

IDEA

Our whole idea revolves around the output specification that was asked us to deduce it from given datasets. The first approach that we thought to merge every dataset then try to deduce the parameters on which we can classify a particular student performance and level of study, but this couldn't allowed us take any intuition from data, then later on we come up with a conclusion that we don't need all data, we need only relevant data through which we can deduce something out of it and try to predict output as per the requirement asked to us.

So we had used only three dataset i.e. test submission data, test details data and students data. As there was no bench mark was mentioned on dataset so it was bit difficult for us at that time to categorize any student on the basis of performance, so we had some assumptions on the basis of those we had prepared some bench marks to categorized student.

Following are our assumptions:-

Assumption 1: Check whether a student is able to achieve marks above average marks in targeted batch, if marks is more than average then good otherwise bad performance.

Assumption 2: Check if the student is not good in targeted batch but if the marks of a student is equal or above of other average targeted batch marks then we will suggest those courses to that student.

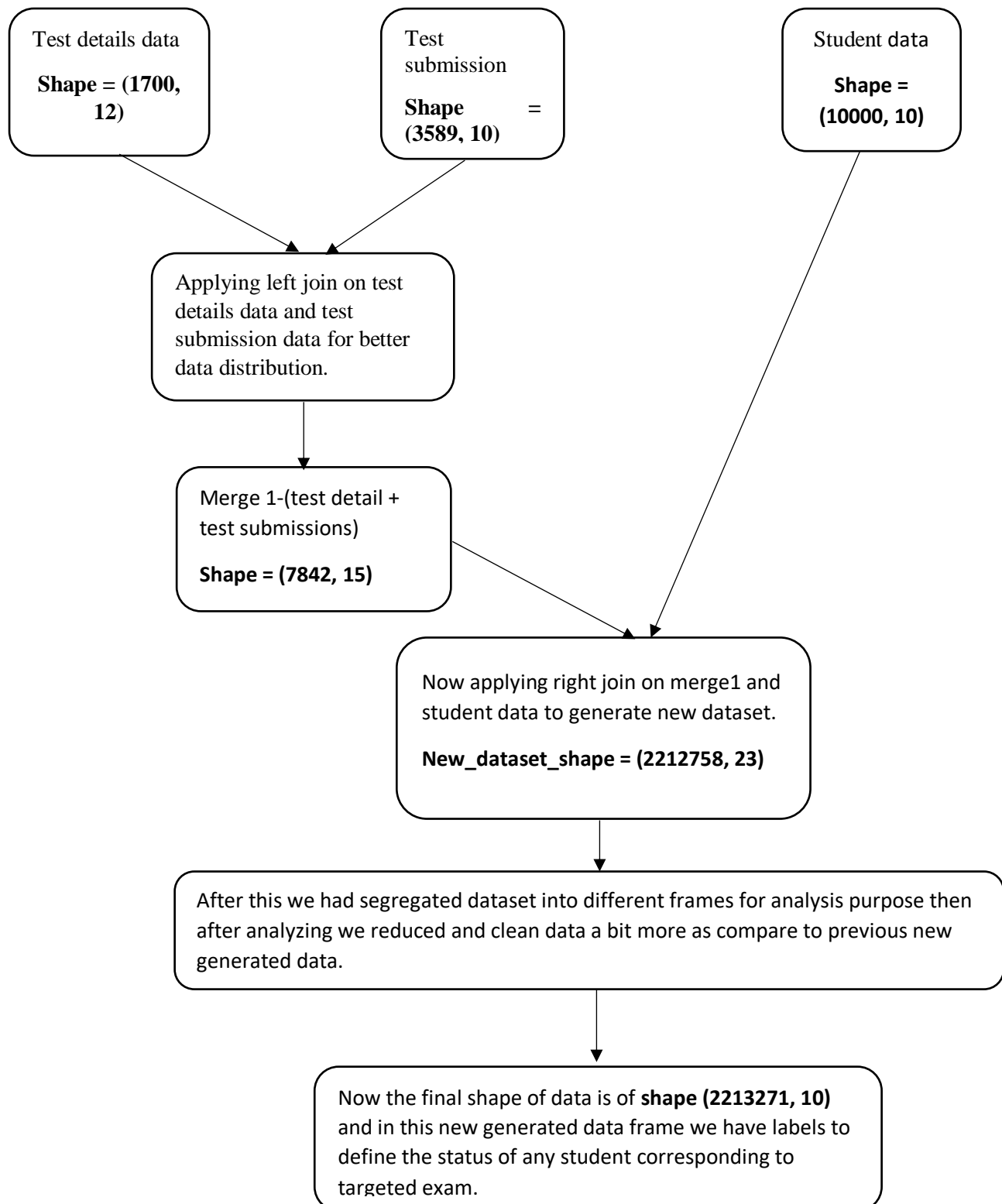
The above mentioned is our way to check the level of performance of student and to suggest whether they are eligible for any other courses. This is our whole idea and we had prepared data by segregating every targeted exam, so that we can analyze them correctly and using those analysis we can segregate students into two category.

