**TASK 1:**

**SUMMARIZING THE DATASET:**

**Observation Of the Data**

| | Id | Price | Age_08_04 | Mfg_Month | Mfg_Year | KM | HP | Met_Color | Automatic | CC |
|---|---|---|---|---|---|---|---|---|---|---|
| **count** | 1436.00 | 1436.00 | 1436.00 | 1436.00 | 1436.00 | 1436.00 | 1436.00 | 1436.00 | 1436.00 | 1436.00 |
| **mean** | 721.56 | 10730.82 | 55.95 | 5.55 | 1999.63 | 68533.26 | 101.50 | 0.67 | 0.06 | 1576.86 |
| **std** | 416.48 | 3626.96 | 18.60 | 3.35 | 1.54 | 37506.45 | 14.98 | 0.47 | 0.23 | 424.39 |
| **min** | 1.00 | 4350.00 | 1.00 | 1.00 | 1998.00 | 1.00 | 69.00 | 0.00 | 0.00 | 1300.00 |
| **25%** | 361.75 | 8450.00 | 44.00 | 3.00 | 1998.00 | 43000.00 | 90.00 | 0.00 | 0.00 | 1400.00 |
| **50%** | 721.50 | 9900.00 | 61.00 | 5.00 | 1999.00 | 63389.50 | 110.00 | 1.00 | 0.00 | 1600.00 |
| **75%** | 1081.25 | 11950.00 | 70.00 | 8.00 | 2001.00 | 87020.75 | 110.00 | 1.00 | 0.00 | 1600.00 |
| **max** | 1442.00 | 32500.00 | 80.00 | 12.00 | 2004.00 | 243000.00 | 192.00 | 1.00 | 1.00 | 16000.00 |

1. Price: The average car price is $10,730.82, with prices ranging from $4,350 to $32,500. A standard deviation of $3,626.96 indicates a significant variation in prices.

2. Age: The average age of the cars (as of August 2004) is 55.95 months, with most cars between 44 and 70 months old.

3. Manufacturing Year: The average manufacturing year is 1999.63, with cars manufactured between 1998 and 2004, and 75% of cars manufacured before 2001.

4. Kilometers (KM): The average mileage is 68,533 km, with cars ranging from as low as 1 km to 243,000 km. The interquartile range shows that most cars have driven between 43,000 and 87,000 km.

5. Horsepower (HP): The cars have an average horsepower of 101.5, ranging from 69 to 192 horsepower, indicating a variety of vehicle power levels.

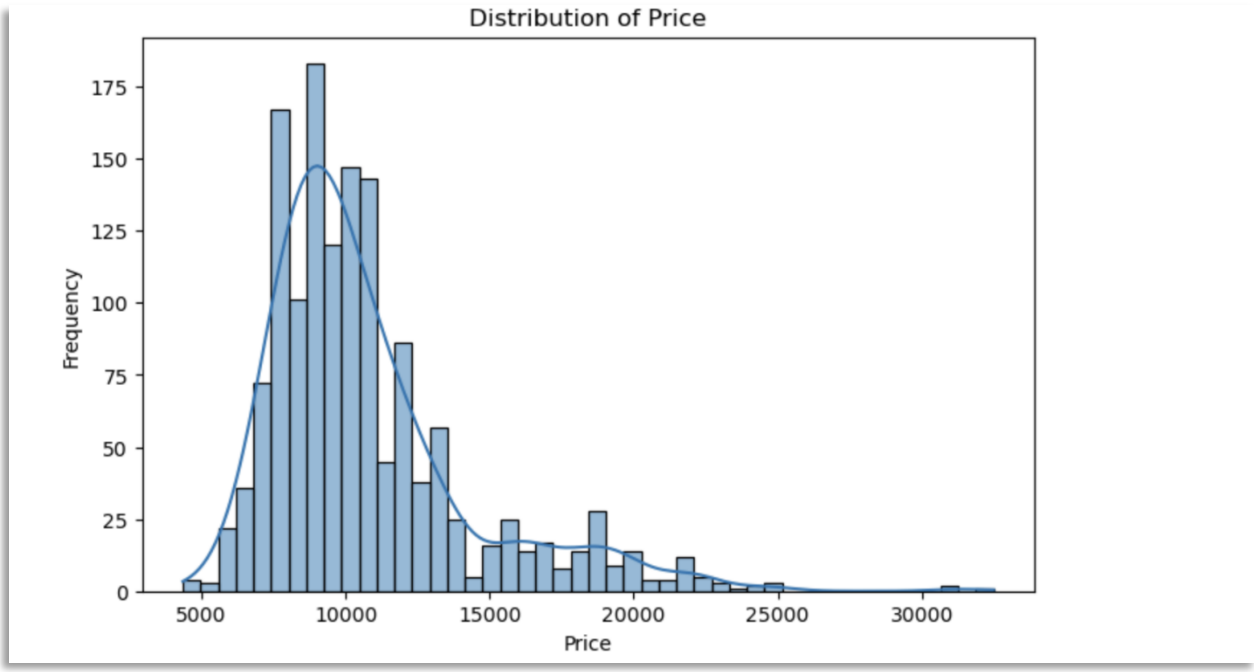
6. Metallic Color: 67% of the cars have metallic paint, based on the binary variable.

7. Automatic Transmission: Only 6% of the cars are automatic, as shown by the low average of this binary feature.

8. Engine Capacity (CC): The average engine capacity is 1,576.86 cc, with most cars having engine sizes between 1,400 and 1,600 cc.

