

Problem set 1

Cristopher Urbina H.

2024-08-27

Table of contents

1	Estadística Descriptiva	1
1.1	Pregunta 1	1
1.1.1	Diccionario de variables	1
1.1.2	Información del Dataframe	2
1.2	Pregunta 2	3
1.3	Pregunta 3	4

1 Estadística Descriptiva

1.1 Pregunta 1

La unidad de análisis son los estudiantes desvinculados del sistema escolar, y como impacta sobre ellos la política de reportes del MINEDUC en la probabilidad de revinculación y graduación. Para esto, se cuenta con una muestra de 8516 estudiantes desvinculados entre abril y junio del 2022.

Respecto, a las variables IDstudent esta se refiere al identificador único de cada estudiante, y la variable IDschool se refiere al identificador único de cada escuela.

1.1.1 Diccionario de variables

- **IDstudent**: Identificador único de cada estudiante.
- **IDschool**: Identificador único de cada escuela.
- **SchoolGrade2022**: Grado escolar en el que se encontraba el estudiante al momento de la desvinculación. -**DropoutDate**: Fecha de desvinculación del estudiante.
- **DayOfWeekDroppedOut**: Día de la semana en que se desvinculó el estudiante.
- **PrioritaryStudentIn2021**: Variable binaria que indica si el estudiante fue considerado prioritario en el año 2021.
- **Female**: Variable binaria que indica si el estudiante es mujer.
- **Age**: Edad del estudiante al momento de la desvinculación.
- **GPAin2021**: Promedio de notas del estudiante en el año 2021.

- **AttendanceIn2021:** Asistencia del estudiante en el año 2021.
- **GraduatedIn2021:** Variable binaria que indica si el estudiante se graduó en el año 2021.
- **Migrant:** Variable binaria que indica si el estudiante es migrante.
- **ListedInDropoutReport:** Variable binaria que indica si el estudiante fue listado en el reporte de desvinculación del MINEDUC.
- **SchoolPrincipalADP:** Variable binaria que indica si el director del colegio es ADP.
- **EnrolledByAug312022:** Variable binaria que indica si el estudiante se matriculó en algún establecimiento educacional antes del 31 de agosto del 2022.
- **GraduatedIn2022:** Variable binaria que indica si el estudiante se graduó en el año 2022.
- **MunicipalSchool:** Variable binaria que indica si el establecimiento educacional es municipal.
- **SchoolDaysFromMay302022:** Días que han pasado desde el 30 de mayo del 2022 hasta su desvinculación.
- **DroppedOutMay30Before:** Variable binaria que indica si el estudiante se desvinculó antes del 30 de mayo del 2022.

1.1.2 Información del Dataframe

#	Column	Non-Null Count	Dtype
0	IDstudent	8516 non-null	float32
1	IDSchool	8516 non-null	int32
2	SchoolGrade2022	8516 non-null	float32
3	DropoutDate	8516 non-null	datetime64[ns]
4	DayOfWeekDroppedOut	8516 non-null	float32
5	PrioritaryStudentIn2021	8516 non-null	int8
6	Female	8516 non-null	int8
7	Age	8516 non-null	int8
8	GPAIN2021	8416 non-null	float64
9	AttendanceIn2021	8516 non-null	int8
10	GraduatedIn2021	8516 non-null	float32
11	Migrant	8516 non-null	float32
12	ListedInDropoutReport	8516 non-null	float32
13	SchoolPrincipalADP	8516 non-null	float32
14	EnrolledByAug312022	8516 non-null	float32
15	GraduatedIn2022	8516 non-null	float32
16	MunicipalSchool	8516 non-null	float32
17	SchoolDaysFromMay302022	8516 non-null	int16
18	DroppedOutMay30Before	8516 non-null	float32

1.2 Pregunta 2

A continuación se muestra un resumen estadístico de las variables **Age**, **Female**, **AttendanceIn2021**, **GPain2021** y **MunicipalSchool**.

	Age	Female	AttendanceIn2021	GPain2021	MunicipalSchool
count	8516.000	8516.000	8516.000	8416.000	8516.000
mean	12.054	0.531	82.943	5.514	0.539
std	3.460	0.499	22.169	1.018	0.499
min	6.000	0.000	0.000	1.000	0.000
25%	9.000	0.000	85.000	4.900	0.000
50%	13.000	1.000	86.000	5.600	1.000
75%	15.000	1.000	96.000	6.300	1.000
max	19.000	1.000	100.000	7.000	1.000

Ahora, se presenta la distribución univariada de cada variable.

