Name: Mubina Ashrafi

Designation/Study: Completed B. Sc. in Computer Science and Engineering

I explored different data science roles, their responsibilities, and the required skills.

Role: Data Analyst

Responsibilities:

1. Primary Responsibilities:

Data Analysts are responsible for collecting, processing, and analyzing data to help organizations make informed decisions. Their primary responsibilities include:

- **Data Collection:** Gathering and collecting data from various sources.
- **Data Cleaning:** Cleaning and organizing raw data for analysis.
- **Data Analysis:** Using statistical methods and tools to analyze data sets.
- **Data Visualization:** Presenting findings through visualizations (charts, graphs, dashboards).
- **Reporting:** Creating reports and insights based on data analysis.
- **Identifying Trends:** Recognizing patterns and trends within the data.

Duties and Responsibilities of a Data Analyst

- Collect data
- Design and maintain databases and systems
- Analyze and interpret collected data
- Understand data trends and patterns
- Prepare reports
- Determine areas of improvement

indeed

2. Contribution to Business Decision-Making:

Data Analysts contribute to business decision-making processes by providing actionable insights derived from data analysis. They help in:

- **Risk Assessment:** Identifying potential risks and opportunities.
- **Performance Evaluation:** Analyzing business performance metrics.
- **Strategy Development:** Supporting strategic decision-making.
- **Forecasting:** Predicting future trends based on historical data.
- **Optimization:** Recommending improvements and optimizations.
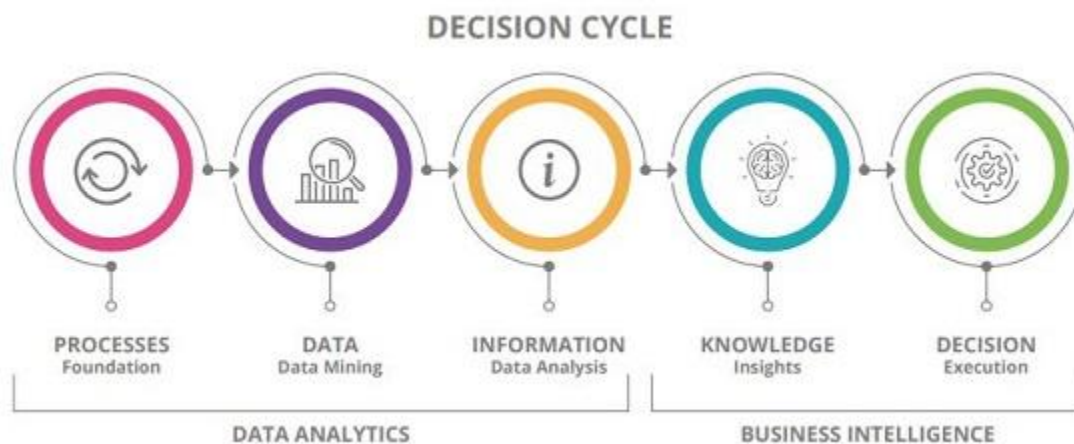
3. Examples of Projects:

Data Analysts may be involved in various projects, such as:

- **Market Analysis:** Analyzing market trends and consumer behavior.
- **Customer Segmentation:** Dividing customers into groups for targeted strategies.
- **Product Performance:** Evaluating the success of products/services.
- **Financial Analysis:** Assessing financial data for budgeting and forecasting.

Required Skills:

1. Technical Skills:

Crucial technical skills for a Data Analyst include:

- **Statistical Analysis:** Applying statistical methods to interpret data.
- **Programming:** Using languages like Python or R for data manipulation.
- **Database Management:** Proficiency in SQL for querying databases.
- **Data Visualization:** Creating compelling visualizations using tools like Tableau or Power BI.

2. Importance of Tools:

Proficiency in tools like Excel, SQL, and data visualization is crucial because:

- **Excel:** Widely used for data cleaning, analysis, and basic visualization.
- **SQL:** Essential for querying and managing databases.
- **Data Visualization Tools:** Aid in presenting complex data in a comprehensible manner.
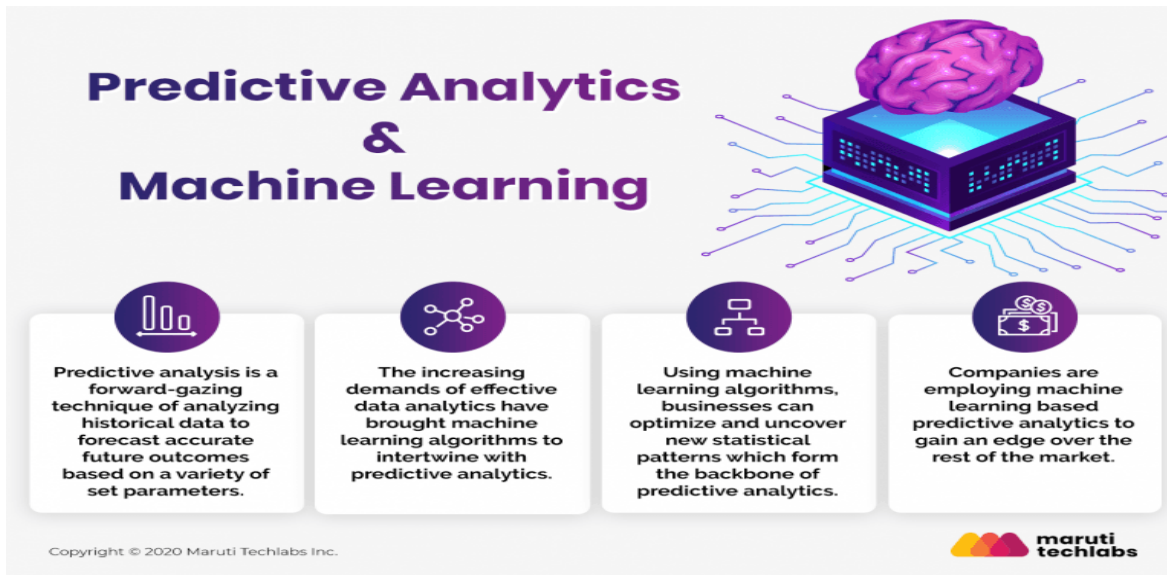
3. Soft Skills:

Soft skills that enhance a Data Analyst's effectiveness:

- **Communication:** Clearly conveying findings to non-technical stakeholders.
- **Critical Thinking:** Analyzing data critically to draw meaningful conclusions.
- **Attention to Detail:** Ensuring accuracy in data analysis.
- **Problem-Solving:** Addressing challenges in data processing and analysis.

4. Importance of Machine Learning:

While not always a primary focus, knowledge of machine learning is beneficial for Data Analysts because:

- **Advanced Insights:** ML can reveal complex patterns and insights.
- **Predictive Analytics:** Machine learning models can predict future trends.
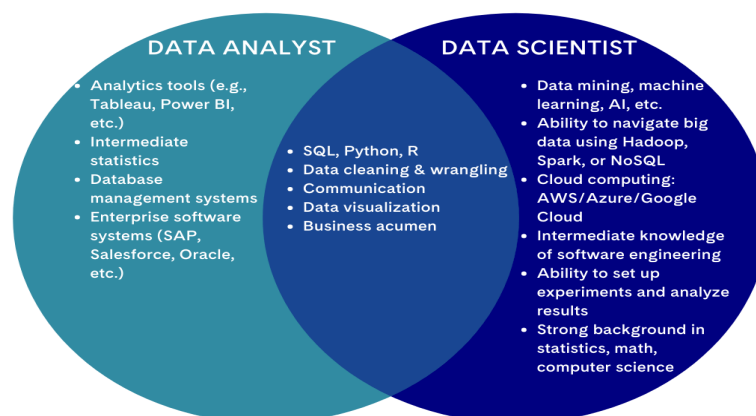- **Automation:** ML can automate certain aspects of data analysis.

Predictive Analytics & Machine Learning

Predictive analysis is a forward-gazing technique of analyzing historical data to forecast accurate future outcomes based on a variety of set parameters.

The increasing demands of effective data analytics have brought machine learning algorithms to intertwine with predictive analytics.

Using machine learning algorithms, businesses can optimize and uncover new statistical patterns which form the backbone of predictive analytics.

Companies are employing machine learning based predictive analytics to gain an edge over the rest of the market.

Copyright © 2020 Maruti Techlabs Inc.

Role: Data Scientist

Responsibilities:

1. **Distinguishing Data Scientist from Data Analyst:**

   Data Scientist: Focuses on extracting insights and knowledge from both structured and unstructured data. They are involved in developing and deploying machine learning models, conducting advanced statistical analysis, and often work with big data technologies.

   Data Analyst: Primarily deals with analyzing and interpreting complex datasets to provide actionable insights. They may use statistical analysis and visualization tools to present findings.

DATA ANALYST
- Analytics tools (e.g., Tableau, Power BI, etc.)
- Intermediate statistics
- Database management systems
- Enterprise software systems (SAP, Salesforce, Oracle, etc.)

- SQL, Python, R
- Data cleaning & wrangling
- Communication
- Data visualization
- Business acumen

DATA SCIENTIST
- Data mining, machine learning, AI, etc.
- Ability to navigate big data using Hadoop, Spark, or NoSQL
- Cloud computing: AWS/Azure/Google Cloud
- Intermediate knowledge of software engineering
- Ability to set up experiments and analyze results
- Strong background in statistics, math, computer science

2. **Using Statistical Modeling and Machine Learning**
   Data Scientists utilize statistical modeling to identify patterns, trends, and relationships in data. Machine learning is employed for predictive modeling, classification, clustering, and other tasks. They leverage algorithms to make sense of large datasets and create models that can make predictions or automate decision-making.

