CSE3020 – Data Visualization (ELA), Winter Semester 2021-2022

Lab Assignment IA3 – Slot L43-L44

By: Jonathan Rufus Samuel (20BCT0332) Dos: 8.02.2022

Lab Assignment – IA3 Scalar Visualizations – 1

Note on Software used for following Visualizations: (Tableau)

Tableau is a visual analytics platform transforming the way we use data to solve problems—empowering people and organizations to make the most of their data.

It includes:

- Easy to access from different sources.
- No need for any technical or programming knowledge, and Quick response for making a dashboard.
- In terms of connecting and sharing, it has various inbuilt advanced features such as: Collaboration and distribution, highly securable, Multiple data sources connection, Easy importation and exportation of the massive size of data.
- For easy accessibility and analysis, the data file can be downloaded locally on mobile or desktop, multilingual representation of data, real-time exploration of any dataset, etc.

Q) Create a dataset of 20 rows and 10 columns of data associated with any of your interested domain. The dataset shall include data of types: Qualitative and Quantitative (ordinal, nominal, interval and ratio – continuous / discrete).
For the created dataset, perform the following visualization of:

a. Precise Comparison of Two or more categorical data
b. Two or more Continuous Data over a period of time
c. Numerical data across one or more categorical data
d. Relative Proportion of one or more categorical data

Answer: Some Key points to note before visualization process.

1. Dataset Used:

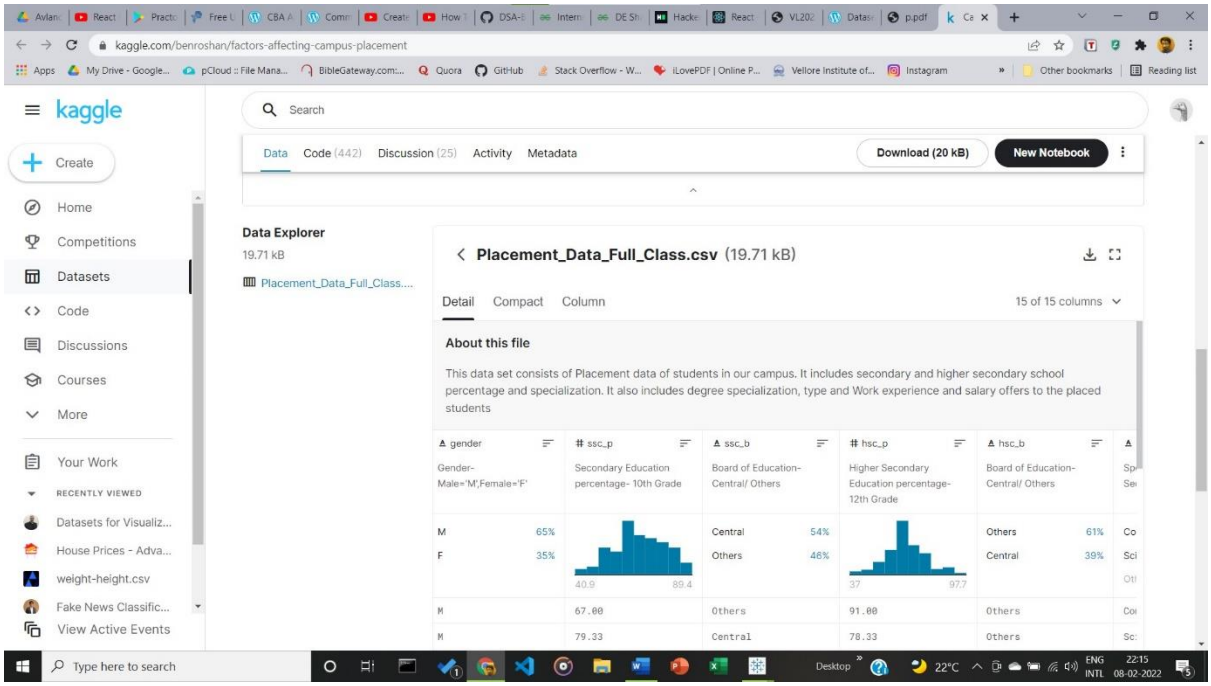
The dataset used is that of the Placement Data at a given College. As mentioned on the parent website Kaggle; *"This data set consists of Placement data of students in a XYZ campus. It includes secondary and higher secondary school percentage and specialization. It also includes degree specialization, type and Work experience and salary offers to the placed students".* Contains 20 Rows and 10 Columns as required by the question.

