

# SHETH L.U.J. AND SIR M.V. COLLEGE

## SUBJECT: Data Analysis with R

Aim: creating datasets from raw data (text files, CSV files, Excel sheets) and importing data into R

### For xls file

The screenshot displays the RStudio interface during the import of an Excel file. The 'Import excel data' dialog is open, showing the file path 'C:/Users/itlab/Downloads/HR.xlsx'. The 'Data Preview' section shows a table with columns: Age, Attrition, BusinessTravel, DailyRate, Department, DistanceFromHome, Education, EducationField, EmployeeCount, EmployeeNumber, and EnvironmentSatisfaction. The 'Import Options' section shows 'Name: HR', 'Sheet: Default', 'Range: A1:D10', and 'First Row as Names' checked. The 'Code Preview' section shows the R code: 

```
library(readxl)
HR <- read_excel("C:/Users/itlab/Downloads/HR.xlsx")
View(HR)
```

The 'Environment' pane on the right shows the data frame 'HR' with 1470 observations and 35 variables. The 'Data' pane shows the first 26 rows of the data frame.

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction
1	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	1
2	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	2
3	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	4
4	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	5
5	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	7
6	32	No	Travel_Frequently	1005	Research & Development	2	2	Life Sciences	1	8	8
7	59	No	Travel_Rarely	1324	Research & Development	3	3	Medical	1	10	10
8	30	No	Travel_Rarely	1358	Research & Development	24	1	Life Sciences	1	11	11
9	36	No	Travel_Frequently	216	Research & Development	23	3	Life Sciences	1	12	12
10	36	No	Travel_Rarely	1299	Research & Development	27	3	Medical	1	13	13
11	35	No	Travel_Rarely	809	Research & Development	16	3	Medical	1	14	14
12	29	No	Travel_Rarely	153	Research & Development	15	2	Life Sciences	1	15	15
13	31	No	Travel_Rarely	670	Research & Development	26	1	Life Sciences	1	16	16
14	34	No	Travel_Rarely	1346	Research & Development	19	2	Medical	1	18	18
15	28	Yes	Travel_Rarely	103	Research & Development	24	3	Life Sciences	1	19	19
16	29	No	Travel_Rarely	1389	Research & Development	21	4	Life Sciences	1	20	20
17	32	No	Travel_Rarely	334	Research & Development	5	2	Life Sciences	1	21	21
18	22	No	Non-Travel	1123	Research & Development	16	2	Medical	1	22	22
19	53	No	Travel_Rarely	1219	Sales	2	4	Life Sciences	1	23	23
20	36	No	Travel_Rarely	371	Research & Development	2	3	Life Sciences	1	24	24
21	24	No	Non-Travel	673	Research & Development	11	2	Other	1	26	26
22	36	Yes	Travel_Rarely	1218	Sales	9	4	Life Sciences	1	27	27
23	34	No	Travel_Rarely	419	Research & Development	7	4	Life Sciences	1	28	28
24	21	No	Travel_Rarely	391	Research & Development	15	2	Life Sciences	1	30	30
25	34	Yes	Travel_Rarely	699	Research & Development	6	1	Medical	1	31	31
26	53	No	Travel_Rarely	1282	Research & Development	5	3	Other	1	32	32

# SHETH L.U.J. AND SIR M.V. COLLEGE

## SUBJECT: Data Analysis with R

For text base

The screenshot displays the RStudio environment. The 'Import Dataset' dialog box is open, showing the 'loan\_approval' dataset. The 'Input File' field contains the path 'D:\SYCS\loan\_approval.txt'. The 'Encoding' is set to 'Automatic', 'Heading' to 'Yes', 'Row names' to 'Automatic', 'Separator' to 'Comma', 'Decimal' to 'Period', 'Quote' to 'Double', and 'Comment' to 'None'. The 'na.strings' are set to 'NA'. The 'Data Frame' preview shows the following columns: name, city, income, credit\_score, loan\_amount, years\_employed, points, and loan\_approved. The console shows the following code:

```
R - R 4.3.2 ~\bin
> loan_approval <- read.csv("D:/SYCS/loan_approval.txt")
> view(loan_approval)
> |
```

The 'Data Frame' preview shows the following data:

name	city	income	credit_score	loan_amount	years_employed	points	loan_approved
Allison Hill	East Jill	113810	389	39698	27	50	False
Brandon Hall	New Jameside	44592	729	15446	28	55	False
Rhonda Smith	Lake Roberto	33278	584	11189	13	45	False
Gabrielle Davis	West Melanview	127196	344	48823	29	50	False
Valerie Gray	Mariastad	66048	496	47174	4	25	False
Darren Roberts	Port Jesseville	62098	689	19217	29	65	True
Holly Wood	Lake Joseph	59256	373	40920	40	35	False
Nicholas Martin	Nelsonside	48289	524	45866	20	25	False
Patty Perez	Port Leslieview	126530	367	14826	36	55	False
Emily Rios	Wilkersomouth	43434	446	18359	8	20	False
Justin Baker	Hurstfurf	118696	670	15373	8	75	True
Ann Williams	East Courtneychester	127080	365	26216	24	35	False
Julie King	Lake Jenniferside	146939	573	43006	21	50	False
Jeffrey Chavez	Teresaburgh	101482	819	7973	40	100	True
Mark Lynch	West Kathryn	41395	843	1037	38	80	True
Alec Hickman	Johnbury	107397	380	6613	31	60	True
Vanessa Patel	East Donna	85302	469	18370	33	50	False
Jenny Lewis	Lake Larry	34165	682	29711	33	50	False
David Brown	New Anglashire	33905	525	1471	4	30	False
Amy Jones	Port Markhaven	42280	543	19250	31	25	False
Carlos Walls	Sanchezfort	58657	716	13478	6	65	True
Jeffrey Henderson	Shieldston	60495	400	4972	25	50	False
Rebecca Rodriguez	East Amanda	96237	443	28648	11	45	False

# SHETH L.U.J. AND SIR M.V. COLLEGE

## SUBJECT: Data Analysis with R

### From text reader

The screenshot displays the RStudio interface. The top pane shows the 'Import Text Data' dialog box with the file 'Student Stress Factors.csv' selected. The 'Data Preview' section shows the first 10 rows of the dataset. The 'Import Options' section shows the 'Name' as 'Student\_Stress\_Factors' and the 'Delimiter' as 'Comma'. The 'Code Preview' section shows the R code used to load the dataset.

**Data Preview:**

Timestamp (character)	Kindly Rate your Sleep Quality (double)	How many times a week do you suffer headaches (double)	How would you rate your academic performance (double)	how would you rate your study load? (double)	How many times a week do you suffer headaches (double)
27/10/2023 21:54:15	3	1	3	4	1
28/10/2023 12:24:40	4	1	2	3	1
28/10/2023 12:24:51	2	1	2	1	1
28/10/2023 12:26:11	3	2	3	2	2
28/10/2023 12:26:45	2	3	1	5	1
28/10/2023 12:31:02	3	1	3	2	2
28/10/2023 12:34:45	3	5	1	4	4
28/10/2023 12:35:43	4	3	1	4	4
28/10/2023 12:36:07	2	1	4	4	4
28/10/2023 12:36:20	1	2	3	2	2
28/10/2023 12:37:22	2	3	5	5	5
28/10/2023 12:38:40	3	1	5	1	1

**Import Options:**

Name: Student\_Stress\_Factors  
Skip: 0  
First Row as Names: ☒  
Trim Spaces: ☒  
Open Data Viewer: ☒  
Delimiter: Comma  
Quotes: Default  
Escape: None  
Comment: Default  
NA: Default

**Code Preview:**

```
library(readr)
Student_Stress_Factors <- read_csv("C:/Users/itlab/Desktop/Student Stress Factors.csv")
view(Student_Stress_Factors)
```

The bottom pane shows the 'Student\_Stress\_Factors' dataset loaded into the environment. The console shows the R code used to load the dataset.

**Console:**

```
R - R 4.3.2 - ~/
> specify the column types or set 'show_col_types = FALSE' to quiet this message.
> view(Student_Stress_Factors)
> |
```