



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

PROYECTO GRUPAL 1

PREDICCIÓN DE FUGA DE CLIENTE EN INDUSTRIA DE TELECOMUNICACIONES

INTEGRANTES

SEBASTIÁN MICHEA
CONSTANZA PARRA
BERNARDO RIFFO
GONZALO TAMBURINI

04 OCTUBRE 2023



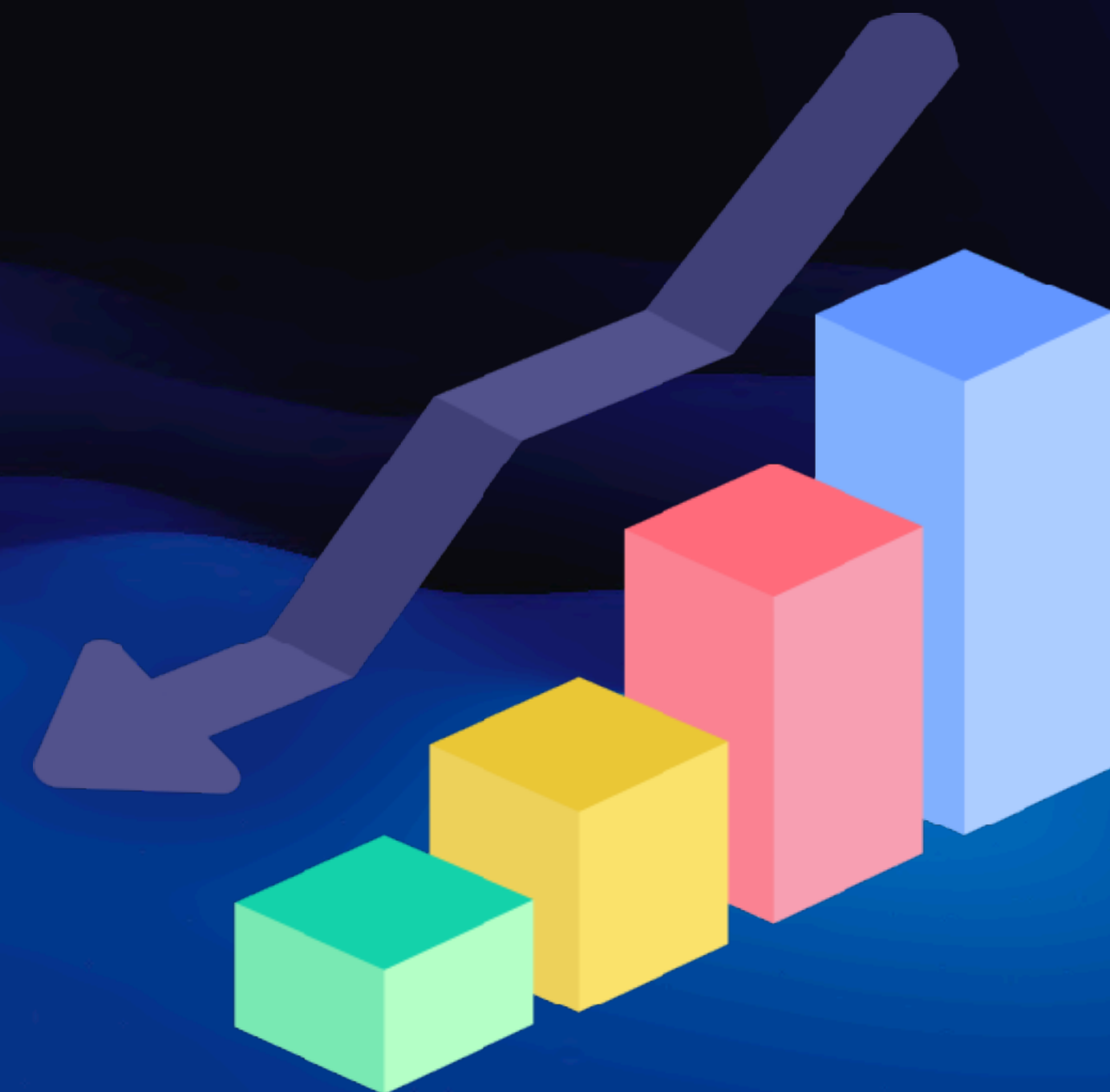
OBJETIVOS

OBJETIVO GENERAL

DESARROLLAR UN MODELO PREDICTIVO DE FUGA DE CLIENTES EN LA INDUSTRIA DE LAS TELECOMUNICACIONES MEDIANTE EL USO DE ALGORITMOS DE APRENDIZAJE AUTOMÁTICO Y PROPONER ESTRATEGIAS DE RETENCIÓN BASADAS EN LOS RESULTADOS OBTENIDOS.

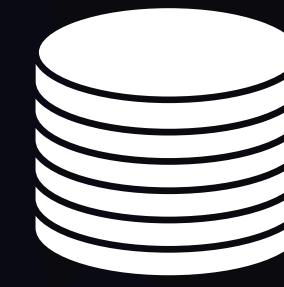
OBJETIVOS ESPECÍFICOS

- 1.-REALIZAR UN ANÁLISIS EXPLORATORIO DE DATOS EXHAUSTIVO SOBRE EL CONJUNTO DE DATOS DE CLIENTES.
- 2.-LLEVAR A CABO LA LIMPIEZA Y PREPARACIÓN DE DATOS PARA ASEGURAR LA CALIDAD Y COHERENCIA DE LOS DATOS UTILIZADOS EN EL MODELADO
- 3.-ENTRENAR MODELO DE ÁRBOL DE DECISIÓN, SVM (SUPPORT VECTOR MACHINE) Y FOREST RANDOM PARA PREDECIR LA FUGA DE CLIENTES.
- 4.-UTILIZAR MÉTRICAS DE EVALUACIÓN COMO ACCURACY, SENSIBILIDAD Y ESPECIFICIDAD PARA VALIDAR Y COMPARAR LOS MODELOS.
- 5.-PROPONER CUATRO ACCIONES ESPECÍFICAS DE RETENCIÓN DE CLIENTES BASADAS EN LOS RESULTADOS DEL MODELO CONSIDERANDO EL CONTEXTO DE LA INDUSTRIA DE LAS TELECOMUNICACIONES.





