

NAME OF THE PROJECT

**Ratings Prediction Project**

Submitted by:

Ram Kumar

**ACKNOWLEDGMENT**

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals. We would like to extend my sincere thanks to SME. Khushboo Garg .

We are highly indebted to Flip Robo technology for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

I thanks and appreciations also go to our colleague in developing the project and people who have willingly helped us out with their abilities.

Thanks all.

Ram kumar

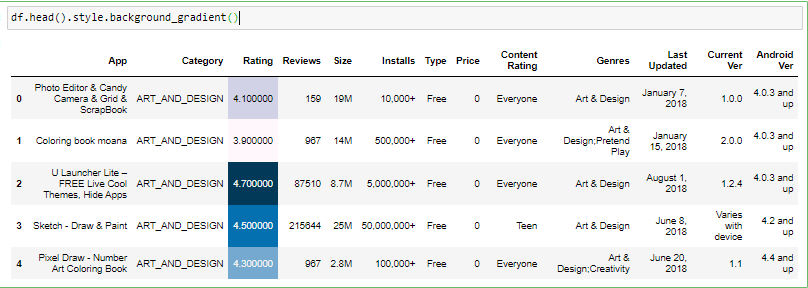
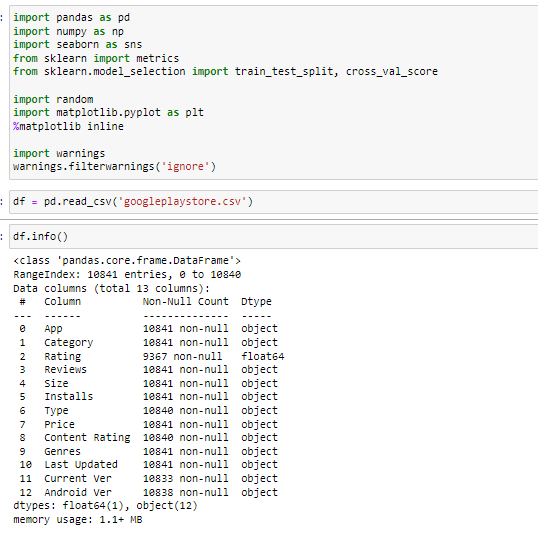
.

**INTRODUCTION**

* We have a client who has a website where people write different reviews for technical products. Now they are adding a new feature to their website i.e. The reviewer will have to add stars(rating) as well with the review
* The rating is out 5 stars and it only has 5 options available 1 star, 2 stars, 3 stars, 4 stars, 5 stars. Now they want to predict ratings for the reviews which were written in the past and they don’t have a rating. So, we have to build an application which can predict the rating by seeing the review.

**Analytical Problem Framing**

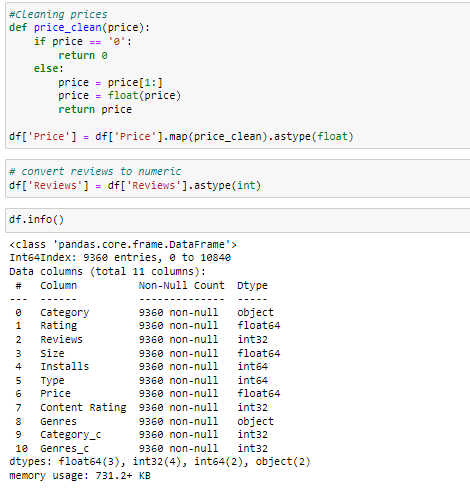
Import library and load the dataset



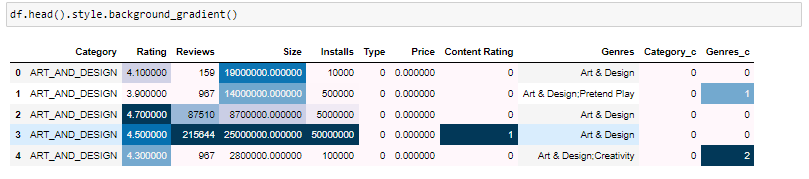


* Display all column name of dataset.

