



DREXEL UNIVERSITY

LeBow
College of Business

precisely

BSAN 710 - FALL 2024

**DREXEL LEBOW CAPSTONE
PROJECT**

SEGMENTATION OF RESIDENTIAL CUSTOMERS

Professor Dr. Hande Y. Benson

Anusha R
Hema Gunapati
Nithin Venkatesh
Rohith Kalluri
Sai Rahul Munigala

Table Of Contents

<i>Executive Summary</i>	<i>1</i>
<i>Introduction to Business Challenge.....</i>	<i>2</i>
<i>Data Review</i>	<i>3</i>
<i>Data Analysis</i>	<i>4-5</i>
<i>K-means Clustering.....</i>	<i>4.1</i>
<i>Feature Importance</i>	<i>4.2</i>
<i>Regional Analysis</i>	<i>4.3</i>
<i>Business Insights & Recommendations.....</i>	<i>6</i>
<i>Appendix</i>	<i>7-8</i>
<i>References</i>	<i>9</i>

EXECUTIVE SUMMARY

Our project delivers actionable business insights for the telecom industry by analyzing customer data across the United States to understand demographics and connectivity needs. The primary goals were to understand the customer demographics of the current base customers, focus on three states—Nevada, Florida, and Vermont—to identify regional trends, and provide recommendations to help telecom providers acquire new customers. Using a dataset of approximately 40,000 records enriched with Precisely's APIs, we integrated property-specific attributes and PsyteGeodemographics information, adding features such as income levels, property tenure, urban/rural classifications, and household composition. This enriched data offered a comprehensive view of customer characteristics nationwide.

To extract meaningful insights, we applied K-Means Clustering, grouping customers into distinct clusters based on demographic and geographic traits. Supporting methods included Chi-Square Tests to assess the significance of categorical variables like property tenure and urban/rural classification, and Principal Component Analysis (PCA) to highlight key segmentation factors such as socioeconomic indicators, property characteristics, and geographic distinctions. The resulting dataset, comprising approximately 38,000 rows and 33 columns, served as a robust foundation for analysis, ensuring a precise understanding of diverse customer segments.

Our analysis revealed five distinct customer clusters, each highlighting unique preferences and characteristics. These included rural and semi-rural customers, urban professionals, and suburban families seeking tailored packages. Regional patterns showed Nevada dominated by urban professionals (Cluster 2), Florida by suburban families (Cluster 1), and Vermont by rural characteristics (Cluster 4). Sub-clustering further refined insights, with Cluster 4 segmented into three sub-clusters showing diverse rural connectivity needs, and Cluster 5 into two sub-clusters reflecting an urban-rural blend.

Strategic recommendations based on these insights include expanding broadband in underserved rural areas like Vermont, offering flexible packages for tech-savvy urban professionals in Nevada, and enhancing family-oriented services for suburban families in Florida. These insights equip telecom providers with targeted strategies to improve customer satisfaction and expand market reach. By combining clustering techniques, statistical analysis, and enriched data, this project successfully meets its objectives. It underscores the importance of data-driven approaches in uncovering actionable insights, equipping telecom providers with targeted strategies to improve customer satisfaction and expand market reach.

INTRODUCTION TO THE BUSINESS CHALLENGE

The telecom industry serves as the backbone of modern communication, playing a pivotal role in connecting people, businesses, and communities. Despite its significance, the industry faces persistent challenges, particularly in addressing the diverse needs of customers across the United States. Issues such as unreliable broadband services, with 26% of households experiencing connectivity problems, and a lack of high-speed internet access for 50% of rural Americans underscore the pressing need for targeted service improvements. As digital transformation accelerates, the demand for robust and reliable connectivity continues to grow, making it imperative for telecom providers to understand their customers deeply.

This project was motivated by the growing complexity of customer demographics and the pressing need for telecom providers to enhance their market reach and service delivery. The U.S. is a vast and diverse market, where customer preferences vary significantly based on geographic, economic, and social factors. To remain competitive, telecom providers must move beyond generic strategies and adopt data-driven approaches to identify, understand, and cater to distinct customer profiles. The U.S. telecom services market is predicted to rise by 6.6% between 2024 to 2030 and telecom companies that act first can take advantage of the growing consumer needs. By analyzing enriched customer data and leveraging advanced techniques like clustering and statistical analysis, this study aims to uncover actionable insights that can guide providers in bridging connectivity gaps and aligning their services with customer needs. The insights derived from this project will not only help improve customer satisfaction but also enable telecom providers to capitalize on untapped market opportunities, ensuring sustainable growth in an increasingly competitive landscape.

DATA REVIEW

Data Enrichment, Random Sampling and Data Cleaning

To effectively analyze customer demographics across the United States, we focused on residential units within the PIF dataset and employed a structured, multi-step process involving sampling, data enrichment, and rigorous cleaning to prepare the data for advanced analysis. The initial dataset was extensive, and to make it manageable while maintaining representativeness, we utilized random sampling. This approach reduced the dataset to approximately 40,000 rows while ensuring a balanced representation of customer demographics and geographic regions. By retaining diversity across states and regions, the sample allowed us to perform meaningful analyses of customer trends and patterns.

Using Precisely's APIs, we enriched the dataset with critical geodemographic and property-specific attributes to create a detailed and comprehensive view of customer profiles. The enrichment process integrated attributes such as income levels, urban and rural classifications, property tenure, household composition, and other key demographic variables. Each record in the dataset was uniquely identified and linked using PBKey and address identifiers, ensuring accuracy and enabling granular analysis. This enrichment added significant value to the dataset, providing a multidimensional perspective on customer characteristics.

Following enrichment, we performed an extensive data cleaning process to enhance the accuracy, reliability, and consistency of the dataset. This process included removing duplicate entries to eliminate redundancies and ensure data uniqueness. Missing values in categorical columns were addressed using mode imputation, while missing values in numerical columns were handled using median imputation to preserve the integrity of the data distribution. Standardizing formats across all fields ensured that the dataset was compatible with advanced analytical techniques. These cleaning steps resolved inconsistencies and prepared the dataset for seamless integration into subsequent analyses.

The resulting dataset consisted of approximately 38,000 rows and 33 columns, offering a robust foundation for clustering and statistical methods. It encapsulated a wide range of customer-specific attributes, including geodemographic and property details, enabling a comprehensive understanding of diverse customer segments. By structuring the dataset effectively, we created a reliable basis for identifying distinct customer profiles, assessing their connectivity needs, and uncovering actionable insights to inform strategic decisions in the telecom industry. This systematic approach to sampling, enrichment, and cleaning ensured that our analysis was both precise and impactful, positioning telecom providers to meet the varied demands of their customer base more effectively.

DATA ANALYSIS

K-means Clustering

To achieve the goal of segmenting customers based on demographics and property characteristics, we employed K-Means Clustering, a scalable and widely used method for customer segmentation in industries like telecom. This approach efficiently processed a dataset of 38,000 rows and 33 features, producing intuitive centroids that represented distinct customer groups. The Elbow Method validated the selection of K=5 clusters, capturing segmentation without overfitting shown in **Figure 1**.

The clustering analysis identified five distinct customer segments:

- Cluster 1: Semi-Rural Mix – Moderate-income families prioritizing affordability and multi-user functionality.
- Cluster 2: Affluent Urbanities – High-income, middle-aged individuals requiring premium services and smart solutions.
- Cluster 3: Comfortable Suburbanites – Affluent families valuing reliable, family-oriented broadband plans.
- Cluster 4: Rural Resilient Customers – Underserved rural areas with significant connectivity gaps and infrastructure needs.
- Cluster 5: Affordable Renters – Predominantly urban, price-sensitive customers with some rural traits.

To refine the analysis, we performed sub-clustering on Cluster 4 and Cluster 5. Cluster 4, entirely rural, was divided into three sub-clusters reflecting varying connectivity needs, property types, and income levels. Cluster 5, predominantly urban but with rural influences, was split into two sub-clusters capturing differences in income and service preferences. These sub-clusters provided a granular understanding of customer connectivity needs.

Feature Importance

To simplify the high-dimensional dataset and isolate the most influential features, we applied Principal Component Analysis (PCA) to reduce the data into three principal components as shown in **Figure 2**. By applying a threshold of 0.1 to PCA loadings, we identified the top contributing features for each principal component. The analysis identified three principal components that drove meaningful segmentation:

- 1.PC1: Socioeconomic Factors – Included Adult Age, Household Income, and Property Type as key contributors.
- 2.PC2: Household and Property Characteristics – Highlighted Property Value and Household Composition as significant factors.
- 3.PC3: Geographic and Tenure Distinctions – Emphasized Urban/Rural Classification and Property Tenure as critical variables.

Additionally, a Chi-Square Test identified Property Tenure, Urban/Rural Classification, and Adult Age as the most influential categorical variables, providing further support for their role in segmentation. These findings ensured the clustering model effectively captured demographic and geographic diversity in customer profiles.

DATA ANALYSIS

Regional Analysis

Focusing on Nevada, Florida, and Vermont, regional analysis revealed state-specific trends shown in **Figure 3**

- Nevada: Dominated by Cluster 2 (Affluent Urbanities), highlighting a wealthy urban demographic.
- Florida: Aligned with Cluster 1 (Semi-Rural Mix), representing moderate-income, family-oriented households.
- Vermont: Predominantly Cluster 4 (Rural Resilient Customers), showcasing high demand for broadband and connectivity challenges.

BUSINESS INSIGHTS & RECOMMENDATIONS

Our clustering analysis revealed significant trends and opportunities for the telecom industry, providing actionable insights to address customer connectivity preferences and unlock growth potential. By leveraging this segmentation framework, telecom providers can target specific customer segments with tailored strategies, ensuring improved customer satisfaction and market expansion.

Cluster Insights

Cluster 1 Semi-Rural Mix: A diverse, semi-rural-like cluster with family-oriented households, stable housing, and older demographics.

Cluster 2 Affluent Urbanites: Wealthy urban dwellers in densely populated areas with stable housing and financial independence.

Cluster 3 High-Income Suburbanites: Affluent, family-oriented suburban households with stable, high-value properties.

Cluster 4 Rural Resilience: A rural, cost-conscious cluster characterized by stable residences, family households, and lower-value properties.

Cluster 5 Budget-Conscious Urban Residents: Urban or mixed-density renters living in affordable housing, often financially constrained but demographically diverse.

Regional Recommendations

Nevada (Cluster 2): Launch premium, high-speed internet plans for tech-savvy urban professionals with VIP support.

Florida (Cluster 1): Offer affordable, family-friendly bundles and improve network access in semi-rural areas.

Vermont (Cluster 4): Focus on basic broadband plans with public-private partnerships to fund rural infrastructure.

Key Business Benefits

- Customer Loyalty: Tailored plans meet customer needs and build trust.
- Revenue Stability: Diversified offerings stabilize income streams.
- Competitive Advantage: Focused strategies position providers to dominate underserved markets.

This model equips telecom providers with actionable strategies to address diverse customer needs, ensuring long-term growth and competitive leadership.

