

**For [i], where i is number, please kindly refer to Appendix.*

Key insights from web-scraped data

The bar chart[1] on education levels reveals that master's degrees dominate job postings, appearing in over 700 instances, followed by bachelor's degrees at approximately 300. Doctoral degrees and high school diplomas are mentioned fewer than 50 times each. This trend emphasizes the need for graduate-level education, justifying the focus on a master's program to meet employer expectations for advanced expertise.

The pie chart[2] on remote job opportunities shows that 84.1% of roles require on-site presence, while only 15.9% are remote. This highlights the importance of preparing students for hybrid and in-office roles, with curriculum elements focusing on collaborative tools and workplace adaptability to address this imbalance.

In the bar chart[3] of skills in FAANG job postings, SQL is the most in-demand skill, with over 60 mentions, followed by modeling and cloud computing at nearly the same level. Leadership, Python, and big data rank slightly lower but are still prominent. These insights emphasize the need for courses in database management, cloud computing, and leadership to prepare students for competitive positions.

The frequency chart[4] for hard skills highlights C++ as the most frequently mentioned programming language, appearing over 17,000 times, followed by Python and SQL with around 10,000 mentions each. TensorFlow and PyTorch lag behind with fewer than 5,000 mentions each. These numbers underline the necessity of focusing on widely used programming languages and tools to ensure students are well-prepared for technical roles in the data science job market.

Hierarchical clustering implementation

The dendrogram [5] provides a hierarchical clustering of skills, which informs the development of a comprehensive course curriculum by grouping related topics. However, manual adjustments were necessary to ensure logical course structure, skill progression, and balanced coverage without repetition. The curriculum is as attached [6] For instance, foundational programming skills like Python, SQL, and MATLAB, originally spread across clusters, were grouped into a single course, "Programming for Applied AI," to consolidate technical instruction. Similarly, advanced technical skills like TensorFlow, PyTorch, and deep learning were combined under "Advanced Machine Learning and AI" for focused expertise. Business-related clusters such as leadership, project management, and strategic thinking were grouped into "Leadership and Strategic Management" to provide cohesive soft skill development. Visualization and storytelling tools, essential for communicating data insights, were combined into "Data Visualization and Storytelling." These adjustments avoided redundancy while ensuring logical skill alignment, creating a curriculum with 8-12 courses that comprehensively address the technical, business, and soft skills needed for successful careers in data science and AI.

K-means clustering implementation

Elbow method graph [7] shows the optimal value of k is 4, with the k value and clustering result [8][9], the course curriculum has been thoughtfully structured to align with the distinct skill clusters identified[10]. For example, Semester 1 emphasizes foundational technical skills like Python, SQL, and MATLAB, paired with essential analytical concepts such as statistics and modeling. These align with Cluster 0, which focuses on programming and analytics. Semester 2 introduces tools for data visualization (Tableau, Power BI) and cloud computing (Hadoop, Spark), reflecting Cluster 1's orientation towards applied data science and infrastructure. Semester 3 progresses to advanced machine learning frameworks (TensorFlow, PyTorch) and strategic leadership, addressing both technical mastery and managerial insights. Finally, Semester 4 integrates advanced business intelligence, ethics, and a capstone project, ensuring holistic skill application and alignment with

real-world industry needs. This curriculum mirrors the clustering insights to systematically build technical and strategic competencies. The labeled scatter plot is also attached [11].

Interpretation of results using ChatGPT API

Please refer to [12] for prompt 1 and [13] for prompt 2 for details.

Prompt 1 response [12]:

The Master of Business and Management in Data Science and Artificial Intelligence (MBM-DSAI) at the University of Toronto is designed for ambitious professionals who seek to master the intersection of business strategy, advanced analytics, and AI-driven decision-making. This comprehensive six-semester program equips students with a unique blend of technical expertise, business acumen, and leadership skills to excel in a data-centric world.

Through a carefully curated curriculum, students will gain hands-on experience in cutting-edge tools like Python, Tableau, TensorFlow, and cloud platforms such as AWS, while also learning strategic market analysis, ethical AI practices, and project management. The program's structure ensures a seamless integration of technical knowledge—including machine learning, big data engineering, and natural language processing—with business intelligence and consulting skills essential for leadership roles.

