Project 2

Team Name: Ninety9Bottles- Bianca Worthy, Jerome Amose, Radhika Anusuri, Brunel Thimothe, and Tracey Fudge

Extract: We used 2 data sources, <https://covid19datahub.io/articles/python.html> and <https://www.kaggle.com/datasets/gabrielramos87/iowa-sales-liquor-jan-2021jan-2022>, both formatted by CSV.

Transform: We cleaned both CSV files using Pandas. From the Iowa Sales Liquor dataset, we dropped every row with no values, then reset the index to tidy it up. Finally, we created all the relational tables and removed duplicates from both the store and alcohol tables. For the Covid 19 dataset, we only grabbed our needed columns: *date*, *confirmed*, and *deaths.*

Load: Using Python we exported each table into Excel, converted to CSV then uploaded to PostgreSQL. Using the uploaded tables, we made different queries identifying correlations between Iowa sales and Covid confirmed cases.