|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continuous |
| Weight of Gold | Continuous |
| Distance between two places | Continuous |
| Length of a leaf | Continuous |
| Dog's weight | Continuous |
| Blue Color | Nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Discrete |

Q1) Identify the Data type for the Following:

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

|  |  |
| --- | --- |
| Data | Data Type |
| Gender | Nominal |
| High School Class Ranking | Ordinal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Nominal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Nominal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Nominal |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Ratio |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?

Answer: Total number of outcomes = 2^3 = 8

Total combinations are {HHH,HHT,HTH,THH,THT,HTT,TTT}

Having two head and one tail ={HHT,HTH,THH}= 3

Probability of getting two heads and one tail = 3\8 = 0.375 =37.5%

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1
2. Less than or equal to 4
3. Sum is divisible by 2 and 3

Answer: total number of outcomes=36

The possible sum of dice rolls are (2,3,4,5,6,7,8,9,10,11,12)

The possible sum of dice rolls are (2,3,4,5,6,7,8,910,11,12)

A)0% Probability that sum is equal to 1

B)There are 3 possible number of combinations when two dice are rolled the sum

Is less that or equal to 4

Probability = 3/36=1/12=0.0833=8.33%

C) Two sum 6 and 12 are divisible by 2 and 12 sum having 5 possible that is {(6,6)}

Probability of getting sum is divisible by 2 and 3 is : 6/36=0.1667=16.7%

Q5) A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn is blue?

Answer:

Two balls are drawn at random

Number of balls without the blue colour :2+3+2=7Number of ways drawing 2 balls out of total balls = 21Number of ways drawing 2 balls out of total balls without blue balls = 10

Probability of none balls drawn is blue ball = 10/21=0.4761 =47.61%

Q6) Calculate the Expected number of candies for a randomly selected child

Below are the probabilities of count of candies for children (ignoring the nature of the child-Generalized view)

|  |  |  |
| --- | --- | --- |
| CHILD | Candies count | Probability |
| A | 1 | 0.015 |
| B | 4 | 0.20 |
| C | 3 | 0.65 |
| D | 5 | 0.005 |
| E | 6 | 0.01 |
| F | 2 | 0.120 |

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

Answer:

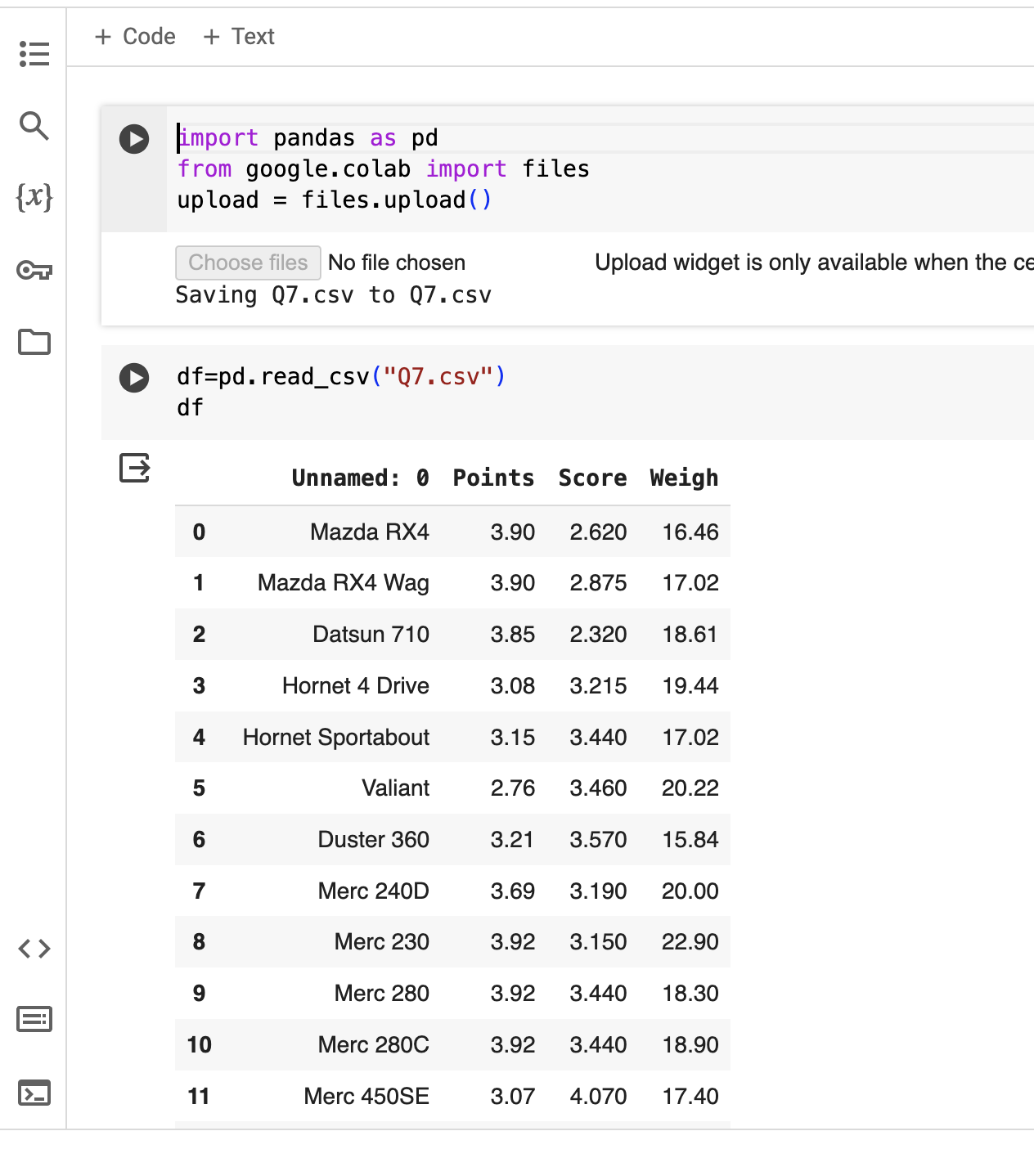
|  |  |  |  |
| --- | --- | --- | --- |
| CHILD | Candies count | Probability | Expected number of candies |
| A | 1 | 0.015 | 0.015 |
| B | 4 | 0.20 | 0.8 |
| C | 3 | 0.65 | 1.95 |
| D | 5 | 0.005 | 0.025 |
| E | 6 | 0.01 | 0.06 |
| F | 2 | 0.120 | 0.24 |

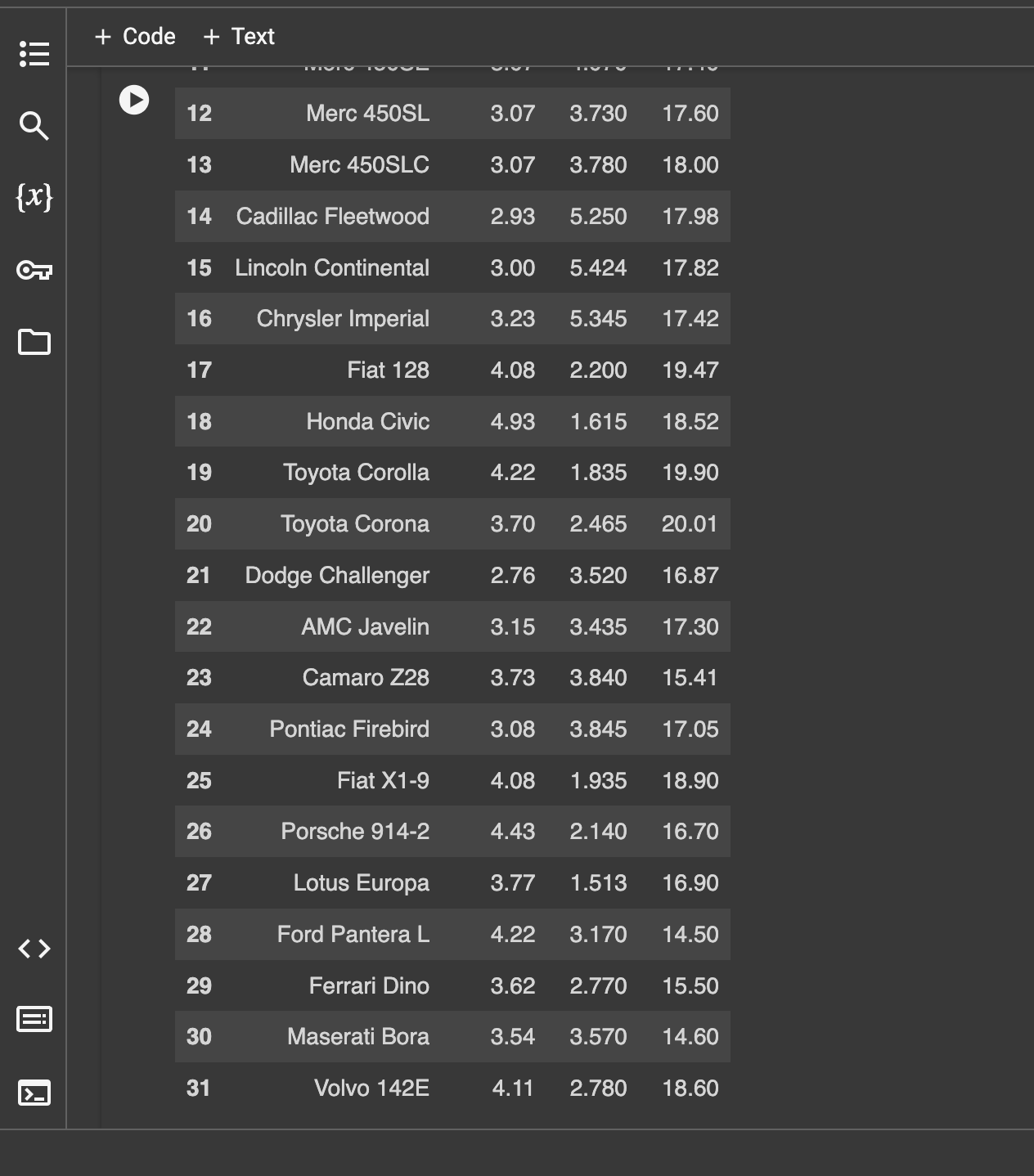
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

