January 2025

Interactive Dashboard for E-Commerce Sales Analysis

Data Science Lab

Mahmoud Ferig	918535
Mohammad Noaman Khan	919598
Manesh Manesh	910029
Gauri Ramdasji Thote	910129

Professor Marco Fattore - Professor Silvio Gerli

UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Abstract

This report presents the development and findings of an interactive dashboard built using Streamlit for the analysis of e-commerce time series sales data. The primary objectives were to analyze sales trends, assess the impact of external factors, and provide actionable insights to improve decision-making for e-commerce businesses.

Key aspects of the project include:

- Comprehensive data enrichment, including temporal features, lagged metrics, rolling averages, economic indicators, and holiday effects, to enhance the dataset's analytical utility.
- Advanced time series decomposition to isolate trends, seasonal patterns, and residuals, offering a detailed understanding of sales behavior for each sector.
- Correlation analysis and statistical tests to quantify the influence of external factors, such as holidays, GDP, inflation, and COVID-19 lockdown periods, on sales performance.
- Integration of findings into a dynamic, interactive dashboard, allowing users to explore sector-specific patterns, assess correlations, and compare sales across different conditions.

The dashboard provides practical insights, such as identifying peak sales periods during holidays, analyzing the recovery patterns of different sectors post-lockdown, and uncovering long-term trends across the dataset.

Resources:

- **GitHub Repository:** GitHub Link
- **Colab Notebook:** Colab Notebook Link

This work demonstrates the value of combining enriched datasets, advanced analytics, and interactive visualizations for effective e-commerce decision-making. Recommendations for future work include incorporating predictive analytics and automating real-time data processing for continuous insights.

Introduction

Dataset

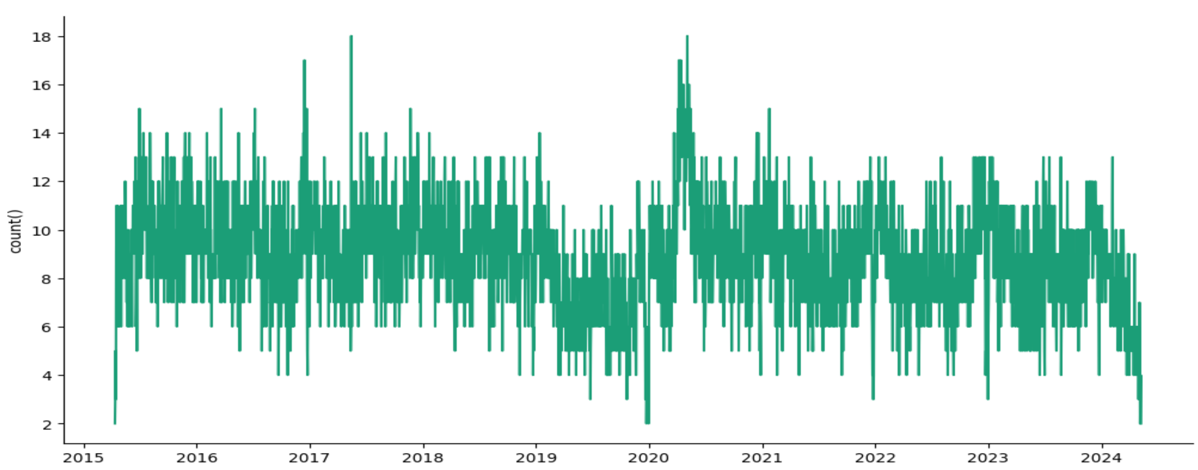
The dataset used in this project consisted of historical sales data from an e-commerce platform, covering a period of **11 years (2013–2024)**. However, due to inconsistencies and missing data in earlier years, the analysis focused on the period from **April 2015 to May 2024**, which offered more reliable and complete records.

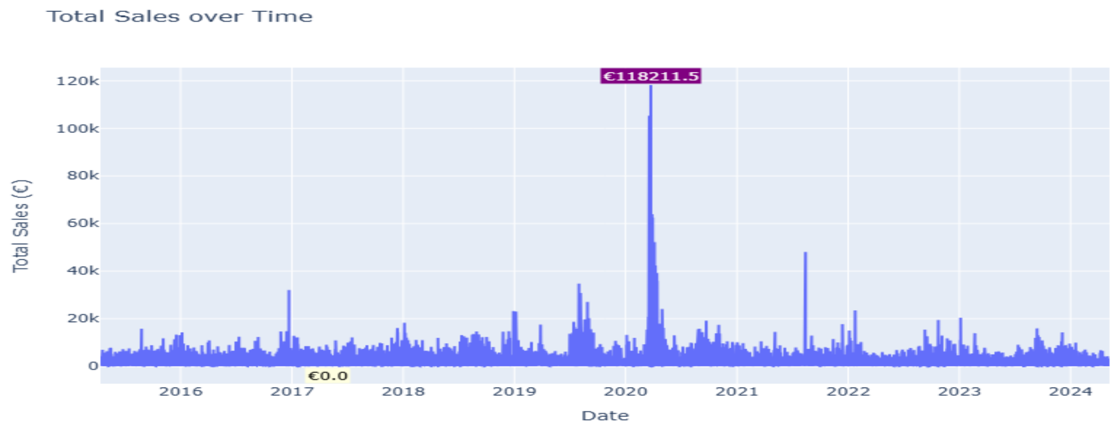
Key steps in dataset preparation included:

- **Data Cleaning:**
  - Removed null values and duplicate records to ensure data integrity.
  - Addressed inconsistencies in date formats and variable definitions.
- **Data Transformation:**
  - Normalized numerical variables to improve comparability.
  - Encoded categorical features, such as holidays and event types, into binary indicators for analysis.
- **Data Enrichment:**
  - **Temporal Features:** Added fields such as day of the week, week of the year, weekend indicators, and seasonality markers.
  - **Lagged Metrics and Rolling Averages:** Calculated metrics like sales from the previous day or week and applied 7-day rolling averages to smooth fluctuations.
  - **Holiday and Event Data:** Integrated external datasets for major holidays (e.g., Black Friday, Christmas) and events like the World Fishing Championship and Serie A Matches.
  - **Economic Indicators:** Merged data on GDP, inflation, and unemployment from ISTAT, aligning these with the dataset's timeline.

Descriptive Statistics

- **Number of Records: 29,536**
- **Number of Features: 3** original, expanded to **32** through enrichment.
- **Key Variables:**
  - **Total Sales:** Mean = **€2,581.63**, Std Dev = **€5,683.05**
  - **Customer Transactions:** Max = **18/day**, Min = **1/day**.
- **Sales Concentration:** The top six sectors accounted for **83.49%** of total sales.