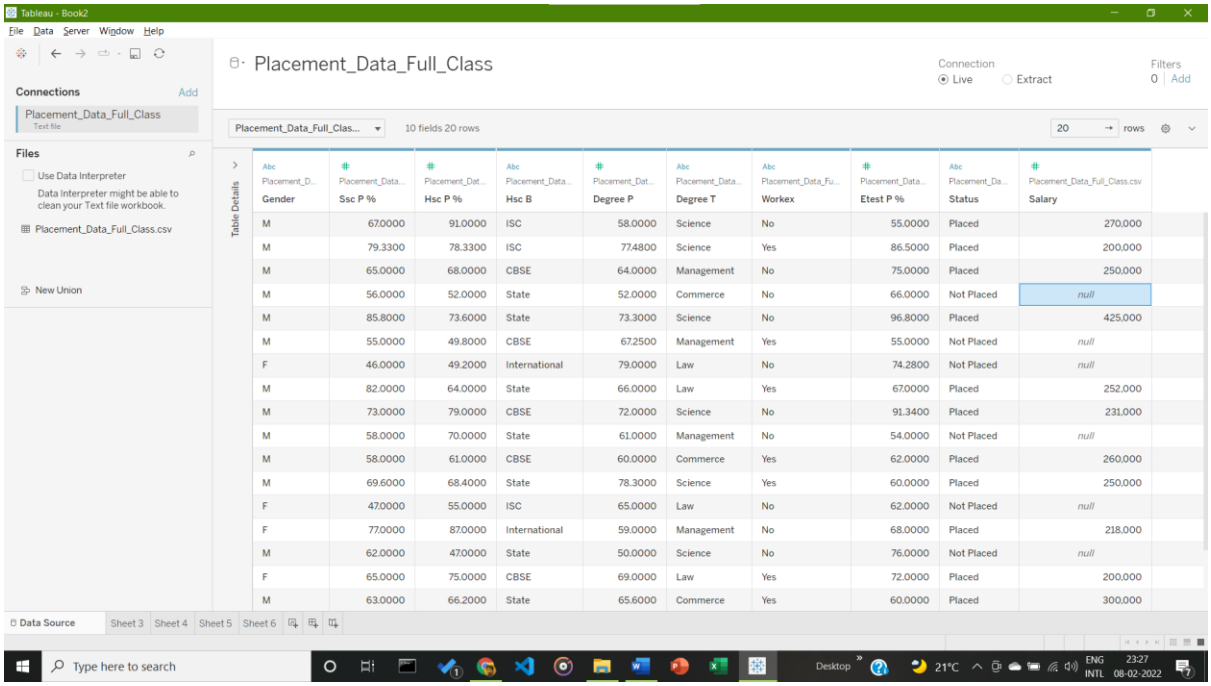
Link(s): https://www.kaggle.com/benroshan/factors-affecting-campus-placement
https://mksaad.wordpress.com/2020/06/30/datasets-for-visualization/

Categories Present in Dataset:

- **Qualitative: (Categorical Data)**
  - Degree T
  - Gender
  - Highschool Board (hsc b)
  - Status
  - Work Experience (Work ex)

- **Quantitative:**
  - Degree Percentage (Degree P)
  - Employability Test (Etest P%)
  - Secondary Education Percentage (ssc p%)
  - Higher Secondary Education Percentage (hsc p%)
  - Salary



