Individualizing and Operationalizing Energy Benchmark using Collaborative Filtering to Create A Market Momentum

Te Du, Ben Steers, Ian Xiao

Center for Urban Science and Progress
New York University
New York, US

td928@nyu.edu, bs3639@nyu.edu, ixx200@nyu.edu

Executive Summary

Benchmarking is a standard practice in business and policy making to highlight gaps, formulate quantitative goals, and track progress. Having an energy benchmarking solution that helps building owners to assess the potential return of investment in energy technology and understand their market competitiveness is a key to fuel the energy reduction initiative with market-driven incentives. However, developing a comprehensive, effective, and fair benchmarking metric is difficult because there is a significant amount of subjectivity and conflicting interests involved. The team believes that the cornerstone for a fair, insightful, and effective benchmarking framework is the ability to identify the right group of peers for comparison.

The team explored a new peer grouping mechanism, which was inspired by a technique called Collaborative Filtering (CF), for benchmarking building energy with a focus on multi-family buildings. The team argued that the CF approach can provide a more flexible and tailored measurement based on building characteristics while maintaining overall fairness. Sophisticated e-commerce companies, such as Amazon, eBay, and Alibaba, use CF to identify similar users for product recommendation based on high dimensional characteristics. The team also proposed and discussed a metric for assessing the quality of peer group in relation to common benchmarking techniques such as Linear Regression and K-Means clustering. In addition, this study includes a preliminary ensemble model idea that aims to combine the strength and minimize weaknesses of Linear Regression, K-Means, and Collaborative Filtering models.

Finally, the team developed an interactive dashboard prototype using the state of art Open Source visualization tool from Uber and data generated from the proposed CF approach. The dashboard was designed based on two hypothetical, but specific, use cases with underlying decision process in mind: a city manager who wants to prioritize building audit with limited resources and a building owner who wants understand her building's relative performance.

***Abstract* — The goal of this study is to compare the different benchmarking methods by examining the effectiveness of their abilities to identify benchmarking peer groups. The three approaches compared are the linear regression, K-means clustering methods, and collaborative filtering. The results found that the collaborative filtering has the best performance among the three. Based on the findings from collaborative filtering, a interactive dashboard is built to facilitate managing operations for city agency and also buildings owners to identify their peer group.**

Keywords—benchmarking, energy disclosure, Local Law 84, multifamily housing, office, energy usage, water use, natural gas use, optimization, interface

I. Introduction

Background

Buildings occupy a large portion of the energy use worldwide, estimates ranging from 20-40%, which are expected to increase in the coming years (Pérez-Lombard et al., 2008). Therefore, monitoring building energy use and improve energy efficiency in buildings are absolutely paramount in reducing overall global energy use and Greenhouse Gas (GHG) emissions. New York City has committed to an ambitious energy reduction plan by setting its goal to reduce 80% of its GHG emission by year 2050. And reducing energy use and improving energy efficiency in buildings is on top of the list for where to look to make city more sustainable [1].

To achieve the energy reduction goal, the city has put various signature solutions in place: mandating self energy consumption reporting for auditing through Local Law 84, providing incentives and guidelines for owners to upgrade and install better energy technologies in the existing and future buildings, launching constructive competition of energy reduction in some large offices used by public agencies and private companies, and etc [2].

Overall Challenge of Energy Reduction

All the initiatives of encouraging energy reduction that the city undertook are necessary and important steps to create momentum. However, such momentum, the team argues, is only artificial because of mandatory compliance to policies. The momentum is likely to stop once the regulatory and incentive forces retract. Split incentive is a key reason. It refers to the situation when building owners pay for the cost of upgrading energy technology while tenants reap the cost saving benefits in their energy bills [3].

Opportunities: A Different Benchmarking Use Case and Peer Selection Mechanism

So, what does it take to create a market-driven atmosphere so that building owners will cut back energy reduction voluntarily and proactively? Although the answer is not a simple one, the obvious solution is to develop a mechanics to tie energy reduction to profitability and assess competitiveness in the sales and rental markets. With this in mind, the team saw two opportunities. First, more accurate sales and rental premium can be further correlated with granular energy data from Local Law 84 (LL84) OpenData sources. An existing study by the Appraisal Institute already shown high level quantitative connection between energy efficiency, which is measured by EnergyStar or LEED grades, and the premium of sales and rental per unit square [4]. A further calibration for granularity and accuracy can be beneficial to support important investment decisions. A team at NYU CUSP is currently investigating this topic. Therefore, the team decided to focus on the second opportunity: a benchmarking mechanism that anchors on an individualized peer grouping process enabled by the well-researched Collaborative Filtering technique. The report argues that the peer searching process, which is distinctly different from the common clustering algorithm (e.g. K-Means) allows more relevant grouping considering high dimensional building characteristics.

