



UNIVERSIDAD PERUANA DE CIENCIAS APLICADAS

FACULTAD DE NEGOCIOS

PROGRAMA ACADÉMICO DE ADMINISTRACIÓN DE EMPRESAS

Aplicación de ciencia datos para pronosticar la compra de certificados de
revisión técnica de vehículos de la empresa FORA

TRABAJO DE INVESTIGACIÓN

Para optar el grado de bachiller en Administración de Empresas

AUTOR(ES)

Chavez Cajo, Eva Sofia ([0000-0002-0577-6161](tel:0000-0002-0577-6161))

Sedano Chuquizuta, Melany Rosario ([0000-0002-0447-7709](tel:0000-0002-0447-7709))

ASESOR

Palacios Ruiz, Julio Cesar ([0000-0003-1610-5205](tel:0000-0003-1610-5205))

Lima, 16 de julio de 2021

DEDICATORIA

El presente trabajo de investigación se lo dedicamos a nuestras familias y seres queridos que nos apoyaron incondicionalmente en favor de nuestro crecimiento profesional.

AGRADECIMIENTOS

Queremos expresar nuestro especial agradecimiento a la empresa Fora por brindarnos su apoyo y facilidades en la elaboración de este proyecto.

RESUMEN

El objeto de estudio del presente trabajo es la empresa Fora S. A. C., empresa dedicada a revisiones técnicas, donde al realizar los análisis tanto externo como interno se logró identificar el problema esencial de la compañía. De esta manera, se diagnosticó que existía un inadecuado método en solicitar cierta cantidad de certificados vehiculares al momento de realizar las compras por parte del área Logística. Por esta razón, esta investigación se basa en identificación de las variables que influyen en la compra de estos certificados y en emplear la mejor técnica de ciencia de datos que permita anticipar la compra y disminuir las consecuencias negativas en el inventario y en lo económico de la empresa. Para lo cual, se empleó la metodología de ciencia de datos mediante un análisis el explicativo y exploratorio; asimismo, la creación de la base de datos utilizada en el trabajo se empleó a través de fuentes internas y externas a la empresa, lo que permite responder a las preguntas de Data Science formuladas en el proyecto. También, es importante mencionar que se realizó un análisis estadístico cuantitativo, esto con la finalidad de encontrar las correlaciones entre todas las variables. De las cuales, se lograron hallar notables descubrimientos que sin duda aportarán valor a la organización y permitirá que el área logre los objetivos, debido a que se proponen diversas soluciones y sugerencias que se deberán tomar para emplear de mejor manera la decisión de compra y reducir al mínimo las consecuencias que afectan a la compañía.

Palabras clave: Objetivos, Metodología, *Data Science*.

Data science application to forecast the purchase of technical inspection certificates for vehicles from the FORA company

ABSTRACT

The object of study of this work is the company Fora S. A. C., a company dedicated to technical reviews, where by performing both external and internal analyzes it was possible to identify the essential problem of the company. In this way, it was diagnosed that there was an inadequate method in requesting a certain amount of vehicle certificates at the time of making purchases by the Logistics area. For this reason, this research is based on the identification of the variables that influence the purchase of these certificates and on using the best data science technique that allows anticipating the purchase and reducing the negative consequences on the inventory and on the economy of the company. business. For which, the data science methodology was used through an explanatory and exploratory analysis; Likewise, the creation of the database used in the work was used through internal and external sources to the company, which allows answering the Data Science questions formulated in the project. Also, it is important to mention that a quantitative statistical analysis was carried out in order to find the correlations between all the variables. Of which, notable discoveries were found that will undoubtedly add value to the organization and allow the area to achieve its objectives, because various solutions and suggestions are proposed that should be taken to better use the purchase decision and reduce the consequences that affect the company to a minimum.

Keywords: Objectives, Methodology, Data Science.

TABLA DE CONTENIDO

1.	11
1.1.	11
1.2.	12
1.3.	14
2.	15
2.1.	15
2.2.	17
2.3.	17
2.4.	17
2.5.	18
2.6.	18
3.	18
3.1.	18
4.	18
4.1.	18
4.2.	19
4.3.	19
5.	19
5.1.	19
5.2.	19
5.3.	20
5.4.	20
5.5.	21
6.	21
6.1.	21
6.2.	24
7.	25
7.1.	25
7.2.	25
7.3.	26

7.4.	26
7.5.	29
8.	31
8.1.	31
8.1.1.	31
8.2.	33
9.	33
9.1.	33
9.2.	34
10.	35
10.1.	35
11.	38
11.1.	38
11.2.	38
12.	38
12.1.	38
12.2.	39
12.3.	40
13.	42
13.1.	42
13.2.	47
13.3.	48
14.	48
14.1.	48
14.2.	49
14.3.	49
14.4.	50
14.5.	50
14.6.	50

REFERENCIAS

ÍNDICE DE FIGURAS

<i>Figura 1 Modelo de Negocio</i>	14
<i>Figura 2 Diccionario de la base de datos de Fora</i>	23
<i>Figura 3 Medidas de tendencia central de afluencia vehicular del 2016 al 2020</i>	32
<i>Figura 4 Medidas de tendencia central de certificados usados del 2016 al 2020</i>	33
<i>Figura 5 Medidas de tendencia central de vehículos vendidos en el 2017</i>	33
<i>Figura 6 Arquitectura Funcional de datos</i>	35
<i>Figura 7 Estadística de regresión de las variables</i>	37
<i>Figura 8 Excel con los datos a recolectar</i>	42
<i>Figura 9 Panel del modelo final para Fora</i>	43
<i>Figura 10 Estadísticas de regresión calibrado al 40%</i>	44
<i>Figura 11 Estadísticas de regresión calibrado al 70%</i>	45
<i>Figura 12 Estadísticas de regresión calibrado al 100%</i>	47

ÍNDICE DE GRÁFICOS

<i>Gráfico 1 Cantidad de certificados mensual solicitados por cada sede en el 2020</i>	17
<i>Gráfico 2 Gráfico de pérdidas económicas mensuales en el 2020</i>	17
<i>Gráfico 3 La afluencia de vehículos de los años 2016 al 2020 en la sede Surco</i>	23
<i>Gráfico 4 Certificados Usados de los años 2016 al 2020 en la sede Surco</i>	24
<i>Gráfico 5 Cantidad de vehículos vendidos en el 2017</i>	25
<i>Gráfico 6 Campana de Gauss de la afluencia de vehículos de los años 2016 al 2020 en la sede Surco</i>	27
<i>Gráfico 7 Campana de Gauss de certificados usados de la sede Surco del 2016 al 2020</i>	28
<i>Gráfico 8 Campana de Gauss de los vehículos vendidos en el 2017</i>	29
<i>Gráfico 9 Correlación de las variables</i>	30
<i>Gráfico 10 Regresión Lineal de Cantidad de certificados usados según la Afluencia</i>	36
<i>Gráfico 11 Regresión Lineal de Cantidad de certificados usados según periodo</i>	37
<i>Gráfico 12 Pronóstico de certificados a comprar para el 2021</i>	40
<i>Gráfico 13 Correlación lineal entre las variables Cantidad de certificados usados y Afluencia calibrado al 40%</i>	43
<i>Gráfico 14 Correlación lineal entre las variables Cantidad de certificados usados y el Periodo calibrado al 40%</i>	44
<i>Gráfico 15 Correlación lineal entre las variables Cantidad de certificados usados y Afluencia calibrado al 70%</i>	45
<i>Gráfico 16 Correlación lineal entre las variables Cantidad de certificados usados y el Periodo calibrado al 70%</i>	45
<i>Gráfico 17 Correlación lineal entre las variables Cantidad de certificados usados y Afluencia calibrado al 100%</i>	46
<i>Gráfico 18 Correlación lineal entre las variables Cantidad de certificados usados y el Periodo calibrado al 100%</i>	47

1. Presentación de la empresa para el proyecto

1.1. Presentación y descripción de la empresa

Fora es una empresa peruana que brinda los servicios de inspección técnica de vehículos de mayor demanda en todo el país, el operador de inspección técnica de vehículos ha establecido altos estándares de calidad a nivel nacional bajo la supervisión de entidades oficiales. El funcionario explicó que la autorización otorgada por el Ministerio de Transporte (MTC) les permite brindar servicios que abarcan todo el país (Fora, *s.f.*).

Según Fora (*s.f.*), su misión es contribuir a la seguridad vial en el Perú y a la reducción de los índices de accidentes de tránsito y contaminación en el medio ambiente a través del servicio de revisiones técnicas vehiculares.

Asimismo, su visión se basa en llegar a estar cada vez más cerca de sus clientes, creciendo cada día más, manteniendo el liderazgo en el servicio de Revisiones Técnicas Vehiculares en el Perú siendo así la empresa que genere el cambio en seguridad vial a nivel nacional (Fora, *s.f.*).

Figura 1 Modelo de Negocio



Fuente: Fora

Elaboración: Propia

1.2. PESTEL

- **Político**

Actualmente se vive un ambiente político inestable debido a las próximas elecciones generales (Gestión, 2021) la gran cantidad de candidatos y las propuestas que ofrecen no son tan alentadoras y se espera que la situación cambie en medida que se acerca la fecha de votación.

- **Económico**

El tipo de cambio es el más alto registrado en la historia del país (3.72) y sigue en aumento por la inestabilidad política, la pandemia y factores externos (Gestión, 2021).

Muchos de los productos que compra la Fora son importados y esta subida de precio afecta directamente a la economía de la empresa.

- **Social**

Lima es calificada como una de las peores ciudades para conducir (El Diario Vasco, 2020). Sumado a la cultura peruana de querer siempre evadir las leyes, cerca de un millón de vehículos no cuentan con la revisión técnica vigente (García, A., 2018). Las calles son inseguras al tener tantos vehículos con posibles fallas técnicas por no contar con la ficha de revisión técnica.

- **Tecnológico**

En la actualidad, se ha aprobado el Decreto Supremo N° 005-2021-MTC respecto al Reglamento Nacional para el Fomento del Chatarreo, publicado en el diario oficial El Peruano (Andina, 2021). El propósito de este reglamento es promover la eliminación final o renovación de flotas de automóviles para reducir la emisión de gases de efecto invernadero (GEI) y contaminantes locales que afectan la salud pública, y ayudar a reducir los accidentes en la vía pública y mejorar el sistema de transporte terrestre. Para Fora, esta es una oportunidad porque la empresa tendrá más clientes a largo plazo, no solo por las revisiones técnicas, sino también por el mantenimiento de las nuevas adquisiciones de automóviles.

- **Ecológico**

La contaminación que producen los vehículos en Perú favorece a muchas enfermedades respiratorias y, por ende, a un porcentaje de muertes. Si se compara el nivel de rigurosidad por la que pasan los autos en Perú con los de otros países como Estados Unidos, aún falta mucho por mejorar para tener los niveles que estos exigen. Por otro lado, en otros países como Chile, se limita el uso de vehículos que tienen alto grado de contaminación, una práctica que favorece mucho al ciudadano, pero que al mismo tiempo perjudica al conductor al ser su medio de transporte (Asociación Automotriz Peruana, 2019).

