Date: 30th Mar 2024

**CSE 523: Machine Learning**

**Weekly Report 8**

**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.
* We then dropped the features and reduced it to 53 features and 3111 records with the help of feature dropping technique.

**This week:**

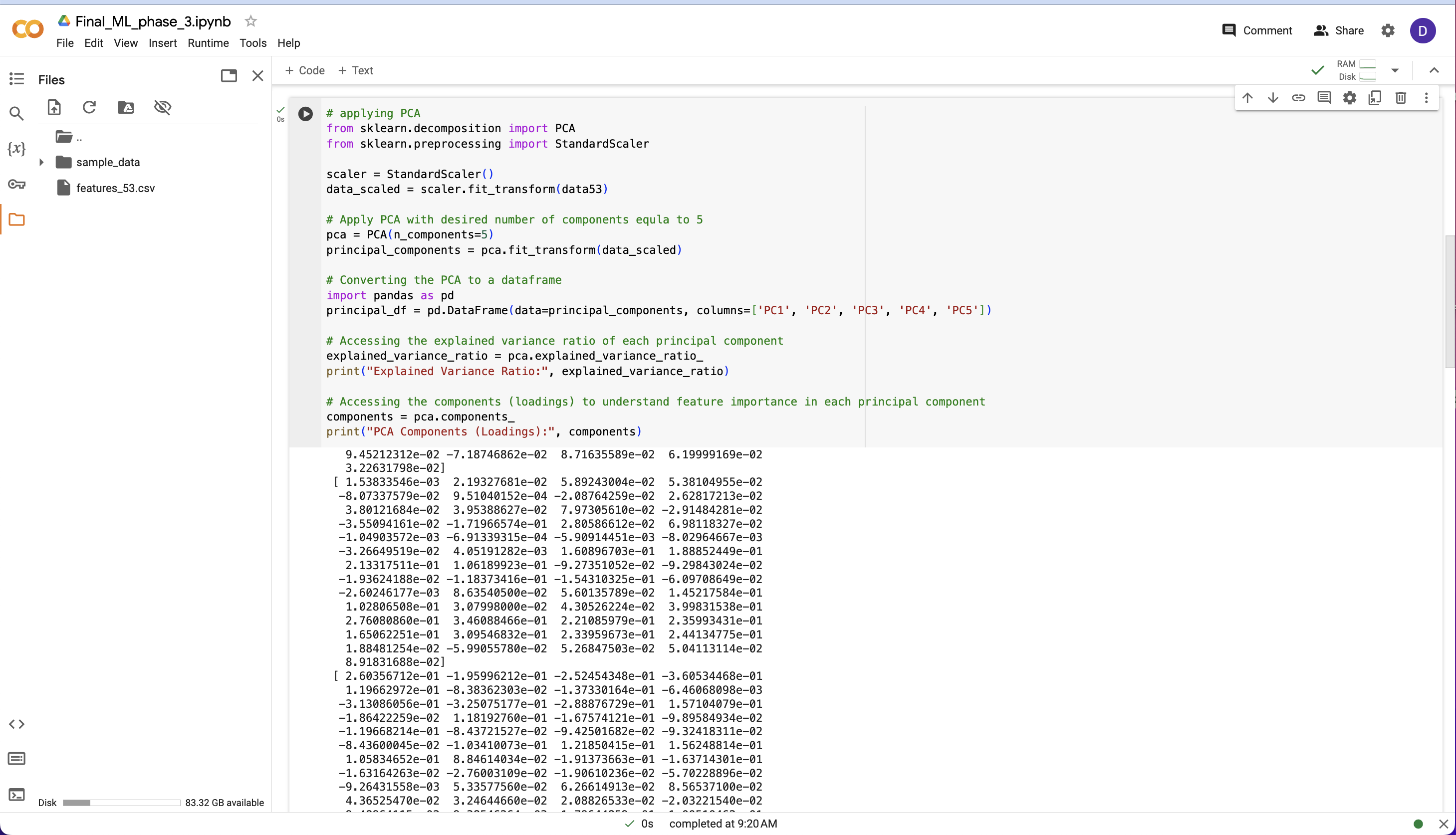
We applied PCA on these 53 features and extracted 5 features which are asked in the problem statement. We then proceeded to calculate the performance score using linear regression on both the dataset of 5 features and 53 features.

This is the MSE before applying PCA.

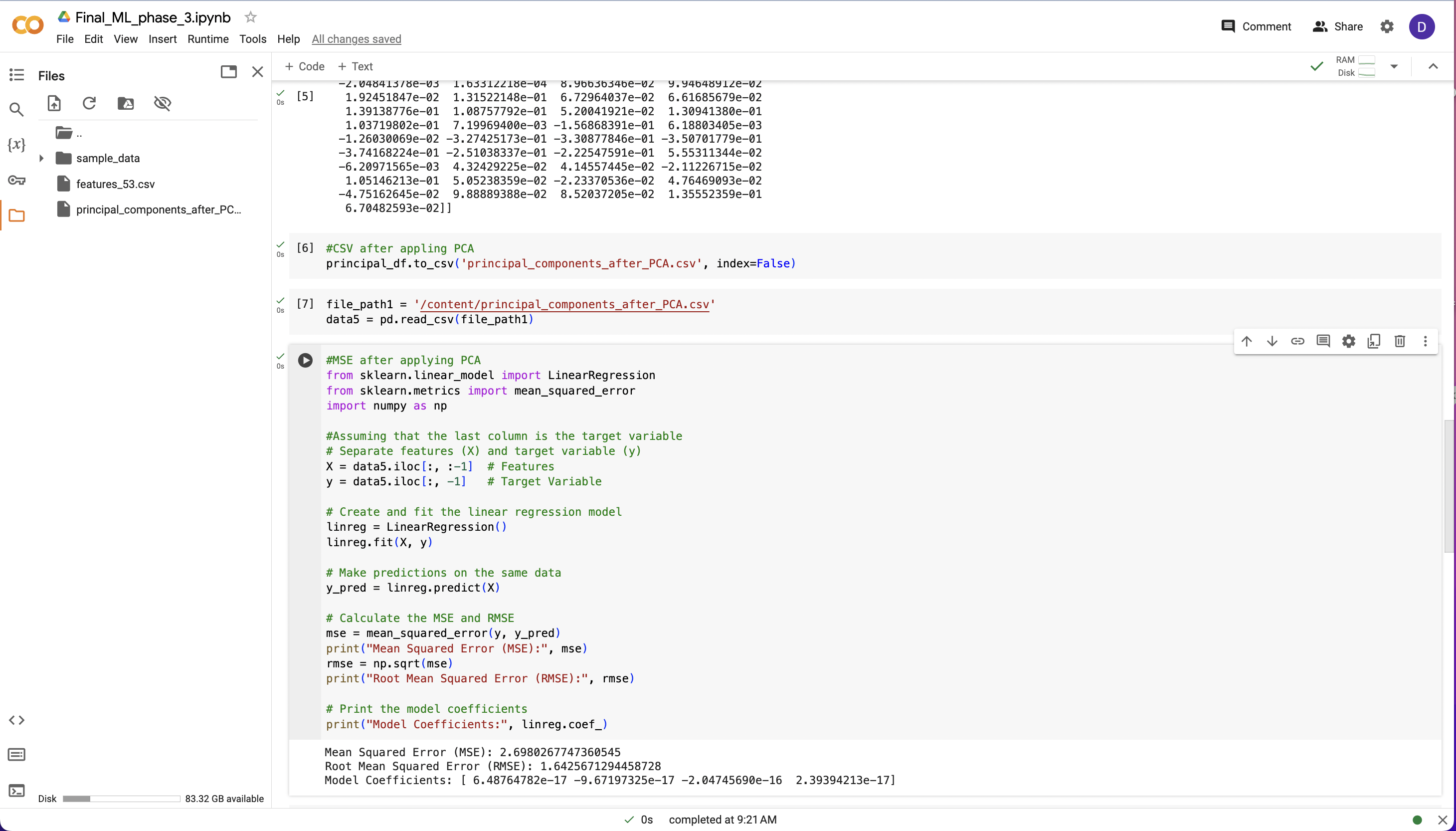
A screenshot of a computer program

Description automatically generated

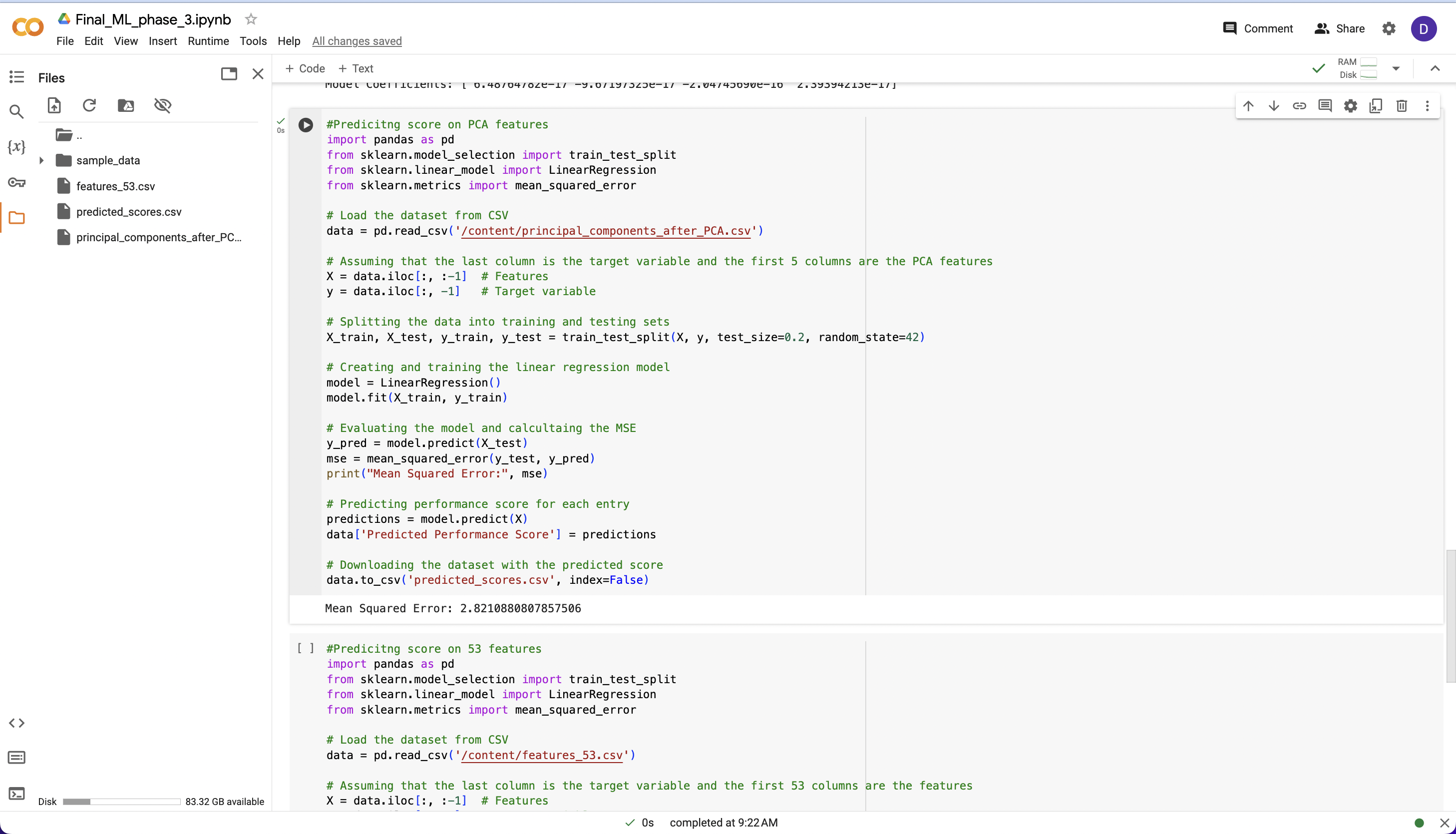
This is applying PCA.



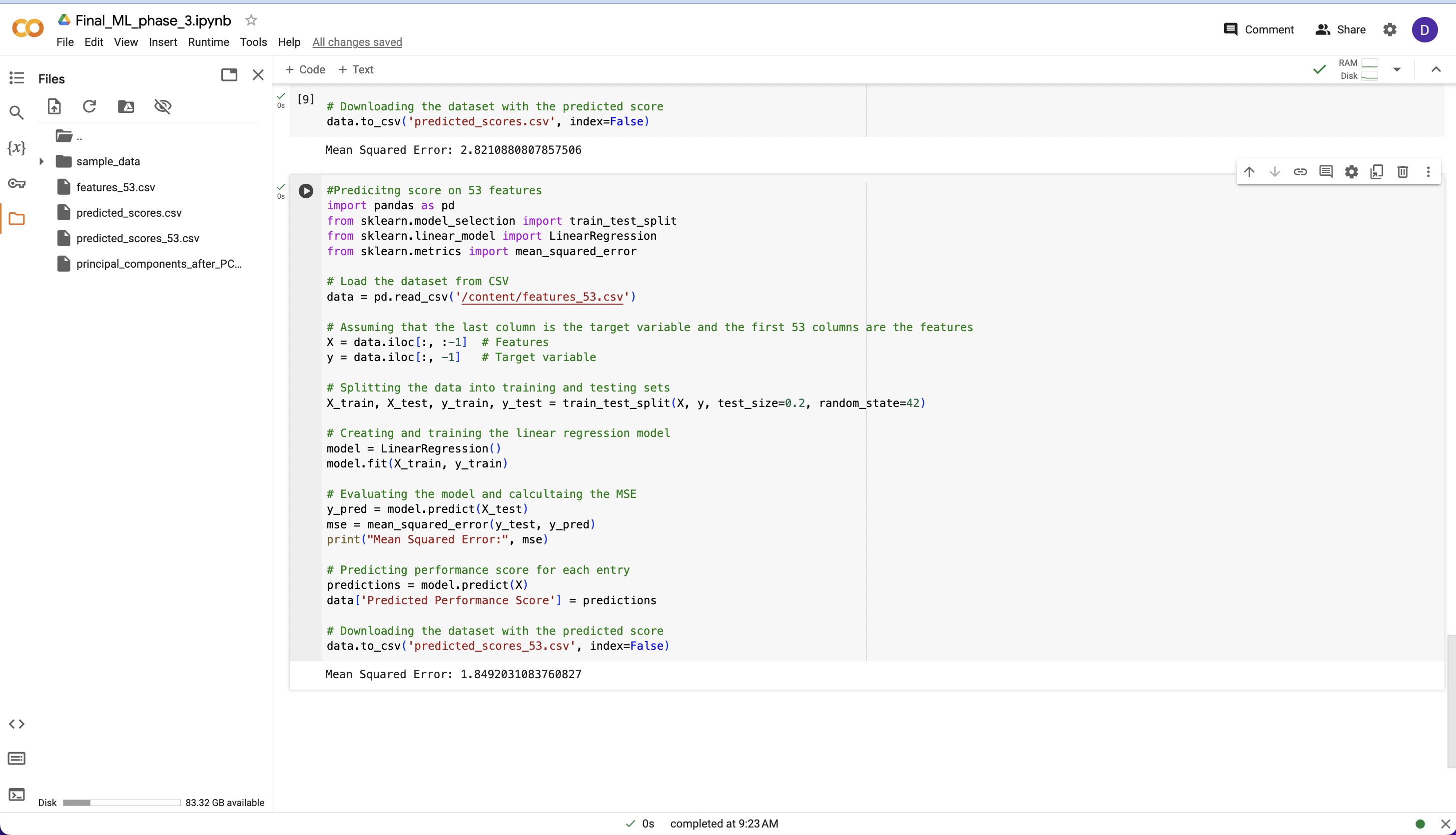
This is MSE after applying PCA.



Predicting score on PCA features.



Predicting score on 53 features.



Now for the next week, we would like to work on the model fitting so that the MSE is less, we will also start building an interactive dashboard for the same.