

**IMPACTO DE LAS TIC EN EMPRESAS COLOMBIANAS DEL SECTOR  
SERVICIOS: 2008 - 2019**

**LAURA DANIELA CASTILLO HERNÁNDEZ**

**WORKING PAPER**

**TUTOR  
GUSTAVO ESCOBAR**



**UNIVERSIDAD COLEGIO MAYOR DE CUNDINAMARCA  
FACULTAD DE ADMINISTRACIÓN Y ECONOMÍA  
PROGRAMA DE ECONOMÍA  
BOGOTÁ D.C.  
2023**

**Tabla de contenido**

Resumen .....	5
Introducción .....	7
Revisión de la literatura .....	10
Desarrollo .....	16
Referencias .....	32

**Lista de figuras**

Figura 1.....	18
Figura 2.....	20
Figura 3.....	20
Figura 4.....	21
Figura 5.....	23
Figura 6.....	24
Figura 7.....	25
Figura 8.....	25
Figura 9.....	26
Figura 10.....	27
Figura 11.....	28
Figura 12.....	28

**Lista de anexos**

Anexo A.....	34
Anexo B .....	35
Anexo C .....	36
Anexo D .....	37
Anexo E .....	38
Anexo F .....	39

## Resumen

Los sectores económicos se han visto beneficiados por el avance en las ciencias informáticas, debido a que han permitido que las Tecnologías de la Información y Comunicación (TIC) se implementen en áreas económicas de forma masiva (Bachicoria, 2018). La necesidad de competir en el mundo moderno, implica a las empresas aumentar el uso de las herramientas proporcionadas por las TIC, generando crecimiento, y a su vez disminuyendo los costos de producción. Teniendo en cuenta un contexto nacional, el gobierno de Álvaro Uribe Vélez y los seguidos a éste, se comprometieron con el Plan Nacional de TIC 2008-2019 (PNTIC) a beneficiar a los colombianos, con el fin de que se informen y se comuniquen haciendo uso eficiente y productivo de las TIC para mejorar la inclusión social y aumentar la competitividad, es decir, la productividad. El PNTIC se coordinó y estuvo alineado con la visión Colombia 2019, el Plan Nacional de Desarrollo, la Política Nacional de Competitividad, el Plan de Ciencia y Tecnología y el Programa Estratégico de Uso de Medios y Tecnologías de la Información y Comunicaciones (MTIC) en la Educación, y con otros programas y proyectos que buscaron tener un efecto positivo en la competitividad del país. Así mismo, el Plan Nacional de TIC tuvo en cuenta los resultados de la Agenda de Conectividad, los Convenios de Competitividad Exportadora y la Agenda Interna, así como los proyectos sectoriales adelantados en las diferentes entidades del Estado que tuvieron a las TIC como eje central de sus acciones en el periodo de 2008-2019 (Ministerio de Comunicaciones, 2008-2019).

A partir de lo anterior, se usaron 2 herramientas aprendidas en el diplomado de *Big Data* de la Universidad Javeriana de Cali, donde se plantea un modelo de regresión lineal y una visualización en *Power Bi*, para observar el impacto de las TIC en empresas del sector servicios en el periodo 2008 - 2019, teniendo en cuenta que a lo largo de estos años aumentó

el acceso a internet por parte de los colombianos, siendo el año 2015 el que demostró un mayor acceso a esta herramienta. De mismo modo, se probó por medio del modelo de regresión lineal un resultado del score del 88%, el cual señala una relación positiva entre la variable dependiente, “total del personal ocupado” y las variables independientes, “uso de internet”, “distribución de productos en línea”, “hacer pedidos por internet”, “recibir pedidos por internet”, “servicio al cliente”, “llamadas telefónicas”, “búsqueda de información” y “uso de aplicaciones”, mostrando el impacto de las TIC en empresas del sector servicios, y a su vez su por medio del análisis de *Machine Learning*, se puede inferir un aumento en la productividad, debido a la implementación de las TIC en empresas del sector servicios.

Se recomienda para futuras investigaciones, centrarse en un sector en específico para entender mejor cuál sector tiene más dependencia relacionado a las TIC, entendiendo también si esta última, favorece a todos de la misma manera.

## Introducción

El estudio del impacto de las tecnologías de la información y las comunicaciones (TIC) en empresas colombianas del sector servicios: 2008 – 2019, las considera herramientas que inciden en las actividades económicas, con el fin de impulsar la productividad en distintos países del mundo (Huarng, 2011); de esta manera, se favorece el empleo. Desde el año 2008 en Colombia se hace la medición sobre su efecto en diferentes variables económicas a través del Índice de Desarrollo de las TIC; el cálculo de los indicadores básicos de tenencia y uso de tecnologías de la información y comunicación en las empresas colombianas cumple con hacer seguimiento de manera anual al proceso de la informatización de la economía, esto como consecuencia de la globalización de estas tecnologías en la década del siglo XX, para ello la información se recolecta a partir de encuestas económicas manufactureras, de comercio y de servicios (DANE, 2008).

A partir de las diferentes necesidades que surgieron en la economía se detectó un cambio en la estructura ocupacional de las economías de los países industrializados; esto generó la necesidad de investigar a cerca de la incidencia de las TIC en el desarrollo de los diferentes sectores económicos, dando origen al término: “economía de la información”, éste ha sido interpretado de distintas formas por varios economistas. Fritz Machlup (1962) fue el economista que dio origen al concepto “economía del conocimiento”, y midió la dimensión de la misma teniendo en cuenta cómo contribuye a sectores como educación, medios de comunicación, equipamiento de información, servicios de información y otras actividades. De este modo, se observó que la fuerza de trabajo en Estados Unidos aumentó en un 31% (Katz, 2009). En consecuencia, se ha podido demostrar que en Colombia ha existido un aumento de la demanda laboral de diferentes puestos de trabajo en específico los cuales de

acuerdo con el Ministerio del Trabajo son los sectores de ciberseguridad, *Big Data*, computación en la nube e inteligencia artificial, así como la productividad laboral en Colombia tuvo un aumento del 27% gracias a la implementación de las TIC en los diferentes sectores económicos, entre los cuales destacan principalmente el sector de servicios, industria y comercio los cuales cerca del 60% utilizan las TIC como herramienta para mejorar sus procesos productivos (MinTic, 2019).

Por otro lado, Jonscher plantea que hay una relación directa entre la proporción de trabajadores de la información y el desarrollo económico; del mismo modo, plantea que el impacto de las TIC en la productividad produce un efecto de rezago temporal. Lo que indica que la inversión en TIC no genera un impacto inmediato en la productividad, si no que ésta está sujeta al cambio en diferentes variables como por ejemplo el conocimiento de los colaboradores de las empresas en cuanto a las TIC, así como los procesos de coordinación de la implementación de las TIC en las diferentes empresas, así mismo a partir del desarrollo de la presente investigación se evidenció como variables independientes ayudan y potencializan al crecimiento económico generando empleo ya que a partir del uso de las TIC en variables como el servicio al cliente, distribución de mercancías, búsqueda de información del mercado y comunicación a través de videoconferencias se mejora operativamente las condiciones de la empresa lo que por ende conduce al crecimiento de la compañía a su vez.