A highlight of the program is the Capstone Project, where students apply their learning to solve real-world business challenges, demonstrating their readiness to lead data-driven transformation across industries. Whether aspiring to be a data scientist, AI strategist, or business leader, the MBM-DSAI positions graduates as trailblazers in the fast-evolving AI landscape.

Unlock your potential to lead with innovation—enroll today and shape the future of business and AI.

Prompt 2 response [13]:

The skill clusters in data science and AI curricula emphasize both technical and soft skills essential for well-rounded professional development. Cluster 1 focuses on foundational and advanced technical abilities, such as Python and SQL for data management and machine learning, while also integrating soft skills like leadership and communication to foster adaptability. Cluster 2 dives into the frequency and industry relevance of technical tools like Matlab and Java, showcasing their utility across varying experience levels and industries.

A key similarity between the two clusters lies in their alignment around technical programming utility. Skills like Python and SQL, foundational to data manipulation and model building, are pivotal in both clusters. These tools form the backbone of data science operations, from basic queries to complex machine learning pipelines. The focus on skill popularity by industry further validates their applicability, reinforcing their role as universal assets for data-driven roles.

Soft skills like teamwork and communication in Cluster 1 complement these technical capabilities by enabling professionals to lead projects and collaborate effectively. Leadership qualities enhance the adaptability required for dynamic AI-driven environments, as highlighted by adaptability metrics in Cluster 2.

Together, these clusters create a cohesive educational structure where technical mastery supports functional expertise, and soft skills enhance cross-disciplinary collaboration. This synergy fosters graduates capable of excelling in data science and AI roles, driving innovation and strategic impact.

Discussion and final course curriculum

Comparing Section 3 (raw hierarchical clustering) and Section 4 (second clustering algorithm),

Section 4 is better suited for designing the final course curriculum due to its logical grouping of skills, minimizing redundancy, and ensuring cohesive course design.

Section 3 fragments related topics into isolated clusters, making curriculum development complex and less intuitive. For instance:

- Technical skills like Python, SQL, TensorFlow, and PyTorch are separated, requiring manual reorganization to form coherent courses.
- Core business skills such as Negotiation, Risk Management, and Project Management are scattered, complicating the creation of integrated business and management courses.
- Leadership, Communication, and Teamwork are treated as separate entities, reducing the coherence of leadership-focused content.

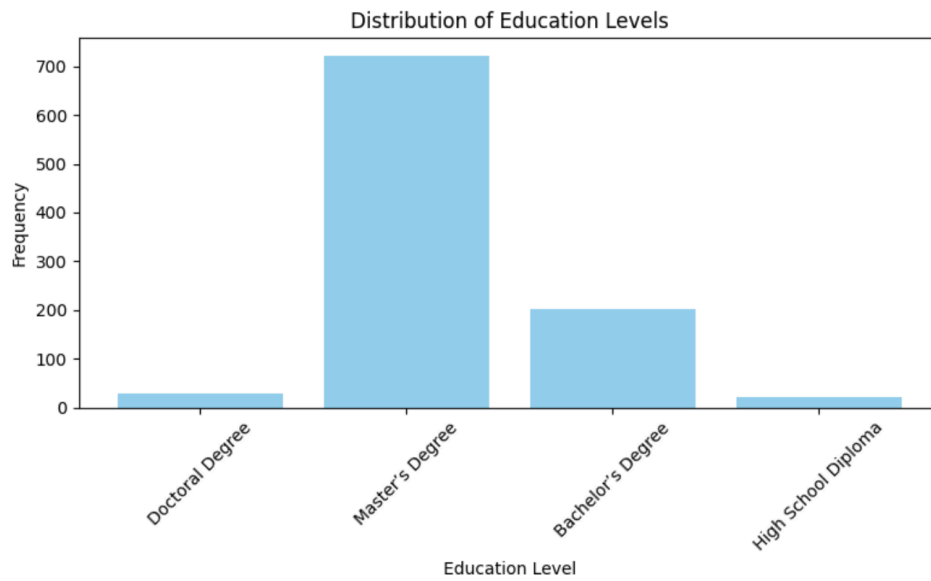
In contrast, Section 4 effectively groups related skills into cohesive, practical courses, streamlining curriculum development:

- SQL, Big Data, and Cloud Computing are grouped under a single course for technical depth.
- Leadership, Team Management, and Strategic Thinking are aligned into a comprehensive leadership course.
- Advanced Machine Learning naturally integrates TensorFlow, PyTorch, and deep learning, avoiding the disjointed structure seen in Section 3.