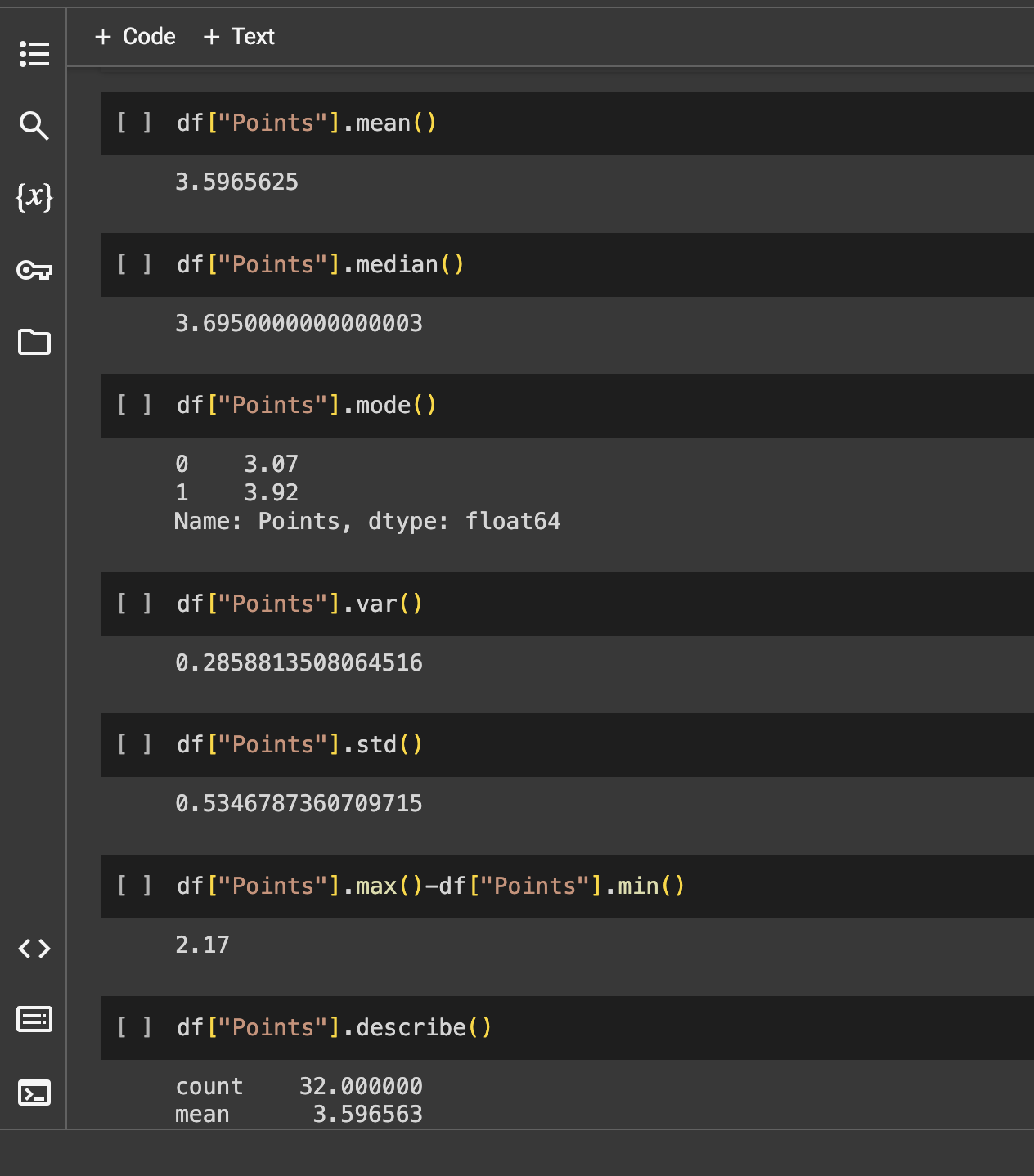
* For Points,Score,Weigh>

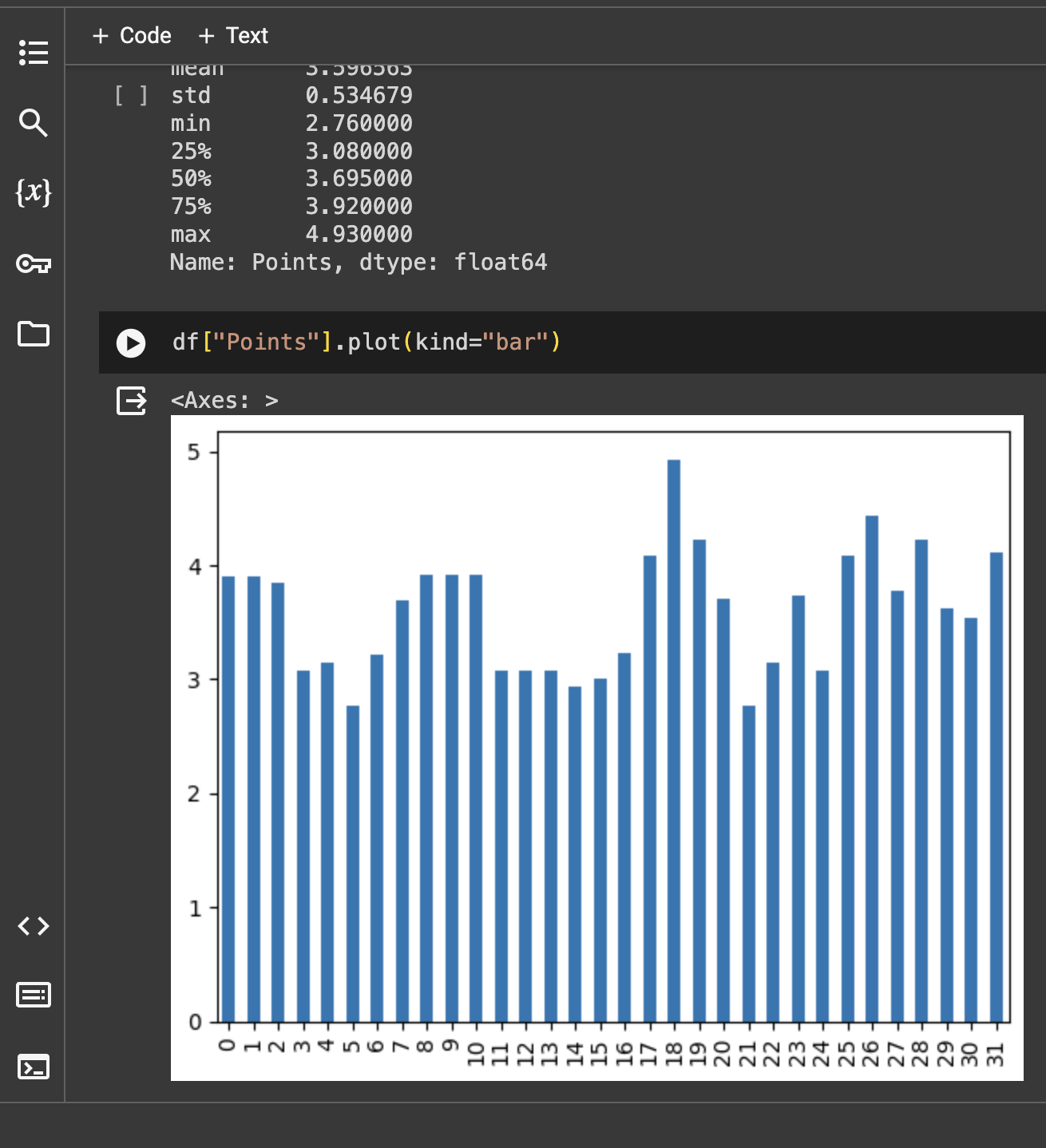
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

