

sample

Amal Tom

2024-11-10

```
options(repos = c(CRAN = "https://cran.rstudio.com"))
```

HR Data Analytics

The below document consist of exploration and analysis of a HR dataset. Contains various data exploration methods,visualizations to understand the data in a better way.

Navigate to the bottom for inferences.

Loading the dataset

```
data=read.csv("E:/RCSS/SEM 3/R/hrdata.csv")
data
```

##	Name	Age	Degree	Year.of.Joining	Degree.Percentage
## 1	ARJUN R	28	BA	2006	63
## 2	SUNIL CHERUVATHOOR SUNNY	42	B.com	2015	58
## 3	ARJUN SANKAR	51	Bsc	2006	75
## 4	RINOY LAL	50	B.com M.com	2008	60
## 5	ANGEL JOSE	34	B.com	2008	71
## 6	GOKUL A MENON	51	PDC	2006	42
## 7	BINCY VARGHESE	31	B.com M.com	2006	68
## 8	DESILVA WILSON W	34	Bsc	2007	75
## 9	VINNAKOTA CHAITANYA	40	B.Tech	2007	60
## 10	SHAHEEN PARWEEN	42	B.com	2008	65
## 11	ALDUS AREECKAL	38	LLB	1988	80
## 12	ABIJITH K V	21	B.com MBA	2000	60
## 13	THANKOM PETER	38	Bsc	2009	61
## 14	SERA MATHEW MULAMOOTTIL	38	BA	2008	58
## 15	RAHUL K LAL	40	BA	2008	70
## 16	HARIKRISHNAN S	42	BA	1993	63
## 17	FAIZ V A	44	B.Tech	2000	68
## 18	ABRAHAM GEORGE	38	PDC	2009	51
## 19	RESHMA RAJEEV	32	BBA	2009	78
## 20	VISAL S UNNITHAN	28	B.com	2009	52
## 21	DRISHYA RAJAN	51	M.tech	2009	66
## 22	CLEMENCE PRASAD	24	B.com Mcom	2000	60
## 23	SAPNA CHOUDHARY	24	Bsc	2010	75
## 24	ANSA SARA SAJU	31	Bsc	2010	71

## 25	JUSTIN JOY	25	BA	2010	63
## 26	ANUSREE P	34	B.Tech	2000	57
## 27	MARIA PAUL	44	BA	2010	65
## 28	KEZIAH MARY MATHEW	44	B.Tech	2001	71
## 29	NANS RAJAN PHILIP	40	BA	2010	63
## 30	ARUN JOSE	28	B.Tech	2001	69
## 31	MITHUL MOTTY	22	BBA MBA	2001	65
## 32	S KIRAN KRISHNA	24	PMIR	2011	75
## 33	MARIA ALINA MATHEW	24	BBA	2012	80
## 34	JEFIN JOSEPH	23	Bsc Msc	2001	71
## 35	JAYASANKAR P G	42	M.tech	2002	71
## 36	MANU N MOHAN	21	BA	2013	60
## 37	KEERTHANA V	32	MSW	2002	70
## 38	GAYATHRI VIJAYANAND	28	B.com	2013	52
## 39	HAZEL MARY KURIEN	32	LLB	2003	65
## 40	SANJAY C P	25	M.com	2003	57
## 41	PRISCILLA THOMAS	23	Bsc	2014	75
## 42	NAUREEN LAILA AZEEZ	51	PMIR	2004	65
## 43	BHAVYA PRASAD	28	LLb	2004	61
## 44	JITHIN GEORGE	24	B.com	2005	66
## 45	SUJIN SAJI KURIAN	31	BA MA	2006	68
## 46	DEVANKUMAR K	28	B.com	2006	60
## 47	SAGAR M	34	Bsc	2014	65
## 48	JACOB J PYNADATH	50	B.com	2006	69
## 49	ATHUL KM	50	sslc	2014	40
## 50	MOBIN MOHAN	22	B.com	2015	68
## 51	MILI GEORGE	23	Bsc	2014	75
## 52	NIKITHA C MURALI	51	PMIR	2004	65
## 53	NAVANEETA PILLAI	28	LLb	2004	61
## 54	GEO DAVIS	24	B.com	2005	66
## 55	KARTHIK R	31	BA MA	2006	68
## 56	NISHANT SURENDRAN	28	B.com	2006	60
## 57	JOSHNA GEORGE	34	Bsc	2014	65
## 58	VINU D	50	B.com	2006	69
## 59	NAMITA ELIZABETH TERANCE	51	Bsc	2006	75
## 60	DIVYANAND L	50	B.com M.com	2008	60
##	experience	Department	Salary		
## 1	8	Marketing	18000		
## 2	22	Finance	52000		
## 3	31	Marketing	91000		
## 4	30	Finance	50000		
## 5	14	Marketing	24000		
## 6	31	HR	18000		
## 7	11	Marketing	21000		
## 8	14	Marketing	34000		
## 9	20	Marketing	30000		
## 10	22	Finance	32000		
## 11	18	marketing	28000		
## 12	1	Finance	12000		
## 13	18	IT	48000		
## 14	18	Marketing	44000		
## 15	20	Marketing	30000		
## 16	22	Marketing	32000		
## 17	24	Project	34000		

```
## 18      18      HR 28000
## 19      12 Marketing 42000
## 20       8  Finance 48000
## 21      31  Project 61000
## 22       4 Marketing 14000
## 23       4 Marketing 14000
## 24      11 Marketing 38000
## 25       5 Marketing 35000
## 26      14      IT 44000
## 27      24 Marketing 74000
## 28      24  Project 54000
## 29      20 Marketing 50000
## 30       8      IT 28000
## 31       2 Marketing 14000
## 32       4      HR 14000
## 33       4  Finance 14000
## 34       3 Marketing 13000
## 35      22  Finance 42000
## 36       1 Marketing 11000
## 37      12      HR 22000
## 38       8 Marketing 18000
## 39      12  legal 32000
## 40       5  Finance 15000
## 41       3 Marketing 23000
## 42      31      HR 41000
## 43       8      HR 18000
## 44       4 Marketing 28000
## 45      11 Marketing 21000
## 46       8 Marketing 38000
## 47      14 Marketing 24000
## 48      30 Marketing 90000
## 49      30 Marketing 14000
## 50       2  Finance 12000
## 51       3 Marketing 23000
## 52      31      HR 41000
## 53       8      HR 18000
## 54       4 Marketing 28000
## 55      11 Marketing 21000
## 56       8 Marketing 38000
## 57      14 Marketing 24000
## 58      30 Marketing 90000
## 59      31 Marketing 91000
## 60      30  Finance 50000
```

```
data1=data
```

