

Curso de Big Data

Trabajo Practico 4

Integrantes del grupo: Sebastián Einstoss, Sofía Marinkovic y Pilar Ruiz Orrico

Parte 1: Análisis de la base de hogares y cálculo de pobreza

Ejercicio 1

	CODUSU	ANO4	TRIMESTRE	NRO_HOGAR	REALIZADA	REGION	MAS_500
0	TQRMNOQUPHMKKUCDEJAH00701956	2021	1	1	1	44	N
1	TQRMNOPWXHMKRCDEGLDF00701361	2021	1	1	1	41	N

2 rows × 88 columns

	CODUSU	ANO4	TRIMESTRE	NRO_HOGAR	REALIZADA	REGION	MAS_500
15	TQRMNOQXUJHJMMUCDEJAH00693031	2021	1	1	1	1	S
21	TQRMNOPRPHLLNRCDEJAH00651171	2021	1	1	1	1	S
24	TQRMNOPVUHKLMNCDEIAD00655817	2021	1	1	1	1	S
36	TQRMNOPRYTHLOPMCDEJAH00655933	2021	1	1	1	1	S
43	TQRMNOOPVHLMNNCDEIAD00650542	2021	1	1	1	1	S

5 rows × 88 columns

	CODUSU	ANO4	TRIMESTRE	NRO_HOGAR	COMPONENTE	H15	REGION	M
0	TQRMNOPQHLLKMCDEFIAH00646702	2021	1	1		1	1	43
1	TQRMNOPPRHLLKMCDEFIAH00665104	2021	1	1		1	1	43

2 rows × 177 columns

	CODUSU	ANO4	TRIMESTRE	NRO_HOGAR	COMPONENTE	H15	REGIOI
36799	TQRMNOPPRHKMLNCDEIAD00655703	2021	1	1		1	1
36800	TQRMNOPPRHKMLNCDEIAD00655703	2021	1	1		2	1
36801	TQRMNOPPWHKMNUCDEIAD00655837	2021	1	1		1	1
36802	TQRMNOPPWHKMNUCDEIAD00655837	2021	1	1		2	1
36803	TQRMNOPPWHKMNUCDEIAD00655837	2021	1	1		3	1

5 rows × 177 columns

Ejercicio 2

Longitud base hogares: 1474 ;Longitud base individual: 4082

4082

Ejercicio 3

En este inciso filtramos aquellos valores que no tiene sentido, como por ejemplo: Edades negativas y variables de ingreso que tomen valores negativos.

También reemplazamos los valores faltantes por la mediana y dropeamos variables que no consideramos relevantes. En el código se aclara cuales son.

Finalmente para realizar el analisis convertimos las variables categoricas en dummies.

	CODUSU	ANO4	TRIMESTRE	NRO_HOGAR	COMPONENTE	H15	REGION	M
0	TQRMNOPPRHKMLNCDEIAD00655703	2021	1	1		1	1	1
1	TQRMNOPPRHKMLNCDEIAD00655703	2021	1	1		2	1	1

2 rows × 242 columns

	CODUSU	NRO_HOGAR	REGION	MAS_500	AGLOMERADO	CH04	CH06	PP0
0	TQRMNOPPRHKMLNCDEIAD00655703	1	1	1		32	1	29
1	TQRMNOPPRHKMLNCDEIAD00655703	1	1	1		32	2	29

