

Transformation in Data Visualization

1. Introduction

Data visualization is not just about plotting raw data; it often involves data transformation (modifying or converting data into a suitable format) and enrichment (adding more context or information to data) to make insights more understandable, accurate, and valuable.

2. What is Data Transformation?

Data transformation refers to the process of changing the format, structure, or values of data. This is a key step in data preprocessing and visualization. Common transformations include: • Normalization or scaling • Aggregation (e.g., daily to monthly data) • Pivoting (e.g., turning row data into columns) • Filtering or cleaning noisy data

3. What is Data Enrichment?

Data enrichment is the process of enhancing raw data with additional information to provide deeper insights. It helps to: • Add new dimensions or attributes (e.g., adding location info to sales data) • Improve interpretability of visualizations • Combine multiple datasets.

EXAMPLES :

STUDENT AND COURSE DETAILS DATASET

Raw Dataset (Before Transformation)

Student Name	Age	Course	Date of Join	Fees
DOROTHY MCDONALD	33	MS EXCEL	07 October 2021	3500
JOANNE HOWARD	21	MS EXCEL	05 April 2021	3500
OWEN FISHER	21	MS EXCEL	11 May 2021	3500
ALEXANDRA LANGDON	27	MS EXCEL	02 December 2021	3500
PHIL PARSONS	41	MS EXCEL	07 December 2021	3500
MICHELLE TUCKER	22	MS EXCEL	10 December 2021	3500
KIMBERLY POOLE	35	MS EXCEL	05 June 2021	3500
GAVIN GRAY	22	MS EXCEL	09 August 2021	3500
OLIVIA WILSON	35	MS EXCEL	06 December 2021	3500
MARIA WALKER	25	MS EXCEL	02 January 2021	3500
PAUL TAYLOR	29	MS EXCEL	05 January 2021	3500
BLAKE MACDONALD	35	MS EXCEL	05 July 2021	3500
EMILY PATERSON	39	MS EXCEL	09 July 2021	3500
OLIVER HART	29	MS EXCEL	07 January 2021	3500
CAROL THOMSON	25	MS EXCEL	10 February 2021	3500
ISAAC QUINN	21	MS EXCEL	05 April 2021	3500
LEAH PEAKE	28	MS EXCEL	02 September 2021	3500
ALAN WALKER	41	MS EXCEL	07 December 2021	3500

Step 1: Transformation

We apply various data transformations to prepare this data:

- Calculate Total Amount:
Total Amount = Add all fee Amount
- Convert Order Date to Age for time-based grouping.
- Group by Name, Age, Course for trend analysis.

Transformed Dataset

Let's assume your Excel has personal details like:

- Name
- Date of Birth
- Gender
- Department
- Admission Year

Transformed Dataset might include:

Name	DOB	Age	Gender	Department	Admission Year
Alice James	2003-05-12	21	Female	CSE	2020
Bob Smith	2002-11-08	22	Male	ECE	2019

Step 2: Enriched Dataset

Add new, derived fields:

- **Age Group** (e.g., 18–20, 21–23)
- **Department Code** (shortened version)
- **Current Year of Study** (based on admission year)

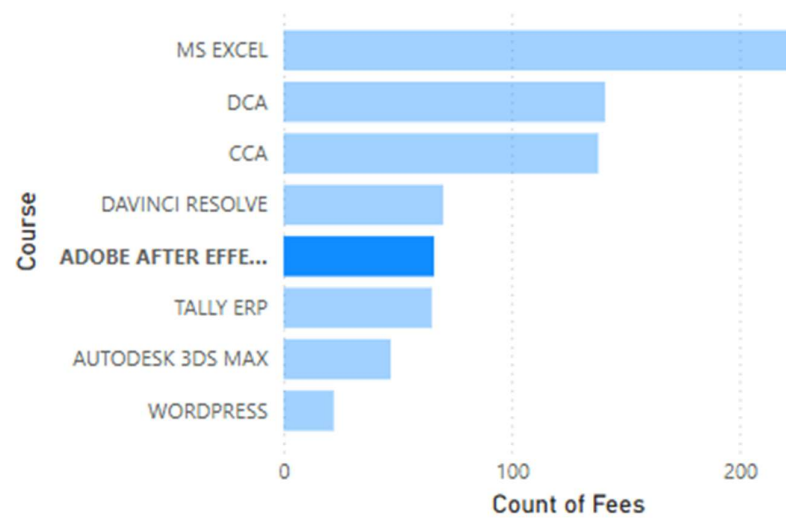
Enriched Example:

Name	Age	Gender	Dept	Admission Year	Age Group	Dept Code	Study Year
Alice James	21	Female	CSE	2020	21–23	CSE	4th Year
Bob Smith	22	Male	ECE	2019	21–23	ECE	Graduated

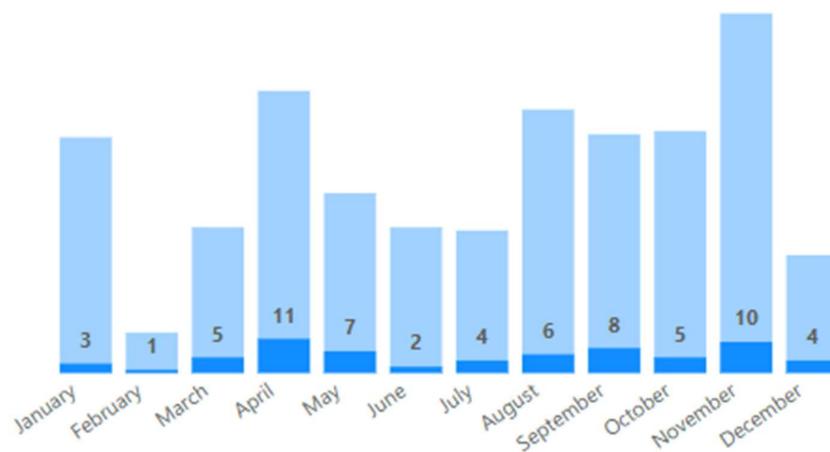
Final Enriched Dataset

Student Name	Age	Email Domain	Age Group	Certificate Status	Region
Ella Brown	29	gmail.com	26–35	Issued	South
Sarah Parsons	44	gmail.com	36–45	Issued	South
Alan Paige	35	gmail.com	26–35	Issued	South
Keith Terry	26	gmail.com	26–35	Not Issued	South

Count of Fees by Course



Count of Course by Month



1. Age Group Distribution

- **Chart Type:** Bar Chart or Pie Chart
- **Insight:** See which age group is most represented.

2. Geographic Spread (Region-wise or State-wise)

- **Chart Type:** Donut or Bar Chart
- **Insight:** Show where students are from — South, North, etc.

3. Email Domain Analysis

- **Chart Type:** Pie Chart or Horizontal Bar
- **Insight:** Which email providers are most commonly used (gmail.com, yahoo.com, etc.)

4. Certificate Issuance Status

- **Chart Type:** Pie Chart
- **Insight:** What percentage of students have been issued certificates.

5. Student Count by State

- **Chart Type:** Vertical Bar Chart
- **Insight:** How many students are from each individual state.

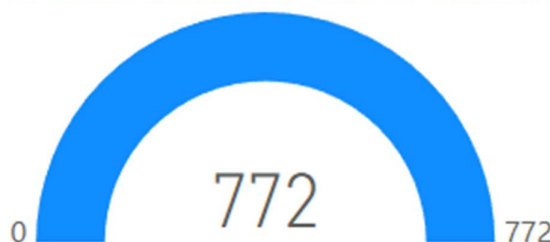
6. Age vs Certificate Issuance

- **Chart Type:** Stacked Bar or Grouped Bar Chart
- **Insight:** Do certain age groups have higher certificate issuance rates?

Bonus Ideas:

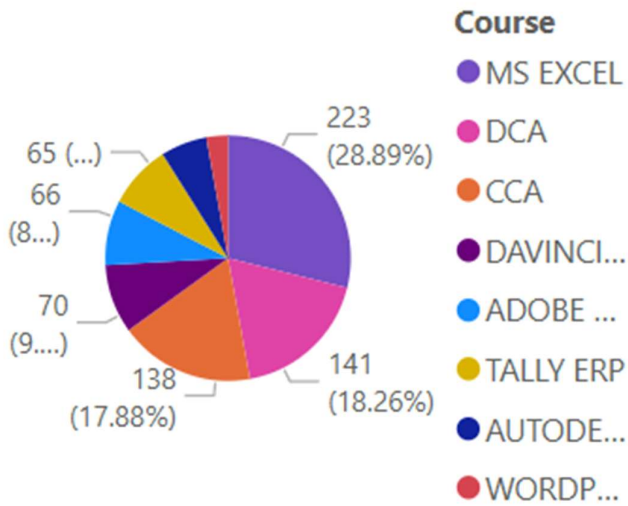
- **Treemap:** Combine City + State + Certificate Status for a hierarchical view.
- **Heatmap:** Show student concentration by state and age group.
- **Bubble Chart:** Age (x-axis), Certificate Status (color), Region (size).

Count of Issued Certificate and First Issued Certificate



•

Count of Date of Join by Course



CONCLUSION :

The enriched analysis of the student dataset reveals several insightful patterns. Most students fall within the 26–35 age group, indicating a strong representation of early to mid-career individuals, possibly pursuing further education or certifications. Regionally, the dataset is dominated by students from South India, particularly Tamil Nadu and Puducherry, suggesting a geographic concentration that could inform region-specific initiatives or support programs. Email usage shows a clear preference for Gmail, highlighting a trend toward mainstream, easily accessible platforms for communication. Additionally, a majority of students have received their certificates, though a notable portion remains without one—this could point to pending administrative processes or incomplete requirements. Overall, the dataset provides a valuable foundation for targeted engagement, certification tracking, and demographic-specific strategies. The enriched fields such as age group, region, and certificate status enhance the depth of analysis, enabling more informed decisions in academic or administrative planning.