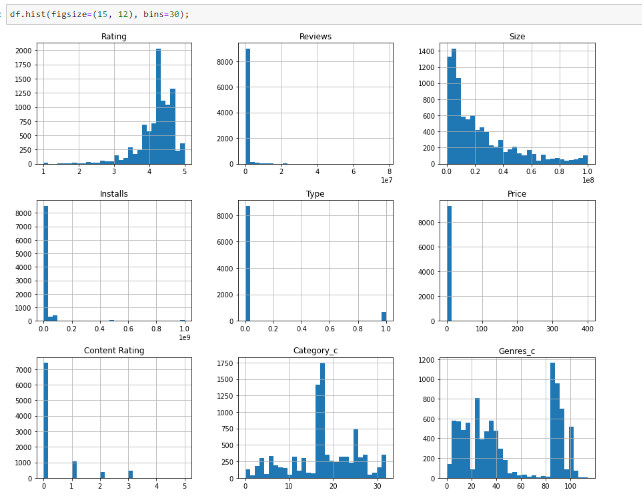




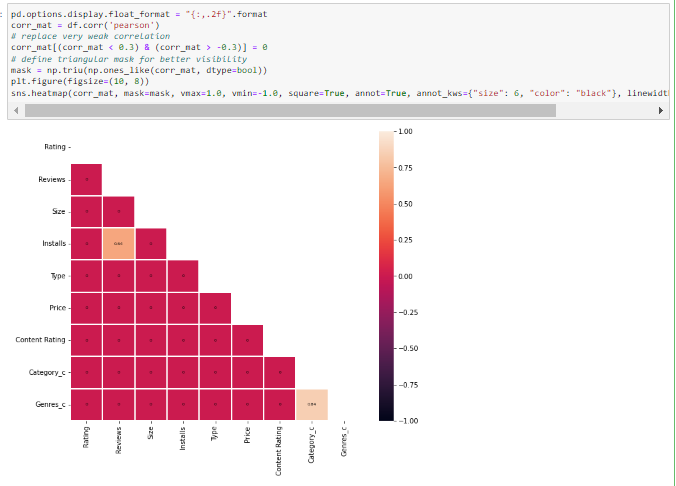
* Display statistical summary.



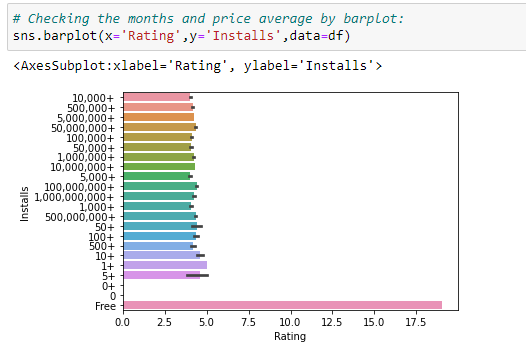
* Display histplot of all columns.

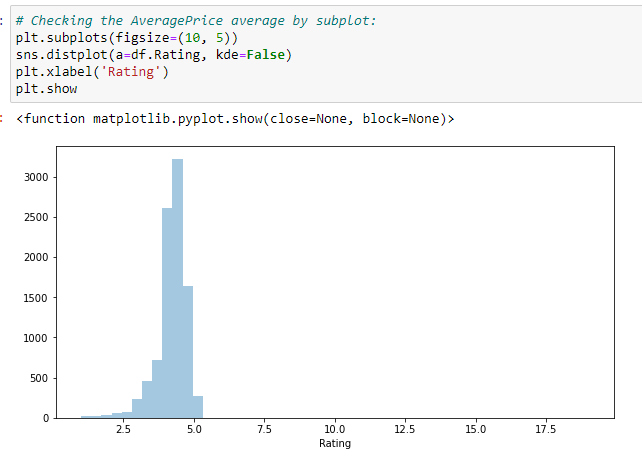


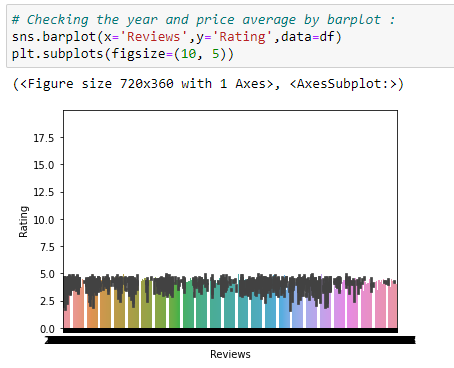
* Display correlation of columns using heatmap.



