Project -1

**Customized Data Engineering Roadmap**

Trainee Name:- Nachiketbhai Prajapati

Batch:- C19 (ADE)

1. **Review the First Session:**

* Data engineering focuses on designing and managing systems for efficiently collecting, processing, storing, and analyzing large volumes of data. It ensures that data pipelines are robust, scalable, and capable of delivering accurate insights.
* Data ingestion involves the process of collecting and importing data from various sources into a data storage system for further processing and analysis. It ensures the continuous and efficient flow of data from disparate sources into a centralized repository.
* Data ingestion involves the process of collecting and importing data from various sources into a data storage system for further processing and analysis. It ensures the continuous and efficient flow of data from disparate sources into a centralized repository.
* **Processing**: Data processing transforms raw data into a usable format through cleaning, filtering, and aggregation.  
  **Storage**: Data storage involves saving processed data in databases or data lakes for secure and efficient retrieval.  
  **Analysis**: Data analysis interprets stored data to extract insights and support decision-making through techniques like querying, visualization, and statistical modeling.

4. **Self-Assessment:**

* I hold a B.Tech in Civil Engineering and have since transitioned to the tech field by pursuing Project Management and currently studying AI and Data Analytics at Saskatchewan Polytechnic. My skill set includes critical thinking, time management, statistics, presentation skills, model validation, Microsoft Office, programming concepts, communication, data analytics, and various tools like Microsoft Power BI and SQL. I also excel in problem solving, data visualization, and reporting, alongside a strong work ethic and proficiency in English.
* **Strengths:**  
  I possess a solid foundation in critical thinking, analytical skills, and effective communication, which are crucial in both technical and collaborative environments. My proficiency in data visualization and analytics tools enables me to transform complex data into meaningful insights. Additionally, my time management skills allow me to balance multiple projects and deadlines effectively.
* **Weaknesses:**  
  One of my main challenges is the lack of hands-on experience in projects directly related to data analytics, primarily due to my previous focus on civil engineering. This gap has motivated me to embrace learning opportunities, particularly in Azure concepts, as I strive to bridge my background in civil engineering with my current studies. I view this as an exciting challenge to expand my skill set and enhance my qualifications in the tech industry.
* **Areas For improvement:**
* To enhance my transition into the tech field, I recognize several key areas for improvement. First, gaining practical experience through hands-on projects is essential. Engaging in real-world data analytics projects will not only solidify my theoretical knowledge but also build my portfolio.
* Second, I aim to deepen my understanding of cloud technologies, particularly Azure. Familiarizing myself with Azure tools and services will be crucial for my career in data engineering and analytics. I plan to seek online courses and certifications to strengthen my expertise in this area.
* Additionally, improving my programming skills, particularly in languages like Python and SQL, will allow me to manipulate and analyze data more efficiently. I intend to dedicate time to practice coding regularly and work on relevant projects to boost my proficiency.
* Lastly, enhancing my networking skills by connecting with professionals in the data analytics field can provide valuable insights and potential mentorship opportunities, guiding me as I navigate this career transition. By focusing on these areas, I aim to become a more well-rounded candidate in the competitive landscape of data analytics and engineering.

**5. Roadmap Development:**

#### 1. Skills to Acquire

* **Cloud Technologies**: Azure, AWS
* **Data Analytics Tools**: Microsoft Power BI, Tableau, Data bricks
* **Programming Languages**: Python, SQL, Pyspark
* **Data Management**: ETL processes, Data Warehousing
* **Machine Learning**: Basic algorithms and model evaluation
* **Data Visualization**: Advanced visualization techniques
* **Soft Skills**: Communication, Teamwork, Problem Solving

#### 2. Prioritization Based on Job Market Demand

* **High Demand**: Cloud Technologies (Azure), Python, SQL, Data Analytics Tools (Power BI, Tableau)
* **Moderate Demand**: Data Management (ETL), Machine Learning basics
* **Emerging Demand**: Advanced Data Visualization, Soft Skills (Communication and Teamwork)

#### 3. Desired Proficiency Levels

* **Beginner**: Basic understanding of Azure and AWS, introductory Python and SQL
* **Intermediate**: Proficient in SQL and Python for data manipulation, capable of creating basic dashboards in Power BI and Tableau
* **Advanced**: Ability to design ETL processes, perform advanced data analysis, and implement machine learning algorithms

#### 4. Documentation Format

* Create a structured Excel sheet with the following columns:
  + - Skill Name
    - Priority Level (High/Medium/Low)
    - Desired Proficiency Level (Beginner/Intermediate/Advanced)
    - Resources for Learning (Online courses, books, etc.)
    - Timeline for Acquisition (e.g., 3 months, 6 months)