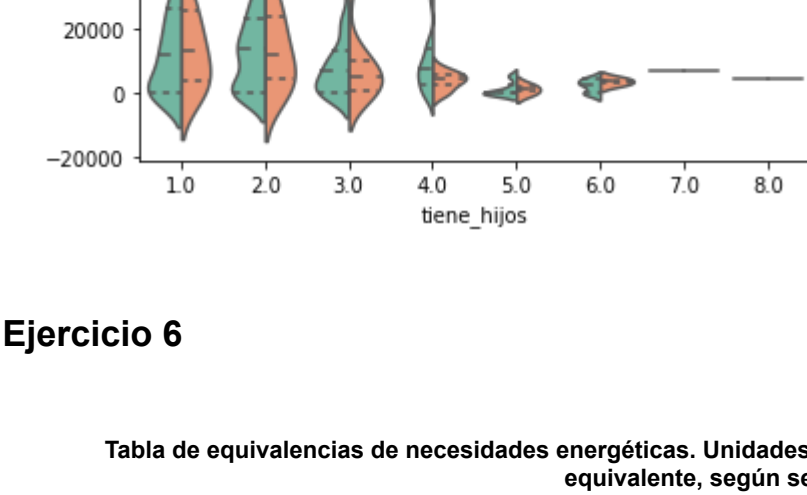
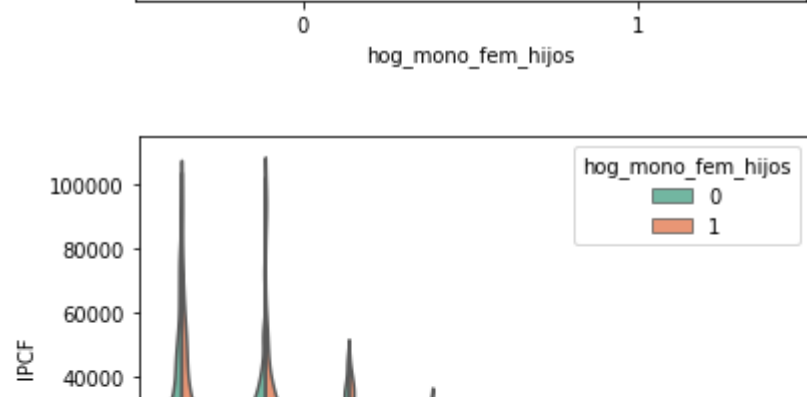
2 rows × 1033 columns

Ejercicio 4 - Construcción de variables

Generamos dos variables para tener en cuenta a la hora de analizar la pobreza. La primera es si hay niños en el hogar (los cuales representan "una boca más que alimentar" y podría pensarse que hay una relación entre la cantidad de hijos y la pobreza), y la segunda es si se trata de un hogar del tipo monoparental con jefe de hogar femenino (intentando captar un efecto de la feminización de la pobreza)

(210, 1035)

Ejercicio 5 - Gráfico



Ejercicio 6

	Tabla de equivalencias de necesidades energéticas. Unidades de adulto equivalente, según sexo y edad	Unnamed: 1	Unnamed: 2	Unnamed: 3
0		NaN	NaN	NaN
1		NaN	NaN	NaN
2	Edad	Mujeres	Varones	NaN
3		NaN	NaN	NaN
4	Menor de 1 año	0.35	0.35	NaN

	Tabla de equivalencias de necesidades energéticas. Unidades de adulto equivalente, según sexo y edad	Unnamed: 1	Unnamed: 2
0		NaN	NaN
1		NaN	NaN
2	Edad	Mujeres	Varones
3		NaN	NaN
4	Menor de 1 año	0.35	0.35

	Tabla de equivalencias de necesidades energéticas. Unidades de adulto equivalente, según sexo y edad	Unnamed: 1	Unnamed: 2
2	Edad	Mujeres	Varones
4	Menor de 1 año	0.35	0.35
5	1año	0.37	0.37
6	2 años	0.46	0.46
7	3 años	0.51	0.51
8	4 años	0.55	0.55

	Edad	sexo	value
0	Menor de 1 año	Mujeres	0.35
1	1año	Mujeres	0.37
2	2 años	Mujeres	0.46
3	3 años	Mujeres	0.51
4	4 años	Mujeres	0.55

	Edad	sexo	value
0	Menor de 1 año	Mujeres	0.35
1	1año	Mujeres	0.37
2	2 años	Mujeres	0.46
3	3 años	Mujeres	0.51
4	4 años	Mujeres	0.55

	Edad	sexo	value
0	Menor de 1 año	Mujeres	0.35
1	1año	Mujeres	0.37
2	2 años	Mujeres	0.46
3	3 años	Mujeres	0.51
4	4 años	Mujeres	0.55

	Edad	sexo	value
0	Menor de 1 año	Mujeres	0.35
1	1año	Mujeres	0.37
2	2 años	Mujeres	0.46
3	3 años	Mujeres	0.51
4	4 años	Mujeres	0.55

	CH06	adulto_equiv	CH04
0	0	0.35	2
1	1	0.37	2
2	2	0.46	2
3	3	0.51	2
4	4	0.55	2
5	5	0.60	2
6	6	0.64	2
7	7	0.66	2
8	8	0.68	2
9	9	0.69	2

	CH06	adulto_equiv	CH04
0	0	0.35	2
1	1	0.37	2
2	2	0.46	2
3	3	0.51	2
4	4	0.55	2
5	5	0.60	2
6	6	0.64	2
7	7	0.66	2
8	8	0.68	2
9	9	0.69	2

	CH06	adulto_equiv	CH04
0	0	0.35	2
1	1	0.37	2
2	2	0.46	2
3	3	0.51	2
4	4	0.55	2
5	5	0.60	2
6	6	0.64	2
7	7	0.66	2
8	8	0.68	2
9	9	0.69	2

	CH06	adulto_equiv	CH04
0	0	0.35	2
1	1	0.37	2
2	2	0.46	2
3	3	0.51	2
4	4	0.55	2
5	5	0.60	2
6	6	0.64	2
7	7	0.66	2
8	8	0.68	2
9	9	0.69	2

	CODUSU	NRO_HOGAR	REGION	MAS_500	AGLOMERADO	CH04	CH06	PPi
0	TQRMNOPPRHKMLNCDEIAD00655703	1	1	1		32	1	29
1	TQRMNOPPRHKMLNCDEIAD00655703	1	1	1		32	2	29
2	TQRMNOPPWHKMNUCDEIAD00655837	1	1	1		32	2	53
3	TQRMNOPPWHKMNUCDEIAD00655837	1	1	1		32	2	22
4	TQRMNOPPWHKMNUCDEIAD00655837	1	1	1		32	1	19

5 rows × 1036 columns

	ad_equiv_hogar
CODUSU	
TQRMNOPPQHJMLQCDEJAH00702455	1.67
TQRMNOPPQHJNMQCDEJAH00698190	3.10
TQRMNOPPQHJONQCDEJAH00693114	1.76
TQRMNOPPQHKMOSCEJAH00656008	2.53
TQRMNOPPQHLMPPCDEJAH00701610	3.92

	CODUSU	NRO_HOGAR	REGION	MAS_500	AGLOMERADO	CH04	CH06	PPi
0	TQRMNOPPRHKMLNCDEIAD00655703	1	1	1		32	1	29
1	TQRMNOPPRHKMLNCDEIAD00655703	1	1	1		32	2	29
2	TQRMNOPPWHKMNUCDEIAD00655837	1	1	1		32	2	53
3	TQRMNOPPWHKMNUCDEIAD00655837	1	1	1		32	2	22
4	TQRMNOPPWHKMNUCDEIAD00655837	1	1	1		32	1	19

5 rows × 1037 columns

Ejercicio 7

2474

821

Ejercicio 8

	CODUSU	NRO_HOGAR	REGION	MAS_500	AGLOMERADO	CH04	CH06	PP0
0	TQRMNOPPRHKMLNCDEIAD00655703	1	1	1		32	1	29
1	TQRMNOPPRHKMLNCDEIAD00655703	1	1	1		32	2	29
7	TQRMNOPPXHLLNCDEIAD00690221	1	1	1		32	1	66
8	TQRMNOPPXHLLNCDEIAD00690221	1	1	1		32	1	27
9	TQRMNOPPXHLLNCDEIAD00690221	1	1	1		32	2	65