Idea and Motivations

The primary motivation behind this analysis is to gain insights into the temporal patterns of e-commerce sales and identify potential factors influencing these patterns. Understanding such trends is crucial for businesses to optimize inventory management, marketing strategies, and overall sales forecasting. The enriched dataset allows stakeholders to:

- Identify patterns and trends effectively.
- Perform advanced analytics using dynamic dashboards.
- Make informed decisions with actionable insights.

Aim

This report accompanies an interactive dashboard, providing a dynamic and explorative view of the data and analysis results. The dashboard allows users to delve deeper into specific sectors, time periods, and external factors, fostering a more comprehensive understanding of the e-commerce sales dynamics.

The main aim of this project is to:

1. Perform a comprehensive time series analysis of e-commerce sales, exploring trends, seasonality, and cyclical patterns.
2. Investigate the impact of external factors, such as holidays, economic indicators, and the COVID-19 pandemic, on e-commerce sales.
3. Develop a robust understanding of the dynamics driving sales in different product sectors.

Methodologies

Algorithms and Methods Used

1. **Data Enrichment Techniques:**
   - **Principles:** Enhance dataset utility by incorporating derived features, external data, and contextual information to capture temporal, economic, and behavioral patterns.
   - **Theory/Formulas:**
     - **Temporal Features:** Extract patterns based on the day of the week, week of the year, weekend, and seasonal trends.
     - **Lagged and Smoothed Metrics:**
       - Lagged Metrics: Use prior values of key variables to capture temporal dependencies.
         - Rolling Averages: Smooth out short-term fluctuations by calculating the average over a specific window.
     - **Holiday Features:** Incorporate information about holidays, pre-holiday periods, and their potential effects on behavior.
     - **Economic Indicators:** Use macroeconomic data such as GDP, inflation, and unemployment to provide contextual understanding.
     - **Contextual Events:** Capture the impact of significant external events, such as lockdown periods and high sales sectors events.
   - **Pros and Cons:**
     - Pros:
       - Enhances dataset depth and analysis capabilities.
       - Provides richer insights by integrating contextual and external factors.
     - Cons:
       - Requires domain expertise for effective feature engineering.
       - Risk of over-engineering or introducing noise if features are not relevant.

2. **Data Visualization:**
   - **Principles:** Dynamic and interactive representation of trends and patterns.
   - **Tools:** Plotly, Matplotlib.
   - **Pros and Cons:**
     - Pros: High interactivity and user engagement.
     - Cons: Potential performance issues with very large datasets.

3. **Time Series Decomposition:**
   - **Principles:** Time series decomposition separates a time series into its fundamental components: trend, seasonality, and residual. This helps in understanding the underlying patterns and isolating the effects of each component.
   - **Theory and Formulas:** The additive model, used here, assumes the time series can be expressed as the sum of its components:

$$Y(t) = Trend(t) + Seasonality(t) + Residual(t)$$

where $Y(t)$ is the observed value at time t.

- **Pros and Cons:** Decomposition provides a simple and intuitive way to visualize and interpret time series data. However, it may not be suitable for complex, non-linear patterns. Alternatives include more advanced methods like ARIMA modeling or state-space models, which can capture more intricate relationships within the data.

4. **Correlation Analysis:**
   - **Principles:** Correlation analysis measures the strength and direction of the linear relationship between two variables. It helps in identifying potential influencing factors on sales, such as holidays or economic indicators.
   - **Theory and Formulas:** The Pearson correlation coefficient (r) quantifies the linear association between two variables:

$$r = \frac{Cov(X,Y)}{(\sigma_X \times \sigma_Y)}$$

     where $Cov(X, Y)$ is the covariance between variables X and Y, and $\sigma_X$and σ_Y are their standard deviations.
   - **Pros and Cons:** Correlation analysis is a widely used and easily interpretable method for assessing relationships between variables. However, it only detects linear relationships and may not capture more complex dependencies. Alternatives include non-linear correlation measures or causal inference techniques.

5. **Statistical Tests (t-test):**
   - **Principles:** Statistical tests assess the significance of observed differences between groups. In this project, the t-test was used to compare sales during lockdown and transitional periods with non-lockdown periods.
   - **Theory and Formulas:** The t-test compares the means of two groups:

$$t = \frac{(mean_1 - mean_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

     where $mean_1$ and $mean_2$ are the group means, S_1 and S_2 are the group standard deviations, and n_1 and n_2 are the group sizes.
   - **Pros and Cons:** The t-test is a robust and widely applicable test for comparing group means. However, it assumes normality of the data and may not be appropriate for non-normal distributions. Alternatives include non-parametric tests like the Mann-Whitney U test, which are less sensitive to distributional assumptions but may have lower power.

Experiments

Detailed methodologies and outcomes are outlined for each experiment, with implementation details available in the accompanying notebook file.

Experiment 1: Data Enrichment

- **Objective:**
  Enhance dataset utility by deriving additional features and metrics for deeper insights and analysis.
- **Method:**
  - **Self Data Extension:** Derived temporal features and lagged sales metrics to capture patterns and seasonality:
    - Day of Week (`Day_of_Week`)
    - Week of Year (`Week_of_Year`)
    - Weekend Indicator (`Is_Weekend`)
    - Seasonal Indicator (`Season`)
    - Lagged Sales Metrics (`Total_Sales_Lag1`, `Total_Sales_Lag7`)
    - Rolling Average for 7 Days (`Total_Sales_Rolling7`)
  - **Holiday Enrichment:** Incorporated holiday-related indicators using API to account for their impact on sales:
    - Is Holiday (`Is_Holiday`)
    - Is Pre-Holiday (7 days before a holiday, `Is_Pre_Holiday`)
  - **Economic Indicators:** Enriched the dataset with external economic data from ISTAT to understand broader influences:
    - Gross Domestic Product (`GDP`)
    - Inflation Rate (`Inflation`)
    - Unemployment Rate (`Unemployment`)
  - **Covid-19 Lockdown Impact:** Added lockdown status to assess pandemic effects (`Lockdown Status`).
  - **High Sales Sectors Events:** Incorporated significant events to analyze their impact on top-performing sales sectors. For example, added events like the *World Fishing Championship* and *National Fishing Day* for the *Pesca* sector, and events such as *Serie A Matches, Champions League Matches, Europa League Matches*, and *Euro/World Cup Events* for the *Calcio* sector.
  -
- **Parameters:**
  - Moving average window: 7 days.
- **Reasoning:**
  By enriching the data with temporal, external, and contextual features, we can better model patterns, account for anomalies, and support robust visualization and decision-making.

Experiment 2: Time Series Decomposition of Sector Sales

- **Objective:**
  The primary goal of this experiment was to decompose the total sales time series for each product sector into three key components—trend, seasonal, and residual. This decomposition provides a clearer understanding of underlying sales patterns, helping to identify long-term movements, periodic fluctuations, and irregular variations.

