

Data processing :

Library

Select Datasets

Select Data Source or Upload local file

Click here or drag-and-drop files to upload

Your options for CAREGIVERS.csv

Name

CAREGIVERS.csv

Description

Rows to process for schema

1000

Add column to show source file

Character encoding

UTF-8

Tags

search

You selected

Total Datasets: 1

Total Size: 198.7 KB

Delimited CAREGIVERS.csv 198.7 KB

Show preview Edit columns

	ROW_ID	CGID	LABEL	DESCRIPTION
1	2228	16174	RO	Read Only
2	2229	16175	RO	Read Only
3	2230	16176	Res	Resident/Fellow/PA...
4	2231	16177	RO	Read Only
5	2232	16178	RT	Respiratory
6	2233	16179	Res	Resident/Fellow/PA...
7	2234	16180	Rehab	Rehabilitation
8	2235	16181	MD	
9	2236	16182	CM	Case Manager
10	2237	16183	Rehab	Rehabilitation
11	2238	16184	RN	RN
12	2239	16185	MD	Attending
13	2240	16186	RN	RN
14	2241	16187	RN	RN

Library

Datasets +

19 items in list

Export Logs

Latest

Last 7 days

Last 30 days

Yes

No

Yes

No

Mine

search

search

Data Source

Tags

Name

Type

Version #

Status

of Rows

CAREGIVERS.csv

CREATE PROJECT

EXPORT

MORE ACTIONS

1

completed

7,567

SHOW PREVIEW

The screenshot shows the DataRobot Datasets page. At the top, there's a navigation bar with 'DataRobot' and a search bar. Below that, a 'Datasets' section shows '19 items in list'. A table lists several datasets:

Name	Type	Version #	Status	# of Rows	Tags	Created
CAREGIVERS.csv AnswerSet for Project "caregivers" (version 1)	CSV	1	completed	7,567	+ add tags	a minute ago by kaboub.profo@emsi-edu.ma
cahrtventsFINAL AnswerSet for Project "cahrtvents final" (version 2)	CSV	2	completed	475,776	+ add tags	2 days ago by kaboub.profo@emsi-edu.ma
outputfinal AnswerSet for Project "outputfinal" (version 1)	CSV	1	completed	498,477	+ add tags	3 days ago by kaboub.profo@emsi-edu.ma
fourthprescriptionwithallcusstays AnswerSet for Project "fourth" (version 12)	CSV	1	completed	323,914	+ add tags	3 days ago by kaboub.profo@emsi-edu.ma
fifththelayoutand patients AnswerSet for Project "fifth" (version 7 at Step 6)	CSV	1	completed	58,148	+ add tags	3 days ago by kaboub.profo@emsi-edu.ma
secondDataset AnswerSet for Project "second" (version 3)	CSV	2	completed	80,183	+ add tags	3 days ago by kaboub.profo@emsi-edu.ma
PRESCRIPTIONS.csv	CSV	1	completed	500,000	+ add tags	by kaboub

234 PM 5/12/2021

Projects CareGiversBasedProject

Tools

Steps

Remove Rows

Remove Rows

Click to annotate

Start with CAREGIVERS.csv

5 COLUMNS - 7,567 ROWS AT STEP 2

REFRESH DATASETS VIEW SCRIPT

Remove rows

Sources	ROW_ID	CGID	LABEL	DESCRIPTION
1	2226	16174	RD	
2	2229	16175	RD	
3	2230	16176	Res	
4	2231	16177	RD	
5	2232	16178	RT	
6	2233	16179	Res	
7	2234	16180	Reha	
8	2235	16181	MD	
9	2236	16182	CM	
10	2237	16183	Reha	
11	2238	16184	RN	
12	2239	16185	MD	
13	2240	16186	RN	
14	2241	16187	RN	
15	2242	16188	SW	Social Worker
16	2243	16189	RO	Read Only
17	2244	16190	RO	Read Only
18	2245	16191	MD	
19	2246	16192	RD	Read Only
20	2247	16193	Res	Resident/Fellow/PA/NP
21	2248	16194	Res	Resident/Fellow/PA/NP
22	2249	16195	RD	Read Only
23	2250	16196	RPH	Pharmacist
24	2251	16197	CM	Case Manager
25	2252	16198	MDs	Resident/Fellow/PA/NP
26	2253	16199	MD	
27	2254	16200	RN	RN
28	2255	16201	Res	Resident/Fellow/PA/NP
29	2256	16202	NP	Read Only
30	2257	16203	RO	Read Only
31	2258	16204	Res	Resident/Fellow/PA/NP