El objetivo de este trabajo de investigación es aplicar un modelo de regresión lineal con base a las diferentes construcciones teóricas encontradas a través de la revisión de literatura, que permita ver la relación entre las diferentes variables que inciden en el desarrollo de la labor del personal ocupado que trabajan en empresas que usan las TIC y las distintas actividades que dichas empresas realizan, como lo son: hacer, recibir y distribuir productos en línea, enviar o recibir correos electrónicos, realizar búsquedas de información

sobre bienes y servicios, atención de servicio al cliente por medio de las TIC, realizar llamadas telefónicas y el uso de aplicaciones. Para ello se lleva a cabo un estudio de la regresión lineal en la cual se utiliza el “total de personal ocupado” como variable dependiente, y el “uso de internet”, “distribución de productos en línea”, “hacer pedidos por internet”, “recibir pedidos por internet”, “servicio al cliente”, “llamadas telefónicas”, “búsqueda de información” y “uso de aplicaciones” (DANE)<sup>1</sup>, como variables independientes. Se tienen en cuenta empresas del sector servicios, donde se muestran empresas como: hoteles y restaurantes, actividades de edición, actividades complementarias de transporte y comunicaciones, inmobiliarias, educación superior privada, salud humana privada, radio, televisión, juegos de azar, actividades deportivas, recreativas y otros servicios, en los periodos comprendidos entre el 2008 al 2019. Esto con el fin de aplicar lo aprendido en el diplomado de *Big Data: para la toma de decisiones* de la Universidad Javeriana de Cali.

De acuerdo con los resultados obtenidos a partir del análisis de variables, se pudo determinar que el modelo contiene un R-cuadrado del 0.88 y las predicciones se ajustan bien al modelo, teniendo en cuenta que se usaron 2.197 datos para llevar a cabo la elaboración del modelo, el cual contiene 15 variables en las cuales se evalúa el uso de servicios tecnológicos dentro de las empresas del sector servicios. Las variables utilizadas dentro del modelo dan cuenta de que tanto las empresas hacen uso de herramientas tecnológicas, dichas variables son: uso de internet, distribución de productos en línea, recibir productos por internet, capacitación de personal, servicio al cliente, llamadas telefónicas, búsqueda de información sobre bienes y servicios; cabe resaltar, que para algunos años no hay disponibilidad de datos con respecto a algunas variables.

---

<sup>1</sup> Información suministrada de la base de datos del DANE, *Indicadores Básicos de TIC en empresas*

También se observa que, ante una mayor frecuencia en el uso de internet en el desarrollo de las actividades en una empresa, independientemente del área en el que estén enfocados sus servicios, así como una mayor distribución de sus productos en línea, tendrá relaciones positivas con el personal ocupado. Esto favoreciendo al sector empresarial debido a que los avances tecnológicos, actualmente se cuenta con más herramientas para impulsar sus ingresos y de mismo modo incentivar a las pequeñas empresas (pymes) a usar las TIC con el fin de incrementar sus ganancias, generar empleo y tener un negocio estable.

### **Revisión de la literatura**

La revisión de antecedentes se tomará teniendo en cuenta cada uno de los conceptos a estudiar, que en este caso son aquellos estudios o revisiones relacionadas con la evolución de las TIC, así como aquellas investigaciones que brindan un panorama más amplio de lo que es el *Machine Learning*.

En primera instancia, es necesario entender que una regresión lineal múltiple desde un punto de vista matemático surge con el fin de mostrar la distribución de probabilidades de variables aleatorias, o se usa para ajustar alguna función a un conjunto de datos; en esta investigación se utiliza el segundo punto de vista. Teniendo en cuenta lo anterior, (Granados, 2016) a lo largo de su estudio explica la forma correcta de realizar un modelo con creación de variables *Dummies* configurando un individuo de referencia, factores de ponderación, variables de interacción, interrelación, entre otros.

Al respecto, Maisueche (2019) tiene como principal objetivo en su investigación, determinar las diferentes variables que inciden en el éxito de la aplicación de las herramientas tecnológicas dentro de una compañía a través del *Machine Learning*, y a su vez hace un

estudio de la importancia del *Machine Learning* en industrias como: logística, mantenimiento, negocios del entorno empresarial, marketing, entre otros. Teniendo en cuenta los beneficios que esperan obtener las industrias por medio de la digitalización como: aumentar la eficiencia de los procesos, mejorar la eficiencia en la gestión y el ahorro de tiempo junto con la accesibilidad ágil a la información. Así mismo, Maisuche añade que no hace falta ser una gran empresa para superar la transformación digital, sino que la clave consiste en realizar un análisis interno para identificar aquellos puntos fuertes y débiles que son susceptibles de ser mejorados ya sea con esta tecnología o con otras.

Así mismo, el trabajo de Alderete (2014) realiza un aporte a la literatura empírica sobre los determinantes de la adopción de las TIC en empresas industriales de Argentina. A partir de una muestra con micro datos, en una encuesta tecnológica realizada a 163 empresas de estas tramas en el año 2006, derivada de un proyecto de investigación previo, se estiman diferentes modelos econométricos para identificar las variables explicativas de la adopción de las TIC en las empresas de las industrias de Argentina. En el cual se evidencia una relación positiva entre la incorporación de las TIC y el desempeño de las empresas, al aumentar la productividad, con la mejora de su posición competitiva en el mercado.

De acuerdo a Alderete (2014), hay estudios que constatan la presencia de procesos dinámicos de adopción donde, existen restricciones de acceso a las TIC por el lado de la demanda, como la capacidad financiera reducida de las empresas, según el autor, aquellas empresas con menor capacidad de inversión tardarán más en adoptar la innovación tecnológica. Así mismo, se observa un conjunto de barreras que surgen del lado de la oferta (infraestructura de conectividad, calidad y costos de los servicios), que deberían considerarse para el análisis del uso y difusión de las TIC.

Por otra parte, Pulido (2020) en su investigación, tiene como objetivo principal determinar factores que influyen en la adopción del *Machine Learning* en empresas colombianas, realiza un análisis de factores técnicos y organizacionales que influyen en el uso de esta herramienta en el entorno empresarial. Determina que los factores organizacionales son aquellos que vienen determinados por la estructura de la organización, así como sus objetivos, principios y valores siendo la suma de diferentes variables internas, por otro lado, los factores técnicos son procesos correspondientes a la operación de la organización, de esta manera, tanto los factores organizacionales como los técnicos son influyentes en la adopción de *Machine Learning*. Como resultado de la investigación se obtuvieron ocho factores los cuales fueron propuestos en un modelo de investigación, el cual fue validado a través de una encuesta aplicada a empresas colombianas. Mediante estadística multivariante con modelamiento de ecuaciones estructurales. Pulido evidenció que las condiciones facilitadoras, facilidad de uso, utilidad, apoyo a los directivos y gestión de datos son influyentes en la adopción de *Machine Learning* en un entorno empresarial colombiano.

Con lo anterior, Pulido (2020) sostiene que la *multimetodología* la cual define como una metodología de investigación fusionado el cual permite hacer el uso de diferentes herramientas en una misma investigación, así éstas no compartan el mismo paradigma, lo que se conoce como *multimetodología multiparadigma*. La primera de las limitaciones de dicha metodología está ligada al uso de la Teoría Fundamentada en las observaciones en sitio. Esto limita el uso de algunos conceptos atados a la Teoría Fundamentada, como lo es el criterio de la saturación teórica.

En lo que respecta a otros estudios relacionados con *Machine Learning*, Ramírez (2018) profundiza sobre el subcampo de las ciencias de la computación, teniendo en cuenta que una de sus ramas es la inteligencia artificial, cuya finalidad es implementar técnicas para tratar bases de datos de gran volumen. En el transcurso del documento presenta las diferentes herramientas que se pueden usar para realizar predicciones como lo son principalmente los estudios estadísticos, y/o otras herramientas que facilitan el manejo de información. Su estudio finaliza mostrando los campos de aplicación en los cuales se ha enfocado *Machine Learning*, el cual corresponde a la medicina, la construcción, las finanzas, la robótica, la educación, entre otras, a su vez revisa cuáles son sus grandes fortalezas, oportunidades, debilidades y amenazas para la humanidad como lo son el cambio climático y el aumento de la población.