**DISTRIBUTION OF PRICE:**

```
Mean of Price: 10730.824512534818
Median of Price: 9900.0
```

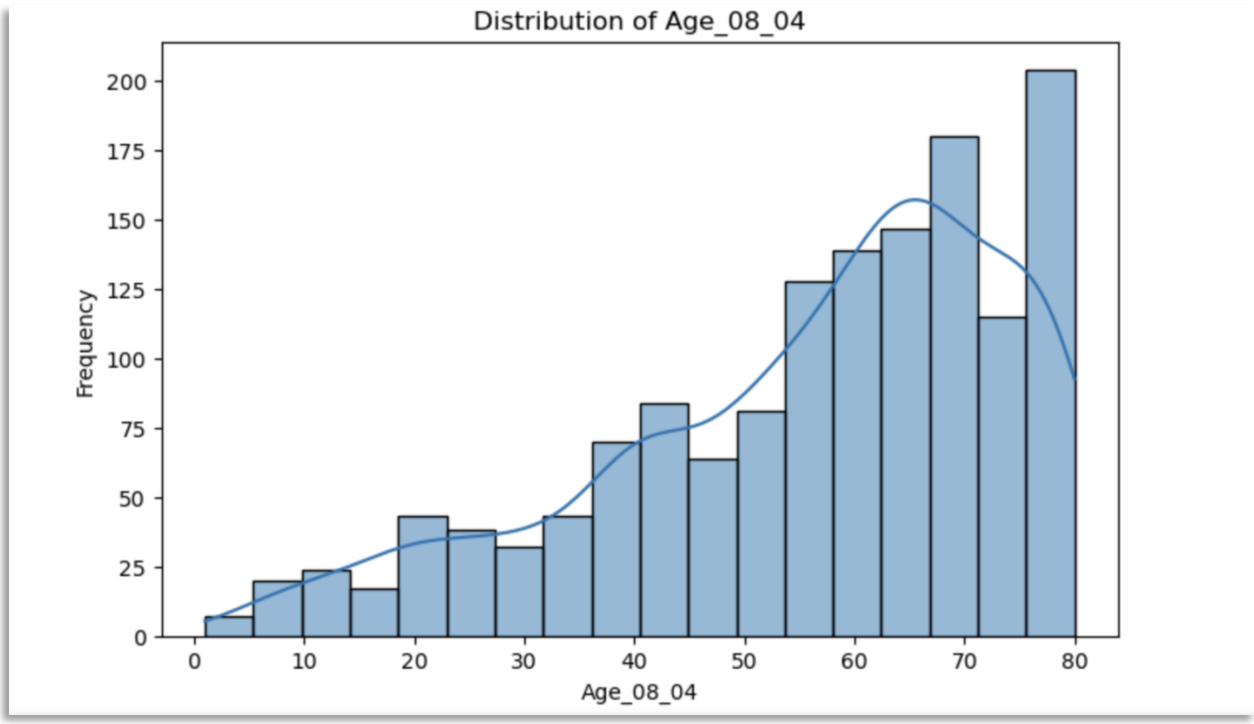



**INFERENCE:**
The median price is 9,900, which suggests that half of the cars are priced below 9,900 and half are priced above.
This means that a significant portion of the cars are priced lower than the average price, i.e., 10730. Therefore, most cars are likely to be more affordable than the average is suggesting.
The average price could have been influenced by a few high-priced cars(outliers), making it appear that cars are generally more expensive than they are.

We can also look at Skewness to understand the asymmetry of the distribution.
In this Mean > Median therefore it means it's Right Skewed – implying that there are many cheap cars but only few expensive ones.

**DISTRIBUTION OF AGE:**

```
Mean of Age: 55.94707520891365
Median of Age: 61.0
```

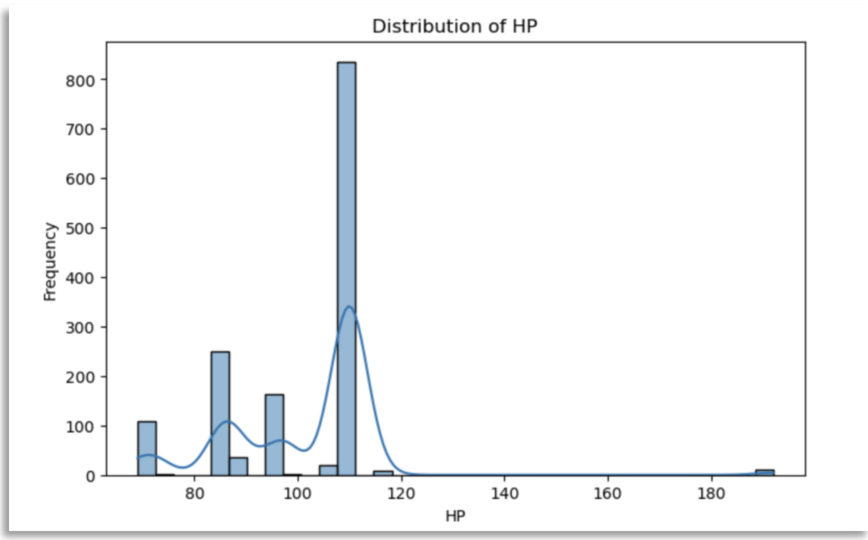



**INFERENCE:**

In this Mean < Median which means it's Left Skewed – implying that there are a few newer cars in the dataset but the majority of cars are older**.**

