



Prepared by Nikshita Ranganathan, Archit Barua
Shyamala Venkatakrishnan, Heejae Roh
Professor Venkata Duvvuri

June 22th, 2023

3D HEALS®

Project Draft



Executive Summary

01

04

Business Problem

Building Customer email lookalikes

02

05

Analytics & Visuals

Exploratory Data Analysis
& Predictive Model

03

06

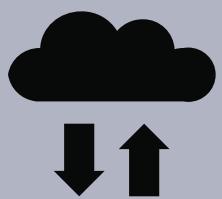
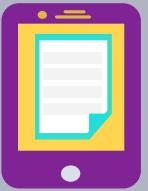
Clear Concise Flow

Analysis and synthesis
of the data

Recommendations & Findings
& Future Research

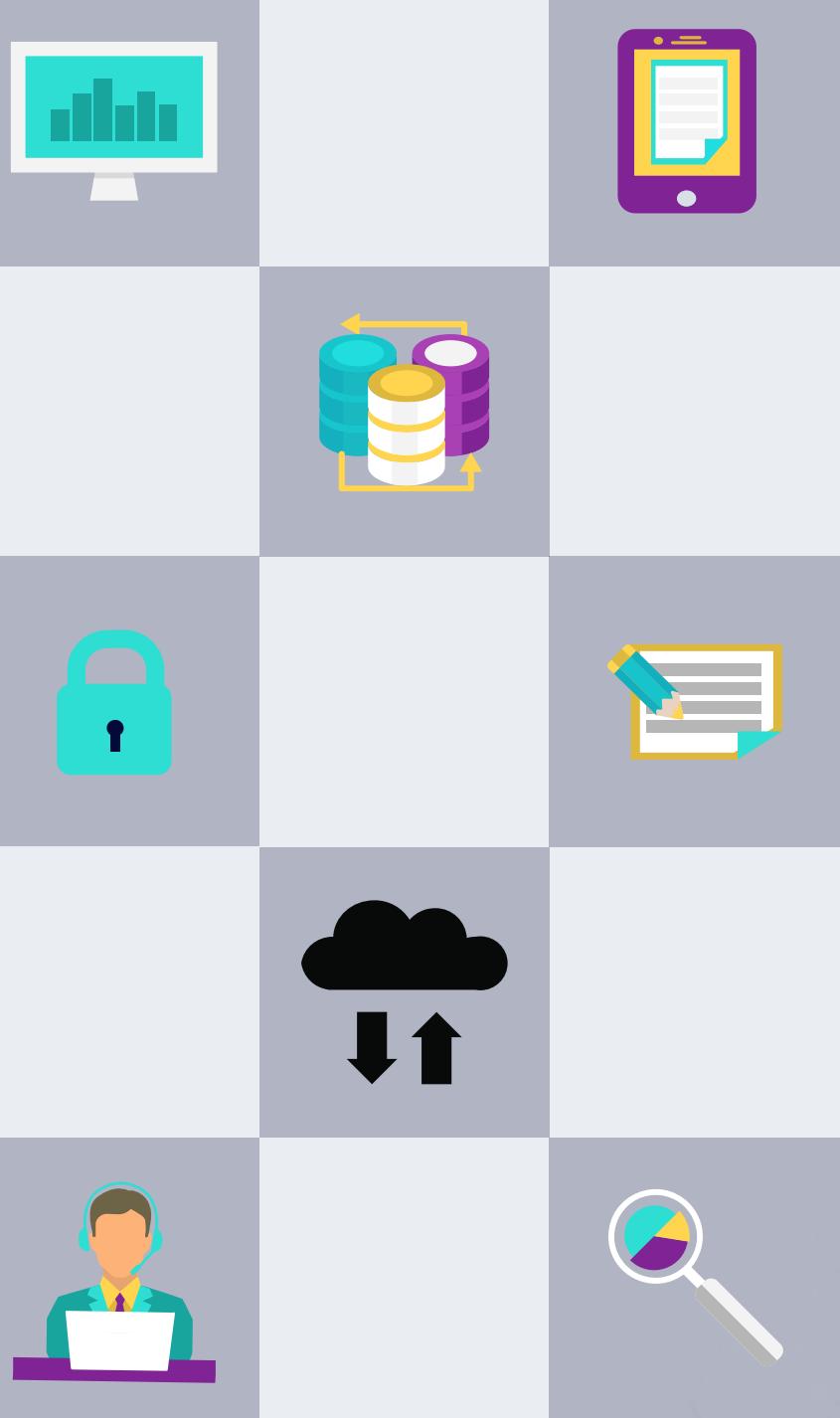
Executive Summary

01



Executive Summary

- Our project aims to address the business objectives of our client through advanced analytics and machine learning techniques.
- We employ Python for data cleansing, manipulation, and model creation, Excel for initial data exploration, and Tableau for data visualization.
- Key milestones include project setup, data acquisition and exploration, model development, applying models to address business objectives, visualization of analysis results, impact analysis, and quality assurance.
- Each team member has specific assignments aligned with their skills, including data analysis, model building, business research, and project management.
- Risks such as data quality, model inaccuracy, overfitting, team coordination, changing business objectives, and security breaches are identified and addressed through mitigation strategies.
- Success metrics include model performance, alignment with business objectives, user acceptance, value created, timely delivery, and risk mitigation.
- The proof of concept will be presented through a structured approach, incorporating clear communication, compelling visualizations, live demonstrations, and interactive sessions.
- The executive summary highlights the project's focus, approach, milestones, team assignments, risk mitigation strategies, success metrics and the presentation method for the proof of concept.
- Overall, our project aims to deliver actionable insights and recommendations that align with the client's business objectives and contribute to their strategic goals.



Business Problem

02

SWOT & Building Customer email lookalikes

SWOT Analysis of 3DHEALS

Strengths

- Domain expertise: Specialized knowledge in 3D printing and bioprinting.
- Innovation and technology adoption: Staying at the forefront of advancements like AI.
- Customization and personalization: Tailored solutions for individual patients.
- Strong network and partnerships: Relationships with hospitals, research institutions, and manufacturers.

Weaknesses

- Limited market awareness: Challenges in increasing visibility and expanding customer base.
- High production costs: Expensive technologies affecting affordability in certain markets.
- Dependence on skilled workforce: Difficulty scaling operations due to specialized expertise.

Opportunities

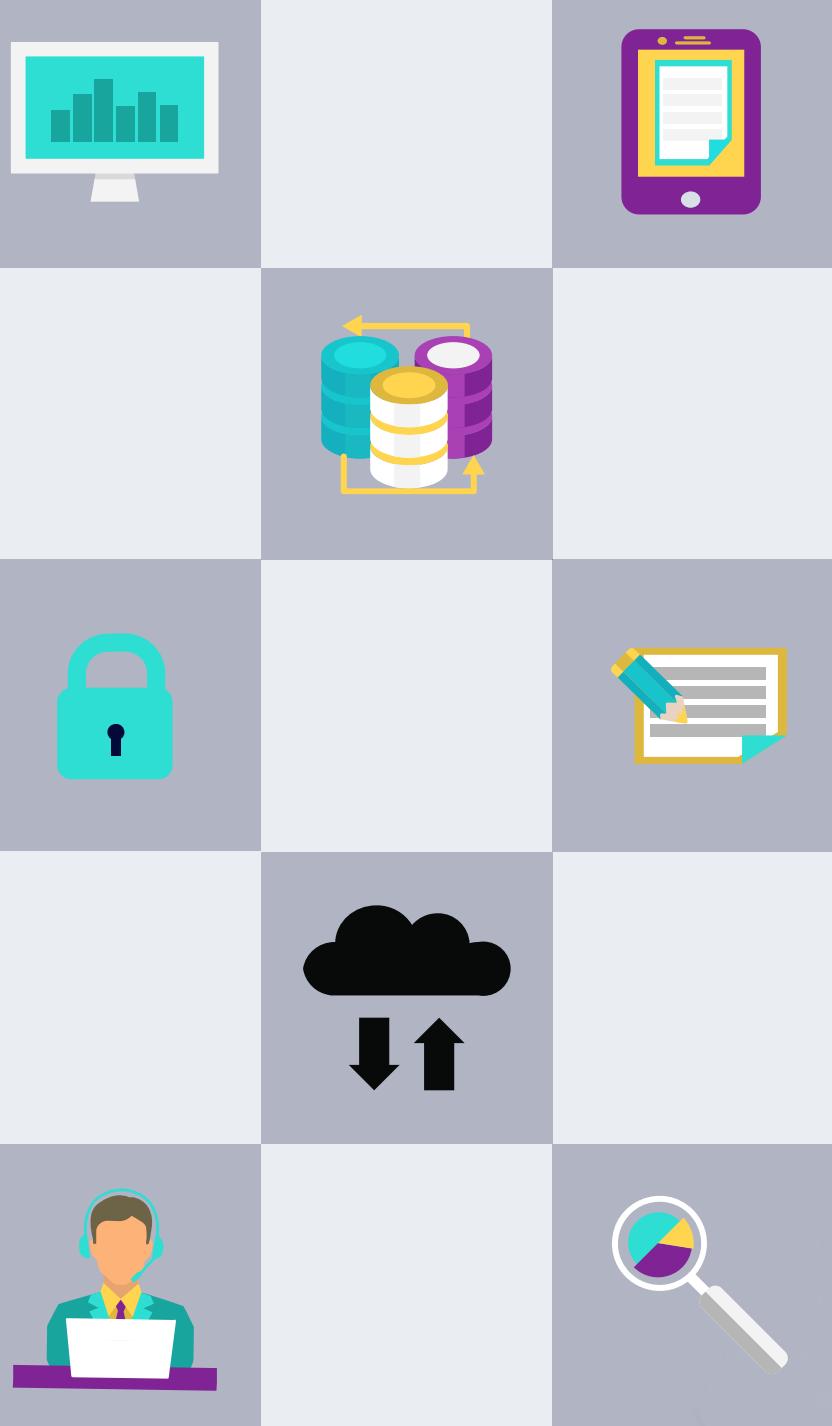
- Growing demand for personalized medicine: Expanding market share with tailored healthcare solutions.
- Advancements in AI and data analysis: Improving efficiency and accuracy in the 3D printing process.
- Collaboration with healthcare providers: Accessing new markets and increasing sales through partnerships.
- Research and development: Driving growth and competitiveness through innovative products and technologies.

Threats

- Competition: Challenges from other companies offering similar products and services.
- Regulatory challenges: Strict regulations impacting speed and cost of bringing new products to market.
- Technological disruption: Rapid advancements potentially outpacing current capabilities.
- Intellectual property protection: Ensuring protection of proprietary technologies, designs, and processes.

Business Problem

- **Attendee Segmentation:** Cluster attendees based on 'Industry', 'Organization', 'Job Title', 'Country/Region', and 'Source Name' for enhanced understanding of attendee demographics.
- **Event Attendance Prediction:** Utilize features like 'Industry', 'Organization', 'Job Title', 'Country/Region', etc., to predict likely attendance for future events.
- **Webinar/Session Optimization:** Analyze 'Time in Session (minutes)', 'Join Time', 'Leave Time', and 'Questions & Comments' to optimize the timing and content of sessions.
- **Consent Management:** Observe fields related to consent to understand attendee comfort levels with data sharing and recording, thereby improving data usage planning and addressing privacy concerns.
- **Sponsorship Analysis:** If 'Source Name' indicates sponsors, analyze the success of different sponsors in attracting attendees.
- **Attendee Origin Analysis:** Examine 'Country/Region', 'City', 'State/Province', and 'Zip/Postal Code' to gain insights into the geographical distribution of attendees for targeted marketing efforts.
- **Marketing Channel Effectiveness:** Analyze 'Email' domain, 'Linkedin Link', and 'Source Name' to determine which marketing channels drive the highest event attendance.
- **Session Feedback Analysis:** Mine 'Questions & Comments' for insights about the sessions' strengths and areas for improvement.



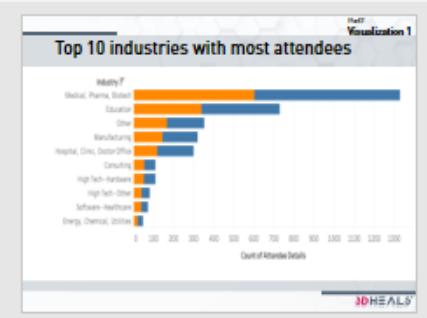
03

Analytics & Visuals

Predictive Model & Exploratory Data Analysis

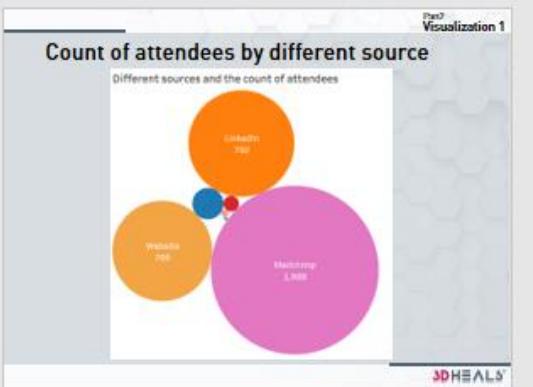
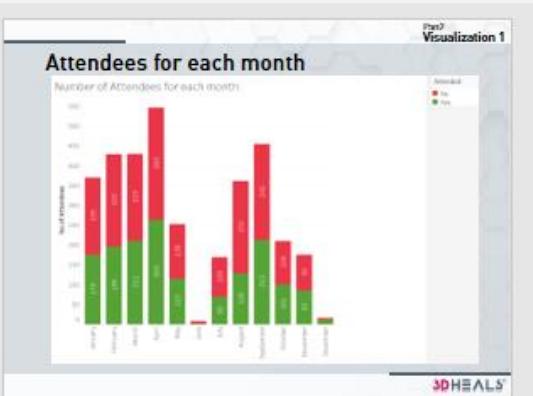
Summary of EDA 1

- Over the course of four seasons, there was a gradual decrease in both the number of registrations and attendees for the meetings, with Season 1 having 1388 and Season 4 having 686 registrations.
- Almost half of the registered members have not attended the meetings across all seasons.
- The medical industry had the highest number of attendees at the meetings, followed by the education and manufacturing industries.
- The topic '3D Bioprinting for Cardiovascular System' had the highest average duration of stay among attendees at 65 minutes. In contrast, the topic 'Healthcare 3D Printing Ecosystem: India' had the least average duration of stay among attendees at 7 minutes.
- The meeting about the product '3D Bioprinting Skin', held during Season 1, had the maximum number of attendees at 250.



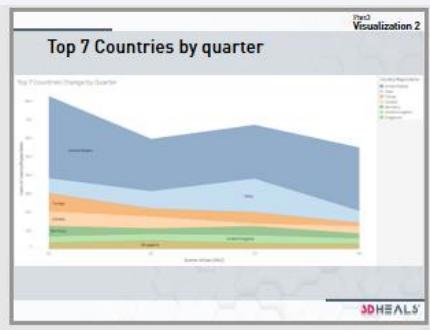
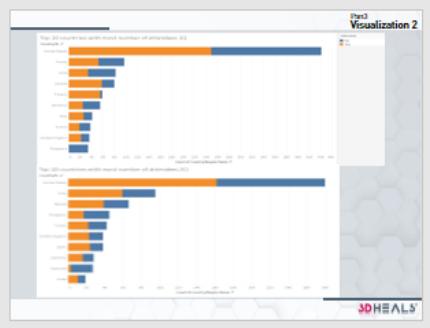
Summary of EDA 2

- April had the highest number of attendees who answered "No" (283) and "Yes" (263), followed by September and March. The lowest attendance rates were in June and December. This may be due to a variety of factors like holidays, weather conditions, and marketing efforts.
- The meeting had a diverse range of attendees, including many students and professionals such as CEOs, Directors, professors, and researchers. Students likely attended to gain knowledge and networking opportunities in the 3D printing and healthcare industries.
- The attendees of the meeting learned about the 3DHeals event from a variety of sources, with the highest number coming from Mailchimp, followed by LinkedIn and other websites.



Summary of EDA 3

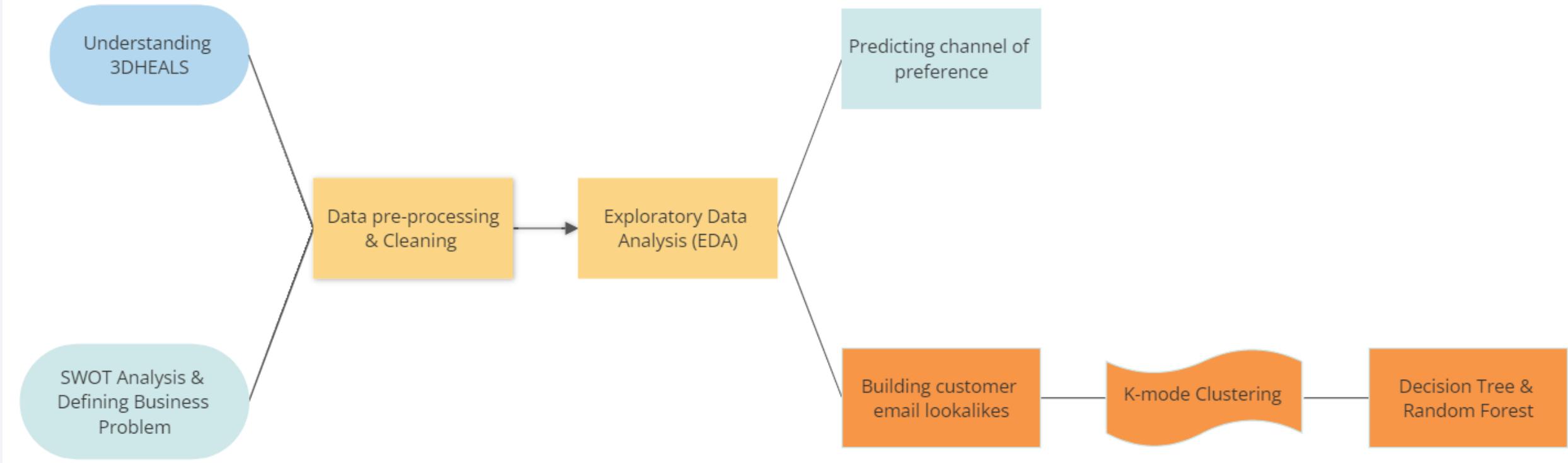
- The top 5 registration by country was the United States with 1,349, followed by India with 426, followed by Turkey, Canada and Germany.
- Checking the top 10 countries by each quarter, except top 5 countries in world wide, there are mainly European countries such as the UK and Spain. There are also Singapore, Australia, Israel and Brazil.
- Top 7 countries by the number of registrations in each country by quarter fluctuated rather than constant.
- In attendance rate by country, most of them were almost 50% attending and 50% not attending. Among the top 10 countries, Singapore shows 81%, Australia shows 85% not attended. On the other hand, Canada has the highest attendance rate at 63%, followed by the UK and Germany with 59% and 55%.
- By proportions of the total attendees, the United States accounted for a quarter, followed by India with 8%, Turkey with 3.5% and Canada with 4.5%. Overall registrations were higher in Turkey, but due to attendance rates, the number of attendees in Canada was higher.