APPENDIX

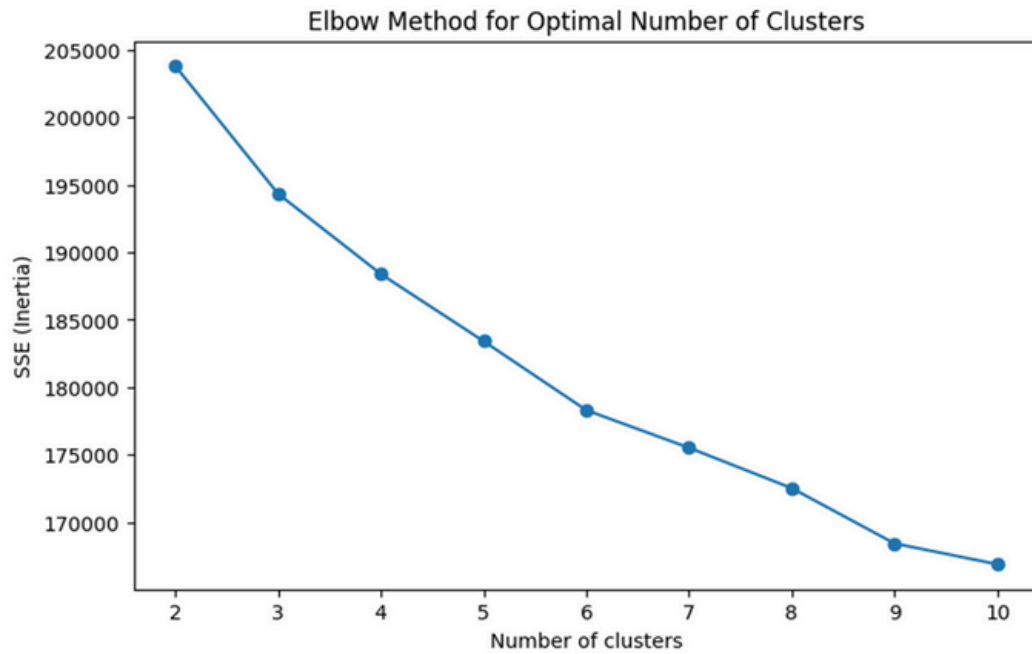


Figure 1: Elbow Method

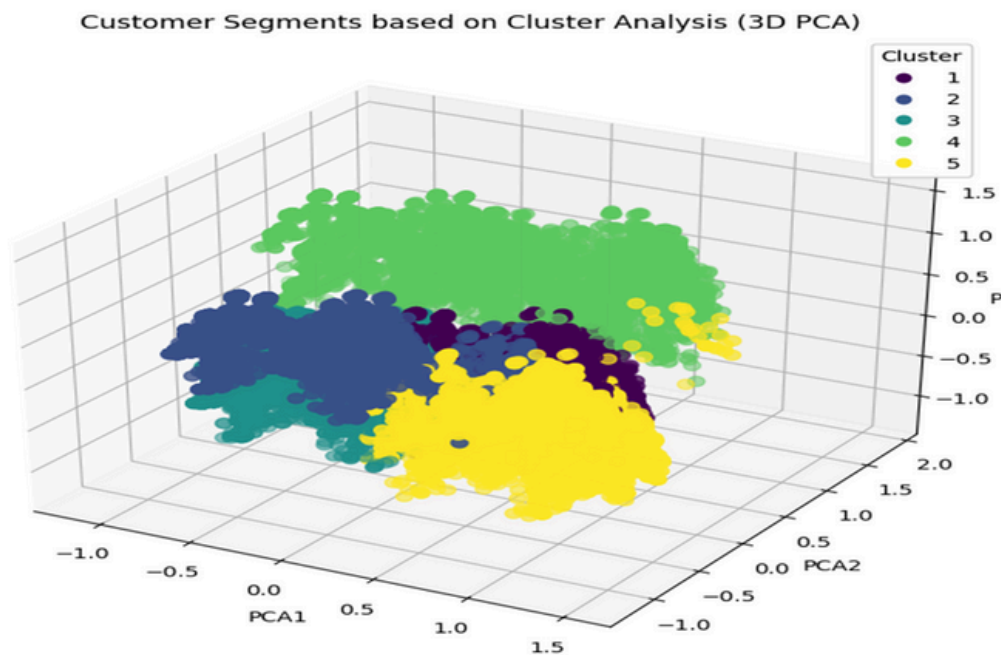


Figure 2: 3D Principal Component Analysis

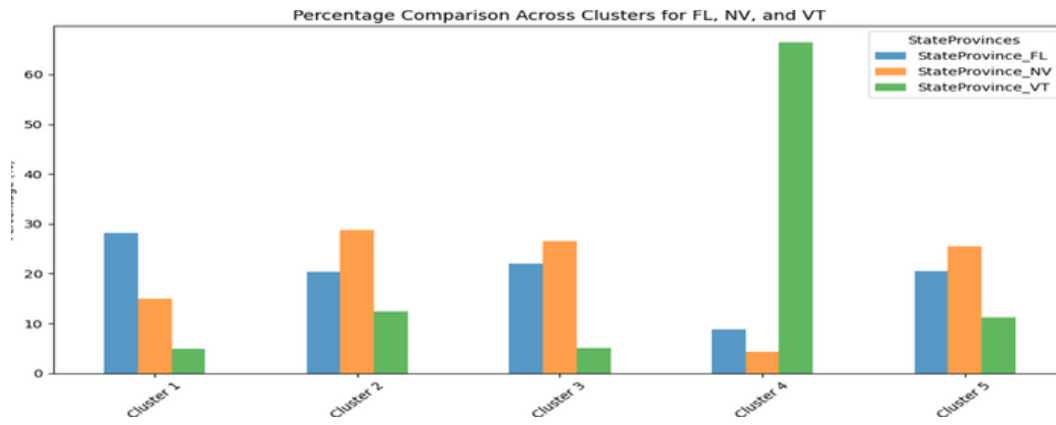


Figure 3: Cluster Comparison for FL,NV & VT

REFERENCES

- Baschnonga, A and Loozen, T. Is the digital home immune to the cost-of-living crisis? 2023, [LINK](#)
- Pugh, Cameron. Millions in rural America lack reliable internet. How Massachusetts towns got online. 2024, [LINK](#)
- Precisely. PSYTE™ US Data Guide. 2024, [LINK](#)
- PwC (2023). The connected consumer: Redefining the role of telecom in a digital era, [LINK](#)
- Measuring America's People, Places, and Economy, [LINK](#)
- U.S. Telecom Services Market Size, Share & Trends Analysis Report By Service Type (Mobile Data Services, Machine-to-Machine Services), By Transmission (Wireline, Wireless), By End-use, And Segment Forecasts, 2024 - 2030, [LINK](#)