Por otro lado, (Cuadrado, 2019) concluye que las empresas aprovechan todos los avances que se generan en el mundo a través de la tecnología. De esta manera, dichas empresas utilizan el *Machine Learning*, el cual no es más que una técnica asociada a la detección automática de patrones relevantes dentro de un conjunto de datos; así mismo, el recurso cuenta con distintas herramientas para manipular y analizar datos a través del uso sistémico de herramientas tecnológicas. Así pues, una de las técnicas más usadas por *Machine Learning* son las regresiones lineales, en este orden de ideas, resulta importante mencionar los aportes de esta investigación ya que se puede relacionar el crecimiento de las diferentes industrias a partir de la aplicación del *Machine Learning*, realizando una regresión lineal múltiple que permita identificar la relación entre las diferentes variables del uso de las TIC en las empresas.

Con respecto a estudios relacionados con la tecnología de la información (García, 2021), expone en su análisis, los medios informativos utilizan nuevos conceptos para mostrar la evolución de la economía, mostrando que la sociedad del conocimiento es uno de estos nuevos conceptos teniendo en cuenta que se fundamenta en las Tecnologías de la Información y Comunicaciones (TIC). El análisis económico que realiza (García, 2021) con respecto al impacto social se realiza de manera horizontal y vertical, así pues, en el horizontal se tiene en cuenta el comportamiento de los agentes económicos y de los individuos frente al desarrollo y accesibilidad a las TIC. De esta manera, el enfoque horizontal nos plantea el crecimiento de la productividad como un determinante de la elevación del nivel de vida, concluyendo que el incremento de la productividad inducido por las TIC es una fuente de generación de empleo.

Por último, para tener un análisis de la influencia de las TIC en la economía chilena, (Leiva, 2003), plantea que el surgimiento de la Nueva Economía basada en el uso de las TIC ha creado una nueva área de estudios con impactos macroeconómicos en tres sectores: el mercado laboral, el nivel de precios y los efectos en la productividad total de los factores. En este sentido la innovación de las TIC, tiene consecuencias en el PIB, debido a que favorece el proceso productivo y a su vez incrementa la tasa de crecimiento de la economía. El impacto en el mercado laboral es positivo, debido a que se crea empleo asociado al sector de las TIC. Por esto, se ha evidenciado que el Ministerio de las TIC ha invertido en programas que fomenten la capacitación en ciencia de datos o fundamentos de analítica para toda la población en general, esto con tal de impulsar el trabajo en estos campos y expandir las oportunidades en la población.

Según las encuestas realizadas por el DANE se muestra que en Colombia cerca del 40% de las empresas han adoptado las TIC como una herramienta para su crecimiento y desarrollo. Por ende, el retraso que presenta el país en las TIC se ve reflejado en las pymes en todos los niveles, y, en forma reiterada, confirmado por las encuestas y los estudios, ya que las empresas grandes cuentan con un mayor porcentaje de adhesión al uso corporativo de las TIC. Según las encuestas, el uso que se les da a las TIC es superficial y hay desaprovechamiento y desconocimiento de las capacidades de los sistemas computacionales y de los diferentes productos para usos empresariales.

Teniendo en cuenta lo anterior, en la revisión de la literatura se observa la necesidad de analizar el impacto de las TIC en las empresas, dado que se ha mostrado una ventaja importante dentro de la economía nacional, como internacional, tal y como se evidencia con trabajos de los autores mencionados.

## Desarrollo

### Proceso metodológico

En primera instancia, para iniciar con el proceso metodológico del trabajo, se aclara que en el diplomado de *Big Data: para la toma de decisiones*, se aprendió a tener un buen manejo de las bases de datos, y a su vez realizar visualizaciones en *Power Bi* de cualquier base de datos. Debido a lo anterior, se tomaron las estadísticas proporcionadas por el Departamento Administrativo Nacional de Estadística (DANE, s.f.), en los Indicadores Básicos de TIC en empresas – Históricos, donde se presenta información a partir del año 2008 hasta el año 2019. Luego de obtener los datos se exportaron estos mismos a la interfaz *Colab* la cual fue la herramienta utilizada para generar el código de *Python*, de esta manera luego de haber procesado los datos en dicha interfaz se realizaron los cálculos para determinar la relación entre las variables utilizando un modelo de regresión lineal múltiple.

Los Indicadores Básicos de TIC en empresas realiza una clasificación en el personal ocupado de las empresas que usaron herramientas TIC para su trabajo, esta categoría provee información del aumento del empleo con respecto a los años de estudio. A partir de lo anterior, se descargan todos los anexos que se encuentran en la plataforma del DANE, del año 2008 al 2019.

Al realizar la descarga de las bases de datos, se realizan diferentes filtros en los que se elimina la información que no es necesaria para el análisis, ya que una de las herramientas principales del *Big Data* es limpiar la base de datos con el fin de usar únicamente los datos que necesitan y así al momento de realizar la regresión lineal, el programa leerá los datos de una forma sencilla, así mismo, la información cualitativa se convierte en cuantitativa para así facilitar el cálculo de relaciones entre variables. Seguidamente, se utiliza el Coeficiente de

Correlación en el cual se toma como perspectiva que, si el mismo resulta ser un valor mayor o igual 0.7, se pueden ajustar los datos mediante una regresión lineal, teniendo en cuenta que puede existir cierto error, en caso tal de que el porcentaje resulte menor a 0.7, la linealidad no será evidente y se podrá proceder a buscar algún modelo no lineal que mejor ajuste a la predicción que se desea obtener (D'Angiolo, 2019).

Para llevar a cabo el análisis, se define una base de datos como un conjunto de valores, que se pueda mostrar de la siguiente manera:

$$X = X_1, X_2, \dots, X_n \text{ donde } x_n \in R^n$$

Teniendo en cuenta que cada dato de entrada está señalado por un valor real  $y_1$

$$Y = y_1, y_2, \dots, y_n \text{ donde } y_n \in R^n$$

El modelo lineal de predicción toma la siguiente regla para basar la predicción

$$f(x) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n,$$

$$A = \alpha_0, \alpha_1, \alpha_2, \dots, \alpha_n,$$

Lo anterior teniendo en cuenta que  $\alpha$  es una constante que en el caso de este modelo al ser constantes no nulas corresponde a una regresión lineal multivariable (D'Angiolo, 2019).

### **Modelo de regresión lineal múltiple**

Los datos que se usaron para llevar a cabo el modelo de regresión lineal son el total de personal ocupado en el sector servicios como variable dependiente, teniendo en cuenta que en dicho sector se usan herramientas de las TIC, y como variables independientes el total de empresas que usan TIC, y las empresas que usaron plataformas electrónicas para vender, comprar insumos, enviaron o recibieron correos electrónicos, realizaron búsqueda de

información, usaron aplicaciones, realizaron llamadas telefónicas por medio de la implementación de TIC.

Teniendo en cuenta lo anterior, se realizó la importación de las librerías tal y como se evidencia en el anexo a, con el fin de empezar a codificar para crear una interfaz independiente.

Al realizar el modelo de predicción, anteriormente se procedió a importar la base de datos modificada, es decir teniendo en cuenta las variables cuantitativas y cualitativas debido a que *Python* es sensible a errores en el formato de los datos. Véanse los anexos.

Continuamos con la verificación del tipo de cada una de las variables, lo que indica que la información cuantitativa esté designada por float o int64 que nos corresponde a información numérica y la información cualitativa este designada por object, tal y como se muestra en la figura 1.

