



Real world sales data analysis using python

Submitted to:

Prof. Talha Ali Khan & Prof. Sayyed Auwn Muhammad

Subject:

Data Analytics (B)

Presented by:

Chirag Suryakanth Patole (24220357)

Yash Vaibhav Chavan (38777287)

Sawroop Shivarai Teli (97323729)

Vignesh Kumar Veeramani (84084352)

Real world sales data analysis using python

Abstract:

This report is focused on the analysis and visualization of electronics sales data from Amazon, with the aim of answering pertinent business questions. Employing Python libraries such as Pandas, Numpy, Matplotlib, and Seaborn, the analysis delved into user ratings, categories, and time of sale. The dataset provides a comprehensive overview of electronics store purchases. The report uncovered valuable insights from the past sales data.

Introduction:

Embarking on the analysis of Amazon's electronics sales data, the aim of this study is to employ Python-based Product Sales Analysis and Visualizations to optimize internal operations, refine marketing strategies, and minimize costs. Central to our endeavor is the exploration of evolving consumer consumption patterns to inform strategic decisions and drive economic efficiencies. Navigating the transition toward data-driven decision-making, we consider both past methodologies and future aspirations. The overarching goal is to provide actionable answers to key business questions, quantifying success metrics and offering a strategic roadmap for sustained growth and informed decision-making in a dynamic market landscape.

Methodology:

The Exploratory Data Analysis (EDA) commenced with the importation of necessary libraries and the loading of the electronics sale's dataset using Pandas. Initial exploration involved inspecting the first and last few rows of the dataset and examining its shape. The `info()` function was utilized to obtain an overview of the data types and null values in each column. Identifying discrepancies in data types and null values, the methodology proceeded to address these issues.

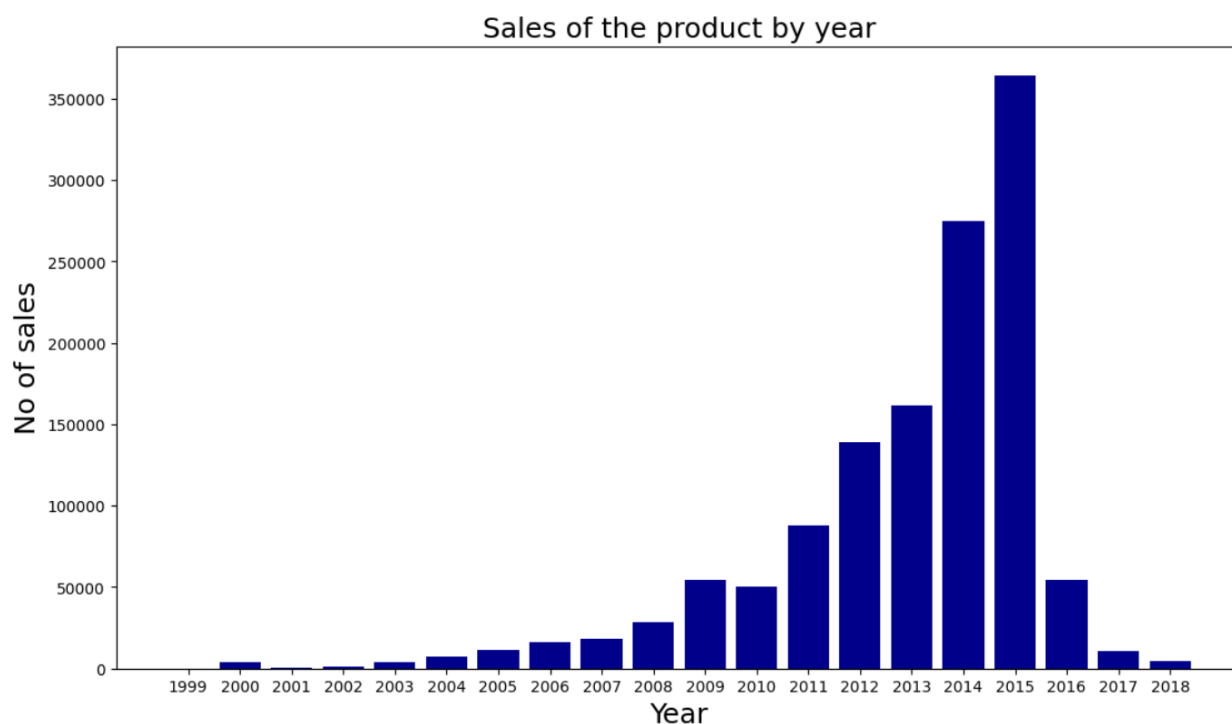
To handle null values, the `astype` function was employed to convert specific columns to their corresponding types. Additionally, the timestamp column was transformed using the `pd.to_datetime` function to ensure the correct format. A statistical summary of the dataset was generated using `describe()` to gain insights into the distribution of numerical features.

