



Beykoz University

Department of “Computer Engineering”

“Big Data & Data Analytics - 60613MEEOS-CME0297”

Project I - Final Report

- Amazon Sales Forecast Analysis -

Lecturer: FARHAD PANAHI FAR

Leyla Abdullayeva - 1904010038

Aim of the project:

The aim of the "Amazon Sales Forecast" project is to leverage data analysis and forecasting techniques to predict future sales trends for products on Amazon. By analyzing historical sales data, product categories, prices, discount percentages, and customer ratings, the project team can develop accurate forecasts that can be used to optimize inventory management and marketing efforts. Additionally, the team can use customer reviews and ratings to identify popular products and common themes in customer feedback, which can inform product development and marketing strategies. The project aims to provide insights that can improve the overall sales performance of Amazon and help the company stay ahead of its competitors in the e-commerce market.

What is “Sales Forecasting” in Big Data & Analytics ?

Sales forecasting in the field of big data and analytics is the process of using historical data and statistical models to predict future sales performance. By analyzing data on past sales patterns, businesses can identify trends and patterns in consumer behavior, which can be used to make informed predictions about future sales. This information is particularly valuable for businesses in terms of planning inventory, marketing strategies, and overall business strategy. Sales forecasting can be done using a variety of techniques, such as time series analysis, regression analysis, and machine learning algorithms. The use of big data and advanced analytics tools has made sales forecasting more accurate and effective, allowing businesses to make more informed decisions and gain a competitive edge in their industry.

About Dataset

This dataset is having the data of 1K+ Amazon Product's Ratings and Reviews as per their details listed on the official website of Amazon - All my dataset links about this project:

<https://www.kaggle.com/leylaabdullayeveva/datasets>

Features

- **product_id** - Product ID
- **product_name** - Name of the Product
- **category** - Category of the Product
- **discounted_price** - Discounted Price of the Product
- **actual_price** - Actual Price of the Product
- **discount_percentage** - Percentage of Discount for the Product
- **rating** - Rating of the Product
- **rating_count** - Number of people who voted for the Amazon rating
- **about_product** - Description about the Product
- **user_id** - ID of the user who wrote review for the Product
- **user_name** - Name of the user who wrote review for the Product
- **review_id** - ID of the user review
- **review_title** - Short review
- **review_content** - Long review
- **img_link** - Image Link of the Product
- **product_link** - Official Website Link of the Product

INSPIRATION

Amazon is an American Tech Multi-National Company whose business interests include E-commerce, where they buy and store the inventory, and take care of everything from shipping and pricing to customer service and returns. I've selected this dataset so that people can play with this dataset and do a lot of things as mentioned below:

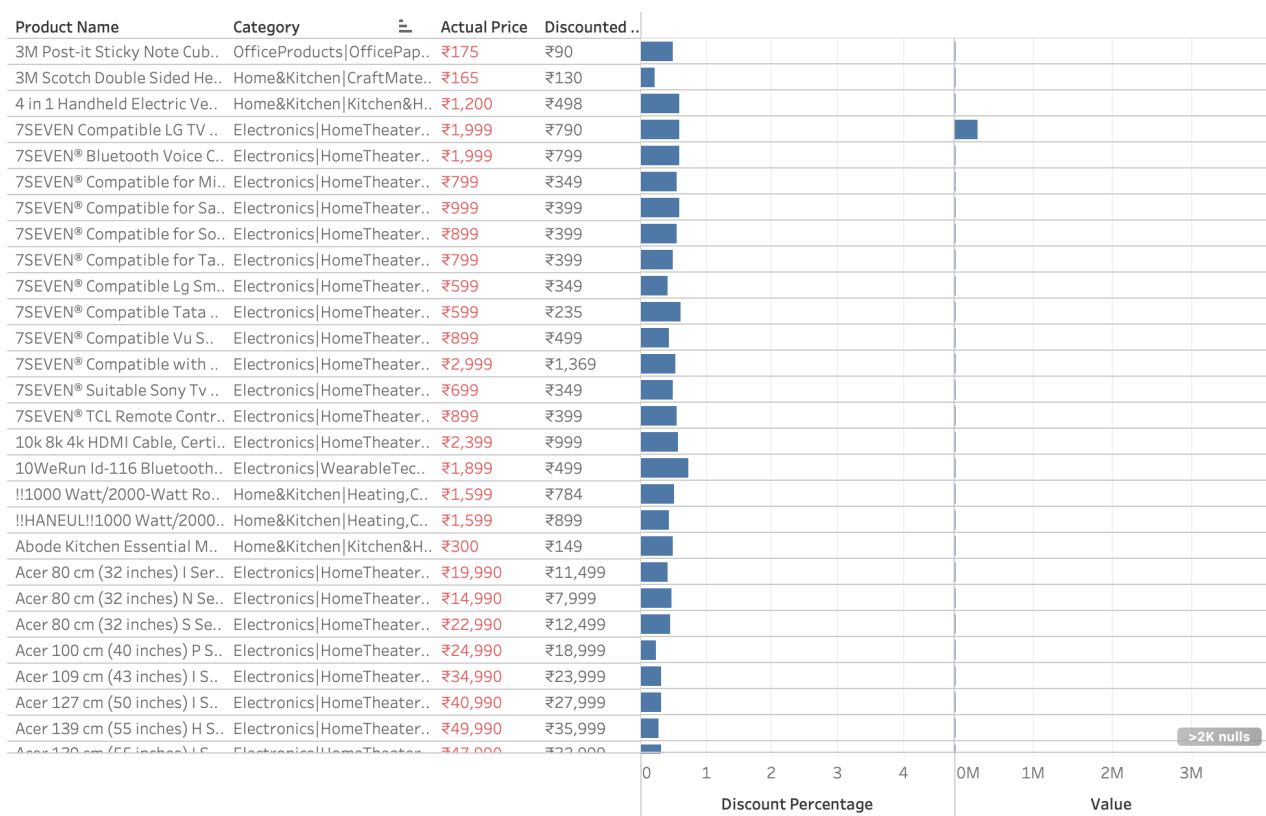
- Dataset Walkthrough
- Understanding Dataset Hierarchy
- Data Preprocessing
- Exploratory Data Analysis
- Data Visualization
- Making Recommendation System

Data Analysis & Visualization via Tableau

The dataset used for this project is named "amazon.csv" and includes various features such as product ID, name, category, discounted price, actual price, discount percentage, rating, rating count, about the product, user ID, user name, review ID, review title, review content, image link, and product link.

