



UNIVERSIDAD
AUSTRAL | INGENIERÍA

Trabajo Práctico Cuatrimestral

Análisis de datos de MIMIC3 para la asistencia en la toma de decisiones

Docentes:

Ignacio Cassol, Nicolas Gamboa, Melina Juarez

Alumnos:

Juan Bautista Buthet, Valentina De la Torre

Algoritmos y Estructura de Datos

2023

INDICE

1. INTRODUCCION.....	2
2. DESARROLLO.....	2
3. PROCEDIMIENTOS.....	3
3.1 Pregunta 1.....	3
3.2 Pregunta 2.....	4
4. RESULTADOS.....	4
5. CONCLUSION.....	5
6. REFERENCIAS.....	6
7. ANEXOS.....	7

1. INTRODUCCION

Se utilizara la base de datos “MIMIC-III” abreviado de Medical Information Mart for Intensive Care, esta misma es una gran base de datos de libre acceso que comprende datos desidentificados relacionados con la salud de pacientes que permanecieron en unidades de cuidados críticos del Beth Israel Deaconess Medical Center entre 2001 y 2012. Precisamente trabajaremos con un subconjunto “demo”, que contiene información sobre 100 pacientes también presentes en la base de datos clínicos MIMIC-III.

MIMIC-III consta de 26 tablas; 6 para seguimiento de pacientes, 8 son de datos colectados en cuidados intensivos, 7 son de datos almacenados en el hospital y 5 son diccionarios

2. DESARROLLO

A partir de la base de datos de MIMIC-III surgen varias preguntas como ¿Cuáles son los diagnósticos más comunes entre los pacientes que fallecieron en la UCI? ¿Cuál es el promedio de readmisiones por pacientes en la UCI y cuáles son los diagnósticos más comunes de estas readmisiones?

Esperamos poder responder estas preguntas e inquietudes surgidas, a través de un completo análisis de la base de datos. Tras un amplio análisis y búsqueda sobre gran parte de los diagnósticos encontrados en la base de datos, nuestra expectativa es observar a la hipertension y a la pneumonia como los diagnósticos mas comunes entre los pacientes fallecidos. “La hipertensión es una de las causas principales de muerte prematura en el mundo.” (Organización Mundial de la Salud, Hipertensión, 16 de marzo de 2023) y el peligro de la pneumonia como enfermedad (véase anexo 1). Esto se debe no solo a su tasa de mortalidad, si no a su basta presencia en seres humanos. A su vez esperamos encontrar un promedio de readmision por paciente mayor o igual a 2. Esto se debe a que para tan solo 100 pacientes (MIMIC3.demo), encontramos una variedad de códigos tanto en las columnas “icd9_code” como en las columnas “icustay_id”. Pudiendo esto significar una entrada y salida (readmisiones) de mismos pacientes.

Es fundamental realizar una identificación de los diagnósticos más comunes entre los pacientes que fallecieron. Es importante para mejorar la atención y el tratamiento de los pacientes críticamente enfermos en la UCI. Gracias a una

correcta identificación de los diagnósticos mas severos es posible enfatizar en en el cuidado de pacientes con dicho diagnostico, conllevando a reducir la mortalidad y mejorar la calidad de vida de los pacientes.

Los administradores y/o directivos de la UCI pueden utilizar el promedio de readmision por paciente para determinar la calidad de la atención médica, lo que indica la eficacia de los tratamientos y los procesos de atención. Mediante la identificación de los diagnósticos más frecuentes a pacientes que reingresaron, los administradores y/o directivos pueden desarrollar estrategias efectivas para contrarrestar posibles errores, mejorar la calidad de atención y consecuentemente reducir una futura tasa de reingreso (“congestionamiento” de gente en la UCI).

3. PROCEDIMIENTOS

3.1

¿Cuáles son los diagnósticos más comunes entre los pacientes que fallecieron en la UCI?

