**Python for Data Engineering**

Why python? Why not SQL?

* Structure and unstructured data difference
* Combination of SQL and python we need

Benefits of python:

* Reduce development time
* OOP
* No compile
* Reduce code length
* Easy to learn easy to understand
* Open source (free)
* Automatic memory management

Python is used to learn different purposes:

ML, DS, Data engineering, Deep learning, application processes, server level, etc.

Basics:

* Python data types: Numbers, strings, …. Etc.
* Data structures (collections): Lists, tuples, dictionaries, sets
* Conditionals and loop control statements: if, for, while, pass, break, continue..
* Functions

Advanced:

* Files Input and output
* Exception handling
* List comprehension
* Lambda expression
* Regular expression
* Modules and logging

Login to Databricks: Create cluster (only 1 cluster we can create in community edition)

* We can learn in data bricks:
  + Python
  + Pyspark
  + Spark SQL
  + Delta lake
* Python is case sensitive

Variables in python:

Local:

Global: