Smart Metering Project at Nairobi Water

PRESENTED BY: **GROUP 3**

MEET OUR TEAM

Vanessa Sandra- Team Lead

Wamuyu Gitonga

Lillian Wangui

Natalie Wanjiku

Leejay Mwakireti

John Njogu

Elvis oduor

AGENDA

- Introduction
- Data
- Modelling
- Deployment
- Model Performance
- Key Insights & Recommendations
- Conclusion

BUSINESS UNDERSTANDING

The Nairobi City Water and Sewerage Company (NCWSC) plays a crucial role in ensuring the realization of the constitutional right to clean and safe water under Article 43 (1)(d) of the Constitution of Kenya. Its mandate includes:

Provision of Water Services Ensuring residents of Nairobi have access to clean, safe, sufficient, and affordable water. Rehabilitating and maintaining water sources to guarantee sustainable supply.

Sanitation Services Supporting County Government efforts in sanitation management as per the Fourth Schedule of the Constitution.

Addressing Water Scarcity Implementing measures to deal with water shortages, especially in informal settlements where untreated and unsafe water businesses have emerged.

STAKEHOLDERS

Some of the major stakeholders include;
NCWSC Management & Operations Teams – oversee production, distribution, and billing; directly responsible for reducing non-revenue water.

Policy Makers & Regulators (e.g., WASREB, Ministry of Water & Sanitation) – set compliance standards, monitor performance, and allocate resources.

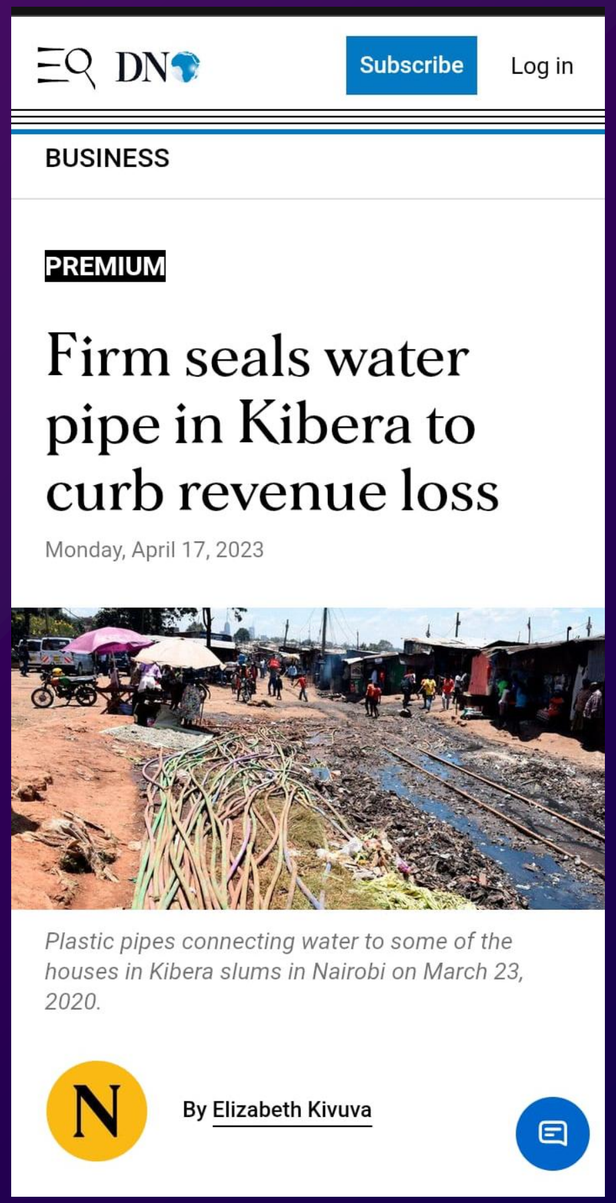
Nairobi City Planners & County Government – align water management with urban development plans.

Customers – end-users who rely on the water supply and are directly affected by service reliability, pricing, and billing accuracy.

Technical Partners/Vendors – providers of smart metering technology and IoT solutions.

Donors/Investors (if applicable) – external organizations funding or supporting water infrastructure projects.

BUSINESS PROBLEM



PROPOSED SOLUTION

To address this, NCWSC is investing in smart metering technology one among the interventions geared towards reduction of NRW. Smart meters offer accurate, real-time consumption data, helping detect leakages, curb illegal connections, and improve billing accuracy. A KES 300 million budget has been allocated for 2,000 smart meters in priority zones, with the expectation of significantly reducing losses and improving efficiency.

