{AUTUMN INTERNSHIP PROJECT REPORT}

# VISUALIZING GLOBAL TEMPERATURE TRENDS

SATYAJIT KUMAR SHAW

SECTION 1/BANGABASI COLLEGE

Period of Internship: 25th August 2025 - 19th September 2025

Report submitted to: IDEAS – Institute of Data Engineering, Analytics and Science Foundation, ISI Kolkata

# 1. Abstract

This Project explores time-series data visualization and analysis using Python. Initially, it focuses on analyzing global temperature data from two sources (GCAG and GISTEMP) to identify historical trends and seasonal patterns. Through line plots and heatmaps, the analysis reveals a significant warming trend since the 1970s. The notebook then extends these techniques to a separate dataset on car sales across different regions and companies over 10 years, demonstrating how similar visualization and analysis methods can be applied to understand sales trends, regional performance, and company-wise distribution. The project highlights the power of data visualization in uncovering insights from diverse time-series datasets.

# 2. Introduction

This project focuses on the crucial skill of analyzing and visualizing time-series data, a fundamental aspect of understanding trends and patterns across various domains. Time-series data, which is collected sequentially over time, is ubiquitous in fields ranging from climate science and economics to sales forecasting and stock market analysis. Understanding how to effectively handle, analyze, and visualize this type of data is essential for making informed decisions and drawing meaningful conclusions.

**Relevance**: The project's relevance lies in its application to real-world datasets. By analyzing climate data, we can observe and interpret the long-term trends in global temperatures, contributing to the broader understanding of climate change. Similarly, analyzing sales data provides insights into market performance, regional strengths, and company-specific trends, which are vital for business strategy.

**Technology Involved**: The project primarily utilizes Python, a powerful and versatile programming language widely adopted in data science. Key libraries employed include:

- **Pandas:** For efficient data manipulation and analysis, particularly with structured data like Data Frames.
- **Matplotlib and Seaborn:** For creating static and aesthetically pleasing statistical visualizations.
- **Plotly Express:** For generating interactive and informative plots, enhancing the exploratory data analysis process.
- **NumPy:** For numerical operations and handling arrays.

**Background Material Survey**: The project draws upon fundamental concepts in data analysis, including data loading, aggregation, and time-series specific techniques like calculating moving averages. It also leverages principles of effective data visualization to communicate findings clearly.

**Procedure Used**: The general procedure followed in this project involves:

- Loading the dataset from a local machine.
- Performing initial data inspection (e.g., viewing the head of the DataFrame).
- Preprocessing the data as needed (e.g., converting date columns, extracting time components).
- Applying grouping and aggregation to summarize data by relevant time periods or categories.
- Generating various types of visualizations (line plots, heatmaps) to explore different aspects of the data.
- Interpreting the visualizations to identify trends, patterns, and anomalies.

**Purpose of Doing the Project**: The primary purpose of this project is to provide hands-on experience in applying data analysis and visualization techniques to time-series data. It aims to build proficiency in using Python and its libraries for data-driven storytelling, enabling the user to extract insights and communicate them effectively through visual means

# Topics Covered in Training (Weeks 1-2)

Based on the content and techniques demonstrated in this notebook, here is a list of topics that were likely covered during your initial training:

- Introduction to Python for Data Science
- Working with Pandas DataFrames (loading, cleaning, manipulation)
- Data Aggregation and Grouping
- Introduction to Data Visualization
- Using Matplotlib and Seaborn for plotting
- Creating interactive visualizations with Plotly Express
- Fundamentals of Time-Series Data Analysis
- Handling Date and Time Data
- Calculating Moving Averages
- Understanding and creating different chart types (Line plots, Bar plots, Heatmaps, Stacked Bar plots)
- Data Import and Export (CSV, Excel, Google Drive, Local Uploads)

# 3. Project Objective

The main objectives of this project are to illustrate and practice the following:

- **Understanding Time-Series Data:** To grasp the concept of time-series data and its importance in analyzing trends over time.
- **Data Loading and Preparation:** To demonstrate how to load datasets from different sources (like Google Drive and local machine) and prepare them for analysis, including handling date formats.
- **Exploratory Data Analysis (EDA) with Visualization:** To perform EDA using various visualization techniques (line plots, heatmaps) to identify patterns, trends, and seasonal variations in time-series data.
- **Applying Smoothing Techniques:** To understand and apply smoothing methods like moving averages to reveal underlying long-term trends in noisy data.
- **Comparative Analysis:** To compare trends from different sources (like GISTEMP and GCAG) or across different categories (like car sales by company and region).
- **Generalizing Analysis Techniques:** To show how the learned data analysis and visualization techniques can be applied to different time-series datasets with varying contexts (e.g., climate data vs. sales data).

# 4.Methodology

This project followed a structured approach to analyze time-series data using Python and its data science libraries. The methodology focused on data acquisition, cleaning, analysis, and visualization to extract meaningful insights from the datasets.

**1. Data Collection:**

Two primary datasets were used in this project:

- **Global Temperature Data:** This dataset, containing historical mean temperature anomalies from two sources (GCAG and GISTEMP), was loaded directly from a publicly shared Google Drive file. The data included 'Source', 'Date', and 'Mean' temperature anomaly values.
- **Car Sales Data:** This dataset, containing car sales information across different companies and regions over 10 years, was uploaded directly from a local machine into the Colab environment. The data included 'Year', 'Company', 'Units_Sold', and 'Region'.

No surveys or specific sampling methodologies were employed as the datasets were obtained from existing sources.

**2. Data Cleaning and Preprocessing:**

The raw data underwent the following preprocessing steps:

- **Loading:** Data was loaded into pandas DataFrames using pd.read_csv and pd.read_excel depending on the file format. For the Google Drive data, a direct download URL was constructed. For local files, google.colab.files.upload() was used.
- **Date Conversion:** For the temperature data, the 'Date' column was converted to datetime objects using pd.to_datetime to enable time-series analysis and extraction of year and month components.
- **Handling Missing Values:** While not explicitly shown as a separate step for these specific datasets, a general approach to handling missing values would involve identification (e.g., df.isnull().sum()) and appropriate strategies like imputation or removal, depending on the nature and extent of missing data.

**3. Data Analysis and Visualization:**

Exploratory Data Analysis (EDA) was a key component of the methodology, heavily relying on visualization to understand trends and patterns. The following steps and techniques were applied:

- **Initial Data Inspection:** The first few rows of the loaded DataFrames were displayed (df.head()) to get a preliminary look at the data structure and content. Basic information like shape, data types, and unique values for categorical columns were also inspected.
- **Trend Analysis (Line Plots):**
    - Line plots were generated to visualize the change in mean temperature over time for both data sources, allowing for a visual comparison of trends and identifying periods of warming or cooling.
    - Similar line plots were created to show the overall car sales trend across years and regional sales trends over time.
- **Smoothing (Moving Averages):** A rolling average with a window of 12 months (and later 120 months for stronger smoothing) was calculated for the temperature data using the .rolling().mean() function in pandas. This technique helped in smoothing out short-term fluctuations and highlighting long-term climate trends.
- **Seasonal Analysis (Heatmaps and Line Plots):**
    - Heatmaps were constructed to visualize seasonal temperature patterns across months and years for specified periods (last 50 and last 20 years). This involved extracting Year and Month from the Date column and pivoting the data to create a matrix suitable for a heatmap.
    - An interactive line plot using Plotly Express was also generated to show monthly temperature variations over time, providing an alternative visualization for seasonal patterns.
- **Categorical Analysis (Bar and Stacked Bar Charts):**

- Bar charts were used to visualize the total units sold per car company, allowing for a comparison of sales performance across different manufacturers.
- A stacked bar chart was created to show company-wise sales broken down by region, providing insights into regional contributions to each company's sales and the overall regional market distribution.

**4. Tools Used:**

The primary tools and libraries used for data analysis and visualization were:

- **Python:** The core programming language.
- **Pandas:** For data manipulation and analysis.
- **NumPy:** For numerical operations.
- **Matplotlib:** For creating static plots.
- **Seaborn:** For creating enhanced statistical plots, particularly heatmaps.
- **Plotly Express:** For creating interactive visualizations.
- **Google Colab:** The cloud-based environment where the code was executed.
- **Google Drive:** Used as a source for one of the datasets.

**5. Code Repository:**

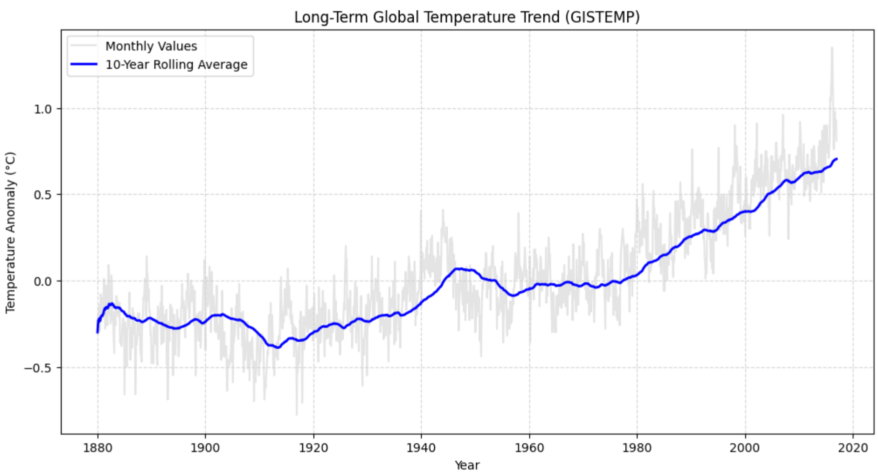
The Python code developed for this project can be found on GitHub at the following link:

[Insert GitHub Repository Link Here]

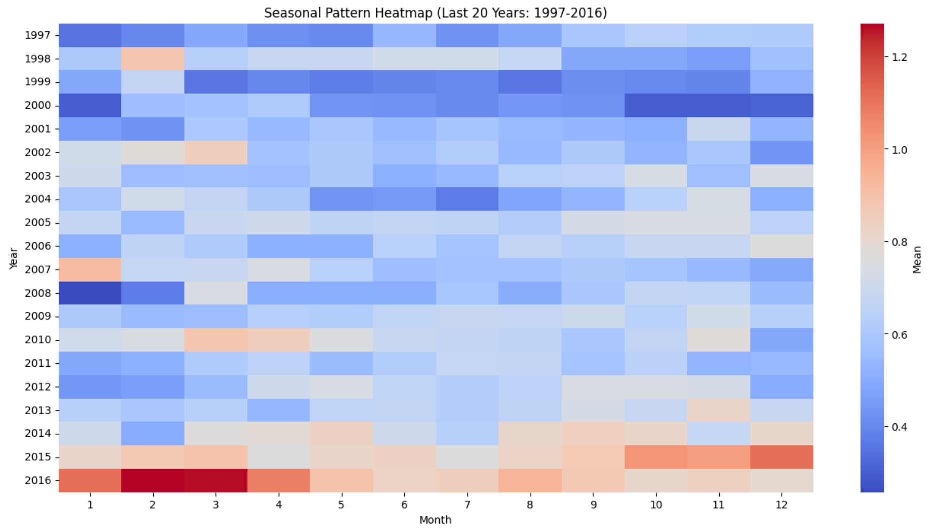
# 5.Data Analysis and Results

.

