Urban Water Management

Introduction

- Water is one of the most essential resources in a person's daily life. Even though **70%** of the earth is covered by water, only **3 %** of the total resource is potable i.e readily available for human consumption.
- Weather extremes brought on by climate change and increasing population have accelerated the rate at which water level is depleting.

Why it matters ?

- Studies suggest by 2025, **half** of the world's population could be living in areas facing water scarcity, and **two-thirds** of the world's population may face water shortage for at least one month every year.
- **2022** was the **third** consecutive year of drought for California.
- **9.7** Million people in California are affected by drought.

California's Reservoirs

- Weeks of heavy snow and rainstorms(**147%** of average rainfall), have essentially ended the state's three-year drought.
- Just **9%** of California is still experiencing "severe drought" conditions, down from almost **33%** a month ago.
- But all is not good since the groundwater has been depleted after years of mismanagement and abuse to a level which even the record breaking rains can't reverse.

Project Plan

- Tackling the problem requires a two pronged approach. We need to look at both the demand and the supply to effectively devise a solution.
- The supply side helps in monitoring the water levels and the intensity of the drought while data points from demand will assist in building a forecasting model.

Data Collection
Demand and Supply of water in California. Data is collected from open source sets released by government bodies and web scraping.

Data Preprocessing
Cleaning data by handling outliers, missing data points and removing inconsistencies. Feature selection by picking variables that best describe the water demand and supply.

Data Visualization and EDA
Ad Hoc queries for extracting insights. Visualizing the data in a meaningful way, using geospatial and graphical data to analyze and extract any insights present in the dataset like how fast water recharges in the neighboring areas.

Predictive Modeling
A model to determine which variables will impact the risk of water shortage the most, and predict the possibility of drought based on new inputs.

Feature selection

- **Explored** all the datasets to look for underlying patterns in the data, potential queries for analysis, and data issues like inconsistency, missing values and outliers.
- **Feature Selection :** After exploring the dataset we picked the following variable for supply : unimpaired flow, current groundwater levels, precipitation, snowpack in the mountain, recycled water and non-revenue water.
- For the demand side, we would be choose residential per-capita water use, water change by county and water sales. We believe these variables are most appropriate for our model since they have **up-to-date** information and tackle **different aspects** of water in California.

Metadata for supplier data

Field Name	Data Type	Description	Optional/Mandatory
Supplier Name	Plain Text	Name of the urban water supplier	Mandatory
Public Water System ID	Plain Text	ID number(s) associated with the water system(s) in the supplier service area	Mandatory
Reporting Month	Date	The pertaining month of the Monitoring Report	Mandatory
County	Plain Text	County/counties encompassing the supplier service area	Mandatory
Hydrologic Region	Plain Text	Name of the principal hydrologic region encompassing the supplier's service area	N/A (Prefilled)
Climate Zone	Plain Text	Name of the principal climate zone encompassing the supplier's service area	N/A (Prefilled)
Total Population Served	Number	Estimate of the number of permanent residents served potable water during the reporting month	Mandatory

