ELECTRIC CONSUMPTION IN NEW YORK CITY

By:
Prerna Yadav
Karan Gaikwad
Mahesh Kate
Ashutosh Deulkar
Harshal Nanavare

INTRODUCTION

In New York City, pricing and usage of electricity are significant variables that have an impact on both residents and companies. New York City is one of the biggest and most populated cities in the world, making it a significant electricity consumer.

DATA

Source: Electric Consumption and Cost 2010 - Feb 2023

The dataset is maintained by the New York City Department of Housing Preservation and Development (HPD) and provides information on electric usage and costs for residential, commercial and public housing properties across the city spanning from 2010 to February 2023.

NUMBER OF ROWS	447849
NUMBER OF COLUMNS	27

PROBLEM STATEMENT

The electricity consumption and cost is increasing year by year in new york city

In this project our goal is to give overview of cost and consumption as well as find out potential influences across the neighbourhood boroughs of the new york city.

Furthermore, we also aim to analyse chances of cost and energy savings

In order to accomplish these goals, we would be performing descriptive and predictive analytics

Glimpse of Data

In this project our goal is to give consumption as well as find out potential influences across the neighbourhood boroughs of the new york city.

Furthermore, we also aim to analyse chances of cost and energy savings

In order to accomplish these goals, we would be performing descriptive and predictive analytics

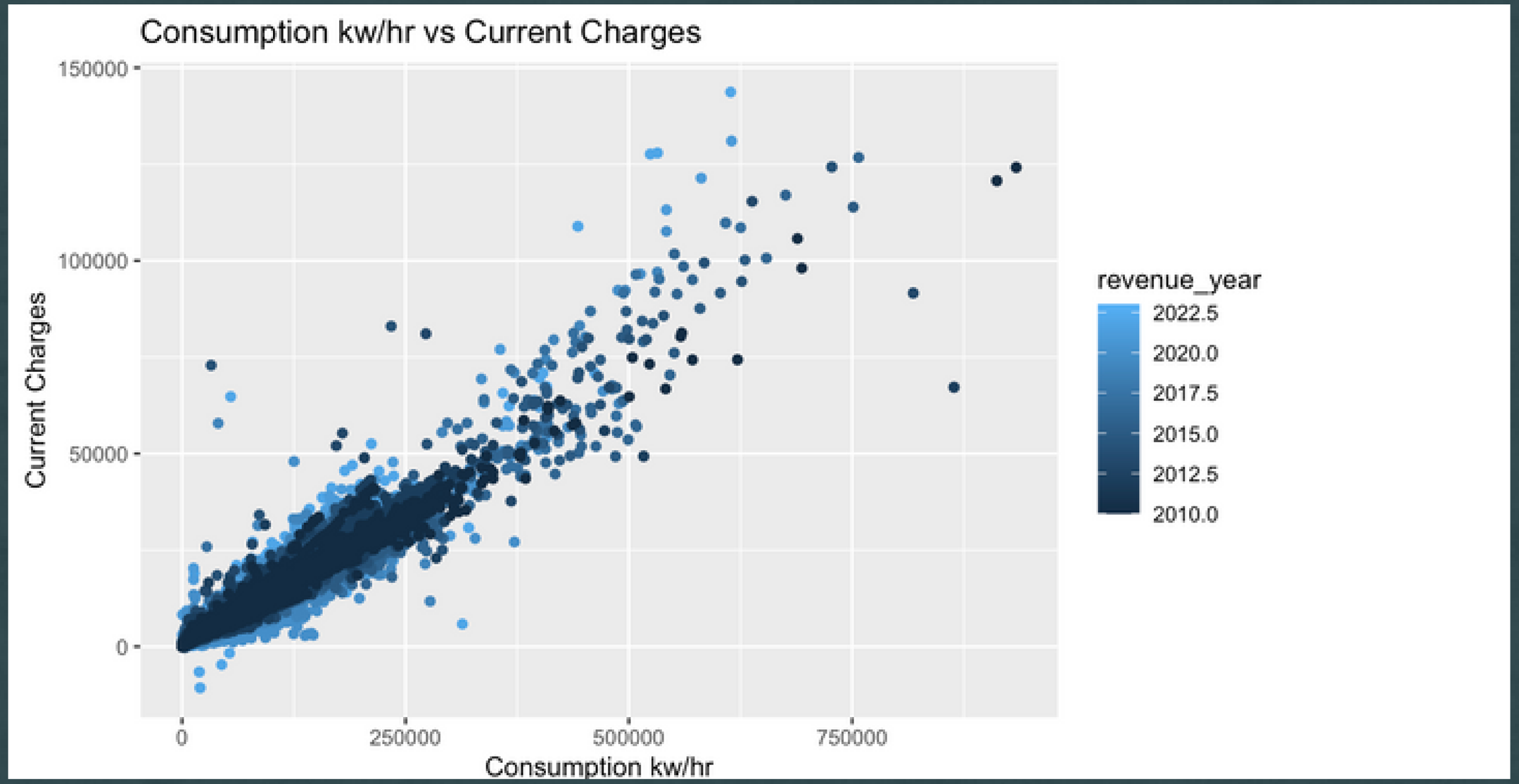
Column Name	Description	Expected Values
Borough	Borough	Bronx, Brooklyn, Manhattan
Development Name	Development name	The name of the housing dev as listed in the Development Book.
Location	Building number. In order to run an analysis by building, you can use a combination of TDS and building number which gives a unique identifier for each building	The building number of the
Meter AMR	Is the meter Automatic Meter Reading (AMR), Interval or none	AMR, Interval, None
Meter Scope	The buildings or areas the account and meter supply	Listing of the buildings/Area
TDS #	TDS (Tenant Data System) number is the unique identifier for all NCYHA developments. It is recommended to use it in order to run analysis by development. The TDS is also the unique link between NYCHA data sets.	Number. The non development facilit identified in the field Fundin don't have a TDS #. For these use EDP or RC Code.
EDP	NYCHA Electronic Data Processing. Number used to identify individual NYCHA developments. EDP is used by NYCHA only to link data issued from a different system (the energy management system that was used by NYCHA before 2010). It is recommended to use the TDS # as a unique identifier of each development.	Three digit number
RC Code	NYCHA budget responsibility code. Code representing a specific development.	Letter indicating the boroug series of numbers.
Funding Source	The development's funding source including Federal, Mixed Finance, or an indication that the facility is a non development facility which means a non residential facility.	Federal, Mixed Finance, or N Development Facility
AMP number	Abbreviation for Asset Management Project (AMP) numbers. HUD Development asset tracking number. An AMP number can consist of more than one development.	NY and a series of numbers
Vendor Name	Utility vendor name	Vendor name
Revenue Month	Year and month of bill: 2016-01	2016-01
Service Start Date	Bill start date	Date
Service End Date	Bill end date	Date
number days	Number of days on bill	Number of days on bill
Meter Number	Meter number	Meter number
Estimated	Meter was not read for the time period. The consumption and cost are estimated. (Data is updated with actual reads once the meter is read)	Yes (Y) or No (N)
Current Charges	Total costs	Dollar value
Rate Class	The rate applied to the account. Details about each rate (dollar value) are available on the vendor web site.	Rate code is listed.
Bill Analyzed	The bill was analyzed for billing errors by NYCHA's Utility Management system during the billing period	Yes (Y) or No (N)
Consumption (KWH)	Total KWH consumption	KWH consumption
KWH Charges	Total KWH charges	Dollar value
Consumption (KW)	Total KW consumption	KW consumption
KW Charges	Total KW charges	Dollar value
Other charges	Total other charges	Dollar value