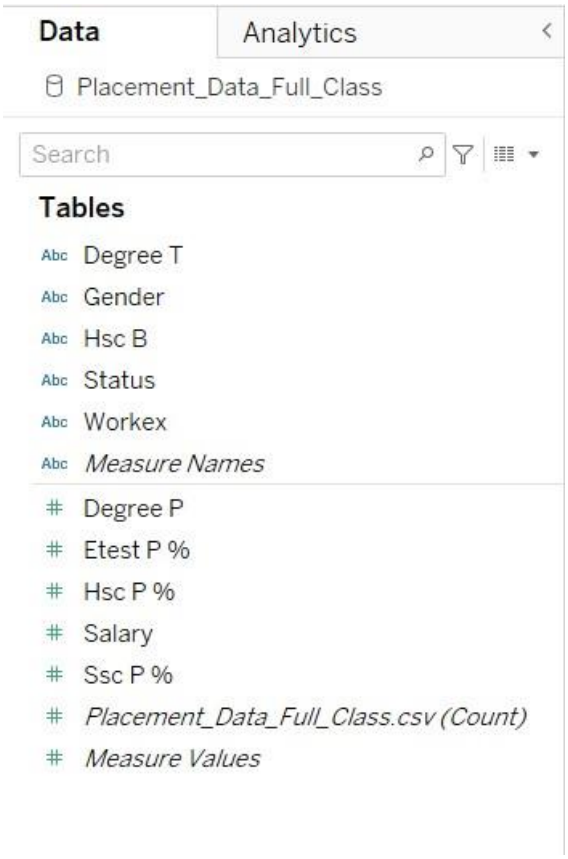


Name

Placement_Data_Full_Class.csv

Fields

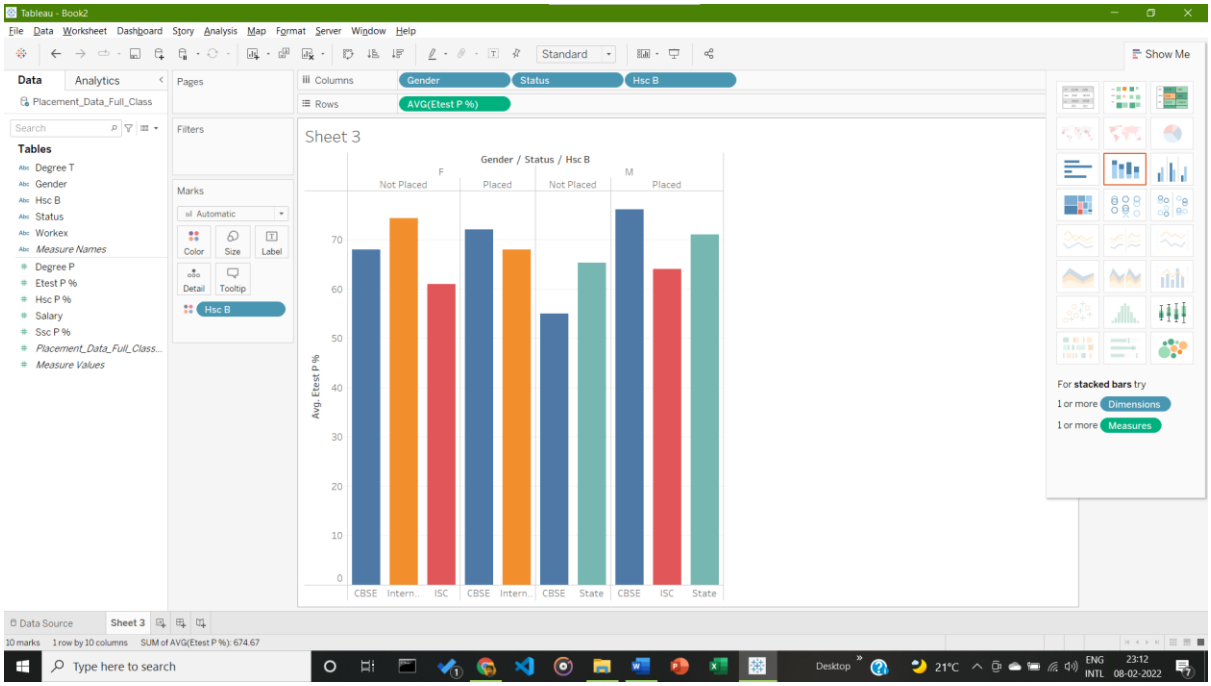
Type	Field Name	Physical Table	Remote Field Name
Abc	Gender	Placement_Data_Full_Class.csv	gender
#	Ssc P %	Placement_Data_Full_Class.csv	ssc_p %
#	Hsc P %	Placement_Data_Full_Class.csv	hsc_p %
Abc	Hsc B	Placement_Data_Full_Class.csv	hsc_b
#	Degree P	Placement_Data_Full_Class.csv	degree_p
Abc	Degree T	Placement_Data_Full_Class.csv	degree_t
Abc	Workex	Placement_Data_Full_Class.csv	workex
#	Etest P %	Placement_Data_Full_Class.csv	etest_p %
Abc	Status	Placement_Data_Full_Class.csv	status
#	Salary	Placement_Data_Full_Class.csv	salary



