

Formateado 2

Pablo Parrilla Cañadas

2025-03-15

Librerías

```
library(readxl)
library(readr)
library(dplyr)
```

```
##
## Adjuntando el paquete: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(lubridate)
```

```
##
## Adjuntando el paquete: 'lubridate'

## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union
```

Carga del conjunto de datos

```
dataset=read.csv("../Datos/Iniciales/dataset_2024.csv",sep = ";")
asignaturas=read.csv("../Datos/Iniciales/asignaturas_2024.csv",sep=";")
titulaciones=read.csv("../Datos/Iniciales/titulaciones_2024.csv",sep=";")
estudiantes=read.csv("../Datos/Iniciales/estudiantes_2024_rev.csv",sep=";")
head(dataset)
```

```
##      dni_hash      tit_hash      asi_hash  anyo_ingreso  tipo_ingreso  nota10
## 1 b54d431164a8 4a8596a7790b 03da56baee66         2017          NTE    7,691
## 2 b54d431164a8 4a8596a7790b 3082f0ff052f         2017          NTE    7,691
## 3 b54d431164a8 4a8596a7790b 32a097cbd16d         2017          NTE    7,691
## 4 b54d431164a8 4a8596a7790b d1050bc418f1         2017          NTE    7,691
```

```

## 5 b54d431164a8 4a8596a7790b d6e3ab95be72          2017          NTE  7,691
## 6 b54d431164a8 4a8596a7790b 000b07fd81c1          2017          NTE  7,691
##  nota14 campus estudios_p estudios_m dedicacion desplazado discapacidad
## 1 10.371      V          5          4          TC          1          NA
## 2 10.371      V          5          4          TC          1          NA
## 3 10.371      V          5          4          TC          1          NA
## 4 10.371      V          5          4          TC          1          NA
## 5 10.371      V          5          4          TC          1          NA
## 6 10.371      V          5          4          TC          1          NA
##  abandono becado preferencia_seleccion baja_fecha caca
## 1      NA      0          1          2024
## 2      NA      0          1          2024
## 3      NA      0          1          2024
## 4      NA      0          1          2024
## 5      NA      0          1          2024
## 6      NA      0          1          2024
##                                grupos_por_tipocredito matricula_activa
## 1 {'PA': 1552185, 'PC': 1552186, 'PL': 1552184, 'TA': 1552183}          1
## 2                                {'PL': 1552090, 'TA': 1552089}          1
## 3                                                                1
## 4                                                                1
## 5                                                                1
## 6                                {'PA': 1551995, 'PL': 1698319, 'TA': 1551989}          1
##  nota_asig          fecha_datos cod_centro curso_mas_bajo curso_mas_alto
## 1      2025-01-31 14:42:43          C          2          4
## 2      2025-01-31 14:42:43          C          2          4
## 3      2025-01-31 14:42:43          C          2          4
## 4      2025-01-31 14:42:43          C          2          4
## 5      0,0 2025-01-31 14:42:43          C          2          4
## 6      0,0 2025-01-31 14:42:43          C          2          4
##  cred_mat1 cred_mat2 cred_mat3 cred_mat4 cred_mat5 cred_mat6 cred_sup_normal
## 1      0,0      9,0      18,0      46,5      0,0      0      0,0
## 2      0,0      9,0      18,0      46,5      0,0      0      0,0
## 3      0,0      9,0      18,0      46,5      0,0      0      0,0
## 4      0,0      9,0      18,0      46,5      0,0      0      0,0
## 5      0,0      9,0      18,0      46,5      0,0      0      0,0
## 6      0,0      9,0      18,0      46,5      0,0      0      0,0
##  cred_sup_espec cred_sup cred_mat_normal cred_mat_movilidad cred_ptes_acta
## 1      0,0      0,0          73,5          0,0          61,5
## 2      0,0      0,0          73,5          0,0          61,5
## 3      0,0      0,0          73,5          0,0          61,5
## 4      0,0      0,0          73,5          0,0          61,5
## 5      0,0      0,0          73,5          0,0          61,5
## 6      0,0      0,0          73,5          0,0          61,5
##  cred_mat_practicas cred_mat_sem_a cred_mat_sem_b cred_mat_anu cred_mat_total
## 1          0,0          30,0          31,5          12,0          73,5
## 2          0,0          30,0          31,5          12,0          73,5
## 3          0,0          30,0          31,5          12,0          73,5
## 4          0,0          30,0          31,5          12,0          73,5
## 5          0,0          30,0          31,5          12,0          73,5
## 6          0,0          30,0          31,5          12,0          73,5
##  cred_sup_sem_a cred_sup_sem_b cred_sup_anu cred_sup_total rendimiento_cuat_a
## 1          0,0          0,0          0,0          0,0          0,0
## 2          0,0          0,0          0,0          0,0          0,0

```

## 3	0,0	0,0	0,0	0,0	0,0	0,0
## 4	0,0	0,0	0,0	0,0	0,0	0,0
## 5	0,0	0,0	0,0	0,0	0,0	0,0
## 6	0,0	0,0	0,0	0,0	0,0	0,0
##	rendimiento_cuat_b	rendimiento_total	exento_npp	anyo_inicio_estudios		
## 1	NA	0,0	NA	2017		
## 2	NA	0,0	NA	2017		
## 3	NA	0,0	NA	2017		
## 4	NA	0,0	NA	2017		
## 5	NA	0,0	NA	2017		
## 6	NA	0,0	NA	2017		
##	es_retitulado	es_adaptado	cred_sup_1o	cred_sup_2o	cred_sup_3o	cred_sup_4o
## 1	NA	NA	60,0	51,0	42,0	9,0
## 2	NA	NA	60,0	51,0	42,0	9,0
## 3	NA	NA	60,0	51,0	42,0	9,0
## 4	NA	NA	60,0	51,0	42,0	9,0
## 5	NA	NA	60,0	51,0	42,0	9,0
## 6	NA	NA	60,0	51,0	42,0	9,0
##	cred_sup_5o	cred_sup_6o	practicass	actividades	ajuste	cred_sup_tit
## 1	0,0	0	4,5	1,0	NA	167,5
## 2	0,0	0	4,5	1,0	NA	167,5
## 3	0,0	0	4,5	1,0	NA	167,5
## 4	0,0	0	4,5	1,0	NA	167,5
## 5	0,0	0	4,5	1,0	NA	167,5
## 6	0,0	0	4,5	1,0	NA	167,5
##	cred_pend_sup_tit	impagado_curso_mat	asig1	pract1	activ1	total1
## 1	72,5	NA	0,0	NA	NA	0
## 2	72,5	NA	0,0	NA	NA	0
## 3	72,5	NA	0,0	NA	NA	0
## 4	72,5	NA	0,0	NA	NA	0
## 5	72,5	NA	0,0	NA	NA	0
## 6	72,5	NA	0,0	NA	NA	0
##	rend_total_ultimo	rend_total_penultimo	rend_total_antepenultimo			
## 1	100,0	0,0	15,789473684210526			
## 2	100,0	0,0	15,789473684210526			
## 3	100,0	0,0	15,789473684210526			
## 4	100,0	0,0	15,789473684210526			
## 5	100,0	0,0	15,789473684210526			
## 6	100,0	0,0	15,789473684210526			
##	pft_events_2024_7	pft_days_logged_2024_7	pft_visits_2024_7			
## 1	NA	NA	NA			
## 2	NA	NA	NA			
## 3	NA	NA	NA			
## 4	NA	NA	NA			
## 5	NA	NA	NA			
## 6	NA	NA	NA			
##	pft_assignment_submissions_2024_7	pft_test_submissions_2024_7				
## 1	NA	NA				
## 2	NA	NA				
## 3	NA	NA				
## 4	NA	NA				
## 5	NA	NA				
## 6	NA	NA				
##	pft_total_minutes_2024_7	n_wifi_days_2024_7	resource_events_2024_7			

##	1		NA	NA
##	2		NA	NA
##	3		NA	NA
##	4		NA	NA
##	5		NA	NA
##	6		NA	NA
##	n_resource_days_2024_7 pft_events_2024_8 pft_days_logged_2024_8			
##	1	NA	NA	NA
##	2	NA	NA	NA
##	3	NA	NA	NA
##	4	NA	NA	NA
##	5	NA	NA	NA
##	6	NA	NA	NA
##	pft_visits_2024_8 pft_assignment_submissions_2024_8			
##	1	NA	NA	
##	2	NA	NA	
##	3	NA	NA	
##	4	NA	NA	
##	5	NA	NA	
##	6	NA	NA	
##	pft_test_submissions_2024_8 pft_total_minutes_2024_8 n_wifi_days_2024_8			
##	1	NA		NA
##	2	NA		NA
##	3	NA		NA
##	4	NA		NA
##	5	NA		NA
##	6	NA		NA
##	resource_events_2024_8 n_resource_days_2024_8 pft_events_2024_9			
##	1	NA	NA	NA
##	2	NA	NA	4
##	3	NA	NA	NA
##	4	NA	NA	NA
##	5	NA	NA	5
##	6	NA	NA	33
##	pft_days_logged_2024_9 pft_visits_2024_9 pft_assignment_submissions_2024_9			
##	1	NA	NA	NA
##	2	2	2	NA
##	3	NA	NA	NA
##	4	NA	NA	NA
##	5	1	1	NA
##	6	5	8	NA
##	pft_test_submissions_2024_9 pft_total_minutes_2024_9 n_wifi_days_2024_9			
##	1	NA		NA
##	2	NA	91,17016666666667	NA
##	3	NA		NA
##	4	NA		NA
##	5	NA	47,7713	NA
##	6	NA	462,5455666666667	NA
##	resource_events_2024_9 n_resource_days_2024_9 pft_events_2024_10			
##	1	NA	NA	NA
##	2	1	1	NA
##	3	NA	NA	NA
##	4	NA	NA	NA
##	5	1	1	NA

## 6	18	4	2
##	pft_days_logged_2024_10	pft_visits_2024_10	pft_assignment_submissions_2024_10
## 1	NA	NA	NA
## 2	NA	NA	NA
## 3	NA	NA	NA
## 4	NA	NA	NA
## 5	NA	NA	NA
## 6	1	1	NA
##	pft_test_submissions_2024_10	pft_total_minutes_2024_10	n_wifi_days_2024_10
## 1	NA		NA
## 2	NA		NA
## 3	NA		NA
## 4	NA		NA
## 5	NA		NA
## 6	NA	45,65503333333336	NA
##	resource_events_2024_10	n_resource_days_2024_10	pft_events_2024_11
## 1	NA	NA	NA
## 2	NA	NA	NA
## 3	NA	NA	NA
## 4	NA	NA	NA
## 5	NA	NA	NA
## 6	1	1	NA
##	pft_days_logged_2024_11	pft_visits_2024_11	pft_assignment_submissions_2024_11
## 1	NA	NA	NA
## 2	NA	NA	NA
## 3	NA	NA	NA
## 4	NA	NA	NA
## 5	NA	NA	NA
## 6	NA	NA	NA
##	pft_test_submissions_2024_11	pft_total_minutes_2024_11	n_wifi_days_2024_11
## 1	NA		NA
## 2	NA		NA
## 3	NA		NA
## 4	NA		NA
## 5	NA		NA
## 6	NA		NA
##	resource_events_2024_11	n_resource_days_2024_11	pft_events_2024_12
## 1	NA	NA	NA
## 2	NA	NA	NA
## 3	NA	NA	NA
## 4	NA	NA	NA
## 5	NA	NA	NA
## 6	NA	NA	NA
##	pft_days_logged_2024_12	pft_visits_2024_12	pft_assignment_submissions_2024_12
## 1	NA	NA	NA
## 2	NA	NA	NA
## 3	NA	NA	NA
## 4	NA	NA	NA
## 5	NA	NA	NA
## 6	NA	NA	NA
##	pft_test_submissions_2024_12	pft_total_minutes_2024_12	n_wifi_days_2024_12
## 1	NA		NA
## 2	NA		NA
## 3	NA		NA

```

## 4          NA          NA
## 5          NA          NA
## 6          NA          NA
## resource_events_2024_12 n_resource_days_2024_12 pft_events_2025_1
## 1          NA          NA          NA
## 2          NA          NA          NA
## 3          NA          NA          NA
## 4          NA          NA          NA
## 5          NA          NA          NA
## 6          NA          NA          NA
## pft_days_logged_2025_1 pft_visits_2025_1 pft_assignment_submissions_2025_1
## 1          NA          NA          NA
## 2          NA          NA          NA
## 3          NA          NA          NA
## 4          NA          NA          NA
## 5          NA          NA          NA
## 6          NA          NA          NA
## pft_test_submissions_2025_1 pft_total_minutes_2025_1 n_wifi_days_2025_1
## 1          NA          NA          NA
## 2          NA          NA          NA
## 3          NA          NA          NA
## 4          NA          NA          NA
## 5          NA          NA          NA
## 6          NA          NA          NA
## resource_events_2025_1 n_resource_days_2025_1
## 1          NA          NA
## 2          NA          NA
## 3          NA          NA
## 4          NA          NA
## 5          NA          NA
## 6          NA          NA

```

Unión de los datasets.

```

dataset=merge.data.frame(dataset,select(titulaciones, c("tit_hash","titnom")))

dataset=merge.data.frame(dataset,select(asignaturas,c("asi_hash","asinom")))

dataset=merge.data.frame(dataset,estudiantes)

dataset_ordenado=dataset[,c(1:3,137:143,4:136)]
head(dataset_ordenado)

```

```

##      dni_hash      asi_hash      tit_hash
## 1 0000355c755a 9bc7684b0d1c d4ee9f58e586
## 2 0000355c755a b5eb854bc3ac d4ee9f58e586
## 3 0000355c755a 6b454452a2b6 d4ee9f58e586
## 4 0000355c755a 7dcf1501322b d4ee9f58e586
## 5 0000355c755a 9877e1af96b8 d4ee9f58e586
## 6 0000355c755a 9abb2626204b d4ee9f58e586
##
##                                     titnom
## 1 Grado en Ingeniería en Diseño Industrial y Desarrollo de Productos
## 2 Grado en Ingeniería en Diseño Industrial y Desarrollo de Productos

```

```

## 3 Grado en Ingeniería en Diseño Industrial y Desarrollo de Productos
## 4 Grado en Ingeniería en Diseño Industrial y Desarrollo de Productos
## 5 Grado en Ingeniería en Diseño Industrial y Desarrollo de Productos
## 6 Grado en Ingeniería en Diseño Industrial y Desarrollo de Productos
##      asinom nacionalitat data_nac sexe alta_universitat
## 1 Taller de modelos y prototipos      E      1996      M      2001
## 2      Taller de Diseño III            E      1996      M      2001
## 3 Mecánica y Teoría de Mecanismos      E      1996      M      2001
## 4      Empresa                        E      1996      M      2001
## 5 Diseño Asistido por Ordenador        E      1996      M      2001
## 6 Diseño Básico y Creatividad          E      1996      M      2001
## prov_origen anyo_ingreso tipo_ingreso nota10 nota14 campus estudios_p
## 1 COMVAL      2024      NAP      7,5      7.5      V      5
## 2 COMVAL      2024      NAP      7,5      7.5      V      5
## 3 COMVAL      2024      NAP      7,5      7.5      V      5
## 4 COMVAL      2024      NAP      7,5      7.5      V      5
## 5 COMVAL      2024      NAP      7,5      7.5      V      5
## 6 COMVAL      2024      NAP      7,5      7.5      V      5
## estudios_m dedicacion desplazado discapacidad abandono becado
## 1      5      TP      NA      NA      NA      0
## 2      5      TP      NA      NA      NA      0
## 3      5      TP      NA      NA      NA      0
## 4      5      TP      NA      NA      NA      0
## 5      5      TP      NA      NA      NA      0
## 6      5      TP      NA      NA      NA      0
## preferencia_seleccion baja_fecha caca grupos_por_tipocredito matricula_activa
## 1      1      2024      1
## 2      1      2024      1
## 3      1      2024      1
## 4      1      2024      1
## 5      1      2024      1
## 6      1      2024      1
## nota_asig      fecha_datos cod_centro curso_mas_bajo curso_mas_alto
## 1      7,5 2025-01-31 14:42:43      E      2      3
## 2      9,5 2025-01-31 14:42:43      E      2      3
## 3      6,6 2025-01-31 14:42:43      E      2      3
## 4      7,0 2025-01-31 14:42:43      E      2      3
## 5      7,5 2025-01-31 14:42:43      E      2      3
## 6      9,3 2025-01-31 14:42:43      E      2      3
## cred_mat1 cred_mat2 cred_mat3 cred_mat4 cred_mat5 cred_mat6 cred_sup_normal
## 1      0,0      12,0      4,5      0,0      0,0      0      0,0
## 2      0,0      12,0      4,5      0,0      0,0      0      0,0
## 3      0,0      12,0      4,5      0,0      0,0      0      0,0
## 4      0,0      12,0      4,5      0,0      0,0      0      0,0
## 5      0,0      12,0      4,5      0,0      0,0      0      0,0
## 6      0,0      12,0      4,5      0,0      0,0      0      0,0
## cred_sup_espec cred_sup cred_mat_normal cred_mat_movilidad cred_ptes_acta
## 1      193,5      193,5      16,5      0,0      16,5
## 2      193,5      193,5      16,5      0,0      16,5
## 3      193,5      193,5      16,5      0,0      16,5
## 4      193,5      193,5      16,5      0,0      16,5
## 5      193,5      193,5      16,5      0,0      16,5
## 6      193,5      193,5      16,5      0,0      16,5
## cred_mat_practicas cred_mat_sem_a cred_mat_sem_b cred_mat_anu cred_mat_total

```

## 1	0,0	6,0	10,5	0,0	16,5		
## 2	0,0	6,0	10,5	0,0	16,5		
## 3	0,0	6,0	10,5	0,0	16,5		
## 4	0,0	6,0	10,5	0,0	16,5		
## 5	0,0	6,0	10,5	0,0	16,5		
## 6	0,0	6,0	10,5	0,0	16,5		
##	cred_sup_sem_a	cred_sup_sem_b	cred_sup_anu	cred_sup_total	rendimiento_cuat_a		
## 1	0,0	0,0	0,0	0,0	0,0		
## 2	0,0	0,0	0,0	0,0	0,0		
## 3	0,0	0,0	0,0	0,0	0,0		
## 4	0,0	0,0	0,0	0,0	0,0		
## 5	0,0	0,0	0,0	0,0	0,0		
## 6	0,0	0,0	0,0	0,0	0,0		
##	rendimiento_cuat_b	rendimiento_total	exento_npp	anyo_inicio_estudios			
## 1	NA	0,0	1	2024			
## 2	NA	0,0	1	2024			
## 3	NA	0,0	1	2024			
## 4	NA	0,0	1	2024			
## 5	NA	0,0	1	2024			
## 6	NA	0,0	1	2024			
##	es_retitulado	es_adaptado	cred_sup_1o	cred_sup_2o	cred_sup_3o	cred_sup_4o	
## 1	NA	NA	60,0	43,5	52,5	37,5	
## 2	NA	NA	60,0	43,5	52,5	37,5	
## 3	NA	NA	60,0	43,5	52,5	37,5	
## 4	NA	NA	60,0	43,5	52,5	37,5	
## 5	NA	NA	60,0	43,5	52,5	37,5	
## 6	NA	NA	60,0	43,5	52,5	37,5	
##	cred_sup_5o	cred_sup_6o	practicas	actividades	ajuste	cred_sup_tit	
## 1	0,0	0	0,0		NA	193,5	
## 2	0,0	0	0,0		NA	193,5	
## 3	0,0	0	0,0		NA	193,5	
## 4	0,0	0	0,0		NA	193,5	
## 5	0,0	0	0,0		NA	193,5	
## 6	0,0	0	0,0		NA	193,5	
##	cred_pend_sup_tit	impagado_curso_mat	asig1	pract1	activ1	total1	ajuste1
## 1	46,5		NA	0,0	NA	NA	0
## 2	46,5		NA	0,0	NA	NA	0
## 3	46,5		NA	0,0	NA	NA	0
## 4	46,5		NA	0,0	NA	NA	0
## 5	46,5		NA	0,0	NA	NA	0
## 6	46,5		NA	0,0	NA	NA	0
##	rend_total_ultimo	rend_total_penultimo	rend_total_antepenultimo				
## 1							
## 2							
## 3							
## 4							
## 5							
## 6							
##	pft_events_2024_7	pft_days_logged_2024_7	pft_visits_2024_7				
## 1	NA		NA		NA		
## 2	NA		NA		NA		
## 3	NA		NA		NA		
## 4	NA		NA		NA		
## 5	NA		NA		NA		

