r

2024-07-11

# Load necessary libraries  
library(readxl)

## Warning: package 'readxl' was built under R version 4.4.1

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.4.1

##   
## 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

# Set the path to your Excel file  
file\_path <-"C:\\Users\\ASUS\\Downloads\\bike data sneha.excel xs.xlsx"

# Read the specific sheet  
bike\_data <- read\_excel(file\_path, sheet = "bike data sneha.excel xs")

## New names:  
## • `temp` -> `temp...10`  
## • `temp` -> `temp...11`

# Display the structure of the dataset  
str(bike\_data)

## tibble [149 × 16] (S3: tbl\_df/tbl/data.frame)  
## $ instant : num [1:149] 1 2 3 4 5 6 7 8 9 10 ...  
## $ dteday : POSIXct[1:149], format: "2011-01-01" "2011-01-02" ...  
## $ season : num [1:149] 1 1 1 1 1 1 1 1 1 1 ...  
## $ yr : num [1:149] 0 0 0 0 0 0 0 0 0 0 ...  
## $ month : num [1:149] 1 1 1 1 1 1 1 1 1 1 ...  
## $ holiday : num [1:149] 0 0 0 0 0 0 0 0 0 0 ...  
## $ weekday : num [1:149] 6 0 1 2 3 4 5 6 0 1 ...  
## $ workingday: num [1:149] 0 0 1 1 1 1 1 0 0 1 ...  
## $ weathersit: num [1:149] 2 2 1 1 1 1 2 2 1 1 ...  
## $ temp...10 : num [1:149] 0.344 0.363 0.196 0.2 0.227 ...  
## $ temp...11 : num [1:149] 0.364 0.354 0.189 0.212 0.229 ...  
## $ hum : num [1:149] 0.806 0.696 0.437 0.59 0.437 ...  
## $ windspeed : num [1:149] 0.16 0.249 0.248 0.16 0.187 ...  
## $ casual : num [1:149] 331 131 120 108 82 88 148 68 54 41 ...  
## $ registered: num [1:149] 654 670 1229 1454 1518 ...  
## $ cnt : num [1:149] 985 801 1349 1562 1600 ...

# Display the first few rows of the dataset  
head(bike\_data)

## # A tibble: 6 × 16  
## instant dteday season yr month holiday weekday workingday  
## <dbl> <dttm> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 2011-01-01 00:00:00 1 0 1 0 6 0  
## 2 2 2011-01-02 00:00:00 1 0 1 0 0 0  
## 3 3 2011-01-03 00:00:00 1 0 1 0 1 1  
## 4 4 2011-01-04 00:00:00 1 0 1 0 2 1  
## 5 5 2011-01-05 00:00:00 1 0 1 0 3 1  
## 6 6 2011-01-06 00:00:00 1 0 1 0 4 1  
## # ℹ 8 more variables: weathersit <dbl>, temp...10 <dbl>, temp...11 <dbl>,  
## # hum <dbl>, windspeed <dbl>, casual <dbl>, registered <dbl>, cnt <dbl>

# Display the first few rows of the dataset  
head(bike\_data)

## # A tibble: 6 × 16  
## instant dteday season yr month holiday weekday workingday  
## <dbl> <dttm> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1 2011-01-01 00:00:00 1 0 1 0 6 0  
## 2 2 2011-01-02 00:00:00 1 0 1 0 0 0  
## 3 3 2011-01-03 00:00:00 1 0 1 0 1 1  
## 4 4 2011-01-04 00:00:00 1 0 1 0 2 1  
## 5 5 2011-01-05 00:00:00 1 0 1 0 3 1  
## 6 6 2011-01-06 00:00:00 1 0 1 0 4 1  
## # ℹ 8 more variables: weathersit <dbl>, temp...10 <dbl>, temp...11 <dbl>,  
## # hum <dbl>, windspeed <dbl>, casual <dbl>, registered <dbl>, cnt <dbl>