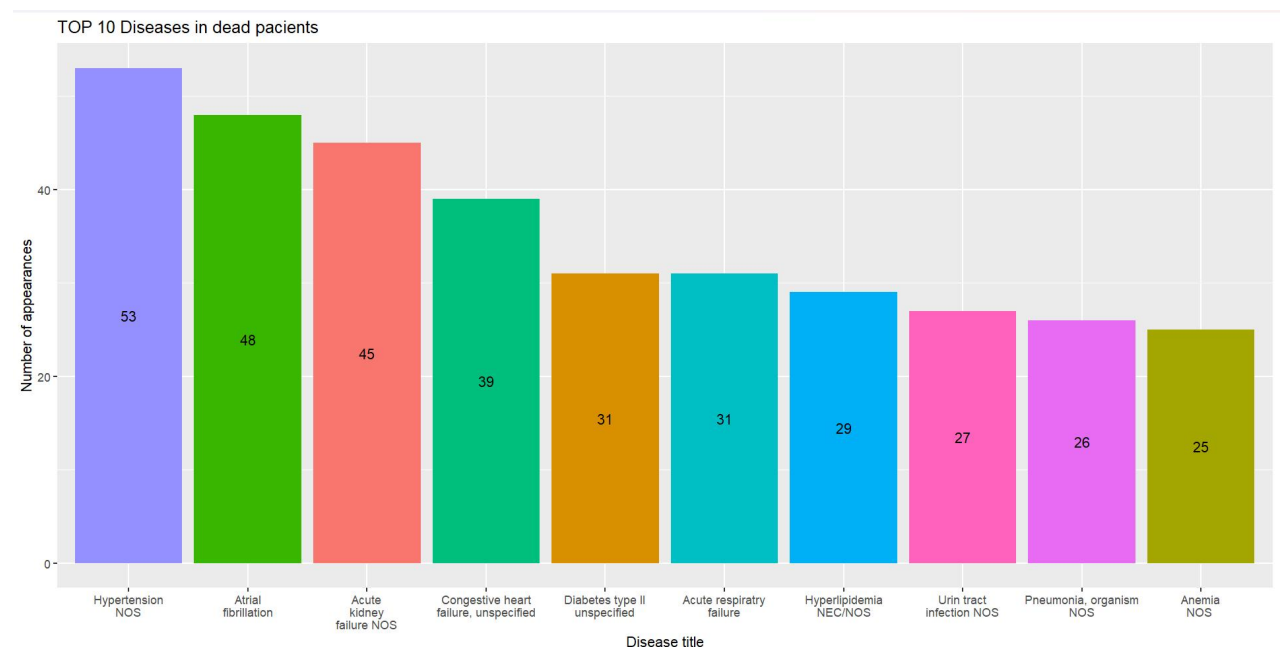


Gráfico 1: representación visual a la pregunta 1, ¿Cuáles son los diagnósticos más comunes entre los pacientes que fallecieron en la UCI? Ver anexo 2 y 4

Notamos en el Grafico 1 que hay mas de 100 diagnósticos totales si se suman todos. Esto se debe a que un paciente puede tener mas de un diagnostico.

3.2

¿Cuál es el promedio de readmisiones por pacientes en la UCI y cuáles son los diagnósticos más comunes de estas readmisiones?

```
> readmission_rate
# Source:   SQL [1 x 1]
# Database: sqlite 3.41.2 [C:\Users\bauti\Austra]
  readmission_rate
      <dbl>
1             1.36
```

Imagen 1: Promedio de readmisiones por paciente.

