**Summary**

**About Dataset**

**Sales Prediction for Big Mart Outlets**

The data scientists at BigMart have collected 2013 sales data for 1559 products across 10 stores in different cities. Also, certain attributes of each product and store have been defined. The aim is to build a predictive model and predict the sales of each product at a particular outlet.

Using this model, BigMart will try to understand the properties of products and outlets which play a key role in increasing sales.

Please note that the data may have missing values as some stores might not report all the data due to technical glitches. Hence, it will be required to treat them accordingly.

**Data Dictionary**

We have a train (8523) and test (5681) data set, the train data set has both input and output variable(s). You need to predict the sales for the test data set.

**Train file:**

**CSV** containing the item outlet information with a sales value

**Variable Description**

**Item\_Identifier** ---- Unique product ID  
**Item\_Weight** ---- Weight of product  
**Item\_Fat\_Content** ---- Whether the product is low fat or not  
**Item\_Visibility** ---- The % of the total display area of all products in a store allocated to the particular product  
**Item\_Type** ---- The category to which the product belongs  
**Item\_MRP** ---- Maximum Retail Price (list price) of the product  
**Outlet\_Identifier** ---- Unique store ID  
**Outlet\_Establishment\_Year** ---- The year in which the store was established  
**Outlet\_Size** ---- The size of the store in terms of ground area covered  
**Outlet\_Location\_Type** ---- The type of city in which the store is located  
**\*Outlet\_Type** ---- Whether the outlet is just a grocery store or some sort of supermarket  
**Item\_Outlet\_Sales** ---- sales of the product in t particular store. This is the outcome variable to be predicted.

**Test file:**

**CSV** containing item outlet combinations for which sales need to be forecasted

**Variable** **Description**  
**Item\_Identifier** ----- Unique product ID  
**Item\_Weight** ---- Weight of product  
**Item\_Fat\_Content** ----- Whether the product is low fat or not  
**Item\_Visibility** ---- The % of the total display area of all products in a store allocated to the particular product  
**Item\_Type** ---- The category to which the product belongs  
**Item\_MRP** ----- Maximum Retail Price (list price) of the product  
**Outlet\_Identifier** ----- Unique store ID  
**Outlet\_Establishment\_Year** ----- The year in which store store was established  
**Outlet\_Size** ----- The size of the store in terms of ground area covered  
**Outlet\_Location\_Type** ---- The type of city in which the store is located  
**Outlet\_Type** ---- whether the outlet is just a grocery store or some sort of supermarket

**Submission file format**

**Variable Description**

**Item\_Identifier** ----- Unique product ID  
**Outlet\_Identifier** ----- Unique store ID  
**Item\_Outlet\_Sales** ----- Sales of the product in t particular store. This is the outcome variable to be predicted.

**About the Ptoject**

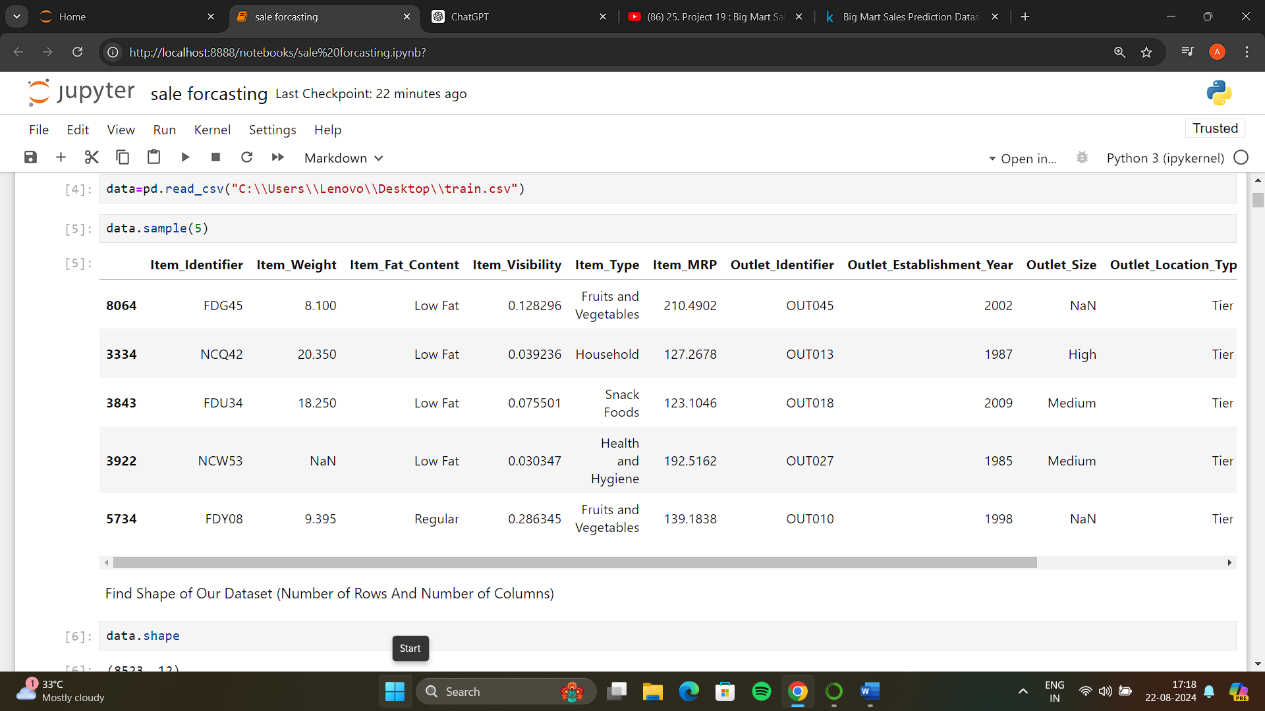
**Summary of Predicting Sales Forecasting with Big Mart Dataset**

