Chicken Breeder ETL data pipeline Project

This project demonstartes an ETL (Extract, Transform, Load) pipeline that extracts a 100 messy chicken dataset.

How it's Made

- Programming Language : Python

- Database: PostresSQL

- Data Source:

    - Created by AI. - As it was my first time using Pandas i wanted a simple set of Data

- Python Libraries:

    - pandas : For data manipulation and cleaning
    - psycopg2 : For Database interaction
    - dotenv: To handle enviromental variables

Lessons Learned :

- Pandas
  This was my first time learning and using pandas. In my previous project, I used python alone to clean the data. For example, using things like .strip() etc. After researching how data engineers/ how data is transfomed in the ETL process, I came across Pandas. I learnt the basics and the syntax through w3schools pandas teaching segment, and i made use of various methods such as isnull(), df.info(), .notna(). Before learning about Pandas i was curious as to how data was cleaned. I used to think, how would you know what parts of the data are missing, incorrect, misrepresented, however, pandas opened my eyes to how data can be analysed without having to look through every single column. It did this by having the data in a dataframe(df). And when you used the method df.info(), it would show details about the entire csv i extracted, specifically, the amount of Nulls in each column/s.

Im excited to develop this newly aquired skill and use it with real life datasets.