The curriculum from Section 4 provides eight balanced courses that emphasize both technical and business skills, ensuring clarity and alignment with industry requirements. Section 3, by comparison, would require significant restructuring to achieve the same level of organization.

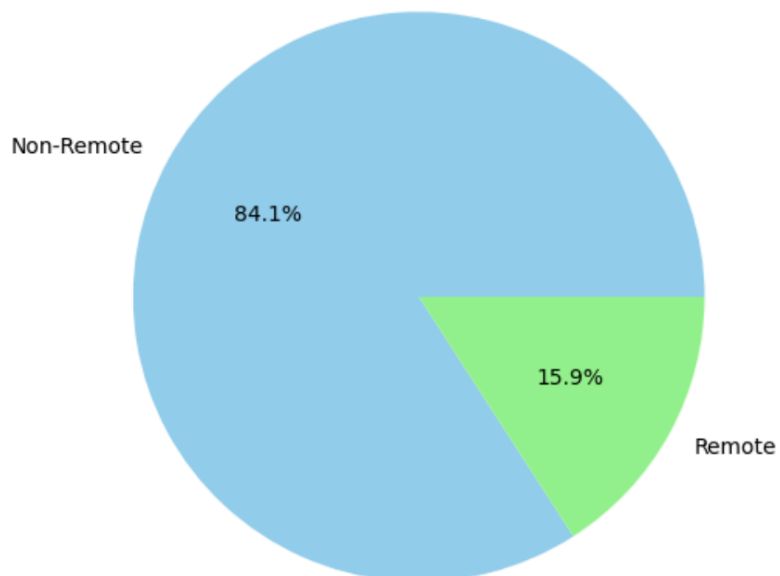
Appendix

[1]

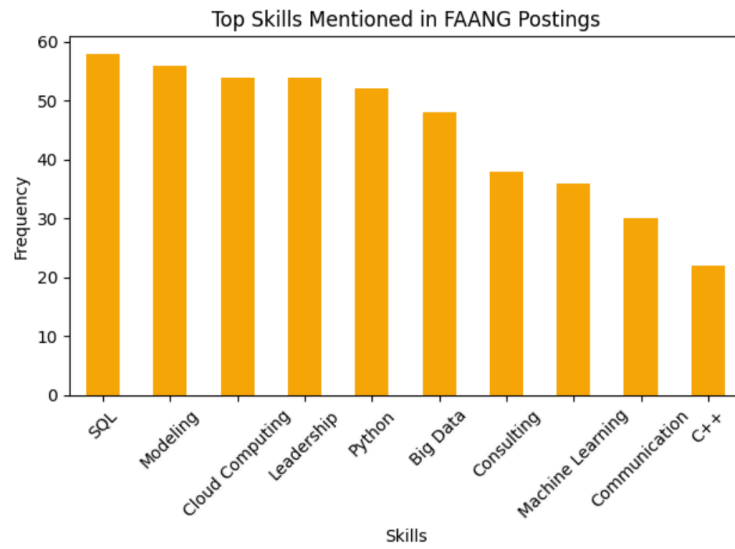


[2]

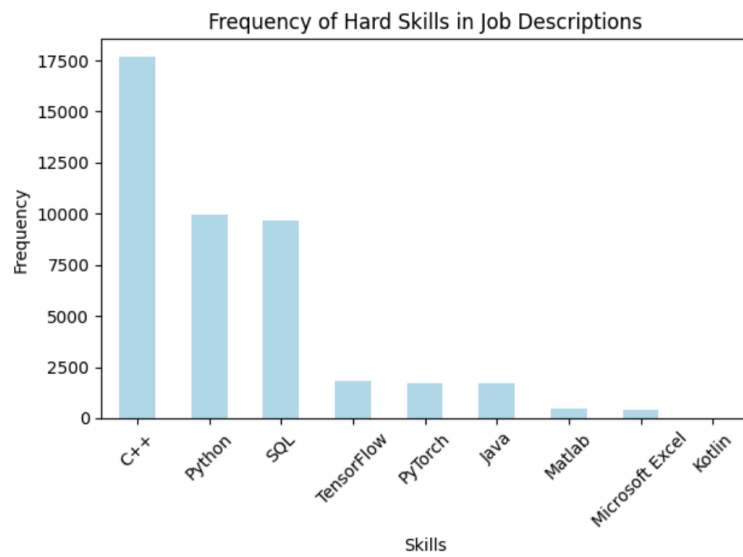
Percentage of Remote Jobs (Based on Descriptions and Location)



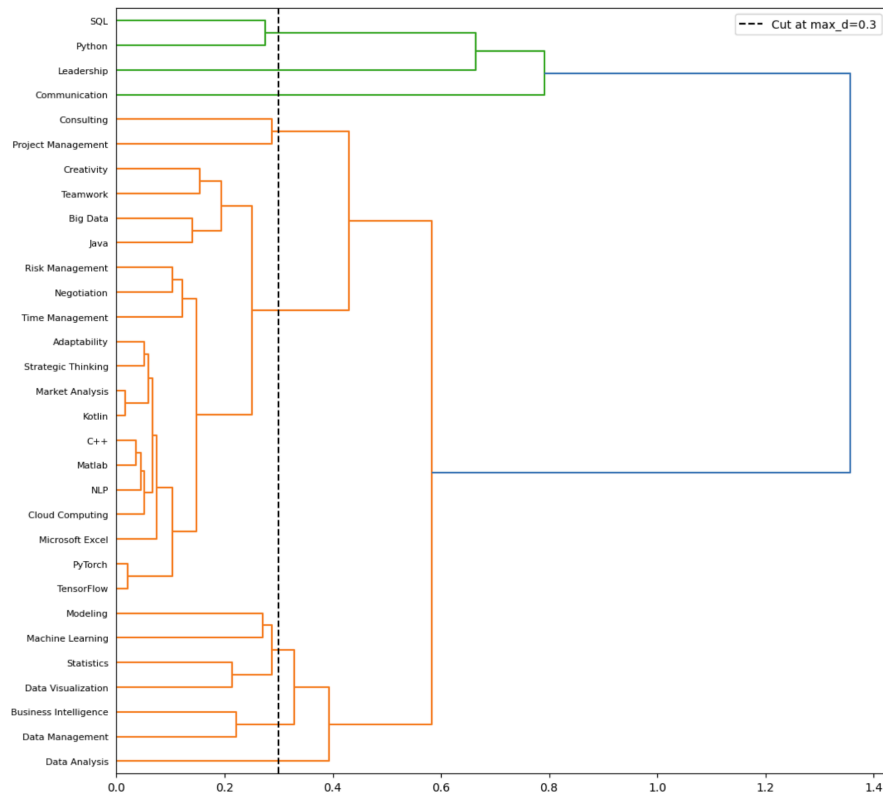
[3]



[4]



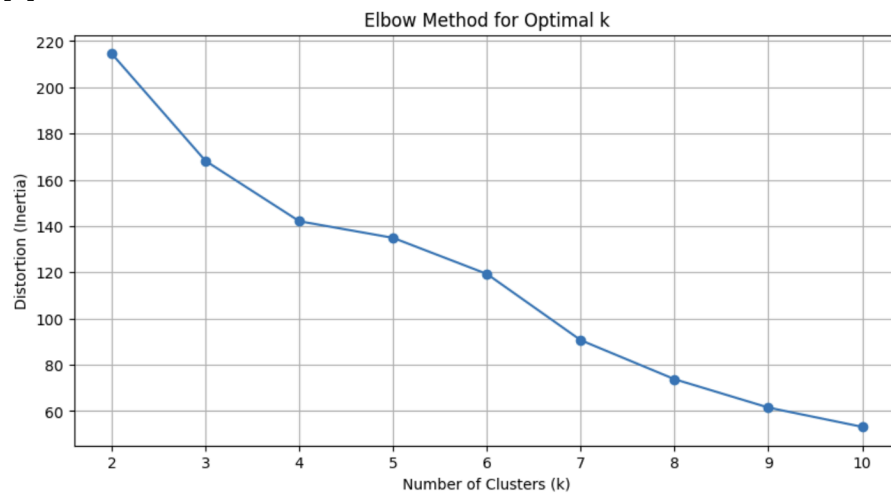
[5]



