**Mini Project Report on**



**Big Mart Sales Prediction**



**Submitted in partial fulfilment of the requirement for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE & ENGINEERING**

**Submitted by:**

**Student Name TANMAY SAXENA University Roll No. 2019179**

***Under the Mentorship of***

**Mentor Name: Dr. SURENDRA KUMAR SUKLA**

**Professor**



**Department of Computer Science and Engineering**

**Graphic Era (Deemed to be University)**

**Dehradun, Uttarakhand**

**July-2023**



**CANDIDATE’S DECLARATION**

I hereby certify that the work which is being presented in the project report entitled **“Big Mart Sales Prediction”** in partial fulfillment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineeringof the Graphic Era (Deemed to be University), Dehradun shall be carried out by the under the mentorship of **Dr. SURENDRA KUMAR SUKLA, Professor**, Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun.

Name TANMAY SAXENA University Roll no 2019179

**Table of Contents**

|  |  |  |  |
| --- | --- | --- | --- |
| **Chapter No.** | | **Description** | **Page No.** |
| Chapter 1 | | Introduction | **1-2** |
| Chapter 2 | | Literature Survey | **3-4** |
| Chapter 3 | | Methodology | **5-12** |
|  |
| Chapter 5 | | Conclusion and Future Work | **13-14** |
|  | | References | **15** |

**Chapter 1**

**Introduction**

The daily competition between different malls as well as big malls is becoming more and more intense because of the rapid rise of international supermarkets and online shopping’s. Every mall or mart tries to provide personal and short-term donations or benefits to attract more and more customers on a daily basis, such as the sales price of everything which is usually predicted to be managed through different ways such as corporate asset management, logistics, and transportation service, etc. Current machine learning algorithms that are very complex and provide strategies for predicting or predicting long-term demand for a company's sales, which now also help in overcoming budget and computer programs. In this report, we basically discuss the subject of specifying a large mart sale or predicting an item for a customer’s future need in a few supermarkets in various locations and products that support the previous record. Various ML algorithms such as linear regression, randomforest, etc. are used to predict sales volume. As we know, good marketing is probably the lifeblood of all organizations, so sales forecasting now plays an important role in any shopping mall. It is always helpful to predict the best, and develop business strategies about useful markets and to improve market knowledge. Regular sales forecasting research can help in-depth analysis of pre-existing conditions and conditions and then, assumptions are often used in terms of customer acquisition, lack of funding, and strength before setting budgets and marketing plans for the coming year. In other words, sales forecasts are predicted on existing services of the past. In- depth knowledge of the past is required to develop and enhance market opportunities no matter what the circumstances, especially the external environment, which allows to prepare for the future 2 needs of the business. Extensive research is ongoing in the retailer’s domain to predict long term sales demand. An important and effective method used to predict the sale of a mathematical method, also called the conventional method, but these methods take more time to predict sales. And these methods could not manage indirect data so to overcome these problems in traditional methods the machine learning techniques used. ML methods can handle not only indirect data but also large data sets well.

**PROBLEM STATEMENT:-** Due to increasing competition many malls and bigmart are trying their best to stay ahead in competition .In order to find out what are the various factors which affect the sales of bigmart and what strategies one needs to employ in order to gain more profit one need to have some model on which they can rely .So a predictive model can be made which could help to gain useful information and increase profit.

**Chapter 2**

**Literature Survey**

Earlier we are having different methods for predicting the sales such as:

Expert’s Opinion Method: Here, marketing and sales professionals will make sales predictions. Analysis and sales forecasting are labor-intensive tasks. However, predictions can also be inaccurate for a variety of reasons, including a lack of enthusiasm for the task being performed and other problems.

Kadam ,et.al [1] have suggested when the prediction for the sales for bigmart was done using the algorithm like random forest and LR for prediction analysis it gave lesser accuracy .So to overcome this problem we can use another algorithm which is XG boost algorithm which not only gives better accuracy but also is more efficient.

Makridakis, et.al [2] have suggested predicting methods and applications containing DataLack and short life cycles. So some data like historical data, consumer-focused markets face uncertain needs, which can be an accurate predictor of outcome.

C. M. Wu , et.al [3] have suggested comparison of Different ML Algorithms for MultipleRegression on Black Friday Sales Data used the concept of neural network to compare the various different algorithms .Using neural network as the concept which is very complex and less efficient concluded that we should use much simpler algorithm for the prediction purposeIn 2012 O. Ajao Isaac, A. Abdullahi Adedeji, I. Raji Ismail Regression analysis is used across business fields for tasks as diverse as systematic risk estimation, production and operations management, and statistical inference. This paper presents the cubic polynomial least square regression as a robust alternative method of making cost prediction in business rather than the usual linear regression. The study reveals that polynomial regression is a better alternative with a very high coefficient of determination.

In 2013 X. Yua, Z. Qi, Y. Zhao Advances in information technologies have changed our lives in many ways. There is a trend that people look for news and stories on the internet. Under this circumstance, it is more urgent for traditional media companies to predict print's sales than ever. Previous approaches in newspapers/magazines sales forecasting are mainly focused on building regression models based on sample data sets. But such regression models can suffer from the over- fitting problem. Recent theoretical studies in statistics proposed a novel method, namely support vector regression (SVR), to overcome the over-fitting problem. This study, therefore, applied support vector regression to the newspaper/magazines sales forecasting problem. The experiment showed that SVR is a superior method.

In 2015 Xinqing Shu, Pan Wang Boosting is one of the algorithms which can boost the accuracy of weak classifiers, and Adaboost has been widely and successfully applied to classification, detection, and data mining problems. In this paper, a new method of calculating parameters, Adaboost-AC, which uses the accelerated good fitness function to acquire the weights of the weak classifiers is presented. The new algorithm is compared with the traditional Adaboost based on the UCI database and its promising performance is shown by the experimental results.

Das, P., Chaudhury Prediction of retail sales of footwear using feedforward and recurrent Neural Networks (2018) Prediction of retail sales of footwear using feedforward and recurrent neural networks used neural networks for prediction of sales. Using the neural network for predicting weekly retail sales, which is not efficient, So XG boost can work efficiently.

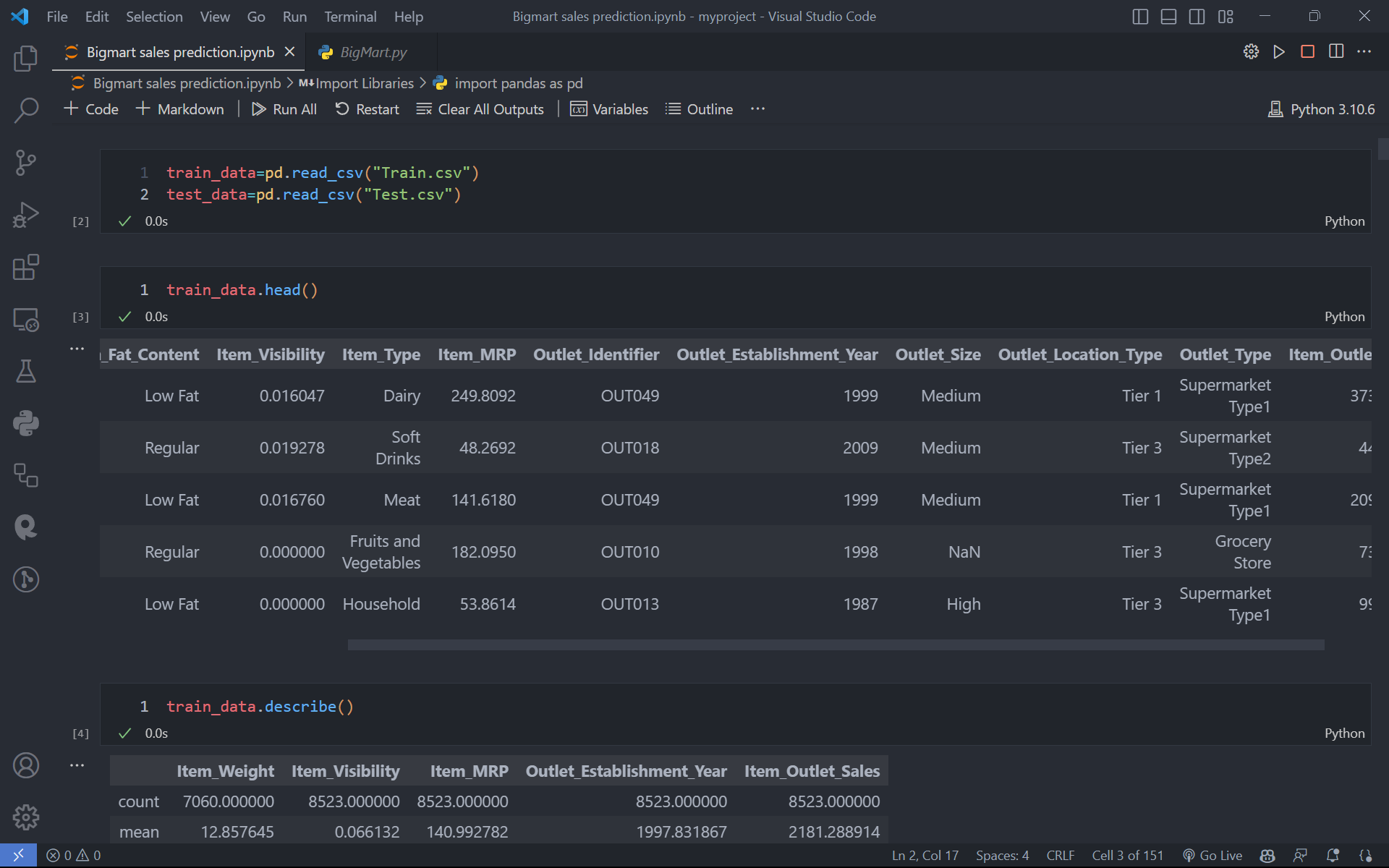
Makridakis.S., Wheelwrigh.S.C., Hyndman. R.J Forecasting methods and applications (2008) Forecasting methods and applications contain a Lack of Data and short life cycles. So, some of the data like historical data, consumer- oriented markets face uncertain demands, can be a prediction for an accurate result.

