

Lab sheet #6

Synthetic Dataset Generation- EDA- Blog Writing

Exercise 1

Explore packages in python for synthetic data generation for numeric and categorical features. List at least three the packages and generate a dataset of 100 records with one numeric and one categorical attribute

Exercise 2

Generate a dataset for blog writing. Perform EDA. Write a blog based on your analysis of the dataset.

a. Evaluation Criteria of dataset generation

- *Minimum number of features: 8*
- *Minimum number of records: 1000*
- *There should be at least one label (discrete or continuous) to apply supervised learning*
- *Select features w.r.t a real-world problem. Clearly define the feature name, its datatype and range of values. Completely random dataset is not recommended.*

b. Evaluation Criteria of EDA

Component	Excellent	Satisfactory	Poor
Breadth of Exploration	Answer more than three questions about the data; target is substantially different aspects of the data.	At least 3 questions were initially asked of the data, but there is some overlap between questions.	Fewer than 3 initial questions were posed of the data.
Depth of Exploration	A sufficient number of follow-up questions were asked to yield insights that helped to more deeply explore the initial questions.	Some follow-up questions were asked, but they did not take the analysis much deeper than the initial questions.	No follow-up questions were asked after answering the initial questions.
Data Quality	Data quality was thoroughly assessed with extensive profiling of fields and records.	Simple checks were conducted on only a handful of fields or records.	Little or no evidence that data quality was assessed.
Visualizations	More than 10 visualizations were produced, and a variety of marks and encodings were explored. All design decisions were both expressive and effective.	At least 10 visualizations were produced. The visual encodings chosen were largely effective and expressive, but some errors remain.	Several ineffective or inexpressive design choices are made. Fewer than 10 visualizations have been produced.

Data Transformation	Effective transformation were used to extend the dataset in interesting or useful ways.	Simple transforms were primarily used.	The raw dataset was used directly, with little to no additional transformation.
Captions	Captions richly describe the visualizations and contextualize the insight within the analysis.	Captions do a good job describing the visualizations, but could better connect prior or subsequent steps of the analysis.	Captions are missing, overly brief, or shallow in their analysis of visualizations.
Creativity & Originality	You exceeded the parameters of the assignment, with original insights or a particularly engaging design.	You met all the parameters of the assignment.	You met most of the parameters of the assignment.

c. Evaluation criteria of blog writing

Criteria	Unsatisfactory – 0%	Limited – 80%	Proficient – 90%	Exemplary – 100%
Content and Creativity	Postings show no evidence of insight, understanding or reflective thought about the topic.	Postings provide minimal insight, understanding and reflective thought about the topic.	Postings provide moderate insight, understanding and reflective thought about the topic.	Postings provide comprehensive insight, understanding, and reflective thought about the topic.
Text Layout, Use of Graphics	Does not insert any graphics, or uses	Selects and inserts many low-quality graphics and multimedia which do not enhance the content.	Selects and inserts graphics and multimedia that are mostly high quality and enhance and clarify the content.	Selects and inserts high quality graphics and multimedia when appropriate to enhance the content's visual appeal and increase readability.
Quality of Writing and Proofreading	How do you organize your ideas and information in a logical and coherent way? How do you use headings, subheadings, paragraphs, sentences, and transitions to guide your readers and make your content easy to scan and read? How do you use images, videos, graphs, tables, bullet points, and other visual elements to enhance your content and break up the text? How do you support your claims and arguments with			

	evidence, examples, and sources? How do you avoid plagiarism, errors, and misinformation?
--	---