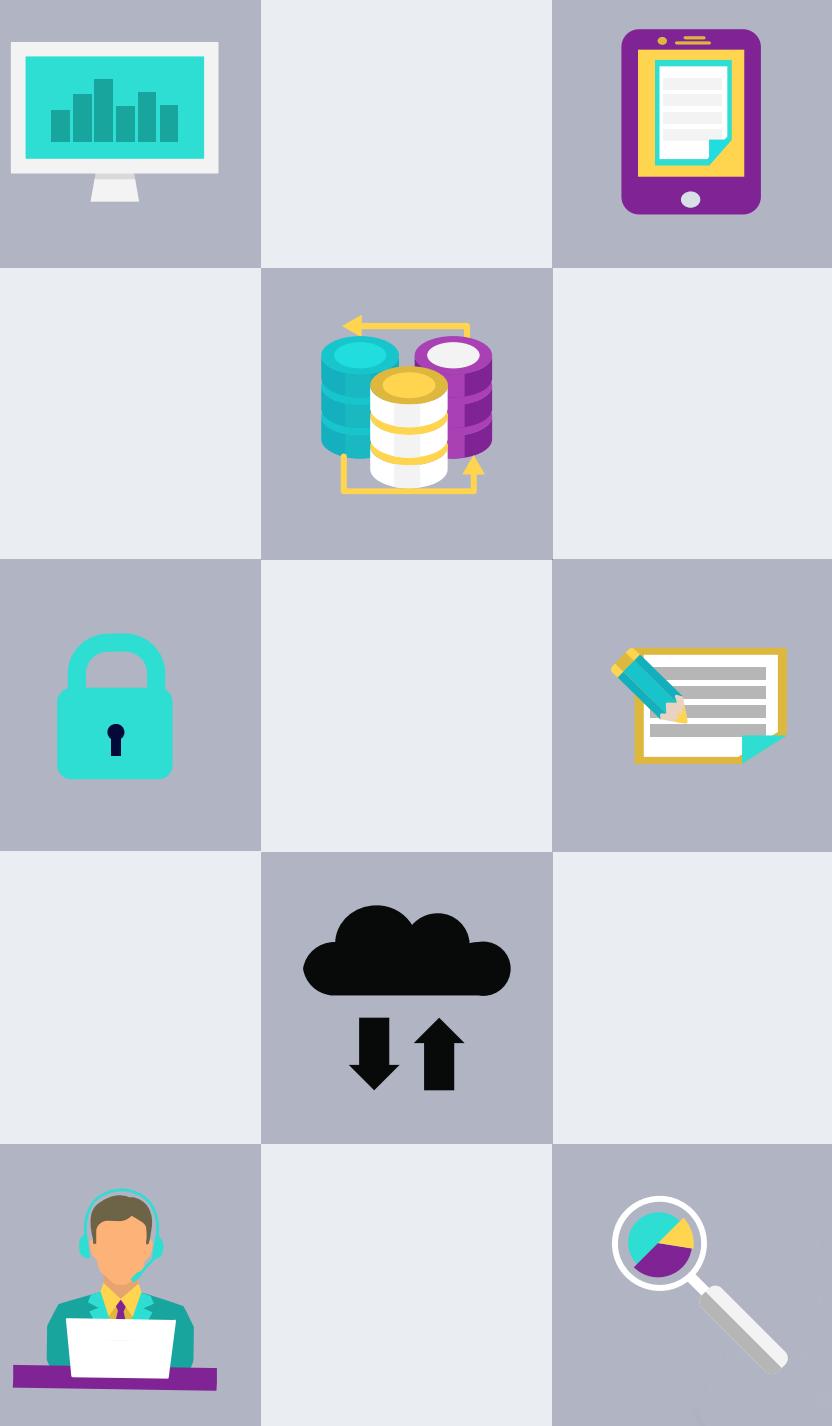


Clear Concise Flow

04

Flow of our Analysis



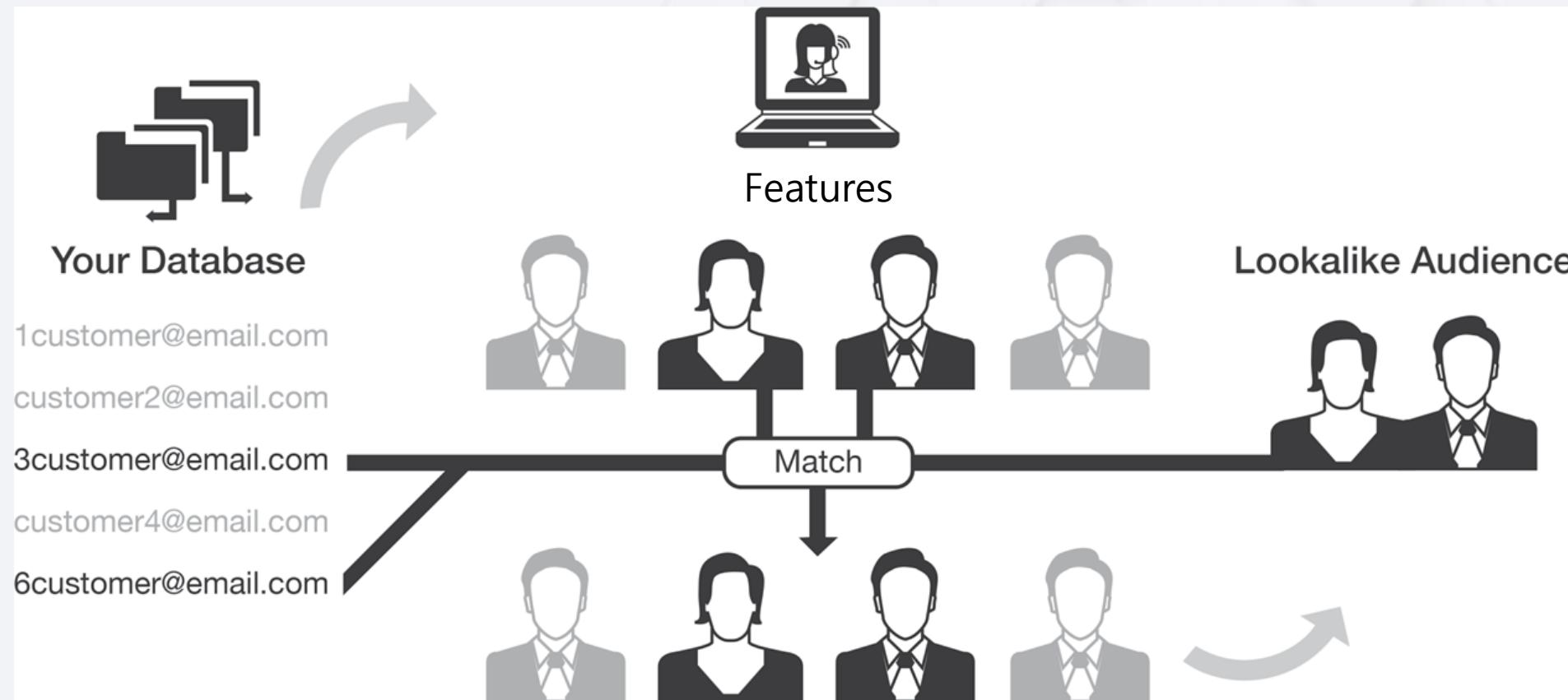


05

Analysis and synthesis

Predictive Model: Building Customer email lookalikes

Model Target: Building customer email lookalike



K-Mode Clustering

What if the data is

Categorical

Master's student
Research Engineer
Director Biomedical Engineer Researcher CTO
CEO PhD student Founder
Associate Professor
Research Assistant President Partner
Professor Engineer Consultant Student
Scientist

| Job Title |
|--------------|
| Country |
| Industry |
| Organization |
| Source Name |
| Domain |

Data Cleaning

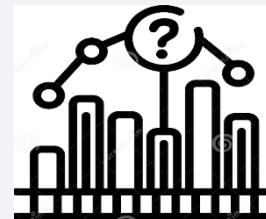
Removing unwanted variables

Removed: 'User Name (Original Name)', 'First Name', 'Last Name', 'Is Guest', ...



01

Removing empty/missing records



Extracting Domain Name From email

03

| domain_name |
|-----------------|
| gmail.com |
| uq.edu.au |
| gmail.com |
| pegamedical.com |
| wakehealth.edu |
| ... |
| clecell.co.kr |
| inobitec.com |
| gmail.com |
| outlook.com |
| gmail.com |

Finding optimal k with elbow method

Select Cluster variables

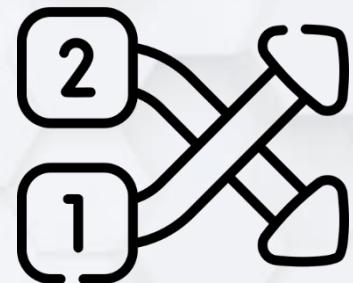
| Column | Unique |
|--------------|--------|
| Country | 87 |
| Industry | 30 |
| Organization | 1,876 |
| Job Title | 1,291 |
| Source Name | 9 |
| Domain | 745 |

01

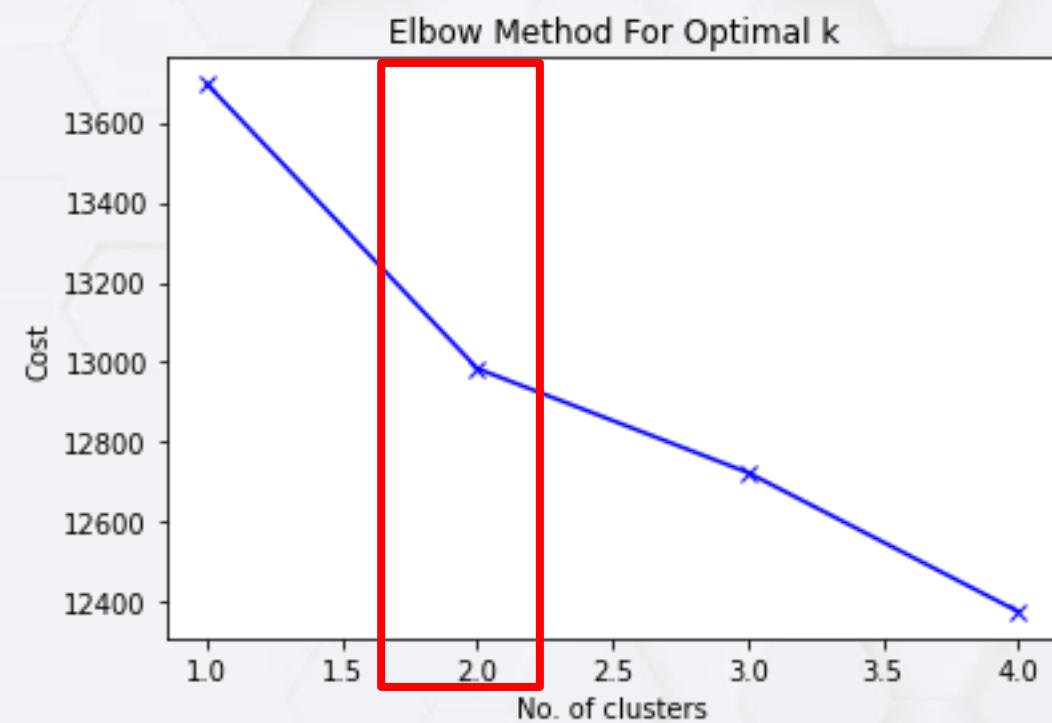
02

03

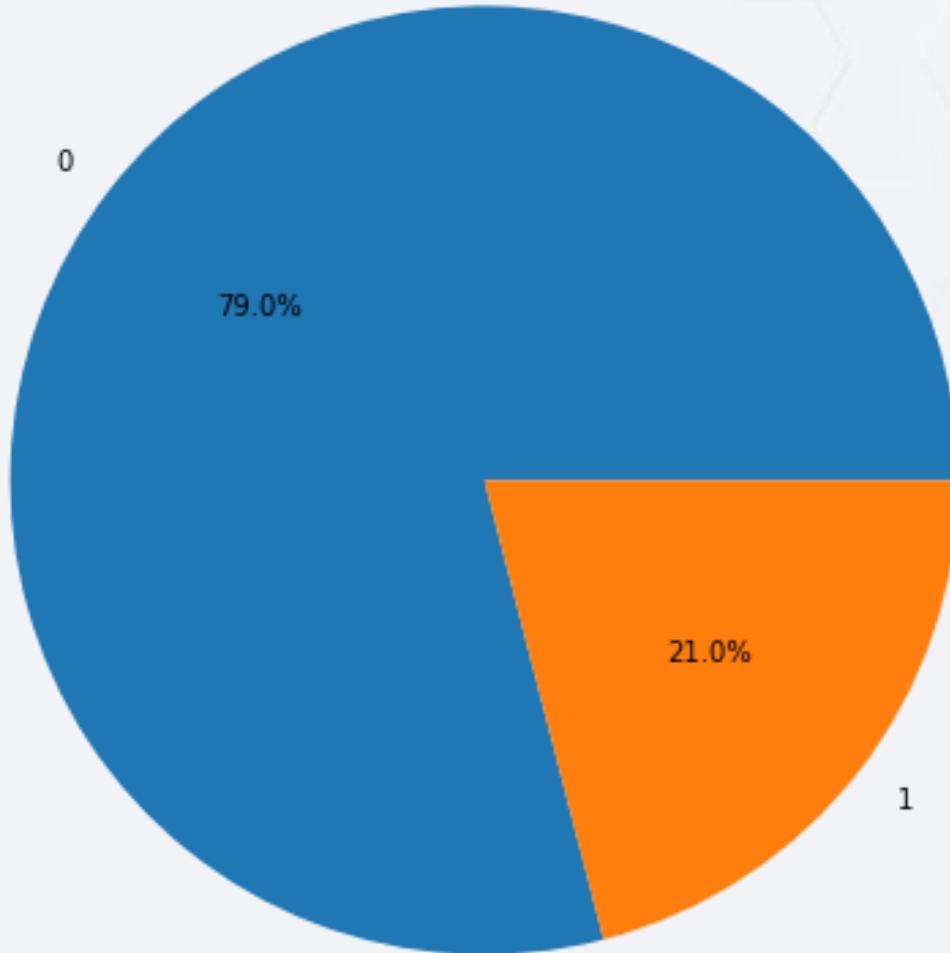
Decide n_Cluster = 2



Elbow curve Finding optimal 'K'



Cluster Ratio



| Cluster | Counts | Percent |
|---------|--------|---------|
| 0 | 2,505 | 79.0% |
| 1 | 573 | 21.0% |

Cluster 'o' and Centroids



US
United States
Country/Region



Medical, Pharma, Biotech
Industry



VoxCell BioInnovation
Organization



CEO
Job Title



Mailchimp
Source Name



Gmail.com
Domain_name

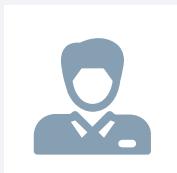
Cluster '1' and Centroids



IN
India
Country/Region



Education
Industry



Student
Job Title



Mailchimp
Source Name

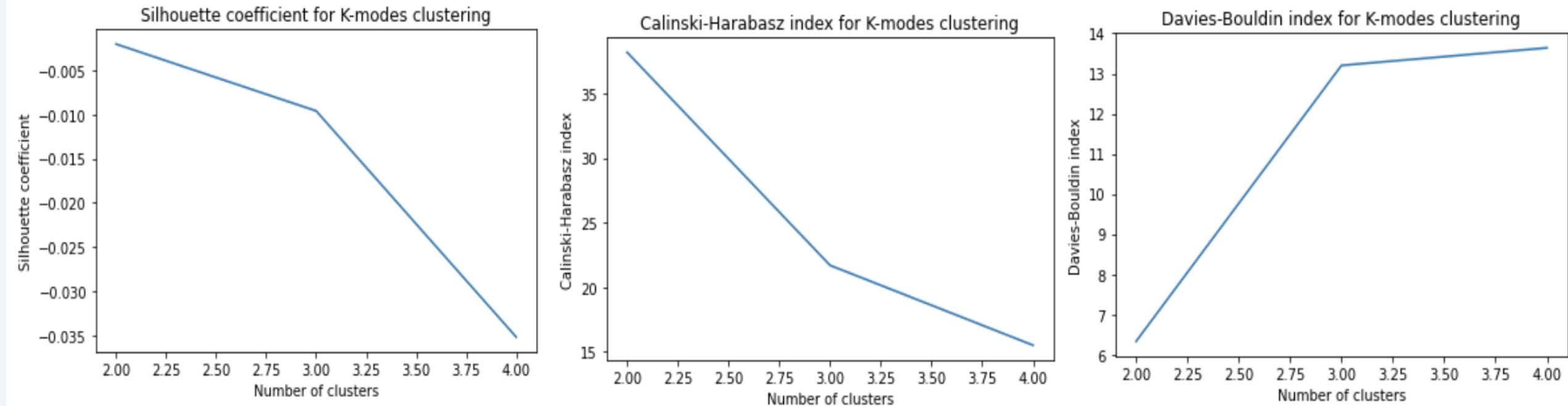


**The Hong Kong
Polytechnic University**
Organization



Poly.edu.hk
Domain_name

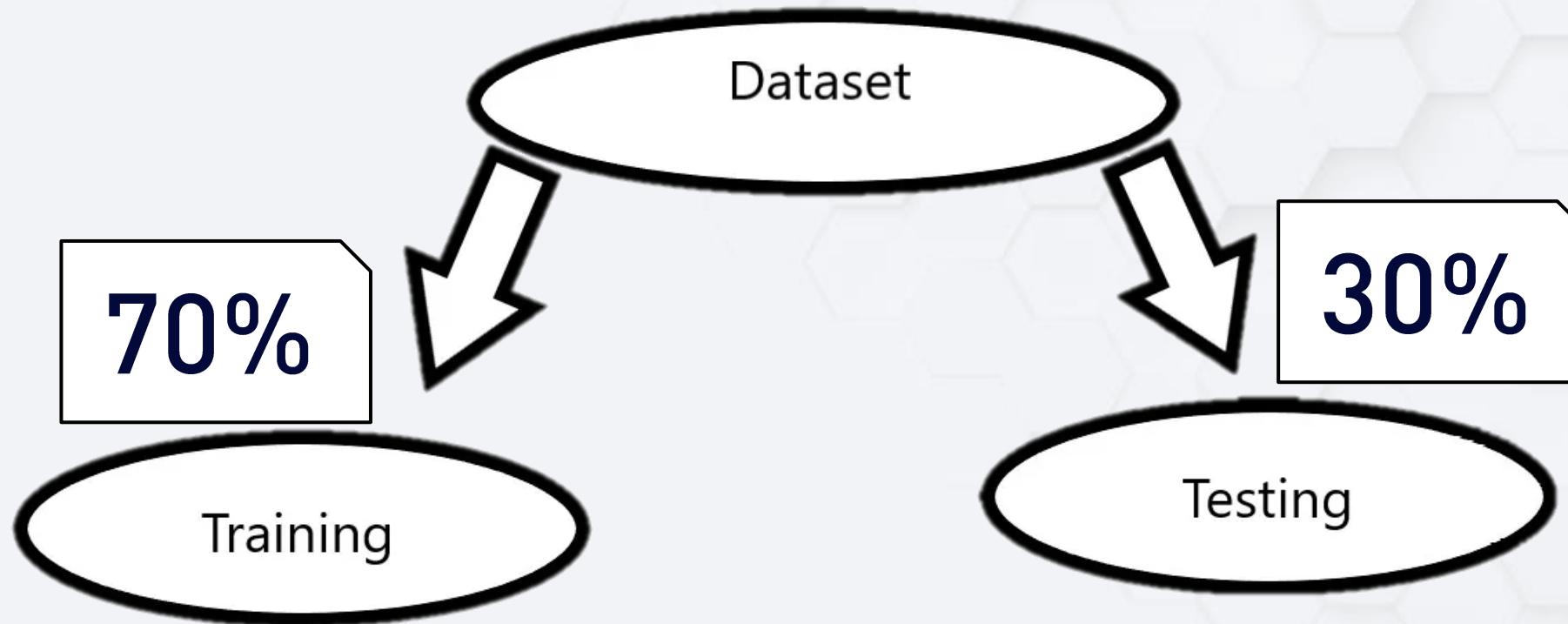
Validating the effectiveness of clusters



