



Facultad de Ciencias Naturales y Matemática

APLICACIÓN DEL ALGORITMO DE REGRESIÓN LINEAL USANDO PYTHON PARA
OPTIMIZAR LOS RECURSOS HUMANOS EN PUESTOS DE INSUMOS QUÍMICOS DE
SUNAT

Líneas de Investigación:

Matemática pura y aplicada

Trabajo de Suficiencia Profesional para optar el Título Profesional de Licenciado en
Matemática

Autora:

Salazar Marín, Gloria Mercedes (0000-0003-1748-5043)

Asesor:

Lic. Absalón Castillo Valdivieso (0000-0002-6083-9321)

Lima - Perú

2022

Título

Aplicación del algoritmo de regresión lineal usando Python para optimizar los recursos humanos en puestos de Insumos Químicos de SUNAT.

Gloria Mercedes Salazar Marín

Dedicatoria

Gracias a los esfuerzos y al amor desmesurado de mis padres por apoyarme en el largo camino de mi formación universitaria. Gracias a mis maestros por guiarme e incentivarme a ser un profesional que aporte a mi país.

Agradecimientos

Agradezco a Dios por su infinito amor de permitir que mi familia siga unida pese a las adversidades de esta pandemia.

Quiero agradecer de forma muy especial a Carlos Palomino, por ser la persona que me motivó a seguir creciendo profesionalmente, por estar en el momento oportuno y poner una luz en mi camino.

Agradezco a mi asesor el Lic. Absalón Castillo por guiarme con sus valiosos aportes en el presente trabajo de investigación, por transmitirme sus enseñanzas y sabios consejos para continuar en mi largo camino de mi carrera profesional.

Índice de Contenido

Resumen.....	9
Abstract.....	10
I. Introducción.....	11
1.1. Trayectoria del Autor.....	12
1.2. Descripción de la Empresa	12
1.2.1. Misión.....	13
1.2.2. Visión.....	13
1.3. Organigrama de la Empresa	15
1.4. Áreas y Funciones Desempeñadas	16
II. Descripción de Actividad Específica.....	18
2.1. Conceptos Básicos para el Desarrollo del Trabajo.....	18
2.1.1. Bienes Fiscalizados.....	18
2.1.2. Zonas Geográficas Bajo Régimen Especial.....	18
2.1.3. Equipos y Maquinarias	18
2.1.4. Rutas Fiscales	19
2.1.5. Puestos de Controles Obligatorios.....	20
2.1.6. Bienes Fiscalizados de Uso Doméstico	20
2.2. Prevención y Detección de Desvío Ilegal de Insumos Químicos.....	23
2.3. Problemática.....	27
2.3.1. Problema General	28
2.4. Objetivo.....	28

2.5. Desarrollo de la Investigación.....	28
2.5.1. Análisis de Información.....	29
2.5.2. Análisis ETL.....	31
2.5.3. Aprendizaje Automático de Datos.....	35
III. Aportes de la investigación.....	50
IV. Conclusiones.....	51
V. Recomendaciones.....	52
VI. Anexos.....	57

Índice de Figuras

Figura 1 - Experiencia laboral del autor	12
Figura 2 - Organigrama SUNAT	16
Figura 3 - Proceso de fiscalización	23
Figura 4 - Proceso de elaboración de cocaína.....	24
Figura 5 - Inspecciones vehiculares	25
Figura 6 - Herramientas de pruebas de campo.....	26
Figura 7 - Equipos de protección personal (EPP).....	27
Figura 8 - Metodología de trabajo	29
Figura 9 - Registro de ocurrencias	30
Figura 10 - Registro de intervenciones	31
Figura 11 - Malla curricular.....	34
Figura 12 - Archivos del proyecto	34
Figura 13 - Ciencia de datos	36
Figura 14 - Algoritmos matemáticos	37
Figura 15 - Conexión de datos	40
Figura 16 - Código para la extracción de información	40
Figura 17 - Histograma de variables.....	41
Figura 18 - Matriz de correlación grafica	44
Figura 19 - Matriz de correlación resultante.....	45
Figura 20 - Grafico inicial de variables	46
Figura 21 - Gráfico de regresión lineal	47

Índice de Tablas

Tabla 1 - Rutas fiscales	19
Tabla 2 - Puestos de control obligatorios.....	20
Tabla 3 - Variables a analizar	42
Tabla 4 - Variables validadas.....	42
Tabla 5 - Matriz de correlación.....	43
Tabla 6 - Promedio de intervenciones	47
Tabla 7 - Resultado de regresión lineal.....	48

Resumen

El presente trabajo titulado Aplicación del algoritmo de regresión lineal usando Python para optimizar los recursos humanos en puestos de Insumos Químicos de SUNAT, tiene como principal objetivo demostrar que mediante la aplicación del algoritmo de regresión lineal usando Python se puede optimizar los procesos de una organización, en este caso la organización es la Superintendencia Nacional de Aduanas y de Administración Tributaria (SUNAT) en su Área de Fiscalización de Insumos Químicos, la cual se encarga de fiscalizar los transportes que puedan llevar productos químicos de forma ilegal hacia las zonas del VRAEM (Valle de los ríos Apurímac, Ene y Mantaro) para luego ser utilizados en la elaboración de drogas ilícitas, ver (Anexo A). El trabajo realizado en esta investigación determina la asignación del personal correcto para la fiscalización de los vehículos que transporten insumos químicos, para ello se recolectaron datos sobre las intervenciones realizadas en los puestos de control, que fueron validados con el objetivo de verificar su consistencia, almacenándose en una base de datos, para luego proceder a realizar el análisis de los datos con el modelo de Regresión Lineal, usando el lenguaje de programación Python, que permitió analizar los días de mayor y menor afluencia de vehículos para predecir la cantidad de personal adecuada con la finalidad de optimizar el recurso humano en los puestos de control. La información recolectada durante las intervenciones en los puestos de control facilitará la investigación del presente trabajo.

Palabras Clave: Lenguaje de programación Python, regresión lineal, insumos químicos.

Abstract

The present work entitled Application of the linear regression algorithm using Python to optimize human resources in Chemical Inputs positions at SUNAT, has as main objective to demonstrate that through the application of the Linear Regression algorithm using Python the processes of an organization can be optimized, in this case the organization is the National Superintendence of Customs and Tax Administration (SUNAT) in its area of control of chemical inputs, which is responsible for supervising the transports that can carry chemical products illegally to the areas of the VRAEM (Valley of the Apurimac, Ene and Mantaro rivers) to then be used in the manufacture of illicit drugs, see (Annex A). The work carried out in this research determines the correct personnel assignment for the inspection of vehicles that transport chemical inputs, for this purpose data were collected on the interventions carried out at the checkpoints, which were validated with the aim of verifying their consistency, stored in a database, and then proceed to perform the analysis of the data with the Linear Regression model, using the Python programming language, which allowed to analyze the days of greater and lesser influx of vehicles to predict the appropriate number of personnel, in order to optimize the human resource in the checkpoints. The information collected during the interventions at the checkpoints will facilitate the investigation of this work.

Keywords: Python programming language, linear regression, chemical inputs.

I. Introducción

En la actualidad muchos de los procesos administrativos dentro de las organizaciones están basados en el uso de herramientas tecnológicas que permiten la optimización de recursos a través del análisis de datos, estas herramientas están basadas en algoritmos matemáticos que permiten analizar tendencias del comportamiento organizacional y, en base a ello, implementar mejoras significativas en los procesos como la asignación correcta del recurso humano.

Precisamente, la óptima asignación del recurso humano es uno de los principales inconvenientes que se ha podido encontrar en los procesos operacionales de la Superintendencia Nacional de Aduanas y de Administración Tributaria (SUNAT) en el área de insumos químicos, cuando se fiscaliza los vehículos que puedan transportar insumos químicos de manera ilegal o al realizar operaciones claves de fiscalización, ya que no cuenta con personal necesario para la ejecución de las intervenciones; asimismo, en días de poca afluencia vehicular, la cantidad de personal asignado es excesiva para la carga laboral registrada. Esto conlleva a que cuando se presenta bastante afluencia vehicular se deje pasar los vehículos de transporte sin ningún tipo de revisión o que los transportistas deban esperar un tiempo prolongado para su fiscalización, causando incomodidad entre ellos, muchas veces se hace necesaria la intervención de la policía para evitar problemas mayores. Por otro lado, se puede encontrar personal que no está realizando ninguna actividad, en los días que hay poca afluencia de vehículos de transporte.

Con el fin de mejorar el proceso de revisión de los vehículos que puedan transportar insumos químicos de forma ilícita, se analizarán los datos usando el algoritmo de regresión lineal con Python, lo que permitirá predecir la afluencia de vehículos y la cantidad de agentes con los que se requerirá trabajar.

1.1. Trayectoria del Autor

La autora Gloria Mercedes Salazar Marín cuenta con el grado de bachiller en la escuela profesional de Matemática y Estadística, con experiencia profesional en diversas áreas del sector público y privado a nivel nacional, dentro de las cuales está la institución gubernamental denominada Superintendencia Nacional de Aduanas y Administración Tributaria (SUNAT) en el área de Bienes Fiscalizados como la más destacada para el presente trabajo ya que ha realizado diversas actividades relacionadas al control y optimización de procesos a través de la aplicación de métodos matemáticos para determinar indicadores que ayuden a gestionar adecuadamente los recursos institucionales. En la Figura 1, se puede apreciar a la autora realizando actividades relacionadas al trabajo descrito.

Figura 1

Experiencia laboral de la autora



Nota. La figura muestra las actividades realizadas por la autora de la investigación.

1.2. Descripción de la Empresa

La Superintendencia Nacional de Aduanas y de Administración Tributaria (SUNAT), según información que encontramos en su portal web es una institución creada por la Ley N°

24829, aprobada por el Decreto Legislativo N° 501, considerada como un organismo técnico especializado vinculado al Ministerio de Economía y Finanzas.

Además, también informa que es un organismo técnico especializado, adscrito al Ministerio de Economía y Finanzas, cuenta con personería jurídica de derecho público, con patrimonio propio y goza de autonomía funcional, técnica, económica, financiera, presupuestal y administrativa que, en virtud a lo dispuesto por el Decreto Supremo No 061-2002-PCM, expedido al amparo de lo establecido en el numeral 13.1 del artículo 13° de la Ley N° 27658, ha absorbido a la Superintendencia Nacional de Aduanas (ADUANAS), asumiendo las funciones, facultades y atribuciones que por ley, correspondían a esta entidad.

1.2.1. Misión

Servir al país proporcionando los recursos necesarios para la sostenibilidad fiscal y la estabilidad macroeconómica, contribuyendo con el bien común, la competitividad y la protección de la sociedad, mediante la administración y el fomento de una tributación justa y un comercio exterior legítimo (SUNAT, 2020).

1.2.2. Visión

Convertirnos en la administración tributaria y aduanera más exitosa, moderna y respetada de la región (SUNAT, 2020).

- Exitosa, porque lograremos resultados similares a los de las administraciones de los países desarrollados.
- Moderna, porque incorporaremos en nuestros procesos las tecnologías más avanzadas y utilizaremos los enfoques modernos de gestión de riesgo y fomento del cumplimiento voluntario para enfrentar con éxito los desafíos.
- Respetada por:

- El Estado: por mantener niveles bajos de evasión y de fraude en la tributación interna y el comercio exterior, y contribuir a financiar los programas sociales y el desarrollo del país.
- Los contribuyentes y usuarios de comercio exterior: porque aquellos que son cumplidores se sienten respetados; reciben todas las facilidades para el cumplimiento de sus obligaciones y tienen confianza en la capacidad de la institución de detectar y tratar los incumplimientos.
- Sus trabajadores: porque laboran en una institución con mística, modelo dentro del estado; orgullosos de pertenecer a la organización y comprometida con su misión.
- Sus trabajadores potenciales: porque es una institución atractiva para trabajar, que compite de igual a igual con las instituciones más atractivas del Estado y con las más respetadas empresas por los mejores egresados de las más prestigiosas instituciones educativas; y es capaz de atraer gente con experiencia que se destaque en el sector público o el privado.
- Otras administraciones: porque la consultan y la toman como referente (SUNAT, 2020).

La finalidad principal de la SUNAT se encuentra en la página web (SUNAT, 2020) en la cual se puede observar que su principal finalidad es administrar los tributos del gobierno y los conceptos tributarios y no tributarios que se le encarguen por Ley, para asegurar la solvencia fiscal y estabilidad macroeconómica, aplicando correctamente la normativa existente.

Además de ello la SUNAT tiene otras funciones como asegurar también la implementación, inspección y control de la política aduanera en el territorio nacional, asegurando la correcta aplicación de tratados y convenios internacionales y normativa en vigor.

La función más relevante que cumple la SUNAT, desde el enfoque de esta investigación, es participar en el combate contra la minería ilegal, así como también con el

narcotráfico. Para ello, realiza la fiscalización de ingreso, permanencia, transporte o traslado y salida de productos de actividad minera, insumos químicos y maquinarias que puedan ser utilizadas directa o indirectamente para la elaboración de drogas ilícitas y otros fines que se establezcan mediante Ley (SUNAT, 2020). Es en esta área de fiscalización de insumos químicos, el área de análisis que se abordará.

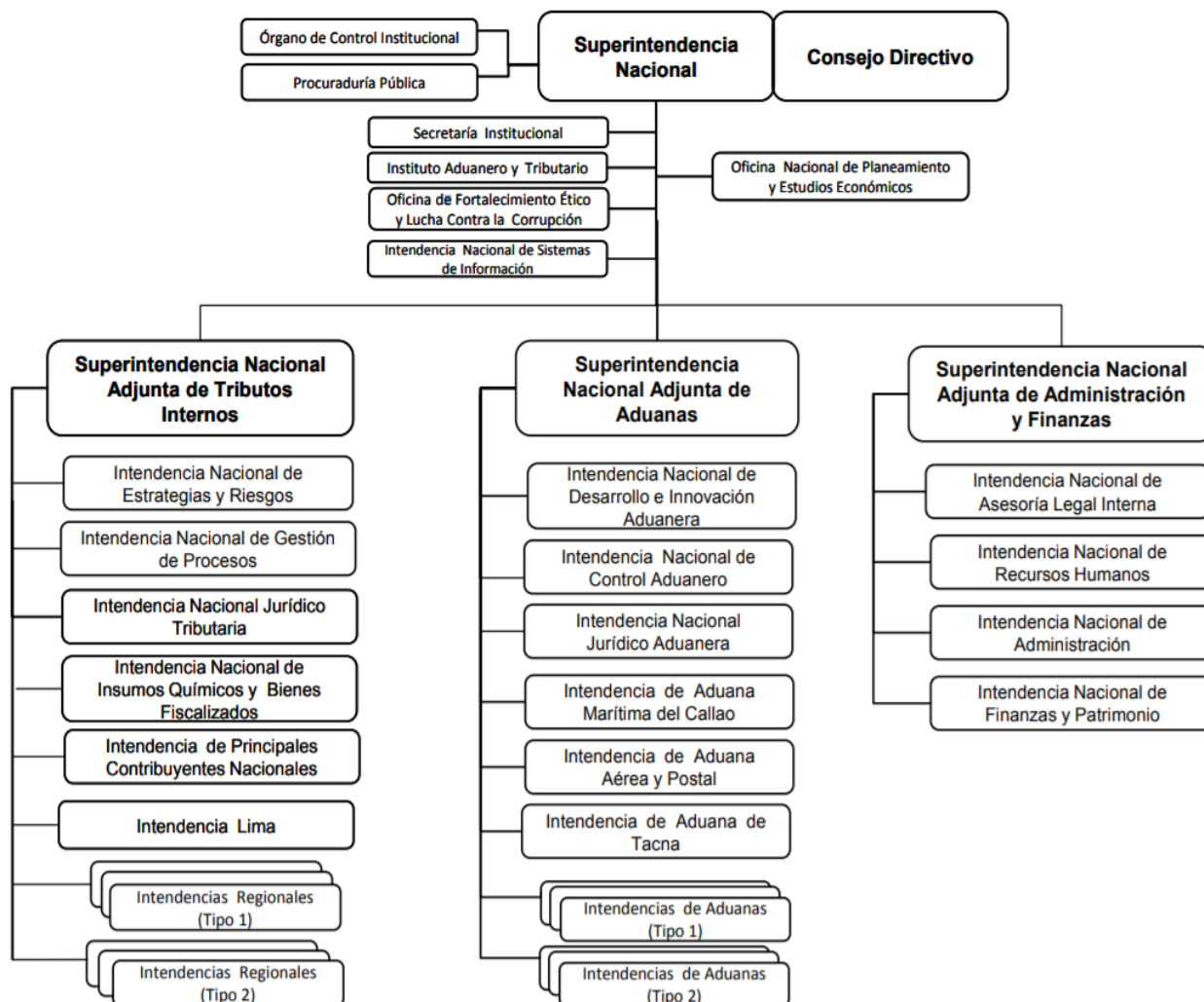
1.3. Organigrama de la Empresa

En la Figura 2 se observa el organigrama de la SUNAT, según la cual tiene tres unidades importantes, la Superintendencia Nacional Adjunta de Tributos Internos, la Superintendencia Nacional Adjunta de Aduanas y la Superintendencia Nacional Adjunta de Administración y Finanzas.

El área que se va a analizar es la Intendencia Nacional de Insumos Químicos y Bienes Fiscalizados, que como señala la misma institución (Decreto Supremo N° 016-2013-EF, 2013) es un órgano dependiente de la Superintendencia Nacional Adjunta de Tributos Internos, encargada de los procesos relativos al registro, control, fiscalización, recaudación, cobranza, devolución y restitución o reintegro sobre insumos químicos y bienes fiscalizados en materia de minería ilegal y elaboración de drogas ilícitas entre otros, a nivel nacional.