Categories: - We had segregated the students into two categories:

1. Category zero: when student marks is less than average and not able to perform well.
2. Category one: when student marks is more than average and able to perform well.

We have used these labels to tag student's level and eligibility both in one go so that later on it help us to see which student is at which level and what other courses we can suggest them so that any particular student should not face any issue.

Flow chart of Data Preparation steps



Solution Approach

Initially we have faced some issue regarding segregation of students on the basis of performance and even in dataset there was nothing mentioned like what are the number question that a particular exam has, what is the base mark on which a particular student had scored marks. So these are the issues that we faced and then after generating new dataset

We had done following tasks to classify student: -

1. We have made a judicial use of **marks** column with respect to every targeted exams.
2. We used **average marks** as a bench mark to classify the student performance and eligible criteria.
3. **Example:** A student targeting **CTET** exam but the marks are not above average of all the marks of students who are targeting **CTET** exam, so we will say this student performance is low and hence need improvement or may be eligible for this exam and this student will be given **label = 0** otherwise **label = 1 for good performance**. In this way we have given label to every student for classification.
4. Now for course suggestion we have done some manual task of comparison like again say a student is targeting **CTET** exam and marks are not above average but if marks of any student is above average of any other course then we will suggest those courses to that student.

Above mentioned are our ways to categorize student performance level and suggesting courses to them, this all process is totally depend on **average marks** of corresponding **target exams**. This parameters may change if we have a solid bench marks to judge any exam, but this our whole idea to categorize any student.

Technology used for analysis and deep learning model Preparation

1. **Tableau:** Used for visualizing the data
2. **Pandas:** Used to create data frames
3. **Deep learning:** Used for making system learn for classifying student in two categories.
4. **Matplotlib:** Used for Visualizing model performance
5. **Dictionaries (Python):** Used to store the suggested courses for any student.

Online Resources used for converting json file to csv file

1. <https://json-csv.com/>
2. <https://apps.thejeshgn.com/json-csv/>

Code and Tool / Frameworks

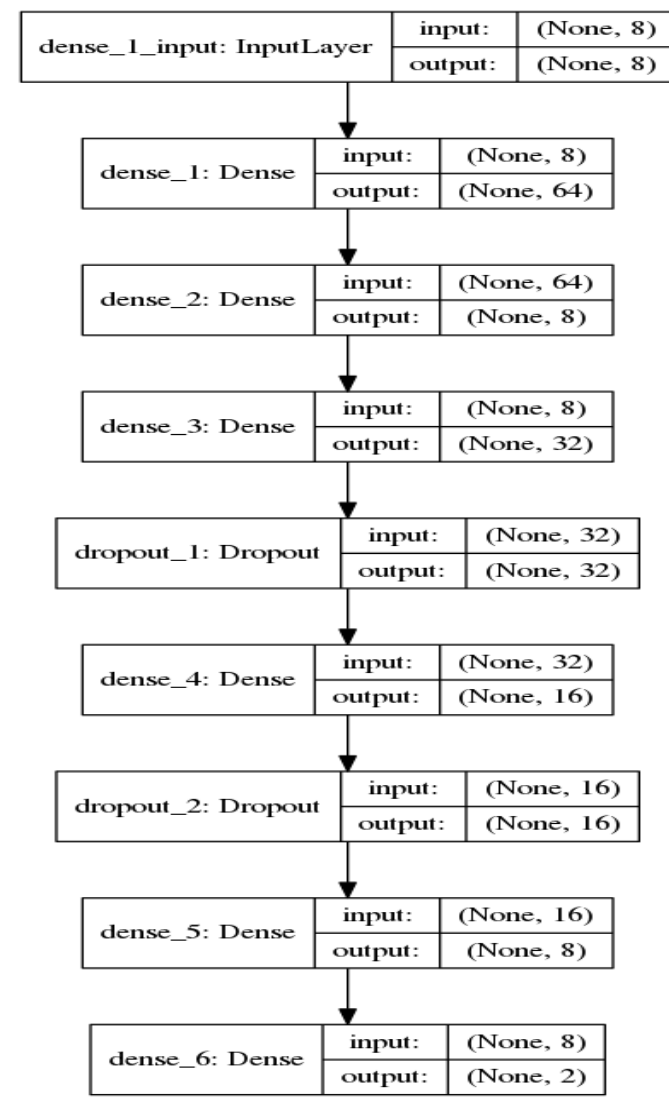
Code link: <https://github.com/GauravDaharia20/Project-Decipher>

We are living in the era of automation where we collaborate our work with machine intelligence, we had tried to use this intelligence to make something automated so that we can achieve our goal for this project.

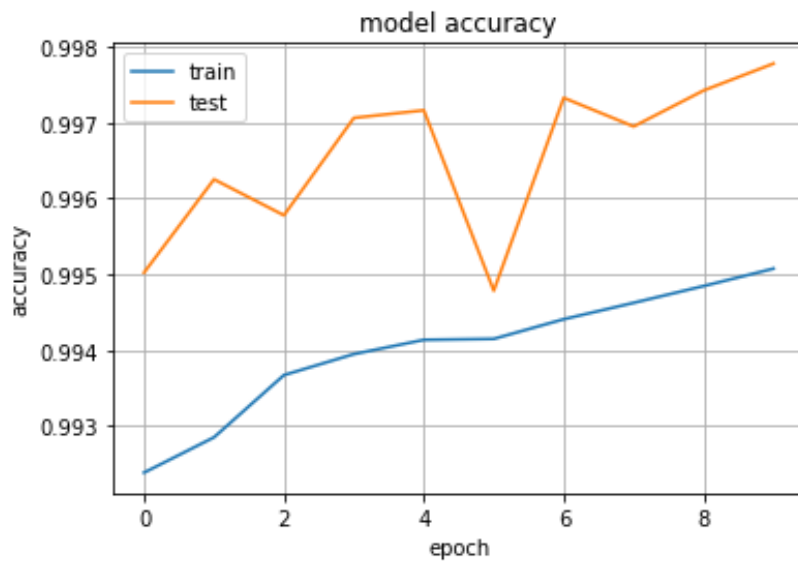
We have developed a Deep Neural Network which classifies the student performance and eligibility in two classes i.e. high (1) or low (0), we have used our prepared data to make system learn and help us to classify any student. It has a wide use in real life and it also cutoff our manual work up to some extent.

We have used **Keras** for preparation of this neural network and **Tableau** for data insight visualization.

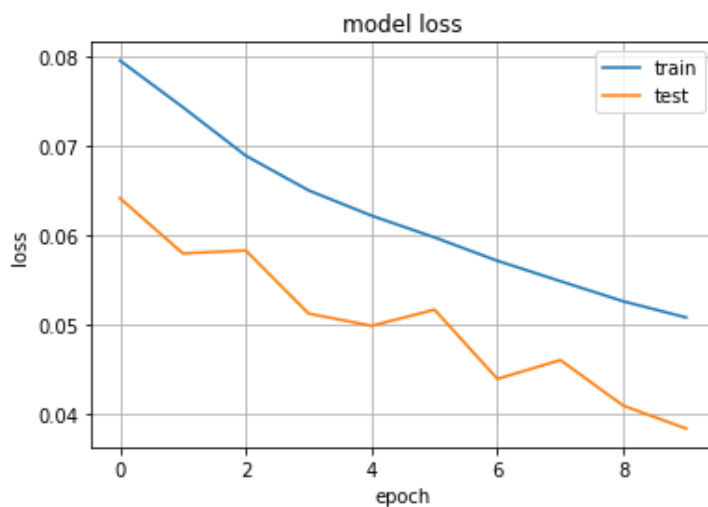
DNN Architecture



Model Accuracy Graph



Model loss Graph



Model accuracy graph shows model accuracy achieved at every iteration and here we have used only 10 iteration to train our model, as data was so clear it was able to achieve good accuracy at minimum iteration.

Accuracy Achieved: 99.7%

Model Loss graph shows the loss get reduced at every iteration and this has to be minimized because we want our model should achieve **good accuracy** and **loss should be minimum**. This model loss is around **3%** which acceptable.

Method for Suggesting Courses as Performance of Any Student

The method that we had used here to suggest other courses to any student is on the basis of marks that a student had obtained in the targeted courses. As we have said in **idea** section of this report that we had done some manual work for suggesting courses. So following are the steps that we followed: -

1. Store every targeted batch average marks in a list for comparison purpose.
2. Create a dictionary (python) of those store average marks with respect to index and these indices will act like a label to identify the course name.
3. Create another dictionary where key will be the student id and values will be the list of courses for which a student is eligible.
4. After storing values in dictionary create dataframe for good data representation.

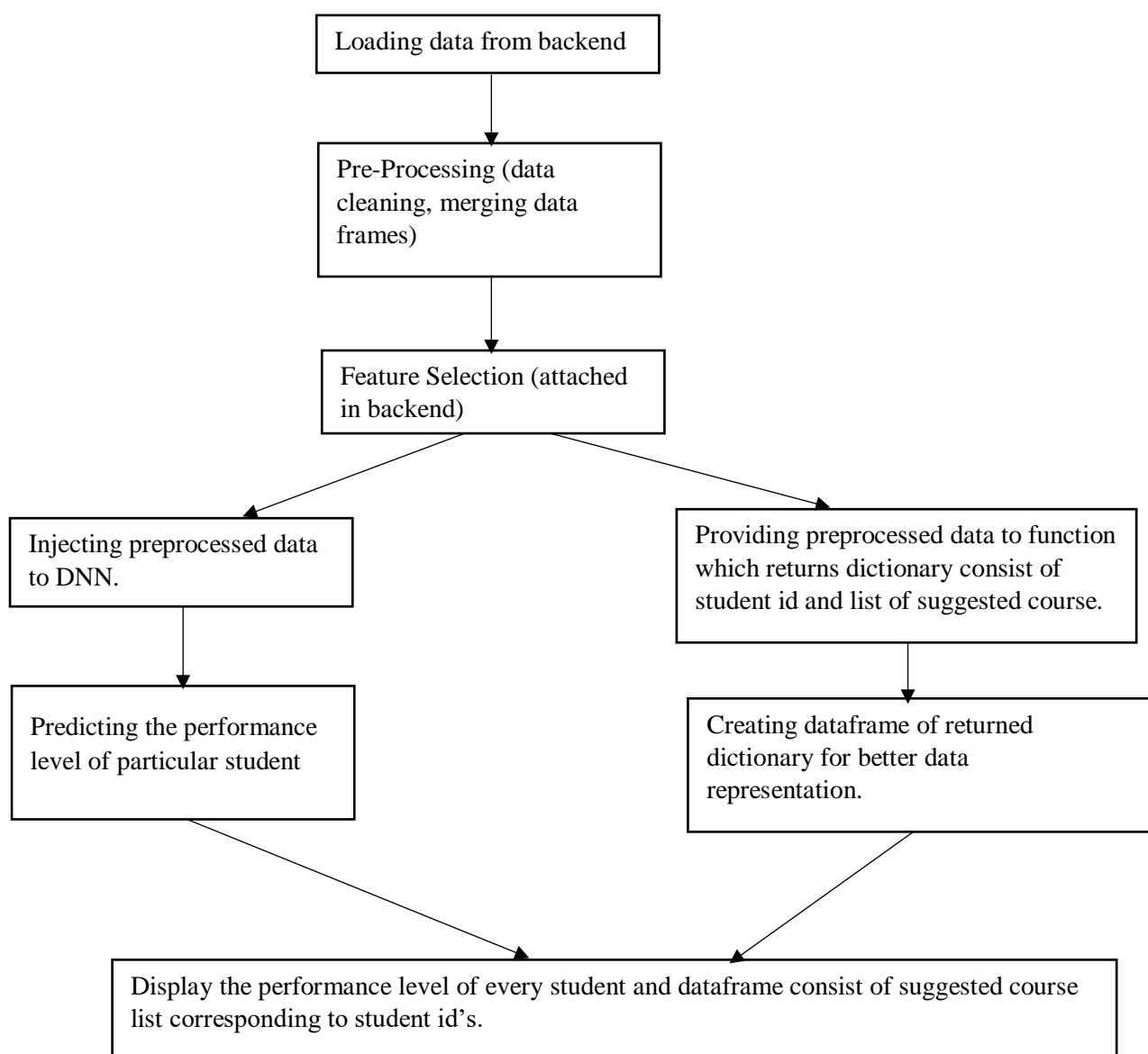
We had suggested other courses to students on the basis of marks, if let's say for an instance a student is in **CTET** target batch but the score obtained by that student is not above average of that particular target batch, so in that case how we are able to suggest other courses to that student as the performance is not that much good, for that reason we decided to compare the marks of every student with the average marks of all target batch courses that we have stored on list and dictionary, if the marks is above average of any other target batch then we will add that index (**i.e. acting like an id for course**) to list and will map that list to corresponding **student id** in dictionary that we have created in **step3** above mentioned steps. In this way any student will be able to know about other available courses.

For more detail please go through **Course suggestion Algo Page taker.ipynb** file which is shared in above mentioned GitHub link. In this file please go through cell number 300 to cell number 364, through this cells it will make more clear to understand our thought process for course suggestion technique.

Strategy to Run in Production Environment

This particular project require few things like GPU support and a server for data pre-processing as we have prepared our dataset by sub sampling the merged data so for that reason it require a server. Secondary thing required is a GPU support because as we have developed a deep neural network which classifies student in two level i.e. good or bad performance and on GPU the model will work efficiently and the cost of prediction will be less on real time.

Below following work will define the running strategy of model and course suggestion part:

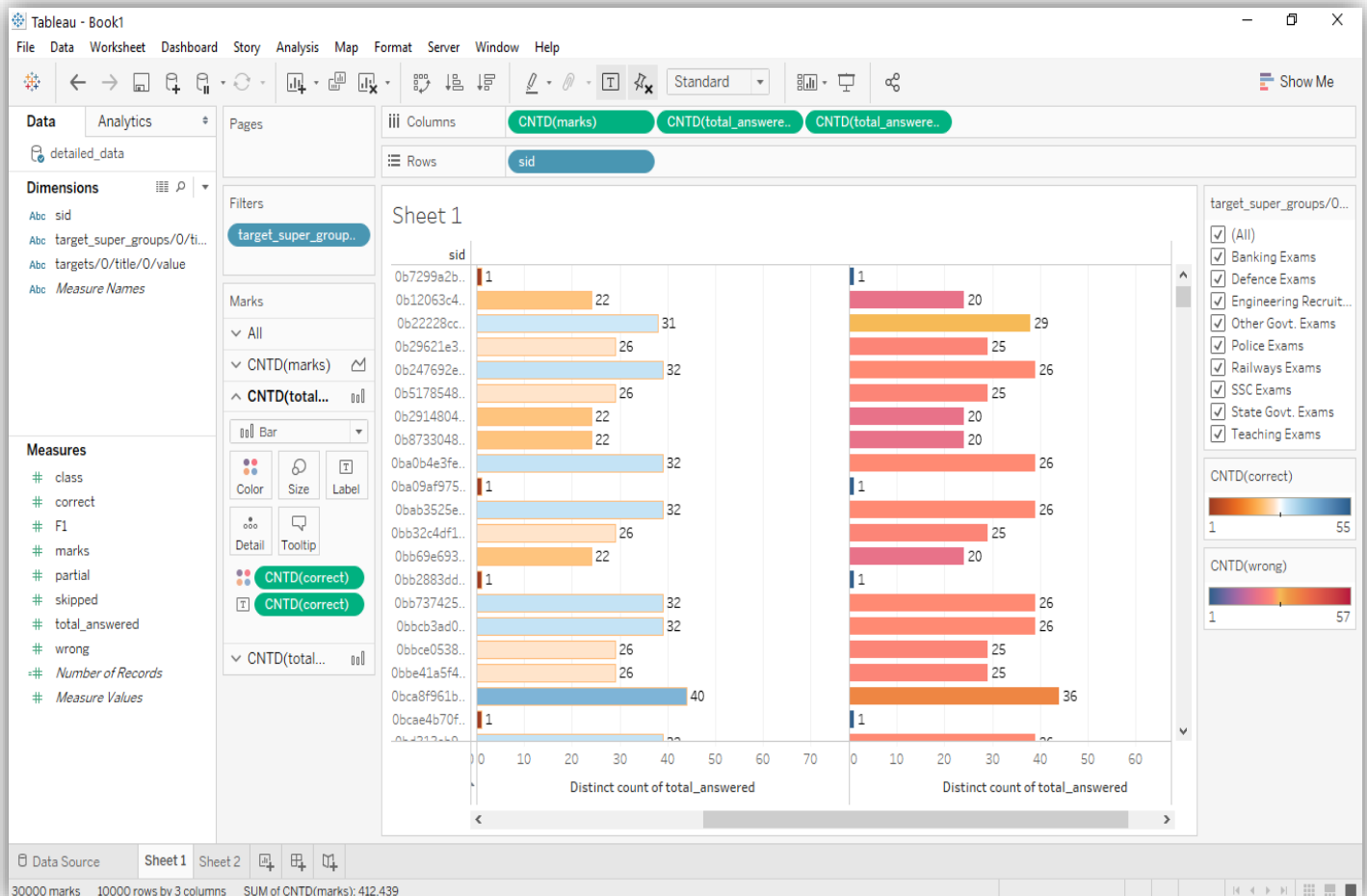


Visualization

1. Objective 1:

Representation of the number of questions answered, out of which numbers of questions are correct and number of questions are wrong.

Tool used here for visualization is **Tableau Desktop**.



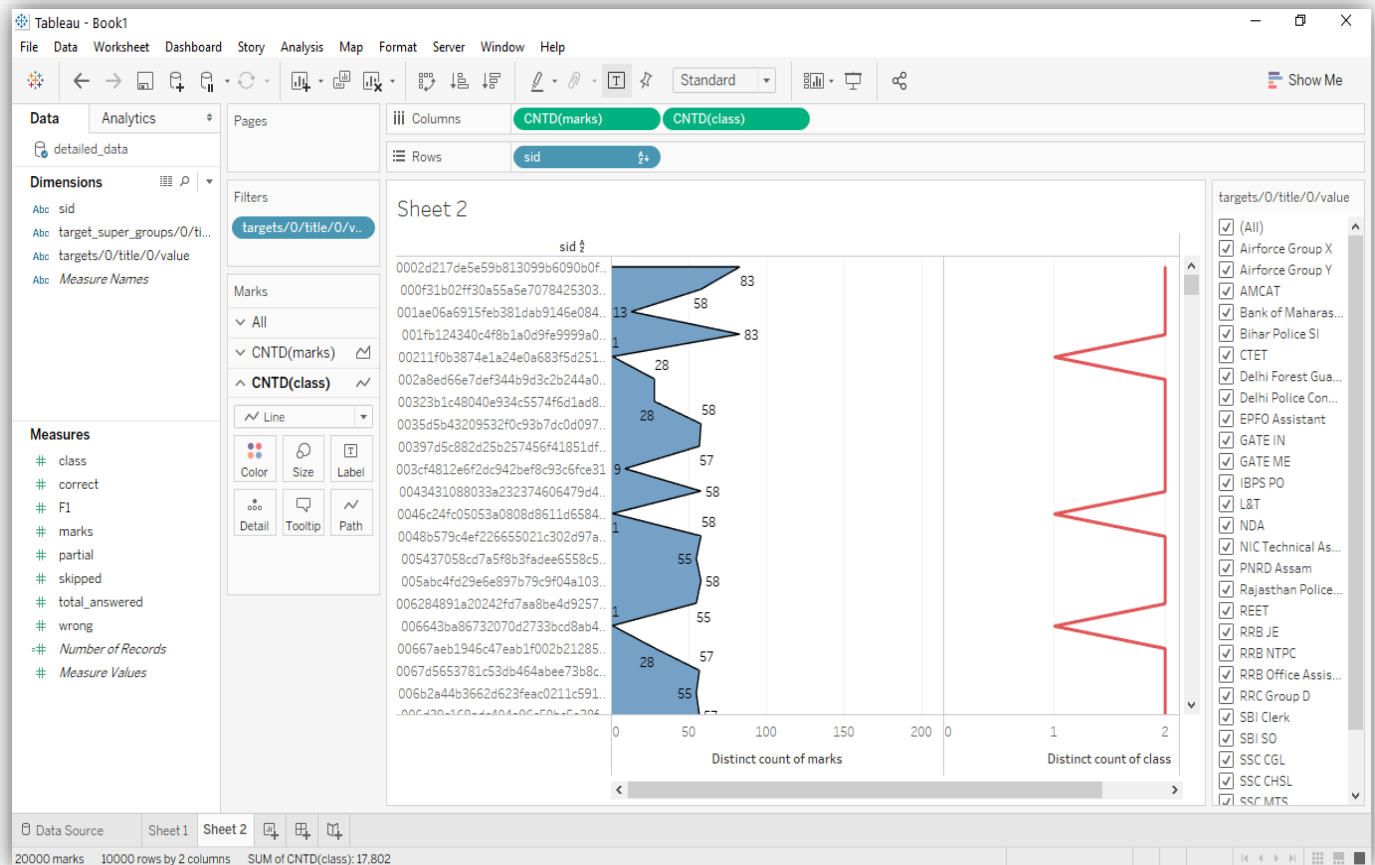
On the basis of super target groups, two Bar graphs showing the total number of question Answered, by the students

1. First bar graph shows the number of questions answered and label showing the correct answers.
2. Second bar graph shows the number of questions answered and label showing the wrong answers.

2. Objective 2

Representation of all the marks of students and their class, whether the students fulfill their targets with the given marks or not.

Tool used for visualization is **Tableau Desktop**.



The Graphs shows the total marks scored along with the class whether the student scored above average or below average.

The value 2 in class shows the student scored below average marks and value 1 shows the student scored above average and fit for their suitable targets.