**Methodology for Crowdfunding Database Project**

* This project involves designing a database for managing crowdfunding campaigns. Below is a detailed methodology outlining the steps taken in this project:

Data Extraction and Transformation

* Identify the unique categories present in the crowdfunding dataset.
* Create a category\_id column with sequential identifiers, such as "cat1", "cat2", etc.
* The resulting DataFrame is exported to a CSV file (category.csv) for further processing.
* Identify the unique subcategories in the dataset.
* Create a subcategory\_id column with sequential identifiers (e.g., "subcat1", "subcat2", etc.).
* The transformed data is exported to a CSV file (subcategory.csv).
* Extract relevant columns from the original dataset, such as cf\_id, contact\_id, company\_name, blurb, goal, pledged, outcome, backers\_count, country, currency, launched\_at, deadline, and others.
* Rename certain columns to make them more readable (e.g., rename blurb to description, launched\_at to launch\_date).
* Convert goal and pledged columns to the float data type.
* Convert the launched\_at and deadline columns to a datetime format.
* Add category\_id and subcategory\_id columns, linking them to the previously created category and subcategory DataFrames.
* The transformed campaign data is saved as campaign.csv.
* Iterate through the rows of the contacts DataFrame and convert each row into a dictionary for easier processing.
* Extract the dictionary values by using Python list comprehensions.
* Create a new DataFrame containing the extracted data.
* Split the name column into two new columns: first\_name and last\_name.
* The cleaned and transformed data is saved as contacts.csv.

Database Design and Schema Creation

* Used QuickDBD to sketch an ERD that represents the relationships among the four primary tables:
* category
* subcategory
* campaign
* contacts
* Define data types for each column (e.g., VARCHAR, INTEGER, FLOAT, DATE, etc.).
* Specify primary keys for each table and foreign keys to establish relationships between the tables.
* Add constraints to ensure data integrity (e.g., NOT NULL, UNIQUE).
* The SQL schema is saved in a file named crowdfunding\_db\_schema.sql.

Database Creation and Data Import

* Create a new database named crowdfunding\_db using PostgreSQL.
* Execute the crowdfunding\_db\_schema.sql script to create the tables in the correct order, ensuring that foreign keys are handled appropriately.
* The table creation script ensures the correct sequence of table creation and links the foreign keys accurately.
* Import the CSV files (category.csv, subcategory.csv, campaign.csv, and contacts.csv) into their respective tables in the PostgreSQL database.
* Used SELECT statements to check that data has been correctly imported into each table.