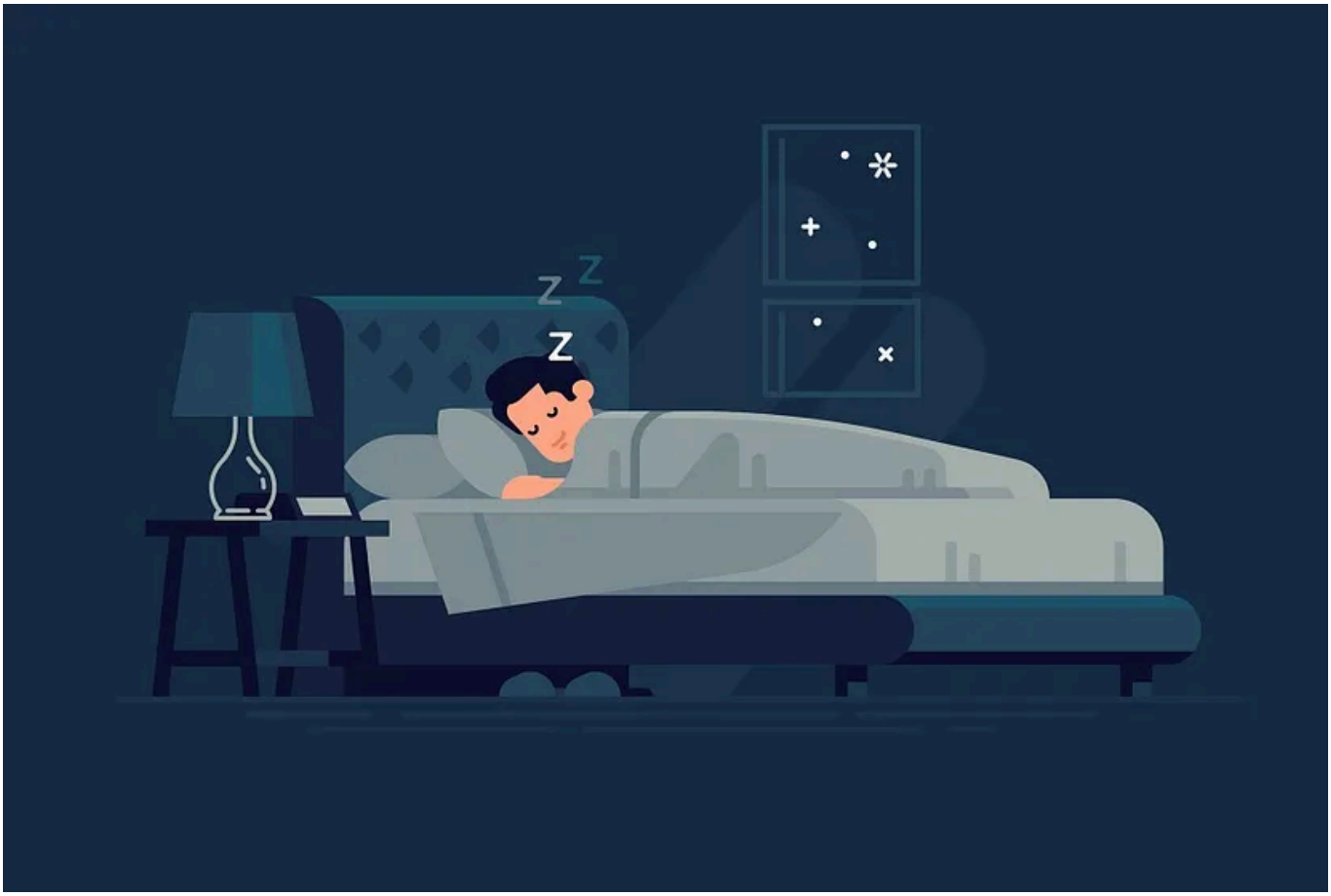
Analyzing Sleep Data & Making Predictions Using R.