CHALLENGE

How NCWSC can strategically identify the zones and customer segments with the greatest water losses, and prioritize smart meter investments in a way that maximizes return on investment while sustainably reducing non-revenue water?

OBJECTIVE

The main goal of this project is to study the gap between the water produced and the water billed by the Nairobi City Water and Sewerage Company (NCWSC). By focusing on areas with the highest losses and exploring the potential of smart meters, the project aims to propose practical ways to reduce non-revenue water (NRW), enhance efficiency, and make the company more financially sustainable.

SPECIFIC OBJECTIVES

1) Quantify Water Losses per region.
2) Cluster customers on water loss.
3) Prioritize High Impact Areas.

DATA UNDERSTANDING

NCWSC Operational Data – records of total water production volumes at treatment plants and distribution points for 10 months

Customer Billing Data – customer-level and billing records reflecting consumption and revenue collection in all zones in Nairobi.

WASREB Tariff Data On Approved Billings – regulatory benchmarks for evaluating NCWSC's efficiency relative to national standards.

The data source

NCWSC

DATA PREPARATION

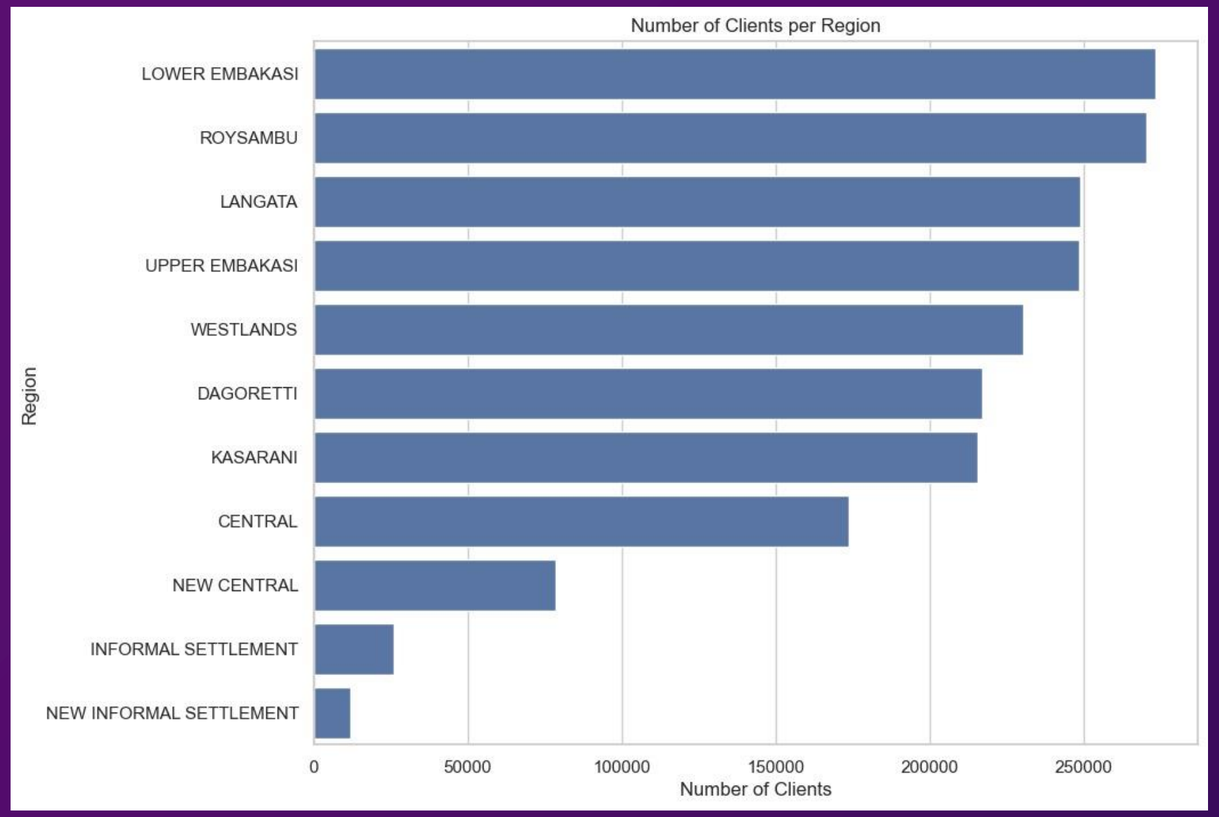
Data preprocessing steps included:

Checking and handling missing values: we employed the following methods to mitigate their effects.

- For columns with greater than 80% of missing values we will drop them.
- For Numeric columns such as the meter size we will fill with the mode.

