**Tableau Project**

**Avocado Sales Analysis**

**Group 3**

Aashish Kurimilla – 0833509

Surbhi Sharma - 0828888

Inderpreet Singh -0836676

Sai Anudeep Janapareddy - 0836698

Abhishek Gudapati – 0816514

**Introduction of Project**

**Objective**: The objective of this dataset is to evaluate avocado sales dynamics, market trends, and regional preferences to inform strategic decision-making in the avocado industry. The motivation behind selecting this topic arises from the significance of the avocado market and the necessity to comprehend consumer behaviour, pricing dynamics, and regional variations impacting avocado sales. Given the growing demand for healthy food options and the avocado's popularity in various culinary applications, understanding sales trends and factors influencing purchasing decisions is crucial for stakeholders in the avocado supply chain. This analysis seeks to uncover insights that can drive pricing strategies, marketing initiatives, and distribution efforts, ultimately contributing to the sustainable growth and competitiveness of businesses operating in the avocado market.

**Goals:**

* **Understand Sales Trends**: Analyse historical sales data to identify patterns, trends, and seasonality in avocado sales.
* **Explore Price Variations**: Investigate the distribution of average prices for different types of avocados to understand price dynamics.
* **Evaluate Regional Performance**: Assess sales performance across different regions and identify regions with high or low sales volumes.
* Explore the relationship between avocado characteristics (such as type, size, and quality) and sales performance.
* Provide insights into the implications of market trends and consumer behaviour on pricing strategies, marketing initiatives, and distribution channels in the avocado industry.
* The fluctuation in avocado sales over time, comparing average sales before, during, and after the COVID-19 pandemic.

**Benefits**:

* This dataset analysis will offer valuable insights for stakeholders in the avocado industry, including growers, distributors, retailers, and consumers, to make informed decisions regarding avocado production, marketing, and consumption practices.
* By understanding the market dynamics, pricing trends strategies to meet consumer demand effectively, enhance profitability, and promote sustainable practices in avocado cultivation and distribution.
* Additionally, these insights can help policymakers in formulating regulations and policies to support the avocado industry's growth while ensuring environmental sustainability and equitable distribution of resources.
* In summary, utilizing the results of this analysis can help create a more reliable and sustainable avocado supply chain that will benefit both industry stakeholders and consumers, and regional preferences influencing avocado sales, stakeholders can optimize their alike.

**Actual Dataset: Avocado Sales and Production**

**Dataset Link:** <https://www.kaggle.com/datasets/vakhariapujan/avocado-prices-and-sales-volume-2015-2023>

**Source**: The dataset is sourced from various sources within the avocado industry, including avocado growers, distributors, retailers, and market research firms. It may also be supplemented by publicly available data on avocado sales and consumption.

**Funding Source**: The funding source for data collection is likely from private industry sources, market research firms, and governmental agricultural departments responsible for collecting agricultural data.

**Variables**:

• Date: The date of avocado sales transactions.

• Average Price: The average price of avocados in the transaction.

• Total Volume: The total volume of avocados sold in the transaction.

• Type: The type or variety of avocados sold (e.g., Hass, Fuerte, etc.).

• Region: The region where the avocados were sold or produced.

• Total Bags: The total number of bags containing avocados in the transaction.

• Small Bags: The number of small bags containing avocados in the transaction.

• Large Bags: The number of large bags containing avocados in the transaction.

• X Large Bags: The number of extra-large bags containing avocados in the transaction.

**Utility and Limitations:**

**Utility**: This dataset offers comprehensive insights into avocado sales metrics, including pricing, volume, and distribution across different regions and types. It enables analysis of market trends, pricing dynamics, and regional preferences, aiding stakeholders in optimizing their strategies for production, marketing, and distribution.

**Limitations**: The dataset may lack coverage of certain niche avocado varieties or regions, potentially limiting the comprehensiveness of the analysis. Additionally, while it provides valuable information on sales metrics, it may not capture factors such as consumer preferences, marketing campaigns, or external market forces that could influence avocado sales.

**Rejected Dataset**: Avocado Sales

**Dataset Link**: <https://www.kaggle.com/datasets/smokingkrils/avacado-price-prediction>

**Source**: The dataset is sourced from Kaggle, a popular platform for sharing datasets and conducting data analysis. However, the specific origin or methodology of data collection is not provided in the dataset description.

**Variables:**

* Date: The date of avocado sales transactions.
* Average Price: The average price of avocados in the transaction.
* Total Volume: The total volume of avocados sold in the transaction.
* 4046, 4225, 4770: These seem to represent specific avocado PLU (Product Lookup Code) numbers, possibly indicating different varieties or categories of avocados.
* Total Bags: The total number of bags containing avocados in the transaction.
* Small Bags, Large Bags, Large Bags: These variables represent the number of bags of different sizes containing avocados.
* Type: The type or variety of avocados sold (e.g., conventional, or organic).
* Year: The year of the transaction.
* Region: The region where the avocados were sold or produced.

**Utility and Limitations:**

**Utility**: This dataset provides valuable information for analysing avocado sales trends, pricing dynamics, and regional preferences. It allows for the exploration of factors influencing avocado sales, such as pricing strategies, seasonal variations, and consumer preferences.

**Limitations**: However, there are several limitations to consider:

* **Incomplete Information:** Some variables, such as PLU numbers and bag sizes, may not be adequately explained, making interpretation challenging.
* **Data Quality**: The dataset may contain missing values, errors, or inconsistencies that could affect the analysis and interpretation of results.
* **Limited Context**: The dataset does not provide information on external factors that could influence avocado sales, such as market competition, economic conditions, or advertising campaigns.
* **Lack of Documentation:** The lack of detailed documentation or background information on data collection methods and sources may hinder the understanding and reliability of the dataset.

**Reason of Rejection:**

The avocado sales dataset, while providing valuable insights into avocado sales trends and regional preferences, is criticized for its incomplete information, data quality concerns, limited context, and lack of documentation. These issues make it unsuitable for robust analyses without significant data preprocessing and additional external information. The dataset's limitations, such as a lack of explanation for variables like PLU numbers and bag sizes, may lead to misunderstandings or incorrect analyses. In addition, the dataset provides old data information till 2015.

**Total volume sales by Region and Type**

A graph of a number of people

Description automatically generated with medium confidence

* The image is a bar graph titled "Total volume sales by Region and Type."
* It compares the total sales volume across different regions, distinguishing between conventional (blue) and organic (orange) sales.
* Conventional sales generally have higher volumes than organic sales. One region stands out with significantly higher sales volume than others.
* The graph is interactive, allowing customization of data display by selecting regions or adjusting parameters.

**Average Sales across the time period**

A graph of lines with text

Description automatically generated with medium confidence

* The chart is a line graph titled "Average Sales across the time period," with the x-axis labelled "Month of Date" and the y-axis labelled "Avg Sales($K)" indicating average sales in thousands of dollars. The x-axis lists months from January 2019 (2019-01) through December 2021 (2021-12), showing a three-year span.
* There are three lines on the graph, each representing average sales during different time periods: pre-COVID, during COVID, and post-COVID.
* These lines are color-coded for clarity: the pre-COVID line is blue, the during COVID line is orange, and the post-COVID line is Gray.
* The pre-COVID line shows fluctuations in sales but generally trends upward over time until early 2020, which is likely when the COVID-19 period starts. During COVID, the orange line shows a significant drop in sales starting around March 2020, with a sharp dip and recovery, followed by a somewhat volatile but generally upward trend. Post-COVID, the Gray line indicates a return to fluctuating sales similar to the pre-COVID pattern, but with overall higher average sales than the pre-COVID period.
* There is a black dot on the "During COVID" line around April 2020, which may indicate a specific data point of interest, such as the lowest sales during the COVID period or an outlier.
* The chart also includes a legend in the upper right corner, which helps identify which colour corresponds to each period.
* At the bottom, there is a note that says, "The trend of average Sales for Month. Colour shows details about Covid Times." This suggests that the primary purpose of the chart is to compare the average monthly sales across the three different periods relative to the COVID-19 pandemic.

**The relation between the Average Price of Avocado and the Average Total Volume sold.**

A graph with blue and orange lines

Description automatically generated

* The image shows a line chart titled "The relation between the Average Price of Avocado and the Average Total Volume sold." The chart plots two variables over time, from 2015 to approximately the start of 2024.
* There are two y-axes on the chart:

1. The left y-axis represents the Average Total Volume sold, measured in units (the numbers are in the thousands, indicated by 'K', ranging from 0 to 1200K).
2. The right y-axis corresponds to the Average Price in dollars, ranging from $0.5 to $1.75.

* The x-axis represents the Month of Date, with each tick presumably representing a month from 2015 through the beginning of 2024.
* Two lines represent the different datasets:

1. The blue line represents the Average Total Volume sold. This line fluctuates significantly over time, with peaks and troughs indicating variability in the volume of avocados sold. The volume appears to have seasonal patterns with regular intervals of increases and decreases.
2. The orange line represents the Average Price of avocados. This line also shows fluctuation but with a less pronounced seasonal pattern than the volume sold. The price peaks at times when the volume dips and vice versa, suggesting an inverse relationship between price and volume sold.