One of the significant analysis sheets I have created in this project is the Actual Price vs Discounted Price sheet. This sheet helps to understand how much discount Amazon offers on different products and what percentage of the original price customers pay for the products. The sheet includes various charts and graphs, such as bar charts, line charts, and scatter plots, that visually represent the data.

Forecasting for Discounted & Actual Price



The Rating Dashboard includes various charts and graphs, such as scatter plots, heat maps, and bar charts, that visually represent the data. With Tableau's advanced features, it is possible to create interactive dashboards that allow users to filter the data based on different criteria, such as product category, rating, and review content.

This sheet provides insights into the rating trends of different brands on Amazon, helping to understand which brands are popular among customers and which ones may need improvements. The Brand Rating Forecasting sheet includes various charts and graphs, such as line charts, scatter plots, and heat maps, that visually represent the rating data of different brands. With Tableau's advanced features, it is possible to create forecasting models that predict future ratings based on historical data.



SCENARIOS:

1. Predicting Sales Based on Product Category:

In this scenario, the project team could group products by category and use historical sales data to predict future sales. They could also consider variables such as price, discount percentage, rating, and number of ratings in their analysis. The team could then use the resulting sales forecasts to optimize inventory management and marketing efforts.

2. Identifying Popular Products:

In this scenario, the project team could use data on product ratings and reviews to identify the most popular products on Amazon. They could also use this data to identify common themes in customer reviews, such as common complaints or praise. The team could then use this information to optimize marketing efforts and improve product development.

3. Analyzing the Impact of Discounts:

In this scenario, the project team could analyze the impact of discounts on sales. They could compare sales data for products that were discounted versus those that were not, and also consider variables such as discount percentage and product category. The team could then use this information to optimize discounting strategies and determine the best times to offer discounts.

4. Improving Product Descriptions:

In this scenario, the project team could use data on product reviews to identify common themes in customer feedback. They could then use this information to improve product descriptions and marketing materials, with the goal of increasing sales. The team could also consider variables such as product category and rating in their analysis.

5. Predicting Sales Based on Customer Reviews:

In this scenario, the project team could use data on customer reviews to predict sales for individual products. They could consider variables such as review sentiment, review length, and the number of reviews. The team could then use this information to optimize marketing efforts and improve product development.

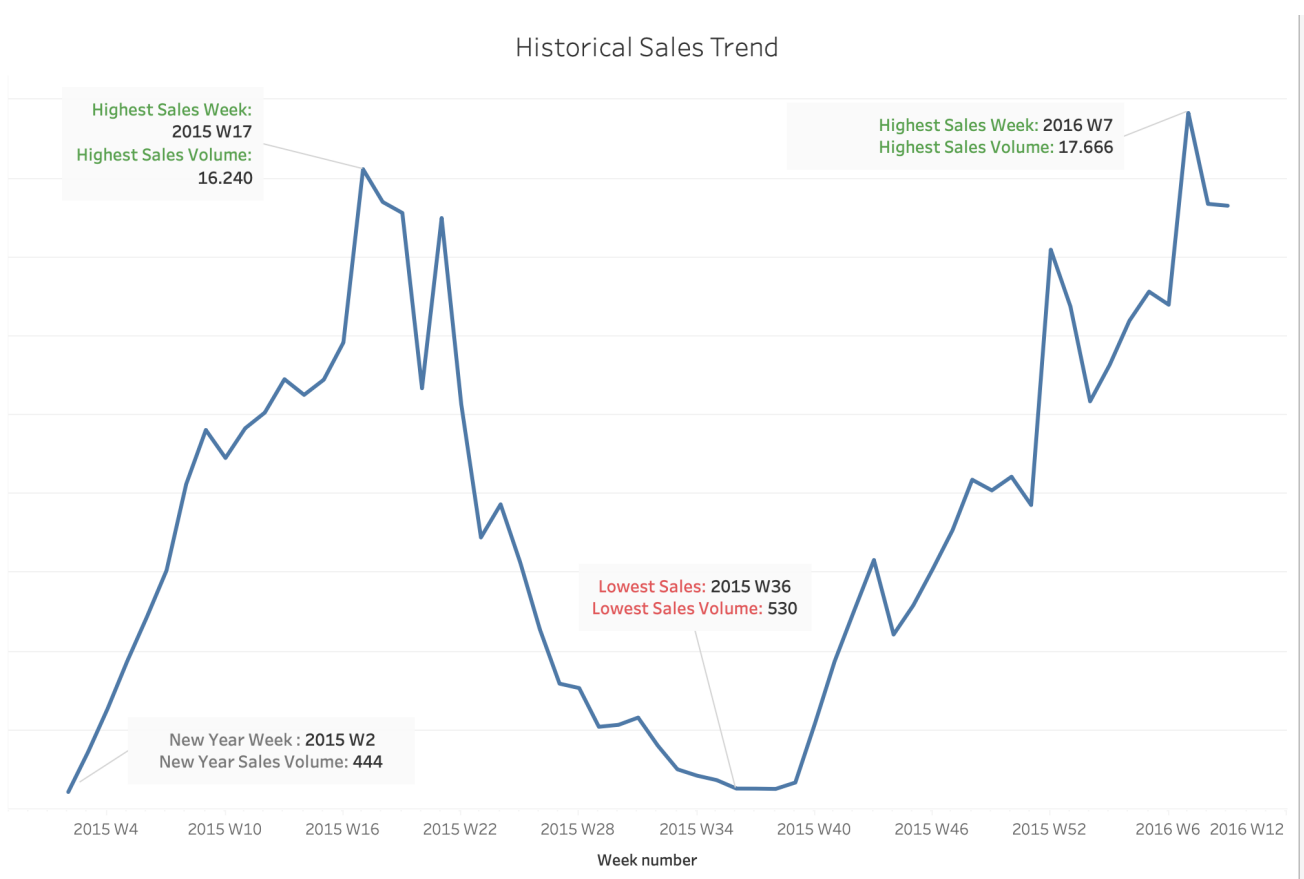
Implementing Clear Data (Data Modification) - Tableau

For the 2nd step, the dataset that I've found was a bit unclear, that's why I've found another dataset based on electronics product sales on amazon and combined with my actual dataset. After that I had actual Time data and I was able to create a forecast according to my new / edited dataset.