Data Explorations:-

```
colnames(data)
```

```
## [1] "Name"      "Age"      "Degree"
```

```
## [4] "Year.of.Joining" "Degree.Percentage" "experience"
## [7] "Department"      "Salary"
```

```
age=unique(data$Age)
degree=unique(data$Degree)
degree_100=unique(data$Degree.Percentage)
experience=unique(data$experience)
department=unique(data$Department)
salary=unique(data$Salary)

age
```

```
## [1] 28 42 51 50 34 31 40 38 21 44 32 24 25 22 23
```

```
degree
```

```
## [1] "BA"          "B.com"       "Bsc"         "B.com M.com" "PDC"
## [6] "B.Tech"      "LLB"         "B.com MBA"   "BBA"         "M.tech"
## [11] "B.com Mcom"  "BBA MBA"     "PMIR"        "Bsc Msc"     "MSW"
## [16] "M.com"      "LLb"         "BA MA"       "sslC"
```

```
degree_100
```

```
## [1] 63 58 75 60 71 42 68 65 80 61 70 51 78 52 66 57 69 40
```

```
experience
```

```
## [1] 8 22 31 30 14 11 20 18 1 24 12 4 5 2 3
```

```
department
```

```
## [1] "Marketing" "Finance"   "HR"        "marketing" "IT"        "Project"
## [7] "legal"
```

```
salary
```

```
## [1] 18000 52000 91000 50000 24000 21000 34000 30000 32000 28000 12000 48000
## [13] 44000 42000 61000 14000 38000 35000 74000 54000 13000 11000 22000 15000
## [25] 23000 41000 90000
```

Finding total salary contribution from each department

```
dep_salary <- tapply(data$Salary, data$Department, sum)
dep_salary
```

```
## Finance      HR      IT      legal marketing Marketing Project
## 327000    200000    120000    32000      28000    1200000    149000
```

Merging 'marketing' with 'Marketing'

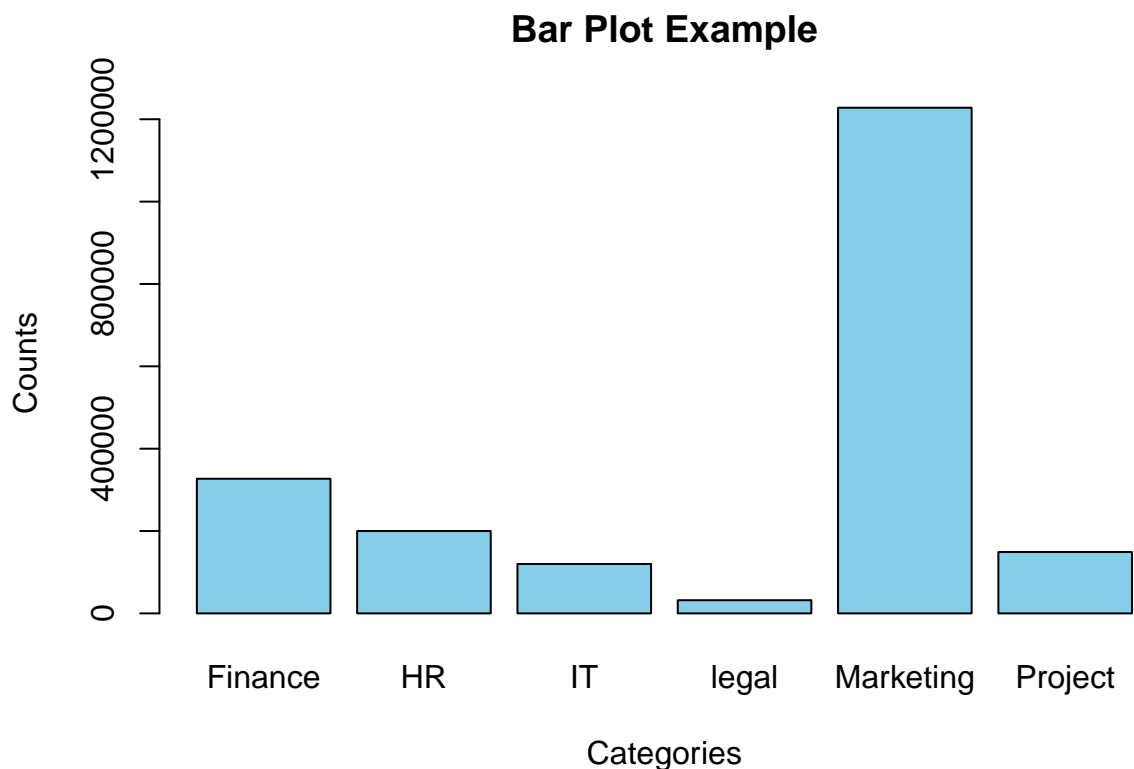
```
data1$Department[data1$Department == "marketing"] <- "Marketing"
dep_salary <- tapply(data1$Salary, data1$Department, sum)
dep_salary
```

```
##      Finance      HR      IT      legal Marketing      Project
##      327000     200000    120000     32000    1228000     149000
```

Inference 1: Most of the salary is being utilised by the Marketing department

#Representing through Bar Plot

```
values <- c(A = 10, B = 15, C = 7)
barplot(dep_salary, main = "Bar Plot Example", xlab = "Categories", ylab = "Counts", col = "skyblue")
```



Number of employees across each departments:-

```
# Count occurrences of each unique value in 'Department' column
dep_count<-table(data1$Department)
dep_count
```

```
##
##   Finance      HR      IT      legal Marketing  Project
##      10       8      3      1      35      3
```

Inference 2:Most of the employees are under the Marketing department

Average salary of employees across different branches:-

```
avg=dep_salary/dep_count
avg
```

```
##   Finance      HR      IT      legal Marketing  Project
## 32700.00 25000.00 40000.00 32000.00 35085.71 49666.67
```

Inference 3:Average salary of employees working under 'Project' department is the highest

Salary Range across different branches:-

```
salary_range_by_dept <- tapply(data1$Salary, data1$Department, range)
print(salary_range_by_dept)
```

```
## $Finance
## [1] 12000 52000
##
## $HR
## [1] 14000 41000
##
## $IT
## [1] 28000 48000
##
## $legal
## [1] 32000 32000
##
## $Marketing
## [1] 11000 91000
##
## $Project
## [1] 34000 61000
```

Inference 4: The Marketing department has the the lowest and highest allotted salary(Range:11000-19000)

Converting categorical 'degree' to numerical values-'Degree':-

```
data1$degree <- as.numeric(as.factor(data1$Degree))
Degree<-unique(data1$Degree)
data1
```

