**OPTIMIZING INVENTORY MANAGEMENT FOR AN ONLINE RETAIL STORE**

UMBC Data 604 – Data Management

Under the guidance of **Dr Sarika Rajeev**

Team:

Hansika Karnati

Koushik Choppakatla

Vishal Balaji Kotrike

Yeshwanth Rayankula

Yeswanth Reddy Chereddy

**ABSTRACT**

The goal of this final project is to optimize inventory management for an online retail company utilizing SQL and linear regression modeling. The project makes use of SQL Server Management Studio to generate multiple tables for data analysis, as well as Jupyter Notebook for data visualizations and training and assessing a linear regression model. This project's target users are online retail shop managers, inventory managers, and decision-makers in charge of inventory optimization. Furthermore, the research uses linear regression as a machine learning technique to develop a predictive model. The model is trained and assessed within the Jupyter Notebook environment, taking advantage of its data analysis and machine learning capabilities. The project intends to assist in making educated decisions about inventory levels, procurement methods, and customer demand forecasting by analyzing the data and leveraging the trained linear regression model.

**TABLE OF CONTENTS**

1. INTRODUCTION

2. OBJECTIVE

3. TECHNOLOGIES USED

4. DATASET

a. FACTORS TO ANALYZE

5. TARGET USERS

6. RESULTS

a. DATA VISUALIZATIONS

b. OUTCOMES

7. BENEFITS OF OPTIMIZING INVENTORY MANAGEMENT

8. CONCLUSION

9. REFERENCES

10. APPENDIX

1. **INTRODUCTION**

Online retail shop performance depends on efficient inventory management because it has a direct impact on costs, customer satisfaction, and operational effectiveness. Ineffective inventory management can result in higher costs, displeased clients, and operational difficulties. It is essential for online merchants to put methods into place that streamline their inventory management procedures in order to get over these challenges and increase revenue.

Compared to typical brick-and-mortar businesses, inventory management in online retail outlets has difficulties. Process complexity is increased by fluctuating client demand, a large variety of product offers, and dynamic market conditions. Online retailers need to handle a wide range of products, different order quantities, and timely fulfillment across many regions. Modern technologies and analytical methods are essential for efficiently managing and optimizing inventory in order to solve these difficulties.

The goal of this project is to address the unique inventory management issues that online retail shops encounter and to offer solutions for improvement. We will do a thorough analysis of key variables that have a substantial impact on inventory management, including total quantity sold, slow-moving products, revenue generation, and high-profit margin items. To undertake data analysis and modification, SQL Server Management Studio will be used. Various tables will be generated in order to obtain thorough insights into inventory patterns, client demand trends, and product performance. Additionally, Jupyter Notebook will be used to visualize data, train machine learning models, and assess their performance, with a focus on using linear regression for inventory forecasting.

This effort intends to provide online retail shop managers, inventory managers, and decision-makers with important information and advice for enhancing inventory management practices. SQL Server Management Studio and Jupyter Notebook are used to accomplish this. Greater inventory control, cost savings, and customer satisfaction are among the anticipated outcomes. This project's ultimate purpose is to provide online merchants with the tools they need to succeed in a competitive market by efficiently managing and optimizing their inventories

**2. OBJECTIVE**

Online retail shop performance depends on efficient inventory management because it has a direct impact on costs, customer satisfaction, and operational effectiveness. Ineffective inventory management can result in higher costs, displeased clients, and operational difficulties. It is essential for online merchants to put methods into place that streamline their inventory management procedures in order to get over these challenges and increase revenue.

Compared to typical brick-and-mortar businesses, inventory management in online retail outlets has particular difficulties. Process complexity is increased by fluctuating client demand, a large variety of product offers, and dynamic market conditions. Online retailers need to handle a wide range of products, different order quantities, and timely fulfillment across many regions. Modern technologies and analytical methods are essential for efficiently managing and optimizing inventory in order to solve these difficulties.

The goal of this project is to address the unique inventory management issues that online retail shops encounter and to offer solutions for improvement. We will do a thorough analysis of key variables that have a substantial impact on inventory management, including total quantity sold, slow-moving products, revenue generation, and high-profit margin items. To undertake data analysis and modification, SQL Server Management Studio will be used. Various tables will be generated in order to obtain thorough insights into inventory patterns, client demand trends, and product performance. Additionally, Jupyter Notebook will be used to visualize data, train machine learning models, and assess their performance, with a focus on using linear regression for inventory forecasting.

This effort intends to provide online retail shop managers, inventory managers, and decision-makers with important information and advice for enhancing inventory management practices. SQL Server Management Studio and Jupyter Notebook are used to accomplish this. Greater inventory control, cost savings, and more customer satisfaction are among the anticipated outcomes. This project's ultimate purpose is to provide online merchants with the tools they need to succeed in a competitive market by efficiently managing and optimizing their inventories.

The primary objective of this project is to use SQL and linear regression modeling to optimize inventory management for an online retail company. The initiative attempts to accomplish the following goals:

Analyze and comprehend inventory data: The project will obtain insights into the present inventory state and discover trends and patterns that affect inventory management by investigating parameters such as unit pricing, revenue, and quantity.

Identify opportunities for inventory optimization: The project will discover slow-moving products, high-profit margin items, and other crucial criteria that can help improve inventory levels and procurement plans using data analysis and visualization approaches.

Develop a linear regression model: Using machine learning techniques, the project will train a linear regression model to estimate future inventory demands using inventory data.This model will help with accurate forecasting and proactive inventory management decisions.

Analyzing and validating the model: The project will analyze the trained linear regression model's performance in anticipating inventory needs to assure its correctness and reliability. This evaluation will aid in determining the model's performance in optimizing inventory management.

Actionable recommendations: The project will give actionable recommendations for inventory control, procurement strategies, and customer demand forecasting based on the analysis, insights, and model findings. These recommendations will help online retail store managers and decision-makers make intelligent inventory management decisions.

**3. TECHNOLOGIES USED**

SQL Server Management Studio (SSMS): Basically SQL Server Management Studio is a client tool to connect to Databases. You can connect to multiple instances from a single Management Studio. You create, delete, and alter any database with the use of this Management Studio.

Jupyter Notebook: Jupyter Notebook for Python is just an Interactive IDE (Integrated Development Environment) for testing, teaching/tutoring, and sharing code between many people.

Pandas: A powerful data analysis and manipulation library used for efficient data cleaning, transformation, and aggregation.

Matplotlib: A library in Python for creating data visualizations like plots, graphs, and charts, and it is also used for data analysis.

Seaborn: A library in python that is used to provide data visualizations on statistical plots like boxplot, scatter plots etc,.

NumPy: It is used for multidimensional arrays and advanced mathematical concepts and provides tools for working with them.

SQL Server Management Studio and Jupyter Notebook are two tools used to modify and analyze data. Pandas, Matplotlib, Seaborn, and NumPy, on the other hand, are scientific computing and visualization tools for effective data analysis and communication.

**4. DATASET**

For this project "Optimizing Inventory Management for an Online Retail Store,” We collected a dataset from Kaggle-an online platform for data scientists to share their data knowledge and discover other datasets. It provides a web-based data science environment that provides users to explore, analyze, and develop models using various machine learning techniques. Through this interface, various data scientists and machine learning experts can come together and collaborate with one another. People also participate in data science competitions to solve posted challenges.

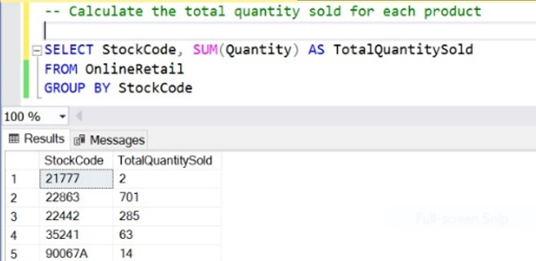
The dataset relates to a popular website, similar to Stack Overflow, that regularly creates a sizable volume of data. This dataset contains a wealth of data and offers important insights into user patterns, trends, and behaviour.

In this research, several tables were considered for analysis. These tables include a variety of user activity-related information, including interactions between questions and answers, user profiles, timestamps, and other pertinent features. We hope to better understand user behaviour, preferences, and trends in relation to the website by examining these tables.

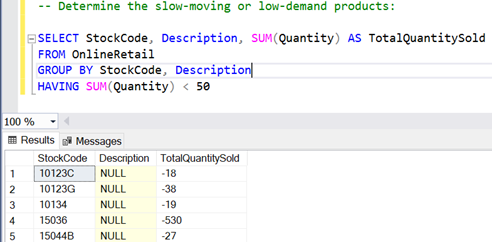
Our goal is to find insightful observations, patterns, and relationships that help guide decision-making and optimize inventory management for online retail stores by utilizing the specified Kaggle dataset and examining the selected tables.

a. FACTORS TO ANALYZE

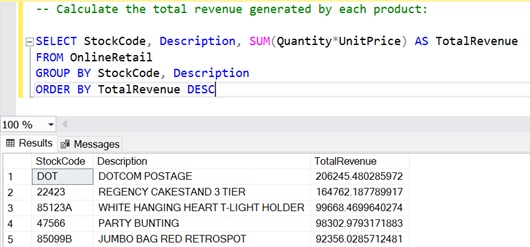
* Total Quantity Sold:
  + Calculate the total quantity sold for each product.
  + Helps in understanding product demand and identifying popular items.



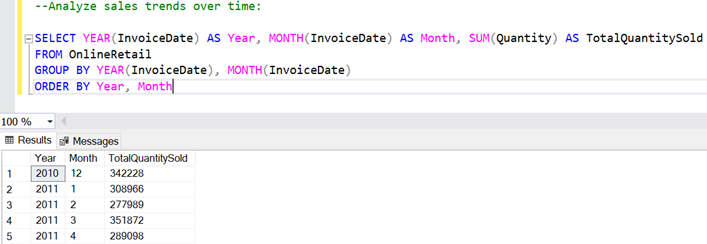
* Slow-Moving and Low-Demand Products:
  + Identify products with low sales velocity.
  + Helps in reallocating resources, avoiding overstocking, and reducing carrying costs.



* Total Revenue Generated:
  + Calculate the total revenue generated by each product.
  + Enables analysis of sales performance and identification of high-revenue products.



* Analyze Sale trends over time:
  + Identify products with high sales over time.
  + Allows for focusing on future demand, plan inventory levels accordingly.



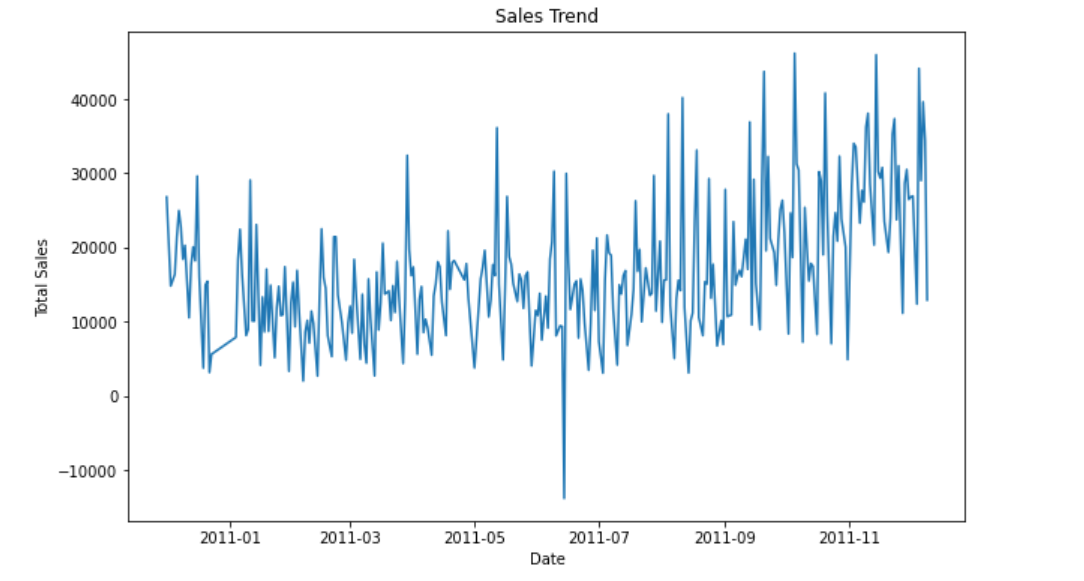
**5. TARGET USERS**

* Sales Representatives must focus on their top-selling products and increase sales efforts for maintaining them.
* E-commerce managers should examine slow-moving products to identify potential marketing issues.
* Product managers must examine the revenue generated by each product and make data-driven decisions and concentrate on high-profit items
* Auditors must identify inconsistencies, investigate potential threats such as theft, and implement corrective actions.
* Supply chain managers must analyze total sold and demand patterns in order to optimize inventory levels and avoid excess inventory or stockouts.
* Sales managers can use the project to evaluate their sales team, identify underperforming products, and take proactive measures to improve overall performance, such as giving out advertisements.
* Logistic managers can use the project to improve the warehouse and distribution processes. Plan efficient distribution routes for high-selling products.

