Futuristic Smart City

---What causes the most carbon footprint?

Group 10 240030602 240022601 240002187
Date April 2025

1 Data & Questions –240030602

The dataset we chose is Futuristic Smart City Citizen Activity Dataset from Kaggle dataset under the filter of "Data Visualization", uploaded by Atharva Soundankar, since it's released recently and related to our growing concern about sustainability. License is CC0: Public Domain thus no copyright constraints.

The dataset contains 1,000 data points and 15 contributes, covering demographics (e.g., Age, Gender), behavior (e.g., Sleep Hours, Shopping Hours), mobility (e.g., Mode of Transport), and energy use (e.g., Carbon_Footprint_kgCO2). The dataset includes both numerical and categorical variables. Basically, it's ready to be visualized.

This visualization project aims to explore which behavioral and demographic factors affect carbon footprints the most. The data exploration and processing work found the correlation of the features and Carbon_Footprint_kgCO2. Highly ranked features are saved first. Besides, we also made our own choice based on the intention of an interesting design. Through interactive visualization webpage, insights can be given about our question: what causes the most carbon footprint?

2 Description of Visualisation – 240022601

The visualization can be seen as three parts: the title and filter section, the overview with three charts, and the detail with two charts. We chose "green" as the core theme color to avoid stereotypical associations and to connect with the topic of carbon footprint and sustainability.

Originally, bar charts, a pie chart, a scatterplot, and a box plot were chosen accordingly to address the open question of "What causes the most carbon footprint?" Based on the nature of each data column, for example, the mode of transport, a bar chart is the most intuitive way to compare the carbon footprint differences across categories. A pie chart was chosen for gender to emphasize proportion. A scatter plot shows how age correlates with carbon footprint across the dataset. Originally, we intended to use a box plot for age group distributions in the detail section, but it was replaced with a bar chart due to implementation issues.

Regarding interactivity, four filters were provided in the top section to allow user filtering by mode of transport, gender, age group, and the social media usage hours, allowing users to explore different data combinations and discover how each factor may relate to carbon footprint.

Additionally, all chart elements all serve as filters. For example, clicking the gender while choosing the mode of transport as walking will change the filter on the top accordingly, and every chart will change dynamically, showing only the related data.

Tooltips are also provided. When hovering over the element, they will show detailed values like total footprint, average footprint, and data counts, helping users explore the data more deeply. This approach should align with Schneiderman's Mantra, "overview first, details on demand," helping users easily grasp the broad patterns and dig into specific insights when needed.

3 Implementation –240002187

Our group used the carbon footprint value as the key point of the study, and analyzed the transportation, social, age, and gender of urban residents' activities.

I added a link to the data source and a brief description on the panel. I also added buttons that can be filtered. The images can also interact with each other and change accordingly. When our mouse clicks on the bar, there will be a filtering function. When our mouse moves to the bar or points, there will be many necessary details.

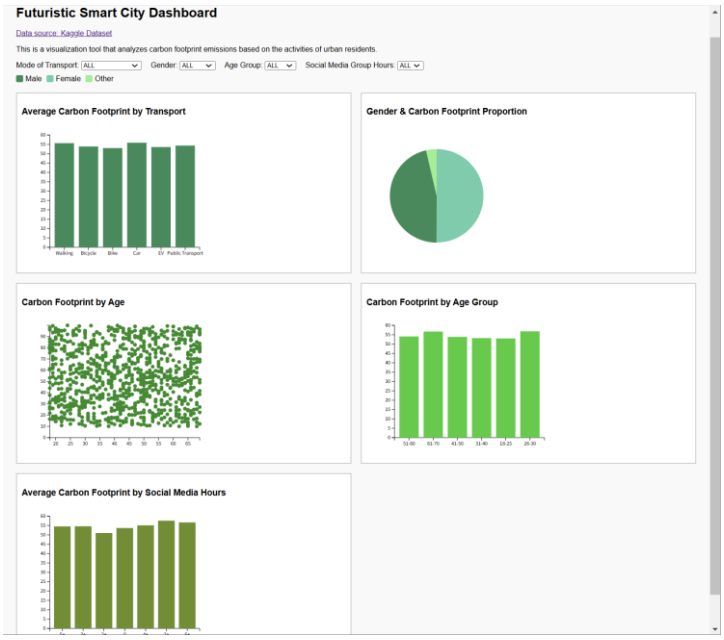

Figure: General appearance of the website

We also thought that this data could be processed first, then visualized and filtered, so the results returned were ideal.

