



# **INPATIENT ANALYSIS AND PREDICTING LENGTH OF STAY**



# OBJECTIVE

- **Predecir el tiempo** que un paciente estará ingresado (regresión)
- Solo **datos disponibles** al ingresarse
- **Factores** relacionados a su **condición**

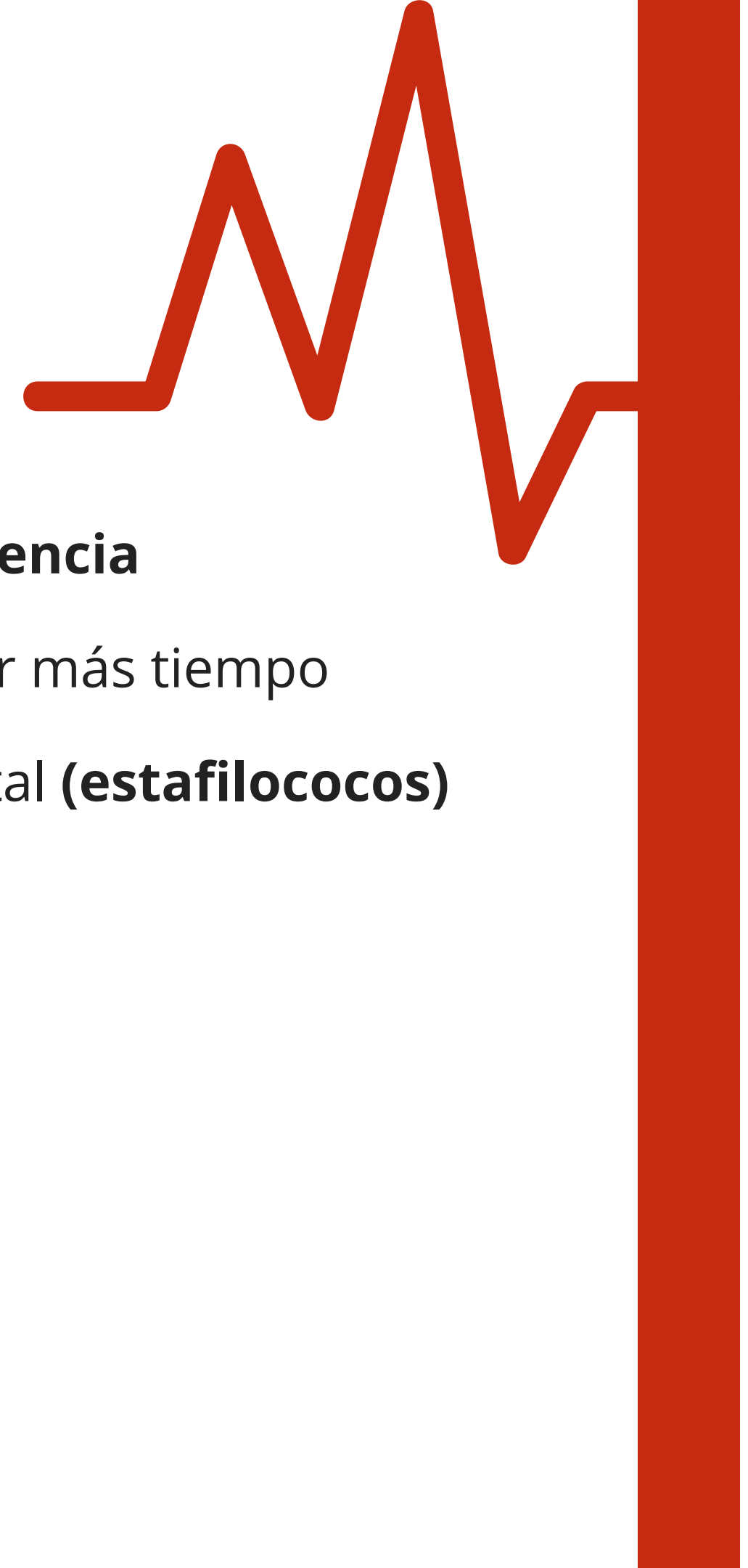




# INTRODUCTION

- Coste medio de operaciones de un hospital es **~\$2500 por paciente por día, ~\$180,000,000 por año** en Los Estados Unidos (200 camas)
- **Precios** de las intervenciones son **estandarizados**
- **Gran incentivo para reducir** tiempos de ingreso
- Pandemia del **Covid-19**