## 6	NA	NA	NA
##	pft_assignment_submissions_2024_7	pft_test_submissions_2024_7	
## 1	NA	NA	
## 2	NA	NA	
## 3	NA	NA	
## 4	NA	NA	
## 5	NA	NA	
## 6	NA	NA	
##	pft_total_minutes_2024_7	n_wifi_days_2024_7	resource_events_2024_7
## 1		NA	NA
## 2		NA	NA
## 3		NA	NA
## 4		NA	NA
## 5		NA	NA
## 6		NA	NA
##	n_resource_days_2024_7	pft_events_2024_8	pft_days_logged_2024_8
## 1	NA	NA	NA
## 2	NA	NA	NA
## 3	NA	NA	NA
## 4	NA	NA	NA
## 5	NA	NA	NA
## 6	NA	NA	NA
##	pft_visits_2024_8	pft_assignment_submissions_2024_8	
## 1	NA	NA	
## 2	NA	NA	
## 3	NA	NA	
## 4	NA	NA	
## 5	NA	NA	
## 6	NA	NA	
##	pft_test_submissions_2024_8	pft_total_minutes_2024_8	n_wifi_days_2024_8
## 1	NA		NA
## 2	NA		NA
## 3	NA		NA
## 4	NA		NA
## 5	NA		NA
## 6	NA		NA
##	resource_events_2024_8	n_resource_days_2024_8	pft_events_2024_9
## 1	NA	NA	NA
## 2	NA	NA	NA
## 3	NA	NA	NA
## 4	NA	NA	NA
## 5	NA	NA	NA
## 6	NA	NA	NA
##	pft_days_logged_2024_9	pft_visits_2024_9	pft_assignment_submissions_2024_9
## 1	NA	NA	NA
## 2	NA	NA	NA
## 3	NA	NA	NA
## 4	NA	NA	NA
## 5	NA	NA	NA
## 6	NA	NA	NA
##	pft_test_submissions_2024_9	pft_total_minutes_2024_9	n_wifi_days_2024_9
## 1	NA		NA
## 2	NA		NA
## 3	NA		NA

## 4	NA	NA
## 5	NA	NA
## 6	NA	NA
## resource_events_2024_9	n_resource_days_2024_9	pft_events_2024_10
## 1	NA	NA
## 2	NA	NA
## 3	NA	NA
## 4	NA	NA
## 5	NA	NA
## 6	NA	NA
## pft_days_logged_2024_10	pft_visits_2024_10	pft_assignment_submissions_2024_10
## 1	NA	NA
## 2	NA	NA
## 3	NA	NA
## 4	NA	NA
## 5	NA	NA
## 6	NA	NA
## pft_test_submissions_2024_10	pft_total_minutes_2024_10	n_wifi_days_2024_10
## 1	NA	NA
## 2	NA	NA
## 3	NA	NA
## 4	NA	NA
## 5	NA	NA
## 6	NA	NA
## resource_events_2024_10	n_resource_days_2024_10	pft_events_2024_11
## 1	NA	NA
## 2	NA	NA
## 3	NA	NA
## 4	NA	NA
## 5	NA	NA
## 6	NA	NA
## pft_days_logged_2024_11	pft_visits_2024_11	pft_assignment_submissions_2024_11
## 1	NA	NA
## 2	NA	NA
## 3	NA	NA
## 4	NA	NA
## 5	NA	NA
## 6	NA	NA
## pft_test_submissions_2024_11	pft_total_minutes_2024_11	n_wifi_days_2024_11
## 1	NA	NA
## 2	NA	NA
## 3	NA	NA
## 4	NA	NA
## 5	NA	NA
## 6	NA	NA
## resource_events_2024_11	n_resource_days_2024_11	pft_events_2024_12
## 1	NA	NA
## 2	NA	NA
## 3	NA	NA
## 4	NA	NA
## 5	NA	NA
## 6	NA	NA
## pft_days_logged_2024_12	pft_visits_2024_12	pft_assignment_submissions_2024_12
## 1	NA	NA

```

## 2          NA          NA          NA
## 3          NA          NA          NA
## 4          NA          NA          NA
## 5          NA          NA          NA
## 6          NA          NA          NA
##  pft_test_submissions_2024_12 pft_total_minutes_2024_12 n_wifi_days_2024_12
## 1          NA          NA          NA
## 2          NA          NA          NA
## 3          NA          NA          NA
## 4          NA          NA          NA
## 5          NA          NA          NA
## 6          NA          NA          NA
##  resource_events_2024_12 n_resource_days_2024_12 pft_events_2025_1
## 1          NA          NA          NA
## 2          NA          NA          NA
## 3          NA          NA          NA
## 4          NA          NA          NA
## 5          NA          NA          NA
## 6          NA          NA          NA
##  pft_days_logged_2025_1 pft_visits_2025_1 pft_assignment_submissions_2025_1
## 1          NA          NA          NA
## 2          NA          NA          NA
## 3          NA          NA          NA
## 4          NA          NA          NA
## 5          NA          NA          NA
## 6          NA          NA          NA
##  pft_test_submissions_2025_1 pft_total_minutes_2025_1 n_wifi_days_2025_1
## 1          NA          NA          NA
## 2          NA          NA          NA
## 3          NA          NA          NA
## 4          NA          NA          NA
## 5          NA          NA          NA
## 6          NA          NA          NA
##  resource_events_2025_1 n_resource_days_2025_1
## 1          NA          NA
## 2          NA          NA
## 3          NA          NA
## 4          NA          NA
## 5          NA          NA
## 6          NA          NA

```

‘ Conversión de columnas numéricas que se han clasificado como cadena de texto al tener como separador una “,”.

```

convertir_columnas_especificas=function(df, columnas_a_convertir) {
  df[columnas_a_convertir]=lapply(df[columnas_a_convertir], function(col) {
    col=as.character(col)
    col=trimws(col)
    col=gsub(",", ".", col, fixed = TRUE)
    num_col=as.double(col)

    if (any(!is.na(num_col))) {
      return(num_col)
    }
  })
}

```

```

    return(col)
  })
  return(df)
}

columnas_a_convertir=c(13, 28, 33:55, 56, 61:68, 70:72, 73,79:143)

dataset_ordenado=convertir_columnas_especificas(dataset_ordenado, columnas_a_convertir)

dataset_ordenado$rend_total_ultimo=as.numeric(gsub(",", ".", dataset_ordenado$rend_total_ultimo))
dataset_ordenado$activ1=as.numeric(gsub(",", ".", dataset_ordenado$activ1))

summary(dataset_ordenado)

```

```

##      dni_hash      asi_hash      tit_hash      titnom
## Length:281540    Length:281540    Length:281540    Length:281540
## Class :character  Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
##      asinom      nacionalitat      data_nac      sexe
## Length:281540    Length:281540    Min.   :1996    Length:281540
## Class :character  Class :character  1st Qu.:2002    Class :character
## Mode  :character  Mode  :character  Median :2004    Mode  :character
##                                     Mean  :2003
##                                     3rd Qu.:2005
##                                     Max.   :2007
##                                     NA's   :12
## alta_universitat prov_origen      anyo_ingreso      tipo_ingreso
## Min.   :1975      Length:281540    Min.   :2009      Length:281540
## 1st Qu.:2021      Class :character  1st Qu.:2021      Class :character
## Median :2022      Mode  :character  Median :2023      Mode  :character
## Mean    :2022
## 3rd Qu. :2024
## Max.    :2025
##                                     Mean    :2022
##                                     3rd Qu. :2024
##                                     Max.    :2024
##                                     NA's    :44039
##      nota10      nota14      campus      estudios_p
## Min.   : 5.00    Min.   : 5.00    Length:281540    Min.   :1.000
## 1st Qu.: 6.97    1st Qu.: 9.15    Class :character  1st Qu.:4.000
## Median : 7.84    Median :10.79    Mode  :character  Median :4.000
## Mean    : 7.78    Mean    :10.43
## 3rd Qu. : 8.64    3rd Qu. :12.00
## Max.    :10.00    Max.    :14.00
## NA's    :52709    NA's    :53509
##                                     Mean    :4.314
##                                     3rd Qu. :5.000
##                                     Max.    :6.000
##                                     NA's    :974
##      estudios_m      dedicacion      desplazado      discapacidad
## Min.   :1.0      Length:281540    Min.   :1      Min.   :1
## 1st Qu.:4.0      Class :character  1st Qu.:1      1st Qu.:1
## Median :5.0      Mode  :character  Median :1      Median :1
## Mean    :4.4
## 3rd Qu. :5.0
##                                     Mean    :1
##                                     3rd Qu. :1
##                                     Mean    :1
##                                     3rd Qu. :1

```

```

## Max.      :6.0          Max.      :1          Max.      :1
## NA's      :974          NA's      :177596    NA's      :279122
## abandono      becado      preferencia_seleccion  baja_fecha
## Mode:logical  Min.      :0.0000  Min.      : 1          Length:281540
## NA's:281540   1st Qu.:0.0000  1st Qu.: 1          Class :character
##              Median :0.0000  Median : 1          Mode  :character
##              Mean   :0.4179  Mean   : 2
##              3rd Qu.:0.0000  3rd Qu.: 2
##              Max.   :2.0000  Max.   :20
##              NA's   :57618
##          caca      grupos_por_tipocredito matricula_activa  nota_asig
## Min.      :2024   Length:281540      Min.      :1          Min.      : 0.00
## 1st Qu.:2024   Class :character      1st Qu.:1          1st Qu.: 0.00
## Median :2024   Mode  :character      Median :1          Median : 0.00
## Mean   :2024          Mean   :1          Mean   : 1.06
## 3rd Qu.:2024          3rd Qu.:1          3rd Qu.: 0.00
## Max.   :2024          Max.   :1          Max.   :10.00
##              NA's   :169393
## fecha_datos      cod_centro      curso_mas_bajo  curso_mas_alto
## Length:281540   Length:281540      Min.      :1.000  Min.      :1.000
## Class :character  Class :character      1st Qu.:1.000  1st Qu.:1.000
## Mode  :character  Mode  :character      Median :2.000  Median :2.000
##              Mean   :1.887  Mean   :2.299
##              3rd Qu.:3.000  3rd Qu.:3.000
##              Max.   :5.000  Max.   :5.000
##
## cred_mat1      cred_mat2      cred_mat3      cred_mat4
## Min.      : 0.0    Min.      : 0.00    Min.      : 0.00    Min.      : 0.000
## 1st Qu.: 0.0    1st Qu.: 0.00    1st Qu.: 0.00    1st Qu.: 0.000
## Median : 0.0    Median : 0.00    Median : 0.00    Median : 0.000
## Mean   :22.5    Mean   :14.53    Mean   : 12.12    Mean   : 8.476
## 3rd Qu.:60.0    3rd Qu.:25.50    3rd Qu.: 15.00    3rd Qu.: 0.000
## Max.   :82.0    Max.   :87.00    Max.   :111.00    Max.   :102.000
##
## cred_mat5      cred_mat6  cred_sup_normal  cred_sup_espec
## Min.      : 0.0000  Min.      :0    Min.      : 0.000  Min.      : 0.00
## 1st Qu.: 0.0000  1st Qu.:0    1st Qu.: 0.000  1st Qu.: 0.00
## Median : 0.0000  Median :0    Median : 0.000  Median : 0.00
## Mean   : 0.9164  Mean   :0    Mean   : 1.776  Mean   : 4.21
## 3rd Qu.: 0.0000  3rd Qu.:0    3rd Qu.: 0.000  3rd Qu.: 0.00
## Max.   :81.0000  Max.   :0    Max.   :40.000  Max.   :327.00
##
## cred_sup      cred_mat_normal  cred_mat_movilidad  cred_ptes_acta
## Min.      : 0.000  Min.      : 0.00  Min.      : 0.000  Min.      : 0.00
## 1st Qu.: 0.000  1st Qu.: 60.00  1st Qu.: 0.000  1st Qu.: 43.50
## Median : 0.000  Median : 60.00  Median : 0.000  Median : 60.00
## Mean   : 5.987  Mean   : 55.08  Mean   : 1.766  Mean   : 49.66
## 3rd Qu.: 4.500  3rd Qu.: 61.00  3rd Qu.: 0.000  3rd Qu.: 60.00
## Max.   :327.000  Max.   :114.00  Max.   :120.000  Max.   :102.00
##
## cred_mat_practicas  cred_mat_sem_a  cred_mat_sem_b  cred_mat_anu
## Min.      : 0.000  Min.      : 0.00  Min.      : 0.0  Min.      : 0.000
## 1st Qu.: 0.000  1st Qu.: 22.50  1st Qu.: 19.5  1st Qu.: 0.000
## Median : 0.000  Median : 30.00  Median : 28.5  Median : 0.000

