separat. A més, en el cas de l'any 2021 les aules que comptaven amb aparells van ser totes a la vegada i hi havia diferències grans de registres per aules. També

es va decidir la creació de *dataframes* intentat que les aules tinguessin els registres més similars possible i evitar d'aquesta manera tornar a tenir molts *missings*.

L'any 2021 les aules que tenien més registres eren l'aula 004 i la 005 de la planta 0, l'aula S01 de la planta -1 i l'aula 101 de la planta 1. A continuació les aules 001 i 003 de la planta 0, i finalment la resta d'aules. Aquesta agrupació probablement ve donada per les aules on es van instal·lar primer els aparells fins arribar a totes les aules tractades en aquest treball.

Als apartats següents es detallaren els anàlisis de les dades una vegada les dades ja havien estat tractades o, en el cas dels primers estadístics, sense tenir en compte els valors NA.

4.3Temperatures

4.3.1 Estadístics principals

Els estadístics principals es van obtenir de la base de dades completa sense modificar, però obviant els valors *missings.* Els resultats en el cas de les temperatures els veiem a la taula resum següent:

Taula 8: *Estadístics l'any 2021*

	Min	Q1	Median	Mean	SD	Q3	Max
Qaire DS U 004[FME]-Temperatura[ºC]	15.3	21.5	25.30	24.28	3.46	27.10	29.5
Qaire DS U 005[FME]-Temperatura[ºC]	15.9	21.9	25.40	24.50	3.48	27.30	29.9
Qaire DS U 101[FME]-Temperatura[ºC]	13.0	22.2	26.90	25.43	4.07	28.70	31.8
Qaire DSU Aula 001[FME]-Temperatura[ºC]	11.8	17.8	19.00	18.86	1.88	19.70	24.5
Qaire DSU Aula 002[FME]-Temperatura[ºC]	14.2	17.8	18.90	18.71	1.29	19.50	22.9
Qaire DSU Aula 003[FME]-Temperatura[ºC]	11.9	17.7	18.90	18.83	1.86	19.70	24.0
Qaire DSU Aula 102[FME]-Temperatura[ºC]	14.2	17.7	18.90	18.73	1.34	19.50	23.8
Qaire DSU Aula 103[FME]-Temperatura[ºC]	13.8	17.7	18.90	18.72	1.32	19.50	23.6
Qaire DSU Aula 7[FME]-Temperatura[ºC]	14.2	17.7	18.80	18.69	1.29	19.45	22.7
Qaire DSU Aula S02[FME]-Temperatura[ºC]	13.9	17.7	18.80	18.58	1.22	19.30	22.4
Qaire DSU Aula S03[FME]-Temperatura[ºC]	13.9	17.7	18.80	18.62	1.25	19.40	22.6
Qaire DSU Aula S04[FME]-Temperatura[ºC]	13.9	17.7	18.75	18.58	1.24	19.30	22.6
Qaire DSU Aula S05[FME]-Temperatura[ºC]	14.0	17.7	18.70	18.60	1.24	19.30	22.4
Qaire DSU Biblioteca[FME]-Temperatura[ºC]	13.9	17.8	18.90	18.68	1.27	19.40	22.5
Qaire DSU PC1[FME]-Temperatura[ºC]	14.3	17.7	18.80	18.68	1.30	19.40	23.3
Qaire DSU PC2[FME]-Temperatura[ºC]	14.3	17.7	18.90	18.70	1.30	19.40	23.2
Qaire DSU PC3[FME]-Temperatura[ºC]	14.4	17.7	18.70	18.60	1.24	19.30	23.0
Qaire DS U S01[FME]-Temperatura[ºC]	15.3	21.0	24.90	23.68	2.93	26.10	28.3
Qaire DSU Sala d'actes[FME]-Temperatura[ºC]	14.5	17.7	18.80	18.61	1.27	19.30	22.9
Qaire DSU Sala de juntes[FME]-Temperatura[ºC]	14.3	17.7	18.80	18.65	1.27	19.40	22.5
Qaire DSU Sala estudis[FME]-Temperatura[ºC]	14.1	17.7	18.75	18.59	1.25	19.30	22.7
Qaire DSU Sala estudis CFIS[FME]-Temperatura[ºC]	13.9	17.7	18.80	18.61	1.26	19.40	23.0
Qaire DSU Seminari[FME]-Temperatura[ºC]	14.3	17.7	18.70	18.61	1.27	19.30	22.8

Observem que al 2021, malgrat la majoria d'aules ronden els 14ºC-15ºC en temperatures mínimes. L'aula 003, de la planta 0, presenta valors de 11.9ºC. Pel contrari l' aula DS0005 té una temperatura mínima més elevada que la mitja. DSU101 té una temperatura màxima bastant per sobre de la resta de sales. En moltes aules veiem que la temperatura mitja ronda uns 4º per sota dels valors recomanats de temperatura, però sense arribar a ser molt extremes. La temperatura màxima ronda els 30ºC a l'aula 005.

Taula 9: *Estadístics l'any 2022*

	Min	Q1	Median	Mean	SD	Q3	Max
Qaire DS U 004[FME]-Temperatura[ºC]	13.6	20.0	23.4	23.38	4.14	26.8	31.7
Qaire DS U 005[FME]-Temperatura[ºC]	14.3	20.2	23.9	23.71	4.11	26.9	32.1
Qaire DS U 101[FME]-Temperatura[ºC]	13.0	20.7	24.2	24.43	4.66	28.7	33.9
Qaire DSU Aula 001[FME]-Temperatura[ºC]	14.5	20.6	23.9	23.84	3.98	27.0	31.7
Qaire DSU Aula 002[FME]-Temperatura[ºC]	13.9	19.7	23.5	23.33	4.15	26.5	31.6
Qaire DSU Aula 003[FME]-Temperatura[ºC]	13.8	19.9	23.2	23.37	4.03	26.8	31.4
Qaire DSU Aula 102[FME]-Temperatura[ºC]	13.2	20.4	23.7	24.07	4.58	28.0	33.8
Qaire DSU Aula 103[FME]-Temperatura[ºC]	13.2	19.7	23.6	23.77	4.81	28.0	33.5
Qaire DSU Aula 7[FME]-Temperatura[ºC]	15.1	20.6	23.5	23.39	3.24	26.3	29.5
Qaire DSU Aula S02[FME]-Temperatura[ºC]	14.0	20.7	23.3	23.17	3.14	25.9	29.4
Qaire DSU Aula S03[FME]-Temperatura[ºC]	16.3	20.6	23.0	23.22	3.13	26.3	28.9
Qaire DSU Aula S04[FME]-Temperatura[ºC]	13.8	20.5	23.6	23.28	3.35	26.0	29.8
Qaire DSU Aula S05[FME]-Temperatura[ºC]	14.8	20.2	23.6	23.08	3.32	25.9	29.1
Qaire DSU Biblioteca[FME]-Temperatura[ºC]	13.1	21.6	24.6	24.41	3.82	27.2	32.6
Qaire DSU PC1[FME]-Temperatura[ºC]	14.2	20.6	23.9	23.86	4.03	27.2	32.0
Qaire DSU PC2[FME]-Temperatura[ºC]	13.8	19.9	23.6	23.59	4.27	27.3	31.8
Qaire DSU PC3[FME]-Temperatura[ºC]	14.9	19.8	23.1	22.91	3.51	25.9	30.0
Qaire DS U S01[FME]-Temperatura[ºC]	16.4	20.4	23.0	23.06	3.19	25.8	29.5
Qaire DSU Sala d'actes[FME]-Temperatura[ºC]	16.4	19.9	22.9	22.71	3.15	25.4	28.7
Qaire DSU Sala de juntes[FME]-Temperatura[ºC]	12.5	18.6	22.2	22.58	4.37	26.5	31.4
Qaire DSU Sala estudis[FME]-Temperatura[ºC]	15.1	20.6	23.8	23.73	3.87	26.7	32.0
Qaire DSU Sala estudis CFIS[FME]-Temperatura[ºC]	11.9	20.3	24.0	24.28	4.62	28.1	35.1
Qaire DSU Seminari[FME]-Temperatura[ºC]	13.2	18.8	22.9	23.13	4.89	27.5	33.0

En aquest cas s'observa que les temperatures mitges van augmentar l'any 2022, destacant que a la sala d'estudis es va arribar a prop dels 35ºC, 10ºC per sobre de la temperatura òptima a l'estiu. Però la mitja general ens mostra que la temperatura es troba dintre dels límits òptims.

Taula 10: *Estadístics l'anys 2023*

	Min	Q1	Median	Mean	SD	Q3	Max
Qaire DS U 004[FME]-Temperatura[ºC]	13.8	15.8	17.1	17.29	1.97	18.6	23.1
Qaire DS U 005[FME]-Temperatura[ºC]	14.8	16.6	18.0	18.44	2.38	20.0	25.1
Qaire DS U 101[FME]-Temperatura[ºC]	13.6	15.7	17.2	17.16	1.82	18.5	21.7
Qaire DSU Aula 001[FME]-Temperatura[ºC]	14.4	16.4	17.2	17.33	1.35	18.2	22.8
Qaire DSU Aula 002[FME]-Temperatura[ºC]	13.7	16.0	17.2	17.30	1.81	18.1	24.6
Qaire DSU Aula 003[FME]-Temperatura[ºC]	14.0	16.0	17.1	17.04	1.38	18.2	20.8
Qaire DSU Aula 102[FME]-Temperatura[ºC]	13.3	15.8	17.1	16.99	1.63	18.1	21.2
Qaire DSU Aula 103[FME]-Temperatura[ºC]	12.8	15.3	16.3	16.32	1.44	17.4	20.1
Qaire DSU Aula 7[FME]-Temperatura[ºC]	15.4	17.3	18.1	18.23	1.38	19.1	23.7
Qaire DSU Aula S02[FME]-Temperatura[ºC]	14.5	18.4	19.3	19.32	1.25	20.1	25.0
Qaire DSU Aula S03[FME]-Temperatura[ºC]	16.9	18.3	18.9	18.92	0.82	19.5	22.0
Qaire DSU Aula S04[FME]-Temperatura[ºC]	17.2	18.7	19.4	19.69	1.47	20.6	25.8
Qaire DSU Aula S05[FME]-Temperatura[ºC]	17.5	18.7	19.6	19.74	1.36	20.6	23.7
Qaire DSU Biblioteca[FME]-Temperatura[ºC]	14.7	16.8	19.0	18.76	2.19	20.6	22.8
Qaire DSU PC1[FME]-Temperatura[ºC]	14.7	16.1	17.4	17.32	1.48	18.5	21.4
Qaire DSU PC2[FME]-Temperatura[ºC]	13.8	15.8	17.0	16.94	1.50	18.0	21.3
Qaire DSU PC3[FME]-Temperatura[ºC]	14.8	16.6	17.4	17.48	1.26	18.3	21.4
Qaire DS U S01[FME]-Temperatura[ºC]	17.5	18.4	19.0	19.13	1.03	19.7	23.6
Qaire DSU Sala d'actes[FME]-Temperatura[ºC]	17.3	18.4	19.0	18.95	0.91	19.3	27.1
Qaire DSU Sala de juntes[FME]-Temperatura[ºC]	12.7	15.5	16.4	16.98	2.48	17.6	27.5
Qaire DSU Sala estudis[FME]-Temperatura[ºC]	13.8	16.1	18.7	18.26	2.32	20.1	23.7
Qaire DSU Sala estudis CFIS[FME]-Temperatura[ºC]	13.1	15.7	16.8	16.96	1.73	18.3	22.9
Qaire DSU Seminari[FME]-Temperatura[ºC]	12.7	15.0	16.0	15.79	1.27	16.6	22.1

En aquest cas les temperatures només serien comparables als mesos corresponents de 2022. Com es tracta de dades de gener i febrer, observem que les temperatures màximes són bastant baixes que les de l'any 2022 i per tant la mitja també es veu reduïda.

4.3.2 Segmentacions

S'han fet diverses taules amb les segmentacions que permeten les variables per tal de veure si hi ha algun element destacable.

En aquest cas ens fixem que tot i que els caps de setmana s'esperaria que la temperatura de les aules fos més baixa, ja que no acostumen a haver-hi tants alumnes i per tant no aporten el seu calor corporal, les temperatures continuen sent bastant similars a les temperatures de la setmana. Què no hi hagi tanta distinció podria ser indicador de que l'activitat humana no afecta tant a l'augment de la temperatura i per tant seria més fàcil poder controlar-la mitjançant els aparells d'aire condicionat o calefaccions per tal de que es mantinguin en les temperatures idònies. O també podria indicar que la quantitat d'esdeveniments que tenen lloc a la facultat els caps de setmana (si hi ha) són equiparables a l'activitat diària.

Taula 11: *Temperatura els caps de setmana*

	Min	Q1	Median	Mean	SD	Q3	Max
P_1_AulaS01	16.70	20.00	22.75	22.94	3.26	25.90	29.50
P_1_AulaS02	14.15	20.15	22.55	22.81	3.27	25.74	29.35
P_1_AulaS03	16.70	20.20	22.40	23.01	3.25	26.35	28.90
P_1_AulaS04	16.30	19.90	22.90	22.91	3.44	25.90	29.80
P_1_AulaS05	16.40	19.90	22.75	22.76	3.34	25.80	29.10
P_1_SalaActes	16.80	19.90	22.70	22.65	3.16	25.40	28.70
P0_Aula001_t22	14.70	19.70	23.65	23.43	4.23	26.95	31.70
P0_Aula002	14.15	19.20	23.00	23.01	4.41	26.60	31.55
P0_Aula003	14.50	19.20	22.90	23.03	4.26	26.70	31.40
P0_A004	14.20	19.20	22.80	23.01	4.39	26.85	31.70
P0_A005	14.60	19.70	23.70	23.49	4.32	27.10	32.10
P0_Aula7	16.05	19.95	22.90	23.07	3.43	26.30	29.45
P0_PC1	14.55	19.80	23.55	23.50	4.26	27.20	31.95
P0_PC2	14.05	19.30	23.05	23.22	4.47	27.40	31.80
P0_PC3	15.55	19.40	22.60	22.65	3.61	25.65	29.10
P0_SalaJuntes	12.55	18.35	21.80	22.37	4.53	26.50	31.40
P0_Seminari	13.20	18.45	22.80	22.96	4.97	27.30	32.80
P1_Aula101	13.00	19.81	24.10	24.01	4.94	28.70	33.70
P1_Aula102	13.65	19.55	23.65	23.75	4.85	28.20	33.60
P1_Aula103	13.60	18.90	23.30	23.35	5.04	28.00	33.40
P1_Biblioteca	14.70	20.10	23.50	23.72	4.29	27.25	32.30
P1_SalaEstudis	16.05	19.65	23.65	23.48	4.14	26.82	31.80
P1_SalaEstudisCFIS	11.95	19.80	23.90	23.98	4.77	28.02	33.95

Taula 12: *Temperatura els dies de setmana*

	Min	Q1	Median	Mean	SD	Q3	Max
P_1_AulaS01	16.45	20.50	23.10	23.11	3.15	25.80	29.40
P_1_AulaS02	14.05	20.90	23.45	23.31	3.07	25.90	29.20
P_1_AulaS03	16.35	20.70	23.10	23.31	3.08	26.30	28.70
P_1_AulaS04	13.95	20.80	23.90	23.43	3.30	25.95	29.70
P_1_AulaS05	15.70	20.40	23.90	23.21	3.31	25.90	29.00
P_1_SalaActes	16.70	19.95	22.90	22.73	3.15	25.41	28.50
P0_Aula001_t22	14.50	21.00	23.98	24.01	3.86	27.00	31.70
P0_Aula002	14.00	20.00	23.55	23.46	4.04	26.45	31.45
P0_Aula003	13.85	20.20	23.30	23.50	3.93	26.80	31.30
P0_A004	13.60	20.35	23.55	23.53	4.02	26.70	31.55
P0_A005	14.30	20.50	24.00	23.80	4.02	26.75	32.00
P0_Aula7	15.10	20.90	23.60	23.53	3.15	26.30	29.40
P0_PC1	14.20	20.90	23.95	24.01	3.92	27.20	31.90
P0_PC2	13.80	20.25	23.80	23.74	4.18	27.30	31.80
P0_PC3	14.90	20.00	23.25	23.02	3.47	25.95	30.00
P0_SalaJuntes	12.55	18.80	22.55	22.66	4.30	26.50	31.10
P0_Seminari	13.20	19.05	22.95	23.20	4.86	27.55	33.00
P1_Aula101	13.40	21.20	24.35	24.60	4.53	28.75	33.90
P1_Aula102	13.25	20.90	23.70	24.20	4.47	27.90	33.80
P1_Aula103	13.25	20.00	23.60	23.94	4.71	28.00	33.50
P1_Biblioteca	14.40	22.25	24.85	24.69	3.58	27.15	32.55
P1_SalaEstudis	15.15	20.95	23.85	23.83	3.74	26.70	32.00
P1_SalaEstudisCFIS	12.05	20.55	24.10	24.40	4.55	28.10	35.05

Les temperatures al primer quadrimestre (al voltant del 23ºC) de l'any 2022 van estar 2ºC per sota de les registrades al segon quadrimestres quadrimestre (al voltant dels 21ºC).

