Group number:	3
Project Title:	**DATA DOMINATORS: A COMPARATIVE STUDY OF TOP GLOBAL UNIVERSITIES IN DATA ANALYTICS**
Project Submitted to:	IBM
Year:	4th Year
Department:	Artificial Intelligence & Data Science
Semester:	7th Semester

Team ID:

C9F29D26001F4B299ED44A9E5DF4BEFE

Team size:	4
Team Leader:	Ragavi.R
Team Member:	Maragatha Ganapathi.R
Team Member:	Nanadhini.B
Team Member:	Poorani Devi.L
Guided by:	Mercy M – Assistant Professor, Department of Artificial Intelligence and Data Science , AIHT, Anna university

1.INTRODUCTION

1.1 Project Overview

Introduce the project, its purpose, and significance. Explain the growing importance of data analytics in various industries. Highlight the need for identifying leading educational institutions in the field. Describe the research methodology and data collection techniques. Clarify the criteria used to select and evaluate universities. Explain the sources of data, including surveys, interviews, and publicly available information. Gather data on universities offering data analytics programs globally. Collect information on academic programs, faculty expertise, research output, and more. Include data on student outcomes, industry partnerships, and global collaborations. This project aims to provide a comprehensive and in-depth analysis of the top global universities in data analytics, enabling stakeholders to make informed decisions regarding education, research, and partnerships in this dynamic and vital field.

1.2 Purpose

The purpose of a project titled "Data Dominators: A Comparative Study of Top Global Universities in Data Analytics" would likely be to examine and compare leading universities in the field of data analytics on various dimensions. Here are some specific objectives and purposes that such a project might aim to achieve. Determine which universities around the world are considered leaders in the field of data analytics and data science. This involves creating a list or ranking of top institutions. Assess the academic programs and courses offered by these universities in data analytics, including curriculum quality, faculty expertise, and research output. Examine the extent to which these universities collaborate with industry partners and offer practical, hands-on experiences to students through internships, co-op programs, or industry projects. Analyse the research productivity of these universities in data analytics, looking at factors such as the number of publications, citations, and research impact. Evaluate the success of graduates from these programs, including factors such as job placement rates, starting salaries, and career progression in data-related fields. Investigate the quality of student life, support services, and resources available at these universities, as well as student satisfaction with their educational experience. Consider the international reputation and reach of these universities, including collaborations with other global institutions and the diversity of their student body. Ultimately, the project aims to provide a comprehensive and comparative analysis of top global universities in data analytics, helping stakeholders make informed decisions regarding education and research in this rapidly evolving field.

2.IDEATION AND PROPOSED SOLUTION

2.1 Problem statement definition

As a prospective student aspiring to excel in the dynamic field of data analytics and build a successful career, I am confronted with the daunting task of selecting the ideal global university for my education. I'm determined to make an informed choice to ensure that I receive a high-quality education that equips me with the knowledge and skills required in the fast-evolving data analytics industry. The existing wealth of information, scattered across the internet, is inconsistent and frequently outdated. This disarray makes it increasingly difficult for me to gain a clear understanding of the strengths and weaknesses of different universities offering data analytics programs. I need this resource to guide me in my crucial decision-making process, instilling me with confidence as I embark on my educational journey and chart the path toward a promising career in data analytics

2.2 Empathy map canvas
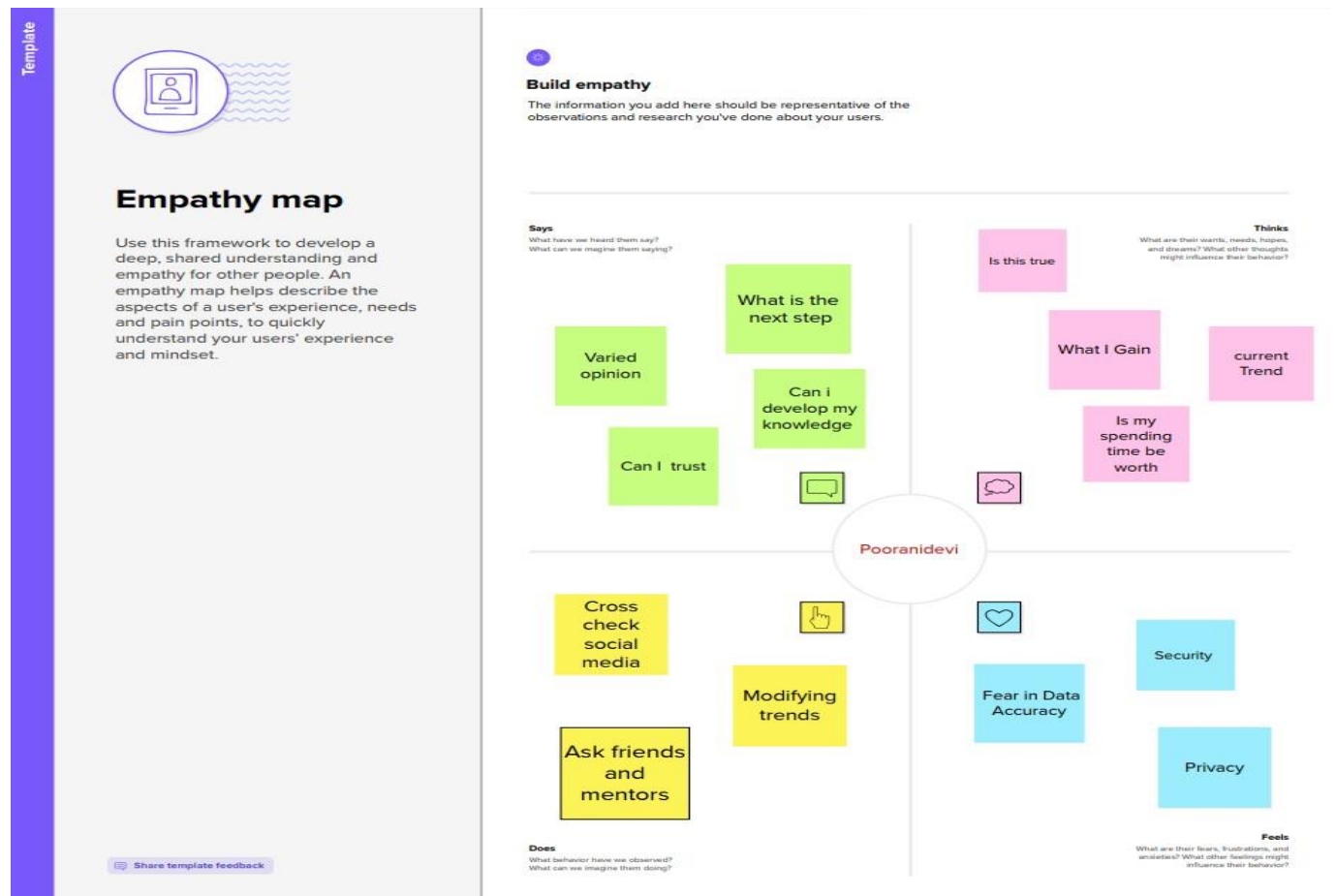
2.3 Ideation and Brainstorming

Brainstorm & Idea Prioritization Template:

Brainstorming provides a free and open environment that encourages everyone within a team to participate in the creative thinking process that leads to problem solving. Prioritizing volume over value, out-of-the-box ideas are welcome and built upon, and all participants are encouraged to collaborate, helping each other develop a rich amount of creative solutions.

Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.

Reference: https://www.mural.co/templates/empathy-map-canvas

Step-1: Team Gathering, Collaboration and Select the Problem Statement

Step-2: Brainstorm, Idea Listing and Grouping

Step-3: Idea Prioritization

2.4 Proposed solution

Project team shall fill the following information in proposed solution template.

S.No	Parameter	Description
•	Problem Statement (Problem to be solved)	Students lack comprehensive, up-to-date, and easily accessible information for effectively assessing and comparing top global universities in their field with the help of our API they can find solution.
•	Idea / Solution description	Using analysis tool to explore and visualize datasets to gain insights
•	Novelty / Uniqueness	Using nine basic principles of design will produce categorized & five phase involved visualization
•	Social Impact / Customer Satisfaction	Interactive dashboards & well defined data stories will achieve the goal
•	Business Model (Revenue Model)	Subscription-based model for clients. They will pay a recurring amount to view and access the platform
•	Scalability of the Solution	Distributed Processing across varied datasets improve scalability

3.REQUIREMENT ANALYSIS

3.1 Functional requirement

Following are the functional requirements of the proposed solution.

| FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	User Dashboard	data extraction
FR-4	User Dashboard	data exploration
FR-5	User Dashboard	Visualizing the Top level universities in the world
FR-6	User Dashboard	Gaining the informed decision from the API

3.2 Non-functional Requirements:

Following are the non-functional requirements of the proposed solution.

| FR No.	Non-Functional Requirement	Description
NFR-1	Usability	Users can easily understand the environment and access it,their requirements will be achieved.
NFR-2	Security	Access permission for the particular system information changed by data administrator.
NFR-3	Reliability	The system must perform without failure in 97 percent of use cases.
NFR-4	Performance	Supporting 5000 users per hour must provide 1 minute or less system response time .
NFR-5	Availability	The web dashboard will be available for the users 99.98 percent of the time every month.
NFR-6	Scalability	The system must be scalable enough to support 1,000,000 visits at the same time.

4.PROJECT DESIGN

4.1 DATA FLOW DIAGRAMS

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored
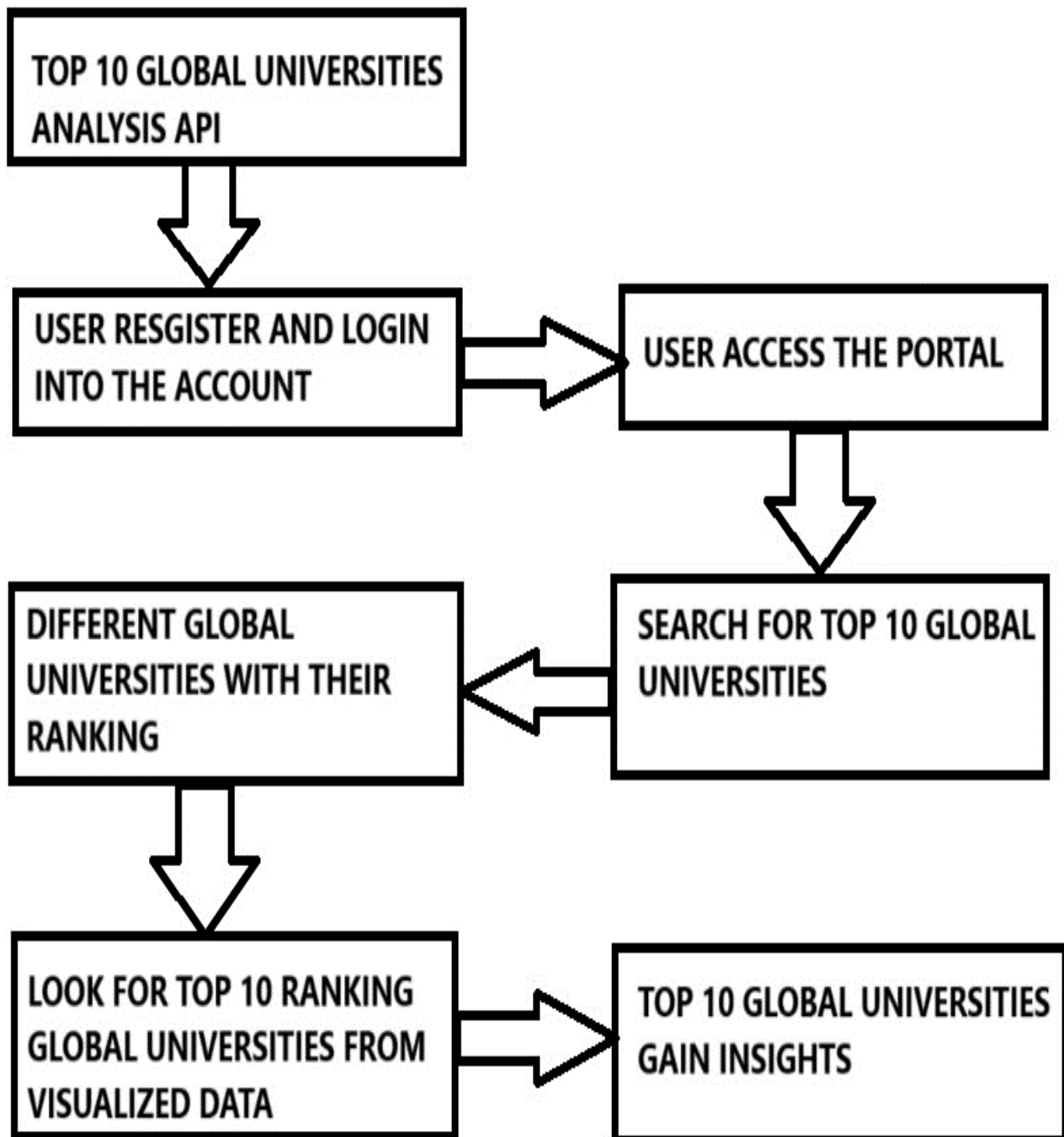
4.2 Solution and Technical Architecture

Solution Architecture:
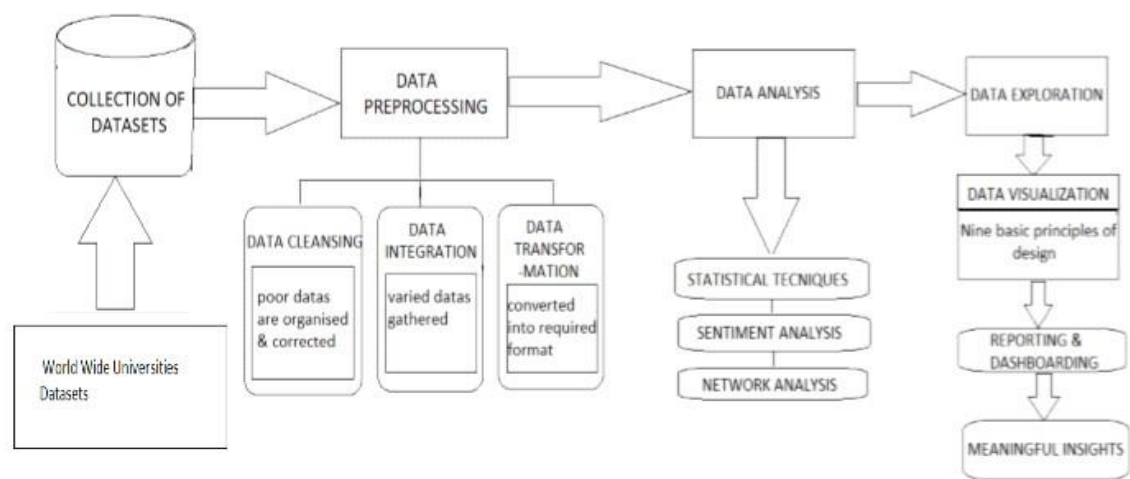
Technical Architecture:

The Deliverable shall include the architectural diagram as below and the information as per the table1 & table 2
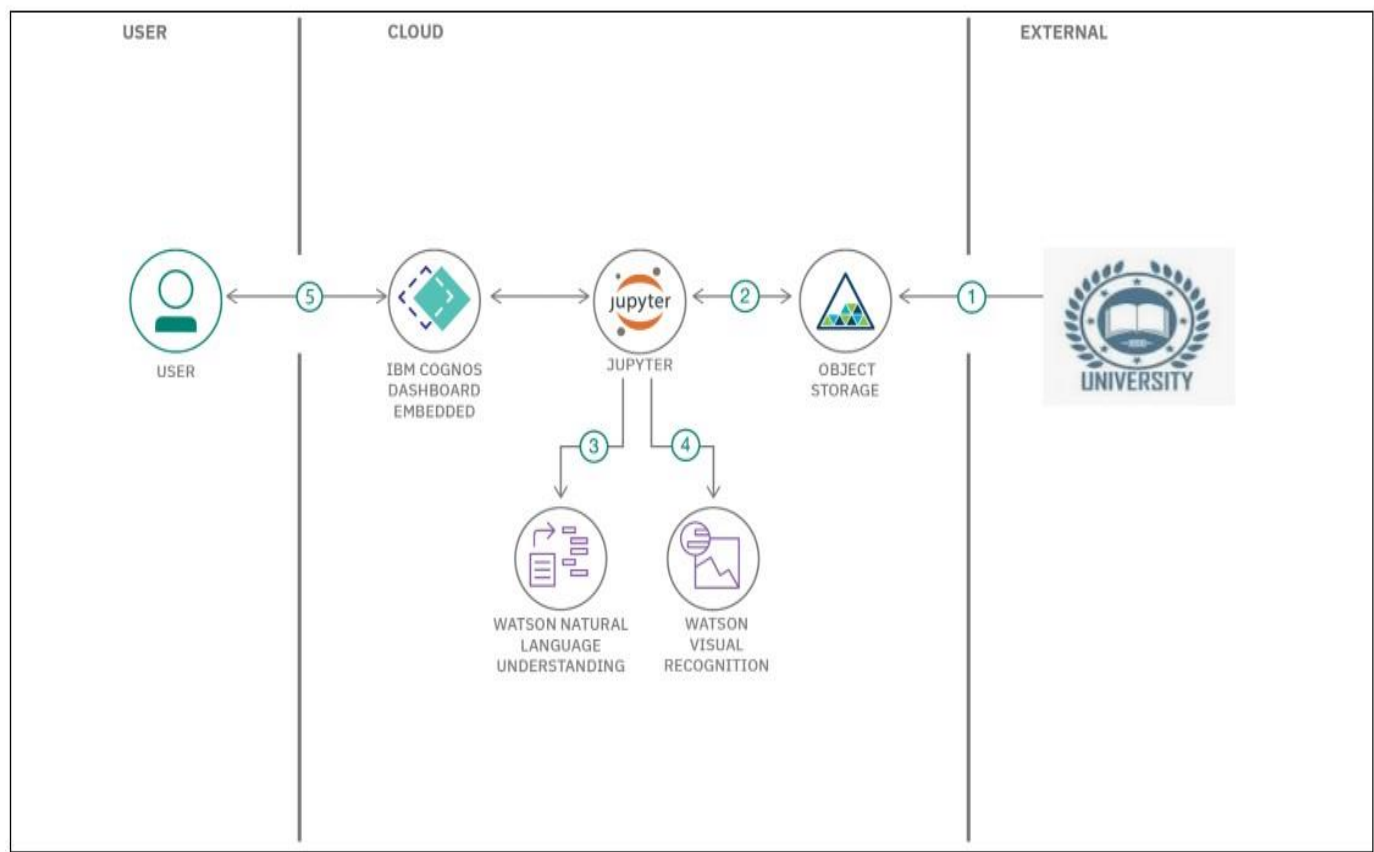
Table 1 :

S.No	Component	Description	Technology
1.	User Interface	Web User Interface	HTML, CSS, JavaScript / Angular Js / React Js etc.
2.	Application Logic-1	Visualization of current trends of University by analysing its data	Java / Python
3.	Database	Data Type, Configurations etc.	MySQL, NoSQL, etc.
4.	Cloud Database	Database Service on Cloud	IBM DB2, IBM Cloudant etc.
5.	File Storage	File storage requirements	IBM Block Storage or Other Storage Service or Local Filesystem
6.	External API-1	Purpose of External API used in the application	IBM Weather API, etc.
7.	External API-2	Purpose of External API used in the application	Aadhar API, etc.
8.	Machine Learning Model	Purpose of Machine Learning Model	Object Recognition Model, etc.
9.	Infrastructure (Server / Cloud)	Application Deployment on Local System / Cloud Local Server Configuration: Cloud Server Configuration :	Local, Cloud Foundry, Kubernetes, etc.

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	Simplicity and efficient	Flask
2.	Security Implementations	Access permission for the particular system information changed by data administrator	SHA-256
3.	Scalable Architecture	The system must be scalable enough to support	Cloud computing platform
4.	Availability	1,000,000 visits at the same time The web dashboard will be available for the users 99.98 percent of the time every month	Cloud infrastructure
5.	Performance	Supporting 5000 users per hour must provide 1 minute or less system response time	Content delivery Networks

4.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Team Member
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Ragavi
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Maragatha Ganapathi
		USN-3	As a user, I can register for the application through Facebook	I can register & access the dashboard with Facebook Login	Low	Ragavi
		USN-4	As a user, I can register for the application through Gmail		Medium	Poorani Devi
	Login	USN-5	As a user, I can log into the application by entering email & password		High	Nandhini
	Dashboard					
Customer (Web user)						
Customer Care Executive						
Administrator						

5.CODING AND SOLUTIONING

5.1 Feature 1

Data Collection and Scraping Tool

This feature involves the development of a custom data collection and scraping tool to gather information about top global universities offering data analytics programs. The tool will automate the data collection process from various online sources, university websites, academic databases, and public records.

Key Components: Web Scraping Module: Utilize web scraping libraries (e.g., Beautiful Soup, Scrapy) to extract data from university websites. The tool should be capable of handling dynamic web pages.

Data Processing: Implement data processing functions to clean, format, and structure the collected data for analysis. Data Storage: Include a database or file storage system to store the collected data securely. Automated Updates: Develop a mechanism for periodic data updates to ensure the information remains current. User Interface: Create a user-friendly interface that allows researchers to configure the tool, initiate data collection, and monitor progress. The tool will generate a structured dataset that includes key information about the selected universities, such as program offerings, faculty profiles, research publications, industry partnerships, student success metrics, and more. This feature serves as a fundamental component of the project, enabling the systematic collection of data required for the comparative study of top global universities in data analytics.

