

Low Level Design

**Black Friday Sales Prediction**

|  |  |
| --- | --- |
| **Written By** | **Aditya Gautam** |
| **Document Version** | **3.0** |
| **Last Revised Date** | **04-Dec-2022** |



**DOCUMENT CONTROL**

## Change Record:

|  |  |  |  |
| --- | --- | --- | --- |
| **VERSION** | **DATE** | **AUTHOR** | **COMMENTS** |
| 1.0 | 22-Nov-2022 | Aditya Gautam | Introduction and architecture defined |
| 2.0 | 30 -Nov-2022 | Aditya Gautam | Architecture & Architecture description appended and updated. |
| 3.0 | 04-Dec- 2022 | Aditya  Gautam | Deployement and Conclusion appended and updated |
|  |  |  |  |

**Reviews:**

|  |  |  |  |
| --- | --- | --- | --- |
| **VERSION** | **DATE** | **REVIEWER** | **COMMENTS** |
|  |  |  |  |

**Approval Status:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **VERSION** | **REVIEW**  **DATE** | **REVIEWED BY** |  | **APPROVED BY** | **COMMENTS** |
|  |  |  |  |  |  |



**Contents**

1. Introduction 1
   1. What is Architecture design document?
   2. Scope

1. Architecture
   1. Architecture of the Black Friday Sales Prediction 2-8
   2. Life Cycle of Machine Learning Project
   3. Detailed Architecture of Black Friday Analysis
   4. Jupyter Notebook Architecture
   5. Power BI Architecture
2. Architecture Description 9-16
   1. Data Description
   2. Importing the Libraries in Jupyter Notebook
   3. Importing the Dataset in Jupyter Notebook
   4. Exploratory Data Analysis
   5. Feature Engineering

3.6 Split into Train and Test Data

3.7 Feature Scaling

3.8 Hyper parameter Tuning

3.9 Model Training

3.10 Model Prediction and Evaluation

3.11 Submission of the Predicted Values

4. Deployment 17-21

4.1 Load the Dataset in Power BI

4.2 Create the interactive dashboard

4.3 Publish to Power Bi account

4.4 Publish to Web

4.5 Create the embedded code

4.6 Share the Public Link to Client

**5. Conclusion**

**5.1 Insight of the Black Friday Sales Prediction 22**

1. **Introduction**

* 1. **What is Low Level document?**

The Low level document is to give the internal logic design of the code for the Black Friday Sales Prediction dashboard. LDD describes the class diagrams with the methods and relations between classes and programs specification. It describes the modules so that the person can directly code the program from the document.

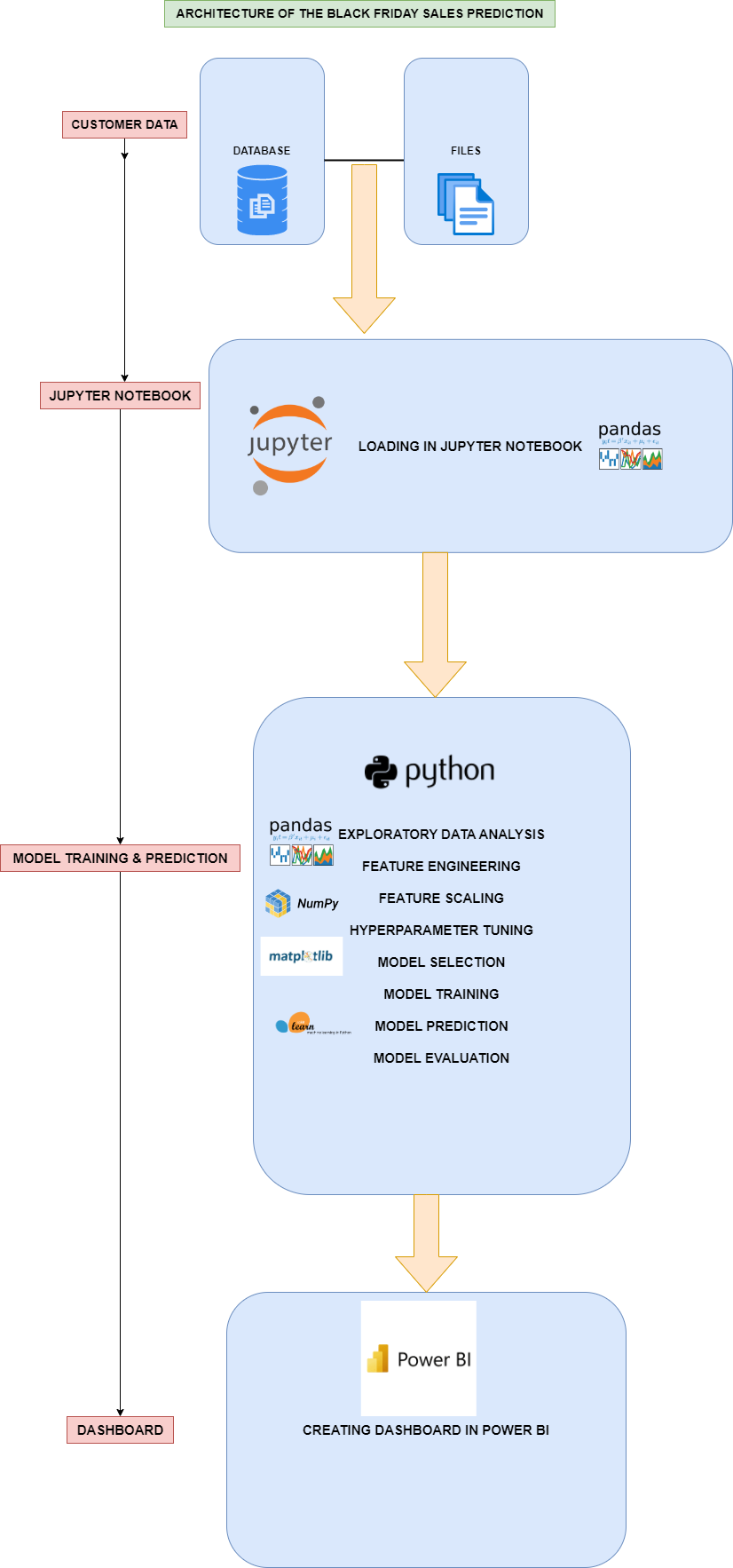
It gives the person a complete overview about the whole project, which helps to understand the step by step process how to proceed the project and complete it.

* 1. **Scope**

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

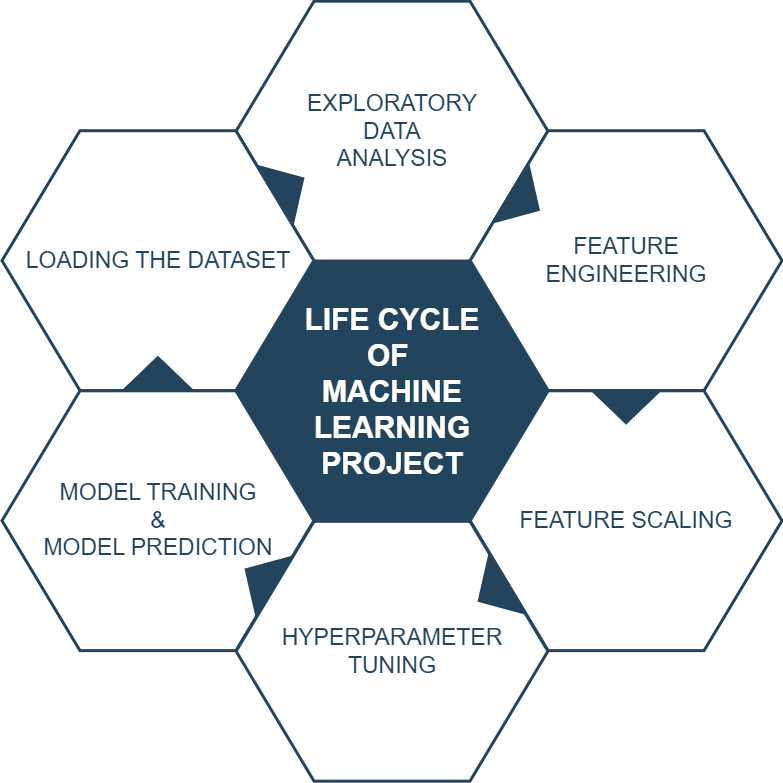
2. **Architecture**

2.1 **Architecture of the Black Friday Sales Prediction**

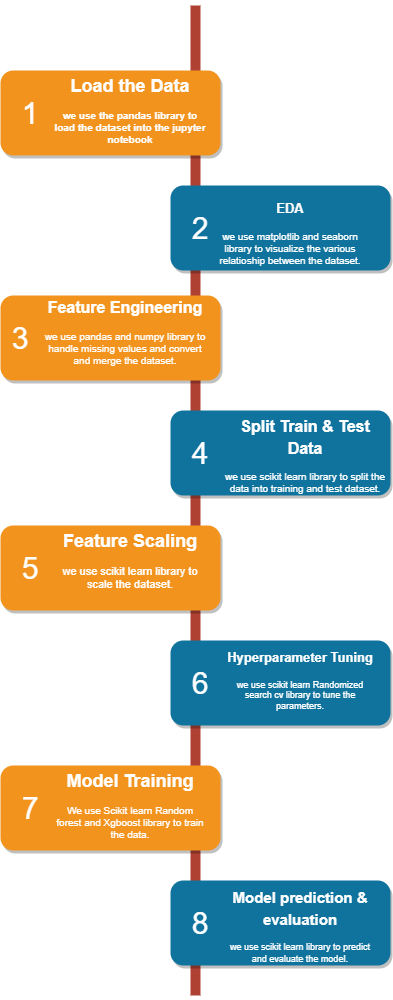


The detailed architecture of the Black Friday Sales Analysis has been discussed in the above architecture diagram which gives a overview of the step by step process of the project which gives an idea about flow of the data from original sources to database, then exporting the data from database to importing the data into jupyter notebook by using pandas library for data cleaning process, then for visualize the data, visualization library such Matplotlib and seaborn is used for the purpose and pandas library is used for Feature engineering. Then scikit learn library is used for feature selection , model training, hyperparameter tuning and model evaluation of the data. And finally, deploying the trained data into Power Bi for creating an interactive dashboard.

