**0. Abstract**

Este estudio, en colaboración con el Máster de formación permanente en Periodoncia y Periimplantologia de la Universidad de Barcelona, se centró en mejorar la eficiencia de los procedimientos de cirugía dentoalveolar, como las extracciones dentales y la colocación de implantes, enfrentando desafíos como la periimplantitis. Los objetivos principales fueron optimizar la gestión de la duración y dificultad de las intervenciones quirúrgicas y mejorar la asignación de operaciones a los estudiantes según su nivel de experiencia.

Los resultados identificaron variables críticas que influyen en la duración y complicaciones de las cirugías, permitiendo la implementación de dos modelos predictivos en una aplicación Streamlit para uso clínico.

Estos modelos no sólo predicen el tiempo necesario para cada procedimiento, sino que también clasifican las intervenciones por dificultad, mejorando la planificación y formación tanto de pacientes como de estudiantes. El estudio demuestra que la ciencia de datos puede optimizar la gestión de recursos y la atención al paciente en entornos de aprendizaje clínico, contribuyendo significativamente a la práctica odontológica.

1. **Introducción:**
   1. Context and need: el contexto junto a las motivaciones para nuestro proyecto (lo de las enfermedades dentales, puede que explicar por qué no hemos podido hacer el objetivo de predecir la enfermedad).

Este proyecto odontológico surge de la colaboración con el Máster de formación permanente en Periodoncia y Periimplantologia de la Universidad de Barcelona, donde los estudiantes llevan a cabo procedimientos de cirugía dentoalveolar bajo la supervisión de profesionales. La cirugía dentoalveolar es una subdisciplina de la cirugía oral y maxilofacial que se concentra en intervenciones relacionadas con los dientes y el hueso alveolar en la cavidad bucal, como extracciones dentales y colocación de implantes.

A pesar de los avances en la técnica quirúrgica y los materiales utilizados en la colocación de implantes dentales, aún persiste la posibilidad de complicaciones, como la periimplantitis. Esta condición inflamatoria puede llevar a la pérdida del implante y del tejido circundante, lo que impacta negativamente en la calidad de vida del paciente y puede requerir intervenciones adicionales.

Además de abordar los problemas de salud de los pacientes, los encargados del máster buscan optimizar la gestión de sus procedimientos para mejorar la calidad de atención y aprovechar mejor los recursos de la clínica. Esto permite evitar desperdicios de tiempo y recursos facilitando acoger a más estudiantes en el máster, con el objetivo de alcanzar el número óptimo de alumnos para una buena formación y mejorar así la calidad asistencial de los pacientes.

* 1. Task and objective: es la contribución científica y los objetivos cubiertos (introducción general con ambos objetivos).

Como se ha mencionado previamente, el director del máster, el doctor Rui Figuereido, ha mostrado su interés en mejorar su funcionamiento. Por ello, el objetivo principal es optimizar la gestión de las consultas clínicas a través de una aplicación Streamlit, mediante el análisis de la duración y las posibles complicaciones de las operaciones, con el fin de administrar de manera más eficiente los recursos utilizados.

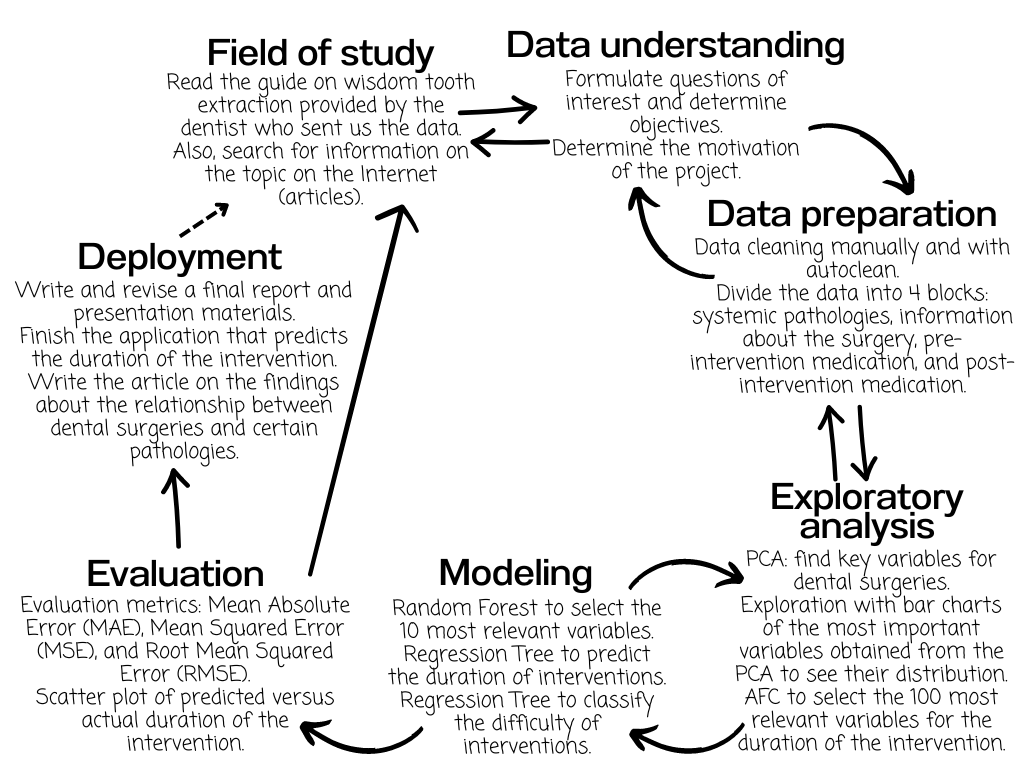
Para alcanzar el objetivo de la manera más precisa, se han llevado a cabo las siguientes subtareas:

1. **Identificar perfiles de pacientes con mayor riesgo**. Con este objetivo, buscamos determinar el perfil de los pacientes con mayor riesgo de experimentar complicaciones postoperatorias. Esto nos permitirá identificar qué tipos de procedimientos pueden ser más delicados y requerir un manejo especializado.
2. **Predecir la duración de la intervención** para cada paciente. Se desarrollarán modelos predictivos basados en datos para estimar la duración de la intervención quirúrgica, la cual tiene una gran relación con la dificultad de esta. Esto implica analizar factores relacionados tanto con las operaciones, como con la salud bucodental de los pacientes.
3. **Asignar operaciones a los distintos alumnos según su experiencia** y según el nivel de dificultad operatoria. Finalmente, al gestionar las operaciones según su duración y dificultad, se asignan a los alumnos según el nivel de estos y de manera óptima para maximizar los recursos de la clínica. Esto podría permitir, en un futuro, recibir a más alumnos en la clínica, logrando que la gestión de las camillas disponibles sea cubierta de forma óptima.

Como se ha señalado previamente, esta información estará disponible para los estudiantes y empleados de la clínica a través de una aplicación. Simplemente deberán ingresar los datos cruciales de la operación y del paciente, y la aplicación proporcionará automáticamente toda la información solicitada.

1. **Materials and methods:** todo lo que se ha usado para analizar y los modelos estadísticos.
   1. **Methodology used**

Respecto a la metodología usada para llevar a cabo el trabajo, se usó CRISP-DM, ya que se trata de una de las metodologías que más se ajustan a los pasos típicos de un proyecto de ciencia de datos y es una de las más populares en el ámbito de la industria. Además, se hicieron unas adiciones a este ciclo típico para ajustarlo más al contexto de este proyecto. Los pasos seguidos son los mostrados en la Figura 1.



* 1. **Technology used** (retos q hemos tenido q afrontar como el autoclean)

Durante el desarrollo de este proyecto, se enfrentaron diversos desafíos tecnológicos debido a la complejidad de la base de datos. La selección de herramientas y lenguajes específicos respondió tanto a las necesidades del proyecto como a la familiaridad con las tecnologías disponibles, garantizando eficiencia y efectividad en cada fase.