**Chapter 3**

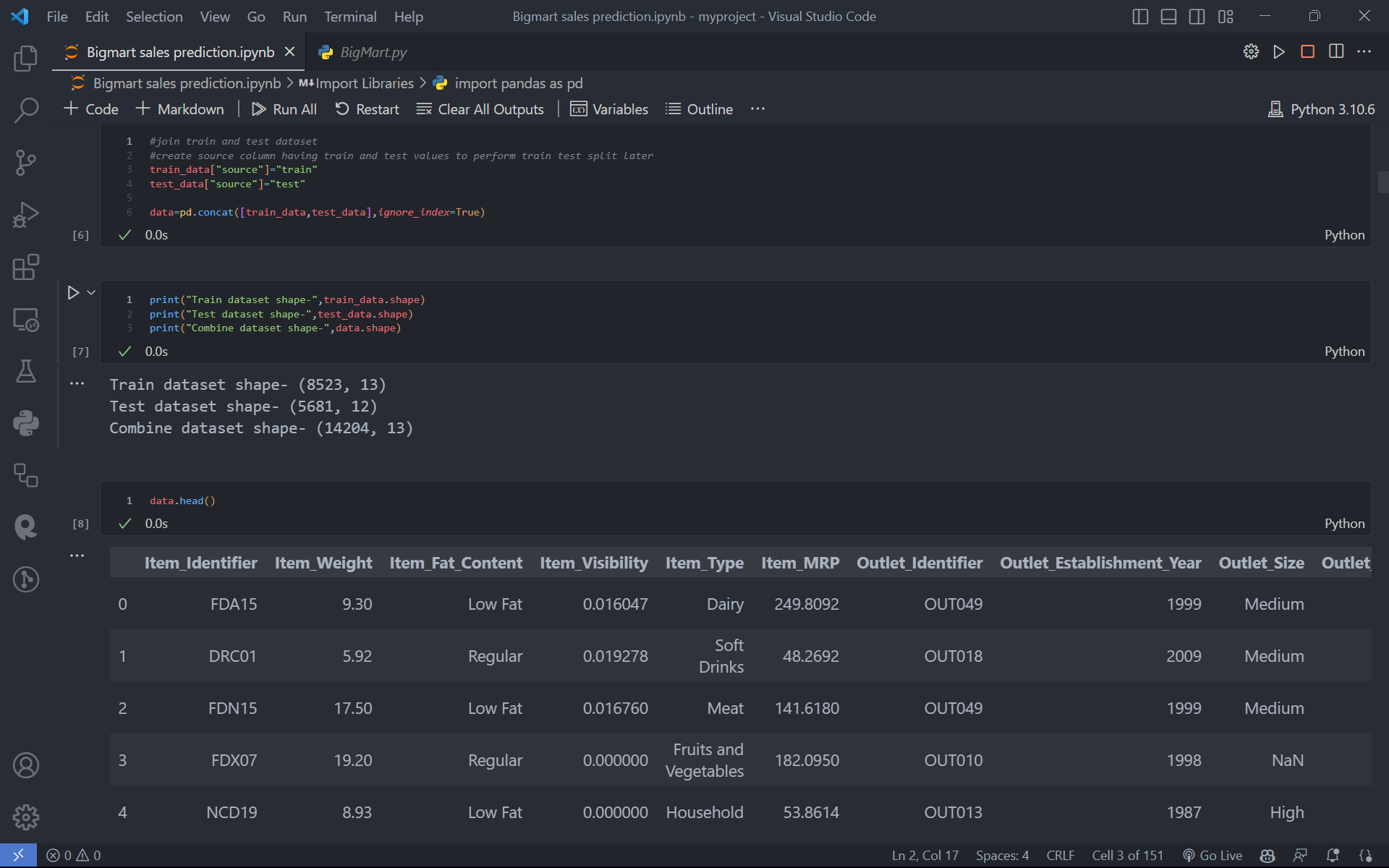
**Methodology**

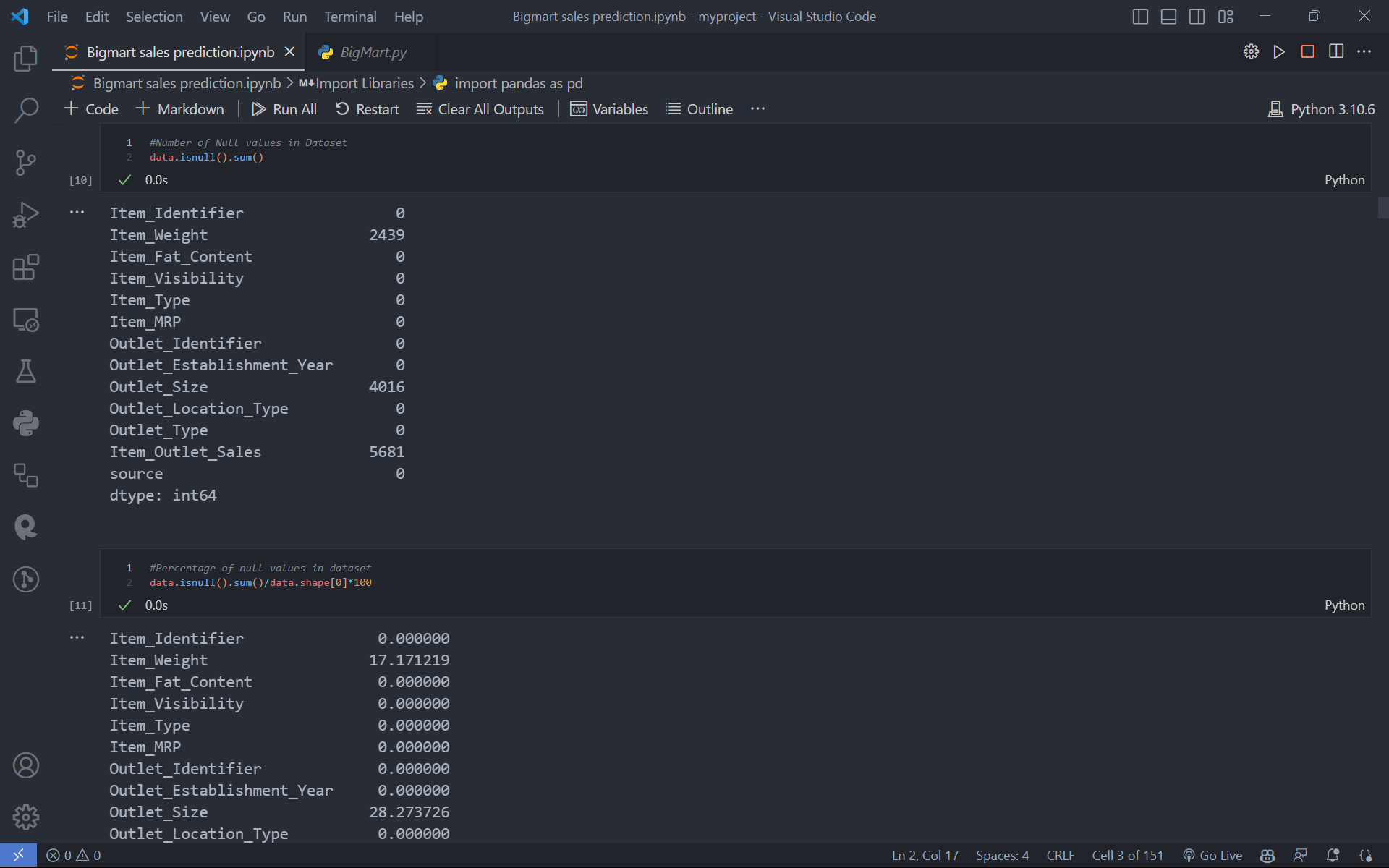
The methodology for predicting sales at Big Mart can vary depending on the specific approach and data available. However, here is a general outline of a common methodology for sales prediction:

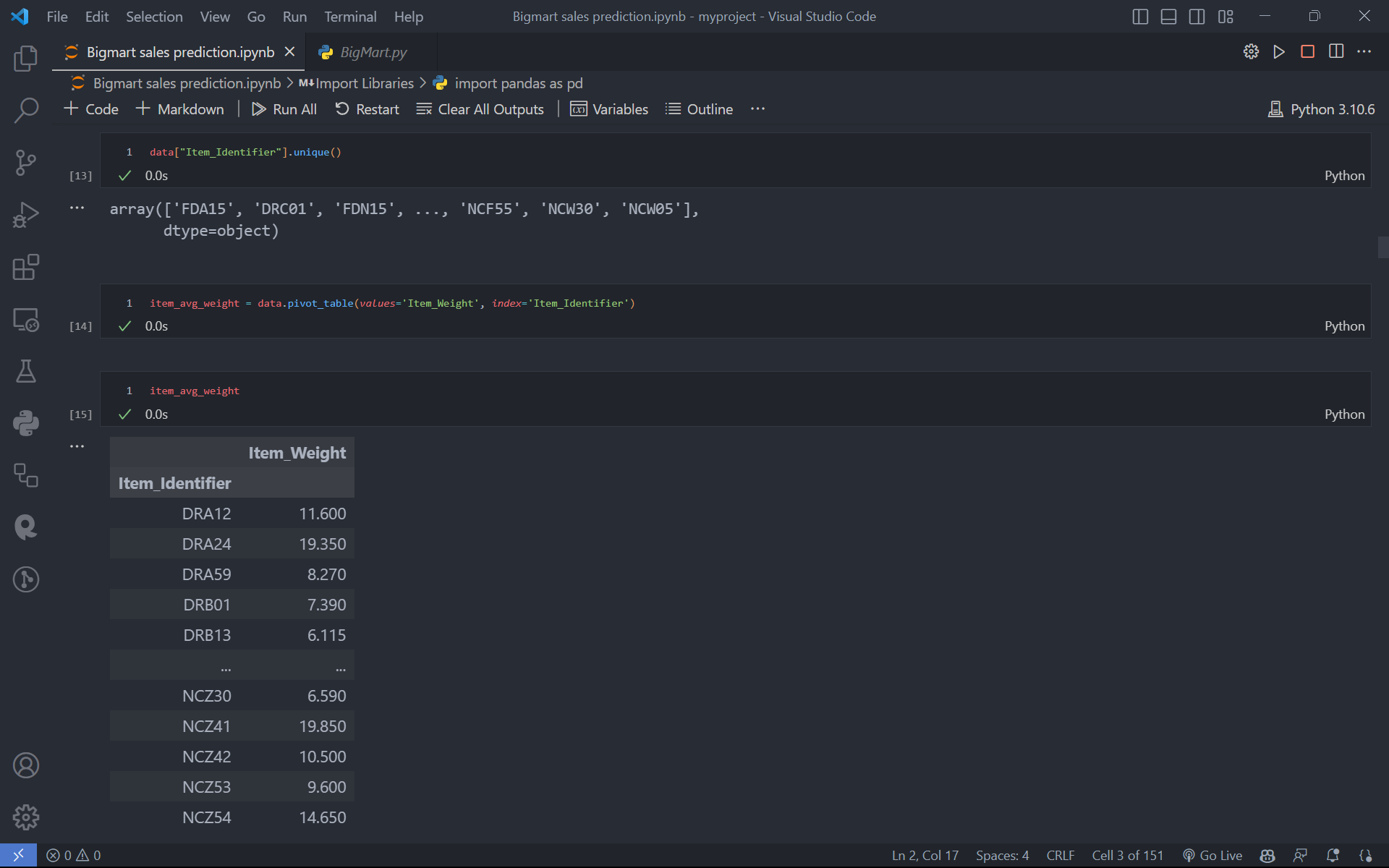
**Data Collection**: Gather relevant data related to sales, such as historical sales records, product information, promotions, customer demographics, store locations, and any other variables that might impact sales. This data can be obtained from internal databases or external sources.