N Nethmi Kulasooriya
6 min read · May 31, 2024



Data analysis involves inspecting, cleaning, transforming, and modeling data to discover useful information, inform conclusions, and support decision-making. It encompasses a variety of techniques and tools to extract insights from datasets, ranging from basic statistical summaries to advanced machine learning algorithms

Sleep is an interesting part of our lives. It's a naturally recurring state of mind and body characterized by altered consciousness, in this article, I hope to share my experience of data exploration and making Predictions using R.

In this article, you'll learn how to:

- Explore and prepare a dataset for analysis.
- Utilize R Shiny to create interactive graphs and charts.
- Identify patterns in the data and interpret their significance.
- Build regression models to make predictions.
- Combine predictions with visualizations for a comprehensive analysis.

Exploring data

The name of my data set is "Sleep Health and Lifestyle".

Dataset Overview:

The Sleep Health and Lifestyle Dataset covers various variables related to sleep and daily habits. It includes gender, age, occupation, sleep duration, quality of sleep, physical activity level, stress levels, B MI category, blood pressure, heart rate, daily steps, and the presence or absence of sleep disorders.

First of all, we need to call the packages

Hide

```
library(ggplot2)
library(datasets)
```

This is a code

Then import the data set

In R, there are several functions you can use to import datasets, depending on the format of the data. It is one of the ways to get a csv file. csv Files: For comma-separated values files, you can use the read. csv() function.

```
Sleep_Health=read.csv("C:/Users/nethmi/Downloads/Sleep_health_and_lifestyle_dataset.csv")
```

This is a code

R's head() function displays the first few rows of a data frame or matrix. It is helpful for quickly inspecting the structure and contents of a dataset without showing the entire dataset

Fetching the First few rows of the data set

```
head(Sleep_Health)
```

	Person.ID <int>	Gender <chr>	Age <int>	Occupation <chr>	Sleep.Duration <dbl>	Quality.of.Sleep <int>
1	1	Male	27	Software Engineer	6.1	6
2	2	Male	28	Doctor	6.2	6
3	3	Male	28	Doctor	6.2	6
4	4	Male	28	Sales Representative	5.9	4
5	5	Male	28	Sales Representative	5.9	4
6	6	Male	28	Software Engineer	5.9	4

The head() function in R displays the first few rows of a data frame.

Getting the data structure

#view the data structuer of the data set

Hide

```
str(Sleep_Health)
```

```
'data.frame':   374 obs. of  13 variables:
 $ Person.ID              : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Gender                 : chr  "Male" "Male" "Male" "Male" ...
 $ Age                    : int  27 28 28 28 28 28 29 29 29 29 ...
 $ Occupation             : chr  "Software Engineer" "Doctor" "Doctor" "Sales Representative" ...
 $ Sleep.Duration         : num  6.1 6.2 6.2 5.9 5.9 5.9 6.3 7.8 7.8 7.8 ...
 $ Quality.of.Sleep       : int  6 6 6 4 4 4 6 7 7 7 ...
 $ Physical.Activity.Level: int  42 60 60 30 30 30 40 75 75 75 ...
 $ Stress.Level           : int  6 8 8 8 8 8 7 6 6 6 ...
 $ BMI.Category           : chr  "Overweight" "Normal" "Normal" "Obese" ...
 $ Blood.Pressure         : chr  "126/83" "125/80" "125/80" "140/90" ...
 $ Heart.Rate             : int  77 75 75 85 85 85 82 70 70 70 ...
 $ Daily.Steps            : int  4200 10000 10000 3000 3000 3000 3500 8000 8000 8000 ...
 $ Sleep.Disorder         : chr  "None" "None" "None" "SleepApnea" ...
```

The str() function in R is short for “structure,” and it provides a compact display of the internal structure of an R object. In my example, str(Sleep. Health) summarizes the data frame, indicating that it is a data frame with 374 observations and 13 variables.

