**CONTENIDO**

[**PROGRAMA DE ACUERDOS DIRECTOS DE LA SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO DE COLOMBIA** 3](#_Toc161568524)

[**1.** **CONTEXTO Y PREGUNTAS DE INTERÉS** 3](#_Toc161568525)

[**2.** **EXPLORACIÓN DE DATOS** 6](#_Toc161568526)

[**3.** **LIMPIEZA PRELIMINAR DE LOS DATOS** 8](#_Toc161568527)

[**4.** **CONSTRUCCIÓN DE LA VARIABLE DE RESPUESTA** 9](#_Toc161568528)

[**5.** **SEPARACIÓN DE VARIABLES PARA ENTRENAR EL MODELO** 17](#_Toc161568529)

[Ilustración 1 4](#_Toc161568504)

[Ilustración 2 8](#_Toc161568505)

[Ilustración 3 9](#_Toc161568506)

[Ilustración 4 10](#_Toc161568507)

[Ilustración 5 10](#_Toc161568508)

[Ilustración 6 11](#_Toc161568509)

[Ilustración 7 12](#_Toc161568510)

[Ilustración 8 12](#_Toc161568511)

[Ilustración 9 13](#_Toc161568512)

[Ilustración 10 14](#_Toc161568513)

[Ilustración 11 15](#_Toc161568514)

[Ilustración 12 16](#_Toc161568515)

[Ilustración 13 17](#_Toc161568516)

[Ilustración 14 18](#_Toc161568517)

[Ilustración 15 18](#_Toc161568518)

[Ilustración 16 19](#_Toc161568519)

[Ilustración 17 21](#_Toc161568520)

[Ilustración 18 22](#_Toc161568521)

[Ilustración 19 23](#_Toc161568522)

[Ilustración 20 25](#_Toc161568523)

## **PROGRAMA DE ACUERDOS DIRECTOS DE LA SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO DE COLOMBIA**

## **CONTEXTO Y PREGUNTAS DE INTERÉS**

El proyecto se enmarca en los servicios de atención al ciudadano prestados por la Superintendencia de Industria y Comercio de Colombia.  Uno de los objetivos de esta entidad es informar a los consumidores sobre sus deberes y derechos y asistirlos en el marco del Estatuto del Consumidor (Ley 1480 de 2011) para estos sean respetados.

Dentro de los servicios ofrecidos por esta entidad se encuentra la facilitación de resolución de conflictos entre consumidores y proveedores: “Arreglos directos”. La ciudadanía recurre a esta modalidad para presentar quejas respecto a los bienes o servicios ofrecidos por empresas de carácter formal e informal.

“Todas las diferencias de naturaleza disponible que puedan surgir con ocasión de una relación de consumo particular y concreta pueden ser solucionadas mediante arreglo directo. Los siguientes son algunos de los temas que pueden ser resueltos con este mecanismo:

* Cumplimiento de garantías y de las condiciones del bien o servicio
* Publicidad engañosa en la promoción de un bien o servicio
* Daño en algún bien del consumidor por motivo de la prestación de un servicio ejemplo: servicio de lavandería
* Por la presencia de cláusulas abusivas”

Fuente: <https://www.sic.gov.co/red-nacional-de-proteccion-al-consumidor/servicios/arreglo-directo>

Cuando un ciudadano recurre a esta modalidad se inicia un proceso de duración variada, donde se convocan a ambas partes del conflicto para que asistan voluntariamente y a través la mediación de la entidad se busca que el proceso llegue a un acuerdo entre las partes.

Contamos con un conjunto de datos que incluye los casos interpuestos ante la entidad por la ciudadanía desde el año 2020 a 2024 (Fecha de corte 20/02/2024) **Estos datos son de carácter confidencial y solo se usan con propósito académico sin permiso de publicación.**

Nuestro proyecto busca aplicar modelos predictivos que se entrenen con los datos disponibles para responder un interrogante:

1. Dadas ciertas variables explicativas cómo lo pueden ser la causal, el tipo de producto, etc. ¿cuál es el resultado probable de respuesta. ¿Se llega a un acuerdo? O ¿A cuál de los posibles resultados será el caso adjudicado?

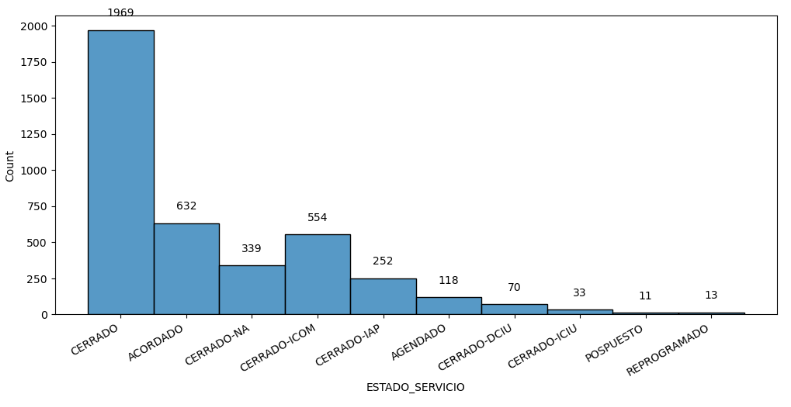


Ilustración 1

1. Dadas ciertas variables explicativas cómo lo pueden ser la causal, el tipo de producto, etc. ¿Cuánto tiempo tardará la entidad en cerrar el caso, más allá del resultado?

**Impacto de predecir cuanto tiempo tardara la Red Nacional de Protección al Consumidor de la SIC en cerrar un caso “Arreglo Directo”**

La pregunta formulada sobre la estimación del tiempo necesario para cerrar un caso en el contexto del Programa de Acuerdos Directos de la Superintendencia de Industria y Comercio de Colombia es de gran relevancia y tiene el potencial de generar un impacto significativo tanto para la entidad como para la ciudadanía. A continuación, se detallan algunas razones que justifican la importancia y el impacto de esta pregunta:

**Para la Entidad (Superintendencia de Industria y Comercio)**

**Mejora de la Eficiencia Operativa**: Al entender qué factores influyen en la duración de los casos, la entidad puede identificar cuellos de botella y áreas de ineficiencia en sus procesos. Esto permite implementar mejoras operativas que aceleren la resolución de casos, optimizando recursos y tiempo.