* Display barplot of all columns.

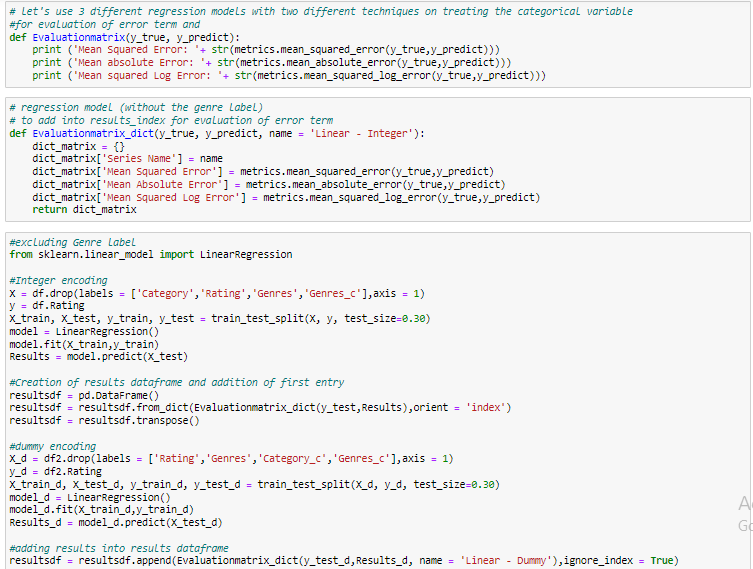


* Display barplot of all columns 
* Display barplot of all columns :

