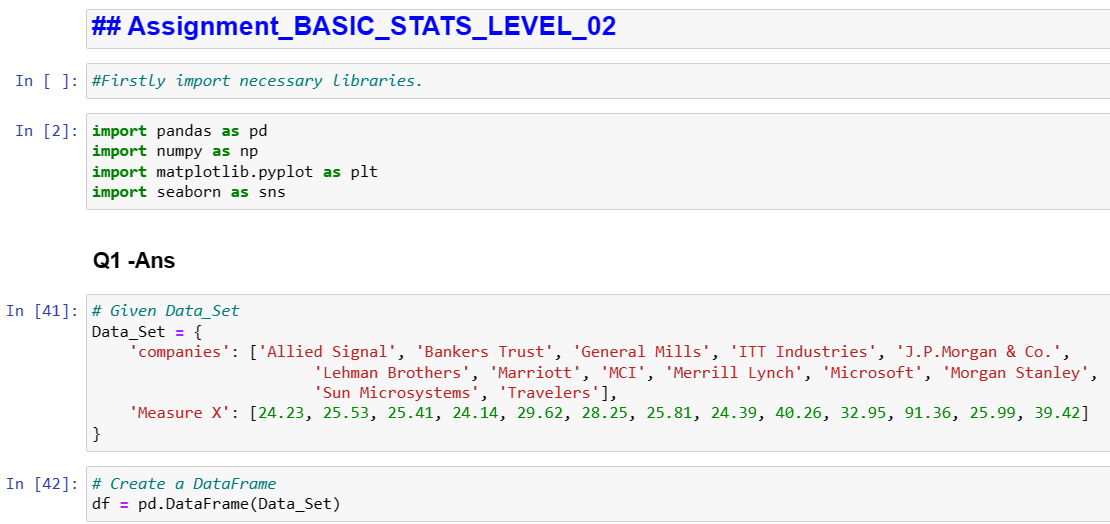
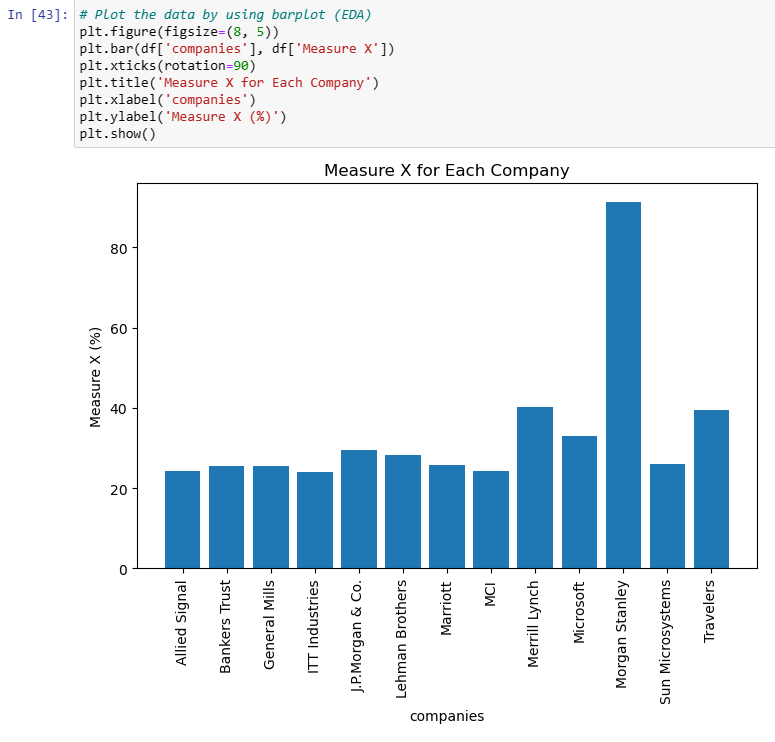
**Topics: Descriptive Statistics and Probability**

