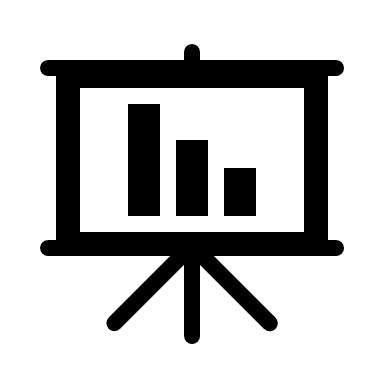
1. PROYECTO EDA EXCEL



1. DESCRIPCIÓN DEL PROYECTO

Partiendo de un archivo de datos en forma csv (**marketing\_campaign.csv)** obtenido a través de la web <https://www.kaggle.com/> procedemos a realizar la importación, depuración de datos y análisis de los mismos con objeto de dar respuesta y demostrar los conocimientos adquiridos en la parte del curso relativo al manejo de EDAs con hojas de cálculo, en este caso, Excel.

1. ESTRUCTURA DEL PROYECTO

Los pasos seguidos durante la elaboración del proyecto han sido los siguientes:

|  |  |
| --- | --- |
| **Volcar los datos** desde fichero csv (marketing\_campaign) | |
| Dejamos una hoja con los valores originales y hacemos una copia donde haremos las modificaciones que determinemos y que luego será la base del resto del dashboard | |
|  | |
| **Análisis descriptivo de los datos**: Contamos con los datos de una empresa consolidada que opera en el sector de la alimentación minorista. Venden productos de 5 categorías principales: vinos, productos cárnicos selectos, | |
| frutas exóticas, pescado especialmente preparado y productos de repostería. Estos se pueden subdividir en productos "gold" y regulares. Los clientes pueden pedir y adquirir | |
| productos a través de 3 canales de venta: tiendas físicas, catálogos y el sitio web de la empresa. A nivel global, la empresa tuvo ingresos sólidos y un margen de beneficio saludable | |
| en los últimos 3 años, pero las perspectivas de crecimiento de beneficios para los próximos 3 años no son prometedoras... Por esta razón, se están considerando varias iniciativas estratégicas | |
| para revertir esta situación. Una de ellas es mejorar el rendimiento de las actividades de marketing, con un enfoque especial en las campañas de marketing. | |
|  | |
| El **objetivo** del equipo es construir un modelo predictivo que genere el mayor beneficio para la próxima campaña de marketing directo, programada para el próximo mes. | |
| La nueva campaña, la sexta, tiene como objetivo vender un nuevo gadget a la base de datos de clientes. Para construir el modelo, se llevó a cabo una campaña piloto con 2.240 clientes. | |
| Los clientes fueron seleccionados al azar y contactados por teléfono en relación con la adquisición del gadget. Durante los meses siguientes, los clientes que compraron la oferta fueron | |
| debidamente identificados. El coste total de la campaña de muestra fue de 6.720MU y los ingresos generados por los clientes que aceptaron la oferta fueron de 3.674MU. Globalmente, | |
| la campaña tuvo un beneficio de -3.046MU. La tasa de éxito de la campaña fue del 15%. El objetivo del equipo es desarrollar un modelo que prediga el comportamiento del cliente y aplicarlo | |
| al resto de la base de clientes. Se espera que el modelo permita a la empresa seleccionar a los clientes con mayor probabilidad de adquirir la oferta, dejando fuera a los que no responden, | |
| haciendo que la próxima campaña sea altamente rentable. Además, aparte de maximizar el beneficio de la campaña, al CMO le interesa estudiar las características de aquellos clientes | |
| que están dispuestos a comprar el gadget. | |
|  | |
| **Revisar los datos** | |
| Ordenamos los datos por el identificador del cliente | |
| Búsqueda de datos outliners y vacíos y tratamiento de los mismos | |
| En la columna income tenemos datos vacíos y datos extremos. Usamos una gráfico de cajas y bigotes (en la hoja "datos originales" para ver los valores extremos que se salen del rango representativo | |
| Eliminamos las filas que contienen en esta columna los valores: 666.666, todos los valores mayores de 150.000 y los valores vacíos | |
| Con el gráfico de cajas y bigotes comprobamos que al quitar los valores indicados ya no quedan valores outliners. | |
| Revisamos el resto de columnas y no encontramos datos anómalos. | |
| Cuando incluimos la columna edad observamos que hay persona más que centenarias. Entendemos que se trata de una errata en la | |
| mecanizazión de los datos y procedemos a corregirlos (1899 ->1999, 1900 -> 1990, 1893 ->1993). Marcamos los datos cambiados con sombreado amarillo en la hoja DatosDepurados. | |
| Al revisar los datos por estado civil y fijarnos en la categoría YOLO (you only live once) vemos que solo hay una persona y que todos los | |
| datos de la misma son iguales salvo el identificador de cliente, por lo que entendemos que dicho cliente está repetido y lo quitamos | |
| de nuestro análisis, dado que solo es un cliente y no nos aporta información relevante desde el punto | |
| de vista de la relación conyugal, lo consideramos del mismo modo que los "absurdos" que vemos a continuación. | |
| Detectamos que tenemos dos clientes con un estado civil "absurdo" que no nos aporta información veraz, por lo que procedemos a | |
| prescindir igualmente de ellos. | |
| También apreciamos tres clientes cuyo estado civil es "alone" que lo equiparamos con "single". | |
|  | |
| **Añadimos columnas** en la tabla de datos depurados para incluir más información (se han sombreado con fondo amarillo) para hacer agrupaciones de datos (ventas en las diferentes | |
| campañas, agrupar clientes por generaciones, calculamos la edad de los clientes) | |
|  | |
| Creamos la diferentes **tablas dinámicas** con los datos relevantes para hacer un análisis de los datos que contemplen las diferentes variables recopiladas así como los | |
| **gráficos** correspondientes. Con ello se puede revisar la efectividad de cada una de las campañas, ver cuáles son los grupos de clientes que han sido más propensos a contratar | |
| los productos, para hacer hincapié en los valores incluidos en cada campaña. Así mismo se pueden ver las generaciones donde las campañas han sido más efectivas para | |
| hacer hincapié en las mismas de cara a promocionar y favorecer la venta en estos grupos que son más afines a la compra de nuestros productos. | |
| Con los gráficos y big numbers del dashboard así como jugando con las segmentaciones incluidas podemos jugar con los datos para señalar las generaciones de personas que van | |
| a ser nuestro grupo objetivo de cara a la nueva campaña, podemos ver tanto en número de aceptación como en términos económicos cuales son los grupos preferentes donde | |
| tenemos que centrar nuestros esfuerzos. | |

1. INSTALACIÓN Y REQUISITOS

La realización del proyecto se ha llevado a cabo con Excel 2019, ya instalado y configurado en el pc de trabajo.

1. RESULTADOS Y CONCLUSIONES

Finalmente, el proyecto relativo a EDAs con Excel ha quedado configurado en tres archivos, que serán los que se entregarán a través de la plataforma Github:

* + 00.ReadmeExcel
  + 01.marketing\_campaign
  + 02.ProyectoEDAexcel.01

1. PRÓXIMOS PASOS

Una vez realizado este proyecto y procedido a la entrega, espero recibir feedback correspondiente al mismo para comprobar el grado de aprendizaje alcanzado y las mejoras a considerar para sacar un mayor provecho de la herramienta, de manera que, no solo se obtengan los resultados exigidos, sino también de la manera más óptima posible de cara a la optimización de los recursos.