The ultimate hope is to maintain and accelerate NYC's energy reduction progress by creating a data feedback that can spark internal motivations of building owners and investor using market forces. To overcome the barrier of user adoption, a benchmarking tool must demonstrate the ability to find relevant peers so that building owners can trust the insights. With this in mind, exploring new techniques and computation process to defining a better peer group is the thesis of this study.

Current Use Cases of LL84 Energy Benchmarking: A Different Application Compare to our Proposal

NYC mandated Local Law 84 to collect energy and water consumption data from buildings over 50,000 square feets since 2011. With this data, the city can benchmark buildings for a different purpose than introducing market forces. The city is using LL84 data to gather quantitative insights to inform building code and upgrade recommendation [2].

Also, Local Laws 85 and 87 further extend the coverage of buildings and renovation projects that are required to audit their energy profile or comply certain energy efficiency standards. The Local Law 88 requires lighting equipment updates to more efficient models and adding sub-meterings that will monitor tenant-level electricity consumption behaviors.

Parallel to these efforts are certificates programs offered by agencies such as EPA and United States Green Buildings Council (USGBC). Leadership in Energy and Environment (LEED) is a scoring system that reward efficient design components and performance statistics by USGBC. Though the model provides a clear indicator of performance, the quantitative value of the actual point-sums is not very informative and do not facilitate comparative analysis between buildings' scores [5]. Also, researches found using the LEED score, while provides correct prediction in overall energy reduction, also gives 28% to 35% buildings actually use more energy than their conventional counterparts [6].

Despite those advancement in understanding and reducing building energy use, it remains a daunting task to reach the goal set by the mayor. Finding an accurate baseline to compare a building's performance is made difficult due to two major problems: the complexity of building systems and the lack of data to describe them. In order to get an accurate assessment, a simulation model of the building could be built, incorporating all of the equipment and anthropogenic loads on the space which would give an estimate based on the direct expected energy usage as opposed to using correlations between attributes [7].

II. Literature Review

Kontokosta performed a similar analysis looking at Local Law 84, PLUTO, and CoStar Group data. The paper analyzed the energy consumption across commercial office buildings and presented a new model for benchmarking as a means of peer-to-peer building energy usage comparison. The study concluded that the model works on a national level, but may have trouble dealing with more localized, heterogeneous markets and that developing localized benchmarking tools would help to give more accurate comparisons to similar buildings in the area [5].

Top-down modeling, as the study above performs, attempts to compare buildings against their peers, describing performance as a relative score. An alternate method, bottom-up modeling, is approached by Burman et al. Bottom-up modeling attempts to look at a building's energy performance from a theoretical perspective. Two models are tested: building physics method and aggregated end-use. These methods are limited by the lack of data that is available and the level of thermal modeling that is required to compute these models [7]. The study concluded saying that both top-down and bottom-up methods have their benefits and drawbacks and can be used to complement each other.

Hsu performed a study, looking into what information is necessary in evaluating building energy performance to determine the most important information for disclosure mandates. Hsu compared the predictions of energy performance based on benchmarking, engineering audit data, and a combination of the two. It was found that benchmarking data predictive power was equal to that of a combination of benchmarking and audit data, and performed better than the engineering audit data alone [8]. This implies that building systems are relatively inconsequential when compared to building-level variation.

III. Data

Local Law 84

The building energy efficiency data comes from the open data source from the Local Law 84. The Local Law 84 is a benchmarking law that mandates owners of large buildings, buildings with a gross floor area greater than 50,000 square feet, to report their water and energy use annually. The mandate includes approximately 15,000 private and public properties that, while only covering 2% of properties city, it includes 47% of total property floor area [2]. Energy use incorporates both electricity, in the form of site energy usage intensity (EUI), source EUI, and weather normalized site and source EUI, water usage, and fuel usage covering multiple fuel types and natural gas.

Of all of the data reported, this report focuses on water, natural gas, and electricity consumption, using municipally supplied potable water intensity (gal/ft^2), natural gas usage intensity (kBtu/ft^2), and site EUI (kBtu/ft^2).

PLUTO

The Primary Land Use Tax Lot Output (PLUTO) dataset presented by the Department of City Planning (DCP), consists of information gathered from DCP along with other city agencies, including the Department of Finance and the Department of Citywide Administrative Services. PLUTO contains detailed building information such as the number of units, floor area, year built/renovated, and zoning information. The PLUTO dataset contains 88 variables and 857,237 samples in total.

Integration

Local Law 84 (LL84) from 2013 to 2016 and Pluto data are joined on Borough, Tax Block, and Lot (BBL) which is a unique identifier for a single real estate unit in New York City. This is the highest resolution available for both datasets being used.

Processing

After an initial join of the LL84 and the PLUTO datasets, the total number of records is 14,906 lots. The data still contains many missing values and are therefore not usable for any of the analytical methods presented below in its current state. Entries were dropped where it was missing a value in any of the critical columns. Due to the relatively large percentage of samples with missing values, methods of filling missing values will be explored, such as imputing an average variable or using a well-chosen dummy variable. This will depend on the variable type and its domain characteristics.

