**Topics: Descriptive Statistics and Probability**

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

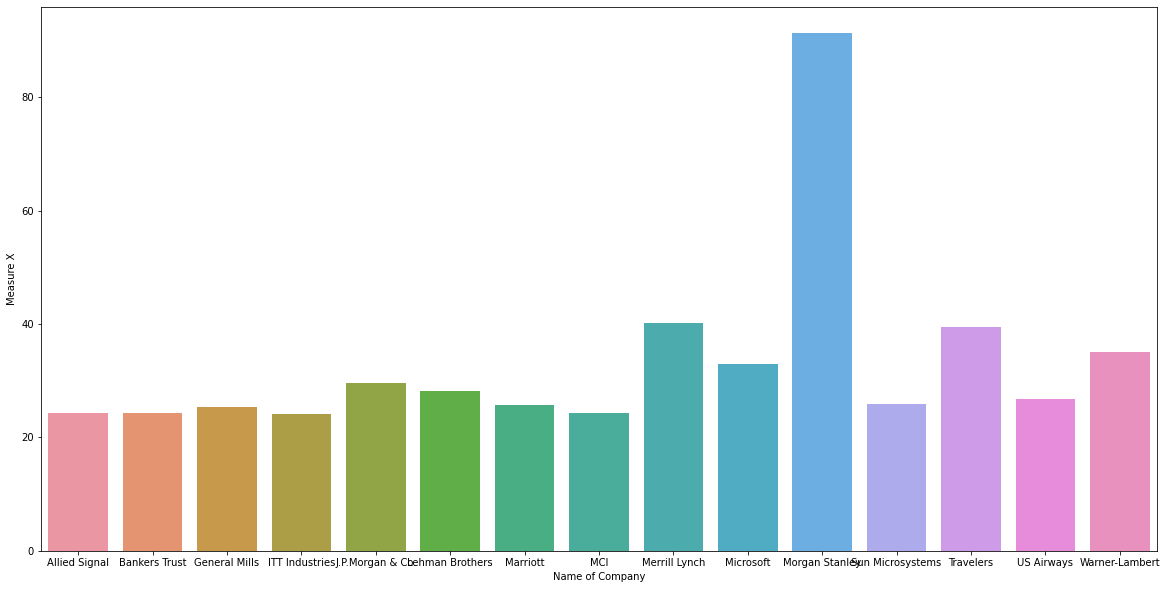
|  |  |
| --- | --- |
| **Name of company** | **Measure X** |
| Allied Signal | 24.23% |
| Bankers Trust | 25.53% |
| General Mills | 25.41% |
| ITT Industries | 24.14% |
| J.P.Morgan & Co. | 29.62% |
| Lehman Brothers | 28.25% |
| Marriott | 25.81% |
| MCI | 24.39% |
| Merrill Lynch | 40.26% |
| Microsoft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

**Ans:**

Step 1: create a dataframe by using pandas of the above data.

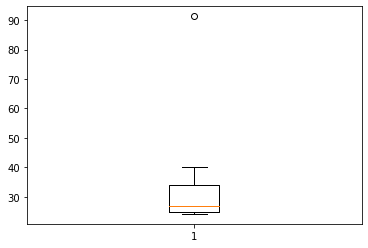
Step 2 : plot the numerical (float 64) data by using boxplot to find the outliers.

Code : sns.barplot(data = data, x = "Name of Company", y = "Measure X")



plt.boxplot(data["Measure X"])

plt.title("Finding Outliers")



**Conclusion :** Here, the Morgan Stanley company contains outlier of value 91.36%

Step 3 : Import numpy and calculate

Mean: data.mean() = 33.184667

Standard Deviation: data.std() = 16.991085

Variance: data.var() = 288.696955



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.

**Ans : Formula for IQR = Q3 – Q1, 12-5 = 7. IQR represents 50% of the data present inside this range.**

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

**Ans : The dataset is positively skewed.**

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?

**Ans: If it was found that the data point with the value 25 is actually 2.5, then the new box-plot will contain no outliers.**



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Ans : Between 5 to 7.**

1. Comment on the skewness of the dataset.

**Ans: As per the visualization the data looks positively/ Right side 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.

**Ans: 1. Both histogram and boxplot shows that the data is positively skewed.**

**2. Boxplot shows that 1 outlier is present in the dataset.**

**3. By histogram we can see the frequency of the data present in the dataset.**

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.)

**Ans : probability is one out of 200 call is misdirected. Let, n(E) = 1/200**

**To find probability(P) that at least one in five attempted reached wrong number is, n(S) = 1/5**

**Formula : P = n(E)/n(S)**

**P = (1/200)/(1/5)**

**P = 0.005/ 0.2**

**P = 0.025**

**Therefore, the probability of 1 in 5 attempted calls reaching wrong number is 0.025.**

1. 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?

**Ans : 2000**

1. Is the venture likely to be successful? Explain

**Ans : Probability of getting profit P(0.2+0.3+0.1) = 0.6**

**There are 0.6 i.e. 60% of positive results are shown in the table. So, we can say that venture is likely to be successful.**

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

**Ans: -2000\*1+-1000\*1+1000\*2+2000\*3+3000\*1 = 8000/6 = 1333.3**

**Long term average earning is 1333.3**

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

**Ans: Here, the good measure is positive results i.e. profits are 60% which is more than negative results i.e. loss 20%**

**Probability of profit is 60%.**

**Probability of loss is 20%.**