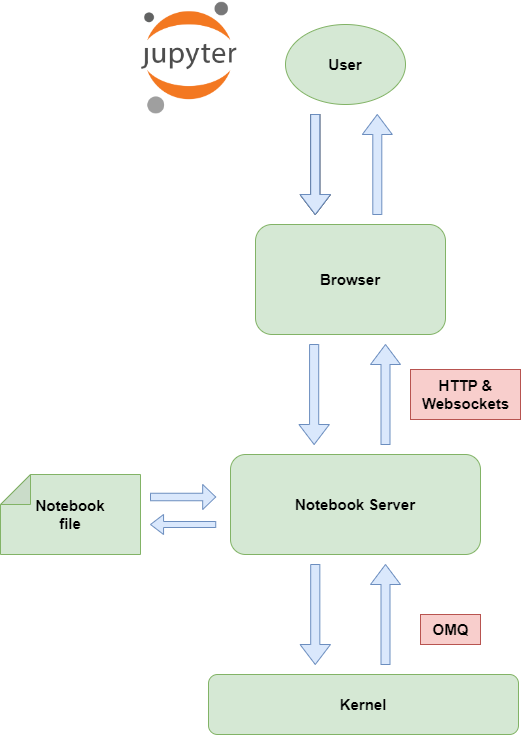
2.2 **Life Cycle of Machine Learning Project**



2.3 **Detailed Architecture of Black Friday Analysis**



**2.4 Jupyter Notebook Architecture**



Jupyter Notebook and its flexible interface extends the notebook beyond code to visualization, Multimedia, collaboration, and more. In addition to running your code and output, together with markdown notes, in an editable document called. When you save it, this is sent your browser to the notebook server, which saves it on disk as JSON file with a .ipynb extension.

The notebook server is a communication hub. The browser, notebook file on disk, and kernel cannot talk to each other directly. They communicate through the notebook server. The notebook server, not the kernel, is responsible for saving and loading notebooks, so you can edit notebooks even if you don’t have the kernel for that language-- you just won’t be able to run code. The kernel doesn’t know anything about the notebook document, it just gets sent cells of code to execute when the user runs them.

2.5 **Power BI Architecture**



**Power BI**

Power BI is a business analytics solution that lets you visualize your data and share insights across your organization, or embed them in your app or website. Connect to hundreds of data sources and bring your data to life with live dashboards and reports.

2.5.1. **Data sources**

The most important component of Power BI is its vast range of data sources.

 You can import data from files in your system, cloud-based online data sources or connect directly to live connections. The some of the important data sources are given below:

* 1. Excel
  2. Text/CSV
  3. JSON
  4. XML
  5. MySQL Database
  6. PostgreSQL Database
  7. Snowflake

#### 2.5.2. Power BI Desktop

Power BI Desktop is a client-side tool known as a companion development and authoring tool.

Power BI Desktop has tools and functionalities to connect the data sources, transform the data, data modeling and creating interactive dashboard.

Using Power BI Desktop features, one can dodatacleansing, create business metrics and data models, define the relationship between data, define hierarchies, create visuals and publish reports.

1. Connecting to your data, Access data from hundreds of supported on-premises and cloud-based sources, such as Dynamics 365, Salesforce, Azure SQL DB, Excel, and SharePoint. Ensure it’s always up to date with automated, incremental refreshes. Power BI Desktop enables you to develop deep, actionable insights for a broad range of scenarios.
2. Preparing and modeling your data with ease, Data prep can take most of your time, but it doesn’t have to with data modeling tools. Reclaim hours in your day using the self-service Power Query experience familiar to millions of Excel users. Ingest, transform, integrate and enrich data in Power BI.
3. Provide advanced analytics with the familiarity of Excel, Dig deeper into data and find patterns you may have otherwise missed that lead to actionable insights. Use features like quick measures, grouping, forecasting, and clustering. Give advanced users full control over their model using powerful DAX formula language. If you’re familiar with Excel, you’ll feel at home in Power BI.
4. Create stunning reports with interactive data visualizations. Tell your data story using a drag-and-drop canvas and hundreds of modern data visuals from Microsoft and partners—or create your own, using the Power BI open source custom visuals framework. Design your report with theming, formatting, and layout tools.

#### 2.5.3. Power BI Service

Power BI Service is a web-based platform from where you can share reports made on Power BI Desktop, collaborate with other users, and create dashboards.

It is available in three versions:

* Free version
* Pro version
* Premium version

#### 2.5.4. Power BI Report Server

The Power BI Report Server is similar to the Power BI Service. The only difference between these two is that Power BI Report Server is an on-premise platform. It is used by organizations who do not want to publish their reports on the cloud and are concerned about the security of their data.

Power BI Report Server enables you to create dashboards and share your reports with other users following proper security protocols. To use this service, you need to have a Power BI Premium license.

#### 2.5.5. Power BI Gateway

This component is used to connect and access on-premise data in secured networks. Power BI Gateways are generally used in organizations where data is kept in security and watch. Gateways help to extract out such data through secure channels to Power BI platforms for analysis and reporting.

#### 2.5.6. Power BI Mobile

Power BI Mobile is a native Power BI application that runs on iOS, Android, and Windows mobile devices. For viewing reports and dashboards, these applications are used.

#### 2.5.7. Power BI Embedded

Power BI Embedded offers APIs which are used to embed visuals into custom applications.

3. **Architecture Description**

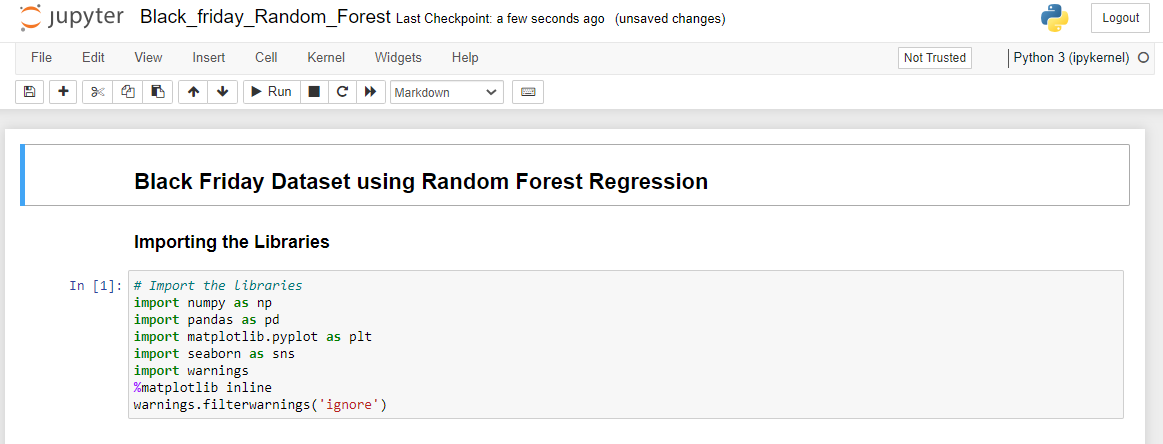
3.1. **Data Description**

Dataset has 537577 rows (transactions) and 12 columns (features) as described below:

* User\_ID: Unique ID of the user. There are a total of 5891 users in the dataset.
* Product\_ID: Unique ID of the product. There are a total of 3623 products in the dataset.
* Gender: indicates the gender of the person making the transaction.
* Age: indicates the age group of the person making the transaction.
* Occupation: shows the occupation of the user, already labeled with numbers 0 to 20.
* City\_Category: User's living city category. Cities are categorized into 3 different categories 'A', 'B' and 'C'.
* Stay\_In\_Current\_City\_Years: Indicates how long the users has lived in this city.
* Marital\_Status: is 0 if the user is not married and 1 otherwise.
* Product\_Category\_1 to \_3: Category of the product. All 3 are already labaled with numbers.
* Purchase: Purchase amount.

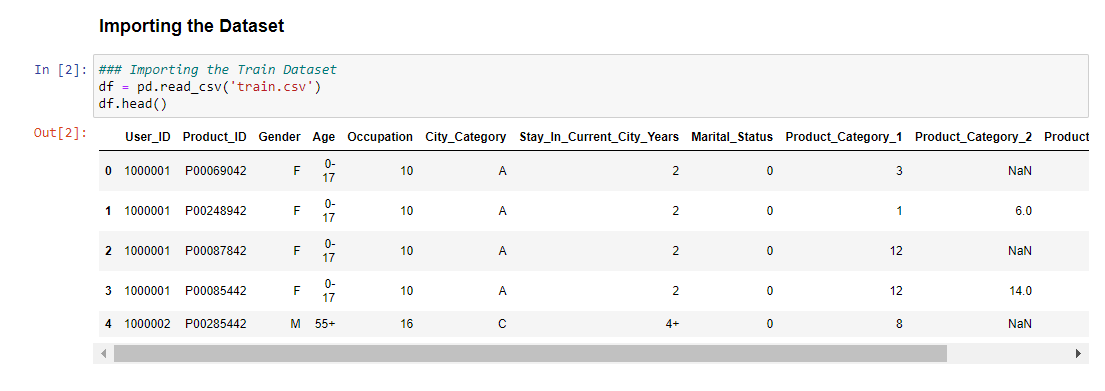
3.2 **Importing the Libraries in Jupyter Notebook**

Importing the Libraries are the most important and intial feature in jupyter notebook, where the libraries such as pandas, numpy, matplotlib , seaborn, Sklearn etc is imported in the notebook.