- **Method and Parameters:**

  - **Decomposition Technique:**
    The $seasonal\_decompose$ function from the $statsmodels$ library was utilized to separate the components of the time series.

  - **Model Assumption:**
    Additive Model (model='additive'): The sales data was assumed to follow an additive structure, where the observed data can be expressed as:

    $Yt = Tt + St + Rt$

    Here, Yt is the observed value, Tt is the trend, St is the seasonal component, and Rt is the residual or noise. This model is suitable for sales data with consistent seasonal fluctuations and a trend that evolves linearly over time.

  - **Frequency Parameter (freq):**

    – A **monthly frequency** was selected to reflect typical sales cycles in e-commerce. This corresponds to the expectation of recurring patterns within each calendar month.

    – The parameter was set based on the dataset's time granularity and domain knowledge of sector-specific sales trends.

  - **Visualization:**

    – Each component (trend, seasonal, and residual) was plotted to provide a visual representation of the decomposition process and to aid interpretation.

    – Separate decompositions were performed for each sector, ensuring tailored insights.

- **Reasoning:**
  Decomposing time series data enables the isolation of distinct sales patterns. By understanding these components:

  – **Trends** highlight long-term movements, such as growth or decline in sector sales.

  – **Seasonality** reveals periodic variations driven by predictable factors like holidays or monthly cycles.

  – **Residuals** capture irregular or unexpected changes, offering insights into anomalies or noise.

Experiment 3: Correlation Analysis of Sales with External Factors

- **Objective:**
  This experiment aimed to assess the relationship between total sales and various external factors, such as holidays, economic indicators (GDP, inflation, unemployment), Covid-19 lockdown periods, and some sectors events. Understanding these correlations helps evaluate how external influences impact sales patterns and can guide strategic decisions for e-commerce operations.

- **Method and Parameters:**
  - **Correlation Metric:**
    - This metric is particularly suitable for capturing linear dependencies between numerical variables such as sales and economic indicators.
    - The **Pearson correlation coefficient** was calculated to measure the linear relationship between total sales and each external factor. The Pearson coefficient (r) ranges from -1 to 1, where:

$$r > 0: Positive\ linear\ relationship.$$
$$r < 0: Negative\ linear\ relationship.$$
$$r = 0: No\ linear\ relationship.$$

  - **Library and Function Used:**
    - The $\boldsymbol{corr}()$ function from the $\boldsymbol{pandas}$ library was employed for efficient computation of correlation coefficients.
    - No specific parameters were required for this function.

  - **Preprocessing:**
    - Ensured all variables were properly aligned with the sales data based on the date.
    - Converted categorical features (e.g., holidays and lockdowns) into binary indicators (1 for presence, 0 for absence).

- **Reasoning:**
  By analyzing correlations, we can:
  - Identify which external factors have the strongest association with sales.
  - Determine whether these relationships are consistent across different sectors or time periods.
  - Uncover insights into how macroeconomic trends, holidays, and pandemic-related disruptions influence consumer behavior.

Experiment 4: Statistical Tests for Impact Analysis

- **Objective:**
  This experiment employed statistical tests to assess the significance of differences in sales under various conditions. The analysis covered multiple aspects, including the impact of lockdown periods, holidays, economic shifts, and other contextual factors, to understand how these conditions influenced sales patterns.

- **Method and Parameters:**
  - **Statistical Test Used:**
    - The **independent samples t-test** was applied to compare the mean sales between groups defined by different contextual factors:
      1. Lockdown vs. non-lockdown periods.
      2. Holiday vs. non-holiday periods.
      3. High vs. low GDP periods.
      4. Other factor-based groupings as defined by the dataset.
    - The test assumes normal distribution and similar variances in the groups, conditions that were checked and addressed as needed.
  - **Library and Function Used:**
    - The $\boldsymbol{ttest\_ind}$ function from the $\boldsymbol{scipy.stats}$ library was used for the analysis.
  - **Parameter Settings:**
    - $\boldsymbol{nan\_policy = 'omit'}$: Missing values in sales data were omitted to maintain accuracy.
    - Default equal variance assumption: Variances between groups were assumed equal unless proven otherwise.

- **Reasoning:**
  Statistical tests like the t-test provide a systematic approach to determining whether observed differences in sales under varying conditions are statistically significant. This allows for:
    - Identifying key external and temporal factors that influence sales.
    - Quantifying the impact of different events or economic scenarios.
    - Informing data-driven strategies to optimize sales under various conditions.

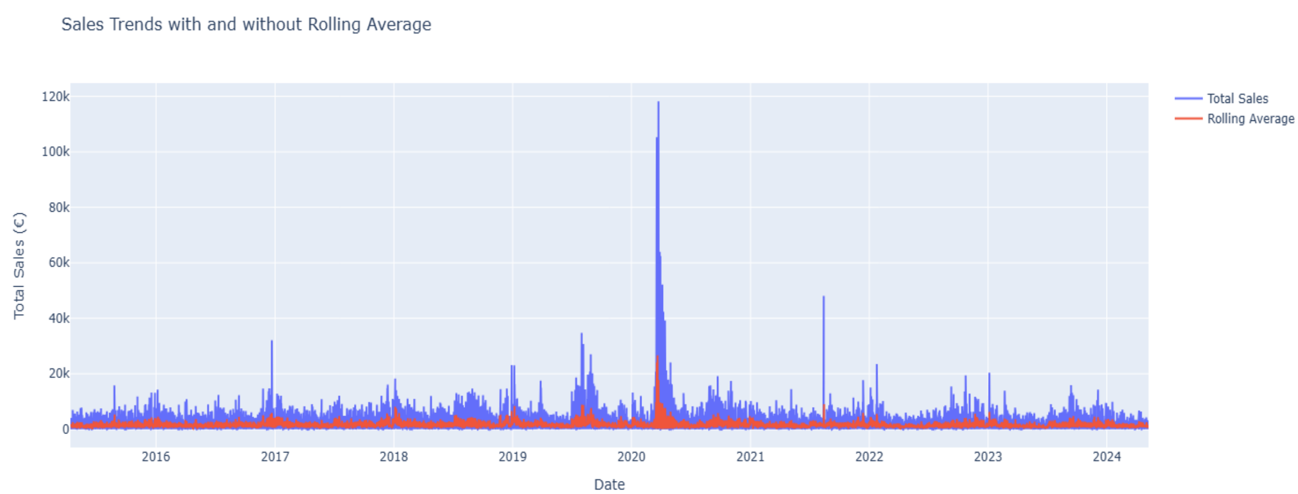
Results

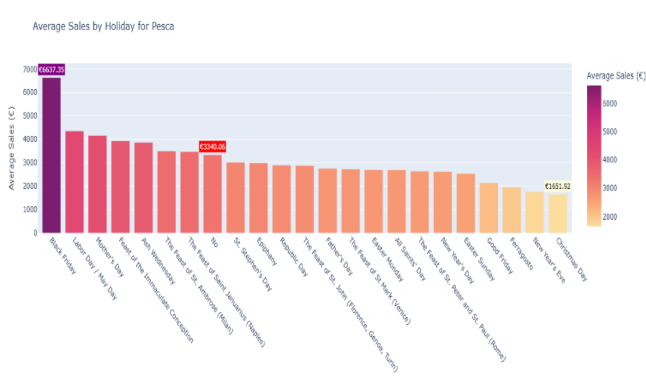
This section presents the results of the experiments conducted, utilizing easily understandable graphs and tables to illustrate the key findings.

