# Utility Sector Performance Optimization

## Project Overview

This project analyzes over two decades of U.S. utility sector data to identify underperforming state–sector combinations and model data-driven strategies for achieving 10%+ revenue growth. Using Python for data cleansing, forecasting, and profitability simulation and Tableau for interactive dashboards, the analysis pinpoints pricing inefficiencies, customer adoption gaps, and revenue improvement opportunities across residential, commercial, industrial, transportation, and other utility sectors.

## Dataset Description

Source: Kaggle – U.S. Electricity Prices Dataset (originally from the U.S. Energy Information Administration)  
Time Span: 2001–2024 (monthly granularity)  
Size: ~85,000 rows, covering all U.S. states and territories  
Features:  
 - year, month – time period  
 - stateDescription – U.S. state or region  
 - sectorName – utility sector (residential, commercial, industrial, transportation, other)  
 - customers – number of customers served  
 - price – average price per kWh  
 - revenue – total sector revenue  
 - sales – total electricity sales (MWh)

## Methodology

1. Data Cleaning  
 - Removed rows with missing/invalid customer counts (~30% of records affected)  
 - Created revenue\_per\_customer metric for cross-state comparability  
 - Applied IQR filtering to remove extreme outliers  
 - Converted to time-series format for trend analysis  
  
2. Exploratory Data Analysis (EDA)  
 - Distribution and correlation analysis of price, revenue, and customers  
 - Outlier detection using Z-score thresholds and log-scale scatterplots  
  
3. Performance Benchmarking  
 - Identified bottom 10 state–sector combinations by revenue per customer  
 - Categorized underperformance causes: No Revenue Data, High Customers–Low Revenue, Low Revenue Base  
  
4. Forecasting  
 - Holt-Winters Exponential Smoothing to project sector-specific revenue  
 - Simulated strategic scenarios: +10% customer growth, +10% price increase, +20% usage growth, and combined strategies  
  
5. Profitability Simulation  
 - Estimated costs to model potential profit margins by state and sector  
 - Identified low-margin markets for further investigation  
  
6. Visualization  
 - Built Tableau dashboards for KPI tracking, benchmarking, and scenario modeling

## Key Insights & Results

- Transportation sector showed strongest price elasticity, suggesting price adjustments could boost revenue.  
- Residential and industrial sectors were largely price-inelastic — non-price interventions are more effective here.  
- Several states (e.g., Utah, New Mexico, Colorado) showed high customers but low revenue per customer, pointing to underutilization.  
- Combined strategy of usage + price + customer growth achieved simulated revenue gains of 10–15% in targeted sectors.

## How to Run the Project

Requirements:  
pip install pandas numpy matplotlib seaborn statsmodels jupyter  
  
Steps:  
1. Clone this repository:  
 git clone https://github.com/yourusername/utility-sector-performance.git  
 cd utility-sector-performance  
2. Download the dataset from Kaggle and place it in the project folder as clean\_data.csv.  
3. Open and run the Jupyter Notebook:  
 jupyter notebook Identifying\_Underperforming\_Utility\_Sectors.ipynb  
4. Tableau dashboard (.twbx file) can be opened with Tableau Desktop or Tableau Public.

## Visualizations

Tableau Dashboards:  
- Outlier State–Sector Analysis  
- Revenue Forecasting & Scenario Modeling  
- Sector & State KPI Tracker  
  
Example Charts from Python Analysis:  
(Add image links/screenshots stored in /images folder)

## Future Improvements

- Integrate actual cost data instead of simulated costs for more accurate profitability modeling.  
- Expand analysis to real-time streaming utility data for near-instant KPI updates.  
- Apply machine learning regression models to refine elasticity estimates.  
- Build a web-based dashboard in Streamlit or Dash for broader accessibility.