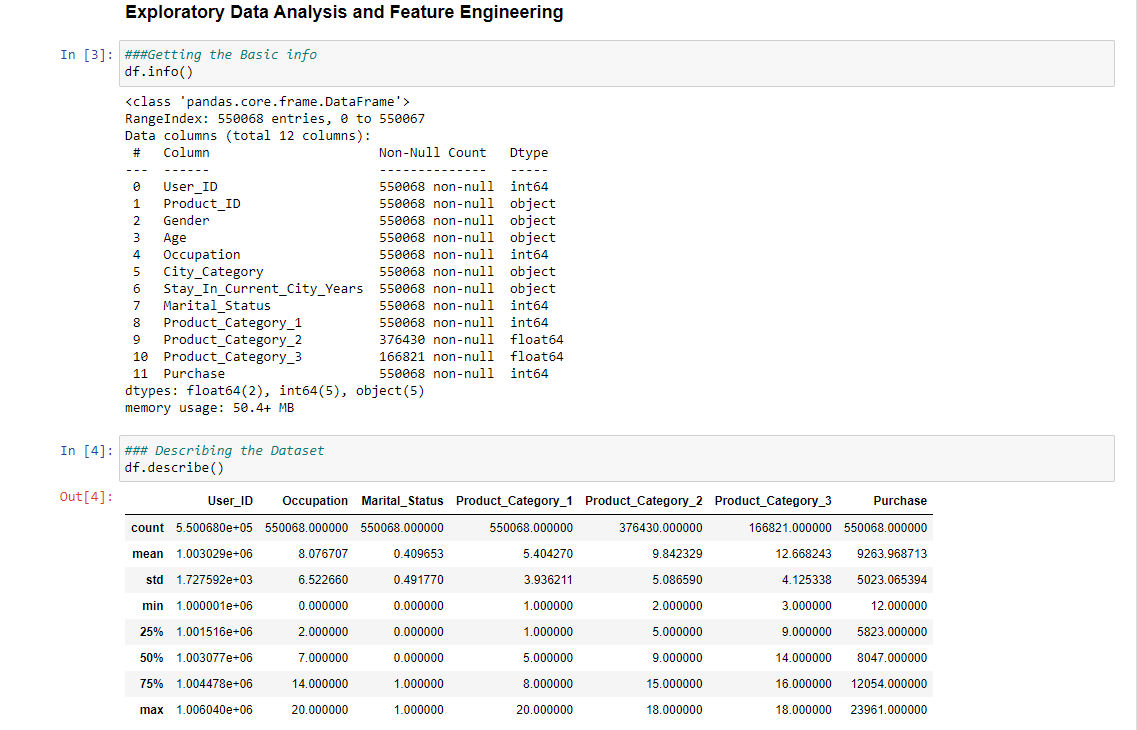
3.3 **Importing the Dataset in Jupyter Notebook**

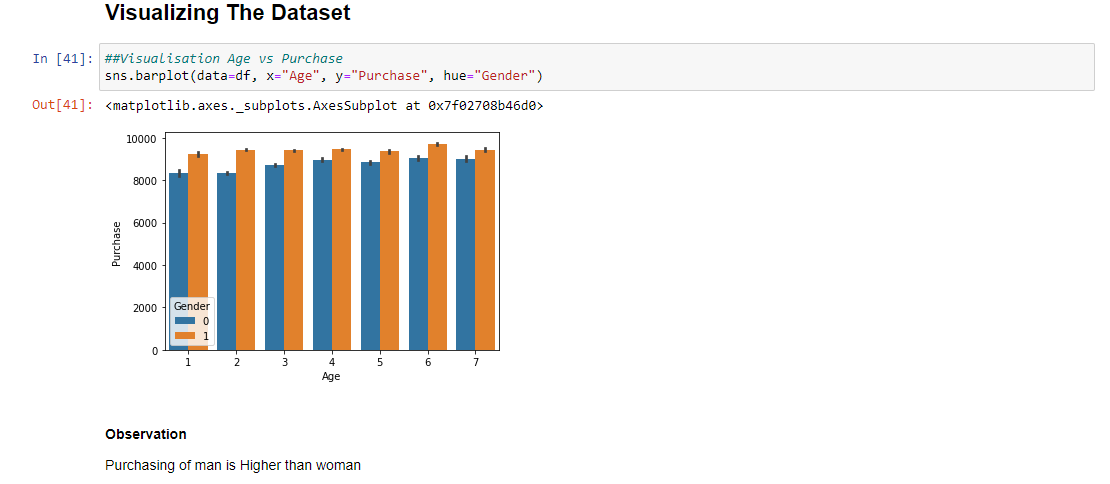
Importing the dataset is the vital step in the jupyter notebook, where the dataset is imported using the pandas library read\_csv is used to import the dataset.



3.4 **Exploratory Data Analysis**

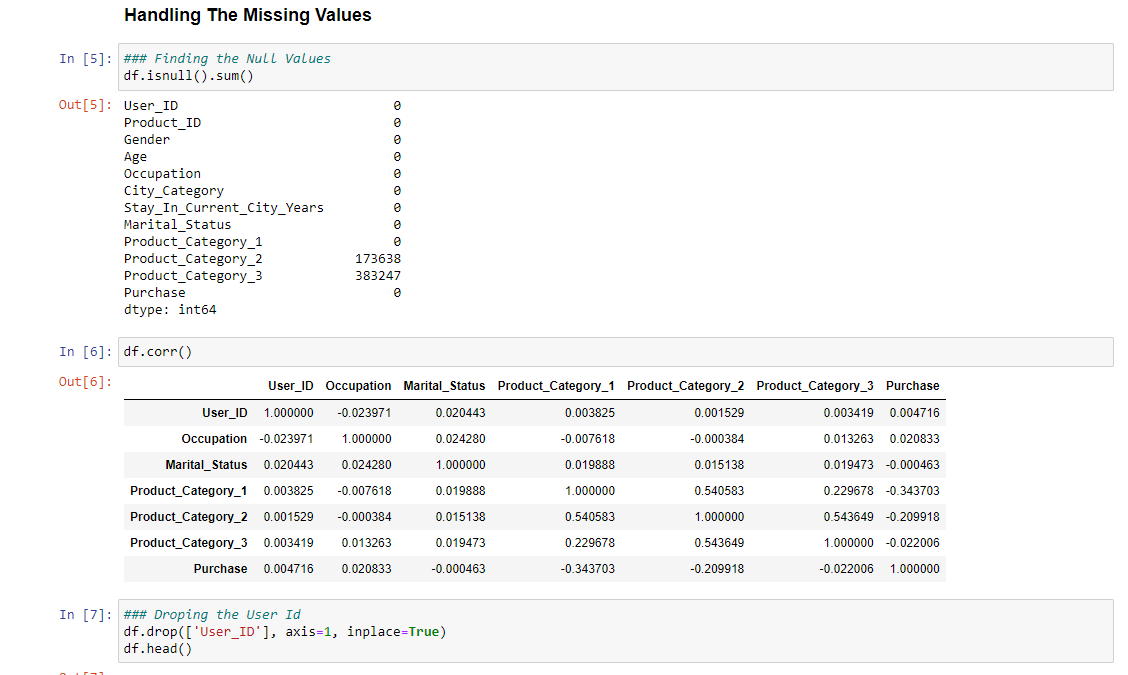
After the dataset is imported, Exploratory Data Analysis is done on the dataset, where various plotting of the data is done using matplotlib and seaborn library. Which helps to understand about correlation of the data, missing values present in the data and outliers present on the data.