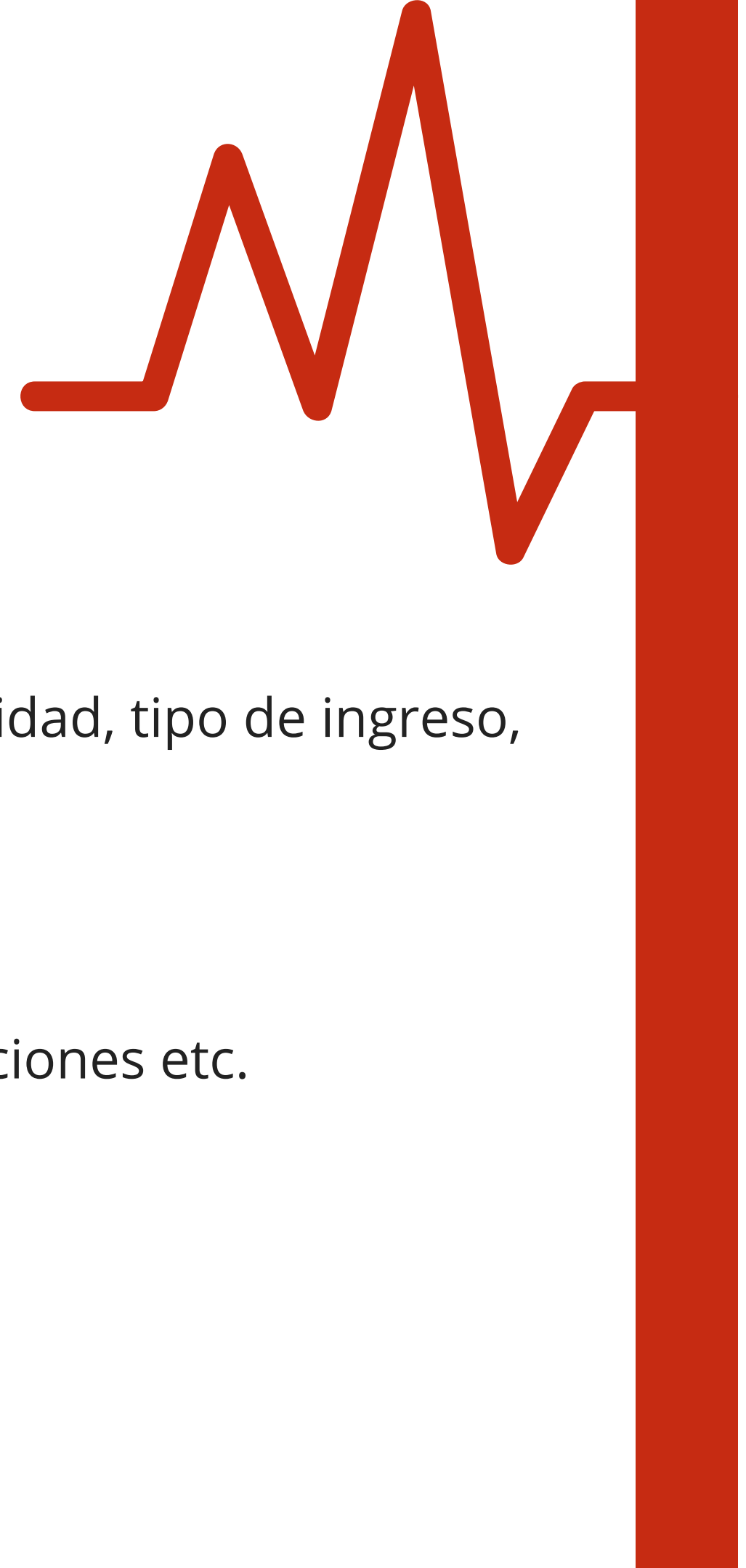
# BENEFITS OF THE PROJECT



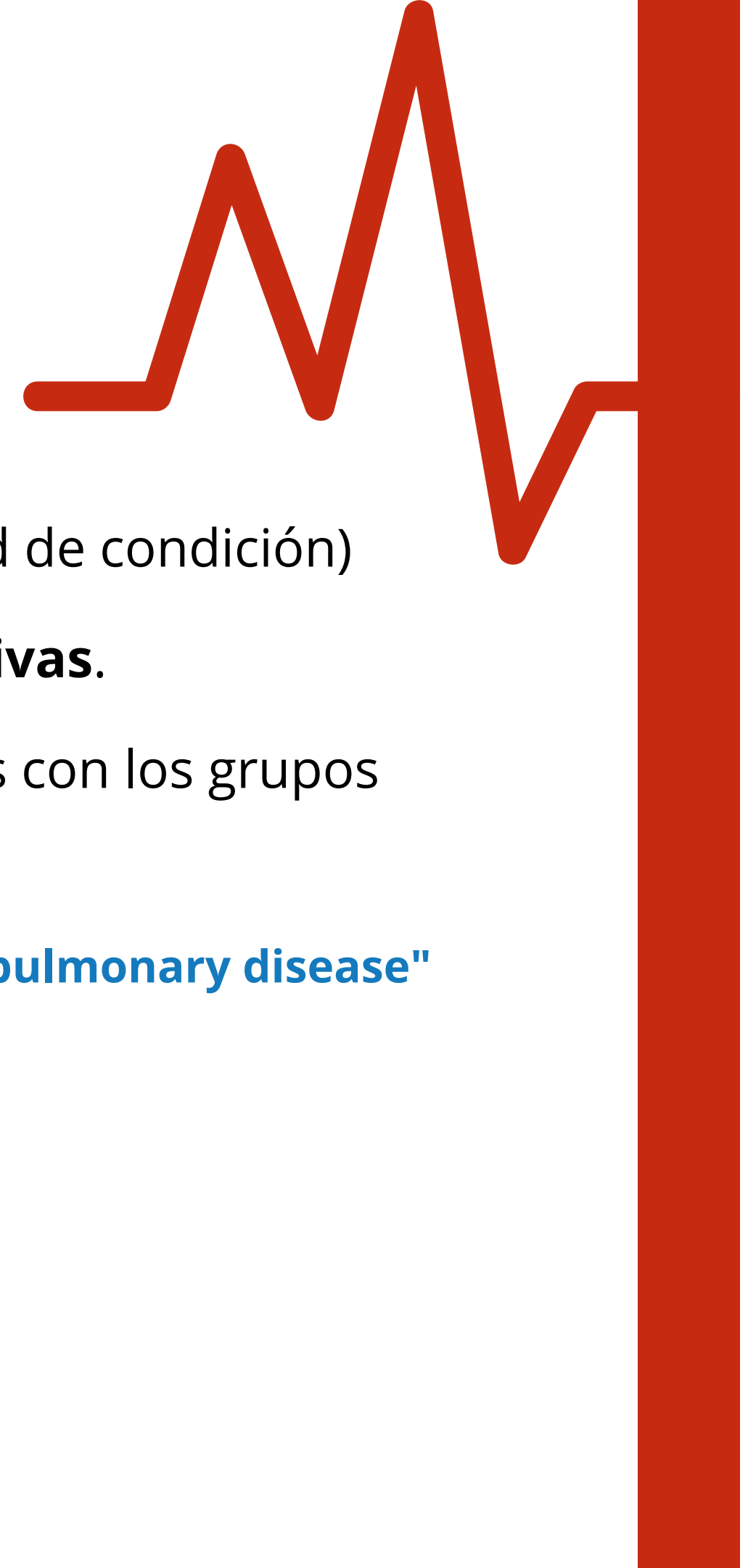
- Nos permitirá analizar los **factores** que tienen **mayor influencia**
- **Identificar** los **pacientes** con **mayor probabilidad** de estar más tiempo
- **Reducciones de infecciones** contagiadas dentro del hospital (**estafilococos**)
- **Efectos no deseados** de los medicamentos
- Menos recursos = **beneficios económicos**
- Reducir **impacto** sobre el **medio-ambiente**