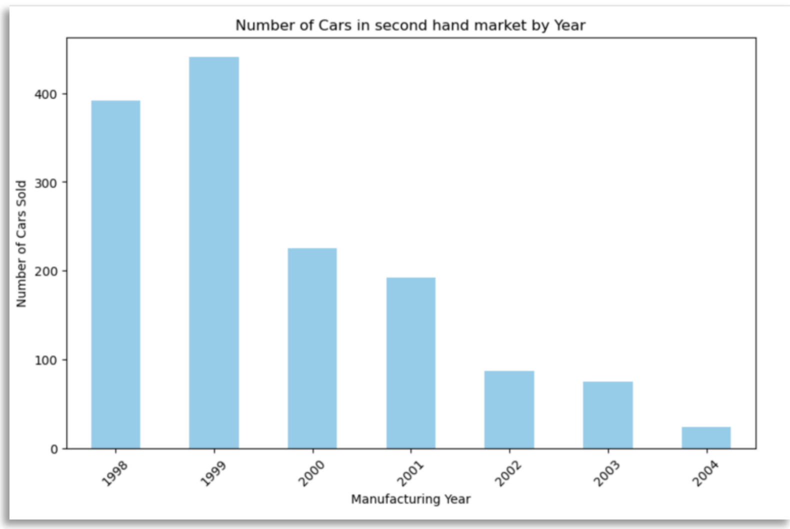
**DISTRIBUTION OF HP:**

```
Mean of HP: 101.50208913649026
Median of HP: 110.0
```



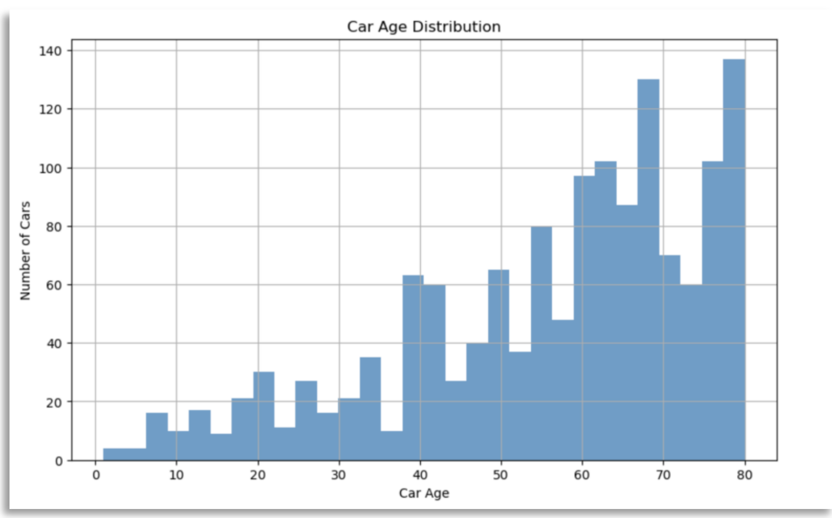

**INFERENCE:**
So, from the above we can assume that Left Skewed - Mean < Median but it is also important that we look at the graph and the skewness calculation which is a better indicator of the distribution shape. So, though the mean < median, it follows a Right skewed distribution.

