

INTERROGANTES EN EL ANALISIS DE DATOS “PRODUCTS_DF

Nos encontramos analizando el Dataset “Products_df”, y se nos plantean una serie de interrogantes tras haber analizado cada variable de manera independiente y habiendo conocido las características generales del mismo (índice, forma, columnas, registros etc...):

1. ¿CUÁL SON LOS REQUISITOS PARA SER CLIENTE DE EASYMONEY?

- a. Es preciso tener una cuenta básica Easy Money?
- b. Se puede escoger el servicio que se desee?
- c. No es preciso tener contratado ningún servicio para ser cliente?

Esta duda nos surge, tras darnos cuenta que en algunos periodos (meses), no hay servicios contratados.

2. Observamos que tenemos en cartera 456.373 clientes, de los cuales;

- el 76,8%, (350.384) tienen algún servicio contratado en algún de los meses
- y un 23,2% (105.989) no tienen ningún servicio contratado.

Puede que signifique, que **¿NO SON CLIENTES REALES? Y SON POTENCIALES CLIENTES, QUE SE HAN INTERESADO POR NUESTRO SERVICIO PERO NO HEMOS CONSEGUIDO CAPTARLOS?**

3. QUE SIGNIFICAN LOS 61 VALORES NULOS?: (en payroll y pension_plan)

- a. No se han introducido los datos?
- b. No se ha contratado ese servicio?

```
In [6]: print('Valores nulos:', df_products.isna().sum(), sep=2*'\n')
```

Valores nulos:

pk_cid	0
pk_partition	0
short_term_deposit	0
loans	0
mortgage	0
funds	0
securities	0
long_term_deposit	0
em_account_pp	0
credit_card	0
payroll	61
pension_plan	61
payroll_account	0
emc_account	0
debit_card	0
em_account_p	0
em_acount	0
dtype:	int64

4. Entendemos que la ingesta de datos, es el volcado de los datos mensuales, pero **¿ES POSIBLE SABER LA FECHA DE APERTURA DE CONTRATACIÓN DEL SERVICIO DEL CLIENTE?**
5. Vemos que cada mes, la cantidad de datos aumenta, **¿A QUÉ ES DEBIDO?**

a. Nuevos servicios contratados, incluyendo los nuevos clientes?

¿Cuántos datos se ingestaron cada día?

```
In [12]: df_products.pk_partition.value_counts()
```

```
Out[12]: 2019-05-28    442995
2019-04-28    439627
2019-03-28    436183
2019-02-28    431727
2019-01-28    426875
2018-12-28    422481
2018-11-28    416387
2018-10-28    402300
2018-09-28    375323
2018-08-28    352922
2018-07-28    339339
2018-06-28    252104
2018-05-28    249926
2018-04-28    247463
2018-03-28    245258
2018-02-28    242521
2018-01-28    239493
Name: pk_partition, dtype: int64
```

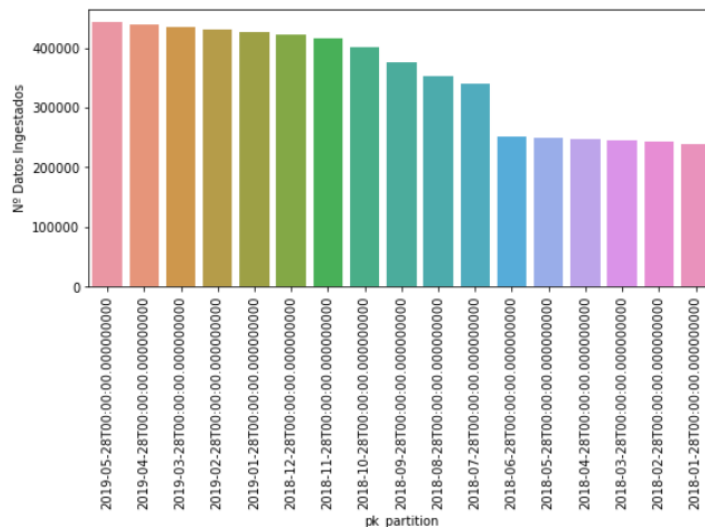
Fecha volcado de datos	cantidad de datos volcados	Dif mes a mes
2018-01-28	239493	
2018-02-28	242521	3028
2018-03-28	245258	2737
2018-04-28	247463	2205
2018-05-28	249926	2463
2018-06-28	252104	2178
2018-07-28	339339	87235
2018-08-28	352922	13583
2018-09-28	375323	22401
2018-10-28	402300	26977
2018-11-28	416387	14087
2018-12-28	422481	6094
2019-01-28	426875	4394
2019-02-28	431727	4852
2019-03-28	436183	4456
2019-04-28	439627	3444
2019-05-28	442995	3368
	TOTAL	203502

Hay un incremento de datos desde el 28 de enero del 2018 hasta el 28 de mayo de 2019, (que es el único registro de 203.502 datos. Siempre con una tendencia

ascendente. Observando que en fechas como julio del 2018, se produce casi el 25% de los volcados totales. **¿A QUE ES DEBIDO?** Y los meses de septiembre y octubre, el volumen de datos también es importante.

(se debe a alguna campana de mk, (descuentos, oferta de captación) nueva apertura de servicios? Nuevas condiciones?....

- b. Observamos que NO se reflejan las bajas de los clientes. **¿DE QUÉ MANERA OBTENERLAS?** Averiguar cuando dejan de volcarse datos de esos clientes...



6. **SI EN UN MES, NO HAY DESCARGA DE DATOS DE UN CLIENTE, ES PORQUE SE HA DADO DE BAJA? O PORQUE NO HA CONTRATADO NINGÚN SERVICIO?**
7. Hemos encontrado, algunos clientes, que solo tienen contratado algunos meses la tarjeta de debito, **¿NO DEBERÍA IR ASOCIADA UNA TARJETA A UNA CUENTA?**

Una vez descubramos estos interrogantes podremos avanzar con los siguientes datasets “comercial activity & sociodemographics” y conocer:

1. **Cuántos clientes tienen contratado un único servicio, o varios? ¿Cuál es el cliente con más servicios contratados?**
2. **% de fidelización de clientes?**
 - Cuántos clientes están desde el inicio?
 - Cuántos clientes se han dado de baja?
 - Cuántos hemos recuperado?
3. De donde son los clientes que compran nuestro servicio y por qué.
4. Que publico objetivo tiene cada servicio
5. Que rango de edad contrata cada servicio
6. Son hombres o mujeres
7. Con que media salarial
8. Etc..