Descriptive Statistics

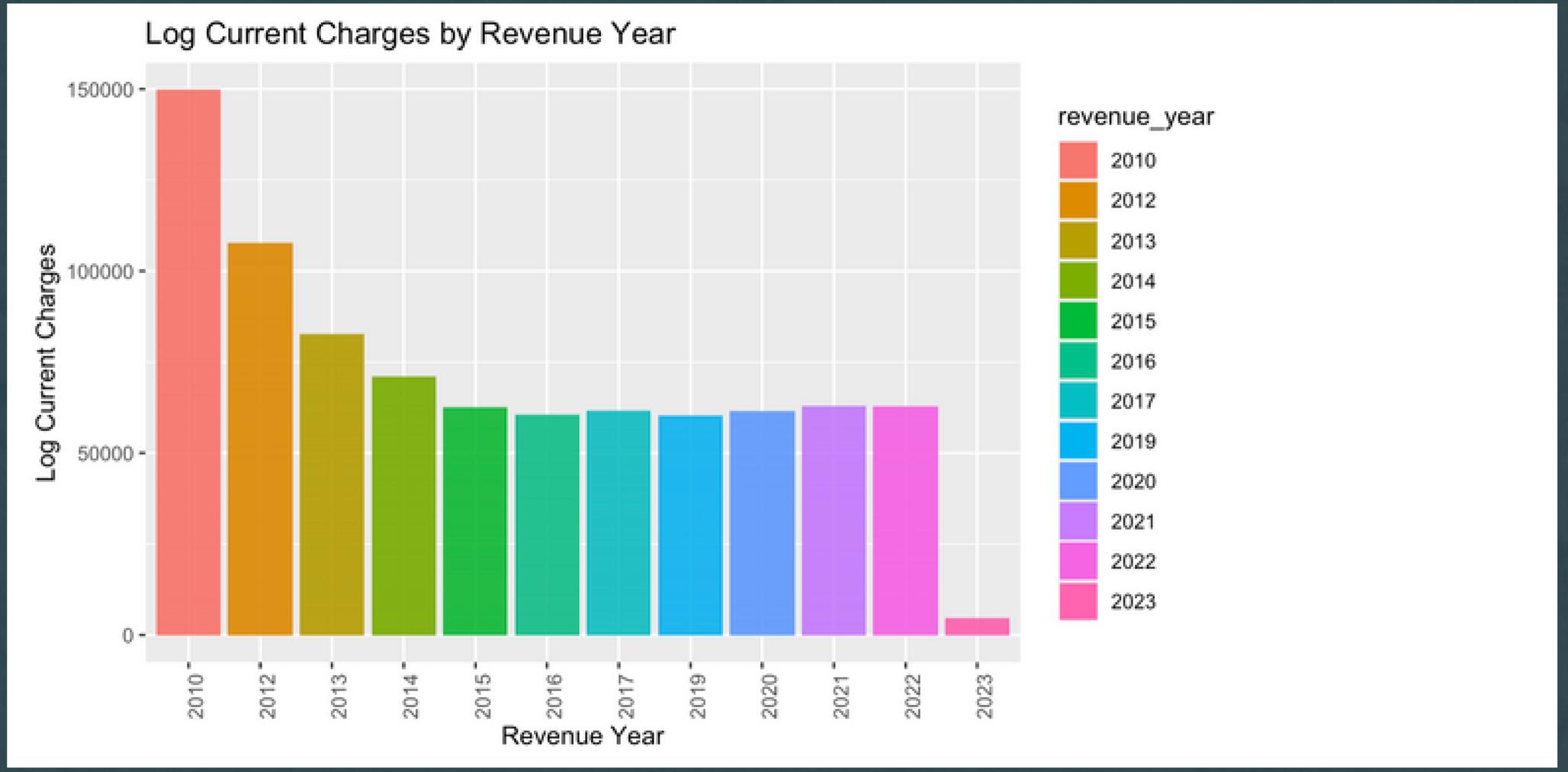
The above mentioned data spans for the period of 23 years and it provides information on electricity consumption and billing for various commercial and rental spaces around NYC. The mean consumption of electricity is 31,000 kwh, with a standard deviation of 72000 kwh. The mean consumption of power(kw) is 64kw, with a standard deviation of 334kw. The average billing period is around 30.4 days, with a standard deviation of 15.8 days. The mean current charges are $4,278, with a standard deviation of 13,726. The average kwh charges are $1,432 with a standard deviation of $11,571. The average kw charges are $912, with a standard deviation of $6,104. The mean other charges are $1,933 with a standard deviation $22,441.

