Analysis of Sustainable Building in Israel

The Art of Analyzing Big Data- The Data Scientist's Toolbox
"Housing Market and Data"

Author: 312141369

Table of Contents

Introduction

This project presents an analysis of sustainable building data published by the Israeli Ministry of Environmental Protection. We focus on residential buildings and the differences throughout the years and between municipalities. Awareness of the need for green construction that preserves the values of sustainability and reduces energy consumption has been growing in recent years. Part of the purpose of the study is to examine how this is reflected in the construction industry. All the code is included in the attached notebook [5].

In this report we present the data used and the preprocessing steps implemented. This is followed by high-level explorations of the data using a variety of visualization tools including choropleth heatmaps, density heatmaps, histograms, scatter plots and more. An analysis of the yearly changes is presented, along with linear regression models that try and derive trends over time. The next chapter explores the differences in sustainable building between local authorities. Similar analysis of additional metrics (score, rank, type of construction projects) is included in the following chapter. Finally, a demonstration of fitting multiple linear regression model that is based on demographic data to predict the volume of sustainable building. This method is used to derive interesting conclusion on the connection between demographic metrics of local authorities and sustainable building.

Background

The Israeli Standard for Sustainable Building (SI-5281) was composed to minimize the adverse effects of the construction process and the building usage on the environment. Established first in 2005, the standard was regularly upgraded every several years, most recently in 2020 [1]. The standard defines metrics and criteria in multiple areas regarding to the environmental effect of building. These areas include energy, ground, water, materials, health and wellbeing, waste, transportation, management and innovation. The standard is based on a point accumulation system for each area based on the building's type and designation (residential, offices, education etc.). In addition, the standard relates to both new buildings and existing buildings that undergo renovation.

As of the end of 2020 there are around 450 building that meet SI-5281, with additional 1000 buildings under inspection. Sustainable buildings that were certified by the Green Building Professional Committee (GBPC) are allowed to publicly display a certificate stating that fact [2]. According to the Ministry of Environmental Protection (MOEP), many privately built buildings that do not hold the SI-5281 certification do meet the standard [3]. However, due to them not undergoing the inspection process are not included in their data. Therefore, the data collected contains only buildings that were inspected and passed the standard's threshold.

In 2020 the National Building and Planning Council approved regulations which dictates that any new building projects conform with SI-5281 (in effect from March 2022). Moreover, hundreds of Local Outline Plans that dictates the same were approved. These together, will result in the construction of thousands of sustainable buildings across the country. The new regulations are expected to decrease the emissions of green-house gases from buildings, the electricity consumption (by 30%) as well as the waste volume. Furthermore, the quality of life is expected to rise due to the use of materials which are more environmentally friendly [3].

Data

In this section we detail the data, the data sources and the data preparation procedures that were used. The main data sources include the Israeli Ministry of Environmental Protection and the Israeli Central Bureau of Statistics (CBS).

The data the MOES provides is divided in two. The first is data on sustainable buildings projects that were previously approved by the GBPC between 2008 to 2019 (hereinafter "sustainable buildings"). The second is data on buildings that are currently being built in compliance with SI-5281 (hereinafter "active sustainable buildings"). both containing extensive information on the attributes of the buildings. The attributes include the full address (with Israel Transverse Mercator (ITM) coordinates), area, designation, number of residential units, number of floors and metrics regarding SI-5281 compliance.

Local authorities in Israel data is sourced from the Israeli CBS [4]. The data file aggregates many metrics regarding diverse aspects for each municipality of the 255 in Israel. The metrics include physical data and population data - on demography, education, welfare, infrastructure, construction, etc. Additionally, the file includes budget data as well as inequality and peripherality metrics.

Preprocess

In this section we describe the data preparation steps used. The steps taken include correcting and filling missing values and converting coordinates.

Both data sets from the MOES are similarly preprocessed. First, missing values are filled using median value for each attribute. For example, missing number of floors was filed with 8 for instances from which it was missing, as it is the median of the existing floors values of other instances. Next, the columns names and local authorities' names were changed to match for all instances. This step consists of getting all the possible value forms describing the same value and setting a single form for it throughout the data. For example, a local authority name may appear in multiple different spelling forms. For each project in the sustainable buildings dataset, the certification year is stated. In the active sustainable buildings no single certification date is stated because the projects are ongoing. Therefore, the earliest date for each project is selected.

The last step of preprocessing the MOES data is to prepare the coordinates of each building. Most of the records have ITM coordinates, which were converted to the more common WGS84 form. Several coordinates were identified as wrong, i.e., outside of the borders of Israel. For these, as well as for the instances with missing attributes the coordinates were extracted from the address information. The address was translated from Hebrew to English and the coordinates were acquired using GeoPy.

Exploration

Following are high-level explorations of the data. Choropleth heatmaps, density heatmaps histograms and 3-dimensional scatter plots are used.

Figure 1 shows a heatmap of Israel with all the sustainable buildings weighted by their area (square meter). The map clearly shows that most sustainable buildings projects are in the central district and along the shoreline. Section Analysis By Municipality inspect the differences between municipalities in depth.

Figure 2 (a-c) shows various density heatmaps of pairs of attributes with each attribute histogram. These help to easily inspect the correlation between two attributes. All three heatmaps shows some correlation between the attributes. Some correlation is expected, for example, a larger area or more floors can accommodate more residential units. This partial correlation can also be observed, in a slightly noisier fashion, in Figure 2 (d). This figure shows a 3-d scatter of project area, number of residential units and floors colored by the project's year.

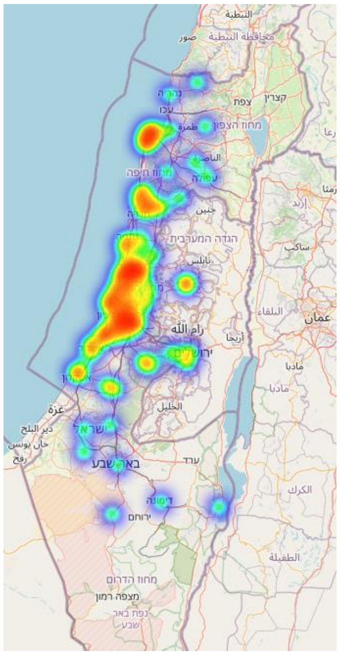

Figure 1 Heatmap showing green building in Israel weighted by area (square meter).

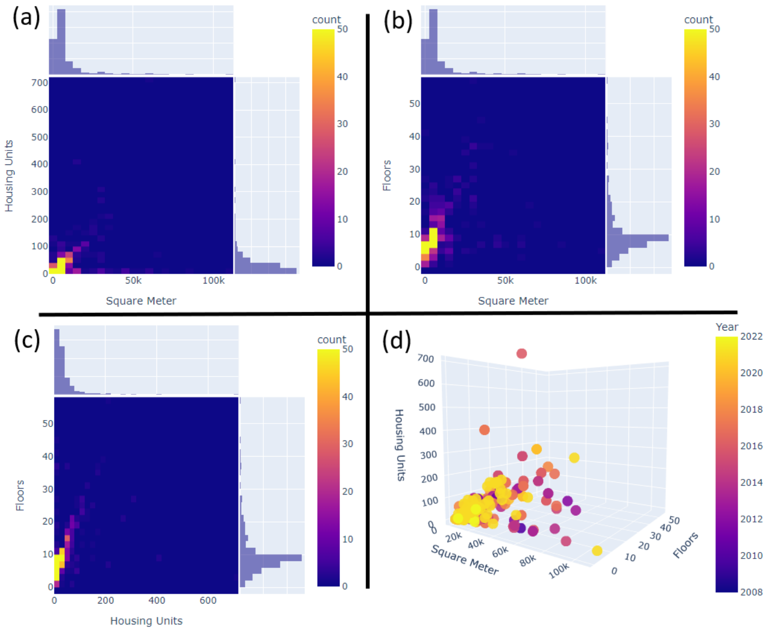

Figure 2 (a) Density heatmap of project area and number of residential units. (b) Density heatmap of project area and number of floors. (c) Density heatmap of floors and number of residential units. (d) 3-dimensional scatter plot of project area, number of residential units and floors

Table 1 displays statistics of the three main attributes of the MOES data. For all three, the standard deviation is large, signifying large variations between projects. This is also reflected in the significant difference between the mean values and the median values. The variance also shown in Figure 2 (d) scatter plot. Figure 3 shows the distribution of sustainable building projects' designations. Residence is the most frequent with 679 projects, followed by offices (49 projects) and education (46 projects).

Attribute\Metric	Mean	σ	Median	Min	Max
Housing Units	33.082	46.063	25.0	0.0	700.0
Floors	9.169	6.143	8.0	0.0	56.0
Area (square meter)	6605.826	9713.662	3961.0	0.0	107668.0

Table 1 Statistics on number of housing units, number of floors and area of projects.

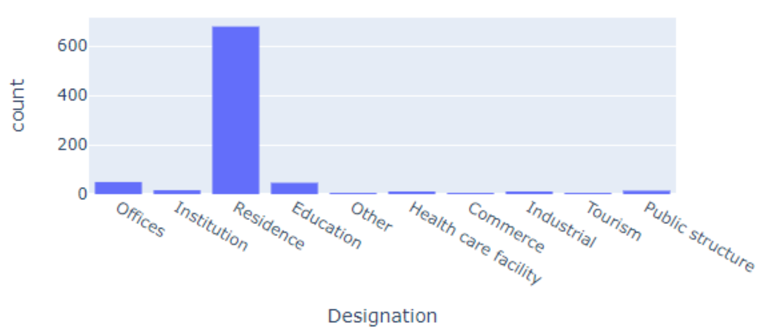

Figure 3 Histogram of the distribution of sustainable building projects' designations.

Analysis of Yearly Changes

This chapter explores the yearly changes and the trends that can be observed over time. Bar plots along with scatter plots are used to visualize the yearly changes. In addition, yearly choropleth heatmaps of the green building project help visualize the trends geographically.

A video presenting choropleth heatmaps of cumulative green building projects in Israel is included in the notebook [5]. Figure 4 shows various histograms of the distributions of several attributes per year. The 'Project Count' histogram shows the amount of sustainable building projects per year. Total yearly sum and yearly average per project for each of the following attributes – Area (square meter), Floors and housing units are also presented in Figure 4. In all the graphs showing totals (Figure 4) a clear decrease can be seen in 2020. A clear and concise reason is hard to determine, however this might be due to COVID-19 related reasons.

Figure 4 Histograms of the distributions of attributes per year.

Linear Regressions

Following are several attempts to fit simple linear regression model to try and derive trends over time. Two setups were used, the first includes the year 2022 and the second without. This is done to reduce the error stemming from the fact that for 2022 the data is partial.

