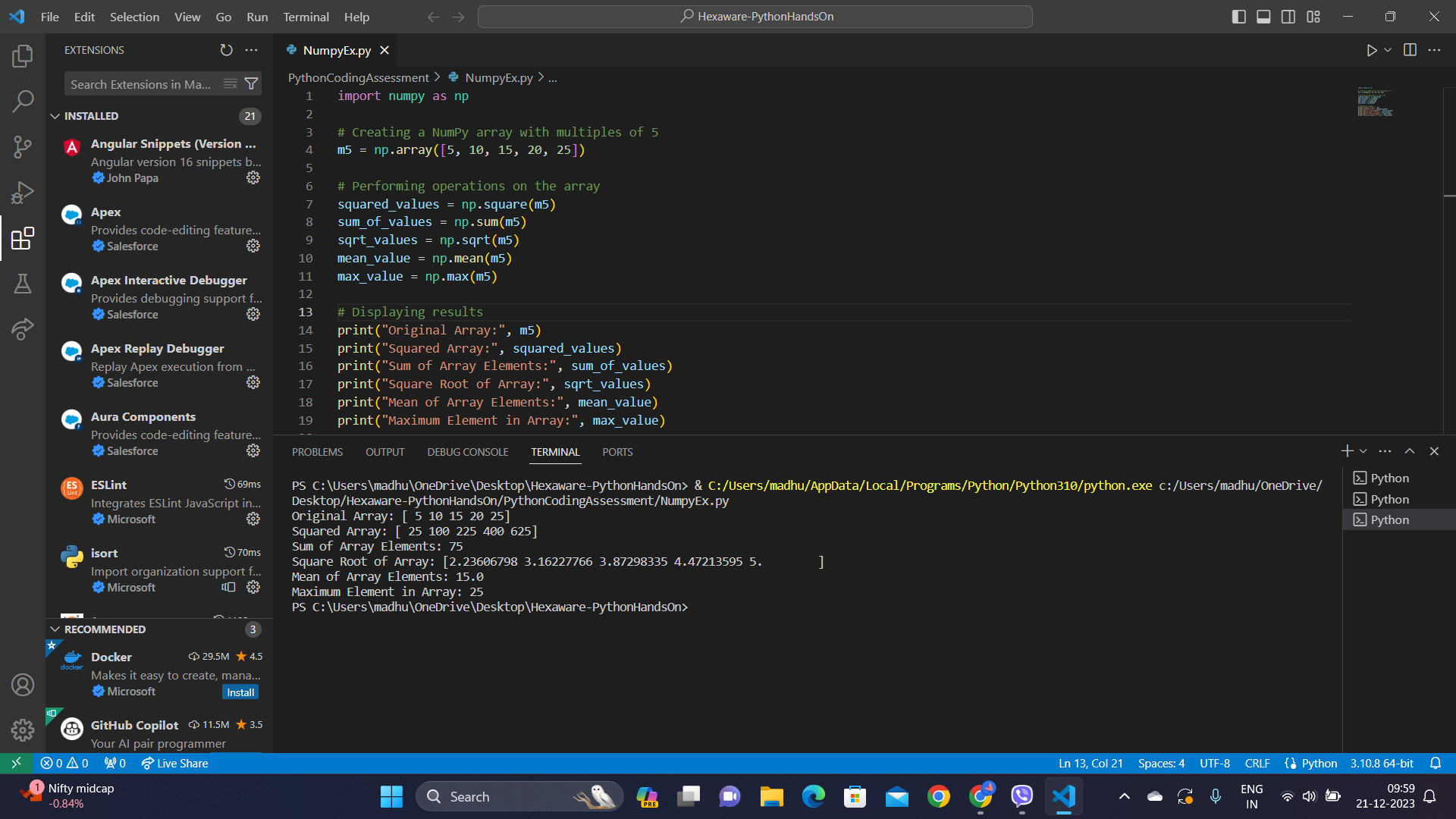
**Day- 15  
Madhu Kalyani Gadi (20-12-2023)**

