Addressing Water Inequality: An Analysis of California's Public Water Systems and Policy Solutions for Disadvantaged Communities

Neha Burri - 11694929

Akhila Reddy Kommiti - 11705751

Pooja Thella - 11713291

Marisa Simha Sai Ravi Kumar - 11684252

University of North Texas

Toulouse Graduate School of Business

ADTA 5940 – Capstone

Dr. Jingjing Tong,

Spring 2025

Contents

CHAPTER 1: INTRODUCTION

Background

Being able to have safe drinking water stands as a fundamental right, yet different communities across the world, particularly in California, experience considerable gaps in their access to water quantity and quality. People from most parts of California benefit from sustainable water services, yet several marginalized communities in rural and low-income urban areas experience elevated risks from contaminated water supplies (Arche et al., 2024). Various social, economic, and environmental elements, together with racial discrimination, unstable economic markets, and outdated infrastructure systems, drive this inequality. The absence of equal clean water access leads to critical effects which cause health deterioration and economic difficulties and sustain negative poverty and social inequality patterns.

California's water system, despite its size and complexity, struggles with issues of equity in water distribution. The afluent parts of cities get steady, clean water supplies (Carchi et al., 2023). Still, poorer areas, mainly inhabited by Latino and African American residents, face insufficient water infrastructure due to poor quality water sources. Inspection results indicate higher exposure of residents in these regions to dangerous chemicals like arsenic and nitrates in their drinking water, which creates severe health concerns (Chen & Franklin, 2023). Historical, along with present social inequality patterns, combine with climate change issues to intensify these challenges so vulnerable groups remain with no sufficient access to vital water resources.

This study investigates the elements contributing to California's public water dispersion inequity, with particular attention paid to its effects on susceptible population groups. The research establishes connections between water quality standing, socioeconomic factors, and neighbourhood location to reveal how some neighbourhoods struggle more with clean water access

(Fernandez-Bou et al., 2023). Existing policies, technological solutions, and potential new approaches will be examined in the research to determine ways of reducing discriminatory water distribution patterns across the state.

Addressing water inequality in California stands as an essential requirement because it defends community health among disadvantaged residents along environmental equality between groups in society. Secure equal access to protected water supplies across California remains an emergency requirement since this western state faces escalating water scarcity during drought periods (Arche et al., 2024). The study adds to water justice discourse through systematic investigation of structural inequalities which maintain racial disparities and through the presentation of practical methods to attain water equity.

This investigation will answer three main questions by determining what social elements, together with economic conditions and environmental aspects, result in unequal clean water access throughout California. Which factors determine that marginalized communities face unequal water treatment, and how do racial, class and geographical locations influence water distribution (Carchi et al., 2023). The state needs new policies and technological innovations to guarantee fair water distribution, particularly for disadvantaged groups. This research investigates water inequality at multiple levels in California while creating strategies for securing water rights for all members of society.

Research Questions

Exploratory Data Analysis (EDA) Questions Geographic Distribution of Water Systems

The spatial distribution of water systems operates in what manner throughout California's regions along with counties and Field Operations Branches (FOBs)? The distribution of water systems

shows clustering according to different system types of particularly Community systems and Transient Non- Community systems.

This analysis aims to discover dense regions or water system-deprived areas for better distribution of resources and regulatory attention.

Population Served by System Type

What percentage of people uses each category of water systems including Community and Transient Non-Community as well as Non-Transient Non-Community? Does uneven distribution exist regarding

water service connectivity along with user population numbers between less- advantaged communities and more advantaged areas?

The goal is to show differences in water system capacity alongside documentation of communities without adequate water services.

Water Source Utilization

The total percentage distribution exists between water systems that draw their water from groundwater, from surface water and from purchased water. Does the selection of water source depend on specific geographic locations and does this quantity pattern match population characteristics and system classifications?

The purpose includes analyzing population dependence on unique water sources in addition to identifying drought and contamination exposure risks.

Regulatory Compliance and Operator Certification

The levels of operator certification for treatment and distribution units differ among water system facilities. What systems lack required certifications and what features do these systems have regarding population quantities and system types?

The goal: Assess operational readiness levels and detect which systems would face potential operational difficulties due to lack of operator skills.

Predictive Modeling Questions System Status Prediction

Using population data alongside service connections and water source types and regulatory compliance information will the status of a water system predict whether it will change from active use to inactive status?

The goal is to preventively detect systems which might fail or deactivate so appropriate protective actions can be executed.

Population Growth Impact on Water Systems

What effect will future demographic growth in California have on local water service needs? Do we have tools to determine which systems will surpass their current capabilities or need equipment updates?

The organization needs to invest in infrastructure while confirming their systems will satisfy future requirements.

Disadvantaged Community Identification

Is there a way to forecast which water systems will meet the criteria for "Disadvantaged Small Community Water Systems" (DAVCS) through mixing population data with fee codes and socioeconomic elements?

The intervention aims to direct available funding resources toward helping disadvantaged communities receive proper support.

Sanitary Survey Frequency Prediction

What are the chances that a sanitary visit will become overdue or needs to occur in the near future for systems across population groups and water sourcing methods when we observe their previous survey dates?

Organizations should optimize their inspection schedules to accomplish prompt monitoring of high- risk systems.

CHAPTER 2: LITERATURE REVIEW

Historical Context: Overview of California's water infrastructure and inequality

The water systems of California deal with multiple complicated challenges which encompass both reliable water availability and water quality together with proper distribution among different populations. The state operates under extensive challenges because it exists across multiple geographic zones and demonstrates financial disparities between regions while dealing with climate change consequences. As the biggest and most intricate water system in the United States exists in California the state faces water shortage alongside pollution and unequal distribution of clean drinking water. Each distinct area within the system shows different water conditions because its dependable water resource supply levels change accordingly.

The state of California retrieves its water supply from Sierra Nevada snowmelt and maintains reservoirs as well as aqueducts to operate its water distribution network. Continuous pressure on these water supply systems continues during drought years because climate change has produced both longer and more intense dry spells (Agustí Pérez-Foguet, 2023). Water resource management in the state must balance sustainable water delivery to its 40 million residents together with broad agricultural operations through scarce resources (Agustí Pérez-Foguet, 2023) Water storage and distribution fails to meet increasing challenges because dry periods affect California with increasing regularity.

The principal water quality problems continue to threaten communities across the entire state of California. The residents throughout rural areas and low-income districts of Central Valley and southern California lack reliable access to water purity. Federal water requirement limits are exceeded by unsafe drinking water, which affects more than one million people across California,

according to national reports. Public health risks plague communities that obtain water from systems contaminated with nitrates, arsenic, and lead contamination. The water systems cannot afford to upgrade their infrastructure because they do not have sufficient financial resources to maintain it.

Groundwater supply represents the primary water source for the state as it attempts to meet escalating water requirements, which creates a major problem. The excessive extraction of groundwater has led to surface land destruction and deteriorating water quality in defined areas. The contamination of groundwater has made it harder for communities to access clean water through these pollution problems.

California faces limitations in solving its water issues because water supply concentrations are unbalanced across the state. The Central California cities of Los Angeles, San Francisco, and San Diego operate secure, clean water supply systems, yet rural communities, especially those located in the Central Valley, encounter water contamination issues. Better funding for infrastructure combined with wider water management involvement will result in equal water accessibility for the people of California.

Environmental Racism & Disparities: Disproportionate effects on Latino and African American populations

The effects of polluted water on Latino and African American populations in California represent a major issue that recent research has started to analyze more intensively. The research conducted by Acquah & Allaire (2023) and Fernandez-Bou et al. (2023) shows that these populations typically reside in places where water pollution rates are high, thus negatively impacting their medical situation and life conditions. Communities with poor water quality mainly consist of residents from lower-income brackets who have little influence on political matters.

The populations seek water from outdated and maintained unsatisfactorily water supply systems. The rural Latino communities located in the Central Valley particularly depend on secondary water service systems that tend to harbour nitrates, arsenic or bacterial contaminants (Chen & Franklin, 2023). These pollutants present significant dangers to public health and affect three main vulnerable groups: children, elderly citizens, and women who are pregnant. Those who continually encounter hazardous substances in their water supply face increases in their potential to develop cancer alongside birth abnormalities and developmental delays.