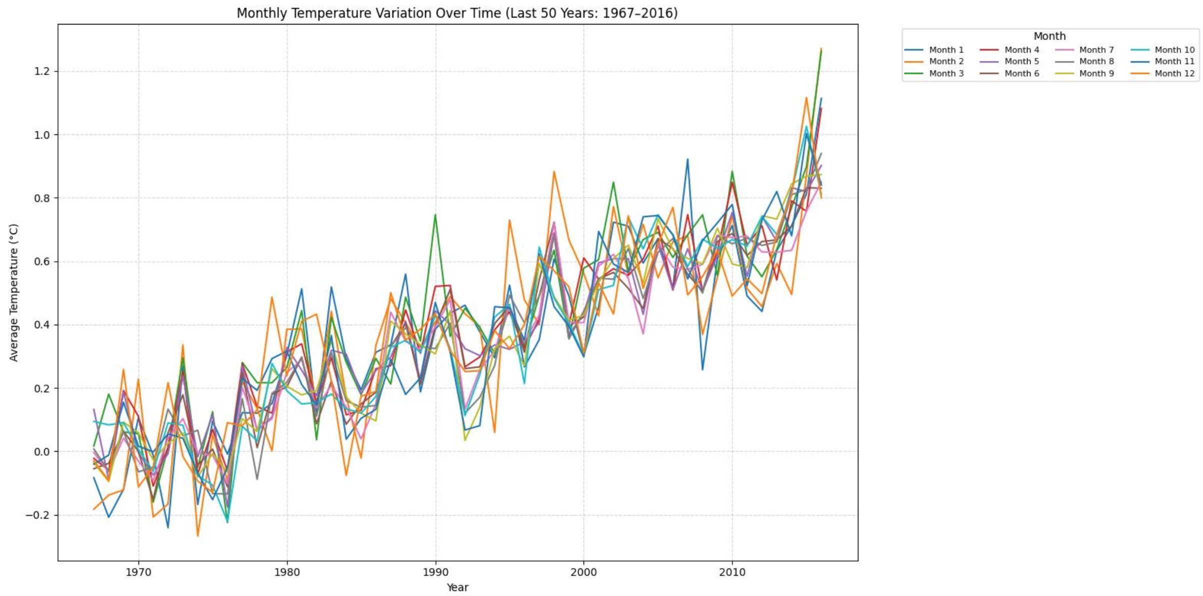
## Long-Term Global Temperature Trend (GISTEMP):-



## Seasonal Pattern Heatmap (Last 20 Years):-




## Monthly Temperature Variation Over Time (Last 50 Years)




# ASSIGNMENT

## CARS SALES IN DIFFERENT REGIONS OVER 10 YEARS

Saving car_sales_10_years_africa (1).csv to car_sales_10_years_africa (1) (7).csv

Data Loaded ✅

| | Year | Company | Units_Sold | Region |
|---|---|---|---|---|
| 0 | 2022 | Kia | 387425 | North America |

1 2018 BMW 192898 South America

2 2016 Kia 390713 North America

3 2018 Honda 387784 North America

4 2019 Ford 103164 Africa

============================================================

CAR SALES DATA ANALYSIS

============================================================

Dataset shape: (50, 4)