- **Legal**

Debido a la pandemia, una de las disposiciones por el MTC son las de las constantes ampliaciones en las vigencias de las licencias de conducir de los vehículos según la Resolución Directoral N° 003-2021-MTC/18, publicada en el diario oficial El Peruano (Andina, 2021). Esto representa una amenaza para Fora, pues la empresa contaría con menos clientes a largo plazo no solo por las ampliaciones de vigencias sino también por las constantes restricciones que impone el gobierno para usar autos particulares.

1.3. FODA

Fortalezas

- Mayor red de locales a nivel nacional.
- Un grupo de talentos capaces, competitivos y empleables.
- En el negocio de la inspección técnica de vehículos, tiene más experiencia que casi todos los competidores.
- Tiene un sistema ERP que puede administrar mejor los recursos y controlar mejor los procesos financieros y contables.
- Empresa autorizada por el MTC.

Oportunidades

- Incremento del sector automotor.
- Acuerdo de cooperación mutua o acuerdo comercial.
- Participar en exposiciones y eventos institucionales que promuevan el desarrollo de la marca.
- La posibilidad de realizar la automatización de procesos de almacén con proveedores estratégicos de ERP.

- Desarrollar otros servicios complementarios, como talleres mecánicos y revisiones técnicas de motocicletas.
- Incremento de clientes por disposiciones políticas favorables.

Debilidades

- Infraestructura para mejorar.
- La tasa de rotación del personal es alta.
- Algunos procesos de la empresa carecen de software de gestión.
- Personal no involucrado con actualizaciones de la normativa.
- Supervisión inadecuada del sistema de revisión de tecnología de vehículos.

Amenazas

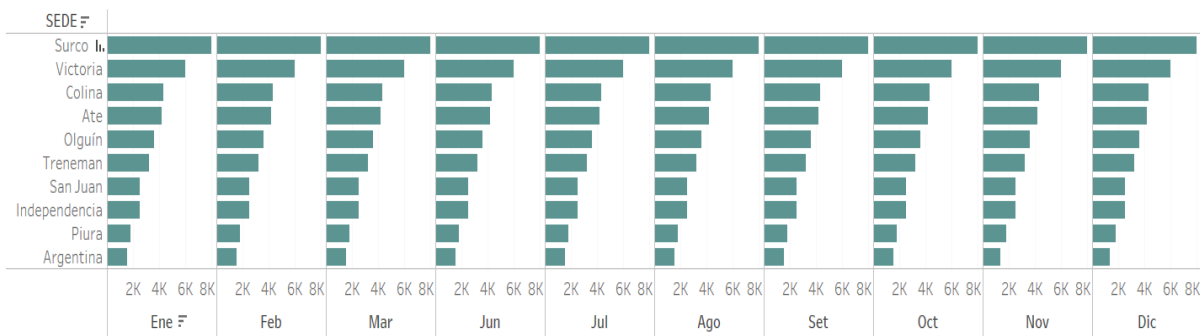
- Nuevos lugares para revisiones técnicas de vehículos de empresas competidoras.
- Campaña de difamación en los medios.
- Los posibles cambios en la normativa afectarán a la continuidad del centro de inspección técnica de vehículos.
- Disminución de clientes por restricciones políticas debido a la pandemia.
- Políticas de restricciones vehiculares por el Covid-19.

2. Comenzando el proyecto

2.1. Identificación del problema - Comprendiendo el negocio y el problema

Fora es una empresa autorizada por el MTC para brindar servicios de revisiones técnicas vehiculares. Cuenta con 10 sedes distribuidas por todo Lima. Mensualmente se hacen compras de una cantidad fija de los certificados que se usarán en cada sede.

Gráfico 1 Cantidad de certificados mensual solicitados por cada sede en el 2020



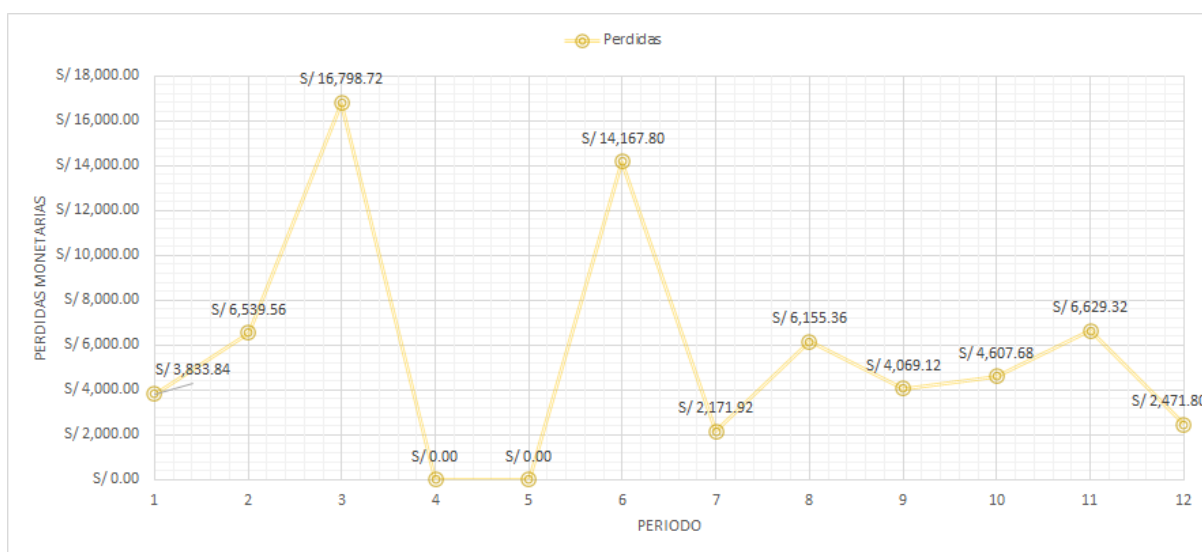
Fuente: Fora

Elaboración: Propia

En el gráfico se puede ver los pedidos de compra de los certificados y resalta que, sin importar el mes, siempre se pide la misma cantidad por sede.

Los certificados se cambian cada 3 meses según la Resolución Directoral 35-2020-MTC y los que no se usan al finalizar el trimestre se desechan. Este sobrestock ocasiona pérdidas económicas para la empresa.

Gráfico 2 Gráfico de pérdidas económicas mensuales en el 2020



Fuente: Fora

Elaboración: Propia

En el gráfico se muestra las pérdidas económicas que ha tenido la empresa en el 2020, si bien es cierto en los últimos meses hay una disminución, en los meses de marzo y junio se nota un aumento considerable por efectos de la pandemia.

2.2. Formulación de preguntas para Data Science

Las preguntas de investigación están enfocadas a predecir cuántos certificados se deben pedir en las órdenes de compra de la sede de Surco. En tal contexto se plantea la siguiente pregunta: **¿Cuánto sería la cantidad mensual necesaria de certificados para abastecer a la sede Surco?**

- ¿Cuál podría ser la afluencia vehicular en la sede Surco?
- ¿Cuántos certificados podrían usarse en la sede Surco?
- ¿Cuántos vehículos podrían ser potenciales clientes según los vehículos vendidos en el 2017?

Es por ello que, el enfoque analítico que se tomará en el proyecto será predictivo.

2.3. Justificación

Con los constantes cambios por parte del Ejecutivo, como el aforo permitido en las distintas regiones y por tipo de negocio. Sumado a la inestabilidad política y sanitaria, For a debe tomar en consideración las pérdidas monetarias que ocasiona no prestar atención al stock de los certificados que mantiene en cada sede. En esta ocasión nos centraremos en la sede Surco porque es la que presenta mayor afluencia.

2.4. Objetivos del proyecto

Objetivo general

Pronosticar la cantidad mensual necesaria de certificados para la sede Surco para el 2021.

Objetivos específicos

Determinar la afluencia de vehículos en la sede Surco.

Determinar la cantidad de certificados que podrían usarse.

Determinar la cantidad de vehículos que podrían ser potenciales clientes a través de vehículos vendidos en el 2017.

2.5. Posibles soluciones al problema, basadas en ciencia de datos

Ante la problemática de la pérdida económica por el sobrestock de certificados que se desechan (en 2020 se desecharon 89,483 certificados ocasionando pérdidas de más de S/60,000) debido a la Resolución Directoral 35-2020-MTC, las posibles soluciones que ayudarían a minimizar la pérdida sería determinar la cantidad necesaria de certificados a comprar a través de la identificación de la afluencia vehicular y los certificados usados para considerar un stock mínimo.

2.6. Hipótesis basadas en el modelo de ciencia de datos

A través de la regresión se predecirán los certificados que se usarán en la sede Surco en el 2021.

3. Marco teórico

3.1. Metodología de Ciencia de Datos

La metodología para utilizar estará basada en las etapas de ciencia de datos con el fin de pronosticar y obtener las respuestas a las preguntas de la investigación a través de un modelo predictivo.

4. Requisitos y Recopilación de datos

4.1. Determinación de datos necesarios para responder a las preguntas

Se requerirán datos como:

- Registros de ingreso de los vehículos
- Listado de certificados usados
- Cantidad de certificados en inventario
- Cantidad de vehículos vendidos en el 2017

4.2. Ubicación de fuentes de datos

Los datos se obtienen de la base de datos de Fora, llamada “Reportes de Inspecciones” desde el año 2016 al 2020.

4.3. Presentación de datos recopilados

La recopilación de datos se obtiene a través de la misma base de datos propiedad de la empresa. Esta base de datos es original, y se utilizará para responder preguntas de investigación y lograr metas generales y metas específicas.

5. Comprensión y preparación de datos

5.1. Uso de herramientas para determinar posibles relaciones entre los datos.

Determinando el punto de partida.

Para determinar las relaciones entre los datos obtenidos se analizaron inicialmente desde Microsoft Excel, debido a que la base de datos se encuentra desde esta plataforma y nos facilita en análisis estadístico. Para un mejor análisis de la base, Power BI y Tableau nos permitió realizar gráficos con una mejor visualización.

5.2. Revisión y depuración de datos que se usarán en el estudio

En el presente estudio se utilizarán los datos desde el 2016 al 2020 para calcular la afluencia vehicular de la sede Surco, los certificados usados durante dichos años. Para obtener la afluencia vehicular se depura de la siguiente manera:

Se descarga el reporte de inspecciones del sistema de Fora en Excel y se eliminan las columnas innecesarias y nos quedamos con las columnas “Fecha”, “Año”, “Día”, “Mes”, “Estado de Inspección” y “Resultado”. Se realiza un filtrado en la columna “Planta” y se deja “Surco”. Luego se crea una tabla dinámica, con filtros de “Año” y “Resultado”, que indique cuántos vehículos han aprobado y desaprobado en todos los meses del año. De esta forma, se obtiene el dato de cuántos vehículos han aprobado o desaprobado su inspección técnica vehicular del 2016 al 2020.

Para hallar la cantidad de certificados anulados mensuales del 2016 al 2020, se debe restar mes a mes a la afluencia vehicular la cantidad de vehículos aprobados y desaprobados. El resultado es la cantidad de certificados anulados por mes.

Los certificados usados se calculan de la suma de los certificados anulados más los vehículos aprobados.