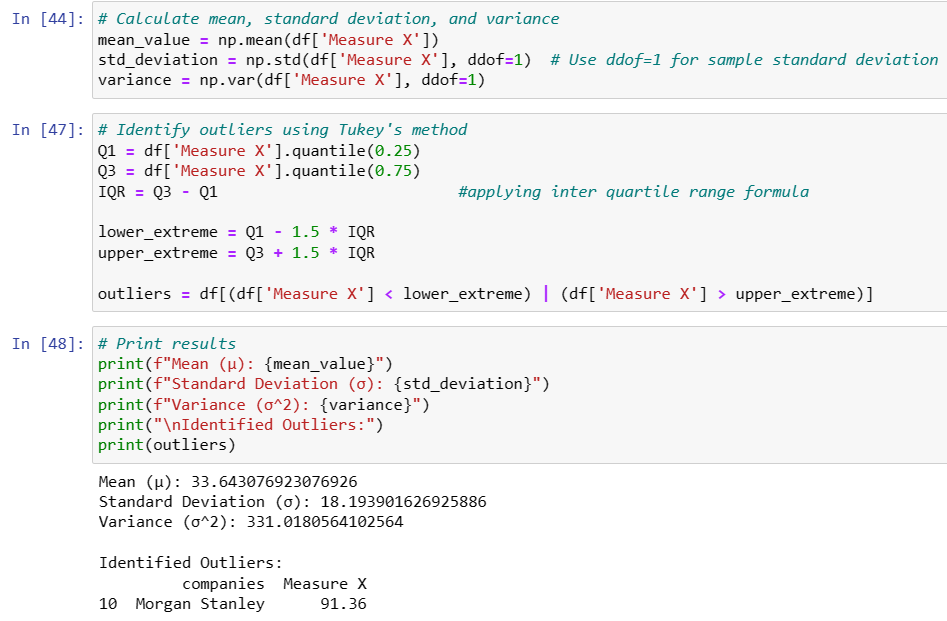
1. Look at the data given below. Plot the data, find the outliers and find out

|  |  |
| --- | --- |
| **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**- Solved in. python notebook







Mean = 33.64

Standard deviation = 18.19

Variance = 331.01

Outliers = 91.36 (Morgan Stanley)



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** - inter-quartile range we can find out by using formula

**IQR =** Q3-Q1 -----1)

Here, Q3 = 12

Q1 = 5 ( approx values)

Put these values in formula 1) to get IQR

IQR = 12-5 = **7**

* IQR generally shows us the spread of 50% of the data. Hence IQR = 7 this value implies that the central 50% of the data falls within a range of approximately 7.

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

**Ans** -

* Skewness refers to the asymmetry of a distribution in the dataset.
* In this case the whisker(1.5\*IQR) value max extending towards right side, hence we can say that the “positively skewed ”or “right skewed” data.

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 the data point with the value 25 is actually 2.5, this would significantly impact the dataset.
* It will create small boxplot value as compare to existing boxplot.
* Its directly impacts on outlier value and in this case there would be no outliers in the

dataset.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Ans –**

* Mode is nothing but the most frequent value in the dataset.
* The mode of this data set lie in between 5 to 10 and approximately between 4 to 8.

1. Comment on the skewness of the dataset.

**Ans –**

* From this above histogram the data is max distributed on the left side of the graph

and tail flows towards the right side, hence it is the “positively skewed data”.

* Mean >median>mode

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 -**

* They both are right-skewed and both have same outliers of value 25.
* The median can be easily visualized in box plot where as in histogram mode is more visible.

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** –

If 1 in 200 long-distance telephone calls are getting misdirected.

probability of call misdirecting = 1/200

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

The probability for at least one in five attempted telephone calls reaches the wrong number

Number of Calls = 5

n = 5

p = 1/200

q = 199/200

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

P(x) = ⁿCₓ pˣ qⁿ⁻ˣ

P(x) = (nCx) (p^x) (q^n-x)

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

P(1) = **0.0245037**

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

Number is **0.0245037.**

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** -

* The most likely monetary outcome is depends on the highest probability (P(x)).
* In this case, the value 2000 has the highest probability of 0.3, so the most likely monetary outcome is $2000.

1. Is the venture likely to be successful? Explain

**Ans** –

* Yes, the probability that the venture will make more than 0 or a profit
* p(x>0)+p(x>1000)+p(x>2000)+p(x=3000) = 0.2+0.2+0.3+0.1 = 0.8 this states that there is a good 80% chances for this venture to be making a profit

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

**Ans –**

* Average earning nothing but the expected value(E). It is calculated by some of the product of x and probability.
* E(x) =(−2000)(0.1)+(−1000)(0.1)+(0)(0.2)+(1000)(0.2)+(2000)(0.3)+(3000)(0.1)

E(x) =−200−100+0+200+600+300

E(x) = 800

* The long-term average earning is $800.

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

**Ans** –

One measure of risk is the standard deviation (σ).

σ=∑(x−μ)2⋅P(x), where μ is the mean (expected value).

Calculate σ=∑(x−800)2⋅P(x)

σ≈1099.67

Now, we will find variance

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