Assignment 3: Portfolio of Work with related report

Alexander Alboukhari

May 6, 2025

Contents

Backup Project Files

All source code files, Jupyter notebooks, static images, data, and backup archives are included in the downloadable ZIP archive below:

- Download ZIP Archive (Code + Data + Images)

1 Task 1:Static Visualization Analysis Rating vs Price by Coffee Roast Type

1.1 Link to Publicly Accessible Visualization

The visualization was generated using Python and Matplotlib, and is available at the following link:

- Coffee Rating vs Price by Roast Type

1.2 Overview

This project analyzes a coffee reviews dataset to explore how coffee quality ratings relate to pricing,with a focus on differences among roast types.

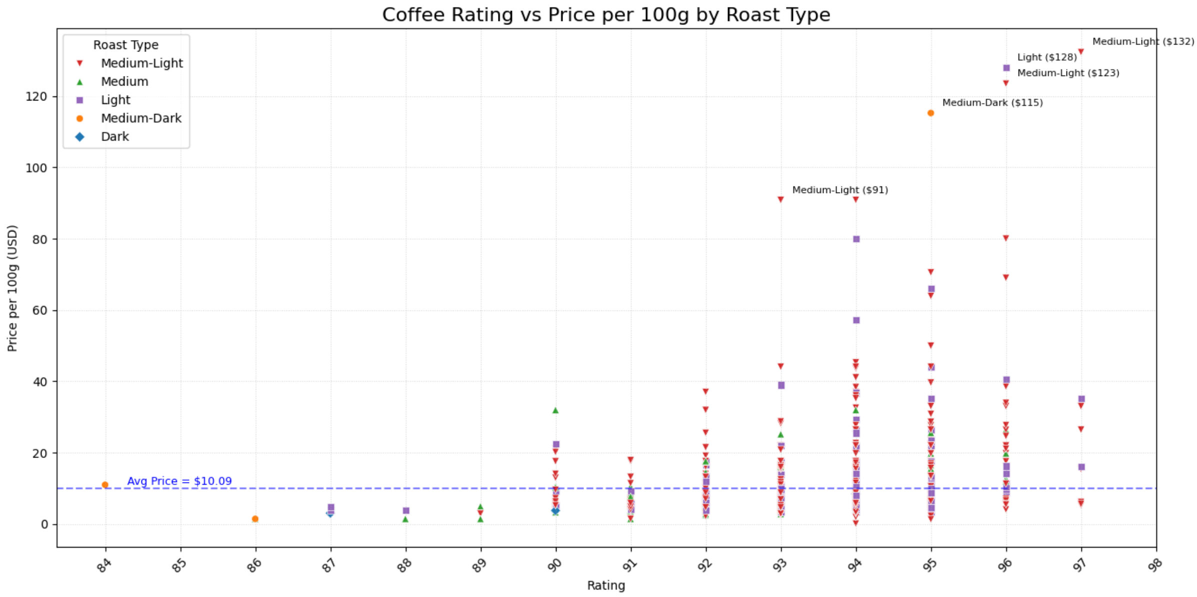

Figure 1: Scatter Plot: Rating vs Price Colored by Roast Type

The goal is to create a clean, static visualization as shown in Figure 1 that highlights patterns and provides insights to specialty coffee professionals and fans.

1.3 Audience

This visualization is designed for:

- Coffee professionals (roasters, buyers, graders).
- Retail managers and specialty distributors.
- Data analysts interested in market price.
- Coffee fans who are interested in understanding the pricing trends based on coffee ratings and roast types.

It is intended for those with a basic understanding of coffee roast categories and data interpretation.

1.4 Purpose

The purpose of this visualization is to clearly analyze and communicate the relationship between coffee ratings and market prices, while also examining whether roast types affect this relationship.

- Explore the relationship between coffee ratings and their price per 100g.

- Investigate if different roast types linked to premium pricing.
- Identify exceptionally priced outlier coffees.
- Support strategic insights into how roast type impact market value

The goal is to support data-driven decision-making among coffee professionals regarding quality and pricing strategies.

1.5 Questions Answered

This visualization addresses several key questions:

- **Is there a positive relationship between rating and price?**
  Yes, the plot shows that coffees rated above 94 tend to achieve higher market prices.
- **Do roast types impact pricing within similar rating ranges?**
  Medium-Light roasts appear slightly more dominant in higher priced brackets compared to darker roasts.
- **Which roasts are priced higher?**
  Medium-Light and Light roasts dominate the top price tier, especially above $100.
- **Are all highly rated coffees expensive?**
  Not necessarily. While many highly rated coffees command premium prices, others remain relatively affordable. However, the five most expensive coffees each priced above $115 per 100g are clearly labeled in the plot, emphasizing their outlier status in both rating and price..
- **Is price variability consistent across rating ranges?**
  No, price spreads widen for coffees rated 95 and above.
- **Are all highly rated coffees expensive?**
  No, several coffees rated 94 – 97 are reasonably priced, suggesting brand, exclusivity, and other factors affect pricing.
- **What is the average price point?**
  A horizontal line shows the average at $10.09.

1.6 Design Choices

1.6.1 Layout

The numerical character of both variables led one to choose a scatter graph. Every point reflects a coffee sample. Using both color and marker form, the roast type helps to decrease visual ambiguity.

1.6.2 Color Scheme and Markers

Roast types were encoded using distinct colors and markers:

- **Dark:** Blue triangle down
- **Medium-Dark:** Orange circle
- **Medium:** Green triangle up
- **Medium-Light:** Red triangle down
- **Light:** Purple square

Color was combined with marker shape to support accessibility

1.6.3 Highlighting Outliers

Labeling directly with (plt.text()) the five most costly coffees displayed roast type and price. This raises knowledge of goods that notably flee from accepted prices.

1.6.4 Reference Lines

- **Average Price Line:** Dashed blue line at $10.09, clearly labeled.

1.6.5 Tick and Axis Formatting

Using (plt.xticks()) the x-axis was manually ticked from 84 to 98 to guarantee that every rating is clear-view. The rotation in labels improved readability. To keep interpretability, the y-axis scale stayed linear.

1.6.6 Annotation Style

To prevent overlap, outlier annotations were positioned a bit above and to the right of their matching spots. For freely view, the average price label found where it belongs at the lower-left.

1.6.7 Design Principles Applied

- **Simplicity:** Focused only on essential variables.
- **Contrast:** High color and shape contrast ensures clarity.
- **Data-Ink Ratio:** No decorative clutter all elements are data-driven.
- **Accessibility:** Redundant encoding (color and shape) makes the plot inclusive.
- **Annotation Hierarchy:** Labels are used only where necessary (top 5), avoiding visual noise.
- **Scalability:** Saved at 300 DPI to ensure print and digital clarity.

1.6.8 Readability Enhancements

- **Gridlines**: Dashed, light gray lines for visual reference.
- **Legend:** Included was a clearly labeled legend headed *Roast Type*, whose color and form matched the relevant data points in the plot. For categorical sections, this offers quickly interpretability.
- **Resolution**: Saved at 300 DPI for publication quality export.
- **Whitespace Control**: `tight_layout()` applied to avoid clipping.

1.7 Conclusion

The relationships among coffee rating, pricing, and roast type is captured in the last scatter plot. It presents a reasonable and easily readable perspective on specialty coffee price. The story lets readers rapidly grasp both general trends and outlier behaviors via careful use of color, shape, label, and layout. Python ensures repeatability and adaptability for customizing. Industry reports, educational research, or consumer-facing information all fit this kind of visualizing. This type of static visualization has useful uses in product marketing, retail pricing strategy, and consumer education. For example, coffee companies could utilize these knowledge to price their high-rated roasts higher. Similarly, purchasers might identify undervalued coffees by comparing ratings to prices. In the future, this plot might be expanded with origin country data or feedback from customers to provide an intricate market profile. Given the Python script's reproducibility, the same approach might be used to other luxury products such as tea, wine, or chocolate to analyze rating-based price dynamics.

2 Task 2 : Global Water Consumption Analysis (2000-2024) Dashboard

2.1 Link to Publicly Accessible Live Project

Power BI Dashboard Link: Power BI Dashboard

2.2 Intended Audience

This dashboard is designed for key stakeholders in the field of water resource management and environmental including:

- **Academic Institutions**: For researchers in environmental science.
- **Urban and Agricultural Planners**: To identify and reduce inefficiencies in sectoral water use.