Further steps involved handling null and zero values. The 'split' column was dropped as it lacked relevance. The 'brand' column underwent a transformation by replacing top frequencies with NaN values. The 'user_attr' column was fixed by converting it to the 'str' type and assigning gender values based on specific conditions.

Throughout the EDA, various Pandas functions and techniques were applied, including dropping columns, converting data types, replacing values, and exploring unique value counts. These steps collectively aimed to prepare the dataset for in-depth analysis and visualization during the subsequent phases of the report.

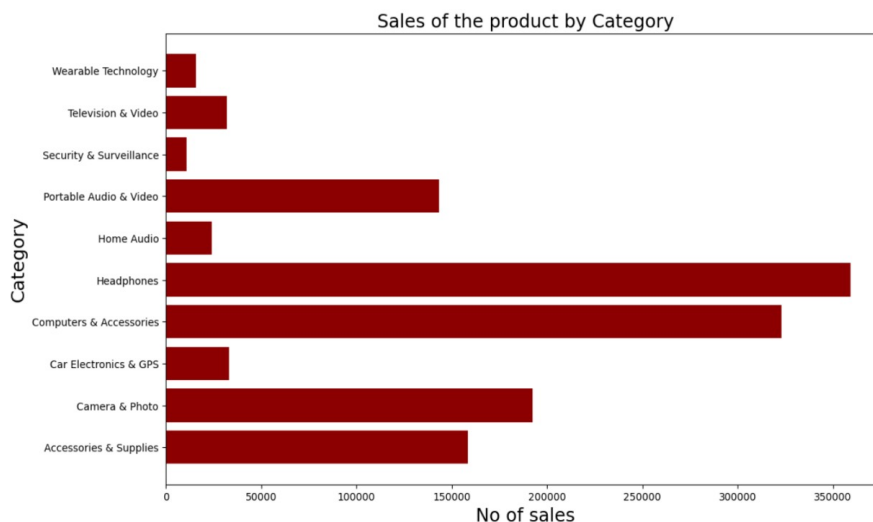
Analysis and Findings:

Q1. Which year has the highest sales?