# THE DATA

- **Departamento de salud** de Nueva York
- **37 variables y 2,500,000 registros**
- **Datos estructurados:** edad, genero, raza, riesgo de mortalidad, tipo de ingreso, gravedad de la condición etc.
- **Datos no estructurados:** diagnosticas, intervenciones.
- **Otras variables:** números ID, numero de certificado, ubicaciones etc.

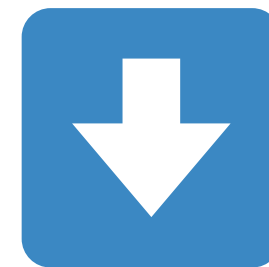
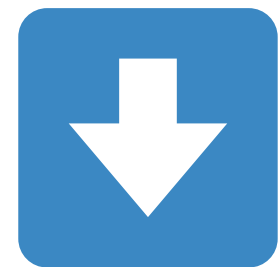


# FEATURE ENGINEERING



- **Selección** de variables **por intuición** (edad, genero, gravedad de condición)
- La mayor parte de la **información** en las **columnas descriptivas**.
- **Agrupaciones** utilizando **palabras clave** y creando columnas con los grupos

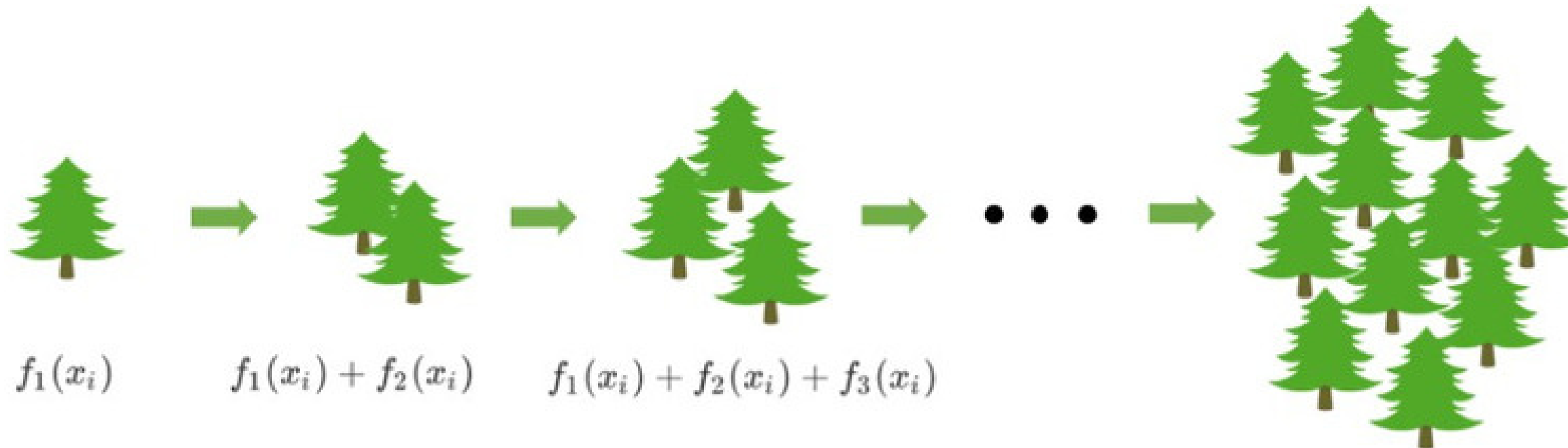
"Major respiratory infections & inflammations" , "Chronic obstructive pulmonary disease"



