DSC 520 - Final Project - Step 1

Emilio Flores

2024-07-11

Introduction - Relationship Between Agricultural Commodity Prices

The agriculture industry in the United States contributes significantly to the country's economy. According to the United States Department of Agriculture ("USDA"), in 2023, US farms contributed more than $200 billion dollars to the economy. All food and agriculture-related industries as a whole contributed more than 1.5 trillion dollars, or 5.6% of the GDP. This industry has attracted several national and international investors in the last couple of decades.

As most of small and medium farms continue to merge and/or be purchased by larger firms, commodity production in the US has become more competitive and lucrative. Nevertheless, specially in California, agricultural productions has been threatened by record-breaking droughts and groundwater pumping restrictions. California land is one of the most valuable and productive assets in the world yet, institutional investors must understand how sensitive certain commodities (almonds, grapes, peaches, etc.) are to drought.

This document will analyze historical precipitation levels in specific regions of California and how certain commodities were affected by the lack and/or abundance of water. Several data science tools will be used to clean and analyze public data sets. This will allow institutional investors to better understand the historical patterns and will facilitate their risk mitigation strategies.

Research Questions

1. What are the most valuable commodities produced in California?
2. What are the historical precipitation levels in specific regions of California?
3. What is the historical production of the most valuable commodities in California?
4. What is the relationship between precipitation and commodity production?
5. What commodities are less affected by drought?
6. What commodities are the most affected by drought?

Approach

Several government organizations track precipitation levels in specif regions of California. Also, the USDA tracks the commodity prices of the most important commodities produced in the same state. At least 10 years of historical data will be gathered, merged and analyzed using R. The analysis will demonstrate the relationship between precipitation and commodity prices in California, if any.

How your approach addresses the problem

This approach will partially address the problem. Precipitation/drought does have a significant impact on commodity prices but it is not the only predictor. Commodity prices are also affected by international affairs, inflation, national regulations, etc.

Data

- National Weather Service https://www.weather.gov/wrh/Climate?wfo=hnx:
  - This site summarizes the total precipitation accumulation in inches, per month and/or year in several regions of the country. The data set will input "T" in any month where the total precipitation was equal to or lower than 0.005 inches.
- California Department of Agriculture https://www.cdfa.ca.gov/Statistics/:
  - This website provides for several reports that have tracked historical production and prices of several commodities produced in the state.
- United States Department of Agriculture - National Agricultural Statistics Service https://www.nass.usda.gov/:
  - This site has several data set on agricultural production and historical prices.

Required Packages

I will need dplyr, scales, ggplot2, readxl, and Metrics.

Plots and Table Needs

I will first gather all data and create one or several data frames to facilitate analysis and the creation of visualizations. I will mostly use histograms, scatter plots and qq-plots to visualize the distributions of my data.

Questions and Future Steps

Right now I know how to import excel files into R and merge several columns into a data frame. I also now how to transform the data to clean and/or extract empty cells. I also know how to calculate new columns and add them to data sets. I also know how to run regression analysis and how to measure a model(s) accuracy.