

**Análisis de comportamiento operacional en agente de *call center*  
empleando inteligencia artificial**

**Camilo Andrés Pérez Ospino**

**Trabajo Dirigido**

**Tutor**

**D.Sc Alvaro David Orjuela Canon**



UNIVERSIDAD DEL ROSARIO  
ESCUELA COLOMBIANA DE INGENIERÍA JULIO GARAVITO  
PROGRAMA DE INGENIERÍA BIOMÉDICA  
BOGOTÁ D.C  
2022

## Agradecimientos

En esta sección quisiera dedicárselo a mi familia quienes han sido mi motor y mi apoyo para hacer mis sueños realidad, desde el primer momento que decidí estudiar Ingeniería Biomédica, todos han estado involucrados en mi crecimiento personal y profesional, dando apoyo, consejos y hasta correcciones en mis entregas. Mis padres son personas muy fuertes y honestas quienes me han dado los valores y la rectitud para cumplir con este objetivo, mi hermana quien es la persona que me acompaña en todos en los sentidos y es mi mayor soporte.

Adicional, aparte de agradecerles a los ingenieros y profesores que hacen parte de las instituciones de la Universidad del Rosario y la Escuela Colombiana de Ingeniería Julio Garavito, por impartir sus conocimientos y experiencias dentro y fuera de las aulas, con la paciencia y el buen criterio que todos como estudiantes reconocemos de ellos. De los cuales me gustaría mencionar a el ingeniero Alvaro Orjula quien es mi tutor en este trabajo y además una gran persona con quien he compartido diferentes espacios como lo son el semillero de investigación SemillIAS y la ingeniera Sanadra Cancino, quien fue mi tutora de énfasis en procesamiento de señales e imágenes médicas, quien además de compartir espacios de investigación en el semillero de PROMISE es la persona que me acercó a Millenium BPO para realizar mis prácticas profesionales.

## Resumen

Este proyecto de investigación trata de solventar una problemática que afrontan hoy en día las empresas de *call center*, la cual es la rotación de sus empleados, dejando brechas operativas y económicas en la compañía.

Desde la división de innovación de Millenium BPO, Datos No Estructuras. Se propone actuar de manera activa ante esta situación por lo que se propone el uso de la analítica de datos y la correlación de variables en sus asesores que demuestren declive en su comportamiento operacional y de estos mediante el uso de modelos de inteligencia artificial predecir la renuncia psicológica, término acuñado para describir a los empleados que antes de su renuncia presentan patrones de comportamiento que van en contra vía de lo esperado por la empresa, como lo puede ser una baja productividad, no relacionarse con sus compañeros, pérdida en la calidad de su trabajo, entre otros. Por diferentes motivos, como lo son la insatisfacción laboral, altas metas operativas, poco crecimiento dentro de la empresa, entre otros.

Por ello, esta investigación está enfocada en la para la extracción y transformación de datos con el fin de describir de patrones operaciones de asesores de un *call center* junto con la implementación de datos socio-demográficos en asesores activos y retirados dentro de la campaña de Icetex durante 12 semanas consecutivas laborales y con estos datos entrenar diferentes modelos de inteligencia artificial para predicción de la renuncia.

Los resultados de esta investigación demuestran una diferencia significativa en el rendimiento de las personas que estaban a punto de renunciar y los asesores activos, en términos de la productividad, eficiencia, ausencia en el trabajo y calidad de gestión ante el usuario, otros. Por lo que se expone el comportamiento característico de la renuncia psicológica en los asesores antes de su retiro voluntario.

Se realizó un entrenamiento de aprendizaje de máquina para la predicción de un asesor en los estados de activo y retirado, donde se presenta que el modelo de bosques aleatorios (*random forest*) es el mejor con un  $89,5\% \pm 4,3\%$  de exactitud en las pruebas realizadas en 58 asesores de *call center*.

# Índice general

Agradecimientos . . . . .	I
Resumen . . . . .	II
<b>1. INTRODUCCIÓN</b>	<b>1</b>
1.1. Planteamiento del problema . . . . .	2
1.2. Antecedentes . . . . .	3
<b>2. OBJETIVOS</b>	<b>5</b>
2.1. General . . . . .	5
2.2. Específicos . . . . .	5
<b>3. MARCO TEÓRICO</b>	<b>6</b>
3.1. <i>Business Process Outsourcing (BPO)</i> . . . . .	6
3.2. Indicadores de rendimiento . . . . .	9
3.3. Exploración de datos . . . . .	10
3.4. Inteligencia artificial . . . . .	12
<b>4. METODOLOGÍA</b>	<b>18</b>
4.1. Bases de datos . . . . .	18
4.1.1. Bases de datos demográficas . . . . .	18
4.1.2. Bases de datos operación . . . . .	20
4.1.3. Selección de participantes . . . . .	21
4.2. Correlación datos con estados . . . . .	22
4.2.1. Análisis exploratorio de los datos . . . . .	22
4.2.2. Exploración aprendizaje no supervisado . . . . .	22
4.3. Modelo de predicción de renuncia con aprendizaje supervisado . . . . .	22
<b>5. RESULTADOS Y DISCUSIÓN</b>	<b>24</b>
5.1. Bases de datos . . . . .	24
5.1.1. Bases de datos demográficas . . . . .	24
5.1.2. Bases de datos operación . . . . .	29
5.1.3. Matriz resultante . . . . .	37
5.2. Correlación de datos . . . . .	37
5.2.1. Análisis exploratorio de los datos . . . . .	37
5.2.2. Exploración con Inteligencia Artificial . . . . .	41
5.3. Modelos de predicción . . . . .	44

<b>6. CONCLUSIONES</b>	<b>47</b>
<b>7. RECOMENDACIONES Y TRABAJOS FUTUROS</b>	<b>48</b>
<b>BIBLIOGRAFÍA</b>	<b>49</b>
<b>ANEXO</b>	<b>52</b>

# Índice de figuras

3.1. Ejemplos de las actividades que se realizan en una BPO. Extraído de [23] . . . . .	6
3.2. Representación jerarquía operación de un call center y representación de lazo de control . . . . .	8
3.3. Vistas CRM Icetex . . . . .	9
3.4. Descripción partes diagrama de caja. Extraído de [28] . . . . .	10
3.5. Ejemplo de estimación densidad del núcleo. Extraído de [31] . . . . .	11
3.6. Descripción diagrama de violín. Extraído de [33] . . . . .	11
3.7. Diagrama estrategias del <i>Machine Learning</i> . Extraído de [34] . . . . .	12
3.8. Agrupamiento con <i>Kmeans</i> [35] . . . . .	13
3.9. Arquitectura de mapas auto-organizados con redes neuronales. Extraído de [36] . . . . .	14
3.10. Arquitectura de máquinas de vectores de soporte. Extraído de [39] . . . . .	14
3.11. Arquitectura de bosques aleatorios. Extraído de [41] . . . . .	15
3.12. Arquitectura de redes neuronales artificiales. Extraído de [44] . . . . .	15
3.13. Matriz de confusión. Extraído de [46] . . . . .	16
4.1. Diagrama de bloques pre-procesamiento: Base de datos demográficos . . . . .	19
4.2. Diagrama de bloques pre-procesamiento: Base de datos operación . . . . .	21
5.1. Distribución edad de los agentes por genero . . . . .	25
5.2. Cantidad de agentes por estrato . . . . .	25
5.3. Distribución de agentes en Bogotá . . . . .	26
5.4. Cantidad de agentes según su estado . . . . .	27
5.5. Densidad del núcleo . . . . .	28
5.6. Distribución variables categóricas . . . . .	28
5.7. Retiros durante el año . . . . .	29
5.8. Productividad promedio hora en Enero 2021 . . . . .	30
5.9. Diagrama de violín: Gestión agentes por hora . . . . .	31
5.10. Hora del día y tráfico en el canal de chat . . . . .	32
5.11. Diagrama de violín: TMO agentes por hora . . . . .	34
5.12. Diagrama de violín: Gestión agentes por día . . . . .	35
5.13. Diagrama de violín: TMO agentes por día . . . . .	36
5.14. Comparación en productividades . . . . .	38
5.15. Total de interacciones . . . . .	38
5.16. Total horas de trabajo por semana . . . . .	39
5.17. Ausentismo por semana . . . . .	39
5.18. Encuestas nivel de servicio . . . . .	40

5.19. Matrices de correlación . . . . .	41
5.20. Gráfica de Davies-Bouldin . . . . .	42
5.21. Mapa auto-organizado . . . . .	43
5.22. Cluster seleccionado por SOM . . . . .	43
5.23. Diagrama de violín, resultados modelos . . . . .	45
5.24. Matrices de confusión . . . . .	46

# Índice de tablas

3.1. Descripción de las líneas de telefonía y chat de Icetex . . . . .	7
4.1. Normalización y categorización . . . . .	19
4.2. Sección días del mes en semanas . . . . .	21
4.3. Hiper- parámetros para la búsqueda cruzada . . . . .	23
5.1. Ejemplos de normalización y categorización . . . . .	24
5.2. Características extraídas por hora . . . . .	30
5.3. Características extraídas por día . . . . .	32
5.4. Características extraídas por semana . . . . .	33
5.5. Kmeans: Relación estados reales de la base con las agrupaciones del cluster . . . . .	42
5.6. SOM: Relación estados reales de la base con las agrupaciones del cluster . . . . .	44
5.7. Resultados modelos en validación . . . . .	45



# Capítulo 1

## INTRODUCCIÓN

El mercado laboral hoy en día es muy volátil y amplio, esto sumado a las nuevas generaciones que buscan diferentes oportunidades y saltan con facilidad de empresa en empresa buscando nuevos desafíos y expectativas laborales, hace que la rotación de personal sea más frecuente. Actualmente, para las empresas es un gran problema ya que realizan todo un proceso de contratación que implica una inversión económica, jurídica y tiempo de capacitación que afecta a la organización cuando se tiene una alta rotación de personal ya que se pierde la inversión, la estabilidad operativa y llegan al punto de no cumplir con sus objetivos empresariales y estrategias planteadas por la falta de personal, además de perder credibilidad en el mercado para nuevos empleados [1] [2].

Aunque la decisión de un empleado de salir de la empresa de forma voluntaria se puede dar por diversas razones como una mejor oferta laboral, no estar a gusto con el ambiente cooperativo, pocas posibilidades de crecimiento profesional, estrés laboral, monotonía, poca satisfacción personal, carga excesiva de trabajo, entre otras; las razones conllevan en general a que el empleado pierde el interés y la motivación de continuar en la compañía y se observa en su baja productividad, disminución en la interacción con el equipo de trabajo, impuntualidad y otros comportamientos en contra de las políticas de la empresa. Esto es lo que Littlewood define como Renuncia Psicológica, la cual indica que el trabajador llega a un estado donde no renuncia [3], pero su comportamiento no contribuye a los fines de la empresa, se puede decir que realizan el mínimo de sus tareas para continuar y con el tiempo la baja motivación de trabajar da como resultado la renuncia de forma anticipada [4] [5] [6].

Una las áreas donde más se ve estos casos es en el sector de *call centers*, en donde la rotación de sus asesores de atención al cliente es alta. Los autores Chevez, R. A., & Inga, L. K. en su estudio validaron que la rotación anual de un centro de atención al cliente en Lima, Perú fue en promedio 18.5% (420 personas aproximadamente) [7]. Esto es debido a la naturaleza del cargo, ya que el empleado siempre está expuesto a la ira, hostilidad y violencia verbal por parte del cliente; mientras tanto a ellos se les exige que manejen y controlen sus comportamientos, incluidas sus verbalizaciones, lenguaje corporal y expresiones faciales, de acuerdo con las reglas y expectativas dadas por la empresa. Sumado a las altas metas que deben cumplir y que en general no es un trabajo deseado sino más por la necesidad de una experiencia laboral o ingreso económico, lo que conlleva a que los asesores renuncien fácilmente [8] [9].

Por estas razones, Millenium BPO, una compañía de externalización de procesos empresariales (*Business Process Outsourcing*), la cual brinda el servicio de atención al usuario a otras empresas con el esquema clásico de un centro de atención al cliente (*call center*) busca nuevas formas de resolver esta problemática. La propuesta de esta investigación es que con la implementación de analítica de datos y el uso de inteligencia artificial se pueda identificar patrones de renuncia psicológica y la posibilidad de estimar el retiro anticipado de un asesor.

La siguiente investigación esta dividía en seis capítulos, el primero es la continuación del contexto de la problemática enfocada a la compañía Millenium BPO, en conjunto de una revisión literaria de los últimos cuatro años que engloben este tema. En el siguiente capítulo se describirá parte de la información que se debe tener en cuenta para el entendimiento de la investigación, la definición de una compañía de externalización de procesos empresariales junto con su método de operación, descripción jerárquica y aplicativos relevantes, también términos básicos con los que se estima la productividad y buen servicio de los asesores y finalmente, definiciones de métodos de exploración de datos y modelos de inteligencia artificial. Posterior a ello, la metodología empleada para el desarrollo de la investigación y sus respectivos resultados, discusión de estos y conclusiones al terminar la ejecución de la investigación.

## 1.1 Planteamiento del problema

El área de recursos humanos (RH) es la encargada de gestión humana dentro de la empresa, principalmente sus objetivos son seleccionar y formar talento humano para que ejerza determinadas tareas dentro de la compañía. Las tareas del área de RH recaen en el organización y planificación del personal, reclutamiento, selección, formación, prevención de riesgos laborales y el bienestar de sus empleados; este último según Littlewood tiene una relación directa con la renuncia psicológica ya que si el empleado no está a gusto con su situación empieza a aumentar la posibilidad de retirarse de la empresa. Por ello, se dice que el área de RH también está encargada de mantener una buena psicología laboral con el fin de mantener a sus empleados en un nivel óptimo de motivación para que realicen sus tareas de forma correcta, ágil y proactiva, con estrategias e incentivos que son repartidos en las demás áreas de la compañía [3] [10] [11].

Sin embargo, este monitoreo no es fácil y más para las empresas grandes ya que las personas que se encuentran en este estado de renuncia psicológica de igual forma están haciendo lo mínimo posible para continuar en la compañía por lo que no es tan fácil de identificar [12].

Millenium BPO es una empresa con una trayectoria de 20 años en el sector de *call centers* en Colombia, se encarga de dar asistencia tercerizada para otras empresas como atención al usuario, ventas, soportes técnicos, cobranzas, entre otros. Estos procesos son impulsados por tecnologías como la automatización robótica de procesos (RPA), aplicativos propios de gestión de relaciones con el cliente (CRM) e implementación de inteligencia artificial junto con inteligencia de negocios. Actualmente, Millenium BPO padece del problema explicado anteriormente, la rotación del personal es un tema crítico, los datos entregados por la empresa indican que en el mes septiembre del 2022 se retiraron de forma voluntaria 289 asesores en

sus diferentes campañas (empresas que contratan a Millenium BPO), esto hablando únicamente de los asesores no de los miembros del *staff* (personal encargado de gestionar, formar y evaluar a los asesores), personal administrativo, ingenieros y otros cargos que conformar a la compañía [12].

En busca de minimizar este problema la división de innovación, inteligencia artificial e inteligencia de negocios denominada Datos No Estructurados (DNE), quiere afrontar esta situación con ayuda del análisis descriptivo y correlación de variables en sus asesores que demuestren declive en su comportamiento operacional cuando entran en el estado de renuncia psicológica. Se tiene a disposición el uso de datos socio-demográficos (edad, estrato, sexo, entre otros) e indicadores de rendimiento (tiempo promedio de interacción, calificación del usuario, número de interacciones atendidas, entre otras). para así seleccionar los más relevantes y con la ayuda de inteligencia artificial estimar si el asesor se encuentra en estado de renuncia psicológica y así dar un estimado a RH de los posibles asesores que tengas estos comportamientos para establecer estrategias de retención, apoyo psicológico o prepararse ante la posibilidad de la renuncia del personal.

La investigación está planteada para ser implementada en la campaña de Icetex, la cual es una empresa del estado que promueve a los ciudadanos el acceso a la educación superior con otorgamiento de créditos y apoyos económicos [13]. Por ello, Millenium BPO cuenta con 490 asesores activos que están en la capacidad de brindar información, acompañamiento y recibir solicitudes de los usuarios que quieran acceder a estas oportunidades, como también la entrega de información del estado de cuenta, gestionar cobranza, acuerdos de pago, desembolso y renovación de créditos ya existentes, estas solicitudes llegan por diferentes módulos de atención como lo son telefónico, página web, redes sociales, radicación de PQRs y chats.

Para esta investigación se decidió utilizar una de sus líneas de atención: chat. Esto debido a que su relación entre asesores activos y retirados es alta y la empresa mostró especial interés por iniciar con dicho segmento ya que es una de las líneas más complejas en la contratación de personal. Por ende, la pregunta de investigación relacionada a esta problemática es:

**¿La inteligencia artificial puede ser implementada en la estimación de la renuncia psicológica de los asesores usando variables operativas y socio-demográficas en la renuncia de asesores en la línea de chat en el periodo 2021 y 2022?**

## 1.2 Antecedentes

Al hacer una revisión literaria de los últimos cinco años entre revisas, artículos y trabajos de grado, se evidencia que el estudio de la predicción y la clasificación del retiro voluntario de empleados no es un tema nuevo, pero si poco explorado.

El primer estudio ha destacar es el realizado por LL, Yu-Ri et Al [14], en esta investigación es una comparación estadística de los niveles de estrés en diferentes sectores de la economía, en esta se da como resultado que el sector de los *call center* tienen mayor índice de estrés en sus trabajadores, dando como resultado síndrome de *burnout* o estrés laboral. El concepto de estrés laboral es retomado por Littlewood [3] para dar explicación a este tipo de estrés

laboral que desencadena en los comportamientos de renuncia psicológica de un empleado.

Ahora, buscando investigaciones sobre la predicción en la rotación de empleados, está la de Raj Chakraborty et al [15] los cuales correlacionaron diferentes variables: satisfacción laboral, edad, distancia entre lugar de residencia y el trabajo, entre otros. Estas son expuestas en el modelo de clasificación de bosques aleatorios (*Random Forest*) dando buenas métricas en la sección de personal que se retira de la empresa. También, Shikha N. Khera et al [16] implementa el mismo modelo pero utilizándolo con diferentes variables: número de proyectos, horas trabajadas, satisfacción laboral, otras. Tuvieron una exactitud en la predicción de renuncia del 98 %.