Exploratory Data Analysis

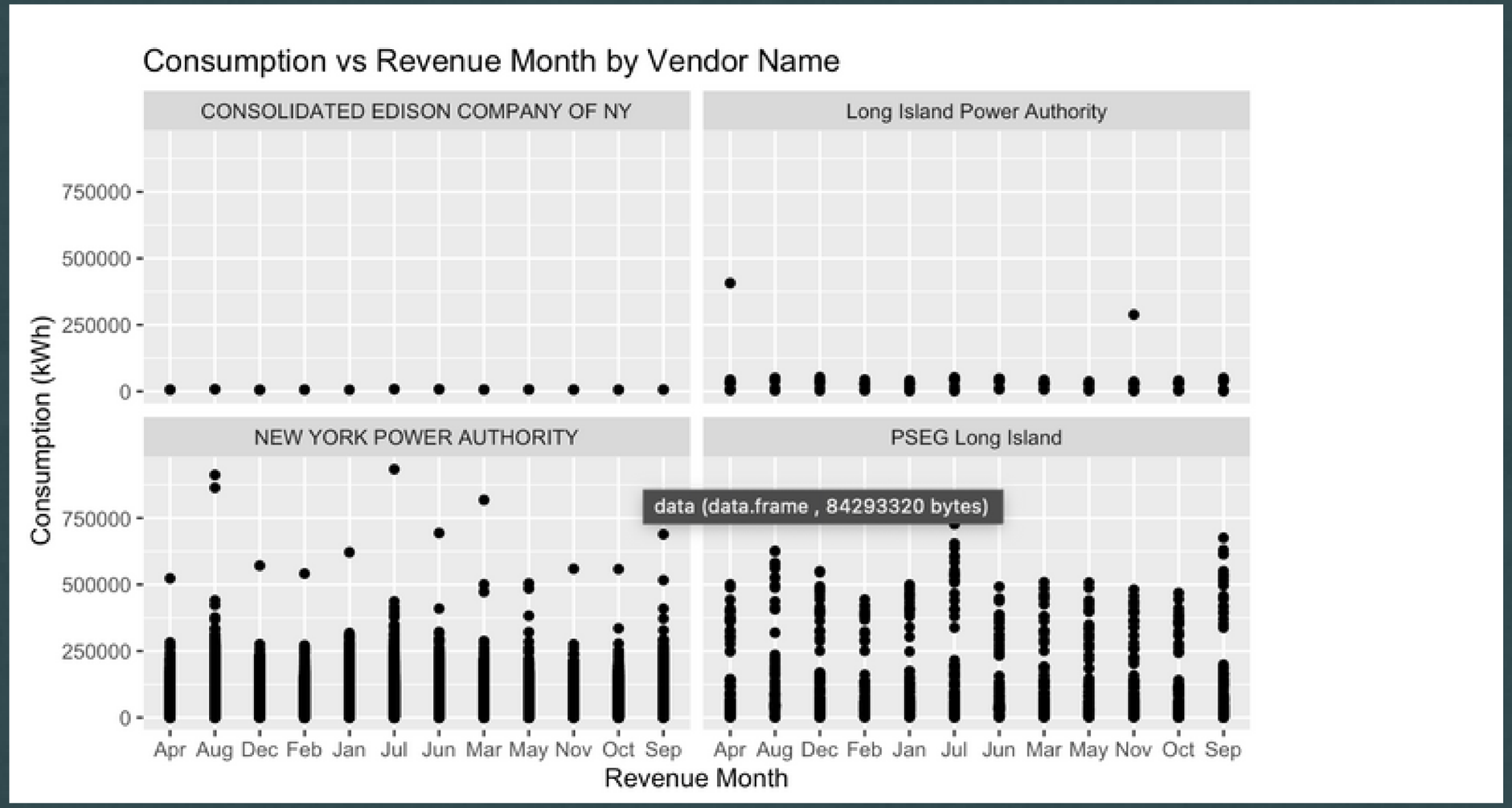
Consumption(kWh) vs Current Charges for all the years



