

| 1  | Sale ID | Date of Sale | Buy Price    | Sale Price   | Profit      | Time of Sale | Dealership Name |
|----|---------|--------------|--------------|--------------|-------------|--------------|-----------------|
| 2  | 10001   | 1/1/21       | \$ 21,274.00 | \$ 27,316.00 | \$ 6,042.00 | 10:48 AM     | East Adams      |
| 3  | 10002   | 1/1/21       | \$ 20,226.00 | \$ 24,441.00 | \$ 4,215.00 | 4:47 PM      | East Adams      |
| 4  | 10003   | 1/1/21       | \$ 20,178.00 | \$ 21,712.00 | \$ 1,534.00 | 10:37 AM     | Westfield       |
| 5  | 10004   | 1/1/21       | \$ 18,784.00 | \$ 23,513.00 | \$ 4,729.00 | 2:27 PM      | Pineville       |
| 6  | 10005   | 1/1/21       | \$ 17,190.00 | \$ 20,645.00 | \$ 3,455.00 | 11:59 AM     | Westfield       |
| 7  | 10006   | 1/2/21       | \$ 21,711.00 | \$ 24,815.00 | \$ 3,104.00 | 11:31 AM     | East Adams      |
| 8  | 10007   | 1/2/21       | \$ 20,598.00 | \$ 25,002.00 | \$ 4,404.00 | 10:51 AM     | Pineville       |
| 9  | 10008   | 1/2/21       | \$ 19,282.00 | \$ 24,309.00 | \$ 5,027.00 | 8:44 AM      | Westfield       |
| 10 | 10009   | 1/3/21       | \$ 18,703.00 | \$ 22,994.00 | \$ 4,291.00 | 12:47 PM     | East Adams      |
| 11 | 10010   | 1/3/21       | \$ 20,937.00 | \$ 22,794.00 | \$ 1,857.00 | 5:49 PM      | East Adams      |
| 12 | 10011   | 1/3/21       | \$ 19,103.00 | \$ 24,070.00 | \$ 4,967.00 | 4:36 PM      | Culver          |
| 13 | 10012   | 1/3/21       | \$ 18,058.00 | \$ 24,382.00 | \$ 6,324.00 | 2:24 PM      | Westfield       |
| 14 | 10013   | 1/4/21       | \$ 17,856.00 | \$ 21,543.00 | \$ 3,687.00 | 2:18 PM      | Pineville       |
| 15 | 10014   | 1/4/21       | \$ 16,116.00 | \$ 18,946.00 | \$ 2,830.00 | 3:08 PM      | East Adams      |
| 16 | 10015   | 1/4/21       | \$ 16,145.00 | \$ 20,564.00 | \$ 4,419.00 | 5:06 PM      | Westfield       |
| 17 | 10016   | 1/5/21       | \$ 17,464.00 | \$ 23,711.00 | \$ 6,247.00 | 4:30 PM      | Oak Park        |
| 18 | 10017   | 1/6/21       | \$ 24,057.00 | \$ 29,580.00 | \$ 5,523.00 | 6:14 PM      | Culver          |
| 19 | 10018   | 1/6/21       | \$ 10,029.00 | \$ 15,496.00 | \$ 5,467.00 | 4:45 PM      | Oak Park        |
| 20 | 10019   | 1/6/21       | \$ 12,810.00 | \$ 17,285.00 | \$ 4,475.00 | 6:45 PM      | Oak Park        |
| 21 | 10020   | 1/6/21       | \$ 20,695.00 | \$ 26,402.00 | \$ 5,707.00 | 1:10 PM      | East Adams      |
| 22 | 10021   | 1/7/21       | \$ 17,623.00 | \$ 24,085.00 | \$ 6,462.00 | 4:10 PM      | Oak Park        |
| 23 | 10022   | 1/7/21       | \$ 20,111.00 | \$ 23,366.00 | \$ 3,255.00 | 11:58 AM     | Pineville       |
| 24 | 10023   | 1/7/21       | \$ 16,481.00 | \$ 20,079.00 | \$ 3,598.00 | 6:46 PM      | Westfield       |
| 25 | 10024   | 1/8/21       | \$ 12,337.00 | \$ 16,266.00 | \$ 3,929.00 | 9:18 AM      | Culver          |

# Discovering Meaning



**Data Literacy Skills**

Discovering meaning is a data literacy skill that is all about identifying patterns in data and then making sense of these patterns.

In this module, you learn about some tools that you can use to conduct an exploratory analysis of data. These tools help you get a sense of some of the patterns in your data set and what they might mean. This stage is critical before moving on to more focused analysis.

# Exploratory Data Analysis

Once your data is clean and the questions that you would like to explore are identified, you are ready to begin to discover meaning in your data. This process begins with **exploratory data analysis**. Exploratory data analysis is the "getting-to-know-you" phase of working with data, and focuses on understanding the distribution of data.



# Understanding Distribution of Data

Exploring the distribution of data, both numerically and graphically, can begin to show the big picture of what your data looks like.

You can think of data distribution as organization of data. It helps you take a messy collection of data and begin to arrange it. Data is organized numerically and graphically - often from smallest to largest. Data visualizations, or graphics, help you see each data value, as well as the data set as a whole.

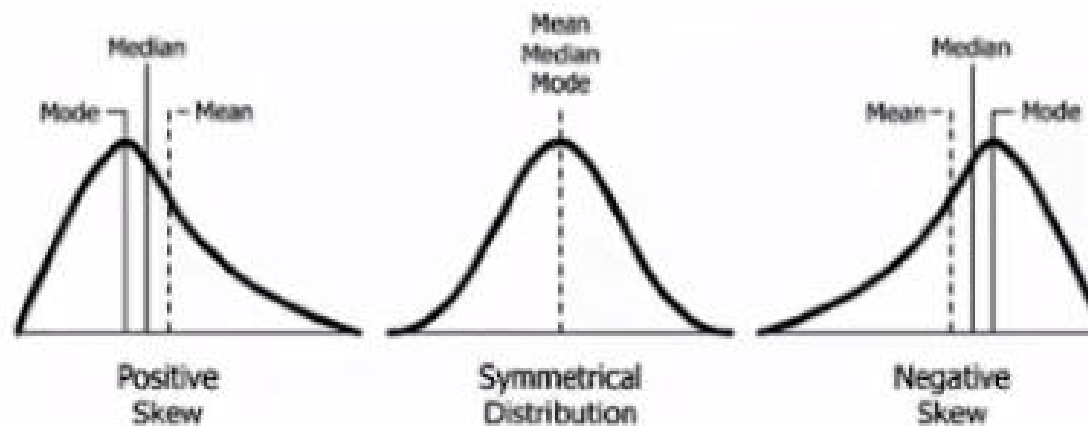
In this module, you learn foundational statistical concepts that help you explore the distribution of your data, including:

- histograms
- mean and median
- and box-and-whisker plots, to name a few.

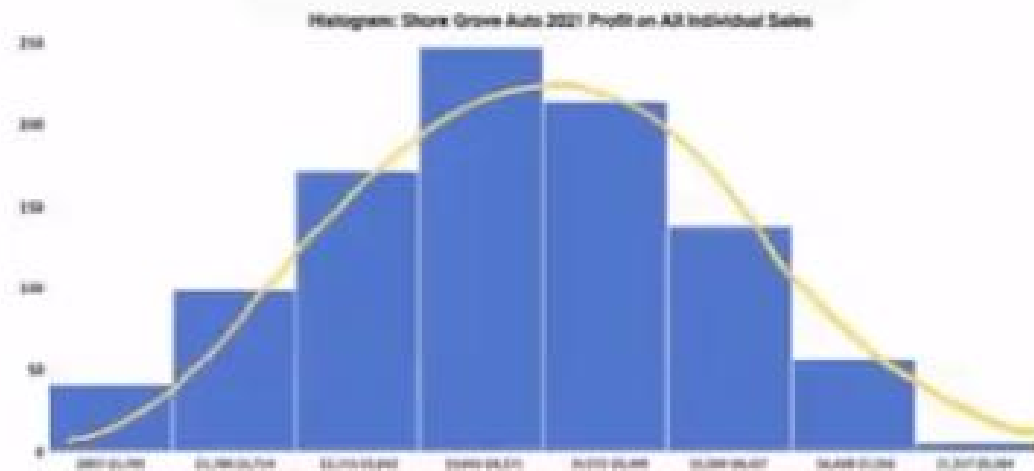
# Exploring Shape

---

What is the **shape** of a distribution?



## What does Kim learn by exploring the shape of the distribution of profit data?

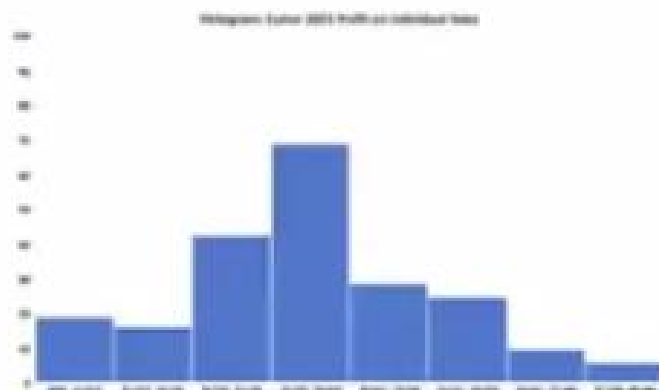


### Key Takeaways

- Data is normally distributed
- Most profit falls in the middle between \$3600 and \$5500
- Kim would like to see this analysis for each dealership

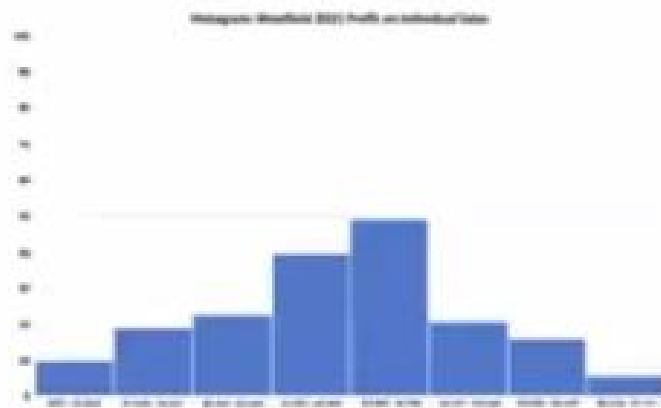
## What does Kim learn by exploring the shape of the distribution of profit data by dealership?

The Culver, Westfield, and East Adams locations look to have relatively normal distribution shape, with much of the data in the center tapering off on either side.



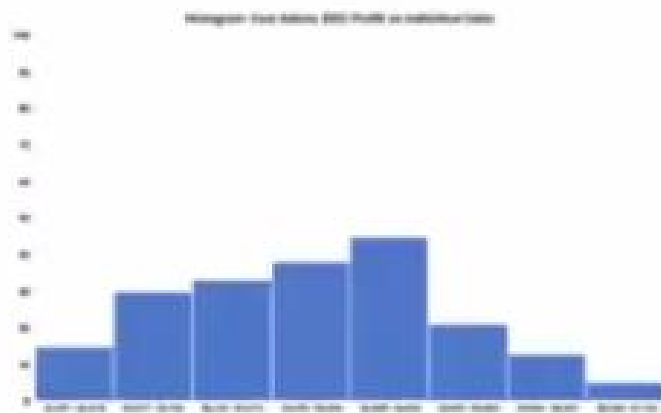
**Culver**

relatively normal distribution



## Westfield

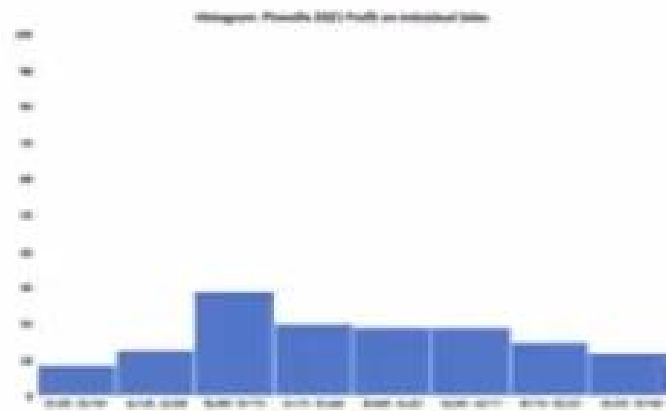
relatively normal distribution



## East Adams

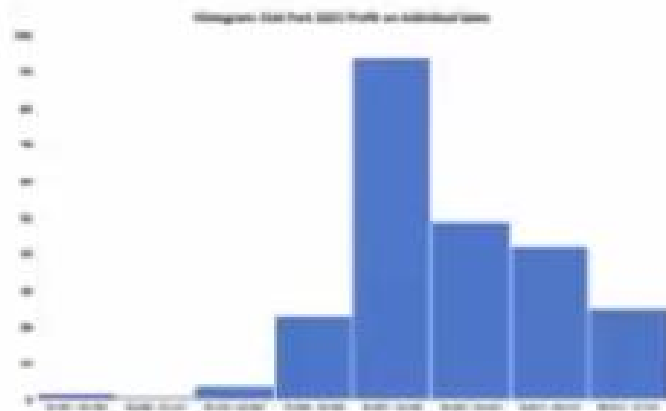
relatively normal distribution





## Pineville

more uniform distribution



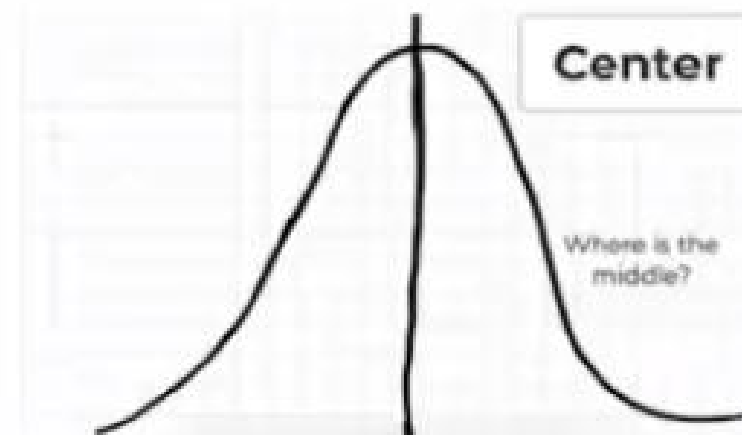
## Oak Park

skewed left distribution

# Exploring Center

---

What does the **center** of the data tell us?



## Mean

The **mean** is the average of the data.

To calculate the mean, you will add all the values and then divide by the number of data values. In this case, you would add all of the exam scores together and divide by the number of data values (5).

The mean exam score, or average exam score, among these students is 88.4.

$$\frac{\text{Sum of Values}}{\# \text{ of values}} = \text{mean}$$

$$\frac{99 + 85 + 85 + 93 + 80}{5} = 88.4$$

$$\boxed{\text{Mean} = 88.4}$$

## Median

The **median** is the exact middle, or midpoint, of the data.

To find the median, you must first organize the data from the smallest value to the largest value. The median is the center value, or the average of the center values if there are two.

**In this case, the median is 85.**

80, 85, 85, 93, 99

80, 85, (85), 93, 99

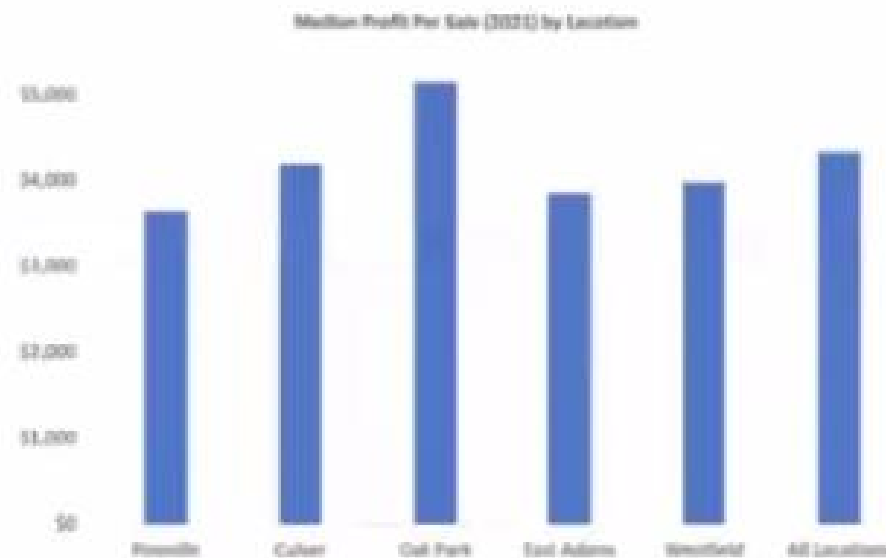
median = 85

---

## When to use Mean versus Median

As a general rule of thumb, use mean **only** for normal (bell-shaped) distributions with no outliers.

## What does Kim learn by exploring the center of the distribution of profit data?



### Key Takeaways

- Oak Park has highest median profit
- Pineville has the lowest median profit
- Kim has more questions related to cumulative profit and number of sales at each dealership

What does Kim learn by exploring the center of the distribution of profit data?



Overall median = \$4,334

1

Range

74

min  
#

70

max  
#

2

Quartiles

74

min  
#

25%

Q1

25%

Q2  
median

25%

Q3

25%

95

max  
#  
Q4



## Interquartile Range

$Q3 - Q1 = \text{Interquartile range (IQR)}$

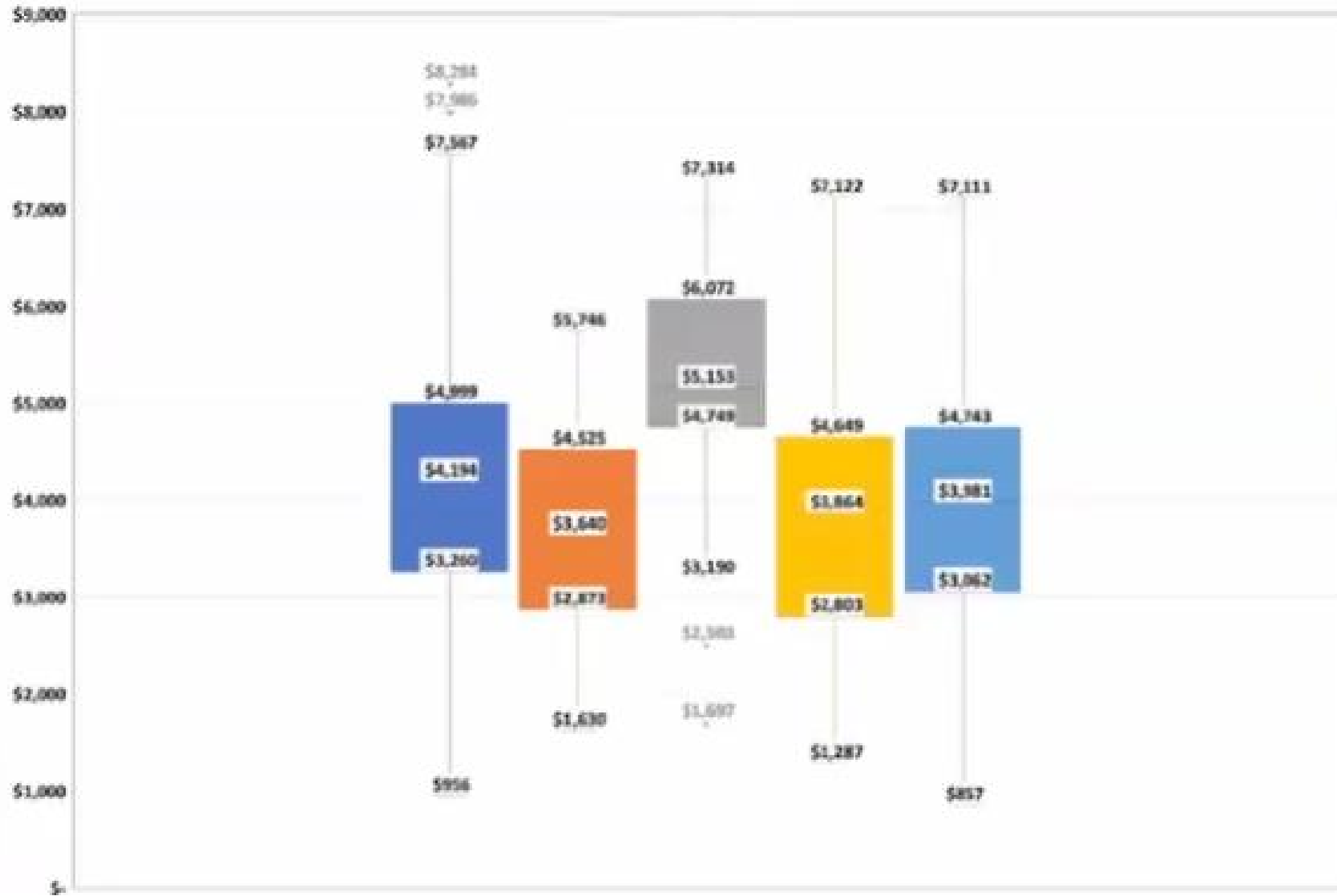


# Exploring Spread of Shore Grove Auto Profit Data Using Box-and-Whisker Plots

Because box-and-whisker plots show the range, display the five-number summary values, and give way to calculating the interquartile range, Kim begins her analysis with box-and-whisker plots.

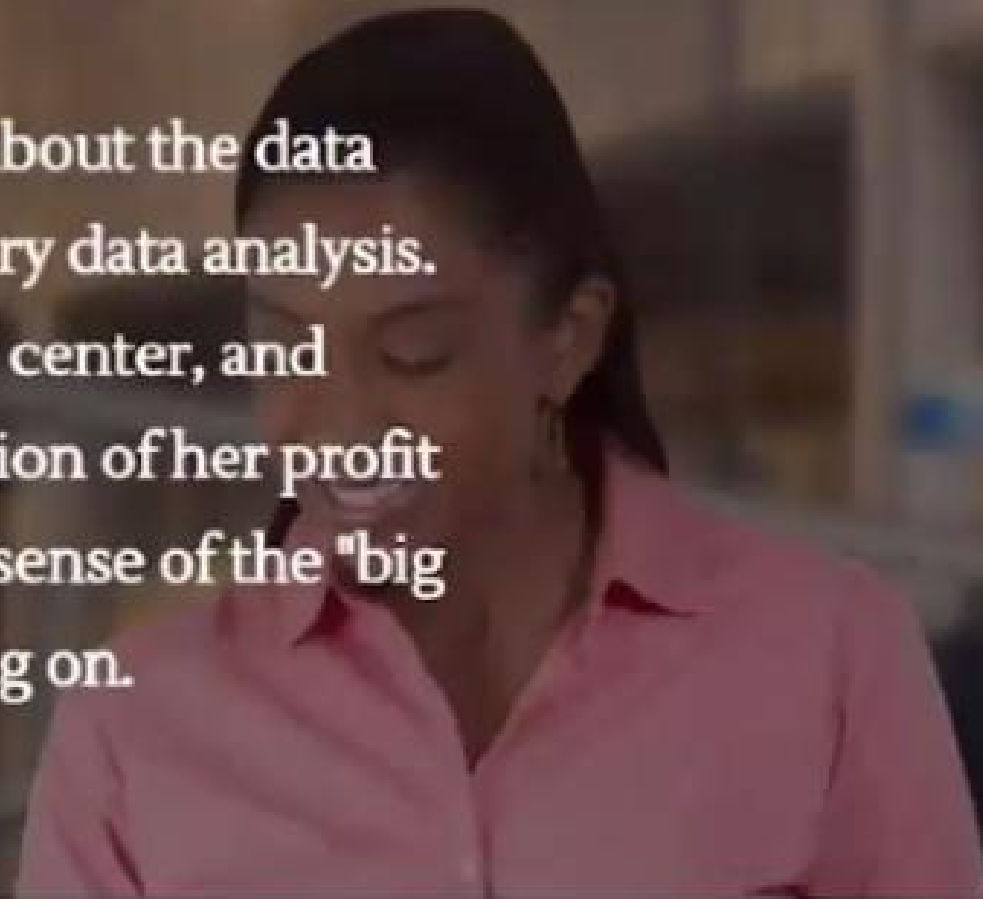
Box and Whisker Plot: Profit by Dealership

■ Culver ■ Pineville ■ Oak Park ■ East Adams ■ Westfield



---

Kim has learned a lot about the data through her exploratory data analysis. Learning about shape, center, and spread of the distribution of her profit data gave her a better sense of the "big picture" of what's going on.



## Key Takeaways from Kim's Exploratory Data Analysis

1

Although the consolidated data is normally distributed, the data at individual dealerships is not. This tells Kim that she should use median rather than mean in future analysis.

2


There are a few outliers in the data. None of them are extreme, so Kim decides that they do not need to be removed from the dataset.

3

She is curious about what is going on at Oak Park, with the highest median profit per sale, and Pineville, with the lowest median profit per sale.

4

Kim realizes she's learned more about the profit at each dealership, but she's curious about how many sales each dealership made in 2021 and how the frequency of sales impacts profit.

A woman with dark hair, wearing a pink button-down shirt, is smiling and standing in front of a large window. The window reflects a car dealership with several cars parked outside. The scene is dimly lit, suggesting dusk or dawn.

Finally, Kim is ready to make data-informed decisions!

Now that Kim has synthesized her findings and visited each dealership to understand contextual factors, she feels ready to make some decisions to increase profit in 2022.

These are *data-informed* decisions!

---

**Thank You**