**Problem Description:**

Here in this dataset the about the sales of the video Game in a year from 2019 to 2020. It contains 9833 records and 18 feature. Here we have to predict the futures sale based of the features of the dataset and need to answer the following questions below:

1. Is it possible to accurately forecast the sales of its video game for each of its customers (local video games stores)?
2. Given a certain marketing budget, how can we best distribute marketing activities to maximize sales?

**Dataset Explorations**

Here in this dataset it have the following columns with datatype and also I have shown the description of the dataset.

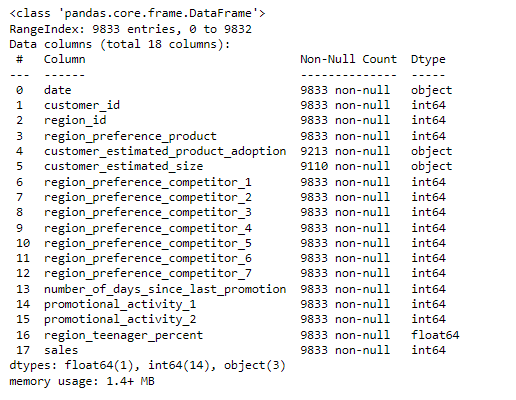


Figure : Each column information

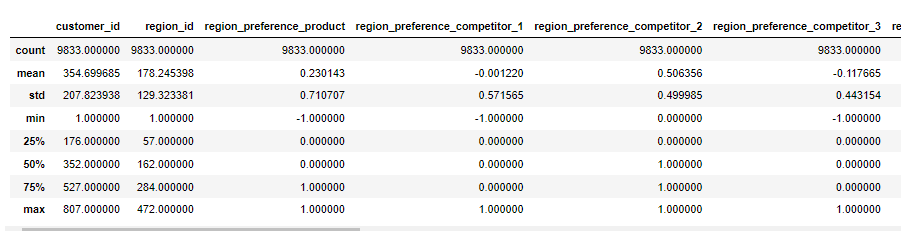


Figure : Each Numeric column different calculations

Here is the frequency of the sales column in the dataset:

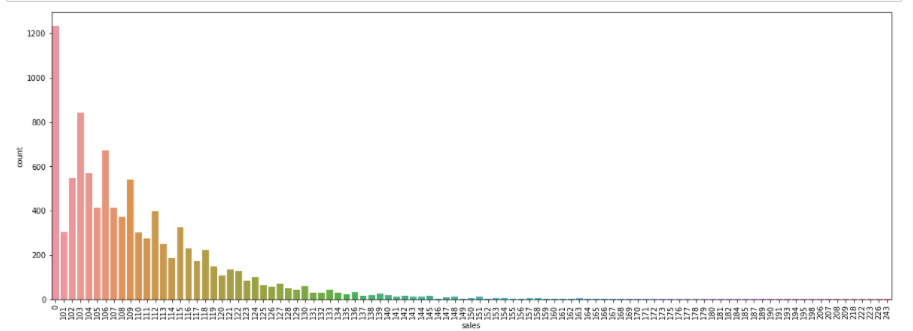


Figure : Sales column Frequencies

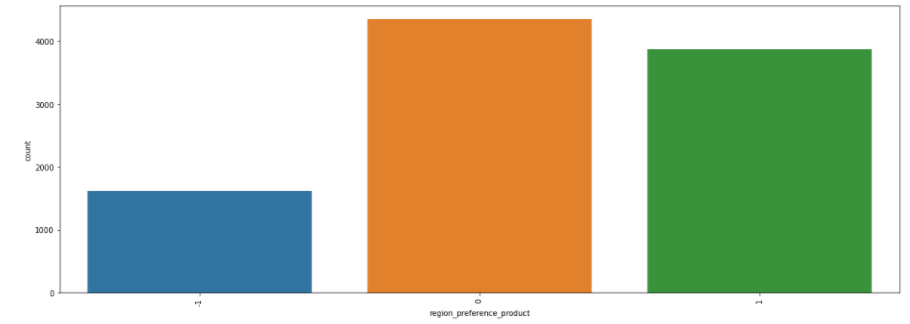


Figure : region\_preference\_product column frequencies of the sales targets achievments

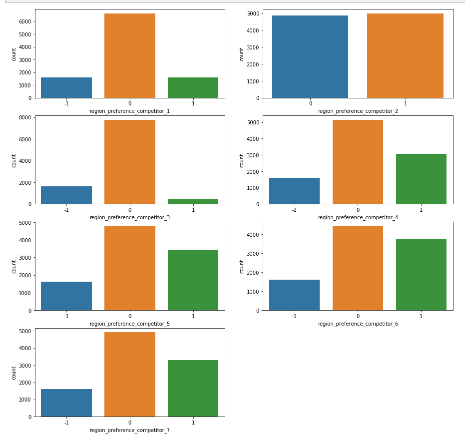


Figure : Each region competitor Target Achievements

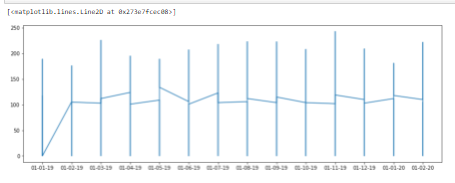


Figure : Time line of sales month wise

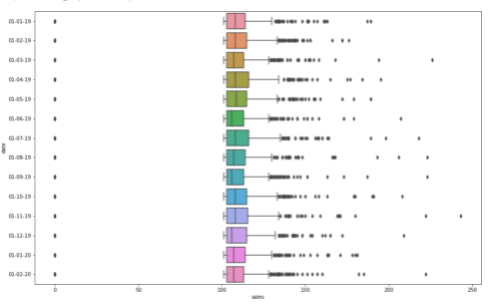


Figure : Box plot of each sale distribution date wise

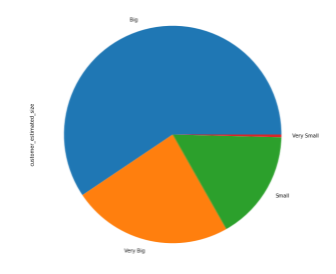


Figure : Customers size Distribution

**Preprocessing for the dataset:**

Here in these steps we have dropped the following columns due to its irrelevance from the dataset.

* 'region\_id’, (As ID could be any number)
* 'region\_preference\_product', (Preferences over region base)
* 'region\_preference\_competitor\_[x]’, (Region of the competitor preferences)
* 'number\_of\_days\_since\_last\_promotion’ (Number of last promotion days)

Encode the categorical label into numeric form in the dataset:

* 'customer\_estimated\_product\_adoption'
* 'customer\_estimated\_size'

For the date column we have splitted the date column into day, month and year.

Here is the Dataset correlation graphically:

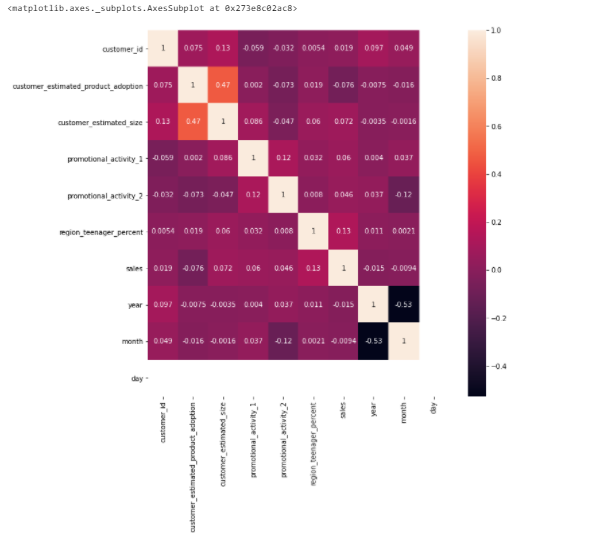


Figure : Correlation graph

From the above correlation graph, we can clearly observe that the dataset is highly correlated to the target column and the dataset ready for the model input.

After the preprocessing of the dataset. We divided the dataset into features and labels. Splitted the dataset into train and test data. So now, the dataset is prepared for the machine-learning model.

**Machine Learning Model:**

Here we have built a model that will train over the provided features and predict over that features for the new sale price value. Here is the code that follows the process of the given model:

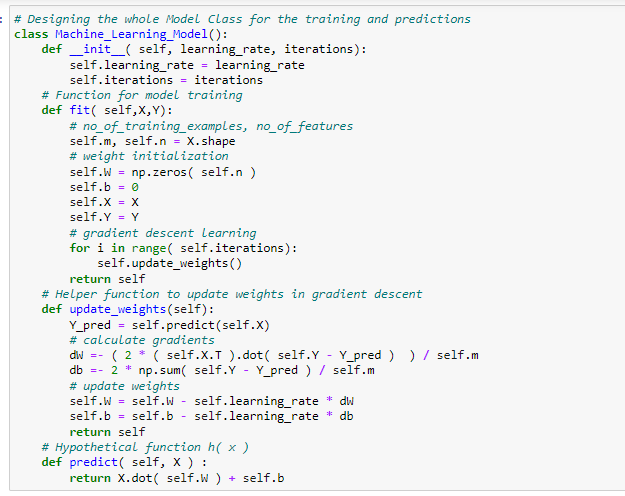
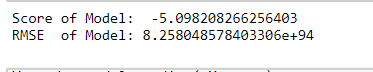


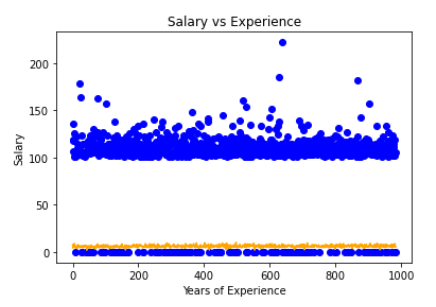
Figure : Machine Learning Model Code

**Model Evaluation:**

Here is the machine Learning model Evaluation Results for the test data.



Here is the evaluation model results graphically:



1. Is it possible to accurately forecast the sales of its video game for each of its customers (local video games stores)?

If we take the data frame of that specified region only then we can accurately forecast the sale for each customers.

1. Given a certain marketing budget, how can we best distribute marketing activities to maximize sales?

The model can be used to produce the future sale according to the dataset features and can be easily use to maximize the sales by tweaking the features values.