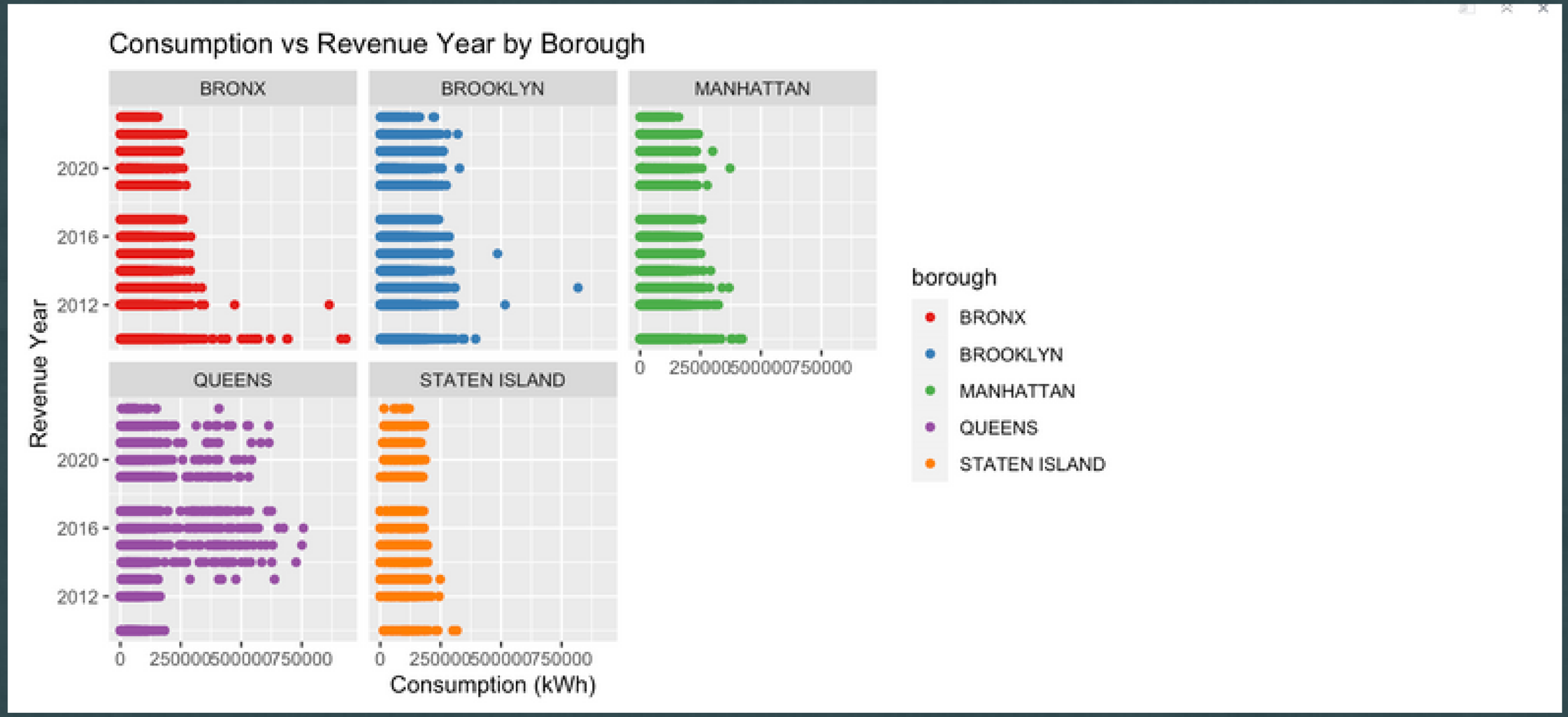
Current charges for all the years



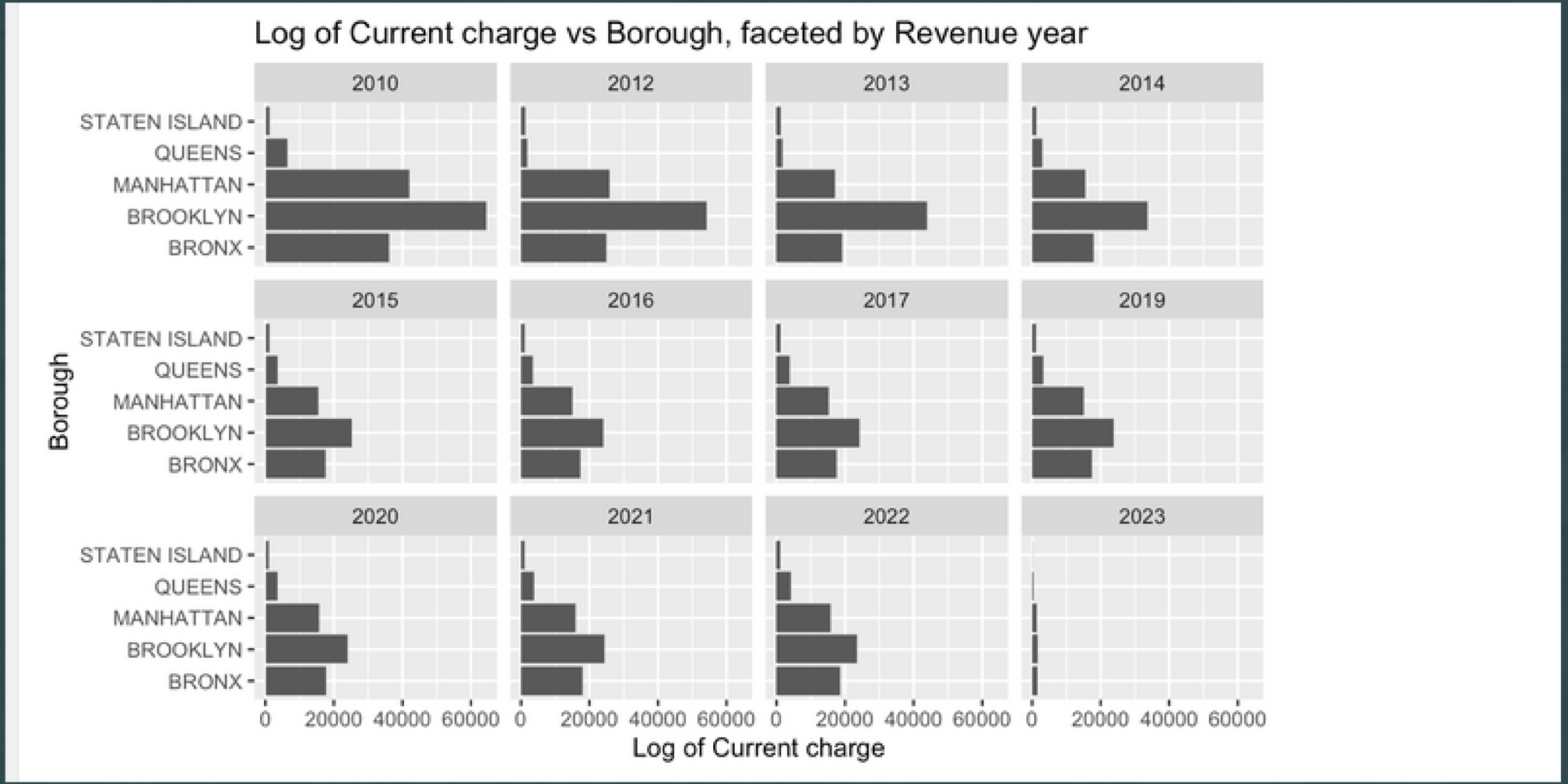
Consumption over Revenue Month by Vendor Name



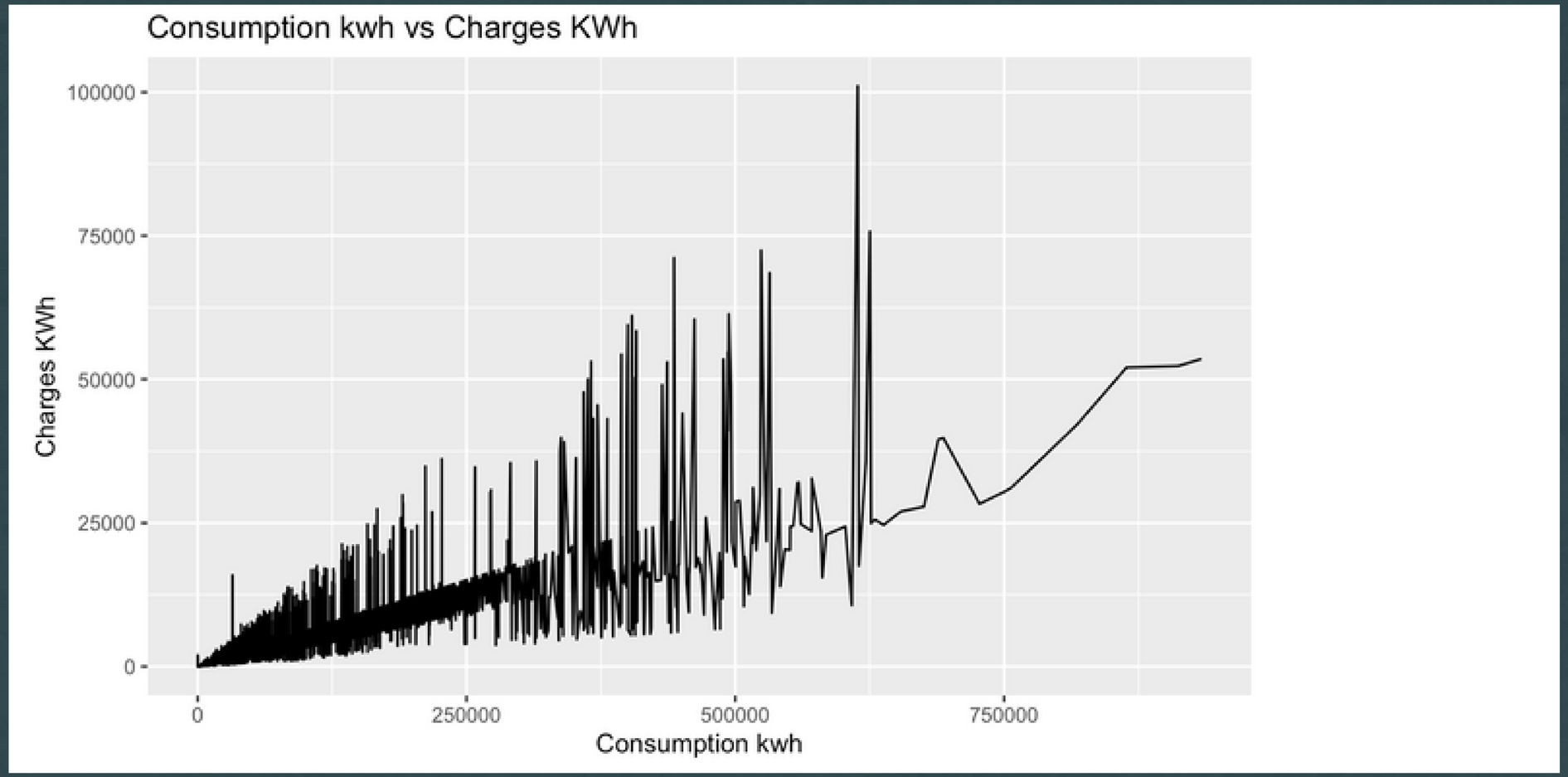
Consumption over Revenue Month by Borough