Citizen_ID	Age	Gender	Mode_of_T	Work_Hou	Shopping_	Entertainm	Home_Ene	Charging_	Carbon_Fo	Steps_Wall	Calories_Bi	Sleep_Hou	Social_Mec	Public_Events_Hours
1001	56	Female	Walking	5	2	0	5.32	0	44.7	15635	975	9.2	5.8	0.5
1002	69	Male	Bicycle	0	2	2	2.19	0	92.39	1671	455	8.1	5.5	1.9
1003	46	Male	Bike	0	4	0	4.68	0	78.57	1777	324	4.7	3.8	2.8
1004	32	Male	Car	7	2	3	3.42	0	55.46	4022	537	4.9	3.5	0.5
1005	60	Male	Walking	3	3	1	2.79	0	98.95	19244	1414	6.6	2.2	0.5
1006	25	Male	Walking	0	1	3	9.65	0	51.87	14817	1104	5.2	0.1	1.3
1007	38	Male	EV	9	0	1	6.79	1	15.24	15260	997	4.2	0.8	2.4
1008	56	Male	EV	6	4	2	7.83	0	33.6	2074	346	8.8	4.5	0.9
1009	36	Male	Bike	4	3	1	4.52	0	49.25	10908	663	4.1	2.8	0.8
1010	40	Female	Car	4	0	0	5.15	0	24.12	10475	853	9	0.9	1.9
1011	28	Female	Walking	0	0	3	3.9	1	95.28	13878	926	4.9	1.6	1.6
1012	28	Female	Bike	9	4	1	2.78	0	79.02	16120	1019	9.8	5.7	0.2
1013	41	Female	Bike	5	3	3	3.43	0	34.15	2986	387	7.6	2.2	0.2
1014	53	Female	Public Trar	8	0	2	8.38	1	63.17	11825	815	8.4	5.4	1.9
1015	57	Male	EV	8	0	2	7.42	0	74.93	10193	812	4.7	1.8	1.4
1016	41	Female	Car	7	3	0	6.38	0	94.72	18430	1081	9.2	3.3	2.9
1017	20	Female	Walking	4	2	2	5.8	0	89.35	3176	507	9	0.1	2
1018	39	Female	Bike	0	0	1	9.38	0	56.37	19722	1383	7.9	3.1	0
1019	19	Male	Walking	8	1	2	2.58	0	79.72	11047	664	7	3.8	1.4
1020	41	Male	Public Trar	0	3	0	4.24	1	60.59	17884	1389	8.3	1.4	1.2
1021	61	Male	Car	7	0	3	4.8	0	59.74	19617	1334	4	3.5	0.3
1022	47	Male	Bike	4	4	3	8.26	1	20.33	5608	464	7.6	3.9	2.8
1023	55	Female	Walking	2	1	2	9.94	0	21.35	12906	764	9.1	4.1	2.4

smart_city_citizen_activity

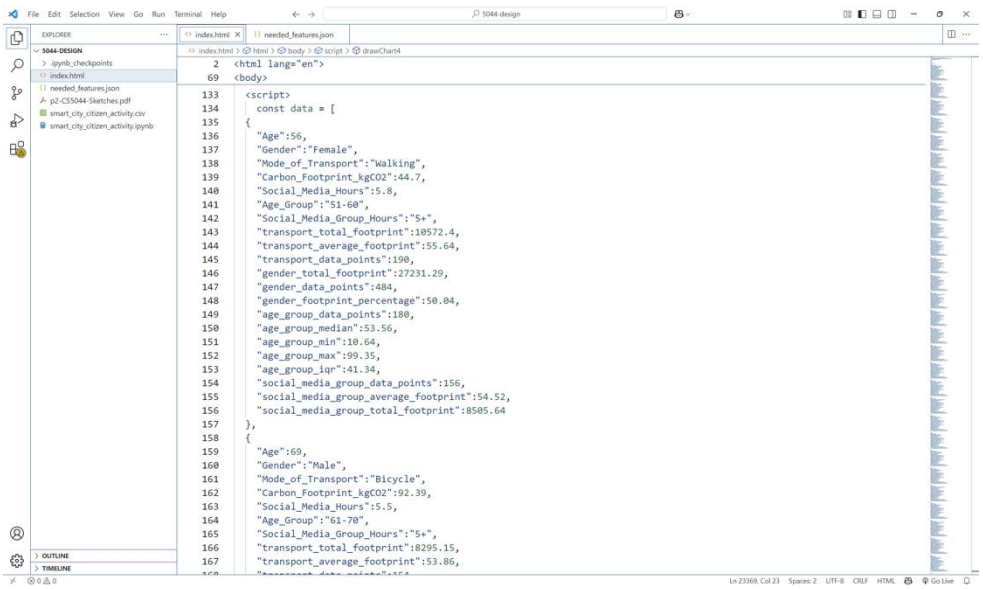

Figure: Data sets

4 Insights from the Visualisation & Critical Discussion –240002187 240022601

There are already several interesting insights that can be seen without using filters, only tooltips are needed. For example, while we might assume walking is the most eco-friendly mode of transport (in contrast to cars), or that the older generation (61–70) contributes less to the carbon footprint due to their lifestyle, the charts show otherwise. Both are actually the second-highest contributors within their categories, just slightly behind cars and the 26–30 age group.

We in fact faced several implementation limitations due to skill and time constraints.

First, during preprocessing, I created new columns (like grouped social media hours and age groups). Because of the use of labels and bins, these columns became categorical instead of numerical, which caused some charts to appear out of order, which might slightly confuse users.

Second, we didn't clearly separate the overview (three charts) from the detail section (two charts), which weakened the flow we originally aimed for.

Third, we had hoped to support multi-value filtering, but this was not successfully implemented. This limited the ability to make comparisons as originally intended.

Fourth, the JSON file from preprocessing isn't dynamically bound using D3, instead, it's currently hardcoded into HTML.

Finally, the box plot hadn't been successfully implemented. It is a compromise decision to switch to the bar chart, considering it can still be presented in a similar way.

If given more time, most of these issues could be addressed. Still, the current visualization already supports enough interactivity and clarity for users to explore the topic meaningfully.