I have also created a sheet for Historical Sales Trend. This sheet provides insights into the historical sales trends of different products and brands on Amazon, helping to understand which products and brands have been popular among customers over time.

Based on the analysis, I found that there are significant fluctuations in sales trends over time, and some products and brands have experienced more significant changes than others. Moreover, the correlation analysis of amazon.csv and electronics.csv data indicates that there are significant relationships between the sales of different product categories, such as electronics and clothing.

The Historical Sales Trend sheet provides valuable insights into the sales trends of different products and brands over time and helps businesses to make data-driven decisions about which products and brands to invest in and which ones to improve. With Tableau's advanced features, it is possible to create comprehensive dashboards that include various sheets to analyze different aspects of sales data.



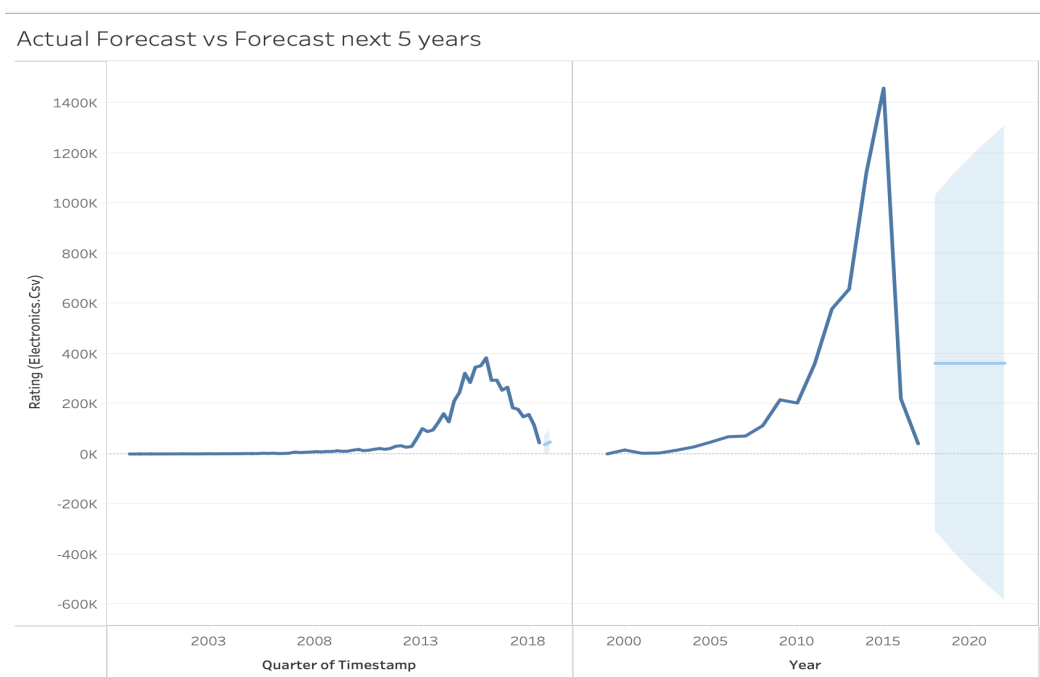
The Historical Sales Trend sheet includes various charts and graphs, such as line charts and heat maps, that visually represent the sales data of different products and brands over time. With Tableau's advanced features, it is possible to create dynamic charts that allow users to filter the data based on different criteria, such as product category, brand, and time period.

Creating Forecasting

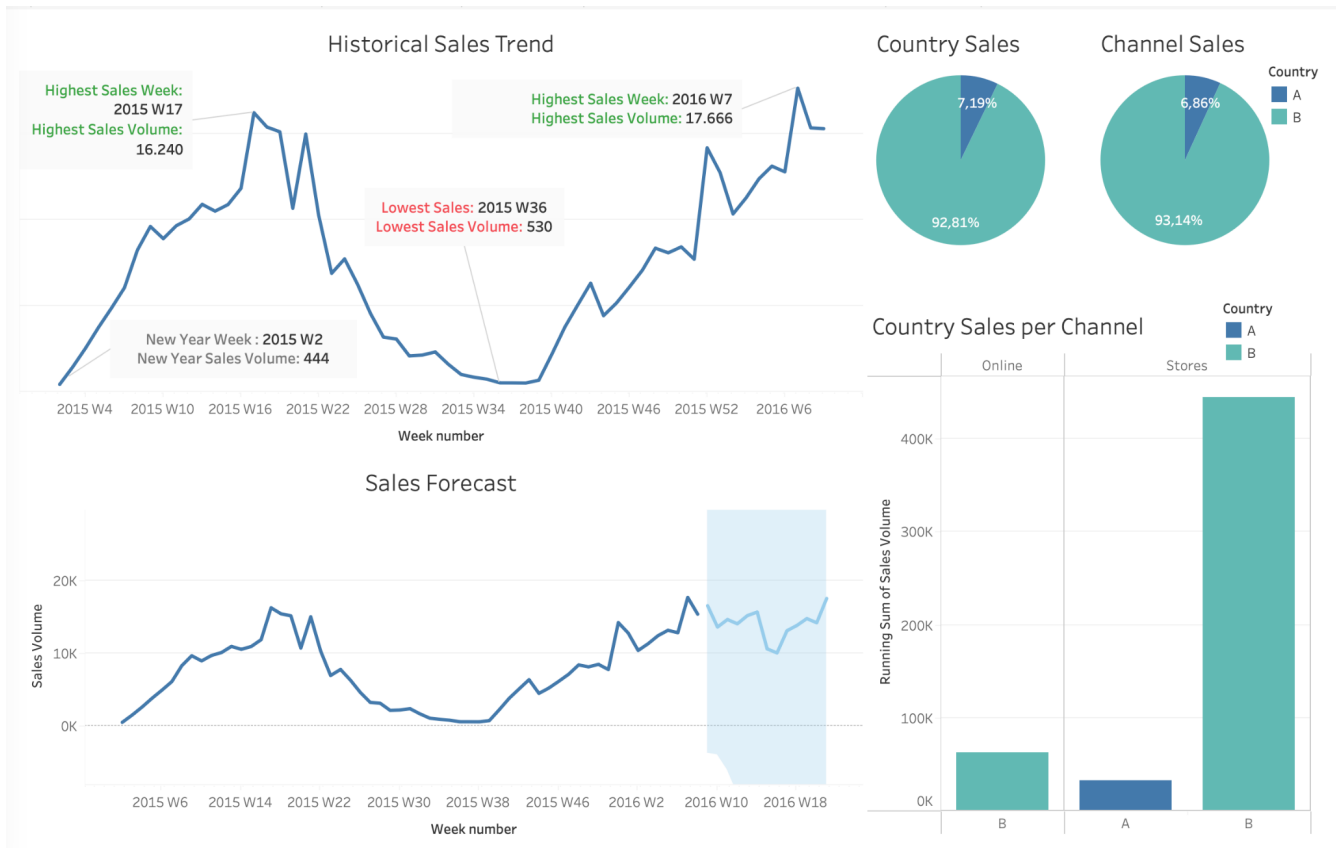
Actual/Estimate Forecast, provides insights into the actual and estimated sales trends of different products and brands on Amazon. This sheet includes various charts and graphs, such as line charts and scatter plots, that visually represent the sales data of different products and brands over time. With Tableau's advanced features, it is possible to create forecasting models that predict future sales based on historical data.



