***Mapping the Depths: A Statistical Analysis of Groundwater Overuse and Depletion across India***

-By Eric Siqueira (23070126041)

FY AIML-A2 (2023-2027)

1. Introduction:

Since the time of the Indus Valley Civilisation, India, has been blessed with vast agricultural landscapes and a dense population. A complex system of irrigation canals, damns and wells were developed across the valley to facilitate the heavy population demand. Even today, India relies heavily on the same ground and surface water resources as a primary source of irrigation, industry and drinking water.

Groundwater plays the most pivotal role in sustaining livelihoods, rural and urban, ensuring food security, and meeting the basic needs of a Billion Indians. However, the sustainability of groundwater resources in India, particularly in rural and agrarian areas, has become a pressing concern with the modern-day problems of over-extraction, pollution, and erratic rainfall patterns exacerbated by climate change. Moreover, indiscriminate use of fertilizers and pesticides has further contributed to groundwater contamination, posing health risks to communities reliant on these sources for drinking water.

Some Relevant Statistics

* According to the Central Ground Water Board (CGWB), groundwater accounts for nearly 65-70% of the total irrigated area in the country.
* Groundwater also serves as a primary source of drinking water for rural households, meeting approximately 85% of their domestic water needs.
* According to the CGWB, over 20% of India's groundwater blocks are in a critical state, indicating a high rate of depletion.
* A World Bank report states that over 60% of India's districts are at risk of reaching critical groundwater depletion levels within the next two decades, indicating extraction happening much faster than replenishment.

To address these challenges, data-driven approaches are essential for understanding groundwater dynamics and formulating effective management strategies. By leveraging data analytics and statistical insights, local and state governments can adequately identify regions facing acute groundwater stress, quantify water availability, and develop sustainable groundwater management plans. Government Bodies such as the Central Groundwater Board (CGWB) and initiatives such as the National Aquifer Mapping and Management Programme (NAQUIM) aim to assess groundwater resources' quality and quantity, facilitating informed decision-making and public awareness.

The aim is to create as comprehensive and robust data analysis and statistics projects on groundwater resources and their depletion across India as possible. Helps provide valuable insights for policymakers, researchers, and local stakeholders.

1. Objectives:

* *Statistically Analyse the loss of groundwater and aquifer resources across Indian States.*
* *Find out the States and UTs of India facing the highest risks of groundwater overexploitation.*
* *Utilise case studies to develop possible future solutions to solve Groundwater overexploitation in drought-prone States and UTs.*

1. Methodology:

Step 1: Data Acquisition:

I utilised 3 datasets of National Compilation on Dynamic GroundWater Resources of India of 2017, 2020 and 2022. These are comprehensive annual publishings by the Central Ground Water Board, Ministry of Jal Shakti, India; that go into detail on the groundwater and aquifer health across every district in every State and Union Territory in India. I specifically used the Annexure of ‘District-wise groundwater resources availability, utilization and stage’ which is a tabular compilation of Groundwater Recharge Rate, Extraction Rate and Future Availability to compute the renewability of groundwater resources across India’s State and District levels.

Step 2:Data Preprocessing

Due to the 3 datasets initially being in pdf format, I was forced to manually input State and District data into a Microsoft Excel Sheet and import that into a common .csv format. Upon compiling all 3 datasets I ensured the column headings were the same across the 3 datasets and accounted for empty/missed data values.

Step 3: Code used for Exploratory Data Analysis

# Mounting Google Drive

from google.colab import drive

drive.mount('/content/drive')

# Install necessary libraries

!pip install pandas matplotlib seaborn

# Importing libraries

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

import numpy as np

# Load CSV files from Google Drive

csv\_2017 = pd.read\_csv('/content/drive/MyDrive/STAIML MiniProj/2017GW.csv')

csv\_2020 = pd.read\_csv('/content/drive/MyDrive/STAIML MiniProj/2020GW.csv')

csv\_2022 = pd.read\_csv('/content/drive/MyDrive/STAIML MiniProj/2022GW.csv')

# Combine the data into a single DataFrame

data = pd.concat([csv\_2017, csv\_2020, csv\_2022], ignore\_index=True)

# Add a 'Year' column

data['Year'] = pd.Series([2017] \* len(csv\_2017) + [2020] \* len(csv\_2020) + [2022] \* len(csv\_2022))

# Checking of each of three datasets

print("Shape of csv\_2017:", csv\_2017.shape)

print("Shape of csv\_2020:", csv\_2020.shape)

print("Shape of csv\_2022:", csv\_2022.shape)

# Check missing values for 2017 dataset

missing\_values\_2017 = csv\_2017.isna().sum()

print("Missing values in 2017 dataset:")

print(missing\_values\_2017)

# Check missing values for 2020 dataset

missing\_values\_2020 = csv\_2020.isna().sum()

print("\nMissing values in 2020 dataset:")

print(missing\_values\_2020)

# Check missing values for 2022 dataset

missing\_values\_2022 = csv\_2022.isna().sum()

print("\nMissing values in 2022 dataset:")

print(missing\_values\_2022)

def calculate\_groundwater\_statistics(df):

state\_list = []

total\_grnd\_water\_recharge = []

curr\_gw\_extr\_list = []

future\_available\_GW\_list = []

for state, subset in df.groupby('Name of State'):

state\_list.append(state)

total\_grnd\_water\_recharge.append(sum(subset['Total Annual Ground Water Recharge']))

curr\_gw\_extr\_list.append(sum(subset['Total Current Annual Ground Water Extraction']))

future\_available\_GW\_list.append(sum(subset['Net Ground Water Availability for future use']))

return pd.DataFrame({"State":state\_list, "GW\_Recharge":total\_grnd\_water\_recharge, "GW\_Extraction": curr\_gw\_extr\_list, "Future\_GW\_Available": future\_available\_GW\_list})

# Apply the function to each dataset

df\_2017\_stats = calculate\_groundwater\_statistics(csv\_2017)

df\_2020\_stats = calculate\_groundwater\_statistics(csv\_2020)

df\_2022\_stats = calculate\_groundwater\_statistics(csv\_2022)

# Print the statistics for 2017 dataset

print("Statistics for 2017 dataset:")

print(df\_2017\_stats)

# Print the statistics for 2020 dataset

print("\nStatistics for 2020 dataset:")

print(df\_2020\_stats)

# Print the statistics for 2022 dataset

print("\nStatistics for 2022 dataset:")

print(df\_2022\_stats)

# Visualizing groundwater dsitribution for 2017

plt.figure(figsize=(10, 6))

sns.histplot(data=csv\_2017, x='Total Annual Ground Water Recharge', bins=20, kde=True, color='blue')

plt.title('Groundwater Level Distribution - 2017')

plt.xlabel('Total Annual Ground Water Recharge [hectare meter (hm³)]')

plt.ylabel('Frequency')

plt.show()

# Visualizing groundwater dsitribution for 2020

plt.figure(figsize=(10, 6))

sns.histplot(data=csv\_2020, x='Total Annual Ground Water Recharge', bins=20, kde=True, color='green')

plt.title('Groundwater Level Distribution - 2020 ')

plt.xlabel('Total Annual Ground Water Recharge [hectare meter (hm³)]')

plt.ylabel('Frequency')

plt.show()

# Visualizing groundwater dsitribution for 2022

plt.figure(figsize=(10, 6))

sns.histplot(data=csv\_2022, x='Total Annual Ground Water Recharge', bins=20, kde=True, color='orange')

plt.title('Groundwater Level Distribution - 2022 ')

plt.xlabel('Total Annual Ground Water Recharge [hectare meter (hm³)]')

plt.ylabel('Frequency')

plt.show()

