Berries Report

MA615

10/16/2020

Introduction

The goal of this data analysis is to examine how different aspects of the production of raspberries are related to the their value. The data was collected from a USDA database and contains a variety of information including production methods, chemicals used, etc. on three type of berries: raspberries, strawberries, and blueberries. For this analysis, raspberries were the only berry of interest.

Data Cleaning

The data cleaning process consisted of several steps. The first step was to eliminate redundancy by removing any variables with only one unique value. This created 8 variables which are displayed below.

Year	Period	State	Commodity	Data Item
2019	MARKETING YEAR	CALIFORNIA	BLUEBERRIES	BLUEBERRIES, TAME - PRICE R
2019	MARKETING YEAR	CALIFORNIA	BLUEBERRIES	BLUEBERRIES, TAME, FRESH M
2019	MARKETING YEAR	CALIFORNIA	BLUEBERRIES	BLUEBERRIES, TAME, PROCESS
2019	MARKETING YEAR	CALIFORNIA	RASPBERRIES	RASPBERRIES - PRICE RECEIVE
2019	MARKETING YEAR	CALIFORNIA	RASPBERRIES	RASPBERRIES, FRESH MARKET
2019	MARKETING YEAR	CALIFORNIA	RASPBERRIES	RASPBERRIES, PROCESSING - P

Next, the dataset is filtered such that only raspberries are the only berry and years are the only period that are included. The columns "Commodity" and "Period" now contain only 1 unique value and can be eliminated from the dataset.

Year	State	Data Item	Domain
2019	CALIFORNIA	RASPBERRIES - ACRES HARVESTED	TOTA
2019	CALIFORNIA	RASPBERRIES - PRODUCTION, MEASURED IN LB	TOTA
2019	CALIFORNIA	RASPBERRIES - YIELD, MEASURED IN LB / ACRE	TOTA
2019	CALIFORNIA	RASPBERRIES, BEARING - APPLICATIONS, MEASURED IN LB	CHEMI
2019	CALIFORNIA	RASPBERRIES, BEARING - APPLICATIONS, MEASURED IN LB	CHEMI
2019	CALIFORNIA	RASPBERRIES, BEARING - APPLICATIONS, MEASURED IN LB	CHEMI

The next step was to clean up the Data Item, and Domain Category variables by splitting each into several columns. In total, these two variables were split into 8 new columns: Berry, Type, Market, Production, Measure, Average, Chemical_Type, and Chemicals. Berry and Average were eliminated because each contained only one unique value.

Year	State	Type	Market	Production	Measure	
2019	CALIFORNIA			ACRES HARVESTED		
2019	CALIFORNIA			PRODUCTION	MEASURED IN LB	
2019	CALIFORNIA			YIELD	MEASURED IN LB / ACRE	
2019	CALIFORNIA	BEARING		APPLICATIONS	MEASURED IN LB	
2019	CALIFORNIA	BEARING		APPLICATIONS	MEASURED IN LB	
2019	CALIFORNIA	BEARING		APPLICATIONS	MEASURED IN LB	

Finally, NA's were removed from the Value column. These were displayed in the original dataset as (D) and (NA).

Year	State	Type	Market	Production	Measure	
2019	CALIFORNIA			ACRES HARVESTED		
2019	CALIFORNIA			PRODUCTION	MEASURED IN LB	
2019	CALIFORNIA			YIELD	MEASURED IN LB / ACRE	
2019	CALIFORNIA	BEARING		APPLICATIONS	MEASURED IN LB	
2019	CALIFORNIA	BEARING		APPLICATIONS	MEASURED IN LB	
2019	CALIFORNIA	BEARING		APPLICATIONS	MEASURED IN LB	

Exploratory Data Analysis

As mentioned previously, the goal of this analysis is to examine. . . To this end, I. . .

Conclusion