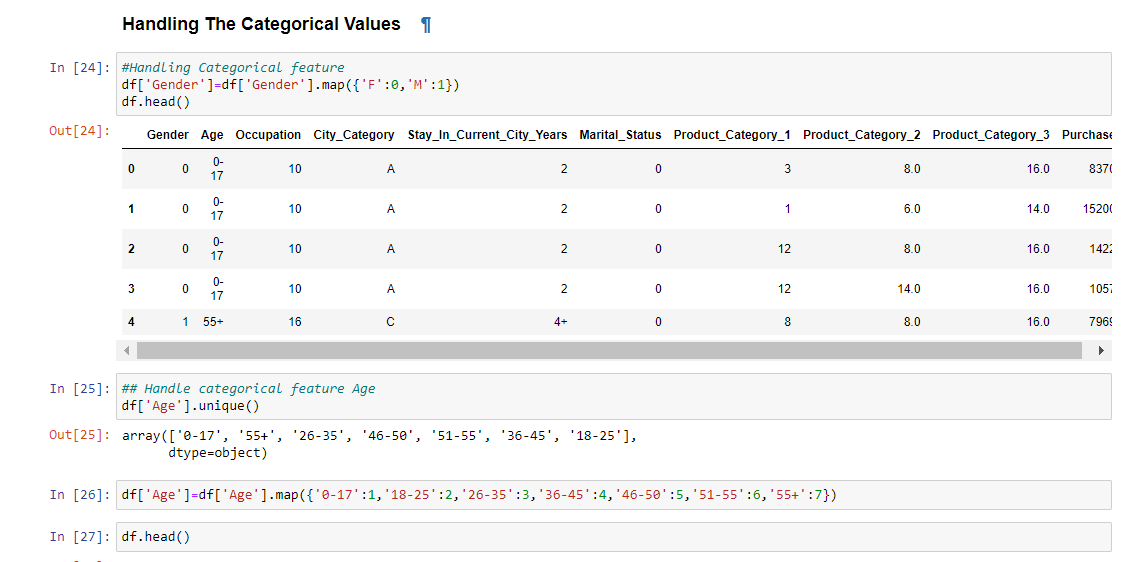




3.5 **Feature Engineering**

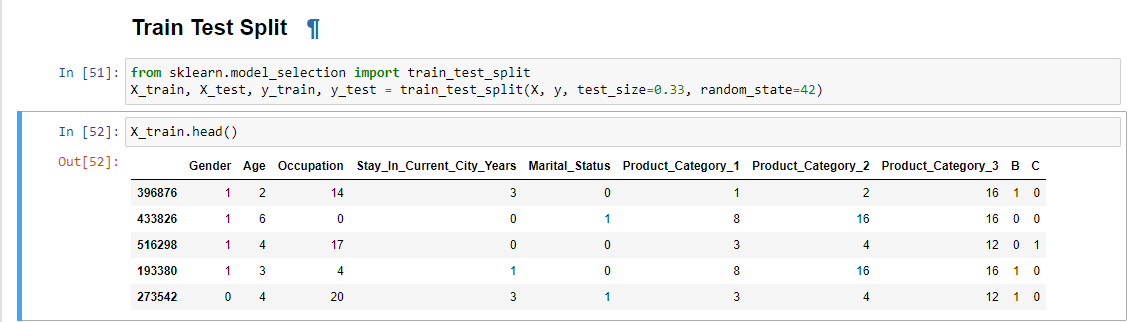
In Feature Engineering, Missing Value is Observed and replaced with other value, Categorical values are handled using pandas library and redundant values are dropped.





3.6 **Split into Train and Test Data**

Now, the clean dataset is split into train and test dataset by using the sklearn library, which is a required process during the model training part.

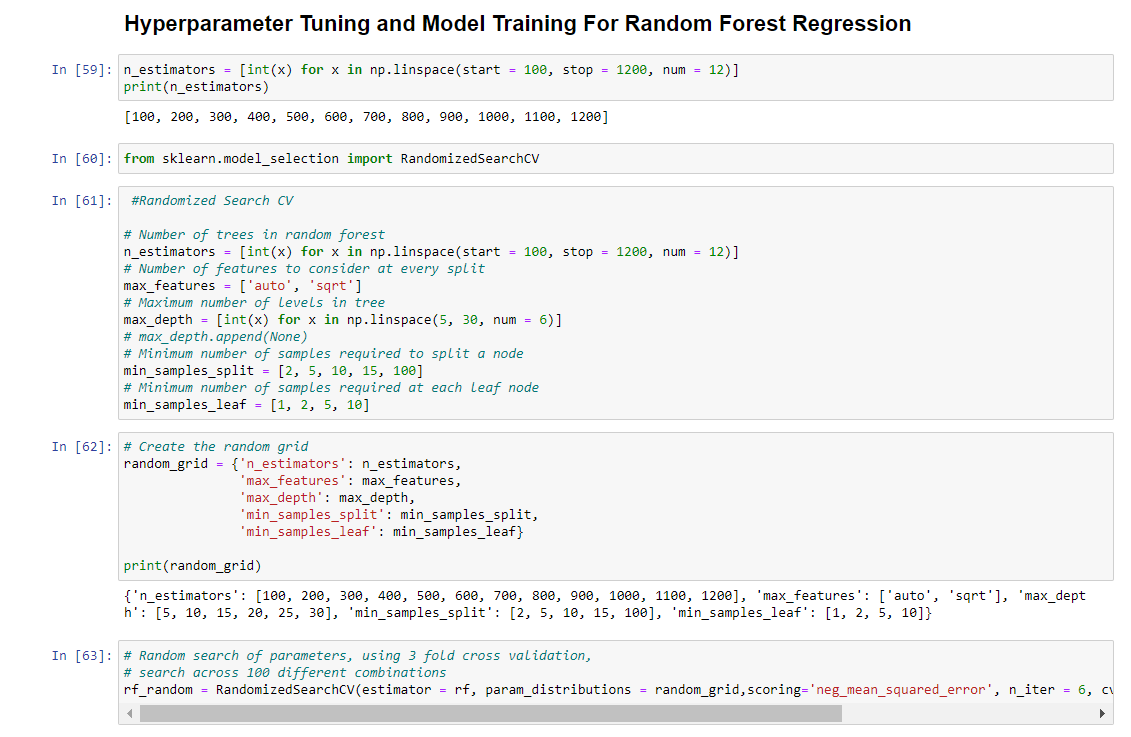


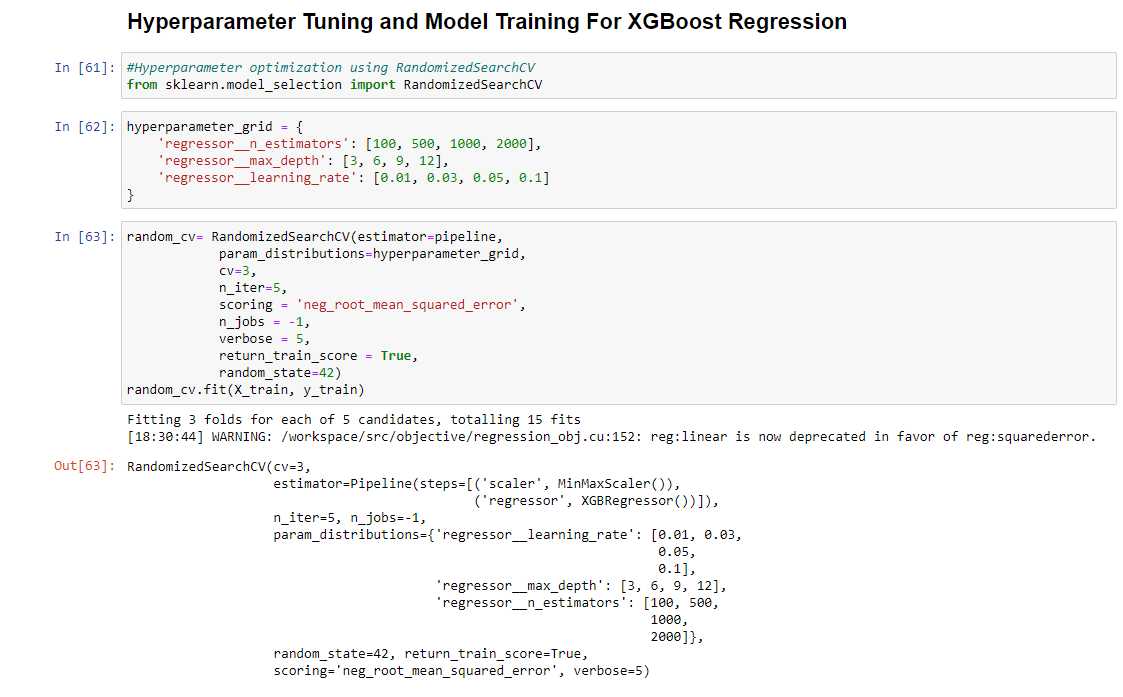
3.7 **Feature Scaling**

Now the dataset is scaled using the sklearn library, which helps to scaled the data if the data is not scaled, which is a required process during the model training part. But in case of Boosting method Scaling is not required it is automatically internally scaled.

3.8 **Hyper parameter Tuning**

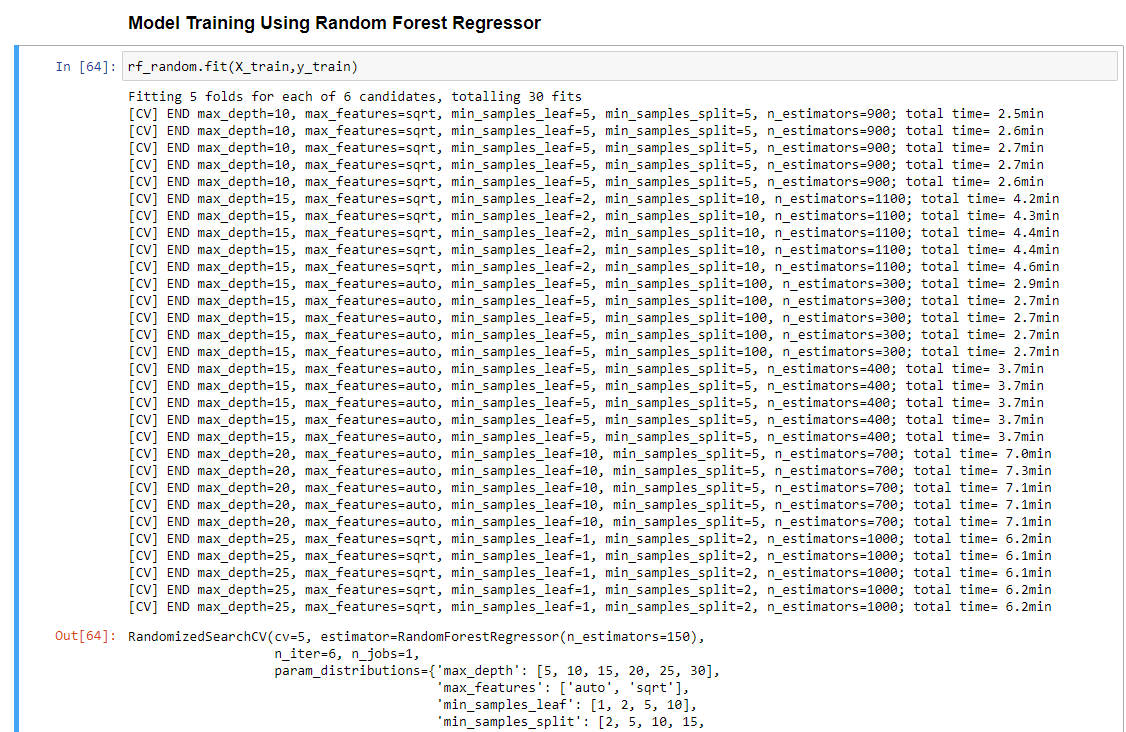
For the efficient and more precise training of the model, Hyper parameter tuning is a vital part before train a model, which helps to attains desired level of accuracy during the training of the model, which elevates the accuracy of the model.