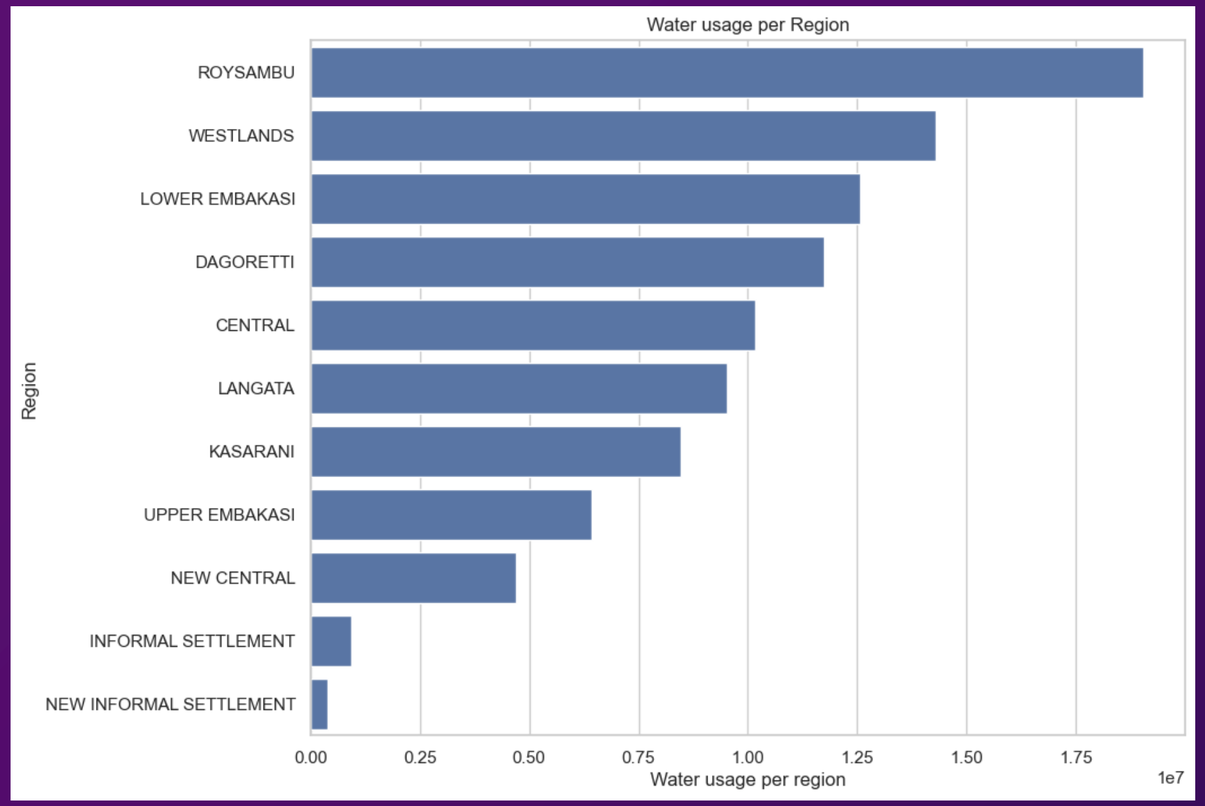
- Removing duplicates
- Filtering outliers
- Feature engineering