Agustí Pérez-Foguet, (2023) argues that historical practices alongside housing segregation and environmental racism together expose Latino and African American communities to unequal amounts of environmental hazards that include polluted water. His claims are seconded by Carch et al., 2023 and (Chen et al., 2023) who argues that Such communities struggle to get clean water because they face different barriers which include restricted funds along with weak political standing coupled with complicated administrative systems that overlook their needs.

Thanks to the environmental justice movement, the nation now recognizes the unequal distribution of water services among communities and seeks justice through improved water infrastructure development in these areas (Acquah & Allaire, 2023). The California Department of Water Resources, together with other state agencies, currently tackles these disparities through their implementation programs. California Directorates focus their water infrastructure investment funds on areas that have higher pollution rates (Fernandez-Bou et al., 2023). Many efforts must be enacted to deliver clean drinking water safety to all people throughout California without discrimination based on either race or economic level.

Governance and Policy Developments: SGMA, California Water Plan 2023

The development of water policies in California spans throughout history to respond to the fast growth rate and complex natural resource management needs across the state. Since the start of the state's water management system, California has built big infrastructure elements, including dams, reservoirs, and aqueducts (Goddard et al., 2021). The State Water Project and Central Valley Project development enabled northern California water allocation to transport water throughout southern regions, supporting agricultural growth and urban expansion.

The centralized water management system authorized to serve agricultural practices and urban development has received growing opposition (Hanak & Chappelle, 2025). California has started implementing more sustainable water management approaches because water scarcity and environmental preservation have become vital issues (Goddard et al., 2021). The Sustainable Groundwater Management Act (SGMA) stands as a major achievement, becoming law in 2014 to control excessive groundwater usage while avoiding additional damage to the environment. Under SGMA, local agencies must write operational plans for managing their groundwater source areas and work toward sustainability by 2040.

The California Water Plan undergoes its most recent restructuring in 2023 to serve as the guiding document for state water resource management. Through this plan, the state focuses on climate change adaptation, water system resilience enhancement, and water equality between all communities (London et al., 2021). The 2023 revision of the California Water Plan emphasizes forward progress that serves disadvantaged communities and native tribes because they typically face exclusion from water management procedures. Building upon previous top-down water management systems, California now adopts an inclusive strategy to serve the entire population's social and cultural requirements.

Water policies in California need to address intricate challenges regarding restoring natural environments and preserving habitats. The Sacramento-San Joaquin Delta, together with other critical ecosystems, makes California its home while supplying water to millions of residents (Siirila-Woodburn et al., 2021). Controlling water management between human supply requirements and environmental sustainability demands rigorous oversight because climate change impacts extend with increasing intensity.

Technology and Tools in Water Management: Predictive analytics, remote sensing, recycling/desalination

The water management issues across California require breakthrough technological solutions to find optimal repair measures. Additions of predictive analytics enable better water quality outcome predictions and water resource availability forecasts through its operation combined with data modelling and real-time monitoring systems for water distribution networks (Archer et al., 2024). Water management systems become smarter and more efficient through implemented technological solutions because they reduce water scarcity during droughts and critical emergencies.

According to (Agustí Pérez-Foguet, 2023) these monitoring systems let water utilities recognize pollutants during their occurrence, resulting in the simultaneous protection of public health systems. Remote sensing allows for combined monitoring of underground aquifers and water surface elevations, enabling the detection of water levels that mostly exist in subsurface zones. Water management effectiveness in California has grown significantly because of advanced predictive modelling technologies now spread throughout the field. Modern predictive modelling systems produce water supply outlooks by analysing recorded data, which combines actual climate measurements with projected climate conditions (Hanak & Chappelle, 2025). The data collection

provides water executives with tools to choose suitable reservoir control systems and decide appropriate water allocations between agriculture, cities, and environmental protection needs.

Water recycling technology in California operates jointly with desalination methods to preserve water supplies. The wastewater facilities in Orange County operate their water recycling operation through state-financed capital and produce potable water from their treatment facilities (Agustí Pérez-Foguet, 2023). The desalination plant currently under development in Huntington Beach could supply water to California, but its implementation creates problems related to finances and environmental effects. California must implement these modern technologies to increase water stability. The state should dedicate funding to infrastructure expansion through targeted budget allocations while simultaneously developing water technology policies which need appropriate financial support.

CHAPTER 3: METHODOLOGY – DATA PREPARATION

Software

This research is dependent on Python to execute its data analysis because this programming language provides exceptional capabilities in data manipulation and analysis features. The data manipulation and preparation tasks could be completed through the panda's library, which enables easy processing of extensive data, including reading, cleaning and summarizing data operations. The data visualization relied on Matplotlib and Seaborn tools to visualize patterns, distributions, and relationships in the data contents. Pattern recognition tasks were carried out using sci-kit-learn for regression analysis, clustering procedures, and cross-validation. These analytical instruments enabled detailed research of the public water systems data while enabling researchers to study trends between the different variables.

Data Collection

Information regarding California water systems came from the Drinking Water Watch Public Water System Facilities dataset. The available dataset consists of 7,815 entries containing 35 columns with essential variables, including Water Source Type, System Status, Residential Population and Regulating Agency. The database originates from publicly available information the California Department of Public Health distributes. The dataset contains numerical and categorical variables, permitting exploratory data analysis (EDA) and machine learning model applications. New data entries continue to update the information base to maintain the highest accuracy and current status in all analyses.

drinking-water-watch-public-water-system-facilities.csv

1 to 10 of 7815 entries Filter

FOB	Region	Regulating Agency	Water System No	Water System Name	Principal County Served	Principal County Served, State	Federal Water Sys
SOUTHERN CALIFORNIA FOB	SECTION V - SOUTHERN CA SECTION	DISTRICT 16 - CENTRAL	CA1900045	1000 TRAILS / SOLEDAD CANYON PRESERVE	LOS ANGELES	LOS ANGELES, California	NC
NORTHERN CALIFORNIA FOB	SECTION II - NORTH COASTAL SECTION	LPA58 - NAPA COUNTY	CA2800014	11:11 WINERY	NAPA	NAPA, California	NC
CENTRAL CALIFORNIA FOB	SECTION III - NORTH CENTRAL SECTION	LPA69 - SAN JOAQUIN COUNTY	CA3901031	132 INVESTMENTS WATER SYSTEM	SAN JOAQUIN	SAN JOAQUIN, California	C
CENTRAL CALIFORNIA FOB	SECTION IV - SOUTH CENTRAL SECTION	LPA70 - SAN LUIS OBISPO COUNTY	CA4000725	141 SUBURBAN ROAD WATER SUPPLY	SAN LUIS OBISPO	SAN LUIS OBISPO, California	NTNC
CENTRAL CALIFORNIA FOB	SECTION IV - SOUTH CENTRAL SECTION	DISTRICT 19 - TEHACHAPI	CA1503684	148 EAST WATER SYSTEM	KERN	KERN, California	C
CENTRAL CALIFORNIA FOB	SECTION IV - SOUTH CENTRAL SECTION	LPA46 - KINGS COUNTY	CA1600293	15TH AVENUE	KINGS	KINGS, California	C

Figure 1 Sample Data from the Drinking Water Watch Public Water System Facilities Dataset: This table displays information about various public water systems in California, including system identifiers, regions, and the counties served.

Overview of the variables

The Field Operations Branch division, designated as FOB, serves the water system region throughout California. During current categorization procedures, the state requires the Field Operations Branch (FOB) variable to set water system divisions across California. Users obtain improved knowledge of California's larger geographical regions through the Region variable since this field determines whether the water system exists in the southern or northern parts of the state. The Regulating Agency variable determines all regulatory agencies together with their governing bodies that control water system operations. The water management agencies discharge their regulatory authority by implementing quality control mandates and regulatory standards. The Water System No is a specific tracking identifier assigned to every individual water system within the dataset.

The Water System Name describes each water system's specific name or identifier. The specific sites that comprise system names are "1000 Trails/Soledad Canyon Preserve" and "11:11 Winery."

The Principal County Served shows which county mainly depends on each water system to fulfill their water needs.

