



**Eficiencia Operativa en AndesBPO SAS: Predicción de Flujo de Llamadas y Gestión de Agentes.**

Daniel Felipe Muñoz Restrepo

Informe final de semestre de industria para optar al título de Ingeniero de Telecomunicaciones

Asesor interno:

Hernán Felipe García Arias, PhD

Asesor externo:

Darshan Ochoa Arroyave, Ingeniero Sistemas

Universidad de Antioquia

Faculta de Ingeniería

Departamento de Ingeniería Electrónica y Telecomunicaciones

Medellín, Antioquia

2024

---

Cita

(Muñoz Restrepo, 2024)

---

Referencia

Muñoz Restrepo D.F. (2024). *Eficiencia Operativa en AndesBPO SAS: Predicción de Flujo de Llamadas y Gestión de Agentes*. Semestre de Industria. Universidad de Antioquia, Medellín, 2024.

Estilo APA 7 (2020)

---



**Repositorio Institucional:** <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - [www.udea.edu.co](http://www.udea.edu.co)

**Rector:** Jhon Jairo Arboleda Céspedes.

**Decano/Director:** Julio César Saldarriaga Molina.

**Jefe departamento:** Eduard Emiro Rodríguez Ramírez.

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

## Resumen

En el ámbito de los contact centers y BPOs, la eficiencia operativa y la optimización de recursos son fundamentales para garantizar un servicio de calidad y maximizar la satisfacción del cliente. En este contexto, Andes BPO, ubicada en La Ceja, Antioquia, gestiona aproximadamente 700 empleados y atiende a cerca de 59 clientes en diversas líneas de negocio como servicio al cliente, cobranza y telemercadeo. Un desafío significativo para la empresa es manejar de manera efectiva el flujo de llamadas entrantes, asegurando una atención oportuna y eficiente.

Para mejorar la gestión de recursos, este proyecto se centró en la aplicación de un modelo de regresión lineal. Este modelo permite analizar y predecir el número de llamadas que se espera ingresarán al contact center Andes BPO para un cliente específico, con una ventana de predicción de 7 días. Las variables consideradas en el modelo incluyen el día del año, el mes, el día de la semana, así como indicadores de fin de semana y festivos.

El proyecto también involucró la creación de un sistema robusto de recolección y almacenamiento periódico de datos, crucial para disponer de la información necesaria para el entrenamiento efectivo del modelo. Los resultados obtenidos mostraron una precisión razonable en las predicciones, respaldada por métricas como el Error Cuadrático Medio (MSE) y el coeficiente de determinación ( $R^2$ ), los cuales demostraron ser aceptables tanto en los conjuntos de entrenamiento como en los de prueba.

Esta iniciativa no solo ha fortalecido la capacidad de Andes BPO para gestionar eficientemente sus recursos y optimizar la atención al cliente, sino que también sienta las bases para futuras mejoras en la planificación operativa y estratégica de la empresa.

## Introducción

La predicción precisa del volumen de llamadas en un contact center es fundamental para optimizar recursos y mejorar la calidad del servicio. La regresión lineal, ampliamente utilizada en la predicción de eventos, ha sido aplicada históricamente en este contexto. Este estudio revisa trabajos relevantes y describe el desarrollo de un mecanismo para predecir el número diario de llamadas, empleando técnicas de machine learning y análisis de series temporales.

En su tesis (Fuentetaja Matey, 2017), Rubén Fuentetaja Matey se centró en dos objetivos principales: la predicción del número de llamadas entrantes a un Call Center en intervalos de 30 minutos y la planificación de los descansos de los agentes para minimizar la necesidad de agentes externos. El trabajo se desarrolló en varias etapas: análisis del Centro De Atención Telefónica y de la base de datos, definición del problema, diseño del modelo de predicción, implementación y validación del modelo, diseño e implementación del modelo de asignación de personal, y propuestas de mejora. Este estudio se destaca por su enfoque sistemático y detallado en la predicción y gestión de llamadas, así como en la optimización de recursos humanos, lo cual fue altamente relevante para mi trabajo. Aunque la predicción de las llamadas se realizó mediante una técnica diferente a la usada en el presente proyecto, la productividad de los agentes se calcula mediante regresión lineal, proporcionando una base sólida para el desarrollo de mi trabajo.

El proyecto "Diseño del método de estimación de demanda para mejorar el desempeño de un Contact Center en Barranquilla" (Arteta et al., 2020) se presenta un método de predicción de demanda utilizando técnicas como ARIMA y VAR. Este enfoque está diseñado para optimizar la programación operativa y reducir la variabilidad en la demanda, validando las predicciones mediante simulaciones. Aunque el modelo ARIMA demostró ser más efectivo con un RMSE de 5,775.991 llamadas por semana, en este proyecto se adoptó el uso de una técnica diferente. A pesar de que la regresión lineal es menos robusta que el ARIMA utilizado por los autores mencionados dado que este último permite modelar de mejor manera un conjunto de datos que muestran una fuerte autocorrelación y patrones complejos, en el presente proyecto se alcanzaron métricas de desempeño cercanas a las obtenidas en dicho estudio, lo cual valida la efectividad del modelo en el enfoque de predicción de llamadas para apoyar la planificación de recursos.

Andes BPO carece de un sistema eficiente respaldado por datos para la programación de turnos y la estimación del volumen de llamadas, crucial para la asignación de personal en sus líneas de atención. En respuesta a esta necesidad, se realizó la implementación de un modelo de regresión lineal para predecir con precisión el tráfico de llamadas. Se implementó un sistema automático de recolección y almacenamiento de datos históricos detallados, integrando variables adicionales relevantes para mejorar la precisión del modelo predictivo. El modelo se entrena con datos nuevos en cada ejecución y envía automáticamente, vía correo electrónico, las predicciones para los siguientes 7 días. Destaca que el sistema opera de manera autónoma, sin intervención humana durante el proceso.

Sin embargo, es importante destacar que el modelo es completamente dependiente de datos históricos por lo tanto su precisión puede verse afectada por cambios abruptos en patrones de llamadas que no están reflejados en los datos

históricos recopilados además, presenta sensibilidad a variables externas que no estén incluidas en el modelo, como eventos especiales o cambios en las campañas de marketing que podrían influir en la precisión de las predicciones por lo tanto la capacidad del modelo para adaptarse a cambios a largo plazo en el comportamiento de los usuarios o en las condiciones del mercado podría ser limitada, requiriendo actualizaciones periódicas del modelo.

El proyecto se alinea con los esfuerzos para mejorar la eficiencia operativa y la planificación de recursos en la empresa AndesBPO. Los datos abarcan un período de año y medio y continúan expandiéndose. La metodología empleada incluye la rigurosa limpieza y transformación de datos, así como la creación de nuevas características, y la construcción y evaluación de un modelo de regresión lineal. Además, el modelo establece una base sólida para que la compañía pueda extenderlo a todos sus clientes en el futuro. Además, el modelo tiene el potencial de evolucionar más allá de la predicción de llamadas para mejorar otros aspectos del negocio.