Por otro lado, de la investigaciones de Vasa, Jalpesh et al [17] y Henao C. [18], tomaron información de los datos socio-demográficos (edad, sexo, estrato, entre otros), antigüedad y nivel dentro de la empresa para la estimación de la rotación del personal, los primeros autores obtuvieron con dichas variables el modelo de máquinas de vectores de soporte (*Support Vector Machines*) las mejores métricas . El segundo autor propone implementar el algoritmo de *XGBoost* el cual obtuvo un 87 % de precisión en sus pruebas. Así mismo la autora Monsalve Lopera, L [19] emplea las mismas características mencionadas anteriormente, pero en el contexto real en una BPO de la ciudad de Medellín, Colombia, dando como resultado de su investigación que las variables más representativas son la antigüedad y el devengado para determinar la rotación del personal.

Ahora bien, el común de estas investigaciones es el uso de alguna encuesta o métrica relacionada directamente con la situación y opinión del empleado con respecto a su trabajo. Como esta investigación está planteada desde la extracción datos operativos es importante resaltar la importancia de la extracción y la calidad de los datos como lo expone Zhang, Heng et al [20] en su investigación. Por tal motivo es de importancia resaltar la investigación de Valero, S. et al [21] la cual toma como estrategia el uso de inteligencia artificial no supervisada para la minería de datos y agrupación de datos para luego implementar el algoritmo de bosques aleatorios en la predicción de la deserción escolar. Finalmente, se encuentra la revisión hecha por parte de Wickramasinghe, Chathurika [22] sobre el uso de mapas auto-organizados con redes neuronales (*Deep Self-Organizing Maps*) como un herramienta robusta y novedosa en la minería y visualización de datos.

## Capítulo 2

# OBJETIVOS

### 2.1 General

Determinar factores relacionados con renuncia de agentes de la campaña Icetex, empleando datos disponibles en Millenium BPO y el uso de algoritmos de inteligencia artificial.

### 2.2 Específicos

1. Emplear herramientas o algoritmos de la inteligencia artificial para determinar variables asociadas a la renuncia de agentes
2. Establecer una base de datos de dos años de los exagentes con su respectiva caracterización de la información en el segmento chat.
3. Determinar la existencia de una relación entre la disminución de los niveles de indicadores de rendimiento de un agente y su renuncia tres meses antes de esta.

## Capítulo 3

# MARCO TEÓRICO

Se tratará los temas relacionados para el desarrollo de la investigación, tales como, la definición de una compañía de externalización de procesos empresariales (*Business Process Outsourcing (BPO)*) su operación, el organigrama de una campaña de *call center*, las tecnologías y aplicativos que se implementan, métricas que se utilizan en la industria y finalmente la implementación de analítica de datos y de modelos de inteligencia artificial.

### 3.1 *Business Process Outsourcing (BPO)*

La asociación colombiana de BPO, define el sector de externalización de procesos empresariales como: “el sector de tercerización de procesos de negocio, se entiende como la delegación de uno o más procesos de negocio, intensivos en el uso de tecnologías de la información, a un proveedor externo, quien a su vez posee, administra y gerencia los procesos seleccionados, basado en métricas definidas y medibles” [23].

Por lo anterior la función de un BPO es proporcionar la infraestructura y el personal para realizar actividades o procesos de atención al usuario como lo realiza un centro de atención al usuario (*Front office*) y de procesos internos que requieren validación o analítica (*Back office*), la figura 3.1 muestra algunos ejemplos de las actividades que se realizan en cada uno de ellos.



Figura 3.1: Ejemplos de las actividades que se realizan en una BPO. Extraído de [23]

Las empresas que contratan a los BPO se denominan dentro del sector "campañas", las cuales dependiendo del tiempo de contratación y los procesos a delegar, tendrán más o menos funciones, por ejemplo, puede existir una campaña exclusiva al registro de emergencias de una red eléctrica, como puede existir una que cubra todas las formas de contacto y de gestiones posibles por el tercero, el cual es el caso de Icetex. Esta es una campaña que se divide en diez líneas y tienen la facultad de gestionar más de 500 tipos de solicitudes.

Las líneas como se ha mencionado es el segmento o el medio por donde se conecta o remite una solicitud el usuario, estos segmentos pueden venir de diferentes canales de comunicación, sin embargo, los canales también se pueden dividir dependiendo de la interacción con el usuario, estas pueden ser de tipo *inbound*, *outbound* o medios digitales, respectivamente el primero es cuando el usuario decide comunicarse y se gestiona de manera inmediata, el segundo es cuando la BPO busca comunicarse en tiempo real con el usuario por los medios registrados, y el último, es la comunicación no en tiempo real con el usuario, se trata en general de peticiones, quejas, reclamos o validaciones con cierto nivel de analítica. Para dimensionar lo dicho, la tabla 3.1 describe los segmentos de telefonía y chat en la campaña de Icetex, en ella se aprecia sus canales con su tipo de interacción y ejemplos de procesos que deben realizar, usualmente [12].

<b>Línea</b>	<b>Canal</b>	<b>Interacción</b>	<b>Procesos</b>
Telefónico	Beneficiarios	<i>Inbound</i>	Información del crédito, gestión de renovación, etc
		<i>Outbound</i>	Cobranza, acuerdos de pagos, etc
	Línea azul	<i>Inbound</i>	Línea directa puntos de atención presencial
	Candidatos	<i>Inbound</i>	Información de productos, acompañamiento del proceso, etc
<i>Outbound</i>		Información de productos y convocatorias (personas que han autorizado la devolución de esta llamada)	
Chat	WhatsApp	<i>Inbound</i>	Información del crédito, gestión de renovación, etc
	Página web (SAV)	<i>Inbound</i>	Información del crédito, gestión de renovación, etc

Tabla 3.1: Descripción de las líneas de telefonía y chat de Icetex

Luego de repasar como se divide y funciona una BPO, a nivel operacional existe una clara estructura jerárquica, la cual es expuesta en la figura 3.6. Inicia con el gerente de cuenta o gerente de campaña, el cual es el encargado de velar por el cumplimiento de las metas a nivel general y es el encargado de presentar los resultados ante la compañía y otras áreas. Él a su vez dirige tres áreas, operación, calidad y formación, estas están encargadas de dirigir el funcionamiento de la campaña y la gestión del personal, supervisar y auditar las interacciones en busca de una mejora continua y dar las capacitaciones y retroalimentación a los asesores. En el siguiente nivel se encuentran los supervisores quienes cumplen las mismas labores enfocadas en el segmento asignado. se continúa con los coordinadores los cuales están disponibles para la resolución de dudas y apoyo a su grupo de asesores asignados. Por último, está el asesor

quien es el que está en contacto con el cliente gestionando su solicitud, queja, reclamo u otros motivos por el cual se da la interacción [12].

Teniendo en cuenta las tareas de las diferentes áreas se puede asimilar a un sistema retroalimentado ya que el área de calidad actúa como sensor, formación como el controlador que aplica cambios sobre los coordinadores de operación y estos afectan a los asesores. Esta es otra de las razones por las que se manifiesta el cansancio psicológico que presentan los asesores ya que están en constante monitoreo y evaluación el cual genera presión a su labor. En la figura 3.6 la ruta que esta en color azul representa esta retroalimentación que se realiza constantemente.

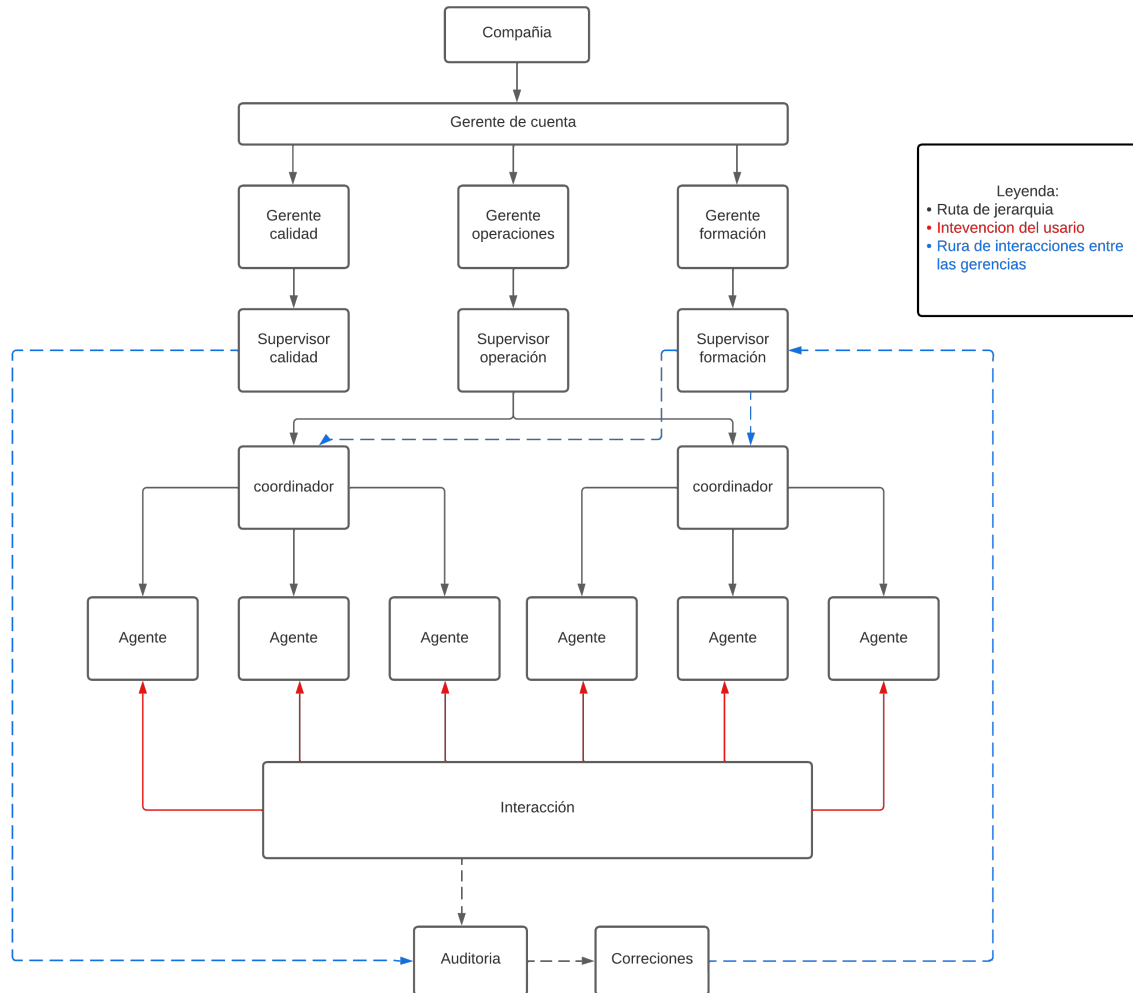


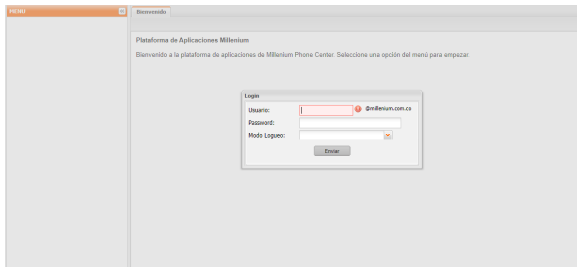
Figura 3.2: Representación jerarquía operación de un call center y representación de lazo de control

Finalmente, la interacción que hace el asesor con el usuario; se hace a través del *Customer Relationship Management* o Gestión de Relación con los Clientes (CRM), que es una plataforma corporativa que funciona como centro de mando para las actividades que deben realizar los asesores, además le permite al asesor, buscar, actualizar y otorga sugerencias de respuestas

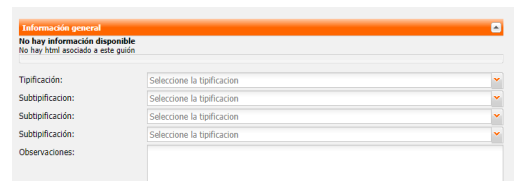


a los usuarios, ya que es un sistema dinámico que reduce procesos repetitivos y obsoletos que antes realizaban los *call center*. Esta es la herramienta principal donde se estructura la información que luego será guardada en los servidores de bases de datos tanto del cliente como del asesor en su comunicación [24].

Millenium BPO cuenta con una gran ventaja ya que estos *software* son construidos por su unidad de desarrollo por lo que le permite ser muy versátil y personalizado a cada campaña. Las figuras 3.3 son vistas del CRM del Icetex, la figura 3.3b es el lugar de gestión del asesor donde documenta su interacción con el usuario y esta al finalizar es cargada a una bases de datos que guarda información como la hora de inicio de la gestión, teléfono de contacto, tiempo de interacción, tipificaciones (motivos de la interacción), calificaciones, entre otros [12].



(a) Vista CRM Icetex: Inicio



(b) Vista CRM Icetex: Lugar de gestión asesor

Figura 3.3: Vistas CRM Icetex

### 3.2 Indicadores de rendimiento

Los indicadores de rendimiento o en ingles *Keys Performance Indicators* (KPIs), como se menciona son indicadores cuantificables que permiten tener mejor visualización del desempeño de los agentes a través de los datos de operación. Hay una gran cantidad de indicadores que dependen del segmento y el tipo de interacción que se tenga con los usuarios [25] y algunos ejemplos de estos son descritos a continuación:

- Satisfacción del cliente: Indica que la respuesta de la solicitud fue del agrado del usuario, generalmente son variables categóricas como: Muy satisfecho, satisfecho, insatisfecho y muy insatisfecho.
- Tiempo de trabajo después de la llamada: Mide cuanto tiempo se tarda en terminar la gestión y poder atender la siguiente interacción (*backoffice*).
- Tasa de llamadas al mes: Son las interacciones que ingresan en un periodo de 30 días.
- Resolución de llamadas: Indica la solución al requerimiento del usuario de manera efectiva y coherente, generando satisfacción en el usuario
- Tasa de llamadas completadas: Se refiere a la tasa de llamadas que no se interrumpen y terminan por decisión del usuario.
- Tasa de ausentismo laboral: Se refiere a la Cantidad de horas de ausencia del asesor contra la cantidad de horas laborales.

- Puntuación neta del promotor (*Net Promoter Score*, NPS): Es una métrica asociada a la satisfacción del cliente ante la gestión y su perspectiva a la campaña
- Productividad: Cálculo de las interacciones que realiza el agente en un tiempo determinado, generalmente es por hora o día.
- Tiempo medio de operación: Este es el tiempo en minutos que tarda el asesor en realizar una interacción completa.

### 3.3 Exploración de datos

En esta investigación la exploración de datos tiene un alto componente de importancia, por lo que la siguiente sección busca traer conceptos comunes de la visualización e interpretación de gráficas que se implementaran a lo largo de la investigación.

- Gráfica de barras: La gráfica de barras representa una estimación de la tendencia central de una variable numérica con la altura de cada rectángulo, excelente para dar el tamaño de un grupo categórico [26].
- Diagrama de caja: Es un tipo de gráfico que se usa para el análisis de datos explicativos, visualización de los cuartiles y la media permite observar la distribución de los datos y sus asimétrica de forma sencilla y rápida, teniendo cinco datos estadísticos importantes: valor mínimo, primer cuartil (25% de los datos), mediana (50% de los datos), tercer cuartil (75% de los datos) y su valor máximo. Adicional, los datos fuera de la caja representan datos atípicos de la muestra [27].

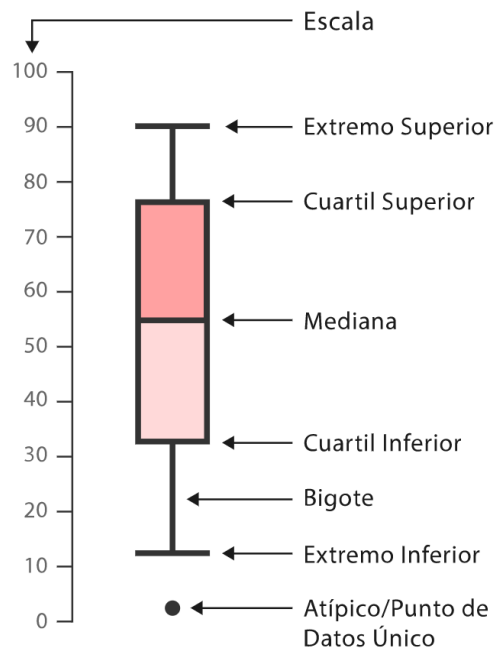


Figura 3.4: Descripción partes diagrama de caja. Extraído de [28]

- Estimación densidad del núcleo (*Kernel density estimate*): Esta es una técnica de visualización de datos, la cual permite estimar una curva continua y suave sobre un histograma. La curva se calcula ponderando la distancia de todos los puntos en cada ubicación específica a lo largo de la distribución. Si hay más puntos agrupados localmente, la estimación es mayor, ya que aumenta la probabilidad de ver un punto en esa ubicación [29] [30].

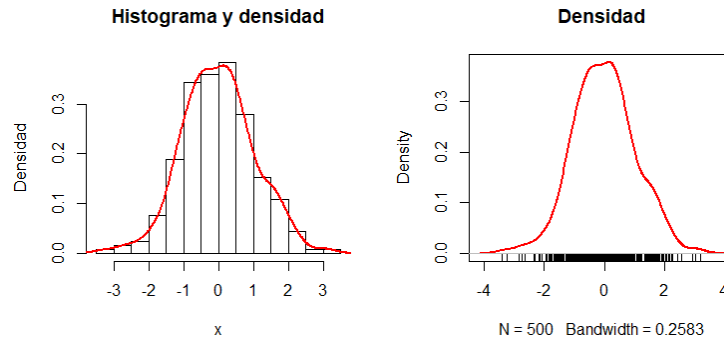


Figura 3.5: Ejemplo de estimación densidad del núcleo. Extraído de [31]

Ahora bien, al tratarse de un suavizado de un histograma, donde se tienen muchos picos en el momento de la visualización se perderá este comportamiento, por lo general se trata de un kernel de tipo Gaussiano en donde su dominio en los extremos salen de datos que originalmente no estaban [30].

- Diagrama de violín: Combina los métodos de diagrama de caja con el de estimación de densidad del núcleo. Es interesante ya que para cada categoría da información de la estadística relevante y su silueta muestra la forma de distribución que tiene la categoría [32].

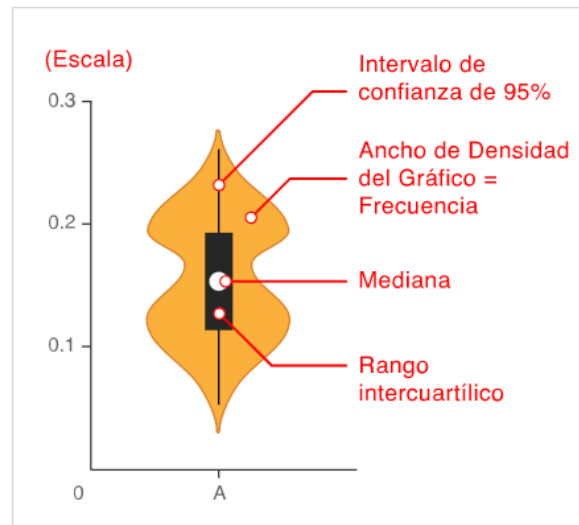


Figura 3.6: Descripción diagrama de violín. Extraído de [33]

- Gráfica línea continua con bandas de error: Este tipo de diagrama representa el error o la incertidumbre como una región sombreada alrededor de la curva principal.

### 3.4 Inteligencia artificial

El término inteligencia artificial (IA) se expresa comúnmente cuando una computadora, robot o una maquina se asemeja al comportamiento o inteligencia humana. Esto significa aprender de la experiencia (*Machine Learning*, ML), reconocer objetos (*Computer Vision*), entender y responder el lenguaje natural (*Natural Leanguje Processing*), hasta tomar decisiones, entre otros.

Por ello, el concepto de inteligencia artificial es muy amplio y depende arduamente de la aplicación que se le quiera dar. En el caso de esta investigación se orienta más hacia la línea de del aprendizaje de maquina (*Machine Learning*), la cual un computador aprende a realizar una tarea estudiando un conjunto de ejemplos para realizar tareas similares con nuevos datos que no ha visto anteriormente. La figura 3.7 explica los dos caminos o estrategias comunes que toma el ML en su implementación.

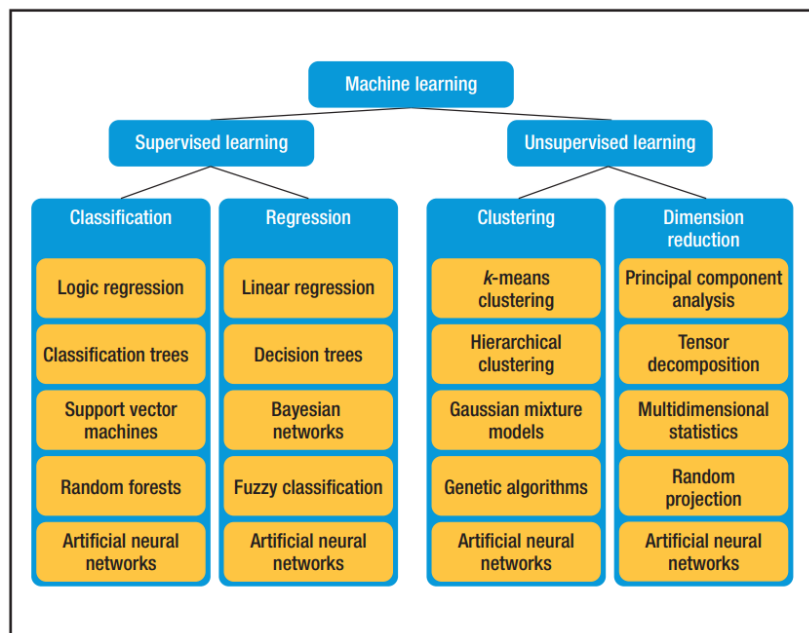


Figura 3.7: Diagrama estrategias del *Machine Learning*. Extraído de [34]

- Estrategia uno: Entrenamiento no supervisado, Para este caso la máquina aprende sin conocer los datos de salida o las soluciones a las que debe llegar. Es común aplicar esta ruta para la descripción de una serie de ejemplos de los cuales no se tiene conocimiento previo sobre ella o para estimar el comportamiento de sus datos agrupándolos (*clustering*) [34]. Asi mismo el algoritmo más usual es el *Kmeans*. Este algoritmo busca minimizar la suma de las distancias entre los puntos y el centroide al que pertenecen, la cantidad de centroides con la que hace la exploración se debe elegir previamente, por

ello se debe realizar a través de los siguientes pasos:

1. Iterar la muestra en diferentes números de agrupamiento.
2. Elegir una de las siguientes métricas:
  - Método del codo: Este se emplea para revisar en que punto la gráfica tiene un punto de inflexión para determinar el número de grupos óptimo. Por lo que se busca el punto cuando su segunda derivada es igual a cero.
  - Índice de Davies Bouldin: En el caso de DB indica el número de grupos óptimo en su mínimo local.
  - Análisis de silueta: Para SLC es lo contrario que DB, en este su máximo local indicará el número de grupos óptimo.
3. Estimar métrica por cada conjunto de agrupamiento.
4. Elegir el número de grupos óptimo según los criterios de cada métrica.
5. Estimar número óptimo de grupos sobre la muestra y observar resultados.

Un ejemplo de agrupamiento con *kmeans* se observa en la figura 3.8 [34]

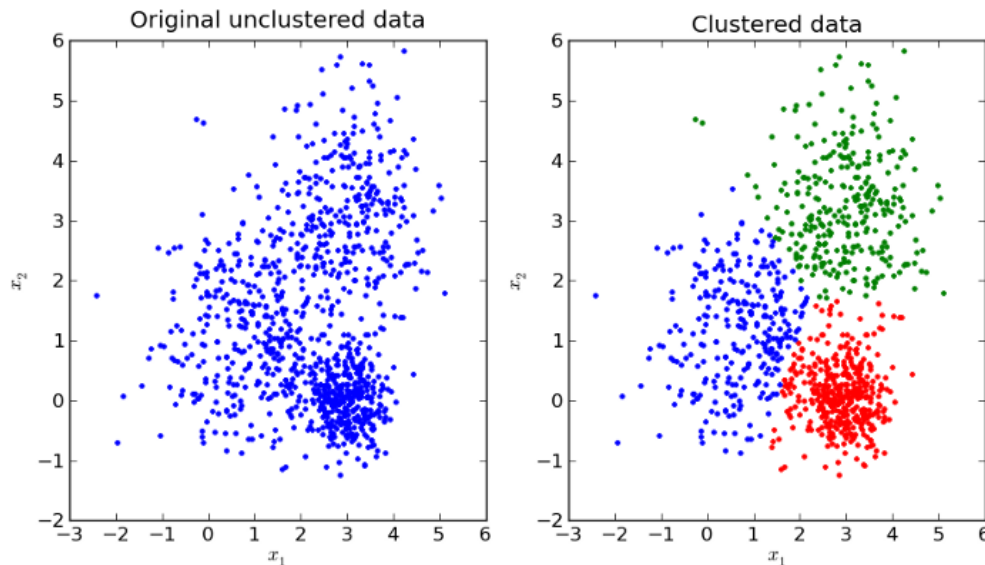


Figura 3.8: Agrupamiento con *Kmeans* [35]

Otro de los algoritmos de agrupamiento es Mapas auto-organizados con redes neuronales (*Deep Self-Organizing Maps, SOM*) como su nombre lo indica es un mapa auto-organizado de los pesos sinápticos de una red neuronal, la figura 3.9 explica de manera gráfica como se conforma este mapa.

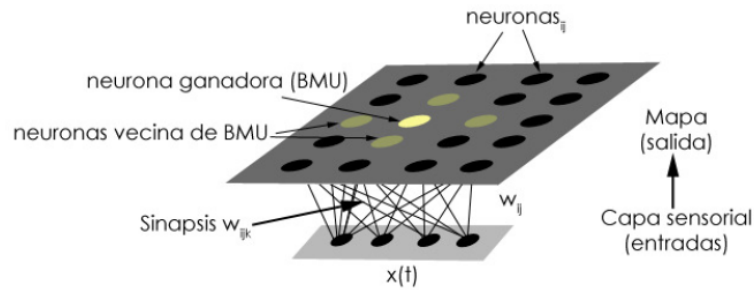
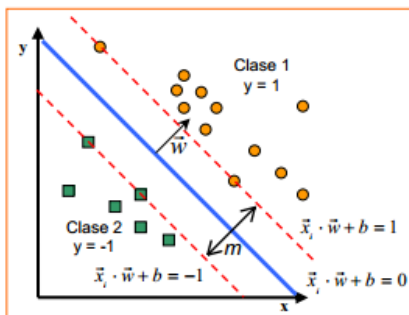


Ilustración 1. Estructura de un mapa autoorganizado

Figura 3.9: Arquitectura de mapas auto-organizados con redes neuronales. Extraído de [36]

Los autores Wickramasinghe, Chathurika et al, en su revisión de como emplear esta herramienta para la visualización de los datos indican que se permite la extracción de características de alto nivel en una sola estructura, permitiendo visualizar las relaciones entre el mapa , mapa de colores que se obtiene en la visualización que permite analizar cualitativamente el número de grupos [22].

- Estrategia dos: Entrenamiento supervisado, para esta estrategia la máquina sí tiene la solución ante los ejemplos que se usan para entrenar. En esta estrategia hay dos caminos comunes, el primero clasificar, donde se le enseña al modelo que al ver los datos nuevos los catalogue en una de sus categorías ya definidas. Y el segundo estimar, la probabilidad en la que se los datos se encuentren en una de las categorías ya definidas, similar a un modelo de regresión [34]. Para el camino de clasificar se tienen diferentes algoritmo, algunos son:
  - Máquinas de vectores de soporte (*support vector machine*), usado por Shikha N. Khera et al [16] en su investigación, este funciona correlacionando datos en un espacio de características de grandes dimensiones de forma que los puntos de datos se puedan categorizar, incluso si los datos no se separan linealmente de otro modo [37] [38].



Hiperplano de separación óptimo

Figura 3.10: Arquitectura de máquinas de vectores de soporte. Extraído de [39]

- Bosques aleatorios (*Random forest*), usados por los autores Raj Chakraborty et al [15] y Vasa, Jalpesh et al [17], este se caracteriza por el uso de varios arboles de decisión para llegar a la respuesta mas óptima. Los arboles de decisión se encuentran en la categoría de regresión y su algoritmo se basa en una serie rutas establecidas hasta llegar al final, por ello la estrategia de bosques aleatorios es dividir la muestra y pasar por estos múltiples arboles para ponderar los resultados de todos como se muestra en la figura 3.12 [40].

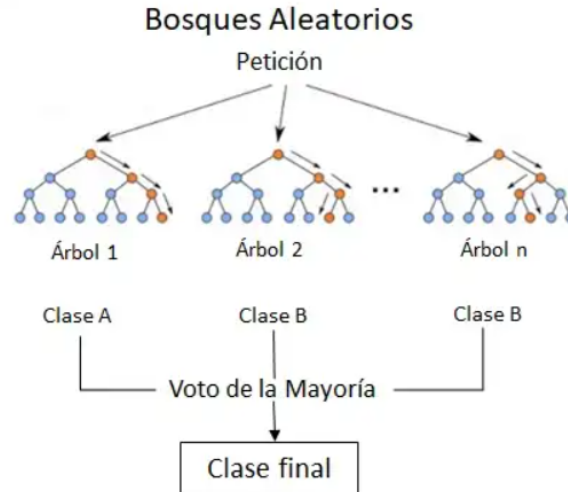


Figura 3.11: Arquitectura de bosques aleatorios. Extraído de [41]

- Redes neuronales artificiales (*artificial neural networks*), este ultimo son algoritmos que están inspirados en el comportamiento del cerebro humano imitando el funcionamiento y la conexión entre si de las neuronas biológicas. Estas se componen de capas de nodos en tres niveles, capa de entrada, capa oculta y capa salida. En entrenamiento las neuronas se guardan en un peso y un umbral de paso (función de activación), donde al llegar nuevos ejemplos y estos superan el umbral, permite el paso de la información al siguiente nodo y así consecutivamente hasta llegar al final del nivel [42] [43].

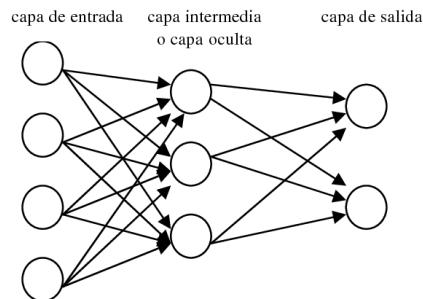


Figura 3.12: Arquitectura de redes neuronales artificiales. Extraído de [44]

Finalmente, para hablar sobre las métricas del rendimiento del modelo hay que tener claridad en los conceptos de falso positivo, falso negativo, verdadero positivo y verdadero negativo [45].

- Verdadero positivo (VP): cuando el valor real es Positivo y la predicción también es Positiva.
- Verdadero negativo (VN): cuando el valor real es negativo y la predicción también es negativa.
- Falso negativo (FN): Cuando el valor real es negativo pero la predicción es positiva. También se conoce como error de tipo 1.
- Falso positivo (FP): Cuando el valor real es positivo pero la predicción es negativa. También se conoce como error de tipo 2.

		<b>Predicción</b>	
		<b>Positivos</b>	<b>Negativos</b>
<b>Observación</b>	<b>Positivos</b>	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	<b>Negativos</b>	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Figura 3.13: Matriz de confusión. Extraído de [46]

Estos conceptos se aprecian con facilidad en la gráfica de matriz de confusión (figura 3.13) y con ella se observan la métricas de rendimiento [45], estas son:

- Exactitud (*Accuracy*): Frecuencia con la que el clasificador realiza una predicción correcta, ecuación 3.1.

$$Exactitud = \frac{VP + VN}{Total} \quad (3.1)$$

- Precisión: Se relaciona con la predicción verdadera. Calcula cuantas predicciones son realmente verdaderas de todas las que el modelo indicó que lo eran, ecuación 3.2.

$$Precisión = \frac{VP}{VP + FP} \quad (3.2)$$

- Sensibilidad: Esta métrica se relaciona con el número de observaciones reales que se predicen correctamente, ecuación 3.3.

$$Sensibilidad = \frac{VP}{VP + FN} \quad (3.3)$$



- *F1-Score*: Este ultimo es la media armónica entre la precisión y la sensibilidad, por lo que parece un equilibrio entre las dos y una métrica mas completa, ecuación 3.4

$$F1 - score = \frac{Precisión * Sensibilidad}{Precisión + Sensibilidad} \quad (3.4)$$

# Capítulo 4

## METODOLOGÍA

En este capítulo se describe los pasos en el tratamiento de bases de datos, exploración de variables y el uso de herramientas de inteligencia artificial.

### 4.1 Bases de datos

Gracias a la contribución y apoyo de Millenium BPO para esta investigación, se tienen diferentes bases de datos con los históricos de los agentes de la campaña de Icetex. Son dos bases de datos de orígenes distintos, la primera contiene toda la información socio-demográfica de los asesores activos e inactivos de todos sus canales y la segunda tiene la información de operación, en ella se registra todas las interacciones realizadas por el agente mediante su interacción con el CRM.

Sin embargo, mucha de esta información no se encuentra estructurada o directamente no aportan información para esta investigación, por lo que cada base de datos paso por una primera etapa de pre-procesamiento, posterior a esta se realizó una exploración de datos con el fin de describir la información resultante y con ello buscar las relaciones, tendencias y describir la población que participa en la investigación. Luego, una etapa de extracción de características, con la información relevante encontrada en lo anterior, se extraen las variables más representativas de cada base de datos. Finalmente, se concilia información de las dos bases en una sola base y así obtener una matriz única.

#### 4.1.1. Bases de datos demográficas

Como su nombre lo indica en esta base de datos demográficos se obtiene información de todos los asesores que han sido contratados por Millenium BPO para la campaña Icetex en los últimos dos años (2021-2022). Estos incluyen: sexo, fecha de nacimiento, dirección, localidad de vivienda, estrato, número de hijos, nivel educativo, fecha de contratación y de retiro, entre otros.

- Preprocesamiento:

Las etapas del procesamiento se explican en la figura 4.1, en ella se puede observar que la primera fase busca eliminar errores e información no relevante y la segunda fase inicia con el bloque denominado tratamiento de variables no estructuradas, su finalidad fue normalizar y categorizar el texto que se encuentran en estas columnas.

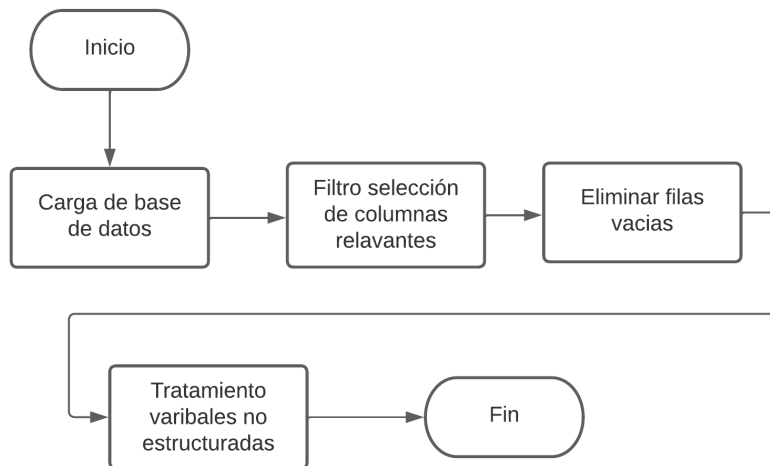


Figura 4.1: Diagrama de bloques pre-procesamiento: Base de datos demográficos

Las variables a las que se le aplicó este método son las presentadas en la tabla 4.1, la primera columna es la variable, la segunda con cuantas etiquetas fueron encontradas en la base de datos y la tercera el nombre de las etiquetas.

Característica	Cantidad de categorías	Nombre categorías
Sexo	2	Femenino, Masculino
Estado civil	4	Soltero, Casado, Unión Libre, Divorciado
Escolaridad	3	Profesional, Tecnólogo, Escolar
Medio de transporte	5	Transporte Publico, Bicicleta, Caminando, Moto, Carro

Tabla 4.1: Normalización y categorización

Por último, se hizo una estandarización de las direcciones para que se encuentren en la estructura general en Bogotá, la estructura es la siguiente:

Vía principal (1) # Vía generadora (2) – Número de placa (3) [47]

1. Muestra donde se ubica el predio: calle, carrera, diagonal + números (00-99) y letras (A-Z)

2. Vía que se cruza con la principal: Números (00-99), letras (A-Z) o cuadrantes (Sur o Este)
  3. Número de placa: Números (00-99)
- Ajuste base de datos:

Con la información normalizada se extrajo cuatro nuevas características, estas fueron:

1. Edad: Para los agentes Retirados es la diferencia entre su fecha de nacimiento y la fecha de retiro y para los agentes Activos, es con respecto a la fecha actual (9/11/2022).
2. Meses en la empresa: Se modifica igual que la edad, para los Retirados es la fecha de contratación contra la de retiro y para los Activos contra la actual.
3. Barrio: Con la calle estandarizada y mediante la geo-codificación se extrae el barrio de donde se encuentra la dirección asignada.
4. Distancia en Km casa a sede de trabajo: Teniendo en cuenta que la sede de trabajo para la campaña de Icetex se ubica en la dirección Calle 72 86-60, Engativá, Bogotá, con latitud y longitud 4.697 y -74.106. Mediante la geo-codificación se extrae la latitud y longitud de la dirección de cada asesor y con ello se calcula la distancia entre dos puntos utilizando la distancia geodésica o la distancia ortodrómica en Kilómetros.[48].