2.3 Purpose of the Dashboard

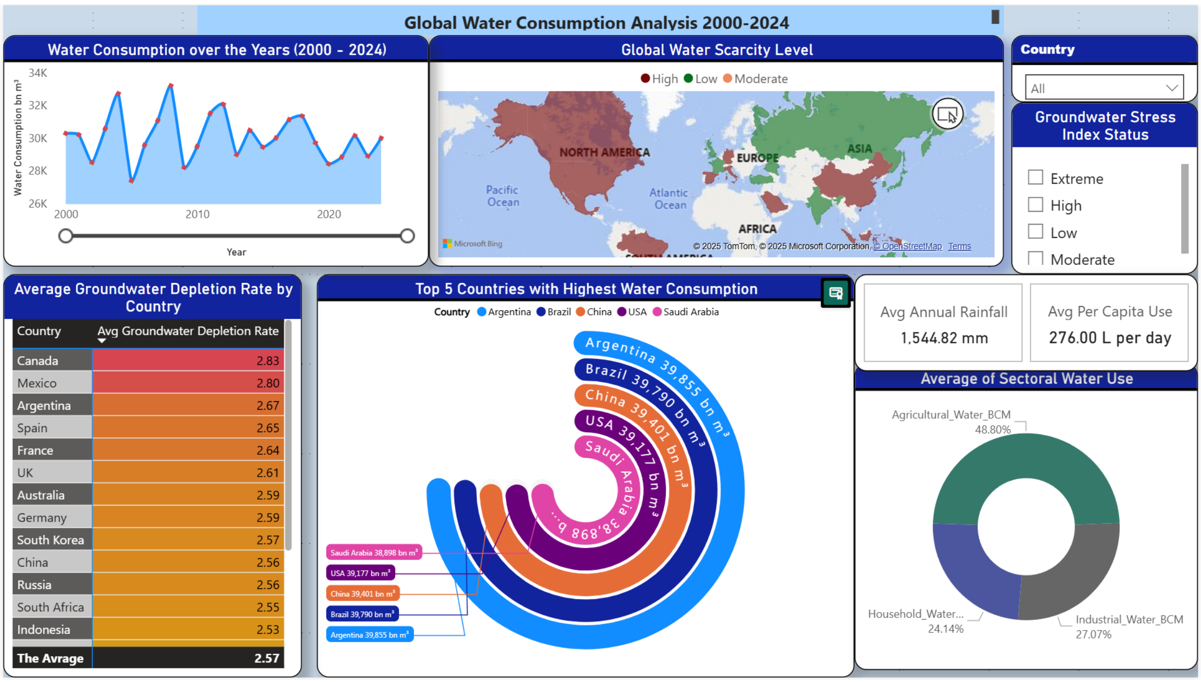

Figure 2: Global Water Consumption Analysis 2000-2024 Dashboard

The goal of the dashboard as shown in Figure 2 is to provide an interactive narrative on global water consumption from 2000 to 2024. It combines multiple usage trends, geographic scarcity, groundwater stress, and sectoral water use to help users identify high risk regions and assess major contributors to global water demand. It offers a c important message that especially in high demand areas, water usage is not fairly spread across sectors and regions, therefore stressing groundwater. Worldwide water use averaged between 26,000 and 34,000 billion cubic meters (bn m³) between 2000 and 2024. Global water consumption reached a new high in 2008 roughly 34,000 billion cubic meters. Though there is some decrease in the years that follow, there is no consistent drop indicating that demand is still high and that change over time moves slowly. The globe is running out of water groundwater depletion is accelerating, so efficiency gains in agriculture have to take center stage. Particularly with regard to resource stress connected to climate, these results should guide national and worldwide strategies for fair and sustainable water management.

2.4 Questions Answered

Each visualization in the dashboard was created to answer a specific analytical question relevant to global water usage and stress. The following table summarizes the question and the resulting insights:

1. **Global Water Consumption Over Time (2000–2024)**
   Question: How has global water consumption changed over the past two decades?
   Result: Steady upward trend in water use reflects growing demand resulting from industrialization, population development, and agricultural expansion.

2. **Global Water Scarcity Map**
   Question: Which regions are experiencing high levels of water scarcity?
   Result: Countries in EU like (Germany,Spain,Italy), the Saudi Arabia ,North Amarica, and parts of South Asia for Scarcity water, with very limited renewable water resources.

3. **Top 5 Countries by Water Consumption**
   Question: Which nations are the largest consumers of water globally?
   Result: Argentina and the Brazil lead in water consumption, followed by China, United States,and Saudi Arabia.

4. **Groundwater Depletion Rate by Country**
   Question: Which countries are depleting groundwater the fastest?
   Result: Countries like Canada, Mexico, and Argentina show high average depletion rates, indicating unsustainable extraction practices.

5. **Sectoral Water Use Distribution**
   Question: Which sectors are the primary consumers of global water resources?
   Result: Agriculture is the dominant water consuming sector (nearly 49%), followed by industrial (27%) and household use (24%).

6. **Groundwater Stress Index Distribution**
   Question: What is the global distribution of groundwater stress categories?
   *Result:*The countries fall into the high like Canada and China along with the extreme like UK in groundwater stress categories, signaling a looming global sustainability concern.

7. **Country Drill-Down Table**
   Question: What are the detailed metrics for each country regarding water use and stress?
   Result: Users may look into for every country factors including sectoral breakdown, per capita use, depletion rate, and rainfall. The evidence validates that water stress varies by context and is multifaceted.