Silhouette coefficient

-
-

Split the train & test data



Decision Tree

Industry:
Education

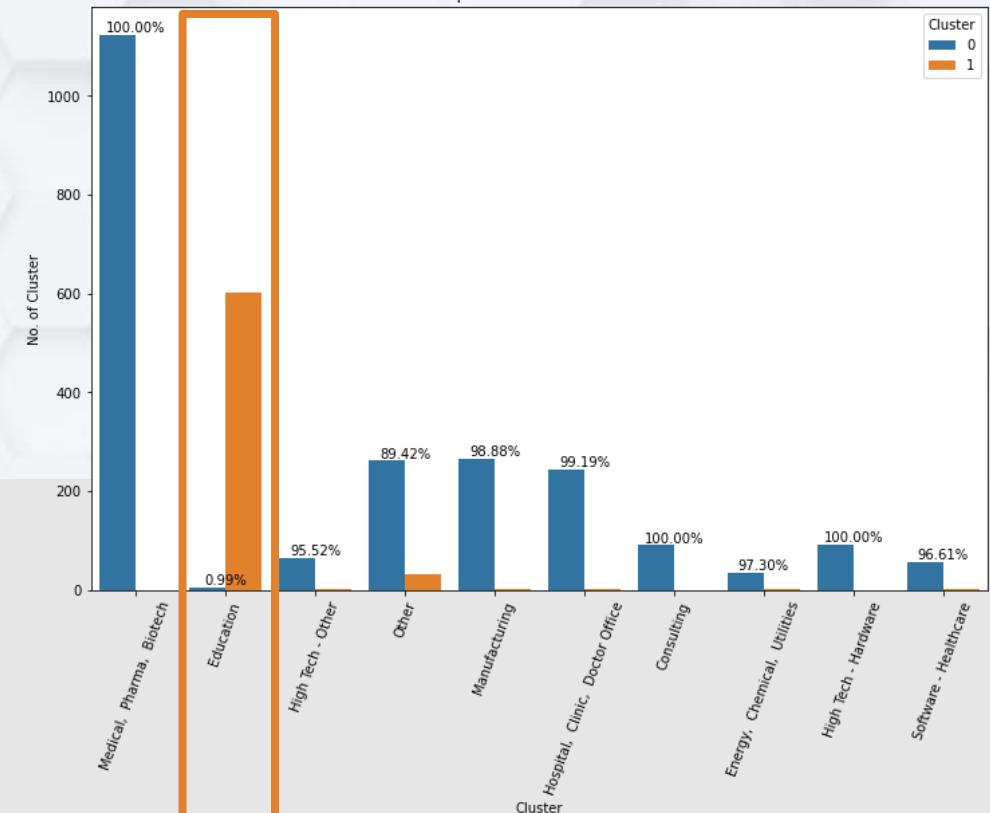
encoder_Industry_8 <= 0.5
gini = 0.308
samples = 2154
value = [1745, 409]

Country/Region:
India

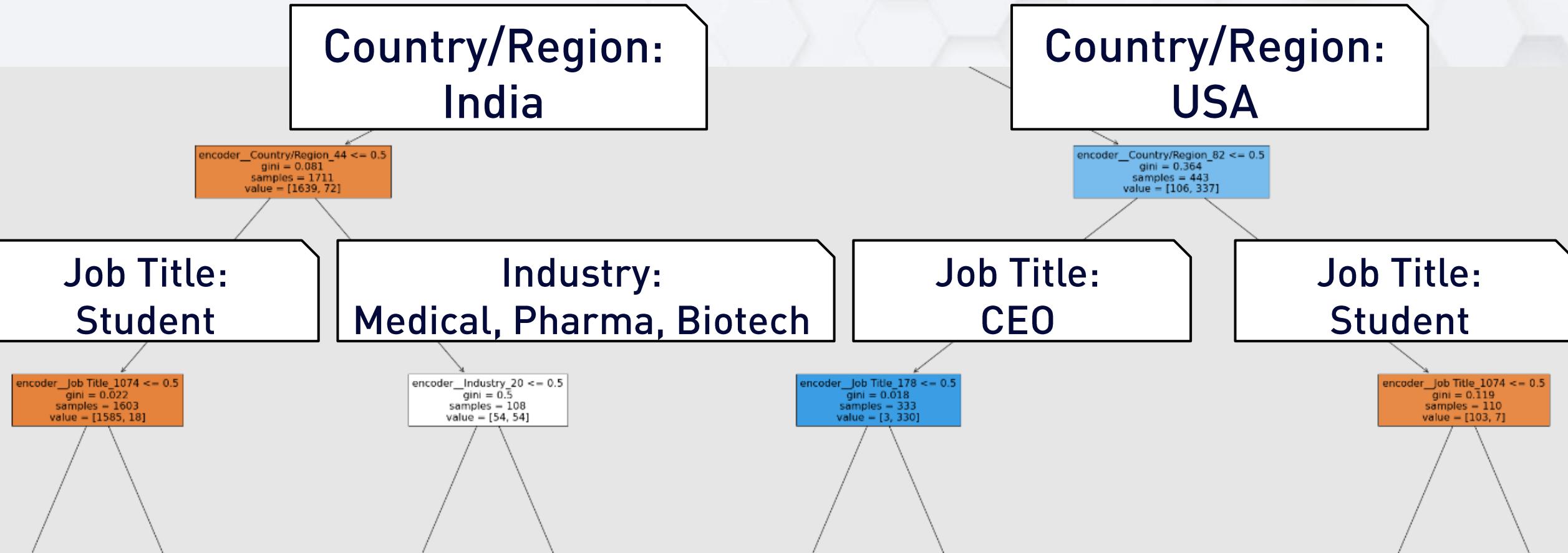
encoder_Country/Region_44 <= 0.5
gini = 0.081
samples = 1711
value = [1639, 72]

Country/Region:
USA

encoder_Country/Region_82 <= 0.5
gini = 0.364
samples = 443
value = [106, 337]



Decision Tree



Decision Tree

Country/Region:
India

```
encoder_Country/Region_44 <= 0.5  
gini = 0.081  
samples = 1711  
value = [1639, 72]
```

Cluster '1' and Centroids

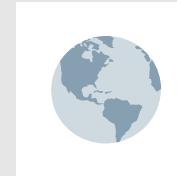


IN
India
Country/Region

Country/Region:
USA

```
encoder_Country/Region_82 <= 0.5  
gini = 0.364  
samples = 443  
value = [106, 337]
```

Cluster '0' and Centroids



US
United States
Country/Region

Decision Tree

Job Title:
Student

```
encoder_Job Title_1074 <= 0.5
gini = 0.022
samples = 1603
value = [1585, 18]
```

Industry:
Medical, Pharma, Biotech

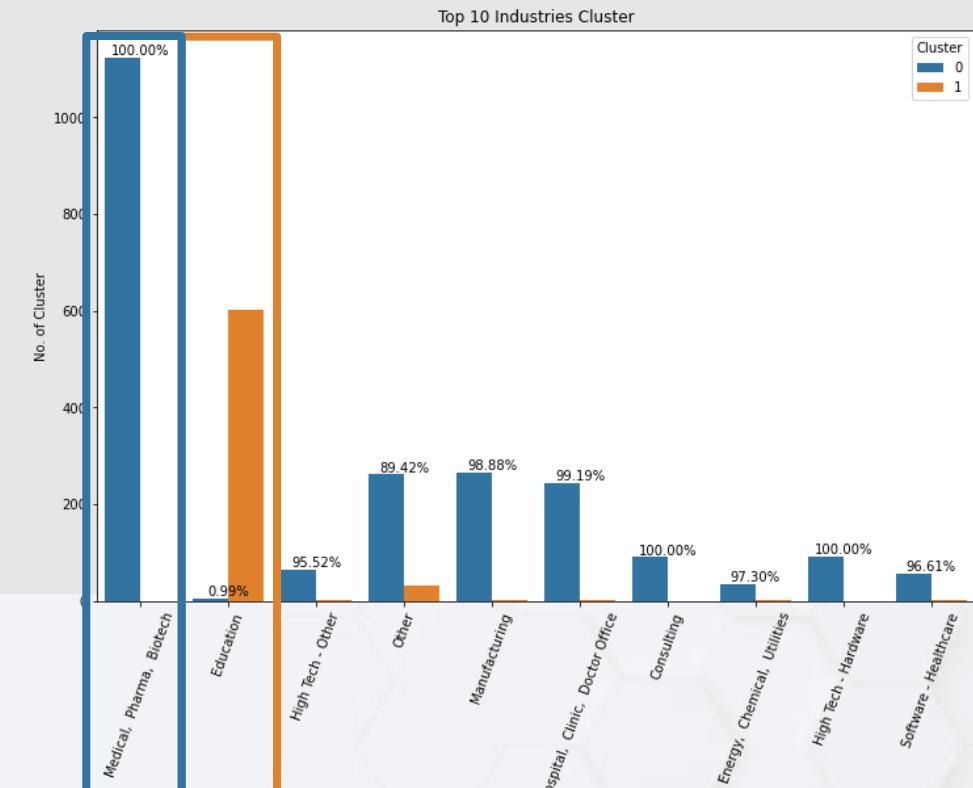
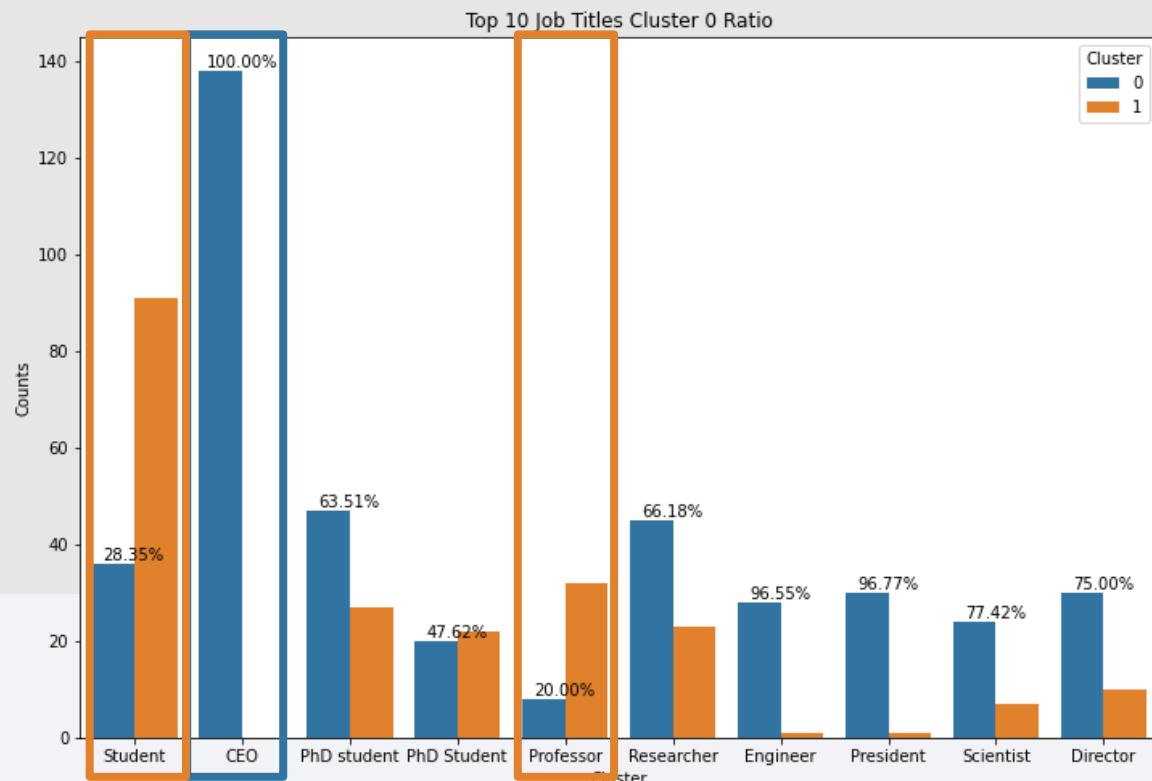
```
encoder_Industry_20 <= 0.5
gini = 0.5
samples = 108
value = [54, 54]
```

Job Title:
CEO

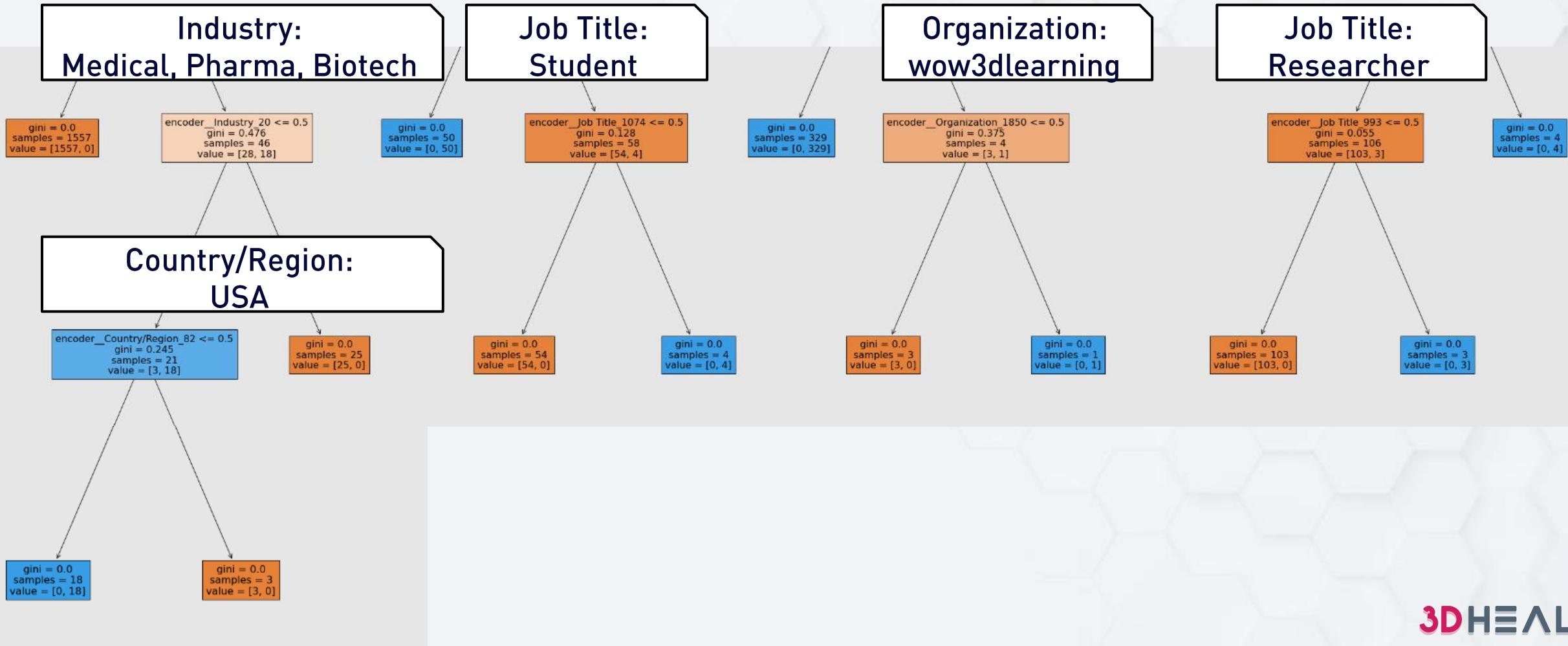
```
encoder_Job Title_178 <= 0.5
gini = 0.018
samples = 333
value = [3, 330]
```

Job Title:
Student

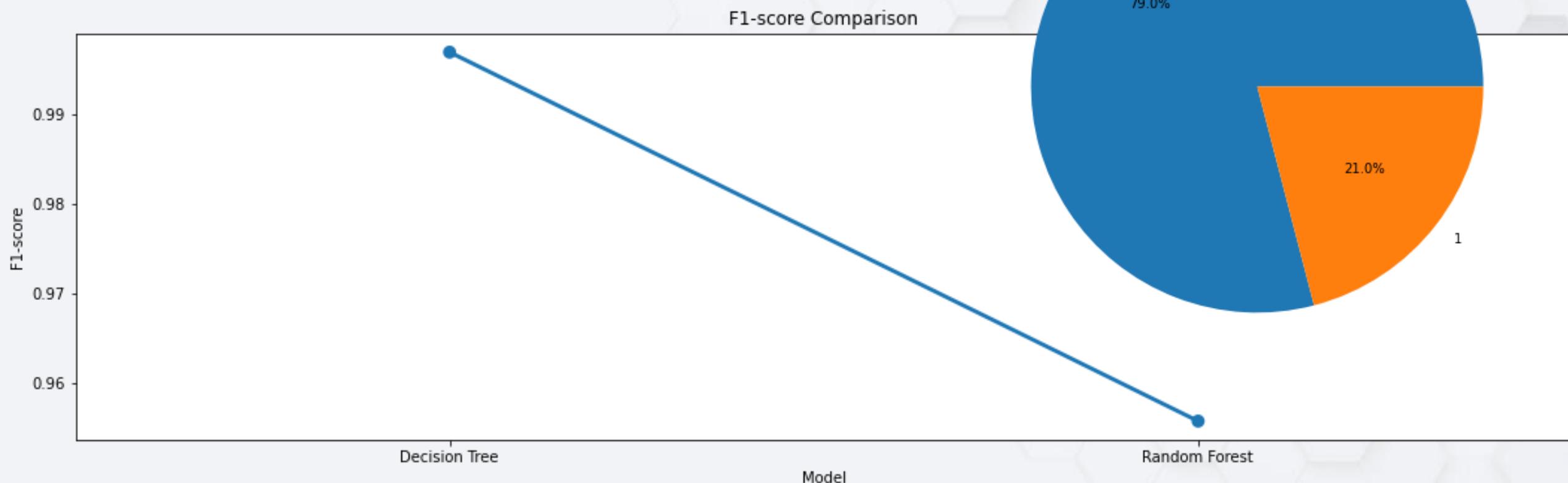
```
encoder_Job Title_1074 <= 0.5
gini = 0.119
samples = 110
value = [103, 7]
```



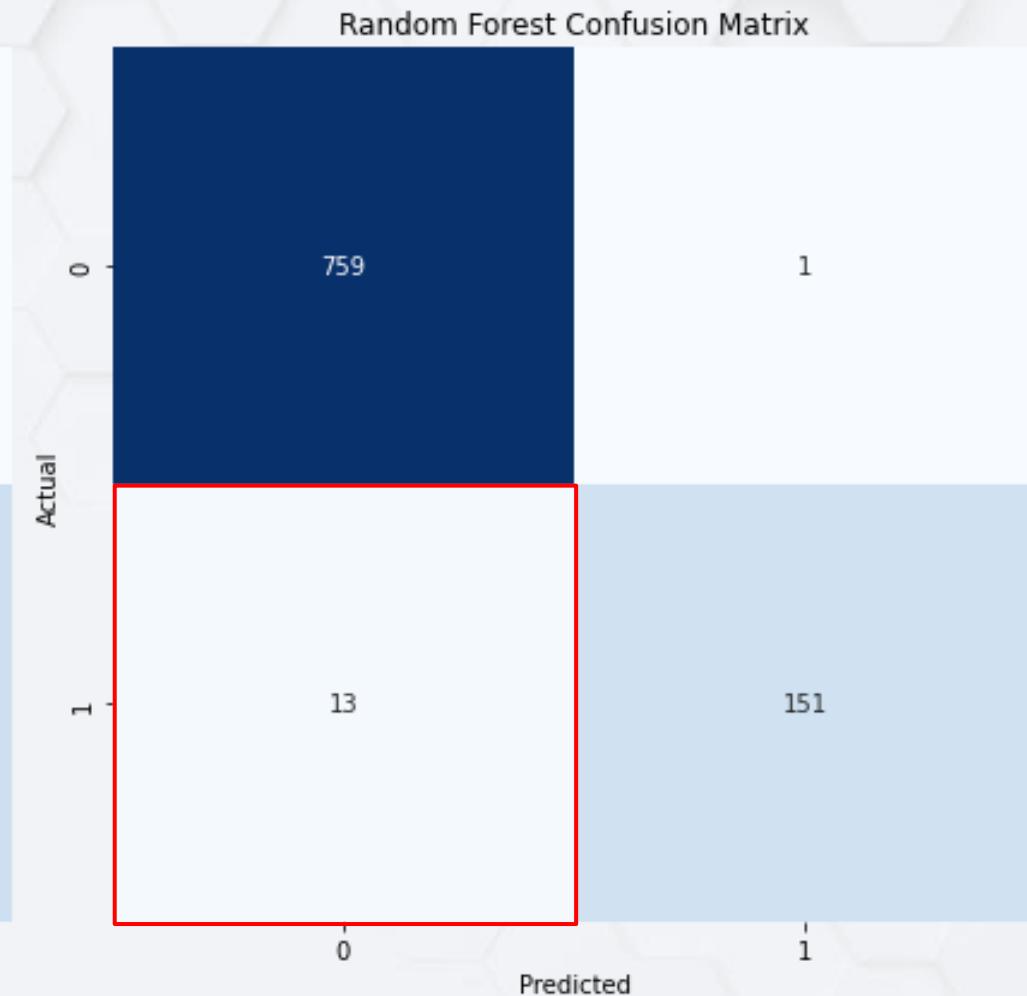
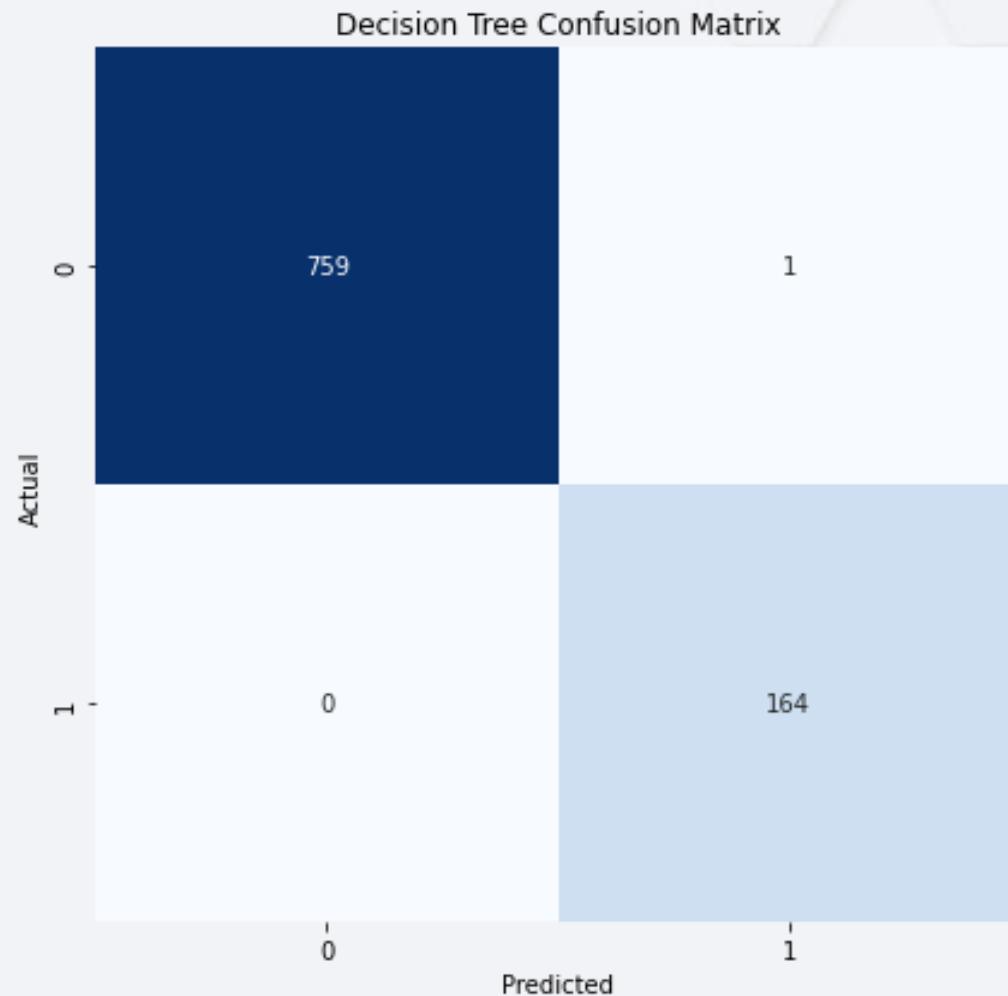
Decision Tree



F-1 Score comparison

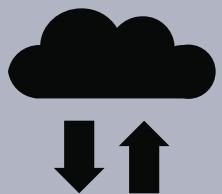
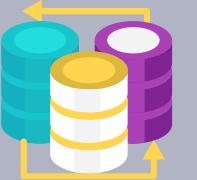
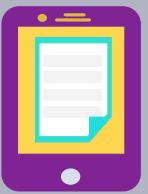


Confusion Matrix



06

Recommendations & Findings & Future plans



Recommendations

- **Customized Marketing:** Leverage attendee segmentation for designing personalized marketing strategies to increase engagement and event attendance.
- **Predictive Modelling:** Employ predictive models to forecast event attendance for efficient resource planning and management.
- **Event Timing and Content:** Evaluate session/webinar engagement to identify areas for content enhancement or timing adjustment.
- **Data Privacy Considerations:** Maintain respect for data privacy, exploring alternatives if attendees express discomfort with recording or information sharing.
- **Sponsorship Collaboration:** Analyze sponsor influence on event attendance for potential strategic collaborations.
- **Geographic Targeting:** Consider focusing on regions with higher attendance rates for event hosting or increased marketing efforts.
- **Marketing Channels:** Assess the effectiveness of various marketing channels to concentrate efforts on the most impactful ones.
- **Feedback Analysis:** Apply natural language processing techniques to 'Questions & Comments' for valuable attendee sentiment analysis and direct feedback.

Finding: Centroids

Cluster '0'

US
United States

Medical, Pharma,
Biotech

VoxCell BioInnovation

CEO

Mailchimp

Gmail.com

Country/Region



Industry



Organization



Organization



Source Name



Domain



Cluster '1'

IN
India

Education

The Hong Kong
Polytechnic University

Student

Mailchimp

Poly.edu.hk

Future Plans

Sentiment Analysis: Using Natural Language Processing (NLP) techniques in the 'Questions & Comments' field might reveal valuable insights about the attendees' opinions and attitudes. This can help in improving the event experience further.

Predictive Modeling for Attendance: Build a predictive model using machine learning techniques to forecast an individual's attendance for future events based on their past records and engagement.

Network Analysis: If data on the relationships between attendees is available (such as connections on LinkedIn), network analysis could be used to identify influencers and key clusters within the attendee community.

Time Series Analysis: This can be used to understand the patterns in attendee registrations and dropout rates over time. This can help in better planning and predicting future event attendance.

Churn Analysis: Perform a churn analysis to identify individuals who have stopped attending the events. Understanding these individuals' characteristics can help devise strategies to re-engage them.

Q&A

Thank you!
3DHEALS®

REFERENCE

- 3DHEALS. (n.d.). About us. Retrieved from <https://3dheals.com/about-us>
- Deloitte. (2020). 2020 global health care outlook: Laying a foundation for the future. Retrieved from <https://www2.deloitte.com/us/en/pages/life-sciences-and-health-care/articles/global-health-care-sector-outlook.html>
- Machado, L. (2018, August 20). Artificial intelligence and 3D printing. Medium. <https://medium.com/healthcare-3d-printing-stories/artificial-intelligence-and-3d-printing-94f45f3e45dd>
- PwC. (2019). What's next for the pharmaceuticals industry, amid digital disruption and rapid technological advances? Retrieved from <https://www.pwc.com/gx/en/pharma-life-sciences/pdf/pwc-pharma-2020.pdf>
- Statista. (2021). 3D printing - Statistics & Facts. Retrieved from <https://www.statista.com/topics/1174/3d-printing>
- 3D Systems. (n.d.). Healthcare. Retrieved from <https://www.3dsystems.com/healthcare>
- 3DHEALS. (n.d.). About us. Retrieved from <https://3dheals.com/about>
- Gartner. (2018). Gartner says 3D printing is changing the landscape of the medical device market. Retrieved from <https://www.gartner.com/en/newsroom/press-releases/2018-04-03-gartner-says-3d-printing-is-changing-the-landscape-of-the-medical-device-market>
- Huang, S. H., Liu, P., Mokasdar, A., & Hou, L. (2020). Additive manufacturing and its societal impact: A literature review. *The International Journal of Advanced Manufacturing Technology*, 67(5-8), 1191-1203. doi: 10.1007/s00170-012-4558-5

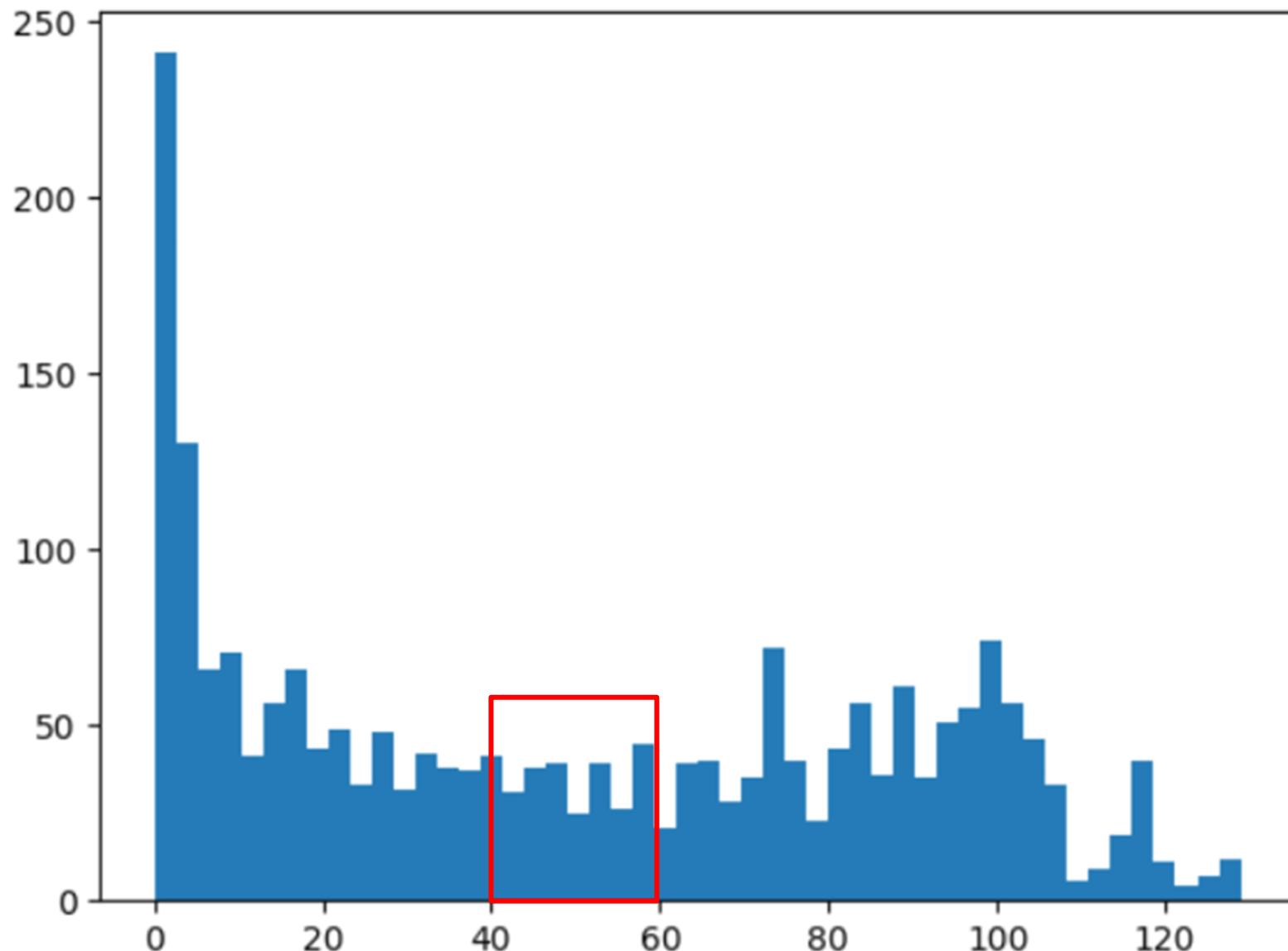
REFERENCE

- Organovo. (n.d.). About Organovo. Retrieved from <https://organovo.com/about/>Rengier, F., Mehndiratta, A., von Tengg-Kobligk, H., Zechmann, C. M., Unterhinninghofen, R.,Kauczor, H. U., & Giesel, F. L. (2010). 3D printing based on imaging data: Review of medical applications. International Journal of Computer Assisted Radiology and Surgery, 5(4), 335-341.doi: 10.1007/s11548-010-0476-x
- Stratasys. (n.d.). Medical solutions. Retrieved from <https://www.stratasys.com/medical>
- Suganya Karunamurthy. (2022). K-Mode Clustering | solved example | implementation. YouTube. Retrieved from <https://www.youtube.com/watch?v=EVI2ejcsTfg>
- Harika Bonthu. (2021, June 13). KModes Clustering algorithm for categorical data. Analytics Vidhya. Retrieved from <https://www.analyticsvidhya.com/blog/2021/06/kmodes-clustering-algorithm-for-categorical-data/#h-2-scree-plot-or-elbow-curve-to-find-optimal-kvalue>
- PRASHANT BANERJEE. (2022). K-Means clustering with Python. Kaggle. Retrieved from <https://www.kaggle.com/code/prashant111/k-means-clustering-with-python>
- ASHISH. (2020). Bank customer clustering (K-Modes Clustering). Kaggle. Retrieved from <https://www.kaggle.com/code/ashydv/bank-customer-clustering-k-modes-clustering/notebook>
- rakshithvasudev. (2017, August 2). What is one hot encoding? why and when do you have to use it?. Hackernoon. Retrieved from <https://hackernoon.com/what-is-one-hot-encoding-why-and-when-do-you-have-to-use-it-e3c6186d008f>

Appendix 1

Exploratory Data Analysis

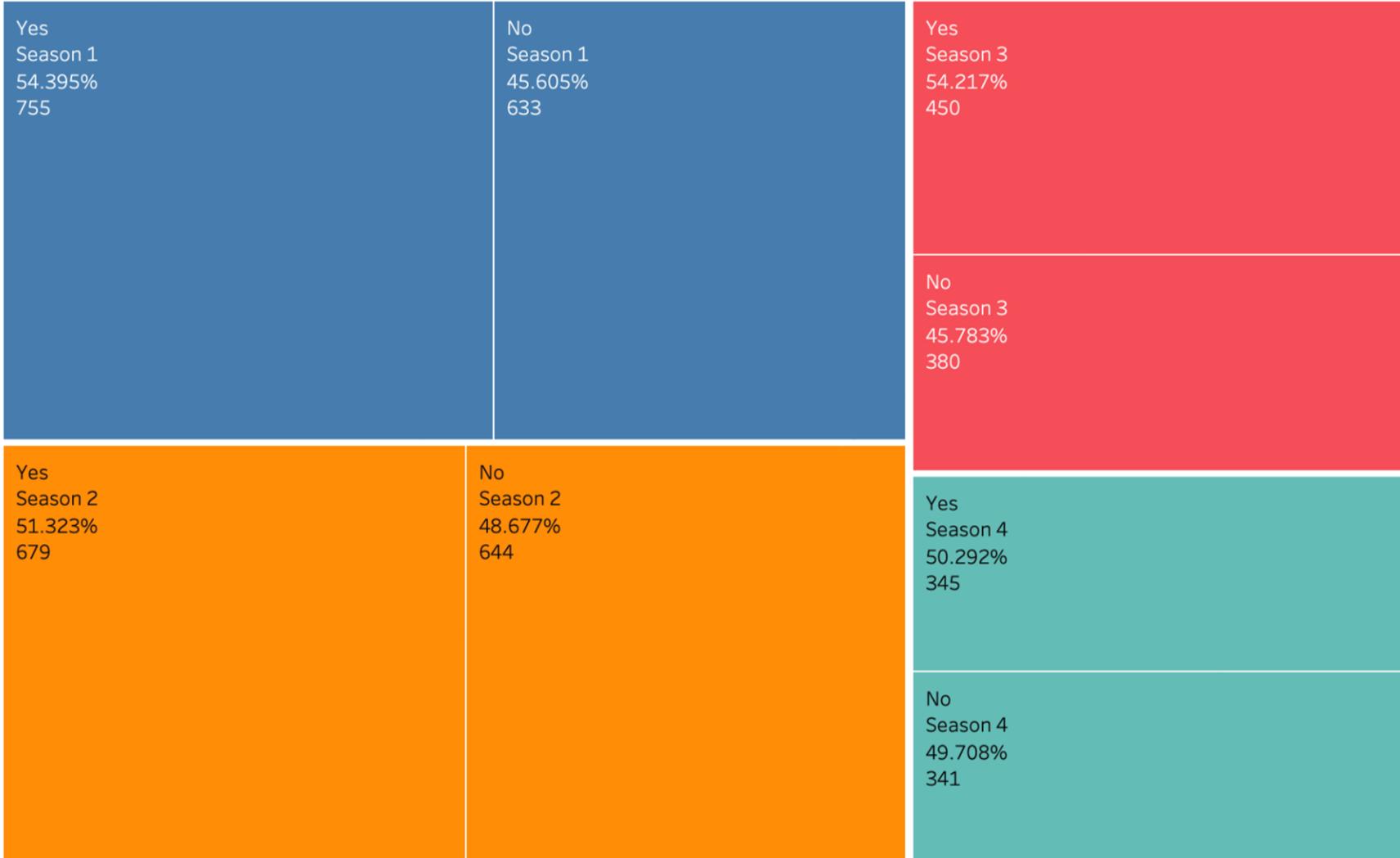
Time in Session (minutes)



Count: 2229.0
Mean: 50.4
Std: 37.9
Min: 0.0
25%: 13.0
50%: 47.0
75%: 85.0
Max: 129.0

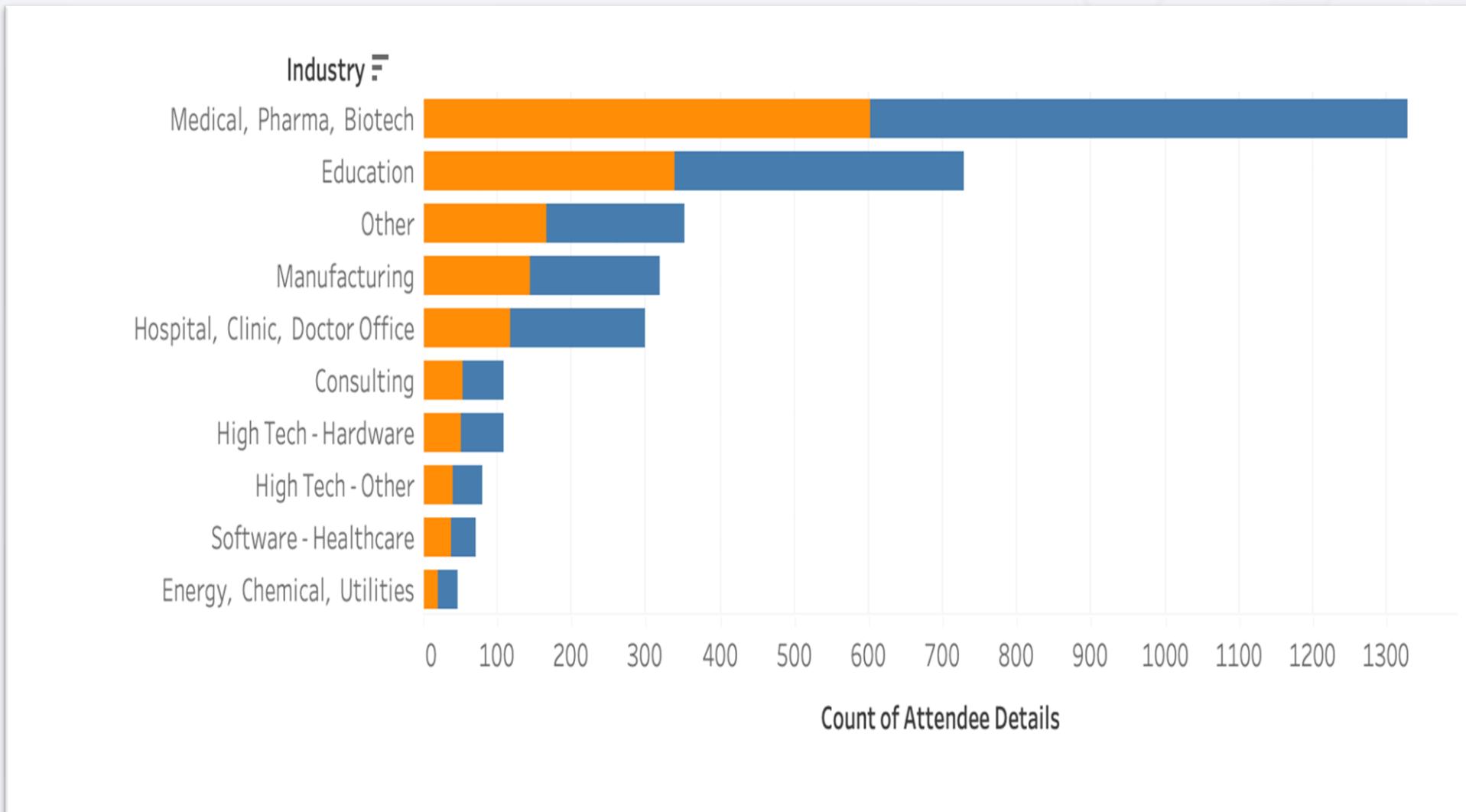
Season wise, Attended

What percentage of registered participants attended the meetings, season wise



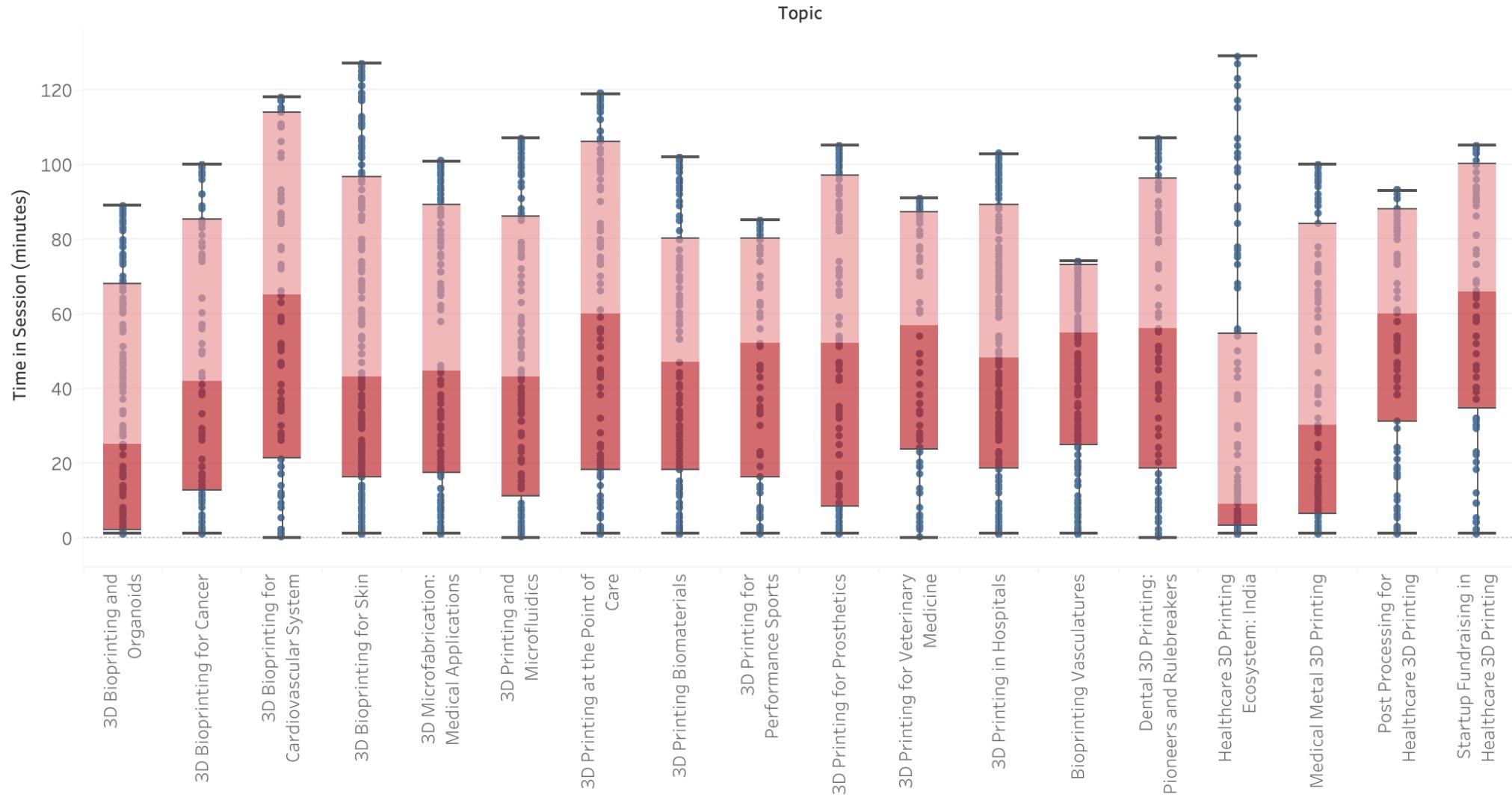
APPENDIX

Top 10 industries with most attendees

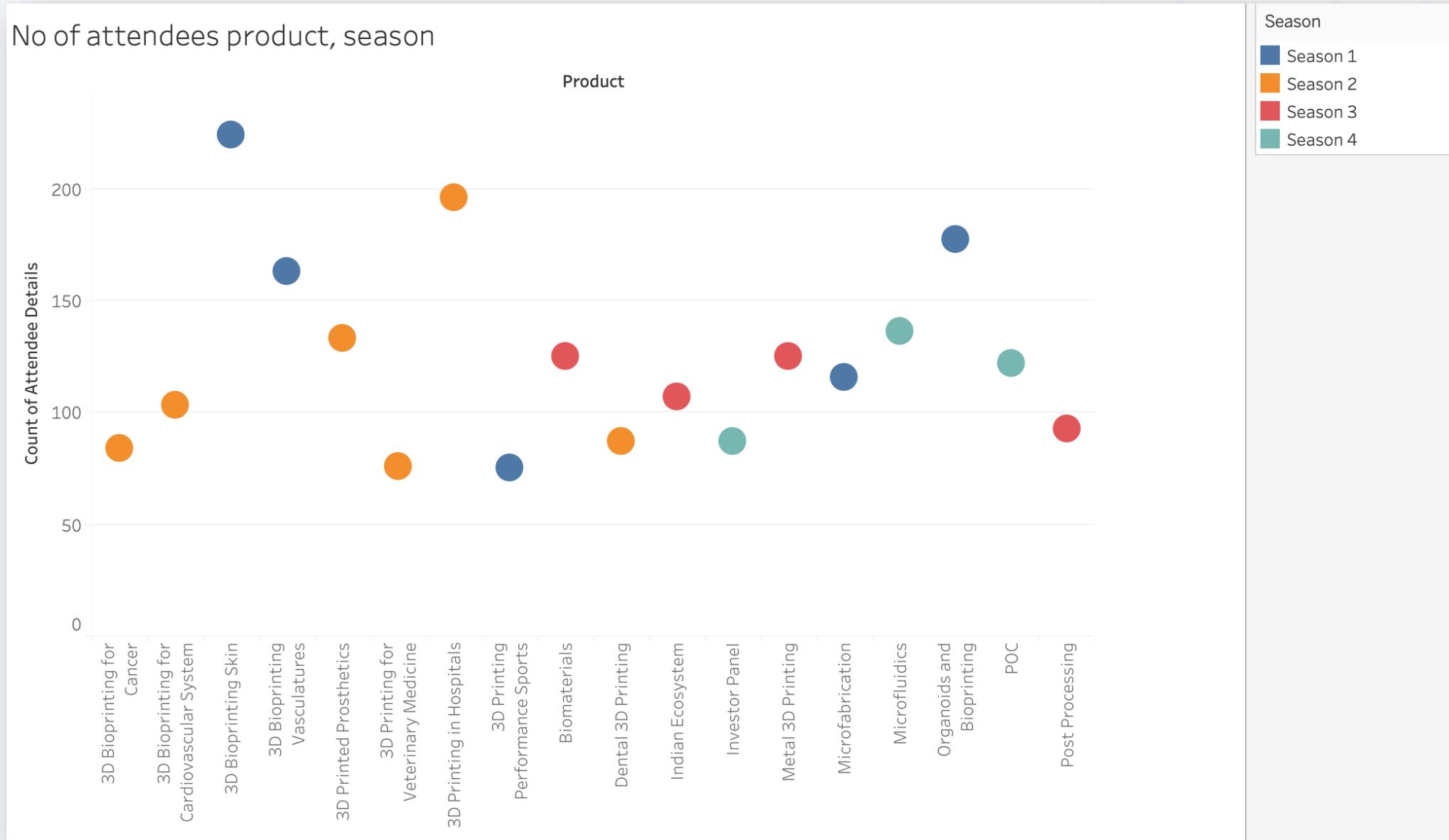


Average Duration

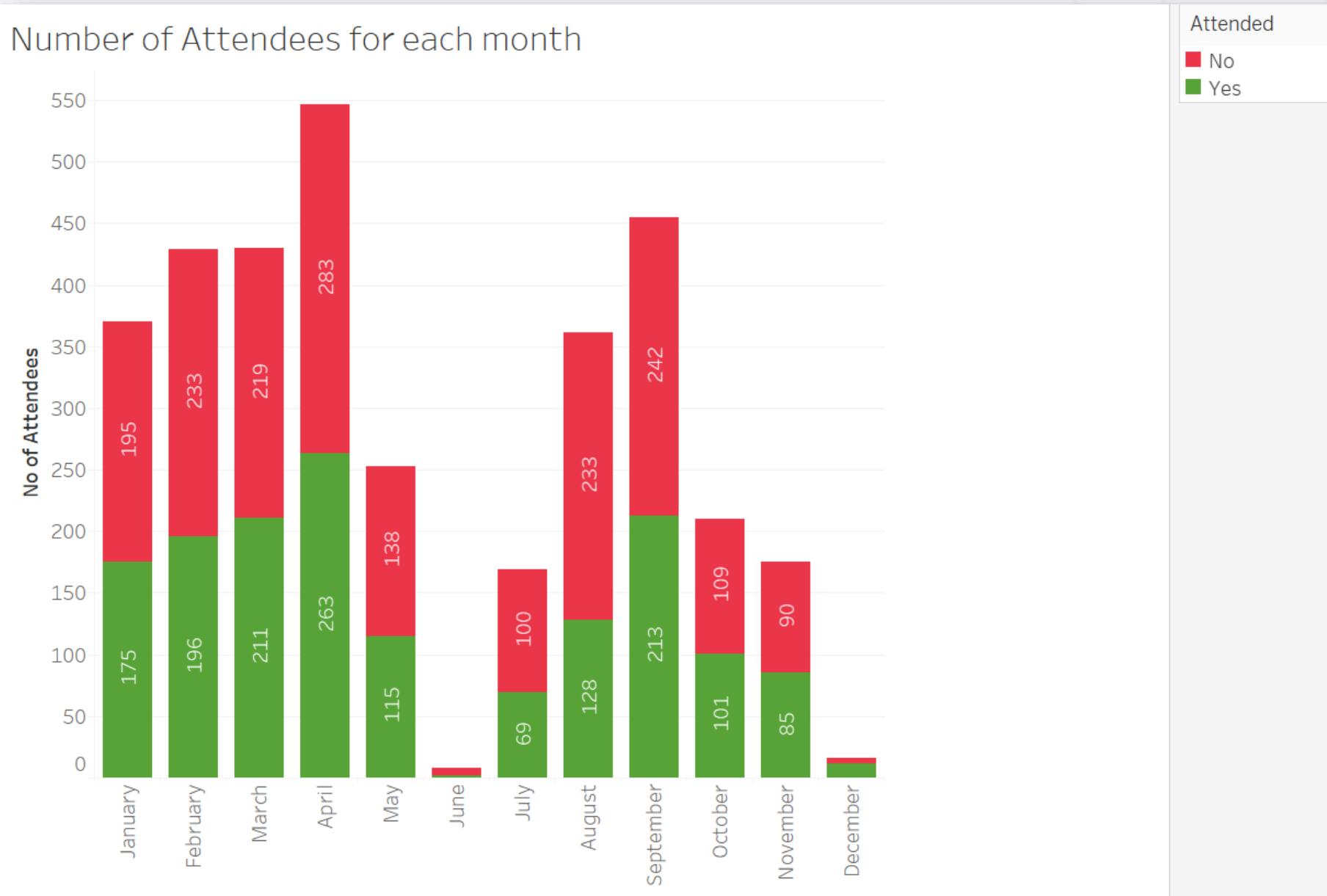
What is the average duration of stay in meetings by the attendees in each topic?



