

Maharaja Agrasen Institute of Technology Department of Management

Report

on

Faculty Development Program cum Workshop on "Data Analytics using R" (23rd July to 27th July 2021)

> Venue: Zoom online Time: 9.30 am to 5.00pm

Resource Person: Dr Vishal Shukla, Associate Professor School of Commerce, Presidency University, Bengaluru

Convener: Dr. Amit Gupta, HOD, MBA, MAIT

Faculty Coordinators of FDP: Dr. Neeru Gupta, Dr. Ritu Gupta, Dr. Sukhvinder Singh.





Maharaja Agrasen Institute of Technology Department of Management, MBA

DAY 1, 23rd July 2021

The Department of Management, MBA, MAIT organized a Faculty Development Program cum Workshop on "Data Analytics using R" (23rd July to 27th July 2021). The resource person was Dr. Vishal Shukla, Associate Professor, School of Commerce, Presidency University, Bengaluru. The group of participants included faculty, research scholars, corporate professionals and students from different states of India and also international participants on "Data Analytics Using R".

Inaugural:

Faculty Development Program commenced with an inaugural program and then followed by three sessions Data Analytics Using R, delivered by the Resource Person. The program started at 10:30 AM on online mode at Zoom Platform. The inaugural program was started with saraswativandana . Then the one week FDP program was introduced to the dignitaries, guests and participants by the FDP Coordinators - Dr. Neeru Gupta, Dr. Ritu Gupta and Dr. Sukhvinder Singh. The welcome speech was given by Director, MAIT, Prof. (Dr.) Neelam Sharma. It was followed by the speech of Dean Academics, MAIT, Prof. (Dr.) S.S. Deswal . Then the resource person for the FDP Dr. Vishal Shukla introduced the FDP Programme along with its importance and relevance in today's world to the participants. Then the event was addressed by an inspirational and motivational speech of Founder Chairman, MAIT, Dr. Nand Kishore Garg. The inaugural session concluded with the votes of thanks by Dr. Amit Gupta, HOD-MBA ,MAIT. A total of 105 participants were present in the inaugural session.





Session 1:

The resource person enlightened the participants with the understanding of R environment in session 1. Session focused on uses of R Software & how to start with software, session 1 was attended by 94 Participants and continued for duration of 90 minutes.



Session 2:

The resource person delivered the practical session on handling Data in R during session 2. Importing of resources & important directories with software, Participants attended session-2 with full zeal; session 2 was attended by 92 Participants and continued for 90 minutes.



Session 3:

The resource person delivered the training on data cleaning, coding and manipulation in R during session 3. The session 3 was attended by 89 Participants and continued for duration of 90 minutes.

The faculty, research scholars, corporate professionals and students learnt a lot through knowledgeablesessionsbyresourceperson.