**UNDERSTANDING CARS BY YEAR:**



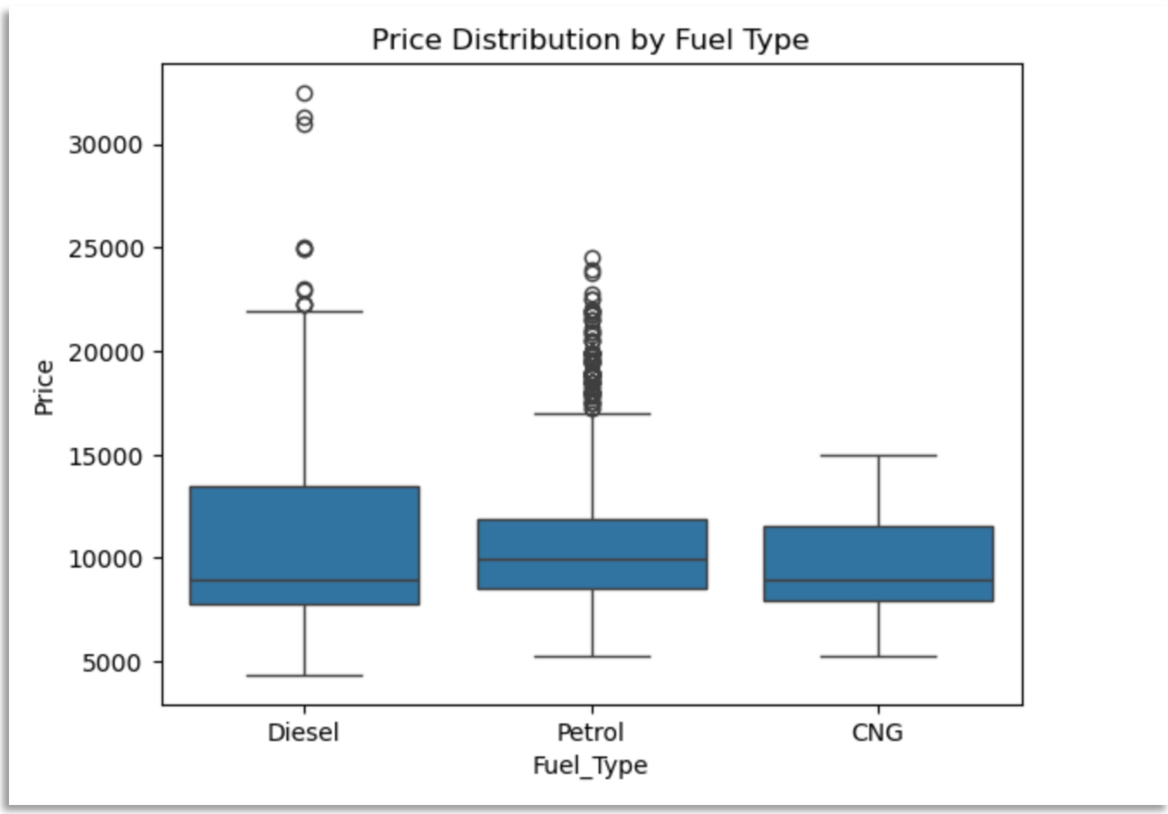

There was a greater presence of 1998 and 1999 Toyota Corolla models in the second-hand market, while newer models were less commonly available. This trend is clearly illustrated in the graph, with the most recent data available up to 2004.

**CAR AGE DISTRIBUTION:**



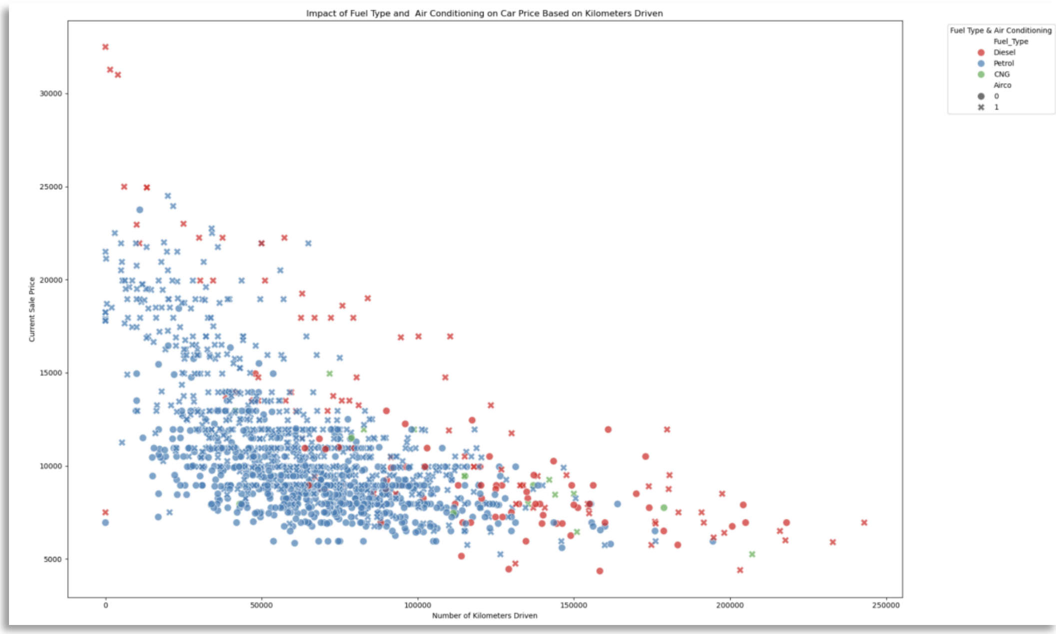

The Bar graph shows us that the number of cars is higher with age greater than 40 months. As the car age declines the number of cars gets less. Thats why the median price of car is less than mean price.

**PRICE DISTRIBUTION BY FUEL TYPE:**




The box plot shows that the price distribution of cars based on fuel type, showing that Diesel cars generally have higher prices and more outliers, with a median of around 10,000 and outliers up to 30,000. Petrol cars have a median price just below 10,000, with outliers reaching up to 25,000. CNG cars are the most consistently priced, with a median around 7,500 and no significant outliers.

## IMPACT OF FUEL TYPE & AC ON CAR PRICE BASED ON THE KM DRIVEN:




Diesel cars (red) tend to cluster at higher price points compared to Petrol (blue) and CNG (green), particularly when the car has driven fewer kilometers. However, this gap closes as kilometers driven increases

Petrol cars (blue) dominate the lower price range across the spectrum, and their prices drop more quickly with kilometers driven increases

CNG cars (green) appear sparsely and are mostly concentrated in the middle price range, with fewer data points overall

Cars with air conditioning (cross markers) tend to have higher sale prices compared to cars without air conditioning (circle markers). This is noticeable across different fuel types.

This impact is especially pronounced in the lower Kilometer driven cars, where the price difference between cars with and without air conditioning is more visible.

