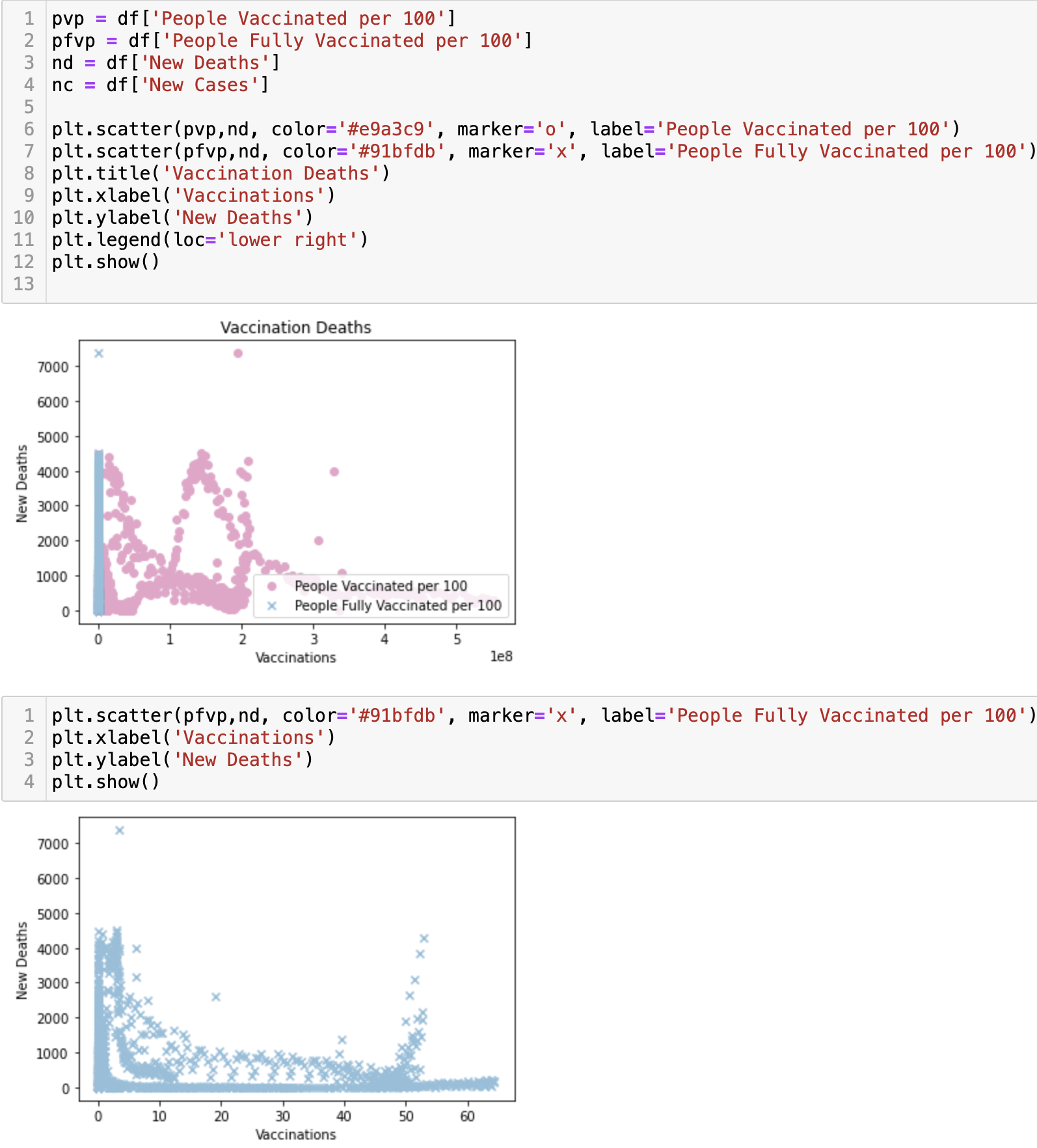
**Part A**

ziyu9240(510119533):