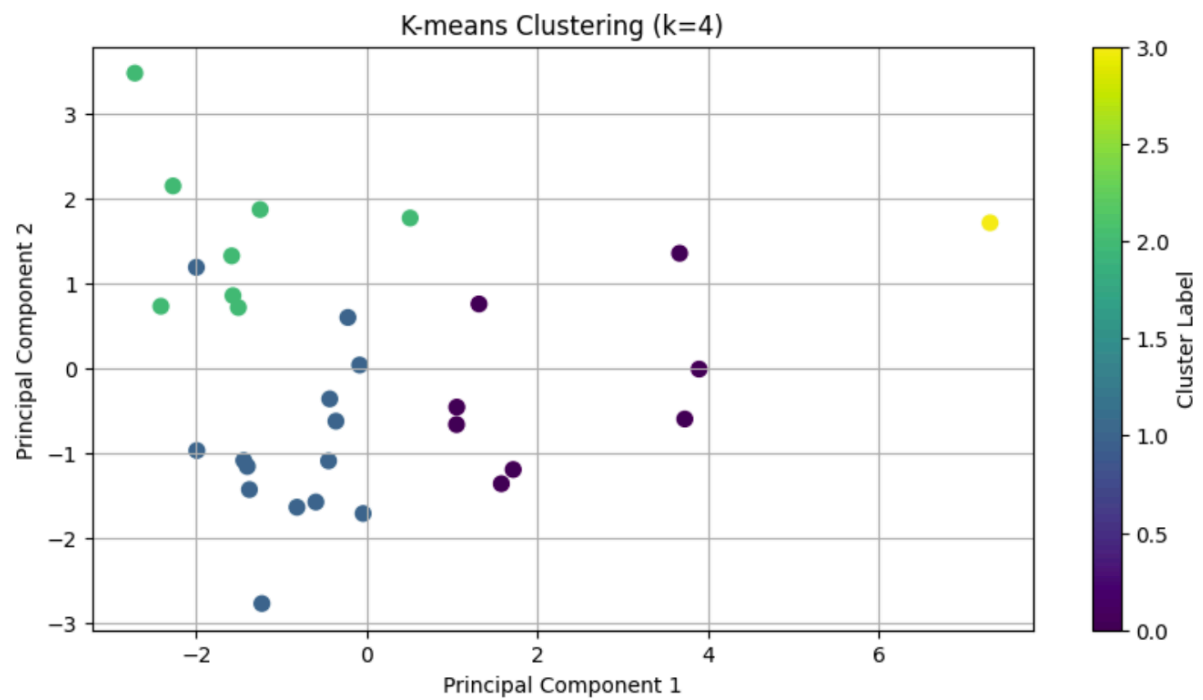
[6]

```
section_3_curriculum_sequence = {
  "Semester 1": [
    {"Course": "Foundations of Data Management and Business Intelligence", "Topics": ["Data Storage and Retrieval", "SQL", "Business Intelligence Tools"]},
    {"Course": "Programming for Data Science", "Topics": ["Python", "Advanced SQL", "Version Control (Git)"]},
  ],
  "Semester 2": [
    {"Course": "Statistical Foundations and Data Analysis", "Topics": ["Descriptive and Inferential Statistics", "Hypothesis Testing", "Exploratory Data Analysis"]},
    {"Course": "Data Visualization and Storytelling", "Topics": ["Tableau", "Power BI", "Data Storytelling Techniques"]},
  ],
  "Semester 3": [
    {"Course": "Machine Learning and Predictive Modeling", "Topics": ["Regression and Classification", "Clustering", "Model Evaluation and Tuning"]},
    {"Course": "Business Strategy and Market Analysis", "Topics": ["Strategic Thinking", "Market Analysis", "Risk Management"]},
  ],
  "Semester 4": [
    {"Course": "Advanced Machine Learning and AI", "Topics": ["TensorFlow", "PyTorch", "Natural Language Processing"]},
    {"Course": "Cloud Computing and Big Data Engineering", "Topics": ["AWS/Azure/GCP", "Hadoop", "Big Data Processing"]},
  ],
  "Semester 5": [
    {"Course": "Leadership and Team Management", "Topics": ["Leadership Skills", "Team Collaboration", "Project Management"]},
    {"Course": "Ethics and Governance in AI", "Topics": ["Ethical AI", "Data Privacy", "Regulatory Compliance"]},
  ],
  "Semester 6": [
    {"Course": "Business Intelligence and Consulting", "Topics": ["Consulting Techniques", "Strategy Implementation", "Negotiation"]},
    {"Course": "Capstone Project", "Topics": ["Integrated Application of All Skills"]},
  ],
}
```