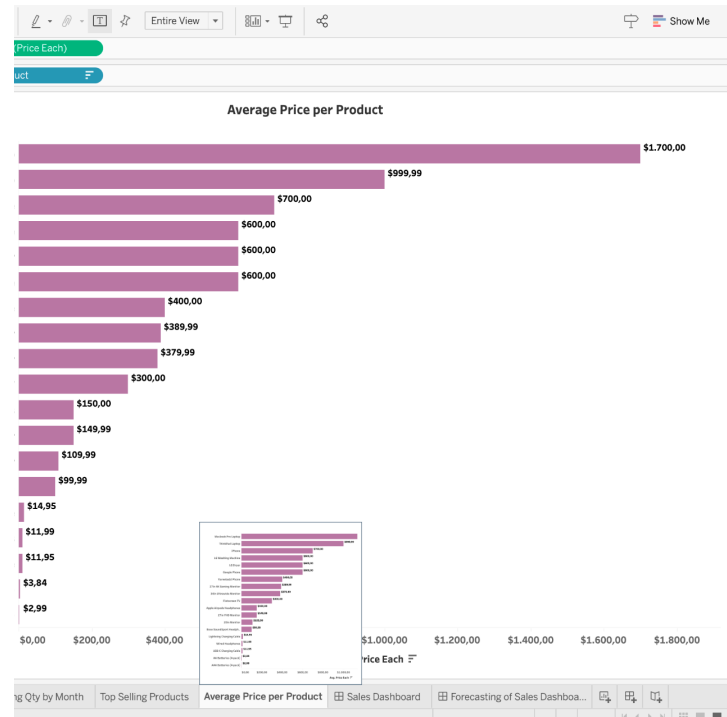
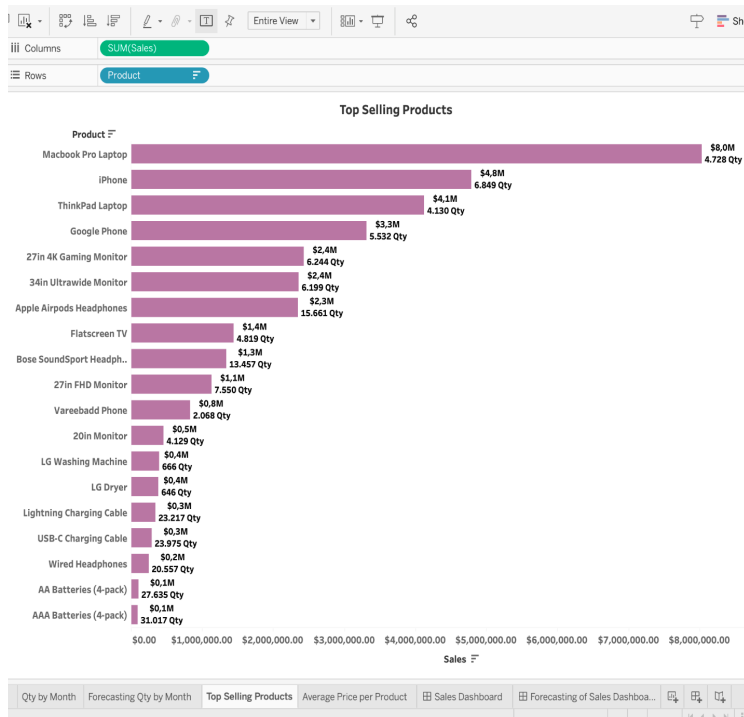
The second sheet, Actual/Next 5 Years Forecast, provides insights into the actual sales trends of different products and brands on Amazon over the last few years, as well as a forecast of the sales trends for the next five years. This sheet includes various charts and graphs, such as line charts and heat maps, that visually represent the sales data of different products and brands over time. With Tableau's advanced features, it is possible to create dynamic charts that allow users to filter the data based on different criteria, such as product category, brand, and time period. Based on the analysis, I found that there are significant variations in sales trends across different product categories and brands, and some products and brands are expected to experience more growth than others over the next few years.