1. **Dataset Overview:**

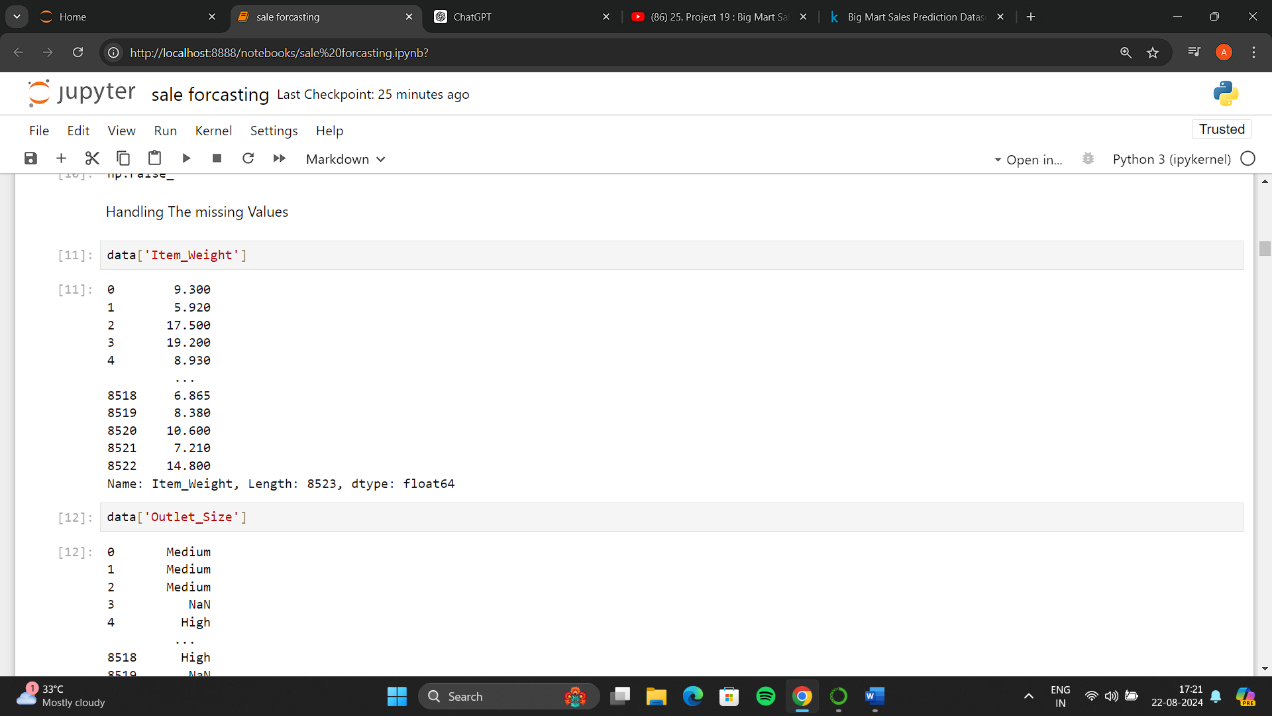
* **Big Mart Sales Dataset**: Contains data about sales of products from various Big Mart stores.
* **Features**: Typically includes attributes such as Item\_Identifier, Item\_Weight, Item\_Fat\_Content, Item\_Visibility, Item\_Type, Item\_MRP, Outlet\_Identifier, Outlet\_Establishment\_Year, Outlet\_Size, Outlet\_Location\_Type, and Outlet\_Type.
* **Target Variable**: Item\_Outlet\_Sales (the sales amount).

2. **Data Preprocessing:**

* **Loading Data**: Read the dataset using pandas.

