Pyspark Assignment 2

1. Create a program to store json dataset and perform following:

# Read JSON file into dataframe

df = spark.read.json("sample\_books.json")

df.printSchema()

1. Count number of records

2. Show title,price of the book.

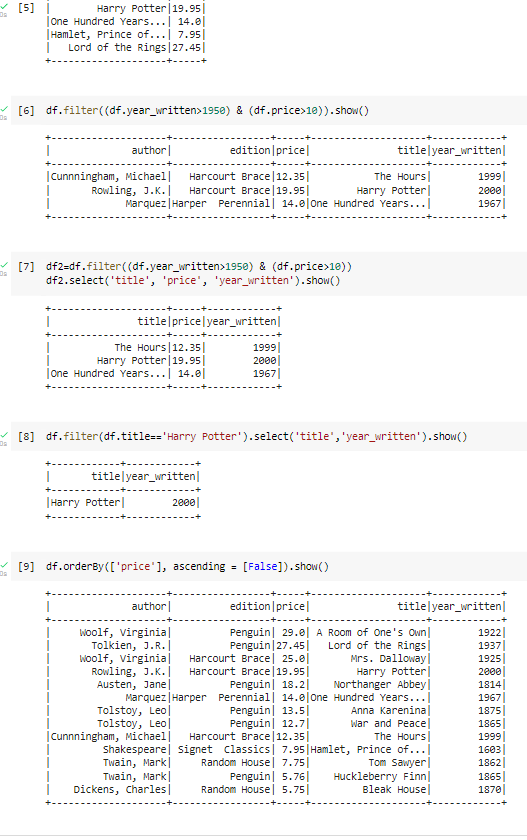
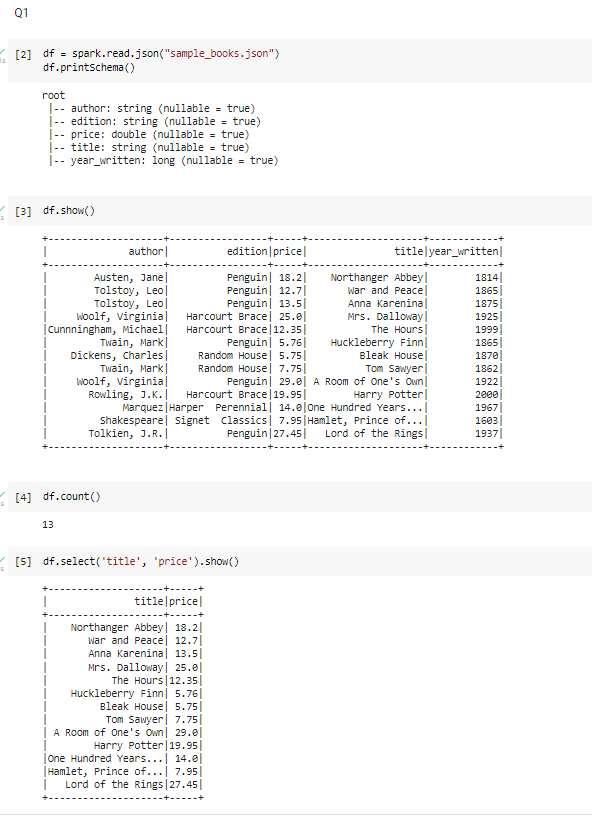
3. Show the records books which have been written after 1950 and price of book is greater than 10.

4. Show the title ,price and year of books of question 3

5. Count the number of question 4 records.

6. Show the title year written of 'Harry Potter' Book.

7. Find the costliest book.



2. for mtcars.csv file:

1. Find the top five records of mpg attribute.

2. Find out the records where the value of mpg is less than 18.

3. Add a new column to named as "wtTon" to dataframe with values of weight

column multiplied by 0.45.

4. Compute the average weight of cars by their cylinders.

5. Also sort the output of q.4

6. Write a sql query to find gears where the value of cylinder lies

between 4 and 9.