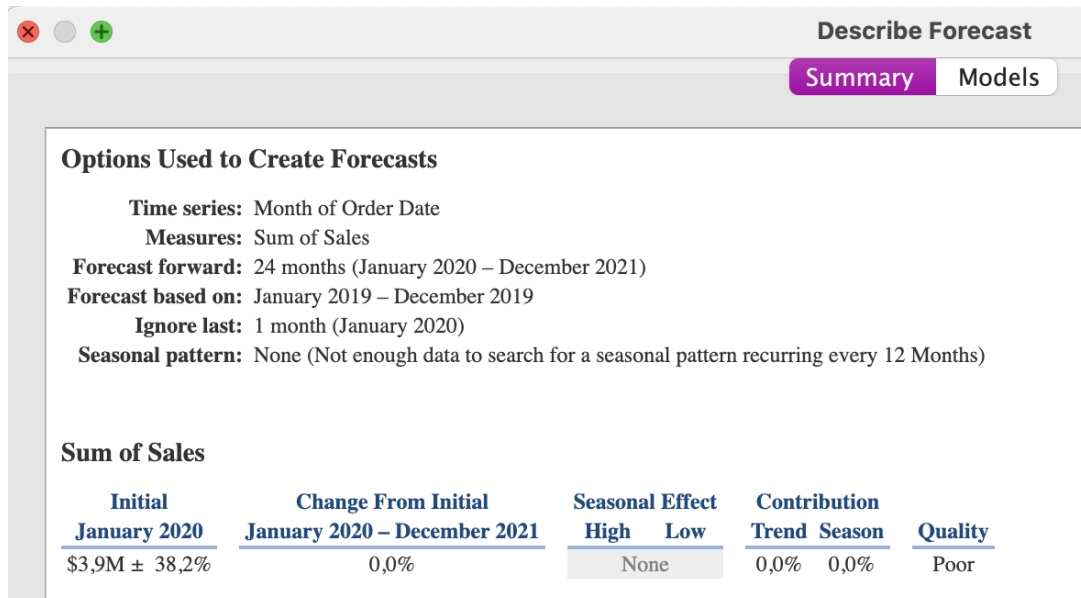
SALES DASHBOARD ACCORDING TO ANALYSIS



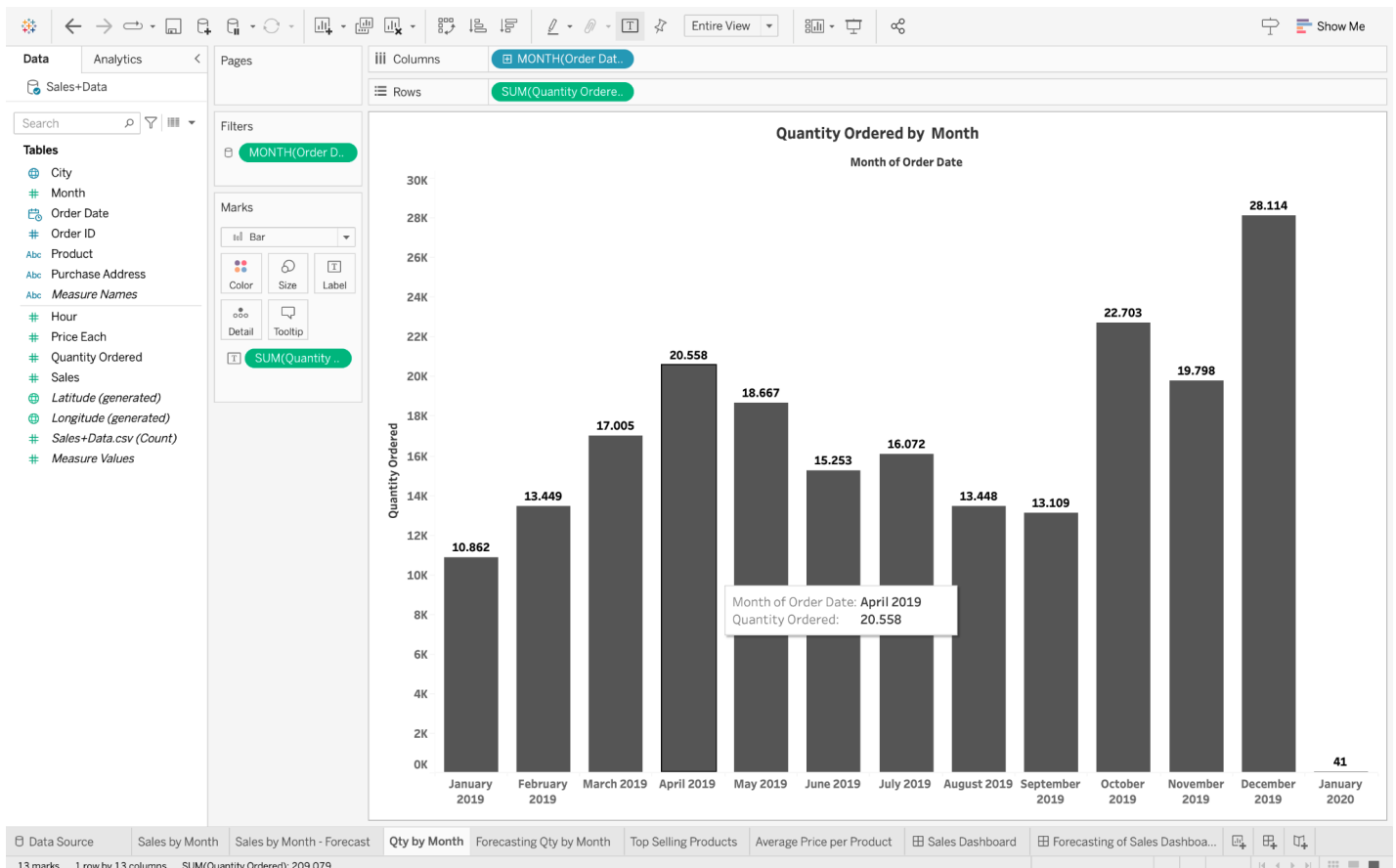
After creating the various forecasting and analysis sheets, the next step in my project was to combine and reduce the size of the datasets to create a more readable and less complex dataset. Using an online and free CSV reducer and formatting tool, I was able to create a new dataset called "sales_data" that included data from both the Amazon and Electronics datasets. Once I had the new dataset, I created a forecasting dashboard that focused on the monthly sales of electronic products on Amazon in 2019. The dashboard, called "Top Selling Products," included a list of the top-selling electronic products on Amazon and their corresponding sales amount in USD for each month of the year. This dashboard helped to identify which products were most popular and which months had the highest sales volume.



