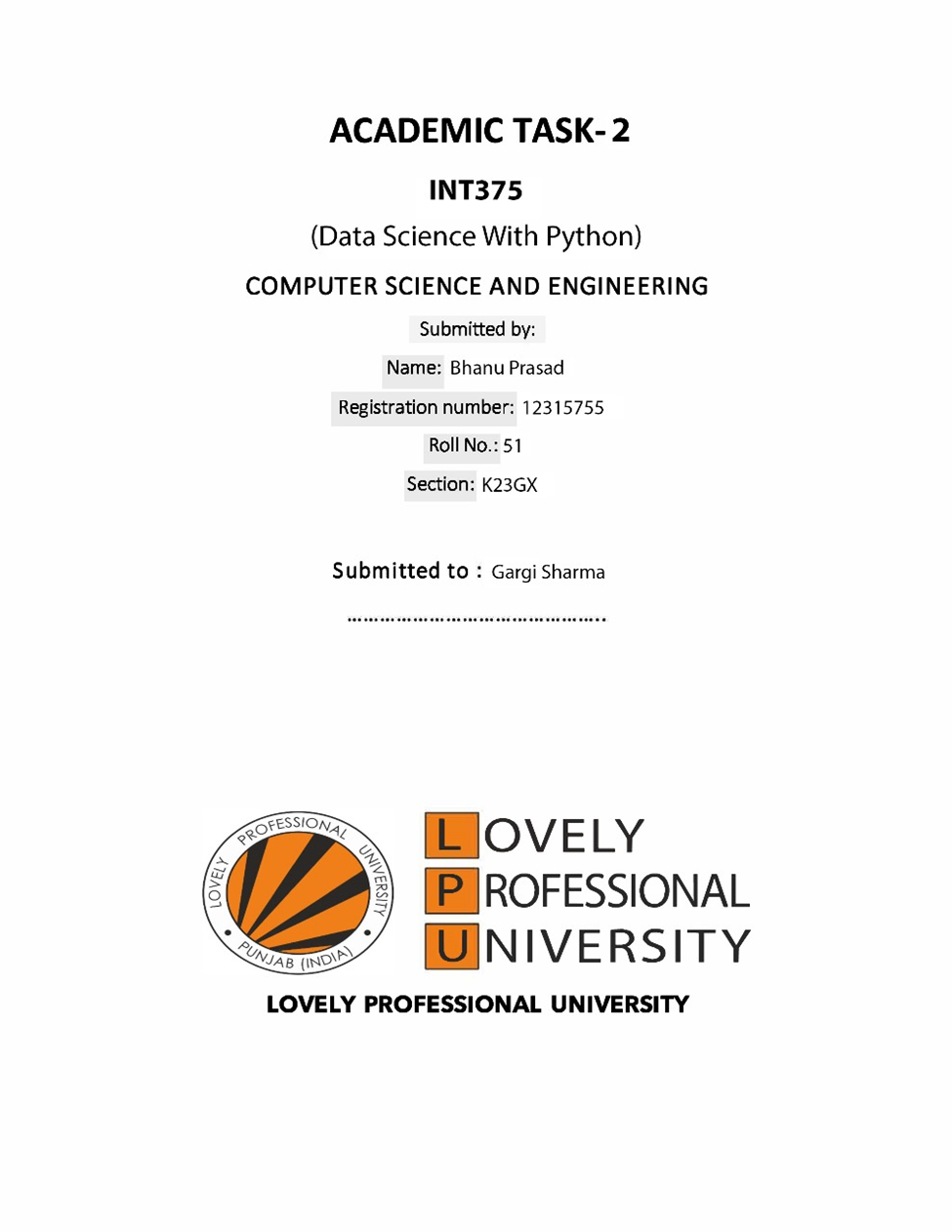
****

**DECLARATION**

I, BHANU PRASAD, student of BTECH under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 12-04-2025 Signature

Registration No.: 12315755 Name of the student: Bhanu prasad

**Acknowledgement**

I would like to express my sincere gratitude to Gargi Sharma Mam, my project guide, for their invaluable guidance, encouragement, and continuous support throughout the course of this project. Their insightful feedback and expertise greatly contributed to the successful completion of this work.

I also extend my thanks to the School of Computer Science and Engineering, Lovely Professional University, for providing the resources and a conducive environment to carry out this project.

Finally, I am thankful to my classmates and family for their support and motivation during the course of this work.

