NYU Center for Urban Science + Progress

Assessing the Circular Economy Opportunity in NYC

Capstone - Team:
Yang Li
Yong Fan
RongJian Yang
Mentor: Martina Balestra, NYU CUSP
Sponsors: Nicole Spina, NYCEDC; Kate Kitchener, DSNY

Date: June, 22 2020

Table of Contents

1.Abstract

This capstone project studied the circular economy concept mainly focused on understanding and predicting waste generation patterns from a residential waste perspective in the normal period approach and the Covid-19 period approach. Technical models including LSTM, Cluster, and Regression models were applied in the two major analyses to see the trend and predict the waste generation volume for each waste type as well as understand how socioeconomic factors would affect waste generation.

Our analyses showed that waste generation fluctuates over the years. There was seasonality in all five boroughs, while there wan not any obvious trend in these time-series data. People tend to generate more waste during the summer than wintertime. And there was a high positive correlation between the three main types of waste generation. As for socioeconomic factors, Income level has a negative correlation with refuse generation during the normal period, and both refuse, and recyclable waste has a negative correlation with income during the Covid-19 Stay at Home period.

This project also provided an initial business model for EDC, private sectors, and communities to collaborate to promote the development of New York City circular economy.

2.Introduction

In the traditional linear economy, raw materials go straight from production to consumption and disposal. In this scenario, economic values are created by producing and selling as many products as possible, which inevitably result in unnecessary waste, and that creates many problems for urban environments across the globe, such as scarcity of landfills and various forms of pollution. In contrast to the 'take-make-waste' linear model, a circular economy promotes a regenerative model to keep materials in-use for as long as possible and gradually decouple growth from the consumption of finite resources. A transition to a circular economy would drive $380B in cost savings for businesses worldwide and could reduce carbon emissions by 70% (Dana Rohrabacher, 2013). It is designed to benefit the environment, business, and society, and NYC is taking the spot to be the leader of the circular economy in the US. Within the One NYC Plan, Mayor Bill de Blasio pledged that NYC would send zero waste to landfills by 2030 (NYC

Office of Sustainability, 2016). A city of 8 million residents, how is it possible to have zero waste sent to landfills? The answer is the circular economy. NYC EDC (Economic Development Corporation) was currently touching base with generating this regenerative economy model, and they wanted to understand how individuals, business entities, and related departments could participate in this model while benefiting from the value that "unwanted waste" creates. Given the special period of the Covid-19 pandemic, the pattern of urban mobility shifted dramatically, stay-at-home order shut down business in the city, and strained people at home. Thus, the usual pattern of residential waste generation also changed dramatically. Under such context, the team took separate analysis on month-on-month waste generation pattern for March and May only, and general waste generation pattern during normal times. The expected result of research and analysis on NYC residential waste generation would benefit circular economy businesses in dispatch resources and more efficient operation and understanding how the pandemic affects the pattern of residential waste generation.

3. Problem Statement

According to a new analysis by the global risk consulting firm Verisk Maplecroft, the United States, with 4 percent of the world's population, produces 12 percent of the world's municipal solid waste with only a recycling rate 35 percent(Smith, 2019). As one of the country's biggest cities, New York City should take primary responsibility for this problem. The incomplete waste collection and recycling systems have left the town missing out on many reusable resources, which could generate substantial economic value. To better put those waste back into the economic cycle, the goal of this project was to understand and predict the waste generation pattern to introduce a business model and make recommendations for DSNY and related private sectors to corporate together to recycle more reusable waste.

Here are the lists of our research questions and corresponding hypothesis:

- [RQ1]The team's first mission is to develop a deep understanding of waste generation patterns during the normal period as well as to find out the relationship between socioeconomic background and waste generation. Based on that, the team could predict the volume of waste generation to make a proper business model and give suggestions.

[H1] The team supposed the waste generation pattern in the normal period steadily fluctuates over the years. Community districts with higher median household income will generate more waste than community districts with a lower median household income.

- [RQ2]The second mission of the team is to explore how the normal waste generation pattern changed during the Covid-19 period on the geographical aspect and how socioeconomic factors play its role at this time. By digging insights from waste generation patterns in both the Covid-19 period and the normal time, the team implemented those insights into our business model and recommendations.
  [H2] During the stay at home period, the team supposed the amount of residential refuse generated would skyrocket due to limited mobility in all five boroughs, paper, and mix recyclable would slightly decline due to more reuse than disposal. Neighborhoods with more residential zonings experienced a more significant increase in waste generation. High-income areas would see less of an increase because residents left the city to avoid pandemic hot zones, while more impoverished neighborhoods would see more of an increase because people can't afford to leave.

4. Literature Review

There is already a large body of publications analyzed and predicted the waste generation pattern. We will refer to some of these works with relevant methodologies.

Navarro and Meza, who live in two different decades, have adopted the same approach to predict the waste generation volume - Time Series. Navarro, in a 2002 paper, has conducted dynamic multiple solid waste(MSW) generation analysis using time series data of solid waste generation quantities. Methodologies, including a seasonal AutoRegressive and Moving Average(sARIMA) model and dynamic system modeling, were applied to forecast the MSW generation. Both forecasting techniques have demonstrated good results in terms of accumulative error and prediction length within an established accuracy level. And the sARIMA methodology gave outstanding results for daily and monthly data(Navarro, 2002). In 2019, Meza proposed to use recurrent neural networks as one of their forecasting alternatives to predict the waste generation in the city of Bogota. One of the main advantages of these networks (also known as long-term memory networks, LSTM) is that they can adjust the behavior of non-linear data and maintain memory and forget states which take into account past time information. Based on the robust database they submitted, their results confirmed that LSTM is a suitable model for this type of analysis (Meza,2019)

To understand and analyze how social-economic background will affect waste generation patterns, Bandara, in a 2007 paper, has performed a regression analysis to delineate a relationship for the different income groups separately by taking the amounts of per capita waste. In this case, they took income level as their independent variable, and the results showed that income level is an essential social-economic factor that has a high correlation with waste generation(Bandara,2007).

5. Data

All data sourced from NYC Open Data Portal.

5.1. **DSNY Monthly Tonnage Data**

DSNY Monthly Tonnage Data provides monthly collection tonnages that the Department of Sanitation collects from NYC residences and institutions. For more information about DSNY services, see: http://www1.nyc.gov/assets/dsny/site/services.This data set is collected in community district scope and in the frequency of every month. Three main types of waste(refuse, paper, and MGP) are reported in this data set, while there are also organic waste data but with many NAs. Columns of this data set include boroughs, community district id, tons of refuse, tons of paper, tons of MGP, tons of source-separated organic waste, etc. The data set has 11 columns in total. This data set is used to conduct the LSTM time series to identify the time patterns of waste generation. It's also connected with the community geo data to analyze on a spatial scale.

5.2. NYC Planning | Community Profiles

This interactive data portal contains demographic, economic data, and other resources describing New York City's 59 community districts, each represented by a community board. Among all the profile data provided by this data source, only Community District Median Income(based on American Community Survey 2013-2017) and Population (based on 2010 Census) were selected for this project exclusively as demographic features. The community geo data are used to conduct the spatial analysis of waste generation.

6. Methodology

To make plausible suggestions on how to reduce the economic losses caused by waste and to boost the circular economy, we must dive deep into the waste generation to understand how it is like and what its patterns are. Our research first focused on how waste generation and the circular economy was like in both long-term and short-term scales. The general analysis was to analyze overall residential waste generation throughout the years. Based on all the waste generation patterns, we could make plans ahead for the smarter circular economic model. Analysis during the Covid-19 period will mainly be based on the current shock of the Covid-19 and how the unprecedented circumstance of the Covid-19 pandemic would influence the waste generation and the circular economy.

6.1 General Waste Analysis:

6.1.1 Time Series Analysis:

The team used the NYC tonnage dataset and focused on the near-term data collected after January 2010. We summed up the data into boroughs while the original was in community district scale. Then we applied an LSTM model to visualize the trend of all types of waste generation and how they changed during the last decade. And the team also created a heatmap of correlation between three main waste types.