En primer lugar, para la planificación del proyecto, se utilizó The Gantt Project. Esta herramienta fue elegida por ser uno de los planificadores más conocidos y fáciles de entender. Además, la aplicación ofrece plantillas predefinidas, lo que simplifica la gestión de tiempos y tareas, facilitando así la organización y seguimiento del proyecto.

Para la limpieza, preparación y exploración de los datos, se recurrió al lenguaje R. Este lenguaje está especializado en estadística y análisis de datos, lo que lo hace ideal para estas tareas. R cuenta con una gran cantidad de paquetes y bibliotecas que se utilizaron, como `dplyr`, `factoextra` y `ggplot2`, permitiendo realizar análisis y visualizaciones de datos de manera eficiente.

En cuanto al modelado, se optó por Python debido a su versatilidad y familiaridad. Python es extremadamente útil para una amplia gama de aplicaciones, desde desarrollo web hasta análisis de datos y aprendizaje automático. Se usaron bibliotecas poderosas y ampliamente utilizadas en estos campos, como `sklearn` (DecisionTreeRegressor, mean\_absolute\_error y mean\_squared\_error) y `scipy.stats`. Para el desarrollo y la ejecución de los modelos en Python, se empleó Google Colab, ya que proporciona recursos computacionales gratuitos, incluidos GPUs, lo cual es especialmente beneficioso dada la dimensión de la base de datos.

Finalmente, para convertir el trabajo en un producto funcional, se exploró y comenzó a utilizar Streamlit. Esta es una biblioteca de Python de código abierto que permite crear aplicaciones web interactivas y personalizadas para proyectos de ciencia de datos y machine learning. La principal ventaja de Streamlit es su facilidad de uso y enfoque en la simplicidad. Permite a los científicos de datos y desarrolladores crear interfaces de usuario interactivas de manera rápida y sencilla, utilizando solo unas pocas líneas de código Python. Con Streamlit, se pudieron integrar visualizaciones, controles de usuario y modelos de machine learning en una web de datos de forma intuitiva.

La elección de estas herramientas y lenguajes estuvo orientada a maximizar la eficiencia y efectividad en cada fase del proyecto, asegurando así la calidad y funcionalidad del producto final.

* 1. **Data**

Respecto a los datos utilizados, estos fueron proporcionados por el responsable de una clínica dental en Barcelona, Rui Figueiredo. Los datos consisten en tres conjuntos de datos: uno de 2016, otro de 2017 y el último de 2018. Todos ellos fueron el resultado de una encuesta realizada entre los dentistas de la clínica en Barcelona, y cada conjunto de datos contiene 936 variables.

Dado que los conjuntos de datos tienen las mismas variables y no hay continuidad entre ellos, es decir, no hay pacientes repetidos, se ha decidido considerar la totalidad de ellos como datos transversales y concatenarlos todos por filas. Esto resultó en un conjunto de datos más grande con 2106 observaciones, más de 930 variables y sin distinción de tiempo, ya que no era útil.

Además, como la cantidad de variables es demasiado alta, la gran mayoría son binarias y se puede identificar una distinción temática entre ellas, se han dividido en 4 bloques teniendo en cuenta qué se desea hacer con ellas y lo que el objetivo principal requiere. Estos bloques son: las patologías sistémicas de los pacientes, que contiene características como el género, la fecha de nacimiento, el número de cigarrillos por día si son fumadores, si consumen alcohol o no, niveles de colesterol, entre otros; los detalles de la intervención, que contiene variables como su duración, herramientas utilizadas y sus características, tipo de intervención y mucho más; la medicación pre-intervención, que son los medicamentos que los pacientes tomaban antes de someterse a la intervención; y, finalmente, la medicación post-intervención.

En términos de protección de datos, el acceso inicial estará restringido exclusivamente a los investigadores clínicos. Los datos personales utilizados en el estudio deben ser salvaguardados de conformidad con las disposiciones del Reglamento General de Protección de Datos (UE) 2016/679 y la Ley Orgánica 3/2018 de Protección de Datos Personales y garantía de los derechos digitales, que ajusta la legislación española al Reglamento General de Protección de Datos de la Unión Europea. El nombre del paciente no debe ser incluido en ningún documento del estudio, únicamente se le asignará un código de paciente al inicio del mismo. La divulgación de información se realizará exclusivamente en forma agregada y en ningún caso se revelará la identidad de los participantes en ninguna publicación o presentación en congresos derivada de los resultados de este estudio.

* 1. **Data preparation**

Inicialmente, como la primera subtarea se centra en el análisis de los pacientes, se han seleccionado las 111 primeras columnas de la base de datos, las cuales contienen las características de los pacientes, así como sus patologías y medicamentos.

A continuación, con el propósito de generar gráficos de barras que representen la distribución de las variables de interés en la base de datos, hemos procedido a transformar dichas variables en categóricas, agrupando las características de acuerdo al siguiente criterio.

Primero se ha reducido las variables binarias para cada tipo de patología a una columna llamada ‘Patologia\_1stemica’ que indicará que tipo de patología tiene el paciente. De la misma forma, se agrupan las distintas medicaciones en la variable ‘Medicacion\_Actual2’. A continuación, para recopilar la edad de los pacientes se han creado rangos de 10 en 10 años en la variable ‘Edad’ a partir de la fecha de nacimiento; y se agrupa a los fumadores en la variable ‘nivel\_de\_fumador’.

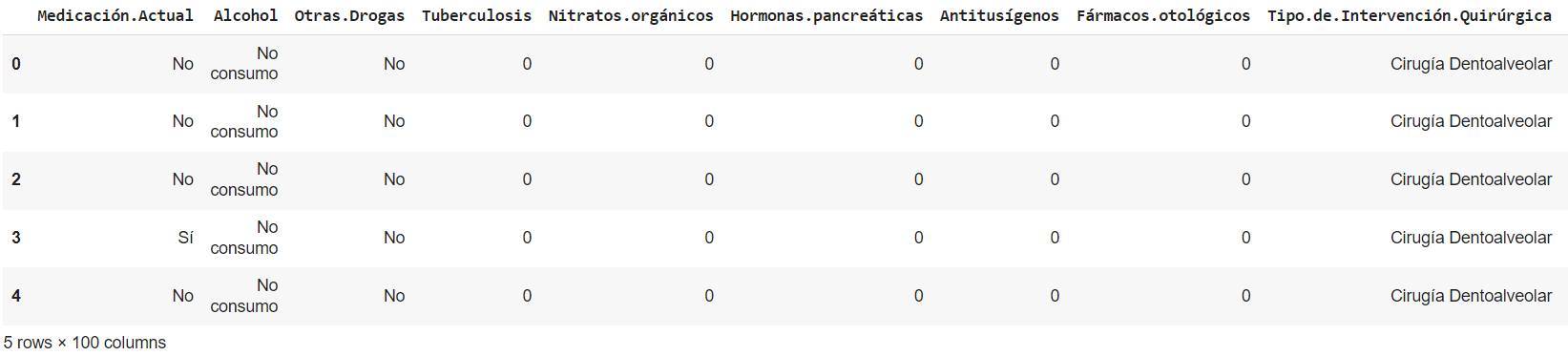
Finalmente, se eliminan las variables que no interesan para este análisis (-Fecha\_Nacimiento, -Date\_Create, -IP\_Adress, -Operador, -Auxiliar, -Jefe\_de\_dia, -Fecha\_intervencion, -Fecha\_Nacimiento, -No.fumador.a, -Exfumador.a, -Fumador.a, -Otras\_Drogas, -Medicacion\_actual, -Otro..especifique., -edad, -Tipo\_De\_cirugia, -Otro..especifique.\_17); y se modifican las variables ‘Patologia\_1stemica’ y ‘Medicacion\_Actual2’ agrupando los valores que aparecen menos de 5 veces en la categoría ‘otros’. A continuación se explican las variables resultantes:

* Género: Agrupa las diferentes categorías de género en una sola columna.
* Patología Sistémica: Combina todas las patologías sistémicas en una única categoría.
* Alcohol: Agrupa los diferentes tipos de consumo de alcohol en una sola columna.
* Medicación Actual: Combina todas las medicaciones actuales en una única categoría.
* Tipo de Intervención Quirúrgica: Resume los distintos tipos de intervenciones quirúrgicas en una sola categoría.
* Categoría de Edad: Clasifica las edades en grupos significativos en rangos de 10 años (entre 10 y 20 años, entre 20 y 30 años…).
* Nivel de Fumador: Agrupa los diferentes niveles de consumo de tabaco en una sola columna.