**6. RESULTS**

a. DATA VISUALISATIONS

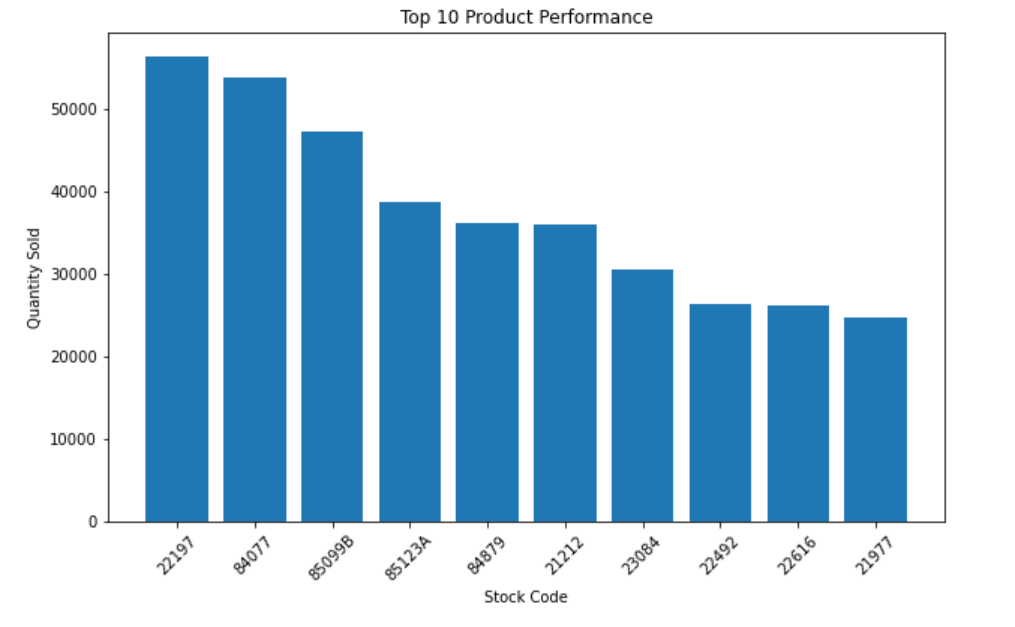
1. **Total Sales vs Date**



The data visualization illustrating the relationship between total sales and date provides valuable insight into sales trends over time. The image shows a distinct pattern of fluctuation in total sales based on date, demonstrating seasonality or periodic variations in client demand.

The visualization shows that certain periods have higher sales volumes, while others have lower sales volumes. The online retailer can use this information to connect inventory management and marketing tactics with observed sales patterns. Recognizing seasonal peaks and troughs allows the retailer to plan inventory levels, promotional activities, and operational resources in advance.

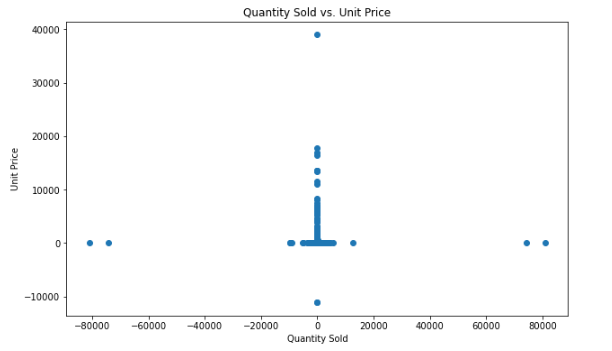
1. **Quantity Sold vs Stock Code**



The data visualization combining the quantity sold with stock codes gives useful insights into the performance of various products, with a particular focus on the top ten products. Notably, the stock code '22197' is the most popular in terms of quantity sold.

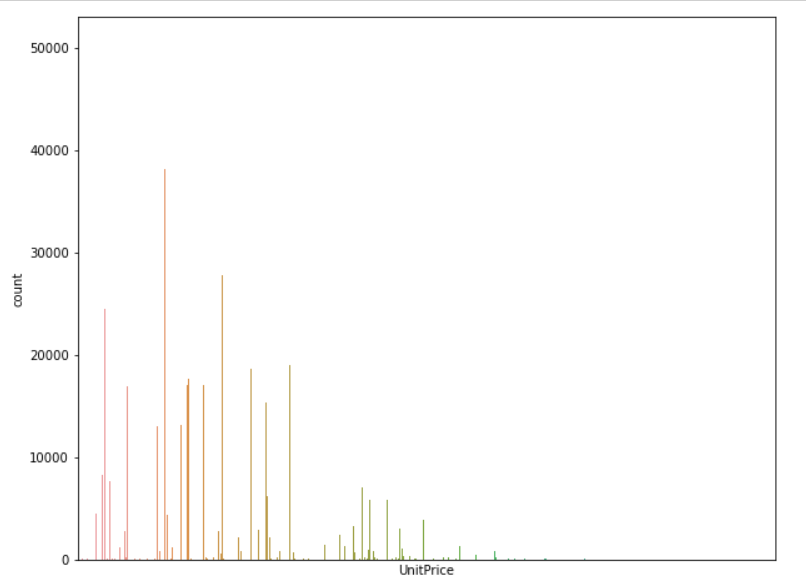
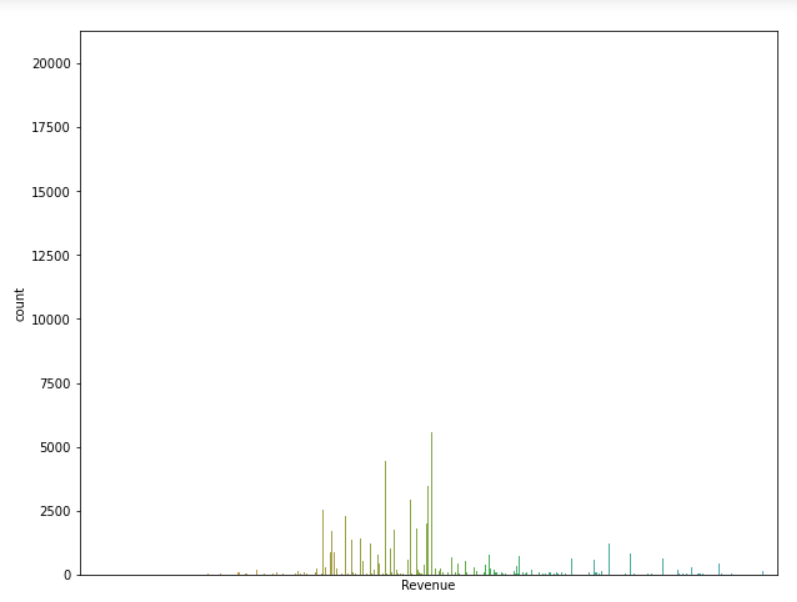
The visualization shows that the product linked with stock code '22197' has outperformed other products in the dataset in terms of sales success. This knowledge can be expanded upon to further understand the elements that contribute to its popularity and strong demand. The online retail store can acquire insights into the elements that make this product a top-performing product by studying its aspects, such as its features, pricing, marketing efforts, or customer feedback.

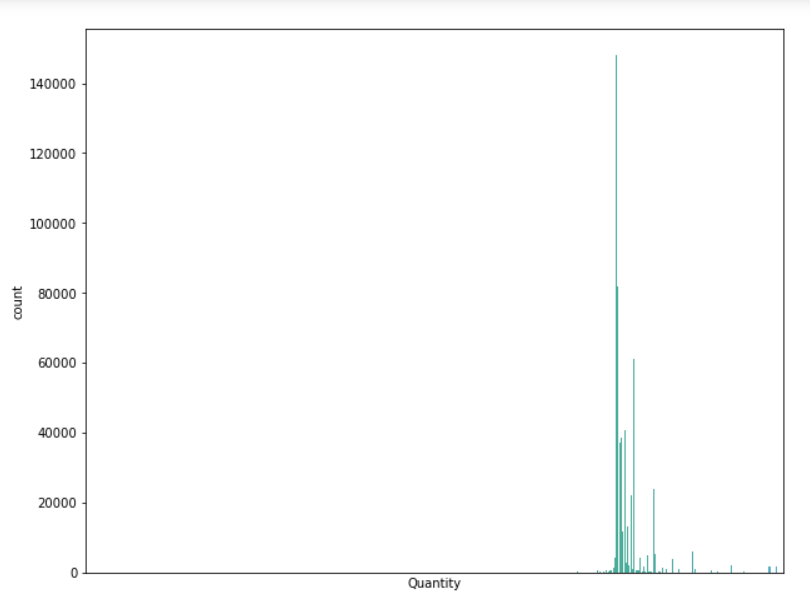
1. **Unit Price vs Quantity Sold**

****

The data visualization demonstrating the link between unit price and quantity sold provides useful insights into the pricing dynamics of various products. Examining this image reveals remarkable patterns and trends, offering light on the effect of unit pricing on the quantity of products sold.

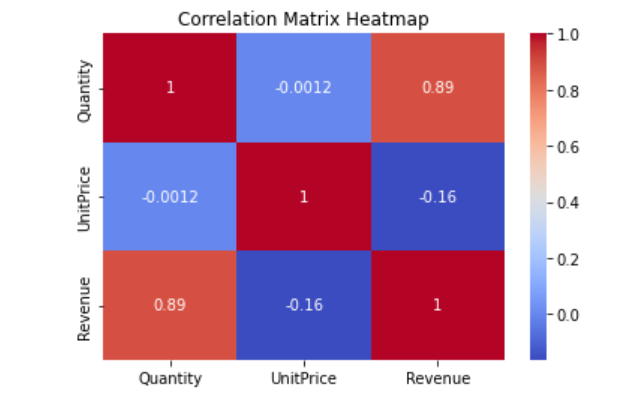
The graphic shows a dispersed distribution of data points, demonstrating variable sales performance across price points. Certain products with lower unit pricing sell in greater quantities, implying a possible link between affordability and client demand. This finding implies that price sensitivity is important in generating sales volume, and competitive pricing strategies may be helpful in recruiting customers and increasing sales.

1. **COUNT PLOTS OF THE NUMERIC VARIABLES FOR DATA ANALYSIS**



UnitPrice, Revenue, and Quantity count graphs provide useful information on the distribution and frequency of these variables. Within the dataset, these visualizations aid in identifying pricing patterns, revenue segmentation, and demand trends. Using these insights, the online retailer may optimize pricing tactics, focus on high-revenue categories, and align inventory management with consumer preferences, ultimately driving sales growth and maximizing profitability.

**V. CORRELATION**

****

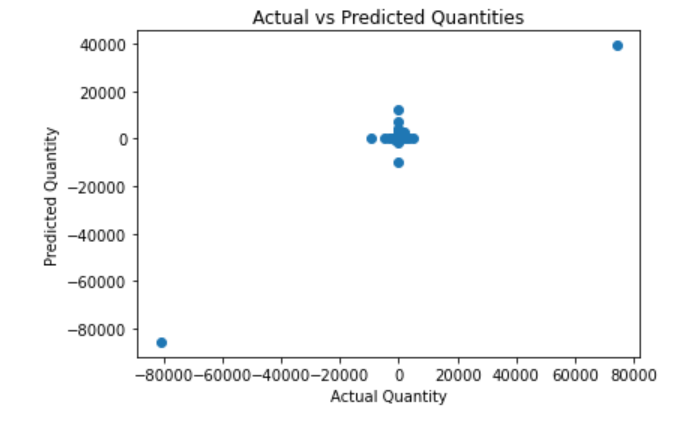
The data visualization of our project's correlation matrix demonstrates insightful links between numerous variables, with the strongest correlation measured at 0.89 between quantity and income. This strong positive correlation indicates that there is a significant relationship between the number of products sold and the money produced.

