Data Science

Ruichun Liu

Problem Set 4

1. **Data sources**

   1.1 Housing price in my hometown.

   1.2 Country music songs in recent years.

   1.3 Top 5 Econ journal papers within 5 years.

2. **Question 5**

   5.d Check what type of object mydf is. What type of an object is mydf$players?
   - class(mydf): "list".
   - class(mydf$players):"data.frame".

3. **Question 6**

   6.4 Verify that the two dataframe are different types: type class(df1) and class(df). What is the class of each?
   - class(df1): "data.frame".
   - class(df):"SparkDataFrame".

   6.5.d If you try the previous command on df1, what happens? Why do you think that happened?
   - There is an error message. The reason might be that they are different types of files. The df1 is data.frame file, but the df is SparkDataFrame.

   6.6.b If you again try the previous command on df1, what happens? Why do you think that happened?
   - There is an error message. The reason might be that they are different types of files. The df1 is data.frame file, but the df is SparkDataFrame.