"Respiratory System Illnesses"

# EL MODELO: XGBOOST

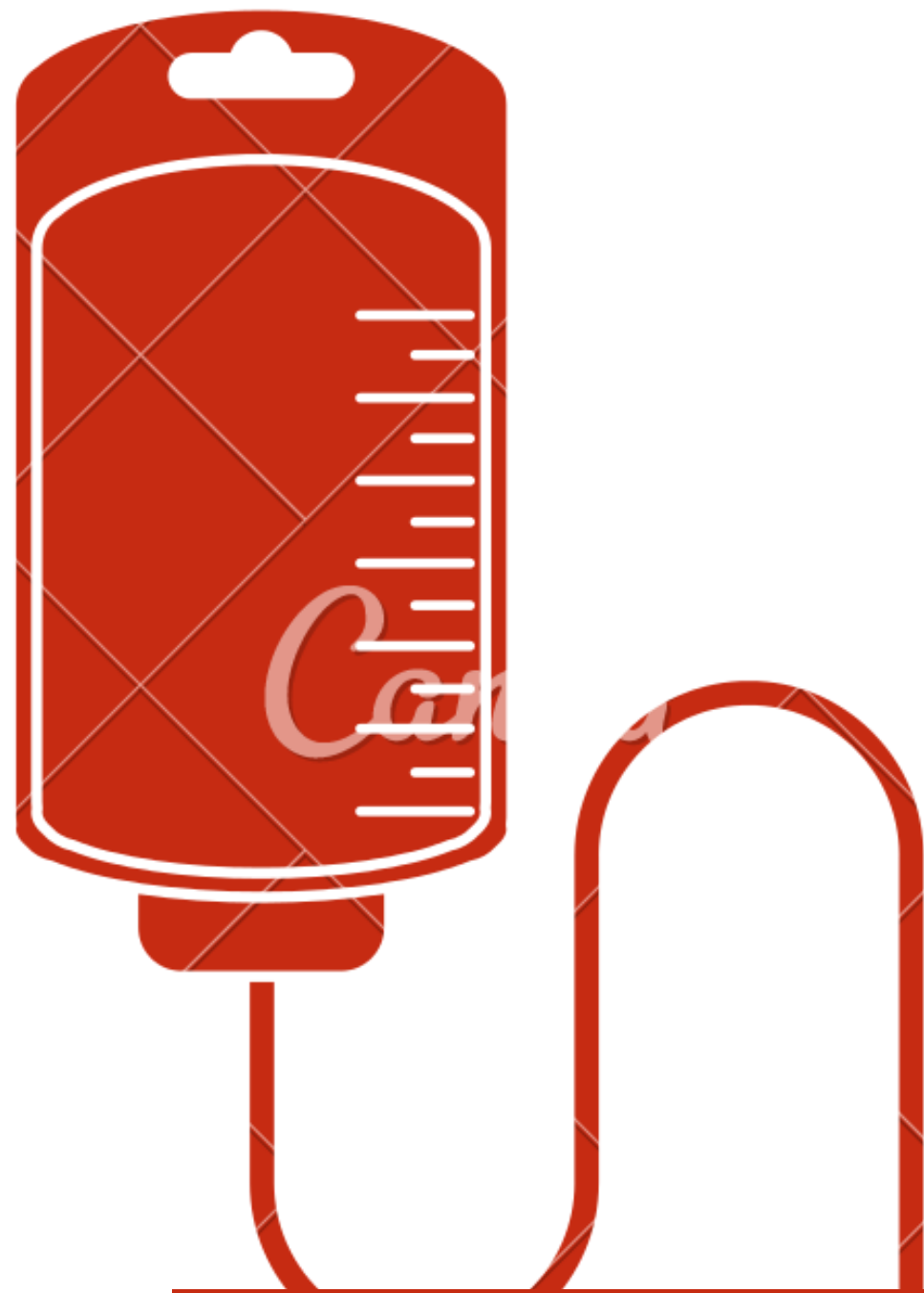
- Crear muchos **arboles** de decision **muy poco profundos**
- Arboles son "**weak learners**"
- Cada arbol coge una **muestra del conjunto**
- Cada arbol nuevo **corrige el error** del arbol anterior



# THE METRIC

- RMSE = root mean squared error (**raíz del error cuadrático medio**)

$$\text{RMSE} = \sqrt{\sum \frac{(y_{pred} - y_{ref})^2}{N}}$$

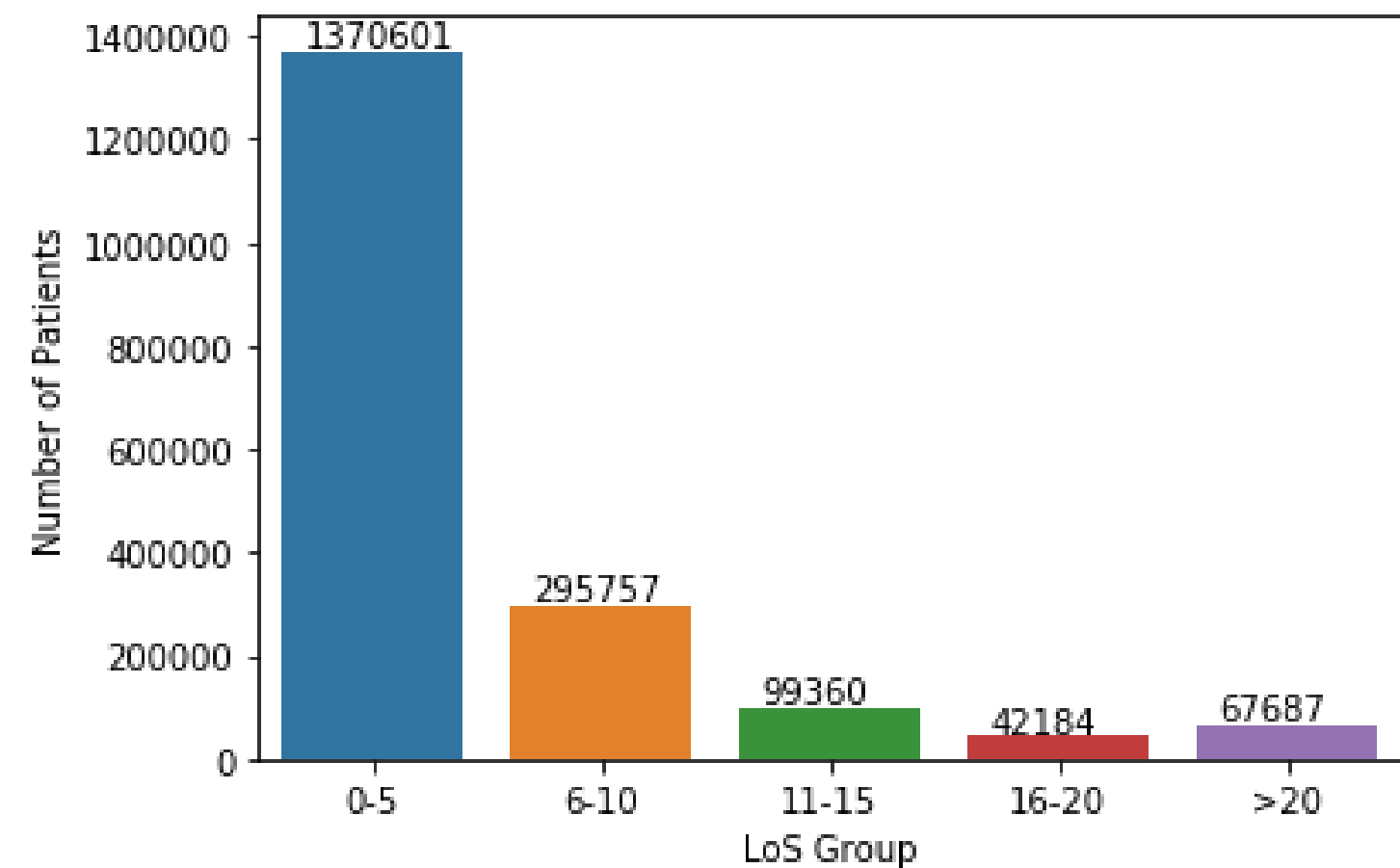
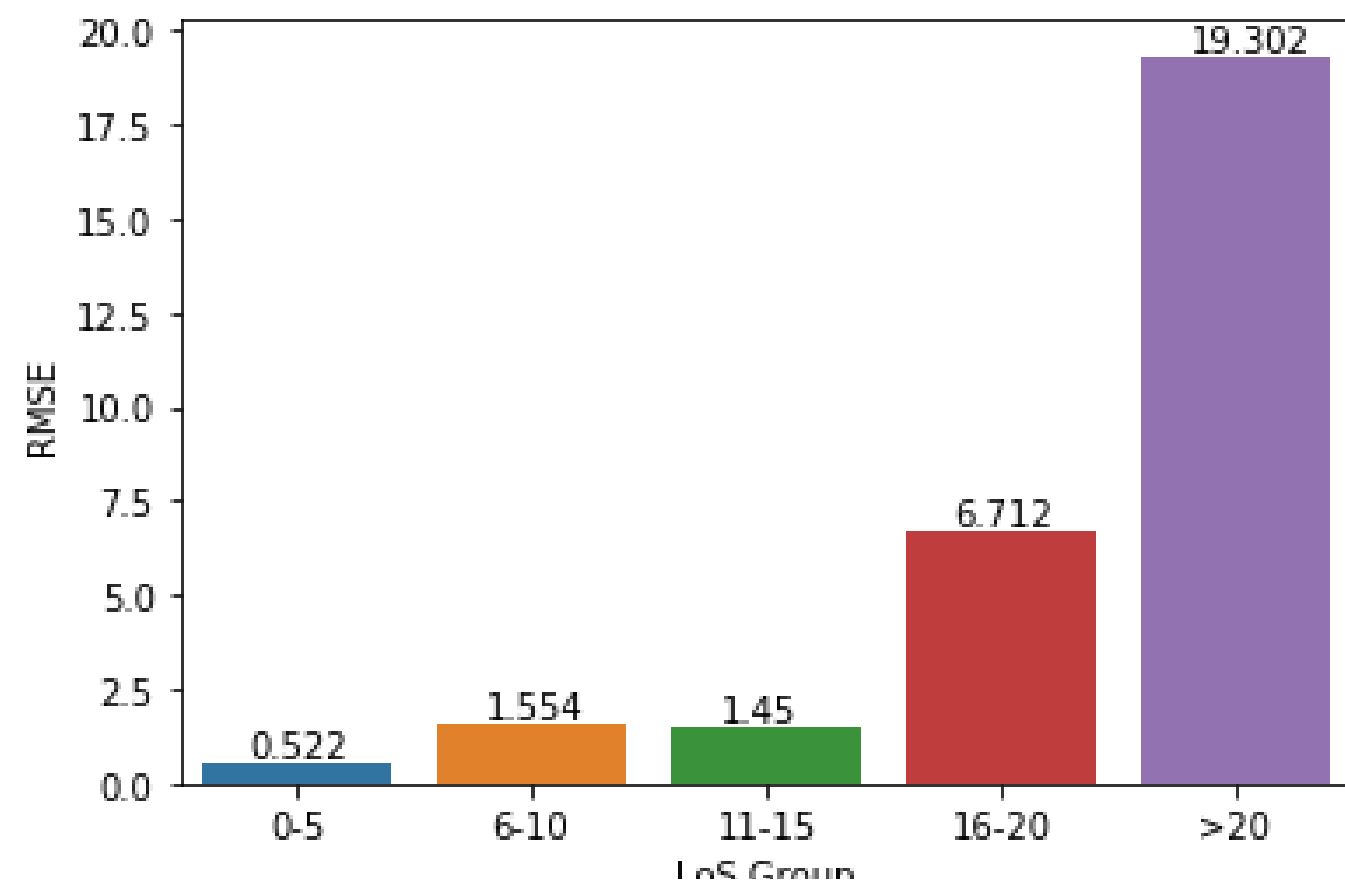