# Define function to plot heatmap for each year

def plot\_heatmap\_for\_year(df, year):

# Check if columns exist before dropping them

columns\_to\_drop = ['S.no.', 'Name of State', 'Name of District', 'State']

columns\_to\_drop = [col for col in columns\_to\_drop if col in df.columns] # Filter existing columns

if columns\_to\_drop:

# Drop specified columns

df = df.drop(columns\_to\_drop, axis=1)

# Convert non-numeric values to numeric

numeric\_df = df.apply(pd.to\_numeric, errors='coerce')

# Plot heatmap

plt.figure(figsize=(10, 7))

sns.heatmap(numeric\_df.corr(), annot=True)

plt.title(f'Heatmap of Groundwater Usage - {year}')

plt.show()

# Plot heatmap for each year

plot\_heatmap\_for\_year(df\_2017\_stats, 2017)

plot\_heatmap\_for\_year(df\_2020\_stats, 2020)

plot\_heatmap\_for\_year(df\_2022\_stats, 2022)

# Defining function to plot state-wise statistics to compare recharge rate v/s extraction v/s availability

def plot\_statewise\_statistics(df, year):

plt.figure(figsize=(12, 6))

plt.bar(df['State'], df['GW\_Recharge'], color='blue', label='Groundwater Recharge')

plt.bar(df['State'], df['GW\_Extraction'], color='red', label='Groundwater Extraction')

plt.bar(df['State'], df['Future\_GW\_Available'], color='green', label='Future Groundwater Availability')

plt.xlabel('State')

plt.ylabel('Volume [hectare meter (hm³)]')

plt.title(f'State-wise Groundwater Statistics - {year}')

plt.xticks(rotation=90)

plt.legend()

plt.tight\_layout()

plt.show()

# Plot state-wise statistics for 2017 dataset

plot\_statewise\_statistics(df\_2017\_stats, '2017')

# Plot state-wise statistics for 2020 dataset

plot\_statewise\_statistics(df\_2020\_stats, '2020')

# Plot state-wise statistics for 2022 dataset

plot\_statewise\_statistics(df\_2022\_stats, '2022')

# Calculate the mean ratio for each year

mean\_ratio\_2017 = df\_2017\_stats['GW\_Extraction'].sum() / df\_2017\_stats['GW\_Recharge'].sum()

mean\_ratio\_2020 = df\_2020\_stats['GW\_Extraction'].sum() / df\_2020\_stats['GW\_Recharge'].sum()

mean\_ratio\_2022 = df\_2022\_stats['GW\_Extraction'].sum() / df\_2022\_stats['GW\_Recharge'].sum()

# Print the mean ratios for each year

print("Mean Extraction-Recharge ratio for 2017:", mean\_ratio\_2017, " or, ",round(mean\_ratio\_2017\*100,2), " percent.")

print("Mean Extraction-Recharge ratio for 2020:", mean\_ratio\_2020, " or, ",round(mean\_ratio\_2020\*100,2), " percent.")

print("Mean Extraction-Recharge ratio for 2022:", mean\_ratio\_2022, " or, ",round(mean\_ratio\_2022\*100,2), " percent.")

# Function to calculate extraction-recharge ratio and plot top and bottom states

def calculate\_ratio\_and\_plot(df, year):

# Sort DataFrame by extraction-recharge ratio

sorted\_df = df.sort\_values(by=f'Extraction\_Recharge\_Ratio\_{year}', ascending=False)

# Get top and bottom 3 states

top\_states = sorted\_df.head(3)

bottom\_states = sorted\_df.tail(3)

# Plot top 3 states with highest extraction-recharge ratio

plt.figure(figsize=(12, 6))

plt.bar(top\_states['State'], top\_states[f'Extraction\_Recharge\_Ratio\_{year}'], color='green')

plt.title(f'Top 3 States with Highest Extraction-Recharge Ratio - {year}')

plt.xlabel('State')

plt.ylabel('Extraction-Recharge Ratio')

plt.xticks(rotation=45)

plt.tight\_layout()

plt.show()

# Plot bottom 3 states with lowest extraction-recharge ratio

plt.figure(figsize=(12, 6))

plt.bar(bottom\_states['State'], bottom\_states[f'Extraction\_Recharge\_Ratio\_{year}'], color='red')

plt.title(f'Top 3 States with Lowest Extraction-Recharge Ratio - {year}')

plt.xlabel('State')

plt.ylabel('Extraction-Recharge Ratio')

plt.xticks(rotation=45)

plt.tight\_layout()

plt.show()

# Calculate ratio and plot top and bottom states for each year

calculate\_ratio\_and\_plot(df\_2017\_stats, '2017')

calculate\_ratio\_and\_plot(df\_2020\_stats, '2020')

calculate\_ratio\_and\_plot(df\_2022\_stats, '2022')

def visualize\_top\_10\_recharge\_comparison(df, year):

# Select relevant columns for comparison

comparison\_df = df[['Name of State',

'Recharge from rainfall During Monsoon Season',

'Recharge from other sources During Monsoon Season',

'Recharge from rainfall During Non Monsoon Season',

'Recharge from other sources During Non Monsoon Season',

'Total Annual Ground Water Recharge']]

# Sort DataFrame by total annual groundwater recharge

comparison\_df.sort\_values(by='Total Annual Ground Water Recharge', ascending=False, inplace=True)

# Set 'Name of State' column as index

comparison\_df.set\_index('Name of State', inplace=True)

# Selecting and plotting top 10 states

top\_10\_df = comparison\_df.head(10)

plt.figure(figsize=(12, 8))

width = 0.15

x = np.arange(len(top\_10\_df.index))

# Bar plots for each source of recharge

for i, column in enumerate(top\_10\_df.columns[:-1]): # Exclude 'Total Annual Ground Water Recharge'

plt.bar(x + i\*width, top\_10\_df[column], width=width, label=column)

# Bar plot for total annual groundwater recharge

plt.bar(x + 4\*width, top\_10\_df['Total Annual Ground Water Recharge'], width=width, color='black', label='Total Annual Ground Water Recharge')

plt.title(f'10 Highest GW Recharge States - {year}')

plt.xlabel('State')

plt.ylabel('Volume (in hectare meters)')

plt.xticks(x + 2\*width, top\_10\_df.index, rotation=45, ha='right') # Adjust rotation angle

plt.legend()

plt.tight\_layout()

plt.show()

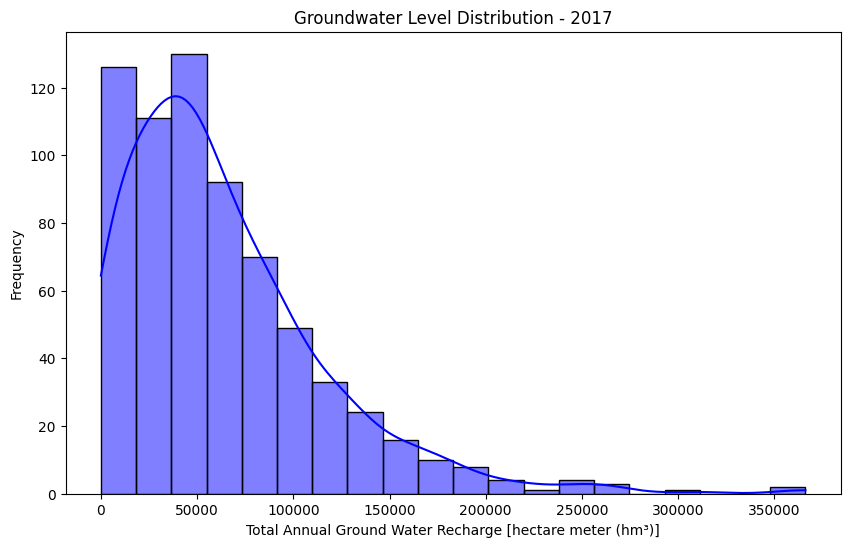
# Visualize top 10 recharge comparison for each dataset

