**Assignment Documentation:**

Input Datasets: whatson.csv, started\_streams.csv

Tool: Eclipse

Code: Scala

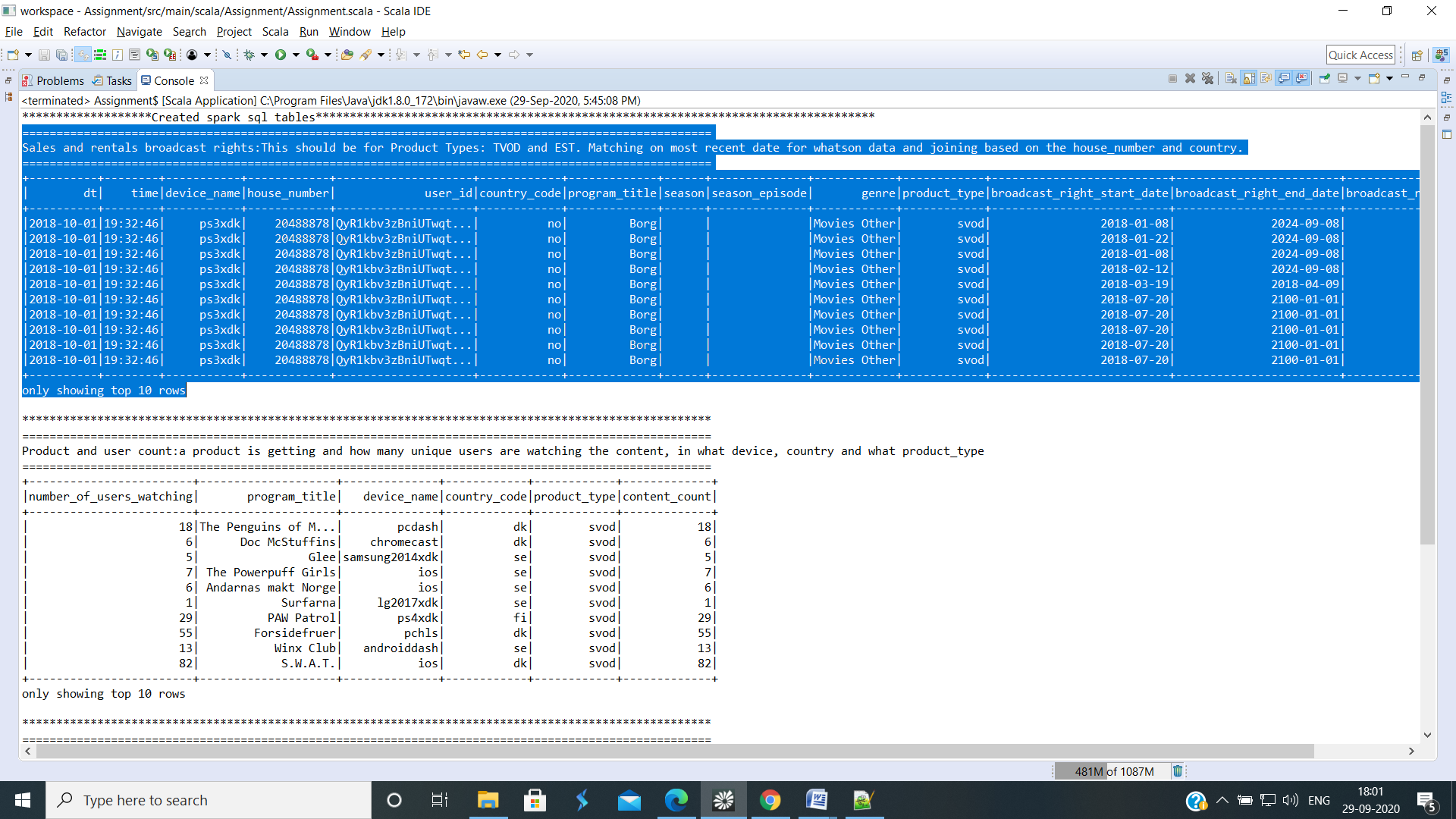
I developed following below steps:

**Task 1:Sales and rentals broadcast rights:**

We need to find the broadcast rights for a title to be able to expose it to the analytics team for further analysis. This should be for Product Types: TVOD and EST. Matching on most recent date for whatson data and joining based on the house\_number and country.

* Initialize the case classes with datatype of string given for all columns to both datasets and read into schema RDD string then change to Data Frame.
* Using spark.sql import can possible to write sql transformations in scala and develop the inner join query on dt and house\_number columns .