2. Visualization:

a. Precise Comparison of Two or more categorical data

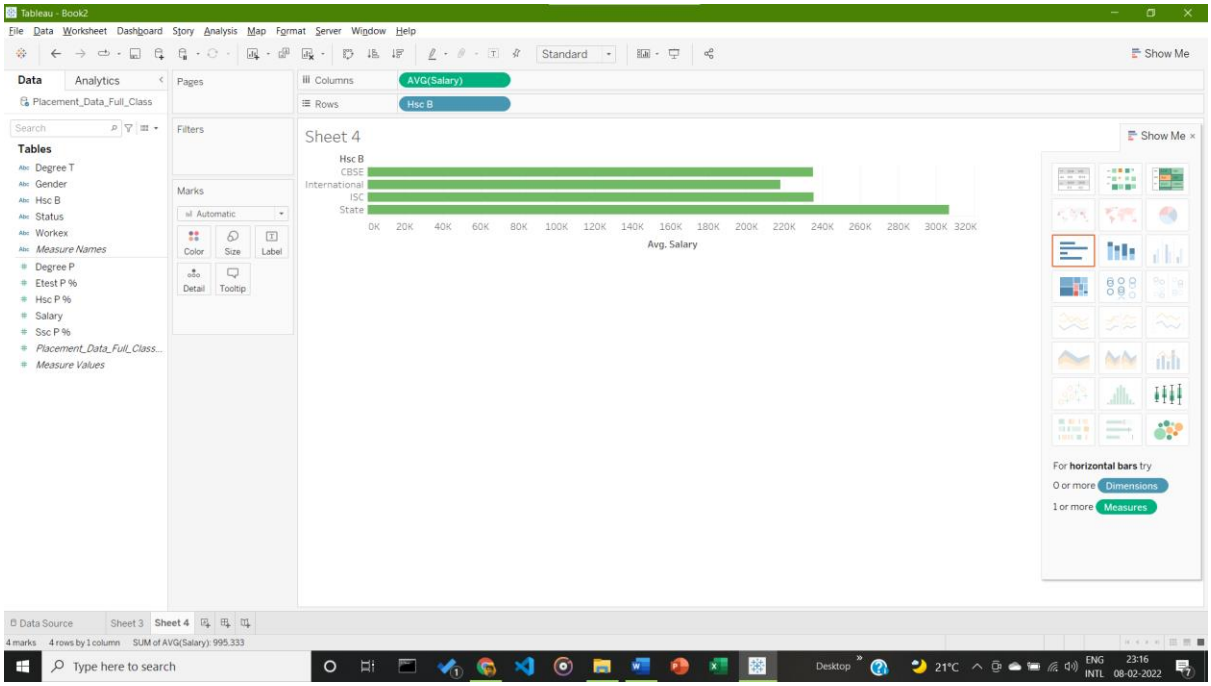
Generalization based on the three following Categorical Data: Status, Gender and Higher Secondary Scores, with data on x axis being Employability Test scores, to give a relation A student's gender, board marks and test scores with overall chance of getting placed.

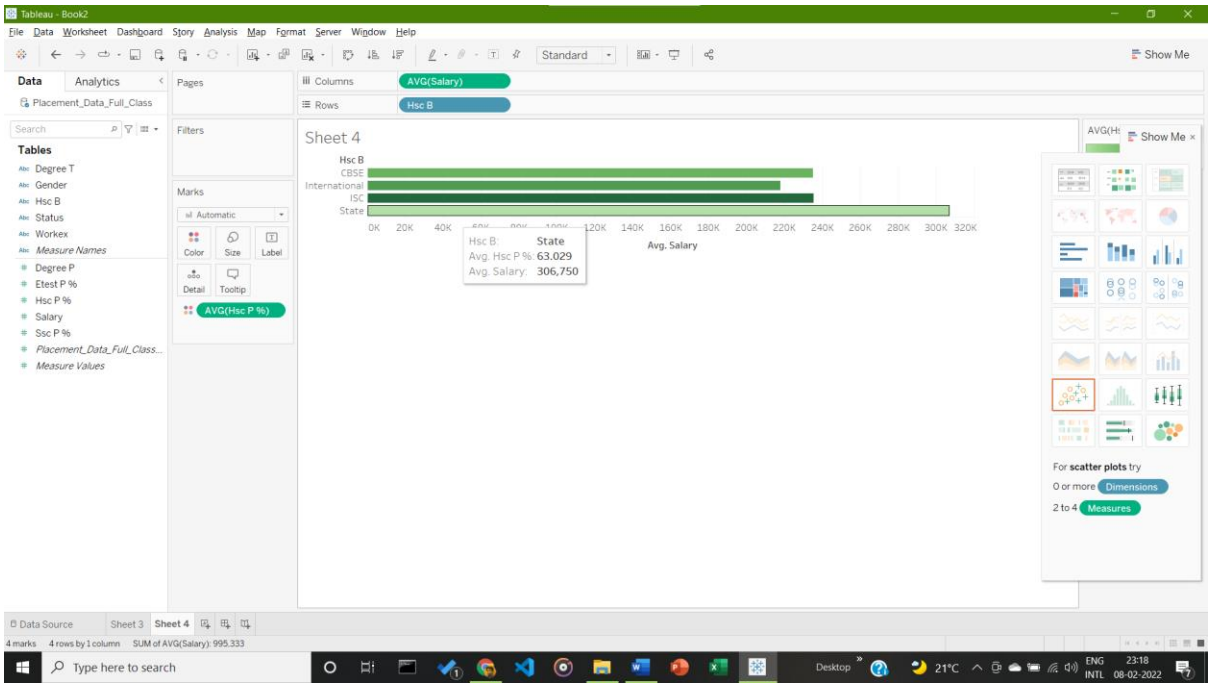


Conclusion from Given Visualization: Highest Summation of Employability in given Dataset goes to Male who have are from CBSE at 76.11%

b. Two or more Continuous Data over a period of time

Generalization based on the two following Continuous Data: Salary and Board (Bo time related fields available)