Cal d'estacar que l'aula on s'han observat valors més elevats en la majoria de segmentacions ha estat la sala d'estudis CFIS, però també és la que registre els valors més baixos en quant a temperatures mínimes.

Finalment també es va fer una segmentació que no es va considerar en crear les variables. Es tracta de la diferència entre els horaris lectius i els no lectius de la setmana.

Taula 13: *Temperatures horari lectiu*

	Min	Q1	Median	Mean	SD	Q3	Max
P_1_AulaS01	16.85	18.88	20.46	20.14	1.63	21.43	23.35
P_1_AulaS02	14.85	18.70	20.77	20.27	2.07	21.80	23.75
P_1_AulaS03	17.00	19.65	20.90	20.64	1.50	21.77	23.75
P_1_AulaS04	16.60	17.94	20.08	19.82	1.99	21.20	25.80
P_1_AulaS05	15.70	17.60	19.40	19.50	2.35	20.45	27.90
P_1_SalaActes	17.00	17.88	18.96	18.93	1.15	19.76	22.35
P0_Aula001_t22	15.10	19.44	21.25	20.76	1.91	22.10	24.45
P0_Aula002	14.50	19.20	20.20	20.22	1.90	21.39	24.75
P0_Aula003	14.45	19.75	20.85	20.80	2.02	22.11	24.70
P0_A004	14.25	19.64	20.85	20.42	1.99	21.65	23.95
P0_A005	15.00	20.15	20.98	20.77	2.01	22.01	24.50
P0_Aula7	15.65	21.25	22.05	21.90	1.82	23.06	25.20
P0_PC1	14.80	20.45	21.40	21.18	1.89	22.45	25.55
P0_PC2	14.35	19.39	20.75	20.66	2.23	22.16	25.20
P0_PC3	15.60	19.40	20.30	20.09	1.43	21.00	23.30
P0_SalaJuntes	14.05	18.25	19.23	19.43	1.99	20.95	25.70
P0_Seminari	13.80	16.95	18.65	18.51	1.95	19.70	21.90
P1_Aula101	14.15	21.05	22.30	21.88	2.32	23.21	25.75
P1_Aula102	13.90	20.74	21.70	21.51	2.15	22.92	25.97
P1_Aula103	14.10	19.54	20.50	20.46	1.91	21.80	24.65
P1_Biblioteca	16.10	22.45	23.80	23.19	2.28	24.75	26.35
P1_SalaEstudis	16.25	21.29	21.92	21.89	1.62	23.15	24.50
P1_SalaEstudisCFIS	14.05	20.20	21.00	20.86	1.92	22.01	24.60

Es consideren:

- **Horaris lectius:** 8h del matí fins a les 20h de la tarda i només dies de setmana.
- **Horaris no lectius:** la resta d'hores del dia i també els caps de setmana.

Amb aquestes condicions s'observa que la mitjana de temperatures en horari lectiu ronda els 20ºC-21ºC, molt propers al límits inferior òptim a l'hivern. Mentre que als horaris no lectives aquesta temperatura mitja es veu augmentada prop de 2ºC. Sembla que les temperatures màximes són més elevades quan no es fa classe.

Taula 14: *Temperatures horari NO lectiu*

	Min	Q1	Median	Mean	SD	Q3	Max
P_1_AulaS01	16.70	20.00	22.80	22.96	3.26	25.90	29.50
P_1_AulaS02	14.15	20.15	22.60	22.85	3.27	25.75	29.35
P_1_AulaS03	16.70	20.25	22.50	23.04	3.25	26.40	28.90
P_1_AulaS04	16.30	19.95	22.90	22.95	3.43	25.95	29.80
P_1_AulaS05	16.40	19.90	22.90	22.80	3.33	25.80	29.10
P_1_SalaActes	16.80	19.90	22.75	22.66	3.16	25.40	28.70
P0_Aula001_t22	14.70	19.75	23.70	23.49	4.22	27.00	31.70
P0_Aula002	14.15	19.20	23.15	23.05	4.38	26.60	31.55
P0_Aula003	14.50	19.25	22.90	23.07	4.24	26.70	31.40
P0_A004	14.20	19.25	22.90	23.07	4.37	26.90	31.70
P0_A005	14.60	19.70	23.70	23.53	4.30	27.10	32.10
P0_Aula7	16.05	20.00	22.90	23.13	3.42	26.40	29.45
P0_PC1	14.55	19.85	23.70	23.56	4.25	27.25	31.95
P0_PC2	14.05	19.35	23.10	23.27	4.45	27.40	31.80
P0_PC3	15.55	19.40	22.70	22.69	3.62	25.70	29.10
P0_SalaJuntes	12.55	18.35	21.90	22.41	4.54	26.55	31.40
P0_Seminari	13.20	18.45	22.80	23.01	4.97	27.31	32.80
P1_Aula101	13.00	19.85	24.05	24.07	4.91	28.75	33.70
P1_Aula102	13.65	19.60	23.65	23.81	4.83	28.20	33.60
P1_Aula103	13.60	18.90	23.35	23.43	5.04	28.00	33.40
P1_Biblioteca	14.70	20.20	23.60	23.82	4.27	27.30	32.30
P1_SalaEstudis	16.05	19.75	23.60	23.53	4.12	26.90	31.80
P1_SalaEstudisCFIS	11.95	19.85	23.90	24.04	4.76	28.10	33.95

4.3.3 Normalitat

Es va fer un primer intentat amb les dades per aules separades amb R. Amb Shapiro Wilk la mostra era massa gran i les dades no eren concloents, per aquest motiu es va intentar prendre una mostra de 5000 dades. Els resultats indicaven que les temperatures no seguien una distribució normal. S'ha de tenir en compte que els test es van fer d'una de les aules concretes, però ens serveix com a indicador de la distribució de la resta de dades, com es veurà en les visualitzacions. Sempre considerant un 0.05 de confiança.

```
##  Shapiro-Wilk normality test
##
## data:  sample_data
## W = 0.8815, p-value < 2.2e-16
```

El test de Anderson-Darling va confirmar la mateixa situació, que la distribució de les dades no era normal.

```
        Anderson-Darling normality test

data:  humi22_df$P_1_AulaS01
A = 38.339, p-value < 2.2e-16
```

4.4 Humitat

4.4.1 Estadístics principals

Taula 15: *Estadístics d'humitat any 2022*

	Min	Q1	Median	Mean	SD	Q3	Max
Qaire DS U 004[FME]-Humitat[%]	16	46	51	50.56	7.83	56	75
Qaire DS U 005[FME]-Humitat[%]	25	46	51	50.37	7.29	55	71
Qaire DS U 101[FME]-Humitat[%]	19	41	46	46.32	7.85	52	66
Qaire DSU Aula 001[FME]-Humitat[%]	20	41	47	46.46	7.33	52	67
Qaire DSU Aula 002[FME]-Humitat[%]	21	45	50	49.20	7.03	54	73
Qaire DSU Aula 003[FME]-Humitat[%]	13	44	50	49.18	7.53	55	70
Qaire DSU Aula 102[FME]-Humitat[%]	21	41	46	46.39	7.83	52	69
Qaire DSU Aula 103[FME]-Humitat[%]	25	43	47	47.72	7.18	53	69
Qaire DSU Aula 7[FME]-Humitat[%]	12	43	50	48.78	9.30	56	69
Qaire DSU Aula S02[FME]-Humitat[%]	17	44	51	50.74	10.06	60	74
Qaire DSU Aula S03[FME]-Humitat[%]	16	45	51	51.06	9.20	59	68
Qaire DSU Aula S04[FME]-Humitat[%]	8	44	51	50.80	9.86	59	73
Qaire DSU Aula S05[FME]-Humitat[%]	12	45	51	51.23	9.74	60	75
Qaire DSU Biblioteca[FME]-Humitat[%]	13	40	46	45.62	8.21	52	68
Qaire DSU PC1[FME]-Humitat[%]	15	42	47	46.74	7.18	52	68
Qaire DSU PC2[FME]-Humitat[%]	11	43	48	47.62	6.82	52	71
Qaire DSU PC3[FME]-Humitat[%]	15	45	52	50.48	8.33	57	72
Qaire DS U S01[FME]-Humitat[%]	20	46	52	52.12	9.29	60	73
Qaire DSU Sala d'actes[FME]-Humitat[%]	24	44	51	51.53	10.03	60	76
Qaire DSU Sala de juntes[FME]-Humitat[%]	21	46	52	51.62	8.00	58	72
Qaire DSU Sala estudis[FME]-Humitat[%]	17	41	47	46.54	8.67	53	77
Qaire DSU Sala estudis CFIS[FME]-Humitat[%]	10	41	46	45.93	7.55	51	67
Qaire DSU Seminari[FME]-Humitat[%]	25	44	49	49.40	7.49	54	76

S'observa com en quant a la mitjana els nivells d'humitat es troben dins dels valors òptims. Tampoc han canviat gaire entre els anys. La zona amb menys humitat és la biblioteca en mitja. I el valor més baix el registrat el trobem a l'aula S04.

