Building a more energy efficient Chicago together

Ada Jing, Sushmita Singha, Bibind Vasu

Executive Summary

- **Goal:**
  - Identify relevant, granulated information from each neighborhood to determine their potential of benefiting from a city-funded energy efficiency program
- **Implementation:**
  - Extracting, cleaning, combining, and analyzing various public datasets from City of Chicago
- **Outcome:**
  - An analytical tool that can be used by the City of Chicago to estimate a neighborhood's potential in energy savings

Executive Summary

The context:
As part of the Sustainable Chicago Action Agenda launched in 2012, the City of Chicago is looking to improve Energy Efficiency in its communities.

One of the most important means in this improvement is to **subsidize poorer communities** in its transition into implementing energy efficiency.

The question:
With limited amount subsidy budget and manpower every year, **which neighborhood should be prioritized?**

The solution:
Develop a scoring system that will help the city identify the neighborhood that will benefit the most. The system will take into following factors in each community, as divided by community areas.

Summary | Data source | Analysis/exploration | Learning

Executive Summary

The context:
As part of the Sustainable Chicago Action Agenda launched in 2012, the City of Chicago is looking to improve Energy Efficiency in its communities.

One of the most important means in this improvement is to **subsidize communities** in its transition into implementing energy efficient appliances.

The question:
With limited amount subsidy budget and manpower every year, **which neighborhood should be prioritized?**

The solution:
Develop a scoring system that will help the city identify the neighborhood that will benefit the most. The system will take into following factors in each community, as divided by community areas.

Executive Summary

The context:
As part of the Sustainable Chicago Action Agenda launched in 2012, the City of Chicago is looking to improve Energy Efficiency in its communities.

One of the most important means in this improvement is to **subsidize communities** in its transition into implementing energy efficient appliances.

The question:
With limited amount subsidy budget and manpower every year, **which neighborhood should be prioritized?**

The solution:
Develop a scoring system that will help the city identify the neighborhood that will **benefit the most** from the subsidy program.

Tools and data

Data cleaning

Data integration

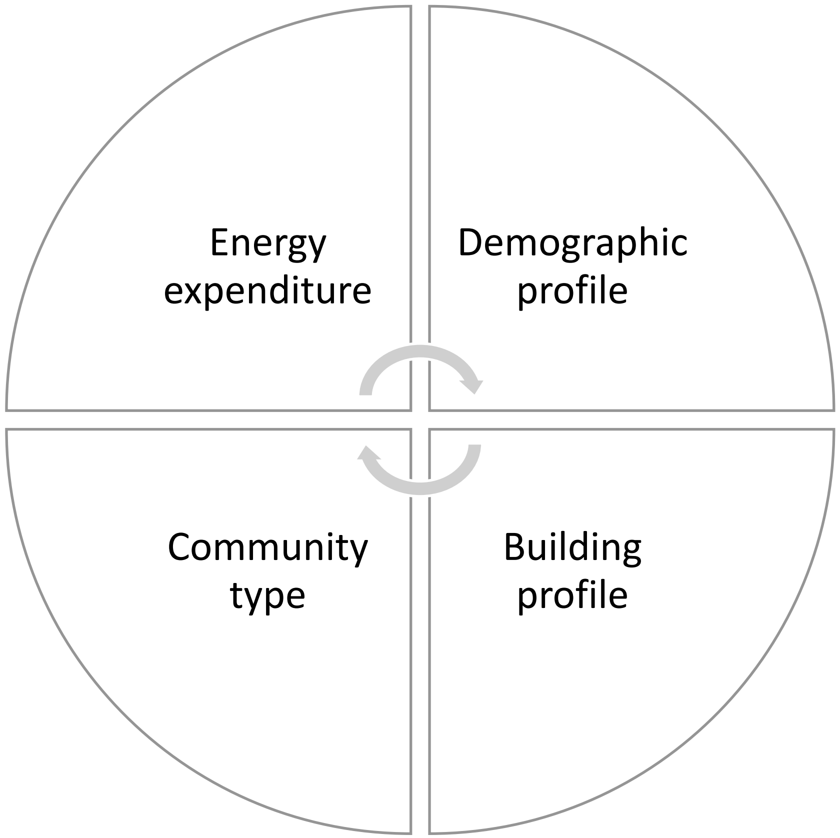
Data collection → ETL → Data wrangling → Data analysis → Data visualization → Reporting

tableau

MySQL

Google Data Studio

Four factors

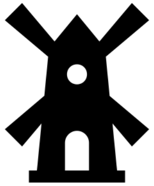


Summary | Data source | Analysis/exploration | Learning

Data source

How energy efficient are the existing buildings in the neighborhood?

Energy_star_score
0-100



Site_EUI/Source EUI
Energy per square foot per year

Electricity_usage
kbtu

Natural_gas_usage
kbtu

Data source

What type of buildings are in the neighborhood?

<u>Building_type/subtype</u>
Residential/commercial
Mall, hospital, school, etc

<u>Year_built</u>
year

<u>Location</u>
Community area
Zip code
Longitude
Latitude
address

Data source

Is this community socio-economically challenged to proactively implement energy efficiency programs?

Income_per_capita

No High School Diploma

Age

Race

Rental profile

Health profile

Unemployment rate

Data preparation/analysis

Loading/analyzing data using R

```r
#normalize the table
present_norm <- as.data.frame(apply(e_2018_1[, 3:10], 2, function(x) (x - min(x))/(max(x)-min(x))))
#reverse variables
present_norm1 <- as.data.frame(apply(present_norm[, c(4,7,8)], 2, function(x) (1-x)))
#binding tables
e_2018_2<-cbind(e_2018_1[,0:2],present_norm[,c(0,1,2,3,5,6)],present_norm1)
head(e_2018_2)
write.csv(e_2018_2,"/Users/apple/Desktop/e_2018.csv")

con <- dbConnect(MySQL(), host = "104.197.215.141", port= 3306,user = "root", password = "rootroot",db="energydb")
metrics <- dbListTables(con)
dbSendQuery(con, "SET GLOBAL local_infile = true;")
dbWriteTable(con, "energy_demographics", value = read.csv("/Users/apple/Desktop/energy-demographics.csv"))
dbWriteTable(con, "2018_normalized_energy_demographics", value = read.csv("/Users/apple/Desktop/e_2018.csv"))
```

Loading data using GCP

Resources

Storage

2 buckets

SQL

1 instance

Data manipulation using Excel

SUM =K2*0.2+J2*0.2+I2*0.1+H2*0.1+G2*0.1+F2*0.1+E2*0.1+D2

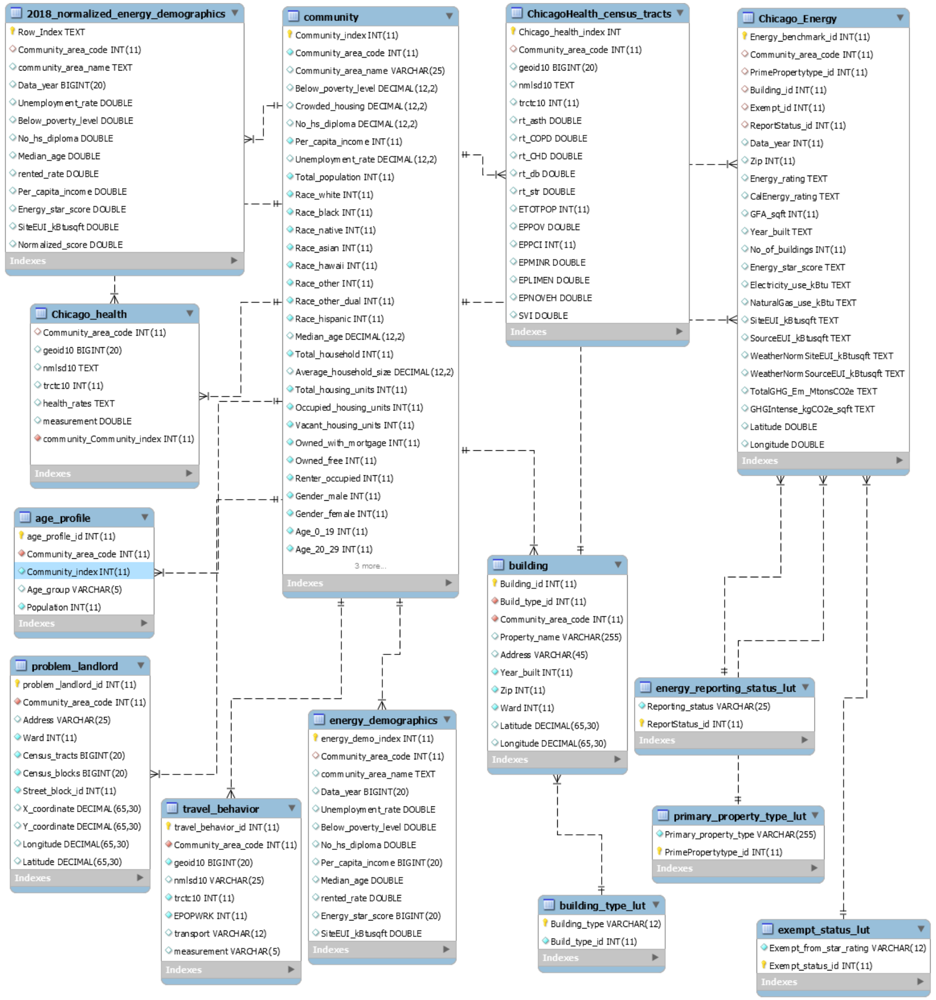
	A	B	C	D	E	F	G	H	I	J	K	L
1		community_	Data_year	Unemployme	Below_pover	No_hs_diplo	Median_age	rented_rate	Per_capita_i	Energy_star	SiteEUI_kBtu	Normalized_score
2	5	Albany Park	2018	0.20361991	0.32162162	0.55597015	0.24742268	0.68298969	0.76396696	0.875	0.65638348	0.1+D2
3	10	Archer Heigh	2018	0.43891403	0.21081081	0.58395522	0.21649485	0.41666667	0.86685566	0.30208333	0.77978168	0.489742726
4	15	Armour Squa	2018	0.32126697	0.82702703	0.60447761	0.7628866	0.7545819	0.84737768	0.4375	0.64451827	0.628165433
5	20	Ashburn	2018	0.19457014	0.11621622	0.24626866	0.37113402	0	0.72185835	0.80208333	0.71191267	0.467803939
6	24	Auburn Gres	2018	0.89140272	0.52162162	0.26865672	0.55154639	0.56414662	0.86986168	0.63541667	0.67062174	0.627931255
7	28	Austin	2018	0.74660634	0.58918919	0.37126866	0.3556701	0.66079611	0.87235447	0.32291667	0.74655909	0.573483637
8	31	Avalon Park	2018	0.54751131	0.31081081	0.15298508	0.8556701	0.22551546	0.68722812	0.14583333	0.83056478	0.473251711
9	35	Avondale	2018	0.21719457	0.25405405	0.38432836	0.25257732	0.70418099	0.76069212	0.75	0.53298529	0.513899798
10	40	Belmont Cra	2018	0.31674208	0.36216216	0.59514925	0.20618557	0.54682131	0.88882643	0.28125	0.8372093	0.515280541
11	44	Beverly	2018	0.14932127	0	0	0.75257732	0.07660367	0.28124542	0.29166667	0.82249644	0.348807388
12	48	Bridgeport	2018	0.30316742	0.32702703	0.38246269	0.45360825	0.63831615	0.65120485	0.69791667	0.65353583	0.545869138
13	53	Brighton Par	2018	0.30316742	0.48108108	0.80410448	0.04639175	0.60337915	0.9403441	0.54166667	0.69814903	0.565809937

Data aggregation using SQL

Result Grid Filter Rows: Search Export:

Community_area_name	Energy_star_score	Unemployment_rate	Per_capita_income	Crowded_housing	Below_poverty_level	SVI
Albany Park	69	9.00	20355	11.20	17.10	0.7983
Archer Heights	68	14.20	16145	8.50	13.00	0.6219
Armour Square	55	11.60	16942	5.90	35.80	0.974
Ashburn	20	8.80	22078	4.20	9.50	0.3807
Auburn Gresham	36	24.20	16022	4.10	24.50	0.7141
Austin	66	21.00	15920	5.70	27.00	0.9537
Avalon Park	83	16.60	23495	0.60	16.70	0.7642
Avondale	25	9.30	20489	5.80	14.60	0.5837
Belmont Cragin	70	11.50	15246	10.00	18.60	0.7857
Beverly	69	7.80	40107	0.70	5.20	0.0206
Bridgeport	30	11.20	24969	4.80	17.30	0.7931
Brighton Park	45	11.20	13138	13.20	23.00	0.7729
Burnside	70	23.40	13756	5.50	22.50	0.9155
Calumet Heights	64	17.20	28977	1.80	12.00	0.5262
Chatham	39	19.00	20320	2.20	25.30	0.6765
Chicago Lawn	96	11.90	14405	6.50	22.20	0.9965
Clearing	65	9.60	23920	3.40	5.90	0.5686
Douglas	15	16.70	23098	1.60	26.10	0.5997
Dunning	41	8.60	26347	4.80	8.30	0.4963

