Project 3 - Team 2

Commercial
Building Disclosure Visualisations

Our Team Members & GitHub Repository

Team Members: Ben, Erin, Violet and Archana

GitHub Repository: https://github.com/VeeBui/project-three-team-two

What is our visualisation about?

We are working for a building leasing and marketing company, and have been tasked with researching the Commercial Building Disclosure project, which can benefit from data visualisation.

Searchable with data, you can view a map as well as specific building metrics such as ratings, energy intensity, net lettable areas and more.

The data delivery is manual to up-date and requires processing to make easy to understand, this clearly shows the certified buildings and their ratings in an easy and visual way.

The primary objective of this project is to perform a visualisation analysis of up to 40,000 ratings within Victoria using available data from the Australian Government site www.cbd.gov.au.





About the CBD Program & Ethics

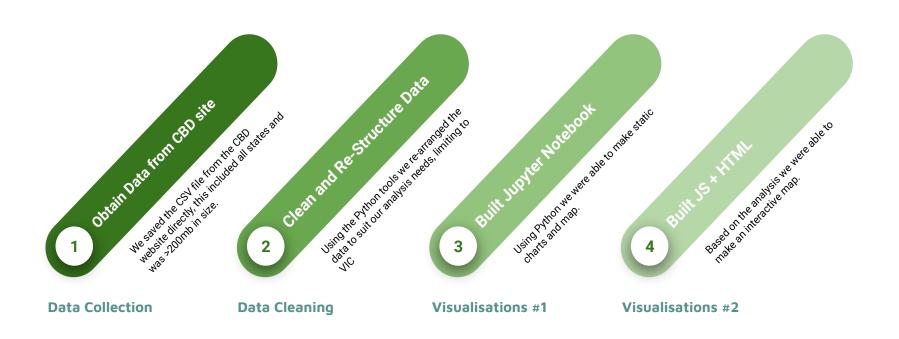
A National Energy Efficiency Program

The Commercial Building Disclosure (CBD)
Program is a national regulatory program that
requires energy efficiency information to be
provided in most cases when commercial office
space of 1000 square metres or more is offered for
sale or lease

Ethical Considerations

The CBD Program is an example of transparency in data collection. This is an instance where you would want the data that has been collected to be shared. The data provided to clients allows them to make an informed choice regarding the energy efficiency of the building they would like to lease which could ultimately affect their fixed costs.

Project Process



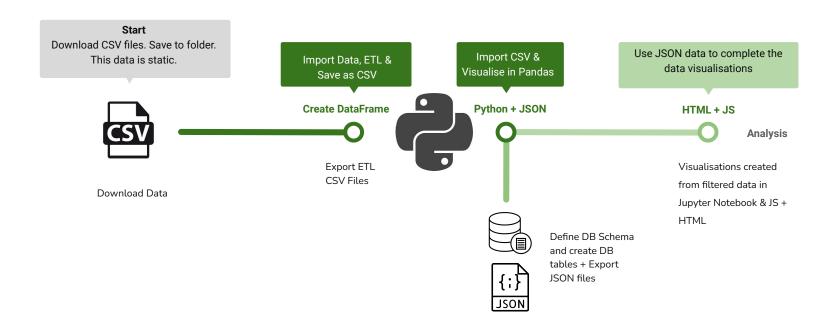
Dependency Review

Our project included the following stack and dependencies;

- Publicly available CSV data from Ethical Government published site
- Jupyter (Pandas) for data loading and transformation
- Jupyter for static data visualisation
- Create JSON Files from Pandas
- Create SQLIte DB to store the data
- JS to make the interactive map.
- The new python library Plotly Express was used in Visualisations #1 Jupyter Notebook.

Data Collection

Method For Data Collection



Method | Dependencies

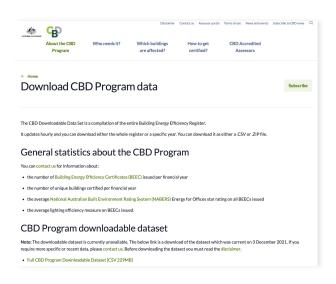
Our project included the following setup and dependencies;

from pathlib import Path from sqlalchemy.ext.automap import automap_base from sqlalchemy.ext.declarative import declarative_base from sqlalchemy import Column, String, Float, Integer from sqlalchemy import create_engine from sqlalchemy.orm import Session

CBD | Website Link

Our project included the Commercial Building Disclosure site;

Given the file size of the download >200mb, we have not uploaded this to out GitHub Repository. Please see link to file for reference.



