



Universidad Nacional
SAN LUIS GONZAGA



Atribución-NoComercial-SinDerivadas 4.0 Internacional

Esta licencia es la más restrictiva de las seis licencias principales Creative Commons, permitiendo a otras solo descargar sus obras y compartirlas con otras siempre y cuando den crédito, pero no pueden cambiarlas de forma alguna ni usarlas de forma comercial.

<http://creativecommons.org/licenses/by-nc-nd/4.0>



UNIVERSIDAD NACIONAL "SAN LUIS GONZAGA"
FACULTAD DE INGENIERÍA DE SISTEMAS
DIRECCIÓN DE INVESTIGACIÓN
EVALUACIÓN DE ORIGINALIDAD



CONSTANCIA

El que suscribe, deja constancia que se ha realizado el análisis con el software de verificación de similitud al documento cuyo título es:

INTERVENCION DE LA ANALÍTICA DE DATOS EN EL ANÁLISIS DE LA VACUNACIÓN COVID-19 EN LA PROVINCIA ICA.

Presentado por:

- **AROSTIGUE SIANCAS KATHERINE CECIL**

BACHILLER en **PREGRADO** de la facultad de Ingeniería de Sistemas. El resultado obtenido es (**porcentaje de similitud 2%**) por el cual se otorga el calificativo de:

APROBADO, según el Reglamento de Evaluación de la Originalidad.

Se adjunta al presente el reporte de evaluación con el software de verificación de originalidad.

Ica, 31 de julio de 2024


Dr. JAVIER ORLANDO GUTIÉRREZ FERREYRA
Director de la Unidad de Investigación
Facultad de Ingeniería de Sistemas

UNIVERSIDAD NACIONAL “SAN LUIS GONZAGA”

VICERRECTORADO DE INVESTIGACIÓN

FACULTAD INGENIERÍA DE SISTEMAS



INTERVENCION DE LA ANALÍTICA DE DATOS EN EL ANÁLISIS DE
LA VACUNACIÓN COVID-19 EN LA PROVINCIA ICA.

Línea de investigación: Ciencias naturales, ingeniería y tecnologías sostenibles

INFORME FINAL DE TESIS, PARA OPTAR EL TITULO
PROFESIONAL DE INGENIERO DE SISTEMAS

Autor: KATHERINE CECIL AROSTIGUE SIANCAS

Ica, Perú

2023

DEDICATORIA

Me gustaría dedicar mi tesis a mis padres y hermana, quien ha sido nuestros pilares para seguir adelante, es para nosotros una gran satisfacción poder dedicarles a ellos, que con mucho esfuerzo, esmero y trabajo lo hemos ganado.

A mi asesor Ing. Erwin Peñas Casas por su apoyo constante en guiarme a poder culminar con éxito mi tesis, siempre ha sido un excelente docente en la universidad, con sus consejos y enseñanza.

Para mi hijo Adrián Howard, que es el pilar y el motor que tengo para poder salir adelante.

AGRADECIMIENTO

Quisiera agradecer primero a Dios por darme salud, bienestar y ganar para seguir adelante con el propósito de ser una mejor persona.

Luego a mi Asesor ING. Erwin Peña Casas, por su esfuerzo y dedicación, sus orientaciones, su manera de trabajar, su paciencia y motivación han sido fundamentales para mi formación como investigadora.

También me gustaría agradecer a mis padres y hermana: Ambrosio Arostigue (que desde cielo me guía y me cuida), mi madre Cecilia Siancas y Jesús Arostigue S. Que siempre ha estado apoyándome en las buenas y en las malas, en cada tropiezo y logro, con su amor, guía y consejo ha estado ahí en todo momento.

A mi hermana Rosario Arostigue S. que siempre está pendiente de mí, con sus consejos sabios, su gran ejemplo como persona y profesional me ayudado a ser mejor persona.

A mi esposo Roni Adrián Peve, por su apoyo en todo el proceso de la universidad, y también un aliento para que pueda seguir adelante y pueda culminar mi tesis.

A mi hijo, Adrián Howard P. por ser una bendición y por él salir adelante para poder darle una mejor calidad de vida, darle amor y protección.

PARA ELLOS,

MUCHAS GRACIAS POR TODO.

INDICE DE CONTENIDOS

Dedicatoria	ii
Agradecimientos	iii
Índice	iv
Índice de contenidos	iv
Índice de tablas	v
Índice de figuras	vi
RESUMEN	viii
ABSTRACT	ix
I. INTRODUCCIÓN	1
1.1 Antecedentes	2
1.2 Realidad problemática	8
1.3 Justificación	11
II. ESTRATEGIA METODOLÓGICA	12
2.1. Diseño metodológico	12
2.2. Población y muestra	12
2.3. Material	12
2.4. Método	13
2.4.1. Obtención de los datos	14
2.4.2. Preparación de los datos	17
2.4.3. Analítica de los datos, comunicación y visualización	26
III. RESULTADOS	48
3.1. Análisis de Frecuencias de las variables categóricas	48
3.2. Análisis descriptivo de las variables numéricas	52
IV. DISCUSIÓN	53
V. CONCLUSIONES	54
VI. RECOMENDACIONES	55
VII. REFERENCIAS BIBLIOGRÁFICAS	56
VIII. ANEXOS	58
Anexo 01: Incremento de muertes en la Perú	58

Índice de tablas

TABLA I: EVOLUCIÓN DEL CORONAVIRUS DICIEMBRE 2022 – ENERO 2023	8
TABLA II: DICCIONARIO DE LOS DATOS	16
TABLA III: FRECUENCIA DE VACUNAS SEGÚN FABRICANTE	40
TABLA IV: FRECUENCIA DE VACUNAS SEGÚN SEXO	41
TABLA V: FRECUENCIA DE VACUNAS POR PROVINCIAS	41
TABLA VI: FRECUENCIA DE VACUNAS POR DISTRITOS	42
TABLA VII: FRECUENCIA DE VACUNAS SEGÚN GRUPO DE RIESGO	43
TABLA VIII: MEDIDAS DE TENDENCIA CENTRAL	44
TABLA IX: MEDIDAS DE DISPERSIÓN	44
TABLA X: INTERVALOS DE CONFIANZA PARA LOS DATOS (95%)	44
TABLA XI: INCREMENTO DE MUERTES EN EL PERU	50

Índice de figuras

Fig. 1. Estado actual del coronavirus	2
Fig. 2. Evolución del coronavirus en el Perú	10
Fig. 3. Método para el análisis de los datos	13
Fig. 4. Portal de los datos abiertos de la vacunación del MINSA	14
Fig. 5. Archivos descargados del MINSA	15
Fig. 6. Intento de carga de archivo de datos totales de vacunación	15
Fig. 7. Archivo cargado, este abre solo hasta la información posible de carga del Excel	16
Fig. 8. Datos iniciales del Dataset	18
Fig. 9. Datos finales del Dataset	19
Fig.10. Datos iniciales del Dataset vacunaIca	20
Fig. 11. Datos finales del Dataset vacunaIca	20
Fig. 12. Resultados descriptivos	21
Fig. 13. Estadística descriptiva de los datos numéricos	22
Fig. 14. Datos del Dataset	23
Fig. 15. Dataset con las variables para análisis	23
Fig. 16. Set de datos limpios, no existen valores perdidos o vacíos	24
Fig. 17. Extracción de los datos en R	24
Fig. 18. Datos iniciales del Dataset final	25
Fig. 19. Datos finales del Dataset final	26
Fig. 20. Instalación de la librería sweetviz	27
Fig. 21. Reporte analítico creado	27
Fig. 22 Archivo HTML generado por la librería	28
Fig. 23. Descarga del archivo HTML	28
Fig. 24. Reporte generado por la librería Sweetviz	29
Fig. 25. Análisis de frecuencia de las dosis por cada provincia	30
Fig. 26. Vacunación en los distritos de la región	31
Fig. 27. Vacunación según los grupos de riesgo	32
Fig. 28. Vacunación según las dosis aplicadas	33
Fig. 29. Vacunación según el fabricante	34
Fig. 30. Vacunación según el tipo de edad	35
Fig. 31. Dosis aplicadas por fabricante en cada provincia	35
Fig. 32. Dosis aplicadas por fabricante según el sexo	36
Fig. 33. Dosis aplicadas por fabricante en los distritos	38

Fig. 34. Dosis máximas y mínimas aplicadas por sexo en cada provincia	38
Fig. 35. Dosis aplicadas por fabricante según grupos de riesgo	40
Fig. 36. Evolución de la vacunación	41
Fig. 37. Comportamiento de vacunación en la provincia de Ica	41
Fig. 38. Comportamiento de vacunación provincia de Chincha	42
Fig. 39. Comportamiento de vacunación provincia de Pisco	42
Fig. 40. Comportamiento de vacunación provincia de Palpa	43
Fig. 41. Comportamiento de vacunación provincia de Nazca	43
Fig. 42. Modelo de Widgets Orange	44
Fig. 43. Estructura del <i>widget Pivote Table</i>	45
Fig. 44. Configuración de la variable de tiempo	46
Fig. 45. Conteo de los datos de las dosis	46
Fig. 46. Selección de las vacunaciones en la provincia de Ica	47

RESUMEN

El presente trabajo de tesis tuvo como objetivo Evaluar la INTERVENCION DE LA ANALÍTICA DE DATOS EN EL ANÁLISIS DE LA VACUNACIÓN COVID-19 EN LA PROVINCIA ICA, la misma que fue ampliado el análisis a toda la región por existir los datos abiertos del MINSA. La metodología del tipo aplicada tecnológica, del nivel descriptivo, con un enfoque cuantitativo, del tipo retrospectivo, la muestra estuvo conformada por todos los datos extraídos del portal del MINSA y filtrados para la región Ica con un total de 1048,554 registros; el archivo original fue reducido por medio de la herramienta de RStudio por ser demasiado grande; el procesamiento requirió del uso de la plataforma de Google colaboratory, con el lenguaje Python y las librerías de Pandas, Numpy, Sweetviz para el análisis básico de los datos. se utilizó igualmente para el análisis el software para minería de datos Orange Datamining. El método aplicado consistió en la extracción de los datos del portal del MINSA, reducción y filtrado de los datos con software RStudio, preparación y análisis de los datos en la plataforma de Google colaboratory y Orange Data Mining. Los resultados muestran que la población mayormente vacunada está en el rango de edades de 20 a 59 años con cerca del 70%, y las vacunas mayormente empleadas fueron de los fabricantes Pfizer y Sinopharm con más del 80% y la población mayormente vacunada se encuentra en la provincia de Ica con 48%. Existe una población similar en vacunación según el sexo con aproximadamente 50%. Se concluye que la analítica de datos ha sido positiva en obtener información valiosa sobre la vacunación del COVID-19.

Palabras clave: Analítica de datos, COVID-19, Librerías, Orange Data Mining

ABSTRACT

The objective of this thesis work was to evaluate the INTERVENTION OF DATA ANALYTICS IN THE ANALYSIS OF COVID-19 VACCINATION IN THE PROVINCE OF ICA, which was extended to the whole region due to the existence of MINSA's open data. The methodology of the applied technological type, descriptive level, with a quantitative approach, retrospective type, the sample consisted of all data extracted from the MINSA portal and filtered for the Ica region with a total of 1048,554 records; the original file was reduced through the RStudio tool for being too large; the processing required the use of the Google Collaboratory platform, with the Python language and the Pandas, Numpy, Sweetviz libraries for the basic analysis of the data. The data mining software Orange Datamining was also used for the analysis. The method applied consisted of extracting data from the MINSA portal, reducing and filtering the data with RStudio software, preparing and analyzing the data in the Google Collaboratory and Orange Data Mining platform. The results show that the population mostly vaccinated is in the age range of 20 to 59 years with about 70%, and the vaccines mostly used were of the manufacturers Pfizer and Sinopharm with more than 80% and the population mostly vaccinated is located in the province of Ica with 48%. There is a similar vaccination population according to sex with approximately 50%. It is concluded that the data analysis has been positive in obtaining valuable information on COVID-19 vaccination.

Keywords: Data analytics, COVID-19, Libraries, Orange Data Mining

I. INTRODUCCIÓN

Las variedades del COVID-19 o CoV-2 del SRAS siguen mutando [1], originando brotes de la enfermedad y obligando a los gobiernos a tomar medidas de prevención y aplicación de nuevas vacunas como la vacuna bivalente. El 5 de mayo de presente año, el director de la Organización Mundial de la Salud (OMS) de la ONU, Tedros Adhanom Gebreyesus, ha señalado que el coronavirus ya no es una emergencia pública internacional, esto no representa que el VIRUS deje de ser una amenaza para la salud de la población mundial [2]. En tal sentido en el estudio se aplica la analítica de datos para poder conocer sobre la evolución de la vacunación y cuáles son las características de dicha vacunación, aprovechando de la disponibilidad de los datos de manera abierta puesto a disposición por el Ministerio de Salud (MINS) del Perú. La analítica de datos se aplicó a los datos de la población de la región de Ica, evaluando el tipo de vacuna según el fabricante, y la tasa de vacunación realizado en los diversos distritos de la región.

La analítica de datos (DA) es el proceso empleado para analizar datos y encontrar tendencias con la finalidad de obtener información de lo que contienen. Para obtener esta información se apoya en software especializados, con herramientas y técnicas de análisis de los datos, usados mayormente por las empresas para poder tomar decisiones más informadas. También es empleada por investigadores para verificar o rechazar modelos, teorías e hipótesis [3]. De otro lado, “los coronavirus (CoV) son una gran familia de virus que suelen causar enfermedades respiratorias, desde el resfriado común hasta el síndrome respiratorio agudo severo (SRAS)” [2]. Siendo la salud de las personas el bien más preciado, es de que realizo este estudio con la finalidad de aportar en el conocimiento sobre el estado de la vacunación en nuestra región y que sirva de referencia a las instituciones de salud, profesionales e investigadores para seguir aportando en el conocimiento sobre las vacunaciones. Además, es de importancia porque se siguen presentando miles de casos nuevos en el mundo como lo reporta la OMS y que se muestra en la Fig. 1.



Fig. 1. Estado actual del coronavirus

Fuente: OMS - <https://covid19.who.int> – 26/09/2023

En la Fig.1, se muestra que en nuestro país se confirmaron 4'518,007 casos, de los cuales fallecieron a causa de esta enfermedad 221,674 ciudadanos. A la fecha de acceso se tiene que se han reportado 13,511 nuevos casos en los últimos 7 días, de ello se deduce la importancia de realizar todos los estudios sobre la vacunación de las personas.

1.1 Antecedentes

Antecedentes internacionales

En el estudio en el que se promueve la vacunación en Corea del Sur [4], debido a las dificultades que le permitan aumentar la tasa de vacunas, por el efecto negativo de estas además de las políticas de la vacunación. Se investigó los sentimientos y conocimientos de la población hacia las vacunas. Se realizó un estudio exploratorio de la evolución de los informes públicos en el tiempo, y se realizó una comparación con otros países. El estudio utilizó las vacunas Moderna, AstraZeneca, Jansen y Pfizer, y algunas palabras como Google, Twiter, Daum, Nave en el recojo de los datos sobre las variables de sentimientos y conocimientos relacionados con la vacuna y las políticas de vacunación del estado. Los resultados evidenciaron que los surcoreanos con relación a la vacunación fue de neutral a negativa y de negativa a positiva en los dos años anteriores. Este comportamiento hacia positivo de la opinión pública se dio a las expectativas esperanzadoras y a la tasa de evaluación en crecimiento. Los resultados con relación a otros países fueron similares.

Según [1] El virus del CoV-2 del SRAS siguen creciendo, y creando brotes de la enfermedad, demorando o parando la apertura en la sociedad y las economías de los países. Las altas tasas de vacunación en algunos países han registrado igualmente la reducción significativa de enfermedades graves, hospitalización incluida la muerte. Lamentablemente la disponibilidad de vacunas no es igual a nivel mundial, donde se cuenta entre el 1% hasta el 70%, dependiendo de la economía del país. El trabajo se enfoca en el análisis de los datos y algoritmos de predicción y su impacto e intensidad en la predisposición en el mundo de la vacuna COVID-19 comparada con Malasia. El análisis de los resultados de vacunación con otros países miembros del G7 (Francia, Alemania, Japón, Italia, Estados Unidos, Reino Unido y Canadá), igualmente también se comparó la vacunación con países de Europa Sudoriental, el análisis de da en las variables: tipo de vacuna, y tasas de vacunación. En este camino, el estudio tuvo como objetivo el modelo de predicción, para el pronóstico de futuras tendencias en vacunaciones en Malasia.

En el estudio de [5], las vacunas incorporan un medio prometedor para reducir la COVID-19; a pesar de ello esta vacunación muestra retos importantes como la actitud hacia el nuevo biológico, en el estudio se tuvo como objetivo saber cuál es la actitud de los ciudadanos del estado de Guerrero (México), para el control del COVID-19 por medio de la vacunación, la investigación del tipo descriptivo, cuantitativa no experimental, de corte transversal, en la que la muestra estuvo conformada por 296 personas de diversos lugares del estado de Guerrero mayores de 18 años y que dieron su aceptación para el estudio. Se empleo un formulario con Google Forms para la recolección de los datos. El formulario presentó dos secciones: con 25 preguntas en escala de Likert sobre datos sociodemográficos. Los resultados arrojan que la actitud hacia la vacuna es negativa en 52,7%. En cuanto a la dimensión de conducta se puede considerar de aceptación hacia la vacunación contra la COVID-19.

De otro lado en el trabajo de [6] sobre la trombosis venosa y la trombocitopenia ocasionada posterior a la vacunación de la vacuna COVID-19 dada por el adenovirus ChAdOx1 nCov-19 de la vacuna AstraZeneca con anticuerpos séricos se han relacionado hacia el factor plaquetario 4 (PF4) complejos polianión. Se reporta un caso de este síndrome inducida por la vacuna de AstraZeneca. Se evidencian descubrimiento de imágenes y laboratorio para las decisiones de tratamiento de este caso específico.

En cuanto a las reacciones de las vacunas, en el estudio de [7] para el caso de la vacunación para el COVID-19 se han dado reacciones adversas. La investigación tuvo como objetivo determinar estas reacciones adversas que se dan posterior a la vacunación contra el COVID-19 en los pobladores adultos. El estudio cualitativo de revisión sistemática de estudios con un

diseño fenomenológico se basó en el método prisma. Se realizó la indagación en diversas bases de datos como Google Scholar, Science Direct, y otras relacionadas con el tema. Los resultados arrojan que las vacunas más utilizadas en estas investigaciones con Pfizer, AstraZeneca y Sinovac. En el caso de la vacuna AstraZeneca, presento reacciones como fatiga (75.9%), fiebre (41%) y dolor articular (48%). Para el caso de la vacuna Pfizer se han presentado reacciones como fatiga (59%), dolor en la zona de inyección. Por lo tanto, de la revisión se ha podido determinar que la vacuna AstraZeneca, es la que más ha sido utilizada, pero la que además ha tenido mayores reacciones adversas.

En otro contexto, el estudio de los autores [8], las vacunas contra la COVID-19 han demostrado cierta hiperreactividad del sistema inmune, lo que puede ocasionar falsos diagnósticos en el dictamen de los estudios PET-TC de pacientes oncológicos. El trabajo de nivel experiencia, describe estos hallazgos en relación con el tipo de vacuna utilizada.

Conocer los factores asociados a la aceptación de las vacunas, es de suma importancia el presente estudio de [9] tuvo como objetivo determinar los factores asociados a la aceptación de la vacuna contra el COVID-19 realizado en el cantón Azogues en el año 2022. Estudio que debido a que la vacuna es el mejor medio de prevención para el control de las epidemias, a falta de terapias más eficaces. El estudio descriptivo, transversal de enfoque cuantitativo, correlacional. Participaron del estudio 228 personas, para lo cual se utilizó un instrumento con 22 preguntas para conocer el nivel de aceptación y los factores de influencia. El estudio arrojó los siguientes resultados, la edad promedio fue de 32.5 años, de los cuales participaron hombres en 52.6%, con estudios secundarios 45.6%, su estado civil soltero fueron 59.2%, cuya orientación religiosa católica fueron 78.9%. Todos aceptaron ser vacunados con la COVID-19. Todos indicaron tener conocimiento sobre la vacuna, documentándose por medio del internet. Existieron razones para no vacunarse como: la poca efectividad, no estar en el grupo de riesgo, además de los efectos adversos. Se concluye en el estudio que la vacunación es la única forma de reducir el COVID-19, pero que se deben realizar estudios más profundos y en diversas poblaciones y evaluar la aceptación para estos casos.

La distribución de las vacunas se reflejan en el trabajo de [10] en la que diversos tipos de vacunas contra el COVID-19 han sido distribuidas a nivel mundial. Pero también se han reportado efectos adversos y complicaciones neurológicas (encefalopatía aguda). A pesar de ser aún supuestos, si estas vacunas están relacionadas de alguna manera con la vacuna contra el COVID-19. Los dos casos reportados de encefalopatía aguda se han dado después de la tercera dosis de la vacuna Moderna y Pfizer, se descartó meningoencefalitis por otras causas. Se concluye en la que el personal de salud debe estar vigilante ante posibles efectos adversos de

estas vacunas. Por lo que el estudio recomienda nuevos estudios para determinar si existe asociación entre las vacunas contra el COVID-19 y las complicaciones neurológicas.

Estudios para un mejor control de la vacunación como el presente trabajo de investigación de [11], cuyo objetivo se basa en construir un modelo epidémico para el control de vacunación del COVID-19 en El Salvador. El estudio promueve una combinación de modelos SEIR (Susceptibles, Expuestos, infectados y Recuperados) mediante técnicas de Machine Learning y redes de contacto. El procedimiento incluyó las fases de a) análisis identificando las variables críticas para el estudio, con la que se diseñó y construyó un modelo con sus parámetros o variables b) Simulación se puso a prueba el modelo, con lo que se pudo variar los parámetros e implementar otras soluciones sin que se altere el modelo original tomando las decisiones más adecuadas, las simulaciones fueron realizadas con los datos de la población de El Salvador, y c) Optimización se realizaron pruebas en diversos escenarios aplicando medidas de control para la vacunación y del distanciamiento social, para tener la mejor estrategia. El resultado se dio como mejorar la estrategia en el monitoreo de la enfermedad: Continuar con las medidas de distanciamiento, priorizar la vacuna de las personas vulnerables, siendo estas las que dieron los resultados más efectivos para reducir el impacto de la pandemia y reducir los costos del tratamiento. Finalmente, se presentan las recomendaciones para las autoridades.

Conocer la voluntad para vacunarse es importante, por lo cual la presente investigación de [12] cuyo objetivo fue identificar la predisposición para recibir la vacuna contra el COVID-19 en Paraguay, se aplicó un estudio del nivel exploratorio online, siendo el estudio observacional, transversal realizado con una encuesta en línea. El instrumento para la encuesta se diseñó en Google Forms, poniéndose a disposición en todo el país por medio de las redes sociales como Twitter, Facebook e Instagram. Además, se emplearon aplicaciones de mensajería como Telegram y Whatsapp, la aplicación del cuestionario se realizó entre el 20 de enero al 08 de febrero del 2021.

En una orientación más amplia, la investigación de [13] se evaluó la respuesta humoral de anticuerpos IG del personal de salud, para la proteína Spike del COVID-19, posterior a la vacuna en el hospital. la investigación se realizó después de 20 días de la puesta de la segunda dosis de la vacuna; se extrajo n257 de sangre del personal que abarco el 70.9% del total, siendo esta participación de forma voluntaria. Se evidenciaron positivamente el 97.6% en n251 muestras. Los resultados de AC en n5 que corresponde el 2.0% fueron negativos y 0.4% con resultado indeterminado para n1. Posterior a esta prueba a los 120 días se realizó la segunda para evaluar los anticuerpos, con n132 muestras (37.3%) de asistencia. Hubo una reducción en n121 lo que representó 89.6%. 8.1% en n11 fueron negativos y nunca se presentaron

anticuerpos. Se mantuvo el 15.7% en los anticuerpos con el historial de haber tenido la infección natural.

Similarmente, la investigación en la República Dominicana de [14] cuyo propósito fue identificar los determinantes de aceptación de la vacuna para el COVID-19 en pobladores mayores a 18 años del Distrito Nacional. El estudio se realizó entre febrero y marzo del 2021 por medio de una plataforma en línea. Participaron del estudio 120 ciudadanos residentes de forma voluntaria sobre la aceptación de la vacuna del COVID-19. El cuestionario incluyó preguntas sociodemográficas en las que el 71.2% estaban dispuestos a vacunarse, el 17.7% no estaban seguros de la vacunación y solo el 11% no tenían intención de ponerse la vacuna. En el caso del sexo, 71.5% de los hombres indicaron mayor predisposición para la vacuna en relación con las mujeres. Los hombres mayores de 45 años aceptaron vacunarse en relación con los más jóvenes. Se evidenció la recomendación de un médico, sobre confiar en el sistema de salud y un alto riesgo de contraer el virus fueron determinantes para que haya una mayor intención de vacunarse. El estudio concluye que las tres cuartas partes de las personas aceptaron vacunarse. El predominio del nivel académico, ocupación, género, edad y nacionalidad fueron los factores demográficos más importantes. En la actualidad el aumento de los brotes de la COVID-19 se debe a la no intención de vacunarse, motivo por el cual se concluye en la importancia de la educación sanitaria relacionada con las consecuencias del contagio del virus, las que debe ser difundida por profesionales de la salud calificados.

La importancia de la vacunación en la reducción de los contagios, es de suma importancia, debido a ello, los autores del trabajo [15] sobre el freno que ha sufrido el contagio del virus se ha dado con la llegada de la vacuna, siendo esta la respuesta para detener el COVID-19. Sin embargo, existen conflictos de tipo éticos en un escenario de inestabilidad social, económica y social de la salud pública, lo que desencadena entre la obligación de las instituciones y la disposición de las personas en un entorno globalizado. Por lo que es de suma importancia realizar una evaluación sobre los aspectos éticos que incorporen elementos como: justicia social, distribución de la vacuna, prioridad cuando la oferta sea menor a la demanda y que actitudes desde los poderes del estado se deben adoptar. Punto a parte son los casos de reticencia de las personas no dispuestas a que se les vacune.

El control de la pandemia con relación a las vacuna se reportan en la investigación de [16] donde se han planteados varios criterios para establecer el control de la pandemia por medio de las vacunas para el COVID-19. La Organización Mundial de la Salud han establecidos parámetros mínimos de para los productos para la vacunación los que demuestren

inmunogenicidad y sobre todo la seguridad necesaria con un equilibrio entre el riesgo y el beneficio, además de que no se tenga efectos adversos con un 50% mínimo de protección.

Antecedentes nacionales

En la investigación de [17] en las que diversas ideas sobre las vacunas para el COVID-19 se han dado durante la pandemia, aunque la mayoría de estas son equivocadas, han limitado el efecto de la protección de las vacunas, esta percepción representa un grave problema de la salud pública. El trabajo tuvo como objetivo determinar las percepciones de la aceptación de la vacuna contra el COVID-19 en la ciudad de Arequipa en el 2021. Planteamiento metodológico utilizado para un estudio prospectivo, observacional, transversal, cualicuantitativo empleando para ello una entrevista semiestructurada, la muestra estuvo conformada por las personas que visitó o estuvo en los alrededores del mercado San Camilo en el mes de mayo del 2021. El resultado indica que los jóvenes entre 18-25 años no están dispuestos a la vacuna, por otro lado, el 65.60% de los adultos si están dispuestos. Igualmente, las mujeres no aceptan vacunarse en un 64.80%. los del nivel secundario no se vacunarán con un 53.40%. En general el 73.30% de los entrevistados indican que momento de la decisión, su intención es no vacunarse. La muestra reflejó que mayormente los jóvenes solteros del sexo femenino no están dispuestos a aceptar la vacuna para el COVID-19 por temor a posibles efectos adversos basados en fuentes equivocadas de información.

Aceptar la vacuna ha sido todo un reto en los países, como también en el nuestro, la investigación de [18] investiga si existe una relación entre el nivel de conocimiento y la aceptación a vacunarse. La falta de conocimiento o información errada, limitan la protección de la vacuna contra el COVID-19, la presente investigación está orientada a comparar el nivel de conocimiento y la aceptación de la vacuna en el personal de salud del hospital de nivel II-1. El estudio se efectuó a 259 profesionales asistenciales de salud lo que fueron sometidos a un cuestionario relacionado con el conocimiento de la vacuna COVID-19, los resultados indican que el mayor porcentaje del conocimiento sobre la vacuna se ubica en el hospital Jerusalén con 75.6% y solo un 6.3% con conocimiento bajo ubicado en el hospital Santa Isabel. En cuanto a la aceptación de la vacuna, el 96.5% se ubican en el hospital Jerusalén, y con el mismo 6.3% de rechazo en el hospital Santa Isabel, se concluyen que el nivel de conocimiento y la aceptación de la vacuna tienen una alta relación ($p=0.000$), por lo que se concluye que un alto conocimiento de las vacunas se tendrá un nivel de aceptación igualmente alto. Caso contrario se tendrá un rechazo hacia la vacuna para el COVID-19.

En el mismo camino en la investigación de [19] las teorías e hipótesis indicaban una falta de eficacia de la vacuna contra el COVID-19 creándose una serie de creencias entre la población por desconocimiento del comportamiento epidemiológico del virus, por lo cual se creó desconfianza, temor, ansiedad, angustia por el desconocimiento y la errada información en las redes sociales. Esta situación utilizada por los grupos antivacuna para generar mayor desinformación. La investigación permitió dar a conocer a los usuarios del centro de salud sobre la efectividad de las vacunas contra el COVID-19. El estudio de tipo cualitativo de diseño fenomenológico logró profundizar sobre como los usuarios percibían la información que circulaban en las redes sociales y que tomaban la decisión de aceptar o rechazar la vacuna. Los resultados en las que se aceptaba la vacuna eran: mantener la salud personal y familiar, cuidado de contraer la COVID-19 y tener síntomas graves, otro factor el no poder ingresar a determinados lugares sin el carné de vacunación. Se llegó a la conclusión que mitos religiosos, temor, vacunas ineficientes fueron los mayores indicadores para el rechazo a la vacuna, a pesar de finalmente tener que vacunarse.

1.2 Realidad problemática

El coronavirus ha generado nuevos contagios y muertes en el mundo por la mutación de variantes. A pesar de ser reducido los contagios y muertes no deja de preocupar esta condición según el portal de Datosmacro.com, se incrementa día a día los contagios y muertes en el Perú.