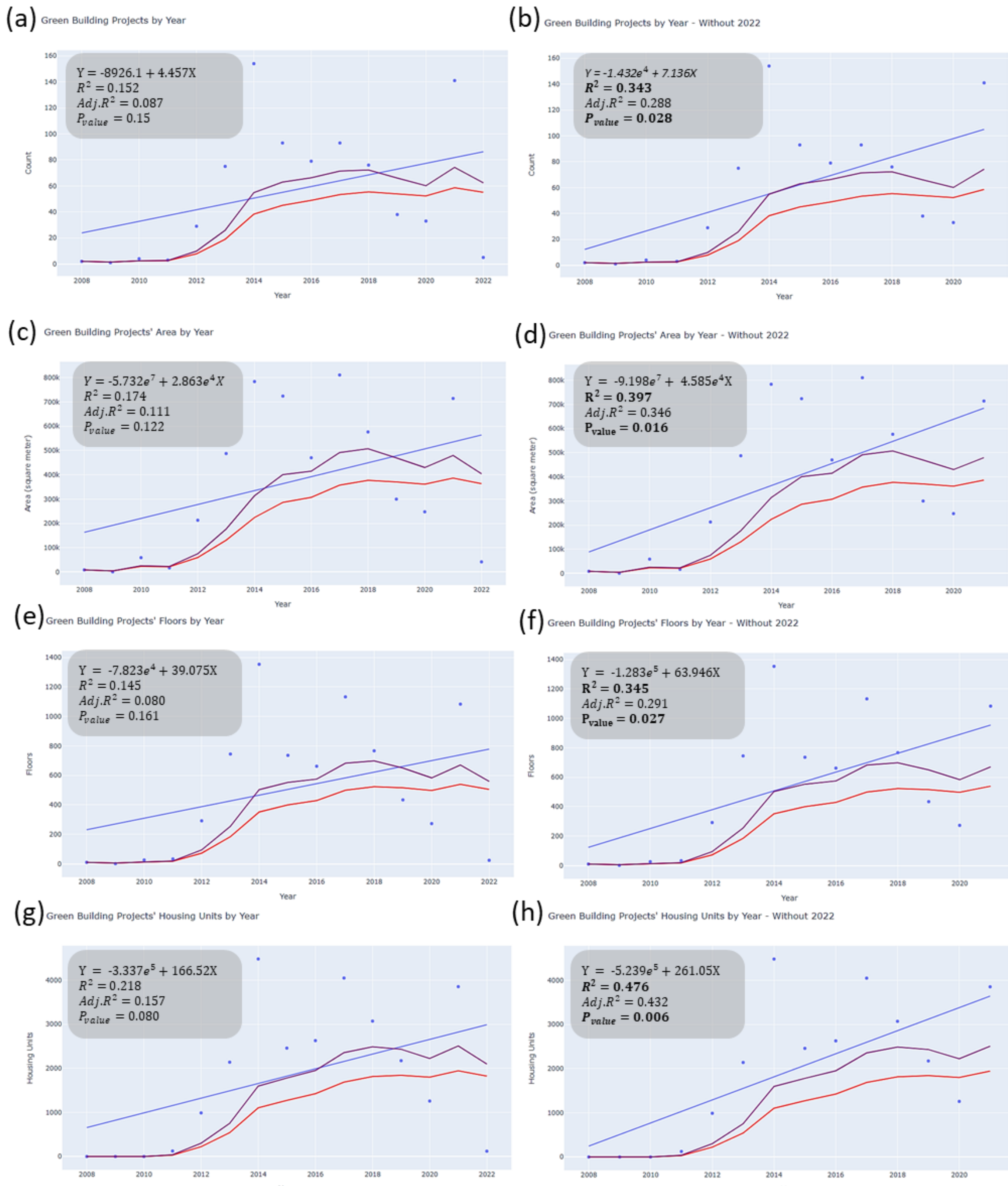

Figure 5 Scatter plots of different metrics in dependence of year. Each row represents a different metric – yearly projects total count (a, b), yearly projects total area (c, d), yearly projects total floors (e, f), yearly projects total residential units (g, h). Three lines are plotted in each graph – liner regression line (blue), exponentially weighted mean (purple) and expanding mean (red). The liner regression line equation, r-squared, adjusted r-squared and the p-value are specified for each graph.

Figure 5 shows the scatter plots of different metrics in dependence of year. Each row represents a different metric – yearly projects total count, yearly projects total area, yearly projects total floors, yearly projects total residential units. Three lines are plotted in each graph – liner regression line, exponentially weighted mean and expanding mean. The liner regression line equation, R-squared, adjusted R-squared and the P-value are specified for each graph. These values can also be seen in Table 2.

	Total Count		**Total Area**		**Total Floors**		**Total Residential Units**	
2022	with	without	with	without	with	without	with	without
Coef.	4.457	*7.136*	$2.863e^4$	$4.585e^4$	39.075	63.946	166.52	261.05
$\boldsymbol{R^2}$	0.152	0.343	0.174	0.397	0.145	0.345	0.218	0.476
$\boldsymbol{Adj.R^2}$	0.087	0.288	0.111	0.346	0.080	0.291	0.157	0.432
$\boldsymbol{P_{value}}$	0.15	$\mathbf{0.028}$	0.122	$\mathbf{0.016}$	0.161	$\mathbf{0.027}$	0.080	$\mathbf{0.006}$

Table 2 Linear regression line R-squared, adjusted R-squared and the P-value are specified for each

We can draw several conclusions from the linear regression fitting and the resulting metrics shown in Table 2. The first observation is that all the lines show positive linear coefficient between the dependent metric and the years. None of the lines fitted to data that included the partial 2022 records were significant ($\alpha = 0.05$). On the contrary, **all** the of the lines fitted to data that did not include the partial 2022 records were significant. Therefore, we can conclude that there are, in fact, statistically significant positive correlation between the year to total count, total area, total floors and total residential units. This indicates that the volume of sustainable building is increasing, or at the very least that the certification volume is increasing.