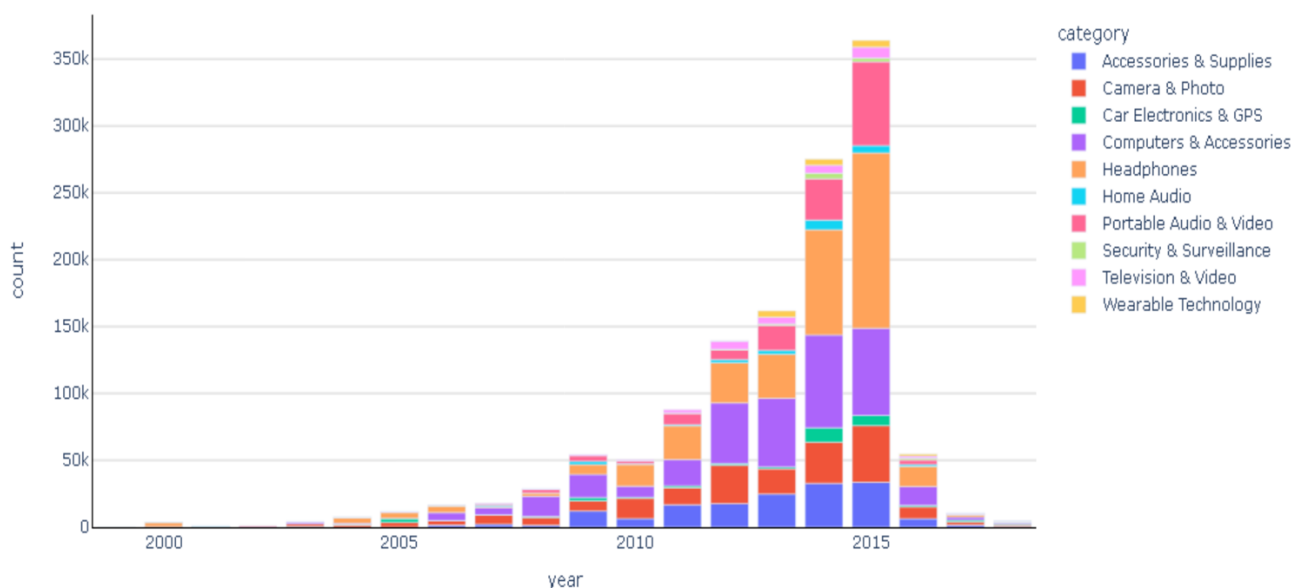
The graph depicting the number of sales over the years clearly indicates that 2015 stood out as the year with the highest sales. Following closely, 2014 secured the position of the second-highest sales, while 2013 claimed the distinction of the third-highest sales during that period. These trends reveal the chronological order of sales performance, with 2015 being the pinnacle in terms of sales figures, followed by subsequent years in descending order.

Q2. Which product has the most sales?



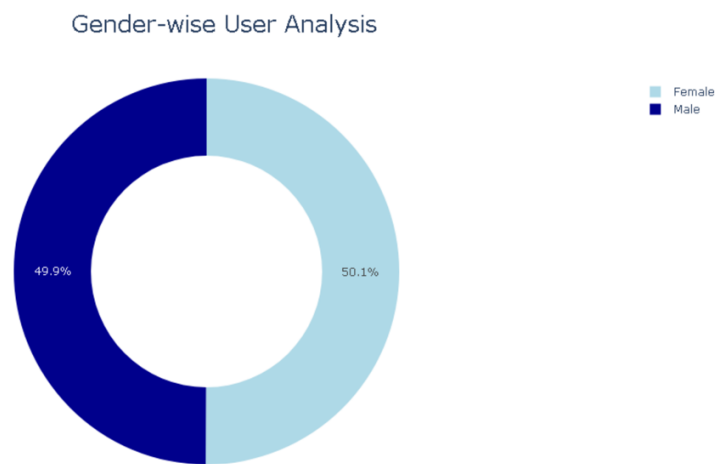
The product that experienced the highest sales was headphones, boasting a remarkable sales figure surpassing 350,000 units. Following closely behind were computer & accessories, recording sales exceeding 300,000 units. Additionally, camera & photo products achieved commendable sales, surpassing the 150,000 mark. This data indicates that headphones emerged as the top-selling category during the specified period.

Q3. Which product has the highest sales in which year, and which product sales are increasing or decreasing?



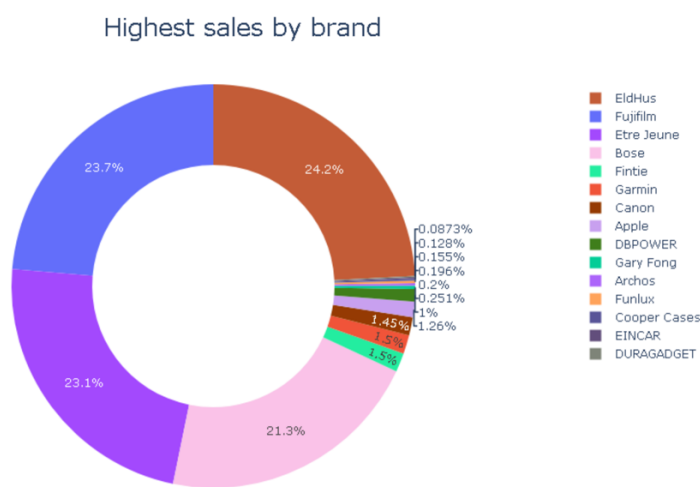
Between 2010 and 2015, there was a notable surge in headphone sales, making them the top-selling product in 2015 among the considered categories. However, post-2015, a decline in sales indicated a shift in consumer preferences or market dynamics. Simultaneously, Portable Audio & Video demonstrated a consistent increase in sales from 2011 to 2015, reflecting a growing demand and positive market response. In contrast, Accessories & Supplies experienced a declining sales trend after 2010, suggesting a shift in consumer interest or market dynamics for this product category.