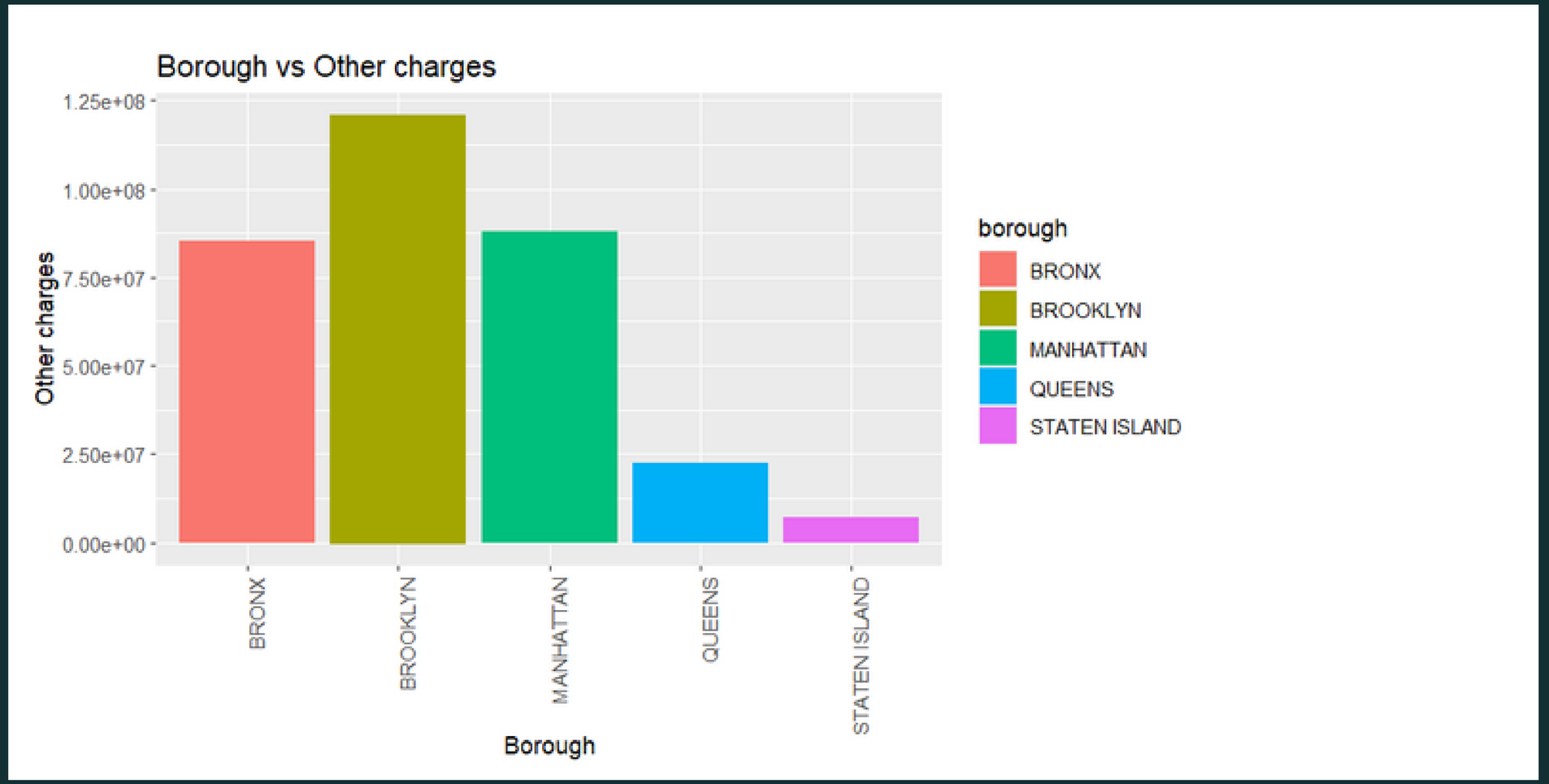
Current charge vs Borough with respect to revenue

