Electricity Consumption And Cost (2010-2023)

NYC OpenData

Ayman Shahin
CS301-101

Introduction

Overview of Dataset

- Records electricity consumption and associated cost for a specific billing instance
- Captures monthly consumption of electricity by development, borough, and vendor from 2010-2023

Westchester
Passaic
Bergen
Long Island Sound
BRONX
NEW YORK
Hudson River
MANHATTAN
East River
Essex
Nassau
La Guardia Airport
Hudson
NEW JERSEY
Newark International Airport
QUEENS
Union
Upper Bay
John F. Kennedy International Airport
BROOKLYN
Jamaica Bay
STATEN ISLAND
Middle-esex
Lower Bay
ATLANTIC
OCEAN
Raritan Bay

Objective

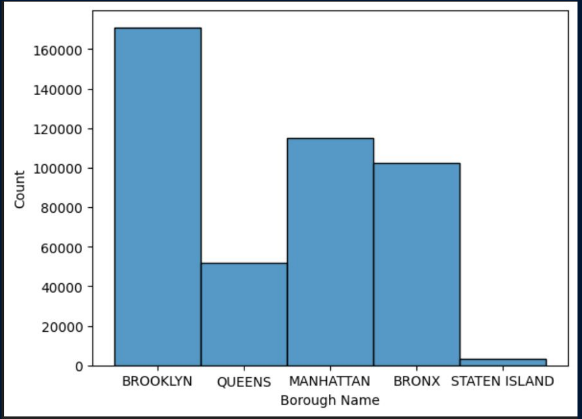
- Examine the relationship between electricity consumption and cost across different boroughs

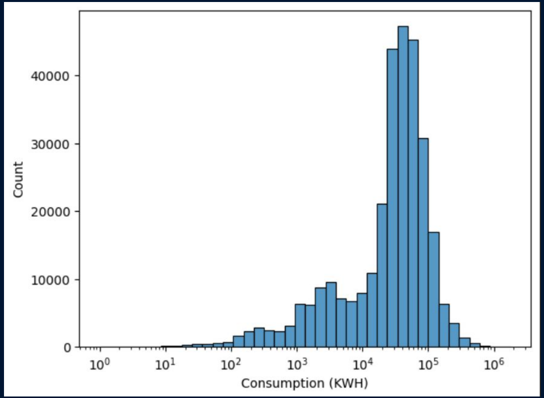
Dataset Description

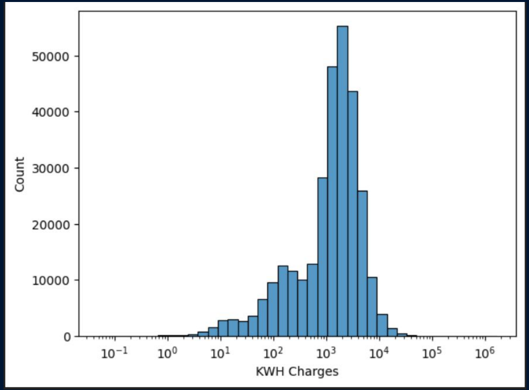
Features

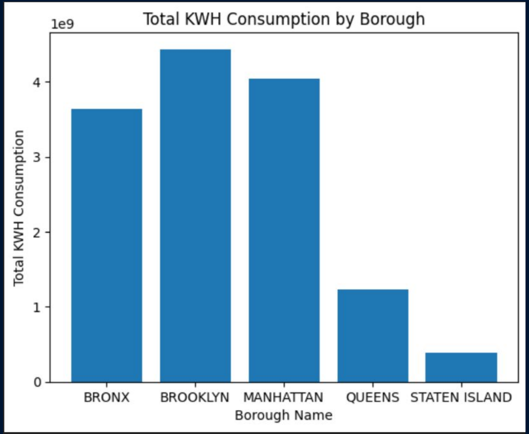
- Development Name: Name of housing development
- Borough: Name of borough (Categorical Variable)
- Account Name: Name of account paying for the service
- Location: Address of housing development
- Meter AMR: Type of automatic reading meter
- Meter Scope: Name of building the meter supplies power to
- TDS #: Tenant Data System
- EDP: Unique identifier for each housing development
- RC Code: Budget responsibility code
- Funding Source: Source of funding for each housing development
- AMP #: Tracking number
- Vendor Name: Name of vendor which supplies power
- UMIS BILL ID: Unique billing identifier
- Revenue Month: Year and month of recorded bill
- Service Start Date: Service start date
- Service End Date: Service end date
- # days: Number of days billed
- Meter Number: Identifier for the meter
- Estimated: Whether or not the meter was read for the time period
- Current Charges: Total cost of consumption for a billing instance
- Rate Class: Type of rate applied to a given account
- Bill Analyzed: Whether or not the bill was analyzed again for billing errors
- Consumption (KWH): Total consumption in kilowatt-hours (Numerical Variable)
- KWH Charges: Kilowatt-hours charges for a particular billing instance (Numerical Variable)
- Consumption (KW): Total consumption in kilowatts
- KW Charges: Kilowatt charges for a particular billing instance
- Other Charges: Other charges charged to the account

