**DATA WAREHOUSING AND DATA MINING**

**LAB MODELS**

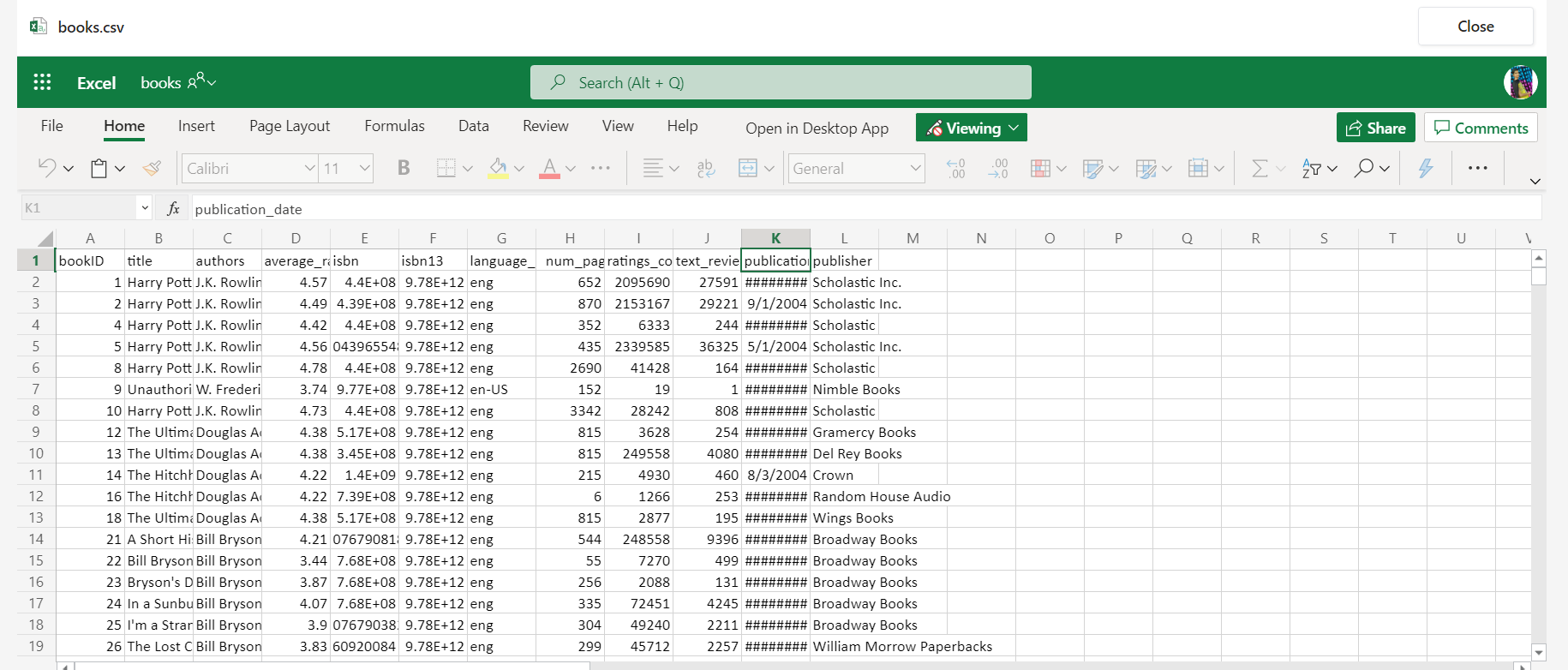
