



CAFEINOMANOS

Un pequeño sorbo hacia el mundo de la ciencia de datos.



Contenido

QUIENES SOMOS

Acerca de nuestra empresa y servicios.

01

CLIENTE

Descripción del cliente y requerimientos

02

PUNTO DE PARTIDA

Con qué datos partimos para nuestro análisis

03

ANÁLISIS

Principales análisis univariados, bivariados y multivariados

04

ALGORITMO

Elección del modelo y breve explicación

05

CONCLUSIONES

06





01

QUIENES SOMOS

Acerca de nuestra empresa y servicios.



Maria Victoria Yaconisi



Mateo Aguiar

EQUIPO



Nicolás Lopez



Mauro Rodriguez Vidal

Acerca de nosotros

Somos una organización ad honorem, creada por 4 integrantes fanatizados y motivados por el café. El nombre de la misma es 'Cafeinomanos' y nos dedicamos a brindarles soluciones en cuanto a lo que la ciencia de datos abarque y esté al alcance, a distintas empresas, sobre todo aquellas relacionadas con el ambiente y mundo del café, con el fin de resolver ciertas necesidades haciendo nada más y nada menos que hablar a los datos.



CLIENTE

Descripción del cliente y requerimientos

02



Cadena de café.

Nuestro cliente es una **cadena de café** que nos brinda sus datos sobre ventas de abril 2019 junto con otros datos esenciales de la organización.

El hito principal y motivación por parte del cliente es la **búsqueda de una acertada campaña de marketing** orientando la misma a aquellos nichos y segmentos de clientes y así atraer a los mismos hacia la empresa.

Logrando como resultado un **incremento en la totalidad de sus clientes**. A partir de las ventas brindadas por parte del cliente sobre abril 2019.

¿A qué nichos y perfiles de clientes necesitarán orientar sus próximas campañas publicitarias de manera que los resultados se vean reflejados en un incremento en ventas?

03

PUNTO DE PARTIDA

Con qué datos partimos para nuestro análisis



DATASET

Contamos con un conjunto de distintos datasets orientados al café y a lo que su negocio respecta.

Customers ☞ Información referida en cuanto a los clientes con sus datos principales

Generations ☞ Información referida en cuanto a los nombres de generaciones a partir de sus años de nacimiento

Product ☞ Información referida en cuanto a los productos con sus características principales

Sales Target ☞ Información referida en cuanto a los targets establecidos

Sales Outlet ☞ Información referida a los sales outlet y algunas características

Pastry Inventory ☞ Información referida a los manejos de stock de algunos productos de manera diaria

Sales Recieps (201904) ☞ Información referida en cuanto a las ventas ocurridas en el período de abril 2019

Dates ☞ Información referida a fechas, intentando ampliar la información a partir de una fecha determinada

Staff ☞ Información referida a los empleados de la cadena y a sus principales datos



04

ANALISIS

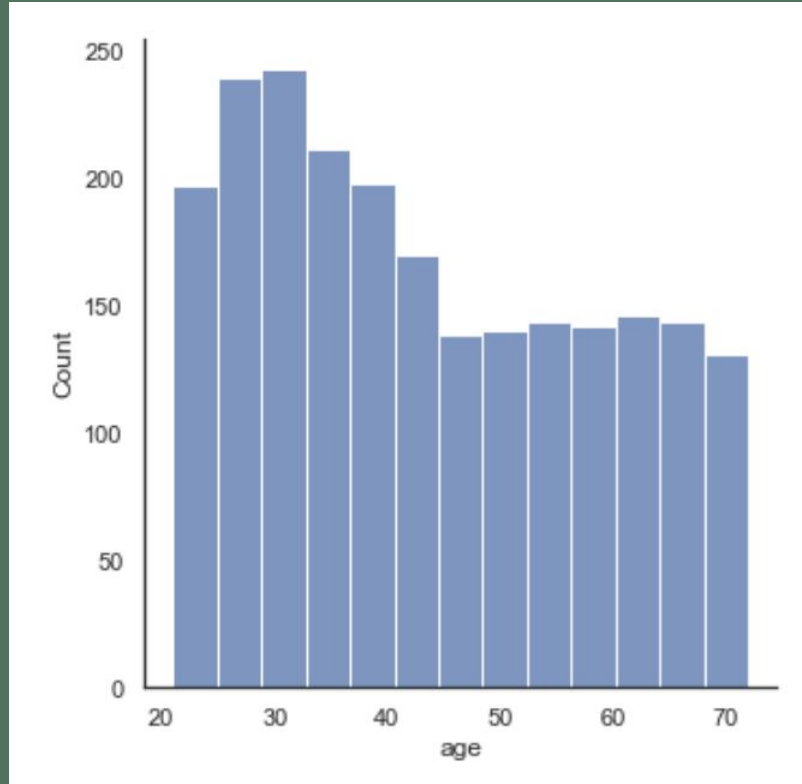
Principales análisis univariados, bivariados y multivariados.

