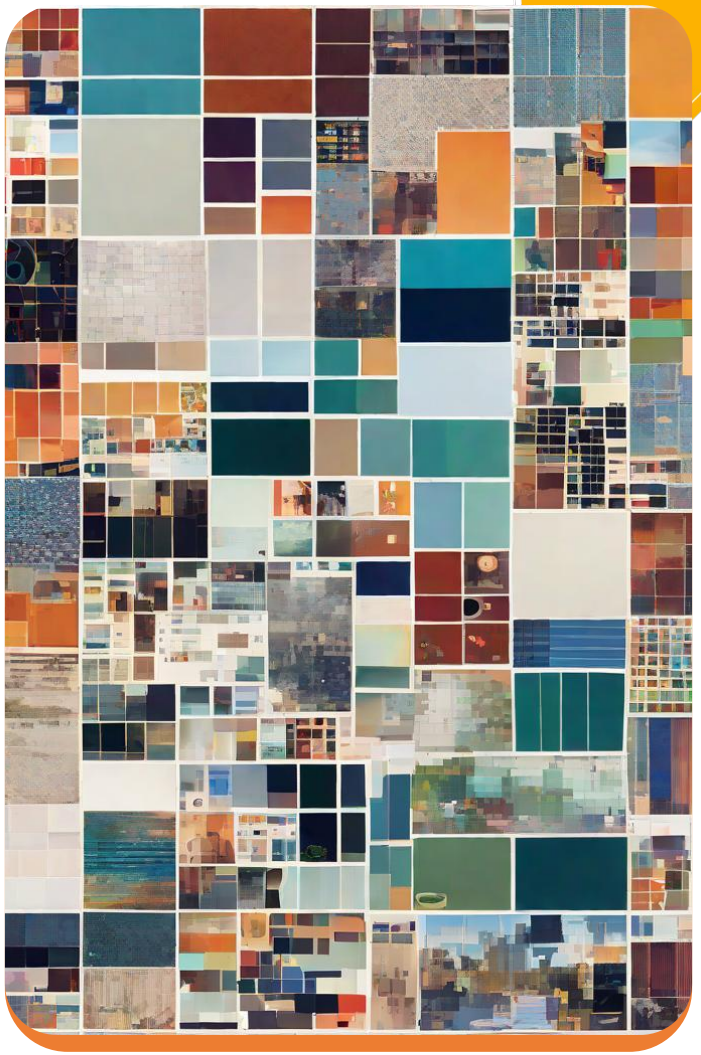
Data Concatenation

This lesson provides an overview of data concatenation, its purpose, usage, and examples. It explains the concepts of rows concatenation and column concatenation, and concludes with the importance of data concatenation in data manipulation tasks.