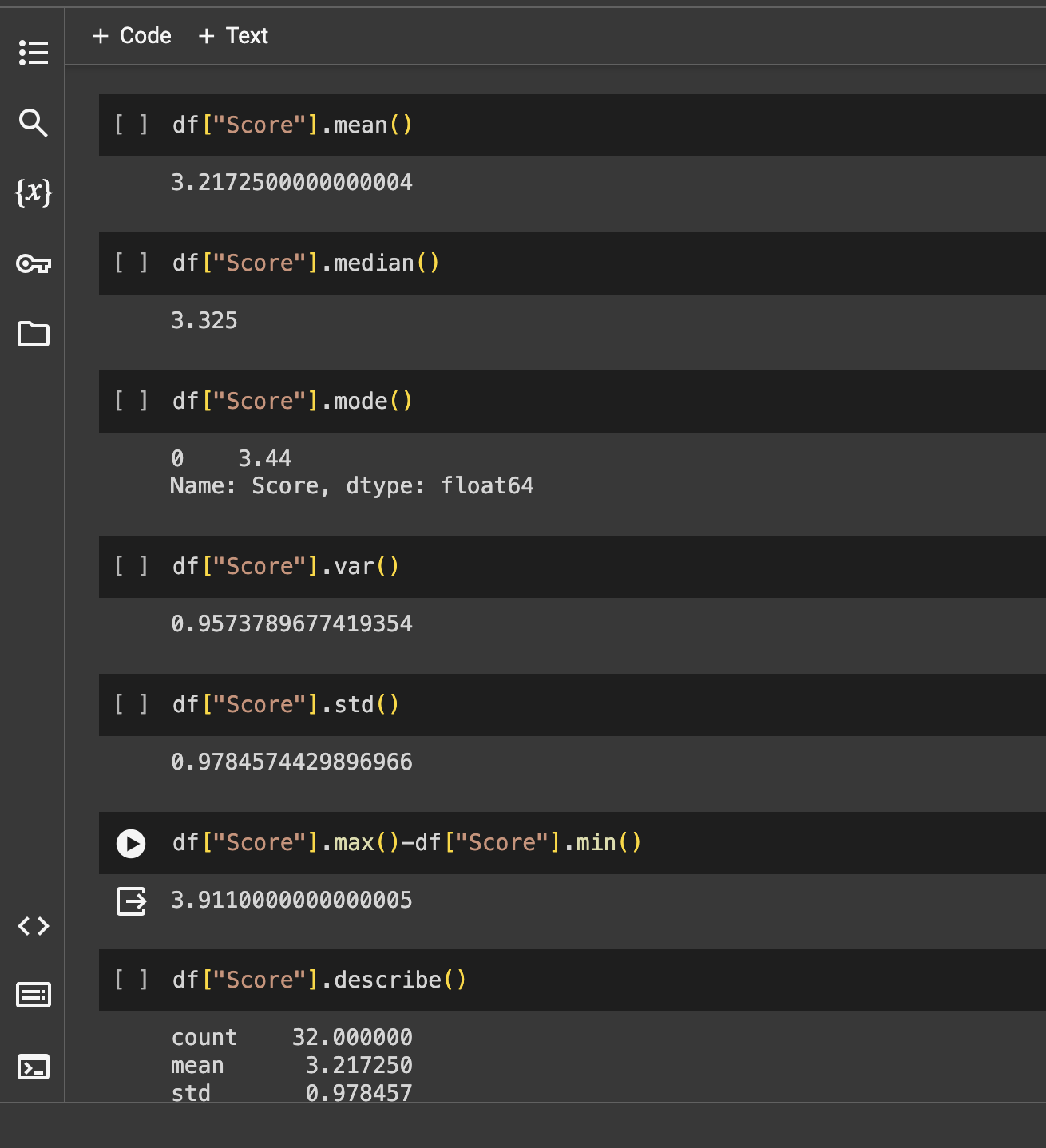
**Use Q7.csv file**

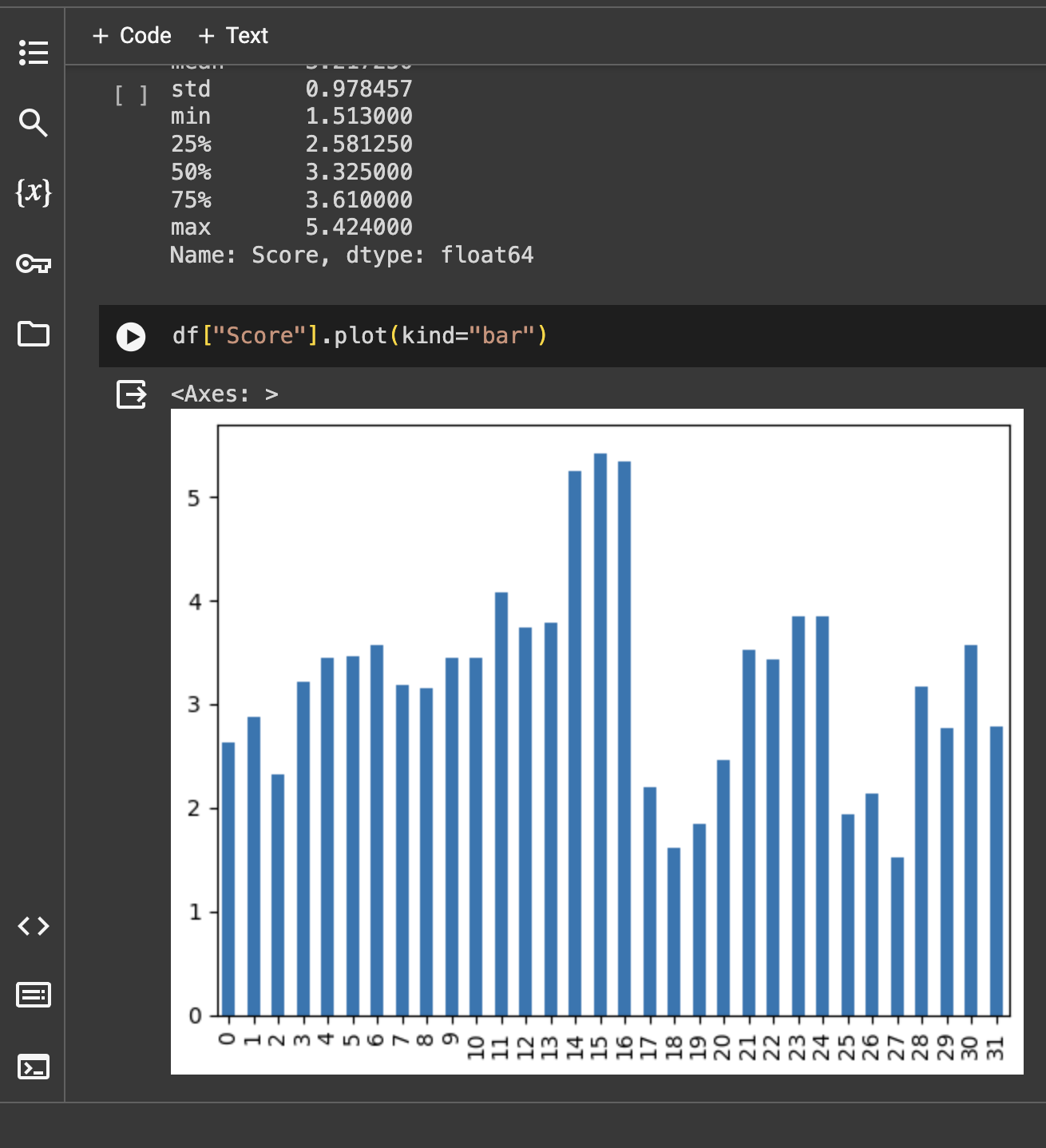
****

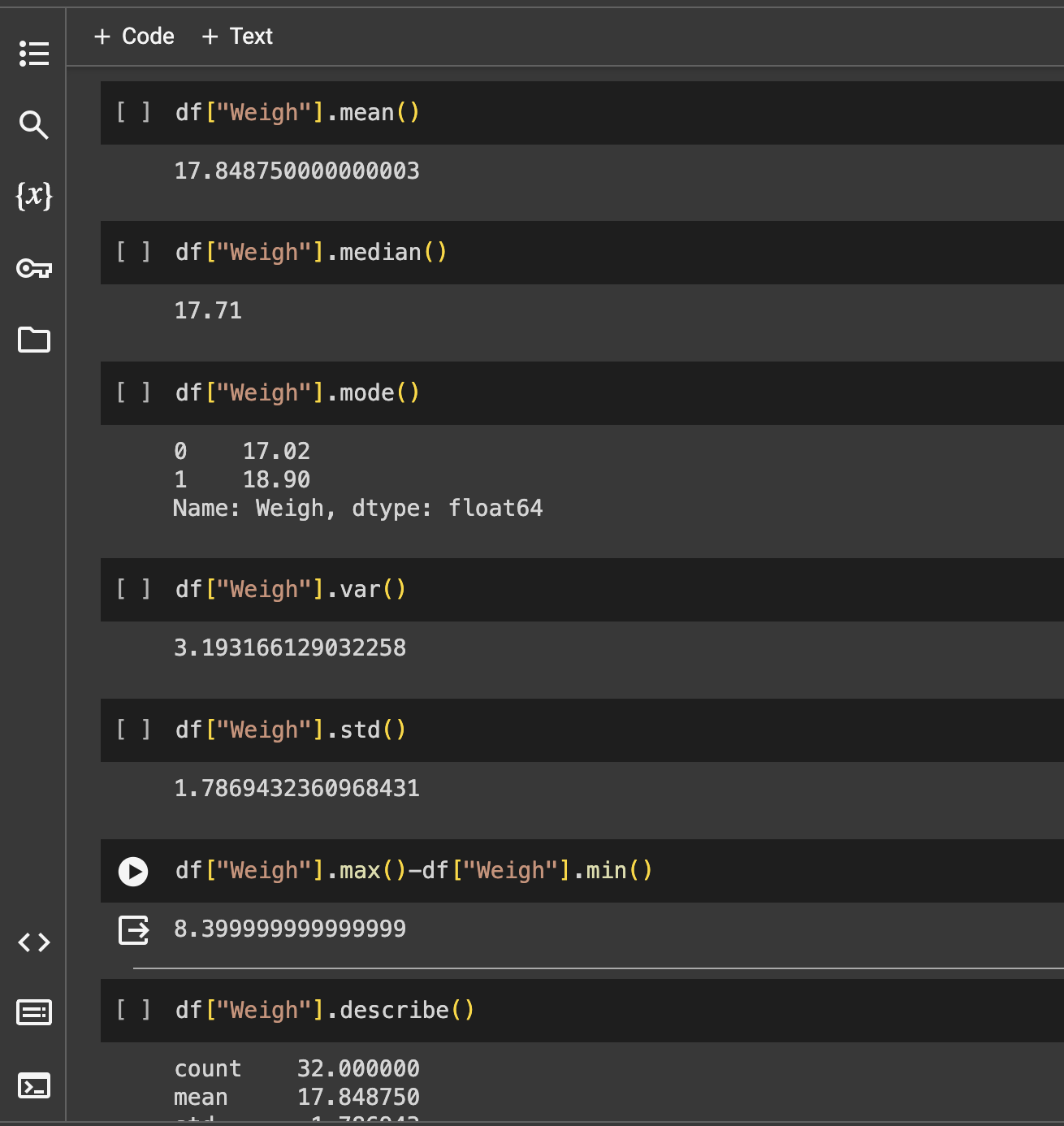
****

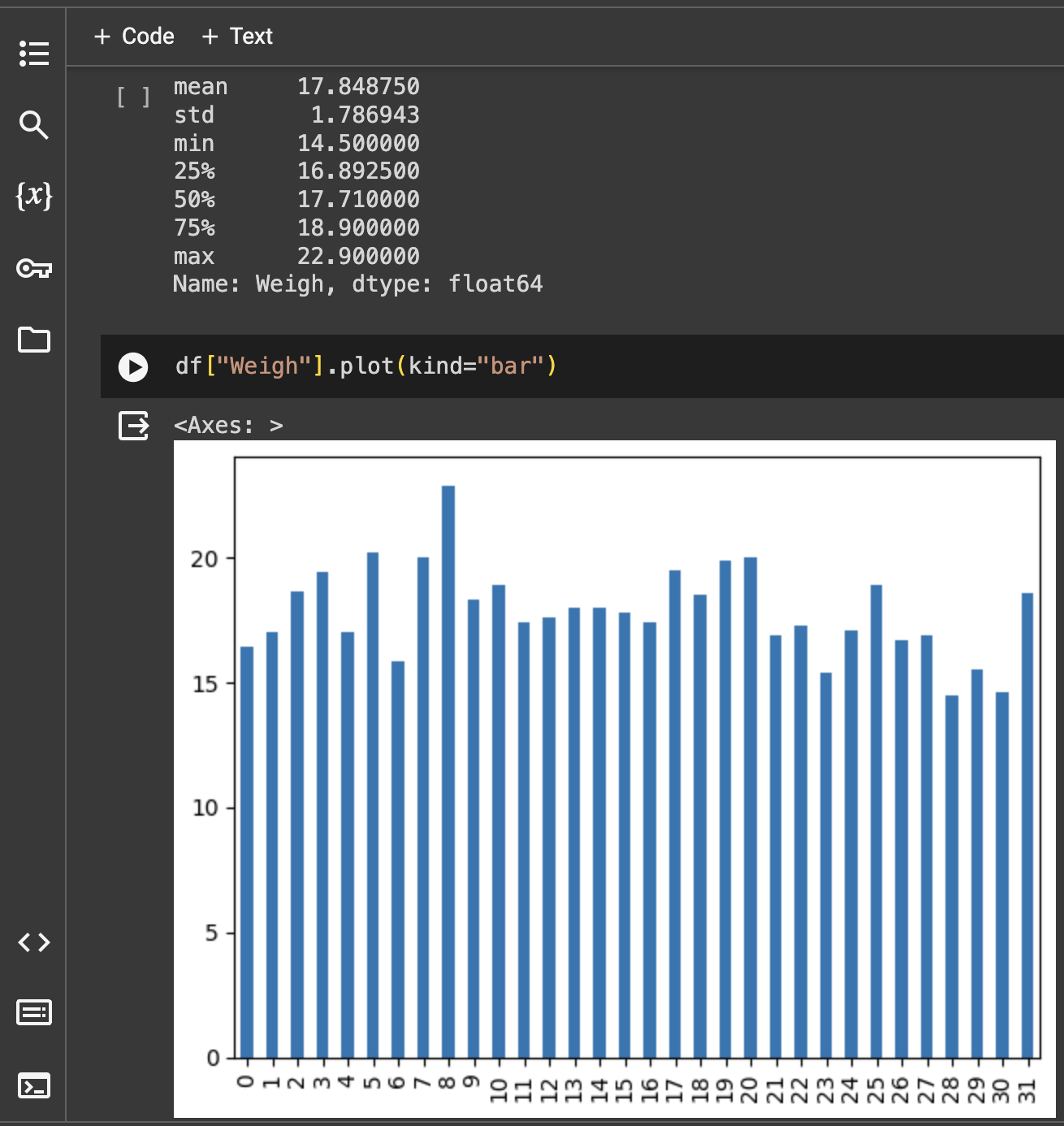
****

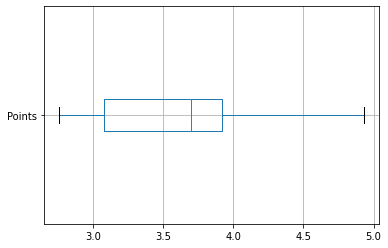
****

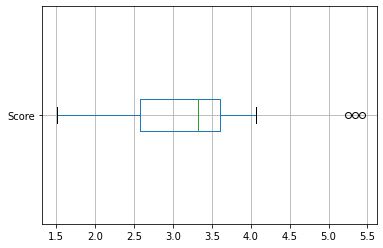
****

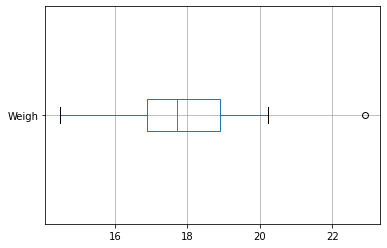
****

****

****

****

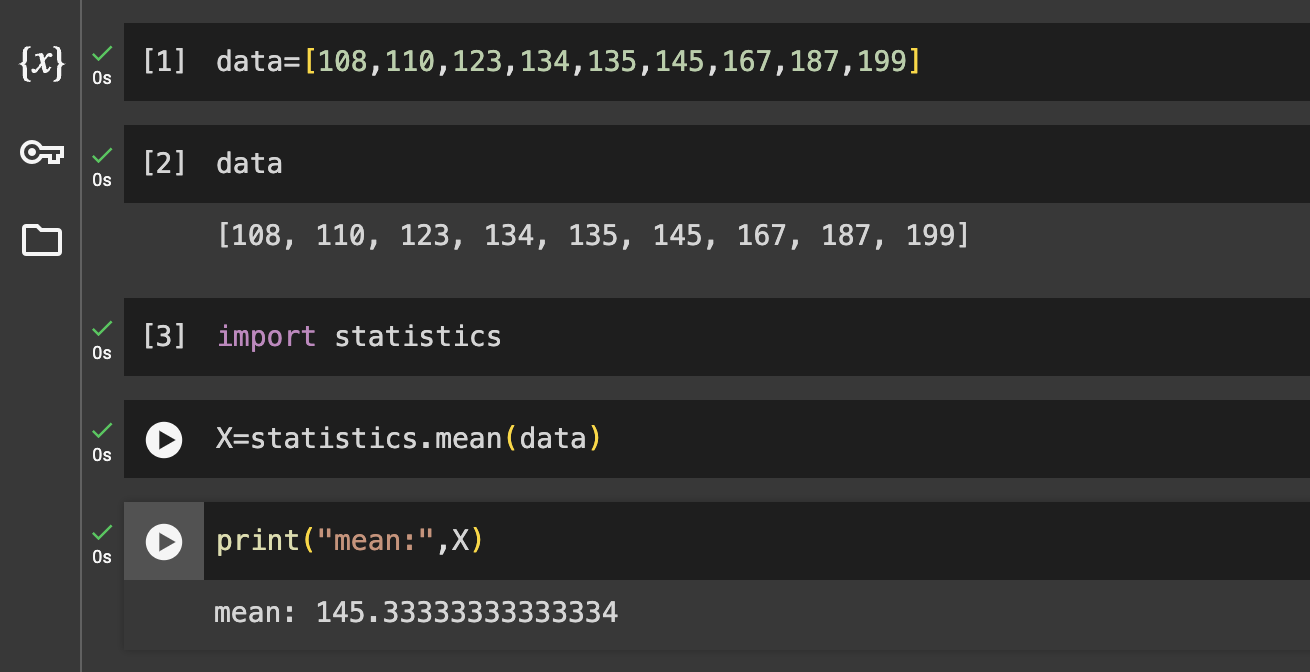
****

****

Q8) Calculate Expected Value for the problem below

1. The weights (X) of patients at a clinic (in pounds), are

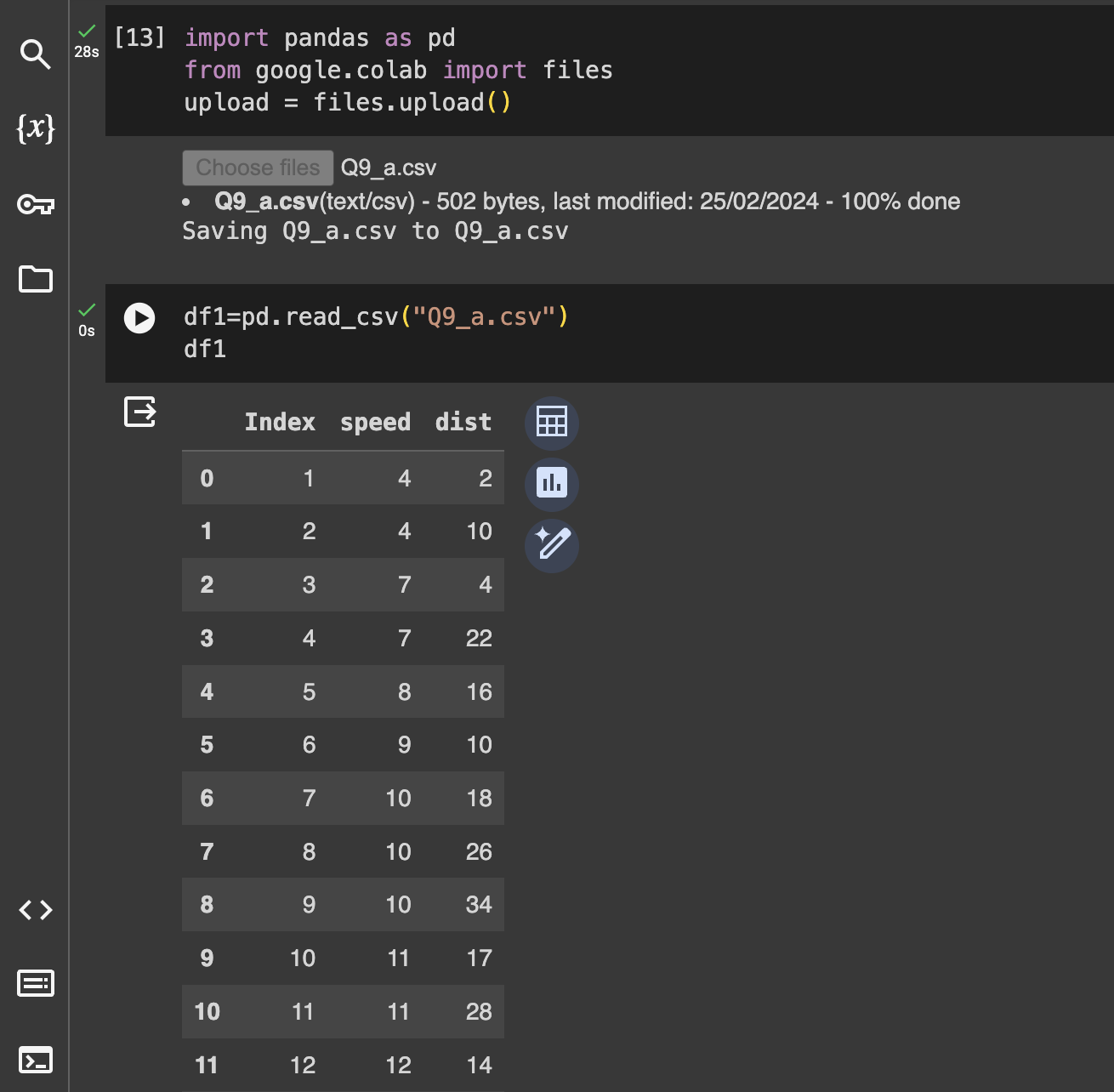
108, 110, 123, 134, 135, 145, 167, 187, 199