[7]



[8]



[9] (* only first few rows of record is shown, please check the code for detailed)

| | Frequency | Avg_Education_Level | Pct_Remote | Avg_Experience_Years | \ |
|--------|-----------|---------------------|------------|----------------------|---|
| Python | 2.354833 | -0.179191 | 0.210339 | -0.569834 | |
| Matlab | -0.804471 | -1.844441 | -0.392659 | 0.010397 | |
| Kotlin | -0.964707 | -0.119137 | -1.900762 | -0.170541 | |
| Java | -0.574801 | -0.220543 | -1.039973 | -0.786300 | |
| C++ | -0.572130 | -0.572526 | 0.082684 | -0.470513 | |

| | Cooccurrence | Unique_Titles | Job_Mention_Diversity | \ |
|--------|--------------|---------------|-----------------------|---|
| Python | -2.354833 | 1.029669 | 1.029669 | |
| Matlab | 0.804471 | -0.830390 | -0.830390 | |
| Kotlin | 0.964707 | -0.914938 | -0.914938 | |
| Java | 0.574801 | -0.514744 | -0.514744 | |
| C++ | 0.572130 | -0.187824 | -0.187824 | |

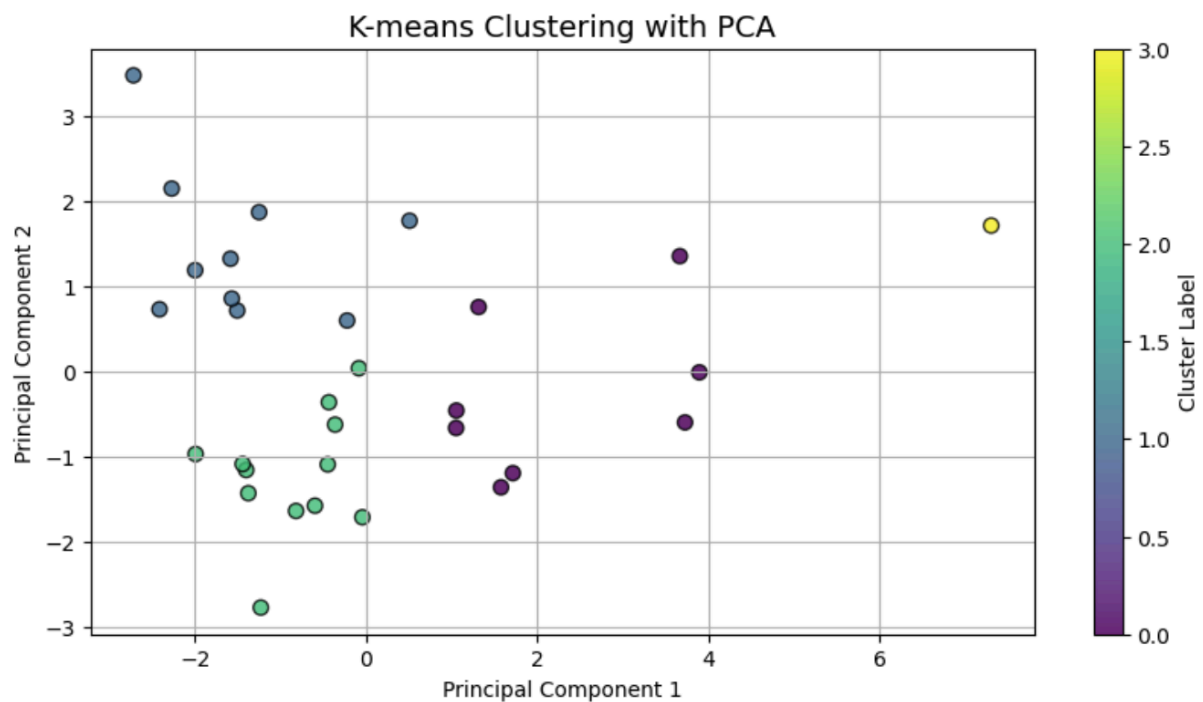
| | Skill_Popularity_By_Industry | Unique_Locations | Industry_Mentions | \ |
|--------|------------------------------|------------------|-------------------|---|
| Python | -0.964160 | 0.998092 | -0.964160 | |
| Matlab | -1.316990 | -0.953133 | -1.316990 | |
| Kotlin | -0.275625 | -1.125691 | -0.275625 | |
| Java | -0.952446 | -0.608018 | -0.952446 | |
| C++ | -0.589492 | -0.130167 | -0.589492 | |