Then getting the summary of the data set

#summary of the data set

Hide

```
summary(Sleep_Health)
```

```
  Person.ID        Gender               Age        Occupation        Sleep.Duration  Quality.of.Sleep
Min.   :  1.00   Length:374         Min.   :27.00   Length:374         Min.   :5.800   Min.   :4.000
1st Qu.: 94.25   Class :character   1st Qu.:35.25   Class :character   1st Qu.:6.400   1st Qu.:6.000
Median :187.50   Mode  :character   Median :43.00   Mode  :character   Median :7.200   Median :7.000
Mean   :187.50                      Mean   :42.18                      Mean   :7.132   Mean   :7.313
3rd Qu.:280.75                      3rd Qu.:50.00                      3rd Qu.:7.800   3rd Qu.:8.000
Max.   :374.00                      Max.   :59.00                      Max.   :8.500   Max.   :9.000
Physical.Activity.Level  Stress.Level   BMI.Category       Blood.Pressure       Heart.Rate
Min.   :30.00           Min.   :3.000   Length:374         Length:374         Min.   :65.00
1st Qu.:45.00           1st Qu.:4.000   Class :character   Class :character   1st Qu.:68.00
Median :60.00           Median :5.000   Mode  :character   Mode  :character   Median :70.00
Mean   :59.17           Mean   :5.385                                         Mean   :70.17
3rd Qu.:75.00           3rd Qu.:7.000                                         3rd Qu.:72.00
Max.   :90.00           Max.   :8.000                                         Max.   :86.00
 Daily.Steps    Sleep.Disorder
Min.   : 3000   Length:374
1st Qu.: 5600   Class :character
Median : 7000   Mode  :character
Mean   : 6817
3rd Qu.: 8000
Max.   :10000
```

This summary provides descriptive statistics for each variable in the dataset “Sleep_Health”.

Each variable’s summary includes measures such as minimum, maximum, median, mean, and quartiles, providing insights into the distribution and characteristics of the data. For categorical variables like gender and occupation, the summary includes counts of unique categories.

Then check out if are there any missing values in our dataset, there are two method

Method 1

Hide

```
is.na(Sleep_Health)
```

Method 2

#idintify missing data

Hide

```
colSums(is.na(Sleep_Health))
```

```
                Person.ID                    Gender                       Age                Occupation 
                        0                         0                         0                         0 
          Sleep.Duration        Quality.of.Sleep Physical.Activity.Level            Stress.Level 
                        0                         0                         0                         0 
           BMI.Category           Blood.Pressure                Heart.Rate               Daily.Steps 
                        0                         0                         0                         0 
          Sleep.Disorder 
                        0 
```

The output from colSums(is. na(Sleep_Health)) indicates the number of missing values (NA) in each column of the dataset “Sleep_Health.” In this case, it appears that there are no missing values in any of the columns of the dataset, as all counts are zero. This suggests that the dataset is complete and there are no missing observations to be addressed.

Working with R Shiny

How I used R Shiny to make interactive graphs and charts

Interactive Web App for “Sleep Health and Lifestyle Dataset”

• Loading Libraries: Towards loading ``shiny” and ggplot2 for creating and plotting the application.

• Reading the dataset: The dataset was read into R using the read.csv() function.

• Defining the UI: Defined the UI using the fluidPage() function with a title panel and sidebar layout.

• Defining the server logic: The server logic is defined using the ‘server()’ function.