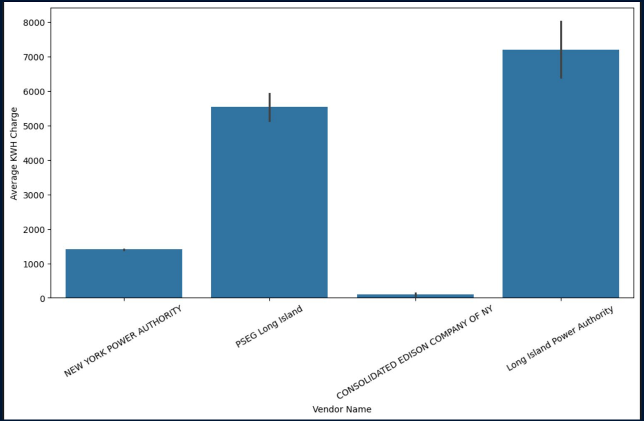
Exploratory Data Analysis



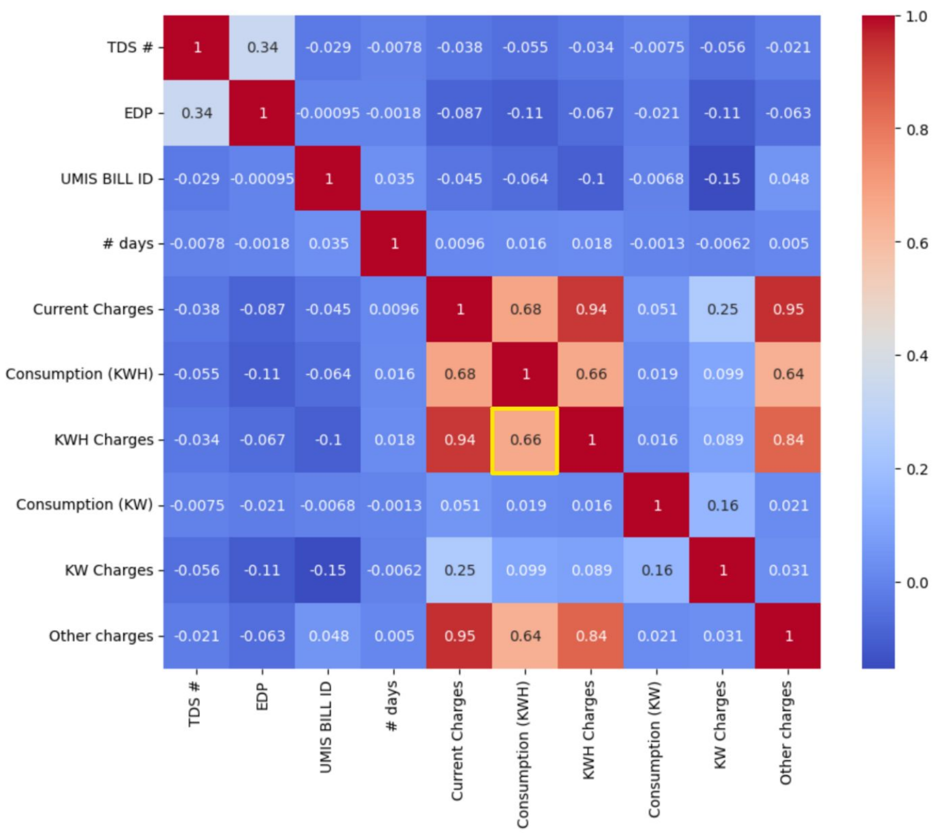




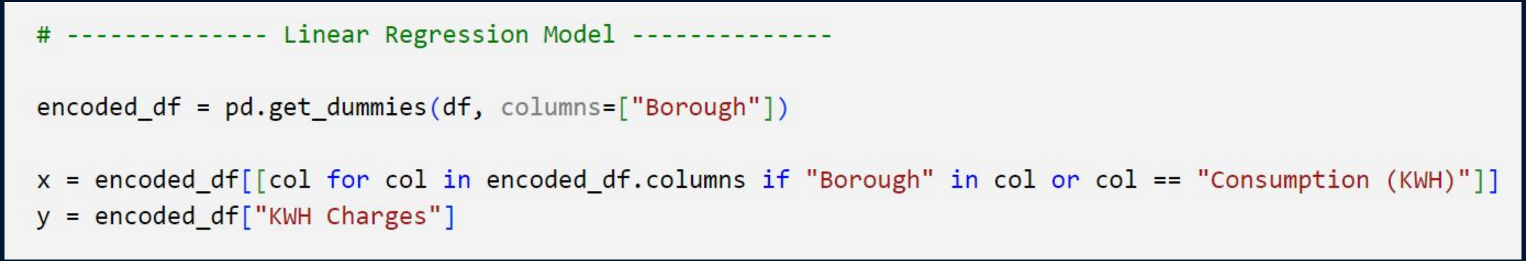




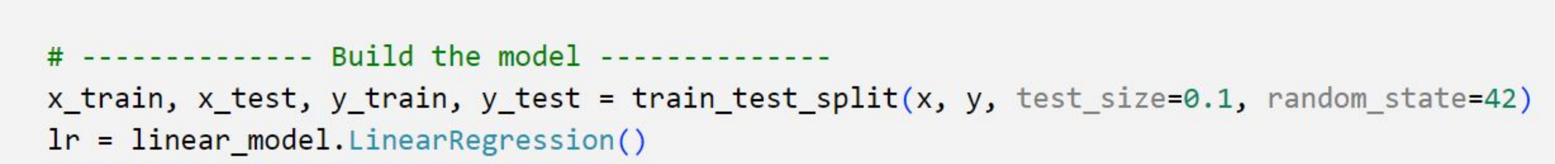
Exploratory Data Analysis



Linear Regression

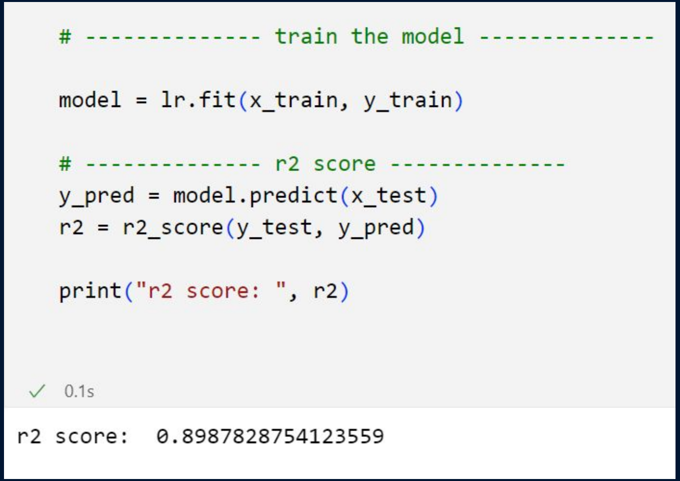


```
# ------------- Linear Regression Model -------------

encoded_df = pd.get_dummies(df, columns=["Borough"])

x = encoded_df[[col for col in encoded_df.columns if "Borough" in col or col == "Consumption (KWH)"]]