## 1. Objetivos

### 1.1. Objetivo General:

Se desarrollo un sistema para la predicción del flujo entrante de llamadas en el contact center de Andes BPO, específicamente para el cliente Gobernación de Antioquia, con el fin de optimizar la asignación de recursos humanos y mejorar la eficiencia operativa.

### 1.2. Objetivos Específicos:

- Se implemento un sistema automatizado de captura y almacenamiento de datos que recopilo información relevante sobre el tráfico de llamadas durante un año y medio (enero de 2023 – julio de 2024), incluyendo cantidad diaria, por hora y por mes, duración y motivo de las llamadas, para el cliente Savia Salud.
- Se implemento un modelo de regresión lineal utilizando los datos recopilados para predecir el flujo de llamadas esperado en los próximos 7 días, con esta información el líder de campaña está en la capacidad de determinar la cantidad óptima de agentes requeridos en cada periodo.
- Se implemento un sistema automatizado de envío de correos electrónicos, utilizando el servidor SMTP de Office 365, para notificar a los responsables de la operación del contact center sobre las recomendaciones de personal para cada periodo, se entrega en cada envío una tabla con las predicciones y un gráfico que le permite observar la precisión que está teniendo el modelo, para que pueda generar alertas en caso de alguna anomalía con el modelo.
- Se evaluó la eficacia del sistema implementado mediante las métricas MSE y  $R^2$ , además se entrega al líder de operación en cada ejecución el grafico donde se compara la eficacia del modelo en sus predicciones en los últimos 7 días vs la cantidad real de llamadas para mantener una permanente vigilancia del comportamiento del modelo.

## 2. Marco Teórico

Un Contact Center, también conocido como centro de interacción con el cliente, es un departamento dentro de una empresa que se encarga de gestionar y atender todas las comunicaciones con los clientes (Licari, s. f.). Estas comunicaciones pueden realizarse a través de diferentes canales como llamadas, videollamadas, chats, correo electrónico, SMS, comentarios en redes sociales, etc (Licari, s. f.). Los agentes de Contact Center utilizan un programa de gestión de relación con el cliente (CRM) para guardar y consolidar todas las conversaciones de diferentes canales en un solo lugar(Licari, s. f.).

El BPO, o Subcontratación de Procesos de Negocios (Business Process Outsourcing), consiste en la subcontratación de funciones o procesos de negocio a proveedores de servicios especializados (TEAM, s. f.). Se trata de asignar a entidades externas actividades específicas que tienen un impacto directo en el negocio con el objetivo de mejorar su calidad, reducir costos y aumentar la productividad (TEAM, s. f.).

Los Contact Centers ofrecen un soporte rápido y eficiente, mientras que los proveedores de BPO pueden gestionar eficazmente las operaciones comerciales fundamentales. Los Contact Centers BPO son proveedores de servicios especializados que gestionan la comunicación con los clientes para otras empresas. Son fundamentales para gestionar las interacciones con los clientes, mejorar la eficiencia y garantizar una comunicación fluida (¿Qué Es un Contact Center BPO y Cómo Funciona? | Five9, s. f.).

La optimización de recursos, ya sean humanos o tecnológicos, es fundamental para mejorar la eficacia operativa y la satisfacción del cliente en cualquier empresa, especialmente en un contexto de BPO. Una gestión eficiente del personal puede resultar en la reducción de costos al minimizar los tiempos de espera y utilizar la cantidad óptima de agentes, evitando así el exceso de personal para atender el flujo de llamadas.

Por otro lado, una falta de personal adecuado y tiempos de espera prolongados pueden llevar a una percepción negativa del servicio por parte de los clientes finales. Esto se traduce en mayores tiempos de espera y en la prestación de un servicio deficiente.

Según la página (Introducción a la regresión lineal, s. f.) define la regresión lineal como una técnica estadística fundamental que se utiliza para modelar la relación entre una variable de respuesta continua y una o varias variables predictoras. Este modelo es esencial para comprender y predecir el comportamiento de sistemas complejos o analizar datos en diversos campos como experimentales, financieros y biológicos.

En la regresión lineal, se busca describir la relación entre una variable dependiente (la respuesta, Y) y una o varias variables independientes (predictores, Xi) mediante una función lineal. La ecuación general de un modelo de regresión lineal es:

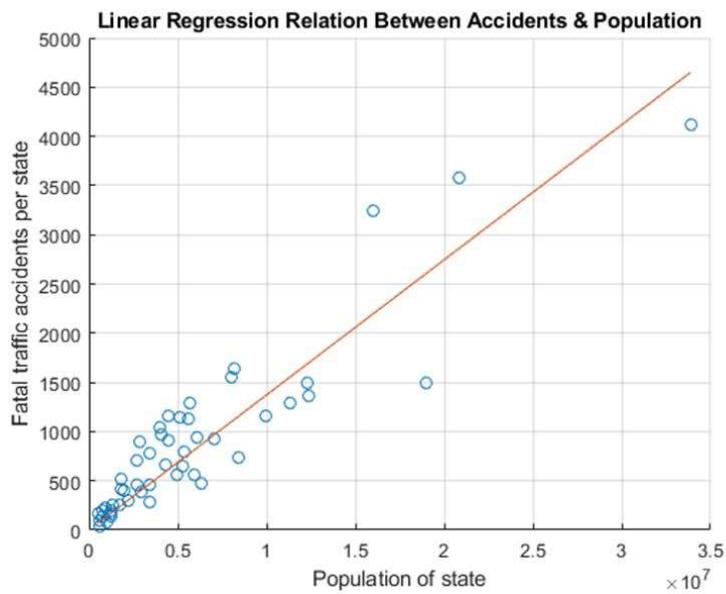
$$Y = \beta_0 + \sum \beta_k \chi_k + \epsilon_i$$

Donde  $\beta$  representa los coeficientes que se deben estimar y  $\epsilon$  representa el término de error.