Analysis By Municipality

This chapter explores the differences between local authorities. Bar plots along with scatter plots are used to visualize differences.

Figure 6 shows the cumulative counts of projects throughout the years, aggregated and colored by municipality. From the analysis know that Raanana is the city with most sustainable building projects (with 133), followed by Yavne (with 102) and Tel-Aviv-Jaffa (with 72). We can see a major difference between the top municipalities, Raanana having almost twice as much projects as Tel-Aviv-Jaffa. The mean amount of sustainable building projects per local authority is 13.8 with standard deviation of 23.2 and a median of 7. The mean number of years in which sustainable building projects were certified per local authority is 3.7 with standard deviation of 2.8 and a median of 3.

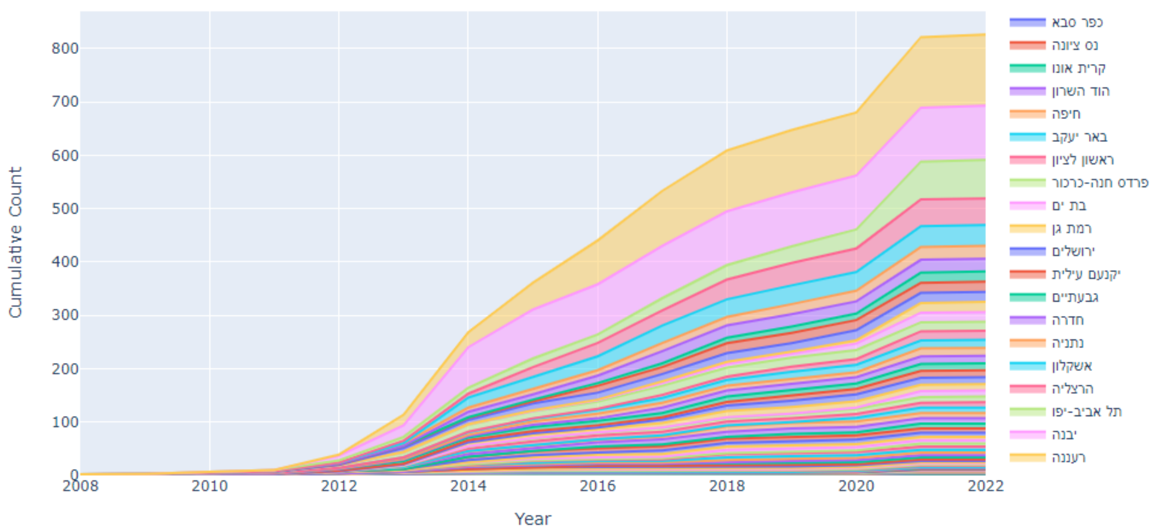

Figure 6 Cumulative counts of projects throughout the years, grouped by and colored by municipality.

Figure 7 shows various histograms of the distributions of several attributes per local authority. The ‘Project Count’ histogram shows the amount of sustainable building projects per local authority. Total sum and average per local authority for each of the following attributes – Area (square meter), Floors and housing units are also presented in Figure 7. In addition, the total number of years in which sustainable building projects were certified per local authority is presented.

