

Title: Dry Eye Analysis Report

Author: Musa Amjad

Date: r Sys.Date()

Output: html_document

- library(tidyverse)
- library(ggplot2)
- library(knitr)
- library(readr)
- knitr::opts_chunk\$set(echo = TRUE)

Background Information

This report analyzes the Dry Eye Dataset to understand key patterns and insights related to dry eye conditions.

Research Questions

1. What are the main characteristics of patients suffering from dry eye?
2. Are there any significant correlations between different symptoms and dry eye severity?
3. How do demographic factors (age, gender) influence dry eye symptoms?

Dataset Overview

```
{r load-data}
```

```
df <- read_csv("Dry_Eye_Dataset.csv")
```

Display basic structure and summary

```
str(df) summary(df)
```

Exploratory Data Analysis

- Checking for Missing Values {r missing-values}
- colSums(is.na(df))

Variable Distributions

```
{r histograms}
```

- ggplot(df, aes(x = Age)) + geom_histogram(binwidth = 5, fill = "blue", alpha = 0.7) + theme_minimal() + ggtitle("Age

Distribution")

```
{r gender-distribution}
```

- `ggplot(df, aes(x = Gender, fill = Gender)) + geom_bar() + theme_minimal() + ggtitle("Gender Distribution")`

Data Visualizations

Symptom Severity vs. Age

```
{r severity-vs-age}
```

- `ggplot(df, aes(x = Age, y = Severity, color = Gender)) + geom_point() + theme_minimal() + ggtitle("Dry Eye Severity by Age and Gender")`

Correlation Matrix

```
{r correlation-matrix} library(GGally)
```

- `ggpairs(df[, apply(df, is.numeric)], title = "Correlation between Numeric Variables")`

Conclusion

- Key findings: The dataset provides insights into the distribution of dry eye symptoms across different demographics.
- Recommendations: Further research could explore machine learning models for predicting dry eye severity.

This concludes the exploratory analysis of the Dry Eye Dataset.

```
In [ ]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

sns.set(style="whitegrid")
```

```
In [ ]: # importing dataset
file_path = "/Dry_Eye_Dataset.csv"
df = pd.read_csv(file_path)

df.head()
```

Out[2]:

	Gender	Age	Sleep duration	Sleep quality	Stress level	Blood pressure	Heart rate	Daily steps	Physical activity	Height	...	Smc
0	F	24	9.5	2	1	137/89	67	3000	31	161	...	
1	M	39	9.6	2	3	108/64	60	12000	74	164	...	
2	F	45	5.4	1	5	134/81	95	12000	93	179	...	
3	F	45	5.4	4	5	110/90	78	19000	32	160	...	
4	F	42	5.7	3	2	99/67	72	4000	173	179	...	

5 rows × 26 columns



```
In [ ]: df.info()

df.isnull().sum()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20000 entries, 0 to 19999
Data columns (total 26 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Gender                                20000 non-null  object
1   Age                                   20000 non-null  int64
2   Sleep duration                        20000 non-null  float64
3   Sleep quality                         20000 non-null  int64
4   Stress level                          20000 non-null  int64
5   Blood pressure                        20000 non-null  object
6   Heart rate                            20000 non-null  int64
7   Daily steps                           20000 non-null  int64
8   Physical activity                     20000 non-null  int64
9   Height                                20000 non-null  int64
10  Weight                                20000 non-null  int64
11  Sleep disorder                         20000 non-null  object
12  Wake up during night                  20000 non-null  object
13  Feel sleepy during day                 20000 non-null  object
14  Caffeine consumption                  20000 non-null  object
15  Alcohol consumption                   20000 non-null  object
16  Smoking                               20000 non-null  object
17  Medical issue                         20000 non-null  object
18  Ongoing medication                    20000 non-null  object
19  Smart device before bed                20000 non-null  object
20  Average screen time                   20000 non-null  float64
21  Blue-light filter                     20000 non-null  object
22  Discomfort Eye-strain                  20000 non-null  object
23  Redness in eye                        20000 non-null  object
24  Itchiness/Irritation in eye           20000 non-null  object
25  Dry Eye Disease                       20000 non-null  object
dtypes: float64(2), int64(8), object(16)
memory usage: 4.0+ MB
```

Out[3]:

	0
Gender	0
Age	0
Sleep duration	0
Sleep quality	0
Stress level	0
Blood pressure	0
Heart rate	0
Daily steps	0
Physical activity	0
Height	0
Weight	0
Sleep disorder	0
Wake up during night	0
Feel sleepy during day	0
Caffeine consumption	0
Alcohol consumption	0
Smoking	0
Medical issue	0
Ongoing medication	0
Smart device before bed	0
Average screen time	0
Blue-light filter	0
Discomfort Eye-strain	0
Redness in eye	0
Itchiness/Irritation in eye	0
Dry Eye Disease	0

dtype: int64

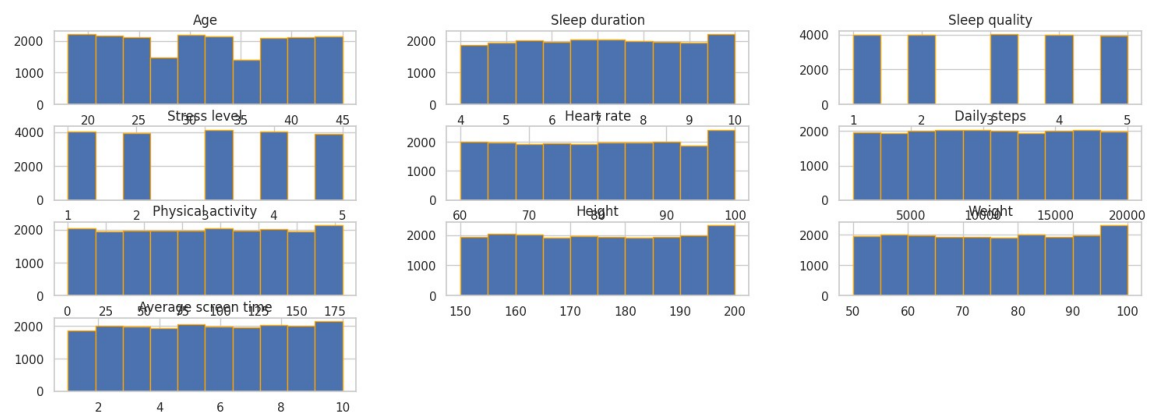
```
In [ ]: df_cleaned = df.copy()

df_cleaned.describe()
```

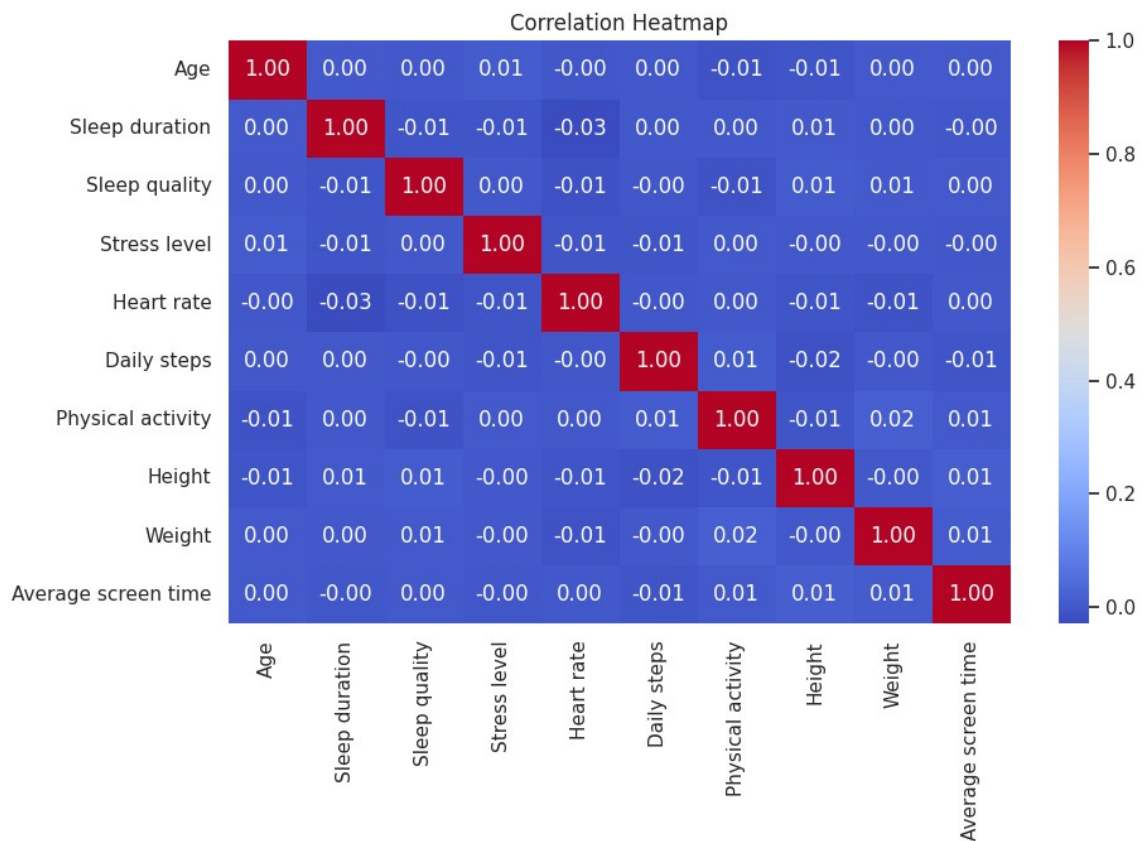
Out[7]:

	Age	Sleep duration	Sleep quality	Stress level	Heart rate	Daily steps
count	20000.000000	20000.000000	20000.000000	20000.000000	20000.000000	20000.000000
mean	31.422800	6.998245	2.997250	2.993750	79.912200	10536.900000
std	8.103717	1.731723	1.412283	1.407235	11.808279	5752.729186
min	18.000000	4.000000	1.000000	1.000000	60.000000	1000.000000
25%	24.000000	5.500000	2.000000	2.000000	70.000000	6000.000000
50%	31.000000	7.000000	3.000000	3.000000	80.000000	11000.000000
75%	39.000000	8.500000	4.000000	4.000000	90.000000	16000.000000
max	45.000000	10.000000	5.000000	5.000000	100.000000	20000.000000

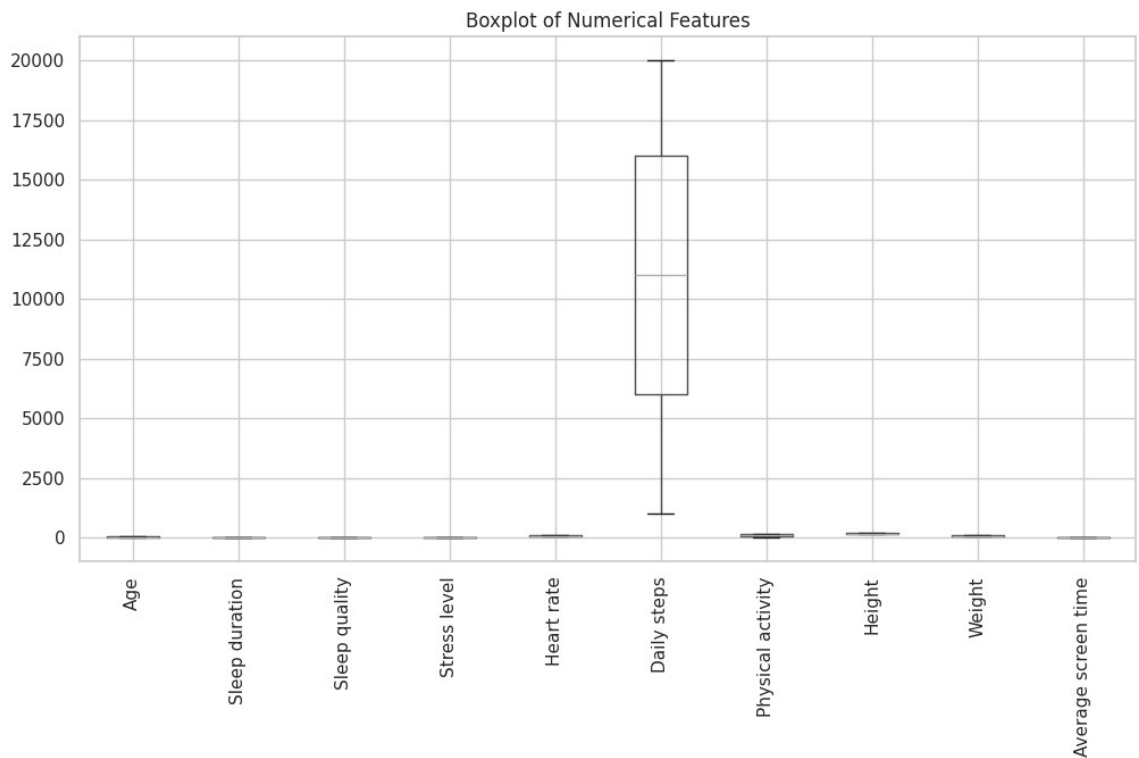
```
In [ ]: df_cleaned.hist(figsize=(18, 6), bins=10, edgecolor="Orange")
plt.show()
```



```
In [ ]: numeric_df = df_cleaned.select_dtypes(include=['number'])
plt.figure(figsize=(10, 6))
sns.heatmap(numeric_df.corr(), annot=True, cmap="coolwarm", fmt=".2f")
plt.title("Correlation Heatmap")
plt.show()
```



```
In [ ]: plt.figure(figsize=(12, 6))
df_cleaned.boxplot(rot=90)
plt.title("Boxplot of Numerical Features")
plt.show()
```



```
In [ ]:
```

```
File "<ipython-input-1-51deb2d1e312>", line 2
    import matplotlib as np import numpy.Pyplot as PLT
                                ^
SyntaxError: invalid syntax
```

```
In [ ]:
```