**EMPLOYEE PERFORMANCE ANALYSIS**

1. **Project Summary**

Scope of this project is INX Future Inc. Employee Performance analysis and Factors effecting the same. It mainly focuses on employee performance across various Departments and the major factors influencing their performance.

In this analysis feature selection plays an important role which is based on the correlation of predictor variables with the target variable(Performance Rating).

Among the several multi class classification models available in machine learning, Random Forest, XGboost and SVC were evaluated for suitability. Out of these Random Forest is found to be better in terms of efficiency and count of error values in predicted V/s actual performance rating being lower compared to other models.

To get better visualisation of results ‘Tableau’ is used along with Matplotlib in Python.

* 1. **Requirements**

The Employee Performance data from INX Future Inc. is provided along with the objectives specified by the CEO of the company.

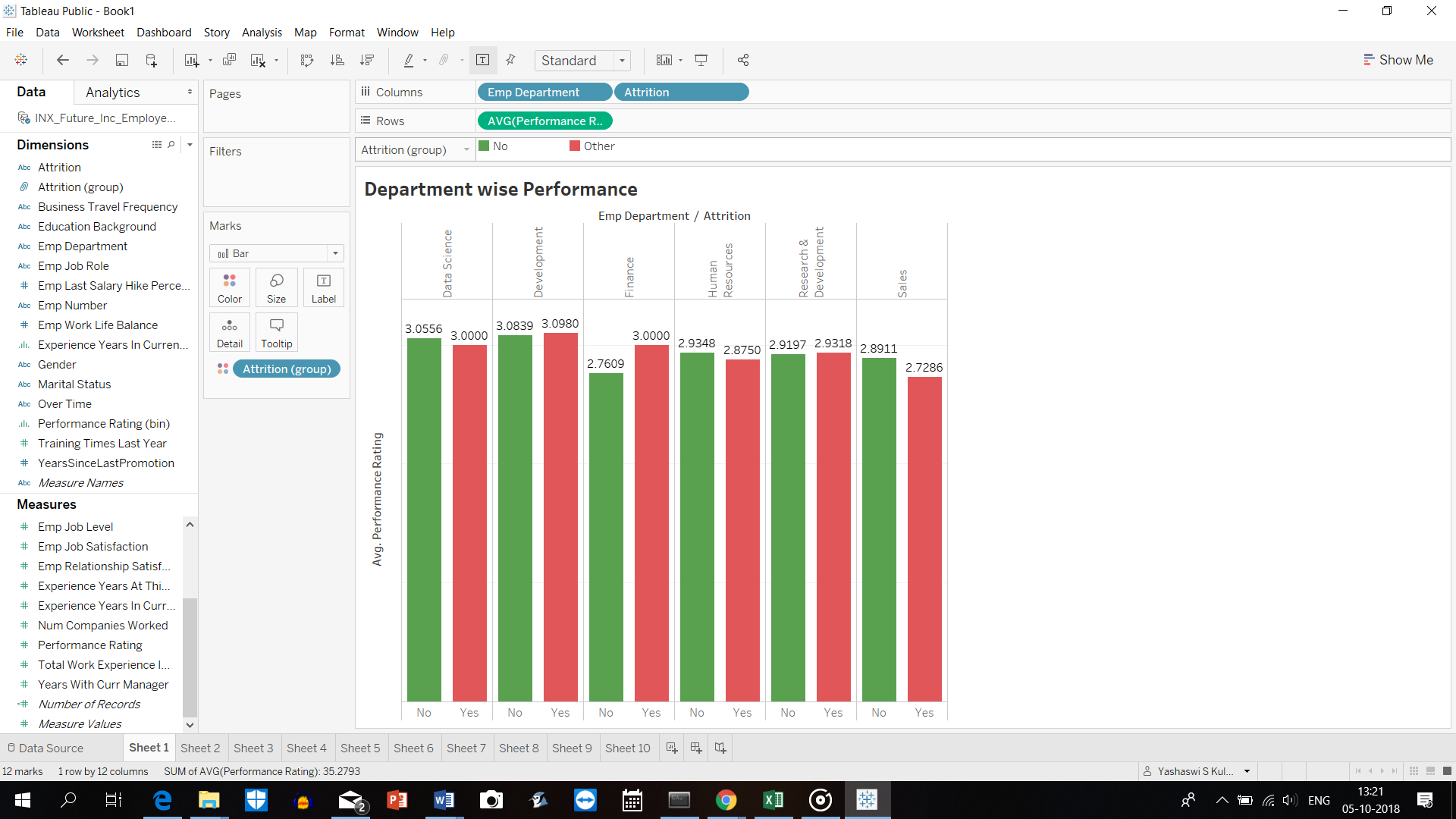
The following insights are expected from this project:

1. Department wise performances
2. Top 3 Important Factors effecting employee performance
3. A trained model which can predict the employee performance based on factors as inputs, which would while hiring new employees in future.
4. Recommendations to improve the employee performance based on insights from analysis.
   1. **Analysis**

Analysis of the data is done using Tableau for better results visualization and improved understandability.

1. **Department wise Performances:**

From the analysis of the data, we can visualize the department wise performance in two ways – ‘Tableau’ and ‘Matplotlib’ visualization in Python.



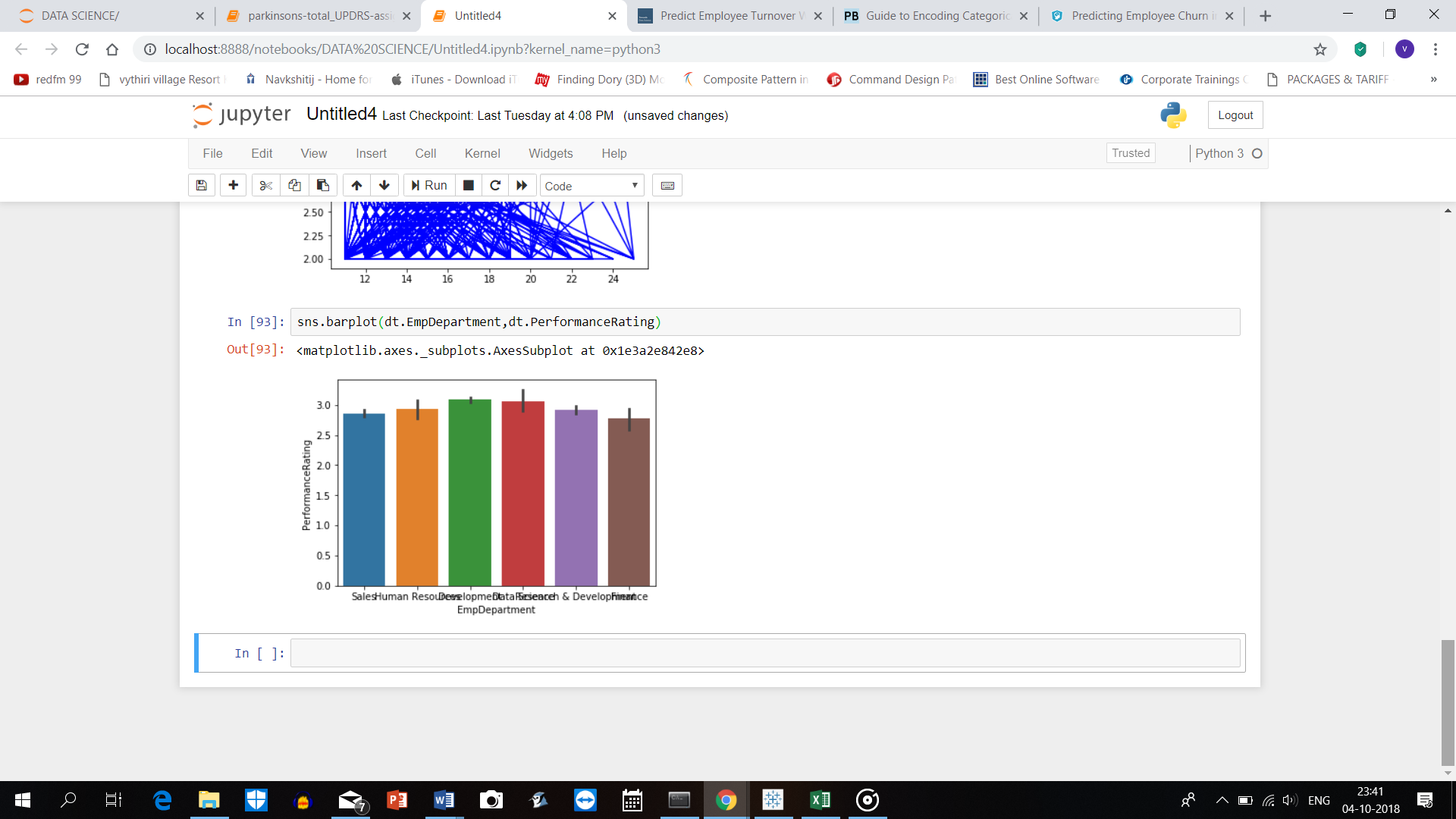
Graph of Employee Department v/s Performance Rating(Tableau)