Existen varios tipos de regresión lineal:

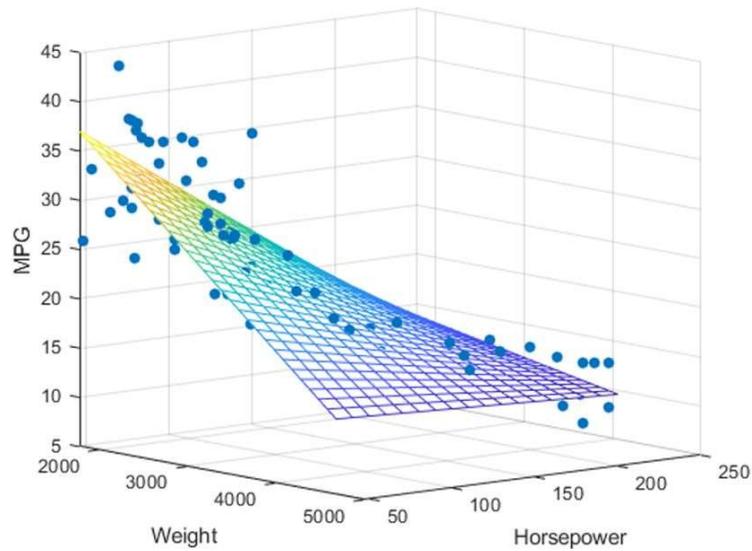
- Regresión lineal simple: Utiliza un único predictor. La ecuación correspondiente es  $Y = \beta_0 + \beta_1 \chi + \epsilon$

**Figura 1.** *Ejemplo de regresión lineal simple (Introducción a la regresión lineal, s. f.)*



- Regresión lineal múltiple: Utiliza múltiples predictores. La ecuación es  $Y = \beta_0 + \beta_1\chi_1 + \beta_2\chi_2 + \epsilon$

**Figura 2.** Ejemplo de regresión lineal múltiple (Introducción a la regresión lineal, s. f.)



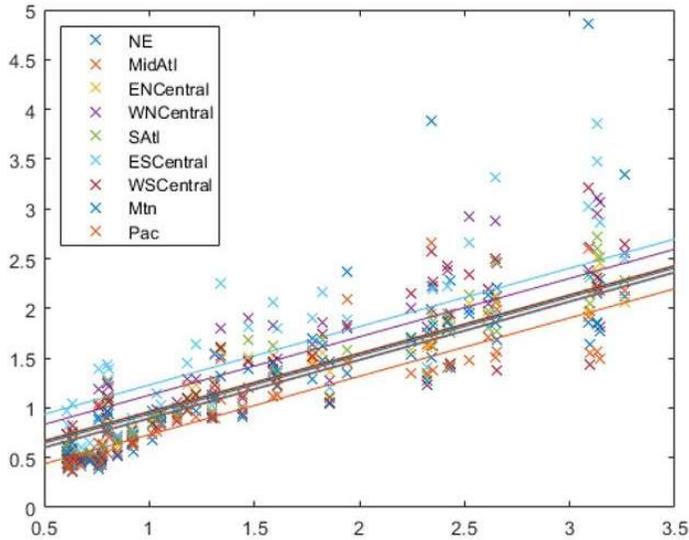
- Regresión lineal multivariante: Modela múltiples variables de respuesta derivadas de los mismos datos. Por ejemplo, si se tienen dos ecuaciones para predecir dos respuestas  $Y_1$  y  $Y_2$  a partir de los mismos predictores  $X_1, X_2$ , etc.

$$Y_1 = \beta_{01} + \beta_{11}\chi_1 + \epsilon_1$$

$$Y_2 = \beta_{02} + \beta_{12}\chi_1 + \epsilon_2$$

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} \beta_{01} & \beta_{11} \\ \beta_{02} & \beta_{12} \end{pmatrix} \begin{pmatrix} 1 \\ \chi_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix}$$

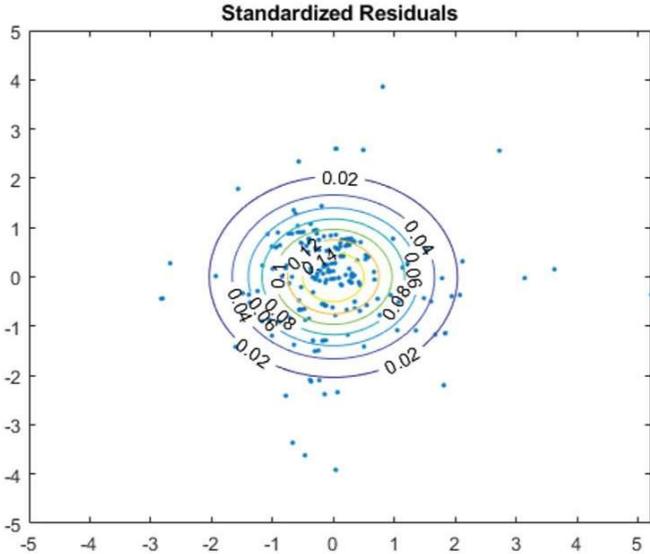
**Figura 3.** Ejemplo de regresión lineal multivariante (Introducción a la regresión lineal, s. f.)



- Regresión lineal múltiple multivariante: Modela varios predictores para múltiples variables de respuesta.

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1q} \\ 1 & X_{21} & X_{22} & \dots & X_{2q} \\ 1 & X_{31} & X_{32} & \dots & X_{3q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nq} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

**Figura 4.** Ejemplo de regresión lineal múltiple multivariante (Introducción a la regresión lineal, s. f.)



La regresión lineal es ampliamente utilizada en diversas aplicaciones, como la predicción o pronóstico de valores de respuesta donde solo se conocen los predictores, y para determinar la fuerza y la naturaleza de la relación entre una variable y sus predictores.

La extracción, transformación y carga (ETL) es un proceso fundamental en la integración de datos que implica combinar datos de diversas fuentes en un repositorio central para su análisis. A lo largo del tiempo, ETL ha evolucionado para abordar las necesidades cambiantes de las organizaciones en términos de volumen, variedad y velocidad de los datos. La importancia de ETL radica en su capacidad para mejorar la calidad de los datos, proporcionar una vista consolidada de los datos para análisis profundos, ofrecer un contexto histórico a los datos y automatizar tareas repetitivas de procesamiento de datos(¿Qué es ETL?, s. f.).

ETL se ha adaptado a las nuevas tecnologías y demandas, como la incorporación de datos no estructurados, el procesamiento en tiempo real y la migración a la nube. La transformación de datos desempeña un papel clave en ETL al limpiar, organizar y optimizar los datos para su análisis. En la actualidad, existen herramientas y servicios especializados, como AWS Glue, que facilitan la implementación y gestión de ETL en entornos empresariales. Con la evolución continua de ETL y la adopción de tecnologías emergentes, como la virtualización de datos, se espera que este proceso siga desempeñando un papel fundamental en la gestión y análisis de datos en las organizaciones modernas(¿Qué es ETL?, s. f.).

Según la web de Amazon (¿Qué es Python?, s. f.), Python es un lenguaje de programación interpretado ampliamente utilizado en diversas áreas como el desarrollo web, la ciencia de datos y el machine learning. Fue creado por Guido Van Rossum en 1989 y se caracteriza por su sintaxis sencilla y legible, lo que facilita su aprendizaje y comprensión. Una de las ventajas de Python es su gran biblioteca estándar, que contiene códigos reutilizables para casi cualquier tarea, lo que permite a los desarrolladores escribir programas con menos líneas de código en comparación con otros lenguajes. Además, Python es compatible con otros lenguajes de programación como Java, C y C++, y se puede ejecutar en múltiples plataformas como Windows, macOS, Linux y Unix.