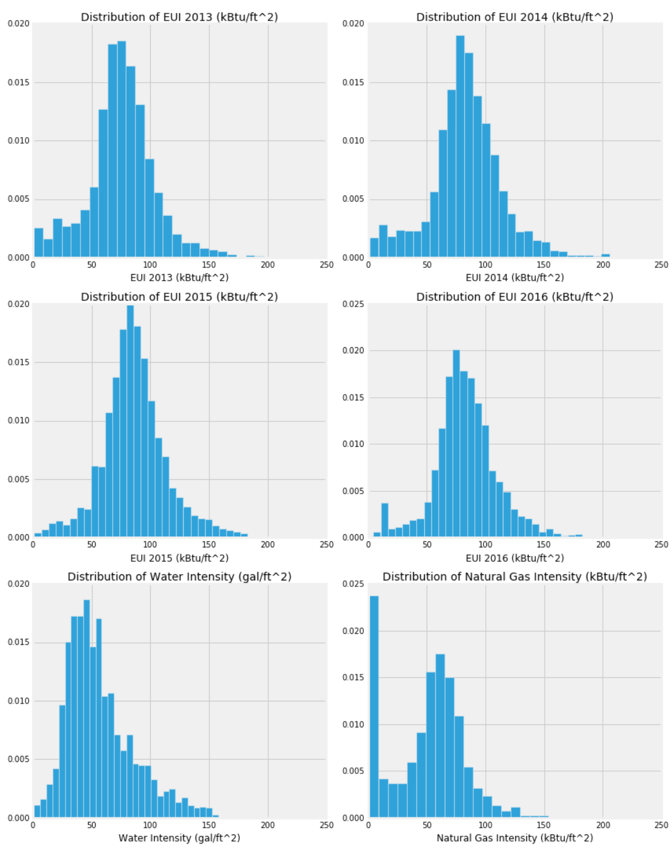

Fig. 1. *Normalized distribution of EUI for the 4 years being considered, 2013, 2014, 2015, and 2016, the water usage intensity, and the natural gas usage intensity. Outliers above 2 log standard deviations are removed.*

There are two methods being considered for removing outliers. One method is where samples with values above 95 percentile are filtered out, and one is where samples with log values outside of two standard deviations of the log mean are filtered out. This method has the benefit of not having the mean be swayed as heavily by extreme outliers and instead removes outlier based on their order of magnitude. Variables that are inherently non-negative will have any negative values removed. The distributions for 2013-2016 can be seen in 2016 with extreme outliers removed using the log-filtering approach.

In addition to outlier removal, there are duplicated BBLs due to primary and secondary buildings in the same lot. In this case, there will be multiple records with the same BBL, but different building details, such as EUI and Floor Area. In this analysis, only the records with the highest values in EUI were kept.

Once outliers and duplicated records were removed, normalization (scaling value between 0 to 1) was applied to continuous numerical features for proper distance calculation in the downstream process.

The geographic position of buildings given by PLUTO are in the New York/Long Island Coordinate System (EPSG:2263). This was projected to WGS84 (EPSG:4326) for increased compatibility with mapping programs.

The final dataset contains 3,199 unique multi-family buildings with key buildings features: floor area from 2013-2016, EUI from 2013-2016, occupancy, building type, borough, longitude / latitude, total units, address, building age, and % of residential / commercial / retail areas.

Descriptive Analysis

This section provides a quantitative overview of the dataset and key modeling features.

The distribution of energy, water, and natural gas usage intensity are given in Figure 1. The distributions of EUI for 2013-2016 remain relatively consistent between years, showing low global variability of building energy usage. Water and natural gas usage are only available for 2016, and therefore cannot be compared across multiple years. Natural gas usage distribution has strange behavior, where there is a high concentration of reported usages in the lowest bin. It was initially assumed that these were assigned zero either due to missing values or because a building didn't use natural gas, however after investigation it was found that there is a peak around 0.1 kBtu/ft^2. Further investigation should be made into the cause of this.

