STUDENT GRADE IMPROVEMENT INHIGHER STUDIES

Sandhya P. Pandey
Assistant Professor,
The S.I.A college of Higher Education, Dombivili(E), Thane, Maharastra.

Abstract: *In India Higher educational institutions grown rapidly. These Institutions focused on quality education for students. Educational sector has a lot of data that can produce valuable and relevant information for improvement of quality Education. This data can be used to increase the quality of educational instutions. Big Data tools plays important role for creation of strategic data.This research is focused on improvement of student's performance in final year of Higher Education based on previous performance using Big data tools.In different courses different heads are there for marks and grades. In this research data assumption is based on one specific BSc(Information Technology of Mumbai University) course.*

Keywords:*Higher Education, Grades, Big Data*

INTRODUCTION

Now a days many big data tools are available to extract useful information. It allows users to analysedata(which is huge in number) from many different dimensions or angles, categorize it, and summarize the relationships identified.

Education field is very important area in terms of any country's growth and development. To achieve this, data could be collected with different sources, pre-processing can be done using big data and data mining tools to prepare results to improve performance of students.This research is focused on improvement of student's performance in Higher Education using Big data tools. It is continuation(with data analysis tool) of research paper which published earlier (without using any tool).

REVIEW OF LITERATURE

Search engines helps to study to find out current trends of results and future scope for improvement in this. After searching Grades improvement in education, all data was related to

School grades or how to improve grades of institutes. Some data was related to improvement in education policy. But for different higher studies courses grades studies were very less.

RESEARCH METHODOLOGY

In this research the secondary sources of data and related study were used. Sampling method is Random Sampling. For security purpose real data is not used for analysis.

DATA ANALYSIS

Tools Used for analysis

- Data Ingestion -- Datasets
- Data Storage -- HDFS
- Data cleaning,processing and analysis -- Mapreduce(Pig,HIVE)
- Visualization –Tableau

Areas covered for analysis

Based on the specific course following data created and analysed. Data heads can be varied based on course requirement:

Variable	**Description**
10^{th}	Grade 10^{th} marks
12^{th} / Diploma	Grade 12^{th} marks/Diploma
CT1-sem1	Class Test Sem 1
CT1-sem2	Class Test Sem 2
CT1-sem3	Class Test Sem 3
CT1-sem4	Class Test Sem 4
Att1	Sem 1 average Attendence
Att2	Sem 2 average Attendence
Att3	Sem 3 average Attendence

Att4	Sem 4 average Attendence
Grade1	Sem 1 Grade
Grade2	Sem 2 Grade
Grade3	Sem 3 Grade
Grade4	Sem 4 Grade
Prac1	Sem 1 Practical performance
Prac2	Sem 2 Practical performance
Prac3	Sem 3 Practical performance
Prac4	Sem 4 Practical performance

Table 1.0

DATA ANALYSIS-IMPLEMENATATION

Step for Analysis:

1. Login Putty, transfer Files
2. Store the files in HDFS
3. Pig processing
4. Hive Processing
5. Rename the part files
6. Copy them to Local Unix (Putty) & Transfer it to the System
7. Visualization

Login Putty, Transfer Files

The primary goal of Putty is to become a multi-platform application capable of executing in most operating systems. Before begin, log in to Ambari Console using the IBM DEMO CLOUD, copy the ssh link and paste it into Putty.

And then Save & Load.

- Using username "sandhyapramod".
- sandhyapramod@systemt.datascientistworkbench.com's password:

After login to Putty server, create a folder either by Command or WinSCP.

Folders are created just to provide a clear thought processing rather than providing confusion while locating the files. In WinSCP, the first half is for System, and the other one is Putty, we can drag and drop all the files we want to use for processing purpose into the folder of choice. And we can rename the files by the same way we do it on our system, through WinSCP. Once done, Check For the Files in the directory.

Store the files in HDFS

The default big data storage layer for Apache Hadoop is HDFS. HDFS component creates several replicas of the data block to be distributed across different clusters for reliable and quick data access. It comprises of 3 important components-NameNode, DataNode and Secondary NameNode. HDFS operates on a Master-Slave architecture model where the NameNode acts as the master node for keeping a track of the storage cluster and the DataNode acts as a slave node summing up to the various systems within a Hadoop cluster.

Creating a College Directory and this Directory Includes Multiple Subdirectories based on the Categories like Attendence, Practicals, Hsc& Tenth, Grades and Class Tests.

Once the directories are created as per differentiating them according to the datasets, we can then transfer the files from putty server storage to HDFS storage by using the following command.

[sandhyapramod@iop-bi-master ~]$

hdfsdfs-put/mnt/home/sandhyapramod/Datasets/Attsem1.csv'/user/sandhyapramod/College/Attend'

This command could be used for transferring the other datasets as per the subdirectories created for them and also by changing the dataset name.

Pig Processing

The pig was designed to make Hadoop more user-friendly and approachable by power-users and nondevelopers. Apache Pig is a high level extensible language designed to reduce the complexities of coding MapReduce applications.