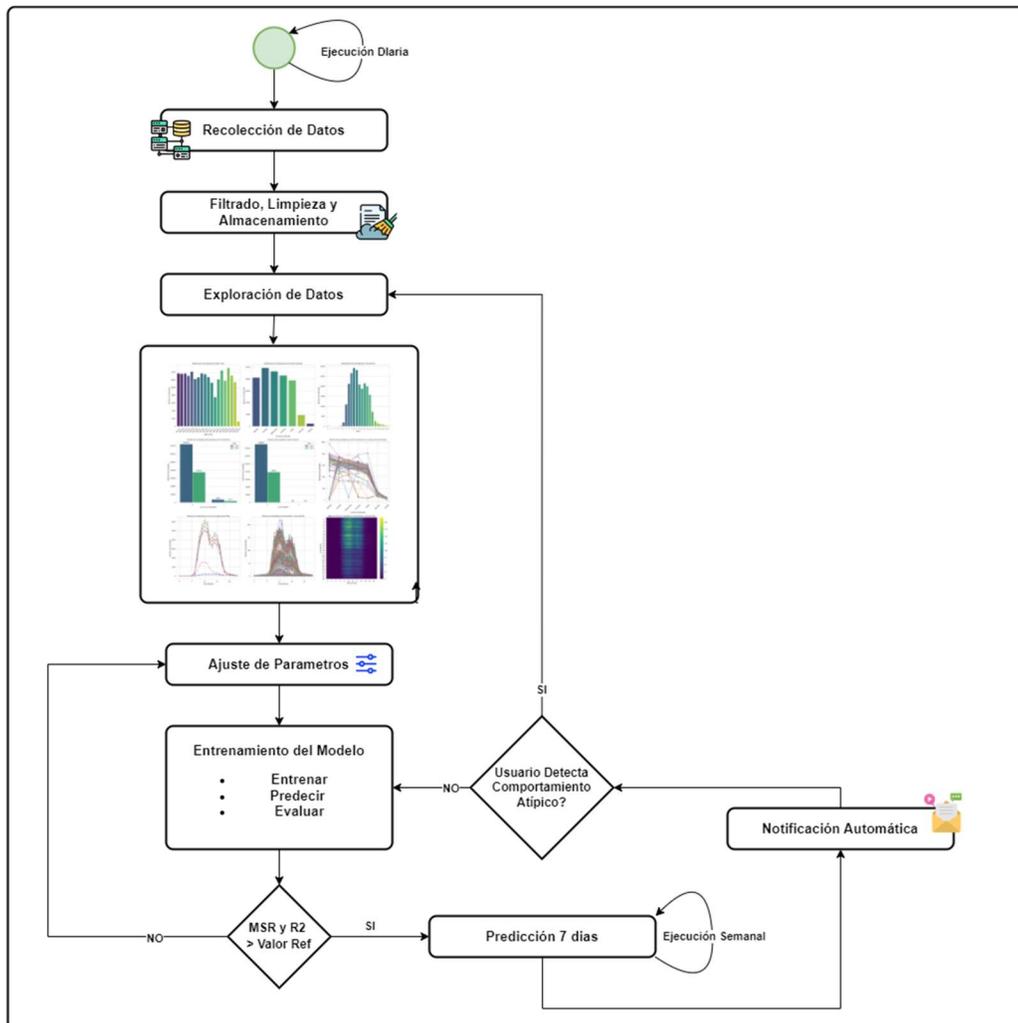
En el desarrollo web, Python se utiliza en el lado del servidor para crear funciones complejas de backend. Ofrece numerosas bibliotecas y frameworks que facilitan la creación de aplicaciones web rápidas y eficientes. Por ejemplo, Django y Flask son dos de los frameworks más populares para el desarrollo web con Python.

En el ámbito de la automatización, Python es ampliamente utilizado para escribir scripts que automatizan tareas diarias como la manipulación de archivos, el envío de correos electrónicos y el análisis de datos. Su sintaxis sencilla y su capacidad para integrarse con otros sistemas hacen que sea una opción popular para la automatización de tareas.

En la ciencia de datos y el machine learning, Python es una herramienta fundamental. Se utiliza para limpiar y procesar datos, crear modelos de machine learning y realizar análisis estadísticos. Su amplia gama de bibliotecas especializadas, como Pandas, NumPy y Scikit-learn, lo convierten en una opción popular entre los científicos de datos y los desarrolladores de machine learning.

### 3. Metodología

Figura 5. Diagrama de Flujo del Proceso (Elaboración Propia)



#### 3.1. Recolección de Datos

Mediante el uso de Python, se construyó un script que diariamente genera una consulta vía API REST al aplicativo encargado de la gestión telefónica de la empresa (Wolkvox). La respuesta en formato JSON es manipulada, estructurada y limpiada para luego almacenarla en un archivo CSV con la siguiente estructura en su nombre: "general-cdr\_1\_2024-07-05", donde "general-cdr\_1" hace referencia al tipo de reporte (CDR - registro detallado de llamadas) y la fecha de la consulta. Así, se genera un archivo por cada día del año para todos los clientes de la empresa.

Para asegurar la integridad de los datos, una vez el archivo se encuentra alojado localmente en el servidor, este se sincroniza con OneDrive para mantener una copia de seguridad y evitar una posible pérdida de datos.

Esta rutina se desarrolló con suficiente antelación para que al momento de realizar el primer entrenamiento del modelo ya se contara con un año y medio de datos históricos.

### 3.1.1. Filtrado y Almacenamiento de Datos

Se tomaron todos los archivos generados (549 al momento de redactar el informe, aumentando a diario) y se filtró cada archivo para tomar solo los datos del cliente objetivo, en este caso, Savia Salud. Todos los datos filtrados se almacenaron en un archivo CSV llamado “1.calls\_details.csv”, que contenía información de 230,958 llamadas al momento de la redacción. El archivo incluye la siguiente información para cada llamada:

- agent\_name – nombre del agente que atendió la llamada.
- date – fecha completa en la que ingreso la llamada.
- telephone – número telefónico del cliente que se comunica.
- time\_min – duración en minutos de la llamada.
- cod\_act – código numérico que representa el motivo de la llamada.
- description\_cod\_act – motive de la llamada.
- ype\_interaction – tipo de llamada (entrante o saliente).
- customer\_id – identificación del cliente.
- hang\_up – Quien finalizó la llamada (cliente o agente)
- campaign\_id – identificación numérica de la campaña de salida.
- conn\_id – identificación única de la llamada.
- skill\_id – identificación numérica de la cola por la cual ingreso la llamada.
- skill\_name – nombre de la cola por la cual ingreso la llamada.
- agent\_id – identificación numérica del agente que atendió la llamada.
- Module – cliente corporativo al que pertenece la llamada

Se realizó una depuración del archivo para tomar solo las características relevantes para el entrenamiento del modelo, utilizando únicamente date, time\_min, y cod\_act. Adicionalmente, se crearon nuevas características derivadas de la fecha:

- year
- month
- day
- hour
- day\_of\_week
- month\_of\_year
- is\_weekend
- is\_holiday