- Análisis descriptivo

Con esta información se buscó relaciones entre las variables, primero realizando un análisis descriptivo a la población y posterior separando la base entre asesores activos y retirados para analizar su comportamiento entre las categorías.

#### **4.1.2. Bases de datos operación**

Esta base de datos es más amplia ya que contiene la gestión del día a día del canal chat, en promedio a este canal ingresan 120.000 interacciones al mes, las cuales un 40 % porcentaje están cubiertas por un robot con inteligencia artificial y el restante un 60 % es tramitado por el asesor. En esta base hay datos como fecha de ingreso de la gestión, tiempo de la conversación, la tipificación, es decir, el motivo del contacto, encuesta que realizan los usuarios sobre la gestión, entre otros datos.

- Pre-procesamiento:

Para esta base de datos el pre-procesamiento fue más sencillo y similar al planteado anteriormente con la base de datos socio-demográfica. La diferencia es que no tiene etapa de tratamiento de datos no estructurados y antes de hacer el filtro de las columnas relevantes se hace un filtro de solo las interacciones del agente, es decir, excluye lo gestionado por el robot tal como lo muestra la figura 4.2.

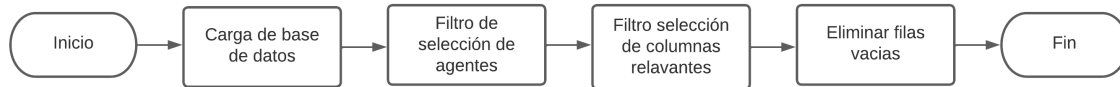


Figura 4.2: Diagrama de bloques pre-procesamiento: Base de datos operación

- **Análisis descriptivo y extracción de características:**  
Debido a la cantidad de datos y para hacer un análisis descriptivo de la generalidad de la operación, se hizo un muestreo de cuatro asesores en un mismo mes, esto en tres meses para observar diferentes comportamientos. La estrategia es la división del mes en cuatro semanas como se presenta en la tabla 4.2 para extraer su comportamiento en diferentes métodos de visualización de datos.

Semana	Intervalo de días
1	1 - 8
2	9 - 16
3	17 - 24
4	25 - 30

Tabla 4.2: Sección días del mes en semanas

Con la división por semanas se extrajo información en parámetros que tengan relación con los KPI's descritos en el marco teórico, excluyendo la puntuación neta del promotor (NPS), ya que se menciona que es una medida que se relaciona con la empresa contratante y no directamente con el asesor.

### 4.1.3. Selección de participantes

Esta es la última sección que relaciona la minería de datos, en ella se concilia toda la información extraída para que quede una única matriz resultante con los datos agrupados de cada asesor. El primer filtro, son los criterios de admisión: agentes de la campaña de Icetex, que los asesores se encuentre en el segmento de chat, que tengan tres o más meses en la empresa. Este tiempo es establecido por el interés de la empresa, debido a que es muy frecuente el retiro de los asesores antes de este plazo, ya que no se adaptan a la presión laboral o no se sienten a gusto con el empleo.

Por lo que se realizó una búsqueda cruzada entre los asesores de la base socio-demográfica en la base operacional para extraer sus características los tres meses. Para el caso de los asesores activos, es más sencilla debido a que ya se extrajeron las características de la base de datos operativos de los meses de julio, agosto y septiembre del 2022, ahora, para los retirados se tomó como inicio el mes antes de la fecha de retiro y dos meses anteriores a partir de este. Se realiza de esta manera para mantener los datos de las cuatro semanas completas. Al final, se obtendrá una base resultante únicamente operativa con la característica de cada asesor

durante los tres meses. Esta base fue la empleada en la correlación de datos con los estados de **Activo** y **Retirado** para determinar si hay diferencia entre ellas.

## 4.2 Correlación datos con estados

### 4.2.1. Análisis exploratorio de los datos

Para la correlación de datos se hizo dos tipos de exploraciones, la primera es un análisis exploratorio de los datos sobre los KPI's extraídos anteriormente para ser comparados entre las dos categorías. La estrategia fue emplear métodos estadísticos como la extracción de promedios y desviaciones estándar entre los agentes, para ver una generalidad del comportamiento y con ello ir comprobando la aparición de este síntoma de la renuncia psicológica. Adicional, se empleó modelos de regresión logística lineal para ver la existencia de tendencias entre las variables que no se observan por separado.

### 4.2.2. Exploración aprendizaje no supervisado

La segunda exploración fue emplear la inteligencia artificial no supervisada por el método tradicional de *Kmeans* y por medio de mapas auto-organizados con redes neuronales (SOM) como lo emplearon los autores Valero, S. et al [21] y Wickramasinghe, Chathurika et al [22], con el fin de describir la base de datos de otra forma y si en relación a sus datos poder diferenciar estas categorías. Es importante destacar que antes de este procedimiento, se realizó un ajuste de pre-procesamiento a la base de datos ya que no pueden tener valores en cero.

- Para el caso del *Kmeans* se itero entre dos y diez grupos buscando el número óptimo de grupos para separar la muestra. La métrica para seleccionar la cantidad de grupos fue el índice de Davies-Bouldin.
- Ahora para el método SOM al entrenar las neuronas estas visualizan su activación a través de un mapa de colores, con él se analizó el número de grupos y hacía que zona del mapa se encuentra agrupados.

## 4.3 Modelo de predicción de renuncia con aprendizaje supervisado

En esta última sección de la investigación fue probar y seleccionar el mejor modelo de ML para la estimación de del estado de **Retiro** o **Activo** de un asesor.

Ya se ha descrito varios de los modelos de ML que son de gran interés para esta investigación, sin embargo, en el momento de ejecutarse se deben optar por otras consideraciones para tener un buen resultado en el entrenamiento del modelo, un ejemplo de esto es la elección adecuada de hiper-parámetros que modifican el comportamiento del modelo, otro es establecer el grupo de ejemplos con el cual el modelo aprende y genera su validación y finalmente las métricas con las cuales se mide el rendimiento modelo.

Para solventar la primera consideración se empleó una búsqueda exhaustiva combinando diferentes hiper-parámetros establecidos a probar y seleccionar la mejor combinación, luego de esto se establece el grupo de entrenamiento y los grupos de validación. Usualmente la estrategia es seleccionar un porcentaje aleatorio de población para entrenar y luego para validar la información, sin embargo, para el caso de esta investigación la estrategia fue implementar el método de validación cruzada, este método separa la base de datos en diferentes sets o  *folds*, logrando escoger un set como prueba dejando los otros sets en entrenamiento, de tal manera que se obtienen las métricas asociadas, repetir el procedimiento variando el set de pruebas hasta que completar todos. Con los resultados se tiene el comportamiento en los diferentes ejemplos para ratificar que no quede sobre entrenado el modelo y además con el método certificar que no importa que dato se entrene siempre mantendrá su rendimiento [49].

Para esta investigación los datos fueron estandarizados por la ecuación 4.1 la cual elimina la media y escala su varianza de los datos por característica.  $\bar{x}$  es la media y  $s$  la desviación.

$$z = \frac{x - \bar{x}}{s} \quad (4.1)$$

Ya con la información consolidada se aplicó el método de validación cruzada con 5  *folds* en los modelos de máquinas de vectores de soporte (SVM), bosques aleatorios (RF) y redes neuronales artificiales (ANN). Y los parámetros para la búsqueda exhaustiva cruzada fueron los establecidos en la tabla 4.3.

Hiper-parámetro				
Modelo	1	2	3	4
SVM	C = {0.01, 0.1, 1, 10, 100, 1000, 10000}	kernel = {linear, poly, rbf, sigmoid}		
ANN	hidden layer sizes = {5, 10, 15, 20, 25, 30}	early stopping = {True, False}	solver = {lbfgs, adam}	max iteracions = {2000, 2500, 3000, 3500}
RF	no estimators = {10, 20, 30}	max features = {autp, sqrt, log2}	criterion = {gini, entropy}	max depth = {5, 10}

Tabla 4.3: Hiper- parámetros para la búsqueda cruzada

## Capítulo 5

# RESULTADOS Y DISCUSIÓN

### 5.1 Bases de datos

#### 5.1.1. Bases de datos demográficas

La base de datos inicia con 140 asesores y 26 características. Al terminar la primera etapa del pre-procesamiento (eliminar información no relevante o que afecte el tratamiento) quedo información de 130 asesores y 14 características, luego en la etapa de Tratamiento de datos no estructurados, aumento a 29 características. La tabla 5.1 muestra un ejemplo del resultado del tratamiento de las variable no estructuradas siguiendo lo establecido de la tabla 4.1, donde la primera columna es la información original y las siguientes son la normalización y caracterización.

<b>Ejemplo</b>	<b>Transporte Público</b>	<b>Bicicleta</b>	<b>Caminando</b>	<b>Moto</b>	<b>Carro</b>
sitp, transmilenio	1	0	0	0	0
bus - carro	1	0	0	0	1
moto electrica	0	0	0	1	0
publico (bus)	1	0	0	0	0

Tabla 5.1: Ejemplos de normalización y categorización

#### 1. Análisis descriptivo de la población

La base de datos contiene una población de 130 agentes, de los cuales el 87 son mujeres (66.9%) y 43 (33.1%) son hombres. La figura 5.1a muestra la diagrama de caja de la edad de los asesores dividido por sexo. En ella se observa que hay valores atípicos en cero, por lo que son considerados errores y se eliminan, siendo así, quedan 126 agentes y la nueva distribución se observa en la figura 5.1b.

Por lo que se determina que la edad promedio de la población es de  $30 \pm 6$  años, dividido por género, sería  $30 \pm 6$  años y para los hombres  $30 \pm 7$ , por lo que la distribución es igual para ambos casos. Ahora la figura 5.3 muestra la distribución del género por el estrato socio-económico de la persona. Como era de esperarse en hay mayor participación de las mujeres como se aprecia en los diagramas de barras.



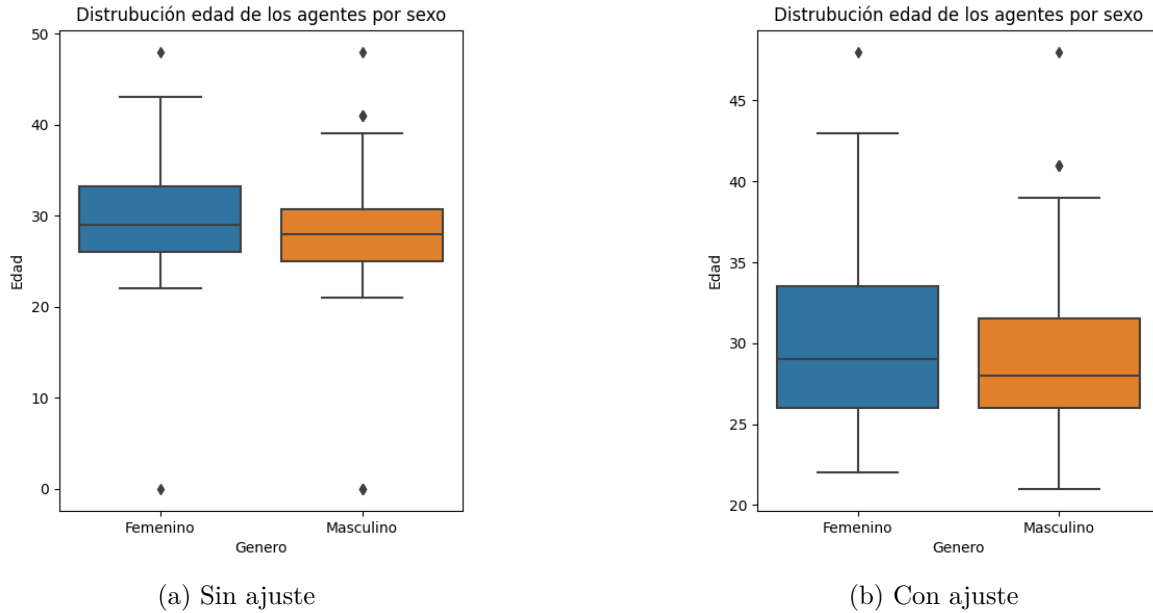


Figura 5.1: Distribución edad de los agentes por genero

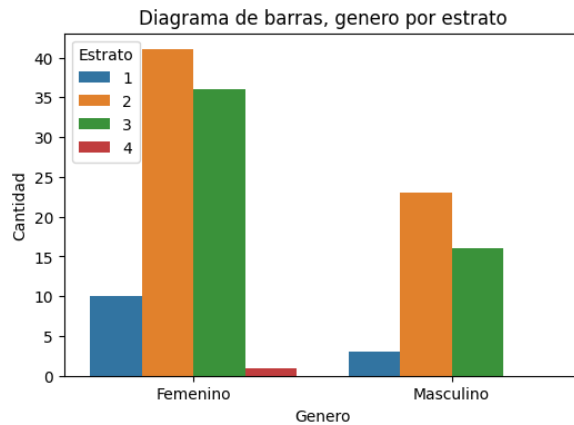
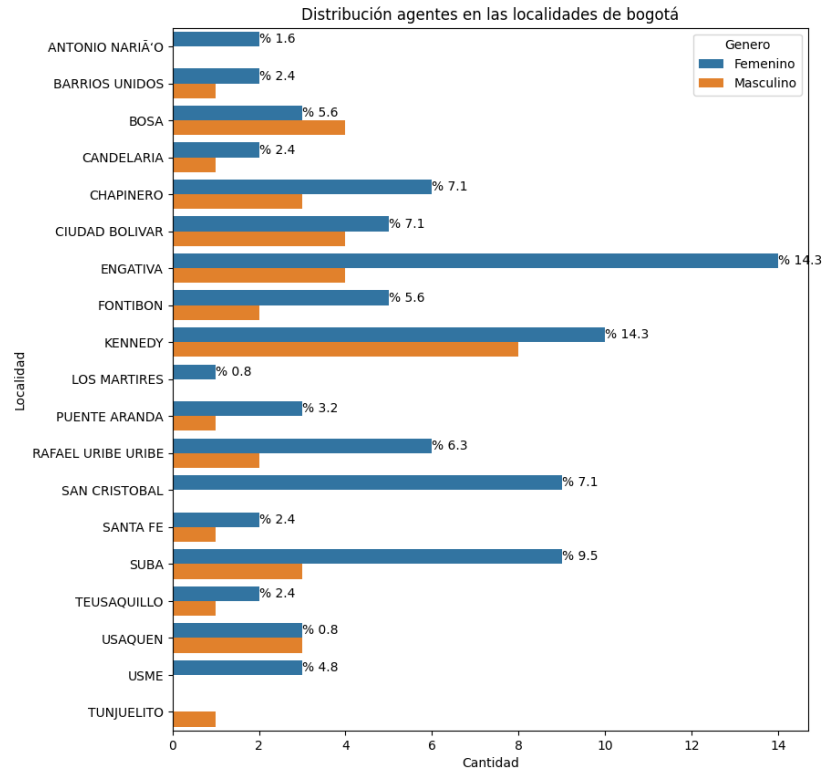


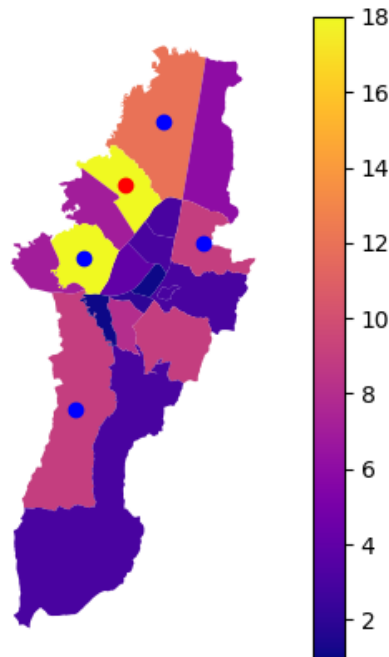
Figura 5.2: Cantidad de agentes por estrato

De esta misma figura 5.3 se observa que los asesores por lo general se encuentran en estrato 2 y 3, siendo entre los dos el 88% de la población. Las localidades mas representativas, son la de Kennedy, Engativá, Suba, Ciudad Bolívar y Chapinero, con 20, 18 y 13 personas respectivamente. Los valores al final de cada barra, corresponde al porcentaje de participación uniendo mujeres y hombres sobre el total de agentes. Y la figura 5.3b es un mapa de calor de donde se encuentran los agentes, se excluyó la localidad de Sumapaz, debido a que no había agentes ahí y al quitarla permitía tener una mejor visualización. De esta gráfica el punto rojo indica la sede principal de Millennium BPO para la campaña de Icetex y los puntos azules las localidades anteriormente mencionadas, menos la de Engativá ya que en esa localidad es donde encuentra la sede.



(a) Cantidad de agentes por localidad

### Asesores en Bogotá - Chat



(b) Mapa de calor

Figura 5.3: Distribución de agentes en Bogotá

## 2. Análisis descriptivo según su estado (Activo o Retirado)

Como se mencionó en la introducción la elección de realizar esta investigación con el canal de chat es debido que su tasa de rotación es bastante alta, actualmente hay 28 agentes activos (28%) y como se observa en la figura 5.4, los asesores activos tienen o menos de un año o ya son bastante antiguos y cuando se compara con los retiros, se observa que la mayoría abandona la campaña antes del 6to mes.

