**Assignment2(1)**

**Topics: Descriptive Statistics and Probability**

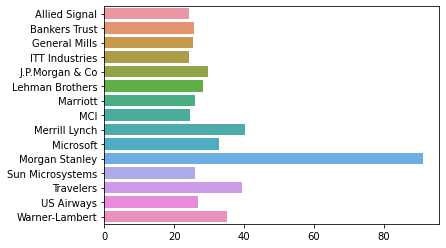
Que.1 Look at the data given below. Plot the data, find the outliers and find out

|  |  |
| --- | --- |
| **Name of company** | **Measure X** |
| Allied Signal | 24.23% |
| Bankers Trust | 25.53% |
| General Mills | 25.41% |
| ITT Industries | 24.14% |
| JPMorgan & 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% |

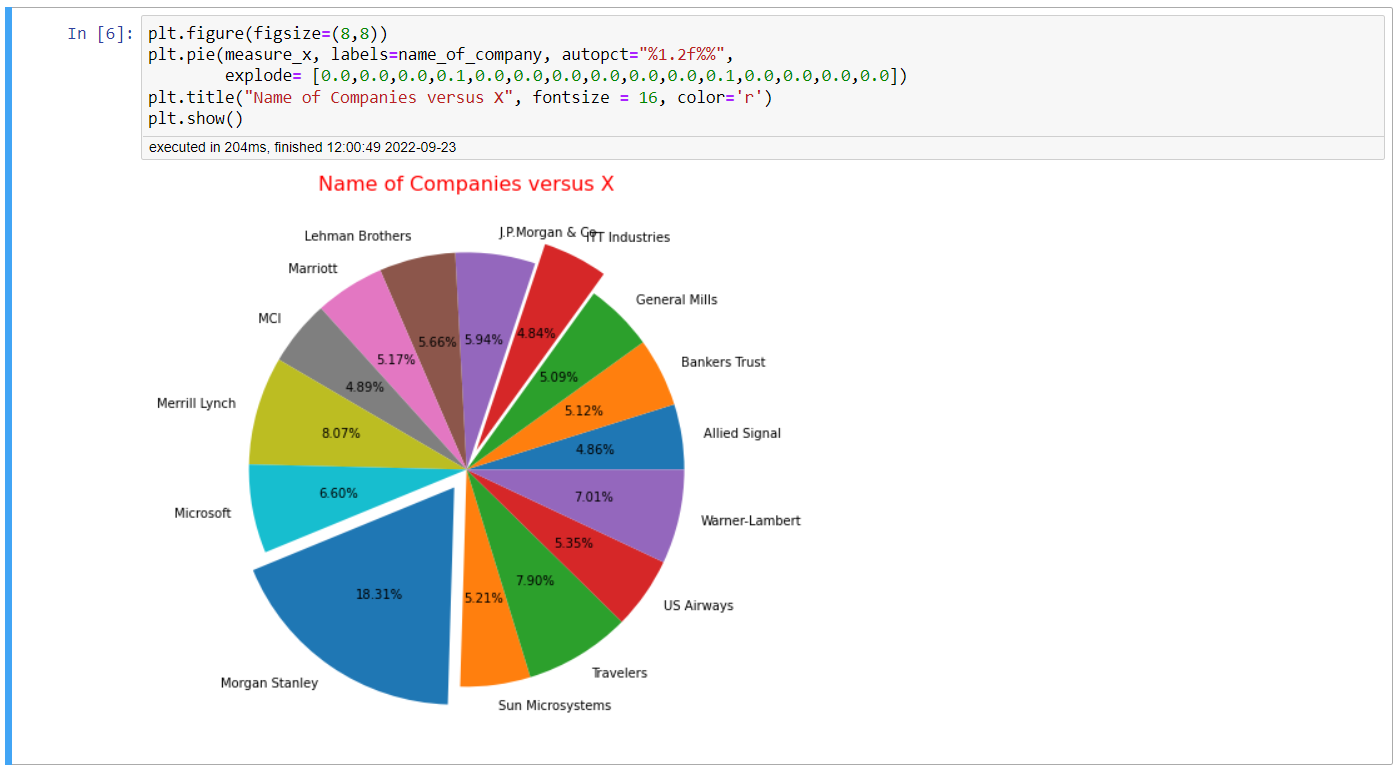
Ans:

Plotting bar chart:





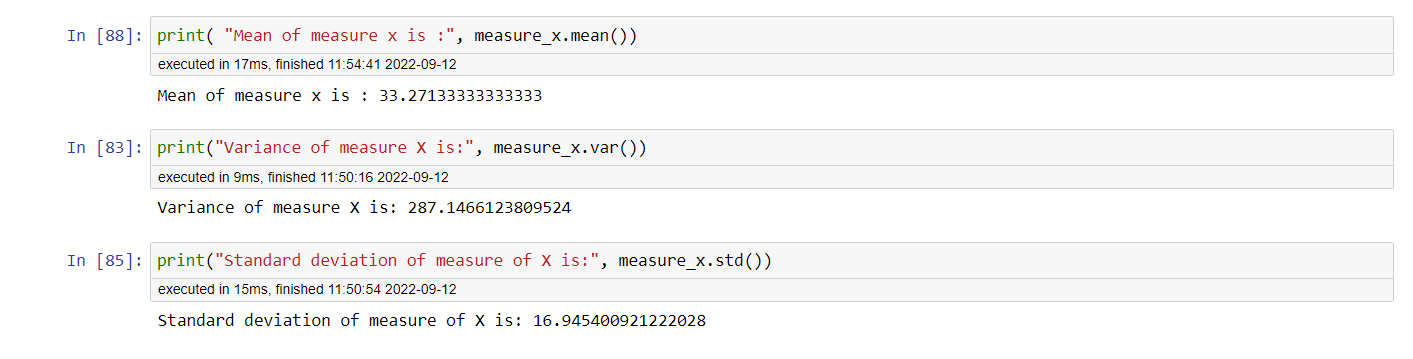
Plotting pie chart:



Finding outliers:



Finding Mean, Variance, Standard deviation:



Summary:

Outlier on Upper side is: 91.36(Morgan Stanley)

Mean of measure x is: 33.2713

Variance of measure X is: 287.1466

Standard deviation of measure of X is: 16.9454

Que.2



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

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

Ans: Approximating 3rd Quartile (Q3) =12 & 1st Quartile (Q1) = 5. Hence, Inter Quartile Range is Q3-Q1= 12-5=7

IQR measures the spread of the middle half of your data. It is the range for the middle 50% of your sample.

Here, middle 50% data of variable X has spread over 7units.

B) What can we say about the skewness of this dataset?

Ans: As median is shifted to the left, it is right skewed data.

C) 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: Data point having value 25 is a outlier in our boxplot. If its actual value is 2.5, then it will lie below 1st quartile instead of being outlier at upper end. So new box-plot will be affected.

Que.3



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans: Mode of this dataset would be between 5 to 10

1. Comment on the skewness of the dataset.

Ans: Here, Mean> Median due to right tail. Hence, Data is Right 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: Both histogram and boxplot are right skewed having outlier on 25. Besides, boxplot shows that initial 50% of data is concentrated, and later 50% data is spreading. This is supported by histogram where height of initial few bars is longer indicating concentration data points.

Que4.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:

In a binomial distribution,

P(x) = ⁿCₓ pˣ qⁿ⁻ˣ

where n = the number of trials

x = number of times a particular outcome is attained

p = probability of success

q = probability of failure

P (X= at least one in five attempted telephone calls reaches the wrong number)

n= Number of Calls = 5

x=1

p = probability of call misdirecting   = 1/200

q = Probability of call not Misdirecting = 1-1/200 = 199/200

P (1) = (5C1) \* (1/200)^1 \* (199/200)^5-1

P (1) = 0.0245037

The probability that at least one in five attempted telephone calls reaches the wrong number is 0.0245037

Que5.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: Most likely monetary outcome of business venture is $2000, since it has the highest probability among rest of the outcomes.

1. Is the venture likely to be successful? Explain

Ans: Yes, Venture is likely to be successful since probability of not making loss is 0.8

i.e P (X=0) + P(X=1000) + P(X=2000) + P(X=3000) is 0.2+0.2+0.3+0.1=0.8=80%

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

Ans: Long term average earning can be known by Expected value.

E(X)= ΣX\*P(X) = -2000\*0.1 -1000\*0.1 +0 +1000\*0.2 +2000\*0.3 +3000\*0.1 = 800

Hence, long-term average earning of business ventures would be $800

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

Ans: Risk involved in a venture of this kind can be gauged by the Variance in earnings.

Given data can be written as X= [-2000, -1000, 0, 0, 1000, 1000, 2000, 2000, 2000, 3000]

Mean=800 and Standard deviation= 1549.19

Coefficient of Variation = Std Deviation / Mean = 1549.19/800 = 1.93

Coefficient of variation allows investors to determine how much volatility, or risk Involved in venture. CV<1 signifies volatility is less since variation in data is less.

But here, CV is 1.9 indicating high variation and hence high volatility.

*Thank You*