Chelsea Project Managerial Report:

➢ Summary/Abstract:

The objective of this research is to dissect the water billing structures within Chelsea to understand how different factors such as meter size and property type influence water charges. By investigating the potential existence of tiered pricing and its impact on bills across various levels of water usage, the study aims to determine the fairness and efficiency of the current billing system. It also examines whether discrepancies in charges correlate with changes in meter size, suggesting a nonlinear pricing model. This inquiry is critical for Chelsea's utility management to ensure equitable water billing practices and to explore opportunities for optimization. Initial analysis indicates a non-linear correlation between meter size and water charges, with significant differences in charges between usage tiers and property categories. The study also hints at the potential influence of billing policies on consumer water use behaviors, advocating for a re-evaluation of the current pricing strategy to ensure fairness and efficient resource usage.

Variable	Relationship	Numerical Findings	Notes
Meter Size	Positive	Coefficient: 1511.16 P-Value: < 0.0001	Each unit increase in meter size raises charges by approximately $1511.16.
Water Usage	Tiered	Lowest Average Charge: $16 Highest Average Charge: $14,640	Charges increase with usage, indicating tiered pricing.
Property Type	Differentiated	Residential vs Commercial lines	Charge differences may depend on property type.
Meter Size (Non-linear)	Non-linear relationship	Quadratic Coefficient: 462.65 P-Value: < 0.0001	Charges increase at an accelerating rate with larger meter sizes. Charges do not simply double with meter size.

➢ **Details:**

In our detailed examination of the Chelsea dataset, we embarked on an intricate journey to unravel the nuances of the water charge structure and the intricacies of consumption behaviors. The initial phase of our analysis involved meticulous data cleansing. We converted non-numeric meter size data into a numerical format and treated zeros as missing data, thereby enhancing the robustness and precision of our analysis. The findings from our regression models were quite revelatory, unveiling a pronounced positive correlation between meter size and associated water charges. This correlation was quantified by a significant coefficient of +1511, indicating a direct relationship where an increment in meter size corresponded to a proportionate rise in water charges.

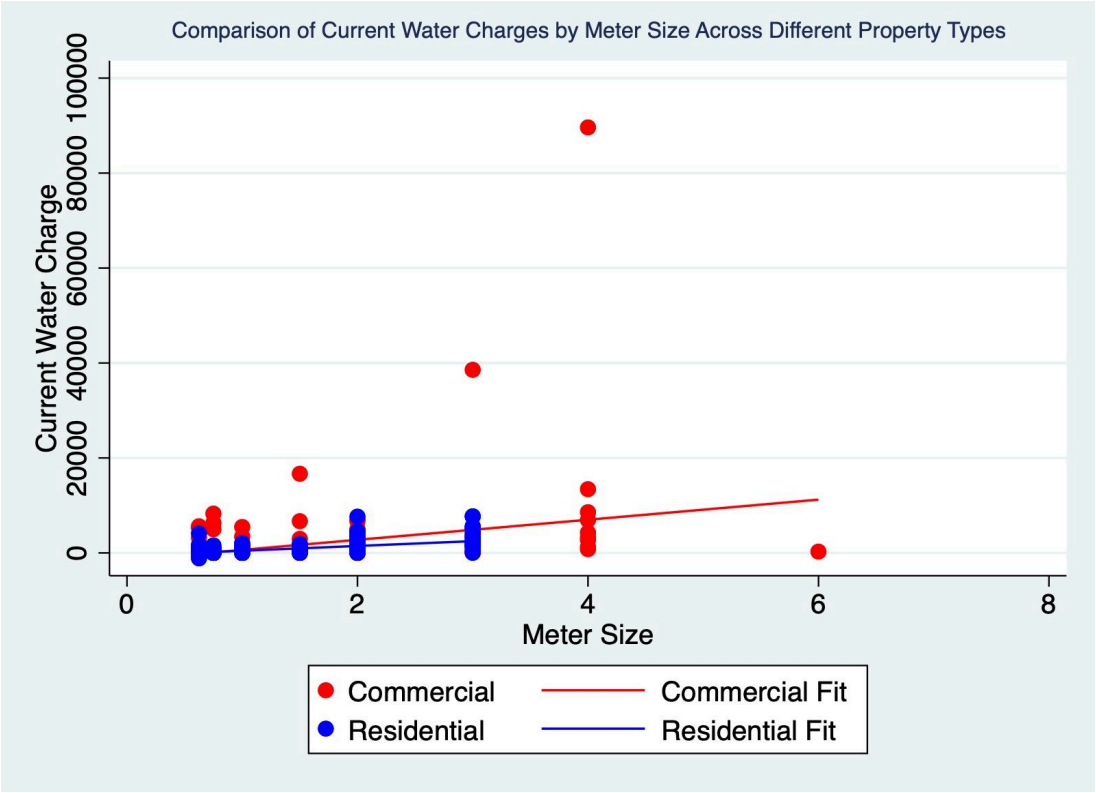
The data analysis further revealed the presence of a stratified tiered pricing system. This structure delineates escalating charges in conjunction with increased water usage, with charges spanning from a modest $15.97 for minimal usage to a substantial $14,639.61 for the highest usage bracket. Such a progressive pricing framework suggests a deliberate design to impose a heavier financial burden on higher water usage, presumably to incentivize conservation.