Output screen:



**Task 2: Product and user count:**

Input Datasets: started\_streams.csv

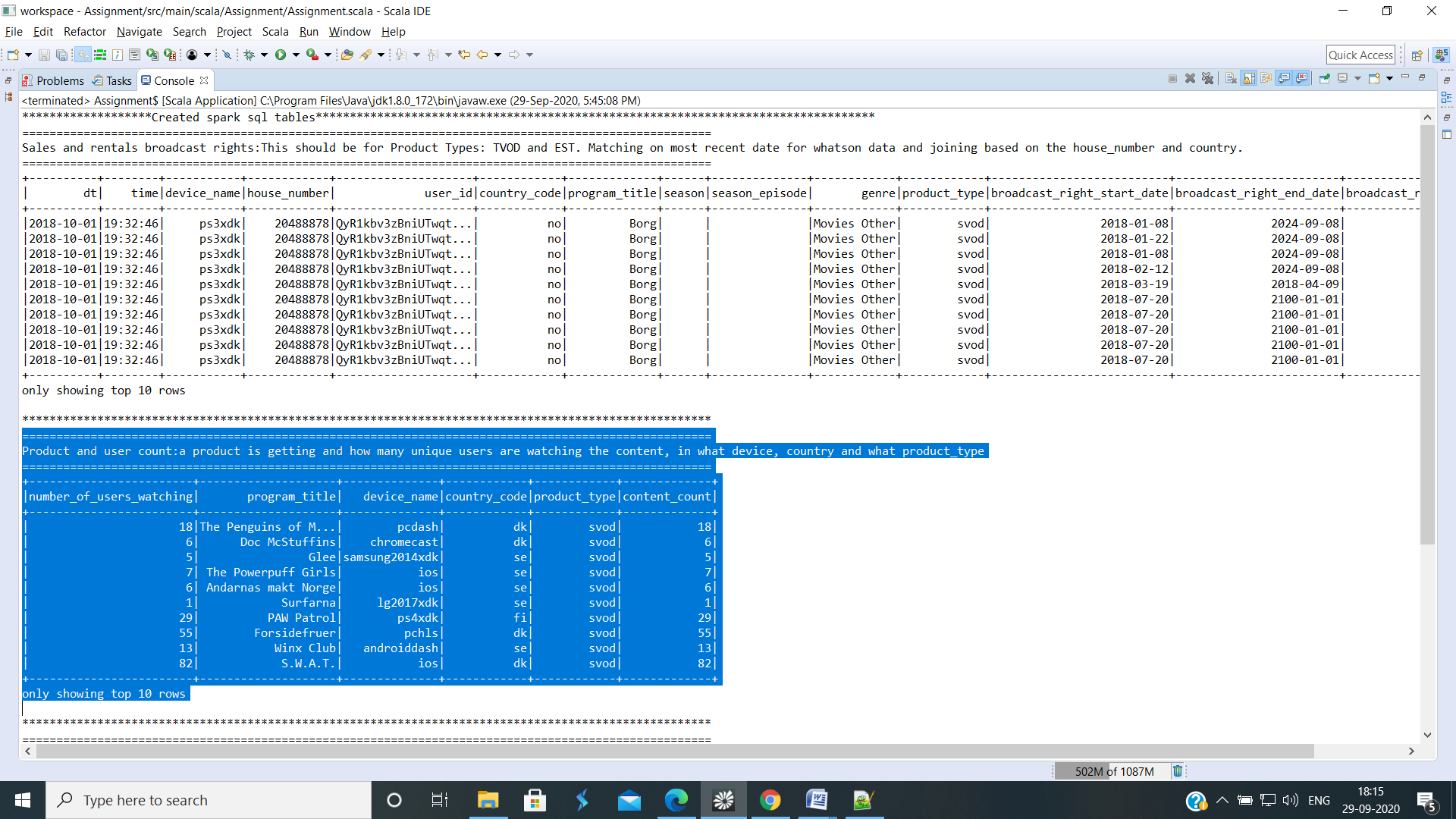
Tool: Eclipse

Code: Scala

We need to know how many watches a product is getting and how many unique users are watching the content, in what device, country and what product\_type .

* Initialize the case class with datatype of string given for all columns to started\_streams dataset and read into schema RDD string then change to Data Frame.
* Using spark.sql import can possible to write sql transformations in scala and retrieve the data for how many users watch the which content

Output Screen:



**Task 3: Genre and time of day:**

Input Datasets: started\_streams.csv

Tool: Eclipse

Code: Scala

We need a list with the most popular Genre and what hours people watch?

* Initialize the case class with datatype of string given for all columns to started\_streams dataset and read into schema RDD string then change to Data Frame.
* Using spark.sql import can possible to write sql transformations in scala and retrieve the data for which genre watch most hours.

Output Screen:

