



SYMBIOSIS INSTITUTE OF TECHNOLOGY, PUNE

Symbiosis International (Deemed University)

(Established under section 3 of the UGC Act, 1956) **Re-accredited by NAAC**

with 'A' grade (3.58/4) | Awarded Category – I by UGC

Founder: Prof. Dr. S. B. Mujumdar, M. Sc., Ph. D. (Awarded Padma Bhushan and Padma Shri by President of India)

Assignment No. 03

Subject: Data Science Lab

Name of Student

Achyut Shukla

PRN No.

20070122005

Branch

CS

Class

A1

**Academic Year &
Semester**

2023-24 _ 7th semester

Date

11th August

Title of Lab Assignment

READING AND WRITING DIFFERENT TYPES OF DATASETS

Theory:

READING AND WRITING DIFFERENT TYPES OF DATASETS

- a. Reading different types of data sets (.txt, .csv) from web and disk and writing in file in specific disk location.
- b. Reading Excel data sheet in R.
- c. Reading XML dataset in R.

Answer:

```
#Load csv data from web
```

```
csv_data_web <- read.csv("https://github.com/datablist/sample-csv-  
files/raw/main/files/organizations/organizations-100.csv")
```

```
View(csv_data_web)
```

Index	Organization Id	Name	Website	Country	Description	Founded	Industry	Number of employees
1	FAB0041d5d5d22c	Ferrell LLC	https://price.net/	Papua New Guinea	Horizontal empowering knowledgebase	1990	Plastics	
2	6A7EdEA9fADC52	McKinney, Riley and Day	http://www.hall-buchanan.info/	Finland	User-centric system-worthy leverage	2015	Glass / Ceramics / Concrete	
3	06FED1ADaE4bcC1	Hester Ltd	http://sullivan-read.com/	China	Switchable scalable moratorium	1971	Public Safety	
4	2bFC18e8a4bc42f	Holder-Sellers	https://becker.com/	Turkmenistan	De-engineered systemic artificial intelligence	2004	Automotive	
5	9eE8A6a4Eb96C24	Mayer Group	http://www.brewer.com/	Mauritius	Synchronized needs-based challenge	1991	Transportation	
6	cC757116f1C085	Henry-Thompson	http://morse.net/	Bahamas	Face-to-face well-modulated customer loyalty	1992	Primary / Secondary Education	
7	219233e8aFf18C3	Hansen-Everett	https://www.kidd.org/	Pakistan	Seamless disintermediate collaboration	2018	Publishing Industry	
8	cc93DCf81a11CD	Montosh-Mora	https://www.brooks.com/	Heard Island and McDonald Islands	Centralized attitude-oriented capability	1970	Import / Export	
9	0B4F93aA06ED03e	Carr Inc	http://ross.com/	Kuwait	Distributed impactful customer loyalty	1996	Plastics	
10	738b5aDe681C6A5	Gaines Inc	http://sandoval-hooper.com/	Uzbekistan	Multi-lateral scalable protocol	1997	Outsourcing / Offshoring	
11	AE61b8f9ebbc476	Kidd Group	http://www.lyons.com/	Bouvet Island (Bouvetoya)	Proactive foreground paradigm	2001	Primary / Secondary Education	
12	eb387D06cCd4609	Crane-Clarke	https://www.sandoval.com/	Denmark	Front-line clear-thinking encryption	2014	Food / Beverages	
13	8D0c29189C97988	Keller, Campos and Black	https://www.garner.info/	Liberia	Ameliorated directional emulation	2020	Museums / Institutions	
14	D2c91cc03CA394c	Glover-Pope	http://www.silva.biz/	United Arab Emirates	Persevering contextually-based approach	2013	Medical Practice	
15	CBAC1ea9C036f4	Pacheco-Spears	https://agular.com/	Sweden	Secured logistical synergy	1984	Maritime	
16	b5D10A1a47a8Afe	Hodge-Ayers	http://www.archer-elliott.com/	Honduras	Future-proofed radical implementation	1990	Facilities Services	
17	6819365c4De0384	Bowers, Guerra and Krause	http://www.carrillo-nicholson.com/	Uganda	De-engineered transitional strategy	1972	Primary / Secondary Education	
18	5c2E9f6ba28dF	McKenzie-Melton	http://montoya-thompson.com/	Hong Kong	Reverse-engineered heuristic alliance	1998	Investment Management / Hedge Fund / Private Equity	
19	ba179f19f7925f5	Branch-Mann	http://www.lozano.com/	Botswana	Adaptive intangible frame	1999	Architecture / Planning	
20	c1Cv98350BAd66b	Weiss and Sons	https://barrett.com/	Korea	Sharable optimal functionalities	2011	Plastics	
21	8de40Ac4e6EaCa4	Velez, Payne and Coffey	http://burton.com/	Luxembourg	Mandatory coherent synergy	1986	Wholesale	
22	Aad96a4f0385f2d	Harrell LLC	http://www.frey-rosario.com/	Guadeloupe	Reverse-engineered mission-critical moratorium	2018	Construction	
23	22aC3FFa4f0703	Eaton, Reynolds and Vargas	http://www.freeman.biz/	Monaco	Self-enabling multi-tasking process improvement	2014	Luxury Goods / Jewelry	
24	5Ec4C272bCf085c	Robbins-Cummings	http://donaldson-wilkins.com/	Belgium	Organic non-volatile hierarchy	1991	Pharmaceuticals	
25	5fD8aA88B91a000	Jenkins Inc	http://www.kirk.biz/	South Africa	Front-line systematic help-desk	2002	Insurance	

```
setwd("~/Documents/college/ds_lab/assignment_3/")
```

```
#Save csv data to disk
```

```
write.csv(csv_data_web, 'sampledata.csv')
```

```
#Load csv data from disk csv_data_local <-
```

```
read.csv("~/Documents/college/ds_lab/assignment_3/sampledata.csv")
```

```
View(csv_data_local)
```

