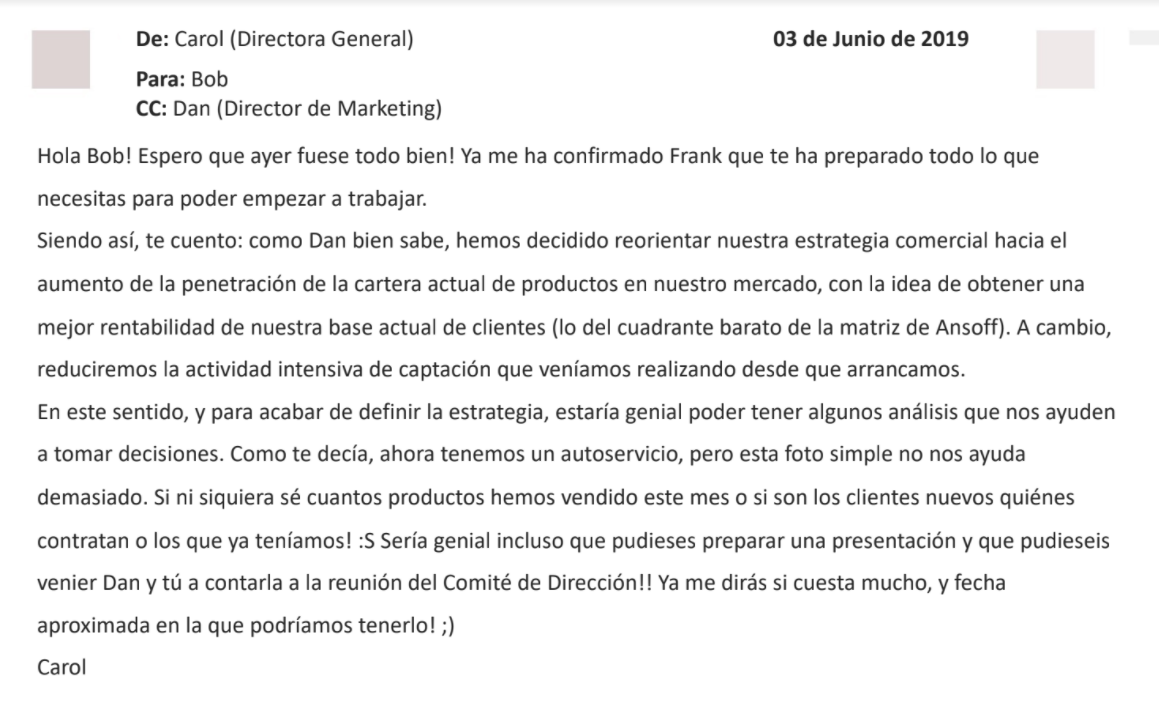
Tarea 1: Análisis



La directora general quiere definir una estrategia de marketing para poder tomar decisiones y adecuarlas a una mejor rentabilidad, partiendo desde el punto en el que tenemos un autoservicio pero información obtenía no es de gran ayuda, nos solicita que realicemos un análisis de las ventas del último mes para tener una mejor visión de las ventas de los productos.

A continuación expondremos una descripción técnica de los pasos que hemos seguido:

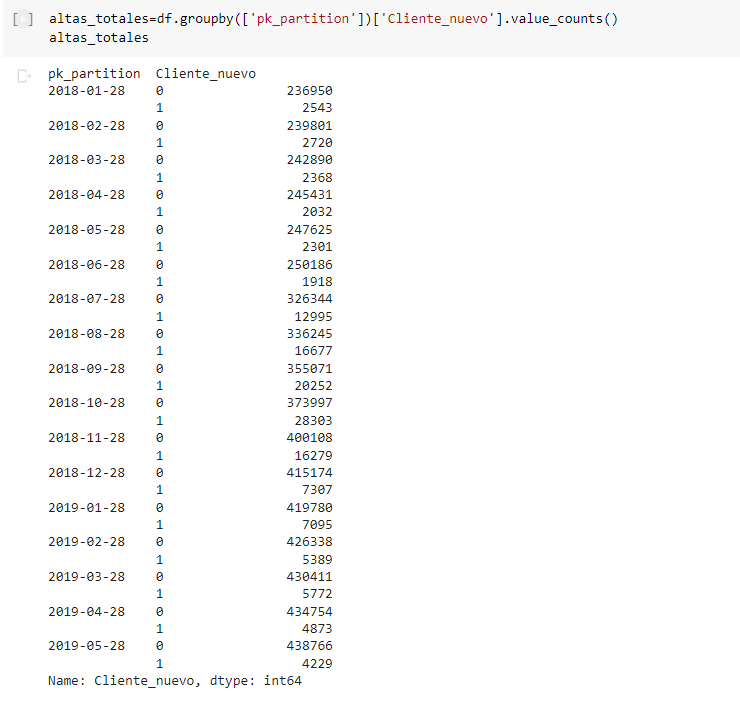
Luego de realizar el EDA y dejar los datos limpios de nulos y duplicados iniciamos con la convirtiendo los campos fechas en el formato correcto para luego crear el campo mes\_alta para poder obtener el campo cliente\_nuevo.

