Assignment - 1

Data Science with R

Deadline : May 31, Wednesday, 23:59 IST

Instructions :

1. You need to go through the whole assignment and attempt all problems in a single R script file.

2. Make a Github Repository ,and upload your R Script file in that Repository. We will Circulate a Google form, in which all of you need to push the link of the repository, for the grading of your assignment.

3. We have **Strict Advisory** for **Cheating, Copying & Plagiarizing** from others assignments. Anyone should not share their solutions of assignment with others. If caught, both(or the group if more than 2 are involve) will be awarded with 0 marks in this assignments, no arguments will be entertain. This will effect your ratification chances badly.

4. You can discuss problems with your friends and welcome to discuss with mentors over discord.

5. This assignments contains 25% weightage of overall project.

6. Partial Grading will be done, if your solution doesn't give desired output.

Questions :

a. Go to https://www.moneyworks4me.com/best-index/nse-stocks/top-nifty50-companies-list/ . Examine the webpage. Your goal is to extract the following information for all 50 companies comes under **Nifty50** from webpage:

1 Company Name (M.Cap)
2 CMP
3 Price Change
4 Market Cap (Cr)
5 52 Week High
6 52 Week Low

7 ROE

8 P/E

9 P/BV

10 EV/EBITDA

11 5YSales Gr(%)

12 5YProfit Gr(%)

Make the dataset with this information. Your dataset should have 50 rows and 12 co0lumns. The appearance will be as following:

	Company Name (M.Cap)	CMP	Price Change	Market Cap (Cr)	52 Week High	52 Week Low	ROE	P/E	P/BV	EV/EBITDA	5YSales Gr(%)	5YProfit Gr(%)
21	Hindustan Unilever Ltd (L)	2,652	2.00%	6,23,018	2,741 09 Dec, 22	2,101 07 Jun, 22	18.21	61.56	12.39	42.23	9.60	13.68
22	Nestle India Ltd (L)	21,590	0.55%	2,08,161	22,296 08 May, 23	16,433 17 Jun, 22	108.73	81.68	65.14	51.66	11.04	14.28
23	Tata Consumer Products Ltd (L)	789.8	1.26%	73,369	861.4 14 Sep, 22	685.0 16 Mar, 23	7.99	77.22	5.76	48.70	21.56	7.85
24	Asian Paints Ltd (L)	3,127	0.16%	2,99,960	3,590 28 Sep, 22	2,560 17 Jun, 22	22.65	73.02	18.75	45.12	13.92	9.64
25	Britannia Industries Ltd (L)	4,597	0.71%	1,10,725	4,706 25 May, 23	3,272 20 Jun, 22	49.87	47.69	31.33	37.09	9.32	11.35
26	ITC Ltd (L)	443.7	0.59%	5,51,431	444.8 26 May, 23	258.1 17 Jun, 22	24.69	29.40	8.16	20.75	7.04	7.99
27	Eicher Motors Ltd (L)	3,691	1.02%	1,00,992	3,886 01 Nov, 22	2,585 17 Jun, 22	13.90	34.64	6.73	24.35	7.92	-0.16
28	Wipro Ltd (L)	401.5	1.85%	2,20,367	488.0 03 Jun, 22	351.9 17 Apr, 23	15.38	19.42	2.84	11.83	10.68	8.95
29	Apollo Hospitals Enterprise Ltd (L)	4,610	0.88%	66,286	4,900 05 Dec, 22	3,528 07 Jun, 22	10.22	65.85	9.77	39.44	-0.65	14.27
30	UPL Ltd (L)	688.2	1.85%	51,657	807.0 05 Dec, 22	607.8 23 Jun, 22	17.99	14.46	1.92	6.69	23.17	16.57
31	Adani Enterprises Ltd (L)	2,543	0.29%	2,89,942	4,190 21 Dec, 22	1,017 03 Feb, 23	18.64	178.68	20.81	95.30	25.56	26.00
32	Sun Pharmaceutical Industries Ltd (L)	970.6	2.75%	2,32,891	1,072 30 Jan, 23	789.8 17 Jun, 22	13.93	27.48	4.16	18.66	4.31	-0.26
33	ICICI Bank Ltd (L)	950.5	1.14%	6,64,608	958.0 23 May, 23	670.4 17 Jun, 22	14.99	20.84	3.38	14.50	16.91	17.03
34	IndusInd Bank Ltd (L)	1,268	0.36%	98,403	1,276 24 May, 23	763.8 23 Jun, 22	10.20	13.32	1.81	9.07	19.86	4.42
35	Axis Bank Ltd (L)	927.0	0.65%	2,85,392	970.5 04 Jan, 23	618.1 23 Jun, 22	12.03	29.79	2.28	11.75	12.86	22.53
36	HCL Technologies Ltd (L)	1,139	2.02%	3,08,978	1,157 03 Feb, 23	875.6 29 Aug, 22	21.77	20.81	4.72	12.52	12.48	10.57
37	Bharti Airtel Ltd (L)	817.9	-0.63%	4,88,115	877.1 25 Nov, 22	629.0 14 Jul, 22	-2.84	0.00	6.16	12.73	2.55	-198.53
38	Divis Laboratories Ltd (L)	3,511	2.26%	93,213	3,977 10 Aug, 22	2,730 14 Mar, 23	27.91	51.00	7.28	33.33	17.13	21.96
39	Maruti Suzuki India Ltd (L)	9,397	1.24%	2,83,869	9,769 31 Oct, 22	7,558 07 Jun, 22	7.34	34.57	4.59	21.37	5.34	-12.08
40	Ultratech Cement Ltd (L)	7,724	1.55%	2,22,976	7,909 12 May, 23	5,158 17 Jun, 22	15.00	43.99	4.10	20.93	15.69	20.30
41	Tata Consultancy Services Ltd (L)	3,329	1.04%	12,18,098	3,575 16 Feb, 23	2,926 26 Sep, 22	46.87	28.90	13.48	19.25	12.86	11.93
42	NTPC Ltd (L)	174.4	-0.03%	1,69,110	182.8 01 Nov, 22	135.0 24 Jun, 22	12.70	9.51	1.22	7.15	8.17	10.21
43	Tech Mahindra Ltd (L)	1,115	1.57%	1,08,660	1,197 01 Jun, 22	944.1 17 Jun, 22	19.06	22.49	3.88	12.14	8.91	15.30
44	Power Grid Corporation Of India Ltd (L)	237.2	-0.27%	1,65,493	249.4 12 May, 23	186.3 26 Sep, 22	20.38	10.98	2.00	7.45	9.20	14.82
45	Adani Ports and Special Economic Zone Ltd (L)	726.9	0.26%	1,57,021	987.9 20 Sep, 22	394.9 03 Feb, 23	16.22	30.85	3.51	17.50	13.55	8.05
46	Bajaj Auto Ltd (L)	4,611	-0.71%	1,30,472	4,664 26 May, 23	3,462 30 Sep, 22	19.21	21.56	4.45	16.97	8.77	6.09
47	Bajaj Finserv Ltd (L)	1,438	0.75%	2,29,055	1,846 13 Sep, 22	1,078 01 Jul, 22	22.03	35.69	4.94	12.97	22.79	15.04

b. Now pick any 5 Stocks out of these 50 stocks. Go to them by cliciking on the stock names, showing in the provided webpage. For each of choosen 5 companies do the following:

extract the following information for May 13 to May 22 (Dates may have changed, extract data for whatever dates available). But there must be of 10 days data, of following attributes.

- Sales

- YoY Gr. Rt.

- Adj EPS

- YoY Gr. Rt.

- BVPS

- Adj Net Profit

- Cash Flow from Ops.

