

Optimal Manufacturing Site Analysis

1. Introduction

This project focuses on the analysis of **Metropolitan Statistical Area (MSA) economic data**, particularly in the manufacturing sector. The goal is to clean, transform, and analyze data to extract insights on economic trends, manufacturing output, and income levels across various U.S. metropolitan areas. The analysis is performed using **R and the tidyverse ecosystem**.

2. Objectives

- Process raw economic data for **data consistency and accuracy**.
- Identify **high-performing MSAs in manufacturing output**.
- Determine the **most cost-effective MSA for manufacturing** based on personal income and economic indicators.
- Enable **state-based comparisons** of economic performance.

3. Methodology

The analysis follows a structured data pipeline:

Data Processing & Cleaning

- The dataset (`final_2.csv`) is imported and **MSA names are standardized** by removing unnecessary text.
- Key variables are **selected, renamed, and transformed** for better usability.
- Suppressed or missing data values (`(D)`) are replaced with **NA**.
- Numeric values are **cleaned and converted** into appropriate formats.

Data Transformation

- The dataset is **reshaped using `spread()` and `gather()`** to create structured data for analysis.
- Additional **data manipulations** include:
 - Filtering and ranking MSAs based on **manufacturing output**.
 - Identifying the **lowest per capita personal income (PCPI) MSA** among high-manufacturing regions.
 - Splitting **MSA names into city and state** for regional analysis.

Custom Functions for Analysis

Two custom **R functions** (`msa_large()` and `msa_large2()`) are developed to:

- Retrieve the **MSA with the highest value** for each economic category within a given state.

- Allow **state-based queries** for comparing economic indicators.

4. Key Findings

- The top 25% highest manufacturing output MSAs were identified.
- Among them, the lowest-cost metropolitan area for manufacturing was determined based on per capita income.
- The dataset allows comparative analysis across states to identify top-performing MSAs for different economic indicators.

5. Tools & Technologies Used

- **R Programming** for data analysis and visualization.
- **tidyverse (dplyr, tidyr, ggplot2)** for efficient data manipulation.
- **Statistical analysis techniques** for economic trend evaluation.

6. Conclusion

This project demonstrates how **data analytics and statistical techniques** can be leveraged to **extract insights from economic data**. The findings provide valuable information for **businesses, policymakers, and investors** looking to understand regional economic performance. With further enhancements, the analysis could be expanded to include **predictive modeling and interactive reporting tools**.