## Figura 1

### *Verificación del tipo de datos en Python*

```

✓ [25] datos = datos.dropna()
0s      datos.dtypes

Index          int64
Sector         object
Ingresos       int64
Total de personal ocupado  int64
Total de empresas  int64
Uso de Internet  int64
Distribuir productos en línea  float64
Recibir pedidos por Internet  float64
Hacer pedidos por Internet  float64
Capacitación de Personal  float64
Enviar o recibir correo electronico  float64
Busqueda de Informacion Sobre bienes y servicios  float64
Llamadas telefónicas por Internet VoIP  float64
Servicio al Cliente  float64
Uso de aplicaciones  float64
dtype: object

```

*Nota.* En esta imagen se muestra la importación de la librería y la clasificación de las variables en Python. Fuente: Elaboración propia.

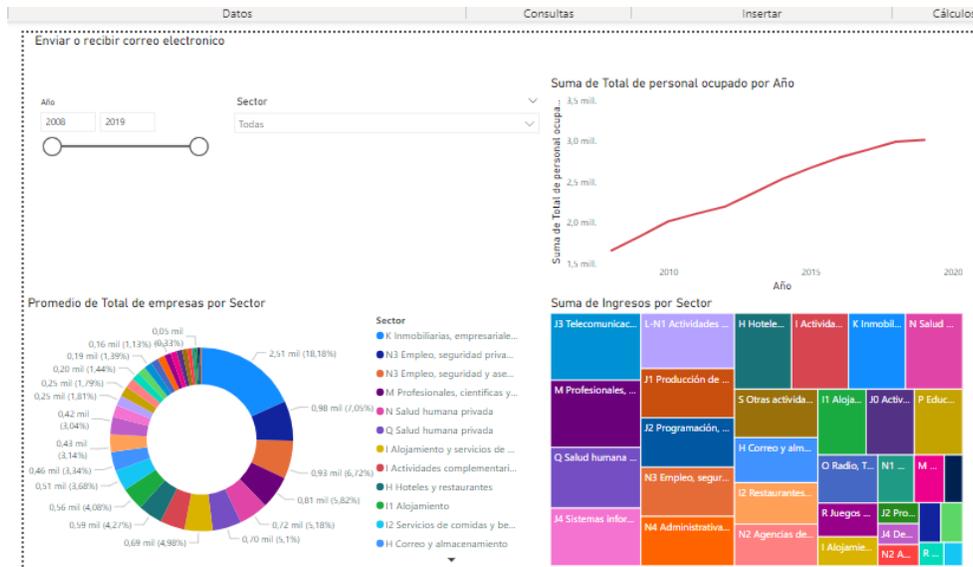
### **Visualización en *Power Bi***

Las visualizaciones en *Power Bi* son una herramienta clave del *Big Data*, ya que se usan con el fin de reflejar información contenida en las bases de datos de una forma rápida y concisa, para ello se importó al aplicativo la misma base de datos que se usó para llevar a cabo la regresión lineal en *Python*.

Teniendo en cuenta la base de datos que se importó, en el costado derecho del aplicativo se selecciona el tipo de visualización o visualizaciones que se adecuen para proyectar los datos. En este trabajo se realizaron 2 paneles, los cuales están compuestos de 3 visualizaciones cada uno, el primero se elaboró teniendo en cuenta el total del personal ocupado, el total de empresas por sector y el total de ingresos por sector, tal y como se muestra en la figura 2, considerando que es un panel aplicativo en el cual el lector puede navegar y seleccionar los años que desea visualizar y de mismo modo puede seleccionar el sector que dese analizar. Por otro lado, el segundo panel se elaboró teniendo en cuentas las actividades que implican el uso de las TIC, tal y como se evidencia en el anexo c.

### **Figura 2**

*Panel No.1 en Power Bi*

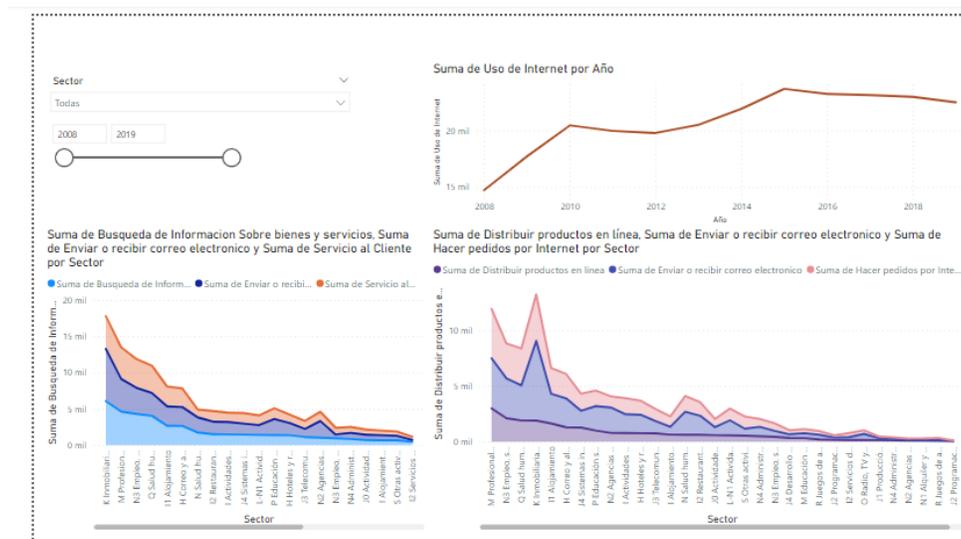


Nota. En esta imagen se muestran de manera gráfica los datos tomados del DANE.

Fuente: Elaboración propia

Figura 3

Panel No.2 en Power Bi



Nota. En esta imagen se muestran de manera gráfica los datos tomados del DANE.

Fuente: Elaboración propia

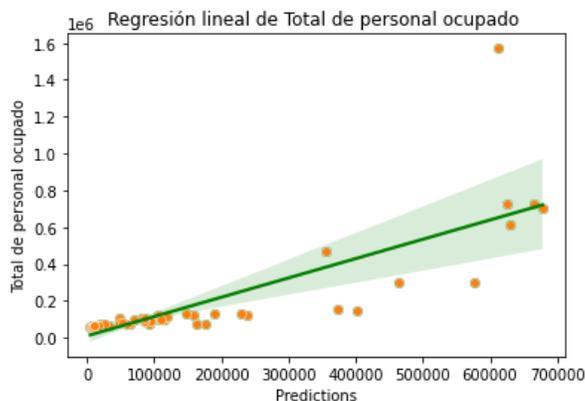
## Resultados

### Análisis de resultados del modelo de regresión lineal múltiple

El modelo de regresión lineal múltiple que se llevó a cabo en la elaboración de este trabajo se usó con el fin de explicar el comportamiento del personal ocupado con respecto al uso de herramientas de las TIC en empresas del sector servicios. Para ello, se tuvo en cuenta la relación de la variable dependiente con las variables independientes. A partir de lo anterior, el modelo ayuda a reflejar la relación de las variables, dividiendo los datos en un 70% para generar el modelo y un 30% para la prueba y error de este, con el fin de tener un modelo preciso, tal y como lo vemos en la figura 4, en la cual se muestra una relación positiva entre la variable dependiente y las variables independientes.