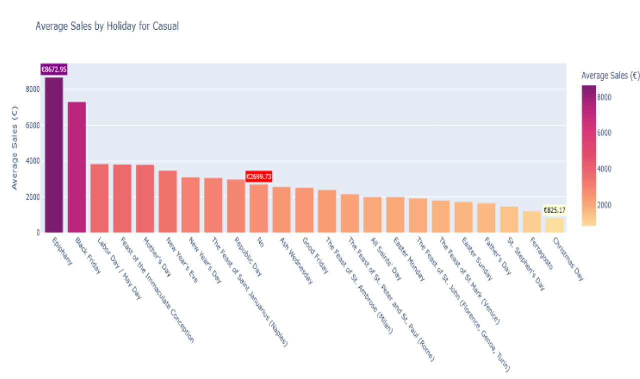
1. Data Enrichment

Temporal Patterns, Sales Trends, and Holidays

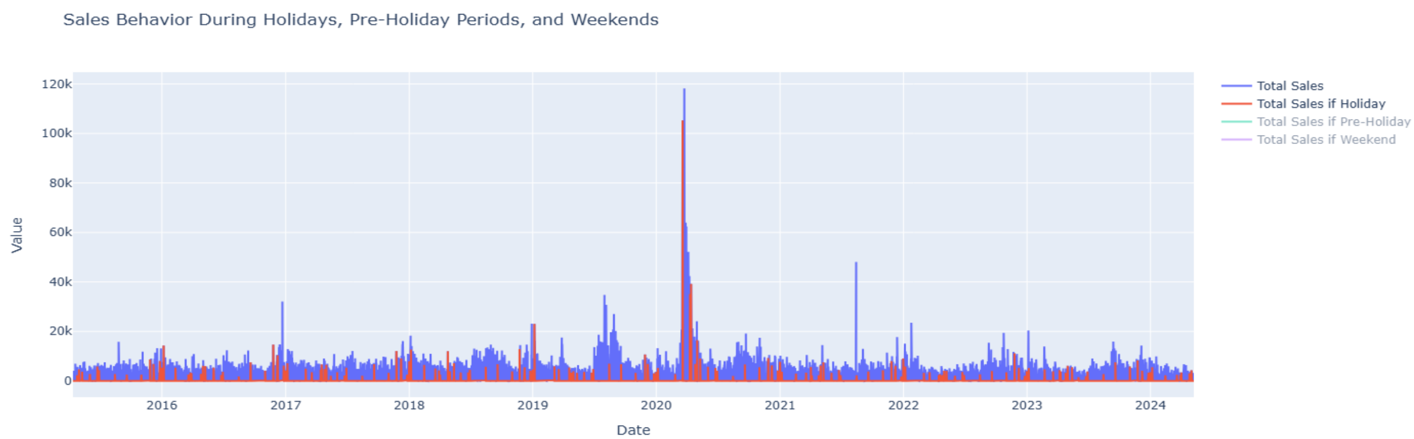
- **Findings:**
  - The **7-day rolling average** smoothed short-term fluctuations, making it easier to spot underlying trends. The rolling average was particularly useful in identifying long-term growth or declines in sales, removing noise from the data.
  - Temporal features such as **Holiday**, **Pre Holiday**, and **Is Weekend** revealed trends in sales behavior. Sales tended to increase during weekends and before holidays, with marked seasonality in certain sectors.
  - **Lagged sales metrics** (Total Sales from the previous day or week) helped capture short-term fluctuations and trends, proving useful in time series forecasting.
- **Visualization:**

