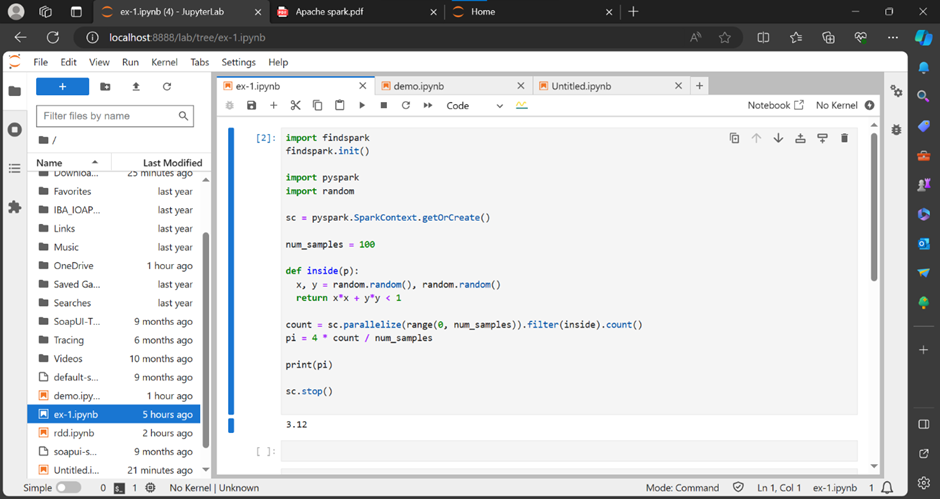
**G. Rupa Manasa 26-12-2023**

**Day 19 Assessment**

**Installation of Jupyter:**

Jupyter is installed by using the command **py -m pip install jupyter.**

It is launched by using **jupyter lab.**



**Why Choose PySpark for ETL?**

**Performance:**

PySpark leverages in-memory computing, making ETL processes faster than ever.

**Ease of Use:**

Python developers can seamlessly transition to PySpark due to its Pythonic syntax.

**Scalability:**

Handle massive datasets with ease, thanks to Spark’s distributed processing.

**Rich Ecosystem:**

PySpark integrates with popular tools and libraries, making it versatile for various data tasks.

**ETL Workflow:**

Extract: Retrieve data from various sources like databases, files, or APIs.

Transform: Clean, aggregate, and manipulate data to fit your analysis needs.

Load: Store the transformed data into a database or data warehouse for analysis.

**Real-World Applications:**

Perform large-scale data cleansing and preparation.

Analyze streaming data in real-time.

Create data pipelines for machine learning and AI applications.

Handle structured and unstructured data effortlessly.