EDA



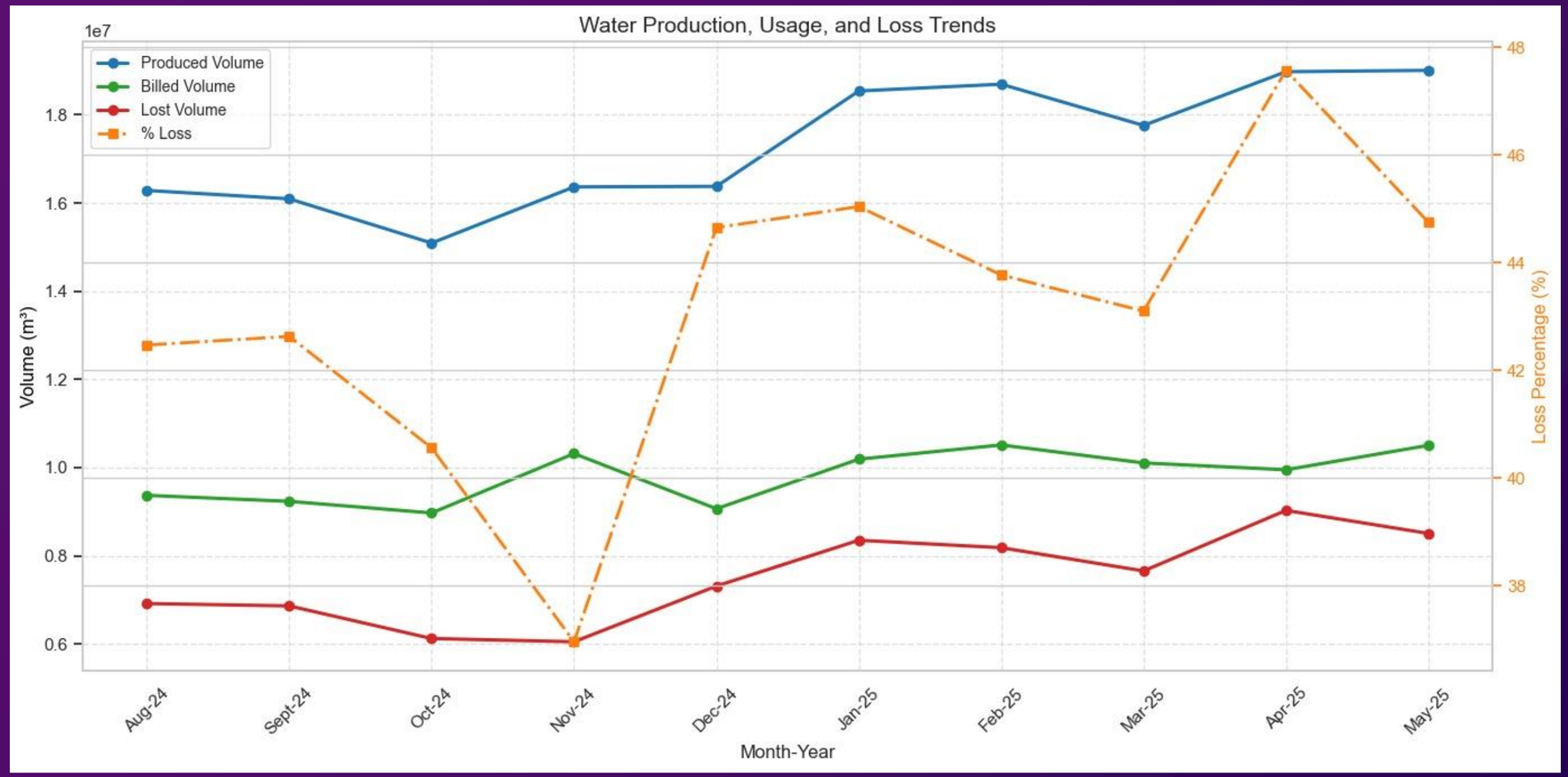
The Lower Embakasi, Roysambu, and Langata regions account for the largest share of the customer base, indicating that these areas have the highest number of connections or billed installations. In contrast, informal settlements represent the smallest share of customers, reflecting fewer registered connections or metered accounts.