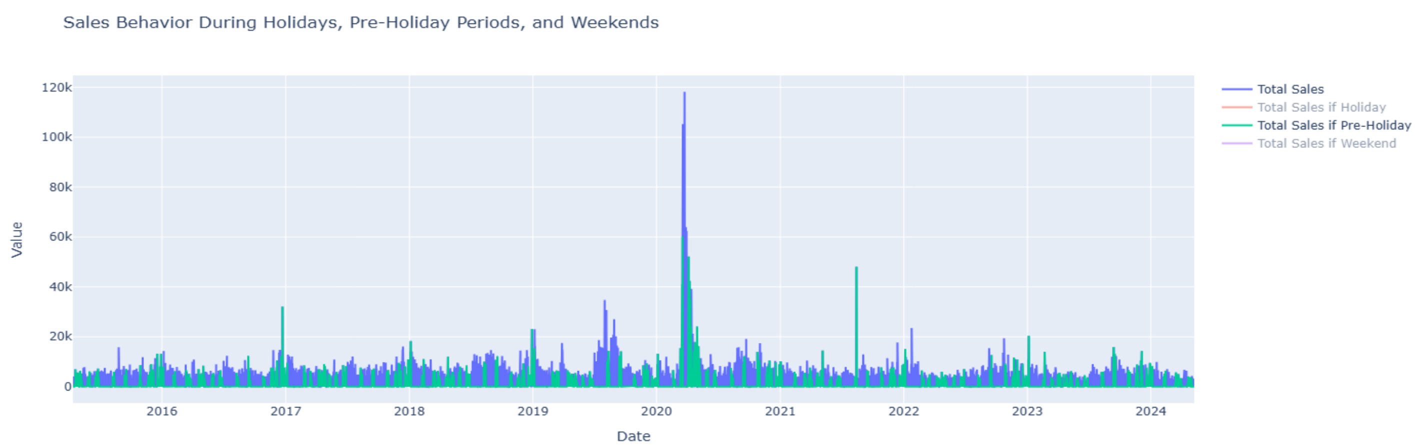


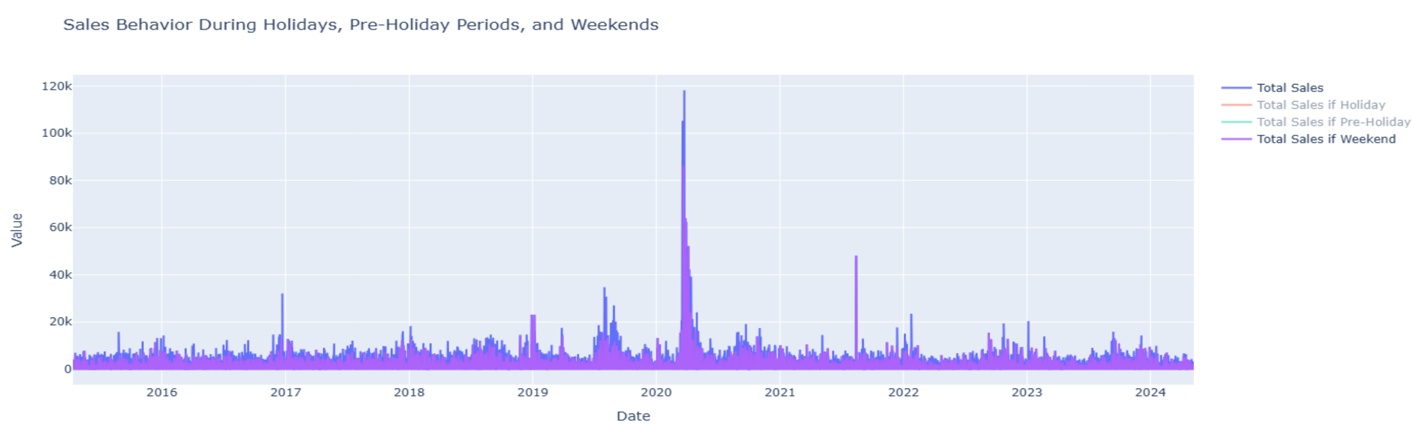




	Total Sales	***On Holiday***	***Pre-Holiday***	***Weekend***
Total Sales	50783718.45	2980218.18	15116446.89	12924971.91
Percentage	100%	5.87%	29.77%	25.45%





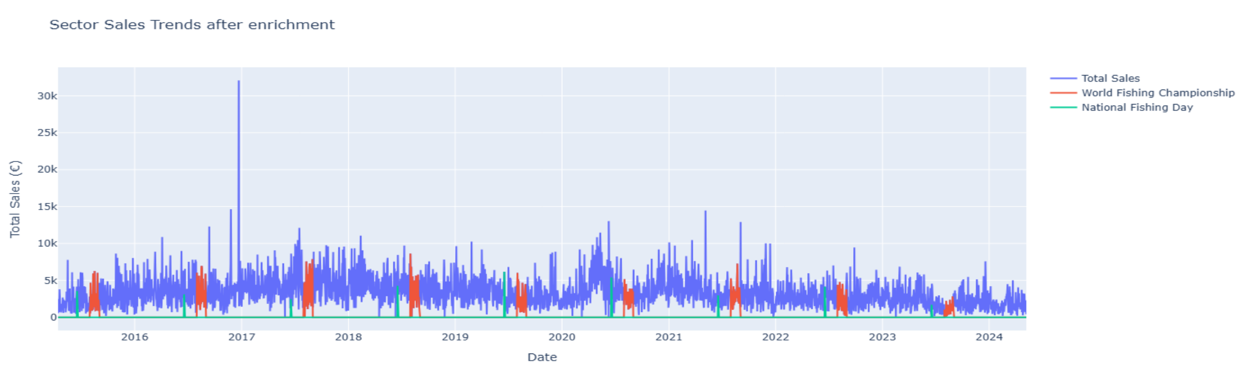


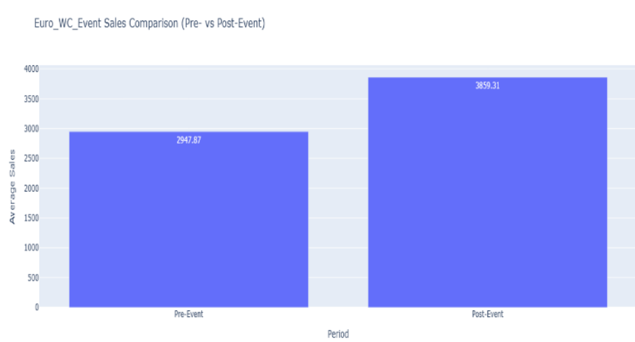
Sectors Events Impacts

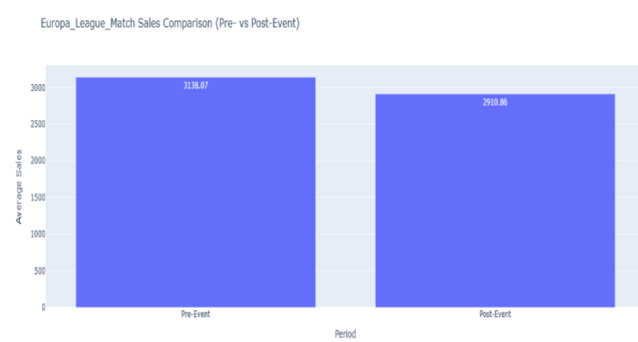
- **Findings:**
    - Events like **National Fishing Day** (Pesca) and **Serie A Matches** (Calcio) had substantial positive impacts on sales, as shown by increased sales on these days.
    - Specific events were integrated, showing higher spikes during these events compared to normal periods. For instance the nation football tournaments showed increase in sales after the event while for the **Serie A Matches**, it happens before and in the beginning of the event.

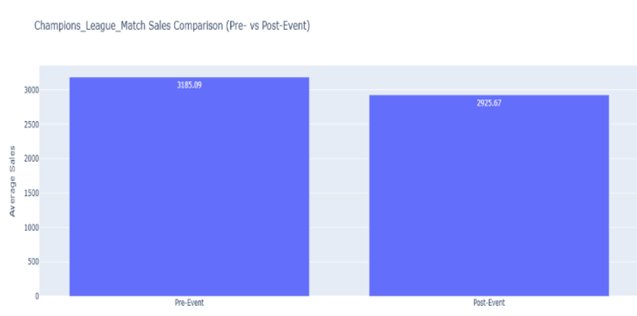
	Normal Days	World Fishing Championship	National Fishing Day
Average Sales	3372.93	2812.63	
Average Sales	3324.33		3754.40
T-statistic		-4.54	0.65
P-value		5.842632542604447e-06	0.52