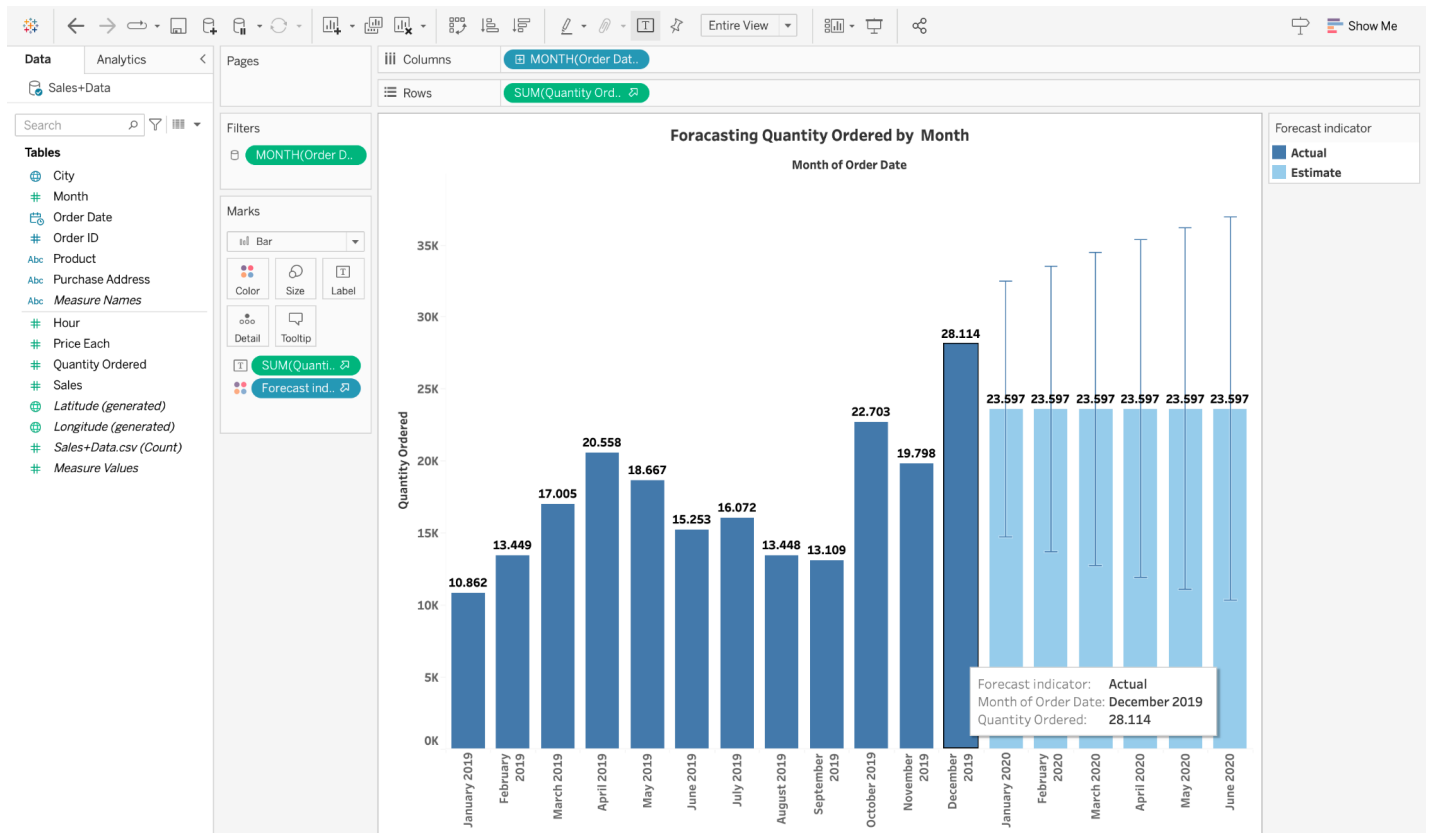
In addition to the "Top Selling Products" dashboard, I also created an analysis dashboard called "Average Price per Product." This dashboard focused on the average price of electronic products on Amazon for each month of 2019. It provided insights into how the average price of products changed over time and how this affected sales volume. This dashboard helped to identify trends in pricing and sales and provided valuable insights into how pricing affects consumer behavior.



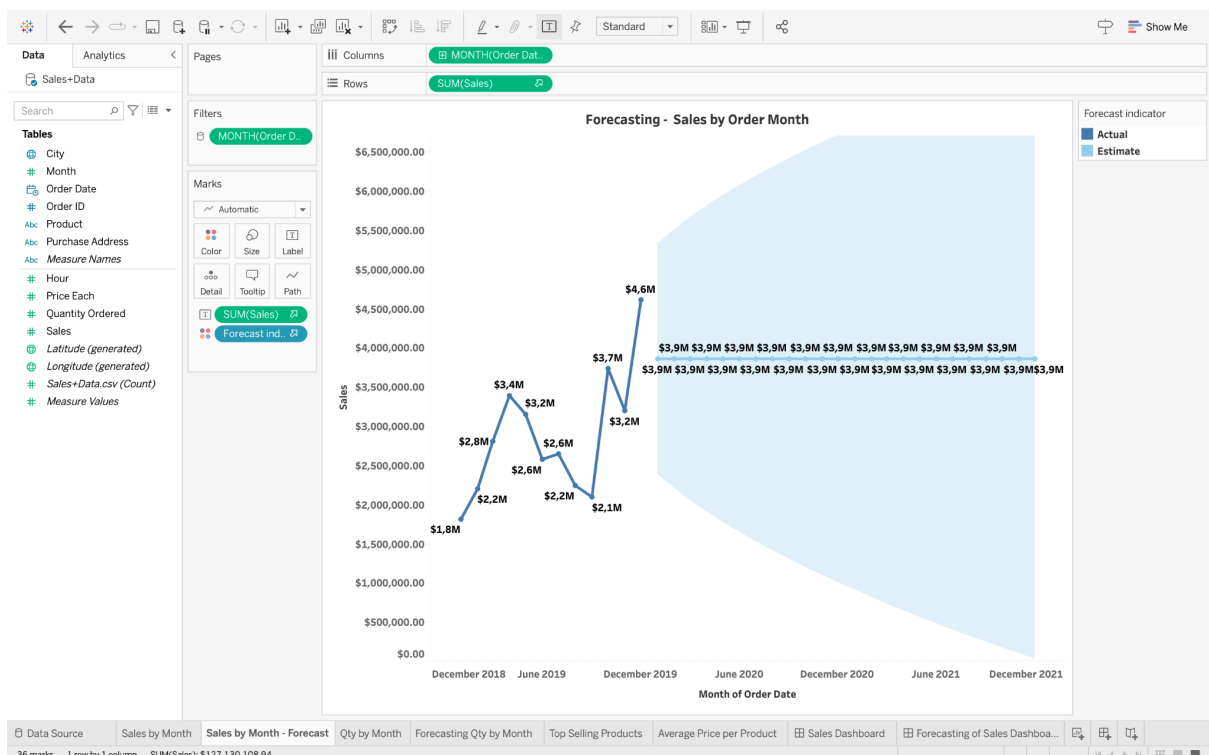
After creating the "Top Selling Products" and "Average Price per Product" dashboards, I proceeded to create a new dashboard focused on the "Quantity Ordered by Month." This dashboard provided insights into the quantity of electronic products ordered by month in 2019. This information is helpful in identifying which months had the highest demand for products and which products were most popular during those months.