**Planificación y Gestión de Recursos**: La capacidad de predecir el tiempo de resolución de los casos facilita una mejor planificación y asignación de recursos. Esto es crucial para la gestión de personal, la distribución de la carga de trabajo y la asignación de recursos financieros.

**Mejora de la Calidad del Servicio**: Al reducir los tiempos de espera para los consumidores, la entidad mejora la calidad de su servicio. Esto puede llevar a una mayor satisfacción del usuario y confianza en la entidad, reforzando su imagen como un órgano eficaz y responsable.

**Análisis Predictivo para la Toma de Decisiones**: La capacidad de prever tiempos de resolución ayuda en la toma de decisiones estratégicas, permitiendo a la entidad anticiparse a situaciones de alto volumen de casos o identificar tendencias que requieran atención especial.

**Para la Ciudadanía**

**Transparencia y Previsibilidad**: Proporcionar a los ciudadanos una estimación del tiempo que podría tomar resolver sus casos añade un nivel de transparencia y previsibilidad al proceso. Esto puede reducir la ansiedad y la incertidumbre asociadas a la espera de una resolución.

**Mejora de la Experiencia del Usuario**: La capacidad de obtener una estimación del tiempo de resolución permite a los ciudadanos planificar mejor y gestionar sus expectativas, mejorando su experiencia general con el proceso.

**Fomento de la Participación Ciudadana**: Al mejorar la transparencia y eficiencia del proceso de resolución de casos, se puede fomentar una mayor participación de los ciudadanos en el uso de los mecanismos de protección al consumidor, fortaleciendo así la cultura de los derechos del consumidor.

**Empoderamiento de los Consumidores**: La información sobre los tiempos de resolución empodera a los consumidores para tomar decisiones informadas sobre si iniciar o no un proceso de reclamación, basándose en una comprensión clara del compromiso temporal que esto implica.

**Expectativas Claras**: Proporcionar estimaciones de tiempo ayuda a establecer expectativas claras desde el inicio. Cuando los ciudadanos tienen una idea aproximada de cuánto tiempo llevará resolver sus casos, pueden ajustar sus expectativas y sentirse más seguros sobre el proceso. Esta claridad en las expectativas es fundamental para construir una relación de confianza entre la ciudadanía y la entidad.

**Percepción de Eficacia**: Los tiempos de respuesta rápidos y predecibles son a menudo interpretados por la ciudadanía como un indicador de eficacia y compromiso por parte de la entidad. Una gestión eficiente y transparente de los tiempos de resolución contribuye a una imagen positiva de la Superintendencia, lo cual es esencial para mantener y aumentar la confianza pública en sus servicios.

**Fomento de la Utilización de Servicios**: La confianza en los tiempos de respuesta puede alentar a más ciudadanos a utilizar los servicios de la entidad para resolver sus conflictos de consumo. Saber que sus casos serán atendidos de manera oportuna puede motivar a los consumidores a buscar activamente sus derechos, promoviendo así una mayor justicia y equidad en el mercado.

En resumen, los tiempos de respuesta no solo afectan la eficiencia y percepción del servicio proporcionado por la Superintendencia de Industria y Comercio, sino que también juegan un papel crucial en la construcción y mantenimiento de la confianza de la ciudadanía. Esta confianza es fundamental para asegurar la efectividad de los mecanismos de protección al consumidor y fomentar una cultura de respeto y cumplimiento de los derechos de los consumidores.

## **EXPLORACIÓN DE DATOS**

El conjunto de datos '**ARREGLO\_DIRECTO.csv**' inicialmente contenía **4,049** registros distribuidos en **31** columnas. El diccionario de datos de la fuente original, se muestra a continuación:

Diccionario de Datos

* **FECHA\_REGISTRO**: Fecha y hora en la que se registró el caso o la interacción.
* **PROYECTO\_RNPC**: Identificador o nombre del proyecto o programa bajo el cual se registra el caso.
* **UNIDAD\_DEPARTAMENTO**: Departamento o unidad administrativa donde se originó o registró el caso.
* **DESTINO\_DEPARTAMENTO**: Departamento de destino relacionado con el caso, si aplica.
* **DESTINO\_CIUDAD**: Ciudad de destino o implicada en el caso.
* **ATENCION\_TIPO\_TRAMITE**: Tipo de trámite o servicio solicitado por el ciudadano.
* **ATENCION\_DESTINO**: Entidad o destino hacia donde se dirige la atención o el trámite.
* **ATENCION\_TEMA**: Tema específico de la atención o consulta realizada.
* **ATENCION\_MOTIVO**: Motivo de la atención o consulta.
* **ATENCION\_ESTADO**: Estado actual de la atención (ejemplo,
* **PERSONA\_RANGO\_EDAD**: Rango de edad de la persona que realiza la consulta o reporte.
* **PERSONA\_CONDICION\_DISCAPACIDAD**: Indica si la persona tiene alguna condición de discapacidad.
* **PERSONA\_POBLACION\_ESPECIAL**: Si la persona pertenece a alguna población especial (podría referirse a grupos vulnerables o minoritarios).
* **PERSONA\_GENERO**: Género de la persona.
* **PERSONA\_PROFESION**: Profesión u ocupación de la persona.
* **PERSONA\_ACEPTA\_HABEAS\_DATA**: Indica si la persona acepta el tratamiento de sus datos personales (Habeas Data).
* **TIPO\_EMPRESA**: Clasificación de la empresa involucrada (Formal, Informal).
* **EMPRESA\_TIPO\_DOCUMENTO**: Tipo de documento de identificación de la empresa (NIT, CC, etc.).
* **EMPRESA\_NUMERO\_DOCUMENTO**: Número de documento de identificación de la empresa.
* **EMPRESA\_RAZON\_SOCIAL**: Razón social de la empresa involucrada.
* **ESTADO\_SERVICIO**: Estado actual del servicio o caso (Agendado, Cerrado).
* **CAUSAL**: Causa o razón detrás de la consulta o problema reportado.
* **TIPO\_PRODUCTO**: Tipo de producto implicado en el caso (Bien, Servicio).
* **VALOR\_PRODUCTO**: Valor monetario del producto o servicio implicado.
* **FECHA\_ADQUISICION**: Fecha en la que se adquirió el producto o servicio.
* **FECHA\_ULTIMO\_ENCUENTRO**: Fecha del último encuentro o interacción relacionada con el caso.
* **FECHA\_FIN\_ULTIMO\_ENCUENTRO**: Fecha de finalización del último encuentro, si aplica.
* **FECHA\_FIN\_REAL**: Fecha de finalización del último encuentro, si FECHA\_FIN\_ULTIMO\_ENCUENTRO es vacío entonces toma el valor FECHA\_ULTIMO\_ENCUENTRO.
* **ENCUENTRO\_ACORDADO**: Indica si se acordó algún encuentro o seguimiento.
* **FECHA\_CONTRATO**: Fecha en la que se firmó un contrato relacionado con el caso, si aplica.
* **DURACION**: Cantidad de días que se demora un registro en ser finalizado.