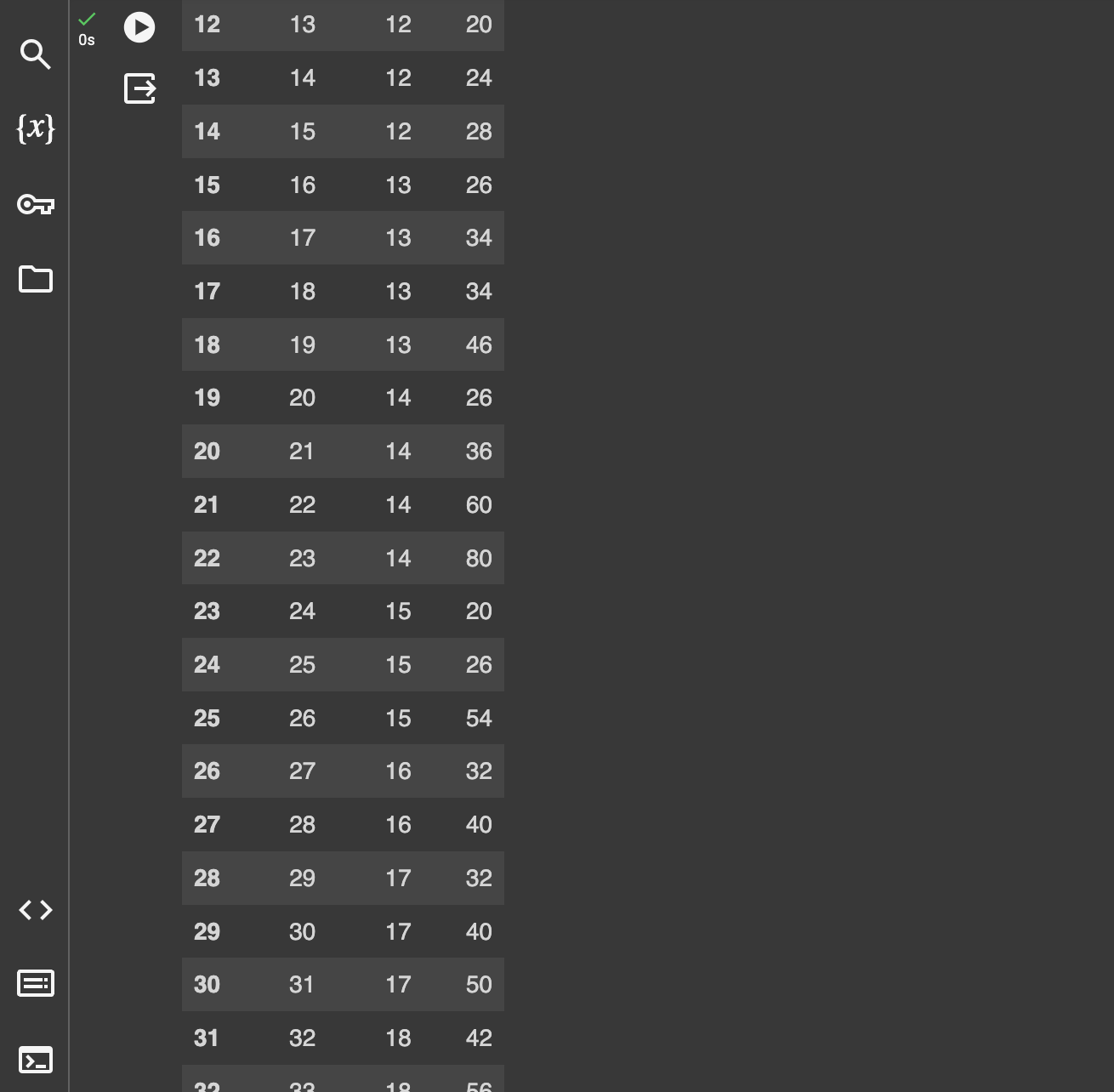
Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

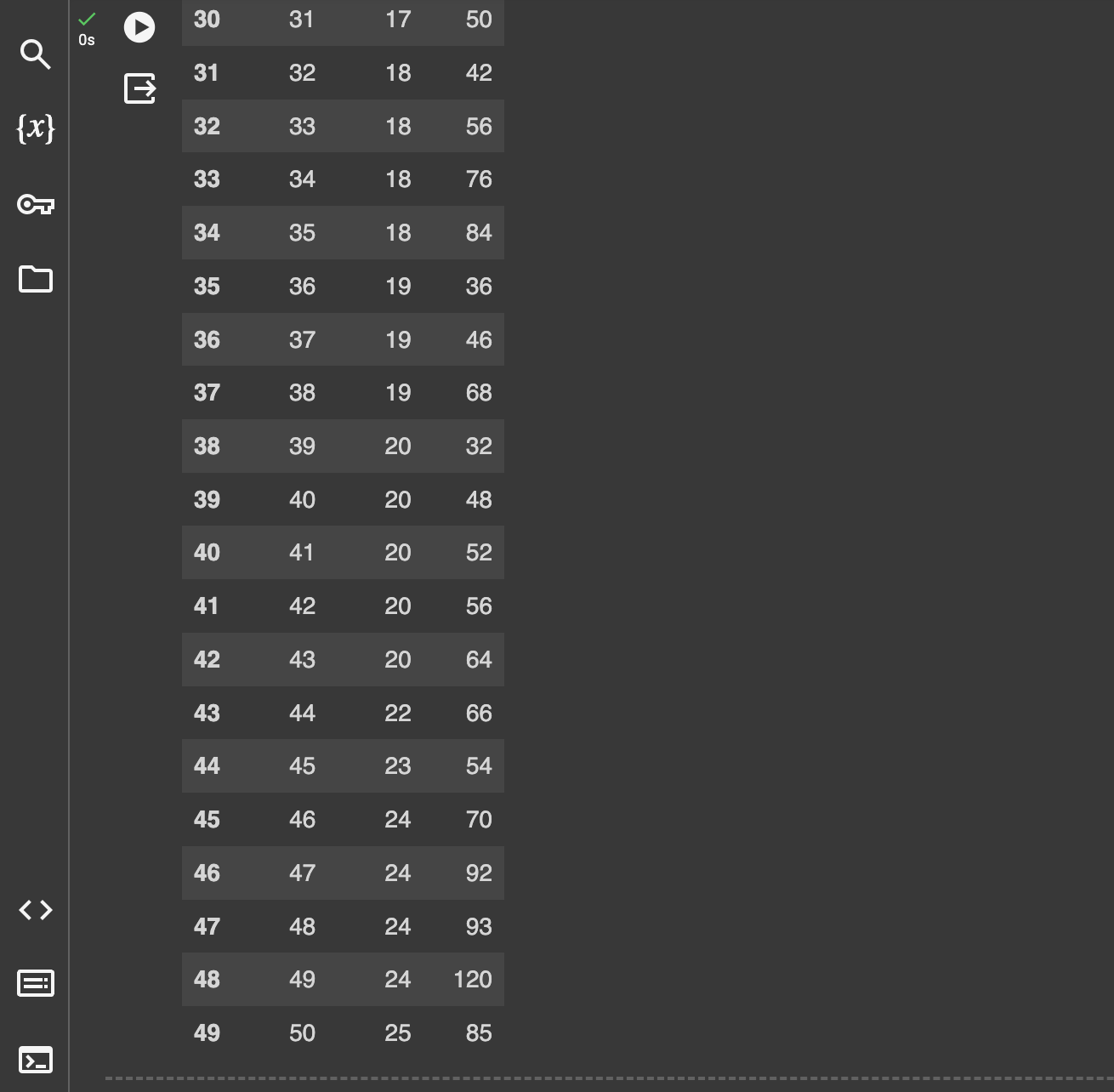
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