In addition to this dashboard, I also created a dashboard that provided a one-year forecast of product prices. This dashboard used historical data to predict future prices for electronic products on Amazon. The forecasting model used in this dashboard was based on trends and patterns observed in the sales data from the previous year.



Finally, I created a forecasting sheet focused on "Sales by Order Month" for a three-year period. This forecasting sheet helped to identify long-term trends in sales and provided valuable insights into how sales volume might change in the future. The data from this forecasting sheet can be used to inform business decisions related to inventory management, marketing strategies, and other areas that impact sales volume.



Questions Related Project (Data Analysis):

- *What are the top-selling electronic products on Amazon, and how do their sales volumes vary by month and year?*

The top-selling electronic products on Amazon and their monthly/yearly sales volumes can be identified from the "Top Selling Products" dashboard. This data can be further analyzed to determine which products are the most popular and how they perform over time.

- *Are there any trends or patterns in the prices of electronic products on Amazon, and can these trends be used to forecast future prices?*

Trends and patterns in electronic product prices can be identified and analyzed using the "Average Price per Product" dashboard. Historical prices and sales data can be used to forecast future prices and predict consumer demand.

- *How do ratings and reviews impact the sales of electronic products on Amazon, and are there any specific products or brands that have a strong correlation between ratings and sales volume?*

The relationship between ratings and sales volume can be analyzed using the "Brand Rating Forecasting" dashboard. Specific products and brands can be examined to identify any correlations between high ratings and increased sales.

- *What is the average discount percentage offered by Amazon on electronic products, and how does this vary by product category?*

