

Robert Pavlik

D596 – The Data Analytics Journey

LFN1 Task 2: The Data Analytics Life Cycle

A - (1)(a)(2)

Data Engineer

- A Data Engineer focuses on the architecture and infrastructure of data systems. They have the responsibility of building pipelines that allow data to flow to different sources, such as data warehouses or lakes. They also work on large-scale data integration, ensuring the data is stored, processed, and made accessible efficiently.
- They play a critical role in the Data Acquisition and Data Cleaning phase of the life cycle. The data that they gather is stored and structured for later use. They also clean and standardize the data, ensuring the removal of inconsistencies and errors before it is passed on for analysis.
- They work with tools such as SQL, NoSQL, Hadoop, Spark, ETL tools, and cloud platforms.

Machine Learning Engineer

- A Machine Learning Engineer works with implementing machine learning models, to ensure that they are capable of handling production-level tasks. They also work on the development, deployment, and maintenance of models that can automate predictions and decision-making based on the data.
- They are part of the Data Mining and Predictive Modeling phase of the life cycle. During Data Mining they apply algorithms and Machine Learning to large datasets to extract meaningful patterns and insight that can assist later in business decisions. They can also have the responsibility of developing models to analyze historical data used to predict future outcomes. The models are then optimized to ensure accuracy and performance.
- They use programming languages such as Python and/or R, and machine learning libraries such as TensorFlow, and PyTorch. They can also use cloud-based services for model deployment such as GoogleAI, and AWS SageMaker.

Business Intelligence Developer

- The BI Developer creates tools, reports, and dashboards that turn data into business insights. They translate complex datasets into readable visualizations that can be used to make decisions in the organization.
- They are part of the Reporting and Visualization phase of the life cycle. They create visualization and dashboards that help the stakeholders understand the data. They work on transforming raw data into reports that aid in decision-making. Part of their job might also entail assisting non-technical stakeholders in understanding complex datasets.
- BI Developers typically use business intelligence tools like Tableau, Power BI, Looker, and SQL to create visual reports.

While all the roles work together in the life cycle, they each serve a different and distinct purpose. Each role requires a unique skill set, tools, and responsibilities to achieve their respective goals.

B

Data Mining

- It involves exploring large datasets to discover hidden patterns and relationships. It involves techniques such as classification, clustering, and association. Which works well with both structured and unstructured data. There is no decision-making or predictions made during this process.

Machine Learning

- Automating the process of pattern recognition and taking it a step further by making predictions and decisions based on historical data. After the model is set up it requires very little human interaction, which allows the system to learn and adapt as new data comes in. Unlike Data Mining this excels in predictive tasks and real-time decision-making based on learning patterns.

Statistics

- This discipline focuses on understanding the data through mathematical analysis. The data can be tested using statistical methods such as mean, variance, and/ or regression models. Unlike the other two disciplines, this usually focuses on smaller clean datasets. Emphasizing data collection and hypothesis testing before applying models.

B - (1)(2)

Data Scientist

- Data Scientists gather, analyze, and interpret large datasets to assist in making informed decisions.
- They must have a strong background in programming languages such as Python, and R, as well as knowing machine learning. Usually, a bachelor's degree in data science, computer science, or another related field, and can seek advanced roles with a master's degree.

Computer and Information Research Scientist

- They research to advance computer technology and solve complex computing problems. This usually involves developing new algorithms, software systems, and machine learning applications.
- They must have a deep understanding of algorithms, data structures, and computing theories. An advanced knowledge of machine learning, AI, and scientific programming. Most positions in this field will require an advanced degree such as a master's or Ph.D. in computer science or a related field.

Operations Research Analyst

- This is a focus on using mathematical models and statistical analysis to solve operational and logistical problems for businesses. They use data analytics tools to improve processes and resource management.
- Must have strong mathematical skills, particularly in statistics, calculus, and linear algebra. Experience with analytical software is also a must. A bachelor's degree in operational research, mathematics, or industrial engineering is required.

C - (1)

- 1. Ideation – Designing innovative solutions to complex data problems would allow me to explore and generate ideas that add value to an organization. Having the ability to see connections between unrelated data sets or phenomena would be critical.
- 2. Arranger – My organizational skills would be beneficial in managing data teams and resources. Being flexible will allow me to enable optimizing workflows, and ensuring productivity by arranging the right tools, people, and data sets.
- 3. Restorative – Bringing my troubleshooting and problem-solving skills to the table to resolve data inconsistencies, optimize algorithms, and address scalability issues would be a strength I could offer to an organization.
- 4. Learner – Being able to continuously learn. This is a strength that aligns with staying updated on the latest tools, technologies, and analytical techniques which would bring both personal and team success.
- 5. Adaptability – Projects often have shifting requirements and tight deadlines. Being able to stay calm under pressure and flexible when responding to unexpected changes in project direction or data needs.

Data Science compared with different analytics disciplines. ProjectPro. (2024, September 6). <https://www.projectpro.io/article/data-science-compared-with-different-analytics-disciplines/175>

U.S. Bureau of Labor Statistics. (2024, September 6). *Data scientists: Occupational outlook handbook*. U.S. Bureau of Labor Statistics. <https://www.bls.gov/ooh/math/data-scientists.htm>