Carlos Cristiano

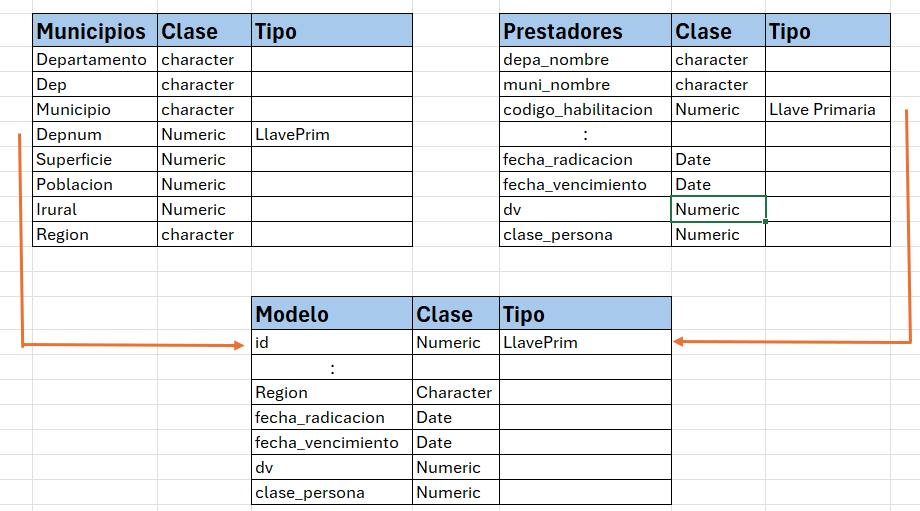
**PRUEBA TECNICA**

**Exploración de los Datos**

La primera fase del análisis consistió en una exploración de los datos provistos en dos conjuntos principales: `Prestadores` y `Municipios`. El conjunto de `Prestadores`, compuesto por 31 variables en formato de texto (tipo `character`), comprende un total de 60946 registros de prestadores de servicios de salud, e incluye detalles significativos como departamento, municipio, nombre, tipo de entidad y datos de contacto. Especial atención se prestó a los campos de fecha (`fecha\_radicacion`, `fecha\_vencimiento` y `fecha\_cierre`), los cuales ameritaban una conversión a formato de fecha adecuado para posibilitar análisis temporales subsecuentes. Por otra parte, `Municipios` se componia de 1118 entidades territoriales, revelando características clave como la superficie, la población, índice de ruralidad y Región.

La base de datos llamada Modelo, es la union de las bases Prestadores y municipios, y esta unida por sus llaves primarias Depnum y codigo\_habilitacion. En el caso de Código habilitación se extrajeron los cinco primeros dígitos los cuales corresponde a Depnum, es decir re realizo un pegue de izquierda con los datos de municipio, esta nueva base de datos tiene los mismo 60946 registros y adicional tiene las variables: Superficie, Población, Virtual y Región.