#### Figura 4

*Regresión lineal Total de personal ocupado en Python*



*Nota.* En esta imagen se muestra la relación positiva entre las variables dependientes e independientes. Fuente: Elaboración propia

La línea ajustada verde muestra gráficamente que, si uno se mueve hacia la izquierda o hacia la derecha en el eje x, una cantidad que represente el uso del internet será reflejado en el total de personal ocupado en empresas del sector servicios. Para ello, se tiene en cuenta

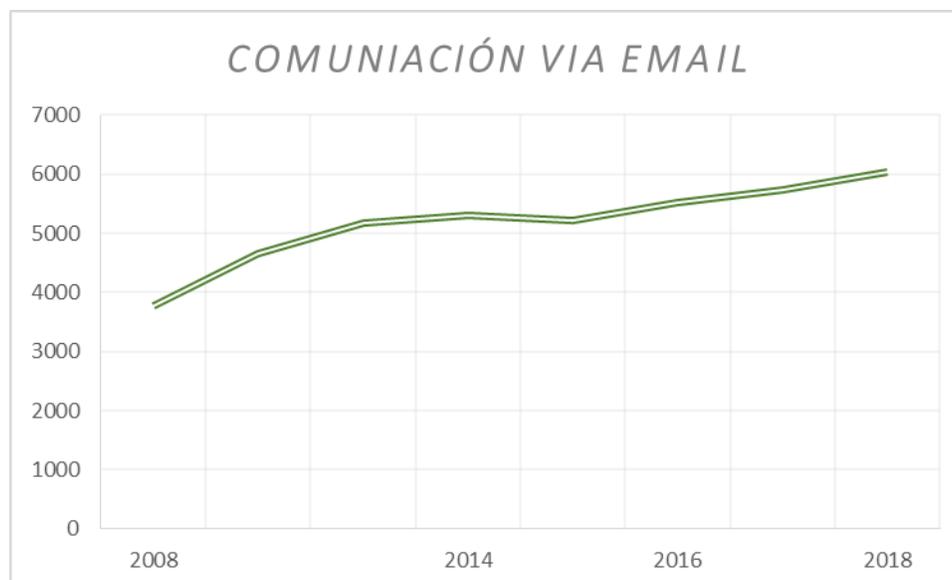
que son objeto de este análisis las empresas de servicios que cumplen los siguientes criterios y se dedican a las actividades de: hoteles y restaurantes (sección H) con personal ocupado igual o superior a 40 personas e ingresos anuales iguales o superiores a \$3.000 millones; almacenamiento, comunicaciones y actividades auxiliares de transporte (sección I) con personal ocupado igual o superior a 40 personas o ingresos anuales iguales o superiores a \$3.000 millones; actividades inmobiliarias, empresariales y de alquiler (sección K) con personal ocupado igual o superior a 75 personas o ingresos anuales iguales o superiores a \$3.000 millones; actividades de educación superior privada (sección M) con personal ocupado igual o superior a 20 personas o ingresos anuales iguales o superiores a \$1.000 millones; actividades relacionadas con la salud humana (sección N) con personal ocupado igual o superior a 40 personas o ingresos anuales iguales o superiores a \$3.000 millones; y entretenimiento y otros servicios (sección O) con personal ocupado igual o superior a 40 personas o ingresos anuales iguales o superiores a \$2.000 millones. La relación de la variable dependiente e independiente se da únicamente dentro este rango de datos.

Por otro lado, debemos tener en cuenta la significación del F-Test, que, en este caso, es mayor a 0.8890, lo que quiere decir que las variables independientes explican y tienen relación con la variable dependiente. El  $r$  cuadrado, explica la varianza de la variable dependiente, en este caso corresponde a 0.8890, siendo un valor mayor al dato mínimo para llevar a cabo la regresión lineal, mostrando que es un modelo explicativo en el cual la relación entre las variables independientes con la variable dependiente muestra que a lo largo de los años el empleo se vio impulsado y este comportamiento se ve reflejado en el implemento de actividades de las TIC, en actividades como realizar uso del internet, enviar o recibir correos electrónicos, búsqueda de información sobre bienes y servicios, llamadas telefónicas, distribuir, recibir y hacer pedidos por internet.

Dado lo mencionado anteriormente, se evidenció que en los años de estudio la actividad que más se desarrolló en empresas del sector servicios fue la comunicación vía email tal y como se proyecta en la figura 5.

### **Figura 5**

*Total de la comunicación vía email en empresas del sector servicios*



*Nota.* En esta imagen se proyecta la actividad que más se realizó en empresas del sector servicios. Fuente: Elaboración propia

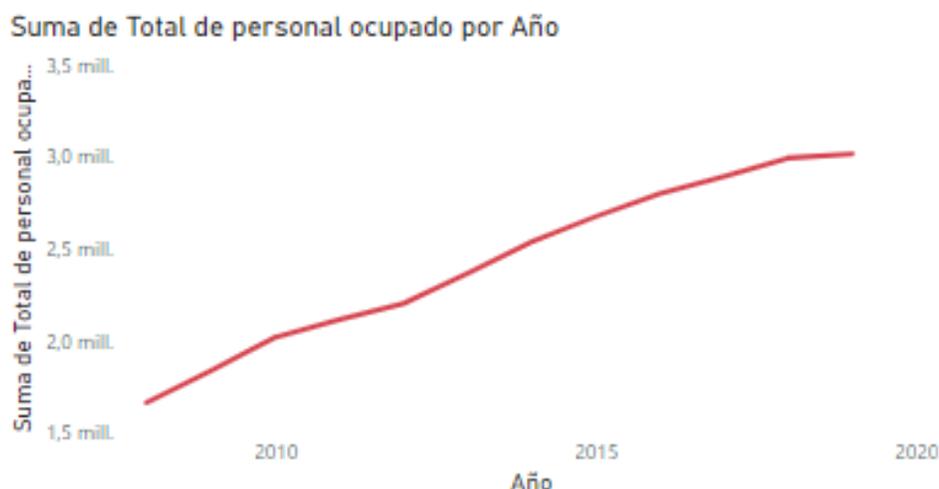
### **Análisis de resultados de la Visualización en *Power Bi***

Las visualizaciones que se elaboraron con el fin de detallar de una forma sencilla la información contenida en la base de datos muestra la relación positiva de las variables dependientes con respecto a las variables independientes. De esta forma, se evidencia el incremento del personal ocupado en empresas del sector servicios; el año 2019 es aquel que muestra la mayor cantidad de personal ocupado con 1.689.658 personas trabajando en el

sector servicios tal y como se muestra en la figura 6, siendo el sector de salud humana y privada el que contó con mayor personal.

### Figura 6

*Total personal ocupado en empresas del sector servicios*

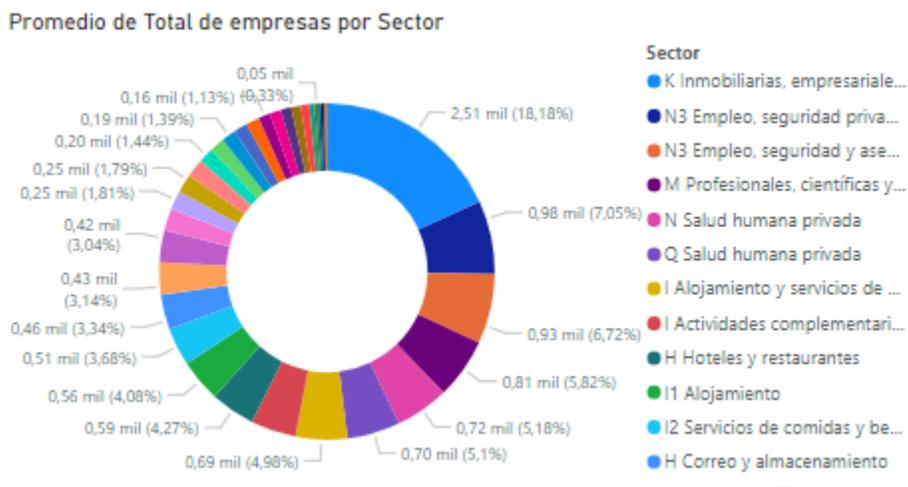


*Nota.* En esta imagen se expone el total del personal ocupado en empresas del sector servicios en 2008-2019. Fuente: Elaboración propia

Por otro lado, se realiza una visualización en donde se evidencia que el sector de servicios inmobiliarios es aquel que cuenta con una cantidad mayor de empresas, seguido del sector de servicios de seguridad privada y servicios de aseo tal y como se evidencia en la figura 7. A su vez se tiene en cuenta en la visualización de la figura 8, las actividades de implementación de las tic en empresas del sector servicios, esta siendo una visualización interactiva en la que el lector puede elegir tanto el sector como el año donde quiera ver información.