The combination between central service counties and their associated states becomes accessible to all users through the Principal County county-served state variable. The Federal Water Sys variable system performs official federal-type classification assignments to all water systems. Community systems are designated as "C," whereas both "NC" and "NTNC" represent non-community as well as non-Transient Non-Community systems. The classification system demonstrates what type of service logistics water systems provide regarding community-wide distribution versus temporary or residential group population delivery.

CHAPTER 4: EXPLORATORY DATA ANALYSIS (EDA)

The dataset we selected has Seven thousand nine hundred and fifteen records from the Drinking Water Watch Public Water System Facilities contain 35 datasets that supply complete information about California's public water systems. The dataset provides fundamental information about water distribution networks, their supply origins, and operational regions under regulatory management. The database contains a mix of categories and numeric types of variables. The dataset classifies the water systems through categorical variables, including FOB (Field Operations Branch) and Region, which identify Southern and Northern California areas. Each water system receives monitoring from the authority listed under the Regulating Agency column, which contains the system-specific identifiers Water System No and Water System Name. Other categorical variables in the dataset, including Federal and State Water System Types, organize systems into categories to mark their service types as Community, Non-Transient Non-Community, and Transient Non-Community.

Numerical Variables

Numerical variables collected from the data offer both operational performance results and characteristics of California's water systems. The metrics in the data set contain numerical values suitable for statistical assessment and predictive modeling functions. The Residential Population, Non-Transient Population, Transient Population, and other numerical variables form the primary section of data points. A subset of variables allocates population statistics based on residential and non-transient residents and transient residents who use water services from a system. The total population in water systems is determined by serving all population groups together. The analysis

encounters low data completion rates since the Residential Population contains numerous missing elements from its population-based variables.

Service connections consist of three numerical variables, which monitor connections between agriculturally, commercially, and residentially specialized facilities. These values supply essential indications about service requirements within the agricultural area alongside residential growth. These systems operate between zero-serving agricultural users and maximizing service to 2,462 agricultural users among numerous other systems with residential and commercial service areas. This dataset introduces the connections_per_capita measure to determine the service connection rate by dividing service connections by population numbers. The developed mathematical measurement shows the installation density of water systems throughout specified areas.

The days_since_last_survey numerical feature shows the computed number of days between the present time and the last water system inspection dates. Researchers utilize this feature to identify inspection-due systems within the dataset by analyzing monitoring frequencies. Numerical variables give detailed information about the populations served and distribute service connections throughout water systems while reporting operational status.

Categorical Variables

Numerical information within multiple important categorical variables allows the data segmentation to create significant data clusters. Water system characteristics manifest in this data through organizational standards related to geographic regions, regulatory authorities, and service delivery requirements. The field operations branch (FOB) and region variables serve as the primary geographic classifying variable that organizes Southern and Northern California areas. The chosen

divisions establish conditions to understand the functional and organizational differences between distinct geographical locations.

Users who access the Regulating Agency field can access regulatory oversight details through the listed authorization bodies in this column. Water System No and Water System Name serve as unique identifiers for each system, enabling precise tracking and analysis of individual water systems. All water systems require two categorical essential attributes that outline their Federal Water System Type and State Water System Type status. The system types for permanent residential and transient population services originate from classifications of the community and non-transient non-community and transient non-community services.

The system status categorical measurement determines the operational status of water systems between normal operations, maintenance, and inactive status. Standard performance analytics of water system reliability demand this information for proper examination. The classification of water systems relies on ownership information in the Owner Type and water supply origins from the Primary Water Source Type. The Treatment Plant Class teamed up with the Distribution System Class, allows organizations to assess operation complexity across all water systems through their treatment process complexity and distribution network complexity dimensions.

Feature Engineering

Machine learning model prediction quality improves through feature engineering by changing unprocessed data inputs into optimized format characteristics. The database held multiple newly developed variables since their goal focused on discovering patterns and enhancing analytical operations. The core engineered feature named connections_per_capita obtains its value by performing a population-based division of the Total Number of Service Connections by Total

Population. The new variable shows the available service connections that serve one individual using the designated area services. This supplemental variable assists in assessing water system capacity for new population growth while allowing analysts to predict areas likely to experience water system failures.

Evaluating Transient Population against Residential Population relies on ratio calculation to merge available numbers. The ratio allows organizations to analyze residential and temporary population size differences to compare water usage rates between these population groups.

The days passed since the last inspection appear in the Days_since_last_survey variable. The regulatory bodies use the days_since_last_survey indicator to create inspection periods that evaluate system needs for inspections and heightened oversight. Project team members utilized existing information to generate the Status_Label variable since it acted as the final output. The Status_Label function applies the surveyed intervals and their median value to establish water system status as either Active or Inactive.

The maintained variables boost the dataset's value, enabling analysts to create better analytical models that achieve improved performance outcomes. The present dataset supports predictive modeling applications while operational water pattern analyses and disparity detection become achievable by adding useful variables, including connections_per_capita transient_to_residential_ratio and days_since_last_survey.

Handling Missing Values

Data gaps are a regular phenomenon in datasets that require suitable treatment methods to produce accurate modeling and analysis results. The researcher would face analysis distortion from untreated missing data points throughout several variables within this dataset. All numerical and categorical data attributes required systematic procedures for handling missing values.

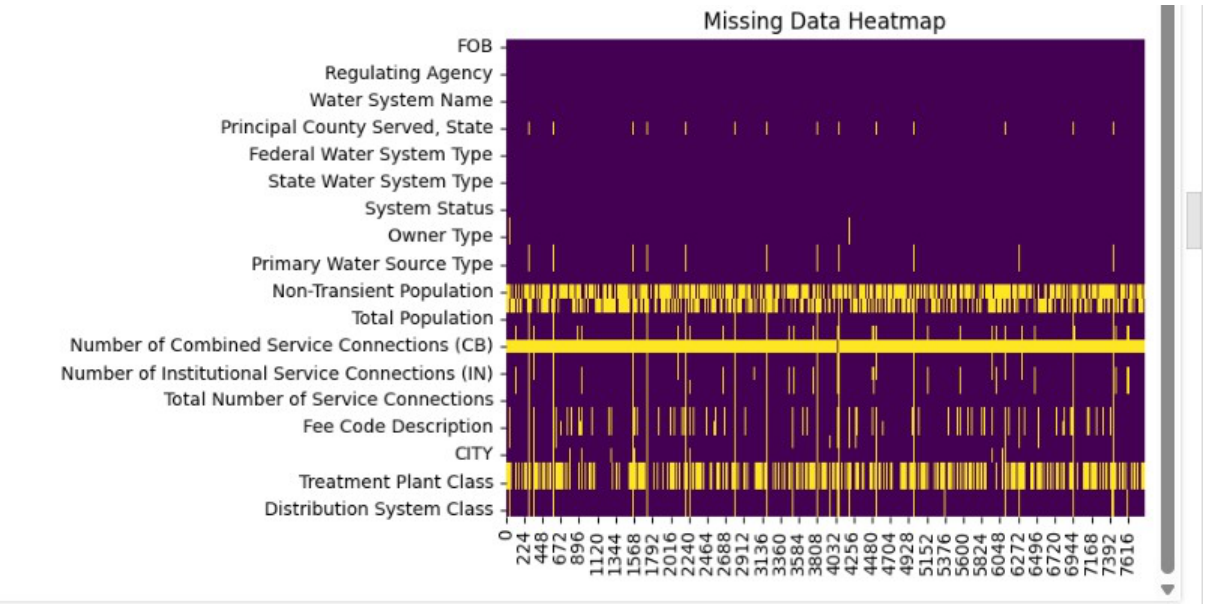

Figure 2 A missing data heatmap showing which data columns contain missing data. The visual display of yellow bars points out essential gaps in data within columns for example 'Number of Combined Service Connections (CB)' along with 'Fee Code'

