

Predict Customer's preferred Brand

Presentación: Luis Sánchez Peña
sanchezisp@gmail.com

Fecha: 25 de Septiembre de 2017

Customer's preferred Brand

Introducción

Blackwell Electronics' CTO Danielle Sherman quiere que completemos la encuesta hecha por el departamento de marketing.

La encuesta busca conocer la marca preferida por los clientes pero parte quedó incompleta. Se desea conocer la marca preferida de todos los clientes encuestados, incluyendo los incompletos.

Debe entregarse un informe con los métodos usados y las salidas devueltas por R.

Customer's preferred Brand

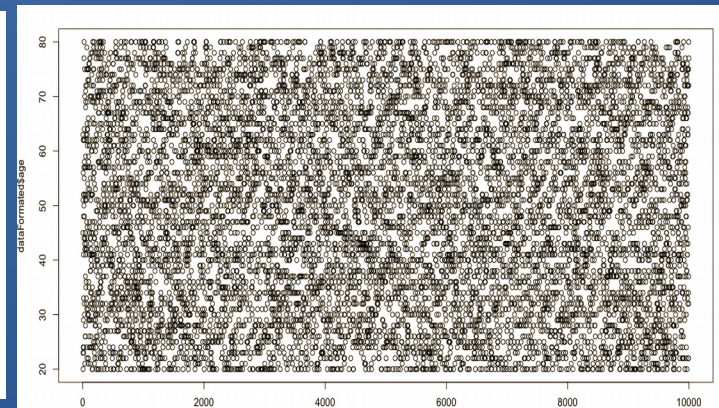
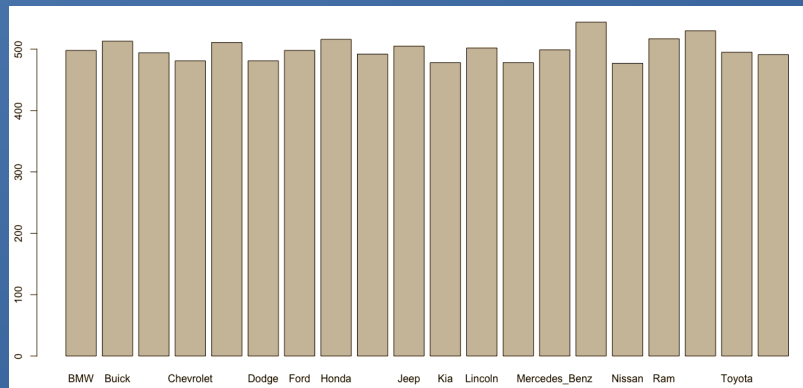
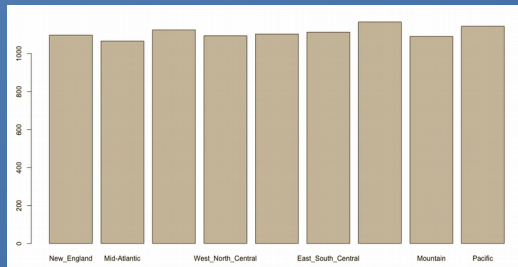
Fuente de datos

- La empresa entrega dos ficheros resultantes de la encuesta:
 - Un CSV con 10.000 instancias completas: tabla1.
 - Una hoja de cálculo tipo Excel con 5.000 instancias defectuosas (falta el campo “Brand”): tabla2.
- En uno de los documentos tenemos las leyendas de cada campo.

Customer's preferred Brand

Fuente de datos

- El fichero completo parece ser una encuesta perfecta. Tiene las instancias perfectamente distribuidas en todos los campos.
- ¡El fichero incompleto también tiene todas las instancias perfectamente distribuidas! ¿-?
- No se observa correlación directa entre datos numéricos.
- ¿Es posible una encuesta perfecta?



Customer's preferred Brand

Análisis realizados

Buscamos el mejor modelo para completar la tabla2.
Comparé los métodos knn y Random Forest (RF) ambas con distintas semillas estadísticas y cross-validation (10)
Los resultados fueron:

modelo	semilla	k	Acc train	Kappa train	Acc test	Kappa test
KNN998	998	15	0.5989332	0.0653966	0.6166467	0.1063153
KNN1234	1234	19	0.6004803	0.0603191	0.6030412	0.0613679
RF1234	998		0.9224516	0.8352687	0.9231693	0.8373113
RF998	1234		0.9224898	0.8352859	0.5166066	-0.024852

Customer's preferred Brand

Reflexión

- ¿Tiene sentido usar RF para completar una tabla y luego inferir relaciones?
 - Si la tabla 1 ya da buenos resultados ¿Para qué mas datos?
 - Si la tabla1 no aproxima bien.¿Por qué usarla? Solo amplificará el error al rellenar la tabla2.

¡Mejor volver a la pregunta!

¿Qué marca quieren nuestros clientes?

Customer's preferred Brand

Reflexión

¿Qué marca quieren nuestros clientes?

Respuesta: La que más compran.

Customer's preferred Brand

Reflexión

Los datos que necesito para responder a esa pregunta son los de ventas.

Customer's preferred Brand

2ª Reflexión

¿Para qué sirve la encuesta?

Para saber quiénes son nuestros clientes.

Customer's preferred Brand

Quiénes compran qué

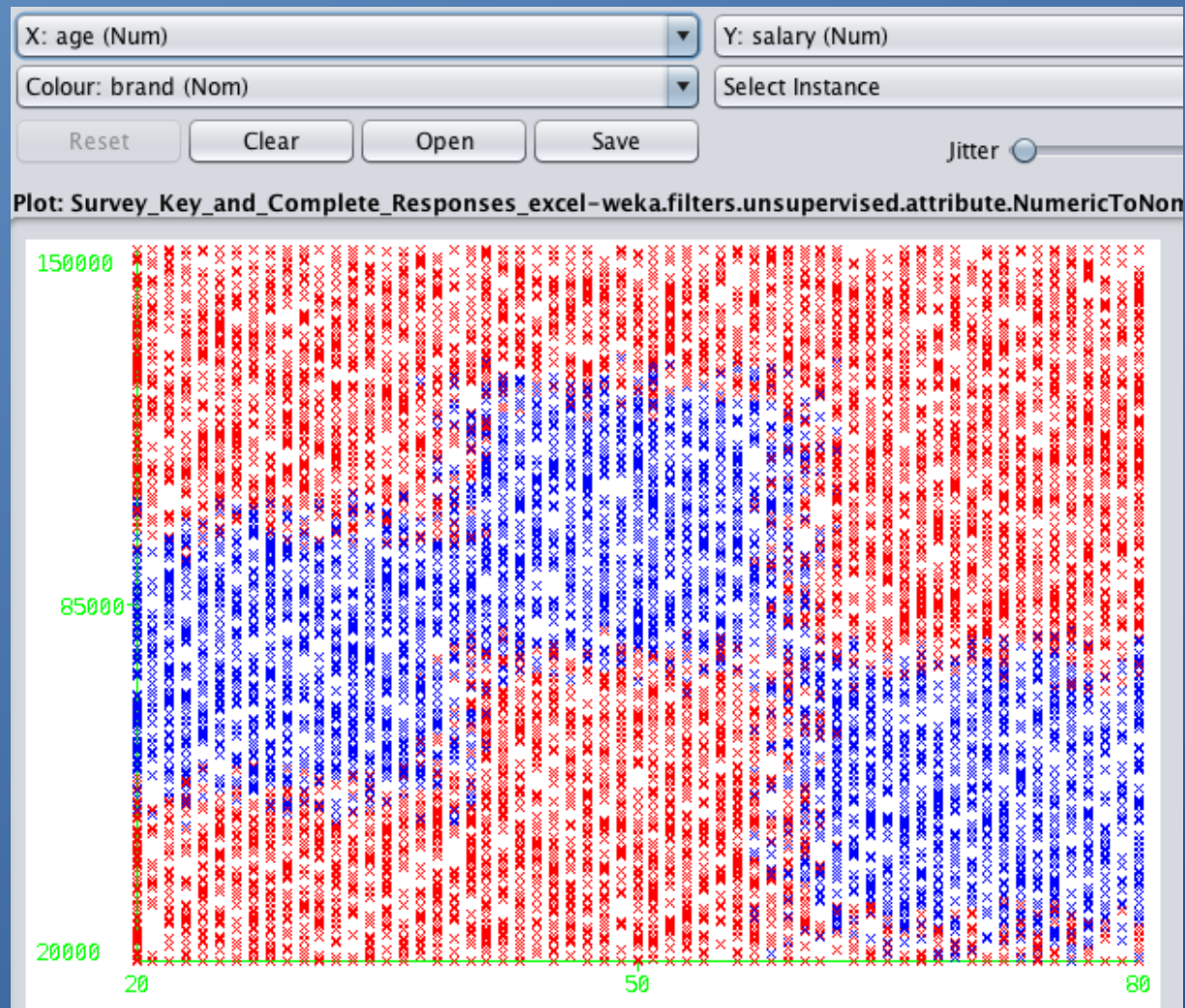
Para poder identificar a nuestros clientes he hecho dos análisis fundamentalmente:

- Uno visual.
- Un árbol de decisión (J48 con WEKA y C50 con R).

Customer's preferred Brand

Quiénes compran qué

Estudio visual en WEKA:



Customer's preferred Brand

Quiénes compran qué

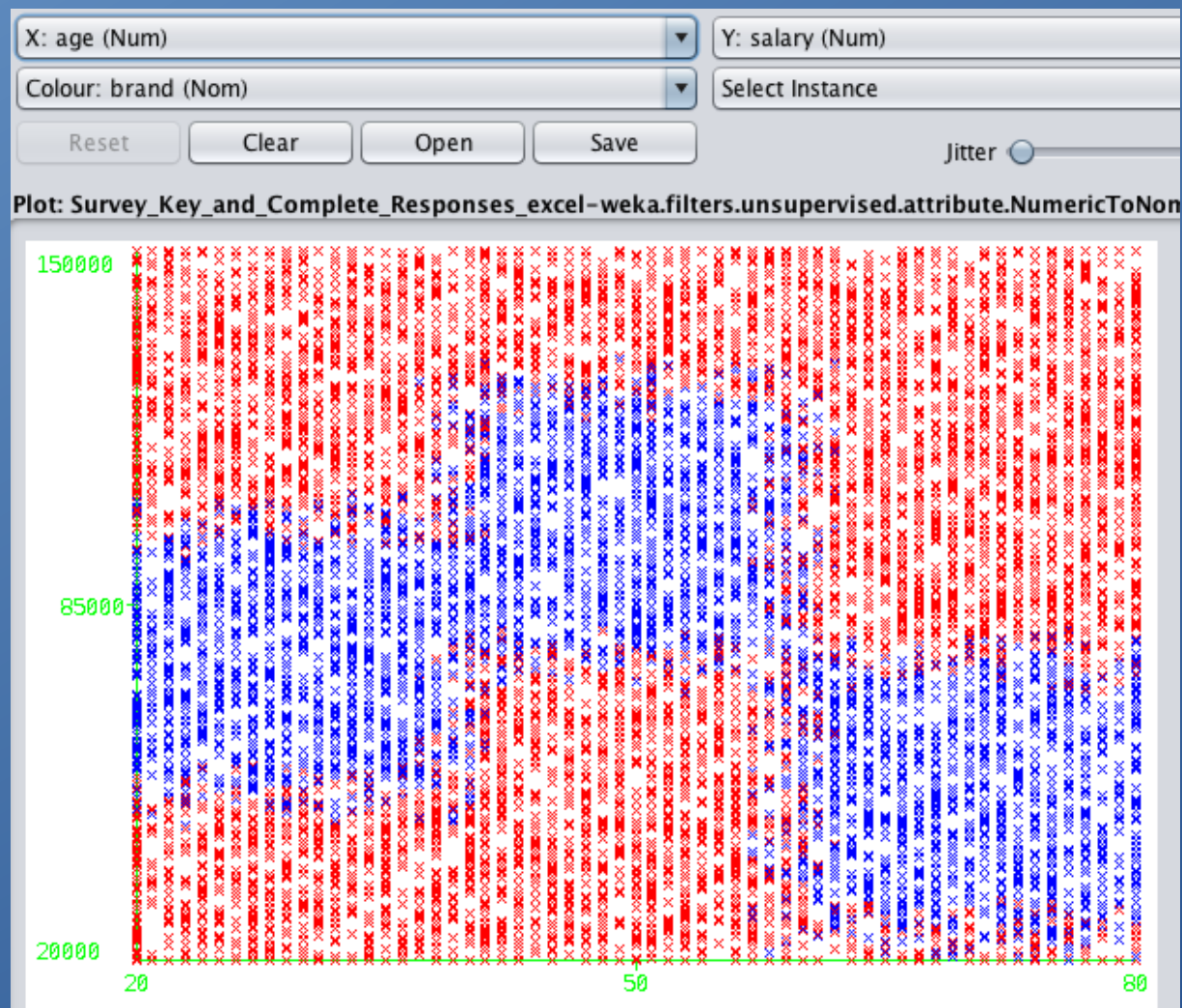
Decision tree J48 en WEKA:

```
salary <= 125058.22
|
| salary <= 45570.48
| |
| | age <= 59: 1 (1340.0/23.0)
| | age > 59
| | |
| | | salary <= 30848.55
| | | |
| | | | salary <= 24972.87: 1 (148.0/39.0)
| | | | salary > 24972.87
| | | | |
| | | | | age <= 61
| | | | | |
| | | | | | salary <= 25660.98: 0 (3.0/1.0)
| | | | | | salary > 25660.98: 1 (11.0)
| | | | | age > 61: 0 (158.0/45.0)
| | | salary > 30848.55: 0 (348.0/14.0)
| salary > 45570.48
| |
| | salary <= 100191.79
| | |
| | | age <= 40
| | | |
| | | | salary <= 51922.38: 1 (160.0/52.0)
| | | | salary > 51922.38: 0 (1308.0/98.0)
| | | age > 40
| | | |
| | | | age <= 59
| | | | |
| | | | | salary <= 76745: 1 (718.0/101.0)
| | | | | salary > 76745: 0 (567.0/40.0)
| | | | age > 59
| | | | |
| | | | | salary <= 76556.47: 0 (778.0/75.0)
| | | | | salary > 76556.47: 1 (646.0/42.0)
| | | salary > 100191.79
| | | |
| | | | age <= 38: 1 (591.0/42.0)
| | | | age > 38
| | | | |
| | | | | age <= 58: 0 (594.0/75.0)
| | | | | age > 58
| | | | | |
| | | | | | age <= 59
| | | | | | |
| | | | | | | salary <= 118339.09: 0 (22.0/4.0)
| | | | | | | salary > 118339.09: 1 (8.0/1.0)
| | | | | age > 59: 1 (666.0/21.0)
| salary > 125058.22: 1 (1934.0/36.0)
```

Customer's preferred Brand

Quiénes compran qué

Juntando ambos:



Cross-selling ofert on Purchase Request

- Distado.
- Disado.
- El cDD.

Cross-selling ofert on Purchase

Resumen

Se puede realizar un análisis de los productos más consumidos por un cliente y ofrecer otros productos para hacer obtener un cross-selling.

Los resultados animan a continuar con análisis de más productos para aumentar los beneficios de la empresa.

Una clasificación mayor de los productos y e incluir estas clasificación mejoraría la probabilidad de acierto. Incluir el precio del producto y el margen obtenido por la empresa aumentaría aún más los beneficios.