Partiendo de la pregunta comentada previamente, nos pareció relevante enfocar nuestros distintos análisis a la información propia de los clientes. Como así lo es su edad, sexo, generación y totales gastados.

UNIVARIADO

Edad

Podría decirse que existe una mayoría sobre aquellos clientes que rondan entre los 21 y 40 años de edad, pero vemos que la diferencia con las edades que le siguen no son cambios muy abruptos.

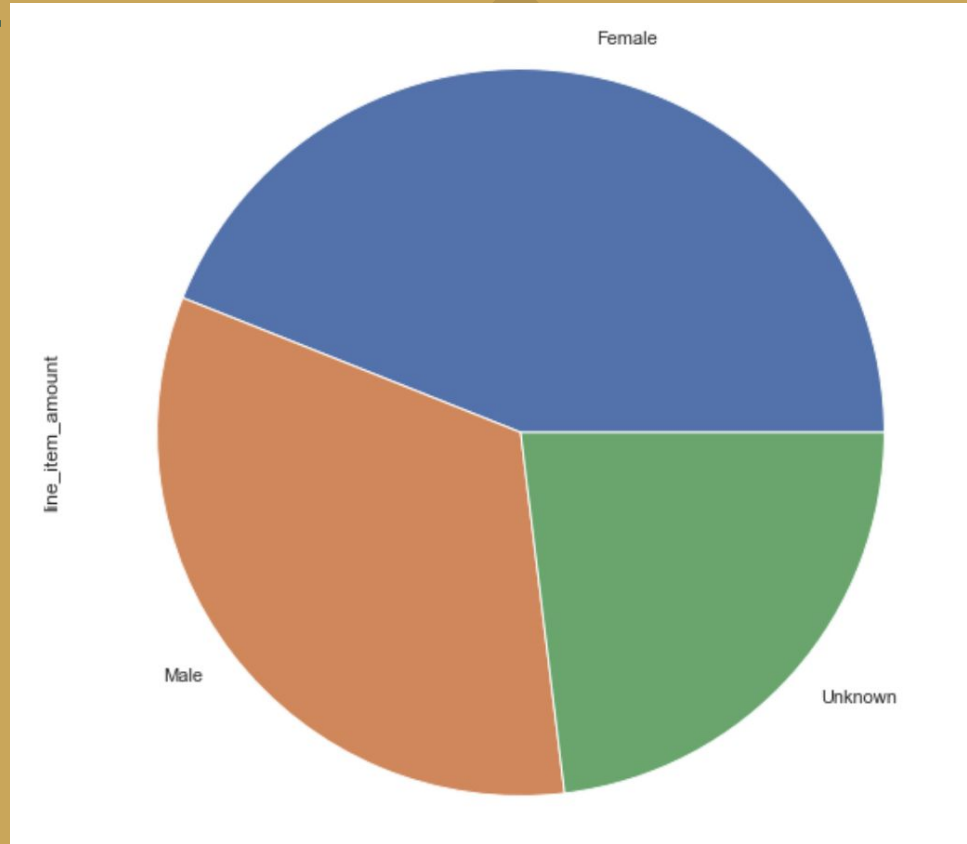


Lo que podemos deducir acerca del análisis hecho sobre la edad de nuestros clientes, es que se mantienen un promedio de edad bastante calmo, sin ver picos extremistas en ningún sector de edades en particular.