EER



2018_normalized_energy_demographics
Row_Index TEXT
Community_area_code INT(11)
community_area_name TEXT
Data_year BIGINT(20)
Unemployment_rate DOUBLE
Below_poverty_level DOUBLE
No_hs_diploma DOUBLE
Median_age DOUBLE
rented_rate DOUBLE
Per_capita_income DOUBLE
Energy_star_score DOUBLE
SiteEUI_kBtusqft DOUBLE
Normalized_score DOUBLE
Indexes
community
Community_index INT(11)
Community_area_code INT(11)
Community_area_name VARCHAR(25)
Below_poverty_level DECIMAL(12,2)
Crowded_housing DECIMAL(12,2)
No_hs_diploma DECIMAL(12,2)
Per_capita_income INT(11)
Unemployment_rate DECIMAL(12,2)
Total_population INT(11)
Race_white INT(11)
Race_black INT(11)
Race_native INT(11)
Race_asian INT(11)
Race_hawaii INT(11)
Race_other INT(11)
Race_other_dual INT(11)
Race_hispanic INT(11)
Median_age DECIMAL(12,2)
Total_household INT(11)
Average_household_size DECIMAL(12,2)
Total_housing_units INT(11)
Occupied_housing_units INT(11)
Vacant_housing_units INT(11)
Owned_with_mortgage INT(11)
Owned_free INT(11)
Renter_occupied INT(11)
Gender_male INT(11)
Gender_female INT(11)
Age_0_19 INT(11)
Age_20_29 INT(11)
3 more...
Indexes
ChicagoHealth_census_tracts
Chicago_health_index INT
Community_area_code INT(11)
geoid10 BIGINT(20)
nmlsd10 TEXT
trctc10 INT(11)
rt_asth DOUBLE
rt_COPD DOUBLE
rt_CHD DOUBLE
rt_db DOUBLE
rt_str DOUBLE
ETOTPOP INT(11)
EPPOV DOUBLE
EPPCI INT(11)
EPMINR DOUBLE
EPLIMEN DOUBLE
EPNOVEH DOUBLE
SVI DOUBLE
Indexes
Chicago_Energy
Energy_benchmark_id INT(11)
Community_area_code INT(11)
PrimePropertytype_id INT(11)
Building_id INT(11)
Exempt_id INT(11)
ReportStatus_id INT(11)
Data_year INT(11)
Zip INT(11)
Energy_rating TEXT
CalEnergy_rating TEXT
GFA_sqft INT(11)
Year_built TEXT
No_of_buildings INT(11)
Energy_star_score TEXT
Electricity_use_kBtu TEXT
NaturalGas_use_kBtu TEXT
SiteEUI_kBtusqft TEXT
SourceEUI_kBtusqft TEXT
WeatherNormSiteEUI_kBtusqft TEXT
WeatherNormSourceEUI_kBtusqft TEXT
TotalGHG_Em_MtonsCO2e TEXT
GHGIntense_kgCO2e_sqft TEXT
Latitude DOUBLE
Longitude DOUBLE
Indexes
Chicago_health
Community_area_code INT(11)
geoid10 BIGINT(20)
nmlsd10 TEXT
trctc10 INT(11)
health_rates TEXT
measurement DOUBLE
community_Community_index INT(11)
Indexes
age_profile
age_profile_id INT(11)
Community_area_code INT(11)
Community_index INT(11)
Age_group VARCHAR(5)
Population INT(11)
Indexes
problem_landlord
problem_landlord_id INT(11)
Community_area_code INT(11)
Address VARCHAR(25)
Ward INT(11)
Census_tracts BIGINT(20)
Census_blocks BIGINT(20)
Street_block_id INT(11)
X_coordinate DECIMAL(65,30)
Y_coordinate DECIMAL(65,30)
Longitude DECIMAL(65,30)
Latitude DECIMAL(65,30)
Indexes
travel_behavior
travel_behavior_id INT(11)
Community_area_code INT(11)
geoid10 BIGINT(20)
nmlsd10 VARCHAR(25)
trctc10 INT(11)
EPOPWRK INT(11)
transport VARCHAR(12)
measurement VARCHAR(5)
Indexes
energy_demographics
energy_demo_index INT(11)
Community_area_code INT(11)
community_area_name TEXT
Data_year BIGINT(20)
Unemployment_rate DOUBLE
Below_poverty_level DOUBLE
No_hs_diploma DOUBLE
Per_capita_income BIGINT(20)
Median_age DOUBLE
rented_rate DOUBLE
Energy_star_score BIGINT(20)
SiteEUI_kBtusqft DOUBLE
Indexes
building
Building_id INT(11)
Build_type_id INT(11)
Community_area_code INT(11)
Property_name VARCHAR(255)
Address VARCHAR(45)
Year_built INT(11)
Zip INT(11)
Ward INT(11)
Latitude DECIMAL(65,30)
Longitude DECIMAL(65,30)
Indexes
building_type_lut
Building_type VARCHAR(12)
Build_type_id INT(11)
Indexes
energy_reporting_status_lut
Reporting_status VARCHAR(25)
ReportStatus_id INT(11)
Indexes
primary_property_type_lut
Primary_property_type VARCHAR(255)
PrimePropertytype_id INT(11)
Indexes
exempt_status_lut
Exempt_from_star_rating VARCHAR(12)
Exempt_status_id INT(11)
Indexes

Describe

Multi-family residential housing is one of the least efficient

```sql
SELECT
    primary_property_type_lut.Primary_property_type,
    building_type_lut.Building_type,
    e.Energy_star_score,
    e.SiteEUI_kBtusqft
FROM
    (SELECT
        Building_id,
            SiteEUI_kBtusqft,
            Energy_star_score,
            PrimePropertytype_id
    FROM
        Chicago_Energy) AS e
        JOIN
    (SELECT
        Building_id, Build_type_id
    FROM
        building) AS bu ON bu.Building_id = e.Building_id
        JOIN
    building_type_lut ON building_type_lut.Build_type_id = bu.Build_type_id
        JOIN
    primary_property_type_lut ON primary_property_type_lut.PrimePropertytype_id = e.PrimePropertytype_id
GROUP BY primary_property_type_lut.Primary_property_type,Building_type;
```

Primary_property_type	Building_type	Energy_star_sc...	SiteEUI_kBtusqft
Retail Store	Commercial	96	38.2
Wholesale Club/Supercenter	Commercial	92	23
Senior Care Community	Residential	9	170.7
Courthouse	Commercial	85	77.5
K-12 School	Commercial	8	121.9
Hotel	Residential	75	49.4
Financial Office	Commercial	74	77
Residence Hall/Dormitory	Commercial	73	68.3
Residence Hall/Dormitory	Residential	71	73.6
Office	Municipal	56	132.8
Hotel	Commercial	50	128.9
Hospital (General Medical &...	Commercial	34	251
Worship Facility	Commercial	33	71.4
Senior Care Community	Commercial	28	111.6
Multifamily Housing	Commercial	19	79.2
K-12 School	Municipal	16	109
Bank Branch	Commercial	14	141.6
Multifamily Housing	Municipal	100	7.4
Office	Commercial	100	47.3
Multifamily Housing	Residential	10	57.8