##	Name	Age	Degree	Year.of.Joining	Degree.Percentage
## 1	ARJUN R	28	BA	2006	63
## 2	SUNIL CHERUVATHOOR SUNNY	42	B.com	2015	58
## 3	ARJUN SANKAR	51	Bsc	2006	75
## 4	RINOY LAL	50	B.com M.com	2008	60
## 5	ANGEL JOSE	34	B.com	2008	71
## 6	GOKUL A MENON	51	PDC	2006	42
## 7	BINCY VARGHESE	31	B.com M.com	2006	68
## 8	DESILVA WILSON W	34	Bsc	2007	75
## 9	VINNAKOTA CHAITANYA	40	B.Tech	2007	60
## 10	SHAHEEN PARWEEN	42	B.com	2008	65
## 11	ALDUS AREECKAL	38	LLB	1988	80
## 12	ABIJITH K V	21	B.com MBA	2000	60
## 13	THANKOM PETER	38	Bsc	2009	61
## 14	SERA MATHEW MULAMOOTTIL	38	BA	2008	58
## 15	RAHUL K LAL	40	BA	2008	70
## 16	HARIKRISHNAN S	42	BA	1993	63
## 17	FAIZ V A	44	B.Tech	2000	68
## 18	ABRAHAM GEORGE	38	PDC	2009	51
## 19	RESHMA RAJEEV	32	BBA	2009	78
## 20	VISAL S UNNITHAN	28	B.com	2009	52
## 21	DRISHYA RAJAN	51	M.tech	2009	66
## 22	CLEMENCE PRASAD	24	B.com Mcom	2000	60
## 23	SAPNA CHOUDHARY	24	Bsc	2010	75
## 24	ANSA SARA SAJU	31	Bsc	2010	71
## 25	JUSTIN JOY	25	BA	2010	63
## 26	ANUSREE P	34	B.Tech	2000	57
## 27	MARIA PAUL	44	BA	2010	65
## 28	KEZIAH MARY MATHEW	44	B.Tech	2001	71
## 29	NANS RAJAN PHILIP	40	BA	2010	63
## 30	ARUN JOSE	28	B.Tech	2001	69
## 31	MITHUL MOTTY	22	BBA MBA	2001	65
## 32	S KIRAN KRISHNA	24	PMIR	2011	75
## 33	MARIA ALINA MATHEW	24	BBA	2012	80
## 34	JEFIN JOSEPH	23	Bsc Msc	2001	71
## 35	JAYASANKAR P G	42	M.tech	2002	71
## 36	MANU N MOHAN	21	BA	2013	60
## 37	KEERTHANA V	32	MSW	2002	70
## 38	GAYATHRI VIJAYANAND	28	B.com	2013	52
## 39	HAZEL MARY KURIEN	32	LLB	2003	65
## 40	SANJAY C P	25	M.com	2003	57
## 41	PRISCILLA THOMAS	23	Bsc	2014	75

## 42	NAUREEN LAILA AZEEZ	51	PMIR	2004	65
## 43	BHAVYA PRASAD	28	LLb	2004	61
## 44	JITHIN GEORGE	24	B.com	2005	66
## 45	SUJIN SAJI KURIAN	31	BA MA	2006	68
## 46	DEVANKUMAR K	28	B.com	2006	60
## 47	SAGAR M	34	Bsc	2014	65
## 48	JACOB J PYNADATH	50	B.com	2006	69
## 49	ATHUL KM	50	sslc	2014	40
## 50	MOBIN MOHAN	22	B.com	2015	68
## 51	MILI GEORGE	23	Bsc	2014	75
## 52	NIKITHA C MURALI	51	PMIR	2004	65
## 53	NAVANEETA PILLAI	28	LLb	2004	61
## 54	GEO DAVIS	24	B.com	2005	66
## 55	KARTHIK R	31	BA MA	2006	68
## 56	NISHANT SURENDRAN	28	B.com	2006	60
## 57	JOSHNA GEORGE	34	Bsc	2014	65
## 58	VINU D	50	B.com	2006	69
## 59	NAMITA ELIZABETH TERANCE	51	Bsc	2006	75
## 60	DIVYANAND L	50	B.com M.com	2008	60
##	experience	Department	Salary	degree	
## 1	8	Marketing	18000	6	
## 2	22	Finance	52000	1	
## 3	31	Marketing	91000	10	
## 4	30	Finance	50000	2	
## 5	14	Marketing	24000	1	
## 6	31	HR	18000	17	
## 7	11	Marketing	21000	2	
## 8	14	Marketing	34000	10	
## 9	20	Marketing	30000	5	
## 10	22	Finance	32000	1	
## 11	18	Marketing	28000	13	
## 12	1	Finance	12000	3	
## 13	18	IT	48000	10	
## 14	18	Marketing	44000	6	
## 15	20	Marketing	30000	6	
## 16	22	Marketing	32000	6	
## 17	24	Project	34000	5	
## 18	18	HR	28000	17	
## 19	12	Marketing	42000	8	
## 20	8	Finance	48000	1	
## 21	31	Project	61000	15	
## 22	4	Marketing	14000	4	
## 23	4	Marketing	14000	10	
## 24	11	Marketing	38000	10	
## 25	5	Marketing	35000	6	
## 26	14	IT	44000	5	
## 27	24	Marketing	74000	6	
## 28	24	Project	54000	5	
## 29	20	Marketing	50000	6	
## 30	8	IT	28000	5	
## 31	2	Marketing	14000	9	
## 32	4	HR	14000	18	
## 33	4	Finance	14000	8	
## 34	3	Marketing	13000	11	


```
## 35      22    Finance  42000    15
## 36       1  Marketing  11000     6
## 37      12      HR   22000    16
## 38       8  Marketing  18000     1
## 39      12     legal  32000    13
## 40       5    Finance  15000    14
## 41       3  Marketing  23000    10
## 42      31      HR   41000    18
## 43       8      HR   18000    12
## 44       4  Marketing  28000     1
## 45      11  Marketing  21000     7
## 46       8  Marketing  38000     1
## 47      14  Marketing  24000    10
## 48      30  Marketing  90000     1
## 49      30  Marketing  14000    19
## 50       2    Finance  12000     1
## 51       3  Marketing  23000    10
## 52      31      HR   41000    18
## 53       8      HR   18000    12
## 54       4  Marketing  28000     1
## 55      11  Marketing  21000     7
## 56       8  Marketing  38000     1
## 57      14  Marketing  24000    10
## 58      30  Marketing  90000     1
## 59      31  Marketing  91000    10
## 60      30    Finance  50000     2
```