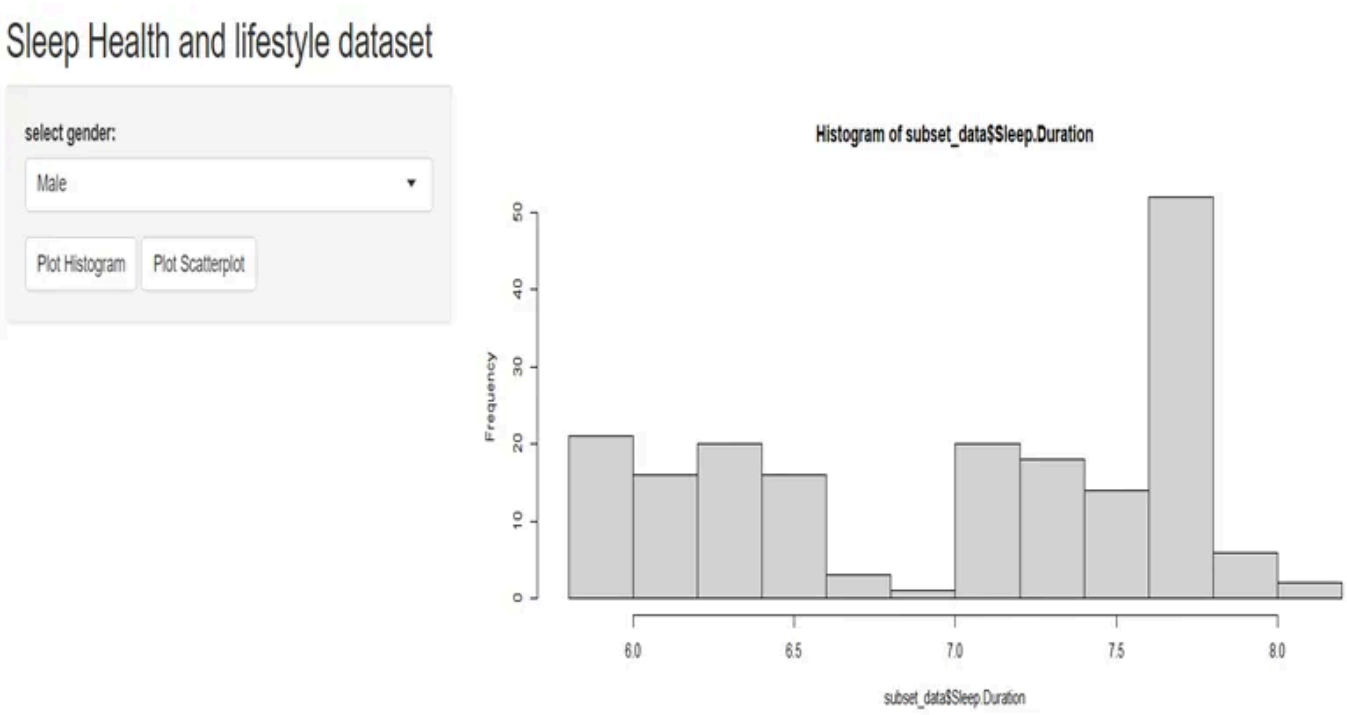
• Observer Events: Created two observer events to respond to action buttons.

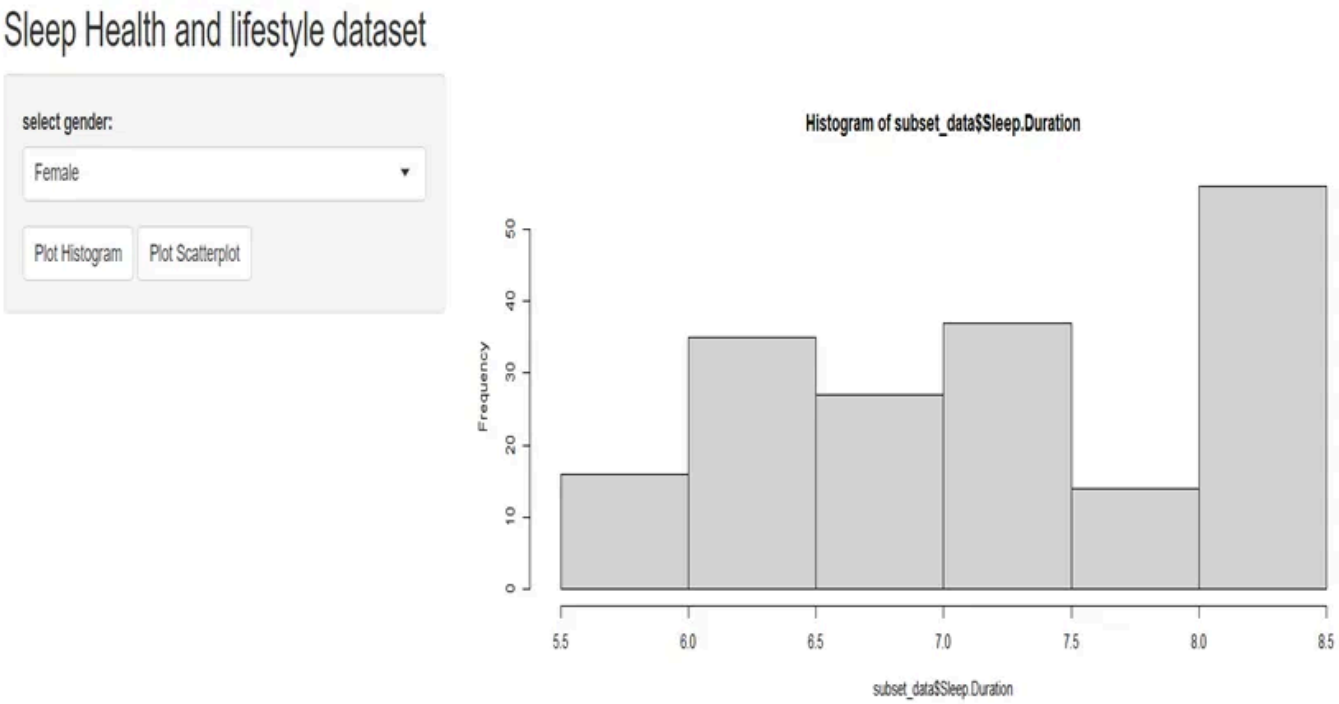
• Running the application: The application is run using the shinyApp() function.