**NAME: ANUVRINDA R**

**ROLL NO:18BCS077**

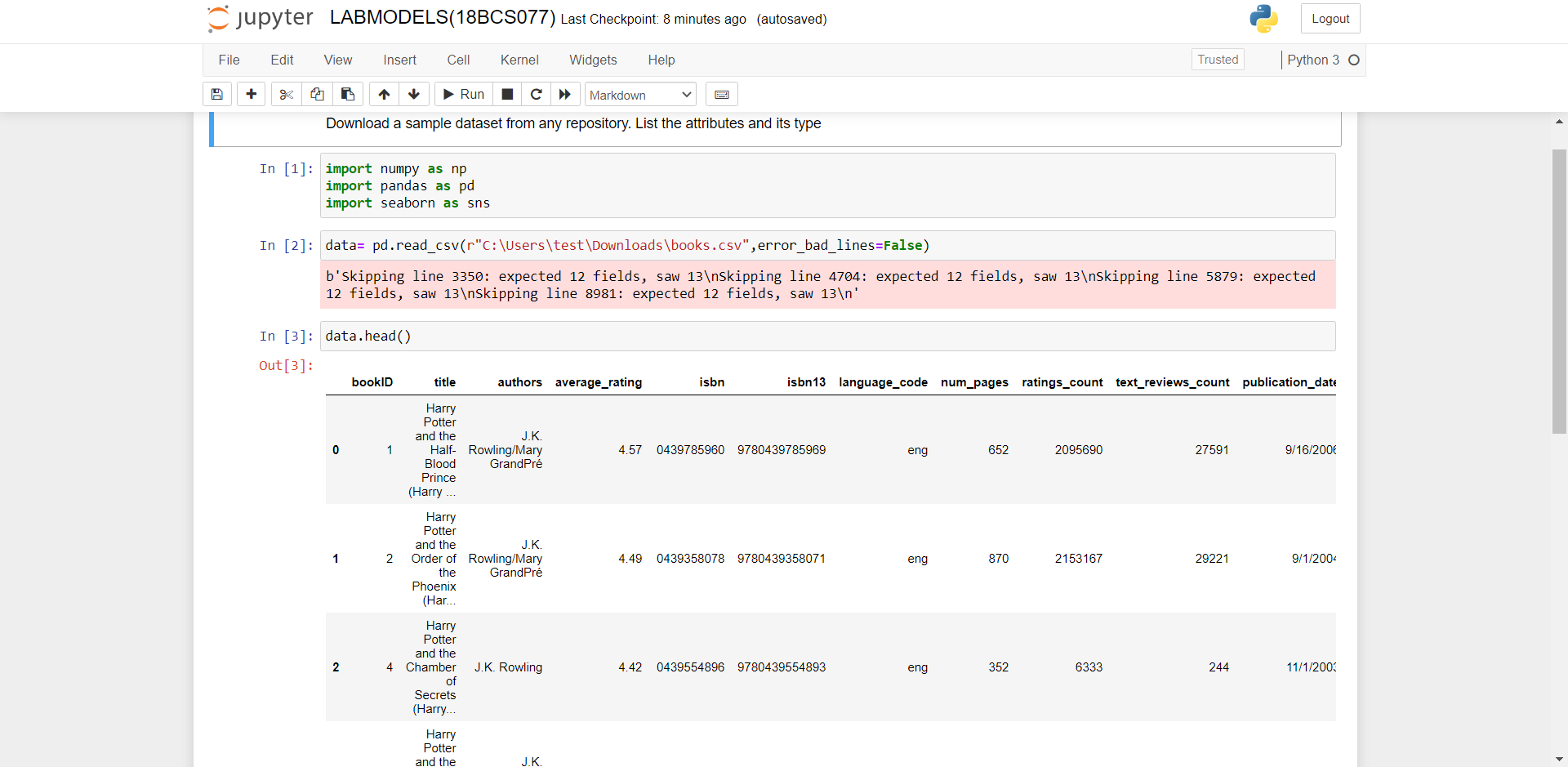
**DATE:10.04.2021**