4.4.2 Normalitat

En aquest cas al igual que passava amb la temperatura, el test de shapiro no es va poder realitzar de forma decuada i els resultats diferien. En aquest cas el test de Anderson-Darling ens diu que no teníem prou evidències per descartar que segueixi una distribució normal.

```
        Anderson-Darling normality test

data:  humi22_df$P_1_AulaS01
A = 38.339, p-value < 2.2e-16
```

4.5 Nivell de Co2

4.5.1 Estadístics principals

Taula 16: *Estadístics de CO2*

	Min	Q1	Median	Mean	SD	Q3	Max
Qaire DS U 004[FME]-Qualitat aire (CO2)[]	322	404.00	434	497.17	216.15	495	4685
Qaire DS U 005[FME]-Qualitat aire (CO2)[]	315	402.00	430	514.86	289.58	485	4559
Qaire DS U 101[FME]-Qualitat aire (CO2)[]	311	395.00	423	471.90	186.55	470	3040
Qaire DSU Aula 001[FME]-Qualitat aire (CO2)[]	288	434.00	463	530.69	216.31	511	2822
Qaire DSU Aula 002[FME]-Qualitat aire (CO2)[]	0	435.75	473	563.42	264.36	565	4565
Qaire DSU Aula 003[FME]-Qualitat aire (CO2)[]	0	430.00	462	517.25	169.35	530	3162
Qaire DSU Aula 102[FME]-Qualitat aire (CO2)[]	287	430.00	457	495.06	168.90	497	3507
Qaire DSU Aula 103[FME]-Qualitat aire (CO2)[]	272	436.00	473	531.10	219.12	529	5000
Qaire DSU Aula 7[FME]-Qualitat aire (CO2)[]	343	435.00	460	479.28	89.18	492	3388
Qaire DSU Aula S02[FME]-Qualitat aire (CO2)[]	263	441.00	477	554.99	263.88	543	5000
Qaire DSU Aula S03[FME]-Qualitat aire (CO2)[]	322	426.00	456	501.31	186.31	502	4369
Qaire DSU Aula S04[FME]-Qualitat aire (CO2)[]	296	442.00	474	519.81	174.87	526	3429
Qaire DSU Aula S05[FME]-Qualitat aire (CO2)[]	324	435.00	463	511.26	177.01	517	4335
Qaire DSU Biblioteca[FME]-Qualitat aire (CO2)[]	307	444.00	471	488.22	81.11	510	1408
Qaire DSU PC1[FME]-Qualitat aire (CO2)[]	276	435.00	473	524.00	196.17	537	3435
Qaire DSU PC2[FME]-Qualitat aire (CO2)[]	0	443.00	487	554.34	216.75	573	2820
Qaire DSU PC3[FME]-Qualitat aire (CO2)[]	348	432.00	455	492.28	151.42	486	2869
Qaire DS U S01[FME]-Qualitat aire (CO2)[]	272	363.00	395	498.46	332.44	469	5000
Qaire DSU Sala d'actes[FME]-Qualitat aire (CO2)[]	346	437.00	461	509.44	236.53	495	5000
Qaire DSU Sala de juntes[FME]-Qualitat aire (CO2)[]	344	430.00	451	475.49	100.26	482	1907
Qaire DSU Sala estudis[FME]-Qualitat aire (CO2)[]	317	441.00	476	529.34	163.53	552	2753
Qaire DSU Sala estudis CFIS[FME]-Qualitat aire (CO2)[]	314	437.00	471	508.37	140.36	525	1797
Qaire DSU Seminari[FME]-Qualitat aire (CO2)[]	290	435.00	469	543.18	271.31	522	5000

Pel que fa als nivells de CO2, els valor mitjans es troben dintre dels rangs de bona IDA, nivell òptim. Però observem que els valors màxims de totes les aules són bastant més elevats dels límits adequats Fins i tot, arribant a ser indicadors de bones condicions de CO2.

4.5.2 *Segmentacions*

Taula 17: *Hores lectives*

	Min	Q1	Median	Mean	SD	Q3	Max
P_1_AulaS01	342.0	416.25	458.25	570.05	263.69	629.12	1990.0
P_1_AulaS02	393.5	463.50	498.50	567.91	163.58	636.12	1515.5
P_1_AulaS03	409.0	464.50	487.50	633.93	370.16	574.25	2390.5
P_1_AulaS04	403.5	476.00	530.00	557.48	112.28	593.25	1107.5
P_1_AulaS05	404.0	469.25	511.50	607.71	228.85	666.00	1708.5
P_1_SalaActes	387.0	443.88	472.75	514.31	181.52	505.75	2030.0
P0_Aula001_c22	376.0	452.38	482.25	513.52	108.53	526.88	1130.5
P0_Aula002	389.5	478.12	538.00	611.45	198.16	678.62	1409.0
P0_Aula003	409.5	496.38	567.75	650.09	219.73	752.38	1509.0
P0_A004	374.0	452.25	495.75	537.33	168.69	545.50	1509.5
P0_A005	372.0	451.88	536.00	569.78	143.88	664.25	1107.0
P0_Aula7	398.0	454.88	515.25	598.91	221.43	694.50	2022.5
P0_PC1	379.5	486.00	544.00	568.35	131.61	606.75	1382.0
P0_PC2	404.5	509.00	593.25	705.65	336.13	772.00	2733.5
P0_PC3	396.0	440.75	458.50	499.69	158.11	492.88	1818.0
P0_SalaJuntes	361.5	446.38	479.00	511.79	106.51	542.62	1094.5
P0_Seminari	403.5	474.62	507.50	608.74	237.22	648.25	1977.0
P1_Aula101	366.0	464.75	497.50	545.10	153.56	564.38	1365.0
P1_Aula102	398.5	477.88	509.50	525.26	73.04	565.12	799.0
P1_Aula103	379.5	487.12	563.25	732.18	426.49	780.25	2898.5
P1_Biblioteca	379.0	460.88	485.25	487.84	43.22	515.62	597.0
P1_SalaEstudis	425.5	501.88	554.25	596.57	150.89	626.62	1210.0
P1_SalaEstudisCFIS	382.0	474.88	511.50	518.96	73.19	551.25	882.0

Un cop més on hi ha més diferències és en les hores lectives i no lectives. Els valors mitjans són més elevats durant les classes que quan no hi ha, com s'esperaria per la presencia dels alumnes i professors.

Taula 18: *Hores no lectives*

	Min	Q1	Median	Mean	SD	Q3	Max
P_1_AulaS01	302.0	355.50	376.00	437.78	217.29	411.0	3683.0
P_1_AulaS02	269.5	425.00	447.50	481.34	229.48	481.5	5000.0
P_1_AulaS03	340.5	407.50	428.50	463.42	177.20	457.5	4073.0
P_1_AulaS04	310.5	429.00	453.00	469.97	102.69	483.5	1908.5
P_1_AulaS05	347.0	419.00	440.00	469.61	189.68	465.0	4158.0
P_1_SalaActes	359.5	428.00	446.00	470.00	134.72	467.5	2266.5
P0_Aula001_c22	307.5	419.00	439.50	454.25	88.61	465.5	1766.0
P0_Aula002	226.0	419.00	445.50	493.24	166.16	490.0	2043.0
P0_Aula003	213.0	413.00	437.00	469.75	136.93	467.5	2751.0
P0_A004	344.0	397.00	414.50	437.71	108.69	440.5	2203.5
P0_A005	350.0	396.00	412.00	433.74	96.94	435.0	1688.5
P0_Aula7	364.0	429.00	450.50	461.86	56.95	478.0	1209.5
P0_PC1	323.5	414.00	439.00	461.94	122.31	470.5	2436.5
P0_PC2	220.5	417.00	447.50	478.00	136.03	486.0	1883.0
P0_PC3	363.0	423.50	443.00	457.87	101.40	465.5	2254.0
P0_SalaJuntes	354.5	419.00	437.50	442.65	50.26	455.0	1082.5
P0_Seminari	292.0	414.50	440.00	452.33	101.16	467.0	3364.5
P1_Aula101	325.0	387.00	403.50	423.27	98.58	427.0	1659.0
P1_Aula102	302.5	414.00	430.84	445.00	82.03	455.0	1352.0
P1_Aula103	291.0	415.50	437.00	450.77	72.65	470.0	1296.5
P1_Biblioteca	320.0	435.38	452.50	456.02	41.95	473.0	1080.5
P1_SalaEstudis	324.5	425.50	448.50	475.41	100.43	481.0	1199.5
P1_SalaEstudisCFIS	325.5	421.50	445.00	476.84	132.19	476.5	1737.0

4.5.3 Outliers

Els *outliers* d'aquesta variable es comentaran directament amb l'ajuda dels gràfics. On queda visible la quantitat de valors extrems que conté.