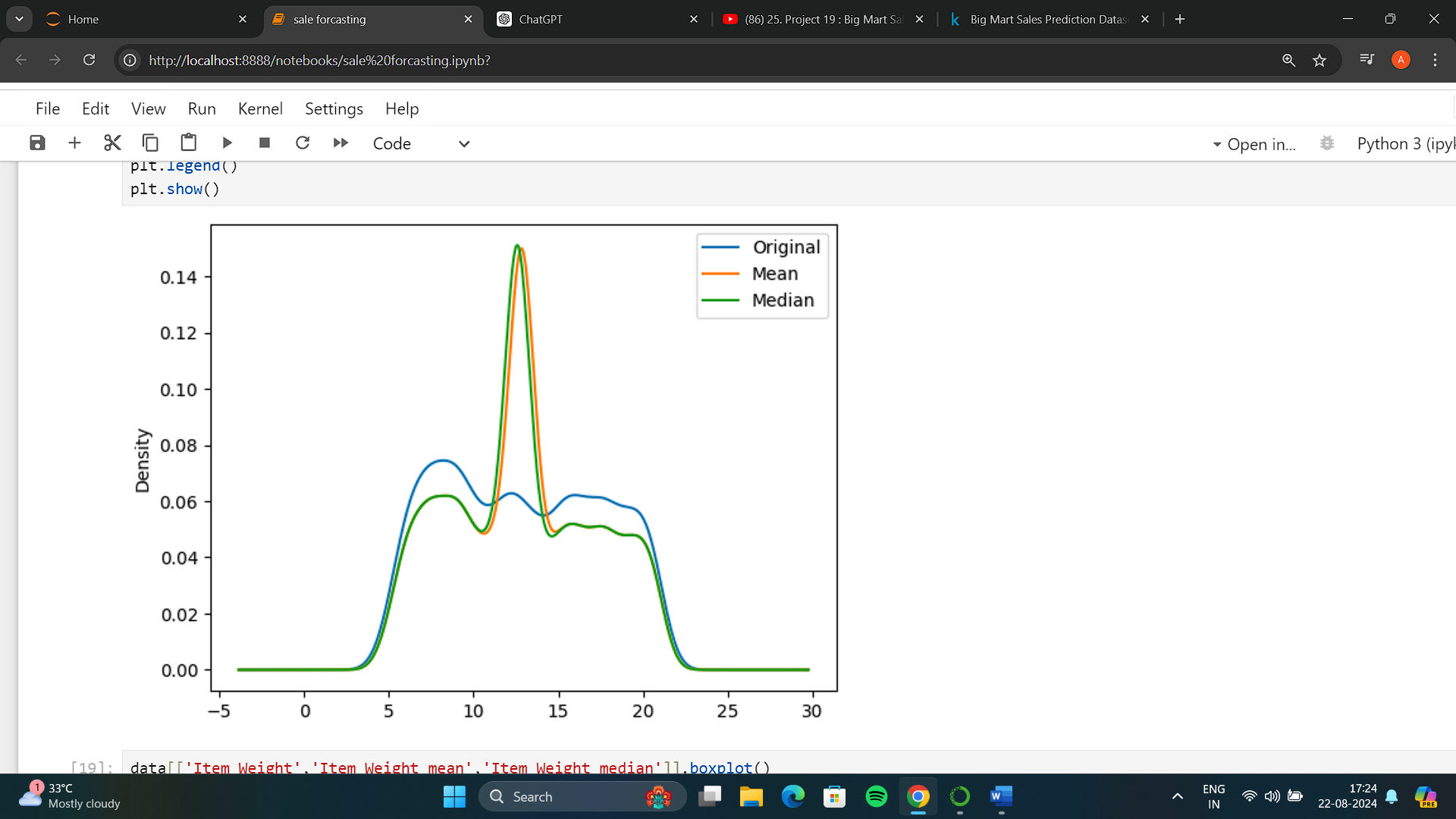


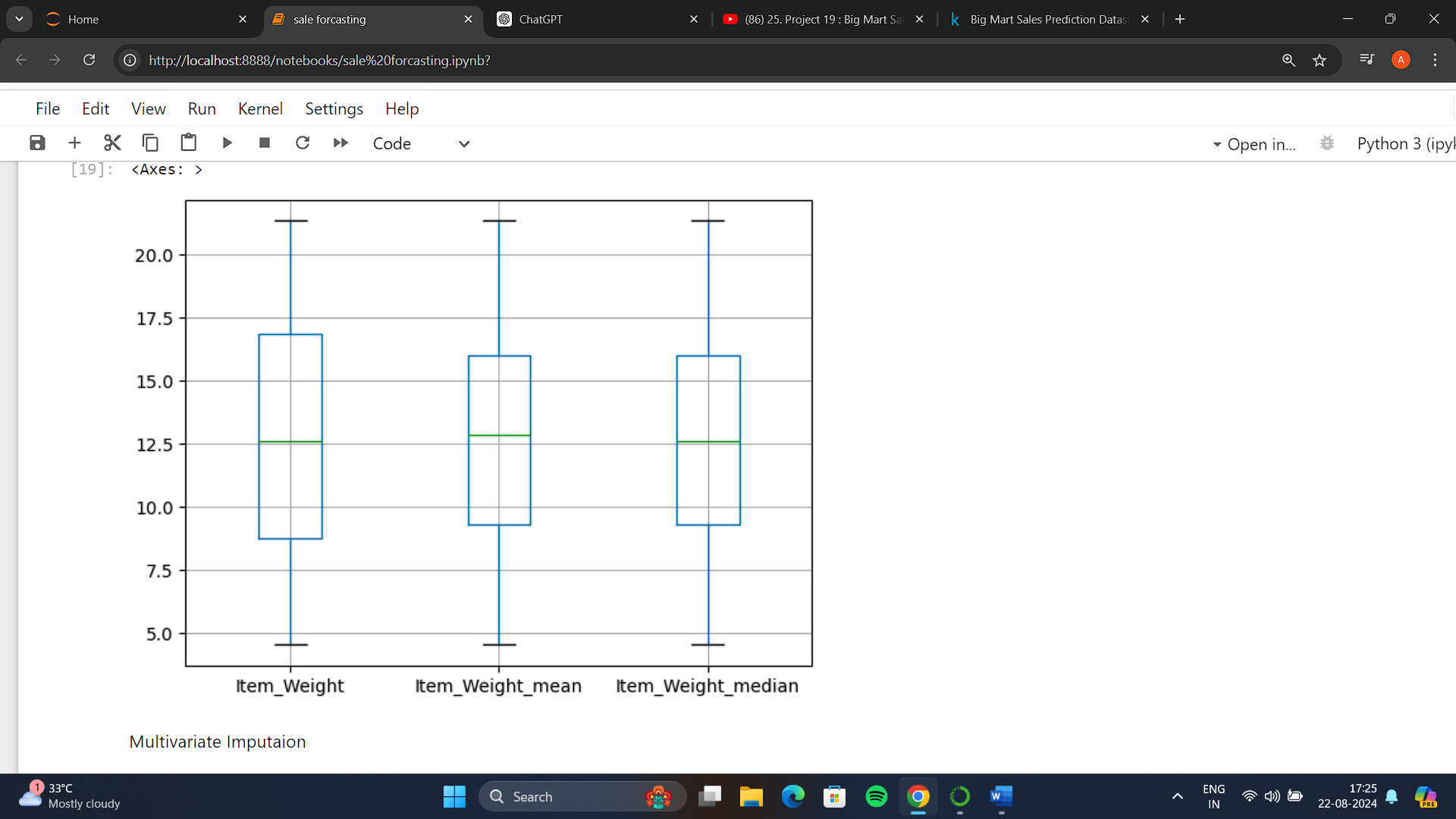
**Handling Missing Values**: Fill or impute missing values as appropriate



