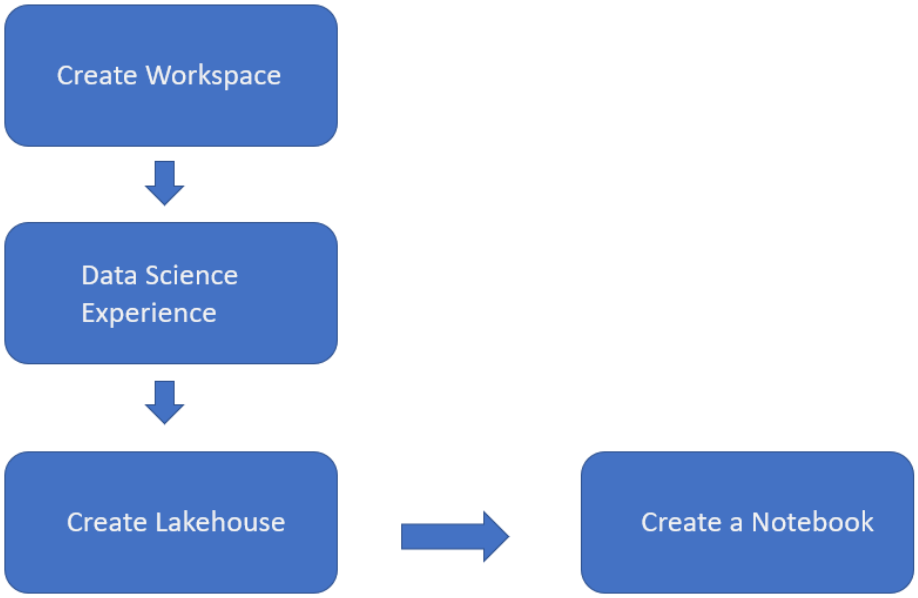
How is it easy to train machine learning models in Fabric?



Setting Up the Environment

A lakehouse was set up in Microsoft Fabric to store and organize the data in one place. It made accessing and managing datasets much easier and kept the workflow smooth.

The lakehouse is created under the Data science experience.

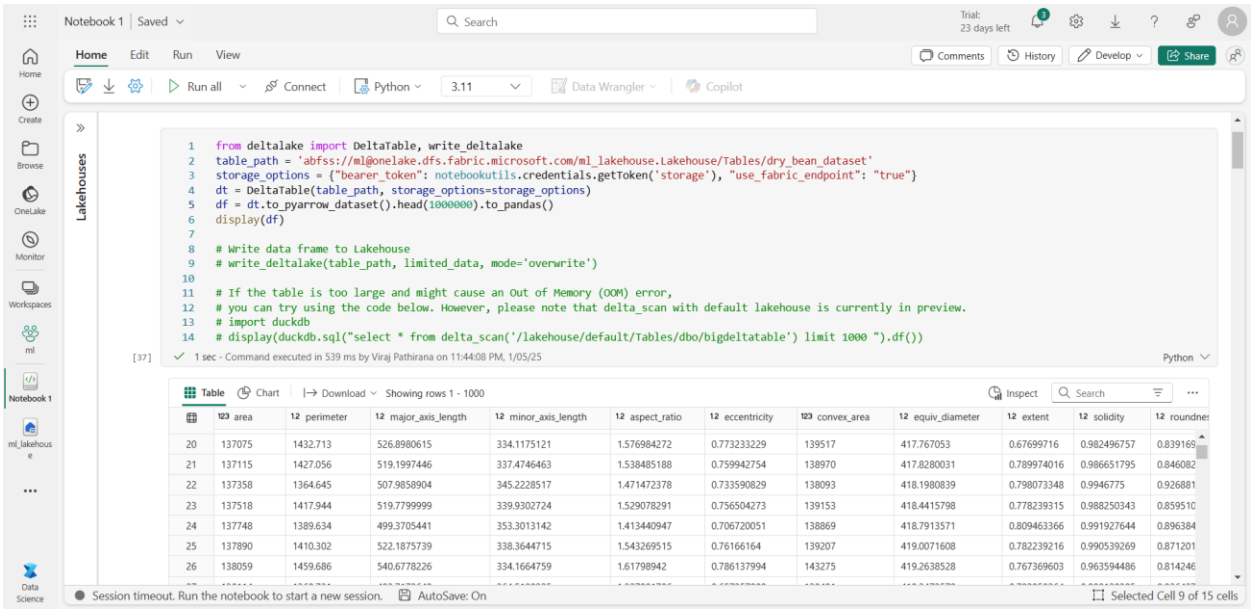
Utilizing Python in Notebooks

We can use notebooks to train machine learning models.

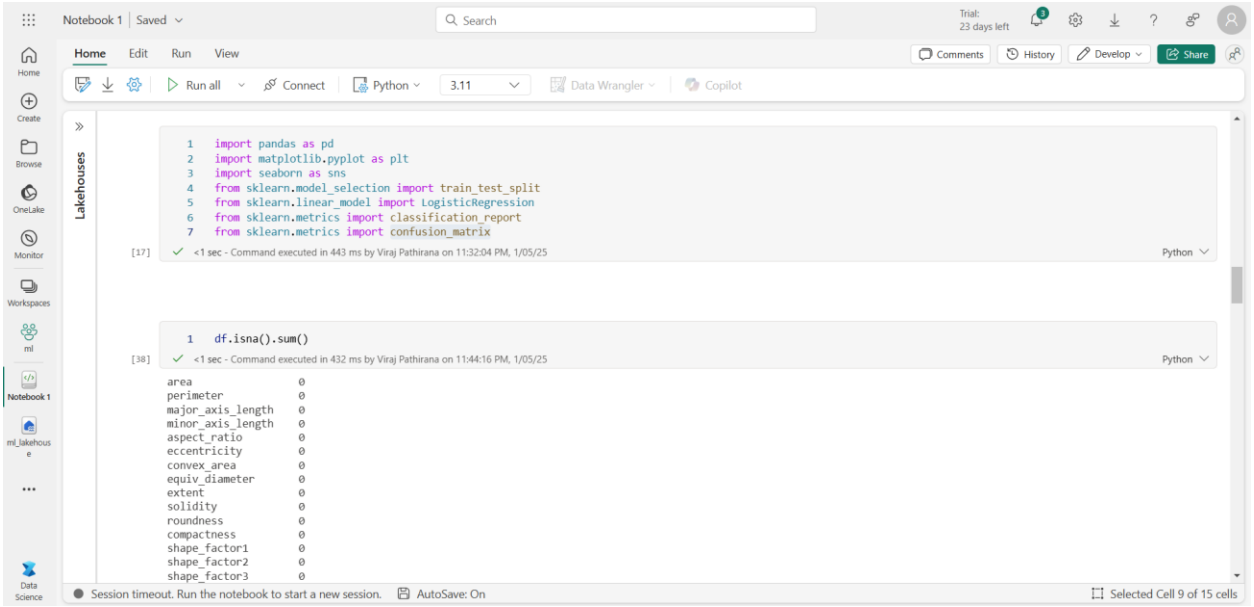
Python notebooks within Fabric were used to implement the logistic regression model. The flexibility of Python and its libraries, such as pandas, matplotlib, and scikit-learn, made data preprocessing, model training, and evaluation efficient and effective.

The workflow involved key steps, including

- Importing the Data from the lakehouse
- Data cleaning and preparation.
- Splitting data into training and testing sets.
- Training and evaluating the logistic regression model with relevant metrics.