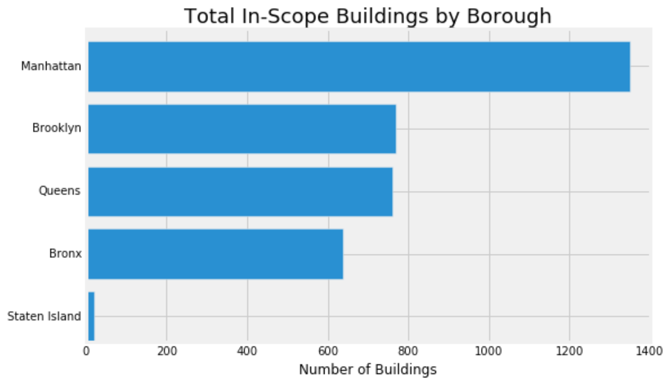

Fig. 2. *Building count breakdown by borough. The total number of buildings is 3,199. Samples with missing values, and extreme measurements in energy, age, office, and commercial spaces are excluded. Staten Island is under-represented compared to other boroughs.*

The distribution of buildings being considered in this report can be seen in Figure 2. Most buildings used are found in Manhattan, meaning that the results of this analysis may be biased towards building characteristics that are present in Manhattan. Brooklyn, the Bronx, and Queens are all comparable in magnitude. Staten Island is largely underrepresented in this sample, most likely due to the low number of large multi-family buildings compared to downtown areas like Manhattan. This is most likely also the cause of the difference between all of the borough building counts.

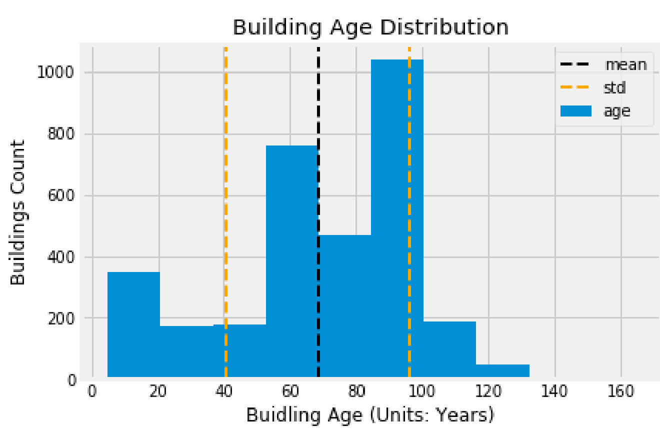

Fig. 3. *histograms or the buildings age. Majority of buildings falls between fifty to hundred years old.*

The building age is included in the model on the assumption that the buildings in different ages might choose different materials and construction methods due to engineering advancement and aesthetic difference. These could potentially contribute to difference in energy efficiency.

IV. Methods

The methodology of this study was designed to achieve two objectives: 1) to compare how well Collaborative Filtering (CF) selects peer group against Linear Regression global selection and K-Means clustering. Cluster level average EUIs were calculated and the standard deviation of them were used to compare clustering performance across models. 2) to calculate the Energy Efficiency Ratio (EER) as a metric used to rank how well each building performs against the standard and other buildings. EERs from Linear Regression, K-Means, and Collaborative Filtering were used to experiment an ensemble technique and support the interactive visualization.

Each model has distinct clustering process to support comparison in this study. The Linear Regression takes a global selection, which means the model fits across all samples without breaking the sample space into subpopulations. The predicted EUI value is used as the benchmark standard. K-Means breaks down the global sample into various clusters with certain cluster sizes, which can be thought of as a top-down Approach. Then, all the buildings in the same cluster are compared against the cluster average. In this study, K-Means provided meaningful insights with the number of clusters range from 2-5 with average sample size of 1,600 to 640 respectively.

In contrast, with a bottom-up paradigm, Collaborative Filtering uses each individual building as the center; it finds the nearest neighbours to form a peer group based on a pairwise distance metric. Then, the building used as the center is compared against the peer group. The process repeats for every building in the final dataset. Theoretically, CF can provide 3,199 distinct clusters (the total samples in the final dataset) with consistent and meaningful cluster sizes. For example, one can create 3,199 distinct clusters with 100 samples in each cluster. This is the fundamental different CF can offer compare to other clustering techniques such as K-Means, DBScan, etc. Conceptually, having the ability to use individual sample to form its own cluster with consistent sample sizes is what allows fair individualization.

Lastly, the team argues that a good peer group should have consistent energy consumption characteristics. Therefore, average standard deviation across all clusters was used as a benchmarking metric for the three modelling approaches. High standard deviation in EUI in a cluster may suggest sub-optimal clustering due to two main reasons: poor sample selection and insufficient population in the cluster.

Using Linear Regression, K-Means, and CF for benchmarking set the foundation for ensembling. Each model has strengths and weaknesses. Ensemble allows one to combine the EER output of each model through a simple or weighted averaging. The ensemble mechanism reduces the risk of relying on any single approach that may lead to inconsistency and unfairness. The holistic EER score allows agencies to prioritize building investigation under different resource constraints. For example, an agency may want to only focus on top 10 high risk buildings because they only have 5 investigators and 2 weeks.

With such use case in mind, the secondary objective of this study is to explore the potential of creating operational tools. The team designed an interactive dashboard for two hypothetical use cases: city operation managers who want to plan resources for inspection and building owners who want to assess their competitiveness. A prototype tool was developed using the latest Open Source Visualization package, deck.gl, from Uber.

