Create a short document (1-2 pages) in your github describing the data wrangling steps that you undertook to clean your capstone project data set. What kind of cleaning steps did you perform? How did you deal with missing values, if any? Were there outliers, and how did you decide to handle them? This document will eventually become part of your milestone report.

1. Imported following packages

html from lxml

requests

pandas

1. Need to make dynamic web link , so it can pass different year in which different week with football season to collect data for 4 year, 16 regular game, 4 week of postseason and 4 week of preseason.
2. Created LIST variables for year , season, seasonname, week, url, columnname and DICT weekname.

year = [2013,2014,2015,2016,2017]

season=[1,2,3]

seasonname = ['Regular','Playoffs','Preseason']

week = [1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17]

weekname ={'Regular' : ['wk1','wk2','wk3','wk4','wk5','wk6','wk7','wk8','wk9','wk10','wk11','wk12','wk13','wk14','wk15','wk16','wk17']

,'Playoffs':['Wild Card Playoffs','Divisional Championships','Conference Championships','Pro Bowl Week','Super Bowl']

,'Preseason' : ['Hall of Fame Game','Preseason Week 1','Preseason Week 2','Preseason Week 3','Preseason Week 4']

}

#week = [1,2,3]

url=[]

COLUMN\_NAMES =['Player','Injury','Game Status','season','seasonType','week']

1. Created nested for loop to pass year , season and week to create url .
2. Read each webpage and stored in a list of dataframe with team index and player information.
3. Extracted data frames from list.
4. Passed this list to for loop with length of list to extract data frame.
5. Added fields to data frame to identify year, season and week.
6. Appended these data frames to main data frame with ignore\_index= True.
7. Also, within the nested for loop extracted node for team names.
8. Assigned Team Index to team name to complete the data frame such as way that each row in data frame represent player name with team, type of injury and week of injury.
9. Saves final data frame to csv file using to\_csv.