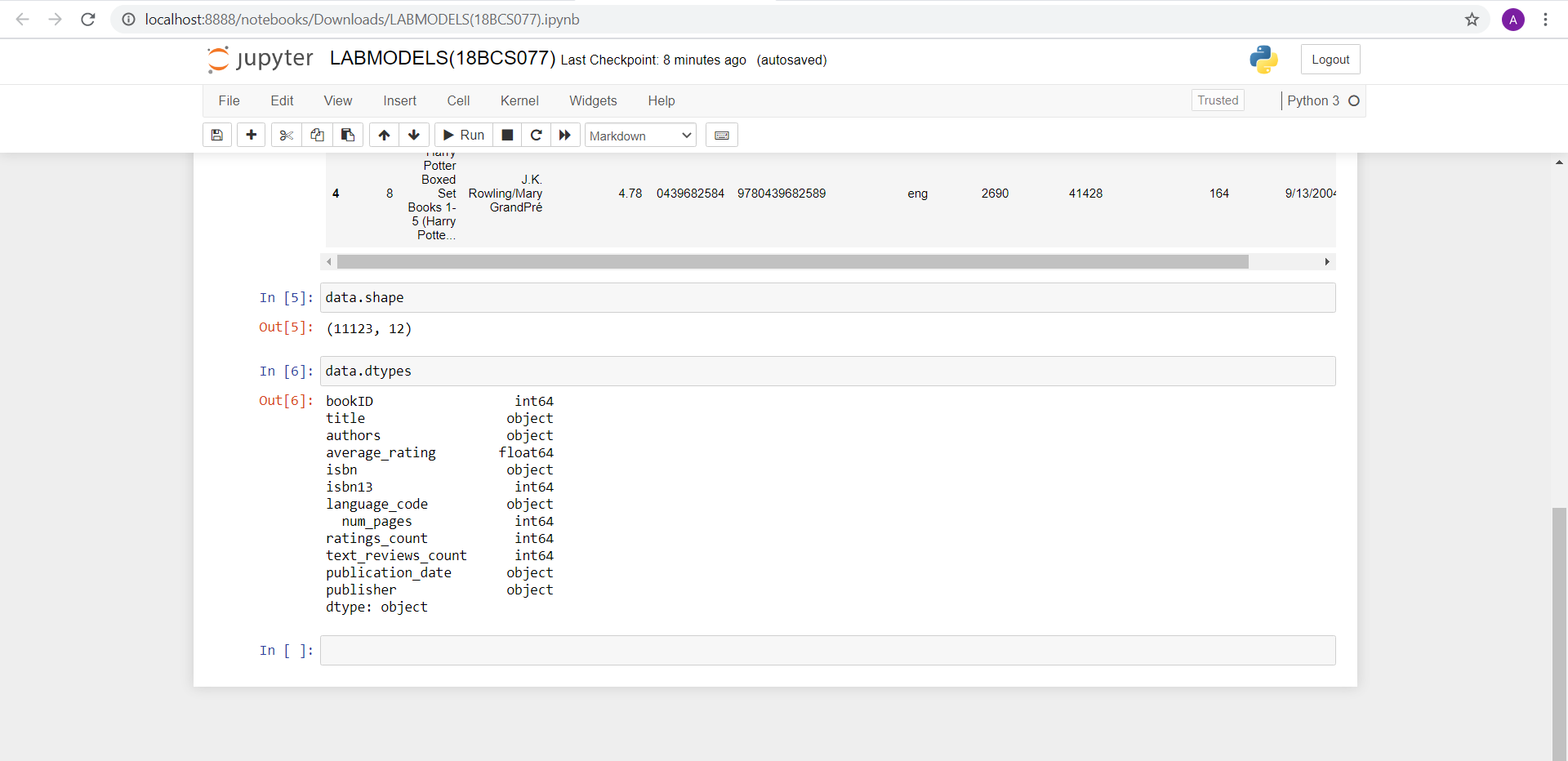
**1.Download a sample dataset from any repository. List the attributes and its type.**