5.3. Calidad de datos

Para desarrollar el presente trabajo de investigación, la información fue extraída de las bases de datos de Fora. En donde para garantizar la calidad de los mismos se empleó la metodología del consolidado en donde se unieron los diferentes reportes depurados y el monitoreo donde se verificó que los datos fueran relevantes para el proyecto.

5.4. Gobierno de datos

Toda empresa debe tener los datos representados y protegidos por un área única que asegure la confiabilidad, calidad y mantener la data actualizada para los distintos requerimientos. El área de sistemas es la que se encuentra a cargo de la custodia de los datos y es responsable de ponerla a disponibilidad de las otras áreas para su respectivo análisis.

5.5. Diccionario de datos

El presente diccionario ayudará a la empresa a entender con qué parámetros trabajar, de tal manera que tendrá un mejor control de la información.

Figura 2 Diccionario de la base de datos de Fora

	Tipo	Longitud	Índice	Único	Descripción	Regla de negocio
Planta	Cadena	5	No	No	Nombre de la sede	Surco
Fecha	Cadena	10	No	No	Ninguno	Son válidas a partir del 01/01/2016 al 31/12/2020
Día	Entero	2	No	No	Día del mes	Del 1 al 31
Mes	Cadena	10	No	No	Ninguno	Ninguno
Año	Entero	4	No	No	Ninguno	2016, 2017, 2018, 2019 y 2020
Estado de Inspección	Cadena	11	No	No	Vehículos que hayan completado el proceso de inspección técnica vehicular	Consolidado
Resultado	Cadena	1	No	No	Vehículos aprobados y desaprobados durante su inspección técnica vehicular	A y D

Fuente: Fora

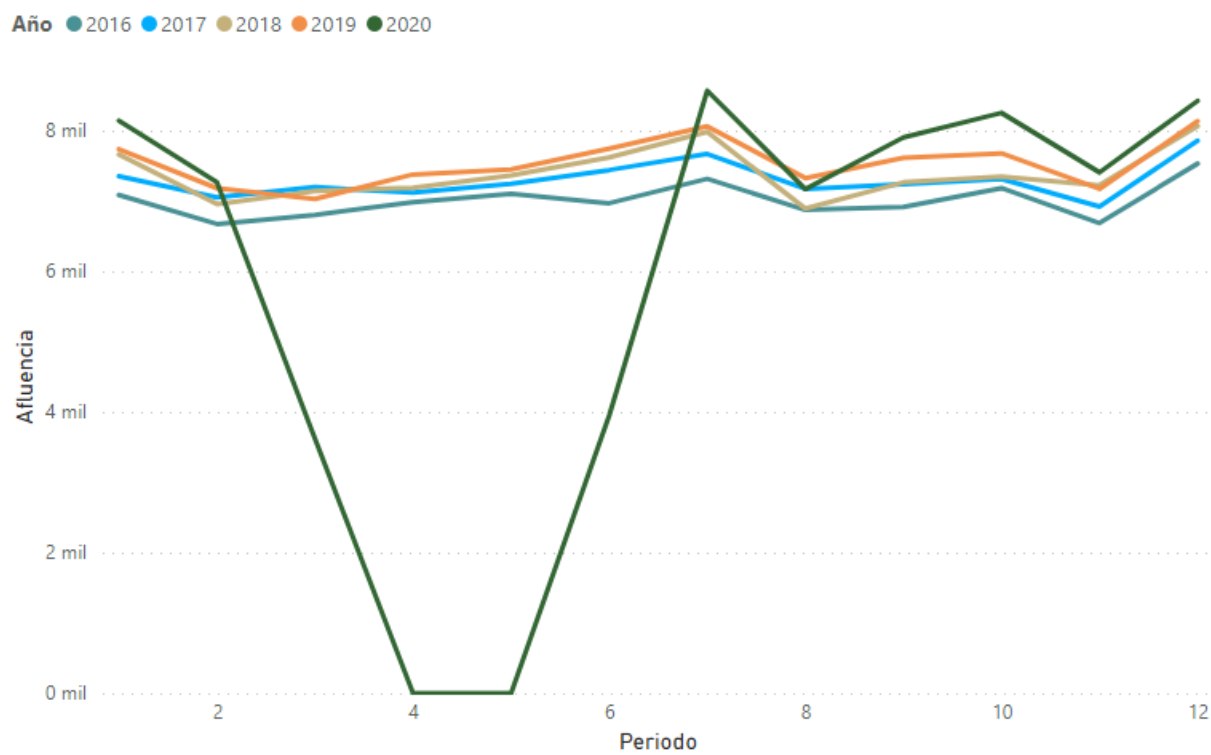
Elaboración: Propia

6. Visualización

6.1. Presentación de datos a utilizar para el estudio

En esta parte se emplearán las herramientas como el Power BI, Excel y Tableau para responder las preguntas de ciencia de datos que facilitarán la toma de decisiones por el área correspondiente de las compras de los certificados de la empresa Fora.

Gráfico 3 La afluencia de vehículos de los años 2016 al 2020 en la sede Surco



Fuente: Fora

Elaboración: Propia

En el gráfico se puede comparar la afluencia mensual de los vehículos en los últimos 5 años en la sede Surco, en donde se ve que el 2020 tiene picos atípicos por la suspensión perfecta en los meses de abril y mayo, teniendo una recuperación en los siguientes meses.

Gráfico 4 Certificados Usados de los años 2016 al 2020 en la sede Surco

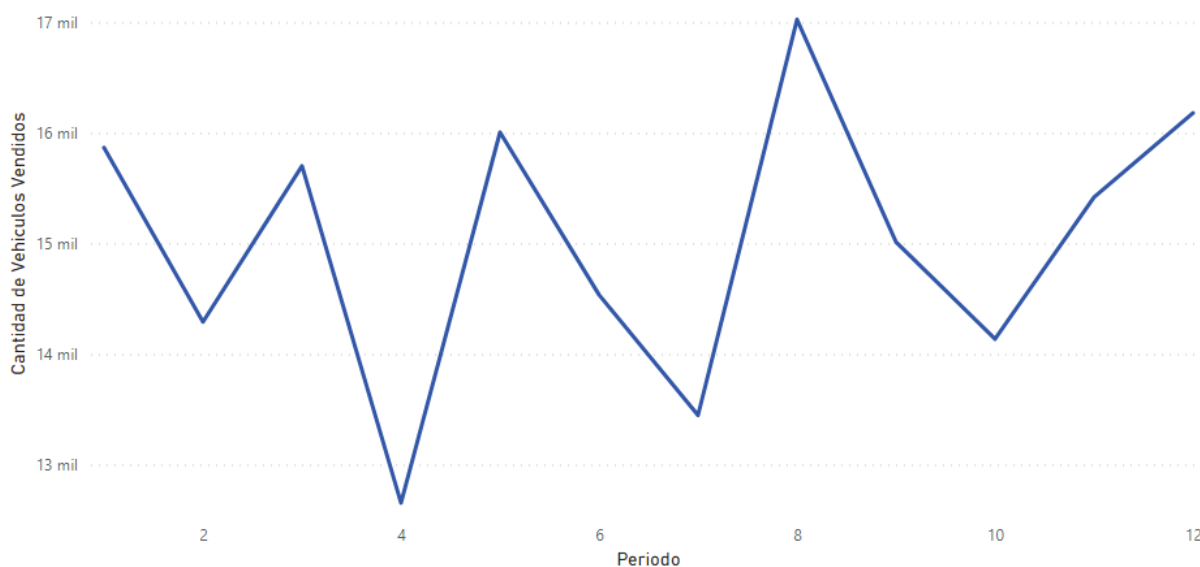


Fuente: Fora

Elaboración: Propia

En este gráfico se puede visualizar el comportamiento de los certificados aprobados, de igual manera que en el gráfico anterior se ve el golpe de la pandemia; asimismo, aparentemente no tiene un patrón por lo que se concluye que no son estacionales.

Gráfico 5 Cantidad de vehículos vendidos en el 2017



Fuente: Fora

Elaboración: Propia

En el gráfico se puede visualizar la cantidad de vehículos que mensualmente podrán representar nuevos clientes para la empresa, puesto que tiene que pasar revisión técnica obligatoria al menos una vez al año por lo que cumplieron más de 3 años de fabricación. Este número es un apoyo para que el pronóstico sea más certero.

6.2. Determinación de las herramientas de visualización a utilizar para el proyecto

Para este proyecto se ha utilizado dos herramientas las cuales son Excel, Tableau y Power BI, ya que por su versatilidad y su funcionalidad amistosa permite crear diversas visualizaciones a través de la transformación de datos.

7. Producción y Análisis exploratorio de datos

7.1. Plan de muestreo y tamaño de la muestra

Para el plan de muestreo, se optó por uno sencillo, es decir, se escogió un 60% de la sede más representativa, Surco. Por ser una base con un gran número de datos, la proporción seleccionada permitirá trabajar con un nivel de confianza del 95%. Por ejemplo, en el 2020 la muestra tiene un valor de 55,927 de vehículos ingresados, 243,704 certificados usados y 108,169 de vehículos vendidos en el 2017. Con estos datos más los de los años 2016 al 2019 se espera obtener un modelo que ayude a pronosticar la cantidad óptima de certificados a adquirir, y por ende la cantidad de certificados que se deben comprar para esta sede.

7.2. Diseño del estudio de muestreo

El muestreo que se escogió es el de tipo probabilístico, dentro de este modelo se optó por el muestreo aleatorio simple. La población que se toma en cuenta es la cantidad de vehículos de los clientes que asistieron a Fora a realizar una revisión técnica en la sede de Surco durante el 2016 al 2020. La muestra se determinará a través de un juicio experto, el porcentaje a considerar será un 60% de la sede con mayor afluencia, es decir, Surco. Debido al tamaño de la población, el porcentaje es más que suficiente para determinar el modelo. Dentro de esta muestra se analizará la afluencia, los certificados usados, vehículos vendidos en el 2017 y el periodo, siendo estas variables cuantitativas de escala discreta. El estudio del muestreo sería de esta manera:

$f(x) = [\text{Cantidad de certificados usados, los certificados usados, cantidad de vehículos vendidos en el 2017, Periodo desde el 2016}]$

7.3. Ejecución del muestreo

Al emplear el método aleatorio simple, la fórmula sería la siguiente:

$$\text{Tamaño de Muestra (n)} = \frac{\frac{z^2 \times p(1-p)}{e^2}}{1 + \left(\frac{z^2 \times p(1-p)}{e^2 N} \right)}$$

Donde:

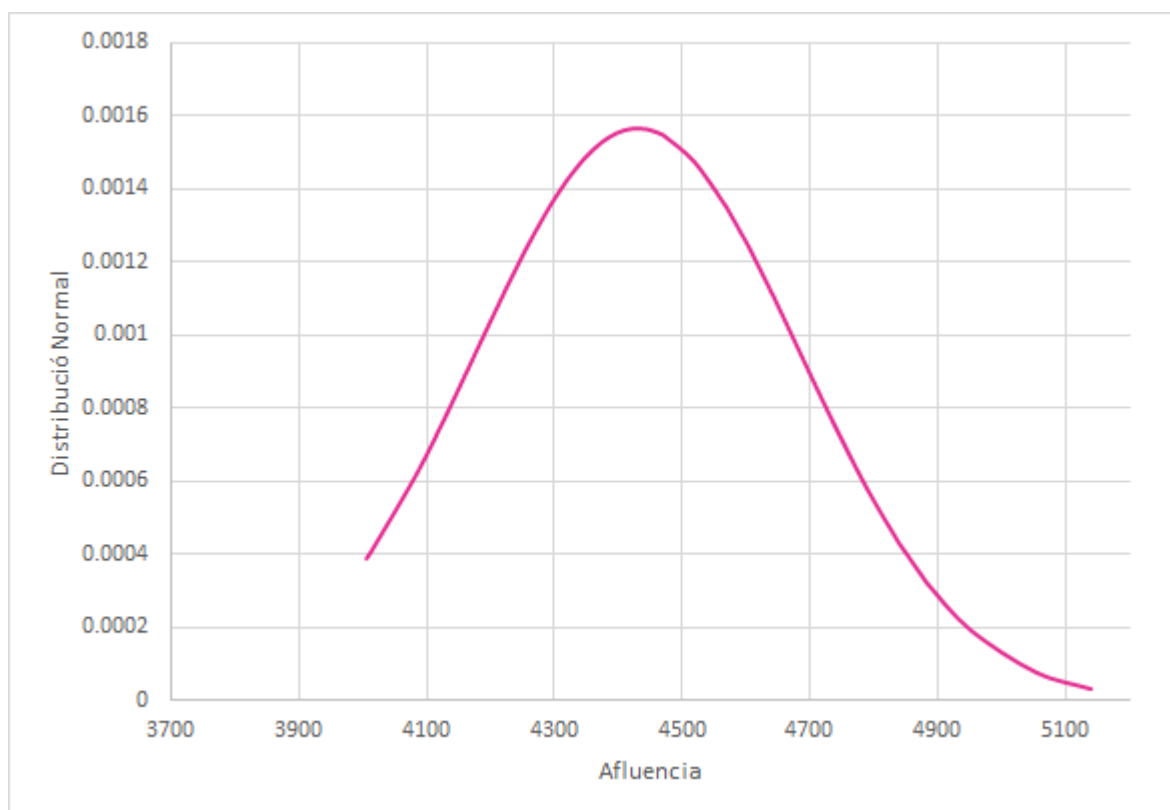
- N = tamaño de la población
- n = tamaño de la muestra
- p = probabilidad de éxito
- q = probabilidad de fracaso
- Z = nivel de confianza deseado
- e = margen de error

Para este proyecto se trabajará con el 60% del tamaño de muestra que nos permitirá trabajar con un nivel de confianza del 95%.

7.4. Resumen y examen de la distribución de variables

En este punto se emplea el análisis estadístico de cada variable.

Gráfico 6 Campana de Gauss de la afluencia de vehículos de los años 2016 al 2020 en la sede Surco

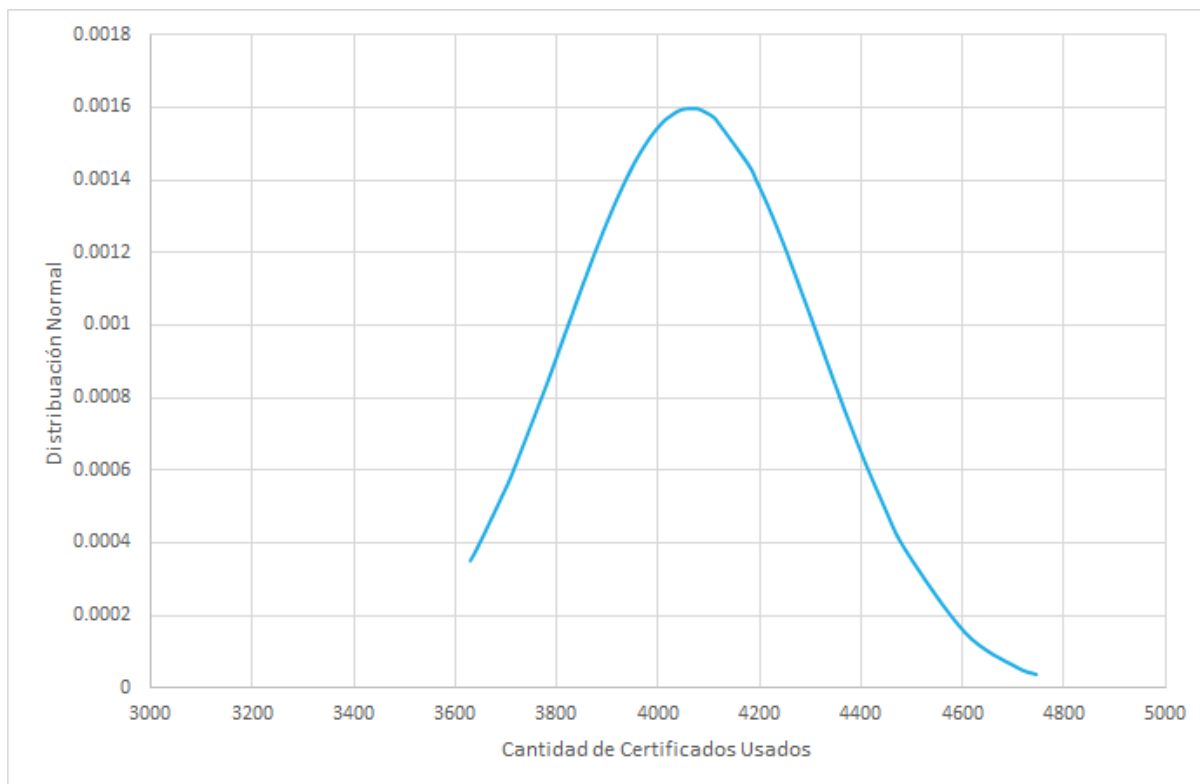


Fuente: Fora

Elaboración: Propia

Según el gráfico 6, observamos el comportamiento mensual de la afluencia de los años 2016 al 2020 donde la cantidad de vehículos que asistieron se concentra entre 4,175 y 4,685, donde la desviación estándar de 255.26 ayuda a comprender que tan alejados están los valores de la media, que tiene un valor de 4,429.87.

Gráfico 7 Campana de Gauss de certificados usados de la sede Surco del 2016 al 2020

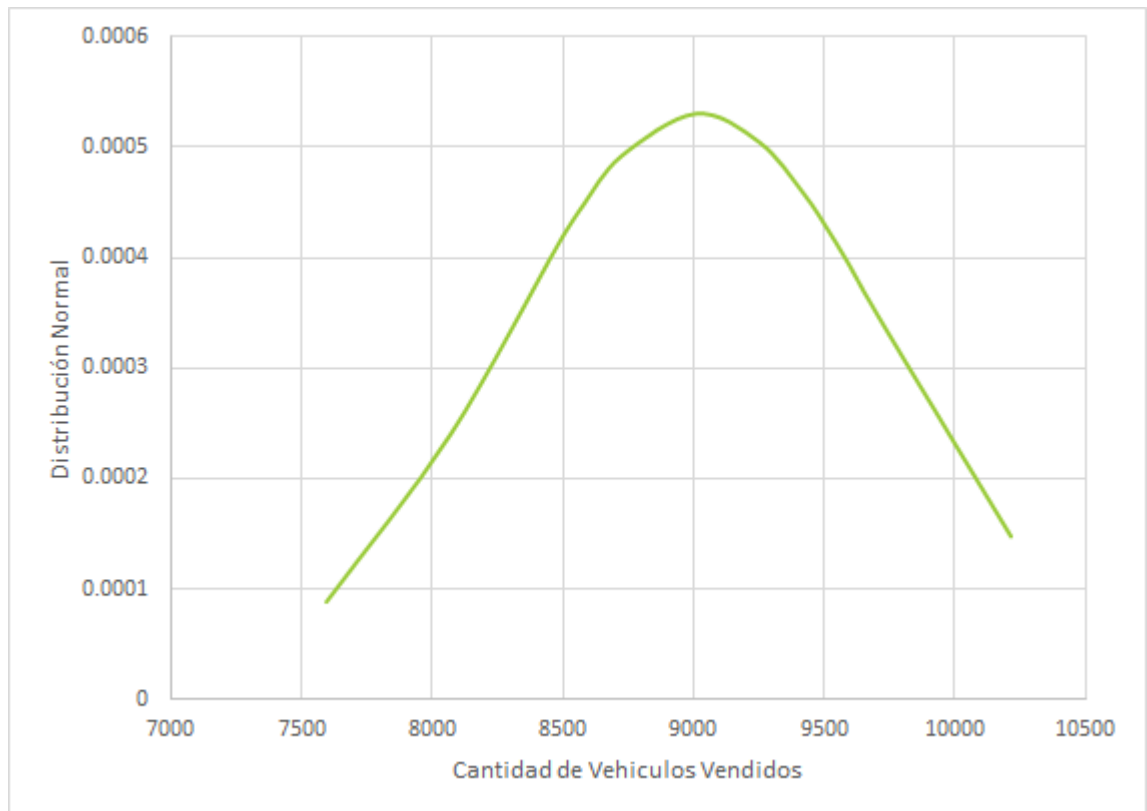


Fuente: Fora

Elaboración: Propia

Según el gráfico 7, durante los meses de los años 2016 al 2020, la mayor concentración de la cantidad de certificados usados está entre los valores de 3812 y 4310, donde la desviación estándar, que tiene un valor de 249.77, ayuda a conocer qué tan alejados están los valores del promedio (4061).

Gráfico 8 Campana de Gauss de los vehículos vendidos en el 2017



Fuente: Fora

Elaboración: Propia

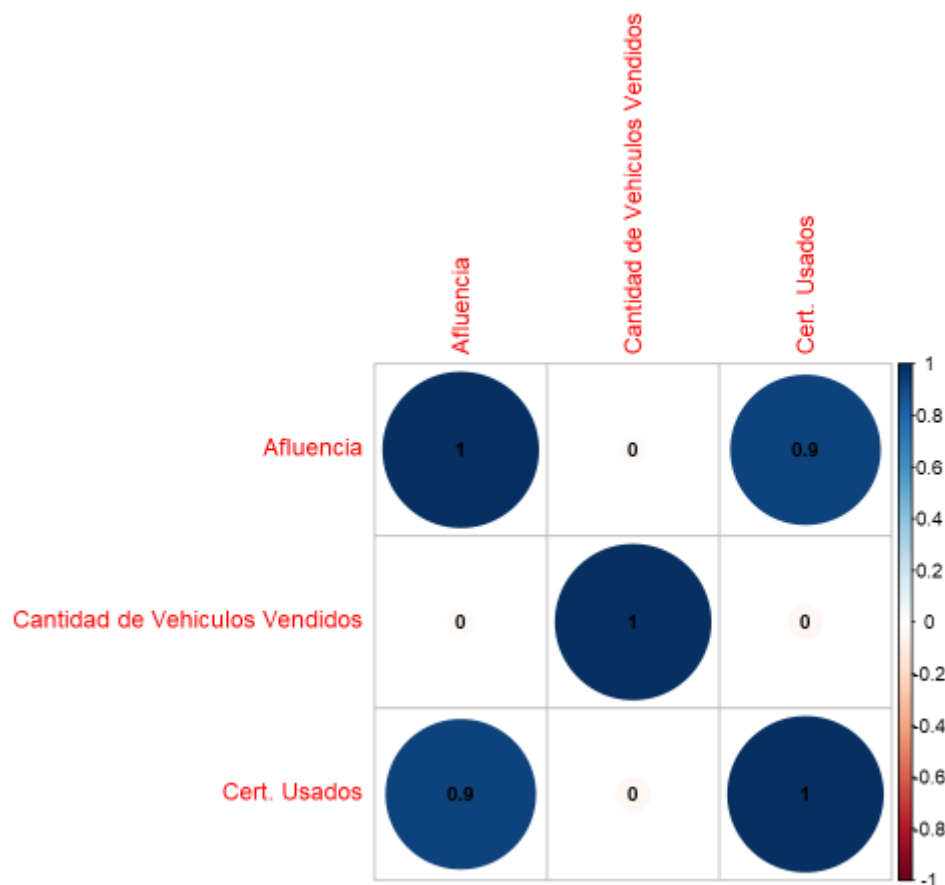
En este gráfico, el comportamiento de la muestra de los vehículos vendidos mensualmente en el 2017 indica que la concentración fluctúa alrededor de 9000 autos. Estas ventas elevadas son en meses como enero, mayo, septiembre y diciembre, los cuales están cerca de fechas importantes como navidad y año nuevo.

