**STAT 112 - Introduction to Data Processing and Visualization Project**

**Urban Population, Economical Factors, and Their Effect on Sales Patterns on Countries with Least Co2 Emissions**

Umut Deniz Taş

December 8, 2024

*Abstract*— This research explores the relationship between urbanization, economic elements, and sales trends in nations with the lowest CO2 emissions. By examining critical factors such as the percentage of urban populations, fuel prices, taxation levels, and product performance, we reveal patterns that demonstrate how urbanization and economic conditions affect consumer behavior. Our results suggest that countries with larger urban populations typically show better sales performance, whereas economic factors like elevated gasoline prices and tax rates can hinder consumer spending. Furthermore, the performance of product lines varies considerably among countries with different average order values, and seasonal trends indicate fluctuations in order volumes and deal sizes over time. This study offers practical insights for businesses aiming to tailor their sales strategies to align with urbanization and economic trends in eco-conscious markets. Keywords— Urbanization, Sales Performance, Economic Factors, Gasoline Prices, Tax Rates, Product Performance, Seasonality, Consumer Behavior, Market Trends, Order Values, Data Visualization, Tableau Analysis, Business Insights, Market Strategy, CO2-Emissions.

# Introduction

In this research, we utilized two dataset containing data from countries around the world, each with various features that impact sales and consumer behavior and joined them. The main goal of this study is to identify the factors that influence sales patterns the most, focusing on safety, reliability, and economic variables.

1.1 **DATA DESCRIPTION**

**1.**1.1    **Automobile Sales Dataset**

The dataset contains Sales data of an Automobile company. In the Automobile dataset, there exist 20 variables and 2747 observations. The dataset is saved as **“autosale.xlsx”**. The variable information is as follows:

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| **ORDERNUMBER** | This column represents the unique identification number assigned to each order. |
| **QUANTITYORDERED** | It indicates the number of items ordered in each order. |
| **PRICEEACH** | This column specifies the price of each item in the order. |
| **ORDERLINENUMBER** | It represents the line number of each item within an order. |
| **SALES** | This column denotes the total sales amount for each order, which is calculated by multiplying the quantity ordered by the price of each item. |
| **ORDERDATE** | It denotes the date on which the order was placed. |
| **DAYS\_SINCE\_LASTORDER** | This column represents the number of days that have passed since the last order for each customer. It can be used to analyze customer purchasing patterns. |
| **STATUS** | It indicates the status of the order, such as "Shipped," "In Process," "Cancelled," "Disputed," "On Hold," or "Resolved." |
| **PRODUCTLINE** | This column specifies the product line categories to which each item belongs. |
| **MSRP** | It stands for Manufacturer's Suggested Retail Price and represents the suggested selling price for each item. |
| **PRODUCTCODE** | This column represents the unique code assigned to each product. |
| **CUSTOMERNAME** | It denotes the name of the customer who placed the order. |
| **PHONE** | This column contains the contact phone number for the customer. |
| **ADDRESSLINE1** | It represents the first line of the customer's address. |
| **CITY** | This column specifies the city where the customer is located. |
| **POSTALCODE** | It denotes the postal code or ZIP code associated with the customer's address. |
| **COUNTRY** | This column indicates the country where the customer is located. |
| **CONTACTLASTNAME** | It represents the last name of the contact person associated with the customer. |
| **CONTACTFIRSTNAME** | This column denotes the first name of the contact person associated with the customer. |
| **DEALSIZE** | It indicates the size of the deal or order, which are the categories "Small," "Medium," or "Large." |

**1.1.2**     **Global Country Information**

This comprehensive dataset provides a wealth of information about **all countries worldwide**, covering a wide range of indicators and attributes. It encompasses demographic statistics, economic indicators, and much more. In the dataset, there exist 14 variables and 195 observations. The dataset is saved as **“world-data-2023.xlsx”**. The variable information is as follows:

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| **Country** | Name of the country. |
| **Latitude** | Latitude coordinate of the country's location |
| **Longitude** | Longitude coordinate of the country's location |
| **Birth Rate** | Number of births per 1,000 population per year |
| **CO2-Emission** | Carbon dioxide emissions in tons |
| **CPI** | Consumer Price Index, a measure of inflation and purchasing power. |
| **Gasoline Price** | Price of gasoline per liter in local currency |
| **GDP** | Gross Domestic Product, the total value of goods and services produced in the country |
| **Life expectancy** | Average number of years a newborn is expected to live |
| **Population** | Total population of the country |
| **Tax Revenue(%)** | Tax revenue as a percentage of GDP |
| **Total tax rate** | Overall tax burden as a percentage of commercial profits. |
| **Unemployment rate** | Percentage of the labor force that is unemployed |
| **Urban\_population** | Percentage of the population living in urban areas |

The joined datasets contain 36 variables but in our research, we focused on 10 of them. These variables are: Urban population, Country, Gasoline price,Total tax rate, Productline, Priceeach, Dealsize, Quantityordered, Orderofdate and Sales.

