Week 5 Report

Topic: Types of Data in a Dataset, Visualization Basics, Data Modeling

1. Types of Data in a Dataset

Understanding the types of data in a dataset is crucial for analysis and visualization. The main types include:

1.1 Quantitative Data

Quantitative data is numerical and can be measured or counted. - **Continuous data:** Can take any value within a range (e.g., temperature, sales revenue). - **Recommended Visualization:** Line Chart - **Example:** Showing monthly sales revenue trends. - **Discrete data:** Only certain values are possible (e.g., number of customers, disaster occurrences). - **Recommended Visualization:** Column Chart - **Example:** Number of disasters per state in a year.

1.2 Qualitative Data

Qualitative data is descriptive and categorizes information. - **Nominal data:** Categories with no natural order (e.g., disaster type, state name). - **Recommended Visualization:** Bar Chart - **Example:** Count of each disaster type in the dataset. - **Ordinal data:** Categories with a meaningful order (e.g., risk levels: low, medium, high). - **Recommended Visualization:** Stacked Bar Chart - **Example:** Distribution of low, medium, high-risk states.

1.3 Time-Series Data

Data collected over time intervals to observe trends or patterns. - **Recommended Visualization:** Line Chart - **Example:** Yearly disaster occurrences trend.

1.4 Geospatial Data

Data related to locations, often represented with coordinates or maps. - **Recommended Visualization:** Choropleth Map - **Example:** Disaster frequency by state.

2. Visualization Basics

Visualization is key for understanding patterns, trends, and insights in datasets. - Helps in identifying patterns and anomalies. - Supports decision-making. - Tools: Python (Matplotlib, Seaborn), Power BI, Excel.

3. Data Modeling

Organizes and structures data to optimize analysis and visualization. - **Fact Tables:** Contain measurable data (e.g., number of affected people, cost). - **Dimension Tables:** Contain descriptive attributes (e.g., disaster type, state, year). - **Relationships:** Connect fact and dimension tables for data integrity.

4. Key Learnings

- Correct visualization choice depends on data type.
- Line charts are useful for continuous or time-series data.
- Bar/column charts are ideal for discrete or categorical data.
- Data modeling supports accurate visualizations.

5. Outcome of Week 5

- Understood data types and their suitable visualizations.
- Learned visualization best practices.
- Prepared groundwork for dashboards in Power BI.

6. Plan for Next Week

- Apply visualizations to real datasets.
- Build dashboards in Power BI using fact and dimension tables.
- Explore advanced visualizations and interactive reports.