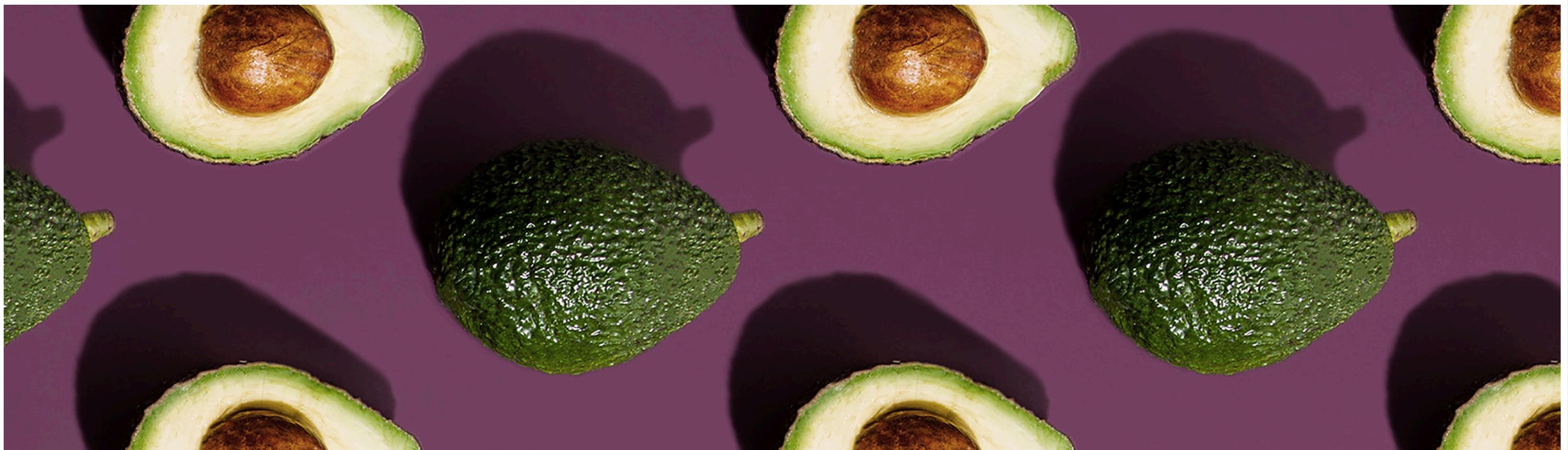


## What's in an Avocado Toast: A Supply Chain Analysis



You find yourself in London, crafting a delectable avocado toast, a dish that has risen dramatically in popularity on breakfast menus since the 2010s. This straightforward recipe requires just five ingredients: a ripe avocado, half a lemon, a generous pinch of salt flakes, two slices of sourdough bread, and a good drizzle of extra virgin olive oil. Most of these ingredients are now staples in grocery stores, and as you will find with this project, that is no small feat!

In this project, you'll conduct a supply chain analysis of three ingredients used in avocado toast using the Open Food Facts database. This database contains extensive, openly-sourced information on various foods, including their origins. Through this analysis, you will gain an in-depth understanding of the complex supply chain involved in producing a single dish.

Three pairs of files are provided in the data folder:

- A CSV file for each ingredient, such as `avocado.csv`, with data about each food item and countries of origin.
- A TXT file for each ingredient, such as `relevant_avocado_categories`, containing only the category tags of interest for that food.

Here are some other key points about these files:

- Some of the rows of data in each of the three CSV files do not contain relevant data for your investigation. In each dataset, you will need to filter out rows with irrelevant data, based on values in the `categories_tags` column. Examples of categories are fruits, vegetables, and fruit-based oils. Filter the DataFrame to include only rows where `categories_tags` contains one of the tags in the relevant categories for that ingredient.
- Each row of data usually has multiple category tags in the `categories_tags` column. There is a column in each CSV file called `origins_tags`, which contains strings for the country of origin of each item.

After completing this project, you'll be armed with a list of ingredients and their countries of origin and be well-positioned to launch into other analyses that explore how long, on average, these ingredients spend at sea.

[Open Food Facts database ↗](#)

```

import pandas as pd

# Load the data
avocado = pd.read_csv('data/avocado.csv', sep='\t')

# Keep only relevant columns
relevant_columns = ['code', 'lc', 'product_name_en', 'quantity', 'serving_size', 'packaging_tags', 'brands', 'brands_tags', 'categories_tags', 'labels_tags', 'countries', 'countries_tags', 'origins', 'origins_tags']

avocado = avocado[relevant_columns]

# Helper function to clean origin names
def clean_origin(origin):
    if pd.isna(origin):
        return None
    # Keep only letters and spaces
    cleaned = re.sub(r'^[A-Za-z ]+', '', origin)
    return cleaned.strip()

def get_top_origin(csv_file, txt_file):
    # Load product data
    df = pd.read_csv(csv_file, sep='\t')

    # Filter for United Kingdom
    df = df[df['countries_tags'].astype(str).str.contains('united-kingdom', case=False, na=False)]

    # Load relevant categories
    with open(txt_file, 'r') as file:
        relevant_cats = [line.strip() for line in file.readlines()]

    # Filter by relevant category tags
    df['categories_tags'] = df['categories_tags'].astype(str)
    df_filtered = df[df['categories_tags'].apply(lambda x: any(cat in x for cat in relevant_cats))]

    # Clean and count origins
    df_filtered['cleaned_origin'] = df_filtered['origins'].dropna().apply(clean_origin)
    top_origin = df_filtered['cleaned_origin'].value_counts().idxmax()
    return top_origin

# Load the data for each product
avocado_df = pd.read_csv('data/avocado.csv', sep='\t')
olive_oil_df = pd.read_csv('data/olive_oil.csv', sep='\t')
sourdough_df = pd.read_csv('data/sourdough.csv', sep='\t')

top_avocado_origin = get_top_origin('data/avocado.csv', 'data/relevant_avocado_categories.txt')
top_olive_oil_origin = get_top_origin('data/olive_oil.csv', 'data/relevant_olive_oil_categories.txt')
top_sourdough_origin = get_top_origin('data/sourdough.csv', 'data/relevant_sourdough_categories.txt')

# Print results
print("Top Avocado Origin:", top_avocado_origin)
print("Top Olive Oil Origin:", top_olive_oil_origin)
print("Top Sourdough Origin:", top_sourdough_origin)

```

Top Avocado Origin: Peru  
 Top Olive Oil Origin: Greece  
 Top Sourdough Origin: United Kingdom