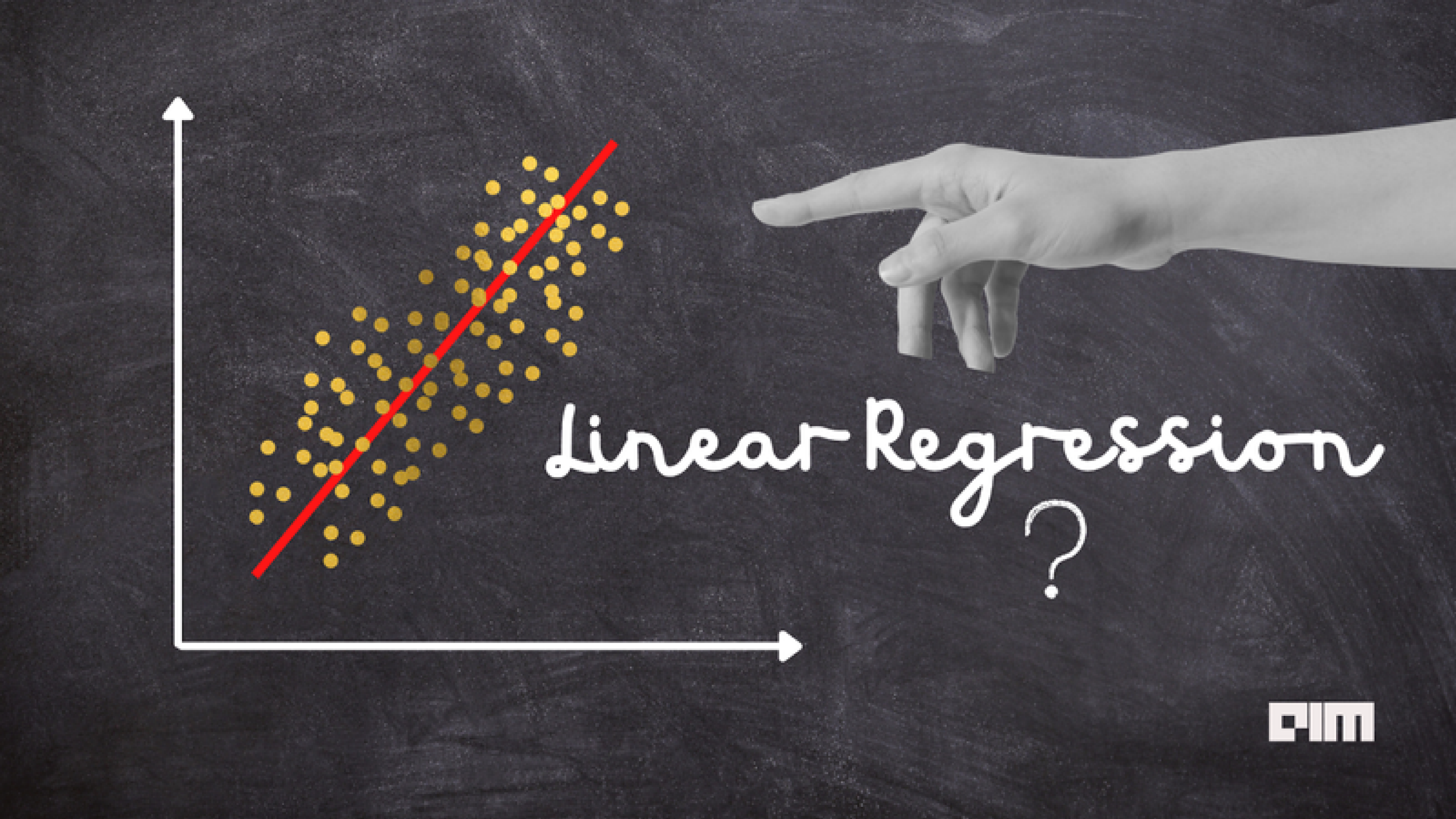


Consumption in KWh compared to Charges KWh



Borough vs other charges




Linear Regression
?

