"Which Country Provides the Best Coffee?"

Problem Description

We are a coffee company based in Southampton, UK, and are looking to improve where we source our coffee from. The quality of coffee sourced from our existing suppliers has significantly reduced in recent years, so we want to change suppliers whilst maintaining our reputation as the best artisanal coffee place in town.

We have obtained a wealth of coffee supplier data from the independent Coffee Quality Institute, but we lack the skills required to analyse the data to help us decide where to send our coffee buyers to negotiate a new supply.

Dataset

The data comes from the Coffee Quality Database courtesy of Data Scientist James LeDoux. The original data can be found on James' github. These data were collected from the Coffee Quality Institute's review pages in January 2018.

There is data for both Arabica and Robusta beans, across many countries and professionally rated on a 0-100 scale. All sorts of scoring/ratings for things like acidity, sweetness, fragrance, balance, etc - may be useful for either separating into visualizations/categories or for modeling/recommenders.

The data is not perfect, which may present some challenges for analysis. For example:

- Some of the data is missing, largely indicated by NA (not available) for data in some columns, e.g. "mill", "variety" and "processing_method"

- The units used in some columns is inconsistent, e.g. "bag_weight" contains data in both kilograms and pounds

Goals

In order to know where to send our buyers, we need to analyse the data to understand the following:

- Which country supplies the best coffee given the measures which our customers find important (flavor, aroma, body, uniformity)?

- How many suppliers exist in each country? This would enable us to simplify our supplier chain to a single export market.

- The owner has a preference for the Washed/Wet processing of beans, since it results in a more consistent and predictable flavour profile, so this needs to be taken into account

Existing Code

We have some existing code that was used in the past for previous analyses that was developed by previous employees that left the company, and would like to build on this existing code, whilst ensuring we're able to run the updated code ourselves in the future.

This code generates a visual plot of United States coffee suppliers in terms of flavor and cupper points, and can be found at https://github.com/softwaresaved/coffee-analysis. It is not currently documented, but can be run using the following commands:

```
python3 -m venv venv
source venv/bin/activate
pip install -r requirements.txt
python analyse-coffee.py
```

Deliverables

We would like an answer to the following question: "Which country should we send our buyers to?"

- Provide evidence for your choice, including statistics and visualisations of all relevant variables that have impacted your choice

- The modified code, supplied in a GitHub repository, with short documentation that provides instructions for us to easily re-run your analysis code to reproduce your findings