Tras una revisión posterior a la primera entrega, se descubrieron **58** registros que habían sido ingresados como '**pruebas**'. La eliminación de estas observaciones de prueba resultó en una actualización del conjunto de datos, dejándolo con un total de **3,991** registros válidos.

Adicionalmente, se identificó la necesidad de enriquecer el conjunto de datos con tres nuevas columnas para ampliar la información disponible y mejorar el análisis. Estas columnas son:

**PAID**: Identificador del registro.

**UNIDAD\_NOMBRE**: Esta variable hace referencia a la casa del consumidor “Centro de atención de RNPC.

**NOMBRE\_PRODUCTO**: Producto adquirido por el ciudadano que eleva la reclamación ante la SIC.

Estos ajustes aseguran que la base de datos sea más completa y se ajuste mejor a los requisitos del análisis.

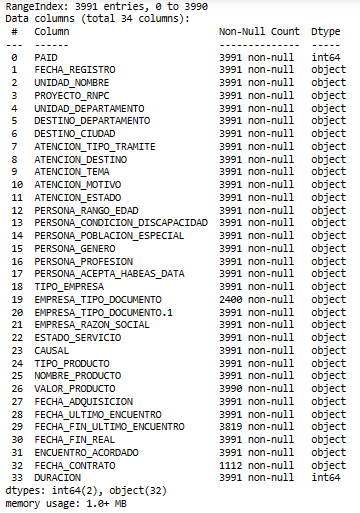


Ilustración 2

## **LIMPIEZA PRELIMINAR DE LOS DATOS**

Para optimizar nuestro modelo predictivo y mejorar la eficiencia computacional, se implementó una estrategia de 'reducción de dimensionalidad'. Este enfoque implica la eliminación deliberada de variables que se consideran redundantes o que aportan poco valor predictivo al modelo. Las variables seleccionadas para su eliminación fueron:

* **ATENCION\_TIPO\_TRAMITE**
* **ATENCION\_ESTADO'**
* **ATENCION\_DESTINO**
* **EMPRESA\_TIPO\_DOCUMENTO**
* **EMPRESA\_RAZON\_SOCIAL**
* **PERSONA\_ACEPTA\_HABEAS\_DATA**
* **FECHA\_ADQUISICION**
* **FECHA\_FIN\_REAL**
* **FECHA\_CONTRATO**
* **DURACION**

Tras esta cuidadosa selección y eliminación de variables, obtenemos un conjunto de datos más adaptable y enfocado, ideal para la construcción de modelos predictivos más precisos y eficientes. A continuación, se presenta el conjunto de datos actualizado tras la aplicación de la reducción de dimensionalidad:

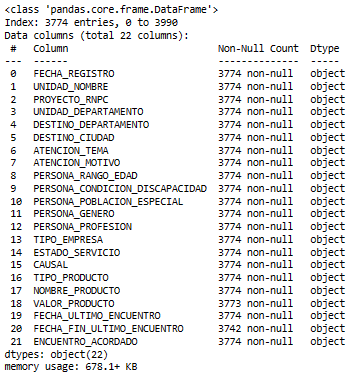


Ilustración 3

## **CONSTRUCCIÓN DE LA VARIABLE DE RESPUESTA**

La construcción del modelo predictivo, orientado a predecir el tiempo de resolución de un arreglo directo, requiere el desarrollo de una variable específica que represente esta duración. En este sentido, la variable diseñada para calcular el número de días que transcurren para resolver un caso se fundamenta en la selección de la fecha más reciente entre las variables **FECHA\_ULTIMO\_ENCUENTRO** y **FECHA\_FIN\_ULTIMO\_ENCUENTRO**, la cual se registra en la nueva variable denominada **FECHA\_ULTIMA**.

Posteriormente, para determinar la duración del proceso, se realiza una operación de sustracción entre la variable **FECHA\_ULTIMA** y **FECHA\_REGISTRO**, obteniendo así el tiempo total invertido en el caso. Este resultado se almacena en la variable **DURACION\_PROCESO**, la cual es clave para el análisis predictivo que busca estimar el período necesario para llegar a un acuerdo en el marco de un arreglo directo.

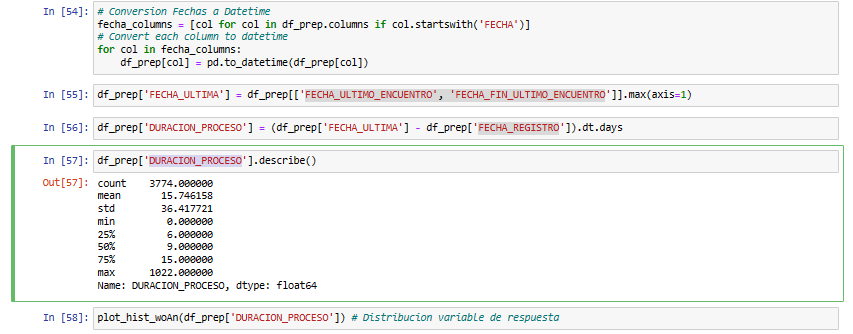


Ilustración 4

**AGRUPANDO VALORES PARA HACER ANALISIS**

En esta parte de preparación de los datos. Se realiza una categorización basada en rangos predefinidos de días y almacenamos estas categorías en una nueva columna '**DURACION\_DSC**'. Esto es útil para análisis que requieren agrupar los datos en rangos específicos de duración, facilitando la visualización, la interpretación y el análisis estadístico.:

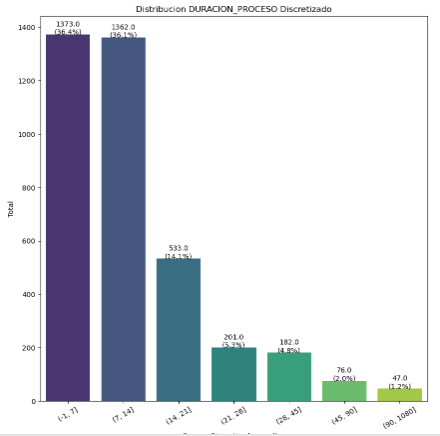


Ilustración 5

**COMPORTAMIENTO DE LA VARIABLE DE RESPUESTA Vs VARIABLES PREDICTORAS**

Una vez creada la variable de respuesta DURACION\_ACO con sus distintas transformaciones, podemos hacer análisis más detallados frente a su comportamiento con otras variables del predictoras del dataframe. En la gráfica se muestra a continuación se hace un análisis de con la variable **CAUSAL**:

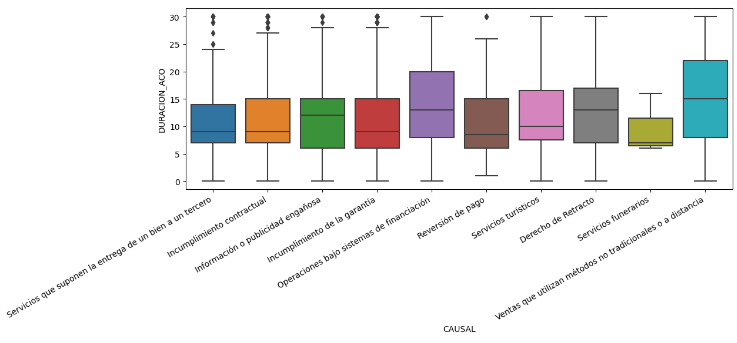


Ilustración 6

A partir de esta gráfica, se pueden realizar varias observaciones:

**Variabilidad de la Duración**: Existe variabilidad en los tiempos de duración ('DURACION\_ACO') entre las diferentes causales. Esto sugiere que la causa del arreglo directo podría influir en el tiempo que toma resolver el caso.

**Medianas y Rangos**: Algunas categorías muestran medianas más altas que otras, y los rangos (las diferencias entre los cuartiles superior e inferior) también varían. Esto indica que ciertas causales pueden, en promedio, tomar más o menos tiempo en encontrar una resolución.

**Outliers o Valores Atípicos**: Hay presencia de valores atípicos en varias categorías, como se evidencia por los puntos que se encuentran fuera del rango de las cajas. Estos outliers pueden indicar casos que son excepcionalmente largos en su duración, lo que podría deberse a complicaciones o características únicas de esos casos en particular.

**Comparación entre Causales**: Algunas categorías, como "**Revisión de pago**" y "**Derecho de Retracto**", parecen tener una menor mediana y un rango intercuartil más compacto en comparación con categorías como "**Incumplimiento contractual**" y "**Servicios suponen la entrega de bien en un tercero**", que muestran una mayor dispersión y mediana más alta. Esto podría interpretarse como que ciertos tipos de causales son resueltos más rápidamente que otros.

En conclusión, la gráfica indica que la duración para llegar a un acuerdo varía significativamente dependiendo de la causal del arreglo directo. La entidad podría usar estos insights para mejorar la eficiencia de sus procesos y para establecer expectativas realistas con los ciudadanos sobre los tiempos de resolución esperados.

**PREPARACIÓN DE LOS DATOS**

Una de las variables que puede ser de alto valor para el análisis es VALOR\_DEL\_PRODUCTO, se tiene una particularidad, pues dicha variable es categórica, pero optamos por volverla numérica.

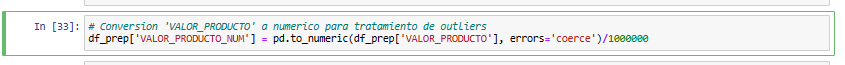


Ilustración 7

Así las cosas, con esta transformación podemos hacer mejor el tratamiento de los outliers.

Con relación a esta variable, identificamos que existen valores categorizados como “INDETERMINADO”. Sin embargo, para dar un adecuado tratamiento a esta variable, en conjunto con el cliente, decidimos hacer una imputación de sobre esta variable, convirtiendo los registros cuyo valor es “INDEPERMINADO” por NaN.

Al igual que la variable XXX realizamos un agrupamiento de los valores, dejándolos de la siguiente forma:

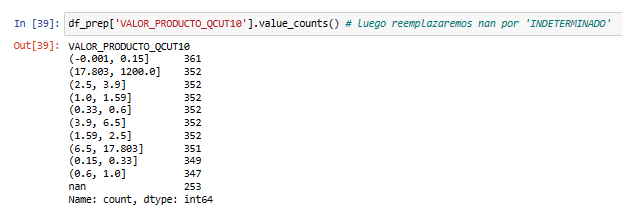


Ilustración 8

El grafico que se muestra a continuación, representa la distribución de una variable de valor del producto, la cual ha sido discretizada en rangos de precios. El eje x muestra los rangos de valor del producto en moneda colombiana (COP), y el eje y muestra el total de productos que caen en cada uno de estos rangos.