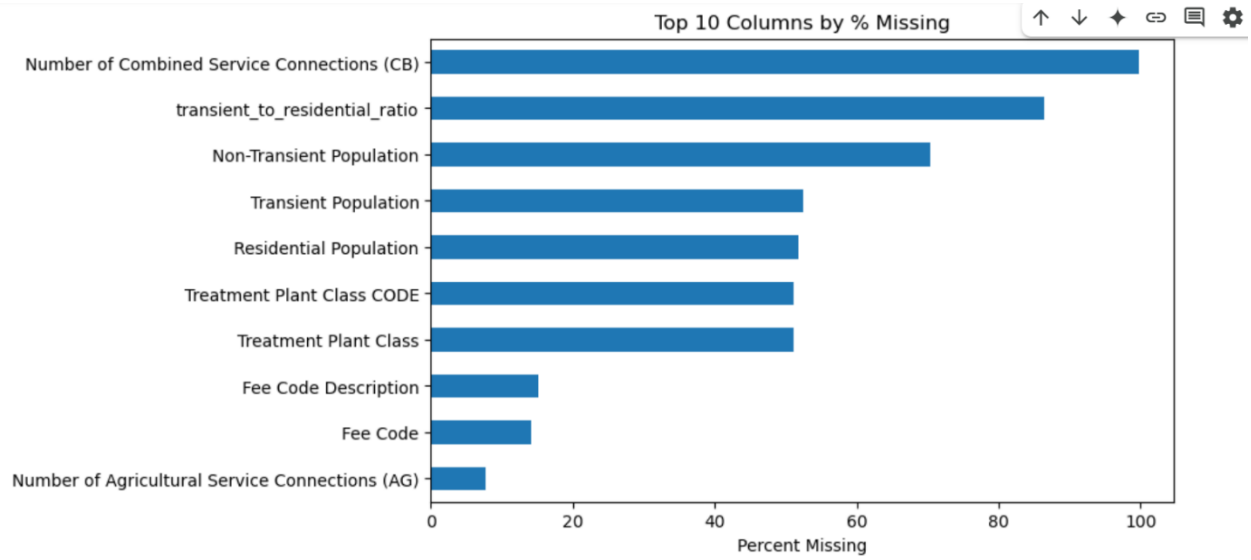

Figure 3 The bar chart displays the ten dataset columns with the most significant percentages of data missing. Analysis challenges will arise due to data completeness issues in the 'Number of Combined Service Connections (CB)' column along with 'transient_to_res

Handling Missing Numerical Variables

The Simple Imputer function from sci-kit-learn performed median imputation to fill in missing numerical values. The team selected median values as their preferred imputation method since they work well with outliers yet preserve the fundamental statistical distribution of the data for appropriate replacement of missing values that will not alter the data pattern. The Residential Population and Number of Agricultural Service Connections (AG) variables had major missing entries among three essential numerical datasets. The Residential Population variable contained 4051 missing entries, but the Number of Agricultural Service Connections had 609 records with missing data. The insertion of median values into these chosen columns-maintained data coherence and accessible relevant information, which would be helpful for future analytical and modeling applications.

Handling Missing Categorical Variables

The approach for handling missing categorical variables differed from the other methodology. Simple Imputer function used its constant strategy to fill in missing data points with a "Missing" value, which served as a replacement method for the data. The categorical data structure preservation method protected data integrity while preventing the elimination of anything due to unknown values. Three categorical variables within the dataset, Primary Water Source Type, Owner Type, and Fee Code Description, contained missing data entries. The Primary Water Source Type variable included 172 unrecorded entries, while the Fee Code Description recorded 1,178 empty fields. We inserted the term "Missing" into blank data points to build a complete dataset for analysis.

Handling missing feature engineered missing values

The new data features added to the dataset through variables connections_per_capita and transient_to_residential_ratio enhanced analytical capabilities for the analysis. Engineered variables can receive missing cell data that originates from source variables. The calculation resulted in null values for connections_per_capita whenever the Total Population or Total Number of Service Connections contained null values. The data scientists assigned numerical median values with a "Missing" category to replace missing entries in these features.

Impact of Missing Data Handling

The applied imputation methods produced an analytical dataset with complete information, minimal bias, and relevant complexity. The analytical use of median value protected the natural distribution patterns of numerical attributes without endangering the original makeup of categorical attributes. The adopted strategies safeguarded most original data entries, which enabled precise modeling and accurate predictions without data corruption due to missing values.

Conclusion

This chapter presented the dataset and included a detailed description of both data structure and significant variables. The dataset included both categorical types and numerical types to examine water system components, regional population trends, and service network conditions in California. Through the investigation, we discovered different data correlations that explained service infrastructure differences across different system types and water access variations among regions.

We removed missing values from the dataset using an imputation method that remained simple. The numerical data types containing missing values were replaced with median-based values to preserve the original data center points. Using the dummy placeholder "Missing," the model kept

the original format of categorical variables when some data points were absent. The dataset implemented these procedures to achieve data completeness before entering into analytical procedures that serve as a base for future predictive modeling work. Further analysis starts in the following chapters, thanks to the database foundation established during this chapter's work.

CHAPTER 5: PRELIMINARY RESULTS AND DISCUSSION

The first analysis involved the California public water system dataset. The research analyzed how various variables were arranged in the dataset and identified patterns. It also evaluated variations in water service delivery among geographic regions and system classification types. The first procedure included conducting Exploratory Data Analysis to visualize variable distributions, and then researchers searched for missing data and outliers in the dataset.

Feature engineering generated two variables that strengthened the relationship between water service infrastructure and population shifts through connections_per_capita and transient_to_residential_ratio. Our correlation analysis studied the numerical relationship between data variables that measured population numbers and service connections as inputs to discover water system performance factors. The methods in this chapter enabled us to extract vital findings from original data sources for future predictive modeling throughout subsequent chapters.

Descriptive Statistics and Data Overview

Several numerical factors in the dataset deliver essential information regarding the dimensions of public water infrastructure across California. The population statistics represent the most vital variables, with the residential and non-transient populations joining the transient and total populations. Each water system serves a designated number of permanent residents to be considered its residential Population. Water systems across California serve minimal groups of residents and gigantic populations that exceed thousands of people. Total Population incorporates all groups of residents, non-residential visitors, and other staying populations in its calculation. The third variable allows users to see the full scope of those who utilize the water systems. Most water supply systems distribute their Total Population to smaller groups of people throughout their

service areas. Most of the Population relies on the few large water systems among many smaller systems that serve only limited people. Most histogram data points stay within the lower section of the scale, demonstrating this skewed distribution pattern.

Service connection statistics and population information provide valuable information about water system infrastructure and user demands. Each water system has various agricultural users named Agricultural Service Connections (AG) ranging from zero to above 2,400. Most water systems provide service to only a small number of rural users and virtually no agricultural users despite some systems serving regions with higher numbers of agricultural and residential users. The Residential Service Connections database records residential customer counts in values similar to those in agricultural service connections. Most systems provide water to several hundred residential customers, whereas some networks reach tens of thousands of domestic users. Diverse population sizes exist between the communities dependent on varied water supply systems.