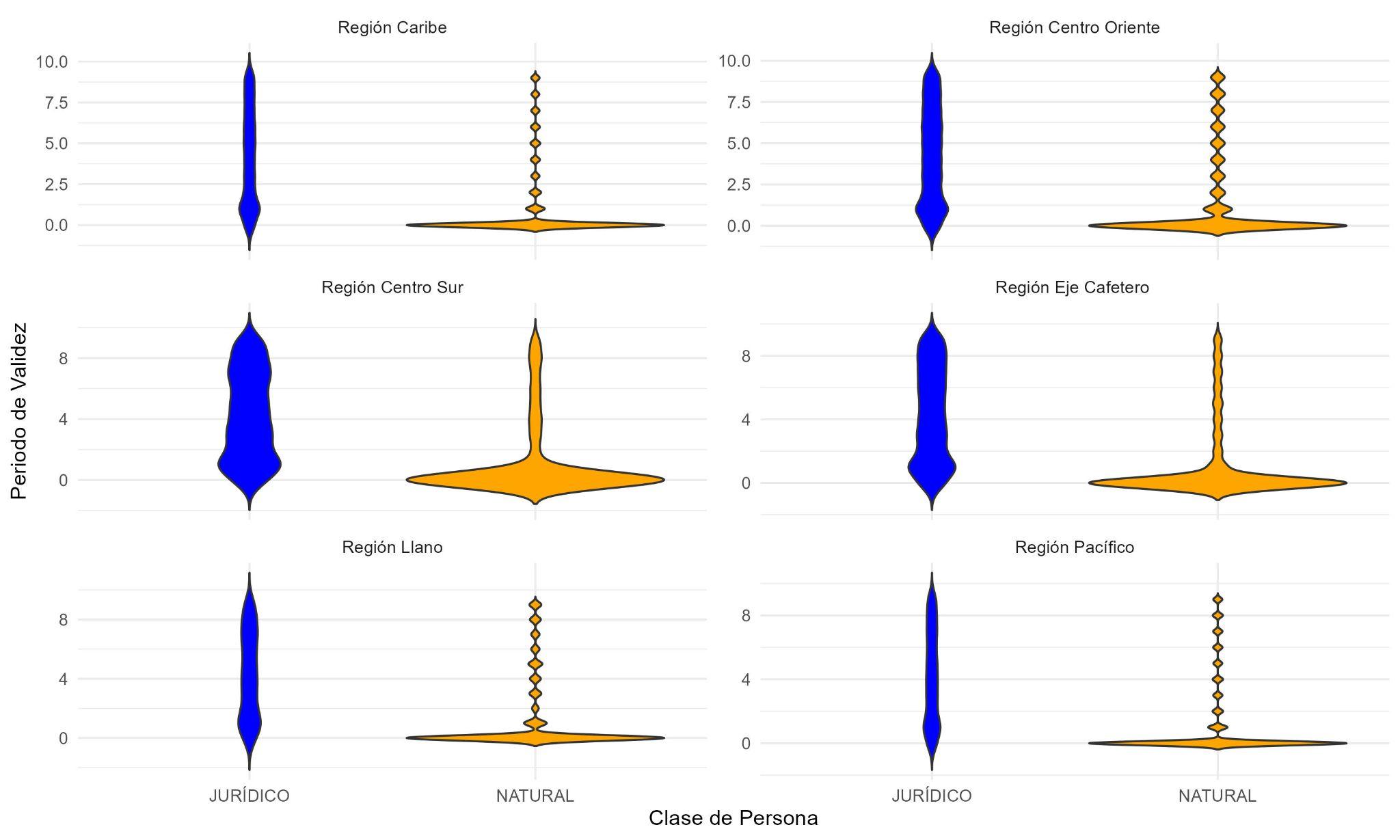
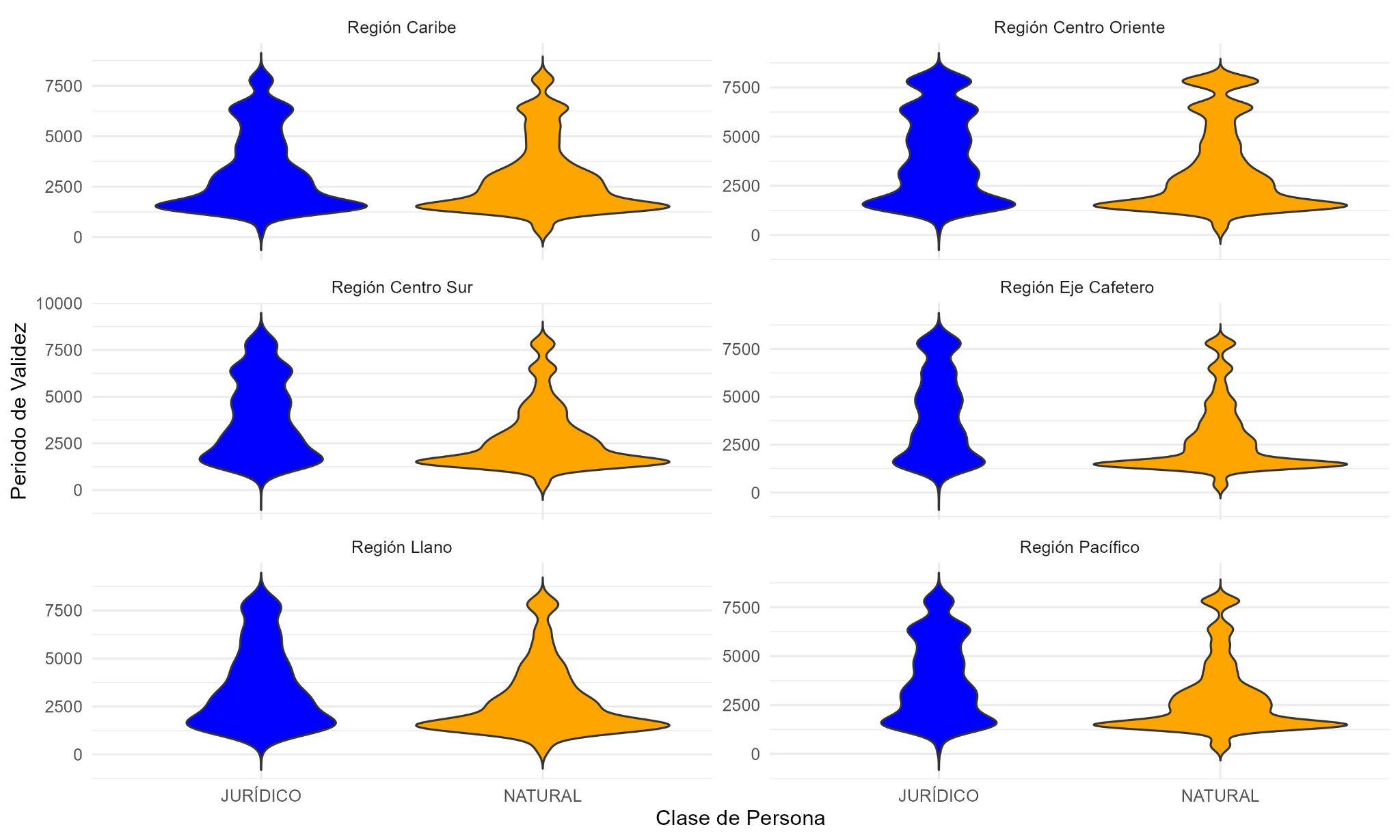
Imagen 1: Exploracion de Datos