Num of Attendees by product & Season



Attendees for each month



APPENDIX

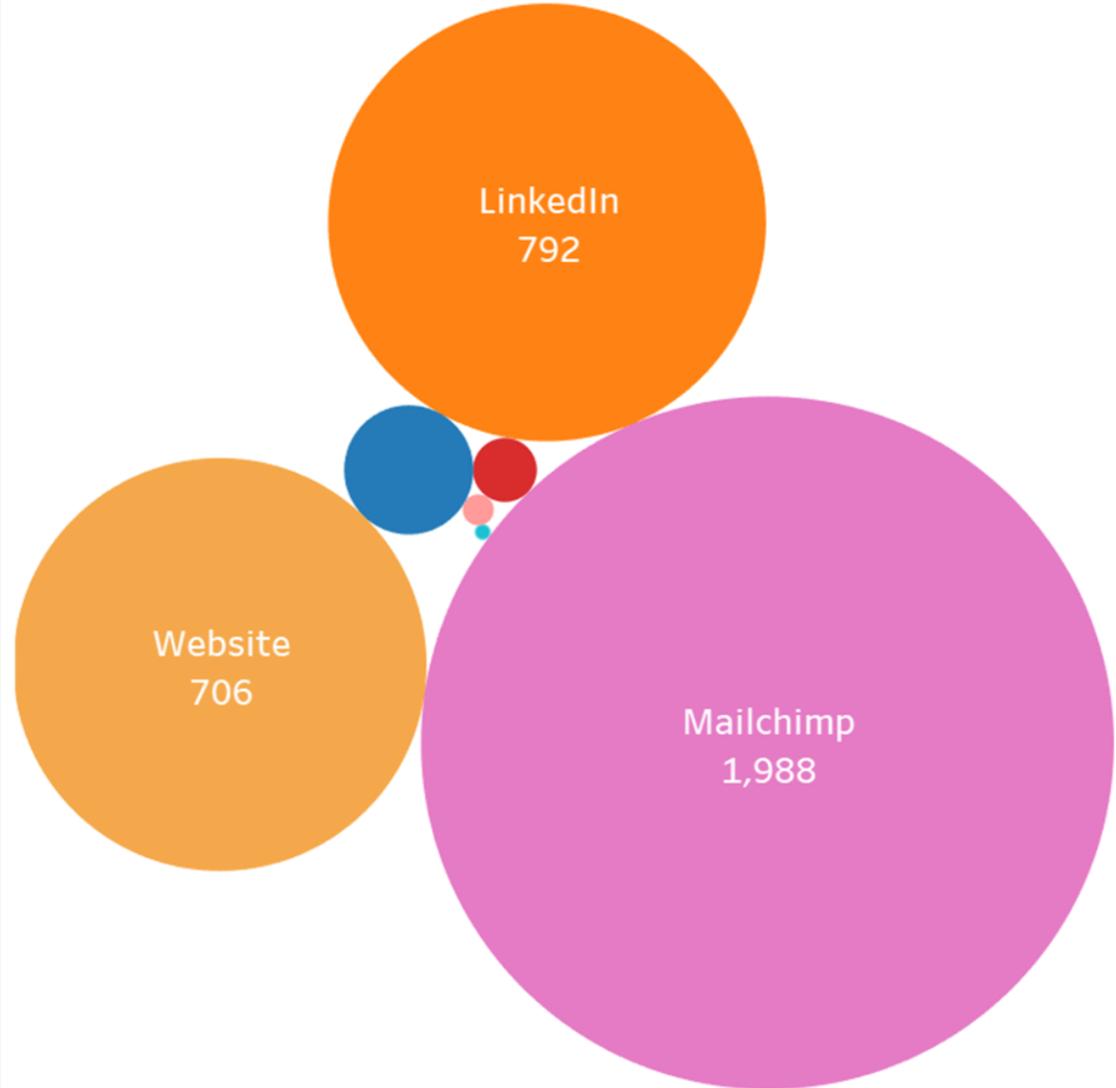
Wordcloud of Top 20 Attendee's Job title