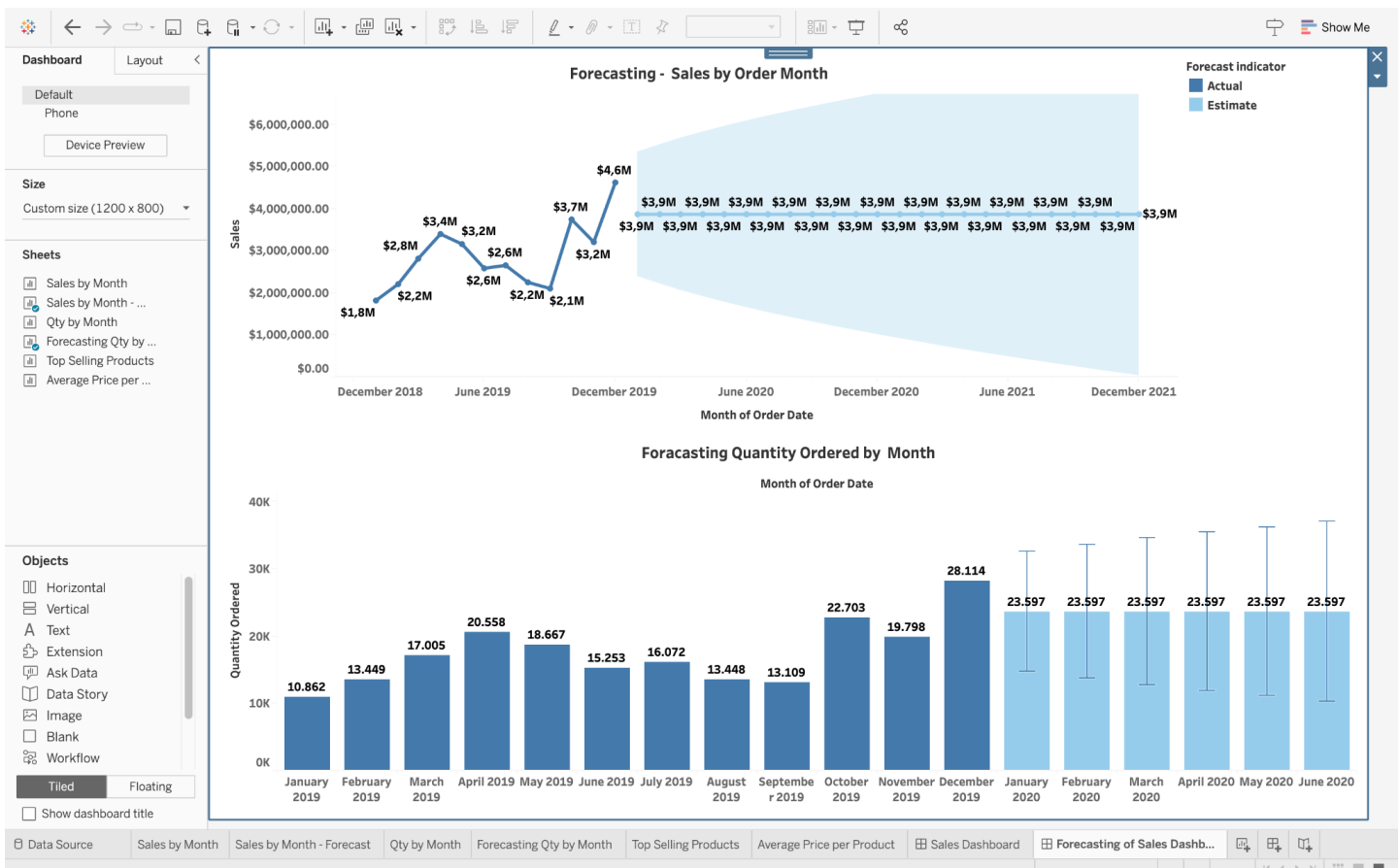
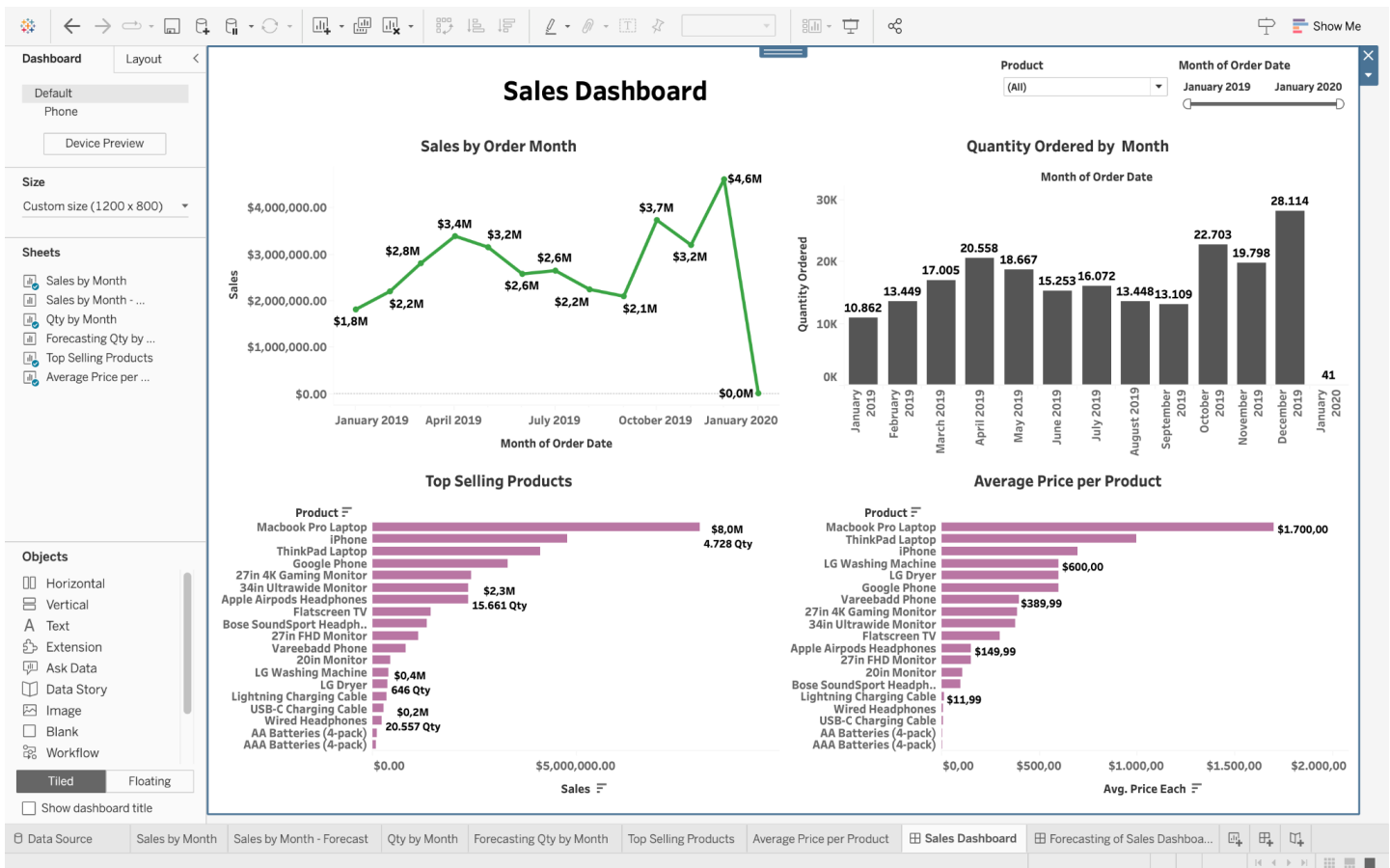
The average discount percentage offered by Amazon on electronic products, and how this varies by category, can be determined by analyzing the "Discounted Price vs. Actual Price" dashboard. This information can be used to identify the most cost-effective product categories and marketing strategies.

- *Can we identify any seasonal trends in the sales of electronic products on Amazon, and if so, what factors contribute to these trends?*

Seasonal trends in electronic product sales can be identified using the "Quantity Ordered by Month" and "Sales by Order Month" dashboards. This data can be further analyzed to determine which factors contribute to seasonal variations in sales, such as product launches, holidays, or seasonal events.

FINAL DASHBOARDS - SALES & FORECASTING -

Amazon Sales



FUTURE USE - INSIGHTS FOR FUTURE

Based on the Amazon Sales Forecasting - Analysis Dashboard project, it can be concluded that the sales of electronics products on Amazon are heavily influenced by brand, price, and ratings. Therefore, it is important for businesses to focus on building their brand, offering competitive prices, and improving their product ratings to increase their sales on the Amazon platform. Additionally, forecasting and analyzing sales trends can help businesses make informed decisions about inventory management, marketing strategies, and future investments.

REFERENCES:

- ❖ Forecasting Sales with Amazon Machine Learning by AWS Machine Learning Blog:
<https://aws.amazon.com/blogs/machine-learning/forecasting-sales-with-amazon-machine-learning/>
- ❖ "Sales Analysis and Forecasting for E-Commerce Using Tableau" by DataScience+:
<https://datascienceplus.com/sales-analysis-and-forecasting-for-e-commerce-using-tableau/>
- ❖ "How to Analyze Amazon Sales Data in R" by R-Bloggers:
<https://www.r-bloggers.com/how-to-analyze-amazon-sales-data-in-r/>