### Figura 7

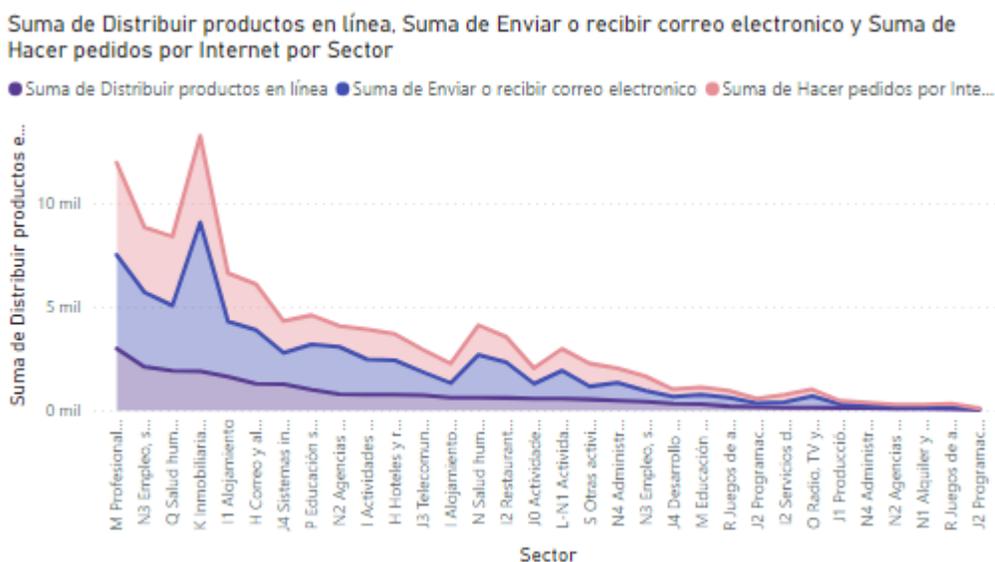
Total de empresas por sector servicios



Nota. En esta imagen se muestra por medio de una visualización el total de empresas del sector servicios en 2008-2019. Fuente: Elaboración propia

Figura 8

Actividades de implementación de las TIC en sector servicios

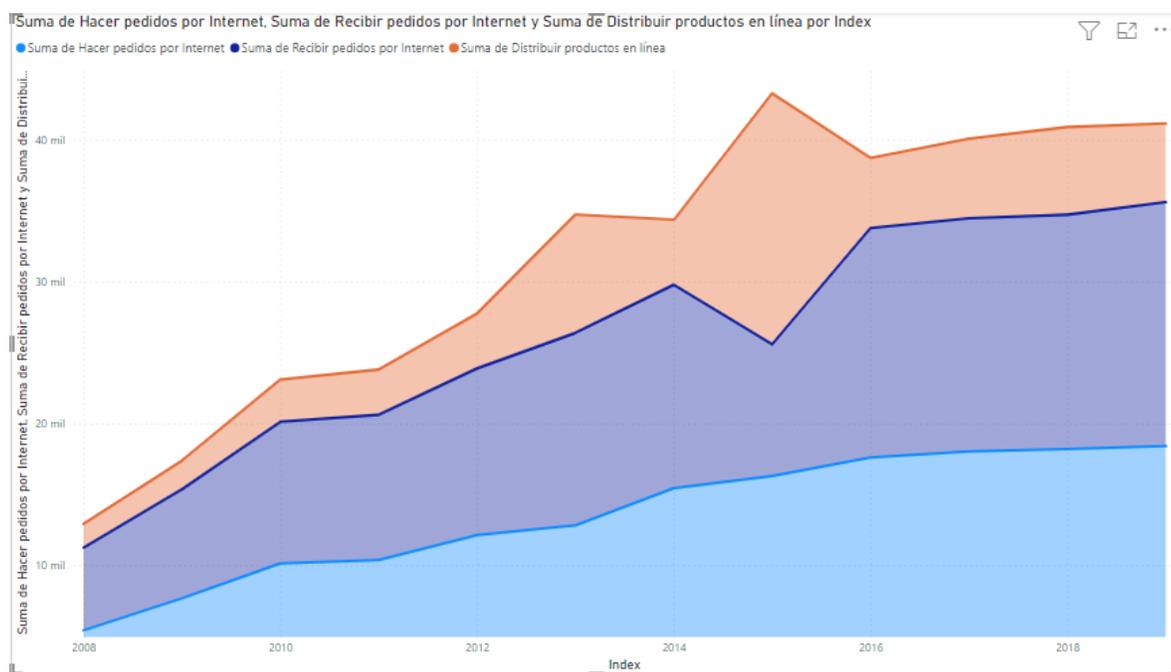


Nota. En esta imagen se expone por medio de una visualización las actividades que desarrollan las empresas del sector servicios en 2008-2019. Fuente: Elaboración propia

La realización de un análisis más detallado de la suma al hacer, recibir y distribuir pedidos por medio de Internet, se tiene en cuenta que el procedimiento se puede realizar por cualquier dispositivo móvil y que se realizó de manera automática la suma de dichas variables en todos los sectores comprendidos en el estudio conociendo que los periodos comprendidos son del 2008 al 2019 tal y como se evidencia en la figura 9. Además de esto se demuestra el crecimiento a lo largo del tiempo de cada una de las variables, afirmando que cada vez el acceso a las TIC es mayor.

### Figura 9

*Hacer, recibir y distribuir pedidos por internet del 2008 al 2009*



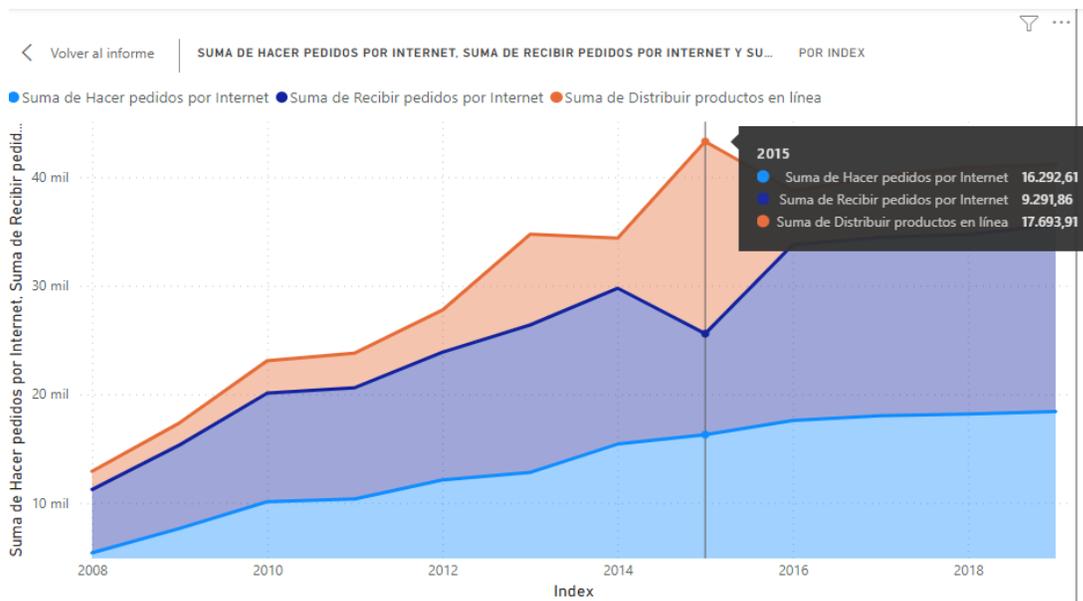
*Nota.* En esta visualización se refleja las actividades de hacer, recibir y distribuir pedidos por internet en empresas del sector servicios 2008-2019. Fuente: Elaboración propia

La elaboración con un análisis detallado de la figura 10, donde se puede observar que en todos los años es mayor la cantidad de productos que distribuyen comparado con hacer y recibir productos. por internet, siendo el año 2015 con 17.693 distribuciones de

productos en línea, lo que se puede reflejar como ventas hechas por internet, esto debido a que el cuarto trimestre del año en estudio, represento el 21,30% del total de accesos a internet banda ancha en el país (MinTic, 2016).

