

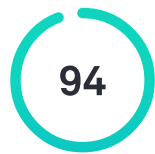
Untitled

by Manjot Singh

General metrics

15,608	2,340	263	9 min 21 sec	18 min 0 sec
characters	words	sentences	reading time	speaking time

Score



This text scores better than 94% of all texts checked by Grammarly

Writing Issues

42		42
Issues left	Critical	Advanced

Unique Words

Measures vocabulary diversity by calculating the percentage of words used only once in your document

31%
unique words

Rare Words

Measures depth of vocabulary by identifying words that are not among the 5,000 most common English words.

44%
rare words

Word Length

Measures average word length

5

characters per word

Sentence Length

Measures average sentence length

8.9

words per sentence

Untitled

Abstract—Sleep plays a critical component in human health and well-being. Adequate, high-quality sleep is essential for physical, emotional, and mental functioning. Poor quality and quantity of sleep can lead to a variety of issues including decreased productivity and increased risk of chronic conditions such as obesity, diabetes, and heart-related diseases. Sleeping is intricately linked with various body functions such as metabolism and has a significant impact on the functioning of the immune system. Optimal sleep promotes a greater immune system and improves defense against infections and other diseases. Furthermore, disorganized sleeping patterns can seriously impact an individual's mental health, creating issues like anxiety and depression. Therefore, promoting superior sleep health through analytics, campaigns, and lifestyle adjustments can help lead to a better understanding of the importance of quality sleep and make informed decisions to enhance one's sleep habits and lifestyle choices.

Keywords: Sleep, Lifestyle, Chronic Conditions, Databases, Data Analytics, SQL, E-R Diagrams

I. INTRODUCTION

The goal of the Sleep Health and Lifestyle Analysis Database Project aims to comprehend how sleep patterns and health outcomes are related. The study intends to find patterns and connections that might guide treatments and encourage

healthy sleep habits by utilizing a broad dataset that includes person-specific information, occupational details, and metrics linked to sleep. Substantial analysis employing database approaches will be employed throughout the project.

II. SIGNIFICANCE TO THE REAL WORLD

This research advances knowledge of various factors affecting the health of sleep. The analysis can provide and recommend:

- Public health campaigns that support relaxed sleep practices.
- Make specific recommendations for workplace well-being.
- Determine the sleep schedules needed for various professions.
- Creation of personalized sleep-improvement plans.
- Studies on the identification and management of sleep disorders
- Educate policymakers on the benefits of sleep.

III. LITERATURE REVIEW

The study integrates data from previous research on the on the link between optimal sleep and the physical health of an individual. The project's plan for data processing and interpretation is based on this evaluation of the literature. Some of the existing research is listed as follows:

- 1) Lifestyle Factors and Sleep Health across the Lifespan
- Joseph M. Dzierzewski
- 2) Measuring Subjective Sleep Quality: A Review - Marco Fabbri
- 3) Effects of morning and evening physical exercise on

subjective and objective sleep quality: an ecological study - Mathias Goldberg

4) Rethinking the sleep-health link - Lisa Matricciani

5) Sleep health epidemiology in low and middle-income countries: a systematic review and Meta-analysis of the prevalence of poor sleep quality and sleep duration - Guido Simonelli

6) Long sleep duration and health outcomes: A systematic review, meta-analysis, and meta-regression - Maki Jike

7) A meta-analysis of the association between insomnia with objective short sleep duration and risk of hypertension - Yanyuan Dai

IV. METHODOLOGY

We have collected Sleep Health and Lifestyle Analysis data from Kaggle and synthesized it as per our requirements. Based on this dataset, we created an initial ER Diagram which would be further enhanced based on our data modeling.

We established a connection to MySQL Server using MySQL.connector library in Python, created and loaded the tables needed for our analysis from the original dataset. These new tables were created to normalize the themes of the tables and for usage in Data warehousing. We utilized MongoDB to compare the querying systems of both types of databases and gain a better insight into the database which would be better for our analysis.

We created a MongoDB cluster in Atlas to give all team members access to the database and query it at any time.

Once our data modification was completed, we utilized Google BigQuery for Data warehousing purposes. This allowed us to store past queries and historical data in one place while also being able to perform new analyses on it.

We performed analysis using MySQL, MongoDB and Google BigQuery. This allowed us to compare all the tools to gain detailed insights into the effects of Lifestyle factors on Sleep Health.

Visualizations were created using Matplotlib and Seaborn by loading the data queried by MySQL queries into a Pandas Dataframe. We utilized Github to coordinate the progress and maintain version control and connected through Google Meets several times to discuss the next steps.

V. DATABASE STRUCTURE

Our Sleep Health and Lifestyle data has been broken into 5 tables.

1. person
2. sleepmetrics
3. metrics
4. physicalactivity
5. fact

The main table is person.

A. person

- person ID: An identifier for each individual in the dataset.
- gender: The gender of the person: male or female.
- age: The age of the person in years.

- occupation: The occupation or profession of the person.

B. sleepmetrics

- metrics ID: An identifier for each individual in the dataset.
- sleepDuration: The # of hours the person sleeps per day.
- qualityofSleep: A rating to scale the quality of sleep, ranging from 1-10.
- sleepDisorder: Identifies if a sleep disorder is present or absent in the person - none, insomnia, sleep apnea.

C. metrics

- metrics ID: Identifier for a set of metrics or measurements
- stressLevel: A rating scale for stress level in which they experienced, ranging from 1-10.
- BMICategory: The BMI category of the person: underweight, normal, or overweight.
- bloodPressure: The blood pressure measurement of the person: systolic over diastolic pressure.
- heartRate: The resting heart rate of the person in beats per minute.

D. physicalactivity

- physicalActivity ID: Identifier for a physical activity-related record
- physicalActivityLevel: The # of minutes the person engages in physical activity on a daily.
- dailySteps: The of steps the person takes per day.

E. fact

- fact ID: Identifier for a fact observation
- person ID: Identifier to identify a person

- physicalActivity ID: Identifier for a physical activity-related record
- metrics ID: Identifier for a set of metrics or measurements

VI. ENTITY RELATIONSHIP DIAGRAM

An E-R diagram helps visualize the data requirements and specifications in a database.

Fig. 1. Entity Relationship Diagram

Figure 1. shows the E-R diagram with 5 tables.

VII. DESIGN STEPS

VIII. TOOLS & TECHNOLOGY STACK

1. Data Source: Kaggle
2. Data Modeling: mySQLWorkbench, MySQL
3. Database: MongoDB
4. Visualization: Seaborn, Matplotlib
5. DB Connectivity: Python – MySQL
6. ETL and Data Warehouse: Google BigQuery
7. Coding: Visual Studio, Jupyter Notebook
8. Documentation: Microsoft Word, PowerPoint, LATEX
9. Version Control: GitHub
10. Agile: Trello, Lucidcharts

IX. TECHNICAL DIFFICULTY

Throughout the project, We came across many challenges.

A few of the notable ones are listed below: After breaking up the dataset into 4 tables, it was challenging to create primary keys for new tables and establish a relationship between them. Faced connection challenges while connecting the MySql database to the Python notebook for data

visualizations.

X. DATA ANALYSIS / VISUALIZATION

We have done Data Visualization using the following python libraries: Numpy, Pandas, Matplotlib, and Seaborn.

Fig. 2. Distribution of Sleep Disorders by Age Groups.

In Figure 2, a stacked bar chart shows the number of persons with different sleep disorders distributed across different Age groups. We can see that the majority of people in age group 30 to 39 are not diagnosed with any sleep disorder. Fifty percent of people in the age group 40 to 49 are suffering from insomnia.

Fig. 3. Average Sleep Duration by Occupation.

