1. Data processing, feature engineering, and visualization:

Appendix 1 illustrates that Python, SQL, and cloud platforms are the three most commonly used programming languages. Appendix 2 identifies modeling, data management, and machine learning as the top technical skills based on job postings. The pie chart in Appendix 3 reveals that project management is the most dominant skill in business intelligence and project management, followed by consulting, business intelligence, and negotiation. Additionally, Appendix 4 presents the distribution of teamwork and communication skills, highlighting creativity as the most prevalent, followed by communication, teamwork, leadership, and adaptability.

2. Hierarchical clustering implementation:

The hierarchical clustering dendrogram, displayed in the appendices (Appendix 5), informed the structure of the proposed curriculum. Based on this analysis, the curriculum is organized into eight distinct courses, each tailored to develop specific skills and address particular areas of expertise:

Advanced Programming and Systems covers C++, Java, and Linux/Unix, essential for systems programming and low-level software development. **Computational Tools and Simulation** includes MATLAB, Simulation, and Excel, integral for scientific computation and applications in engineering and finance. **Data Analytics and Visualization** focuses on Data Visualization, Business Intelligence, and R, equipping students for data analyst roles by enabling effective data analysis and insight generation. **Big Data and Machine Learning** combines Big Data, Machine Learning, and Optimization, addressing large-scale data processing and advanced analytical techniques. **Statistical Analysis and Modeling** utilizes Statistical Analysis, Modeling, and Python to provide a cohesive approach to data-driven decision-making. **Cloud Platforms and Data Management** encompasses Cloud Platforms, SQL, and Data Management, essential for managing and querying data in modern cloud-based systems. **Leadership and Communication in Teams** develops Teamwork, Consulting, Leadership, and Communication skills, vital for effective team and project management. Lastly, **Project Management and Adaptability** includes Project Management, Adaptability, and Creativity, preparing students to handle dynamic projects with innovation and flexibility.

Adjustments to the clustering results were made to balance course content and ensure clear focus areas. Technical skills such as C++, Java, and Linux/Unix were grouped to reflect their synergy, while data-related skills were combined to emphasize their complementary roles. Similar skills like Python, R, and MATLAB were distributed across different courses to avoid redundancy and broaden learning outcomes. Additionally, business-oriented skills like negotiation and business intelligence were logically grouped to highlight their relevance in professional settings. These adjustments ensure comprehensive skill coverage, maintain curriculum coherence, and align the courses with industry expectations and real-world job roles.

3. K-means or DBSCAN clustering implementation:

The K-means elbow graph (Appendix 6) and the scatterplot from the K-means clustering algorithm (Appendix 7) were instrumental in designing the proposed curriculum. Based on these analyses, the curriculum comprises eight courses, each tailored to specific skill sets.

Course 1: Data Management and Cloud Technologies covers SQL for data querying, introductions to cloud platforms like AWS, Azure, and Google Cloud for scalable computing, and best practices in data management for handling large datasets. **Course 2: Advanced Programming and Optimization Techniques** includes R for statistical computing and data visualization, C++ for high-performance computing and complex simulations, and optimization techniques for efficient problem-solving in business and data contexts. **Course 3: Systems Development and Adaptability** focuses on Java for enterprise-level object-oriented programming, Linux/Unix for system management and scripting, and adaptability training to excel in fast-changing tech environments. **Course 4: Business Intelligence and Negotiation** brings together advanced Excel functionalities for financial modeling, negotiation techniques for effective stakeholder communication, and business intelligence tools like Power BI and Tableau for informed decision-making.**Course 5: Consulting and Team Collaboration** emphasizes creating impactful data visualizations, consulting methodologies for problem-solving and client management, and teamwork skills to build effective collaboration and leadership abilities. **Course 6: Computational Modeling and Programming** incorporates Python for data analysis, machine learning, and automation; modeling techniques using mathematical and statistical approaches to simulate scenarios; and MATLAB for applied numerical computation in engineering and science problems. **Course 7: Big Data and Statistical Analysis** delves into simulation for predictive modeling, big data frameworks like Hadoop and Spark for large-scale data processing, and statistical analysis methods including inferential statistics and hypothesis testing. **Course 8: Artificial Intelligence and Communication** focuses on building supervised and unsupervised machine learning models, fostering creativity in AI system design, and developing communication strategies to present technical concepts to non-technical audiences.