cording_							
eytti/-Estadio	field Robins Robins Train Links				-	a x	
· 90 @• 8 8 8 = 1 +	Cuto Hefunden					1 Dey112 *	
Gernight - Nectors - de			Environment History	Connections Tutorial			
I B Source Sever	4 / - C	And the Manager of State	😅 🔂 🐨 Ingert Dataset + 🔏			are: [4]	
15 class(vecl) 15 str(vecl)			t + 🦓 Gistel During O data3	2571 obs. of 23 variables	a,	-	
1 is vector(vecl)	i is vector (vecl)		O data4	2571 obs. of 23 variables			Dr. Vishal Studa
19 vecdata = read.csu(fi)	le.choose(), header - TRUE)		O vecdata	10 obs. of 2 variables		1.Egg)	A DESCRIPTION OF THE OWNER OWNER OF THE OWNER OWNE OWNER OWNER OWNER OWN
20 ver15 21 22 ver2<-e(2, 3, 6, 5, 8, 9, 10, 11, 14, 16) 23 24 ver1ver2 25 ver1ver2 26 ver1ver2 27 27 28 ver1ver2 29 ver1ver2 20 ver15 21 21 21 21 21 21 21 21 21 21		i	Values logdata vec1 vec2	num [1:10] 1.609 0.693 1.099 1.5 num [1:10] 5 2 1 7 8 9 1 4 10 15 num [1:10] 2 3 6 5 8 9 10 11 14	46 2.079 16		
			Tiles Plots Package	en Hele Viewer			MAREDAINEMAN
27 atr(logdata)			20 instal @ Linsta	9		112	
28 vec1[7]			Aare	. Descration	Amount.		1000 CO. 1000 CO. 1000 CO.
30 Emple			User Library			-	santosh Kumar
satole	sample(x, size, replace - skuse	L, prob = MILL)	atured	Combine Muttermensional Arrays	14-2	20 5	Constant Manager
# sample_int [hote]	a signification a cample of the specified of	to from the elements of a using	(1) asRetation	Safe Password Entry for R. Ott. and SSH	1.1	0.0	a santosn kumar
nucle Terminal John -	wither with or within a replacement.	- E	. assetthat.	Eavy Fre and Post Assertions	0.2.1	0.0	
Conserved Society of Conserved Society (14) for the annual loss S x: (int S 2 1 7 8 9 1 4 10 15 S			 Backports 	feimplementations of Functions introduced Since 8-300	12.1	0 0	
			haiettäent	Tools for base64 encoding	15.5-2	0.0	
			12.14	Ecost C++ Header Nies	1,75.0-0	4.0	🥂 Deeptanshu 💴 🖥
18			E Billions	Edwite Operations	10-7	0.0	and the second sec
<pre>> vec1;7] [1] 1 subset(vec1, vec2 > =5) srubret(vec1, vec2 + =5) subset(vec1, vec2 + =5) subset(vec1, vec2 + =5)</pre>		101 2 8400	A simple saluation representing vectors of Ensity plata (BLOBS)	121			
		i troom	Convert Statistical Objects into Tidy Tiblicies	0.7.6	0.0	375	
			C) BERRY	Custom Bootstrap 'Sais' Themes for shirty' and 'meandown'	92.53	2.0	🗶 Dr. Neera Gu
1 1 C 1 A 1 4 10 12			U : chiten	Cache R Objects with Automatic Pruning	1.0.5	9.0	
			1.1 CAR	Call R Trons K	3.7.0		

Faculty Development Program cum Workshop on "Data Analytics using R" DAY 2, 24th July 2021

Session 1:

Session 1 of day 2 of the FDP started at 9:30 AM on Zoom Software (online mode). FDP resource person Dr. Vishal Shukla enlightened the participants with the Exploratory Data Analysis in R. Session focused on uses of R Software for calculating descriptive statistics and it continued for duration of 90 minutes.



Session 2:

Session 2 commenced at 11:30 am and the session focused on creating graphs and plots in R. The resource person stressed that R software can generate graphical pictures in very high resolution. FDP participants attended the session very enthusiastically. The session continued for 120 minutes.



Session 3:

Last session of the day i.e. Session 3 started at 2:30 pm and the session focused on Advanced Data visualization in R. The participants were very interactive during the session and were curious to learn more and more. The session was uneventful and went on for 120 minutes. The participants requested for the r codes for the previous day, which were readily shared by Dr. Vishal Shukla.



Faculty Development Program cum Workshop on "Data Analytics using R" DAY 3, 25th July 2021

Session 1:

Session 1 of Day 3 started with Testing of bar charts. Bar plots, Histograms including how breaks are affecting the histogram. Dr Vishal also discussed regarding the difference between SPSS and R and why in today's scenario R is more impactful than SPSS. He also discussed about Box Plot, concept of outliers and pie charts and how the different colors can be embedded in the charts.



Session 2:

Session 2 started with Scatter Plots to analyze the linear relationship between two variables for correlation and linear regression. Dr. Shukla explained directional and non-directional hypothesis by highlighting the surety between the relationships. He discussed about parametric and nonparametric test, concept of null hypothesis and type 1 and type 2 errors. Thereafter he started with one sample t test and discussed about Bartlett's test.