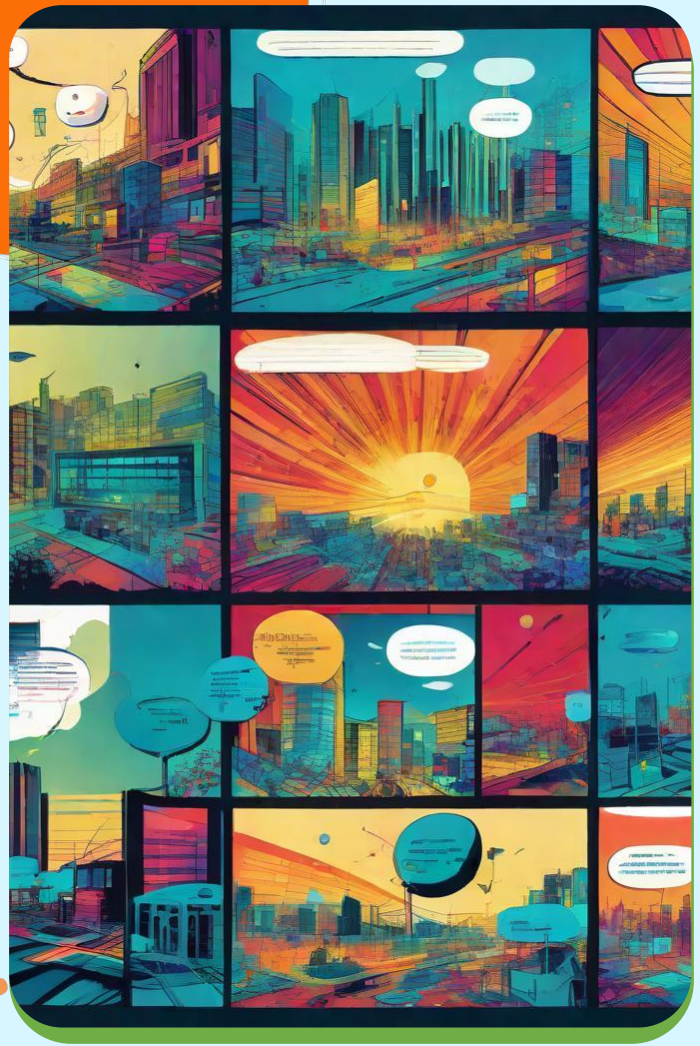
Introduction

01 It's a fundamental operation in data manipulation and is often used in data preprocessing and analysis tasks.

02 Data concatenation is the process of combining two or more datasets by appending them either along rows or columns.

Definition

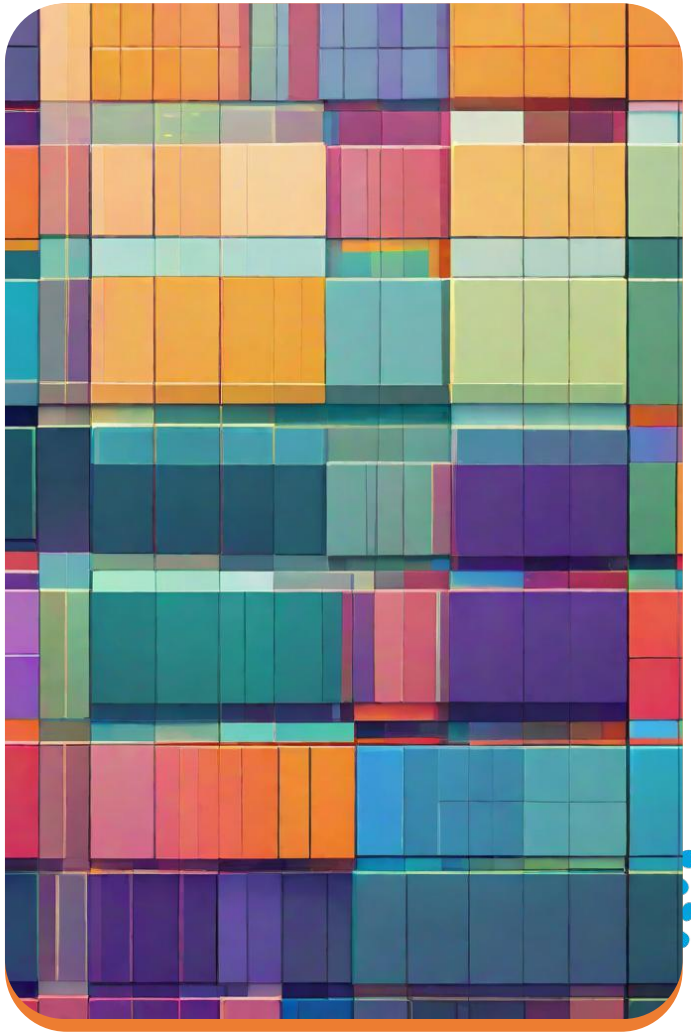
- Data concatenation refers to the merging of datasets either by adding rows (stacking vertically) or by adding columns (joining horizontally).
- This operation does not perform any alignment of data other than along the concatenation axis, which means it simply stitches datasets together without any additional processing.



Purpose

01 This allows for easier analysis, visualization, and manipulation of the combined data.

02 The primary purpose of data concatenation is to combine multiple datasets to form a single, larger dataset that contains all the information from the original datasets.



Usage

- Data concatenation is commonly used in various data science tasks such as data cleaning, merging different sources of data, and preparing data for analysis.
- It is particularly useful when dealing with datasets that have the same structure but different observations or when merging data from different time periods or sources.

Rows Concatenation

01 This results in a larger dataset with more rows, where the observations from the second dataset are added below the observations from the first dataset.

02 Rows concatenation involves stacking datasets on top of each other, essentially appending one dataset's rows to another's.

