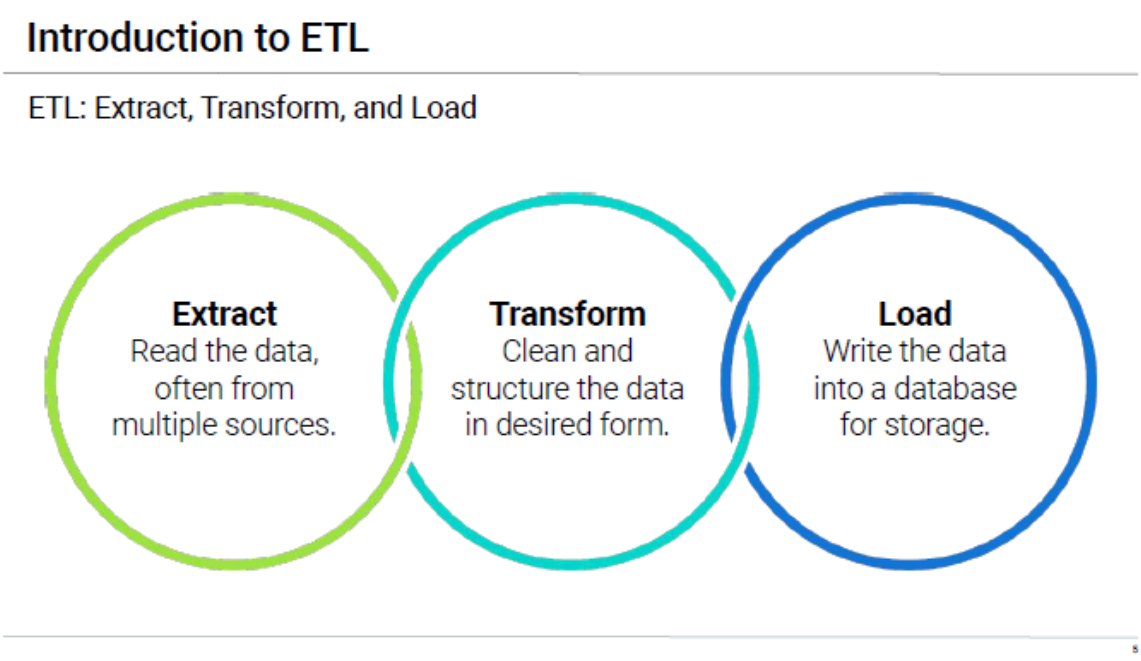
Project 2 – Crowdfunding ETL

Shiva Bajelan

Hail Nijo

Jeremy Hooper

ETL – Mini Project



Introduction

For our ETL mini project, we worked as a threesome to practise building an ETL pipeline using Python, Pandas, and either Python dictionary methods or regular expressions to extract and transform data. After transforming the data, we created four CSV files and used the CSV file data to create an ERD and a table schema. Finally, we uploaded the CSV file data into a Postgres database. In this report, each of us describes our contributions to the work.

Create the Category and Subcategory DataFrames - Shiva Bajelan

Create the Campaign DataFrame – Shiva Bajelan

Create the Contacts DataFrame – Jeremy Hooper

My task was to extract and transform the data from the contacts .xlsx Excel data file.

After Shiva had created the category, subcategory, and campaign DataFrames, she passed the program to me to continue to create the contacts DataFrame.

Our instructions were to **create a Contacts DataFrame that had the following columns:**

- A column named "contact_id" contains the contact person's unique number.
- A column named "first_name" containing the contact person's first name.
- A column named "last_name" that contains the contact person's first name.
- A column named "email" that contains the email address of the contact person

We were then to export the DataFrame as a *contacts.csv* CSV file.

I selected option 1, using Pandas to create the contact Data Frame. It proceeded as follows:

- Create a Contacts DataFrame that has the following columns:
- A column named "contact_id" that contains the unique number of the contact person.
- A column named "first_name" that contains the first name of the contact person.
- A column named "last_name" that contains the first name of the contact person.
- A column named "email" that contains the email address of the contact person

Then export the DataFrame as a contacts.csv CSV file.

- The work proceeded as follows:
- Iterate through the contact_info_df and convert each row to a dictionary.
- Use the try/except method in case error handling was needed.
- Print out the list of values for each row. This is organised to data into contact-id, name, and email.
- I created a “first_name” and “last_name” column, and then filled those columns using the “split” method, splitting the first and last names from the name by splitting the string on the “ “.
- To complete the transformation, I reordered to columns to contact_id, first_name, last_name, and email. The final DataFrame looks like this:

Out[51]:

	contact_id	first_name	last_name	email
0	4661	Cecilia	Velasco	cecilia.velasco@rodrigues.fr
1	3765	Mariana	Ellis	mariana.ellis@rossi.org
2	4187	Sofie	Woods	sofie.woods@riviere.com
3	4941	Jeanette	Iannotti	jeanette.iannotti@yahoo.com
4	2199	Samuel	Sorgatz	samuel.sorgatz@gmail.com

The data types are:

In [52]:

```
1 # Check the datatypes one more time before exporting as CSV file.
2 clean_contact_info_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 4 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   contact_id  1000 non-null   int64 
 1   first_name  1000 non-null   object
 2   last_name   1000 non-null   object
 3   email       1000 non-null   object
dtypes: int64(1), object(3)
memory usage: 31.4+ KB
```

- Lastly, I exported the DataFrame to a CSV file.

Create the ERD and Schema for the Crowdfunding Database – Hail Nijo, Shiva Bajelam, Jeremy Hooper.

Create PostgreSQL database Crowdfunding.db

Jeremy Hooper - Submission with Description of Documents and Images submitted.nb

JUPITER Notebook

- ETL-Mini_Project_JHooper.ipynb ETL contacts data for contracts.csv
- ETL-Mini_Project_SBajelan.ipy ETL crowdfunding and category data for category.csv, category.csv, subcategory.csv

CROWDFUNDING DATABSE Creation

- Crowdfunding.db ERD Diagram.png
- Crowdfunding.db schema.sql – Creating Tables

- Campaign with Table-data.png - Creation of Tables.
- Category with Table-data.png - Creation of Tables.
- Contacts with Table-data.png - Creation of Tables.
- Subcategory with Table-data.png - Creation of Tables.
- Confirm creation category table.png
- Confirm creation contacts table.png
- Confirm creation subcategorytable.png
- Confirm creation campaign table.png