Performance rating with respect to attrition:

* In Data Science, Development, Research & Development,Human Resources Departments the difference between ratings w.r.t Attrition is negligible.
* In Sales department, performance rating of employees leaving the organization is relatively lower
* Finance department is losing high performing employees.

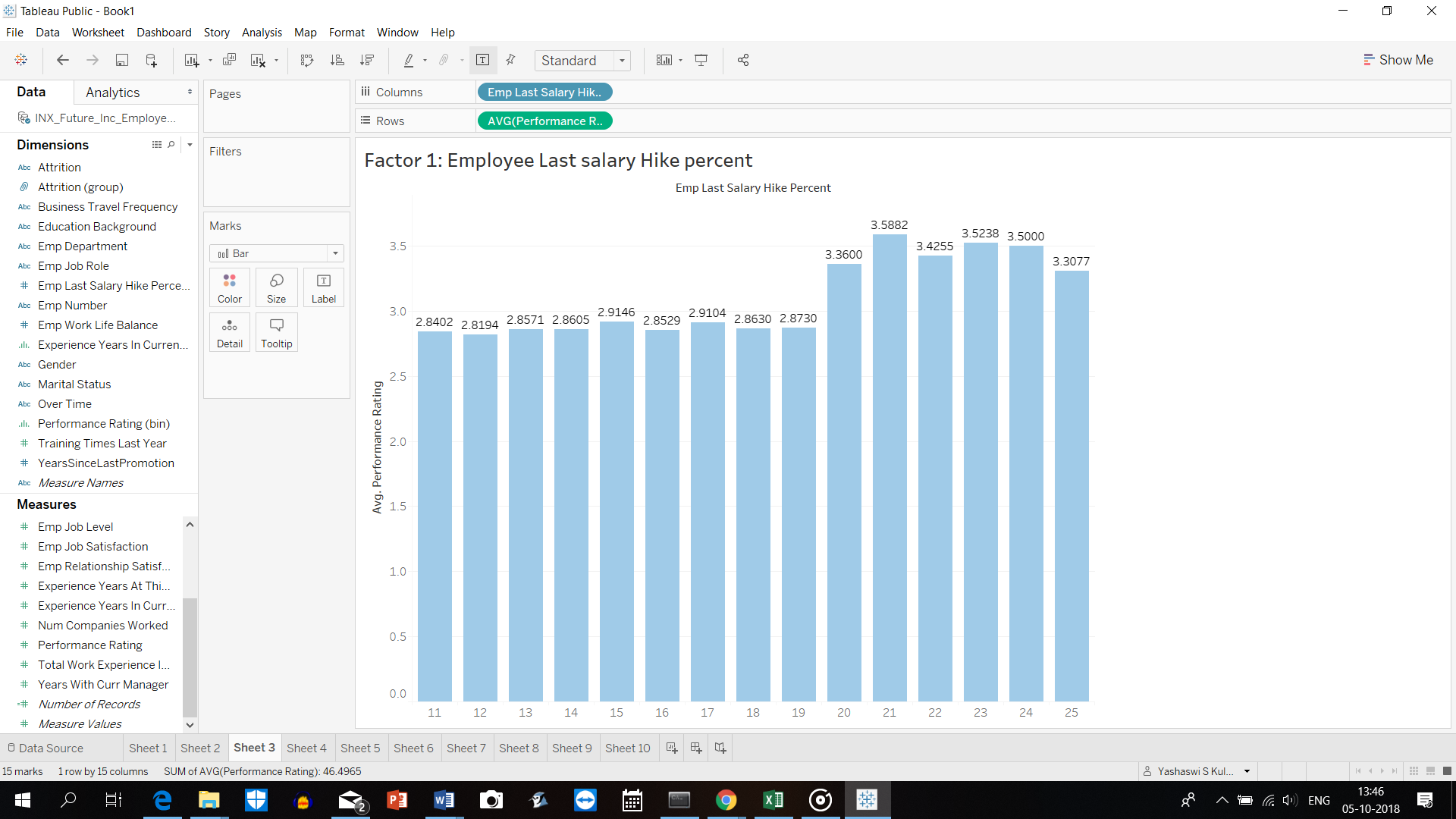
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Employee Department | Count of Employees Based on Performance Rating | | | |  |
|  | **Good**  **(2)** | **Excellent**  **(3)** | **Outstanding**  **(4)** | **Grand Total** | **% of 3 Ratings** |
| Data Science | 1 | 17 | 2 | 20 | 85 |
| Development | 13 | 304 | 44 | 361 | 84.21052632 |
| Finance | 15 | 30 | 4 | 49 | 61.2244898 |
| Human Resources | 10 | 38 | 6 | 54 | 70.37037037 |
| Research & Development | 68 | 234 | 41 | 343 | 68.22157434 |
| Sales | 87 | 251 | 35 | 373 | 67.2922252 |
| Grand Total | 194 | 874 | 132 | 1200 |  |
|  |  |  |  |  |  |

All in all, Data Science and Development Departments have 85% of their employees in the Excellent(3.00) Rating category. In other Departments Excellent performing employees are in the range of 61-75% .

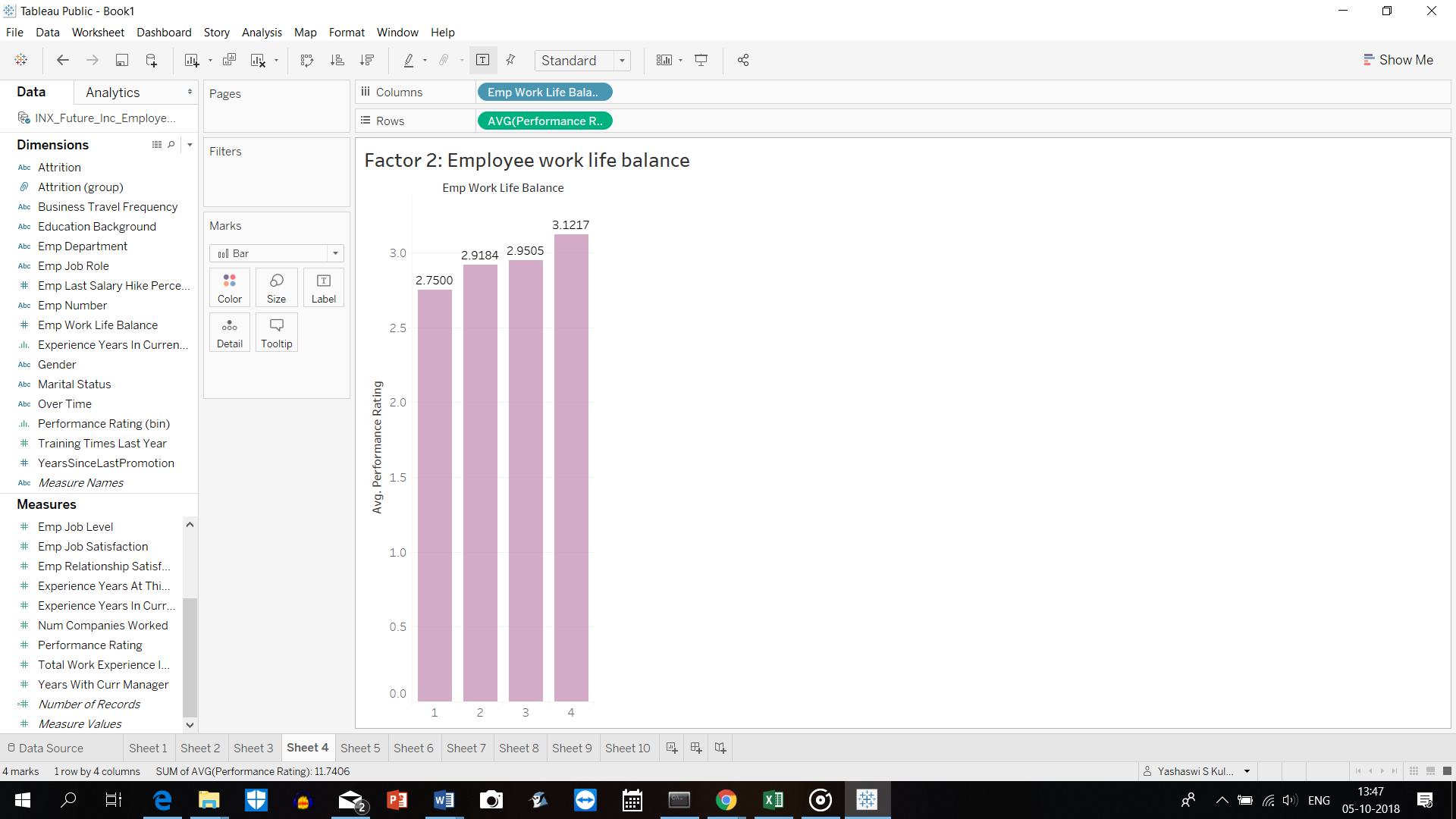


Graph of Employee Department v/s Performance Rating(Matplotlib)

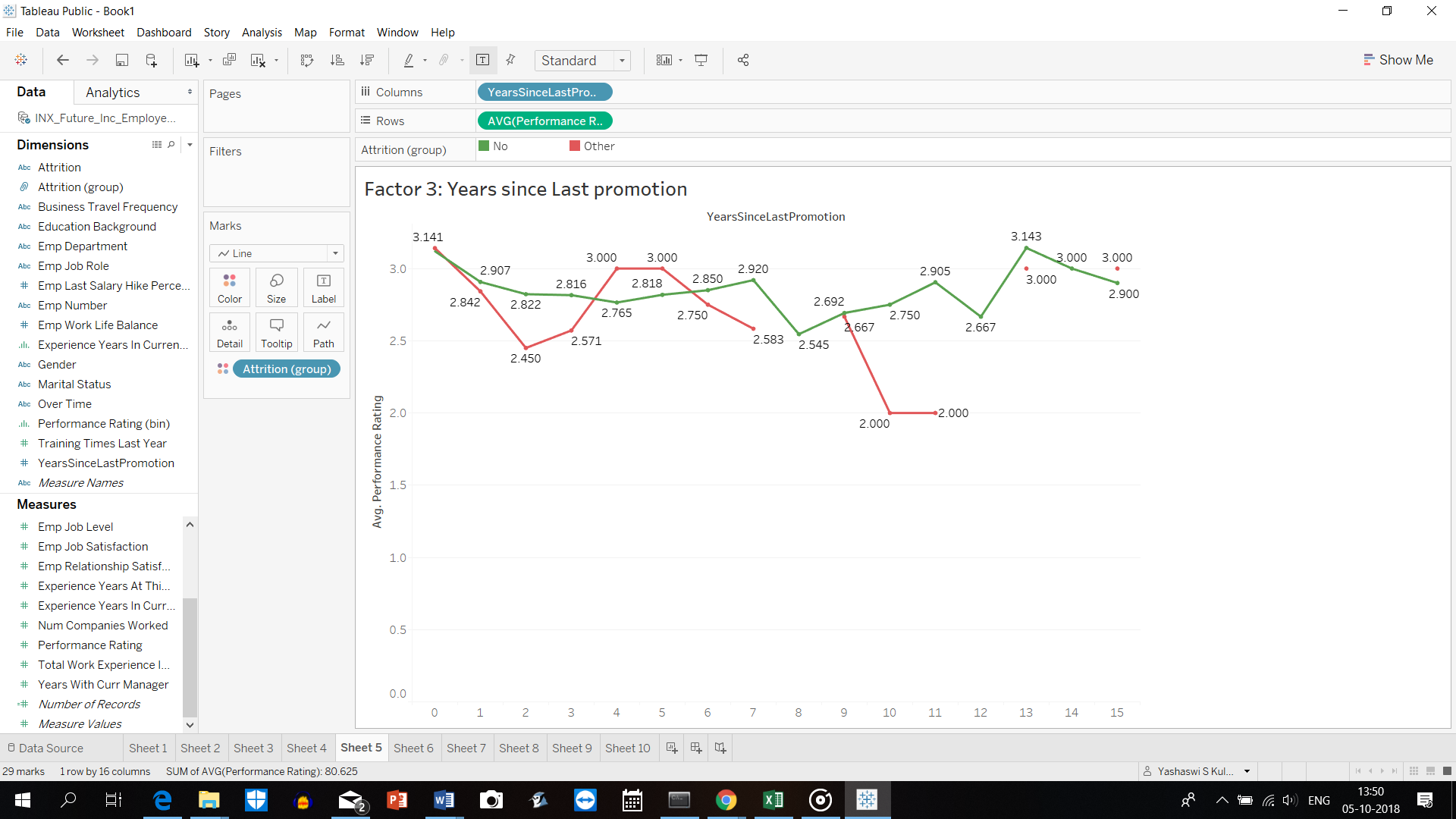
1. **Top 3 Factors effecting Employee Performance:**



Above graph indicates that Performance Rating is highly proportional to the Employee Last salary Hike percent provided.



As the graph indicates, the Performance Rating is highly proportional to the Employee Work Life Balance.



This shows that, Higher the gap since Last promotion , lower is the performance rating leading to Attrition of employees.

* 1. **Summary**

Considering the objective, post evaluation of different models, Random Forest is found to be suited better with an efficiency of 93.88% and count of error values predicted Vs actual performance rating being lower. Hence it is used for the analysis.

**Recommendations:**

1. In Finance department people with higher performance rating(3.000) are leaving the company which is leading to knowledge erosion and creating gap in subject matter expertise. So, to improve the employee performance, additional training and knowledge transition sessions from high performers to low performers must be ensured.
2. The Development department although having (3.08)rating for NO attrition, people having similar rating(3.09) are on the higher side of attrition, which is not a good sign. So, it is recommended to improve the work life balance which further contributes to improved performance of existing employees and retention of high performing employees.
3. Sales department has a good opportunity to hire employees, as there exists a high churn of low performing employees.

It is beneficial to hire employees having relevant experience and more creative, innovative ideas.(Increased Job involvement)

1. For improving the performance of low performers(2.00) in all other departments, a minimum of 3 trainings must be provided to all the employees.
2. **Data**

The data for modelling is provided from external sources such as excels sheets.

* 1. **External data**

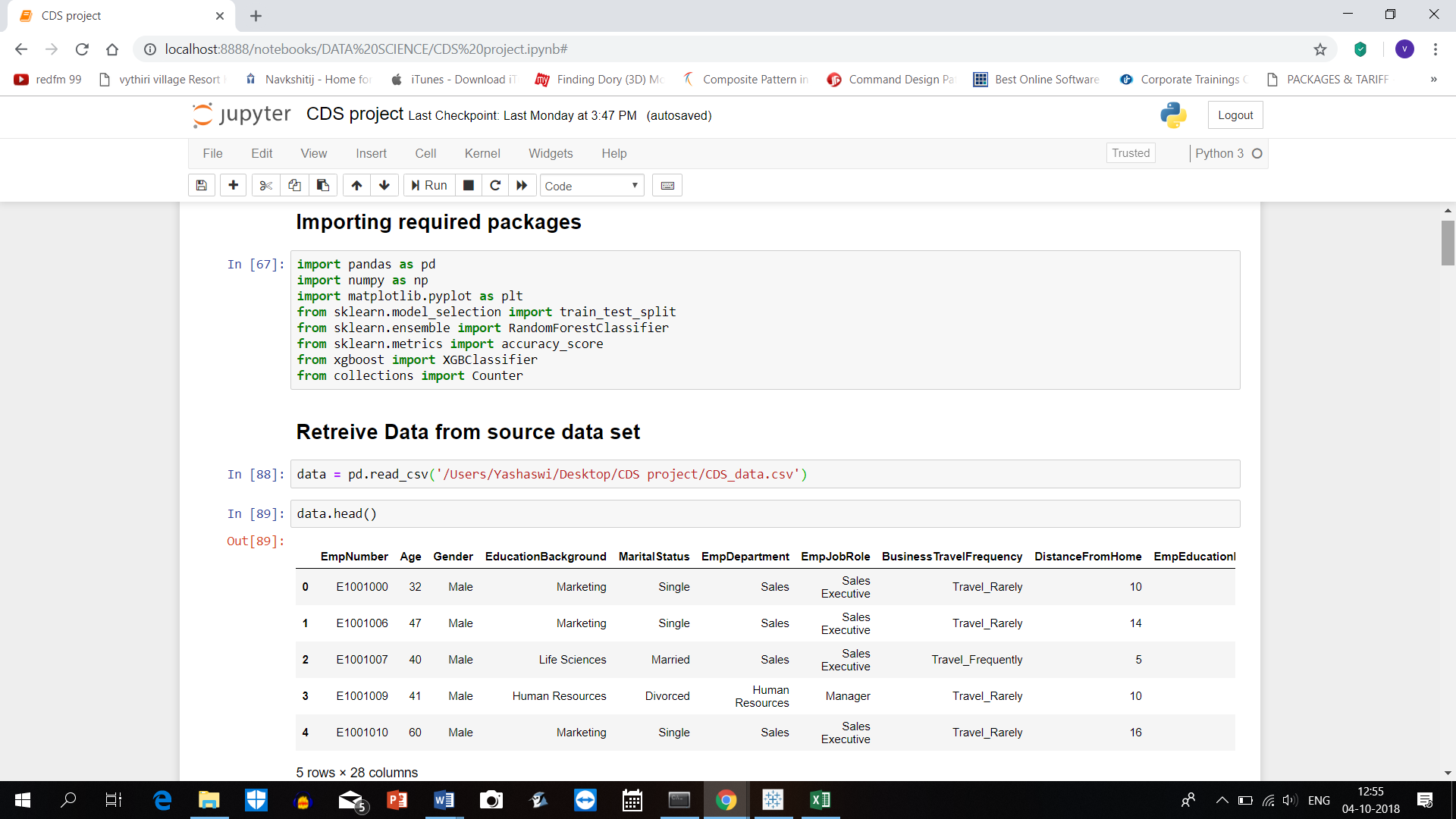
The data is loaded from external third-party sources which requires processing. Here the third-party source is INX Future Inc. and data in excel sheet format. The given excel file has the tabular data in one sheet and the Data definitions in the other sheet.

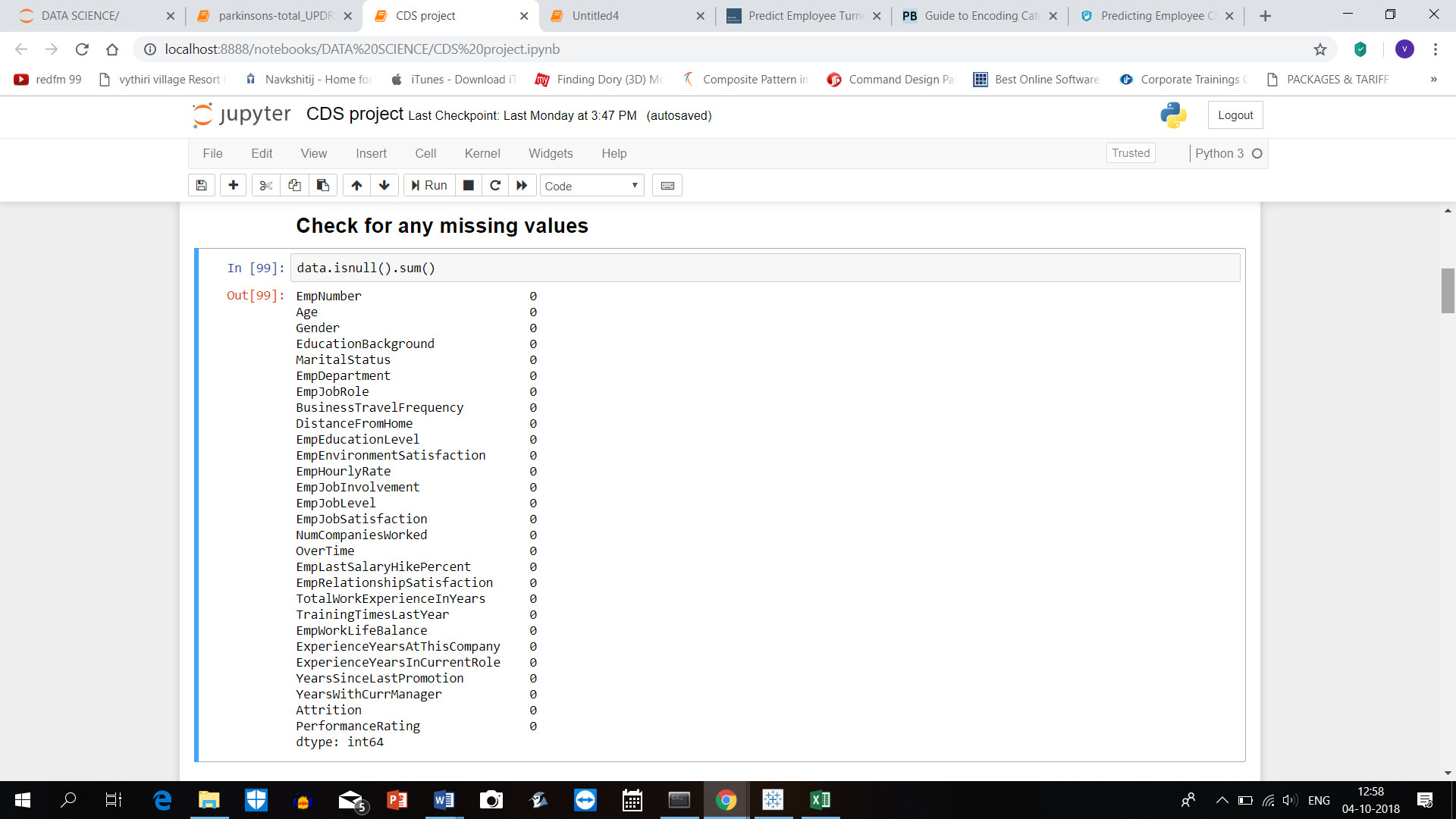
* 1. **Processed data**

Before modelling the data, it requires pre-processing. The filters applied to tabular data were removed and the categorical data was studied. For model building and prediction, only the data sheet was extracted from the excel sheets.

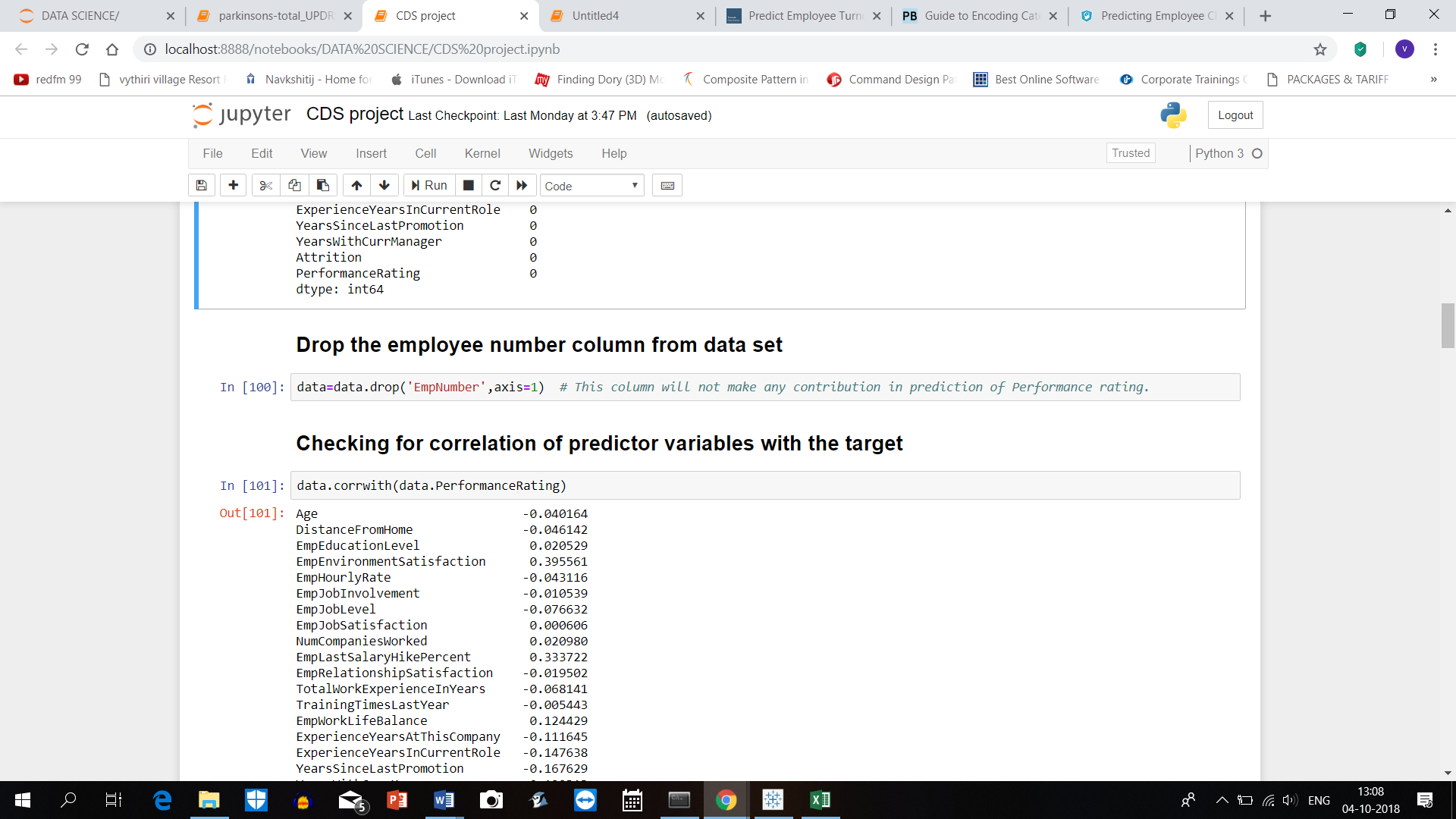
1. **Data Processing**

In this section ,the data processing, data mugging, exploratory analysis is done using Python in Jupyter notebook. The processing includes converting data from excel to csv file, loading the data set to the jupyter file using the **pd.read\_csv** command in python as shown below.

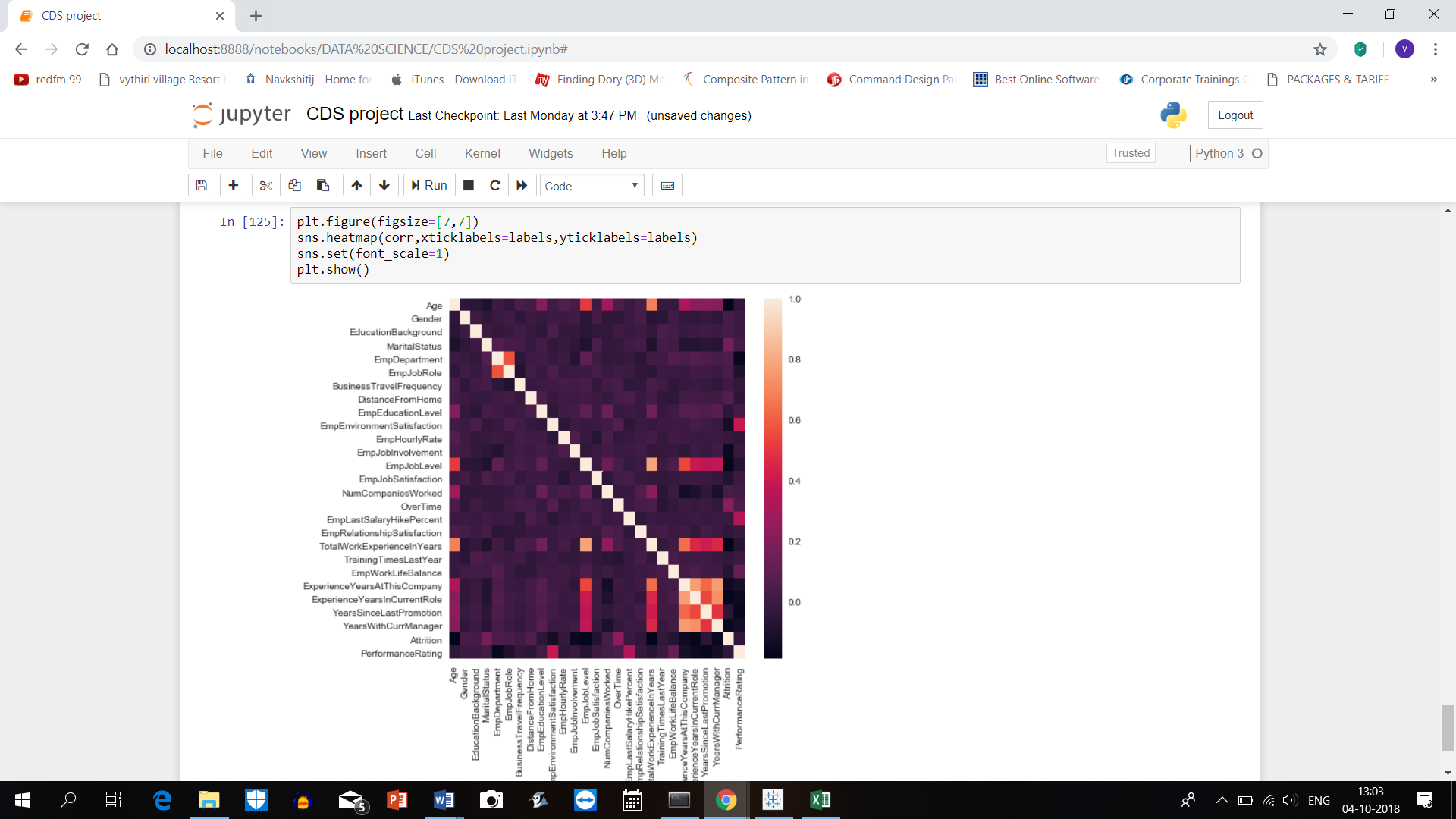




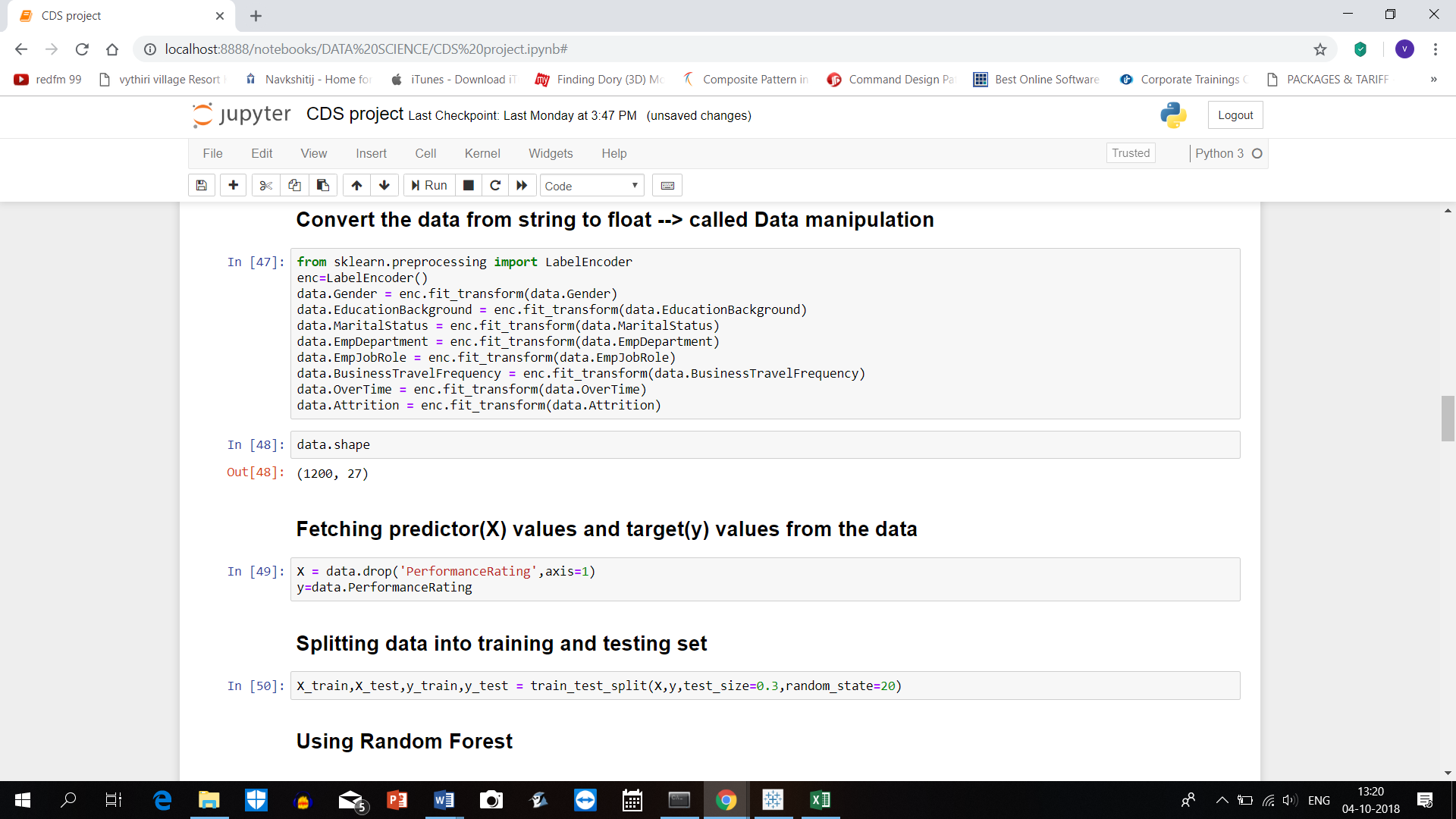
During exploratory analysis, the Employee number column is dropped from data, as there is no contribution for model and prediction.



The further feature exploration is done using a heatmap of the correlation between the predictors/ features and the target i.e, Performance Rating.



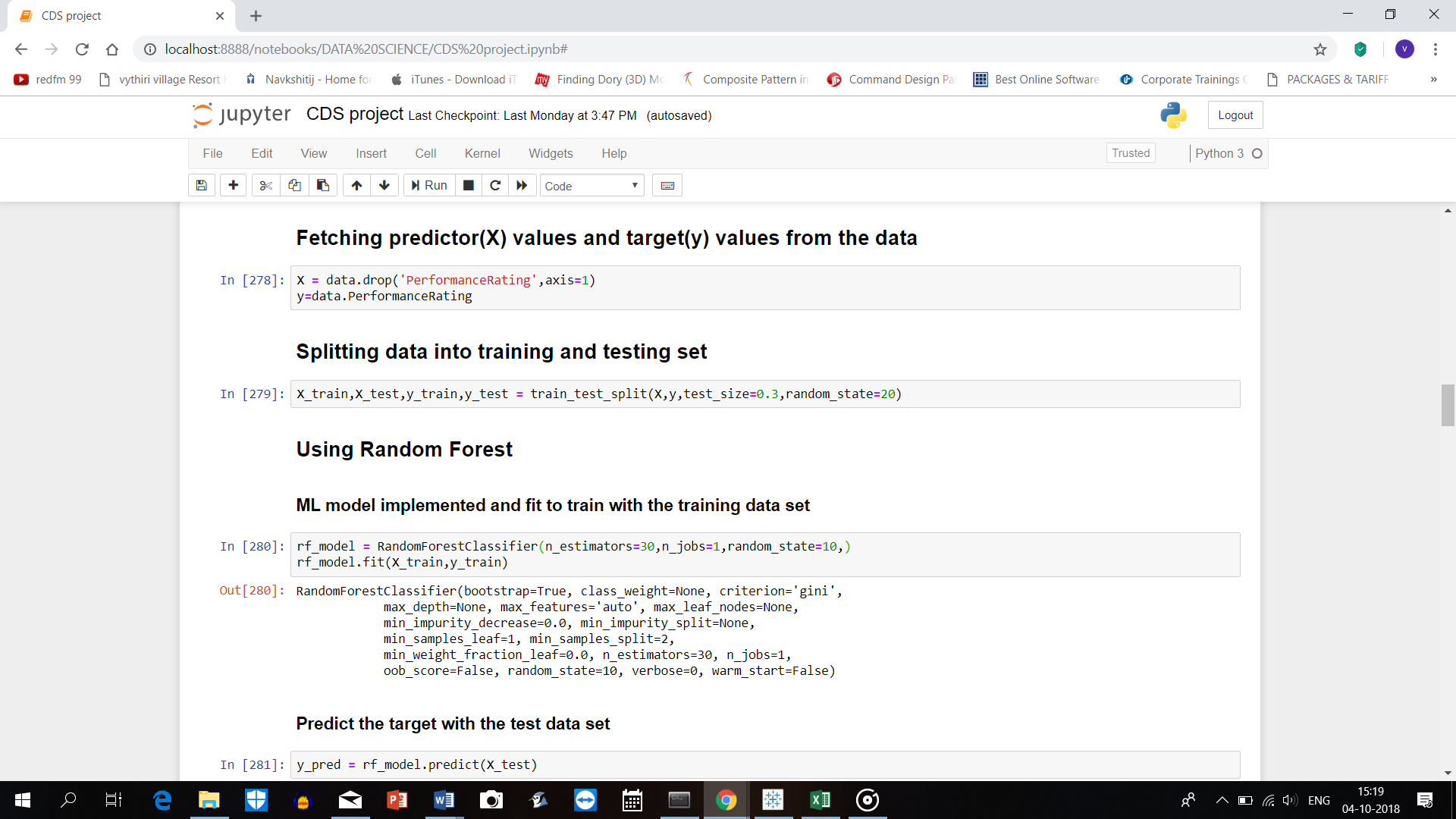
By selecting the features having higher correlation to target and converting the categorical data to numerical data using Data manipulation technique such as Label Encoding.



1. **Model**

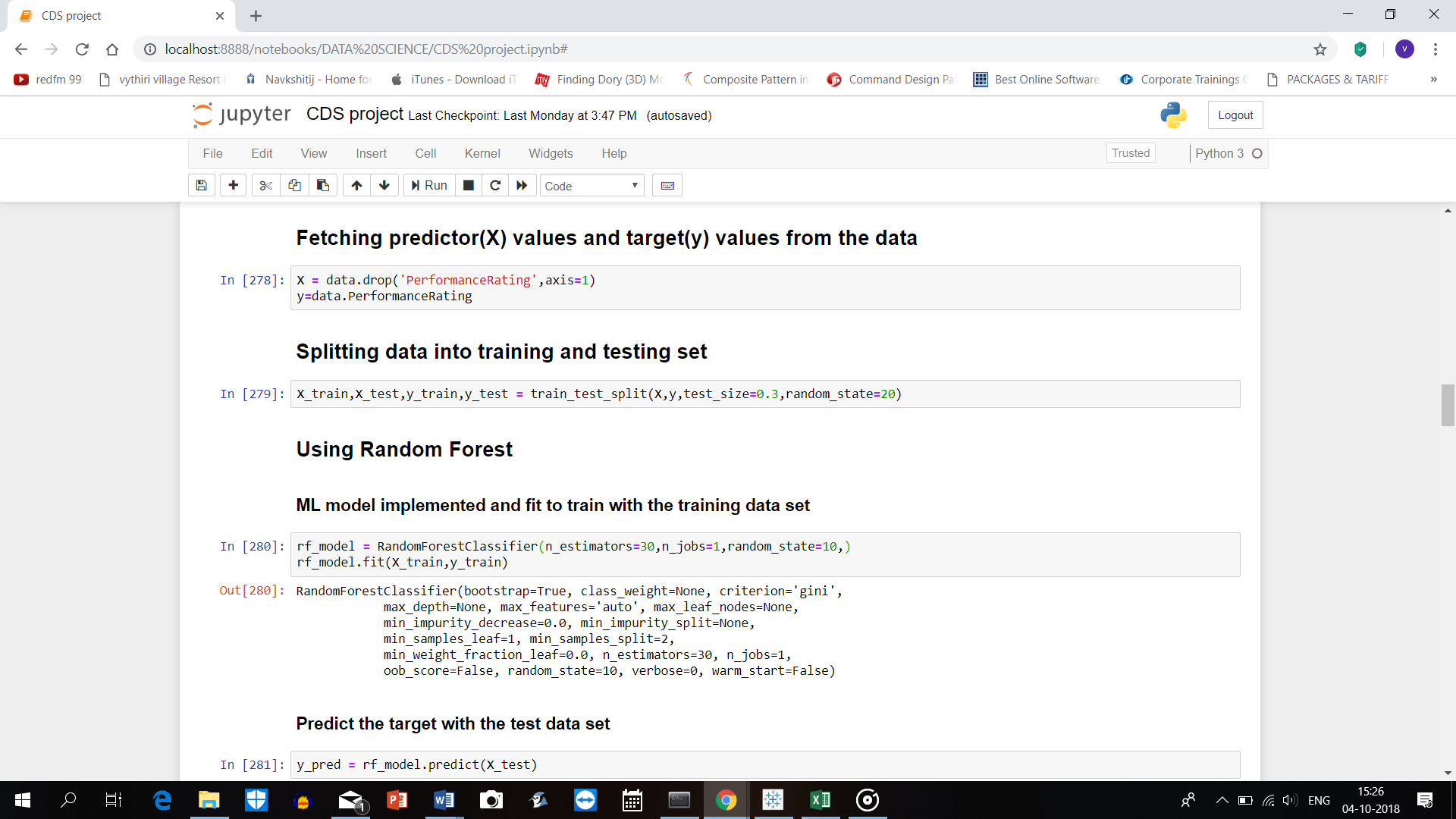
From the available machine learning models for multi class classification, Random Forest is more efficient in various aspects when compared to other models such as support Vector machine.

For implementing the Random Forest model, firstly the predictors(X) and target(y) is defined. Next the data set is divided into 2 parts: Training set , Testing set using the code as shown below.



* 1. **Training model**

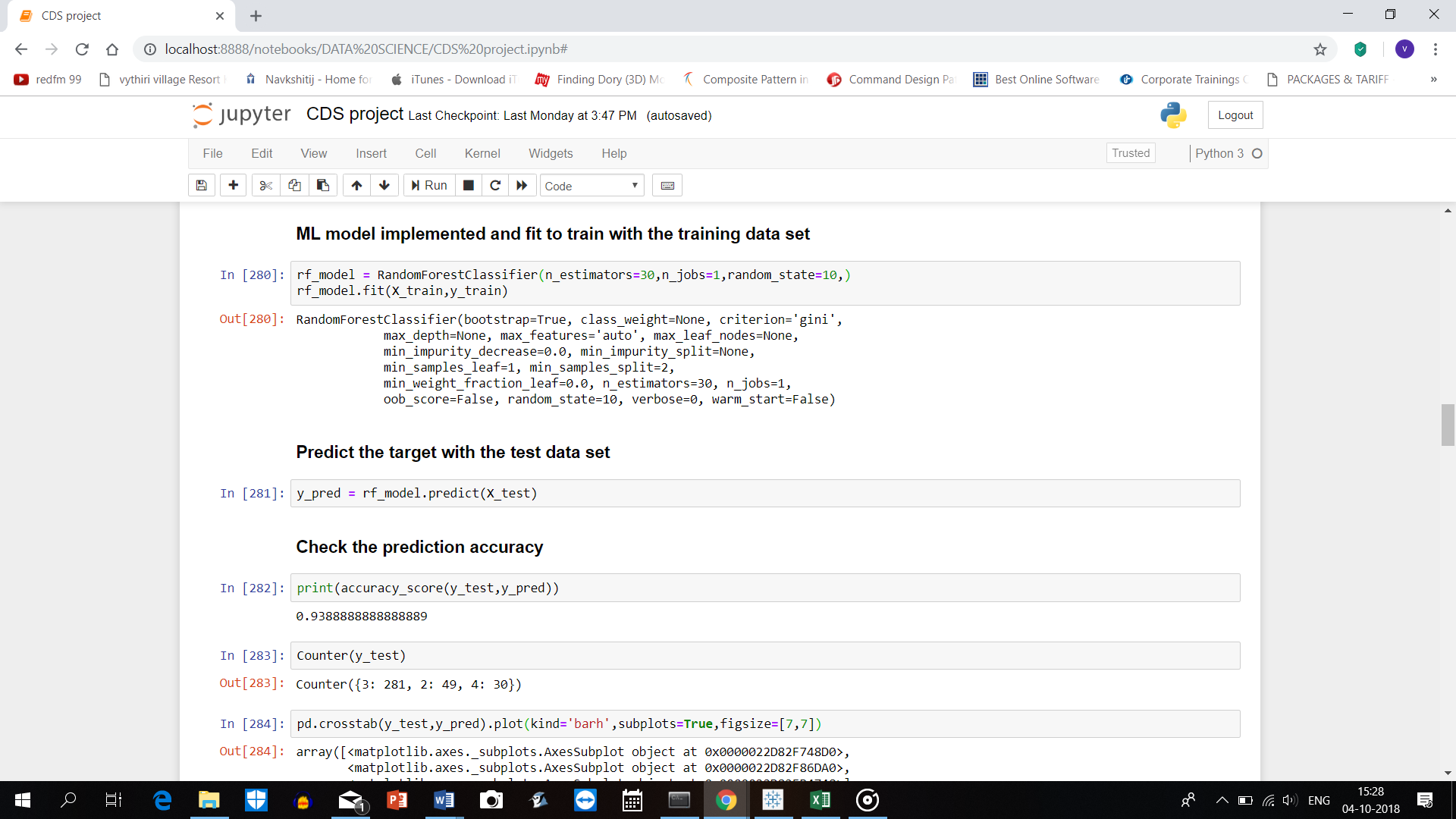
The Random Forest Classifier model is called and is trained with the Training data set by using the **rf\_model.fit(X\_train,y\_train)** command.



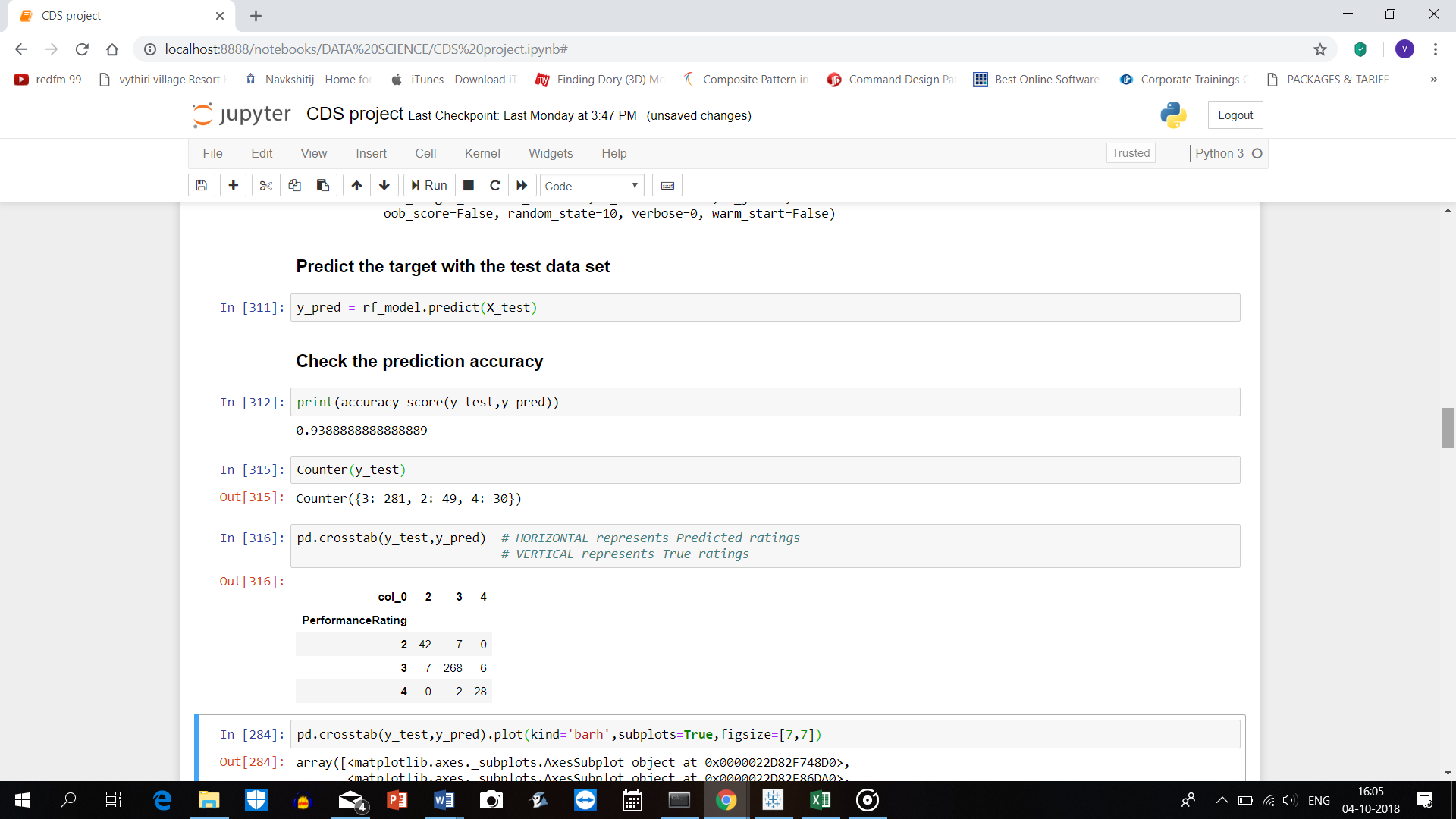
* 1. **Predict model**

The trained model is fed with the Testing data set using **rf\_model.predict(X\_test)** command.

To check the prediction accuracy rate, we use the accuracy\_score function which shows a 93.88% efficiency for Random Forest model.



To compare the predicted rating with the True ratings, crosstab is used as shown below.



1. **Visualization of results**

The prediction results in comparison with True values can be visualised as shown below.

