

Business Data Analysis Project Documentation

ONESIGHT

Names

Maria Remon

Sara Mostafa

Rawan Ahmed

Eman Alaa

1. Project Overview

This project provides a comprehensive business intelligence framework by combining analysis across four critical business dimensions: Global Inflation, Company Demographics, E-commerce Sales, and Consumer Behavior.

The primary goal is to visualize and interpret complex global data to support strategic decision-making in market positioning, operational efficiency, and economic risk assessment.

The project utilized a pipeline involving Python for cleaning and transforming complex datasets (Global Inflation and Companies) and Power BI for final data preparation and visualization.

2. Objectives

- **Economic Risk Assessment:** To understand global economic stability and identify regions facing extreme inflation or volatility.
- **Corporate Analysis:** To establish key metrics around company size, age, employee distribution, and industry contribution to the global workforce.
- **E-commerce Performance Tracking:** To monitor key sales metrics, identify seasonal trends, and evaluate product and regional performance.
- **Customer Behavior Profiling:** To profile customer purchasing habits, payment preferences, and sales distribution by demographic and product categories.

3. Tools Used

| Tool | Purpose |
|----------|--|
| Excel | Exploration: Used for initial viewing and validating the raw data structure before complex processing. |
| Python | Cleaning & Transformation: Employed for rigorous data cleaning, standardizing formats, and transforming data into a usable structure. |
| Power BI | Visualization: Served as the platform for preparing the cleaned data and designing the final interactive dashboards using calculated KPIs. |

4. Datasets Used

The project is built upon four primary data domains visualized across four distinct Power BI dashboards:

- **Global Inflation Data:** Macroeconomic data used to track inflation rates over time across various countries.
- **Companies Data: Demographic data on companies, including founding year, employee estimates, and geographic location.**
<https://www.kaggle.com/datasets/peopledatalabssf/free-7-million-company-dataset> (before cleaning)
(couldn't be able to upload before cleaning on GitHub as the file is so big)
- **E-commerce Sales Data:** Transactional data covering revenue, orders, regional sales, and product performance.
- **Shopping Trends Data:** Data focusing on customer purchasing habits, promo code usage, ratings, and gender distribution.

5. Methodology

The project followed a structured, multi-step methodology, with a heavy focus on the Python cleaning phase for the two complex datasets.

5.1. Data Cleaning (Python)

Global Inflation Data Cleaning Steps (Tidy Data Conversion)

The script focused on preparing the data for time-series analysis by converting its structure and cleaning the values.

| Step | Simple Explanation |
|--------------------------|--|
| Column Standardization | Removed extra spaces and capitalized headers to ensure uniform column names (e.g., Countryname, Indicatorname). |
| Duplicate Removal | Dropped duplicate rows based on Country and Indicator names to maintain data uniqueness. |
| Missing Value Imputation | Filled all missing (NaN) values with zero to avoid errors in numeric calculations. |
| String Cleaning | Standardized Countryname to Title Case and Indicatorname to lowercase for consistency. |
| Data Reshaping (Melt) | Converted the data from a wide format (years in separate columns) to a long format (all years in one column, Year, and all values in one column, InflationRate), which is ideal for visualization. |
| Date Conversion | Converted the Year column (initially a string) into a proper datetime object. |

Companies Data Cleaning Steps

| Step | Simple Explanation |
|------------------------|---|
| Column Renaming | Renamed complex headers (e.g., current employee estimate) to simplified, consistent names (e.g., CurrentEmployees). |
| String Standardization | Cleaned and standardized text fields (e.g., name, industry, country) to Title Case and replaced missing values with the string 'Unknown'. |
| Date Handling | Converted the FoundedYear to a proper date format and filled missing years with a placeholder date (1900-01-01). |
| Duplicate Removal | Dropped duplicate rows based on the combination of name and domain to ensure each company is counted once. |

5.2. Data Preparation & Visualization

- KPI Calculation: Key Performance Indicators (KPIs) like Total Revenue, Average Order Value, and Average Inflation were calculated directly from the prepared data.
- Dashboard Design: Designed four separate, interactive dashboards to clearly present the insights for each domain, allowing users to apply slicers and filters for in-depth analysis.

6. Key Insights and Analytical Questions

The dashboards were specifically designed to answer the following detailed analytical questions, leveraging the calculated KPIs and data visualizations.