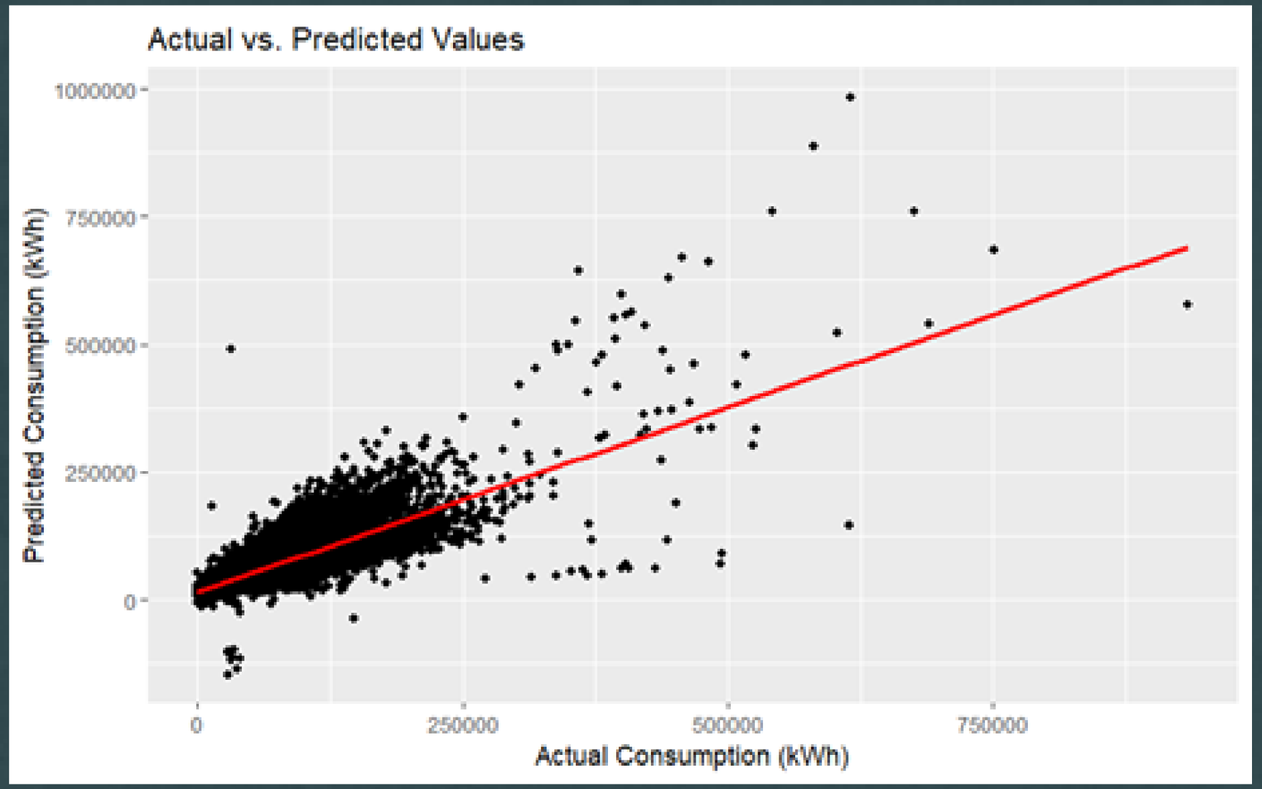
Prediction Model

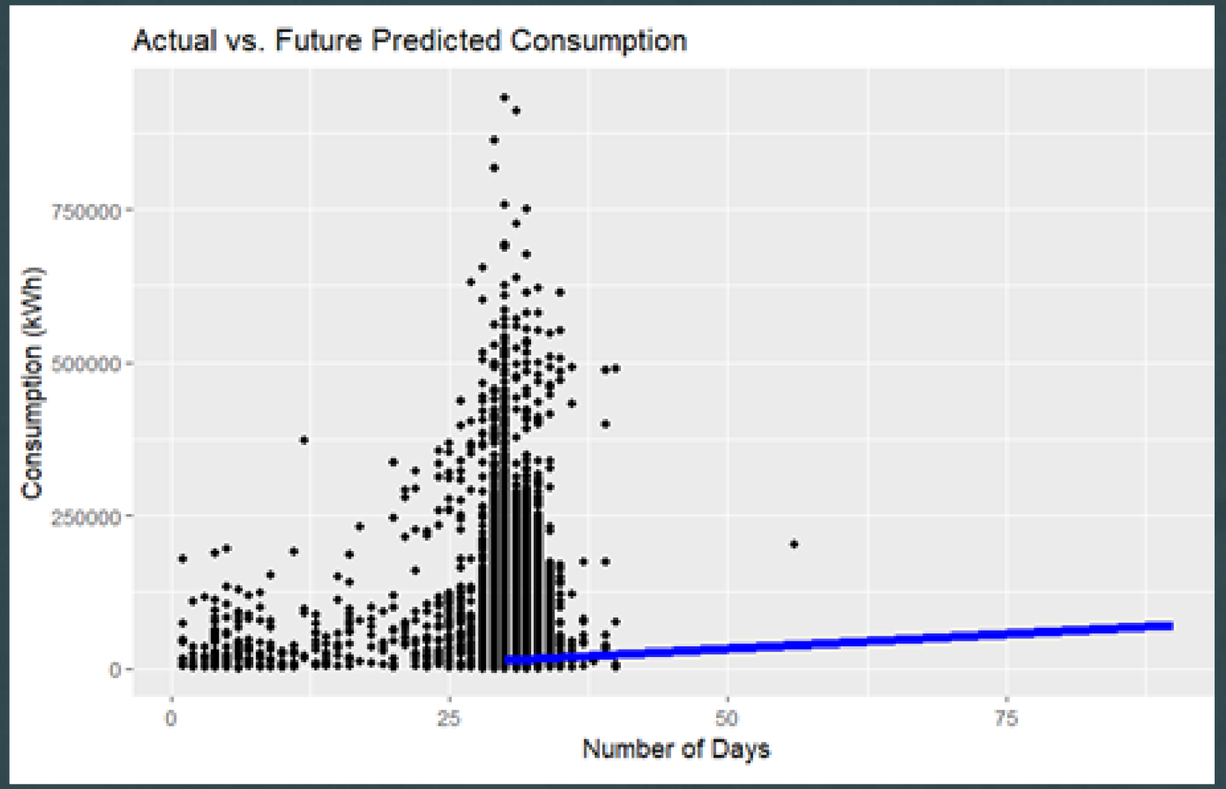
1. Traning and Testing the model.
2. Using Lm function for linear regression.
3. Predicting the model.
4. Calculating the mean squares error and absolute error.
5. Ploting the graph

```r
# Load the data
dataset <- read.csv("C:/Users/py96230n/Desktop/cleaned_datanew.csv")

# Split the data into training and testing sets
set.seed(123) # for reproducibility
train_indices <- sample(nrow(dataset), 0.7 * nrow(dataset))
train_data <- dataset[train_indices, ]
test_data <- dataset[-train_indices, ]

# Fit a linear regression model using the training data
model <- lm(consumption_kwh ~ number_of_days + meter_amr + other_charges, data = train_data)
summary(model)

# Make predictions on the testing data
predictions <- predict(model, newdata = test_data)

# Calculate the root mean squared error (RMSE) of the predictions
rmse <- sqrt(mean((test_data$consumption_kwh - predictions)^2))

# Calculate the mean absolute error (MAE) of the predictions
mae <- mean(abs(test_data$consumption_kwh - predictions))

# Print the MSE, RMSE, MAE, and R-squared value of the model
cat("MSE:", mean((test_data$consumption_kwh  - predictions)^2), "\n")
cat("RMSE:", rmse, "\n")
cat("MAE:", mae, "\n")
cat("R-squared:", summary(model)$r.squared, "\n")

# Plot the actual and predicted values
ggplot(test_data, aes(x = consumption_kwh, y = predictions)) +
  geom_point() +
  #geom_line() +
  ggtitle("Actual vs. Predicted Values") +
  xlab("Actual Consumption (kWh)") +
  ylab("Predicted Consumption (kWh)")+
  geom_smooth(method = "lm", color = "red", se = FALSE)
```

