

HR ANALYTICS TO PREDICT THE DEMAND FOR HOURLY-EMPLOYEES

May 13, 2023

0.0.1 Import dataset

```
[1]: data<-read.csv('hr.csv')
```

0.0.2 Data Summary

```
[2]: summary(data)
```

satisfaction_level	last_evaluation	number_project	average_monthly_hours
Min. :0.0900	Min. :0.3600	Min. :2.000	Min. : 96.0
1st Qu.:0.4400	1st Qu.:0.5600	1st Qu.:3.000	1st Qu.:156.0
Median :0.6400	Median :0.7200	Median :4.000	Median :200.0
Mean :0.6128	Mean :0.7161	Mean :3.803	Mean :201.1
3rd Qu.:0.8200	3rd Qu.:0.8700	3rd Qu.:5.000	3rd Qu.:245.0
Max. :1.0000	Max. :1.0000	Max. :7.000	Max. :310.0
time_spend_company	Work_accident	left	promotion_last_5years
Min. : 2.000	Min. :0.0000	Min. :0.0000	Min. :0.00000
1st Qu.: 3.000	1st Qu.:0.0000	1st Qu.:0.0000	1st Qu.:0.00000
Median : 3.000	Median :0.0000	Median :0.0000	Median :0.00000
Mean : 3.498	Mean :0.1446	Mean :0.2381	Mean :0.02127
3rd Qu.: 4.000	3rd Qu.:0.0000	3rd Qu.:0.0000	3rd Qu.:0.00000
Max. :10.000	Max. :1.0000	Max. :1.0000	Max. :1.00000
sales	salary		
Length:14999	Length:14999		
Class :character	Class :character		
Mode :character	Mode :character		

0.0.3 RENAMING THE IRRELEVANT VARIABLE NAME

```
[3]: install.packages('plyr')
```

Installing package into 'C:/Users/JASHWANTH/AppData/Local/R/win-library/4.2'
(as 'lib' is unspecified)

package 'plyr' successfully unpacked and MD5 sums checked

Warning message:

"cannot remove prior installation of package 'plyr'"

Warning message in file.copy(savedcopy, lib, recursive = TRUE):

"problem copying C:\Users\JASHWANTH\AppData\Local\R\win-

library\4.2\00LOCK\plyr\libs\x64\plyr.dll to

C:\Users\JASHWANTH\AppData\Local\R\win-library\4.2\plyr\libs\x64\plyr.dll:

Permission denied"

Warning message:

"restored 'plyr'"

The downloaded binary packages are in

C:\Users\JASHWANTH\AppData\Local\Temp\Rtmpu0i02M\downloaded_packages

```
[4]: library(plyr)
data<-rename(data, c("sales"="role"))
data<-rename(data, c("time_spend_company"="exp_in_company"))
names(data)[10]<-"salary"
head(data)
```

Warning message:

"package 'plyr' was built under R version 4.2.3"

		satisfaction_level	last_evaluation	number_project	average_monthly_hours	exp_in_company
		<dbl>	<dbl>	<int>	<int>	<int>
A data.frame: 6 × 10	1	0.38	0.53	2	157	3
	2	0.80	0.86	5	262	6
	3	0.11	0.88	7	272	4
	4	0.72	0.87	5	223	5
	5	0.37	0.52	2	159	3
	6	0.41	0.50	2	153	3