Zoom Meeting						92- -	٥	×
🦻 👄 Recording 🐽				# 3	¢≞w Q	Participants (63)		
					Q Fin	d a participant		
(m).		N	e	Kainana Sharma	80 1	lajiv Jain (Co-host, me)	æ ;	₩ 126
			0	i cuiparis situinis	DS I	ir Sukhvinder Singh (Host)	۲	8° 570
A Yash	a knoka yeo	Neba Sharma	Y Subtro Sen Gupta		04 1	r. Amit Gupta H (Co-host)		R 128
Abhinav kataria	Punam Ahlawat	Mayank Pruthi	Sangeeta Rawal	B	🙆 t	or. Ritu Gupta.M., (Co-host)		16 (al
*				bhaskar kappor	🤮 e)r Sanjeev Kumar (Co-host)	æ ;	¥ 🕬
			Ass to Linmute		e ,	AAIT Delhi (Co-host)	•	8 (C)
Manpreet Kaur	Deeptanshu Ra	Ankur Ahuja	Dr. Souvik Bane	Annu Janghu	D 10	or. Mrinal Kanti Das	3	Q (26)
2/3					DV p	ur. Vishal Shuida		÷.
					AK A	bhinav kataria	7	¥ 🖂
Jagbir Ahlawat	Mayank Aggarwal	NITIN WALIA	Akshita Sharma	Mahima Chauhan	99	ditya Ghai		AF 126
					(3)	kxhita Sharma_20bsp0179	2	16 TZA
					AW A	anzit Walia	1	<i>₩</i> 128
Soumya	Ν	Dr. Sushil Kuma	Shailika Rawat	Chandan Taneja		oju Bharti	3	ar (28).
*	🍠 Neha Dhawan					nkur Ahışa	1	<i>§</i> (26
Unmute Start Video	Security Parts	19 63 ~ 🗰 🧧	7 ~ 11 =	estions	ie le	witz MiderAll		1

Session 3:

Session 3 started with a Paired sample t test, analysis of variance (ANOVA), correlation (bivariate and partial correlations), regressions, multiple regression. He explained various assumptions like Multi collinearity, Autocorrelation, Homoscedasticity.

The entire day was very much interactive and all the participants were curious to learn more and more.

Zoom Meeting					120.1750	0 <u>2</u>	٥	×
🧭 🜰 According 👖 🖬				2	# View	Participants (43	(
						Q Find a participant		
Shubbarn Babl	Raily Jain	Mayank Aggarwal		Dr. Jagat Ivoti B		Rajiv Jain (Co-host me)	2	ş (20
	ruji vani	mayern oggarmar	1	the formation of the second se		DF Sakhvinder Singh (Host		¥ (76
<u>.</u>			of Dr.Sanjeev Kumar		- 1	Dr. Vishal Shukla (Co-host	1	• ÷
Dr. Mrinal Kanti Sur	Suman bala	Dr. Vishal Shukla	Varun Goel	Dr Sukhvinder S		Dr. Neeru Gupta (Co-host		# C0
			-	ý		Dr.Sanjeev Kumar (Co-hos	e) 👁 ;	ar rati
Amrit Walia Dr.	· ·	41 41	Ask to Unmute	. T		Ritu Gupta (Co-host)	æ.,	¥ cai
	Dr. Anju Bharti	nju Bharti dhiraj kumar pal	Mayank Pruthi An	Ankur Ahuja	1	DM Or. Mrinal Kanti Das	9	ą cań
					102	Vikram Singh (Asst. Profes	sor	Q (26
Annu Janghu Jagi		Jagbir Ahlawat Soumya				Amrit Walia	į	¥ (2)
	Jagbir Ahlawat		N 🥂			Ankur Abuja	1	% çalı
			🗶 Neha Dhawan	🕴 Dr. Andep Kuma		Annu Janghu	ļ	¥ (2)
						Apurva Shukla, MININIT Pra	¥y≊graj ,	# can
Chhavi Malhotra		Punam Ahlawat	Deepakshi vashi			C Chandan Taneja	j.	<i>ii</i> (56
	🗶 Rite Gupta			👔 Dr. Neeru Gupta		CM Chhavi Malhotra)	¥ (2)
🍨 🔷 📝 ~ Mute – Start Video	Security Partie	te 43 - 💭 🦉		🐨 🚺	-	Invite Mute All	+	

Faculty Development Program cum Workshop on "Data Analytics using R" DAY 4, 24th July 2021

Session 1:

Logistic Regression and Non – Parametric test in R

The session started with the introduction in logistic regression that a predictive modelling algorithm that is used when the Y variable is binary categorical. That is, it can take only two values like 1 or 0. The goal is to determine a mathematical equation that can be used to predict the probability of event 1. Once the equation is established, it can be used to predict the Y when only the Xs are known.

```
Sample of R codes learnt during session 1:
hsb2<- read.csv(file.choose(), header = TRUE)</pre>
attach (hsb2)
race<- as.factor(race)</pre>
schtyp<- as.factor(schtyp)</pre>
prog<- as.factor(prog)</pre>
#Univariate Normality
shapiro.test(write)
# One -Sample Median Test (NPar of One-Sample t-test)
wilcox.test(write, mu = 50)
boxplot(write) # (Graphical representation of Univariate Normality)
# Wilcoxon-Mann-Whitney Test (NPar of Intependent t-test )
wilcox.test(write~female)
wilcox.test(write, read, paired = TRUE)
#Kruskal-Wallis Test(NPar of One-Way ANOVA)
kruskal.test(write, prog )
 #Multiple Comparison/Post-hoc Test
install.packages("pgirmess")
library(pgirmess)
kruskalmc(write~prog)
```

Attendance: 62



Session 2:

The process of Cluster Analysis was discussed in session II. This session started at 12pm after the short break.