Full CBD Program Downloadable Dataset [CSV 229MB]

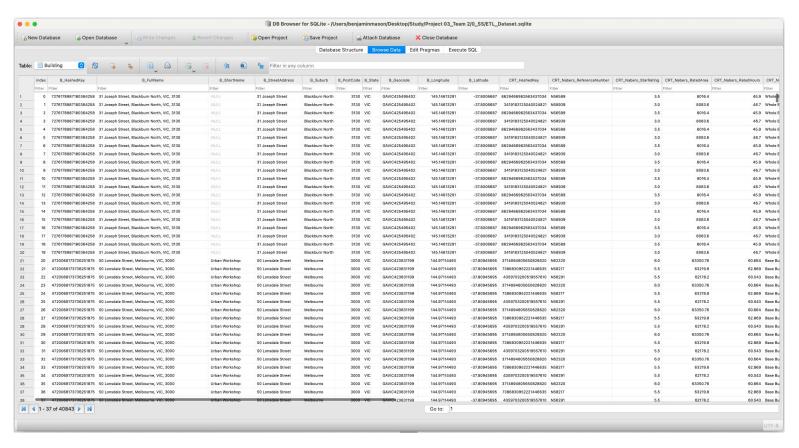
The Jupiter Lab file showing the first CSV read;

P. Uas	hodKov	P. FullName	P. ShortName	B_StreetAddress	P. Suburb	R PostCode	B State	R George	B Longitude	R Latitudo	EC D
0 3292694929379		1 Thynne Street, Bruce, ACT, 2617	NaN	1 Thynne Street	Bruce	2617		GAACT714853532	149.093585	-	
1 3292694929379	9159574	1 Thynne Street, Bruce, ACT, 2617	NaN	1 Thynne Street	Bruce	2617	ACT	GAACT714853532	149.093585	-35.240431	
2 329269 4 929379	9159574	1 Thynne Street, Bruce, ACT, 2617	NaN	1 Thynne Street	Bruce	2617	ACT	GAACT714853532	149.093585	-35.240431	
3 329269 4 929379	9159574	1 Thynne Street, Bruce, ACT, 2617	NaN	1 Thynne Street	Bruce	2617	ACT	GAACT714853532	149.093585	-35.240431	
4 3292694929379	9159574	1 Thynne Street, Bruce, ACT, 2617	NaN	1 Thynne Street	Bruce	2617	ACT	GAACT714853532	149.093585	-35.240431	
5 rows × 62 column	s										
df.info()											
<pre><class 'pandas.="" (to<="" 204'="" column="" data="" rangeindex:="" td=""><td>749 entr otal 62 y</td><td>ies, 0 to 2</td><td>Non-Nu 204749 204749 142420</td><td>ll Count Dtype non-null int64 non-null objec non-null objec</td><td>t</td><td></td><td></td><td></td><td></td><td></td><td></td></class></pre>	749 entr otal 62 y	ies, 0 to 2	Non-Nu 204749 204749 142420	ll Count Dtype non-null int64 non-null objec non-null objec	t						