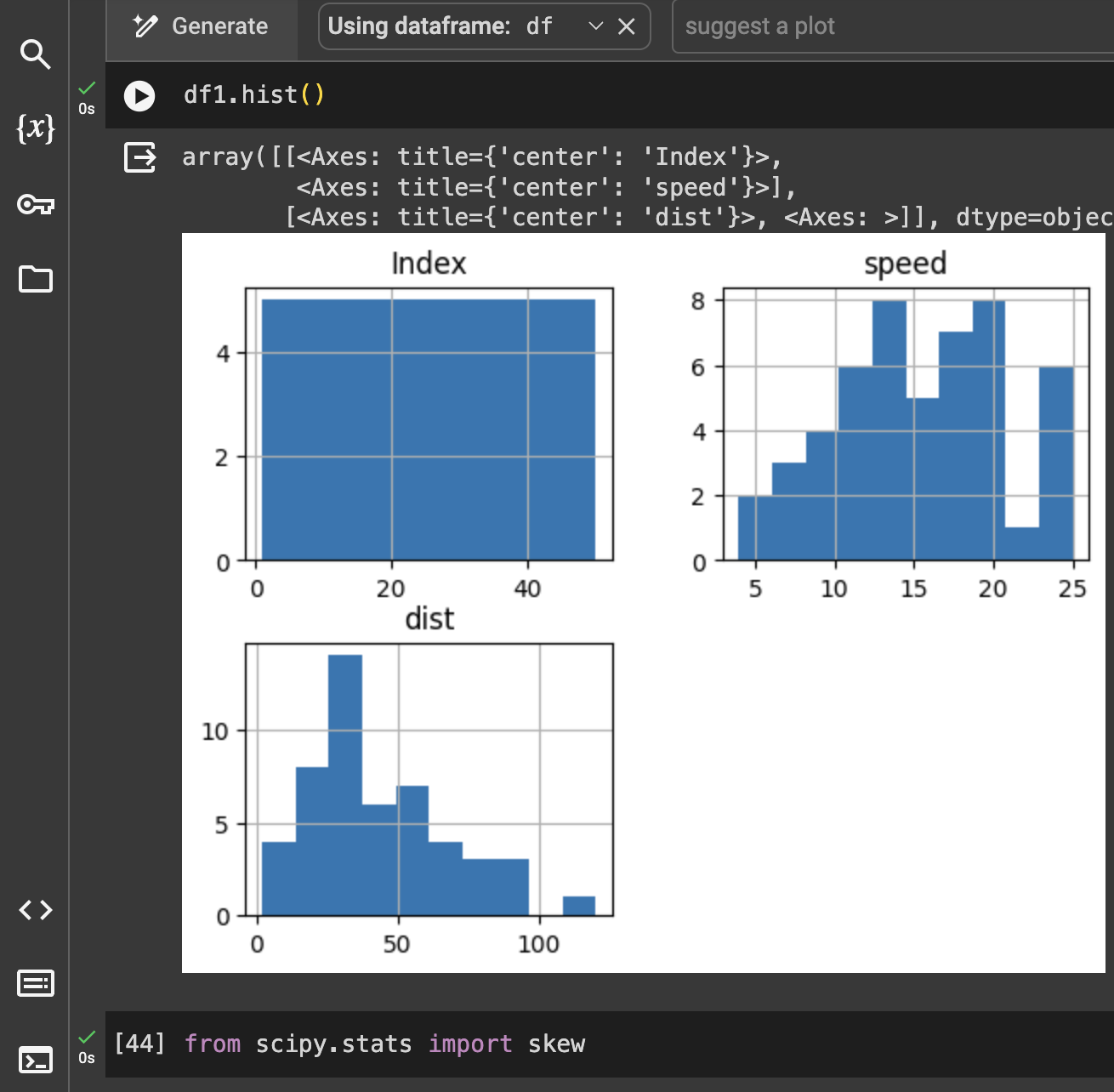
**Cars speed and distance**

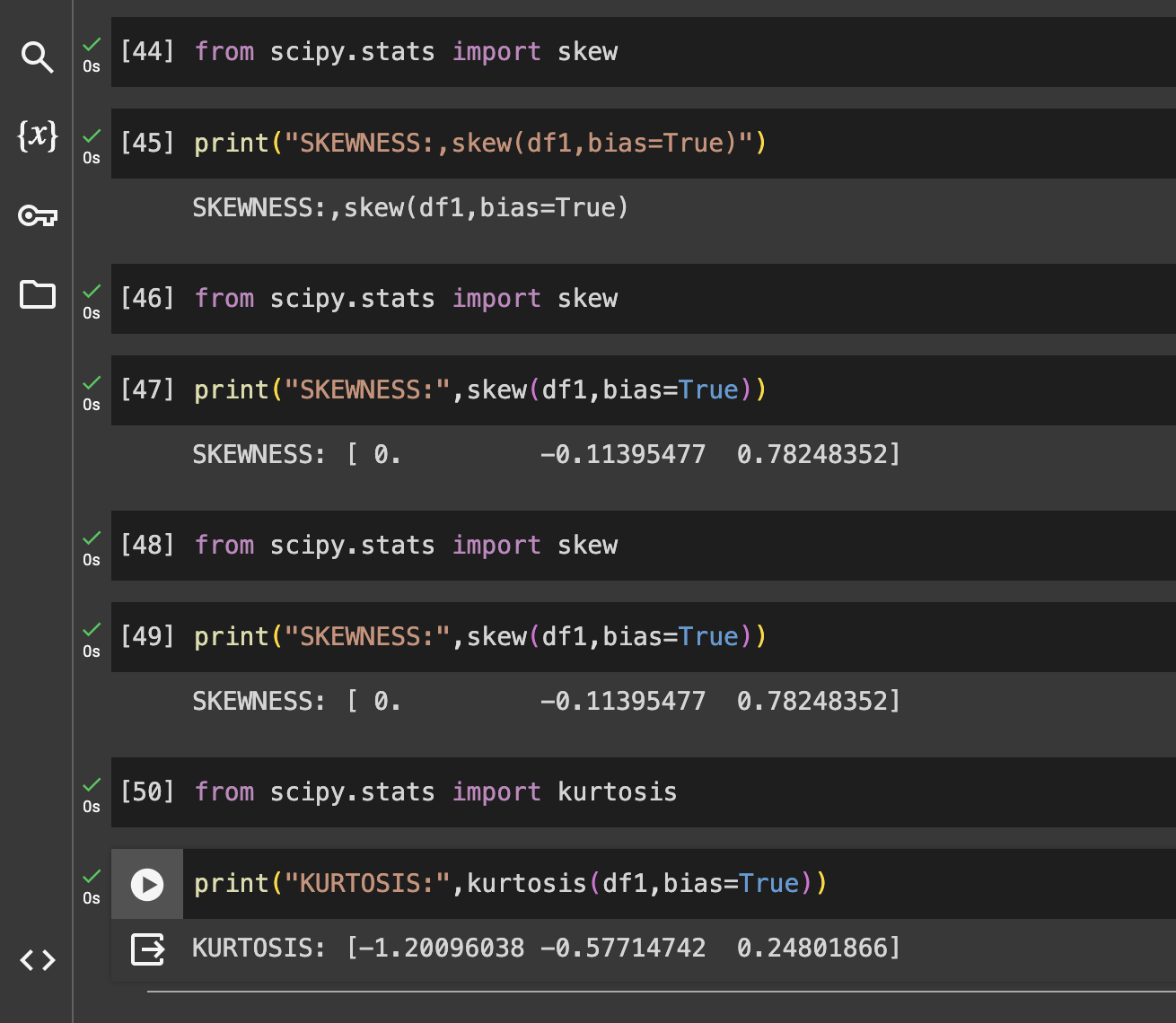
**Use Q9\_a.csv**

****

****

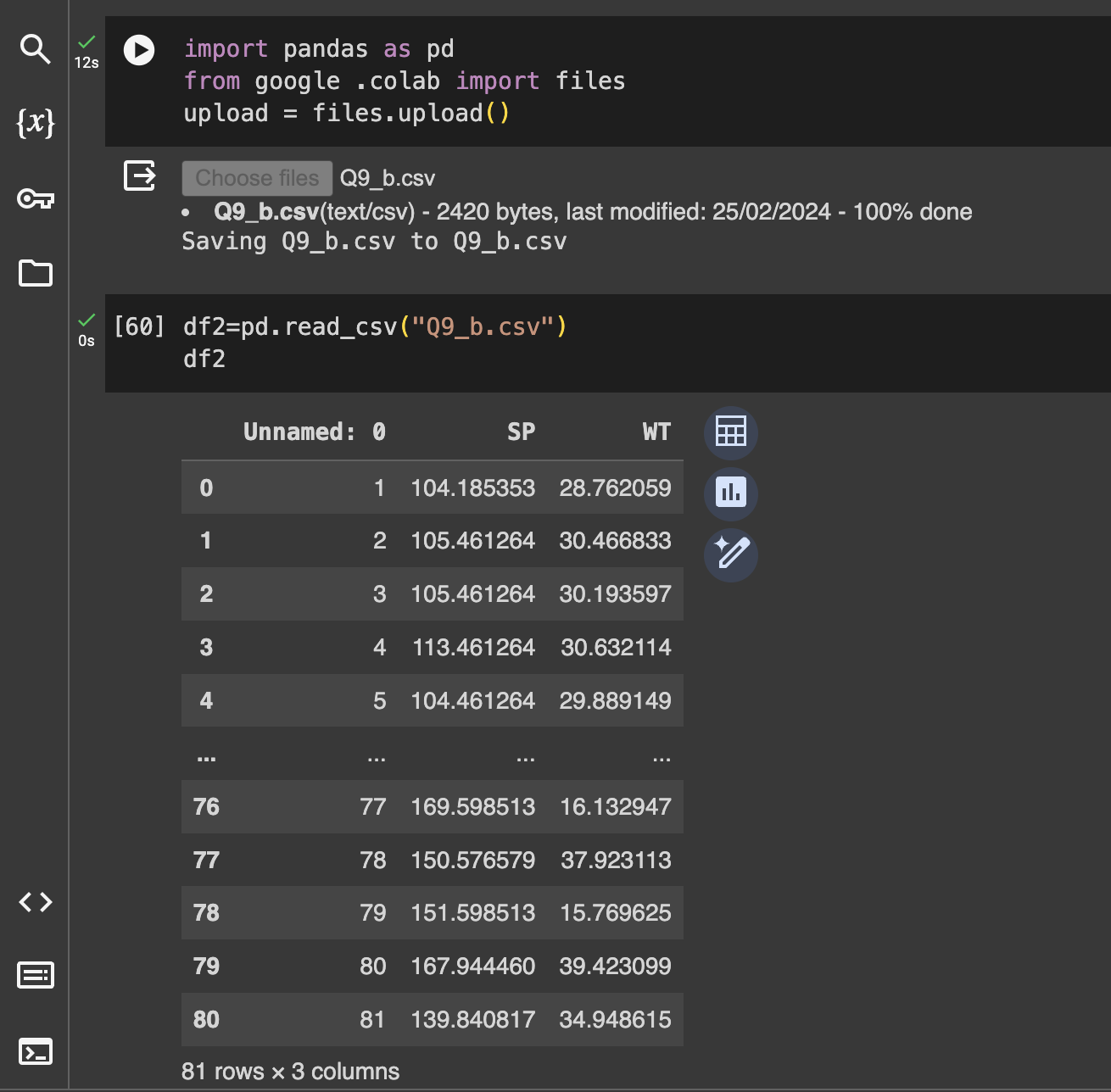
****

****

****

**SP and Weight(WT)**

**Use Q9\_b.csv**

****



****

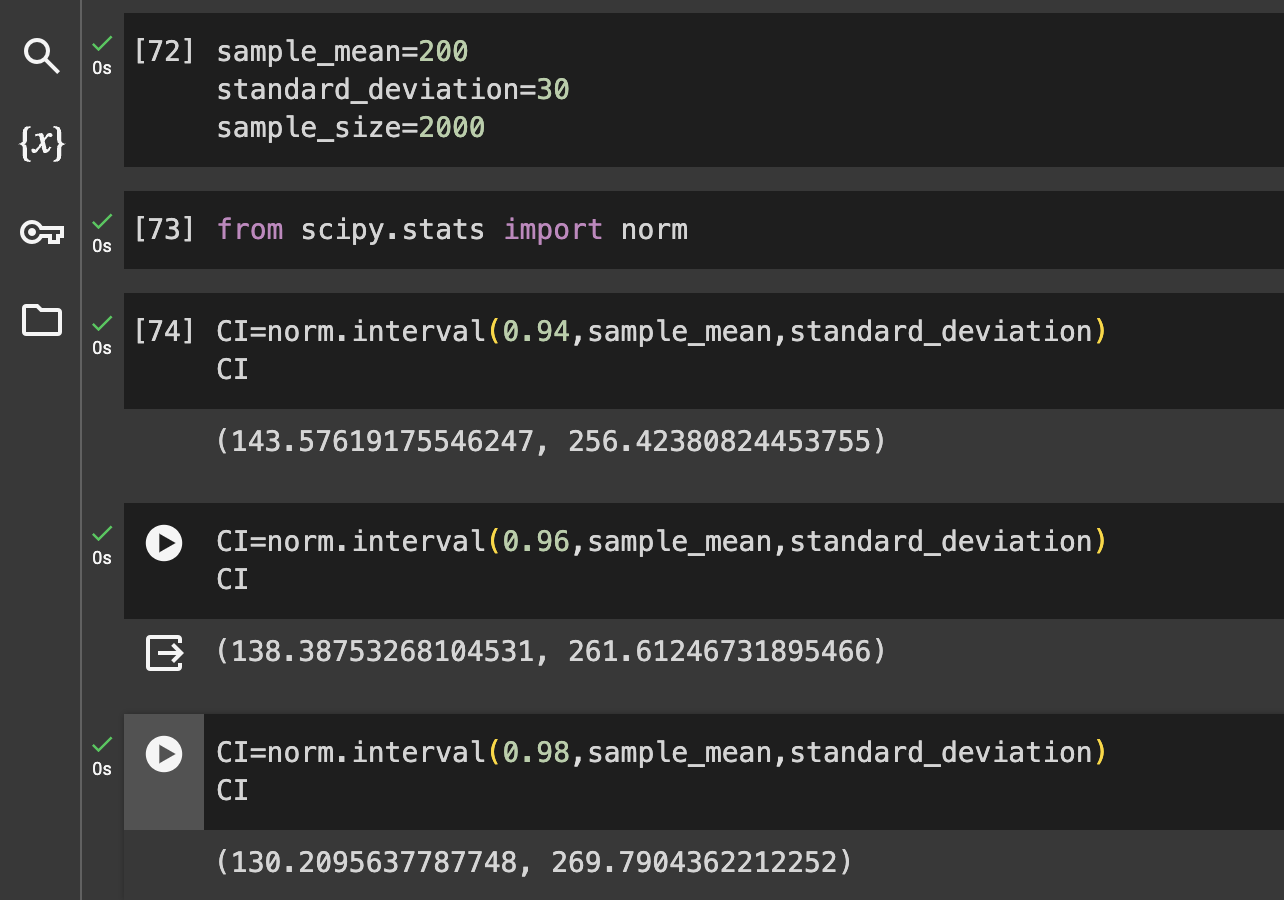
**Q10) Draw inferences about the following boxplot & histogram**

**Answer: The Histogram show that the data is positive skewness(mode<median<mean**



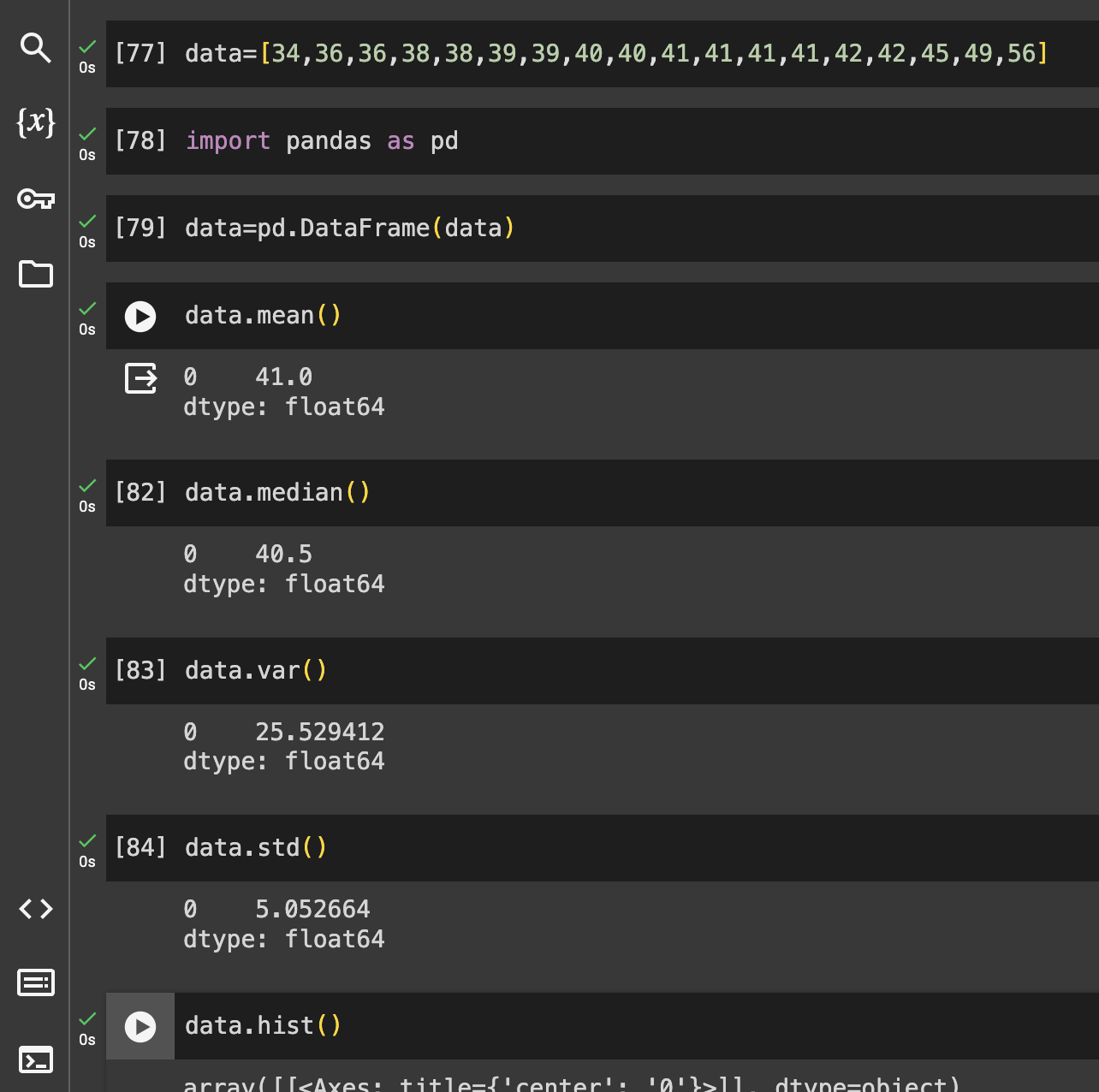


Answer: From the above Box Plot we conclude that the total outliers are on top side

