10 Dec 2022 | IPBA 14 | R Data Exploration and Viz + Data Sanity - Session 1

10 December 2022 10:01 AM

SQL - R? - EDA, Stats, DataViz,

DB

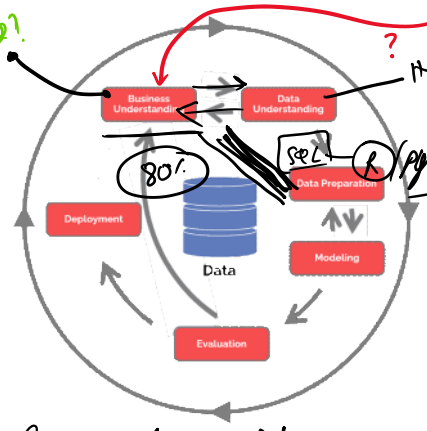
Bit

orders 1000

Rs, 5000 APO

$ PROFIT → Revenue - COSTS

value?



? Historical (SQL)

PM : CRISP-DM

1000 - 100 - 50-60 (online)

① ②

OID	Fraud orders	Channel	Product Group	Payment method
101	Yes	online	Electronic	COD
102	No	offline	Grocery	UPI
103	Yes	mobile	Body care	(NULL)
104	No	Desktop	Gym equip.	Cred
		Hybrid		Debit.

→ Dimentions (400)

NOT NORMAL → ?

① EDA (Business) + Data Quality

1) Univariate analysis
2) Bivariate
3) Multivariate.

② Data Pre processing (ML Ready Data)

- Columns
- Dimentions
- Factors
- Attri

1) Feature Selection (Reduction)
2) Missing strategy < NUM, CAT
3) Outlier strategy (NUM - SCALE)
4) TRANSFOMATION - NUM (NON Normal → NORMAL)
5) Catergorical Encoding - Dummy, ordical.

%>%

data → mutate (FF + NUM) → → → modified data