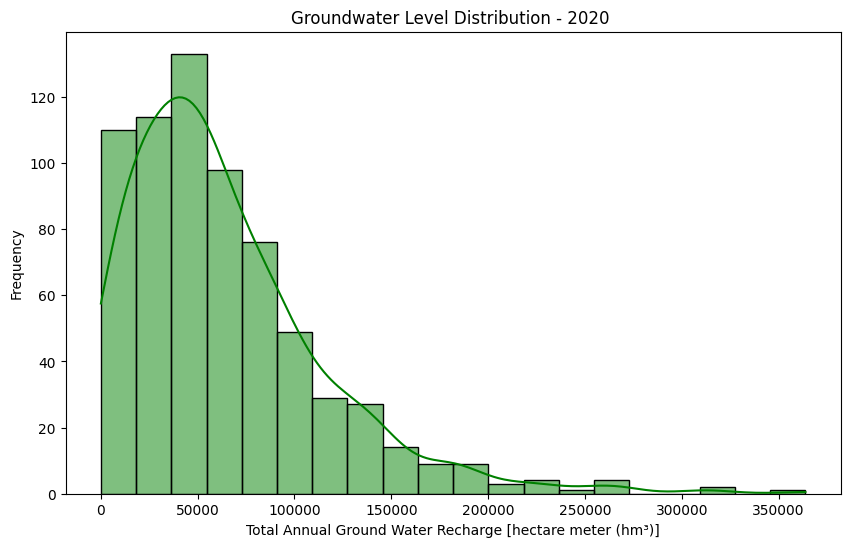
visualize\_top\_10\_recharge\_comparison(csv\_2017, '2017')

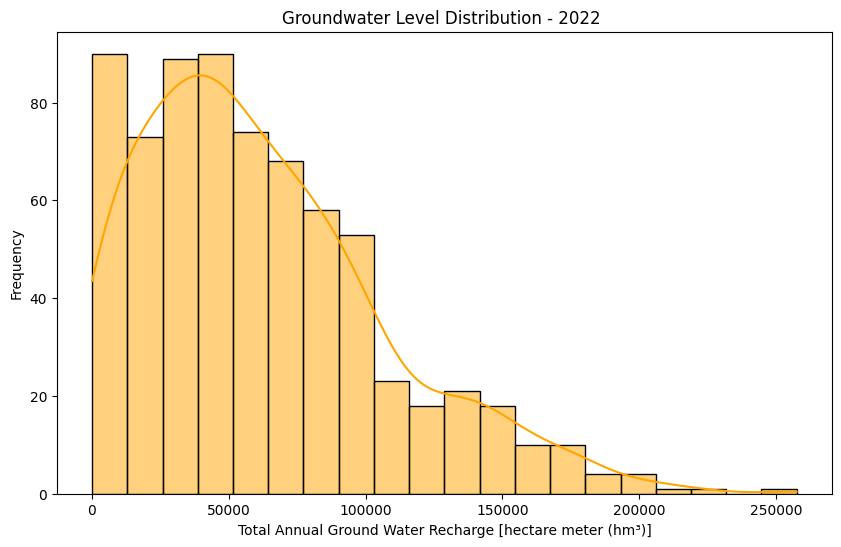
visualize\_top\_10\_recharge\_comparison(csv\_2020, '2020')

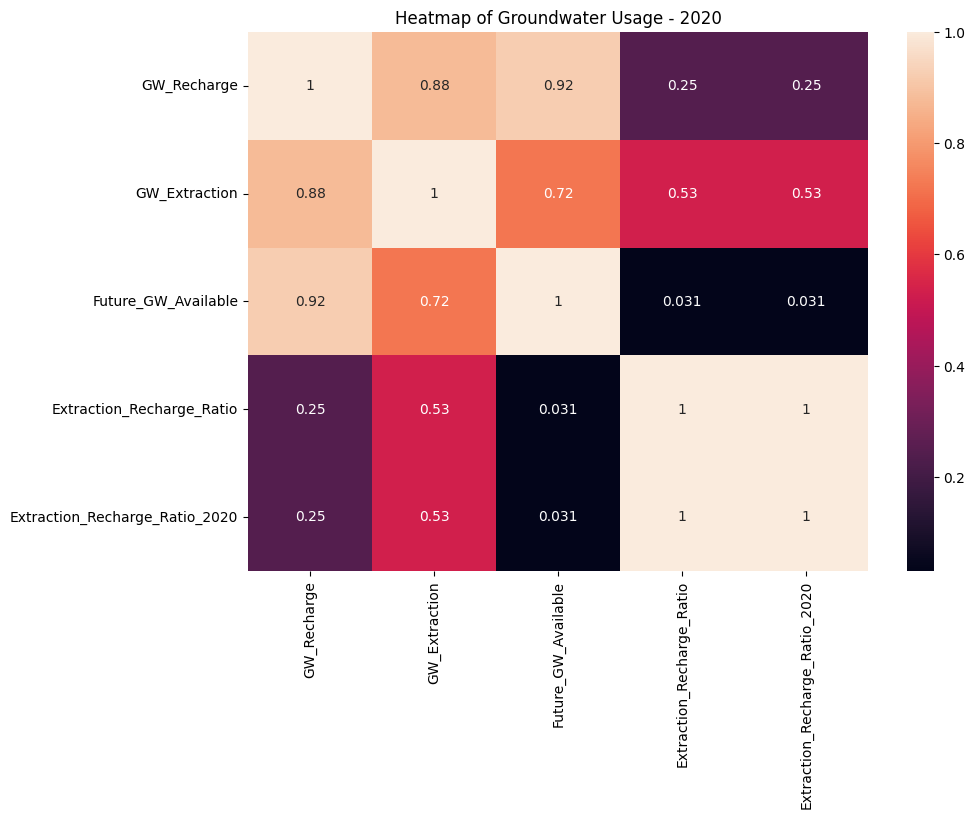
visualize\_top\_10\_recharge\_comparison(csv\_2022, '2022') Create visualizations to explore trends and patterns