3. **Real-world Applications:**
   Healthcare: Predictive analytics for disease diagnosis.
   Finance: Fraud detection and risk assessment.
   Retail: Customer segmentation and recommendation systems.
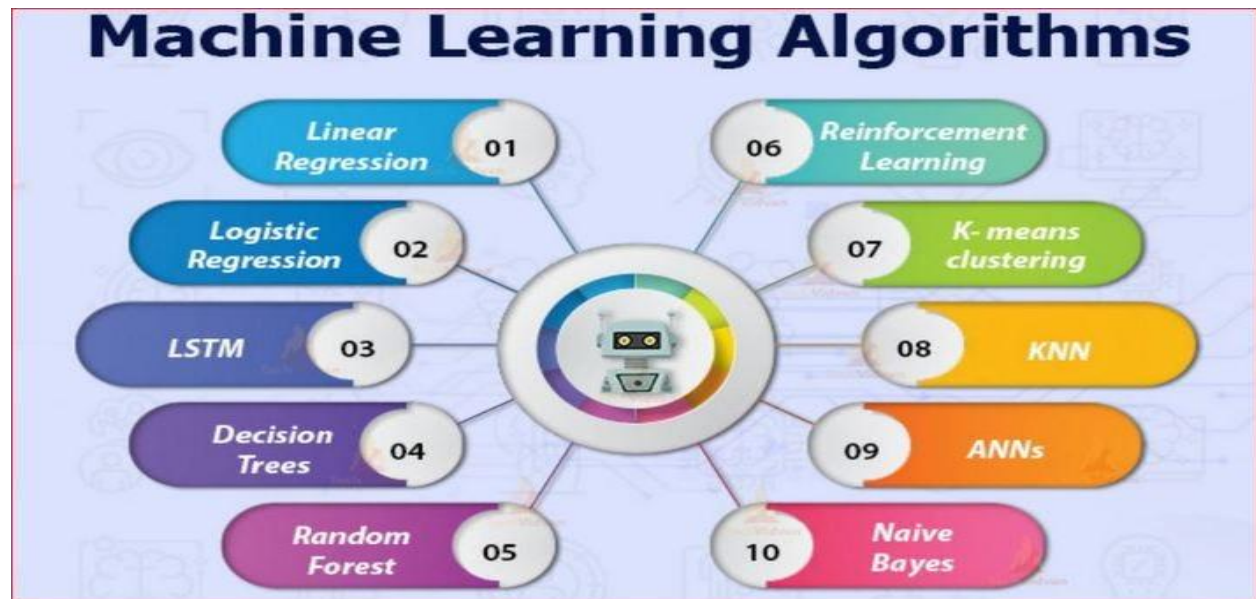   **Manufacturing:** Predictive maintenance for machinery.
   Marketing: Customer behavior analysis and targeted advertising.

Required Skills:

1. Commonly Used Machine Learning Techniques:
   - Supervised learning (e.g., regression, classification).
   - Unsupervised learning (e.g., clustering, dimensionality reduction).
   - Deep learning (neural networks).
   - Ensemble methods (e.g., random forests, boosting).
   - Reinforcement learning.

Machine Learning Algorithms

- 01 Linear Regression
- 02 Logistic Regression
- 03 LSTM
- 04 Decision Trees
- 05 Random Forest
- 06 Reinforcement Learning
- 07 K- means clustering
- 08 KNN
- 09 ANNs
- 10 Naive Bayes

2. Importance of Understanding Statistics and Probability:

- Statistical methods help in hypothesis testing and validating findings.
- Probability theory underlies many machine learning algorithms.
- A solid understanding enables Data Scientists to make informed decisions and assess the reliability of their models.