The following sections will discuss the specific setup of each modelling and clustering techniques. Please refer to the Results section for details on the interactive tool.

Modeling

The objective of this modeling is to identify outliers, or buildings that seem to be underperforming in energy efficiency compared to their building peers. This is difficult to do because building efficiency is a complex system consisting of magnitudes more variables that contribute to or indicate efficiency than are currently available. To try and capture the complexity of the system based on the available information, the building system is measured using multiple regression methods which are combined using a bagging technique. Energy efficiency ratios will be ranked by percentile and above certain percentile would be identified as outliers. The two methods to be used are linear regression which compares building-use peers and K-Means which compares building-characteristic peers.

Building-use peer benchmarking is the conventionally used way of benchmarking, which uses a set of linear regressions, one for each building use classification, i.e. multi-family housing, office, etc. This is based on the assumption that similar building usage types will have similar energy profiles due to the similarity in internal operations. Segmenting based on the building use alone may not always provide the best results, as buildings may have energy usages that correlate more with buildings that have similar characteristics such as size or age. Another segmentation technique is to cluster buildings based on their building characteristics and use the peers within each clusters as a benchmark to compare a building's performance to. This will be performed using K-Means clustering and Collaborative Filtering where the number of clusters and number of samples in a cluster, respectively, will be determined using cross-validation. The clustering approaches have the benefit of not assuming linear, independent relationships, as linear regression does.

The bagging mechanism reduces the risk of relying on any single approach that may lead to inconsistency and unfairness. The holistic outlier score allows agencies to prioritize building investigation under different resource constraints. For example, an agency may want to only focus on top 10 high risk buildings because they only have 5 investigators and 2 weeks.

Linear Regression

To calculate the average standard deviation of EUI 2016 clustering comparison, it is simply the standard deviation of EUI 2016 across all sample because Linear Regression does not reply on clustering. The average standard deviation of EUI 2016 in this case is 27.47.

In addition, the regression model is used to calculate a EUI standard for energy benchmarking. It is important to realize that the choices of characteristics used for regression should be based on the building types. There are some variables included here that are useful for any building types such as lot area, age, and locations. But variables such as workers density, while important for regression model of offices buildings, might not be appropriate for multifamily model. The formula for the regression model is as following:

$$Pred\ EUI = \beta_1 BlgArea + \beta_2 age + \beta_3 UnitsTotal + \beta_4 Occupancy \quad (1)$$

The dependent variable y is the energy use intensity from the Local Law 84 data. After the model is fitted, then the expected values for the building EUI are predicted. Which can then be used to calculate the energy efficiency ratio (EER below) using the formula:

$$EER = \frac{Actual\ EUI}{Predicted\ EUI} \quad (2)$$

which is a measure of a building's energy performance compared to its benchmark. Higher values represent worse energy performance, a value of 1 represents neutral performance and values between zero and one represent better than expected performance.

K-Means Clustering

The k-means algorithm partitions the observations into k groups. The team created k-means clusters using the following key building characteristics: Floor Area, Occupancy, Number of Units, Building Age, and Average EUI from 2013-2015. The standard deviations of EUI 2016 at various number of clusters were then calculated for model comparison.

The team ran a Silhouette Analysis to identify the most appropriate number of clusters for clustering comparison across models. Based on the results below in , the team selected k=4 and the corresponding Average Standard Deviation of EUI 2016 for benchmarking with Linear Regression and Collaborative Filtering.

Table 1. *Results of K-Means cluster cross-validation.*

Clusters	Avg. Cluster Size	Max Cluster Size	Min Cluster Size	Avg. STDEV of EUI 2016
2	1599.5	2921	278	27.8
3	1066.3	2663	91	28.1
4	799.8	2629	63	27.2
5	639.8	2118	61	28.1
6	533.2	2039	24	28.5
7	457.0	2038	2	27.6
8	399.9	2040	2	28.1
9	355.4	2037	1	28.1
10	319.9	1598	1	28.3
11	290.8	1619	1	29.4
12	266.6	1250	1	29.1

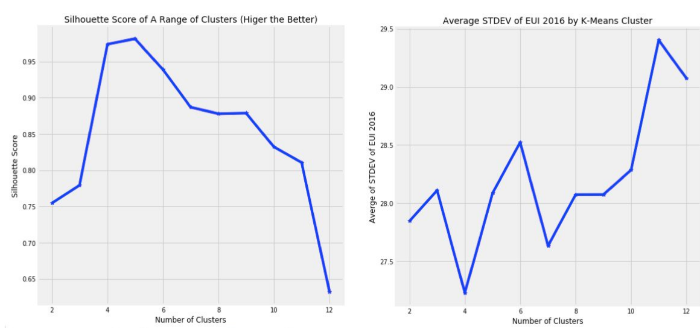

Fig. 4. *The silhouette score (left) and Average standard deviation of EUI 2016 by the number of K-Means clusters (right).*

The EERs for each building were calculated as the following:

$$Predicted\ EUI_k = \frac{\sum_i^{M_k} EUI_i}{M_k} \qquad (3)$$

where k is the cluster index that range from 2 to 12 in this study and M is the size of cluster k. Predicted EUI is then used in the EER equation defined above.

Collaborative Filtering

Collaborative Filtering (CF) is a technique that is being used widely by e-commerce companies, such as Amazon, eBay, and Alibaba, to personalize products recommendation. This is a much more elaborated process than clustering. This study only utilized the clustering in CF - and described as CF in general throughout this paper - for identifying peer buildings.

To identify peers for buildings, a pairwise distance matrix was first computed. The team created a 3,199 by 3,199 matrix with Euclidean Distance as values. The pairwise distances were calculated using the same building parameters used in K-Means.

Table 2. *An illustrative example of a pairwise distance matrix for a sample of 3 buildings. The value is the Euclidean Distance based on building parameters.*

	Building 1	**Building 2**	**Building 3**
Building 1	0	0.2	0.4
Building 2	0.2	0	0.1
Building 3	0.4	0.1	0

Subsequently, for each building (each row of the matrix), the near neighbours were identified by sorting the Euclidean Distance of all other building (the corresponding column of the matrix). Various peer group size can be developed by specifying the number of closest neighbour. Each building may have different or overlapping neighbours depending on the distance ranking. For this study, the team analyzed different CF clusters with different peer sizes range from 30 to 2200. Each instance have 3,199 clusters (e.g. each building is its own cluster with 30 peers).

Similar to K-Means, the EERs are calculated on a group level using, but with different elements and process:

$$Predicted\ EUI_b = \frac{\sum_i^{p} EUI_i}{p} \qquad (4)$$

where b is the building index and p is the number of peers. Predicted EUI is then used in the EER equation defined above.

Ensemble Approach to EER

Each building has three EERs calculated based on Linear Regression, K-Means, and CF. Discussing the accuracy of the EER is out of scope in this study because the lack of ground truth data (EUIs are self reporting). However, the team need to choose a single EER measure for interactive visualization. Given this situation, there two options: 1) to use the best technique that offers the best clustering 2) to combine outputs of all models to reduce risk of relying on any specific model, which is known as Ensembling.

The team proposed a simple approach, which is by taking the average of EER from Linear Regression, K-Means, and CF to create the final EER. A better approach is to use weighted average, but the discussion of weight assignment is not in the scope of this study.

The final EER is going to be indexed and ranked by their values from the highest to the lowest. (1) shows the EER should be close to one to indicate the building's performance is aligned with the expected and it should be considered as energy efficient. The higher the EER for a building is, the worse it is performing compared to its expected values.

As part of the interactive dashboard, a parameter will be available where the user can select the percentile of EER to threshold by, allowing for a tunable parameter that represents what defines an outlier based on the user's specifications. The threshold will be determined using the cumulative distribution of all EERs for that year, where the EER threshold is equal to the point on the cumulative distribution function equal to the percentage specified by the user.

V. Results & Discussion

Which modelling technique offers a better peer selection?

As the team proposed in earlier section, average standard deviation of EUI 2016 (note as *avg. std EUI 2016*) across clusters can be used to assess if there is consistency in peer selection and grouping. A lower *avg. std EUI 2016* suggests the clusters have more relevant peers similar to the building of interest given its key characteristics (e.g. Floor Area, age, Occupancy, Age, and Average EUI from 2013-2015).

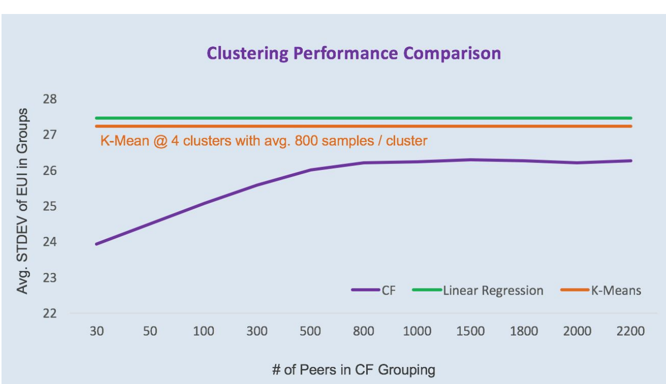

Fig. 5. *Avg. std EUI 2016 comparison of Linear Regression, K-Means, and CF in a spectrum of CF peer sizes*

The graph above suggests that CF provides a better peer selection. This technique demonstrated consistent lower *avg. std EUI 2016* across a wide range of peer sizes. In addition, CF offers a lower avg. std EUI compared to the best result from K-Means which has 4 clusters, average cluster sizes of 800, and an avg. std EUI of 27.2.

Is the difference between K-Means and CF statistically significant?