3. **Exploratory Data Analysis (EDA):**

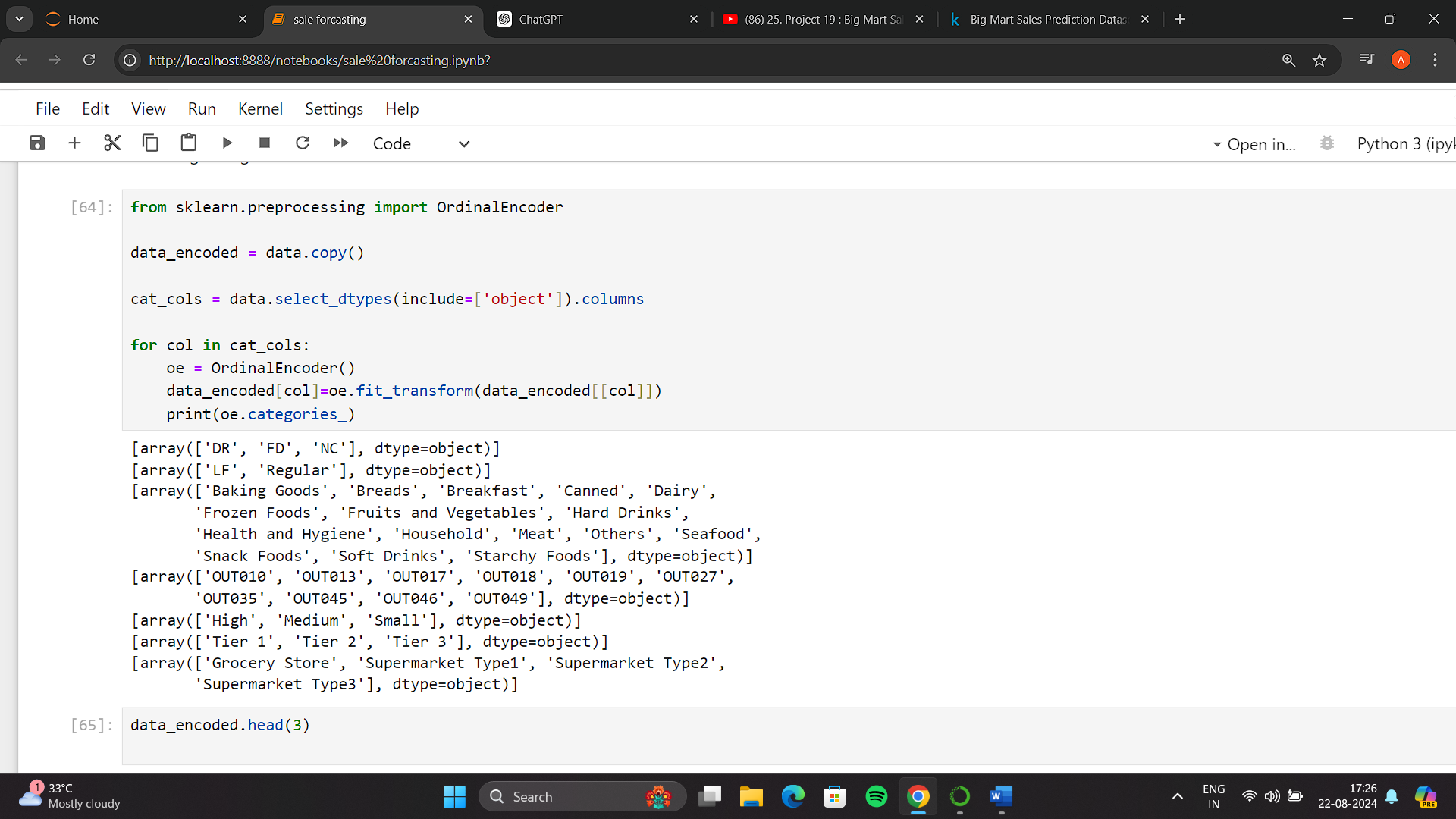
* **Understand Data Distribution**: Use visualization to understand the distribution of features and target variable.





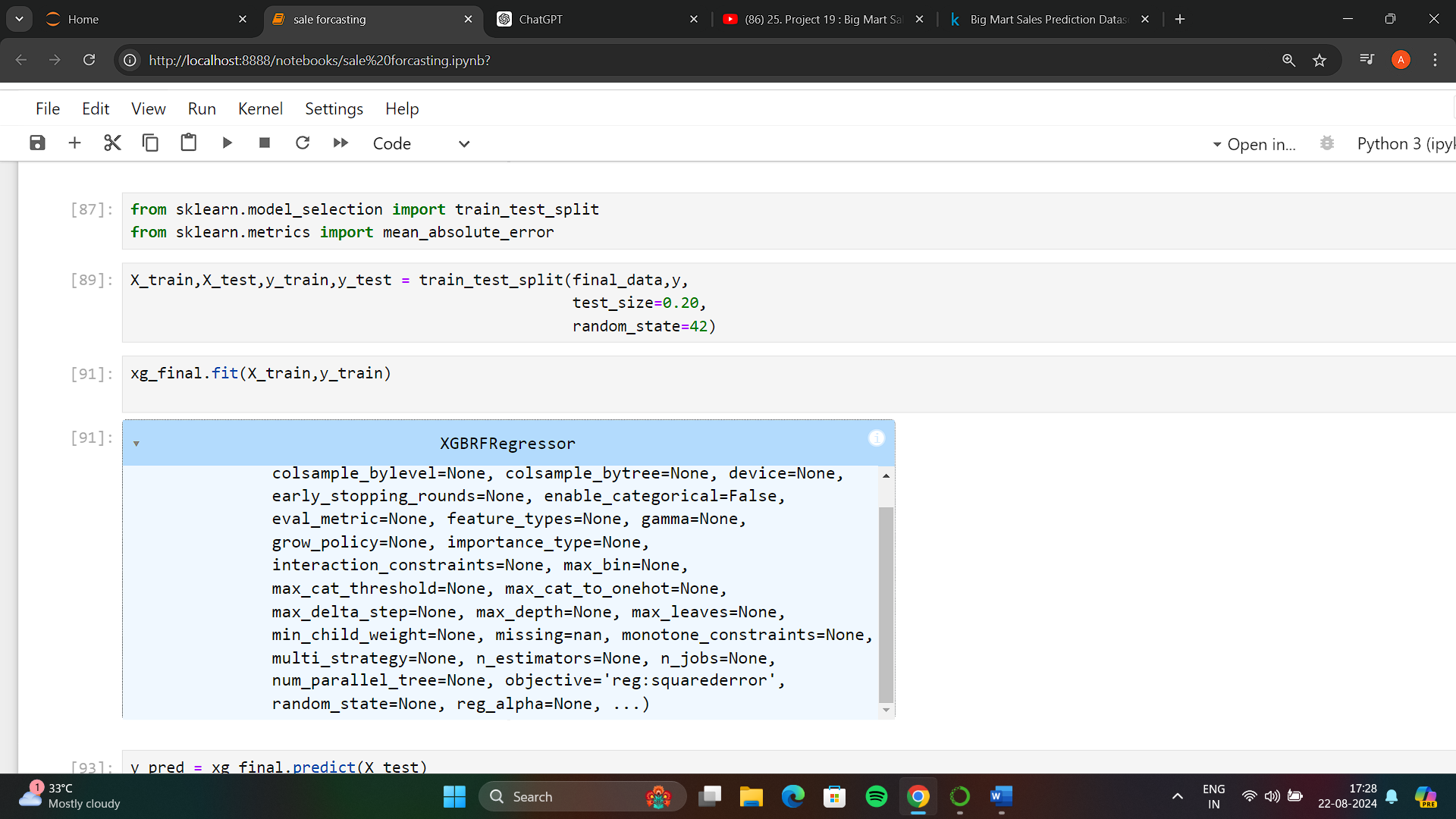
**Feature Encoding:**

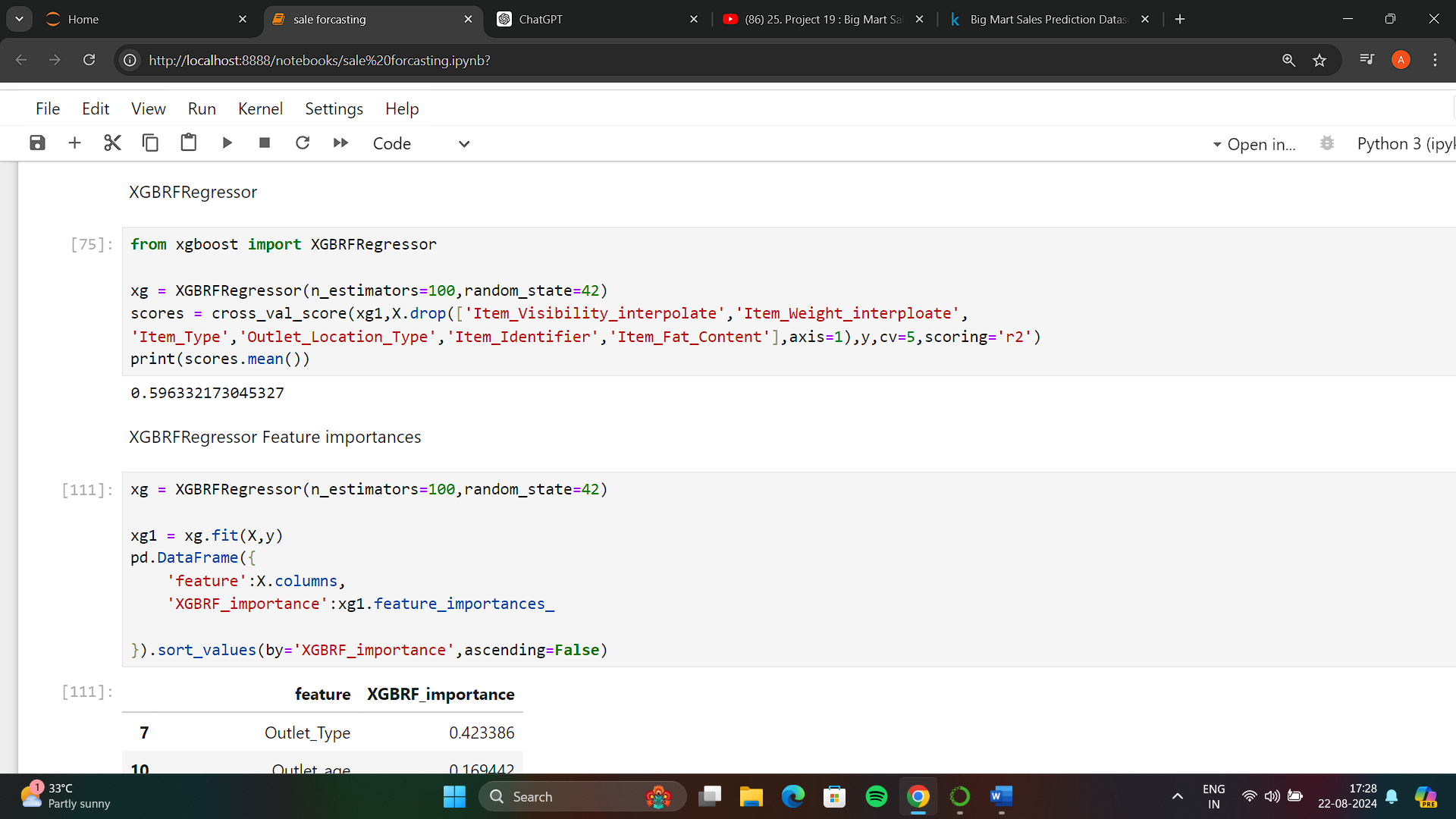
* **Categorical Encoding**: Convert categorical variables into numerical form using one-hot encoding or label encoding.



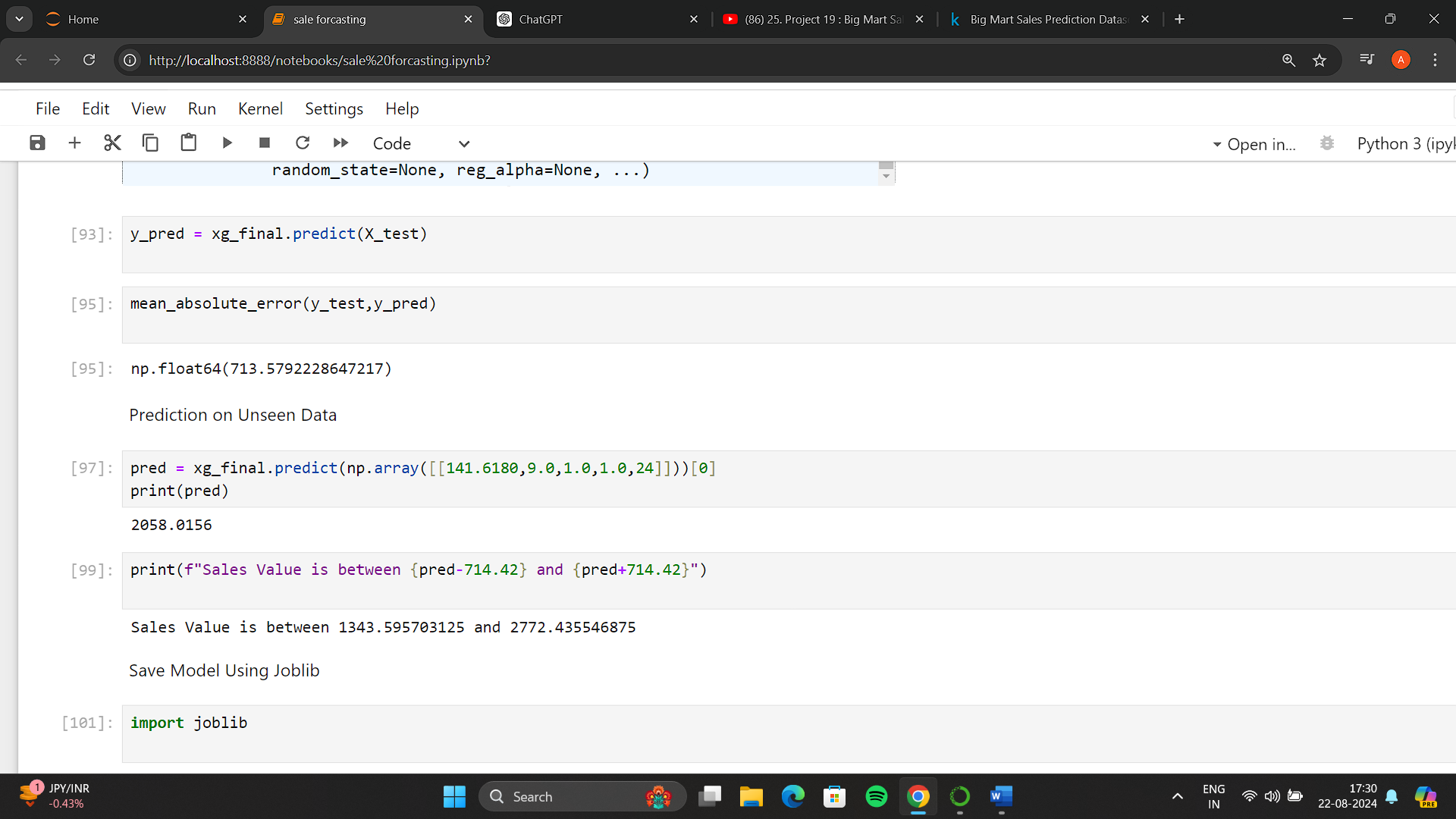
**Model Building:**

* **Split Data**: Divide the data into training and testing sets



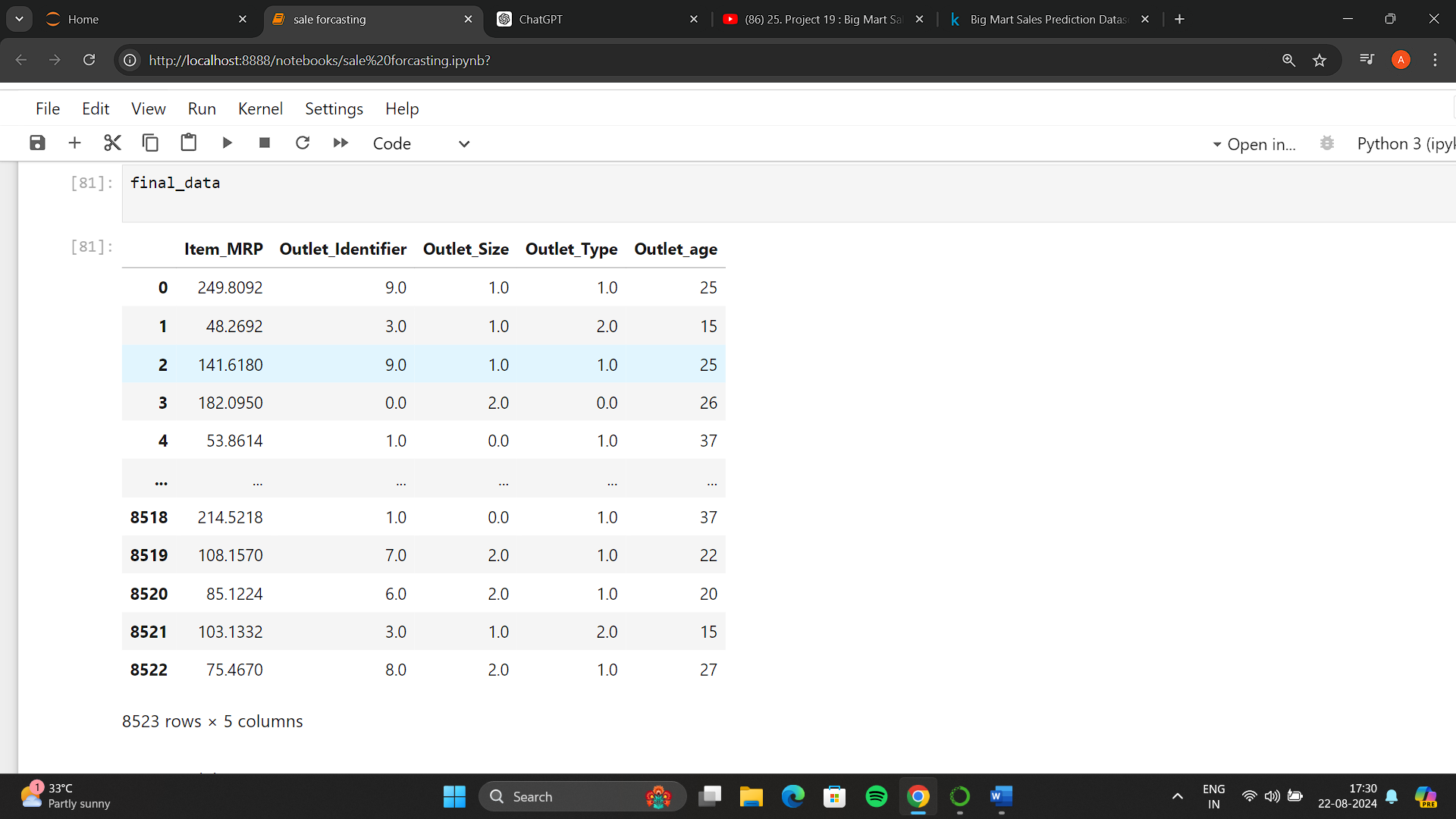
**Choose a Model**: Common models for regression include Linear Regression, Decision Trees, Random Forest, and Gradient Boosting. 

