

IDENTIFY THE HOSPITAL PATIENTS WITH SEVERE INFECTION USING MEDICAL DATA

TASK 1:

Reading the antibioticDT dataset which includes the columns patient_id, day given, antibiotic type, and route.

patient_id	day_given	antibiotic_type	route
1	2	ciprofloxacin	IV
1	4	ciprofloxacin	IV
1	6	ciprofloxacin	IV
1	7	doxycycline	IV
1	9	doxycycline	IV
1	15	penicillin	IV
1	16	doxycycline	IV
1	18	ciprofloxacin	IV
8	1	doxycycline	PO
8	2	penicillin	IV
8	3	doxycycline	IV
8	6	doxycycline	PO
8	8	penicillin	PO
8	12	penicillin	IV
9	8	doxycycline	IV
9	12	doxycycline	PO
12	4	doxycycline	PO
12	9	doxycycline	IV
16	1	doxycycline	IV
16	4	amoxicillin	IV
19	3	doxycycline	PO
19	5	amoxicillin	IV

TASK 2:

A new variable called 'antibiotic_new' added to the data. Table that indicates if it has been only one or two days since the antibiotic was given. The data is now sorted according to patient_id, antibiotic_type and day_given.

patient_id	day_given	antibiotic_type	route
1	2	ciprofloxacin	IV
1	4	ciprofloxacin	IV
1	6	ciprofloxacin	IV
1	18	ciprofloxacin	IV
1	7	doxycycline	IV
1	9	doxycycline	IV
1	16	doxycycline	IV
1	15	penicillin	IV
8	1	doxycycline	PO
8	3	doxycycline	IV
8	6	doxycycline	PO
8	2	penicillin	IV

The new variable 'antibiotic_type' is added.

patient_id	day_given	antibiotic_type	route	last_administration_day	days_since_last_admin	antibiotic_new
1	2	ciprofloxacin	IV	NA	NA	1
1	4	ciprofloxacin	IV	2	2	0
1	6	ciprofloxacin	IV	4	2	0
1	18	ciprofloxacin	IV	6	12	1
1	7	doxycycline	IV	NA	NA	1
1	9	doxycycline	IV	7	2	0
1	16	doxycycline	IV	9	7	1
1	15	penicillin	IV	NA	NA	1
8	1	doxycycline	PO	NA	NA	1
8	3	doxycycline	IV	1	2	0
8	6	doxycycline	PO	3	3	1
8	2	penicillin	IV	NA	NA	1
8	8	penicillin	PO	2	6	1
8	12	penicillin	IV	8	4	1
9	8	doxycycline	IV	NA	NA	1
9	12	doxycycline	PO	8	4	1
12	4	doxycycline	PO	NA	NA	1
12	9	doxycycline	IV	4	5	1
16	4	amoxicillin	IV	NA	NA	1
16	1	doxycycline	IV	NA	NA	1
19	5	amoxicillin	IV	NA	NA	1
19	6	ciprofloxacin	IV	NA	NA	1

TASK 3:

Importing and Printing the first 30 rows of bloodcultureDT.

patient_id	blood_culture_day
1	3
1	13
8	2
8	13
23	3
39	10
45	4
45	9
45	11
51	3
51	6
59	2
64	1
66	9
66	10
69	2
69	6
69	7
69	11
69	16
76	1
77	3

TASK 4:

A combined data set of antibioticDT and bloodcultureDT sorted by patient_id, blood_culture_day, day_given, antibiotic_type.

```
In [4]: # Merge antibioticDT with blood_cultureDT
combinedDT <-merge(antibioticDT,blood_cultureDT,by.antibioticDT="patient_id",by.blood_cultureDT="patient_id")
#combinedDT

# Sort by patient_id, blood_culture_day, day_given, and antibiotic_type
# .... YOUR CODE FOR TASK 4 .....
setorder(combinedDT,patient_id,blood_culture_day,day_given,antibiotic_type)
#combinedDT

# Print and examine the first 30 rows
# .... YOUR CODE FOR TASK 4 .....
combinedDT[1:30]
```