BIVARIADO

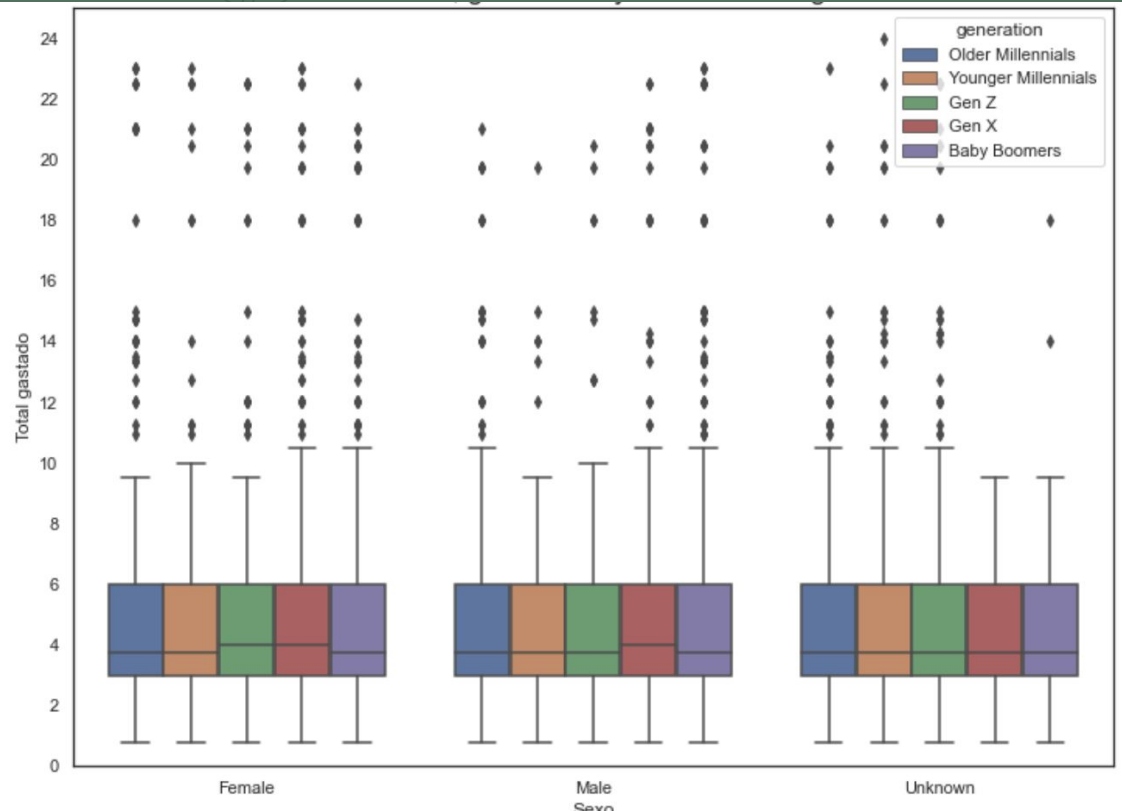
Sexo - Total gastado



Podemos visualizar que el género femenino es el que más gasta en las compras del café para las compras de abril de 2019

MULTIVARIADO

Sexo – Generación – Total gastado



Con este gráfico de caja y bigotes podemos apreciar que las distintas generaciones en sus distintos sexos consumen un monto similar, hablando de una media de \$3,5 o \$4.

05

ALGORITMO

Elección del algoritmo y breve explicación



El dataset se compone de 12 columnas de las cuales 11 corresponden a variables que usaremos para predecir el target.

