Alexandra Blitch
Uzor Francis
Misha Borunda

Project Two Crowdfunding – Written Report

The purpose of this project was to build an ETL (Extract, Transform, Load) pipeline using Python and Pandas to transform crowdfunding data from Excel files. Four different CSV files were created and then used to generate an ERD (Entity Relationship Diagram) and Table Schema. This information was then extracted and loaded into a Postgres database to analyze the data and get insight information.

For the cleaning portion of the project, we had to create new columns to show category and subcategory ids to have unique identifiers for the data and be able to have primary and foreign keys in the new tables that were created. In addition to this, we also had to split a column that included both the category and subcategory items together. After we split the information, the combined column was deleted and therefore we ended up with two clean columns with single items. Another step in the cleaning process was to transform the data into the correct formatting such as dates and integers. The main reason why we had to create four different tables is to eliminate data redundancy by creating primary keys outside the main data table.

Our Crowdfunding Data Base included four different tables (campaign, contacts, category, and subcategory). Using these four tables we were able to get insight information by making joins and aggregations to understand the relationship between datasets. The campaign table has the majority of the data. It shows the Crowdfunding data such as pledge amounts, goals, countries, and company name among other information. We used the foreign keys category_id, subcategory_id, and contact_id to connect to the category, subcategory, and contacts tables. Some of the insight information we were able to discover

includes the category with the highest single pledged which turned out to be “Film & Video”. Another interesting finding was that the US had both the minimum and maximum pledged amounts. Also that Canada has the highest average pledged (Fig .1)amount with $63,927 and that there were 201 more successful outcomes compared to failing ones.

For future work on this data base we suggest using Python to update CSV files into Postgres in order to automate the process of updating new information.

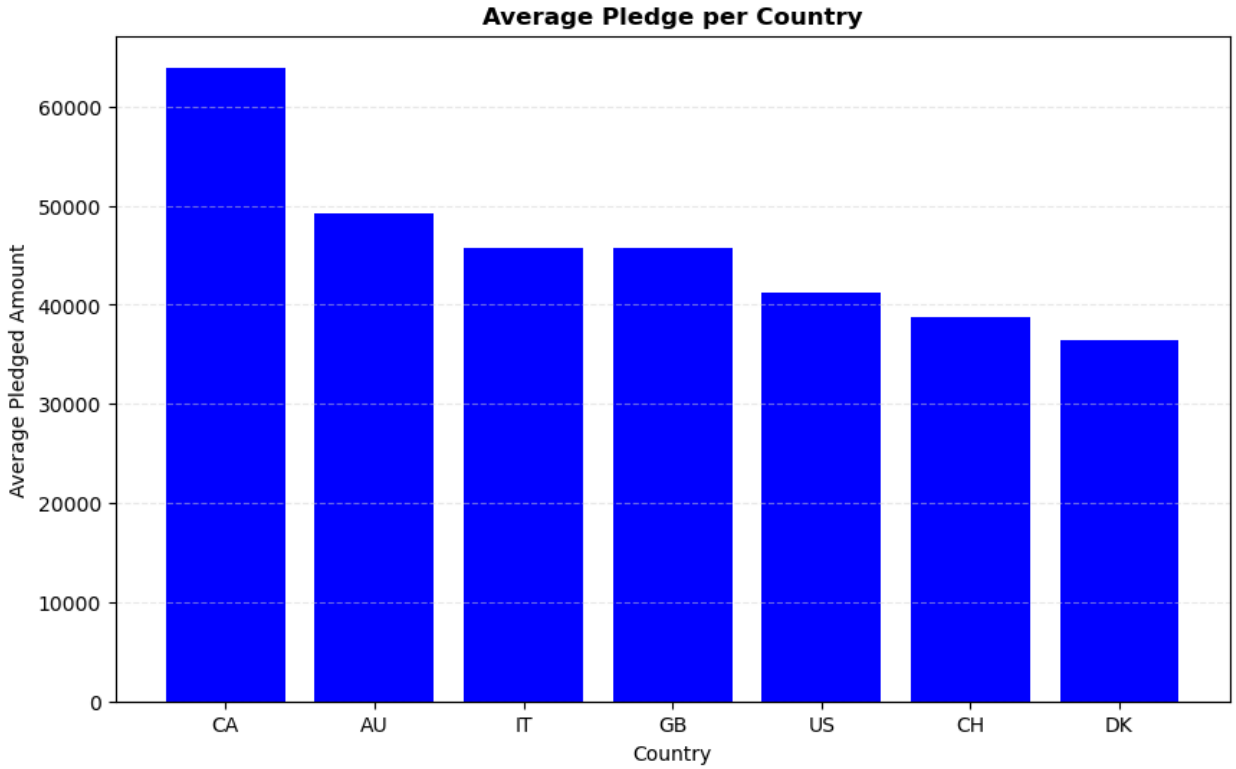


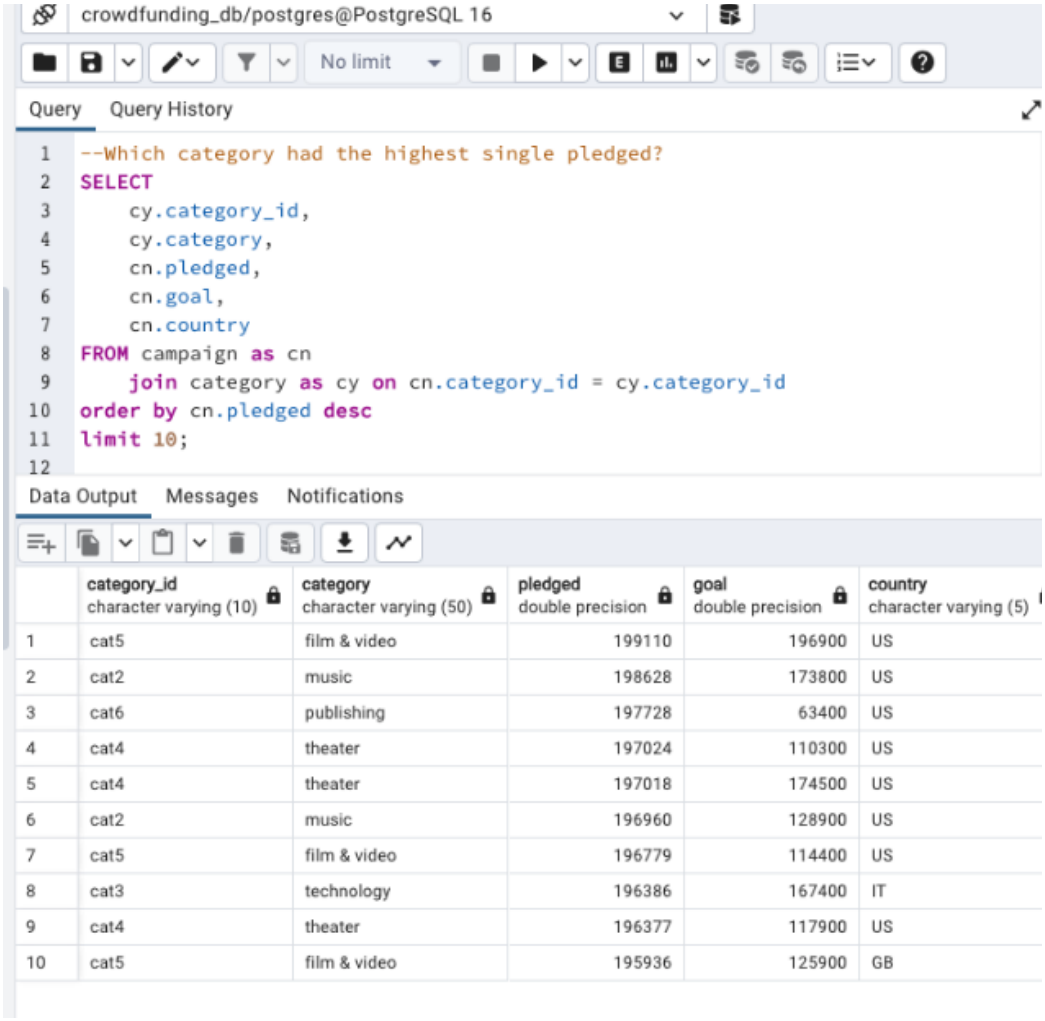
Fig. 1

Alexandra Blitch
Uzor Francis
Misha Borunda

Query Summary:

Query 1: Which category had the highest single pledged?

A: “Film & Video”

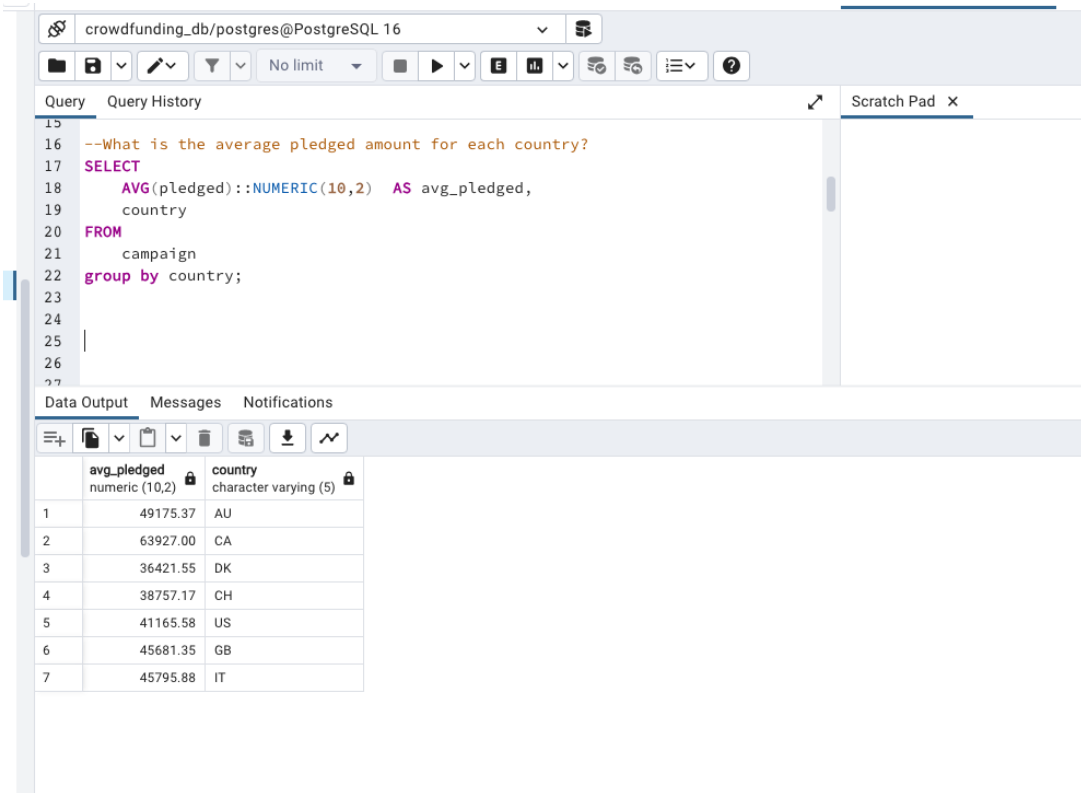


	category_id character varying (10)	category character varying (50)	pledged double precision	goal double precision	country character varying (5)
1	cat5	film & video	199110	196900	US
2	cat2	music	198628	173800	US
3	cat6	publishing	197728	63400	US
4	cat4	theater	197024	110300	US
5	cat4	theater	197018	174500	US
6	cat2	music	196960	128900	US
7	cat5	film & video	196779	114400	US
8	cat3	technology	196386	167400	IT
9	cat4	theater	196377	117900	US
10	cat5	film & video	195936	125900	GB

Alexandra Blitch
Uzor Francis
Misha Borunda

Query 2: What is the average pledge amount for each country?