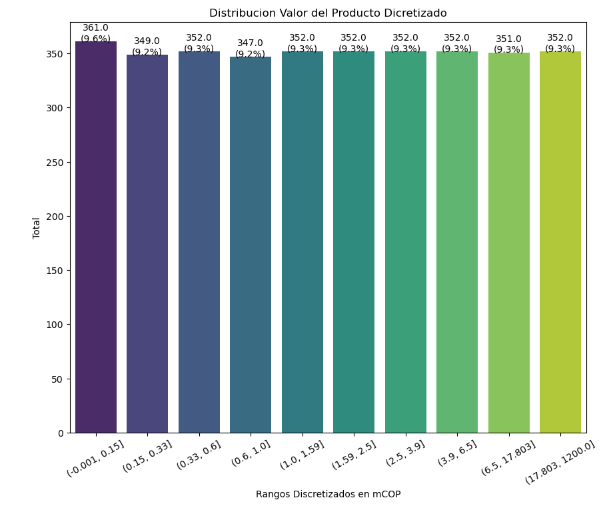


Ilustración 9

Observaciones clave:

**Distribución Uniforme**: Los rangos discretizados tienen una cantidad muy similar de productos, lo que sugiere una distribución bastante uniforme. Cada barra representa un rango de valor y contiene aproximadamente la misma cantidad de productos, indicando que no hay una concentración notable de productos en un rango de precio específico.

**Etiquetas de Rango de Precios**: Los rangos están etiquetados con los valores mínimos y máximos de cada uno y están visualmente codificados por colores para una diferenciación clara.

**Frecuencia y Porcentaje**: Cada barra proporciona dos números: el total de productos en ese rango (parte superior de la barra) y el porcentaje que representan del total (entre paréntesis). Esto da una doble perspectiva sobre el conteo absoluto y su proporción relativa.

**No hay Outliers Visibles**: Dado que todos los rangos tienen una representación similar de productos, no hay indicaciones visibles de outliers o de rangos de precio que sean inusuales en comparación con los demás.

**Uniformidad en la Distribución de Precios**: La uniformidad sugiere que los productos están distribuidos de manera equitativa a través de los distintos rangos de precios, lo cual podría indicar una estrategia de precios que apunta a cubrir un mercado amplio o una diversificación en la oferta de productos.

En resumen, la gráfica muestra una distribución equitativa de productos a lo largo de varios rangos de precios discretizados, sin ninguna concentración notable en un rango específico. Para interpretaciones más profundas y acciones basadas en esta información, sería importante considerar el contexto de los datos, como el tipo de producto y la estrategia de mercado de la entidad.

Por último, con relación a esta variable se ha podido identificar que existen errores con el valor de algunos productos y existen registros con los mismos valores, lo cual sugiere que puedan ser registros duplicados.

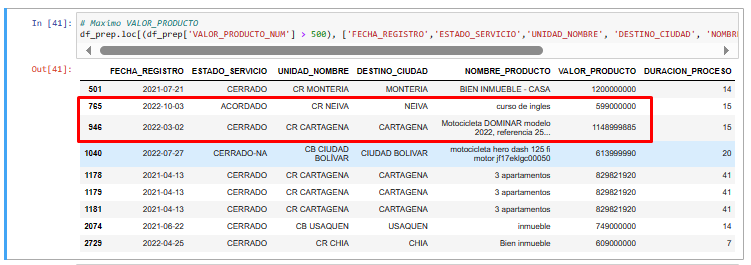


Ilustración 10

**PRESELECCIÓN DE VARIABLES**

Se ha realizado la selección de las variables, en la cual se descartaron varias de ellas, pues en análisis realizado con anterioridad, pudimos percibir que las mismas no generan mayor valor al modelo.

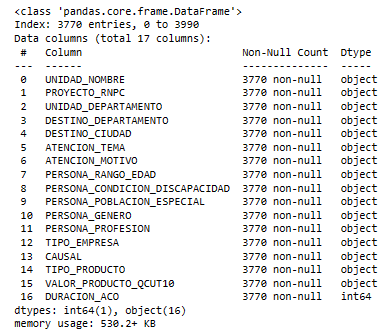


Ilustración 11

**ANALISIS DE LA VARIABLE PERSONA\_RANGO\_EDAD**

La gráfica de barras representa la distribución de la cantidad de personas agrupadas por rangos de edad. Los datos indican que la mayoría de las personas se encuentran en los rangos de 41-59 años y 26-40 años, con 1550 y 1279 personas respectivamente, lo que sugiere que estos son los grupos de edad más comunes en el conjunto de datos. El grupo de 60-80 años tiene una presencia considerablemente menor con 688 personas, y la cantidad disminuye significativamente en los rangos de edad más jóvenes y más viejos, con solo 229 personas de 19-25 años, 9 de 1-18 años y 15 de más de 81 años.

El perfil de edad predominante parece ser de adultos en etapas medianas de la vida, lo que podría indicar que estas personas están más activas en el contexto de acudir a la SIC para hacer una petición, por un Arreglo Directo. La presencia reducida de jóvenes y ancianos podría sugerir que estos grupos son menos activos o están menos representados en el conjunto de datos. Esta información puede ser relevante para la toma de decisiones o la planificación de servicios dirigidos a grupos de edad específicos.

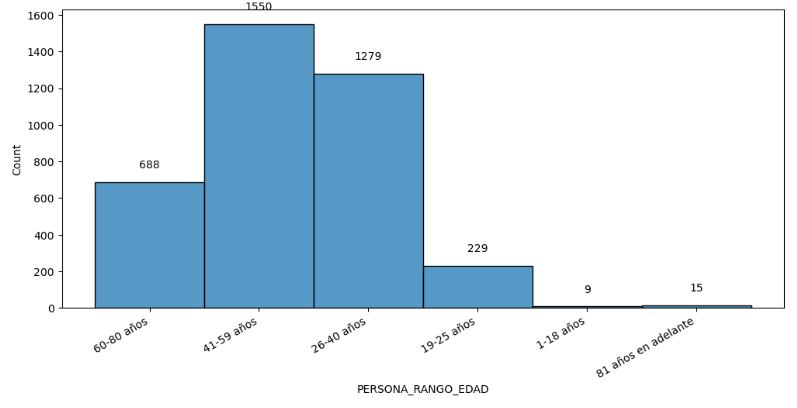


Ilustración 12

**EXPLORACIÓN DE LA VARIABLE RANGO EDAD**

La gráfica tipo boxplot que se relaciona a continuación muestra la distribución de la variable 'PERSONA\_RANGO\_EDAD', la cual representa la duración del acuerdo directo, para diferentes rangos de edad. Aquí están algunas interpretaciones que se pueden hacer basadas en un boxplot típico:

**Distribución por Edad**: La variable de duración se ha dividido entre distintos rangos de edad, lo que sugiere que se está intentando analizar si la edad tiene algún impacto en el tiempo que toma resolver un acuerdo directo.