0.0.4 Exploratory Data Analysis

```
[5]: dim(data)
```

1. 14999 2. 10

```
[6]: str(data)
```

```
'data.frame': 14999 obs. of 10 variables:
 $ satisfaction_level : num 0.38 0.8 0.11 0.72 0.37 0.41 0.1 0.92 0.89 0.42
...
 $ last_evaluation : num 0.53 0.86 0.88 0.87 0.52 0.5 0.77 0.85 1 0.53 ...
 $ number_project : int 2 5 7 5 2 2 6 5 5 2 ...
 $ average_monthly_hours : int 157 262 272 223 159 153 247 259 224 142 ...
 $ exp_in_company : int 3 6 4 5 3 3 4 5 5 3 ...
 $ Work_accident : int 0 0 0 0 0 0 0 0 0 0 ...
 $ left : int 1 1 1 1 1 1 1 1 1 1 ...
```

```
$ promotion_last_5years: int  0 0 0 0 0 0 0 0 0 0 ...
$ role                  : chr  "sales" "sales" "sales" "sales" ...
$ salary                : chr  "low" "medium" "medium" "low" ...
```

```
[7]: attrition<-as.factor(data$left)
      summary(attrition)
```

```
0                11428 1                3571
```

```
[8]: perc_attrition_rate<-sum(data$left/length(data$left))*100
      print(perc_attrition_rate)
```

```
[1] 23.80825
```

```
[9]: cor_vars<-data[,c("satisfaction_level","last_evaluation","number_project",
  "average_montly_hours","exp_in_company","Work_accident","left","promotion_last_5years")]

      aggregate(cor_vars[,c("satisfaction_level","last_evaluation","number_project",
  "average_montly_hours","exp_in_company","Work_accident","promotion_last_5years")],
  by=list(Category=cor_vars$left), FUN=mean)
```

	Category	satisfaction_level	last_evaluation	number_project	average_montly_hours
	<int>	<dbl>	<dbl>	<dbl>	<dbl>
A data.frame: 2 × 8	0	0.6668096	0.7154734	3.786664	199.0602
	1	0.4400980	0.7181126	3.855503	207.4192

```
[10]: install.packages('reshape2')
```

```
Installing package into 'C:/Users/JASHWANTH/AppData/Local/R/win-library/4.2'
(as 'lib' is unspecified)
```

```
package 'reshape2' successfully unpacked and MD5 sums checked
```

```
Warning message:
```

```
"cannot remove prior installation of package 'reshape2'"
```

```
Warning message in file.copy(savedcopy, lib, recursive = TRUE):
```

```
"problem copying C:\Users\JASHWANTH\AppData\Local\R\win-
library\4.2\00LOCK\reshape2\libs\x64\reshape2.dll to
```

```
C:\Users\JASHWANTH\AppData\Local\R\win-
library\4.2\reshape2\libs\x64\reshape2.dll: Permission denied"
```

```
Warning message:
```

```
"restored 'reshape2'"
```

```
The downloaded binary packages are in
```

```
C:\Users\JASHWANTH\AppData\Local\Temp\Rtmpu0i02M\downloaded_packages
```

```
[11]: library(reshape2)
      library(ggplot2)
```

Warning message:

"package 'reshape2' was built under R version 4.2.3"

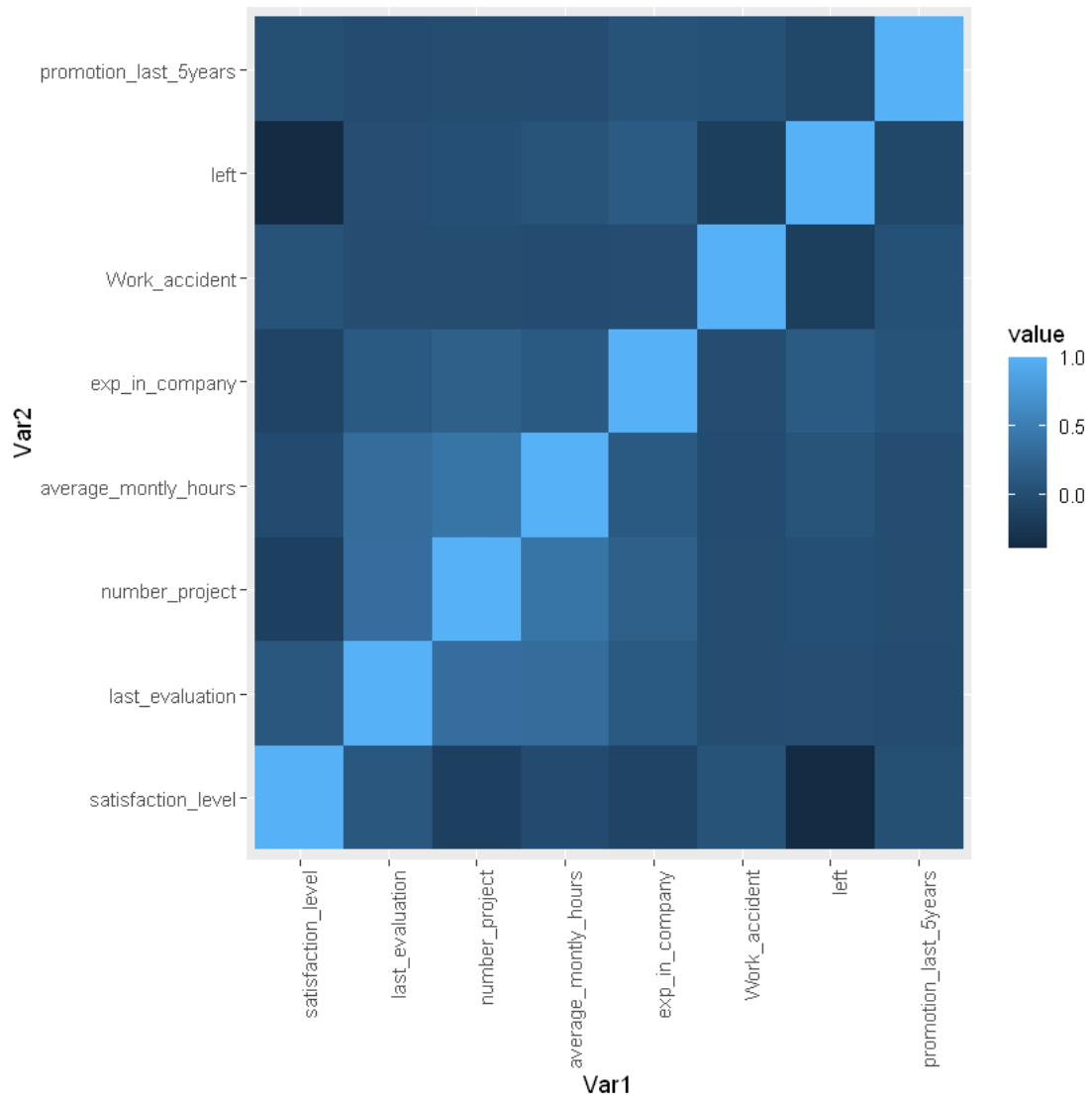
```
[12]: cor_vars<-data[,c("satisfaction_level","last_evaluation","number_project",  
                        ↵  
                        ↪"average_monthly_hours","exp_in_company","Work_accident","left","promotion_last_5years")]
```

```
[13]: cor(cor_vars)
```

A matrix: 8 × 8 of type dbl

	satisfaction_level	last_evaluation	number_project	av
satisfaction_level	1.00000000	0.105021214	-0.142969586	-0.
last_evaluation	0.10502121	1.000000000	0.349332589	0.3
number_project	-0.14296959	0.349332589	1.000000000	0.4
average_monthly_hours	-0.02004811	0.339741800	0.417210634	1.0
exp_in_company	-0.10086607	0.131590722	0.196785891	0.3
Work_accident	0.05869724	-0.007104289	-0.004740548	-0.
left	-0.38837498	0.006567120	0.023787185	0.0
promotion_last_5years	0.02560519	-0.008683768	-0.006063958	-0.

```
[14]: trans<-cor(cor_vars)  
melted_cormat <- melt(trans)  
ggplot(data = melted_cormat, aes(x=Var1, y=Var2, fill=value))+  
geom_tile() +theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



0.0.5 STATISTICAL TEST FOR CORRELATION.