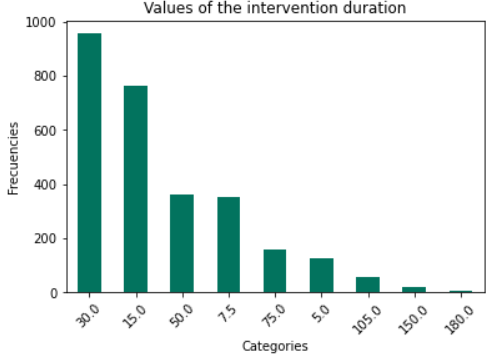
Después del análisis exploratorio categórico de la base de datos general, se va a realizar un análisis numérico separando la base de datos según el tipo de operación (Implantología Bucal y Cirugía Periimplantaria). En esta base de datos, se ha incorporado una nueva categoría denominada 'Género', donde se asignará el valor 1 para 'Mujer' y el valor 0 para 'Hombre'. Además, de nuevo se eliminan variables que no interesan en el análisis, estas son: IP\_Adress, -Operador, -Auxiliar, -Jefe\_de\_dia, -No.fumador.a, -Medicacion\_actual, -NÃºmero\_de\_cigarrillos.dÃ.a. Con la base resultante, se han analizado las diferencias más significativas entre las variables de los dos tipos de operación.

Finalmente, para realizar un PCA explicativo, se utiliza la base de datos numérica anterior, a la cual se le han eliminado las variables 'Patologia\_Sistematica' y 'Medicacion\_Actual' dado que esta información ya está representada a través de las distintas patologías y medicaciones registradas. Además, hemos eliminado las medicinas o enfermedades que aparecen en la base de datos menos de tres veces, lo cual ha contribuido a mejorar los resultados obtenidos. De esta forma, obtenemos una base de datos binaria para el análisis con el género, el nivel de fumadores u otras drogas y las enfermedades y medicamentos actuales.

Para la segunda subtarea, que es predecir la duración de la intervención, se tienen en cuenta las variables seleccionadas con el PCA como relevantes y, previamente al análisis estadístico, de toda la base de datos se seleccionan las 100 variables más influyentes a través de otra técnica de reducción de la dimensionalidad similar al PCA, un Análisis Factorial de Correspondencias Múltiple (MCA) múltiple. Este último no solo ayudaría a reducir las opciones que se abarquen a lo esencial, sino que también ayudaría a confirmar si las variables obtenidas con el PCA tienen algún sentido. Una parte del resultado es el que se muestra en la Figura 1, que se resume en que muchas de las variables sí coinciden con las del PCA, como por ejemplo: medicación actual, alcohol, tipo de intervención quirúrgica, entre otras que no se muestran por las dimensiones.



Posteriormente a este paso, se aplican limpiezas como la eliminación de aquellas intervenciones con la duración sin especificar o errónea. Tras esto, se convierten los rangos de duraciones en valores numéricos calculando la media de cada rango. Por ejemplo, para el rango de 20 a 40 minutos se asigna el valor 30 minutos. Con la distribución ordenada por frecuencia queda tal como se muestra en la Figura 2.



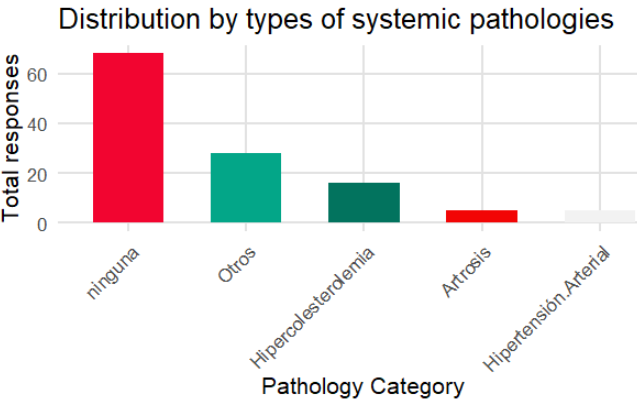
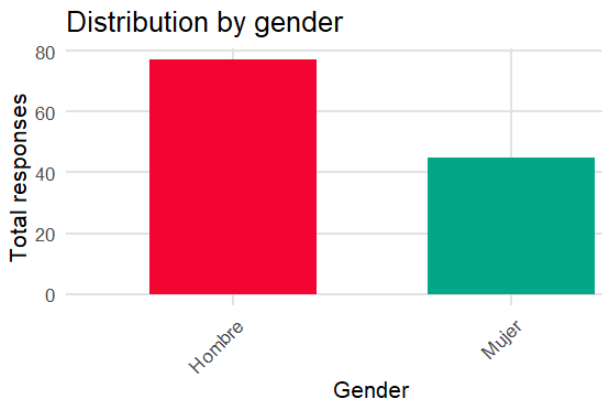
A continuación, se aplica Autoclean, que es una función que limpia los datos de manera automática y aplica un proceso similar al de one-hot encoding.

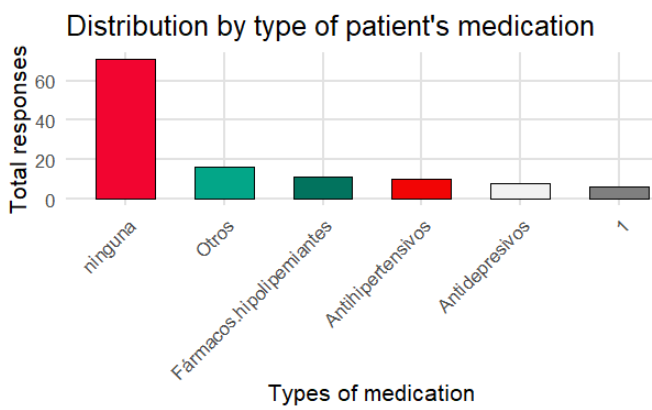
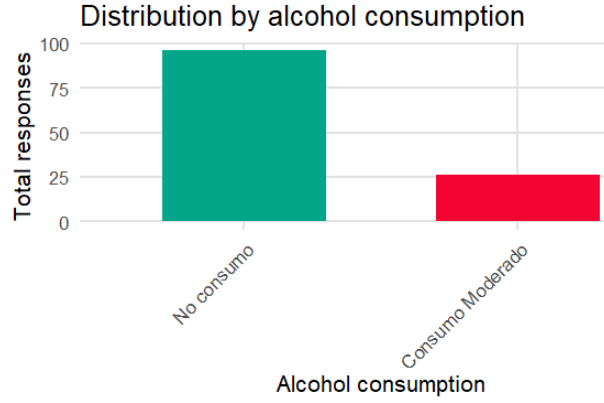
Seguidamente, se identificó un desbalance enorme en los datos, sobre todo en la variable de mayor peso (Tipo de intervención Quirúrgica). Para abordar este problema, se experimentó con diversas técnicas de reequilibrio de datos, incluyendo undersampling y oversampling, además de la implementación de modelos de aprendizaje profundo y ajustes en los hiperparámetros. La estrategia que más impacto tuvo fue la aplicación de la técnica Synthetic Minority Over-sampling Technique (SMOTE).