The correlation coefficient of 0.89 suggests a strong positive linear relationship between quantity and revenue. This observation means that as the number of things sold increases, so does the amount of income generated. It emphasizes the direct impact of sales volume on the online retail store's financial performance.

b. OUTCOMES

* After Training, Testing, and Evaluating our Linear Regression Model

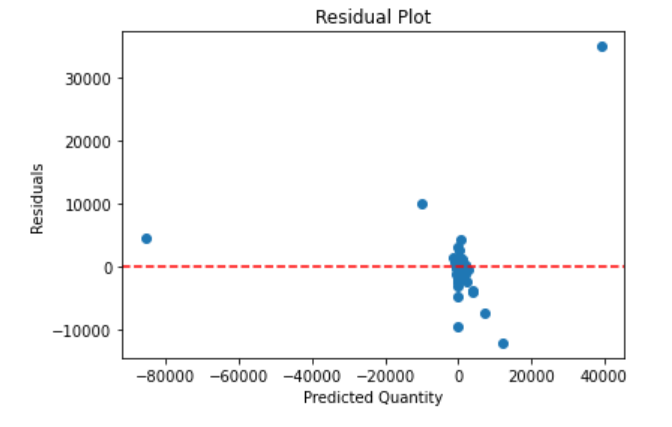
1. **Scatter Plot of Predicted Quantity vs Actual Quantity**



A scatter plot between projected and actual quantity provides vital insights into the accuracy and effectiveness of our quantity prediction model. The scatter plot illustrates the relationship between predicted and actual quantities for various products, with an emphasis on spotting outliers or inconsistencies.

The scatter plot's relatively modest number of outliers reveals that the quantity forecasts closely match the actual numbers for the bulk of the products. This indicates that our quantity prediction model is working effectively and generating accurate estimates in the majority of cases. Having fewer outliers suggests that our predictions are more precise and reliable, which is critical for effective inventory management and meeting consumer demand.

1. **Residual Plot**

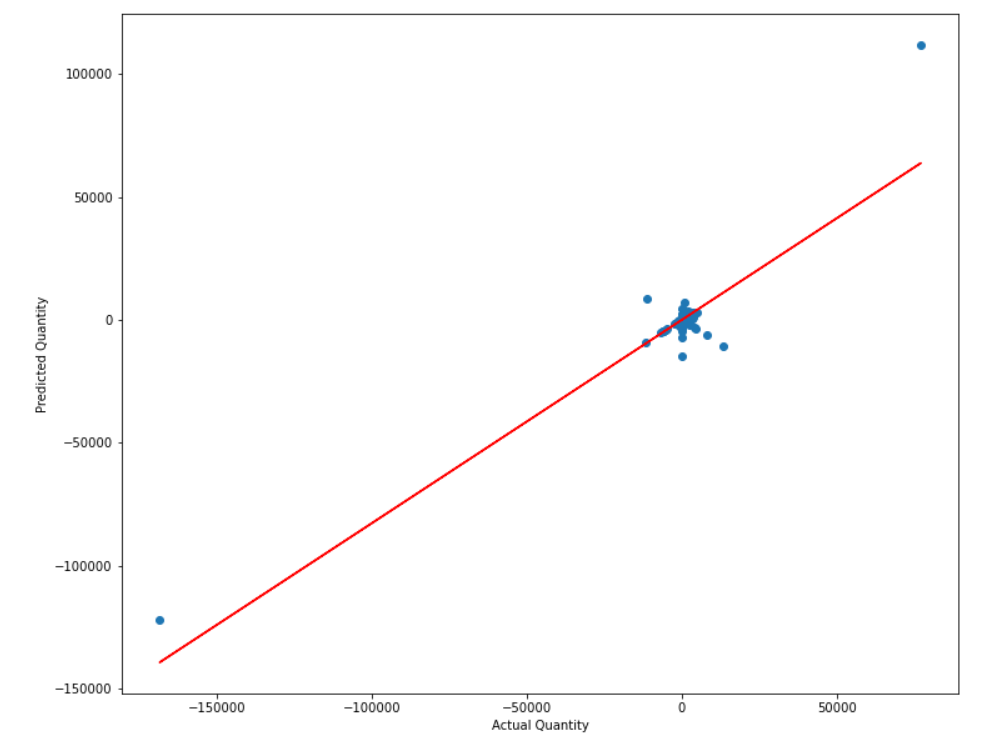
****

A residual plot between residuals and anticipated quantity data visualization provides vital insights into the performance and accuracy of our quantity prediction model. The residual plot allows us to evaluate the accuracy of our predictions by studying the distribution of residuals, which indicates the disparities between the actual and expected numbers.

The presence of a greater number of data points at or near the zero horizontal line is one remarkable finding from the residual plot. This shows that a major fraction of our forecasts match the real amounts very closely, resulting in residuals near to zero. This trend indicates that our quantity prediction model is accurate in guessing the amounts for these products for a large number of data points. A higher concentration of data points near the zero line suggests that our predictions are more accurate and reliable.

1. **Accuracy and Final Plot of The Model**

* Accuracy of the Linear Regression Model after Training, Testing, and Evaluating is 82.57 %
* Plot between Predicted Quantity vs Actual Quantity of the Linear Regression Model

****

The data visualization of a plot between the predicted quantity and actual quantity highlights the performance and accuracy of our quantity prediction model, which is represented by a linear regression graph. The linear graph shows a strong linear relationship between the predicted and actual numbers, emphasizing our model's accuracy in estimating quantity values.

The alignment of the data points along the linear trendline, suggesting a high level of correlation and predictive accuracy, is one remarkable observation from the linear regression graph. The linear relationship implies that our model captures the underlying patterns and trends in the data, allowing for accurate quantity value forecasts. The linear shape of the graph means that as the anticipated quantity increases, so does the real quantity, further verifying our model's prediction powers.

**iv. To optimize inventory management using the quantity column, we can consider the following approaches:**

Demand Forecasting: By analyzing historical sales data and considering the correlation between quantity and revenue, you can forecast the demand for different products. This information can help you plan inventory levels accordingly, ensuring that you have enough stock to meet customer demand without excessive overstocking.

Reorder Point Calculation: Based on the desired service level and lead time, you can calculate the reorder point for each item using the historical quantity data. The reorder point is the inventory level at which you should place a new order to replenish the stock. By considering the correlation between quantity and revenue, you can set the reorder point at a level that ensures sufficient stock availability while minimizing the risk of stockouts.

Inventory Optimization: By analyzing the correlation between quantity and revenue, you can identify the top-selling products that contribute significantly to revenue. Focus on optimizing the inventory management of these high-impact items by implementing strategies such as just-in-time inventory, safety stock optimization, or implementing automated replenishment systems.

Seasonality Analysis: Analyzing the quantity data over different time periods can help you identify seasonal trends in customer demand. By understanding the seasonal fluctuations in quantity, you can adjust your inventory levels accordingly to ensure sufficient stock during peak demand periods and minimize excess inventory during slower periods.

By leveraging the correlation between quantity and revenue, you can make data-driven decisions to optimize inventory management, improve forecasting accuracy, minimize stockouts, reduce carrying costs, and ultimately enhance overall operational efficiency.

**7. BENEFITS OF OPTIMIZING INVENTORY MANAGEMENT**

1. Reduced stock-outs: Our solution gives the online retail business the ability to precisely estimate inventory needs and limit stock-outs, ensuring that products are readily available to satisfy customer demand. This is done by assessing important criteria including total quantity sold and demand trends.
2. Reduced surplus inventory: Our project assists the online retail business in optimizing inventory levels, and decreasing excess inventory and related carrying costs. This is done by gaining insights from studying slow-moving products and high-profit margin items.

Optimized inventory management based on data analysis enables the online retail company to cut costs associated with maintaining goods, freeing up funds for other business operations and enhancing overall cash flow.

1. Enhanced operational performance: The online retail business can streamline operations, boost inventory turnover, and enhance overall operational performance with the use of effective inventory management solutions informed by the findings of our project.
2. Increased customer happiness: Our project assists the online retail business in quickly meeting consumer demand by limiting stockouts and ensuring efficient order fulfillment. This leads to increased customer satisfaction and loyalty.
3. Data-Driven Decision Making: Our solution enables the online retail store to optimize purchasing, product offerings, and pricing strategies for better business outcomes. These decisions will be based on data analysis.
4. Effective order fulfillment: Our project enables effective order fulfillment, assuring timely delivery and satisfying customer expectations, by properly anticipating inventory demands and maintaining optimal inventory levels.
5. Improved forecasting and planning: Our project use data analysis and visualization to provide useful insights into demand trends and product performance. This information helps the online retailer store to improve forecasting and planning procedures, which in turn leads to better inventory management and decision-making.

**8. CONCLUSION**

To summarize, the goal of our project was to improve inventory control for an online retailer by analyzing and visualizing data with the help of programs like SQL Server Management Studio and Jupyter Notebook. We were able to gather important insights that might inform effective inventory management methods by examining important factors like total quantity sold, slow-moving products, income earned, and high-profit margin items.

Based on our findings, the online retail store can gain a lot by using optimized inventory management procedures. First, fewer stock shortages and prompt order fulfillment can lead to higher levels of customer satisfaction. Improved client loyalty and experience result.

Furthermore, good inventory management saves money by reducing surplus inventory and preventing unnecessary carrying costs. The online retail company can optimize procurement, reduce storage costs, and enhance cash flow by precisely estimating inventory demands using techniques such as linear regression.

In addition, the insights generated from our investigation enable data-driven decision-making. The Internet retailer can make educated decisions about product offerings, pricing, and supply chain management. This ensures that resources are appropriately allocated, resulting in enhanced operational performance and profitability.

In a nutshell, our study emphasizes the necessity of optimizing inventory management for online retail companies and shows how data analysis and visualization tools such as SQL Server Management Studio and Jupyter Notebook may help with this process. The online retail store can get benefits such as greater customer happiness, cost reduction, improved cash flow, efficient order fulfillment, and data-driven decision-making by exploiting the insights gathered, ultimately leading to a competitive edge in the online retail industry.

**9. REFERENCES**

*Capella*. (2022, December 30). Retrieved from The Power of Data Management in Retail and e-Commerce:

[https://www.capellasolutions.com/blog/data-management-in-retail-and-e-commerce.](https://www.capellasolutions.com/blog/data-management-in-retail-and-e-commerce.%20)

Dee, C. (2023, March 15). *Algolia*. Retrieved from <https://www.algolia.com/blog/ecommerce/how-to-increase-your-sales-and-roi-with-optimized-ecommerce-merchandising/>