EDA



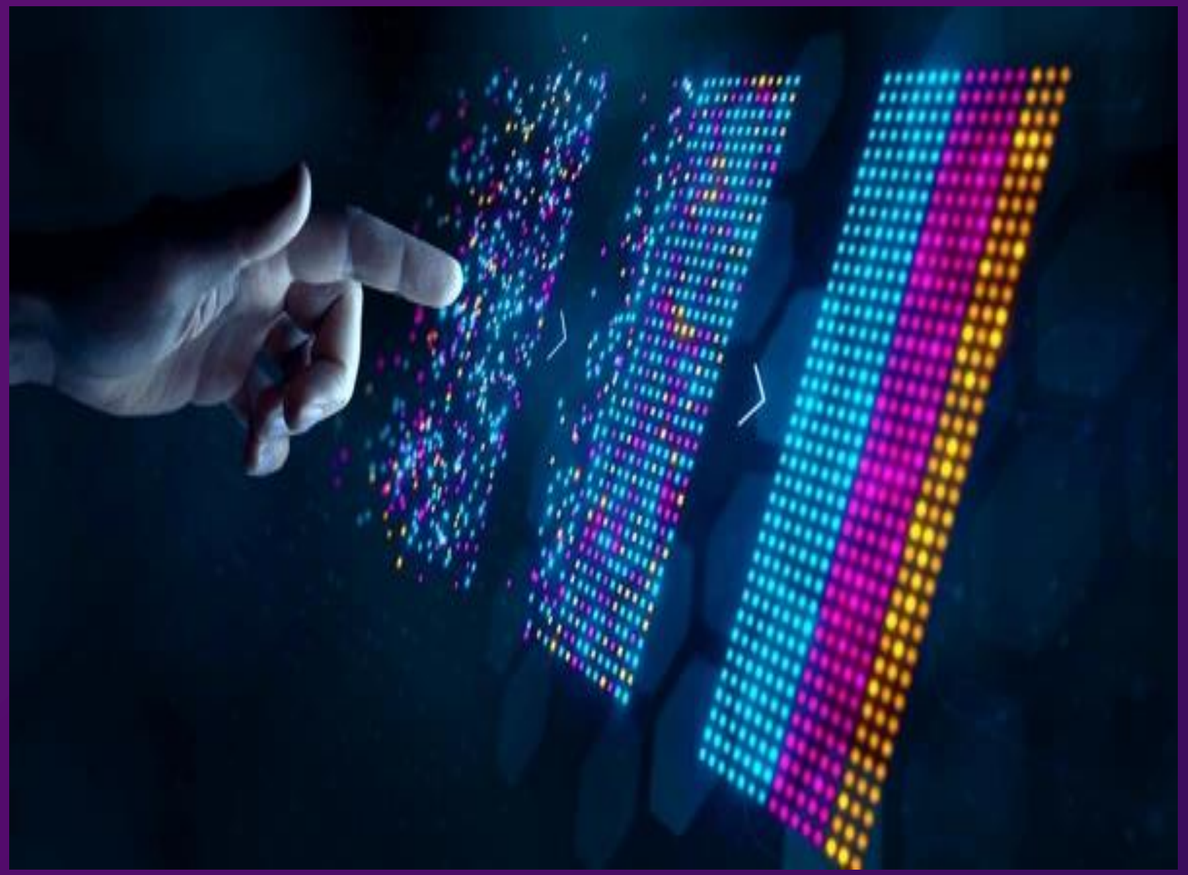
The distribution of billed water volumes across regions highlights significant disparities in consumption and demand. Roysambu recorded the highest billed volume at over 19 million units, standing out as the single largest contributor. Westlands followed with nearly 14.3 million units, while Lower Embakasi and Dagoretti also reported substantial billed volumes, exceeding 12.5 million and 11.7 million units respectively.

EDA



The share of water lost fluctuated between about 37 percent and 48 percent during the period under review. November 2024 