Out[3]:

	date	number_of_game	day_of_week	v_name	v_league	v_game_number	h_name	h_league	h_game_number	v_score	...	h_player_7_name	h_player_
0	18710504	0	Thu	CL1	na	1	FW1	na	1	0	...	Ed Mincher	
1	18710505	0	Fri	BS1	na	1	WS3	na	1	20	...	Asa Brainard	
2	18710506	0	Sat	CL1	na	2	RC1	na	1	12	...	Pony Sager	
3	18710508	0	Mon	CL1	na	3	CH1	na	1	12	...	Ed Duffy	
4	18710509	0	Tue	BS1	na	2	TRO	na	1	9	...	Steve Bellan	

5 rows × 161 columns

We have game_logs.csv as example. This log file has 161 columns. (First 5 rows shown)

Write some code to do the following:

1. Read this file into a Dataframe df
2. Use two different way to modify this dataframe, letting it show only the first 19 columns (1 to the 19)
3. Remove the column “number_of_game”
4. Get first 10 rows of the column with the name ”length-minutes” and find out the avg, min and max of them.
5. Find out all the game that play on the weekends(Sat and Sun)(day_of_week)
6. Based on the data frame of question5, find out what time have the higher average attendance, night game or day game (column: “attendance”, “day_night”)

7. Using groupby to group these record by day_of_week.
8. What are the first five popular host name of these record, how many time they host the game?(h_name)
9. Use lambda expression to find out all the records that host team win. (h_score>v_score)
10. Plot a scatter plot of the first 1000 records using ‘length_minutes’ and ‘h_score’