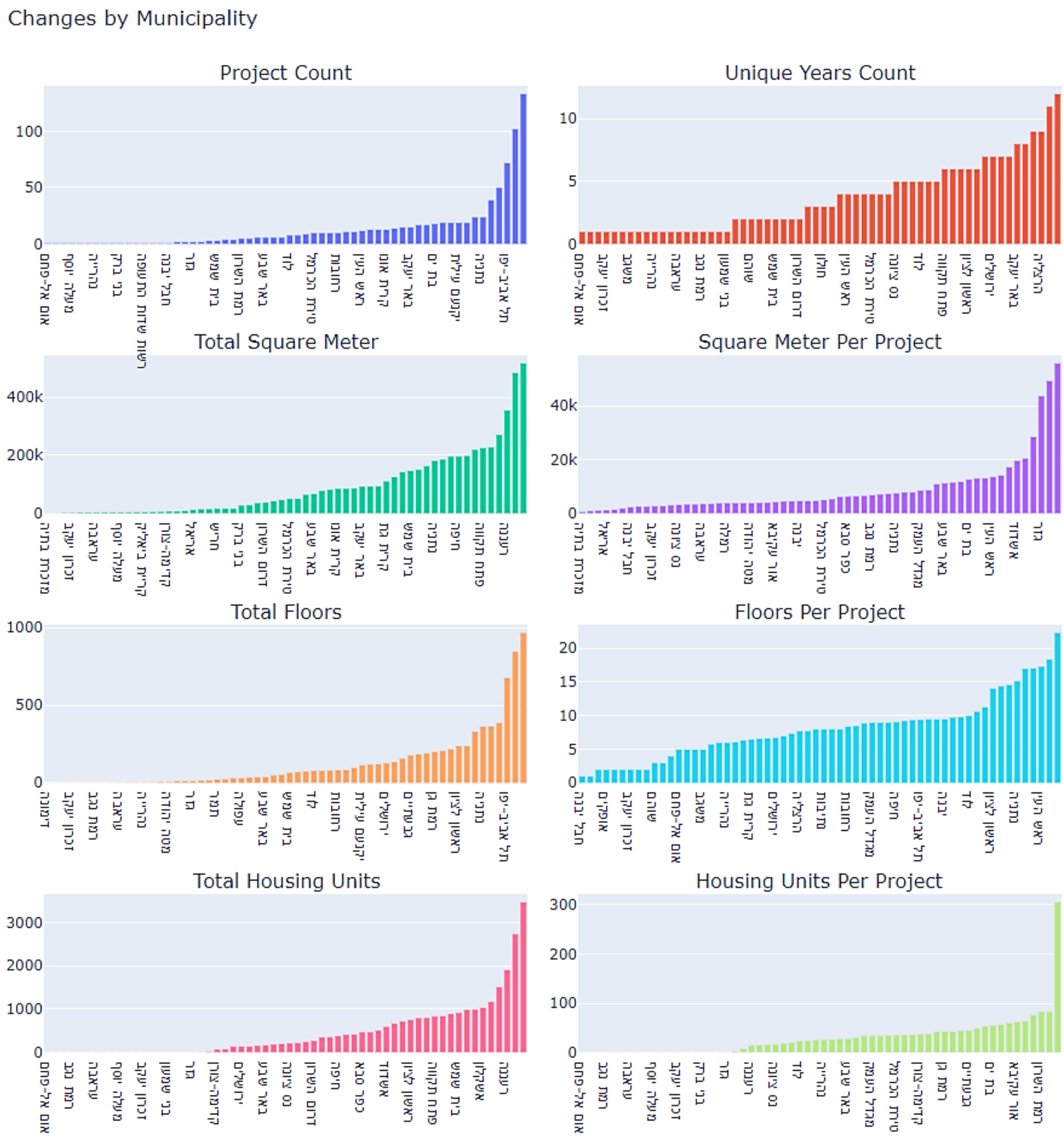

Figure 7 Histograms of the distributions of several attributes per local authority. The 'Project Count' histogram shows the amount of sustainable building projects per local authority. The total number of years in which sustainable building projects were certified per local authority is presented (top right). Total sum and average per local authority for each of the following attributes – Area (square meter), Floors and housing units are also presented.

Additional Metrics Analysis

In this section we analyze three additional attributes – the final score and rank of the project and whether the project is a construction of a new building or a renovation of an existing one. The two first attributes are only available for fully completed projects, therefore, only the "sustainable buildings" dataset records are used.

The final score of a building is an accumulation of points earned for meeting the standard. The mean final score is 60.15 with standard deviation of 5.46 and a median of 59. The maximum final score is 90 while the lowest score is only 43 points. In the attached notebook [5], there is an average score per year. Figure 8 shows a violin plot of the scores (with an internal box plot). This visualization is similar to histogram but helps better visualize the spread of data points throughout the value space. We can clearly see that most of the projects received scores of around 59. In addition, a second small cluster can be seen at around 75-76.