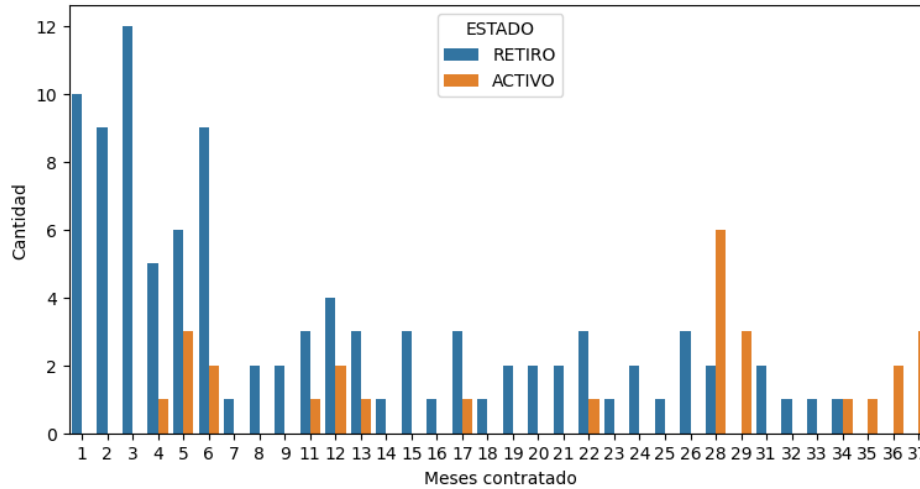
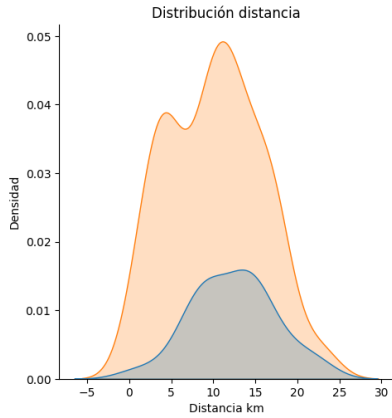
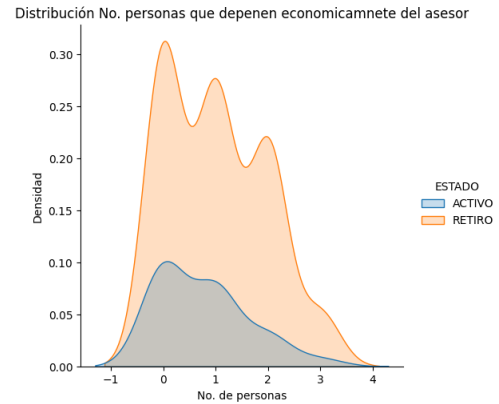


Figura 5.4: Cantidad de agentes según su estado

Ahora bien, hay dos tipos de variables en esta base de datos, las numéricas y las categóricas. De las numéricas se puede observar los resultados en la figura 5.5 es la comparativa de los dos estados con la densidad del núcleo; al ser semejante a un histograma su amplitud está fuertemente relacionado con la cantidad de datos por lo que se detallar es el comportamiento de la curva. La figura 5.5a los datos se concentran en entre los 7 y 20 km, lo cual es un espectro alto, sin embargo, para el caso de los retiros también hay acumulación entre los 0 y 10 km por lo que infiere que aún estando muy cerca de la sede no aporta en la decisión de quedarse lo que va en contra de lo planteado en la investigación de Raj Chakraborty et al [15], en donde esta fue una variable con mucha influencia para el abandono. Ahora, la característica de número de personas que dependen del trabajado en la figura 5.5b, indica que en ambos casos la concentración esta entre cero, una o dos personas, no indica una tendencia para diferenciar las categorías de Activo o Retiro.



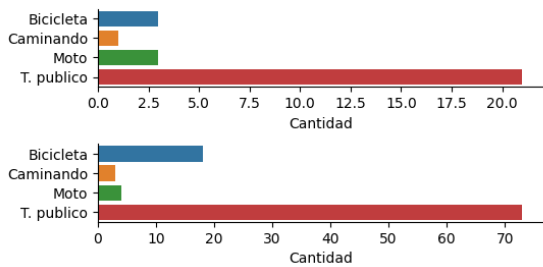
(a) Distancia de residencia a lugar de trabajo



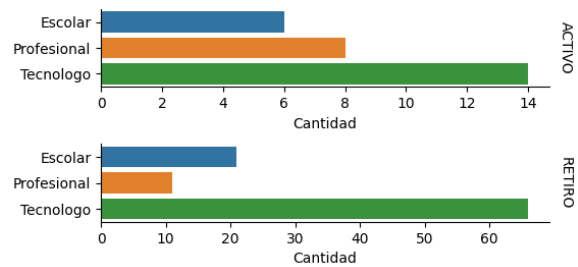
(b) No de personas que dependen economicamente del asesor

Figura 5.5: Densidad del núcleo

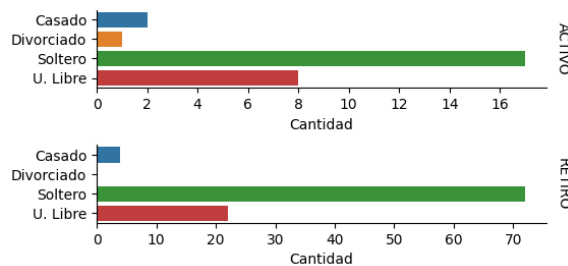
Para el caso de las variables categóricas, la mejor forma de visualizar su comportamiento es con histogramas (figura 5.6). Para las tres categorías: Medio de transporte (figura 5.6a), escolaridad, (figura 5.6b) y estado civil (figura 5.6c); en ambos estados se tienen el mismo comportamiento, en la primera categoría predomina que el medio de transporte es el transporte público, en la segunda categoría el consenso se encuentra que la mayoría de los asesores tienen un nivel de educación de tecnólogo y la última, la tendencia es que tiene un estado civil de solteros.



(a) Medio de transporte



(b) Escolaridad



(c) Estado civil

Figura 5.6: Distribución variables categóricas

Al terminar con la exploración de la base de datos socio-demográfico se tiene un panorama de cuál es la población que se encuentra trabajando como asesor de *call center*. Sin embargo, los datos socio-demográficos no dan alguna tendencia entre sus categorías y el retiro de un asesor. Al terminar con esta etapa queda la información de 126 asesores en 22 características.

### 5.1.2. Bases de datos operación

Esta base de datos tiene un tamaño de 1'456758 instancias para todo el año 2021 y 1'1622123 para el 2022 hasta de 15 de octubre del 2022 y cada instancia esta descompuesta en 30 características iniciales. Siguiendo el diagrama de bloques del pre-porcesamiento de esta base se filtra para omitir los datos del robot, se elimino los datos nulos y se selecciono las columnas de interés, siendo así quedan 1'627088 instancias (juntando ambos años) con 7 características. Algunas de estas características son:

- Fecha y hora de la gestión
- Tiempo de conversación
- Tiempo de espera
- Encuesta de satisfacción
- Encuesta de solución

Como se comentó en la metodología se tomaron cuatro asesores aleatorios, la única condición es que tenga notas de las encuestas para ver su comportamiento a la hora de dar una respuesta al cliente. Ahora, para la selección de los meses se tomará el mes de agosto del 2022, este debido a que se encuentra dentro de la ventana de los últimos tres meses trabajados para los agentes que se encuentren en activo y los otros dos serán el mes marzo 2022 y enero del 2021, ya que según la figura 5.7, son los que tienen mayor participación en retiros, el cual es el foco de esta investigación.

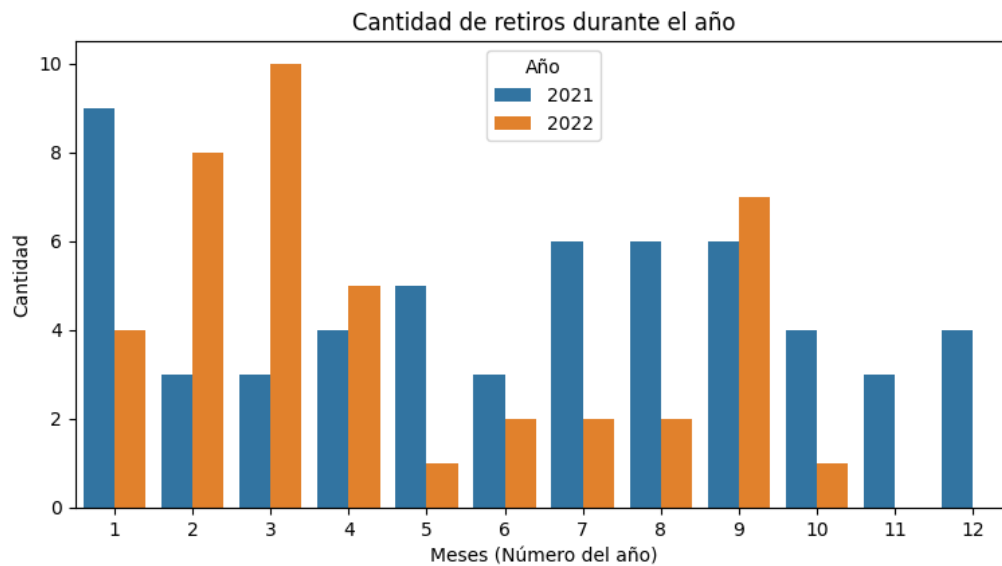


Figura 5.7: Retiros durante el año

Siendo así, se continua con la estrategia establecida, la división del mes en 4 secciones iguales como se indica en la tabla 4.2, esto permite comparar a los agentes en periodos iguales de tiempo y ver su comportamiento. La figura 5.8 es el ejemplo de aplicar este método en el mes de enero 2021, donde las líneas representan la media de la sección y el área sombreada su desviación estándar. Ahora agrupando esta serie de datos por semana los resultados se observan en la figura 5.9 para la productividad/hora.

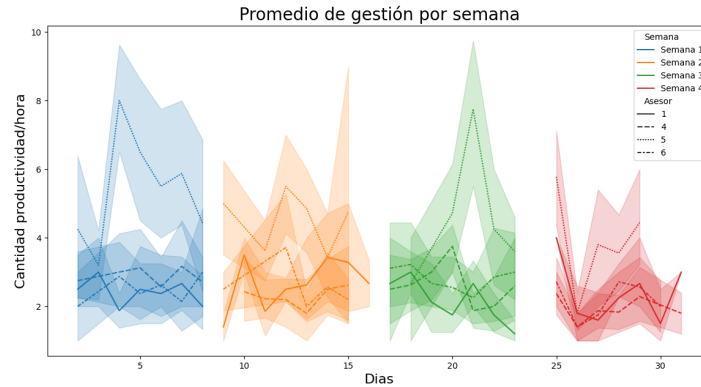


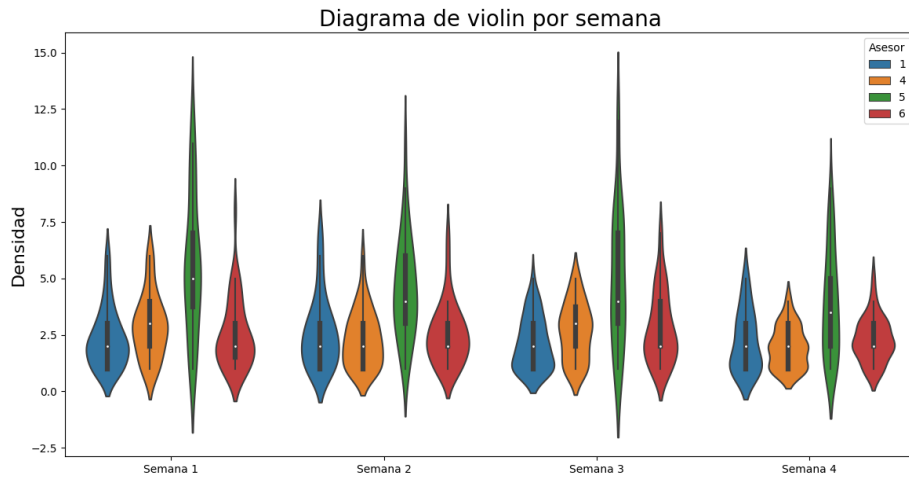
Figura 5.8: Productividad promedio hora en Enero 2021

En el diagrama de violín se puede observar que en enero (figura 5.9a) los datos se encuentran con más concentrados hacia los 2.5 interacciones/hora, excepto por el agente 5, que su promedio de gestiones es más alto y alcanza valores superando las 10 gestiones; por otro lado en marzo (figura 5.9b) el promedio de casos es menor que en enero pero el comportamiento entre los agentes es casi el mismo durante las semanas, menos por el agente 3 que tiene una distribución mayor, como en el caso del mes anterior se observa gracias a la estimación del kernel que tiende hacer más alta y su media también; por ultimo en agosto, que es el mes más cercano a la actualidad, los datos empiezan a hacer más uniformes, con una media casi igual en todas las semanas una distribución entre los 1 y 4 gestiones, menos por el caso del agente 9 en la semana 3, que parece ser un caso atípico por poco grosor del violín. Por otro y replicando la misma estimación la figura 5.11 es TMO/hora.

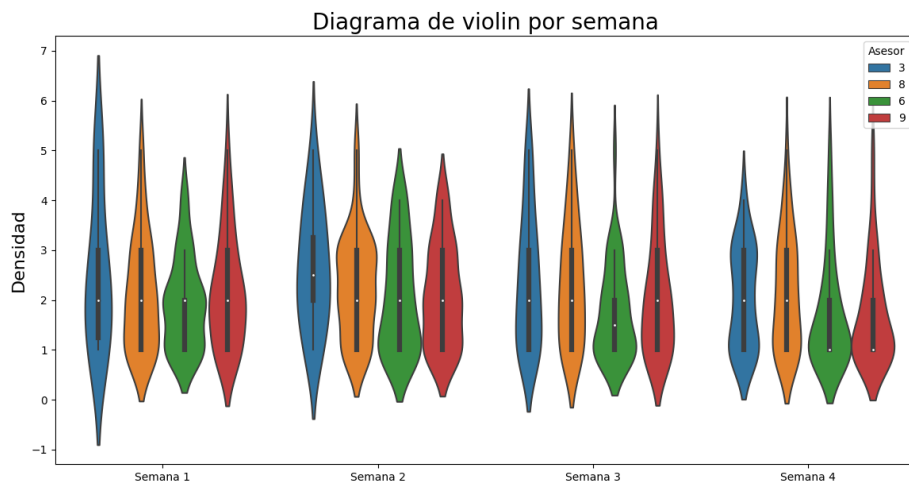
En ella se puede observar mayor diferencia entre los asesores y las semanas, excepto en la figura 5.13b, donde es muy parecido la distribución de todos sus agentes en la semana, pero en la figura 5.13a se ve como los datos se concentran en tiempos bajos de gestión menos la última semana, donde el agente 4 no tiene una distribución homogénea, sino que se concentra en datos altos o bajos.

Agente	Total interacciones mes	Semana 1					Total interacciones
		No. días trabajados	Promedio interacciones hora	Desviación estándar interacciones hora	Promedio TMO hora	Desviación estándar TMO hora	
1	860	7	4.53	2.77	4.46	2.57	240
4	1005	6	4.56	2.11	2.05	3.22	251
5	896	6	4.8	2.46	2.45	3.44	240
6	1017	6	6.33	2.79	1.79	2.01	285

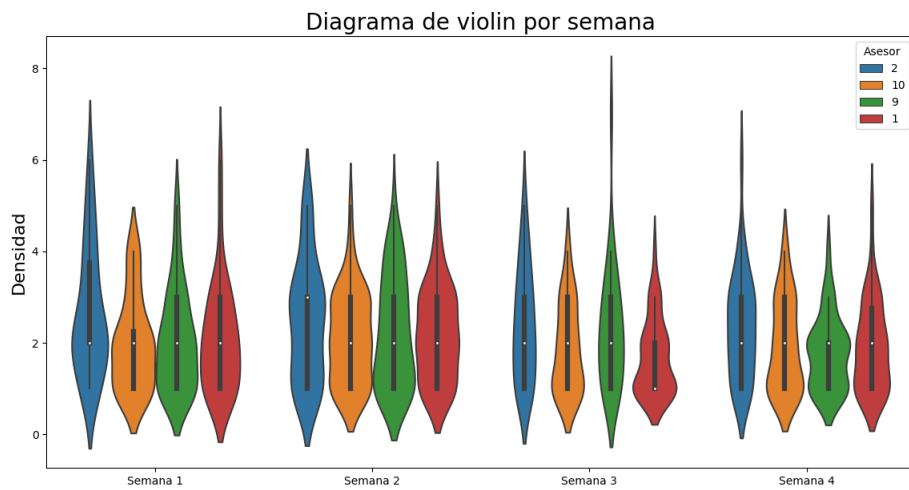
Tabla 5.2: Características extraídas por hora



(a) Enero 2021



(b) Marzo 2022



(c) Agosto 2022

Figura 5.9: Diagrama de violín: Gestión agentes por hora

Lo siguiente es replicar el mismo procedimiento, pero no por hora sino por promedio de gestiones al día, ya que, aunque en la práctica se estima multiplicando el dato de interacciones/hora por las ocho de trabajo, este no es un dato exacto debido a que durante el día se tiene pico de gestión de tráfico y estas variaciones afectan el comportamiento operativo. La gráfica 5.10 es una tabla comparativa del tráfico en el mes de julio del 2022 que fue compartida por parte de Millenium BPO para dar sustento a lo mencionado [12].