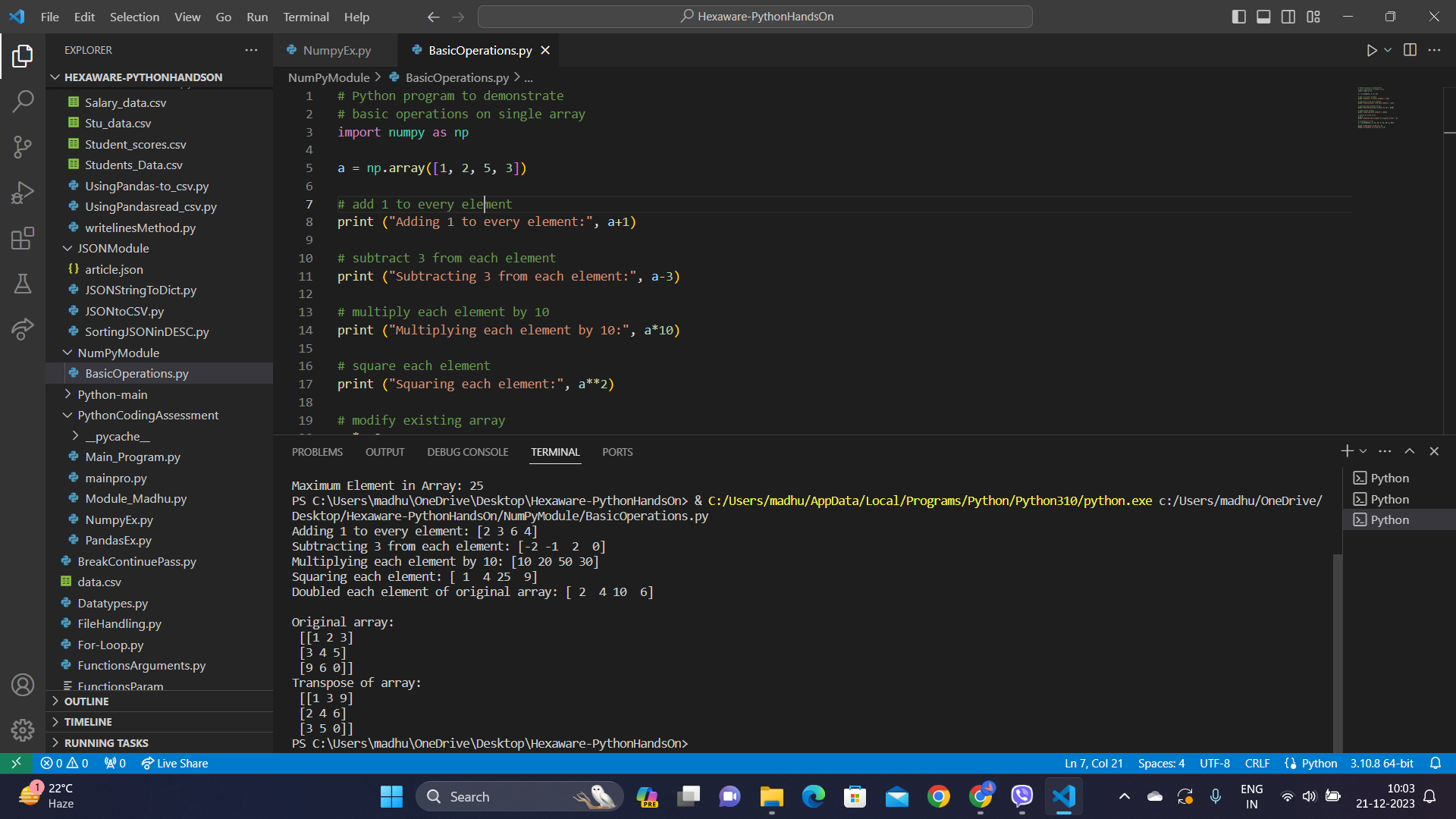
**Numpy Module**  
  
NumPy is a powerful tool in Python for working with arrays, which are like lists but more efficient for mathematical operations. It helps with scientific computing and has features like handling multi-dimensional arrays, functions, and integration with other languages.

**Numpy Arrays**

Numpy arrays are efficient, homogeneous, multidimensional data structures in Python, created using np.array(). It supports powerful element wise operations, making numerical computations concise and fast.