**Mediana**: La línea horizontal dentro de cada caja indica la mediana de la duración para cada grupo de edad. La posición de la mediana proporciona una indicación de la tendencia central de los datos de duración dentro de cada grupo de edad.

**Rango Intercuartil (IQR)**: La altura de cada caja, que va desde el primer cuartil (Q1) al tercer cuartil (Q3), representa el rango intercuartil. Un IQR más amplio significa que hay una mayor dispersión en la duración de los acuerdos dentro de ese rango de edad.

**Valores Atípicos (Outliers)**: Los puntos individuales que se ven por encima y por debajo de los 'bigotes' (las líneas que se extienden desde la caja) son valores atípicos. Estos representan casos que están fuera del rango típico y pueden ser anomalías o casos especiales que podrían requerir una atención más detallada.

**Comparación entre Grupos de Edad**: Se puede observar cómo se compara la mediana y la dispersión de la duración de los acuerdos directos entre los diferentes rangos de edad. Por ejemplo, si algunas cajas tienen medianas significativamente más altas o más bajas que otras, o si algunos grupos de edad tienen más outliers que otros.

**Anomalías Específicas**: El grupo de 1-18 años parece tener una mediana de duración más alta que los otros grupos, lo que podría sugerir que los acuerdos involucrando a personas más jóvenes toman más tiempo en promedio. Además, este grupo tiene una variabilidad relativamente alta en la duración y una cantidad significativa de outliers, indicando que hay casos que se desvían considerablemente de la mediana.

**Datos Consistentes**: Los rangos de edad de adultos medios parecen tener medianas relativamente consistentes entre sí, aunque con diferente cantidad de outliers, lo que sugiere que la edad, en estos rangos, no tiene un gran impacto en la duración del acuerdo directo.

## **SEPARACIÓN DE VARIABLES PARA ENTRENAR EL MODELO**

El subconjunto que se relaciona, está destinado a ser el conjunto de variables independientes o explicativas que se usarán para entrenar un modelo de aprendizaje automático.

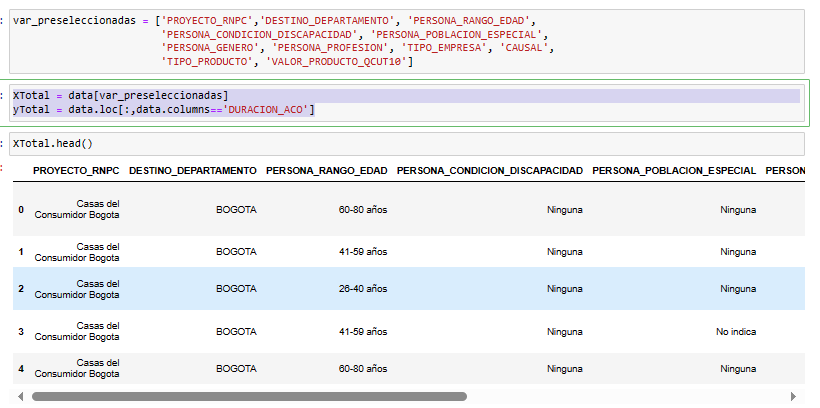


Ilustración 13

Para continuar con la exploración eliminaremos del set de datos algunas variables que no consideramos relevantes, cómo las variables de fecha o variables donde solo existe una categoría. En el anexo se pueden observar los histogramas de todas las variables.

La variable valor del producto la ajustamos para que exprese el valor en millones de pesos colombianos.

**HISTOGRAMAS DE LAS VARIABLES:**

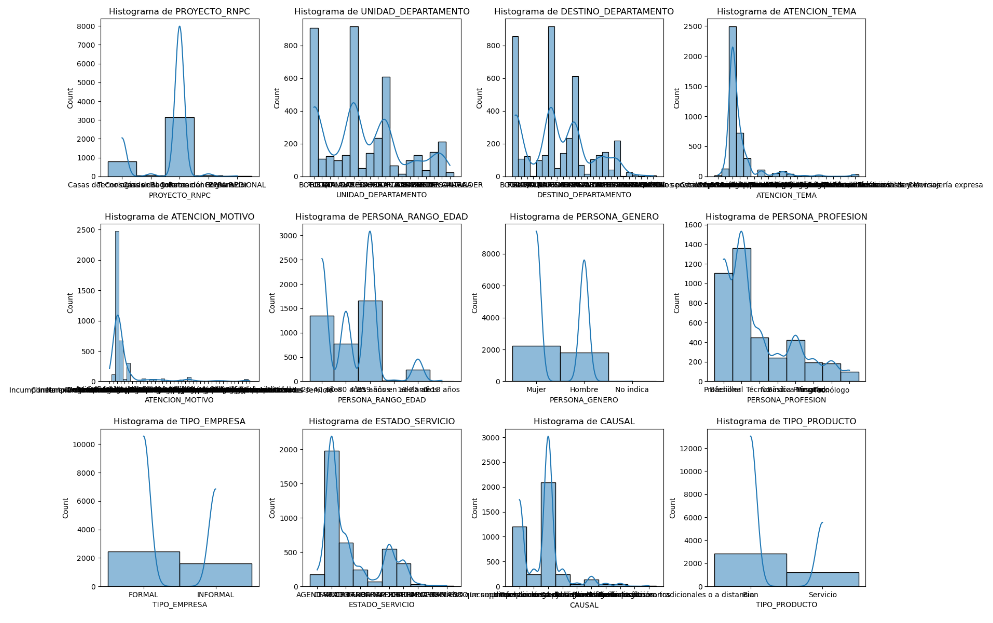


Ilustración 14

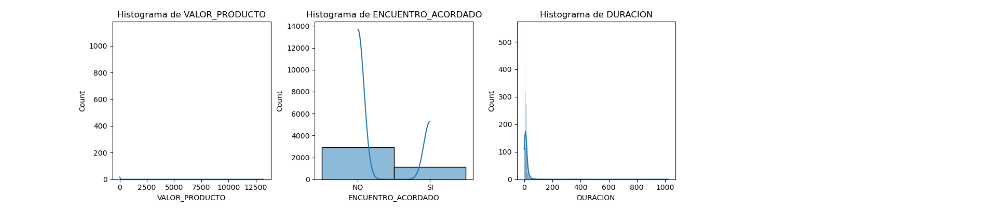
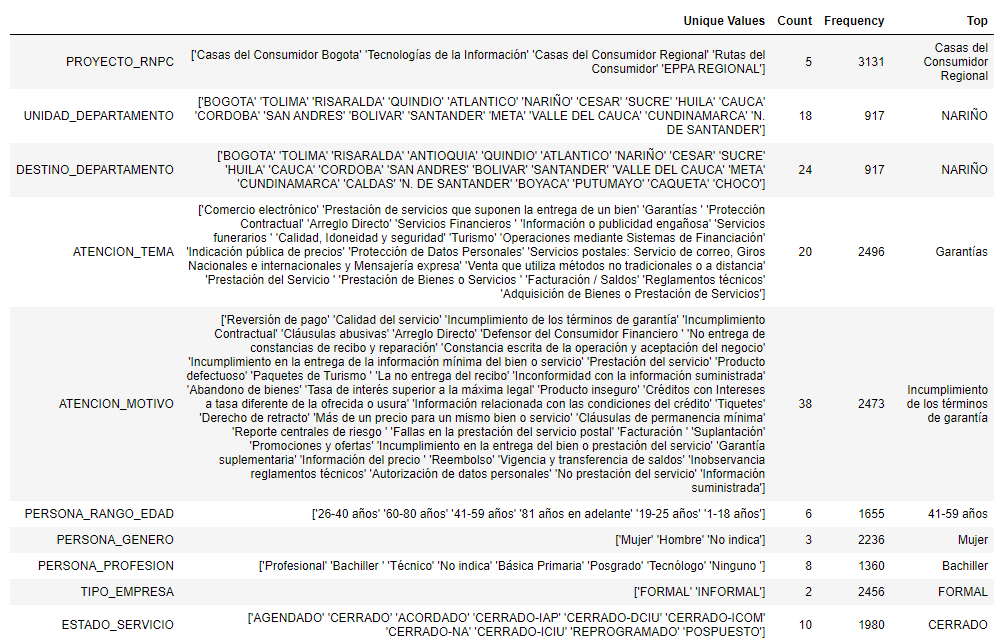


Ilustración 15

En los histogramas presentados antes se observa un desbalance entre los valores únicos en muchas variables incluidas en el análisis.

Para complementar los histogramas tenemos la siguiente tabla donde podemos ver la cantidad de Valores únicos en cada categoría, el valor con mayor frecuencia y la cantidad de observaciones en dicho valor:



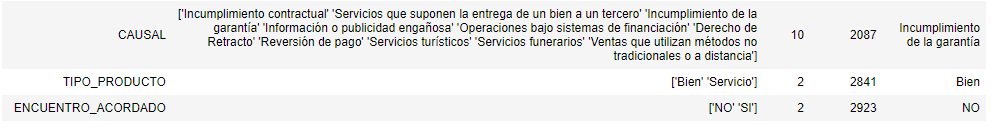


Ilustración 16

**DISTRIBUCIÓN DE FRECUENCIAS PARA LA VARIABLE ESTADO\_SERVICIO**

El gráfico que se anexa a continuación, muestra la distribución de frecuencias para diferentes categorías de la variable '**ESTADO\_SERVICIO**'.

**AGENDADO**: Esta categoría tiene la frecuencia más alta, lo que indica que es el estado más común en el conjunto de datos para los servicios. Esto sugiere que hay muchos servicios que están programados para ser realizados.

**CERRADO' y 'ACORDADO**: Estas categorías tienen frecuencias menores en comparación con '**AGENDADO**', lo cual indica que hay menos servicios que han llegado a estos estados.

**CERRADO/APO', 'CERRADO/ARCU', 'CERRADO/COM', 'CERRADO/VA', y 'REPROGRAMADO**: Estas categorías tienen frecuencias más bajas aún, lo que podría sugerir que son estados finales o que reflejan acciones específicas dentro del proceso de servicio, y que ocurren con menos frecuencia que el agendamiento de servicios.

**PROPUESTO**: Tiene la frecuencia más baja entre las categorías mostradas en el histograma, lo que podría indicar que esta etapa del proceso de servicio es menos común en comparación con las otras categorías.

Este histograma puede ser útil para entender cómo se distribuyen los estados de servicio y podría proporcionar insights para mejorar la eficiencia en la gestión de servicios, al identificar los estados que requieren atención o ajustes en el proceso.

**IDENTIFICANCO CORRELACIONES ENTRE VARIABLES**

El mapa de calor muestra la relación entre dos variables categóricas: **ESTADO\_SERVICIO** y '**PROYECTO\_RNPC**'. Los colores más oscuros o más intensos representan frecuencias más altas de ocurrencias, mientras que los colores más claros indican frecuencias más bajas.

Hallazgos relevantes:

La combinación **CERRADO** en **Casas del Consumidor Regional** tiene una frecuencia muy alta (1498), lo que sugiere que muchos servicios en esta categoría y proyecto han sido finalizados.

**ACORDADO** tiene una alta frecuencia (482) en 'Casas del Consumidor Bogotá', indicando que se alcanzaron muchos acuerdos en este lugar.

Algunas combinaciones tienen valores muy bajos o ceros (indicados por colores más claros o ausencia de color), lo que significa que ciertos estados de servicio ocurren raramente o nunca en algunos proyectos.

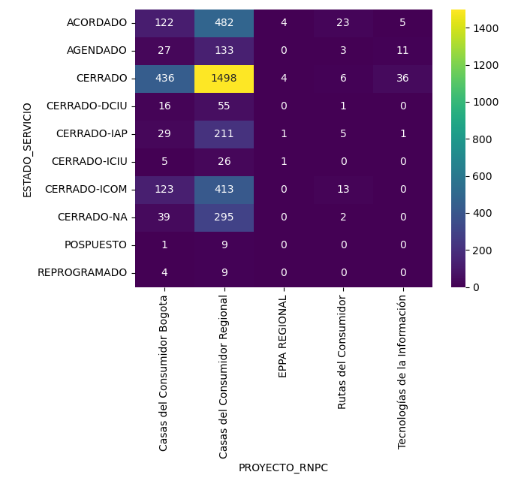


Ilustración 17

**ESTADO\_SERVICIO Vs CAUSAL**

Las áreas de los rectángulos en el gráfico son proporcionales a la frecuencia de las combinaciones de las categorías de estas variables.