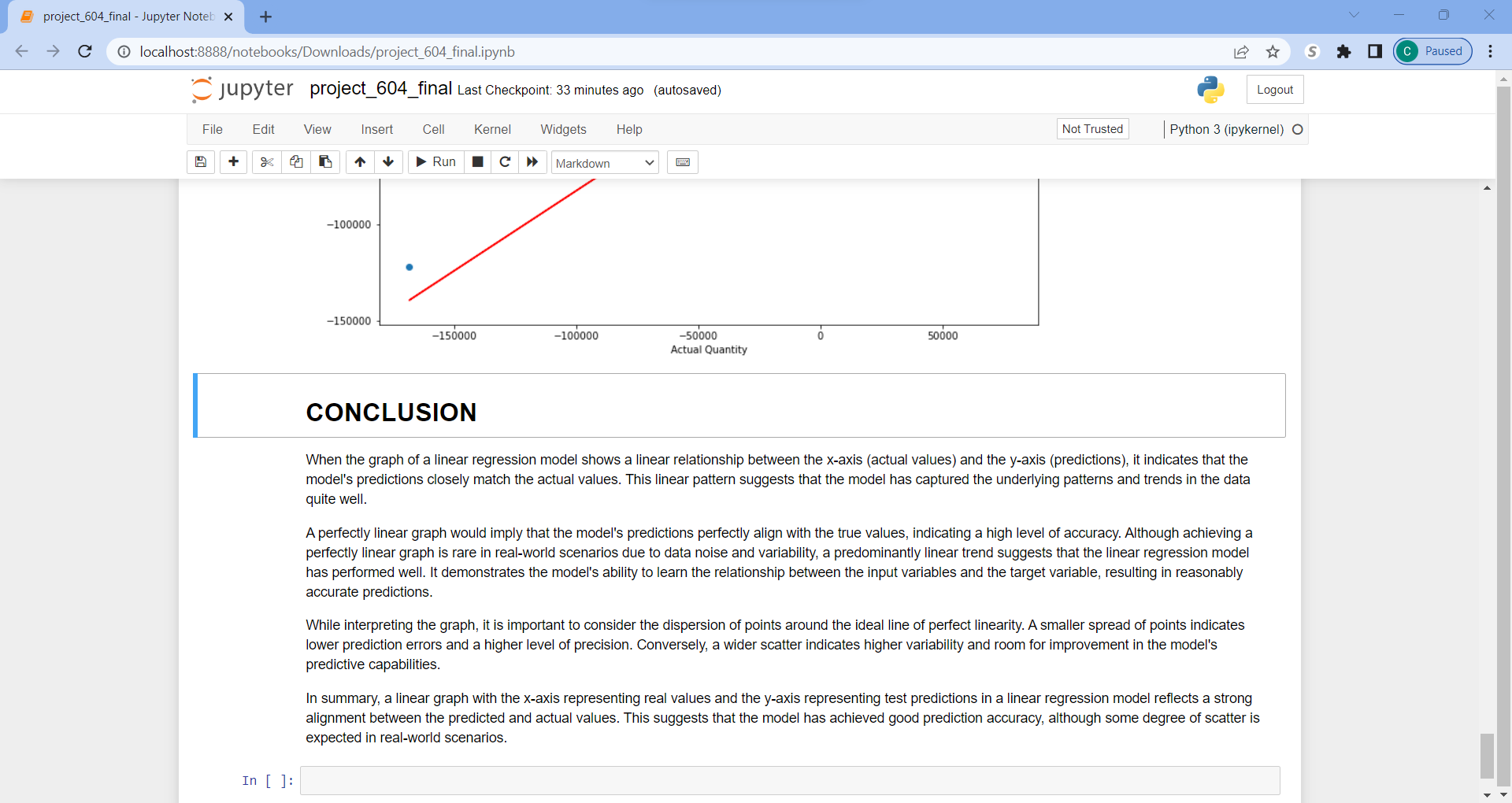
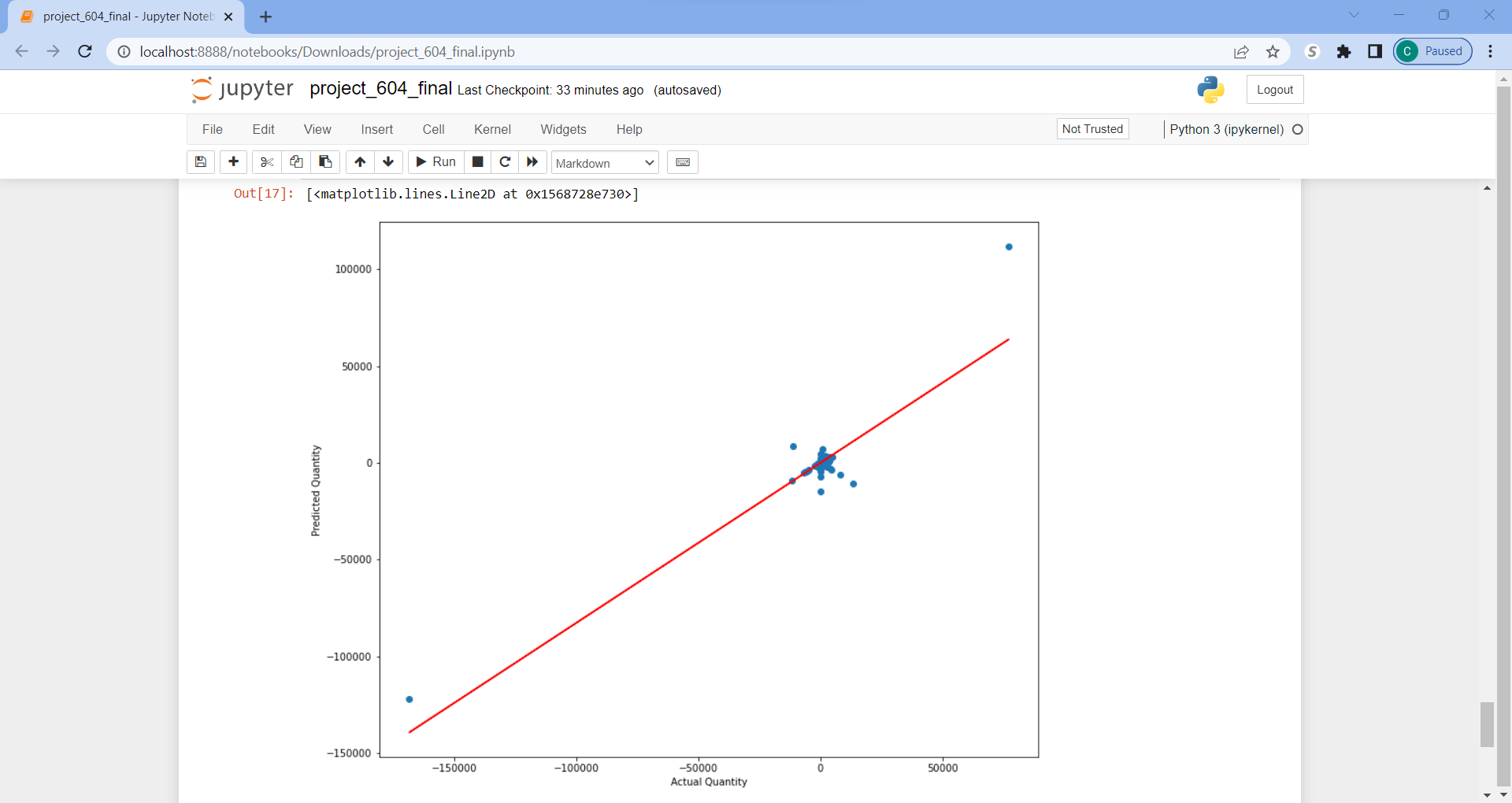
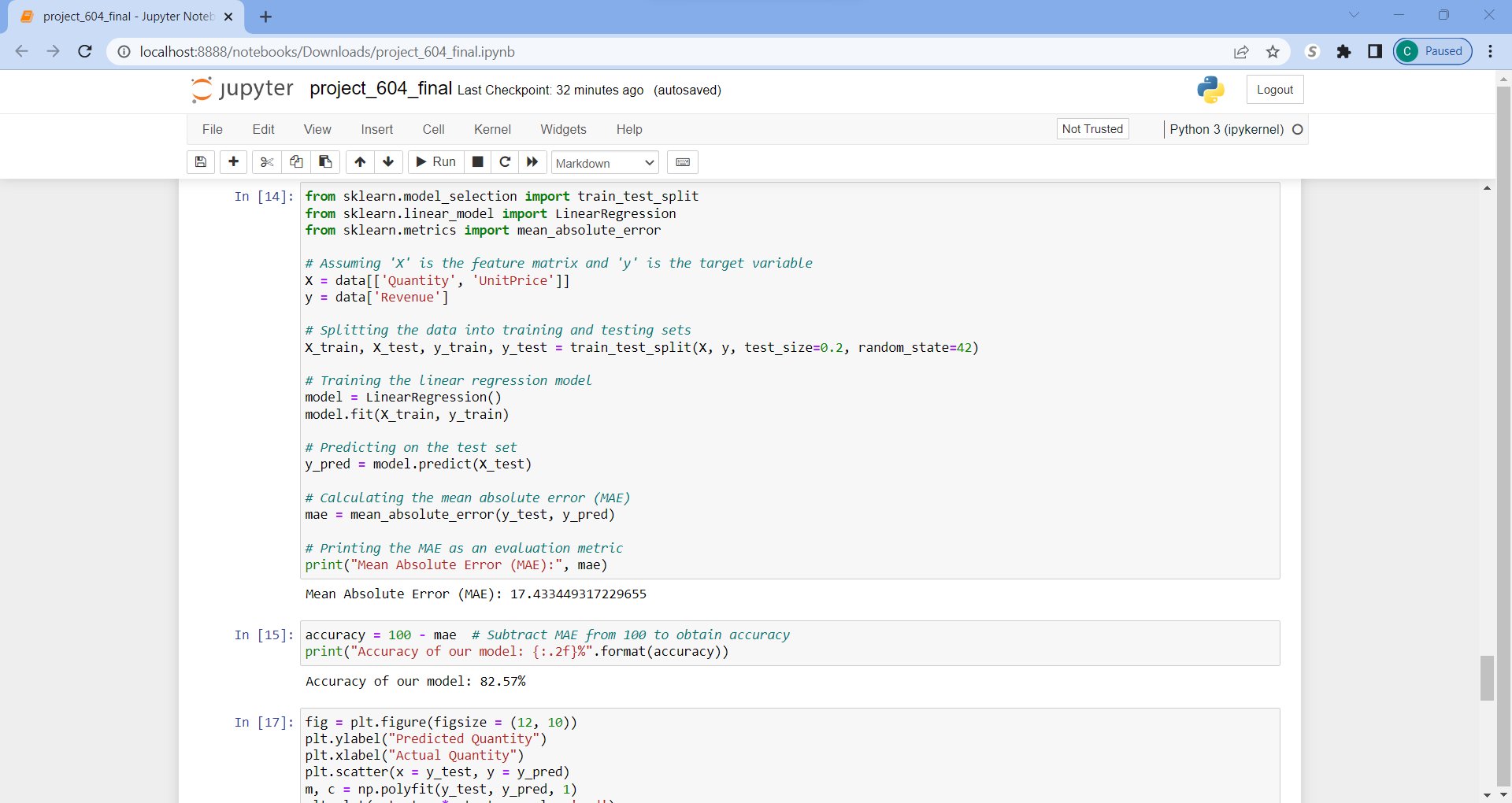
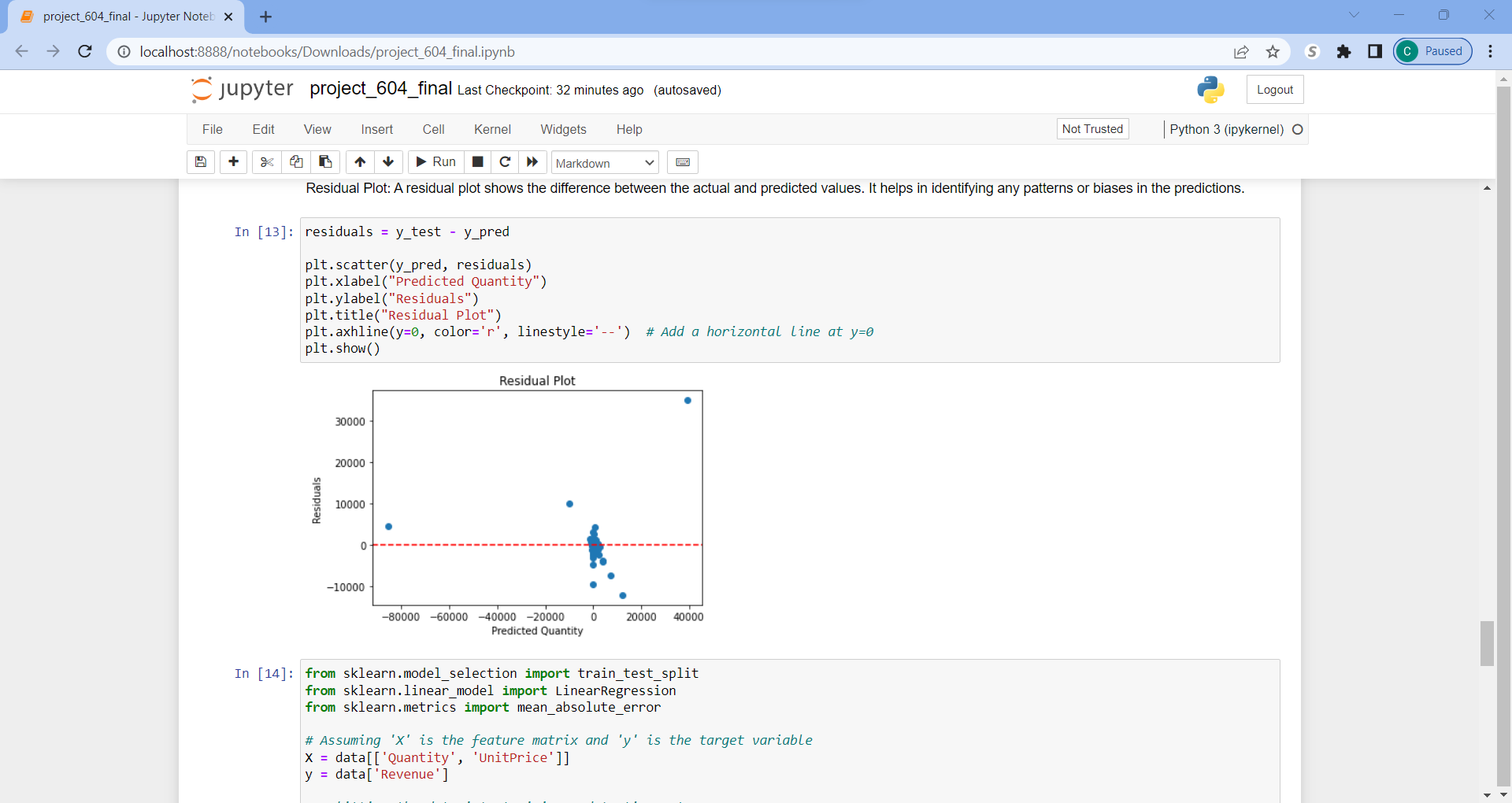
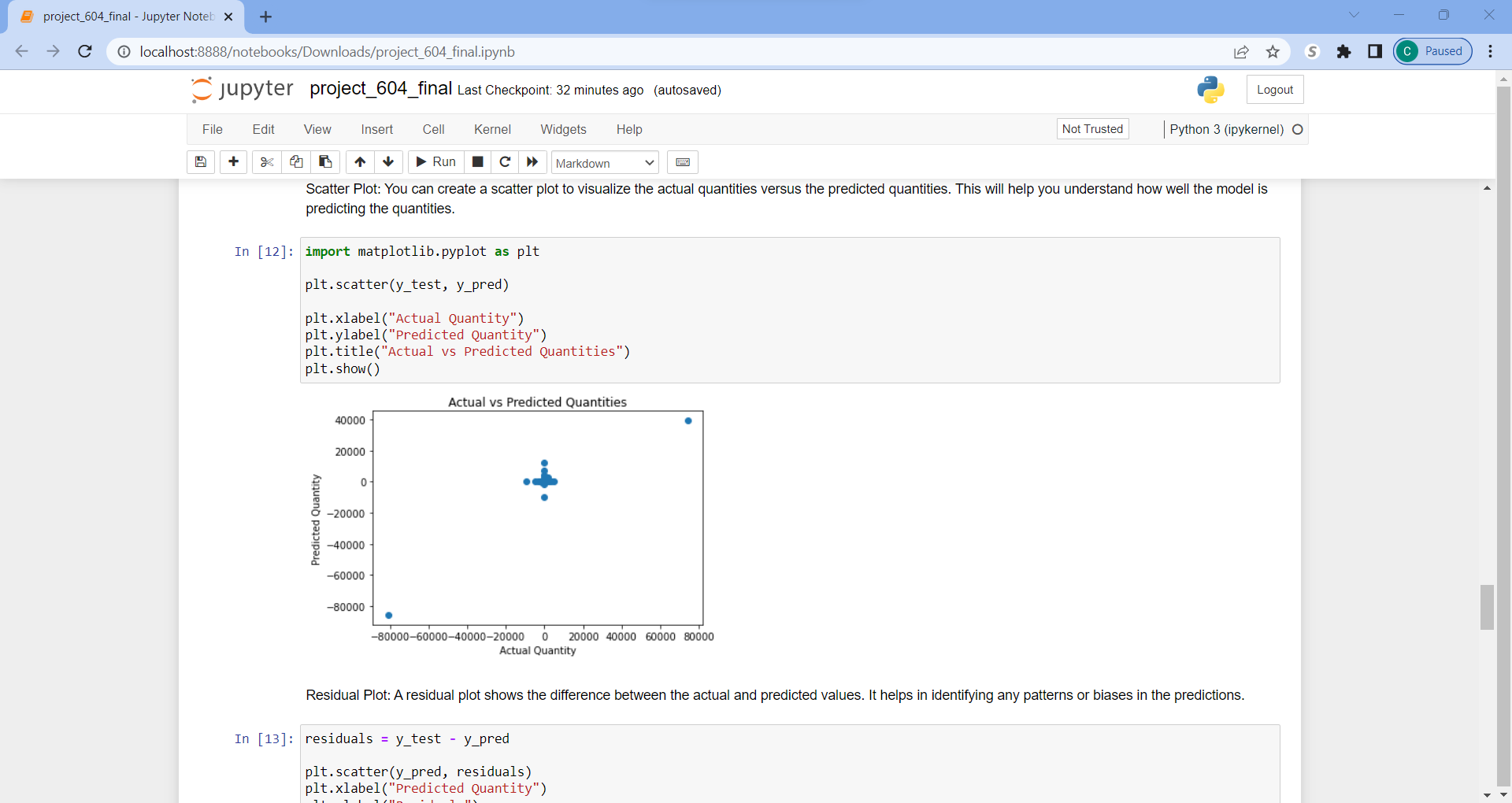
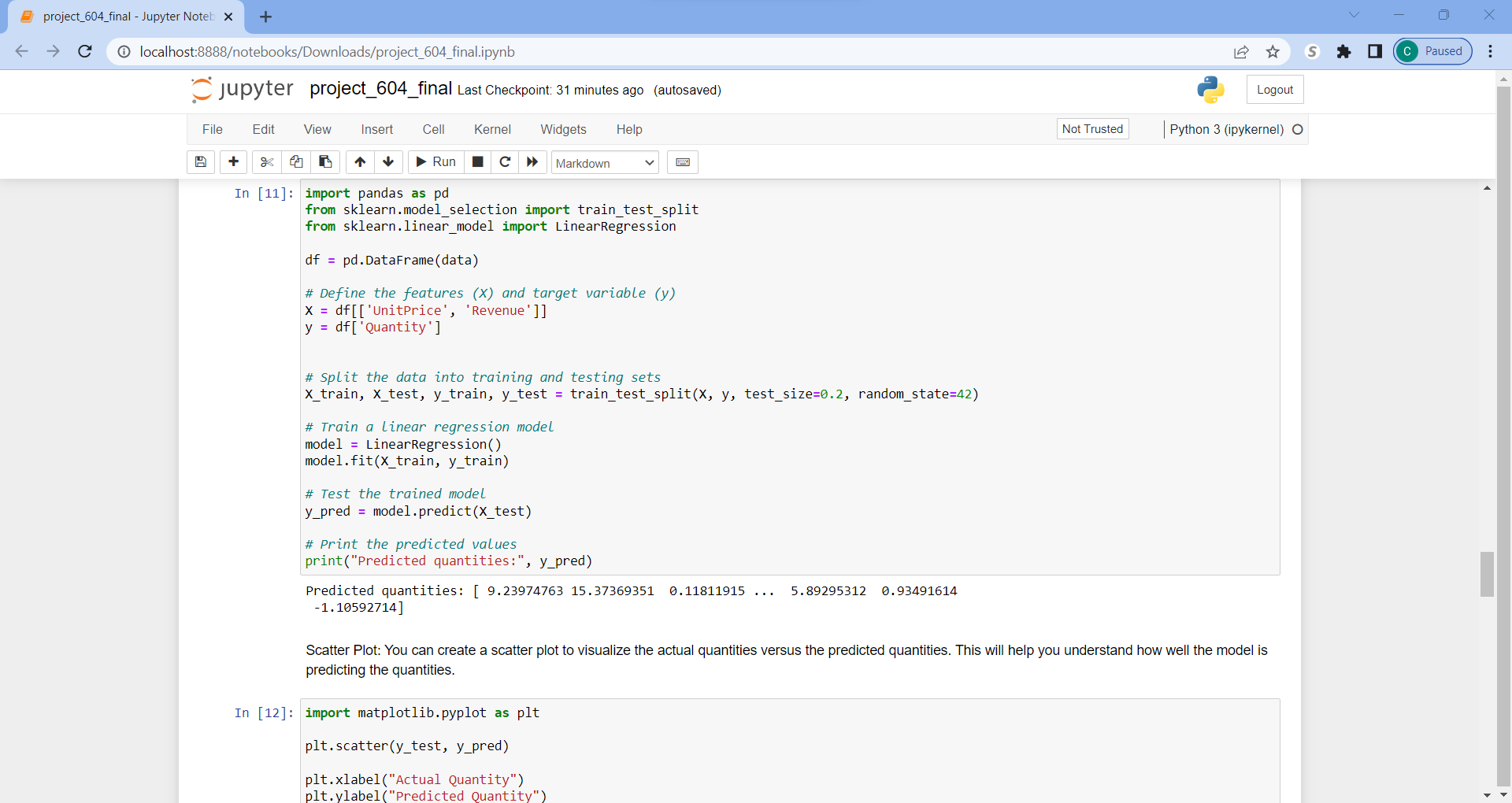
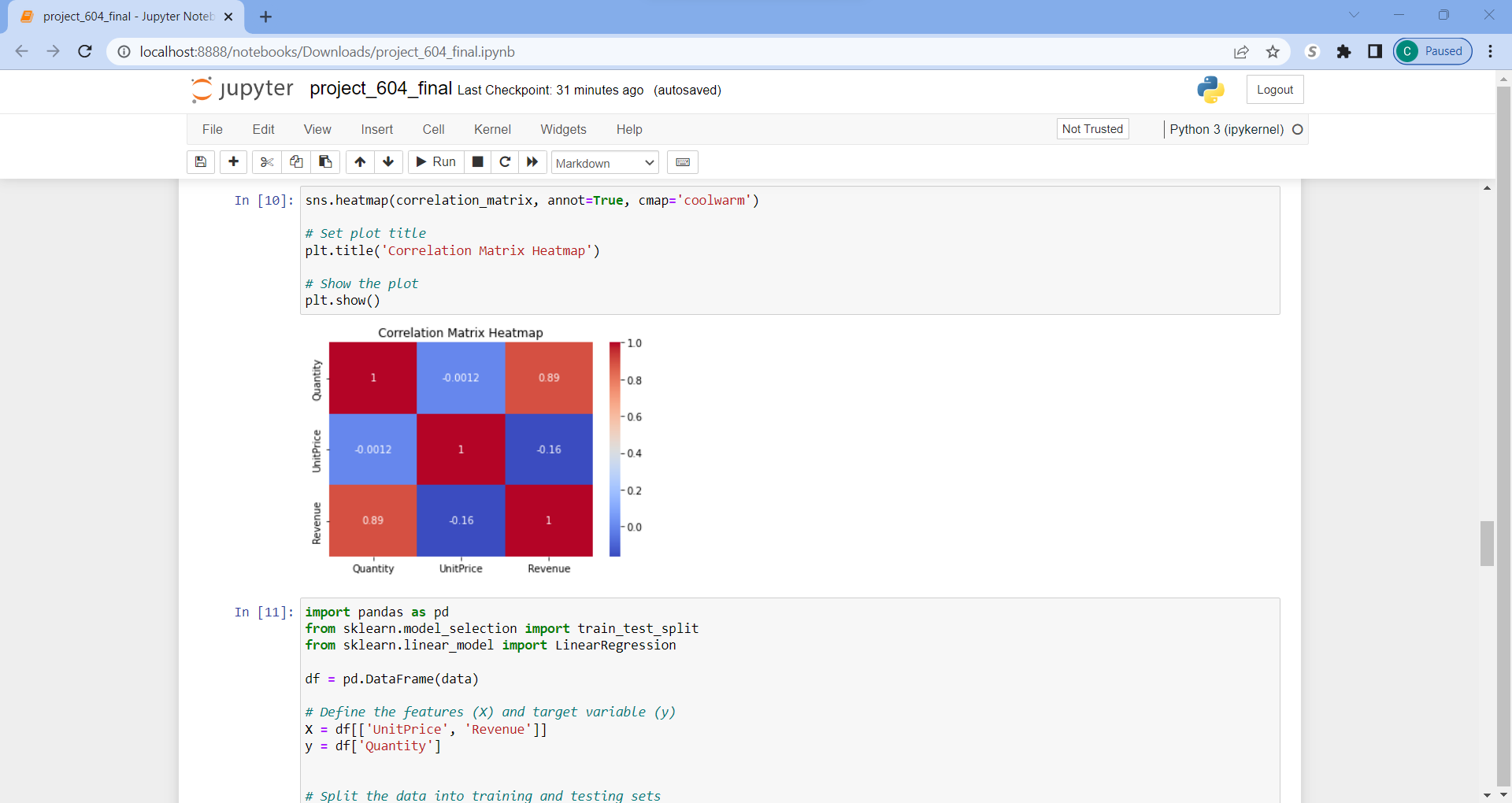