4.5.4 Normalitat

Una vegada més el dos test realitzats ens indiquen no podem afirmar que les dades siguin normals.

```
        Anderson-Darling normality test

data:  tempe22_df$P_1_AulaS01
A = 98.04, p-value < 2.2e-16
```

V. VISUALITZACIONS

En aquests capítol es mostraran totes les visualitzacions fetes al llarg d'aquest treball. Algunes d'aquestes s'han aplicat només a la base de dades pels motius comentats al capítol IV, on es comenta que la base de dades més completa és la de l'any 2022. Tot i que s'ha intentat que totes les bases de dades tinguessin algun tipus de representació.

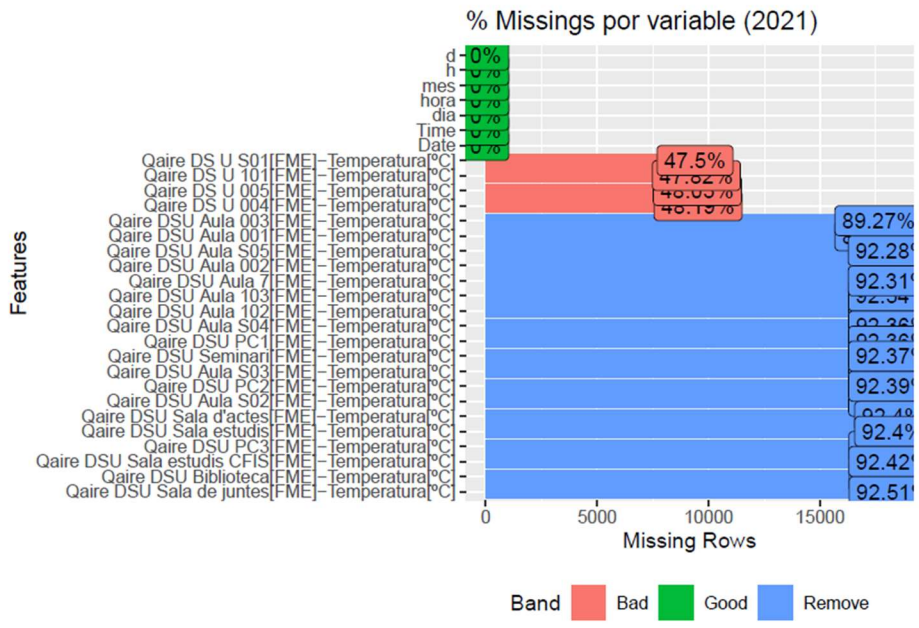
Moltes de les visualitzacions que es mostraran aquí, seran complementaries als resultats de l'anàlisi estadístic. No es posaran totes les que s'han creat però aquestes es poden consultar al *link* de GitHub que podeu trobar l'*Annex 1*.

5.1 Temperatures

5.1.1 Valors missings

Les primeres visualitzacions que es van fer van ser les de anàlisis de valors faltants, tot i que en aquest treball es comentin primer les taules, van ser aquests gràfics els que realment van donar la veu d'alarma sobre la quantitat elevada de valors inexistents.

Gràfic 1: *Gràfic de barres dels missings l'anys 2021*



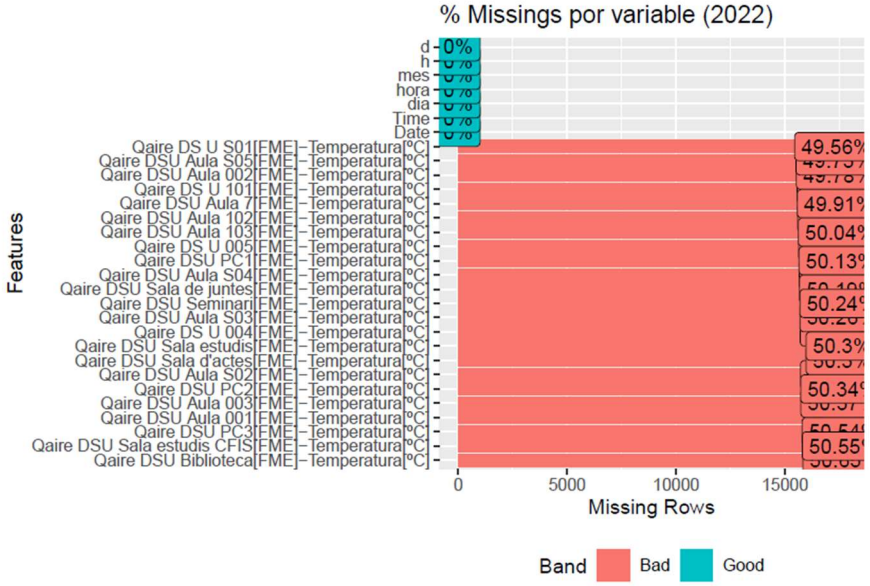
Aquesta primera visualització està generada directament per una funció de R, de la llibreria DataExplorer, que ens permet veure de forma ràpida quants valors *missings* hi ha als diferents

atributs. Es tracta d'un gràfic de barres que no només et mostra el percentatge de *missings* sinó que et dona una indicació, basada en ens paràmetres que pots definir, sobre si les variables s'haurien de treure o si la quantitat de missings és bona o dolenta. En aquest cas els paràmetres són els que venien per defecte.

Com es pot observar la majoria de les aules superen el 80% de dades faltants, algunes poques estan al voltant del 50% i

Al gràfic corresponent de l'any 2022 veiem que la situació millora però el contingut de dades faltants es troba encara al voltant del 50%. Per l'any 2023 observem el mateix tipus de comportament, per aquest motiu no s'inclou el gràfic.

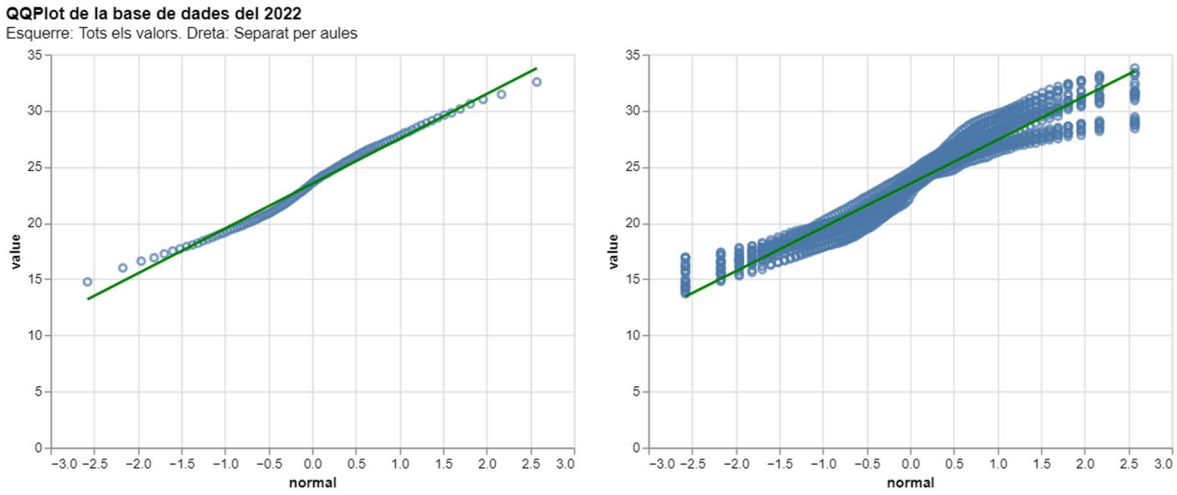
Gràfic 2: *Gràfic de barres dels missings l'anys 2022*



