**Task – add column with the number of branches**

There is a dataframe with two columns:

branch\_id - id of a branch

parent\_id - id of a parent company

|  |  |
| --- | --- |
| **branch\_id** | **parent\_id** |
| 1 | 10 |
| 2 | 10 |
| 3 | 10 |
| 4 | 15 |
| 5 | 15 |
| 6 | 20 |
| 6 | 20 |
| 7 | 25 |
| 7 | 25 |
| 8 | 25 |

Each parent company may have one or more branches.

The task is to add a third column, which says how many branches have the parent company (including the branch for which is the record):

|  |  |  |
| --- | --- | --- |
| **branch\_id** | **parent\_id** | **number\_of\_branches** |
| 1 | 10 | 3 |
| 2 | 10 | 3 |
| 3 | 10 | 3 |
| 4 | 15 | 2 |
| 5 | 15 | 2 |
| 6 | 20 | 1 |
| 6 | 20 | 1 |
| 7 | 25 | 2 |
| 7 | 25 | 2 |
| 8 | 25 | 2 |

Branch #1 has parent #10. There are two other branches that have the same parent #10. Total number of branches with parent #10 is 3, so in the column number\_of\_branches we put 3.

(same for all yellow records)

The green records have parent #15, so for all these records number\_of\_branches is 2.

We count only number of unique branch\_ids, so if there are many records with the same branch\_ids, we count the as one – e.g. blue records are the same, so it’s still only one branch.

In pink records with parent #25 there are two 7s and one 8, so there are two unique branches – in the last column we put 2.

The records are not in order, neither by branch\_id, nor by parent\_id.

Use attached spreadheet Branches.xsls as an input dataframe for testing.