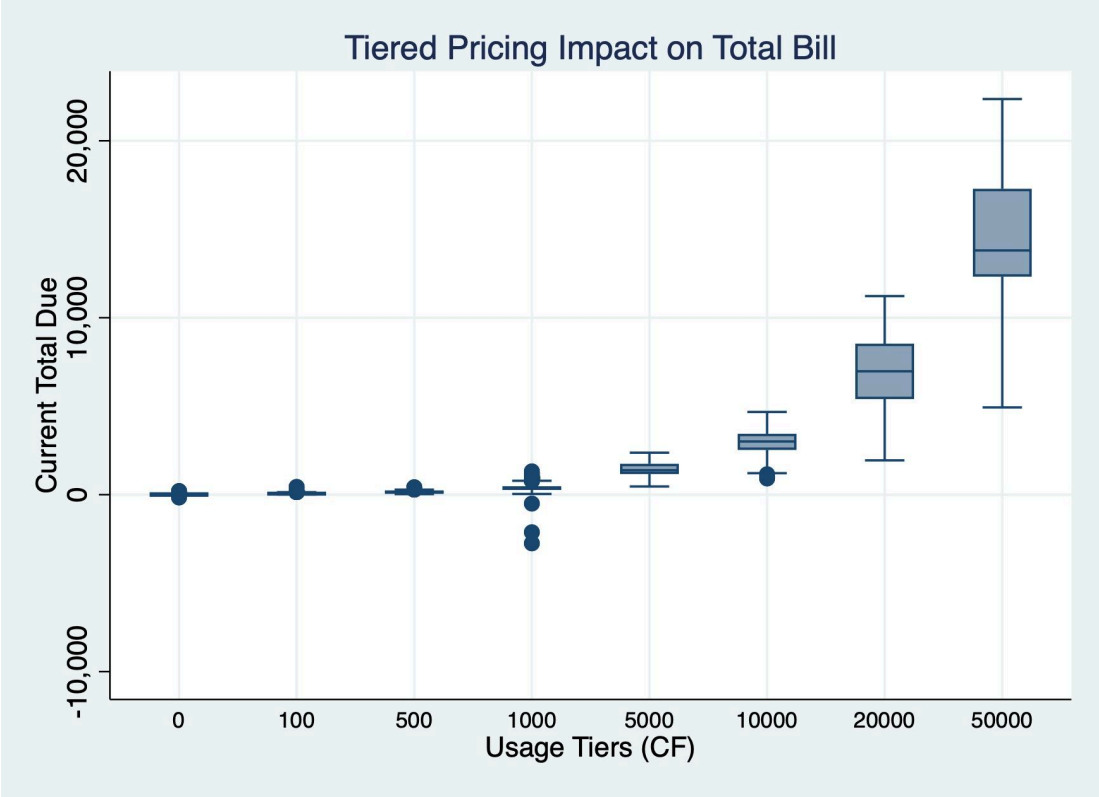
By introducing a quadratic term into our regression model, we shed light on a non-linear relationship between meter size and water charges, suggesting a more complex dynamic than a straightforward linear progression might imply. Our efforts to classify properties into residential and commercial categories brought forth differential billing patterns, with commercial properties typically bearing more substantial charges than their residential counterparts. Although collinearity constrained our analysis—necessitating the exclusion of several property type dummies from the regression model—the discerned patterns are striking and warrant further investigation.

