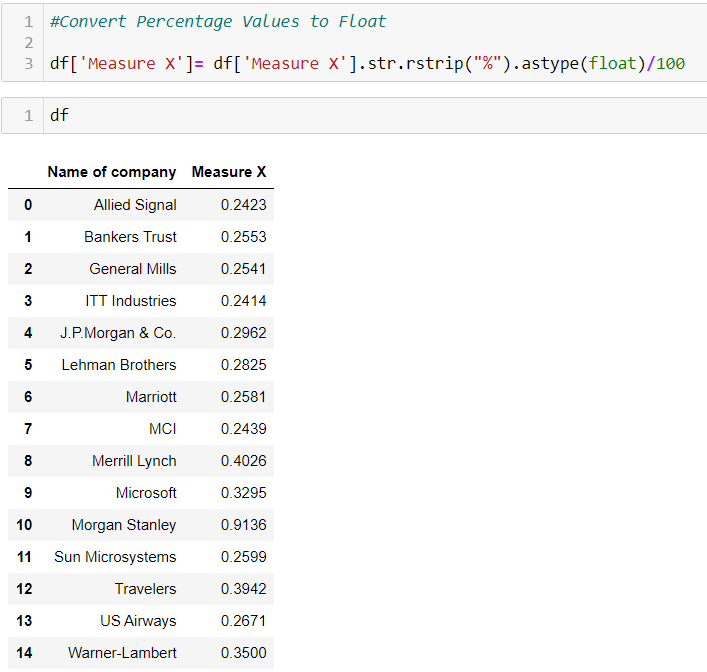
**Topics: Descriptive Statistics and Probability**

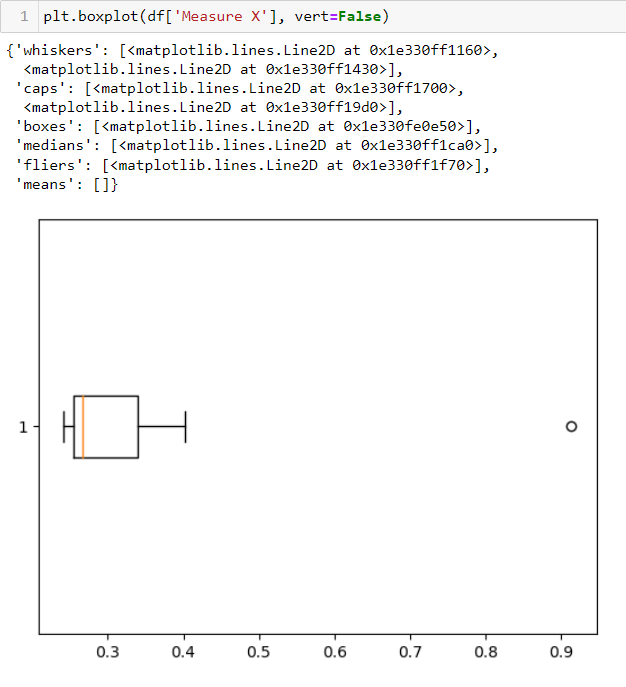
1. Look at the data given below. Plot the data, find the outliers and find out

|  |  |
| --- | --- |
| **Name of company** | **Measure X** |
| Allied Signal | 24.23% |
| Bankers Trust | 25.53% |
| General Mills | 25.41% |
| ITT Industries | 24.14% |
| J.P.Morgan & Co. | 29.62% |
| Lehman Brothers | 28.25% |
| Marriott | 25.81% |
| MCI | 24.39% |
| Merrill Lynch | 40.26% |
| Microsoft | 32.95% |
| Morgan Stanley | 91.36% |
| Sun Microsystems | 25.99% |
| Travelers | 39.42% |
| US Airways | 26.71% |
| Warner-Lambert | 35.00% |

* Answer:



* Mean = **µ** = df['Measure X'].mean() = **0.33271**
* Standard Deviation = = df['Measure X'].std() = **0.16945**
* Variance = = df['Measure X'].var() = **0.02871**
* **BOXPLOT for the given data:**

****

From the Boxplot, we can conclude that there is a **single outlier** and looking at the data, the outlier is Morgan Stanley and its value is 91.36%



Answer the following three questions based on the box-plot above.

1. What is inter-quartile range of this dataset? (please approximate the numbers) In one line, explain what this value implies.

* **Answer**:

IQR for the given data is:

Q3 = 12, Q1 = 5.

IQR= Q3 –Q1 = 12-5 = 7

IQR= 7.

IQR represents 50% of the values in data.

1. What can we say about the skewness of this dataset?

* **Answer**:

The data has an outlier on the right side of the boxplot meaning the data is Right Skewed. We observe Positive Skewness.

1. If it was found that the data point with the value 25 is actually 2.5, how would the new box-plot be affected?

* **Answer:**

The new boxplot will have no outliers.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

* **Answer:** Since, the frequency count of the data in bins 4-6 and 6-8 is repeated twice, the mode will possibly lie between bins 4-6 and 6-8.

1. Comment on the skewness of the dataset.

* **Answer:** There is a long tail towards right. The data is Right Skewed meaning Positively Skewed.

1. Suppose that the above histogram and the box-plot in question 2 are plotted for the same dataset. Explain how these graphs complement each other in providing information about any dataset.

* **Answer:**  Both, Histogram and Boxplot provide idea about the skewness in the data. Histogram makes use of bins to show the frequency counts of the data points while boxplot gives us the idea of the Q1, Q3, IQR, median, minimum value, maximum value and outliers in the data.

1. AT&T was running commercials in 1990 aimed at luring back customers who had switched to one of the other long-distance phone service providers. One such commercial shows a businessman trying to reach Phoenix and mistakenly getting Fiji, where a half-naked native on a beach responds incomprehensibly in Polynesian. When asked about this advertisement, AT&T admitted that the portrayed incident did not actually take place but added that this was an enactment of something that “could happen.” Suppose that one in 200 long-distance telephone calls is misdirected. What is the probability that at least one in five attempted telephone calls reaches the wrong number? (Assume independence of attempts.)

* **Answer:**

Let X be the probability that 1 call is being misdirected out of 200 calls.

P(X) = 1/200

P(X) = 0.005

Probability of having at least one successful call will be = 1-P(X)

1-P(X) = 1-(1/200)

= 199/200

= 0.995

Now, we have been told that every event is independent of other event, so the probability will be:

P= 1 - [(0.995) ^ 5]

P= 1 – 0.97524

P= 0.02476

Therefore, the probability that at least one in five attempted telephone calls reaches the wrong number is 0.02476 that is 2.47%.

* Returns on a certain business venture, to the nearest $1,000, are known to follow the following probability distribution

|  |  |
| --- | --- |
| x | P(x) |
| -2,000 | 0.1 |
| -1,000 | 0.1 |
| 0 | 0.2 |
| 1000 | 0.2 |
| 2000 | 0.3 |
| 3000 | 0.1 |

1. What is the most likely monetary outcome of the business venture?

**Answer:**

$2000 is the most likely monetary outcome because it has high probability of occurrence i.e, 0.3.

1. Is the venture likely to be successful? Explain.

**Answer:**

To be successful is nothing but getting positive returns.

From the above data, we see that the chances of getting positive returns is:

0.2 + 0.3 + 0.1 = 0.6

Yes, there is a **probability** of **60%** that the **venture** **will be** **successful**.

1. What is the long-term average earning of business ventures of this kind? Explain.

**Answer:**

To calculate the long-term average earning of business ventures:

µ = E(X) = X \* P(X)

Here, n = 6.

µ = (-2000\*0.1) + (-1000\*0.1)+ (0\*0.2)+(1000\*0.2)+ (2000\*0.3)+ (3000\*0.1)

µ = (-200-100+0+200+600+300)

µ = **$** **800** is the **long-term average earning** of business ventures of this kind.

1. What is the good measure of the risk involved in a venture of this kind? Compute this measure.

**Answer:**

**Variance (X) = = E () –**

E () = [(-2000^2)\*0.1] + [(-1000^2)\*0.1] + [(0^2)\*0.2] + [(1000^2)\*0.2] + [(2000^2)\*0.3] + [(3000^2)\*0.1]

E () = (400000) + (100000) + (0) + (200000) + (1200000) + (900000)

**E () = 2800000.**

= 800^2 = **640000**

Variance = E () –

Variance= 2800000 – 640000

**Variance= 2160000.**

Now, Standard deviation = = =

= 1469.6938

**Standard Deviation = $1470**

The Variability is quite high. Also, the standard deviation value is $1470 for a mean of $800, so the risk is high.