patient_id	day_given	antibiotic_type	route	last_administration_day	days_since_last_admin	antibiotic_new	blood_culture_day
1	2	ciprofloxacin	IV	NA	NA	1	3
1	4	ciprofloxacin	IV	2	2	0	3
1	6	ciprofloxacin	IV	4	2	0	3
1	7	doxycycline	IV	NA	NA	1	3
1	9	doxycycline	IV	7	2	0	3
1	15	penicillin	IV	NA	NA	1	3
1	16	doxycycline	IV	9	7	1	3
1	18	ciprofloxacin	IV	6	12	1	3
1	2	ciprofloxacin	IV	NA	NA	1	13
1	4	ciprofloxacin	IV	2	2	0	13
1	6	ciprofloxacin	IV	4	2	0	13
1	7	doxycycline	IV	NA	NA	1	13
1	9	doxycycline	IV	7	2	0	13
1	15	penicillin	IV	NA	NA	1	13
1	16	doxycycline	IV	9	7	1	13

TASK 5:

A new variable called 'drug_in_bcx_window' is created which is 1 if the drug was given in the two day window and 0 otherwise.

patient_id	day_given	antibiotic_type	route	last_administration_day	days_since_last_admin	antibiotic_new	blood_culture_day	drug_in_bcx_window
1	2	ciprofloxacin	IV	NA	NA	1	3	NA
1	4	ciprofloxacin	IV	2	2	0	3	1
1	6	ciprofloxacin	IV	4	2	0	3	1
1	7	doxycycline	IV	NA	NA	1	3	NA
1	9	doxycycline	IV	7	2	0	3	0
1	15	penicillin	IV	NA	NA	1	3	NA
1	16	doxycycline	IV	9	7	1	3	0
1	18	ciprofloxacin	IV	6	12	1	3	0
1	2	ciprofloxacin	IV	NA	NA	1	13	NA
1	4	ciprofloxacin	IV	2	2	0	13	1
1	6	ciprofloxacin	IV	4	2	0	13	1
1	7	doxycycline	IV	NA	NA	1	13	NA
1	9	doxycycline	IV	7	2	0	13	1
1	15	penicillin	IV	NA	NA	1	13	NA
1	16	doxycycline	IV	9	7	1	13	1
1	18	ciprofloxacin	IV	6	12	1	13	1
8	1	doxycycline	PO	NA	NA	1	2	NA
8	2	penicillin	IV	NA	NA	1	2	NA
8	3	doxycycline	IV	1	2	0	2	1
8	6	doxycycline	PO	3	3	1	2	1
8	8	penicillin	PO	2	6	1	2	1
8	12	penicillin	IV	8	4	1	2	0

TASK 6:

Created a new variable 'any_iv_in_bcx_window' indicating whether or not an I.V drug was given within a +/-2 day window of a blood culture day.

_type	route	last_administration_day	days_since_last_admin	antibiotic_new	blood_culture_day	drug_in_bcx_window	any_iv_in_bcx_window
xxacin	IV	2	2	0	3	1	TRUE
xxacin	IV	4	2	0	3	1	TRUE
xxacin	IV	2	2	0	13	1	TRUE
xxacin	IV	4	2	0	13	1	TRUE
ycline	IV	7	2	0	13	1	TRUE
ycline	IV	1	2	0	2	1	TRUE
ycline	IV	1	2	0	13	1	TRUE
ycline	IV	1	2	0	3	1	TRUE
ycline	IV	3	1	0	3	1	TRUE
xxacin	PO	3	2	0	3	1	TRUE
ycline	IV	4	1	0	3	1	TRUE
ycline	IV	5	1	0	3	1	TRUE
ycline	IV	1	2	0	4	1	TRUE
ycline	IV	3	2	0	4	1	TRUE
ycline	IV	1	2	0	9	1	TRUE
ycline	IV	3	2	0	9	1	TRUE
ycline	IV	1	2	0	11	1	TRUE
ycline	IV	3	2	0	11	1	TRUE
ycline	IV	2	1	0	9	1	TRUE
ycline	IV	2	1	0	10	1	TRUE
ycline	IV	3	2	0	1	1	TRUE
nicillin	PO	3	2	0	3	1	TRUE