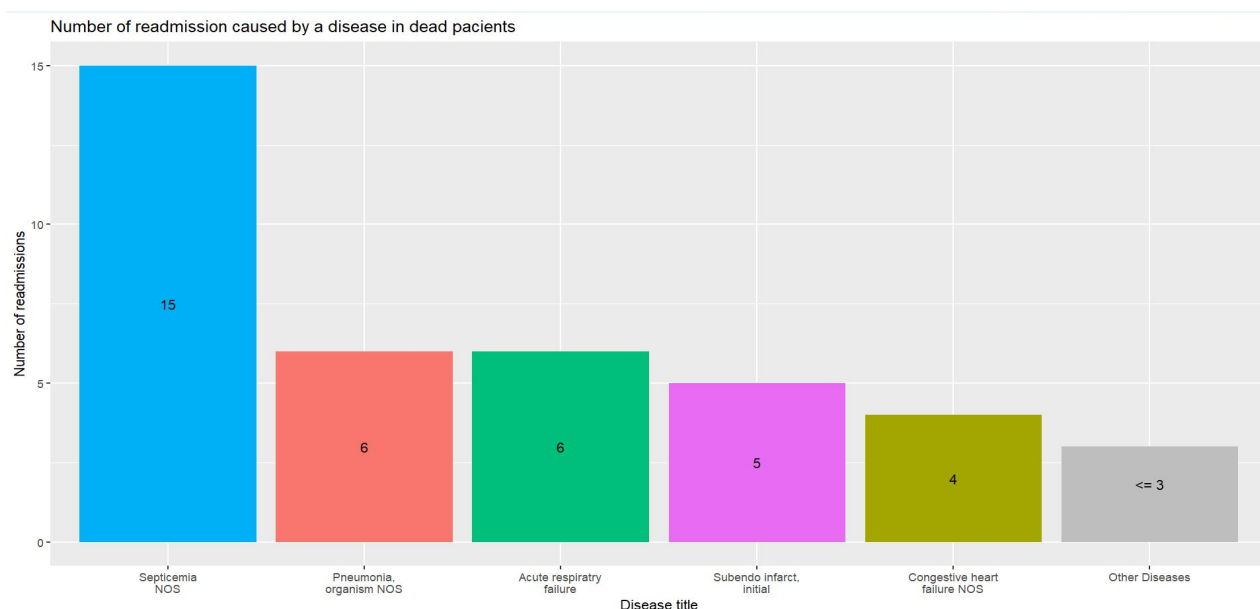


Gráfico 2: representación visual a la pregunta 2, cuáles son los diagnósticos más comunes de pacientes con readmisiones? Ver anexo 3 y 5

4. RESULTADOS

A través del Gráfico 1, vemos que la hipertensión NOS (NOS: no especificado) es el diagnóstico más frecuente en pacientes fallecidos en la UCI superando los 50 diagnósticos y al segundo diagnóstico más frecuente, atrial fibrillation¹, por 5.

El promedio de readmisión por paciente es 1.36, esto indica que de una muestra de 100 pacientes se estiman 136 readmisiones y por tanto significan 236 ingresos totales al hospital. Este resultado es significativo ya que deja en evidencia el elevado flujo de ingreso al hospital.

11% de las readmisiones al hospital son consecuencia de pacientes diagnosticados con Septicemia NOS. Esta información se muestra evidente en el Gráfico 2.

Los resultados son significativos debido a la cantidad de sujetos que había en las agrupaciones realizadas.

¹ <https://www.mayoclinic.org/es-es/diseases-conditions/atrial-fibrillation/symptoms-causes/syc-20350624>

5. CONCLUSIONES

Aunque lo esperado era ver a la hipertension y a la pneumonia como las enfermedades mas diagnosticadas en pacientes fallecidos y si bien la hipertension se presento de manera acertada como la enfermedad mas diagnosticada, la segunda enfermedad mas diagnosticada no fue la pneumonia, sino que fue la atrial fibrillation (ver Grafico 1). Ambas representando en conjunto un total de 101 diagnosticos, superando por una unidad al total de la muestra de pacientes (100 pacientes). Le encontramos respuesta a este resultado en las características puntuales de la hipertension y la pneumonia tanto en sus causas como en sus implicaciones. A su vez, justificamos el numero de diagnosticos a través del dato de que un mismo paciente puede tener mas de un único diagnostico.

Por otro lado, si bien esperábamos un promedio de readmision por paciente mayor o igual a 2, la misma fue de 1.36 (vease Imagen 1). Si bien no es 2, la misma se considera elevada. Un alto porcentaje podría indicar que la atención es de baja calidad e ineficaz.

Por consiguiente, en el Grafico 2 encontramos que de las readmisiones al hospital, el 11% fueron realizadas por pacientes diagnosticados con septicemia². A su vez, notamos la importante diferencia con la segunda enfermedad, pneumonia, siendo la misma diagnosticada en pacientes 9 readmisiones por debajo que la septicemia. Justificamos este resultado basándonos en la gravedad de los síntomas de esta enfermedad (septicemia) y en su frecuencia inusual. Esto se debe a que algunos de sus síntomas, como confusión o respiración rápida, requieren atención médica de emergencia. Para agregar, tan solo 5 enfermedades son causantes de mas de 3 readmisiones al hospital. Esto ultimo afirma la relevancia de este estudio, de esta información, ya que se deduce que el caudal de readmisiones a la UCI no esta ligado a todas las enfermedades sino que a unas pocas.

