**7PAM2000 Applied Data Science 1**

**Assignment 1: Visualisation**

**Name:JAYULKUNDEKAR LAKHAN KRISHNA**

**Student ID:21081427**

Dataset is downloaded from the below link:

<https://www.kaggle.com/datasets/jarredpriester/california-wine-production-19802020>

**GitHub link:**

**Dataset description:**

California is known for its wine, and throughout the years, a lot of wine has been made there. The acres harvested, yield, quantity produced, and total price for each county in California from 1980 through 2020 are included in this dataset.

**1.line Graph**

For the plot, I have used the years and harvested acres columns in the dataset. I have used a Pandas dataframe to slice the required counties and plot the graph. The below graph shows the change in harvest in various counties from 1980 to 2020.

Chart, line chart

Description automatically generated

Figure : Line graphs representing harvested acres

In the above graph, I have observed that the Lake County harvest has increased significantly from 1980 to 2020. Harvests have steadily increased under the kings and Amador counties. The alameda did not show a major increase in harvest size.

**2. Bar graph**

For the bar plot I have used Yield(unit/acre) column and year column. I have used pandas dataframe to slice the required counties and plot the graph. The below graph shows the difference in yield in various counties from year 2010 to 2020. I have used bar plot because, Bars in a bar plot indicate discrete values of a categorical variable. On one axis of the graph, values are exhibited, while on the other, levels are shown. Each metric value "owns" one bar, and the length of each bar is proportional to its value. Bars are consistently drawn from the same starting point to facilitate easy value comparisons.

Chart, bar chart

Description automatically generated

Figure : bar plot representing yield of crop from 2010 to 2020

The above graphs shows that the yield in kings county is higher in each year compared to the other two counties. The lake county is in the second position and the Amador County is the last. From the line graph and bar graph I have observed that lake county harvests more crop but yields less crop compared to the kings county.

**3. Pie chart.**

For the Pie chart I have used production, year columns from the dataset. For the plotting the graph I have developed a plot\_production function produces a pie chart of top selected producing counties in the selected year, the function takes year, data and num\_county to be selected as the parameters.

Chart, pie chart

Description automatically generated

Figure : pie chart of Production of wine in year 2020

I have used pie chart as it is useful for representing the comparison of the data in a slice. A pie chart, also known as a circular bar graph or radial bar graph, is a graphical representation of a categorical variable showing the breakdown of a whole across its levels. Each criterion value is represented by an individual slice of the circle, and the relative size of the slices indicates the proportion of the whole that each criterion occupies.

From the above pie plot, I observed that Fresno, San Joaquin, Madera counties produced 50% of the total production in year 2020. The other counties kern, Sacramento, yolo, Sonoma and Merced contributed to almost 28% of the production.