In Figure 3, this bar chart shows the average sleep duration of people working across different professions. Sales representatives and scientists sleep less than 6 hours on average which is the least across professions.

XI. INNOVATION

This project can bring various benefits to the field of healthcare and wellness using database analytics of sleep health and lifestyle aspects. It helps to make many sleep health-related decisions. A few are listed below.

Customized Advice: The project may be able to provide people with planned advice on how to enhance their general well-being and quality of sleep by evaluating data on lifestyle choices, sleep habits, and health outcomes.

Behavioral Insights: Conducting extensive dataset analyses can yield important insights on how lifestyle decisions, the

Fig. 4. Distribution of BMI Categories by Occupations.

In Figure 4, the stacked graph shows we can infer that most of the Nurses, Teachers, salespersons, and scientists are overweight. We can also see that Doctors, Engineers, Accountants and Lawyers have a Normal BMI.

Fig. 5. Sleep Duration and Stress Level by BMI Category

In Figure 5, the graph indicates as sleep duration decreases there is an increase in stress level. Most of the people sleeping for less duration are overweight.

quality of sleep and general health are related to one another.

Develop behavioral interventions to encourage healthy habits using the information provided here.

Research Advancements: By offering researchers a dataset for studies on the effects of numerous variables on sleep health and well-being, the project may make a significant contribution to the fields of lifestyle medicine and sleep medicine.

Telemedicine: By combining database analytics with telemedicine technology, medical professionals may deliver virtual consultations and interventions by remotely monitoring people's sleep habits and other lifestyle variables.

Wearable Technology: By combining data from fitness trackers and smartwatches, the project can collect real-time

Fig. 6. Relationship Between Sleep Duration and Stress Level by Gender

Figure 6. showcases that females are the majority among sleep-deprived people. This graph also shows that as sleep duration increases, Stress level decreases.

data on lifestyle behaviors and sleep measurements, which improves the analytics' accuracy.

One may develop a useful tool that will help people, healthcare professionals, academics, and policymakers improve health outcomes and quality of life by adding these cutting-edge components to a project that focuses on database analytics of sleep health and lifestyle aspects.

XII. PAIR PROGRAMMING

Our team was able to achieve pair programming by collaborating and troubleshooting on a weekly basis through WhatsApp, meeting in person, and on Google Meets. Our team utilized MongoDB, a real-time online code-sharing text editor for developers. Most of our code was done through MongoDB and saved within Google Docs.

XIII. TEAMWORK

Each team member played a vital role in contributing to the project. We used Trello to track tasks, meetings, and to practice agile workflow. We took the iterative approach as we learned new topics each week. Through group discussions, our team was able to delegate and balance each task.

Role & Responsibilities:

- 1.2. • Chelsea Jaculina - Data Cleaning, Data Pre-Processing, Data Analysis, Agile Master, Documentation in LaTeX, PowerPoint
- Manjot Singh - Data Cleaning, Data Pre-Processing, Data Extraction, Data Loading, Data Modeling, Python, MongoDB
- Sai Prasad Thalluri - Data Visualization, PowerPoint, Data Analysis, Documentation in LaTeX, NoSQL Queries

XIV. AGILE/SCRUM

Google Meet Links:

- 1/31: Meeting 1
- 2/5: Meeting 2
- 2/6: Meeting 3
- 2/7: Meeting 4
- 3/12: Meeting 5
- 3/22: Meeting 6
- 4/10: Meeting 7
- 4/11: Meeting 8
- 4/22: Meeting 9
- 4/25: Meeting 10
- 4/26: Meeting 11

XV. LESSONS LEARNED/KEY LEARNINGS

1. Creating an ER diagram helped us to understand data and establish relationships between tables.
2. Data cleaning, validation, and pre-processing are important to minimize errors and to ensure the integrity and quality of data is ensured.

XVI. GRAMMARLY

After finalizing our report, our team used Grammarly to ensure our paper was spelling, grammar, punctuation, and mistakes free.

XVII. BEST PRACTICES

- Breaking up the data for better understanding and data processing.
- Connecting to the MySQL database eases the process of

creating data visualizations

XVIII. VERSION CONTROL

Our team uploaded our project files and materials to Github for version control. The repository has been made publicly accessible to foster transparency, keep track of project progress, and for any future work collaboration. The Github repository link can be found in the Appendix.

XIX. CONCLUSION

Overall, this analysis allowed us to examine the relationships between sleep duration, quality of sleep, and sleep order, with corresponding metrics such as occupation, gender, and

XX. FUTURE WORK

Our team can further expand our analysis with some features such as:

1. Predictive Analytics: Incorporating machine learning algorithms will elevate this project even further. By predicting health issues or risks using one's sleep data and their lifestyle habits could help prevent health issues in the long run.
2. Data Security: With health data being sensitive information, further work on this project can include access controls, encryption, and providing robust privacy measures.

XXI. REFERENCES

Dataset <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset/data>

- 1) Fabbri, M.; Beracci, A.; Martoni, M.; Meneo, D.; Tonetti, L.; Natale, V. Measuring Subjective Sleep Qual-

ity: A Review. *Int. J. Environ. Res. Public Health* 2021, 18, 1082. <https://doi.org/10.3390/ijerph18031082>

2) Dzierzewski, J.M.; Sabet, S.M.; Ghose, S.M.; Perez, E.; Soto, P.; Ravyts, S.G.; Dautovich, N.D. Lifestyle Factors and Sleep Health across the Lifespan. *Int. J. Environ. Res. Public Health* 2021, 18, 6626. <https://doi.org/10.3390/ijerph18126626>

3) Goldberg, M., Pairot de Fontenay, B., Blache, Y., & Debarnot, U. (2024). Effects of morning and evening physical exercise on subjective and objective sleep quality: an ecological study. *Journal of Sleep Research*, 33(1), e13996. <https://doi.org/10.1111/jsr.13996>

4) Lisa Matricciani, Yu Sun Bin, Tea Lallukka, Erkki Kronholm, Melissa Wake, Catherine Paquet, Dorothea Dummer, Tim Olds, Rethinking the sleep-health link, *Sleep Health*, Volume 4, Issue 4, 2018, Pages 339-348, ISSN 2352-7218, <https://doi.org/10.1016/j.sleh.2018.05.004>

5) Guido Simonelli, Nathaniel S. Marshall, Antigone Grillakis, Christopher B. Miller, Camilla M. Hoyos, Nick Glozier, Sleep health epidemiology in low and middle-income countries: a systematic review and meta-analysis of the prevalence of poor sleep quality and sleep duration, *Sleep Health*, Volume 4, Issue 3, 2018, Pages 239-250, ISSN 2352-7218, <https://doi.org/10.1016/j.sleh.2018.03.001>

6) Maki Jike, Osamu Itani, Norio Watanabe, Daniel J. Buysse, Yoshitaka Kaneita, Long sleep duration

and health outcomes: A systematic review, meta-analysis and meta-regression, Sleep Medicine Reviews,

Volume 39, 2018, Pages 25-36, ISSN 1087-0792,

<https://doi.org/10.1016/j.smr.2017.06.011>

7) Yanyuan Dai, Alexandros N. Vgontzas, Le Chen, Dan-

dan Zheng, Baixin Chen, Julio Fernandez-Mendoza,

Maria Karataraki, Xiangdong Tang, Yun Li, A meta-

analysis of the association between insomnia with objective short sleep

duration and risk of hypertension, Sleep

Medicine Reviews, Volume 75, 2024, 101914, ISSN

1087-0792, <https://doi.org/10.1016/j.smr.2024.101914>

XXII. APPENDIX

LINKS:

- Sleep, Health, and Lifestyle Dataset Link
- Version Control: Github Link
- Agile: Trello Board Link
- Term Project Rubric Link