**Numpy Basic operations:**



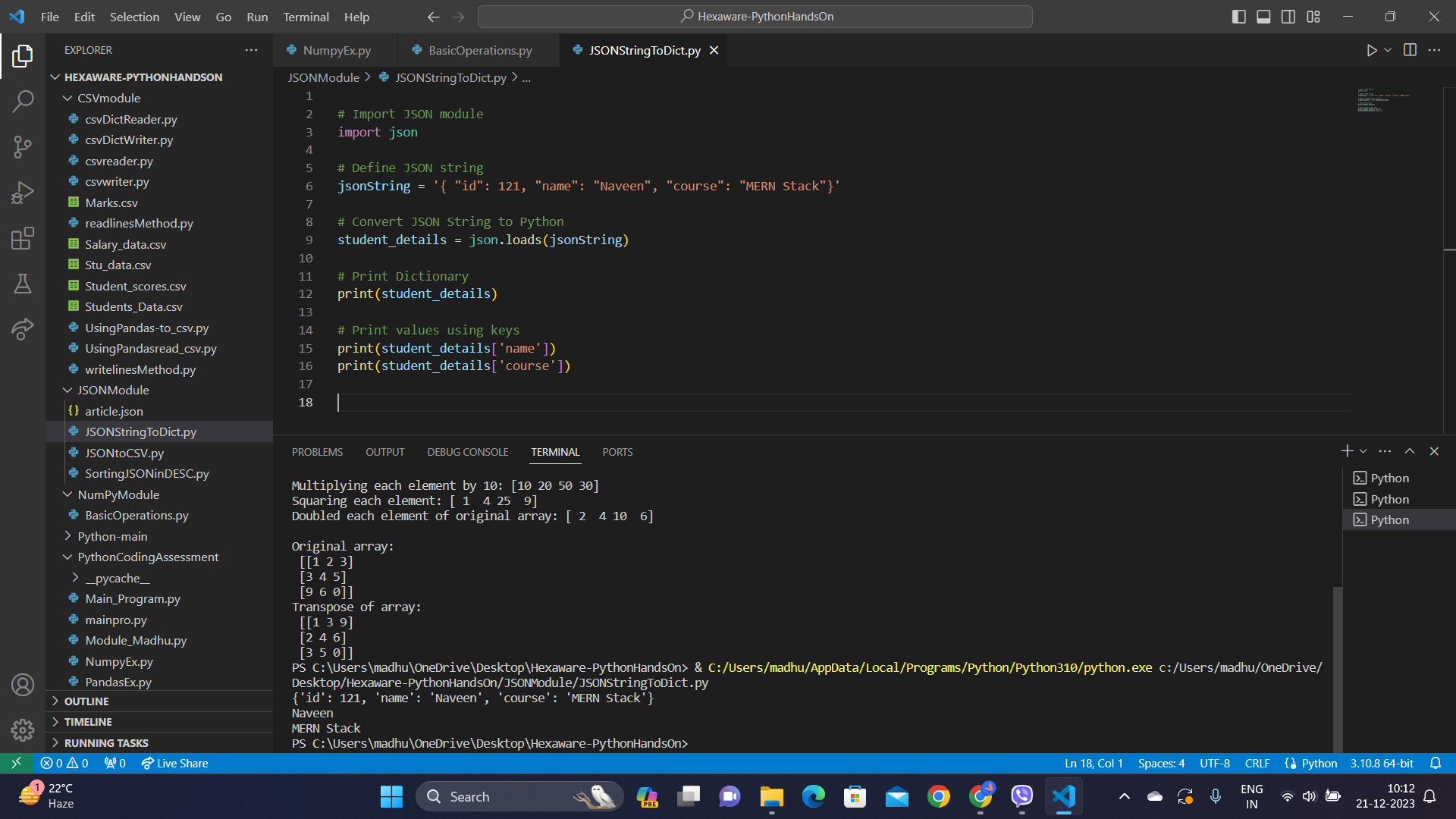
- An array a is created and basic operations are demonstrated, including adding 1, subtracting 3, multiplying by 10, and squaring each element.

- The original array is then modified in place by doubling each element using the a \*= 2 operation.

- Additionally, the transpose of a 2D array is shown using a.T, switching rows with columns, providing a concise demonstration of NumPy's array manipulation capabilities.

**JSON**

JSON is like a language for computers to talk to each other, which stands for JavaScript Object Notation, is a lightweight and widely used data interchange format. It serves as a way to represent and exchange structured data between systems.



**BIG DATA:** (intro)

Big Data refers to the vast amount of information that is too large and complex to be processed by traditional data processing tools. The challenges with Big Data involve not only the large size but also the speed at which it is generated and the variety of formats it can take. Traditional methods can struggle to handle this large data, so specialized tools and technologies are used to store, process, and analyze the data.

Three kinds of data in Big data. They are

1. Unstructured data

2. Semi structured data

3. Structured data

**HDFS:**

Hadoop Distributed File System (HDFS) is the storage component of the Hadoop framework, designed for distributed and scalable storage of large datasets. It breaks data into blocks and distributes them across a cluster of machines for redundancy and efficient processing.