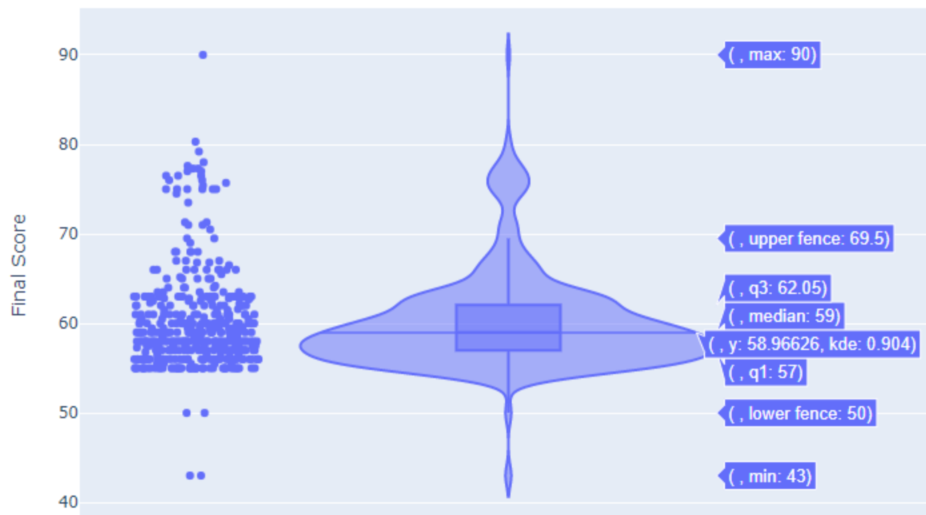

Figure 8 Violin plot of the projects final scores, with an internal box plot.

The final rank of a building is determined by meeting a number of minimum criteria and points. There is no direct conversion from points to rank or vice versa. Moreover, besides the star-based system detailed in the technical guide for green building [6], the dataset also contains instances with textual ranks. "*green*", "*excellent*" and "*excellent green*" are all appearing in the data with no mention in the guide and no direct conversion to the star-based system. Despite this, we elected to keep these records and include them in the analysis.

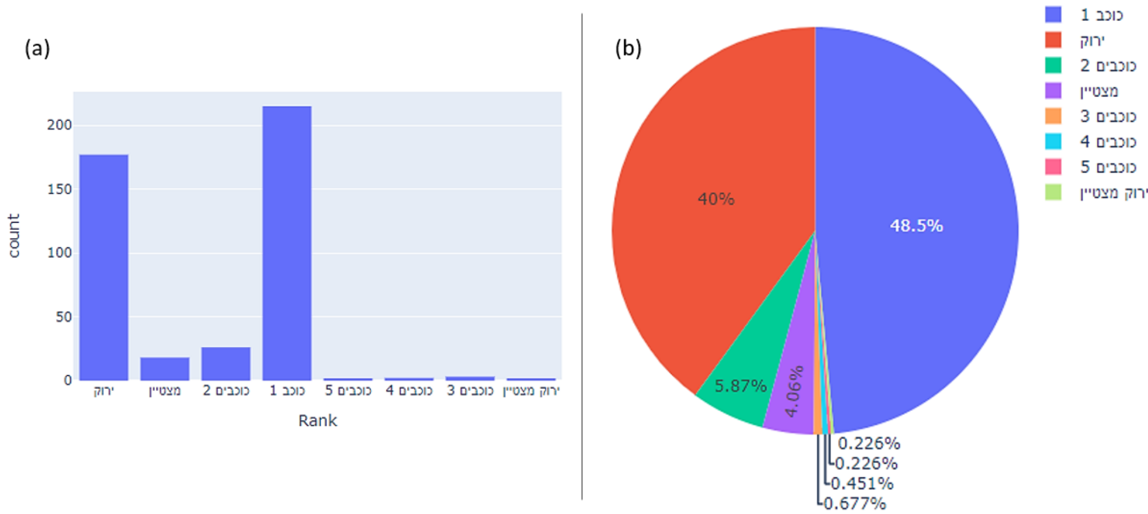

Figure 9 (a) Histogram of the ranks. (b) Pie chart displaying the relative part of each rank.