TABLA I
EVOLUCIÓN DEL CORONAVIRUS DICIEMBRE 2022 – ENERO 2023

Fecha	Incremento Muertos	Muertos	Muertos / millón	Incremento Confirmados	Confirmados	Confirmados / 100.000 - 14 días
02/02/2023	8	219.012	6.565,22	247	4.482.120	13,7
01/02/2023	73	219.004	6.564,98	627	4.481.873	14,42
31/01/2023	68	218.931	6.562,79	-44	4.481.246	15,36
30/01/2023	53	218.863	6.560,76	44	4.481.290	17,03
29/01/2023	60	218.810	6.559,17	145	4.481.246	17,45
28/01/2023	21	218.750	6.557,37	273	4.481.101	19,55
27/01/2023	18	218.729	6.556,74	374	4.480.828	21
26/01/2023	25	218.711	6.556,20	2.903	4.480.454	21,76
25/01/2023	30	218.686	6.555,45	-1.524	4.477.551	17,46
24/01/2023	0	218.656	6.554,55	0	4.479.075	24,45
23/01/2023	7	218.656	6.554,55	104	4.479.075	28,4
22/01/2023	24	218.649	6.554,34	279	4.478.971	28,09
21/01/2023	19	218.625	6.553,62	737	4.478.692	37,05

20/01/2023	9	218.606	6.553,05	404	4.477.955	34,84
19/01/2023	17	218.597	6.552,78	487	4.477.551	52,31
18/01/2023	26	218.580	6.552,27	941	4.477.064	50,85
17/01/2023	7	218.554	6.551,49	513	4.476.123	44,95
16/01/2023	17	218.547	6.551,28	184	4.475.610	46,49
15/01/2023	21	218.530	6.550,77	846	4.475.426	50,31
14/01/2023	19	218.509	6.550,14	759	4.474.580	47,77
13/01/2023	13	218.490	6.549,58	627	4.473.821	50,35
12/01/2023	22	218.477	6.549,19	1.468	4.473.194	56,83
11/01/2023	41	218.455	6.548,53	807	4.471.726	61,91
10/01/2023	1	218.414	6.547,30	1.318	4.470.919	75,45
09/01/2023	26	218.413	6.547,27	0	4.469.601	76,72
08/01/2023	31	218.387	6.546,49	3.270	4.469.601	277,87
07/01/2023	0	218.356	6.545,56	0	4.466.331	72,36
06/01/2023	52	218.356	6.545,56	6.231	4.466.331	82,59
05/01/2023	0	218.304	6.544,00	0	4.460.100	76,35
04/01/2023	20	218.304	6.544,00	-1.029	4.460.100	89,75
03/01/2023	33	218.284	6.543,40	1.029	4.461.129	108,58
02/01/2023	19	218.251	6.542,41	1.457	4.460.100	129,73
01/01/2023	0	218.232	6.541,84	0	4.458.643	127,92
31/12/2022	31	218.232	6.541,84	1.617	4.458.643	139,4
30/12/2022	23	218.201	6.540,91	2.790	4.457.026	153,43
29/12/2022	32	218.178	6.540,22	3.162	4.454.236	165,36
28/12/2022	33	218.146	6.539,26	5.326	4.451.074	182,01
27/12/2022	21	218.113	6.538,27	1.742	4.445.748	206,37
26/12/2022	20	218.092	6.537,64	67.101	4.444.006	244,98
25/12/2022	64	218.072	6.537,04	-65.286	4.376.905	53,05
24/12/2022	1	218.008	6.535,13	3.413	4.442.191	263,61
23/12/2022	9	218.007	6.535,10	4.148	4.438.778	264,59
22/12/2022	38	217.998	6.534,83	4.470	4.434.630	268,75
21/12/2022	19	217.960	6.533,69	5.254	4.430.160	298,68
20/12/2022	15	217.941	6.533,12	8.083	4.424.906	373,51
19/12/2022	32	217.926	6.532,67	852	4.416.823	348,47
18/12/2022	39	217.894	6.531,71	3.830	4.415.971	345,91
17/12/2022	34	217.855	6.530,54	6.298	4.412.141	362,02
16/12/2022	39	217.821	6.529,52	6.770	4.405.843	380,24
15/12/2022	44	217.782	6.528,35	8.718	4.399.073	398,15
14/12/2022	34	217.738	6.527,03	13.450	4.390.355	413,59
13/12/2022	25	217.704	6.526,01	14.624	4.376.905	416,26

12/12/2022	22	217.679	6.525,26	3.072	4.362.281	398,99
11/12/2022	21	217.657	6.524,60	4.958	4.359.209	394,98
10/12/2022	12	217.636	6.523,98	3.738	4.354.251	411,23
09/12/2022	20	217.624	6.523,62	5.536	4.350.513	400,02
08/12/2022	38	217.604	6.523,02	14.456	4.344.977	404,9
07/12/2022	24	217.566	6.521,88	30.215	4.330.521	400,17
06/12/2022	46	217.542	6.521,16	-270	4.300.306	309,6
05/12/2022	0	217.496	6.519,78	0	4.300.576	328,63
04/12/2022	26	217.496	6.519,78	9.203	4.300.576	328,63
03/12/2022	29	217.470	6.519,00	12.376	4.291.373	321,29
02/12/2022	13	217.441	6.518,13	12.746	4.278.997	284,2
01/12/2022	14	217.428	6.517,74	13.868	4.266.251	256,05

En la Tabla I, nuestro país reporta, 4.482.120 personas confirmadas de coronavirus, 247 personas confirmadas más que el valor anterior.

Fuente: <https://datosmacro.expansion.com/otros/coronavirus/peru>

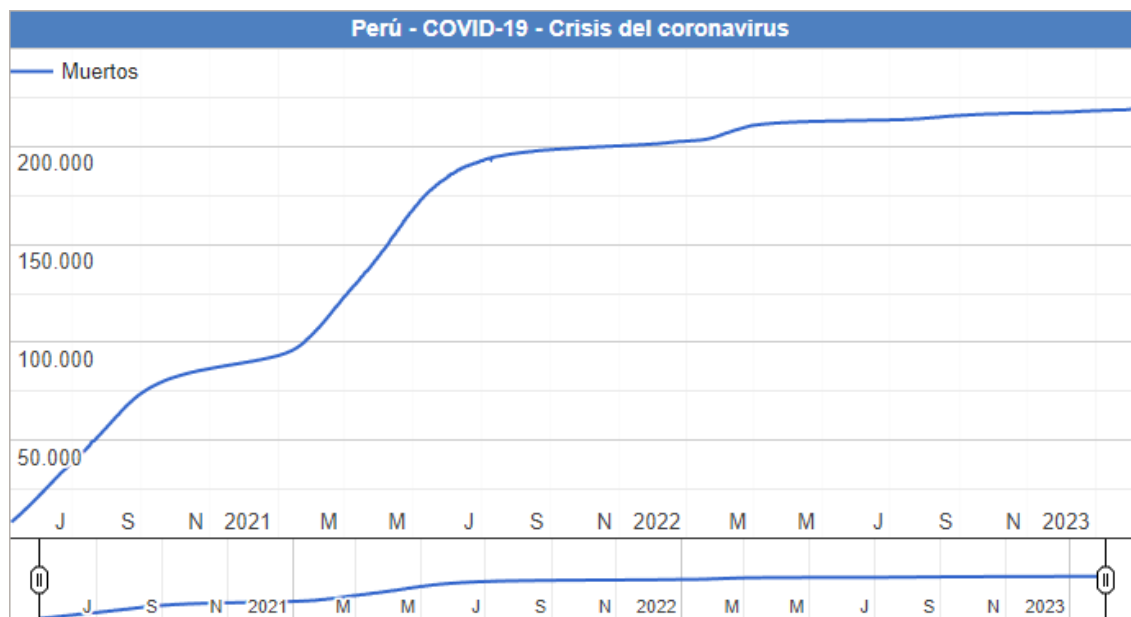


Fig. 2. Evolución del coronavirus en el Perú

Fuente: <https://datosmacro.expansion.com/otros/coronavirus/peru>

Como se desprende de la evolución del coronavirus en nuestro país, es importante aprovechar los datos registrados sobre las vacunaciones, para tener conocimiento sobre el estado de la vacunación, los tipos de vacunas empleadas, la edad de la persona vacunada, y el desconocer cuáles serán los efectos adversos que pueden presentar la vacunación es una necesidad primordial, sin embargo, a pesar de que en el transcurso de la pandemia se generaron diversos

temores y rechazo a la vacuna por una parte de la población. Muchos de estos temores por el desconocimiento, o mala información que circulaba por las redes sociales. A estas alturas cuando ya el mayor porcentaje de la población en nuestro país y en nuestra región han sido vacunados. Al contar con datos sobre la vacunación es importante aprovechar estos datos para convertirlos en conocimiento, aprovechando la analítica de datos. Datos como la edad de la persona, marca de la vacuna, cantidad de dosis aplicada y otros parámetros adicionales, nos permiten poder hacer un análisis de estos datos que servirán de referencia para posibles casos que se presente de respuestas adversas a la vacunación.

Como consecuencia de esta realidad, se planteó la siguiente formulación del problema ¿Será eficaz la intervención de la analítica de datos en la evaluación de la vacunación COVID-19?

1.3 Justificación

El proyecto de investigación es de suma importancia, ya que existen datos sobre la vacunación en la ciudad de Ica, pero es necesario obtener de esos datos el conocimiento necesario para encontrar conocimiento y patrones en dichos datos. Además de encontrar conocimiento en una de las pandemias más grandes que ha pasado la humanidad y que aún a pesar de retorno a la aparente normalidad, están reportándose casos de contagios y muertes por este virus. Por lo cual es proyecto permitirá evaluar la vacunación realizada en nuestra provincia.

Se justifica el proyecto en la medida que las técnicas de análisis de datos están cobrando mayor relevancia en el tratamiento de estos, con la aplicación de técnicas y herramientas avanzadas de estadística.

Realizamos esta investigación con el objetivo de Evaluar la INTERVENCION DE LA ANALÍTICA DE DATOS EN EL ANÁLISIS DE LA VACUNACIÓN COVID-19 EN LA PROVINCIA ICA, la misma que se hizo extensiva a toda la región de Ica, por estar disponible los datos de manera abierta. con la que realizamos una evaluación descriptiva de la vacunación en sus diversas variables.

II. ESTRATEGIA METODOLÓGICA

2.1. Diseño metodológico

El diseño metodológico está determinado por el tipo de investigación aplicada tecnológica, cuantitativa, según la intervención corresponde a una investigación observacional, al no tener el control de los datos la investigación es retrospectiva, según el corte de la investigación es transversal y al medir una variable el estudio es descriptivo [20].

2.2. Población y muestra

Unidad de investigación: ciudadano que ha recibido la vacuna

Población: todos los ciudadanos que han sido vacunado en la región de Ica.

Muestra: la muestra está determinada por los registros de los vacunados en la región de Ica, que según el reporte de la DIRESA se cuenta con 1048,554 registros de vacunación correspondientes al periodo de mayo 2021– agosto del 2023 mostrados en la Fig. 34.

2.3. Material

Para el desarrollo del proyecto este se realizó en dos partes: por un lado, la revisión bibliográfica en diversas bases de datos sobre el tema a desarrollar, para darle el sustento soporte teórico respectivo.

Debido al tamaño del archivo (+ de 10 MB) obtenido del portal del MINSA, Se empleó el software R y R-Studio, esto debido a que con Excel o Python no se pudo cargar el archivo. Con R-Studio se pudo filtrar los datos para tener solo los que corresponden a la Región Ica, reduciendo significativamente el tamaño de la Base de Datos (BD), haciendo que esta pueda ser cargada en Python y en Orange, para el procesamiento.

En segundo lugar, se emplearon diversas técnicas y algoritmos para el análisis de los datos mediante el uso de la plataforma tecnológica del Google Colaboratory en la nube, con la

aplicación del lenguaje Python, con librerías como Pandas y Numpy y la librería de sweetviz. Además, se utilizó el software para el análisis de datos de Orange Data Mining, el cual es un software gráfico por medio de widgets que facilitan el procesamiento y análisis de los datos no siendo necesarios conocimientos del lenguaje de programación como Python.

2.4. Método

Para el análisis de datos, se seguirá el siguiente procedimiento, detallado en la Fig. 2:

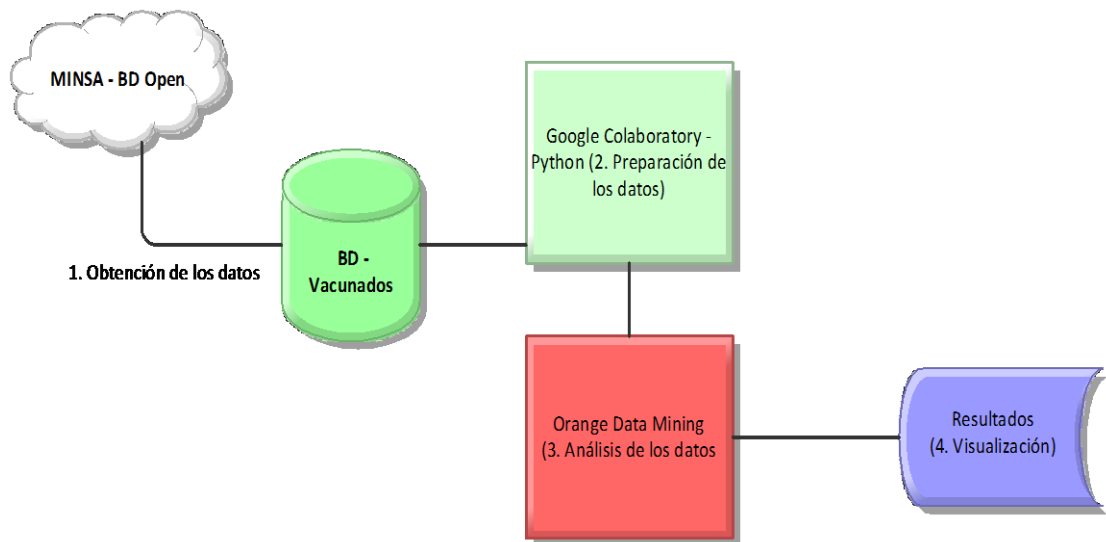


Fig. 3. Método para el análisis de los datos

1. Obtención de datos: Los datos sobre la vacunación empleados fueron obtenidos del portal del Minsas. Estos datos son de acceso libre (open access).
2. Preparación de datos: Los datos originales, no siempre están en el formato correcto para el análisis por lo cual hubo la necesidad de la preparación y transformación de estos. Aquí hay que seleccionar los datos que aportan al estudio, ver si los datos están en el formato correcto (tipo de datos) de no estarlo de debe transformar el formato adecuado. Verificar que en los datos no existan datos vacíos o perdidos, de serlo así se debe realizar la *imputación* de los datos o eliminarlos. La culminación de esta fase es tener la base de datos limpia y con las variables que aporten a la analítica de datos, también se le conoce como el nombre de *dataSet* o *dataFrame*.
3. Analítica de los datos: Se realizó el análisis, descriptivo de los datos, la clasificación y agrupamiento de los datos y ver el comportamiento de la vacunación en el tiempo para ello se utilizó de diversos algoritmos y librerías para este análisis.

- Comunicación y visualización: para una mayor comprensión de los resultados se obtuvieron visualizaciones gráficas y estadísticas para entender las características, distribución, clasificación y agrupamiento de los datos.

Es importante considerar que el análisis de datos es un proceso iterativo y que requiera regresar a cualquiera de estos pasos y efectuar los cambios en función de los resultados mostrados.

2.4.1. Obtención de los datos

Los datos de la vacunación del COVID-19, fueron obtenidos del portal del ministerio de salud MINSa del país, en el portal de los vacunados el estado peruano pone a disposición el Dataset de los mismos.

Vacunación contra COVID - 19 [Ministerio de Salud - MINSa]

COVID-19
Los datos publicados corresponden a las vacunas reportadas contra COVID-19 a nivel nominal.

Dato y Medio de Distribución

- Data Completa de Vacunación COVID-19 - [Ministerio de Salud]**
Data de Vacunación COVID-19 - [Ministerio de Salud] [Ir al recurso](#)
- Diccionario de Datos - Vacunas contra Covid - 19**
Data de Vacunación COVID-19 - [Ministerio de Salud] [Previsualizar](#) [Descargar](#)
- Data de Muestra Vacunación Covid19 - [Ministerio de Salud]**
Data de Muestra Vacunación Covid19 - [Ministerio de Salud] [Ir al recurso](#)

Dataset Info

These fields are compatible with **DCAT**, an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web.

Field	Value
Publisher	Ministerio de Salud
Fecha modificada	2022-09-17
Fecha de lanzamiento	2021-04-23
Identificador	24af4ff4-226d-4e3d-90cb-d26a1849796e
License	Open Data Commons Attribution License
Contact Email	mesadeayuda@minsa.gob.pe
Public Access Level	Public

Fig. 4. Portal de los datos abiertos de la vacunación del MINSa

Fuente: <https://www.datosabiertos.gob.pe/dataset/vacunación-contra-covid-19-ministerio-de-salud-minsa>

Como se puede comprobar en la Fig. 4, el MINSa pone a disposición los datos de la vacunación para cualquier persona que desee conocer la información registrada en sus sistemas. El portal cuenta con la posibilidad de tener la información siguiente:

- ✓ Data completa de la vacunación Covid-19
- ✓ Data de muestra de la vacunación Covid-19

✓ Diccionario de los datos sobre la vacunación

En la información del *Dataset*, se puede comprobar la licencia de ser una base de datos abierta *Open Data Commons Attribution License*. Este tipo de licencia le confiere a cualquier persona el poder utilizar los datos para los fines que estime conveniente, por ser datos de acceso público.