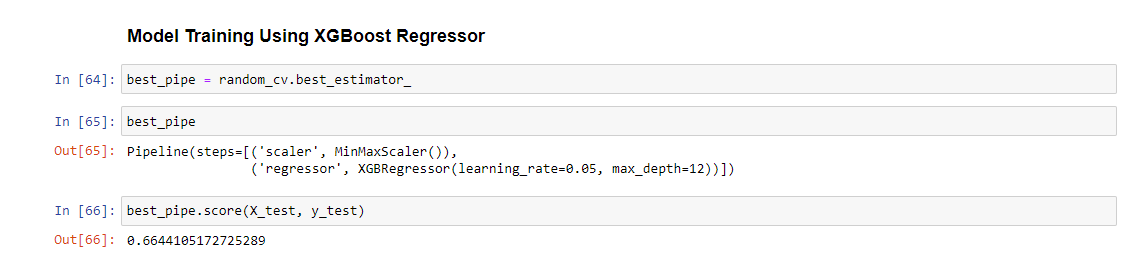




3.9 **Model Training**

The dataset is trained using two different model one is Random forest regressor and other one is XGBoost regressor which trained the dataset into desires level of accuracy and helps to attain precise level of prediction.

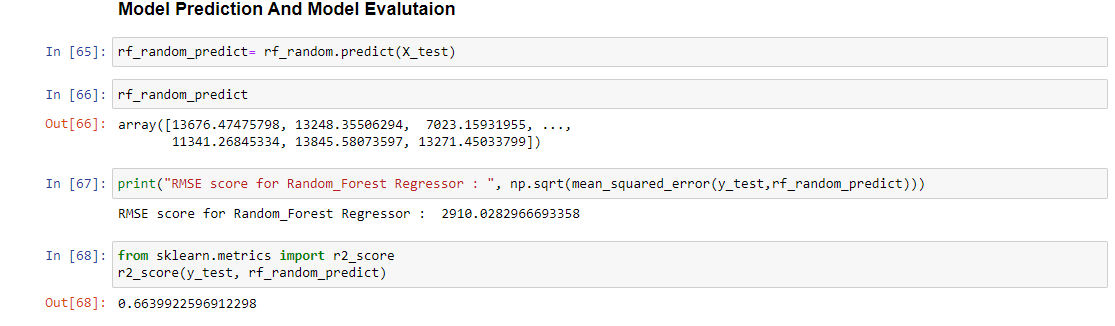




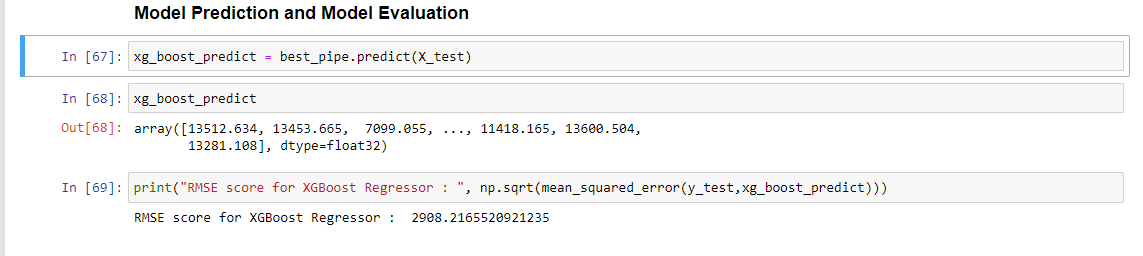
3.10 **Model Prediction and Evaluation**

Now the Trained model is use to find the predicted values of the model and the model is evaluated using the two estimator one is RMSE and other one is R2 Score, which gives an idea at what level of precision the model is trained.

**Random Forest Regressor**



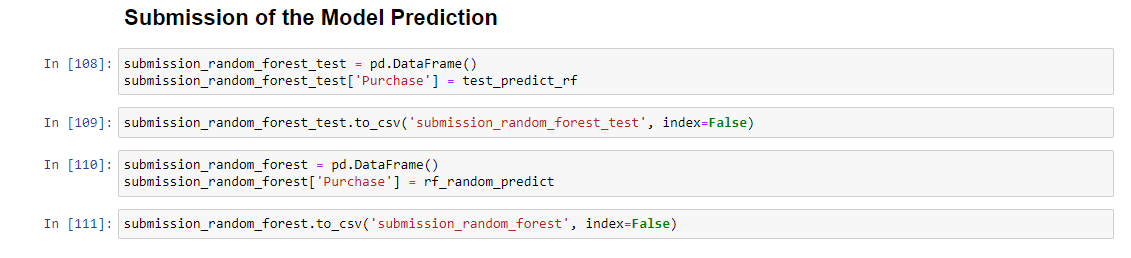
**XGBoost Regressor**



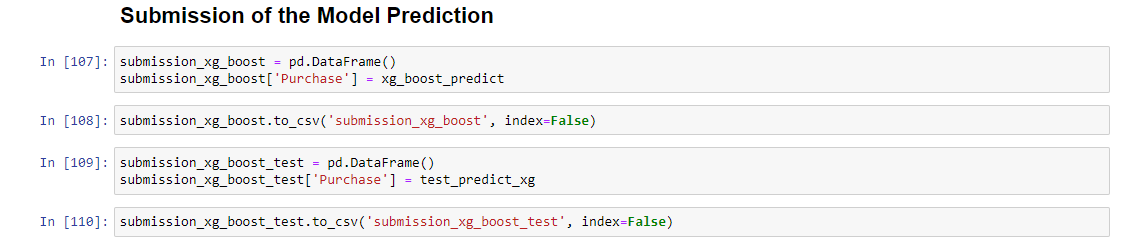
3.11 **Submission of the Predicted Values**

The final step is to submit the predicted values in the CSV form, so that it can be reuse for many other purposes.