Community energy overview

Energy Snapshot of Chicago At a Glance

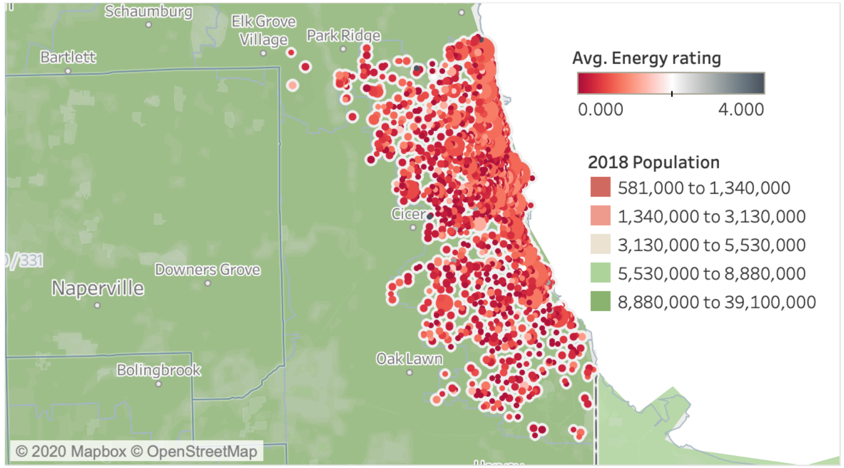
Avg. Energy star score	Avg. GFA sqft	Avg. GHGIntense kgCO2e sqft	Avg. SiteEUI kBtusqft	Avg. SourceEUI kBtusqft
46	269,263	10	96	17

Community Area High & Low Energy Rating

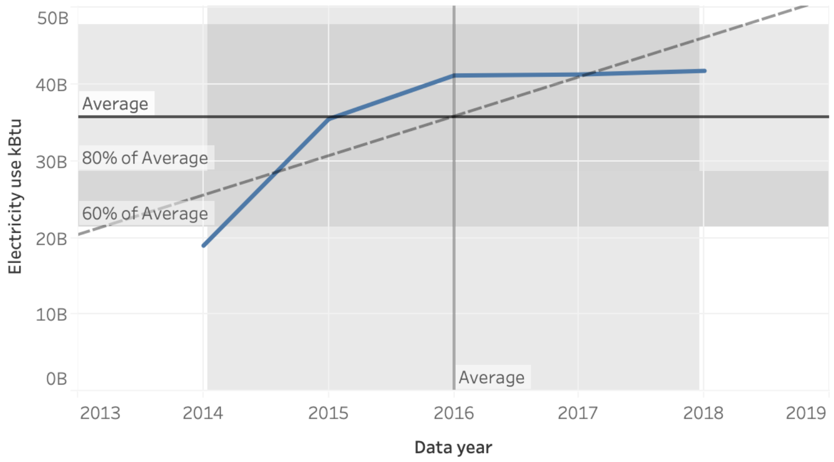
In / Out of Set 1HighStar / In / Out of Set 1LowStar

Community area name (gr..	In: Out	Out: In	Out: Out
UPTOWN	25	40	2
NEAR NORTH SIDE	25	50	2
LOOP	25	50	2
EDGEWATER	25	46	2
AVONDALE, LAKE VIEW, LI..	25	49	2
AUSTIN, EAST GARFIELD P..	25	50	2
ROGERS PARK	23	27	1
OHARE	23	16	1
HYDE PARK	23	41	2
NEAR SOUTH SIDE	20	43	2
KENWOOD	19	29	2
SOUTH SHORE	17	36	2
IRVING PARK	17	15	1
PORTAGE PARK	16	3	1
NEW CITY	16	27	
GRAND BOULEVARD	16	17	1
BELMONT CRAGIN	16	24	1
WEST RIDGE	15	33	1
DOUGLAS	14	38	1
NORTH PARK	13	9	
ARMOUR SQUARE	12	9	1
MORGAN PARK	11	4	
ALBANY PARK	11	12	1
ENGLEWOOD	10	22	1

Chicago Average EnergyRating



Electricity Consumption in Chicago



Community area name

- ✓ ALBANY PARK
- ✓ ARCHER HEIGHTS
- ✓ ARMOUR SQUARE
- ✓ ASHBURN
- ✓ AUBURN GRESHAM
- ✓ AUSTIN
- ✓ AVALON PARK
- ✓ AVONDALE
- ✓ BELMONT CRAGIN
- ✓ BEVERLY
- ✓ BRIDGEPORT
- ✓ BRIGHTON PARK
- ✓ BURNSIDE
- ✓ CALUMET HEIGHTS
- ✓ CHATHAM
- ✓ CHICAGO LAWN
- ✓ CLEARING
- ✓ DOUGLAS
- ✓ DUNNING
- ✓ EAST GARFIELD PARK
- ✓ EAST SIDE
- ✓ EDGEWATER
- ✓ EDISON PARK
- ✓ ENGLEWOOD
- ✓ FOREST GLEN
- ✓ FULLER PARK
- ✓ GAGE PARK
- ✓ GARFIELD RIDGE

Data year 2014 to 2018

Energy star score

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

https://prod-useast-a.online.tableau.com/#/site/bibindvasu/views/EP6/Dashboard1?:iid=3

Community demographics

Energy-Health-Demographics

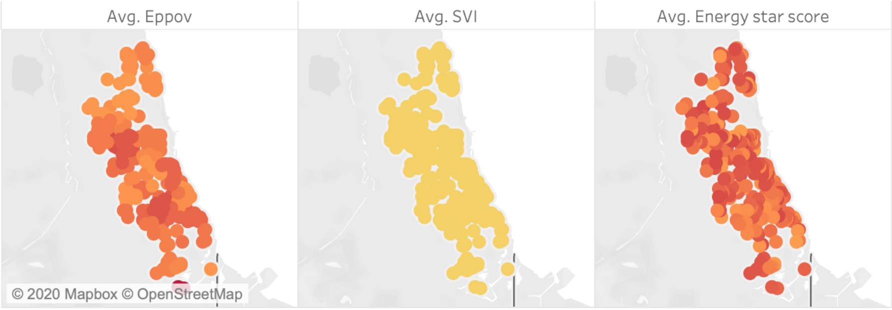
Chicago- Health Status at a Glance

Avg. Eppov	Avg. SVI	Avg. Rt Asth	Avg. rt CHD	Avg. rt COPD	Avg. Rt Db	Avg. Rt Str
19.29	0.45	9.24	4.18	4.49	8.21	2.53

Communities Above Poverty Ratio of 19%

Community area name	Avg. Ep..	Avg. SVI	Avg. En..
ENGLEWOOD	45.71	0.94	36.53
RIVERDALE	65.75	0.94	37.14
WEST GARFIELD PARK	47.65	0.93	23.38
BURNSIDE	32.80	0.92	28.00
FULLER PARK	32.05	0.91	22.33
WASHINGTON PARK	43.80	0.91	40.95
NORTH LAWNDALE	44.53	0.91	47.73
HUMBOLDT PARK	27.60	0.89	40.61
EAST GARFIELD PARK	44.05	0.89	26.42
AUSTIN	30.52	0.89	37.06
CHICAGO LAWN	32.55	0.88	40.70
SOUTH CHICAGO	32.23	0.88	41.38
NEW CITY	34.70	0.88	44.40
SOUTH LAWNDALE	34.38	0.88	37.81
SOUTH DEERING	29.23	0.87	33.51
WEST ENGLEWOOD	37.93	0.86	38.95
WOODLAWN	40.50	0.86	40.42
GAGE PARK	21.00	0.86	43.07
SOUTH SHORE	37.76	0.85	44.19
GREATER GRAND CROSSI..	37.75	0.85	32.30
BELMONT CRAGIN	[illegible]	[illegible]	[illegible]

