

# Forage Data Analytics with GenAI.

Task 1:- EDA (Exploratory Data Analysis)  
Techniques to handle missing values  
leverage synthetic data generation when data is limited.

Predict Creditcard delinquency.

Need of Company:- More structured data driven approach  
current method:- historical trends & broad segmentation models  
AI driven insights & model should be made

My work:

① Review dataset

- structure
- completeness
- key attributes
- missing, incorrect values identity

② Use GenAI Tools:

- generate insights
- ensure confidentiality

③ Summarize findings:-

- patterns
- anomalies
- key risk indicators
- report

\* EDA: identify patterns, trends, inconsistencies, missing values to ensure data quality and reliability.

\* Data integrity: correct data without missing values, duplicates, inconsistencies.

\* Steps in EDA:

① Understanding dataset

Key variables

categorical  
or  
numerical  
data points

missing  
or  
inconsistent  
values

}  
Use GenAI tool  
"ChatGPT"

prompt: "Analyze this dataset & provide a summary of key columns, including common patterns and missing values."



## ② Identity missing values & outliers: techniques

- ① Statistical imputation: Replace missing values with mean, median or regression based imputation.
- ② Understanding missingness patterns:
  - data is Missing completely at random (MCAR)
  - Missing at random (MAR)
  - Missing not at random (MNAR)
- ③ Removing Irrelevant data: ensure no bias.

GenAI prompt: Identity missing values in this dataset and recommend the best imputation strategy based on industry best practices.

## ③ Understanding relationships between variables:

GenAI prompt: Analyse the correlation b/w customer income and delinquency risk summarizing Key findings in simple terms.

## ④ Detecting patterns & risk factors:

GenAI prompt: Analyse trends in late payments and identify the top 3 risk factors associated with delinquency.

Use Mean = if data is well balanced without extreme values  
Use Median = if data contains outliers that effects average  
Use Mode = if single value is repeated many times.

## Applying

\* GenAI generated imputation without understanding data context can introduce significant bias.



## \* Techniques for handling missing values

### Before page techniques

#### → Causes of Missing data:

- ① Random errors
- ② Skewed data collection
- ③ Customer behaviour

#### → How to handle:

##### ① Deleting missing data

##### ② Imputation

Mean, Median, Mode  
Forward or backward filling - uses existing data trends to estimate missing entries.

##### ③ AI assisted Imputation.

detect patterns & provides suggestions of statistical imputation.  
suggest synthetic data

⇒ Check for duplicates, inconsistent formatting & logical errors.

## \* Synthetic data generation

↳ when real world data is limited or sensitive

synthetic data generation is done to fill gaps.

artificially generated data

is created by statistical models (or) AI driven techniques

this will provide security, privacy for real data.

averages & historical trends

uses ML models.

#### → Traditional Statistical Simulation Techniques:

##### ① Monte Carlo simulations

##### ② Bootstrapping

##### ③ Probabilistic modeling

When to use synthetic data:

- ① Enhancing small datasets
- ② Filling in missing values
- ③ Testing AI Models
- ④ Ensuring privacy compliance

GenAI Example prompt: Generate synthetic payment history data for customers with missing records while ensuring that distributions align with historical patterns observed in the dataset (eg, standard deviation, typical payment behaviors).

\* Make sure synthetic data doesn't introduce bias.

\*\* Use "Quadratic" free Tool for getting insights & analysis.