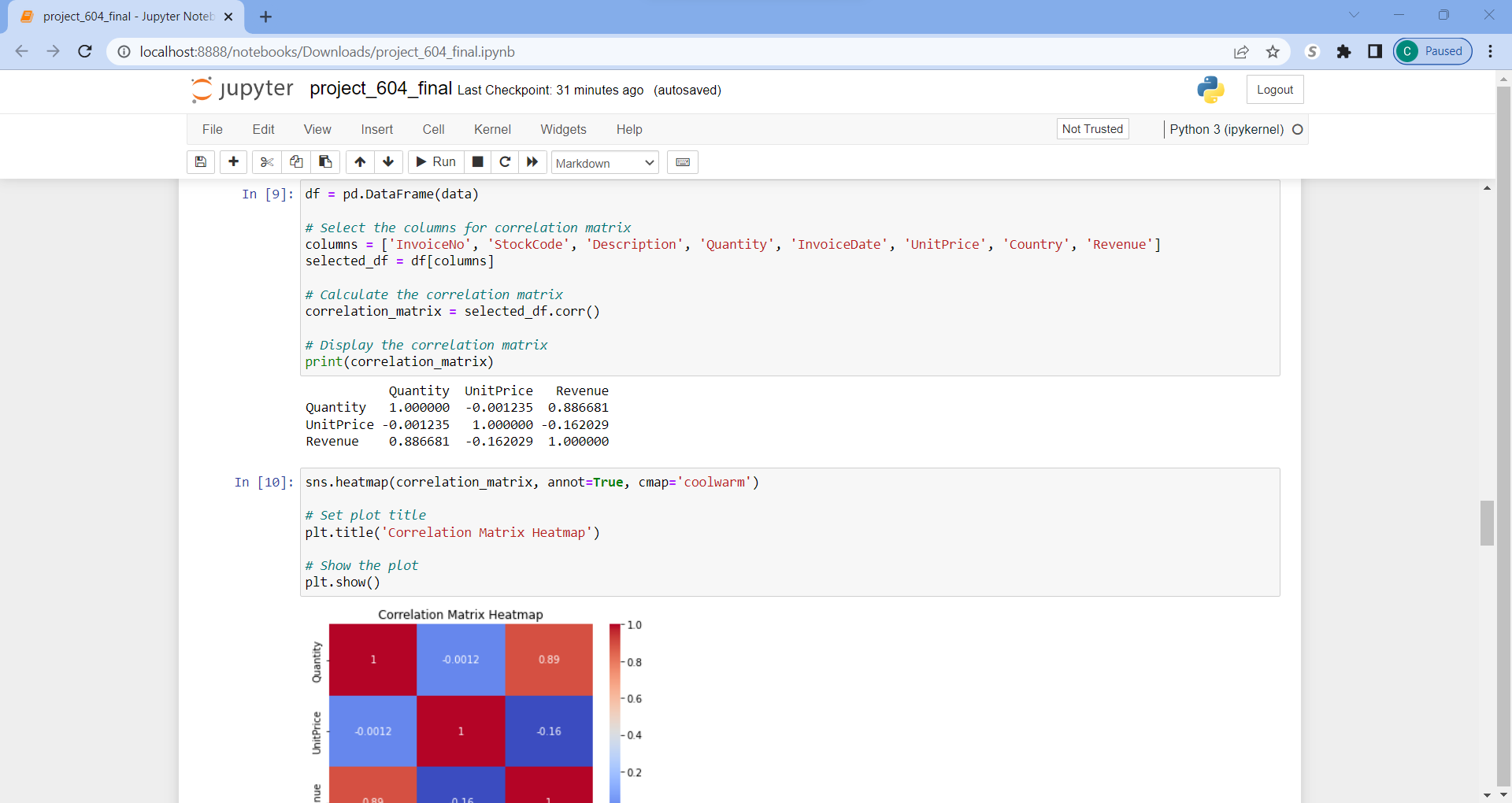
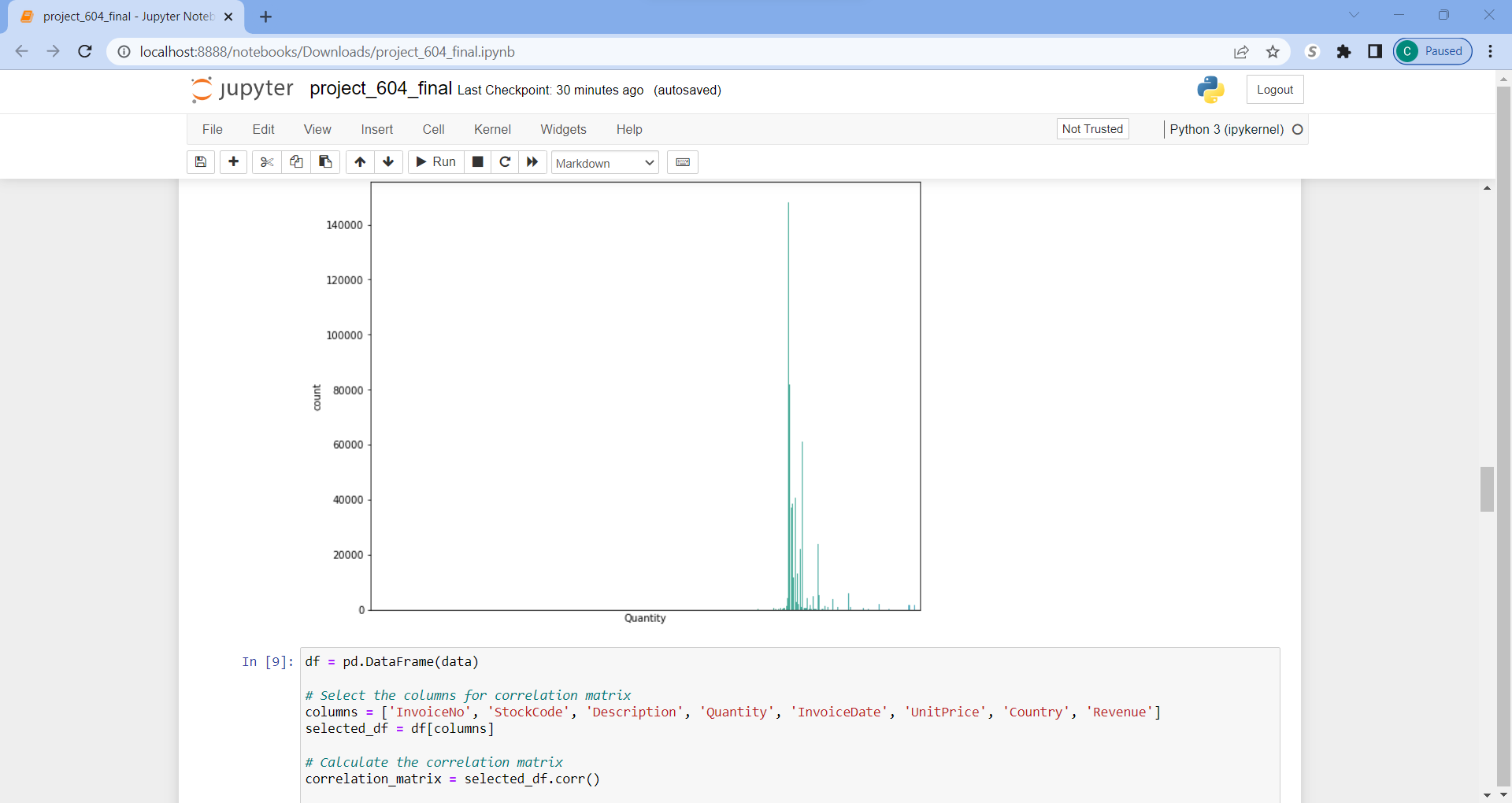
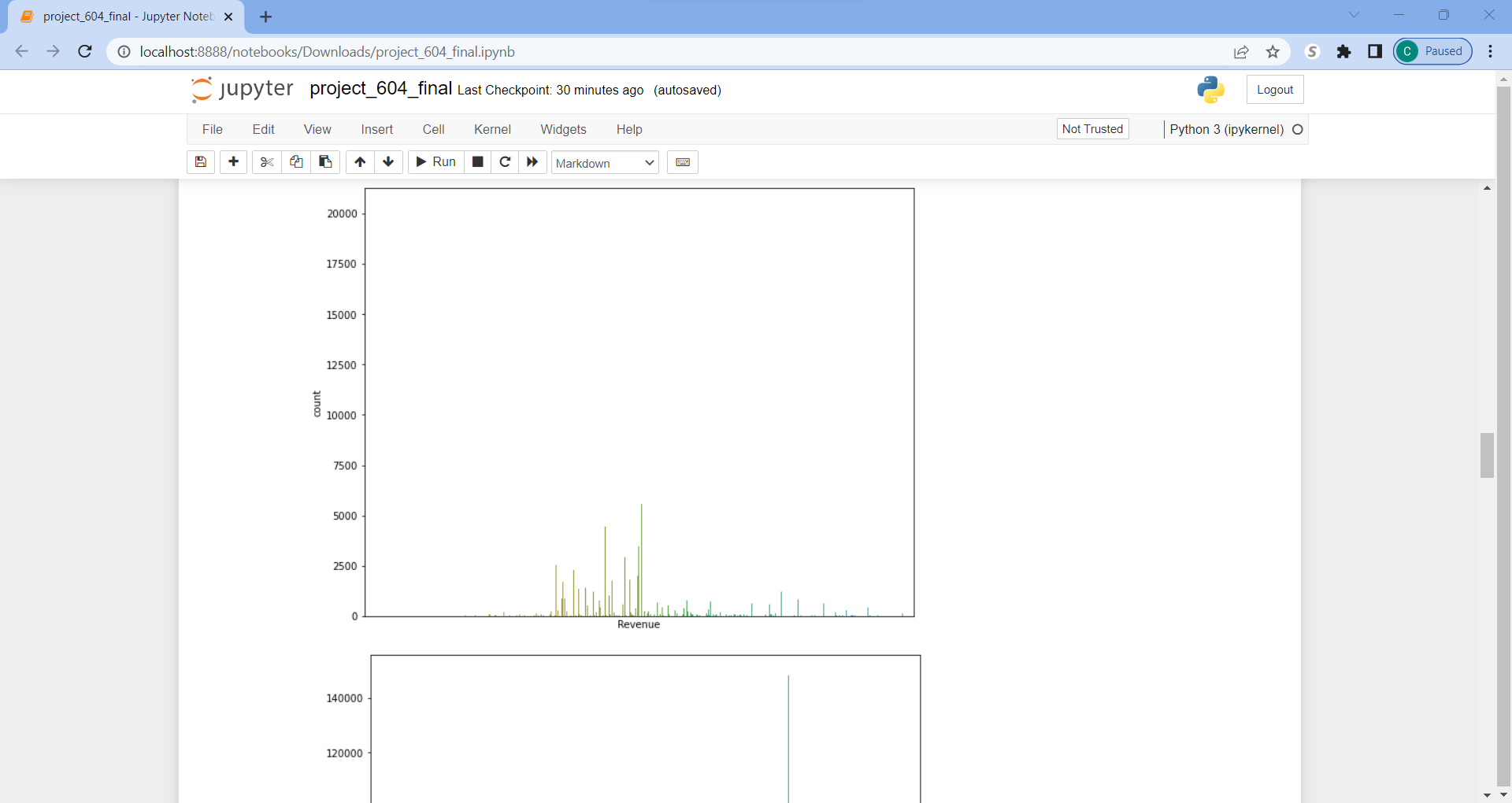
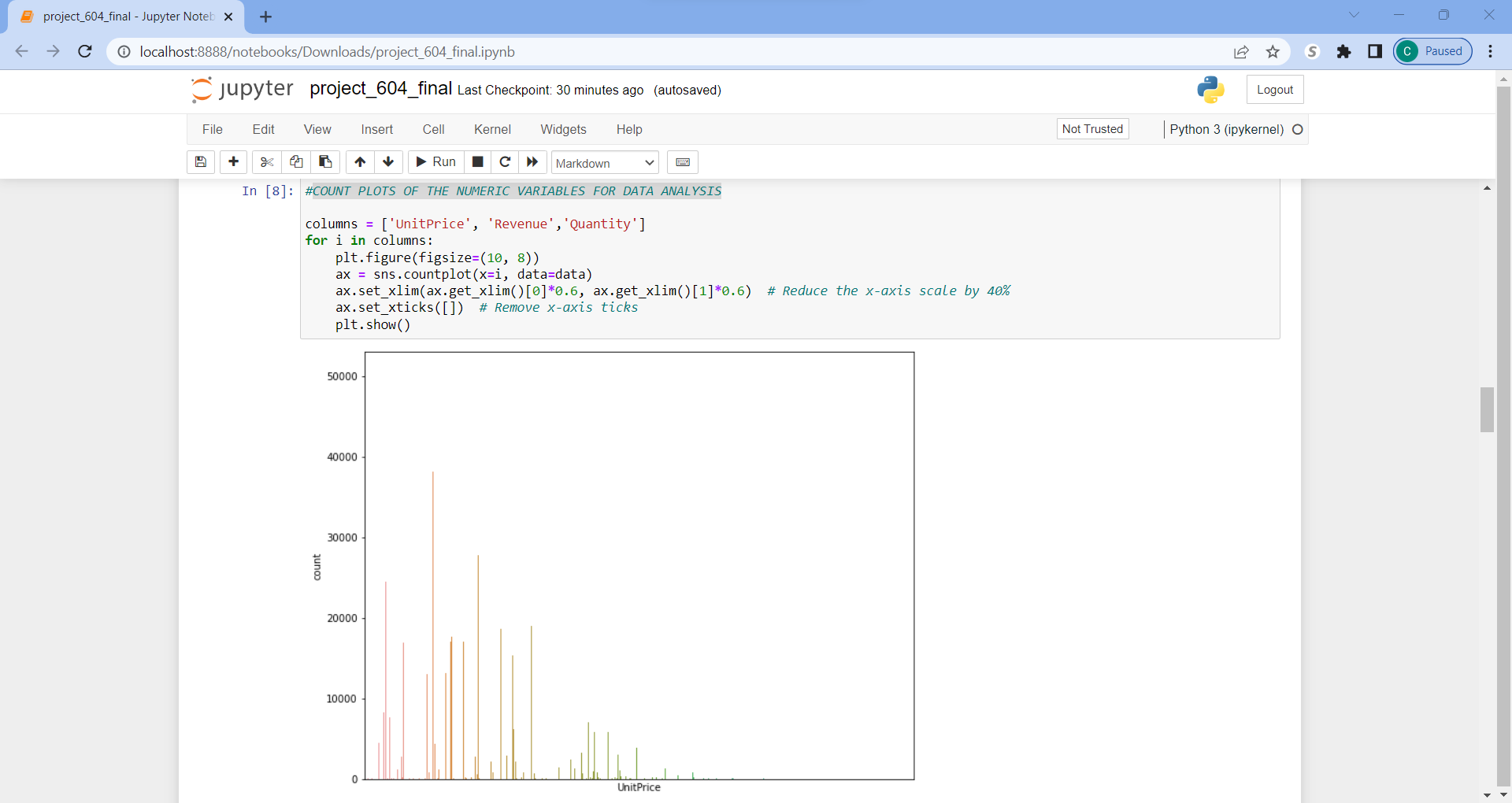
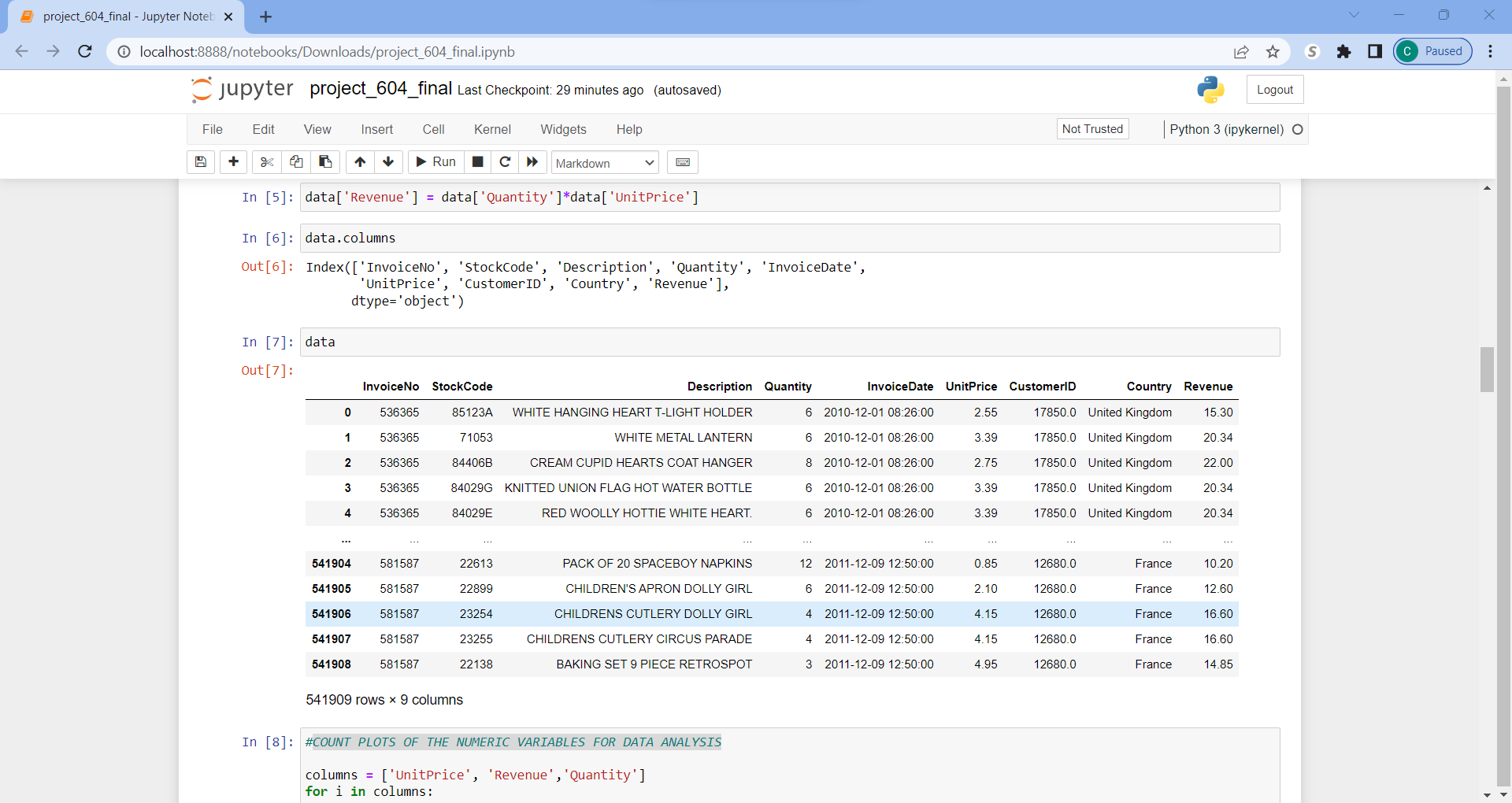
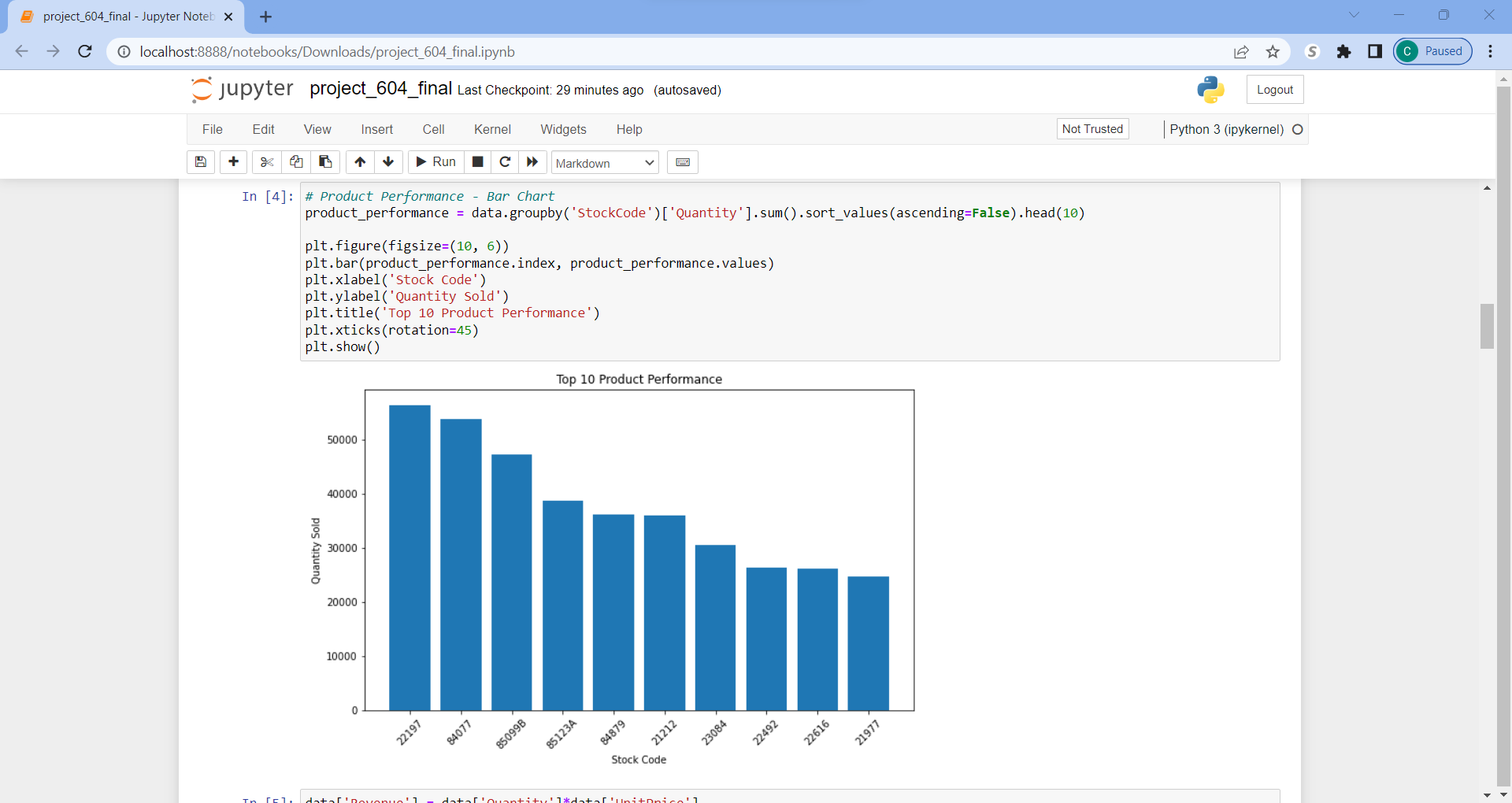