TASK 7:

A new variable called 'day_of_first_new_abx_in_window' created indicating the first day of potential antibiotic 4 day sequences.

last_administration_day	days_since_last_admin	antibiotic_new	blood_culture_day	drug_in_bcx_window	any_iv_in_bcx_window	day_of_first_new_abx_in_window
2	2	0	3	1	TRUE	4
4	2	0	3	1	TRUE	4
2	2	0	13	1	TRUE	4
4	2	0	13	1	TRUE	4
7	2	0	13	1	TRUE	9
1	2	0	2	1	TRUE	3
1	2	0	13	1	TRUE	3
1	2	0	3	1	TRUE	3
3	1	0	3	1	TRUE	3
3	2	0	3	1	TRUE	5
4	1	0	3	1	TRUE	3
5	1	0	3	1	TRUE	3
1	2	0	4	1	TRUE	3
3	2	0	4	1	TRUE	3
1	2	0	9	1	TRUE	3
3	2	0	9	1	TRUE	3
1	2	0	11	1	TRUE	3
3	2	0	11	1	TRUE	3

TASK 8:

A new dataset 'Simplified_data' with only patient_id, blood_culture_day, and day_given from the above data.

combinedDT.patient_id	combinedDT.blood_culture_day	combinedDT.day_given
1	3	4
1	3	6
1	13	4
1	13	6
1	13	9
8	2	3
8	13	3
23	3	3
23	3	4
23	3	5
23	3	5
23	3	6

TASK 9:

A new variable 'num_antibiotic_days' added which indicates the number of each patient/blood culture day combinations. It consists of only blood culture days with more than or equal to four antibiotic days. The total no of rows extracted is 656.

combinedDT.blood_culture_day	combinedDT.patient_id	combinedDT.day_given	num_antibiotic_days
3	23	3	4
3	23	4	4
3	23	5	4
3	23	6	4
6	164	1	6
6	164	2	6
6	164	4	6
6	164	6	6
7	164	1	6
7	164	2	6
7	164	4	6
7	164	6	6
3	200	3	4
3	200	4	4
3	200	5	4
3	200	7	4
7	200	3	5
7	200	4	5

TASK 10:

A new variable 'four_in_seq' created which indicates the antibiotic sequence has no skips of more than one day.

combinedDT.blood_culture_day	combinedDT.patient_id	combinedDT.day_given	num_antibiotic_days	lagday	four_in_seq
3	23	3	4	NA	NA
3	23	4	4	3	1
3	23	5	4	4	1
3	23	6	4	5	1
6	164	1	6	NA	NA
6	164	2	6	1	1
6	164	4	6	2	0
6	164	6	6	4	0
7	164	1	6	NA	NA
7	164	2	6	1	1
7	164	4	6	2	0
7	164	6	6	4	0
3	200	3	4	NA	NA
3	200	4	4	3	1
3	200	5	4	4	1
3	200	7	4	5	0
7	200	3	5	NA	NA
7	200	4	5	3	1
7	200	5	5	4	1
7	200	7	5	5	0
11	206	3	8	NA	NA
11	206	4	8	3	1
11	206	6	8	4	0
11	206	7	8	6	1

TASK 11:

A new data frame with an infection indicator is created considering patient_id from previous data set where four_in_seq is equal to 1.

combinedDT.patient_id	infection
23	1
164	1
200	1
206	1
237	1
294	1
311	1
372	1
402	1
415	1
442	1
460	1
462	1
490	1
515	1
530	1
608	1
627	1
634	1
681	1
685	1
686	1
728	1
758	1

TASK 12:

The all patient's data is imported and examined. It has all patient id with a total of 890 patient_id.

With the calculation of total patients with infection, the count is 102.

Therefore it is clear that the percentage of people who met the criteria for presumed infection is 11.46 %

patient_id	infection
23	1
164	1
200	1
206	1
237	1
294	1
311	1
372	1
402	1
415	1
442	1
460	1
462	1
490	1
515	1
530	1
608	1
627	1
634	1
681	1
685	1
686	1