CONOCIENDO LOS DATOS ORIGINALES



DATASET ORIGINAL : 6.590 REGISTROS DE CLIENTES CON INFORMACIÓN DEMOGRÁFICA, DATA HISTÓRICA DEL CLIENTE EN LA COMPAÑÍA, PRODUCTOS CONTRATADOS Y COMPORTAMIENTO.

VARIABLES DEMOGRÁFICAS DATASET

-SEXO, EDAD, ESTADO CIVIL, N° DE DEPENDIENTES.

VARIABLES DE INTERÉS

-FUGADO, CAUSA FUGADOS



DATA HISTÓRICA DE CLIENTE/COMPORTAMIENTO

-PLAN, MESES COMO CLIENTE, CARGO MENSUAL LLAMADA, HISTÓRICO DE COBROS, SERVICIO INTERNET, SERVICIOS ADICIONALES, ENTRE OTROS.





ANÁLISIS EXPLORATORIO, LIMPIEZA E INGENIERÍA DE DATOS

PREPARACIÓN DE LOS DATOS

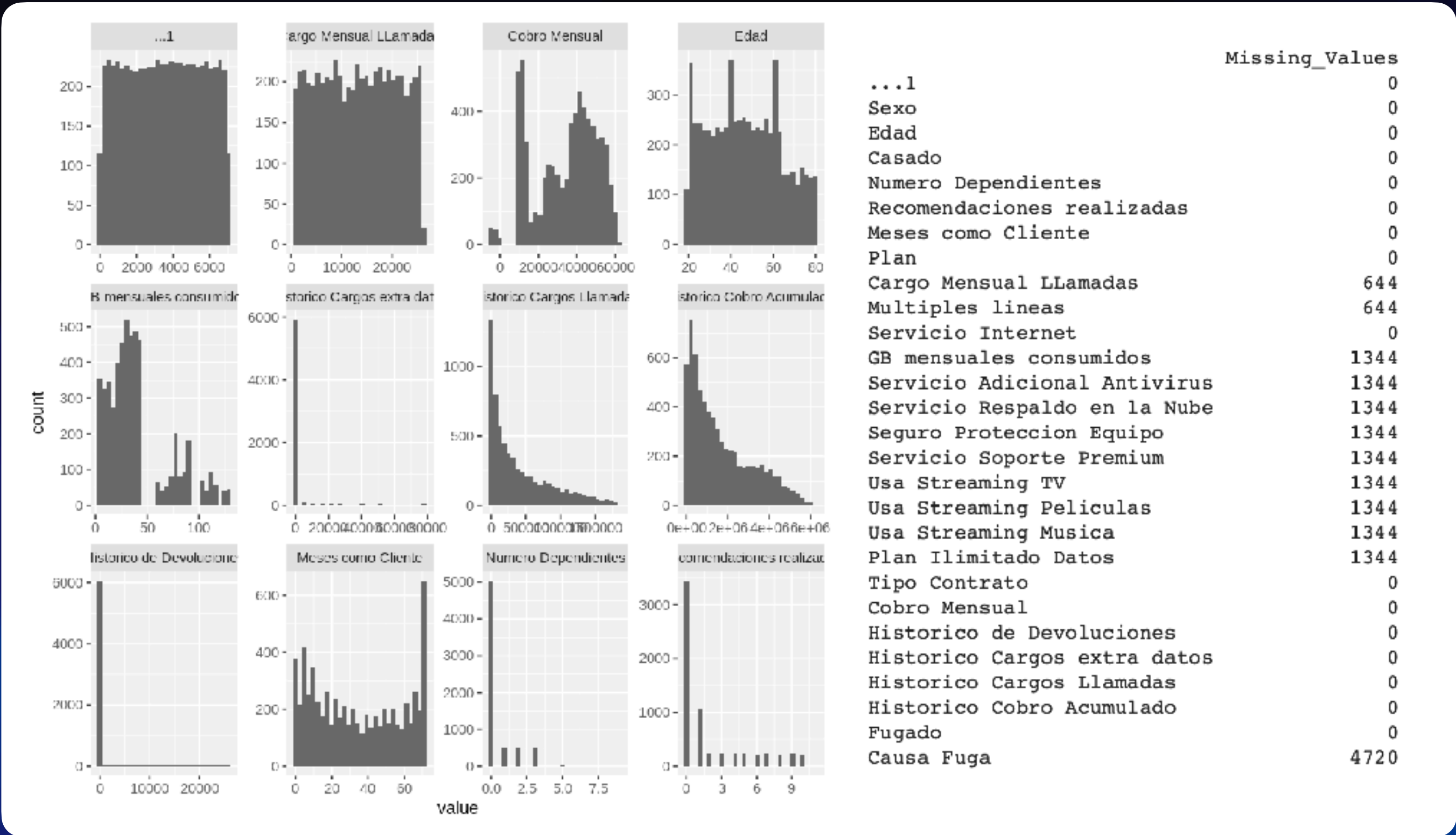
PARA PODER CONSTRUIR EL MODELO, EN PRIMER LUGAR NECESITAMOS REALIZAR UNA EXPLORACIÓN PREVIA DE LA INFORMACIÓN, DE ESTA MANERA CHEQUEAMOS POR POSIBLES DATOS FALTANTES E INCOMPLETOS QUE EXISTAN EN EL DATASET ORIGINAL.