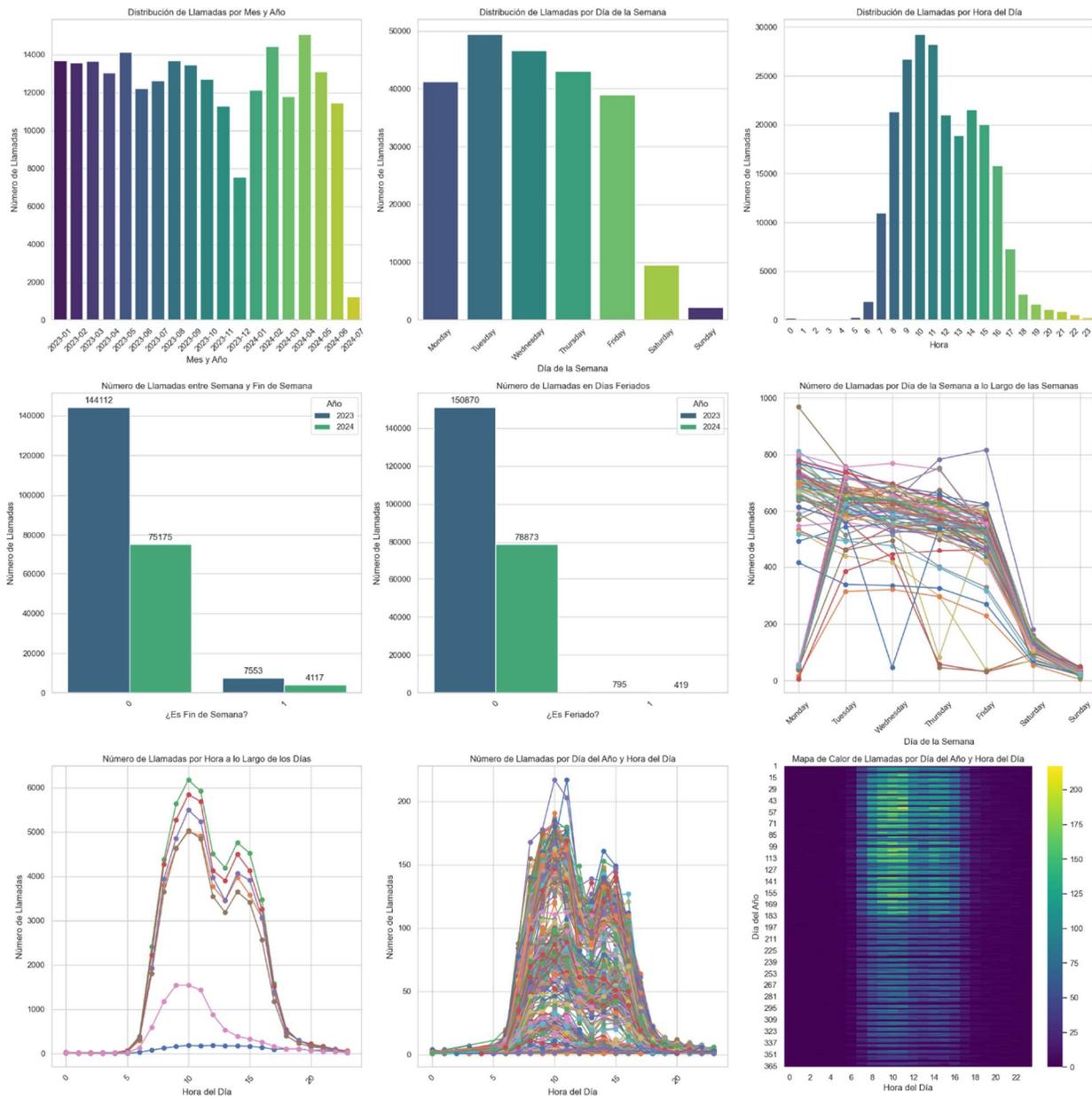
Estas características se crearon para mejorar la precisión del modelo y manejar correctamente las predicciones en fines de semana y días festivos, los cuales representan un comportamiento atípico.

### 3.1.2. Exploración de Datos

Se procedió con la exploración de los datos obtenidos usando las librerías matplotlib y seaborn, con el fin de analizar los datos y verificar que efectivamente los datos tenían comportamientos relativamente periódicos o secuenciales y determinar si realmente con el modelo de regresión lineal se lograría implementar el modelo o si por el contrario se iba a requerir un modelo más robusto.

A continuación, se presentan las gráficas del comportamiento de los datos y el análisis que se hizo de cada una de ellas:

**Figura 6. Comportamiento temporal del flujo de llamadas (Introducción a la regresión lineal, s. f.)**



Voy a exponer las características más relevantes que se observan en las diferentes gráficas construidas, si se analizan las gráficas de las casillas 1,3 – 3,1 y 3,2 (Distribución de Llamadas por Hora del Día, Número de Llamadas por Hora a lo Largo de los Días de la Semana y Número de Llamadas por Día del Año y Hora del Día) vemos una clara periodicidad en el flujo entrante de llamadas el cual siempre a las mismas horas del día presenta picos y valles similares ya sea si lo analizamos por días, días de la semana o semanas del año siempre se repite el mismo patrón de flujo.

Ahora al analizar el comportamiento presentado en la gráfica de la cuadrícula 2,3 (Número de Llamadas por Día de la Semana a lo Largo de las Semanas) encontramos un comportamiento nuevamente periódico y muy consistente en el cual a inicio de semana siempre está el pico más fuerte de llamadas y este flujo va disminuyendo al pasar los días hasta tener su menor valor los días domingos, además se observa con claridad los días lunes festivos los cuales concentran el flujo de llamadas en torno a la misma cantidad (no más de 50) y luego en su mayoría continúan la tendencia habitual de los días martes.

Posterior a estos análisis realizados a la distribución de los datos se determinó que la regresión lineal efectivamente podía ser un modelo que pudiera ajustarse bien a este conjunto de datos dado ese comportamiento tan periódico dependiente del tiempo y de la particularidad de si un día es festivo o no.

Este nuevo conjunto de datos con las características adicionales creadas se almacena en el archivo “2.calls\_details\_with\_date\_info.csv”

### **3.1.3. Entrenamiento del Modelo**

Para este paso se realizó el entrenamiento del modelo, para lograrlo se requirió del uso de la librería sklearn de Python, específicamente los módulos `train_test_split`, `LinearRegression` y `mean_squared_error`, `r2_score`. El proceso consistió en leer el dataset contenido en el archivo `2.calls_details_with_date_info.csv` y generar una agrupación por día ya que el conjunto de datos original tenía el detalle de cada llamada individual pero no el consolidado del día; luego cada fila generada quedó representada la cantidad de llamadas totales que ingresaron ese día y se le adjuntaron las columnas adicionales relevantes que ya se tenían `dia_del_ano`, `mes`, `dia_de_la_semana`, `es_fin_de_semana` y `is_holiday`.

Con el conjunto ya preparado con los datos resumidos y las columnas adicionales listas se procedió con el entrenamiento del modelo, los pasos fueron los que habitualmente se realizan en este tipo de implementaciones.

- Definir las variables independientes [`'dia_del_ano'`, `'mes'`, `'dia_de_la_semana'`, `'es_fin_de_semana'`, `'is_holiday'`] y dependiente [`'numero_de_llamadas'`]
- Dividir los datos en conjuntos de entrenamiento y prueba con el uso del módulo `train_test_split`, se determinó que el 80% de los datos se usarían para entrenamiento y el 20% para prueba.
- Crear y entrenar el modelo de regresión lineal con el conjunto de datos de entrenamiento.
- Hacer predicciones.
- Evaluar el modelo.

Este paso requirió de varios intentos de ejecución y afinación de parámetros, por ejemplo, acá fue donde se encontró la necesidad de que el conjunto de datos contara con información acerca de si el día pertenecía o no a fin de semana o si era un día festivo o no, ya que sin estas características el modelo no estaba ajustándose de manera adecuada al comportamiento de los datos y estas falencias se hacía muy evidentes al momento de evaluar el modelo.

El modelo se evaluó con las métricas de desempeño MSR y R2, la idea es que el MSR sea lo mas bajo posible y R2 lo mas cercano posible a 1, de esta manera las variables se van ajustando hasta lograr estos valores lo mas cercano a lo ideal como sea posible.

#### **3.1.4. Predicción y Planificación**

Se socializó con el personal encargado de usar el producto y se determinó que se necesitaban datos de los siguientes 7 días para una planificación efectiva de los turnos. Por tanto, se definió que el modelo entregara la predicción de los próximos 7 días posteriores a su ejecución.

#### **3.1.5. Notificación Automática**

Usando Python y la librería smtplib, se construyó un modelo que, utilizando el servidor de correo de la empresa (Office 365), envía al correo electrónico del usuario responsable del proceso la predicción de las llamadas por día y adjunta una gráfica del desempeño del modelo en los últimos 7 días transcurridos, en caso de que el usuario detecte alguna inconsistencia(predicción notablemente desviada) en la gráfica se procedería con la notificación y posterior revalidación de parámetros y reentrenamiento del modelo ya que esto podría indicar un cambio abrupto en el comportamiento normal del flujo de llamadas.

#### 4. Resultados y análisis

Existen dos métricas de evaluación del desempeño del modelo de regresión lineal muy utilizadas: el MSR (Mean Squared Error) y el R2 (Coeficiente de Determinación).

**MSR (Mean Squared Error):** Es una métrica que cuantifica la diferencia promedio al cuadrado entre los valores predichos por el modelo y los valores reales. Se calcula como la media de los cuadrados de los errores (diferencias entre los valores predichos y reales). Un MSE más bajo indica un mejor ajuste del modelo a los datos.

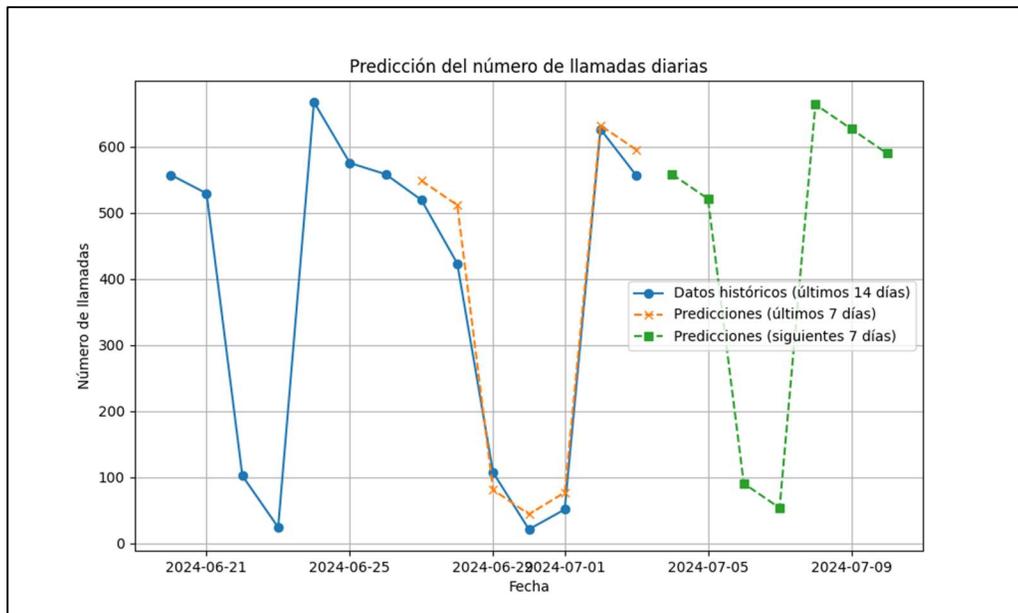
**R2 (Coeficiente de Determinación):** Es una métrica que indica la proporción de la varianza en la variable dependiente que es explicada por las variables independientes del modelo. Su valor varía entre 0 y 1, donde un valor más cercano a 1 indica un mejor ajuste del modelo.

- El modelo de regresión lineal implementado mostró un rendimiento satisfactorio luego de haber afinado de mejor manera las características adicionales para tener en cuenta para el entrenamiento como lo fueron los días festivo y determinar si un día era fin de semana o no, a continuación, se relacionan el resultado de los parámetros de evaluación del modelo.
  - MSE en entrenamiento: 4623.7376245450405
  - R2 en entrenamiento: 0.9312095494911848
  - MSE en prueba: 5883.098195345913
  - R2 en prueba: 0.9176402679748454
- El MSE obtenido con el conjunto de entrenamiento muestra que el modelo se ajusta bien a los datos con los que fue entrenado y el MSE de prueba, aunque el valor es ligeramente mayor que el MSE de entrenamiento, aun así, muestra que el modelo tiene un buen rendimiento en datos no vistos.
- El R2 para el conjunto de entrenamiento fue de 93.12% que representa la variabilidad en el número de llamadas diarias es explicada por el modelo en el conjunto de entrenamiento. Esto indica un ajuste muy fuerte del modelo a los datos de entrenamiento mientras que el R2 en prueba fue de 0.9176, valor que sugiere que el 91.76% de la variabilidad en el número de llamadas diarias es explicada por el modelo en el conjunto de prueba. A pesar de ser un poco menor que el R2 de entrenamiento, aun así, muestra un ajuste muy fuerte, lo que sugiere que el modelo generaliza bien a datos no vistos.
- Similarmente el modelo entrega el detalle del comportamiento en los últimos 7 días, ya que una variabilidad muy fuerte en el rendimiento del modelo generaría una alerta de que la línea está teniendo un comportamiento anómalo a lo que generalmente ocurre esto puede ser un exceso de llamadas por algún evento particular o la ausencia de esta por posible inestabilidad en la numeración que recibe las llamadas u otro tipo de problema.
  - MSE en los últimos 7 días: 1721.6987676898932
  - R2 en los últimos 7 días: 0.9703232865955459

Similar a lo visto con el resultado de las métricas obtenido en el conjunto de entramiento y prueba vemos un ajuste bastante certero en la predicción del flujo de llamadas.

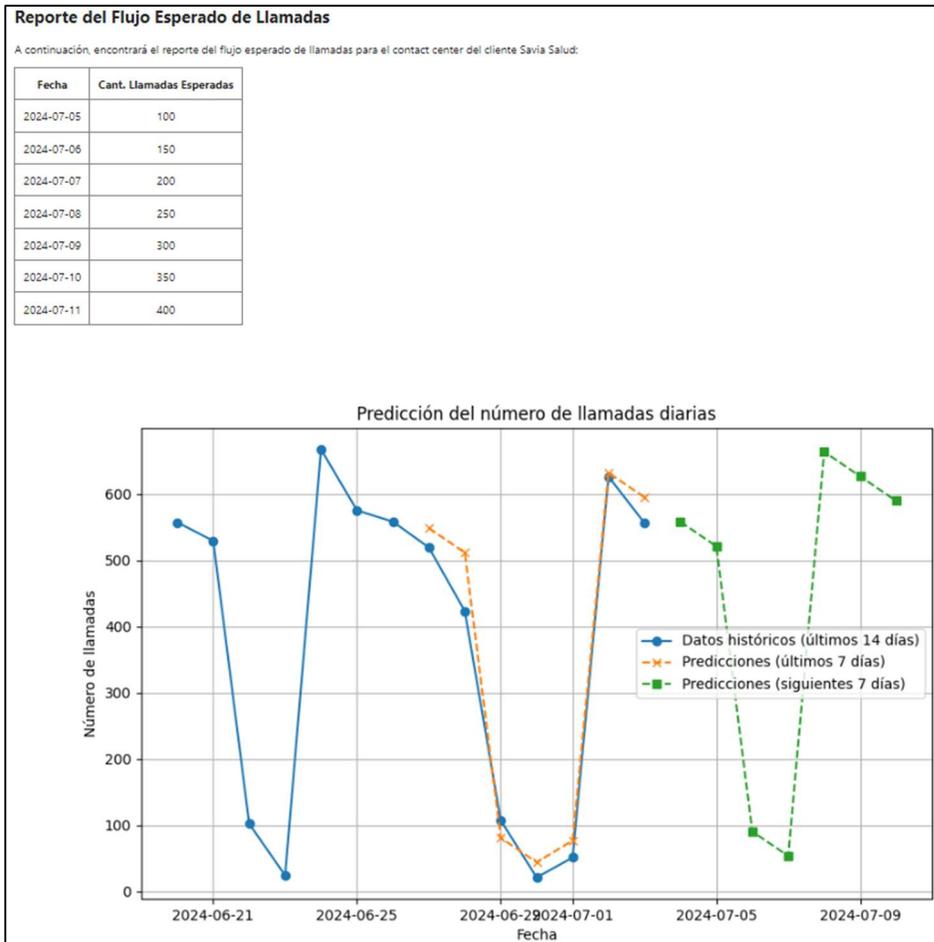
- Resultado final de las predicciones pasadas y futuras del modelo, en la gráfica la línea azul representa el flujo real de llamadas, la línea naranja la predicción de días pasado para contrastar con el flujo real y finalmente la línea verde representa la predicción para los siguiente 7 días.

**Figura 7.** Estimación del flujo de llamadas (Elaboración Propia).



- El producto final genera el reporte que se muestra en la siguiente imagen, este llega de manera automática y periódica al usuario final sin ninguna intervención ya que el sistema funciona de manera completamente autónoma.

**Figura 8.** Correo electrónico entregado al usuario del producto (Elaboración Propia)



## 5. Conclusiones

- Mediante el proyecto se desarrolló un sistema eficaz para predecir el flujo de llamadas. La alta precisión del modelo, evidenciada por los bajos valores de MSE y los altos valores de  $R^2$ , demuestra que el sistema cumple con su objetivo general. Esto permitirá a los responsables de la operación optimizar la asignación de recursos humanos, mejorando así la eficiencia operativa, ya que previamente no se contaba con una base sólida para tomar decisiones en la asignación de agentes para la atención de llamadas.
- El sistema automatizado implementado captura y almacena una cantidad significativa de datos, esenciales tanto para el entrenamiento del modelo predictivo como para su posterior precisión. Además, está diseñado para reunir diversos tipos de datos útiles para replicar y expandir el modelo a otros enfoques del negocio. Se establecieron mecanismos de respaldo que aseguran la integridad y conservación de los datos ante cualquier falla del servidor de almacenamiento, garantizando así la fiabilidad del sistema.
- Para asegurar la completa autonomía del sistema de predicción, se añadió un módulo de envío automático de predicciones. Este utiliza el servidor SMTP de Office 365 de la empresa para enviar correos electrónicos anticipando el flujo de llamadas a los responsables de la operación. Los correos incluyen una tabla con las predicciones y un gráfico de precisión del modelo, facilitando la previsión y gestión eficiente de los turnos de los agentes. Este proceso automatizado evita reprocesos y demoras que podrían ocurrir si el envío del reporte dependiera de una persona.
- Al evaluar la eficacia del sistema utilizando las métricas MSE y  $R^2$ , se obtuvieron resultados muy satisfactorios, con un MSE bajo y un  $R^2$  alto tanto en los conjuntos de entrenamiento como de prueba. Estas métricas fueron cruciales para ajustar los parámetros del modelo, mejorando su precisión. Además, se proporciona regularmente un gráfico comparativo que permite al líder de operación monitorear la precisión del modelo en tiempo real. Esto asegura una vigilancia constante, detectando anomalías oportunamente y permitiendo reajustes necesarios, dado que el modelo es susceptible a cambios repentinos en el flujo de llamadas.
- La capacidad del modelo para generalizar nuevos datos se ha validado, mostrando consistencia tanto en los conjuntos de entrenamiento como en los de prueba. La estructura automatizada y la inclusión de mecanismos de respaldo aseguran la integridad y disponibilidad de los datos. Este sistema no solo optimiza la asignación de recursos humanos, sino que también proporciona una base sólida para futuras mejoras y expansiones en la gestión operativa de la compañía.
- Con la implementación de este sistema predictivo, se ha proporcionado una herramienta robusta para la gestión proactiva de recursos en la compañía. Esto ha permitido aprovechar grandes cantidades de datos para

tomar decisiones fundamentadas, impactando positivamente los costos operativos y mejorando la experiencia de atención a los usuarios. Además, se fortalece la relación con los clientes corporativos, ya que el sistema contribuye a una asignación de recursos más eficiente y una atención más satisfactoria al cliente final.