Distribution of Total Population Served

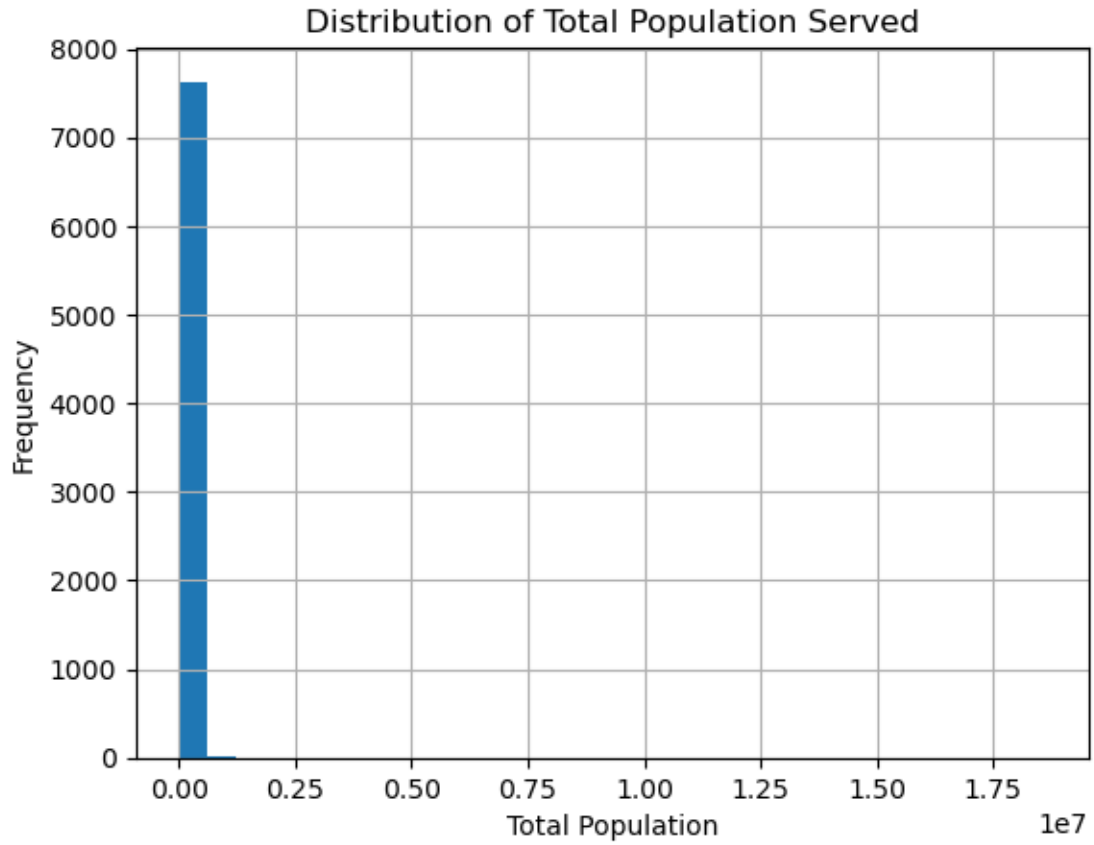

Figure 4 The histogram depicts where the public water systems in California supply their service to their populations. Most public water systems operate within small districts, with very few systems serving broad areas with a population. The data distribution pattern presents strong skewness because most water systems provide services to minimal population numbers.

Insights from the Distribution of Total Population Served histogram

Most of the data points in the histogram display strong right skewness due to the Total Population served by water systems. The distribution chart shows that nearly all facilities serve minor population segments since the large bar appears at the leftmost area. The distribution extends to the right through only a limited number of water systems that handle large metropolitan areas. Major metropolitan areas account for the population numbers controlled by few large water systems because smaller water systems primarily serve regional communities throughout

California. Databases, including infrastructure systems, show this skewed pattern because one or several enormous entities usually operate with significantly greater capacity than numerous smaller systems.

System by federal Type

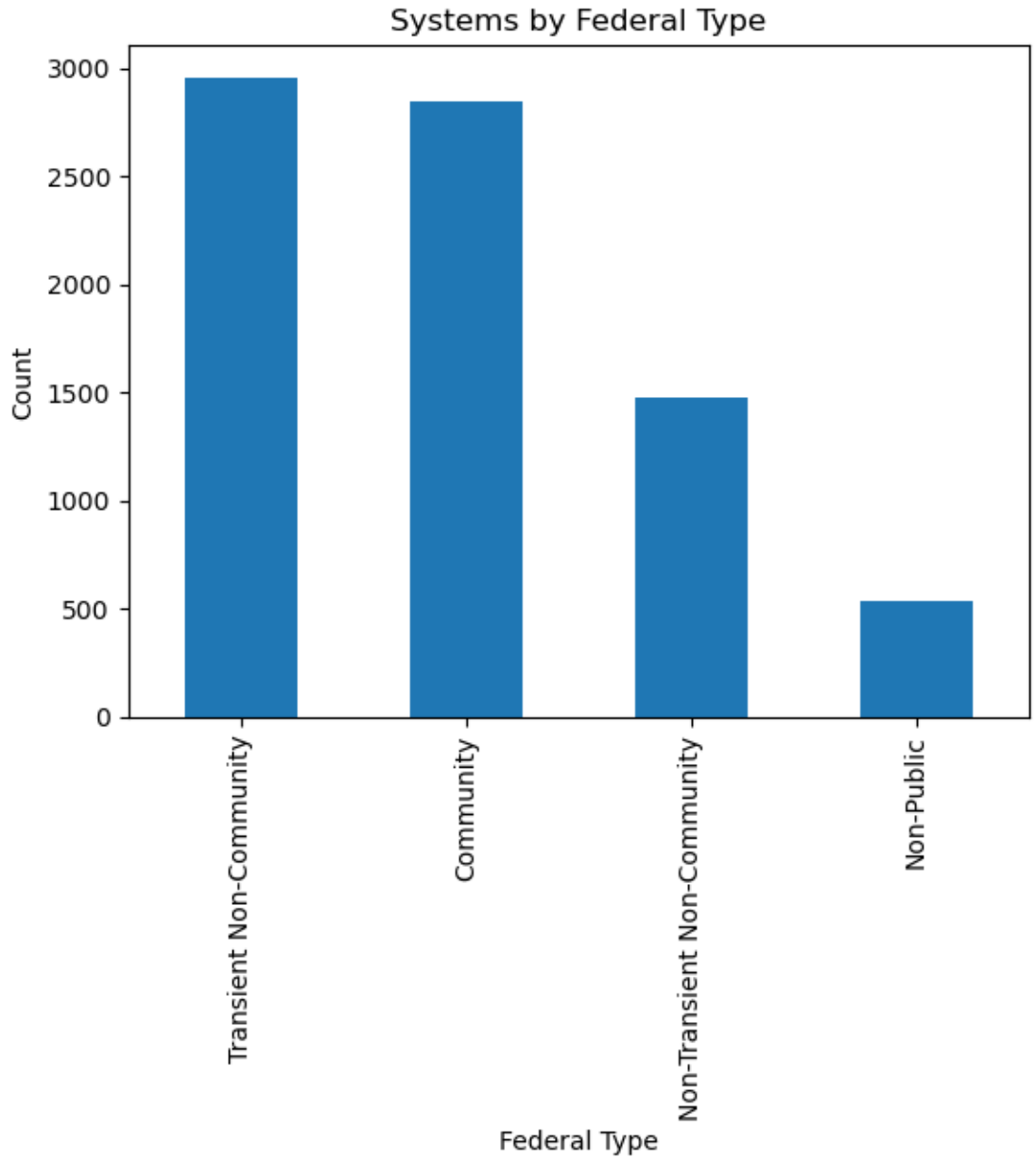

Figure 5 The distribution of water systems across California appears through bars according to their federal classifications. Among the water systems analyzed in the dataset, Transient Non-Community and Community are the biggest groups. In contrast, Non-Transient, Non-Community, and Non-Public systems comprise tiny fractions.

Insights from the System by federal Type bar chart

The data distribution shows that the transient non-community and community water systems are the primary systems since they encompass more than 2,500 systems. These system types deliver water to large population groups who occupy their facilities seasonally or permanently in residential areas. The data indicates that Non-Transient, Non-Community, and Non-Public systems appear infrequently because they serve constrained demographic populations or maintain private ownership. This statistical pattern reveals how California's water system infrastructure primarily serves communities and transient population needs across broad geographic areas.