```

```

## Mean      : 0.969      Mean      : 26.63      Mean      : 25.1      Mean      : 6.842
## 3rd Qu.: 0.000      3rd Qu.: 30.00      3rd Qu.: 30.0      3rd Qu.:11.000
## Max.      :30.000      Max.      :148.00      Max.      :136.5      Max.      :99.000
##
## cred_mat_total      cred_sup_sem_a      cred_sup_sem_b      cred_sup_anu
## Min.      : 0.00      Min.      : 0.000      Min.      : 0.00000      Min.      : 0.00000
## 1st Qu.: 60.00      1st Qu.: 0.000      1st Qu.: 0.00000      1st Qu.: 0.00000
## Median : 60.00      Median : 0.000      Median : 0.00000      Median : 0.00000
## Mean      : 58.58      Mean      : 1.556      Mean      : 0.01021      Mean      : 0.01508
## 3rd Qu.: 60.00      3rd Qu.: 0.000      3rd Qu.: 0.00000      3rd Qu.: 0.00000
## Max.      :290.50      Max.      :39.000      Max.      :30.00000      Max.      :30.00000
##
## cred_sup_total      rendimiento_cuat_a      rendimiento_cuat_b      rendimiento_total
## Min.      : 0.000      Min.      : 0.000      Length:281540      Min.      : 0.000
## 1st Qu.: 0.000      1st Qu.: 0.000      Class :character      1st Qu.: 0.000
## Median : 0.000      Median : 0.000      Mode  :character      Median : 0.000
## Mean      : 1.581      Mean      : 4.872      Mean      : 2.502
## 3rd Qu.: 0.000      3rd Qu.: 0.000      3rd Qu.: 0.000
## Max.      :39.000      Max.      :100.000      Max.      :100.000
## NA's      :24829      NA's      :20173
## exento_npp      anyo_inicio_estudios      es_retitulado      es_adaptado
## Min.      :1      Min.      :2009      Min.      :1      Min.      :1
## 1st Qu.:1      1st Qu.:2022      1st Qu.:1      1st Qu.:1
## Median :1      Median :2023      Median :1      Median :1
## Mean      :1      Mean      :2023      Mean      :1      Mean      :1
## 3rd Qu.:1      3rd Qu.:2024      3rd Qu.:1      3rd Qu.:1
## Max.      :1      Max.      :2024      Max.      :1      Max.      :1
## NA's      :240079      NA's      :281302      NA's      :274689
## cred_sup_1o      cred_sup_2o      cred_sup_3o      cred_sup_4o
## Min.      : 0.00      Min.      : 0.00      Min.      : 0.00      Min.      : 0.000
## 1st Qu.: 0.00      1st Qu.: 0.00      1st Qu.: 0.00      1st Qu.: 0.000
## Median :60.00      Median : 0.00      Median : 0.00      Median : 0.000
## Mean      :37.88      Mean      :23.33      Mean      : 11.26      Mean      : 2.719
## 3rd Qu.:60.00      3rd Qu.:60.00      3rd Qu.: 6.00      3rd Qu.: 0.000
## Max.      :90.00      Max.      :84.00      Max.      :127.50      Max.      :87.000
##
## cred_sup_5o      cred_sup_6o      practicas      actividades
## Min.      : 0.0000      Min.      :0      Min.      : 0.0000      Min.      : 0.03
## 1st Qu.: 0.0000      1st Qu.:0      1st Qu.: 0.0000      1st Qu.: 1.00
## Median : 0.0000      Median :0      Median : 0.0000      Median : 1.33
## Mean      : 0.3461      Mean      :0      Mean      : 0.4391      Mean      : 3.16
## 3rd Qu.: 0.0000      3rd Qu.:0      3rd Qu.: 0.0000      3rd Qu.: 4.92
## Max.      :81.0000      Max.      :0      Max.      :20.0000      Max.      :18.00
## NA's      :254456
## ajuste      cred_sup_tit      cred_pend_sup_tit      impagado_curso_mat
## Min.      : 0.1      Min.      : -999.00      Min.      : 0.0      Min.      :1
## 1st Qu.: 6.0      1st Qu.: 0.00      1st Qu.: 81.0      1st Qu.:1
## Median : 9.0      Median : 60.00      Median :144.0      Median :1
## Mean      : 9.9      Mean      : 76.37      Mean      :150.5      Mean      :1
## 3rd Qu.:13.5      3rd Qu.: 122.33      3rd Qu.:225.0      3rd Qu.:1
## Max.      :48.0      Max.      : 373.50      Max.      :999.0      Max.      :1
## NA's      :274951      NA's      :281501
## asig1      pract1      activ1      total1
## Min.      : 0.000      Min.      : 0.50      Min.      : 0.18      Min.      : 1.67

```

```

## 1st Qu.: 0.000    1st Qu.: 4.50    1st Qu.: 1.00    1st Qu.:10.00
## Median : 0.000    Median : 6.00    Median : 2.00    Median :13.50
## Mean   : 1.581    Mean   : 7.31    Mean   : 3.15    Mean   :14.51
## 3rd Qu.: 0.000    3rd Qu.: 9.00    3rd Qu.: 4.65    3rd Qu.:18.97
## Max.   :39.000    Max.   :18.00    Max.   :18.00    Max.   :31.84
##                NA's   :278573    NA's   :270786    NA's   :281145
##      ajuste1  rend_total_ultimo rend_total_penultimo rend_total_antepenultimo
## Min.   :0      Min.   : 0.00    Min.   : 0.00    Min.   : 0.00
## 1st Qu.:0      1st Qu.: 73.91    1st Qu.: 69.44    1st Qu.: 62.50
## Median :0      Median :100.00    Median :100.00    Median :100.00
## Mean   :0      Mean   : 83.36    Mean   : 81.69    Mean   : 78.86
## 3rd Qu.:0      3rd Qu.:100.00    3rd Qu.:100.00    3rd Qu.:100.00
## Max.   :0      Max.   :100.00    Max.   :100.00    Max.   :100.00
##                NA's   :104955    NA's   :169548    NA's   :218991
## pft_events_2024_7 pft_days_logged_2024_7 pft_visits_2024_7
## Min.   : 1.00    Min.   :1.00    Min.   : 1.00
## 1st Qu.: 2.00    1st Qu.:1.00    1st Qu.: 1.00
## Median : 5.00    Median :1.00    Median : 1.00
## Mean   : 7.78    Mean   :1.39    Mean   : 2.13
## 3rd Qu.:10.00    3rd Qu.:1.00    3rd Qu.: 2.00
## Max.   :191.00    Max.   :9.00    Max.   :24.00
## NA's   :280893    NA's   :280893    NA's   :280898
## pft_assignment_submissions_2024_7 pft_test_submissions_2024_7
## Length:281540    Length:281540
## Class :character    Class :character
## Mode  :character    Mode  :character
##
##
##
##
## pft_total_minutes_2024_7 n_wifi_days_2024_7 resource_events_2024_7
## Min.   :-29.00    Min.   : 1.00    Min.   : 1.0
## 1st Qu.: 45.56    1st Qu.: 1.00    1st Qu.: 1.0
## Median : 47.64    Median : 2.00    Median : 2.0
## Mean   : 80.21    Mean   : 3.05    Mean   : 3.7
## 3rd Qu.: 91.77    3rd Qu.: 3.00    3rd Qu.: 4.0
## Max.   :779.45    Max.   :30.00    Max.   :47.0
## NA's   :280906    NA's   :214252    NA's   :281170
## n_resource_days_2024_7 pft_events_2024_8 pft_days_logged_2024_8
## Min.   :1.00    Min.   : 1.00    Min.   : 1.00
## 1st Qu.:1.00    1st Qu.: 2.00    1st Qu.: 1.00
## Median :1.00    Median : 4.00    Median : 1.00
## Mean   :1.25    Mean   : 6.22    Mean   : 1.27
## 3rd Qu.:1.00    3rd Qu.: 7.00    3rd Qu.: 1.00
## Max.   :6.00    Max.   :212.00    Max.   :17.00
## NA's   :281170    NA's   :278469    NA's   :278469
## pft_visits_2024_8 pft_assignment_submissions_2024_8
## Min.   : 1.00    Length:281540
## 1st Qu.: 1.00    Class :character
## Median : 1.00    Mode  :character
## Mean   : 1.81
## 3rd Qu.: 2.00
## Max.   :29.00
## NA's   :278473