## 2. NORMALIZING THE VARIABLE KILOMETERS:

```
Original KM values:
0    46986
1    72937
2    41711
3    48000
4    38500
Name: KM, dtype: int64

Standard Normalized KM values:
0   -0.574695
1    0.117454
2   -0.715386
3   -0.547650
4   -0.801028
Name: KM, dtype: float64
```

## 3. CREATING DUMMIES FOR FUEL TYPES:

```
Original Fuel_Type values:
0    Diesel
1    Diesel
2    Diesel
3    Diesel
4    Diesel
Name: Fuel_Type, dtype: object

DataFrame with Fuel_Type dummies:
   Id                                          Model  Price  Age_08_04  \
0   1  TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors  13500         23
1   2  TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors  13750         23
2   3  TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors  13950         24
3   4  TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors  14950         26
4   5    TOYOTA Corolla 2.0 D4D HATCHB SOL 2/3-Doors  13750         30

   Mfg_Month  Mfg_Year        KM Fuel_Type  HP  Met_Color  ... Radio  \
0         10      2002 -0.574695    Diesel  90          1  ...     0
1         10      2002  0.117454    Diesel  90          1  ...     0
2          9      2002 -0.715386    Diesel  90          1  ...     0
3          7      2002 -0.547650    Diesel  90          0  ...     0
4          3      2002 -0.801028    Diesel  90          0  ...     0

   Mistlamps  Sport_Model  Backseat_Divider  Metallic_Rim  Radio_cassette  \
0          0            0                 1             0               0
1          0            0                 1             0               0
2          0            0                 1             0               0
3          0            0                 1             0               0
4          1            0                 1             0               0

   Parking_Assistant  Tow_Bar  Fuel_Type_Diesel  Fuel_Type_Petrol
0                  0        0              True             False
1                  0        0              True             False
2                  0        0              True             False
3                  0        0              True             False
4                  0        0              True             False

[5 rows x 41 columns]
```

4. **PARTITIONING THE DATASET:**

```
Training   :  (718, 39)
Validation :  (431, 39)
Test       :  (287, 39)

Training   :  (718, 39)
Validation :  (430, 39)
Test       :  (288, 39)
```

## TASK 2:

## SUMMARY OF THE CASE STUDY: MARKET STREET WINE CASE

### CHANGES IN CUSTOMER BASE:

- It varied from people who wanted an experience with wines from classic regions to people who are curious about local and organic wines.
- The age group used to be mid 40s and 50s which shifted to 30s and 40s.
- Types of customers:
  1. Knowledgeable and enthusiasts.
  2. Curios but not widely informed.
  3. Who just came to grab something.

### CHANGES IN REGULATIONS:

- In 1933, a "Three tier system" was introduced: Manufacturers, Wholesalers and retailer (all require license)
- If a retailer/restaurant want to do business with a winery(manufacturer) – they need to obtain license in whole of Virginia to open their warehouses and then send their representatives to the endpoint which is retailer/ restaurants.
- Retailers have to only buy from reps.

### CHANGE IN WINE TRENDS:

- In 2006 wine with high alcohol was preferred.
- A movement called "In pursuit of balance", pushed to prefer lower alcohol which meant now grapes had to be picked earlier for heartier wines
- Rose wines were less popular but now there is sharp increase in popularity especially those that from Provence, France.

- To better understand the growth: Rose wine grew 118% times in 5 years whereas still wine grew 1.5%.
- In 2021, it came to a stop. Now the focus is on few high-quality labels.
- Rose wine sales according to places: Cold – seasonality, warm- all through the year.
- One more shift was Bordeaux was shifting to local Burgundy, Champagne, California etc.

**ANOTHER SHIFT:**

- Shift towards sustainability in US led to an interest in natural wines.
- Natural wines are now ahead of organic and sparkling wines.
- Millennials were actually converting them into something more than just buzz words.
- They are coming into the store and persistently asking for those wines.
- But understanding what exactly is organic is proving difficult. Does it mean wines low on sulphate?

**COVID 19 TRENDS:**

- Covid 19 led to supply chain issues like importing French produced sparkling wine was now difficult. Many hoped that these issues short term, but they went on for several years.
- Hence US produced prosecco and sparkling wine, the local substitutes. Now they became less of a holiday thing and more of an everyday thing.
- There was a new energy to the domestic production therefore sparkling wine category grew 13% among US drinkers in 2022.
- As the supply chain issues continued it gave opportunity for wines from lesser-known regions like Washington, Oregon, New York, Finger Lakes-Food and drinks from Idaho, Colorado and Virginia.
- But still many Charlottesville wine consumers preferred old world options.

**CHANGES IN PURCHASING WINES:**

- Apart from the supply chain issues, method of purchasing has also changed now.
- They have started using ecommerce platforms, stores became warehouses and hence there is no issue of space.
- But it would be difficult to retain the customers they have gained during pandemic "online only" shopping.
- This digital engagement allowed to collect more data on customer behaviors
- Millenials were buying more age than any other age.

**BUSINESS PROBLEMS:**

The below are the business problems to be addressed:

1. While classic wines (French and Italian) have become difficult to obtain. What other alternatives consumers might be open to trying?

2. Were there wine producing regions more easily available in US? That consumers might be interested in?
3. Could we find good ones that hit the popular price point of under 20$?
4. Should they think about additional strategies to entice millennials?
5. What other wine trends would it be beneficial to consider?

## BREIF ANALYSIS OF THE DATASET

## THROUGH EXCEL USING POWER QUERY:

## INFERENCE 1:

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | title | points | badge | price_usd | organic* | vintage_year | alcohol_content | bottle_size | wine_style | grape_variety | designation | vineyard_state | vineyard_region_ |
| 2 | Frey NV Organic Red (California) | 86 | Best Buy | 9 | 1 | | 0.139 | 750 ml | Red | Red Blends | Organic | California | California |
| 3 | Dark Horse 2018 Sauvignon Blanc (California) | 90 | Best Buy | 8 | 0 | 2018 | 0.13 | 750 ml | White | Sauvignon Blanc | | California | California |
| 4 | Dark Horse 2016 Merlot (California) | 89 | Best Buy | 8 | 0 | 2016 | 0.135 | 750 ml | Red | Merlot | | California | California |
| 5 | Dark Horse 2019 Chardonnay (California) | 89 | Best Buy | 8 | 0 | 2019 | 0.145 | 750 ml | White | Chardonnay | | California | California |
| 6 | Dark Horse 2019 Sauvignon Blanc (California) | 89 | Best Buy | 8 | 0 | 2019 | 0.13 | 750 ml | White | Sauvignon Blanc | | California | California |
| 7 | Dark Horse 2019 Rosé (California) | 88 | Best Buy | 8 | 0 | 2019 | 0.13 | 750 ml | Rose | Rosé | | California | California |
| 8 | Dark Horse 2020 Pinot Grigio (California) | 88 | Best Buy | 8 | 0 | 2020 | 0.13 | 750 ml | White | Pinot Grigio | | California | California |
| 9 | Dark Horse NV Double Down Red (California) | 88 | Best Buy | 8 | 0 | | 0.145 | 750 ml | Red | Red Blends | Double Down | California | California |
| 10 | Dark Horse 2017 Pinot Noir (California) | 88 | Best Buy | 8 | 0 | 2017 | 0.14 | 750 ml | Red | Pinot Noir | | California | California |
| 11 | Sutter Home NV Pinot Grigio (California) | 88 | Best Buy | 8 | 0 | | 0.13 | 750 ml | White | Pinot Grigio | | California | California |
| 12 | Dark Horse 2019 Merlot (California) | 88 | Best Buy | 8 | 0 | 2019 | 0.145 | 750 ml | Red | Merlot | | California | California |
| 13 | Dark Horse NV Big Red (California) | 87 | Best Buy | 8 | 0 | | 0.135 | 750 ml | Red | Red Blends | Big | California | California |
| 14 | Dark Horse 2017 Cabernet Sauvignon (California) | 87 | Best Buy | 8 | 0 | 2017 | 0.135 | 750 ml | Red | Cabernet Sauvignon | | California | California |
| 15 | Dark Horse 2016 Chardonnay (California) | 87 | Best Buy | 8 | 0 | 2016 | 0.135 | 750 ml | White | Chardonnay | | California | California |
| 16 | Twisted 2015 Pinot Grigio (California) | 87 | Best Buy | 8 | 0 | 2015 | 0.13 | 750 ml | White | Pinot Grigio | | California | California |
| 17 | Dark Horse 2017 Sauvignon Blanc (California) | 87 | Best Buy | 8 | 0 | 2017 | 0.13 | 750 ml | White | Sauvignon Blanc | | California | California |

With the help of power query, we have sorted the "Best buy" badge data in excel. First, we sorted the price column in ascending order and then we sorted the points column in descending order. We have also moved organic column next to the price_usd column, after that we sorted the organic column. After the sorting we can understand the data better.

As there was a trend of organic wine we can see that the only best buy badged wine which is also organic is Frey NV which is of 9$ and has 86 points. All other wine has the best badge but are not organic, but they have higher points and still many of the wines come under 20$.

## INFERENCE 2:

| | title | points | badge | price_usd | natural* | vintage_year | alcohol_content | bottle_size | wine_style | grape_variety | designation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | Halcyon 2019 Camino Alto Vineyard Cabernet Franc (El Dorado) | 93 | Cellar Selection | 35 | 1 | 2019 | 0.137 | 750 ml | Red | Cabernet Franc | Camino Alto Vineyard |
| 3 | Cristom 2013 Louise Vineyard Pinot Noir (Eola-Amity Hills) | 94 | Cellar Selection | 58 | 1 | 2013 | 0.135 | 750 ml | Red | Pinot Noir | Louise Vineyard |
| 4 | Roco 2016 Wits' End Vineyard Chardonnay (Chehalem Mountains) | 92 | Cellar Selection | 60 | 1 | 2016 | 0.135 | 750 ml | White | Chardonnay | Wits' End Vineyard |
| 5 | Patricia Green Cellars 2016 Freedom Hill Vineyard Coury Clone Pinot Noir (Willamette Valley) | 95 | Cellar Selection | 75 | 1 | 2016 | 0.14 | 750 ml | Red | Pinot Noir | Freedom Hill Vineyard Coury Clo |
| 6 | The Eyrie Vineyards 2013 Original Vines Estate Pinot Noir (Dundee Hills) | 94 | Cellar Selection | 80 | 1 | 2013 | 0.13 | 750 ml | Red | Pinot Noir | Original Vines Estate |
| 7 | Kendall-Jackson 2013 Vintner's Reserve Summation Red (California) | 91 | Cellar Selection | 17 | 0 | 2013 | 0.145 | 750 ml | Red | Red Blends | Vintner's Reserve Summation |
| 8 | Styring 2014 Whimsy Estate Riesling (Ribbon Ridge) | 95 | Cellar Selection | 20 | 0 | 2014 | 0.126 | 750 ml | White | Riesling | Whimsy Estate |
| 9 | McNab Ridge 2017 Petite Sirah (Mendocino) | 92 | Cellar Selection | 20 | 0 | 2017 | 0.148 | 750 ml | Red | Petite Sirah | |
| 10 | Long Shadows 2019 Poet's Leap Riesling (Columbia Valley (WA)) | 92 | Cellar Selection | 20 | 0 | 2019 | 0.123 | 750 ml | White | Riesling | Poet's Leap |
| 11 | Soos Creek 2017 Soleil Red (Columbia Valley (WA)) | 91 | Cellar Selection | 20 | 0 | 2017 | 0.141 | 750 ml | Red | Bordeaux-style Red Blend | Soleil |
| 12 | St. Amant 2018 Mohr-Fry Ranch Petite Sirah (Lodi) | 94 | Cellar Selection | 21 | 0 | 2018 | 0.143 | 750 ml | Red | Petite Sirah | Mohr-Fry Ranch |
| 13 | Barra of Mendocino 2016 Estate Grown Petite Sirah (Mendocino) | 93 | Cellar Selection | 22 | 0 | 2016 | 0.145 | 750 ml | Red | Petite Sirah | Estate Grown |
| 14 | Sokol Blosser 2015 Pinot Gris (Dundee Hills) | 91 | Cellar Selection | 22 | 0 | 2015 | 0.13 | 750 ml | White | Pinot Gris | |
| 15 | Barra of Mendocino 2011 Estate Grown Petite Sirah (Mendocino) | 89 | Cellar Selection | 22 | 0 | 2011 | 0.145 | 750 ml | Red | Petite Sirah | Estate Grown |
| 16 | Keuka Spring 2014 Humphreys Vineyard Riesling (Finger Lakes) | 88 | Cellar Selection | 22 | 0 | 2014 | 0.12 | 750 ml | White | Riesling | Humphreys Vineyard |
| 17 | Tolosa 2019 No Oak Heritage Chardonnay (Central Coast) | 91 | Cellar Selection | 23 | 0 | 2019 | 0.134 | 750 ml | White | Chardonnay | No Oak Heritage |

In this image, we have filtered the Cellar Selection badge data. as we are focusing on the natural wine we sorted the natural column in descending order, then sorted the price column in ascending order, after that, we sorted the points column in descending order so that we can get a clear view on our data.

As we can see that there are 5 natural wine with the price range of $35 - $80 and with points range of 92-95. If the wine is not natural then the price range is $17 - $525 and the point ranges from 88 - 100.

**INFERENCE 3:**

| | title | points | badge | price_usd | natural* | vintage_year | alcohol_content | bottle_size | wine_style | grape_variety | designation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | J. Scott Cellars 2017 Sauvignon Blanc (Oregon) | 91 | Editors' Choice | 18 | 1 | 2017 | 0.138 | 750 ml | White | Sauvignon Blanc | |
| 7 | 2Hawk 2020 Grenache Rosé (Rogue Valley) | 91 | Editors' Choice | 18 | 1 | 2020 | 0.127 | 750 ml | Rose | Rosé | Grenache |
| 8 | Barra of Mendocino 2016 Estate Grown Cabernet Sauvignon (Mendocino) | 91 | Editors' Choice | 20 | 1 | 2016 | 0.145 | 750 ml | Red | Cabernet Sauvignon | Estate Grow |
| 9 | Owen Roe 2020 Crawford-Beck Vineyard Pinot Gris (Eola-Amity Hills) | 92 | Editors' Choice | 22 | 1 | 2020 | 0.13 | 750 ml | White | Pinot Gris | Crawford-B |
| 10 | Charlie & Echo 2019 Witherfall Tempest Red (San Diego County) | 91 | Editors' Choice | 25 | 1 | 2019 | 0.118 | 750 ml | Red | Red Blends | Witherfall T |
| 11 | J. Scott Cellars 2016 Syrah (Rogue Valley) | 91 | Editors' Choice | 28 | 1 | 2016 | 0.149 | 750 ml | Red | Syrah | |
| 12 | Amador Cellars 2015 Estate Reserve Grenache Noir (Amador County) | 93 | Editors' Choice | 30 | 1 | 2015 | 0.15 | 750 ml | Red | Grenache Noir | Estate Rese |
| 13 | Fields Family 2013 Lodi Native Stampede Vineyard Zinfandel (Clements Hills) | 91 | Editors' Choice | 30 | 1 | 2013 | 0.139 | 750 ml | Red | Zinfandel | Lodi Native |
| 14 | Troon 2014 Estate Tannat (Applegate Valley) | 90 | Editors' Choice | 35 | 1 | 2014 | 0.144 | 750 ml | Red | Tannat | Estate |
| 15 | Thacher 2020 Hastings Ranch Viognier (Paso Robles) | 92 | Editors' Choice | 36 | 1 | 2020 | 0.129 | 750 ml | White | Viognier | Hastings Ra |
| 16 | Davies 2015 Pinot Noir (Anderson Valley) | 94 | Editors' Choice | 40 | 1 | 2015 | 0.143 | 750 ml | Red | Pinot Noir | |
| 17 | Birichino 2019 Mr. Natural No. 3 Enz Vineyard Old Vine Mourvèdre (Lime Kiln Valley) | 93 | Editors' Choice | 40 | 1 | 2019 | 0.125 | 750 ml | Red | Mourvèdre | Mr. Natural |
| 18 | Ricochet 2019 Pinot Noir (Willamette Valley) | 91 | Editors' Choice | 40 | 1 | 2019 | 0.13 | 750 ml | Red | Pinot Noir | |
| 19 | Day 2019 Belle Pente Vineyard Chardonnay (Yamhill-Carlton) | 94 | Editors' Choice | 48 | 1 | 2019 | 0.123 | 750 ml | White | Chardonnay | Belle Pente |
| 20 | Brooks 2015 Crannell Pinot Noir (Eola-Amity Hills) | 93 | Editors' Choice | 48 | 1 | 2015 | 0.135 | 750 ml | Red | Pinot Noir | Crannell |
| 21 | Winderlea 2013 Crawford Beck Vineyard Pinot Noir (Eola-Amity Hills) | 91 | Editors' Choice | 48 | 1 | 2013 | 0.128 | 750 ml | Red | Pinot Noir | Crawford Be |
| 22 | King Estate 2013 Antiquum Vineyards Pinot Noir (Willamette Valley) | 92 | Editors' Choice | 55 | 1 | 2013 | 0.135 | 750 ml | Red | Pinot Noir | Antiquum V |
| 23 | Dutton-Goldfield 2014 Fox Den Vineyard Pinot Noir (Green Valley) | 94 | Editors' Choice | 62 | 1 | 2014 | 0.135 | 750 ml | Red | Pinot Noir | Fox Den Vir |
| 24 | Brick & Mortar 2017 Rosé (California) | 91 | Editors' Choice | 8 | 0 | 2017 | 0.115 | 375 ml | Rose | Rosé | |
| 25 | Brick & Mortar 2018 Rosé (California) | 91 | Editors' Choice | 8 | 0 | 2018 | 0.115 | 375 ml | Rose | Rosé | |
| 26 | Maker 2020 Ser Winery Rosé of Grenache (Monterey) | 90 | Editors' Choice | 8 | 0 | 2020 | 0.13 | 250 ml | Rose | Grenache | Ser Winery |
| 27 | Al 2016 Sauvignon Blanc (Monterey County) | 88 | Editors' Choice | 8 | 0 | 2016 | 0.129 | 375 ml | White | Sauvignon Blanc | |
| 28 | Maker 2019 Nicole Walsh Cabernet Pfeffer (Cienega Valley) | 91 | Editors' Choice | 12 | 0 | 2019 | 0.13 | 250 ml | Red | Cabernet Pfeffer | Nicole Wals |
| 29 | Coopers Hall 2018 Rosé of Pinot Noir (Willamette Valley) | 90 | Editors' Choice | 12 | 0 | 2018 | 0.139 | 500 ml | Rose | Pinot Noir | Rosé of |
| 30 | Castle Rock 2014 Pinot Noir (Mendocino County) | 87 | Editors' Choice | 13 | 0 | 2014 | 0.135 | 750 ml | Red | Pinot Noir | |
| 31 | Vinum 2015 Chenin Blanc (Clarksburg) | 91 | Editors' Choice | 14 | 0 | 2015 | 0.135 | 750 ml | White | Chenin Blanc | |
| 32 | J. Lohr 2018 Riverstone Chardonnay (Arroyo Seco) | 88 | Editors' Choice | 14 | 0 | 2018 | 0.14 | 750 ml | White | Chardonnay | Riverstone |
| 33 | Bonny Doon 2020 Le Cigare Volant Red (Central Coast) | 92 | Editors' Choice | 15 | 0 | 2020 | 0.136 | 750 ml | Red | Rhône-style Red Blend | Le Cigare Vo |

The last Badge is the Editor's choice in this data, we have more natural options with the prices ranging from $18 - $62 and points ranging from 91 - 94. if we talk about non-natural options the prices are higher ranging from $8 - $475 with points ranging from 87 - 100.

**THROUGH PYTHON:**

**UNDERSTANDING ROWS AND COLUMNS:**

```
(41731, 65)
Index(['title', 'points', 'badge', 'price_usd', 'vintage_year',
       'alcohol_content', 'bottle_size', 'wine_style', 'grape_variety',
       'designation', 'vineyard_state', 'vineyard_region_1',
       'vineyard_region_2', 'winery', 'winery_longitude', 'winery_latitude',
       'winery_state', 'review_url', 'review_year', 'taster_name', 'natural*',
       'organic*', 'cherr*', 'sweet*', 'tanni*', 'earth*', 'fresh*', 'melon',
       'yeast', 'tart', 'bright*', 'jam', 'plum*', 'perfume*', 'fruit*',
       'toast*', 'dry*', 'crisp*', 'acid*', 'vanilla', 'apple', 'lime', 'oak*',
       'spic*', 'strawberr*', 'tropical*', 'smok*', 'meat*', 'berr*',
       'mineral', 'apricot', 'tobacco', 'leather*', 'forest', 'pepper*',
       'herb*', 'floral', 'rich*', 'citrus*', 'nut*', 'grass*', 'lemon*',
       'cream*', 'wood*', 'honey'],
      dtype='object')
```

The dataset has 41,731 rows and 65 columns. The columns represent various attributes of wine, including details like title, points, price, vintage year, and winery location, along with tasting notes (e.g., fruit, tannin, oak) and descriptors. The data spans a wide range of characteristics used for wine reviews and analysis.

**CHECKING FOR NULL VALUES:**

```
title                0
points               0
badge            35138
price_usd          334
vintage_year       470
                 ...
grass*               0
lemon*               0
cream*               0
wood*                0
honey                0
Length: 65, dtype: int64
```

The output shows the count of missing values for each column in the wine_df DataFrame. Notably, the 'badge' column has 35,138 missing entries, while 'price_usd' and 'vintage_year' have 334 and 470 missing values, respectively, indicating areas where data may need to be handled before analysis.

**CHECKING FOR DUPLICATE ROWS:**

```
print(wine_df.duplicated().sum())
```

```
0
```

The output indicates that there are no duplicate entries in the wine_df DataFrame