|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continous |
| Weight of Gold | Continous |
| Distance between two places | Continous |
| Length of a leaf | Continous |
| Dog's weight | Continous |
| Blue Color | Discrete |
| 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 | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ordinal |

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

Ans) Total oberservations=23=8

Chance of to came two heads=(H,T,H),(T,H,H),(H,H,T)

Probability=3/8=0.375

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

Ans)

Total obersavations=62=36

1. Equal to 1 obersavations =0

Probability=0/36=0

1. Less than or equal to 4 obeservations=(1,1),(1,2),(1,3),(2,1),(2,2),(3,1).

Probability=6/36=1/6=0.16666 (or) 0.17.

1. Divisible by 2 and 3 obeservations=(1,5),(2,4),(3,3),(4,2),(5,1),(6,6).

Probability=6/36=1/6=0.16666 (or) 0.17.

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) Probality = 10/21

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

= (3.09)

= 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.

Q7 )

|  |
| --- |
| Points Score Weight |
| Mean 3.596563 3.21725 17.84875 |
| Mode 3.07-3.92 3.44 17.02-18.90 |
| Median 3.695 3.25 17.71 |
| Variance 0.2858814 0.957379 1.786943 |
| SD 0.5346787 0.9784574 1.786943 |
| Range 2.76-4.93 1.513-5.425 14.5-22.9 |



Q8) Calculate Expected Value for the problem below

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

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

Ans8) Expected value is = (Probability \*value)

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

= 145.33

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

**Cars speed and distance**



9a.Ans)

|  |
| --- |
| speed distance |
| Skewness -0.1139548 0.7824835 |
| Kurtosis 2.422853 3.248019 |



**SP and Weight(WT)**



****



**9b.Ans**

|  |
| --- |
| **WT SP** |
| **Skewness -0.6033099 1.581454** |
| Kurtosis 3.819466 5.723521 |

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



Ans) The histograms peak has right skew and tail is on right.

Mean>Median. We have outlliers on the higher side

Ans) 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 ?

Q11Ans)

Mean x=200

Standerd diviviation s=30

No. of oberservation n=2000

Df=(n-1)=2000-1

**Confidence interval=94%**

Stats.t.ppf(ci,df) # we have get in t value

Stats.t.ppf(0.97,1999)=t=1881

Confidence interval

Max interval=200+1.88×30/=201.2611

Min interval=200-1.88198.7388

**Confidence interval 98%:**

Stats.t.ppf(0.99,1999)=t=2.3282

Confidence interval

Max interval=200+2.32×30/=201.5563

Min interval=200-2.32198.443

**Confidence interval 96%:**

Stats.t.ppf(0.98,1999)=t=2.055

Confidence interval

Max interval=200+2.055×30/=201.3785

Min interval=200-2.055198.621

**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?

Ans12.1) mean=41,median=40.5,variance=25.52941,sd=5.052664.

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

Ans13) If mean =median , Skewness=0, Perfectly Symetrical (or) Normal Distribution.

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

Ans14) If Mean > Median,Skewness value is positive then it is called Rightly skeweed.

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

Ans15) If Mean < Median , Skewness value is negative then it is called Left skeweed.

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

Ans16) Sharper peaks & Fatter tails.

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

Ans17) Flater peak & Thinner tails

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



What can we say about the distribution of the data?

Ans) The above Boxplot is not normally distribued the median is towards the higher value

What is nature of skewness of the data?

Ans) The data is skewed towards left.The whisker range of minimum value is greater than maximum.

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

Ans) the Inter quantile Range =Q3 Upper quartile-Q1 Lower Quartile.

= 18-10

= 8

Q19) Comment on the below Boxplot visualizations?



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

Q19Ans) First there are no outliers. Second both the boxplot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zerontonnonnskewness 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)

20QAns) the mean value of MPG=µ=34.42

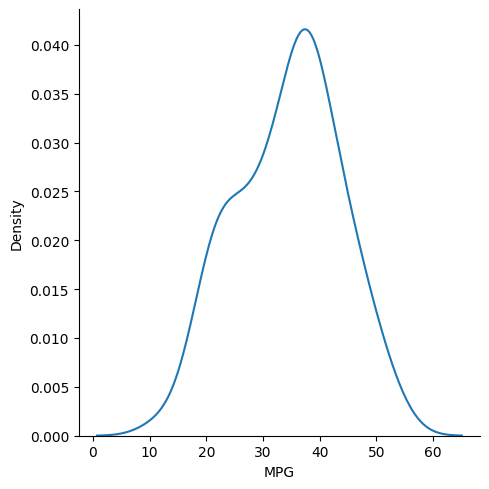
The standerd diviation value of MPG=¬=34.4220

1. P(MPG>38)=0.4585
2. P(MPG<40)=0.5643
3. P(20<MPG<50)=0.3369

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



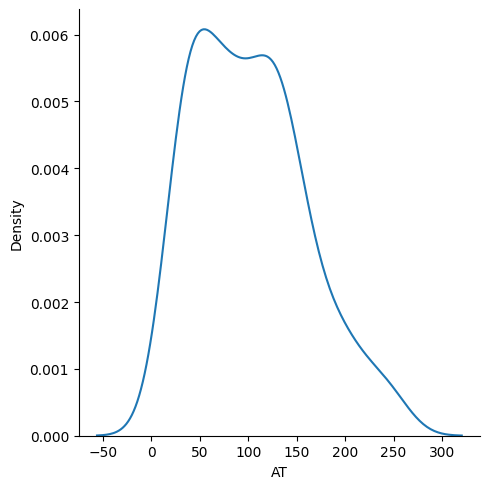
This is not normal distribution why because mpg column skew ness value =! 0

Then it left skewed distribution

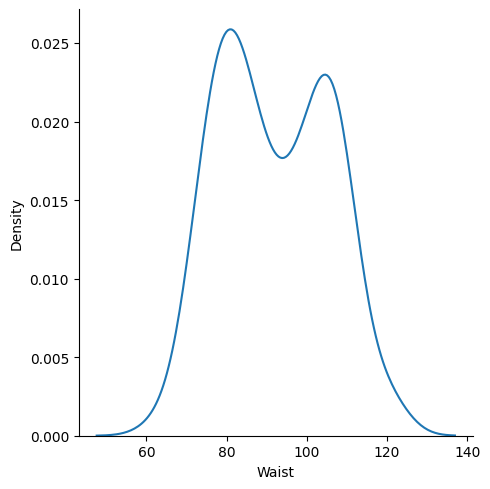
The skewness value is negative ( -0.1779)

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

Dataset: wc-at.csv



This is not normal distribution . this one is right skewed why because this skewness value is (0.58486)



This is also not a normal distribution . this one is a right skewed distribution

The skewness value is =(0.1340

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

Ans) To import scipy library

Command of z score is=stats.norm.ppf(ci+α/2)

90%

Stats.norm.ppf(90+10/2)=stats.norm ppf(0.95)=1.64485 (or) 1.65

94%

Stats.norm.ppf(94+6/2)=stats.norm.ppf(0.97)=1.88079 (or) 1.89

60%

Stats.norm.ppf(60+40/2)=stats.norm.ppf(0.80)=0.841621 (or) 0.85

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

Ans) to import scipy libraries

And numpy libraries

Pt(tscore,df)

Df=sample size -1=25-1=24

Command of t score:

St.t.ppf(confidence interval,df)

Confidence interval=95%

St.t.ppf(0.975,24)

t-score value =2.0638

confidence interval=96%

st.t.ppf(0.98,24)

t-score value=2.171544

confidence interval=99%

st.t.ppf(0.995,24)

t-score value=2.79693

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

Ans)

Sample size = n=18

Sample mean=x=260

Pop mean=270

Sample standered deviation=s=90

Degree of freedom (df)=n-1=17

t-score value= (sample mean-pop mean)/(std deviation/(n \*\* 0.5)

=(260-270)/(90/(0.5 \*18))

= -0.4714

Probability= t.cdf(t-score, df)

The probability value is 0.32167

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -215.9815 21.7963 -9.909 <2e-16 \*\*\*

Waist 3.4589 0.2347 14.740 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 33.06 on 107 degrees of freedom

Multiple R-squared: 0.67, Adjusted R-squared: 0.667

F-statistic: 217.3 on 1 and 107 DF, p-value: < 2.2e-16

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -533.34 42.86 -12.45 <2e-16 \*\*\*

sqrt(Waist) 66.44 4.47 14.86 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 32.88 on 107 degrees of freedom

Multiple R-squared: 0.6737, Adjusted R-squared: 0.6706

F-statistic: 220.9 on 1 and 107 DF, p-value: < 2.2e-16

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1328.34 95.92 -13.85 <2e-16 \*\*\*

log(Waist) 317.14 21.26 14.92 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 32.8 on 107 degrees of freedom

Multiple R-squared: 0.6753, Adjusted R-squared: 0.6723

F-statistic: 222.6 on 1 and 107 DF, p-value: < 2.2e-16

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -7.8240714 1.4729616 -5.312 6.03e-07 \*\*\*

Waist 0.2288644 0.0322008 7.107 1.43e-10 \*\*\*

I(Waist \* Waist) -0.0010163 0.0001731 -5.871 5.03e-08 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.308 on 106 degrees of freedom

Multiple R-squared: 0.779, Adjusted R-squared: 0.7748

F-statistic: 186.8 on 2 and 106 DF, p-value: < 2.2e-16

Machine Learning PAP with Flask

Spam Identifier for Messages

Type in your Message Here

Predict

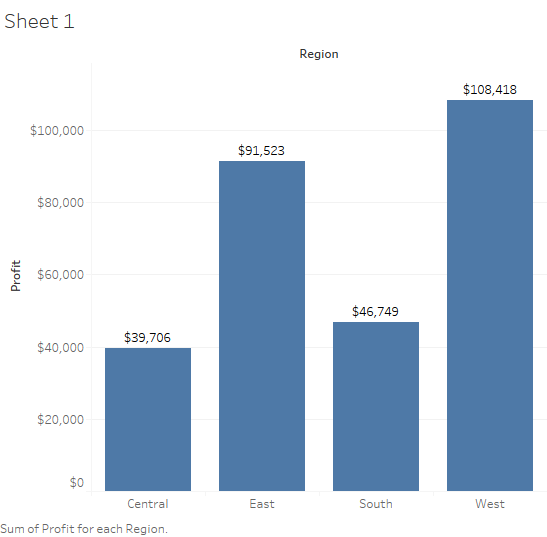


Tableau Desktop -14 days

Tableau Public- Free (no time limit)

Tableau students – (Free 1 year)

Basic steps- In real life

1. Connect to your dataset (vcsv, excel, ms access etc) -Job done (sample superstore)
2. Look at the data source( yes , looked at the data source, what variables do I have , do secondary research (what is ship mode , what is sales , what categories are there, what data type is my variable)
3. Identify your dimensions and measures (dimesnions-Rowid, column id, customer id)
4. Decide on what charts to build ( we are free to perform any chart analysis) (this could be tree maps, text tables or, heat maps depending on requirement)
5. Make interactive dashboards (do an exploratory analysis, which is playing around with dashboards for sake of learning)
6. Write a story on dashboards

Discrete vs continuous

In tableau , measures and dimensions

Measure , any data that can be represented in the form of numbers and which makes sense

Dimension , any data that can be represented in the form of numbers and which doesnot make sense

Ed, rowed, column id,

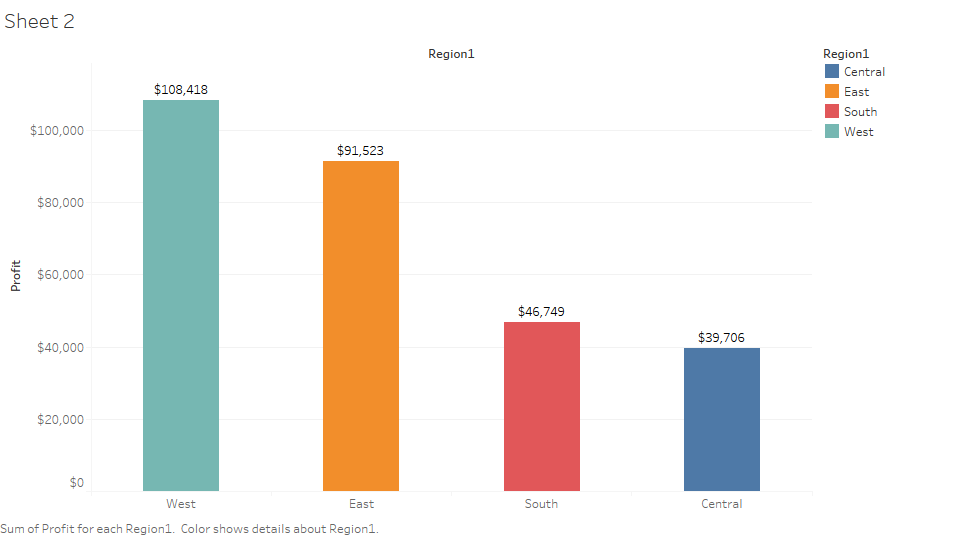
When you want to add additional measures(loss)

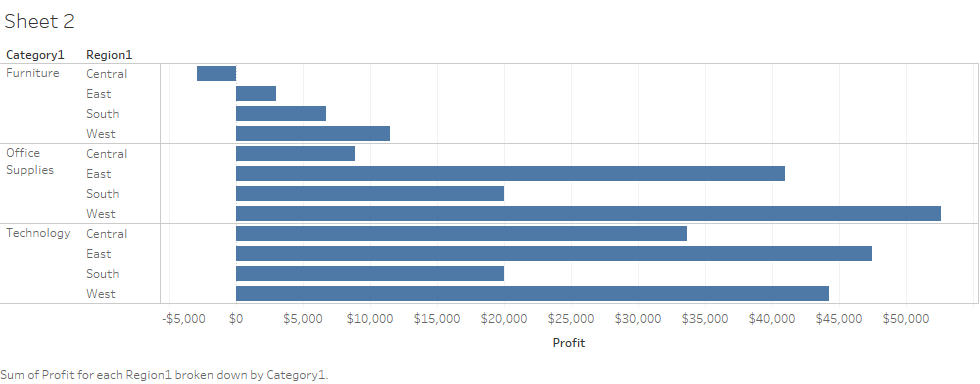
If (profits>0) then “profit”

Else

Then “loss

end





24 type of charts in tableau

Apart form this custom made charts are available

Waterfall chart

Doughnut chart

Funnel chart

Pareto chart

Example that we will discusss

Then we will also discuss , paramters, groups, sets, LOD, connection with R, hierarchies, custom made charts to draw polygon maps, etc , calculated fields,

Filters, pages,

Top 10 customers

Bottom 10 customers