

Data cleaning → Missing values

↓

lat long columns → -ve values → abs

↓  
0 invalid → NaN

Date → Datetime features

cat columns → Trailing whitespace, lower case

Use → python → Upload

↙ Import → clean.

DVC pipeline → cmd

Detailed EDA → Functions → Univariate  
Bivariate  
Multivariate  
↓  
Hypothesis  
↙ ↘ ↘  
ANOVA Chi-2 JB

EDA → Feature Selection → Target impact  
↓  
Imputation