



Incumplimiento de Pago de Tarjeta de Crédito

Search



Proyecto 2

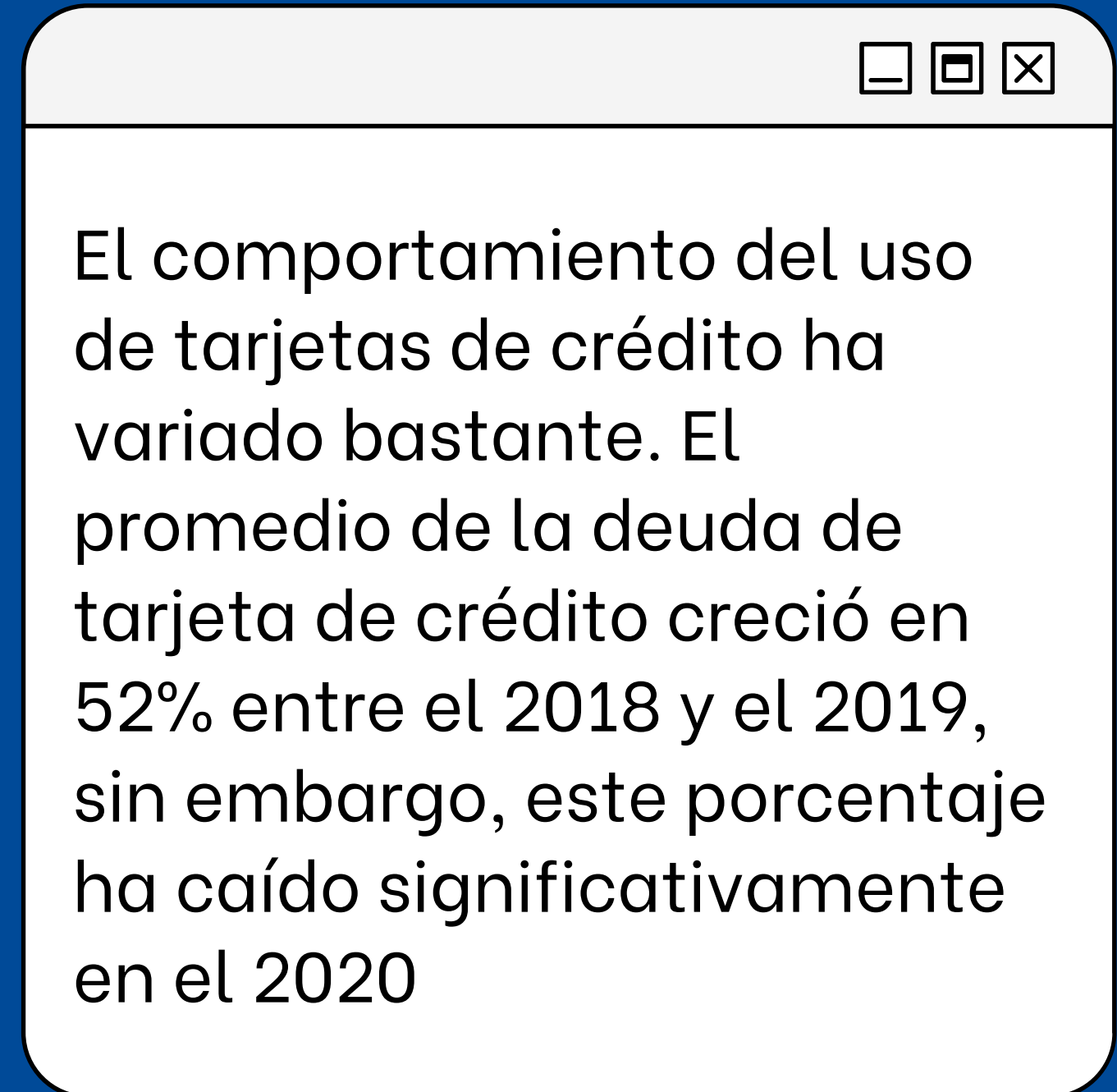
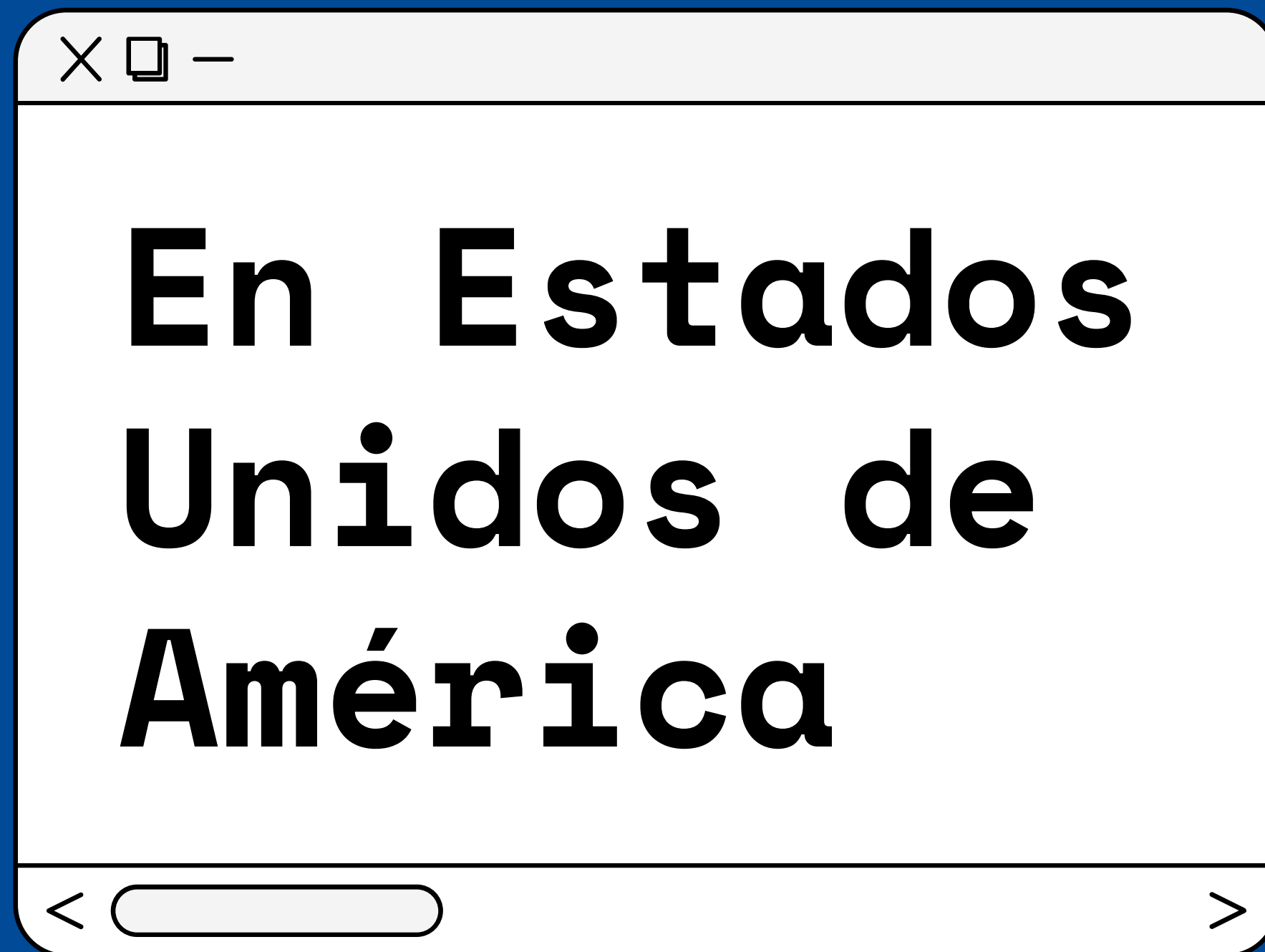
Diego Solorzano

Amado Garcia

Ricardo Valenzuela

Sara Zavala







Estados Unidos tiene una deuda de \$807 mil millones distribuidas en 506 millones de tarjetas de crédito mientras que la deuda promedio de una familia Estadounidense es de \$6,270

Estos números incrementan cada año hasta el 2020 donde se vio un decremento debido al COVID-19



Algunas Soluciones



A través de las compañías de liquidación de deudas.



El individuo con impago negocie directamente con los acreedores.



Cabe mencionar que maximizar su flujo de efectivo es crucial.

Objetivos

Search



1. Identificar a los tipos de clientes que pueden caer en impago
 - a. Identificar a los tipos de clientes que pueden caer en impago
 - b. Desarrollar modelos vistos en clase y predicciones sobre los clientes con deuda de tarjeta de crédito
 - c. Identificar a qué se debe el comportamiento que demuestran los porcentajes de deuda en los últimos años

Generales

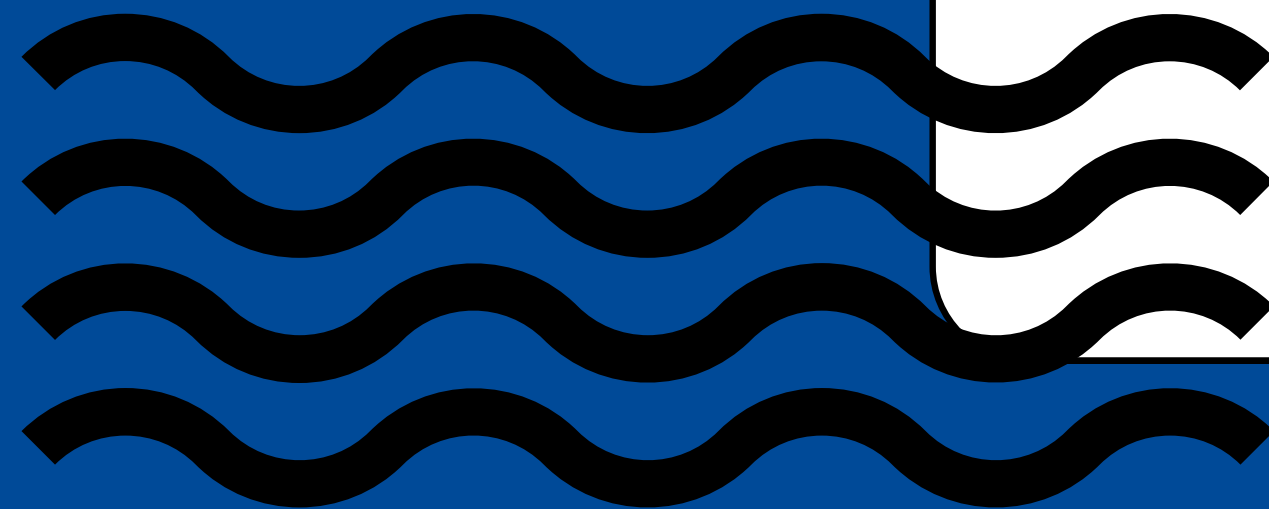
Específicos

Python



Analysis Exploratorio

Hay 25 variables
(22 cuantitativas,
3 categóricas)





Estandarizar nombres de columnas

```
df = clean(df, method = "standardize")
df.head()
```

	id	limit_bal	sex	education	marriage	age	pay_0	pay_2	pay_3	pay_4	...	bill_amt4	bill_amt5	bill_amt6	pay_amt1	pay_amt2	pay_amt3	pay_amt4	pay_amt5
0	1	20000.0	2	2	1	24	2	2	-1	-1	...	0.0	0.0	0.0	0.0	689.0	0.0	0.0	0.0
1	2	120000.0	2	2	2	26	-1	2	0	0	...	3272.0	3455.0	3261.0	0.0	1000.0	1000.0	1000.0	1000.0
2	3	90000.0	2	2	2	34	0	0	0	0	...	14331.0	14948.0	15549.0	1518.0	1500.0	1000.0	1000.0	1000.0
3	4	50000.0	2	2	1	37	0	0	0	0	...	28314.0	28959.0	29547.0	2000.0	2019.0	1200.0	1100.0	1100.0
4	5	50000.0	1	2	1	57	-1	0	-1	0	...	20940.0	19146.0	19131.0	2000.0	36681.0	10000.0	9000.0	9000.0

Cambiar a tipo categorica las variables adecuadas

```
: to_categoric = ["sex", "education", "marriage",
                  "pay_0", "pay_2", "pay_3", "pay_4", "pay_5", "pay_6", "default.payment.next.month"]
```

```
: df = clean(df, method = 'dtypes', columns = to_categoric,
            dtype='category')
```

```
: df = df.rename(columns = {'pay_0': 'pay_1'}, inplace = False)
```

Eliminando columna de Id

```
1): df = clean(df, method = 'dropcols', columns = ['id'])
```

Cambiando a solo un valor desconocido los valores 0,5,6 de educacion

```
2): df = clean(df, method = "replaceval",  
               columns = ["education"],  
               to_replace = [0,5,6],  
               value = 5)
```

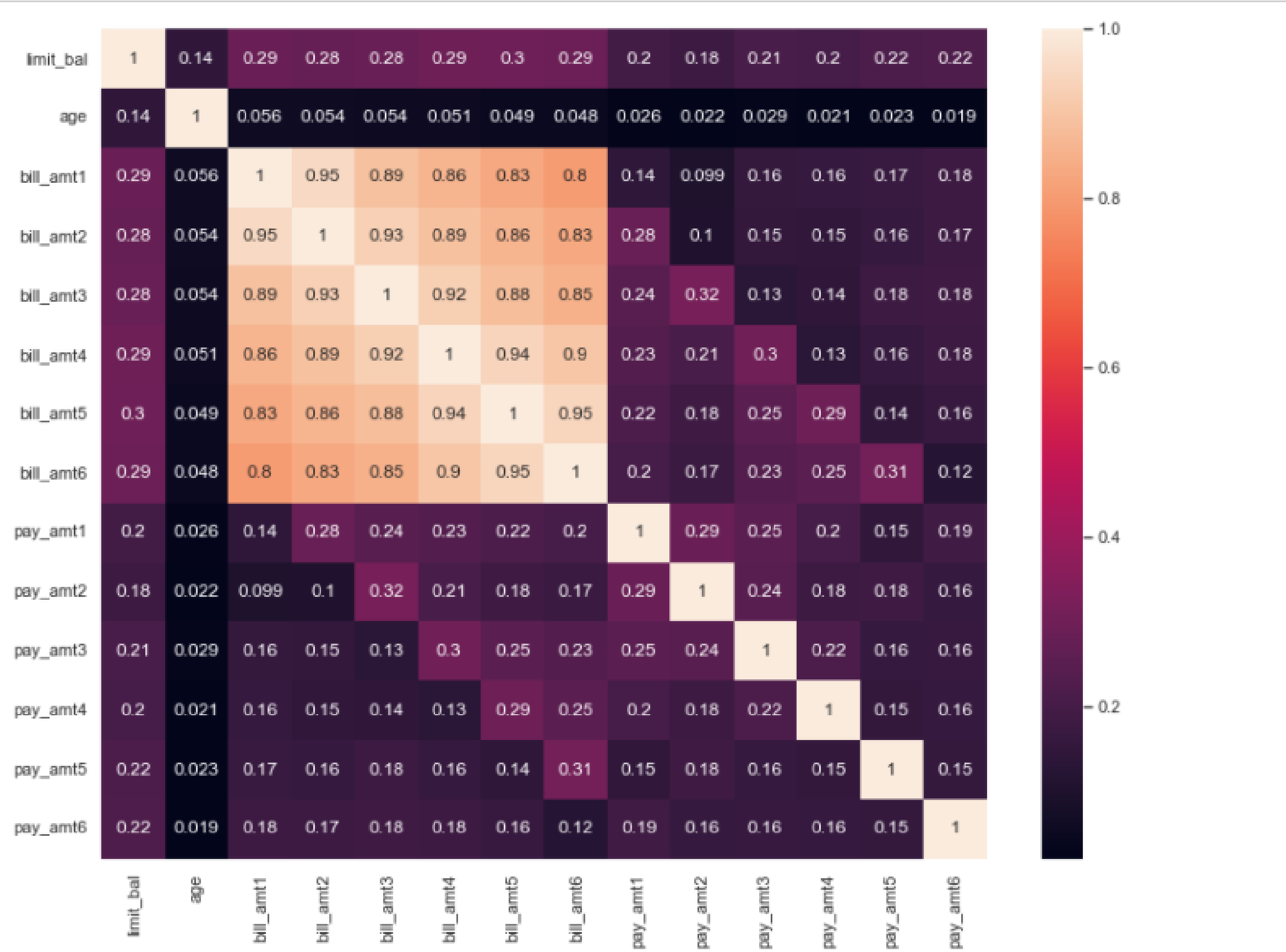
Cambiando el valor 0 a 3 (otros) en marrige

```
3): df = clean(df, method = "replaceval",  
               columns = ["marriage"],  
               to_replace = [0],  
               value = 3)
```

Aclaracion sobre los valores -2 y 0 de las columans pay

Estos valores si tienen un significado en el dataset con respecto a la realizacion de los pagos por lo que no se pueden modificar a nulo.

```
correlation = df.corr()
plt.figure(figsize = (13, 10))
sn.heatmap(correlation, annot=True)
plt.show()
```



Obtener columna de bill_amt con valores promediados

```
dfSimp = df
```

```
promedio = (dfSimp["bill_amt1"]+dfSimp["bill_amt2"]+dfSimp["bill_amt3"]  
            +dfSimp["bill_amt4"]+dfSimp["bill_amt5"]+dfSimp["bill_amt6"])/6  
dfSimp["prom_bill_amt"] = promedio
```

```
dfSimp = clean(dfSimp, method = 'dropcols', columns = ['bill_amt1',"bill_amt2","bill_amt3","bill_amt4","bill_amt5","bill_amt6"])
```

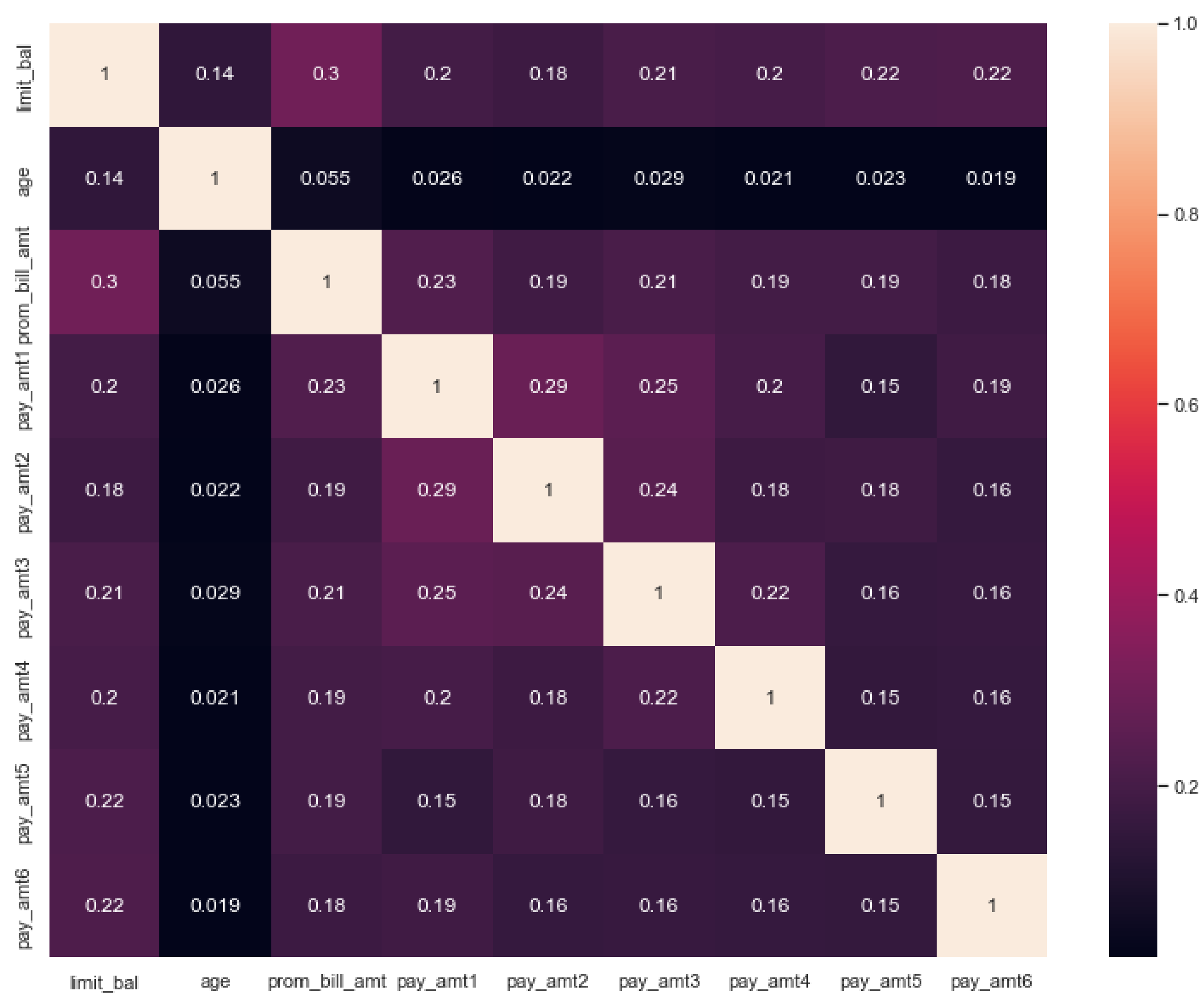
Cambiando de orden columnas

```
cols = ['limit_bal', 'sex','education','marriage','age','pay_1','pay_2','pay_3',  
        'pay_4','pay_5','pay_6','prom_bill_amt','pay_amt1','pay_amt2','pay_amt3','pay_amt4','pay_amt5',  
        'pay_amt6','default.payment.next.month']
```

```
dfSimp = dfSimp[cols]
```

Analizar nuevo dataset

```
correlation = dfSimp.corr()  
plt.figure(figsize = (13, 10))  
sn.heatmap(correlation, annot=True)  
plt.show()
```



Datos Atipicos

```
dfSimp = clean(dfSimp, method='outliers', columns=["pay_amt1","pay_amt3","pay_amt5"])
```

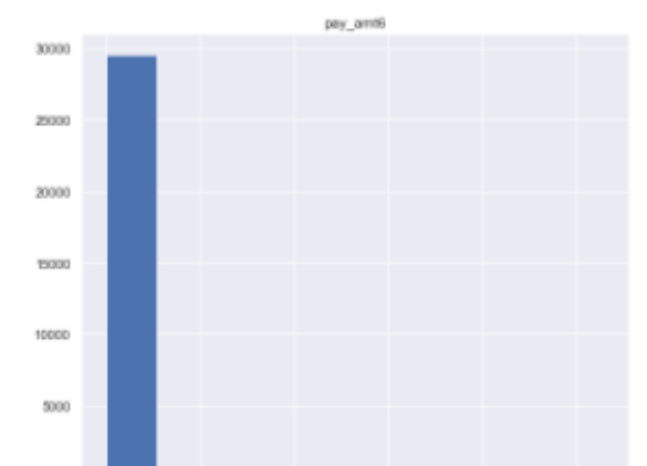
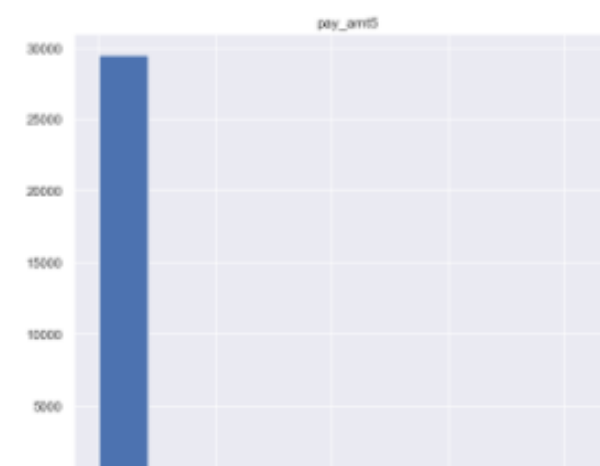
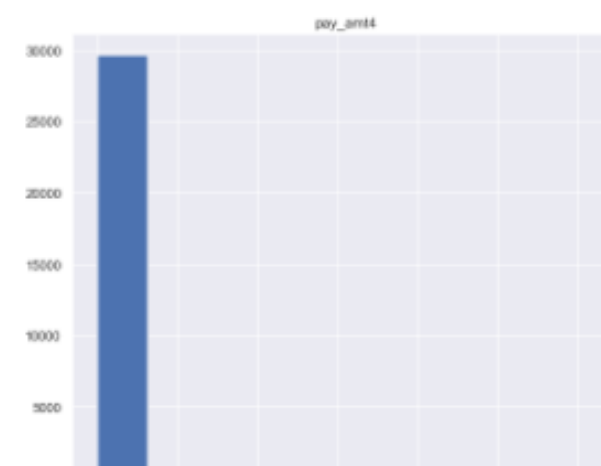
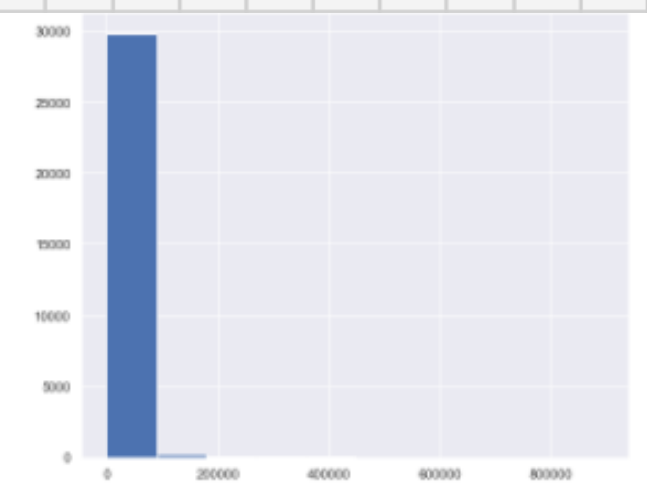
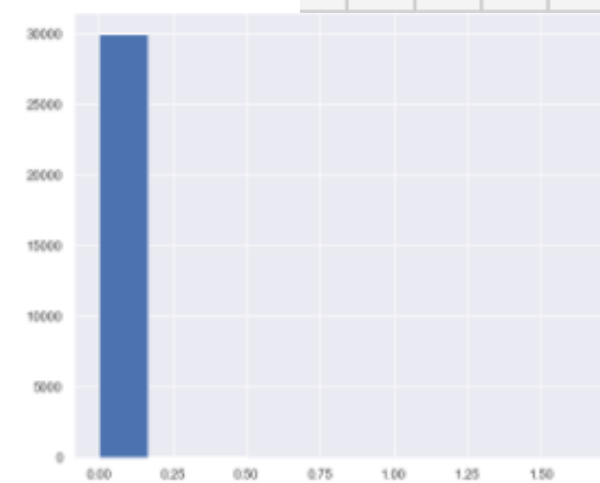
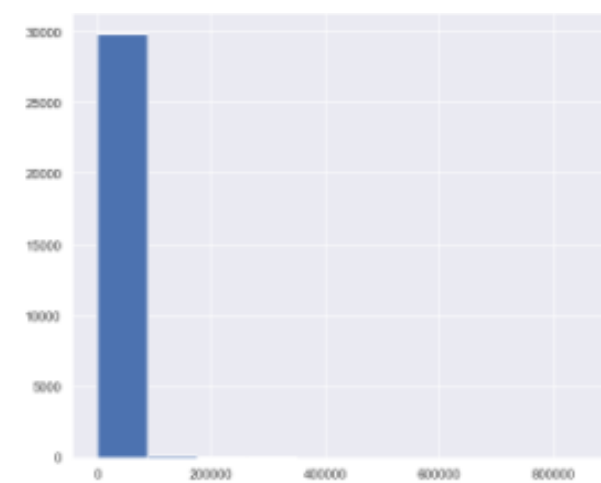
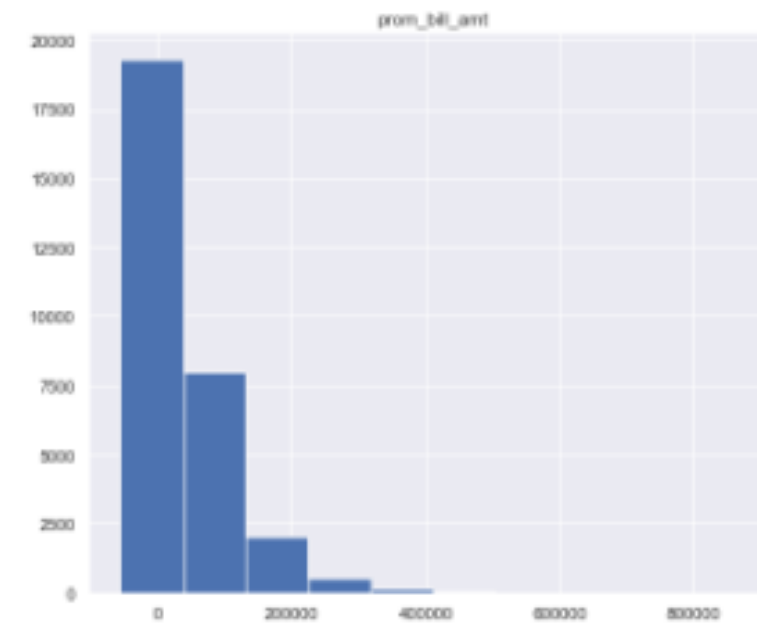
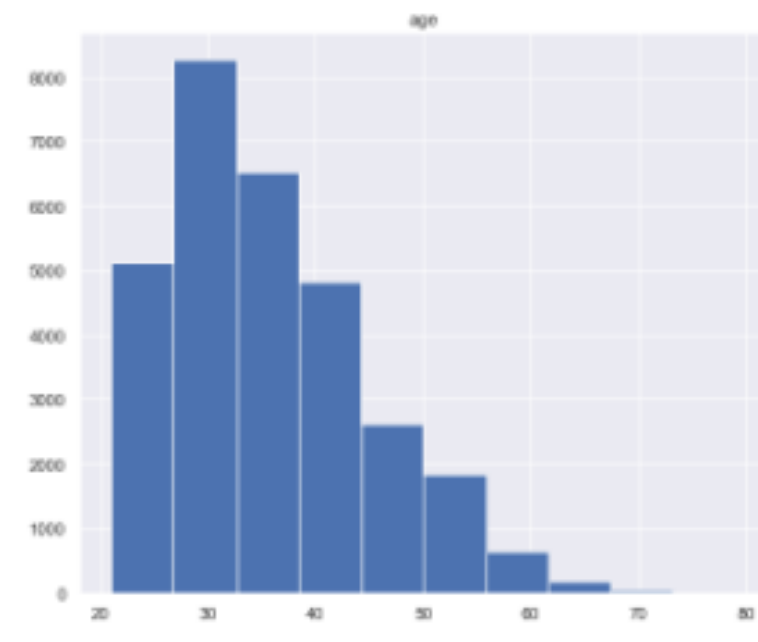
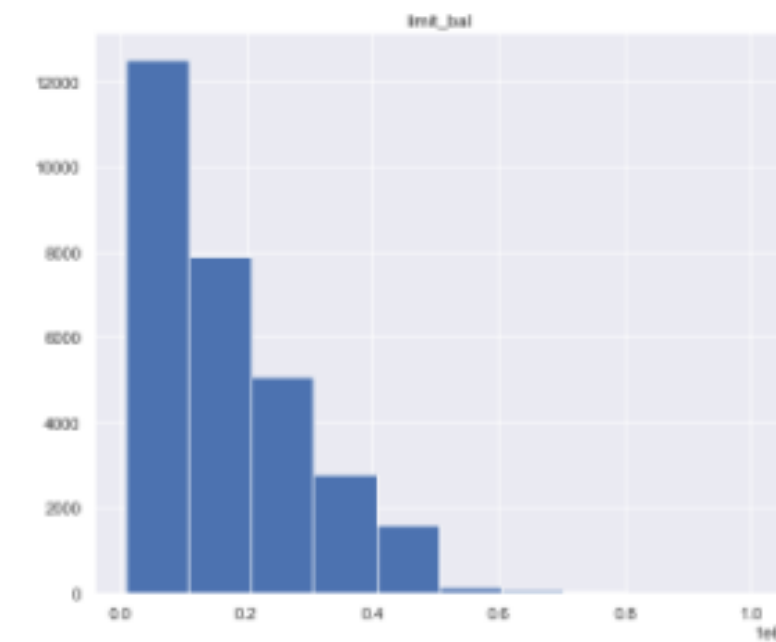
```
dfSimp.index = list(range(0,23241))
```

```
explore(dfSimp, method="summarize")
```

	dtypes	count	null_sum	null_pct	nunique	min	25%	50%	75%	max	mean
age	int64	23241	0	0.0	54	21.0	28.0	34.0	41.0	75.0	35.303903
default.payment.next.month	category	23241	0	0.0	2	-	-	-	-	-	-
education	category	23241	0	0.0	5	-	-	-	-	-	-
limit_bal	float64	23241	0	0.0	73	10000.0	50000.0	110000.0	200000.0	800000.0	143288.670883
marriage	category	23241	0	0.0	3	-	-	-	-	-	-
pay_1	category	23241	0	0.0	11	-	-	-	-	-	-
pay_2	category	23241	0	0.0	11	-	-	-	-	-	-
pay_3	category	23241	0	0.0	11	-	-	-	-	-	-
pay_4	category	23241	0	0.0	10	-	-	-	-	-	-
pay_5	category	23241	0	0.0	10	-	-	-	-	-	-
pay_6	category	23241	0	0.0	10	-	-	-	-	-	-
pay_amt1	float64	23241	0	0.0	4975	0.0	403.0	1950.0	3500.0	11012.0	2457.650876
pay_amt2	float64	23241	0	0.0	5412	0.0	400.0	1865.0	3585.0	385228.0	3596.274171
pay_amt3	float64	23241	0	0.0	4517	0.0	150.0	1376.0	3000.0	9072.0	1902.397659
pay_amt4	float64	23241	0	0.0	4831	0.0	90.0	1100.0	3000.0	256662.0	3057.287208
pay_amt5	float64	23241	0	0.0	4029	0.0	0.0	1001.0	2600.0	7464.0	1657.465384
pay_amt6	float64	23241	0	0.0	4781	0.0	0.0	1041.0	3000.0	528666.0	3269.936492
prom_bill_amt	float64	23241	0	0.0	20895	-56043.166667	2789.833333	18756.666667	48071.666667	456957.5	34871.408301
sex	category	23241	0	0.0	2	-	-	-	-	-	-

```
dfSimp.hist(figsize=(30, 25))
```

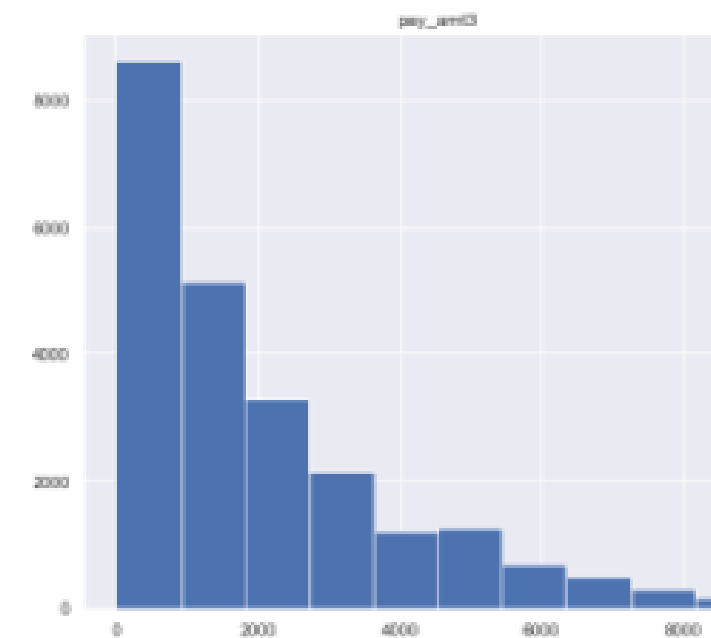
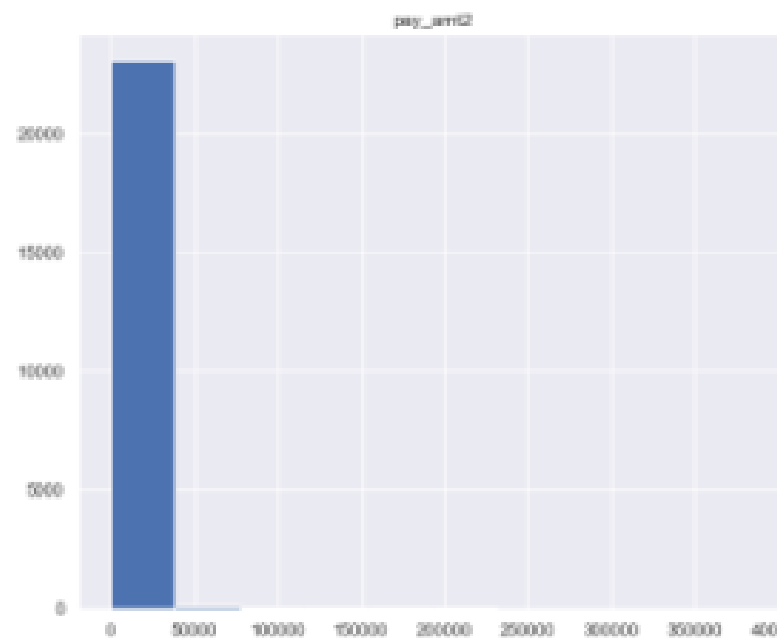
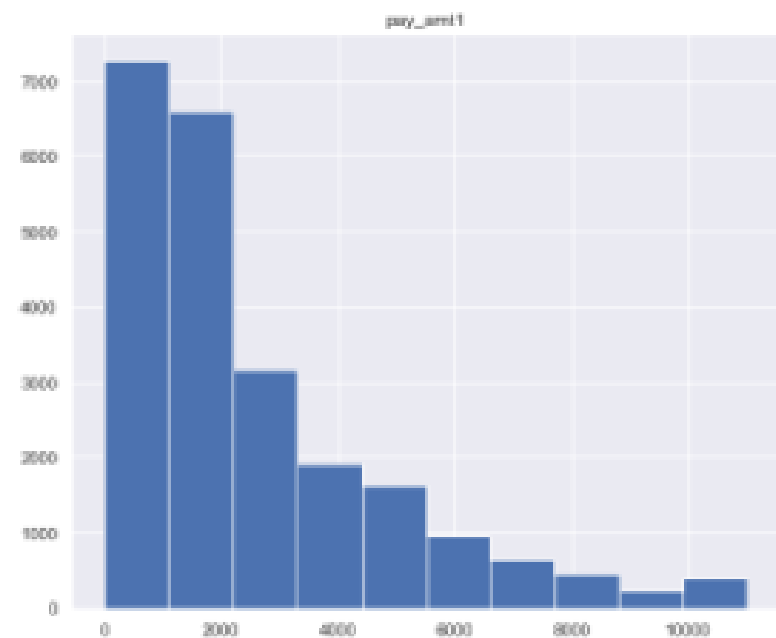
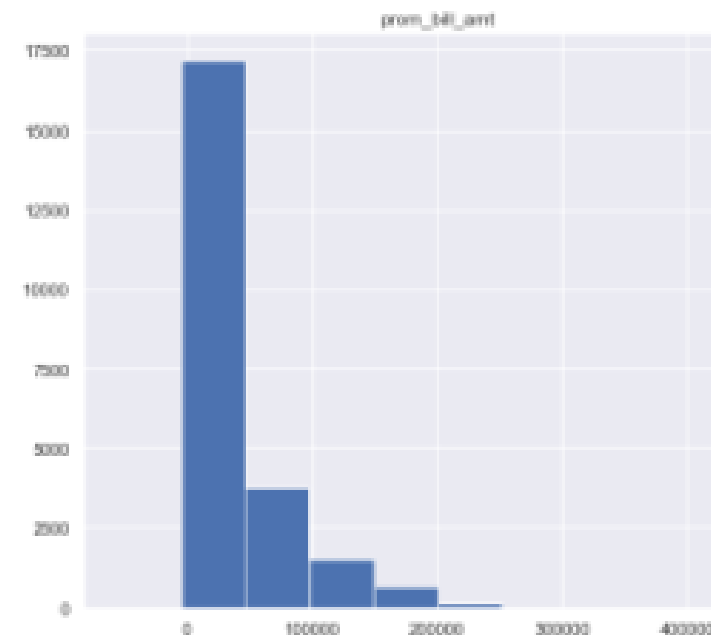
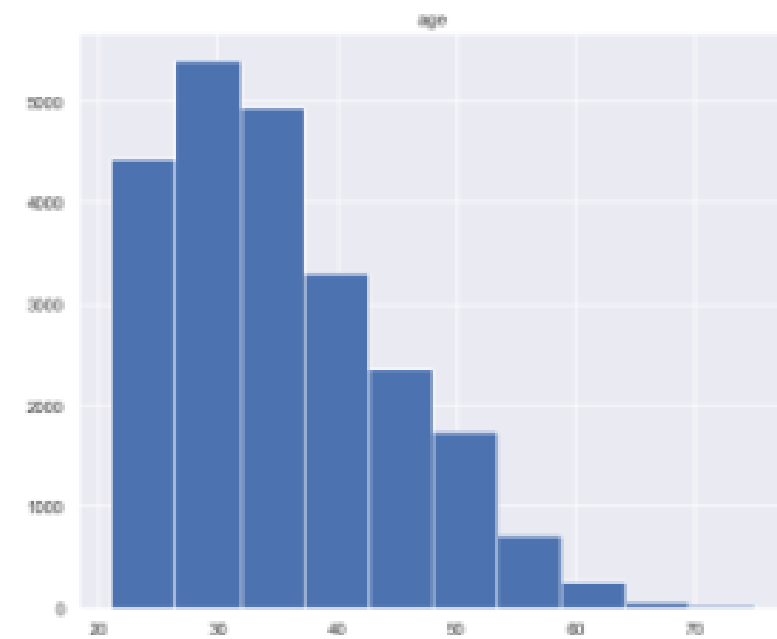
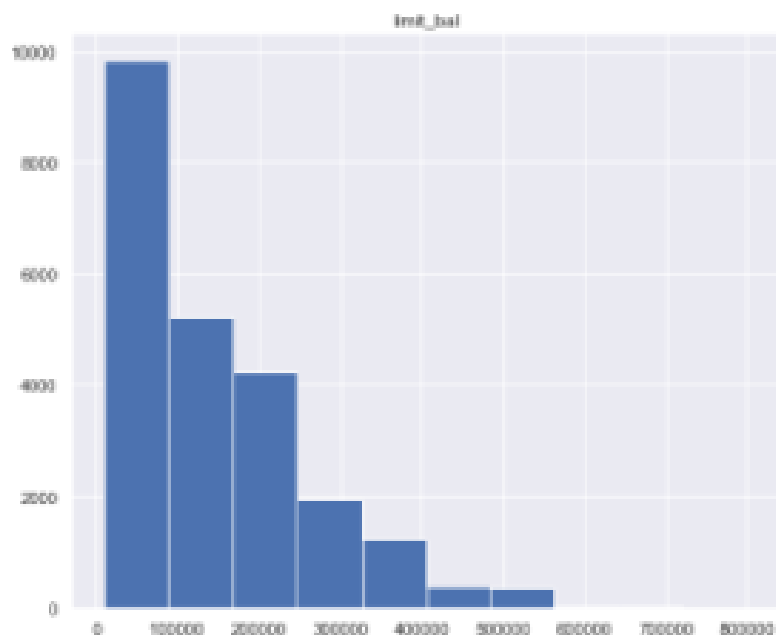
```
array([[<AxesSubplot:title={'center':'limit_bal'}>,  
       <AxesSubplot:title={'center':'age'}>,  
       <AxesSubplot:title={'center':'prom_bill_amt'}>],  
      [<AxesSubplot:title={'center':'pay_amt1'}>,  
       <AxesSubplot:title={'center':'pay_amt2'}>,  
       <AxesSubplot:title={'center':'pay_amt3'}>],  
      [<AxesSubplot:title={'center':'pay_amt4'}>,  
       <AxesSubplot:title={'center':'pay_amt5'}>,  
       <AxesSubplot:title={'center':'pay_amt6'}>]], dtype=object)
```



Representacion visual de datos cuantitativos

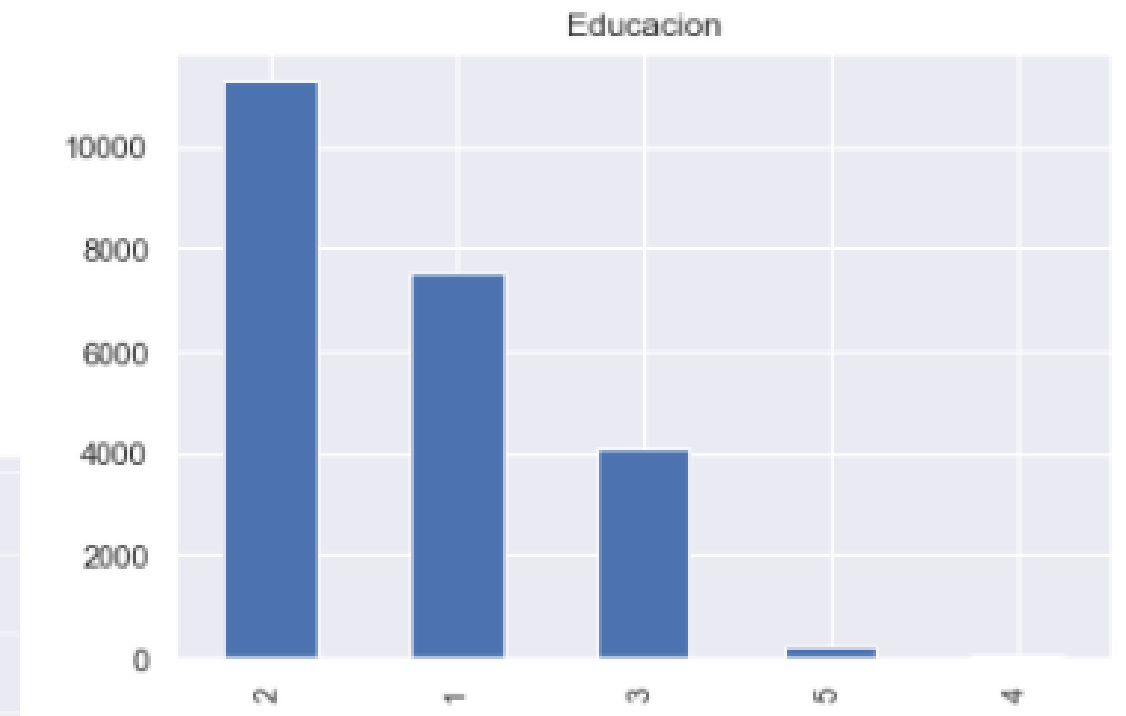
```
#Creacion de graficas para visualizar cada columna del dataset
dfSimp.hist(figsize=(30, 25))
```

```
array([[<AxesSubplot:title={'center':'limit_bal'}>,
       <AxesSubplot:title={'center':'age'}>,
       <AxesSubplot:title={'center':'prom_bill_amt'}>],
      [<AxesSubplot:title={'center':'pay_amt1'}>,
       <AxesSubplot:title={'center':'pay_amt2'}>,
       <AxesSubplot:title={'center':'pay_amt3'}>],
      [<AxesSubplot:title={'center':'pay_amt4'}>,
       <AxesSubplot:title={'center':'pay_amt5'}>,
       <AxesSubplot:title={'center':'pay_amt6'}>]], dtype=object)
```

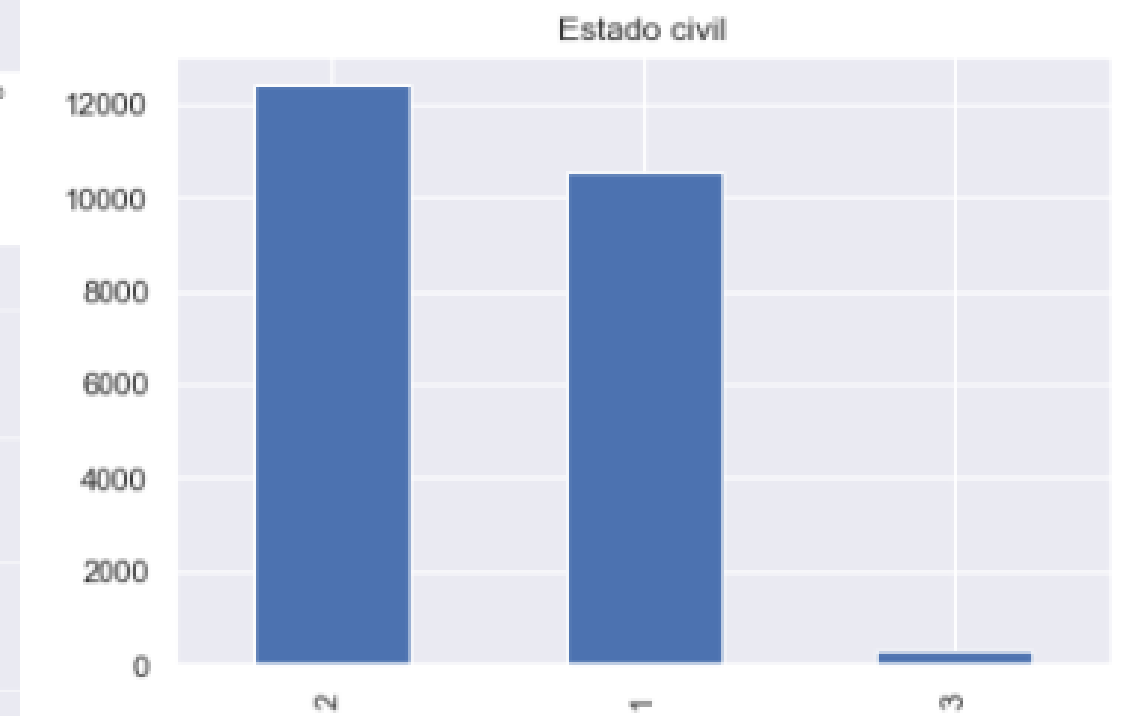


Visualizacion datos categoricos

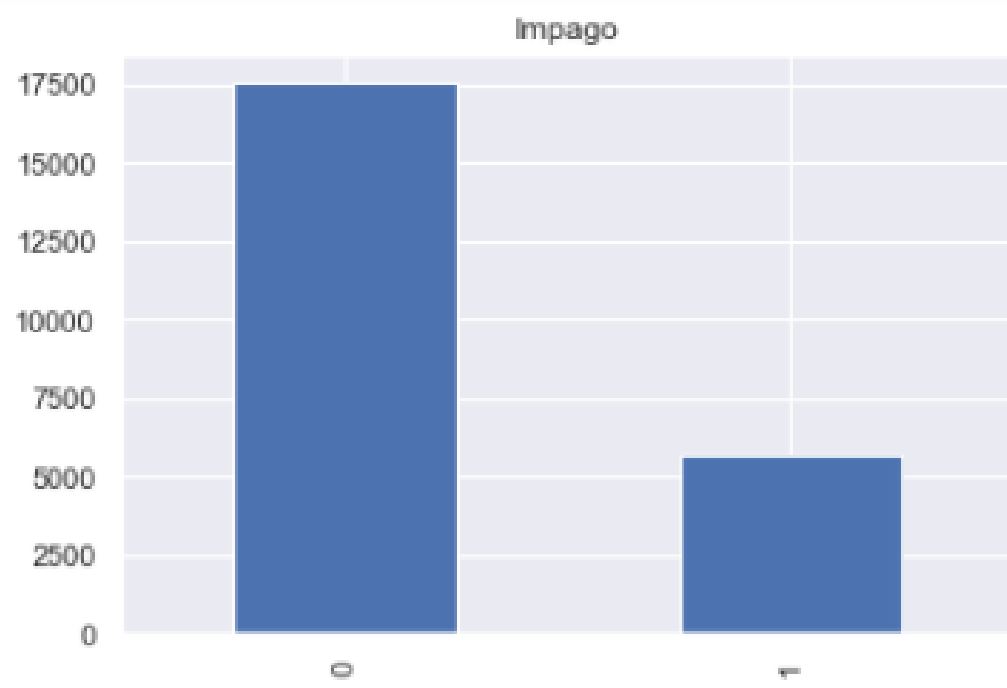
```
dfSimp['education'].value_counts().plot(kind='bar', title =
<AxesSubplot:title={'center':'Educacion'}>
```



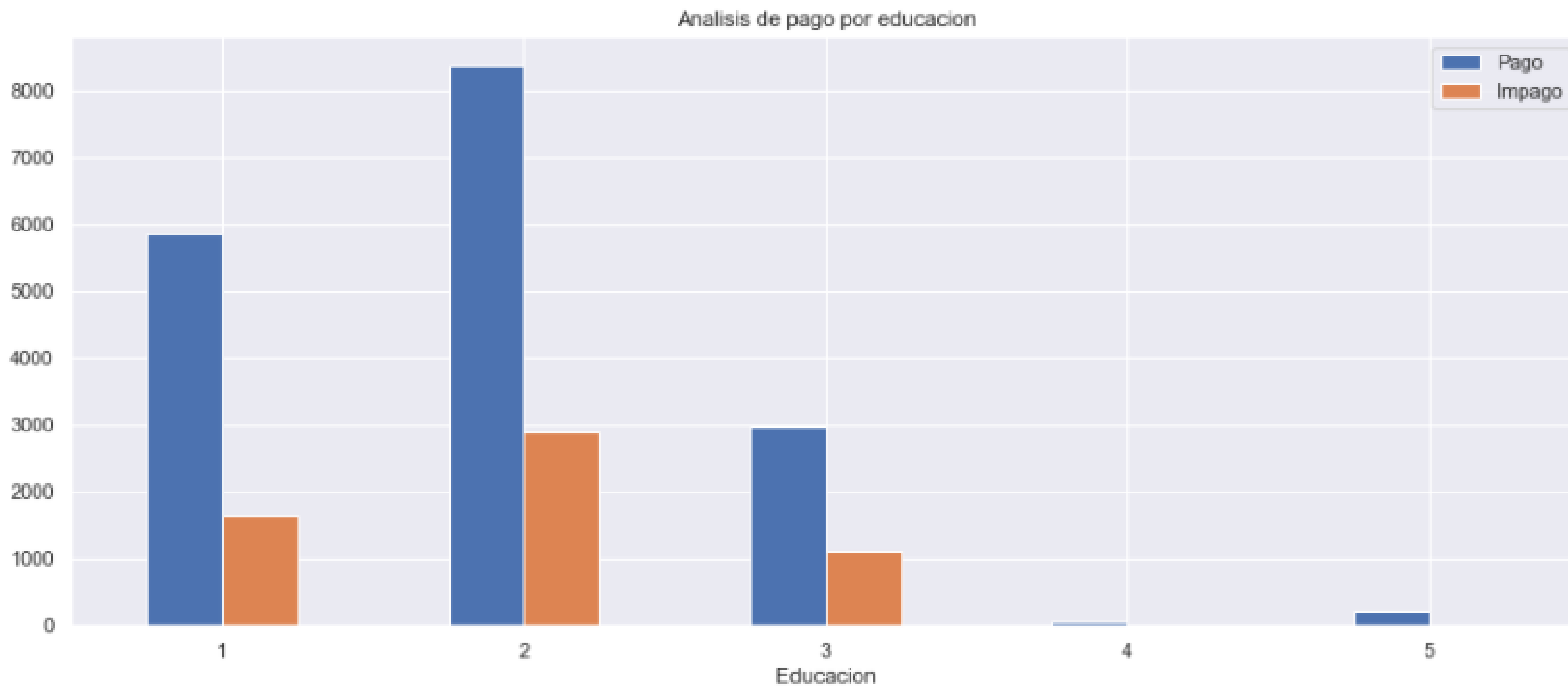
```
dfSimp['marriage'].value_counts().plot(kind='bar', title =
<AxesSubplot:title={'center':'Estado civil'}>
```



```
dfSimp['sex'].value_counts().plot(kind='bar', title = "Sexo")
```

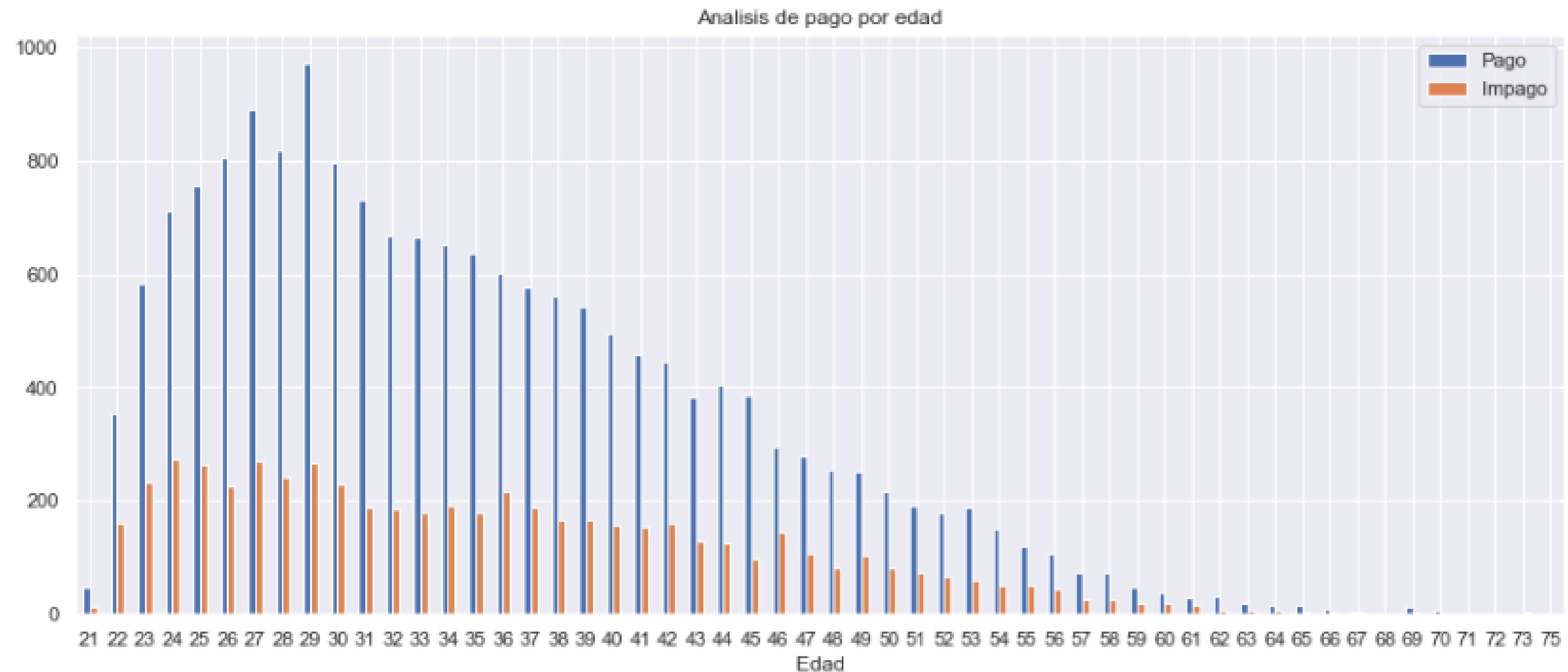
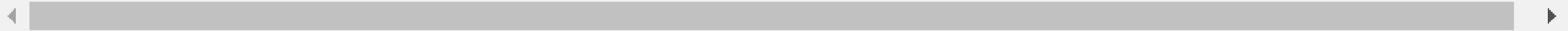


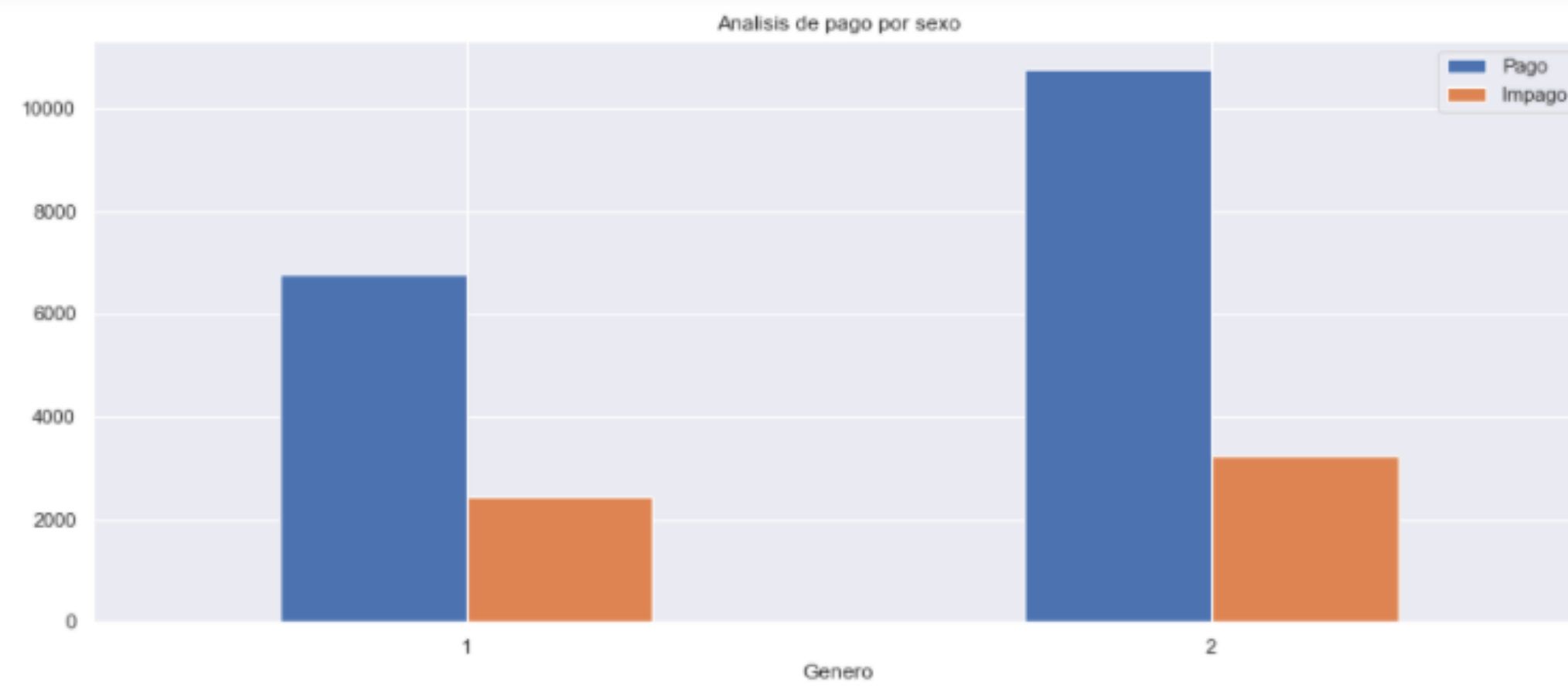
```
In [30]: pd.crosstab(dfSimp.education, dfSimp['default.payment.next.month']).plot(kind="bar",figsize=(15,6), title = "Análisis de pago por  
plt.xlabel('Educacion')  
plt.xticks(rotation=0)  
plt.legend(["Pago", "Impago"])  
plt.show()
```



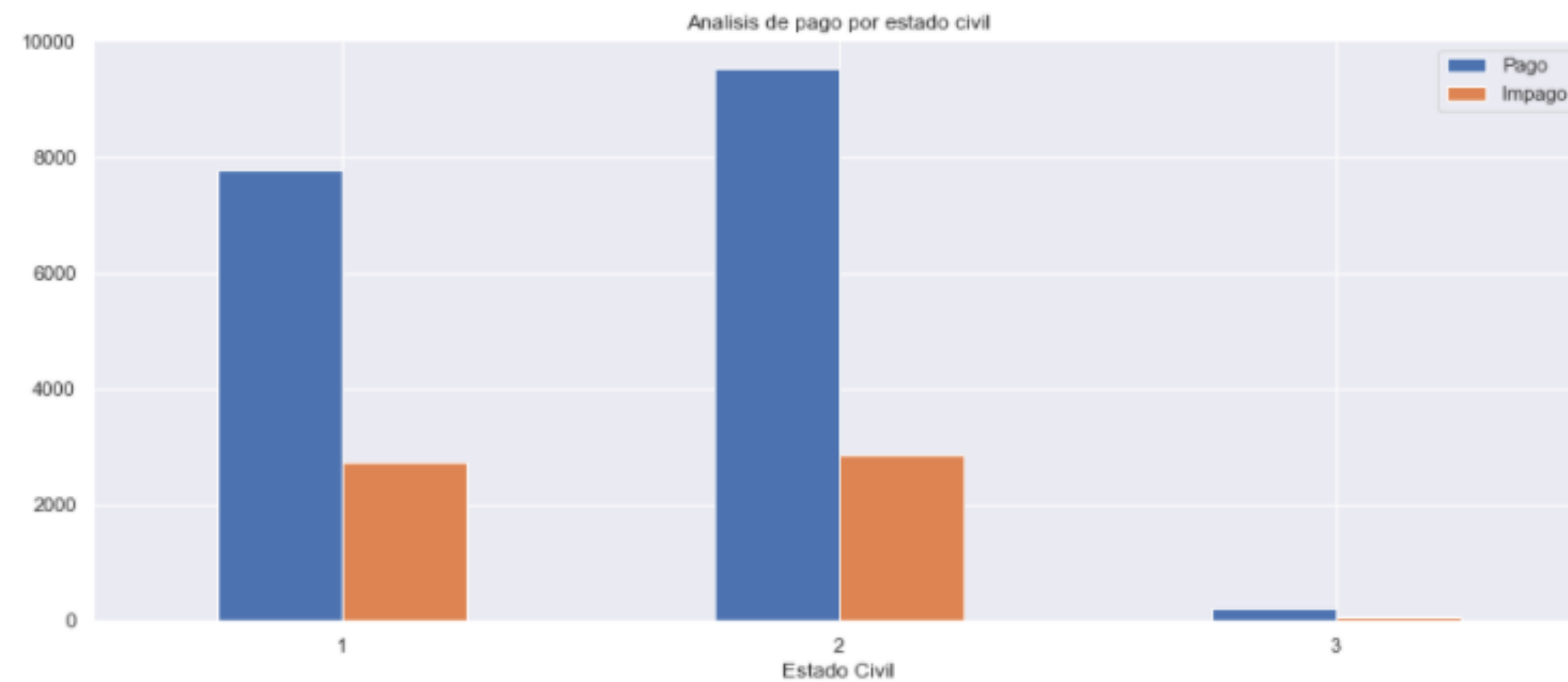
Cruce de variables

```
pd.crosstab(dfSimp.age, dfSimp['default.payment.next.month']).plot(kind="bar",figsize=(15,6), title = "Análisis de pago por edad"  
plt.xlabel('Edad')  
plt.xticks(rotation=0)  
plt.legend(["Pago", "Impago"])  
plt.show()
```



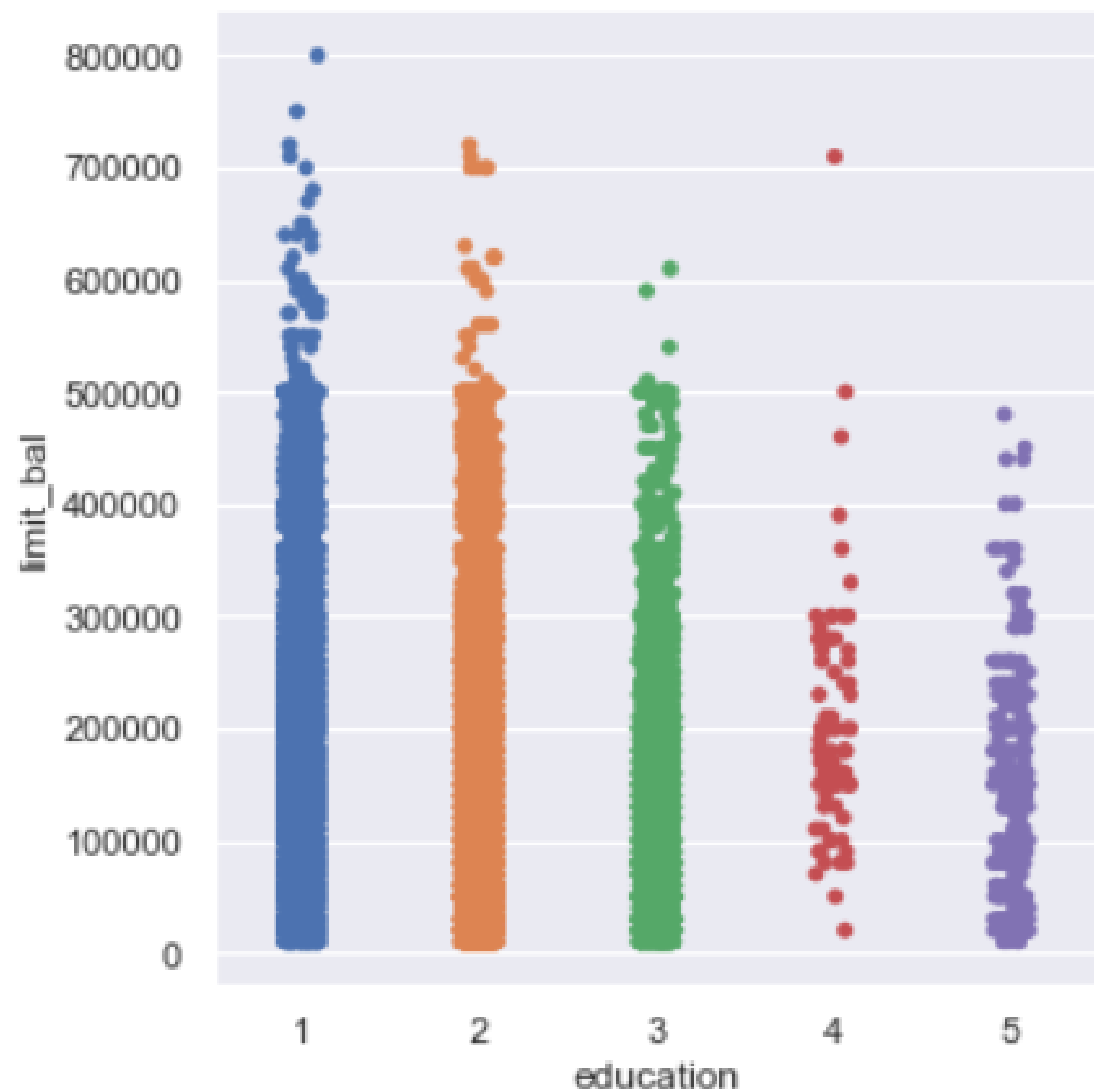


```
pd.crosstab(dfSimp.marriage, dfSimp['default.payment.next.month']).plot(kind="bar",figsize=(15,6),title = "Analisis de pago por e  
plt.xlabel('Estado Civil')  
plt.xticks(rotation=0)  
plt.legend(["Pago", "Impago"])  
plt.show()
```



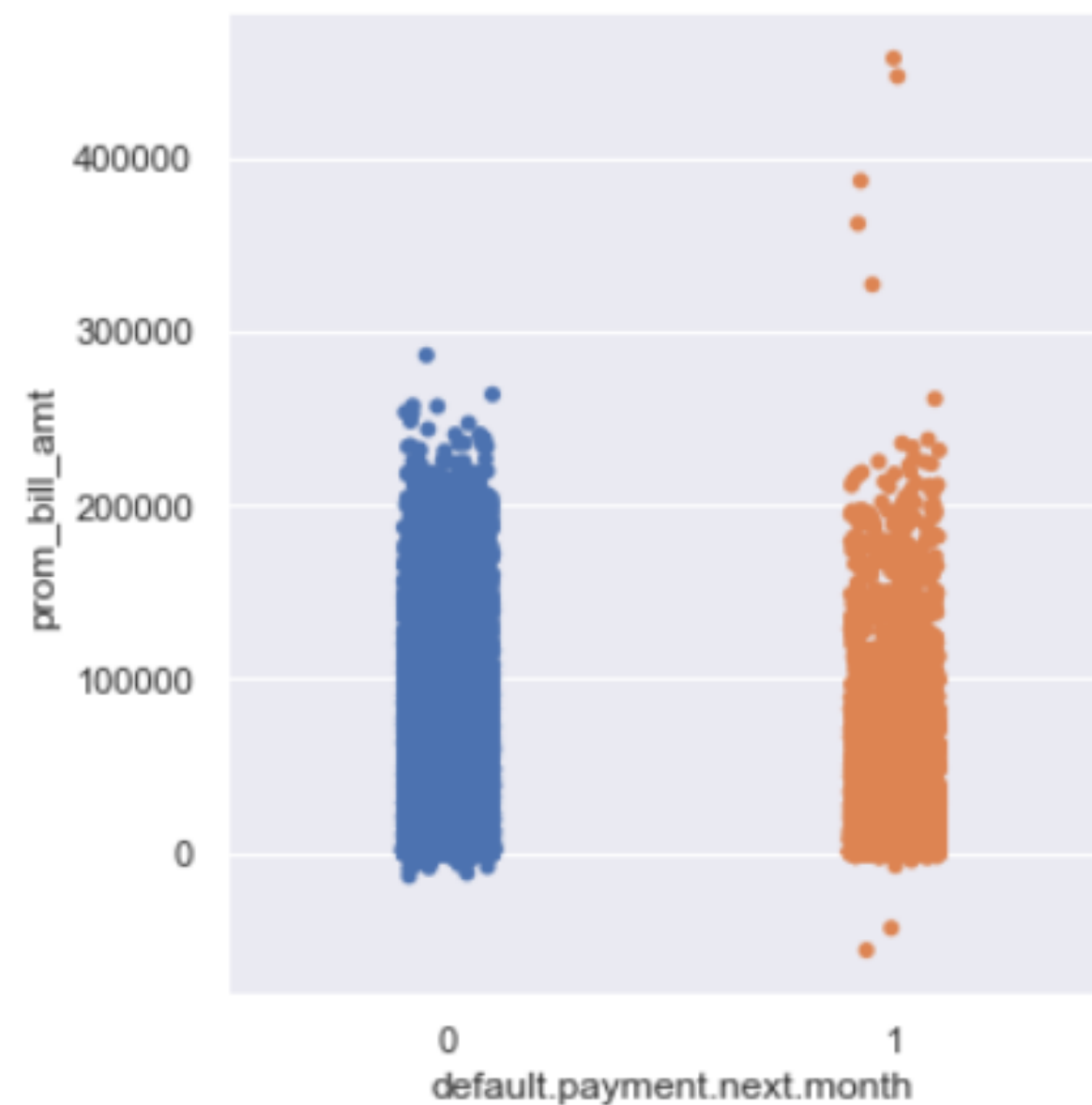

```
sns.catplot(x="education", y="limit_bal", data=dfSimp)
```

```
<seaborn.axisgrid.FacetGrid at 0x2a707394d90>
```



```
sns.catplot(x="default.payment.next.month", y="prom_bill_amt", data=dfSimp)
```

```
<seaborn.axisgrid.FacetGrid at 0x2a7097810a0>
```



Round 4

Descripción y Conclusión Final



El conjunto de datos tras las operaciones de limpieza cuenta con 23241 observaciones y 19 variables. 10 son categóricas y 9 son cuantitativas:



Características importantes

- El nivel de educación que predomina en los datos es universidad seguido de postgrado.
- El estado civil predominante es soltero
- El sexo mas presente es el femenino



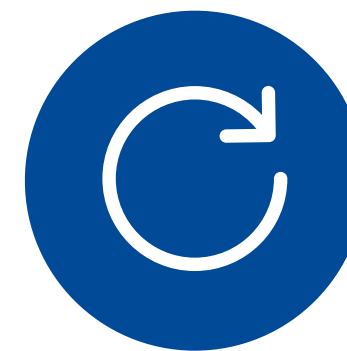


Variables cuantitativas

- La media de edad de las personas es de 35 años.
- La media de cantidad de crédito es alrededor de 143000.
- Los datos de pay_amt están muy sesgados, cuentan con altos valores atípicos



Al cruzar variables podemos observar lo siguiente:



1

Los que más probablemente paguen su deuda son personas alrededor de 23 a 36 años.

2

Hay un buen porcentaje de impago en todas las edades.

3

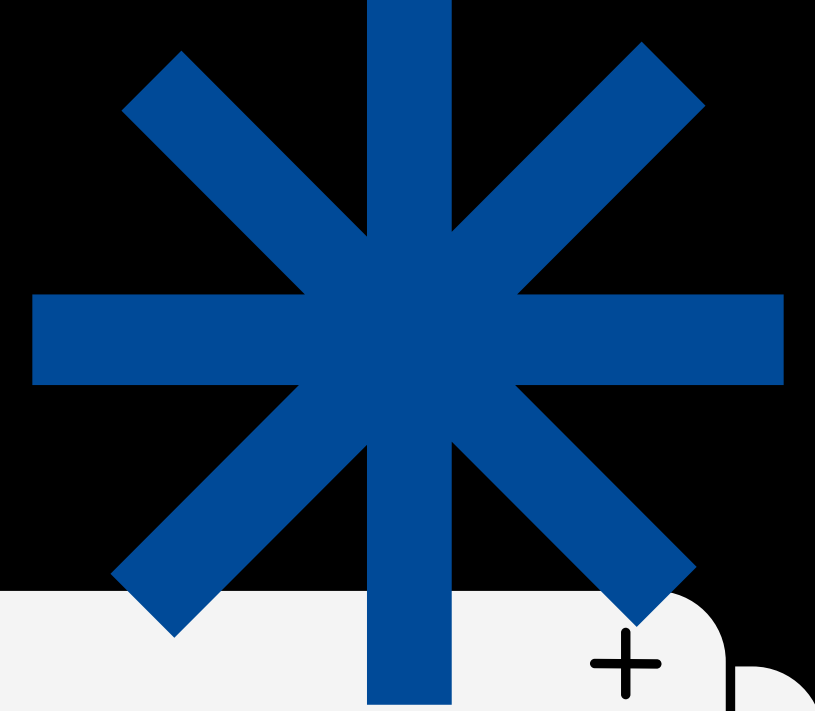
El estado civil casado presenta una cantidad de impagos muy similar a la de soltero a pesar de tener menos observaciones

4

El sexo femenino presenta tener una cantidad de impagos muy similar al masculino a pesar de tener mas observaciones.

5

Los niveles de educación estudios de posgrado, universidad, bachillerato suelen pedir un valor más alto de crédito que otros y desconocidos.



Gracias !

:D

