

Introduction / Welcome

Welcome to Data Engineering Professional!

In order to build and validate a valuable Machine Learning model, you need prepared data. In this course we look in detail how to use RapidMiner Studio to perform some of the most common data preparation steps. The course is appropriate for anyone that will build RapidMiner processes, because almost everyone needs to do at least some data preparation. This uses a hands-on approach and you should plan to spend time actively building data prep processes in RapidMiner Studio.

The course can help you prepare for the [Data Engineering Professional Exam](#). We do not cover the answers to all questions on the exam, instead we ask you to take ownership for learning, understanding, and practicing the topics that we outline.

Please remember that we frequently update these pathways and are happy to hear any suggestions for improvement.

Auto Cleaning

Define Target Improve Quality Change Types Handle Numbers Summary

This table is just for your information. Rapidminer will automatically remove the columns highlighted below since they have a very low quality for machine learning. We will also replace all missing values for the remaining columns.

duration Number	wage_inc.1st Number	wage_inc.2nd Number	wage_inc.3rd Number	col.adj Category	working.hours Number	pension Category	standby.pay Number
1	5	?	?	?	48	?	?
2	4.500	5.800	?	?	35	ret_aliv	?
?	?	?	?	?	38	empl_centr	?
3	3.700	4	5	tc	?	?	?
3	4.500	4.500	5	?	48	?	?
2	3	2.500	?	?	35	?	?
3	4	5	5	tc	?	empl_centr	?
3	6.900	4.900	2.300	?	48	?	?
2	3	7	?	?	38	?	12
1	5.700	?	?	none	48	empl_centr	?

Many missing

Many missing

BACK NEXT