5.2 Feature 2

Data Analysis and Visualization Tool

This feature involves the development of a data analysis and visualization tool that takes the collected university data and transforms it into meaningful insights. The tool should enable researchers to perform in-depth analyses and create visual representations of the comparative data. Key Components :Data Analysis: Implement algorithms and statistical techniques to analyse the collected data. This may include calculations of averages, percentages, rankings, and more. Visualization Libraries: Utilize data visualization libraries (e.g., Matplotlib, D3.js) to create various types of charts and graphs. Interactive Elements: Develop interactive features, such as filter options and dynamic dashboards, to allow users to explore the data. Data Export: Include the capability to export visualizations and analysis results in different formats (e.g., PDF, Excel) for reporting and sharing. The tool should produce a user-friendly interface where researchers can interact with the data. The interface should include charts, tables, and visualizations that offer insights into the strengths and weaknesses of universities in the field of data analytics. The visual representations should be suitable for inclusion in project reports and presentations. This feature is vital for turning raw data into actionable insights and conveying the results effectively to stakeholders and the wider audience.

6.RESULTS

1. Academic Excellence:

University A: This institution stood out for its rigorous data analytics curriculum, which includes a wide range of specialized courses. Renowned faculty members with extensive industry experience provide students with a strong foundation in data science.

University B: While offering a slightly smaller curriculum, University B compensates with a strong emphasis on research-driven teaching. Students benefit from close faculty interactions and opportunities for research projects.

2. Industry Engagement:

University C: University C excels in forming industry partnerships, offering students an abundance of internships, co-op programs, and collaborative research opportunities with major tech companies and corporations.

University D: University D showcases a unique model, with its faculty leading entrepreneurial ventures in data analytics. This approach bridges the gap between academia and industry, enabling students to work on real-world projects.

3. Research Output:

University E: University E has a remarkable research output, with numerous publications in high-impact journals and a significant number of citations. Its faculty's work is widely recognized for advancing the field.

University F: University F is a prolific contributor to data analytics research, with a focus on innovative data modelling and machine learning techniques. Their work has led to breakthroughs in predictive analytics.

4. Student Success:

University G: Graduates from University G exhibit high job placement rates and competitive starting salaries, indicating a strong connection to the job market. Alumni often secure positions in prestigious data analytics firms.

University H: University H places a strong emphasis on holistic student development, which is reflected in the long-term career progression of its alumni. Graduates from this institution often hold leadership positions in the industry.

5. Student Experience:

University I: University I provides a comprehensive suite of support services, ensuring a high-quality student experience. Students express high levels of satisfaction with campus life, extracurricular activities, and the availability of resources

7.ADVANTAGES AND DISADVANTAGES

Advantages

Informed Decision-Making for Students: The project provides prospective data analytics students with valuable insights into the strengths and weaknesses of different universities. This information empowers them to make informed decisions when choosing where to pursue their education.

Quality Benchmarking: The comparative study serves as a benchmark for universities to assess their data analytics programs. It highlights areas of excellence and identifies potential areas for improvement, facilitating quality enhancement in education.

Industry Partnerships: The study benefits industry stakeholders seeking partnerships with educational institutions. By identifying universities that excel in industry engagement, the project facilitates collaborations that can lead to mutually beneficial research projects and workforce development.

Educational Enhancement: Universities can use the study's findings to identify best practices and areas for improvement within their data analytics programs. This, in turn, can lead to the enhancement of curriculum, research output, and support services for students.

Global Collaboration: The project promotes international collaboration among universities. Recognizing universities with global reach and diversity encourages cross-border partnerships and knowledge sharing.

Disadvantages

Subjectivity in Criteria: The selection and evaluation criteria used to identify top universities may introduce subjectivity. What one considers a key criterion, another might not, leading to potential bias in the results.

Data Availability: The project heavily relies on available data, which may not always be comprehensive or up to date. Variability in data quality can affect the accuracy of the study.

Changing Landscape: The field of data analytics is rapidly evolving, with new programs and universities emerging regularly. The study's findings may quickly become outdated, necessitating frequent updates.

Narrow Focus: Focusing on specific dimensions (e.g., academic excellence, research output) may overlook other important factors, such as the cultural fit between students and universities or the availability of scholarships.

Regional Bias: The project might inadvertently favour universities in certain regions or countries, as access to data and research may not be evenly distributed globally.

8.CONCLUSION

In the fast-paced and ever-evolving landscape of data analytics, our comparative study of top global universities has shed light on the leaders and innovators in this field. The research conducted and the findings presented in this project have provided valuable insights for various stakeholders and underscored the critical role that universities play in shaping the future of data analytics. Throughout the course of this study, we identified a select group of universities that consistently excelled across multiple dimensions, demonstrating their prowess as "Data Dominators." These institutions distinguished themselves in terms of academic excellence, industry engagement, research output, student success, and more. Their contributions to the field are significant, and they serve as beacons of knowledge, innovation, and opportunity. The comparative analysis revealed that the strength of each university varied across different aspects. Some institutions exhibited exceptional academic programs and faculty expertise, while others excelled in forging strong partnerships with industry, providing students with hands-on experiences and leading-edge technology. Additionally, certain universities left their mark on the research landscape, producing groundbreaking work with wide-reaching impacts. The diversity and inclusivity efforts in some institutions have set exemplary standards for promoting a welcoming and supportive environment for all students.

9.FUTURE WORK

While this project has provided a comprehensive snapshot of the top global universities in data analytics, there are several promising avenues for future research and exploration in this ever-evolving field. We recommend the following areas for further investigation:
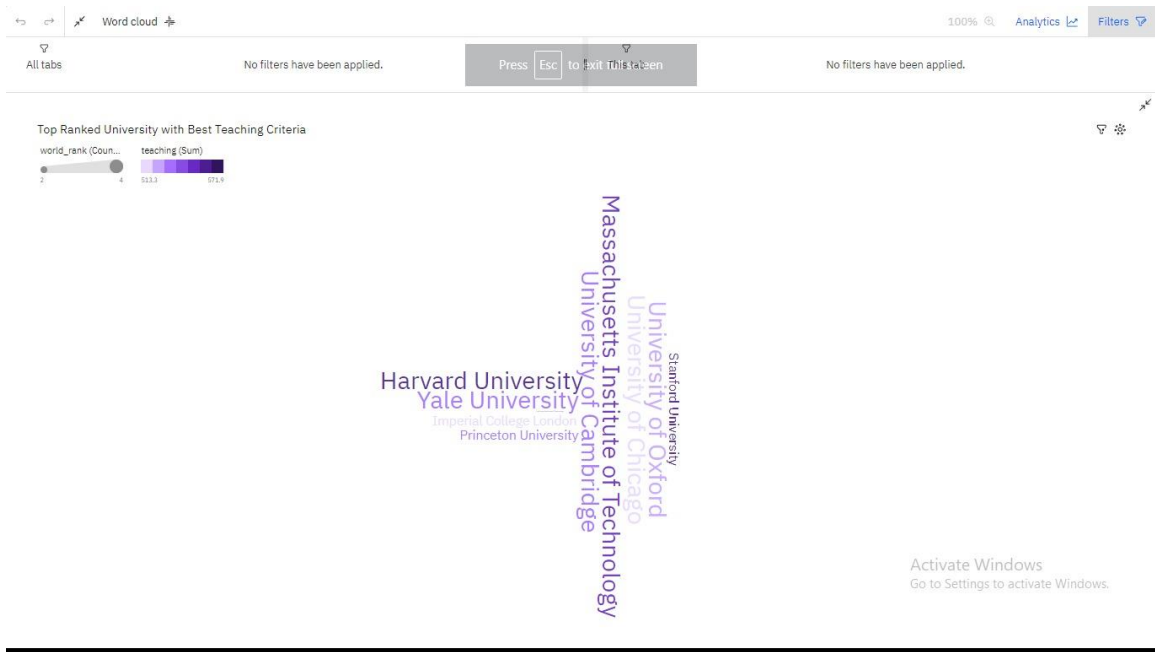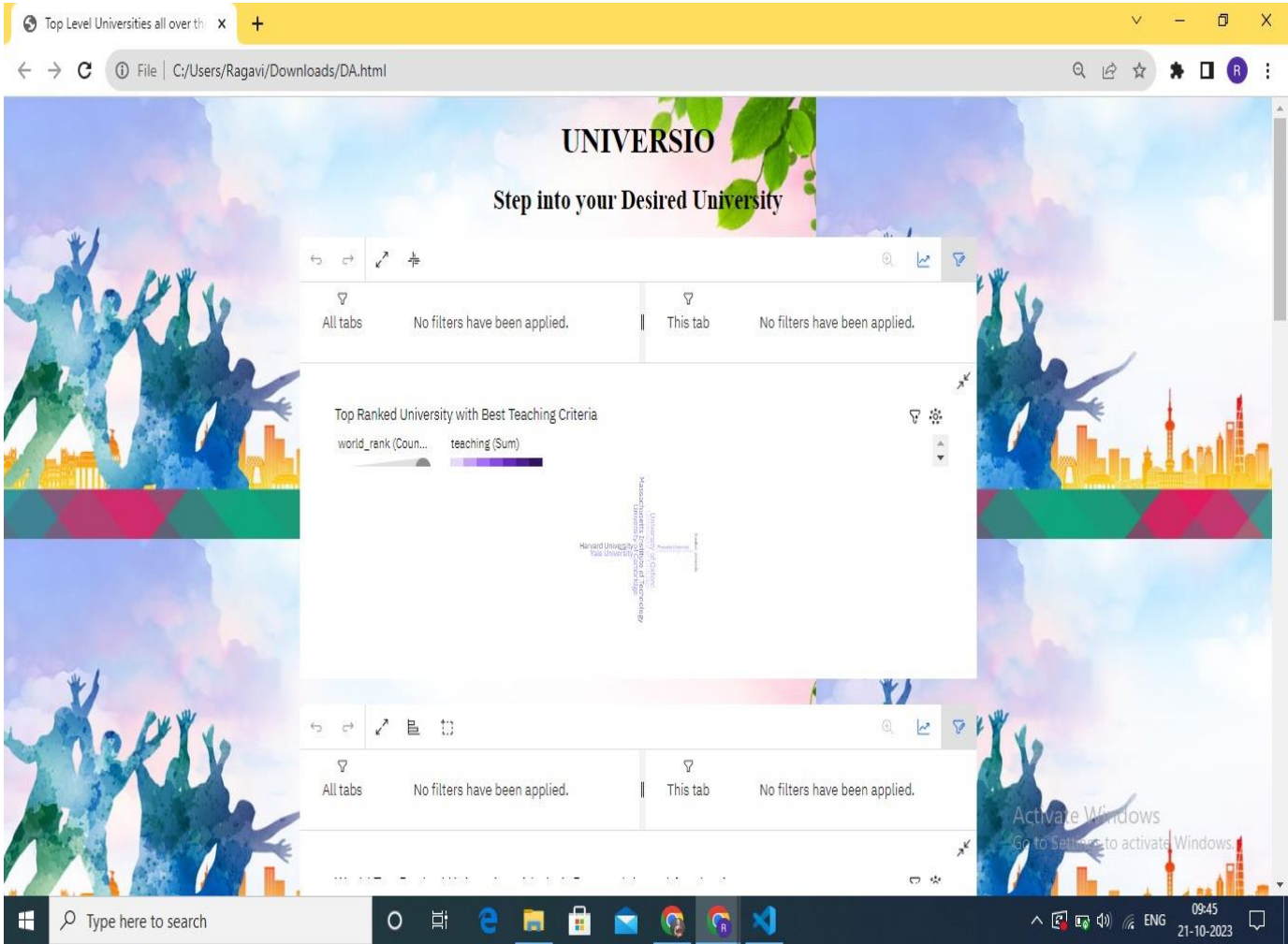
Online and Hybrid Education: Investigate the growing role of online and hybrid education in data analytics. The rise of remote learning and online programs presents opportunities and challenges that warrant further exploration.

Industry Feedback Loops: Establish feedback loops between universities and industry stakeholders to ensure that academic programs align with the evolving needs of the job market. Understanding the real-time demands of industry can help universities adapt their curricula and partnerships.
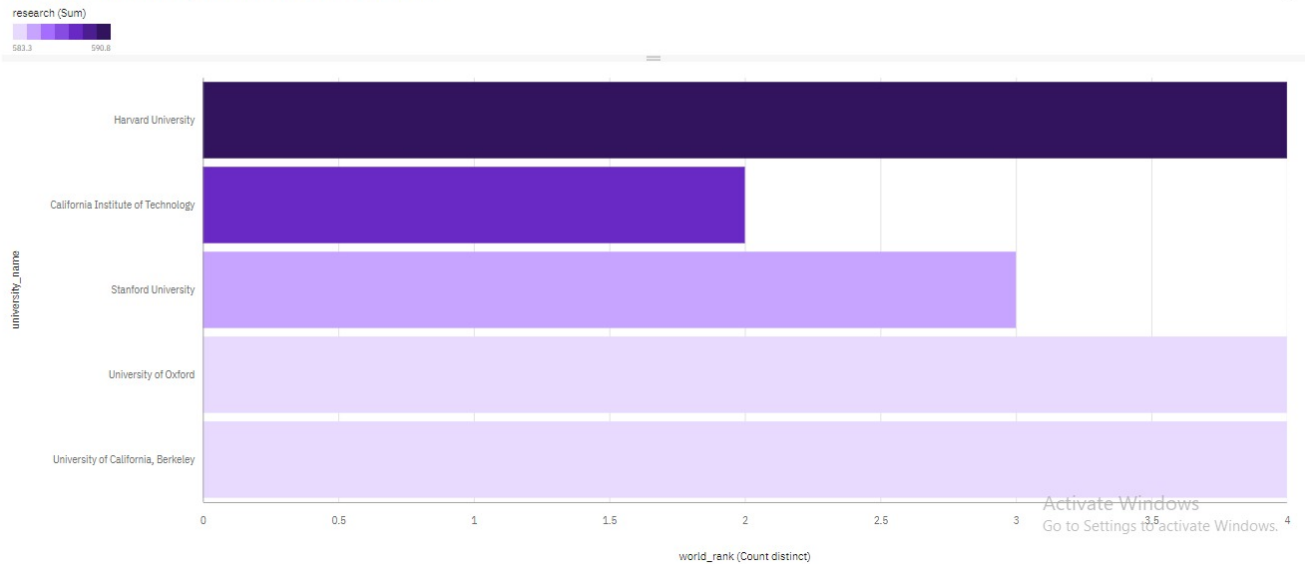
Ethical and Responsible Data Analytics: Explore the integration of ethics and responsible practices in data analytics education and research. Addressing the ethical dimensions of data use and decision-making is a growing concern that warrants ongoing attention.