3. Approaching and Solving Complex Business Problems:

- **Problem Framing:** Clearly define the business problem.
- **Data Exploration:** Understand the available data.
- **Model Building:** Select and build appropriate models.
- **Evaluation:** Assess model performance.
- **Deployment:** Implement the solution in a real-world setting.
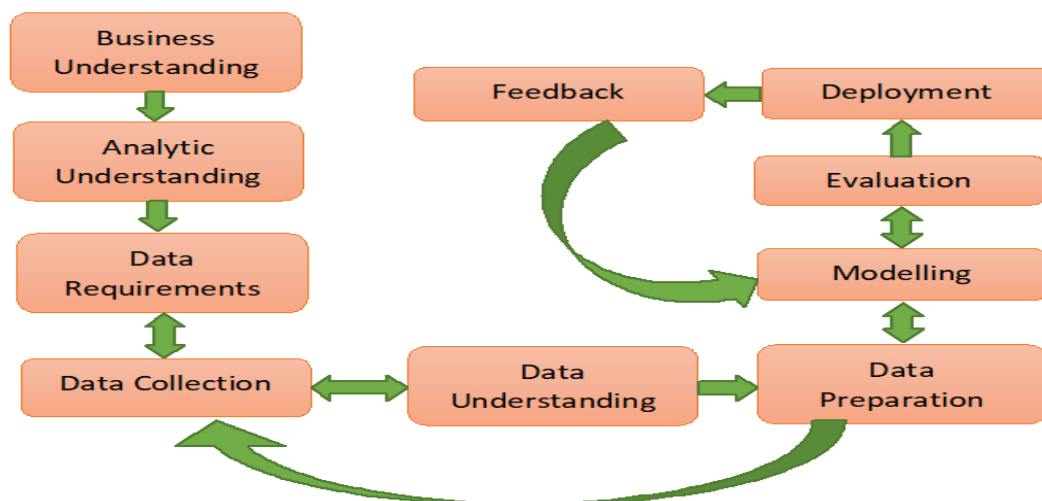
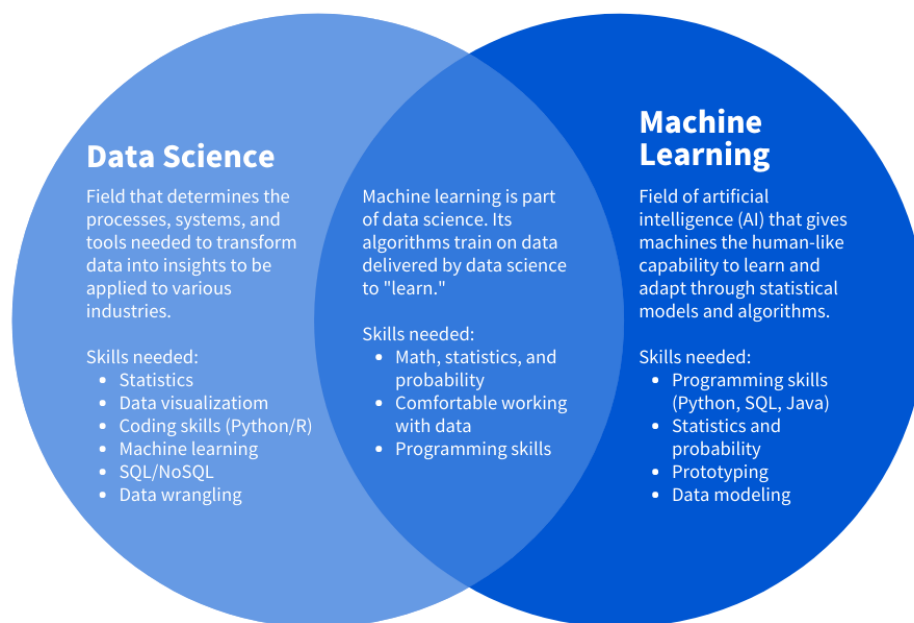**4. Importance of Machine Learning:**

- **Predictive Analytics:** Predict future trends and outcomes.
- **Automation:** Enable systems to learn and improve without explicit programming.
- **Pattern Recognition:** Identify complex patterns in data.
- **Decision Support:** Assist in making data-driven business decisions.

Role: Machine Learning Engineer

Responsibilities:

1. Difference from Data Scientist:

- A Machine Learning Engineer primarily focuses on developing and deploying machine learning models into production systems. They are more involved in the implementation and optimization of these models. In contrast, a Data Scientist often deals with the entire data science pipeline, including data exploration, feature engineering, and model interpretation. While there is overlap, Machine Learning Engineers typically have a stronger emphasis on the engineering aspects of deploying models in real-world applications.

Data Science

Field that determines the processes, systems, and tools needed to transform data into insights to be applied to various industries.

Skills needed:
- Statistics
- Data visualizatiom
- Coding skills (Python/R)
- Machine learning
- SQL/NoSQL
- Data wrangling

Machine learning is part of data science. Its algorithms train on data delivered by data science to "learn."

Skills needed:
- Math, statistics, and probability
- Comfortable working with data
- Programming skills

Machine Learning

Field of artificial intelligence (AI) that gives machines the human-like capability to learn and adapt through statistical models and algorithms.

Skills needed:
- Programming skills (Python, SQL, Java)
- Statistics and probability
- Prototyping
- Data modeling

2. Main Focus:
- The main focus of a Machine Learning Engineer is to design, build, and deploy machine learning models that can perform specific tasks or make predictions based on data. This involves selecting appropriate algorithms, optimizing model performance, and integrating models into production systems.