6.1. Companies Analysis

| Analytical Questions / KPI Insights | Metrics/Findings |
|---|--|
| Highest Number of Companies (Top 5 Countries) | Revealed by the Top 5 Countries by Number of Companies chart. |
| Highest Number of Current Employees (Top 10 Companies) | Identified by the Top 5 Companies by Current Employees chart (e.g., IBM, Accenture). |
| Total Number of Companies (KPI) | 98K |
| Highest Current Employees in a Single Company (KPI) | 274K |
| Founding Year of the Oldest Company (KPI) | 1789 |
| Total Number of Current Employees (KPI) | 67M |
| Average Number of Current Employees per Company (KPI) | 672.69 |
| Industries Employing the Most People | Identified by the Top Company per Country and Total Current Employees by Industry charts (e.g., Information Technology, Hospital & Health Care). |
| Top Company in Each Country by Current Employees | Answered by drill-down on the Top Company per Country chart. |
| Difference between Total and Current Employees for each company | Calculated metric on the backend data. |
| Total Gap between Total and Current Employees | Calculated metric on the backend data. |
| Top 10 Countries with the highest total employees | Identified by the Top 3 Countries by Total Employees chart (e.g., United States, Unknown, United Kingdom). |

6.2. Global Inflation Analysis

| Analytical Questions / KPI Insights | Metrics/Findings |
|--|--|
| Highest, Lowest, and Average Global Inflation Rates (KPIs) | Max: 65.37K, Min: -72.70, Avg: 37.93 |
| Countries Suffering from the Highest Inflation Rates | Identified by the Most Inflation-Intense Countries chart (e.g., Venezuela, Congo, Dem. Rep. Of The). |
| Inflation Change over Time across different countries | Visualized in the Average of Inflation Rate by Year and Countryname line chart, showing high volatility. |
| Inflation Difference between Countries in terms of Proportion/Percentage | Visualized in the Revenue by Region donut chart, showing the relative impact. |
| Countries showing Stable vs. Highly Volatile Inflation Patterns | Derived from analyzing the trend lines in the Average of Inflation Rate by Year and Countryname chart. |
| Effect of Selecting a Specific Country on the Inflation Trend | Enabled through the country slicer feature on the dashboard. |

6.3. E-commerce Sales & Revenue

| Analytical Questions / KPI Insights | Metrics/Findings |
|--|--|
| Total Revenue Generated (KPI) | 73.41M |
| Total Orders Placed (KPI) | 100K |
| Average Order Value (AOV) (KPI) | 734.15 |
| YoY Revenue Growth/Decline (KPI) | Revenue is growing YoY (50.7%). |
| Months with Highest/Lowest Revenue | Identified by the Monthly Revenue line chart (Highest: November/December). |
| Regions Generating Highest Revenue | Identified by the Revenue by Region chart (e.g., Oceania, Asia, North America). |
| Regions Underperforming | Identified by the lowest bars in the Revenue by Region chart (e.g., Africa). |
| Product Categories Generating the Most Revenue | Identified by the Revenue by Product Category chart (e.g., Automotive, Sports). |
| Categories with the Highest Pricing Contribution | Identified by the Product Category Pricing chart. |
| Products Ordered the Most | Identified by the Most Ordered Products chart (e.g., Automotive, Electronics). |
| Categories with the Highest Number of Returns | Identified by the Products Returned chart. |
| Most Used Payment Method | Identified by the Orders by Payment Method donut chart (e.g., Credit Card, Bank Transfer). |
| Percentage of Products Returned (KPI) | 6% |

6.4. Consumer Behavior

| Analytical Questions / KPI Insights | | Metrics/Findings |
|---|--|--|
| Total Sales (KPI) | | 233K |
| Product Categories Generating the Highest Sales | | Identified by the Total Sales by Category chart (e.g., Clothing). |
| Revenue Distribution across Product Categories | | Visualized in the Revenue by Product Category donut chart. |
| Average Purchase Amount per Transaction (KPI) | | 59.76 |
| Percentage of Purchases including Promo Code Usage (KPI) | | 43% |
| Total Number of Customers (KPI) | | 3.9K |
| Shipping Type Most Commonly Used across Product Categories | | Identified by the Shipping Type Usage Across Product Categories chart (e.g., Blouse, Skirt). |
| Sales Variation across Different Seasons (Highest Sales Season) | | Identified by the Total Sales by Season line chart (e.g., Fall). |
| Gender Distribution of Customers | | Visualized in the Customer count by Gender donut chart (e.g., Female 1.2K, Male 2.7K). |
| Average Rating per Item Purchased | | Visualized in the Average Rating by Item Purchased chart (e.g., Shoes, Skirt, Jeans). |

7. Conclusion

This project successfully created an interconnected business intelligence ecosystem across four diverse domains: Corporate, Economic, and E-commerce.

By leveraging Python for robust data cleaning and Power BI for dynamic visualization, the dashboards provide immediate, actionable answers to critical business questions.

The insights range from identifying global economic instability (inflation hotspots) to granular operational details (peak sales seasons and top-performing products), providing stakeholders with a centralized source of truth for strategic planning and risk mitigation.