¿CÓMO LO HACEMOS?

A PARTIR DE HISTOGRAMAS Y CONTANDO VALORES FALTANTES, NOTAMOS QUE EL GRUPO DE CLIENTES QUE NO TIENE SERVICIOS INTERNET, POSEE INFORMACIÓN INCOMPLETA, POR LO QUE DECIDIMOS FILTRAR LA INFORMACIÓN Y CONSIDERAR ITERAR MODELOS PARA CLIENTES QUE CUENTEN CON SERVICIO INTERNET.

TRATAMIENTO DE COLUMNAS IRRELEVANTES

A PRIORI, NOTAMOS QUE COLUMNAS COMO “...1” ,”HISTÓRICO DE DEVOLUCIONES” Y “HISTÓRICO CARGOS EXTRA DATOS”, PODRÍAN PERJUDICAR A NUESTROS MODELOS. PRINCIPALMENTE, DADO QUE “...1” ACTÚA COMO UN ID DE CLIENTE, MIENTRAS QUE ”HISTÓRICO DE DEVOLUCIONES” Y “HISTÓRICO CARGOS EXTRA DATOS” SON VARIABLES CON POCA VARIANZA, LO QUE LIMITA LA INFORMACIÓN ENTREGADA POR ESTAS VARIABLES.



LIMPIEZA

- SE FILTRA POR SERVICIO INTERNET = 'SI' , ELIMINANDO VALORES 'NA'
- SE QUITAN COLUMNAS IRRELEVANTES PARA NUESTROS MODELOS



ANÁLISIS EXPLORATORIO, LIMPIEZA E INGENIERÍA DE DATOS

RESULTADO: DATASET LIMPIO

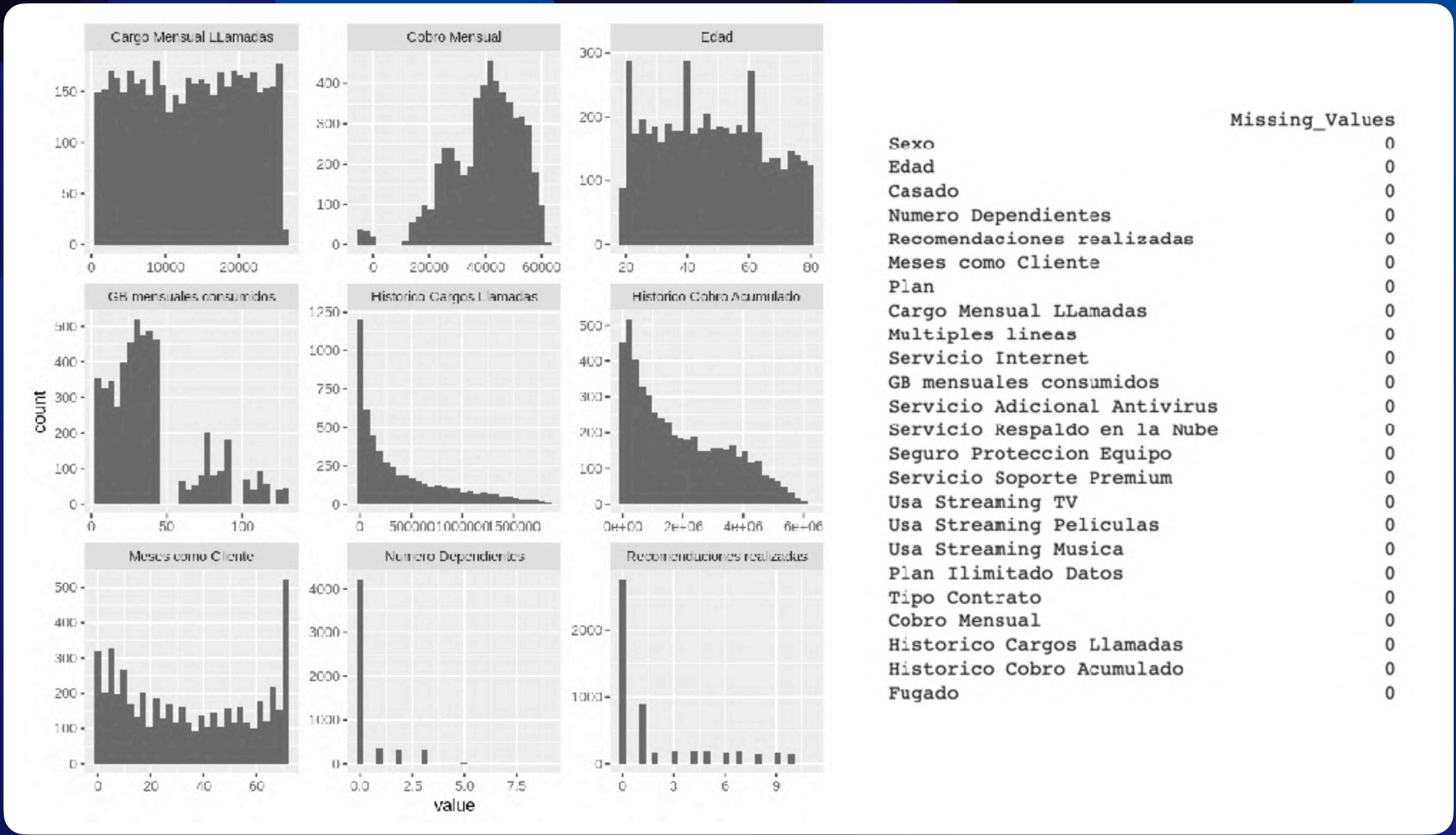
UNA VEZ REALIZADA LA LIMPIEZA, OBSERVAMOS LAS VARIABLES RESULTANTES, DONDE OBSERVAMOS QUE NO TENEMOS VALORES FALTANTES, ESTO, PRODUCTO DEL FILTRO : “SERVICIO INTERNET = SI”. LO QUE NOS REDUJO EL DATASET EN APROXIMADAMENTE 87% DE LOS REGISTROS ORIGINALES, QUEDANDO EN 4601 REGISTROS.