## Figura 10

*Hacer, recibir y distribuir pedidos por internet en el año 2015*

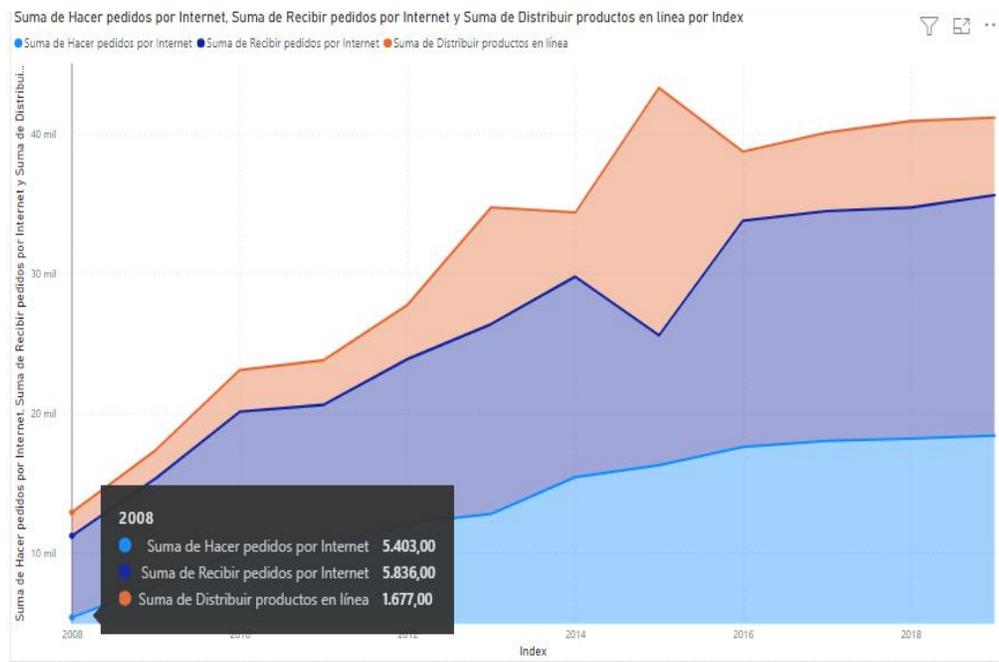


*Nota.* En esta visualización se muestran las actividades de la figura 9 en el año 2015. Fuente: Elaboración propia

Por otro lado, en el año que iniciaron los estudios, es decir en el 2008 la suma de hacer pedidos por internet correspondía a 5403 unidades, distribuir productos correspondía a 5836 unidades y recibir productos por este medio correspondía a 1677 unidades, tal y como se evidencia en la figura 11. Si se realiza la comparación con el último año de estudio que es el año 2019, se evidencia un aumento exponencial en las 3 variables, debido al avance tecnológico el país, tal y como se evidencia en la figura 12.

## Figura 11

### Hacer, recibir y distribuir pedidos por internet en el año 2008

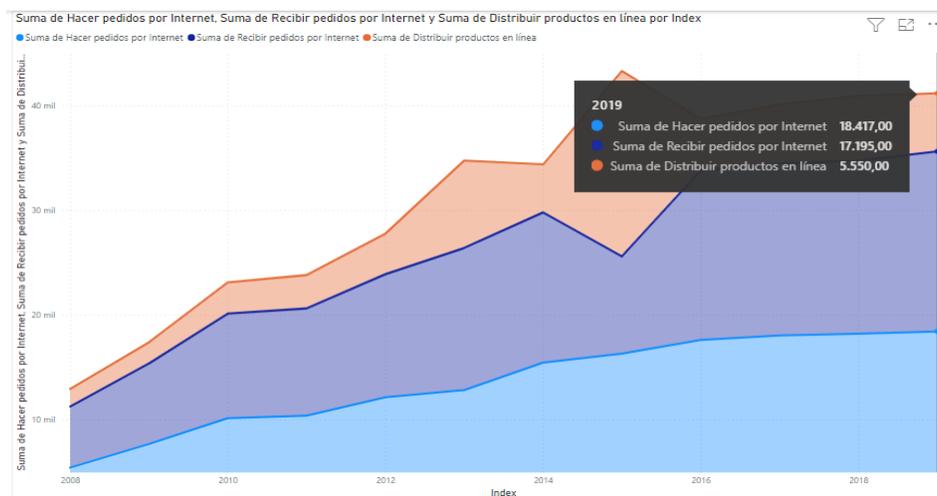


*Nota.* En esta visualización se muestran las actividades de la figura 9 en el año 2008. Fuente:

Elaboración propia

### Figura 12

### Hacer, recibir y distribuir pedidos por internet en el año 2019



*Nota.* En esta visualización se muestran las actividades de la figura 9 en el año 2019. Fuente:

Elaboración propia

## Conclusiones

1. En primer lugar, por medio del modelo se evidencia que las Tecnologías de la Información y Comunicación, impulsan el empleo, mostrando el aumento de personal ocupado en el sector servicios, siendo el sector de servicios inmobiliarios, servicios de aseo y salud humana privada los sectores con mayor personal ocupado, que usan en sus ocupaciones diarias herramientas de las TIC.

2. Realizar el modelo de regresión lineal y la visualización se puede concluir que el país debe seguir con los mecanismos para impulsar a las empresas con herramientas de las TIC, con el fin de incrementar el desarrollo de la tecnología, teniendo en cuenta que en la actualidad el comercio electrónico es cada vez más importante en la economía del país, según la cámara de comercio de Bogotá

3. El *Machine Learning* como herramienta del *Big Data*, para mostrar y visualizar la evolución de las TIC en las empresas de distintos sectores colombianos, funcionó como herramienta de predicción para llevar un análisis más preciso enfocado en la economía colombiana.

4. Teniendo en cuenta el enfoque horizontal de la nueva economía, se puede concluir que la elevación del nivel de calidad de vida, depende de la estrategia que adopta una empresa cuando ofrece sus productos en el mercado. Así, se muestra la importancia de realizar un análisis enfocado en el aumento de la producción creada por las TIC, siendo una fuente generadora de empleo para los colombianos.

5. La inversión en TIC no genera un impacto inmediato en la productividad, si no que este va acompañado de otras variables independientes, para ver el resultado de este,

tal y como se evidenció a lo largo de esta investigación con el fin de potencializar el crecimiento económico aumentando la cantidad de personal ocupado.

6. Se recomienda para futuras investigaciones, centrarse en un sector en específico del sector servicios para entender mejor cuál sector tiene más dependencia relacionado a las TIC, entendiendo también si esta última, favorece a todos de la misma manera. Así mismo, se puede explorar la relación que hay de las redes sociales en las empresas colombianas.

7. Se evidencia que a pesar de que hay disponibilidad de datos en cuanto a variables como el uso del internet, aún existe una deficiencia en cuanto a los datos que son obtenidos por el Ministerio de las TIC, por lo que se propone aumentar el número de datos con el fin de obtener mejores resultados en cuanto al análisis del impacto del uso de las TIC en el aumento de la productividad en el país.

