|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Categorical |
| 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 | Categorical |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Categorical |

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 | Ordinal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Ordinal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

**Ans:**

When three coins are tossed the total number of possible combinations are 23 = 8.

These combinations are HHH, HHT, HTH, THH, TTH, THT, HTT, TTT.

The number of combinations which have two heads and one tail are:

HHT, HTH, TTH which makes them 3 in number.

Therefore the [Probability](https://www.cuemath.com/data/probability/) of getting two heads and one tails in the toss of three coins simultaneously is defined as:

P (Two heads and One Tail) = Number of desired outcomes

= 3 = 0.375

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

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

**Ans:**

1. Equal to 1: it is not possible because the sum will always exceed the sum of 1.

Possible outcomes=0

Total outcomes=6^2=36

P=0/36=0.

1. Less than or Equals to 4:

Possible outcomes=6(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)

N(s)Total outcomes=36

P=6/36=0.166.

1. Sum is divisible by 2 and 3:

Possible outcomes=5(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)

N(s)Total outcome=36

P=6/36=0.166.

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?

**Ans:**

Total balls (2+3+2)=7

N(s)Total outcomes= 7C2=7\*6/2\*1=21

Possible Outcomes= 5C2=5\*4/2\*1=10

P=10/21=0.47

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

**Ans:**

Excepted values of candies for the randomly selected child:

=1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

=0.015+0.8+1.95+0.025+0.06+0.24

=3.09.

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**

**Ans:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Title** | **Points** | **Scores** | **Weigh** |
| **Mean** | 3.596563 | 3.21725 | 17.84875 |
| **Median** | 3.695 | 3.325 | 17.71 |
| **Mode** | 3.92 | 3.44 | 17.02 |
| **Variance** | 0.648655 | 1.203509 | 12.36097 |
| **Standard Deviation** | 0.526258 | 0.963048 | 1.758801 |
| **Range** | 2.17 | 3.911 | 8.4 |
| **Highest Value** | 4.93 | 5.424 | 22.9 |
| **Lowest Value** | 2.76 | 1.513 | 14.5 |

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?

**Ans:**

Expected Value= sum(probability\*value)

Sum P(x)\*V(x)

Probability=1/9

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

Sum= (1/9\*108) +(1/9\*110) +(1/9\*123) +(1/9\*134) +(1/9\*135) +(1/9\*`45) +(1/9\*167) +(1/9\*187) +(1/9\*199)

Sum = (1/9) \*(108+ 110+ 123+ 134+ 135+ 145+ 167+ 187+ 199)

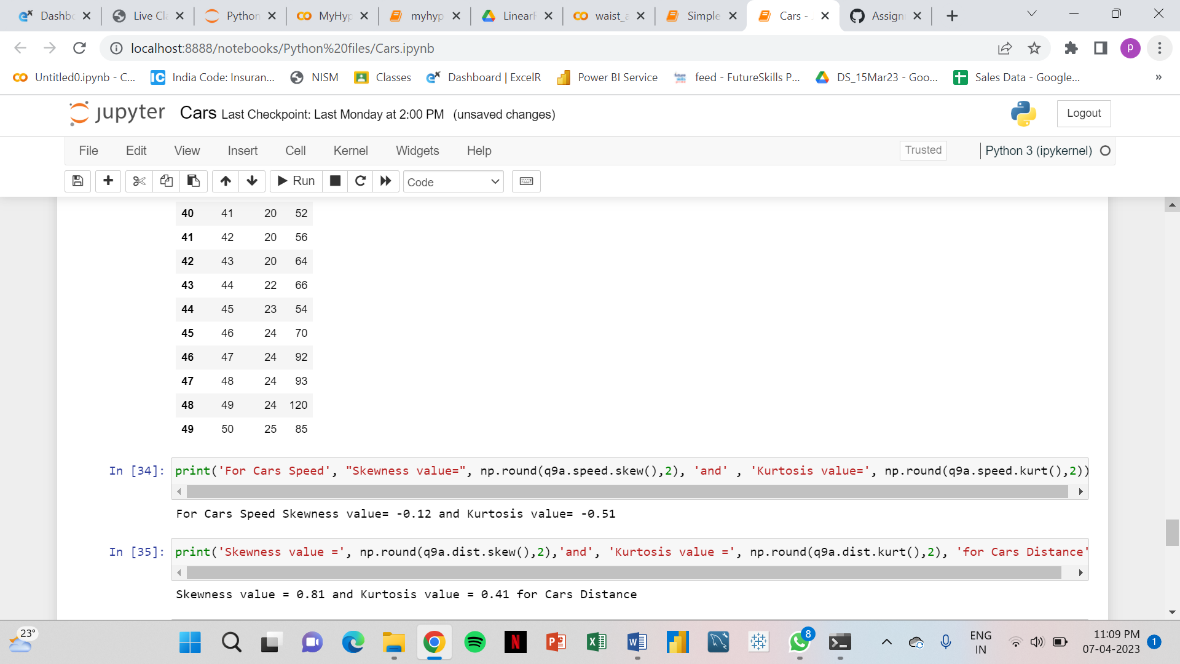
Sum = (1/9) \*(1308)