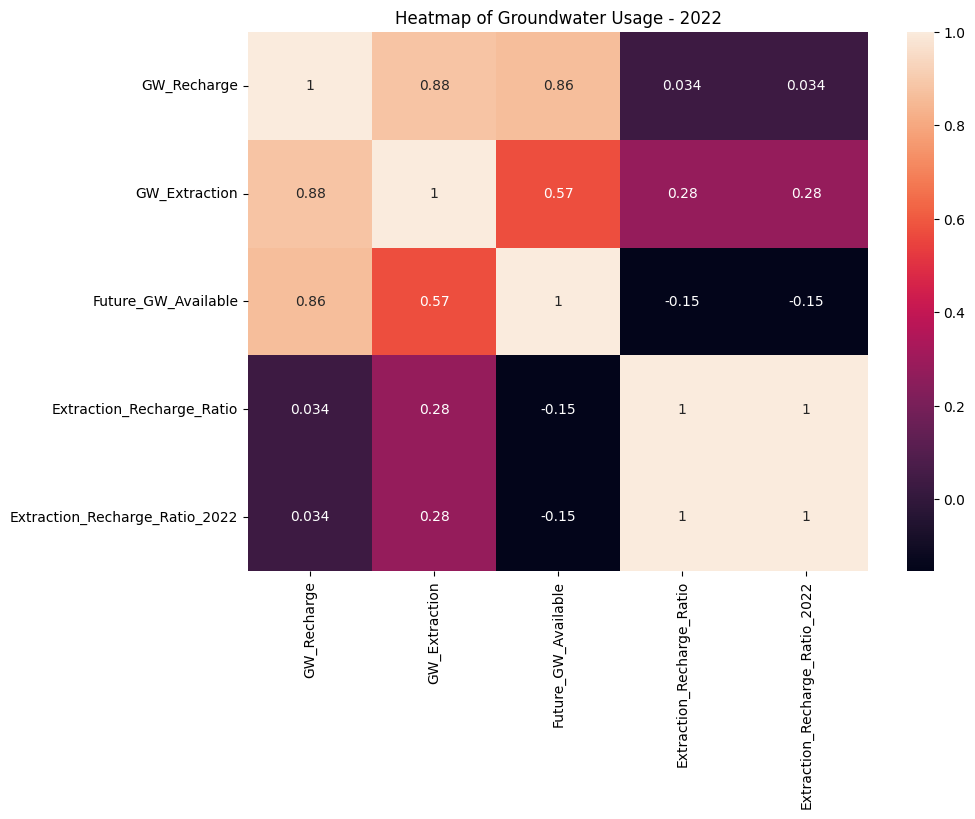
Step 4: Visualization and Reporting:

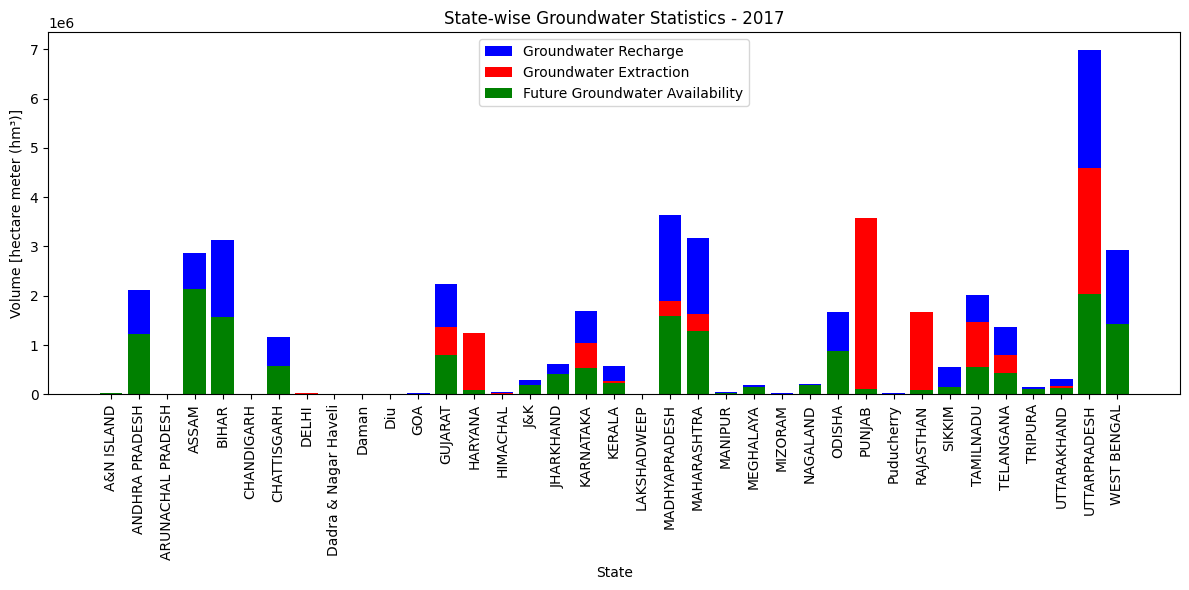


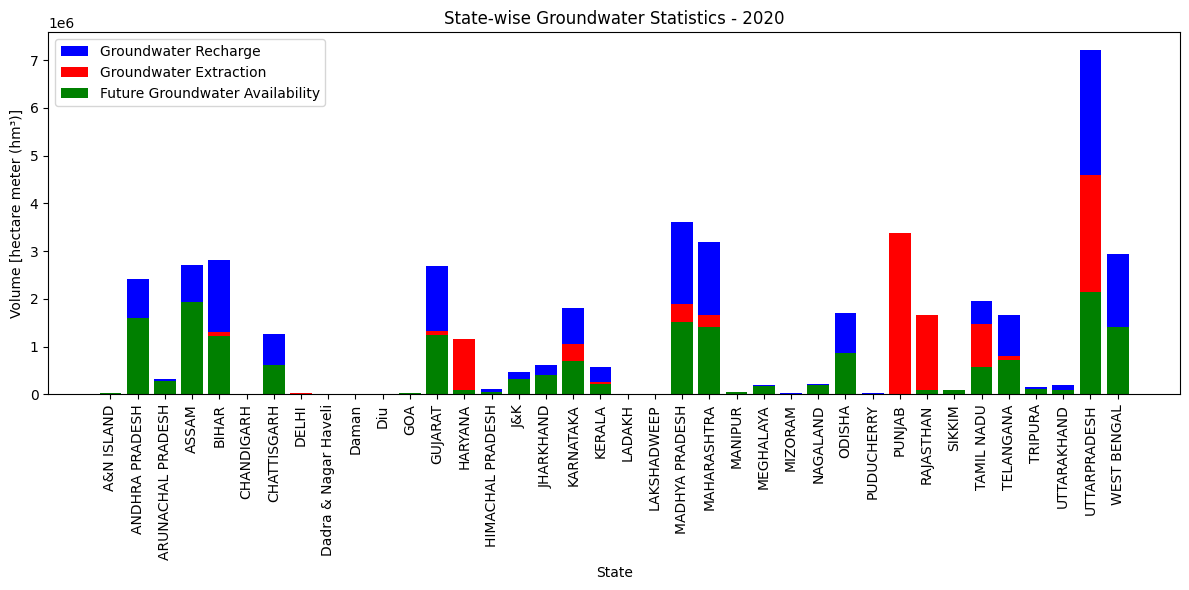


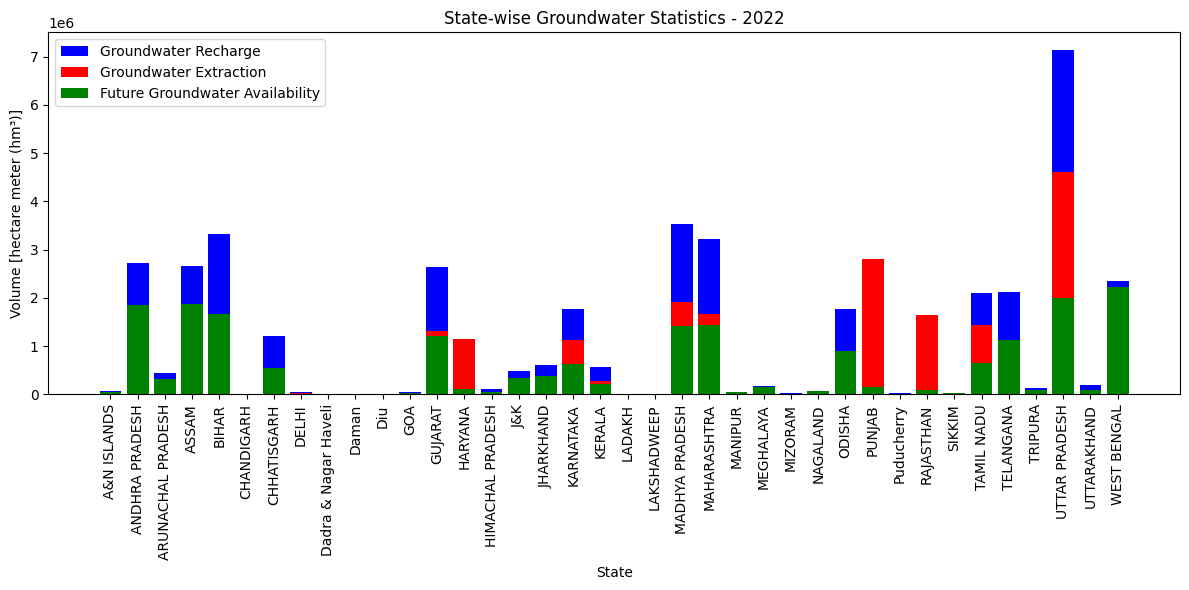


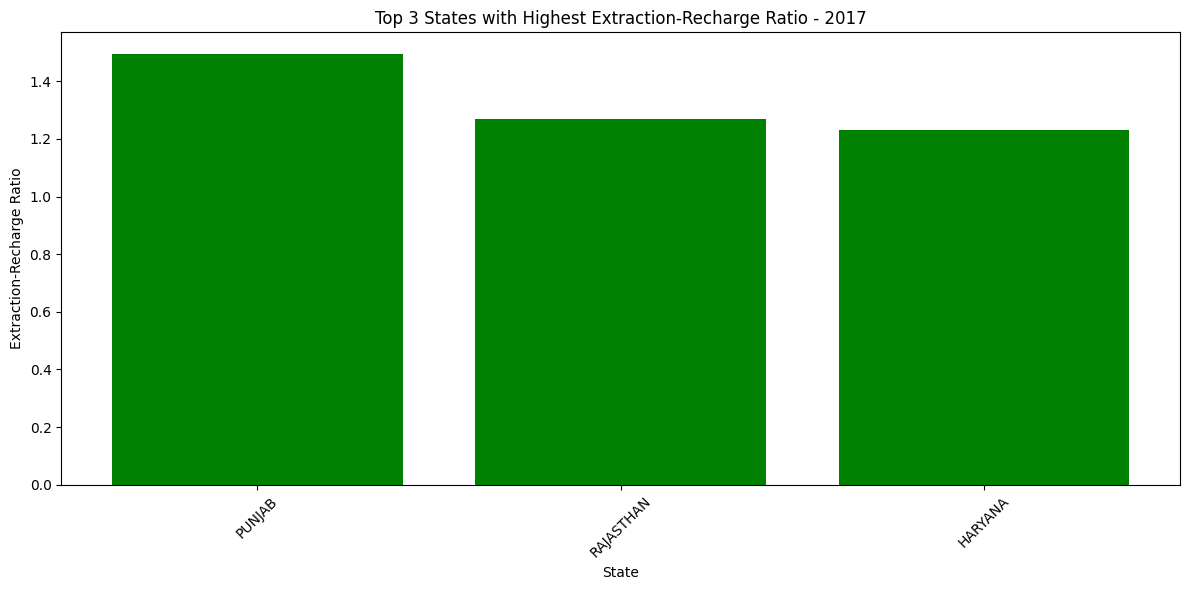
 