**Random Forest Regressor**



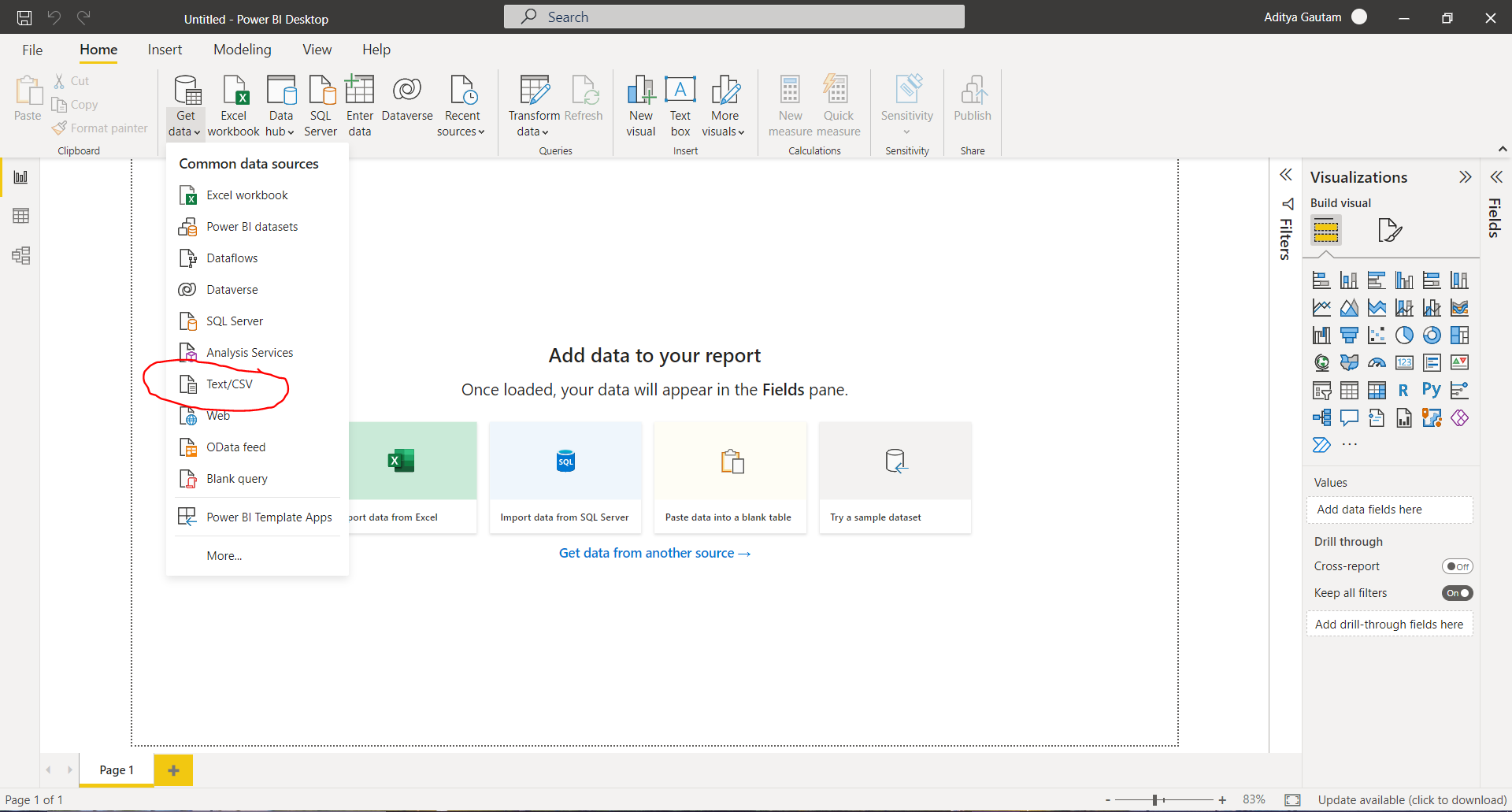
**XGBoost Regressor**



4. **Deployment**

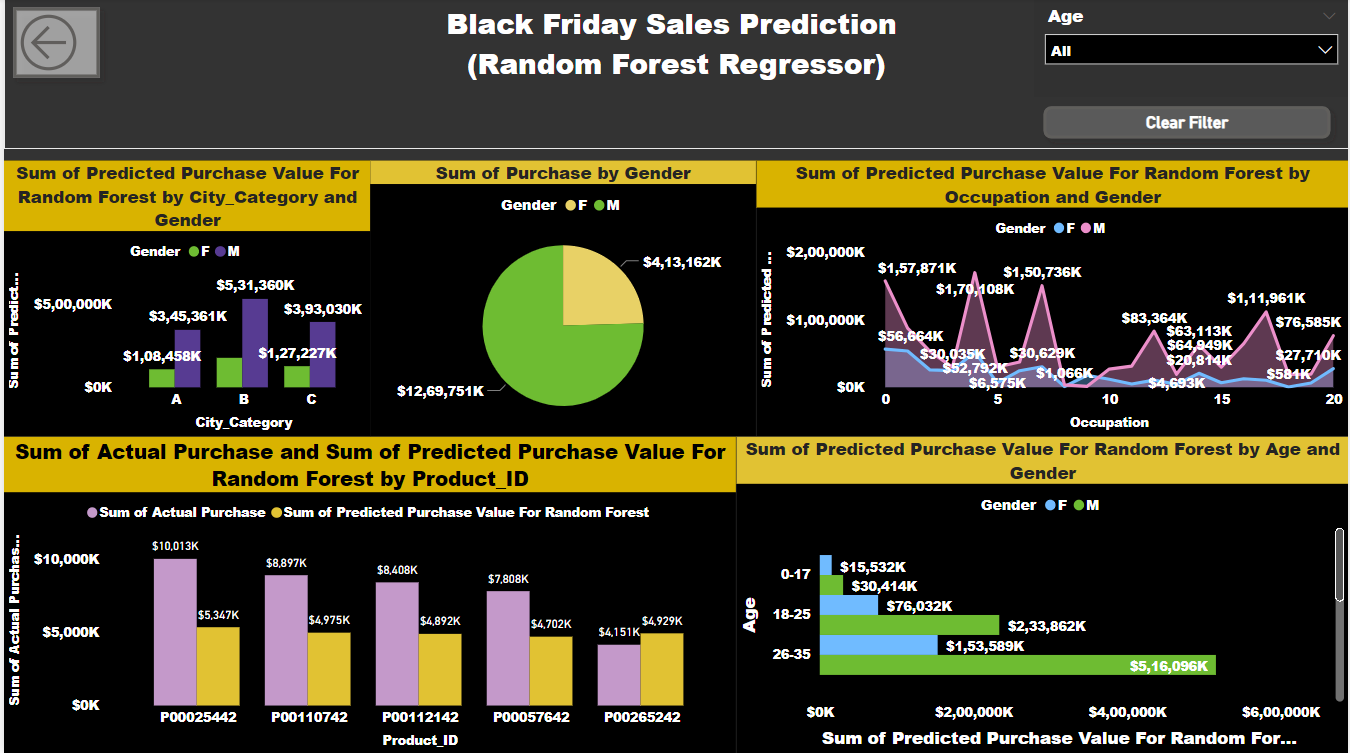
4.1 **Load the Dataset in Power BI**

Now, Load the Dataset into Power Bi for creating the dashboard.