**DATASET:**



**LOADING THE DATASET:**

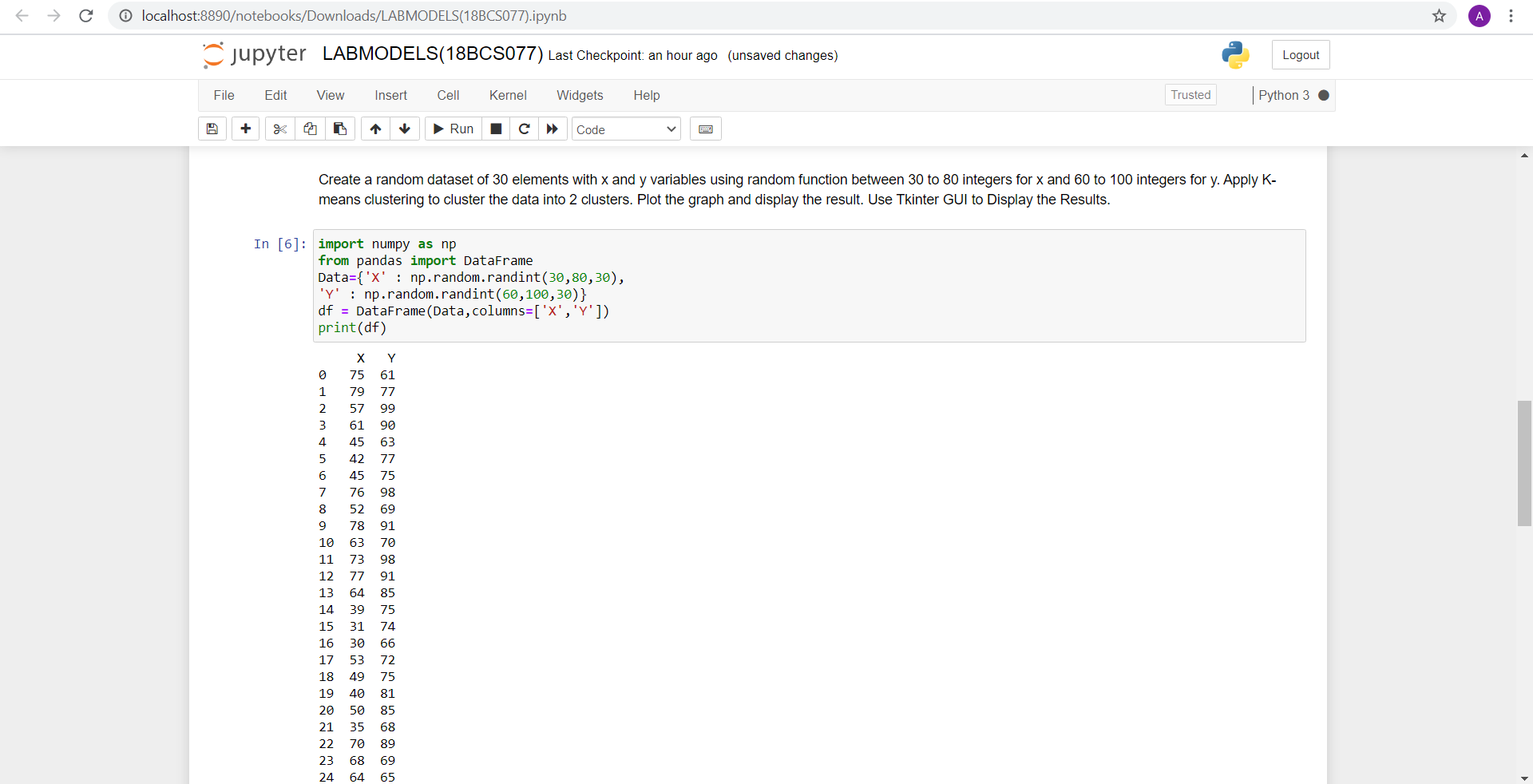


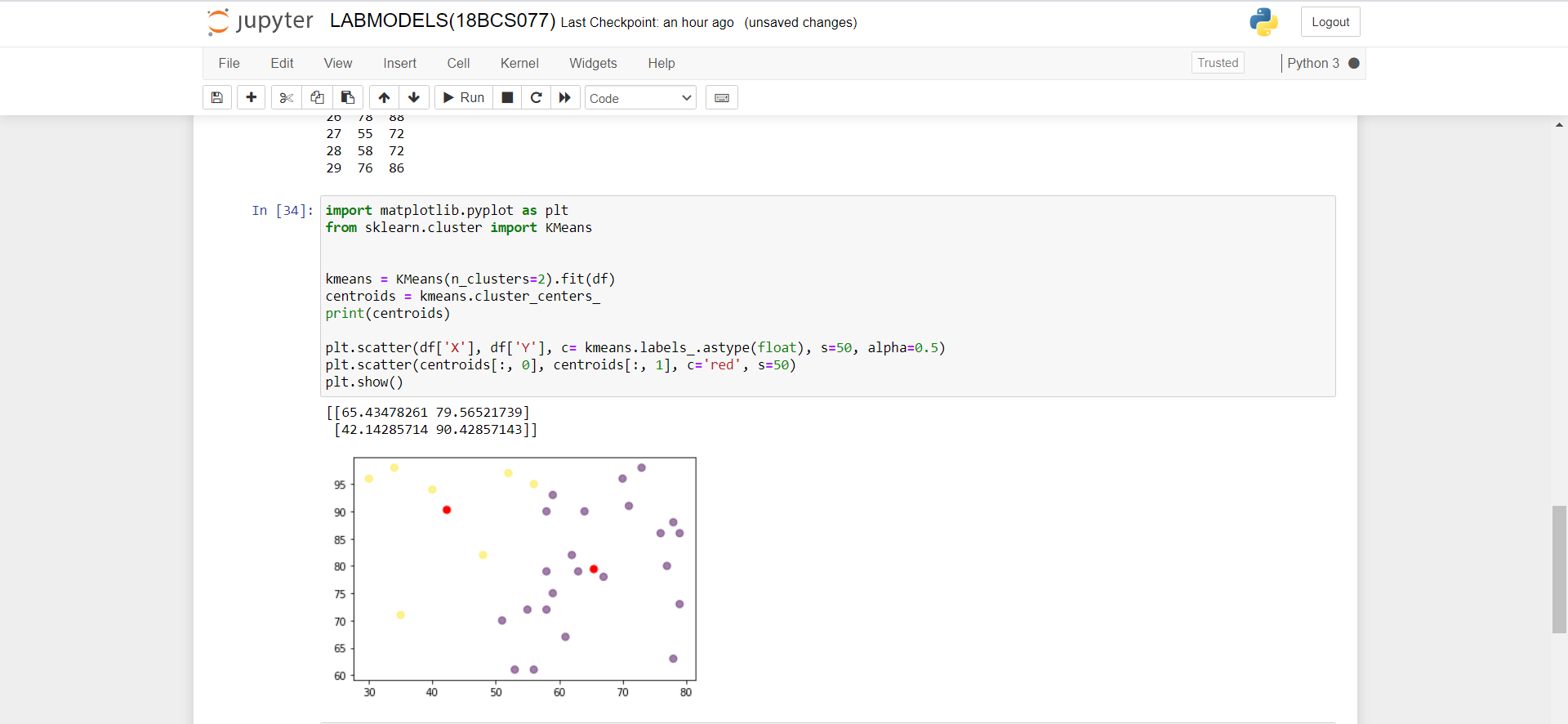


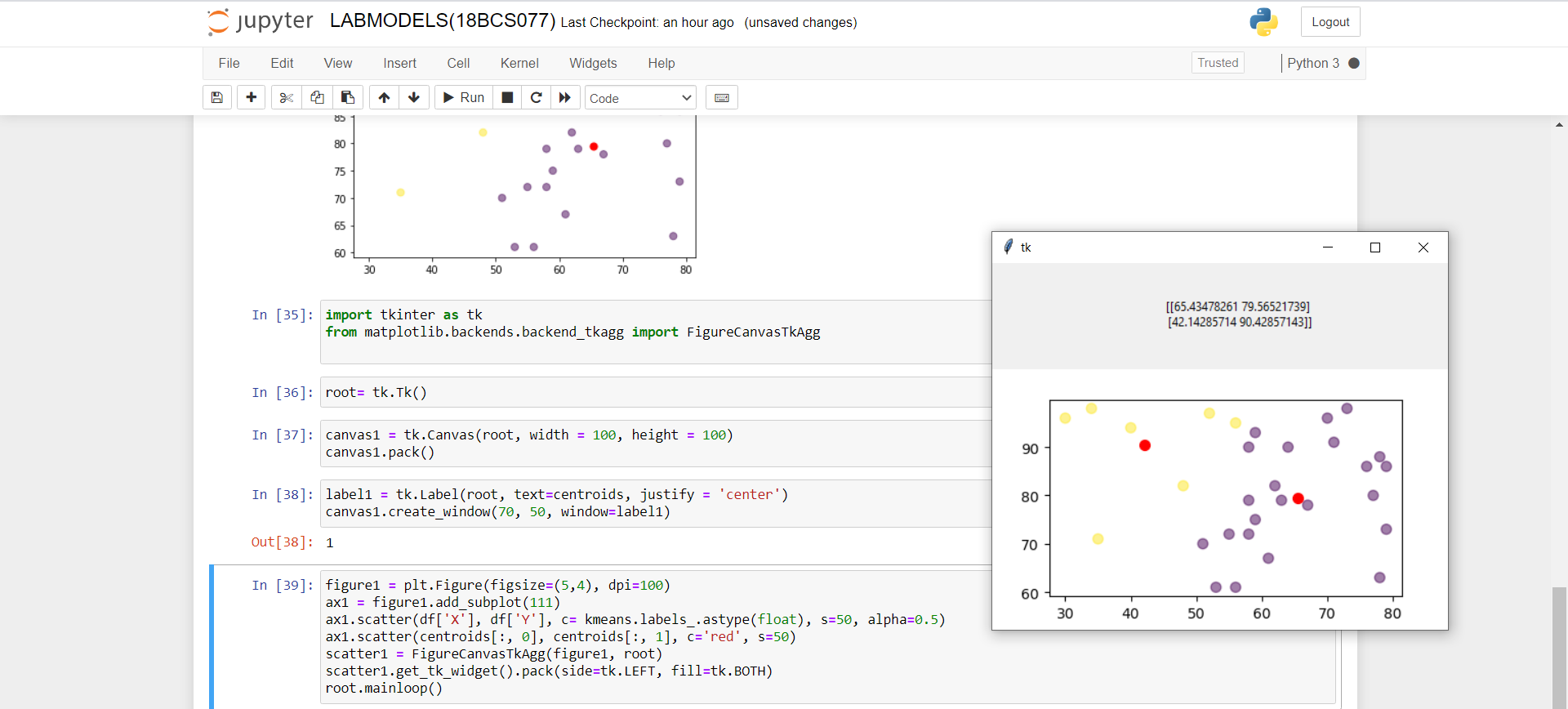
**ATTRIBUTES DESCRIPTION:**

|  |  |  |
| --- | --- | --- |
| ATTRIBUTE | TYPE | DESCRIPTION |
| **bookID** | Numerical | Contains the unique ID for each book/series |
| **title** | Categorical-Nominal | Contains the titles of the books |
| **authors** | Categorical-Nominal | Contains the author of the particular book |
| **average\_rating** | Numerical-Continuous | The average rating of the books, as decided by the users |
| **ISBN** | Numerical-Discrete | ISBN(10) number, tells the information about a book - such as edition and publisher |
| **ISBN 13** | Numerical-Discrete | The new format for ISBN, implemented in 2007. 13 digits |
| **language\_code** | Categorical-Nominal | Tells the language for the books |
| **Num\_pages** | Numerical-Discrete | Contains the number of pages for the book |
| **Ratings\_count** | Numerical-Discrete | Contains the number of ratings given for the book |
| **text\_reviews\_count** | Numerical-Discrete | Has the count of reviews left by users |
| **publication\_date** | Categorical-Nominal | Contains the publication dates of the books |
| **publisher** | Categorical-Nominal | Contains the publisher names of the books |

2.Create a random dataset of 30 elements with x and y variables using random function between 30 to 80 integers for x and 60 to 100 integers for y. Apply K- means clustering to cluster the data into 2 clusters. Plot the graph and display the result. Use Tkinter GUI to Display the Results.







3.Upload in your github account. Provide the link for access.

<https://github.com/Anuvrinda/DWDM-labmodels>