5.1.2 Normalitat

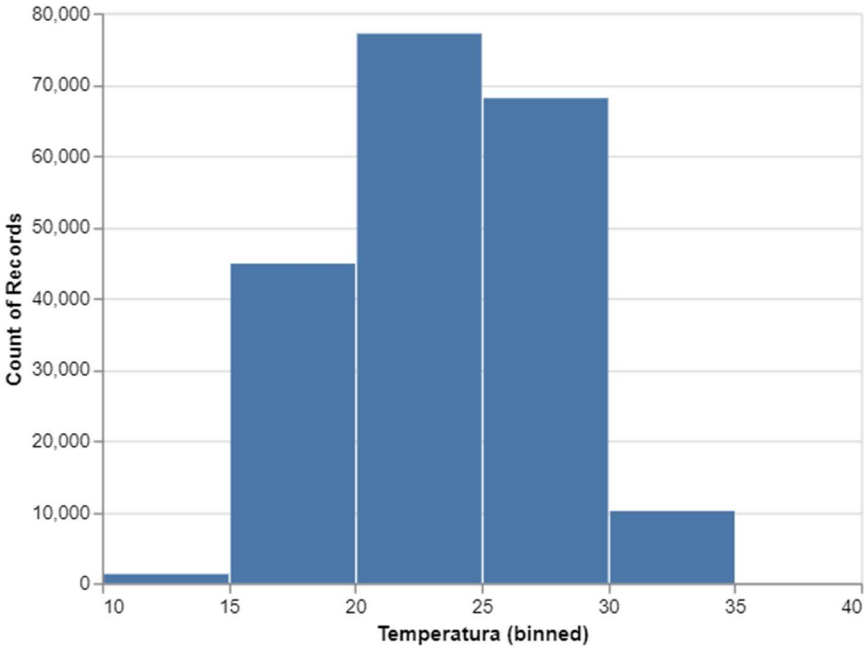
Pel poder comprovar la normalitat de les dades es van ser servir diferents gràfics. El més destacat és el QQPlot. La línia en verd es va crear mitjançant un procés de regressió. El que podem observar és que no es pot assegurar que els valors segueixin una distribució normal. Al gràfic esquerre podem veure que graficats tots els valors mentre a la dreta es separen per aules.

Gràfic 3: *Gràfics de temperatures i normalitat*



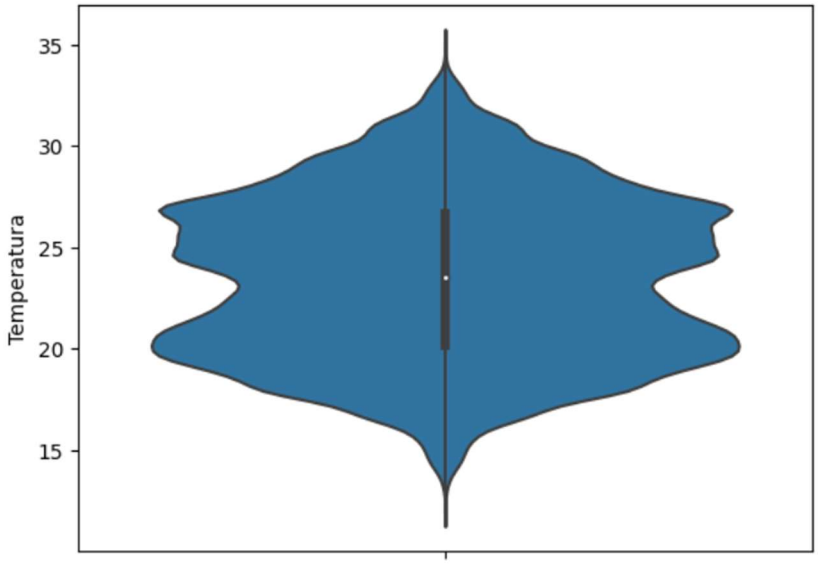
Si s'observa el boxplot es veu que els valors centrals sembla que sí segueixen aquesta distribució per els extrems no.

Gràfic 4. *El boxplot de temperatures*



També es van generar violin plots per comprovar la densitat, el conjunt de gràfics amb les aules separades es pot trobar al document notebook de python i aquí és on es pot apreciar millor que els valors no acaben d'adaptar-se a una normal.

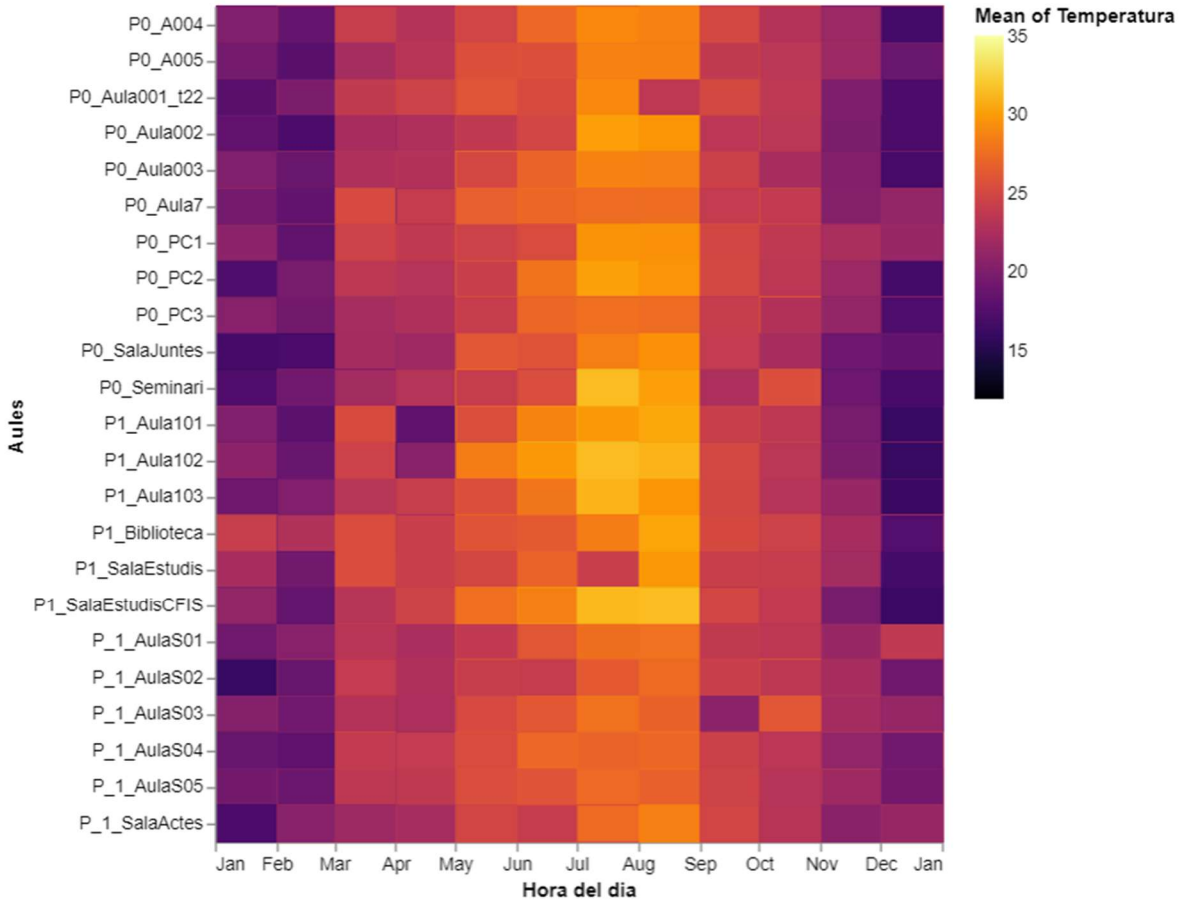
Gràfic 5: *Violin plot temperatura*



5.1.3 Heatmaps

Un gràfic que ens permet veure clarament la tendència de les temperatures és el heatmpa.