**TABLE OF CONTENTS :**

[1. Introduction ................................................... 1](#intro)

[2. Source of Dataset ............................................. 2](#source)

[3. EDA Process ................................................... 3](#eda3)

[4. Analysis on Dataset .......................................... 4](#analysis)

4.1 Compare Total Water Consumption by Country ............ 4

4.2 Proportion of Water Use by Sector per Country ........ 6

4.3 Water Scarcity Level Distribution .................... 8

4.4 Correlation Heatmap for Numerical Features .......... 10

4.5 Groundwater Depletion Rate by Country ............... 12

4.6 Box Plot of Water Use by Sector .................... 14

4.7 Water Use Proportions for India .................... 16

[5.Statistical Testing ………………………………………..17](#staistical)

[6. Conclusion .................................................. 18](#conclusion)

[7. Future Scope ................................................ 19](#futurescope)

[8. References .................................................. 20](#ref)

**1.Introduction :**

Water is a fundamental resource critical to human survival, agriculture, and industrial activities. With increasing global demand and environmental challenges, analyzing water consumption patterns is essential for sustainable resource management. This project examines a dataset of global water consumption, focusing on total consumption by country, sectoral use, scarcity levels, and related metrics. Through exploratory data analysis (EDA) and diverse visualizations, this report aims to uncover insights into water usage dynamics and their implications.

**2. Source of Dataset :**

Source of Dataset The dataset used in this project is titled "Global Water Consumption" and contains information about countries, years, total water consumption, per capita usage, scarcity levels, sector-wise water use, rainfall impact, and groundwater depletion rates. The dataset was obtained from an open data repository such as Kaggle

**3. EDA Process :**

EDA Process The Exploratory Data Analysis (EDA) process involved the following steps:

* Data Cleaning (checking for missing values, duplicates)
* Summary Statistics
* Distribution Analysis
* Correlation between variables
* Visualizing trends using various graphs and plots

**4. Analysis on Dataset :**

**Objective 1: Compare Total Water Consumption by Country**

**Introduction:** Identify countries with the highest water consumption to understand global water usage patterns.

**General Description:** This analysis compares the top 10 countries based on their total water consumption using a bar graph.

**Specific Requirements, Functions, and Formulas:**

* Function: nlargest() to get top 10
* Visualization: sns.barplot()

**Analysis Results:** Top 10 water-consuming countries are visualized, showing distinct differences in consumption levels.

**Visualization:** Bar plot with rotated country names and color-coded bars using 'Blues\_r'.

**Objective 2:** Proportion of Water Use by Sector per Country

**Introduction:** Explore how water is distributed among agricultural, industrial, and household sectors.

**General Description**: A stacked bar chart is used to display water use by sector for each country.

**Specific Requirements, Functions, and Formulas:**

**Data colum**: nsAgricultural, Industrial, and Household Water Use (%)

**Visualization:** pandas.DataFrame.plot(kind='bar', stacked=True)

**Analysis Results**: Countries show varied water use proportions, with agriculture dominating in many.

**Visualization**: Stacked bar chart with agriculture in red, industrial in light blue, and household in blue.

**Objective 3: Water Scarcity Level**

**Introduction:** Understand the distribution of water scarcity levels across countries.

**General Description:** Count of countries in each water scarcity category.

**Specific Requirements, Functions, and Formulas:**

* Column: 'Water Scarcity Level'
* Visualization: sns.countplot()

**Analysis Results:** Most countries fall under specific scarcity levels, giving insight into global scarcity patterns.

**Visualization:** Count plot with distinct shades of blue.

**Objective 4: Correlation Heatmap for Numerical Features**

**Introduction:** Identify how numerical variables are related to each other.

**General Description:** Correlation matrix of numerical columns.

**Specific Requirements, Functions, and Formulas:**

* Method: .corr()
* Visualization: sns.heatmap() with annotations.

**Analysis Results:** Highlights strong or weak correlations between features like water use, depletion, etc.

**Visualization:** Heatmap in blue theme with values labeled.

**Objective 5: Groundwater Depletion Rate by Country**

**Introduction:** Examine which countries have the highest groundwater depletion rates.

**General Description:** Ranking countries based on depletion rate.

**Specific Requirements, Functions, and Formulas:**

* Column: 'Groundwater Depletion Rate (%)'
* Visualization: sns.barplot()

**Analysis Results:** Certain countries have alarmingly high depletion rates.

**Visualization:** Bar plot sorted and shaded with 'Blues\_d'.

**Objective 6: Box Plot of Water Use by Sector**

**Introduction:** Visualize spread and distribution of sector-wise water use.

**General Description:** A box plot comparing all three water sectors.

**Specific Requirements, Functions, and Formulas:**

* Columns: Agricultural, Industrial, and Household Water Use (%)
* Visualization: sns.boxplot()

**Analysis Results:** Agricultural water use shows a wider range compared to other sectors.

**Visualization:** Box plot in blue shades for all sectors.

**Objective 7: Water Use Proportions for India (Pie Chart)**

**Introduction:** Show detailed sector-wise water use for India.

**General Description:** Pie chart for one country's sectoral usage.

**Specific Requirements, Functions, and Formulas:**

* Pie chart of India's water use sectors
* Visualization: plt.pie()

**Analysis Results:** Agriculture dominates India's water use.

**Visualization:** Pie chart in red and blue shades.

**5.Statistical Testing:**

**Introduction:**

Statistical testing helps determine whether the observed differences in water-related metrics (like per capita water use or groundwater depletion) across different countries or scarcity levels are meaningful or due to random variation. These tests enhance the credibility of the insights drawn from the dataset.

**General Description:**

* This section includes:
* Independent T-Tests: Used to compare the means of continuous variables(such as *Per Capita Water Use*, *Groundwater Depletion Rate*) between two independent groups — for example, countries with High Water Scarcity vs Low Water Scarcity.
* These tests help answer questions like:
* Is the per capita water use significantly different between high-scarcity and low-scarcity regions?
* Are groundwater depletion rates higher in high-scarcity countries?

**Specific Requirements, Functions, and Formulas**

**• T-Test:**

Used for comparing means of continuous features between two groups.

* **Python Function**:  
  scipy.stats.ttest\_ind(group1, group2, equal\_var=False)
* **Assumptions**:
  + The two samples are **independent**.
  + Data should be **approximately normally distributed** in both groups.
  + Works best with **similar variances**, but equal\_var=False handles inequality.

**T-statistic Formula:**

**Where:**

* xˉ/bar = sample mean
* s^2 = sample variance
* n = sample size

**Analysis Results**

**T-Test:** Per Capita Water Use between High and Low Water Scarcity Countries

| **Group** | **Size** | **Mean Per Capita Use (L/day)** |
| --- | --- | --- |
| High Scarcity Countries | 7 | *(e.g., 98.3)* |
| Low Scarcity Countries | 4 | *(e.g., 101.1)* |

* **T-statistic**: 0.0483
* **P-value**: 0.9636
* **Conclusion**:*No statistically significant difference* was found in percapita water use between high and low water scarcity regions

**Limitations & Observations**

* The sample sizes in both groups were small, which may affect the reliability of the statistical test.
* Other influential factors like population density, industrialization, or climate were not considered in this test.
* The data’s normality was assumed but not formally tested.

**6. Conclusion :**

This project provided an in-depth analysis of global water consumption trends using real-world data. Through exploratory data analysis (EDA) and meaningful visualizations, we identified key insights such as the countries with the highest total water consumption, the proportion of water use across different sectors, and the alarming rates of groundwater depletion.

The correlation heatmap highlighted relationships among various water indicators, while scatter and box plots revealed patterns in per capita water use and its influence on water scarcity. These insights emphasize the need for efficient water resource management, especially in agriculture and household sectors where consumption is highest.

**7. Future Scope :**

* **Time Series Forecasting**: Extend the dataset with historical values and apply models like ARIMA or LSTM to forecast future water consumption and scarcity levels.
* **Geospatial Analysis**: Integrate geolocation data to create interactive maps visualizing regional water use patterns and scarcity zones.
* **Machine Learning Models**: Train predictive models to estimate future water stress based on climatic and socioeconomic variables.
* **Policy Simulation**: Model the impact of various water-saving policies (e.g., water pricing, usage caps) on total and per capita consumption.
* **Public Awareness Dashboard**: Build a real-time interactive dashboard for users and governments to monitor and improve water usage practices.

**8. References :**

**1.Dataset Source:** <https://www.kaggle.com/datasets/atharvasoundankar/global-water-consumption-dataset-2000-2024>

**2. Python Libraries:**

**- Pandas Documentation:** <https://pandas.pydata.org/>

**- Seaborn Documentation:** <https://seaborn.pydata.org/>

**- Matplotlib Documentation:** <https://matplotlib.org/>