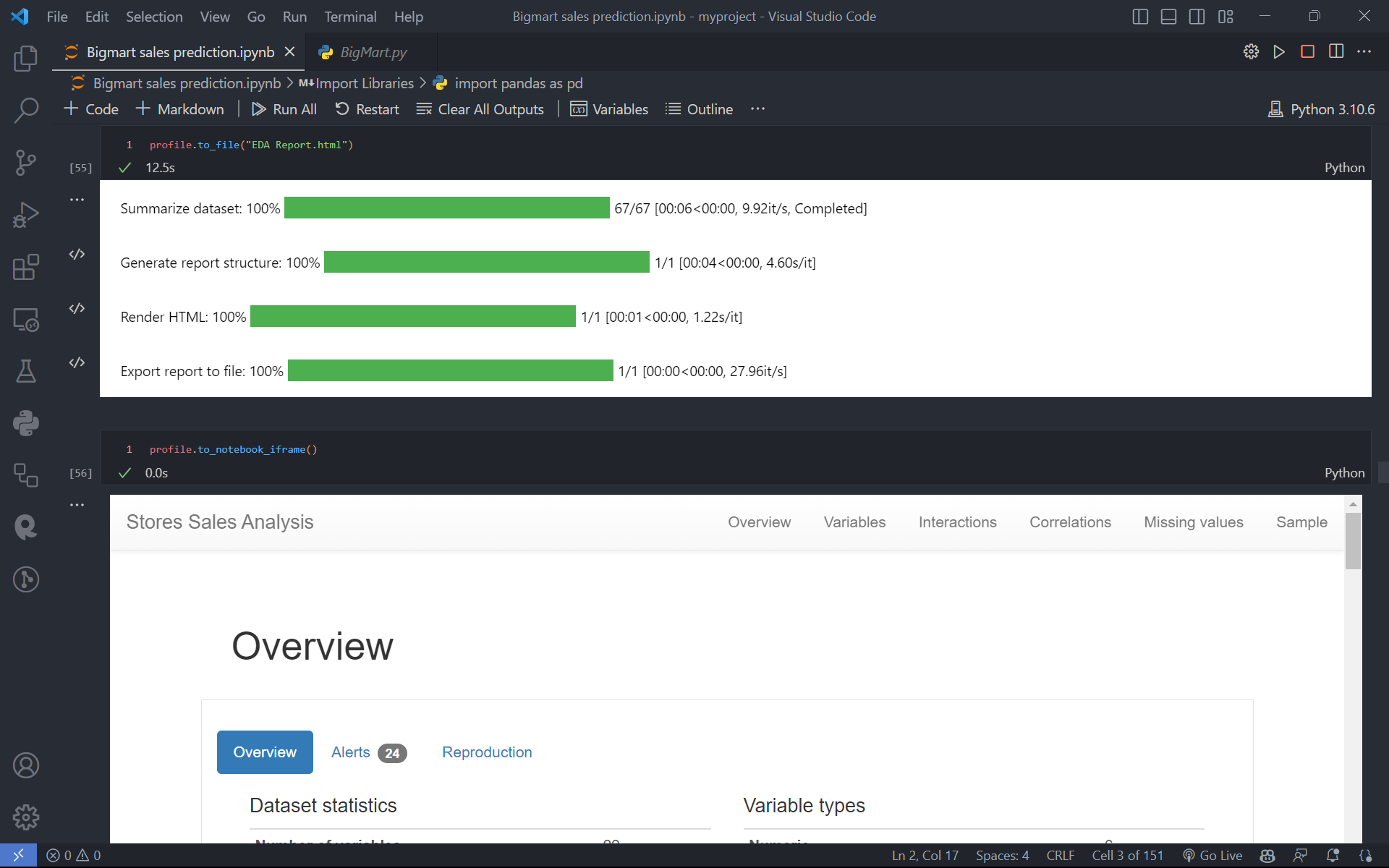
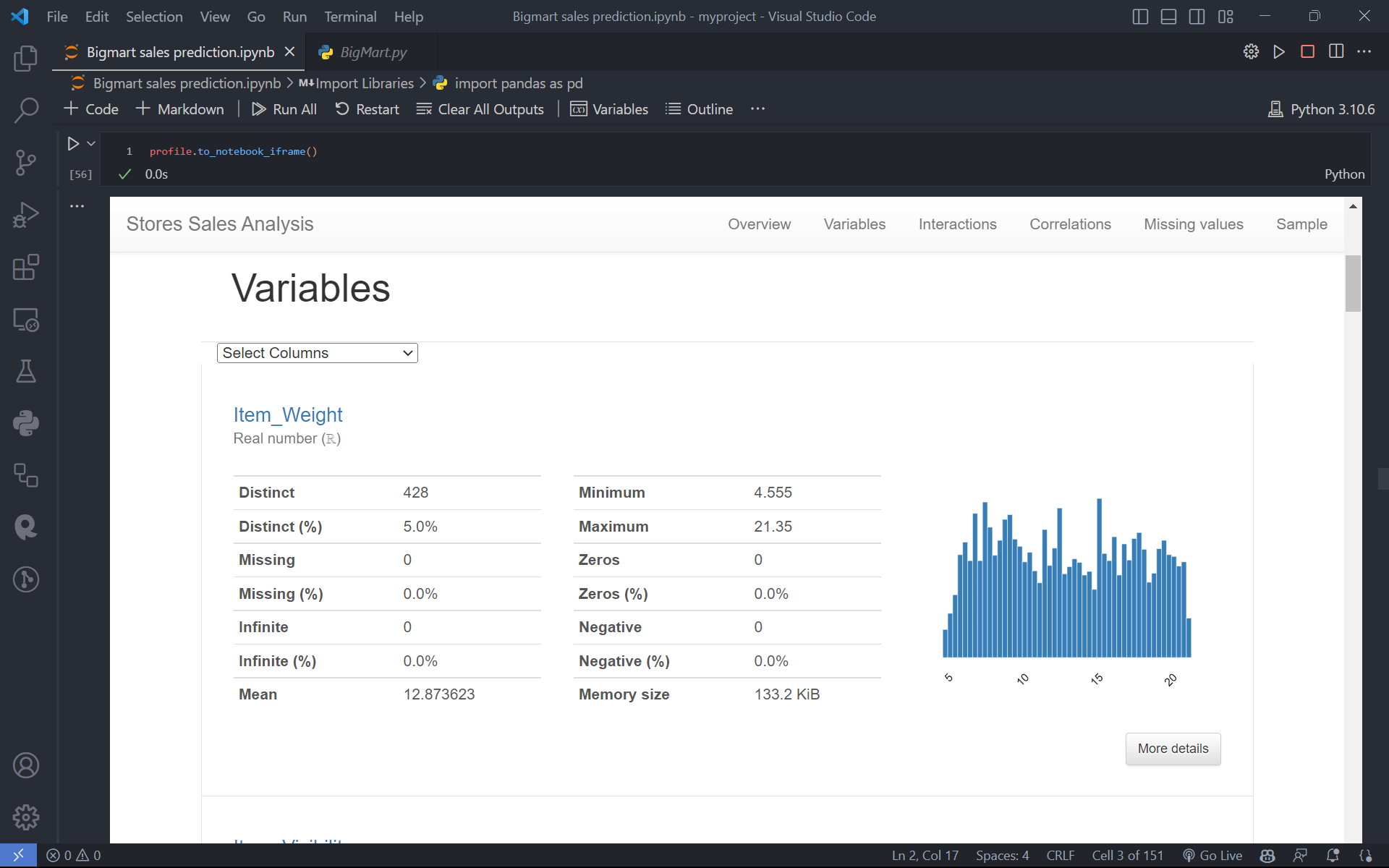
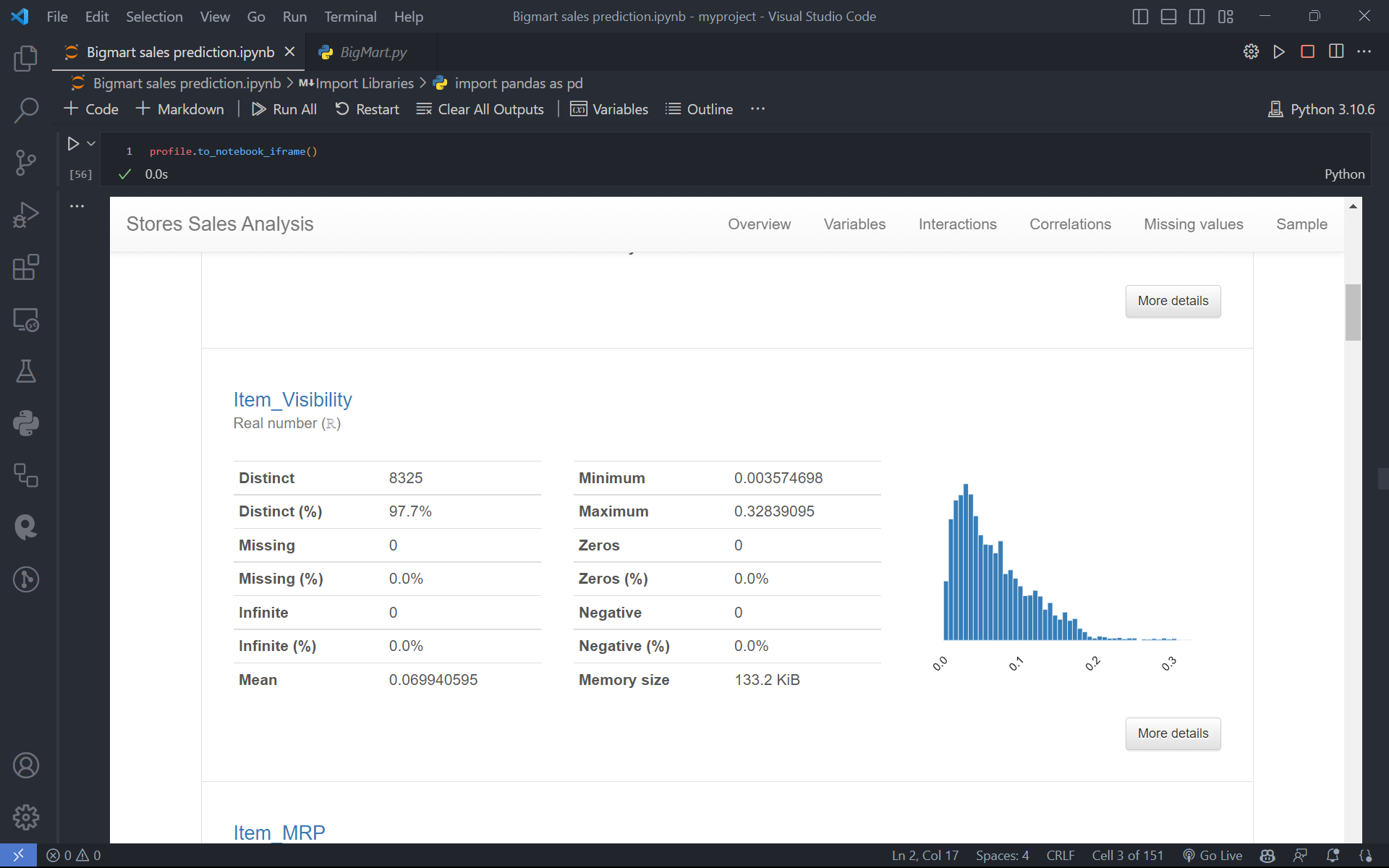