Primary water sources types

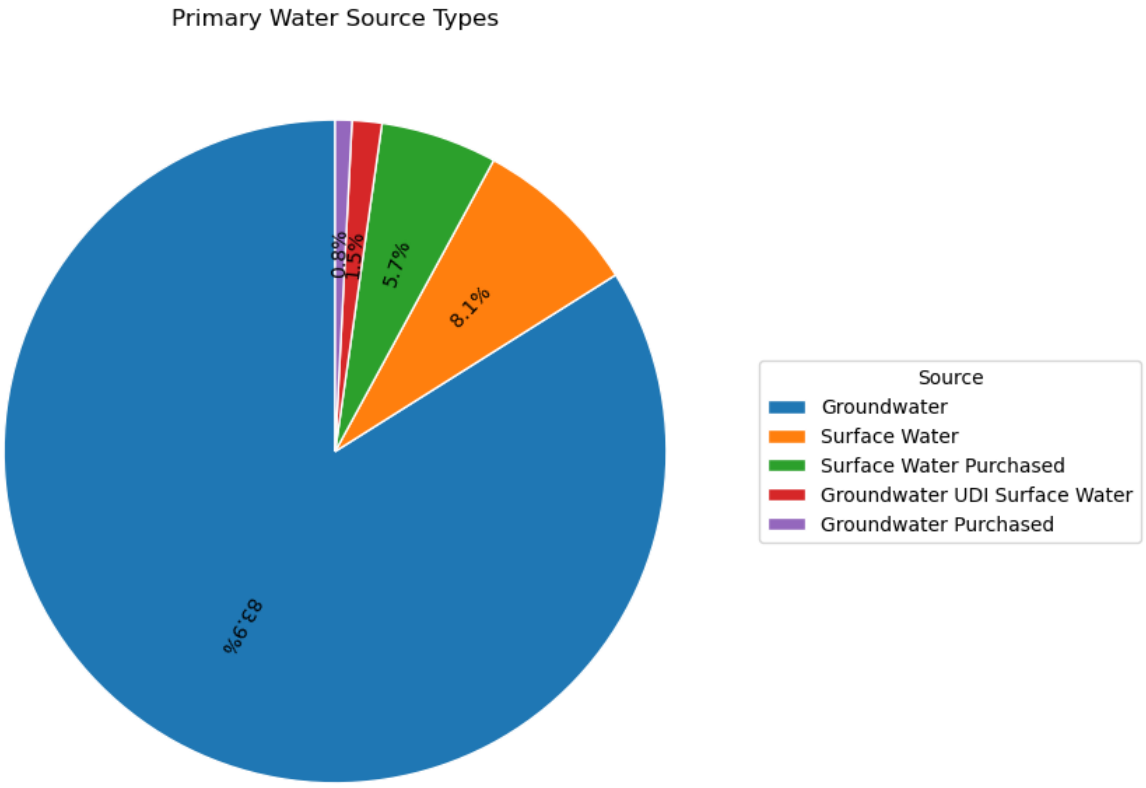

Figure 6 The chart shows how California public water systems obtain their main water supplies. The water infrastructure of public systems in California depends primarily on Groundwater at 83.9%, while Surface Water covers 8.1% and Surface Water Purchased

makes us up to 5.7%. The composition includes three minor sources, which comprise only a small fraction of the total UDI surface water and Groundwater purchased.

Insights from the Primary water sources types from the pie chart

Public water systems in California depend on groundwater for their main water supply because it represents 84 percent of all public water systems in the state. The state depends heavily on groundwater resources because surface water resources often face restrictions in the water supply or reach maximum utilization limits. The water supply of Surface Water and Surface Water Purchased represents minor ratios compared to the larger part of the water source. The remaining water supply sources, which include Groundwater UDI Surface Water and Groundwater Purchased, represent a minimal portion compared to the total supply because they are used less frequently. The allocation of water between different sources illustrates the necessity of groundwater regulation to ensure water supply sustainability throughout California.

Total Population by Federal System Type

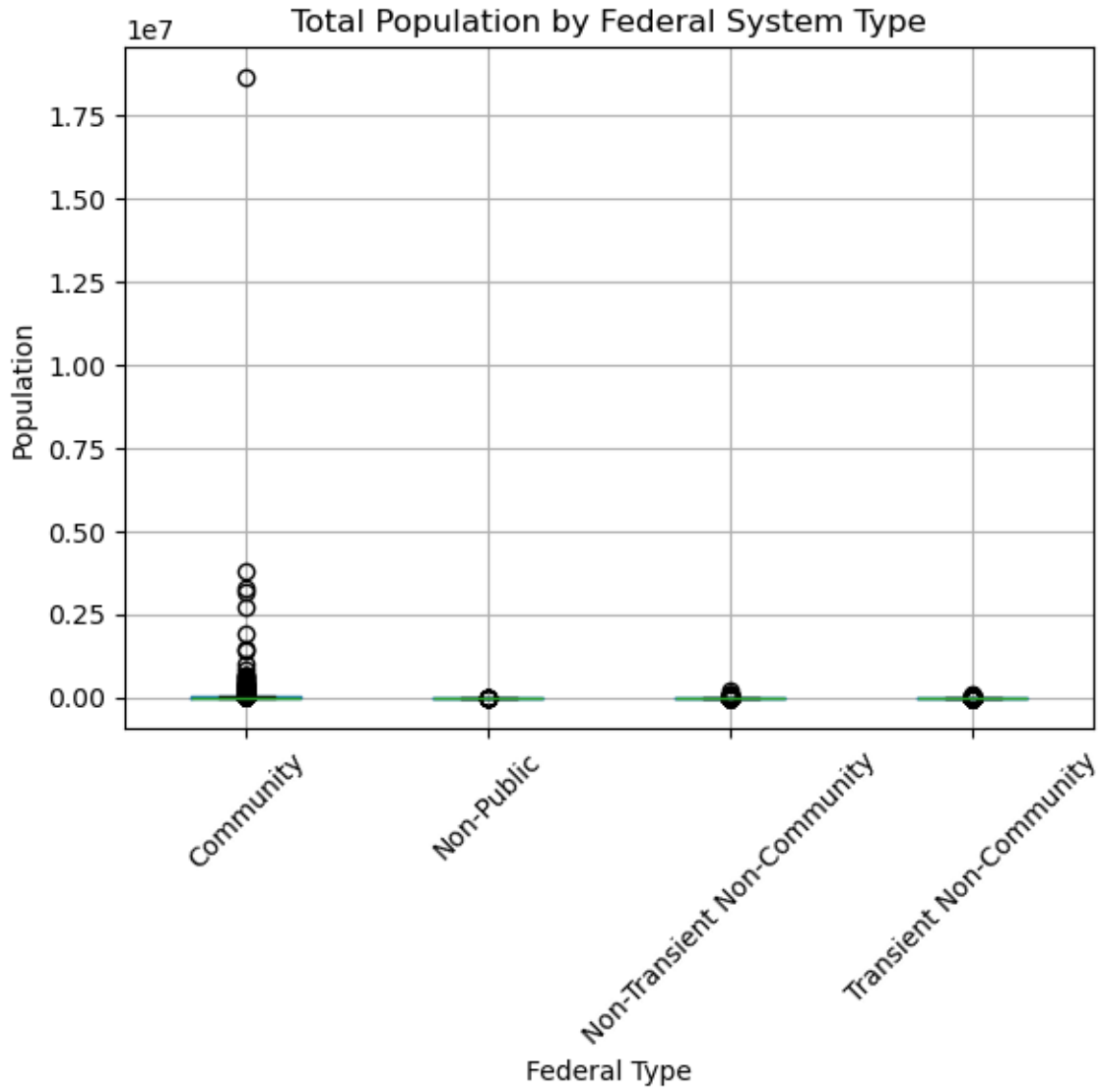

Figure 7 The box plot shows how different federal system types distribute their service populations. The community systems control the biggest population sizes, although some individual systems have unusually extensive populations. The total populations served by Non-Public, Non-Transient, Non-Community, and Transient Non-Community systems remain lower in numbers with fewer extreme population values.

Insights from the Total Population by Federal System Type box plot

The box plot shows substantial unevenness between populations who use the different water systems. Small urban districts and large cities provide services to most of the population, as shown

by the boxes and outlier points in the chart. The Systems of Non-Public type and Non-Transient Non-Community and Transient Non-Community categories serve much smaller population sizes than Community systems based on box plot data. Large urban communities form only a few outliers within the Community category, yet most supported population centers remain small. The current water infrastructure in California uses two different scales of facilities to distribute water across the state.