REDUCCIÓN DE FILAS : 6589 -> 4601

TENIENDO ESTO EN CONSIDERACIÓN, GRAFICAMOS LAS VARIABLES CONTINUAS, NOTANDO UN COMPORTAMIENTO COHERENTE Y SIN ANOMALÍAS.

CON ESTE DATASET OPTIMIZADO EVITAMOS ERRORES DE MODELADOS Y A LA VEZ HACEMOS MÁS EFICIENTE EL PROCESO.

AHORA, PODEMOS PROCEDER AL PASO SIGUIENTE: BALANCEO DE CLASES, ELECCIÓN DE PROPORCIÓN DE ENTRENAMIENTO/ VALIDACIÓN Y ELECCIÓN DEL MODELO A UTILIZAR.





BALANCEO, PROPORCIÓN Y MODELAMIENTO

BALANCEAR LAS CLASES

DADA LA NATURALEZA DE LA INFORMACIÓN, NOTAMOS QUE LA CLASE MAYORITARIA ERAN LOS "NO FUGADOS" POR LO QUE PROBAMOS DISTINTOS BALANCEOS DE CLASES PARA DETERMINAR EL MEJOR BALANCE RATIO. ITERAMOS CON RATIOS DE 1.0, 1.2 Y 1.5.

PROPORCIÓN DE ENTRENAMIENTO Y VALIDACIÓN


PROBAMOS LOS MODELOS CON DISTINTAS PROPORCIONES DE ENTRENAMIENTO Y VALIDACIÓN (70/30, 80/20, 90/10), OBTENIENDO COMO CONCLUSIÓN QUE LA MEJOR PROPORCIÓN PARA ENTRENAR EL MODELO ES UN 80%, MIENTRAS QUE PARA VALIDACIÓN UN 20%. DE ESTA MANERA CONSIDERAMOS UN VOLUMEN DE DATOS CONSIDERABLE PARA EL ENTRENAMIENTO Y UN GRUPO CONSIDERABLE PARA VALIDACIÓN.

APLICAR MODELOS

ELEGIDA LA PROPORCIÓN DE ENTRENAMIENTO Y VALIDACIÓN (80/20), PROCEDAMOS A ITERAR NUESTROS MODELOS CON LOS BALANCE RATIO DE 1.0, 1.2 Y 1.5. LOS RESULTADOS DE ESTAS ITERACIONES SE VERÁN EN LA SIGUIENTE LÁMINA.