6.1.2 Social-Economic Spatial Analysis:

Two datasets were used in this section, including DSNY monthly tonnage collection and Community District demographic profile. We merged demographic information with waste data. Then, we generated several heatmaps to visualize waste collection tonnage. Later, the team conducted clustering analysis seeking the relationship between median household income and per capita waste generation among all community districts. After that, the team applied a regression model from the statsmodels module to community district clusters found in the previous clustering result to test the initial hypothesis about waste-demographic correlation and made conclusions.

Here are some of the technical details. Before putting data into clustering models, we first computed per capita waste generation by categories by taking the division of a community district's gross waste generation and its corresponding population. We also removed some of the features; only four were left, including Median Household Income, Refuse per capita, Recycle Paper per capita, and Mix-recyclables per capita. Then, we normalized the dataset by StandardScaler and used PCA to reduce the dimensionality of the data to two components for better visualization while maintaining a 91% ratio of variance explained.

After finishing preparing data, we applied three different clustering algorithms, including DBScan, K-Means, and Gaussian-Mixture, to evaluate the best performing algorithm by measuring the optimal silhouette score in each algorithm. Silhouette score is a method of interpretation and validation of consistency within clusters of data.

$$silhouette_{s}ample_{s}core = \frac{b - a}{max(a, b)}$$

"The Silhouette Coefficient is calculated using the mean intra-cluster distance (a) and the mean nearest-cluster distance (b) for each sample. The Silhouette Coefficient for a sample is (b - a) / max(a, b). To clarify, b is the distance between a sample and the nearest cluster that the sample is not a part of. "(sklearn.metrics.silhouette_score. (n.d.). , 2019)

After thorough parameter optimization for each algorithm, Gaussian-Mixture proved to be the best clustering method in this case.

6.2 Covid-19 Waste Analysis:

6.2.1 Time Series Analysis :

Using the time series model built in part 6.1.1, we compared the actual waste generation values with the predicted ones. We could also see that waste generation in different boroughs changed differently during the Covid-19 Stay at Home period. Regarding the current shock of the

Covid-19, we built a time series model to get the predicted value from March to June and then compared it with the actual tonnage of residential waste generation.

Considering the large proportion of refuse and high correlation among these three types of wastes, we chose refuse as an example and built a time series model for each borough. Narrow et al (2002) used ARIMA models to help waste management. Meza et al(2019) used recurrent neural networks to predict the waste generation in Bogota, which proved robust for this application. We also used Long Short Term Memory, which is a subcategory of RNN, to build the time series model. We used two layers of LSTM stacking together. The structure was shown in Figure 5(in appendix). Here the X(t) refers to the time series data. The Tk refers to the LSTM layers' output, which is the hidden state of the network. The outputs of the first layer would come directly to the second layer. The second LSTM layer's outputs would then go to a linear layer and get the final outputs. We used the first 70% data as training data and the remaining as the test data. For computing convenience, we also scale the data. The scale function is as follows:

$$Scaler(X) = \frac{X - min(X)}{max(X) - min(X)}$$

X: variable vector

max(x)/min(x): the max/min value of all elements in vector X

We used lagging of 12 in this model, which aimed to capture the yearly seasonality of the data. The hyperparams of these models are: hiddeen_state = 18, learning_rate = 5e-5, number of training epochs = 1500. Considering it’s a regression model, we used MSE as the loss function.

6.2.2 Social-Economic Spatial Analysis:

The analytical methodology in this section is identical to section 6.1.2. But in this section, the team will use the analysis result to test different sets of hypotheses under the timeframe of Covid-19 NYC Stay At Home order period.

6.3 Analysis on Suggestions and Potential Business models

Based on the previous analythese, the team estimated on what kind of waste is causing the most economic losses as well as looked at entities which have been playing important roles in the circular economy, including startups, academic institutions, venture capitals to help reduce the economic losses of waste. After that, the team also came up with related suggestions and a potential business model to give a more straightforward explanation.

7. Results

7.1General Waste Analysis

7.1.1 Time series analysis

From the visualizations in Figure 1-3(in appendix), we can see that refuse occupies the most significant proportion in all wastes. There is also seasonality in all five boroughs, while there isn't any obvious trend in these time-series data. People tend to generate more waste during summer than winter. Moreover, due to the same-direction change of all these three types of wastes shown in the above visualization, there are high correlations (Figure 4 in appendix) among the three types, which can also simplify our model of time series in part 6.2.
The results of six time-series models mentioned in par are shown in Table 1(in appendix). The real and predicted values are shown in Figure 6-8(in appendix).
As we can see above, the LSTM structure has captured the fluctuation of the time series successfully. All five boroughs have similar trends as well as seasonality. The waste generation reached its maximum during the summer, while it reduced to its lowest point in winter. This pattern may suggest different manipulation of waste and the circular economy in different seasons.

7.1.2 Analysis on DSNY tonnage collection in space(2015-2019)

In Figure 9(in appendix), Districts in Queens, Brooklyn have noticeable drop in Paper collection tonnage, and districts which had noticeable dropped clustered together in Queens.

In Figure 10(in appendix), MGP collection largely remains the same level of collection tonnage across all five boroughs.

The preliminary result is that economic loss on recyclable paper increases from 2015 to 2019 because factors that affect DSNY collection were the same when it comes to refuse and recyclable waste collection. But for some districts in north-east Queens, the amount of paper that got collected dropped significantly in 2019. This could indicate DSNY recyclable paper collection plan has room for improvement in the future and awareness of public recycling.

In the clustering analysis, the silhouette-score from Gaussian Mixture equals 0.3988. The algorithm provided two clusters that it calculated to have relations, Figure 7.1.2-1 is a geographic representation of the two clusters.

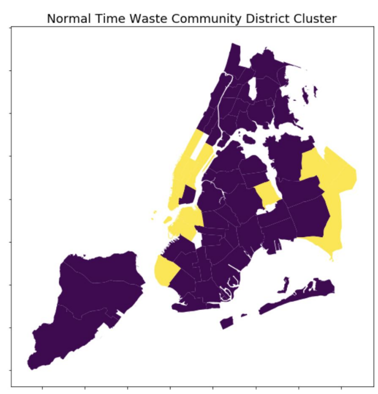

Figure 7.1.2-1 Normal Time geographic Cluster Results

the first cluster(dark blue) contained most of the community districts that has lower household income, and the median household income in this cluster was $52,490 , and most of the community districts distributed in the 50k-65k range. (figure 11 in appendix).

After applying regression model to this cluster, the result shows that there was a positive correlation between income and refuse generation per capita (figure 12 in appendix) with coefficient=0.3964, P=0.006 (significance level = 0.05) which is significant. Also, there is a positive correlation between income and recyclable (paper + mix-recyclable) generation per capita (figure 13 in appendix) with coefficient=0.8617, P=0.0001which is also statistically significant.

The second cluster (light yellow) contained most of the community districts that had higher household income and the median household income in this cluster was $105,616 (Figure 14 in appendix).

The result of regression analysis showed differences between the second and the first cluster. There was a negative correlation between income and refuse generation per capita (Figure 15 in appendix) with coefficient=-0.557, P=0.048 which is significant. And there is a positive correlation between income and recyclable (paper + mix-recyclable) generation per capita (figure 15 in appendix) with coefficient=0.7175, P=0.006.

From the above analysis result, we can conclude that Income does affect waste generation. In the lower income cluster, with the rise of household income, refuse generation per capita increases in tandem, so does recyclables (paper+mix-recyclables) generation per capita. But in the higher income cluster, refuse generation per capita decreases while household income rises even though the tandem of recyclables generation remains positive while income increases. In short, social economic background has different influences across various income levels, especially in refuse generation, but not much influences on recyclables generation.

7.2 Covid-19 Waste Analysis

7.2.1 Time Series Prediction

Based on our time series model in 7.1.1, it can be seen that all the boroughs except Manhattan had a surge in waste generation in the Covid-19 periods, especially in June. This may be partly explained by the fact that people who lived outside Manhattan didn't have to go to Manhattan for work during stay-at-home policy.

7.2.2 Analysis on DSNY tonnage collection in space(Stay At Home Period ~ March-May)

In Figure 16(in appendix), it visualized the income level of each community district in New York City. Most of the high income districts (>$100,000) are in Manhattan and DownTown Brooklyn.

So, we calculated past 5-year waste tonnage average (March-May) and used it to subtract the stay at home period waste tonnage and generated some heatmaps.

