Sasha Porter
2 May 2025
DS 4002
Professor Rasero

Using Data Science to Analyze Alcohol Consumption Patterns Across the World and Predict Future Consumption Behavior

Overview

You're a rising data scientist at the forefront of global health policy. Your latest mission? Use forecasting models to uncover future patterns in alcohol consumption across three continents. Alcohol consumption is more than just a cultural practice, it's a critical public health issue with profound implications for both individual well-being and societal outcomes. Around the globe, people's drinking behaviors are shaped by shifting economic conditions, policy interventions, and health campaigns. But how do these patterns change over time, and what can they tell us about the future? That's where you come in. You're joining a global data science team tasked with forecasting alcohol consumption trends to inform future public health efforts. Across regions like Europe, Africa, and Asia, alcohol use has shifted in complex ways, rising in some places, falling in others, driven by economic change, culture, and policy. Now it's your turn to help make sense of these patterns. Your team has access to an internationally standardized dataset from the United Nations Global SDG Indicator Database. It contains alcohol consumption per capita (ages 15+) for dozens of countries, collected in five-year intervals (2000–2019). Your challenge is to turn this data into a time series model that not only captures historical trends, but also predicts future drinking behavior. By looking at trends across three major world regions and making predictions about the future, you'll uncover insights that can make a real difference. Public health leaders need reliable predictions to guide policy and allocate resources—it is your work that will help public health leaders generate a solution.

Deliverable

You'll use real-world data to explore how alcohol consumption has changed across Europe, Africa, and Asia, and apply time series forecasting techniques, focusing on the ARIMA model, to predict future trends. Your challenge is to preprocess and aggregate the data, implement and evaluate your forecast, and interpret your results in a public health context. You'll complete this work using R for data cleaning and visualization, and Python for modeling. Starter files and resources are provided in the repository to help guide your process. Your final product will combine technical analysis with clear communication: a short written summary that presents your key findings and visualizations, along with a GitHub repository that documents your work. This is your chance to take the lead as a data scientist and contribute insights that matter while gaining experience in modeling, interpretation, and public impact.