SNS COLLEGE OF ENGINEERING

Kurumbapalayam (po), Coimbatore – 641 107

Accredited by NAAC-UGC with 'A' Grade

Approved by AICTE & Affiliated to Anna University, Chennai

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

19AD504 – DATA VISUALIZATION

UNIT –I

INTRODUCTION TO DATA VISUALIZATION

1.4 UNDERSTAND THE CONTEXT,EXPLORATORY VS EXPLANATORY ANALYSIS

Context:

In simple language, context means the setting of an event. You can think of context as all the information you need to know to truly understand something.

Context is information that helps viewers of your visualisation better understand what they're looking at. In the example below, the text "Record High" helps the person viewing the graph know that the data item shown is the highest ever recorded.

Without this small bit of text for context, it would be difficult to show this visually without showing the entire history of data for this item. Even then, the text might do the job better.

Adding context is often helpful, so consider when it can add to your presentation.

Here are three simple rules for adding context

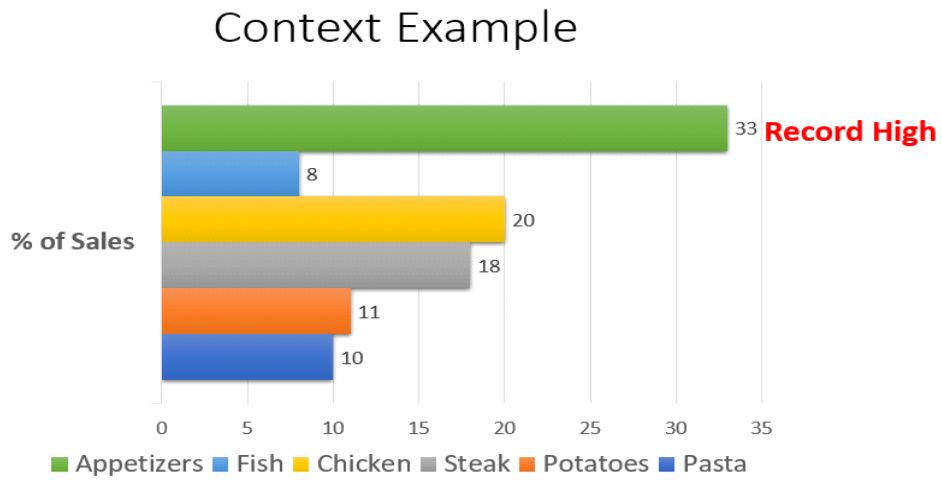
Brevity- Use as few words/letters as possible. Keep it to one line when possible.

For example, use TTM instead of trailing twelve months if you can.

Simplicity- Use simple words or acronyms. For example, use vs. utilise.

Clarity - Be clear and specific. Don't say good or bad if you can provide an objective and specific piece of information that speaks for itself.

For example, when you watch a movie from the very start, you learn about the names of the characters, where the movie is set, and what the plot line is. You can follow the movie's storyline because you've learned about the context.



But when someone walks into the movie half-way through, they don't really know what's going on. They don't know the characters' names, the storyline, where it's set, or what the relationships between the characters are. They're confused because they don't understand the context of the movie.

Often times, when thinking about the context of a situation, it's a good idea to think about what you'd need to explain to an alien about a situation to get them caught-up on what's going on.

Exploratory data analysis

Exploratory data analysis (EDA) is the process of summarising the data using statistical and visualisation approaches in order to focus on key elements of the data for additional analysis.

This involves studying the dataset from a variety of perspectives and characterising and summarising it without assuming anything about what it contains.

Exploratory data visualisation typically starts with a broad view of the data, using plots such as histograms, scatter plots, and box plots to understand the distribution and relationships among variables.

As the analyst gains a deeper understanding of the data, they may create more specialised visualisations to focus on specific relationships or patterns.

One of the key goals of exploratory data visualisation is to identify areas of interest or anomalies in the data that may warrant further investigation.

By creating a variety of plots and comparing them to one another, the analyst can gain a more comprehensive understanding of the data and identify trends, patterns, and relationships that may not be immediately apparent from looking at the raw data.

It's important to note that exploratory data visualisation is just one step in the data analysis process.

After identifying patterns and relationships through visualization, the analyst may need to conduct additional analyses, such as statistical tests or machine learning models, to confirm and further understand the findings.

Two main aspects of EDA are:

Openness:

meaning a person exploring the data should be open to all possibilities prior to its exploration.

Skepticism:

meaning one must ensure that the obvious story the data tells is not misleading

Types

There are four primary types of EDA:

Univariate non-graphical : This is simplest form of data analysis, where the data being analyzed consists of just one variable. Since it's a single variable, it doesn't deal with causes or relationships. The main purpose of univariate analysis is to describe the data and find patterns that exist within it.

Univariate graphical : Non-graphical methods don't provide a full picture of the data. Graphical methods are therefore required. Common types of univariate graphics include:

Stem-and-leaf plots, which show all data values and the shape of the distribution.

Histograms, a bar plot in which each bar represents the frequency (count) or proportion (count/total count) of cases for a range of values.

Box plots, which graphically depict the five-number summary of minimum, first quartile, median, third quartile, and maximum.

Multivariate nongraphical: Multivariate data arises from more than one variable. Multivariate non-graphical EDA techniques generally show the relationship between two or more variables of the data through cross-tabulation or statistics.

Multivariate graphical: Multivariate data uses graphics to display relationships between two or more sets of data. The most used graphic is a grouped bar plot or bar chart with each group representing one level of one of the variables and each bar within a group representing the levels of the other variable.

Explanatory data analysis

Explanatory visualizations help managers communicate the results of their analyses to influence changes and improvements.

Explanatory data visualization refers to the use of data visualization techniques to present data and information in a way that helps to explain or clarify a specific idea or concept.

This type of visualization is often used to communicate the findings or insights of an analysis or study, and is typically accompanied by a narrative or explanation to provide context and meaning to the data.

Explanatory data visualizations often involve the use of charts, graphs, maps, and other visual elements to represent data in a clear and concise manner. They may also include annotations, labels, and other elements to help explain the data and its significance.

Explanatory data visualization is an important tool for effectively communicating complex data and ideas to a broad audience, and can be used in a variety of fields including business, science, economics, and more.

To create an Explanatory Data Visualization, you might simplify your charts by excluding irrelevant data and you might use styling features to highlight the key message.

Explanatory	Exploratory
Data explanation	Data exploration
Meant for non-expert audience, having no background knowledge of the subject matter	Meant for expert user group, having prior knowledge in the subject matter
Represents understandable visual data	Represents complexities of big data
Doesn’t have analytical purposes	Has analytical purposes