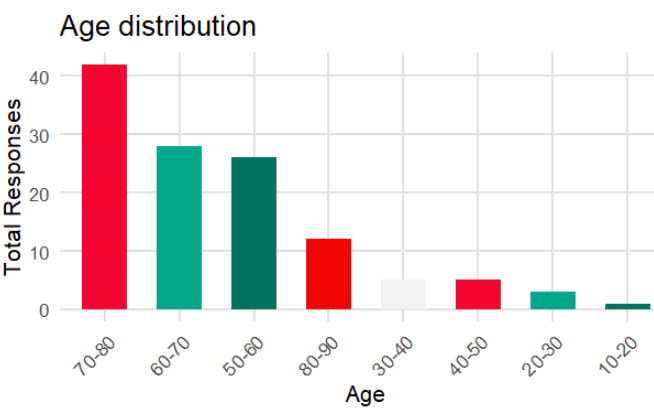
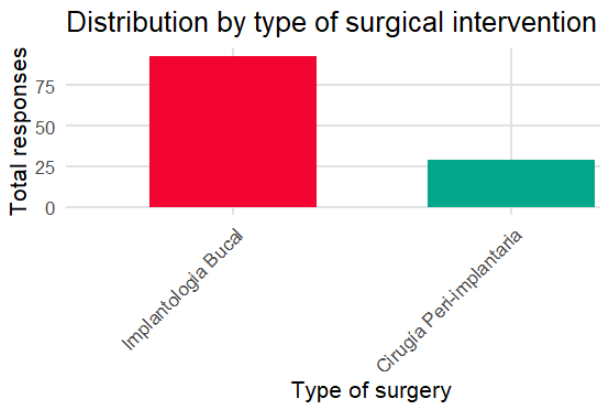
Por lo tanto, se decidió aplicar SMOTE (Synthetic Minority Over-sampling Technique) a la variable indicada, con el propósito de generar ejemplos sintéticos para las clases minoritarias y equilibrar la distribución de esta variable. Balancear esta variable en lugar de la duración de la intervención (la variable de estudio) nos proporcionó mejores resultados de predicción. SMOTE se eligió específicamente porque, a diferencia de simplemente duplicar los datos existentes, crea nuevas muestras que representan combinaciones lineales de los vecinos más cercanos de las clases minoritarias, lo que ayuda a mantener la diversidad del conjunto de datos. Esto es crucial para evitar el sobreajuste y mejorar la capacidad del modelo para generalizar. Además, SMOTE aborda directamente el problema de las clases desbalanceadas, permitiendo que el modelo tenga un mejor rendimiento en la predicción de estas clases. La mejora en el rendimiento del modelo después de aplicar SMOTE es notoria, proporcionando una mayor precisión y estabilidad en las predicciones del futuro modelo.

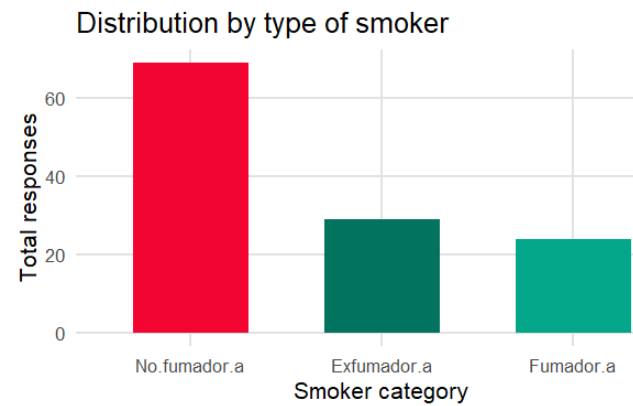
* 1. **Statistical analysis**

En una primera instancia, se presentarán los resultados del análisis exploratorio para examinar los datos contenidos en la base categórica.









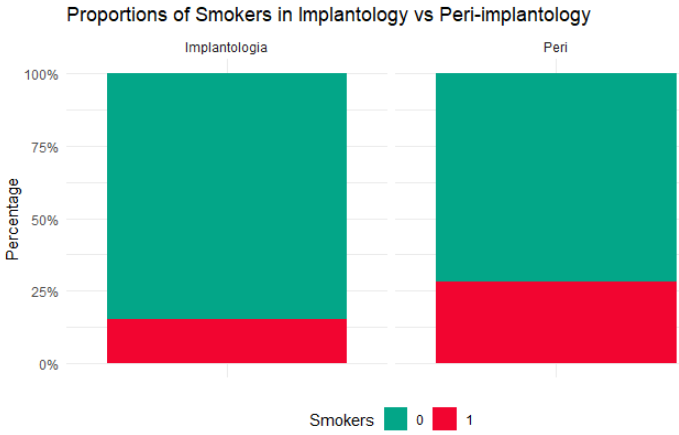
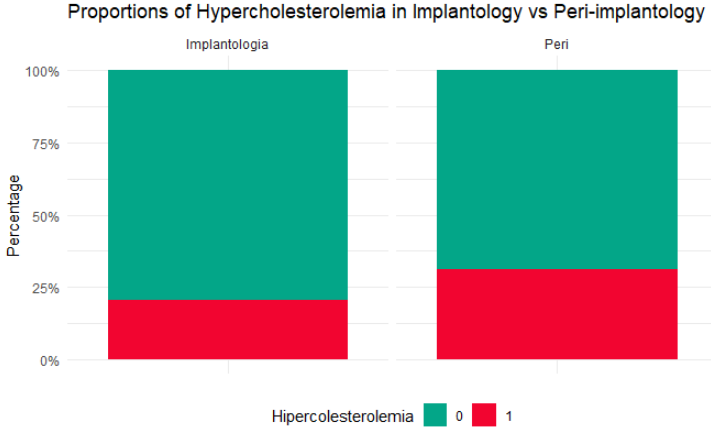
Se advierte una notable disparidad en la cantidad de datos entre la cirugía de implantología bucal y otros tipos de operaciones. Además, se observa una predominancia masculina en los registros, mientras que las edades más frecuentes oscilan entre los 60 y 80 años. Asimismo, se destaca una menor proporción de no fumadores en comparación con fumadores. Por último, se identifican como enfermedades más comunes la hipercolesterolemia, la artrosis y la hipertensión arterial.

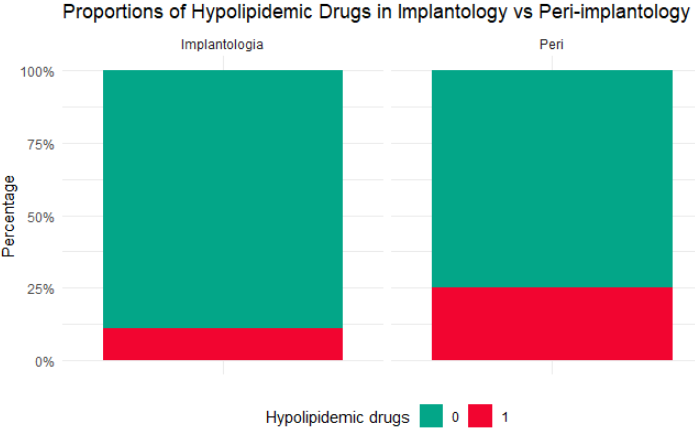
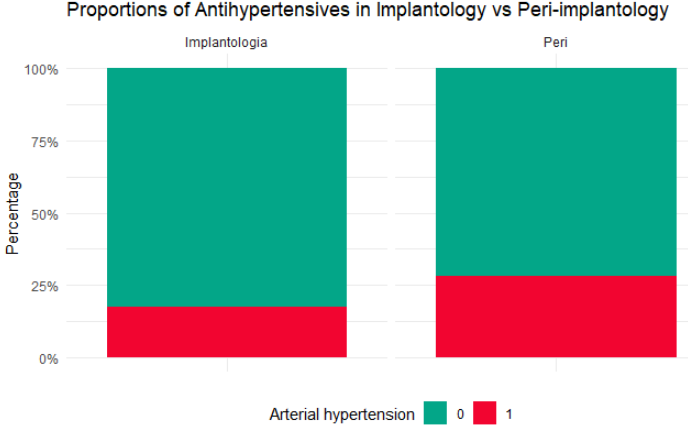
A su vez, para abordar el primer objetivo, hemos generado la base de datos numérica de los pacientes descrita anteriormente. Esta base de datos incluye dos tipos de procedimientos dentoalveolares: la implantología bucal, que implica la colocación de implantes dentales; y la cirugía peri-implantaria, destinada a abordar posibles complicaciones de peri-implantitis surgidas tras la implantología.

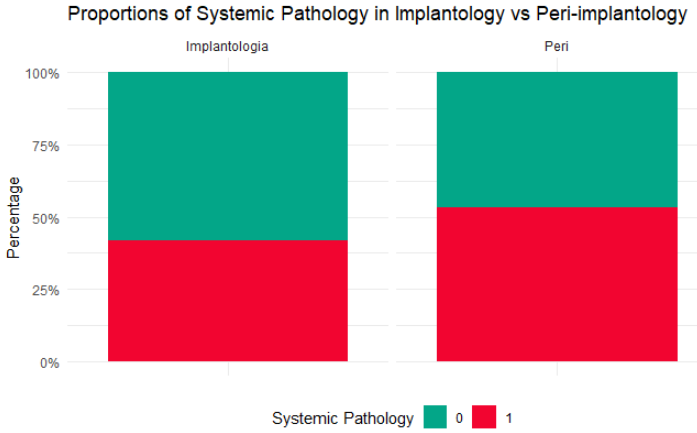
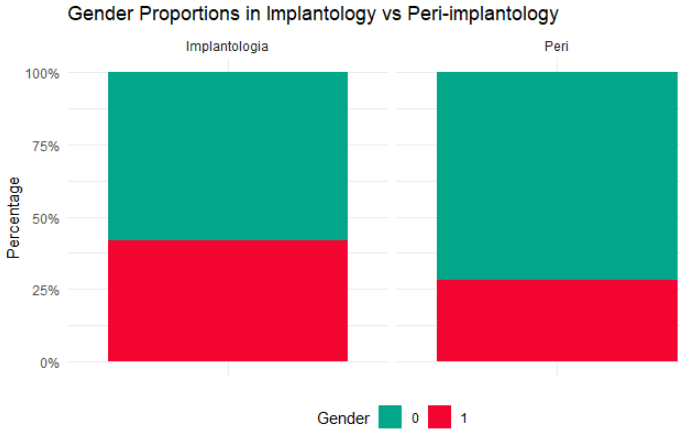
Dado que los pacientes sometidos a ambos tipos de operaciones (implantología bucal y cirugía peri-implantaria) no están relacionados, es imposible emplear un modelo para predecir si una implantología resultará en peri-implantitis. Sin embargo, es factible analizar y comparar las diferencias entre los pacientes que han experimentado periimplantitis y aquellos que han recibido implantes dentales, incluso considerando que estos últimos también podrían enfrentar problemas de periimplantitis en el futuro. Este análisis nos permitirá identificar si existe una variabilidad significativa en alguna variable específica entre ambos grupos, lo cual podría señalar un factor diferenciador relevante.

Para ello, inicialmente, optamos por emplear un test de chi-cuadrado para evaluar las diferencias entre la misma variable en cada base de datos. Sin embargo, este método generalmente exige que las frecuencias esperadas en cada celda de la tabla de contingencia sean al menos de 5. En nuestro caso, la mayoría de las variables no cumplían con este criterio, lo que podría comprometer la efectividad del test. Por consiguiente, decidimos realizar una diferencia de medias entre las variables de cada base de datos. Este enfoque facilita una comparación directa y más clara entre los dos grupos, centrada específicamente en la métrica de interés. Además, este método resulta más intuitivo y está directamente vinculado con la pregunta de investigación, ofreciendo una interpretación más directa de los resultados.

De esta forma, después de seleccionar las 10 variables con más diferencia entre sus medias entre los pacientes de periimplantitis y toda la población, obtenemos que las características que más diferencian a los dos tipos de paciente son: "Fármacos.hipolipemiantes", "Genero", "Fumador.a", "Patologia\_1stemica", "Hipercolesterolemia", "Hipertensión.Arterial", "Antihipertensivos", "Medicacion\_Actual2", "Artrosis" y "Insulina". Más adelante, se analizará el significado y la relevancia de estos resultados en comparación con estudios previos sobre el tema.

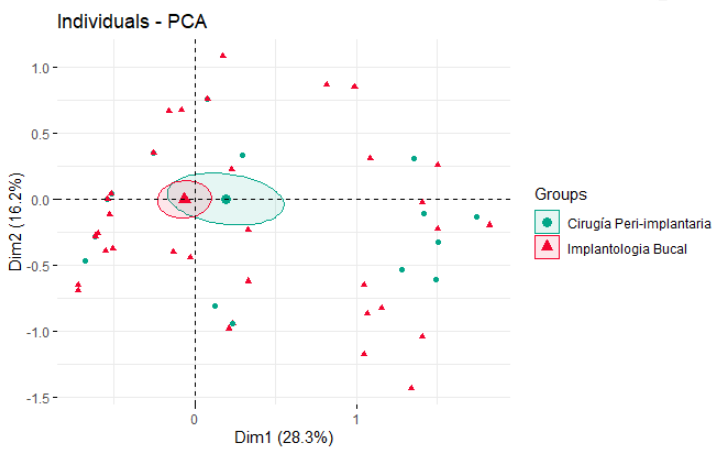
A continuación se muestra la diferencia de proporciones de las variables más significativas para nuestro estudio: 

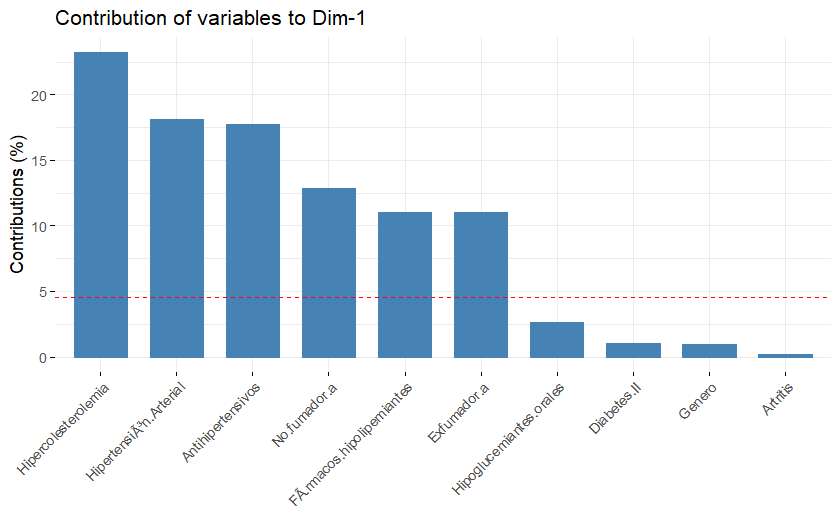




Otro enfoque que consideramos para determinar si existen variables significativas que diferencian a los pacientes es la implementación de un análisis de componentes principales (PCA) explicativo tras unir las dos bases de datos. Este método nos permite, a parte de identificar patrones ocultos en los datos, observar si alguna dimensión resultante logra separar a los grupos de pacientes de manera efectiva. Sin embargo, es importante reconocer las limitaciones inherentes a este enfoque. Como hemos mencionado anteriormente, algunos pacientes que actualmente poseen implantes podrían desarrollar periimplantitis en el futuro. Esta eventualidad impide una separación clara y definitiva entre los grupos mediante el PCA.