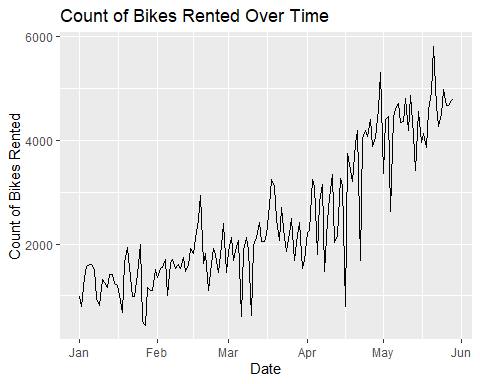
# Summary statistics of the dataset  
summary(bike\_data)

## instant dteday season yr month   
## Min. : 1 Min. :2011-01-01 Min. :1.00 Min. :0 Min. :1.000   
## 1st Qu.: 38 1st Qu.:2011-02-07 1st Qu.:1.00 1st Qu.:0 1st Qu.:2.000   
## Median : 75 Median :2011-03-16 Median :1.00 Median :0 Median :3.000   
## Mean : 75 Mean :2011-03-16 Mean :1.47 Mean :0 Mean :2.987   
## 3rd Qu.:112 3rd Qu.:2011-04-22 3rd Qu.:2.00 3rd Qu.:0 3rd Qu.:4.000   
## Max. :149 Max. :2011-05-29 Max. :2.00 Max. :0 Max. :5.000   
## holiday weekday workingday weathersit   
## Min. :0.00000 Min. :0 Min. :0.0000 Min. :1.000   
## 1st Qu.:0.00000 1st Qu.:1 1st Qu.:0.0000 1st Qu.:1.000   
## Median :0.00000 Median :3 Median :1.0000 Median :1.000   
## Mean :0.02013 Mean :3 Mean :0.6846 Mean :1.463   
## 3rd Qu.:0.00000 3rd Qu.:5 3rd Qu.:1.0000 3rd Qu.:2.000   
## Max. :1.00000 Max. :6 Max. :1.0000 Max. :3.000   
## temp...10 temp...11 hum windspeed   
## Min. :0.05913 Min. :0.07907 Min. :0.0000 Min. :0.04541   
## 1st Qu.:0.23333 1st Qu.:0.24572 1st Qu.:0.4948 1st Qu.:0.16045   
## Median :0.34250 Median :0.33794 Median :0.6142 Median :0.21394   
## Mean :0.36807 Mean :0.36087 Mean :0.6194 Mean :0.21722   
## 3rd Qu.:0.50583 3rd Qu.:0.49305 3rd Qu.:0.7408 3rd Qu.:0.26188   
## Max. :0.70833 Max. :0.65469 Max. :0.9483 Max. :0.50746   
## casual registered cnt   
## Min. : 9 Min. : 416 Min. : 431   
## 1st Qu.: 121 1st Qu.:1367 1st Qu.:1536   
## Median : 307 Median :1705 Median :2056   
## Mean : 492 Mean :2012 Mean :2504   
## 3rd Qu.: 695 3rd Qu.:2570 3rd Qu.:3429   
## Max. :2355 Max. :4238 Max. :5805

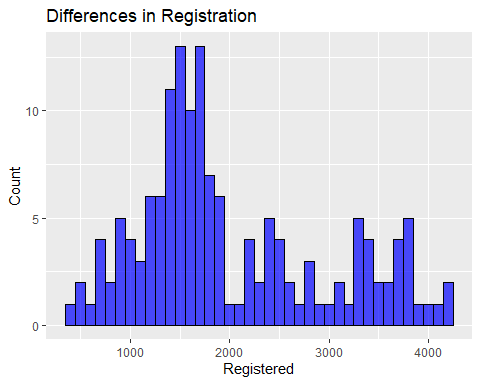
# Example of data visualization using ggplot2  
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.4.1

# Plotting the count of bikes rented over time  
ggplot(bike\_data, aes(x = dteday, y = cnt)) +  
 geom\_line() +  
 labs(title = "Count of Bikes Rented Over Time",  
 x = "Date",  
 y = "Count of Bikes Rented")



ggplot(bike\_data, aes(x = registered)) +  
 geom\_histogram(binwidth = 100, fill = "blue", color = "black", alpha = 0.7) +  
 labs(title = "Differences in Registration", x = "Registered", y = "Count")



ggplot(bike\_data, aes(x = registered, y = cnt)) +  
 geom\_point(color = "blue", alpha = 0.6) +  
 labs(title = "Registered vs CNT", x = "Registered", y = "CNT")

