



**BHARATIYA VIDYA BHAVAN'S**  
**SARDAR PATEL INSTITUTE OF TECHNOLOGY**  
(Empowered Autonomous Institute Affiliated to University of Mumbai)  
[Knowledge is Nectar]

**Department of Computer Engineering**

Name	Ms Neerja Doshi
UID	2021300029
DIV	BE COMPS [ ADV -> BATCH F ]
ADV EXP 10	

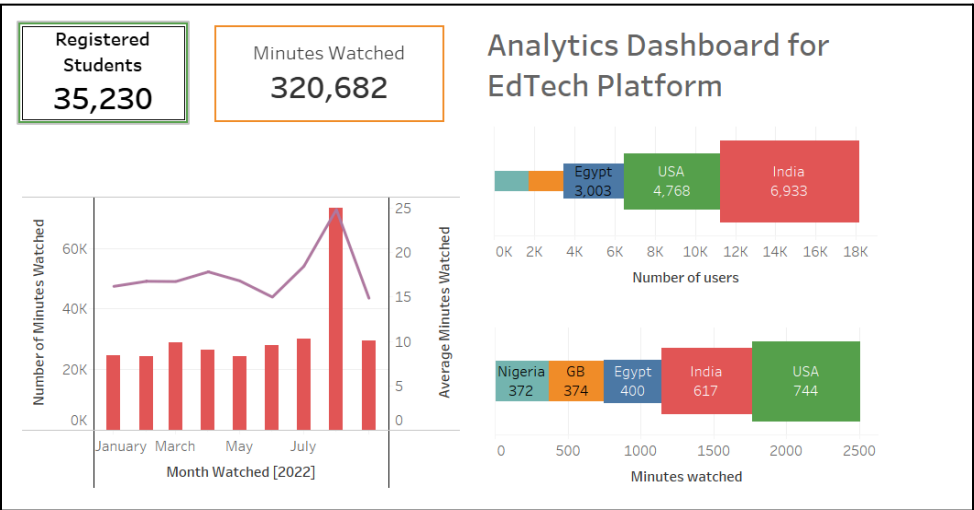
AIM	10. Design Big Data Dashboard using Tableau on the dataset -Education Sector																																								
Dataset Particulars	<b>Name: Edtech Data Analysis : A list of online courses on edx.org learning platform</b>																																								
	Link: <a href="https://www.kaggle.com/code/sagarpratap23/edtech-data-analysis">https://www.kaggle.com/code/sagarpratap23/edtech-data-analysis</a>																																								
	<b>Column Details :</b>																																								
	<table><tr><th>Sr No.</th><th>Field Name</th><th>Description</th></tr><tr><td>1</td><td>serial_number</td><td>Unique identifier for each course</td></tr><tr><td>2</td><td>course_enrolled</td><td>Number of students enrolled</td></tr><tr><td>3</td><td>course_type</td><td>Type of course (e.g., verified, micro)</td></tr><tr><td>4</td><td>institution</td><td>Institution that created the course</td></tr><tr><td>5</td><td>instructors</td><td>Comma-separated names of course instructors</td></tr><tr><td>6</td><td>level</td><td>Difficulty level (e.g., beginner, intermediate, advanced)</td></tr><tr><td>7</td><td>subject</td><td>Subject of the course</td></tr><tr><td>8</td><td>language</td><td>Language of instruction</td></tr><tr><td>9</td><td>subtitles</td><td>Available subtitle languages (comma-separated)</td></tr><tr><td>10</td><td>course_effort</td><td>Estimated time and effort required to complete</td></tr><tr><td>11</td><td>course_length</td><td>Length of the course in weeks</td></tr><tr><td>12</td><td>price</td><td>Course price</td></tr></table>			Sr No.	Field Name	Description	1	serial_number	Unique identifier for each course	2	course_enrolled	Number of students enrolled	3	course_type	Type of course (e.g., verified, micro)	4	institution	Institution that created the course	5	instructors	Comma-separated names of course instructors	6	level	Difficulty level (e.g., beginner, intermediate, advanced)	7	subject	Subject of the course	8	language	Language of instruction	9	subtitles	Available subtitle languages (comma-separated)	10	course_effort	Estimated time and effort required to complete	11	course_length	Length of the course in weeks	12	price
Sr No.	Field Name	Description																																							
1	serial_number	Unique identifier for each course																																							
2	course_enrolled	Number of students enrolled																																							
3	course_type	Type of course (e.g., verified, micro)																																							
4	institution	Institution that created the course																																							
5	instructors	Comma-separated names of course instructors																																							
6	level	Difficulty level (e.g., beginner, intermediate, advanced)																																							
7	subject	Subject of the course																																							
8	language	Language of instruction																																							
9	subtitles	Available subtitle languages (comma-separated)																																							
10	course_effort	Estimated time and effort required to complete																																							
11	course_length	Length of the course in weeks																																							
12	price	Course price																																							

13	course_description	Description of the course content
14	course_syllabus	Link to the course syllabus
15	course_url	URL of the course on edx.org

- 1) Here is a tabular representation of the courses we are studying from the dataset
- 2) The dataset has the following 5 courses:

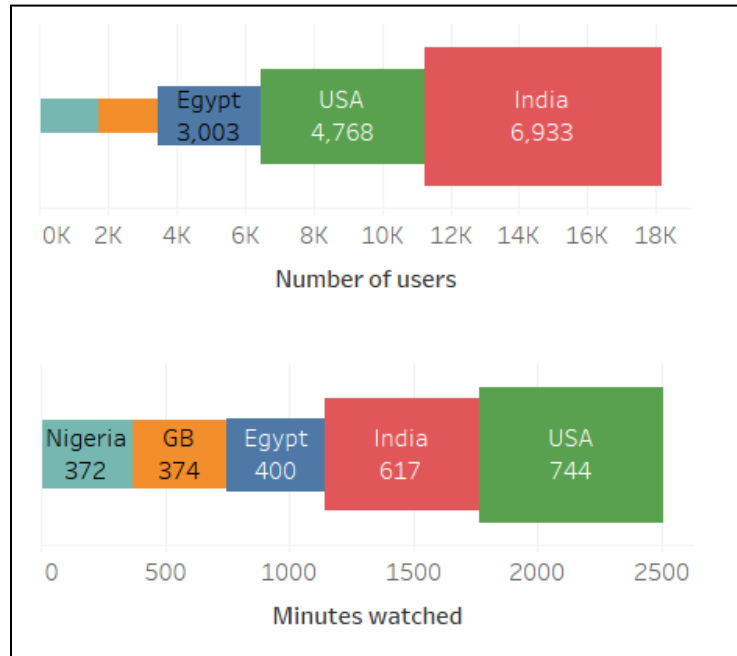
Course Name				
Course Title	Minutes Watched:	Average Minutes Watched:	Average Rating:	Number of Ratings:
Introduction to Data and Data Science	333,266	22	5	572
Introduction to Excel	164,472	27	5	242
Python Programmer Bootcamp	155,164	34	5	101
SQL	234,825	36	5	125
Statistics	193,280	28	5	215

Dashboard



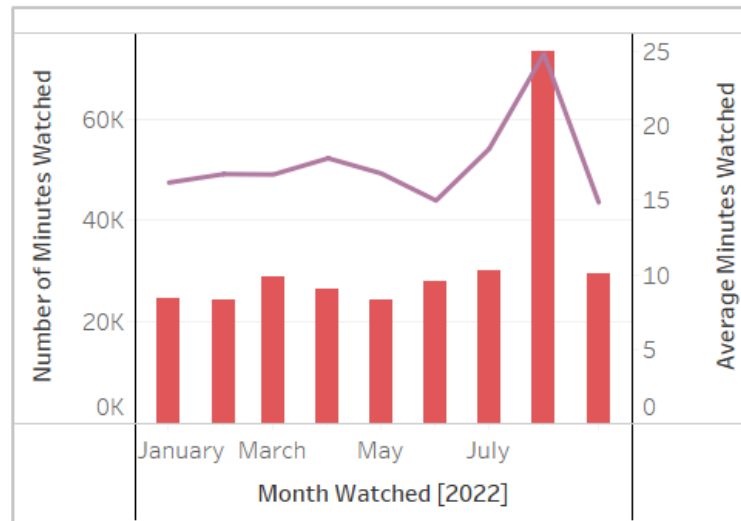
## Analysis

### 1) Clustered Bar graph



- 1) The dataset we have is predominantly talking about course related to the field of computers.
- 2) As per dataset, India has the most number of users followed by USA
- 3) The explanation to this is :
  - a) High population inclined towards learning computer related subjects
  - b) Availability of Internet connectivity
- 4) Egypt, Nigeria are countries having low GDP but are developing
- 5) Hence, we can say , students there are trying to adapt to computer related literacy

## 2) Histogram with density plot :



- 1) As the most number of users + highest number of minutes watched is both in India and USA; the graph is greatly affected by these two countries
- 2) Exams in USA of Spring and Summer Semesters occur in the month of June and July
- 3) In India, June is the start of most schools
- 4) Hence this explains the peak in of **“Most Number of minutes watched”** for the month of July

### Conclusion

By performing this experiment I learnt hands on on Tableau software. I've understood how education subscription data helps in understanding education, interests levels, concentration amongst students, and how this varies with countries.