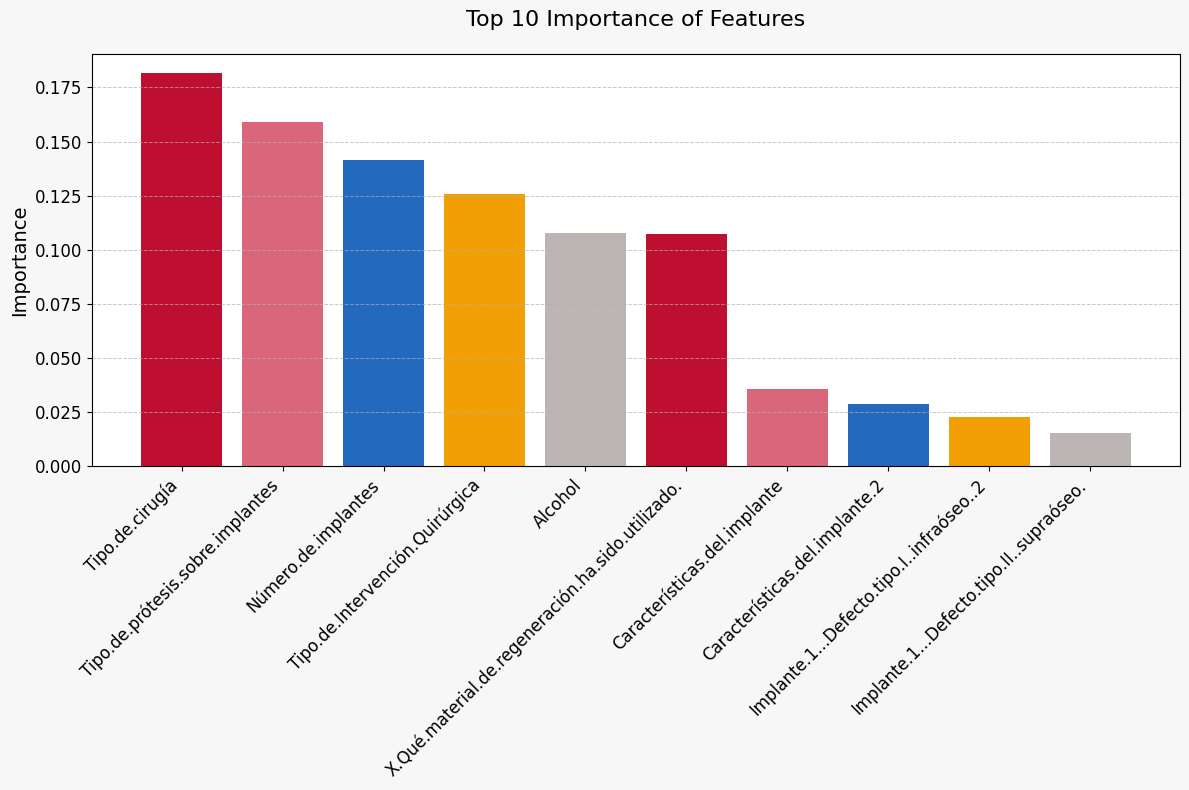
Después de realizar el análisis de componentes principales (PCA), encontramos que las dos primeras dimensiones explican hasta un 44.5% de la variabilidad total de los datos. Al visualizar estos resultados mediante un gráfico de dispersión, donde los puntos representan a los pacientes y están coloreados según el tipo de procedimiento, observamos ciertas tendencias. Aunque los pacientes de implantología bucal no se distinguen claramente, lo cual podría deberse a la presencia de casos futuros de periimplantitis, la mayoría de los pacientes de periimplantitis tienden a dividirse por la primera dimensión.

Conociendo las limitaciones de nuestro conjunto de datos, comprobamos que las variables que más influyen en esta primera dimensión, coinciden con aquellas que previamente identificamos como las más diferenciadoras entre los pacientes de periimplantitis y los de implantología bucal. Este hallazgo refuerza la relevancia de estas variables en la caracterización de los grupos de estudio. 



Así, se observa como las variables como Hipercolesterolemia, Antihipertensivos, Hipertensión Arterial, Fumador, Fármacos hipolipemiantes, Insulina y Diabetes tipo 2, las cuales coinciden en los dos análisis, destacan como las más relevantes para distinguir entre los pacientes con peri-implantitis.

Para predecir el tiempo de intervención se procede con el método Random Forest, con el objetivo de reducir aún más las variables predictoras a las realmente determinantes para predecir la duración de la intervención. Las 10 variables resultantes del Random Forest son las que se muestran en la Figura 3 y las que se escogen para analizar y predecir las duraciones con la ayuda del modelo Decission Tree Regressor el cual nos devolvía el error más pequeño, como se explicará más adelante.



En la continuación del estudio sobre la estimación del tiempo de intervención quirúrgica, se emprendió un tercer objetivo: desarrollar un modelo manual para clasificar la dificultad de las intervenciones quirúrgicas. Este modelo se fundamenta en la selección de variables críticas que influyen directamente en el proceso quirúrgico, proporcionadas por el Dr. Rui Figuereido.

Las variables seleccionadas incluyen patologías del paciente, tipo de operaciones, género del paciente y la duración de la intervención, esta última calculada previamente mediante el modelo de predicción de tiempos de intervención desarrollado. Para clasificar las operaciones, se implementó un método de puntuación ponderada. Siguiendo los criterios de la Asociación Americana de Anestesiología (ASA) para las variables patológicas y asignando una mayor puntuación a las intervenciones de mayor duración, se estableció un sistema de puntos que refleja la complejidad y el riesgo asociado con cada intervención.

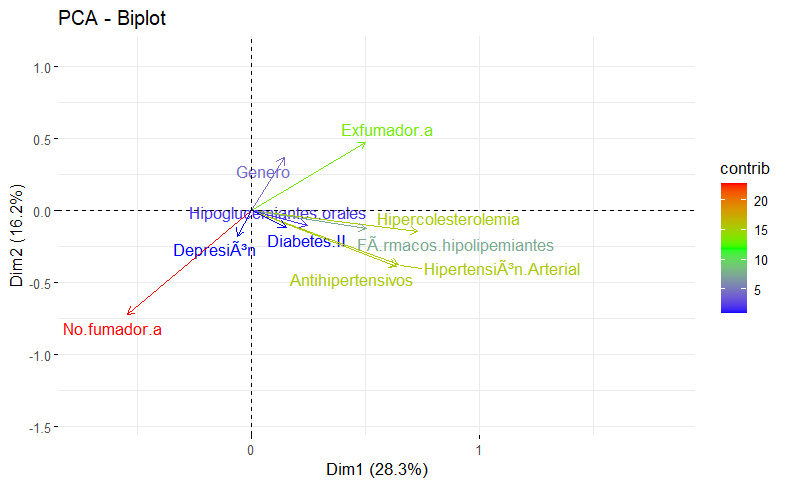
Con estas puntuaciones, se procedió a calcular los cuartiles de la distribución de los datos, asignando a cada uno de estos rangos una categoría de dificultad: Muy fácil, Fácil, Moderado y Difícil. Esta clasificación cuantitativa permitió estructurar las intervenciones en cuatro niveles claros de dificultad, facilitando así la gestión de recursos y la preparación del personal médico.

Para validar la efectividad del modelo de clasificación, se ajustó un modelo de Random Forest (RF) utilizando estas categorías como variables objetivo. El resultado que se observará más adelante, fue un ajuste casi perfecto, lo cual indica que el modelo es capaz de replicar y predecir de manera precisa la clasificación de dificultad basada en los parámetros establecidos. Sin embargo, es importante reconocer que este alto grado de ajuste puede ser indicativo de un sobreajuste, dado que el modelo fue diseñado de manera casi manual y está altamente especializado para las condiciones del conjunto de datos actual. Esta especificidad puede limitar la generalización del modelo a otros contextos clínicos.