Figura 2

Organigrama de la SUNAT



Nota. Estructura orgánica de la SUNAT. Tomado de SUNAT (2020).

1.4. Áreas y Funciones Desempeñadas

El área en la cual se realizó la investigación es la Gerencia de Fiscalización de Bienes Fiscalizados; el cargo se denominó profesional para el control y fiscalización de bienes fiscalizados; las funciones realizadas en el puesto fueron las siguientes (solo se mencionarán las más relevantes para los objetivos de esta investigación):

- Proponer a la Superintendencia Nacional Adjunta de Tributos Internos, la creación, fusión, traslado y desactivación de puestos de control fijo o móvil, a nivel nacional en materia de competencia.

- Promover acciones de conducta responsable en la materia de insumos químicos y productos fiscalizados y orientar al sector público y privado sobre la correcta aplicación de la normatividad de insumos químicos y bienes fiscalizados.
- Supervisar la definición de los recursos necesarios para el desarrollo de las funciones, para su atención por parte de los órganos competentes.
- Proponer la actualización de la lista de insumos químicos y bienes fiscalizados.
- Realizar estudios e investigaciones vinculados al control de bienes fiscalizados.
- Elaborar y elevar las propuestas de mejora e informes sobre el desempeño de los sistemas y procedimientos aplicados en el ámbito de competencia.
- Formular y proponer planes operativos, así como las acciones relativas a las funciones a cargo a ser incorporadas en el plan de acción y las estrategias en el ámbito de competencia.
- Acopiar, analizar, procesar y sistematizar información de diversas fuentes, así como realizar investigaciones en materia de control y fiscalización que permitan prevenir y detectar modalidades de desvío y uso ilegal de insumos químicos y bienes fiscalizados.
- Seleccionar, programar y ejecutar acciones de control y fiscalización.

La lista anterior es una relación resumida del Decreto Supremo N° 016-2013-EF.

(Decreto Supremo N° 016-2013-EF, 2013), con las actividades que son relevantes demostrar en este trabajo, que es poder optimizar la asignación de recursos humanos (personal) a los puestos de control, a través de análisis y propuestas de mejoras y procedimientos.

II. Descripción de Actividad Específica

En esta sección se procederá a explicar la actividad específica sobre la cual está basada la presente investigación, “prevención y detención de desvío ilegal de insumos químicos”. Iniciaremos explicando algunos conceptos y definiciones importantes sobre el proceso, así como las leyes que sustentan la implementación e importancia de la actividad.

2.1. Conceptos Básicos para el Desarrollo del Trabajo

2.1.1. Bienes Fiscalizados

El Decreto Legislativo N° 1126 (2012) establece medidas de control en los insumos químicos y productos fiscalizados, maquinarias y equipos utilizados para la elaboración de drogas ilícitas, y detalla sobre bienes fiscalizados lo siguiente:

- Autorización. Permiso que otorga SUNAT al usuario para el ingreso y salida legal de Bienes Fiscalizados hacia o del territorio nacional.
- Bienes Fiscalizados. Insumos químicos, productos y sus subproductos o derivados, maquinarias y equipos utilizados, directa o indirectamente, en la elaboración de drogas ilícitas.

Los Bienes Fiscalizados se encuentran en el Anexo B.

2.1.2. Zonas Geográficas Bajo Régimen Especial

Hidrocarburos en zona geográficas bajo Régimen Especial (ver Anexo C). Se establece que la inscripción en el Registro de Hidrocarburos del Osinergmin y en el Registro de la SUNAT tendrá un plazo de vigencia máximo de un año, para los usuarios que realizan actividades con hidrocarburos en las zonas geográficas sujetas al Régimen Especial (Decreto Legislativo N° 1126, 2012).

2.1.3. Equipos y Maquinarias

Equipos. Todos aquellos que sirvan o se utilicen, directa o indirectamente, para la elaboración de drogas ilícitas (Decreto Legislativo N° 1126, 2012).

Maquinarias. Todas aquellas que sirvan o se utilicen, directa o indirectamente, para la elaboración de drogas ilícitas (Decreto Legislativo N° 1126, 2012).

2.1.4. Rutas Fiscales

Ruta Fiscal. Vía de transporte de uso obligatorio autorizada por el Ministerio de Transportes y Comunicaciones – MTC a propuesta de la SUNAT para el traslado de insumos químicos, productos, maquinarias y equipos (Decreto Legislativo N° 1126, 2012), ver Tabla 1.

Tabla 1

Rutas fiscales

Rutas Fiscales
Panamericana Norte
Carretera Central
Panamericana Sur
Ica-Ayacucho-San Francisco
Cusco-Quillabamba-Echarate
La Oroya-Huancayo-Huanta-Pacaycasa
Cusco-Abancay-Ayacucho
La Oroya-Tarma-Satipo-Mazamari
La Oroya-Huánuco-Tingo María-Pucallpa
Von Humbolt-Pto Inca
Tingo María-Tarapoto
Juliaca-Purina-DV. Ananea-Sandia
Juliaca-Azángaro-Crucero-Quiscupunco
DV Huancané-Vilquechico-Suches-Ananea
Lima-Nazca-Ático

Nota. Vías de transporte obligatorios para transportistas. Tomado de Ministerio de Transportes y Comunicaciones (2013).

2.1.5. Puestos de Controles Obligatorios

Puesto de Control Obligatorio. Lugar de control permanente que la SUNAT utiliza para el ejercicio regular de sus funciones y la verificación de los bienes sujetos a control y fiscalización. (Decreto Legislativo N° 1126, 2012), ver Tabla 2.

Tabla 2

Puestos de control obligatorios

Puesto de Control Obligatorio	Ubicación
Herrería	Av. Perú N 1661 Mz N lote 1, Pampa del Carmen (s-1) distrito y provincia de Chanchamayo, departamento de Junín
Muyurina	Kilómetro 278.847 de la carretera Huanta - Ayacucho, distrito de Jesús Nazareno, provincia de Huamanga, departamento de Ayacucho.
Pucusana (Salida Lima)	Kilómetro 56.6 de la Panamericana Sur, margen izquierda, distrito de Pucusana, provincia y departamento de Lima.
Ancón (Salida Lima)	Kilómetro 41 de la Panamericana Norte, distrito de Ancón, provincia y departamento de Lima

Nota. Puestos de control obligatorios para transportistas. Tomado de Resolución de Superintendencia N° 162-2019/SUNAT (2019).

2.1.6. Bienes Fiscalizados de Uso Doméstico

Según el Decreto Supremo N° 268-2019-EF (2019) se define los bienes fiscalizados considerados de uso doméstico y artesanal, conforme a lo establecido en los artículos 5 y 16 del Decreto Legislativo N°1126.

Mezclas. Son las que contengan dentro de su composición algunos de los siguientes insumos químicos en la concentración que a continuación se señalan:

- Del ácido clorhídrico en una concentración superior al 10%.
- Del ácido sulfúrico en una concentración superior al 10%.

- Del permanganato de potasio en una concentración superior al 2%.
- Del carbonato de sodio en una concentración superior al 10%.
- Del carbonato de potasio en una concentración superior al 10%.
- Del sulfato de sodio en una concentración superior al 30%.
- Del óxido de calcio en una concentración superior al 40%.
- Del hidróxido de calcio en una concentración superior al 40%.
- Del ácido nítrico en una concentración superior al 10%.
- Del ácido fórmico en una concentración superior al 10%.
- Del hidróxido de calcio y óxido de calcio en concentraciones que sumadas superen al 40%.

Disolventes. Toda mezcla líquida orgánica, capaz de disolver o disgregar otras sustancias, que contenga uno o más insumos químicos fiscalizados tales como acetona, acetato de etilo, acetato de n-propilo, benceno, éter etílico, hexano, metil etil cetona, metil isobutil cetona, tolueno y xileno, que hayan sido incorporados directa o indirectamente, en concentraciones que sumadas sean superiores al 20% en peso. Dichos disolventes se encuentran sujetos al registro, control y fiscalización, aun cuando contengan un aditivo de cualquier naturaleza que le dé coloración, en tanto no pierdan sus características de disolvente.

Bienes Fiscalizados. considerados para uso doméstico artesanal, son considerados bienes fiscalizados para uso doméstico, los siguientes:

1. Carbonato de sodio decahidratado, comercialmente denominado como sal de soda o sal de sosa cristalizada.
 - a. En envases de hasta doscientos gramos.
 - b. Cantidad máxima por mes: cincuenta kilogramos por adquiriente.
2. Óxido de calcio
 - a. En envases de hasta veinticinco kilogramos.

- b. Cantidad máxima por mes: cincuenta kilogramos por adquirente.
3. Hidróxido de calcio
 - a. En envases de hasta veinticinco kilogramos.
 - b. Cantidad máxima por mes: cincuenta kilogramos por adquirente.
4. Gasolinas, gasoholes, diésel y sus mezclas con biodiésel
 - a. Aquellas surtidas directamente en envases que no sean de vidrio o de material frágil, hasta un máximo de diez galones de diésel y sus mezclas con biodiésel o hasta cinco litros de gasolina y/o gasoholes, por día, para su utilización en actividades de uso doméstico.
5. Mezclas que dentro de su composición contengan carbonato de sodio o carbonato de potasio en una concentración superior al 10%.
 - a. En envases de hasta cinco kilogramos.
 - b. Cantidad máxima por mes: cinco kilogramos por adquirente.
6. Disolventes sujetos a control y fiscalización
 - a. En envases de hasta un galón.
 - b. Cantidad máxima por mes: un galón por adquirente.
7. El ácido nítrico considerado para uso artesanal debe ser comercializado de acuerdo con las disposiciones siguientes:
 - a. En concentración porcentual de hasta 65%.
 - b. En envases de hasta un litro.
 - c. Cantidad máxima por mes: un litro por adquirente.

En el Decreto Legislativo N° 1103 (2012) se establecen medidas de control y fiscalización en la distribución, transporte y comercialización de insumos químicos que puedan ser utilizados en la minería ilegal. Asimismo, se define:

- a. Hidrocarburos. Comprende Diésel, Gasolinas y Gasoholes.

- b. Insumos Químicos. El mercurio, cianuro de potasio, cianuro de sodio y los Hidrocarburos.

Las personas que trasladen bienes sujetos a control y fiscalización a que hace referencia el referido Decreto Legislativo, están obligadas a presentarse en los puestos de control obligatorio ubicados en la ruta fiscal que le corresponda utilizar para la verificación y visado correspondiente.

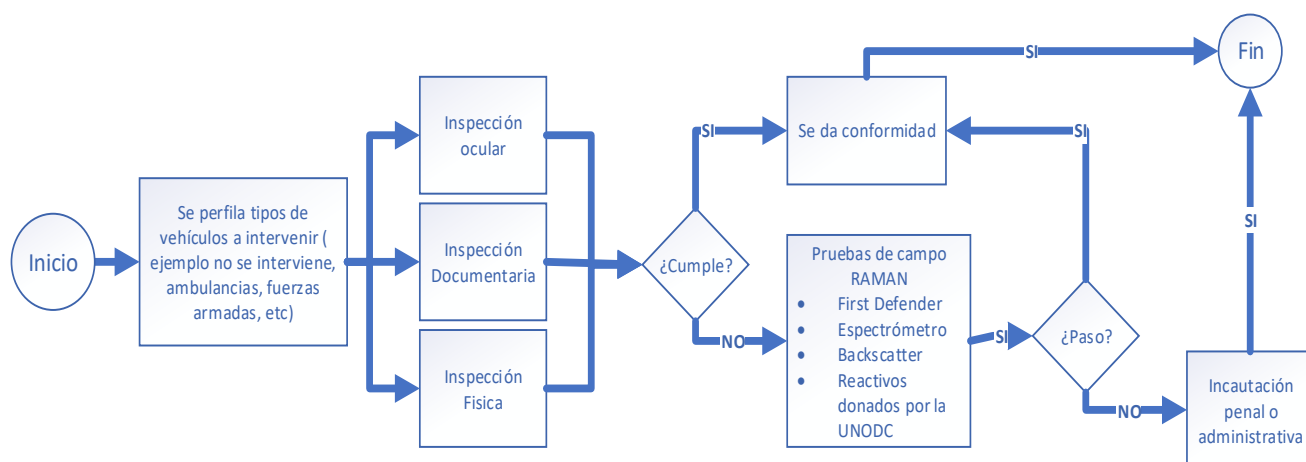
2.2. Prevención y Detección de Desvío Ilegal de Insumos Químicos

Respecto a la fiscalización en los puestos de control y en los operativos móviles en carretera se establecen rutas fiscales (ver Tabla 1) que, según el Decreto Legislativo N° 1126 (2012), se define como vía de transporte de uso obligatorio autorizada por el Ministerio de Transportes y Comunicaciones (MTC) a propuesta de la SUNAT para el traslado de insumos químicos, productos, maquinarias y equipos: se puede observar la Figura 3 que explica cuáles son los pasos a seguir durante las intervenciones.

Además, en la Figura 4, se puede apreciar el proceso de la elaboración de cocaína y como intervienen cada uno de los diferentes tipos de insumos químicos en ella.

Figura 3

Proceso de fiscalización

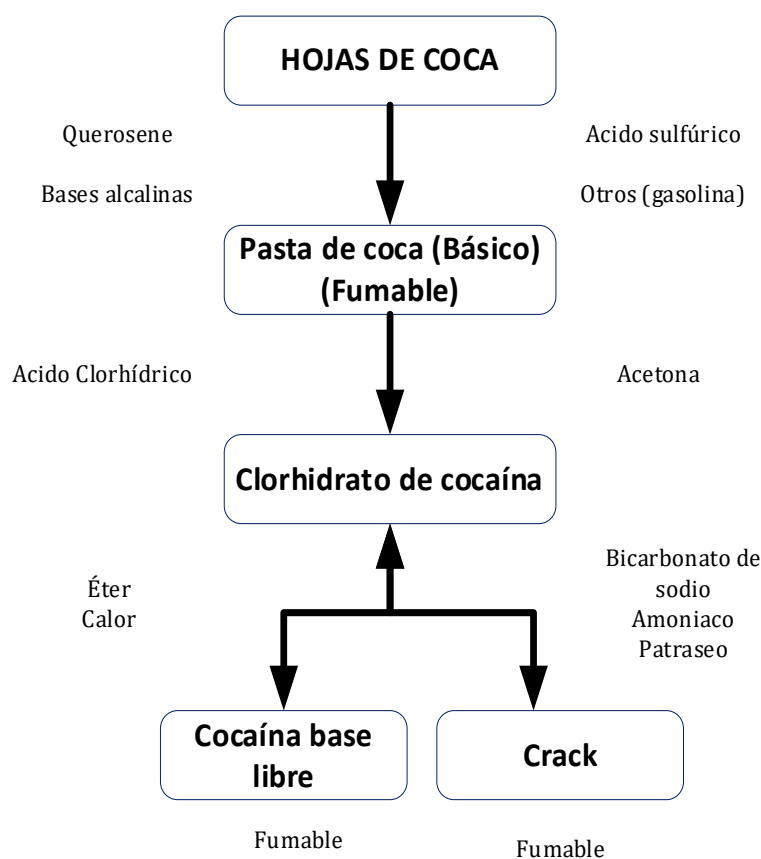


Nota. Proceso de fiscalización de insumos químicos detallado.

Las intervenciones de fiscalización en los puestos de control y operativo móvil en carreteras se realizan los 365 días del año de manera consecutiva y consiste en verificar que los bienes y el transporte se encuentren sustentados documentariamente con el objetivo de generar riesgo en los usuarios que no cumplen con presentar los documentos necesarios y que puedan trasladar productos fiscalizados de manera ilegal. Asimismo, permite relevar información para el desarrollo de futuras fiscalizaciones en carreteras y puestos de control.

Figura 4

Proceso de elaboración de cocaína



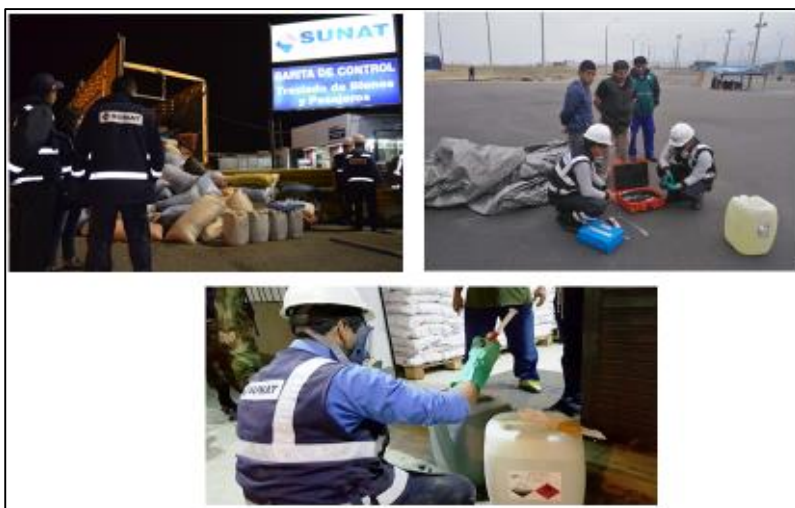
Nota. Proceso detallado de la elaboración de cocaína.

Para realizar dicha operatividad se requiere perfilar los vehículos que no serán materia de intervención de verificación como vehículos vacíos, ambulancias, vehículos blindados de empresas que presten servicio de transporte de valores, vehículos de transporte destinados para el servicio de limpieza pública, vehículos de las fuerzas armadas, policiales y vehículos de

bomberos. Las modalidades de fiscalización en los transportes que serán verificados e inspeccionados son: inspección ocular, revisión documentaria y revisión física. En la Figura 5 se puede apreciar como se realizan estas verificaciones.

Figura 5

Inspecciones vehiculares



Nota. Se aprecia los diferentes tipos de inspecciones realizadas.