Q4. Which gender tends to buy more products?



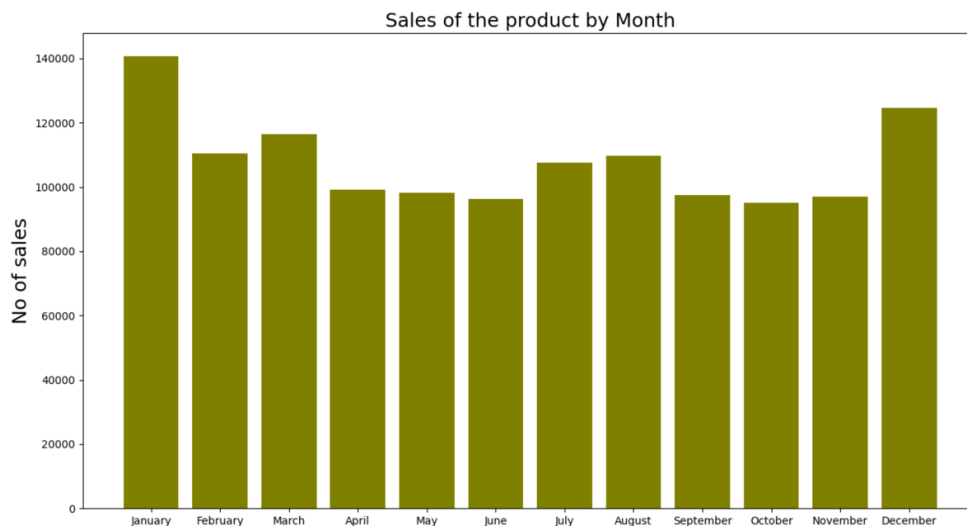
The chart revealed a noteworthy trend in consumer behavior, indicating that females held a slight lead in purchasing products, constituting 50.1% of the market share. Conversely, males accounted for 49.9% of the purchases. This distribution suggested a relatively balanced buying pattern between the two genders during the period under consideration.

Q5. Which brand has the highest sales?



In the analysis of the provided chart depicting sales distribution among various brands, it was discerned that Eldhus emerged as the brand with the highest sales, commanding a significant share of 24.2%. Following closely were Fujifilm, Etre Jeune, and Bose, with respective sales percentages of 23.7%, 23.1%, and 21.1%. This data, gathered from the chart, underscored Eldhus' superior performance, positioning it as the leading brand in terms of sales during the period under consideration.

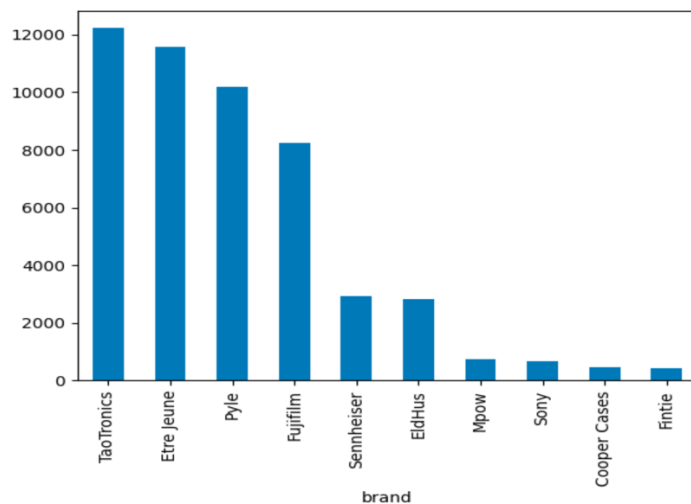
Q6. Which month has the highest sales?



