**Final Report: Analysis and Forecasting of Water Consumption, Rainfall, and Reservoir Levels in Istanbul**

**Project Overview**

This project investigates how Istanbul's daily water consumption is influenced by rainfall patterns and dam reservoir levels over the period from 2011 to 2024. By analyzing data from the Istanbul Metropolitan Municipality (IBB) Open Data Portal, the study identifies historical trends, evaluates correlations among key variables, and forecasts future scenarios using machine learning (ML) and statistical methods.

Water resource management is increasingly important for cities like Istanbul, where growing urbanization and climate change impact the sustainability of water systems. This project explores the following core questions:

* How strongly is water consumption correlated with rainfall and dam fill rates?
* Are there seasonal or long-term trends in consumption, rainfall, and reservoir levels?
* Can data-driven models predict future water usage patterns and sustainability challenges?

**Motivation**

With global warming increasing the frequency of droughts and urban expansion raising water demand, Istanbul faces growing challenges in water management. Understanding the interplay between rainfall, reservoir capacity, and consumption is crucial for proactive planning. This project offers insights that:

* Illuminate the key drivers of water use and storage fluctuations,
* Inform future-focused policy and conservation strategies,
* Provide tools for predicting supply-demand mismatches.

**Data Sources**

All data was obtained from the **IBB Open Data Portal**:

* **Daily Water Consumption** (in m³/day) for the city of Istanbul.
* **Daily Rainfall Amounts** (kg/m³) for 10 major dams.
* **Daily Reservoir Fill Percentages** for the same 10 dams.

The datasets were cleaned, timestamp-aligned, and merged using Python for a unified time-series analysis. Missing values were addressed through filtering or computed rolling means.

**Objectives**

1. Explore correlations between rainfall, reservoir levels, and water usage.
2. Identify annual and seasonal trends to understand climate-related shifts.
3. Predict next-day water consumption using machine learning models.
4. Forecast water usage for the next 10–100 years using both ML and statistical methods.
5. Visualize findings to support data-driven water management.

**Tools and Techniques**

* **Environment:** Google Colab (Python)
* **Data Handling:** pandas, numpy
* **Visualization:** matplotlib, seaborn
* **ML Models:** LinearRegression, RandomForestRegressor
* **Statistical Modeling:** ARIMA via statsmodels
* **Evaluation Metrics:** RMSE, R², correlation coefficients

**Key Findings and Analysis**

**1. Rainfall vs Reservoir Fill Levels (Per Dam)**

Each dam’s rainfall values were compared with its corresponding reservoir fill percentage. Pearson correlation was calculated, followed by scatter plots and heatmaps. We found out the results as the hypothesis cannot be rejected: There is no significant relationship between the amount of rainfall and the occupancy rate of the dam.

A graph with blue dots

AI-generated content may be incorrect.A graph with different colored squares

AI-generated content may be incorrect.

**Ömerli dam is brought here as an example,**

**There are 9 more dams like this result.**

**2. Water Consumption vs Reservoir Fill Levels**

We examined the relationship between citywide daily water consumption and each dam's fill percentage. The results as follows: Hypothesis rejected: There is a significant relationship between water consumption and the fullness rate of dams.

A chart with numbers and a red and blue box

AI-generated content may be incorrect.

A graph with blue dots

AI-generated content may be incorrect.

**Ömerli dam is brought here as an example, there are 9 more dams like this result.**

**3. Water Consumption vs Rainfall (Per Dam)**

This part measured how consumption is influenced by localized rainfall. Most correlations were weak, but varied by dam. The results as follows: Hypothesis rejected: There is a significant relationship between water consumption and the amount of rainfall into the dams.

A graph of blue dots

AI-generated content may be incorrect.A chart with numbers and a number of objects

AI-generated content may be incorrect.

**Ömerli dam is brought here as an example, there are 9 more dams like this result.**

**4. Total Rainfall vs Total Consumption**

Daily summed rainfall across all dams was compared to citywide consumption. Resulted in Hypothesis rejected: There is a significant relationship between total precipitation and water consumption.

A grey background with black text

AI-generated content may be incorrect.

A green dotted chart with numbers and a white background

AI-generated content may be incorrect.

