**Project 2 Proposal**

1. **Creating DataFrames and Exporting CSVs:** *Iana*

Create the Category and Subcategory DataFrames by extracting and transforming data from the crowdfunding.xlsx Excel file. Export them as category.csv and subcategory.csv and save to your GitHub repository.

Software: Microsoft Excel (to open and work with the crowdfunding.xlsx file), Jupyter Notebook

Libraries: pandas: For data extraction, transformation, and DataFrame creation, numpy: For any numerical operations, csv: For exporting data to CSV files.

Skills: Extract data from an Excel file, data cleaning, filtering, and formatting, using pandas library for data manipulation, numpy for numerical operations, exporting data to CSV files.

1. **Creating the Campaign DataFrame:** *Peter*

Extract and transform data from crowdfunding.xlsx to create a campaign DataFrame. This involves renaming columns, data type conversion, and matching category and subcategory IDs. Export this DataFrame as campaign.csv and save it to your GitHub repository.

Software: Microsoft Excel (to open and work with the crowdfunding.xlsx file), Jupyter Notebook

Libraries: pandas: For data extraction, transformation, and DataFrame creation, numpy: For any numerical operations, csv: For exporting data to CSV files.

Skills: Extract data from an Excel file, data cleaning, filtering, and formatting, using pandas library for data manipulation, numpy for numerical operations, exporting data to CSV files, database concepts for ID matching.

1. **Creating the Contacts DataFrame and Crowdfunding Database:** *Sophie*

Choose either Option 1 (using Python dictionary methods) or Option 2 (using regular expressions) to extract and transform data from contacts.xlsx to create the Contacts DataFrame. Clean and export this DataFrame as contacts.csv to your GitHub repository.

Sketch an Entity-Relationship Diagram (ERD) based on the four CSV files and use it to create a table schema. Save the schema as crowdfunding\_db\_schema.sql and put it on your GitHub repository.

Create a new Postgres database named crowdfunding\_db, create tables based on the schema, and verify data import from the CSV files by running SELECT statements.

Software: Microsoft Excel (to open and work with the contacts.xlsx file), Postgres (for creating the database and managing data), QuickDB for ERD

Libraries: pandas: For data extraction, transformation, and DataFrame creation, re (for Option 2, using regular expressions), csv: For exporting data to CSV files.

Skills: Extract data from an Excel file, data cleaning, parsing, and formatting, pandas library, dictionary methods for data transformation regular expressions for data extraction, PostgreSQL, creating table schemas and running SELECT statements, Entity-Relationship Diagram (ERD) based on the project's data structure.

To do:

1. Add collaborators
2. Clone repos on computer
3. Each create own part of jupyter notebook or create a branch