3. High Demand Industries/Applications:

- Machine Learning Engineers are in high demand across various industries, including:
    - **Finance:** Fraud detection, algorithmic trading.
    - **Healthcare:** Predictive analytics for patient outcomes, image analysis.
    - **E-commerce:** Recommendation systems, demand forecasting.
    - **Automotive:** Autonomous vehicle development.
    - **Technology:** Natural language processing, computer vision.
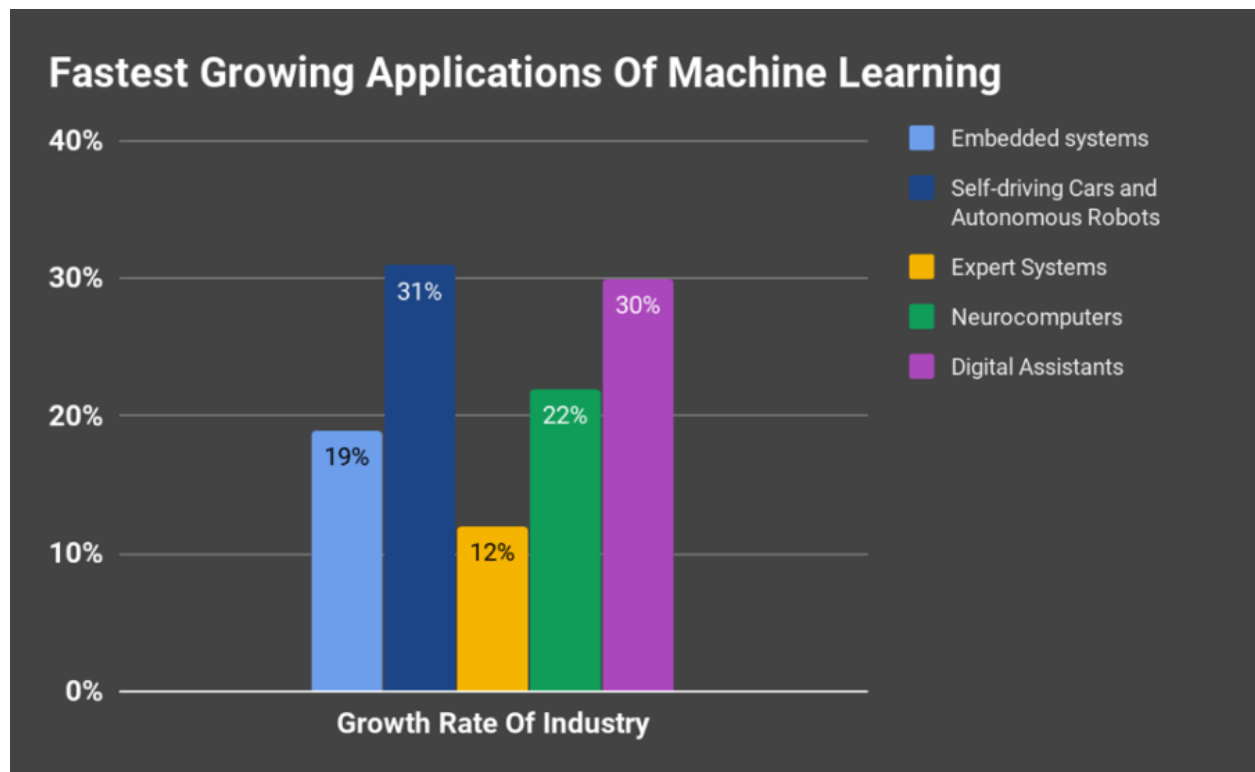
Required Skills:

1. Programming Languages and Frameworks:

- Languages: Python, R.
- Frameworks: TensorFlow, PyTorch, Scikit-learn.
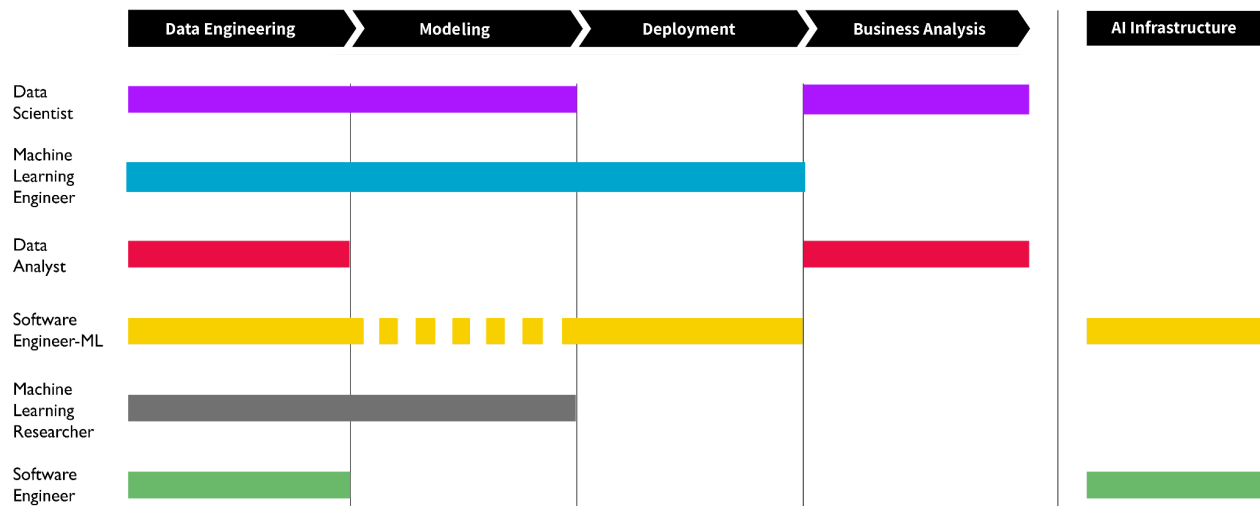- Libraries: NumPy, pandas.