A Welch T-Test suggests that there is a statistically significant differences in the average standard deviation of EUI 2016 based on the K-Means and CF approach. The lower average standard deviation in EUI 2016 among peers in all CF clusters suggests a more relevant grouping as we argue in the Modelling section.

$$t = \frac{m_A - m_B}{\sqrt{\frac{s_A{}^2}{N_A} - \frac{s_B{}^2}{N_B}}} \qquad (5)$$

where *m* is the mean of the average standard deviation EUI 2016 across various setup, *s* is the variance of average standard deviation of EUI 2016, and *N* is the sample size.

To calculate perform the Welch T-Test, the team perform K-Means and CF under 11 different conditions by varying k in K-Means and peer group size in CF. The Null Hypothesis is that there is no significant difference in avg. std EUI.

The Welch T-Test shows the Null Hypothesis can be rejected. In other words, there is a statistically significant difference in the avg. std EUI.

Table 3. *Detailed results of K-Means and CF using different cluster and peer group sizes.*

	K-Means		Collaborative Filtering	
Experiment	# of Clusters	Avg. STDEV of EUI 2016	# of Peers	Avg. STDEV of EUI 2016
1	2	27.85	30	23.93
2	3	28.11	50	24.49
3	4	27.23	100	25.07
4	5	28.09	300	25.57
5	6	28.52	500	26.01
6	7	27.63	800	26.20
7	8	28.07	1000	26.24
8	9	28.07	1500	26.29
9	10	28.29	1800	26.27
10	11	29.40	2000	26.21
11	12	29.07	2200	26.26
	Average	28.21	Average	25.68
	STD	0.61	STD	0.83
	Absolute T Score = 8.13			

In addition to quantitative measurement, CF provides various qualitative advantages compared to other clustering techniques. First, the pairwise Euclidean distance between buildings in CF offers traceability. It is imperative to be able to provide clear and easy-to-understand explanation of why certain buildings are grouped together if this method is being used to inform decision making and obtain buy-in from building owners. Second, using each building as the center of its own cluster offer individualization based on a building's specific characteristics. This can help to make a case for building owners for adoption. Lastly, CF offers scalability if the peer grouping requires high dimension parameters.

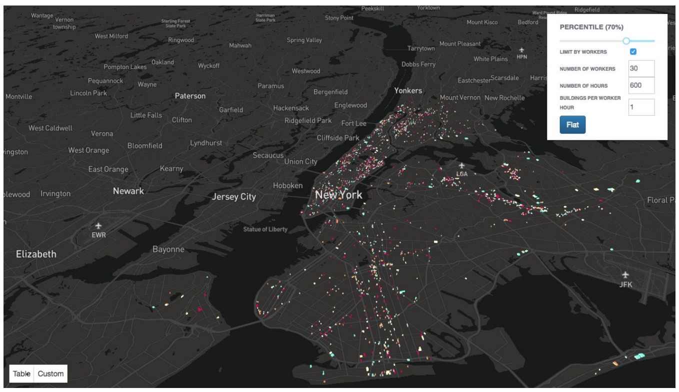

Fig. 6. *Interactive operational dashboard. Buildings close to their benchmarked efficiency are shown in blue, while buildings with poor energy efficiency are highlighted in red. Heights of the buildings are also used to display the energy efficiency where larger heights signify poorer efficiency.*

VI. Conclusions

In this study, the team proposed a new method, inspired by Collaborative Filtering, to identify individualized peer grouping while maintaining overall fairness. Individualized peer grouping allows stakeholders, such as building owners and investors, to connect the dots between energy reduction and key business indicators, such as return of investment and competitiveness in sales and rental premiums. With the help of an easy-to-use dashboard, internal and market-driven motivations can be generated to sustain and accelerate the momentum of NYC's 50x30 energy reduction initiative.

The team presented the process, benefits, and scientific validation of using a peer grouping mechanism from Collaborative Filtering (CF). In addition, multiple EERs for each building were calculated with Linear Regression, K-Means, and CF. These EERs are averaged, based on the Ensemble method, to create a single metric to support visualization.

The final visualization and operation optimization tool produced in the project with two specific use cases in mind: one for city operators and the other for building owners. With the ensembled EER, certain buildings would be chosen to be inspected for further investigations. Fig. 6. gives a screenshot of the dashboard for the operational officers to get a quick feedback on where to look for inefficiency buildings. The slider allows users to choose a level of significance for outliers in the building efficiency by filtering based on the population percentile above a certain threshold of EER. The three input fields allow the users to choose the number of workers, number of hours and number buildings per worker hour. Then, depending on the personnels resources and day-to-day variance of available time. There are couple advantages to this type of operational tool: 1) the minimizing the decision to make for the administrators. The only inputs needed are the number of personnels and their hours of working. 2) cross-validation between multiple models to ensure the based-line logic behind the decisions is to safeguard essential biases. The complexity of the models are hidden away from the users. 3) iterative approach that will allow the administration and building owners to improve on their own decisions and considerable shorten the cycle of receiving the results from an operational standpoint allowing more rapid corrections.