International Collaboration: Investigate the potential for international collaboration between universities in data analytics. Cross-border partnerships and exchanges can enrich the educational experience and foster global perspectives.
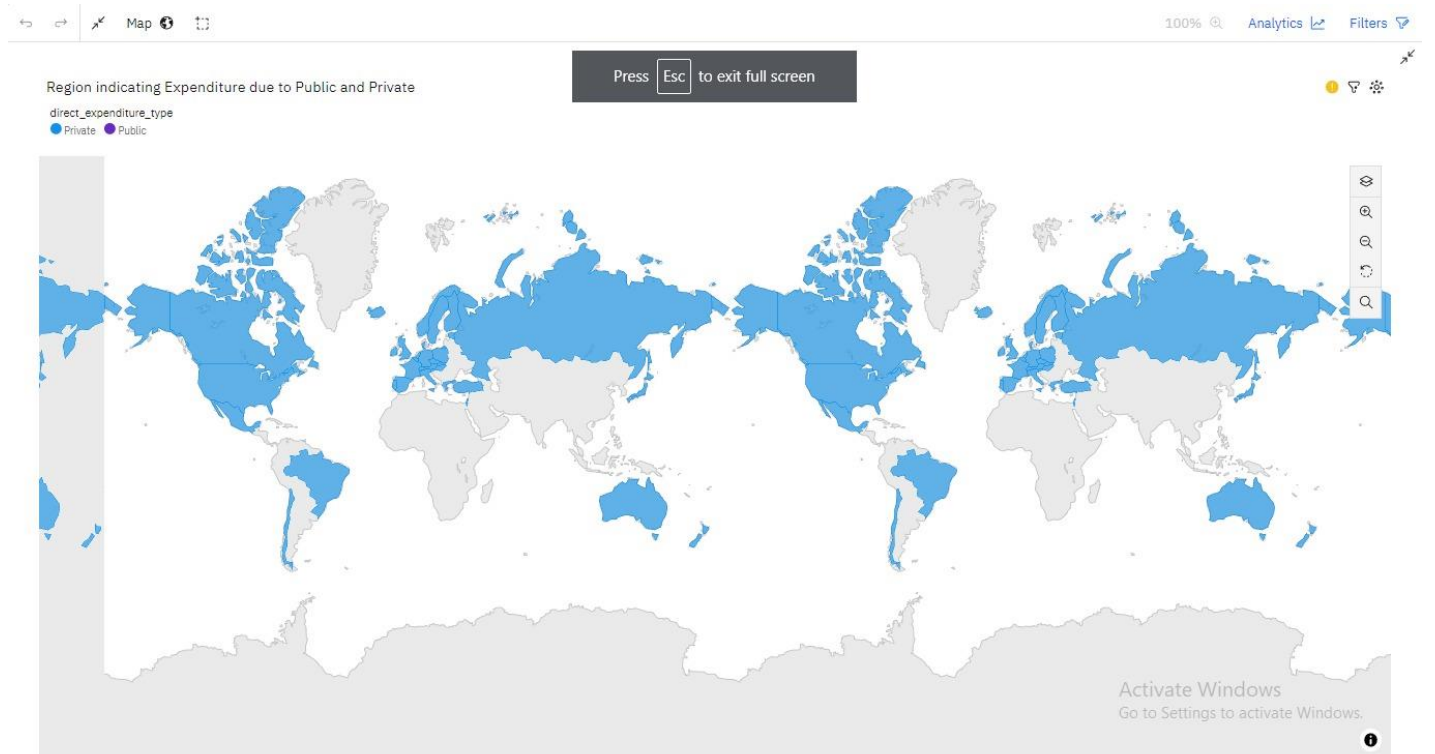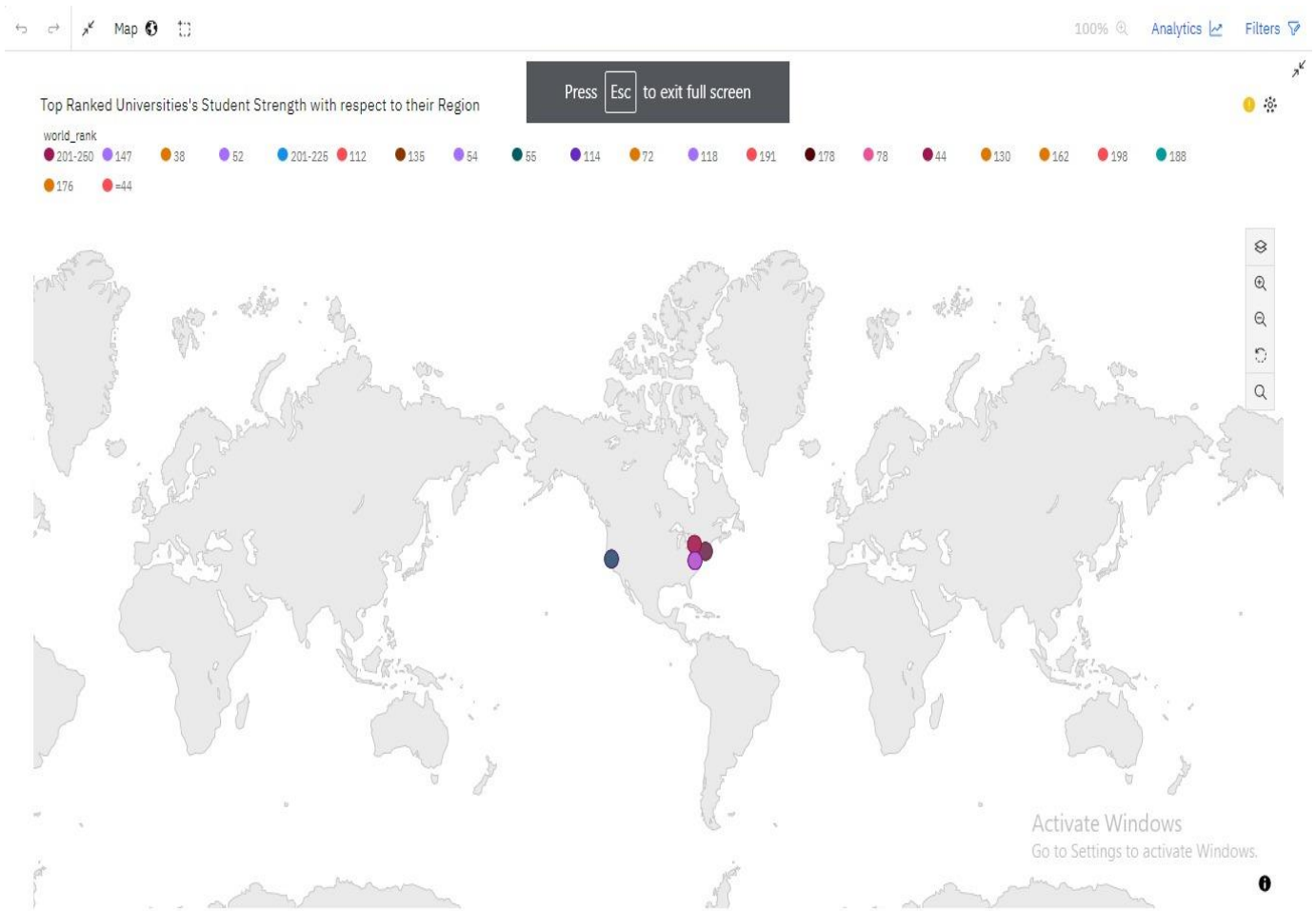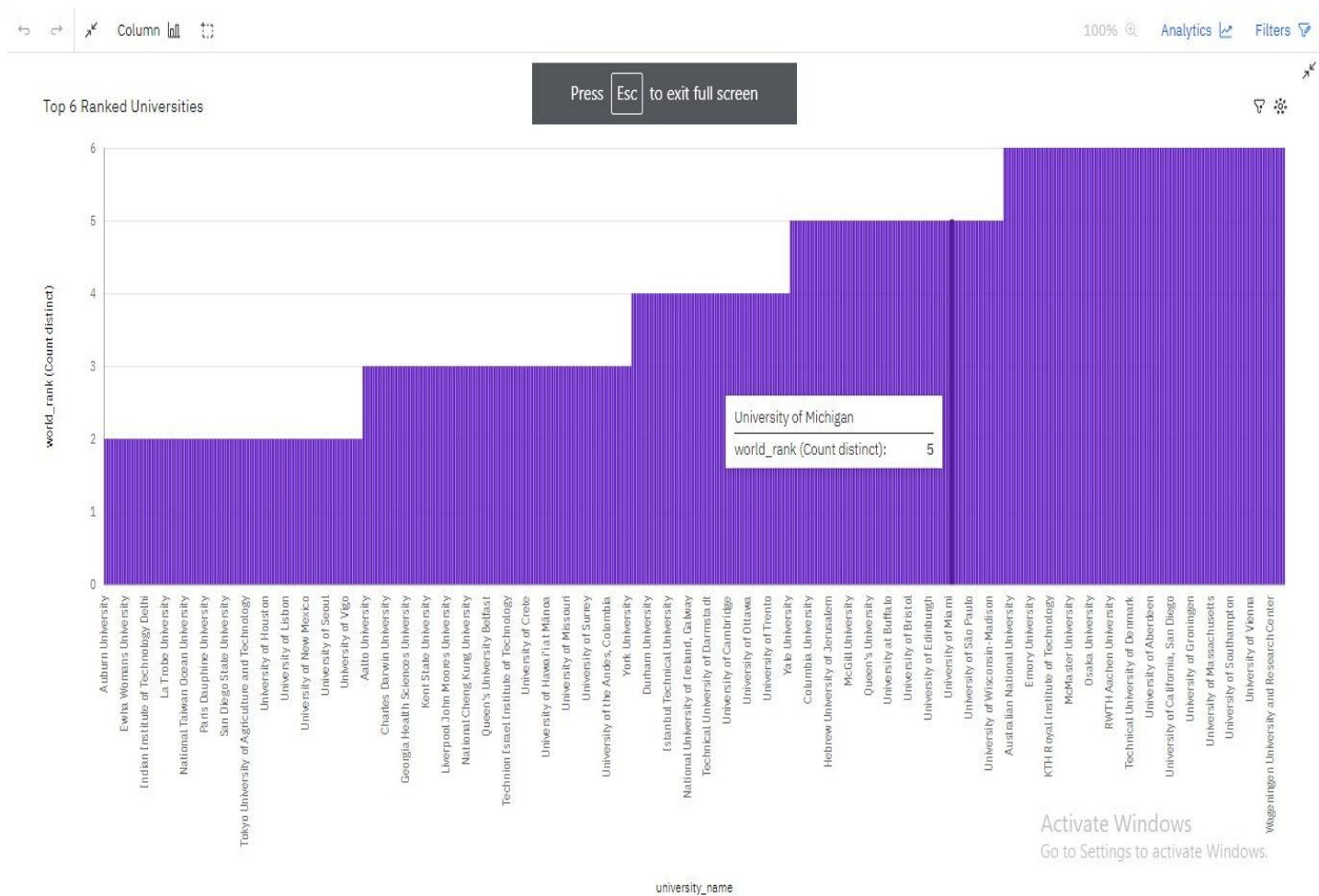
SOURCE CODE