| | Cluster |
|--------|---------|
| Python | 0 |
| Matlab | 1 |
| Kotlin | 1 |
| Java | 1 |
| C++ | 1 |

[10]

```
K_mean_clustering_curriculum_sequence = {
    "Semester 1": [
        {"Course": "Programming for Applied AI", "Topics": ["Python", "SQL", "MATLAB"]},
        {"Course": "Statistics and Advanced Analytics", "Topics": ["Statistics", "Modeling", "Optimization", "NLP"]},
    ],
    "Semester 2": [
        {"Course": "Data Visualization and Storytelling", "Topics": ["Tableau", "Power BI", "Data Storytelling"]},
        {"Course": "Big Data and Cloud Computing", "Topics": ["Hadoop", "Spark", "Cloud Computing"]},
    ],
    "Semester 3": [
        {"Course": "Advanced Machine Learning and AI", "Topics": ["TensorFlow", "PyTorch", "Deep Learning"]},
        {"Course": "Leadership and Strategic Management", "Topics": ["Leadership", "Project Management", "Strategic Thinking"]},
    ],
    "Semester 4": [
        {"Course": "Business Intelligence and Consulting", "Topics": ["Business Intelligence", "Consulting", "Strategy Implementation"]},
        {"Course": "Ethics and Governance in AI", "Topics": ["Ethical AI", "Data Privacy", "Regulatory Compliance"]},
        {"Course": "Capstone Project", "Topics": ["Integrated Application of All Skills"]},
    ],
}
```



[11]



[12]

"You are an expert education consultant. Write a compelling yet concise description for below new master's program in around 200 words, 'Master of Business and Management in Data Science and Artificial Intelligence,' offered by the University of Toronto. Highlight how the program combines technical expertise, business acumen, and leadership skills to entice potential students to enrol in this program. Overall word count must less than 220.

< 4/5 >

 **Master of Business and Management in Data Science and Artificial Intelligence**
University of Toronto

The Master of Business and Management in Data Science and Artificial Intelligence (MBM-DSAI) program at the University of Toronto equips aspiring leaders with a unique blend of technical expertise, business acumen, and leadership skills. Designed to address the growing demand for data-driven decision-makers, this cutting-edge program bridges the gap between AI innovation and real-world business challenges.

Through a rigorous curriculum, students gain proficiency in advanced data analytics, machine learning, and AI technologies while mastering essential business strategies such as market analysis, financial decision-making, and organizational leadership. The program emphasizes hands-on learning, offering experiential projects and industry partnerships that prepare graduates to tackle complex challenges across diverse sectors.