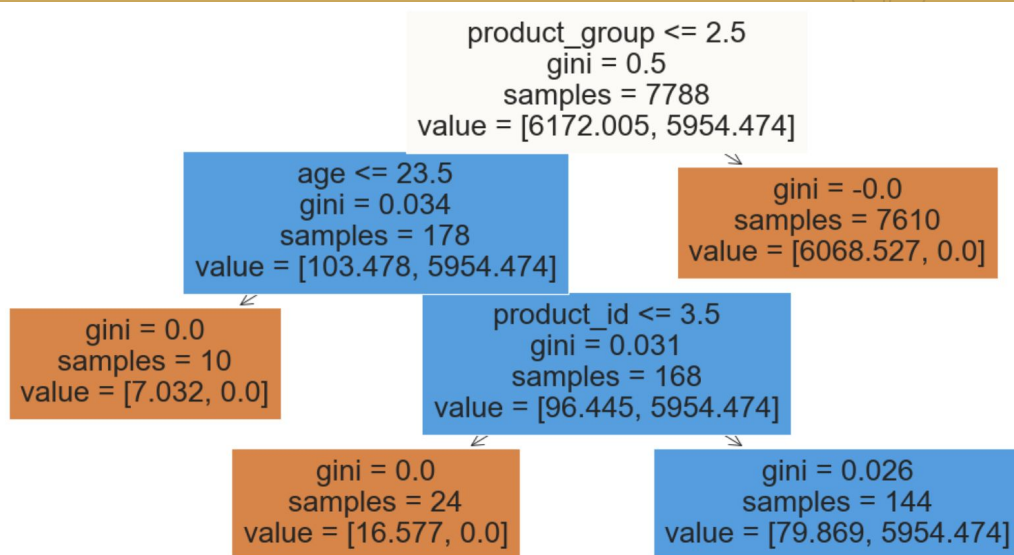
MoreThan20 es la variable a predecir. Esta variable se trata de aquellas transacciones de los clientes en las que se gasta más de 20 usd. Se coloca un 1 cuando se supera los 20 y un 0 en el caso contrario.

Todas las variables son numéricas, ya sea enteras o reales y no tiene valores nulos.

RANDOM FOREST

Test accuracy: 0,986065

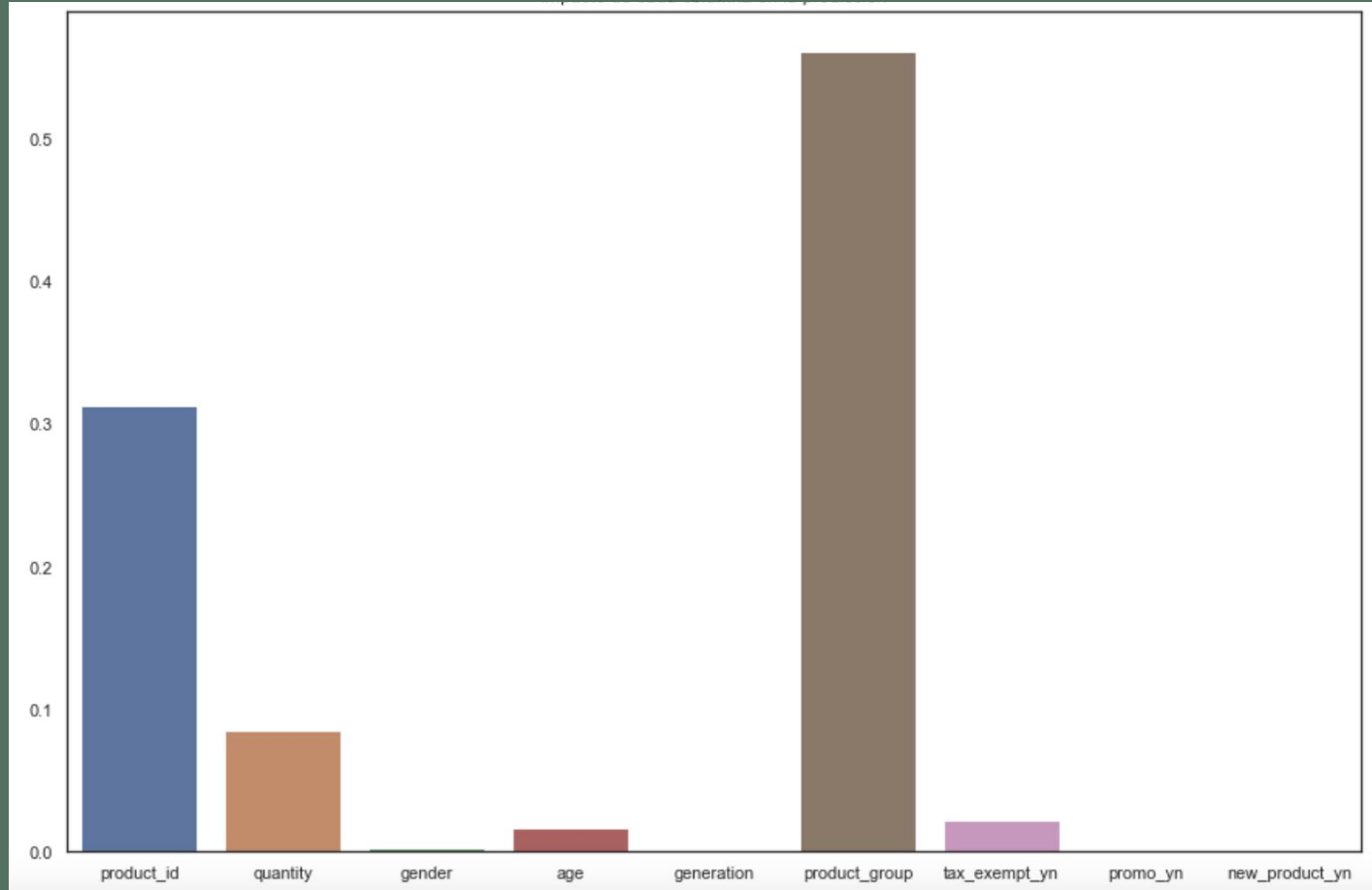
Graficamos el modelo:



	product_id	quantity	unit_price	gender	age	generation	product_group	tax_exempt_yn	promo_yn	new_product_yn	MoreThan20
0	52	1	2.5	2	39	0	5	1	0	0	0
1	52	1	2.5	1	39	0	5	1	0	0	0
2	52	1	2.5	1	39	0	5	1	0	0	0
3	52	1	2.5	0	39	0	5	1	0	0	0
4	52	2	2.5	1	39	0	5	1	0	0	0

RANDOM FOREST

Impacto de cada
columna en la
predicción



KNN



```
: escalador = MinMaxScaler()

x = dataset_final.drop(["MoreThan20", "unit_price"], axis=1)
y = dataset_final["MoreThan20"]

x = escalador.fit_transform(x)

(x_train, x_test, y_train, y_test) = train_test_split(x, y, stratify=y, test_size=0.50, random_state=11)
```

Genero el modelo y lo entreno

```
: clasificador = KNeighborsClassifier(n_neighbors=3)

clasificador.fit(x_train, y_train)

: KNeighborsClassifier(n_neighbors=3)
```

Evaluo el desempeño del modelo

```
: clasificador.score(x_test, y_test)

: 0.9953819978935429
```

A person wearing a blue denim shirt and brown suspenders is pouring water from a brass kettle into a glass coffee dripper. The dripper has a white paper filter and is filled with coffee grounds. The background is dark and out of focus.

06



CONCLUSIONES

Para el dataset elegido, podemos ver que ambos modelos funcionan de forma muy similar. Tanto el Random Forest como el KNN dan predicciones muy cercanas a 1.



En relación a lo analizado, consideramos que las dos variables más importantes sobre las cuales definir una campaña de Marketing fructífera son sexo y producto.

Tal como se vió en el análisis Bivariado, las mujeres son los clientes que más consumen, por lo tanto, iniciar una campaña en función de ellas ayudará a aumentar las ventas a nivel general.



En cuanto a los productos, consideramos que publicidad dedicada a aumentar los grupos de elementos más vendidos y reducir la cantidad total de opciones es la mejor estrategia para mejorar los ingresos.





GRACIAS!

[Link a nuestro proyecto](#)