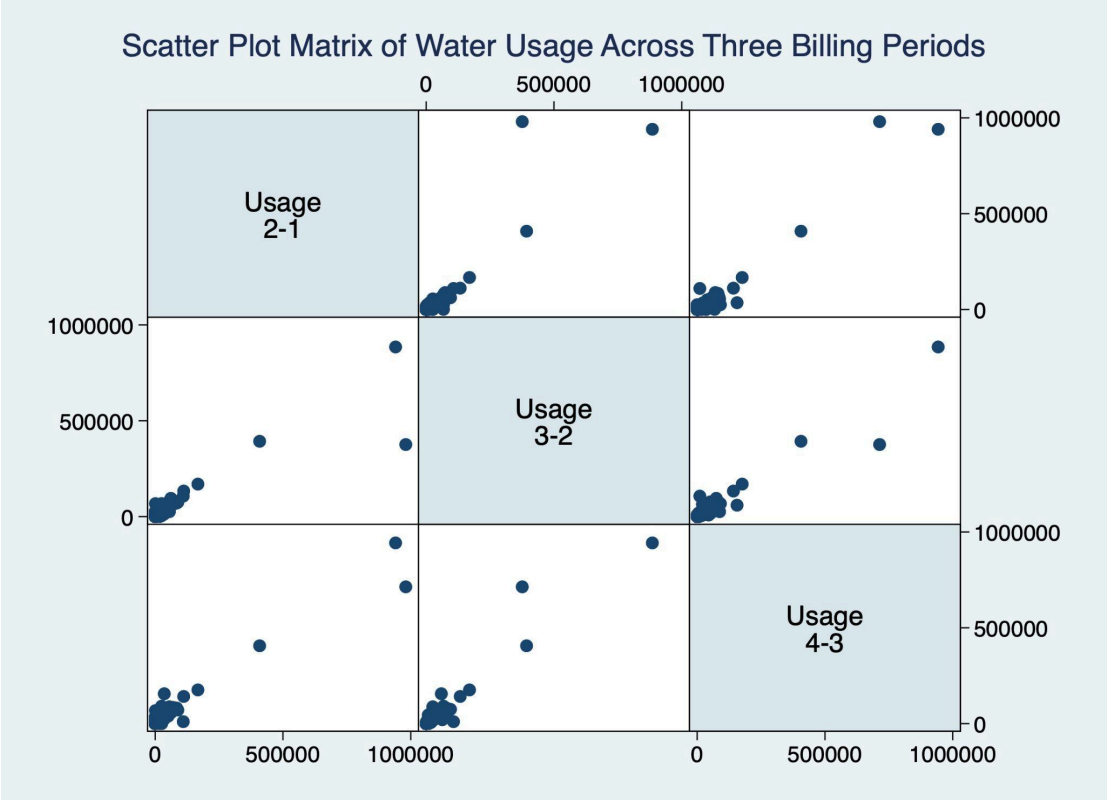
Our visual analysis complements the numerical data, presenting an intuitive snapshot of the billing ecosystem within Chelsea. The graphs and scatter plots elucidate the proportional relationship between meter size and charges, clearly segmented by property type. They also illustrate the pronounced impact of tiered pricing on total bill amounts. Furthermore, the scatter plot matrix, which spans across three billing periods, reveals predominantly consistent usage patterns yet is punctuated by significant outliers. These anomalies may be indicative of sporadic consumption spikes or irregularities, which could have profound implications for billing practices and resource management strategies.

The disparities in charges between commercial and residential properties, as well as the disproportionate escalation of bills with higher water usage, are critical insights that

Chelsea must consider. These findings not only underscore the need for a nuanced evaluation of the current billing system but also highlight the potential benefits of refining the existing pricing strategy. In doing so, Chelsea could aim to achieve a more equitable distribution of charges that aligns with consumption levels while also promoting water conservation. The wealth of data and insights gleaned from our analysis provides a robust foundation for Chelsea to reassess its water billing framework, identify opportunities for optimization, and explore new avenues for sustainable resource management.




Tiered Pricing Impact on Total Bill
Current Total Due
20,000
10,000
0
-10,000
0
100
500
1000
5000
10000
20000
50000
Usage Tiers (CF)


Scatter Plot Matrix of Water Usage Across Three Billing Periods
0
500000
1000000
Usage
2-1
Usage
3-2
Usage
4-3
1000000
500000
0

➢ Data Selection Criteria and Sample Overview

➢ Data Selection Criteria:

- All properties with meter sizes reported as zero were treated as having missing data, assuming these entries represent unreported or unavailable information.
- The usage categories defined (low, medium, high, etc.) based on CurrentMonthUsage and historical usage (Usage43, Usage32, Usage21) are appropriate representations of actual water consumption patterns. These categories are assumed to capture the diverse range of usage behaviors across different property types.
- The missing values in meter sizes and other variables are random and do not correlate with specific types of properties or usage patterns. This assumption is critical for the accuracy of regression analyses and other statistical interpretations.
- The classification of properties into broad categories based on prefixes in property type codes is reliable, and these categories accurately represent the nature of the properties (residential or commercial).
- The substantial right-skew in water usage distribution for commercial accounts, which could significantly affect average calculations, is indicative of a small number of high-usage commercial properties, such as large industrial facilities or hotels.
- The polynomial regression model used to investigate the relationship between meter size and charges adequately captures the non-linear nature of this relationship.