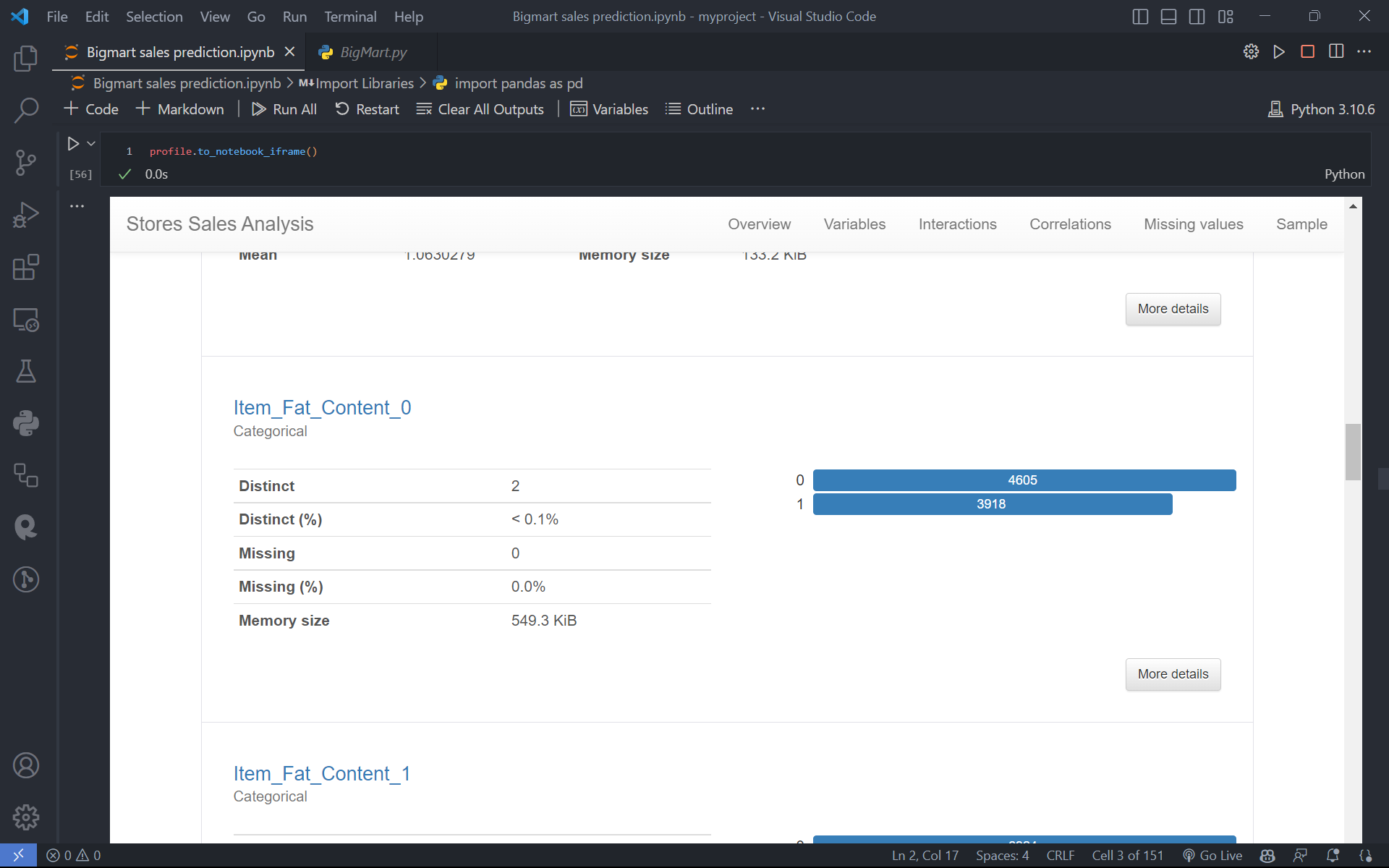
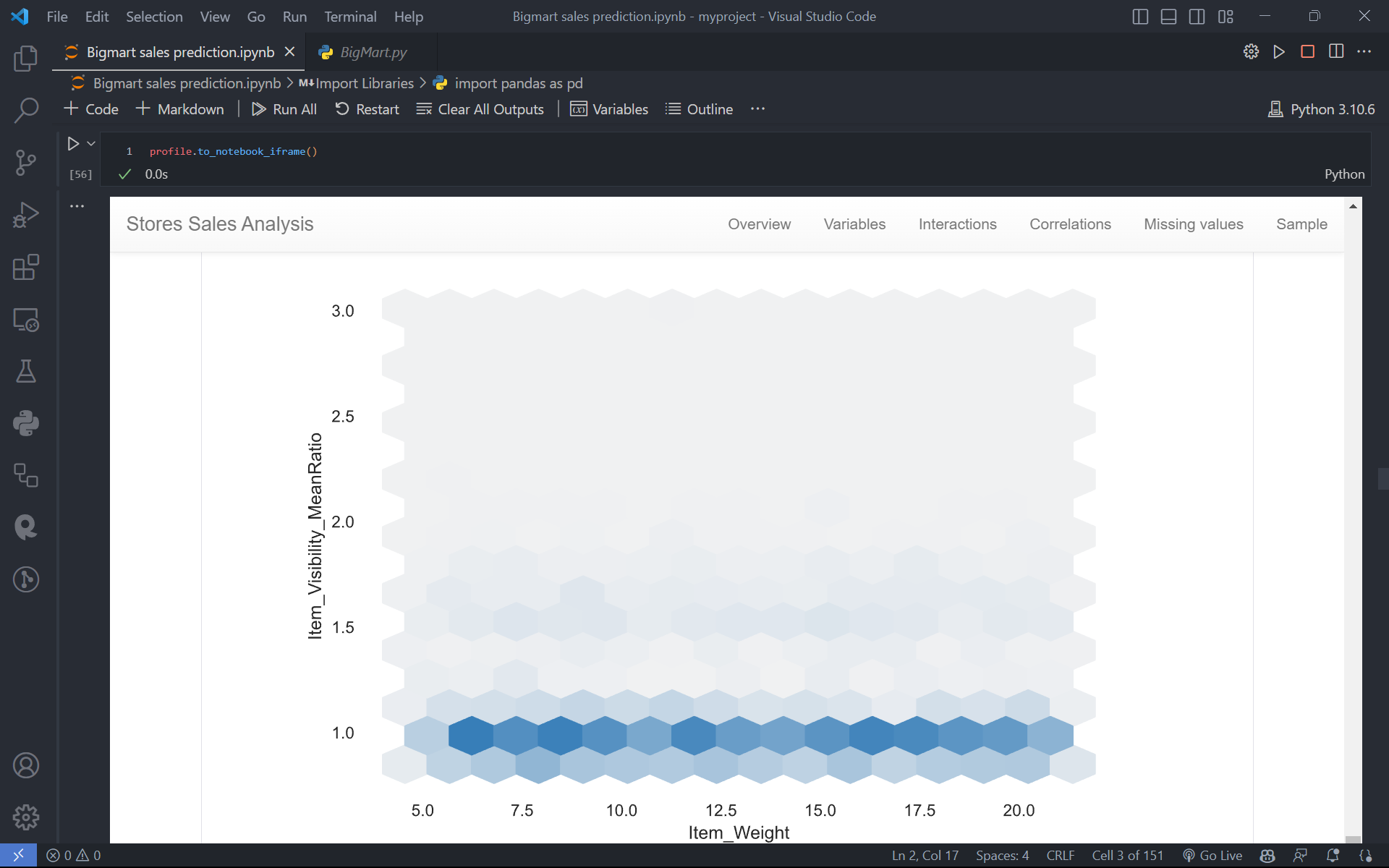
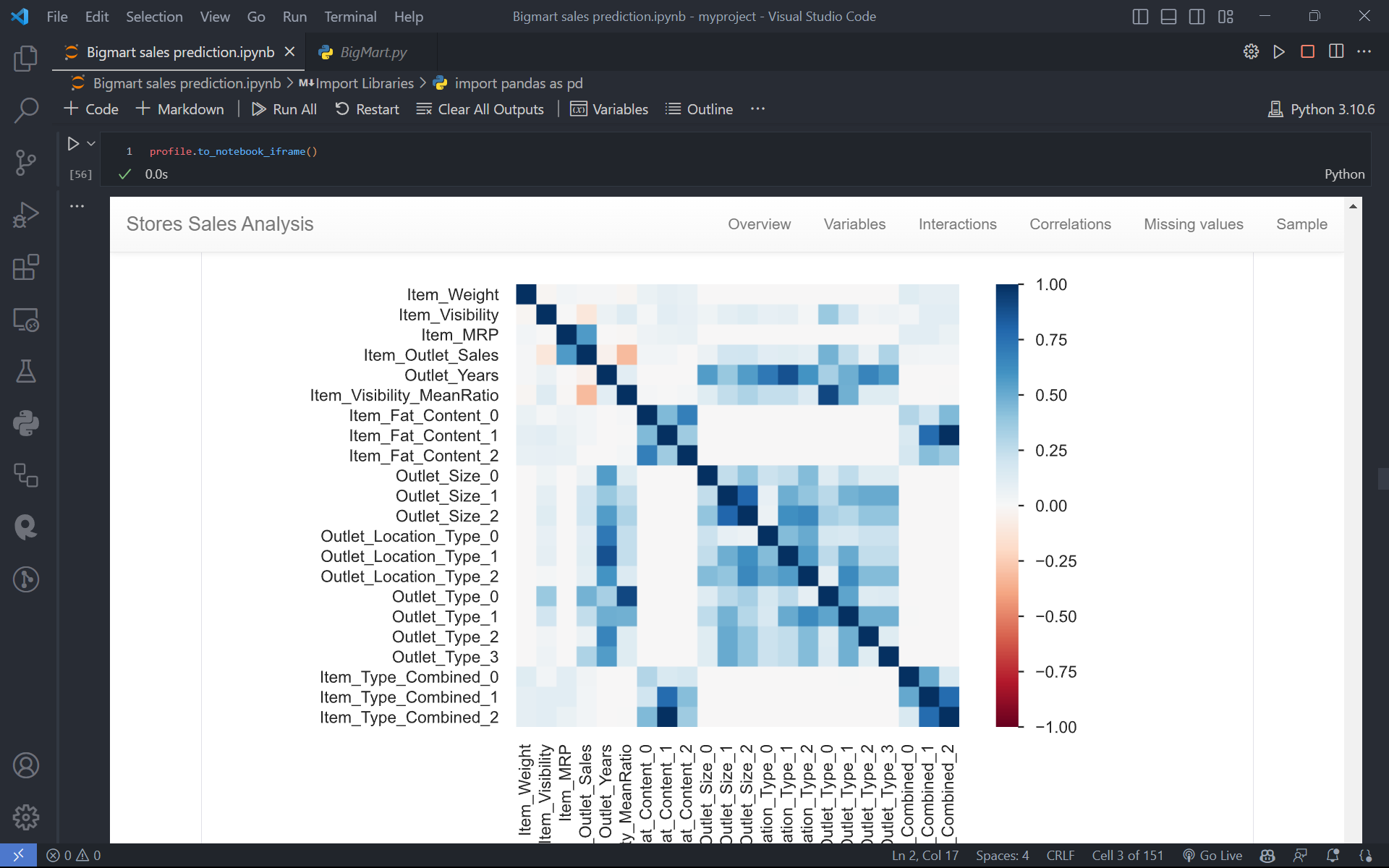


