upGrad

#LifeKoKaroLift

Data Science Program

upGrad
Course : Data Science
Lecture On : EDA Coding
Instructor : Sumit Shukla

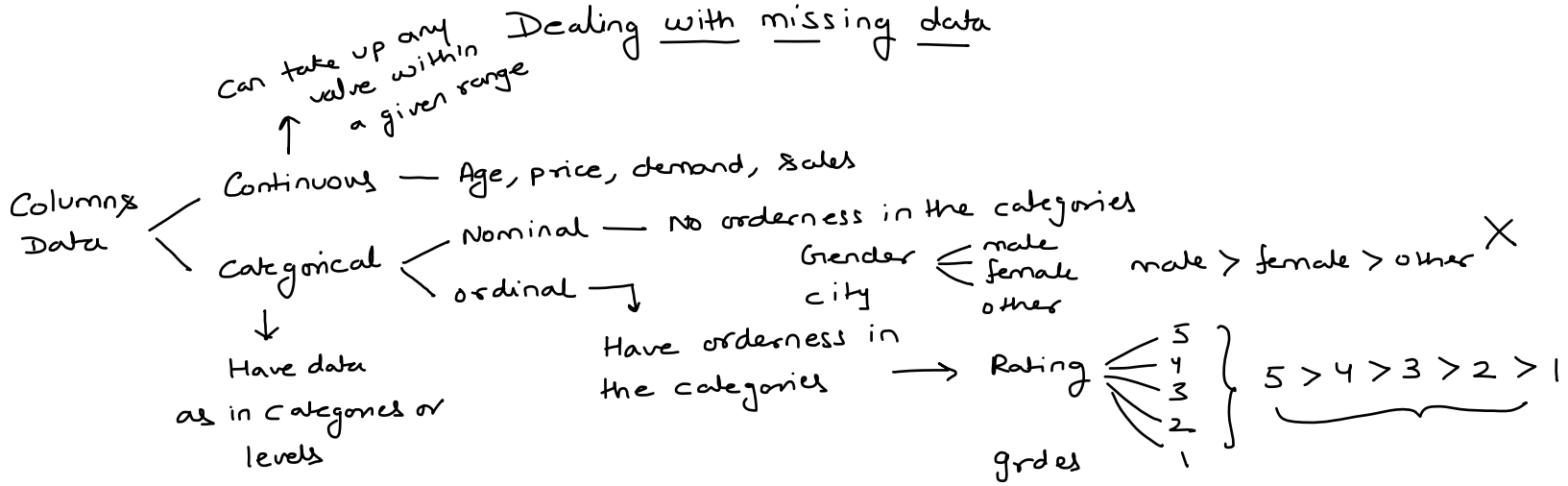
upGrad

Topics:

1. How to impute missing values
2. Kaggle: Video Games Data
3. QNA

Missing Value Imputation

upGrad

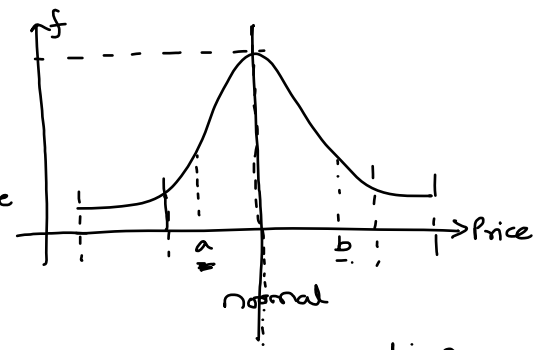


Missing Value Imputation

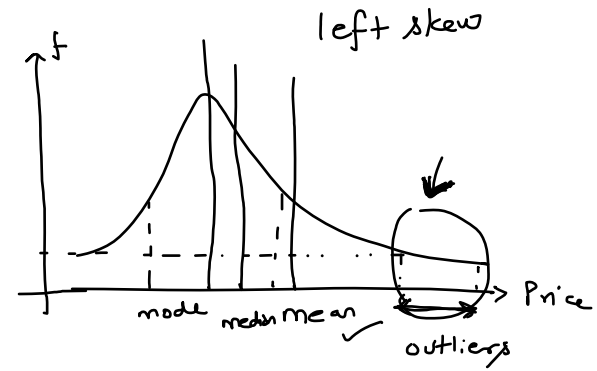
upGrad

Imputation

Continuous — mean / median



mean median mode



Price

Price

1.2
2.3
2.4
2.7
2.7
2.7
⋮

Continuous Column — Having outlier → avoid using mean as technique for imputation

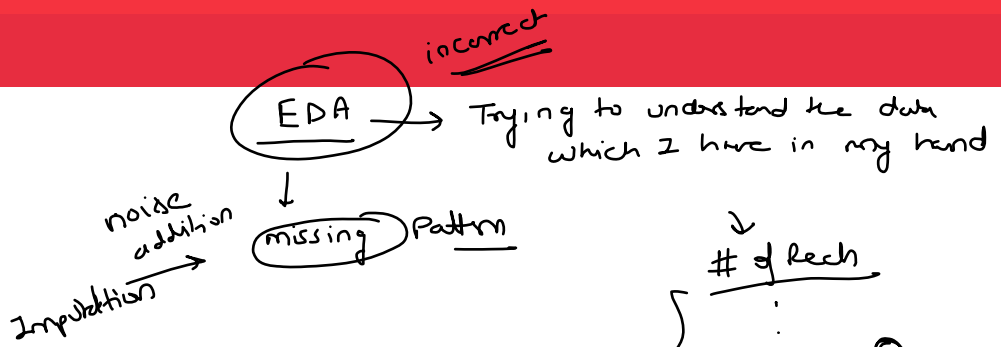
No outliers → anything

imputation

i/p clean data → ML → Model

Missing Value Imputation

upGrad



Category wise imputation

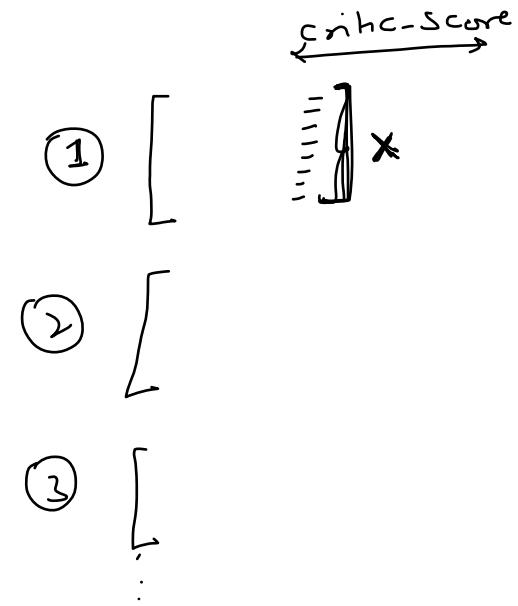
1. median (Rate)
2. median (Rate) where category = X

Product	Category	Rate
A1	X	10
A2	V	11
B3	X	
H5	V	12
R5	X	13
G5	X	12
V9	X	12

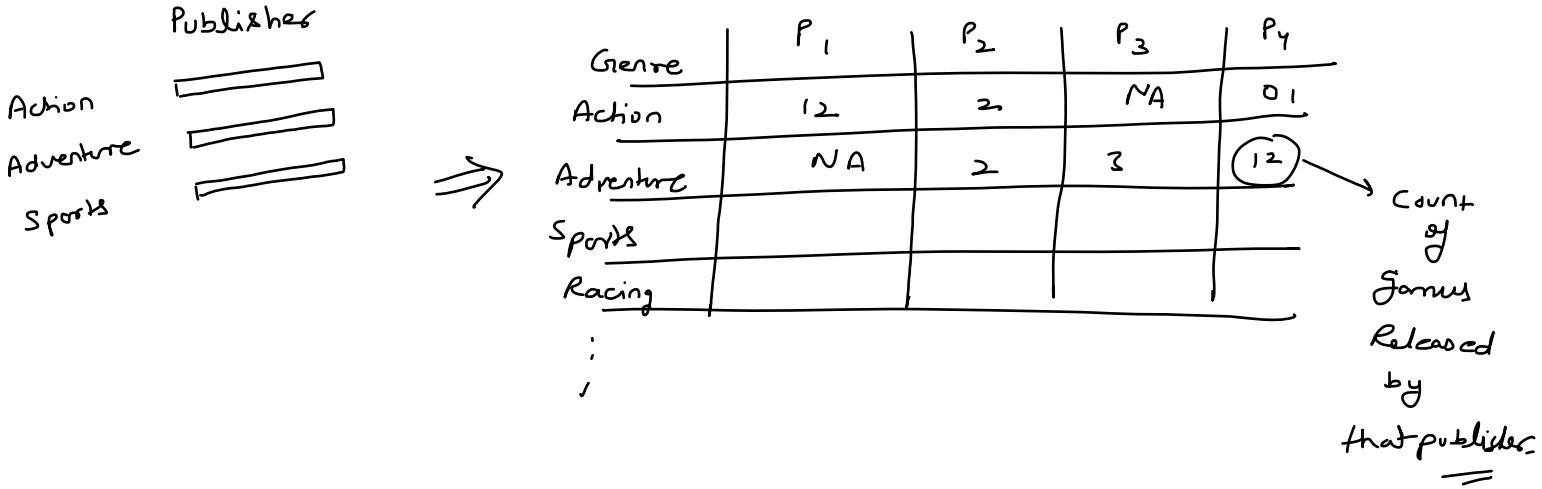
Missing Value Imputation

upGrad

```
df.groupby('Publisher')['Critic_Score'].apply(lambda x: x.fillna(x.median()))
```



Product	Category	Rate
A1	X	10
A2	V	11
B3	X	na
H5	V	12
R5	X	13
G5	X	12
V9	X	12

Missing Value Imputation

upGrad