2. Importance of Model Deployment and Optimization:

- Model deployment ensures that the developed machine learning models are integrated into real-world systems, allowing them to make predictions on new data. Optimization is crucial for improving model efficiency and reducing

inference time, ensuring that models can scale to handle large datasets and deliver results in a timely manner.

3. Collaboration with Data Scientists and Software Engineers:
- Data Scientists: Machine Learning Engineers collaborate with Data Scientists in the early stages of model development, receiving insights into data characteristics and requirements. They work together to ensure that models are aligned with business goals.
- Software Engineers: Collaboration involves integrating machine learning models into software applications. Machine Learning Engineers and Software Engineers work together to develop scalable, maintainable, and efficient systems.

Data Engineering	Modeling	Deployment	Business Analysis		AI Infrastructure
Data Scientist					
Machine Learning Engineer					
Data Analyst					
Software Engineer-ML					
Machine Learning Researcher					
Software Engineer					

4. Importance of Machine Learning:
- Machine learning is at the core of a Machine Learning Engineer's role, as they leverage algorithms and models to make predictions or decisions based on data. The importance lies in creating intelligent systems that can adapt and improve their performance over time, ultimately providing valuable insights or automation in various domains.

**Role: Data Engineer**

Responsibilities:

1. Core Responsibilities of a Data Engineer:
- Design, construct, install, and maintain efficient and scalable data pipelines.
- Develop, construct, test, and maintain data architectures (e.g., databases, large-scale processing systems).
- Collaborate with data scientists, analysts, and other stakeholders to understand data needs and implement solutions.
- Ensure data availability, integrity, and security.

2. Contributions to Data Infrastructure:

- Build and optimize data architecture, infrastructure, and tools for data generation, transformation, and consumption.
- Implement and manage ETL (Extract, Transform, Load) processes to move and integrate data across systems.
- Collaborate with IT teams to ensure seamless integration with existing systems.
- Continuously monitor and improve the performance, reliability, and scalability of data infrastructure.

ETL PIPELINE
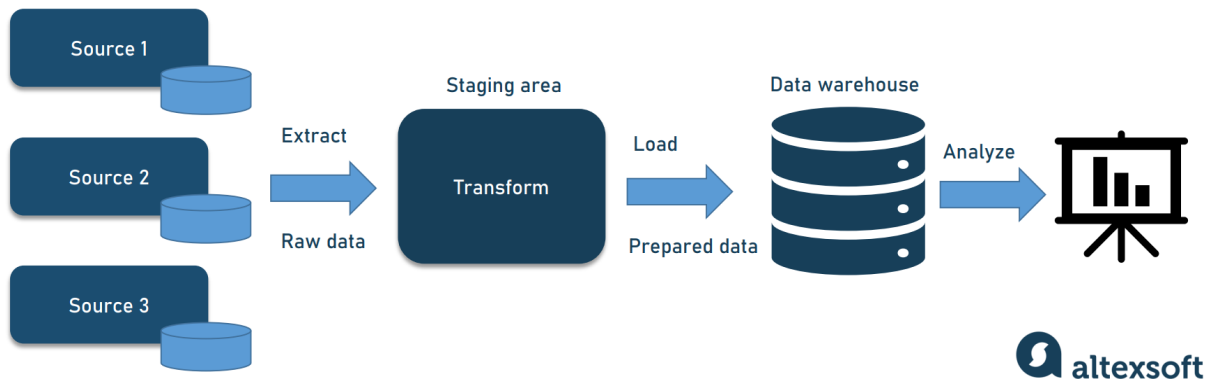
3. Examples of Daily Tasks:
- Writing and optimizing SQL queries for data extraction.
- Developing and maintaining ETL pipelines.
- Troubleshooting data-related issues.
- Collaborating with cross-functional teams to understand data requirements.
- Implementing data security and access controls.
- Keeping up with industry best practices and emerging technologies.
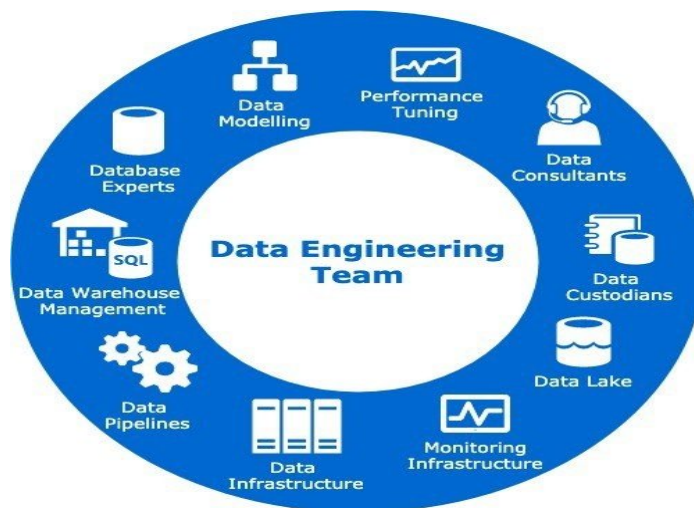
Required Skills:

1. Tools and Technologies:
- **Data Ingestion:** Apache Kafka, AWS Kinesis, Apache NiFi.
- **Storage:** Apache Hadoop, Amazon S3, Google Cloud Storage.
- **Databases:** MySQL, PostgreSQL, MongoDB, Cassandra.
- **ETL:** Apache Spark, Apache Airflow, Talend.
- **Cloud Platforms:** AWS, Azure, Google Cloud.

2. Importance of Knowledge:
- **Databases:** To design and optimize database schemas for efficient data storage.
- **ETL Processes:** Essential for extracting, transforming, and loading data between systems.
- **Cloud Platforms:** Modern data engineering often involves cloud-based solutions for scalability and flexibility.

3. Ensuring Data Quality:
- Implementing data validation checks in ETL processes.
- Setting up data monitoring and error handling.
- Collaborating with data stakeholders to define and enforce data quality standards.

Data Quality Pipeline
(Powered by Collibra DQ & Cloudera DataFlow)
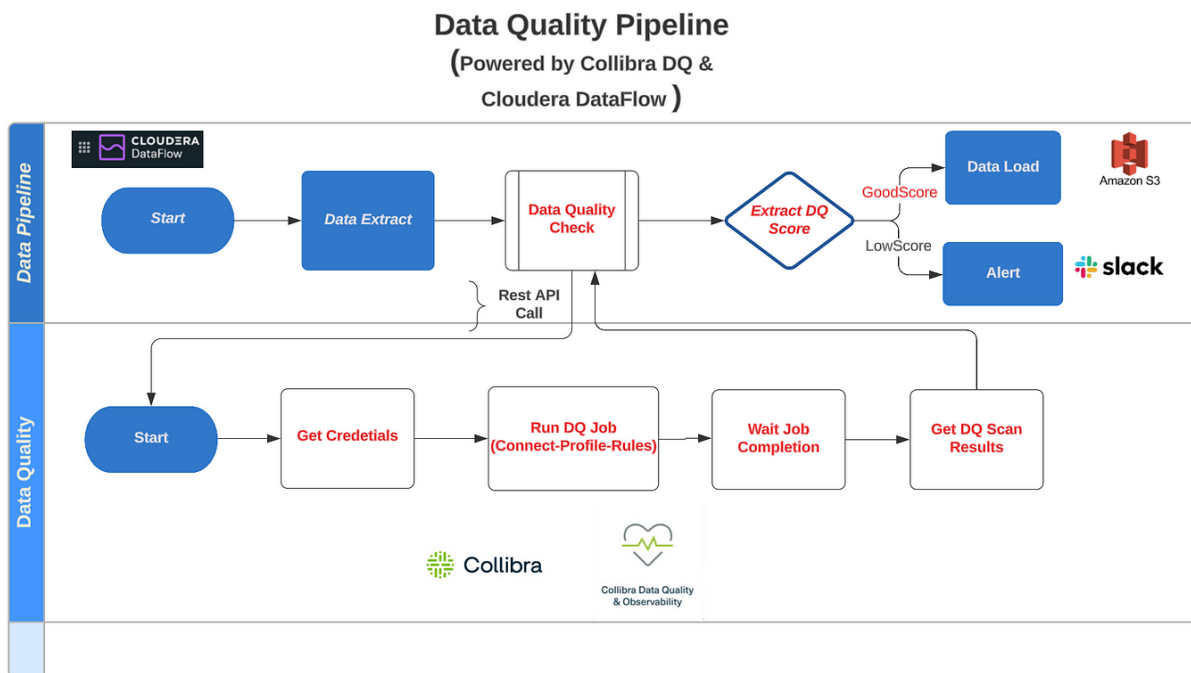
4. Machine Learning (ML):
- While not always a core responsibility, understanding ML concepts can be beneficial.
- Data Engineers may be involved in integrating ML models into data pipelines for real-time decision-making.