Cuando la revisión del vehículo cumpla con todos los requerimientos solicitados se dará conformidad a los documentos con el sello del profesional fiscalizador de la SUNAT. En caso no cumpla con los controles o los documentos necesarios para sustentar el traslado de los bienes fiscalizados y se presuma que el vehículo traslade insumos químicos, se realizarán las pruebas de campo con tecnología Raman como el First Defender, espectrómetro, Backscatter y reactivos donados por la Oficina de las Naciones Unidas Contra la Droga y el Delito (UNODC) como se puede apreciar en la Figura 6.

Figura 6

Herramientas de pruebas de campo



Nota. Se aprecia las diferentes herramientas que acompañan el trabajo de fiscalización día a día.

Al ser estos insumos reactivos químicos el personal utiliza equipos de protección personal los cuales se aprecian en la Figura 7.

Si después de las pruebas preliminares se detecta y confirma la presencia de productos fiscalizados transportados de forma irregular, se procederá a realizar la incautación penal o administrativa de ser el caso.

La incautación administrativa se realiza cuando la presentación de la documentación no tiene concordancia con los sistemas de la SUNAT o con la verificación e inspección física de los bienes fiscalizados transportados. Este procedimiento tiene sustento en el artículo 47 del Decreto Legislativo N° 1339:

Artículo 47. Infracciones y sanciones: El incumplimiento de las obligaciones señaladas en el presente decreto legislativo, constituye infracción a este, independientemente de las acciones de naturaleza civil o penal a que hubiere lugar. La sanción aplicable a las referidas infracciones es la incautación. Mediante decreto supremo, a propuesta de la SUNAT, refrendado por el titular del Ministerio de Economía y Finanzas en el marco de sus

competencias, se establece la tabla de infracciones y sanciones administrativas por el incumplimiento del presente decreto legislativo, así como el procedimiento sancionador respectivo a cargo de la SUNAT (Decreto Legislativo N° 1339, 2017).

Figura 7

Equipos de protección personal (EPP)



Nota. Se aprecia los equipos de protección personal usados en las pruebas de campo.

2.3. Problemática

Actualmente el Perú es considerado como el segundo productor a nivel mundial de cocaína. Para realizar esta producción se requiere de los insumos químicos para poder elaborarla, Por lo tanto, se realiza la actividad de importación y producción nacional, usando diferentes mecanismos para poder transportar de forma ilegal a través de vías no autorizadas por el MTC o de manera irregular camuflándola debajo de diversos productos para así poder pasar los controles de la SUNAT e ingresarlos hacia las zonas del VRAEM y poder utilizarlos en la producción de drogas ilícitas.

Por este motivo, la asignación del personal es muy importante para lograr una eficacia en los operativos, en los puestos de control tanto internos como externos, pudiendo así lograr revisar la mayor cantidad de transportes de acuerdo con la afluencia, esto ayudará a minimizar

el riesgo que implica dejar pasar un vehículo que este transportando insumos químicos fiscalizados de forma irregular.

2.3.1. Problema General

Según la problemática descrita de la investigación, el problema general se plantea con la interrogante:

“¿Cómo determinar el número óptimo de agentes asignados en los puestos de control para los operativos de fiscalización?”

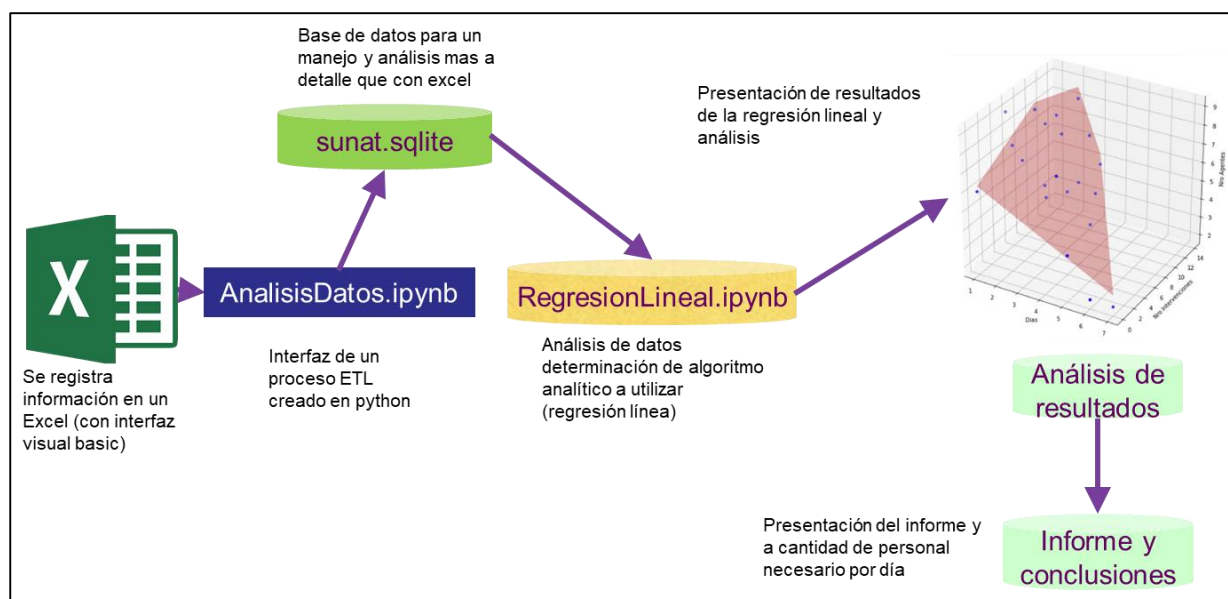
2.4. Objetivo

El objetivo principal de la presente investigación se plantea:

“Determinar el número óptimo de agentes asignados a cada puesto de control para los operativos de fiscalización aplicando modelos matemáticos en el lenguaje de programación Python”

2.5. Desarrollo de la Investigación

Para analizar y determinar los factores que más influyen en la asignación de personal en los puestos de control para la fiscalización se procedió a realizar un análisis de la información que se registra durante los operativos de fiscalización, para luego realizar un análisis de estos y determinar el modelo matemático adecuado para poder optimizar la cantidad de agentes asignados. El trabajo realizado para este trabajo se puede apreciar en la Figura 8, el cual se procederá a detallar en las siguientes secciones.

Figura 8*Metodología de trabajo*

Nota. Se puede apreciar los pasos que se realizaron para el análisis de los datos y la presentación de resultados.

2.5.1. Análisis de Información

Como señala Peláez Carrillo Betzari:

“En la actualidad, las organizaciones acumulan una gran cantidad de información empresarial, la cual debe ser llevada de una manera adecuada para conocer el pasado, presente y hacia dónde va la organización, sobre todo en la ola tecnológica en la que la mayoría se encuentra incursionando. El contar con información exacta, y precisa es un recurso muy valioso para las empresas, ya que de ella pueden depender variables en el crecimiento de esta; a partir de una información se toman decisiones y se proyectan metas u objetivos.” (SofOS Comunicaciones, 2014)

Las empresas ahora necesitan gestionar su información pues es un recurso valioso que ayuda a su crecimiento, puesto que ayuda a la toma de decisiones, estas decisiones lo que hacen es maximizar sus beneficios, optimizar procesos y por supuesto evitar pérdidas.

Lo primero que se realizó es el análisis de información, para ello se analizaron los datos que se generan diariamente y son almacenados en dos archivos Excel.

El primer archivo Excel es un registro de ocurrencias, el cual se aprecia en la Figura 9, en este registro de ocurrencia se almacenan los sucesos más importantes ocurridos durante la intervención, como por ejemplo lugar de la intervención, hora inicio y hora fin, la cantidad de trabajadores que se encontraron en ese momento y anotaciones generales sobre el trabajo realizado.

Y el segundo archivo Excel con el que se trabaja es el registro de intervenciones que podemos ver en la Figura 10, en el cual tenemos el detalle de la intervención, datos de transportista, datos de empresas que envían los insumos y la relación de insumos que se han intervenido, la actualización de este registro se realiza diariamente a través de una interfaz de VB (Visual Basic) en la cual los agentes ingresan la información.

Figura 9

Registro de ocurrencias

A	B	C	D	E	F	G	H	I	J	K	L
1	INTENDENCIA	PUESTO DE CONTROL SUPERVISIÓN	EQUIPO	TURNO	FECHA	HORA INICIO	HORA FIN	N° AGENTES	N° SUPERVISORES	INASISTENCIAS (REGISTRO)	CANTIDAD DE PNP
2	INIQBF	CONTROL MOVIL LIMA S1	EM2	T1	1/2/2020	08:00:00	20:00:00	3	-	-	
3	INIQBF	CONTROL MOVIL LIMA S1	EM1	T1	2/2/2020	08:00:00	20:00:00	1	-	-	
4	INIQBF	CONTROL MOVIL LIMA S1	EM1	T1	3/2/2020	08:00:00	20:00:00	1	-	-	
5	INIQBF	CONTROL MOVIL LIMA S7	EM4	T1	3/2/2020	08:00:00	20:00:00	1	-	-	
6	INIQBF	CONTROL MOVIL LIMA S1	EM2	T1	3/2/2020	08:00:00	20:00:00	1	-	-	
7	INIQBF	CONTROL MOVIL LIMA S1	EM1	T1	4/2/2020	08:00:00	20:00:00	1	-	-	
8	INIQBF	CONTROL MOVIL LIMA S7	EM4	T1	4/2/2020	08:00:00	20:00:00	1	-	-	
9	INIQBF	CONTROL MOVIL LIMA S1	EM2	T1	4/2/2020	08:00:00	20:00:00	1	-	-	
10	INIQBF	CONTROL MOVIL LIMA S1	EM1	T1	5/2/2020	08:00:00	20:00:00	2	-	-	
11	INIQBF	CONTROL MOVIL LIMA S7	EM4	T1	5/2/2020	08:00:00	20:00:00	1	-	-	
12	INIQBF	CONTROL MOVIL LIMA S1	EM1	T1	6/2/2020	08:00:00	20:00:00	1	-	-	
13	INIQBF	CONTROL MOVIL LIMA S7	EM4	T1	6/2/2020	08:00:00	20:00:00	1	-	-	
14	INIQBF	CONTROL MOVIL LIMA S7	EM3	T1	6/2/2020	08:00:00	20:00:00	1	-	-	
15	INIQBF	CONTROL MOVIL LIMA S7	EM3	T1	7/2/2020	08:00:00	20:00:00	1	-	-	
16	INIQBF	CONTROL MOVIL LIMA S1	EM2	T1	7/2/2020	08:00:00	20:00:00	1	-	-	
17	INIQBF	CONTROL MOVIL LIMA S1	EM2	T1	8/2/2020	08:00:00	20:00:00	1	-	-	
18	INIQBF	CONTROL MOVIL LIMA S7	EM3	T1	9/2/2020	08:00:00	20:00:00	1	-	-	
19	INIQBF	CONTROL MOVIL LIMA S7	EM3	T1	10/2/2020	08:00:00	20:00:00	1	-	-	
20	INIQBF	CONTROL MOVIL LIMA S7	EM4	T1	10/2/2020	08:00:00	20:00:00	1	-	-	
21	INIQBF	CONTROL MOVIL LIMA S7	EM4	T1	11/2/2020	08:00:00	20:00:00	1	-	-	
22	INIQBF	CONTROL MOVIL LIMA S1	EM2	T1	11/2/2020	08:00:00	20:00:00	2	-	-	
23	INIQBF	CONTROL MOVIL LIMA S1	EM1	T1	11/2/2020	08:00:00	20:00:00	1	-	-	
24	INIQBF	CONTROL MOVIL LIMA S1	EM1	T1	12/2/2020	08:00:00	20:00:00	1	-	-	
25	INIQBF	CONTROL MOVIL LIMA S7	EM4	T1	12/2/2020	08:00:00	20:00:00	1	-	-	
26	INIQBF	CONTROL MOVIL LIMA S1	EM2	T1	12/2/2020	08:00:00	20:00:00	1	-	-	
27	INIQBF	CONTROL MOVIL LIMA S7	EM3	T1	12/2/2020	08:00:00	20:00:00	1	-	-	
28	INIQBF	CONTROL MOVIL LIMA S7	EM4	T1	13/2/2020	08:00:00	20:00:00	1	-	-	
29	INIQBF	CONTROL MOVIL LIMA S1	EM2	T1	13/2/2020	08:00:00	20:00:00	1	-	-	
30	INIQBF	CONTROL MOVIL LIMA S7	EM3	T1	13/2/2020	08:00:00	20:00:00	1	-	-	
31	INIQBF	CONTROL MOVIL LIMA S1	EM1	T1	14/2/2020	08:00:00	20:00:00	1	-	-	

Nota. Se observa el registro de ocurrencia de las intervenciones diarias.

En estos archivos se puede visualizar toda la información necesaria para el análisis de datos correspondiente y determinar cuáles de estos factores serán las variables predictoras¹, es decir, las que ayudaran a determinar el número de agentes óptimos.

Después de analizar la información de los archivos, se determinó que muchos de los registros debían ser validados (datos faltantes, ilegibles) y ordenados correctamente, por lo que el siguiente paso consistió en realizar el proceso ETL², que como resultado dio una información limpia, lista para su procesamiento.