A: AU $49K, CA $64K, DK $36K, CH $39K, US $41K, GB $46K, and IT $46K

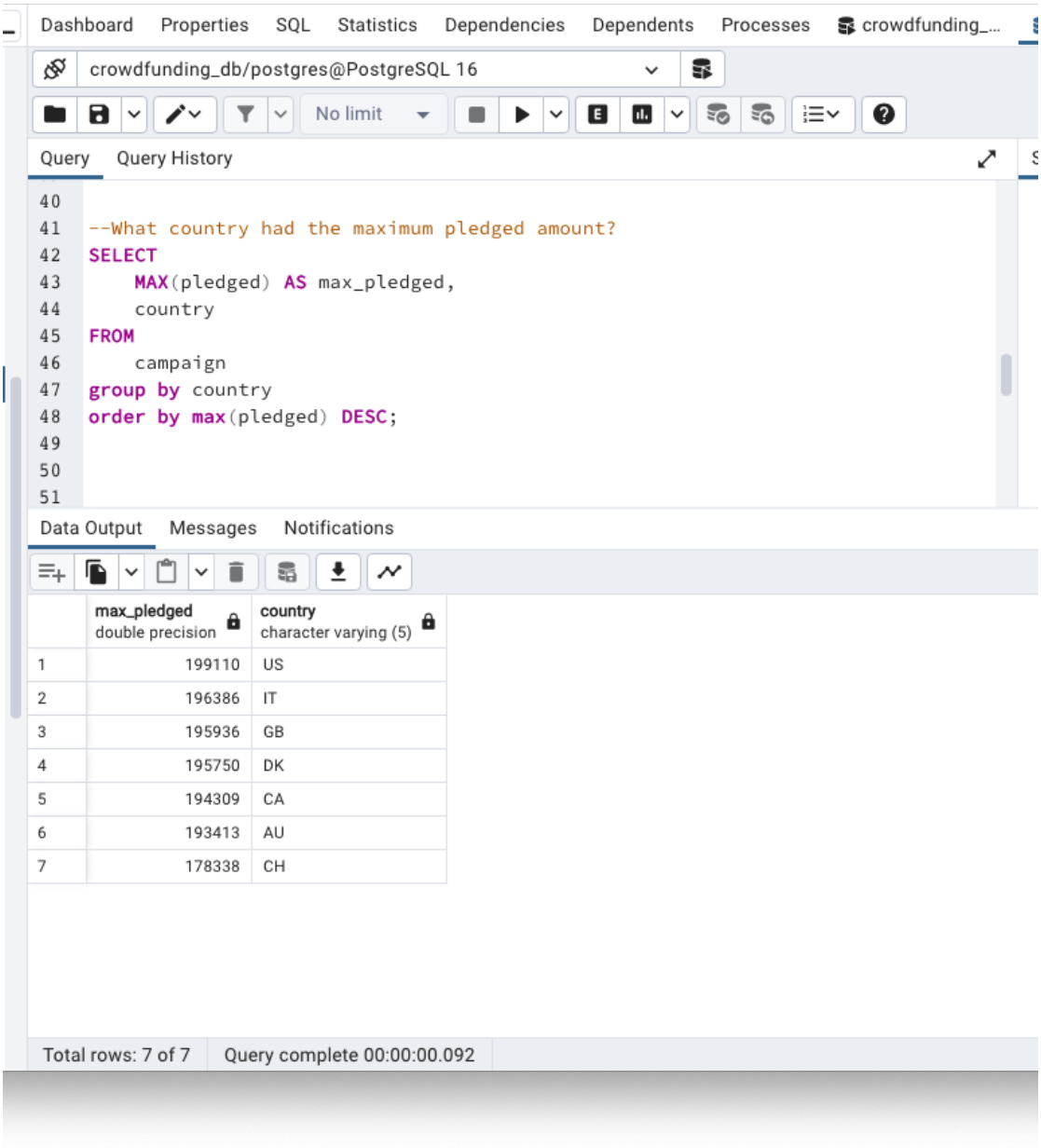


	avg_pledged numeric (10,2)	country character varying (5)
1	49175.37	AU
2	63927.00	CA
3	36421.55	DK
4	38757.17	CH
5	41165.58	US
6	45681.35	GB
7	45795.88	IT