The SQLite Building Table Schema;

```
    Search
    Se
                                                                                                                                                                                                                                                                                                                                                                                                                                       ზე Ⅲ ...
buildings_schema.sql U X
Users > benjaminmason > Desktop > Study > Project 03 Team 2 > project-three-team-two old > static > data > buildings schema.sql
                    DROP TABLE IF EXISTS Building;
                  CREATE TABLE Building (
                             B_HashedKey decimal,
                             B_FullName varchar,
                             B_ShortName varchar,
                             B_StreetAddress varchar,
                             B_Suburb varchar,
                             B_PostCode integer,
                             B_State varchar,
   11
                             B_Geocode varchar,
   12
                             B_Longitude decimal,
   13
                             B_Latitude decimal,
   14
                             CRT_HashedKey decimal,
   15
                             CRT_Nabers_ReferenceNumber varchar,
    16
                             CRT_Nabers_StarRating decimal,
   17
                             CRT_Nabers_RatedArea decimal,
    18
                             CRT_Nabers_RatedHours decimal,
    19
                             CRT_Nabers_RatingScope varchar,
   20
                             CRT Nabers AnnualEmissions decimal,
   21
                             CRT_Nabers_AnnualEmissionIntensity decimal,
   22
                             CRT Nabers AnnualConsumption decimal,
   23
                             CRT_Nabers_OwnerName_varchar,
   24
                             TLA_Name varchar,
   25
                             TLA_AssessorName varchar,
   26
                             TLA_NetLettableSpace decimal,
   27
                             FS_Name varchar,
   28
                             FS_Level decimal,
   29
                             FS_NLA decimal,
    30
                             CRT_BuildingNla decimal,
   31
                             CRT_NumberOfLevels decimal
    32
    33
    34
                   -- Verify successful data import
    35
                   SELECT * FROM Building;
    36
```

The SQLite DataBase (Showing Buildings Table);



Data Cleaning

Method | Data Cleaning

For our project we needed to clean the data using Python;

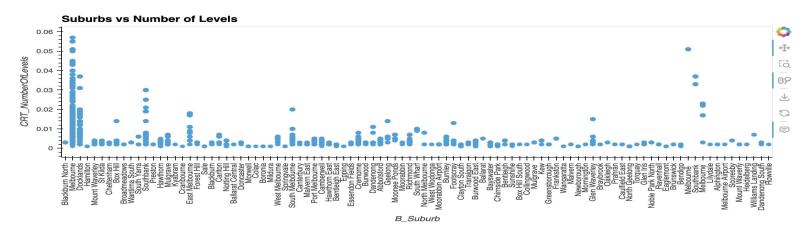
- Drop columns that had "dirty data" meaning various data object type in the same column. This was mainly date and columns with additional separators that couldn't be split without making errors.
- Drop rows that didn't include Lat + Long for adding to the Map.
- Reduced dataset to limit the visualisations to buildings in Victoria.
- Reduced total columns from 62 to 28.
- Reduced total rows from 204,000 to 40,800.

The Jupiter Lab file showing the leaned up dataset;

```
[46]: # Import the dependencies.
      import sqlite3
      from sqlalchemy import create_engine
      from sqlalchemy import Column, Integer, String, Float
      from sqlalchemy.ext.declarative import declarative_base
      from pathlib import Path
      from sqlalchemy import create engine, text
      import matplotlib.pyplot as plt
[23]: Base = declarative_base()
      conn = sqlite3.connect('Buildings.sqlite')
      cursor = conn.cursor()
[24]: cursor.execute('''DROP TABLE Building''')
[24]: <sqlite3.Cursor at 0x306af9240>
[25]: # Create a Buidling Class for the table.
      cursor.execute('''CREATE TABLE Building(
      B HashedKey decimal, B FullName varchar, B ShortName varchar, B StreetAddress varchar, B Suburb varchar, B PostCode integer,
      B_State varchar, B_Geocode varchar, B_Longitude decimal, B_Latitude decimal,
      CRT_HashedKey decimal, CRT_Nabers_ReferenceNumber varchar,
      CRT Nabers StarRating decimal, CRT Nabers RatedArea decimal,
      CRT Nabers RatedHours decimal, CRT Nabers RatingScope varchar,
      CRT_Nabers_AnnualEmissions decimal, CRT_Nabers_AnnualEmissionIntensity decimal,
      CRT_Nabers_AnnualConsumption decimal, CRT_Nabers_OwnerName varchar,
      TLA Name varchar, TLA AssessorName varchar, TLA NetLettableSpace decimal,
      FS_Name varchar, FS_Level decimal, FS_NLA decimal, CRT_BuildingNla decimal, CRT_NumberOfLevels decimal)''')
[25]: <sqlite3.Cursor at 0x306af9240>
[26]: Building_ETL = pd.read_csv('/Users/benjaminmason/Desktop/Study/Project 03_Team 2/project-three-team-two_old/resources/ETL_Dataset.csv')
[27]: Building_ETL.to_sql('Building', conn, if_exists='append', index = False)
[27]: 40843
[28]: cursor.execute('''SELECT * FROM Building''').fetchall()
[28]: [(7276178987180364258,
        '31 Joseph Street, Blackburn North, VIC, 3130',
        '31 Joseph Street',
```