Figura 10

Registro de intervenciones

E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
HORA	REC	TIPO DOC REM	N. DOCUMENTO REM	TIPO COP REM	SERIE REM	CORRELATIVO REM	DEPARTAM TO PARTIDA	PROVINCIA PARTIDA	DISTRITO PARTIDA	PUNTO DE PARTIDA (Calle,Av.,Prta,Mo,Lo)	TIPO DOC DEST	N. DOCUMENTO DEST	DEPARTAMENTO LLEGADA	PROVINCIA
10:58	RD76	RUC	20100113610	GRE-REMITENTE	T011	00002040280	LIMA	LIMA	ATE		RUC	2.01E+10	ICA	ICA
11:07	RD76	RUC	2055643496	GUIA DE REMISION FISICA REMITENTE	0017	00000009314	CALLAO	PROV. CONST. DEL CALLAO	VENTANILLA	AV. NESTOR GAMBETTA 1265	RUC	2.0566E+10	AYACUCHO	PARINAC
11:23	RD76	RUC	20107034413	GUIA DE REMISION FISICA REMITENTE	0016	00000001663	CALLAO	PROV. CONST. DEL CALLAO	VENTANILLA	AV. NESTOR GAMBETTA 1265	RUC	2.0107E+10	TACNA	TAC
11:23	RD76	RUC	20107034413	GUIA DE REMISION FISICA REMITENTE	0016	00000001663	CALLAO	PROV. CONST. DEL CALLAO	VENTANILLA	AV. NESTOR GAMBETTA 1265	RUC	2.0107E+10	TACNA	TAC
11:58	RD76	RUC	20100113610	GRE-REMITENTE	T011	00002040274	LIMA	LIMA	ATE		RUC	2.01E+10	AYACUCHO	HUAMA
14:07	RD76	RUC	20662135714	GUIA DE REMISION FISICA REMITENTE	0001	00000000379	LIMA	LIMA	LURIN	ZI CAR. PANAMERICANA SUR KM. 26.5	RUC	2.0602E+10	CUSCO	QUISPIC
14:07	RD76	RUC	20662135714	GUIA DE REMISION FISICA REMITENTE	0001	00000000379	LIMA	LIMA	LURIN	ZI CAR. PANAMERICANA SUR KM. 26.5	RUC	2.0602E+10	CUSCO	QUISPIC
14:18	RD76	RUC	20370146994	GRE-REMITENTE	T075	00000000936	LIMA	LIMA	LURIGANCHO		RUC	2.037E+10	ICA	PISC
14:51	RD76	RUC	20494040596	GUIA DE REMISION FISICA REMITENTE	0002	00000001074	LIMA	LIMA	LURIN	ZI CAR. PANA. SUR KM. 26.5	RUC	2.0494E+10	ICA	ICA
15:04	RD76	RUC	20100113610	GRE-REMITENTE	T641	000000011302	LIMA	LIMA	ATE		RUC	2.01E+10	CUSCO	CUS
09:09	RD76	RUC	20536039717	FACTURA ELECTRONICA	F001	00000005946	LIMA	LIMA	COMAS		RUC	2.016E+10	LIMA	CAÑE
09:18	RD76	RUC	10282649522	GUIA DE REMISION FISICA REMITENTE	0001	00000000848	LIMA	LIMA	LURIN	PANAMERICANA SUR KM. 26.5	RUC	1.0283E+10	AYACUCHO	HUAN
09:18	RD76	RUC	10282649522	GUIA DE REMISION FISICA REMITENTE	0001	00000000848	LIMA	LIMA	LURIN	PANAMERICANA SUR KM. 26.5	RUC	1.0283E+10	AYACUCHO	HUAN
09:18	RD76	RUC	10282649522	GUIA DE REMISION FISICA REMITENTE	0001	00000000848	LIMA	LIMA	LURIN	PANAMERICANA SUR KM. 26.5	RUC	1.0283E+10	AYACUCHO	HUAN
09:21	RD76	RUC	20450700771	GUIA DE REMISION FISICA REMITENTE	0006	00000000366	CALLAO	PROV. CONST. DEL CALLAO	CALLAO	AV. NESTOR GAMBETTA KM. 11.5	RUC	2.0451E+10	ICA	PISC
09:21	RD76	RUC	20450700771	GUIA DE REMISION FISICA REMITENTE	0006	00000000366	CALLAO	PROV. CONST. DEL CALLAO	CALLAO	AV. NESTOR GAMBETTA KM. 11.5	RUC	2.0451E+10	ICA	PISC
09:42	RD76	RUC	20662907351	GRE-REMITENTE	EG01	00000000012	LIMA	LIMA			RUC	2.0603E+10	ICA	ICA
09:48	RD76	RUC	10420929779	GUIA DE REMISION FISICA REMITENTE	0002	00000000027	CALLAO	PROV. CONST. DEL CALLAO	CALLAO	AV. NESTOR GAMBETTA N° 1265	RUC	1.0421E+10	ICA	NAS
09:48	RD76	RUC	10420929779	GUIA DE REMISION FISICA REMITENTE	0002	00000000027	CALLAO	PROV. CONST. DEL CALLAO	CALLAO	AV. NESTOR GAMBETTA N° 1265	RUC	1.0421E+10	ICA	NAS
09:48	RD76	RUC	10417457220	GUIA DE REMISION FISICA REMITENTE	0001	00000000011	CALLAO	PROV. CONST. DEL CALLAO	CALLAO	AV. NESTOR GAMBETTA N° 1265	RUC	1.0417E+10	AREQUIPA	AREQU
09:57	RD76	RUC	20600170822	GUIA DE REMISION FISICA REMITENTE	0001	00000000280	LIMA	LIMA	LURIN	ANTIGUA CARRETEREA PANAMERICANA SUR LUJRN KM. 26.5	RUC	2.06E+10	CUSCO	LA CONVI
09:57	RD76	RUC	20600170822	GUIA DE REMISION FISICA REMITENTE	0001	00000000280	LIMA	LIMA	LURIN	ANTIGUA CARRETEREA PANAMERICANA SUR LUJRN KM. 26.5	RUC	2.06E+10	CUSCO	LA CONVI
09:57	RD76	RUC	20600170822	GUIA DE REMISION FISICA REMITENTE	0001	00000000280	LIMA	LIMA	LURIN	ANTIGUA CARRETEREA PANAMERICANA SUR LUJRN KM. 26.5	RUC	2.06E+10	CUSCO	LA CONVI
10:11	RD76	RUC	20100055237	FACTURA ELECTRONICA	F002	00000159729	CALLAO	PROV. CONST. DEL CALLAO	CALLAO		RUC	2.01E+10	ICA	ICA
10:11	RD76	RUC	20100055237	FACTURA ELECTRONICA	F002	00000159729	CALLAO	PROV. CONST. DEL CALLAO	CALLAO		RUC	2.01E+10	ICA	ICA
10:27	RD76	RUC	10154526027	GUIA DE REMISION FISICA REMITENTE	0004	00000003211	LIMA	LIMA	LURIN	PLANTA CONCHAN KM. 26.5	RUC	1.0155E+10	LIMA	CAÑE
10:27	RD76	RUC	10154526027	GUIA DE REMISION FISICA REMITENTE	0004	00000003211	LIMA	LIMA	LURIN	PLANTA CONCHAN KM. 26.5	RUC	1.0155E+10	LIMA	CAÑE
12:06	RD76	RUC	20603558899	GUIA DE REMISION FISICA REMITENTE	0001	00000000151	CALLAO	PROV. CONST. DEL CALLAO	CALLAO	AV. NESTOR GAMBETTA 1265	RUC	2.0604E+10	CUSCO	CUS
12:06	RD76	RUC	20603558899	GUIA DE REMISION FISICA REMITENTE	0001	00000000151	CALLAO	PROV. CONST. DEL CALLAO	CALLAO	AV. NESTOR GAMBETTA 1265	RUC	2.0604E+10	CUSCO	CUS
12:06	RD76	RUC	20603558899	GUIA DE REMISION FISICA REMITENTE	0001	00000000151	CALLAO	PROV. CONST. DEL CALLAO	CALLAO	AV. NESTOR GAMBETTA 1265	RUC	2.0604E+10	CUSCO	CUS
12:06	RD76	RUC	20603762356	GUIA DE REMISION FISICA REMITENTE	0002	00000000035	CALLAO	PROV. CONST. DEL CALLAO	CALLAO	AV. NESTOR GAMBETTA 1265	RUC	2.0604E+10	CUSCO	CUS
12:16	RD76	RUC	20100128218	GUIA DE REMISION FISICA REMITENTE	0010	00000431104	LIMA	LIMA	LURIN	CARRETERA PANAMERICANA SUR KM. 26.5	RUC	2.01E+10	AREQUIPA	ISLA
12:43	RD76	RUC	20554545743	FACTURA ELECTRONICA	F401	000000011532	CALLAO	PROV. CONST. DEL CALLAO	VENTANILLA	AV. NESTOR GAMBETTA 1265	RUC	2.0128E+10	LIMA	CAÑE
12:43	RD76	RUC	20554545743	FACTURA ELECTRONICA	F401	000000011533	CALLAO	PROV. CONST. DEL CALLAO	VENTANILLA	AV. NESTOR GAMBETTA 1265	RUC	2.0128E+10	LIMA	CAÑE
12:43	RD76	RUC	20554545743	FACTURA ELECTRONICA	F401	000000011533	CALLAO	PROV. CONST. DEL CALLAO	VENTANILLA	AV. NESTOR GAMBETTA 1265	RUC	2.0128E+10	LIMA	CAÑE
12:52	RD76	RUC	20563973537	GUIA DE REMISION FISICA REMITENTE	0001	00000000465	CALLAO	PROV. CONST. DEL CALLAO	VENTANILLA	AV. NESTOR GAMBETTA 1265	RUC	2.0564E+10	APURIMAC	COTABA

Nota. Se observa el registro de intervenciones realizadas.

2.5.2. Análisis ETL

Según Pearlman (2019), en todo proceso de negocios la información que se maneja es importante, por lo que esa información debe trasladarse y prepararse para su uso, además muestra una serie de casos en los cuales el proceso ETL es útil, de los cuales se puede resaltar (para el uso de la investigación) la sincronización con los sistemas clave de la organización.

¹ Variables que, en base a su análisis, ayudan a determinar de acuerdo a su cantidad un elemento futuro a predecir.

² Proceso de extracción, transformación y carga, donde se valida la información, se corrige y se consolida.

Para el presente trabajo, por tanto, se requiere integrar información del proceso clave de intervenciones y analizar los datos.

Pearlman (2019) también comenta que este proceso ETL consta de tres fases: extracción, transformación y carga:

- Fase 1 extracción. En esta fase se logra obtener datos limpios desde diferentes fuentes de información relevante para el tema que se quiere lograr.
- Fase 2 transformación. En esta fase se realiza la validación de datos (y es la fase en la que al momento del análisis de información se tuvo mayor trabajo para poder validar los diversos datos que se encuentran dentro de los archivos Excel), esta validación consiste en la normalización de datos, eliminación de duplicados, verificación de datos y clasificación de datos.
- Fase 3 carga. Finalmente, la carga de datos ya limpios, validados dentro de un repositorio de información, que permitirá el análisis correcto de la data, para luego poder presentar la información para la toma de decisiones.

Para la realización e implementación del proceso ETL, se utilizó el lenguaje de programación Python que es una herramienta libre muy usada para este tipo de análisis de datos y predicciones como se puede apreciar en la siguiente cita:

Python es un lenguaje de programación interpretado multiparadigma, es decir; soporta hacer programación orientada a objetos y programación imperativa. Su creador fue Guido Van Rossum y debido a que es de código libre, la comunidad de desarrolladores ha creado librerías o módulos para hacer casi cualquier cosa. En general la gente que hace análisis de datos o estadística conoce bien R Project, SAS o SPSS. Pero pocos se han acercado a Python y sus librerías para análisis de datos, las ventajas pueden ser cuestionables con respecto al software que específicamente fueron diseñados con una perspectiva estadística, pero sin duda la potencia y calidad de

librerías es muy buena. Principalmente para hacer uso de algoritmos de Machine Learning, las librerías en Python son mejores que las de R Project, sobre todo las técnicas de Deep Learning (Dlegorreta, 2015).

En la definición de Dlegorreta (2015), se indica que Python es un lenguaje de programación muy orientado a lo que es el análisis de datos, especialmente a los temas relacionados de Big Data³, además compara este lenguaje de programación con otros existentes, dándolo como preferencia en comparación sobre los demás.

El uso de lenguaje de programación es una herramienta útil y relevante como se puede observar en la Figura 11, que es la malla curricular de la carrera de matemática de la Universidad Nacional Federico Villarreal (Universidad Nacional Federico Villarreal, 2020) la cual se puede visualizar en rectángulos rojos a los cursos de programación que se deben llevar durante la carrera, esto es debido, a que como matemáticos dentro de las funciones, se encuentran poder analizar los resultados de los algoritmos aprendidos durante toda la carrera y el análisis de estos algoritmos conllevan a hacer pruebas con ellos y para ello se necesita la lectura y prueba de gran cantidad de datos, los lenguajes de programación ayudan en la automatización de esas tareas, además que muchos de los algoritmos de Big Data se basan en postulados matemáticos y estadísticos. Permitiendo entender la lógica de la programación y el manejo de base de datos.

³ Permite el análisis de gran cantidad de datos, permitiendo predecir resultados, enfocándolo a diversos tipos de negocios, actualmente la proyección que se hace de la pandemia de la COVID - 19 se realiza con un tipo de proyecto de *Big Data* (análisis de regresiones).

Figura 11

Malla curricular

PRIMER SEMESTRE		SEGUNDO SEMESTRE		TERCER SEMESTRE		CUARTO SEMESTRE		QUINTO SEMESTRE		SEXTO SEMESTRE		SÉPTIMO SEMESTRE	
ESTUDIOS GENERALES				ESTUDIOS ESPECÍFICOS									
01	INGLÉS I	08	INGLÉS II	15	INGLÉS III	22	CÁLCULO MULTIVARIABLE	27	ANÁLISIS REAL I	33	ANÁLISIS REAL II	32	ANÁLISIS REAL III
CRÉDITOS:	1	CRÉDITOS:	1	CRÉDITOS:	1	CRÉDITOS:	5	CRÉDITOS:	5	CRÉDITOS:	5	CRÉDITOS:	5
02	LENGUAJE Y COMUNICACIÓN	09	LIDERAZGO Y DESARROLLO PERSONAL	16	PSICOLOGÍA ORGANIZACIONAL	23	ÁLGEBRA LINEAL I	28	CÁLCULO DE PROBABILIDADES II	34	ESTRUCTURAS ALGEBRAICAS I	33	ESTRUCTURAS ALGEBRAICAS II
CRÉDITOS:	3	CRÉDITOS:	3	CRÉDITOS:	2	CRÉDITOS:	4	CRÉDITOS:	4	CRÉDITOS:	4	CRÉDITOS:	4
03	MATEMÁTICA BÁSICA	10	MEDIO AMBIENTE Y DESARROLLO SOSTENIBLE	17	CÁLCULO VECTORIAL	24	INVESTIGACIÓN OPERATIVA	29	ECUACIONES DIFERENCIALES ORDINARIAS I	35	MATEMÁTICA FINANCIERA Y ACTUARIALES	36	MODELOS MATEMÁTICOS II
CRÉDITOS:	6	CRÉDITOS:	3	CRÉDITOS:	5	CRÉDITOS:	4	CRÉDITOS:	5	CRÉDITOS:	4	CRÉDITOS:	4
04	FILOSOFÍA Y ÉTICA	11	SOCIOLOGÍA	18	METODOLOGÍA DE LA INVESTIGACIÓN CIENTÍFICA	25	LENGUAJE DE PROGRAMACIÓN I	30	ÁLGEBRA LINEAL II	36	MODELOS MATEMÁTICOS I	31	ANÁLISIS DE DATOS
CRÉDITOS:	3	CRÉDITOS:	2	CRÉDITOS:	3	CRÉDITOS:	4	CRÉDITOS:	4	CRÉDITOS:	4	CRÉDITOS:	4
05	ACTIVIDADES CULTURALES Y DEPORTIVAS	12	FÍSICA I	19	GEOPOLÍTICA Y REALIDAD NACIONAL	26	CÁLCULO DE PROBABILIDADES I	31	LENGUAJE DE PROGRAMACIÓN II	37	ECUACIONES DIFERENCIALES ORDINARIAS II	27	ANÁLISIS DE DATOS
CRÉDITOS:	1	CRÉDITOS:	5	CRÉDITOS:	3	CRÉDITOS:	4	CRÉDITOS:	4	CRÉDITOS:	5	CRÉDITOS:	5

Nota. Se observa la malla curricular de la carrera de matemáticas (Universidad Nacional Federico Villarreal, 2020).

Además, se trabajó con la herramienta JupyterLab⁴, para la implementación del proceso ETL, el resultado de este es la implementación de una base de datos SQLite⁵.

Los archivos creados, para el proceso ETL y la base de datos de la SUNAT. SQLite se detallan en la Figura 12.

⁴ Herramienta, tipo block de notas que permite, la codificación en el lenguaje de programación Python.

⁵ Base de datos portátil, personal de tamaño adecuado para este tipo de proyecto (al menos por el inicio de este).

Figura 12*Archivos del proyecto*

Nombre	Estado	Fecha de modificación	Tipo	Tamaño
ipynb_checkpoints	📁	3/6/2020 12:01	Carpeta de archivos	
Data	📁	4/6/2020 14:42	Carpeta de archivos	
AnalisisDatos.ipynb	📄	26/5/2020 13:05	Archivo IPYNB	17 KB
Bien.ipynb	📄	12/5/2020 16:21	Archivo IPYNB	1 KB
debug.log	📄	2/6/2020 16:52	Documento de texto	1 KB
Departamento.ipynb	📄	10/5/2020 18:52	Archivo IPYNB	1 KB
DetalleIntervencion.ipynb	📄	15/5/2020 17:04	Archivo IPYNB	1 KB
Distrito.ipynb	📄	12/5/2020 15:59	Archivo IPYNB	1 KB
Familia.ipynb	📄	10/5/2020 21:17	Archivo IPYNB	1 KB
Intervenciones.ipynb	📄	15/5/2020 14:56	Archivo IPYNB	2 KB
Ocurrencias.ipynb	📄	10/5/2020 09:47	Archivo IPYNB	2 KB
Provincia.ipynb	📄	12/5/2020 11:42	Archivo IPYNB	1 KB
Proyectomatrix_scatter.png	🖼️	2/6/2020 23:17	Archivo PNG	25 KB
PuestoControl.ipynb	📄	10/5/2020 12:33	Archivo IPYNB	1 KB
RegresionLineal - Jupyter Notebook.pdf	📄	3/6/2020 00:15	Archivo PDF	609 KB
RegresionLineal.ipynb	📄	3/6/2020 00:13	Archivo IPYNB	334 KB
Sqlite.ipynb	📄	26/5/2020 13:02	Archivo IPYNB	19 KB
Sunat - copia.sqlite	📄	26/5/2020 12:36	Archivo SQLITE	908 KB
Sunat.sqlite	📄	2/6/2020 12:59	Archivo SQLITE	908 KB
TipoDocumento.ipynb	📄	26/5/2020 12:55	Archivo IPYNB	1 KB
Transportista.ipynb	📄	10/5/2020 21:07	Archivo IPYNB	1 KB
Utilitarios.ipynb	📄	14/5/2020 21:13	Archivo IPYNB	2 KB
Variedad.ipynb	📄	13/5/2020 19:23	Archivo IPYNB	1 KB

Nota. Se observa los archivos generados para la realización del proyecto.

El proceso ETL implementado es un proceso incremental, es decir que a medida que se va generando más información de las intervenciones con el trabajo del día a día, se tendrá que ir ejecutando para cargar mayor información, la información que se carga es la que alimenta el modelo matemático predictivo, que se va refinando a medida que se analiza mayor información, detectando cambios en las tendencias de afluencia de transportes por días, determinando óptimamente el número de agentes necesarios para los puestos de control.

2.5.3. Aprendizaje Automático de Datos

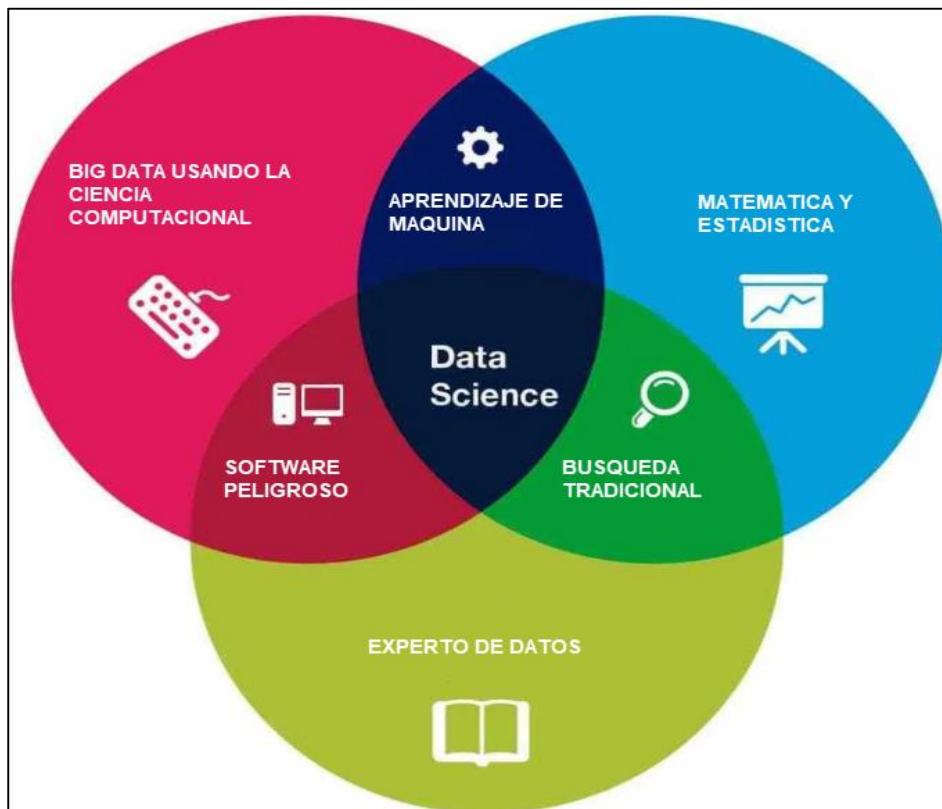
Según el portal especializado Velogig (2020), las técnicas de análisis de datos se incrementaron con el concepto de Big Data, pero a partir de ahí surgieron diversos tipos de proyecto como por ejemplo el aprendizaje de máquina, más conocido como Machine Learning, este aprendizaje de máquina permite realizar gran cantidad de proyectos relacionados con la predicción de información y todos ellos basados en algoritmos matemáticos.

