**APPROACH**

The basic idea was to get the target variable **id\_prop** feature values as our for the **id\_accs** feature of the dataset **Test**. The method followed to predict the target variable **id\_prop** is as follows:

1. **Method** **1**: First took the features from the **Accounts** dataset and merge with the **Accounts\_Properties** on the basis of **id\_accs** feature to obtain the features of **Accounts** into **Accounts\_Properties**. Did the same for the **Test** dataset on the basis of the feature **id\_accs**.
2. Then merged the two datasets **Account\_Properties** and **Test** on the basis of the feature **id\_accs**, hence the id\_prop feature was obtained.
3. **Method 2:** I had taken the categorical features of the dataset **Properties**- **sale\_status, portfolio and region\_\_c** and added them to the dataset Accounts\_Properties. There were some missing values after merging, so I predicted those values by separating the null and not null values as train and test data respectively. The algorithm I had used to predict them was XGBoost Classifier and Multi Label Classification.
4. Then I predicted the same categorical values for the test dataset by using Account\_Properties as train data and Test\_Data dataset as test data using **XGBoost Classifier and Multi Label Classification.**
5. After this I had the test dataset having 4 columns **id\_accs, sale\_status, portfolio, region\_\_c**. On the basis on these three categorical values I merged the Test dataset with the Properties dataset. Hence the id\_prop feature was obtained for the test data.
6. But there were few id\_accs that had Null values and very few id\_prop in its id\_prop column. To solve that I concatenated the id\_prop column of the test dataset that was obtained by using the above 2 approaches, thus obtaining a new dataset having **frequencies of all the id\_prop values** obtained. Then I filled the id\_prop column of those few id\_accs with the id\_prop having the **higher frequencies’ by** checking eachid\_accs with their frequencies **id\_prop values**. Hence all the **id\_prop** values for all the **id\_accs** of the test data were obtained.

**Quality checks performed / Errors found:**

No errors were found**.**

**Feature Engineering:**

Categorical values of the dataset were created into dummies thus increasing the number of features. The categorical features that were merged from Account \_Properties to the Accounts table created Null values. These values were predicted using the XGBoost Classifier and Multi label Classification.

**Key Trends:**

The id\_prop obtained with the help of two approaches was concatenated into a single dataset and their frequencies were calculated. The trend of these frequencies played a major role to predict the most popular id\_prop.

**Tools Used:**

I used **Anaconda jupyter (language Python)** notebook for this solution. The algorithm used were XGBoost Classifier and Multi Label Classification.