Visualisations #1 Jupyter Notebook

Number of levels of buildings per suburbs

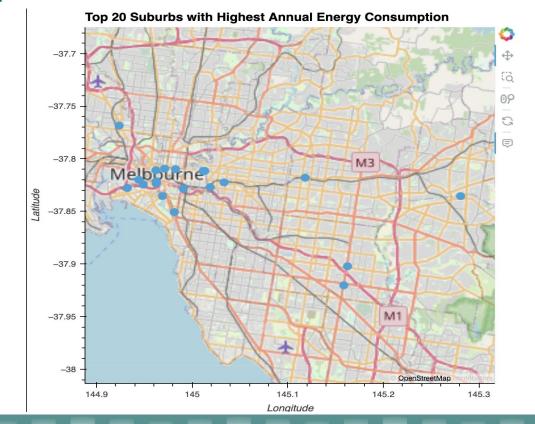


Summary:

- The scatter plot reveals the relationship between suburbs and greenhouse gas emissions.
- Most data points cluster at the lower end of GHG emissions, with a few outliers showing higher emissions.
- Without context, we can't definitively determine the correlation or causation between suburbs and emissions.

TOP 20 Suburbs with Highest Annual Energy Consumption

Here's a map plot displaying the top 20 suburbs with the highest annual energy consumption. The plot uses latitude and longitude coordinates to visualise these suburbs on the map.



	B_Suburb
0	North Geelong
1	Mildura
2	Geelong
3	Brunswick
4	Mornington
5	Williams Landing
6	Docklands
7	South Wharf
8	Sale
9	Southbank
10	Broadmeadows
11	Burwood
12	Melbourne
13	Boronia
14	Caulfield East
15	Colac
16	Cranbourne
17	Hamilton
18	Heidelberg
19	Lilydale

"Average Star Ratings for Top 20 Suburbs"

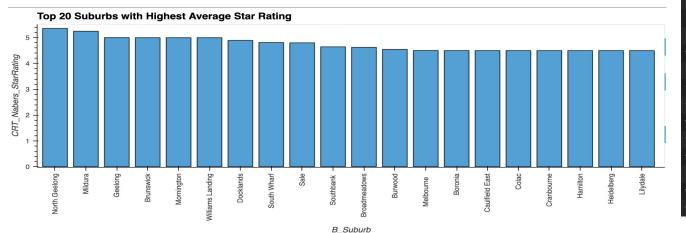
```
# Calculate the average 'CRT_Nabers_StarRating' for each suburb
average_ratings = building_df.groupby('B_Suburb')['CRT_Nabers_StarRating'].mean()

# Select the top 20 suburbs with the highest average rating
p_20_suburbs = average_ratings.nlargest(20)

# Convert the Series to a DataFrame
top_20_suburbs_df = top_20_suburbs.reset_index()

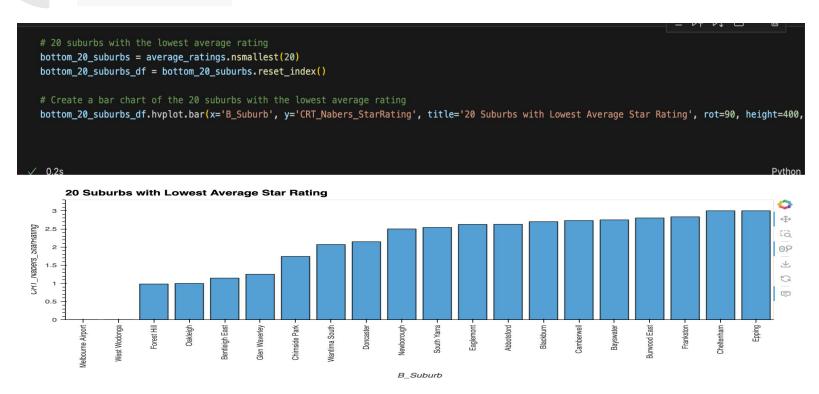
# Display the DataFrame
top_20_suburbs_df

✓ 0.0s
```