Para tener un mejor panorama en la Figura 13 se presenta a detalle una explicación de lo que es la ciencia de datos y como se relaciona con las matemáticas.

Los profesionales en matemáticas se encuentran dentro del gran círculo celeste que implica conocimientos detallados de matemática y estadística para el uso del aprendizaje de máquina, que permitirá determinar tendencias después de un análisis a todas las variables relacionadas a ella.

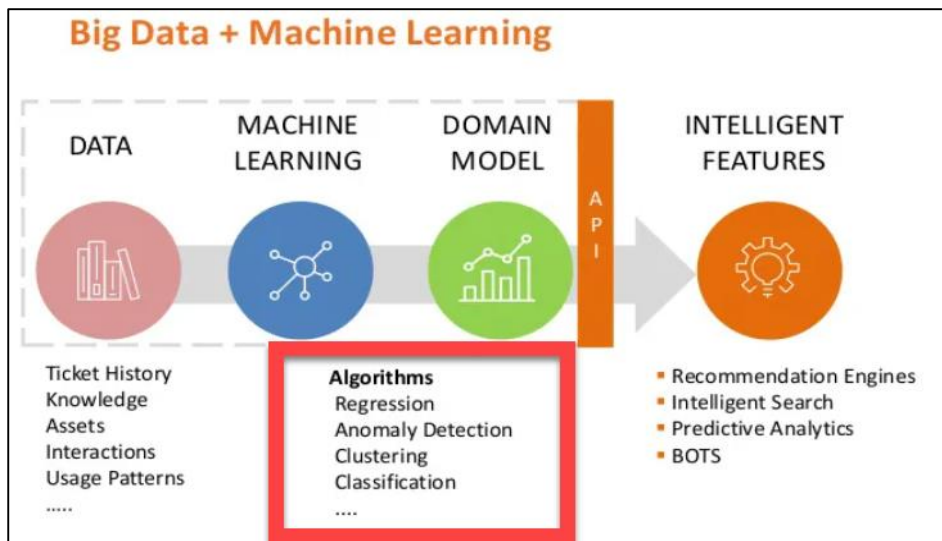
Figura 13

Ciencia de datos



Nota. Se observa la relación que hay entre la ciencia de datos y las diferentes áreas y profesiones. Tomado de Velogig (2020).

En la Figura 14, se puede ver los algoritmos matemáticos que intervienen en los proyectos de Machine Learning y que conllevan a un modelo de dato analítico que ayudan a predecir información.

Figura 14*Algoritmos matemáticos*

Nota. Se observa los algoritmos matemáticos que se usan en la ciencia de datos. Tomado de Velogig (2020).

Muchos de esos algoritmos matemáticos predictivos están basados en el aprendizaje automático que se definen de la siguiente manera:

Un conjunto de métodos que pueden detectar automáticamente patrones en los datos, y luego utilizar los patrones detectados para predecir los datos futuros, o realizar otros tipos de toma de decisiones bajo incertidumbre. (Murphy, 2012)

Un ejemplo muy común de este aprendizaje automático es la personalización de propaganda en los celulares, a medida que se hacen búsquedas de algún tipo de producto o servicio estos empiezan a aparecer con mayor frecuencia en los banners de las páginas web o redes sociales de los usuarios, he ahí el aprendizaje automático, que también como dice Murphy (2012) ayudan a predecir escenarios de datos futuros.

Es así que CISCO (2019) señala que el aprendizaje automático puede estar relacionado a:

- **Problemas de Regresión.** Es el cálculo de las relaciones matemáticas entre una variable continua y una o más variables. Esta relación matemática luego puede utilizarse para calcular los valores de una variable desconocida dados los valores conocidos de las demás. Los ejemplos de regresión son el cálculo de la posición del automóvil y su velocidad con el GPS, la predicción de la trayectoria de un tornado con datos meteorológicos, o la predicción del valor futuro de una acción mediante datos históricos y otras fuentes de información. La ejecución de la regresión significa encontrar la línea que interpola mejor los valores. La línea puede tomar varias formas y se expresa como función de regresión. Una función de regresión le permite estimar el valor de una variable dado el valor de la otra, para los valores que no se han obtenido antes.
- **Problemas de Clasificación.** Se utilizan cuando la variable desconocida es discreta. Por lo general, el problema comprende el cálculo al cual, de un conjunto de clases predefinidas, pertenece un ejemplo específico. Los ejemplos típicos de clasificación son reconocimiento de la imagen, o diagnóstico de las patologías de exámenes médicos, o identificación de rostros en una imagen. Una interpretación visual de un problema de clasificación se puede considerar en dos dimensiones, donde los puntos que pertenecen a diferentes clases se marcan con otro símbolo. El algoritmo “aprende” ejemplos de la ubicación y la forma de la línea fronteriza entre las clases. Esta línea fronteriza luego puede utilizarse para clasificar nuevos ejemplos.
- **Métodos de Agrupamiento.** Estos se pueden ver como la detección automática de grupos de ejemplos que tienen características similares, que pueden indicar posiblemente el hecho de que un miembro del grupo pertenece a una clase bien definida. Por ejemplo, los algoritmos de agrupamiento se utilizan para identificar grupos de

usuarios basados en su historial de compras en línea, y luego envían avisos dirigidos a cada miembro.

- **Métodos de Asociación.** Este es un problema muy relevante para los comerciantes en línea, y consiste en detectar grupos de elementos que se observan con frecuencia en conjuntos. Se usan para sugerir compras adicionales a un usuario, según el contenido de su carrito de compras.

Asimismo, CISCO (2019) comenta que el análisis matemático más usado para poder predecir datos es el análisis de regresión lineal, el cual consiste en analizar la correlación de variables, normalmente se usa el método Pearson, en donde un alto nivel de correlación entre las variables indicará que estas pueden aplicarse para el análisis de regresión lineal. De no tener una correlación alta se podrá optar por otro algoritmo de datos.

El análisis de regresión lineal indica que, al analizar variables predictoras y su grado de correlación, se podrá determinar una variable predecible que permitirá obtener valores a futuro, esta regresión lineal puede ajustarse continuamente a medida que la cantidad de datos aumenta y se requiera aplicar nuevamente el algoritmo,

Además, CISCO (2019) señala que se utiliza con frecuencia en el ámbito empresarial y el análisis financiero con datos históricos para informar las estrategias de acción futura. Por lo que este es el algoritmo que se utilizará para implementar una nueva estrategia de asignación óptima de recurso humano en los puntos de control, basado en el análisis de datos generados de las intervenciones de fiscalización realizadas.

El documento donde se encuentra la programación en lenguaje Python del análisis de regresión lineal se encuentra en el archivo RegresiónLineal.ipynb que se puede apreciar en la Figura 12. A continuación, se procede a explicar:

Lo primero que ha de realizarse es conectarse a la base de datos SUNAT. SQLite en la cual se guarda los datos validados por el proceso ETL, esto se logra usando algunos comandos Python y sus librerías pandas como se puede apreciar en la Figura 15.

Figura 15

Conexión de datos

```

▶ import matplotlib.pyplot as plt
  from pandas.plotting import scatter_matrix
  import numpy as np
  import pandas as pd
  import sqlite3 as sql
  from sklearn import linear_model

```

Creamos la conexión y el cursor a la base de datos sunat.sqlite

```

▶ con=sql.connect('sunat.sqlite')
  cur=con.cursor()

```

Nota. Se observa el proceso de conexión de datos usando Python.

Seguidamente, en la Figura 16 se procede a extraer la información más relevante sobre las variables que intervienen en el proceso de intervenciones a los transportistas. Además, en la Figura 17 se puede ver el primer análisis realizado mostrando el histograma de variables usadas para la predicción del número de agentes.

Figura 16

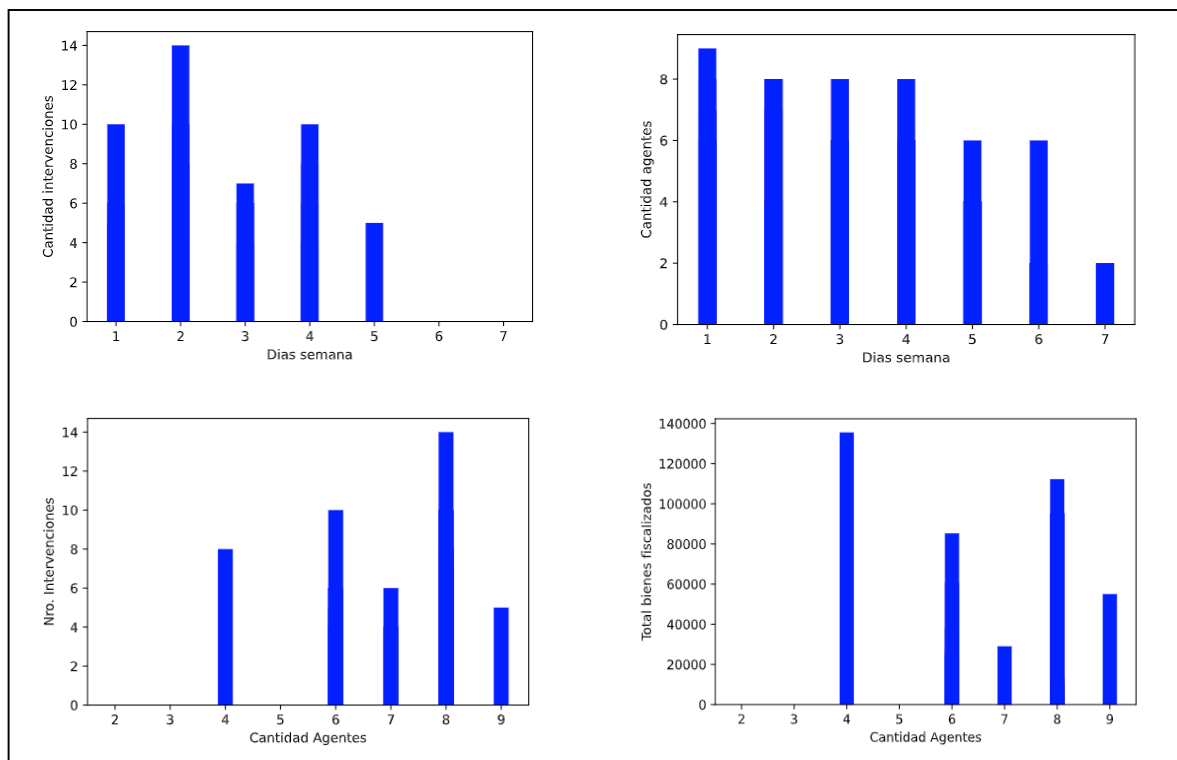
Código para la extracción de información

```

data=pd.read_sql('select fechaintervencion,nombreDia,nrointervenciones,agentes,pnp,total from AnalisisPersonas where
data.head()

```

Nota. Se observa la forma de extraer información usando Python.

Figura 17*Histograma de variables*

Nota. Se observa los histogramas de las variables realizado en Python.

Los datos que se obtienen para el análisis se presentan en la Tabla 3, se ha determinado que la variable a predecir es el número de agentes y para ello la información que brinda el proceso, es la fecha de intervención, el nombre del día de intervención, el número de intervenciones, cantidad de agentes de la Policía Nacional del Perú (PNP) asignados y el total de bienes fiscalizados (en toneladas), que serán las variables predictoras; en la Tabla 3 solo se muestra una de ellas para los datos recolectados en los últimos 3 meses, puesto que cada tres meses se recalcula el número de agentes necesarios para cada día de fiscalización, esto debido a que cada tres meses hay cambio de estación y las variables a analizar varían.

Tabla 3*Variables a analizar*

Fecha Intervención	Nombre Día	Nro. Intervenciones	Nro. Agentes	Nro. PNP	Total de Bienes Fiscalizados
2020-02-03	Lunes	5	9	12	55031
2020-02-04	Martes	4	7	12	220
2020-02-05	Miércoles	6	6	10	24707
2020-02-06	Jueves	6	6	12	37632
2020-02-10	Lunes	6	7	12	29032

Nota. Se puede visualizar las 5 primeras filas de los datos a analizar.

Para realizar el análisis de predicción algunas de las variables deben convertirse en variables numéricas, por ejemplo, el número de días se convertirá en variables numéricas, eso se puede visualizar en la Tabla 4.

Tabla 4*Variables validadas*

Fecha Intervención	Nombre Día	Nro. Intervenciones	Nro. Agentes	Nro. PNP	Total de Bienes Fiscalizados
2020-02-03	1	5	9	12	55031
2020-02-04	2	4	7	12	220
2020-02-05	3	6	6	10	24707
2020-02-06	4	6	6	12	37632
2020-02-10	1	6	7	12	29032

Nota. Se puede visualizar las 5 primeras filas de los datos validados.

El siguiente paso es determinar si se puede aplicar el algoritmo de regresión lineal en la distribución de datos, es decir si las variables tienen una relación correlacional, para ello se crea la matriz de correlación utilizando Pearson, que se presenta en la Tabla 5 (la cual se realiza con comandos Python).

Tabla 5

Matriz de correlación

Variable	Nombre Día	Nro. Intervenciones	Nro. Agentes	Nro. PNP	Total de Bienes Fiscalizados
Nombre día	1	-0.56	-0.77	-0.32	-0.48
Nro. Intervenciones	-0.56	1	0.71	0.32	0.66
Nro. Agentes	-0.77	0.71	1	0.50	0.39
Nro. PNP	-0.332	0.33	0.50	1	0.04
Total bienes fiscalizados	-0.48	0.66	0.39	0.04	1

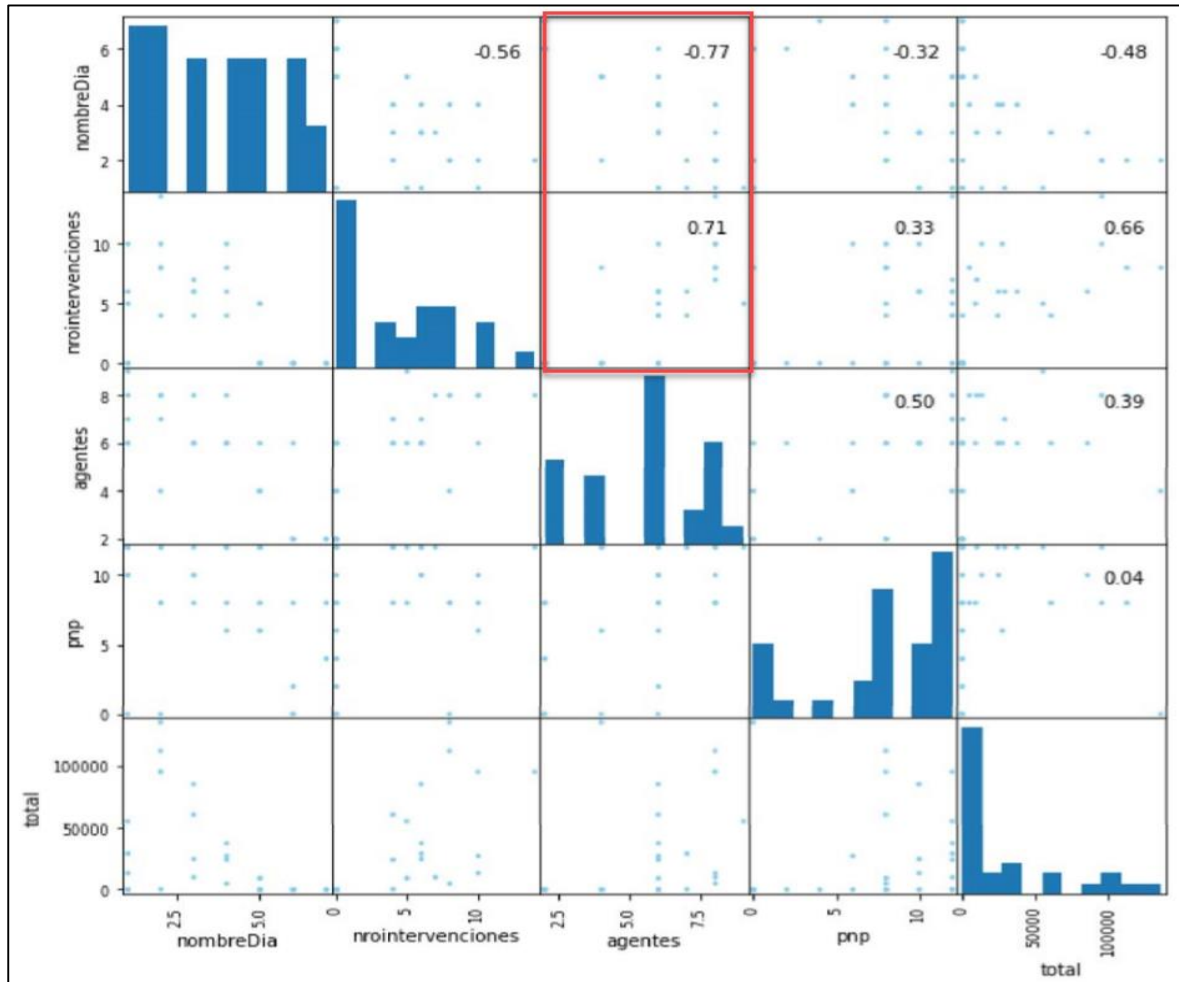
Nota. Se observa la matriz de correlación elaborada con Python.