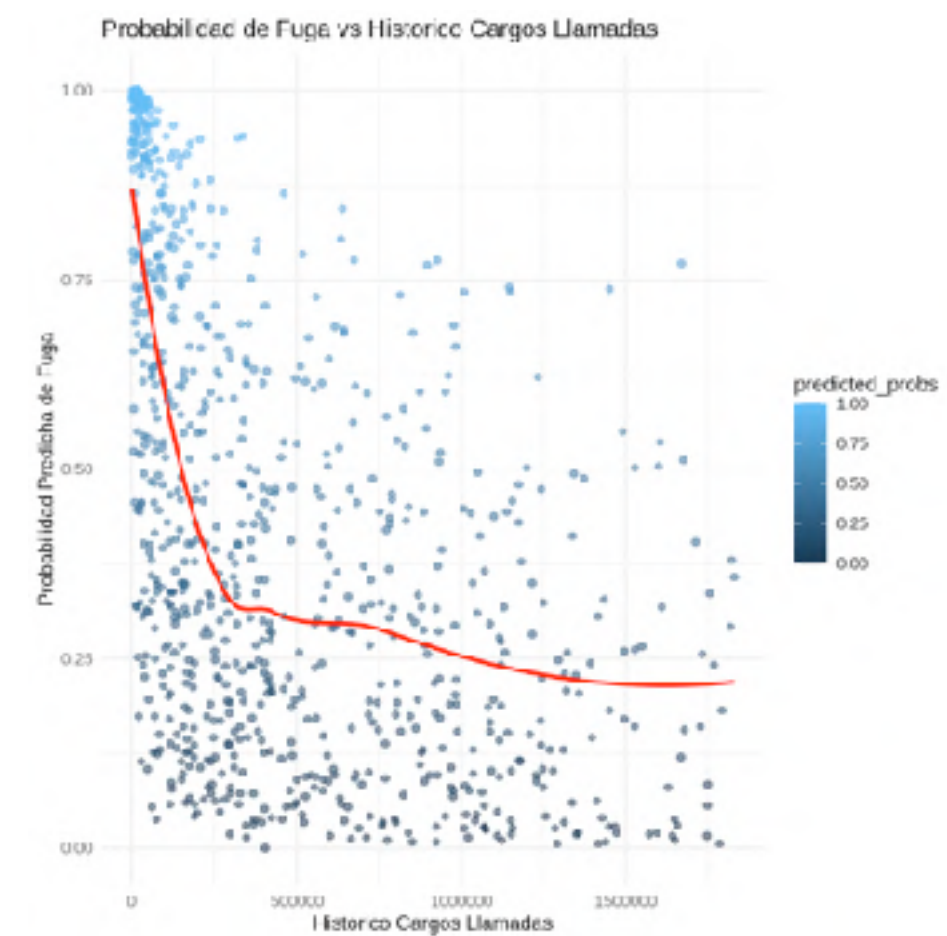
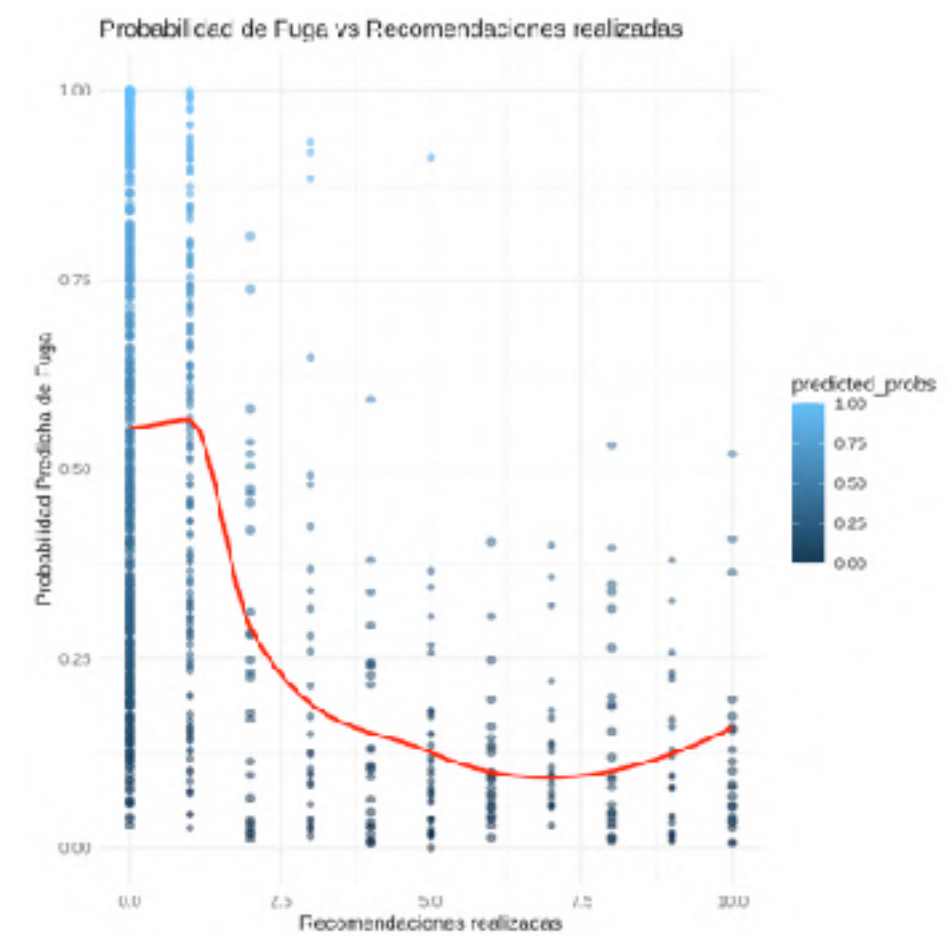
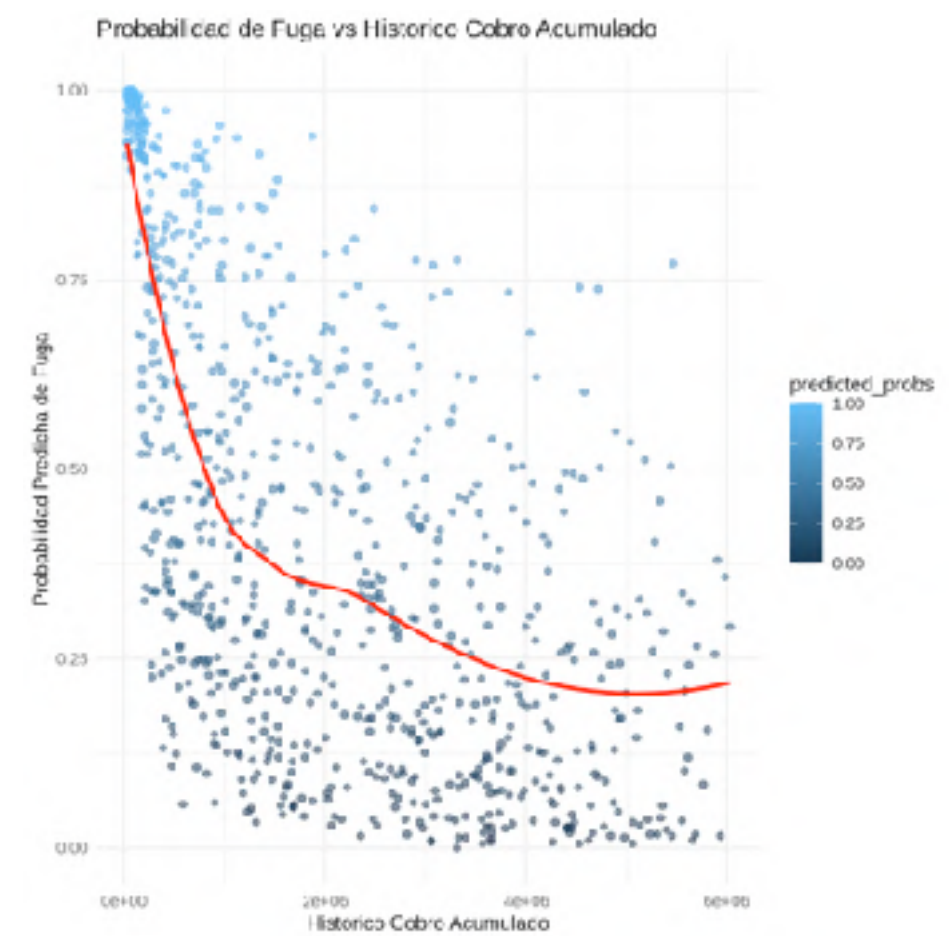