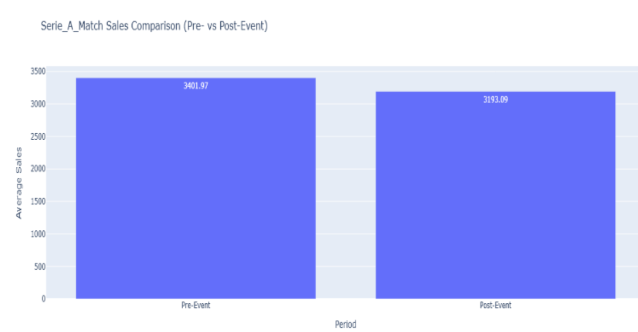
- **Visualization:**











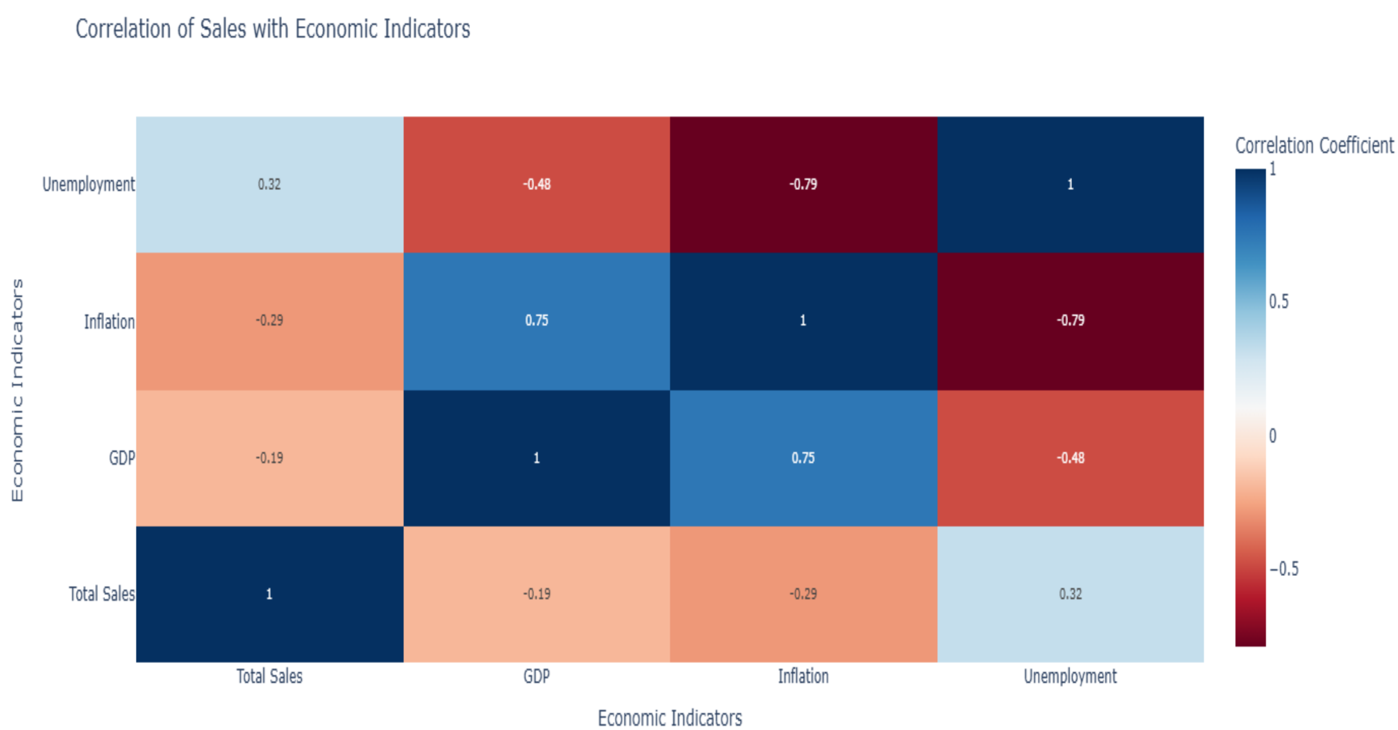
Economic Indicators and Their Influence on Sales

- **Findings:**
  - Economic indicators such as **GDP** and **Inflation** showed strong correlations with sales, though their impact varied by sector. GDP increases were positively correlated with sales, while inflation showed mixed effects.
  - **Unemployment** was inversely related to sales, with higher unemployment associated with mixed effects on sales, particularly in Casual sector.
  - **If we take a closer look at Casual Sector results:**
    - **Total Sales and GDP: -0.1933**

      This negative correlation indicates a slight inverse relationship between the Casual Sector sales and GDP, meaning that as the economy grows, the sales in the Casual Sector tend to decrease slightly.
    - **Total Sales and Inflation: -0.2853**

      This moderate negative correlation suggests that inflation negatively affects sales in the Casual Sector. As prices rise, consumer spending might decrease, affecting sales.
    - **Total Sales and Unemployment: 0.3209**

      This positive correlation indicates that when unemployment increases, total sales in the Casual Sector also tend to increase. This could be due to people spending on basic needs during times of economic downturn.
- **Visualization:**



Contextual Interpretation:

The enrichment of the dataset allowed us to identify multiple drivers of sales patterns, both short-term and long-term.

- **Temporal features** revealed clear weekly and seasonal trends, while **holiday and event impacts** demonstrated specific days of high sales.
- **Economic indicators** highlighted how broader economic conditions influence sector sales, which is crucial for forecasting and strategy.
- The **rolling averages** allowed for better trend visualization by reducing volatility, making it easier to predict future sales.

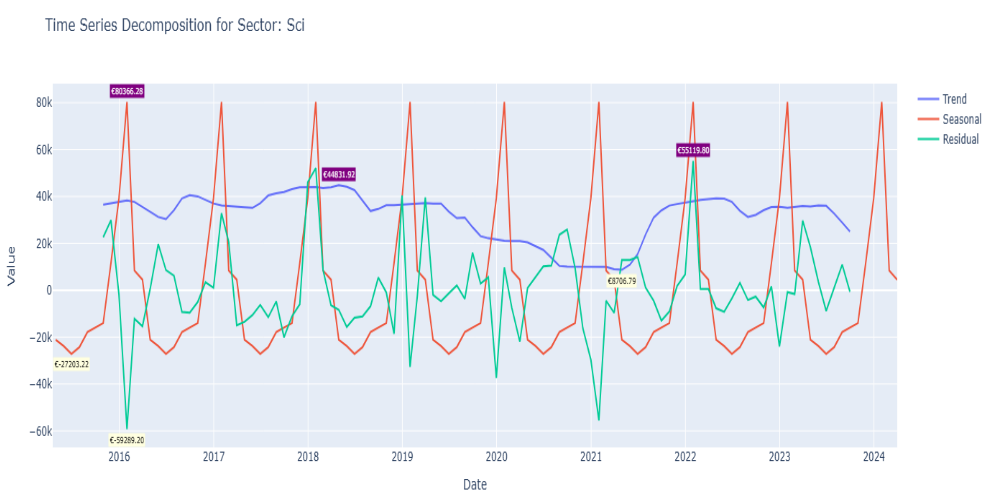
2. Time Series Decomposition of Sector Sales

Findings: Time series decomposition revealed distinct patterns in total sales for different product sectors.