8. Lo aprendido en el diplomado de *Big Data*: para la toma de decisiones de la Universidad Javeriana de Cali, contribuyó a mi conocimiento personal para realizar un aporte y así realizar un estudio del impacto de las TIC en empresas colombianas del sector servicios en 2008 al 2019, por medio de las herramientas proporcionadas por este diplomado, las cuales fueron el uso del aplicativo *Power Bi*, para un análisis detallado de la base datos y a su vez realizar la limpieza de la misma, para realizar el modelo de regresión lineal que se llevo a cabo a lo largo del trabajo.

9. Pese a que se evidencia una relación positiva en el modelo en cual se mostraron variaciones que incrementan año tras año en el personal ocupado de empresas del sector servicios. Colombia debe implementar políticas en el área de las TIC con el fin de aumentar la competitividad a nivel internacional.

10. Si bien este trabajo, proporciona un modelo de regresión lineal, cabe resaltar que fue desarrollado en el entorno colombiano, por lo que da pie a realizar estudios en otros países o incluso tomar otras variables del uso de las TIC y así tener una muestra mayor para Colombia.

## Referencias

- Alderete, M. V., Jones, C., & Morero, H. A. (2014). Factores explicativos de la adopción de las TIC en las tramas productivas automotriz y siderúrgica de Argentina. <https://doi.org/10.14482/pege.37.7019>
- Cuadrado, A. M. (2019). UTILIZACIÓN DEL MACHINE LEARNING EN LA INDUSTRIA. España.
- DANE. (2008). Indicadores básicos de TIC en Empresas - Históricos. Obtenido de <https://www.dane.gov.co/index.php/estadisticas-por-tema/tecnologia-e-innovacion/tecnologias-de-la-informacion-y-las-comunicaciones-tic/indicadores-basicos-de-tic-en-empresas/indicadores-basicos-de-tic-en-empresas-historicos>
- D'Angiolo, F. G. (2019). Algoritmos de Regresión Lineal aplicados al mantenimiento de un Data ceter. Argentina.
- Departamento Administrativo Nacional de Estadística. (s.f.). Obtenido de <https://www.dane.gov.co/>
- Díaz, J. G. (2021). Importancia y evolución del mercado de las tic.
- Hempell, T. (2002). Does experience matter? Innovations and the productivity of ICT in German services. Discussion paper, 02-43. Centre for European Economic Research. Recuperado de <ftp://ftp.zew.de/pub/zew-docs/dp/dp0243.pdf>
- Huarng. (2011). A comparative study to classify ICT developments by economies.
- IBM. (s.f.). Obtenido de <https://www.ibm.com/co-es/analytics/machine-learning>
- Katz, R. (2009). El Papel de las TIC en el Desarrollo. Recuperado de: [https://books.google.com.co/books?id=-PaSgi8M7n8C&printsec=frontcover&hl=es&source=gbs\\_ge\\_summary\\_r&cad=0#v=onepage&q&f=false](https://books.google.com.co/books?id=-PaSgi8M7n8C&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false)
- Leiva, D. A. (2003). Impacto de las tecnologías de información y comunicación en la economía chilena. Santiago.
- Maisueche, A. (2019). Utilización del machine learning en la industria 4.0. Universidad de Valladolid, 1, 27–36
- Marchese, A. & Jones, C. (2011). Sistemas de información y las Tecnologías de Información y Comunicaciones (TIC) como procesos de aprendizaje en organizaciones de Córdoba y Rosario. Estudio de Indicadores. 40JAIIO – Simposio de la Sociedad de la Información. Córdoba, Argentina.
- Ministerio de Telecomunicaciones de Colombia. (2008-2019). Obtenido de [https://www.mintic.gov.co/portal/715/articles-125156\\_recurso\\_00.pdf](https://www.mintic.gov.co/portal/715/articles-125156_recurso_00.pdf)
- MinTic. (2016). Mintic. Obtenido de <https://www.mintic.gov.co/portal/inicio/Sala-de-prensa/Noticias/15180:Colombia-cerro-2015-con-12-4-millones-de-conexiones-a-internet-banda-ancha>

- Puentes, R. (2016). Análisis de la apropiación y uso de las TIC por parte de las pymes colombianas. Recuperado de <https://revistas.usantomas.edu.co/index.php/iusta/article/view/3524>
- Pulido M. (2020). Factores que influyen en la adopción de machine learning en empresas colombianas.
- Ramírez, D. H. (2018). El machine learning a través de los tiempos, y los aportes a la humanidad. Tomado de: <https://repository.unilibre.edu.co/bitstream/handle/10901/17289/EL%20MACHINE%20LEARNING.pdf?sequence=1>

## Anexos

### Anexo A

*Importación de librerías en Python. Elaboración propia*

```
[ ] ###Importación de librerías  
    ##procesamiento  
    import pandas as pd  
    import pandas as pd  
    import datetime  
    import numpy as np  
    import datetime as dt  
    import missingno as msno  
    import time  
    import itertools  
    import warnings
```

## Anexo B

*Visualización del modelo en Python. Elaboración propia*

```
##Visualización
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D
from statsmodels.graphics.factorplots import interaction_plot
import seaborn as sns
import plotly.graph_objects as go
```

## Anexo C

*Predicción del modelo en Python. Elaboración propia*

```
##Predicción
from scipy import stats
from sklearn import metrics
from math import sqrt
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error as mse

import statsmodels.api as sm
import statsmodels.formula.api as smf

import matplotlib.pyplot as plt
import numpy as np
from sklearn import datasets, linear_model, metrics
```

## Anexo D

*Importación de los datos en Python. Elaboración propia*

```
✓ [21] datos = pd.read_excel("/df_empresas")
0s datos.head()
```

	Index	Sector	Ingresos	Total de personal ocupado	Total de empresas	Uso de Internet	Distribuir productos en línea	Recibir pedidos por Internet	Hacer pedidos por Internet	Capacitación de Personal	Envío electrónico
0	2008	Industria	182000000000	652576	7453	7122	775.0	3586.0	3028.0	7453.0	71
1	2008	Comercio	53900000000	314714	3774	3774	325.0	1244.0	1244.0	3774.0	37
2	2008	H Hoteles y restaurantes	30000000000	70892	501	493	92.0	170.0	166.0	0.0	4
3	2008	I Actividades complementarias de transporte y ...	30000000000	98266	492	492	111.0	164.0	177.0	0.0	4
4	2008	K Inmobiliarias, empresariales y de alquiler	30000000000	357178	1939	1931	257.0	498.0	538.0	0.0	18

## Anexo E

*Verificación del tipo de datos en Python. Elaboración propia*

```
✓ [25] datos = datos.dropna()
0s      datos.dtypes
```

Index	int64
Sector	object
Ingresos	int64
Total de personal ocupado	int64
Total de empresas	int64
Uso de Internet	int64
Distribuir productos en línea	float64
Recibir pedidos por Internet	float64
Hacer pedidos por Internet	float64
Capacitación de Personal	float64
Enviar o recibir correo electrónico	float64
Busqueda de Información Sobre bienes y servicios	float64
Llamadas telefónicas por Internet VoIP	float64
Servicio al Cliente	float64
Uso de aplicaciones	float64
dtype:	object

## Anexo F

*Definir variables en Python. Elaboración propia*

```
[ ] regressor = LinearRegression()  
    regressor.fit(X_train, y_train)
```

```
▾ LinearRegression  
LinearRegression()
```

```
[ ] y_pred = regressor.predict(X_test)
```

```
[ ] mse = mean_squared_error(y_test, y_pred)  
    r_squared = r2_score(y_test, y_pred)  
  
    print('Mean Squared Error:', mse)  
    print('R-squared:', r_squared)
```

```
Mean Squared Error: 4009751722.1463304  
R-squared: 0.8890540580178936
```