Este enfoque de clasificación no solo proporciona una herramienta práctica para la evaluación preoperatoria, sino que también contribuye al cuerpo académico con un método innovador para gestionar y anticipar las necesidades en el entorno quirúrgico.

1. **Results:** conclusiones que nos salen (para el primer objetivo por un lado y el segundo por el otro, incluye métricas y evaluación y todo eso).

Tras concluir los análisis mencionados para la subtarea 1, identificamos que las variables más significativas en la diferenciación de pacientes con periimplantitis son la hipercolesterolemia, los antihipertensivos, la hipertensión arterial, los fármacos hipolipemiantes, la insulina, la Diabetes tipo 2 y el ser fumador o no. El análisis mediante el biplot de las variables que contribuyen a la primera dimensión corroboran una influencia de estas variables en la separación de los pacientes, así como que el hecho de no fumar también puede ayudar a diferenciarlos.

Además, se observan relaciones intrínsecas entre nuestros datos, derivadas de sus características. Para comprender la relevancia de las variables identificadas como importantes, es crucial considerar ciertos aspectos. Se nota la coexistencia de enfermedades como la hipertensión arterial, la hipercolesterolemia y la diabetes, junto con los fármacos asociados a su tratamiento (antihipertensivos, fármacos hipolipemiantes y la insulina respectivamente). Estas tres enfermedades guardan una relación médica estrecha. Por lo tanto, se puede inferir que los pacientes afectados por hipertensión arterial, hipercolesterolemia o diabetes, así como los fumadores, representan un grupo de mayor riesgo en las intervenciones quirúrgicas y podrían requerir un seguimiento más cercano.

Predicción del Tiempo de Intervención Quirúrgica

Este objetivo se abordó utilizando un modelo de árbol de decisión mejorado con la técnica de Synthetic Minority Over-sampling Technique (SMOTE) para equilibrar los datos. Presentamos las métricas de evaluación para dos iteraciones del modelo, enfatizando la importancia de considerar tanto el RMSE como el MAE:

Primera Iteración del Árbol de Decisión antes del SMOTE:

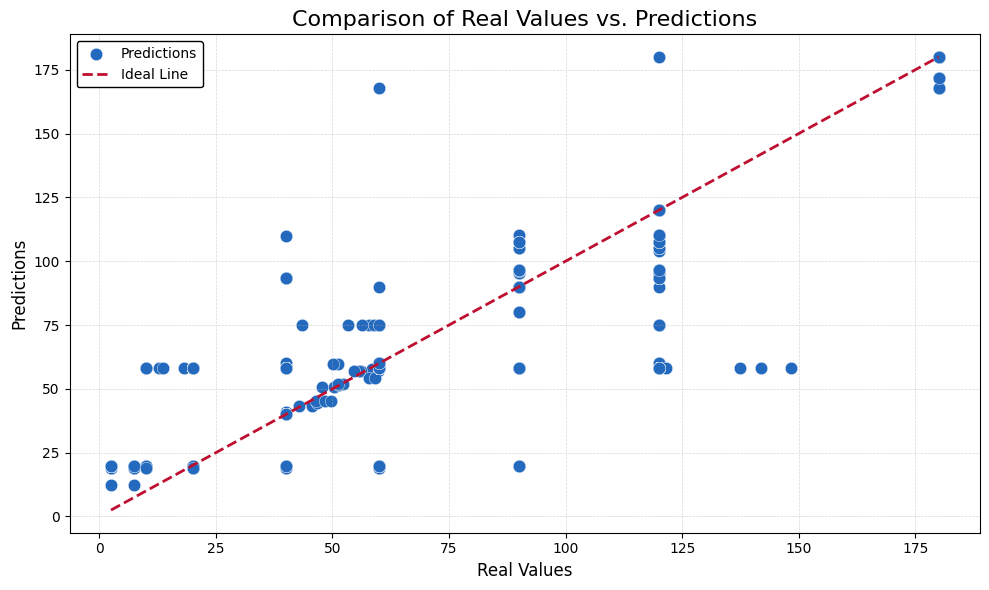
* Mean Squared Error (**MSE**): 407.52
* Root Mean Squared Error (**RMSE**): 20.19
* Mean Absolute Error (**MAE**): 14.74

Segunda Iteración del Árbol de Decisión después del SMOTE:

* Mean Squared Error (**MSE**): 374.26
* Root Mean Squared Error (**RMSE**): 19.35
* Mean Absolute Error (**MAE**): 11.84

El MSE y el RMSE son métricas comunes para evaluar modelos de regresión, sin embargo, ambos son especialmente sensibles a valores atípicos porque elevan al cuadrado las diferencias entre los valores predichos y reales. En contextos donde la distribución de los datos incluye valores extremos o el rango de la variable objetivo es muy amplio, el RMSE puede dar una impresión engañosa de un mal rendimiento del modelo, puesto que magnifica el impacto de los errores grandes.

Adjuntamos un gráfico para visualizar las predicciones y los datos reales.



Por otro lado, el MAE proporciona una medida más intuitiva y menos sensible a los valores extremos, ya que se basa en la media aritmética de los errores absolutos. Esto lo hace más representativo de la tipicidad de los errores en la predicción, proporcionando una mejor visión del rendimiento típico del modelo en situaciones cotidianas. Por esta razón, aunque el RMSE puede ser útil para destacar problemas en presencia de valores atípicos, el MAE ofrece una perspectiva más robusta y equilibrada para evaluar la precisión de las predicciones en nuestro estudio.

Clasificación de la Dificultad de las Intervenciones Quirúrgicas

El segundo objetivo empleó un modelo de Random Forest para clasificar la dificultad de las intervenciones quirúrgicas con las siguientes métricas de desempeño:

Accuracy del modelo: 97.15%

Métricas de Clasificación:

* Difícil: **precisión** 0.88, **recall** 0.99, **f1-score** 0.93
* Fácil: **precisión** 0.99, **recall** 0.96, **f1-score** 0.97
* Moderada: **precisión** 0.98, **recall** 0.95, **f1-score** 0.96
* Muy Fácil: **precisión** 0.98, **recall** 0.99, **f1-score** 0.99

La precisión general y los f1-scores del modelo de Random Forest resaltan su capacidad para clasificar acertadamente las intervenciones quirúrgicas en diferentes niveles de dificultad, destacando particularmente en la identificación de casos desafiantes. Sin embargo, es importante señalar que este modelo fue diseñado de manera casi manual y está optimizado específicamente para nuestro conjunto de datos, lo cual limita su aplicabilidad a otros contextos sin ajustes adicionales. A pesar de esta limitación, los resultados confirman que el modelo es efectivo y valioso para la planificación y formación médica dentro de nuestro ámbito de estudio.

1. **Discussion and value assessment:**

Estas tres enfermedades guardan una relación médica estrecha. Por lo tanto, se puede inferir que los pacientes afectados por hipertensión arterial, hipercolesterolemia o diabetes, así como los fumadores, representan un grupo de mayor riesgo en las intervenciones quirúrgicas y podrían requerir un seguimiento más cercano.

Se han investigado numerosos artículos y estudios científicos sobre la periimplantitis y sus causas con el objetivo de prevenirla, destacando la creciente importancia de esta enfermedad, cada vez más frecuente y menos conocida(citar artículo 3). Estos estudios proporcionan información interesante y variada, pero coinciden en que tanto el tabaquismo como la diabetes son factores de riesgo significativos para el desarrollo de la periimplantitis.

El tabaquismo modifica varios aspectos de la inmunidad innata y adaptativa del huésped (citar artículo 1), ya que se conoce que la nicotina daña la síntesis de proteínas y afecta la capacidad de adhesión de los fibroblastos de la encía (citar artículo 2). En cuanto a la diabetes, los niveles elevados de glucosa en sangre producen productos finales de glicación avanzada (citar estudio 1), lo que hace que los pacientes presenten una respuesta inmune y cicatrización deficiente, por lo que son pacientes de riesgo a la hora de realizar un tratamiento quirúrgico (citar estudio 4).