Alexandra Blitch
Uzor Francis
Misha Borunda

Query 3: Which country pledged the highest amount?

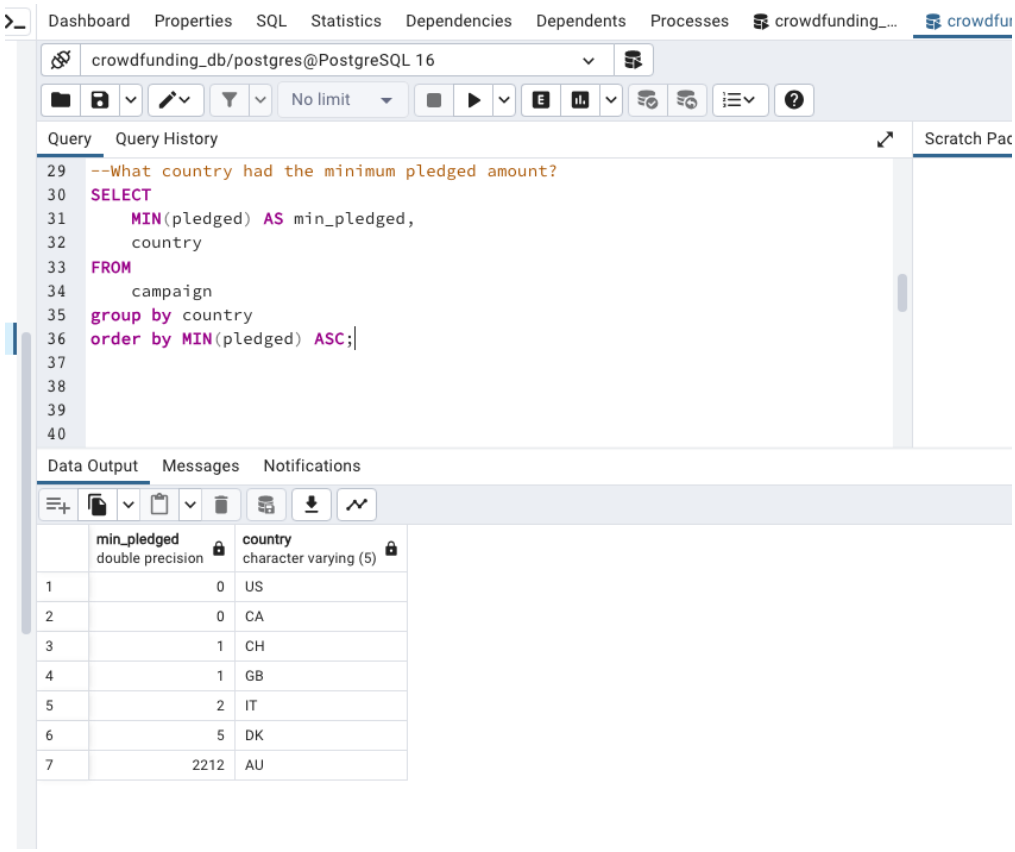
A: US with $199,110



Alexandra Blitch
Uzor Francis
Misha Borunda

Query 4: Which country pledged the lowest amount?