Communities 2Invest:
Poverty Ratio of 19%, Energy Star Score <50



Communities-Vulnerability&Energy

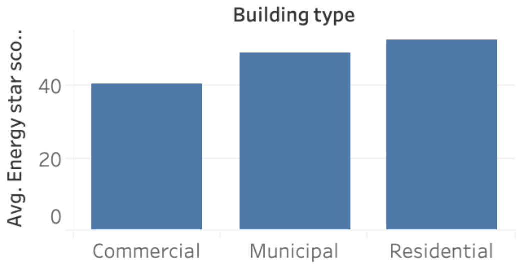
Community ..	..g. Epl..	Avg. Ep..	Avg. Ep..	Avg. Rt ..	Avg. Rt ..	Avg. Rt ..	Avg. SVI	Avg. rt ..	Avg. rt ..	Avg.
AUSTIN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
CHICAGO L..	12.76	28.04	32.55	11.68	14.23	4.32	0.88	5.98	7.38	40
SOUTH CHI..	5.88	37.07	32.23	12.18	17.51	5.65	0.88	7.73	8.46	41
NEW CITY	15.47	28.62	34.70	11.34	14.31	4.38	0.88	6.29	7.67	44
SOUTH LA..	25.19	22.55	34.38	10.10	13.55	3.54	0.88	6.18	6.74	37
SOUTH DEE..	5.35	20.38	29.23	11.43	16.83	5.20	0.87	7.63	8.03	33
WEST ENGL..	1.69	35.36	37.93	13.48	18.72	6.63	0.86	7.86	9.44	38
WOODLAWN	0.91	44.23	40.50	12.48	16.14	5.51	0.86	6.67	7.75	40
GAGE PARK	23.32	15.95	21.00	9.71	11.77	2.99	0.86	5.30	6.00	43
SOUTH SHO..	0.70	42.26	37.76	12.63	17.29	5.78	0.85	7.08	7.97	44
GREATER G..	0.84	39.60	37.75	13.06	17.77	6.19	0.85	7.47	8.51	32
BELMONT C..	24.49	14.90	19.80	9.33	11.04	2.84	0.83	5.44	5.89	49
AUBURN G..	0.25	33.69	31.41	12.94	18.53	6.29	0.82	7.79	8.68	30
HERMOSA	25.80	18.12	22.57	9.43	11.62	2.80	0.81	5.35	5.70	42
WEST RIDGE	11.26	16.19	22.44	9.07	10.56	3.09	0.79	5.87	6.10	44
WEST PULL..	2.24	18.67	28.48	11.10	15.56	5.00	0.78	6.40	7.17	29
MCKINLEY ..	17.94	15.40	25.28	8.88	11.32	2.86	0.77	5.56	5.74	45
CHATHAM	0.36	33.80	28.16	12.30	17.96	5.99	0.76	7.53	7.91	33
PULLMAN	2.80	26.70	24.80	11.70	14.67	4.57	0.76	6.43	6.97	17

https://prod-useast-a.online.tableau.com/#/site/bibindvasu/views/EP5/Dashboard1?:iid=4

Energy Efficiency by property type

EnergyAnalysis -Buildingwise

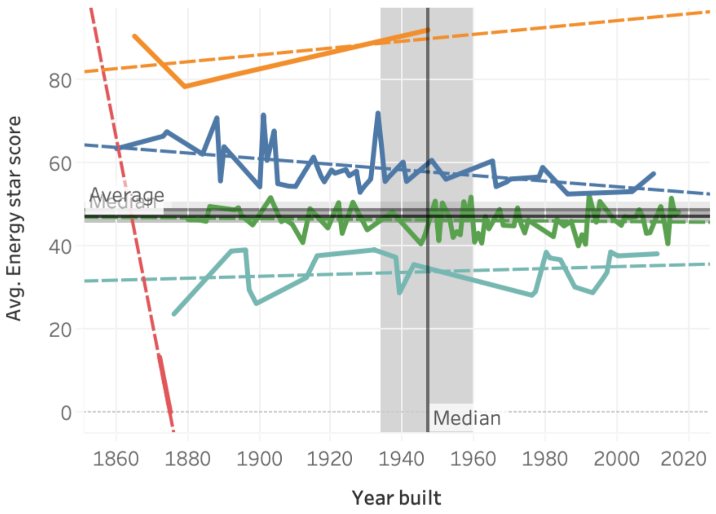
BuildingClassEnergyScore



CommunityHouseholdTypes

Community area name	Owned free	Owned wit..	Renter occ..	Total hous..
Near North Side	14,677,950	42,230,650	69,790,700	126,699,300
Lake View	2,647,392	12,629,760	26,031,423	41,308,575
Near West Side	1,042,144	10,856,041	17,634,469	29,532,654
The Loop	2,281,474	9,609,449	17,036,375	28,927,298
Lincoln Park	1,851,875	7,130,625	11,977,500	20,960,000
West Town	707,328	4,696,704	9,260,928	14,664,960

Year Built Vs Energy Score Clusters



BuildingEnergy

	Building type		
Primary property type	Commercial	Municipal	Residential
Office	62,337,712,187	550,829,234	69,878,865
Hospital (General Medical..	15,147,522,164		
Mixed Use Property	7,401,690,326		463,503,646
College/University	5,066,922,119	1,282,158,311	
Convention Center	4,108,558,777		
Laboratory	3,492,754,880		
Retail Store	2,813,214,213		
Supermarket/Grocery Sto..	2,503,678,923		
Financial Office	2,184,767,836		
Prison/Incarceration	1,556,606,189		
Enclosed Mall	1,502,775,656		
Medical Office	1,252,035,892		
Other	1,043,877,158	13,741,949	11,893,587
Strip Mall	1,022,529,643		
K-12 School	980,692,937	7,268,376,650	

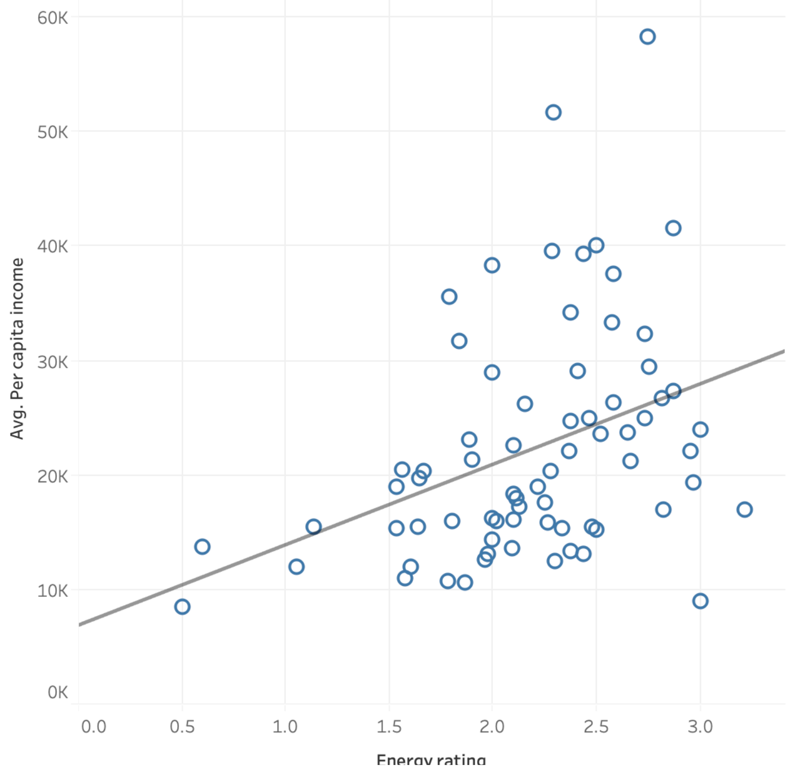
Sum of NaturalGas use ..
0 68B

https://prod-useast-a.online.tableau.com/#/site/bibindvasu/views/EP8/Dashboard1_4?:iid=1

Explore

Higher income communities are more energy efficient

Income/energy rating per community



Avg. Per capita income = 7021.27*Avg. CalEnergy rating (copy) + 6872.51
R-Squared: 0.139667
P-value: 0.0013259

Excluded outliers

Lifting the High Energy Burden in America's Largest Cities: How Energy Efficiency Can Improve Low Income and Underserved Communities

This report provides a snapshot of energy burdens in cities across the US. The authors focus on the high home energy burdens faced by select groups in major metropolitan areas. Years of analysis by the firm of Fisher Sheehan & Colton determined that low-income households pay proportionally more than the average household for energy costs. This analysis builds on this research as it

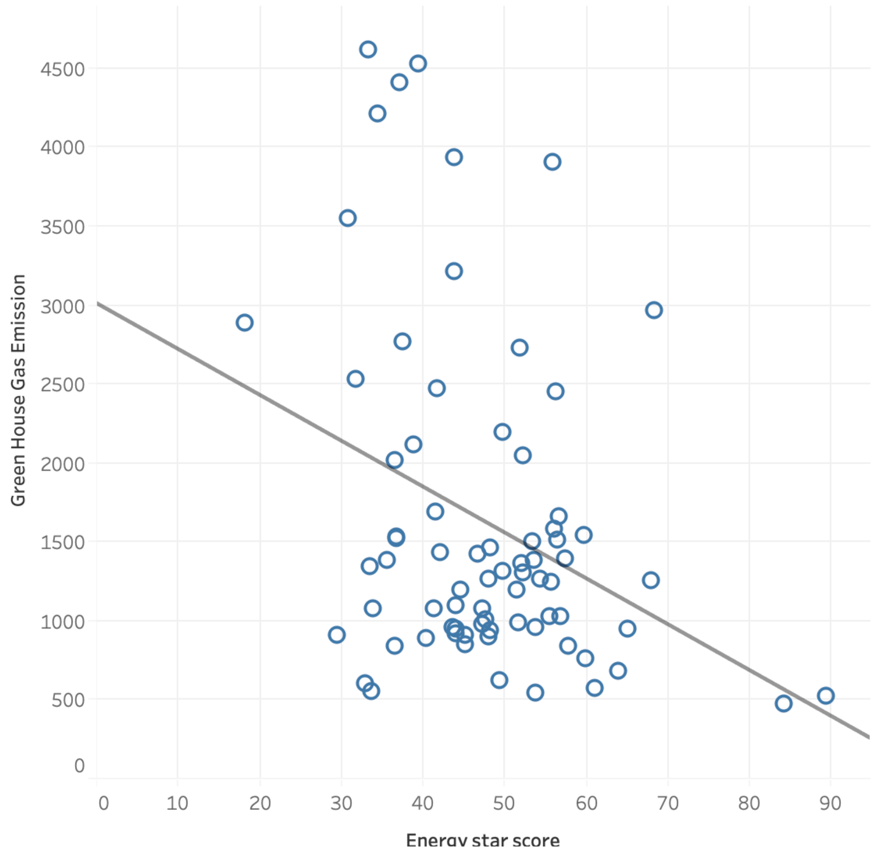
Research

Impacts of energy-efficiency investments on internal conditions in low-income households

Wouter Poortinga, Shiyu Jiang, Charlotte Grey & Chris Tweed

Pages 653-667 | Published online: 27 Apr 2017

Energy inefficiency linked to high GHG emissions



Avg. TotalGHG Em MtonsCO2e = -29.0565*Avg. Energy_star_score + 3011.16
R-Squared: 0.110086
P-value: 0.0038785

Energy efficiency improvement opportunities and associated greenhouse gas abatement costs for the residential sector

Summary | Data source | Analysis/exploration | Learning

Excluded outliers

Discover

Combining multiple factors

```sql
#community and energy profile
SELECT
        c.community_area_name,
        e.Data_year,
        c.Unemployment_rate,
        c.Below_poverty_level,
        c.No_hs_diploma,
        c.Per_capita_income,
        c.Median_age,
        c.Renter_occupied / Occupied_housing_units AS rented_rate,
        e.Energy_star_score,
        e.SiteEUI_kBtusqft
    FROM
        building AS b
            JOIN
        Chicago_Energy AS e ON b.Building_id = e.Building_id
            JOIN
        community AS c ON c.Community_area_code = b.Community_area_code
            JOIN
        exempt_status_lut ON exempt_status_lut.Exempt_status_id = e.Exempt_id
    WHERE
        Exempt_from_star_rating = FALSE
    GROUP BY c.community_area_name,e.Data_year;
```

community_area_name	Data_year	Unemployment_rate	Below_poverty_level	No_hs_diploma	Per_capita_income	Median_age	rented_rate	Energy_star_score	SiteEUI_kBtusqft
Albany Park	2014	9	17.1	34.9	20355	31.1	0.6113	71	80
Albany Park	2015	9	17.1	34.9	20355	31.1	0.6113	64	64.7
Albany Park	2016	9	17.1	34.9	20355	31.1	0.6113	NULL	172.7
Albany Park	2017	9	17.1	34.9	20355	31.1	0.6113	67	60.9
Albany Park	2018	9	17.1	34.9	20355	31.1	0.6113	13	102.7
Archer Heights	2014	14.2	13	36.4	16145	30.5	0.4253	19	96
Archer Heights	2015	14.2	13	36.4	16145	30.5	0.4253	32	88.7
Archer Heights	2016	14.2	13	36.4	16145	30.5	0.4253	68	77.7
Archer Heights	2017	14.2	13	36.4	16145	30.5	0.4253	82	73
Archer Heights	2018	14.2	13	36.4	16145	30.5	0.4253	68	76.7
Armour Square	2014	11.6	35.8	37.5	16942	41.1	0.6613	NULL	90
Armour Square	2015	11.6	35.8	37.5	16942	41.1	0.6613	92	46.1
Armour Square	2016	11.6	35.8	37.5	16942	41.1	0.6613	NULL	91.2
Armour Square	2017	11.6	35.8	37.5	16942	41.1	0.6613	83	45.6
Armour Square	2018	11.6	35.8	37.5	16942	41.1	0.6613	55	105.2
Ashburn	2014	8.8	9.5	18.3	22078	33.5	0.1343	68	84
Ashburn	2015	8.8	9.5	18.3	22078	33.5	0.1343	64	71.5
Ashburn	2016	8.8	9.5	18.3	22078	33.5	0.1343	55	51
Ashburn	2017	8.8	9.5	18.3	22078	33.5	0.1343	46	57
Ashburn	2018	8.8	9.5	18.3	22078	33.5	0.1343	20	91
Auburn Gresham	2015	24.2	24.5	19.5	16022	37	0.5283	60	73.3
Auburn Gresham	2016	24.2	24.5	19.5	16022	37	0.5283	38	84.4
Auburn Gresham	2017	24.2	24.5	19.5	16022	37	0.5283	46	76.9
Auburn Gresham	2018	24.2	24.5	19.5	16022	37	0.5283	36	99.7

Summary | Data source | Analysis/exploration | Learning

Measuring on the same level field

Min-max normalization for each measurement

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Assigning weight for each normalized measurement

$$\bar{x} = \sum_{i=1}^{n} w'_i x_i = \sum_{i=1}^{n} \frac{w_i}{\sum_{j=1}^{n} w_j} x_i = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{j=1}^{n} w_j}$$
$$= \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}.$$

W=10% W=20%

community_area_name	Data_year	Unemployment_rate	Below_poverty_level	No_hs_diploma	Median_age	rented_rate	Per_capita_income	Energy_star_score	SiteEUI_kBtusqft	Normalized_score
Albany Park	2018	0.20361991	0.321621622	0.555970149	0.24742268	0.682989691	0.763966958	0.875	0.656383484	0.583835798
Archer Heights	2018	0.438914027	0.210810811	0.583955224	0.216494845	0.416666667	0.866855663	0.302083333	0.77978168	0.489742726
Armour Square	2018	0.321266968	0.827027027	0.604477612	0.762886598	0.754581901	0.847377682	0.4375	0.644518272	0.628165433
Ashburn	2018	0.194570136	0.116216216	0.246268657	0.371134021	0	0.721858351	0.802083333	0.711912672	0.467803939
Auburn Gresham	2018	0.891402715	0.521621622	0.268656716	0.551546392	0.564146621	0.869861675	0.635416667	0.670621737	0.627931255
Austin	2018	0.746606335	0.589189189	0.371268657	0.355670103	0.660796105	0.872354465	0.322916667	0.746559089	0.573483637
Avalon Park	2018	0.547511312	0.310810811	0.152985075	0.855670103	0.225515464	0.687228115	0.145833333	0.830564784	0.473251711
Avondale	2018	0.21719457	0.254054054	0.384328358	0.25257732	0.704180985	0.760692116	0.75	0.532985287	0.513899798
Belmont Cragin	2018	0.316742081	0.362162162	0.595149254	0.206185567	0.546821306	0.888826433	0.28125	0.837209302	0.515280541
Beverly	2018	0.149321267	0	0	0.75257732	0.076603666	0.281245418	0.291666667	0.82249644	0.348807388
Bridgeport	2018	0.303167421	0.327027027	0.382462687	0.453608247	0.638316151	0.651204849	0.697916667	0.653535833	0.545869138
Brighton Park	2018	0.303167421	0.481081081	0.804104478	0.046391753	0.603379152	0.940344103	0.541666667	0.698149027	0.565809937

Weighting adjustment

Min-max normalization for each measurement

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Assigning weight for each normalized measurement

$$\bar{x} = \sum_{i=1}^{n} w'_i x_i = \sum_{i=1}^{n} \frac{w_i}{\sum_{j=1}^{n} w_j} x_i = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{j=1}^{n} w_j}$$

$$= \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}.$$

W=10%

W=20%

community_area_name	Data_year	Unemployment_rate	Below_poverty_level	No_hs_diploma	Median_age	rented_rate	Per_capita_income	Energy_star_score	SiteEUI_kBtusqft	Normalized_score
Albany Park	2018	0.20361991	0.321621622	0.555970149	0.24742268	0.682989691	0.763966958	0.875	0.656383484	0.583835798
Archer Heights	2018	0.438914027	0.210810811	0.583955224	0.216494845	0.416666667	0.866855663	0.302083333	0.77978168	0.489742726
Armour Square	2018	0.321266968	0.827027027	0.604477612	0.762886598	0.754581901	0.847377682	0.4375	0.644518272	0.628165433
Ashburn	2018	0.194570136	0.116216216	0.246268657	0.371134021	0	0.721858351	0.802083333	0.711912672	0.467803939
Auburn Gresham	2018	0.891402715	0.521621622	0.268656716	0.551546392	0.564146621	0.869861675	0.635416667	0.670621737	0.627931255
Austin	2018	0.746606335	0.589189189	0.371268657	0.355670103	0.660796105	0.872354465	0.322916667	0.746559089	0.573483637
Avalon Park	2018	0.547511312	0.310810811	0.152985075	0.855670103	0.225515464	0.687228115	0.145833333	0.830564784	0.473251711
Avondale	2018	0.21719457	0.254054054	0.384328358	0.25257732	0.704180985	0.760692116	0.75	0.532985287	0.513899798
Belmont Cragin	2018	0.316742081	0.362162162	0.595149254	0.206185567	0.546821306	0.888826433	0.28125	0.837209302	0.515280541
Beverly	2018	0.149321267	0	0	0.75257732	0.076603666	0.281245418	0.291666667	0.82249644	0.348807388
Bridgeport	2018	0.303167421	0.327027027	0.382462687	0.453608247	0.638316151	0.651204849	0.697916667	0.653535833	0.545869138
Brighton Park	2018	0.303167421	0.481081081	0.804104478	0.046391753	0.603379152	0.940344103	0.541666667	0.698149027	0.565809937

Summary | Data source | Analysis/exploration | Learning

Potential

Key takeaways

- Among property types, multi-family residential perform worst with the lowest average energy star score in Chicago.
- Understanding the City's progress in energy efficiency starts with comprehensive data collection, and we are currently facing the challenge of lacking continuous data to demonstrate progress for individual buildings.
  - Additional opportunity to quantify the progress with energy expense per neighborhood.
- For Chicago, on a community level, social-economical indicators, such as per capita income, education, along with the demographic makeup all appear to contribute to energy efficiency.
  - To establish a more accurate understanding of how all the factors interact to impact the benefit of energy efficiency transformation, the City can explore using PCA.

