

Data Visualization on Honey Production dataset using Seaborn and Matplotlib libraries.

Domain: Food and agriculture

Context:

In 2006, a global concern was raised over the rapid decline in the honeybee population, an integral component to American honey agriculture. Large numbers of hives were lost to “Colony-Collapse-Disorder”, a phenomenon of disappearing “worker-bees” causing the remaining “hive-colony” to collapse. Speculation around the cause of this disorder points to hive-diseases and pesticides harming the pollinators, though no overall consensus has been reached. Twelve years later, some industries are observing recovery but the American honey industry is still largely struggling. The U.S. used to locally produce over half the honey it consumes per year. Now, honey mostly comes from overseas, with 350 of the 400 million pounds of honey consumed every year originating from imports. This dataset provides insight into honey production supply and demand in America by state from 1998 to 2012.

Objective:

The Goal is to use Python visualization libraries such as seaborn and matplotlib to investigate the data and get some useful conclusions.

Attribute Information:

Sln.	Attribute	Description
1.	Numcol	Number of honey producing colonies.
2.	Yield percol	Honey yield per colony. (Unit is pounds)
3.	Total prod	Total production (numcol x yieldpercol). (Unit is pounds)
4.	Price per lb	Refers to average price per pound based on expanded sales. Unit is dollars.
5.	Prod value	Value of production (total prod x priceperlb). Unit is dollars.
6.	Stocks	Refers to stocks held by producers. Unit is pounds
7.	Year	Calendar year.
8.	State	Different states' names.

Questions:-

1. Import required libraries and read the dataset.
2. Check the first few samples, shape, info of the data and try to familiarize yourself with different features.
3. Display the percentage distribution of the data in each year using the pie chart.
4. Plot and Understand the distribution of the variable "price per lb" using displot, and write your findings.
5. Plot and understand the relationship between the variables 'numcol' and 'prodval' through scatterplot, and write your findings.
6. Plot and understand the relationship between categorical variable 'year' and a numerical variable 'prodvalue' through boxplot, and write your findings.
7. Visualize and understand the relationship between the multiple pairs of variables throughout different years using pairplot and add your inferences. (use columns 'numcol', 'yield percol', 'total prod', 'prodvalue','year')
8. Display the correlation values using a plot and add your inferences. (use columns 'numcol', 'yield percol', 'total prod', 'stocks', 'price per lb', 'prodvalue')