As discussed during the session, Cluster analysis being a Multivariate analysis technique used to club entities/objects with similar characteristics or attributes. Cluster analysis or clustering was the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups.

Clustering analysis was broadly used in many applications such as market research, pattern recognition, data analysis, and image processing. Clustering can also help marketers discover distinct groups in their customer base. And they can characterize their customer groups based on the purchasing patterns.

Following R codes for cluster analysis was mentioned:

```
#Cluster Analysis
attach(iris)
head(iris)
str(iris)
install.packages("cluster")
library(cluster)
hcluster<- hclust(dist(iris[, 1:2]))</pre>
plot(hcluster)
clustcut<- cutree(hcluster, 3)</pre>
rect.hclust(hcluster, k = 3, border = "red")
table(clustcut, Species)
str(iris)
hcluster<- hclust(dist(iris[, 1:2]), method = "average")</pre>
plot(hcluster)
clustcut<- cutree(hcluster, 3)</pre>
rect.hclust(hcluster, k = 3, border = "red")
table(clustcut, Species)
#K-Means Clustering
set.seed(20)
kmcluster<- kmeans(iris[, 1: 2], 3, nstart = 20)</pre>
kmcluster
kmcluster$centers
kmcluster$totss
table(kmcluster$cluster, Species)
Attendance : Total 52
```



Session 3:

The last session of the day was started after lunch break, at 3.00pm and ended at 5.10pm. It was based on a case study of telecom industry. The topic was 'Analysis of factors affecting churn of customers at the leading Telecom company at Delhi'. There were 5,000 Records, 7 factors out of which the factors were ranked according to Discriminant Analysis.

The objective of the research was to identify the variable of churning intention in telecom sector and their relative worth. Discriminant analysis was the appropriate statistical technique to be used when the dependent variable was the categorical (nominal or nonmetric) variable and the independent variable are metric variables.

```
Following R codes mentioned for session 3 :
```

```
install.packages("ggplot2")
install.packages("caTools")
install.packages("MASS")
install.packages("DiscriMiner")
telecom<- read.csv(file.choose(), header = T)</pre>
attach(telecom)
#Converting the target variable to categorical variable
telecom$churn<- as.factor(telecom$churn)</pre>
str(telecom)
str(churn)
#Baseline Churn rate
prop.table(table(telecom$churn))
#Explorartory Data Analysis
library(gqplot2)
ggplot(telecom, aes(x = churn, y = account length, fill=churn)) +
geom boxplot()
ggplot(telecom, aes(x = churn, y = number vmail messages, fill=churn)) +
```

Attendance : Total 56



Faculty Development Program cum Workshop on "Data Analytics using R" DAY 5, 27thJuly 2021

Session 1:

Session 1 of day 5 of the FDP started at 9:30 AM on Zoom Software (online mode). Session started with the welcome note by Dr. Ritu Gupta. Session was focused on how to get data from Twitter for research purpose with help of R Software. Participants were guided that how they can create a developer account on Twitter and use the same for getting data and apply data analysis techniques. Session 1 continued for duration of 90 minutes.



Session 2:

Session II of the day 5 of the FDP started at12:15 PM on Zoom Software (online mode). Session two started with query resolution of the participants. During the session participants learned to install packages and access tokens from Twitter. Twitter API was used for getting tweets and performing sentiment analysis on tweets for academic research purposes. The outcome of these researches can be taken up as case studies to help the students get insights on real world problems and develop the skills to solve these problems.

Session 3:

Session III of the day 5 started at 2:00 PM by continuing with the coding done in session 2. During this session the participants learned how to scrap data from twitter using R programming. Further they learned that how the scraped data can be used for sentiment analysis by researchers.



Valedictory Session:

A brief valedictory session was conducted at the end of the session 3. Dr. Amit Gupta, Head of Department delivered a vote of thanks to the resource person, participants and all the people who were directly and indirectly associated with this one-week Faculty Development Program cum Workshop. The session was ended with the feedback from some of the participants. Overall it was a good learning Experience.