1. Data Description:

The Dataset used here contained several aspects related to vaccinations, recorded relevant data from 22nd,Jan. 2020 to 10th Sep. 2021. This section aims for making prediction the relationships between ‘People Fully vaccinated per 100’ (as input/independent variables) and ‘New deaths’(as output/dependent variables)

1. Code for Prediction:
2. Import sklearn, pandas, target csv,etc.
3. Observe the relationships between ‘People Fully vaccinated per 100’ and ‘New deaths’ , (discrete quantitative for each variables )



1. Approaches:
2. Using Linear regression to predict

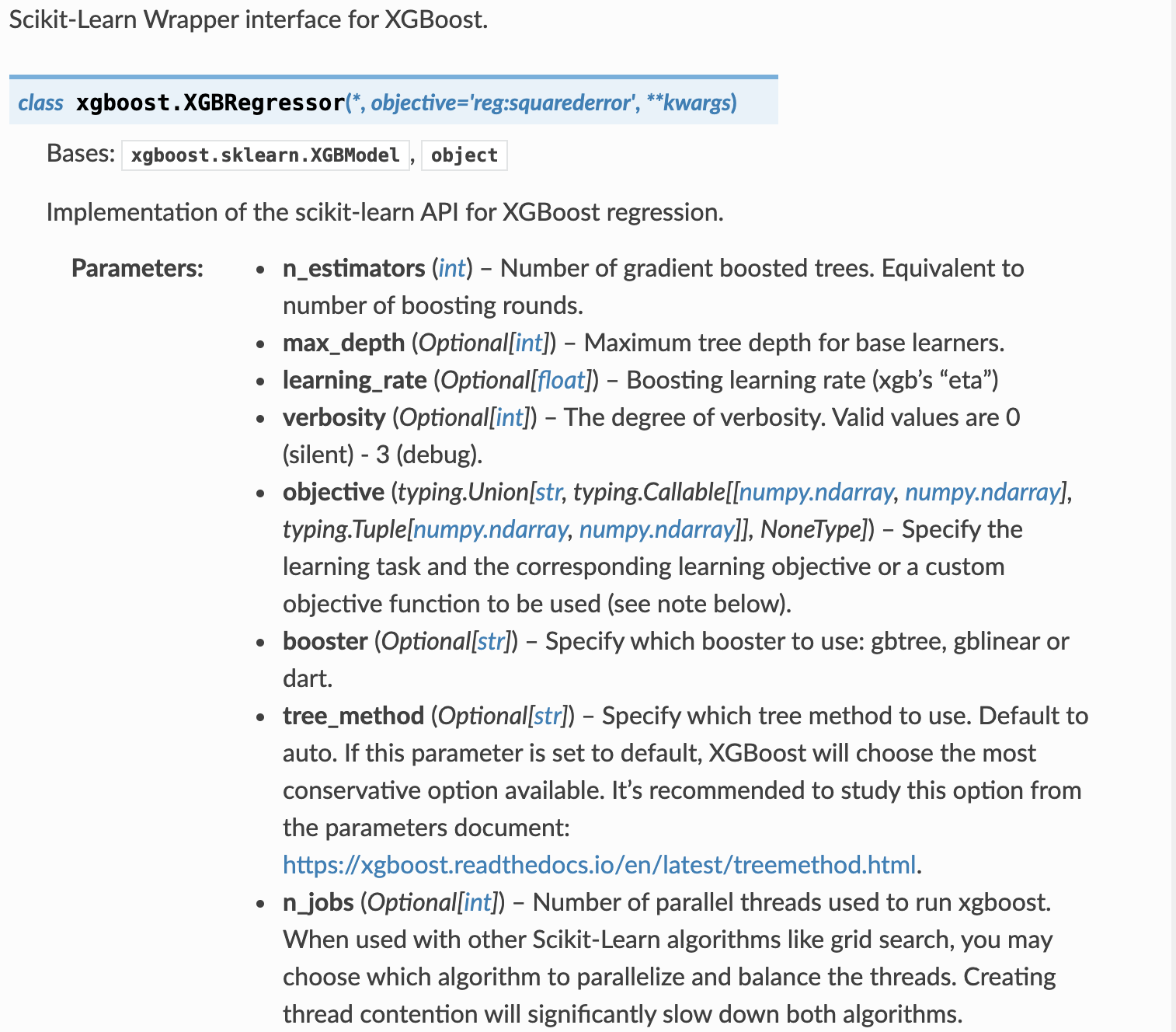