,

Creating correlation matrix, to find dependency between attributes:-

```
cor_matrix <- cor(data1[, c("Salary","experience","Degree.Percentage","Year.of.Joining","Age","degree")])
print(cor_matrix)
```

```
##           Salary  experience Degree.Percentage Year.of.Joining
## Salary      1.00000000  0.68841221    0.1526392716   -0.01002601
## experience   0.68841221  1.00000000   -0.1503898050   -0.07942463
## Degree.Percentage 0.15263927 -0.15038981    1.0000000000   -0.12811997
## Year.of.Joining -0.01002601 -0.07942463   -0.1281199682    1.00000000
## Age          0.68841221  1.00000000   -0.1503898050   -0.07942463
## degree      -0.18052219  0.17938637   -0.0006814493   -0.06892154
##           Age      degree
## Salary      0.68841221 -0.1805221895
## experience   1.00000000  0.1793863659
## Degree.Percentage -0.15038981 -0.0006814493
## Year.of.Joining -0.07942463 -0.0689215366
## Age          1.00000000  0.1793863659
## degree      0.17938637  1.0000000000
```

Heatmap representation of the correlation matrix:-

```
install.packages("corrplot")
```

```
## Installing package into 'C:/Users/i/AppData/Local/R/win-library/4.4'  
## (as 'lib' is unspecified)
```

```
## package 'corrplot' successfully unpacked and MD5 sums checked  
##
```

```
## The downloaded binary packages are in  
## C:\Users\i\AppData\Local\Temp\RtmpIFN07e\downloaded_packages
```

```
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.4.2
```

```
## corrplot 0.95 loaded
```

```
corrplot(cor_matrix,  
  method = "color",  
  addCoef.col = "black",  
  number.cex = 0.7,  
  tl.cex = 0.8,  
  main = "Correlation Matrix Heatmap (corrplot)")
```



Inference 5: Experience has a greater influence on salary compared with other attributes.

Finding the age range and number of employees across different departments:-

```
library(dplyr)

## Warning: package 'dplyr' was built under R version 4.4.2

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
age_range_by_dept <- data1 %>%
  group_by(Department) %>%
  summarise(
    min_age = min(Age, na.rm = TRUE),
    max_age = max(Age, na.rm = TRUE),
    mean_age = mean(Age, na.rm = TRUE),
    median_age = median(Age, na.rm = TRUE),
    employee_count = n()
  )

print(age_range_by_dept)
```

```
## # A tibble: 6 x 6
##   Department min_age max_age mean_age median_age employee_count
##   <chr>      <int>   <int>   <dbl>     <dbl>         <int>
## 1 Finance         21     50    34.6       35            10
## 2 HR              24     51    37.9       35             8
## 3 IT              28     38    33.3       34             3
## 4 Marketing       21     51    33.5       31            35
## 5 Project        44     51    46.3       44             3
## 6 legal          32     32     32        32             1
```

Inference 6:The employees under Project department are very much senior to the other department employees

```
project_legal_details <- data1 %>%
  filter(Department %in% c("Project", "legal", "IT"))
print(project_legal_details)
```

##	Name	Age	Degree	Year.of.Joining	Degree.Percentage	experience
## 1	THANKOM PETER	38	Bsc	2009	61	18
## 2	FAIZ V A	44	B.Tech	2000	68	24
## 3	DRISHYA RAJAN	51	M.tech	2009	66	31
## 4	ANUSREE P	34	B.Tech	2000	57	14
## 5	KEZIAH MARY MATHEW	44	B.Tech	2001	71	24
## 6	ARUN JOSE	28	B.Tech	2001	69	8
## 7	HAZEL MARY KURIEN	32	LLB	2003	65	12

##	Department	Salary	degree
## 1	IT	48000	10
## 2	Project	34000	5
## 3	Project	61000	15
## 4	IT	44000	5
## 5	Project	54000	5
## 6	IT	28000	5
## 7	legal	32000	13

Inference 7:- people working under the departments-Project,Legal and IT contributes lesser number of employees.

```
salary_intervals <- cut(data1$Salary,
                        breaks = seq(11000, 91000, by = 10000),
                        include.lowest = TRUE,
                        right = FALSE, # If you want intervals like [18000, 29000)
                        labels = c("11k-21k", "21k-31k", "31k-41k", "41k-51k", "51k-61k", "61k-71k", "71k-81k"))
data1$SalaryInterval <- salary_intervals
data1
```

##	Name	Age	Degree	Year.of.Joining	Degree.Percentage
## 1	ARJUN R	28	BA	2006	63
## 2	SUNIL CHERUVATHOOR SUNNY	42	B.com	2015	58
## 3	ARJUN SANKAR	51	Bsc	2006	75
## 4	RINOY LAL	50	B.com M.com	2008	60
## 5	ANGEL JOSE	34	B.com	2008	71
## 6	GOKUL A MENON	51	PDC	2006	42
## 7	BINCY VARGHESE	31	B.com M.com	2006	68
## 8	DESILVA WILSON W	34	Bsc	2007	75
## 9	VINNAKOTA CHAITANYA	40	B.Tech	2007	60
## 10	SHAHEEN PARWEEN	42	B.com	2008	65
## 11	ALDUS AREECKAL	38	LLB	1988	80
## 12	ABIJITH K V	21	B.com MBA	2000	60
## 13	THANKOM PETER	38	Bsc	2009	61
## 14	SERA MATHEW MULAMOOTTIL	38	BA	2008	58
## 15	RAHUL K LAL	40	BA	2008	70
## 16	HARIKRISHNAN S	42	BA	1993	63
## 17	FAIZ V A	44	B.Tech	2000	68
## 18	ABRAHAM GEORGE	38	PDC	2009	51
## 19	RESHMA RAJEEV	32	BBA	2009	78
## 20	VISAL S UNNITHAN	28	B.com	2009	52
## 21	DRISHYA RAJAN	51	M.tech	2009	66
## 22	CLEMENCE PRASAD	24	B.com Mcom	2000	60