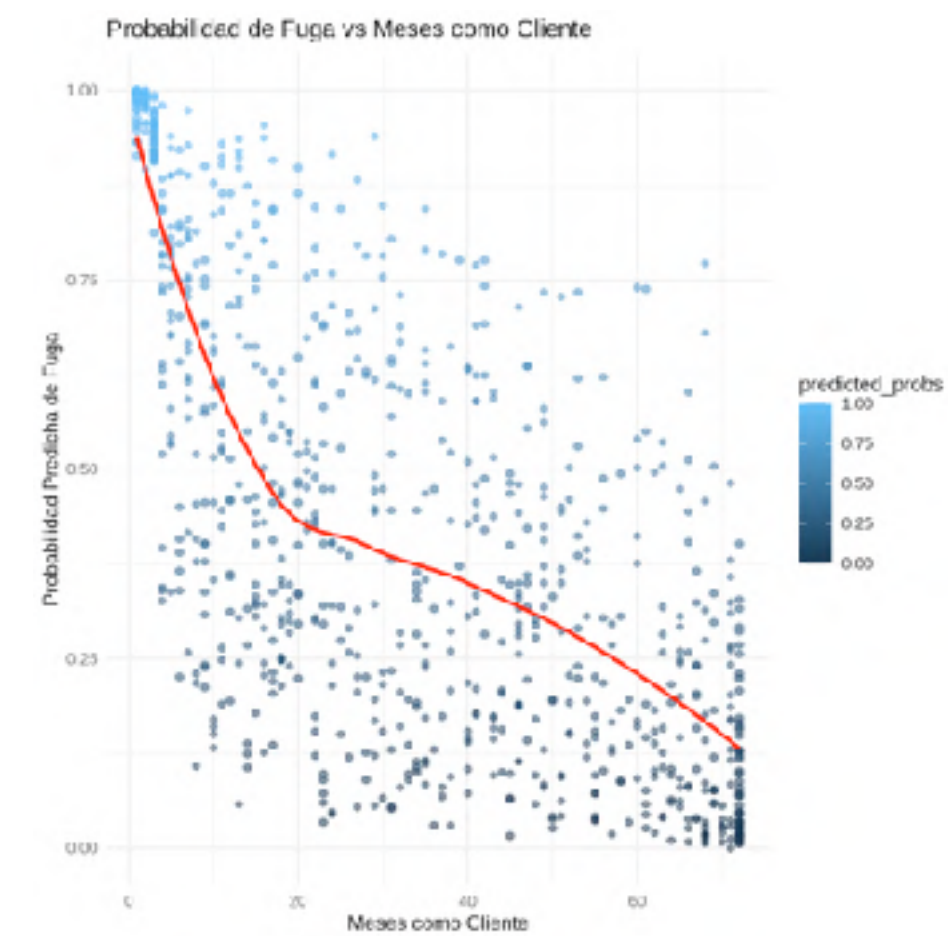
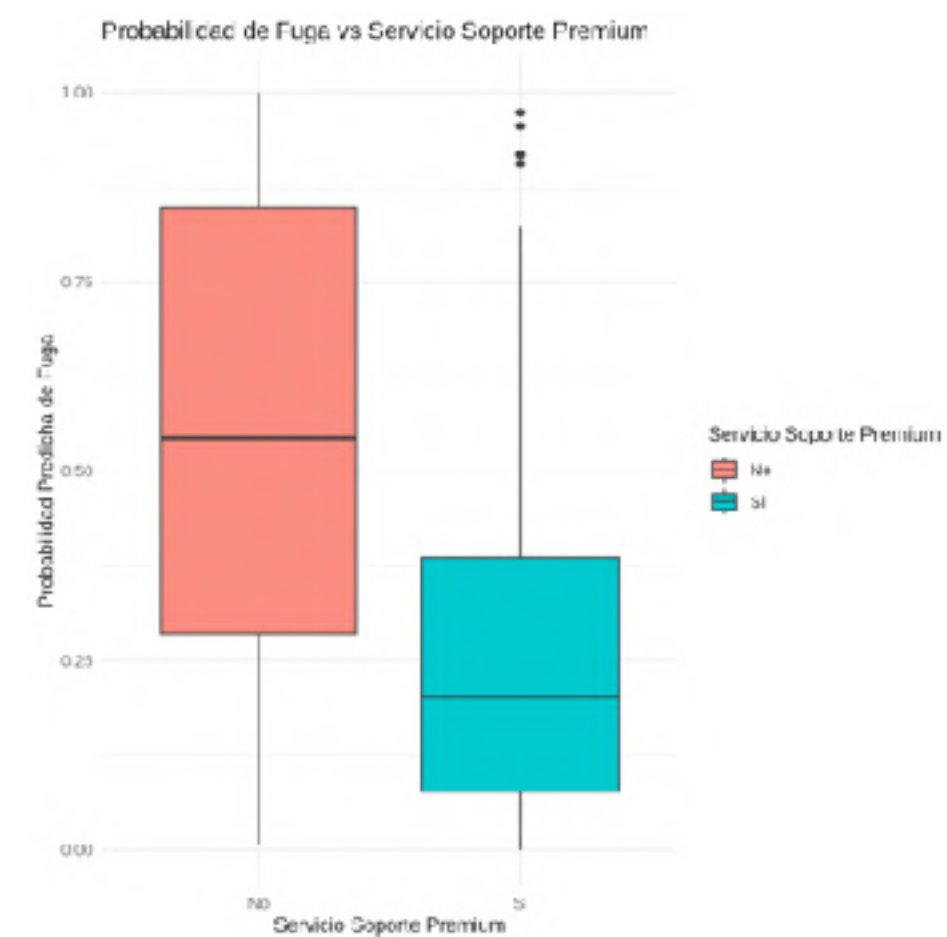