* At the bottom of the chart, a legend indicates the blue line as "Avg. Total Volume" and the orange line as "Avg. Price."
* A note at the bottom of the chart states: "The trends of Avg. Total Volume and Avg.Price are for Date Month. Colour shows details about Avg. Total Volume and Avg.Price."
* Overall, the chart is intended to visualize the relationship between the average price of avocados and the volume sold, suggesting that these two variables may be inversely related over time.

**Sales by Bag Size**

A pie chart with a blue and orange color

Description automatically generated

* The image is a pie chart titled "Sales by Bag Size." It represents the distribution of sales across three different bag size categories: Small Bags, Large Bags, and X-Large Bags.
* The chart is divided into three segments, each coloured differently to represent a specific bag size category.
* Here is the breakdown of the sales by bag size category:

1. Small Bags (orange segment): Account for 79.98% of total sales.
2. Large Bags (blue segment): Constitute 17.94% of total sales.
3. X-Large Bags (Gray segment): Represent 2.08% of total sales.

* The chart provides a clear visual representation of how sales are distributed among the various bag sizes, with small bags being the most popular choice among consumers, followed by large bags and then X-Large bags.
* The legend on the right side of the chart helps identify each bag size category by colour, making it easier to understand the distribution briefly.

**Average sales and the total number of bags sold**

A graph of different colored bars

Description automatically generated with medium confidence

* The image shows a combination chart with two types of data visualization: a bar chart and a line chart, representing average sales and the total number of bags sold according to different seasons.
* The chart has a horizontal x-axis representing the seasons and two vertical y-axes, one on the left and one on the right, representing different scales for the two datasets.
* The bar chart displays the average sales for each season with four bars, each corresponding to a different season: Spring, Summer, Fall, and Winter. The colours of the bars are blue for Spring, orange for Summer, Gray for Fall, and dark blue for Winter. The scale on the left y-axis corresponds to the average sales, which ranges from 0 to 1100K (presumably in a currency unit).
* The line chart overlays the bar chart and shows the total number of bags sold, with points corresponding to each season. The line is coloured black and has markers at each data point. The scale on the right y-axis corresponds to the total bags sold, which ranges from 0 to 220K.
* From the bar chart, we can observe that the average sales are highest in Spring (just below 1100K), followed by Fall (approximately 900K), then Summer (around 800K), and lowest in Winter (just above 700K).
* From the line chart, we can see that the total number of bags sold is highest in Spring (approximately 210K), followed by Summer (just below 200K), then Fall (around 180K), and lowest in Winter (around 160K).
* The title of the chart is "Average sales and Total bags sold according to Seasons." There is also a note at the bottom of the chart which reads, "The trend of average sales and average of Total Bags for Season. For average of Sales - Colour shows details about Season."
* Overall, the chart suggests that both the average sales and the total number of bags sold tend to be higher in the Spring season, with a general decrease in sales and volume as the year progresses towards Winter.