**5. Total Rainfall vs Reservoir Fill (All Dams)**

We tested if total rainfall impacted overall reservoir fill levels. Some strong and some negligible correlations were observed. Results 🡪 The hypothesis cannot be rejected: There is no significant relationship between total precipitation and the fullness rate of dams.

A screenshot of a graph

AI-generated content may be incorrect.

A graph with blue dots

AI-generated content may be incorrect.

**Ömerli dam is brought here as an example, there are 9 more dams like this result.**

**6. ML Modeling: Predicting Next-Day Consumption**

**Feature Engineering:**

* Date decomposition: Year, Month, Day, Weekday
* Lag features: 1-day and 7-day lags for rainfall and fill
* 7-day rolling means

**Models Used:**

* **Linear Regression**: Baseline model
* **Random Forest Regressor**: Outperformed linear model in accuracy

**Results**:

The Random Forest model outperformed Linear Regression by achieving lower RMSE and higher R² scores. Features such as “Year”, “Ay” and “Roll7\_Yağış,” were among the most influential in predicting next-day water consumption. Average results as follows:

LinearRegression RMSE: 155700.5232215537 # Root Mean Sqr. Error

LinearRegression R^2: 0.7337980849896617 # Coefficient of Determination

RandomForest RMSE: 74054.42921005894

A graph of blue dots

AI-generated content may be incorrect.A graph of a line with orange dots

AI-generated content may be incorrect.RandomForest R^2: 0.939781059645995

A graph with a bar

AI-generated content may be incorrect.

A graph of a number of individuals

AI-generated content may be incorrect.

**7. ML-Based Simulation for Future (100-Year)**

Using future dates and climatological averages, RF predicted average daily consumption trends from 2024 to 2124. Three separate simulations for 10, 40, and 70-year intervals were plotted. The simulation using the Random Forest model projected a steady increase in daily water consumption over the next century, with higher trends evident in 40- and 70-year forecasts.

A graph with blue dots and numbers

AI-generated content may be incorrect.

A graph with blue dots and numbers

AI-generated content may be incorrect.

A graph with blue dots and numbers

AI-generated content may be incorrect.

**8. Global Warming Trend Visuals (Yearly Averages)**

Yearly means for rainfall, consumption, and fill percentages were computed. Combined line and bar charts visualize long-term climate impact. Yearly averages revealed a gradual rise in consumption, while rainfall showed variability and a slight downward trend—suggesting potential climate-related stress on water resources.

A graph and chart of a graph

AI-generated content may be incorrect.

**9. Feature Importance & Partial Dependence**

The "Year" feature had notable influence in predictions—indicating climate-linked progression. The “Year” feature ranked among the most impactful in the model, and the partial dependence plot confirmed that predicted consumption increases consistently with time.

A graph with blue and white bars

AI-generated content may be incorrect.

A graph with a line

AI-generated content may be incorrect.

**10. Linear Trend Forecasts (5, 10, 20 Years)**

A graph with numbers and lines

AI-generated content may be incorrect.Separate LinearRegression models were used to project average consumption, rainfall, and fill levels. Linear regression forecasts indicated a continued rise in water consumption and a possible decline in reservoir levels, supporting concerns about long-term sustainability.

**11. ARIMA Forecasts (10, 50, 100 Years)**

We used ARIMA(1,1,0)+trend models for all three variables. Results showed:

* Steady increase in consumption
* Fluctuations in rainfall
* Variable, generally declining reservoir fill percentages

A graph with a line and dots

AI-generated content may be incorrect.

A graph with blue and orange dots

AI-generated content may be incorrect.

A graph with a line and a dotted line

AI-generated content may be incorrect.

**10 years graphs are brought as an example. There are also graphs for 50 and 100 years outcomes in the code.**

**Limitations and Future Work**

* Static climatology assumptions in forecasting
* No integration of temperature or humidity data
* No spatial modeling for dam-specific climate effects

**Future Recommendations:**

* Integrate external climate datasets (ERA5/CMIP6)
* Include socio-economic and population growth indicators
* Develop interactive dashboards for city use

**Conclusion**

This project combines open-source municipal data and predictive modeling to understand Istanbul’s water dynamics. Through detailed correlations and long-range simulations, the findings demonstrate the increasing stress on water systems due to both demand and climate volatility.

The project offers a foundational framework for municipalities to forecast, prepare, and adapt urban water systems for a warming world.