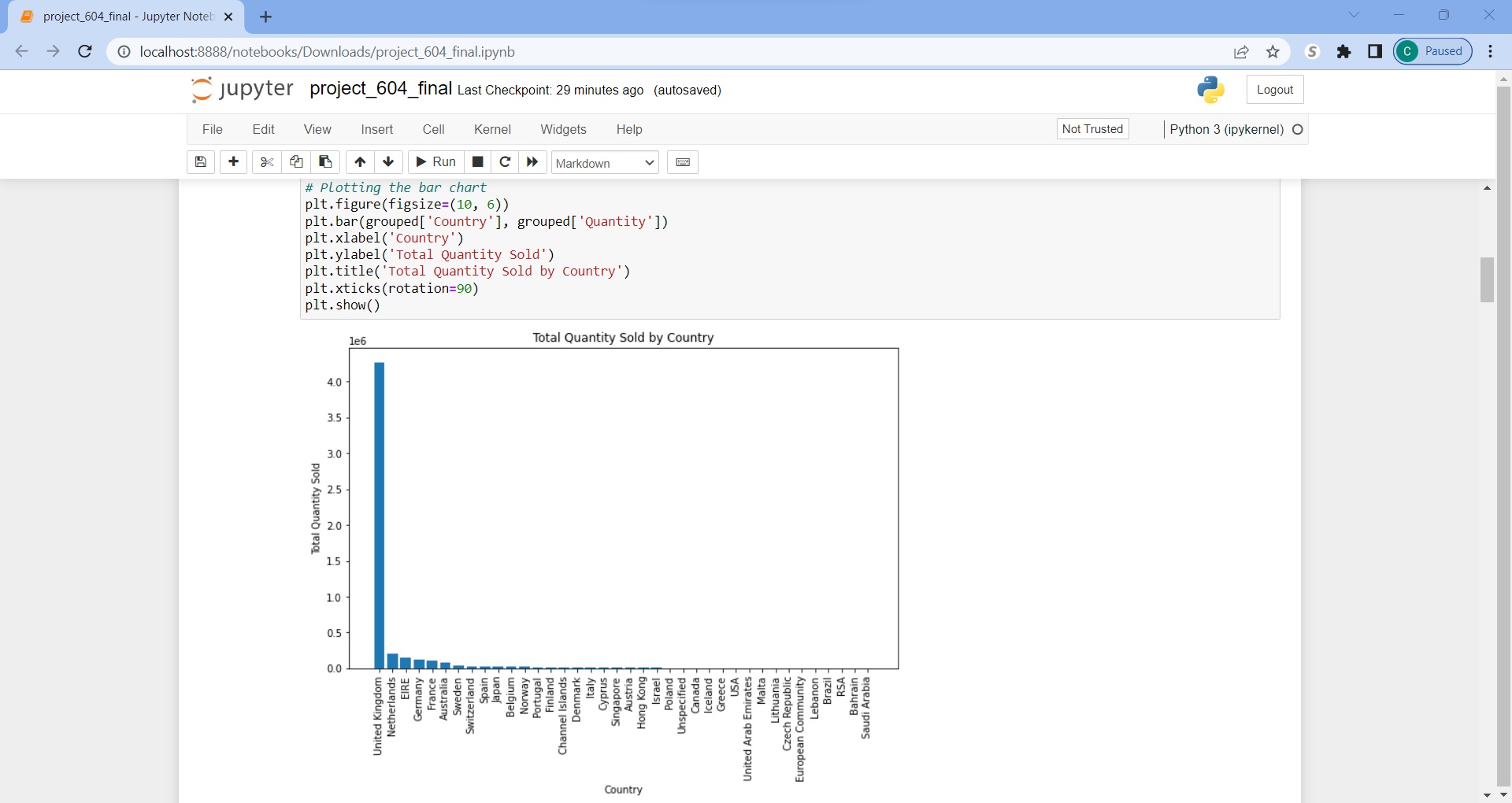
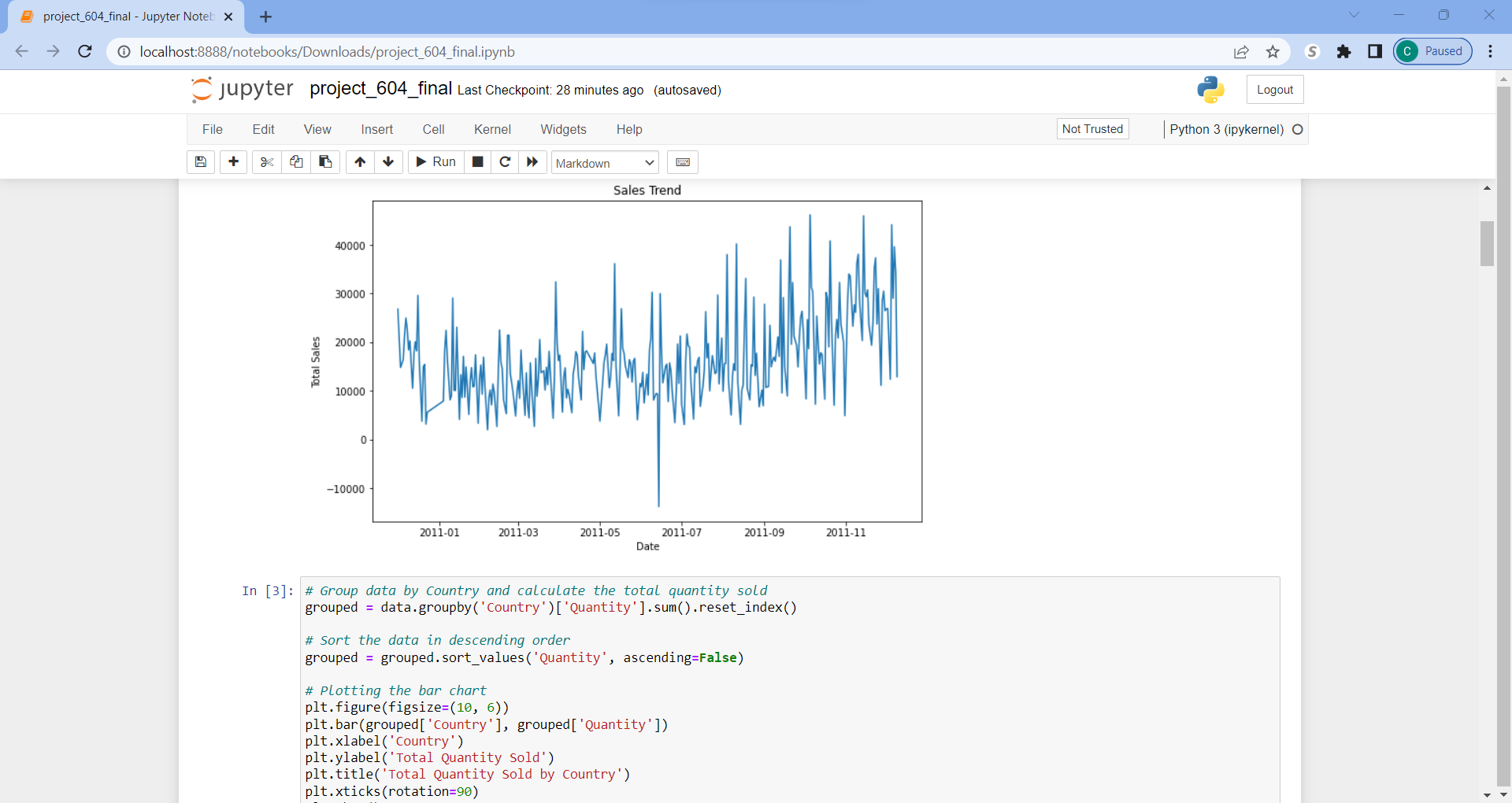
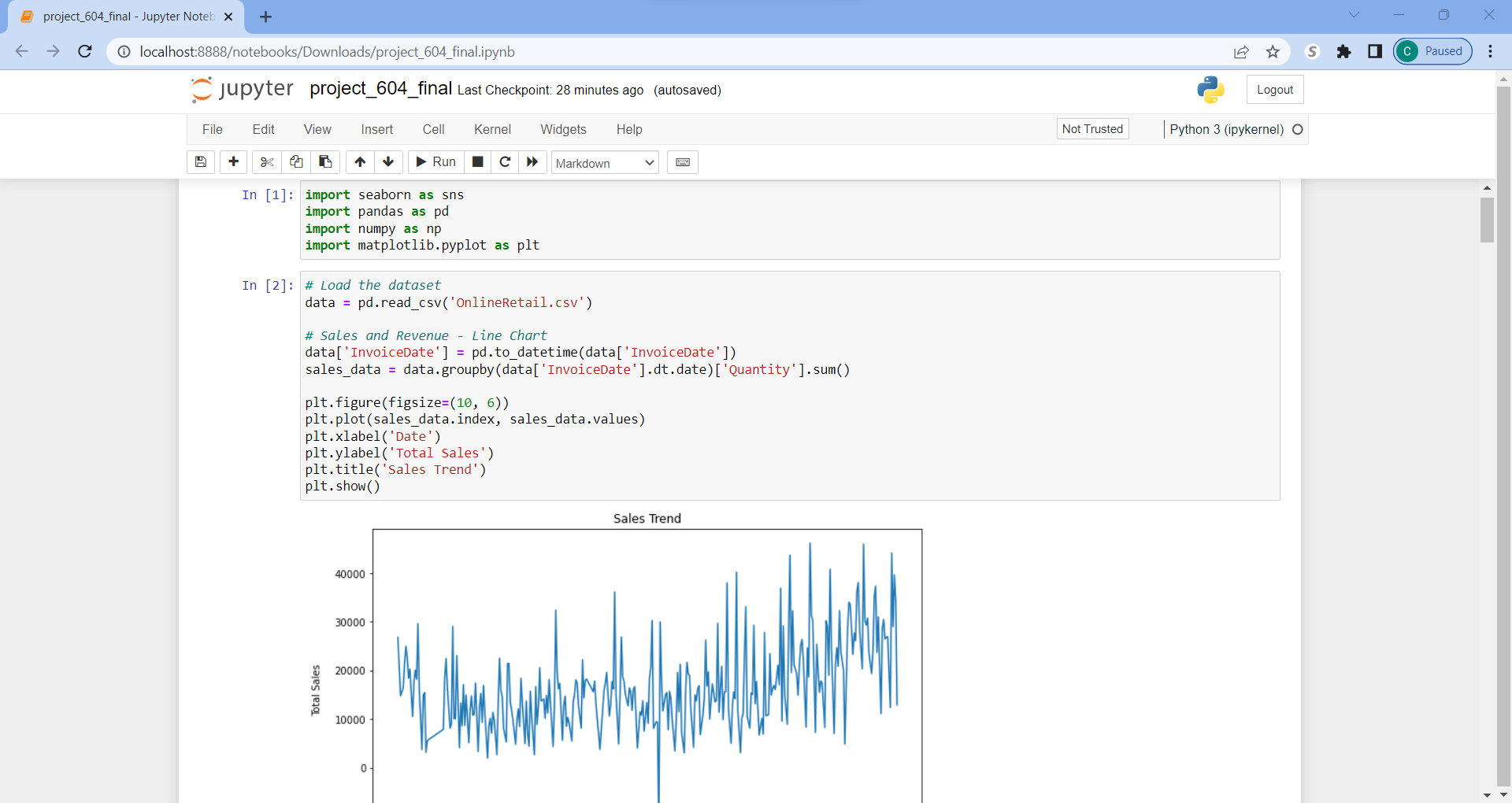
Eliaçık, E. (2022, August 31). *Data Conomy*. Retrieved from [https://dataconomy.com/2022/08/31/machine-learning-in-retail-use-cases/#:~:text=A%20machine%20learning%20model%20for,consumer%20needs%20through%20appropriate%20segmentation](https://dataconomy.com/2022/08/31/machine-learning-in-retail-use-cases/%23:~:text=A%20machine%20learning%20model%20for,consumer%20needs%20through%20appropriate%20segmentation)