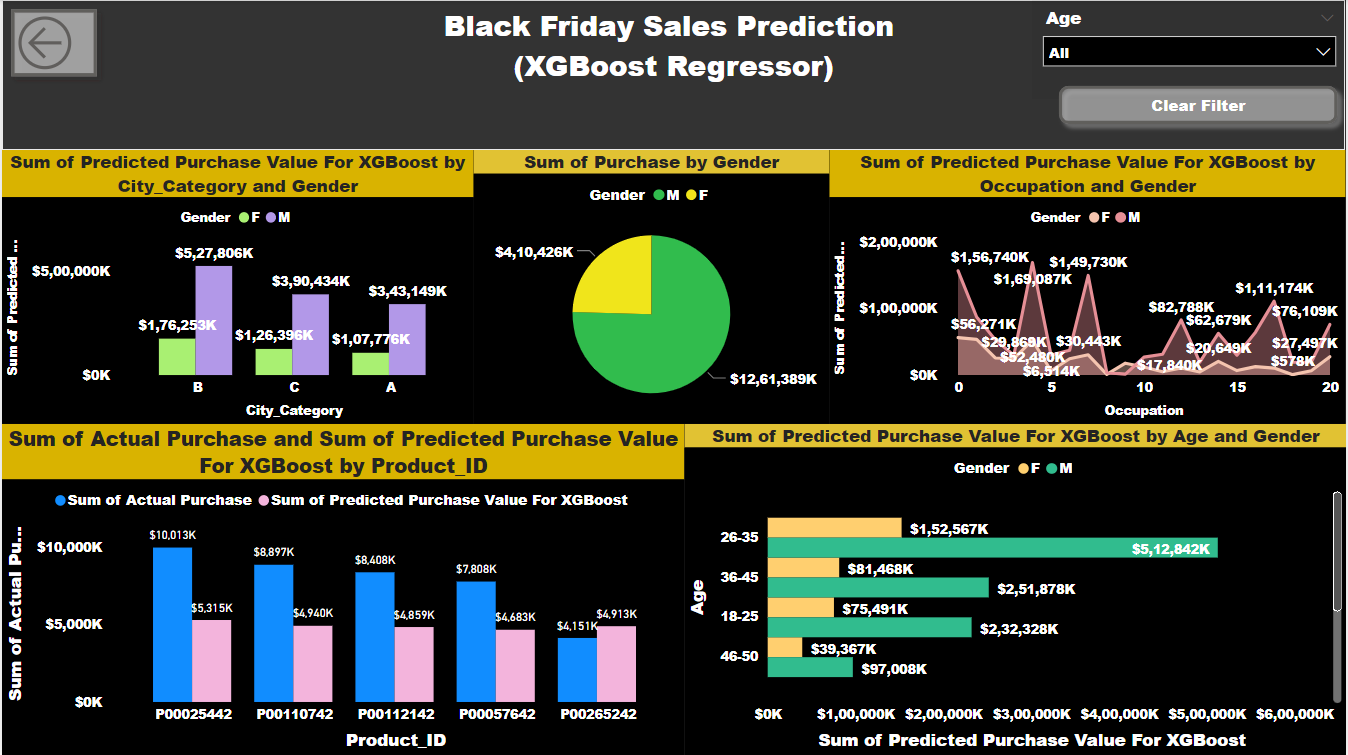


4.2 **Create the interactive dashboard**

**Random Forest Regressor**

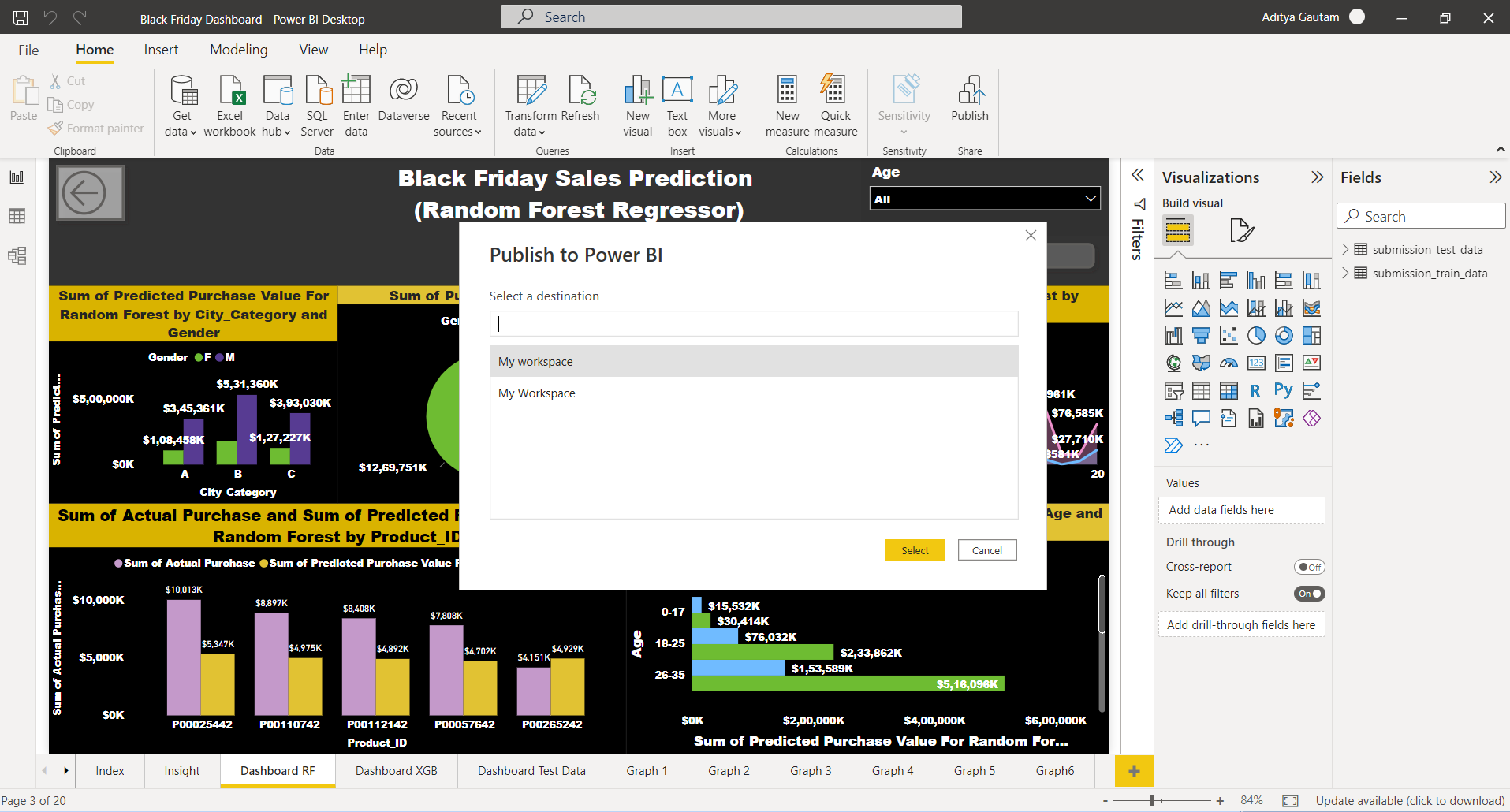


**XGBoost Regressor**

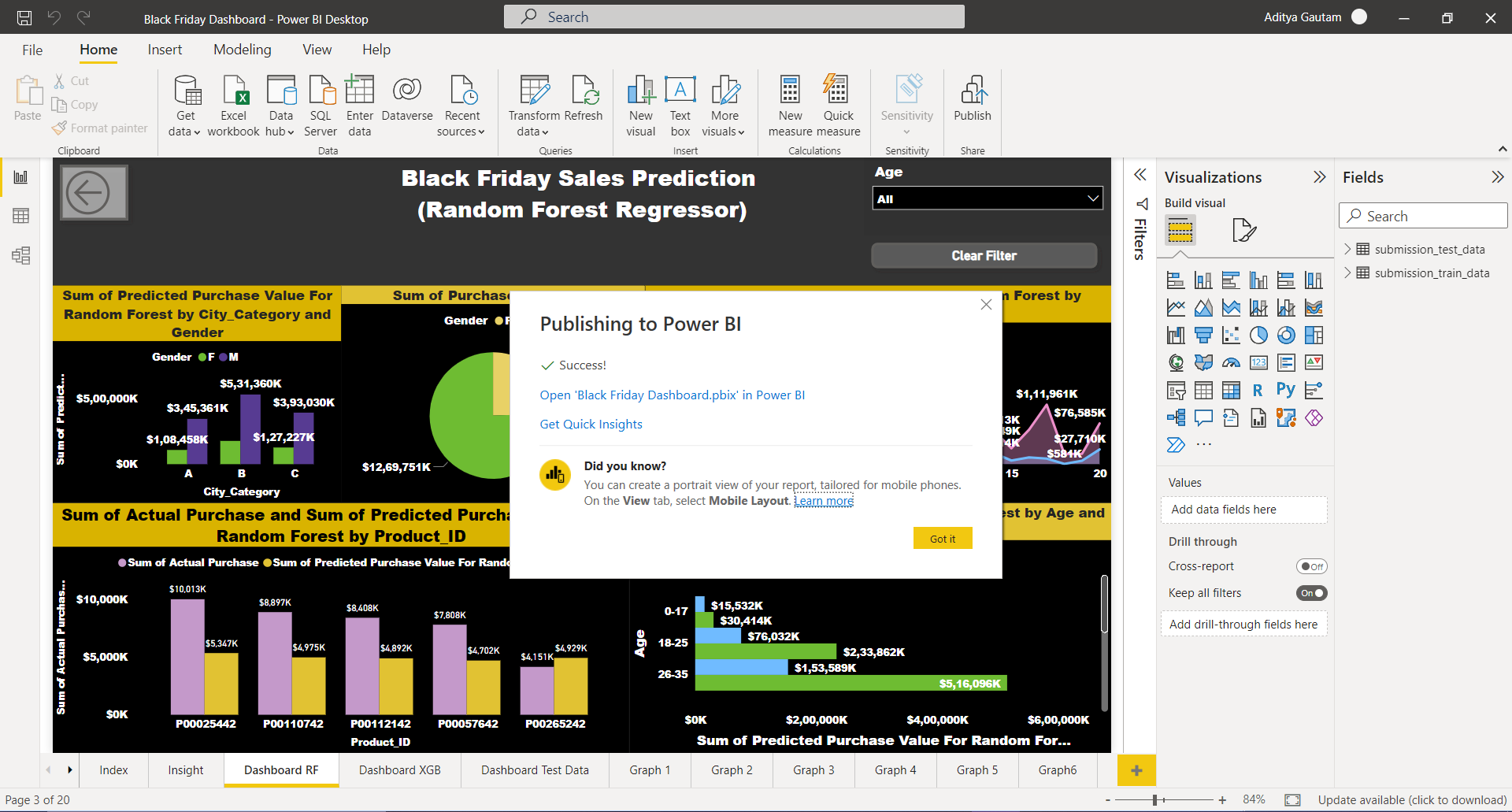


4.3 **Publish to Power Bi account**

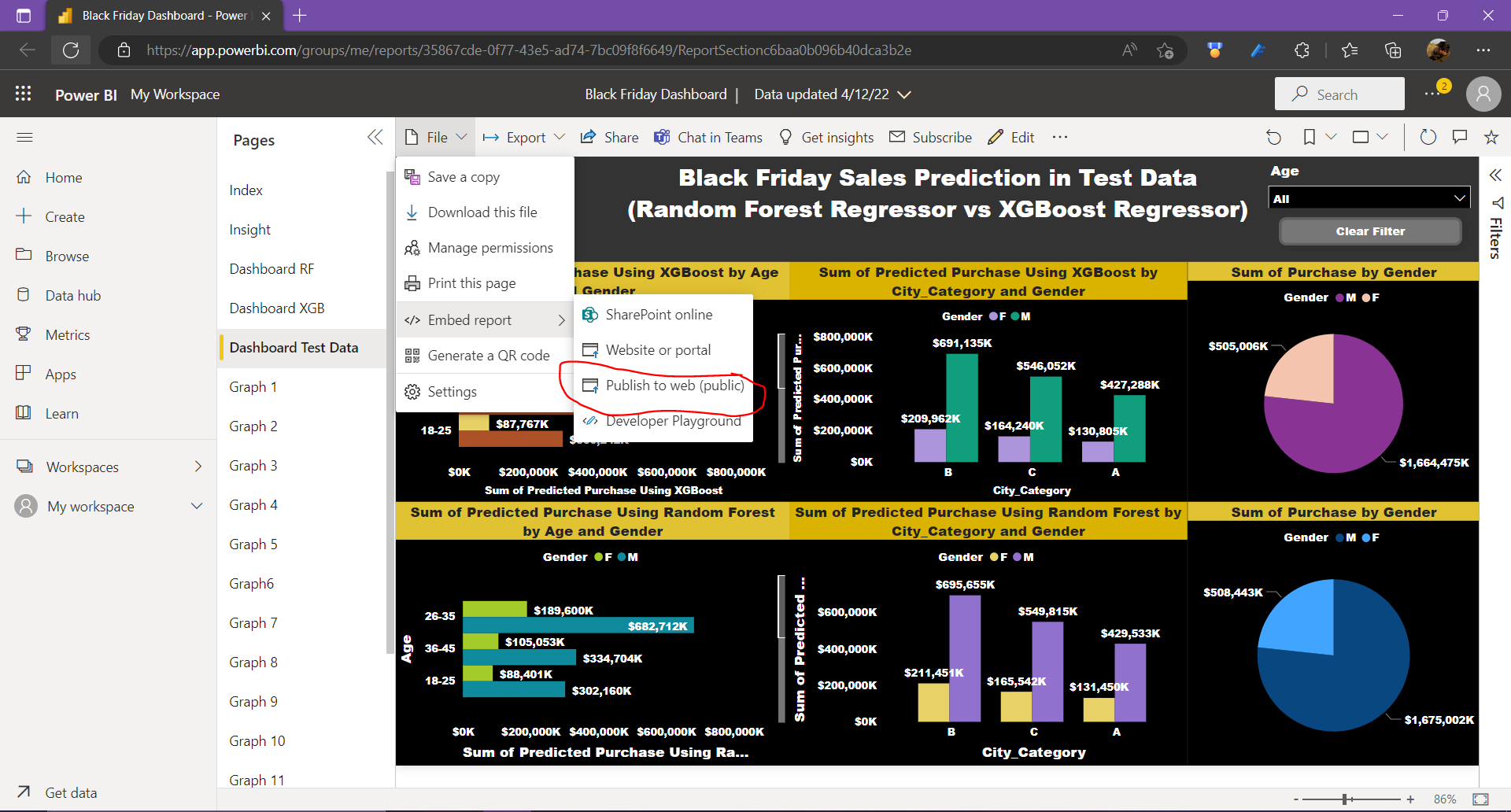
(1)



(2)