2.5 Interactivity

The dashboard offers a interactivity, allowing users to:

- Filter by year, country, continent, or groundwater stress level.
- Use a time range slider to adjust the analysis window.
- Click on charts to dynamically update related views.
- Hover over visual elements for detailed tooltips and exact values.
- Choose different views using dropdown menus.

These features provide exploratory analysis and a better understanding of water usage.

2.6 Linkages Between Visualizations

The dashboard is designed with cohesive interconnections:

- **Geospatial Linkage:** Clicking a country on the water scarcity map updates charts on groundwater depletion and sectoral use.
- **Temporal Syncing:** The time range slider affects all visualizations simultaneously, ensuring consistent analysis.
- **Category Linkage:** Selecting a sector (e.g., Agriculture) dynamically filters the sectoral usage chart and table.
- **Stress-Level Filters:** Selecting a groundwater stress category refines both the map and statistical table.

These linkages transform the dashboard from a static view into an interactive data story.

2.7 Design Choices

2.7.1 Layout

The dashboard is structured in a logical narrative flow:

- Top-level visuals focus on global trends and spatial comparisons.
- Mid-level visuals break down country and sector metrics.
- Bottom-level visuals provide drill-down and comparative analysis.

2.7.2 Color Scheme

A data-driven color strategy was employed:

- **Blues and greens** for neutral indicators such as water volume and rainfall.
- **Reds and oranges** for critical indicators such as high depletion and stress.
- Colorblind-safe palettes ensure accessibility for all users.

2.7.3 Typography and Readability

- Clear, sans-serif fonts ensure readability across all screen sizes.
- Font hierarchy is used effectively for titles, subtitles, and legends.
- Minimal clutter and white space improve focus and user experience.

2.7.4 Best Practices Applied

Supported by purposeful design choices and data preparation processes, every dashboard show was created with a specific analytical goal in mind. Details for every one of the eight visualizations are included

- **Line charts**: for identify long term trends across time and visualize global consumption of water patterns,a line chart was chosen. Interactive sliders let you filter time. Data was filtered yearly for overall consumption.
- **Filled map**: to explain geographical differences in water shortage between nations.For areas of stress, a Filled map with a differing color scale offers quick visual signals.The country data integrated with scarcity scores. Nations without data were marked as white color.
- **Dount chart**: for compare water use by agriculture, industry, and households. A donut chart offers a view with labeled sectors in Global averages per sector were aggregated .
- **Advance Dount chart**: To pick out the countries with highest consumption. Comparative sizing and clear ranking were achieved with a bar chart. Color gradients improve visual appeal and in general overall Year was ordered, top nations chosen, and units defined.
- **KPI cards:** for summarizing rainfall and usage statistics.
- **Matrix table:** to identify countries that extract groundwater excessively.Red indicates unsustainable use, while horizontal bars make labels easier to read and each country average depletion rate was calculated.
- **Slicer:** to show world distribution of groundwater stress and colors complement the map in consistency.
- **Country Drill-Down List** to go thorough overview of metrics at the country level. A list that can be searched and sorted gives the user complete control. Every important metric was combined by country and organized into a single table.

The dashboard follows Edward Tufte's principles ([2]) of data visualization: maximizing data-to-ink ratio, maintaining integrity, and reducing chart junk.

This work used the Groundwater Stress Index (GSI) ([1]) in order to compare groundwater pressure across countries. Derived from both groundwater depletion rates and the moderating influence of rainfall on aquifer recharge, the GSI addresses It computed using the formula:

$$\text{Groundwater Stress Index (GSI)} = \frac{\text{Groundwater Depletion Rate (\%)}}{\text{Rainfall Impact (mm)}} \tag{1}$$

2.8 Conclusion

Finally, the Global Water Consumption Analysis (2000–2024) dashboard presents a data-driven and interactive analysis of globally water usage trends, therefore exposing a consistent increase requests resulting from industrialization, population increase, and agricultural development. It draws attention to important geographical inequalities, with nations like China, Brazil, and Argentina leading in usage while others suffer extreme shortage and groundwater pressure. Confirming agriculture as the biggest consumer, sectoral analysis emphasizes the critical necessity of more effective water use policies. Combining visual analytics with important metrics like the Groundwater Stress Index (GSI) helps the dashboard provide actionable information to equip planners, and academics with means to support fair and sustainable water management globally.

References

[1] P.M. Nussbaum, M. Somogyvári, C. Conrad, M. Sauter, and I. Engelhardt. Calculating groundwater stress and climate change-induced vulnerability of karst aquifers on a global scale. *Environmental Research Letters*, 2021.

[2] Edward R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 2nd edition, 2001.