² <https://www.mayoclinic.org/es-es/diseases-conditions/sepsis/symptoms-causes/syc-20351214#Descripci%C3%B3n%20General>

REFERENCIAS

1. <https://physionet.org/content/mimiciii-demo/1.4/>
2. <https://mimic.mit.edu/docs/iii/tables/patients/>
3. <https://www.who.int/es/news-room/fact-sheets/detail/hypertension>
4. <https://elpais.com/planeta-futuro/red-de-expertos/2021-11-23/pongamos-fin-a-la-pandemia-permanente-de-enfermedades-respiratorias.html>
5. <https://www.mayoclinic.org/es-es/diseases-conditions/atrial-fibrillation/symptoms-causes/syc-20350624>
6. <https://www.mayoclinic.org/es-es/diseases-conditions/sepsis/symptoms-causes/syc-20351214#Descripci%C3%B3n%20General>

ANEXOS

Anexo 1

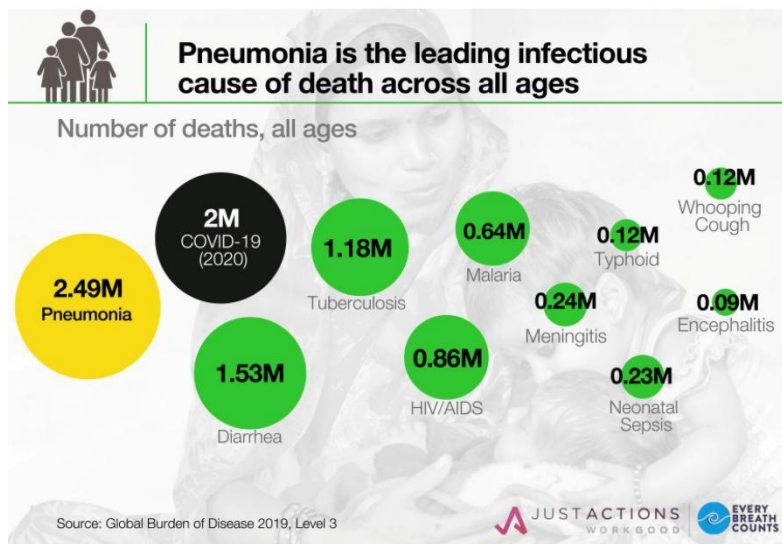


Imagen 2: La/el importancia/peligro de la neumonía como enfermedad

Anexo 2

Obtención de los datos para la pregunta 3.1

```
> # ¿Cuáles son los diagnósticos más comunes entre los pacientes que fallecieron en la UCI?
> deaths <- tbl(mimic_con, "PATIENTS") %>% filter(!is.na(dod))
> diagnoses <- tbl(mimic_con, "DIAGNOSES_ICD") %>% inner_join(tbl(mimic_con, "D_ICD_DIAGNOSES"), by = "icd9_code") %>%
+   select(subject_id, icd9_code, short_title)
> deaths_with_diagnoses <- deaths %>% inner_join(diagnoses, by = "subject_id") %>% select(icd9_code, short_title)
> # Contar los diagnósticos más comunes
> top_diagnoses <- deaths_with_diagnoses %>% group_by(icd9_code, short_title) %>% summarise(count = n()) %>% arrange(desc(count)) %>% head(10)
```

Imagen 3: Código empleado para recolectar y filtrar la información correspondiente a los diagnósticos mas frecuentes entre fallecidos.

```
> top_diagnoses
`summarise()` has grouped output by "icd9_code". You can override
# Source:   SQL [10 x 3]
# Database: sqlite 3.41.2 [C:\Users\bauti\Austral\2023_01\AyED\]
# Groups:   icd9_code
# Ordered by: desc(count)
  icd9_code short_title      count
  <chr>      <chr>          <int>
1 4019      Hypertension NOS        53
2 42731     Atrial fibrillation      48
3 5849      Acute kidney failure NOS 45
4 4280      CHF NOS                  39
5 25000     DMII wo cmp nt st uncnt 31
6 51881     Acute respiratory failure 31
7 2724      Hyperlipidemia NEC/NOS   29
8 5990      Urin tract infection NOS  27
9 486       Pneumonia, organism NOS   26
10 2859     Anemia NOS              25
```

Imagen 4: Tabla indicando los 10 diagnósticos mas frecuentes en pacientes fallecidos en la UCI.