```
[15]: emp_population_satisfaction <-mean(data$satisfaction_level)
      left_pop<-subset(data,left==1)
      print( c('The mean for the employee population is: ',
      ↪emp_population_satisfaction) )
```

```
[1] "The mean for the employee population is: "
[2] "0.612833522234816"
```

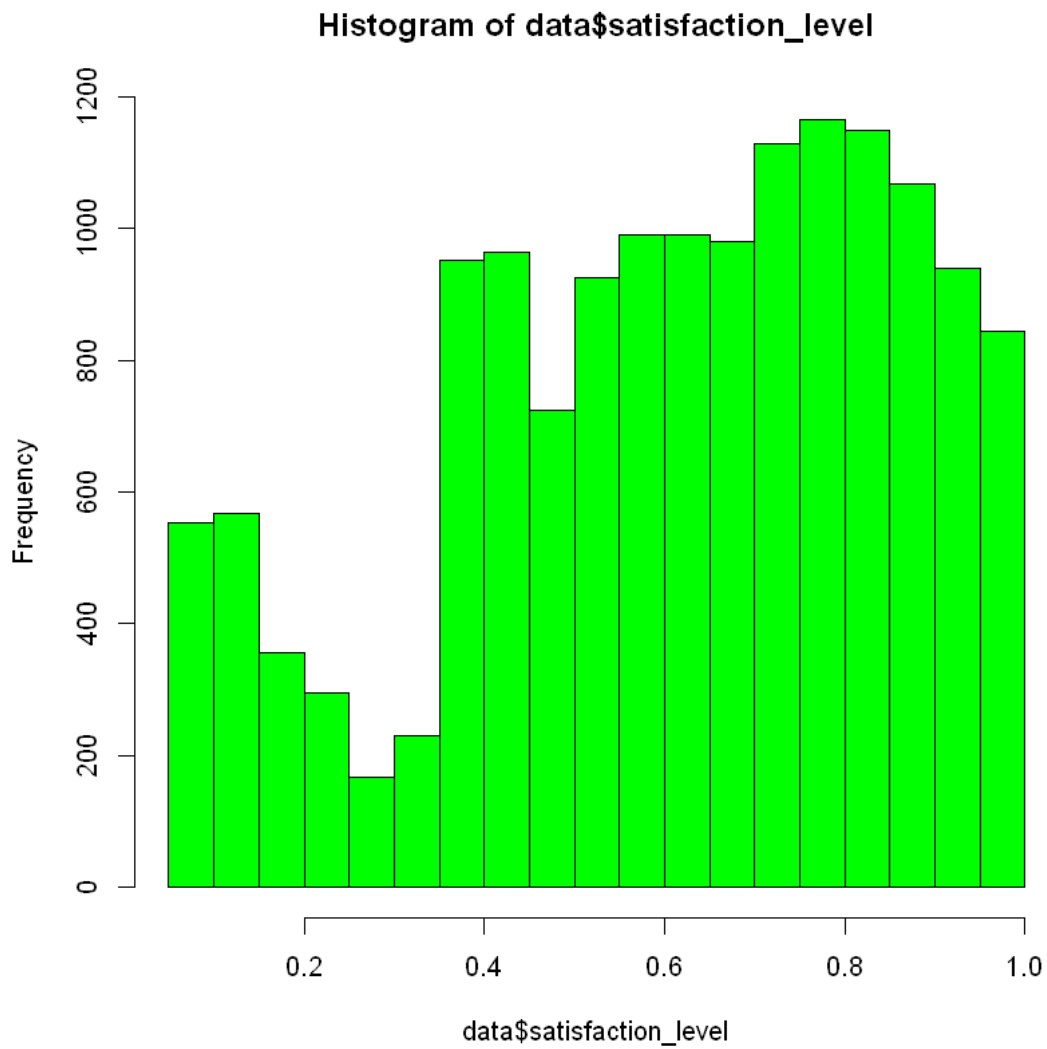
```
[16]: t.test(left_pop$satisfaction_level,mu=emp_population_satisfaction)
```

One Sample t-test

```
data: left_pop$satisfaction_level
t = -39.109, df = 3570, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0.6128335
95 percent confidence interval:
 0.4314385 0.4487576
sample estimates:
mean of x
 0.440098
```

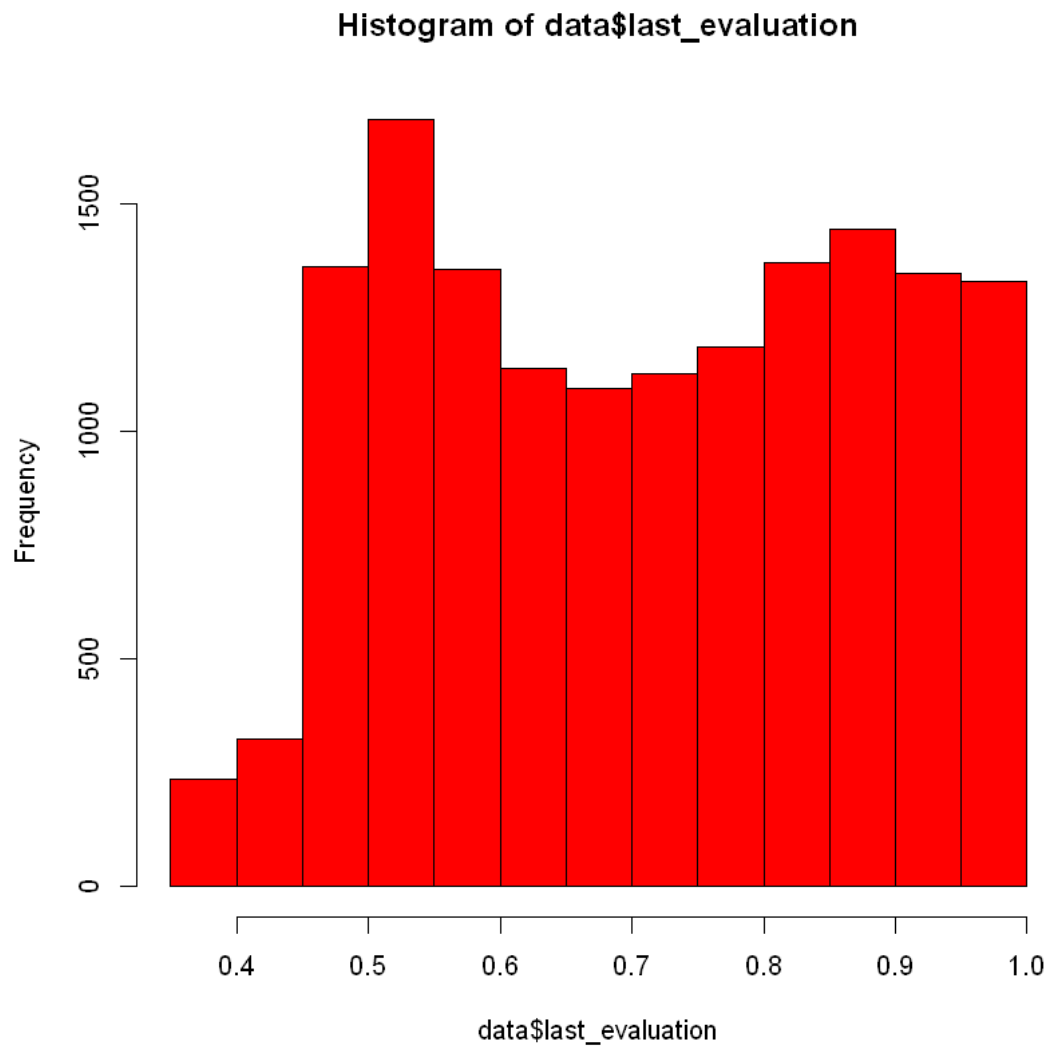
0.0.6 DISPLAY DISTRIBUTION PLOTS ON THE BASIS OF SATISFACTION, EVALUATION AND AVERAGE MONTHLY.

```
[17]: hist(data$satisfaction_level, col="green")
```



```
[18]: par(mfrow=c(1,3))
```

```
[19]: hist(data$last_evaluation, col="red")
```



```
[20]: hist(data$average_monthly_hours, col="blue")
```