In Figure 17(in appendix), lighter color indicates a bigger drop of refuse tonnage collected during the Stay at Home period compared to the 5-year refuse tonnage average. Blue and dark blue were negative number which means there was an increase in collected refuse tonnage. Every district in Manhattan experienced a significant drop in refuse collection (>800 tons) during the Stay at Home period, so does downtown Brooklyn and some districts in the Bronx.

In Figure 18(in appendix), yellow indicates tonnage a drop in recyclables(Paper+Mix-recyclables) tonnage collected during the Stay at Home period compared to the 5-year average. Other two colors indicate an increase. Again, all community districts in Manhattan experienced a recyclable waste tonnage drop along with refuse tonnage drop.

We applied the same concept as we used in section 7.1.2, Gaussian Mixture clustering with silhouette-score equals 0.4702. The algorithm provided two clusters that it calculated to have relations, figure 7.2.2-1 is a geographic representation of the two clusters,

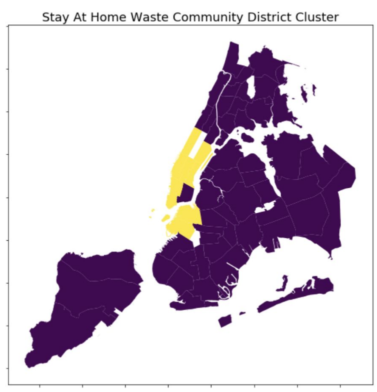

Figure 7.2.2-1 Covid-19 geographic Cluster Results

In the first cluster(dark blue), it contained community districts that had median households as $53,205, and most of the community districts distributed in the 50k-65k range. (figure 19 in appendix)

The result of regression model shows that there was a moderate positive correlation between income and refuse generation per capita (figure 20 in appendix) with coefficient=0.3638, P=0.009 which is significant (Significance Level = 0.05), and strong positive correlation between income and recyclable (paper + mix-recyclable) generation per capita (figure 21 in appendix) with coefficient=0.8469, P=0.0001 which is significant.

The second cluster (light yellow) contained most of the community districts that had higher household income and the median household income in this cluster was $115,993 (figure 22 in appendix).

The result of regression showed there was a strong negative correlation between income and refuse generation per capita (figure 23 in appendix) with coefficient=-0.7784, P=0.013 which is significant. Also, a very weak negative correlation between income and recyclable (paper + mix-recyclable) generation per capita (figure 24 in appendix) with coefficient=-0.0136, P=0.972 which is not significant, and that further prove the abnormal waste generation pattern during Covid-19 Stay at Home period.

In conclusion, compared to the analysis result in section 7.1.2, we can see that household income has even stronger influences on waste generation in high-income community districts. The coefficient of income & refuse per capita during stay at home period(-0.7784) is much steeper than normal time(-0.557), and coefficient of income & recyclable waste per capita during stay at home(-0.0136) is completely different compared to normal time(0.7175). Meanwhile, the lower income cluster group maintained a relatively identical pattern of waste generation during the Stay at Home period compared to normal time. During the stay at home period, most of the businesses including restaurants were closed for good, which means residential waste would increase as mentioned in [H2]. With the evidence of a dramatic shift in waste generation pattern in two cluster groups, we can conclude that residents' chances of fleeing NYC before/during the Stay At Home period has a positive correlation to their income level thus proves [H2] to be true. In another word, if a New Yorker lived in Manhattan with annual income greater than $100,000, the chance of that New Yorker leaving the city before/during the Stay at Home period is higher than New Yorkers in other communities.

7.3 Potential Business Model

7.3.1 Data Collection Method

The current data collection method by DSNY is good, but not sufficient enough to be used as promoting waste recovery business models. For example, the data does not contain a detailed classification of MGP, an abbreviation of recyclable metal glass, plastic and beverage cartons, and it only contains gross tonnage of whatever the department collected from NYC blue recycle bins even though these materials will be separated and weighted at one of the DSNY processing facilities. And that is a data gap. Data gaps constrain business cases and government measures to promote resource efficiency. For the DSNY per se, this gap could be filled with more comprehensive data collection. For example, the department could collect its data from Sims Municipal Recycling, the corporation that handles 100% of NYC recyclable plastic & metal and 50% of recyclable paper. To initiate this procedure, the city government should follow the law and policy of data sharing between private sectors and government agencies. The framework should be DSNY supply residential recyclables to Sims Municipal, Sims Municipal process the mixed waste and weigh them each of the categories at their facility, then DSNY receives updated weight data of each category of the recyclable waste and aggregates them to NYC open data portal.

7.3.2 Technology Approaches

BASF SE is a German chemical company and the largest chemical producer in the world, which offers compostable plastics based on renewable raw materials like Ecovio Plastic bags. This kind of new technology plastic product could improve organic waste recovery and utilization at the source when they were introduced to restaurants and households. Although NYC is cutting its budget on organic waste collection, once the city recovers from the pandemic, NYC organic waste treatment plan could encourage restaurants and households to replace traditional plastic tableware and refuse bags into compostable plastic products by Increasing publicity efforts and providing corresponding subsidies. The average cost of conventional plastic (fossil-based) is 1350 EUR/ton vs. average cost Ecovio: 3600 EUR/ton. The New York City government can weigh the pros and cons in the subsequent implementation of the policy and achieve maximum efficiency recovery at the lowest priceLeuven (2018, November 29).

Besides that, NYC has temporarily suspended organic recycling programs, the economic values and environmental benefits of organic waste are quite significant if more private sectors could get involved in organic waste treatment. To get the private sectors to make up the absence of the city government in organic waste recycling, subsidies and technician support are essential at the beginning. Technology like enzyme reaction, which produces quality fertilizer, is good for small-batch organic waste treatment that produces good end products in the meantime. This tech is good for an urban setting. One of the recommendations is to build a voluntary based community organic recycling program, and the city government could use the remaining budget from the original organic recycling plan to pay for equipment like an enzyme reaction kit that is small and odor-free. And the entire program will be volunteer-based to minimize the cost to sponsor the program. Another approach could be distributing biowaste fermentation bins (avg cost $70 each, 50 pounds capacity) to households willing to participate in the self recycling program sponsored by the city.

Another approach could be separating recyclables at the source of collection, such as AI-powered recycling bins (CleanRobotics https://cleanrobotics.com/), which can recognize the type of recyclables and automatically sort them into designated bins. Each bin's cost ranges from $1,500 to $5,000 depending on the models, with similar size compared to ordinary recycling bins that are usually seen in malls and airports, and they are built for indoor purposes. The scalability of this emerging technology is limited due to the cost of operation and maintenance, but it could be trailed at education and public facilities to validate its feasibility.

7.3.4 Policy Approaches

One of the most efficient ways to reduce waste from the consumer side is to encourage repair. Rather than repairing broken items, many people choose to buy new. Repairing can get broken things to work again. This is one of the easiest ways to recycle unusable items. Contrast to recycle waste and later make it to other products. Repairing is more efficient, time-saving, and convenient. Repair is one of the most important principles of the circular economy and should be widely supported. There are two cases in the EU as follows:

> Tax reduction on repairing: To encourage the repair of goods, the Swedish government has introduced a tax reduction that allows people to get back half of the repairing cost on appliances such as refrigerators, ovens, dishwashers, and washing machines. For repairing bicycles, shoes, leather goods, clothing, and household linen, the VAT also decreased from 25% to 12%.

Repair Café: Established in Amsterdam in 2009, Repair Café aims to bring people together to socialize, use their skills, and reduce waste at the same time. In a Repair Café, different tools and materials are provided to help people repair clothes, furniture, appliances, bicycles, toys, etc. People can also get help from specialists such as electricians, seamstresses, carpenters, and bicycle mechanics.

Both cases above are applicable in NYC. While tax reduction is relatively hard to implement due to law and policy restrictions, city administration can help to foster Repair Café in residential districts. Repair Café can hire volunteers to reduce the cost of maintenance for the cafe.

8. Conclusion/limitations/implications

This project aims to understand and predict the waste generation pattern in New York City. The project also focused on the Covid-19 period waste analysis and tried to understand the correlation between socioeconomic background and waste generation behaviors. We tackled the situation with various models and methods, including LSTM, clustering, and regression, to test the initial hypotheses.

In the normal period, we found waste generation fluctuates over the years. There is seasonality in all five boroughs, while there isn't any noticeable trend in these time-series data. People tend to generate more waste during the summertime than in wintertime. And there was a high positive correlation between the three main types of waste generation. Besides showing the trend and relationship, our LSTM model could predict the waste generation volume in each borough by waste type.

We also found out that income level has a negative correlation with refuse generation during the normal time. Both refuse and recyclable waste has a negative correlation with income during the Covid-19 period. In other words, wealthier people tend to leave the city during emergencies, and that behavior reflects on the waste generation pattern we learned from our analysis.

And based on those analyses, we also provided potential business models to support private sectors and government to better utilize recyclable waste resources and develop the circular economy.

However, our study also has some limitations; for example, we lack commercial waste generation data. If we could start our analyses from the commercial waste branch, our study will create more economic value. Also, to get some business data in the future, we can consider more socioeconomic factors and models to get more accurate results.

9. Project Planning

Figure 25(in appendix) shows the Gantt chart represents our project timeline, the works we have done in this project. Although this project mainly focuses on the recycling and utilization of residential waste, once New York City recovers from the pandemic, our sponsors including DSNY and NYCEDC could collect more commercial waste data and involve more businesses in circular activities to make the city more environmentally friendly and sustainable.

10. Team Roles

Team Member	**Team Roles**
Yang Li	Data Collecting, Web Developing, Research, Report and Presentation Preparing
Rongjian Yang (Team Leader)	Data Collecting, Model developing, Research
Yong Fan	Data Collecting, Lead model developing

Reference

- Dana Rohrabacher. (2013, March). Examining the effectiveness of the NIST laboratories : hearing before the Subcommittee on Technology, Committee on Science, Space, and Technology, House of Representatives, One Hundred Thirteenth Congress, first session, Wednesday, March 20, 2013. Retrieved from https://www.worldcat.org/title/examining-the-effectiveness-of-the-nist-laboratories-hearing-before-the-subcommittee-on-technology-committee-on-science-space-and-technology-house-of-representatives-one-hundred-thirteenth-congress-first-session-wednesday-march-20-2013/oclc/855998726
- Zero Waste Challenge. (2016, February). Retrieved March 31, 2020, from https://www1.nyc.gov/site/sustainability/initiatives/zero-waste-challenge.pag
- Niall Smith. (2019, February). US Tops List of Countries Fuelling The Mounting Waste Crisis. Retrieved April 30, 2020, from https://www.maplecroft.com/insights/analysis/us-tops-list-of-countries-fuelling-the-mounting-waste-crisis/
- People recycle more when they know what recyclable waste becomes. (2019, May). Retrieved from https://www.sciencedaily.com/releases/2019/05/190516103712.htm
- DSNY. (2020, April 8). DSNY Monthly Tonnage Data. Retrieved from https://data.cityofnewyork.us/City-Government/DSNY-Monthly-Tonnage-Data/ebb7-mvp5
- Navarro-Esbrí, J., Diamadopoulos, E., Ginestar, D.(2002). Time series analysis and forecasting techniques for municipal solid waste management. Resour. Conserv. Recycl. 35, 201–214. https://doi.org/10.1016/S0921-3449(02)00002-2.
- Bandara, N. J. G. J., Hettiaratchi, J. P. A., Wirasinghe, S. C., & Pilapiiya, S. (2007). Relation of waste generation and composition to socio-economic factors: a case study. *Environmental Monitoring and Assessment*, *135*(1-3), 31–39. doi: 10.1007/s10661-007-9705-3
- Meza, J. K. S., Yepes, D. O., Rodrigo-Ilarri, J., & Cassiraga, E. (2019). Predictive analysis of urban waste generation for the city of Bogotá, Colombia, through the implementation of decision trees-based machine learning, support vector machines and artificial neural networks. *Heliyon*, *5*(11). doi: 10.1016/j.heliyon.2019.e02810
- sklearn.metrics.silhouette_score¶. (n.d.). Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html
- Leuven. (2018, November 29). Business case - Ecovio® compostable polymers with biobased content. http://afdimpact.org/wp-content/uploads/2018/11/basf-impact-case-day.pdf.

Appendix

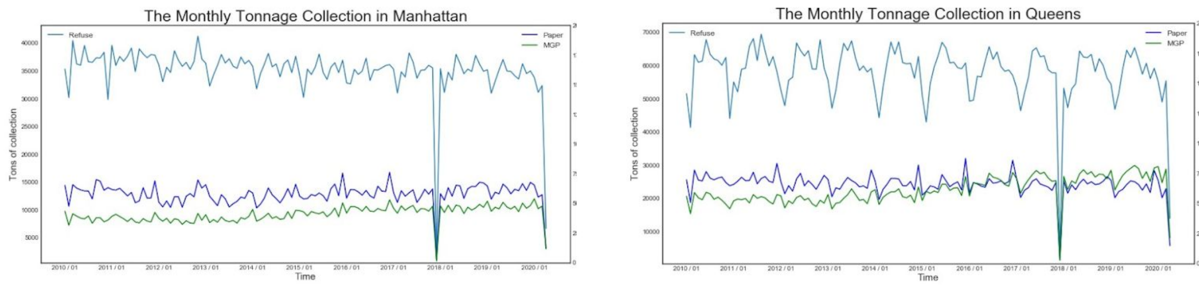

Figure 1: Monthly Tonnage Collection in Manhattan & Queens (From 2010 until now)

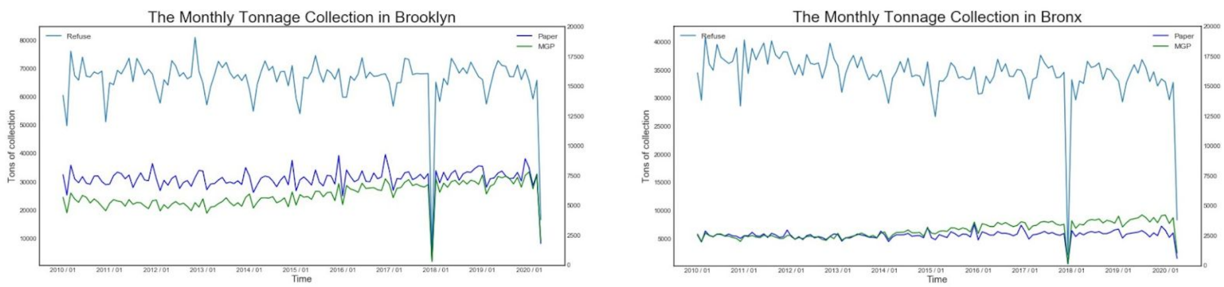

Figure 2: Monthly Tonnage Collection in Brooklyn & Bronx (From 2010 until now)

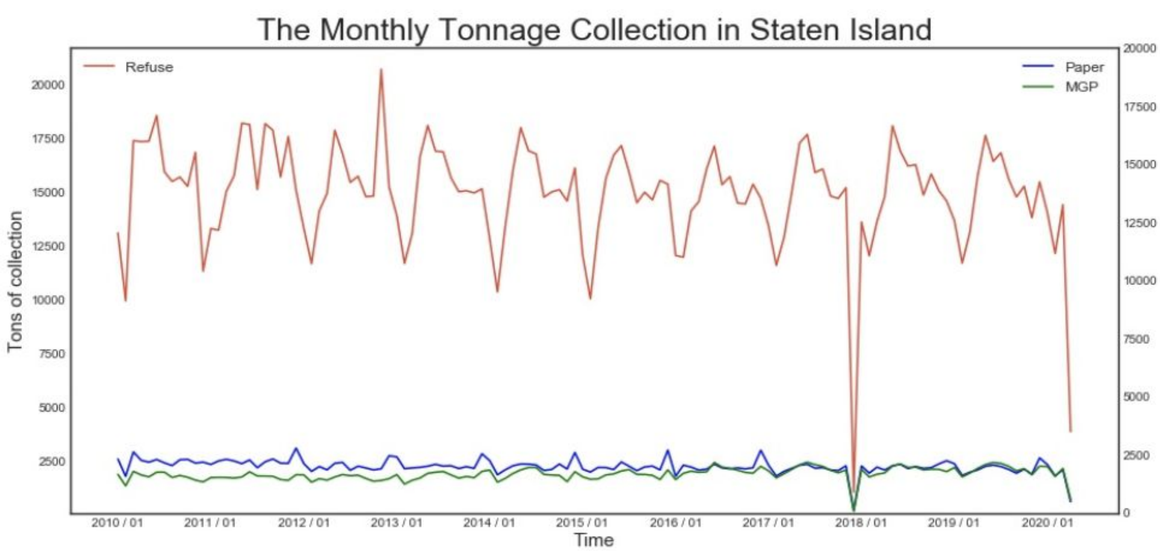

Figure 3: Monthly Tonnage Collection in Staten Island (From 2010 until now)

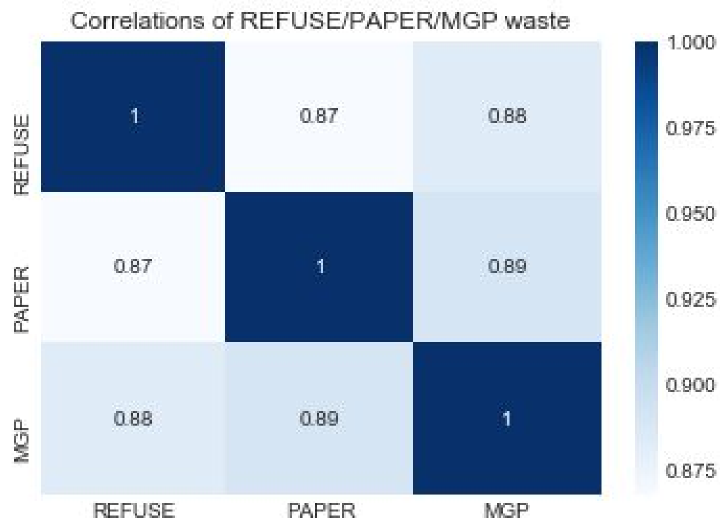

Figure 4: The correlation among three waste types

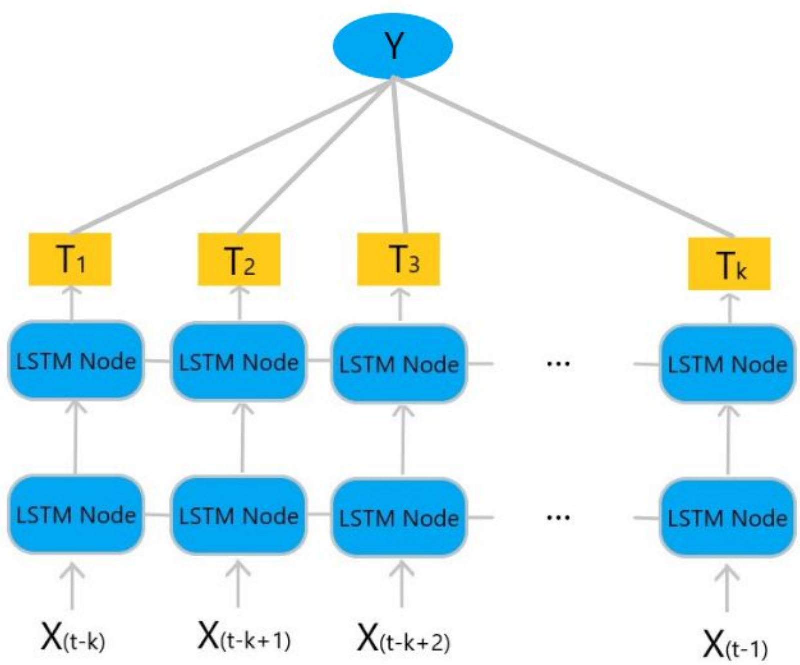

Figure 5: Structure of the LSTM model

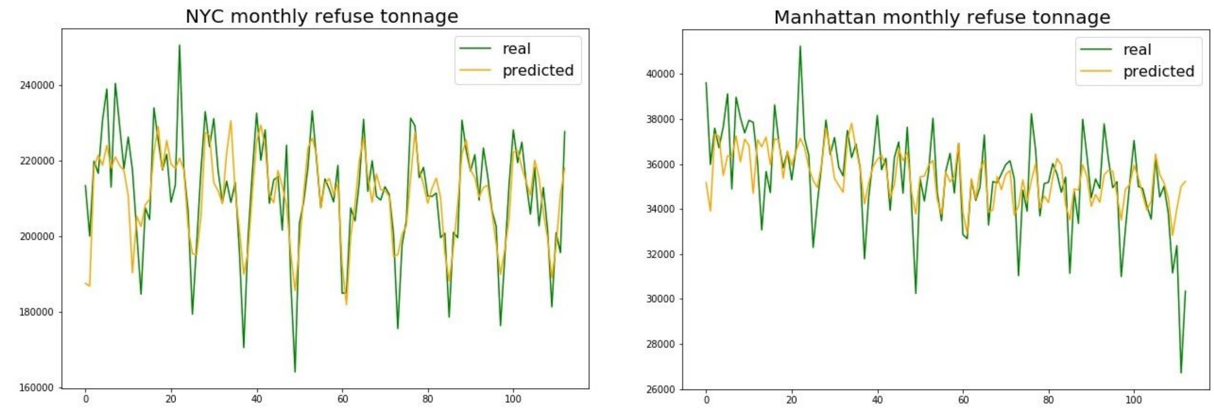

Figure 6: Real and Predicted Value in NYC and Manhattan (from 2010 until now)

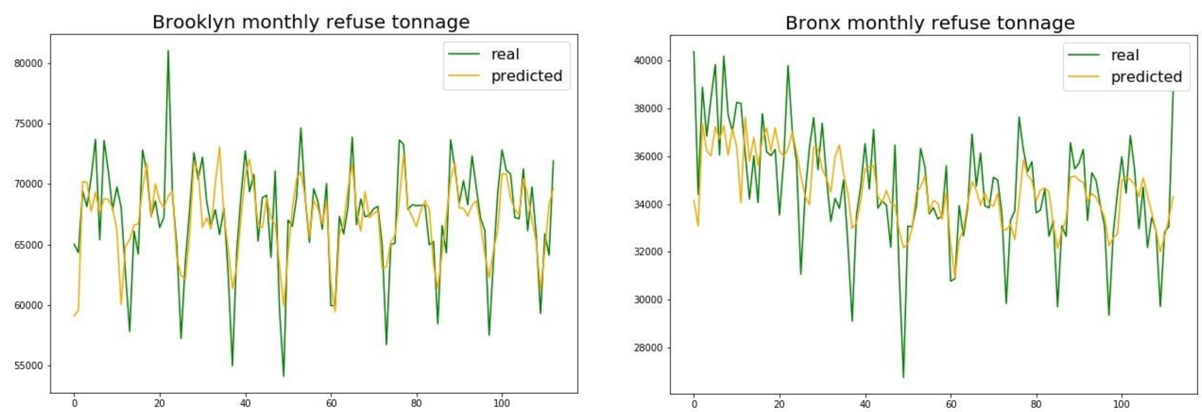

Figure 7: Real and Predicted Value in Brooklyn and Bronx (from 2010 until now)

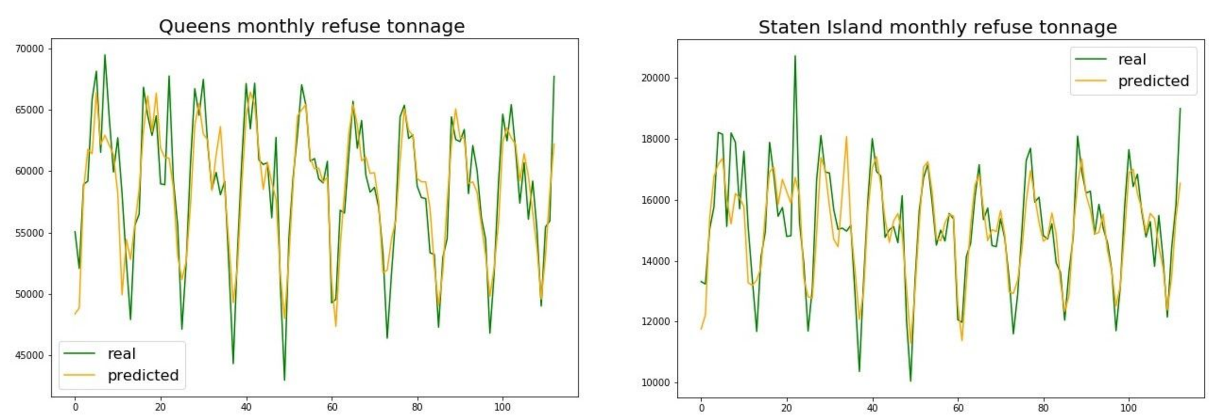

Figure 8: Real and Predicted Value in Queens and Staten Island (from 2010 until now)