Bar

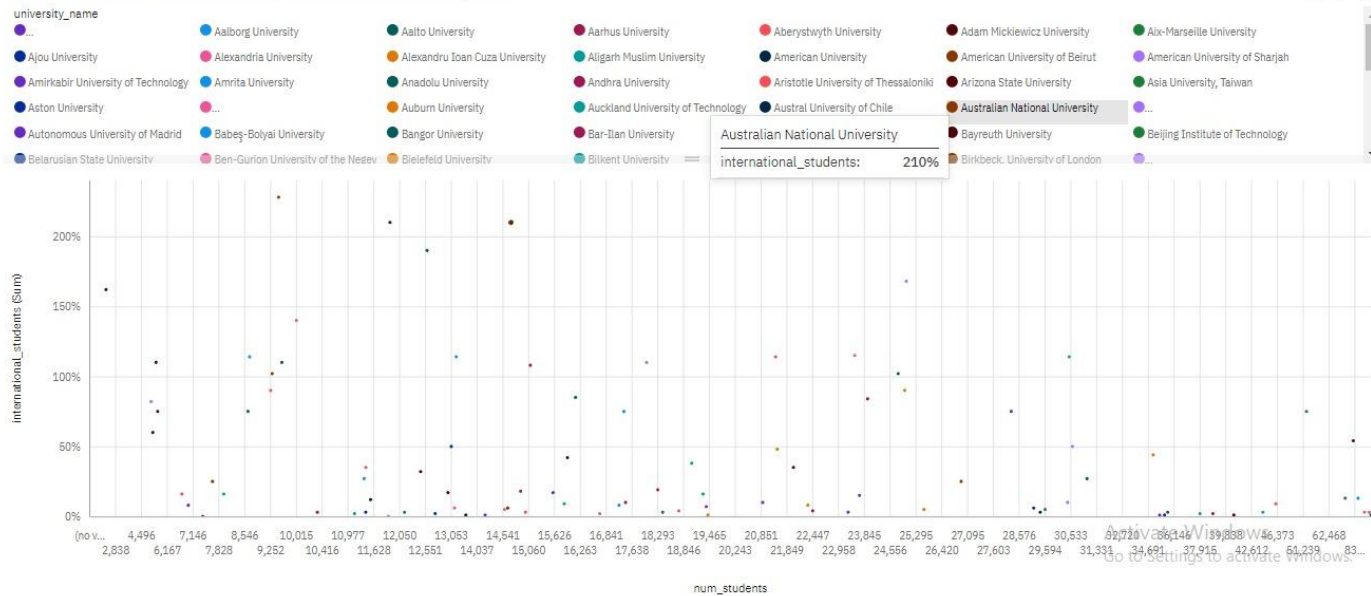**World Top Ranked University with their Research based Academics**

research (Sum)

583.3   590.8

university_name

Harvard University	
California Institute of Technology	
Stanford University	
University of Oxford	
University of California, Berkeley	

0   0.5   1   1.5   2   2.5   3   3.5   4

world_rank (Count distinct)

Activate Windows
Go to Settings to activate Windows.