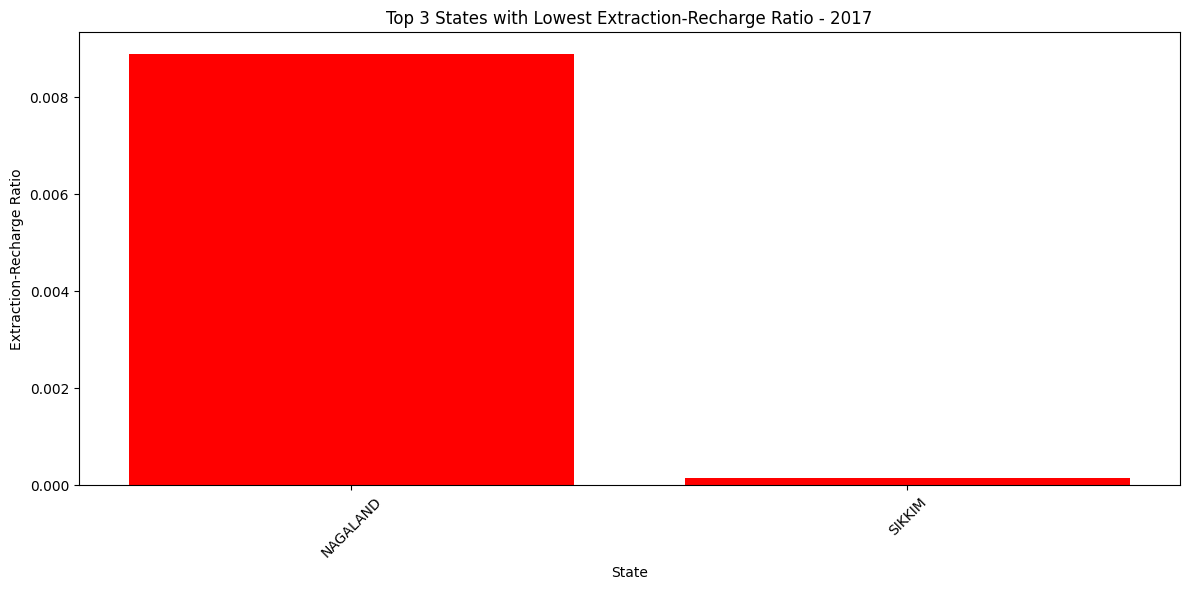


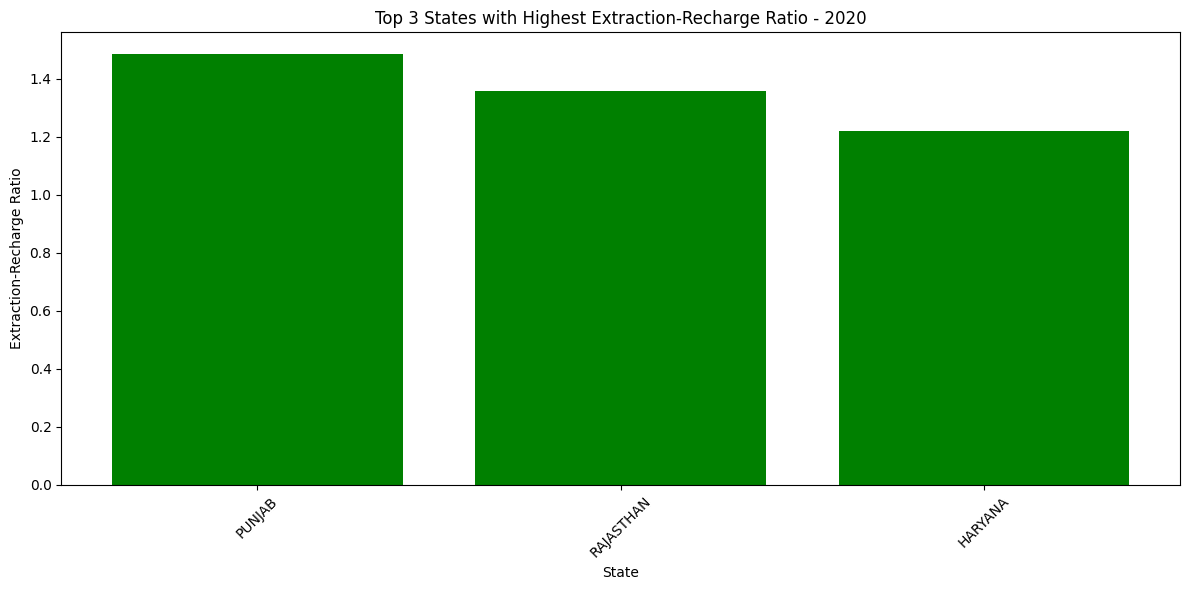


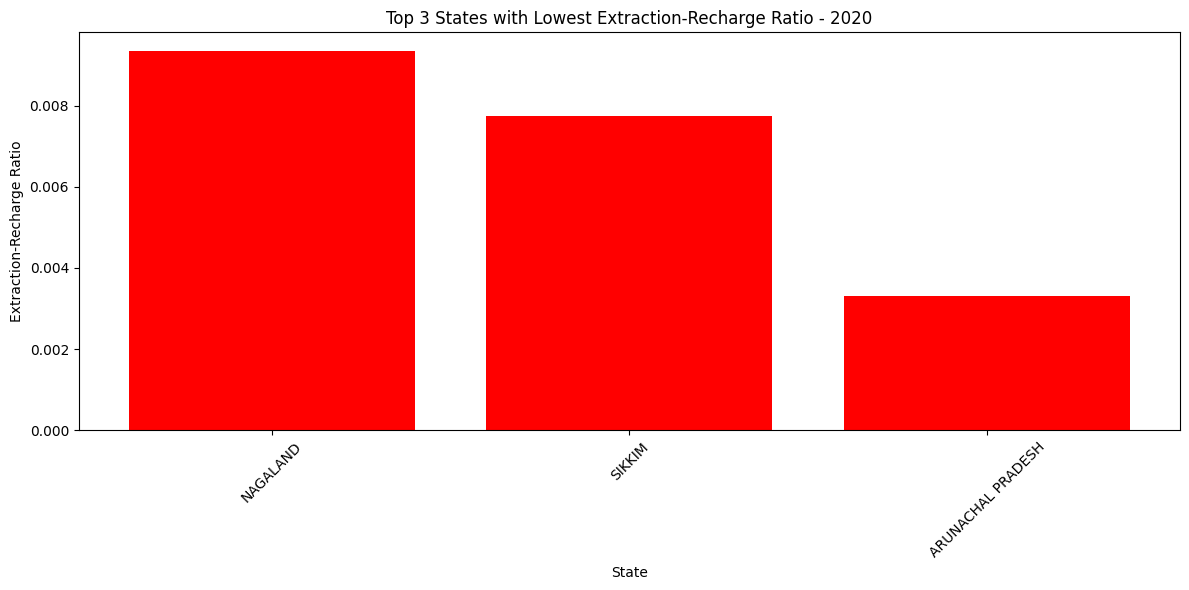
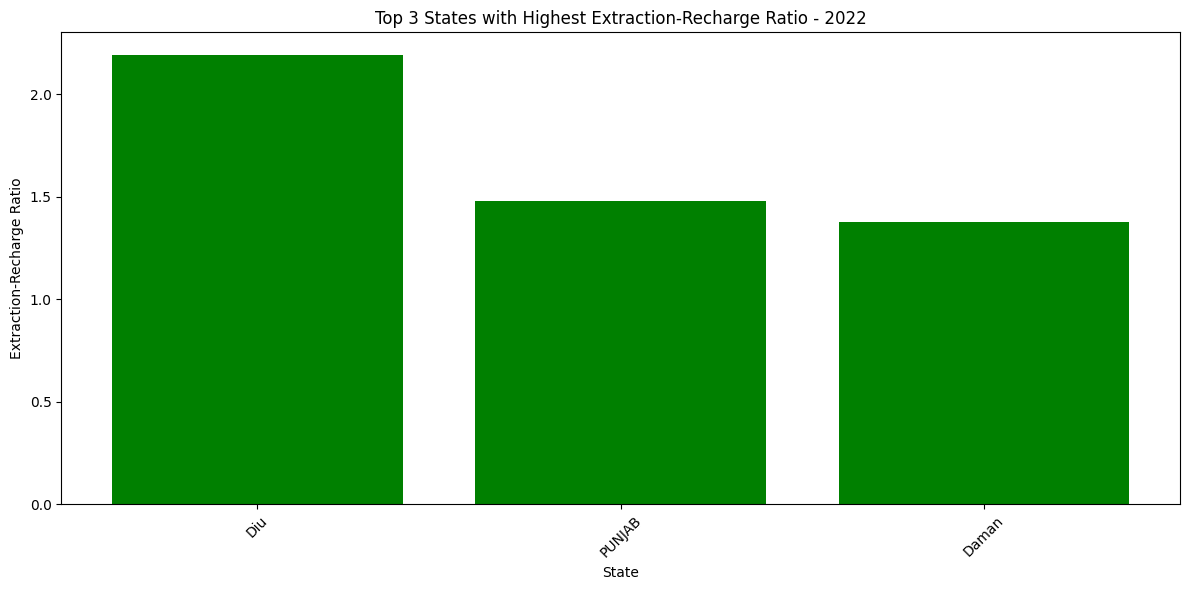


