Fortune 100 Companies: Business & Revenue Analysis

1. Overview

This project analyzes **Fortune 100 companies in the United States** to uncover insights related to revenue distribution, industry contribution, growth trends, workforce size, and geographic presence.

The project demonstrates an **end-to-end data analytics workflow**, starting from **web scraping raw data from scratch**, followed by **data cleaning, SQL-based analysis**, and **interactive dashboarding using Power BI**.

2. Data Source & Web Scraping

Source Website

The dataset was scraped from the following publicly available Wikipedia page:

https://en.wikipedia.org/wiki/List_of_largest_companies_in_the_United_States_by_revenue

This page provides structured tabular information on the largest U.S. companies ranked by revenue.

3. Web Scraping Methodology

Rather than using a ready-made dataset, the data was **collected from scratch using Python**, simulating a real-world analytics scenario.

Approach:

- Used Python's requests library to fetch the webpage
- Parsed HTML content using BeautifulSoup
- Identified and extracted the main table containing company details
- Collected key fields such as:
    - Company Name
    - Revenue
    - Industry
    - Number of Employees
    - Headquarters Location
- Stored the scraped data into a structured **CSV file** for downstream analysis

This step demonstrates the ability to work with **real-world web data**, handle unstructured sources, and convert them into analytics-ready datasets.

4. Dataset Summary

The cleaned dataset contains the following attributes:

- Company name
- Industry classification
- Total revenue
- Employee count
- Revenue growth rate
- Headquarters state (USA)

Dataset scale:

- **100 companies**
- **$13.0 trillion total revenue**
- **16.2 million total employees**
- **7.69% average revenue growth**

5. Tools & Technologies

- **Python:** Web scraping, data cleaning, EDA
- **BeautifulSoup & Requests:** HTML parsing and data extraction
- **Pandas & NumPy:** Data transformation and aggregation
- **SQL (PostgreSQL):** Querying and metric validation
- **Power BI:** Interactive dashboard and reporting

6. Analysis & Methodology

1. Data Collection

- Scraped company-level data directly from Wikipedia
- Converted unstructured HTML tables into structured datasets

2. Data Cleaning & Preparation

- Standardized industry names
- Handled missing and inconsistent values
- Converted revenue and employee figures into numeric formats

3. SQL Analysis

- Created tables in PostgreSQL

- Ran queries to validate KPIs (total revenue, top companies, industry contribution)
- Used SQL for aggregation and ranking logic

4. Visualization & Reporting

- Built an interactive Power BI dashboard
- Designed KPI cards, charts, maps, and filters
- Enabled drill-down analysis using slicers for **Industry** and **HQ State**

7. Dashboard Highlights

Key KPIs

- **Total Revenue:** $13.0T
- **Total Employees:** 16.2M
- **Average Revenue Growth:** 7.69%
- **Total Companies:** 100

Key Visualizations

- **Top 10 Fortune 100 Companies by Revenue**
- **Revenue Contribution by Industry**
- **Revenue vs Growth by Company (Scatter Plot)**
- **U.S. Map Showing Company Headquarters by State**

Key Insights

- Revenue is **highly concentrated** among the top companies
- Financials, Healthcare, and Retail are the largest contributors
- Several mid-revenue companies exhibit strong growth potential
- Fortune 100 headquarters are clustered in major economic hubs such as California, Texas, and New York

8. Business Value

- Provides a clear view of corporate revenue concentration
- Helps identify dominant and high-growth industries
- Supports strategic analysis and benchmarking
- Demonstrates practical analytics skills aligned with real business use cases

9. How to Run the Project

1. Scrape data from the Wikipedia source using Python
2. Clean and preprocess the dataset using Pandas
3. Store and query data using SQL
4. Load cleaned data into Power BI
5. Build visuals and publish insights

10. Conclusion

The analysis shows that Fortune 100 revenue is **concentrated among a small group of leading companies**, with Financial Services, Healthcare, and Retail dominating overall contribution. While large firms lead in scale, several mid-sized companies exhibit **higher growth rates**, indicating future potential. Geographic clustering of headquarters further highlights key economic hubs within the U.S.