Adjustments to the clustering results were made to enhance the curriculum's practicality and alignment with industry needs. To ensure focused learning, programming language clusters were divided so each course maintains a manageable scope while offering diverse skills. Logical grouping was applied by combining business-oriented skills like negotiation and business intelligence to highlight their professional applications. To avoid redundancy, similar skills such as Python, R, and MATLAB were distributed across different courses, broadening the learning outcomes. Lastly, the courses were adjusted to align with industry expectations and real-world job roles, ensuring that the curriculum meets practical professional demands.

4. Discussion and final course curriculum:

The final course curriculum was developed using the k-means clustering algorithm with an optimal number of ten clusters (k=10). This approach effectively grouped skills based on their similarities and logical relationships, ensuring that each course encompassed a diverse yet cohesive set of topics. To create a well-rounded curriculum, adjustments were made to eliminate redundancy and balance technical, business, and soft skills, aligning with the program’s

objective of preparing students for various roles in data science, analytics, and artificial intelligence. The k-means algorithm provided clear distinctions between clusters, facilitating logical groupings of technical skills such as programming languages and statistical methods alongside business-oriented skills like negotiation and business intelligence. However, minor manual adjustments were necessary to ensure that each cluster meaningfully contributed to the curriculum, particularly for those with fewer than three distinct skills.

The final course curriculum comprises eight specialized courses. **Data Management and Cloud Technologies** covers SQL for data querying, introductions to cloud platforms like AWS, Azure, and Google Cloud for scalable computing, and best practices in data management for handling large datasets. **Advanced Programming and Optimization Techniques** includes R for statistical computing and data visualization, C++ for high-performance computing and complex simulations, and optimization techniques for solving real-world business and data problems efficiently. **Systems Development and Adaptability** focuses on Java for enterprise-level object-oriented programming, Linux/Unix for system management and scripting, and adaptability training to excel in dynamic, fast-changing tech environments.

Business Intelligence and Negotiation encompasses advanced Excel functionalities for financial modeling, negotiation techniques for effective stakeholder communication, and business intelligence tools like Power BI and Tableau for informed decision-making. **Consulting and Team Collaboration** emphasizes data visualization, consulting methodologies for problem-solving and client management, and teamwork skills to build effective collaboration and leadership abilities. **Computational Modeling and Programming** incorporates Python for data analysis, machine learning, and automation; modeling techniques using mathematical and statistical approaches to simulate scenarios; and MATLAB for applied numerical computation in engineering and science problems. **Big Data and Statistical Analysis** delves into simulation for predictive modeling, big data frameworks like Hadoop and Spark for large-scale data processing, and statistical analysis methods including inferential statistics and hypothesis testing. Finally, **Artificial Intelligence and Communication** combines machine learning, creativity in AI system design, and communication strategies for presenting technical concepts to non-technical audiences.

The curriculum design is grounded in several key factors. It balances technical skills, such as programming and machine learning, with non-technical skills like communication and negotiation to meet industry demands. Each course covers unique aspects of data science and artificial intelligence, minimizing redundancy and ensuring comprehensive skill coverage. Industry relevance is achieved by aligning courses with real-world applications, such as integrating cloud technologies with data management and combining business intelligence with negotiation skills. Logical groupings of related skills prepare students for roles like data scientist, business analyst, and AI system designer. Additionally, smaller skill clusters were merged into comprehensive courses to maintain coherence and integrity. By leveraging k-means clustering and expert refinement, the final curriculum provides a robust educational foundation for students aspiring to excel in data science and AI fields.

