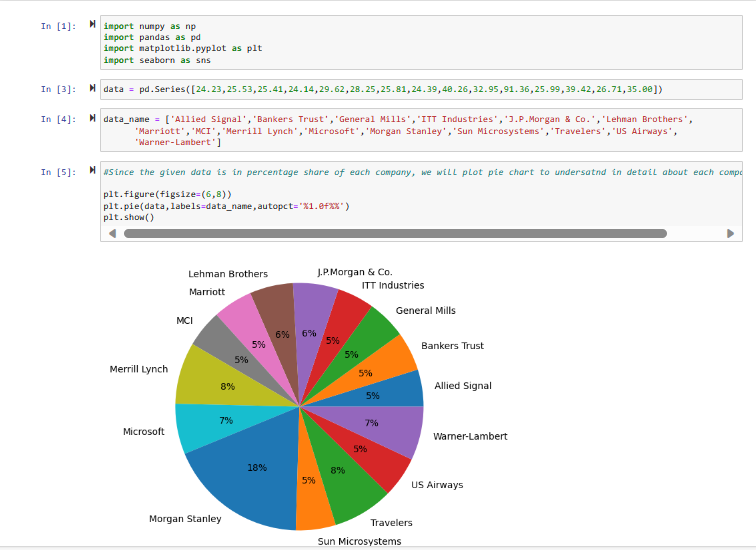
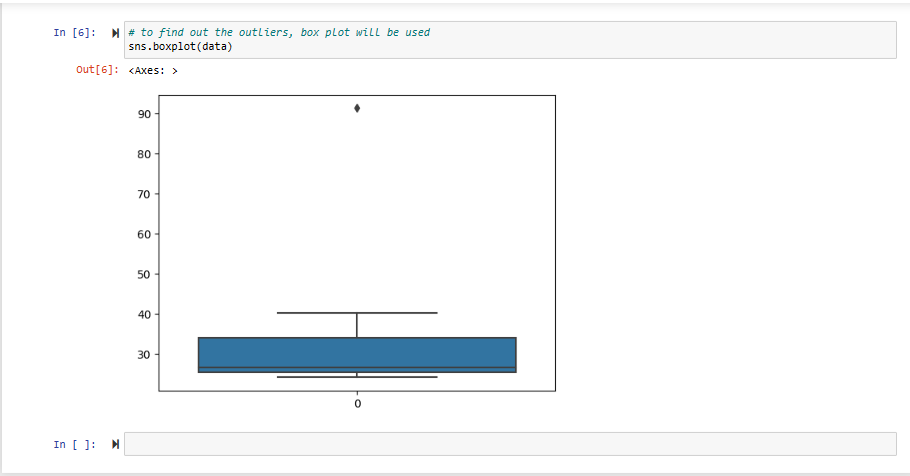
**Topics: Descriptive Statistics and Probability**

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

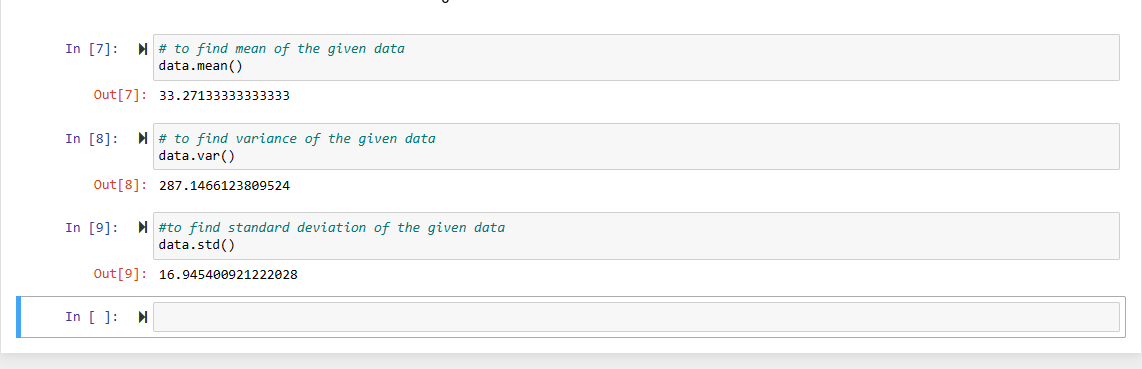
|  |  |
| --- | --- |
| **Name of company** | **Measure X** |
| Allied Signal | 24.23% |
| Bankers Trust | 25.53% |
| General Mills | 25.41% |
| ITT Industries | 24.14% |
| J.P.Morgan & Co. | 29.62% |
| Lehman Brothers | 28.25% |
| Marriott | 25.81% |
| MCI | 24.39% |
| Merrill Lynch | 40.26% |
| Microsoft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



Since the given data is in percentage form from multiple industries, choosing pie chart will be appropriate graph to illustrate the data comprehensively. The above pie chart suggests that Morgan Stannley has the largest measure of X with 18% followed by Travelers & Merrill Lynch with each have 8% measure of X.



There is one outlier in the entire data set which the above box plot is suggesting and the outlier name is Morgan Stanley with 91.36%.



**Answers:**

* Outlier – Morgan Stanley with 91.36%
* Mean – 33.271
* Variance – 287.146
* Standard Deviation – 16.945



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:**

The Q1 range = 5 (approximately)

The Q3 range = 12 (approximately)

Inter-quartile range = Q3 – Q1 = 12-5 = 7

The inter Quartile range suggests that approximately half of the data are within 7 units of the median

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

**Answer:**

The above box suggests that the median is slightly towards lower quartile range suggesting that the data may be slightly 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?

**Answer:**

If the data point with the value is actually 2.5 then there won’t be outlier in the box plot also the lower quartile range will shift towards left and maybe there is a chance of decrease in the median value.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Answer:**

The mode of the data set approximately lies between 5 to 10

1. Comment on the skewness of the dataset.

**Answer:**

The skewness of the dataset is right skewed or 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:**

The boxplot & the histogram, are both skewed right side with more clear reflection of outliers and its value. The box plot helps us to understand about the Q1, Q3 ranges while the histogram suggests each unit’s value more clearly. In this way they both complement each other in explaining the data and about its distribution.

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:**

Probability of 1 in 200 long distance telephone calls getting misdirected (p) = 1/200

Probability of calls not getting misdirected (q) = (1-(1/200)) = 199/200

Now, the probability that at least one in five attempted telephone calls reaches the wrong number

Number of calls (n) = 5

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

Therefore, X=1

P(X) = nCx px qn-x

P (1) = (5C1) (1/200)^1 (199/200)^5-1 …. nCr = n! / r! \* (n - r)!

= ((5\*4\*3\*2\*1)/(1!\*(5-1)!)) \* (1/200) \* (199/200)^4

= (5) \* (0.002) \* (0.981)

= 0.0098

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?

**Answer:**

The mostly likely monetary outcome of the business venture is $2000 because it has the maximum probability i.e., 0.3 reflecting 30% chance that the venture will have monetary outcome of $2000.

1. Is the venture likely to be successful? Explain

**Answer:**

The data suggests that the venture will be successful. Because the probability that the venture will earn more than 0 is

P(X>0) + P(X>1000) + P(X>2000) + P(X>3000) = 0.2+0.2+0.3+0.1 = 0.8 – which reflects that there is 80% for the venture to earn more than 0.

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

**Answer:**

To find the long-term average earning of business venture, we will calculate the Expected Value

Expected Value = (-2000 \*0.1) + (-1000 \* 0.1) + (0\*0.2) + (1000\*0.2) + (2000 \* 0.3) + (3000 \* 0.1) = (-200) + (-100) + 200 + 600 + 300 = 800

Therefore, the long term average expected earning is $800

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

**Answer:**

The good measure of the risk involved in a venture of this kind depends on the Variability in the distribution. Higher Variance means more chances of risk

Var (X) = E(X^2) –(E(X))^2

= 2800000 – 800^2

= 2160000

Std (X) = Square root of Variance

= 1469.6938