ETL Mini Project

by Gonzalo and Anika

Crowdfunding Dataframe

Out[2]:

id	company_name	blurb	goal	pledged	outcome	backers_count	country	currency	launched_at	deadline	staff_pick	spotlight	category & sub-category
61	Baldwin, Riley and Jackson	Pre-emptive tertiary standardization	100	0	failed	0	CA	CAD	1581573600	1614578400	False	False	food/food trucks
65	Odom Inc	Managed bottom-line architecture	1400	14560	successful	158	US	USD	1611554400	1621918800	False	True	music/rock
87	Melton, Robinson and Fritz	Function-based leadingedge pricing structure	108400	142523	successful	1425	AU	AUD	1608184800	1640844000	False	False	technology/web
41	Mcdonald, Gonzalez and Ross	Vision-oriented fresh-thinking conglomeration	4200	2477	failed	24	US	USD	1634792400	1642399200	False	False	music/rock
99	Larson-Little	Proactive foreground core	7600	5265	failed	53	US	USD	1608530400	1629694800	False	False	theater/plays

Out[5]:

blurb	goal	pledged	outcome	backers_count	country	currency	launched_at	deadline	staff_pick	spotlight	category & sub-category	category	subcategory
re-emptive tertiary dardization	100	0	failed	0	CA	CAD	1581573600	1614578400	False	False	food/food trucks	food	food trucks
Managed bottom-line rchitecture	1400	14560	successful	158	US	USD	1611554400	1621918800	False	True	music/rock	music	rock
Function-based adingedge pricing structure	108400	142523	successful	1425	AU	AUD	1608184800	1640844000	False	False	technology/web	technology	web
n-oriented sh-thinking lomeration	4200	2477	failed	24	US	USD	1634792400	1642399200	False	False	music/rock	music	rock
Proactive foreground core	7600	5265	failed	53	US	USD	1608530400	1629694800	False	False	theater/plays	theater	plays

Category & Subcategories DataFrames

Out[12]:

	category_id	category
0	cat1	food
1	cat2	music
2	cat3	technology
3	cat4	theater
4	cat5	film & video
5	cat6	publishing
6	cat7	games
7	cat8	photography
8	cat9	journalism

Out[14]:

	subcategory_id	subcategory
0	subcat1	food trucks
1	subcat2	rock
2	subcat3	web
3	subcat4	plays
4	subcat5	documentary
5	subcat6	electric music
6	subcat7	drama
7	subcat8	indie rock
8	subcat9	wearables
9	subcat10	nonfiction
10	subcat11	animation
11	subcat12	video games
12	subcat13	shorts
13	subcat14	fiction
14	subcat15	photography books
15	subcat16	radio & podcasts
16	subcat17	metal
17	subcat18	jazz
18	subcat19	translations
19	subcat20	television
20	subcat21	mobile games
21	subcat22	world music
22	subcat23	science fiction
23	subcat24	audio

Category and Subcategory lists merged with Crowdfunding.xslx

Creating Contacts DataFrame

- Pandas method
  - Using json.loads() method to convert data and columns to Python dictionary
    - dict_values and column list
  - str.split() method to split first and last names from name column, then used df.drop() method to drop original name column
  - Created clean contacts_df by exporting dataframe to csv file

contact_id	first_name	last_name	email		
4661	Cecilia	Velasco	cecilia.velasco@rodrigues.fr		
3765	Mariana	Ellis	mariana.ellis@rossi.org		
4187	Sofie	Woods	sofie.woods@riviere.com		
4941	Jeanette	Iannotti	jeanette.iannotti@yahoo.com		
2199	Samuel	Sorgatz	samuel.sorgatz@gmail.com		
5650	Socorro	Luna	socorro.luna@hotmail.com		
5889	Carolina	Murray	carolina.murray@knight.com		
4842	Kayla	Moon	kayla.moon@yahoo.de		

Snippet of finalized clean contacts.csv

SQL

- Used QuickDBD to sketch out ERD of the columns
  - Serial primary key → contacts.contact_id
  - Other primary keys → category.category_id, subcategory_id
- Imported csv files to crowdfunding_db with CREATE TABLE with constraints and primary keys
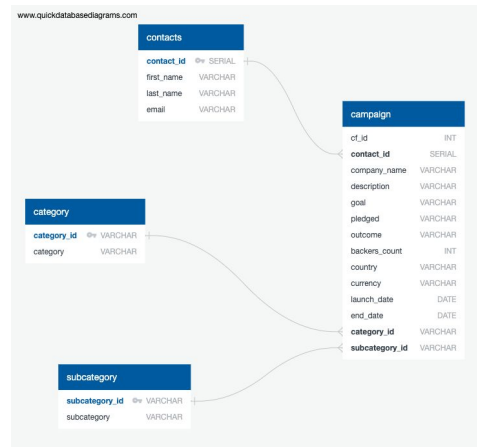
QuickDBD Sketch

Testing out successful import with campaign table in postgres