In addition to the user interface for city operators, building owners can retrieve a list similar buildings, enabled by Collaborative Filtering, in the city with key characteristics of their properties, such as Floor Area, Building Age, Number of Units, and Estimated EUI for past 2 years. For more information and technical details, please refer to the project on GitHub [10].

Limitation and Next Steps

The analysis is still severely limited by the lack of data available. Also the data cleaning process currently involves mostly eliminating the data that are ranked above 95 percentile in EUI for the analysis. But, each parameters should require more careful selection to allow the models to perform optimally. For example, there are buildings over two thousand years old in the current dataset. For multifamily benchmarking modeling, tenant characteristics could make a significant difference in the the accuracy of the models. The team also wanted to explore the effects of submetering on benchmarking. Moreover, the availability of interval data for buildings can also be helpful in benchmarking buildings energy. Building energy consumption should follow a consistent pattern. The disruption in regularity of energy consumption pattern could potentially indicate buildings to be inefficient in its operation and it could be identified and incorporated into benchmarking process.

Both K-Means and CF rely on distance calculation. Euclidean Distance is being used as the distance metric. The shortfall of this metric is that it is not applicable to categorical data (i.e. building category, borough, etc.). A distance metric that can process a mixture of numerical and categorical data need to be investigated in order to augment the solution to select peers based on more complicated building features.

As next steps, the dashboard (see Fig. 6.) could be expanded to provide more relevant user experience for aiding in operational processes. These features could be derived from interviews and suggestions from the target users. The operations challenges could be much more than allocating limited personnels to do field inspections, but the decision can be influenced by the values and other priorities for the agency. Those constraints and objectives can all be packaged into the dashboard to give immediate recommendations to agency operators.

Improvements on the end of buildings owners are also desired. Identifying their own energy benchmarking peers group is only the first steps to inform the building owners. Later iterations of the dashboard should include peer group descriptive characteristics for the peer group. Based on those peer group characteristics, a checklist of specific programs and areas of improvements could be recommended to the building owners based on their peer groups.

References

[1] New York City Mayor's Office of Sustainability. (2016). *New York City's Roadmap to 80 x 50*(Rep.). Retrieved from NYC Mayor's Office of Sustainability website: http://www1.nyc.gov/assets/sustainability/downloads/pdf/publications/New%20York%20City%27s%20Roadmap%20to%2080%20x%2050_20160926_FOR%20WEB.pdf

[2] City of New York. (2017). *New York City's Energy and Water Use 2014 and 2015 Report.* October 2017. Retrieved from New York City Mayor's Office of Sustainability website:http://www.nyc.gov/html/gbee/downloads/pdf/UGC-Benchmarking-Report-101617-FINAL.pdf

[3] Tawanna Dillahunt, Jennifer Mankoff, Eric Paulos. "Understanding Conflict Between Landlords and Tenants: Implications for Energy Sensing and Feedback." 12th ACM international conference. Ed. 12th ACM international conference. Copenhagen: 12th ACM international conference, 2010. 149-158 .

[4] Institute for Market Transformation & Appraisal Institute. (2013). *Green Building and Property Value A Primer For Building Owners and Developers.* Retrieved from Appraisal Institute website: https://www.appraisalinstitute.org/assets/1/7/Green-Building-and-Property-Value.pdf

[5] Kontokosta, C. E. (2014). From Transparency to Transformation: A Market-Specific Methodology for a Commercial Building Energy Performance Rating System. *SSRN Electronic Journal*. doi:10.2139/ssrn.2346414

[6] Newsham, G. R., Mancini, S., & Birt, B. J. (2009). *Do LEED-certified buildings save energy? Yes, but…. Energy and Buildings*,41(8), 897-905. doi:10.1016/j.enbuild.2009.03.014

[7] Burman, E., Hong, S., Paterson, G., Kimpian, J., & Mumovic, D. (2014). A comparative study of benchmarking approaches for non-domestic buildings: Part 2 – Bottom-up approach. *International Journal of Sustainable Built Environment*, 3(2), 247-261. doi:10.1016/j.ijsbe.2014.12.001

[8] Hsu, D. (2014). How much information disclosure of building energy performance is necessary? *Energy Policy*, 64, 263-272. doi:10.1016/j.enpol.2013.08.094

[9] Von Neida, B., & Hicks, T. (n.d.). *Building Performance Defined: the ENERGY STAR National Energy Performance Rating System.* Retrieved from energy star website: https://www.energystar.gov/ia/business/tools_resources/aesp.p

[10] Xiao, Ian, Te Du, Ben Steers, CAUI Team 1 Final Project, (2017), GitHub repository, https://github.com/td928/CAUI_team1_Final_Project.git