Non-exempt buildings that haven't reported energy benchmarking reports in 2018

```sql
SELECT
        c.community_area_name,
        b.Property_name,
        e.Data_year,
        e.Energy_star_score,
        energy_reporting_status_lut.Reporting_status
    FROM
        (SELECT
        Building_id,Property_name,Community_area_code
    FROM
        building) AS b
            JOIN
        (SELECT
        Building_id,
            Energy_star_score,
            Data_year,
            ReportStatus_id
    FROM
        Chicago_Energy) AS e ON b.Building_id = e.Building_id
            JOIN
            (SELECT
            Community_area_name,
            Community_area_code
    FROM
        community) AS c ON c.Community_area_code = b.Community_area_code
            JOIN
        energy_reporting_status_lut on energy_reporting_status_lut.ReportStatus_id=e.ReportStatus_id
    WHERE
        energy_reporting_status_lut.Reporting_status="not submitted" AND e.Data_year="2018"
    GROUP BY c.community_area_name,e.Data_year;
```

Community_area_name	Property_name	Data_year	Energy_star_sc...	Reporting_status
Archer Heights	Acero Schools - Veterans Memorial Campus	2018		Not Submitted
Avondale	Scientific Games-2727 Roscoe	2018		Not Submitted
Belmont Cragin	Acero Schools - Roberto Clemente	2018		Not Submitted
Bridgeport	Union Lofts Condominium Association (3500 S....	2018		Not Submitted
Brighton Park	Acero School - Officer Donald J. Marquez	2018		Not Submitted
Burnside	Greenwood Place	2018		Not Submitted
Chatham	Sherman Plaza	2018		Not Submitted
Douglas	Stratford at South Commons Condominiums	2018		Not Submitted
Dunning	Ridgemoor Estates 1 Condominium	2018		Not Submitted
East Garfield Park	Garfield Park Conservatory, Fieldhouse & Trades	2018		Not Submitted
Edgewater	Edgewater Presbyterian Church	2018		Not Submitted
Gage Park	3594	2018		Not Submitted
Grand Boulevard	IRG Bronzeville	2018		Not Submitted
Greater Grand Crossing	STEWART BUSINESS CENTER	2018		Not Submitted
Hermosa	Hall Plaza East	2018		Not Submitted
Humboldt Park	Franklin Blvd Partners, LLC	2018		Not Submitted
Irving Park	DePaul College Prep	2018		Not Submitted
Kenwood	811 E. 46th	2018		Not Submitted
Lake View	451 W. Melrose	2018		Not Submitted
Lincoln Park	Lincoln Park Commons Condominium Association	2018		Not Submitted

Deeper dive into the socio-economic status of community and energy efficiency

```sql
SELECT
    c.community_area_name,
    e.Energy_star_score,
    c.Unemployment_rate,
    c.Per_capita_income,
    c.Crowded_housing,
    c.Below_poverty_level,
    hc.SVI
FROM
    (SELECT
        Building_id, Property_name, Community_area_code
    FROM
        building) AS b
        JOIN
    (SELECT
        Building_id,
            Energy_star_score,
            Data_year,
            TotalGHG_Em_MtonsCO2e,
            SiteEUI_kBtusqft,
            ReportStatus_id   FROM
        Chicago_Energy) AS e ON b.Building_id = e.Building_id
        JOIN
    (SELECT
        Community_area_name,
            Community_area_code,
            Unemployment_rate,
            Per_capita_income,
            Crowded_housing,
            Below_poverty_level
    FROM
        community) AS c ON c.Community_area_code = b.Community_area_code
        JOIN
    Chicago_health AS h ON h.Community_area_code = c.Community_area_code
    join
    ChicagoHealth_census_tracts as hc on hc.geoid10=h.geoid10
        JOIN
    energy_reporting_status_lut ON energy_reporting_status_lut.ReportStatus_id = e.ReportStatus_id
WHERE
    energy_reporting_status_lut.Reporting_status = 'Submitted'
        AND e.Data_year = '2018'
GROUP BY c.community_area_name , e.Data_year;
```

Community_area_name	Energy_star_score	Unemployment_rate	Per_capita_income	Crowded_housing	Below_poverty_level	SVI
Albany Park	69	9.00	20355	11.20	17.10	0.7983
Archer Heights	68	14.20	16145	8.50	13.00	0.6219
Armour Square	55	11.60	16942	5.90	35.80	0.974
Ashburn	20	8.80	22078	4.20	9.50	0.3807
Auburn Gresham	36	24.20	16022	4.10	24.50	0.7141
Austin	66	21.00	15920	5.70	27.00	0.9537
Avalon Park	83	16.60	23495	0.60	16.70	0.7642
Avondale	25	9.30	20489	5.80	14.60	0.5837
Belmont Cragin	70	11.50	15246	10.00	18.60	0.7857
Beverly	69	7.80	40107	0.70	5.20	0.0206
Bridgeport	30	11.20	24969	4.80	17.30	0.7931
Brighton Park	45	11.20	13138	13.20	23.00	0.7729
Burnside	70	23.40	13756	5.50	22.50	0.9155
Calumet Heights	64	17.20	28977	1.80	12.00	0.5262
Chatham	39	19.00	20320	2.20	25.30	0.6765
Chicago Lawn	96	11.90	14405	6.50	22.20	0.9965
Clearing	65	9.60	23920	3.40	5.90	0.5686
Douglas	15	16.70	23098	1.60	26.10	0.5997
Dunning	41	8.60	26347	4.80	8.30	0.4963
East Garfield Park	7	16.40	13596	7.50	39.70	0.8387
East Side	26	14.50	15347	8.30	18.70	0.9576
Edgewater	81	9.00	33364	3.90	16.60	0.7382
Edison Park		7.40	38337	0.60	5.10	0.0623
Englewood	24	21.30	11993	4.80	42.20	0.9509
Forest Glen	47	5.50	41509	1.30	6.10	0.2605
Fuller Park		40.00	9016	4.50	55.50	0.9958
Gage Park	20	14.00	12014	17.40	20.80	0.8522
Garfield Ridge	48	8.10	24684	2.60	9.00	0
Grand Boulevard		20.60	22056	2.70	28.30	0.5169
Greater Grand Crossing	54	18.90	17213	4.20	25.60	0.7636
Hegewisch	61	9.60	22561	4.40	12.10	0.2506
Hermosa	50	12.90	15411	8.40	19.10	0.9146
Humboldt Park	79	12.30	13391	11.20	32.60	0.9133

Learn

Things we learned/struggled with

- Data loading
- Merging shapefile with mySQL data
- Database formatting tradeoff

Design considerations

- Data preparation
  - Data validation: since the data came from multiple data sources with various time range, we first compared the time range to ensure that there is an overlap of the time period of our data.
  - Data cleaning: we used OpenRefine to clean and prepare our raw data separately. This is to ensure that all the source data hosted in the base tables are merge-ready, and the data types of the common variables are compatible.
  - Data wrangling: using the cleaned data, we normalized the base tables to eliminate dependencies within our data. Our objective is to ultimately provide the City officials with granular information in determining the possibility of subsidizing buildings under its energy efficiency programs (weatherization, or energy efficient appliance subsidies). Therefore, we have provided energy consumption data, while also including relevant socioeconomic/physical factors related to these buildings, such as the rental makeup of the community area, and the type/age of the building.
- Platform consideration:
  - We used a combination of R, Excel, and OpenRefine to execute data cleaning and normalization because it allowed us to work on separate data sources simultaneously. Since an important piece of information we rely on is the geographic information, we need to ensure consistency in presenting information such as address and zip code.
  - We used MySQL to conduct data wrangling since it allowed us to first define our data structure (EER) visually, before generating the schema that we'll use to insert our normalized data and run queries.
  - We used Google Could Platform to enable group work on mySQL
  - Data prepared in mySQL was further connected to Tableau for storytelling through data visualization