1. Using XGBoost Regressor to predict:

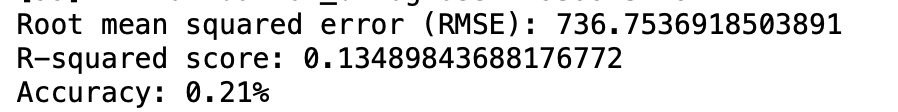
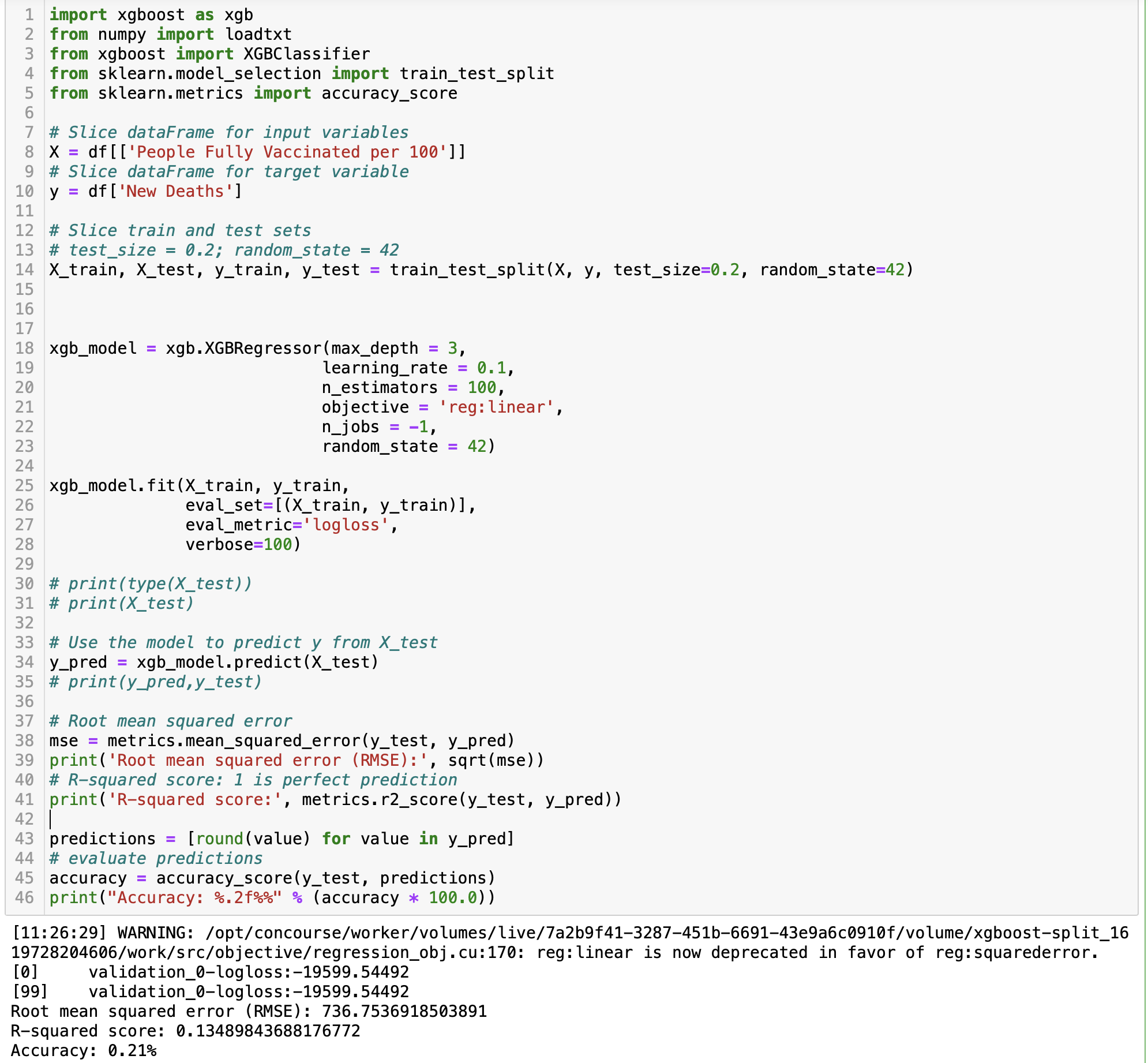
*‘*[*XGBoost*](https://xgboost.ai/) *is a decision-tree-based ensemble Machine Learning algorithm that uses a* [*gradient boosting*](https://en.wikipedia.org/wiki/Gradient_boosting) *framework. In prediction problems involving unstructured data (images, text, etc.) artificial neural networks tend to outperform all other algorithms or frameworks. However, when it comes to small-to-medium structured/tabular data, decision tree based algorithms are considered best-in-class right now.*

