**Extract, Transform, Load Group Project**

**Purpose:**  As a newly founded coffee shop built on strong, ethical morals, we strive not only for the best coffee bean, but delivering the bean to you, the consumer, in a way that reflects what we genuinely believe. Our task is to identify the best vendor that meets both needs.

**Extract Data:** Our dataset was taken directly from Kaggle at the below links:

<https://www.kaggle.com/volpatto/coffee-quality-database-from-cqi>

<https://www.kaggle.com/yamaerenay/ico-coffee-dataset-worldwide?select=disappearance.csv>

**Transform Data:** The transformation needed was dataset dependent here. After initial review, the data was relatively “clean”. Minor adjustments were made to the data sets, including:

* Analyzing Data Types
* Checking for and removing null values
* Removing all zero values
* Reducing sample size to latest 10 years
* Rounding for relevant numerical data where applicable

**Load Data:** Our group is loading the data into a relational database using Postgres. The final tables can be created on any local machine by running the Jupyter notebook file.

* Not all tables were needed to analyze the purpose above, such as all the raw data or pivoted versions.
* The final production tables loading are:
  + ICO\_Exports\_cal\_yr\_df\_10
  + ICO \_exports\_crop\_yr\_df\_10
  + ICO\_exports\_gross\_open\_stocks\_df\_10
  + Disappearances
  + Domestic\_consumption
  + Imports
  + Indicators
  + Merged\_data\_df
  + Re\_exports
  + Retail\_prices
  + Total\_production

To access our GitHub repository: <https://github.com/delery7/ETL-Project---Group-2>