5 rows × 1038 columns

	CODUSU	NRO_HOGAR	REGION	MAS_500	AGLOMERADO	CH04	CH06	PP0
0	TQRMNOPPRHKMLNCDEIAD00655703	1	1	1		32	1	29
1	TQRMNOPPRHKMLNCDEIAD00655703	1	1	1		32	2	29
7	TQRMNOPPXHLLNCDEIAD00690221	1	1	1		32	1	66
8	TQRMNOPPXHLLNCDEIAD00690221	1	1	1		32	1	27
9	TQRMNOPPXHLLNCDEIAD00690221	1	1	1		32	2	65

5 rows × 1039 columns

	CODUSU	NRO_HOGAR	REGION	MAS_500	AGLOMERADO	CH04	CH06	PP0
0	TQRMNOPPRHKMLNCDEIAD00655703	1	1	1		32	1	29
1	TQRMNOPPRHKMLNCDEIAD00655703	1	1	1		32	2	29
7	TQRMNOPPXHLLNCDEIAD00690221	1	1	1		32	1	66
8	TQRMNOPPXHLLNCDEIAD00690221	1	1	1		32	1	27
9	TQRMNOPPXHLLNCDEIAD00690221	1	1	1		32	2	65

5 rows × 1039 columns

	CODUSU	NRO_HOGAR	REGION	MAS_500	AGLOMERADO	CH04	CH06	PP0
0	TQRMNOPPRHKMLNCDEIAD00655703	1	1	1		32	1	29
1	TQRMNOPPRHKMLNCDEIAD00655703	1	1	1		32	2	29
7	TQRMNOPPXHLLNCDEIAD00690221	1	1	1		32	1	66
8	TQRMNOPPXHLLNCDEIAD00690221	1	1	1		32	1	27
9	TQRMNOPPXHLLNCDEIAD00690221	1	1	1		32	2	65

5 rows × 1039 columns

1868

680

	CH06	adulto_equiv	CH04
0	0	0.35	2
1	1	0.37	2
2	2	0.46	2
3	3	0.51	2
4	4	0.55	2
5	5	0.60	2
6	6	0.64	2
7	7	0.66	2
8	8	0.68	2
9	9	0.69	2

	CH06	adulto_equiv	CH04
0	0	0.35	2
1	1	0.37	2
2	2	0.46	2
3	3	0.51	2
4	4	0.55	2
5	5	0.60	2
6	6	0.64	2
7	7	0.66	2
8	8	0.68	2
9	9	0.69	2

	CH06	adulto_equiv	CH04
0	0	0.35	2
1	1	0.37	2
2	2	0.46	2
3	3	0.51	2
4	4	0.55	2
5	5	0.60	2
6	6	0.64	2
7	7	0.66	2
8	8	0.68	2
9	9	0.69	2

	CH06	adulto_equiv	CH04
0	0	0.35	2
1	1	0.37	2
2	2	0.46	2
3	3	0.51	2
4	4	0.55	2
5	5	0.60	2
6	6	0.64	2
7	7	0.66	2
8	8	0.68	2
9	9	0.69	2

	CH06	adulto_equiv	CH04
0	0	0.35	2
1	1	0.37	2
2	2	0.46	2
3	3	0.51	2
4	4	0.55	2
5	5	0.60	2
6	6	0.64	2
7	7	0.66	2
8	8	0.68	2
9	9	0.69	2

	CH06	adulto_equiv	CH04
0	0	0.35	2
1	1	0.37	2
2	2	0.46	2
3	3	0.51	2
4	4	0.55	2
5	5	0.60	2
6	6	0.64	2
7	7	0.66	2
8	8	0.68	2
9	9	0.69	2

Ejercicio 2

	accuracy	auc	ecm	n
0	0.800000	0.760410	0.190403	