DEEP ANALYSIS

Problem with the dataset :

- The super store in the dataset is giving discounts for almost all the products.
- Discounts on all the products will surely lead to loss for the store.
- It is observed that people were buying products of low sale values more in more quantities.
- So we can consider these as the daily needs of people… which they will buy even though we do not give any discount.
- Hence giving discount for such products are of no use ,infact it throws the store into loss.

Solution :

- **First of all we must remove the discounts for all the low sale valued product (Daily need).**
- **For this we first split the dataset into three different quartiles say 0-25%,25%-UpperWhisker,>UpperWhisker based on the sales.**
- **Let these be stored in three different variables say df1,df2,df3. Derived from main dataset Store.**
- **Now for the products in the df1 should not be given the discounts , for that we have to calculate the actual amount of the products without discount which is the MRP. MRP for the products is calculated.**
- **The ultimate solution is that the products which are having low sale values and are not much purchased by customers are given as a combo pack with the products having high sale values that are more purchased.**
- **For this we have to combine the two data sets which are df1,df3.**
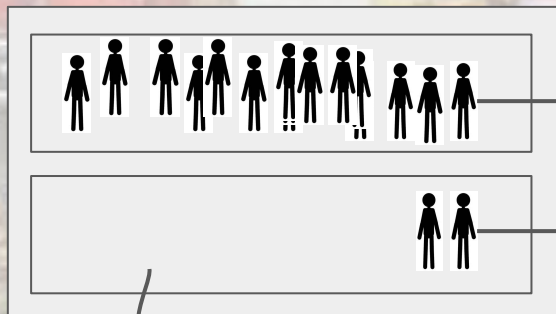
Solution (continued..)

- **Df3 has 164** rows and **df1 has 329 rows**. To combine these sets the number of rows in both the datasets should be either same or in proportion. Hence df1 is split into two sets df_extract1,df_extract2.
- Such as **df_extract1** is having a quartile of upto 15% but still it has **180 rows != 164 rows**.
- Therefore it is first sorted in ascending order and then the last 16 columns is dropped and kept in **low_df1** and these 16 columns are kept in **high_df1**.
- Then the rows of **low_df1 and df3** (both having 164 rows) are combined as combos (as explained before)
- Now high_df1 and df_extract2 are combined such that no data is wasted to get the best accuracy. Finally all the datasets are extended to the combo set.
- **By implementing this model the store can see profits eventually.**
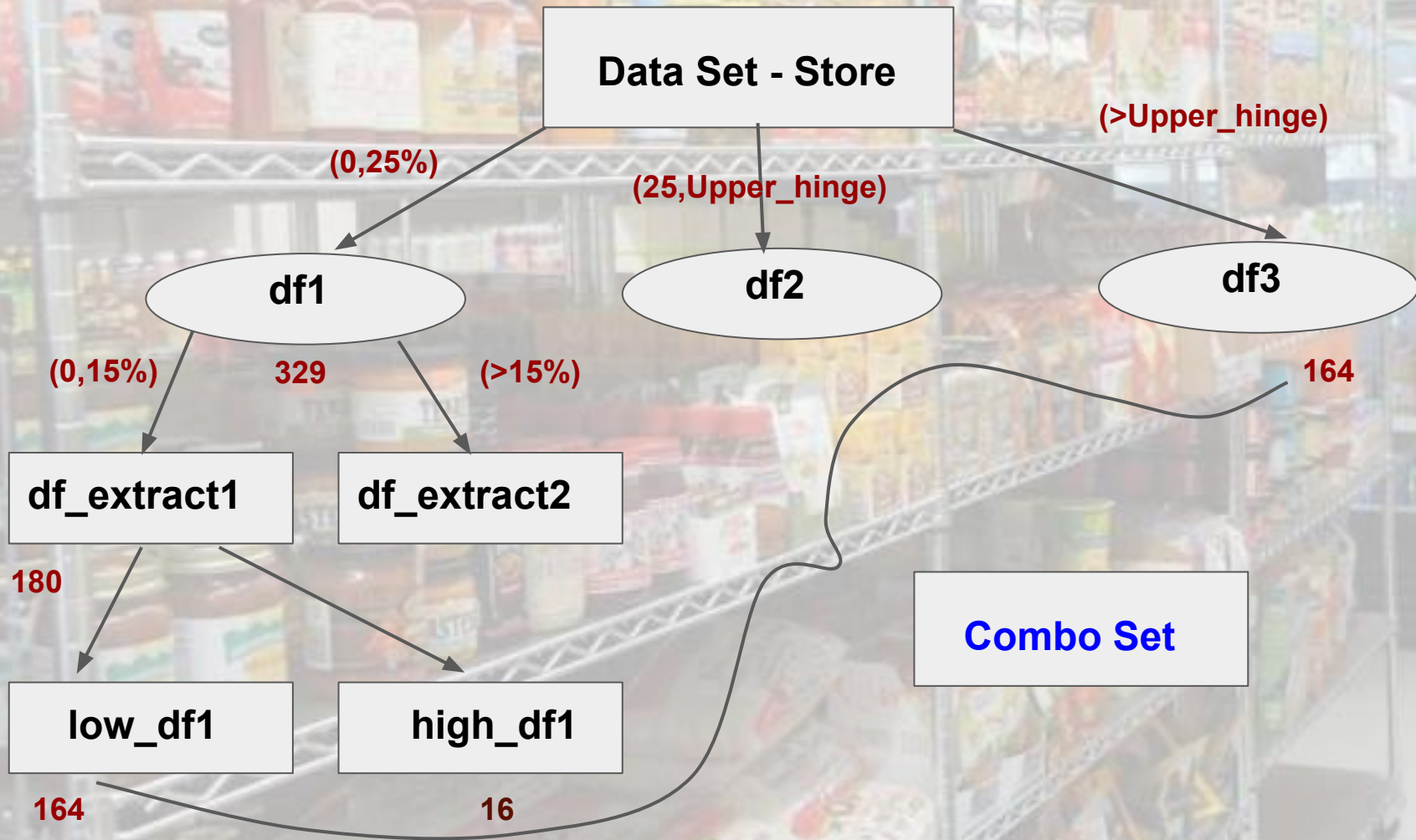
FINAL YIELD

- **After implementing this model, we can conclude that the model is successful by looking at the ACCURACY we got finally.**
- **From our solution we get the Accracy = 32.8%. Previously it is 19% in multi linear regression. The model almost doubled the accuracy which is surely a posive sign for the store with no loss for customers.**

Hence we can conclude that the solution is worth followed.