Wordcloud of Top 20 Attendee's Job title



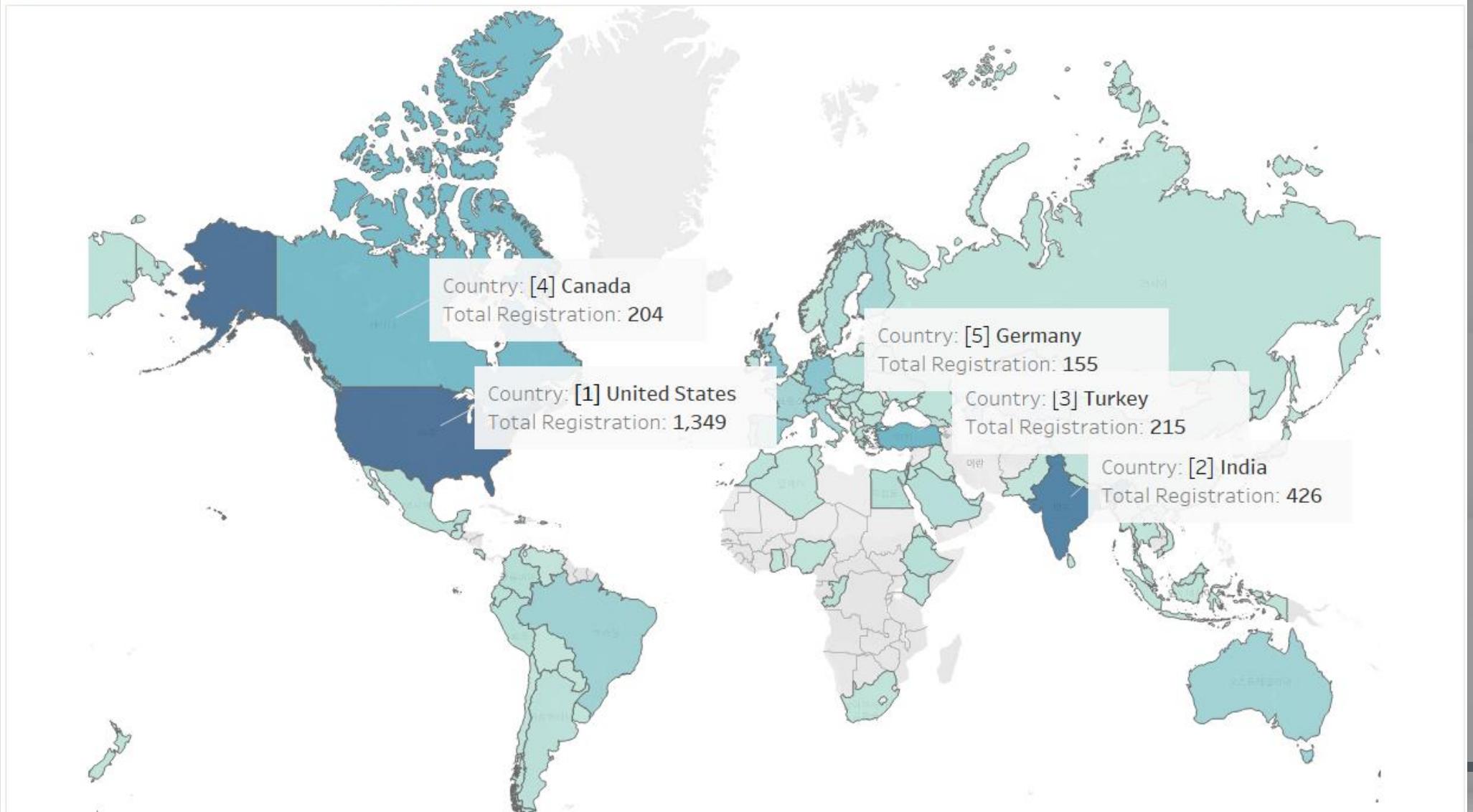
Count of attendees by different source

Different sources and the count of attendees

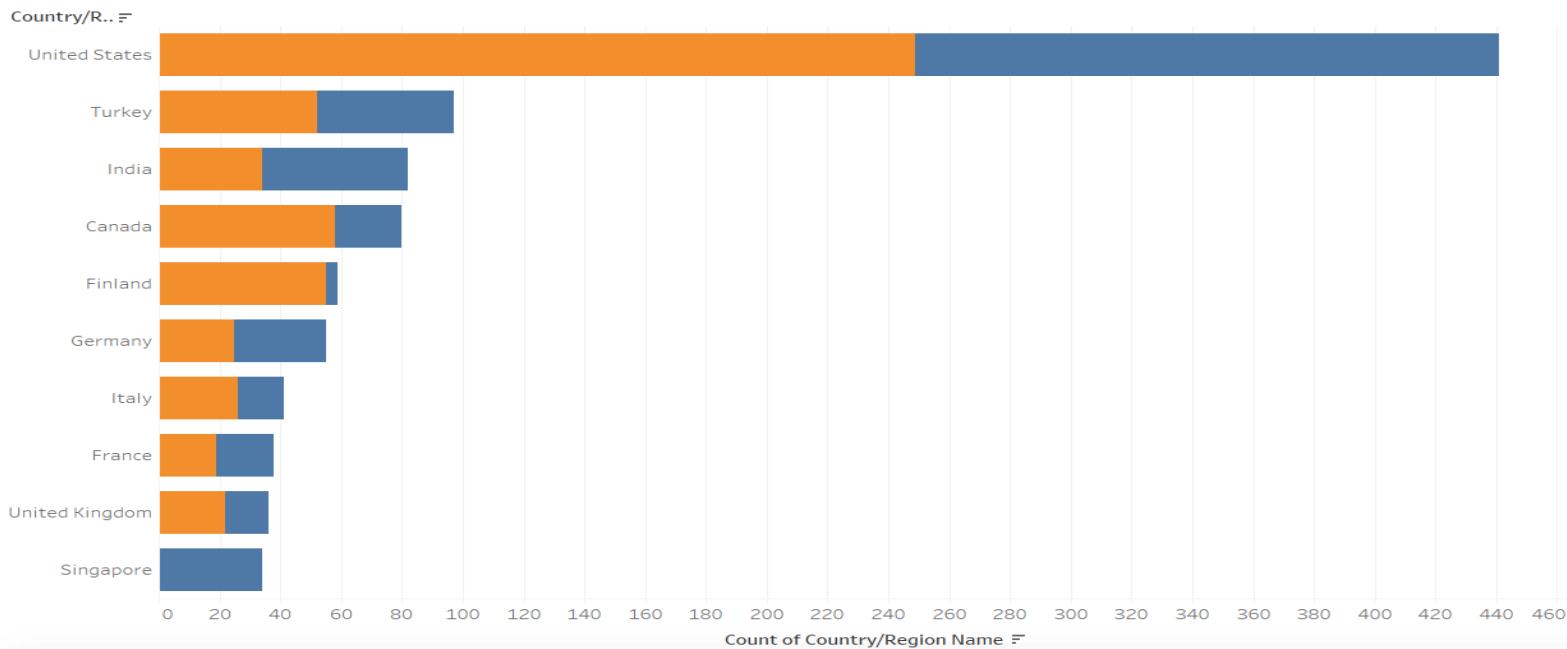


Top 5 Countries by Registration

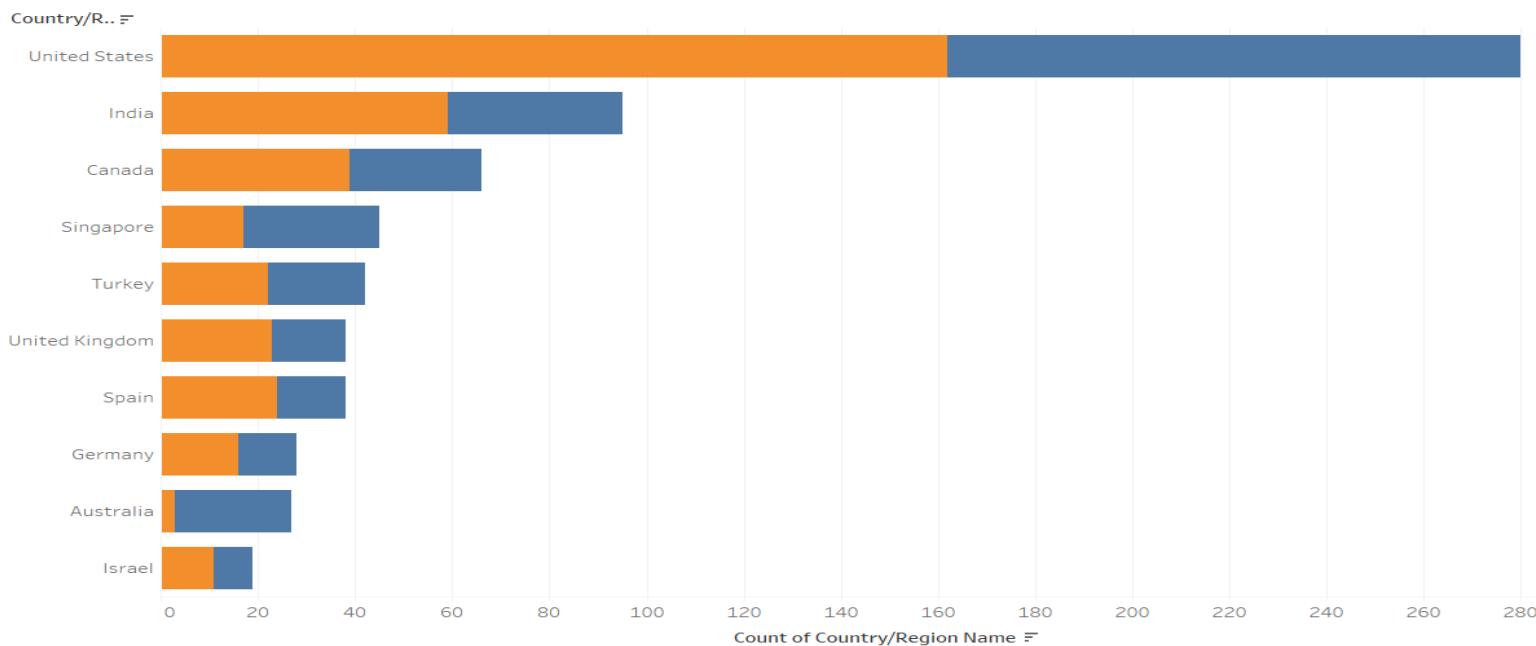
Top 5 countries World wide map



Top 10 countries with most number of attendees 1Q

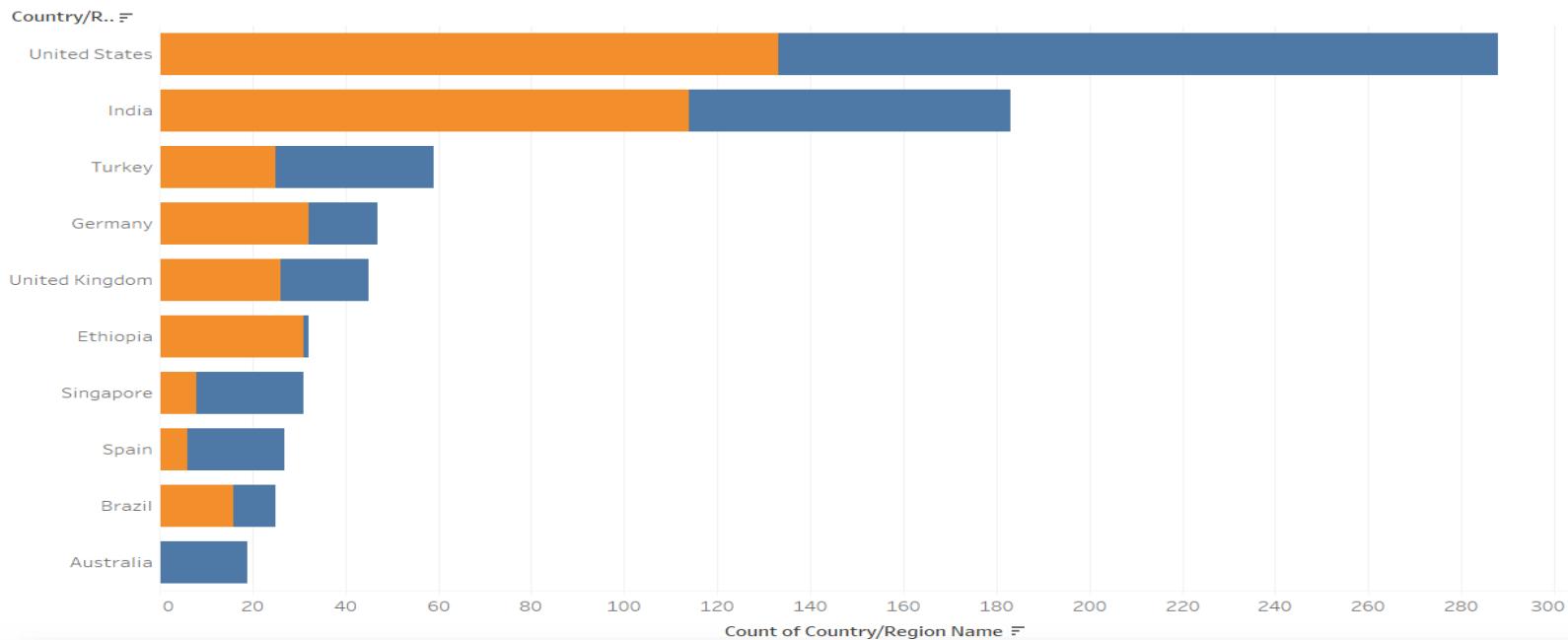


Top 10 countries with most number of attendees 2Q

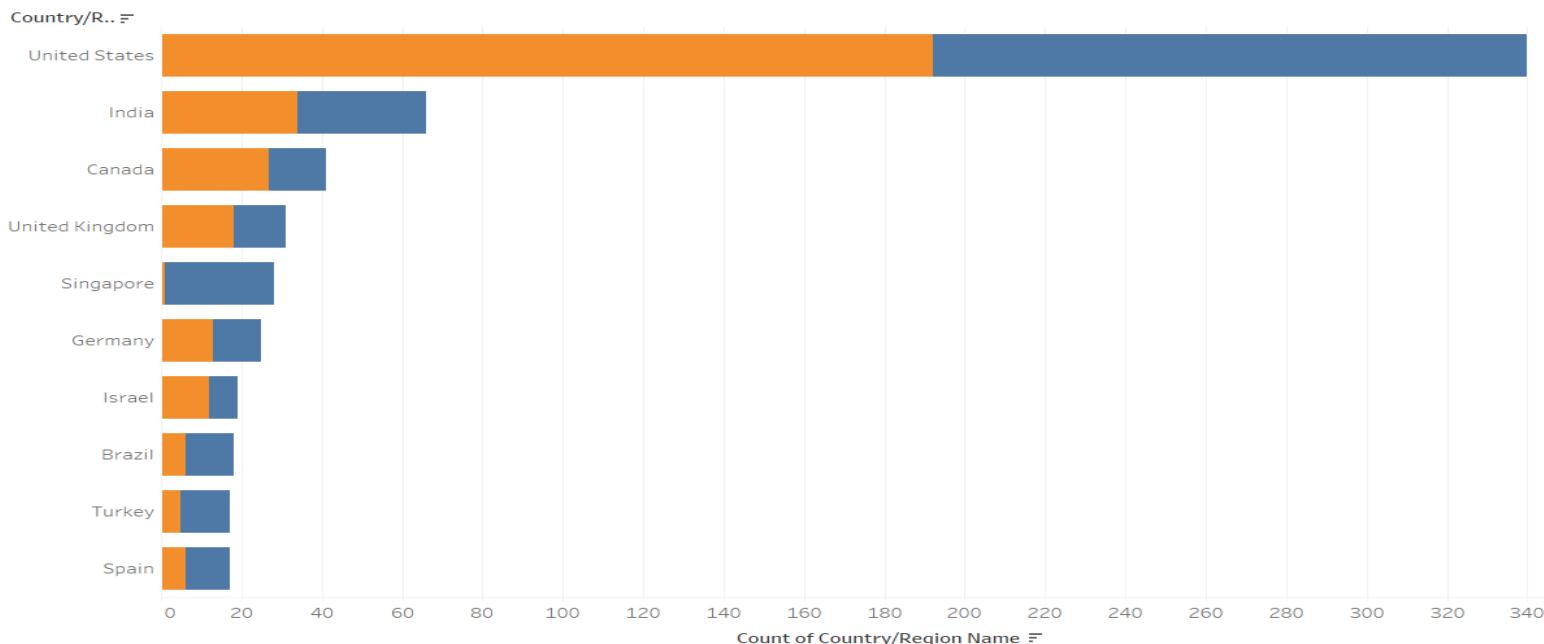


APPENDIX

Top 10 countries with most number of attendees 3Q

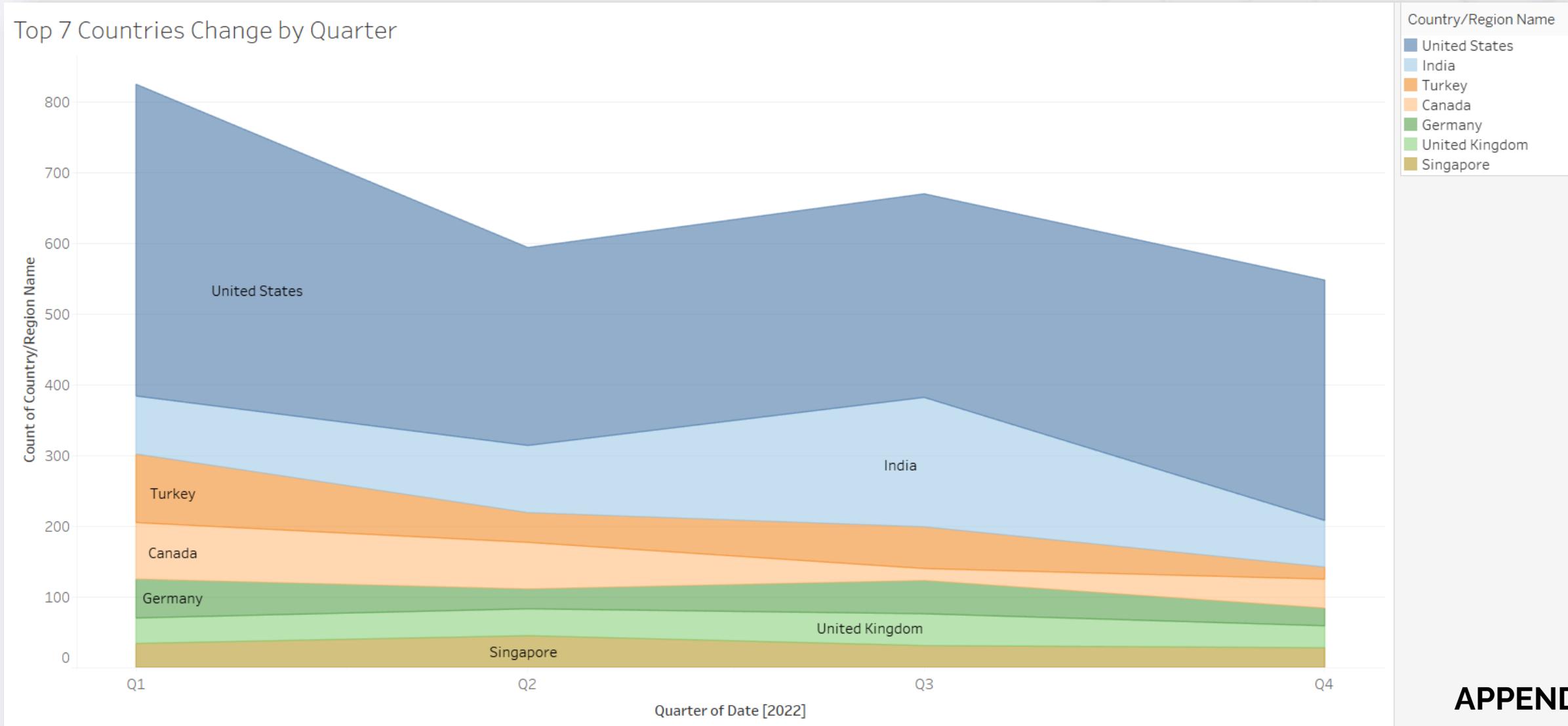


Top 10 countries with most number of attendees 4Q



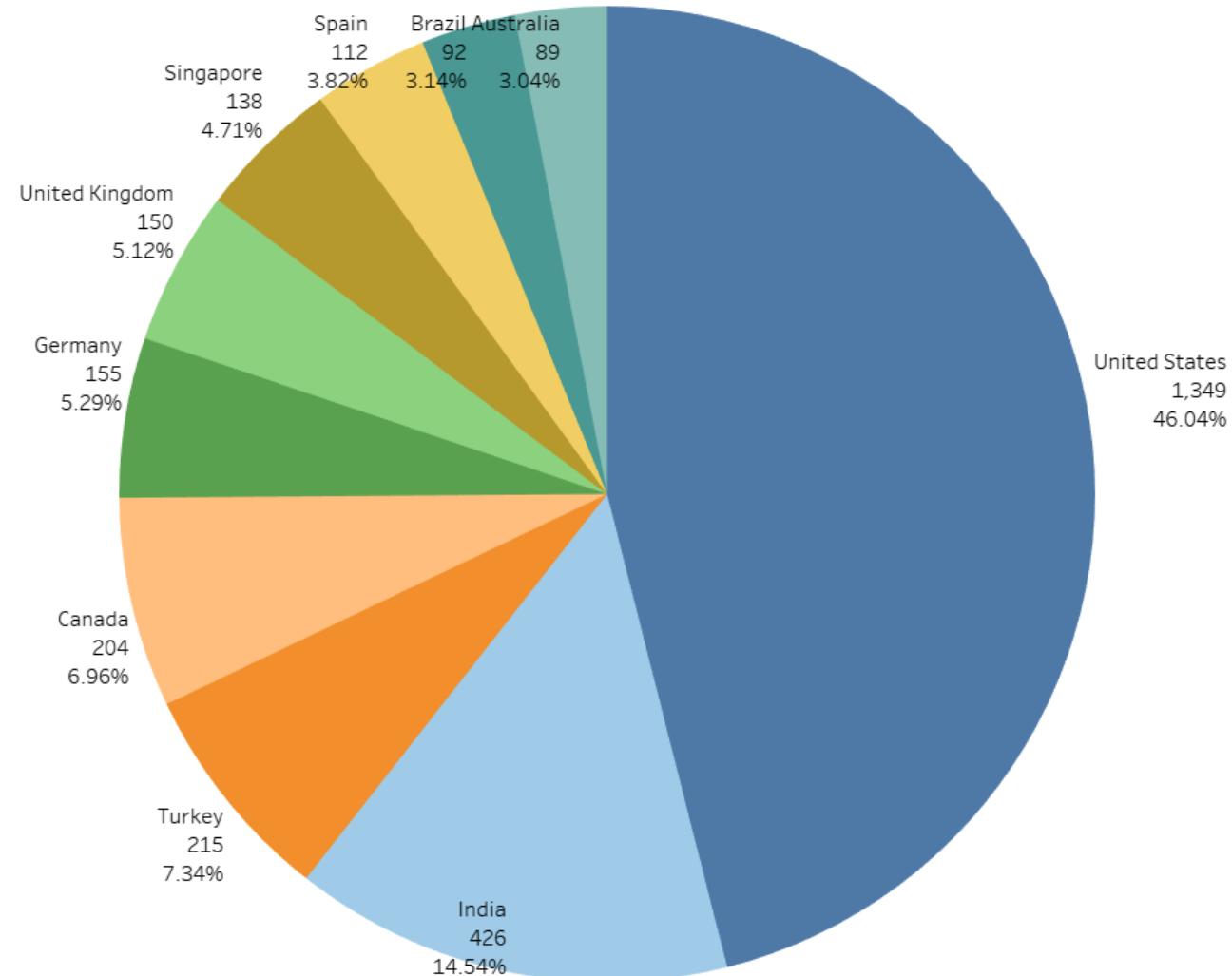
APPENDIX

Top 7 Countries by quarter



Pie Chart of Top 10 Countries

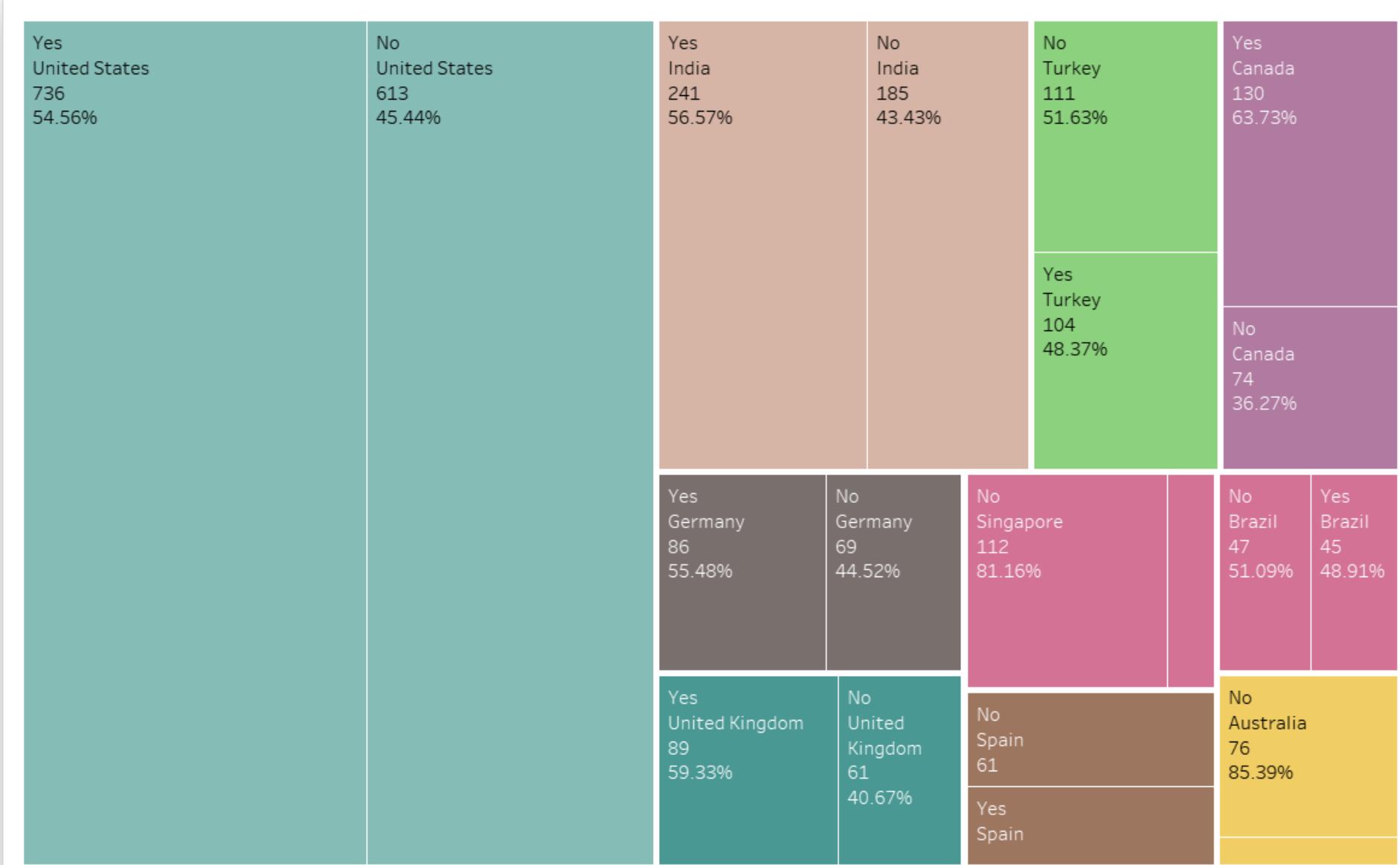
Top 10 Countries and Percentage of Total Registration



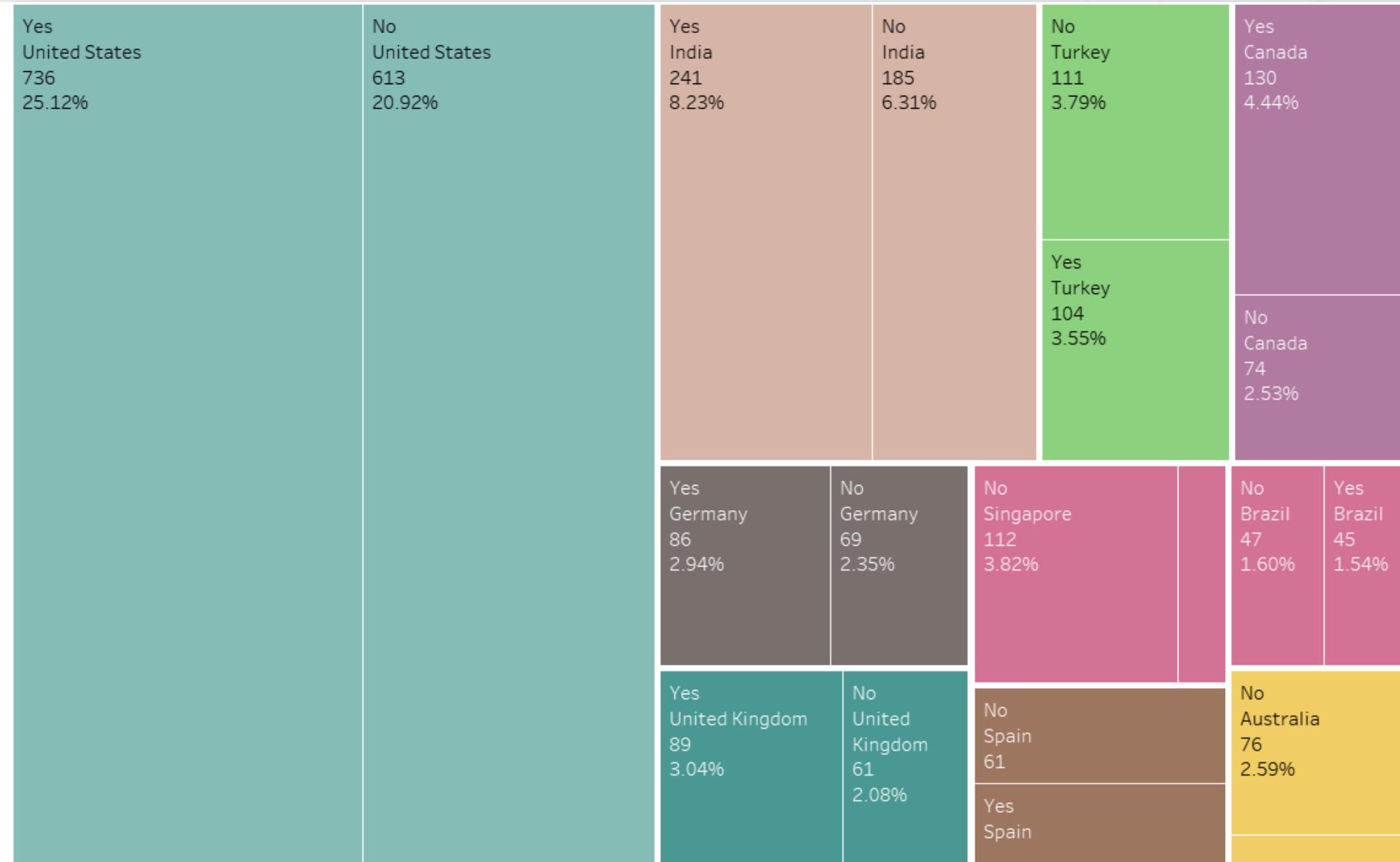
| Country/Region | CNT(Country) |
|----------------|--------------|
| United States | 1,349 |

