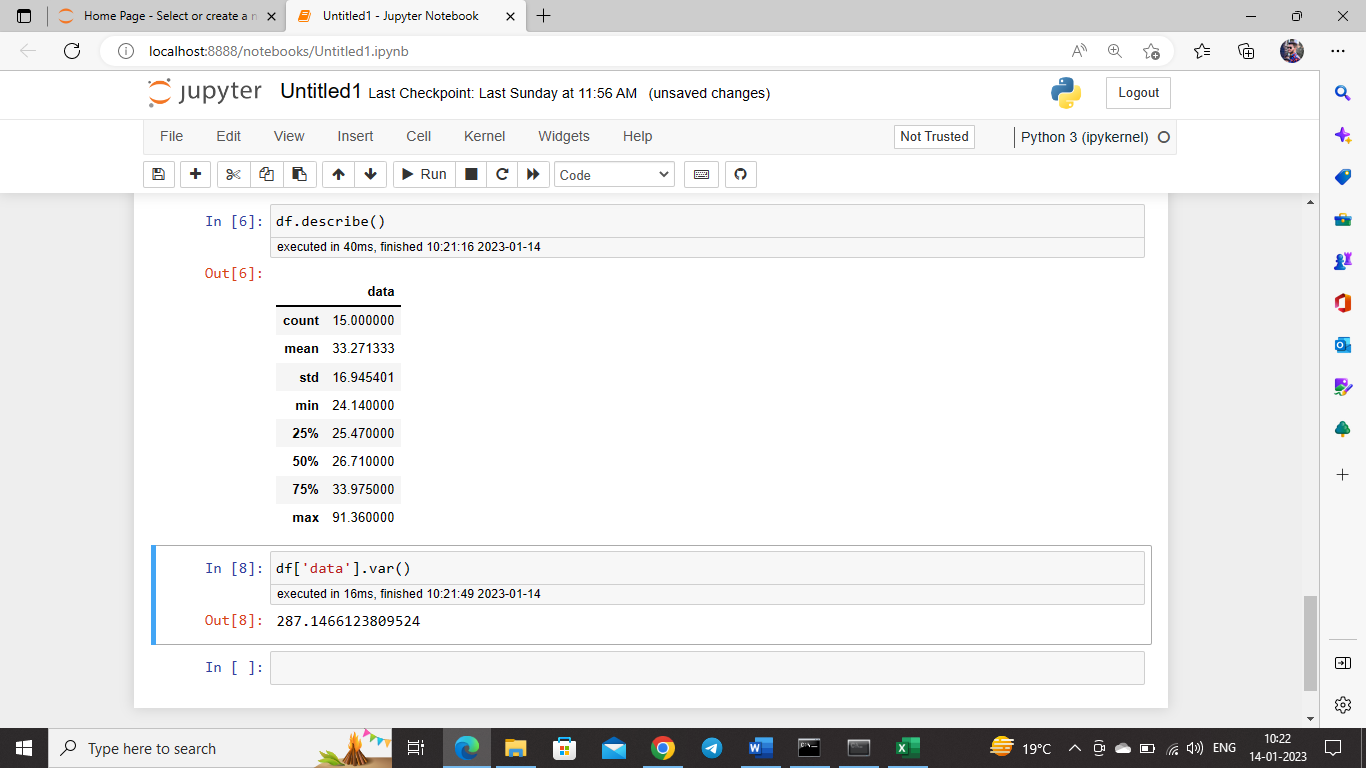
**Topics: Descriptive Statistics and Probability**

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:** From the data, we see that there is one outlier i.e., 91.36 at 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:** IQR = Q3 – Q1

= 12 – 5

= 7

IQR tells us the spread of middle 50% observations. In our problem, the IQR value is 7, which implies spread of our middle 50% observations is 7 unit.

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

**Ans:** In this boxplot, we see that, the median is closer to the bottom of the box, and the whisker shorter on the lower end of the box, this shows that our data is positively skewed.

Majority of the data points are on the left side of the distribution.

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 value is actually 2.5 then there would be no outliers on our dataset and our median is also slightly shifted to the closer to the bottom of the box.



**Answer the following three questions based on the histogram above.**

1. **Where would the mode of this dataset lie?**

**Ans:** The mode of the dataset would be lie between 5 to 6.

1. **Comment on the skewness of the dataset.**

**Ans:** From above histogram, we see that majority of the data points are lies on the left side of the distribution and the longer tail to the right side, so we can say that the data is 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.**

**Ans:** Both graphs show positively skewed data and both graphs show outliers are present in right side of the distribution. In boxplot median is clearly visible and in histogram we see the mode value easily.

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:**

We have that, one in 200 long-distance telephone calls is misdirected.

Then probability of long-distance telephone calls is misdirected= 1/200 = 0.005 = P

Probability of long-distance telephone calls is not misdirected is = 1-1/200= 0.995 = q

We have to find that, at least one in five attempted telephone calls reaches the wrong number,

Then this follows Binomial distribution,

Here,

Number of phone calls is = n= 5

P=0.005

q = 0.995

Let, X be the event that, one in five attempted telephone calls reaches the wrong number is,

P(X)= ⁿCₓ pˣ qⁿ⁻ˣ

= 5c1 \* (0.005)1 \* (0.995)5-1

= 120 \* 0.005 \* 0.9801

= 0.02487

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 of the business venture is, $2000. Because it has highest probability as compared to others.

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

**Ans:** Yes. Venture is likely to be successful because, probability of making more than 0 profit is p(x>0) + p(x>1000) + p(x>2000) + p(x>3000) = 0.2 + 0.2 + 0.3 + 0.1 = 0.8 **i.**e. There is 80% chances of this venture making profit.

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

**Ans:** The long-term average earning is nothing but Expected value

E(X) = sum(x\*p(x))

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

= -100 + -200 + 200 + 600 + 300

= 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:** The good measure of risk involved is variance. Higher the variance more the risk involved.

V(X) = E(X^2) – E(X)^2

= 2800000 -800^2

= 2160000