CASE STUDY 054
Visualizations with Seaborn

Difficulty Level: 1of 3

In this scenario you are a data scientist working to understand specific characteristics of the following dataset to answer so you can solve the challenge questions.

The dataset we are using is a very useful tool and relevant to many industries and organization due to the impact and importance of food production. Overall, the world population is expected to grow to roughly 9.7 billion in the year 2050. Finding solutions for feeding the growing world population has become a hot topic for food and agriculture organizations, entrepreneurs and philanthropists. These solutions range from changing the way we grow our food to changing the way we eat. To make things harder, the world's climate is changing and it is both affecting and affected by the way we grow our food – agriculture.

This dataset provides an insight on our worldwide food production - focusing on a comparison between food produced for human consumption and feed produced for animals. To obtain the file needed you can visit https://www.kaggle.com/dorbicycle/world-foodfeed-production/data from the Food and Agriculture Organization of the United Nations has worldwide food and feed distribution by country and food item from 1961 to 2016 including geocoding.

You are tasked with using Seaborn to create visualizations that will answer the following challenge questions:

1) Graph and use Seaborn to show the largest producers for the top 20 amounts in the year 2013.

    a) Import the CSV

    b) Create a data frame

    c) Get the top 20 in 2013 which will give you the producers.

    d) Build the visualization with Seaborn

2) Find the top 5 countries for coffee for the year 2013

    a) Check the CSV for how coffee is labeled

    b) Group the top 5 countries for 2013

    c) Build the visualization with Seaborn

3) Compare the top 5 countries for coffee from the year 2012 to 2013

    a) Build a separate data frame

    b) Grab the top 5 countries for 2012

c) Build the visualization with Seaborn (look to use a barplot to build them on top of each other)

4) Build a model to visualize the item distribution in the year 2013 and enhance it to the top 1000 values in the 2013 column.

a) Build a new data frame

b) Focus on item distribution and 2013

c) Enhance it to the largest 1000

d) Build the visualization in Seaborn

e) * change the settings so that the visualization is more clear and readable.

a) Print the R-Squared of your model

b) print the comparison between the prediction of the model and the actual data

Good luck!

Difficulty note: this is a fairly straight forward assignment if you have some operational knowledge of Pandas. In addition you have a range of options to use to build the visualizations. Some may work better than others and for the purpose of this we will be aiming to keeping things simple but if you get stuck do not worry, check the clues file for some tips!