



Caso 2 - Telco Co.

Universidad Austral, MEDGC | Cohorte 2020-2021 | Introducción a Data Mining

Grupo 5B

Agosto 24, 2020

INTEGRANTES

FARACH, VERONICA

HEREDIA BAEK, GABRIELA

KINA, HERNAN

LANZA, EZEQUIEL

LAXI, IGNACIO

MANZL, JUAN JOSE

Introducción

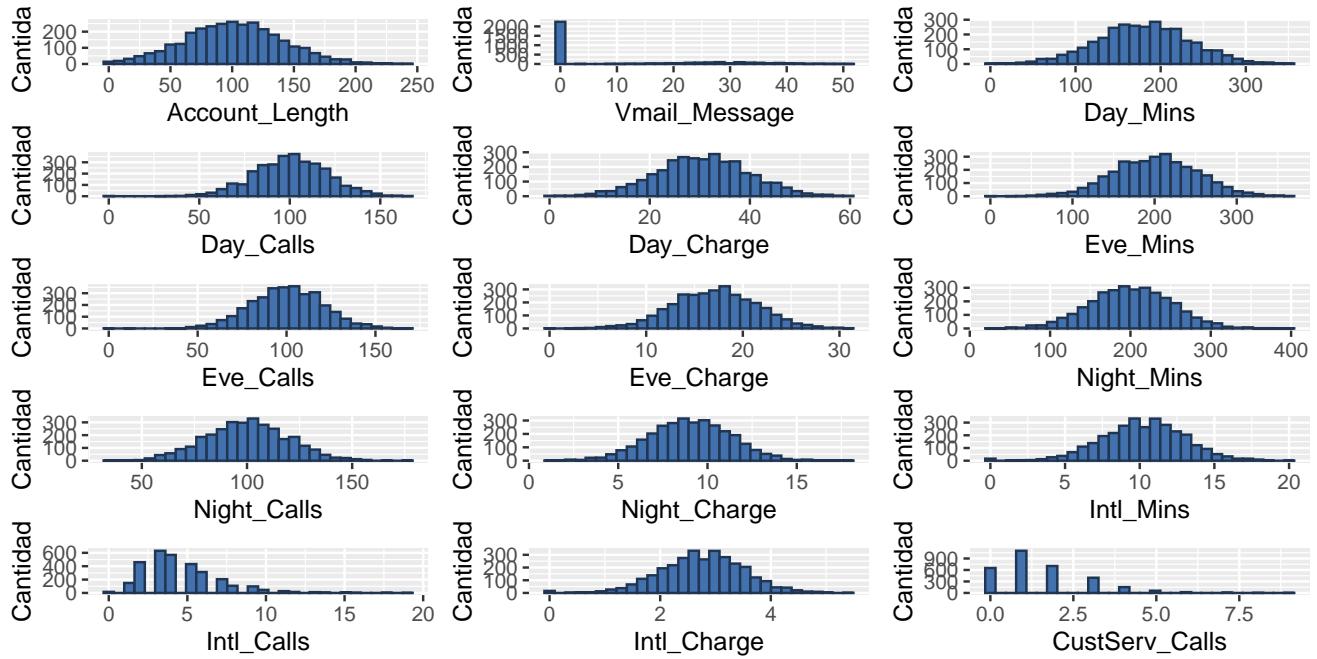


Figure 1: Histograma

Graficos multivariados para variables cualitativas- Observaciones sobre el grafico

1. La proporcion de subscriptores que no tienen Plan de Voice Mail es mayor, en linea con lo observado en los barplots
2. La cantidad de subscriptores que no tienen Plan Internacional es mayor independientemente de si tienen o no Plan de Voicemail.
3. La proporcion de churn es menor entre los subscriptores SIN Voice Mail y SIN Plan Internacional en comparacion con los subscriptores SIN Voice Mail y CON Plan Internacional donde la proporcion de churn y no churn son mas parejas
4. La proporcion de churn es menor entre los subscriptores CON Voice Mail y SIN Plan Internacional en comparacion con los subscriptores CON Voice Mail y CON Plan Internacional donde la proporcion de churn y no churn son mas parejas
5. Por los puntos 3 y 4 las proporciones de churn y no churn son mas similares en subscriptores con Planes internacionales, independientemente de si tienen o no Plan de Voicemail.

De entre las 8 variables correlacionadas se pueden eliminar 4 variables. Las variables seleccionadas para eliminacion son las correspondientes a los cargos: Day_Charge, Eve_Charge, Night_Charge, Intl_Charge.

Creamos un nuevo dataset reducido en 4 columnas: subscriptores.red

Hago una matriz de dispersion indicando en rojo las observaciones que son outliers segun las distnacias de Mahalanobis

Feature Engineering

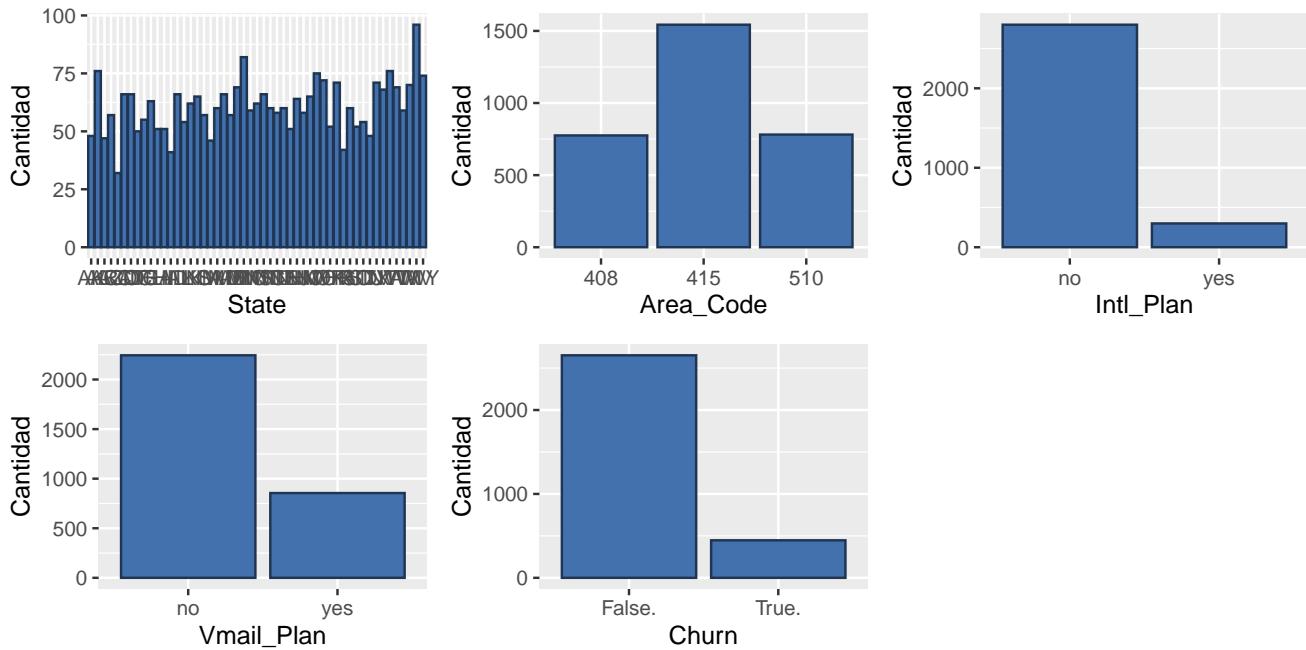


Figure 2: Bar Plots

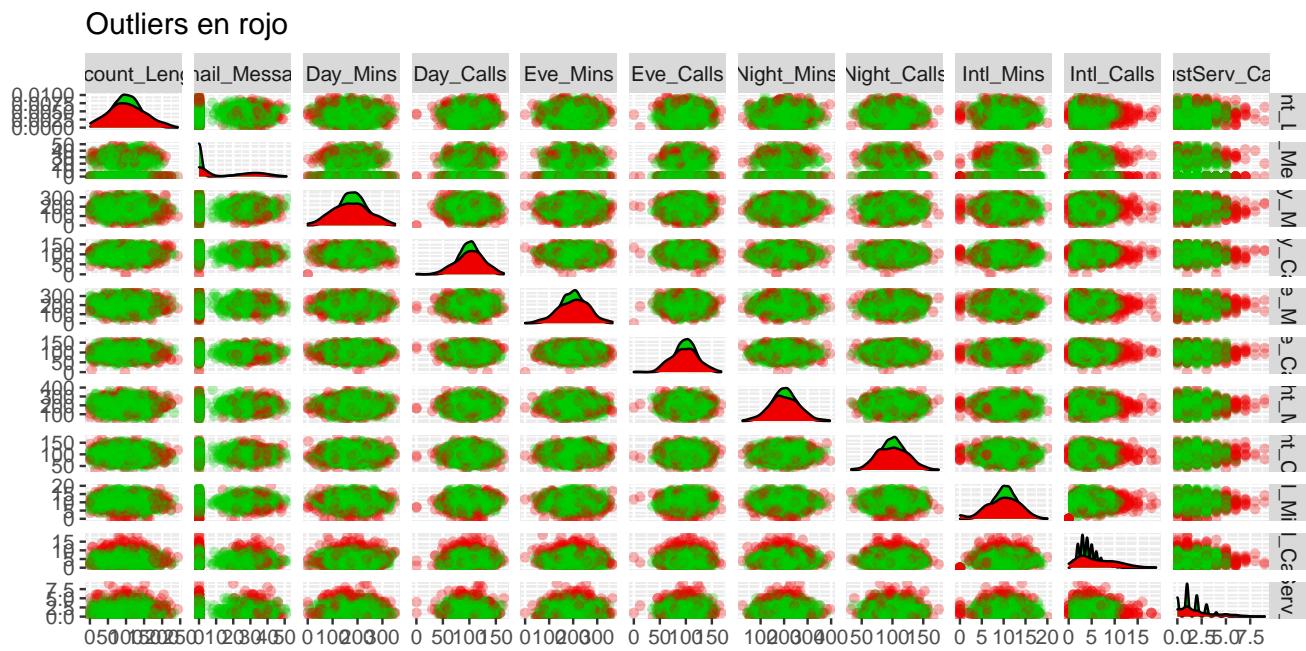


Figure 3: Bar Plots