```

```

## pft_test_submissions_2024_8 pft_total_minutes_2024_8 n_wifi_days_2024_8
## Length:281540      Min.   : -93.98      Min.   : 1.00
## Class :character    1st Qu.: 45.39      1st Qu.: 1.00
## Mode  :character    Median : 46.77      Median : 1.00
##                      Mean    : 68.82      Mean    : 2.14
##                      3rd Qu.: 86.84      3rd Qu.: 2.00
##                      Max.    :1361.40     Max.    :30.00
##                      NA's    :278524     NA's    :275508
## resource_events_2024_8 n_resource_days_2024_8 pft_events_2024_9
## Min.   : 1.00      Min.   :1.00      Min.   : 1.00
## 1st Qu.: 1.00      1st Qu.:1.00      1st Qu.: 9.00
## Median : 2.00      Median :1.00      Median : 22.00
## Mean   : 3.71      Mean   :1.15      Mean   : 33.86
## 3rd Qu.: 4.00      3rd Qu.:1.00      3rd Qu.: 41.00
## Max.   :117.00     Max.   :8.00      Max.   :1140.00
## NA's   :279890     NA's   :279890     NA's   :223406
## pft_days_logged_2024_9 pft_visits_2024_9 pft_assignment_submissions_2024_9
## Min.   : 1.00      Min.   : 1.00      Min.   : 1
## 1st Qu.: 2.00      1st Qu.: 2.00      1st Qu.: 1
## Median : 4.00      Median : 6.00      Median : 1
## Mean   : 4.86      Mean   : 7.95      Mean   : 2
## 3rd Qu.: 7.00      3rd Qu.: 11.00     3rd Qu.: 2
## Max.   :30.00      Max.   :145.00     Max.   :18
## NA's   :223370     NA's   :223608     NA's   :274317
## pft_test_submissions_2024_9 pft_total_minutes_2024_9 n_wifi_days_2024_9
## Min.   : 1.00      Min.   : -17.27     Min.   : 1.00
## 1st Qu.: 1.00      1st Qu.: 94.25      1st Qu.: 8.00
## Median : 2.00      Median : 255.01     Median :11.00
## Mean   : 2.33      Mean   : 343.17      Mean   :10.45
## 3rd Qu.: 3.00      3rd Qu.: 475.60     3rd Qu.:14.00
## Max.   :50.00      Max.   :4712.84     Max.   :29.00
## NA's   :274995     NA's   :224090     NA's   :113911
## resource_events_2024_9 n_resource_days_2024_9 pft_events_2024_10
## Min.   : 1.00      Min.   : 1      Min.   : 1.00
## 1st Qu.: 4.00      1st Qu.: 2      1st Qu.: 15.00
## Median : 9.00      Median : 3      Median : 40.00
## Mean   : 12.59     Mean   : 4      Mean   : 61.39
## 3rd Qu.: 16.00     3rd Qu.: 5      3rd Qu.: 80.00
## Max.   :435.00     Max.   :25      Max.   :2791.00
## NA's   :232897     NA's   :232897     NA's   :227098
## pft_days_logged_2024_10 pft_visits_2024_10 pft_assignment_submissions_2024_10
## Min.   : 1.00      Min.   : 1.00      Min.   : 1.00
## 1st Qu.: 3.00      1st Qu.: 5.00      1st Qu.: 1.00
## Median : 7.00      Median : 11.00     Median : 2.00
## Mean   : 7.68      Mean   : 14.14      Mean   : 2.19
## 3rd Qu.:11.00      3rd Qu.: 20.00     3rd Qu.: 3.00
## Max.   :31.00      Max.   :201.00     Max.   :19.00
## NA's   :227065     NA's   :227219     NA's   :267963
## pft_test_submissions_2024_10 pft_total_minutes_2024_10 n_wifi_days_2024_10
## Min.   : 1.00      Min.   : -3.98      Min.   : 1.00
## 1st Qu.: 1.00      1st Qu.: 194.17     1st Qu.:11.00
## Median : 2.00      Median : 471.77     Median :15.00
## Mean   : 2.75      Mean   : 625.34      Mean   :14.41
## 3rd Qu.: 3.00      3rd Qu.: 862.40     3rd Qu.:18.00

```


## Max. :69.00	Max. :8495.65	Max. :30.00
## NA's :271082	NA's :227562	NA's :108935
## resource_events_2024_10	n_resource_days_2024_10	pft_events_2024_11
## Min. : 1.00	Min. : 1.00	Min. : 1.0
## 1st Qu.: 8.00	1st Qu.: 3.00	1st Qu.: 9.0
## Median : 17.00	Median : 5.00	Median : 27.0
## Mean : 23.16	Mean : 6.04	Mean : 48.1
## 3rd Qu.: 30.00	3rd Qu.: 8.00	3rd Qu.: 58.0
## Max. :1084.00	Max. :29.00	Max. :1908.0
## NA's :232783	NA's :232783	NA's :232018
## pft_days_logged_2024_11	pft_visits_2024_11	pft_assignment_submissions_2024_11
## Min. : 1.00	Min. : 1.00	Min. : 1.00
## 1st Qu.: 2.00	1st Qu.: 3.00	1st Qu.: 1.00
## Median : 5.00	Median : 8.00	Median : 1.00
## Mean : 5.85	Mean : 10.28	Mean : 1.93
## 3rd Qu.: 8.00	3rd Qu.: 14.00	3rd Qu.: 2.00
## Max. :30.00	Max. :264.00	Max. :17.00
## NA's :231988	NA's :232241	NA's :269783
## pft_test_submissions_2024_11	pft_total_minutes_2024_11	n_wifi_days_2024_11
## Min. : 1.00	Min. : -49.44	Min. : 1.00
## 1st Qu.: 1.00	1st Qu.: 137.37	1st Qu.: 1.00
## Median : 2.00	Median : 324.97	Median : 3.00
## Mean : 2.38	Mean : 446.73	Mean : 5.07
## 3rd Qu.: 3.00	3rd Qu.: 611.83	3rd Qu.: 7.00
## Max. :184.00	Max. :9029.56	Max. :27.00
## NA's :274423	NA's :232654	NA's :174609
## resource_events_2024_11	n_resource_days_2024_11	pft_events_2024_12
## Min. : 1.0	Min. : 1.00	Min. : 1.00
## 1st Qu.: 5.0	1st Qu.: 2.00	1st Qu.: 16.00
## Median : 10.0	Median : 4.00	Median : 43.00
## Mean : 15.6	Mean : 4.35	Mean : 68.91
## 3rd Qu.: 20.0	3rd Qu.: 6.00	3rd Qu.: 87.00
## Max. :714.0	Max. :23.00	Max. :3657.00
## NA's :238713	NA's :238713	NA's :228890
## pft_days_logged_2024_12	pft_visits_2024_12	pft_assignment_submissions_2024_12
## Min. : 1.00	Min. : 1.00	Min. : 1.00
## 1st Qu.: 3.00	1st Qu.: 4.00	1st Qu.: 1.00
## Median : 6.00	Median : 10.00	Median : 1.00
## Mean : 7.01	Mean : 13.44	Mean : 1.86
## 3rd Qu.:10.00	3rd Qu.: 19.00	3rd Qu.: 2.00
## Max. :31.00	Max. :179.00	Max. :24.00
## NA's :228859	NA's :229058	NA's :266911
## pft_test_submissions_2024_12	pft_total_minutes_2024_12	n_wifi_days_2024_12
## Min. : 1.00	Min. : -4.7	Min. : 1.00
## 1st Qu.: 1.00	1st Qu.: 187.2	1st Qu.: 5.00
## Median : 1.00	Median : 451.3	Median : 9.00
## Mean : 2.67	Mean : 600.7	Mean : 8.61
## 3rd Qu.: 2.00	3rd Qu.: 836.4	3rd Qu.:12.00
## Max. :131.00	Max. :6680.8	Max. :29.00
## NA's :271635	NA's :229355	NA's :113690
## resource_events_2024_12	n_resource_days_2024_12	pft_events_2025_1
## Min. : 1.00	Min. : 1.00	Min. : 1.00
## 1st Qu.: 7.00	1st Qu.: 3.00	1st Qu.: 10.00
## Median : 17.00	Median : 5.00	Median : 33.00

```
## Mean      : 23.75      Mean      : 5.23      Mean      : 54.02
## 3rd Qu.: 31.00      3rd Qu.: 7.00      3rd Qu.: 70.00
## Max.      :1131.00    Max.      :24.00     Max.      :4158.00
## NA's      :234454    NA's      :234454    NA's      :228902
## pft_days_logged_2025_1 pft_visits_2025_1 pft_assignment_submissions_2025_1
## Min.      : 1.00      Min.      : 1.00     Min.      : 1.00
## 1st Qu.: 2.00      1st Qu.: 3.00     1st Qu.: 1.00
## Median : 5.00      Median : 8.00     Median : 1.00
## Mean      : 5.55      Mean      : 11.41    Mean      : 1.52
## 3rd Qu.: 8.00      3rd Qu.: 16.00    3rd Qu.: 2.00
## Max.      :23.00     Max.      :150.00   Max.      :12.00
## NA's      :228873    NA's      :229355   NA's      :274538
## pft_test_submissions_2025_1 pft_total_minutes_2025_1 n_wifi_days_2025_1
## Min.      : 1.00      Min.      : -8.4     Min.      : 1.00
## 1st Qu.: 1.00      1st Qu.: 136.4     1st Qu.: 4.00
## Median : 1.00      Median : 361.2     Median : 7.00
## Mean      : 3.79      Mean      : 520.3     Mean      : 8.11
## 3rd Qu.: 3.00      3rd Qu.: 732.6     3rd Qu.:11.00
## Max.      :202.00    Max.      :6295.5    Max.      :29.00
## NA's      :276292    NA's      :229700    NA's      :121714
## resource_events_2025_1 n_resource_days_2025_1
## Min.      : 1.00      Min.      : 1.0
## 1st Qu.: 6.00      1st Qu.: 2.0
## Median : 14.00     Median : 4.0
## Mean      : 21.41    Mean      : 4.3
## 3rd Qu.: 29.00     3rd Qu.: 6.0
## Max.      :3314.00   Max.      :21.0
## NA's      :236086    NA's      :236086
```

Eliminación de variables o que no tiene sentido en este contexto, así como la conversión a tipo fecha tanto ed_baja_fecha como fecha_datos

```
s=apply(dataset_ordenado, function(col) all(is.na(col) | (is.character(col) & col == "")))
cat("Eliminadas:", names(dataset_ordenado)[s], "\n")
```

```
## Eliminadas: abandono rendimiento_cuat_b pft_assignment_submissions_2024_7 pft_test_submissions_2024_7
```

```
# Eliminar esas columnas
dataset_ordenado=dataset_ordenado[, !s, drop = FALSE]

dataset_ordenado=select(dataset_ordenado, -c("impagado_curso_mat", "ajuste1", "cred_mat5", "cred_mat6",

dataset_ordenado$fecha_datos=as.POSIXct(dataset_ordenado$fecha_datos, format="%Y-%m-%d %H:%M:%S")
dataset_ordenado$baja_fecha=as.POSIXct(dataset_ordenado$baja_fecha, format="%Y-%m-%d")
```

Extracción de las variables de comportamiento digital en un dataset aparte.

```
#estudiantes=select(dataset_ordenado, c(1:12,15:26,30:31,55:56,58))
poliformat=select(dataset_ordenado, c(1:5,73:131))
summary(poliformat)
```

```

##      dni_hash          asi_hash          tit_hash          titnom
## Length:281540      Length:281540      Length:281540      Length:281540
## Class :character    Class :character    Class :character    Class :character
## Mode  :character    Mode  :character    Mode  :character    Mode  :character
##
##
##
##      asinom          pft_events_2024_7 pft_days_logged_2024_7 pft_visits_2024_7
## Length:281540      Min.   : 1.00      Min.   :1.00      Min.   : 1.00
## Class :character    1st Qu.: 2.00      1st Qu.:1.00      1st Qu.: 1.00
## Mode  :character    Median : 5.00      Median :1.00      Median : 1.00
##                      Mean   : 7.78      Mean   :1.39      Mean   : 2.13
##                      3rd Qu.:10.00      3rd Qu.:1.00      3rd Qu.: 2.00
##                      Max.   :191.00      Max.   :9.00      Max.   :24.00
##                      NA's   :280893      NA's   :280893      NA's   :280898
## pft_total_minutes_2024_7 n_wifi_days_2024_7 resource_events_2024_7
## Min.   :-29.00      Min.   : 1.00      Min.   : 1.0
## 1st Qu.: 45.56      1st Qu.: 1.00      1st Qu.: 1.0
## Median : 47.64      Median : 2.00      Median : 2.0
## Mean   : 80.21      Mean   : 3.05      Mean   : 3.7
## 3rd Qu.: 91.77      3rd Qu.: 3.00      3rd Qu.: 4.0
## Max.   :779.45      Max.   :30.00      Max.   :47.0
## NA's   :280906      NA's   :214252      NA's   :281170
## n_resource_days_2024_7 pft_events_2024_8 pft_days_logged_2024_8
## Min.   :1.00      Min.   : 1.00      Min.   : 1.00
## 1st Qu.:1.00      1st Qu.: 2.00      1st Qu.: 1.00
## Median :1.00      Median : 4.00      Median : 1.00
## Mean   :1.25      Mean   : 6.22      Mean   : 1.27
## 3rd Qu.:1.00      3rd Qu.: 7.00      3rd Qu.: 1.00
## Max.   :6.00      Max.   :212.00      Max.   :17.00
## NA's   :281170      NA's   :278469      NA's   :278469
## pft_visits_2024_8 pft_total_minutes_2024_8 n_wifi_days_2024_8
## Min.   : 1.00      Min.   : -93.98      Min.   : 1.00
## 1st Qu.: 1.00      1st Qu.: 45.39      1st Qu.: 1.00
## Median : 1.00      Median : 46.77      Median : 1.00
## Mean   : 1.81      Mean   : 68.82      Mean   : 2.14
## 3rd Qu.: 2.00      3rd Qu.: 86.84      3rd Qu.: 2.00
## Max.   :29.00      Max.   :1361.40      Max.   :30.00
## NA's   :278473      NA's   :278524      NA's   :275508
## resource_events_2024_8 n_resource_days_2024_8 pft_events_2024_9
## Min.   : 1.00      Min.   :1.00      Min.   : 1.00
## 1st Qu.: 1.00      1st Qu.:1.00      1st Qu.: 9.00
## Median : 2.00      Median :1.00      Median : 22.00
## Mean   : 3.71      Mean   :1.15      Mean   : 33.86
## 3rd Qu.: 4.00      3rd Qu.:1.00      3rd Qu.: 41.00
## Max.   :117.00      Max.   :8.00      Max.   :1140.00
## NA's   :279890      NA's   :279890      NA's   :223406
## pft_days_logged_2024_9 pft_visits_2024_9 pft_assignment_submissions_2024_9
## Min.   : 1.00      Min.   : 1.00      Min.   : 1
## 1st Qu.: 2.00      1st Qu.: 2.00      1st Qu.: 1
## Median : 4.00      Median : 6.00      Median : 1
## Mean   : 4.86      Mean   : 7.95      Mean   : 2
## 3rd Qu.: 7.00      3rd Qu.:11.00      3rd Qu.: 2

```

```

## Max. :30.00 Max. :145.00 Max. :18
## NA's :223370 NA's :223608 NA's :274317
## pft_test_submissions_2024_9 pft_total_minutes_2024_9 n_wifi_days_2024_9
## Min. : 1.00 Min. : -17.27 Min. : 1.00
## 1st Qu.: 1.00 1st Qu.: 94.25 1st Qu.: 8.00
## Median : 2.00 Median : 255.01 Median :11.00
## Mean : 2.33 Mean : 343.17 Mean :10.45
## 3rd Qu.: 3.00 3rd Qu.: 475.60 3rd Qu.:14.00
## Max. :50.00 Max. :4712.84 Max. :29.00
## NA's :274995 NA's :224090 NA's :113911
## resource_events_2024_9 n_resource_days_2024_9 pft_events_2024_10
## Min. : 1.00 Min. : 1 Min. : 1.00
## 1st Qu.: 4.00 1st Qu.: 2 1st Qu.: 15.00
## Median : 9.00 Median : 3 Median : 40.00
## Mean : 12.59 Mean : 4 Mean : 61.39
## 3rd Qu.: 16.00 3rd Qu.: 5 3rd Qu.: 80.00
## Max. :435.00 Max. :25 Max. :2791.00
## NA's :232897 NA's :232897 NA's :227098
## pft_days_logged_2024_10 pft_visits_2024_10 pft_assignment_submissions_2024_10
## Min. : 1.00 Min. : 1.00 Min. : 1.00
## 1st Qu.: 3.00 1st Qu.: 5.00 1st Qu.: 1.00
## Median : 7.00 Median : 11.00 Median : 2.00
## Mean : 7.68 Mean : 14.14 Mean : 2.19
## 3rd Qu.:11.00 3rd Qu.: 20.00 3rd Qu.: 3.00
## Max. :31.00 Max. :201.00 Max. :19.00
## NA's :227065 NA's :227219 NA's :267963
## pft_test_submissions_2024_10 pft_total_minutes_2024_10 n_wifi_days_2024_10
## Min. : 1.00 Min. : -3.98 Min. : 1.00
## 1st Qu.: 1.00 1st Qu.: 194.17 1st Qu.:11.00
## Median : 2.00 Median : 471.77 Median :15.00
## Mean : 2.75 Mean : 625.34 Mean :14.41
## 3rd Qu.: 3.00 3rd Qu.: 862.40 3rd Qu.:18.00
## Max. :69.00 Max. :8495.65 Max. :30.00
## NA's :271082 NA's :227562 NA's :108935
## resource_events_2024_10 n_resource_days_2024_10 pft_events_2024_11
## Min. : 1.00 Min. : 1.00 Min. : 1.0
## 1st Qu.: 8.00 1st Qu.: 3.00 1st Qu.: 9.0
## Median : 17.00 Median : 5.00 Median : 27.0
## Mean : 23.16 Mean : 6.04 Mean : 48.1
## 3rd Qu.: 30.00 3rd Qu.: 8.00 3rd Qu.: 58.0
## Max. :1084.00 Max. :29.00 Max. :1908.0
## NA's :232783 NA's :232783 NA's :232018
## pft_days_logged_2024_11 pft_visits_2024_11 pft_assignment_submissions_2024_11
## Min. : 1.00 Min. : 1.00 Min. : 1.00
## 1st Qu.: 2.00 1st Qu.: 3.00 1st Qu.: 1.00
## Median : 5.00 Median : 8.00 Median : 1.00
## Mean : 5.85 Mean : 10.28 Mean : 1.93
## 3rd Qu.: 8.00 3rd Qu.: 14.00 3rd Qu.: 2.00
## Max. :30.00 Max. :264.00 Max. :17.00
## NA's :231988 NA's :232241 NA's :269783
## pft_test_submissions_2024_11 pft_total_minutes_2024_11 n_wifi_days_2024_11
## Min. : 1.00 Min. : -49.44 Min. : 1.00
## 1st Qu.: 1.00 1st Qu.: 137.37 1st Qu.: 1.00
## Median : 2.00 Median : 324.97 Median : 3.00

```

## Mean : 2.38	Mean : 446.73	Mean : 5.07
## 3rd Qu.: 3.00	3rd Qu.: 611.83	3rd Qu.: 7.00
## Max. :184.00	Max. :9029.56	Max. :27.00
## NA's :274423	NA's :232654	NA's :174609
## resource_events_2024_11	n_resource_days_2024_11	pft_events_2024_12
## Min. : 1.0	Min. : 1.00	Min. : 1.00
## 1st Qu.: 5.0	1st Qu.: 2.00	1st Qu.: 16.00
## Median : 10.0	Median : 4.00	Median : 43.00
## Mean : 15.6	Mean : 4.35	Mean : 68.91
## 3rd Qu.: 20.0	3rd Qu.: 6.00	3rd Qu.: 87.00
## Max. :714.0	Max. :23.00	Max. :3657.00
## NA's :238713	NA's :238713	NA's :228890
## pft_days_logged_2024_12	pft_visits_2024_12	pft_assignment_submissions_2024_12
## Min. : 1.00	Min. : 1.00	Min. : 1.00
## 1st Qu.: 3.00	1st Qu.: 4.00	1st Qu.: 1.00
## Median : 6.00	Median : 10.00	Median : 1.00
## Mean : 7.01	Mean : 13.44	Mean : 1.86
## 3rd Qu.:10.00	3rd Qu.: 19.00	3rd Qu.: 2.00
## Max. :31.00	Max. :179.00	Max. :24.00
## NA's :228859	NA's :229058	NA's :266911
## pft_test_submissions_2024_12	pft_total_minutes_2024_12	n_wifi_days_2024_12
## Min. : 1.00	Min. : -4.7	Min. : 1.00
## 1st Qu.: 1.00	1st Qu.: 187.2	1st Qu.: 5.00
## Median : 1.00	Median : 451.3	Median : 9.00
## Mean : 2.67	Mean : 600.7	Mean : 8.61
## 3rd Qu.: 2.00	3rd Qu.: 836.4	3rd Qu.:12.00
## Max. :131.00	Max. :6680.8	Max. :29.00
## NA's :271635	NA's :229355	NA's :113690
## resource_events_2024_12	n_resource_days_2024_12	pft_events_2025_1
## Min. : 1.00	Min. : 1.00	Min. : 1.00
## 1st Qu.: 7.00	1st Qu.: 3.00	1st Qu.: 10.00
## Median : 17.00	Median : 5.00	Median : 33.00
## Mean : 23.75	Mean : 5.23	Mean : 54.02
## 3rd Qu.: 31.00	3rd Qu.: 7.00	3rd Qu.: 70.00
## Max. :1131.00	Max. :24.00	Max. :4158.00
## NA's :234454	NA's :234454	NA's :228902
## pft_days_logged_2025_1	pft_visits_2025_1	pft_assignment_submissions_2025_1
## Min. : 1.00	Min. : 1.00	Min. : 1.00
## 1st Qu.: 2.00	1st Qu.: 3.00	1st Qu.: 1.00
## Median : 5.00	Median : 8.00	Median : 1.00
## Mean : 5.55	Mean : 11.41	Mean : 1.52
## 3rd Qu.: 8.00	3rd Qu.: 16.00	3rd Qu.: 2.00
## Max. :23.00	Max. :150.00	Max. :12.00
## NA's :228873	NA's :229355	NA's :274538
## pft_test_submissions_2025_1	pft_total_minutes_2025_1	n_wifi_days_2025_1
## Min. : 1.00	Min. : -8.4	Min. : 1.00
## 1st Qu.: 1.00	1st Qu.: 136.4	1st Qu.: 4.00
## Median : 1.00	Median : 361.2	Median : 7.00
## Mean : 3.79	Mean : 520.3	Mean : 8.11
## 3rd Qu.: 3.00	3rd Qu.: 732.6	3rd Qu.:11.00
## Max. :202.00	Max. :6295.5	Max. :29.00
## NA's :276292	NA's :229700	NA's :121714
## resource_events_2025_1	n_resource_days_2025_1	
## Min. : 1.00	Min. : 1.0	

```
## 1st Qu.: 6.00      1st Qu.: 2.0
## Median : 14.00     Median : 4.0
## Mean   : 21.41     Mean   : 4.3
## 3rd Qu.: 29.00     3rd Qu.: 6.0
## Max.   :3314.00    Max.   :21.0
## NA's   :236086     NA's   :236086
```

Ordenación de las variables por mes.

```
filtrar_columnas_por_palabra=function(df, palabra) {
  columnas_filtradas=grep(palabra, names(df), value = TRUE, ignore.case = TRUE)

  df_filtrado=df[, columnas_filtradas, drop = FALSE]

  return(df_filtrado)
}

#a=dataset_ordenado
#dataset_ordenado=a

eventos=filtrar_columnas_por_palabra(poliformat, "pft_events")
logged=filtrar_columnas_por_palabra(poliformat, "logged")
visits=filtrar_columnas_por_palabra(poliformat, "visits")
minutes=filtrar_columnas_por_palabra(poliformat, "minutes")
wifi=filtrar_columnas_por_palabra(poliformat, "wifi")
resource_events=filtrar_columnas_por_palabra(poliformat, "resource_events")
resource_days=filtrar_columnas_por_palabra(poliformat, "resource_days")
assignment=filtrar_columnas_por_palabra(poliformat, "assignment")
test=filtrar_columnas_por_palabra(poliformat, "test")

#Con las columnas finales ordenadas, todo será más sencillo
dataset_ordenado=dataset_ordenado[,1:72]
dataset_ordenado=cbind(dataset_ordenado,eventos,logged,visits,minutes,wifi,resource_days,resource_events)

#Formateamos a 0 para poder mostrar gráficas
dataset_ordenado[,72:131 ]=replace(dataset_ordenado[,72:131 ], is.na(dataset_ordenado[,72:131 ]), 0)

dataset_ordenado[,c("discapacidad","desplazado","es_retitulado","exento_npp","es_adaptado") ]=replace(d

#a=dataset_ordenado
#dataset_ordenado=a
```

Vamos a hacer una categoría especial para la preferencia de selección, en la que, a partir de 4, la preferencia es baja.

```
dataset_ordenado$preferencia_seleccion=ifelse(dataset_ordenado$preferencia_seleccion>3,20,dataset_ordenado$preferencia_seleccion)
dataset_ordenado$preferencia_seleccion=as.factor(dataset_ordenado$preferencia_seleccion)
dataset_ordenado$preferencia_seleccion=ifelse(dataset_ordenado$preferencia_seleccion=="20","Baja",dataset_ordenado$preferencia_seleccion)

table(dataset_ordenado$preferencia_seleccion)

##
##      1      2      3  Baja
## 143124 34338 18050 28410
```

Creación de la variable “abandono” y su dataset asociado, ya filtrado para alumnos de ingeniería informática. Debe ser antes de filtrar por cuatrimestres, ya que, si no, se eliminarán asignaturas de las que alumnos se han desmatriculado.

```
mes_abandono_por_alumno=dataset_ordenado[dataset_ordenado$tit_hash=="0fecf9247f3d",] %>%
  filter(!is.na(baja_fecha)) %>%
  mutate(mes = month(baja_fecha, label = TRUE, abbr = FALSE),ordered=FALSE) %>%
  group_by(dni_hash, mes) %>%
  summarise(asi_left = n(), .groups = "drop") %>%
  group_by(dni_hash) %>%
  slice_max(asi_left, n = 1, with_ties = FALSE) %>%
  ungroup()

fecha=dataset_ordenado[dataset_ordenado$tit_hash=="0fecf9247f3d",] %>%
  filter(!is.na(baja_fecha)) %>%
  group_by(dni_hash, baja_fecha) %>%
  summarise(n = n(), .groups = "drop") %>%
  group_by(dni_hash) %>%
  slice_max(order_by = n, n = 1, with_ties = FALSE) %>%
  select(-n)

conteo_abandonos=dataset_ordenado[dataset_ordenado$tit_hash=="0fecf9247f3d",] %>%
  group_by(dni_hash) %>%
  summarise(
    asi_left = sum(!is.na(baja_fecha)),
    abandono = ifelse(asi_left > 5, 1, 0),
    .groups = "drop"
  )

abandono=dataset_ordenado[dataset_ordenado$tit_hash=="0fecf9247f3d",] %>%
  distinct(dni_hash, tit_hash, .keep_all = FALSE) %>%
  left_join(conteo_abandonos, by = "dni_hash") %>%
  left_join(fecha, by = "dni_hash") %>%
  left_join(mes_abandono_por_alumno %>% select(dni_hash, mes), by = "dni_hash")

abandono[abandono$asi_left>12,]
```

```
##          dni_hash      tit_hash asi_left abandono baja_fecha      mes
## 1497 c9c83e27c82f 0fecf9247f3d      29          1 2024-09-24 septiembre
```

```
abandono=abandono[abandono$tit_hash=="0fecf9247f3d" & abandono$abandono==1,]
abandono=select(.data = abandono, -c(tit_hash))

abandono$mes=factor(abandono$mes,ordered =FALSE)
```

Filtramos por asignaturas del primer cuatrimestre, ya que no tenemos las demás. No conseguí sacarlas mediante web scrapping de la web de la UPV, así que de momento la he cogido a mano.

```
codigos_A=c(11540, 11542, 11541, 11545, 11547, 11548, 11549, 11553, 11555, 11557, 11560, 11561, 11563,

dataset_ordenado=dataset_ordenado %>%
  left_join(asignaturas[, c("asi", "asi_hash")], by = "asi_hash") %>%
```

```
filter(asi %in% codigos_A) %>%
select(-asi)
```

Reparación de valores muy extremos.

```
dataset_ordenado$resource_events_2024_10[dataset_ordenado$resource_events_2024_10>400]=400

tests=dataset_ordenado[,127:131]
tests[tests>30]=30
dataset_ordenado[,127:131]=tests
```

Creación de las variables suma de los comportamientos para una asignatura concreta a lo largo del mes.

```
dataset_ordenado$evento_poliformat_asg=rowSums(dataset_ordenado[,73:79])
dataset_ordenado$loggings_asg=rowSums(dataset_ordenado[,80:86])
dataset_ordenado$visitas_asg=rowSums(dataset_ordenado[,87:93])
dataset_ordenado$minutos_asg=rowSums(dataset_ordenado[,94:100])
dataset_ordenado$n_res_day_asg=rowSums(dataset_ordenado[,108:114])
dataset_ordenado$n_res_events_asg=rowSums(dataset_ordenado[,115:121])
dataset_ordenado$assign_asg=rowSums(dataset_ordenado[,122:126])
dataset_ordenado$test_asg=rowSums(dataset_ordenado[,127:131])
```

Creación de las columnas de comportamiento digital por estudiante. Son las suma de todas las interacciones de un ámbito concreto para todas las asignaturas.

```
columnas_a_sumar=colnames(dataset_ordenado[,c(73:100,108:131)])

dataset_ordenado=dataset_ordenado %>%
  group_by(dni_hash) %>%
  mutate(across(all_of(columnas_a_sumar), ~ sum(., na.rm = TRUE), .names = "{.col}_est")) %>%
  ungroup()

head(dataset_ordenado[,c(73:100,108:131)])
```

```
## # A tibble: 6 x 52
##   pft_events_2024_7 pft_events_2024_8 pft_events_2024_9 pft_events_2024_10
##           <dbl>           <dbl>           <dbl>           <dbl>
## 1             0             0             0             0
## 2             0             0            17            38
## 3             0             0             9            79
## 4             0             0             0             0
## 5             0             0             0             0
## 6             0             0             0             0
## # i 48 more variables: pft_events_2024_11 <dbl>, pft_events_2024_12 <dbl>,
## #   pft_events_2025_1 <dbl>, pft_days_logged_2024_7 <dbl>,
## #   pft_days_logged_2024_8 <dbl>, pft_days_logged_2024_9 <dbl>,
## #   pft_days_logged_2024_10 <dbl>, pft_days_logged_2024_11 <dbl>,
## #   pft_days_logged_2024_12 <dbl>, pft_days_logged_2025_1 <dbl>,
## #   pft_visits_2024_7 <dbl>, pft_visits_2024_8 <dbl>, pft_visits_2024_9 <dbl>,
## #   pft_visits_2024_10 <dbl>, pft_visits_2024_11 <dbl>, ...
```

Filtramos ya por informática en el dataset, eliminando el nombre del título y su código hash.


```
dataset_ordenado= dataset_ordenado[dataset_ordenado$tit_hash=="0fecf9247f3d",]
```

```
dataset_ordenado=select(dataset_ordenado, -c("tit_hash", "titnom"))
```