7.5. Resultado del análisis de relaciones

De acuerdo con la ejecución del muestreo, en el gráfico 9 se puede visualizar que existe relación entre las variables que se estaban considerando inicialmente, esto permitirá continuar con la investigación con los certificados usados, por tener la misma relación que la afluencia. Asimismo, se nota claramente que la cantidad de vehículos vendidos no tiene

relación con ninguna de las otras, esto se puede deber a que la data es solo de un año (2017), por eso puede parecer que no se debería contar; sin embargo, por criterio esta variable si afecta a los certificados a comprar, pues representan futuros clientes y por esta razón se añadirá un porcentaje al pronóstico para que ayude con la precisión.

Gráfico 9 Correlación de las variables



Fuente: Fora

Elaboración: Propia

8. Comenzando la interpretación de datos

8.1. Estudio probabilístico de los datos obtenidos

8.1.1. Análisis estadístico

Figura 3 Medidas de tendencia central de afluencia vehicular del 2016 al 2020

<i>Afluencia</i>	
Media	4429.87
Error típico	32.95
Mediana	4367.63
Moda	#N/D
Desviación estándar	255.26
Varianza de la muestra	65159.15
Curtosis	0.34
Coefficiente de asimetría	0.86
Rango	1137
Mínimo	4003.80
Máximo	5140.80
Suma	265792.05
Cuenta	60
Nivel de confianza(95.0%)	65.94

Fuente: Fora

Elaboración: Propia

Según la figura 3, las medidas de tendencia central de la afluencia muestran que el promedio mensual de los años 2016 al 2020 es de 4,429.87. El valor mínimo en la muestra de la afluencia es de 4003 y el máximo es de 5140.

Figura 4 Medidas de tendencia central de certificados usados del 2016 al 2020

<i>Cert. Usados</i>	
Media	4061.73
Error típico	32.25
Mediana	4011.60
Moda	#N/D
Desviación estándar	249.77
Varianza de la muestra	62386.39
Curtosis	0.33
Coefficiente de asimetría	0.74
Rango	1114.8
Mínimo	3631
Máximo	4746
Suma	243704
Cuenta	60

Fuente: Fora

Elaboración: Propia

Durante los meses de los años 2016 al 2020, el valor mínimo de certificados usados fue 3,631 y el valor máximo fue 4,746. Además, el promedio de certificados usados es de 4,061 y la desviación estándar de 249.77.

Figura 5 Medidas de tendencia central de vehículos vendidos en el 2017

<i>Cantidad de Vehiculos Vendidos</i>	
Media	9014.05
Error típico	217.21
Mediana	9129.90
Moda	#N/D
Desviación estándar	752.43
Varianza de la muestra	566150.43
Curtosis	-0.33
Coefficiente de asimetría	-0.38
Rango	2623.20
Mínimo	7592.40
Máximo	10215.60
Suma	108168.60
Cuenta	12
Nivel de confianza(95.0%)	478.07

Fuente: Fora

Elaboración: Propia

Las medidas de tendencia central para la muestra de vehículos comprados indican por ejemplo que el promedio es de 9014.05 y la desviación estándar es de 752.43.

8.2. Distribución de la muestra: Interpretación del comportamiento

En los datos estadísticos, la distribución muestral es el resultado que se obtiene al considerar todas las muestras posibles que se pueden obtener de la población, con esto en mente, se ha verificado que la media es de 4429.87 en el caso de la afluencia vehicular de la sede de Surco. En ese sentido, si le agregamos un rango de 255.26 a ambos lados, es decir, si le sumamos y restamos la desviación estándar, podemos decir que el pronóstico para los siguientes meses puede estar en ese rango.

Del mismo modo, para los certificados usados, donde la media es 4061.73 y la desviación estándar es de 249.77, podemos decir que en los siguientes meses los certificados usados estarán entre los valores de 3,812 y 4,310.

9. Arquitectura de Datos

9.1. Insumos, procesamiento, almacenamiento y destino

- **Insumos**

Como insumos se utilizarán fuentes propias de la empresa a través de diversos reportes generados por las diferentes áreas. Asimismo, sería prudente utilizar fuentes externas para que el proyecto sea más exacto y considerando todos los aspectos que ayuden a resolver el problema.

- **Procesamiento**

Los datos son extraídos de un software que emplea la empresa para poder transformarlos en Microsoft Excel para su posterior análisis.

- **Almacenamiento**

Luego de transformar la data, estos se almacenarán en la Nube como el Drive, ya que es fácil de usar y no requiere gran cantidad de espacio.

- **Destino**

Una vez realizado el almacenamiento, se procede a realizar una exploración informativa mediante modelos de gráficos y análisis plasmados a través de los programas Power BI o Tableau, de ellos se podrán obtener información relevante para la creación de modelos predictivos. Además, se podrá obtener una información procesada y de conocimiento óptimo para la toma de decisiones gerenciales.

Figura 6 Arquitectura Funcional de datos



Fuente: Fora

Elaboración: Propia

9.2. Arquitectura tecnológica: Tecnología, Herramientas e Infraestructura

- **Tecnología**

Con la ayuda de los avances tecnológicos, cada vez se pueden emplear mejores métodos para analizar datos como un científico de datos. Estar a la vanguardia es necesario, pero también se debe utilizar tecnología que se adecue a las necesidades de la empresa.

- **Herramientas**

Las herramientas que se emplean para la arquitectura son las que ayudan a analizar y visualizar mejor los datos y los resultados, como Excel y Power BI.

- **Infraestructura**

En este punto se aprovecha la infraestructura tecnológica ofrecida que sería el hardware y software para el soporte en primera instancia de los procesos principales de la organización.

10. Técnicas de ciencia de datos

10.1. Supervisado

La técnica que se aplica en este trabajo es la de supervisado, ya que se le enseña al modelo mediante la alimentación de datos para predecir datos futuros, y en este caso se pretende predecir la cantidad óptima de certificados a adquirir de la empresa Fora. Para lo cual ya se estableció que la variable a usar será la cantidad de certificados usados como dependiente y la afluencia y el periodo como independientes, esta última variable es porque al no tener estacionalidad es cambiante en el tiempo, por lo que optariamos por adicionar una regresión lineal de serie de tiempo.

La elección de las variables sería:

- Variable dependiente (Y): Cantidad de certificados usados
- Variable independiente (X1): Afluencia
- Variable independiente (X2): Periodo

A partir de las estadísticas anteriores se llega a la siguiente ecuación de regresión lineal múltiple:

$$y = 0.95(X1) - 0.035(X2) - 156.49$$

Sin embargo, esta ecuación no contempla los vehículos vendidos en el 2017, por lo que se trabajará con el 5% de la cantidad total (180,281 unidades) según el periodo, que se agregarían al pronóstico, siendo V_m =cantidad mensual de vehículos vendidos en el 2017, la ecuación quedaría de la siguiente manera:

$$y = 0.95(X1) - 0.035(X2) - 156.49 + 5\%V_m$$

A continuación, se detallan las estadísticas de regresión de las variables mencionadas.

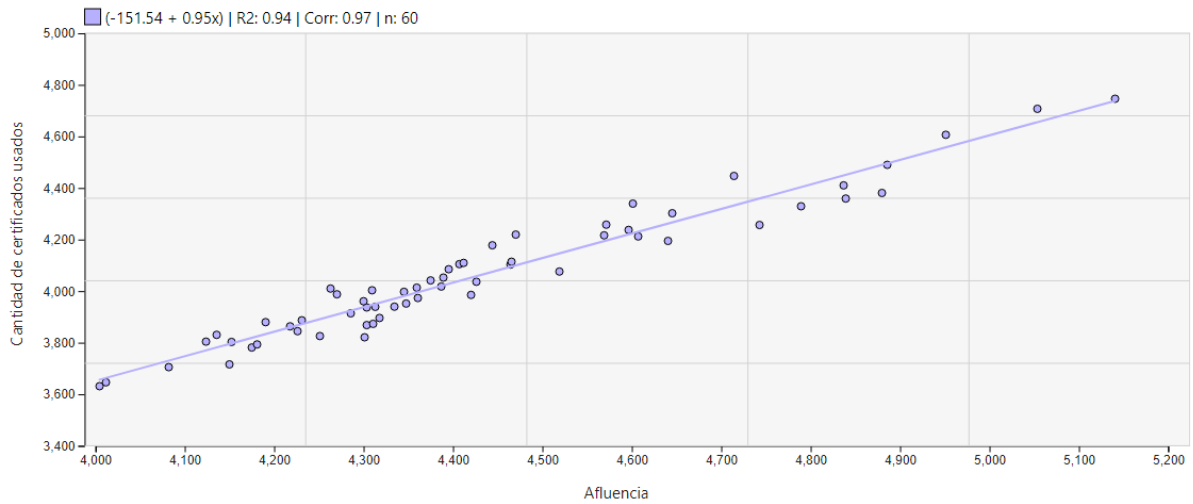
Figura 7 Estadística de regresión de las variables

<i>Estadísticas de la regresión</i>	
Coefficiente de correlación múltiple	0.97201267
Coefficiente de determinación R^2	0.94480864
R^2 ajustado	0.9428721
Error típico	59.6992748
Observaciones	60

Fuente: Fora

Elaboración: Propia

Gráfico 10 Regresión Lineal de Cantidad de certificados usados según la Afluencia

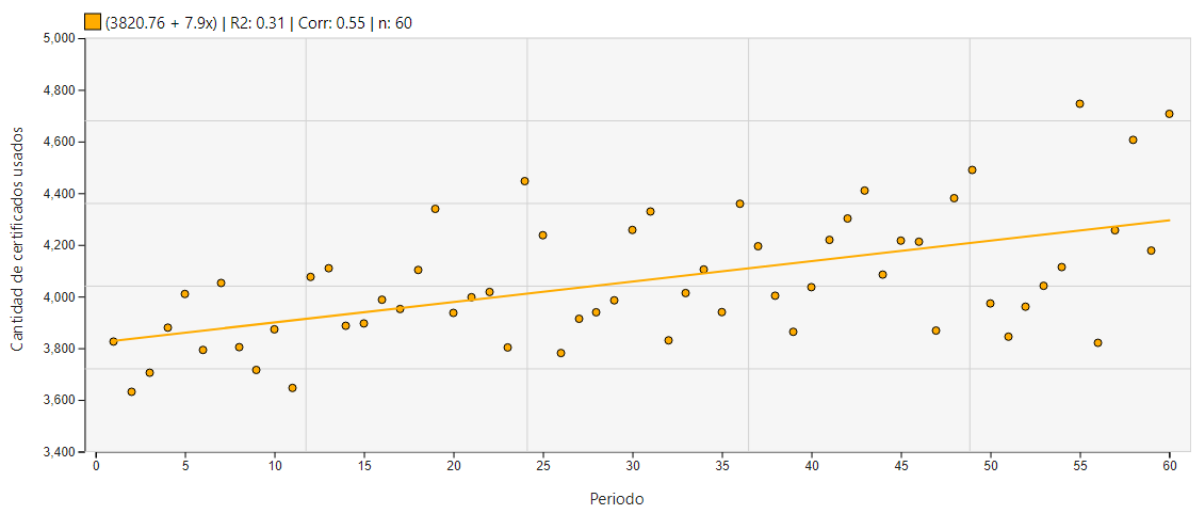


Fuente: Fora

Elaboración: Propia

En esta regresión, el coeficiente de correlación de las variables afluencia y certificados usados es de 0.97, esto indica que están muy relacionadas directa y positivamente una de otra. Entre más alta sea la afluencia, la cantidad de certificados usados también aumenta.

Gráfico 11 Regresión Lineal de Cantidad de certificados usados según periodo



Fuente: Fora

Elaboración: Propia

En el gráfico 10, la cantidad de certificados está un poco dispersa según el periodo. Su coeficiente de correlación es de 0.55, ello indica que no depende mucho del periodo el uso de los certificados.

11. El modelo y la historia

11.1. Determinación del modelo más apropiado para la descripción o predicción.

El modelo más apropiado es el predictivo, debido a que se logró predecir la cantidad de certificados que debe adquirir la empresa Fora.

11.2. Descripción de los hallazgos, ¿Que se descubrió?, comience a contar una historia y planifique la visualización.

Se descubrió que hay una mayor correlación entre la afluencia y los certificados usados que los vehículos vendidos en el 2017 con las otras dos variables, por lo que se incluirá solo un porcentaje de esta última variable para el pronóstico. Además, para suavizar el efecto de la pandemia, se optó por promediar los meses correspondientes a los años anteriores. Con este cálculo se puede llegar a un resultado más preciso.

12. Visualización, Implementación y ajustes

12.1. Definición y justificación de los criterios a tomar para la visualización de los resultados

En base en el análisis se distingue que el enfoque analítico predictivo dirige a la identificación de las más notables variables que impactaría en la cantidad de certificados a

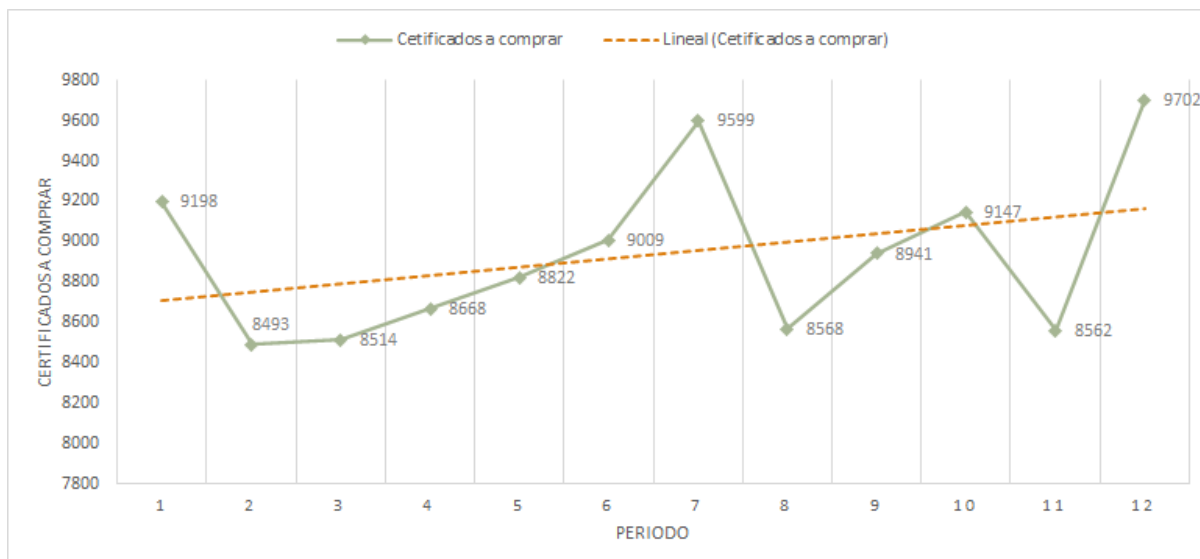
comprar en la empresa FORA. Estas serían: Registros de ingreso de los vehículos, listado de certificados usados, el periodo y los vehículos vendidos en el 2017, este último porque representan posibles clientes.

Para la delimitación y demostración de la visualización de datos en base a los resultados se prefirió usar herramientas como Tableau y Power Bi que son plataformas de análisis visual que transforman la manera en que usamos los datos para resolver problemas. Permite a las personas y organización sacar el máximo partido de los datos.

12.2. Visualización de los hallazgos, grafique lo descubierto y compártanlo

Ejecutando la fórmula para los siguientes periodos del 2021, se obtienen los siguientes resultados, donde los certificados a comprar es la fórmula que nos da la regresión de la afluencia con los certificados usados considerando el periodo más un porcentaje de vehículos vendidos en el 2017.

Gráfico 12 Pronóstico de certificados a comprar para el 2021



Fuente: Fora

Elaboración: Propia

12.3. Estrategia de implementación

La estrategia que se quiere conseguir con este estudio es pronosticar la cantidad idónea para comprar cada mes con la finalidad de que, al terminar el trimestre, el número de certificados no usados sea mínima y con ello lograr una reducción significativa de las pérdidas. Para lo cual se determinó la mejor manera de pronosticar a través de la regresión lineal y al finalizar se presenta una manera en el que los datos deberían ser recolectados en el Excel según como se muestra en la Figura 9 para posteriormente crear las visualizaciones en Power BI siendo la Figura 10 el resultado de la implementación, donde no solo te muestra la cantidad de certificados a comprar sino, también la cantidad de certificados que se están usando, las alertas para tener un aviso de cuándo es momento de comprar certificados, la cantidad de certificados en stock y la afluencia para tener presente cómo va el flujo de afluencia.

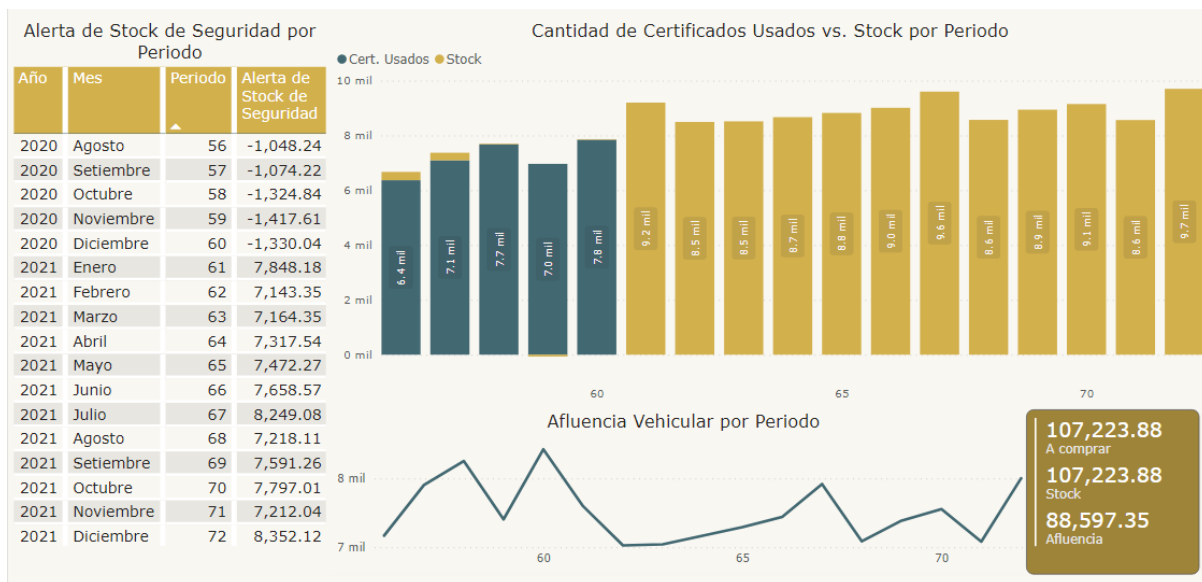
Figura 8 Excel con los datos a recolectar

AÑO	MES	PERIODO	AFLUENCIA	CERT. USADOS	STOCK	CERT. POR DÍA	ALERTA	CERT. A COMPRAR
2016	Enero	1	7,085	6,376	216	213	-1134	6,592
2016	Febrero	2	6,673	6,052	147	202	-1203	6,199
2016	Marzo	3	6,803	6,175	148	206	-1202	6,323
2016	Abril	4	6,984	6,466	29	216	-1321	6,495
2016	Mayo	5	7,104	6,683	-73	223	-1423	6,610
2016	Junio	6	6,968	6,322	158	211	-1192	6,480
2016	Julio	7	7,315	6,754	57	225	-1293	6,811
2016	Agosto	8	6,873	6,340	49	211	-1301	6,389
2016	Setiembre	9	6,916	6,193	237	206	-1113	6,430
2016	Octubre	10	7,184	6,455	231	215	-1119	6,686
2016	Noviembre	11	6,686	6,077	134	203	-1216	6,211
2016	Diciembre	12	7,532	6,793	224	226	-1126	7,017
2017	Enero	13	7,352	6,849	-3	228	-1353	6,846
2017	Febrero	14	7,051	6,478	81	216	-1269	6,559
2017	Marzo	15	7,197	6,493	205	216	-1145	6,698
2017	Abril	16	7,117	6,646	-24	222	-1374	6,622
2017	Mayo	17	7,246	6,586	158	220	-1192	6,744
2017	Junio	18	7,440	6,838	91	228	-1259	6,929
2017	Julio	19	7,669	7,232	-85	241	-1435	7,147
2017	Agosto	20	7,173	6,561	114	219	-1236	6,675
2017	Setiembre	21	7,241	6,662	78	222	-1272	6,740
2017	Octubre	22	7,311	6,696	110	223	-1240	6,806
2017	Noviembre	23	6,921	6,338	97	211	-1253	6,435
2017	Diciembre	24	7,857	7,411	-85	247	-1435	7,326
2018	Enero	25	7,661	7,062	77	235	-1273	7,139
2018	Febrero	26	6,957	6,302	167	210	-1183	6,469

Fuente: Fora

Elaboración: Propia

Figura 9 Panel del modelo final para Fora



Fuente: Fora

Elaboración: Propia

13. Calibrado e Implementación

13.1. Métodos de ajuste

En base al método que se aplicó para el ajuste de la regresión lineal se realizó a través de una evaluación del error típico, en otras palabras, la variación producida por los valores pocos conocidos. Por ello, si el error es más pequeño que el error típico, serán más acertadas las predicciones que se ejecutan a partir de dicho modelo. Es preferente recalcar que se empleó la herramienta de Excel para hacer el método.