El objetivo de esa matriz es determinar qué variables son correlativas, por lo que se analiza primero las variables que son correlativas con la variable que se quiere predecir, es decir: con la variable agentes, se busca índices de correlación altos, sean negativos o positivos, lo que se puede apreciar mejor en una matriz de correlación gráfica y eso está en la Figura 18. En la que se puede apreciar que las variables que tienen mayor correlación respecto a la variable de agentes son: el nombre del día con una correlación de -0.77 , lo que indica que existe una correlación negativa, es decir, que mientras más se acerque el fin de semana, el número de agentes tiende a disminuir, pero no es la única variable con correlación alta, también se tiene la variable número intervenciones con 0.77 de correlación positiva, que indica que a mayor

número de intervenciones se necesitará mayor número de agentes, ambas variables están marcadas en la Figura 18 en rectángulos rojos.

Figura 18

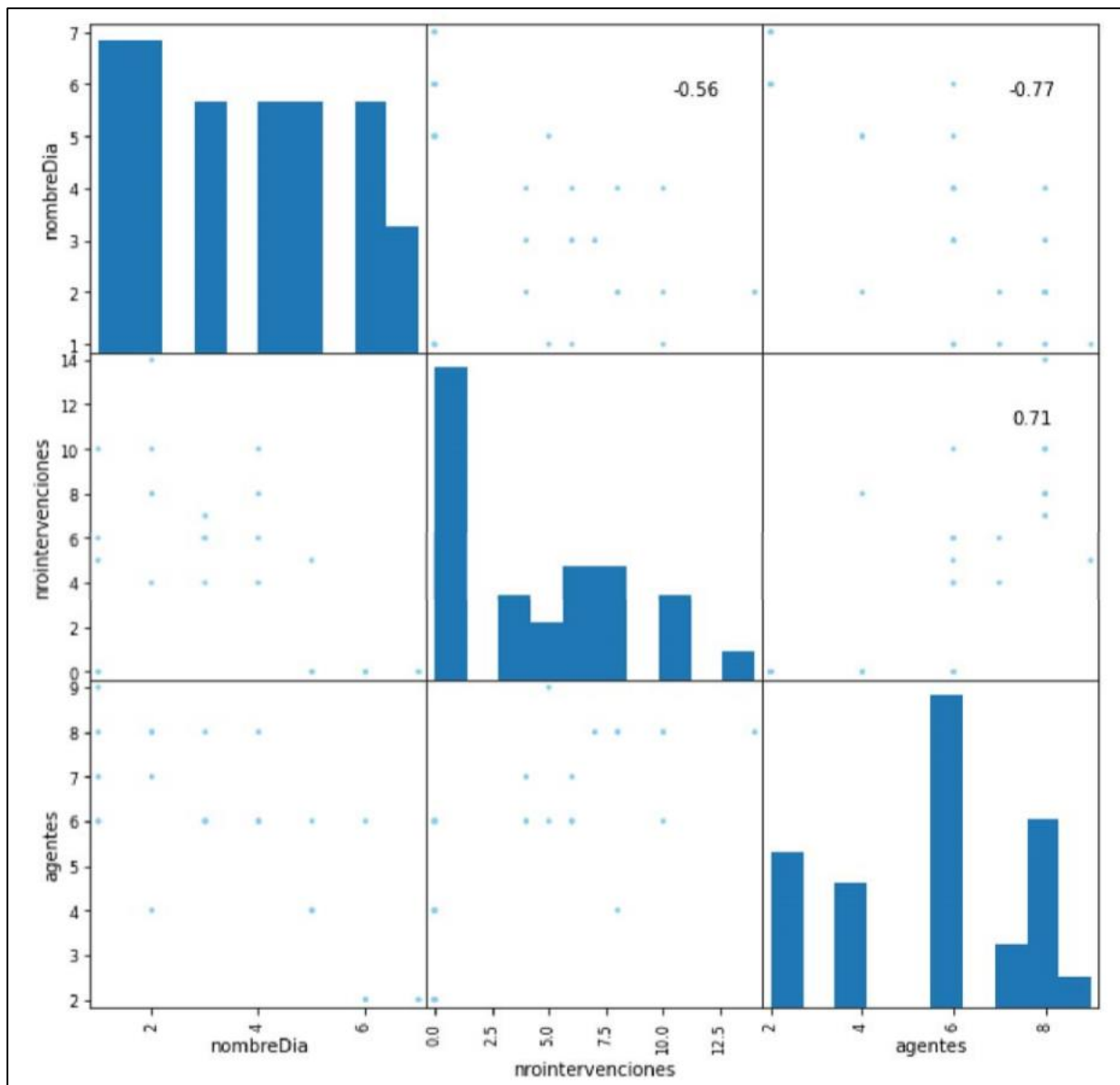
Matriz de correlación gráfica



Nota. Se observa la matriz de correlación gráfica elaborada usando Python.

Las demás variables de la Figura 18 tienen una correlación baja, que indica que no son significativas al momento de determinar el número de agentes necesarios para cada día de la semana.

La Figura 19 se muestra la matriz de correlación con sólo las variables que se utilizaron para realizar el análisis de regresión lineal.

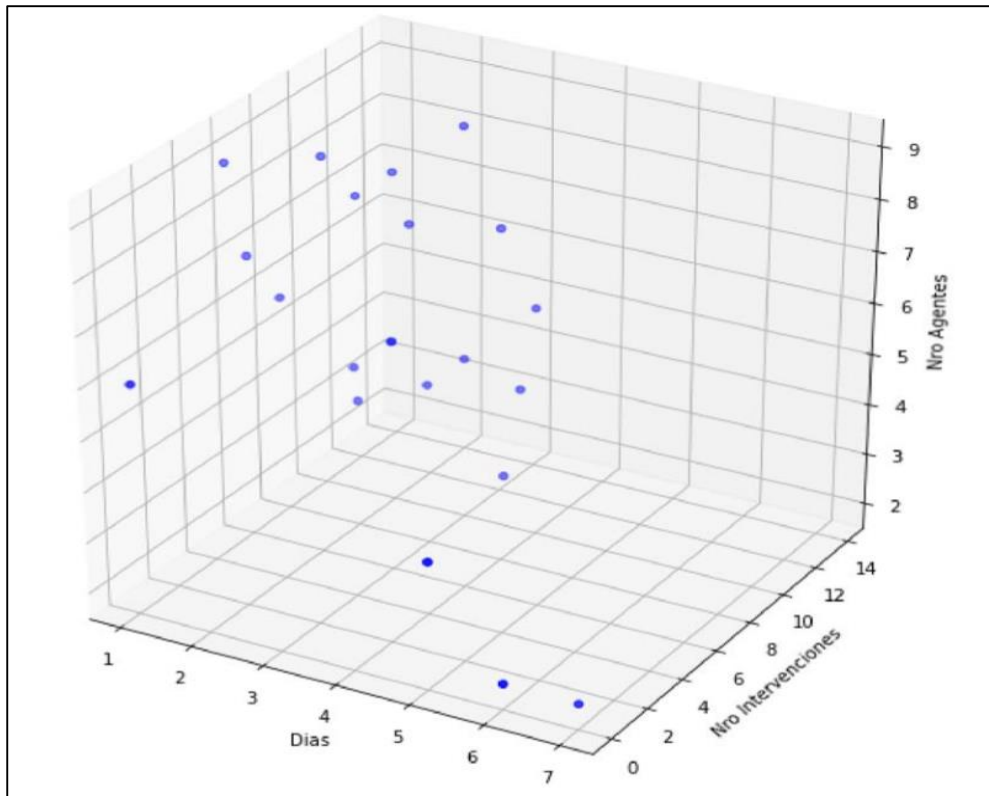
Figura 19*Matriz de correlación resultante*

Nota. Se observa la matriz de correlación resultante elaborada usando Python.

Con las variables predictoras ya definidas, se tiene todas las variables en un gráfico de tres dimensiones para visualizar su correlación y distribución, ese gráfico se encuentra en la Figura 20.

Figura 20

Gráfico inicial de variables



Nota. Se observa el gráfico inicial de variables usando Python.

Se necesita realizar una línea de regresión que indicará la predicción de la variable número de agentes. Para ello la ecuación que se debe determinar debe ser del tipo:

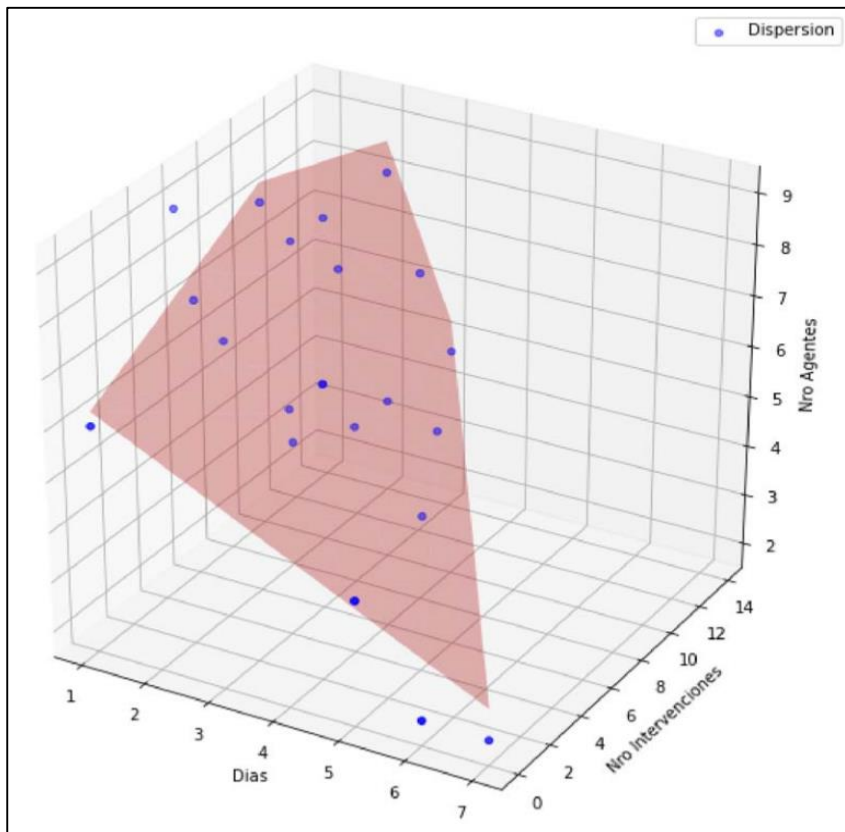
- $nroAgentes = ADias + BNumeroIntervenciones + C$

Donde A, B, C son los coeficientes para cada una de las variables.

Usando las funciones Python (ver Anexo D) se puede determinar que la ecuación es de la siguiente forma:

- $nroAgentes = -0.61Dias + 0.21NumeroIntervenciones + 6.89$

Cuyo gráfico se puede observar en la Figura 21.

Figura 21*Gráfico de regresión lineal*

Nota. Se observa el gráfico de regresión lineal elaborado usando Python.

Con esa ecuación se realiza la predicción usando los promedios de las intervenciones y los números de cada día, que se encuentran en la Tabla 6.

Tabla 6*Promedio de intervenciones*

Nombre Día	Promedio
Lunes	7
Martes	9
Miércoles	6
Jueves	7
Viernes	5

Nombre Día	Promedio
Sábado	3
Domingo	2

Nota. Se observa los promedios de intervenciones por día.

Teniendo como resultado los números correspondientes de los agentes para cada día de la semana como se puede apreciar en la Tabla 7.

Tabla 7

Resultado de regresión lineal

Nombre día	Intervenciones	Agentes
Lunes	7	7
Martes	9	8
Miércoles	6	6
Jueves	7	6
Viernes	5	5
Sábado	3	4
Domingo	2	3

Nota. Se observa los resultados de usar la regresión lineal.

Redondeando el resultado de la Tabla 7 se puede concluir que:

- Lunes 7 agentes
- Martes 8 agentes
- Miércoles 6 agentes
- Jueves 6 agentes
- Viernes 5 agentes
- Sábado 4 agentes

- Domingo 3 agentes

Estos datos indican que durante los fines de semana no hay mucha afluencia de transportes a intervenir, por ello, se puede reducir el personal o en su defecto, planificar alguna intervención móvil, si fuera necesario.

III. Aportes de la investigación

Después de la implementación del proyecto se pueden identificar los siguientes aportes a la institución:

- La organización obtuvo una herramienta de planificación basada en datos existentes, que apoyó al personal de recursos humanos a la toma de decisiones respecto a la mejor asignación de personal por día y puesto de control, que juntamente con su experiencia mejoraron los indicadores de rendimiento y eficiencia de la SUNAT.
- Se implementó una herramienta que ayudó a saber la cantidad exacta de trabajadores necesarios, permitiendo que la SUNAT planifique adecuadamente las convocatorias laborales priorizando el personal realmente necesario, para así poder invertir los recursos en otras áreas que necesitan mayor apoyo de acuerdo con la demanda de trabajo, sobre todo considerando que actualmente todas las instituciones del estado han sido golpeadas por la pandemia del COVID-19
- Se logró analizar y procesar las diversas fuentes de información en materia de fiscalización de los insumos químicos a través de la aplicación del algoritmo de regresión lineal con la herramienta Python.
- La correcta asignación del recurso humano, permitió planificar operativos de fiscalización en zonas diferentes a los puestos de control en días de menor afluencia de vehículos.
- Con esta investigación se logró reducir el desvío ilegal de insumos químicos hacia las zonas VRAEM (Valle de los ríos Apurímac, Ene y Mantaro) ya que los agentes profesionales pueden realizar un control óptimo en el proceso de fiscalización de los vehículos.

IV. Conclusiones

- Se determinó que, a través del uso de un algoritmo matemático (Regresión Lineal) se logra la óptima distribución de los recursos humanos para las intervenciones realizadas en los puestos de control.
- Se implementó una herramienta automatizada para la recolección y análisis de datos en los puestos de control de la SUNAT.
- Se ha demostrado que, los fines de semana, son los días que se requieren menos personal necesitándose solo los días sábados 4 agentes y los días domingos 3 agentes ya que la afluencia de vehículos es menor. En esos días el personal podría descansar o realizar otras actividades propias del puesto.
- Se identificó que, los días que se requiere mayor personal para realizar de manera efectiva la fiscalización de los vehículos son los días lunes y martes con 8 agentes, los días miércoles y jueves se debe contar con 6 agentes y los días viernes con 5 agentes.

V. Recomendaciones

- Este algoritmo está destinado a un solo puesto de control, por lo que se recomienda implementarlo en los otros puestos de control (que tienen su propia distribución de datos), para así poder optimizar la cantidad de recurso humano necesarios en ellos.
- Se recomienda realizar una actualización de datos para el ajuste del algoritmo cada 3 meses ya que las necesidades de personal varían de acuerdo a la estacionalidad actual de cada puesto de control.
- Se recomienda que, los fines de semana se trabaje con la cantidad mínima de personal debido a la poca afluencia de vehículos y además actualizar la proyección cada tres meses para tener mayor precisión en el número de agentes.
- Se recomienda que los días de semana se cuente con mayor cantidad de agentes profesionales de fiscalización en insumos químicos ya que hay mayor afluencia de vehículos.

3. Referencias

- Ahumada E. y Perusquia J. (2016). Inteligencia de negocios: estrategia para el desarrollo de competitividad en empresas de base tecnológica. *Contaduría y Administración*, 61, 127-158.
<https://doi.org/http://dx.doi.org/10.1016/j.cya.2015.09.006>
- Azma F. y Ali Mostafapour M. (2012). Business intelligence as a key strategy for development organizations. (1), 102-106.
<https://doi.org/http://dx.doi.org/10.1016/j.protcy.2012.02.020>
- Bustamante, A., Galvis, E. y Gomez, L. (2016). Perfil de la investigación sobre inteligencia de negocios en América Latina. *UIS Ingenierías*, 15(1), 41-51.
<https://doi.org/10.18273/revuin.v15n1-2016004>
- Cheung, L. (2015). A quantitative correlation coefficient mining method for business intelligence in small and medium enterprises of trading business. *Expert Systems with Applications*, 39(7), 6279-6291.
<https://doi.org/http://dx.doi.org/10.1016/j.eswa.2011.10.021>
- Chung-Kuang, H. (2014). User Acceptance of business intelligence systems in Taiwan's electronics industry. *Social Behavior & Personality: An International Journal*, 42(4), 583-596. <https://doi.org/10.2224/sbp.2014.42.4.583>
- CISCO. (2019). Curso Introducción a Bigdata (2 ed.).
- Código Civil (1984). Decreto Legislativo N° 295. El Peruano.
<https://diariooficial.elperuano.pe/pdf/0090/CODIGOCIVILv03.pdf>
- Comunicaciones SofOs. (2014). Empresas. <http://www.sofoscorp.com/importancia-manejo-datos-una-empresa/>
- Ministerio de Transportes y Comunicaciones (2013). Resolución Ministerial N° 350-2013-MTC/02. El Peruano. chrome-

extension://efaidnbmnnnibpcajpcglclefindmkaj/https://cdn.www.gob.pe/uploads/document/file/358737/1_0_4508.pdf?v=1567804776

Danesh R. y Mohammad H. (2016). A Model for Measuring the Direct and Indirect Impact of Business Intelligence on Organizational Agility with Partial Mediator role of Empowerment (Case Study: Tehran Construction Engineering Organization (TCEO) and ETKA Organization Industries.co). *Procedia - Social and Behavioral Sciences*, 20, 413-421. <https://doi.org/http://dx.doi.org/10.1016/j.sbspro.2016.09.052>

Decreto Supremo 016-2013-EF, (2013). Que modifica Reglamento de Organización y Funciones de la Superintendencia Nacional de Aduanas y de Administración Tributaria. *El Peruano*. <https://busquedas.elperuano.pe/normaslegales/modifican-reglamento-de-organizacion-y-funciones-de-la-super-decreto-supremo-n-016-2013-ef-896145-1/>

Decreto Supremo 268-2019-EF, (2019). Que aprueba las listas de insumos químicos, productos y sus subproductos o derivados que son objeto de control, y definen los bienes fiscalizados considerados de uso doméstico y artesanal, conforme lo establecido en los artículos 5 y 16 del DL 1126. *El Peruano*. <https://busquedas.elperuano.pe/normaslegales/aprueban-las-listas-de-insumos-quimicos-productos-y-sus-sub-decreto-supremo-n-268-2019-ef-1799604-5/>

Decreto Legislativo 1103, (2012). Que establece medidas de control y fiscalización en la distribución, transporte y comercialización de insumos químicos que puedan ser utilizados en la minería ilegal. *El Peruano*. <https://busquedas.elperuano.pe/normaslegales/decreto-legislativo-que-establece-medidas-de-control-y-fisca-decreto-legislativo-n-1103-759548-1/>

Decreto Legislativo 1126, (2012). Que establece medidas de control en los insumos químicos y productos fiscalizados, maquinarias y equipos utilizados para la elaboración de drogas

- ilícitas. El Peruano. <https://busquedas.elperuano.pe/normaslegales/decreto-legislativo-que-establece-medidas-de-control-en-los-decreto-legislativo-n-1126-861637-1/>
- Decreto Legislativo 1339, (2017). Que modifica el Decreto Legislativo N° 1126, sobre medidas de control en los insumos químicos y productos fiscalizados, maquinarias y equipos utilizados para la elaboración de drogas ilícitas. El Peruano. <https://busquedas.elperuano.pe/normaslegales/decreto-legislativo-que-modifica-el-decreto-legislativo-n-1-decreto-legislativo-n-1339-1471014-5/>
- Dlegorreta (2015). Algo sobre Python, análisis de datos y machine learning. <https://dlegorreta.wordpress.com/2015/05/13/algo-sobre-python-analisis-de-datos-y-machine-learning/>
- Murphy, K. (2012). Machine Learning: a Probabilistic Perspective.
- Pearlman, S. (2019). <https://es.talend.com>. <https://es.talend.com/resources/what-is-etl/>
- Ramakrishnan, M. y Sidorova A. (2012). Factors influencing business intelligence (BI) data collection strategies: An empirical investigation. *Decision Support Systems*, 52(2), 486-496. <https://doi.org/http://dx.doi.org/10.1016/j.dss.2011.10.009>
- Resolución de Superintendencia 162-2019/SUNAT, (2019). Que modifica la relación de puestos de control obligatorios. <https://www.sunat.gob.pe/legislacion/superin/2019/162-2019.pdf>
- Rubin, R. (2013). The impact of Business Intelligence systems on stock return volatility. *Information & Management*, 50(2-3), 67-75. <https://doi.org/http://dx.doi.org/10.1016/j.im.2013.01.002>
- Superintendencia Nacional de Aduanas y de Administración Tributaria – SUNAT (2020). Portal Institucional. <http://www.sunat.gob.pe/institucional/quienessomos/index.html>

Trieu, V. (2017). Getting value from Business Intelligence systems: A review and research agenda. *Decision Support Systems*, 93, 111-124.

<https://doi.org/http://dx.doi.org/10.1016/j.dss.2016.09.019>

Universidad Nacional Federico Villareal (2020). Planes Curriculares 2020.

http://web2.unfv.edu.pe/sitio/transparencia_universitaria/informacion_academica/plan_estudio/pregrado/fcnm_mat_p.pdf

Velogig (2020). Blog de Velogig. <https://velogig.com/las-6-mejores-diferencias-entre-big-data-y-machine-learning/>

VI. Anexos

Anexo A - Relación de insumos químicos fiscalizados simples

N°	Insumo químico o producto	Fórmula química	Otras denominaciones
1	Acetato de Etilo	$C_4H_8O_2$	Ester Acético; Ester Etilico del Ácido Acético; Ester Etiloacético; Etanoato de Etilo; Acetidin; Vinagre de nafta.
2	Acetato de n-Propilo	$C_5H_{10}O_2$	Etanoato de Propilo; n-Propil Éster del Ácido Acético; Acetato de Propilo.
3	Acetona	C_3H_6O	Acetonum; Dimetilcetona; Dimethylketone; Propanona; Espíritu Piroacético; 2-Propanona; β -Cetopropana; Éter Piroacético; Ácido Piroacético.
4	Ácido Antranílico	$C_7H_7NO_2$	Ácido Orto Aminobenzoico; 1-Amino-2-Carboxibenceno; Ácido 2-Aminobenzoico; Orto-Carboxianilina.
5	Ácido Clorhídrico y/o Múriatico	HCl	Ácido Hidroclórico; Cloruro de Hidrógeno en solución acuosa.
6	Ácido Fórmico	CH_2O_2	Ácido Metanoico; Ácido Hidroxicarboxílico.
7	Ácido Nítrico	HNO_3	Nitrato de Hidrógeno; Ácido Azótico; Hidróxido Nitrito; Ácido Fumante; Aqua Fortis.
8	Ácido Sulfúrico	H_2SO_4	Ácido Sulfúrico; Oleúm; Ácido Sulfúrico Fumante; Sulfato de Hidrogeno; Aceite de Vitriolo; Ácido Sulfúrico diluido; Ácido Sulfúrico en solución; Ácido Tetraoxosulfúrico (VI); Tetraoxosulfato (VI) de Hidrógeno; Licor de Vitriolo; Espíritu de Vitriolo.
9	Amoniaco	NH_3	Amoniaco Anhidro; Gas Amoniacal; Amoniaco en solución; Solución Amoniacal; Hidróxido de Amonio.
10	Anhídrido Acético	$C_4H_6O_3$	Óxido Acético; Anhídrido del Ácido Acético; Óxido de Acetilo; Anhídrido Etanoico.
11	Benceno	C_6H_6	Benzol; Benzole; Nata de Carbón; Pirabenzol; Naltu Mineral; 1,3,5-Ciclohexatrieno.

12	Carbonato de Sodio	Na_2CO_3	Carbonato bisódico; Sal de Sosa cristalizada; Subcarbonato de Sodio; Carbonato Sódico neutro; Ceniza de Sosa Liviana; Sosa Calcinada; Cenizas de Perla; Carbonato Sódico Anhidro; Sosa del Solvay; Barrilla; Natrón Cristales de Soda.
13	Carbonato de Potasio	K_2CO_3	Sal Tártara; Carbonato Bipotásico; Cenizas de Perla.
14	Cloruro de Amonio	NH_4Cl	Sal de Amoniaco; Sal Amónica; Clorhidrato Amónico.
15	Éter Etilico	$\text{C}_4\text{H}_{10}\text{O}$	Óxido de Etilo; Óxido Dietílico; Éter anestésico; Éter Dietílico; Éter Sulfúrico.
16	Hexano	C_6H_{14}	Hexano Normal; N-Hexano; Hidrido de Caproilo; Hidrido Hexílico; Iso-Hexano y mezcla de isómeros.
17	Hidróxido de Calcio	$\text{Ca}(\text{OH})_2$	Hidrato de Calcio; Cal hidratada; Lechada de Cal Apagada; Cal Muerta.
18	Hipoclorito de Sodio	NaClO	Lejía Hipoclorito Sódico; Agua de Labarraque.
19	Isosafrol	$(\text{CH}_2\text{OO})\text{C}_6\text{H}_3$ $(\text{CH}=\text{CHCH}_3)$	1,2 Metilenodioxi 4-Propenilbenceno; 5-1 Propenil 1,3 Benzodioxol; 1,4-Diacetylbenzene; 1-(4-Acetyl-phenyl)-ethanone.
20	Kerosene	=====	Petróleo Lampante; Aceite Mineral; Kerosina; Keroseno.
21	Metil Etil Cetona	$\text{C}_4\text{H}_8\text{O}$	Methyl Ethyl Ketone; Butanona; 2-Butanona; MEK; Metilpropanona; Etilmetilcetona.
22	Metil Isobutil Cetona	$\text{C}_6\text{H}_{12}\text{O}$	Methyl Isobutyl Ketone; Isopropilacetona; Hexona; 4-Metil-2-Pentanona; MIBK.

23	Óxido de Calcio	CaO	Cal viva; Cal fundente.
24	Permanganato de Potasio	KMnO ₄	Camaleón Mineral; Camaleón Violeta; Permanganato de Potasa; Sal de Potasio del Ácido Permangánico.
25	Piperonal	C ₈ H ₆ O ₃	Heliotropina; 3,4-Metilendioxi-Benzaldehido; Aldehidoperonílico; Carboxaldehyde; Heliotropin.
26	Safrol	CH ₂ OO(C ₆ H ₃) CH ₂ CH=CH ₂	1,2-Metilenodioxi 4-Alilbenceno; 4-Alil- 1,2 Metilenodioxy - Benzol; 5-2 Propenil - 1,3 Benzodioxol.
27	Sulfato de Sodio	Na ₂ SO ₄	En la forma anhidra: Sulfato Sódico Anhidro; Sulfato Sódico Desecado; Torta de Sal; Thenardita (mineral). En la forma hidratada: Sulfato Sódico Decahidratada; Sal de Glauber; Mirabitalita (mineral); Vitriolo de Sosa.
28	Tolueno	C ₇ H ₈	Toluol; Metil Benzol; Hidruro de Cresilo; Fenilmetano; Metilbenceno; Metacida.
29	Xileno	C ₈ H ₁₀	Dimetilbenceno; Xilol; Orto-xileno; Meta-xileno; Para-xileno y Xilenos mixtos.
30	Ácido Sulfámico	HSO ₃ NH ₂	Ácido Sulfamídico o Ácido Amidosulfúrico; Ácido Amidosulfónico; Ácido Amidosulfúrico; Ácido Aminosulfónico.
31	Cloruro de Calcio	CaCl ₂	Cloruro Cálcico; Dicloruro de Calcio; Cloruro de Calcio Anhidro.
32	Hidróxido de Sodio	NaOH	Hidróxido Sódico; Soda Cáustica; Sosa Cáustica; Sosa Lejía; Jabón de piedra; Hidrato de Sodio.
33	Metabisulfito de Sodio	Na ₂ S ₂ O ₅	Pirosulfito Sódico; Disulfito de Sodio.

Anexo B - Relación de insumos químicos fiscalizados compuestos

N°	Insumo químico o producto	Nombre Comercial	Descripción	Familia química
1	Diesel y sus mezclas con Biodiesel	Diesel BX, Diesel BX S50	Mezcla de hidrocarburos y Biodiesel	Mezcla de Hidrocarburos y FAME (Ester Metílico de Ácido Graso)
2	Gasolinas y Gasoholes	Todas las gasolinas y gasoholes	Mezcla de hidrocarburos y Alcohol Carburantes	Mezcla de Hidrocarburo y Alcohol Carburantes
3	Hidrocarburo Alifático Liviano	HAL	Mezcla de hidrocarburos derivados del Petróleo (C ₅ -C ₉)	Condensados del Gas o Gasolina Natural
4	Hidrocarburo Acíclico Saturado	HAS	Mezcla de hidrocarburos derivados del Petróleo (C ₉ -C ₁₂)	Condensados del Gas o Gasolina Natural Corte de nafta virgen compuesto de hidrocarburos acíclicos saturados.
5	Kerosene de aviación Turbo Jet A1	Turbo A1	Mezcla de hidrocarburos derivados de Petróleo	Hidrocarburos
6	Kerosene de aviación Turbo JP5	Turbo JP5	Mezcla de hidrocarburos derivados de Petróleo	Hidrocarburos
7	Solvente N° 1	Solvente 1; Bencina	Mezcla de hidrocarburos derivados del petróleo (C5-C8)	Hidrocarburos
8	Solvente N° 3	Solvente 3; Varsol	Mezcla de hidrocarburos derivados del petróleo (C8-C12)	Hidrocarburos

Anexo C - Relación de zonas de régimen especial

N°	Departamento	Provincia
1	Ayacucho	La Mar, Huanta y Paucartambo
2	Cusco	Paucartambo, Calca, La Convención y Quispicanchi
3	Pasco	Oxapampa
4	Junín	Satipo
5	Puno	Carabaya y Sandia
6	Huánuco	Leoncio Prado, Huamalíes, Puerto Inca, Marañón y Huacaybamba

Anexo D - Análisis de regresión Python

7/6/2020

RegresionLineal (2)

ANÁLISIS DE REGRESIÓN LINEAL

- Parte 1: Leyendo datos de la base de datos sunat.sqlite
- Parte 2: Analizando la correlación de las variables a evaluar
- Parte 3: Realizando la regresión lineal
- Parte 4: Prediciendo resultados

Explicación:

Se desea predecir el número de agentes óptimos para cada día de la semana, para ello se tomará en cuenta el número de intervenciones (se calculará un promedio por día de la semana) y el día de la semana, expresado numéricamente del 1 al 7.

Parte 1: Leyendo datos de la base de datos sunat.sqlite

Vamos a importar las librerías necesarias para poder ejecutar el análisis de regresión lineal correspondiente (instalar esto `pip install ipympl` para poder manejar gráficos interactivos) descomentar

%matplotlib notebook

y ejecutar en jupyter notebook para ver la interactividad

In [39]:

```
import matplotlib.pyplot as plt
from pandas.plotting import scatter_matrix
import numpy as np
import pandas as pd
import sqlite3 as sql
from sklearn import linear_model
```

Creamos la conexión y el cursor a la base de datos sunat.sqlite

In [40]:

```
con=sql.connect('sunat.sqlite')
cur=con.cursor()
```

Leemos la vista que ya existe desde la base de datos sqlite y mostramos los primeros datos de la misma lo almacenamos en el dataframe data el cual luego pasaremos a analizar

7/6/2020

RegresionLineal (2)

In [41]:

```
data=pd.read_sql('select fechaintervencion,nombreDia,nrointervenciones,agentes,pnp,tota
l from AnalisisPersonas where agentes!=0',con)
data.head()
```

Out[41]:

	fehaintervencion	nombreDia	nrointervenciones	agentes	pnp	total
0	2020-02-03	LUNES	5	9	12.0	55031
1	2020-02-04	MARTES	4	7	12.0	220
2	2020-02-05	MIERCOLES	6	6	10.0	24707
3	2020-02-06	JUEVES	6	6	12.0	37632
4	2020-02-10	LUNES	6	7	12.0	29032

El parámetro nombre día no lo podemos analizar en nominal, debemos convertirlo a numérico. Para ello utilizaremos una función lambda y la aplicaremos con el apply, luego mostraremos las primera filas del dataframe data

In [42]:

```
data['nombreDia']=data['nombreDia'].apply(lambda numero: 1 if numero == 'LUNES' else 2
if numero == 'MARTES' else 3 if numero == 'MIERCOLES' else 4 if numero == 'JUEVES' else
5 if numero == 'VIERNES' else 6 if numero == 'SABADO' else 7)
data.head()
```

Out[42]:

	fehaintervencion	nombreDia	nrointervenciones	agentes	pnp	total
0	2020-02-03	1	5	9	12.0	55031
1	2020-02-04	2	4	7	12.0	220
2	2020-02-05	3	6	6	10.0	24707
3	2020-02-06	4	6	6	12.0	37632
4	2020-02-10	1	6	7	12.0	29032

Parte 2: Analizando la correlación de las variables a evaluar

Mostramos una matriz de datos estadísticos (la transponemos) esta matriz específicamente los promedios nos servirá después para poder hacer las predicciones

7/6/2020

RegresionLineal (2)

In [43]:

```
data.describe().transpose()
```

Out[43]:

	count	mean	std	min	25%	50%	75%	max
nombreDia	28.0	3.607143	1.950105	1.0	2.0	3.5	5.00	7.0
nrointervenciones	28.0	4.321429	4.110092	0.0	0.0	4.5	7.25	14.0
agentes	28.0	5.607143	2.166056	2.0	4.0	6.0	7.25	9.0
pnp	28.0	7.928571	4.198135	0.0	6.0	8.0	12.00	12.0
total	28.0	29242.321429	40212.025448	0.0	0.0	9455.0	41981.75	135520.0

Creando la matriz de correlación usando Pearson, nos ayudará a analizar que variables son correlacionales con el número de agentes y esas serán las variables que nos ayudarán a predecir el número óptimo de agentes

In [44]:

```
data.corr('pearson')
```

Out[44]:

	nombreDia	nrointervenciones	agentes	pnp	total
nombreDia	1.000000	-0.561274	-0.774416	-0.320234	-0.480708
nrointervenciones	-0.561274	1.000000	0.709464	0.327646	0.663066
agentes	-0.774416	0.709464	1.000000	0.501847	0.385100
pnp	-0.320234	0.327646	0.501847	1.000000	0.037372
total	-0.480708	0.663066	0.385100	0.037372	1.000000

Como podemos observar en la matriz analizando los agentes con las demás variables tenemos que existen dos variables que tienen una correlación alta mayor a 0.7 y serán las que nos interesen analizar.

Si aun no lo podemos ver claramente a continuación mostraremos la misma matriz pero de manera gráfica donde podemos ver a mayor detalle las variables que tienen mayor correlación.

7/6/2020

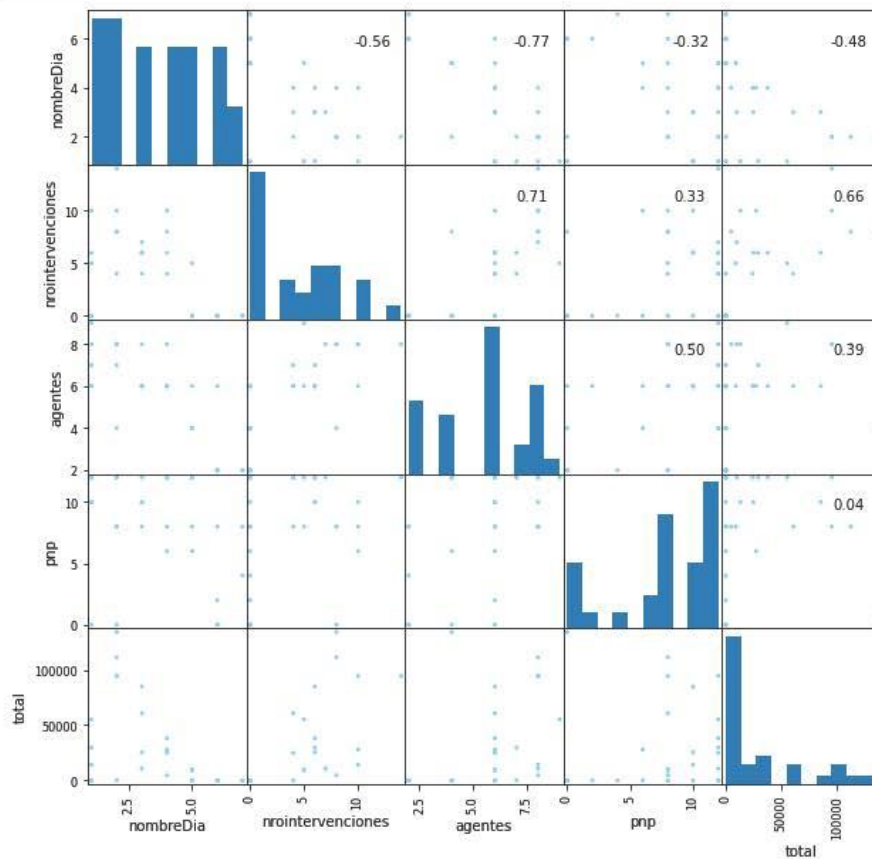
RegresionLineal (2)

In [45]:

```

axes = scatter_matrix(data, alpha=1, figsize=(10,10), color="skyblue")
corr = data.corr("pearson").values
for i, j in zip(*plt.np.triu_indices_from(axes, k=1)):
    axes[i, j].annotate("%.2F" %corr[i,j], (0.8, 0.8), xycoords='axes fraction', ha='center', va='center')
plt.savefig("Proyectomatrix_scatter.png")
#ruta imagen guardada
plt.show()

```



Como podemos ver en la matriz las variables que más correlación tienen con el nro de agentes que es la variable que nos interesa predecir son nombredia y nrointervenciones por lo que solo tomaremos en cuenta dos variables predictoras que ayudarán a determinar el número de agentes que se necesita según el día. eliminaremos las demás columnas y mostramos como quedó nuestro dataframe.

7/6/2020

RegresionLineal (2)

In [46]:

```
data.drop(['pnp', 'total'],axis=1,inplace=True)
data.head()
```

Out[46]:

	fechaintervencion	nombreDia	nrointervenciones	agentes
0	2020-02-03	1	5	9
1	2020-02-04	2	4	7
2	2020-02-05	3	6	6
3	2020-02-06	4	6	6
4	2020-02-10	1	6	7

Mostramos el gráfico nuevamente con solo las columnas solo debemos fijarnos en la correlación de agentes con las otras dos variables

- agentes-nombredia correlación negativa (-0.77)
- agentes-nrointervenciones correlación positiva (0.71)

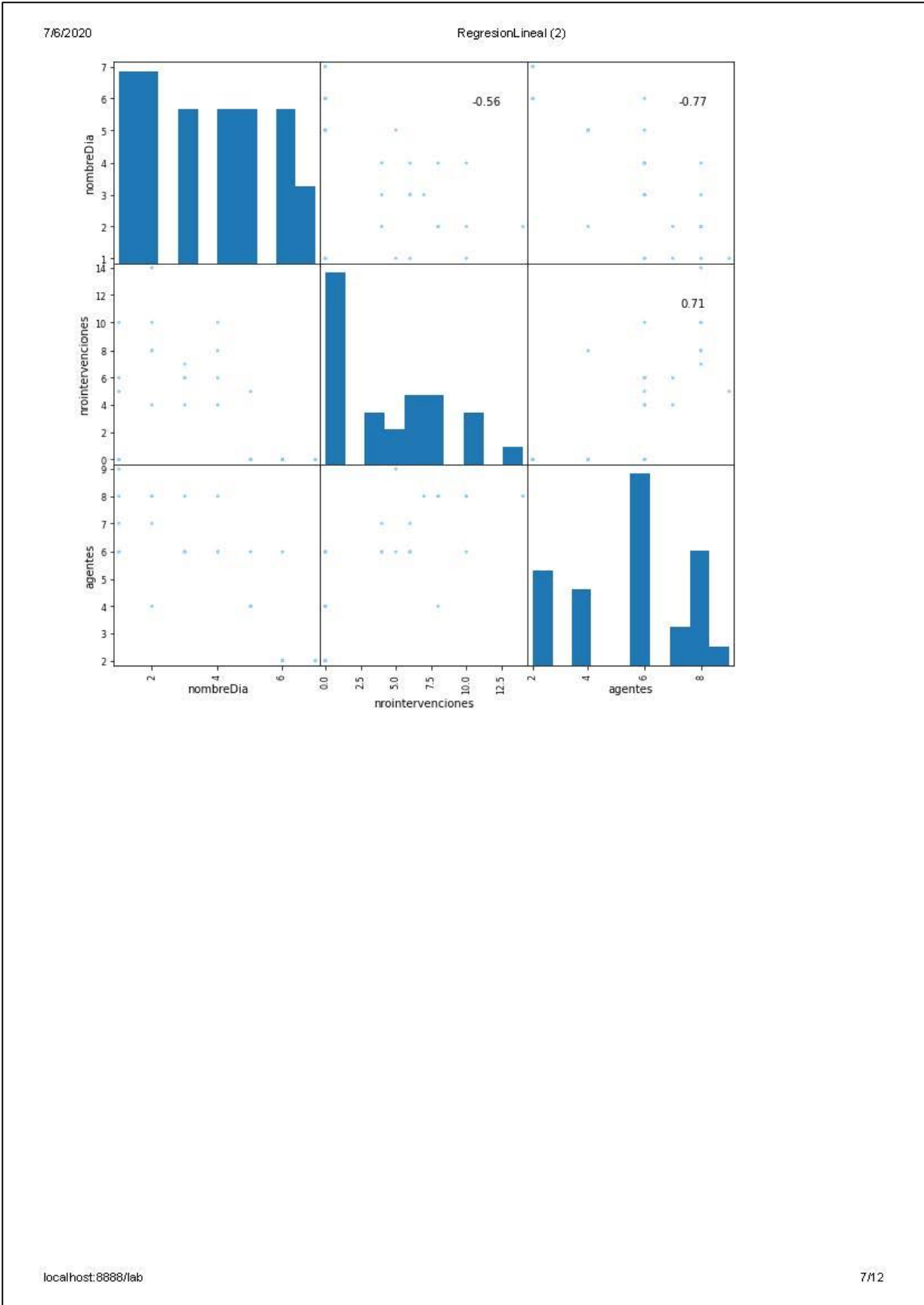
matriz de correlacion grafica

7/6/2020

RegresionLineal (2)

In [47]:

```
axes = scatter_matrix(data, alpha=1, figsize=(10,10), color="skyblue")
corr = data.corr('pearson').values
for i, j in zip(*plt.np.triu_indices_from(axes, k=1)):
    axes[i, j].annotate("%.2f" %corr[i,j], (0.8, 0.8), xycoords='axes fraction', ha='center', va='center')
plt.savefig("Proyectomatrix_scatter.png" )
plt.show()
```



7/6/2020

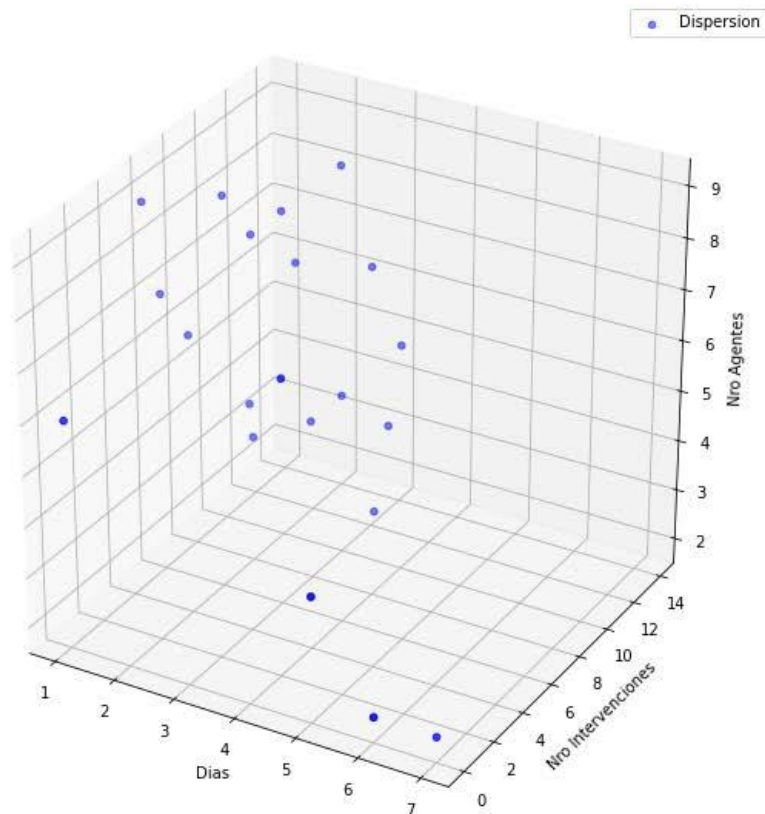
RegresionLineal (2)

Parte 3: Realizando la regresión lineal

Ahora empezaremos a construir la regresión lineal múltiple, para ello primero graficaremos las dos variables esto nos ayudará luego para mostrar la línea de regresión lineal, como son varias variables independientes (2 variables independientes) necesitaremos hacerlo en un plano de 3D

In [48]:

```
fig = plt.figure(figsize=(10, 10))
ax = fig.add_subplot(111, projection='3d')
ax.scatter(data.nombreDia, data.nrointervenciones, data.agentes, c='Blue', marker='o', label='Dispersion', alpha=0.5)
ax.set_xlabel('Dias')
ax.set_ylabel('Nro Intervenciones')
ax.set_zlabel('Nro Agentes')
plt.legend()
plt.show()
```



En base a ello aplicaremos la regresión lineal teniendo como variables independientes a nombre, día y número de intervenciones, mientras que la variable a predecir es la de número de agentes

7/6/2020

RegresionLineal (2)

In [49]:

```
reg=linear_model.LinearRegression()
reg.fit(data[['nombreDia', 'nrointervenciones']], data.agentes)
```

Out[49]:

```
LinearRegression()
```

Después de aplicar la regresión lineal ya podemos ver los coeficientes generados

In [50]:

```
reg.coef_
```

Out[50]:

```
array([-0.61006025,  0.21143098])
```

Con ello ya podríamos hacer predicciones con los valores pre existentes como podemos apreciar en el código

In [51]:

```
Agentes_predi=reg.predict(data[['nombreDia', 'nrointervenciones']])
Agentes_predi
```

Out[51]:

```
array([7.34112811, 6.51963688, 6.33243859, 5.72237834, 7.5525591 ,
       7.36536081, 6.54386957, 6.56810227, 8.39828303, 8.6339467 ,
       6.33243859, 5.29951637, 7.78822277, 5.90957662, 6.1452403 ,
       4.9008871 , 7.36536081, 3.23367193, 2.62361168, 3.84373219,
       3.23367193, 2.62361168, 3.84373219, 3.23367193, 3.84373219,
       3.23367193, 6.2839732 , 6.2839732 ])
```

Además también podemos crear la ecuación de predicción, en base a los coeficientes y al punto de intercepción que nos servirá como la constante de la ecuación

In [52]:

```
a=reg.coef_[0]
b=reg.coef_[1]
i=reg.intercept_

print('Ecuacion de regresion lineal')
print('NroAgentes=', "%.2f" %a, 'nombreDia+', "%.2f" %b, 'nrointervenciones+', "%.2f" %i )
```

```
Ecuacion de regresion lineal
```

```
NroAgentes= -0.61 nombreDia+ 0.21 nrointervenciones+ 6.89
```

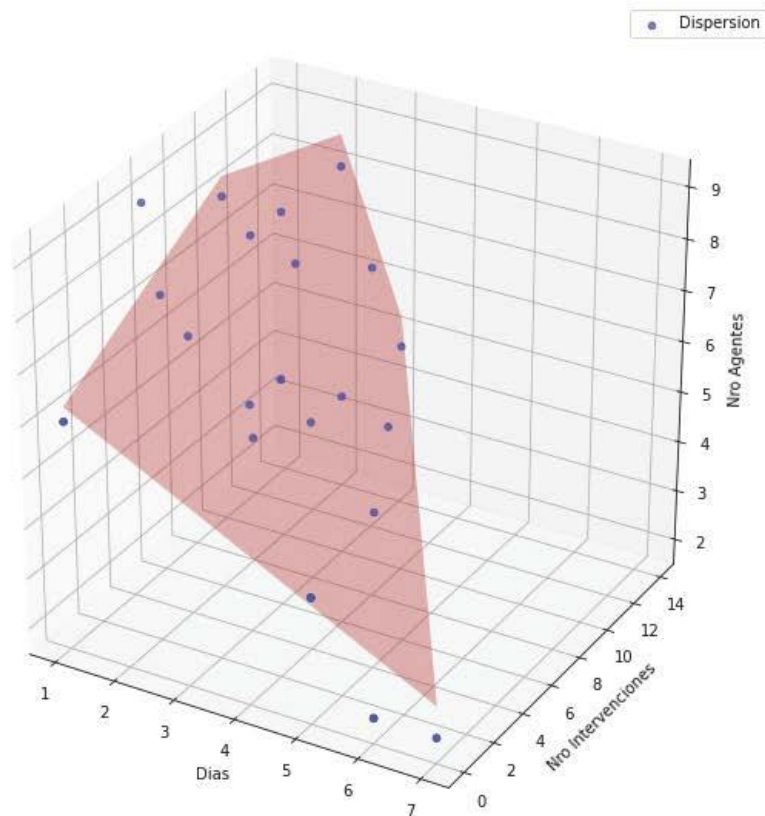
Bien ahora si podemos hacer nuevamente la gráfica y ver la línea de regresión que se forma

7/6/2020

Regresión Lineal (2)

In [53]:

```
fig = plt.figure(figsize=(10, 10))
ax = fig.add_subplot(111, projection='3d')
ax.scatter(data.nombreDia, data.nrointervenciones, data.agentes, c='Blue', marker='o', label='Dispersion', alpha=0.5)
ax.plot_trisurf(data.nombreDia, data.nrointervenciones, Agentes_predi, color='Red', alpha=0.3)
ax.set_xlabel('Días')
ax.set_ylabel('Nro Intervenciones')
ax.set_zlabel('Nro Agentes')
plt.legend()
plt.show()
```



Parte 4: Prediciendo resultados

Necesitamos obtener dos variables el día y el promedio de intervenciones por día, eso lo sacaremos desde a base de datos sunat.sqlite, además se irá actualizando a medida que se vayan cargando los datos mediante nuestro proceso ETL creado en la primera parte del trabajo. y se irán prediciendo con frecuencia mensual, puesto que el plan de trabajo se hace mensual en la sunat.

Crearemos un dataframe que contenga los datos que necesitamos para predecir el número óptimo de agentes y mostraremos los datos cargados

7/6/2020

RegresionLineal (2)

In [54]:

```
datos_predecir=pd.read_sql('select nombreDia,avg(nrointervenciones) prome from AnalisisPersonas group by nombreDia',con)
datos_predecir.head(10)
```

Out[54]:

Nombre día	Promedio
Lunes	7
Martes	9
Miércoles	6
Jueves	7
Viernes	5
Sábado	3
Domingo	2

Nuevamente el nombre de los días lo debemos pasar a numérico para poder usarlos en la ecuación

In [55]:

```
datos_predecir['nombreDia']=datos_predecir['nombreDia'].apply(lambda numero: 1 if numero == 'LUNES' else 2 if numero == 'MARTES' else 3 if numero == 'MIERCOLES' else 4 if numero == 'JUEVES' else 5 if numero == 'VIERNES' else 6 if numero == 'SABADO' else 7)
datos_predecir.sort_values(by=['nombreDia'])
```

Out[55]:

Nombre día	Promedio
1	7
2	9
3	6
4	7
5	5
6	3
7	2

Ahora sí con los valores ya establecidos, predecimos los agentes óptimos por cada día de la semana

7/6/2020

RegresionLineal (2)

In [112]:

```
df = pd.DataFrame(columns=['Dia', 'Intervenciones', 'Agentes'])

for fila in datos_predecir.sort_values(by=['nombreDia']).index:
    x1=datos_predecir.loc[fila,'nombreDia']
    x2=datos_predecir.loc[fila,'prome']
    X=[[datos_predecir.loc[fila,'nombreDia'],datos_predecir.loc[fila,'prome']]
    Y=reg.predict(X)
    df = df.append({'Dia': x1,'Intervenciones':x2,'Agentes':Y[0]}, ignore_index=True)

df
```

Out[112]:

Nombre día	Intervenciones	Agentes
Lunes	7	7
Martes	9	8
Miércoles	6	6
Jueves	7	6
Viernes	5	5
Sábado	3	4
Domingo	2	3

Como podemos ver el cuadro anterior podemos apreciar que la cantidad de elementos óptimos para cada día de la semana es:

- Lunes 7 agentes
- Martes 8 agentes
- Miércoles 6 agentes
- Jueves 6 agentes
- Viernes 5 agentes
- Sábado 4 agentes
- Domingo 3 agentes

In []: