Date: 23rd Mar 2024

**CSE 523: Machine Learning**

**Weekly Report 7**

**Project 7: Athlete Statistics Visualization and Prediction**

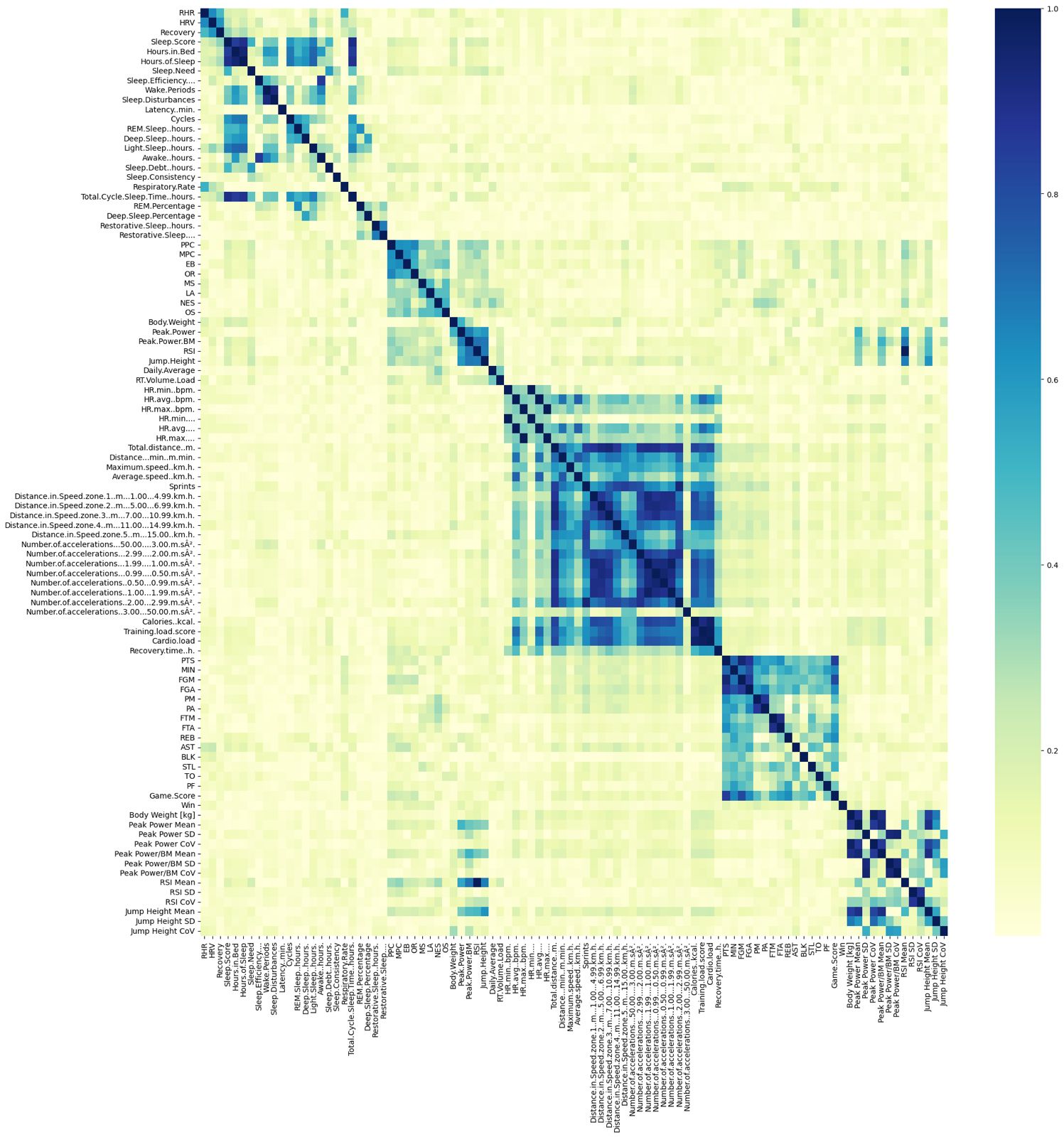
**Team: Byte Busters**

|  |  |
| --- | --- |
| **Enrolment No** | **Name** |
| AU2140060 | Deval Darji |
| AU2140085 | Krishn Patel |
| AU2140103 | Devanshi Shah |
| AU2140178 | Vedant Patel |

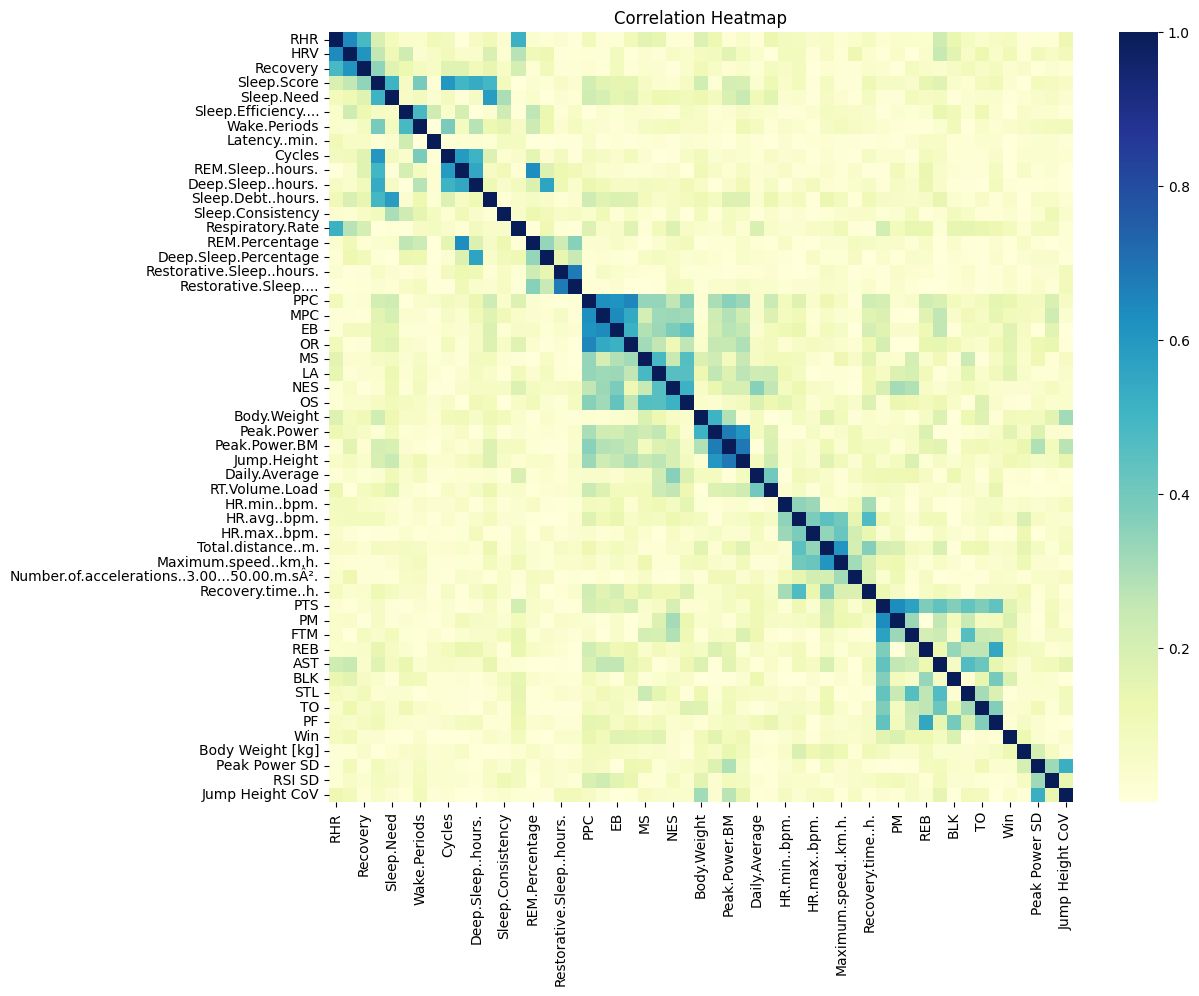
* **Until now:**
* We have merged all the data given to us in different sheets to one common sheet.
* In that we had 113 features and more than 6223 records.
* But the data had almost 50% null rows and more than 15% null columns.
* So, after data imputation using KNN imputation we got a more cleaned data.
* This final imputed data has 96 features and 3111 records.

PTO

* **Feature Dropping**
* We first plotted the heat map from the imputed data.



* We could easily see there were many features which were highly corelated with each other and increased data redundancy in the dataset.
* So, we dropped all the highly correlated features (threshold > 0.7) and reduced the data redundancy.
* By this we brought no. of features from **96 to 53.**
* The new heat map:



* As we can see there are no more features which are highly correlated with each other, and we have successfully dropped the redundant features from 96 to 53 features.
* **Future goal**:
* For the next week, we plan to use PCA and Factor Analysis to further reduce the dimensionality of the dataset.
* And then use linear regression to get some output.