Vale, aquí vamos a separar en varios dataset y los concatenaremos y haremos los últimos retoques.

```
#tabla=as.data.frame(table(df))  
#tabla=tabla[order(-tabla$Freq), ]
```

```
columnas utiles=c(  
  "dni_hash", "asi_hash", "asinom", "nacionalitat",  
  "data_nac", "sexe", "alta_universitat", "prov_origen", "anyo_ingreso", "tipo_ingreso",  
  "nota10", "nota14", "campus", "estudios_p", "estudios_m", "dedicacion",  
  "desplazado", "discapacidad", "becado", "preferencia_seleccion", "baja_fecha", "caca",  
  "grupos_por_tipocredito", "matricula_activa", "nota_asig", "fecha_datos", "cod_centro",  
  "curso_mas_bajo", "curso_mas_alto", "cred_mat1", "cred_mat2", "cred_mat3", "cred_mat4",  
  "cred_sup_normal", "cred_sup_espec", "cred_sup",  
  "cred_mat_normal", "cred_mat_movilidad", "cred_ptes_acta", "cred_mat_practicas",  
  "cred_mat_sem_a", "cred_mat_sem_b", "cred_mat_anu", "cred_mat_total",  
  "cred_sup_sem_a", "cred_sup_sem_b", "cred_sup_total",  
  "rendimiento_cuat_a", "rendimiento_total", "exento_npp", "anyo_inicio_estudios",  
  "es_retitulado", "es_adaptado", "cred_sup_1o", "cred_sup_2o", "cred_sup_3o",  
  "cred_sup_4o", "practicas", "actividades", "ajuste",  
  "cred_sup_tit", "cred_pend_sup_tit", "impagado_curso_mat", "asig1", "pract1", "activ1",  
  "total1", "ajuste1", "rend_total_ultimo", "rend_total_penultimo", "rend_total_antepenultimo",  
  "pft_events_2024_7", "pft_events_2024_8", "pft_events_2024_9", "pft_events_2024_10",  
  "pft_events_2024_11", "pft_events_2024_12", "pft_events_2025_1", "pft_days_logged_2024_7",  
  "pft_days_logged_2024_8", "pft_days_logged_2024_9", "pft_days_logged_2024_10",  
  "pft_days_logged_2024_11", "pft_days_logged_2024_12", "pft_days_logged_2025_1",  
  "pft_visits_2024_7", "pft_visits_2024_8", "pft_visits_2024_9", "pft_visits_2024_10",  
  "pft_visits_2024_11", "pft_visits_2024_12", "pft_visits_2025_1", "pft_total_minutes_2024_7",  
  "pft_total_minutes_2024_8", "pft_total_minutes_2024_9", "pft_total_minutes_2024_10",  
  "pft_total_minutes_2024_11", "pft_total_minutes_2024_12", "pft_total_minutes_2025_1",  
  "n_wifi_days_2024_7", "n_wifi_days_2024_8", "n_wifi_days_2024_9", "n_wifi_days_2024_10",  
  "n_wifi_days_2024_11", "n_wifi_days_2024_12", "n_wifi_days_2025_1",  
  "n_resource_days_2024_7", "n_resource_days_2024_8", "n_resource_days_2024_9",  
  "n_resource_days_2024_10", "n_resource_days_2024_11", "n_resource_days_2024_12",  
  "n_resource_days_2025_1", "resource_events_2024_7", "resource_events_2024_8",  
  "resource_events_2024_9", "resource_events_2024_10", "resource_events_2024_11",  
  "resource_events_2024_12", "resource_events_2025_1", "pft_assignment_submissions_2024_9",  
  "pft_assignment_submissions_2024_10", "pft_assignment_submissions_2024_11",  
  "pft_assignment_submissions_2024_12", "pft_assignment_submissions_2025_1",  
  "pft_test_submissions_2024_9", "pft_test_submissions_2024_10", "pft_test_submissions_2024_11",  
  "pft_test_submissions_2024_12", "pft_test_submissions_2025_1", "evento_poliformat_asg",  
  "loggin_asg", "visitas_asg", "minutos_asg", "n_res_day_asg", "n_res_events_asg",  
  "assign_asg", "test_asg", "pft_events_2024_7_est", "pft_events_2024_8_est",  
  "pft_events_2024_9_est", "pft_events_2024_10_est", "pft_events_2024_11_est",  
  "pft_events_2024_12_est", "pft_events_2025_1_est", "pft_days_logged_2024_7_est",  
  "pft_days_logged_2024_8_est", "pft_days_logged_2024_9_est", "pft_days_logged_2024_10_est",  
  "pft_days_logged_2024_11_est", "pft_days_logged_2024_12_est", "pft_days_logged_2025_1_est",  
  "pft_visits_2024_7_est", "pft_visits_2024_8_est", "pft_visits_2024_9_est",
```

```

"pft_visits_2024_10_est", "pft_visits_2024_11_est", "pft_visits_2024_12_est",
"pft_visits_2025_1_est", "pft_total_minutes_2024_7_est", "pft_total_minutes_2024_8_est",
"pft_total_minutes_2024_9_est", "pft_total_minutes_2024_10_est",
"pft_total_minutes_2024_11_est", "pft_total_minutes_2024_12_est", "pft_total_minutes_2025_1_est",
"n_resource_days_2024_7_est", "n_resource_days_2024_8_est", "n_resource_days_2024_9_est",
"n_resource_days_2024_10_est", "n_resource_days_2024_11_est", "n_resource_days_2024_12_est",
"n_resource_days_2025_1_est", "resource_events_2024_7_est", "resource_events_2024_8_est",
"resource_events_2024_9_est", "resource_events_2024_10_est", "resource_events_2024_11_est",
"resource_events_2024_12_est", "resource_events_2025_1_est", "pft_assignment_submissions_2024_9_est",
"pft_assignment_submissions_2024_10_est", "pft_assignment_submissions_2024_11_est",
"pft_assignment_submissions_2024_12_est", "pft_assignment_submissions_2025_1_est",
"pft_test_submissions_2024_9_est", "pft_test_submissions_2024_10_est",
"pft_test_submissions_2024_11_est", "pft_test_submissions_2024_12_est",
"pft_test_submissions_2025_1_est")

demografía=dataset_ordenado[,c("dni_hash", "nacionalitat",
"data_nac", "sexe", "alta_universitat", "prov_origen", "anyo_ingreso", "tipo_ingreso",
"nota10", "nota14", "estudios_p", "estudios_m", "dedicacion",
"desplazado", "discapacidad", "becado", "preferencia_seleccion")]

creditos=dataset_ordenado[,c( "dni_hash",
"curso_mas_bajo", "curso_mas_alto", "cred_mat1", "cred_mat2", "cred_mat3", "cred_mat4", "cred_sup_norma",
"cred_mat_normal", "cred_mat_movilidad", "cred_ptes_acta", "cred_mat_practicas",
"cred_mat_sem_a", "cred_mat_sem_b", "cred_mat_anu", "cred_mat_total",
"cred_sup_sem_a", "cred_sup_sem_b", "cred_sup_total",
"rendimiento_cuat_a", "rendimiento_total", "exento_npp", "anyo_inicio_estudios",
"es_retitulado", "es_adaptado", "cred_sup_1o", "cred_sup_2o", "cred_sup_3o",
"cred_sup_4o", "practicas", "actividades", "ajuste",
"cred_sup_tit", "cred_pend_sup_tit", "asig1", "pract1", "activ1",
"total1", "rend_total_ultimo", "rend_total_penultimo", "rend_total_antepenultimo")]

poliformat=dataset_ordenado[,c("dni_hash", "asinom", "evento_poliformat_asg",
"loggins_asg", "visitas_asg", "minutos_asg", "n_res_day_asg", "n_res_events_asg",
"assign_asg", "test_asg", "pft_events_2024_7_est", "pft_events_2024_8_est",
"pft_events_2024_9_est", "pft_events_2024_10_est", "pft_events_2024_11_est",
"pft_events_2024_12_est", "pft_events_2025_1_est", "pft_days_logged_2024_7_est",
"pft_days_logged_2024_8_est", "pft_days_logged_2024_9_est", "pft_days_logged_2024_10_est",
"pft_days_logged_2024_11_est", "pft_days_logged_2024_12_est", "pft_days_logged_2025_1_est",
"pft_visits_2024_7_est", "pft_visits_2024_8_est", "pft_visits_2024_9_est",
"pft_visits_2024_10_est", "pft_visits_2024_11_est", "pft_visits_2024_12_est",
"pft_visits_2025_1_est", "pft_total_minutes_2024_7_est", "pft_total_minutes_2024_8_est",
"pft_total_minutes_2024_9_est", "pft_total_minutes_2024_10_est",
"pft_total_minutes_2024_11_est", "pft_total_minutes_2024_12_est", "pft_total_minutes_2025_1_est",
"n_wifi_days_2024_7", "n_wifi_days_2024_8", "n_wifi_days_2024_9", "n_wifi_days_2024_10",
"n_wifi_days_2024_11", "n_wifi_days_2024_12", "n_wifi_days_2025_1",
"n_resource_days_2024_7_est", "n_resource_days_2024_8_est", "n_resource_days_2024_9_est",
"n_resource_days_2024_10_est", "n_resource_days_2024_11_est", "n_resource_days_2024_12_est",
"n_resource_days_2025_1_est", "resource_events_2024_7_est", "resource_events_2024_8_est",
"resource_events_2024_9_est", "resource_events_2024_10_est", "resource_events_2024_11_est",
"resource_events_2024_12_est", "resource_events_2025_1_est", "pft_assignment_submissions_2024_9_est",
"pft_assignment_submissions_2024_10_est", "pft_assignment_submissions_2024_11_est",
"pft_assignment_submissions_2024_12_est", "pft_assignment_submissions_2025_1_est",
"pft_test_submissions_2024_9_est", "pft_test_submissions_2024_10_est",

```

```
"pft_test_submissions_2024_11_est", "pft_test_submissions_2024_12_est",
"pft_test_submissions_2025_1_est"]]
```

Retocamos los datasets. Para poliformat, lo hacemos único por estudiante y hacemos la media de las variables asg, eliminando las originales.

A todas las concatenamos abandono desde el dataset de abandono, y hacemos el filtro de estudiantes que tenían asignaturas del primer cuatri (no tenemos datos del segundo cuatrimestre).

```
poliformat= poliformat %>%
  left_join(select(abandono, dni_hash, abandono), by = "dni_hash") %>% unique()

poliformat=poliformat %>%
  group_by(dni_hash) %>%
  mutate(across(ends_with("_asg"), ~ mean(.x, na.rm = TRUE), .names = "{.col}_media")) %>%
  select(-c("evento_poliformat_asg",
    "loggin_asg", "visitas_asg", "minutos_asg", "n_res_day_asg", "n_res_events_asg",
    "assign_asg", "test_asg", "asinom", "minutos_asg"))

poliformat=unique(poliformat)

sociodemografia=demografia %>%
  distinct(dni_hash, .keep_all = TRUE) %>%
  left_join(select(abandono, dni_hash, abandono), by = "dni_hash") %>% unique()

academicas=creditos %>%
  distinct(dni_hash, .keep_all = TRUE) %>%
  left_join(select(abandono, dni_hash, abandono), by = "dni_hash") %>% unique()

poliformat$abandono[is.na(poliformat$abandono)]=0
sociodemografia$abandono[is.na(sociodemografia$abandono)]=0
academicas$abandono[is.na(academicas$abandono)]=0

dni_abandono_1=sociodemografia %>%
  filter(abandono == 1) %>%
  pull(dni_hash)

abandono=abandono %>%
  filter(dni_hash %in% dni_abandono_1)

poliformat[poliformat<0]=0
poliformat$minutos_asg=rowSums(poliformat[,24:30])
```

Ahora, con todo mucho más dividido, cambiamos los tipos character a tipo factor.

```
sociodemografia[, c("sexe", "prov_origen", "tipo_ingreso",
  "estudios_p", "estudios_m", "dedicacion", "desplazado",
  "discapacidad", "becado", "preferencia_seleccion", "abandono", "nacionalitat")]=
  lapply(sociodemografia[, c("sexe", "prov_origen", "tipo_ingreso",
    "estudios_p", "estudios_m", "dedicacion", "desplazado",
```

```

      "discapacidad", "becado", "preferencia_seleccion", "abandono", "nacionalita
academicas[, c("exento_npp", "es_adaptado", "es_retitulado", "curso_mas_bajo", "curso_mas_alto")] =
  lapply(academicas[, c("exento_npp", "es_adaptado", "es_retitulado", "curso_mas_bajo", "curso_mas_alto")],

```

Guardamos los datos

```

save( academicas, sociodemografia, poliformat, abandono, file="../Datos/Capítulos/Preprocesado.RData".

```