Gràfic 6: *Heatmap de temperatures*

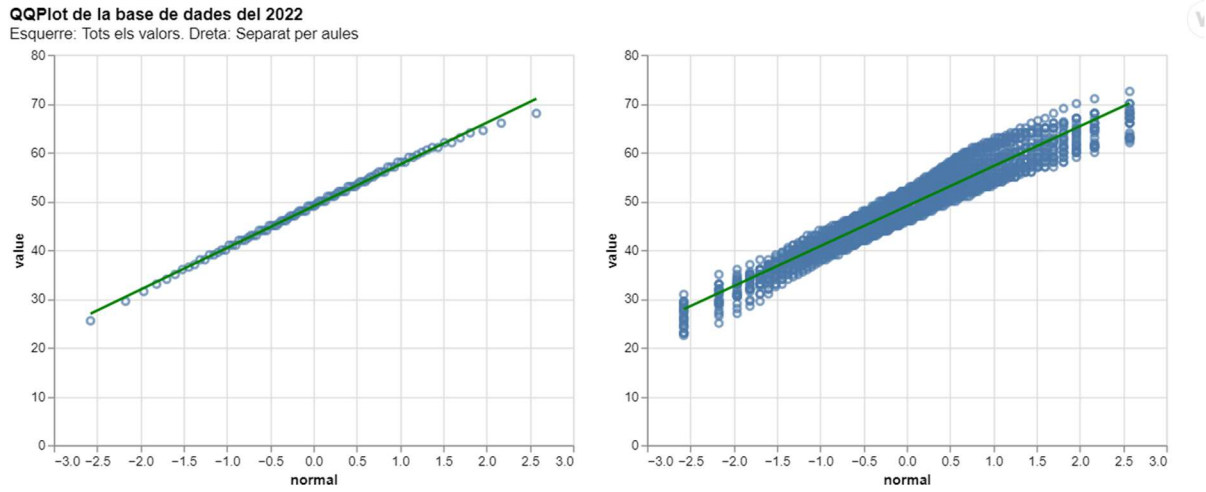


5.2 Humitat

5.2.1 Normalitat

En el cas de la humitat al contrari que el test de Anderson-Darling, ens suggereix que el conjunt de dades si pertany a una distribució normal.

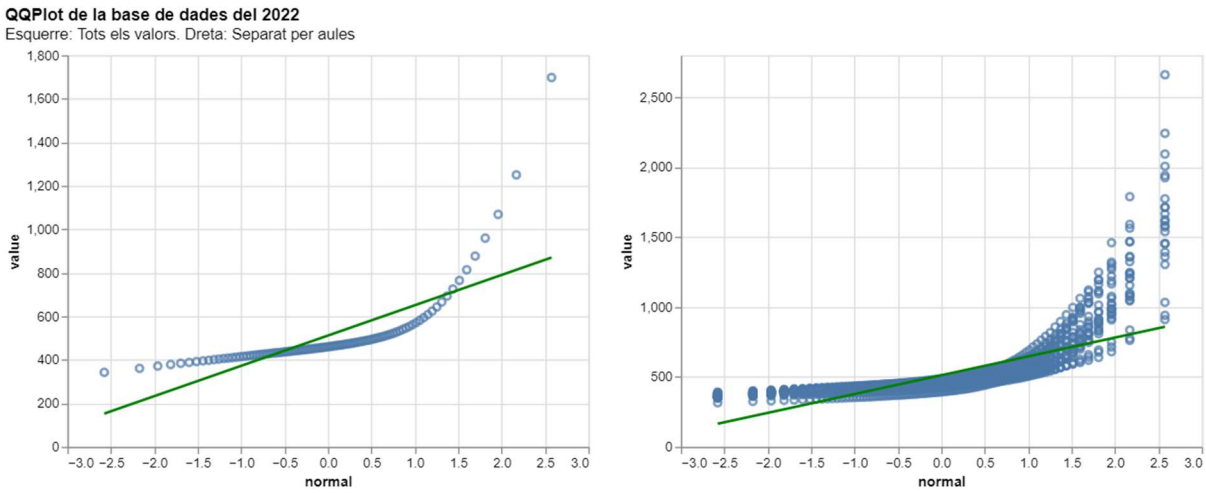
Gràfic 7: *QQPlot humitat*



5.3 Nivells de CO2

5.3.1 Normalitat

Gràfic 8:QQPlot Co2



Amb aquest gràfics és descarta completament la idea de que els valors segueixin una distribució normal.

5.4 Visualització més completa

La visualització més completa creada, ha estat el gràfics de línies que mostra les temperatures al llarg del temps.

El gràfic és manipulable de tal manera que et mostra les aules seleccionades i es pot localitzar una data en concret. Marcant un àrea en el gràfic de sota. És una eina que permet veure de forma ràpida si els valors estan per sobre o per sota o dins dels òptims marcats amb línies de punts horitzontals.

Gràfic 9: *Temperatures al llarg del temps*

VI. CONCLUSIONS

Com a conclusions d'aquest treball cal dir que s'ha pogut veure que en quant a les dades de C02 hi ha molts pics que es superen de lluny els valors òptims i caldria estudiar-ho amb més detall.

També s'ha pogut anar veient que els gràfics han servit de suport durant tot l'estudi, servint per a que indicar el quines coses havia de mirar. Al veure les taules amb tants números algú es podria perdre, però les eines visuals que mostren els pics ajuden a saber on buscar. Però sense deixar de banda l'anàlisi més analític.

També cal destacar que ha estat un repte treballar amb una quantitat tan elevada de dades i amb eines que no havia fet servir tant durant el curs.

En resum podem dir que hi ha moltes aules que no compleixen els valors òptims en quant a mesures puntuals malgrat la mitjana de les dades indiqués una altra cosa. I tot i que la visualització final no era la que esperava, s'ha demostrat que sense visualitzacions la informació que es transmet és menys clara.

VII. BIBLIOGRAFIA

Ajuntament de Barcelona. *Qualitat de l'aire.* https://ajuntament.barcelona.cat/qualitataire/es

Altair (2023). https://altair-viz.github.io/#

Vicepresidencia Tercera del Gobierno (Retrieved September 2, 2023). *Ministerio para la Transición Ecológica y el Reto Demográfico* - Rite - Reglamento instalaciones térmicas en los edificios. (n.d.) Energia.gob.es. https://energia.gob.es/Eficiencia/RITE/Paginas/InstalacionesTermicas.aspx

Comunitat UPC Sostenible. *CampusLab*. https://sostenible.upc.edu/ca/campus-lab

Generalitat de Catalunya (Gencat, 2023). *Vols saber que respires?* Medi Ambient i Sostenibilitat. https://mediambient.gencat.cat/ca/05_ambits_dactuacio/atmosfera/qualitat_de_laire/vols-saber-que-respires/

Nunez, C. (2019, February 4). *Air Pollution Causes, Effects, and Solutions*. Environment. https://www.nationalgeographic.com/environment/article/air-pollution

Projecte SIRENA. (data d'obtenció de les dades febrer 2023). Accés obert a tot el públic. https://serveistic.upc.edu/ca/sirena

Projecte SIRENA. Accés amb compte d'usuari. https://app.dexma.com/dashboard/widgets.htm

QaireUPC. https://sostenible.upc.edu/ca/ca/qaireupc

WHO. (2021, September 22). *Ambient (outdoor) air quality and health*. Who.int; World Health Organization: WHO.https://www.who.int/en/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

Vázquez, Pere-Pau. *Diapositives i material de classe*. Visualització de la Informació.