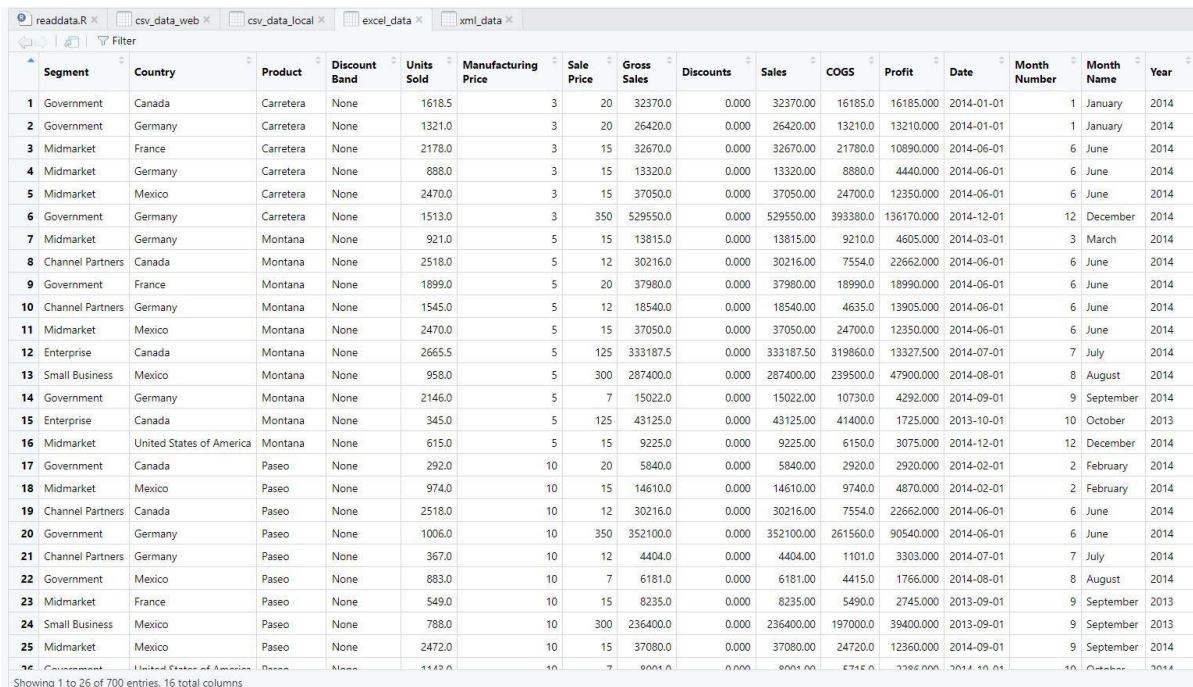
Index	Organization Id	Name	Website	Country	Description	Founded	Industry	Number of employees
1	FAB0041d5b5d22c	Ferrell LLC	https://price.net/	Papua New Guinea	Horizontal empowering knowledgebase	1990	Plastics	
2	6A7EdEA9fADC52	McKinney, Riley and Day	http://www.hall-buchanan.info/	Finland	User-centric system-worthy leverage	2015	Glass / Ceramics / Concrete	
3	06FED1ADaE4bcC1	Hester Ltd	http://sullivan-read.com/	China	Switchable scalable moratorium	1971	Public Safety	
4	2bFC18e8a4bc42f	Holder-Sellers	https://becker.com/	Turkmenistan	De-engineered systemic artificial intelligence	2004	Automotive	
5	9eE8A6a4Eb96C24	Mayer Group	http://www.brewer.com/	Mauritius	Synchronized needs-based challenge	1991	Transportation	
6	cC757116f1C085	Henry-Thompson	http://morse.net/	Bahamas	Face-to-face well-modulated customer loyalty	1992	Primary / Secondary Education	
7	219233e8aFf18C3	Hansen-Everett	https://www.kidd.org/	Pakistan	Seamless disintermediate collaboration	2018	Publishing Industry	
8	cc93DCf81a11CD	Montosh-Mora	https://www.brooks.com/	Heard Island and McDonald Islands	Centralized attitude-oriented capability	1970	Import / Export	
9	0B4F93aA06ED03e	Carr Inc	http://ross.com/	Kuwait	Distributed impactful customer loyalty	1996	Plastics	
10	738b5aDe681C6A5	Gaines Inc	http://sandoval-hooper.com/	Uzbekistan	Multi-lateral scalable protocol	1997	Outsourcing / Offshoring	
11	AE61b8f9ebbc476	Kidd Group	http://www.lyons.com/	Bouvet Island (Bouvetoya)	Proactive foreground paradigm	2001	Primary / Secondary Education	
12	eb387D06cCd4609	Crane-Clarke	https://www.sandoval.com/	Denmark	Front-line clear-thinking encryption	2014	Food / Beverages	