*’https://towardsdatascience.com/https-medium-com-vishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d*

1. Hyperparameters:

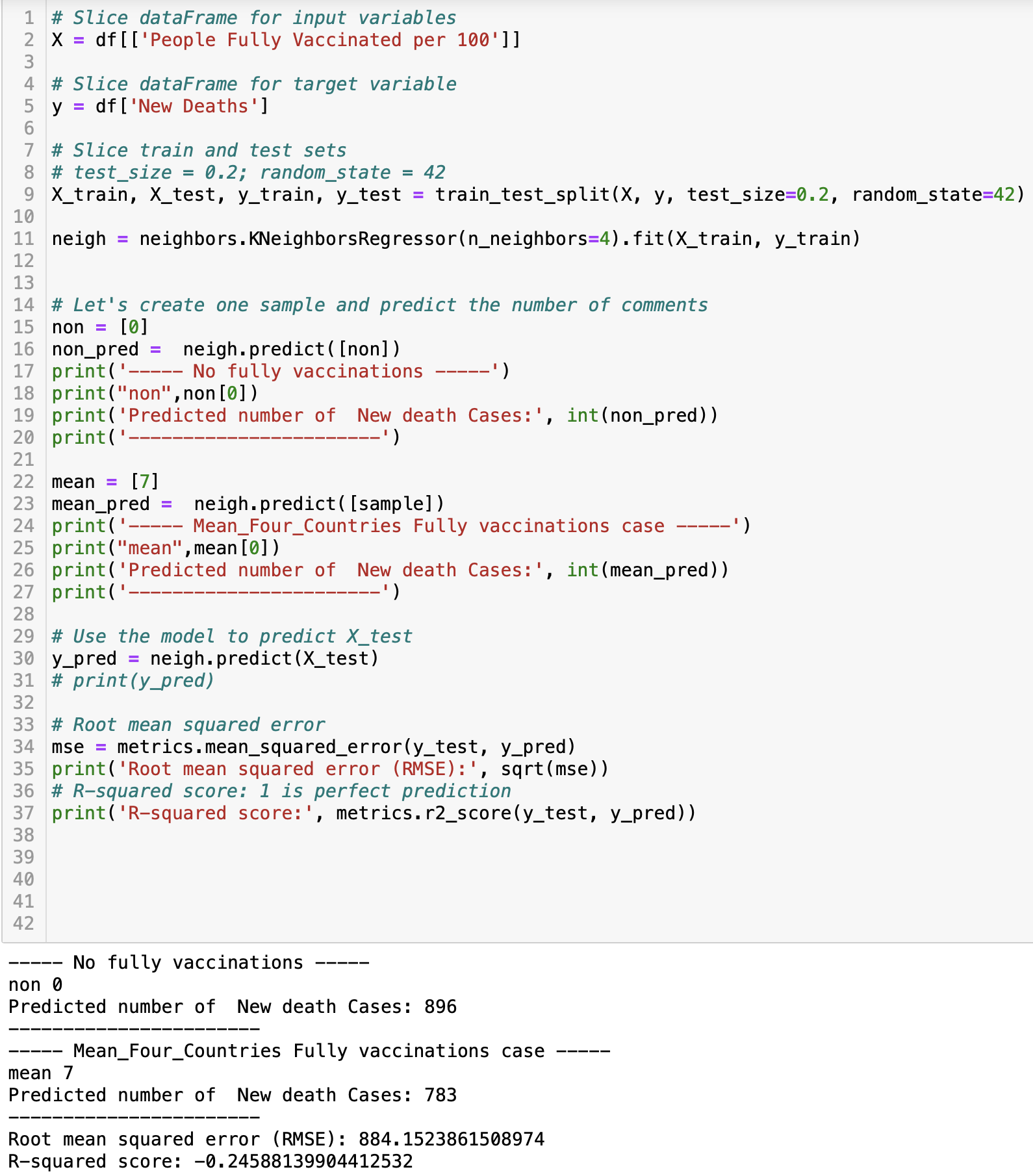


1. Predictions:



3) Using K nearest neighbors Regression to predict:

1. Overall:



1. Separated by Countries

|  | Mean | Predict |
| --- | --- | --- |
| Australia |  |  |
| India |  |  |
| UK |  |  |
| US |  |  |

III. Evaluations and Conclusions:

|  | Evaluation | Conclusion |
| --- | --- | --- |
| Linear Regression | *`Root mean squared error: 796.0082566340341*  *`R-squared score: -0.009851698354347516* | See overall |
| XGBoost Regressor | *`Root mean squared error: 736.7536918503891*  *`R-squared score: 0.13489843688176772*  *`Accuracy: 0.21%* | See overall |
| K nearest neighbors Regression | *`Root mean squared error: 884.1523861508974*  *`R-squared score: -0.24588139904412532* | According to overall, and separated conditions by countries,  (India, Australia, UK, US with  mean fully vaccinated per 100: 1.25; 2.23; 11.48; 11.61;  and corresponding prediction on new deaths:  1995; 960; 558; 371)  It may suggest a negative correlation between ‘People Fully Vaccinated per 100’ and ‘New Deaths’, i.e. more fully vaccinated might link to less new death cases. |
| Overall | Based on the RMSE, R-squared score and Accuracy, we could observe that all of the models above are not so accurate.With K nearest neighbors Regression performs relatively better than the other two. | It may suggest a negative correlation between ‘People Fully Vaccinated per 100’ and ‘New Deaths’.Using the prediction cases ‘0 fully vaccinated’ and ‘mean(four countries) = 7 fully vaccinated’, we could observe that, based on the models above, more fully vaccinated might link to less new death cases. |

