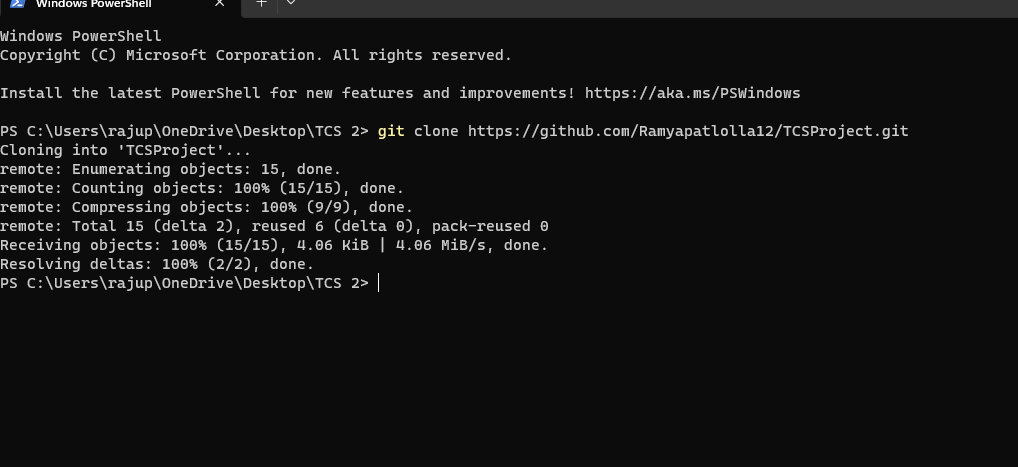
1.Create Scala project through eclipse

* Open eclipse
* Go to file > new > scala project
* Set Scala Library container version to 2.11.11
* Add external spark jars to give the control over spark
* Under src >> create package >> create object task
* Write spark application code.
* Convert the code to maven project.

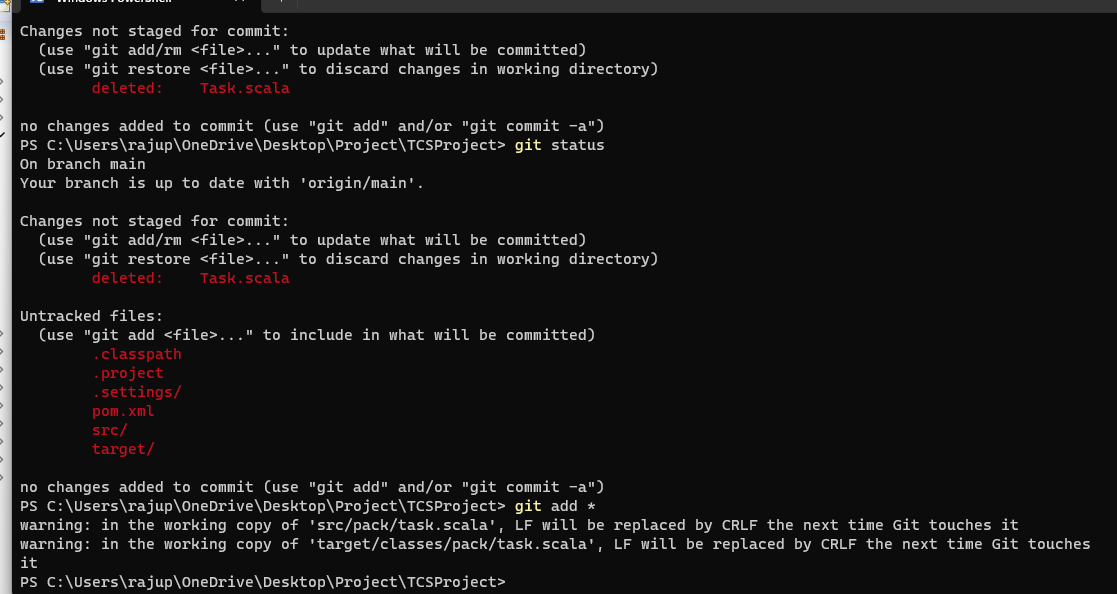
**2.GIT commands**

1. **Git Clone:** cloning my git repository using git clone

**Using command : git clone** [**https://github.com/Ramyapatlolla12/TCSProject/blob/main/Task.scala**](https://github.com/Ramyapatlolla12/TCSProject/blob/main/Task.scala)

****

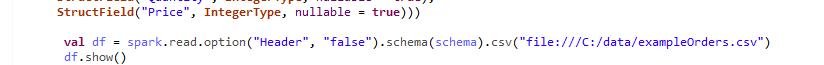
1. **Whatever you do the code changes you need to keep in that floder.**
2. **Then GIT status :** Displays the state of code changes in directory
3. **Git ADD :** To add the files to the staged area.



1. **GIT Commit:**  To commit the code changed with message
2. **GIT push :** pushing the code to repository

**Code walk through :**

* Imported the required packages
* Spark configuration and SparkContext are initialized.
* A SparkSession is created
* A schema is defined for the expected structure of the DataFrame
* Read the Csv file



* Casting the timestamp
* Filtering the Buyers and sellers
* To get matched data used inner join.
* Renamed the seller columns to avoid ambitus
* Creating new columns for matched buyers and matched sellers with existing buyer order id and seller id
* Select required columns and dropping remaining columns.