13	8D0c29189C97988	Keller, Campos and Black	https://www.garner.info/	Liberia	Ameliorated directional emulation	2020	Museums / Institutions	
14	D2c91cc03CA394c	Glover-Pope	http://www.silva.biz/	United Arab Emirates	Persevering contextually-based approach	2013	Medical Practice	
15	CBAC1ea9C036f4	Pacheco-Spears	https://agular.com/	Sweden	Secured logistical synergy	1984	Maritime	
16	b5D10A1a47a8Afe	Hodge-Ayers	http://www.archer-elliott.com/	Honduras	Future-proofed radical implementation	1990	Facilities Services	
17	6819365c4De0384	Bowers, Guerra and Krause	http://www.carrillo-nicholson.com/	Uganda	De-engineered transitional strategy	1972	Primary / Secondary Education	
18	5c2E9f6ba28dF	McKenzie-Melton	http://montoya-thompson.com/	Hong Kong	Reverse-engineered heuristic alliance	1998	Investment Management / Hedge Fund / Private Equity	
19	ba179f19f7925f5	Branch-Mann	http://www.lozano.com/	Botswana	Adaptive intangible frame	1999	Architecture / Planning	
20	c1Cv98350BAd66b	Weiss and Sons	https://barrett.com/	Korea	Sharable optimal functionalities	2011	Plastics	
21	8de40Ac4e6EaCa4	Velez, Payne and Coffey	http://burton.com/	Luxembourg	Mandatory coherent synergy	1986	Wholesale	
22	Aad96a4f0385f2d	Harrell LLC	http://www.frey-rosario.com/	Guadeloupe	Reverse-engineered mission-critical moratorium	2018	Construction	
23	22aC3FFa4f0703	Eaton, Reynolds and Vargas	http://www.freeman.biz/	Monaco	Self-enabling multi-tasking process improvement	2014	Luxury Goods / Jewelry	
24	5Ec4C272bCf085c	Robbins-Cummings	http://donaldson-wilkins.com/	Belgium	Organic non-volatile hierarchy	1991	Pharmaceuticals	
25	5fD8aA88B91a000	Jenkins Inc	http://www.kirk.biz/	South Africa	Front-line systematic help-desk	2002	Insurance	

```
library(readxl)
```

#Load excel data from disk

```
excel_data <- read_excel("~/Documents/college/ds_lab/assignment_3/sampledata.xlsx")
```

View(excel_data)



	Segment	Country	Product	Discount Band	Units Sold	Manufacturing Price	Sale Price	Gross Sales	Discounts	Sales	COGS	Profit	Date	Month Number	Month Name	Year
1	Government	Canada	Carretera	None	1618.5		3	20	32370.0	0.000	32370.00	16185.0	2014-01-01	1	January	2014
2	Government	Germany	Carretera	None	1321.0		3	20	26420.0	0.000	26420.00	13210.0	2014-01-01	1	January	2014
3	Midmarket	France	Carretera	None	2178.0		3	15	32670.0	0.000	32670.00	21780.0	2014-06-01	6	June	2014
4	Midmarket	Germany	Carretera	None	888.0		3	15	13320.0	0.000	13320.00	8880.0	2014-06-01	6	June	2014
5	Midmarket	Mexico	Carretera	None	2470.0		3	15	37050.0	0.000	37050.00	24700.0	2014-06-01	6	June	2014
6	Government	Germany	Carretera	None	1513.0		3	350	529550.0	0.000	529550.00	393380.0	2014-12-01	12	December	2014
7	Midmarket	Germany	Montana	None	921.0		5	15	13815.0	0.000	13815.00	9210.0	2014-03-01	3	March	2014
8	Channel Partners	Canada	Montana	None	2518.0		5	12	30216.0	0.000	30216.00	7554.0	2014-06-01	6	June	2014
9	Government	France	Montana	None	1899.0		5	20	37980.0	0.000	37980.00	18990.0	2014-06-01	6	June	2014
10	Channel Partners	Germany	Montana	None	1545.0		5	12	18540.0	0.000	18540.00	4635.0	2014-06-01	6	June	2014
11	Midmarket	Mexico	Montana	None	2470.0		5	15	37050.0	0.000	37050.00	24700.0	2014-06-01	6	June	2014
12	Enterprise	Canada	Montana	None	2665.5		5	125	333187.5	0.000	333187.50	319860.0	2014-07-01	7	July	2014
13	Small Business	Mexico	Montana	None	958.0		5	300	287400.0	0.000	287400.00	239500.0	2014-08-01	8	August	2014
14	Government	Germany	Montana	None	2146.0		5	7	15022.0	0.000	15022.00	10730.0	2014-09-01	9	September	2014
15	Enterprise	Canada	Montana	None	345.0		5	125	43125.0	0.000	43125.00	41400.0	2013-10-01	10	October	2013
16	Midmarket	United States of America	Montana	None	615.0		5	15	9225.0	0.000	9225.00	6150.0	2014-12-01	12	December	2014
17	Government	Canada	Paseo	None	292.0		10	20	5840.0	0.000	5840.00	2920.0	2014-02-01	2	February	2014
18	Midmarket	Mexico	Paseo	None	974.0		10	15	14610.0	0.000	14610.00	9740.0	2014-02-01	2	February	2014
19	Channel Partners	Canada	Paseo	None	2518.0		10	12	30216.0	0.000	30216.00	7554.0	2014-06-01	6	June	2014
20	Government	Germany	Paseo	None	1006.0		10	350	352100.0	0.000	352100.00	261560.0	2014-06-01	6	June	2014
21	Channel Partners	Germany	Paseo	None	367.0		10	12	4404.0	0.000	4404.00	1101.0	2014-07-01	7	July	2014
22	Government	Mexico	Paseo	None	883.0		10	7	6181.0	0.000	6181.00	4415.0	2014-08-01	8	August	2014
23	Midmarket	France	Paseo	None	549.0		10	15	8235.0	0.000	8235.00	5490.0	2013-09-01	9	September	2013
24	Small Business	Mexico	Paseo	None	788.0		10	300	236400.0	0.000	236400.00	197000.0	2013-09-01	9	September	2013
25	Midmarket	Mexico	Paseo	None	2472.0		10	15	37080.0	0.000	37080.00	24720.0	2014-09-01	9	September	2014
26	Government	United States of America	Paseo	None	1143.0		10	7	8001.0	0.000	8001.00	5715.0	2014-10-01	10	October	2014