Codes & histogram of sleep Duration are shown below

```
library(shiny)
library(ggplot2)

# Read the dataset
dataset <- read.csv("C:/Users/nethmi/Downloads/Sleep_health_and_lifestyle_dataset.csv")

# Define UI for application that draws a histogram
ui <- fluidPage(

  # Application title
  titlePanel("Sleep Health and lifestyle dataset"),

  # Sidebar with a slider input for number of bins
  sidebarLayout(
    sidebarPanel(
      selectInput("Gender",
                  "select gender:",
                  choices = unique(dataset$Gender)),
      actionButton("plot_hist","Plot Histogram"),
      actionButton("plot_scatter","Plot Scatterplot")
    ),

    # Show a plot of the country
    mainPanel(
      plotOutput("Plot")
    )
  )
)

# Define server logic required to draw a histogram
server <- function(input, output) {
  observeEvent(input$plot_hist,{
    subset_data <- subset(dataset, Gender == input$Gender)
    output$Plot <- renderPlot({
      hist(subset_data$Sleep.Duration)
    })
  })

  observeEvent(input$plot_scatter,{
    subset_data <- subset(dataset, Gender == input$Gender)
    output$Plot <- renderPlot({
      ggplot(subset_data, aes(x = Sleep.Duration, y = variable2)) +
        geom_point() +
        xlab("Sleep Duration") +
        ylab("Quality of Sleep") +
        ggtitle("Scatterplot of Sleep Duration vs Quality of Sleep")
    })
  })
}

# Run the application
shinyApp(ui = ui, server = server)
```

This is a code





Data visualization

ggplot2() function is Primarily used for data visualization in R. With ggplot2's flexibility and extensive documentation, we can create a wide range of visualizations to explore and communicate your data effectively.

#data vizualitation

Hide

```
ggplot(Sleep_Health,aes(x=Sleep.Duration ,y=Quality.of.Sleep,color=Gender))+
  geom_point()+
  ggtitle("Scatter plot of Sleep.Duration vs Quality.of.Sleep,colored by Gender")
```

#box plot

Hide

```
library(ggplot2)
ggplot(Sleep_Health,aes(x=Sleep.Duration,y=Quality.of.Sleep, color = Gender))+
  geom_boxplot()+
  ggtitle("scatter plot Sleep.Duration vsQuality.of.Sleep \ n colored by Gender  ")
```

The scatter plot will display individual data points, with each point representing an observation in the dataset. The “x-axis represents sleep duration, while the y-axis represents quality of sleep”. You can observe how sleep duration correlates with quality of sleep. you might notice that individuals with longer sleep durations tend to report higher quality of sleep.

The boxplot also shows the correlation shown by the scatter plot.

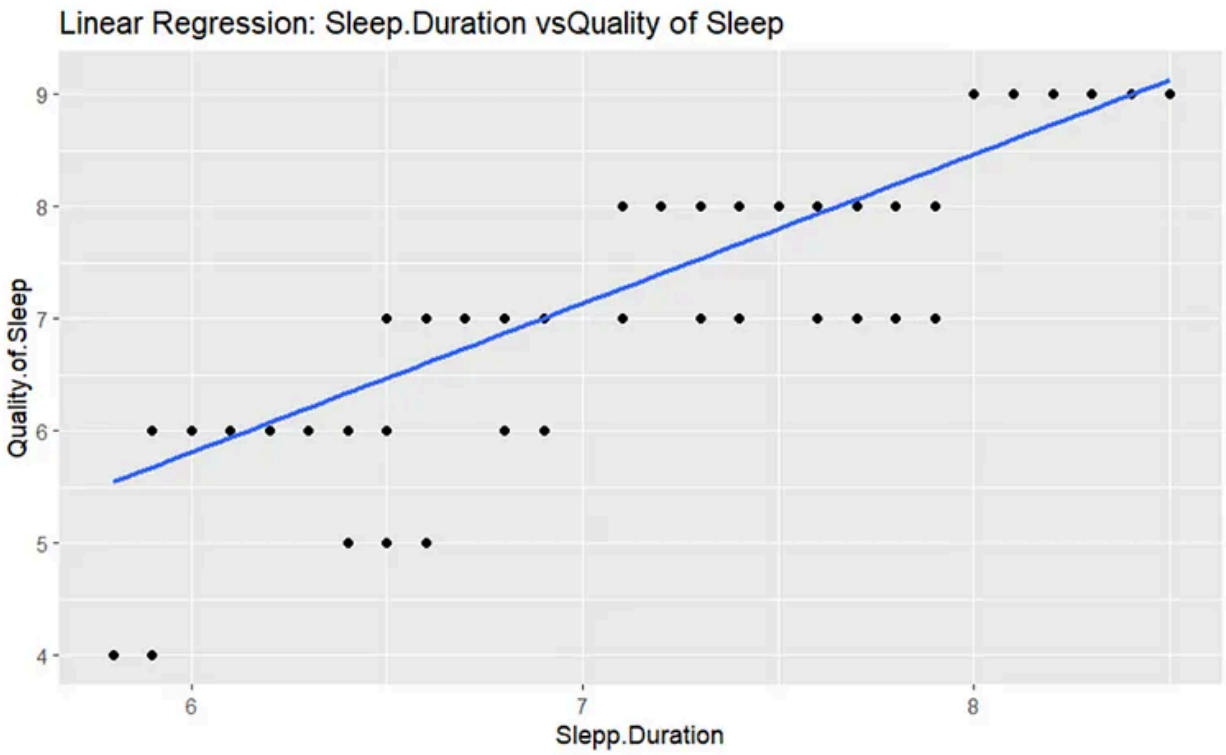
Then Make Predictions

Linear Regression

#linear reggression

Hide

```
ggplot(Sleep_Health, aes(x = Sleep.Duration , y =Quality.of.Sleep)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  labs(x = "Slepp.Duration", y = "Quality.of.Sleep") +
  ggtitle("Linear Regression: Sleep.Duration vsQuality of Sleep")
```



Model regression

Hide

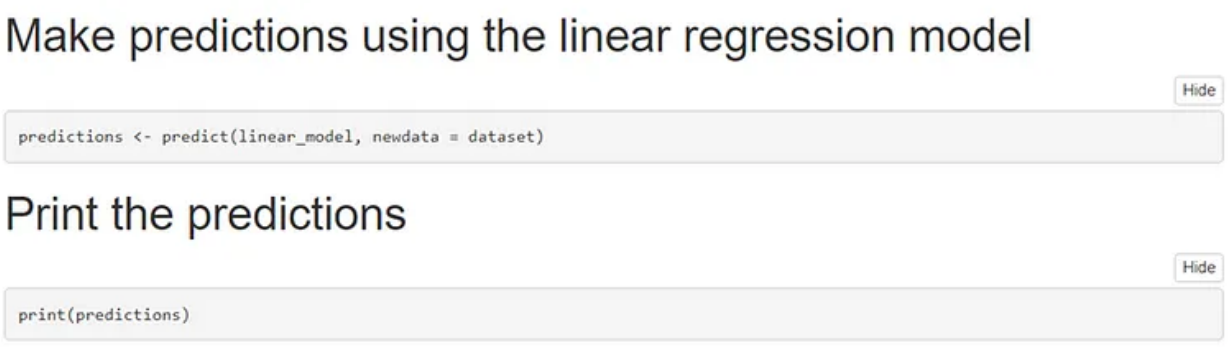
```
linear_model <- lm(Sleep.Duration ~ Quality.of.Sleep, data = dataset)
summary(linear_model)
```

```
Call:
lm(formula = Sleep.Duration ~ Quality.of.Sleep, data = dataset)

Residuals:
     Min       1Q   Median       3Q      Max
-0.46132 -0.33552 -0.06132  0.26448  0.95158

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      2.83871    0.11977   23.70   <2e-16 ***
Quality.of.Sleep 0.58710    0.01616   36.32   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3736 on 372 degrees of freedom
Multiple R-squared:  0.7801,    Adjusted R-squared:  0.7795
F-statistic:  1319 on 1 and 372 DF,  p-value: < 2.2e-16
```

The intercept coefficient represents the estimated mean sleep duration when the quality of sleep is zero. In this case, the estimated intercept is approximately 2.83871.

The coefficient for quality of sleep indicates the change in sleep duration associated with a one-unit increase in quality of sleep. In this case, the estimated coefficient is approximately 0.58710. This suggests that, on average, for every one-unit increase in quality of sleep, sleep duration increases by approximately 0.58710 units.

Prediction



In the provided code, predictions for sleep duration are made using a linear regression model (`linear_model`). Let's analyze the output:

Predicted Sleep Duration: The prediction for sleep duration based on the given quality of sleep (374) is approximately 222.41 units.

This means that, according to the linear regression model, an individual with a reported quality of sleep of 374 is predicted to have a sleep duration of around 222.41 units.t

In this journey through the realm of sleep and health using R programming, we embarked on a fascinating exploration of data and predictive analytics. We began by preparing and exploring a dataset, uncovering intriguing insights into sleep patterns and their correlation with gender. Through interactive visualizations and regression models, we unveiled the importance of adequate sleep in maintaining optimal health and well-being. Our findings underscored the significance of data-driven insights in understanding complex phenomena like sleep quality.

I encourage you, dear readers, to embark on your data exploration adventures. Armed with the powerful tools of R programming and a thirst for knowledge, you can delve into any domain of interest, be it health, finance, or social sciences. By analyzing datasets, visualizing trends, and making predictions, you'll not only gain valuable insights but also develop critical thinking skills essential in today's data-driven world. So, roll up your sleeves, dive into the data, and let the journey begin!

Written by Nethmi Kulasooriya

0 Followers

Edit profile

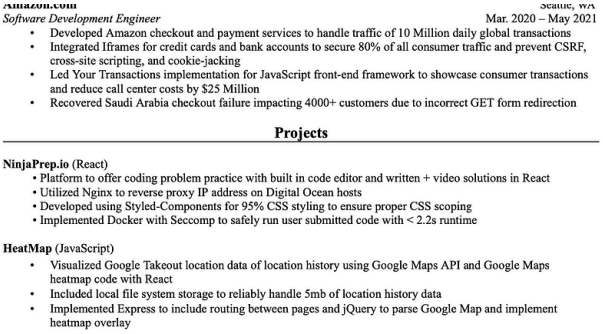
Recommended from Medium

Atif Hussain

Why and How I Use ChatGPT to Create Synthetic Datasets for My...

When teaching a course on data analysis, machine learning (ML), or any related subjec...

Mar 26



Software Development Engineer Mar. 2020 – May 2021

- Developed Amazon checkout and payment services to handle traffic of 10 Million daily global transactions
- Integrated Iframes for credit cards and bank accounts to secure 80% of all consumer traffic and prevent CSRF, cross-site scripting, and cookie-jacking
- Led Your Transactions implementation for JavaScript front-end framework to showcase consumer transactions and reduce call center costs by $25 Million
- Recovered Saudi Arabia checkout failure impacting 4000+ customers due to incorrect GET form redirection

Projects

NinjaPrep.io (React)
- Platform to offer coding problem practice with built in code editor and written + video solutions in React
- Utilized Nginx to reverse proxy IP address on Digital Ocean hosts
- Developed using Styled-Components for 95% CSS styling to ensure proper CSS scoping
- Implemented Docker with Seccomp to safely run user submitted code with < 2.2s runtime

HeatMap (JavaScript)
- Visualized Google Takeout location data of location history using Google Maps API and Google Maps heatmap code with React
- Included local file system storage to reliably handle 5mb of location history data
- Implemented Express to include routing between pages and jQuery to parse Google Map and implement heatmap overlay

Alexander Nguyen in Level Up Coding

The resume that got a software engineer a $300,000 job at Google.

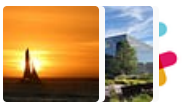
1-page. Well-formatted.

Jun 1 12.1K 162

Lists

Staff Picks

685 stories · 1129 saves



Stories to Help You Level-Up at Work

19 stories · 687 saves

Self-Improvement 101

20 stories · 2287 saves

Productivity 101

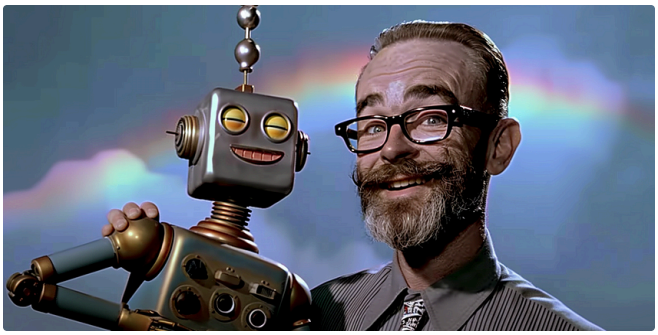
20 stories · 2017 saves

John Gorman

Stop Wasting Your Time

A Simple Framework for Making Better Decisions

Jun 4 16.4K 308

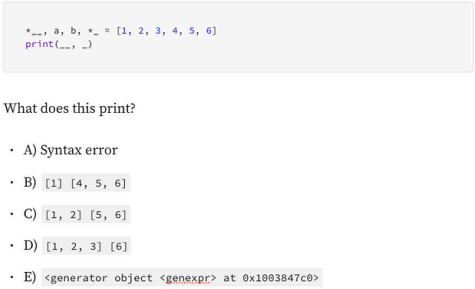


Henrique Centieiro & Bee Lee in The Generator

Top 20 GPT-4o Use Cases That Actually Improve Your Everyday...

These Are Mindblowing AI Use Cases!

Jun 26 953 13

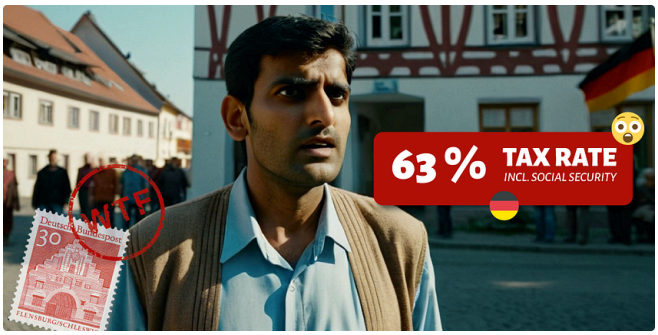


Liu Zuo Lin

You're Decent At Python If You Can Answer These 7 Questions...

No cheating pls!!

Mar 6 6K 29



Jan Kammerath

Why Tech Workers Are Fleeing Germany—A Reality Check

Over the past months and years I have seen a number of friends and colleagues leave...

Jun 23 5K 176

See more recommendations