Type | cancer Analysis → objective

Analyze how features are Associated with Breast cancer.

↓

Data collection → Present in kaggle, UCL Repository sklearn module.

↓

Prefer Python for EDA, statistical Analysis

↓

followed by Power BI Analysis & Dashboard

Create Metadata
↳ of features/columns
↳ explanati explain features & its relevance
↓
& on be used to detect outliers & statistical Analysis

↓

Any other tools ? → Python → code availability
→ & know basics
→ easily detect outlier
→ handle large data easily
→ easily do statistical Analysis

excel → can be :
Baad me Dekhenge ↳ Excel chahiye y a nhi

→ chat GPT
→ Google

Basic Data cleaning steps
↳ Null value → Heatmap use kru?
↳ info of table/Data frame
↳ Target features ?

↳ isme transformation Add krte h.

↳ class Distribution check krenge & source ↳ Internet

need encoding for Power BI

find imbalance

→ irrelevant columns Ko Remove Kaise kru

↳ chatgpt → use Heatmap
↳ Limit / Threshold = 0.9

source code
↳ Stackoverflow

worst values → show disease condition value
↳ extreme
values h
↳ inko outlier ki trh
treat mat kariyo

mean
value se
compare kariyo
↳ general values
↳ benchmark

Both can be
used in
power BI

→ histograms +
kde
use it for
comparison b/w
B & M

kde plot
bnana h

Note
↳ medical data
Right skewed
hota h, so log
transformation
can be used
+
outlier plot.

→ scatter plot → use it to
Analyze or
compare feature

just
to see
how it
will

↳ change data.

→ Analysis

① final thoughts add
kariyo bcoz of
outlier dilemma

Univariate

class Distribution
of Diagnosis
kr skte h

Bivariate
Analysis
↓
compare kr skte
h mean +
worst ko