Filters

values

sort by ascending

by descending

change into

columns split

find + replace

duplicate

rename

fill

whitespace trim leading and trailing

collapse consecutive

group cluster + edit

235 PM 5/12/2021

Projects CareGiversBasedProject

Tools

Steps

Remove Rows

Click to annotate

Start with CAREGIVERS.csv

5 COLUMNS - 2,412 ROWS AT STEP 2

REFRESH DATASETS VIEW SCRIPT

Filters for Remove Rows

DESCRIPTION

Show 1 Selected Items

Read Only 2,412

Resident/Fellow/PA/NP 1,544

RN 600

Read Only 199

Rehabilitation 121

RD/NA 121

Pharmacist 45

Physician

BLANK VALUES

Clear all blanks from "Selected" items

17 UNIQUE VALUES IN 1

BUILD BLANK

Sources	ROW_ID	CGID	LABEL	DESCRIPTION
1	2235	16181	MD	
2	2245	16191	MD	
3	2253	16199	MD	
4	2260	16206	MD	
5	2261	16207	MD	
6	2262	16208	MD	
7	2272	16218	SRT	
8	2279	16225	MD	
9	2284	16230	MD	
10	2289	16235	MD	
11	2303	16249	Cowker	
12	2306	16252	UCO	
13	2307	16253	MD	
14	2328	16274	RRT	
15	2331	16277	MD	
16	2332	16278	MD	
17	2335	16281	MDs	

DataRobot | k8s-trial.paxata.com/domain/drauthy/#/project/72eb92aefdb44a87973c459dd498c39c/ds-72eb92aefdb44a87973c459dd498c39c/steps

Projects: CareGiversBasedProject

Steps | **Compute** | **Aggregate** | **on** **ROW_ID** | **as** **New Column**

Aggregate: New Column as Average(ROW_ID) | Group by: | Compute: Average | Starting from: top of window | Ending at: bottom of window

Remove Rows: | Start with CAREGIVERS.csv

Sources	ROW_ID	New Column	CGID	LABEL	DESCRIPTION
1	2228		16174	RO	Read Only
2	2229		16175	RO	Read Only
3	2230		16176	Res	Resident/Fellow/PA/NP
4	2231		16177	RO	Read Only
5	2232		16178	RT	Respiratory
6	2233		16179	Res	Resident/Fellow/PA/NP
7	2234		16180	Rehab	Rehabilitation
8	2235		16182	CM	Case Manager
9	2237		16183	Rehab	Rehabilitation
10	2238		16184	RN	RN
11	2239		16185	MD	Attending
12	2240		16186	RN	RN
13	2241		16187	RN	RN
14	2242		16188	SW	Social Worker
15	2243		16189	RO	Read Only
16	2244		16190	RO	Read Only
17	2245		16192	RO	Read Only
18	2247		16193	Res	Resident/Fellow/PA/NP
19	2248		16194	Res	Resident/Fellow/PA/NP
20	2249		16195	RO	Read Only
21	2250		16196	RPH	Pharmacist
22	2251		16197	CM	Case Manager

6 COLUMNS - 5,155 ROWS AT STEP 3 | REFRESH DATASETS | VIEW SCRIPT

DataRobot | k8s-trial.paxata.com/domain/drauthy/#/project/72eb92aefdb44a87973c459dd498c39c/ds-72eb92aefdb44a87973c459dd498c39c/steps