## 23	SAPNA CHOUDHARY	24	Bsc	2010	75
## 24	ANSA SARA SAJU	31	Bsc	2010	71
## 25	JUSTIN JOY	25	BA	2010	63
## 26	ANUSREE P	34	B.Tech	2000	57
## 27	MARIA PAUL	44	BA	2010	65
## 28	KEZIAH MARY MATHEW	44	B.Tech	2001	71
## 29	NANS RAJAN PHILIP	40	BA	2010	63
## 30	ARUN JOSE	28	B.Tech	2001	69
## 31	MITHUL MOTTY	22	BBA MBA	2001	65
## 32	S KIRAN KRISHNA	24	PMIR	2011	75
## 33	MARIA ALINA MATHEW	24	BBA	2012	80
## 34	JEFIN JOSEPH	23	Bsc Msc	2001	71
## 35	JAYASANKAR P G	42	M.tech	2002	71
## 36	MANU N MOHAN	21	BA	2013	60
## 37	KEERTHANA V	32	MSW	2002	70
## 38	GAYATHRI VIJAYANAND	28	B.com	2013	52
## 39	HAZEL MARY KURIEN	32	LLB	2003	65
## 40	SANJAY C P	25	M.com	2003	57
## 41	PRISCILLA THOMAS	23	Bsc	2014	75
## 42	NAUREEN LAILA AZEEZ	51	PMIR	2004	65
## 43	BHAVYA PRASAD	28	LLb	2004	61
## 44	JITHIN GEORGE	24	B.com	2005	66
## 45	SUJIN SAJI KURIAN	31	BA MA	2006	68
## 46	DEVANKUMAR K	28	B.com	2006	60
## 47	SAGAR M	34	Bsc	2014	65
## 48	JACOB J PYNADATH	50	B.com	2006	69
## 49	ATHUL KM	50	sslc	2014	40
## 50	MOBIN MOHAN	22	B.com	2015	68
## 51	MILI GEORGE	23	Bsc	2014	75
## 52	NIKITHA C MURALI	51	PMIR	2004	65
## 53	NAVANEETA PILLAI	28	LLb	2004	61
## 54	GEO DAVIS	24	B.com	2005	66
## 55	KARTHIK R	31	BA MA	2006	68
## 56	NISHANT SURENDRAN	28	B.com	2006	60
## 57	JOSHNA GEORGE	34	Bsc	2014	65
## 58	VINU D	50	B.com	2006	69
## 59	NAMITA ELIZABETH TERANCE	51	Bsc	2006	75
## 60	DIVYANAND L	50	B.com M.com	2008	60
##	experience Department Salary degree SalaryInterval				
## 1	8 Marketing	18000	6	11k-21k	
## 2	22 Finance	52000	1	51k-61k	
## 3	31 Marketing	91000	10	81k-91k	
## 4	30 Finance	50000	2	41k-51k	
## 5	14 Marketing	24000	1	21k-31k	
## 6	31 HR	18000	17	11k-21k	
## 7	11 Marketing	21000	2	21k-31k	
## 8	14 Marketing	34000	10	31k-41k	
## 9	20 Marketing	30000	5	21k-31k	
## 10	22 Finance	32000	1	31k-41k	
## 11	18 Marketing	28000	13	21k-31k	
## 12	1 Finance	12000	3	11k-21k	
## 13	18 IT	48000	10	41k-51k	
## 14	18 Marketing	44000	6	41k-51k	
## 15	20 Marketing	30000	6	21k-31k	

## 16	22	Marketing	32000	6	31k-41k
## 17	24	Project	34000	5	31k-41k
## 18	18	HR	28000	17	21k-31k
## 19	12	Marketing	42000	8	41k-51k
## 20	8	Finance	48000	1	41k-51k
## 21	31	Project	61000	15	61k-71k
## 22	4	Marketing	14000	4	11k-21k
## 23	4	Marketing	14000	10	11k-21k
## 24	11	Marketing	38000	10	31k-41k
## 25	5	Marketing	35000	6	31k-41k
## 26	14	IT	44000	5	41k-51k
## 27	24	Marketing	74000	6	71k-81k
## 28	24	Project	54000	5	51k-61k
## 29	20	Marketing	50000	6	41k-51k
## 30	8	IT	28000	5	21k-31k
## 31	2	Marketing	14000	9	11k-21k
## 32	4	HR	14000	18	11k-21k
## 33	4	Finance	14000	8	11k-21k
## 34	3	Marketing	13000	11	11k-21k
## 35	22	Finance	42000	15	41k-51k
## 36	1	Marketing	11000	6	11k-21k
## 37	12	HR	22000	16	21k-31k
## 38	8	Marketing	18000	1	11k-21k
## 39	12	legal	32000	13	31k-41k
## 40	5	Finance	15000	14	11k-21k
## 41	3	Marketing	23000	10	21k-31k
## 42	31	HR	41000	18	41k-51k
## 43	8	HR	18000	12	11k-21k
## 44	4	Marketing	28000	1	21k-31k
## 45	11	Marketing	21000	7	21k-31k
## 46	8	Marketing	38000	1	31k-41k
## 47	14	Marketing	24000	10	21k-31k
## 48	30	Marketing	90000	1	81k-91k
## 49	30	Marketing	14000	19	11k-21k
## 50	2	Finance	12000	1	11k-21k
## 51	3	Marketing	23000	10	21k-31k
## 52	31	HR	41000	18	41k-51k
## 53	8	HR	18000	12	11k-21k
## 54	4	Marketing	28000	1	21k-31k
## 55	11	Marketing	21000	7	21k-31k
## 56	8	Marketing	38000	1	31k-41k
## 57	14	Marketing	24000	10	21k-31k
## 58	30	Marketing	90000	1	81k-91k
## 59	31	Marketing	91000	10	81k-91k
## 60	30	Finance	50000	2	41k-51k

Representing Experience with respect to Salary

```
plot(data1$experience, data1$Salary,
     main = "Marketing Experience vs Salary",
     xlab = "Experience (Years)",
```

```
ylab = "Salary",
pch = 19, col = "blue")
```