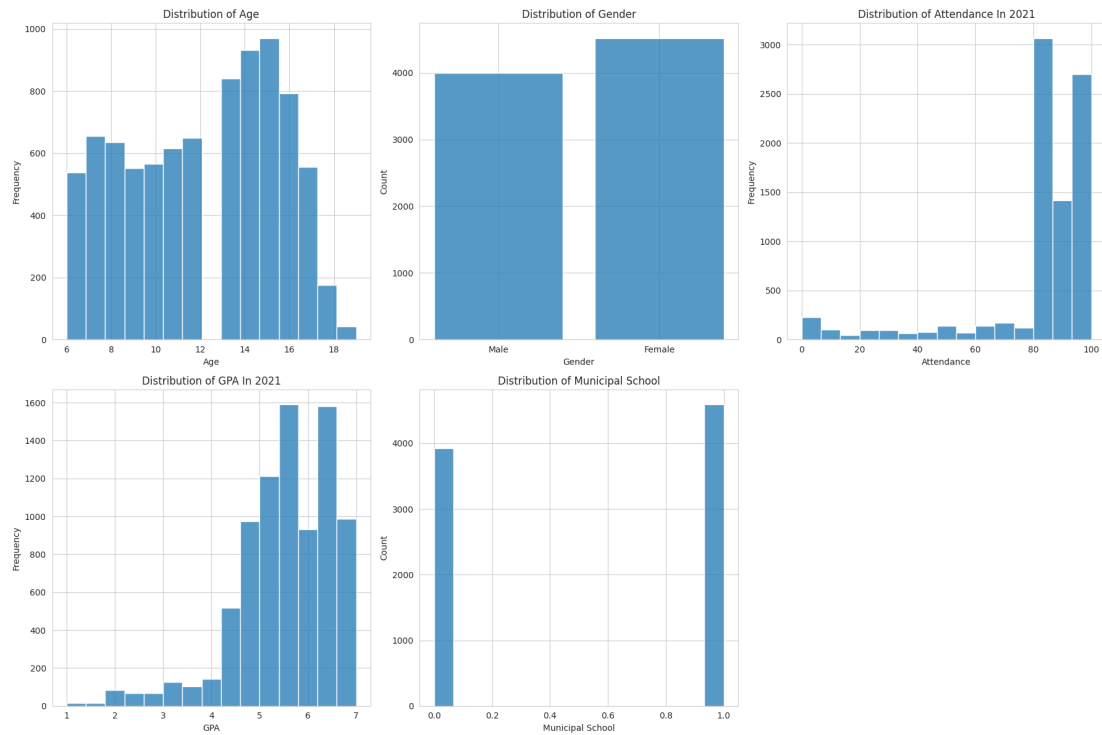


Figure 1: Distribución univariada de cada variable.

Respecto al tamaño muestral de cada variable, solo la variable **GPain2021** presenta valores faltantes, con un total de 100 valores faltantes.

1.3 Pregunta 3

Separando el conjunto de datos en migrantes y no migrantes, se obtiene la siguiente información. Para la población migrante:

	Age	Female	AttendanceIn2021	GPAin2021	MunicipalSchool
count	998.000000	998.000000	998.000000	985.000000	998.000000
mean	11.187375	0.553106	83.623246	5.666701	0.675351
std	3.367787	0.497421	21.544625	0.991215	0.468478
min	6.000000	0.000000	0.000000	1.000000	0.000000
25%	8.000000	0.000000	85.000000	5.100000	0.000000
50%	11.000000	1.000000	85.000000	5.800000	1.000000
75%	14.000000	1.000000	96.000000	6.400000	1.000000
max	19.000000	1.000000	100.000000	7.000000	1.000000

Luego, para la población no migrante:

	Age	Female	AttendanceIn2021	GPAin2021	MunicipalSchool
count	7518.000000	7518.000000	7518.000000	7431.000000	7518.000000
mean	12.169460	0.527534	82.853019	5.493487	0.520883
std	3.455722	0.499275	22.250011	1.020100	0.499597
min	6.000000	0.000000	0.000000	1.000000	0.000000
25%	9.000000	0.000000	85.000000	4.900000	0.000000
50%	13.000000	1.000000	86.000000	5.600000	1.000000
75%	15.000000	1.000000	96.000000	6.300000	1.000000
max	19.000000	1.000000	100.000000	7.000000	1.000000

Al separar la población en migrantes y no migrantes, se observa que la población migrante tiene un promedio de edad menor que la población no migrante, con un promedio de 11.18 años y 12.17 años respectivamente. Además, la población migrante tiene un promedio de asistencia mayor que la población no migrante, con un promedio de 83.62% y 82.85% respectivamente. Por otro lado, la población migrante tiene un promedio de notas mayor que la población no migrante, con un promedio de 5.67 y 5.49 respectivamente. Por último, se observa que la población migrante tiene una mayor proporción de estudiantes en establecimientos municipales, con un 67.53% y 52.08% respectivamente.