- Finalmente, con la implementación de este modelo y la estructuración de los datos almacenados, la compañía ha identificado un campo sin explorar donde deben enfocar sus esfuerzos. Muchos de sus procesos y procedimientos carecen de fundamentos sólidos para la toma de decisiones. Este sistema ha demostrado que, mediante la utilización de datos estructurados y modelos predictivos, es posible optimizar múltiples áreas de la empresa, mejorando así su eficiencia operativa y la toma de decisiones estratégicas.

## Referencias Bibliográficas

Arteta, E., García, L., Guerrero, K., & Pertuz, M. (2020). *Diseño del método de estimación de demanda para mejorar el desempeño de un Contact Center en Barranquilla*. <https://manglar.uninorte.edu.co/handle/10584/9365>

Fuentetaja Matey, R. (2017, junio). *Análisis, desarrollo y valoración de propuestas de mejora para la asignación de personal en un Call Center* [Info:eu-repo/semantics/bachelorThesis]. E.T.S.I. Industriales (UPM). <https://oa.upm.es/47320/>

*Introducción a la regresión lineal*. (s. f.). Recuperado 28 de marzo de 2024, de <https://la.mathworks.com/discovery/linear-regression.html>

Licari, S. (s. f.). *¿Qué es un contact center y qué ofrece?* Recuperado 28 de marzo de 2024, de <https://blog.hubspot.es/service/que-es-contact-center>

*¿Qué es ETL? - Explicación de extracción, transformación y carga (ETL) - AWS*. (s. f.). Amazon Web Services, Inc. Recuperado 28 de marzo de 2024, de <https://aws.amazon.com/es/what-is/etl/>

*¿Qué es Python? - Explicación del lenguaje Python - AWS*. (s. f.). Amazon Web Services, Inc. Recuperado 28 de marzo de 2024, de <https://aws.amazon.com/es/what-is/python/>

*¿Qué Es un Contact Center BPO y Cómo Funciona? | Five9*. (s. f.). Recuperado 28 de marzo de 2024, de <https://www.five9.com/blog/que-es-un-contact-center-bpo-y-como-funciona>

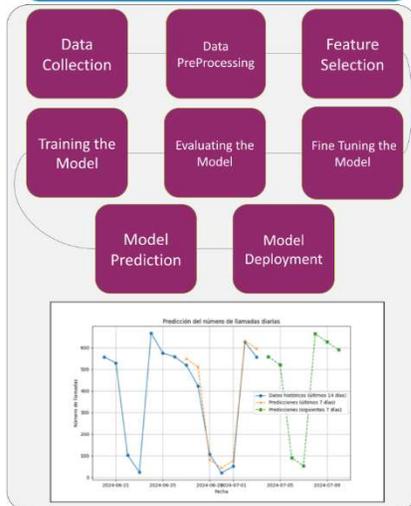
TEAM, A. (s. f.). *BPO: Todo lo que debes saber sobre Business Process Outsourcing*. Recuperado 28 de marzo de 2024, de <https://www.ambit-bst.com/blog/bpo-todo-lo-que-debes-saber-sobre-business-process-outsourcing>

### Introducción

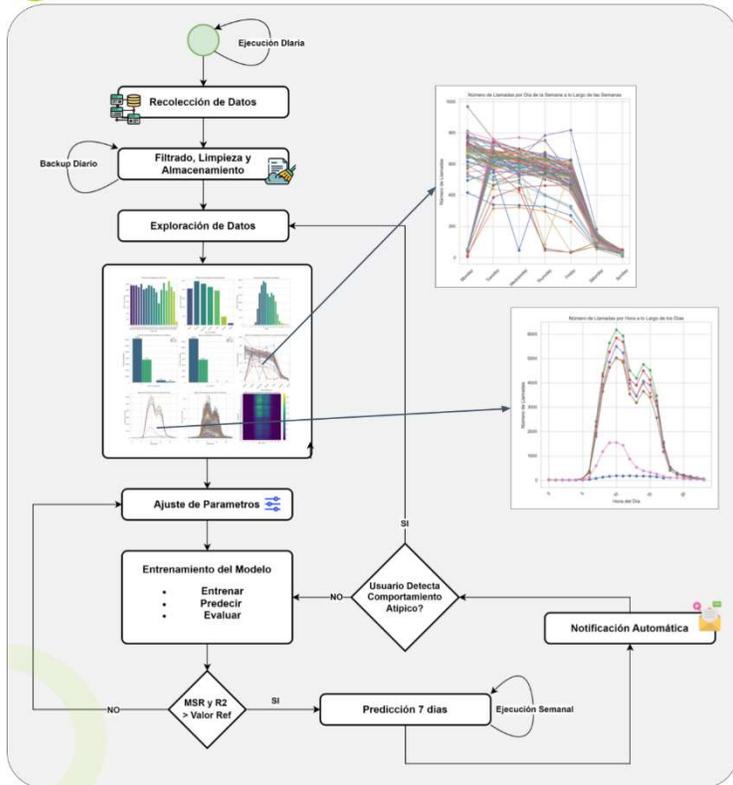
El análisis predictivo del flujo de llamadas es esencial para optimizar la gestión de recursos en centros de atención telefónica. Este proyecto implementa un mecanismo para predecir el número diario de llamadas utilizando regresión lineal, para ello se desarrolló un sistema automático que recolecta y almacena datos históricos detallados, integrando variables adicionales para mejorar la precisión en el entrenamiento del modelo.

### Objetivos

- ✓ Desarrollar un sistema para predecir el flujo entrante de llamadas en el contact center de Andes BPO para un cliente específico.
- ✓ Implementar un sistema automatizado de captura y almacenamiento de datos.
- ✓ Entrenar un modelo de regresión lineal para predecir el flujo de llamadas de los próximos 7 días, permitiendo al líder de campaña determinar la cantidad óptima de agentes.
- ✓ Implementar un sistema automatizado de envío de correos electrónicos que entregará el informe detallado del modelo.



### Metodología



### Resultados

MSE (Mean Squared Error) y  $R^2$  (Coeficiente de Determinación)

MSE en entrenamiento: 4623.74  
 $R^2$  en entrenamiento: 0.93

MSE en prueba: 5883.10  
 $R^2$  en prueba: 0.92

MSE en los últimos 7 días: 1721.70  
 $R^2$  en los últimos 7 días: 0.97

### Conclusiones

- ✓ Tener la predicción del flujo de llamadas, permitirá la optimización al asignar recursos y mejorando la eficiencia operativa.
- ✓ El modelo es altamente sensible a la calidad y cantidad de datos disponibles.
- ✓ La constante supervisión del modelo permite una calibración constante del mismo.
- ✓ Las métricas MSE y  $R^2$  resultan ser parámetros muy adecuados para ajustar el modelo correctamente.
- ✓ Viabilidad en la aplicación del modelo en otras áreas pertinentes de mejora.

#### DATOS DE CONTACTO DEL AUTOR:

3192984032

+57 3192984032

danielf.munoz@udea.edu.co

<http://www.linkedin.com/in/daniel-felipe-munoz-restrepo-05262b184>