

## Project 2

Team Name: Ninety9Bottles

Members: Bianca Worthy, Jerome Amose, Radhika Anusuri, Brunel Thimothe, and Tracey Fudge

### **Executive Summary:**

During the COVID-19 pandemic, U.S. alcohol retail store sales increased compared to usual trends while food services and drinking places sales decreased markedly during the same period. These results indicate an increase in home drinking in the U.S

Team Ninety9Bottles objective was to examine the impact of alcohol during COVID-19. As part of the analysis, the team will review different dataset to find the correlation between alcohol sales and COVID-19.

The following data set was used for the analysis:

1. <https://covid19datahub.io/articles/python.html>
2. <https://www.kaggle.com/datasets/gabrielramos87/iowa-sales-liquor-jan-2021jan-2022>,

The data extracted from both sources were downloaded in csv format.

## Project background:

Team Ninety9Bottles wanted to utilize the dataset mentioned above and determine if COVID-19 impacted the sales of alcohol with a good portion of the population in the US staying at home. To narrow down our analysis and due to time constraints, the team focused the analysis on the states of Iowa and more specifically the county of Kossuth.

## Solutions and approach

The team made the decision to analyze the 2021 data set across both files. In terms of technology, we went with Python(Pandas) and a SQL database for the project.

Python: We used python pandas due to the versatility of the module. Both data set that we used needed to be cleaned and filtered. We had to remove duplicate data from the dataset and also removed all the NULL. Also Pandas allowed us to extract the column that we needed from the spreadsheet and format the data in a way where it can be loaded into the database.

PostgreSQL: The team chose a relational database for a few reasons:

1. The ability to create meaningful information by joining different tables
2. We needed to group and combine different queries to get the information that we needed
3. Reduce redundancy. By going with a relational database, we were able to group all the store information and liquor in distinct tables.
4. The data set that we were using did not meet the key-value pair that is common with No SQL

We first used an ER diagram to simplify the critical database abstraction and design process. This ensures an adequately modeled and designed process.

By adding different constraints within our tables, we were able to extract the data that we needed from different tables using the primary and foreign key.

Approach:

As part of the project, we needed to perform a few steps to arrive at the data that we needed.

The 3 steps process are:

- Transform: Both CSV files were cleaned using Pandas ([Pandas code](#)) As mentioned above, the project focused on the state of Iowa. The data was examined using Pandas and all columns

with no relevant value to our project were dropped. Also as part of the cleanup, we used Pandas to remove any duplicate and NULL values which were not needed for the project.

- Load: Using Pandas, we extracted the data set that we needed for the files above and produced different Excel files which we then used to load to the different table that were created in PostgreSQL. The table creation and load table SQL were used to create the database schema and also setup all the constraint that we needed for the database. ([SQL ETL](#))

- Analysis: Once the data was loaded in PostgreSQL, we created different JOIN queries to extract the data that we needed for our analysis. ([Data Retrieval](#))

The following script was used to load the data in PostgreSQL ([SQL Load](#))

## Conclusion

In conclusion, the analysis performed by Ninety9Bottles showed that during COVID-19, we saw an increase in sales for the county of Kossuth.

To review the actual code used in this project. Please refer to this link: [Github Code](#)