- Debt/CF from Ops

- Return on Equity

- Op. Profit Mgn

- Net Profit Mgn

- Debt to Equity

- Working Cap Days

- Cash Conv. Cycle

Make the data frame. This time your dataset have 14 rows and 11 columns. Example: for Reliance industries the dataset will be as following:

		Mar'13	Mar'14	Mar'15	Mar'16	Mar'17	Mar'18	Mar'19	Mar'20	Mar'21	Mar'22
1	Sales ⓘ	3,97,062	4,34,460	3,75,435	2,73,999	3,05,382	3,91,677	5,69,209	5,97,535	4,66,924	6,99,962
2	YoY Gr. Rt. %	-	9.4%	-13.6%	-27%	11.5%	28.3%	45.3%	5%	-21.9%	49.9%
3	Adj EPS ⓘ	33.1	35.3	37.9	46.1	44.2	54.3	64.8	64.7	61	86
4	YoY Gr. Rt. %	-	6.6%	7.3%	21.8%	-4.1%	22.8%	19.3%	-0.2%	-5.7%	41%
5	BVPS (₹) ⓘ	304.3	333.2	366.3	387.6	439.9	490.9	647.1	701.9	1,023.4	1,151.6
6	Adj Net Profit ⓘ	19,623	20,947	22,499	27,449	26,431	32,477	38,770	41,385	41,242	58,176
7	Cash Flow from Ops. ⓘ	36,918	43,261	34,374	38,134	49,550	71,459	42,346	94,877	26,958	1,10,654
8	Debt/CF from Ops. ⓘ	2.9	3.2	4.7	4.8	4	3.1	6.8	3.5	9.3	2.4
9	Return on Equity % ⓘ	11.2	11	10.8	12.2	10.7	11.7	11.4	9.9	7.2	7.9
10	Op. Profit Mgn % ⓘ	8.3	8.1	10	15.1	15.3	16.4	15	15	17.2	15.9
11	Net Profit Mgn % ⓘ	4.9	4.8	6	10	8.7	8.3	6.8	7	9.7	9.3
12	Debt to Equity ⓘ	0.6	0.7	0.7	0.8	0.8	0.8	0.8	0.8	0.4	0.3
13	Working Cap Days ⓘ	76	75	89	114	104	98	89	107	164	124
14	Cash Conv. Cycle ⓘ	14	4	-6	-24	-34	-41	-30	-22	-30	-27

c. A tennis match runs on the concept of: "best-of-five". The two tennis players play *at most* 5 sets. If a clear winner is determined by set number x (<5), then the rest of the $5 - x$ sets are not played. (For example, if player A plays against player B, and the win sequence is $ABAA$, the winner is clearly A, so the last set will not be played.)

Let p be the probability of player A winning a set.

1. Write an `R` function that returns the number of sets that are played (x) in one simulated tennis match. Call this function `tennis`. It should take as input, the probability of success p and should output the number of games played, x. The function should look like:

```
tennis <- function(p)
{
  ...
  return(x)
}
```

2. Repeat the simulation 1000 times for $p = 0.70$ and save the output in a vector called `matches`. Save the average number of matches, `mean(matches)` in object `ans`. So your code should look like:

```
matches <- ...
for(i in 1:1000)
{
  matches[i] <- ...
}
ans <- mean(matches)
```

d. You are on a game show, being asked to choose between three doors. One door has a car, and the other two have goats. After you choose a door, the host, Monty Hall, opens one of the other doors, which he knows has a goat behind it. Monty then asks whether you would like to switch your choice of door to the other remaining door. Do you choose to switch or not to switch? Whether you switch or not depends on which action has the largest probability of winning the car.

1. Write an R function, MontyHall() to simulate this game show for when the contestant decides to switch doors. The R function should return 1 if the contestant wins and 0 if they lose.

2. Repeat the simulation 1000 times to estimate the probability of winning if the contestant switches.

e. The website https://editorial.rottentomatoes.com/guide/best-netflix-movies-to-watch-right-now/ has the "100 best movies on Netflix" ranked by "Tomatometer". The list goes from "#1" to "#100". Scrape this website and collect the following information:

1. Ranking

2. Name of Movie

3. Tomato % score

4. Year of movie