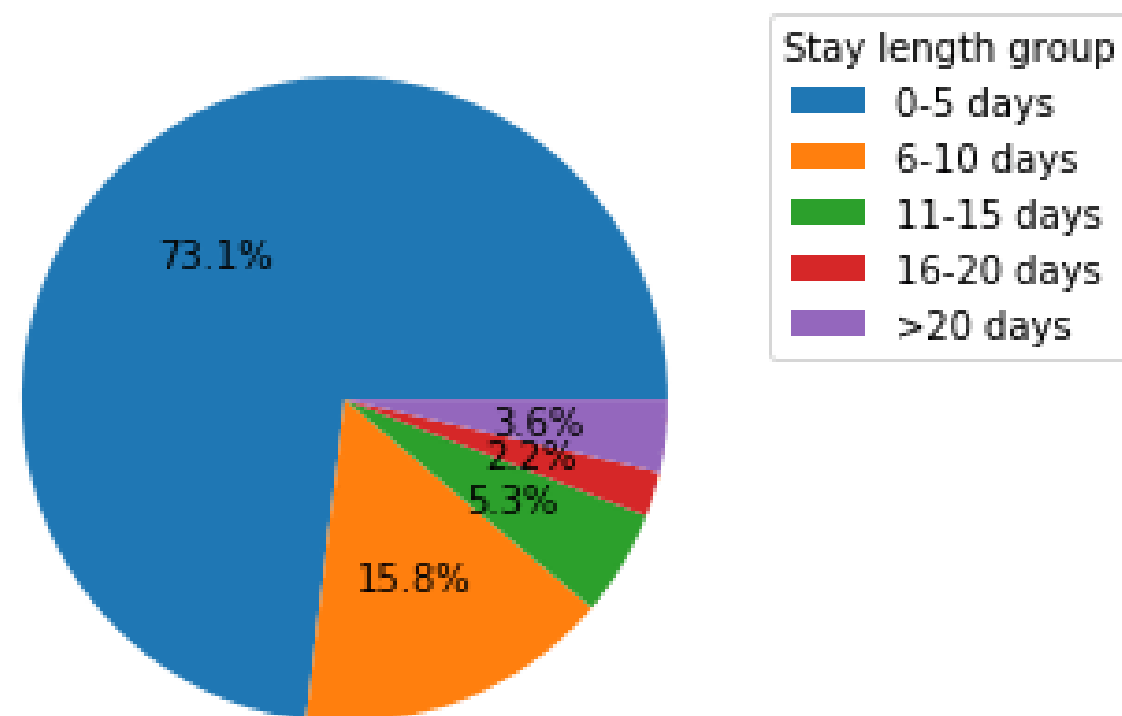


# RESULTS

- Error general de 3.87 días (en media de la estancia de todos los pacientes)
- Error en los pacientes de **0 - 5 dias = 0.522 días**
- Error en los pacientes de **= 0.17 dias**
- Patrón de error: en general, **mas tiempo ingresado = mas error**



Group Proportions



# CONCLUSIONS



# PROPOSALS

- Variabilidad dentro de las categorías - **cada persona es diferente**
- **Complicaciones** que aparecen cuando **pacientes ya están ingresados**

- **Reajustar** agrupaciones
- **Diferentes palabras clave**, siempre mirando error como guía
- Tratar el problema como **multi clasificación**



**THANK YOU!**

Health optimization is our priority