APPENDIX

Top 10 countries & percentage of attendee



Top 10 countries & percentage of Total attendee



APPENDIX

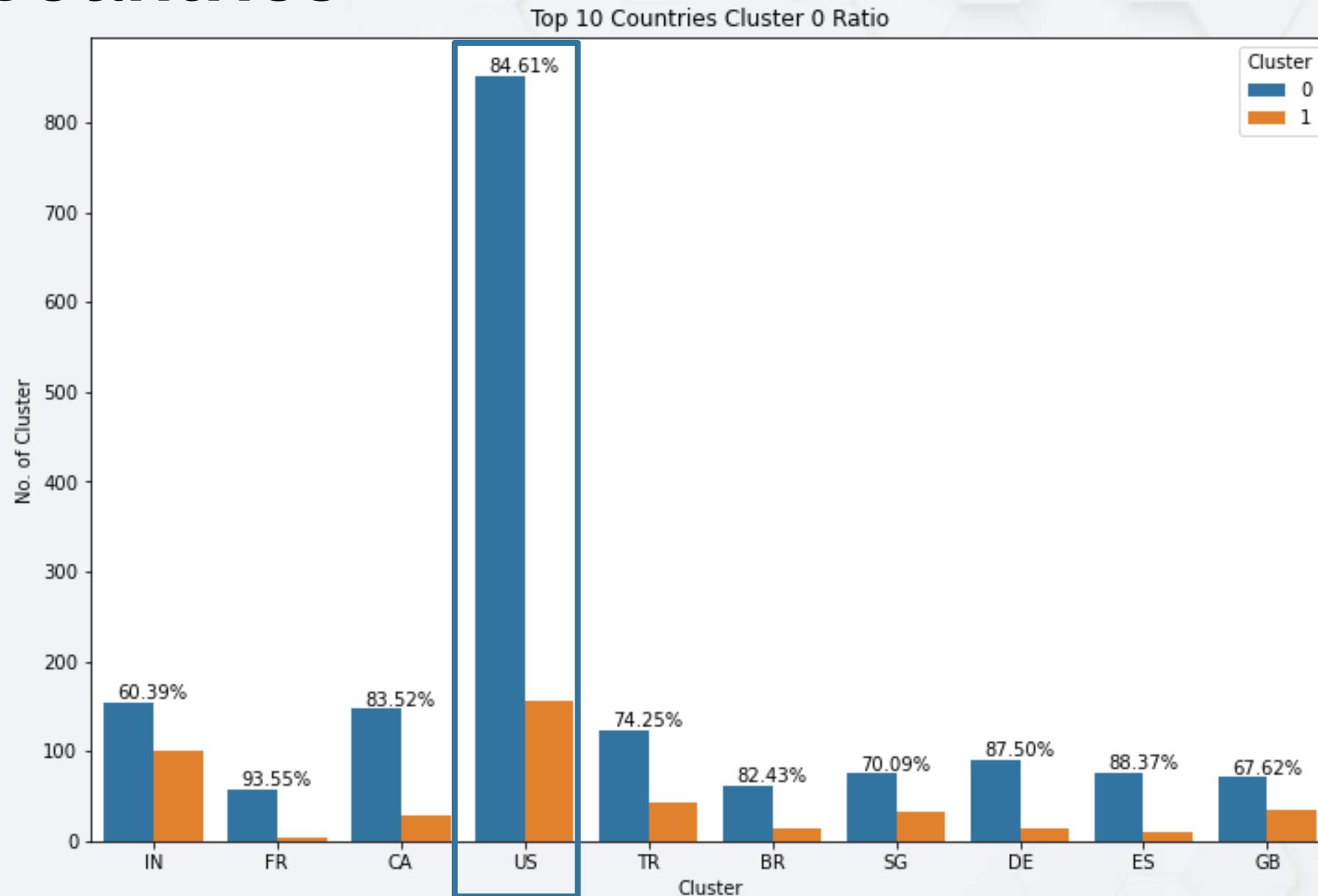
Appendix 2

Clustering Visualization

Attended Yes or No

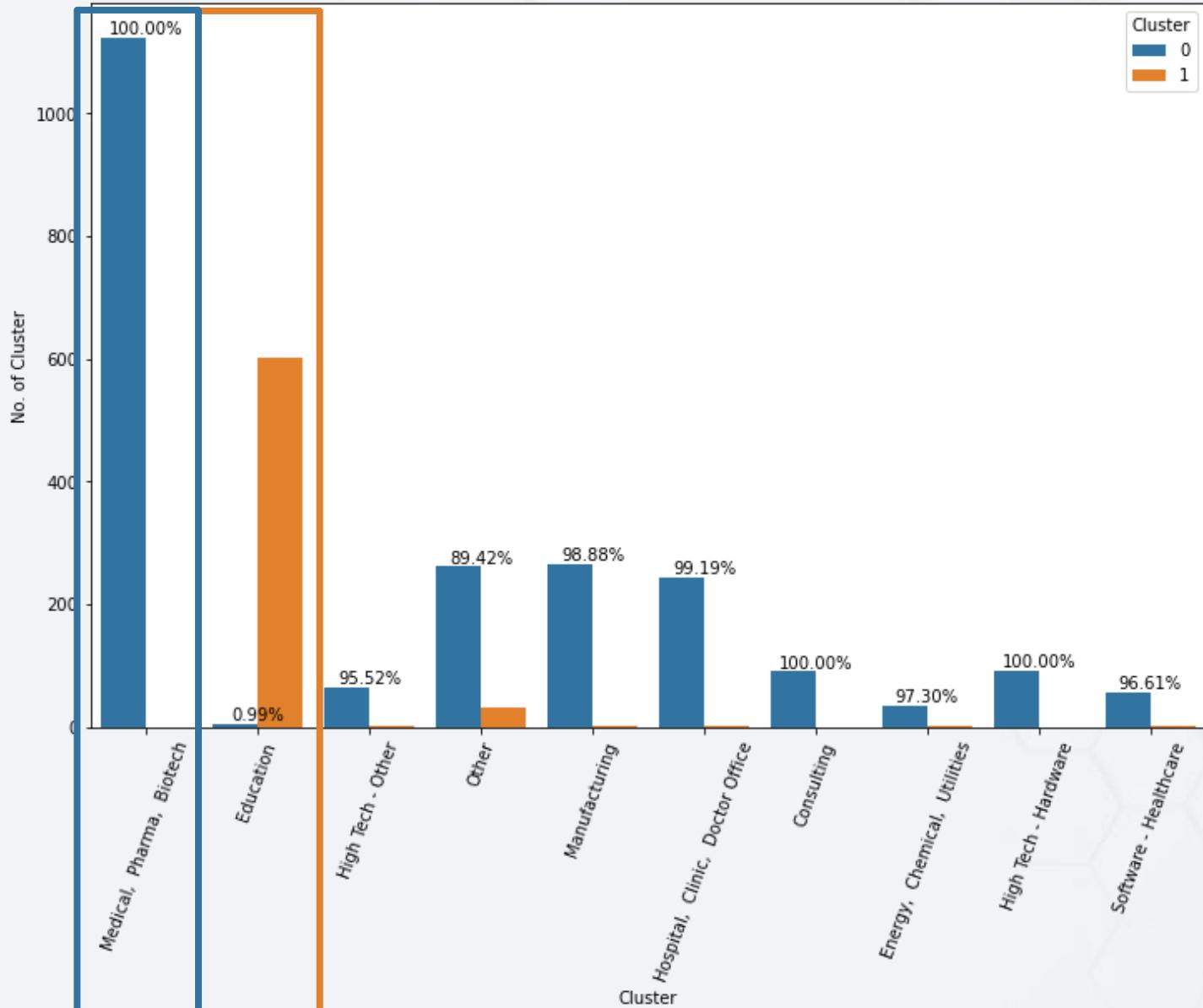


By Countries

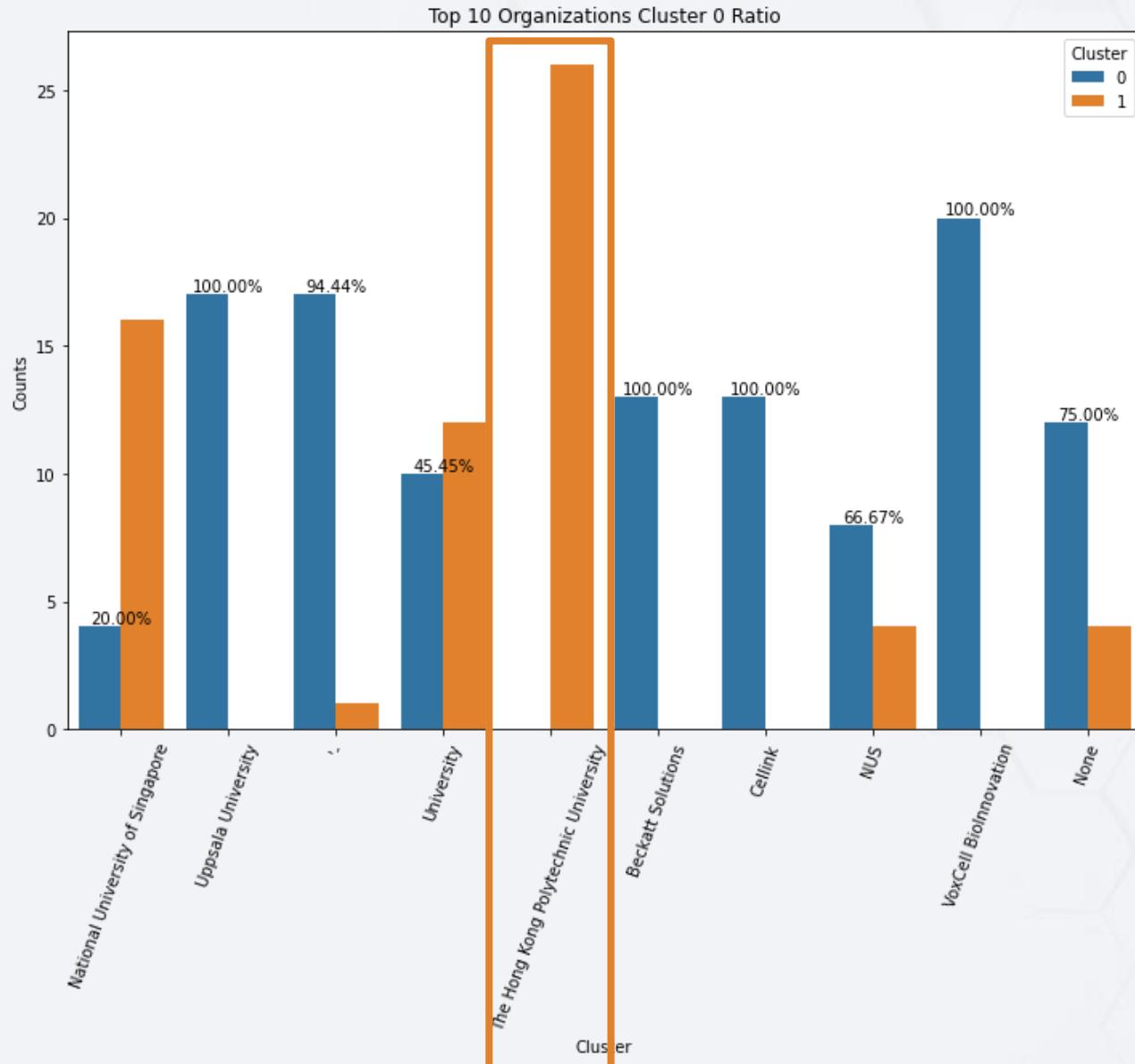


By Industries

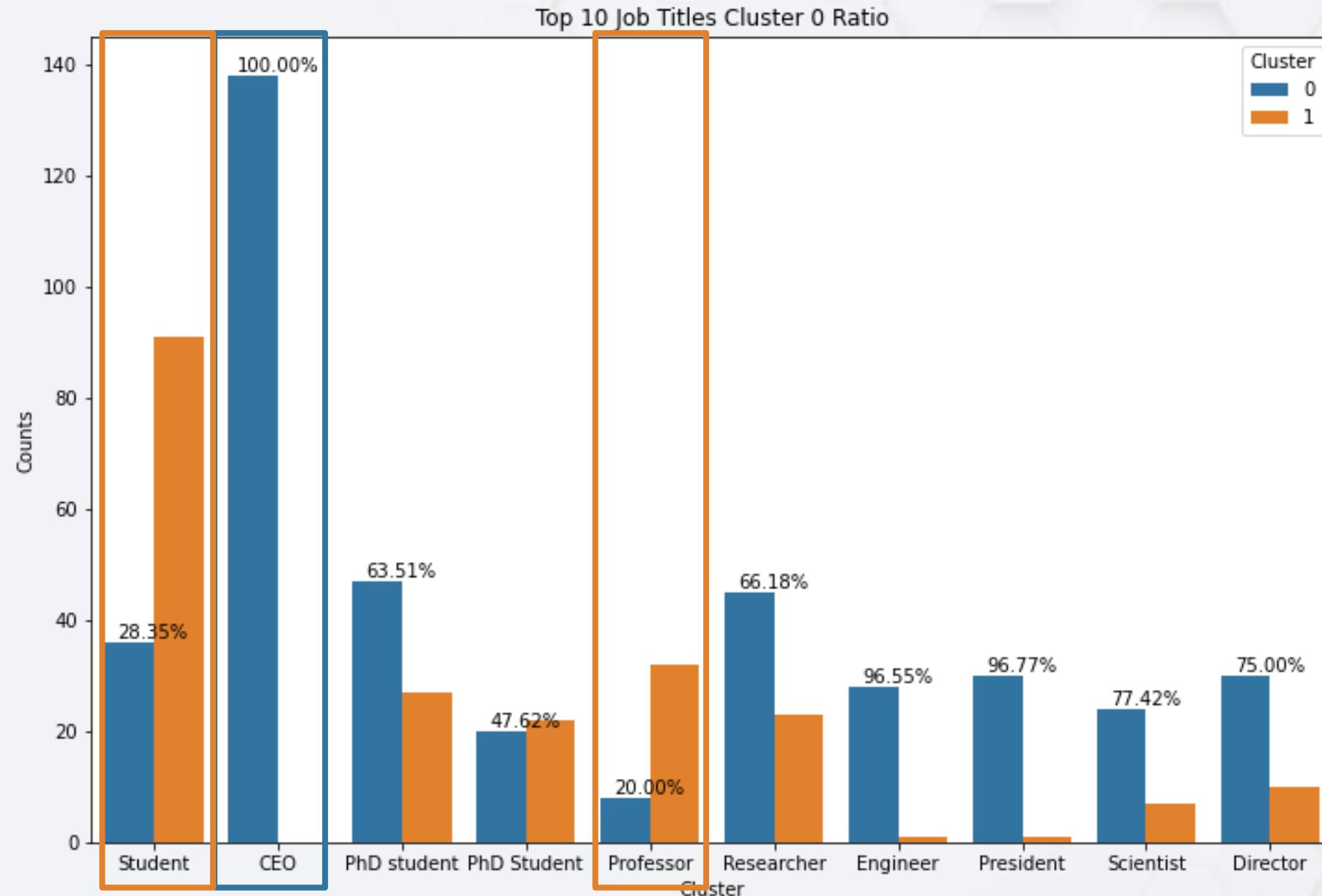
Top 10 Industries Cluster



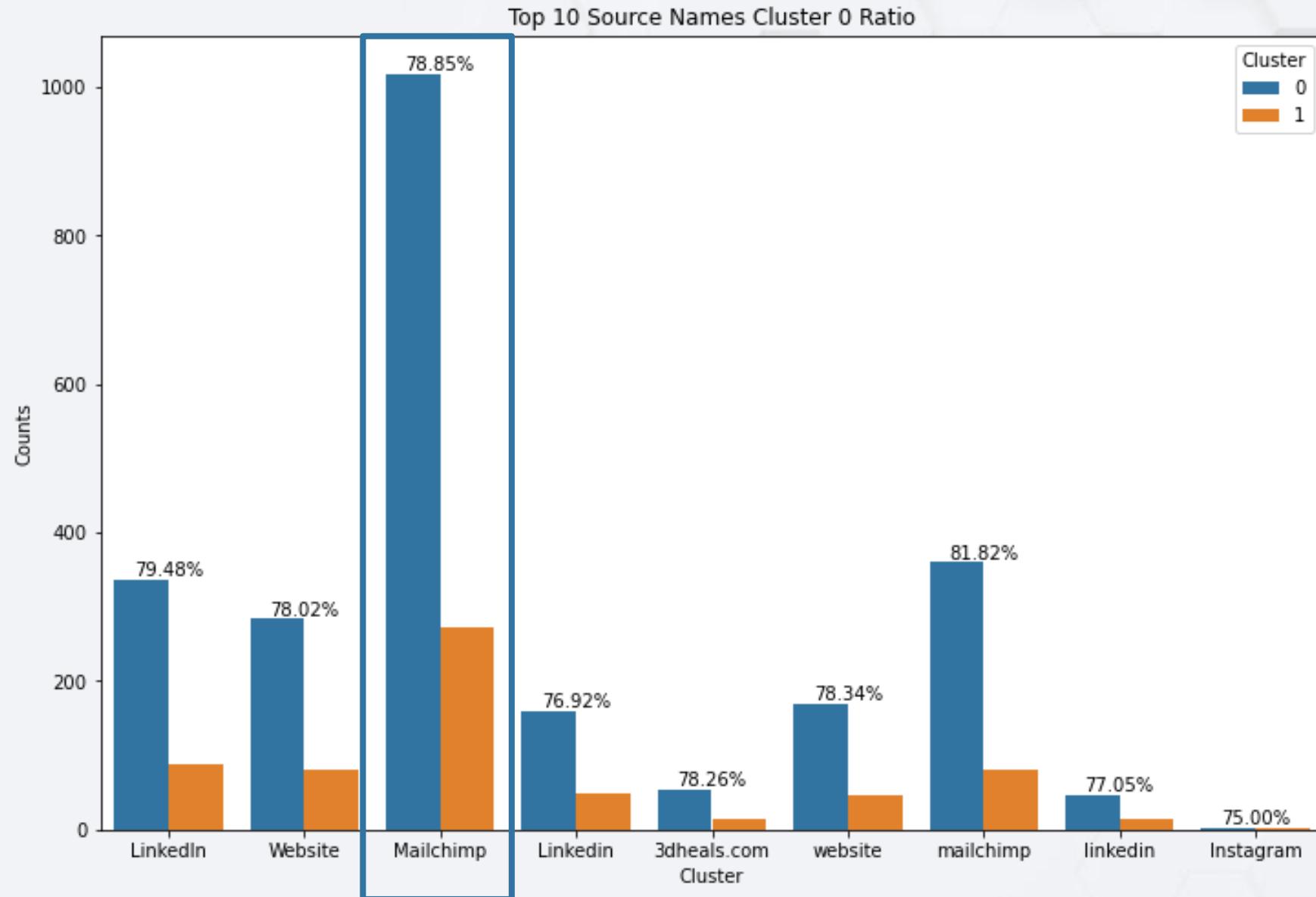
By Organizations



By Job Titles



By Source Name



By Domain_name

Top 10 domains Cluster 0 Ratio