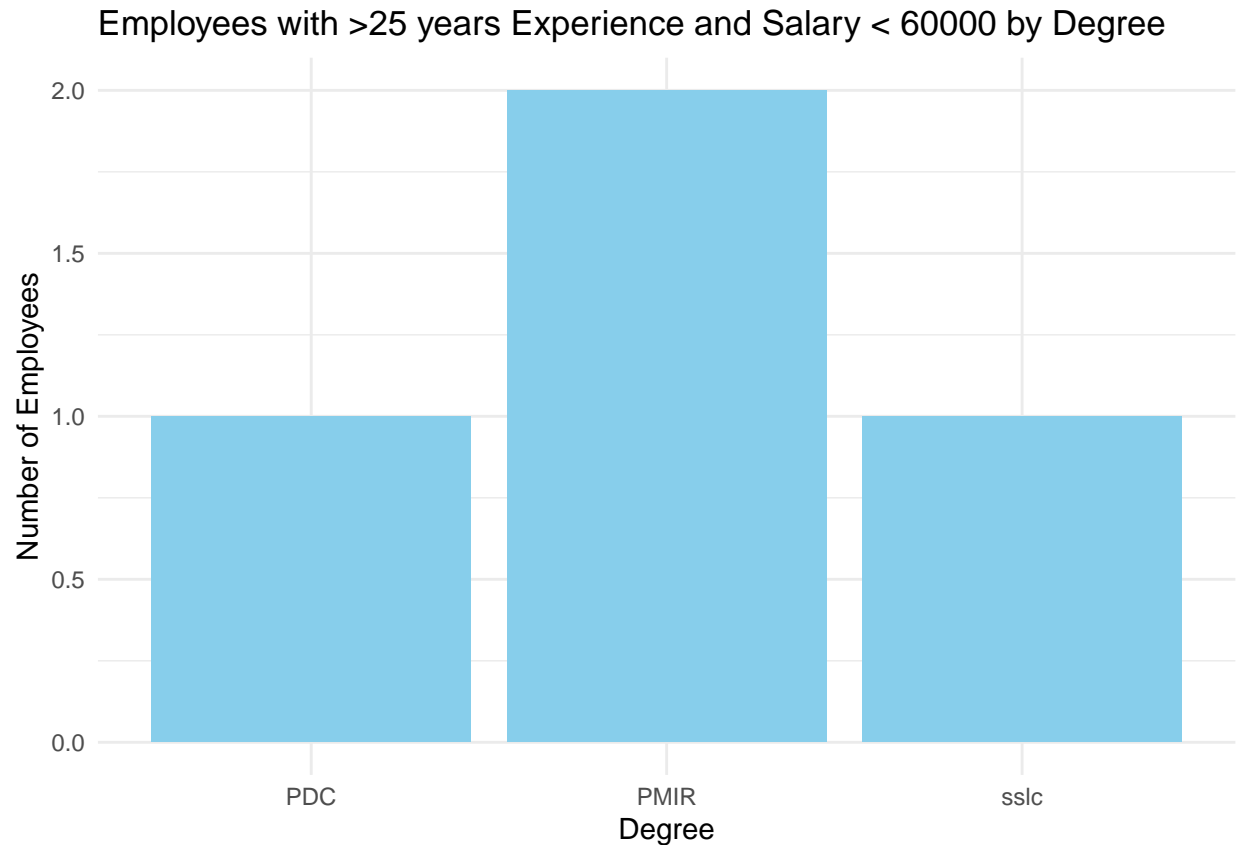
Inference 8:-Even highest experienced employees have salary range from 11000 to 91000

Exploring employees having experience for more than 25 years and below median salary

```
library(ggplot2)
```

Warning: package 'ggplot2' was built under R version 4.4.2

```
library(dplyr)
employees_filtered <- data1 %>%
  filter(experience > 25 & Salary < 45000)
ggplot(employees_filtered, aes(x = Degree)) +
  geom_bar(fill = "skyblue") +
  labs(title = "Employees with >25 years Experience and Salary < 60000 by Degree",
       x = "Degree",
       y = "Number of Employees") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  theme_minimal()
```



Inference 9:- Experience doesn't count for Salary hike with lower Educational qualifications like sslc, PDC

Grouping employees by SalaryInterval and Experience and their respective counts

```
experience_by_salary_interval <- data1 %>%
  group_by(SalaryInterval, experience, Degree, Degree.Percentage)
) %>%
  summarise(Count = n(), .groups = "drop") # .groups = "drop" removes the grouping structure
print(experience_by_salary_interval)
```

```
## # A tibble: 50 x 5
##   SalaryInterval experience Degree      Degree.Percentage Count
##   <fct>          <int> <chr>          <int> <int>
## 1 11k-21k            1 B.com MBA           60     1
## 2 11k-21k            1 BA             60     1
## 3 11k-21k            2 B.com           68     1
## 4 11k-21k            2 BBA MBA           65     1
## 5 11k-21k            3 Bsc Msc           71     1
## 6 11k-21k            4 B.com Mcom          60     1
## 7 11k-21k            4 BBA             80     1
## 8 11k-21k            4 Bsc             75     1
```



```
## 9 11k-21k          4 PMIR          75      1
## 10 11k-21k         5 M.com         57      1
## # i 40 more rows
```

Number of employees under different Salary scale

```
s<-table(experience_by_salary_interval$SalaryInterval)
s
```

```
##
## 11k-21k 21k-31k 31k-41k 41k-51k 51k-61k 61k-71k 71k-81k 81k-91k
##      15      12       8       9       2       1       1       2
```

Inference 10:-Most number of employees are having starting salaries.

Inference 11: Number of employees decreases as salary increases.

Finding the distribution of degree percentage across Salary Scale

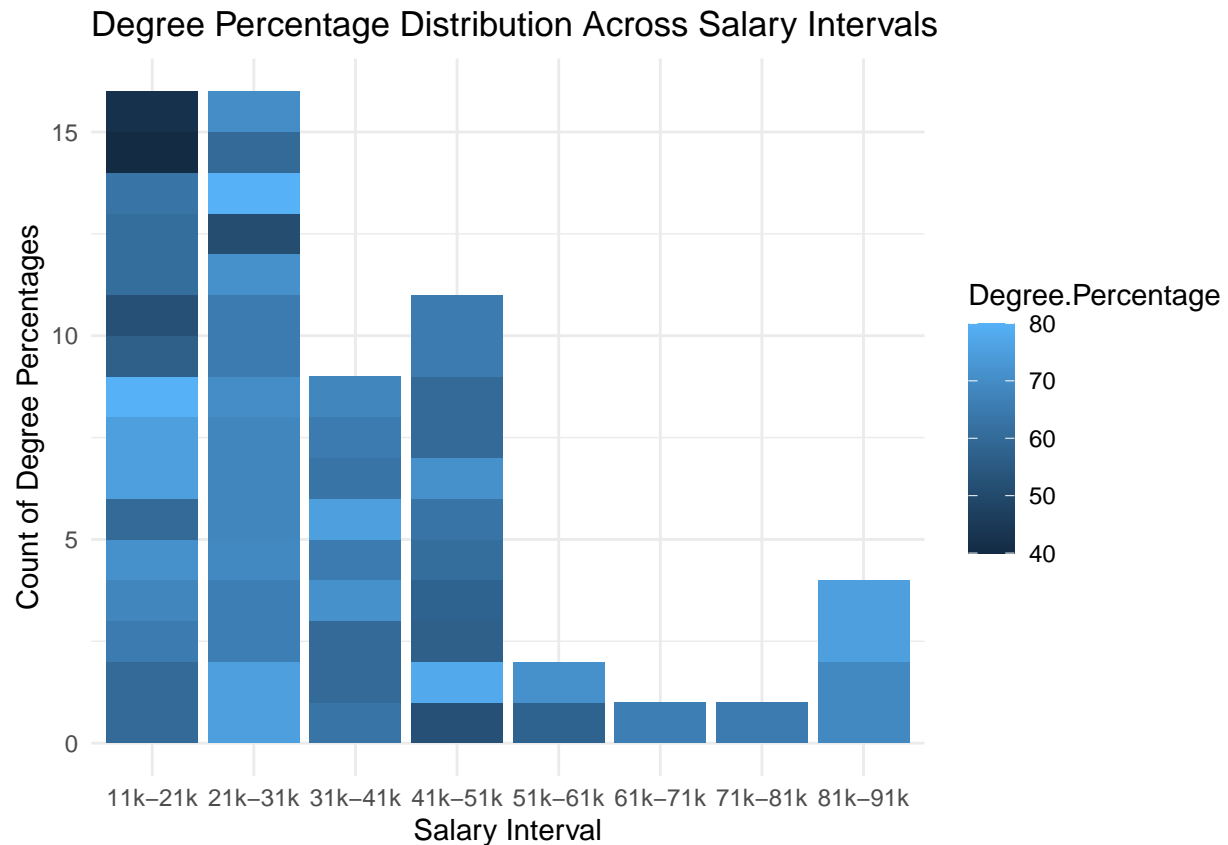
```
library(dplyr)

experience_degree_summary <- data1 %>%
  group_by(SalaryInterval, experience, Degree.Percentage) %>%
  summarise(Count = n(), .groups = "drop")
print(experience_degree_summary)
```

```
## # A tibble: 47 x 4
##   SalaryInterval experience Degree.Percentage Count
##   <fct>          <int>          <int> <int>
## 1 11k-21k         1             60     2
## 2 11k-21k         2             65     1
## 3 11k-21k         2             68     1
## 4 11k-21k         3             71     1
## 5 11k-21k         4             60     1
## 6 11k-21k         4             75     2
## 7 11k-21k         4             80     1
## 8 11k-21k         5             57     1
## 9 11k-21k         8             52     1
## 10 11k-21k        8             61     2
## # i 37 more rows
```

```
ggplot(experience_degree_summary, aes(x = SalaryInterval, y = Count, fill = Degree.Percentage)) +
  geom_bar(stat = "identity", position = "stack") +
  labs(title = "Degree Percentage Distribution Across Salary Intervals",
```

```
x = "Salary Interval",
y = "Count of Degree Percentages" +
theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
theme_minimal()
```