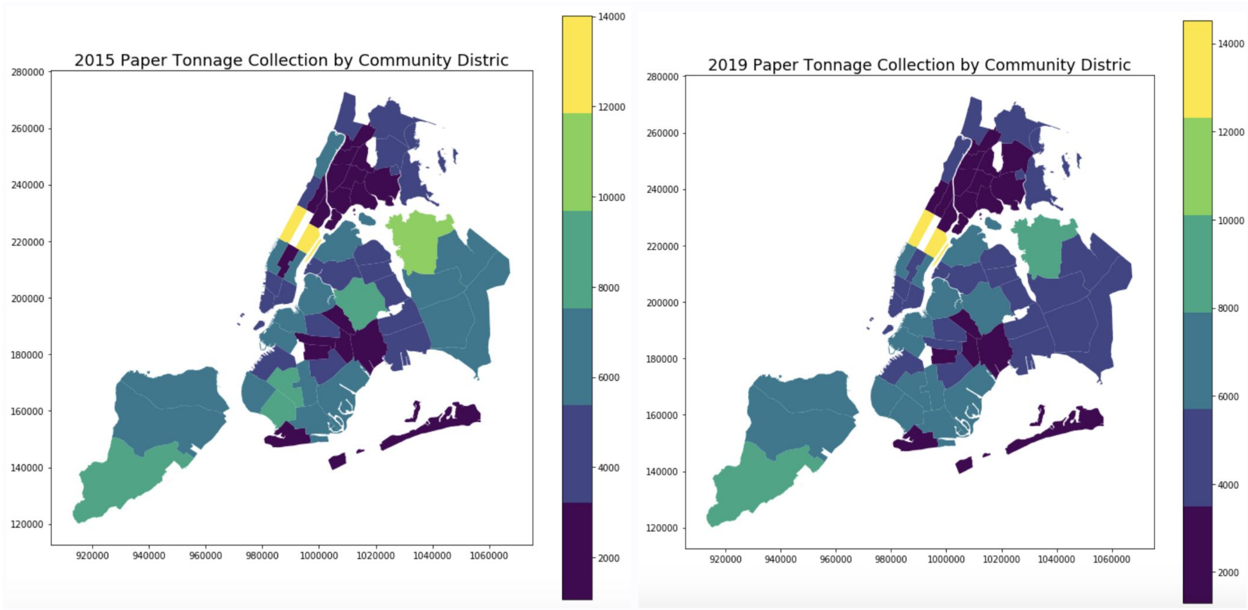

Figure 9:Heatmap of Community District pre-separated Paper tonnage 2015 vs. 2019

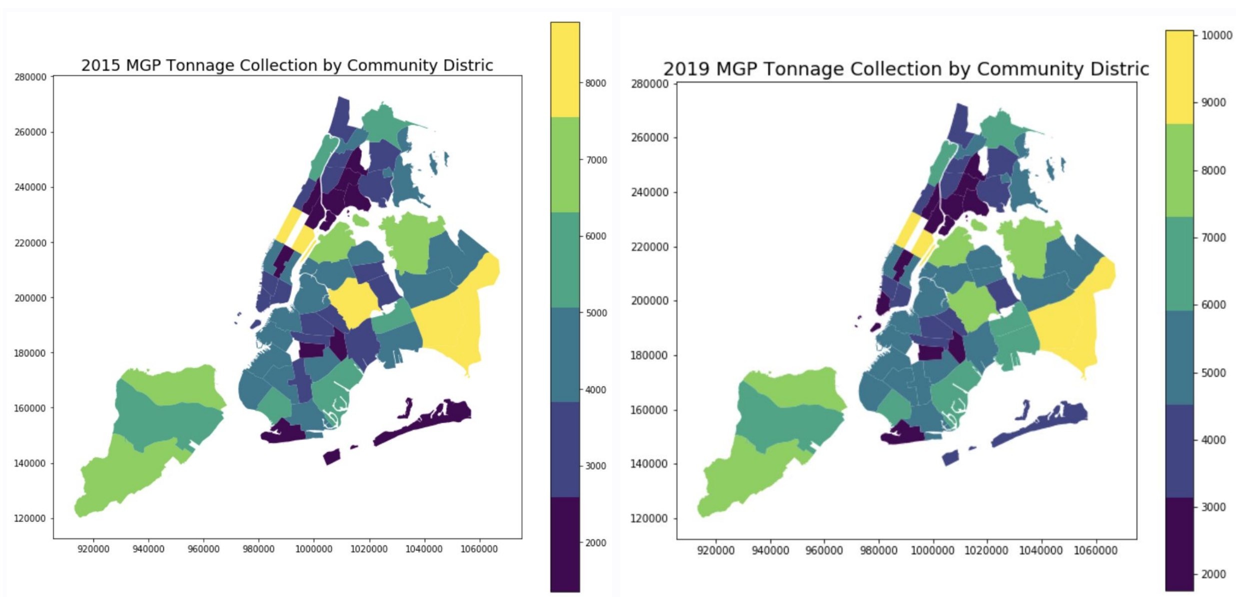

Figure 10: Community District pre-separated Mixed-recyclable tonnage 2015 vs. 2019

