**Simple Introduction to ETL in python.**

**What is ETL**?

ETL stands for extract, transform, load. It is a process of moving the data from source to destination with some data transformation.

Source & Destination here can be Databases, Warehouse, Cloud etc.

**Objective**:

Extracting data from PostgreSQL database, apply data modeling and loading it back to the PostgreSQL database.

Topics covered:

* Database connectivity in python
* Data modeling in python

**Data Modeling**:

Data modeling is building the structure of the database, how data is stored, organized and manipulated.

There are two types of data modeling:

* Star schema
* Snowflake schema

Star schema is applied in the transformation phase.

The data available in PostgreSQL is a retail data.

**Retail Data Overview**:

This data is sample generated data using python.

It consists of 36070 rows and 15 columns. There 36070 distinct customers. It comprises of 3 years of data from 2017 to 2019. There are 5 products, each of product has 3 sub products. It operates in 4 locations. Each Location has 10 stores.

**Extraction Phase**:

Connecting to Postgres SQL using python and extracting the data.

**Transformation Phase**:

Applying data modeling “star schema” and creating fact, dimension tables.

**Loading Phase**:

Loading the transformed tables back to Postgres SQL database.