Creamos la función para poder calcular las diferencias mensuales, luego lanzamos un for loop para crear los campos a diferenciar por cada producto, limpiamos los nulos resultantes del for loop mediante la técnica .fillna igualándolos a 0.

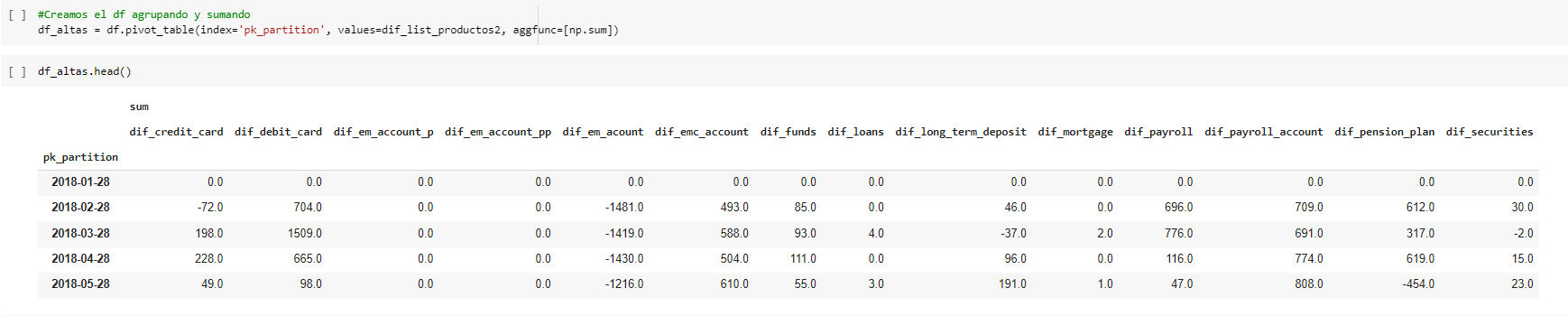
Realizamos varias comprobaciones para estar seguros que la función esta correcta.

Creamos el campo es\_activo para poder identificar los usuarios.

Filtramos las ventas por mes y cliente nuevo mediante el groupby obteniendo la cantidad clientes nuevos por mes.

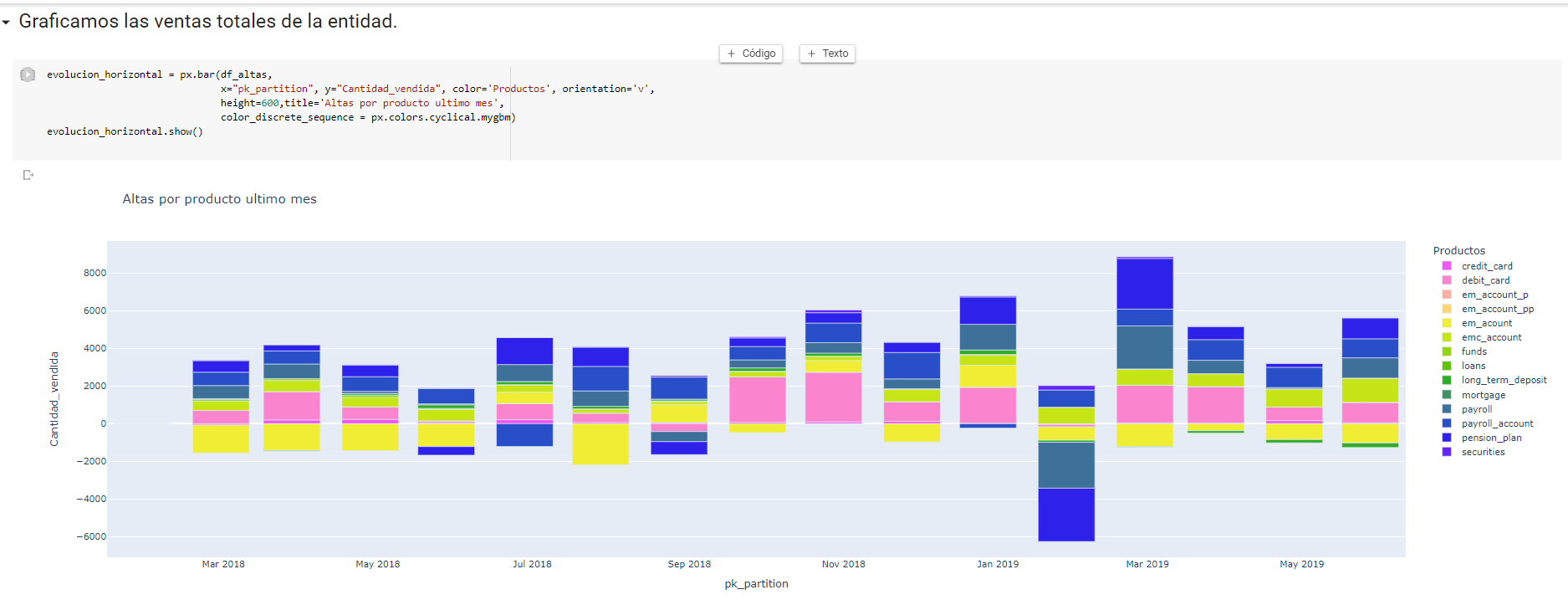


Creamos una lista de los productos dif para luego crear un nuevo dataframe llamado df\_altas agrupándolo vía pivot\_table y así poder obtener el resultado requerido de las altas y bajas de los productos por mes.



Para finalizar el análisis realizamos diferentes filtrados por mes.

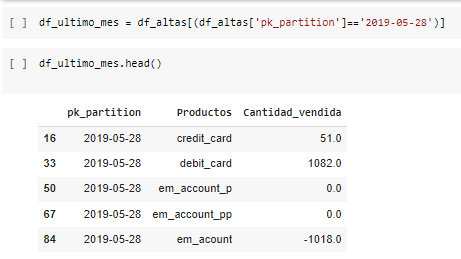
Ventas totales por productos.

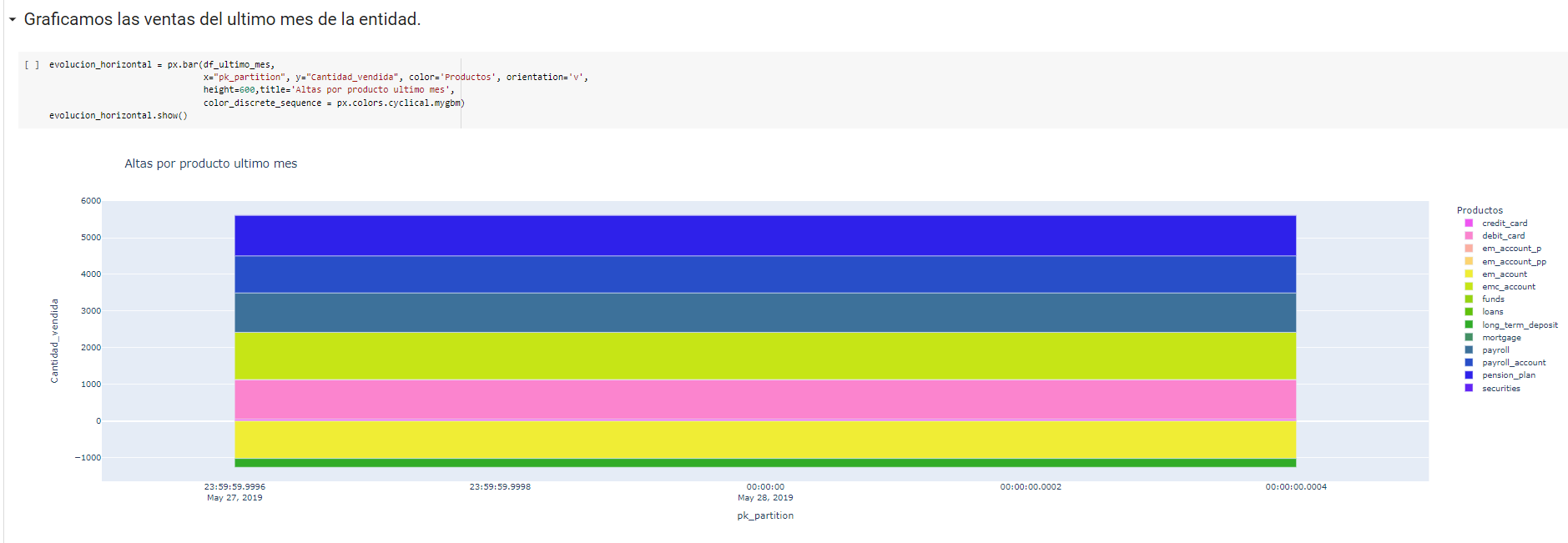


Ventas de los últimos 2 meses.

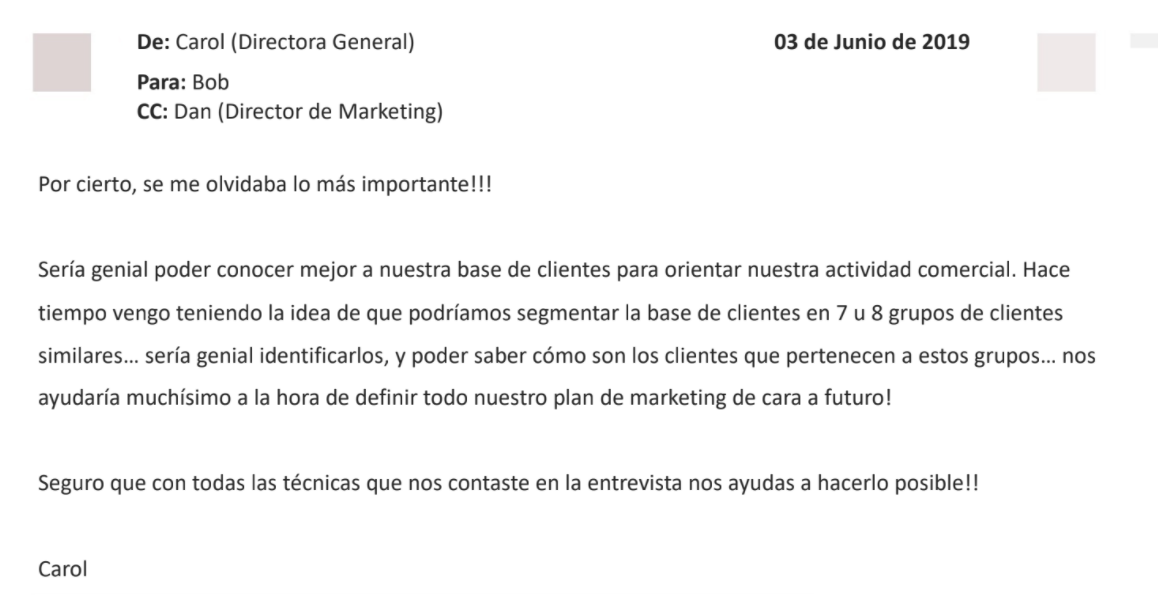


Ventas del último mes.





Tarea 2: Segmentación.



Tarea 3: Recomendación



El responsable del departamento de Marketing nos ha encargado la tarea de clasificar la base de datos de los clientes de tal manera que podamos predecir qué clientes son propensos a contratar los diferentes productos que ofrece la entidad bancaria.

A continuación expondremos una descripción técnica de los pasos que hemos seguido:

Para empezar hemos traído a un nuevo notebook el dataset procesado para la parte del análisis (df\_ final).

Creamos una nueva columna (‘Es\_Activo’). En esta nueva columna hemos querido saber si el cliente tiene o no contratados los diferentes productos y así poder clasificarlo.

Generamos una lista con los diferentes meses (lista\_fechas) y de esta manera podemos hacer el producto cartesiano (‘pk\_partition’ y ‘pk \_cid’). Creamos un data frame (‘df\_final\_’) en el cual indexamos este producto cartesiano.

Cambiamos el formato fecha y juntamos los dos dataframes, el df \_final y el df \_final\_, con merge por la izquierda y nos quedamos con el df \_final y borramos el df \_final\_

Este producto cartesiano genera nulos el mes en el que vamos a predecir, en nuestro caso Junio de 2019 por lo tanto los vamos a rellenar con un for loop y el atributo .map en una lista que contiene ‘entry date’, ‘age’, ‘salary’ , ‘Mes\_Alta’.

Nuestra estrategia será crear a partir de aqui columnas con los dif\_product y de esta manera podemos clasificar a los clientes con 1, 0, y -1. 1 para aquellos que compran el producto en ese mes, 0 si mantiene comprado o si desde el primer mes nunca lo compro y -1 cuando el cliente se da de baja en ese producto. Para generar estos shifts hemos creado 4 shifts por cada dif de producto y así tener un histórico por producto de cada cliente y poder utilizarlo para generar el modelo.

Los nulos que se generan de los diferentes shifts los rellenamos con un valor -100 para que no afecte al modelo posteriormente

Para iniciar el modelo seteamos el ‘pk\_cid’ cómo index y cogemos 6 meses para el train (del 09/2018 al 02/2019) para la validación solo cogemos el mes 03/2019 y para el test el mes 04/2019.

Para continuar debemos eliminar todas aquellas variables que contengan el TARGET, y tambien quitamos todos los -1 ya que queremos plantear el modelo de manera que solo nos prediga las altas, mas adelante podríamos realizar un modelo para las bajas. Pero por ahora lo haremos solo con la alta

Nos damos cuenta que el dataset está muy desbalanceado, encontramos muchos 0 y pocos 1 por lo tanto procedemos a realizar un resample utilizando una técnica de undersampling para evitar el overfitting del modelo.

Procedemos a generar modelos para los diferentes productos y utilizamos los 3 algoritmos de clasificación para ver cual nos puede funcionar mejor (Random Forest, Decision Tree y XGBoost)

Hay algunos productos que aunque realicemos un resampling el dataset se nos queda muy pobre y por lo tanto no podemos hacer un buen estudio. Para ello necesitamos mas datos sobre compras de los productos.

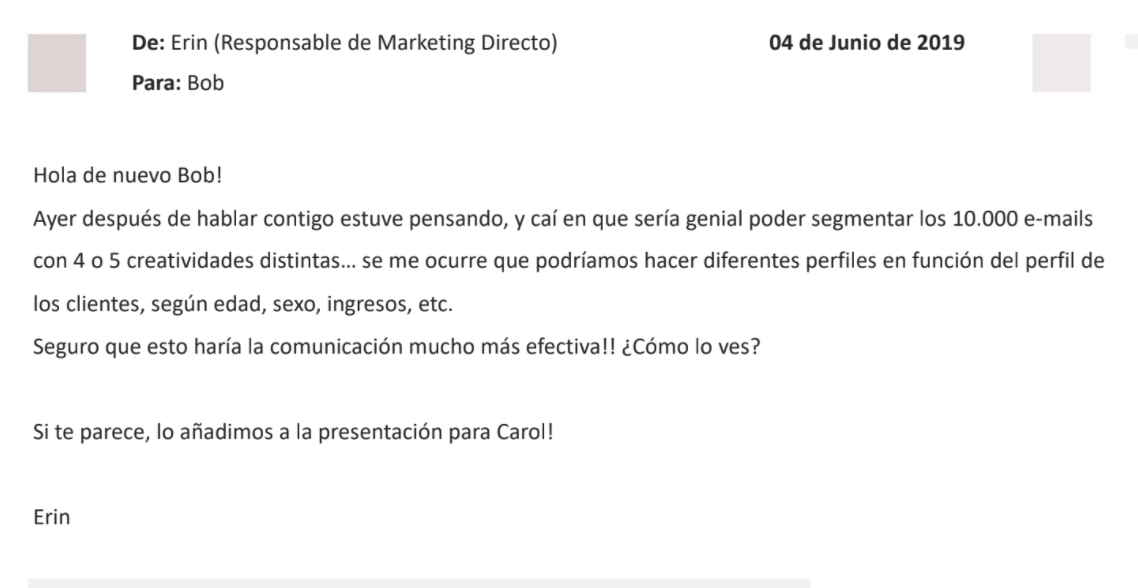
También sacaremos un top features y ver las variables más importantes que el modelo ha considerado para generar los sesgos de la predicción de cada producto.

A continuación compararemos los resultados de los modelos con una ROC curve y realizaremos una matriz de confusión en los diferentes modelos para ver el error entre lo que el modelo ha predicho y el dato real.

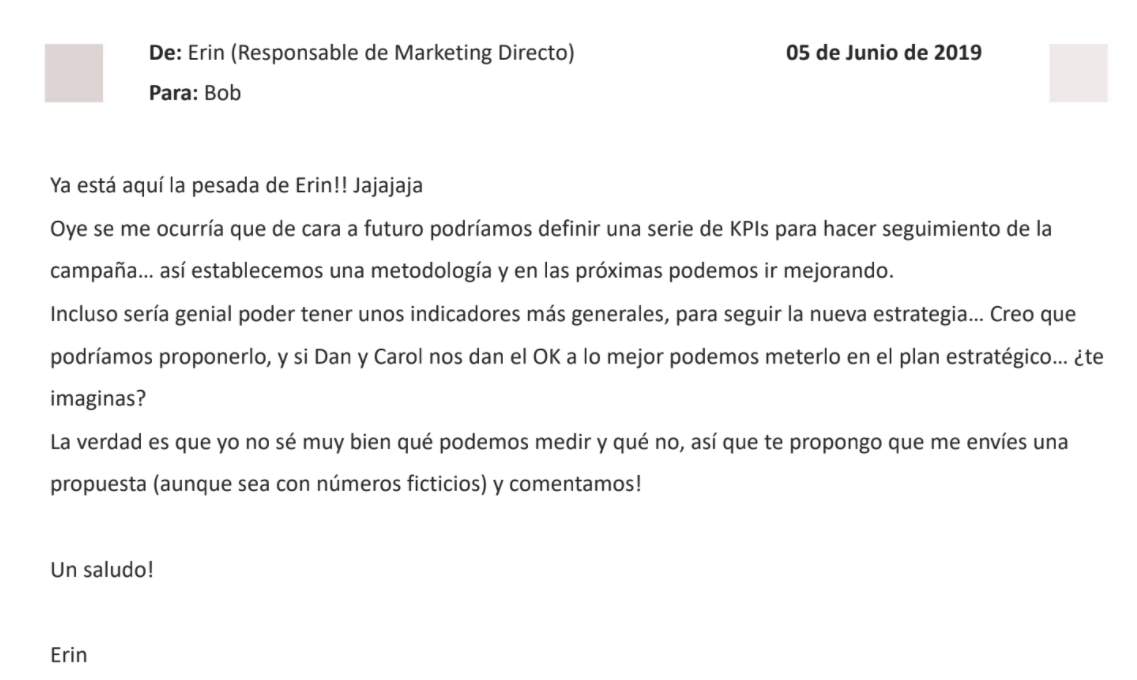
Para finalizar con el estudio utilizaremos la misma lógica para cada producto y de esta manera obtendremos un Data Frame el cual nos muestre todos los clientes que en el dif\_producto sea 0 y su prediccion sea 1.

Tras ver esta tabla y la tarea del producto nos dice que podremos sacar un beneficio por producto “10€ por cada cuenta vendida, 40€ por los productos de ahorro e inversión (planes, fondos, etc.) y 60€ por los de financiación (préstamos y tarjetas)” así que podremos presentar una aproximación de los beneficios sobre cada producto y organizar una campaña de marketing orientado a la compra, mientras que vemos que sería recomendable también ir revisando y completando los datos de este modelo para aquellos productos los cuales aun no podemos predecir por la escasez de datos.

Tarea 4: Personalización.



Tarea 5: Seguimiento.



Tarea 6: Coordinación.