**Part B**

In the last part of making predictions of the relationships between ‘People Fully Vaccinated per 100’ (input) and ‘New Deaths’(Output), several regression approaches have been tried. First, based on the observation of the data types, linear regression prediction was introduced. Though the evaluation quality is not quite optimistic, with root mean squared error 796.01 and R-squared score -0.01, it could suggest there is a negative correlation between ‘People Fully Vaccinated per 100’ and ‘New Deaths’. Then, to try to optimize the prediction model, the XGBoost regression was introduced, giving consideration that it is an efficient implementation of gradient boosting with characteristics like computation speed, parallelization, performance, and can be used for regression predictive modeling. And the accuracy (Root mean squared error: 736.75, R-squared score: 0.13,Accuracy: 0.21%) is relatively higher than the other two . The last approach has been tried, the K nearest neighbors Regression. It has a precision of ‘Root mean squared error: 884.15, R-squared score: -0.25. Which also suggests a negative correlation between ‘People Fully Vaccinated per 100’ and ‘New Deaths’, i.e. more fully vaccinated might link to less new death cases. More insight observations were made by separating the dataset by countries, according to the output of predictions (India, Australia, UK, US with mean fully vaccinated per 100: 1.25; 2.23; 11.48; 11.61;

and corresponding prediction on new deaths:1995; 960; 558; 371) , it still supports the suggestion it demonstrated.