**Data Engineering Roadmap.**

**Python fundamentals**

* Array/Lists: (slicing, sorting, array manipulation, multi dim. array)
* Dictionary
* Functions/ Lambda functions
* List Comprehension / Dictionary Comprehension
* Zip, unpacking
* OOP Concepts

**File handling:**

* Work with csv, Json, parquet files
* Csv parsing, Deal with exceptions

**Database connection and Query management**:

* Establish connection with database (postgres, SQLite)
* Create tables and Perform DML operations

**Data transformation / Data Manipulation**:

* Introduction to pandas
* Load data in dataframe and perform transformation
* Explode/ Flatten JSON
* Aggregate functions
* Build Unit tests for data validations. Business rules.
* Error handling \*\*
* Performance tuning/ optimization \*\*

**Cloud Fundamentals/ Cloud-Based data platform:**

* Snowflake. Create account.
* Azure Fundamentals. / GCP fundamentals.

**Case study:**

Provide relevant case study after every module covered.

* File based handling and performed full load, incremental load and transformations
* Database full load, incremental load and transformations
* Basic data pipeline. And Scheduling.
* Data Engineering best practices

**Prerequisites and Evaluation:**

* All hands on will be done on Git repository. Set up github account and Local git.
* Maintain documentation for all work done. Even for setups.
* After every module conduct a written test covering concepts in the module.
* Plan is to include a code walk through session.