The most common rank given is 1-star, with 215 records receiving it out of 443. The second most common rating is "*green*" with 177 records. The maximum rank of 5-stars was awarded to only a single project. This led us to believe that they might mean the same thing or at least are closely related. In the attached notebook [5], there are the most common scores per year, beginning with "*green*" and changing to 1-star. Given that the scores are low for almost all the projects and given the previous information the argument that "*green*" equals 1-star is more robust. The difference may be due to the changes made to SI-5281. Figure 9 shows a histogram of the ranks and a pie chart displaying the relative part of each rank.

Projects can be a construction of a new building or a renovation of an existing one. The most common is a new construction project, with 803 records out of 823. This shows that most of the sustainable buildings being certified are new. In the attached notebook [5], there are the most common type per year.

Israeli CBS Data Correlation to Sustainable Building - Multiple Linear Regression

In this chapter we demonstrate how connections between demographic data and sustainable building can be discovered and explored. We selected 15 different metrics from the Israeli CBS dataset [4]. Using multiple linear regression, we modeled the connection between the metrics and the amount of sustainable building projects per local authority. This method produced a regression line with R-squared of 0.441 and Adj. R-squared of 0.246.

Table 3 shows the metrics used, their learned coefficients and their p-values. We can see that out of the 15 metrics, two have statistically significant effect on the number of sustainable projects. *"Employees Inequality Index (Gini Index, 0 is complete equality)"* indicates that there is a positive correlation between the financial equality within a local authority to the number of sustainable projects in that authority. "*Percentage of students dropping out*" show similar correlation. This method can be used to identify metric that will help shape profiles of local authorities in relation to the volume of sustainable building. This can help the government focus on assisting and expanding sustainable projects in authorities with less potential of initiating them. This will help increase the overall sustainable buildings

Metric	coefficient	P_{value}
Distance from Tel-Aviv District Border Line (km)	-0.2696	0.065616
Area (square km)	0.0011	0.851042
Population	-3e-05	0.825539
Average Daily Unemployment Benefit	-0.3232	0.638751
Average Monthly Wage of Employees	-0.0149	0.076519
Number of Employees	-0.0002	0.670295
Percentage of Employees Earning Up to Minimum Wage	-2.8696	0.064682
Number of Self-Employed	0.0029	0.187164
Average Monthly Wage of Self-Employed	0.0024	0.428941
Percentage of Self-Employed Earning Up to Half the Average Wage	-0.8047	0.516217
Employees Inequality Index (Gini Index, 0 is complete equality)	**528.7206**	**0.035294**
Percentage of students dropping out	**19.7106**	**0.022375**
Percentage of matriculation certificate eligible among 12th graders	0.6327	0.103564
Percentage of students out of the general population	-1.2413	0.583207
Population density per square kilometer	0.0001	0.862167

Table 3 Multiple linear regression results. Metrics, their learned coefficients and their p-values.

Bibliography

1. The standard institution of israel. אתר לצפייה בתקנים. (n.d.). Retrieved February 28, 2022, from https://ibr.sii.org.il/ibr/#/standards/7
2. Green building certification. Gov.il. (n.d.). Retrieved February 28, 2022, from https://www.gov.il/en/departments/guides/green_building_in_israel?chapterIndex=3
3. Green structures in Israel. Gov.il. (n.d.). Retrieved February 28, 2022, from https://www.gov.il/he/departments/guides/green_building_in_israel?chapterIndex=1
4. קובץ הרשויות המקומיות בישראל - 2019. הלשכה המרכזית לסטטיסטיקה. (n.d.). Retrieved February 28, 2022, from https://www.cbs.gov.il/he/mediarelease/pages/2021/%D7%A7%D7%95%D7%91%D7%A5-%D7%94%D7%A8%D7%A9%D7%95%D7%99%D7%95%D7%AA-%D7%94%D7%9E%D7%A7%D7%95%D7%9E%D7%99%D7%95%D7%AA-%D7%91%D7%99%D7%A9%D7%A8%D7%90%D7%9C-2019.aspx#losExcelos
5. Google. (n.d.). Google colaboratory. Google Colab. Retrieved February 28, 2022, from https://colab.research.google.com/drive/1mx8ZBJHDv09uqqGYVVlkM72-6OgxTvy7?usp=sharing
6. Technical guide for green building. Gov.il. (n.d.). Retrieved February 28, 2022, from https://www.gov.il/he/departments/publications/reports/technical_guide_for_green_building