Además, se puede observar más información relacionada con nuestros resultados relacionada con la hipercolesterolemia y la hipertensión. En relación con el colesterol, se observó un estudio que evaluó su influencia en el nivel óseo marginal periimplantario (MBL) mediante la medición de los niveles de lípidos en la sangre de los pacientes al inicio y su seguimiento posterior. Este estudio concluyó que el colesterol total es una de las variables que más influye en la MBL. (citar estudio 5) Sin embargo, el estudio presentaba limitaciones debido a la falta de una muestra suficientemente grande, una limitación similar a la de los resultados del presente estudio.

Finalmente, se revisó un artículo que observó una tendencia a una mayor prevalencia de periimplantitis en el grupo con enfermedad cardiovascular (citar artículo 6). Sin embargo, el estudio no es concluyente en cuanto a la relación con la hipertensión arterial debido a factores de confusión, lo que sugiere que la verdadera relación podría ser con la diabetes, que a su vez propicia la hipertensión. Para confirmar esta teoría, sería necesario realizar estudios adicionales que verifiquen la hipótesis de que la hipertensión arterial influye directamente en la periimplantitis y no es solo un factor de confusión. Para ello, de nuevo, se necesitaría una muestra de datos más amplia.

Además, se ha demostrado como la predicción de la duración de las intervenciones quirúrgicas dentales y la clasificación de las operaciones por su dificultad, es de gran valor para la gestión clínica. Mediante el uso de técnicas estadísticas avanzadas como el Random Forest y el Análisis Factorial de Correspondencias Multiples (MCA), se identificaron las variables más influyentes que afectan el tiempo de las intervenciones. Entre estas, factores como la posición del diente y características específicas del procedimiento jugaron un papel crucial. Estos elementos son fundamentales dado que por ejemplo: la accesibilidad y la complejidad asociada a diferentes posiciones dentales pueden variar significativamente, influyendo en el tiempo requerido para completar la intervención.

La precisión del modelo, con un error medio absoluto de 14 minutos y posteriormente reducido hasta los 11 minutos de error, fue considerada satisfactoria por el profesional colaborador del estudio, Dr. Rui Figuereido. La relevancia de este nivel de precisión radica en su aplicación práctica, no se busca una precisión minuto a minuto, sino categorizar las intervenciones en duraciones aproximadas (corta, media, larga). Este nivel de precisión es adecuado para planificar las operaciones y asignar recursos de manera efectiva, permitiendo una mejor organización de las camillas y del tiempo del alumnado.

Asimismo, la implementación de los hallazgos del análisis de las complicaciones en las operaciones en una aplicación Streamlit proporciona un valor añadido significativo. Esta herramienta digital permite a los operadores clínicos ingresar datos específicos de cada paciente y procedimiento, recibiendo como salida la duración estimada de la intervención y la clasificación de la dificultad operatoria a la que se van a enfrentar. Esto facilita la planificación diaria, permite ajustar las expectativas del tiempo de ocupación de salas y mejora la eficiencia en la asignación de estudiantes para diferentes tipos de procedimientos según su nivel de experiencia.

Actualmente, la clínica universitaria agenda a todos los pacientes con una asignación estándar de una hora por procedimiento, una práctica que el modelo ayuda a optimizar significativamente. Al prever de manera más precisa la duración de cada intervención, se pueden ajustar los horarios de manera más efectiva, evitando ineficiencias y tiempos muertos, y permitiendo una asignación de recursos más precisa.

Este enfoque no solo mejora la logística y la eficiencia operativa de la clínica, sino que también asegura que los estudiantes de odontología puedan ganar experiencia en intervenciones que se alineen con su nivel de habilidades y conocimientos. Además, al maximizar el uso de los recursos disponibles, la clínica puede potencialmente aumentar el número de pacientes atendidos y, por ende, ampliar las oportunidades educativas para más estudiantes.

1. **Legacy:** aquí hablamos de lo del paper y si vamos a dejar el código o los análisis en alguna plataforma como Github. También hay que hablar de que ODS se verán afectadas.

En cuanto a las conclusiones derivadas de la información de los pacientes susceptibles a sufrir periimplantitis, se propone la redacción de un artículo científico con el objetivo de su publicación en una revista académica. Para ello, se han identificado revistas adecuadas en el Scimago Journal & Country Rank, como la International Journal of Oral Science, que publica investigaciones sobre todos los aspectos de la ciencia oral y campos interdisciplinarios relacionados ([fuente](https://www.scimagojr.com/journalsearch.php?q=19700180533&tip=sid&clean=0)). Además, el código desarrollado podría reutilizarse con muestras de datos más extensas en el futuro, con el fin de obtener resultados más concluyentes.

Además, la web desarrollada con Streamlit aprovecha los resultados de un estudio para mejorar las intervenciones odontológicas. Ofrece funcionalidades para predecir la duración de las cirugías basándose en datos específicos, lo que ayuda a estimar el tiempo requerido de manera más precisa. También analiza la complejidad de las operaciones a partir de la información patológica de los pacientes, clasificando la dificultad de las intervenciones para una mejor planificación y preparación. La plataforma brinda detalles sobre los hallazgos del estudio, permitiendo a los usuarios examinar cómo las cirugías odontológicas se relacionan con diversas patologías sistémicas. Esta herramienta no solo mejora la interpretación y aplicación práctica de los datos, sino que también fomenta el progreso en el ámbito de la odontología y la ciencia de datos. Se puede acceder a la web a través de [este enlace](https://monitormedico.streamlit.app/).

Por último, este proyecto tiene un impacto directo en el ODS 3 (Salud y Bienestar) al enfocarse en mejorar los resultados de la cirugía dentoalveolar y fortalecer la gestión clínica en el sector odontológico. Al identificar y abordar los factores de riesgo en este tipo de cirugía, se pueden minimizar las complicaciones postoperatorias, lo que garantiza un tratamiento odontológico más seguro y efectivo para los pacientes. Además, al optimizar la gestión clínica, se mejora el acceso a los servicios de odontología, se reducen los tiempos de espera y se aumenta la eficiencia en la atención. Este enfoque integral no solo promueve la salud bucal, sino que también contribuye al bienestar general de las personas al proporcionarles una atención médica de calidad y accesible.

1. **Conclusion:** resume lo más esencial de todo el trabajo e igual algún dato interesante.

Este estudio, realizado en colaboración con el Máster de formación permanente en Periodoncia y Periimplantologia de la Universidad de Barcelona, utilizó análisis de datos avanzados para optimizar la gestión clínica y la formación en cirugía dentoalveolar. Se establecieron dos objetivos principales: identificar perfiles de pacientes con alto riesgo de complicaciones postoperatorias y desarrollar modelos predictivos para estimar la duración y dificultad de las intervenciones quirúrgicas.

El análisis de los pacientes de riesgo, permitió prever necesidades específicas de manejo especializado y ajustar los protocolos de intervención para minimizar riesgos para pacientes que sufran hipertensión arterial, diabetes del tipo 2, hipercolesterolemia o sean fumadores. Paralelamente, se implementaron modelos predictivos en una plataforma Streamlit, mejorando la asignación de recursos y la programación de actividades clínicas con una precisión de error medio absoluto reducido de 14 a 11 minutos.

Estos avances podrían mejorar significativamente la eficiencia operativa y la experiencia formativa de los estudiantes dentro de la clínica específica, aunque el modelo está altamente personalizado para esta instalación en particular y no está diseñado para una aplicación global sin adaptaciones considerables.

Es importante señalar que, con una mayor cantidad de datos más completa, en el futuro se podrían obtener conclusiones más robustas y desarrollar aplicaciones útiles para una amplia variedad de clínicas dentales.