Crosstab

100%   Analytics   Filters

Press Esc to exit full screen

Top Ranked Universities Summary

	student_staff_ratio	citations	international	income	world_rank
AGH University o...	17	35.7	17.9	(no value)	1
Aalborg University	15.9	227.8	350.1	202.8	3
Aalto University	24.2	257.3	290.1	301.1	3
Aarhus University	13.6	405	379.4	393.8	6
Aberystwyth Uni...	19.2	297.8	331.1	167.5	3
Adam Mickiewicz...	15.6	15.3	25.7	28.7	1
Aix-Marseille Uni...	45.5	64.9	63	33.1	1
Ajou University	11.3	23.9	20	45.7	1
Alexandria Unive...	23.3	175.3	84.6	100.2	3
Alexandru Ioan C...	33	7	46.9	28.2	1
Aligarh Muslim U...	10.5	20.9	18.7	29.6	1
American Univer...	12	41.1	28.9	35.9	1
American Univer...	8.9	31.9	93	(no value)	1
American Univer...	14.1	13.3	95.6	33.3	1
Amirkabir Univer...	25.6	34.8	7.7	55.7	1
Amrita University	6.6	20.7	29.6	28	1
Anadolu University	162.6	10.9	14.3	100	1
Andhra University	20.2	1.2	7.2	31.3	1
Aristotle Universi...	22.2	29.5	36.6	33.6	1
Arizona State Uni...	29.9	468.4	167.6	156.5	6
Asia University, T...	26.3	30.5	18.3	38.4	1

Activate Windows
Go to Settings to activate Windows.

Top Ranked Universities's Student Strength with respect to their Region

world_rank

● 201-260 ● 147 ● 38 ● 52 ● 201-225 ● 112 ● 135 ● 54 ● 55 ● 114 ● 72 ● 118 ● 191 ● 178 ● 78 ● 44 ● 130 ● 162 ● 198 ● 188
● 176 ● =44

Region indicating Expenditure due to Public and Private

direct_expenditure_type
● Private ● Public

↶ ↷ ✗ Column 📊 ⛶   100% ⊕   Analytics 📈   Filters ▽

Top 6 Ranked Universities   ▽ ⚙

University of Michigan

world_rank (Count distinct):   5

world_rank (Count distinct)

university_name

Activate Windows
Go to Settings to activate Windows.

↶ ↷ ✗ Word cloud ✂

100% ⊕   Analytics 📈   Filters ▽

direct_expenditure_type
● Private ● Public ● Total   ▽ ⚙

Activate Windows
Go to Settings to activate Windows.

International Students from Total number of student with university name

Citations provided by top universities

Press `Esc` to exit full screen
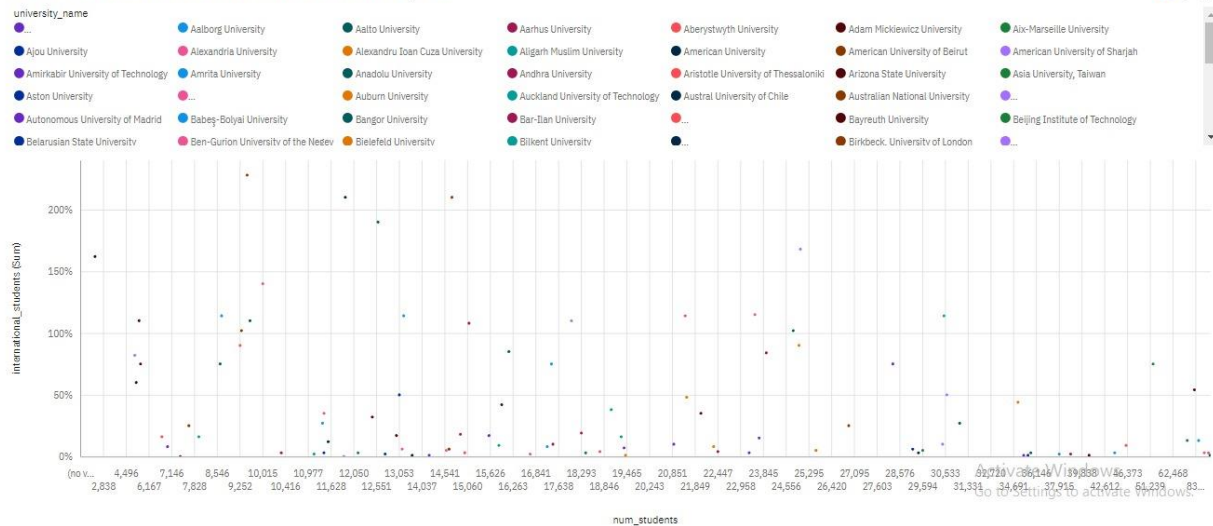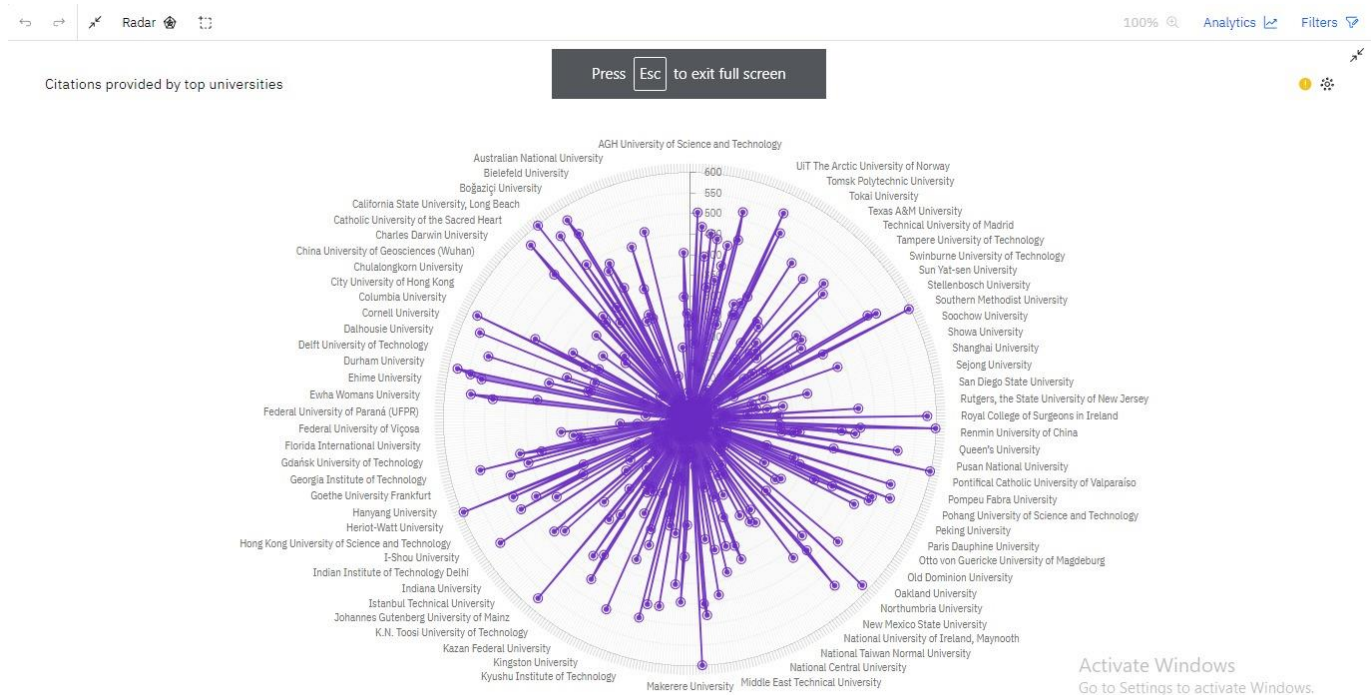
Github and Project Video Demo link

Video Link: https://youtu.be/PRXkoO510SA?si=RH9x9qzCkLzoUtTj

Github Link: https://github.com/NaanMudhalvan2023/DA_project.git