`Cluster Median Income: 52490.39936`

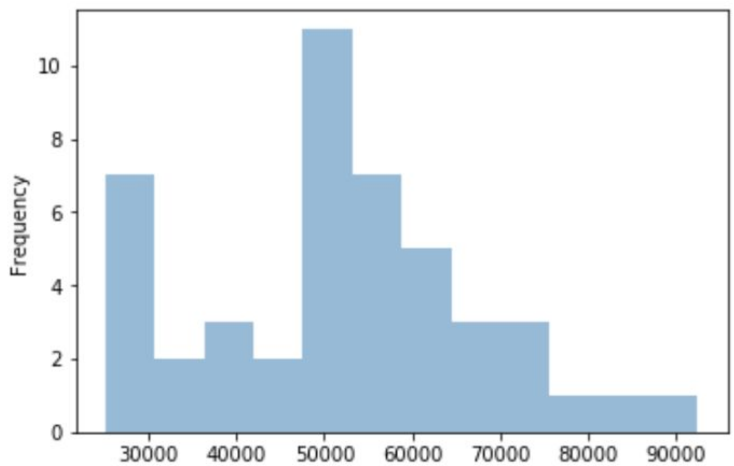

Figure 11: Normal Time Cluster0 Income Distribution

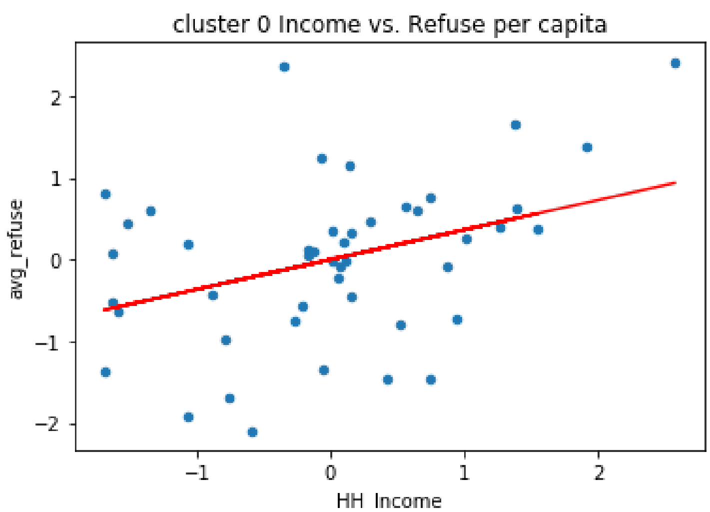

Figure 12: Normal Time Cluster0 Income vs. Refuse per Capita

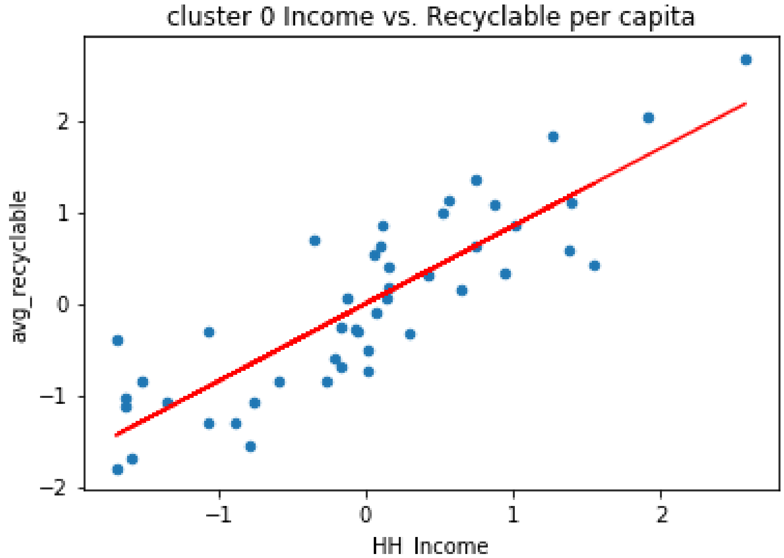

Figure 13: Normal Time Cluster0 Income vs. Recyclables per Capita

```
Cluster Median Income: 105616.53869999999
```

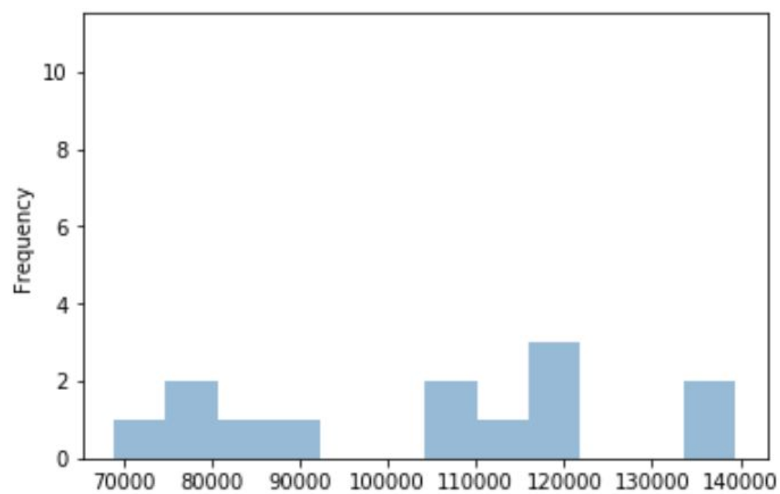

Figure 14: Normal Time Cluster1 Income Distribution

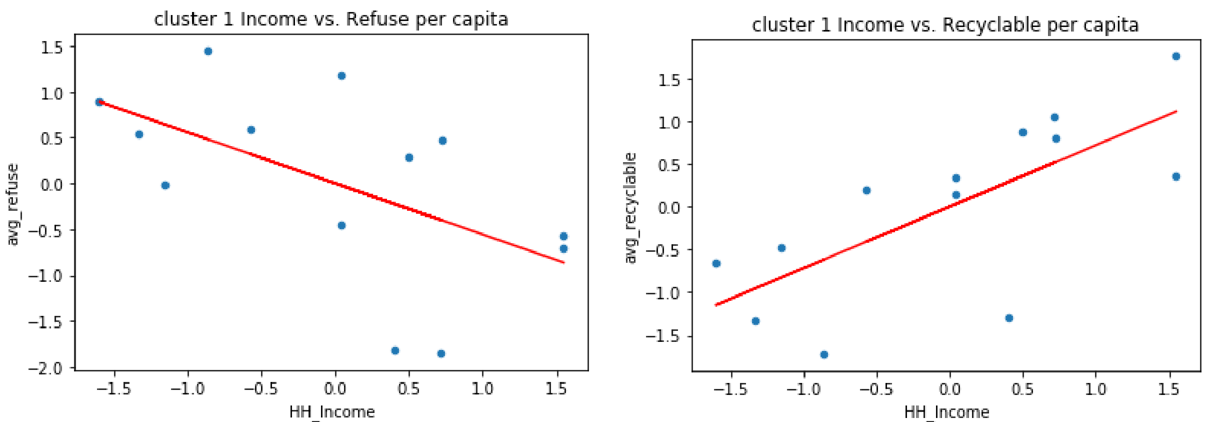

Figure 15: Normal Time Cluster1 Income vs. Refuse/recyclables per Capita

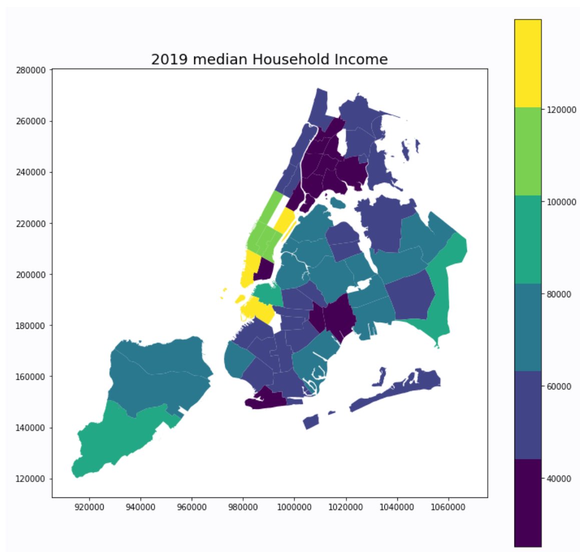

Figure 16: 2019 Median Household Income in NYC

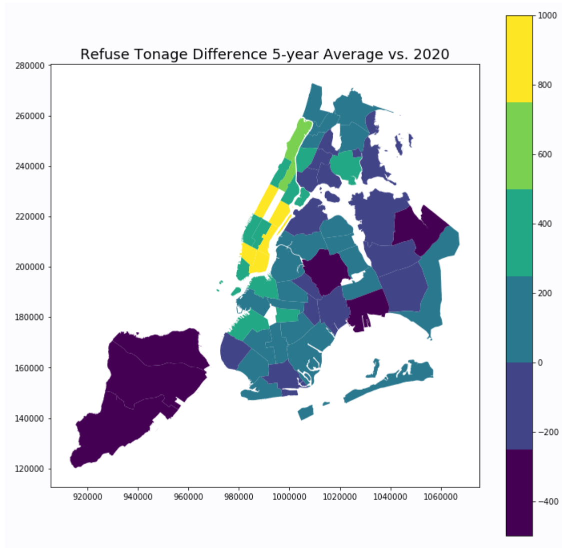

Figure 17: Refuse Tonnage Difference between month on month 5-year average and 2020 Covid-19 period

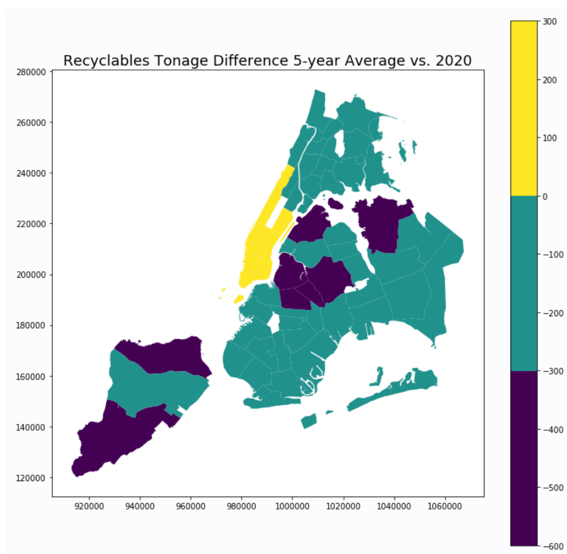

Figure 18: Recyclables Tonnage Difference between month on month 5-year average and 2020 Covid-19 period