Years covered: 2015 to 2024

Companies: Kia, BMW, Honda, Ford, Tesla, Toyota, Hyundai, Tata

Regions: North America, South America, Africa, Europe, Asia

---

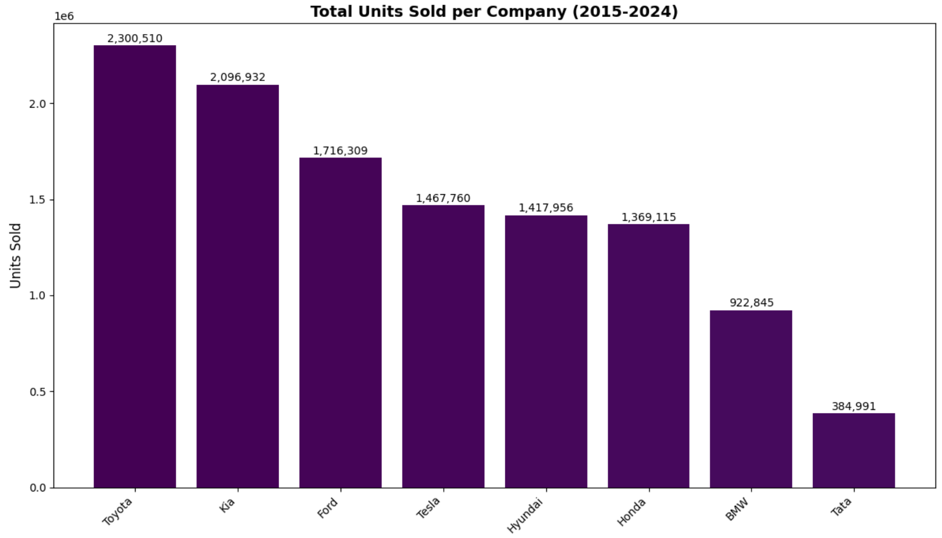
# BAR GRAPH :-

==================================================

COMPANY SALES SUMMARY (HIGHEST TO LOWEST)

==================================================

1. Toyota: 2,300,510 units
2. Kia: 2,096,932 units
3. Ford: 1,716,309 units
4. Tesla: 1,467,760 units
5. Hyundai: 1,417,956 units
6. Honda: 1,369,115 units
7. BMW: 922,845 units
8. Tata: 384,991 units




# Line graph:-

==================================================

YEARLY SALES SUMMARY

==================================================

2015: 1,189,747 units

2016: 776,848 units

2017: 2,086,503 units

2018: 1,575,447 units

2019: 821,553 units

2020: 1,708,526 units

2021: 922,208 units

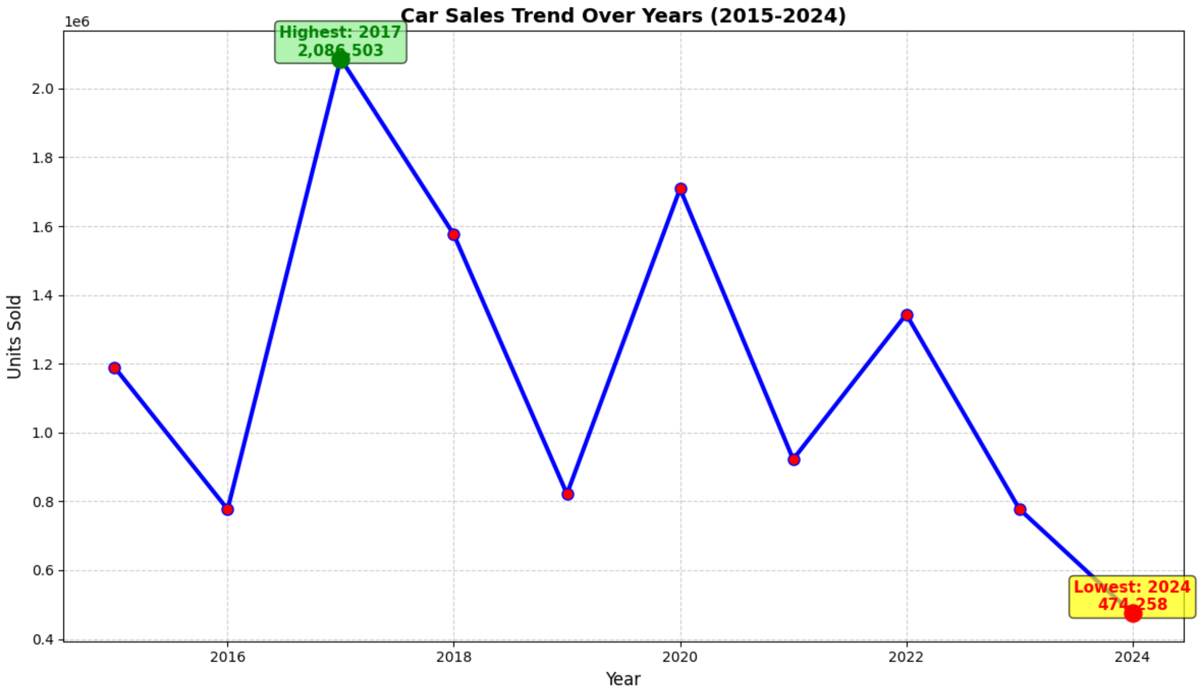
2022: 1,343,676 units

2023: 777,652 units

2024: 474,258 units

► Highest sales year: 2017 (2,086,503 units)

► Lowest sales year: 2024 (474,258 units)




==================================================

REGIONAL SALES SUMMARY

==================================================

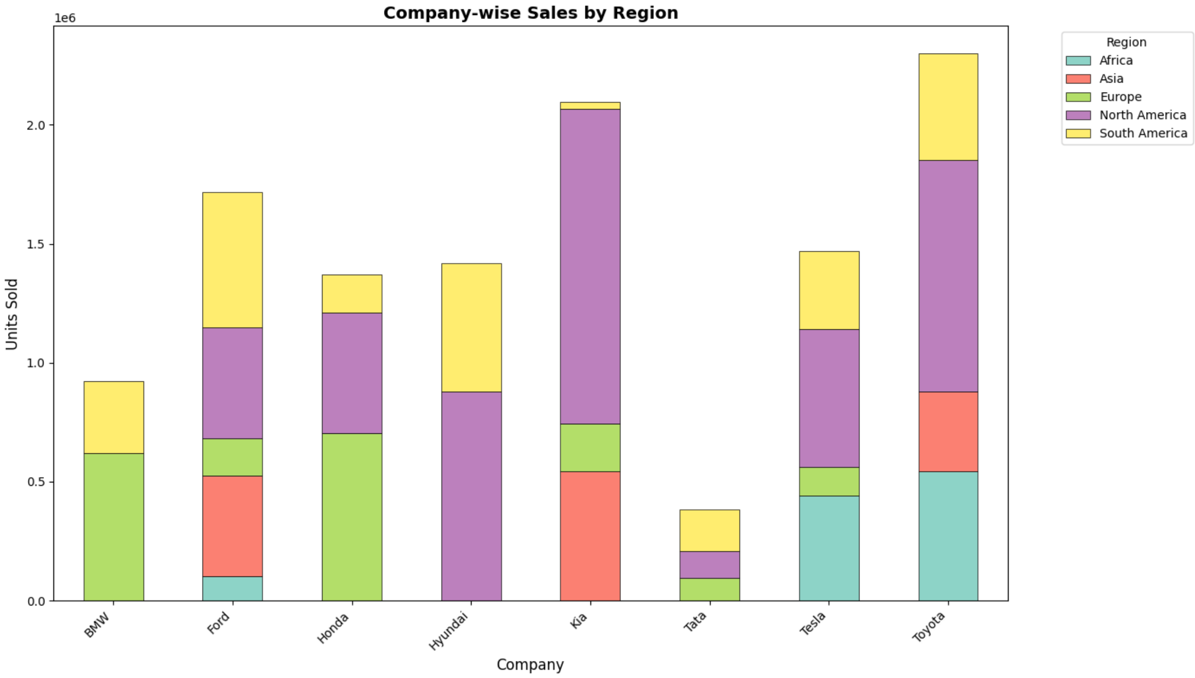
North America: 4,833,468 units

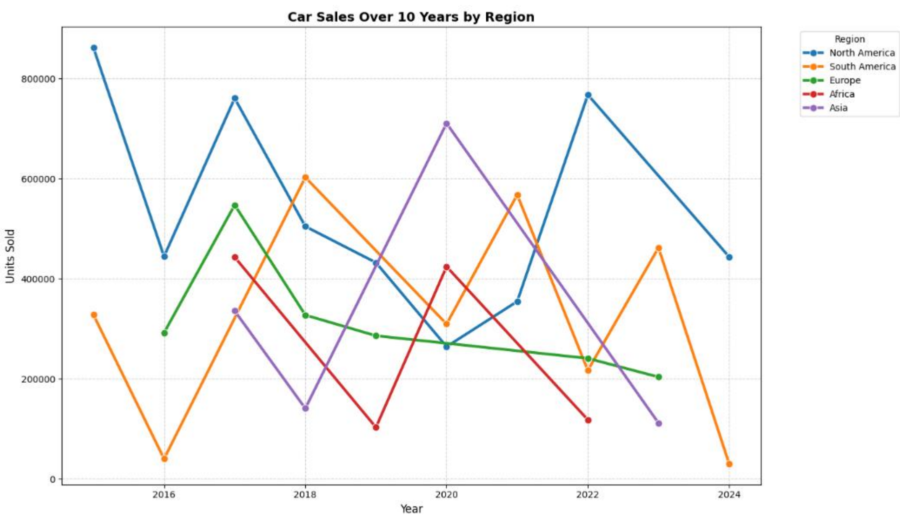
South America: 2,557,627 units

Europe: 1,897,587 units

Asia: 1,299,016 units

Africa: 1,088,720 units




==================================================

ADDITIONAL INSIGHTS

==================================================

Total cars sold (2015-2024): 11,676,418 units

Average yearly sales: 1,167,642 units

Number of sales records: 50

Most active region: North America (4,833,468 units)

Top company: Toyota (2,300,510 units)

# 6.Conclusion

**Global Temperature Data:**

- **Clear Warming Trend:** The analysis of historical temperature anomalies from both GCAG and GISTEMP sources clearly indicates a significant upward trend in global temperatures, particularly accelerating since the 1970s. The line plots and smoothed moving averages consistently demonstrate this long-term warming.
- **Source Consistency:** The high degree of similarity in the trends observed from the two independent data sources (GCAG and GISTEMP) strengthens the reliability of the findings and provides strong evidence for the observed climate change.
- **Seasonal Patterns:** The heatmaps effectively illustrate the seasonal variations in temperature anomalies across different months and years. While there are consistent seasonal cycles, the overall shift towards warmer anomalies in recent decades is evident across most months.

**Car Sales Data:**

- **Market Leaders and Regional Performance:** The bar charts and stacked bar charts provide a clear picture of the top-performing car companies in terms of total units sold and highlight the regional distribution of sales for each company and the overall market. North America appears to be the most active region in terms of total sales.
- **Sales Trends Over Time:** The line plots show the overall trend in car sales over the 10-year period and reveal how sales have varied by region year-on-year. This allows for the identification of periods of growth or decline and understanding regional contributions to the overall market dynamics.

**Overall:**

This project successfully demonstrated the application of various data analysis and visualization techniques to time-series datasets. By employing Python and libraries like pandas, Matplotlib, Seaborn, and Plotly Express, we were able to uncover meaningful trends, patterns, and insights from both climate and sales data. The interactive visualizations, in particular, enhanced the ability to explore the data dynamically.

**Recommendations for Future Work:**

- **More Advanced Time-Series Analysis:** Explore more sophisticated time-series forecasting techniques (e.g., ARIMA, Prophet) to predict future temperature anomalies or car sales.
- **Incorporate Additional Features:** Integrate other relevant datasets, such as economic indicators, population growth, or specific climate events, to potentially identify other factors influencing temperature or sales trends.
- **Deep Dive into Regional Analysis:** Conduct a more in-depth analysis of specific regions in the car sales data to understand localized market dynamics, consumer preferences, or economic impacts.
- **Interactive Dashboards:** Create interactive dashboards using tools like Dash or Streamlit to allow for more dynamic exploration and presentation of the findings.

- **Investigate Anomalies:** Further investigate any significant anomalies or outliers observed in the data to understand their underlying causes.

# 7. APPENDICES

## GITHUB LINK (CONTENTS -DATA STES, VIDEO, GOOGLE COLAB CODE)

[https://github.com/shawsatyajit0-ui/VISUALIZING-GLOBAL-TEMPERATURE-TRENDS-]

## GOOGLE COLAB LINK

**[https://colab.research.google.com/drive/147lke5bICfVfF60qXYev1itHZ7bjLpfs?usp=sharing]**