**Data Preprocessing**: Clean and preprocess the collected data to remove any inconsistencies, missing values, or outliers. This step may involve data imputation, normalization, feature engineering, and other techniques to make the data suitable for analysis.

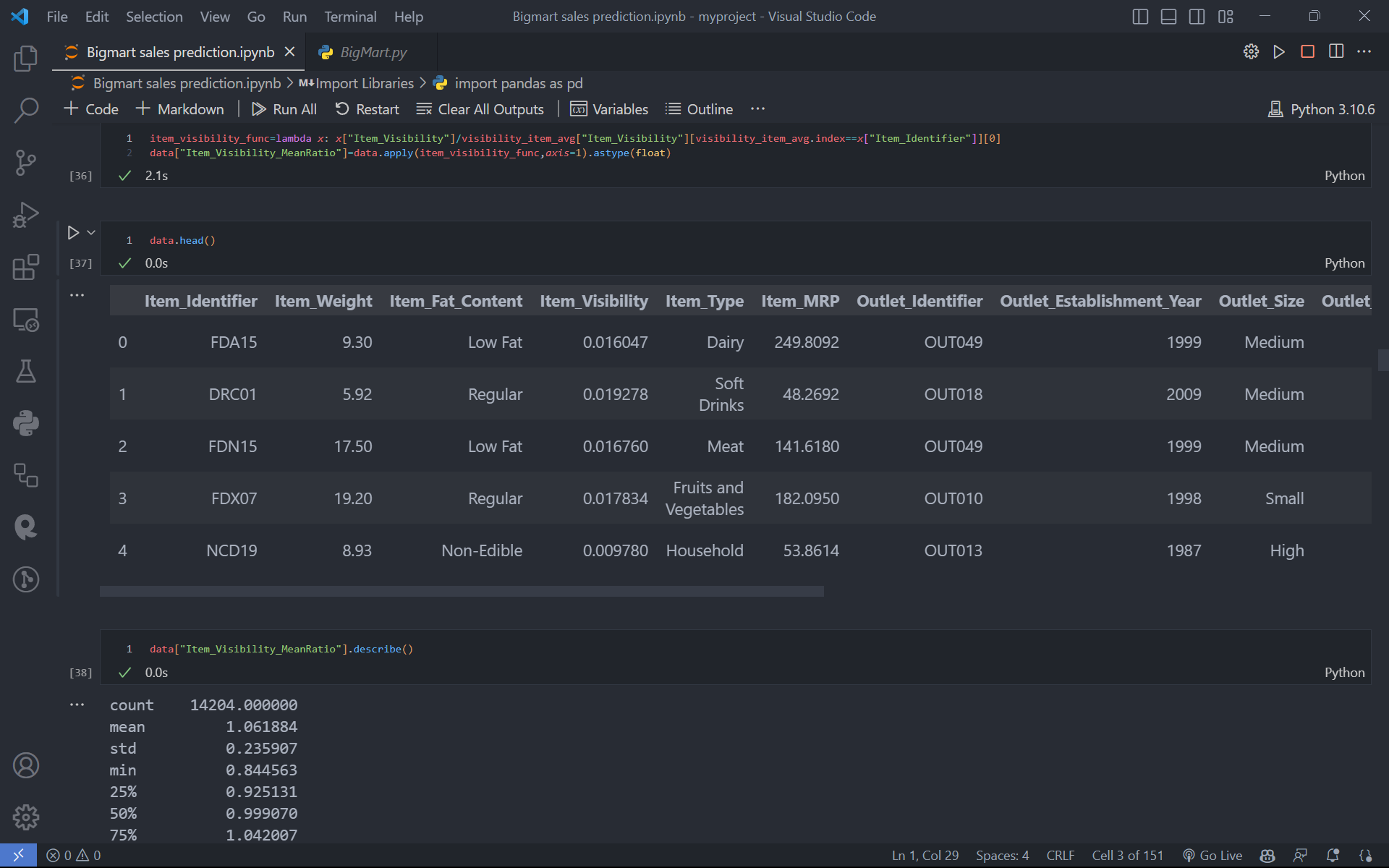




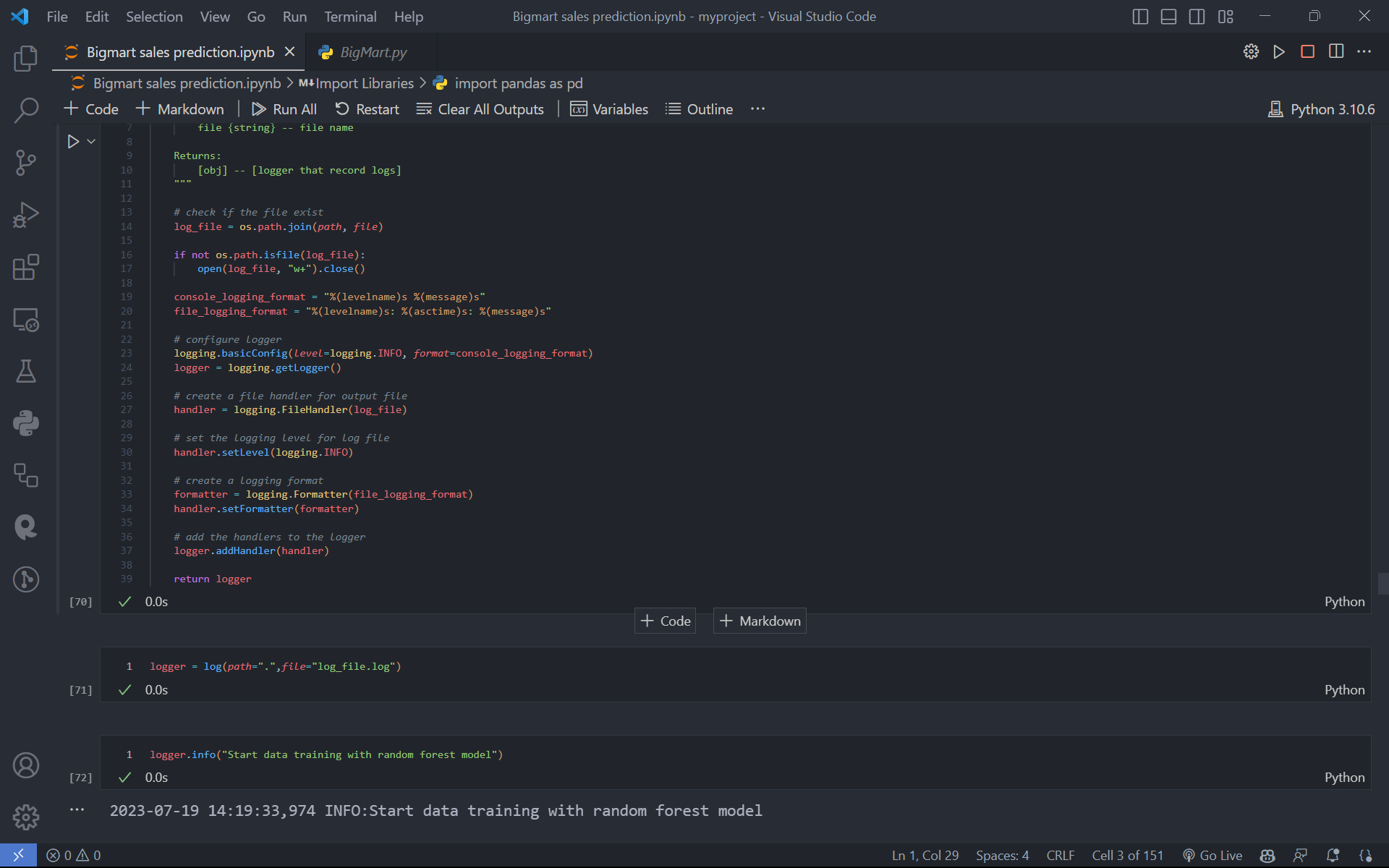
**Exploratory Data Analysis (EDA):** Perform an in-depth analysis of the preprocessed data to understand its characteristics, relationships between variables, and identify any patterns or trends. EDA can involve statistical analysis, data visualization, and hypothesis testing to gain insights into the data.

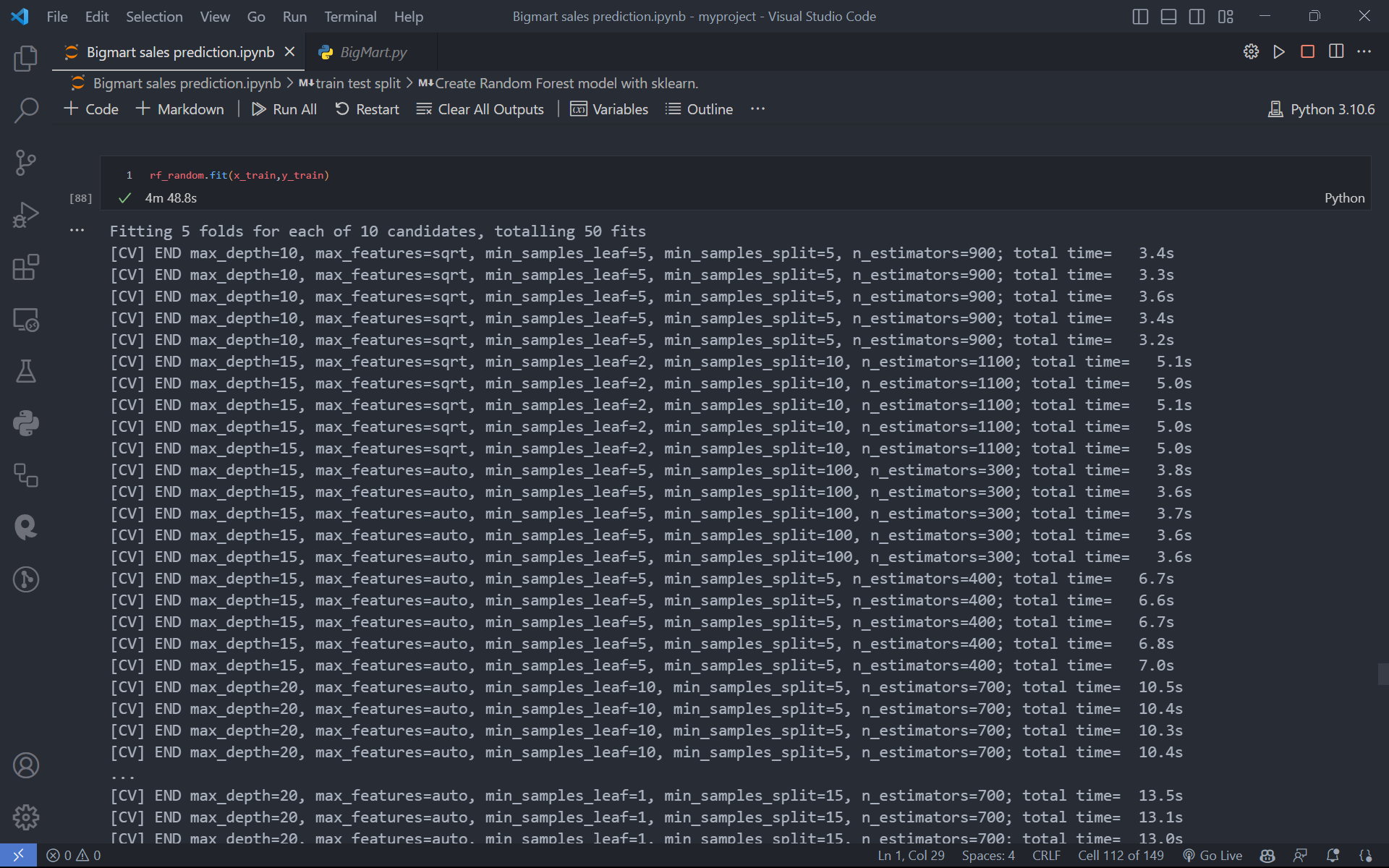
**Feature Selection**: Identify the most relevant features (variables) that have a significant impact on sales. This can be done using statistical methods, correlation analysis, or domain expertise. Selecting the right set of features helps improve the accuracy and efficiency of the sales prediction model.