# Data Preproccessing

In the early phases of data analysis, we encountered multiple variables that had dollar signs preceding their values, which led Tableau to classify them as dimensions instead of measures. To address this issue, we needed to clean and transform these variables by eliminating the dollar signs and modifying the data format, enabling these values to be recognized as numerical measures. This adjustment allowed us to conduct accurate analyses and create effective visualizations.

Furthermore, we revised the names of certain variables to make them more intuitive and comprehensible when presented on the axes of our visualizations. This modification was essential for enhancing the user-friendliness and accessibility of the data for interpretation.

For the remainder of the dataset, we employed Tableau Public's 'Data Interpreter' feature to automatically rectify any inconsistencies, errors, or hidden problems within the data, thereby ensuring the quality and integrity of our analysis. After these preprocessing steps, we were able to dive deeper into the data and begin addressing our research questions, exploring the relationships between variables and uncovering actionable insights related to urbanization, economic factors, and product performance. This allowed us to effectively interpret the data and move forward with the visualization and analysis stages.

# Exploratory Data Analysis

3.1 **Research Questions: We will analyze the data with 5 research questions, and we will try to draw some conclusions.**

3.1.1 *How do sales vary in countries with high urban populations compared to those with low urban populations?*

Comprehending the impact of urbanization on sales can aid businesses in planning market growth and distributing resources effectively. Areas with significant urban development may offer greater sales opportunities owing to increased population density, improved infrastructure, and heightened economic activity. This knowledge can guide companies in focusing on regions with the greatest potential. A scatter plot was used to answer this question. We created a calculated field called ‘Urban Population Percentage’ which is made by dividing urban population to the total population and multiplying it with a hundred. After that we used this calculated field in rows and total number of sales in columns. Then we filtered the data to show us the graph of the ten countries with the least co2-emissions. The results can be seen on figure 1.0.

A graph with blue dots

Description automatically generated

### **Figure 1.0**

### **Interpreting the Figure 1.0:** The chart demonstrates an overall pattern in which nations with elevated urbanization rates (for instance, 90% or above) show increased sales volumes. Conversely, countries with lower urbanization levels (around 60%) generally experience moderate to low sales figures. This suggests a favorable relationship between urbanization and sales, probably fueled by economic activities in densely populated areas. Nevertheless, anomalies indicate that other elements, such as income levels or market traits, might also affect sales outcomes. For instance, Belgium is an outlier in this plot, even though having 98.04% of the population in the urbanized areas, they had fewer total sales then 6 other countries that have less urban population percentage.

3.1.2 *What is the relationship between gasoline prices, tax rates, and total sales across regions?*

Economic elements like fuel costs and taxation levels can greatly affect consumer habits and sales figures. This inquiry examines the ways in which these factors shape buying choices and sales patterns in various areas, assisting in the development of pricing and marketing approaches. A combination of line and bar chart was used to answer this question. We used countries on the columns, total sales and average gasoline prices on rows. To be able to show both average gasoline prices and total sales in the same chart we used a dual axis chart. While the left axis represents average gasoline prices, the right axis represents the total number of sales. Since the difference of the values of these two variables are too big, we did not synchronize left and right axis’ to be able to generate a better visualization. Then we used coloring on our line chart to show the tax rates of countries. This way

the audience can see if the respective country has high or low tax rate by looking at the line. Lastly, we filtered the data to show us the graph of the ten countries with the least co2-emissions. The results can be seen on figure 1.1.

**A screenshot of a graph

Description automatically generated**

A graph of a graph with blue and black lines

Description automatically generated with medium confidence

**Figure 1.1**

**Interpreting the Figure 1.1:** The chart indicates that nations with elevated gasoline prices (such as Norway) often sustain robust sales despite the financial strain, likely attributable to higher income levels or a consumer inclination towards premium offerings. In contrast, nations with reduced gasoline prices (like Austria) exhibit moderate sales, implying that gasoline prices alone do not govern sales patterns. Notably, Denmark, which has one of the lowest tax rates, still records comparatively high sales. This implies that although tax rates and gasoline prices can affect purchasing power, they are not the only determinants of sales results—other elements such as market size, disposable income, and consumer preferences significantly contribute as well.

3.1.3 *Which product lines generate the highest sales in countries with varying average order values?*

Analyzing the relationship between product performance and urbanization can help identify which product lines thrive in specific markets. It can also guide businesses to tailor their offerings to suit the preferences of urban and less urbanized regions. Two different stacked bar charts were used to answer this question. In the first bar chart we used average price of the items on the rows and countries on the columns. In the second chart we used total number of sales instead of average price of the items. Then we used coloring to show product distribution in both charts and since the numbers are close, we labeled each color to see these small differences. Finally, we filtered the data to show us the graph of the ten countries with the least co2-emissions. The results can be seen on figure 1.2.

**Figure 1.2**

**Interpreting the Figure 1.2:** The graph illustrates that "Classic Cars" consistently excel in numerous nations, particularly in those with elevated average order values (for instance, Finland and Norway). These nations also exhibit a wider range of successful product categories, including "Planes" and "Motorcycles." Conversely, countries with lower average order values (such as the Philippines) demonstrate a less varied performance, suggesting restricted purchasing power or varying consumer tastes.

3.1.4 *Which are the top countries by total sales, and how does their percent urban population influence their ranking?*

Understanding the relationship between urbanization and sales success is critical for identifying markets with high sales potential. It helps businesses to decide if prioritizing urbanized regions for future investment and strategic planning is a good idea. A bar chart was used to answer this question. We used total number of sales in the columns and countries on the rows. To show the effect of the urban population, we used the calculated field we created in question 3.2.1. The darker bars indicates that these countries have a high population on urbanized areas of their respective country. Finally, we filtered the data to show us the graph of the ten countries with the least co2-emissions. The result can be seen on figure 1.3.

A graph with blue and white bars

Description automatically generated

**Figure 1.3**

**Interpreting the Figure 1.3**: Finland and Norway excel in total sales, illustrating a strong relationship between elevated urbanization levels and economic prosperity. Their completely urbanized settings create concentrated markets that are conducive to sales expansion.

Singapore and Denmark also achieve commendable results, likely attributable to their significant average urban population ratios. These findings imply that densely populated urban areas offer an advantageous environment for effective sales strategies.

Interestingly, nations such as the Philippines, which exhibit lower urbanization rates, still rank among the top achievers. This may suggest a high population density in urban locales, even with a reduced overall urbanization percentage.

The darker bars in the chart emphasize countries with larger urban populations, which generally excel in sales performance, indicating a direct connection between urban infrastructure and consumer buying power.

3.1.5 *How do order quantities and total order values vary by deal size across different time periods?*

Investigating seasonal trends and the impact of deal size on order behavior helps businesses predict market demand and adjust their sales strategies accordingly. It is important for optimizing inventory, predicting sales, and identifying periods of high or low performance. Two-line charts were used to answer this question. In the first chart, we used years on columns as quarters and total number of items ordered on rows. In the second chart, we used years on columns again and used total number of sales on rows. After that we filtered our data to see how each type of item performed on both charts. Finally, *we filtered the data to show us the graph of the ten countries with the least co2-emissions. The result can be seen on figure 1.4.*

A graph of different colored lines

Description automatically generated

**Figure 1.4**

**Interpreting the figure 1.4:** Order Quantities: Medium-sized transactions exhibit notable seasonal variations, reaching their highest points in Q4 2018 and Q4 2019, indicating that companies prioritize securing medium-sized orders during these times. In contrast, small transactions display a consistent trend, while large transactions show minimal fluctuations over time.

Sales Values: Medium-sized transactions also account for the largest share of total sales during peak periods, underscoring their vital contribution to revenue generation. The large transaction size consistently yields stable yet comparatively lower sales, demonstrating reliability in high-value orders with limited seasonal influence.

The relationship between the Q4 peaks in both order quantities and sales values suggests a significant seasonal impact, likely associated with end-of-year business cycles or holiday demand.

The decline in Q1 2020 across all transaction sizes may reflect global market disruptions, highlighting the necessity of adjusting strategies during uncertain periods. These results emphasize the necessity for customized approaches for mid-sized transactions, especially during peak seasons, while also ensuring strong pipelines for both small and large deals to guarantee consistent stability throughout the year.

# Conclusion and Dıscussion

This examination validates that urbanization, economic influences, and product efficacy are interconnected factors affecting sales results. Although countries that are significantly urbanized and wealthy exhibit robust sales performance, other elements like consumer habits, seasonal variations, and market traits also have essential impacts. Companies can utilize these findings to enhance their market strategies, customize product offerings, and anticipate seasonal needs, thereby ensuring ongoing growth in both mature and developing markets.

Subsequent analyses could build upon this research by integrating further variables, such as income distribution, infrastructure quality, and regional consumer trends, to offer a more thorough insight into sales dynamics.

The results of this analysis offer practical insights for companies aiming to enhance their sales strategies and effectively target their markets.

Urbanization as a Crucial Factor: Companies should focus on areas with significant urbanization rates because of their market potential. Nevertheless, they must also recognize densely populated urban areas in less urbanized nations, such as the Philippines, to seize untapped opportunities.

Economic Considerations: Although gasoline prices and tax rates affect purchasing power, they are less significant than broader elements like income levels and consumer behavior. Companies need to take these economic contexts into account when determining pricing or entering new markets.

Product Diversification: Businesses can concentrate on high-value products in wealthier markets while implementing customized strategies for lower-income areas with limited diversification.

Seasonality and Adaptability: The seasonal surges in the fourth quarter highlight the necessity for proactive planning to capitalize on periods of increased demand.

4.1 Some Surprising Insights:

The performance of countries like the **Philippines**, despite their lower urbanization, was unexpected and highlights the role of population density in driving sales.

The robust sales in regions with high gasoline prices suggest that consumer preferences and disposable income can override economic burdens.

# Links to our project

Tableau Public Dashboard:

<https://public.tableau.com/app/profile/umut.deniz.ta./viz/112-Homework-Dashboard/Dashboard1#1>

Github:

https://github.com/Hybwris/Stat-112-Project-1