Gavett, G. (2014, May 09). *Harvard Business Review*. Retrieved from <https://hbr.org/2014/05/how-data-visualization-answered-one-of-retails-most-vexing-questions>

Valluri, S. (2020, March 10). *Oracle Database Insider*. Retrieved from <https://blogs.oracle.com/database/post/six-retail-dashboards-for-data-visualizations>

**10. APPENDIX**

Below are the screenshots from the code:



# References

*Capella*. (2022, December 30). Retrieved from The Power of Data Management in Retail and e-Commerce: <https://www.capellasolutions.com/blog/data-management-in-retail-and-e-commerce>

Dee, C. (2023, March 15). *Algolia*. Retrieved from <https://www.algolia.com/blog/ecommerce/how-to-increase-your-sales-and-roi-with-optimized-ecommerce-merchandising/>

Eliaçık, E. (2022, August 31). *Data Conomy*. Retrieved from [https://dataconomy.com/2022/08/31/machine-learning-in-retail-use-cases/#:~:text=A%20machine%20learning%20model%20for,consumer%20needs%20through%20appropriate%20segmentation](https://dataconomy.com/2022/08/31/machine-learning-in-retail-use-cases/%23:~:text=A%20machine%20learning%20model%20for,consumer%20needs%20through%20appropriate%20segmentation)

Gavett, G. (2014, May 09). *Harvard Business Review*. Retrieved from <https://hbr.org/2014/05/how-data-visualization-answered-one-of-retails-most-vexing-questions>

Valluri, S. (2020, March 10). *Oracle Database Insider*. Retrieved from <https://blogs.oracle.com/database/post/six-retail-dashboards-for-data-visualizations>