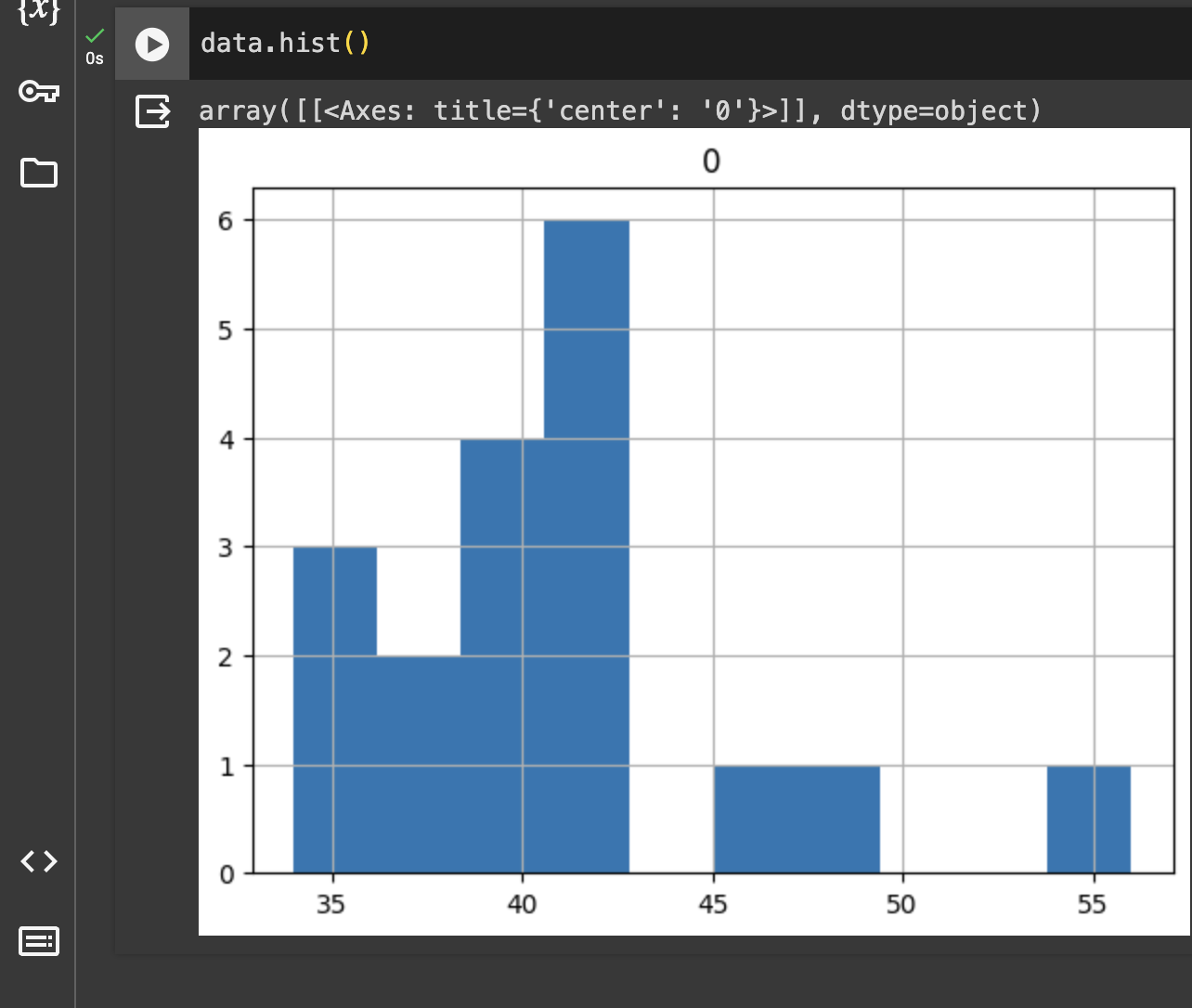
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

**Q12)** Below are the scores obtained by a student in tests

**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

1. Find mean, median, variance, standard deviation.
2. What can we say about the student marks?

Answer:



Q13) What is the nature of skewness when mean, median of data are equal?

Answer: When the mean and median of a dataset are equal,it means that the data is symmetrically distributed.In other words, there is no skewness present in the data.

Q14) What is the nature of skewness when mean > median ?

Answer: When the mean is greater than the median in a dataset, it indicates that the there is positive skewness present.This means that there are some extreme values on the higher end of the range which pull up the mean value, resulting in a longer tail to the right side of the distribution.The median value remains unaffected by the these outliers and stays closer to the bulk of data, giving us a lower representative value.

Q15) What is the nature of skewness when median > mean?

Answer: When the median is greater then the mean in dataset, it indicates that there is negative skewness present.This means that there are some extreme values on the lower end of the data range which pull down the mean value, resulting in a longer tail to the left side of the distribution.

Q16) What does positive kurtosis value indicates for a data ?

Answer: A positive kurtosis value indicates that a dataset has more extreme values

(outliers) than what would be expected from a normal distribution.

Q17) What does negative kurtosis value indicates for a data?

Answer: A negative kurtosis value indicates that a dataset has fewer extreme values (than under normal circumstances) than what would be expected from a normal distribution.

Q18) Answer the below questions using the below boxplot visualization.



What can we say about the distribution of the data?

Answer: Data is not in Symmetric and Negative Skewness

What is nature of skewness of the data?

Answer: Negative Skewness

What will be the IQR of the data (approximately)?

Answer: |QR=Q3-Q1=18-10=18

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

Answer: Boxplot 2 having high speed and whisker length in data is high when compared to Boxplot 1.

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

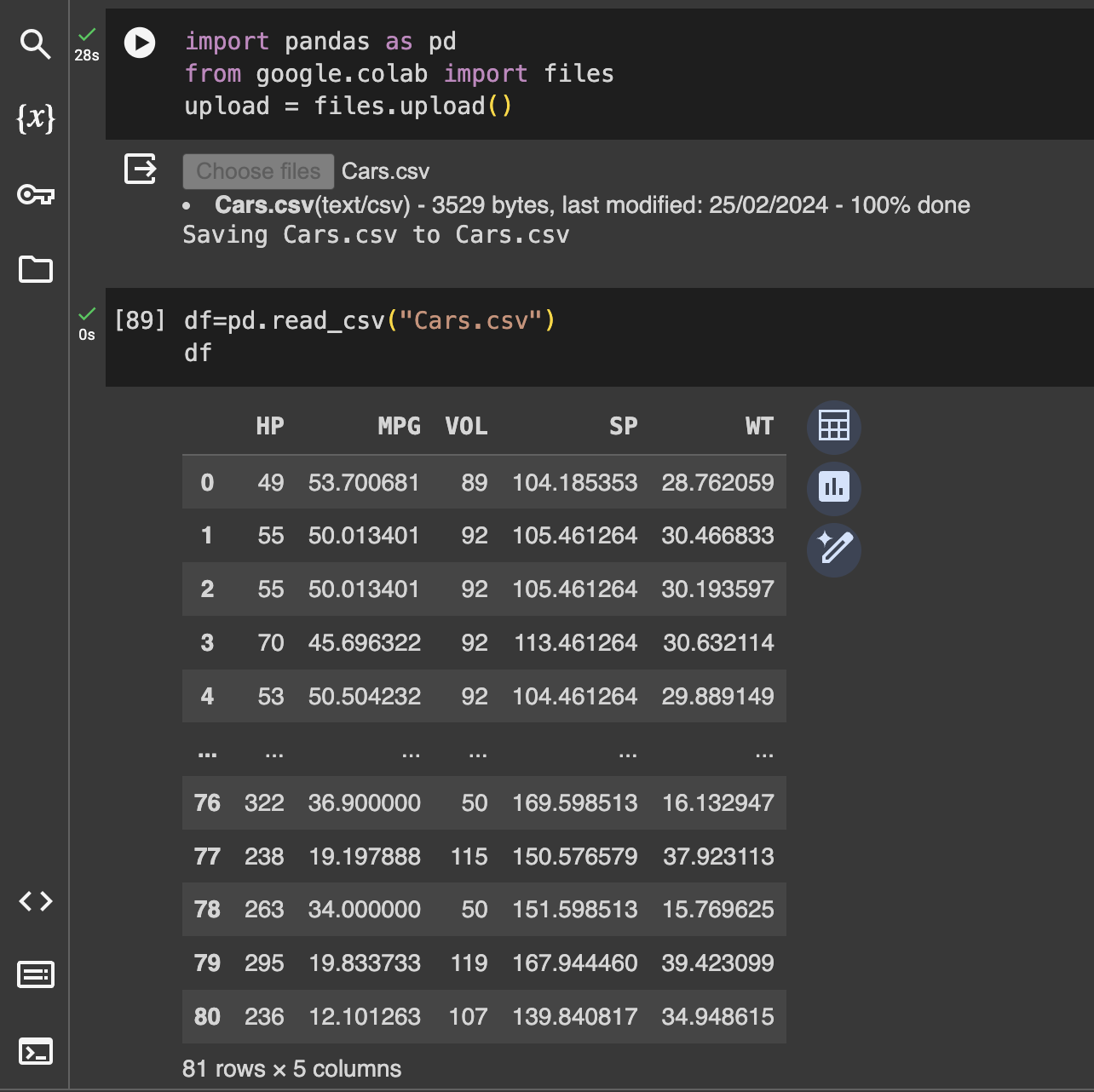
Calculate the probability of MPG of Cars for the below cases.

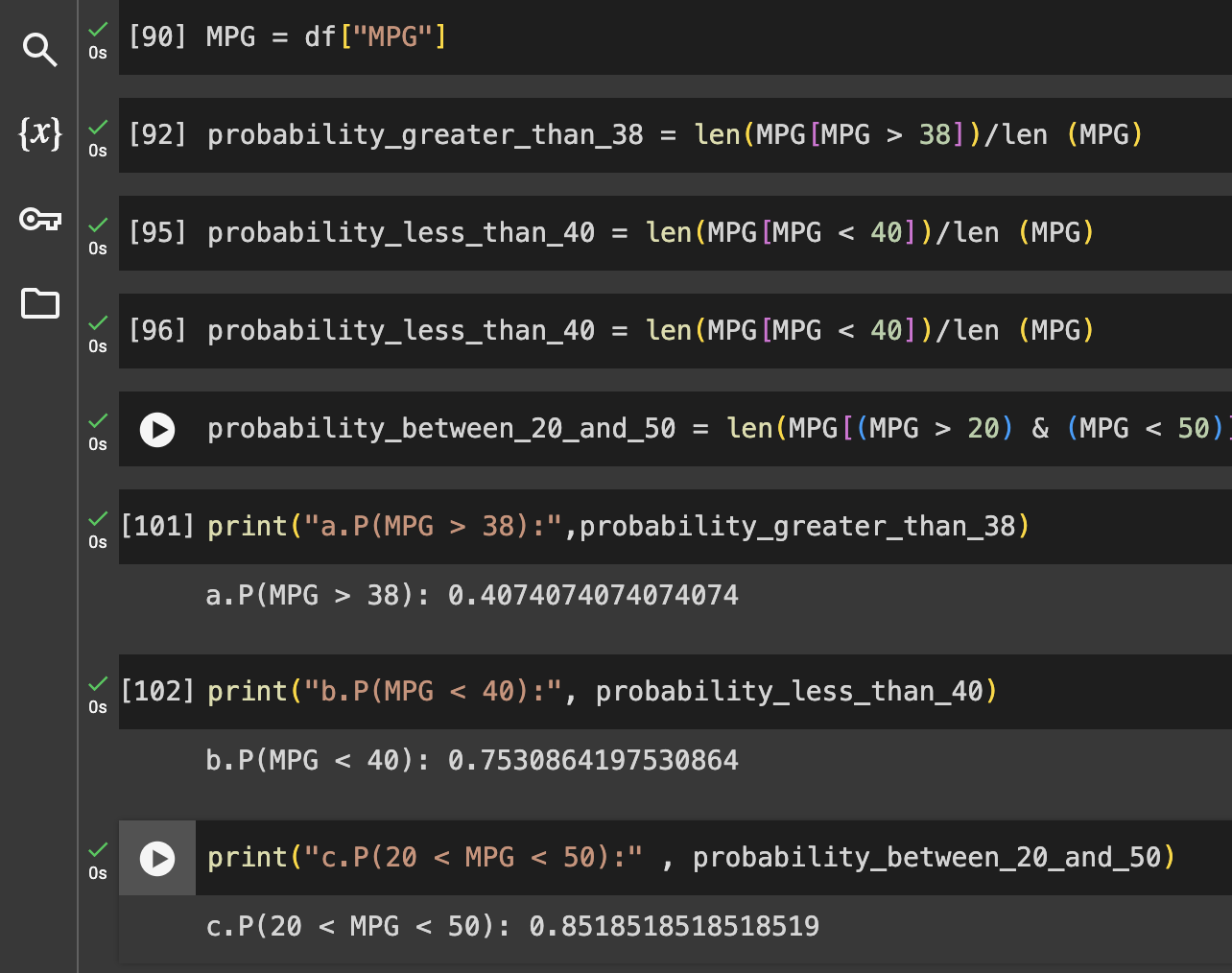
MPG <- Cars$MPG

* 1. P(MPG>38)
  2. P(MPG<40)

c. P (20<MPG<50)