- The **'Pesca' sector** demonstrated a steady upward trend, indicating consistent growth, potentially due to rising interest in fishing-related activities and events.
- The **'Calcio' sector** exhibited pronounced seasonality, with sales peaking during specific months aligned with major sporting events like championships and league matches.
- For the **'Sci' sector**, While the seasonal component was calculated to be very close to zero, the visible peak in winter sales suggests that there is, in fact, a strong seasonal trend despite the statistical results. This kind of pattern often occurs when the seasonal effect is consistent each year but perhaps not large enough to be captured by statistical decomposition methods.

Visualization: Interactive decomposition plots for each sector are provided in the accompanying dashboard. These plots allow for a detailed examination of the trend, seasonal, and residual components, enabling users to identify key periods of sales growth, seasonality patterns, and potential outliers.

Example for 'Sci' sector:



COMPONENT	VALUE
TREND	€31,810.38
SEASONAL	€0.00
RESIDUAL	€178.49

3. Correlation Analysis of Sales with External Factors

Findings: The correlation analysis provided insights into how external factors influenced total sales:

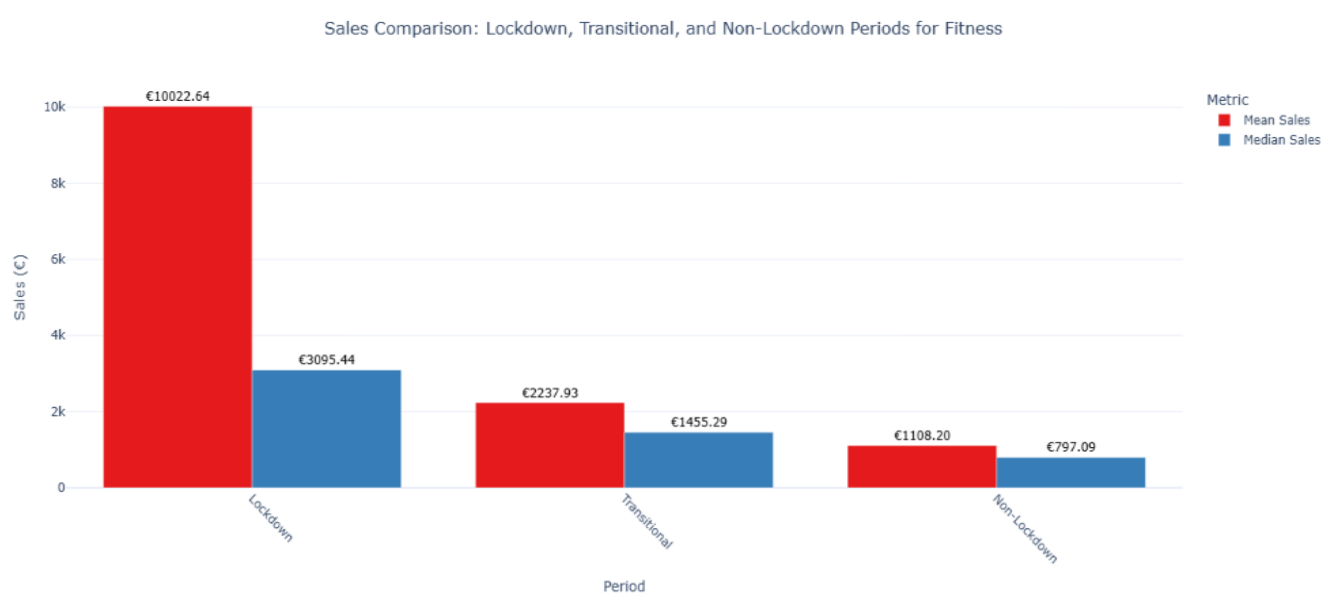
- **Holidays**: Events like **Black Friday** and **Christmas** were strongly positively correlated with sales, confirming their impact on purchasing behavior.
- **Lockdown Periods**: A negative correlation was observed, indicating that restrictions during COVID-19 negatively affected e-commerce sales. However, the degree of this impact varied across sectors.

These correlations highlight the importance of aligning marketing and inventory strategies with external factors to optimize sales outcomes.

Holidays Impact for Fitness Sector:

Metric	Value
Correlation with Holiday (Is_Holiday)	-0.028
Correlation with Pre-Holiday (Is_Pre_Holiday)	-0.01

Visualization: A positive correlation with lockdown status indicates that sales during the lockdown periods are somewhat related to higher sales, possibly due to increased engagement in home fitness activities or fitness-related products during the lockdown.



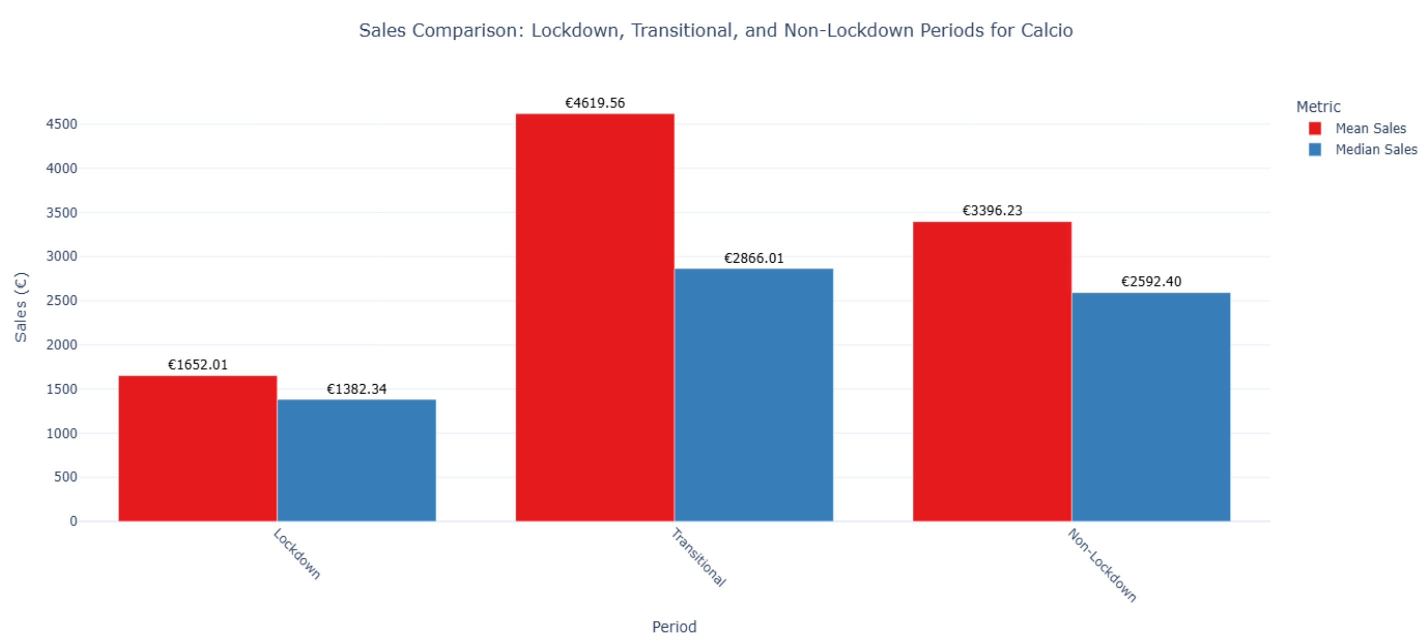
4. Statistical Tests for Impact Analysis