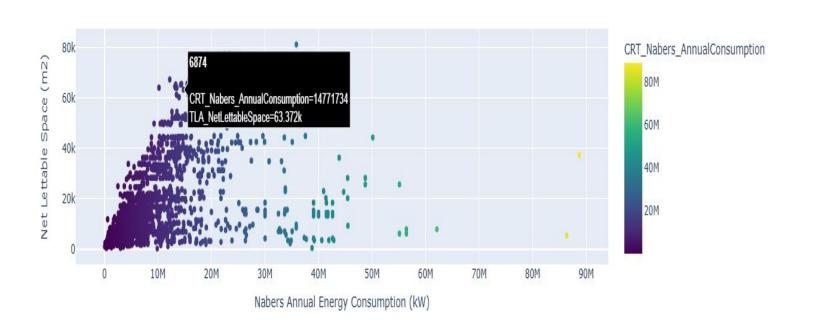


	B_Suburb	CRT_Nabers_StarRating
0	North Geelong	5.360294
1	Mildura	5.250000
2	Geelong	5.004808
3	Brunswick	5.000000
4	Mornington	5.000000
5	Williams Landing	5.000000
6	Docklands	4.894899
7	South Wharf	4.812500
8	Sale	4.800000
9	Southbank	4.645000
10	Broadmeadows	4.625000
11	Burwood	4.545455
12	Melbourne	4.503440
13	Boronia	4.500000
14	Caulfield East	4.500000
15	Colac	4.500000
16	Cranbourne	4.500000
17	Hamilton	4.500000
18	Heidelberg	4.500000
19	Lilydale	4.500000

"Average Star Ratings for lowest 20 Suburbs"



Annual Consumption vs. Net Lettable Space



Visualisations #2 Flask + HTML + JS + JSON

Method | Further Data Cleaning

For this part of the project we needed to reduce the quantity of rows in the data set to make the map load using JavaScript & JSON;

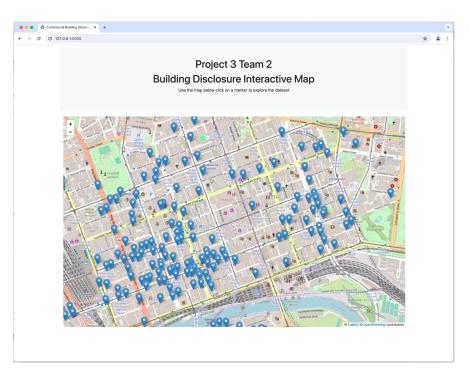
- Drop duplicate building entries, this was done to allow the Map to load.
- Reduced total rows from 40,800 to 2,556

Flask | JS + JSON + CSS

This section of the project includes Flask app.py using HTML, JavaScript and JSON

```
    Search
    Se
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              D ~ $3 Ⅲ ···
app.py 1, U X
 Users > benjaminmason > Desktop > Study > Project 03_Team 2 > project-three-team-two_old > 🏺 app.py > ...
                                from flask import Flask, render_template
                                # Flask Setup
                               app = Flask( name )
                                # Flask Routes
                               12
      13
                                @app.route('/')
      14
                               def index():
      15
                                                 return render_template('index.html')
      16
```

Interactive Map to Explore Data





Lessons learnt

Lessons learnt

- 1. Using GitHub and branch for code revision management is of benefit, but has file limit <100mb.
- 2. Acquiring data from public sites can be "dirty" and require extensive clean up and testing for errors.
- 3. Street Map will not load markers when using excessive entries.
- 4. Generating JSON data within FLASK using the SQLite DB was difficult and as an alternative used a static JSON data file exported from Python.
- 5. Domain knowledge (or lack of time) can play a major role in DB backend & FLASK development, more time was needed to finish the FLASK (SQLite > JSON) back end creation.

Thank You!

Team 2