Role: Business Intelligence (BI) Analyst

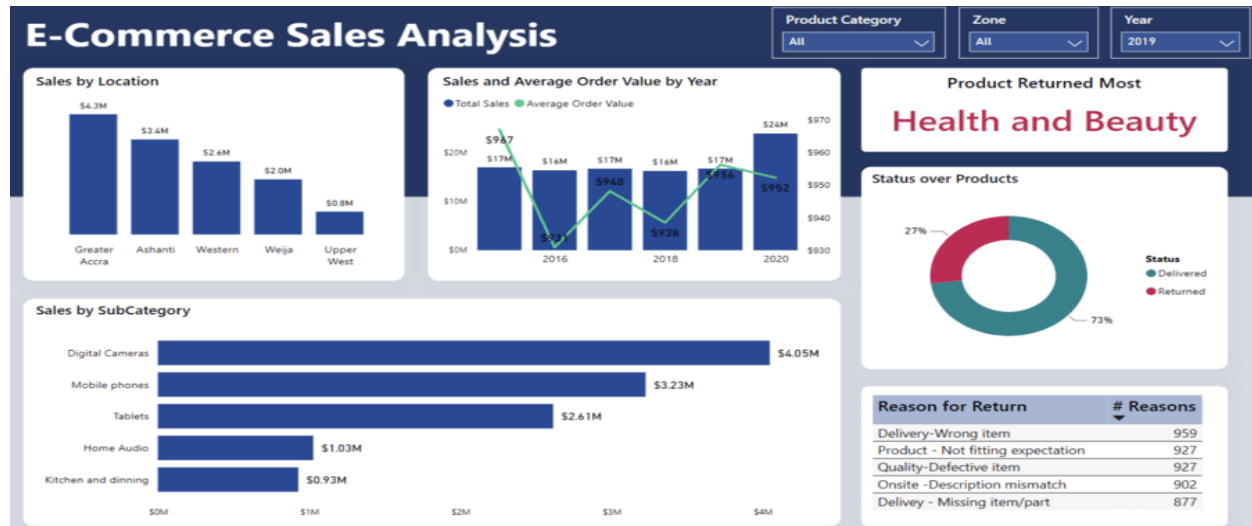**Responsibilities:**

1. Primary Focus:

- The primary focus of a Business Intelligence (BI) Analyst's role is to analyze and interpret complex data sets, transforming them into actionable insights that aid in business decision-making. They work with both internal and external data sources to provide valuable information for strategic planning and decision support.

2. Contribution to Business Performance Improvement:

- BI Analysts contribute to business performance improvement by:
    - Identifying trends and patterns in data to support strategic decision-making.
    - Creating and maintaining dashboards and reports to monitor key performance indicators (KPIs).
    - Conducting data-driven analysis to uncover opportunities for process optimization.
    - Providing insights into market trends and competitor activities.
    - Assisting in the development and implementation of data-driven strategies to enhance overall business efficiency.

3. Examples of Reports/Dashboards:
- **Financial Performance Dashboard:** Summarizes revenue, expenses, and profit margins.
- **Sales and Marketing Reports:** Analyzes customer behavior, conversion rates, and campaign effectiveness.
- **Operational Efficiency Dashboard:** Monitors key operational metrics to identify areas for improvement.
- **Customer Satisfaction Reports:** Analyzes feedback and customer interactions to improve service.

E-Commerce Sales Analysis

Sales by Location: Greater Accra $4.3M, Ashanti $3.4M, Western $2.6M, Weija $2.0M, Upper West $0.8M

Sales and Average Order Value by Year: 2016 $17M, 2017 $16M, 2018 $17M, 2019 $16M, 2020 $24M. Average Order Value: $947, $948, $938, $956, $952

Product Returned Most: **Health and Beauty**

Status over Products: Delivered 73%, Returned 27%

Sales by SubCategory: Digital Cameras $4.05M, Mobile phones $3.23M, Tablets $2.61M, Home Audio $1.03M, Kitchen and dinning $0.93M

Reason for Return	# Reasons
Delivery-Wrong item	959
Product - Not fitting expectation	927
Quality-Defective item	927
Onsite -Description mismatch	902
Delivey - Missing item/part	877

Required Skills:

1. Tools and Technologies:

- BI Analysts commonly use tools such as:
    - Microsoft Power BI
    - Tableau
    - QlikView/Qlik Sense
    - Google Data Studio
    - Excel (for basic reporting and analysis)
- They may also work with SQL for querying databases and extracting relevant data.

2. Understanding of Business Processes and KPIs:

- A deep understanding of business processes and KPIs is crucial because:
    - It allows BI Analysts to align data analysis with organizational goals.
    - Helps in identifying relevant metrics that reflect business performance.
    - Enables the creation of reports and dashboards that provide actionable insights.
    - Facilitates effective communication with stakeholders by framing insights in the context of business objectives.

They monitor company health.

They reveal patterns and trends.

5 Reasons Why KPIs Matter

They measure progress toward goals.

They uncover trouble spots.

They indicate whether a goal needs to be adjusted.

3. Communication with Non-Technical Stakeholders:
- BI Analysts communicate insights to non-technical stakeholders by:
  - Using clear and concise language, avoiding technical jargon.
  - Presenting information in visually appealing dashboards and reports.
  - Providing context to data, explaining its relevance to business objectives.
  - Conducting training sessions or workshops to enhance data literacy among stakeholders.
  - Collaborating with business teams to understand their requirements and tailoring communication accordingly.