Findings: Statistical tests, including **t-tests**, provided quantitative evidence of how various factors impacted sales across different periods and sectors:

- **Lockdown Periods**:
  - Sales were lower for most of the sectors during lockdown periods compared to non-lockdown periods, confirming the negative correlation observed earlier.
- **Transitional Periods**:
  - Some sectors, like **Calcio**, experienced a rebound during transitional periods, likely due to eased restrictions and seasonal factors.
  - Other sectors remained below pre-lockdown levels, reflecting varied recovery dynamics.
- **Holidays and Other Events**:
  - Sales during holiday periods showed statistically increases compared to regular days.

These findings provide a robust basis for identifying critical time frames for operational adjustments and marketing strategies.

Visualization: Bar charts comparing average sales during lockdown, transitional, and non-lockdown periods for each sector are included in the dashboard. These charts provide a clear visual representation of the lockdown's impact on sales across different sectors.



Conclusions

The analysis of e-commerce sales data provided comprehensive insights into sales patterns, the impact of external factors, and the dynamics of different product sectors. Key conclusions drawn from this study include:

1. **Sector-Specific Patterns:**
    - Each product sector exhibited distinct sales behaviors, driven by unique trends and seasonality.
    - **Pesca Sector:** Demonstrated consistent growth, potentially due to increased interest in outdoor activities and targeted events like the World Fishing Championship. This highlights opportunities for sustained growth through event-focused marketing.
    - **Calcio Sector:** Showed pronounced seasonality, with sales peaking during major sporting events like Serie A Matches and the Champions League and After Nation Tournaments. These findings underscore the importance of aligning inventory and promotions with event calendars.
    - **Sci Sector:** Experienced sharp seasonal peaks in winter, driven by skiing activities. Sales trends revealed a steady year-over-year increase, suggesting rising demand for winter sports products.

2. **Impact of External Factors:**
    - **Holidays:** Major holidays, such as Black Friday and Christmas, were strongly associated with increased sales, emphasizing their role as critical periods for marketing and promotions. The integration of holiday data provided actionable insights for optimizing seasonal campaigns.
    - **Economic Conditions:**
        - **GDP:** Sales demonstrated a positive correlation with GDP, reflecting higher consumer spending during periods of economic growth.
        - **Inflation:** The correlation with inflation was mixed, as certain sectors experienced reduced sales during price surges, while others, like essentials, remained stable.
        - **Unemployment:** Higher unemployment showed variable effects, with some sectors (e.g., essential goods) seeing increased sales, likely due to shifts in consumer priorities.

- **Lockdown Periods:**
  - Sales declined by an average of 20% during lockdowns, with significant sectoral variations. While the Calcio sector faced notable declines due to reduced sports activities, the Fitness sector experienced increased engagement during lockdowns, driven by a shift to home-based fitness solutions.
  - Transitional periods after lockdowns showed recovery in some sectors but emphasized the uneven nature of sectoral responses.

3. **Interactive Dashboard:**
   - The dashboard empowered users to dynamically explore trends, assess correlations, and compare sales patterns across time periods and sectors.
   - It enabled data-driven decision-making, such as identifying peak sales periods, planning inventory levels during holidays, and adapting to external disruptions like lockdowns.
   - For instance, users can easily visualize the interplay between external factors, such as GDP trends and holiday impacts, to anticipate sales fluctuations and align strategies accordingly.

4. **Insights on Time Series Dynamics:**
   - Time series decomposition provided a detailed breakdown of sales data into trend, seasonal, and residual components.
   - The insights from decomposition, such as the steady upward trend in Pesca and the seasonal spikes in Sci, provide guidance for long-term planning and targeted marketing.

5. **Quantitative Evidence through Statistical Analysis:**
   - Statistical tests confirmed significant differences in sales during varying conditions, such as lockdown versus non-lockdown periods.
   - The t-tests reinforced the observed trends and provided statistical validation for sector-specific impacts, such as the holiday-driven sales spikes in Pesca and the negative effects of inflation on Casual sector sales.

Recommendations and Future Work

To further enhance this work and address its limitations, the following are recommended:

1. **Advanced Predictive Analytics:**
   - Incorporate **machine learning models** (e.g., ARIMA, Prophet, or LSTM) for forecasting sales trends and predicting the impact of external factors.
   - Implement **sector-specific models** to improve accuracy in sectors with distinct patterns.
2. **Multi-Source Data Integration:**
   - Extend the dataset to include additional data sources, such as customer demographics, weather patterns, and competitor pricing, to enrich the analysis further.
   - Integrate real-time data streams to enhance the dashboard's utility for immediate decision-making.
3. **Automated Data Processing:**
   - Automate the data enrichment pipeline using tools like **Apache Kafka** or **Airflow** to ensure continuous updates and scalability.
   - Develop APIs for seamless integration with external data sources like holiday calendars or economic reports.
4. **Enhanced Visualizations:**
   - Incorporate **dynamic visualizations**, such as interactive maps or network graphs, to explore regional sales patterns and relationships among variables.
   - Include customizable filtering options to allow users to focus on specific sectors, time periods, or external factors.
5. **Behavioral Insights:**
   - Analyze customer transaction data to identify purchasing behaviors and preferences.
   - Conduct sentiment analysis on customer reviews or feedback to correlate product sentiment with sales trends.
6. **Sector-Specific Analysis:**
   - Delve deeper into individual sectors to identify unique factors driving sales.
   - Explore how emerging trends, such as sustainability or digital adoption, affect different sectors.

References

1. **ISTAT - Italian National Institute of Statistics:**
    - **Source for economic indicators such as GDP, inflation, and unemployment.**
    - **Website Link**
2. **World Bank Open Data:**
    - **Global economic indicators, including GDP trends and macroeconomic statistics.**
    - **Website Link**
3. **Calendarific API:**
    - **Source for holiday data, including major events like Black Friday and Christmas.**
    - **Website Link**
4. **World Fishing Championship Official Website:**
    - **Used for event dates to enrich the Pesca sector sales analysis.**
    - **Website Link**
5. **Calcio Data:**
    - **Source for event schedules and data related to Serie A matches**
    - **Champions League and Europa League Data**
6. **COVID-19 Lockdown Data:**
    - **Official government announcements and datasets on lockdown timelines and policies.**