Loading Pig…

COMMANDS –

#pig

For Example Attendence File 1 (Attsem1.csv)

```
grunt>attdata1=LOAD '/user/sandhyapramod/College/Attend/Attsem1.csv' USING
PigStorage(',')  AS (StudentId:chararray,StudentName:chararray,AttendenceSem1:int);

grunt>headless_data = FILTER attdata1 BY(StudentId!='STUDENTID');

grunt>reldata = FOREACH headless_data GENERATE StudentId,StudentName,
AttendenceSem1;

grunt>finaldata = FOREACH reldata GENERATE (chararray)StudentId,
(chararray)StudentName,(int)AttendenceSem1;

grunt>STORE finaldata INTO '/user/sandhyapramod/Attend/AttendS1' Using PigStorage(',');
```

Referring to the above example, the pig is basically loading the data set from the HDFS to Pig Storage, removing the headers if any, and relating each of the column as per the headers and finally generating the final dataset which will be having no junk values and storing it back into the Pig Storage.

HIVE

Hive enables easy data summarization, ad-hoc querying and analysis of large volumes of data. Hive was created to make it possible for analysts with strong SQL skills to run queries on the huge volumes of data that Facebook stored in HDFS. Today, Hive is a successful Apache project used by many organizations as a general-purpose, scalable data processing platform.

Hive developed by Facebook is a data warehouse built on top of Hadoop and provides a simple language known as HiveQL similar to SQL for querying, data summarization and analysis. Hive makes querying faster through indexing.

Example:

```
[sandhyapramod@iop-bi-master ~]$
```

beeline -u jdbc:hive2://iop-bi-master.imdemocloud.com:10000/ -n sandhyapramod

//command for running hive.

Connecting to jdbc:hive2://iop-bi-master.imdemocloud.com:10000/

Connected to: Apache Hive (version 1.2.1-IBM-12)

Driver: Hive JDBC (version 1.2.1-IBM-12)

Transaction isolation: TRANSACTION_REPEATABLE_READ

Beeline version 1.2.1-IBM-12 by Apache Hive

Creating a database with location

0: jdbc:hive2://iop-bi-master.imdemocloud.com>

create database CollegeD location '/user/sandhyapramod/hive/warehouse/CollegeD';

No rows affected (0.028 seconds)

0: jdbc:hive2://iop-bi-master.imdemocloud.com>

useCollegeD;

No rows affected (0.02 seconds)

For Attendence Table 1

0: jdbc:hive2://iop-bi-master.imdemocloud.com>

create external table AS1(StudentId varchar(10), StudentName varchar(60), ASem1 int)

0: jdbc:hive2://iop-bi-master.imdemocloud.com>

row format delimited

0: jdbc:hive2://iop-bi-master.imdemocloud.com>

fields terminated by ','

0: jdbc:hive2://iop-bi-master.imdemocloud.com>

location '/user/sandhyapramod/Attend/AttendS1'

0: jdbc:hive2://iop-bi-master.imdemocloud.com>

TBLPROPERTIES(0:jdbc:hive2://iop-bi-master.imdemocloud.com>'skip.header.line.count'='1');

Referring to the above example, the beeline command triggers to start the hive functionality and creating a database is mandatory as HIVE is using SQL queries so without database the file would be scattered to any default locations, once the database is created,we need to use the database and start generating the tables with reference to the datasets that have been created from PIG process.

Once the tables have been created, we can join all the tables in a database by using joins, clubbing them all together to form a single table which can be used for analysis.

Rename the files

Once the joining is done, the files are saved as '000000_0' file and these files shall be renamed accordingly and copied to the local storage for access and further analysis purpose.

Example:

INSERT OVERWRITE DIRECTORY '/user/sandhyapramod/finalfile/Final'

Row Format Delimited

Fields terminated by ','

select * from final;

[sandhyapramod@iop-bi-master ~]$ hdfsdfs -ls finalfile/final

Found 1 items

-rwxrwx--x+ 3 sandhyapramodsandhyapramod 10908 2017-06-25 08:13 finalfile/final/000000_0

[sandhyapramod@iop-bi-master ~]$ hdfsdfs –mv

'/user/sandhyapramod/finalfile/final/000000_0'

'/user/sandhyapramod/finalfile/final/FinalData.csv'

[sandhyapramod@iop-bi-master ~]$ hdfsdfs -ls finalfile/final

Found 1 items

-rwxrwx--x+ 3 sandhyapramodsandhyapramod 10908 2017-06-25 08:13finalfile/final/FinalData.csv

Referring to the above, the table generated using HIVE will be inserted i.e stored into the HDFS directory from the HIVE database, after the file is stored we can rename that part file to the appropriate naming convention.

Copy them to Local Unix (Putty) & Transfer it to the System

[sandhyapramod@iop-bi-master ~]$ hdfsdfs –copyToLocal

/user/sandhyapramod/finalfile/final/FinalData.csv

/mnt/home/sandhyapramod/Final

The files after being copied to local could therefore can be downloaded to the desktop or simply dragged and dropped using WINScp.

Visualization

For visualization, TABLEAU software was used, with a purpose to clarify the view of the statistics and analysis has been done on a particular file of data.

RESULTS

1. **Student along with attendance in each semesters**

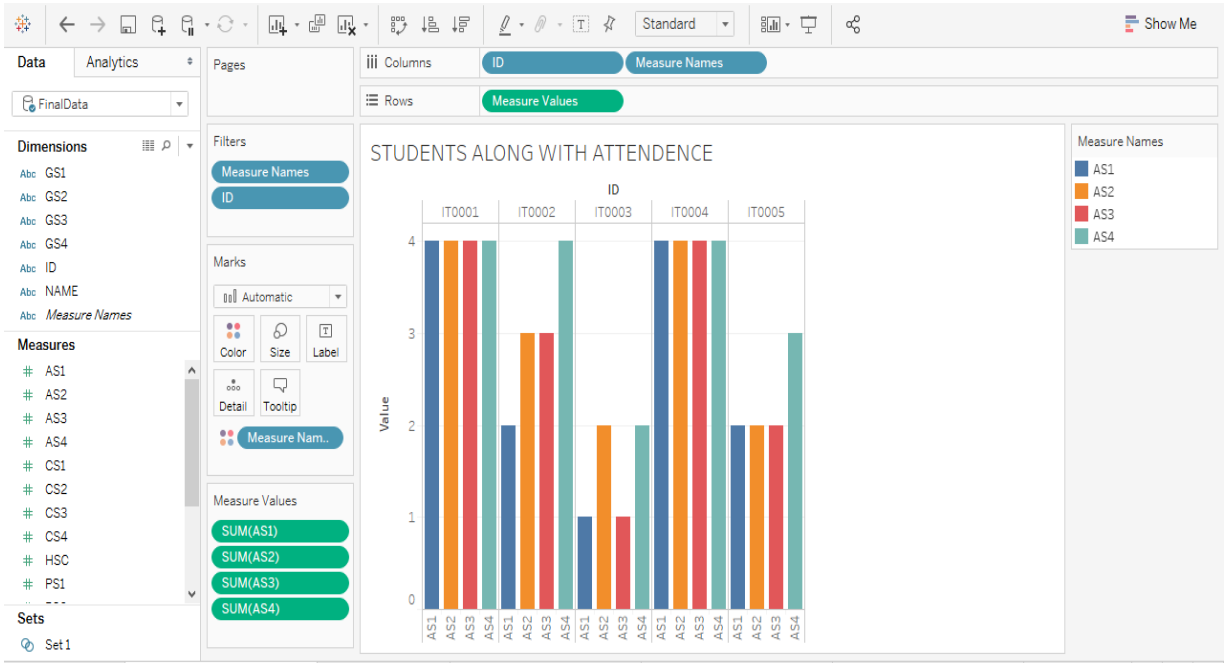

Table 2.0

2. Students along with their grades in each semesters

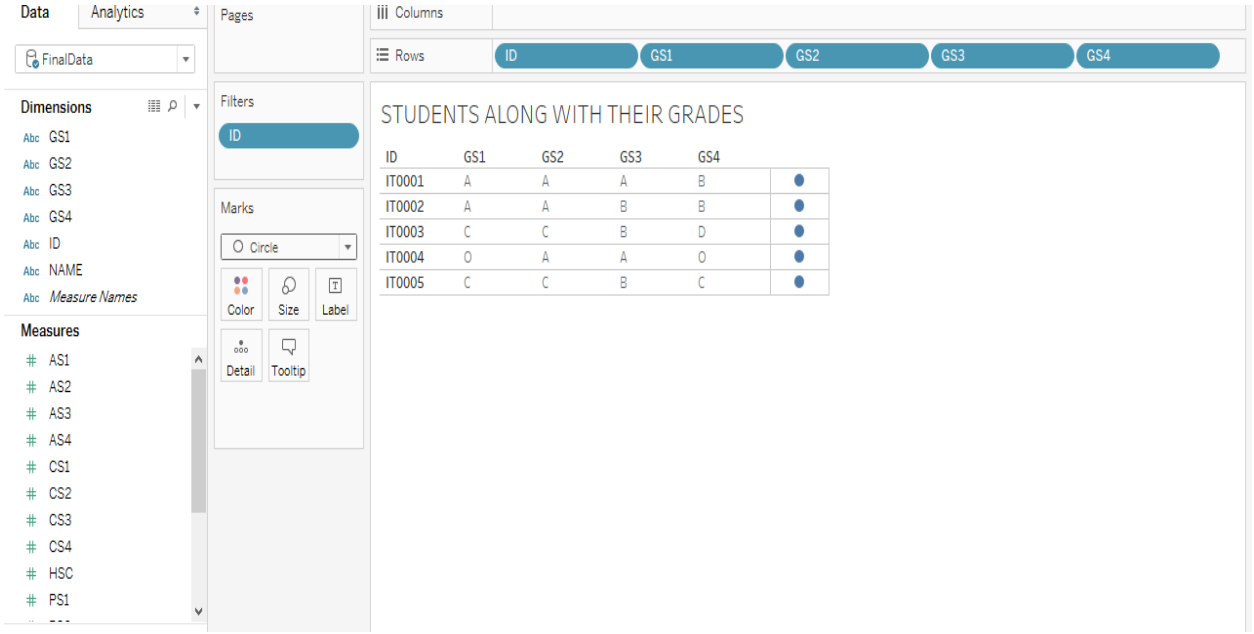

Table 2.1

3. Students along with their marks in Class Tests/Semester

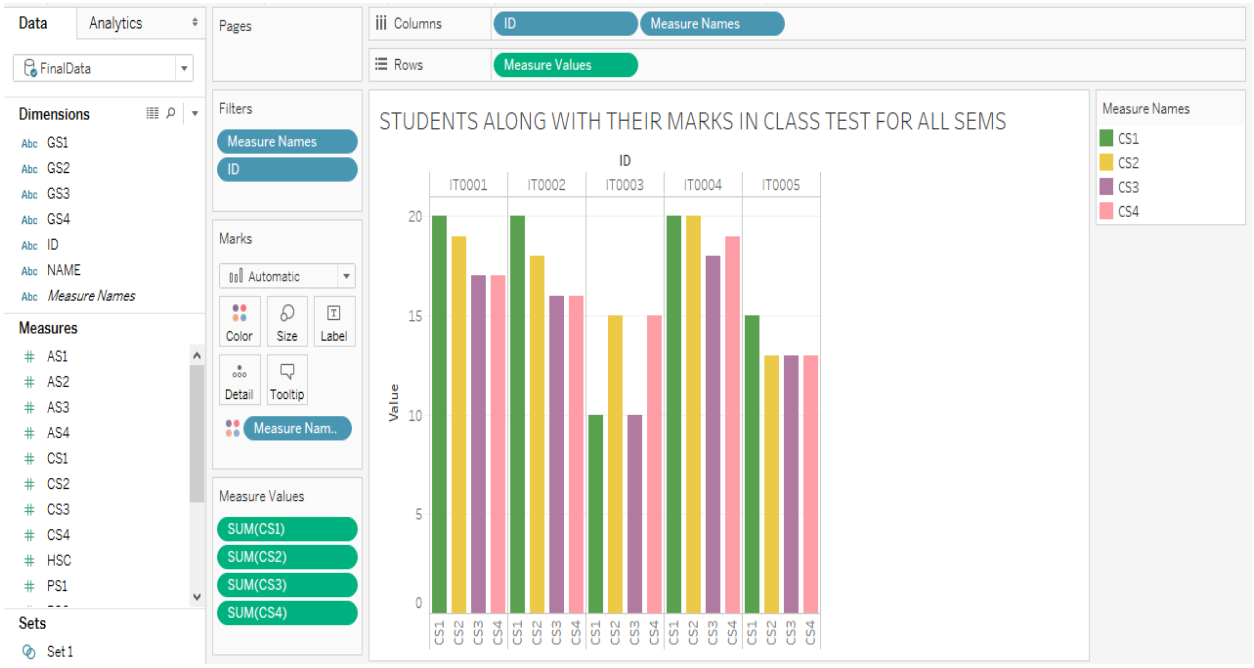

Table 2.2

4. Student along with their Practical Marks/Semester

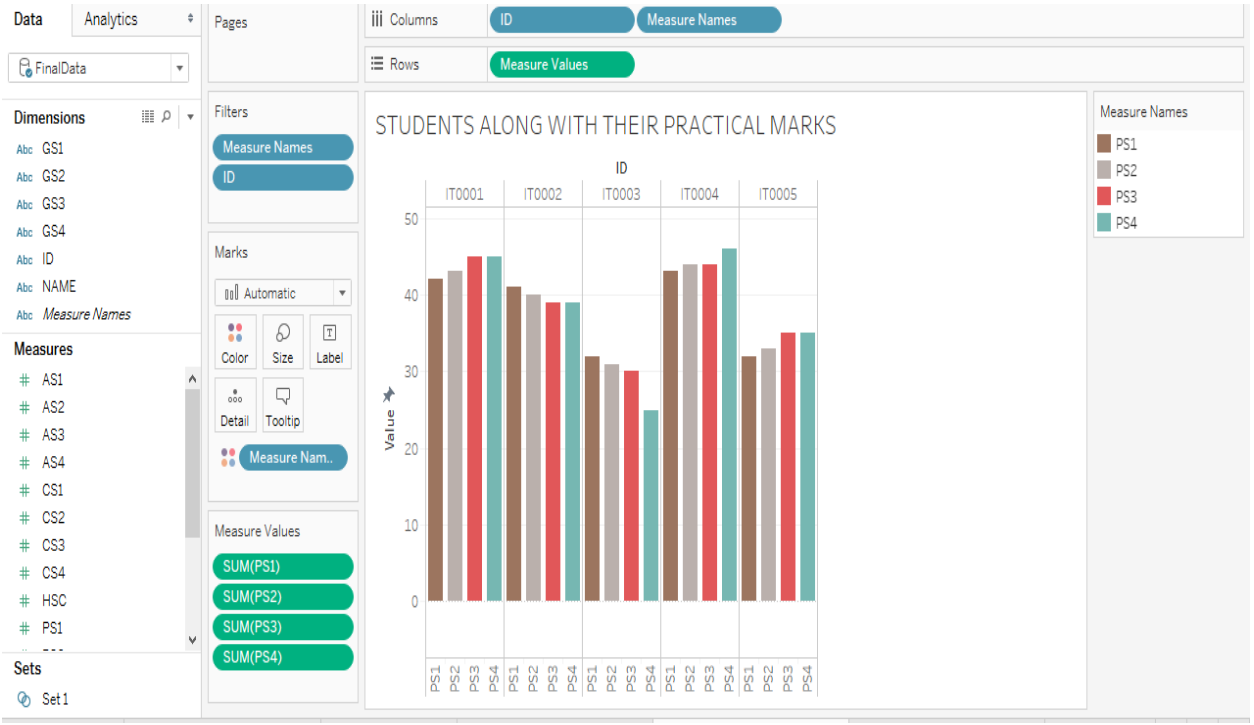

Table 2.3

5. Dashboard 1 (includes every sheet from top in one)