y = encoded_df["KWH Charges"]
```



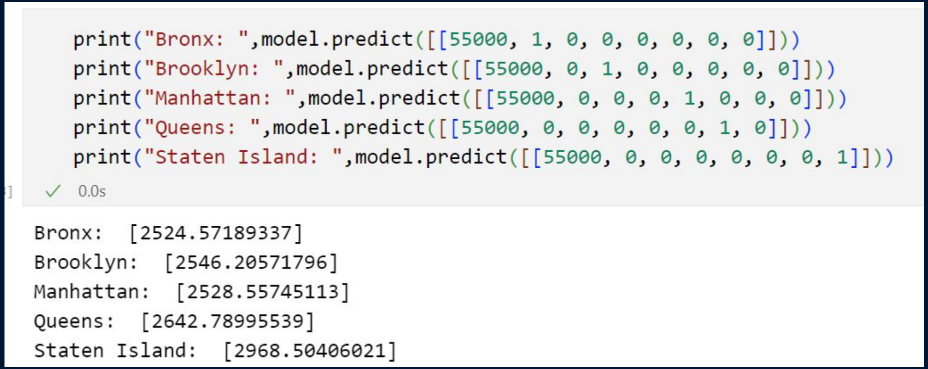
```
# -------------- Build the model --------------
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.1, random_state=42)
lr = linear_model.LinearRegression()
```

Key Findings



```
# -------------- train the model --------------

model = lr.fit(x_train, y_train)

# -------------- r2 score --------------
y_pred = model.predict(x_test)
r2 = r2_score(y_test, y_pred)

print("r2 score: ", r2)
```

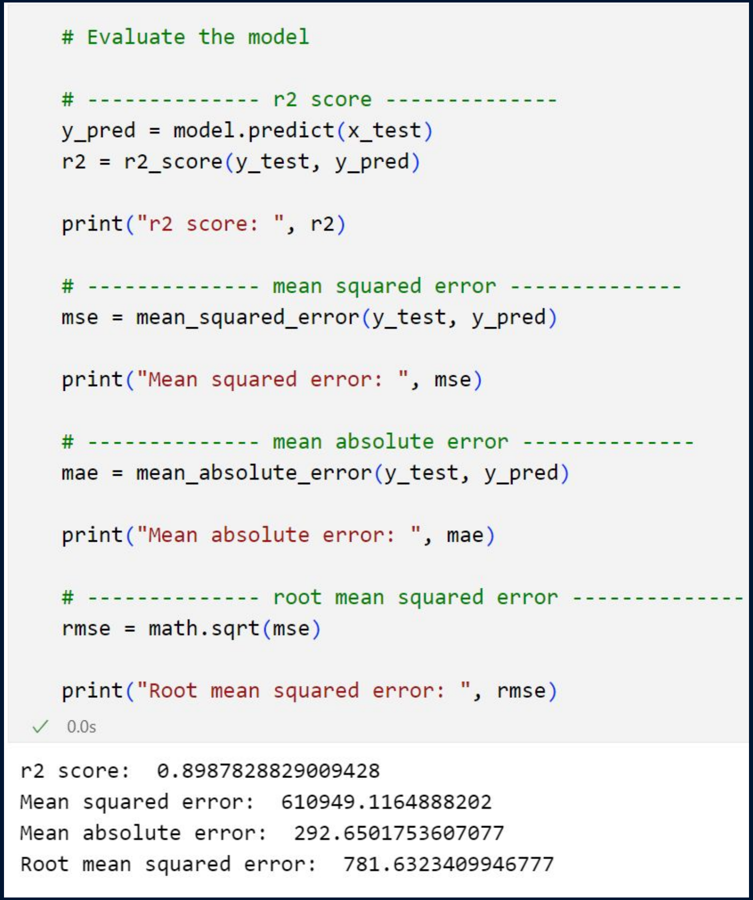
```
r2 score:  0.8987828754123559
```



```
print("Bronx: ",model.predict([[55000, 1, 0, 0, 0, 0, 0, 0]]))
print("Brooklyn: ",model.predict([[55000, 0, 1, 0, 0, 0, 0, 0]]))
print("Manhattan: ",model.predict([[55000, 0, 0, 0, 1, 0, 0, 0]]))
print("Queens: ",model.predict([[55000, 0, 0, 0, 0, 0, 1, 0]]))
print("Staten Island: ",model.predict([[55000, 0, 0, 0, 0, 0, 0, 1]]))
```

```
Bronx:  [2524.57189337]
Brooklyn:  [2546.20571796]
Manhattan:  [2528.55745113]
Queens:  [2642.78995539]
Staten Island:  [2968.50406021]
```

Evaluation



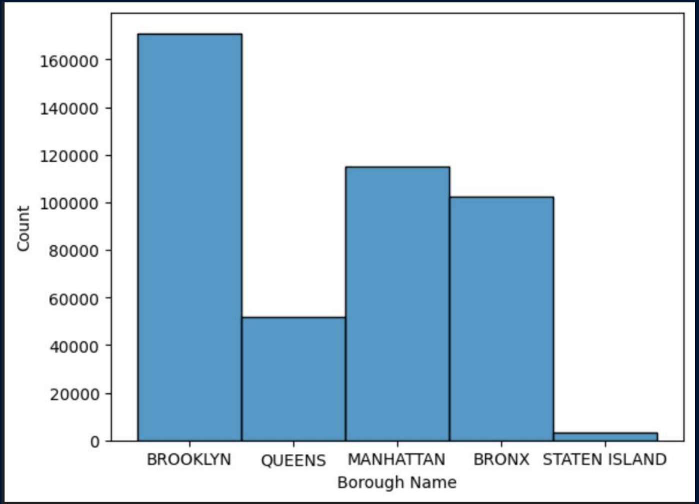
```
# Evaluate the model

# -------------- r2 score --------------
y_pred = model.predict(x_test)
r2 = r2_score(y_test, y_pred)

print("r2 score: ", r2)

# -------------- mean squared error --------------
mse = mean_squared_error(y_test, y_pred)

print("Mean squared error: ", mse)

# -------------- mean absolute error --------------
mae = mean_absolute_error(y_test, y_pred)

print("Mean absolute error: ", mae)

# -------------- root mean squared error -------------
rmse = math.sqrt(mse)

print("Root mean squared error: ", rmse)
```

✓ 0.0s

```
r2 score:  0.8987828829009428
Mean squared error:  610949.1164888202
Mean absolute error:  292.6501753607077
Root mean squared error:  781.6323409946777
```

Flaws

- Small number of observations for Staten Island



References

[1] https://data.cityofnewyork.us/Housing-Development/Electric-Consumption-And-Cost-2010-Feb-2023-/jr24-e7cr