Answer:

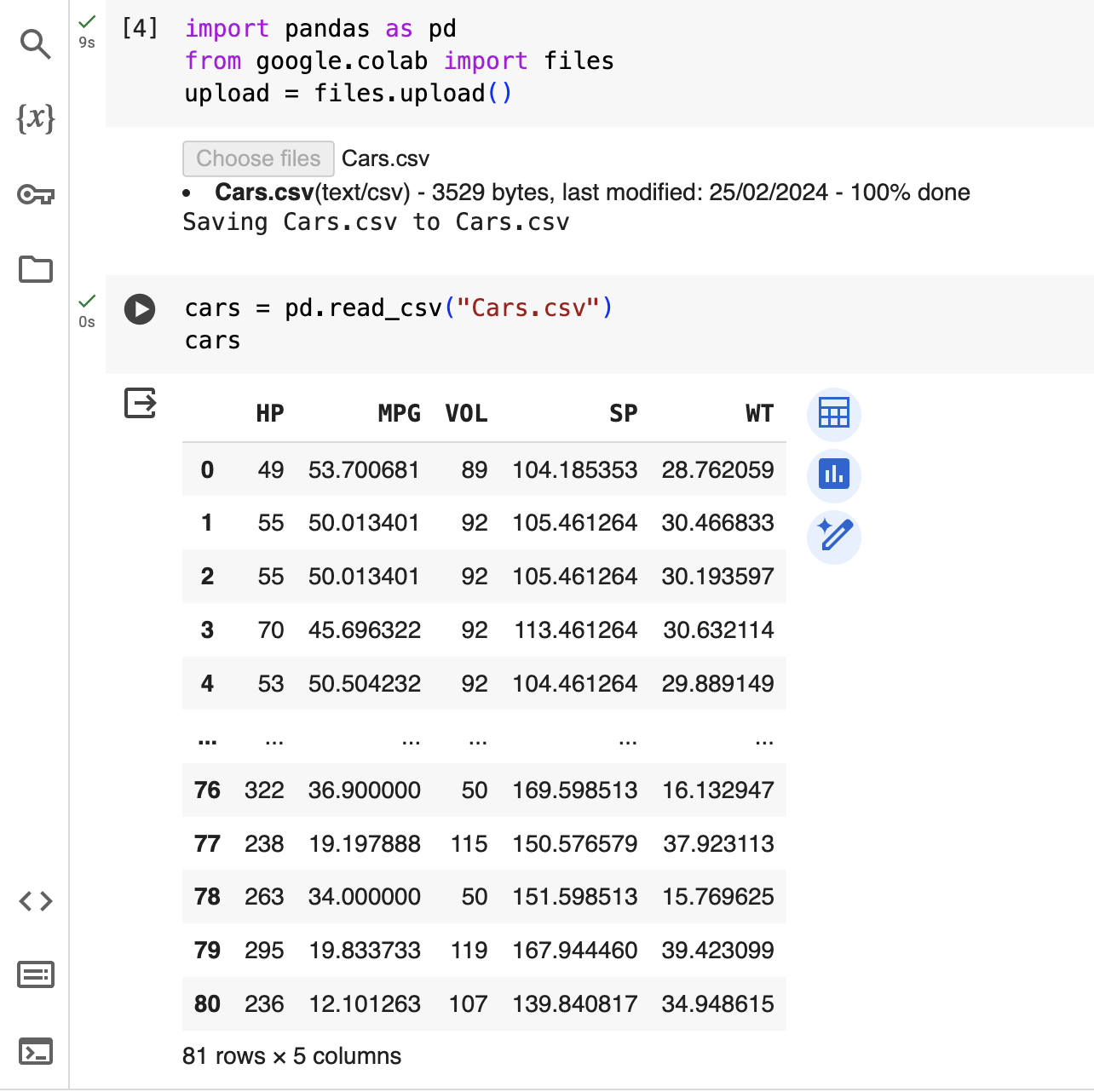


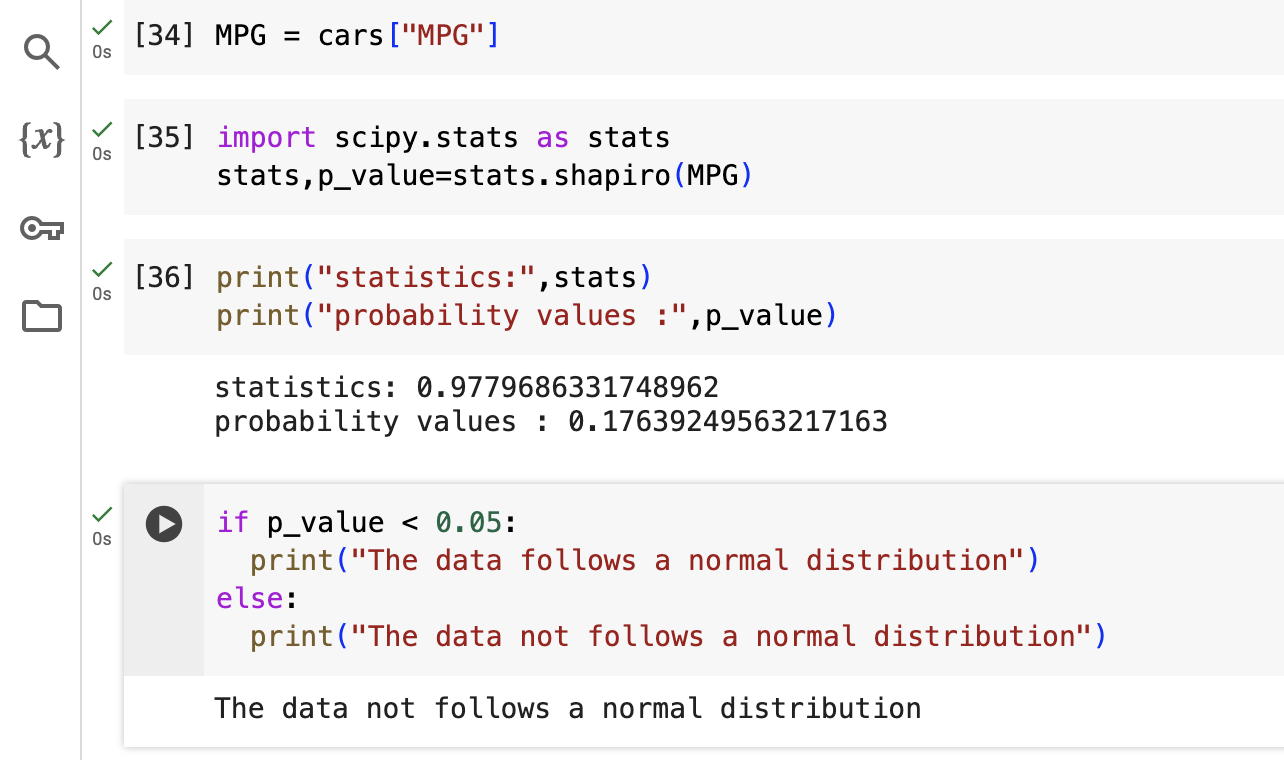


Q 21) Check whether the data follows normal distribution

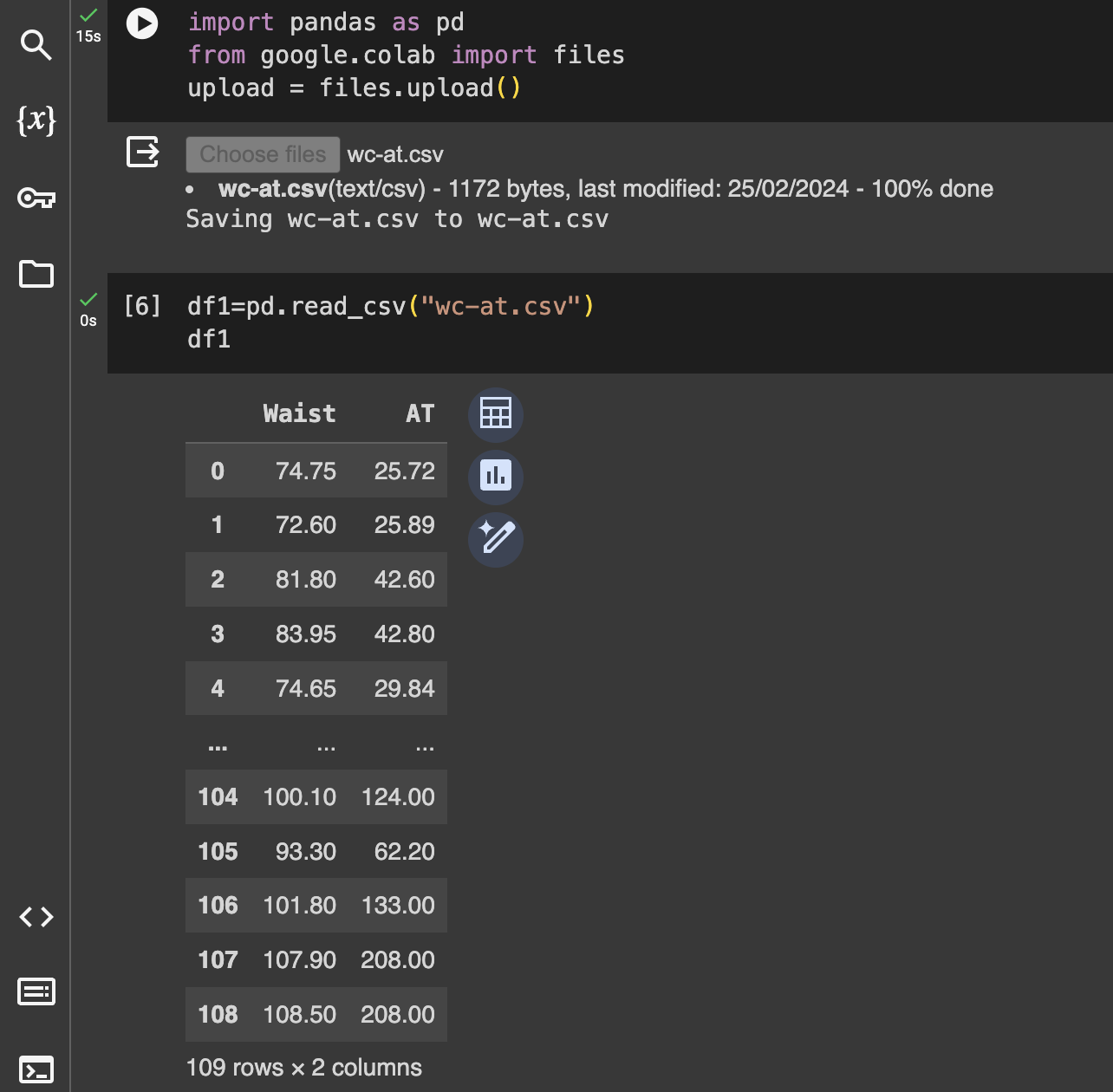
1. Check whether the MPG of Cars follows Normal Distribution

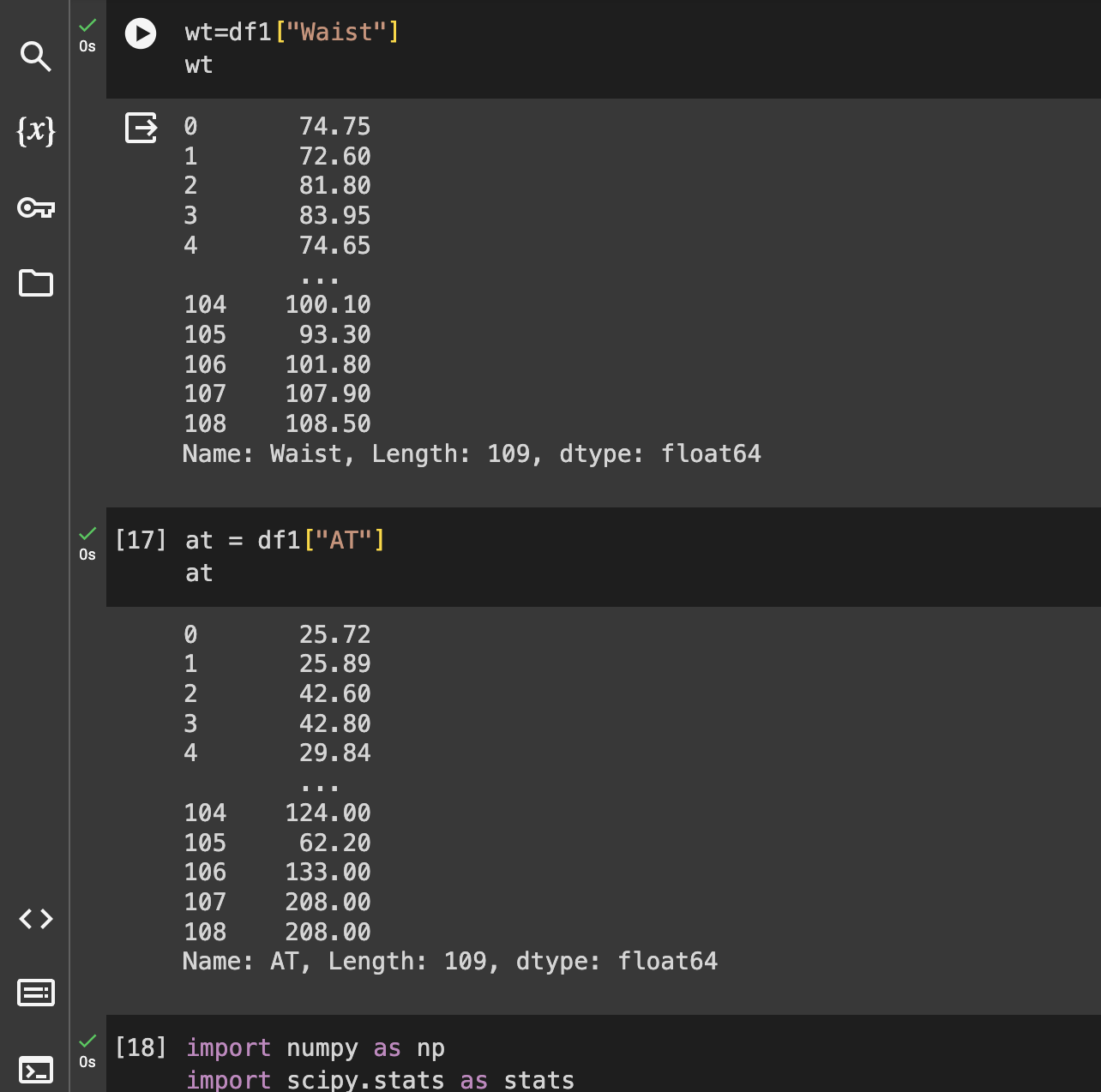
Dataset: Cars.csv

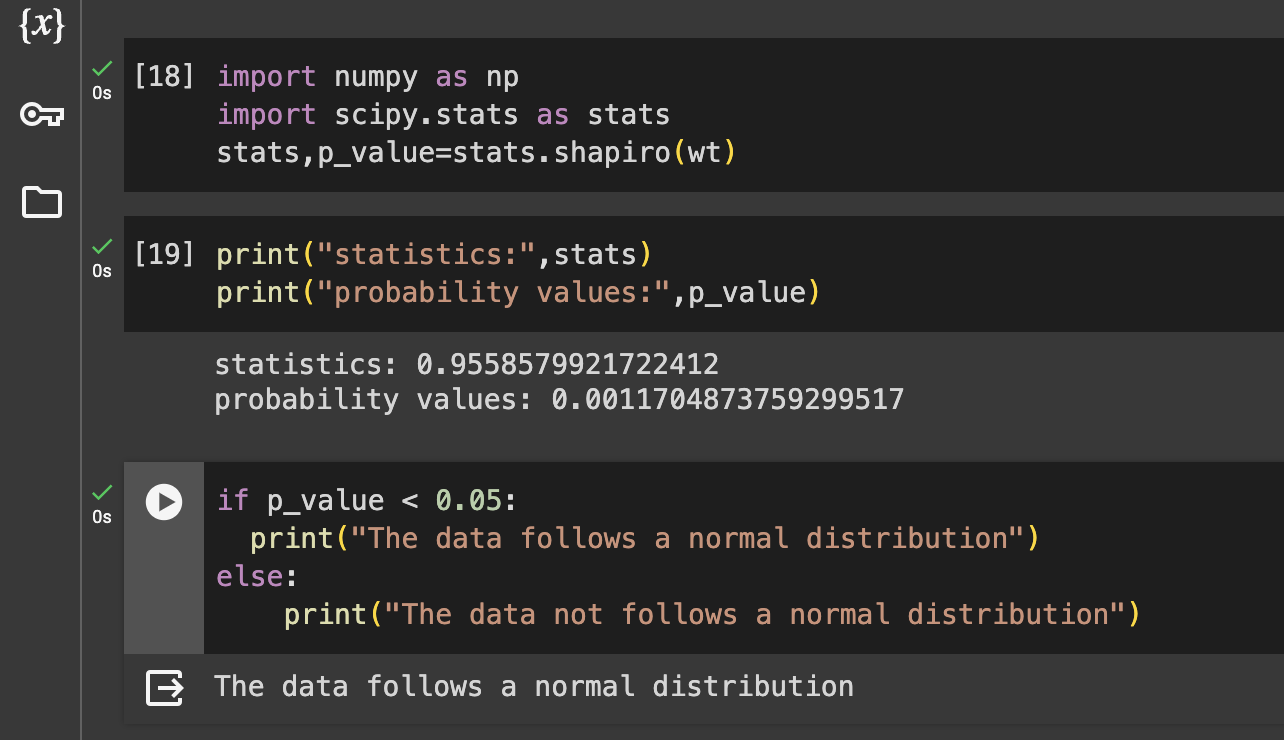


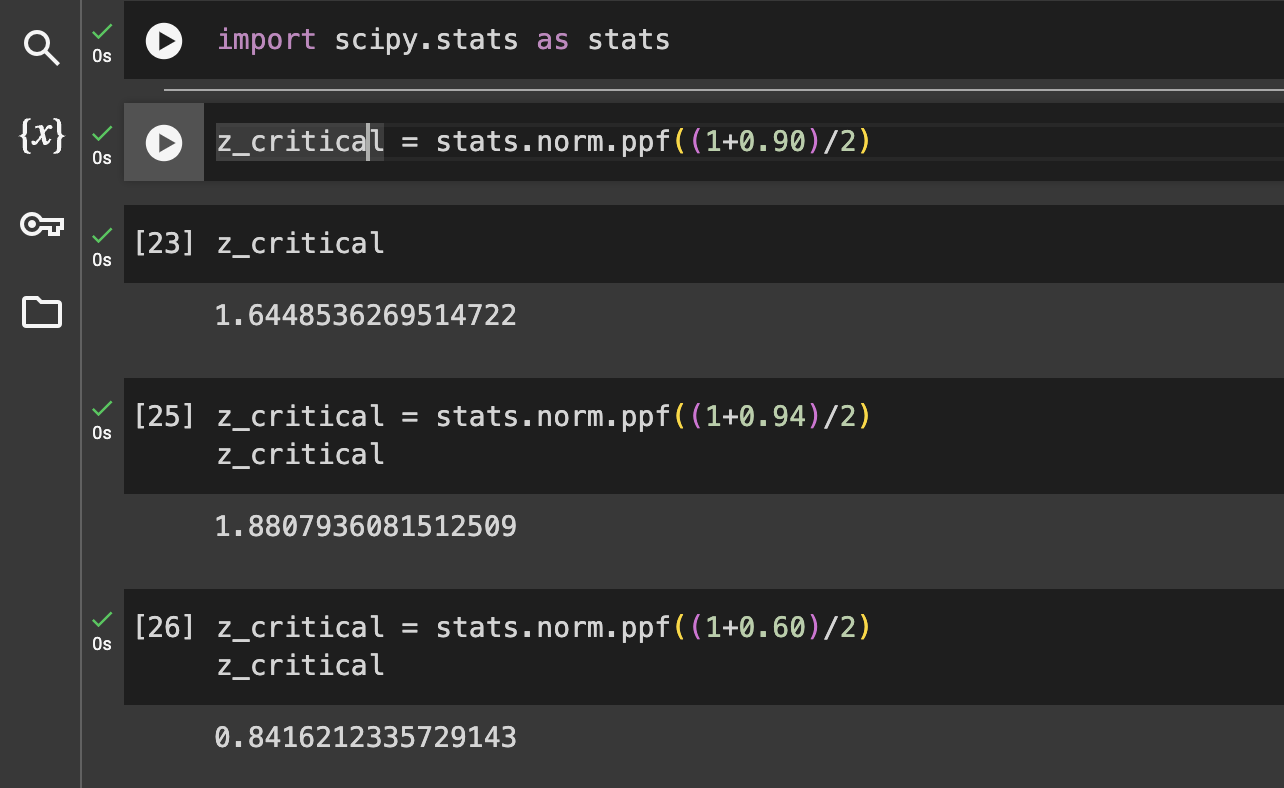


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

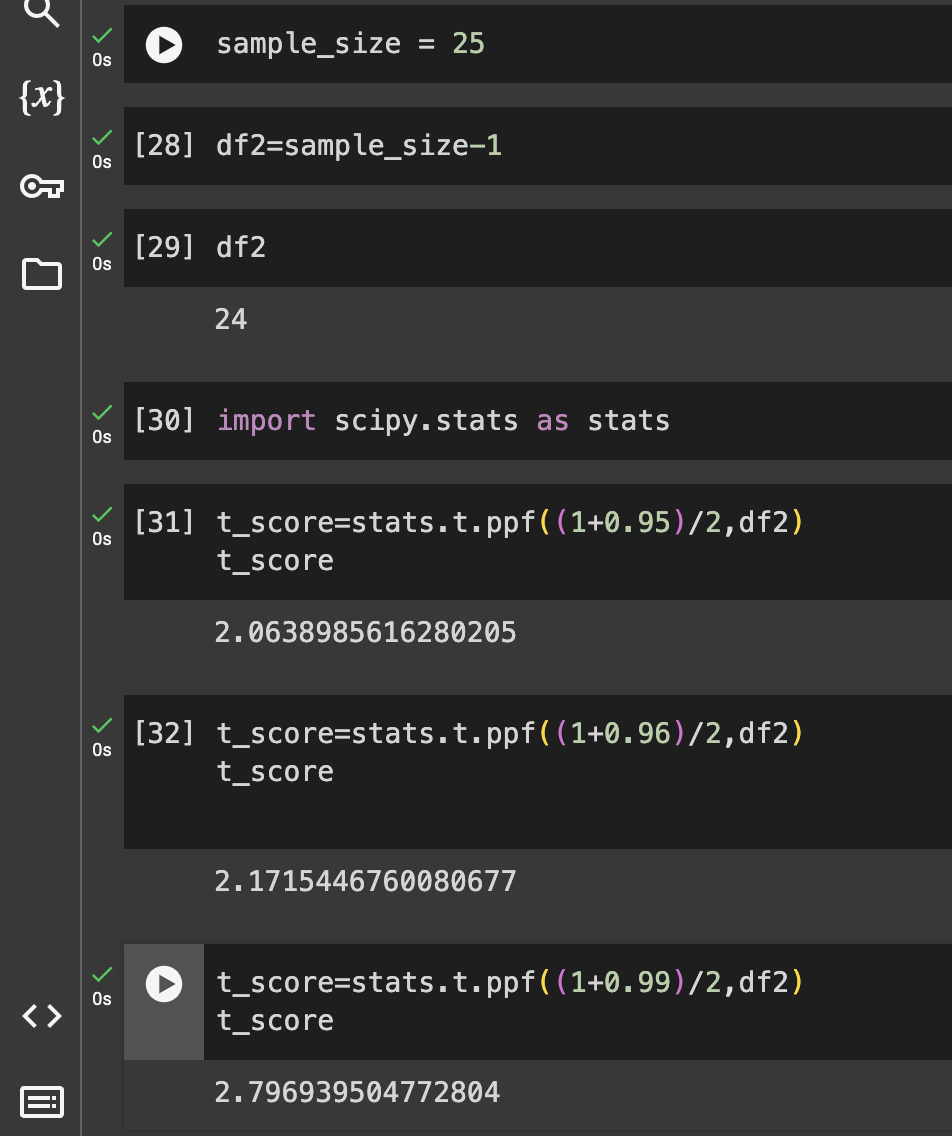
Dataset: wc-at.csv





Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25



Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