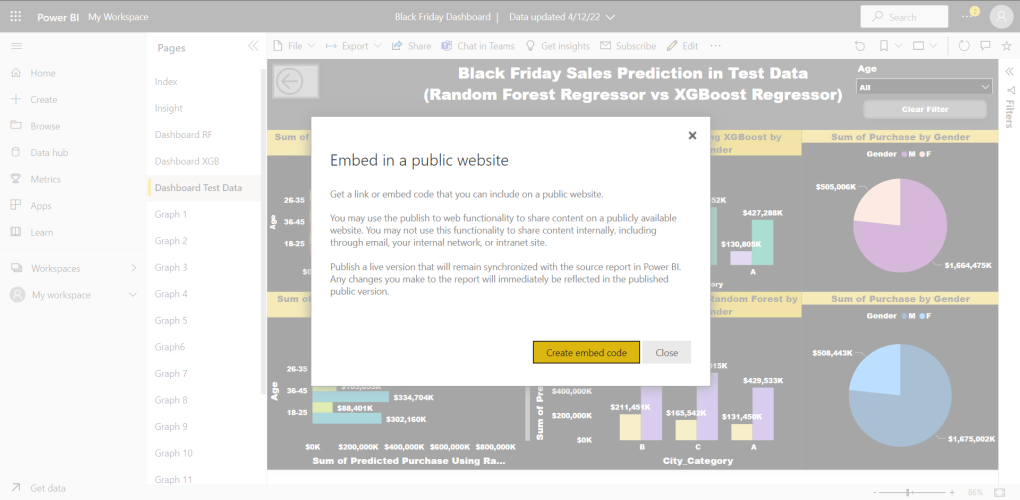


4.4 **Publish to Web**

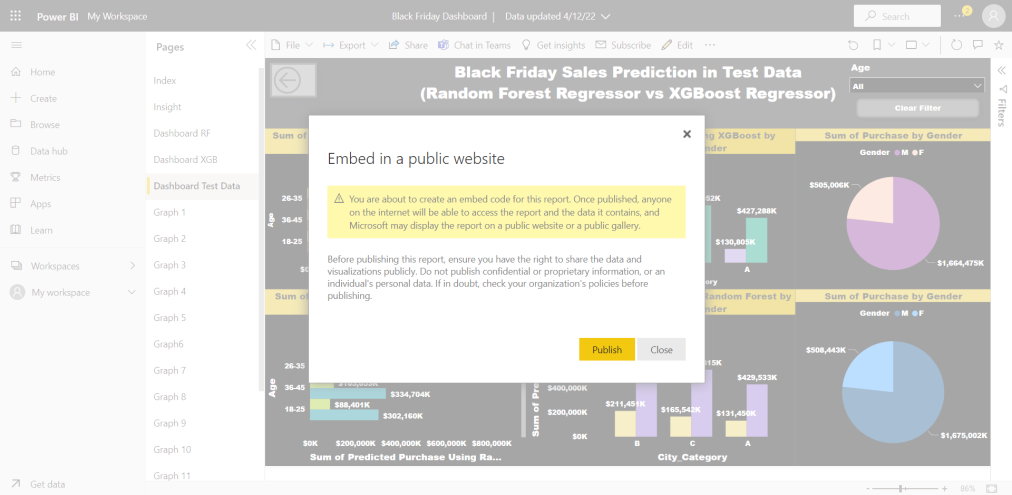


4.5 **Create the embedded code**

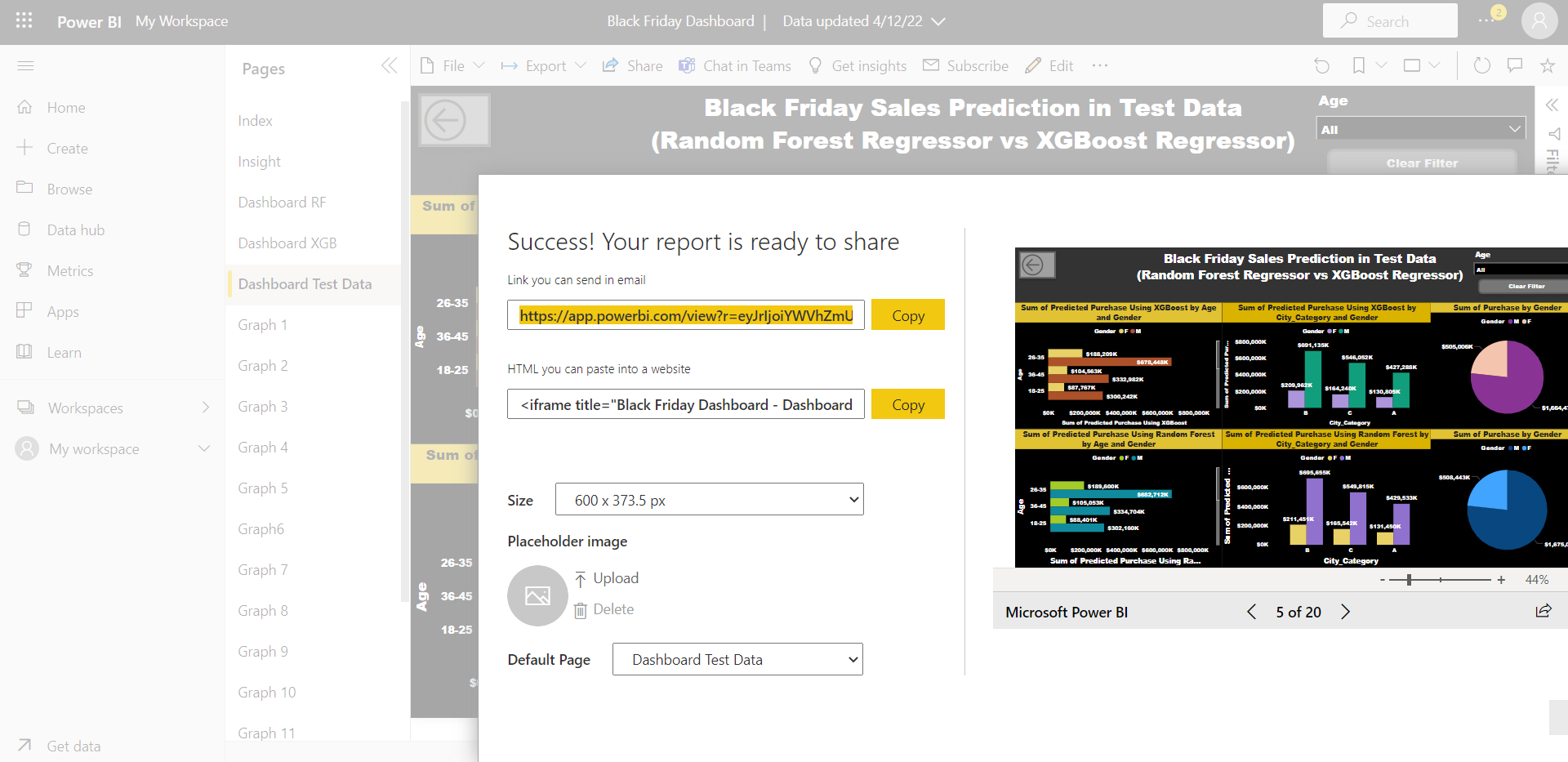
(1)



(2)



4.6 **Share the Public Link to Client**



**5. Conclusion**

**5.1 Insight of the Black Friday Sales Prediction**

**(1) Sum of Actual Purchase and Sum of Predicted purchase using Random Forest Regressor in Train Dataset, it is clear from the graph that the product id - P00025442 has the highest Actual Purchase with $10,013 k and has the highest Predicted Purchase with $ 5,347k.**

**(2) Sum of Actual Purchase and Sum of Predicted purchase using XGBoost Regressor in Train Dataset, it is clear from the graph that the product id - P00025442 has the highest Actual Purchase with $10,013 k and has the highest Predicted Purchase with $ 5,315k.**

**(3) Sum of Predicted Purchase Using Random Forest Regressor and XGBoost Regressor in Test Dataset, it clear from the graph that the product id - P00112142 has the highest Predicted Purchase with $10,243 k.**

**(4) Sum of Predicted Purchase Using Random Forest in Train data, it clear from the graph that the age between 26-35 has the highest Predicted purchase with $5,16,096K for Male and $1,53,589 K for female, where for city category B the predicted purchase is $5,31,360 K for male andfor female $1,77,477 K and for Overall predicted Purchase for male is $12,69,751 K and for female is $4,13,162 K.**

**(5) Sum of Predicted Purchase Using XGBoost Regressor in Train data, it clear from the graph that the age between 26-35 has the highest Predicted purchase with $5,12,842K for Male and $1,52,567 K for female, where for city category B the predicted purchase is $5,27,806 K for**

**male and for female $1,76,253 K and for Overall predicted Purchase for male is $12,61,389 K and for female is $4,10,426 K.**

**(6) Sum of Predicted Purchase Using Random Forest in Test data, it clear from the graph that the age between 26-35 has the highest Predicted purchase with $6,82,712 K for Male and $1,89,600 K for female, where for city category B the predicted purchase is $6,95,655 K for male and**

**for female $2,11,451 K and for Overall predicted Purchase for male is $16,75,002 K and for female is $5,08,443 K.**

**(7) Sum of Predicted Purchase Using XGBoost in Test data, it clear from the graph that the age between 26-35 has the highest Predicted purchase with $6,78,448 K for Male and $1,88,209 K for female, where for city category B the predicted purchase is $6,91,135 K for male and**

**for female $2,09,962 K and for Overall predicted Purchase for male is $16,64,475 K and for female is $5,05,006 K.**

k and has the highest Predicted Purchase with $ 5,347k.