53205.295186666655

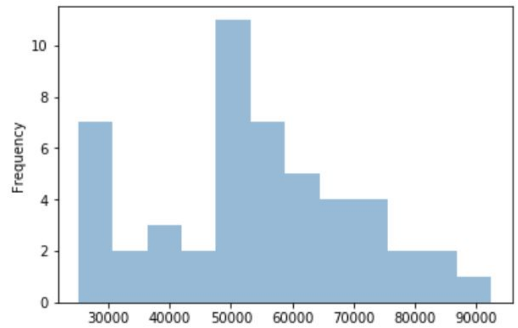

Figure 19: Covid-19 Cluster0 Income Distribution

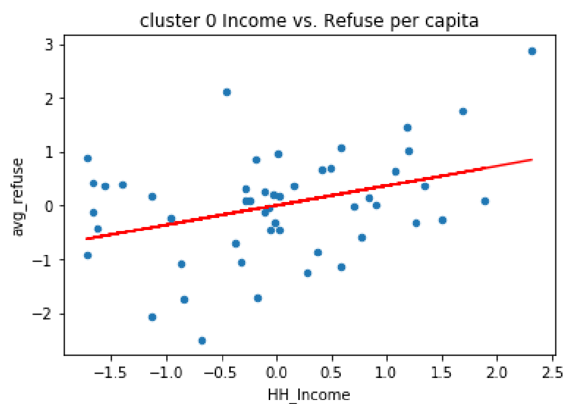

Figure 20: Covid-19 Cluster0 Income vs. Refuse per Capita

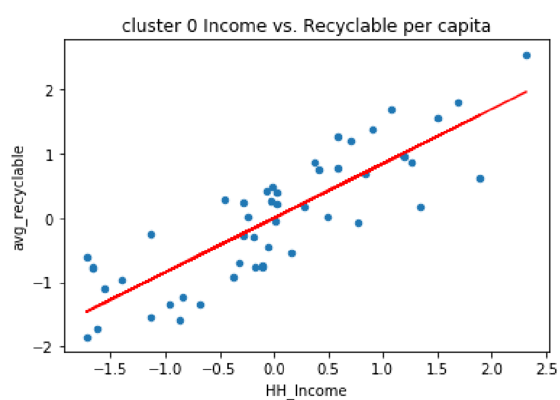

Figure 21: Covid-19 Cluster0 Income vs. Recyclables per Capita

115993.55589999999

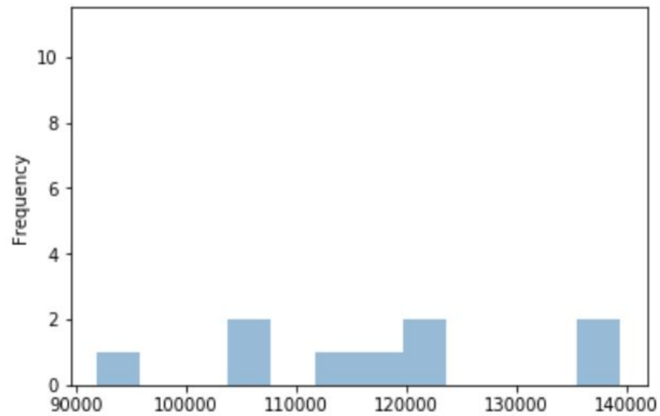

Figure 22: Covid-19 Cluster1 Income Distribution

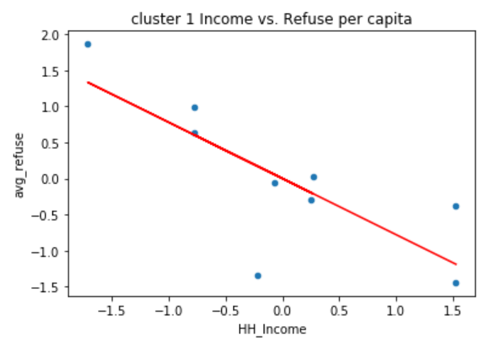

Figure 23: Covid-19 Cluster1 Income vs. Refuse per Capita

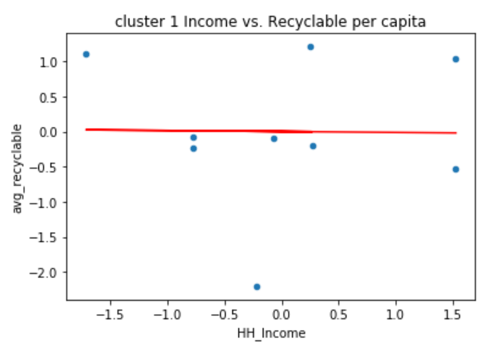

Figure 24: Covid-19 Cluster1 Income vs. Recyclables per Capita)

Tasks	Status	Start Date	Durarion (da	Due Date
Prepare for Final Presentation	Done	7/15/20	8	7/21/20
Document Code in Github	Done	7/11/20	4	7/14/20
Build a Website	Done	7/3/20	8	7/10/20
Finalize Results	Done	6/27/20	6	7/2/20
Conduct Recommendations	Done	6/23/20	4	6/26/20
Conduct Third Progress Report	Done	6/11/20	10	6/20/20
Make Improvement on Models	Done	5/26/20	15	6/10/20
Refine Hypotheses	Done	5/21/20	4	5/25/20
Refine Research Questions	Done	5/16/20	4	5/20/20
Modeling and Analysis	Done	5/2/20	3	5/5/20
Conduct Second Progress Report	Done	5/2/20	3	5/5/20
Request for Plan B Approval	Done	5/1/20	1	5/2/20
Conduct Plan B Proposal	Done	4/30/20	1	5/1/20
Discuss Methods	Done	4/29/20	1	4/30/20
Collect Data	Done	4/29/20	1	4/30/20
Meet with Professors and Sponsors	Done	4/29/20	0	4/29/20
Set a Meeting with Professor	Done	4/27/20	2	4/29/20
Conduct Plan B	Done	4/16/20	9	4/25/20
Report Issues to Sponsors	Done	4/11/20	4	4/15/20
Estimate Missing Data	Done	4/5/20	5	4/10/20
Evaluate the feasibility of Data	Done	4/2/20	2	4/4/20
Present First Progress Report	Done	4/1/20	0	4/1/20
Conduct First Progress Report	Done	3/21/20	7	3/28/20
Discuss Methodology	Done	3/16/20	4	3/20/20
Collect Data	Done	3/10/20	5	3/15/20
Research on Related Materials	Done	3/6/20	3	3/9/20
Approve Project Charter	Done	2/29/20	5	3/5/20
Create Project Charter	Done	2/26/20	2	2/28/20
Identify Business Case	Done	2/20/20	5	2/25/20

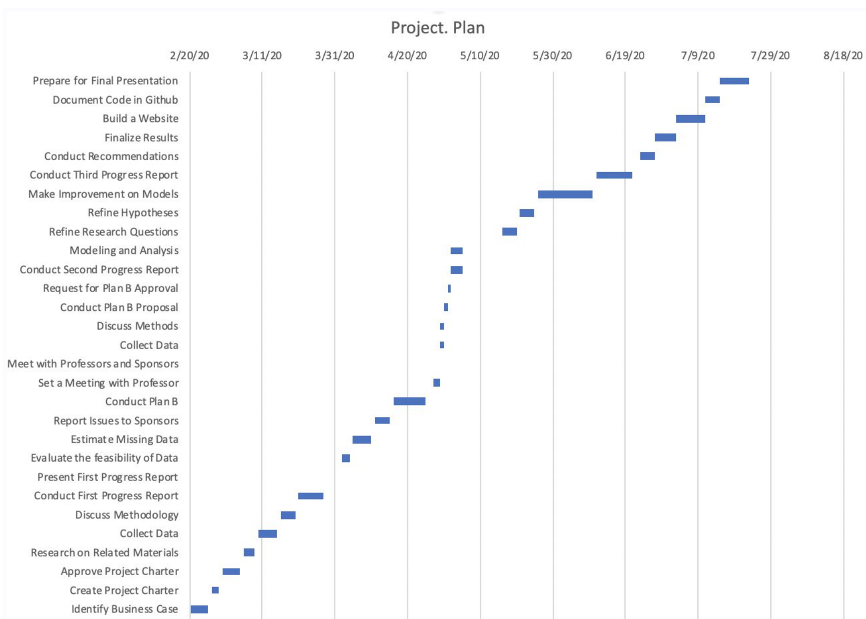

Figure25: Gantt Chart

	Training loss (MSE)	Testing loss (MSE)
NYC	0.01418	0.0064
Manhattan	0.0105	0.0189
Brooklyn	0.0146	0.0067
Queens	0.01182	0.0066
Bronx	0.01850	0.0105
Staten Island	0.00999	0.0052

Table 1: Training and test losses of five models