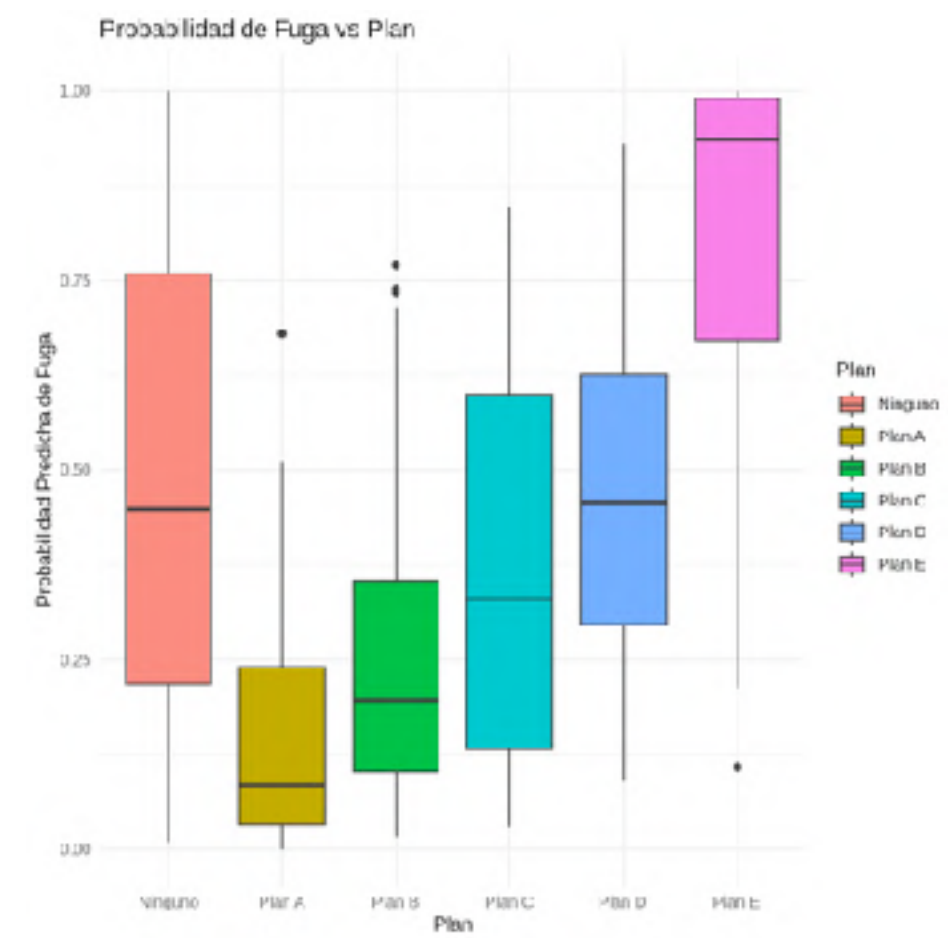
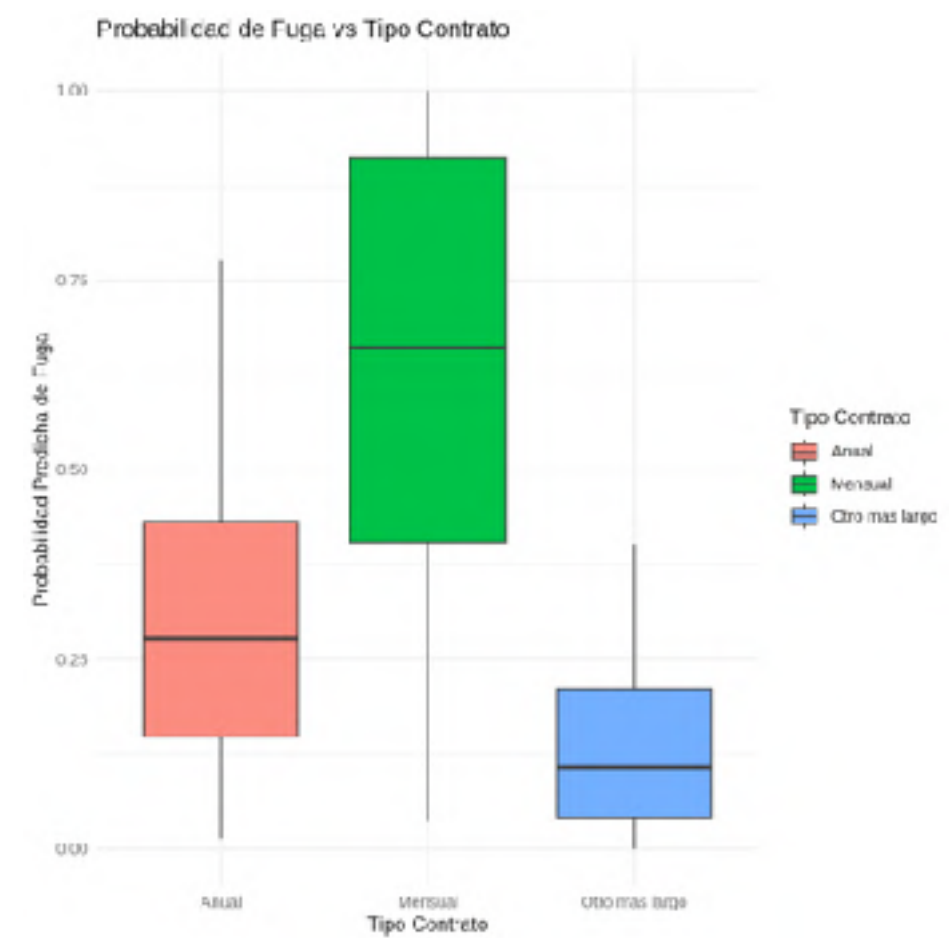
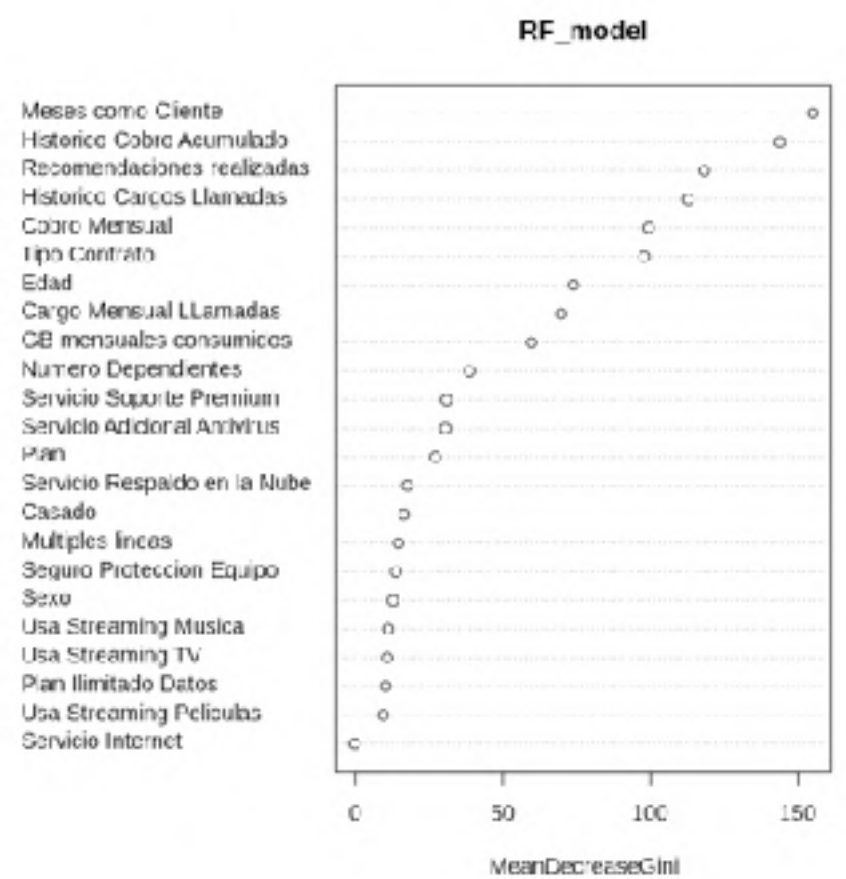