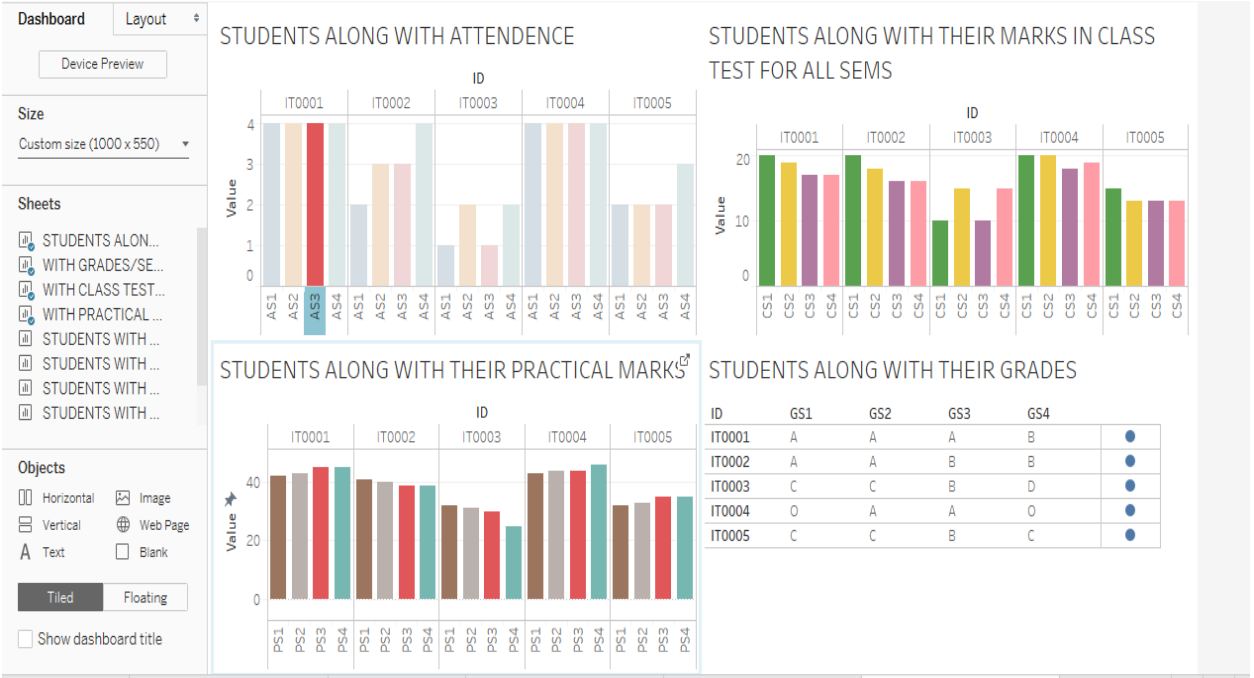

Table 2.4

6. Students with their marks in Semester 1

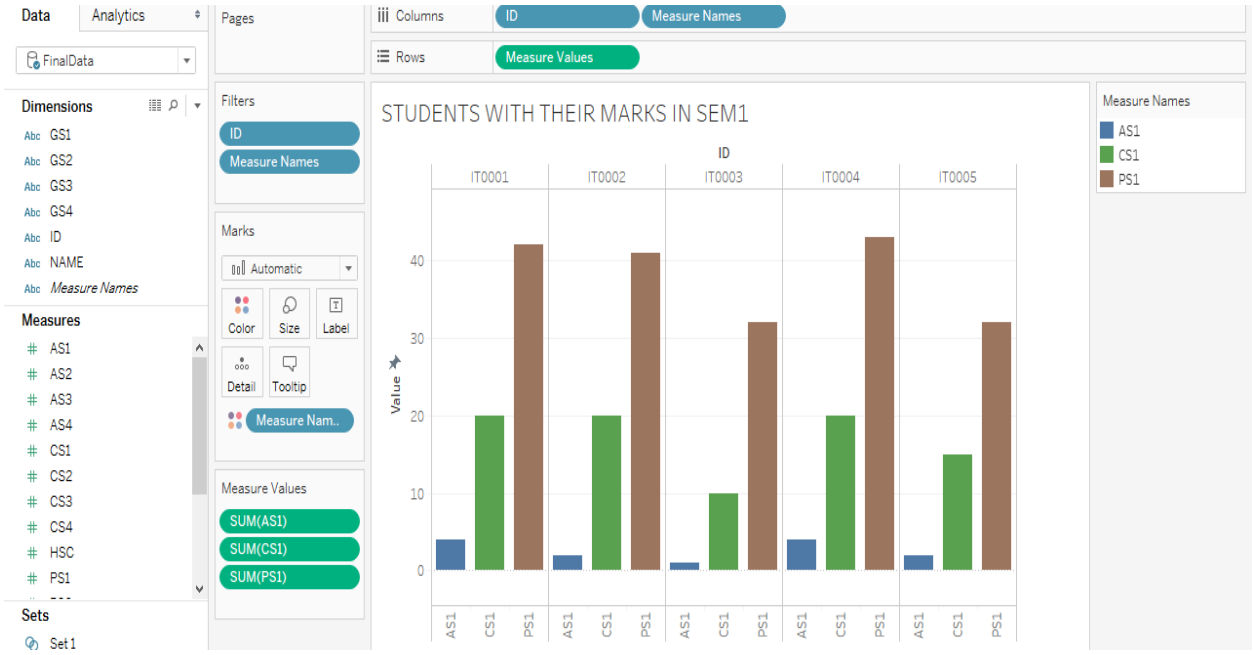

Table 2.5

7. Students with their marks in Semester 2

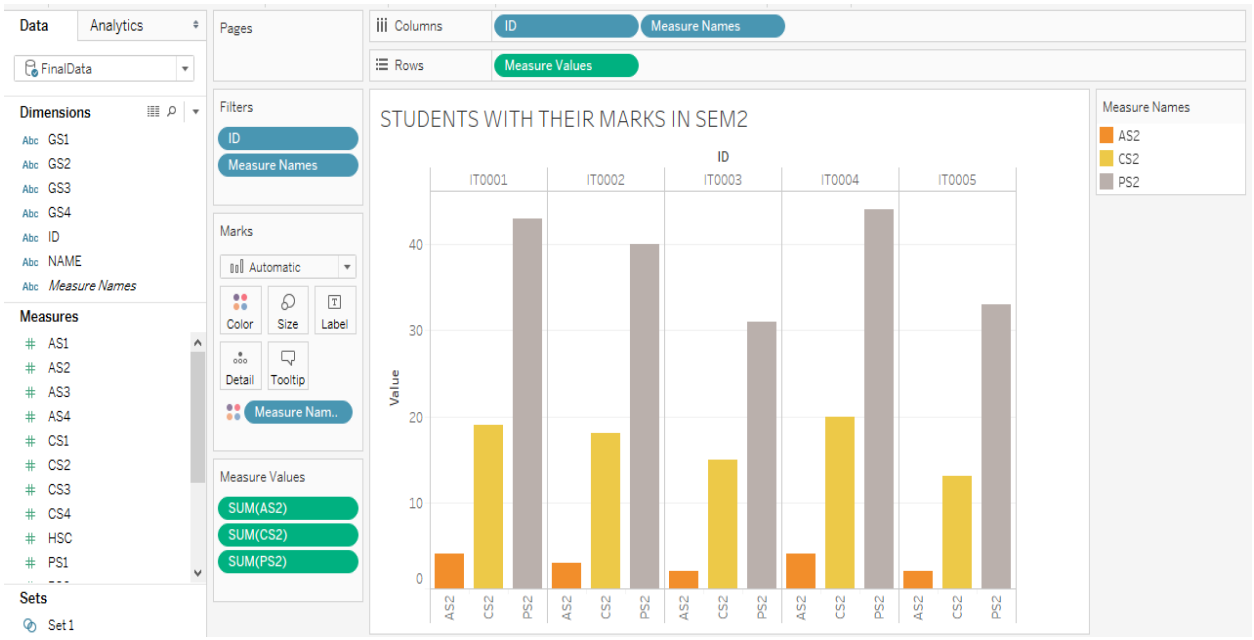

Table 2.6

8. Students with their marks in Semester 3

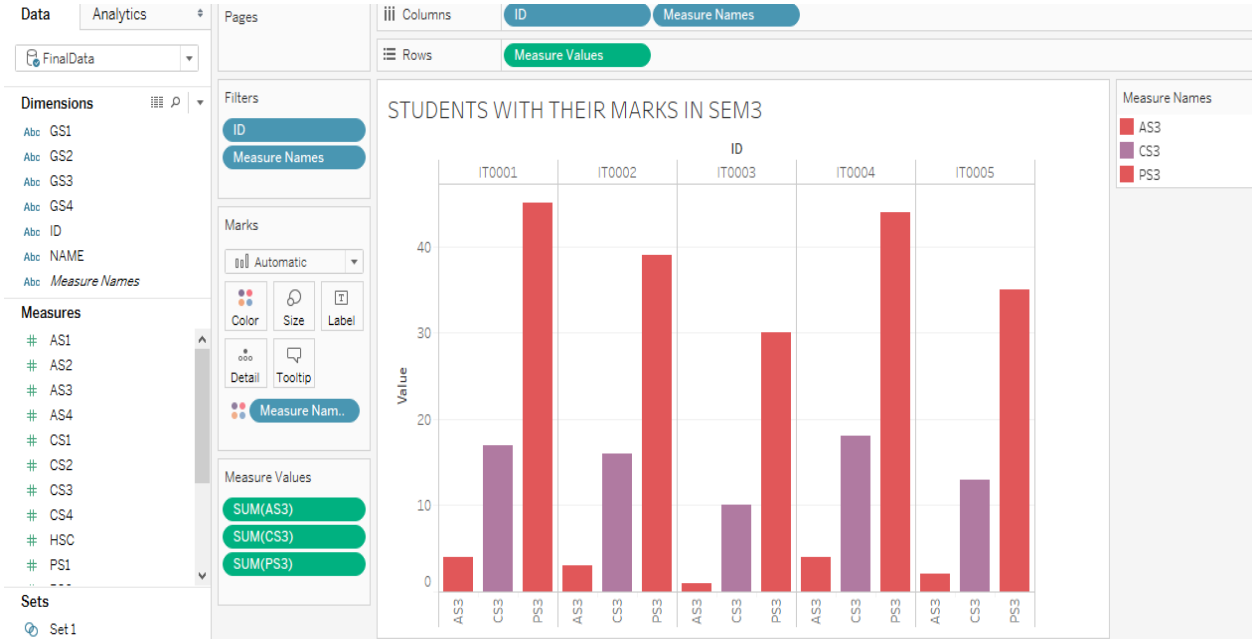

Table 2.7

9. Students with their marks in Semester 4

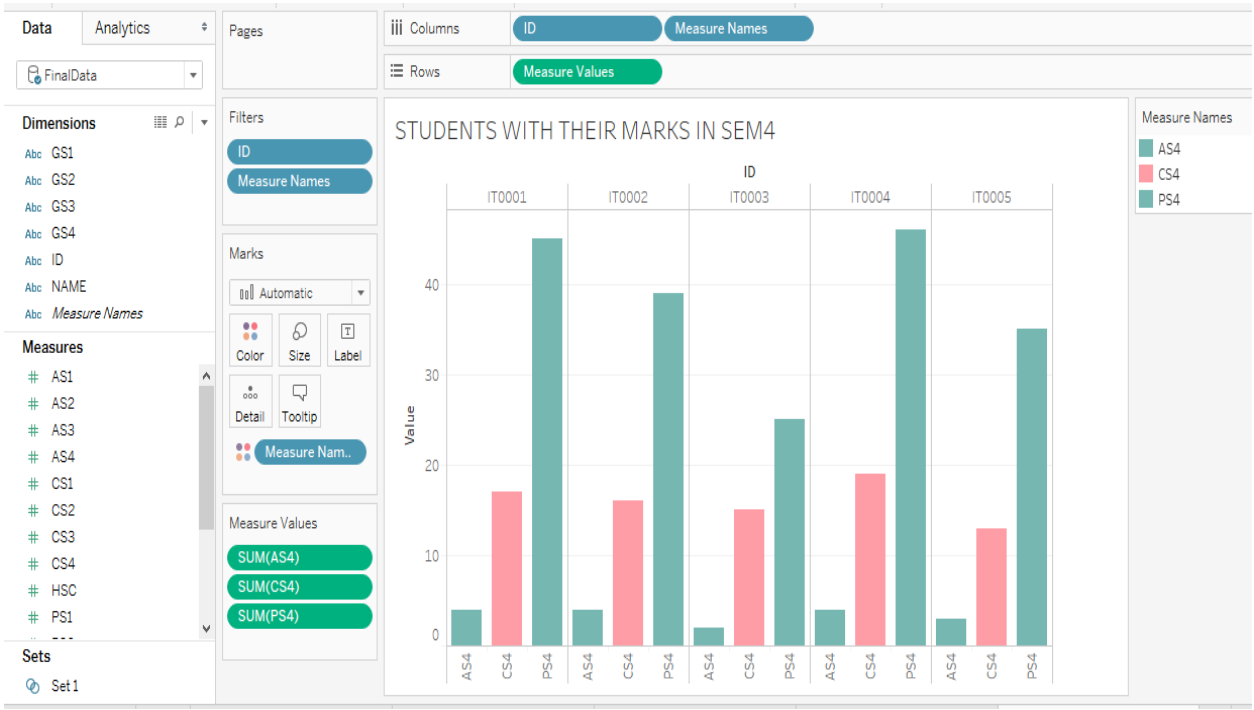

Table 2.8

10. Combining all of the marks together for a student.

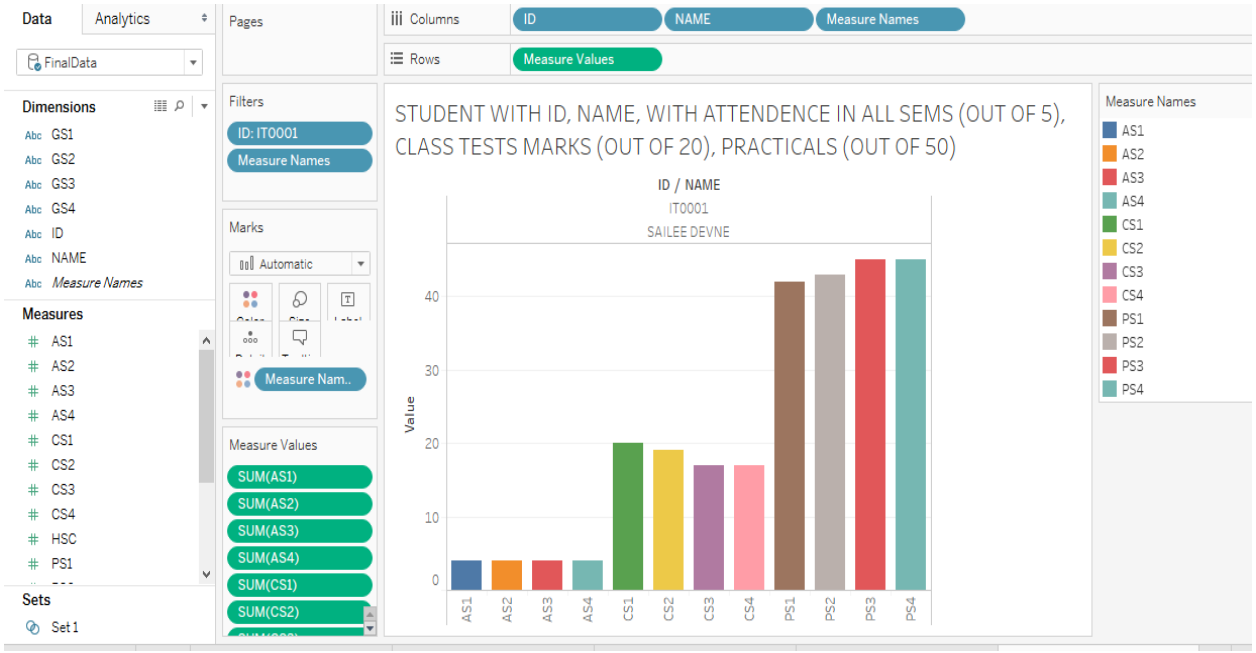

Table 2.9

11. Dashboard 2 (with a single student report)

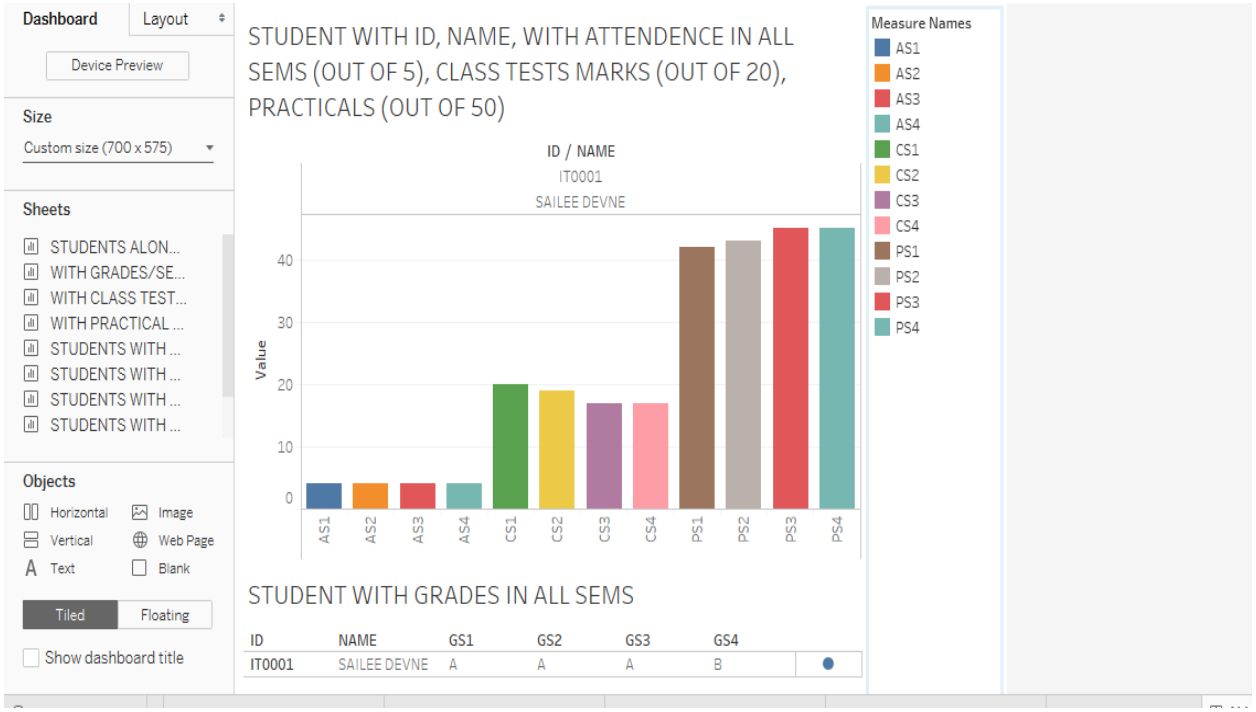

Table 2.11

CONCLUSION

Current Higher education system not predicting grades for students based on student's previous performance. Improvement of grades in higher education is a recent exploring field and this area of research. Big data is new trend and is achieving popularity because of its potentials to handle educational institutes huge and variant data.Because of data set variations and huge size, big data tools can be suitable for analysis.

FUTURE SCOPE

This type of study will help to the students and the teachers to improve the knowledge of the students further. It can also be useful to identify those students which needed special attention in specific area to Improve grades/reduce failure and taking appropriate action for the next semester examination. Big data tools can be used in Higher Education to collect relevant and actionable data of learners. Using this educationalist can find out who is weak in what area, how to help weak student,and how to improve result of college. Current data analysis is related to specific course but for orther courses of all levels this analysis can be implemented.

REFERENCES:

[1] Hadoop- The definitive Guide,Tom White, O'REILLY, Second Edition
[2] http://arxiv.org/ftp/arxiv/papers/1201/1201.3417.pdf
[3] http://airccse.org/journal/ijdms/papers/5313ijdms04.pdf
[4] http://esjournals.org/journaloftechnology/archive/vol2no2/vol2no2_7.pdf
[5] http://www.ijarcsse.com/docs/papers/Volume_3/8_August2013/V3I8-0105.pdf
[6] http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm
[7] https://www.dezyre.com/article/hadoop-components-and-architecture-big-data-and-hadoop-training/114
[8] https://www.ibm.com/big-data/us/en/
[9] http://ercoppa.github.io/HadoopInternals/HadoopArchitectureOverview.html