Hora del día	Interacciones	% interacciones	Bot	Asesores
00	135	0,18558997	135	-
01	63	0,086608653	63	-
02	27	0,037117994	27	-
03	22	0,039867475	22	-
04	29	0,04480279	29	-
05	76	0,45228963	76	-
06	329	2,464909748	314	15
07	1793	5,958125404	665	1128
08	4334	8,845080491	1516	2818
09	6434	10,13733658	2718	3716
10	7374	9,859638993	3378	3996
11	7172	8,659490521	3204	3968
12	6299	7,961122338	2904	3395
13	5791	8,917941738	2126	3665
14	6487	8,825834124	2590	3897
15	6420	7,750787039	3276	3144
16	5638	6,678489435	3466	2172
17	4858	4,811591812	2941	1917
18	3500	3,623816005	2328	1172
19	2636	1,941133611	1894	742
20	1412	1,215270618	1412	-
21	884	0,889457115	884	-
22	647	0,391801048	647	-
23	285	0,030244291	285	-
No data	96	0,13197509	55	41

Figura 5.10: Hora del día y tráfico en el canal de chat

<i>Semana 1</i>				
Agente	Promedio interacciones hora	Desviación estándar interacciones hora	Promedio TMO hora	Desviación estándar TMO hora
1	34.29	11.23	3.0	4.44
4	41.83	6.44	2.16	4.55
5	40.0	4.58	2.69	4.85
6	47.5	13.25	2.34	3.89

Tabla 5.3: Características extraídas por día

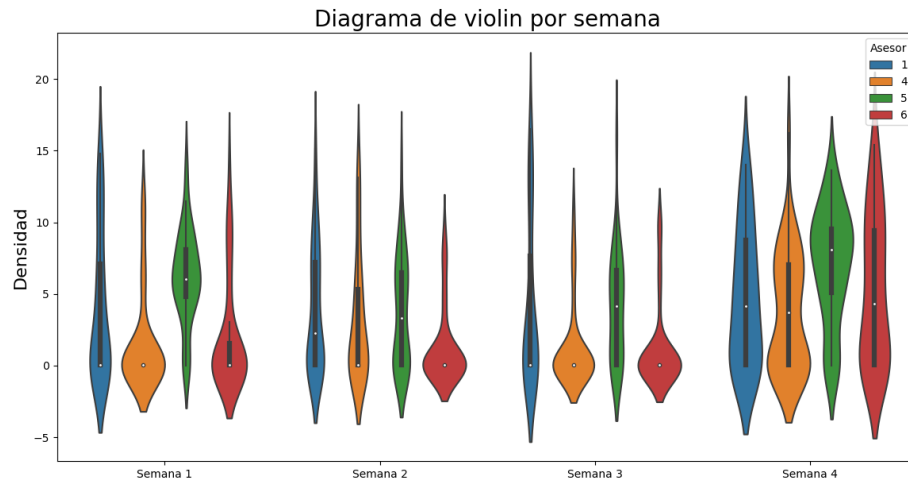
Por último, promedio horas de trabajo día (PHT) y con ello se calcula el ausentismo y las horas extra bajo la ecuación 5.1 y el total de interacciones calificadas, su nota promedio en los ítems de satisfacción y solución, teniendo en cuenta que estas tienen un rango de 1 - 4 puntos. Continuando la tabla 5.4 es un ejemplo de estos resultados.

$$\begin{aligned}
 \text{Ausentismo} &= 8 - PHT \quad \text{si } PHT < 8 \\
 H.Exta &= PHT - 8 \quad \text{si } PHT \geq 8
 \end{aligned}
 \tag{5.1}$$

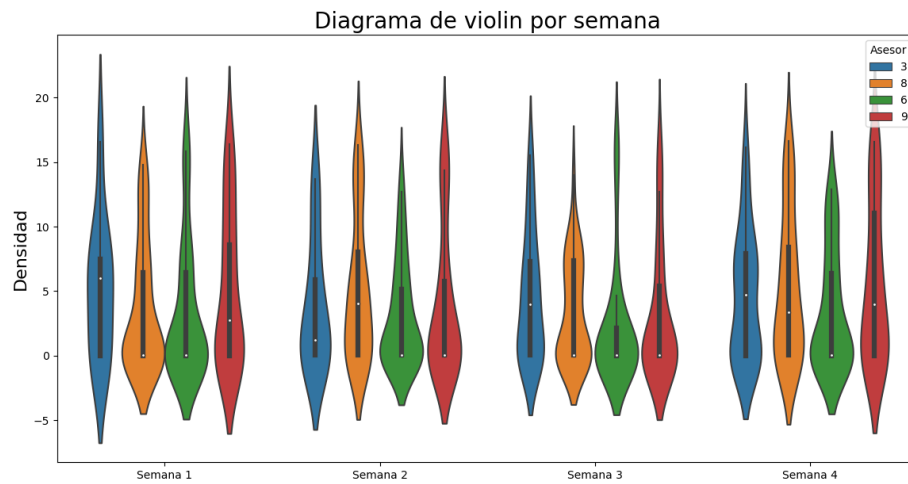


<i>Semana 1</i>						
<b>Agente</b>	<b>Ausentismo</b>	<b>Horas extra</b>	<b>Promedio Calificación: Solución</b>	<b>Total interacciones Calificación: Solución</b>	<b>Promedio Calificación: Satisfacción</b>	<b>Total interacciones Calificación: Satisfacción</b>
1	0.0	1.3	3.34	115	3.43	115
4	0.0	0.36	2.54	95	2.7	95
5	0.13	0.0	2.27	92	2.35	92
6	0.67	0.0	2.81	115	2.98	115

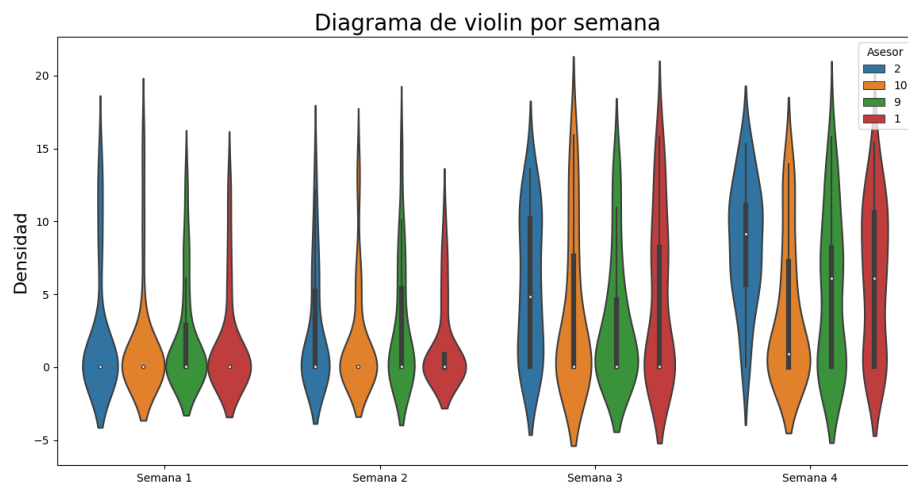
Tabla 5.4: Características extraídas por semana



(a) Enero 2021

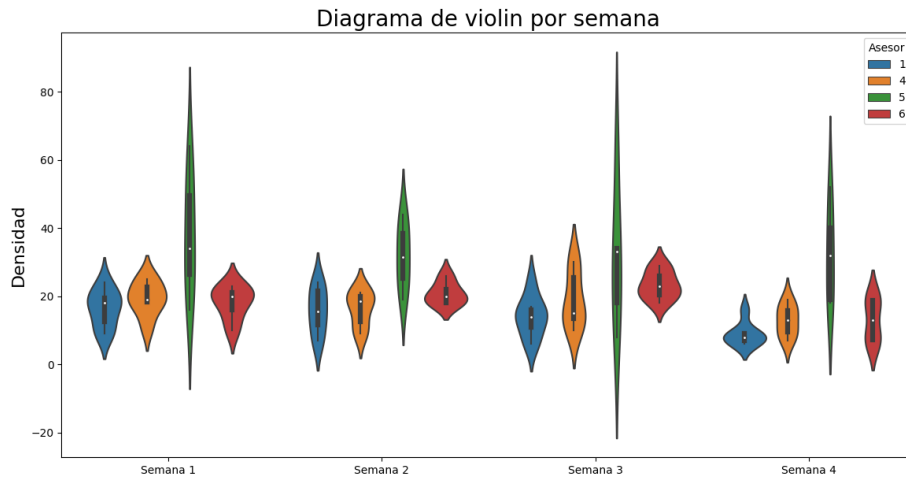


(b) Marzo 2022

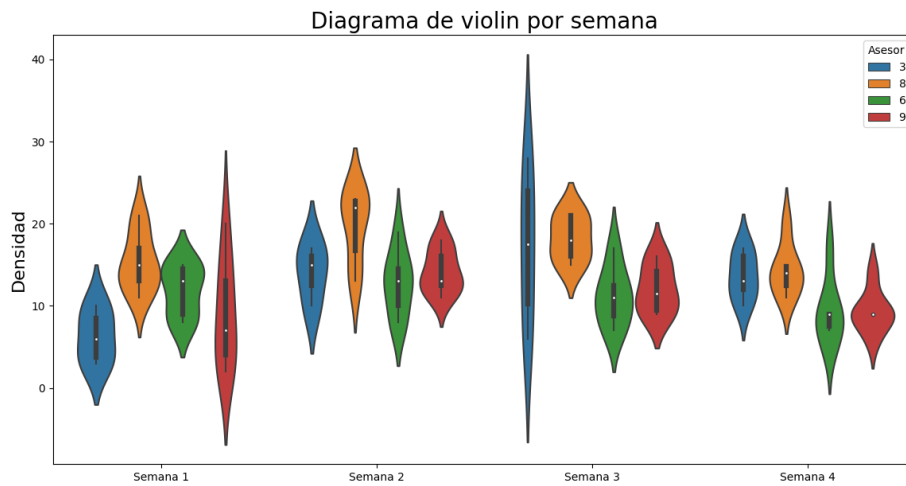


(c) Agosto 2022

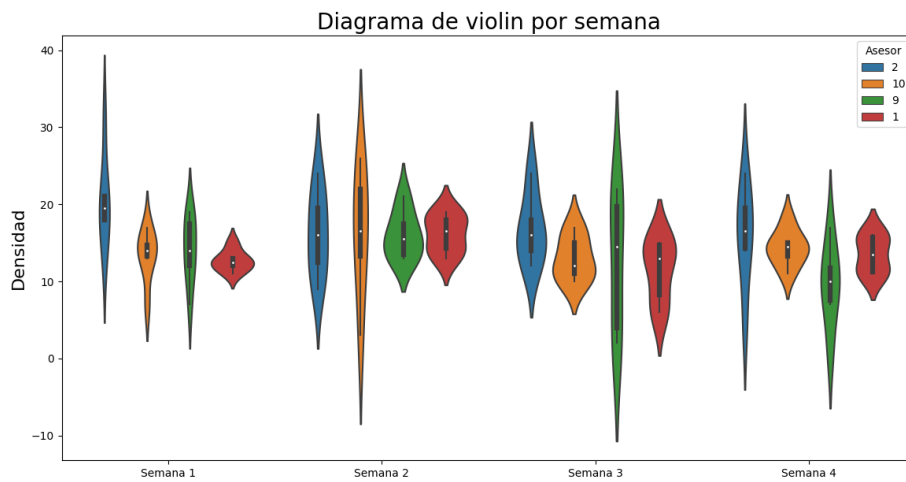
Figura 5.11: Diagrama de violín: TMO agentes por hora



(a) Enero 2021

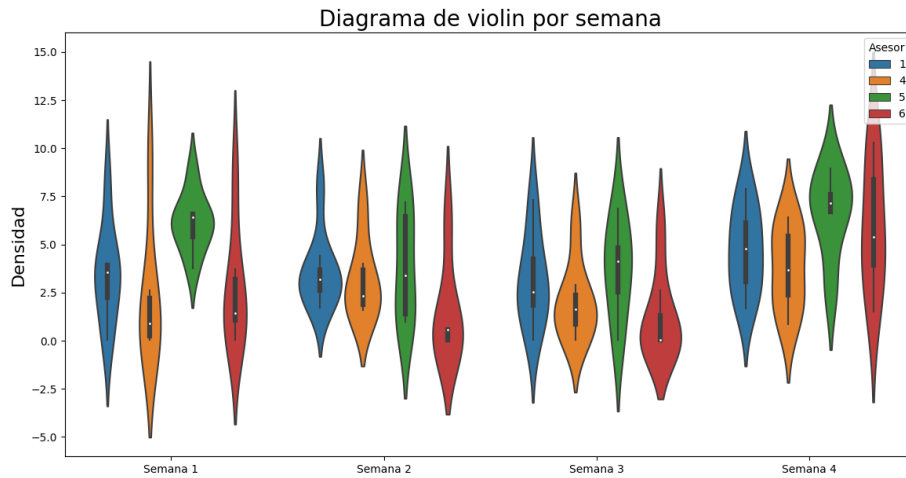


(b) Marzo 2022

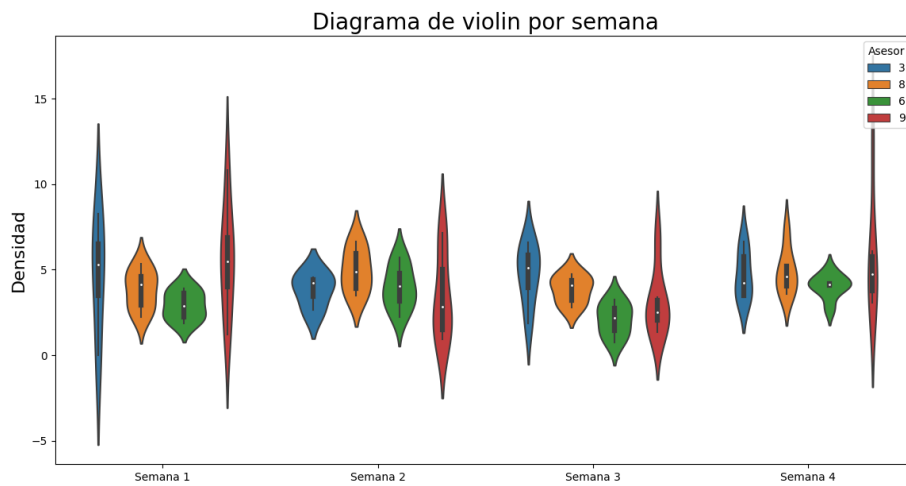


(c) Agosto 2022

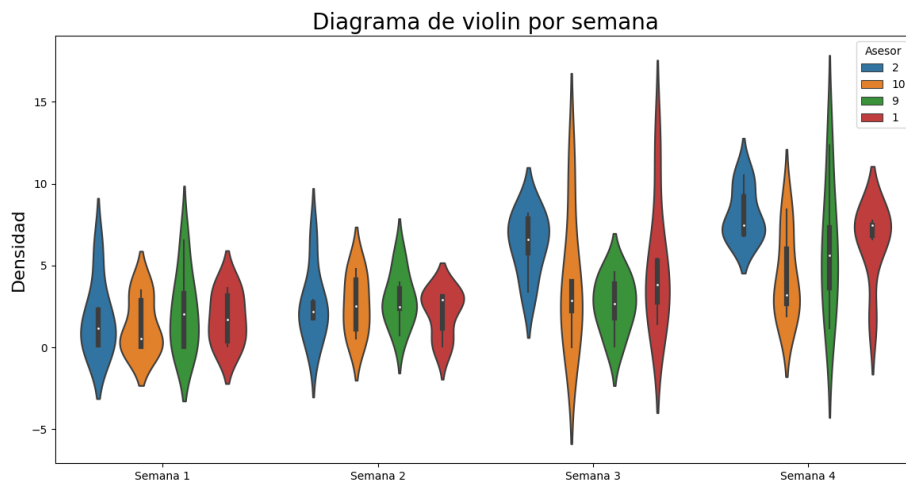
Figura 5.12: Diagrama de violín: Gestión agentes por día



(a) Enero 2021



(b) Marzo 2022



(c) Agosto 2022

Figura 5.13: Diagrama de violín: TMO agentes por día

Por lo que al final de la extracción de características, se tiene por asesor un vector con 73 características en un mes.

### 5.1.3. Matriz resultante

En esta sección se encuentra el resultado de la selección de participantes y el cruce de la información entre ambas bases de datos, lo primero se busca extraer los tres meses consecutivos de la última gestión del asesor para describir el comportamiento teniendo en cuenta si es un asesor activo o retirado.

Siendo así, se extrae la información de 58 asesores (44 asesores retirados y 14 asesores activos) en sus tres ventanas de tiempo, es decir que 68 asesores no se encontró información relacionada de tres meses consecutivos en la base operacional. Esto es debido a que 30 de ellos no cumplieron con el mínimo de meses y los otros se puede explicar en la préstamo interno entre segmentos, esto hace referencia a que dentro de la misma campaña se rota a sus agentes por los diferentes segmentos si se necesita del personal, por lo que con frecuencia un asesor salga de su segmento contratado por un tiempo sin ser modificado en la base socio-demográfica, esto es otro síntoma de la alta rotación del personal, por ello la empresa adopta estas estrategias para suplir el personal y cumplir con los objetivos, por eso se resalta el hecho que fueran meses de gestión consecutiva [12].

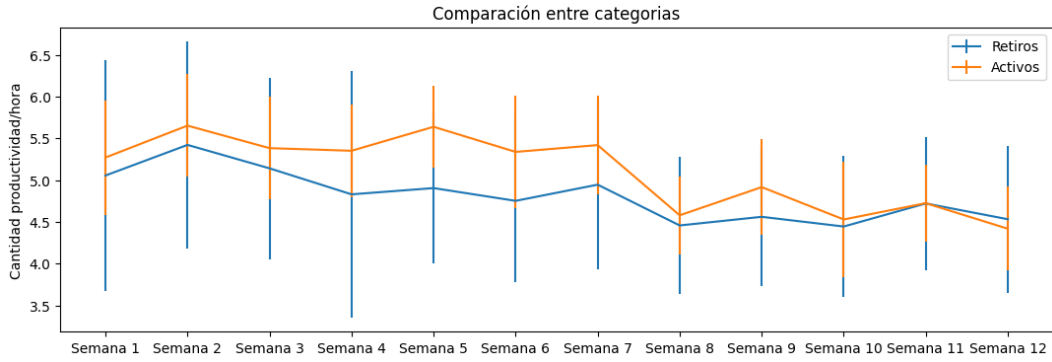
Al tener 3 ventanas de tiempo con 73 características cada una, se unieron todas en un único vector por asesor, dejando así un vector de 220 características por cada asesor (instancia), siendo una de estas características la etiqueta de **Activo** o **Retiro**.

## 5.2 Correlación de datos

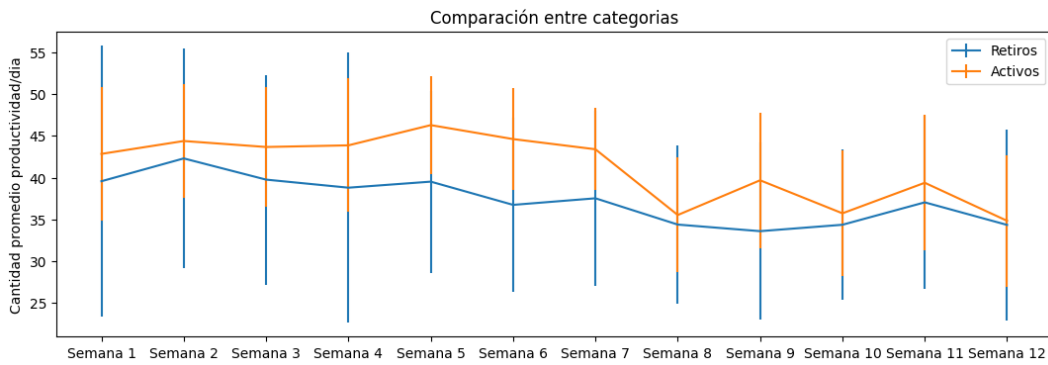
### 5.2.1. Análisis exploratorio de los datos

Para el análisis exploratorio se utiliza la matriz resultante, con esta información se visualiza las más relevantes o que presentaron mayores cambios entre los estados en sus 12 semanas. Estas son:

- **Productividad:** Las figuras 5.14a y 5.14b muestran la distribución media de los datos de todos los agentes en las 12 semanas en productividad hora y día respectivamente. En estas su línea central es el promedio y sus barras verticales es la desviación de cada semana, en ellas se puede observar que si hay una separación a nivel de productividad de los asesores que se retiraron (naranja) con los activos (azul).



(a) Productividad/hora agentes



(b) Productividad/día agentes

Figura 5.14: Comparación en productividades

Siendo así, los agentes que se encuentran en activo realizan aproximadamente 50 interacciones diarias mas que los asesores en sus últimos meses antes de tomar la decisión de renuncia, tal como se describe en la figura 5.15

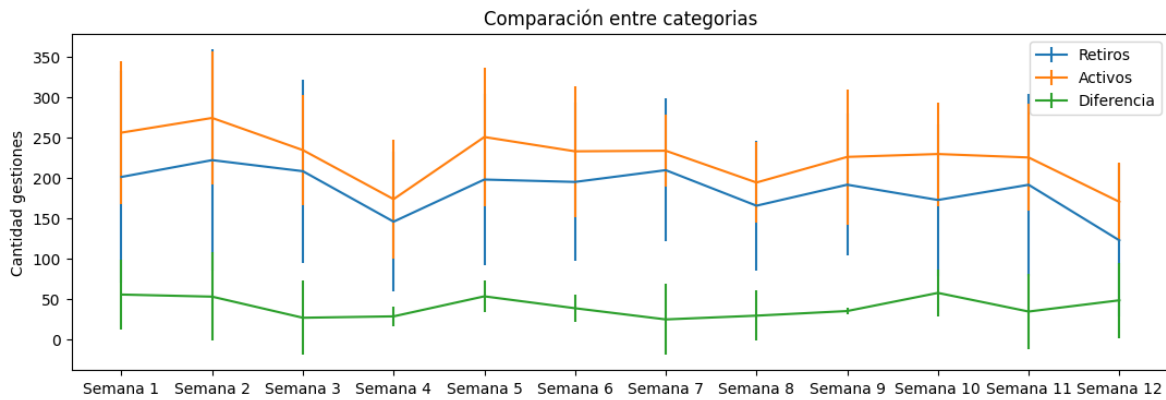


Figura 5.15: Total de interacciones

- Horas de trabajo: En este caso la categoría de agentes activos tiene una menor dispersión en las horas trabajadas, se encuentran menores a las ocho hora en promedio y es lo que

se espera registrarse, ahora bien, sobre la misma gráfica (figura 5.16) se observa como en el caso de los retiros es mucho menor, la semana nueve es un ejemplo.

Esto se puede corroborar con la categoría de ausentismo donde para los asesores retirados su promedio de ausentismos en la semana es de 1 hora (figura 5.17).

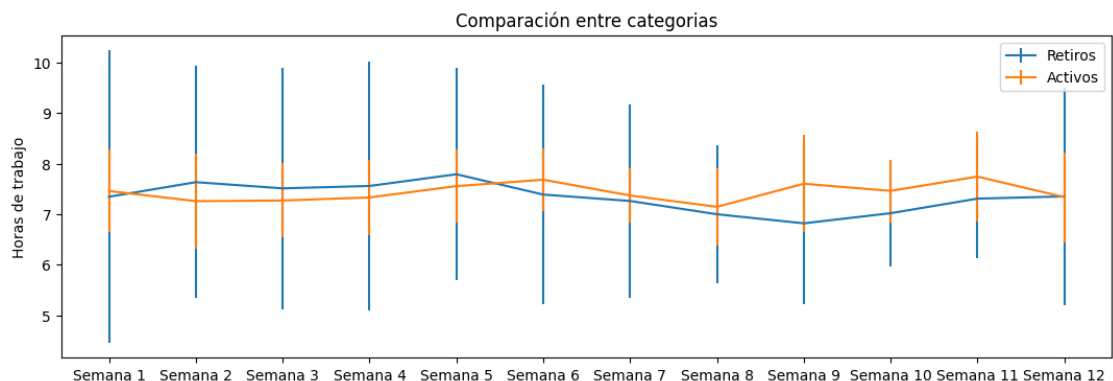


Figura 5.16: Total horas de trabajo por semana

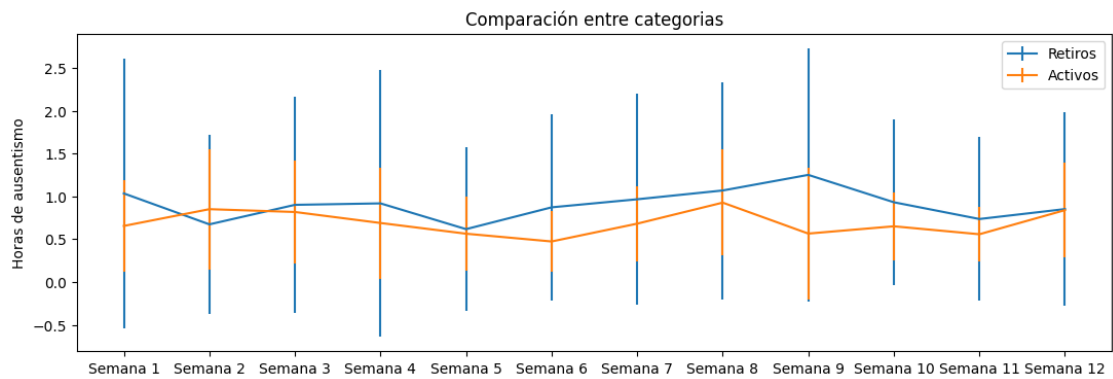
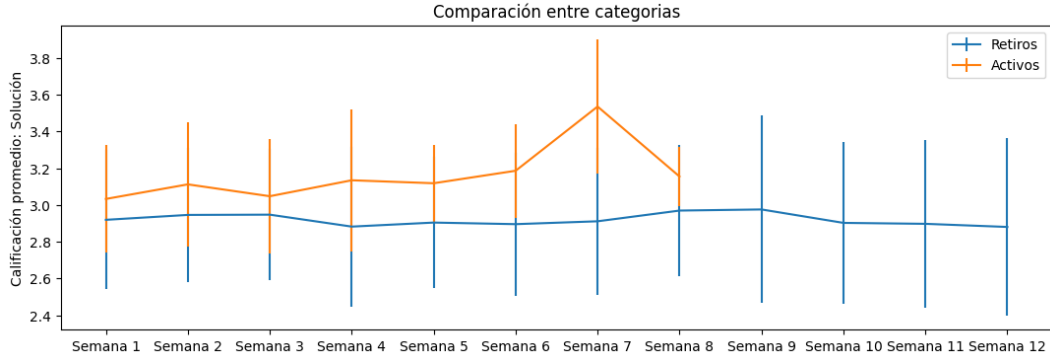
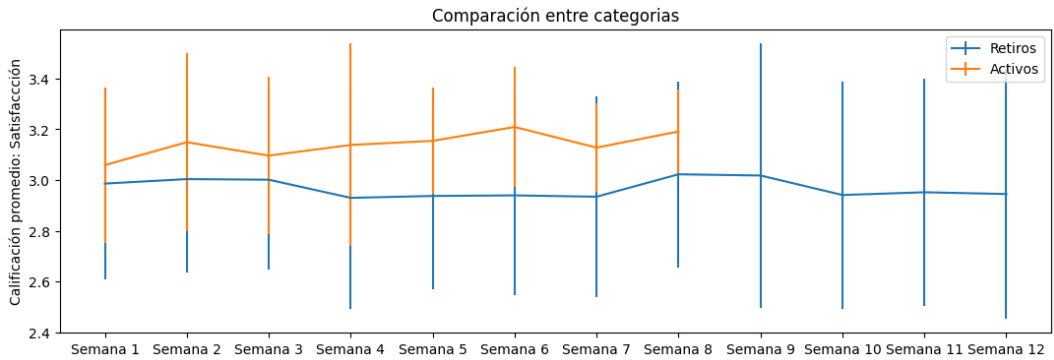


Figura 5.17: Ausentismo por semana

- Encuestas calificaciones: Esta ultima es la comparación entre las calificaciones expuestas del usuario con respecto a su nivel de servicio, estas se dividen en solución brindada y la satisfacción de la respuesta. Las gráficas 5.18a y 5.18b reúnen la información del promedio de calificación a los usuario, teniendo en cuenta que es de 1 a 4 puntos. Teniendo en cuenta que a finales de agosto Icetex retiro estas en cuestras en los canales de chat y telefónico [12].



(a) Solución



(b) Satisfacción

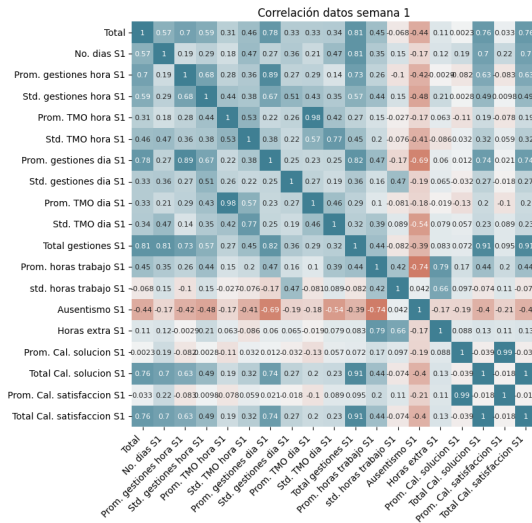
Figura 5.18: Encuestas nivel de servicio

De la figura 5.24 se puede observar que aunque no sea por mucho la diferencia  $\pm 0,3$  puntos, la tendencia es que las personas activas tengan mejor nota que las retiradas.

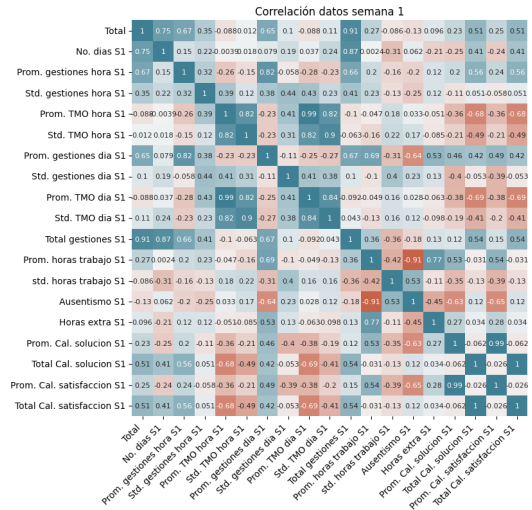
Todos los anteriores comparaciones van de acuerdo a lo expuesto por Littlewood, donde los asesores cerca a su retiro bajan su productividad y buenas practicas laborales en comparación don los asesores que están dispuestos a continuar [3]. Por ello al visualizar sus variables y comprarlas existe esa tendencia a lo largo de los tres meses.

Ahora bien, las figura 5.19a y 5.19b son las matrices de correlación de *Person*, en la primera semana para cada estado. Solo se tomo la primera ya que las demás son las mismas características en diferentes instantes de tiempo. Si se observa altas correlaciones en variables que su aumento depende de otra, ejemplo, la cantidad de casos con la cantidad de calificaciones, sin embargo, estas si no tienen que ver con el comportamiento de la nota dada por el usuario. Y así hay otros ejemplos, las horas trabajadas con el ausentismo, TMO/hora con el TMO/día y así. Sin embargo, al comprarlas entre ellas, las métricas de los retirados tienen mayores relaciones entre ellas, por ejemplo las calificaciones tiene mayor impacto con el resto de variables, ahora esto también se da porque el ultimo mes de los asesores activos no tuvieron calificaciones (figura 5.18a y 5.18b). Por ello, todas las variables son consideradas relevantes para ser llevadas a los modelos de predicción de renuncia.





(a) Retiro



(b) Activo

Figura 5.19: Matrices de correlación

## 5.2.2. Exploración con Inteligencia Artificial

En la anterior sección se demostró que si diferencia en los últimos tres meses de los KPI's de los agentes antes de su retiro, ahora bien, aparte del análisis exploratorio en las características mas relevantes, se busco dar una exploración con mayor impacto en los datos y demostrar si los datos por si mismos pueden representar la brecha entre ambas categorías tiendo en cuenta todas sus variables de la matriz resultante (219 características). Por ello se empelará algoritmos de aprendizaje no supervisados, como lo es *Kmeans* y mapa auto-organizado con redes neuronales.

- *Kmeans*:

La figura 5.20 indica el resultado luego de las diez iteraciones. Con ella se puede determinar que al algoritmo agrupa los datos de los asesores en tres segmentos.

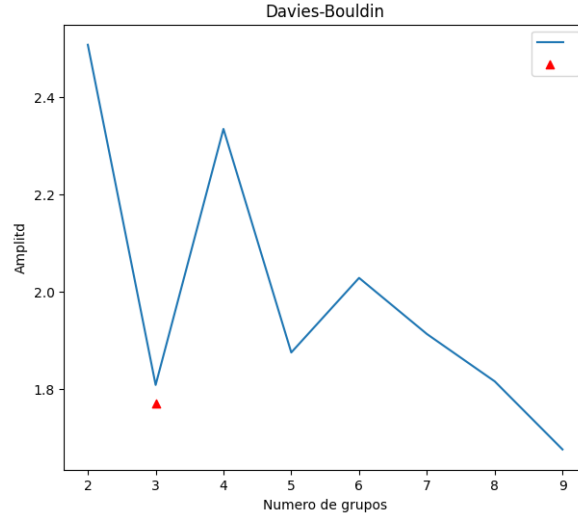


Figura 5.20: Gráfica de Davies-Bouldin

Ahora bien, el contraste de los estados real con la encontrada con el agrupamiento de *Kmeans* se encuentra en la tabla 5.5, en ella se agrega el porcentaje que ocupa el estado de retiro o actual por cada grupo.

Estado cluster	Estado real	Cantidad	% de ocupación estado real en grupo
0	Activo	2	9.09
	Retiro	20	99.91
1	Retiro	20	100
2	Activo	12	75.00
	Retiro	4	25.00

Tabla 5.5: Kmeans: Relación estados reales de la base con las agrupaciones del cluster

Aunque el *Kmeans* no clasifico la información en dos grupos exactos, los datos agrupados demuestran ser lo suficientemente diferentes para que no se mezclen, en especial con los de retiro que es el foco de la investigación.

Siendo así, todos las 219 características siguen siendo relevantes para la predicción de la renuncia.

- Mapa auto-organizado con redes neuronales:

Ahora, para la exploración no supervisada con SOM, se dejaron los parametros de defecto que otorga la libreria *SimpSOM*, y para los hiper-parámetros necesarios es decir el tamaño del mapa se utilizo la ecuación 5.2, en ella el único valor arbitrario es el largo del mapa que por defecto son 10 neuronas. Siendo así la relación de su componentes principales fue de 170 es decir que el mapa sera de 17 de ancho por diez de alto, con un total de 170 neuronas y tasa de aprendizaje del 0.1 a 100 iteraciones.

$$w = \frac{PCA_1}{PCA_2} * 10; \quad h = 10; \quad \text{Tamaño} = w * h \quad (5.2)$$

$PCA_1$  hace referencia a la primera componente resultante de realizar PCA a base de datos, lo mismo para  $PCA_2$  pero con la segunda componente.

La figura 5.21 es el resultado de este método y como se indica en la revisión literaria, aunque proporcione nuevos panoramas y en especial al proyectar información sobre ella, sin embargo se necesita un nivel de experiencia alto para detectar estos nuevos conocimientos de forma directa. La figura mencionada no es muy clara con las agrupaciones de neuronas que se activaron el sobre la zona inferior medial es donde se puede ver un agrupamiento en la activación. Ahora bien, este método no lineal permite también mostrar los resultados en forma de clúster.

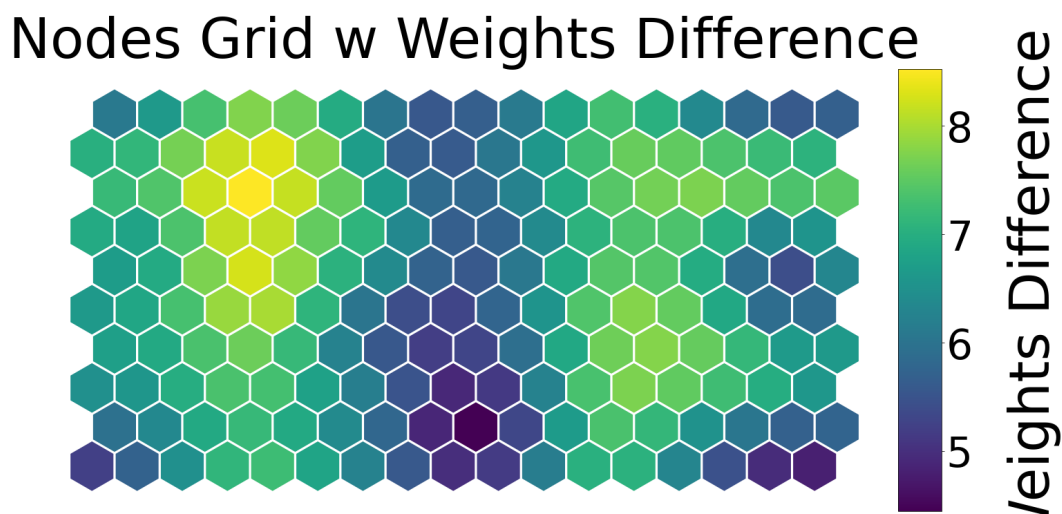


Figura 5.21: Mapa auto-organizado

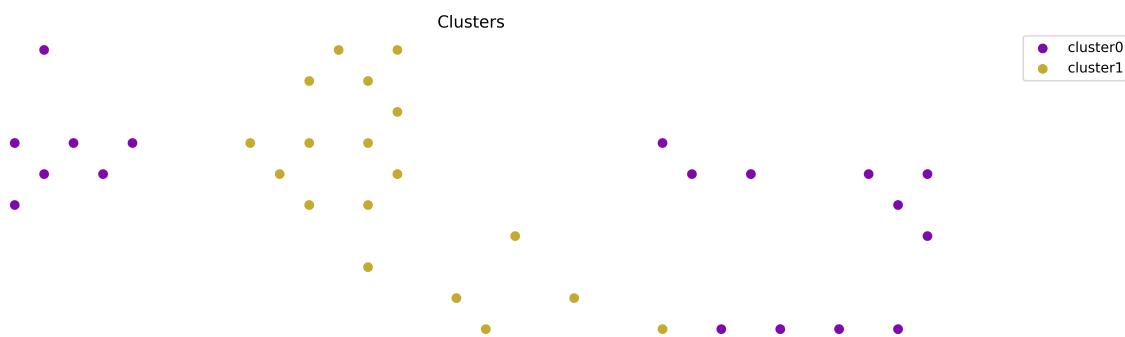


Figura 5.22: Cluster seleccionado por SOM

La figura 5.22 es el agrupamiento que extrae este método, el cual da que encontró dos grupos óptimos sobre la base de datos operacional, los cuales son visualizados de la misma manera que en *Kmeans* en la tabla 5.6. En esta se observa como el algoritmo

aunque detecto únicamente los dos grupos, se confunde mucho mucha en la división de estos datos.

Estado cluster	Estado real	Cantidad	% ocupación estado real en cluster
0	Activo	10	29.41
	Retiro	20	70.59
1	Retiro	20	62.50
	Activo	12	37.50

Tabla 5.6: SOM: Relación estados reales de la base con las agrupaciones del cluster

Al finalizar de realizar el entrenamiento no supervisado en especial al revisar los resultados al implementar *Kmeans* sustenta que los datos empleados separar los dos estados, mientras que SOM aunque es una herramienta mas sofisticada para esta investigación no encuentra de forma efectiva la separación de los estados.

### 5.3 Modelos de predicción

Luego de comprobar que en la exploración de los datos muestran que existen claras diferencias entre los estados y se verifica con la exploración no supervisada que la calidad de la información es suficiente para que un computador pueda diferenciar y separar sus estados. Por lo que se continua con el entrenamiento de modelos de clasificación y determinar cual es el mejor en la predicción de renuncia de un asesor.

Lo primero es la unificación final de las matriz resultante operativa y la socio-demográfica, siendo así queda una matriz resultante de 240 características por agente.

Ahora, se hará la prueba en tres modelos de aprendizaje de maquina: máquinas de vectores de soporte (SVM), bosques aleatorios (RF) y redes neuronales artificiales (ANN) usando el método de validación cruzada con cinco grupos (*folds*) y empleando la búsqueda exhaustiva de los hiper-parámetros. La figura 5.23 es el resultado luego del entrenamiento , junto con la la tabla 5.7, esta reúne los resultados de la exactitud (*accuracy*) de las cinco *folds* en cada modelo.

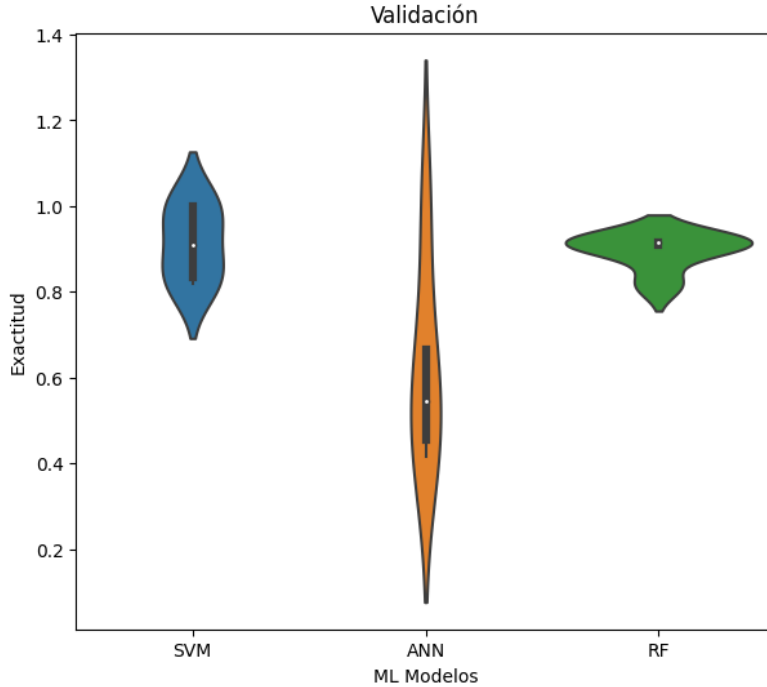


Figura 5.23: Diagrama de violín, resultados modelos

Folds	SVM	ANN	RF
0	<b>1.000</b>	<b>1.000</b>	<b>0.917</b>
1	0.833	0.666	0.917
2	1.000	0.416	0.916
3	0.818	0.545	0.818
4	0.909	0.454	0.909
Promedio ± Desviación	0.912 ± 0.087	0.617 ± 0.235	<b>0.895 ± 0.043</b>

Tabla 5.7: Resultados modelos en validación

Al comparan los tres modelos de la figura 5.23 SVM es el modelo con mejor promedio de exactitud, sin embargo su silueta es mas alargada que RF lo que indica que la distribución de sus datos no esta centralizada sobre su promedio como si pasa en RF, donde su silueta es ancha pero no larga, de hecho tampoco se aprecia la diagrama de caja dentro de ella. Por ultimo, el modelo con ANN es quien presenta resultados mas bajos en exactitud con un 61.7% de acierto. Ahora bien, se decide que el mejor modelo es el de RF, aunque tiene un promedio de exactitud menor a SVM, su dispersión es 4% menor, lo que indica que en siguientes ejemplos tenga menor margen de error. A continuación, se encuentran las figuras de la matriz de confusión de entrenamiento (figura 5.24a) y validación (figura 5.24b).



Figura 5.24: Matrices de confusión

Las métricas presentadas anteriormente para el modelo de RF se obtuvieron bajo los siguientes hiper-parámetros luego de la búsqueda exhaustiva:

- *Class wight* = balanced
- *Max depth* = 10
- *max features* = auto
- *n. estimators* = 20

## Capítulo 6

# CONCLUSIONES

Al completar esta investigación se logró la minería adecuada de los datos, preprocesando y transformando sus variables adecuadamente, dando como resultado una matriz estructurada de 240 características en 58 asesores para el canal de chat, donde se tiene información de variables socio-demográficas junto con sus variables de operación durante 12 semanas consecutivas laborales.

Adicional, al comparar sus características entre los dos estados (Activo y Retirado), como se muestra en las figuras 5.14, 5.15, 5.16, 5.17 y 5.24; donde en ellas se puede analizarla diferencia significativa en el rendimiento de las personas que estaban a punto de renunciar y las asesores activos, por lo que el comportamiento de renuncia psicología, es decir disminuir la productividad, eficiencia, eficacia y calidad en la gestión a un mínimo posible de evidenciar casos de retiro.

Teniendo en cuenta los resultados de la investigación se tomó la decisión de emplear algoritmos de entrenamiento no supervisados con el fin de evaluar la diferencia que puede existir entre los grupos, es decir que sus datos puedan verse separados de manera automática por algoritmos. El resultado de este proceso es satisfactorio ya que al implantar Kmeans se separó la muestra en tres grupos diferentes, sin embargo, la tasa de confusión de las categorías establecidas en estos grupos fue de 74.25 % como se observa en la Tabla 5.5. Siendo así, que los algoritmos de inteligencia artificial son capaces de dividir esta información encontrando tendencias en los datos, situaciones no tan evidentes por la dimensionalidad de sus características.

Finalmente, se implementaron tres modelos de predicción con entrenamiento supervisado para la predicción de la rotación del personal, estos fueron bosques aleatorios (RF), máquinas de vectores de soporte (SVM) y redes neuronales artificiales (ANN), de los cuales los resultados indican una exactitud del  $89,5\% \pm 4,3\%$ ,  $91,2\% \pm 8,7\%$  y  $61,7\% \pm 23,5\%$  respectivamente. Siendo ANN el modelo con menos exactitud por la poca cantidad de datos, se encontró que SVM tiene mejor promedio de exactitud, pero con más desviación. Por lo que el modelo con mejores métricas de predicción es Bosques aleatorios (*Random Forest*) el cual tiene menor desviación asegurando que sus resultados en las predicciones sean más certeros con el 89 % de exactitud, esto además se comparó con las investigaciones de Raj Chakraborty et al [15], Shikha N. Khera et al [16] y Vasa, Jalpesh et al [17] donde sus resultados expresan que también el modelo electo es el RF y SVM.

## Capítulo 7

# RECOMENDACIONES Y TRABAJOS FUTUROS

En el planteamiento del problema se describe que el área de DNE impulsa el uso de estas herramientas tecnológicas para la compañía de Millenium BPO, por lo que al observar buenos resultados se plantea llevar esta metodología de la exploración de datos explicada en diferentes campañas que tengan el canal de chat para extraer sus comportamientos operacionales y sus características socio-demográficas para incrementar la cantidad de instancias de la matriz resultante con el fin de realizar un re-entrenamiento del modelo de clasificación electo para seguir realizando pruebas y obtener un modelo más robusto para apoyar y solventar la esta problemática para RH.



# Bibliografía

- [1] R. Jain y A. Nayyar, «Predicting employee attrition using xgboost machine learning approach,» en *2018 international conference on system modeling & advancement in research trends (smart)*, IEEE, 2018, págs. 113-120.
- [2] Omnisalud. «Los 5 problemas de una empresa con alta rotación de personal.» Omnisalud, ed. (), dirección: <https://omnisalud.co/5-problemas-de-una-empresa-con-alta-rotacion-de-personal/>.
- [3] H. Littlewood, «Efecto Moderador de Abuso, Benevolencia y Afabilidad en el Modelo de Evitación del Trabajo,» Tesis doct., Tesis doctoral, Universidad de las Américas, 2012.
- [4] H. F. Littlewood Zimmerman, «La paradoja del Compromiso,» *Investigación administrativa*, vol. 49, n.º 126, 2020.
- [5] D. Colombo. «Renuncia psicológica: cuando el empleado “se va antes de tiempo.» Daniel Colombo, ed. (), dirección: <https://www.cronista.com/clase/trendy/Renuncia-psicologica-cuando-el-empleado-se-va-antes-de-tiempo-20191129-0006.html>.
- [6] SPHRE. «13 señales de un empleado que está por renunciar.» SPHRE, ed. (), dirección: <https://spheremodel.com/blog/survey-13-senales-renunciar/>.
- [7] R. A. Chevez Mustto y L. K. Inga Cerda, «Rotación de personal en una empresa de call center en la ciudad de Lima, Perú, en el periodo 2016-2017,» 2021.
- [8] L. Saborio Morales y L. F. Hidalgo Murillo, «Síndrome de burnout,» *Medicina Legal de Costa Rica*, vol. 32, n.º 1, págs. 119-124, 2015.
- [9] Y.-R. Lee, J.-Y. Lee, J.-M. Kim, I.-S. Shin, J.-S. Yoon y S.-W. Kim, «A comparative study of burnout, stress, and resilience among emotional workers,» *Psychiatry Investigation*, vol. 16, n.º 9, pág. 686, 2019.
- [10] BBVA. «10 funciones del departamento de recursos humanos.» BBVA, ed. (), dirección: <https://www.bbva.com/es/10-funciones-del-departamento-de-recursos-humanos/>.
- [11] UNIR. «Psicología Laboral: ¿qué es y cómo afecta al entorno de trabajo?» UNIR, ed. (), dirección: <https://www.unir.net/salud/revista/psicologia-laboral/>.
- [12] M. B. S.A. «Millenium BPO S.A.» (), dirección: <https://millenium.com.co/millenium/>.

- [13] ICETEX. «¿Quiénes somos?¿Quiénes somos?» ICETEX, ed. (), dirección: <https://web.icetex.gov.co/el-icetex/informacion-institucional/quienes-somos#:~:text=EL%5C%20ICETEX%5C%20es%5C%20una%5C%20entidad,econ%5C%C3%5C%B3micas%5C%20y%5C%20buen%5C%20desempe%5C%C3%5C%B1o%5C%20acad%5C%C3%5C%A9mico..>
- [14] Y.-R. Lee, J.-Y. Lee, J.-M. Kim, I.-S. Shin, J.-S. Yoon y S.-W. Kim, «A comparative study of burnout, stress, and resilience among emotional workers,» *Psychiatry Investigation*, vol. 16, n.º 9, pág. 686, 2019.
- [15] R. Chakraborty, K. Mridha, R. N. Shaw y A. Ghosh, «Study and prediction analysis of the employee turnover using machine learning approaches,» en *2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)*, IEEE, 2021, págs. 1-6.
- [16] S. N. Khera y Divya, «Predictive modelling of employee turnover in Indian IT industry using machine learning techniques,» *Vision*, vol. 23, n.º 1, págs. 12-21, 2018.
- [17] J. Vasa y K. Masrani, «Foreseeing employee attritions using diverse data mining strategies,» *International Journal of Recent Technology and Engineering (IJRTE)*, vol. 8, n.º 3, págs. 620-626, 2019.
- [18] C. Henao Rios, «Modelo de medición de la rotación de personal como variable de decisión estratégica,»
- [19] L. M. Monsalve Lopera, «Modelo predictivo para identificar factores que influyen en la rotación del personal en una organización del sector business process outsourcing (bpo) en medelln.,»
- [20] H. Zhang, L. Xu, X. Cheng, K. Chao y X. Zhao, «Analysis and prediction of employee turnover characteristics based on machine learning,» en *2018 18th International Symposium on Communications and Information Technologies (ISCIT)*, IEEE, 2018, págs. 371-376.
- [21] S. V. Orea, A. S. Vargas y M. G. Alonso, «Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos,» *Ene*, vol. 779, n.º 73, pág. 33, 2005.
- [22] C. S. Wickramasinghe, K. Amarasinghe, D. Marino y M. Manic, «Deep self-organizing maps for visual data mining,» en *2018 11th International Conference on Human System Interaction (HSI)*, IEEE, 2018, págs. 304-310.
- [23] A. C. de BPO. «¿Qué es el sector BPO?» Asociación Colombiana de BPO, ed. (), dirección: <https://www.bpro.org/que-es-el-sector-bpo>.
- [24] S. L. América. «¿Qué es un CRM?» Salesforce Latin América, ed. (), dirección: <https://www.salesforce.com/mx/crm/>.
- [25] TuDashboard. «30 KPIs para Call Centers ¡Mejora el nivel de tu servicio de llamadas!» TuDashboard, ed. (), dirección: <https://tudashboard.com/kpis-para-call-centers/>.
- [26] G. de barras. «Interpretación de una gráfica de barras.» (), dirección: <https://graficodebarras.com/interpretacion-de-una-grafica-de-barras>.
- [27] BYJU'S. «Box Plot.» (), dirección: <https://byjus.com/maths/box-plot/>.

- [28] T. D. V. Catalogue. «Diagrama Cajas y Bigotes.» ().
- [29] R. Wicklin. «How to visualize a kernel density estimate.» (), dirección: <https://blogs.sas.com/content/iml/2016/07/27/visualize-kernel-density-estimate.html>.
- [30] M. Waskom. «seaborn.kdeplot.» (), dirección: <https://seaborn.pydata.org/generated/seaborn.kdeplot.html#seaborn.kdeplot>.
- [31] R. CODER. «Gráfico de densidad en R.» (), dirección: <https://r-coder.com/grafico-densidad-r/>.
- [32] LabXchange. «How to Interpret Violin Charts.» (), dirección: <https://www.labxchange.org/library/items/lb:LabXchange:46f64d7a.html:1>.
- [33] T. D. V. Catalogue. «Diagrama de Violín.» ().
- [34] C. Ebert y P. Louridas, «Machine learning,» *IEEE Software*, vol. 33, n.º 5, págs. 110-115, 2016.
- [35] C. Rajendran. «Unsupervised Machine Learning (KMeans Clustering) with Scikit-Learn.» (), dirección: <https://medium.com/ascentic-technology/unsupervised-machine-learning-kmeans-clustering-with-scikit-learn-bc8895cd66a8>.
- [36] Unknown. «REDES NEURONALES - MAPAS AUTOORGANIZADOS DE KOHONEN (SOM).» ().
- [37] IBM. «Funcionamiento de SVM.» (), dirección: <https://www.ibm.com/docs/es/spss-modeler/saas?topic=models-how-svm-works>.
- [38] O. C. Carrasco. «Support Vector Machines for Classification.» (), dirección: <https://towardsdatascience.com/support-vector-machines-for-classification-fc7c1565e3>.
- [39] S. A. González. «Detección de Anomalías Cardíacas con Aprendizaje Automático (Machine Learning).» (), dirección: <http://samuelabad1991.blogspot.com/2014/02/analisis-con-maquinas-de-vectores.html>.
- [40] I. Logunova. «Random Forest Classifier: Basic Principles and Applications.» (), dirección: <https://serokell.io/blog/random-forest-classification>.
- [41] H. Y. P. Huacasi. «Bosques Aleatorios.» (), dirección: <https://medium.com/@hpumah/bosques-aleatorios-482163ace92e>.
- [42] M. Banoula. «What is Perceptron: A Beginners Guide for Perceptron.» (), dirección: <https://www.simplilearn.com/tutorials/deep-learning-tutorial/perceptron>.
- [43] IBM. «Neural Networks.» (), dirección: <https://www.ibm.com/cloud/learn/neural-networks>.
- [44] E. A. O. S. V. Botti, «Aplicación de una red neuronal para la predicción de la reacción catalítica isomerización del n-Octano,»
- [45] A. Suresh. «What is a confusion matrix?» (), dirección: <https://medium.com/analytics-vidhya/what-is-a-confusion-matrix-d1c0f8feda5>.
- [46] EcuRed. «Matrices de confusión.» (), dirección: [https://www.ecured.cu/Matrices\\_de\\_confusi%C3%B3n](https://www.ecured.cu/Matrices_de_confusi%C3%B3n).

- [47] A. Puentes. «Así Se Compone Una Dirección.» (), dirección: [https://es.scribd.com/document/420174229/Asi-se-compone-una-direccion#download%5C&from\\_embed](https://es.scribd.com/document/420174229/Asi-se-compone-una-direccion#download%5C&from_embed).
- [48] G. Contributors. «Welcome to GeoPy's documentation!» (), dirección: <https://geopy.readthedocs.io/en/stable/#module-geopy.distance>.
- [49] Scikit-Learn. «Cross-validation: evaluating estimator performance.» (), dirección: [https://scikit-learn.org/stable/modules/cross\\_validation.html](https://scikit-learn.org/stable/modules/cross_validation.html).