Sum=145.33

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

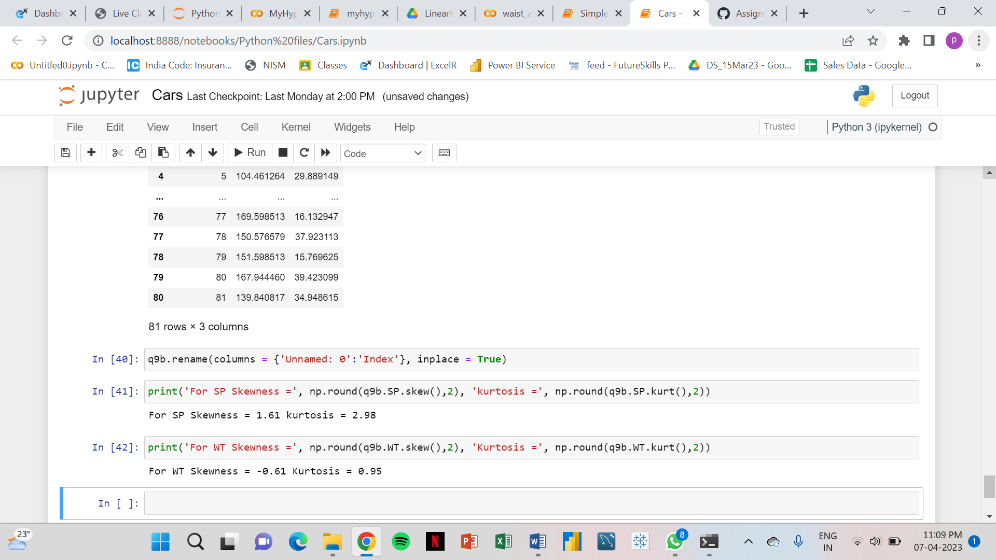
**Cars speed and distance**

**Use Q9\_a.csv**

****

**SP and Weight(WT)**

**Use Q9\_b.csv**

****

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



The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.

The boxplot has outliers on the maximum 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?

**Ans:**

n= 2000

xbar=200

sd=30

Considering a **94%** confidence level, using a calculator, with 200 - 1 = **199 df**, the critical value is **t = 1.8916**, hence:

https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20-%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D%20%3D%20200%20-%201.8916%5Cfrac%7B30%7D%7B%5Csqrt%7B2000%7D%7D%20%3D%20198.73

https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20%2B%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D%20%3D%20200%20%2B%201.8916%5Cfrac%7B30%7D%7B%5Csqrt%7B2000%7D%7D%20%3D%20201.27

The **94%** confidence interval is **(198.73, 201.27).**

Considering a **96%** confidence level, using a calculator, with 200 - 1 = **199 df**, the critical value is **t = 2.0673**, hence:

https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20-%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D%20%3D%20200%20-%202.0673%5Cfrac%7B30%7D%7B%5Csqrt%7B2000%7D%7D%20%3D%20198.61

https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20%2B%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D%20%3D%20200%20%2B%202.0673%5Cfrac%7B30%7D%7B%5Csqrt%7B2000%7D%7D%20%3D%20201.39

The **96%** confidence interval is **(198.61, 201.39).**

Considering a **98%** confidence level, using a calculator, with 200 - 1 = **199 df**, the critical value is **t = 2.3452**, hence:

https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20-%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D%20%3D%20200%20-%202.3452%5Cfrac%7B30%7D%7B%5Csqrt%7B2000%7D%7D%20%3D%20198.43

https://tex.z-dn.net/?f=%5Coverline%7Bx%7D%20%2B%20t%5Cfrac%7Bs%7D%7B%5Csqrt%7Bn%7D%7D%20%3D%20200%20%2B%202.3452%5Cfrac%7B30%7D%7B%5Csqrt%7B2000%7D%7D%20%3D%20201.57

The **98%** confidence interval is **(198.43, 201.57).**

**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?

**Ans:** 1.)

|  |  |
| --- | --- |
| **Title** | **Values** |
| **Mean** | 41 |
| **Median** | 40.5 |
| **Mode** | 41 |
| **Variance** | 24.11111 |
| **Standard Deviation** | 4.910307 |

2.) Maximum Students in the class had obtained 41 marks and the

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

**Ans:**

When a distribution has zero skew, it is symmetrical. Its left and right sides are mirror images.

Normal distributions have zero skew, but they’re not the only distributions with zero skew. Any symmetrical distribution, such as a uniform distribution or some bimodal (two-peak) distributions, will also have zero skew.

The distribution is approximately symmetrical, with the observations distributed similarly on the left and right sides of its peak. Therefore, the distribution has approximately zero skew.

In a distribution with zero skew, the [mean](https://www.scribbr.com/statistics/mean/) and [median](https://www.scribbr.com/statistics/median/) are equal.

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

**Ans:**

Right skew: mean > median

A right-skewed distribution is longer on the right side of its peak than on its left. Right skew is also referred to as positive skew.

You can think of skewness in terms of tails. A tail is a long, tapering end of a distribution. It indicates that there are observations at one of the extreme ends of the distribution, but that they’re relatively infrequent. A right-skewed distribution has a long tail on its right side.

The distribution is right-skewed because it’s longer on the right side of its peak. There is a long tail on the right, meaning that every few decades there is a year when the number of sunspots observed is a lot higher than average.

The [mean](https://www.scribbr.com/statistics/central-tendency/#mean) of a right-skewed distribution is almost always greater than its [median](https://www.scribbr.com/statistics/central-tendency/#median). That’s because extreme values (the values in the tail) affect the mean more than the median.

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

**Ans:**

Left skew: mean < median

A left-skewed distribution is longer on the left side of its peak than on its right. In other words, a left-skewed distribution has a long tail on its left side. Left skew is also referred to as negative skew.

The distribution is left-skewed because it’s longer on the left side of its peak. The long tail on its left represents the small proportion of students who received very low scores.

The [mean](https://www.scribbr.com/statistics/central-tendency/#mean) of a left-skewed distribution is almost always less than its [median](https://www.scribbr.com/statistics/central-tendency/#median).

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

**Ans:**

A positively skewed or right-skewed distribution has a long right tail. It is a sort of distribution where the measures are dispersing, unlike symmetrically distributed data where all measures of the central tendency (mean, median, and mode) equal each other

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

**Ans:**

A distribution with a negative kurtosis value indicates that the distribution has lighter tails than the normal distribution.

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



What can we say about the distribution of the data?

**When the median is closer to the top of the box, and if the whisker is shorter on the upper end of the box, then the distribution is negatively skewed (skewed left).**

What is nature of skewness of the data? **Negatively Skewed**

What will be the IQR of the data (approximately)? **8**

Q19) Comment on the below Boxplot visualizations?



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

**Ans:** First there are no outliers. Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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

Data \_set: Cars.csv

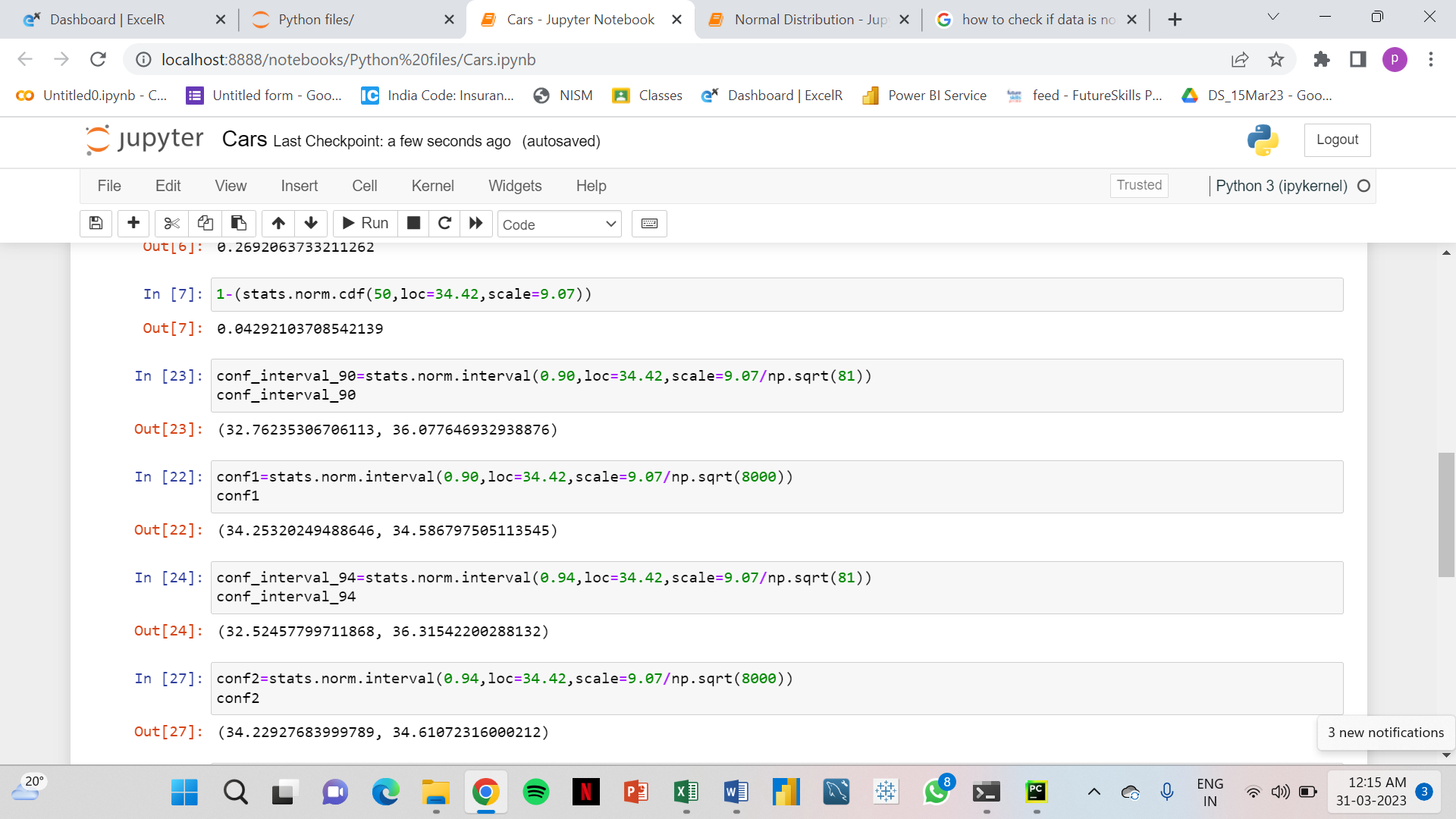
Calculate the probability of MPG of Cars for the below cases.

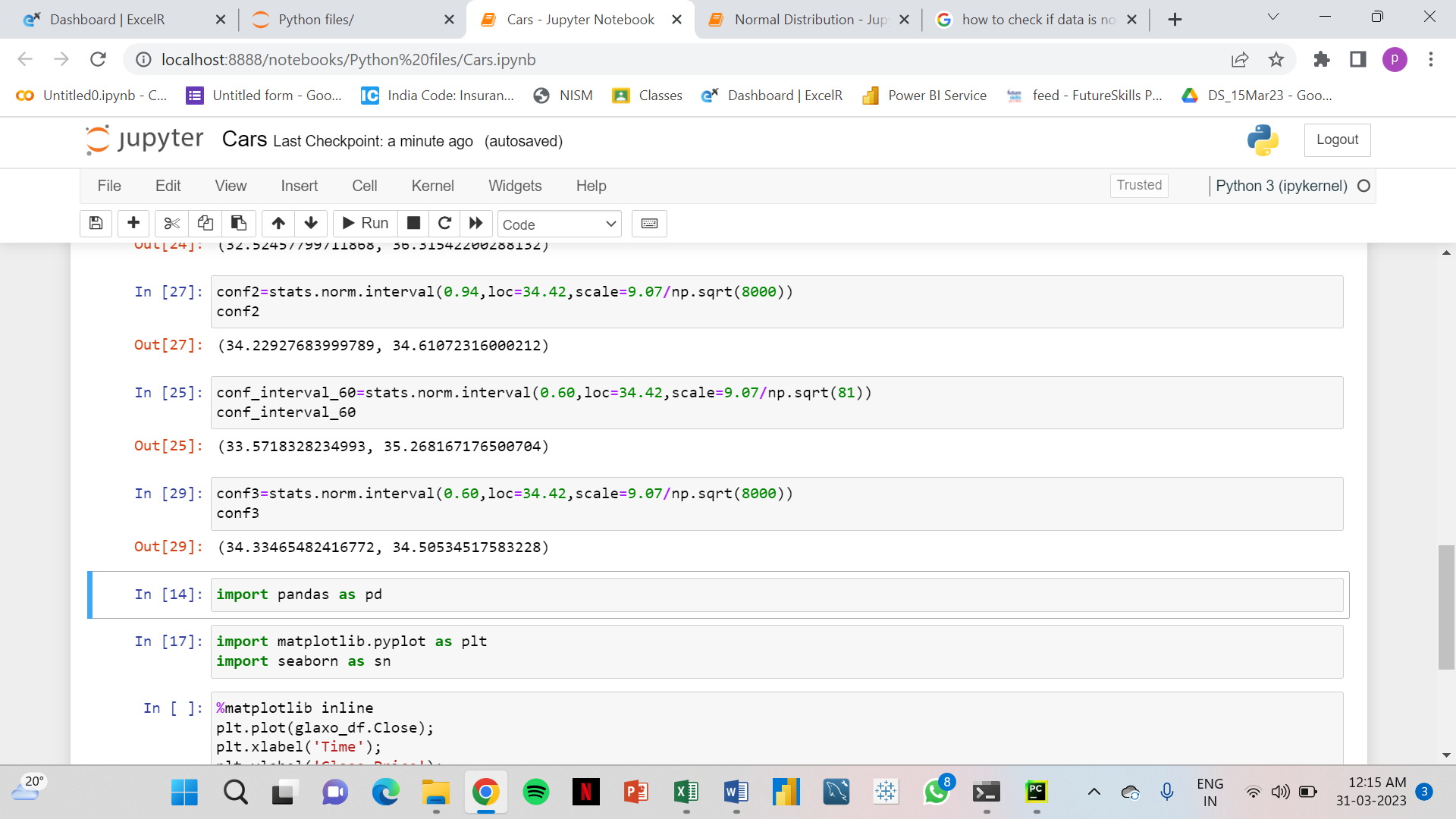
MPG <- Cars$MPG

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

c. P (20<MPG<50)

**Ans:**

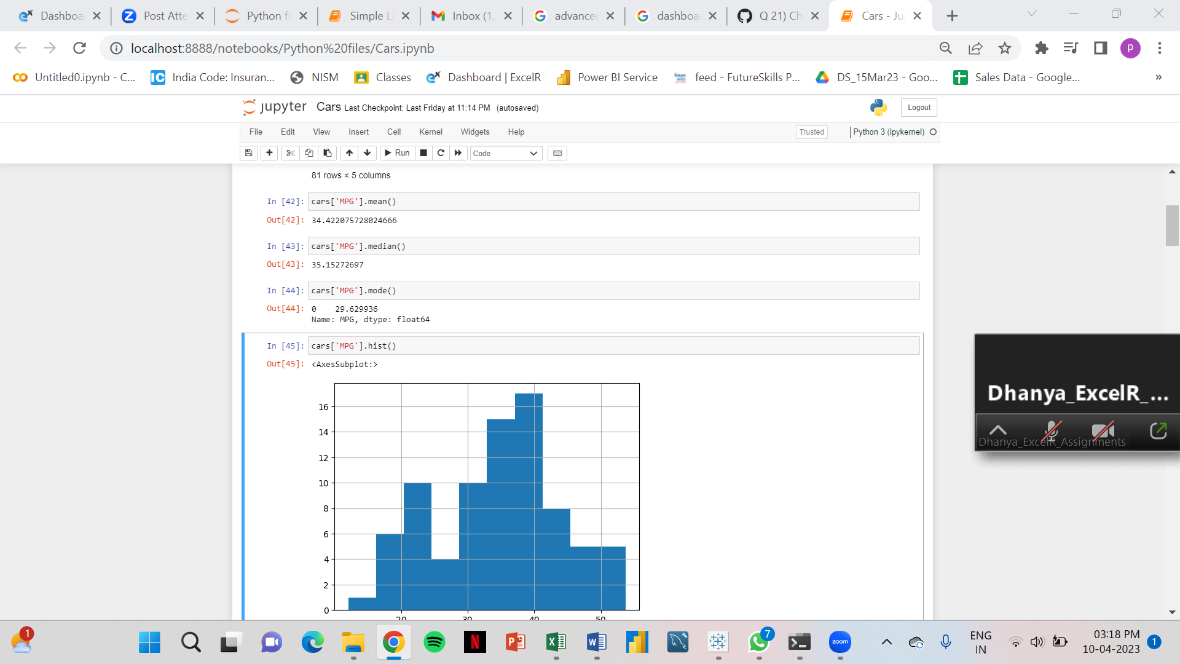
****

****

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

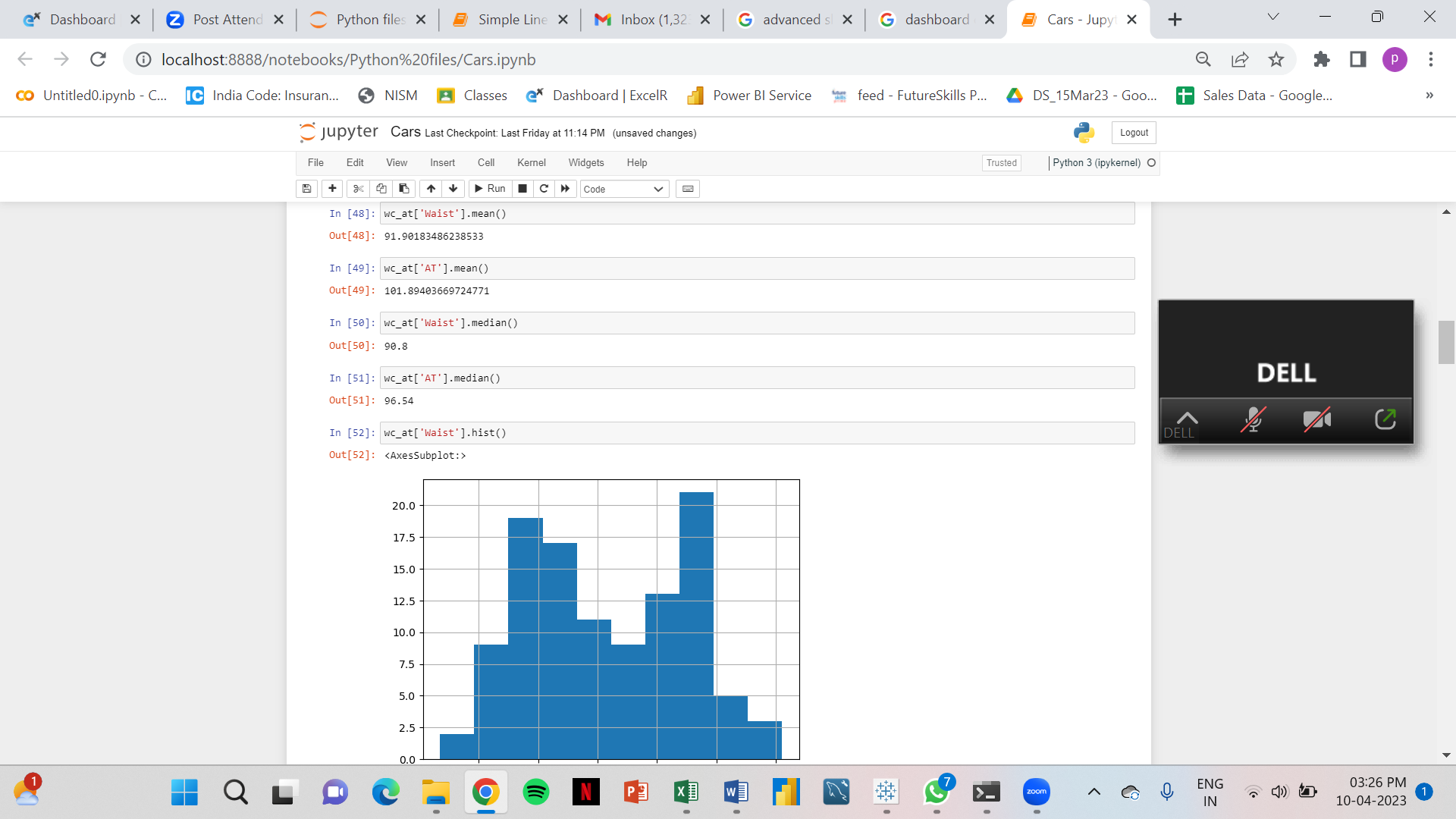
Dataset: Cars.csv



 MPG of cars follows normal distribution

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

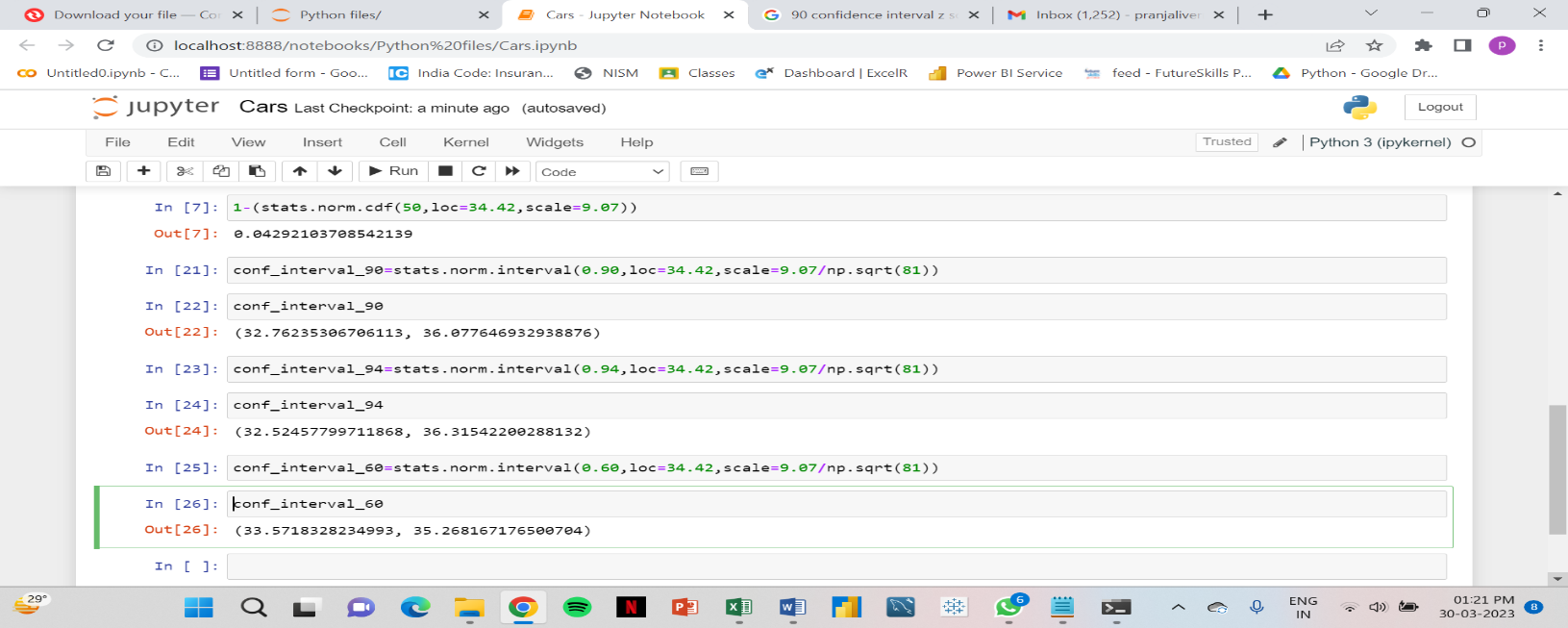
Dataset: wc-at.csv



Adipose Tissue (AT) and Waist does not follow Normal Distribution

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

**Ans:**

**(Cars)**

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

**Ans:**

**t score for 95% confidence interval**

print('T score for 95% Confidence Interval =',np.round(stats.t.ppf(0.025,df=24),4)) T score for 95% Confidence Interval = -2.0639

**t value for 94% confidence interval**

print('T score for 94% Confidence Inteval =',np.round(stats.t.ppf(0.03,df=24),4)) T score for 94% Confidence Inteval = -1.974

**t value for 99% Confidence Interval**

print('T score for 95% Confidence Interval =',np.round(stats.t.ppf(0.005,df=24),4)) T score for 95% Confidence Interval = -2.7969

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 