ESTADÍSTICA COMPARATIVA DE MODELOS DE PREDICCIÓN

MODELO	ARBÓL DE DECISIÓN			SVM			RANDOM FOREST			<div> MODELO ELEGIDO</div>
BALANCE RATIO	1,0	1,2	1,5	1,0	1,2	1,5	1,0	1,2	1,5	
ACCUURACY	0,835	0,8219	0,8219	0,8154	0,8284	0,8122	0.8267	0.8338	0.8496	
SENSITIVITY	0,6688	0,6698	0,6698	0,6977	0,7233	0,6904	0.8577	0.8295	0.7654	
SPECIFICITY	0,9197	0,9022	0,9022	0,8784	0,8839	0,8779	0.7963	0.8375	0.9077	



INSIGHTS DE RANDOM FOREST





RECOMENDACIONES



PARA CLIENTES CON PLANES DE PAGO MENSUAL SE SUGIERE OFRECER PLANES ANUALES PARA MEJORAR PROBABILIDAD DE RETENCIÓN.

OFRECER DE MANERA CONTINUA A LOS USUARIOS DEL PLAN TIPO E, CAMBIOS DE CONTRATOS CON PRESTACIONES ATRACTIVAS O DESCUENTOS POR PERIODOS DE TIEMPO

GENERAR CAMPAÑAS DE CONTRATACIÓN DE SERVICIOS PREMIUM ASOCIADOS A LOS PAGOS DEL PLAN YA QUE ESTA VARIABLE POSEE UNA BAJA PROBABILIDAD DE FUGA

DEBIDO A LA TENDENCIA DE MENOR PROBABILIDAD DE FUGA AL AUMENTAR LAS RECOMENDACIONES DE CLIENTES, SE RECOMIENDA IMPLEMENTAR UN PROGRAMA DE REFERIDOS QUE PREMIE A LOS CLIENTES CON DESCUENTOS POR SUS RECOMENDACIONES EN EL SITIO WEB. ADEMÁS, CONSIDERA OFRECER DESCUENTOS ESPECIALES PARA GRUPOS, INCENTIVANDO ASÍ A LOS CLIENTES A RECOMENDAR EL SERVICIO A AMIGOS Y FAMILIARES.

BONUS: CONSIDERANDO LA VARIABLE CAUSA FUGA, DA CUENTA DE QUE LA MODA OBSERVADA CORRESPONDE A "MEJOR COMPETENCIA", POR ESTA RAZÓN SE PROPONE UNA BÚSQUEDA ACTIVA DE ESTOS POTENCIALES CLIENTES PARA OFRECER MEJORAS EN SUS PLANES ACTUALES Y REVERTIR SU PROBABILIDAD DE FUGA.