Appendix

Appendix 1:

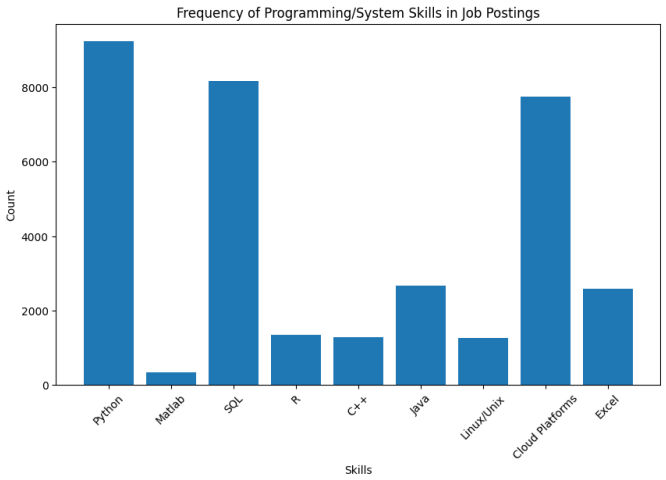

Figure 1: Frequency of Programming/System Skills in Job Postings

Appendix 2:

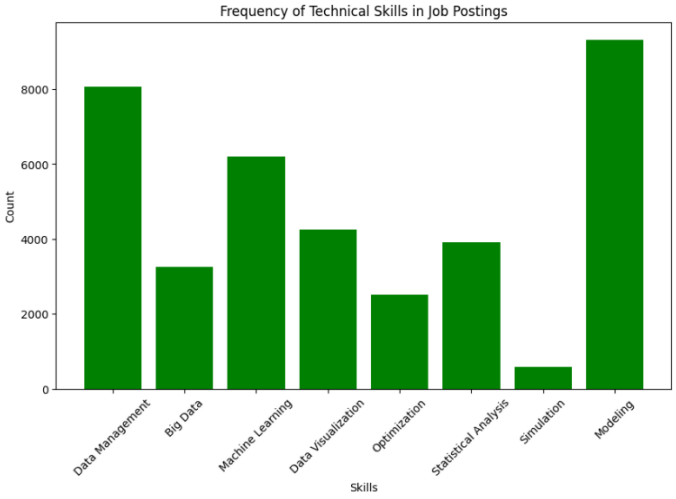

Figure 2: Frequency of Technical Skills in Job Postings

Appendix 3:

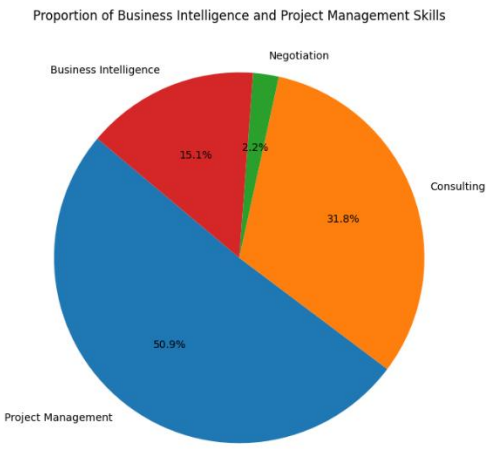

Figure 3: Proportion of Business Intelligence and Project Management Skills

Appendix 4:

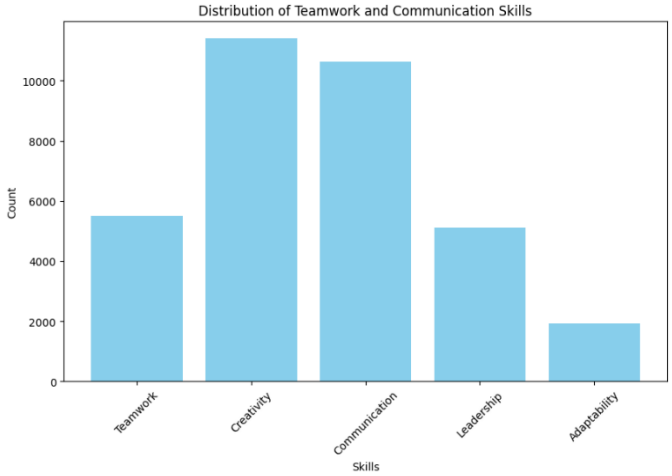

Figure 4: Distribution of Teamwork and Communication Skills

Appendix 5:

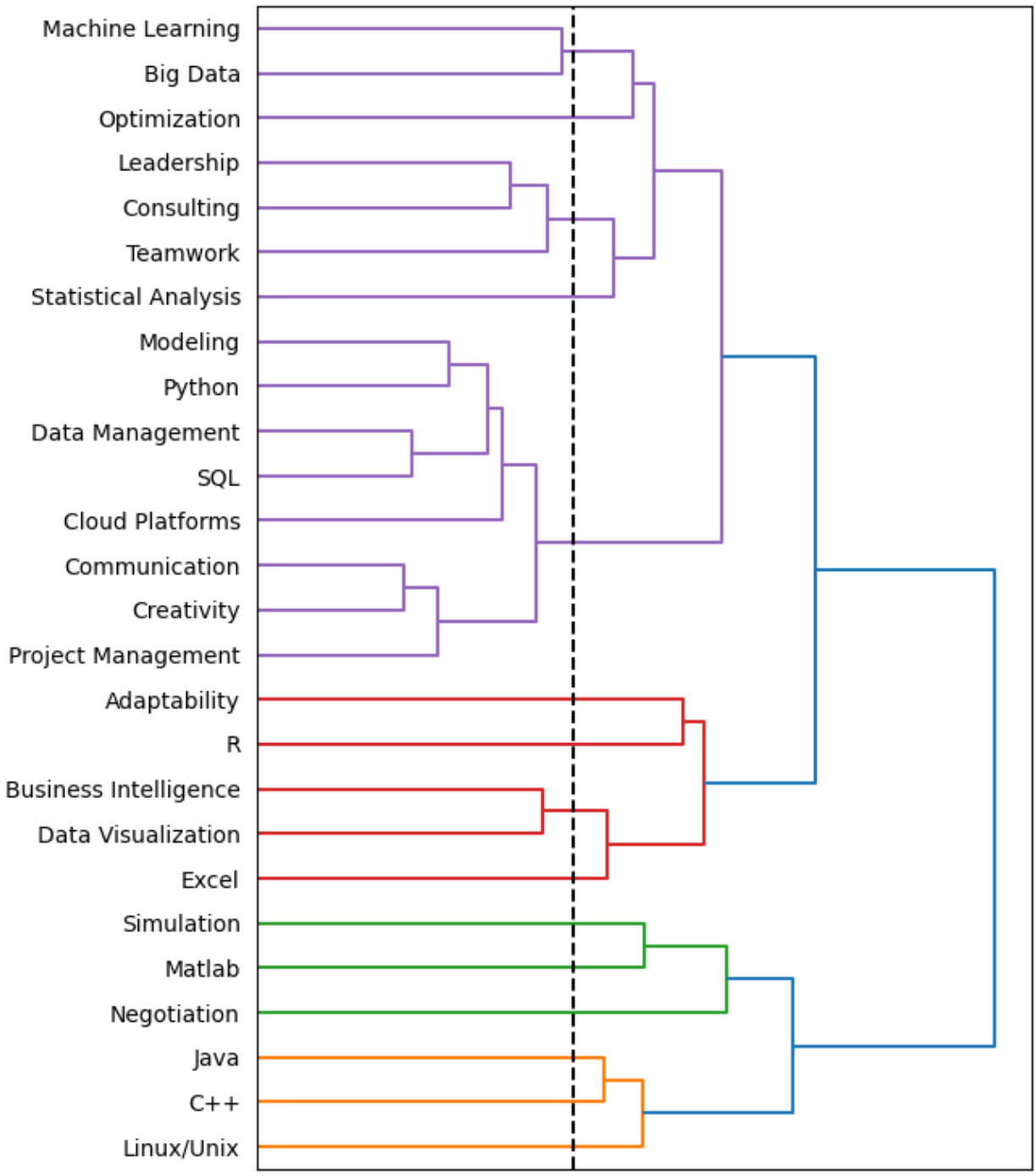

Figure 5: Hierarchical clustering – dendrogram

Appendix 6:

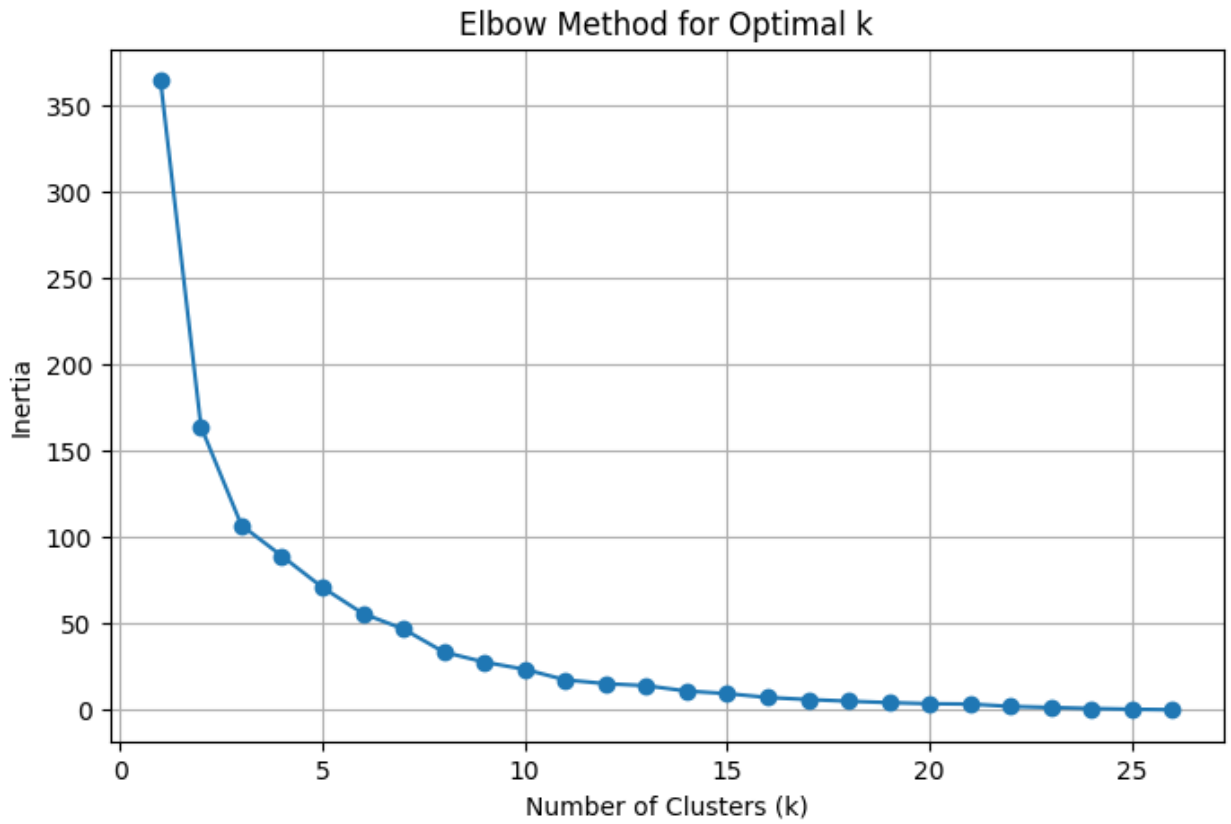

Figure 6: Elbow Method for Optimal k

Appendix 7:

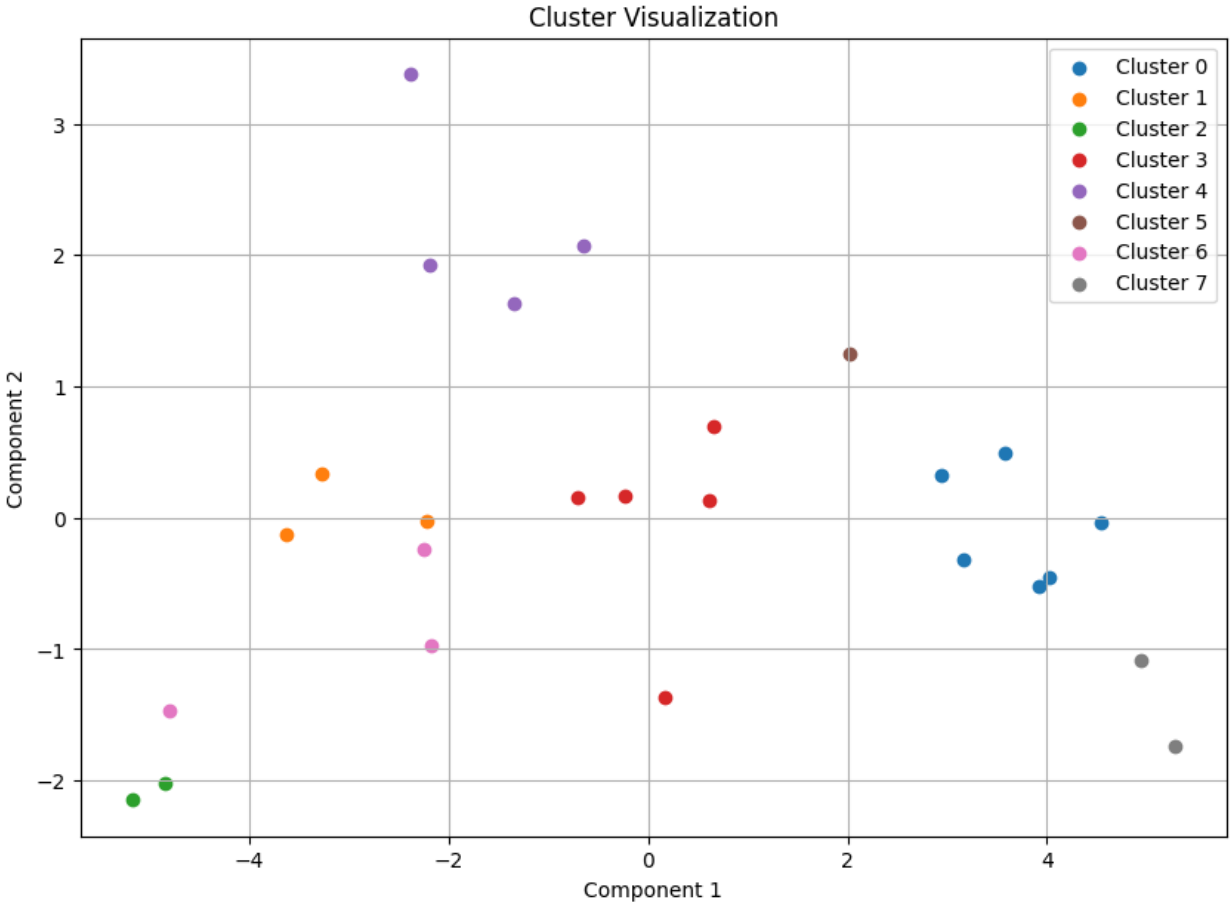

Figure 7: Cluster Visualization