Inference 12:-Employees with lowest degree percentage are having basic Salary(11k-21k) # Inference 13:
Employees with highest degree percentage has highest salary(81k-91k)

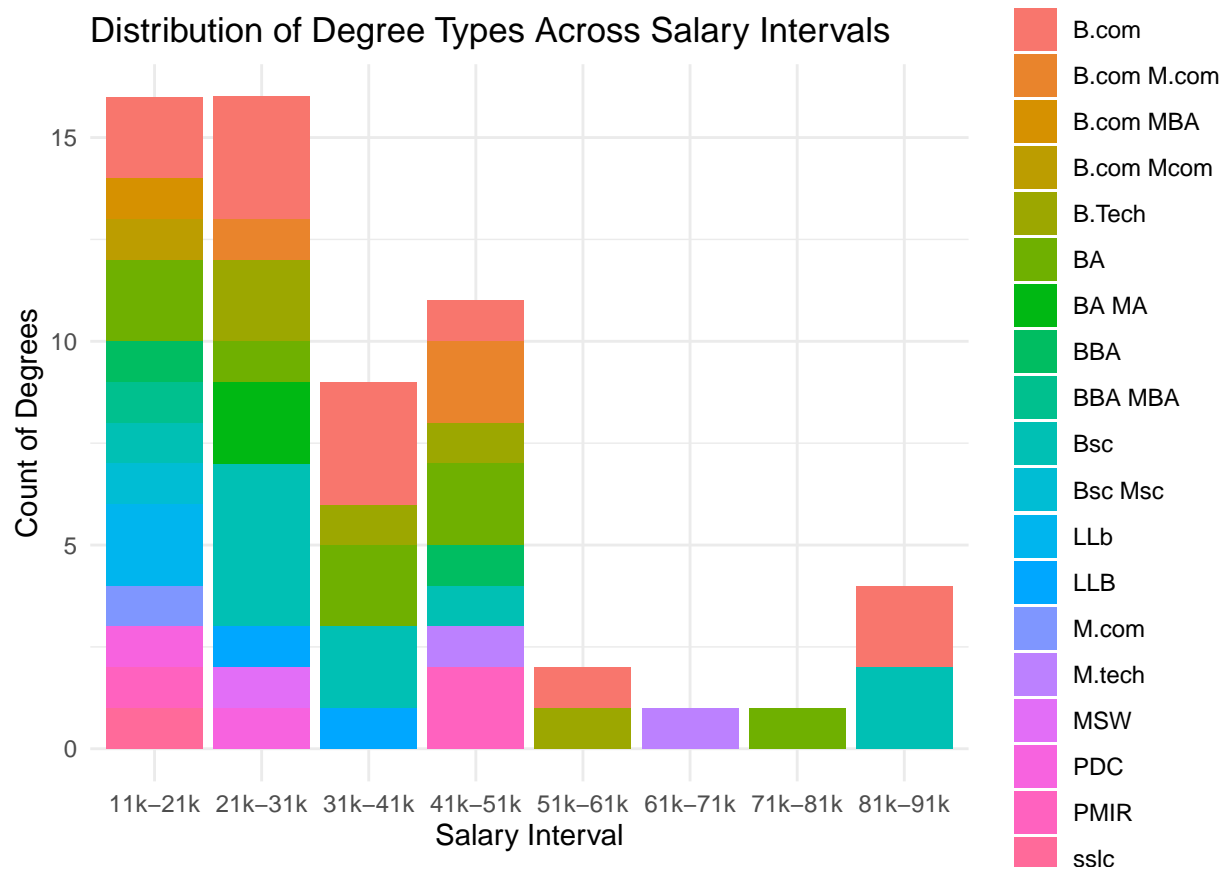
Creating a stacked Bar Chart to observe degree distributions across Salaries

```
library(ggplot2)
degree_salary_summary <- data1 %>%
  group_by(SalaryInterval, Degree) %>%
  summarise(Count = n(), .groups = "drop")
print(degree_salary_summary)
```

```
## # A tibble: 41 x 3
##   SalaryInterval Degree      Count
##   <fct>          <chr>    <int>
## 1 11k-21k      B.com         2
## 2 11k-21k      B.com MBA         1
## 3 11k-21k      B.com Mcom         1
```

```
## 4 11k-21k      BA      2
## 5 11k-21k      BBA     1
## 6 11k-21k      BBA MBA  1
## 7 11k-21k      Bsc     1
## 8 11k-21k      Bsc Msc  1
## 9 11k-21k      LLb     2
## 10 11k-21k     M.com    1
## # i 31 more rows
```

```
ggplot(degree_salary_summary, aes(x = SalaryInterval, y = Count, fill = Degree)) +
  geom_bar(stat = "identity", position = "stack") + # Stack bars by Degree
  labs(title = "Distribution of Degree Types Across Salary Intervals",
       x = "Salary Interval",
       y = "Count of Degrees") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) + # Rotate x-axis labels for readability
  theme_minimal() # Clean theme
```



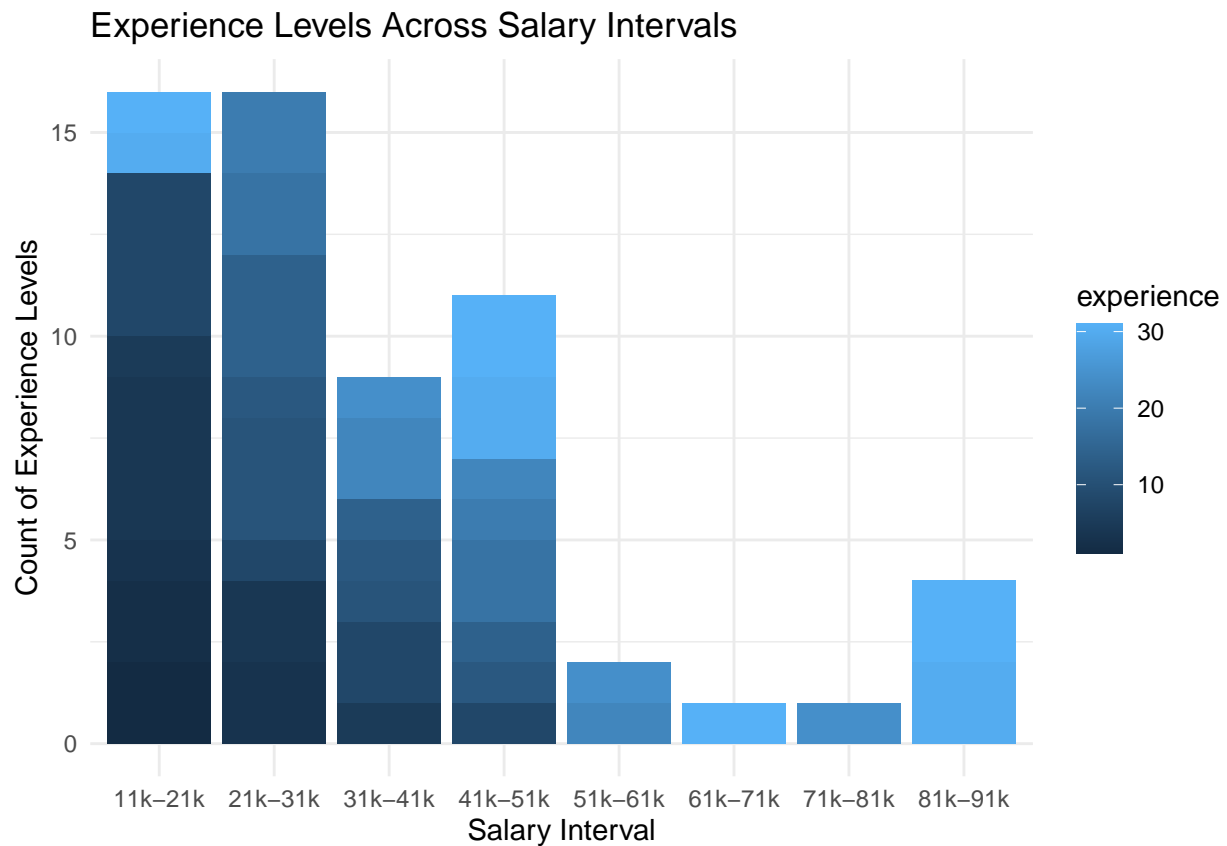
Inference 14:-SSLC and PDC degree holders have lowest pay scale # Inference 15:-Degree doesn't have a significant relationship with salary

```
experience_by_salary_interval <- data1 %>%
  group_by(SalaryInterval, experience, Degree, Degree.Percentage) %>%
  summarise(Count = n(), .groups = "drop")
print(experience_by_salary_interval)
```

```
## # A tibble: 50 x 5
```

```
##      SalaryInterval experience Degree      Degree.Percentage Count
##      <fct>           <int> <chr>           <int> <int>
##  1 11k-21k             1 B.com MBA             60      1
##  2 11k-21k             1 BA                   60      1
##  3 11k-21k             2 B.com                68      1
##  4 11k-21k             2 BBA MBA             65      1
##  5 11k-21k             3 Bsc Msc             71      1
##  6 11k-21k             4 B.com Mcom          60      1
##  7 11k-21k             4 BBA                  80      1
##  8 11k-21k             4 Bsc                 75      1
##  9 11k-21k             4 PMIR                 75      1
## 10 11k-21k             5 M.com                57      1
## # i 40 more rows
```

```
ggplot(experience_by_salary_interval, aes(x = SalaryInterval, y = Count, fill = experience)) +
  geom_bar(stat = "identity", position = "stack") + # Create a stacked bar plot
  labs(title = "Experience Levels Across Salary Intervals",
       x = "Salary Interval",
       y = "Count of Experience Levels") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  theme_minimal()
```



Inference 16:-Employees with 30+ year of experience has the highest Salary(81k-91k) # Inference 17:-Employees with few years of experience has basic salaries.

INFERENCES:-

- 1: Most of the salary is being utilised by the Marketing department(1228000).
- 2: More than half of the total employees are under the Marketing department-58%.(35 out of 60).
- 3: Average salary of employees working under 'Project' department is the highest(49666.67).
- 4: The Marketing department has the lowest and highest allotted salary (Range:11000-19000).
- 5: Experience has a greater influence on salary compared with other attributes.
- 6: The employees under Project department are very much senior to the other department employees.
- 7: People working under the departments-Project, Legal and IT contributes lesser number of employees(3,1,3).
- 8: Even highest experienced employees have salary range from 11000 to 91000.
- 9: Experience doesn't count for Salary hike with lower educational qualifications like SSLC, PDC.
- 10: Most number of employees are having starting salaries.
- 11: Number of employees decreases as salary increases.
- 12: Employees with lowest degree percentage are having basic Salary(11k-21k).
- 13: Employees with highest degree percentage has highest salary(81k-91k).
- 14: SSLC and PDC degree holders have lowest pay scale.
- 15: Degree doesn't have a significant relationship with salary.
- 16: Employees with 30+ year of experience has the highest salary(81k-91k).
- 17: Employees with few years of experience has basic salaries.

Summary

By analysing the given dataset it is clear that employees with basic degree qualification tends to have increase in Salary as their experience increases.