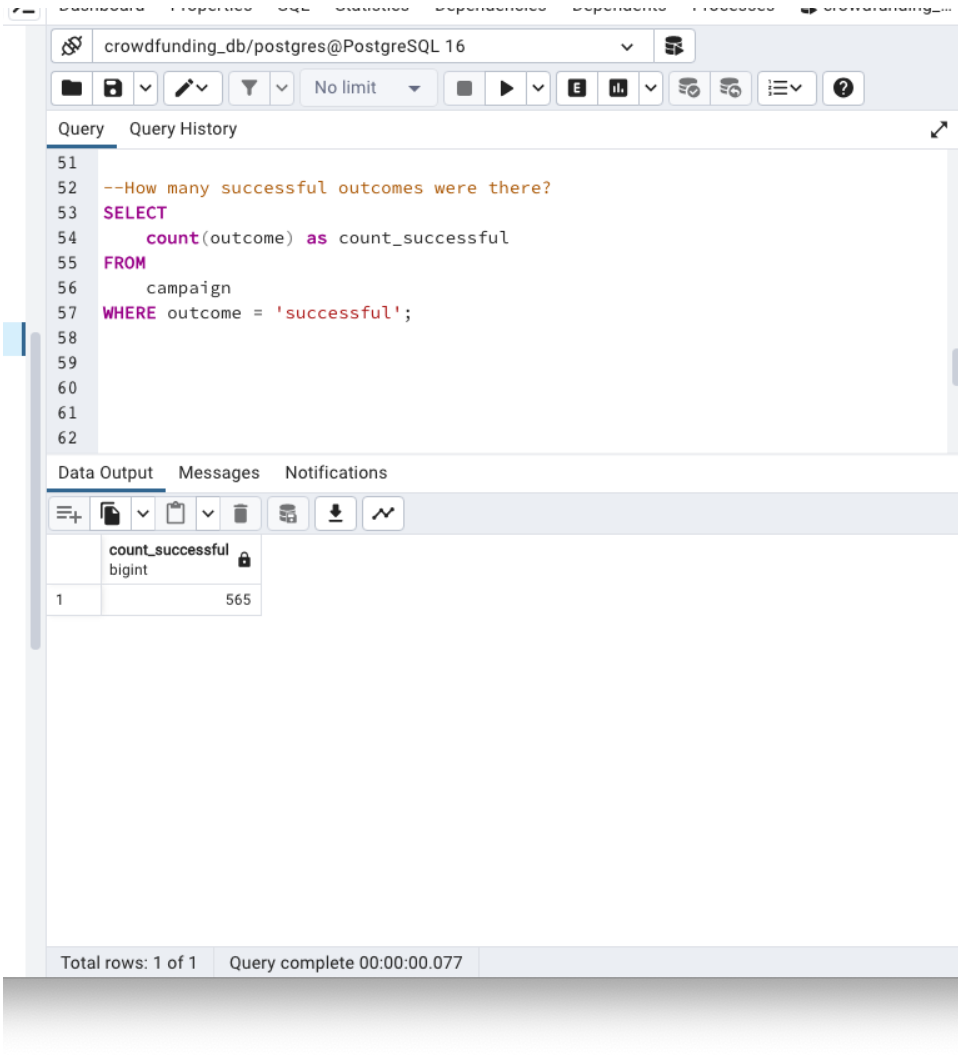
A: US and Canada with $0



Alexandra Blitch
Uzor Francis
Misha Borunda

Query 5: How many successful outcomes were there?

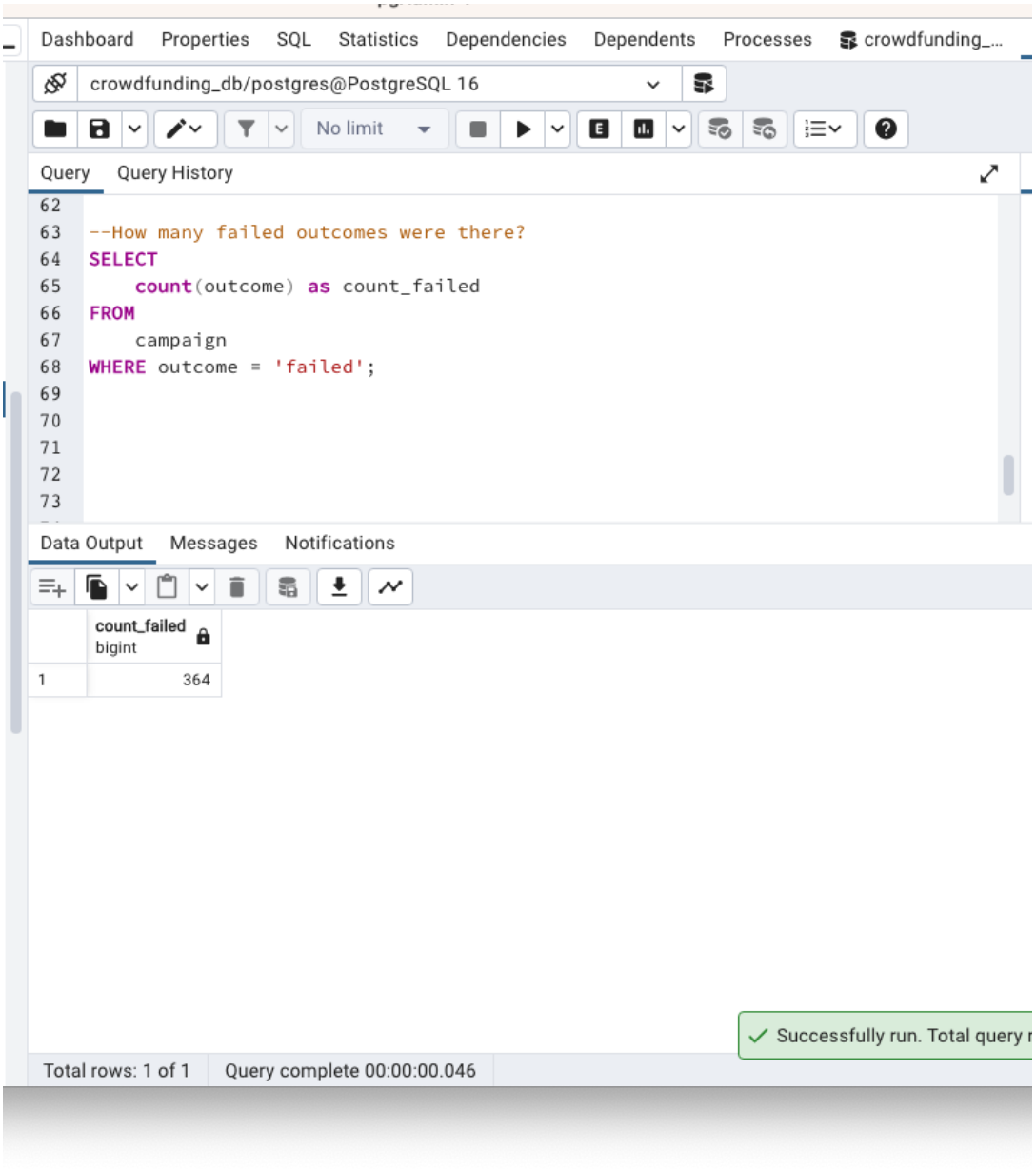
A: 565



Alexandra Blitch
Uzor Francis
Misha Borunda

Query 6: How many failed outcomes were there?

A: 364



Alexandra Blitch
Uzor Francis
Misha Borunda

ERD: