

UNIVERSIDAD SERGIO ARBOLEDA



**UNIVERSIDAD
SERGIO ARBOLEDA**

MÉTODOS DE ANÁLISIS DE DATOS

TALLER 3

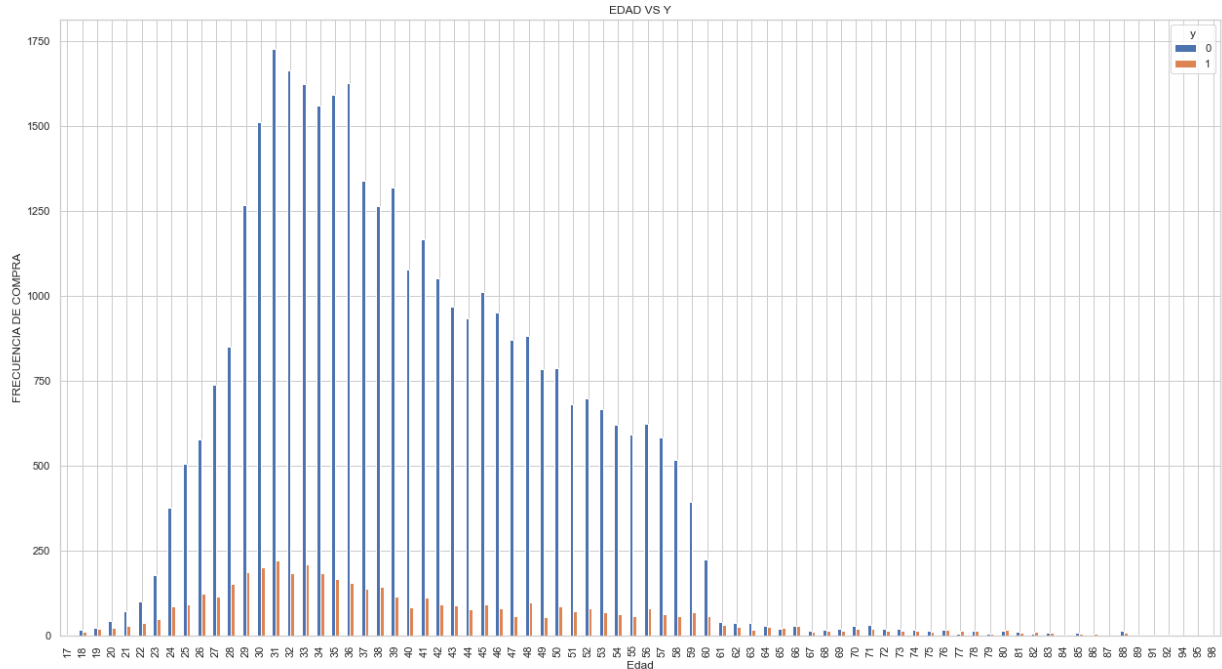
MILLER ALEXANDER QUIROGA CAMPOS

Docente
Luz Stella Gómez Fajardo Ph.D.

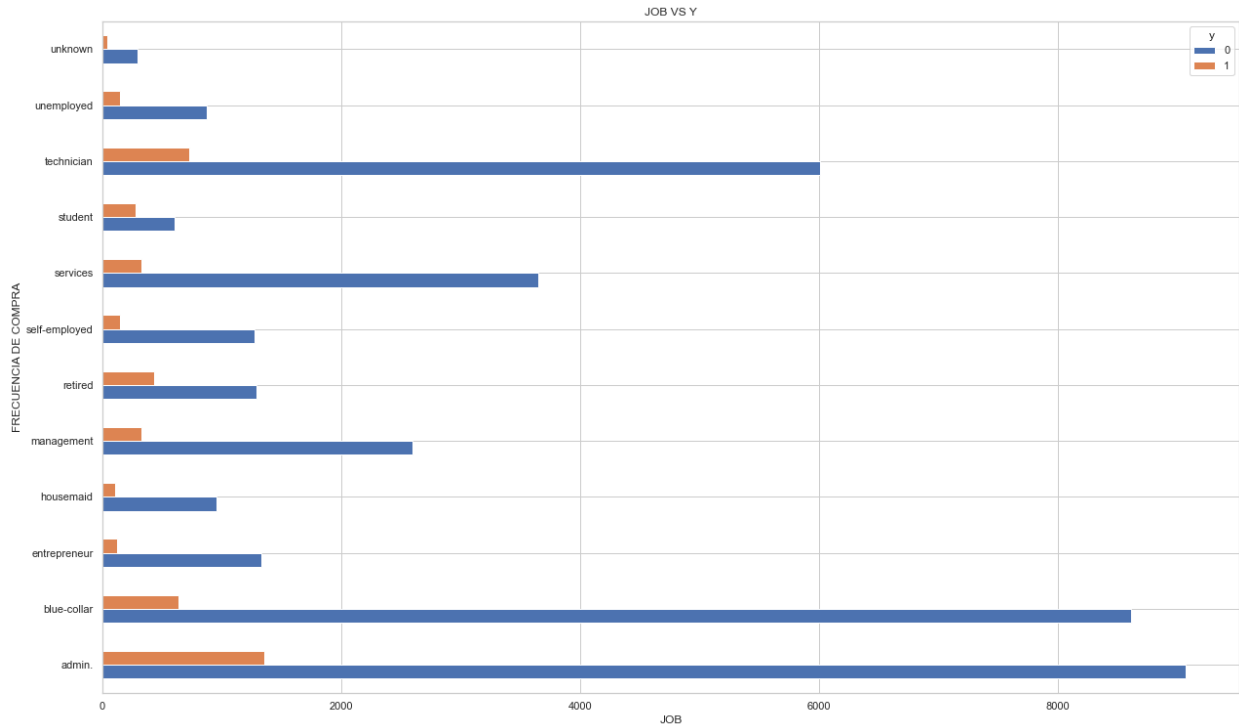
November 28, 2020

Ejercicio

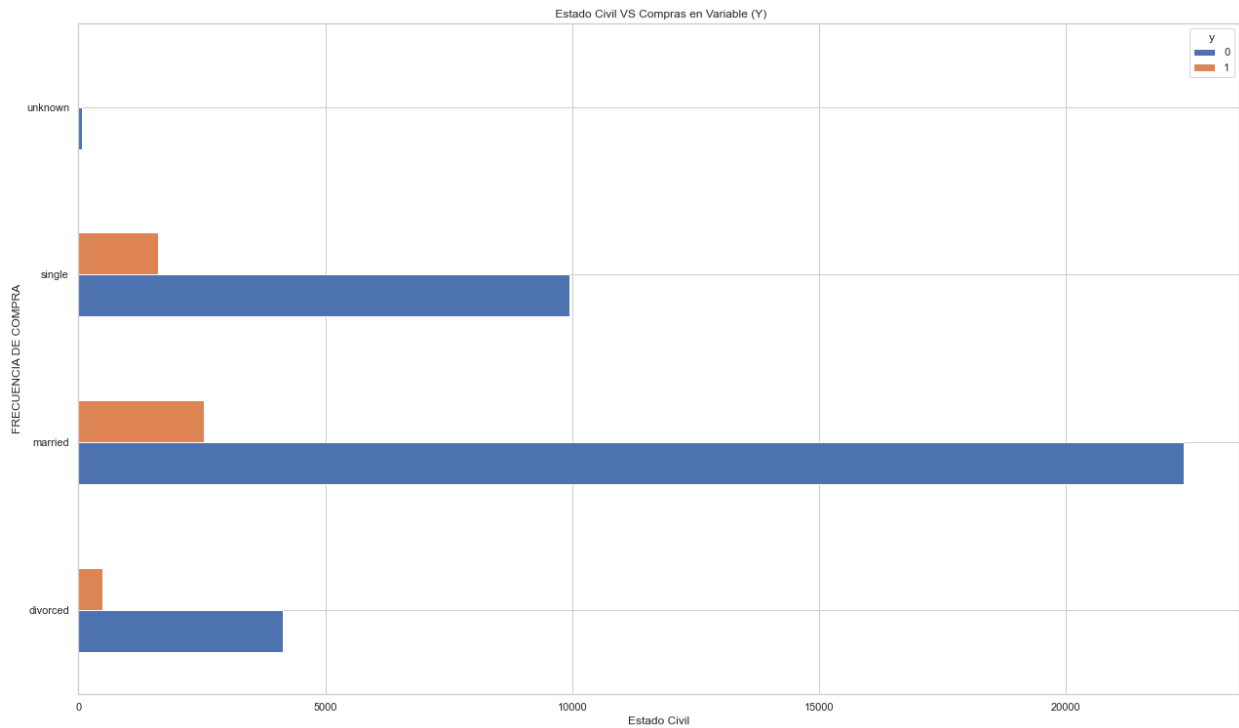
Graficar y analizar del comportamiento y/o distribución de las variables de la data Banking.csv con respecto a la variable respuesta "y", la cual me indica la compra. Con 0 me indica que no hay compra y con 1 hay compra de productos bancarios.



Para esta primera gráfica se analiza que la no compra de productos de un banco se realiza más cuando hay un rango de edad entre 31 y 36 años de edad. En la edad de 29, 37, 38 y 39 hay una frecuencia casi similar en cuanto a la compra de productos en el sector bancario. Para los productos bancarios que se compran, las edades que más frecuencia tiene son las edades entre 28 y 28 años.

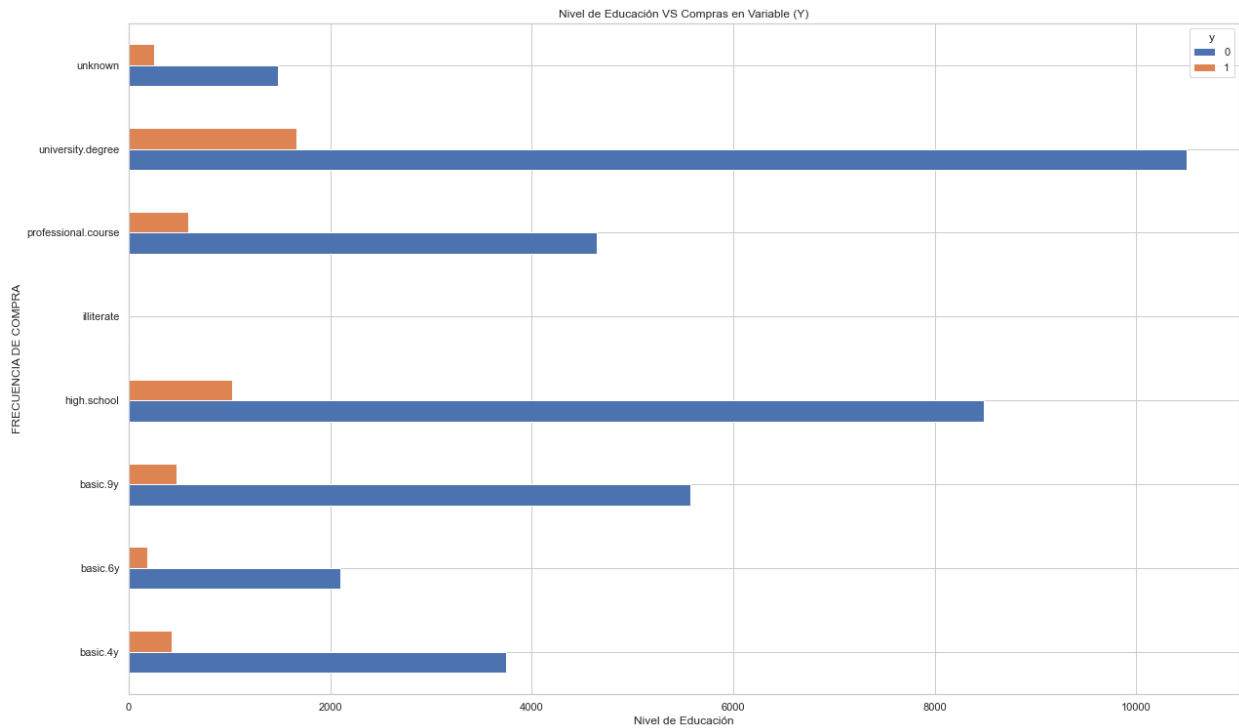


En la segunda gráfica se aprecia la variable Trabajo. En esta gráfica se identifica que los profesionales que no compran son los administradores y cuello azul (blue-collar), pero al igual son las profesiones que tienen una gran compra con respecto a las otras profesiones.



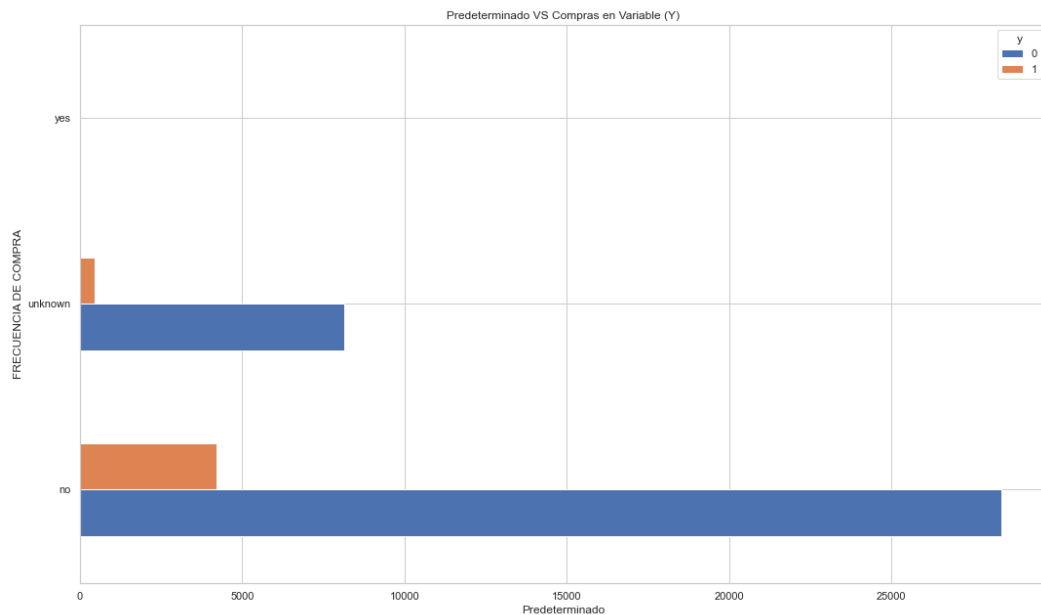
En la 3era gráfica se nota que siempre el no compra de productos con respecto a los que se compra, hay una diferencia bastante alta. En esta Grafica que se analiza el estado civil de

las personas que compran o no compran productos bancarios, las personas que más compran con respecto a las otras son las personas solteras, seguidas a las casadas, Al igual que el no compra de productos.

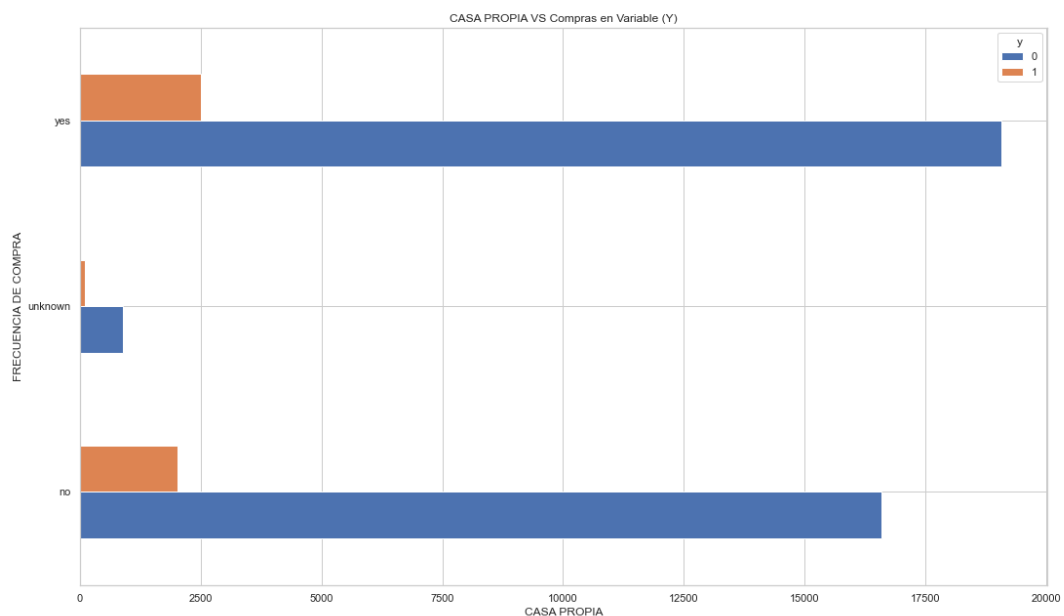


Para este analisis la compra de productos los que más rechazan son los profecionales universitarios seguido de los estudiantes que estan en Pregrado.

Para los productos se compran los que más frecuencia relativa tiene son los profesionales universitarios seguidos de los universitarios.

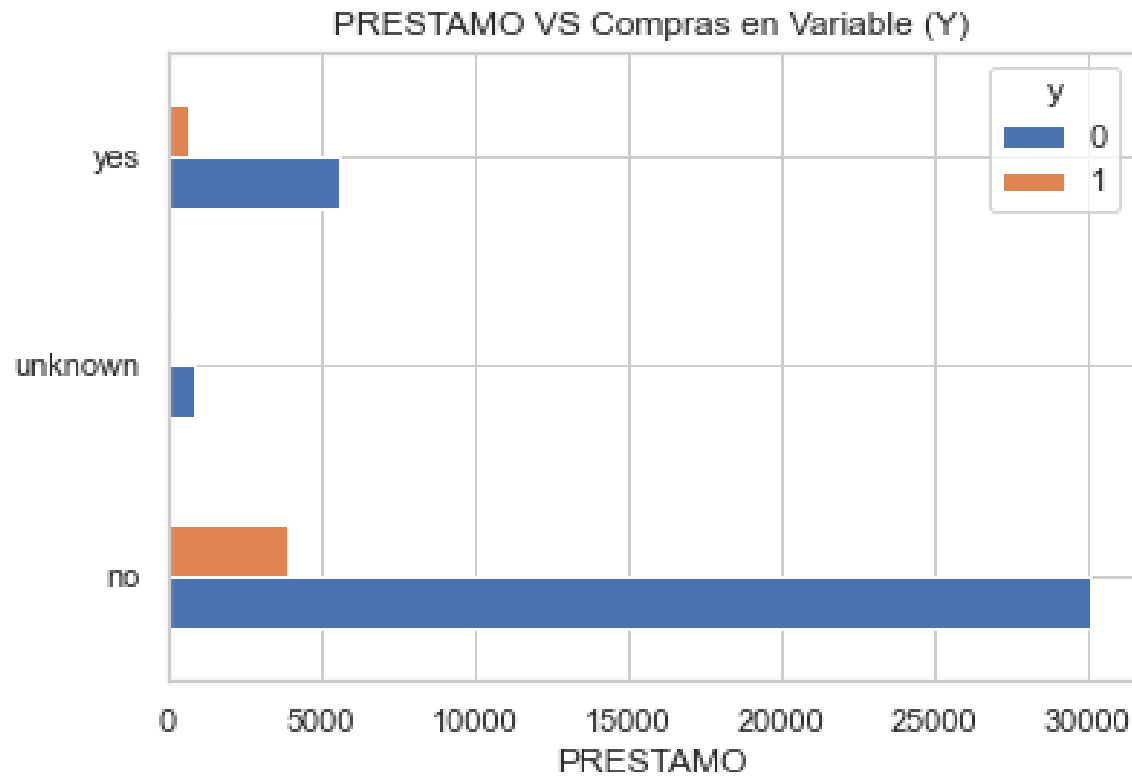


En esta gráfica se contrasta las personas que estan en mora para la salida de compra y no compra. La interpretación es que los que no estan sin ninguna mora, son los que más compran productos.

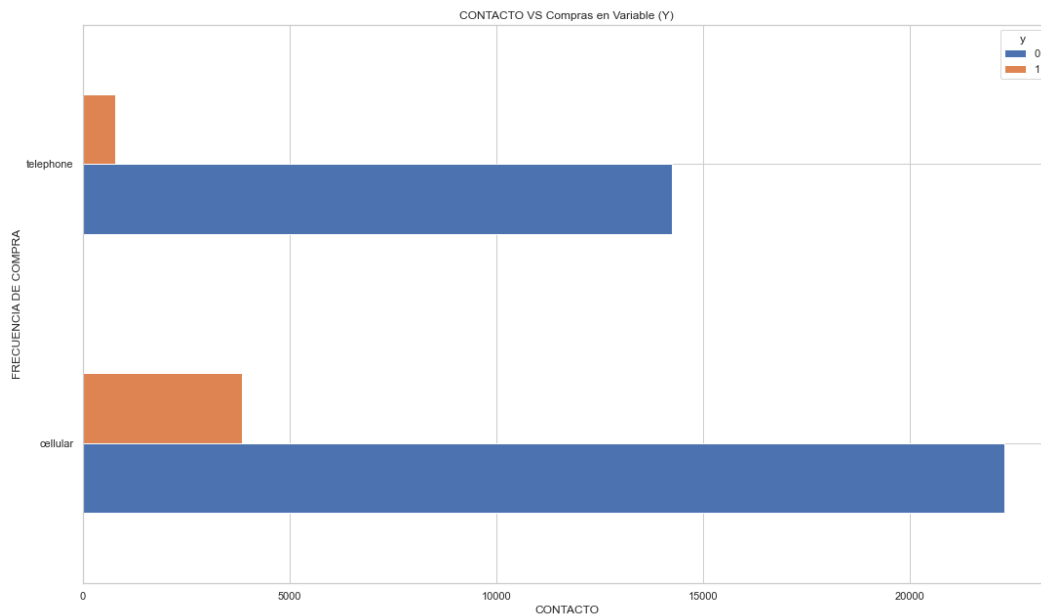


En esta grafica se interpreta que las personas que tienen casa propia y las que no tiene

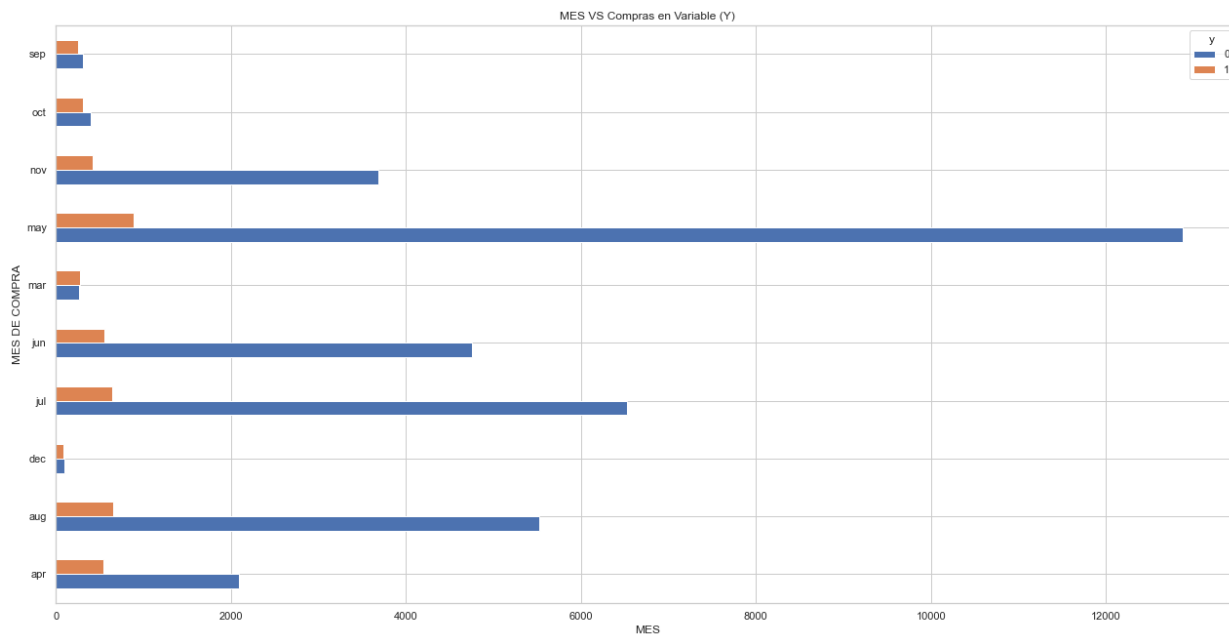
casa propia tienen casi una misma compra de productos de bancarios, pero el que más tiene un poco más de frecuencia son los de casa propia.



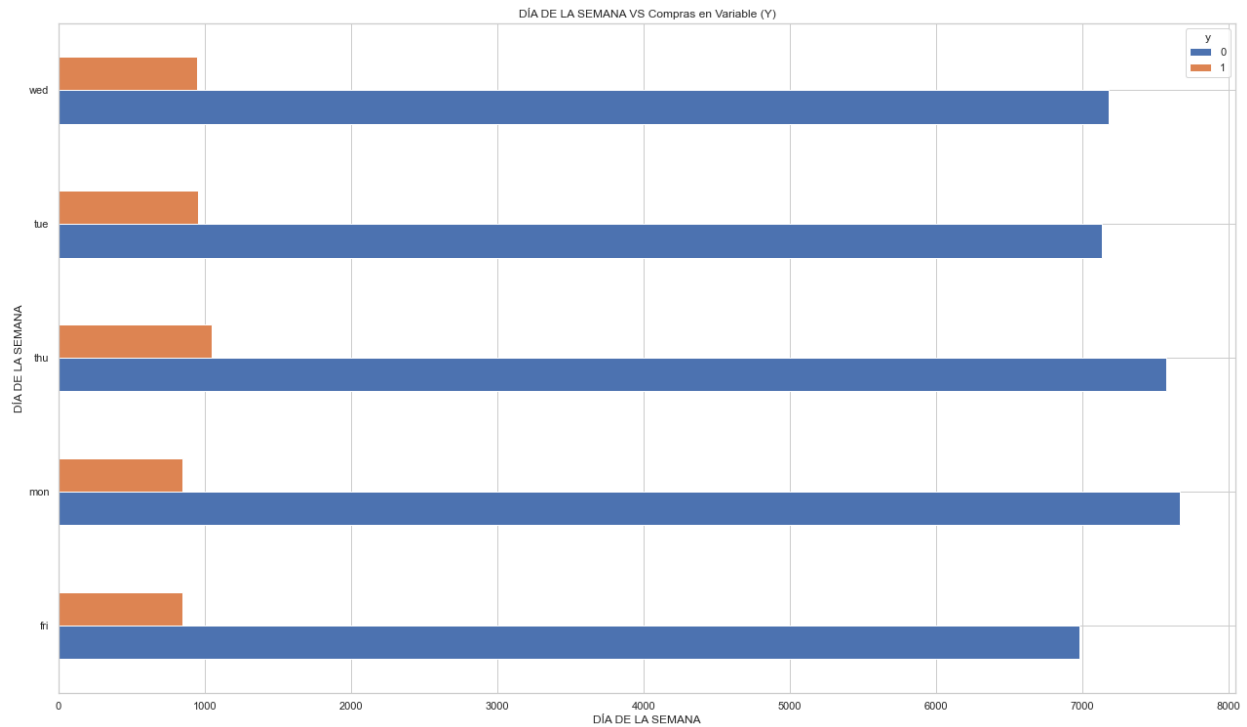
Para esta gráfica se muestra que los que más compran productos son las personas que no tienen algún préstamo.



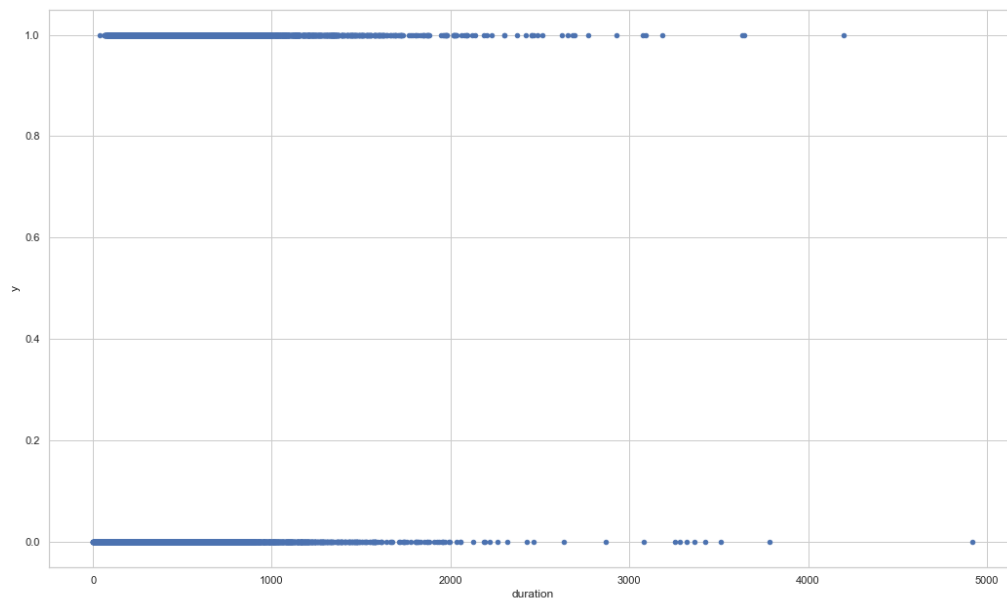
En esta gráfica se observa que los que más compran productos son los que contactan ppor celular, igual son que más rechazan productos.



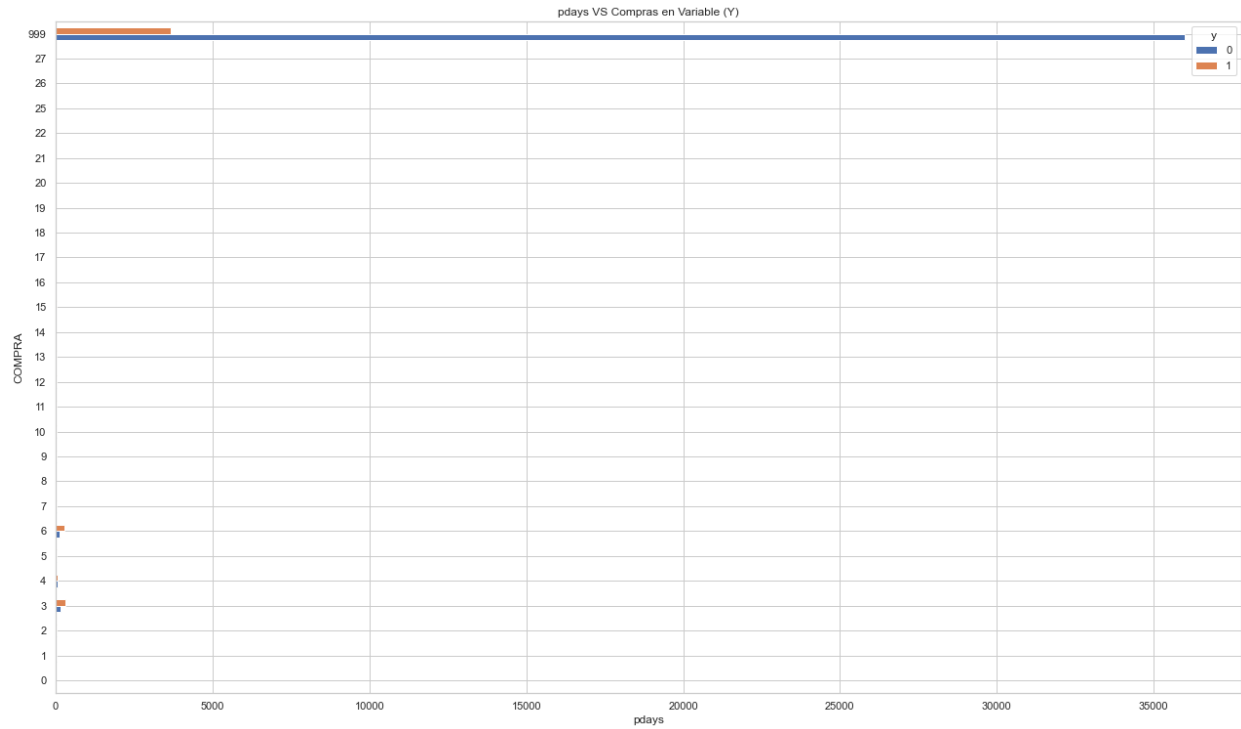
Se observa que en los meses que más se compra son los meses de mayo, agosto, julio y junio.



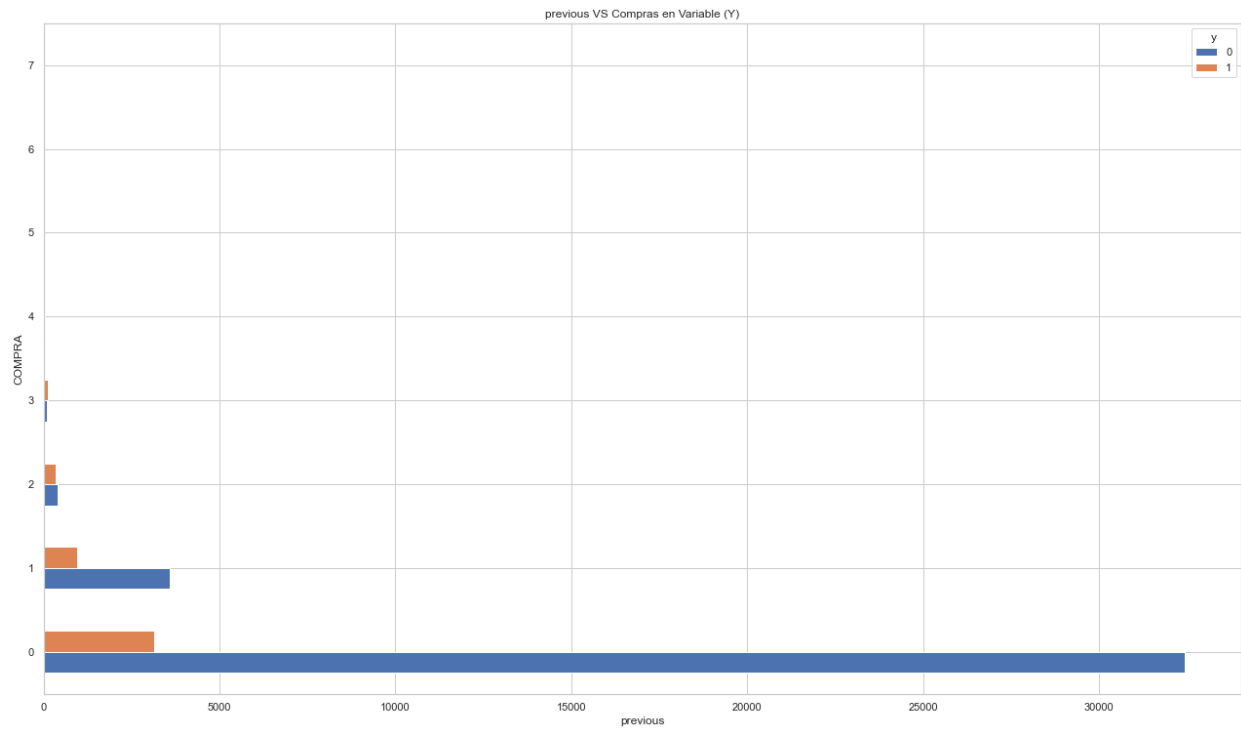
En esta gráfica se observa que todos los días de la semana se realiza casi con la misma frecuencia de compra productos bancarios.



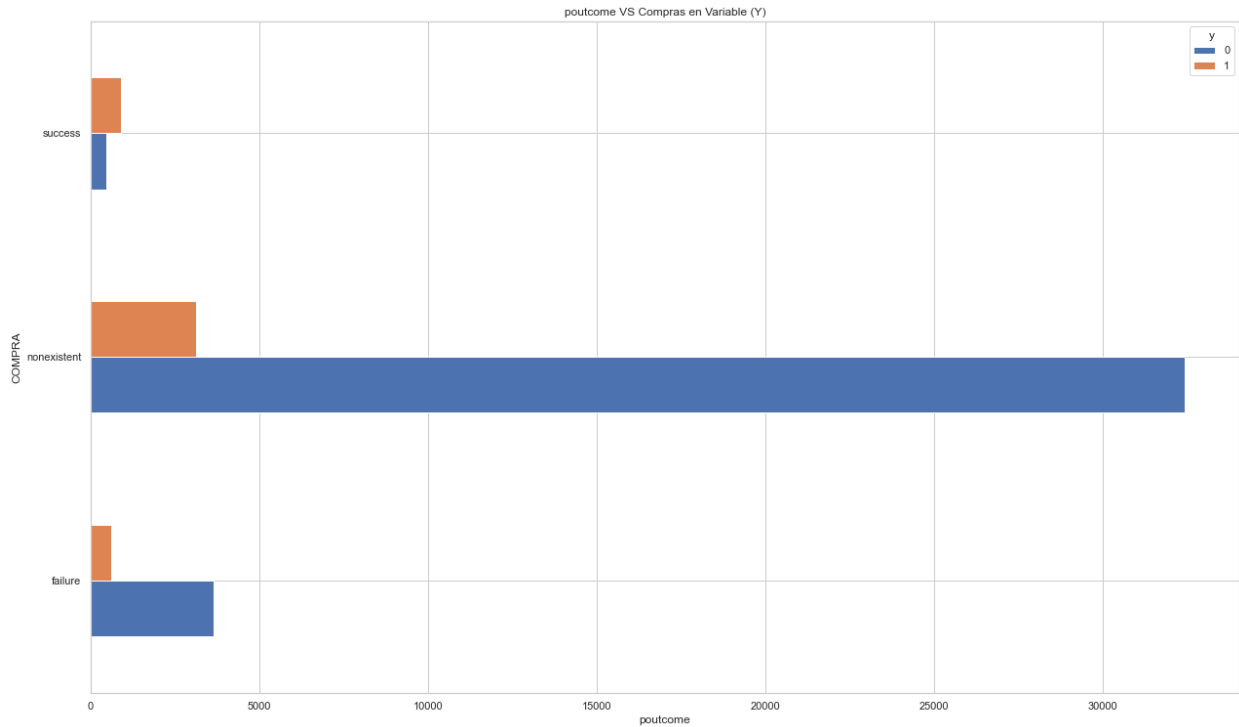
En esta gráfica se observa que la duración de las llamadas fue continua dentro de los 41189 registros por ello no se puede realizar un análisis.



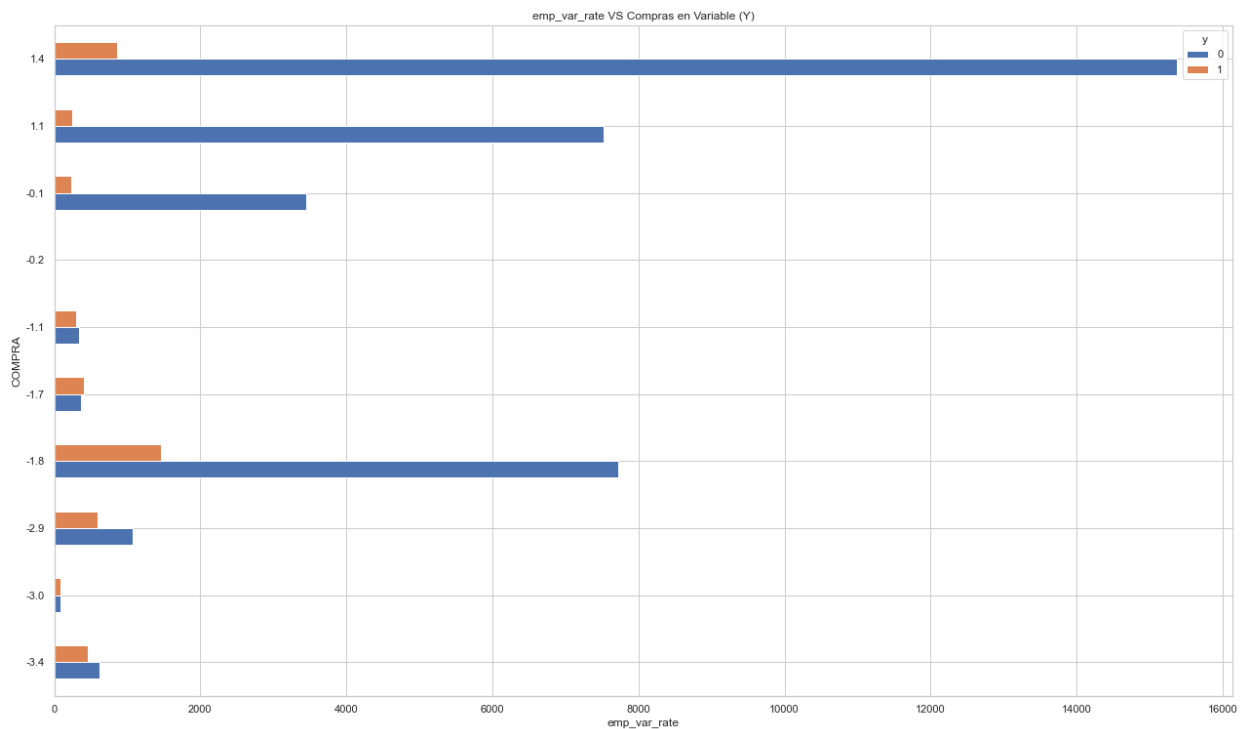
Se refiere esta gráfica a los ultimos días de campaña para ofrecer productos bancarios y se muestra que su mayor frecuencia esta entre los 999.



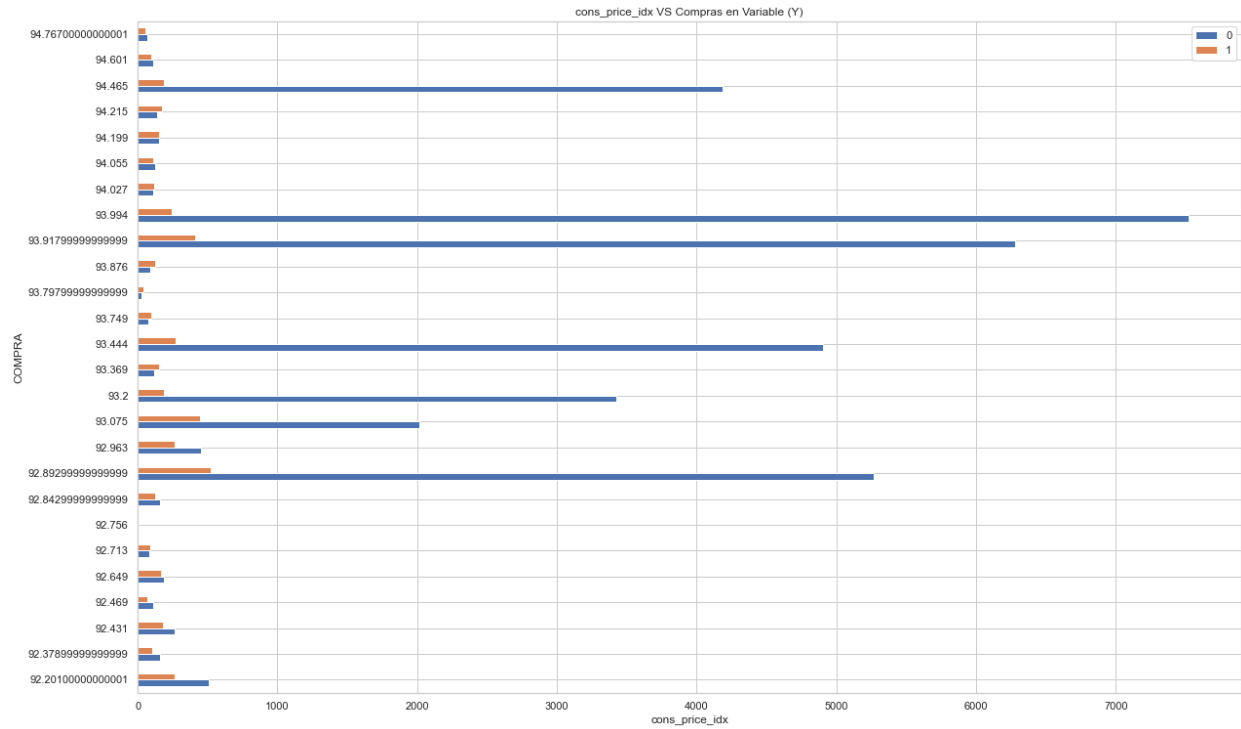
En está gráfica muestra que al número de veces que se realiza contacto antes de que la campaña sea lanzada, y se observa que la hay una mayor compra cuando estan a 0 días de lanzamiento.



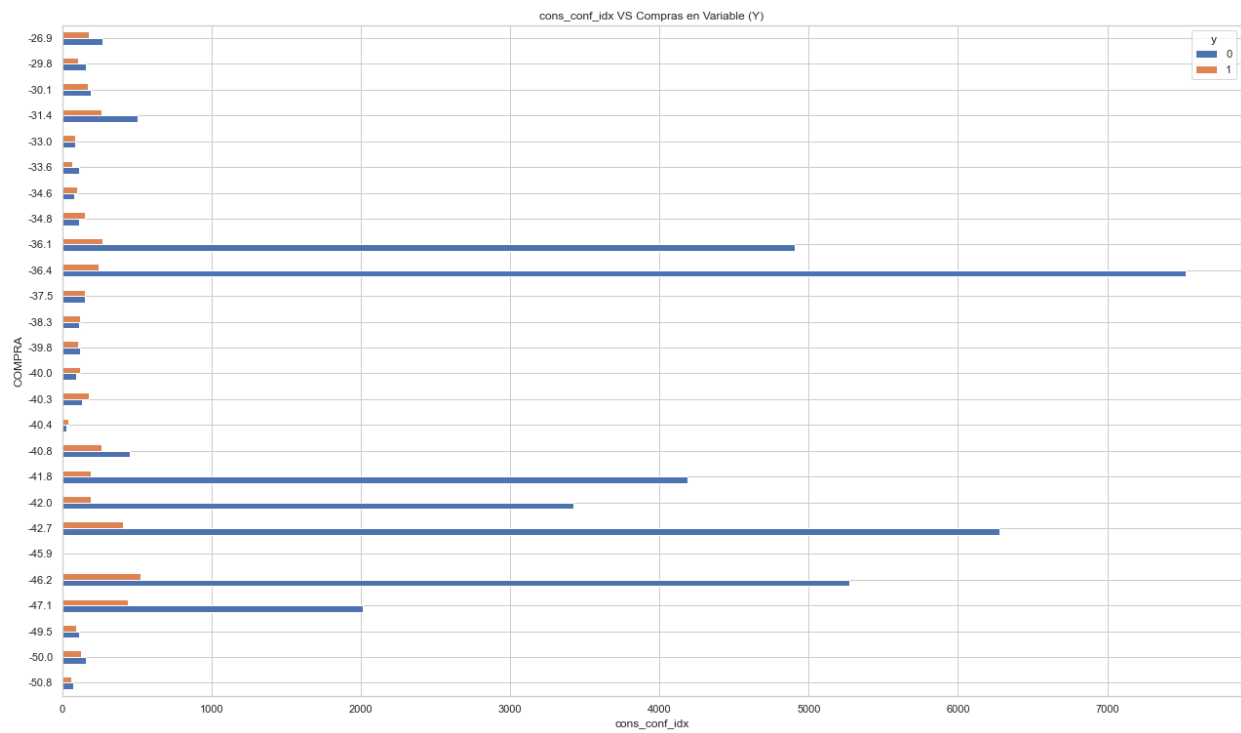
En la gráfica se aprecia la campaña de éxito, y la relación de la variable repuesta de compra o no compra, se realizó cuando no había existente campaña alguna de los productos bancarios.

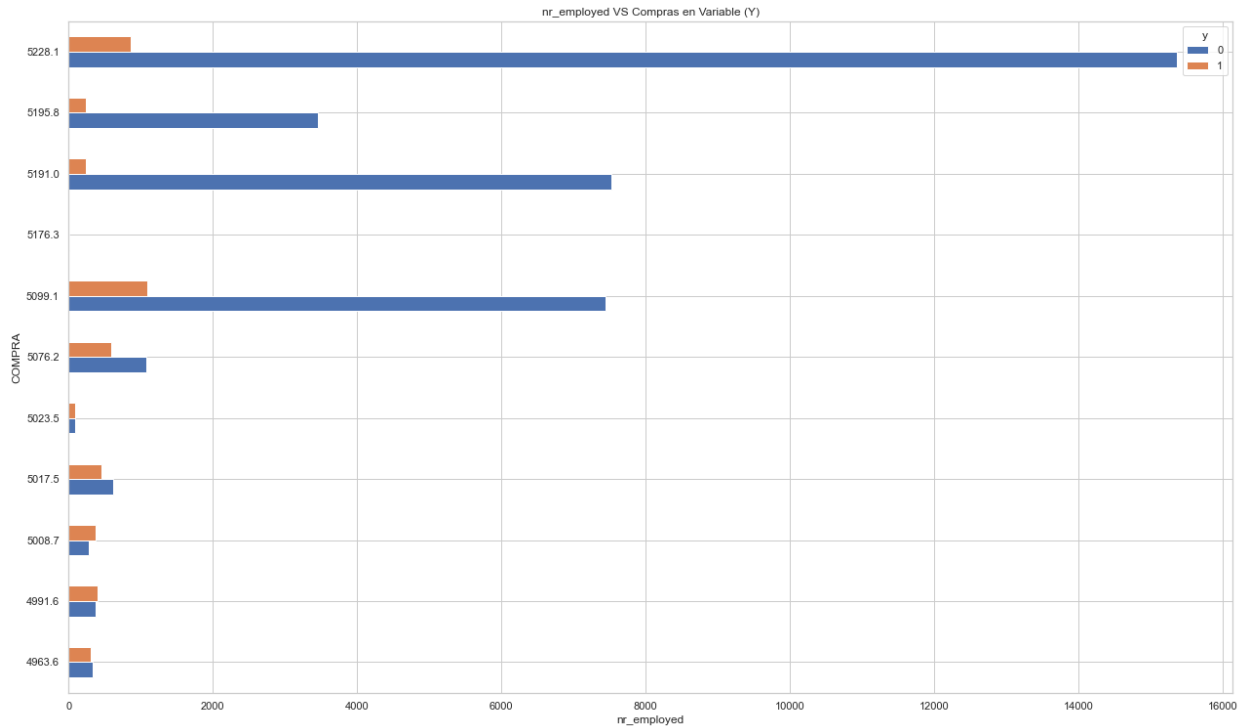


En la gráfica se aprecia la variable de variación de tasa de empleo de -1,8 hay mas frecuencia de compra de productos, seguido de la tasa de 1,4



En esta grafica de indice de precios del consumidor mensual





El código realizado en Python 3.7 fue el siguiente para generar las gráficas:

```
"""
```

Created on Thu Nov 19 18:44:21 2020

@author: milleralexanderquirogacampos

```
"""
```

LIBRERÍAS

```
import pandas as pd importamos librerías para analisis de datos
import numpy as np librería para análisis numérico
import matplotlib.pyplot as plt Librería de visualización
import seaborn as sns graficas más bonitas y estadística
import os Librería para el sistema operativo
from sklearn import preprocessing procesamiento de datos
from pandas profiling import ProfileReport
```

```
plt.rc("font", size=14)
sns.set(style="white")
sns.set(style="whitegrid", color_codes=True)
```

```
ruta del archivo donde esta alojado
os.chdir('/Users/milleralexanderquirogacampos/Documents/Documents - MacBook Pro de
Miller/MaestriaMatematicas/3 - Metodos de Analisis de
datos/LuzStellaGomez/Clase4')
```

```
cwd=os.getcwd() Asigna la variable cwd el directorio de trabajo
text file = "Banking.csv" Asignación del nombre del archivo a una bandera
```

```
sklearn = metodologías de modelos para regresión logística, neuronal,...,
from sklearn.linear model import LinearRegression Regresión Lineal
from sklearn.linear model import LogisticRegression Regresión Logística
```

```
df = pd.read_csv(text file)
```

```
ProfileReport(df)
reporte = ProfileReport(Banking).to_html("report.html")
reporte.to_csv(r'report.txt')
```

```
report = ProfileReport(df)
report.to_file('profile report.html')
```

```
df['duration'].unique()Permite obtener categorías únicas
df['age'].unique()Permite obtener categorías únicas
df['marital'].unique()Permite obtener categorías únicas
df['contact'].unique()Permite obtener categorías únicas
df['day of week'].unique()Permite obtener categorías únicas
df['poutcome'].unique()Permite obtener categorías únicas
df['y'].unique()Permite obtener categorías únicas
```

```
countno sub = len(df[df['y'] == 0])
countsub = len(df[df['y'] == 1])
```

```
pctof no sub = countno sub / (countno sub + countsub)
print("percentage of no subscription", pctof no sub * 100)
pctof sub = countsub / (countno sub + countsub)
print("percentage of subscription", pctof sub * 100)
```

```
pd.crosstab(df.nr_employed, df.y).plot(kind='barh')
plt.title('nr_employed VS Comprases en Variable(Y)')
plt.xlabel('nr_employed')
plt.ylabel('COMPRAS')
plt.savefig('nr_employed')
sns.countplot(x="nr_employed", df=df, palette="Blues")
sns.countplot(y="nr_employed", df=df, palette="GnBu_d")
```

```
df.plot.scatter(x='duration', y='y')
```