Practice Set 4

Grant Naberhaus

February 11, 2020

1 Data I'm Interested In

The data I would be most interested in is NBA data. I know nba.stats.com has a lot of data available about the current and past seasons of the NBA. Additionally, for my capstone, being able to scrap for home pricing data (especially in Norman) would be very beneficial. Outside of these, there are a few stocks I would be interested in obtaining data for.

2 Problem 5

(d) class(mydf(dollarsign(players))): data.frame (e) head(mydf(dollarsign)players): Mike Adams, Lorenzo Alexander, Tyson Alualu, Danny Amendola, Derek Anderson

3 Problem 6

(4) class (df1) "data.frame" Class(df) "SparkDataFrame"

(5) Sepallength species: setosa 5.1 4.9 4.7 4.6 5.0 5,4

(b) An error message occurs; Df1 is a dataframe, not a sparkdataframe

(6) sepallength sepalwidth petallength petalwidth species 5.8 4.0 1.2 0.2 setosa 5.7 4.4 1.5 0.4 setosa 5.7 3.8 1.7 0.3 setosa 7.0 3.2 4.7 1.4 versicolor 6.4 3.2 4.5 1.5 versicolor 6.9 3.1 4.9 1.5 versicolor

(b) an error message occurs; as above, df1 is a dataframe, not a sparkdataframe

(7) df (pipe) select ("SepalLength", "Species") (pipe) filter(df(dollar)SepalLength (greaterthan) 5.5) I could not figure out the pipe operator in SparkR