recorded the lowest loss rate at just under 37 percent, while April 2025 stood out as the month with the highest losses, at nearly 48 percent.

CLUSTERING



1) Clustering is an unsupervised machine learning method that automatically groups similar data points without predefined labels. This project seeks to identify hotspots of water loss and support the strategic rollout of smart metering systems. Clustering provides a powerful approach to this challenge.

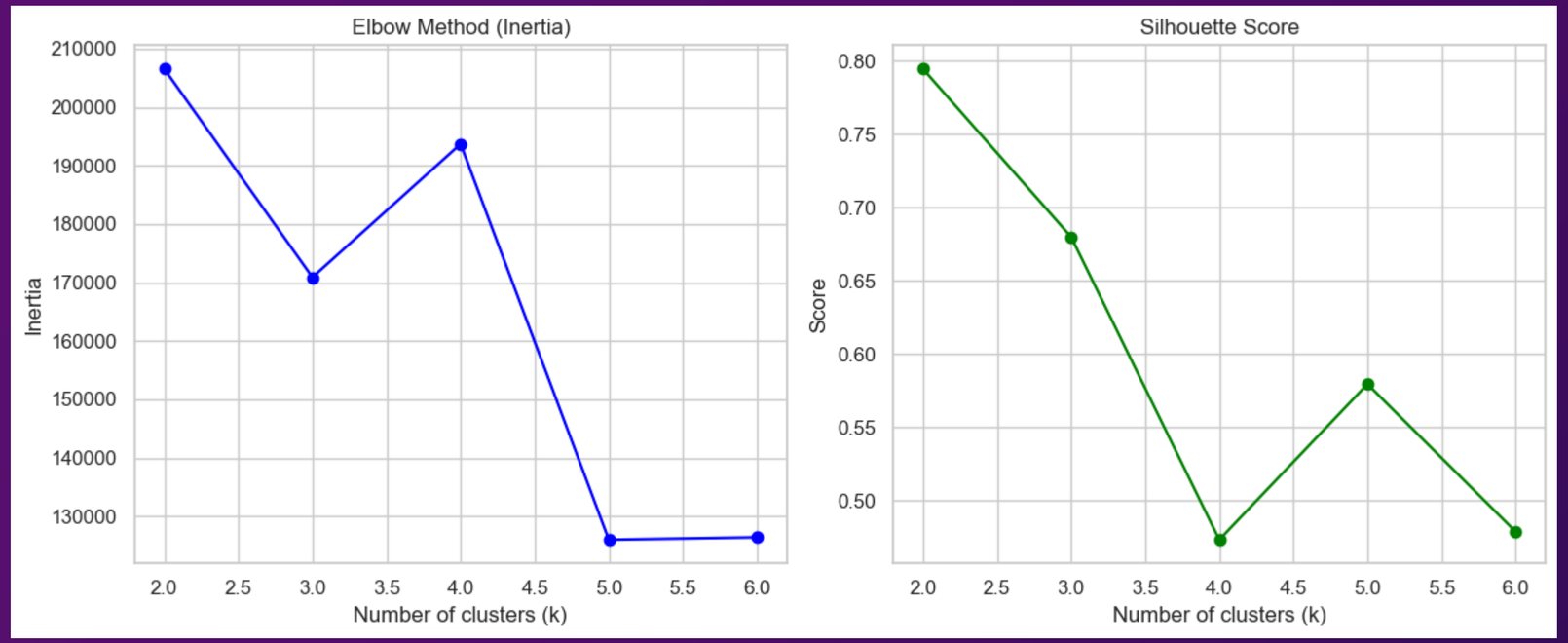
2) By grouping high consumers into three distinct clusters, we can prioritize intervention:

3) Highest consumers – the top priority for smart meter rollout, as losses here have the greatest financial impact.

4) Higher consumers – secondary priority, important for scaling future rollouts.

5) High consumers – still significant but not immediate rollout candidates.

CLUSTERING



1) The elbow method points to k=4 or 5 whereas the Silhouette score strongly favors k=2 (best), but also shows k=4 as a decent balance. Since we are looking for a more detailed segmentation for business insights, we are choosing k=4 since it balances inertia reduction with decent silhouette.

2) Upto this point we have 3 potential optimal n_clusters for our model, to make a good decision, we are employing the use of GridSearchCV to make the best decision that will not only make our model better but also make our recommendations accurate.

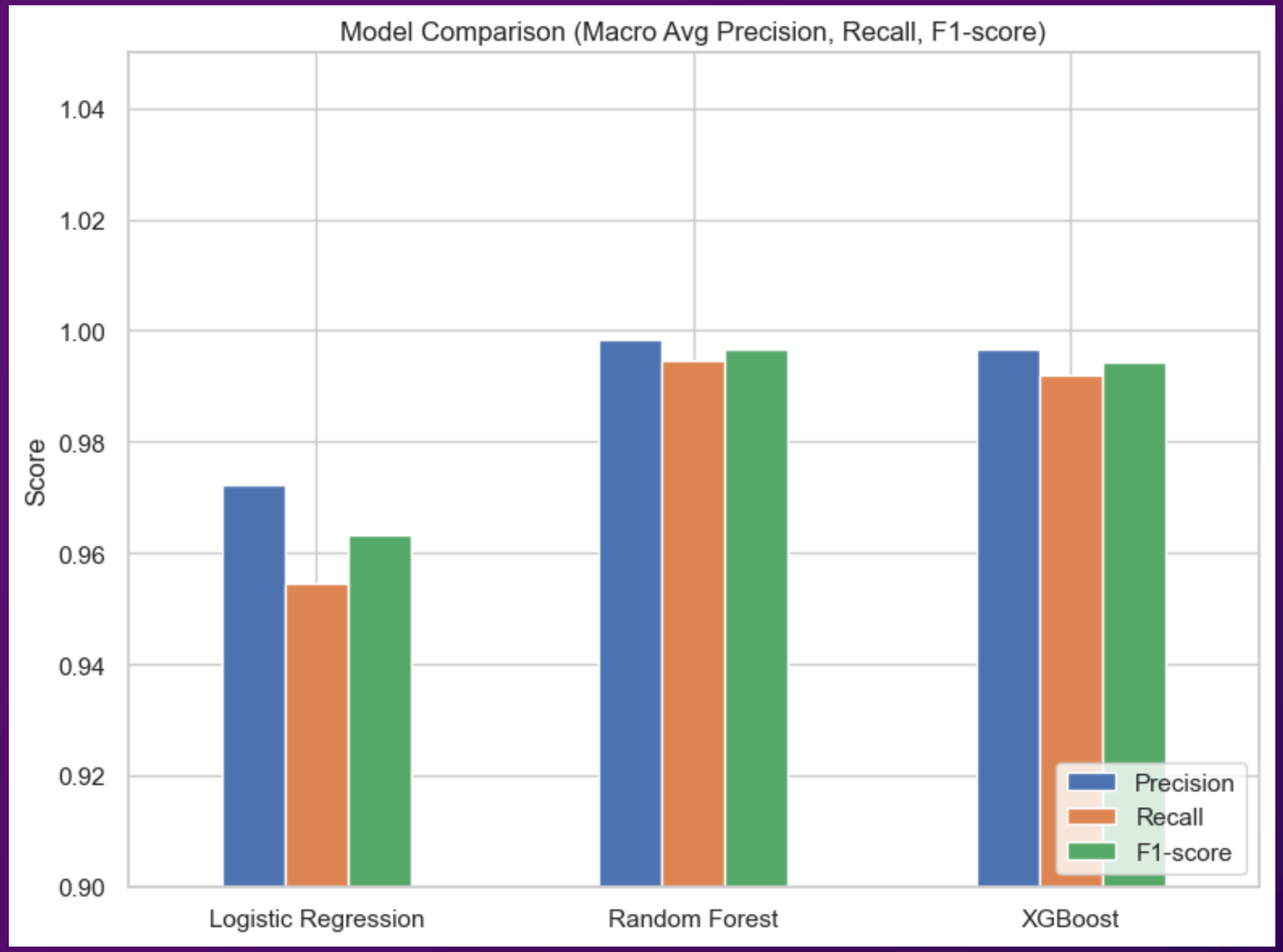
PREDICTIVE MODELLING

After clustering, we treat the cluster labels as "pseudo-classes" and build supervised models to predict them.
This helps answer:

Can we predict which billing category a new customer falls into?
Which features are most important for prediction?

We experiment with different algorithms and compare performance to choose the most reliable model.
The objective is to predict customer billing/consumption and classify customer behavior.

MODEL PERFORMANCE



Summary of Classifier Performance

1) Logistic Regression

Accuracy: ~99.7%

Very high precision, but recall for minority class (class 1) is slightly lower (~0.94).

Best for speed and interpretability, but less robust for imbalanced data.