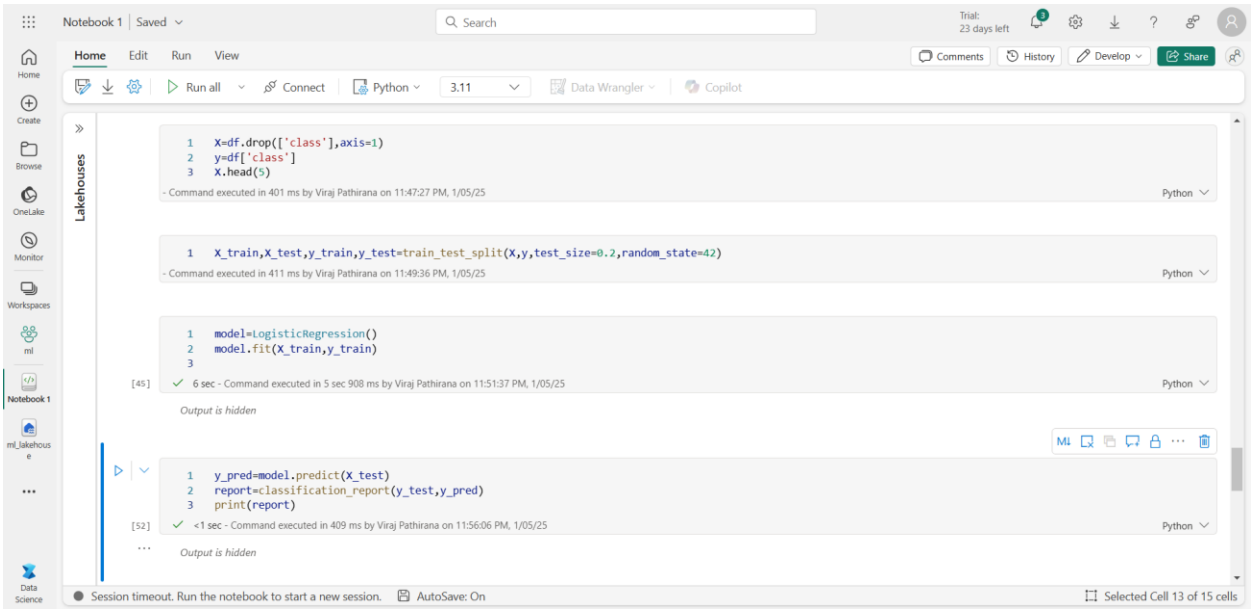


Importing Nessecary Libraries,



Model Training and Evaluation

The performance of the logistic regression model was thoroughly assessed using a range of evaluation metrics, including precision, recall, and F1-score, which provided a deeper understanding of the model's ability to make accurate predictions. These metrics helped identify how well the model performed in terms of minimizing false positives and false negatives, and the F1-score offered a balanced measure of precision and recall. By analyzing these insights, it became possible to pinpoint areas where the model could be improved, whether by fine-tuning hyperparameters, adjusting data preprocessing steps, or exploring alternative approaches to address any performance gaps.




```
1   y_pred=model.predict(X_test)
2   report=classification_report(y_test,y_pred)
3   print(report)
```

[52] ✓ <1 sec - Command executed in 409 ms by Viraj Pathirana on 11:56:06 PM, 1/05/25

```
              precision    recall  f1-score   support

    BARBUNYA       0.61      0.49      0.54       265
      BOMBAY       1.00      0.99      1.00       117
        CALI       0.71      0.74      0.72       316
    DERMASON       0.80      0.85      0.82       671
       HOROZ       0.60      0.53      0.56       391
       SEKER       0.69      0.62      0.65       427
        SIRA       0.57      0.66      0.61       536

    accuracy                           0.69      2723
   macro avg       0.71      0.70      0.70      2723
weighted avg       0.69      0.69      0.69      2723
```

Why Microsoft Fabric Stands Out?

Microsoft Fabric's unified platform simplified every aspect of the machine learning workflow. The integration of lakehouses, notebooks, and computational resources in a single environment eliminated the need for switching between tools.

The Data Science experience in Fabric offers an intuitive and efficient way to manage data science projects while leveraging Python for machine learning tasks.

Lessons and Insights

- The project underscored the importance of effective data preprocessing, rigorous evaluation of model performance, and leveraging a unified platform for data science workflows

- Using a lakehouse for data organization enhanced efficiency and streamlined the data access process

What are the Advantages of using Fabric?

To conclude, there is a variety of benefits that arise from the use of Microsoft Fabric notebooks, making them especially useful tools in data science and machine learning activities. Even though VS Code is an advanced and custom environment for writing code, allowing the user to have endless capabilities, Microsoft Fabric notebooks offer a unified experience that integrates many tools within itself and relieves a great deal of work related to tracing data in extensive projects. Writing and executing python code in microsoft fabric is possible, just as it is in VS Code, however it is made more efficient due to being integrated with other microsoft fabric services such as lakehouses, data pipelines, and power bi.

Working in Fabric notebooks has one of its major advantages in that everything can be accomplished in one place. In VS Code, data management, storage, computation and visualization systems would need to be independently organized, usually needing extra processes and tools to achieve the desired integration. All the numerous services supported by Microsoft Fabric are tightly integrated and collocated in a single system, allowing for better productivity and teamwork.

Microsoft Fabric also has the benefit of being able to scale. When dealing with extensive data or more technical ML Models Machine Managed can automate infrastructure provisioning on-demand without setting up another infrastructure beforehand.

In addition, Microsoft Fabric has a friendly interface and sophisticated analytical features so that even inexperienced users and experienced ones can leverage its capabilities. Microsoft Fabric provides the necessary tools that allow you to customize your solution whether you need to create a straightforward model or execute complicated data science procedures.

Ultimately, with Microsoft Fabric notebooks, activities related to data science are further organized, strengthened by teamwork and are designed in a way so that these activities can grow in quantity. As a result a professional is able to use such tools more comfortably because there is no need to physically operate any structure. Because a single platform is offered to the client for everything, cross-team collaboration, working with big data, and model building are all possible, and the need to use diverse tools and platforms is reduced.