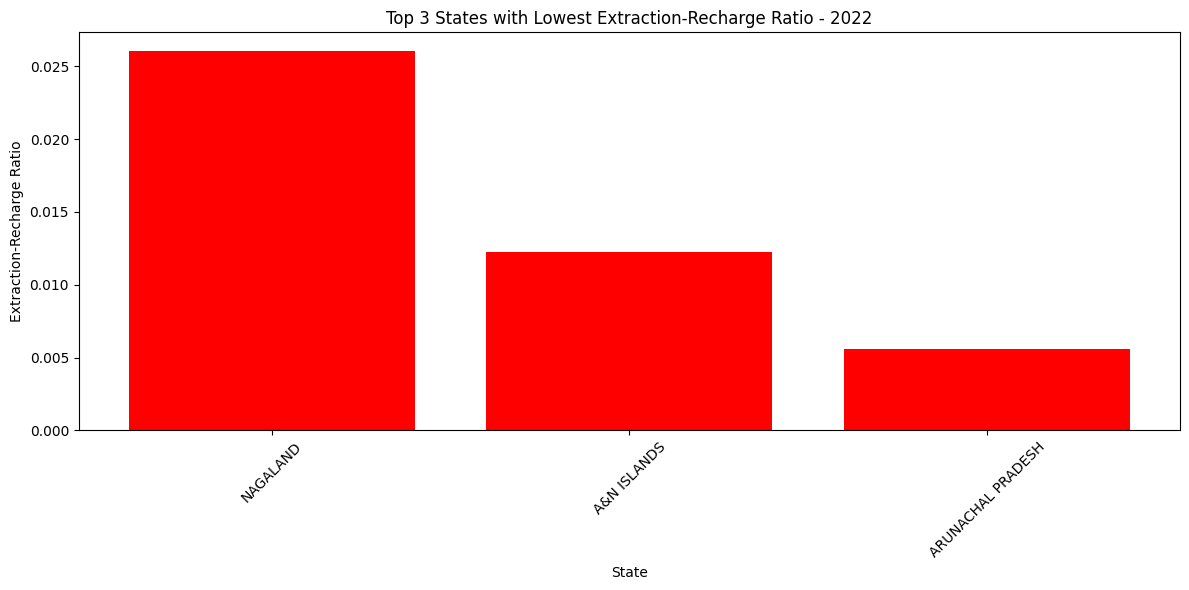


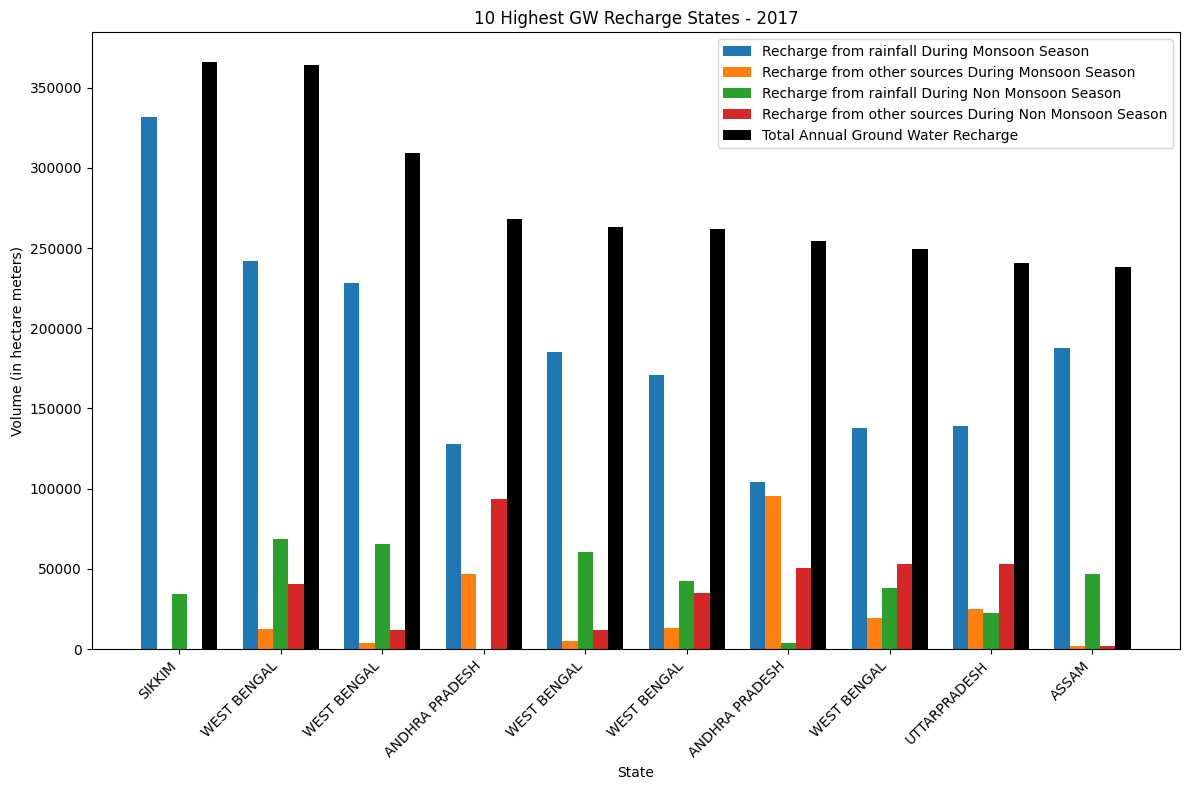


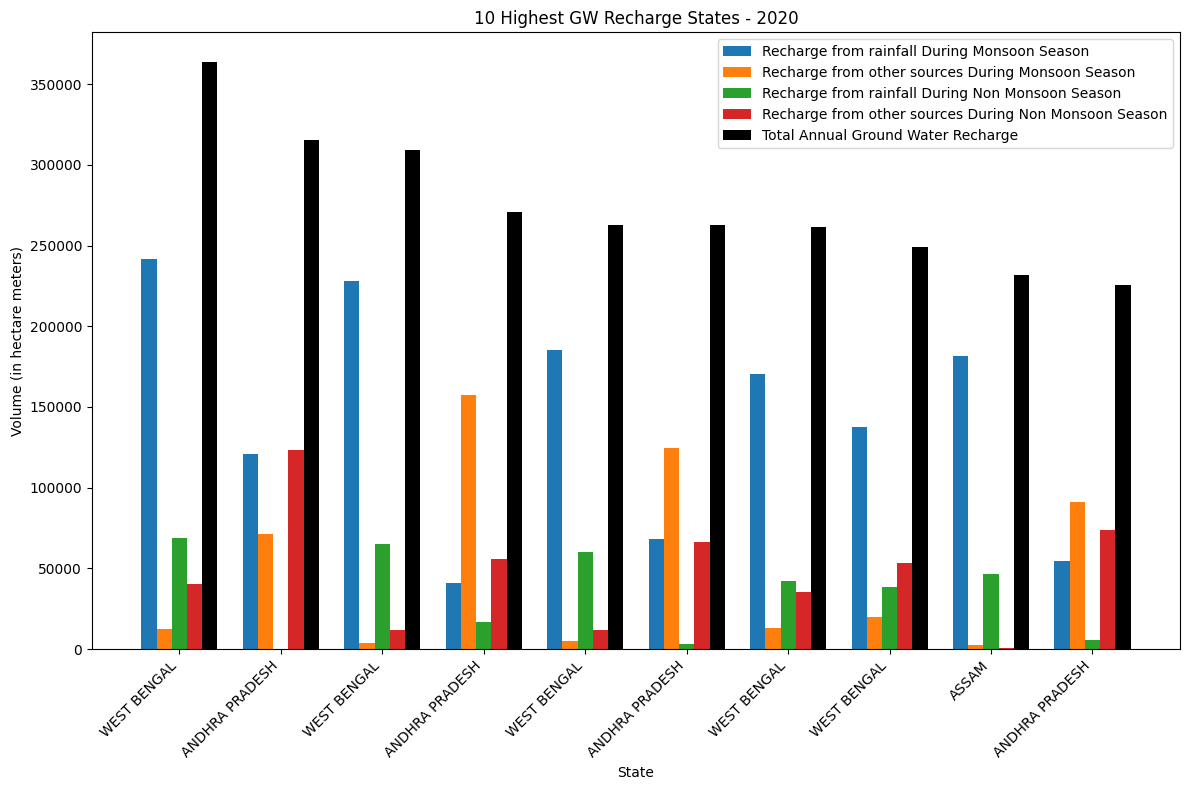


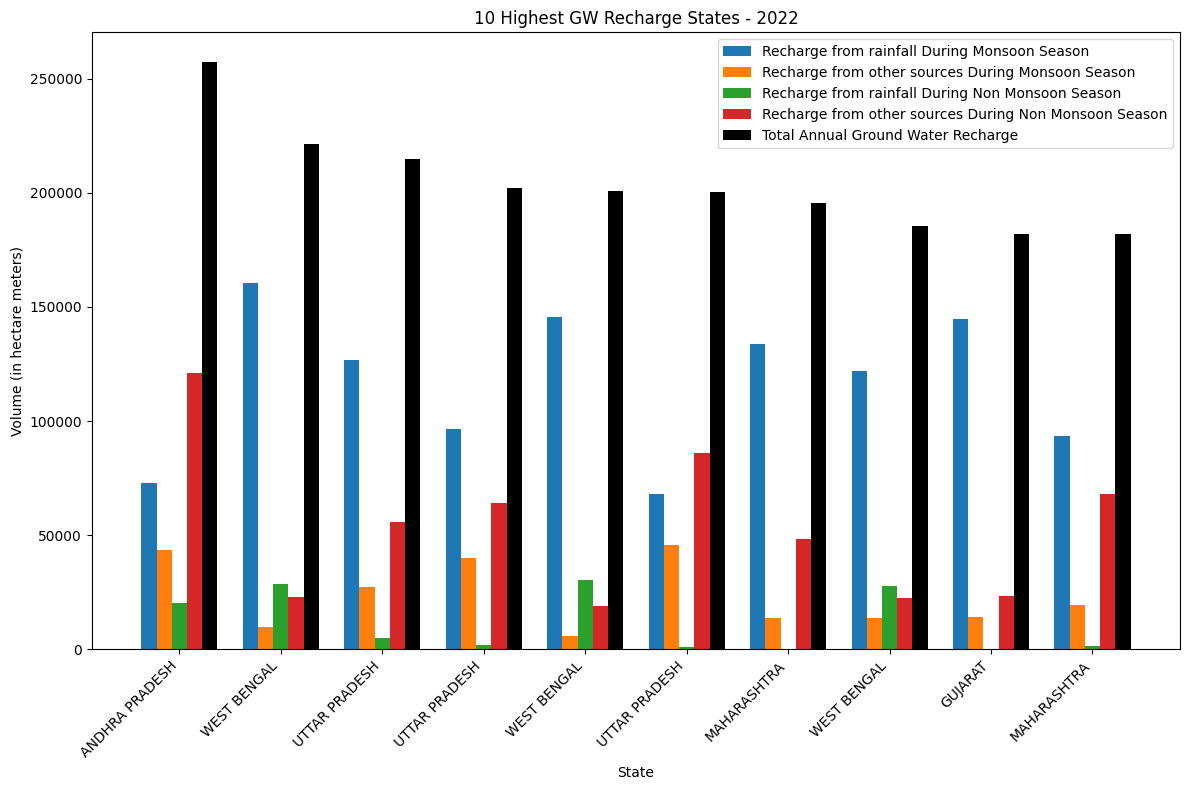










1. Analysis:

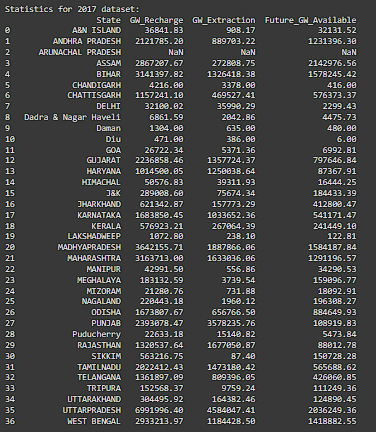
Dataset Size:

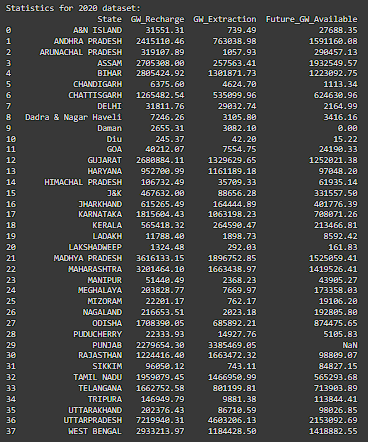
Shape of csv\_2017: (689, 16)

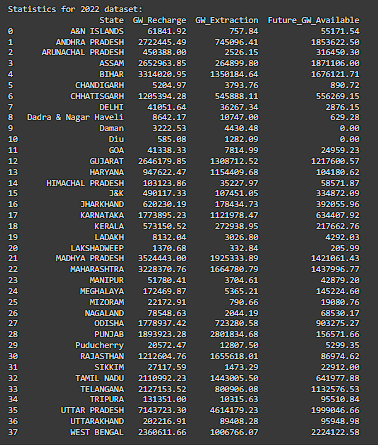
Shape of csv\_2020: (683, 16)

Shape of csv\_2022: (706, 16)

Taking the crucial parameters of State Name, GW\_Recharge, GW\_Extraction and Future\_GW\_Available we can plot several eye opening visualisations.







Mean Extraction-Recharge ratio for 2017: 0.5799154940134049 or, 57.99 percent.

Mean Extraction-Recharge ratio for 2020: 0.5615603386485803 or, 56.16 percent.

Mean Extraction-Recharge ratio for 2022: 0.546594095073245 or, 54.66 percent.

Thus, we see a roughly 2% decrease in the Extraction to Recharge Ratio of GroundWater across Indian States every 3 years.

1. Results:

* Thus we see the most unsustainable groundwater usage practices are the NorthWest States of Punjab, Rajasthan, Haryana Daman and Diu. At times, the exploitation rate is double the annual recharge rate.
* The Seven Sister States of NorthEast India consistently rank among the most renewable and healthy groundwater usages in India especially states such as Nagaland, Sikkim and Arunachal Pradesh.
* Monsoon Rains provide the maximum replenishment to the aquifer and the water table across India.(exception being Andhra Pradesh)

1. Conclusion:

In conclusion, this project sheds light on the critical issue of groundwater sustainability in India, over the years 2017 to 2022. The findings reveal significant disparities in extraction-recharge ratios among states, highlighting the urgent need for sustainable groundwater management practices.