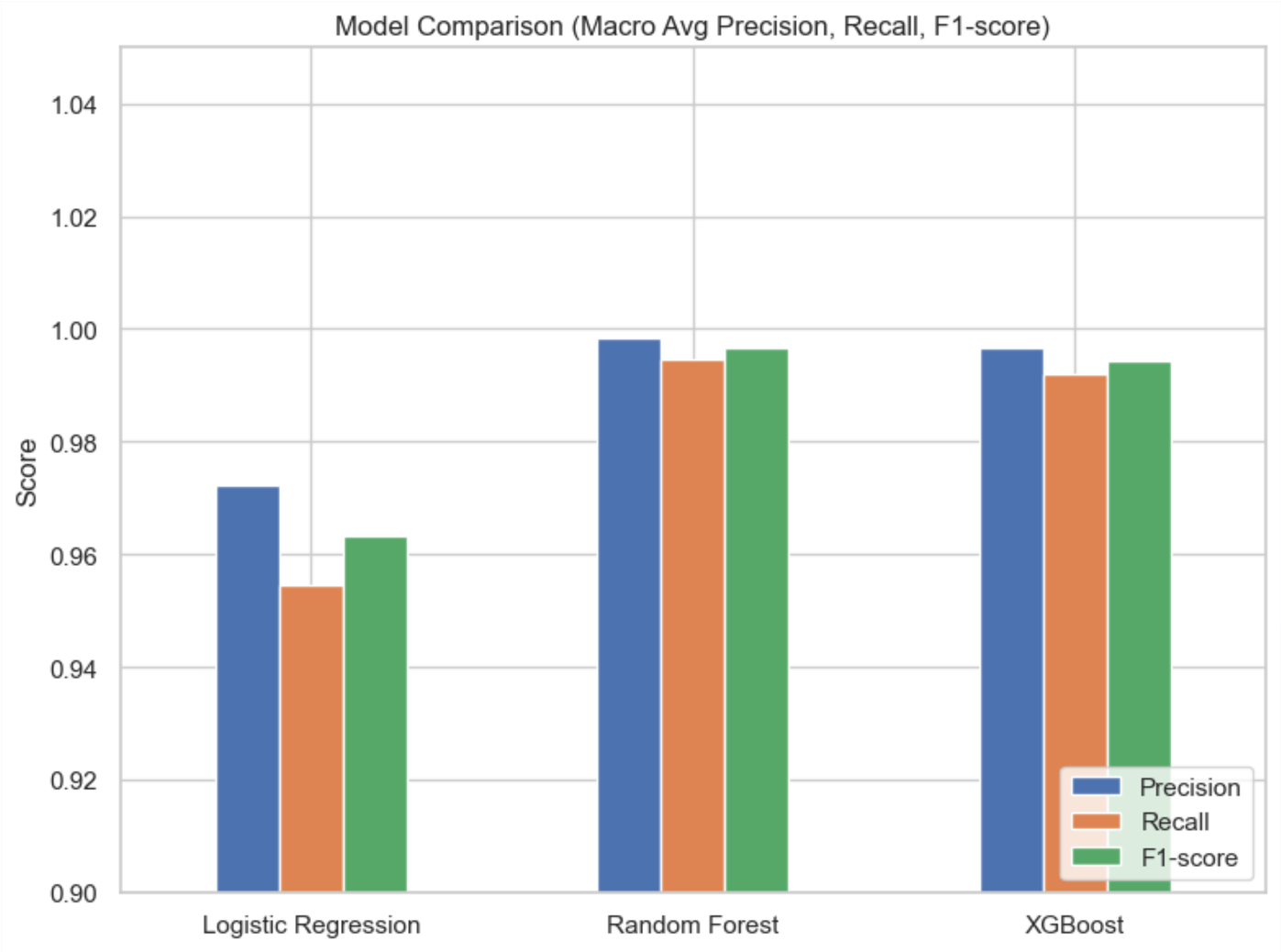
1) Random Forest Classifier

Accuracy: ~99.9%

Excellent balance of precision and recall (~0.99+ for both classes).

Handles complex relationships well; most reliable among the three.

MODEL PERFORMANCE...CONTD



Summary of Classifier Performance

1) XGBoost Classifier
Accuracy: ~99.8%
Performance very close to Random Forest, strong on precision but slightly weaker recall for class 1.
Efficient for large datasets, good at generalization.

Overall:

All three models perform exceptionally well (>99% accuracy).

Random Forest gives the best balance,

XGBoost is nearly as strong with better efficiency,

Logistic Regression is simple and fast but slightly weaker on minority detection.

RECOMMENDATIONS

Technical Recommendations

The rollout of smart meters should begin with high-consumption or high-loss zones such as industries, commercial users, or clusters showing suspicious usage from GMM/KMeans analysis.

A phased approach—moving from pilot projects to wider expansion and eventually citywide coverage—will help reduce costs and test effectiveness. Real-time monitoring is key, with smart meters transmitting hourly or daily consumption data.

Automatic alerts should flag sudden spikes (possible leaks), zero consumption over extended periods (potential illegal connections), or continuous flow during unusual hours (possible burst pipes), enabling quick detection and intervention.

RECOMMENDATIONS

Operational Recommendations

Data Analytics for NRW Reduction

Operational recommendations for reducing Non-Revenue Water (NRW) should focus on leveraging data analytics, proactive maintenance, and automation.

Customer clustering can help identify high-risk groups, where overlapping clusters may signal inconsistent billing and outliers may indicate illegal connections or faulty meters.

Predictive models can be applied to compare expected versus actual usage, flagging anomalies early.

Proactive maintenance should include scheduling field checks whenever smart meters detect unusual patterns and replacing faulty meters promptly to prevent hidden losses from calibration drift.

Finally, integrating smart meters directly with billing systems ensures automated, transparent billing, reducing human interference and minimizing risks of corruption or under-reporting.

RECOMMENDATIONS

Policy and Governance Recommendations

NRW reduction efforts should be guided by clear, measurable targets, such as lowering losses from 45% to 25% within three years, supported by performance dashboards that track and publish progress regularly.

To strengthen implementation, public–private partnerships can be leveraged by collaborating with technology providers to access affordable smart meter solutions, while adopting performance-based contracts that tie vendor payments directly to demonstrated reductions in NRW.

KEY INSIGHTS

NRW reduction efforts should be guided by clear, measurable targets, such as lowering losses from 45% to 25% within three years, supported by performance dashboards that track and publish progress regularly. To strengthen implementation, public–private partnerships can be leveraged by collaborating with technology providers to access affordable smart meter solutions, while adopting performance-based contracts that tie vendor payments directly to demonstrated reductions in NRW.

Contact Us

veesandra30@gmail.com

lilianwangui266@gmail.com

samiellesimple@gmail.com

jnnjogu6@gmail.com

leemwa244@gmail.com

oduorlevi24@gmail.com

Q&A

Thank You

For Your Attention