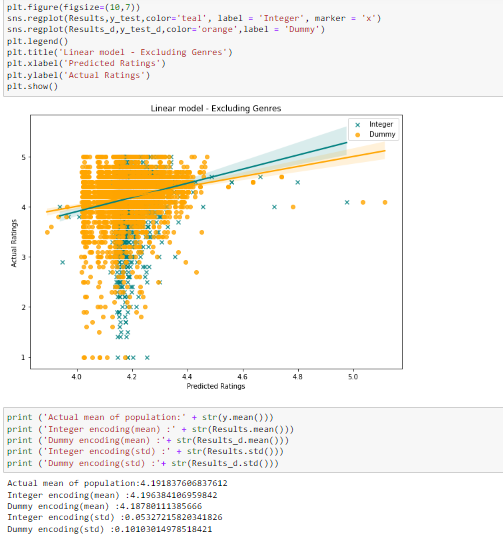


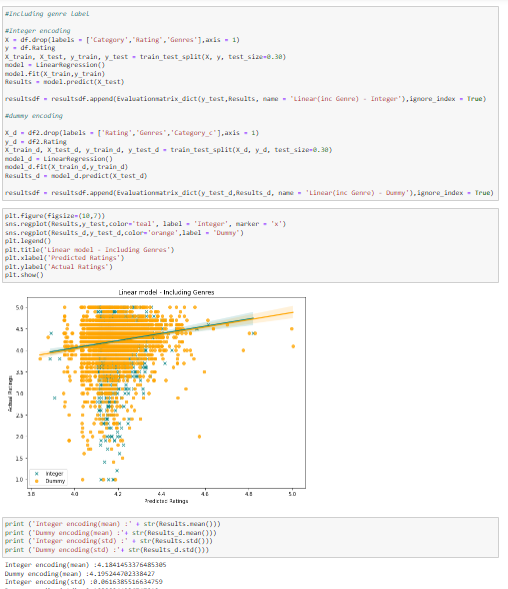
**Model/s Development and Evaluation**

* **Feature engeenering:**

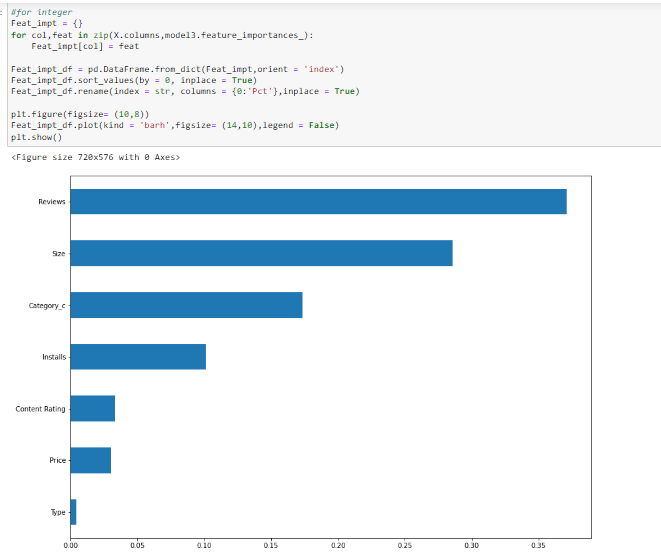


* **Testing of Identified Approaches (Algorithms):**

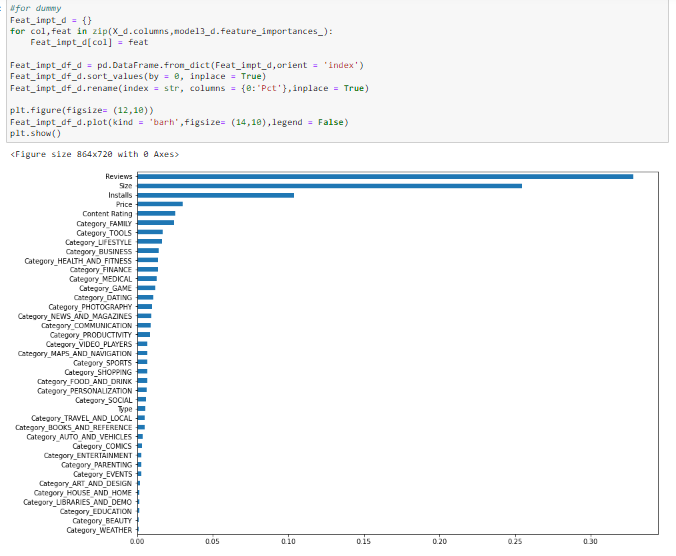




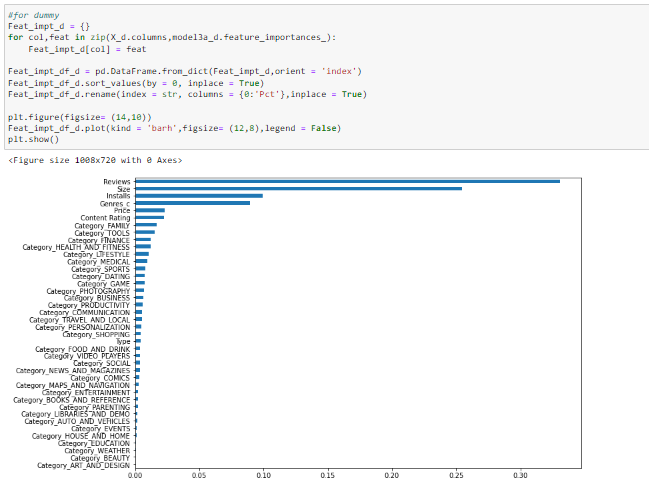
* **Run and Evaluate selected models**



* **higher overall predicted mean** :



* the ratings, the top 4 being reviews, size, category, and number of installs:





* Hardware and Software Requirements and Tools Used
* **Language :-** Python
* **Tool:-** Jupyter
* **OS:-** Windows 10
* **RAM:-** 8gb

**CONCLUSION**

 It is not easy to conclude which model has the best predictive accuracy and lowest error term. Using this round of data as a basis, the dummy encoded SVM model including genres has the lowest overall error rates, followed by the integer encoded RFR model including genes. Yet, all models seem to be very close in terms of it's error term, so this result is likely to change.

What is very surprising to me is how the RFR dummy model has such a significantly more error term compared to all the other models, even though on the surface it seemed to perform very similarly to the RFR integer model.