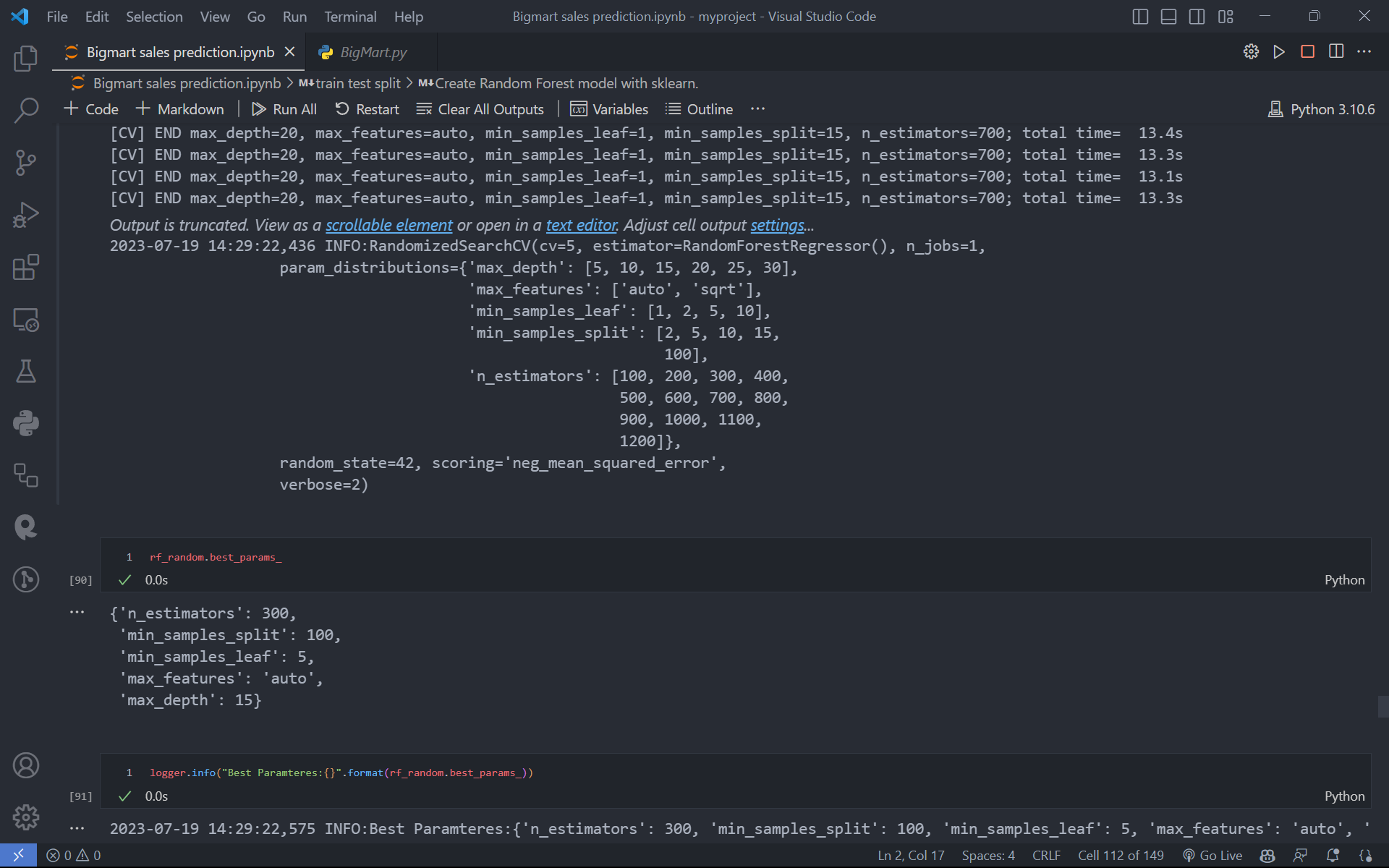


**Model Selection:** Choose an appropriate predictive modeling technique based on the nature of the data and the problem at hand. Commonly used methods for sales prediction include linear regression, decision trees, random forests, support vector machines (SVM), or more advanced techniques like neural networks.

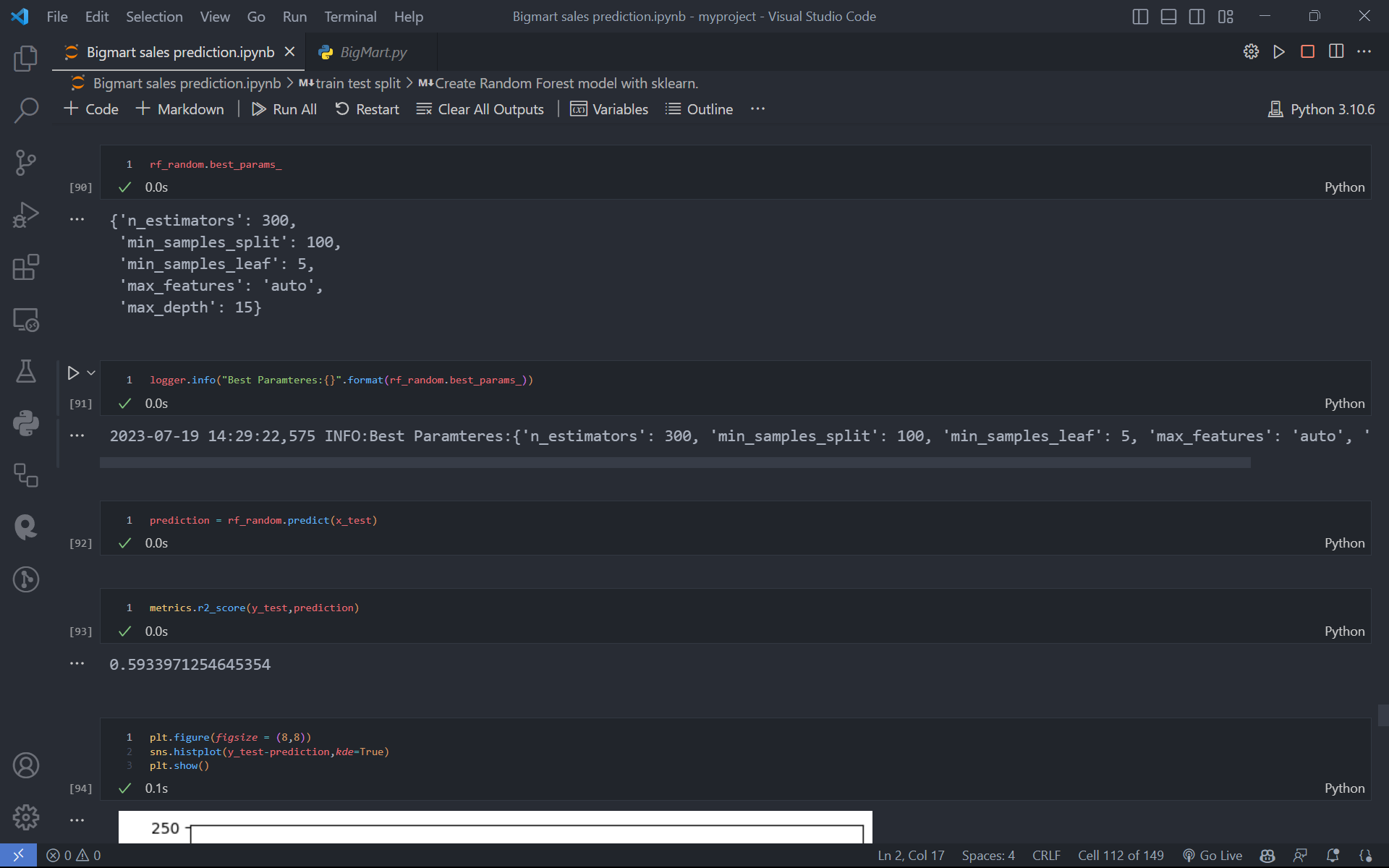


**Model Training:** Split the preprocessed data into training and validation sets. Use the training set to train the selected model by optimizing its parameters. This step involves fitting the model to the data, adjusting the model's weights or coefficients, and minimizing the prediction errors.

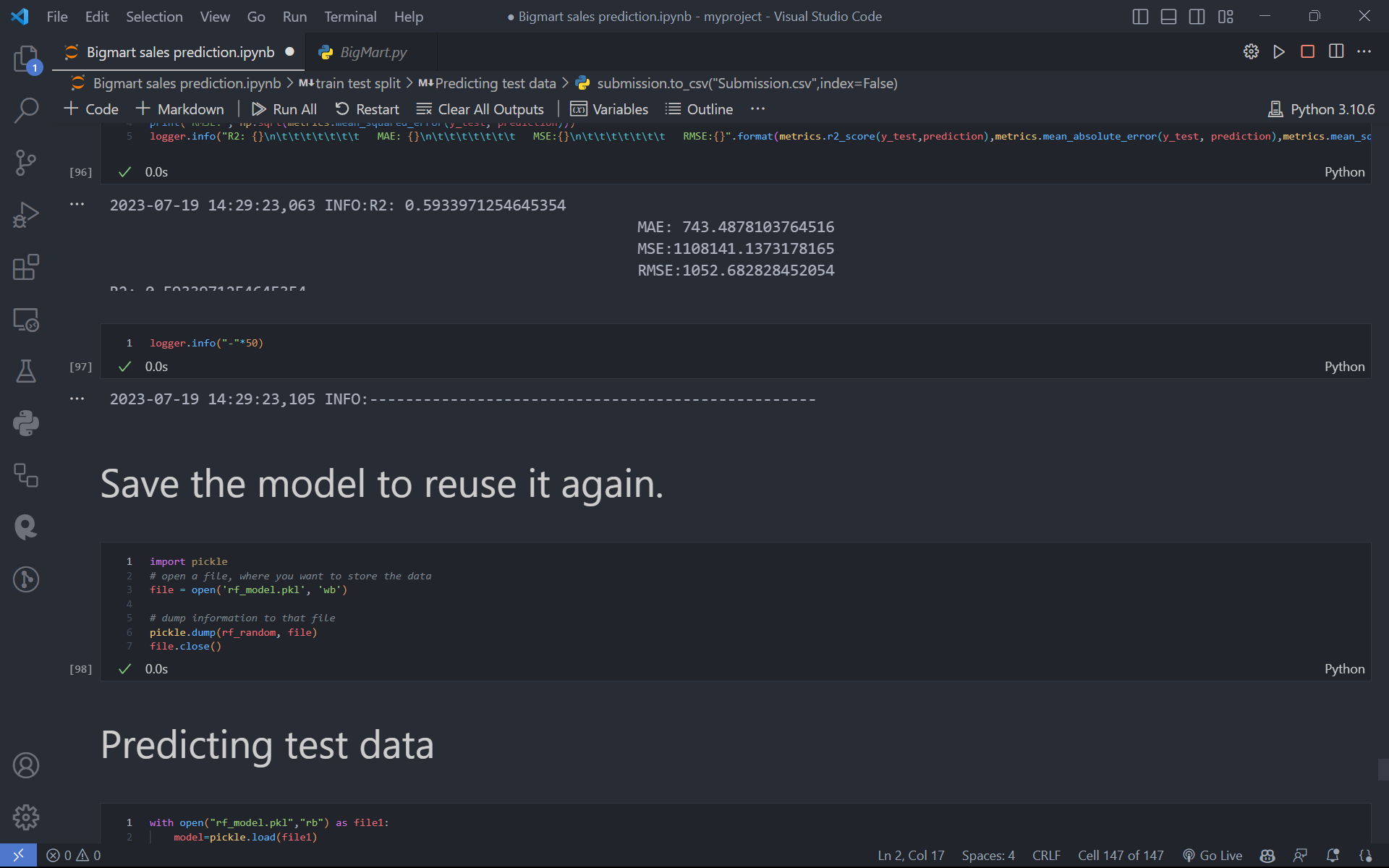


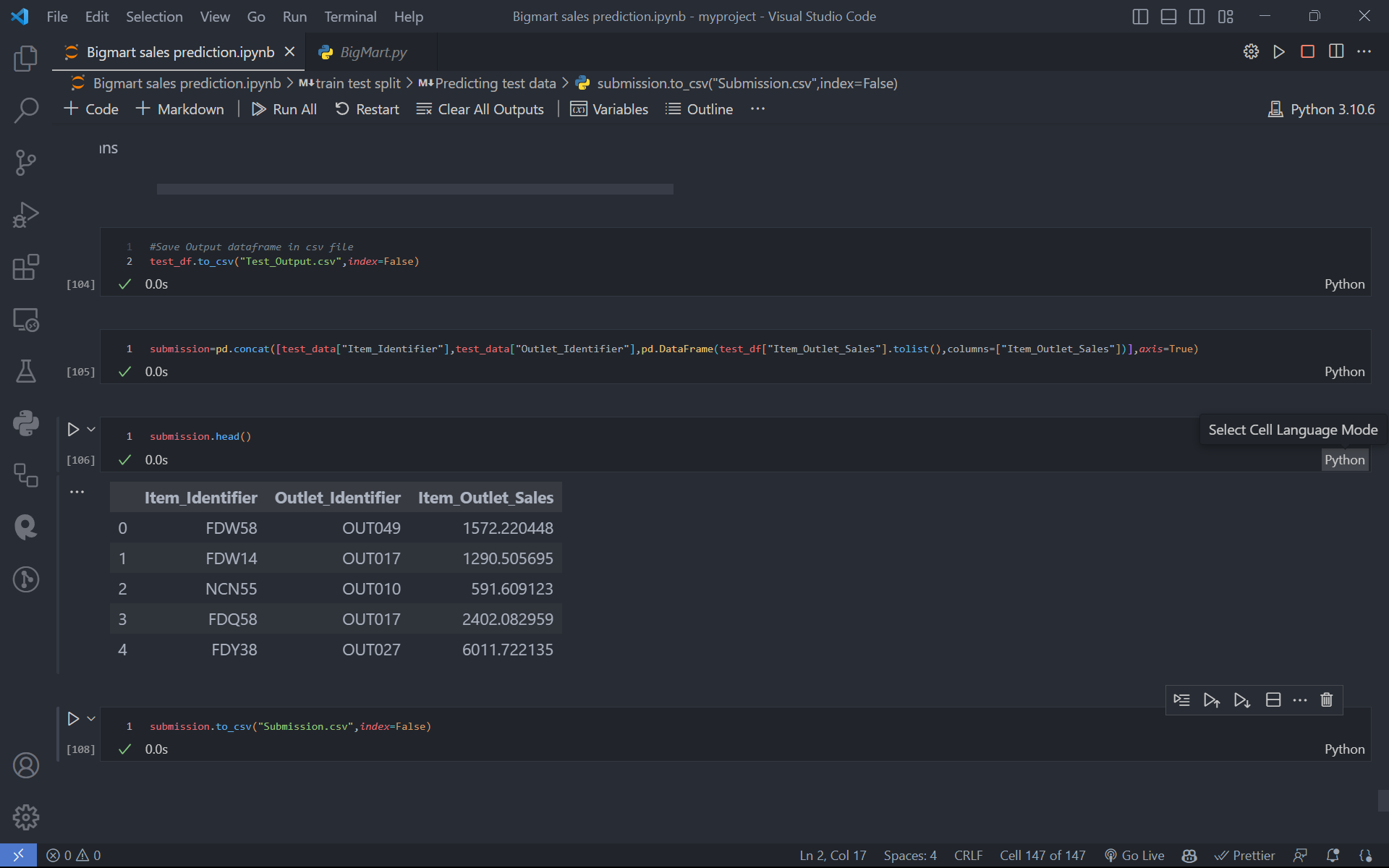


**Model Evaluation**: Evaluate the trained model's performance using the validation set. Common evaluation metrics for sales prediction include mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), or R-squared (coefficient of determination). Adjust the model or try different techniques if the performance is not satisfactory.



**Model Deployment**: Once the model is trained and evaluated, it can be deployed to make predictions on new, unseen data. This could involve integrating the model into a software application, creating an API, or using it in a batch processing system.





**Monitoring and Iteration:** Continuously monitor the performance of the deployed model and collect new data for future predictions. Periodically retrain the model with updated data to ensure its accuracy and relevance over time. Iteratively refine the model based on feedback and new insights.

It's important to note that the specific methodology may vary depending on the specific requirements and constraints of Big Mart's sales prediction project. Additionally, more advanced techniques such as time series analysis or ensemble methods can be applied if the sales data exhibits temporal patterns or complex relationships.

**Chapter 5**

**Conclusion and Future Work**

**Conclusion:**

So from this project we conclude that a smart sales forecasting program is required to manage vast volumes of knowledge for business organizations. The Algorithms which are presented in this report , LR, RF regression, Decision tree and XGBoost regression provide an effective method for data sharing as well as decision-making and also provide new approaches that are used for better identifying consumer needs and formulate marketing plans that are going to be implemented. The outcomes of ML algorithms which are done in this project will help us to pick the foremost suitable demand prediction algorithm and with the aid of which BigMart will prepare its marketing campaigns.

**Future Work:**

The future scope of this project is that this project can further collaborate with any other devices which are supported with an in-built intelligence by virtue of the Internet of Things(I0T) which makes it more feasible to use. Multiple instances parameters and various factors are also make this sales prediction project more innovative and successful. The most important term for any prediction-based system that is accuracy, is often significantly increased because of the increase in the number of parameters.

While the outlined methodology provides a solid foundation for sales prediction at Big Mart, there are several areas for potential future work and improvement:

Customer Segmentation: Perform customer segmentation analysis to identify different customer groups with distinct buying behaviors. This information can be utilized to tailor marketing strategies and promotions to specific customer segments, increasing the effectiveness of sales efforts.

A/B Testing: Implement A/B testing methodologies to evaluate the impact of different strategies or interventions on sales. This would provide valuable insights into the effectiveness of various business decisions and help optimize future actions.

Predictive Analytics Dashboard: Build an interactive and user-friendly dashboard that summarizes the sales predictions and provides visualizations and key performance indicators for decision-makers at Big Mart. This would enable them to monitor sales trends, evaluate model performance, and make data-driven decisions more effectively.

Continuing to refine the sales prediction methodology and incorporating these future work areas can help Big Mart stay competitive, adapt to changing market dynamics, and drive revenue growth by leveraging data-driven insights and predictions

**References**

Here are a few references that can provide you with further insights and information on sales prediction in the context of Big Mart or similar retail businesses:

[1] Bose, I., Mahapatra, R.K.: Business data mining ML perspective. Information & management 39(3) , (2001), pp. 211–225

[2] Punam, K., Pamula, R., Jain, P.K.: A two-level statistical model for big mart sales prediction. In: 2018 International Conference on Computing, Power and Communication Technologies (GUCON), IEEE (2018). pp. 617–620.

[3] Quinlan, J. R. (2014). C4. 5: programs for machine learning. Elsevier.

[4] MacKay, D. J., & Mac Kay, D. J. (2003). Information theory, inference and learning algorithms.

[5] Daumé III, H. (2012). A course in machine learning. Publisher, ciml. info, 5, 69.