



ENTREGA DEL PROYECTO FINAL

# Definición fuente de datos

Fuente de datos – Kaggle – Coffee Quality Database from CQI <https://www.kaggle.com/datasets/volpato/coffee-quality-database-from-cqi>

## 1. Descripción de la temática de los datos

Para el presente trabajo se obtuvo la información de la evaluación y cata de diversas cosechas de café realizados por distintos examinadores de la empresa certificadora internacional Coffee Quality Institute a cosechas de café de 36 distintos países. Se obtuvieron dos bases de datos, una correspondiente a la variedad robusta y otra a la variedad arábica de cafés. Cada base presenta información relacionada a las medidas de calidad de la taza de café como Aroma, Sabor, Acidez, Equilibrio, etc. También presenta información sobre el procesamiento del grano, la localidad de cosecha, entre otros.

Se presentan dos bases de datos que contienen reseñas de evaluadores especializados para ambos tipos de café: arábica y robusta, respectivamente. La información anterior se proporciona en cada conjunto de datos. Las primeras cosechas evaluadas se remontan al 9 de abril del 2010 y el horizonte de las evaluaciones se extiende hasta el 11 de enero del 2018.

## 2. Hipótesis

Comprender si características de la localidad de cosecha, procesamiento y variedad de los granos de café afecta la puntuación final de la taza de café. Para esto se evaluará las puntuaciones otorgadas a los cafés durante toda la historia de evaluaciones de CQI y comprobaremos si ciertas combinaciones de las características de los granos conllevan consistentemente a una mayor puntuación. Después comprobaremos que las dimensiones evaluadas en cada taza de café como aroma, sabor, acidez, entre otras, son consistentemente afectadas por las características de la cosecha del grano de café.

Al saber qué características de la cosecha del café y procesamiento conlleva a un mayor puntaje en la certificación del café se podría generar recomendaciones a las empresas cafetaleras respecto a la zona donde desarrollen sus cosechas para optimizar las características del café que quieren obtener y la puntuación mayor que deberá conllevar a mayores ventas y menores costos de experimentación si la empresa desea crear un café con un perfil de taza específico.

## 2. Datasets

A continuación, se adjunta los csv de las bases de datos organizadas y estructuradas utilizadas en el presente trabajo:



Base de Datos - Tablas de evaluaciones de café.zip