Calibrado al 40%

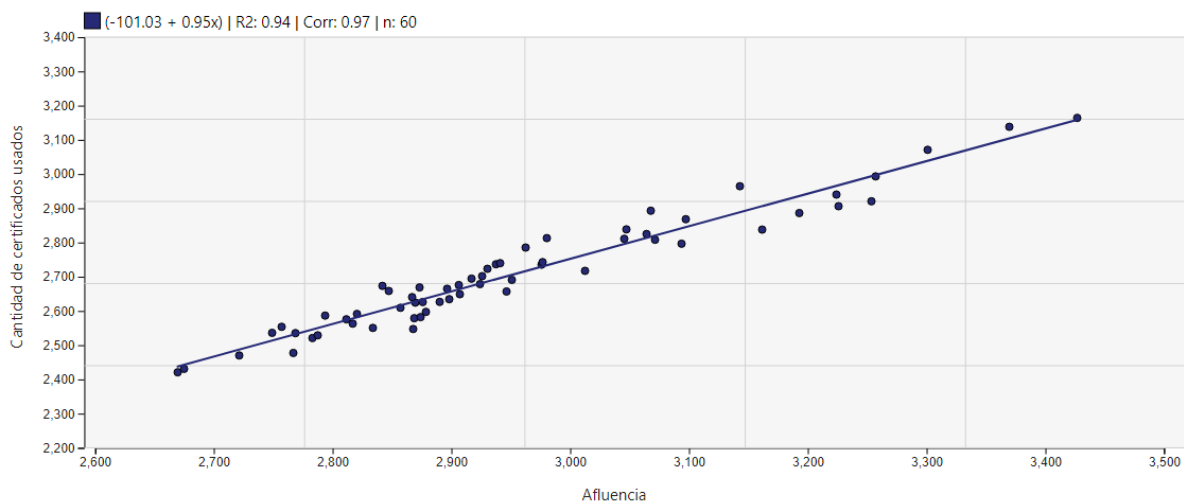
Figura 10 Estadísticas de regresión calibrado al 40%

<i>Estadísticas de la regresión</i>	
Coefficiente de correlación múltiple	0.97201267
Coefficiente de determinación R ²	0.94480864
R ² ajustado	0.9428721
Error típico	39.7995165
Observaciones	60

Fuente: Fora

Elaboración: Propia

Gráfico 13 Correlación lineal entre las variables Cantidad de certificados usados y Afluencia calibrado al 40%

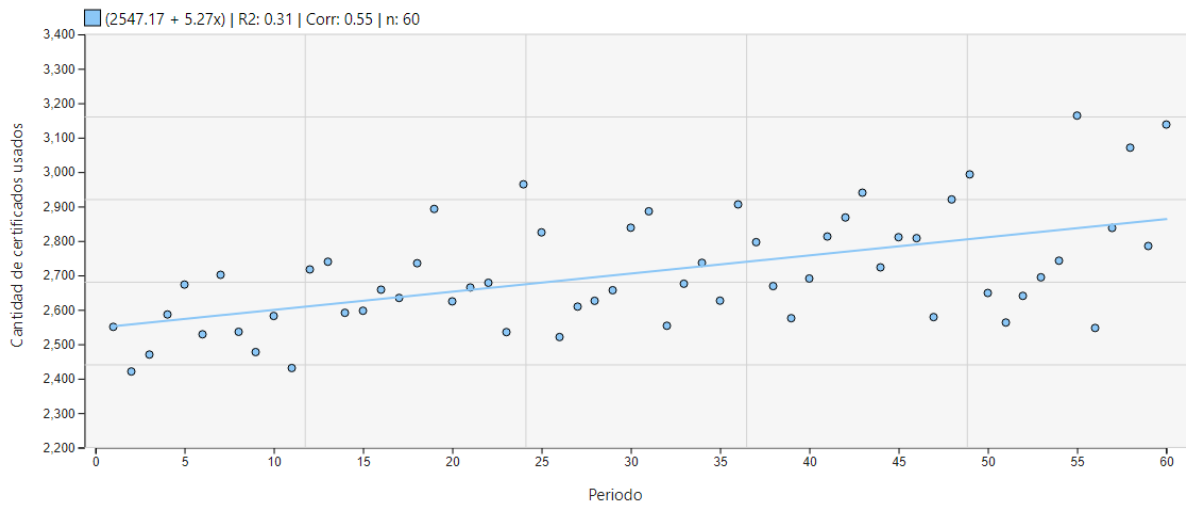


Fuente: Fora

Elaboración: Propia

En el calibrado se usó la diferencia, es decir, el 40% de la base total. A diferencia del muestreo, donde se empleó el 60%, aquí el error típico es de 39.79 con un coeficiente de correlación múltiple del 0.97.

Gráfico 14 *Correlación lineal entre las variables Cantidad de certificados usados y el Periodo calibrado al 40%*



Fuente: Fora

Elaboración: Propia

Calibrado al 70%

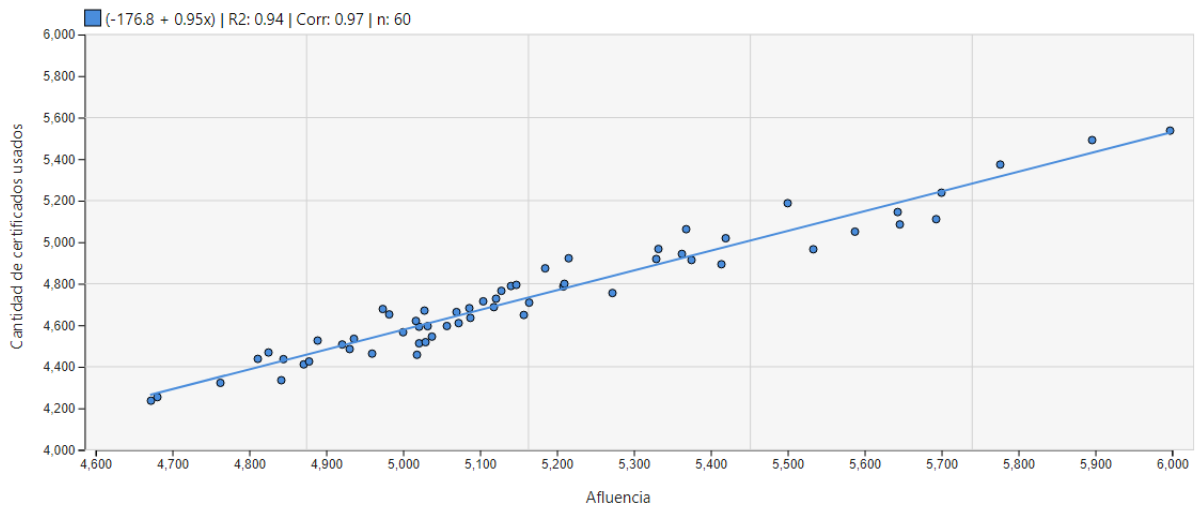
Figura 11 *Estadísticas de regresión calibrado al 70%*

<i>Estadísticas de la regresión</i>	
Coefficiente de correlación múltiple	0.97201267
Coefficiente de determinación R ²	0.94480864
R ² ajustado	0.9428721
Error típico	69.649154
Observaciones	60

Fuente: Fora

Elaboración: Propia

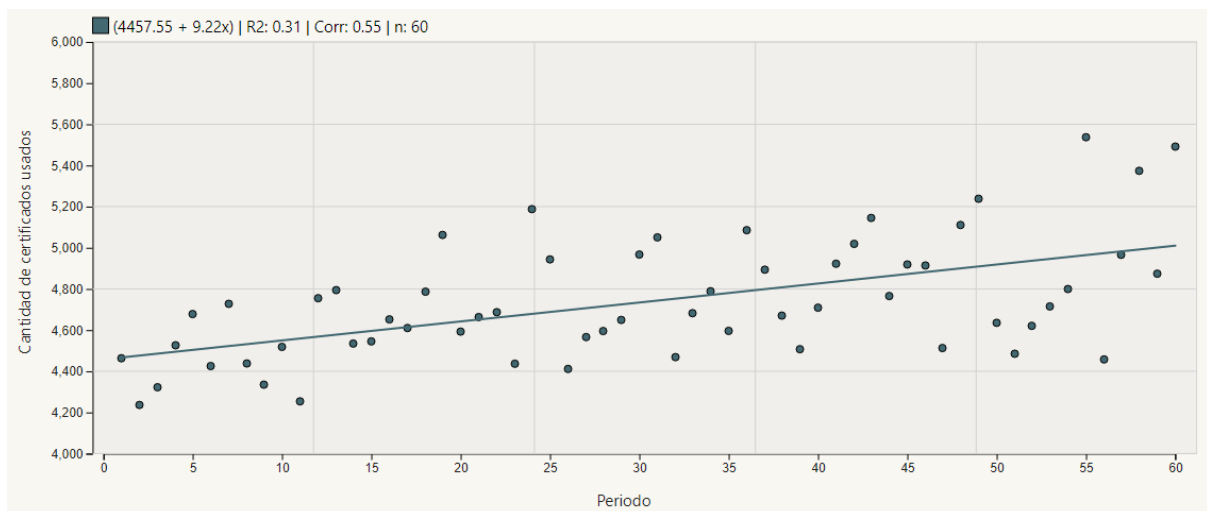
Gráfico 15 *Correlación lineal entre las variables Cantidad de certificados usados y Afluencia calibrado al 70%*



Fuente: Fora

Elaboración: Propia

Gráfico 16 *Correlación lineal entre las variables Cantidad de certificados usados y el Periodo calibrado al 70%*



Fuente: Fora

Elaboración: Propia

En este calibrado representan el 70% de los datos, en donde el coeficiente de correlación múltiple es de 0.97 y el error típico es 69.64 aumentando en 29.85, con respecto al calibrado al 40 %.

Calibrado al 100%

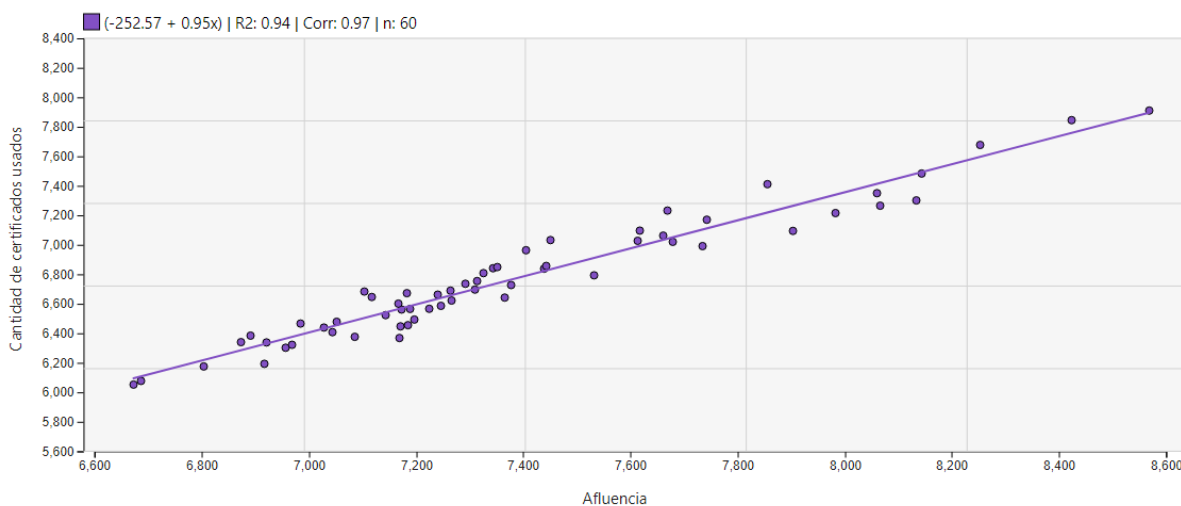
Figura 12 Estadísticas de regresión calibrado al 100%

<i>Estadísticas de la regresión</i>	
Coefficiente de correlación múltiple	0.97201267
Coefficiente de determinación R ²	0.94480864
R ² ajustado	0.9428721
Error típico	99.4987914
Observaciones	60

Fuente: Fora

Elaboración: Propia

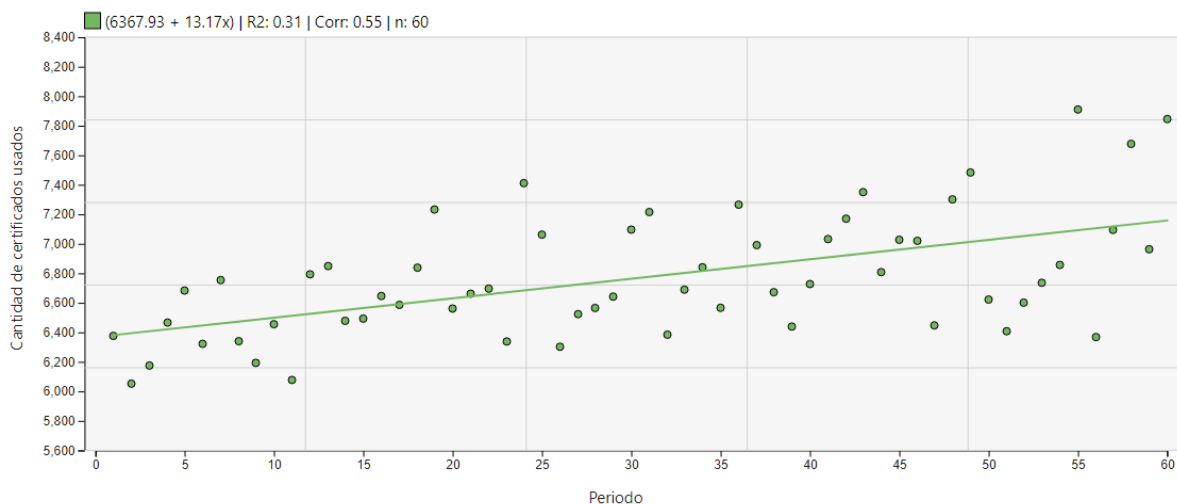
Gráfico 17 Correlación lineal entre las variables Cantidad de certificados usados y Afluencia calibrado al 100%



Fuente: Fora

Elaboración: Propia

Gráfico 18 Correlación lineal entre las variables Cantidad de certificados usados y el Periodo calibrado al 100%



Fuente: Fora

Elaboración: Propia

Debido al calibrado se puede elaborar regresiones múltiples con el fin de encontrar la predicción más certera a través del menor error típico posible conjunto con el mayor coeficiente de correlación. Sin embargo, para el calibrado con el total de los datos, el error típico es de 99.49, mostrando una diferencia mucho mayor a los anteriores calibrados.

13.2. Retroalimentación

Para este proyecto se recolectó una extensa información que ayudó a desarrollar la investigación, en el cual se tuvo que identificar la afluencia vehicular, la cantidad de certificados usados con respecto a los años 2016 al 2020 contemplando el periodo y adicionar la cantidad de vehículos vendidos en el 2017 porque representan futuros clientes para la empresa, todo esto para poder obtener un resultado de cuál sería la cantidad óptima de certificados con la que deberían abastecerse.

En los puntos en donde se aplicó la estadística, se detecta que el método supervisado de regresión lineal múltiple y regresión de serie de tiempo son los modelos más eficaces para la predicción de la cantidad de certificados a comprar de la empresa FORA. Lo cual, se sostiene que el valor del coeficiente R2 ajustado aplicando una calibración del 40%, 70% o 100% nos da un 0.97; por esta razón, el modelo más fiable a utilizar sería basado en el error típico, es decir, con la calibración del 40% con 39.79.

13.3. Conclusiones preliminares

Para el presente trabajo se tuvo como propósito identificar la mejor forma de pronosticar la cantidad de certificados a comprar, pues se observó que existía un problema que es la pérdida monetaria que surge a raíz de la compra excesiva de certificados.

Por otro lado, la pandemia es una condición que afecta el pronóstico, pero se puede contrarrestar con los resultados del histórico de la empresa calculando el promedio de la variable de los meses de años anteriores, de esta manera se puede obtener un resultado cercano a una realidad sin pandemia. Hacer este cálculo fue necesario para realizar los modelos de regresión.

14. Proyecto completo de Ciencia de datos

14.1. Factores críticos de éxito del proyecto

El presente trabajo de investigación se enfocó en encontrar soluciones que ayuden al área de logística a adquirir la cantidad mensual necesaria de certificados en la sede de Surco por medio del modelo propuesto. Por ende, los factores críticos de su éxito fueron aquellos que brindaron una visión integral de la gestión de datos de la empresa Fora, y que pueden llevar a cumplir los siguientes puntos:

- Determinar la afluencia de vehículos de los años 2016 al 2020 en la sede de Surco.
- Determinar la cantidad de certificados usados en la sede de Surco del 2016 al 2020.
- Determinar la cantidad posibles clientes a través de los vehículos vendidos en el 2017.

Por ende, los factores críticos que favorecieron al éxito del presente estudio fueron: el hallazgo de las variables determinantes, el esclarecimiento de los objetivos propuestos, el análisis minucioso de los datos y de sus resultados, y la ejecución de los procesos de la metodología de la ciencia de datos.

14.2. Conclusiones

Este trabajo ayudó a conocer a la empresa FORA e identificar uno de los problemas al que se enfrenta y de cual representa un problema traducido en pérdidas económicas como consecuencia de no contar con el cálculo específico de la cantidad de certificados que debe de comprar según las variables que afectan directamente, sin que haya un exceso de estos, ya que estos documentos expiran después de tres (03) meses, y el excedente tiene que ser eliminado.

Por consiguiente, se concluye que para reducir pérdidas económicas y el sobrestock era necesario poseer un mejor control en la compra de certificados gracias al pronóstico de estos, en donde se encontró factible elaborar la regresión lineal multivariable.

14.3. Nuevas preguntas

Con el desarrollo del proyecto, es posible que aparezcan nuevas inquietudes y hacernos nuevas preguntas como:

- ¿Qué otros inventarios de otras áreas tendrán un problema similar?

- ¿Habrá más datos ocultos que se relacionen con el problema?
- ¿Qué tanto afecta el pronóstico de tener el dato de la participación de mercado?

14.4. Decisiones a tomar a partir de los datos estudiados

La toma de decisiones basada en datos se define como el uso de hechos, indicadores y datos para guiar decisiones estratégicas que sean consistentes con las metas, los objetivos y los planes de la organización. Por ende, las disposiciones basadas en la presente investigación son las siguientes:

1. Se puede considerar optar por implementar un stock mínimo de seguridad, para no perder la oportunidad de colocar un certificado y a la vez se reduce el porcentaje de pérdida.
2. La organización puede desglosar la cantidad de certificados usados, y analizar uno de sus componentes como es certificados anulados y reducir la cantidad de este cada mes.
3. El área de logística incorporará el uso de Power Bi, con la finalidad de brindar un mejor análisis de consumo de certificados.
4. La empresa empleará una ERP o actualizará su sistema, con el propósito de unificar todas las áreas que integran a la misma.

14.5. Autorización de publicación

Anexamos el enlace de la autorización de publicación en el repositorio académico:

<https://cutt.ly/vmKkdi2>.

14.6. Video

Incorporamos el enlace del video corto explicando los detalles del presente trabajo:

<https://cutt.ly/PmKbLiB>.

REFERENCIAS

- Asociación Automotriz del Perú. (2019). Impacto de un sistema de revisiones técnicas eficiente: revisión y experiencias. *Actualizate con la AAP*, (2). <https://aap.org.pe/descarga/revisiones.pdf>
- Blackboard UPC. (s.f.). *Recursos del curso Data Science III - Especialist AP317-2101-QD12*.
- ¿Cuáles son las peores ciudades para conducir del mundo?. (2020, febrero 5). *El Diario Vasco*. <https://www.diariovasco.com/internacional/peores-ciudades-conducir-mundo-20200205145750-nt.html>
- Decreto Supremo que aprueba el Reglamento Nacional para el Fomento del Chatarreo. (2021, enero 11). *El Peruano*. <https://busquedas.elperuano.pe/normaslegales/decreto-supremo-que-aprueba-el-reglamento-nacional-para-el-f-decreto-supremo-n-005-2021-mtc-1925769-4/>
- Elecciones en Perú podrían avivar incertidumbre en los próximos meses, señala Bank of America. (2021, febrero 18). *Gestión*. <https://gestion.pe/economia/elecciones-2021-elecciones-en-peru-podrian-avivar-la-incertidumbre-en-los-proximos-meses-senala-bank-of-america-noticia/>
- Fora. (s.f.). *Página oficial de Fora*.
- García, A. (2018, noviembre, 25). Cerca de un millón de vehículos circulan sin revisión técnica. *Diario Perú21*. <https://peru21.pe/lima/cerca-millon-vehiculos-circulan-revision-tecnica-442603-noticia/>
- MTC amplió vigencia de brevets de vehículos particulares hasta el 30 de junio. (2021, febrero 1). *Andina Agencia Peruana de Noticias*. <https://andina.pe/agencia/noticia-mtc-amplio-vigencia-brevets-vehiculos-particulares-hasta-30-junio-832026.aspx>

Tipo de cambio cierra a S/ 3.72, nueva máximo histórico, por aversión al riesgo. (2021, marzo, 23). *Gestión*. <https://gestion.pe/economia/mercados/dolar-tipo-de-cambio-cierra-a-s-372-nuevo-maximo-historico-por-aversion-al-riesgo-nndc-noticia/>