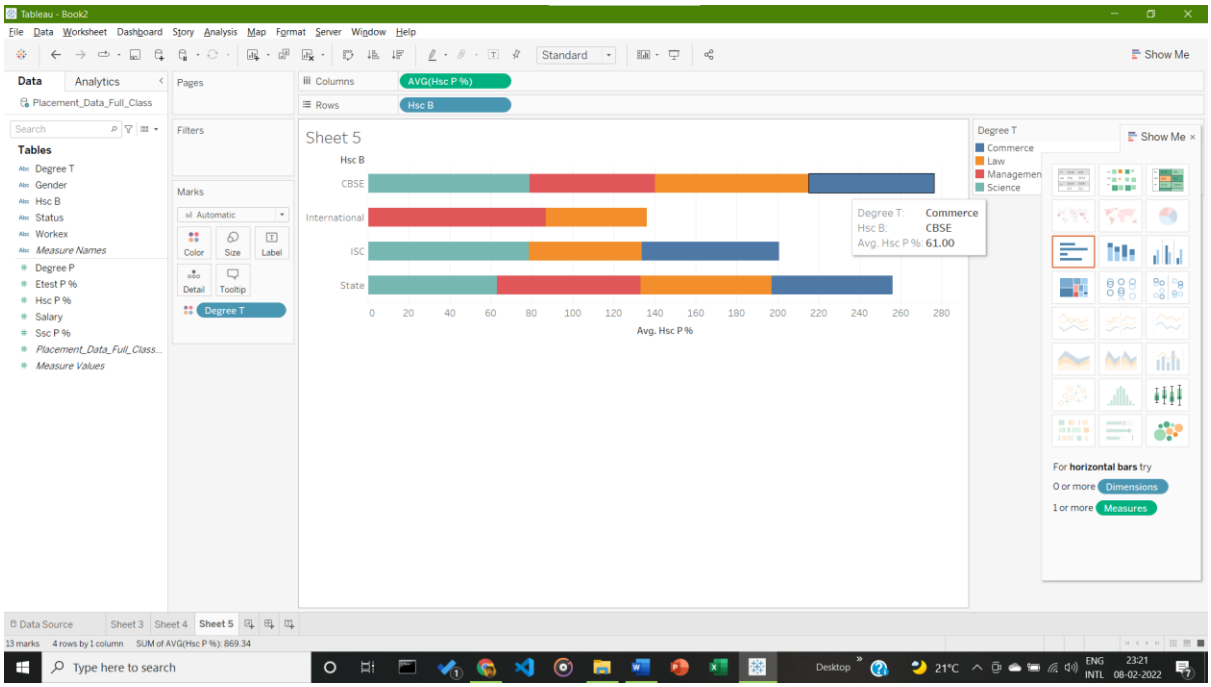




Conclusion from Given Visualization: Highest Average Salary in given Dataset goes to Students from State with an average salary of 306,750.

c. Numerical data across one or more categorical data

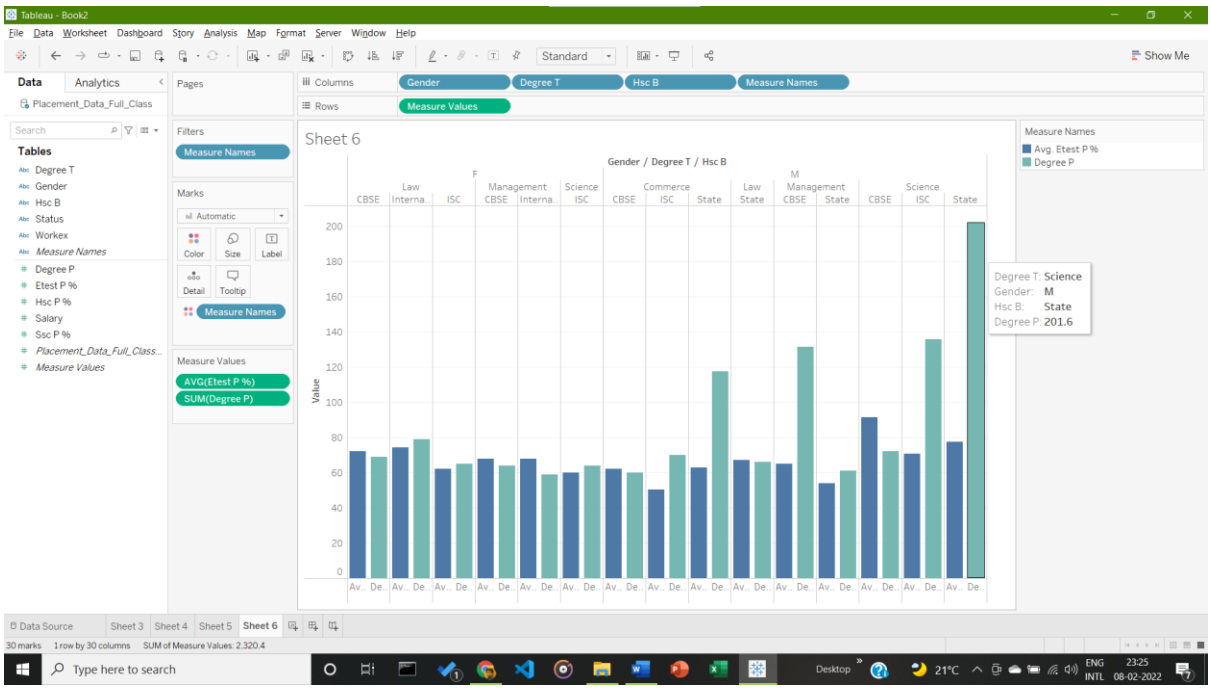
Generalization based on the three following Categorical and Numerical Data: Board, Average High School Board Marks and Field/Degree chosen.



Conclusion from Given Visualization: Highest Average Scores in High School Boards in given Dataset goes to CBSE Students within the Commerce field.

d. Relative Proportion of one or more categorical data

Generalization based on the 6 following Categorical Data: Degree, Gender, Hec B, Measure Names, Degree Scores, and Employability Test scores.



Conclusion from Given Visualization: Highest Cumulative Summation of Degree Test Scores in given Dataset goes to Male who have taken Science as their course and in the CBSE board.