La imagen 1 describe la base resultante usada para el análisis de descriptivo y estadístico llevado a cabo.

**Análisis descriptivos**

A través de las técnicas de exploración de datos, se logró distinguir patrones iniciales y distribuciones de interés, lo cual orientó el enfoque de análisis descriptivo. Entre los hallazgos preliminares, se notó se notó una variabilidad substancial en la cantidad de prestadores por región, lo cual sugiere una posible relación entre los factores demográficos y la accesibilidad a servicios de salud. Adicionalmente se creó una variable llamada “periodo de validez” que comprende la diferencia de días entre fecha\_radicacion y fecha\_vencimiento.

Imagen 2: Distribución de días en las diferentes regiones del país 

La imagen 2 superior, muestra un gráfico de violín de la distribución del período de validez de las habilitaciones de los prestadores en cada región, revelando diferencias notables en la duración de las mismas, mientras que la imagen inferior corresponde a la distribución de los

De la variable dv (digito verificación) es cual es un escalar de 1 a 10. Estos resultados indican variaciones regionales en prácticas administrativas y políticas que podrían afectar la operatividad de los prestadores de servicio.

**Modelo Estadístico**

Para investigar los determinantes de la clasificación de los prestadores de servicios de salud en Colombia, optamos por un modelo lineal mixto generalizado (GLMM). Este modelo nos permitió considerar la naturaleza binaria de la variable de respuesta `clase\_persona`, la cual representa la clasificación del prestador como "NATURAL" (1) o "JURIDICO" (0). El data set fue filtrado con valores donde dv era un numero entero y mayor a 0, por lo cual para el análisis estadístico la base tenía un tamaño de 41495 observaciones.

La variabilidad entre las clasificaciones se modeló en función de la variable dv (`digito verificacion`) y la duración de la habilitación (`periodo\_validez`), como efectos fijos que se espera tengan un efecto sistemático en la clasificación del prestador. Además, se incorporó un efecto aleatorio (`Region`) para discernir la heterogeneidad intragrupo ocasionada por diferencias regionales no capturadas por las variables fijas.

El modelo GLMM con familia binomial y enlace logit se eligió debido a su eficacia para datos agrupados y su capacidad para tratar con respuestas dicotómicas. La inclusión de efectos aleatorios atiende la estructura jerárquica de los datos, permitiendo que los interceptos varíen por región y reconociendo que cada una posee atributos únicos que podrían influenciar la clasificación del prestador. Este enfoque proporciona una manera robusta y flexible de evaluar el papel que juegan las variables clave, mientras se controla por posibles efectos confusos debido a la agrupación de datos.

El intercepto y los coeficientes para dv y periodo\_validez resultaron ser estadísticamente significativos, dado que los valores p asociados son menores que el umbral aceptado de 0.05, lo que sugiere una relación importante de estas variables con la probabilidad de ser un prestador de tipo "NATURAL". Además, la inclusión de Región como un efecto aleatorio refleja la variación en la clasificación de clase\_persona entre regiones, capturando la influencia de factores regionales. La variabilidad intra-clase asociada con las regiones destaca la presencia de heterogeneidad entre las áreas geográficas analizadas. Estos resultados brindan una base cuantitativa para entender cómo determinados factores económicos y regionales influyen en la categoría de los prestadores de servicios de salud en Colombia.

**Conclusiones**

El dígito de verificación implica que a medida que este incrementa, la probabilidad de que un prestador sea clasificado como "NATURAL" disminuye, lo que podría interpretarse como que un dígito de verificación más alto está más comúnmente asociado con entidades jurídicas que con individuos.

Según el modelo, el periodo de validez sugiere que cuanto más larga es la duración de la habilitación o licencia otorgada a un prestador de servicios de salud, mayor es la probabilidad de que este sea clasificado como un individuo "NATURAL".

La variabilidad geográfica impacta significativamente la clasificación de los prestadores, lo que sugiere que las políticas de salud y la regulación pueden necesitar ser contextualizadas a las realidades regionales.

Fuente de Repositorio

https://github.com/cdcristbo/ADRES.git