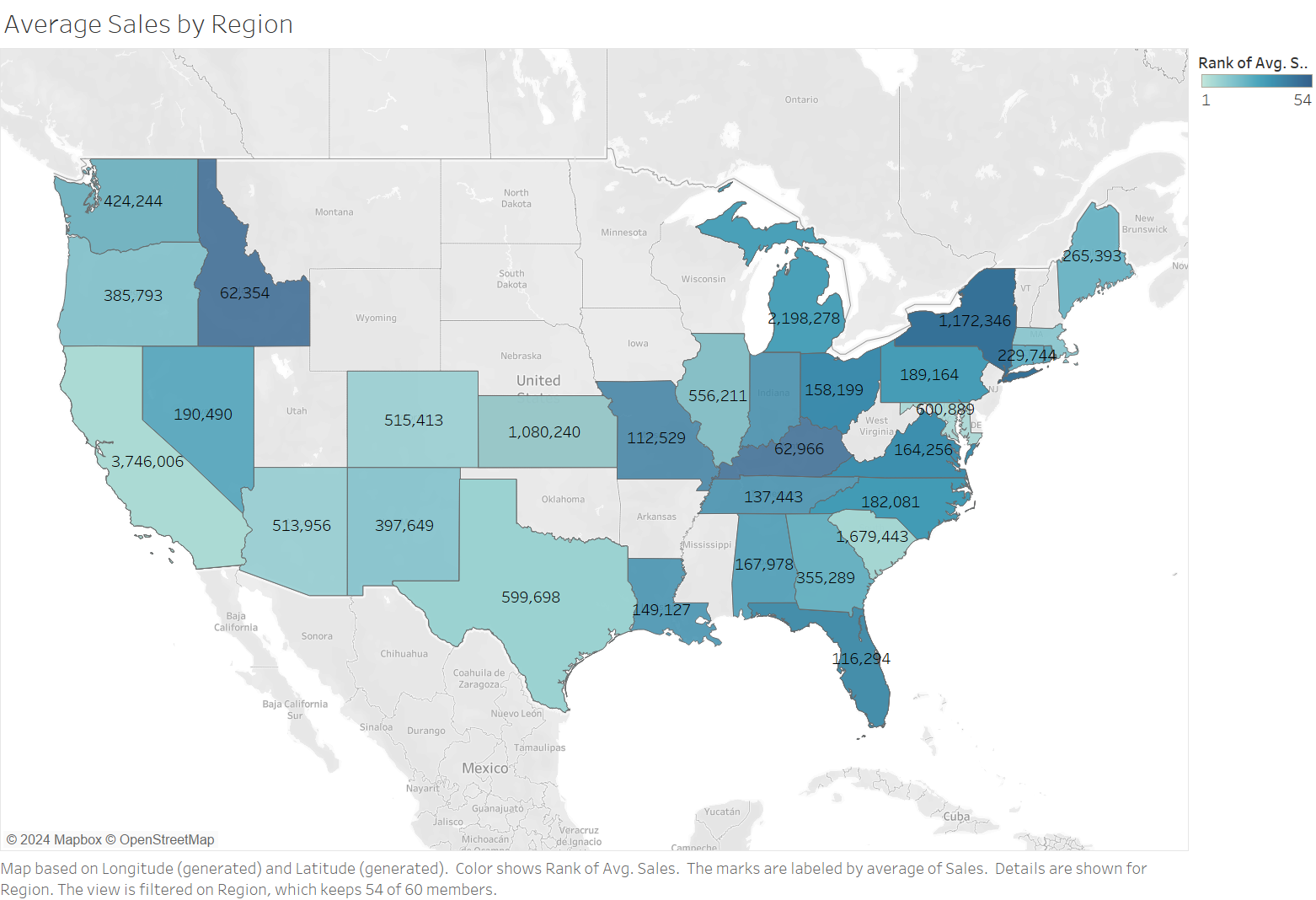
**Region-wise sales of different Avocado types**

A graph of blue and orange bars

Description automatically generated

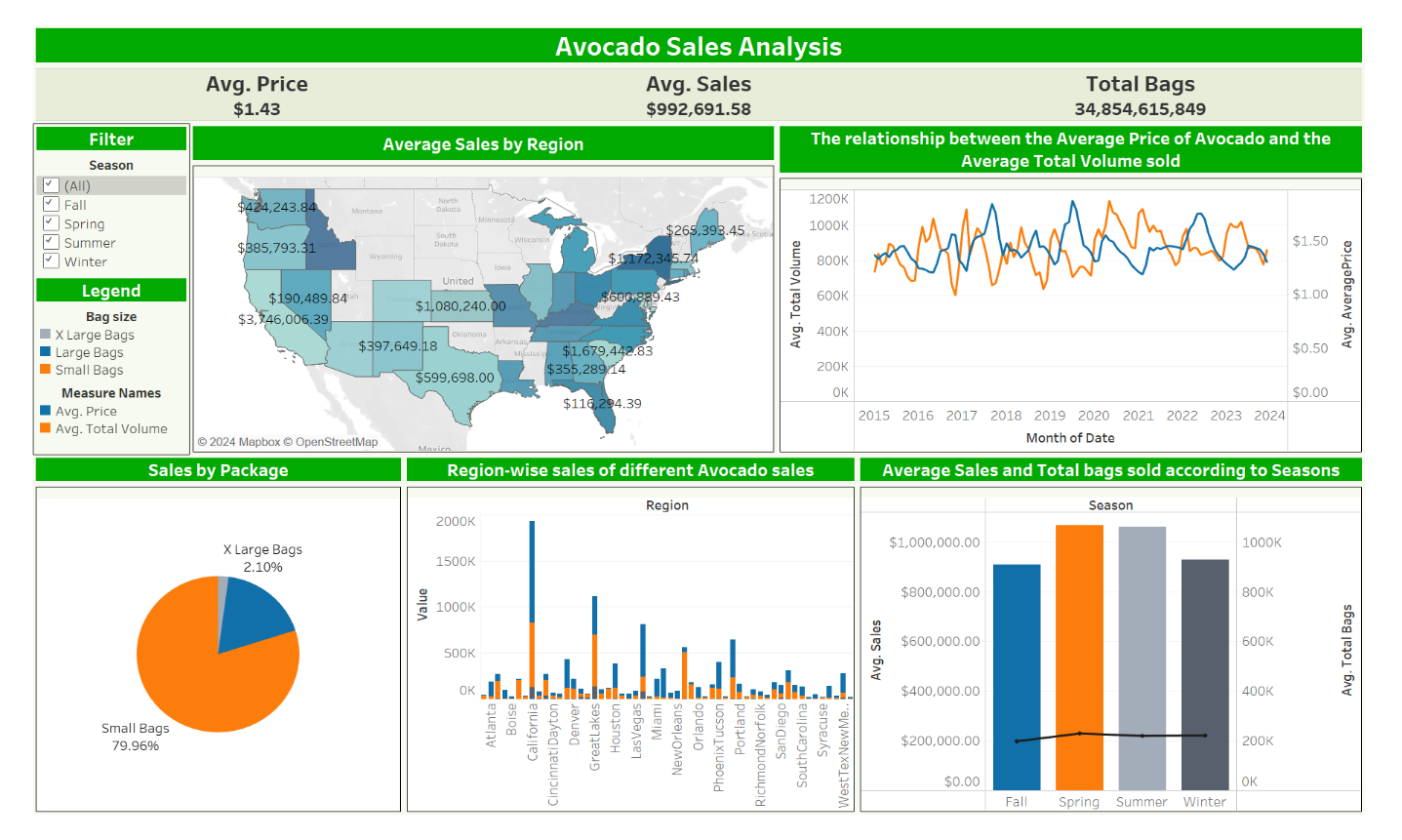
* The image is a clustered bar chart representing region-wise sales of different Avocado types. There are three types of Avocados represented by different colours: PLU4046 in blue, PLU4225 in orange, and PLU4770 in Gray.
* Each cluster of bars corresponds to a specific region, with the x-axis showing the regions and the y-axis representing the sales volume.
* The chart shows the sales volume for each Avocado type in various regions. PLU4046 has the highest sales volume in most regions, followed by PLU4225, and then PLU4770, which has the lowest sales volume compared to the other two types.
* The legend in the top right corner helps identify each Avocado type by colour, and there is a note indicating that the data is filtered on region, including average sales for each Avocado type.
* Overall, the chart provides a visual comparison of the sales volumes of different Avocado types across multiple regions, highlighting the dominance of PLU4046 in most regions.

**Average sales by region**



* The image is a map of the United States colour-coded to show average sales by region. Darker shades of blue indicate higher average sales, while lighter shades represent lower average sales. Each state is labelled with a numerical value representing its average sales figure.
* The map includes a colour key on the right side, showing the rank of states according to sales. It highlights regional differences in sales, with the West Coast and Northeast showing higher averages compared to the Midwest and South regions.
* Notable sales figures include California at $653,539, Washington at $397,698, New York at $182,634, Texas at $153,366, Florida at $137,434, and Illinois at $151,413.
* The map appears interactive, as indicated by the cursor icon over Kentucky with an average sales figure of $122,559. It serves as a visual tool for comparing average sales across U.S. states but lacks additional context such as the time covered or specific product/service sales.

**Dashboard:**



**Conclusion:**

Avocados have experienced a significant increase in popularity and demand, driven by their health benefits, culinary versatility, and appeal to diverse consumer segments. The sales of large packages are not that successful as compared to smaller packages, a family of four prefers small packages containing 5-6 avocados, rather than large packages with 13-15 avocados. Avocado sales are directly linked to the population distribution, as we see in the graph the sales are higher in states with high population density, so cities like California, New York, and Los Angeles have very high sales. The company should consider selling by-products such as oils, sauces, crushes etc, as well, as we don’t see drastic sales increase with a huge difference between years. If the company sets up a farm in the USA, that will benefit more as a large share of avocados are imported into the USA.

**Team Members Contributions:**

* **Aashish Kurimilla** – Total volume sales by Region and Type, The relation between the Average Price of Avocado and the Average Total Volume sold and Dashboard
* **Surbhi Sharma** - Average Sales across the period, Region-wise sales of different Avocado types and Report
* **Inderpreet Singh** -Report and PPT
* **Sai Anudeep Janapareddy** - Sales by Package, Average sales and Total bags sold according to Seasons, KPI and Dashboard
* **Abhishek Gudapati** – Average Sales by Region, Report and PPT