

CSE523 Machine Learning

Weekly Report 3

**Athlete Profiling for Division I Basketball Players**

Submitted to faculty: Mehul Raval

Date of Submission: 02-03-24

Student Details

| Roll No. | Name of the Student |
| --- | --- |
| AU2140024 | Aagam Shah |
| AU2140133 | Dhairya Shah |
| AU2140159 | Aayushi Shah |
| AU2140182 | Aanal Dobariya |

2023-2024 (Winter Semester)

### **Aim:**

Data Visualization for feature engineering

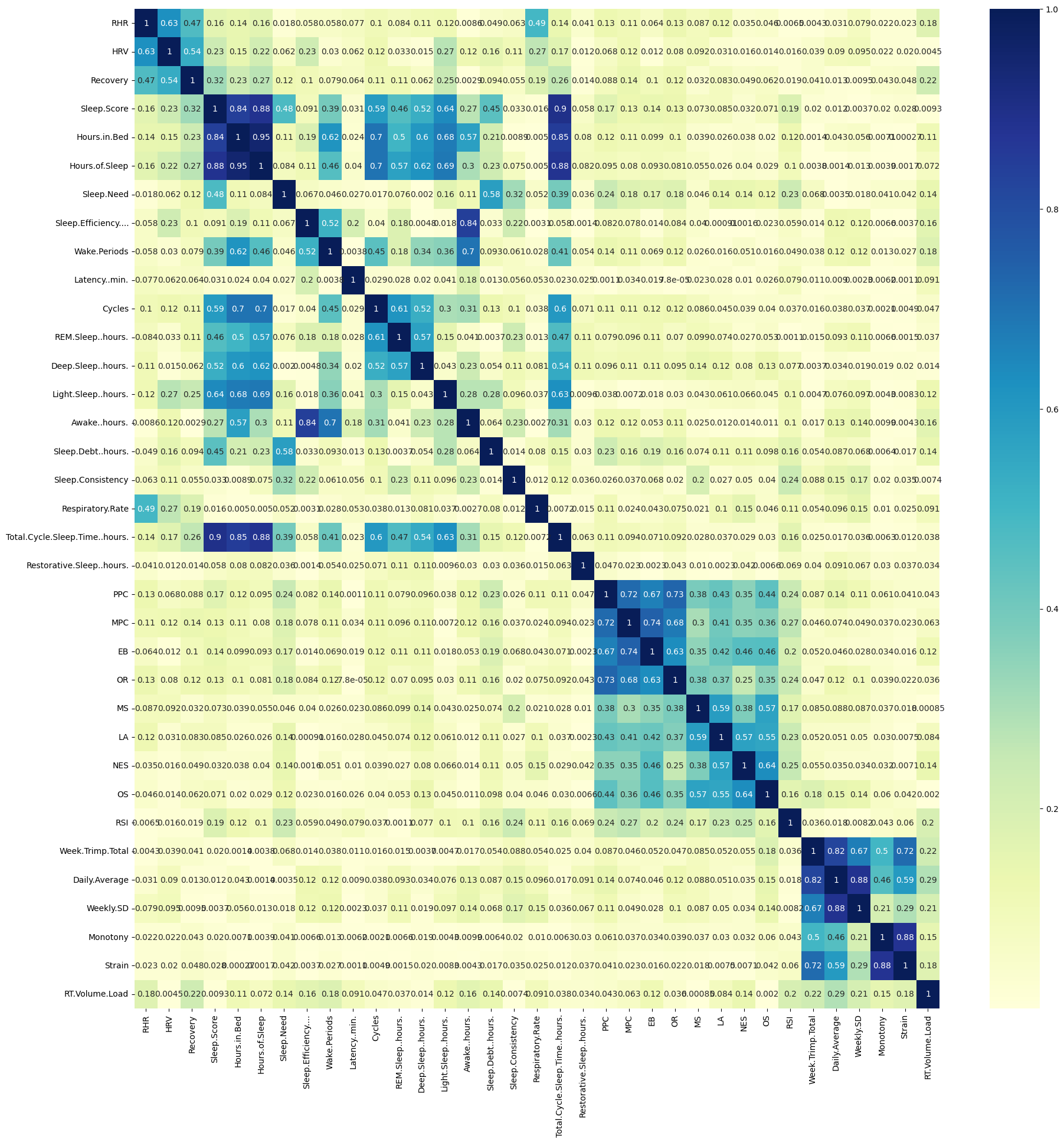
### **Introduction:**

After completing the data imputation process, mean imputation, and multivariate imputation by chained equation (MICE) with random forest regressor as an estimator, we shifted our focus to data visualization and formulated our methodology. We found three primary modalities in the dataset by utilizing correlation heat maps and descriptive statistics, which gave us important information about feature interdependencies. Equipped with these visual aids, we formulated a customized strategy to reveal implicit trends and derive significant data.

### **Work Completed:**

* Correlation Heat Map

As shown in the heat map below, the three blue-shaded clusters have a higher correlation representing the three primary modalities, which are sleep and recovery patterns, training load statistics, and cognitive state



* Inferences from descriptive statistics

Out of the 115 features given, we found a high correlation between the features in three modalities, as discussed before, and found features with high or low variance. So to provide these features with particular weightage, using these inferences, we have planned to do feature engineering like feature clustering or factor analysis.

### **Next steps and goals:**

* Approach formulation
* Factor analysis and feature engineering (Feature clustering)

### **Conclusion:**

Correlation and descriptive statistics gave deeper insights into the dataset distributions and interdependencies. According to these data visualization and descriptive statistics, we concluded that feature engineering should be done on the imputed data; further feature clustering may also be helpful. Furthermore, we decided to follow a particular approach/methodology for athlete profiling and clustering.

Link to the collab : <https://colab.research.google.com/drive/1WMOik8eT5PZ26DLBYIn0FNncvzSAK5GA?usp=sharing>

### **References:**

1. Senbel, S., Sharma, S., Raval, M. S., Taber, C., Nolan, J., Artan, N. S., ... & Kaya, T. (2022). Impact of sleep and training on game performance and injury in division-1 women’s Basketball Amidst the Pandemic. Ieee Access, 10, 15516-15527.

<https://ieeexplore.ieee.org/abstract/document/9690164/>

1. Taber, C.B., Sharma, S., Raval, M.S. et al. A holistic approach to performance prediction in collegiate athletics: player, team, and conference perspectives. Sci Rep 14, 1162 (2024). <https://doi.org/10.1038/s41598-024-51658-8>
2. Sharma, S. U., Divakaran, S., Kaya, T., & Raval, M. (2022, July). A Hybrid Approach for Interpretable Game Performance Prediction in Basketball. In 2022 International Joint Conference on Neural Networks (IJCNN) (pp. 01-08). IEEE. <https://ieeexplore.ieee.org/abstract/document/9892583/?casa_token=Ye3GQJ1JpD0AAAAA:GSlyds24pIa__7Od6UBSNs8nugbEwLvCbI8vG6w-YMYrFX2O-TUlUDo4xej3ulJUAvmO_4ij0J36>