

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.



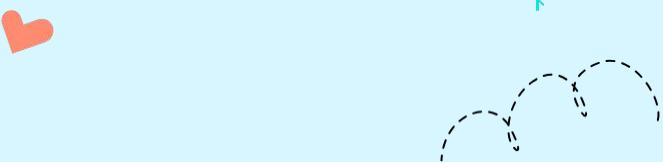
Rows Concatenation

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.



Dataset	DA01	DA02	DA03	DA04
Row 1	Value 1	Value 1	Value 1	Value 1
Row 2	Value 2	Value 2	Value 2	Value 2
Row 3	Value 3	Value 3	Value 3	Value 3
Row 4	Value 4	Value 4	Value 4	Value 4
Row 5	Value 5	Value 5	Value 5	Value 5
Row 6	Value 6	Value 6	Value 6	Value 6
Row 7	Value 7	Value 7	Value 7	Value 7
Row 8	Value 8	Value 8	Value 8	Value 8
Row 9	Value 9	Value 9	Value 9	Value 9
Row 10	Value 10	Value 10	Value 10	Value 10
Row 11	Value 11	Value 11	Value 11	Value 11
Row 12	Value 12	Value 12	Value 12	Value 12
Row 13	Value 13	Value 13	Value 13	Value 13
Row 14	Value 14	Value 14	Value 14	Value 14
Row 15	Value 15	Value 15	Value 15	Value 15
Row 16	Value 16	Value 16	Value 16	Value 16
Row 17	Value 17	Value 17	Value 17	Value 17
Row 18	Value 18	Value 18	Value 18	Value 18
Row 19	Value 19	Value 19	Value 19	Value 19
Row 20	Value 20	Value 20	Value 20	Value 20
Row 21	Value 21	Value 21	Value 21	Value 21
Row 22	Value 22	Value 22	Value 22	Value 22
Row 23	Value 23	Value 23	Value 23	Value 23
Row 24	Value 24	Value 24	Value 24	Value 24
Row 25	Value 25	Value 25	Value 25	Value 25
Row 26	Value 26	Value 26	Value 26	Value 26
Row 27	Value 27	Value 27	Value 27	Value 27
Row 28	Value 28	Value 28	Value 28	Value 28
Row 29	Value 29	Value 29	Value 29	Value 29
Row 30	Value 30	Value 30	Value 30	Value 30
Row 31	Value 31	Value 31	Value 31	Value 31
Row 32	Value 32	Value 32	Value 32	Value 32
Row 33	Value 33	Value 33	Value 33	Value 33
Row 34	Value 34	Value 34	Value 34	Value 34
Row 35	Value 35	Value 35	Value 35	Value 35
Row 36	Value 36	Value 36	Value 36	Value 36
Row 37	Value 37	Value 37	Value 37	Value 37
Row 38	Value 38	Value 38	Value 38	Value 38
Row 39	Value 39	Value 39	Value 39	Value 39
Row 40	Value 40	Value 40	Value 40	Value 40
Row 41	Value 41	Value 41	Value 41	Value 41
Row 42	Value 42	Value 42	Value 42	Value 42
Row 43	Value 43	Value 43	Value 43	Value 43
Row 44	Value 44	Value 44	Value 44	Value 44
Row 45	Value 45	Value 45	Value 45	Value 45
Row 46	Value 46	Value 46	Value 46	Value 46
Row 47	Value 47	Value 47	Value 47	Value 47
Row 48	Value 48	Value 48	Value 48	Value 48
Row 49	Value 49	Value 49	Value 49	Value 49
Row 50	Value 50	Value 50	Value 50	Value 50
Row 51	Value 51	Value 51	Value 51	Value 51
Row 52	Value 52	Value 52	Value 52	Value 52
Row 53	Value 53	Value 53	Value 53	Value 53
Row 54	Value 54	Value 54	Value 54	Value 54
Row 55	Value 55	Value 55	Value 55	Value 55
Row 56	Value 56	Value 56	Value 56	Value 56
Row 57	Value 57	Value 57	Value 57	Value 57
Row 58	Value 58	Value 58	Value 58	Value 58
Row 59	Value 59	Value 59	Value 59	Value 59
Row 60	Value 60	Value 60	Value 60	Value 60
Row 61	Value 61	Value 61	Value 61	Value 61
Row 62	Value 62	Value 62	Value 62	Value 62
Row 63	Value 63	Value 63	Value 63	Value 63
Row 64	Value 64	Value 64	Value 64	Value 64
Row 65	Value 65	Value 65	Value 65	Value 65
Row 66	Value 66	Value 66	Value 66	Value 66
Row 67	Value 67	Value 67	Value 67	Value 67
Row 68	Value 68	Value 68	Value 68	Value 68
Row 69	Value 69	Value 69	Value 69	Value 69
Row 70	Value 70	Value 70	Value 70	Value 70
Row 71	Value 71	Value 71	Value 71	Value 71
Row 72	Value 72	Value 72	Value 72	Value 72
Row 73	Value 73	Value 73	Value 73	Value 73
Row 74	Value 74	Value 74	Value 74	Value 74
Row 75	Value 75	Value 75	Value 75	Value 75
Row 76	Value 76	Value 76	Value 76	Value 76
Row 77	Value 77	Value 77	Value 77	Value 77
Row 78	Value 78	Value 78	Value 78	Value 78
Row 79	Value 79	Value 79	Value 79	Value 79
Row 80	Value 80	Value 80	Value 80	Value 80
Row 81	Value 81	Value 81	Value 81	Value 81
Row 82	Value 82	Value 82	Value 82	Value 82
Row 83	Value 83	Value 83	Value 83	Value 83
Row 84	Value 84	Value 84	Value 84	Value 84
Row 85	Value 85	Value 85	Value 85	Value 85
Row 86	Value 86	Value 86	Value 86	Value 86
Row 87	Value 87	Value 87	Value 87	Value 87
Row 88	Value 88	Value 88	Value 88	Value 88
Row 89	Value 89	Value 89	Value 89	Value 89
Row 90	Value 90	Value 90	Value 90	Value 90
Row 91	Value 91	Value 91	Value 91	Value 91
Row 92	Value 92	Value 92	Value 92	Value 92
Row 93	Value 93	Value 93	Value 93	Value 93
Row 94	Value 94	Value 94	Value 94	Value 94
Row 95	Value 95	Value 95	Value 95	Value 95
Row 96	Value 96	Value 96	Value 96	Value 96
Row 97	Value 97	Value 97	Value 97	Value 97
Row 98	Value 98	Value 98	Value 98	Value 98
Row 99	Value 99	Value 99	Value 99	Value 99
Row 100	Value 100	Value 100	Value 100	Value 100

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

The diagram illustrates the process of column concatenation across three separate tables. Red checkmarks and lines highlight the mapping of columns from the first table to the second, and from the second to the third.

Table 1: Customer Data

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

Table 2: Sales Data

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

Table 3: Final Data

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 😊

