55:10

Purchase Data Study

* Target *

* Whenever we want to change object datatype into numeric data we technique which "Label Encoding"

* if the df["Colname"].astype(int) will not work to convert the data into numeric then we can use the function call ~~Pdnumeric~~. Pd.to_numeric and errors = "coerce"

for example

df ["Stay_in_Current_City_City_years"] = Pd.to_numeric (df ["Coln"], errors = 'Coerce')

* Notes:-
Lets suppose in Categorical data there is 3 gender M, F, T if we are changing it into the numerical data type our machine will asign the Number based on Adphabetical order followed by indexing.

So F → 0
M M → 1
F T — 2
T

* For changing data type able to numeric. we import the library.

→ Label-Encoding

which is:-

from. Sklearn. preprocessing import Label Encoder.

11:29 ;))

* fit transformed (df[i]) changing data object to Numeric

* Statistics part ~~in EDA~~

→ data collection

→ preparing sample

→ null and alternate hypo

→ applying the apis test.

→ if P_value > 0.05 --- fail to reject the null hypothesis

~~Question~~ Ans

It was observed that average purchase made by the men of age 18-25 was 10000. is it still the same. [ ] check through hypothesis

Note :-

① what ever question says it directly goes for null hypothesis.

Null : mean will be equal to 10000 (same question)

All : Mean ≠ 10K (opposit to question)

Question

* Can we use t-test if we have two sample

⟹ T-test is only to check the real problem solving.

If in one popopulation standard is not given which practically not possible and sample size is less than ≤ 30 we will prefer t-test.