Anexo 3

Obtencion de los datos para la pregunta 3.2

```
> #####
> icu_stays <- tbl(mimic_con, "ICUSTAYS")
> diagnoses_icd <- tbl(mimic_con, "DIAGNOSES_ICD")
> d_icd_diagnoses <- tbl(mimic_con, "D_ICD_DIAGNOSES")
> icd9_shortnames <- d_icd_diagnoses %>% inner_join(diagnoses_icd, by = "icd9_code") %>% select(subject_id, hadm_id, icd9_code, short_title, seq_num)
> n_admissions <- icu_stays %>% group_by(subject_id) %>%
+   summarize(n_admissions = n_distinct(icustay_id))
> # calcular el promedio de readmisiones por paciente
> readmission_rate <- n_admissions %>% summarize(readmission_rate = mean(n_admissions, na.rm = TRUE))
```

Imagen 5: Código empleado para recolectar y filtrar información correspondiente al promedio de readmisiones por paciente.

```
> # Obtener las causas más comunes de readmisión con el nombre corto del diagnóstico
> readmissions <- icu_stays %>% group_by(subject_id) %>%
+   summarize(n_admissions = n_distinct(icustay_id)) %>%
+   filter(n_admissions > 0) %>%
+   left_join(icd9_shortnames, by = "subject_id") %>%
+   filter(seq_num == 1) %>% select(icd9_code, short_title) %>%
+   count(icd9_code, short_title, sort = TRUE)
```

Imagen 6: Código empleado para recolectar y filtrar información correspondiente a los diagnósticos mas frecuentes en readmisiones de pacientes.

```
> readmissions
# Source:      SQL [?? x 3]
# Database:    sqlite 3.41.2 [C:\Users\bauti\Austral\2023_01\AyED\Informe
# Ordered by: desc(n)
   icd9_code short_title          n
   <int>    <chr>              <int>
1      389  Septicemia NOS        15
2      486  Pneumonia, organism NOS    6
3     51881 Acute respiratory failure    6
4     41071 Subendo infarct, initial    5
5      4280 CHF NOS                4
6      570  Acute necrosis of liver     3
7      570  Erythema infectiosum        3
8     1510  Mal neo stomach cardia      3
9     1510  TB of hip-unspec           3
10     5715 Cirrhosis of liver NOS      3
# i more rows
# i Use `print(n = ...) ` to see more rows
```

Imagen 7: Tabla indicando los 10 diagnósticos mas frecuentes en readmision de pacientes.

Anexo 4

Grafico 1

```
top_diagnoses %>% collect() %>% ggplot(aes(x = reorder(short_title, -count), y = count)) +
  geom_bar(aes(fill=short_title), stat = "identity") +
  labs(title = "TOP 10 Diseases in dead patients", x = "Disease title", y = "Number of appearances") +
  scale_x_discrete(labels=c("Hypertension\nNOS", "Atrial\nfibrillation", "Acute\nkidney\nfailure NOS", "Congestive heart\nfailure,
  unspecified", "Diabetes type II\nunspecified", "Acute respiratory\nfailure",
  "Hyperlipidemia\nNEC/NOS", "Urin tract\ninfection NOS", "Pneumonia, organism\nNOS", "Anemia\nNOS ")) +
  theme(legend.position = "none") + geom_text(aes(label = count), position = position_stack(vjust = .5))
```

Imagen 8: Código empleado para realizar el grafico 1.

Anexo 5

Grafico 2

```
readmissions %>%
  collect() %>%
  mutate(n = ifelse(n >= 4 & n <= 15, as.character(n), "<= 3")) %>%
  mutate(short_title = ifelse(n == "<= 3", "Other Diseases", short_title)) %>%
  ggplot(aes(x = reorder(short_title, -as.numeric(n)), y = as.numeric(n))) +
  geom_bar(aes(fill = short_title), stat = "identity") +
  geom_text(aes(label = ifelse(n == "<= 3", "<= 3", n)), position = position_stack(vjust = 0.5)) +
  geom_bar(data = data.frame(short_title = "Other Diseases", y = 3),
    fill = "gray", stat = "identity", width = 0.9, aes(x = short_title, y = y)) +
  geom_text(data = data.frame(short_title = "Other Diseases", y = 1.5),
    aes(x = short_title, y = y, label = "<= 3"), vjust = -0.5) +
  scale_x_discrete(labels=c("Septicemia\nNOS", "Pneumonia,\norganism NOS", "Acute respiratory\nfailure", "Subendocardial\ninfarct,\ninitial",
    "Congestive heart\nfailure NOS", "Other Diseases")) +
  labs(title = "Number of readmission caused by a disease in dead patients", x = "Disease title", y = "Number of readmissions") +
  theme(legend.position = "none")
```

Imagen 9: Código empleado para realizar el grafico 2.