**Evaluate the Model**: Assess the model performance using metrics like Mean Absolute Error (MAE) and Mean Squared Error (MSE).



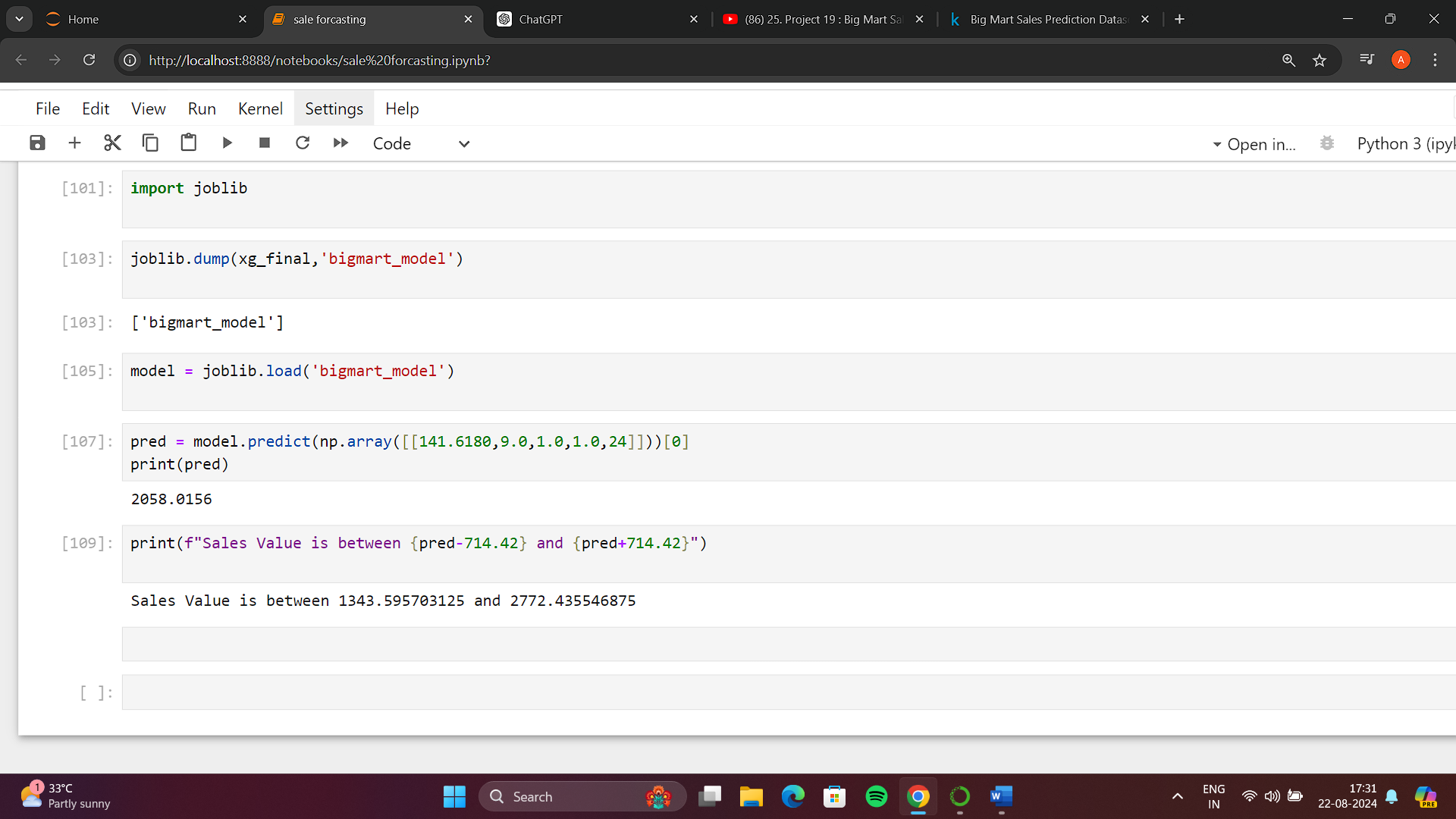
**Final Model and Predictions:**

* **Train Final Model**: Train the final model with the best hyperparameters.
* **Make Predictions**: Predict sales on new or unseen data.



**Deployment:**

* **Deploy Model**: Save the model using joblib or pickle for future use



**Conclusion:** The process of predicting sales forecasting using the Big Mart dataset involves a series of steps from data preprocessing to model evaluation. By properly preparing the data, selecting suitable features, and choosing the right machine learning algorithm, you can build an effective model to forecast sales and make informed business decisions.