Para el caso de los valores **CERRADO** y **Protección al Consumidor**, eso significa que hay una alta frecuencia de casos cerrados asociados con esa causal particular.

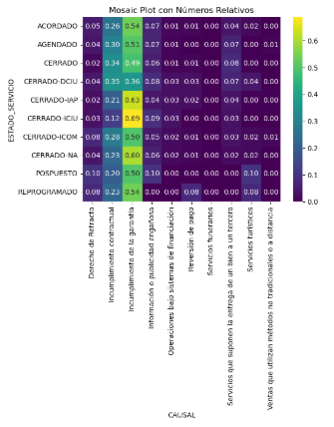


Ilustración 18

**ESTADO\_SERVICIO Vs RANGO\_EDAD**

La mayoría de las observaciones se encuentran en los estados **ACORDADO** (36%) y **CERRADO** (33%). Los estados REPROGRAMADO (8%) y POSPUESTO (0%) son los menos frecuentes.

* El estado "ACORDADO" es más frecuente en el rango de edad "26-40 años".
* El estado "CERRADO" es más frecuente en el rango de edad "41-59 años".
* El estado "REPROGRAMADO" es más frecuente en el rango de edad "19-25 años".
* El estado "POSPUESTO" es más frecuente en el rango de edad "60-80 años".

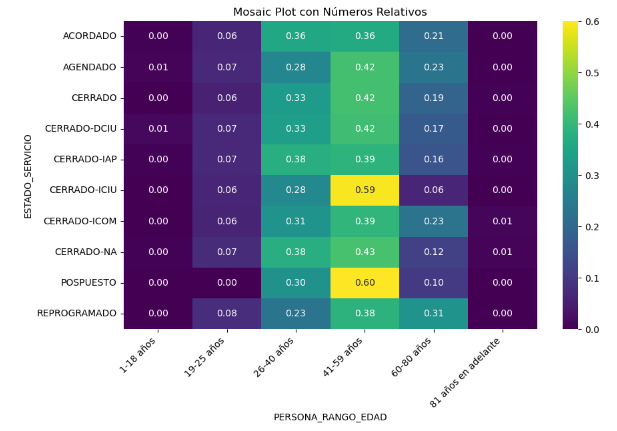


Ilustración 19

**ESTADO SERVICIO Vs ENCUENTRO\_ACORDADO**

A partir de la información del gráfico, podemos realizar las siguientes observaciones:

1. Distribución de la variable "**ESTADO\_SERVICIO**":

* El estado más frecuente es **ACORDADO** (**98%**).
* El estado **AGENDADO** es el menos frecuente (**2%**).
* Los estados **CERRADO**, **CERRADO-DCIU**, **CERRADO IAP**, **CERRADO-ICIU-**, **CERRADO-NA**, **POSPUESTO** y **REPROGRAMADO** tienen una frecuencia baja (entre **0% y 0.8%**).

2. Análisis detallado de los estados con baja frecuencia:

El estado **CERRADO** tiene una frecuencia del **75%**, pero también presenta una desviación estándar del **25%**. Esto indica que hay una variabilidad considerable en la cantidad de observaciones dentro de este estado.

Los estados **CERRADO-DCIU**, **CERRADO IAP**, **CERRADO-ICIU-**, **CERRADO-NA**, **POSPUESTO** y **REPROGRAMADO** tienen una frecuencia muy baja, lo que podría indicar que son estados poco comunes o que hay errores en la recolección de datos.

En general, el gráfico muestra que:

* La variable **ESTADO\_SERVICIO** tiene una distribución no uniforme.
* El estado **ACORDADO** es el más común, mientras que los demás estados tienen una frecuencia baja o muy baja.

**ESTADO\_SERVICIO Vs PERSONA\_GENERO**

El gráfico de barras que se muestra, contiene la distribución de una variable categórica:

Variable: "**GENERO**"

Cada barra del gráfico representa un género posible:

* Hombre
* Mujer
* No indica

La altura de cada barra indica el número de observaciones que pertenecen a ese género.

En este caso, la escala del eje y va de 0 a 2000, lo que indica que se representa el número de observaciones para cada género.

A partir de la información del gráfico, podemos realizar las siguientes observaciones:

1. Distribución de la variable **GENERO**:

* El género más frecuente es "Hombre" (1750).
* El género "Mujer" tiene una frecuencia considerable (1250).
* El género "No indica" tiene la frecuencia más baja (250).

1. Análisis detallado de la distribución:

* La diferencia entre la frecuencia de **Hombre** y **Mujer** es de 500 observaciones.
* La frecuencia de "No indica" es considerablemente menor que la de los otros dos géneros.

1. Consideraciones adicionales:

Es importante tener en cuenta el contexto del problema para interpretar la distribución de la variable **GENERO**.

La frecuencia de "**No indica**" podría deberse a diferentes razones, como la falta de información o la no identificación con ninguno de los dos géneros.

En general, el gráfico muestra que:

* La variable **GENERO** tiene una distribución no uniforme.
* El género **Hombre** es el más frecuente, seguido del género **Mujer**.
* El género **No indica** tiene la frecuencia más baja.

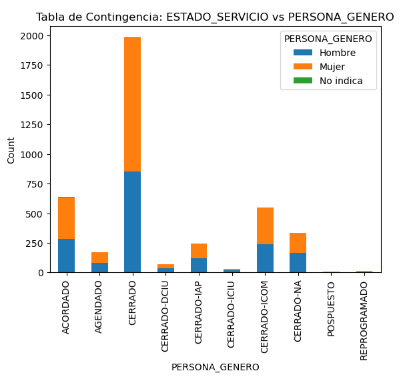


Ilustración 20