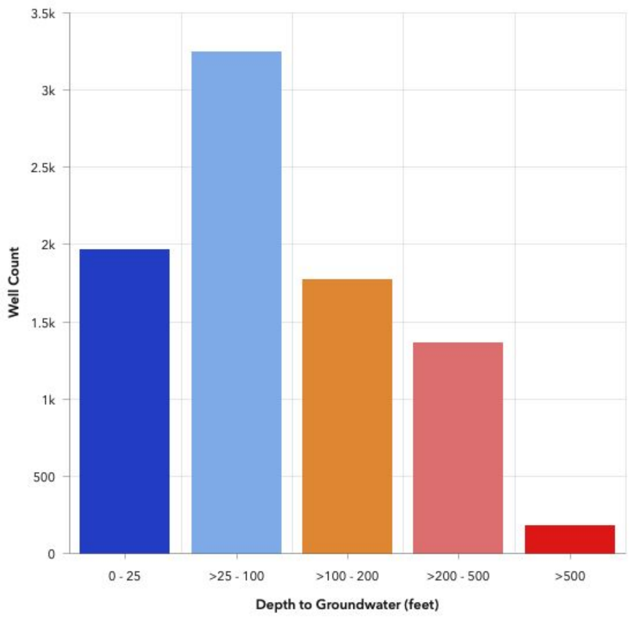
Data Cleaning

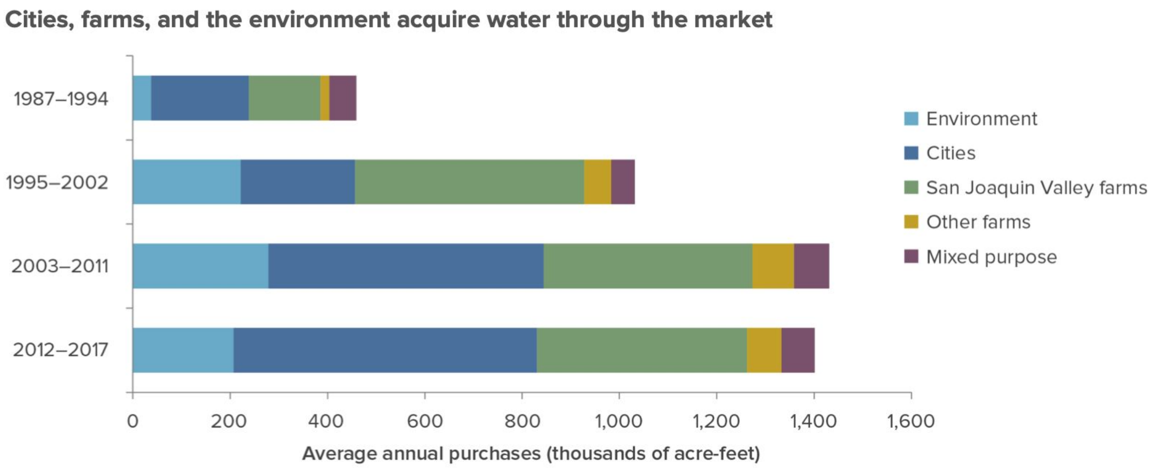
- Removed rows where data in more than **40% of the columns** was missing.
- Removed outliers which may the potential to introduce bias in our predictive model and analysis for the various datasets selected. Kept other outliers which seemed reasonable considering the weather conditions and other variables.
- Made data consistent across the different datasets

supplier_name	public_wate	reporting_month	county	hydrologic_region	climate_zone	total_population_served	r
East Bay Municipal Utilities District	CA0110005	1/15/2023	Alameda,Contra Costa	San Francisco Bay	3	1440000	
East Bay Municipal Utilities District	CA0110005	9/15/2022	Alameda,Contra Costa	San Francisco Bay	3	1430000	
East Bay Municipal Utilities District	CA0110005	8/15/2022	Alameda,Contra Costa	San Francisco Bay	3	1430000	
East Bay Municipal Utilities District	CA0110005	7/15/2022	Alameda,Contra Costa	San Francisco Bay	3	1430000	
East Bay Municipal Utilities District	CA0110005	6/15/2022	Alameda,Contra Costa	San Francisco Bay	3	1430000	
East Bay Municipal Utilities District	CA0110005	5/15/2022	Alameda,Contra Costa	San Francisco Bay	3	1430000	
East Bay Municipal Utilities District	CA0110005	4/15/2022	Alameda,Contra Costa	San Francisco Bay	3	1430000	
East Bay Municipal Utilities District	CA0110005	3/15/2022	Alameda,Contra Costa	San Francisco Bay	3	1430000	
East Bay Municipal Utilities District	CA0110005	2/15/2022	Alameda,Contra Costa	San Francisco Bay	3	1430000	
East Bay Municipal Utilities District	CA0110005	1/15/2022	Alameda,Contra Costa	San Francisco Bay	3	1420000	
East Bay Municipal Utilities District	CA0110005	12/15/2021	Alameda,Contra Costa	San Francisco Bay	3	1420000	
East Bay Municipal Utilities District	CA0110005	11/15/2021	Alameda,Contra Costa	San Francisco Bay	3	1420000	
East Bay Municipal Utilities District	CA0110005	10/15/2021	Alameda,Contra Costa	San Francisco Bay	3	1420000	
East Bay Municipal Utilities District	CA0110005	9/15/2021	Alameda,Contra Costa	San Francisco Bay	3	1420000	
East Bay Municipal Utilities District	CA0110005	8/15/2021	Alameda,Contra Costa	San Francisco Bay	3	1420000	
East Bay Municipal Utilities District	CA0110005	7/15/2021	Alameda,Contra Costa	San Francisco Bay	3	1420000	
East Bay Municipal Utilities District	CA0110005	6/15/2021	Alameda,Contra Costa	San Francisco Bay	3	1420000	
East Bay Municipal Utilities District	CA0110005	5/15/2021	Alameda,Contra Costa	San Francisco Bay	3	1420000	

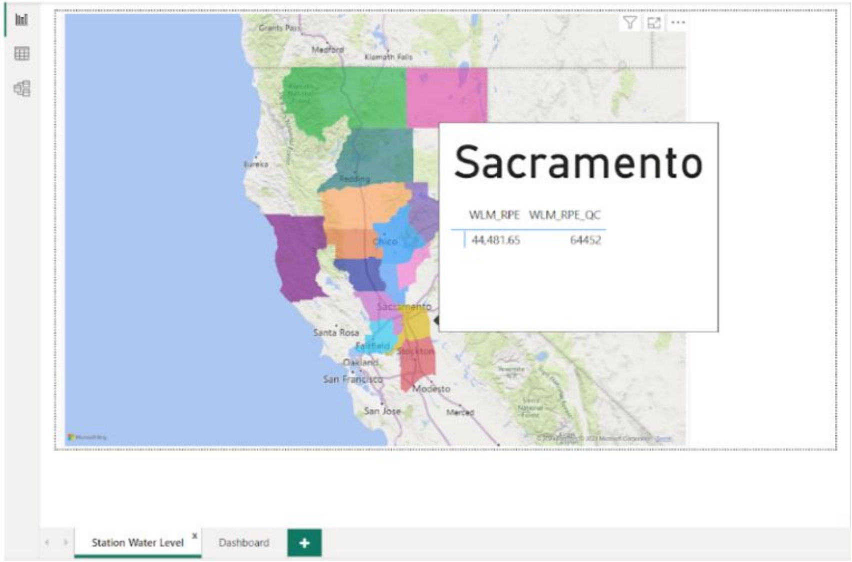
- Using the selected variables allows us to see how the water supplies are **allocated** throughout the year and how the dynamics between different entities work.
- Most of the surface water is **accumulated** in the North so we are looking at how those supplies are being transported to the rest of the state. There will be **leakage** and **inefficiencies**, which are represented by non-revenue water, so we want to see how much that impacts the supplies downstream.

Intermediate Results/Insights





- **Demand Forecasting :** Focusing on **5 main** watersheds that are the most influential predict the changes for the near future (next 3-5 years).
- **Geospatial Analysis :** Dashboard that summarizes the high level metrics and real time monitoring of these metrics.



Recommendations

- Through our analysis we were able to determine the **top water-use categories** by LA county for the public supply, irrigation etc. and then the top usage within each category.
- We observed that the storage level drops over the years although the residential demand was relatively stable. This means that the **depletion** of important water reservoirs comes from irrigation and commercial usage, **not residential**. Thus, although it is recommended for people to save water, it is imperative for farms to use water more efficiently.
- Based on our model we can alert at-risk areas to make **timely intervention** byt the local water agencies.
- Lastly, using the dashboard allows for ad hoc analysis as well as summarized high level metrics for quick analysis. These reproducible visualizations are efficient as they allow us to respond quickly to current events, like storms or sudden rain in May which LA, and the neighboring counties are facing with only minor modifications of the code.

Difficulties in forecasting

- Lastly, we would like to consider the weather conditions. This is by far the most **difficult** part since the forecasts are often short and **not highly accurate** (80% accuracy for a 7 day forecast). Additionally, its granularity and frequency do not match with the slow changing data from the water board.
- Nevertheless, we believe it is still important to incorporate the weather into our analysis since climate change plays a big factor in replenishing the watersheds.
- A combination of both short term and long term forecasts will provide maximum value but the **feasibility** of this model is **still being considered**.

REFERENCES

- Bakker, M., Van Duist, H., Van Schagen, K., Vreeburg, J., & Rietveld, L. (2014). Improving the performance of water demand forecasting models by using weather input. Procedia Engineering, 70, 93-102.
- Alvisi, S., Franchini, M., & Marinelli, A. (2007). A short-term, pattern-based model for water-demand forecasting. *Journal of hydroinformatics*, *9*(1), 39-50.
- Tracking the California drought
- Water shortage vulnerability dashboard
- UNICEF water scarcity report
- Alternate water resources for irrigation