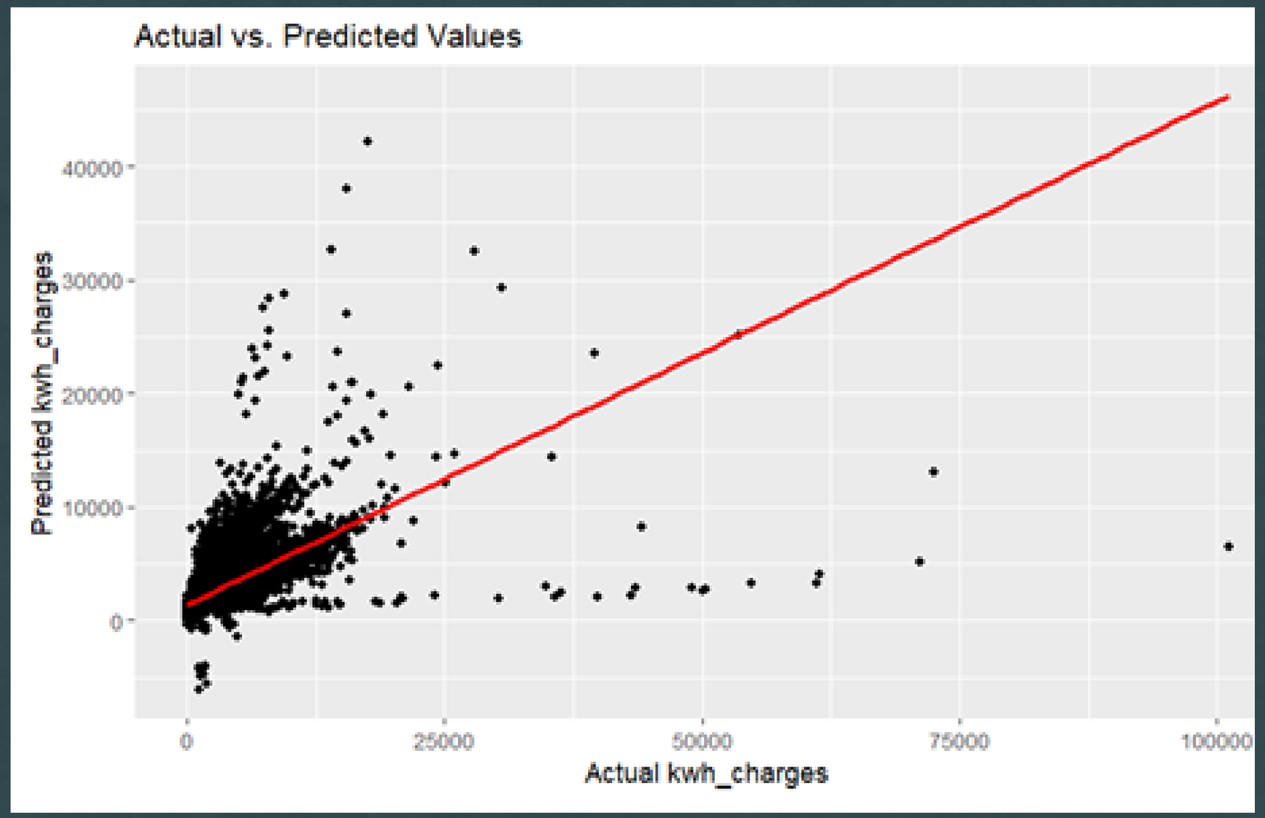
Chunk 3

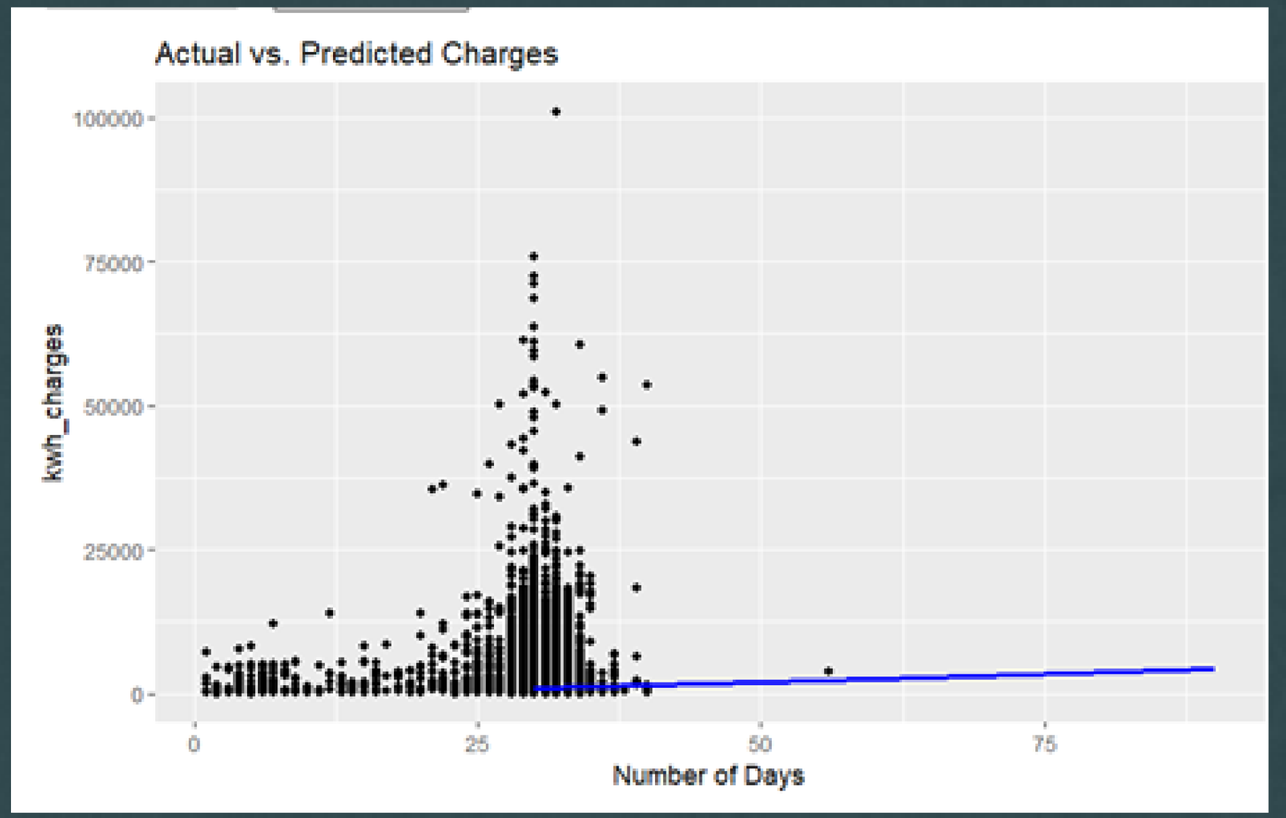
Prediction for Consumption per kWh wrt number of days





Prediction for charges per kWh wrt number of days





RESULT

After analyzing the Electric Consumption and Cost data for New York City from 2010 to February 2023, several conclusions can be drawn:

1. The total electric consumption and cost have increased over time, with some fluctuations.

2. The residential sector consumes the highest amount of electricity, followed by the commercial and industrial sectors.

3. Manhattan has the highest electric consumption and cost among all the boroughs, followed by Brooklyn, Queens, the Bronx, and Staten Island.

4. The winter months have the highest electric consumption and cost, while the summer months have the lowest.

5. There is a positive correlation between electric consumption and cost, which is to be expected.

6. The implementation of energy efficiency measures and renewable energy sources can help reduce electric consumption and cost for the city.

Overall, this data analysis highlights the need for continued efforts to reduce energy consumption and cost in New York City, particularly in the residential sector. The data can be used to inform policies and initiatives aimed at promoting energy efficiency and sustainability in the city.

THANK YOU