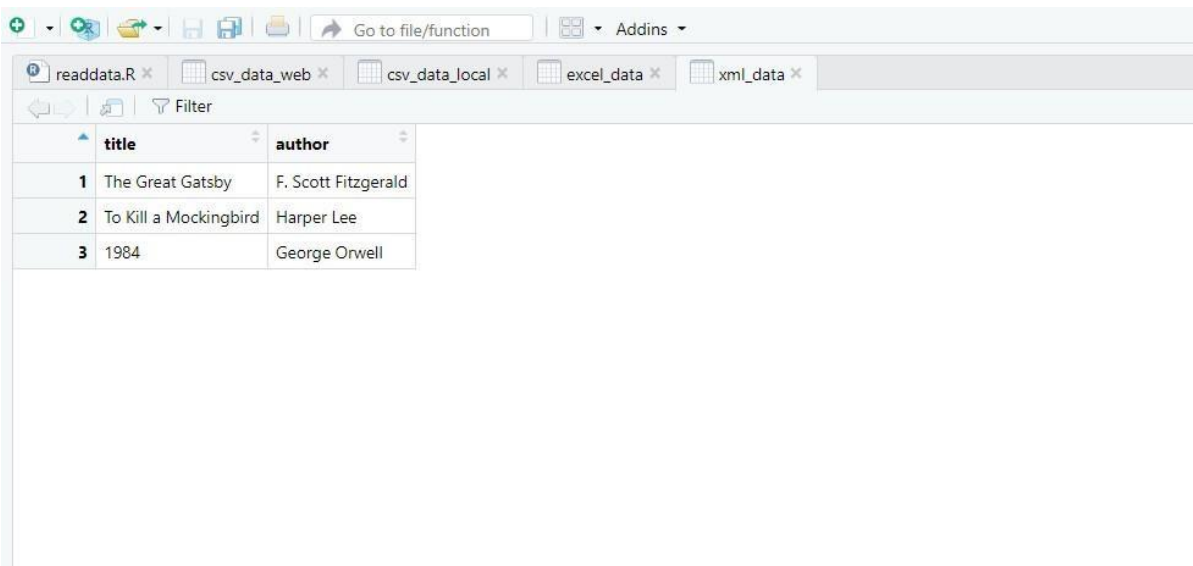
Showing 1 to 26 of 700 entries. 16 total columns

library(XML)

#Load xml data from disk

```
xml_data <- xmlToDataFrame("~/Documents/college/ds_lab/assignment_3/sampledata.xml")
```

View(xml_data)



	title	author
1	The Great Gatsby	F. Scott Fitzgerald
2	To Kill a Mockingbird	Harper Lee
3	1984	George Orwell

Conclusion: *In this lab assignment, we learn to read and write different type of data from different sources.*