What sets the MBM-DSAI apart is its holistic approach. Students are trained not only to harness the power of AI but also to lead cross-functional teams and drive ethical, impactful decisions in an increasingly data-centric world. With access to world-class faculty and Toronto's vibrant AI and tech ecosystem, graduates are uniquely positioned to excel in roles such as data science managers, AI consultants, and innovation leaders.

Take the next step in your career by joining a program that transforms technical skills into strategic impact. Enroll in the MBM-DSAI and redefine the future of business and AI.

In 220 words, Analyze the skill clusters developed for a curriculum in data science and AI. Highlight similarities within 2 clusters, such as how foundational skills (e.g., Python, SQL) align in terms of technical programming utility, or how soft skills (e.g., teamwork, communication) contribute to leadership and adaptability. Discuss how the 2 clusters naturally complement each other to create a cohesive educational structure that supports professional development in data science and AI roles."

Cluster 1:

1 ['Data Management', 'Business Intelligence']

2 ['Data Visualization', 'Statistics', 'Machine Learning', 'Modeling']

3 ['Data Analysis']

4 ['TensorFlow', 'PyTorch', 'Microsoft Excel', 'Cloud Computing', 'NLP', 'Matlab', 'C++', 'Kotlin', 'Market Analysis', 'Strategic Thinking', 'Adaptability', 'Time Management', 'Negotiation', 'Risk Management', 'Java', 'Big Data', 'Teamwork', 'Creativity']

5 ['Project Management', 'Consulting']

6 ['Python', 'SQL']

7 ['Leadership']

8 ['Communication']

Cluster 2:

| | Frequency | Avg_Education_Level | Pct_Remote | |
|------------------------|-----------|---------------------|------------|-----------|
| Avg_Experience_Years \ | | | | |
| Python | 2.354833 | -0.179191 | 0.210339 | -0.569834 |
| Matlab | -0.804471 | -1.844441 | -0.392659 | 0.010397 |
| Kotlin | -0.964707 | -0.119137 | -1.900762 | -0.170541 |
| Java | -0.574801 | -0.220543 | -1.039973 | -0.786300 |
| C++ | -0.572130 | -0.572526 | 0.082684 | -0.470513 |

| | Skill_Popularity_By_Industry | | Unique_Locations | |
|---------------------|------------------------------|-----------|------------------|--|
| Industry_Mentions \ | | | | |
| Python | -0.964160 | 0.998092 | -0.964160 | |
| Matlab | -1.316990 | -0.953133 | -1.316990 | |
| Kotlin | -0.275625 | -1.125691 | -0.275625 | |
| Java | -0.952446 | -0.608018 | -0.952446 | |
| C++ | -0.589492 | -0.130167 | -0.589492 | |

| | Cluster |
|--------|---------|
| Python | 0 |
| Matlab | 1 |
| Kotlin | 1 |
| Java | 1 |
| C++ | 1 |

The skill clusters in data science and AI curricula emphasize both technical and soft skills essential for well-rounded professional development. Cluster 1 focuses on foundational and advanced technical abilities, such as **Python** and **SQL** for data management and machine learning, while also integrating soft skills like **leadership** and **communication** to foster adaptability. Cluster 2 dives into the frequency and industry relevance of technical tools like **Matlab** and **Java**, showcasing their utility across varying experience levels and industries.

A key similarity between the two clusters lies in their alignment around **technical programming utility**. Skills like **Python** and **SQL**, foundational to data manipulation and model building, are pivotal in both clusters. These tools form the backbone of data science operations, from basic queries to complex machine learning pipelines. The focus on **skill popularity by industry** further validates their applicability, reinforcing their role as universal assets for data-driven roles.

Soft skills like **teamwork** and **communication** in Cluster 1 complement these technical capabilities by enabling professionals to lead projects and collaborate effectively. Leadership qualities enhance the adaptability required for dynamic AI-driven environments, as highlighted by adaptability metrics in Cluster 2.

Together, these clusters create a cohesive educational structure where technical mastery supports functional expertise, and soft skills enhance cross-disciplinary collaboration. This synergy fosters graduates capable of excelling in data science and AI roles, driving innovation and strategic impact.