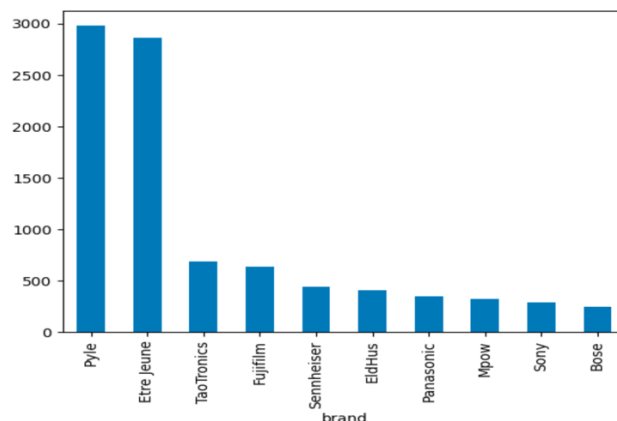
In retrospect, an analysis of the sales trends throughout the year reveals that January emerged as the month with the highest sales. Subsequently, December secured the second position in terms of sales, while March claimed the third spot. Notably, both August and February jointly occupied the fourth position, showcasing comparable peaks in sales. The data from the graph underscores the fluctuating patterns in monthly sales.

Q7. What products sold the most in the last three years (2016, 2017, & 2018)?

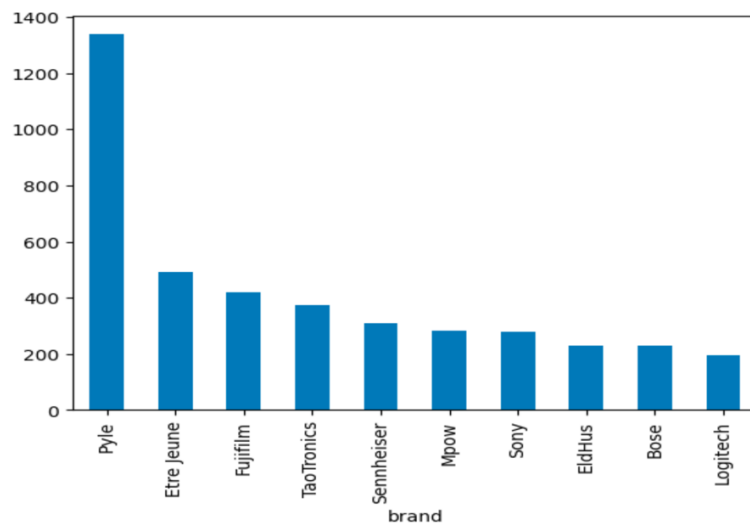
- 2016



- 2017



- 2018



The analysis of the graph, which depicts product sales based on brand ratings over the last three years (2016, 2017, & 2018), reveals notable trends. In 2016, TaoTronics emerged as the top-selling brand, followed by Etire Jeune in the second position and Pyle securing the third spot. Moving to 2017, Pyle claimed the leading position, with Etire Jeune coming in second. The subsequent year, 2018, saw a consistent performance from Pyle, once again securing the top spot. Etire Jeune maintained its presence as the second best-selling brand, while Fujifilm entered the scene, securing the third position in terms of sales. This analysis provides valuable insights into the historical performance and popularity of different brands, emphasizing the dynamic shifts in consumer preferences over the specified period.

Conclusion:

The analysis of electronics sales data reveals key insights for informed decision-making. Identified trends in product sales, ratings, and customer behavior provide a comprehensive understanding of the market landscape. Noteworthy findings include top-selling products and categories, brand performance, and seasonal sales patterns. This data-driven approach allows for targeted marketing, inventory optimization, and strategic planning. Continuous monitoring of these trends is essential for adapting strategies to dynamic market conditions and sustaining business growth.

References:

- <https://www.kaggle.com/datasets>
- <https://www.geeksforgeeks.org/python-programming-language/>
- <https://www.google.com/>