Sanitary survey visits per year

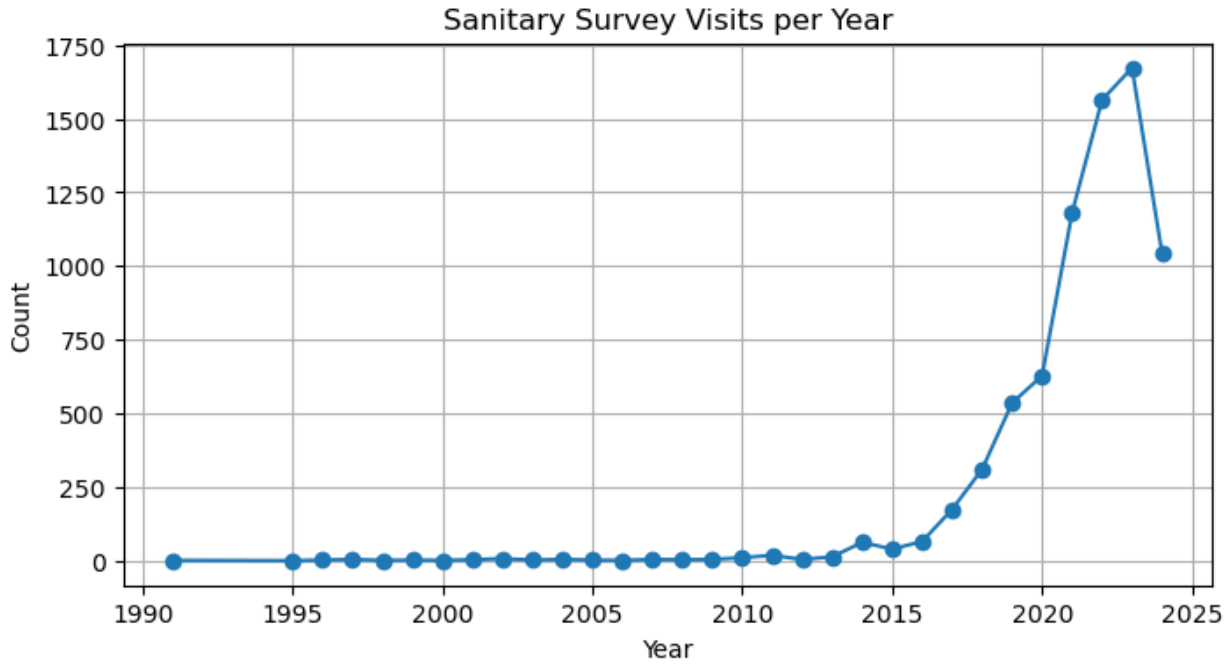

Figure 8 The chart represents sanitary survey visits throughout the period from 1990 to 2025. From 2020, survey activities experienced rapid growth, which reached its highest point until visits registered a significant decline after 2025. Recent years have shown a major escalation in the quantity of sanitary surveys performed according to the plot.

Insights from Sanitary survey visits per year time series

The total number of sanitary surveys showed sustained upward growth since 2020, reaching over 1500 annual inspections. Thorough monitoring of water systems became more common due to regulatory authorities and policy adjustments. Few sanitary surveys took place annually before this rise was recorded. Survey operations seemed to reach completion when the numbers showed a substantial decline during 2025. Water system monitoring shows advancement through the rising number of inspections as these inspections have gained growing importance for maintaining water quality standards and meeting regulatory requirements.

Active vs Inactive Systems

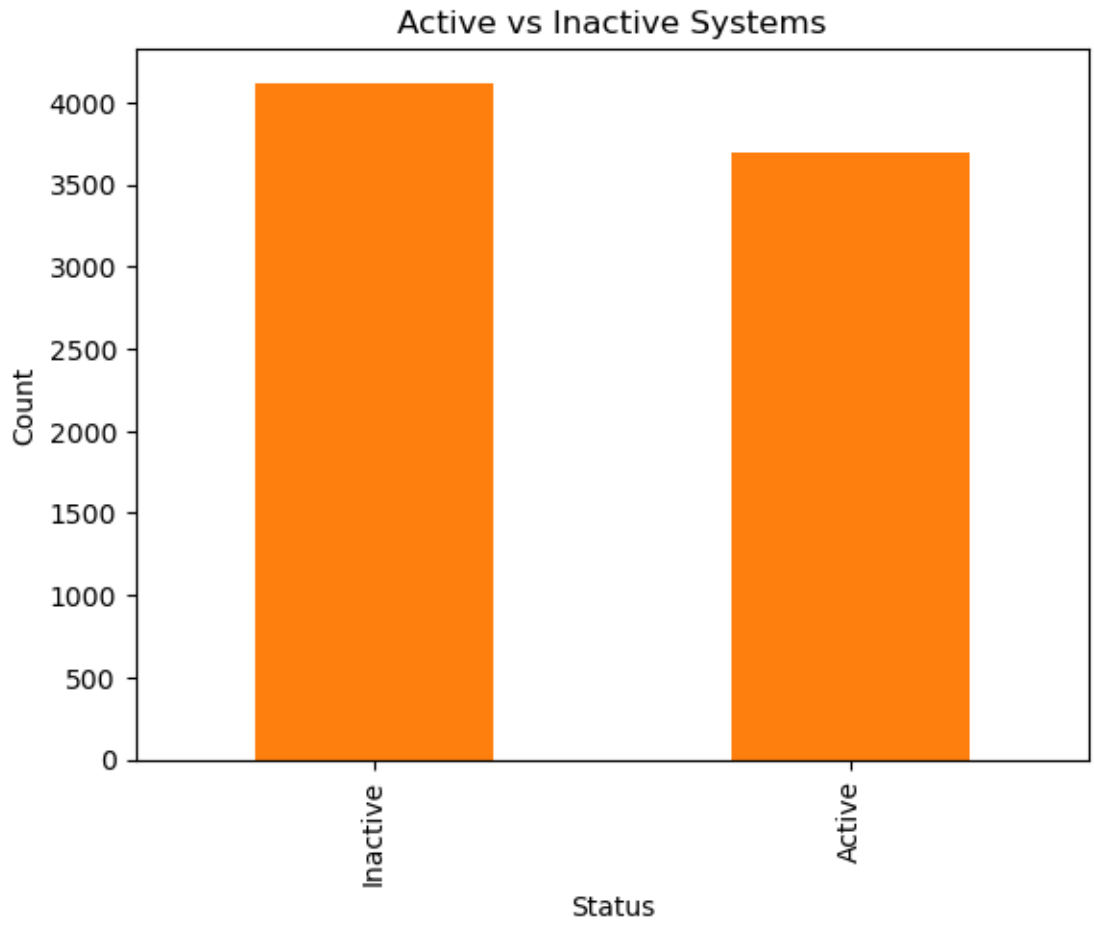

Figure 9 The graphical representation demonstrates the breakdown of active and inactive water systems through bars. The analysis shows that inactive systems outnumber active systems by more than 4,000, indicating that a high proportion of water facilities have ceased operations.

Water source Types by Region

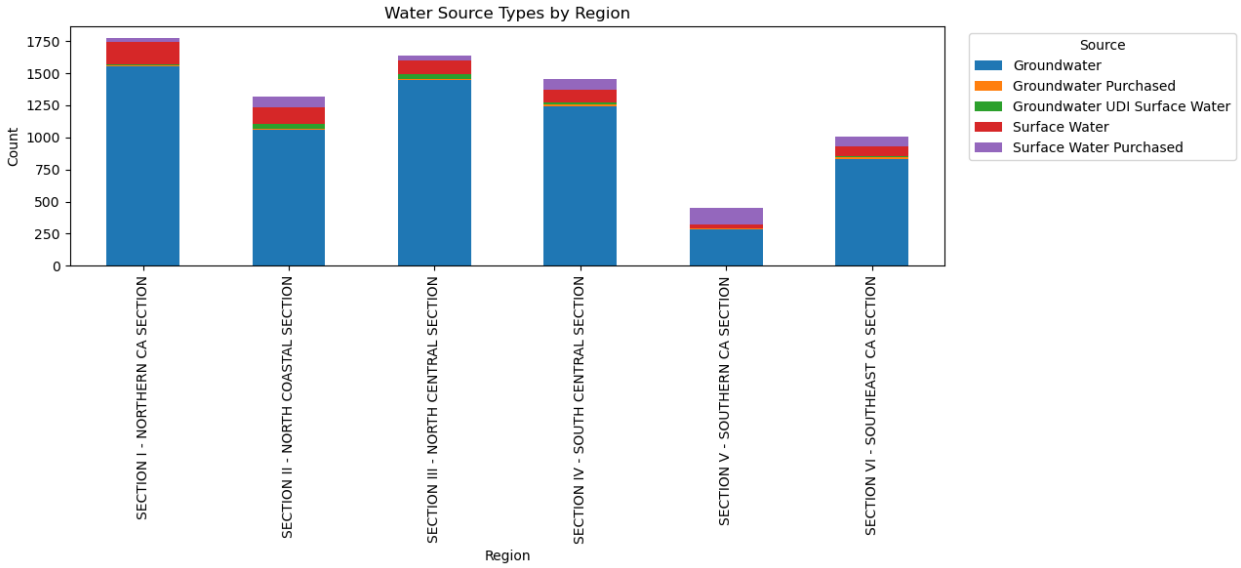

Figure 10 Every bar from the chart displays a different region where the water resources are categorized under Groundwater, Surface Water, and Groundwater Purchased, as well as additional types. The graph shows how each California region uses unique water sources.

Insight from Water source Types by Region from the stacked bar chart

The data displayed significant distinctions between the water resource distributions in California's different areas. Groundwater is the dominant water source throughout Section II - Northern CA Section because Northern Californian areas heavily depend on it. Still, they acquire smaller amounts from surface water and groundwater purchased. The Southern California regions, which comprise Section V - Southern CA Section, show an approximately equal proportion of Groundwater and Surface Water usage because they use imported surface water and local groundwater resources. Groundwater Purchased and Surface Water Purchased together comprise a significant part of water usage in the Southeast CA Section because of scarce water resources in the local area. Southern California's water resource management scenario stands out for its location-specific water challenges and strategic water management that affects water-deficient areas.

Discussion

The examination of California water systems data disclosed vital observations regarding both the spatial pattern of water delivery systems and the fundamental water sources and the economic differences between different regions in water availability and system development. These essential discoveries reveal the actual water delivery situation throughout the state and the problems faced by population groups in respective areas.

The Regional Disparities in Water Access research establishes a remarkable disparity between the water systems of Southern California and those of Northern California. With its large urban populations and agricultural demands, Southern California relies heavily on Groundwater and Surface Water. This part of the region maintains water networks serving millions of people while supplying water for different agricultural operations and commercial activities. The diverse water system in this region creates multiple challenges for Southern Californians, including water shortages, pollution issues, and outdated facilities, which affect Los Angeles and the Central Valley the most. The water management systems of Northern California exist to serve reduced populations and struggle with equal distribution of resources. Surface Water provides the primary water supply for these systems, although such sources are generally abundant yet unevenly situated, particularly within rural and agricultural regions.

The total population data presented in the histogram demonstrates an unbalanced distribution of water system usage requirements. The distribution of population numbers between water systems indicates significant variations since most serve local populations, while extensive facilities serve millions. Total population data reveals through the Box Plot analysis that Community Systems maintain the most significant number of residents, while Transient Non-Community Systems serve lower numbers of temporary inhabitants. The water system requirements of dense urban centers

surpass basic infrastructure requirements as many temporary population areas need simpler water management infrastructure.

The Water Source Types by Region chart demonstrates that different California regions consume water from separate sources. The water usage in Northern California relies heavily on Groundwater, while Southern California and Southeastern California use Surface Water and Groundwater Purchased (purchased Groundwater). The diverse water accessibility throughout California leads certain regions to obtain water from neighboring areas to serve their population needs. The state support for Groundwater remains substantial, but water consumption from surface resources and purchases is of higher importance in drought-prone areas. The vulnerability of Groundwater and the critical importance of sustainable water source management stand out because of the threats to both overuse and contamination.

The Sanitary Survey Visits per Year plot reveals that survey activity experienced considerable growth in 2020 as the authorities elevated their regulatory activities and water quality measures. Survey operations increased significantly since authorities focused more on water safety inspections, especially in rural and agricultural areas contaminated by nitrates and arsenic. The research demonstrates the necessity of continuous water system monitoring and inspection because it ensures drinking water safety for all Californians.

A significant portion of water systems appears as Inactive based on the Active vs. Inactive Systems chart. In contrast, reasons such as regulatory issues and system closures, as well as insufficient maintenance, contribute to this phenomenon. The substantial number of inactive water systems compared to active systems brings concerns about accessing water services for particular regions, primarily in rural locations.

This exploratory research confirms that Californian regions need special approaches to resolve their distinct problems with water access. Some parts of the state consume primarily Groundwater resources, but others satisfy their water demands with imported Surface Water supplies. The state needs to center its initiatives on water quality management, infrastructure modernization, and system operations management through regulatory enforcement to promote equal water services nationwide.

Conclusion

The investigation this study evaluated disparities within California's public water supply networks that affect disadvantaged communities while assessing their exposure to geographic as well as socioeconomic and regulatory factors. A wide data set from California Drinking Water Watch showed that the state exhibits distinct differences in how residents gain access to water and how well water distribution networks function between various districts.

The Analytical Examination of Data showed that small public water systems form the demographic majority but giant community systems supply services to highly populated city districts. Groundwater constitutes the main water source for North California water systems although communities within the states' southern and southeastern regions depend on both groundwater and surface water and imported water supplies. The high number of inactive water systems indicates urgent challenges regarding infrastructure together with maintenance needs that affect primarily rural and economically underprivileged regions.

The predictive modeling system received added strength from feature engineering processes that added variables about connections per capita and transient-to-residential population ratios. Such predictions allow us to find both upcoming population-related system overloads and delayed sanitary inspections. Systematic disparities affect Latino and African American communities together with low-income communities because water supply and distribution systems demonstrate significant regional disparities.

The research data demonstrates that California needs to develop specific reforms which target water management systems. To address the situation strategies need to enhance vulnerable region infrastructure while increasing regulatory controls plus supporting water sustainability through recycled and desalination programs. Predictive analytics and other data-driven tools should remain

in active use because they assess future water challenges and discover ways to ensure equal water distribution.

To make water access universal in California requires governmental cooperation between policy development and technological convergence while providing permanent support to disadvantaged communities who need safe water service above all else.

References

Acquah, S., & Allaire, M. (2023). Disparities in drinking water quality: evidence from California. Water Policy, 25(2). https://doi.org/10.2166/wp.2023.068

Agustí Pérez-Foguet. (2023). Broadening the water affordability approach to monitor the human right to water. Cities, 143, 104573–104573. https://doi.org/10.1016/j.cities.2023.104573

Archer, H., Morello Frosch, R., Pace, C., Baehner, L., & Cushing, L. (2024, August). Drinking water arsenic contamination and COVID-19 outcomes in California, USA. In ISEE Conference Abstracts (Vol. 2024, No. 1). https://ehp.niehs.nih.gov/doi/abs/10.1289/isee.2024.0141\

Carchi, D., Orellana, M., Martínez, A., & Segovia, J. (2023). Affordability and sustainability in the human right to water. Journal of Agribusiness in Developing and Emerging Economies. https://doi.org/10.1108/jadee-06-2023-0151

Chen, H., & Franklin, M. (2023, November 21). Spatio-Temporal Modeling of Surface Water Quality Distribution in California (1956-2023). ArXiv.org. https://doi.org/10.48550/arXiv.2311.12736

Fernandez-Bou, A. S., Rodríguez-Flores, J. M., Guzman, A., Ortiz-Partida, J. P., Classen-Rodriguez, L. M., Sánchez-Pérez, P. A., ... & Medellín-Azuara, J. (2023). Water, environment, and socioeconomic justice in California: A multi-benefit cropland repurposing framework. Science of the Total Environment, 858, 159963. https://www.sciencedirect.com/science/article/pii/S0048969722070632

Goddard, J. J., Ray, I., & Balazs, C. (2021). Water affordability and human right to water implications in California. PLOS ONE, 16(1), e0245237. https://doi.org/10.1371/journal.pone.0245237

Hanak, E., & Chappelle, C. (2025). California's Water Quality Challenges - Public Policy Institute of California. Public Policy Institute of California. https://www.ppic.org/publication/californias-water-quality-challenges/

London, J. K., Fencl, A. L., Watterson, S., Choueiri, Y., Seaton, P., Jarin, J., ... & Bailey, C. (2021). Disadvantaged unincorporated communities and the struggle for water justice in California. Water Alternatives, 14(2), 520-545. https://www.water-alternatives.org/index.php/alldoc/articles/vol14/v14issue2/626-a14-2-4/file

Siirila-Woodburn, E. R., Rhoades, A. M., Hatchett, B. J., Huning, L. S., Szinai, J., Tague, C., ... & Kaatz, L. (2021). A low-to-no snow future and its impacts on water resources in the western United States. Nature Reviews Earth & Environment, 2(11), 800-819. https://www.nature.com/articles/s43017-021-00219-y

Xenarios, S., Edwards, E. Y., & Buurman, J. (2025). Water Tariffs and Affordability in Urban Water Supply and Wastewater Systems. Water Economics and Policy. https://doi.org/10.1142/s2382624x24500164