VIII. ANNEXOS

Annex 1: Script de R per carregar les dades i definir les noves variables

```
# afegir variables
#Lectura de dades TEMPERATURA

tempe21 <- read_excel("TEMPERATURE_01-01-2021_31-12-2021.xls", col_names = TRUE,
col_types = c("date","date",rep("numeric",23)))
tempe22 <- read_excel("TEMPERATURE_01-01-2022_31-12-2022.xls", col_names = TRUE)
tempe23 <- read_excel("TEMPERATURE_01-01-2023_01-02-2023.xls", col_names = TRUE)

attach(tempe21)
attach(tempe22)
attach(tempe23)

#Lectura de dades HUMITAT

humi21 <- read_excel("HUMIDITY_01-05-2021_31-12-2021.xls",col_names = TRUE,
col_types = c("date","date",rep("numeric",23)))
humi22 <- read_excel("HUMIDITY_01-01-2022_31-12-2022.xls", col_names = TRUE)
humi23 <- read_excel("HUMIDITY_01-01-2023_01-02-2023.xls", col_names = TRUE)
attach(humi21)
attach(humi22)
attach(humi23)

#Lectura de dades NIVELL C02
co21 <- read_excel("CO2_01-05-2021_31-12-2021.xls",col_names = TRUE, col_types =
c("date","date",rep("numeric",23)))
co22 <- read_excel("CO2_01-01-2022_31-12-2022.xls", col_names = TRUE)
co22 <- co22[,-3]
co23 <- read_excel("CO2_01-01-2023_01-02-2023.xls", col_names = TRUE)
co23 <- co23[,-3]
#dd <- c(d21, d22, d23)
#summary(dd)

##############################################################################
### CREACIÓ DE NOVES VARIABLES
##############################################################################
# DIES amb NOM

dias <- c("lunes", "martes", "miércoles", "jueves", "viernes", "sábado",
"domingo")
dies <- c("Dilluns", "Dimarts", "Dimecres", "Dijous", "Divendres",
"Dissabte","Diumenge")

tempe21$dia <- as.factor(format(tempe21$Date, "%A"))
tempe21$dia <- factor(tempe21$dia, levels = dias,labels = dies)
tempe22$dia <- as.factor(format(tempe22$Date, "%A"))
tempe22$dia <- factor(tempe22$dia, levels = dias,labels = dies)
tempe23$dia <- as.factor(format(tempe23$Date, "%A"))
tempe23$dia <- factor(tempe23$dia, levels = dias,labels = dies)

humi21$dia <- as.factor(format(humi21$Date, "%A"))
```

```
humi21$dia <- factor(humi21$dia, levels = dias,labels = dies)
humi22$dia <- as.factor(format(humi22$Date, "%A"))
humi22$dia <- factor(humi22$dia, levels = dias,labels = dies)
humi23$dia <- as.factor(format(humi23$Date, "%A"))
humi23$dia <- factor(humi23$dia, levels = dias,labels = dies)


co21$dia <- as.factor(format(co21$Date, "%A"))
co21$dia <- factor(co21$dia, levels = dias,labels = dies)
co22$dia <- as.factor(format(co22$Date, "%A"))
co22$dia <- factor(co22$dia, levels = dias,labels = dies)
co23$dia <- as.factor(format(co23$Date, "%A"))
co23$dia <- factor(co23$dia, levels = dias,labels = dies)

# HORES AMB MINUTS

tempe21$hora <- as.factor(format(tempe21$Time, "%H:%M"))
tempe22$hora <- as.factor(format(tempe22$Time, "%H:%M"))
tempe23$hora <- as.factor(format(tempe23$Time, "%H:%M"))

humi21$hora <- as.factor(format(humi21$Time, "%H:%M"))
humi22$hora <- as.factor(format(humi22$Time, "%H:%M"))
humi23$hora <- as.factor(format(humi23$Time, "%H:%M"))

co21$hora <- as.factor(format(co21$Time, "%H:%M"))
co22$hora <- as.factor(format(co22$Time, "%H:%M"))
co23$hora <- as.factor(format(co23$Time, "%H:%M"))



# MESOS

mes_esp<-
c("ene.","feb.","mar.","abr.","may.","jun.","jul.","ago.","sep.","oct.","nov.","di
c.")
mes_cat <- c("Gener", "Febrer", "Març", "Abril", "Maig", "Juny", "Juliol",
"Agost", "Setembre", "Octubre", "Novembre","Desembre")

tempe21$mes <- as.factor(format(tempe21$Date, "%b"))
tempe21$mes <- factor(tempe21$mes, levels = mes_esp,labels = mes_cat)
tempe22$mes <- as.factor(format(tempe22$Date, "%b"))
tempe22$mes <- factor(tempe22$mes, levels = mes_esp,labels = mes_cat)
tempe23$mes <- as.factor(format(tempe23$Date, "%b"))
tempe23$mes <- factor(tempe23$mes, levels = mes_esp,labels = mes_cat)


humi21$mes <- as.factor(format(humi21$Date, "%b"))
humi21$mes <- factor(humi21$mes, levels = mes_esp,labels = mes_cat)
humi22$mes <- as.factor(format(humi22$Date, "%b"))
humi22$mes <- factor(humi22$mes, levels = mes_esp,labels = mes_cat)
humi23$mes <- as.factor(format(humi23$Date, "%b"))
humi23$mes <- factor(humi23$mes, levels = mes_esp,labels = mes_cat)

co21$mes <- as.factor(format(co21$Date, "%b"))
co21$mes <- factor(co21$mes, levels = mes_esp,labels = mes_cat)
co22$mes <- as.factor(format(co22$Date, "%b"))
co22$mes <- factor(co22$mes, levels = mes_esp,labels = mes_cat)
co23$mes <- as.factor(format(co23$Date, "%b"))
co23$mes <- factor(co23$mes, levels = mes_esp,labels = mes_cat)
```

```
# Hores

tempe21$h <- hour(tempe21$Date)
tempe22$h <- hour(tempe22$Date)
tempe23$h <- hour(tempe23$Date)

humi21$h <- hour(humi21$Date)
humi22$h <- hour(humi22$Date)
humi23$h <- hour(humi23$Date)

co21$h <- hour(co21$Date)
co22$h <- hour(co22$Date)
co23$h <- hour(co23$Date)

# Dia (número)

tempe21$d <- as.factor(format(tempe21$Date, "%d"))
tempe22$d <- as.factor(format(tempe22$Date, "%d"))
tempe23$d <- as.factor(format(tempe23$Date, "%d"))

humi21$d <- as.factor(format(humi21$Date, "%d"))
humi22$d <- as.factor(format(humi22$Date, "%d"))
humi23$d <- as.factor(format(humi23$Date, "%d"))

co21$d <- as.factor(format(co21$Date, "%d"))
co22$d <- as.factor(format(co22$Date, "%d"))
co23$d <- as.factor(format(co23$Date, "%d"))
```

***Annex 2**: Funcions*

```
## Functions

# a la funció se le passen dos elements
# el primer df, es el data frame de la base de dades
# k, és un vector que exclou les variables referents al temps de recollida de les dades
#i les variables que serveixen com a segmentacions

k <- c(1,2,26:28)

data_summary <- function(df){
 dd <- as.data.frame(df)
 table1 <- (as.data.frame(
 cbind(
  Min=apply(dd[,-k],2,min, na.rm=T),
  Q1=sapply(dd[,-k], function(x) quantile(x, probs = 0.25, na.rm = T)),
  Median=apply(dd[,-k],2,median, na.rm=T),
  Mean=apply(dd[,-k],2,mean, na.rm=T),
  SD=apply(dd[,-k],2,sd, na.rm=T),
  Q3=sapply(dd[,-k], function(x) quantile(x, probs = 0.75, na.rm = T)),
  Max=apply(dd[,-k],2,max, na.rm=T))))
 table1 <- round(table1,2)
```

```
  return(table1)
}

# modificació per tal de calcular els mateixos estadístics per segmentacions
# s'afegeix el paràmetre seg de segmentació
# com s'aplica a la nova base de dades cal redefinir les columnes que no es faran servir


segmented_data_summary <- function(df, seg) {
  # Split the data frame into segments based on the 'seg'
  segmented_data <- split(df, df[[seg]])

  # Initialize an empty list to store results for each segment
  results_list <- list()

  # Iterate through each segment and calculate summary statistics for numeric variables
  for (segment_name in names(segmented_data)) {
    segment_df <- segmented_data[[segment_name]]

    # Filter out only the numeric columns
    numeric_columns <- segment_df[, sapply(segment_df, is.numeric)]

    if (ncol(numeric_columns) > 0) {
      summary_stats <- as.data.frame(
        cbind(
          Min = apply(numeric_columns, 2, min, na.rm = TRUE),
          Q1 = sapply(numeric_columns, function(x) quantile(x, probs = 0.25, na.rm = TRUE)),
          Median = apply(numeric_columns, 2, median, na.rm = TRUE),
          Mean = apply(numeric_columns, 2, mean, na.rm = TRUE),
          SD = apply(numeric_columns, 2, sd, na.rm = TRUE),
          Q3 = sapply(numeric_columns, function(x) quantile(x, probs = 0.75, na.rm = TRUE)),
          Max = apply(numeric_columns, 2, max, na.rm = TRUE)
        )
      )
      summary_stats <- round(summary_stats, 2)

      # Store the results in a list with segment name as the key
      results_list[[segment_name]] <- summary_stats
    }
  }

  return(results_list)
}


# aplicat per al cas de les hores per l'any 2022
```

```
k2 <- c(1:4,28:30)

hores_data_summary_22 <- function(df){
 dd <- as.data.frame(df)
 table1 <- (as.data.frame(
  cbind(
   Min=apply(dd[,-k2],2,min, na.rm=T),
   Q1=sapply(dd[,-k2], function(x) quantile(x, probs = 0.25, na.rm = T)),
   Median=apply(dd[,-k2],2,median, na.rm=T),
   Mean=apply(dd[,-k2],2,mean, na.rm=T),
   SD=apply(dd[,-k2],2,sd, na.rm=T),
   Q3=sapply(dd[,-k2], function(x) quantile(x, probs = 0.75, na.rm = T)),
   Max=apply(dd[,-k2],2,max, na.rm=T))))
 table1 <- round(table1,2)
 return(table1)
}
```