

Análisis deudores BCRA

Qué estimador podemos usar para predecir el scoring?

Cómo estimar comportamiento según deuda y según bancos?

Imposible trabajar con 24 meses...alternativa trabajar con 12?

Limpiar datos con $\text{sit} == 0$?

Tomar el máximo de sit? Tomar maximo de $\text{sit} * \text{monto}$?

Debemos tener cuidado en los análisis de % de deuda pq hay muchísimos datos con $\text{monto} == 0$ y alteran la estadística.

Chunksize no funciona con dtypes y eso lo vuelve inmanejable cuando lo lee.

Ejemplo cuit y comportamiento

```
[103]: # df_deudores[fil]
# df_deudores[fil]
df_query.loc[:,df_query.columns.str.startswith('sit')].style.background_gradient(cmap='RdYlGn_r', vmin=0, vmax=5)
```

[103]:

	sit0	sit1	sit2	sit3	sit4	sit5	sit6	sit7	sit8	sit9	sit10	sit11	sit12	sit13	sit14	sit15	sit16	sit17	sit18	sit19	sit20	sit21	sit22	sit23
Banco																								
GALICIA	5	5	4	4	4	4	4	3	3	3	1	1	1	1	1	0	0	0	0	0	0	0	0	0
TRANSATLANTICA	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
TARJETA N	5	4	4	4	4	4	3	3	2	2	1	1	1	1	0	0	0	0	0	0	0	0	0	0
B.PROVINCIA	5	5	4	4	4	4	3	3	3	2	2	1	1	1	1	1	0	0	0	0	0	0	0	0

[105]:

	monto0	monto1	monto2	monto3	monto4	monto5	monto6	monto7	monto8	monto9	monto10	monto11	monto12	monto13	monto14	monto15	monto16	monto17	monto18	monto19	monto20	mon
Banco																						
GALICIA	2298	2318	2331	2101	2088	2064	2038	1979	1959	1922	1868	1857	1848	1829	1393	0	0	0	0	0	0	0
TRANSATLANTICA	0	0	0	0	0	0	0	0	0	0	0	0	1400	2719	0	0	0	0	0	0	0	0
TARJETA N	39	39	39	39	39	39	39	39	39	39	39	39	36	35	0	0	0	0	0	0	0	0
B.PROVINCIA	716	740	719	682	654	629	603	576	566	566	575	575	574	557	547	274	0	0	0	0	0	0

Tiene sentido ver los montos de deuda?

Tiene sentido ver todos los bancos?

Ejemplo cuit y comportamiento

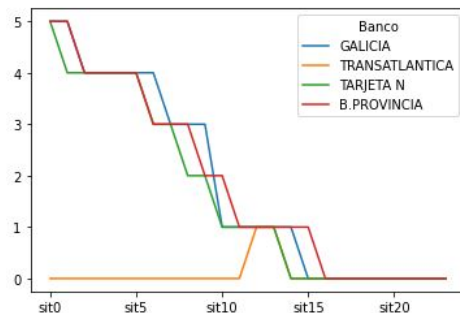
```
[103]: # df_deudores[fil]
# df_deudores[fil]
df_query.loc[:,df_query.columns.str.startswith('sit')].style.background_gradient(cmap='RdYlGn_r', vmin=0, vmax=5)
```

```
[103]:
```

	sit0	sit1	sit2	sit3	sit4	sit5	sit6	sit7	sit8	sit9	sit10	sit11	sit12	sit13	sit14	sit15	sit16	sit17	sit18	sit19	sit20	sit21	sit22	sit23
Banco																								
GALICIA	5	5	4	4	4	4	4	3	3	3	1	1	1	1	1	0	0	0	0	0	0	0	0	0
TRANSATLANTICA	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
TARJETA N	5	4	4	4	4	4	3	3	2	2	1	1	1	1	0	0	0	0	0	0	0	0	0	0
B.PROVINCIA	5	5	4	4	4	4	3	3	3	2	2	1	1	1	1	1	0	0	0	0	0	0	0	0

```
[104]: df_query.loc[:,df_query.columns.str.startswith('sit')].T.plot()
```

```
[104]: <AxesSubplot:>
```



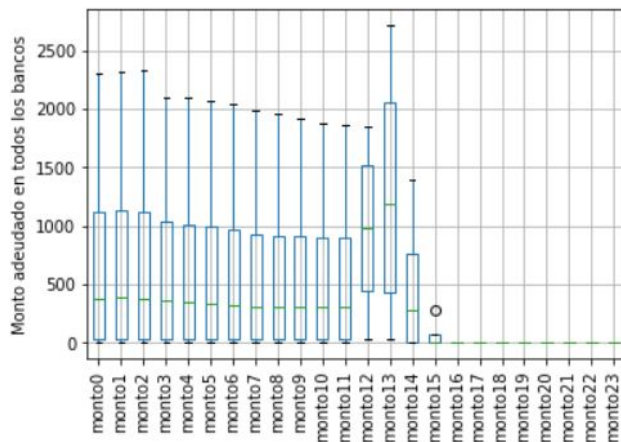
Nos interesa el comportamiento del último año? (más en Argentina)

Nos interesan todas sus sit..? O solamente la máxima?

```
[ ]: # df_query.loc[:,df_query.columns.str.startswith('sit')].T.plot()
```

Ejemplo cuit y comportamiento

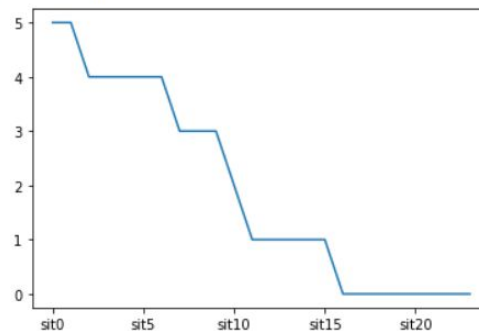
```
[106]: # df_query.columns.str.startswith('monto').boxplot()
df_query.loc[:,df_query.columns.str.startswith('monto')].boxplot()
plt.xticks(rotation=90) ;
plt.ylabel('Monto adeudado en todos los bancos');
```



Valor máximo, mínimo y media de la deuda.

```
[110]: df_deudores[fil].loc[:, 'sit0':'sit23':2].max().g
```

```
[110]: <AxesSubplot:>
```

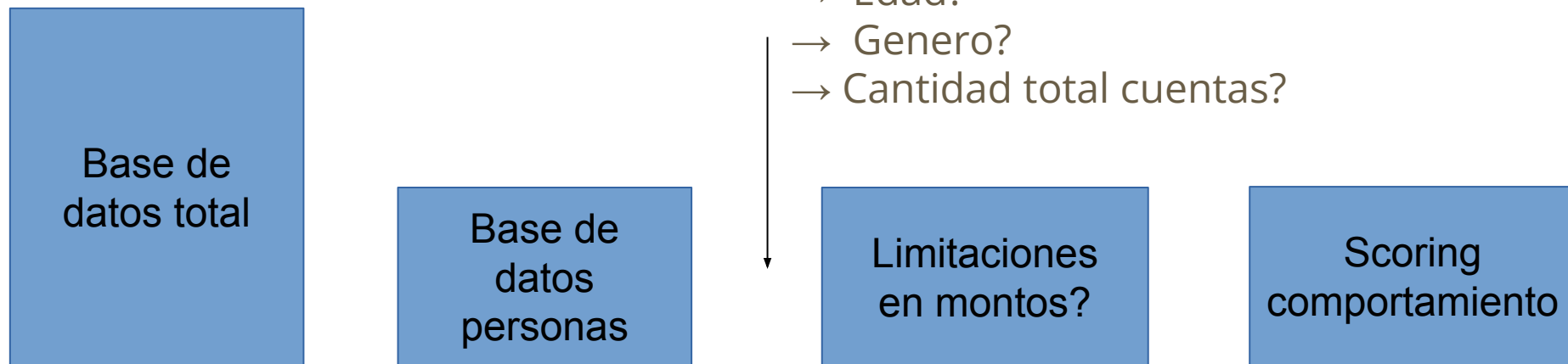


Valor máximo sit

Problemas para leer la base de datos entera

Tiempo de lectura con chunks:

41 segundos para leer 500000 filas --> 31 minutos el total



Ejemplo cuit muchas cuentas

Dr. Cristian Giuppone - Astrónomo (OAC). Data Scientist (B. Galicia)

```
df_query.loc[:,df_query.columns.str.startswith('sit')].style.background_gradient(cmap='RdYlGn_r', vmin=0, vmax=5)
```

	sit0	sit1	sit2	sit3	sit4	sit5	sit6	sit7	sit8	sit9	sit10	sit11	sit12	sit13	sit14	sit15	sit16	sit17	sit18	sit19	sit20	sit21	sit22	sit23
Banco																								
CFN S.A.	2	2	1	1	1	1	1	1	1	1	5	5	5	5	5	5	5	4	4	4	4	4	4	3
tro Litoral	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n Solidaria	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CENCOSUD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4
B.SERV TRAN SA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
BICA	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SISTEMA D	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1
GALICIA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4	4	3	3	3	3	3	3	2
B.ENTRERIOS	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ITAU	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VOII	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SOLUCIONE	2	2	2	2	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Palmares	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0
RESERVA D	0	0	0	0	0	0	0	0	0	0	0	0	0	5	5	5	5	5	5	5	5	5	5	0
CRÉDITOS	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Wenance S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	5	5	5	4	4	4	4	4	4
Credikot	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
UNION PRO	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TARJETA N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4	4	3	3	3	3	3	3	2

Nos interesa el
comportamiento del
último año? (más en
Argentina)

Nos interesan
todas sus sit..? O
solamente la máxima?

3 mil pesos!!!

Ejemplo cuit muchas cuentas

```
[51]: df_query.loc[:,df_query.columns.str.startswith('monto')].style.background_gradient(cmap='RdYlGn_r', vmin=10, vmax=150)
df_query.loc[:,df_query.columns.str.startswith('monto')].style.background_gradient(cmap='RdYlGn_r')
```

[51]:

	monto0	monto1	monto2	monto3	monto4	monto5	monto6	monto7	monto8	monto9	monto10	monto11	monto12	monto13	monto14	monto15	monto16	monto17	monto18	monto19
Banco																				
CFN S.A.	38	35	33	34	35	33	34	38	36	36	22	22	22	22	22	22	22	22	22	22
tro Litoral	223	226	228	230	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n Solidaria	0	0	0	0	19	50	53	55	127	175	60	133	165	97	101	104	107	128	135	142
CENCOSUD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B.SERV TRAN SA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	5	8
BICA	4	5	6	6	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SISTEMA D	119	125	14	21	25	14	25	11	14	15	7	0	0	0	0	0	5	7	17	22
GALICIA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4	4	4	4	4
B.ENTRERIOS	300	285	291	474	294	387	453	459	472	467	500	495	482	497	503	495	510	499	497	525
ITAU	40	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VOII	0	269	246	207	198	47	48	48	48	49	49	49	50	50	50	50	63	17	19	23
SOLUCIONE	7	14	21	24	27	23	20	0	0	0	0	0	0	0	0	0	0	0	0	0
Palmares	4	6	8	12	12	14	21	35	35	35	35	35	35	35	35	35	39	39	43	45
RESERVA D	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3	3	3	3	3	3
CRÉDITOS	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Wenance S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	68	68	68	68	68	68
Credikot	43	44	44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
UNION PRO	51	50	47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TARJETA N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	7	7	7	7	7

Ejemplo cuit muchas cuentas

Dr. Cristian Giuppone - Astrónomo (OAC). Data Scientist (B. Galicia)

```
df_query.loc[:,df_query.columns.str.startswith('sit')].style.background_gradient(cmap='RdYlGn_r', vmin=0, vmax=5)
```

	sit0	sit1	sit2	sit3	sit4	sit5	sit6	sit7	sit8	sit9	sit10	sit11	sit12	sit13	sit14	sit15	sit16	sit17	sit18	sit19	sit20	sit21	sit22	sit23
Banco																								
CFN S.A.	2	2	1	1	1	1	1	1	1	1	5	5	5	5	5	5	5	4	4	4	4	4	4	3
tro Litoral	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
n Solidaria	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CENCOSUD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4
B.SERV TRAN SA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
BICA	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SISTEMA D	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1
GALICIA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4	4	3	3	3	3	3	3	2
B.ENTRERIOS	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
ITAU	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VOII	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SOLUCIONE	2	2	2	2	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Palmares	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0
RESERVA D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	5	5	5	5	5	5	5	5	0
CRÉDITOS	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Wenance S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	5	5	5	4	4	4	4	4	4
Credikot	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
UNION PRO	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TARJETA N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4	4	3	3	3	3	3	3	2

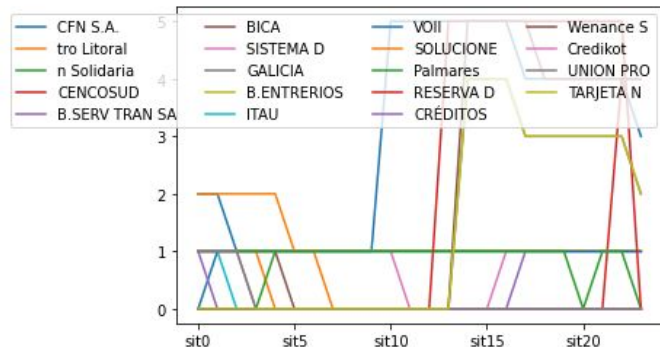
Nos interesa el comportamiento del último año? (más en Argentina)

Nos interesan todas sus sit..? O solamente la máxima?

3 Y 4 mil pesos!!!

Ejemplo cuit muchas cuentas

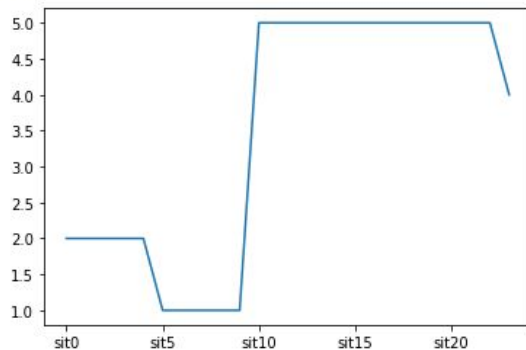
```
[44]: <Figure size 1152x432 with 0 Axes>
```



```
<Figure size 1152x432 with 0 Axes>
```

```
[47]: df_query.loc[:,df_query.columns.str.startswith('sit')].max().plot()
```

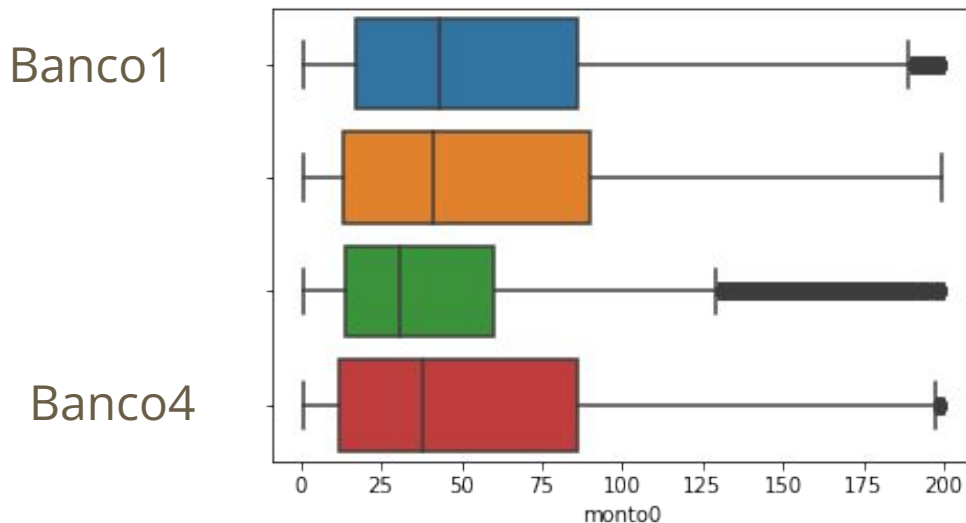
```
[47]: <AxesSubplot:>
```



Máxima situación

Análisis deudores BCRA

Qué estimador podemos usar para predecir el scoring?
Podemos seleccionar scoring en base a 2 o 3 entidades?



Análisis deudores BCRA. Definir!

Qué estimador podemos usar para predecir el scoring?

Cómo estimar comportamiento según deuda y según bancos?

Imposible trabajar con 24 meses...alternativa trabajar con 12?

Limpiar datos con $\text{sit} == 0$?

Tomar el máximo de sit? Tomar maximo de $\text{sit} * \text{monto}$?

Debemos tener cuidado en los análisis de % de deuda pq hay muchísimos datos con $\text{monto} == 0$ y alteran la estadística.

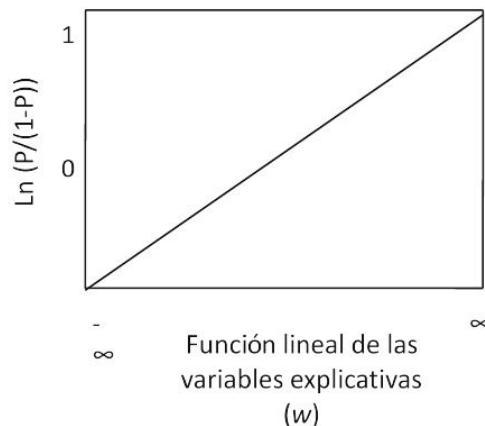
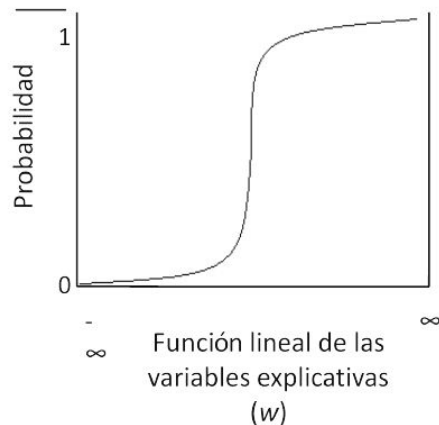
Chunksize no funciona con dtypes y eso lo vuelve inmanejable cuando lo lee.

Score Nosis → Default año siguiente

El modelo se puede linealizar de la siguiente forma:

$$P_t = \frac{1}{1 + e^{-w_t}}$$

$$\text{Ln}\left(\frac{P_t}{1-P_t}\right) = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots \quad (3)$$



La variable dependiente **Pt** es el logaritmo de la relación entre la posibilidad que el individuo incumpla (default=1) y la probabilidad que sea buen pagador (Default=0)

La variable dependiente, *Default*, toma los valores 0 y 1. *Pt* es la probabilidad de que tome el valor 1, por lo tanto la regresión que debe estimarse es:

$$\text{Ln}\left(\frac{\text{Default}_t}{1 - \text{Default}_t}\right) = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots \quad (4)$$

Score Nosis → Default año siguiente

Se analizó la población bancarizada según inscripción en Afip (monotributista, empleadores, autónomos, jubilados, no inscriptos)

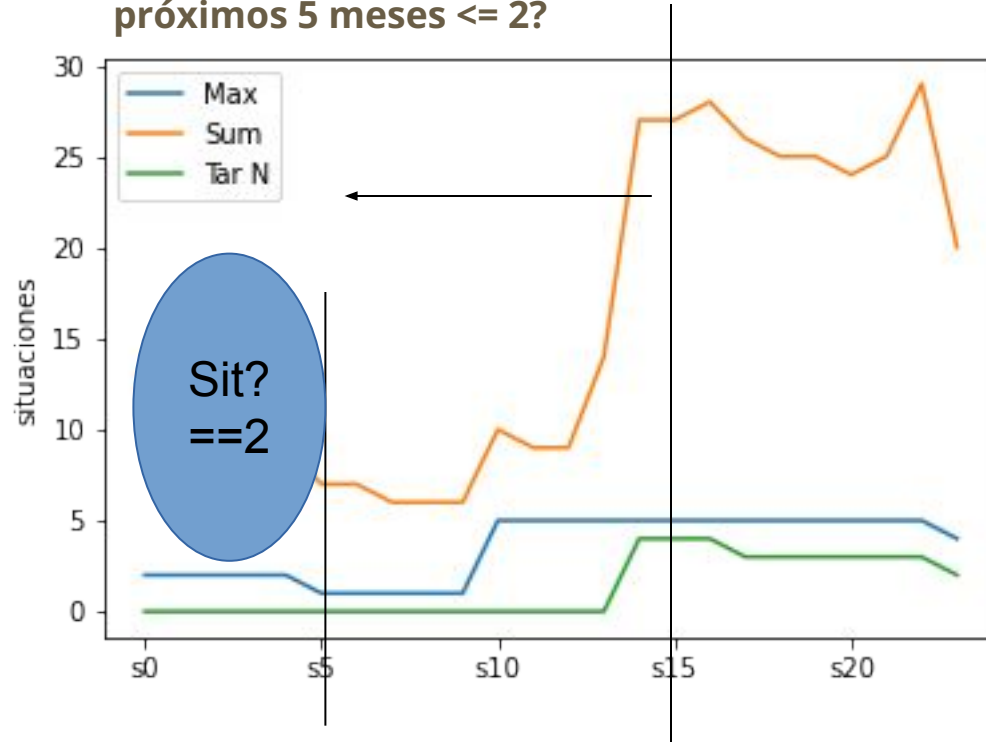
Modelo Versión 4

Porcentaje de deudores que incumplieron sus compromisos en el año siguiente

[illegible]

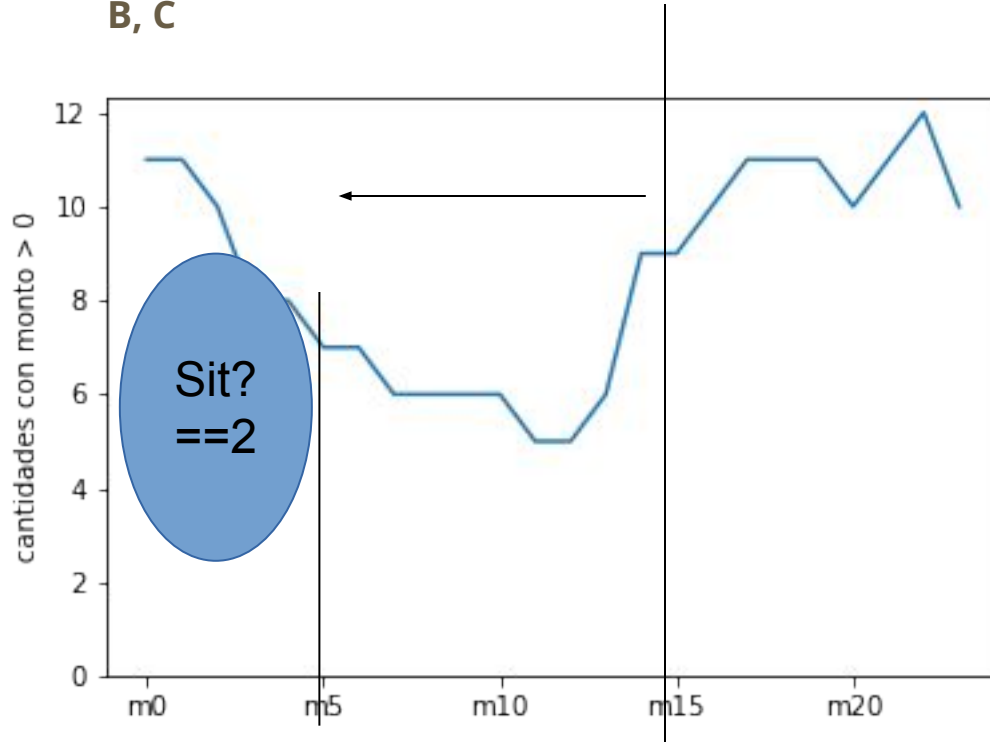
Score ventanas? → Default meses siguientes?

Tomar ventana 10 meses y predecir que en los próximos 5 meses ≤ 2 ?



Score ventanas? → Default meses siguientes?

Generar esto desagregando con tipos de entidades A, B, C



Score ventanas? → Default meses siguientes?

Medir los deltas
de monto por
entidades A, B, C