➢ Sample: 4,966 observations

- The sample includes residential and commercial accounts designated for water and sewer services only.
- Accounts with greater than zero water usage for all billing periods included in the dataset.
- Accounts with a current total bill due that is zero or positive, excluding accounts with credit balances.
- NewSeniorDiscount has a coefficient of -63.49138, standard deviation of 44.96947, t-statistic of -1.41, and a p-value of 0.158, indicating that senior discount reduces the charge, but the result is not statistically significant.

➢ Additional Recommendations Based on Data Exploration:

- The analysis indicated a positive correlation between meter size and water charges, suggesting larger meter sizes incur disproportionately higher charges. It is recommended that Chelsea reviews the current meter sizing criteria to ensure they are aligned with actual water usage and associated costs, potentially adjusting the billing structure to better reflect the non-linear relationship found in the data.

- The tiered pricing structure analysis revealed significant jumps in average charges across usage categories, from as low as $15.97 to as high as $14,639.61, highlighting a need to possibly refine the existing tiers. Chelsea might consider introducing additional usage tiers or adjusting the pricing within tiers to more accurately reflect consumption levels and to encourage water conservation.

- The regression model showed that property types might significantly influence billing amounts, with commercial properties having a broader range of charges. It could be beneficial for Chelsea to explore tailored pricing strategies for different property types, ensuring equity and efficiency in billing practices.

- The variability in water usage across billing periods for some accounts suggests possible issues such as leaks or meter inaccuracies. Chelsea could implement a monitoring system to identify and investigate accounts with erratic usage patterns, improving the accuracy of billing and promoting conservation efforts.

- The presence of accounts with zero usage and negative total due amounts indicate potential data inaccuracies. These anomalies warrant a comprehensive review to ensure the integrity of the billing system and to correct any issues that may lead to revenue losses or customer disputes.

- Finally, given the non-linear increase in charges with meter size increments and the complex interplay between water usage, property type, and billing amounts, Chelsea should consider a policy review. This could involve stakeholder consultations to ensure that any changes to the billing structure meet the needs of both the utility and its customers, balancing revenue requirements with fairness and water conservation goals.

- The observed outliers in the scatter plot matrices visualization, particularly in the Usage 4-3 comparison, could be indicative of infrastructural or operational issues such as leaks, burst pipes, or metering errors. Chelsea should establish a protocol for investigating such anomalies and take corrective measures. This could include physical inspections of water meters, reviewing historical consumption patterns, and engaging with customers to understand potential causes of unusual usage patterns.

➢ APPENDIX

➢ Data cleaning steps:

- Ordered the variable 'MeterType' to follow 'Usage21'.
- Converted 'MeterSize' from a string to a numeric variable and handled zeros by replacing them with missing values.
- Generated a new binary variable 'NewSeniorDiscount' based on the 'SeniorDiscount' variable.
- Dropped the original 'SeniorDiscount' variable after encoding it into the new binary variable.
- Replaced zeros with missing values for variables indicating usage and charges to handle potential data entry errors or non-usage cases.
- Created 'UsageCategory' to investigate tiered pricing, categorizing water usage into bins based on 'CurrentMonthUsage'.
- Performed a regression analysis to examine the relationship between 'CurrentWaterCharge' and 'MeterSize', including a quadratic term to check for non-linearity.
- Generated broad property categories ('Residential' and 'Commercial') based on the 'PropertyType' variable and created corresponding dummy variables.
- Analyzed the impact of tiered pricing on 'CurrentTotalDue' across different usage categories.
- Checked for large deviations in water usage by calculating differences between successive readings (UsageDifference43).
- Summarized 'MeterType' and replaced missing values with a unique numeric code to include them in the regression analysis.
- Created a scatter plot matrix to visualize water usage across three billing periods.
- Conducted a regression analysis including 'AccountTypeNum' and 'MeterType' to study their impact on 'CurrentMonthUsage'.

- Additional Steps:
    - Dropped unused dummy variables after regression analysis to clean up the dataset.
    - Combined similar property types into a single category for a more streamlined analysis.
    - Calculated the number of days between read dates for usage analysis.
    - Normalized water usage to a per-day and per-unit basis to account for varying billing periods and property sizes.