# 

STAT 112 - Introduction to Data Processing and Visualization Project

Best Country to Sell Products

# by

Lara Kayaalp

December 7, 2024

# 

# 

# 

# Abstract

# This study focuses on identifying the best country to sell products by analyzing data from different regions with varying features and price points. To achieve this, economic indicators of countries were compared against market trends to find the factors influencing market demand. By analyzing relationships between numerical and categorical data in the dataset, insights to guide strategic decision-making for product sales were gained.

# Introduction

In this research, a dataset containing information from 196 unique countries was analyzed, each with different market characteristics. The main goal of this project is to determine which countries are the most ideal for selling products based on factors such as market potential, consumer demand, and economic stability. The original dataset of countries included 14 variables for countries, listed below:

|  |  |
| --- | --- |
| **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 dataset for automobile sells, on the other hand, contained 20 variables. Described below:

|  |  |
| --- | --- |
| **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." |

The dataset included categorical columns with some values missing across all variables. These issues were resolved through a data-cleaning process, which is explained below.

# Data Cleaning Process

## During the data pre-processing stage, the aim was to make values consistent within the dataset. This involved detecting and addressing outliers where necessary, and filtering the dataset to handle missing values. These steps had to be completed before the analysis process, to ensure that the dataset was was clean and reliable for analysis.

## Outliers

## Outliers in the dataset were detected using the interquartile range (IQR) method. As there weren’t enough outliers to necessiate automating the process, the outliers were manually filtered in Tableau.

## The results revealed outliers in a column:

## **Birth Rate:** An outlier was identified, Philippines. Its birth rate was significantly higher than the rest. Accordingly, it was filtered in Tableau.

* **GDP:** This time, two outliers were identified. Philippines and USA. They were also filtered.
* **CO2 Emissions:** 4 outliers, Germany, Japan, Philippines and USA, were identified and filtered.
* **Gasoline Price:** The same 4 outliers as above, Germany, Japan, Philipines and USA, were also identified and filtered.

## Missing values

There were missing values in all columns. These missing values were filtered in Tableau to allow further calculations.

While comparing birth rate with deal size score, the deal size had to be converted to a numeric value. To achieve this, a calculated field was created. The field allowed putting the birth rate and deal size on a scatter plot. Essentially converting a categorical column to a numeric column. The function used for this process is given below:

A math equation with orange and black text

Description automatically generated

# Exploratory Data Analysis

After the data cleaning process, the process of analyzing the data set, which contains properties of 196 countries and deal size scores, was ready to be started. Below, you can see the relationship between birth-rate and deal size score, created to investigate and understand the data properly. This way, observing the mean, the standard deviation, the quartiles, the minimum and the maximum values of each quantitative data was possible for further conclusions.

## Research Question

### Does birth rate influence sales in a country?

A graph with green dots and lines

Description automatically generated

Birth rate is a quantitative and continuous variable and may be an indicator of the economic status of a country. Countries with a lower birth rate generally have higher life standards. R-squared value of the chart is 0.043, and the values are slightly left-skewed.

### Does GDP influence deal-sizes in a country?

### A graph with green dots and lines Description automatically generated

After removing the USA outlier, it is visible that as GDP decreases, deal size also decreases. However, the outlier, USA, could require further inspection to understand why it doesn’t fit the current trend. This result validates that as economic status of a country decreases, people focus on more budget alternatives of automobiles.

As explained above, as the GDP of a country decreases, deal sizes also generally decrease. However, this relationship didn’t seem enough to be conclusive, so more relationships were discussed.

### A graph of sales Description automatically generatedAre deal-size and sales correlated?

Deal size and sale number were compared to understand the relationships they have.

This horizontal graph shows that as deal size increases, sales also increase. This relationship was inspected to be able to use sales and deal size interchangeably in further relationships.

Sales with CO2 emissions were compared next.

### Are CO2 emissions and sales correlated?

A graph with green dots and numbers

Description automatically generated

As seen in the graph, it turns out that there is a very strong relationship between sales and carbon emissions in a country. As carbon emissions decrease, sales also decrease. This result was also not surprising, so more questions were considered.

### Does Gasoline Price Influence Sales?

A graph with green dots and numbers

Description automatically generated

Sales by gasoline prices were compared above to understand whether as gas prices increase, sales decrease. The question was to figure out whether increased gas prices scare consumers away from purchasing automobiles. As it turns out, more factors were at play here. There was a correlation between gas prices and other factors which affected sales more than the gas prices do alone.

# Conclusion

In conclusion, throughout the project involving combining two datasets, cleaning some values, filtering and creating more fields through calculation, countries’ economic situation was compared to sales. Gasoline prices don’t seem to be influential, CO2 and sales have a reverse correlation, deal size and sales seem to be closely correlated, GDP and sales also have an inverse correlation, and birth-rate and sales don’t have a conclusive trend line.