Rows Concatenation

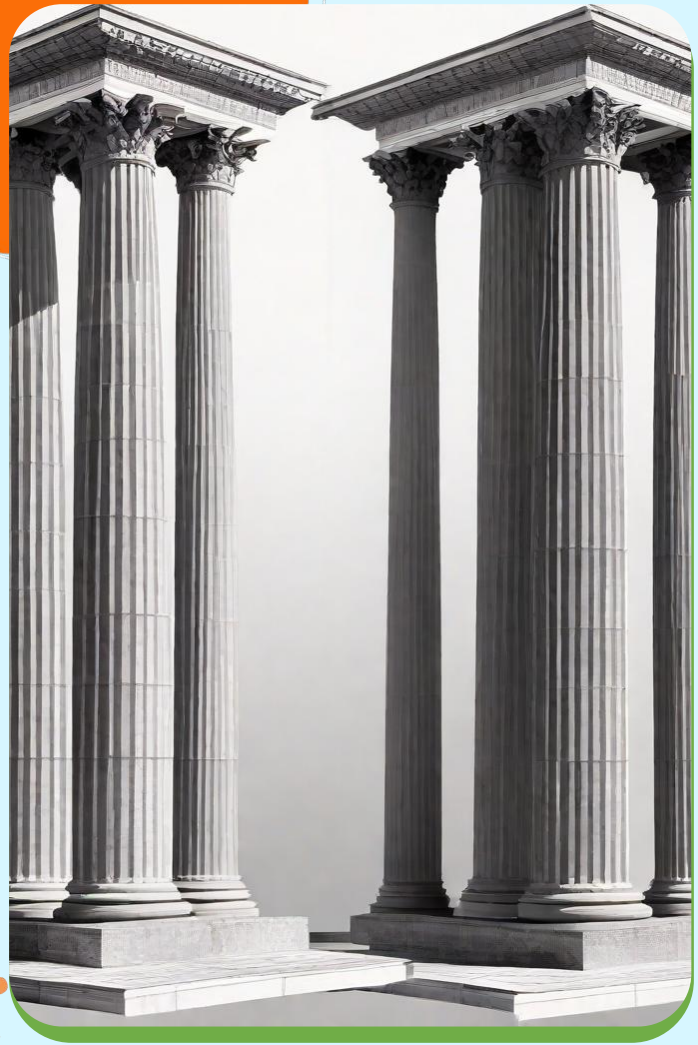
Customer ID	Country	Age	Gender
1	India	33	Male
2	UK	20	Male
3	USA	36	Male
4	UK	21	Female
5	USA	23	Male
6	USA	28	Female
7	UK	36	Female
8	India	24	Female
9	UK	30	Male

Customer ID	Country	Age	Gender
10	UK	26	Female
11	USA	30	Male
12	India	26	Male
13	UK	33	Male
14	USA	31	Female
15	USA	35	Female
16	India	19	Female
17	India	37	Female
18	India	39	Male

Customer ID	Country	Age	Gender
1	India	28	Male
2	UK	18	Male
3	USA	37	Male
4	UK	34	Female
5	USA	35	Male
6	USA	36	Female
7	UK	22	Female
8	India	28	Female
9	UK	37	Male
10	UK	28	Female
11	USA	35	Male
12	India	22	Male
13	UK	33	Male
14	USA	26	Female
15	USA	40	Female
16	India	28	Female
17	India	30	Female
18	India	40	Male

Column Concatenation

- Column concatenation involves joining datasets side by side, adding the columns of one dataset to the columns of another.
- This results in a wider dataset with more columns, where the variables from the second dataset are added next to the variables from the first dataset.



Column Concatenation

Customer ID	Country	Age	Gender
1	India	22	Male
2	UK	36	Male
3	USA	20	Male
4	UK	28	Female
5	USA	36	Male
6	USA	38	Female
7	UK	33	Female
8	India	34	Female
9	UK	24	Male
10	UK	21	Female

Revenue	Cost	Quantity
18579	4254	49
55274	4985	41
57696	3668	39
93685	3106	32
22791	2780	30
73330	4614	40
70055	1962	17
59865	1923	25
52342	2982	34
97516	3771	13

Customer ID	Country	Age	Gender	Revenue	Cost	Quantity
1	India	33	Male	96121	4677	28
2	UK	28	Male	44744	4456	41
3	USA	29	Male	89248	2713	11
4	UK	23	Female	84740	4666	18
5	USA	22	Male	35814	1488	11
6	USA	39	Female	37049	2590	22
7	UK	27	Female	23727	1297	43
8	India	30	Female	61689	2187	43
9	UK	36	Male	93241	3080	20
10	UK	39	Female	81117	4122	45

Conclusion

01 Whether stacking datasets vertically or joining them horizontally, understanding how to concatenate data is essential for anyone working with data in various domains.

02 Data concatenation is a powerful technique in data manipulation, allowing analysts to merge datasets efficiently for analysis and modeling purposes.

Thank you for your time and attention 🙂