Moving forward, it is crucial to integrate these analyses into policy development, emphasizing the adoption of technological innovation, and regulatory frameworks.

7) References/Bibliography:

MINISTRY OF WATER RESOURCES, RIVER DEVELOPMENT AND GANGA REJUVENATION “An Introduction to Real-time Hydrological Information System” <https://cdnbbsr.s3waas.gov.in/s3a70dc40477bc2adceef4d2c90f47eb82/uploads/2023/02/2023021683.pdf>

Ministry of Drinking Water and Sanitation. (2017). "Ground Water Information Booklet". Government of India.

Central Ground Water Board. (2019). "Ground Water Quality in Shallow Aquifers of India". Ministry of Jal Shakti, Government of India.

Ministry of Jal Shakti. (2021). "National Aquifer Mapping and Management Programme (NAQUIM)". Government of India.

<https://ingres.iith.ac.in/gecdataonline/gis/INDIA;parentLocName=INDIA;locname=INDIA;loctype=COUNTRY;view=ADMIN;locuuid=ffce954d-24e1-494b-ba7e-0931d8ad6085;year=2022-2023;computationType=normal;component=recharge;period=annual;category=safe;mapOnClickParams=false>

<https://data.opencity.in/dataset/national-compilation-on-dynamic-ground-water-resources-of-india-2022/resource/categorization-of-states-of-india-by-ground-water-usage---2022>

<https://data.opencity.in/dataset/5cee606a-3134-4a4e-91f5-4de820cbc5cd/resource/0eafe99f-9206-491e-8c45-0cc25e02040a/download/compilation-gw-2023.pdf>

<https://cgwb.gov.in/>

<https://prsindia.org/files/policy/policy_analytical_reports/1455682937--Overview%20of%20Ground%20Water%20in%20India_0.pdf>

<https://data.opencity.in/dataset/81c8bba1-0bdd-48da-ad9b-17cf5a767759/resource/5af44d77-181c-40fd-b5b8-5d2b204f6ef2/download/gwra-2017-national-compilation.pdf>

<https://data.gov.in/resource/stateut-wise-comparison-over-exploited-assessment-units-underground-water-2017-2022>

<https://www.globalwaters.org/sites/default/files/integrated_district-level_water_quality_and_scarcity_estimates_for_india_.pdf>