Projects: CareGiversBasedProject

Steps | **Standard** | **Cross Product** | **Join Using** | **PATIENTS.csv**

CONNECTIONS | OPTIONS | Select connections

selected connection: ROW_ID | PATIENTS.csv

DESCRIPTION	JOIN (LEFT)	JOIN (RIGHT)	ROW_ID (1)	SUBJECT_ID	GENDER	DOB	DOB_HOSP
Read Only	2228	2228	2228	2352	F	2134-12-06 00:00:00	
Read Only	2229	2229	2229	2353	M	2091-03-31 00:00:00	2162-08-10 00:00:00
Resident/Fellow/PA/NP	2230	2230	2230	2354	M	2148-11-01 00:00:00	
Read Only	2231	2231	2231	2355	M	2103-01-07 00:00:00	2158-01-29 00:00:00
Respiratory	2232	2232	2232	2356	F	2087-05-03 00:00:00	
Resident/Fellow/PA/NP	2233	2233	2233	2357	F	2121-05-14 00:00:00	
Rehabilitation	2234	2234	2234	2358	F	2039-07-20 00:00:00	2124-09-17 00:00:00
Case Manager	2235	2235	2235	2360	F	2115-02-27 00:00:00	
Rehabilitation	2237	2237	2237	2361	M	2036-10-08 00:00:00	2108-08-23 00:00:00
RN	2238	2238	2238	2362	M	2072-09-07 00:00:00	2146-03-29 00:00:00
Attending	2239	2239	2239	2363	M	2070-08-24 00:00:00	
RN	2240	2240	2240	2364	F	1867-05-24 00:00:00	2167-07-06 00:00:00
RN	2241	2241	2241	2365	F	2130-11-15 00:00:00	2183-03-10 00:00:00
Social Worker	2242	2242	2242	2366	F	2113-08-18 00:00:00	2150-12-06 00:00:00
Read Only	2243	2243	2243	2367	F	2171-03-27 00:00:00	

15 COLUMNS - 5,155 ROWS AT STEP 3 | REFRESH DATASETS | VIEW SCRIPT

Projects: CareGiversBasedProject

Tools: Steps

Standard Cross Product Join Using PATIENTS.csv

CONNECTIONS OPTIONS

Click the fields in the CONNECTIONS tab and select the columns to perform your Join operation. (change connections)

Lookup type

Keep only matched rows

Intelligence Matching Method

Exact Automatic

DESCRIPTION	JOIN (LEFT) ROW_ID	JOIN (RIGHT) ROW_ID (1)	ROW_ID (1)	SUBJECT_ID	GENDER	DOB	DOD	DOD_HOSP
Read Only	2228	2228	2228	2352	F	2134-12-06 00:00:00		
Read Only	2229	2229	2229	2353	M	2091-03-31 00:00:00	2162-08-10 00:00:00	
Resident/Fellow/PA/NP	2230	2230	2230	2354	M	2140-11-01 00:00:00		
Read Only	2231	2231	2231	2355	M	2103-03-07 00:00:00	2158-01-29 00:00:00	2158-01-29 00:00:00
Respiratory	2232	2232	2232	2356	F	2087-05-03 00:00:00		
Resident/Fellow/PA/NP	2233	2233	2233	2357	F	2121-05-14 00:00:00		
Rehabilitation	2234	2234	2234	2358	F	2039-07-20 00:00:00	2124-09-17 00:00:00	
Case Manager	2236	2236	2236	2360	E	2115-02-27 00:00:00		
Rehabilitation	2237	2237	2237	2361	M	2036-10-08 00:00:00	2108-08-23 00:00:00	2108-08-23 00:00:00
RN	2238	2238	2238	2362	M	2072-09-07 00:00:00	2146-03-29 00:00:00	2146-03-29 00:00:00
Attending	2239	2239	2239	2363	M	2070-08-24 00:00:00		
RN	2240	2240	2240	2364	F	1867-05-24 00:00:00	2167-07-06 00:00:00	
RN	2241	2241	2241	2365	F	2130-11-15 00:00:00	2183-03-10 00:00:00	2183-03-10 00:00:00
Social Worker	2242	2242	2242	2366	F	2113-08-18 00:00:00	2150-12-06 00:00:00	
Read Only	2243	2243	2243	2367	F	2171-03-27 00:00:00		

15 COLUMNS • 5,155 ROWS AT STEP 3

REFRESH DATASETS VIEW SCRIPT

Projects: CareGiversBasedProject

Tools: Steps

Compute Average on ROW_ID as New Column

Window Grouped by: group by column(s)

Window Sorted by: sort column(s)

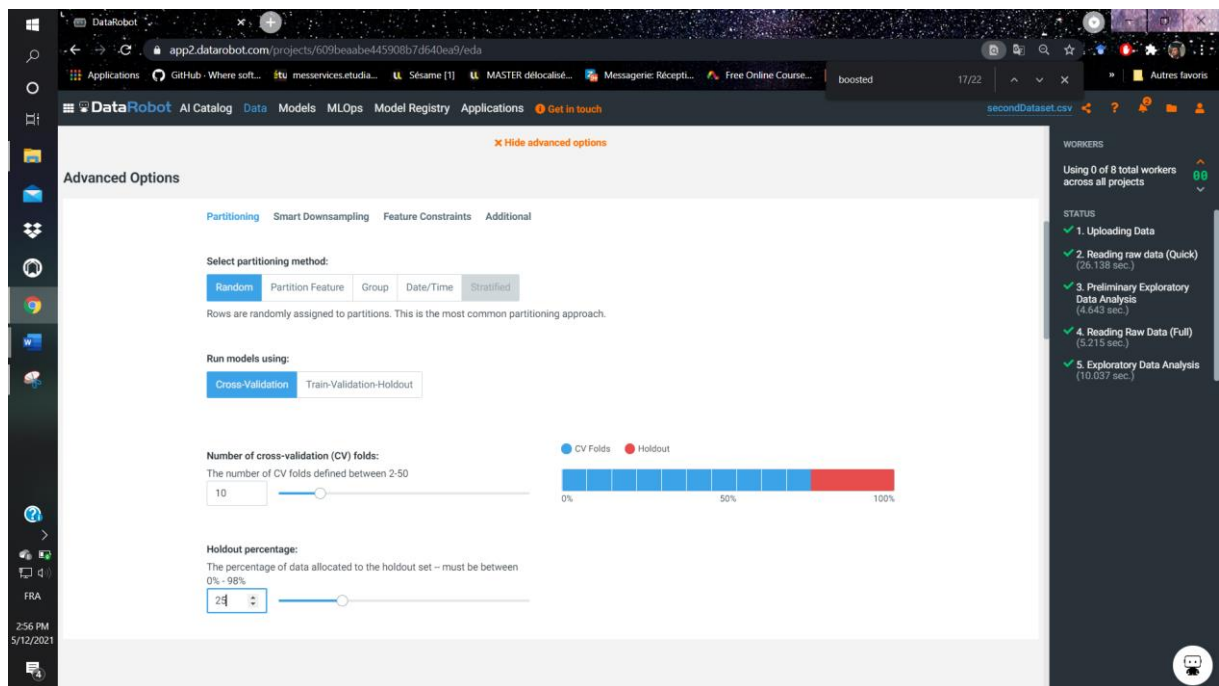
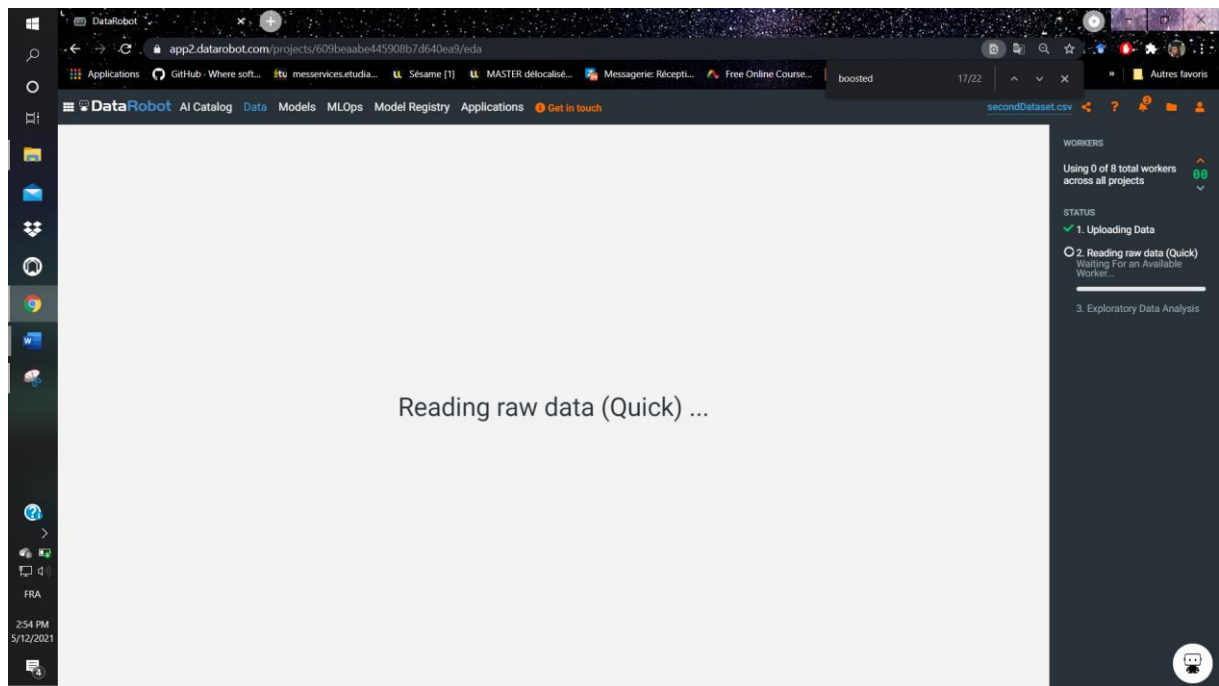
Starting from: top of window

