Un letrero de color negro

Descripción generada automáticamente con confianza baja

***MODELO DE PREDICCION DE PERDIDA DE CLIENTES EN TELECOMUNICACIONES***

*INTEGRANTES*

*Raul Santos Chañe Muñoz*

*Fecha de entrega: 19-02-2023*

***Contenidos***

1. -Descripción del negocio *página 3*
2. -Descripción de los datos *página 4*
3. -Descripción de la metodología y

de los objetivos *página 5*

1. -Consideraciones finales *página 8*

***Descripción del caso de negocio***

Una empresa del rubro de telecomunicaciones que tiene como clientes a la población, presenta un perdida de clientes recientemente,

***Descripción de los datos***

Nuestro dataset contiene 7043 Filas y 21 columnas.

|  |  |
| --- | --- |
| Columna | Descripción |
| customerID | Contiene la abreviación del cliente al que pertenecen. |
| gender | Genero del cliente |
| SeniorCitizen | Persona jubilada |
| Partner | Pareja |
| Dependents | Contiene un 1 en aquellas filas donde la startup tuvo éxito y un 0 en las que han cerrado. |
| tenure | Tipo de tendencia |
| PhoneService | Servicio de Telefono |
| InternetService | Servicio de internet |
| OnlineSecurity | Si el plan contiene seguridad. |
| OnlineBackup | Si el plan contiene respaldo |
| DeviceProtection | Protección en el dispositivo |
| TechSupport | Mantenimiento con garantia |
| StreamingTV | Si el plan contiene TV en vivo |
| StreamingMovies | Si el plan contiene Peliculas en vivo |
| Contract | Tipo de contrato si es mes a mes o anual |
| PaperlessBilling | Facturacion |
| PaymentMethod | Método con el cual se realiza el pago |
| MonthlyCharges | Recargas realizadas en el mes |
| TotalCharges | Cantidad de recargas que se realizaron |
| Churn | Personas que dejaron la operadora |
|  |  |

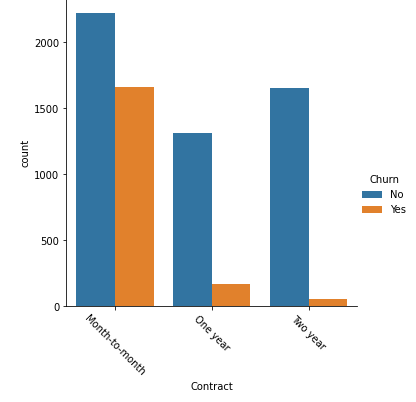
Durante el Data Wrangling se realizo limpieza del dataset, Reemplazando los valores nulos con los promedios de la columna a la cual pertenecen.

***Descripción de la metodología y los objetivos***

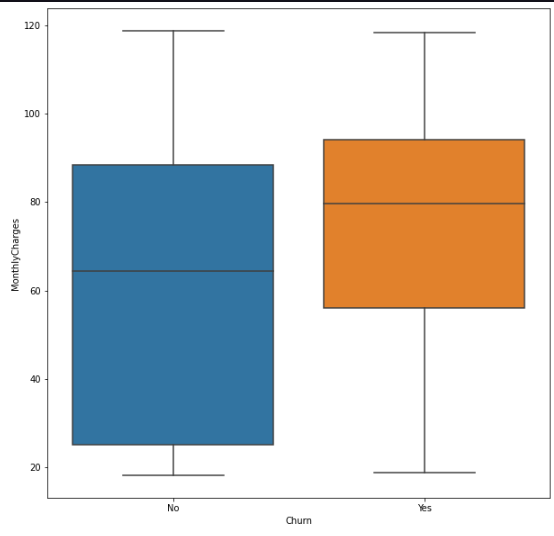
* ***Primera fase: EDA (Exploratory Data Analysis)***

Al inicio del proyecto, lo primero que nos propusimos fue intentar comprender el estado de los datos: aquí entran el equilibrio de categorías, cantidad de nulos o datos sin sentido, varianza, etc. Una vez que eso estuvo cubierto, la siguiente meta fue intentar encontrar algunas particularidades de nuestra temática elegida.

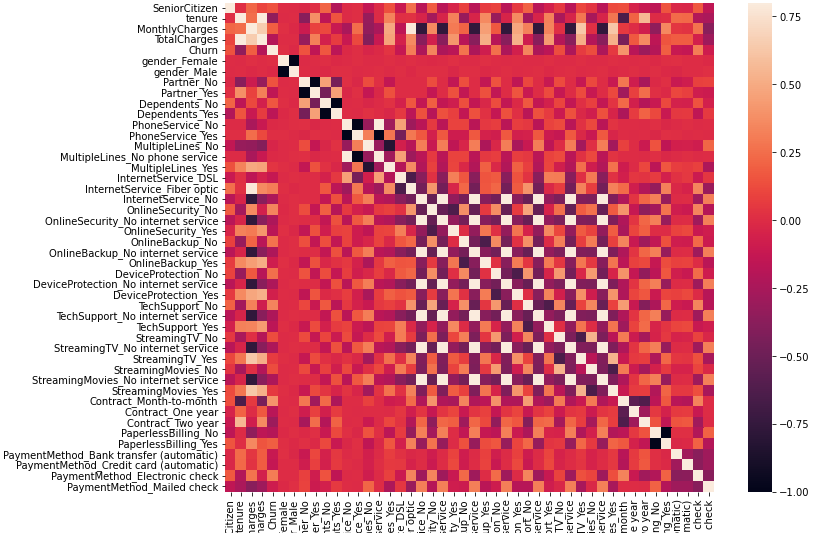
El primer movimiento que realizo es ver como nuestros datos se comportan a partir de las graficas de un valor dependiente e independiente.



Por otro lado, cuando comenzamos a llevar a cabo el análisis univariado descubrimos, tras realizar boxplots, Existe mucha relación entre recargas que se realiza en cada mes y la salida de clientes de la operadora. Ya que los clientes que cumplen con el plan de mes-mes como se muestra arriba tienen mayor índice de retirarse de la operadora.



En el análisis multivariado se busca una correlación alta dentro de nuestro dataset ya tratado.



* ***Segunda fase: búsqueda y aplicación del modelo***

*Random Forest*

Si evaluamos su desempeño mediante los aciertos, este algoritmo presentó un desempeño de 79%.

*Regresión lineal*

Por último, la regresión lineal nos da un desempeño del 80%.

*Modelo elegido*

Se escogio la regresión por le mejor manejo del accuracy, muy aparte que es mucho mas comprensible de análisis y realizar pruebas de desempeño a dicho modelo.

***Consideraciones finales***

Se pudo determinar apartir de la EDA que las columnas o datos que influyen mucho en las perdidas de clientes en el área de telecomunicaciones es tipo de contrato que se realiza. De la misma manera en un análisis multivariado los contratos de mes a mes frente a otras columnas y sus respectivos valores influyen mucho en la salida de clientes en la operadora.

En adición, si se podría agregar a la base de datos los lugares de las personas que desertar de la operadora asi mismo la fecha en la que se realiza las perdidas de clientes. Esto podría ayudar en mucho el análisis.