Ending at: bottom of window

DESCRIPTION	ROW_ID (1)	SUBJECT_ID	GENDER	DOB	DOD	DOD_HOSP	DOD_SSN	EXPIRE_FLAG
nly	2228	2352	F	2134-12-06 00:00:00				0
nly	2229	2353	M	2091-03-31 00:00:00	2162-08-10 00:00:00		2162-08-10 00:00:00	1
Resident/Fellow/PA/NP	2230	2354	M	2140-11-01 00:00:00				0
nly	2231	2355	M	2103-03-07 00:00:00	2158-01-29 00:00:00	2158-01-29 00:00:00	2158-01-29 00:00:00	1
tory	2232	2356	F	2087-05-03 00:00:00				0
Resident/Fellow/PA/NP	2233	2357	F	2121-05-14 00:00:00				0
Rehabilitation	2234	2358	F	2039-07-20 00:00:00	2124-09-17 00:00:00		2124-09-17 00:00:00	1
anager	2236	2360	F	2115-02-27 00:00:00				0
Rehabilitation	2237	2361	M	2036-10-08 00:00:00	2108-08-23 00:00:00	2108-08-23 00:00:00		1
Rehabilitation	2238	2362	M	2072-09-07 00:00:00	2146-03-29 00:00:00	2146-03-29 00:00:00	2146-03-29 00:00:00	1
Rehabilitation	2239	2363	M	2070-08-24 00:00:00				0
Rehabilitation	2240	2364	F	1867-05-24 00:00:00	2167-07-06 00:00:00		2167-07-06 00:00:00	1
Rehabilitation	2241	2365	F	2130-11-15 00:00:00	2183-03-10 00:00:00	2183-03-10 00:00:00		1
Rehabilitation	2242	2366	F	2113-08-18 00:00:00	2150-12-06 00:00:00		2150-12-06 00:00:00	1
Rehabilitation	2243	2367	F	2171-03-27 00:00:00				0
Rehabilitation	2244	2368	M	2068-02-22 00:00:00	2138-03-10 00:00:00	2138-03-10 00:00:00	2138-03-10 00:00:00	1
Rehabilitation	2246	2370	F	2084-11-27 00:00:00	2163-01-08 00:00:00	2163-01-08 00:00:00	2163-01-08 00:00:00	1
Resident/Fellow/PA/NP	2247	2371	F	2138-07-13 00:00:00				0
Resident/Fellow/PA/NP	2248	2372	M	2107-02-20 00:00:00				0
nly	2249	2373	M	2117-11-14 00:00:00				0
Rehabilitation	2250	2374	F	2140-05-25 00:00:00				0
anager	2251	2375	M	2054-11-15 00:00:00				0

14 COLUMNS • 5,155 ROWS AT STEP 4

REFRESH DATASETS VIEW SCRIPT



app2.datarobot.com/projects/609beaabe445908b7d540ea9/eda

DataRobot AI Catalog Data Models MLOps Model Registry Applications Get in touch

Data Quality

Project dataset: secondDataset.csv Features: 45 Datapoints: 80183 Initial downsampling: None [Explore the data](#) [Data Quality Assessment](#) [View info](#)

Project Data Feature Lists

Menu Search Feature List: All Features View Raw Data Create feature list

Feature Name	Data Quality	Index	Var Type	Unique	Missing	Mean	Std Dev	Median	Min	Max
ROW_ID		1	Numeric	58,976	0	29,952	17,082	30,371	1	58,976
SUBJECT_ID		2	Numeric	46,520	0	34,537	28,315	24,872	2	99,999
HADM_ID		3	Numeric	58,976	0	149,914	28,887	149,872	100,001	199,999
ADMITTIME		4	Date	58,651	0	2151-03-16...	10633.37 d...	2151-02-03...	2100-06-07...	2210-08-17...
DISCHTIME		5	Date	58,657	0	2151-03-29...	10633.6 da...	2151-02-23...	2100-06-09...	2210-08-24...
DEATHTIME		6	Date	5,834	71,069	2151-05-15...	10572.802...	2151-04-12...	2100-06-19...	2208-02-05...
ADMISSION_TYPE		7	Categorical	4	0					
ADMISSION_LOCATION		8	Categorical	9	0					
DISCHARGE_LOCATION		9	Categorical	17	0					
INSURANCE		10	Categorical	5	0					

WORKERS

Using 0 of 8 total workers across all projects

STATUS

- 1. Uploading Data
- 2. Reading raw data (Quick) (26.138 sec.)
- 3. Preliminary Exploratory Data Analysis (4.643 sec.)
- 4. Reading Raw Data (Full) (5.215 sec.)
- 5. Exploratory Data Analysis (10.037 sec.)

257 PM 5/12/2021

app2.datarobot.com/projects/609beaabe445908b7d540ea9/eda

DataRobot AI Catalog Data Models MLOps Model Registry Applications Get in touch

Data Quality

Project dataset: secondDataset.csv Features: 45 Datapoints: 80183 Initial downsampling: None [Explore the data](#) [Data Quality Assessment](#) [View info](#)

Project Data Feature Lists

Menu Search Feature List: All Features View Raw Data Create feature list

Feature Name	Data Quality	Index	Importance	Var Type	Unique	Missing	Mean	Std Dev	Median	Min	Max
LOS	TARGET	31	Target	Numeric	35,455	0	4.95	9.26	2.18	0.00	173
DISCHARGE_LOCATION		9		Categorical	16	0					
DIAGNOSIS		17		Text	13,250	1					
LAST_CAREUNIT		26		Categorical	6	0					
CURR_SERVICE		37		Categorical	20	15					
FIRST_CAREUNIT		25		Categorical	6	0					
LAST_WARDID		28		Numeric	17	0	32.53	18.94	33	7	57
ADMISSION_TYPE		7		Categorical	4	0					
FIRST_WARDID		27		Numeric	16	0	32.73	19.04	33	7	57
MARITAL_STATUS		13		Categorical	7	8,707					
LANGUAGE		11		Categorical	72	25,548					
DOB		41		Date	29,150	0	2085-01-28...	22535.82 d...	2092-03-05...	1800-07-02...	2201-07-24...
DBSOURCE		24		Categorical	3	0					
SUBJECT_ID		2		Numeric	39,766	0	34,586	28,318	24,939	3	99,999
SUBJECT_ID (1)		21		Numeric	39,766	0	34,586	28,318	24,939	3	99,999

WORKERS

Using 2 of 8 total workers across all projects

ACTIONS

- Get More Accuracy
- Configure modeling settings
- Unlock project Holdout for all models

PROCESSING

- Keras Slim Residual Neura... (BP44 + 16% + CV#1)
- TensorFlow Deep Learning... (BP46 + 16% + CV#1)
- Waiting for resources
- Obtaining position in line

IN QUEUE

- Gradient Boosted Trees Re... (BP45 + 16% + CV#1)
- Breiman and Cutler Rando... (BP49 + 16% + CV#1)

3:01 PM 5/12/2021