Nombre	Fecha de modificación	Tipo	Tamaño
 Datavacunacovid	08/07/2023 15:34	Archivo de valores separados por comas d...	120.391 KB
 Diccionario_Datos	26/07/2023 8:36	Hoja de cálculo de Microsoft Excel	11 KB
 vacunas_covid	07/07/2023 7:19	Archivo de valores separados por comas d...	10.176.974 KB

Fig. 5. Archivos descargados del MINSAs

En la Fig. 5, se muestra los archivos descargados del portal del MINSAs, los datos de la vacunación como se puede apreciar son archivos de formatos CSV (muy utilizados en analítica de datos) y el archivo del diccionario de datos en formato Excel.

El archivo con los datos totales de la vacunación es un archivo de gran tamaño aproximadamente de 10 GB. El cual es imposible cargar con la hoja de cálculo de Excel, como se muestra en la Fig. 4.

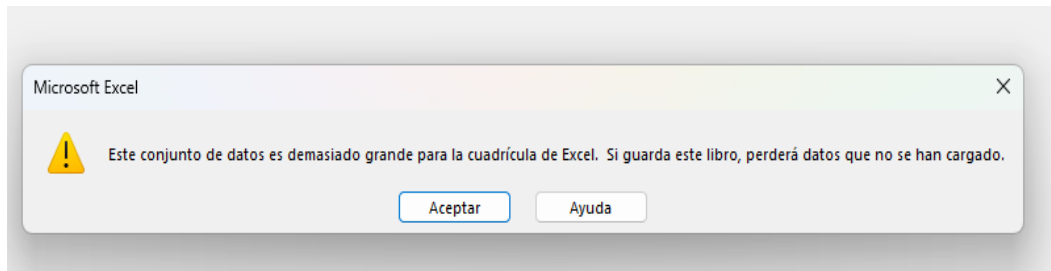


Fig. 6. Intento de carga de archivo de datos totales de vacunación

A pesar de ser imposible abrir el archivo en un computador personal, al aceptar el mensaje el sistema carga la información hasta el máximo disponible del software Excel.

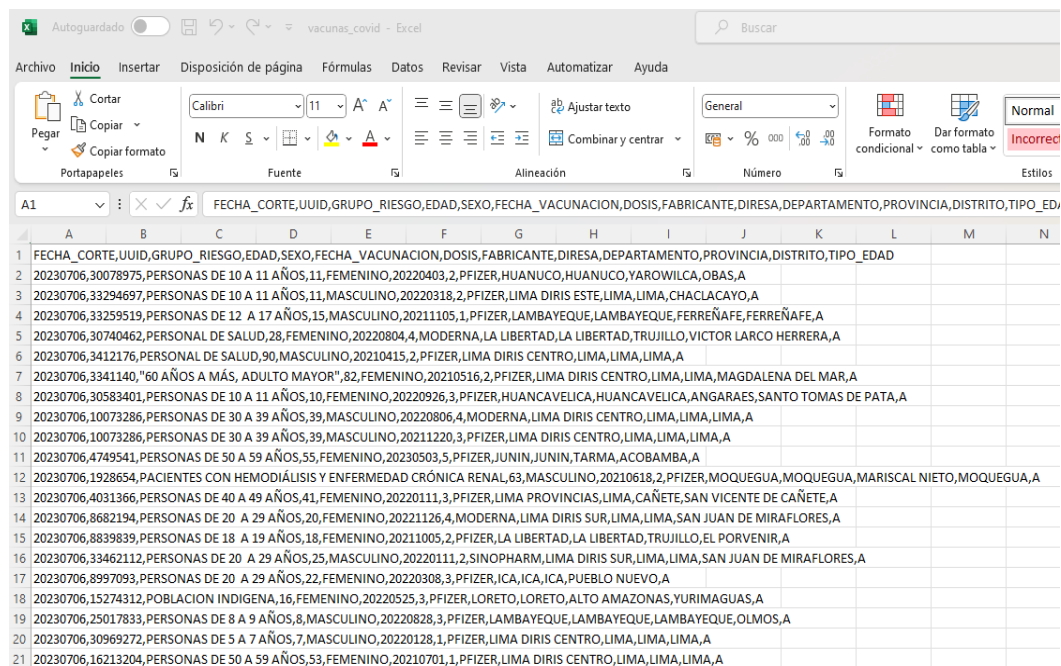


Fig. 7. Archivo cargado, este abre solo hasta la información posible de carga del Excel

El archivo cargado que se muestra en la Fig. 7, abre hasta el máximo de la capacidad del software en donde se comprueba la estructura de los datos del formato .CSV.

Para un mayor entendimiento de los datos almacenados, se abrió el archivo del diccionario de datos, que se encuentra en formato Excel. El contenido se muestra a continuación.

TABLA II
DICCIONARIO DE LOS DATOS

Variable	Descripción
FECHA_CORTE	Fecha de corte de información
UUID	ID de la persona vacunada
GRUPO_RIESGO	Grupo de riesgo al que pertenece la persona vacunada
EDAD	Edad de la persona vacunada
SEXO	Sexo de la persona vacunada
FECHA_VACUNACION	Fecha de vacunación
DOSIS	Dosis aplicada
FABRICANTE	Fabricante de la vacuna
DIRESA	Diresa del EESS donde se aplicó la vacuna

DEPARTAMENTO	Departamento del EESS donde se aplicó la vacuna
PROVINCIA	Provincia del EESS donde se aplicó la vacuna
DISTRITO	Distrito del EESS donde se aplicó la vacuna

2.4.2. Preparación de los datos

Obtenido los datos para la analítica de estos, se realizó una exploración de estos, para ello se utilizó el software *Python* en la plataforma del *Jupyter notebook*.

Para hacer la exploración de los datos iniciales, se debe tener instalado las librerías para el tratamiento de los datos.

```
# Carga de librerías para comprensión de los datos
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import datetime
import os
```

Las instrucciones muestran que se han cargado las librerías para la exploración de los datos. la librería *panda* la misma que permite la manipulación de *Dataset*, *numpy* para cálculos, la librería *matplotlib* para el análisis gráfico y *datetime* para el manejo de fechas.

Una vez cargadas las librerías se procede a la carga de los datos

```
# Carga de los datos completos
vacuna=pd.read_csv('Datavacunacovid.csv', index_col=0, parse_dates=True)
```

El archivo original descargado del MINSA no fue posible poder realizar su lectura, debido al tamaño de más de 10 GB, no pudiendo cargarlo por las limitaciones de la herramienta del *Google colaboratory*.

Se realizó una primera evaluación con los datos instrucción de carga de los datos del archivo de muestra de las vacunaciones del covid19 con el nombre de *Datavacunacovid.csv*, teniendo como columna índice la primera columna.

Conocer la información del Data set mediante la instrucción `vacuna.info ()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 116216 entries, 0 to 116215
Data columns (total 13 columns):
#   Column          Non-Null Count  Dtype
```

```

-----
0  FECHA_CORTE      116216 non-null  object
1  UUID             98730 non-null  object
2  GRUPO_RIESGO     98730 non-null  object
3  EDAD             98728 non-null  object
4  SEXO             98730 non-null  object
5  FECHA_VACUNACION 98729 non-null  float64
6  DOSIS            98729 non-null  object
7  FABRICANTE       98729 non-null  object
8  DIRESA           98729 non-null  object
9  DEPARTAMENTO      98729 non-null  object
10 PROVINCIA        98729 non-null  object
11 DISTRITO         98729 non-null  object
12 TIPO_EDAD;       98726 non-null  object
dtypes: float64(1), object(12)
memory usage: 11.5+ MB

```

La información presentada muestra en la parte superior de que se han encontrado 116216 registros y la fecha de corte (descarga en nuestro caso) el 06 de Julio del 2023, los datos con sus correspondientes tipos de datos, y en la parte inferior el peso del archivo que es de 11.5 MB.

En este primer acercamiento de los datos se puede comprobar los tipos de datos de cada variable.

El contenido de los datos se pudo visualizar con la instrucción `vacuna.head()`, para ver los 5 primeros registros del *Dataset*.

FECHA_CORTE	UUID	GRUPO_RIESGO	EDAD	SEXO	FECHA_VACUNACION	DOSIS	FABRICANTE	DIRESA	DEPARTAMENTO	F
20230706	9659208	PERSONAS DE 50 A 59 AÑOS	58	MASCULINO	20220113.0	1	SINOPHARM	PIURA	PIURA	
20230706	33470872	PERSONAS DE 50 A 59 AÑOS	55	FEMENINO	20220113.0	3	PFIZER	PIURA	PIURA	
20230706	3906284	PERSONAS DE 40 A 49 AÑOS	42	FEMENINO	20210815.0	2	PFIZER	LIMA DIRIS SUR	LIMA	
20230706	34409828	PERSONAS DE 5 A 7 AÑOS	6	FEMENINO	20220203.0	1	PFIZER	LIMA DIRIS NORTE	LIMA	
20230706	9875492	PERSONAS DE 5 A 7 AÑOS	7	MASCULINO	20220315.0	1	PFIZER	LIMA DIRIS NORTE	LIMA	

Fig. 8. Datos iniciales del Dataset

Para visualizar la información de los últimos registros del *Dataset* se usa la instrucción `vacuna.tail()`, los datos se presentan a continuación.

PROVINCIA	DISTRITO	TIPO_EDAD;
TALARA	PARIÑAS	A;
TALARA	PARIÑAS	A;
LIMA	BARRANCO	A;
LIMA	SAN MARTIN DE PORRES	A;
LIMA	SAN MARTIN DE PORRES	A;

Fig. 9. Datos finales del *Dataset*

En las Fig. 8 y 9 se muestran los datos finales del *Dataset*, conociendo la información de forma visual, se tiene conocimiento que los datos registrados y que han sido extraído del portal del MINSA, corresponden a las vacunaciones realizadas en todo el país, por lo que es necesario acceder solo a los datos que correspondan a la región Ica.

Para filtrar los registros para obtener un *Dataset* solo para la región se utilizó la instrucción `vacunaIca=vacuna[vacuna['DEPARTAMENTO']== 'ICA']`. con esta instrucción se ha creado un nuevo *Dataset* con los registros de la región Ica y se le ha asignado el nombre de `vacunaIca`.

Para conocer la información del nuevo *Dataset* `vacunaIca` aplicamos la instrucción `vacunaIca.info()`

```
<class 'pandas.core.frame.DataFrame'>
Index: 28688 entries, 20230706 to 20230706
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   UUID                   28688 non-null  object
1   GRUPO_RIESGO           28688 non-null  object
2   EDAD                   28688 non-null  object
3   SEXO                   28688 non-null  object
4   FECHA_VACUNACION      28688 non-null  float64
5   DOSIS                  28688 non-null  object
6   FABRICANTE             28688 non-null  object
7   DIRESA                 28688 non-null  object
8   DEPARTAMENTO            28688 non-null  object
9   PROVINCIA              28688 non-null  object
10  DISTRITO                28688 non-null  object
11  TIPO_EDAD;             28688 non-null  object
```

```
dtypes: float64(1), object(11)
memory usage: 2.8+ MB
```

La información del nuevo *Dataset* cuenta con la cantidad de 28,688 registros, con un peso de 2.8 MB de almacenamiento.

Para comprobar los registros volvemos a aplicar las instrucciones para visualizar los datos del inicio y fin del *Dataset*, los resultados se muestran a continuación.

`vacunaIca.head()`

	UUID	GRUPO_RIESGO	EDAD	SEXO	FECHA_VACUNACION	DOSIS	FABRICANTE	DIRESA	DEPARTAMENTO	PROVINCIA	DISTRITO
HA_CORTE											
20230706	8997093	PERSONAS DE 20 A 29 AÑOS	22	FEMENINO	20220308.0	3	PFIZER	ICA	ICA	ICA	PUEBLO NUEVO
20230706	13554264	PERSONAS DE 12 A 17 AÑOS	15	MASCULINO	20211031.0	1	PFIZER	ICA	ICA	ICA	SALAS
20230706	8997098	PERSONAS DE 18 A 19 AÑOS	19	FEMENINO	20211227.0	2	SINOPHARM	ICA	ICA	PISCO	PISCO
20230706	9094855	PERSONAS DE 30 A 39 AÑOS	36	MASCULINO	20210828.0	1	SINOPHARM	ICA	ICA	ICA	SAN JOSE DE LOS MOLINOS
20230706	24797758	PERSONAS DE 40 A 49 AÑOS	41	MASCULINO	20210928.0	1	SINOPHARM	ICA	ICA	ICA	PUEBLO NUEVO

Fig.10. Datos iniciales del *Dataset* vacunaIca

`vacunaIca.tail()`

	UUID	GRUPO_RIESGO	EDAD	SEXO	FECHA_VACUNACION	DOSIS	FABRICANTE	DIRESA	DEPARTAMENTO	PROVINCIA	DISTRITO
HA_CORTE											
20230706	8058228	PERSONAS DE 12 A 17 AÑOS	16	MASCULINO	20211120.0	2	PFIZER	ICA	ICA	ICA	SUBTANJALLA
20230706	9474786	PERSONAS DE 30 A 39 AÑOS	33	MASCULINO	20210904.0	1	SINOPHARM	ICA	ICA	PISCO	PISCO
20230706	16634664	PERSONAS DE 40 A 49 AÑOS	47	MASCULINO	20210820.0	2	PFIZER	ICA	ICA	ICA	SANTIAGO
20230706	30297748	PERSONAS DE 18 A 19 AÑOS	18	MASCULINO	20220309.0	3	PFIZER	ICA	ICA	PISCO	HUMAY
20230706	1179648	PERSONAS DE 50 A 59 AÑOS	59	MASCULINO	20210828.0	2	PFIZER	ICA	ICA	PISCO	HUMAY

Fig. 11. Datos finales del *Dataset* vacunaIca.

Los datos ahora se comprueban que corresponden con el estudio, vale decir para la región de Ica, un punto a tomar en cuenta es que los datos no están en el formato del tipo correcto. La información muestra que la EDAD y la DOSIS se presentan con el tipo *Object* que corresponden con datos del tipo categórico, y la FECHA_VACUNACION del tipo float64, sin embargo, esta debería ser del tipo Fecha.

Realizar un análisis descriptivo nos orienta en cuanto a los datos, se aplicó la instrucción `vacunaIca.describe()`, esta instrucción presenta la estadística descriptiva de los datos numéricos.

FECHA_VACUNACION	
count	2.868800e+04
mean	2.021607e+07
std	5.547970e+03
min	2.021021e+07
25%	2.021092e+07
50%	2.021123e+07
75%	2.022042e+07
max	2.023071e+07

Fig. 12. Resultados descriptivos

En la Fig.12 los resultados de esta instrucción nos arrojan que los datos no están debidamente preparados para poder realizar la analítica de estos. Los resultados debieron haber obtenido la estadística descriptiva de los datos numéricos, pero como se puede comprobar solo tomó la fecha de vacunación ignorando datos como la EDAD, la DOSIS, lo que nos indica que no están en el formato correcto.

Para corregir el formato de estos datos, se aplicó las instrucciones correspondientes para convertir los datos a sus tipos respectivos, para los datos de tipo numérico.

```
vacunaIca.EDAD=vacunaIca.EDAD.astype(int)
vacunaIca.DOSIS=vacunaIca.DOSIS.astype(int)
```

Igualmente se corrigió el formato de la fecha con la instrucción

```
vacunaIca['FECHA_VACUNACION']=pd.to_datetime(vacunaIca['FECHA_VACUNACION'], errors='coerce')
```

Verificamos los cambios con la instrucción `vacunaIca.info()`

```
<class 'pandas.core.frame.DataFrame'>
Index: 28688 entries, 20230706 to 20230706
Data columns (total 12 columns):
```

```

#      Column                Non-Null Count  Dtype
---  -
0      UUID                  28688 non-null  object
1      GRUPO_RIESGO          28688 non-null  object
2      EDAD                   28688 non-null  int32
3      SEXO                   28688 non-null  object
4      FECHA_VACUNACION      28688 non-null  datetime64[ns]
5      DOSIS                   28688 non-null  int32
6      FABRICANTE              28688 non-null  object
7      DIRESA                  28688 non-null  object
8      DEPARTAMENTO             28688 non-null  object
9      PROVINCIA               28688 non-null  object
10     DISTRITO                 28688 non-null  object
11     TIPO_EDAD;               28688 non-null  object
dtypes: datetime64[ns] (1), int32 (2), object (9)
memory usage: 2.6+ MB

```

La información que se muestra ahora marca los datos de EDAD y DOSIS de tipo numérico y la fecha en el formato de Fecha.

Volvemos a comprobar la estadística descriptiva para ver los datos `vacunaIca.describe()`

	EDAD	DOSIS
count	28688.000000	28688.000000
mean	32.124721	2.194855
std	17.090345	1.061184
min	1.000000	1.000000
25%	19.000000	1.000000
50%	31.000000	2.000000
75%	45.000000	3.000000
max	102.000000	8.000000

Fig. 13. Estadística descriptiva de los datos numéricos

La aplicación de la instrucción de la estadística descriptiva nos arroja una buena evaluación de los datos. es importante comprender que la edad promedio de los que se han vacunado es de 32 años y la cantidad de dosis de 2 a la fecha de corte. Sin duda llama la atención en cuanto a la edad de que haya casos en la que la edad es de 102 años, y por otro lado en cuanto a la dosis, igualmente exista alguna persona que se haya aplicado 8 vacunas.

ahora, se debe seleccionar solo los datos que nos servirán para la analítica de estos, para ello tomamos con referencia la visualización de los datos.

`vacunaIca.head()`

UUID	GRUPO_RIESGO	EDAD	SEXO	FECHA_VACUNACION	DOSIS	FABRICANTE	DIRESA	DEPARTAMENTO	PROVINCIA	DISTRITO	TIPO_EDAD;
8997093	PERSONAS DE 20 A 29 AÑOS	22	FEMENINO	20220308.0	3	PFIZER	ICA	ICA	ICA	PUEBLO NUEVO	A;
13554264	PERSONAS DE 12 A 17 AÑOS	15	MASCULINO	20211031.0	1	PFIZER	ICA	ICA	ICA	SALAS	A;
8997098	PERSONAS DE 18 A 19 AÑOS	19	FEMENINO	20211227.0	2	SINOPHARM	ICA	ICA	PISCO	PISCO	A;
9094855	PERSONAS DE 30 A 39 AÑOS	36	MASCULINO	20210828.0	1	SINOPHARM	ICA	ICA	ICA	SAN JOSE DE LOS MOLINOS	A;
24797758	PERSONAS DE 40 A 49 AÑOS	41	MASCULINO	20210928.0	1	SINOPHARM	ICA	ICA	ICA	PUEBLO NUEVO	A;

Fig. 14. Datos del *Dataset*

Visualizando los datos de la Fig. 14, se evalúa que hay datos que no son relevantes para el análisis de estos. Se tendrá que eliminar las columnas de UUID, DIRESA, y DEPARTAMENTO que nos son relevantes.

`vacunaIca.drop(columns=['UUID','DIRESA','DEPARTAMENTO'], inplace=True)`, el argumento *inplace* es para hacer que los cambios sean permanentes en el *Dataset*.

FECHA_CORTE	GRUPO_RIESGO	EDAD	SEXO	FECHA_VACUNACION	DOSIS	FABRICANTE	PROVINCIA	DISTRITO	TIPO_EDAD;
20230706	PERSONAS DE 20 A 29 AÑOS	22	FEMENINO	1970-01-01 00:00:00.020220308	3	PFIZER	ICA	PUEBLO NUEVO	A;
20230706	PERSONAS DE 12 A 17 AÑOS	15	MASCULINO	1970-01-01 00:00:00.020211031	1	PFIZER	ICA	SALAS	A;
20230706	PERSONAS DE 18 A 19 AÑOS	19	FEMENINO	1970-01-01 00:00:00.020211227	2	SINOPHARM	PISCO	PISCO	A;
20230706	PERSONAS DE 30 A 39 AÑOS	36	MASCULINO	1970-01-01 00:00:00.020210828	1	SINOPHARM	ICA	SAN JOSE DE LOS MOLINOS	A;
20230706	PERSONAS DE 40 A 49 AÑOS	41	MASCULINO	1970-01-01 00:00:00.020210928	1	SINOPHARM	ICA	PUEBLO NUEVO	A;

Fig. 15. *Dataset* con las variables para análisis

La Fig.15 presenta la información de los datos que corresponden solo con los que son posibles de hacer una analítica de datos.

Finalmente, se realiza la verificación de que no existan datos vacíos o perdidos, por lo cual se utilizó la instrucción `vacunaIca.isnull().sum()`, para asegurar que la data esta completa en todas las columnas.

```

GRUPO_RIESGO      0
EDAD              0
SEXO              0
FECHA_VACUNACION 0
DOSIS            0
FABRICANTE       0
PROVINCIA        0
DISTRITO         0
TIPO_EDAD;      0
dtype: int64

```

Fig. 16. Set de datos limpios, no existen valores perdidos o vacíos.

Como los datos eran muy pocos para una analítica de datos se utilizó el archivo `vacuna.covid`, el mismo que tuvo que ser cargado en la plataforma de *Rstudio*, ya que con Python no se podía cargar por el tamaño. En *Rstudio* se vio la forma de filtrar solo los datos para el DEPARTAMENTO de Ica.

```

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
+ [Go to file/function] [Addins]
conversion.R* x [Untitled1* x] [vacunas x]
Source on Save
1 # importar base de datos con extension csv
2 file.choose()
3 vacunas <- "D:\\Arostegui Katherine\\VACUNAS_COVID\\VACUNAS_COVID\\vacunas_covid.csv"
4 view(vacunas)
5
6
7 #install.packages("tidyverse")_
8 # library(tidyverse)
9
10 library(dplyr)
11 vacunaIca <- vacunas %>%
12   filter(DEPARTAMENTO=="ICA")
13
14 write.table(vacunaIca,"vacunasOK.csv",quote=FALSE, sep=";")
15 view(vacunas)
16
17

```

Fig. 17. Extracción de los datos en R

Para la extracción de los datos cargados en la plataforma del software *RStudio*, en la Fig. 17, en la tres primeras líneas iniciales, se activa la opción de archivos (`file.choose()`), luego se asigna a una variable los datos que se a cargar, indicando la ruta donde se encuentra dicho archivo. El nombre inicial del archivo cargado fue de `vacunas`; una vez cargado los datos se realiza una visualización de los datos del archivo.

Realizada la carga del archivo se tuvo que primero cargar instalar la librería *tidyverse*, y luego ponerla en uso para su aplicación.

Al tener la librería en uso, se procedió a crear el código necesario para la selección de los datos que correspondan solo para el departamento de ICA como se muestra en la opción del filtrado, los datos filtrados se almacenaron en la variable `VacunaIca`, datos que provienen del archivo cargado `Vacunas`.

Finalmente, los datos almacenados en la variable `VacunaIca`, deben ser almacenados en un archivo definitivo al que se le denominó `vacunasOk` en el formato CSV.

Al archivo se quitó las columnas no necesarias, se cambió los tipos de datos para tener el *Dataset* listo para el análisis.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1048554 entries, 0 to 1048553
Data columns (total 9 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   DISTRITO              1048554 non-null object
 1   GRUPO_RIESGO          1048554 non-null object
 2   DOSIS                 1048554 non-null int64
 3   FABRICANTE            1048554 non-null object
 4   TIPO_EDAD             1048554 non-null object
 5   PROVINCIA             1048554 non-null object
 6   FECHA_VACUNACION     1048554 non-null datetime64[ns]
 7   SEXO                  1048554 non-null object
 8   EDAD                  1048554 non-null int64
dtypes: datetime64[ns](1), int64(2), object(6)
memory usage: 72.0+ MB
```

En la información de los datos de la vacunación, se puede apreciar que existen al final en la última columna el tipo de datos: datos de tipo Object, int64, datetime64, los datos del tipo Object representan datos de tipo categórico, datos del tipo int64, representan valores numéricos, y el dato datetime64, valores de fecha.

	DISTRITO	GRUPO_RIESGO	DOSIS	FABRICANTE	TIPO_EDAD	PROVINCIA	FECHA_VACUNACION	SEXO	EDAD
0	PUEBLO NUEVO	PERSONAS DE 20 A 29 AÑOS	3	PFIZER	A	ICA	2022-08-03	FEMENINO	22
1	SALAS	PERSONAS DE 12 A 17 AÑOS	1	PFIZER	A	ICA	2021-10-31	MASCULINO	15
2	PISCO	PERSONAS DE 18 A 19 AÑOS	2	SINOPHARM	A	PISCO	2021-12-27	FEMENINO	19
3	SAN JOSE DE LOS MOLINOS	PERSONAS DE 30 A 39 AÑOS	1	SINOPHARM	A	ICA	2021-08-28	MASCULINO	36
4	PUEBLO NUEVO	PERSONAS DE 40 A 49 AÑOS	1	SINOPHARM	A	ICA	2021-09-28	MASCULINO	41

Fig. 18. Datos iniciales del *Dataset* final

En la Fig. 18, se realiza una visualización de los datos iniciales del archivo en la que se puede comprobar que solo existen datos de la región Ica, además de los primeros 5 datos de los registros de la vacunación del COVID-19.

	DISTRITO	GRUPO_RIESGO	DOSIS	FABRICANTE	TIPO_EDAD	PROVINCIA	FECHA_VACUNACION	SEXO	EDAD
1048549	ICA	PERSONAS DE 30 A 39 AÑOS	1	ASTRAZENECA	A	ICA	2021-11-16	FEMENINO	36
1048550	SANTIAGO	PERSONAS DE 40 A 49 AÑOS	3	ASTRAZENECA	A	ICA	2022-01-26	MASCULINO	42
1048551	ICA	PERSONAS CON TRASTORNOS MENTALES Y DEL NEURODE...	2	PFIZER	A	ICA	2021-09-10	FEMENINO	16
1048552	ICA	OBESIDAD TIPO II (IMC DE 35 A 39.9)	3	PFIZER	A	ICA	2022-12-04	FEMENINO	14
1048553	ICA	PERSONAS DE 30 A 39 AÑOS	2	SINOPHARM	A	ICA	2021-06-10	MASCULINO	32

Fig. 19. Datos finales del *Dataset* final

En la Fig. 19, se realiza una visualización de los datos finales del archivo en la que se puede comprobar que al igual que la Fig. 18, los datos solo corresponden a la región Ica, además de los últimos 5 datos de los registros de las variables de interés de la vacunación del COVID-19.

Es importante resaltar en esta etapa de la metodología que no ha sido necesario realizar otros tipos de técnicas para la preparación de los datos como es la *imputación* de datos para casos en la cual la base de datos tiene datos perdidos o vacíos, cosa que no es el caso. Esto se debe a que el MINSA ya ha puesto a disposición los datos debidamente tratados y sin ningún tipo de estas limitaciones o deficiencias, por lo cual se tiene asegurada la calidad de los datos.

2.4.3. Análítica de los datos, comunicación y visualización

Para el análisis descriptivo de los datos se empleó la librería de *Sweetviz*, una librería especializada que nos ayuda en la exploración y visualización interactiva actividad que se realiza para el análisis exploratorio de los datos (EDA).

Esta librería para el procesamiento de los datos EDA. En el procesamiento esta librería genera un archivo HTML con los resultados y que pueden ser explorados de manera fácil según el deseo del analista de datos.

Es importante tener en consideración que primero hay que instalar la librería antes de usarse, a continuación, se detalla las sentencias para su instalación su uso hasta la generación del archivo HTML y su visualización.

```
#instalación de la librería
!pip install sweetviz
```

Las líneas de código realizan el proceso de instalación de la librería, si la instalación es satisfactoria los resultados se mostrarán según la Fig.20.

```
Collecting sweetviz
  Downloading sweetviz-2.2.1-py3-none-any.whl (15.1 MB)
    15.1/15.1 MB 31.9 MB/s eta 0:00:00
Requirement already satisfied: pandas!=1.0.0,!=1.0.1,!=1.0.2,>=0.25.3 in /usr/local/lib/python3.10/dist-packages (from sweetviz) (1.5.3)
Requirement already satisfied: numpy>=1.16.0 in /usr/local/lib/python3.10/dist-packages (from sweetviz) (1.23.5)
Requirement already satisfied: matplotlib>=3.1.3 in /usr/local/lib/python3.10/dist-packages (from sweetviz) (3.7.1)
Requirement already satisfied: tqdm>=4.43.0 in /usr/local/lib/python3.10/dist-packages (from sweetviz) (4.66.1)
Requirement already satisfied: scipy>=1.3.2 in /usr/local/lib/python3.10/dist-packages (from sweetviz) (1.10.1)
Requirement already satisfied: Jinja2>=2.11.1 in /usr/local/lib/python3.10/dist-packages (from sweetviz) (3.1.2)
Requirement already satisfied: importlib-resources>=1.2.0 in /usr/local/lib/python3.10/dist-packages (from sweetviz) (6.0.1)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from Jinja2>=2.11.1->sweetviz) (2.1.3)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3->sweetviz) (1.1.0)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3->sweetviz) (0.11.0)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3->sweetviz) (4.42.1)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3->sweetviz) (1.4.4)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3->sweetviz) (23.1)
Requirement already satisfied: pillow>=6.2.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3->sweetviz) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3->sweetviz) (3.1.1)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.10/dist-packages (from matplotlib>=3.1.3->sweetviz) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas!=1.0.0,!=1.0.1,!=1.0.2,>=0.25.3->sweetviz) (2023.3)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7->matplotlib>=3.1.3->sweetviz) (1.16.0)
Installing collected packages: sweetviz
Successfully installed sweetviz-2.2.1
```

Fig. 20. Instalación de la librería sweetviz

Instalación de las librerías para el análisis exploratorio de los datos

Una vez instalada la librería esta debe ser cargada en el *notebook de Python*, para luego crear el reporte, las sentencias siguientes realizan dichas actividades.

```
#carga de la librería
import sweetviz as sv

# creación del reporte analitico
report=sv.analyze(vacuna)
report.show_html()
```

```
Done! Use 'show' commands to display/save. [100%] 00:01 -> (00:00 left)
Report SWEETVIZ_REPORT.html was generated! NOTEBOOK/COLAB USERS: the web browser MAY not pop up, regardless, the report IS saved in your notebook/colab files.
```

Fig. 21. Reporte analítico creado.

En la Fig. 21, se comprueba que la librería ha generado el reporte de forma satisfactoria con la culminación del reporte al 100%. Este reporte en formato HTML que se ha creado se debe descargar de drive de Google en donde se ha guardado (en la ruta inicial).

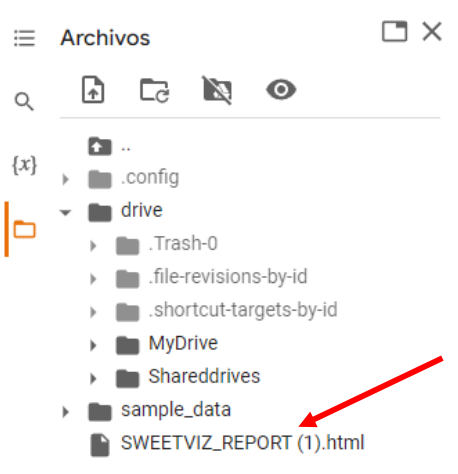


Fig. 22 Archivo HTML generado por la librería
En la sección de las carpetas de nuestro *Google drive*, se muestra el archivo del análisis exploratorio de los datos en el formato html para su posterior análisis.

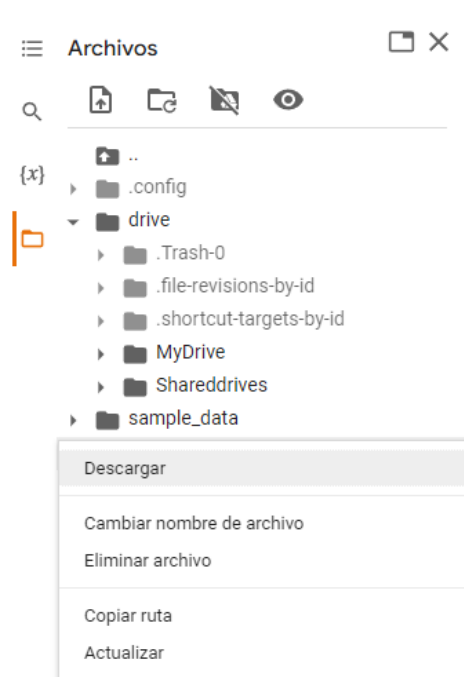


Fig. 23. Descarga del archivo HTML

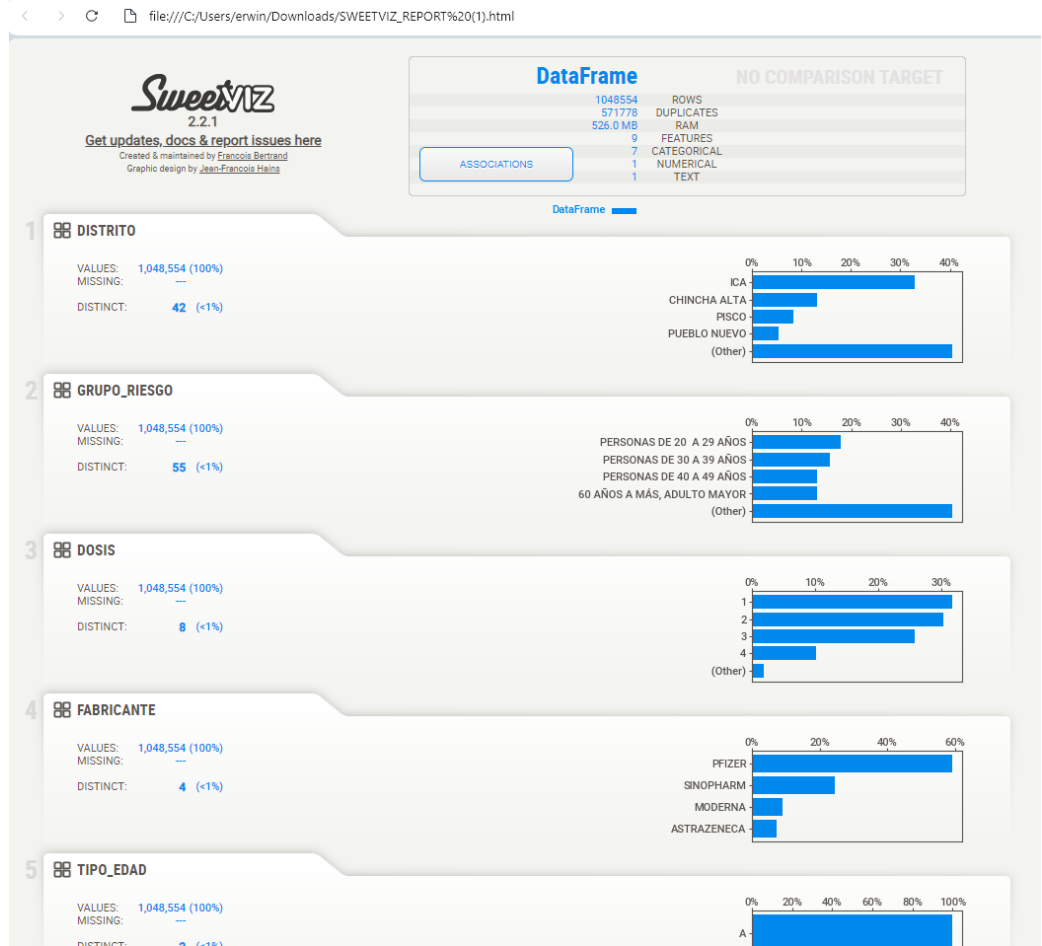


Fig. 24. Reporte generado por la librería Sweetviz

Como se puede apreciar en la Fig. 24, el reporte generado tiene una estadística de cada una de las variables que componen la base de datos. el análisis generado por la librería corresponde a un análisis univariado, y que pasamos a detallar a continuación.

Como se presenta igualmente en la Fig. 24, el análisis reportado por este archivo presenta una doble presentación para que se pueda utilizar a conveniencia del investigador, por una parte, el análisis gráfico y de la otra un análisis de frecuencias que se mostraran en el detalle más adelante.

Para ver el análisis individual de cada variable solo se debe dar clic en la variable a analizar y se mostrara el análisis correspondiente.

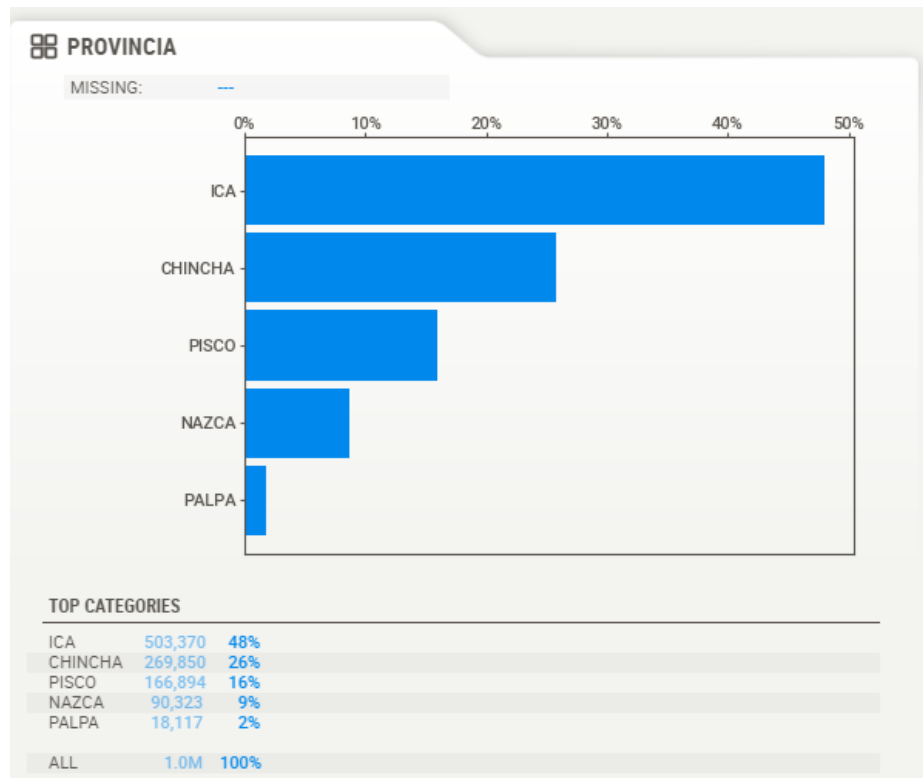


Fig. 25. Análisis de frecuencia de las dosis por cada provincia

Como se mencionó líneas arriba se muestra un resultado gráfico y otro resultado de las frecuencias para un análisis más detallado para el investigador.

En el reporte sobre las vacunas en cada provincia de la Fig.25, se puede ver que entre las provincias de Ica, Chincha y Pisco se han vacunado el 90% de las personas para el periodo estudiado, lo que se refleja igualmente en la gráfica adjunta.

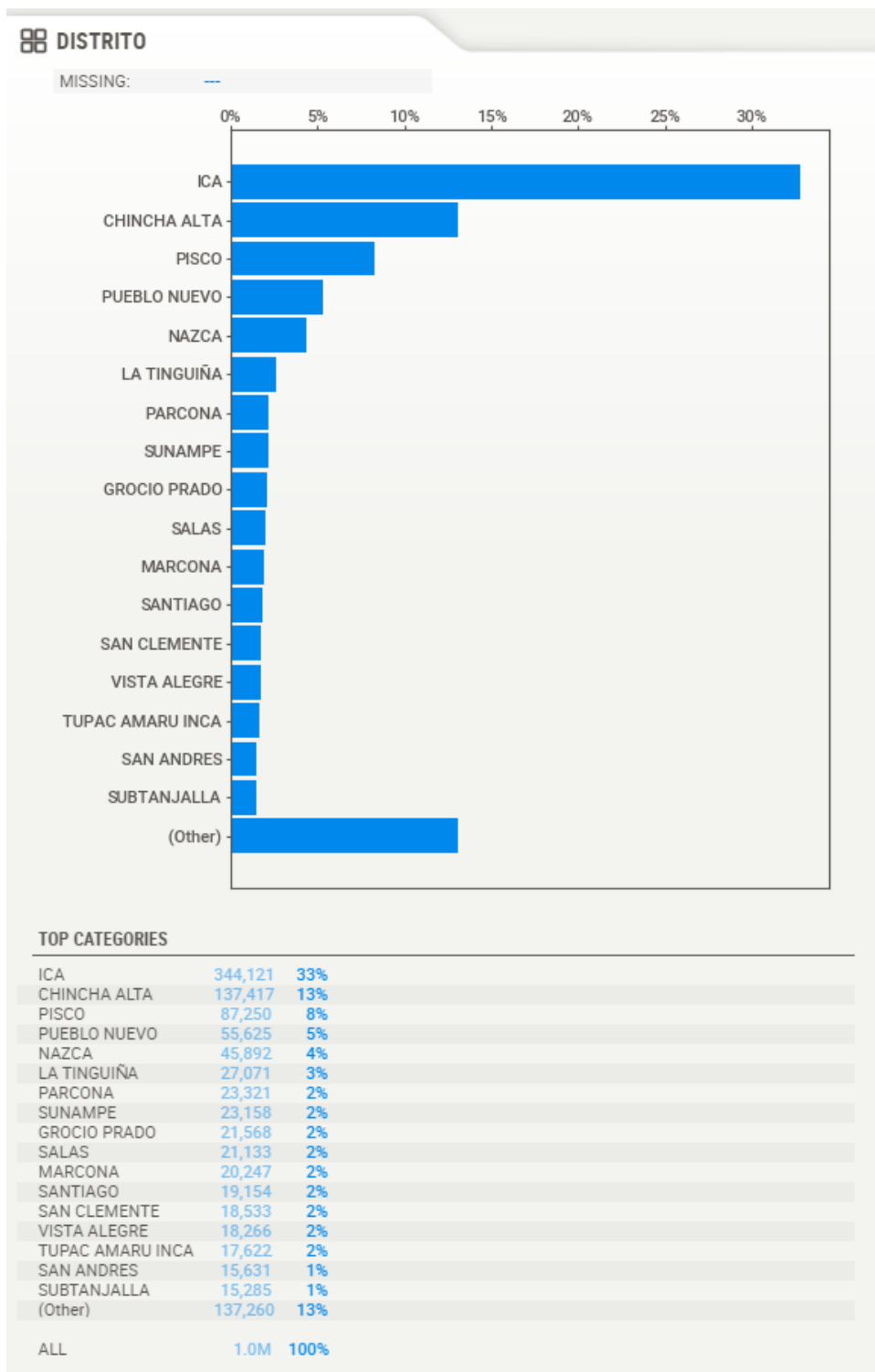
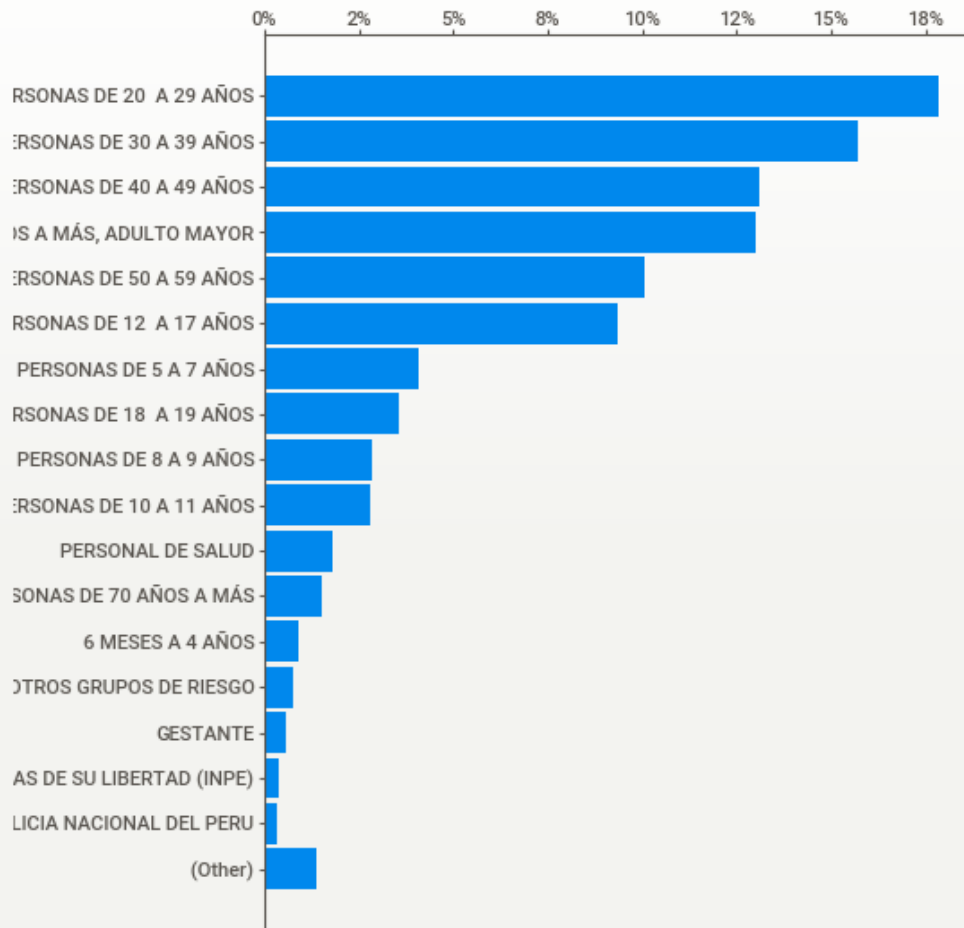


Fig. 26. Vacunación en los distritos de la región

En la Fig. 26, se puede apreciar que las provincias con mayor cantidad de vacunación son la provincia de Ica y Chincha Alta con el 46% de las personas vacunadas, seguido de la provincia de Pisco con solo el 8% de los vacunados, pero que son las más representativas.

GRUPO_RIESGO

MISSING: ---



TOP CATEGORIES

PERSONAS DE 20 A 29 AÑOS	187,284	18%
PERSONAS DE 30 A 39 AÑOS	164,814	16%
PERSONAS DE 40 A 49 AÑOS	137,423	13%
60 AÑOS A MÁS, ADULTO MAYOR	136,516	13%
PERSONAS DE 50 A 59 AÑOS	105,650	10%
PERSONAS DE 12 A 17 AÑOS	98,025	9%
PERSONAS DE 5 A 7 AÑOS	42,855	4%
PERSONAS DE 18 A 19 AÑOS	37,458	4%
PERSONAS DE 8 A 9 AÑOS	29,675	3%
PERSONAS DE 10 A 11 AÑOS	29,173	3%
PERSONAL DE SALUD	18,766	2%
PERSONAS DE 70 AÑOS A MÁS	15,913	2%
6 MESES A 4 AÑOS	9,343	<1%
OTROS GRUPOS DE RIESGO	8,019	<1%
GESTANTE	5,820	<1%
PERSONAS PRIVADAS DE SU LIBERTAD (I...	4,050	<1%
POLICIA NACIONAL DEL PERU	3,395	<1%
(Other)	14,375	1%
ALL	1.0M	100%

Fig. 27. Vacunación según los grupos de riesgo

Las vacunas según el grupo de riesgo, se tiene que los jóvenes entre 20 a 29 años representan el mayor porcentaje con 18% de los vacunados, y por otro lado se resalta la vacunación de las personas privadas de su libertad y la Policía Nacional del Perú con menos del 1% de los vacunados, lo que representa una población muy reducida en relación con el total, pero que igualmente es importante porque se tiene una población vulnerable que también ha sido vacunada.

Es interesante conocer que son 3,395 policías en la región que se han vacunado, siendo un número muy pequeño para la población total.

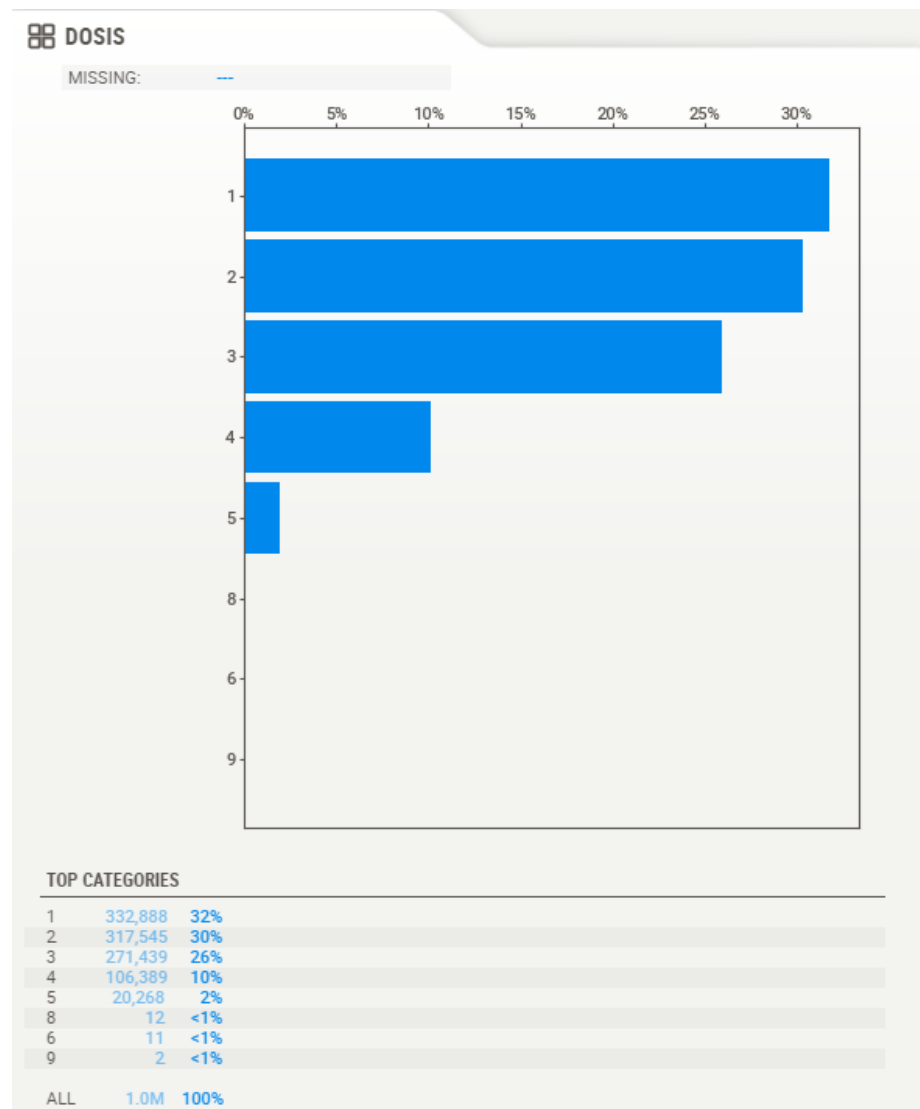


Fig. 28. Vacunación según las dosis aplicadas.

Aquí en este análisis es importante encontrar que, según las frecuencias generadas, solo el 10% había cumplido con la cuarta dosis de vacunación, y un porcentaje muy grande del

88% no cumplió con la obligación de ponerse la cuarta (4ta) dosis, a pesar de las restricciones que se establecieron.

En este mismo análisis se resalta que existen personas que se aplicaron más de las dosis obligatorias con 5, 6, 8 y hasta 9 vacunas dos personas. Es importante esto, porque ha habido personas que se vacunaron por encima de lo establecido en la cantidad de vacunas determinadas por el MINSA.

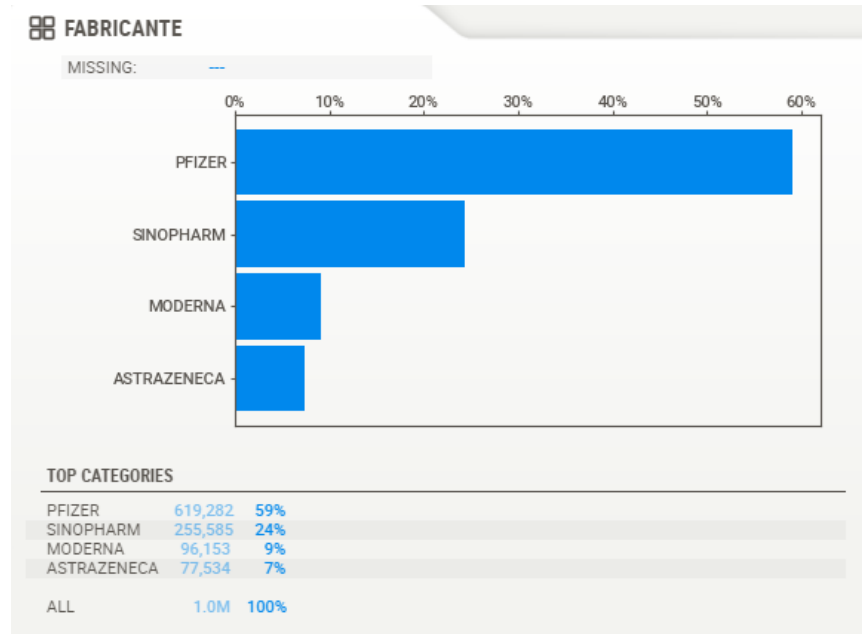


Fig. 29. Vacunación según el fabricante

Las personas que se han vacunado mayormente han sido con las vacunas del fabricante Pfizer con 59% y SinoPharm con 24%, estos dos fabricantes representan más del 80% de la vacunación total y la vacuna de aztrazeneca la de menor aplicación con solo 7% de las vacunas utilizadas.

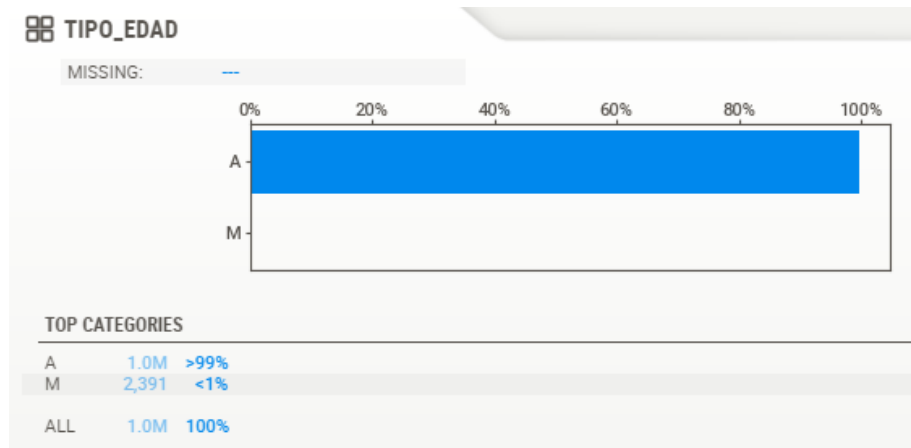


Fig. 30. Vacunación según el tipo de edad

Según lo registrado por el MINSA A(adultos), M(Menores), se puede ver en la Fig. 28 que los adultos representan casi la totalidad de los vacunados mayor a 99%, y menos del 1% menores vacunados.

En este primer análisis de los datos, a pesar de ser una buena herramienta para el análisis de los datos, la librería no cubre con posibilidad de hacer otros análisis más completos como el análisis bivariado o también un análisis multivariado, por lo que para este tipo de análisis utilizamos la herramienta para minería de datos Orange Datamining. A continuación, presentamos el análisis de sus variables.

	FABRICANTE				Total	
	Sum	ASTRAZENECA	MODERNA	PFIZER SINOPHARM		
PROVINCIA	CHINCHA	48255.0	98896.0	357019.0	93586.0	597756.0
	ICA	80611.0	179999.0	680322.0	179040.0	1119972.0
	NAZCA	14060.0	25822.0	118671.0	38583.0	197136.0
	PALPA	2546.0	7008.0	26156.0	5997.0	41707.0
	PISCO	28974.0	46871.0	211436.0	65519.0	352800.0
	Total	174446.0	358596.0	1393604.0	382725.0	2309371.0

Fig. 31. Dosis aplicadas por fabricante en cada provincia

Las dosis aplicadas según la Fig.31, muestra que las provincias donde se han aplicado la mayor cantidad de dosis son en Ica con 1'119,972, seguido de la provincia de Chincha con poco más de la mitad de las vacunaciones en relación con la provincia de Ica, siendo esta 597,756 dosis y la provincia de Pisco con 352,800 dosis. En Ica se destaca los dos

tipos de vacunas como la de Moderna con 179,999 vacuna y Sinopharm con 179,040 vacunas y muy por debajo de esta la vacuna de Pfizer que representaba la mejor opción dentro de las vacunas. En menor proporción siguen esta misma tendencia en la provincia de Chincha.

		FABRICANTE					
		Sum	ASTRAZENECA	MODERNA	PFIZER	SINOPHARM	Total
SEXO	FEMENINO	82850.0	191596.0	716754.0	185271.0		1176471.0
	MASCULINO	91596.0	167000.0	676850.0	197454.0		1132900.0
	Total	174446.0	358596.0	1393604.0	382725.0		2309371.0

Fig. 32. Dosis aplicadas por fabricante según el sexo

Con relación al sexo se han aplicado al sexo masculino 676,850 dosis del fabricante Pfizer y para el sexo femenino 716,754 dosis de esta misma vacuna, de las 2'309,371 dosis totales aplicadas. Siendo la vacunación de menor presencia la del fabricante Astrazeneca para ambos sexos.

FABRICANTE

	Sum	ASTRAZENECA	MODERNA	PFIZER	SINOPHARM	Total
ALTO LARAN	1271.0	2570.0	10404.0	3403.0	17648.0	
CHANGUILLO	9.0	391.0	2393.0	1440.0	4233.0	
CHAVIN	63.0	168.0	583.0	267.0	1081.0	
CHINCHA ALTA	28371.0	60046.0	187258.0	44296.0	319971.0	
CHINCHA BAJA	1592.0	3997.0	16303.0	4897.0	26789.0	
EL CARMEN	1711.0	3350.0	15391.0	5094.0	25546.0	
EL INGENIO	9.0	1191.0	4720.0	2092.0	8012.0	
GROCIO PRADO	3042.0	6160.0	27026.0	8184.0	44412.0	
HUANCANO	119.0	373.0	1893.0	928.0	3313.0	
HUMAY	929.0	1756.0	7800.0	2466.0	12951.0	
ICA	62319.0	141244.0	477848.0	114876.0	796287.0	
INDEPENDENCIA	2335.0	3355.0	15651.0	4989.0	26330.0	
LA TINGUIÑA	3115.0	7451.0	34086.0	10222.0	54874.0	
LLIPATA	329.0	1054.0	3330.0	591.0	5304.0	
LOS AQUIJES	500.0	1773.0	10830.0	3884.0	16987.0	
MARCONA	3557.0	4601.0	24876.0	9101.0	42135.0	
NAZCA	7713.0	16519.0	63224.0	17717.0	105173.0	
OCUCAJE	174.0	664.0	4209.0	2926.0	7973.0	
PACHACUTEC	1282.0	2697.0	10998.0	3018.0	17995.0	
PALPA	1592.0	3765.0	14277.0	3521.0	23155.0	
PARACAS	1349.0	1797.0	7807.0	3576.0	14529.0	
PARCONA	2128.0	2983.0	30455.0	9275.0	44841.0	
PISCO	15629.0	29487.0	117215.0	31260.0	193591.0	
PUEBLO NUEVO	9545.0	20107.0	72219.0	19161.0	121032.0	
RIO GRANDE	429.0	1404.0	5440.0	1237.0	8510.0	
SALAS	4649.0	1336.0	22842.0	9913.0	38740.0	
SAN ANDRES	1967.0	3381.0	20132.0	6270.0	31750.0	
SAN CLEMENTE	3346.0	2798.0	20139.0	8619.0	34902.0	
SAN JOSE DE LOS MOLINOS	766.0	2640.0	10089.0	2508.0	16003.0	

DISTRITO

SAN JUAN BAUTISTA	309.0	936.0	9451.0	2987.0	13683.0
SAN JUAN DE YANAC	48.0	120.0	523.0	457.0	1148.0
SAN PEDRO DE HUACARPANA	15.0	118.0	879.0	393.0	1405.0
SANTA CRUZ	157.0	722.0	2828.0	548.0	4255.0
SANTIAGO	2529.0	4425.0	23366.0	8528.0	38848.0
SUBTANJALLA	1079.0	4358.0	22452.0	5238.0	33127.0
SUNAMPE	2931.0	6694.0	31366.0	7896.0	48887.0
TAMBO DE MORA	422.0	606.0	6753.0	1894.0	9675.0
TATE	858.0	3799.0	9688.0	2054.0	16399.0
TIBILLO	39.0	63.0	281.0	100.0	483.0
TUPAC AMARU INCA	3300.0	3924.0	20799.0	7411.0	35434.0
VISTA ALEGRE	2772.0	3120.0	23458.0	8233.0	37583.0
YAUCA DEL ROSARIO	147.0	653.0	2322.0	1255.0	4377.0
Total	174446.0	358596.0	1393604.0	382725.0	2309371.0

Fig. 33. Dosis aplicadas por fabricante en los distritos

La Fig. 33, muestra la aplicación de las dosis aplicadas en los pacientes según el fabricante en cada uno de los distritos de la región de Ica. Destaca el distrito de Ica con 477,848 vacunas del fabricante Pfizer, seguido del fabricante Moderna con 141,244 dosis aplicadas de las 796,287 vacunas aplicadas, representando esta la mayor cantidad de vacunas puestas.

PROVINCIA	SEXO			
	Aggregate	FEMENINO	MASCULINO	Total
CHINCHA	Min	1.0	1.0	1.0
	Max	6.0	9.0	9.0
ICA	Min	1.0	1.0	1.0
	Max	8.0	9.0	9.0
NAZCA	Min	1.0	1.0	1.0
	Max	8.0	8.0	8.0
PALPA	Min	1.0	1.0	1.0
	Max	5.0	5.0	5.0
PISCO	Min	1.0	1.0	1.0
	Max	8.0	5.0	8.0
Total	Min	1.0	1.0	1.0
	Max	8.0	9.0	9.0

Fig. 34. Dosis máximas y mínimas aplicadas por sexo en cada provincia

La Fig. 34, muestra que ha habido personas que se han vacunado hasta con 9 dosis. En Ica en el sexo femenino se han vacuna hasta 8 dosis y en el sexo masculino hasta 9 dosis, en el caso de Chincha el sexo femenino se reporta casos de hasta 6 vacunas y en el sexo masculino 9 vacunas, en Pisco entre 5 a 8 vacunas como máximo.

	FABRICANTE				Total
	Sum	ASTRAZENECA	MODERNA	PFIZER SINOPHARM	
6 MESES A 4 AÑOS	0.0	12158.0	0.0	0.0	12158.0
60 AÑOS A MÁS, ADULTO MAYOR	2113.0	58682.0	251498.0	5514.0	317807.0
BOMBERO	33.0	944.0	371.0	497.0	1845.0
BRIGADISTAS	0.0	4.0	815.0	12.0	831.0
CASA REPOSO, DE REHABILITACIÓN Y ALBERGUES	0.0	0.0	7.0	2.0	9.0
COLEGIOS PROFESIONALES	201.0	4.0	184.0	664.0	1053.0
CRUZ ROJA	0.0	0.0	54.0	148.0	202.0
DIABETES TIPO I Y TIPO II	0.0	344.0	933.0	6.0	1283.0
ENFERMEDADES CEREBROVASCULARES	0.0	0.0	5.0	0.0	5.0
ESTUDIANTES DE CIENCIAS DE LA SALUD	0.0	68.0	68.0	14.0	150.0
FISCALIZADORES	0.0	0.0	91.0	0.0	91.0
GESTANTE	6.0	64.0	9308.0	118.0	9496.0
INTERNOS DE CIENCIAS DE LA SALUD	6.0	16.0	701.0	53.0	776.0
MIEMBROS DE LAS MESAS ELECTORALES	4.0	0.0	12.0	0.0	16.0
OBESIDAD TIPO II (IMC DE 35 A 39.9)	6.0	516.0	1388.0	2.0	1912.0
OBESIDAD TIPO III (IMC MAYOR O IGUAL A 40)	14.0	24.0	383.0	8.0	429.0
OLÍMPICOS O PARAOLÍMPICOS	8.0	0.0	132.0	139.0	279.0
OTROS GRUPOS DE RIESGO	1129.0	558.0	8009.0	6336.0	16032.0
PACIENTE ESPLÉNICOS O CON TRANSPLANTE DE ORGANOS SÓLIDOS	0.0	0.0	4.0	0.0	4.0
PACIENTE ONCOLÓGICO	0.0	80.0	1400.0	5.0	1485.0
PACIENTE ONCOLÓGICO CON TRATAMIENTO ACTIVO	0.0	4.0	20.0	0.0	24.0
PACIENTE ONCOLÓGICO SIN TRATAMIENTO ACTIVO	0.0	4.0	15.0	0.0	19.0
PACIENTE RECIBIENDO TRATAMIENTO CON INMUNOSUPRESIÓN SEVERA	0.0	0.0	269.0	0.0	269.0
TRUCTIVA CRÓNICA), FIBROSIS QUÍSTICA, FIBROSIS PULMONAR, HIPERTENCIÓN PULMONAR, ASMA GRAVE O NO CONTROLADA.	0.0	4.0	88.0	0.0	92.0
PACIENTES CON HEMODIÁLISIS Y ENFERMEDAD CRÓNICA RENAL	0.0	44.0	447.0	2.0	493.0
PACIENTES ESPLÉNICOS Y ONCOHEMATOLÓGICOS C/TRANSPLANTE DE ÓRGANO SÓLIDO Y PRECURSORES HEMATOPOYÉTICOS	0.0	0.0	4.0	0.0	4.0
PERSONA VIVIENDO CON VIH	3.0	64.0	1654.0	20.0	1741.0
PERSONAL DE SALUD	1203.0	6971.0	21862.0	15939.0	45975.0

PERSONAL DE SEGURIDAD	1.0	36.0	136.0	134.0	307.0
PERSONAL MILITAR Ó FFAA	20.0	528.0	1046.0	2.0	1596.0
PERSONAL QUE LABORA EN INPE	3.0	36.0	382.0	12.0	433.0
PERSONAS CON ENFERMEDADES RARAS Y HUÉRFANAS	3.0	2599.0	3164.0	14.0	5780.0
PERSONAS CON TBC	0.0	8.0	116.0	0.0	124.0
PERSONAS CON TRASTORNOS MENTALES Y DEL NEURODESARROLLO	0.0	232.0	1673.0	14.0	1919.0
PERSONAS DE 5 A 7 AÑOS	0.0	317.0	69566.0	0.0	69883.0
PERSONAS DE 8 A 9 AÑOS	0.0	235.0	51230.0	0.0	51465.0
PERSONAS DE 10 A 11 AÑOS	0.0	267.0	50913.0	0.0	51180.0
PERSONAS DE 12 A 17 AÑOS	158.0	512.0	187268.0	79.0	188017.0
PERSONAS DE 18 A 19 AÑOS	27327.0	5672.0	40225.0	7600.0	80824.0
PERSONAS DE 20 A 29 AÑOS	56228.0	35304.0	241356.0	69076.0	401964.0
PERSONAS DE 30 A 39 AÑOS	43086.0	51238.0	149778.0	123198.0	367300.0
PERSONAS DE 40 A 49 AÑOS	29670.0	65552.0	122653.0	103904.0	321779.0
PERSONAS DE 50 A 59 AÑOS	12367.0	64497.0	136555.0	45331.0	258750.0
PERSONAS DE 70 AÑOS A MÁS	205.0	43542.0	21773.0	72.0	65592.0
PERSONAS EN ESPERA O CON TRASPLANTE DE ÓRGANO	3.0	20.0	36.0	2.0	61.0
PERSONAS PRIVADAS DE SU LIBERTAD (INPE)	481.0	5391.0	8637.0	292.0	14801.0
POBLACION INDIGENA	0.0	5.0	102.0	6.0	113.0
POLICIA NACIONAL DEL PERU	12.0	1716.0	5234.0	108.0	7070.0
PROFESORES MINEDU	86.0	124.0	367.0	2648.0	3225.0
PSORIASIS, ARTRITIS PSORIÁSICA, ARTRITIS REUMATOIDE	0.0	8.0	745.0	0.0	753.0
SERENAZGO	2.0	0.0	38.0	2.0	42.0
SIN GRUPO DE RIESGO	0.0	0.0	0.0	3.0	3.0
SÍNDROME DOWN	0.0	84.0	666.0	2.0	752.0
TRABAJADOR Ó PERSONAL DE LIMPIEZA	5.0	120.0	100.0	720.0	945.0
ZONAS ALTOANDINAS	63.0	0.0	123.0	27.0	213.0
Total	174446.0	358596.0	1393604.0	382725.0	2309371.0

Fig. 35. Dosis aplicadas por fabricante según grupos de riesgo

El análisis que se reporta según los ítems determinados por el MINSA, de que existen una diversidad muy grande asignados a los grupos de riesgo, *consideramos que todos estos ítems, se pueden simplificar.*

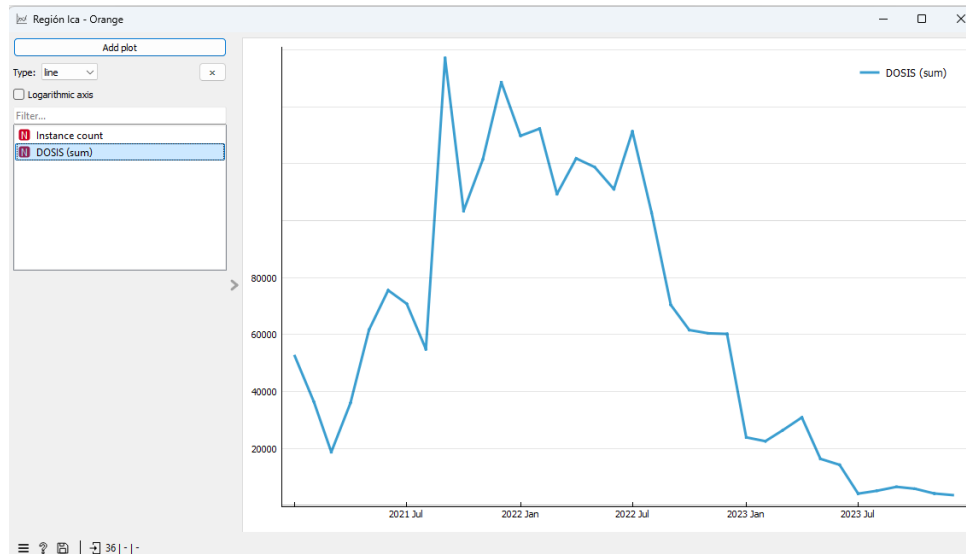


Fig. 36. Evolución de la vacunación a nivel de la región Ica

En el periodo estudiado la Fig. 36, muestra la evolución de la vacunación, con un pico muy alto de 157073 vacunas en el mes de septiembre del 2021 manteniéndose con altibajos hasta el mes de Julio del 2022; a partir de esa fecha ha ido reduciéndose paulatinamente hasta vacunaciones muy pequeñas desde Enero del 2023 con solo 23858 dosis.

Comportamiento de las dosis aplicadas a nivel de provincias

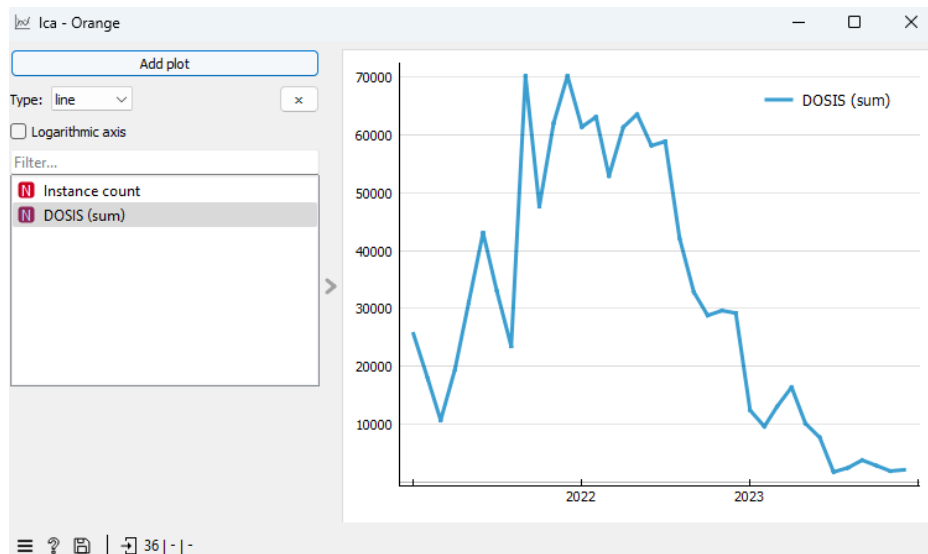


Fig. 37. Comportamiento de vacunación en la provincia de Ica

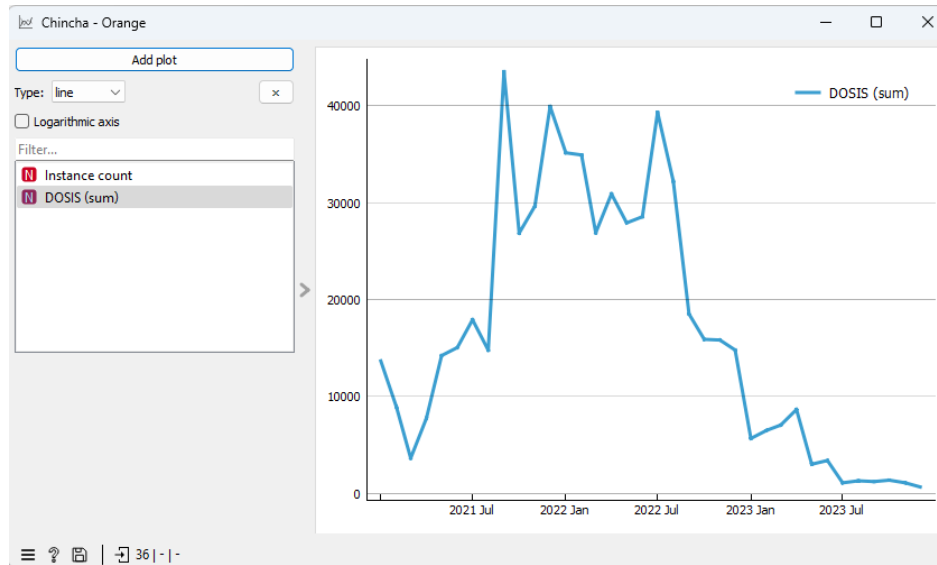


Fig. 38. Comportamiento de vacunación provincia de Chíncha

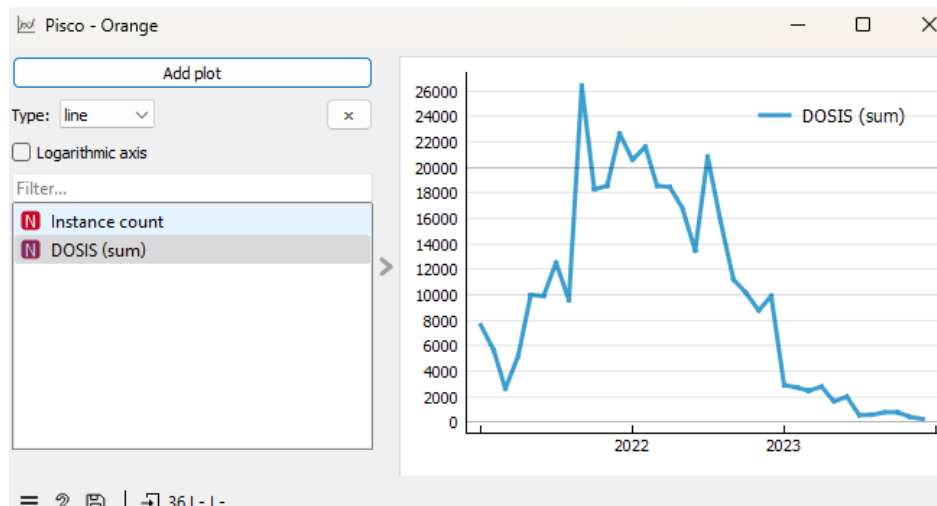


Fig. 39. Comportamiento de vacunación provincia de Pisco

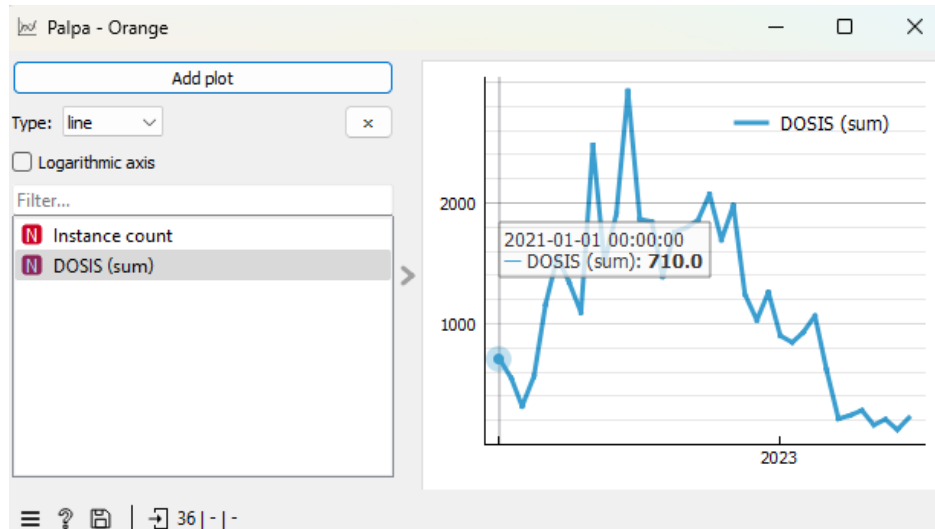


Fig. 40. Comportamiento de vacunación provincia de Palpa

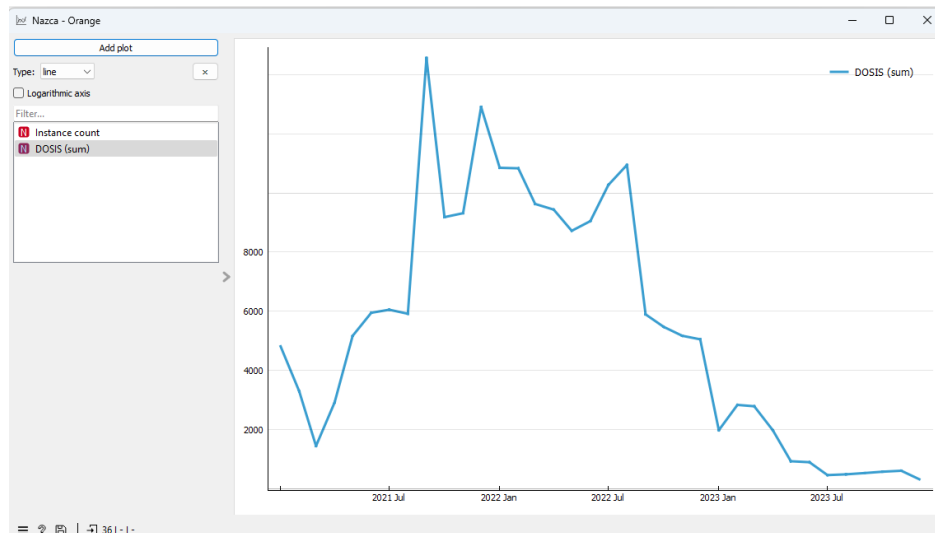


Fig. 41. Comportamiento de vacunación provincia de Nazca

Como se puede apreciar en las Figs. 37 a la 41, el comportamiento de la vacunación es similar en todas las provincias lo que significa que ha habido una buena distribución de estas además de una adecuada planificación de vacunación. Además, se puede apreciar como desde el inicio de la pandemia, la vacunación en la región y en todas las provincias tuvieron un inicio bajo de vacunaciones que se fue incrementando a fines del año 2021 y un pico en el primer semestre del 2021.

A continuación, se presenta el diseño del modelo realizado en la herramienta Orange Datamining.

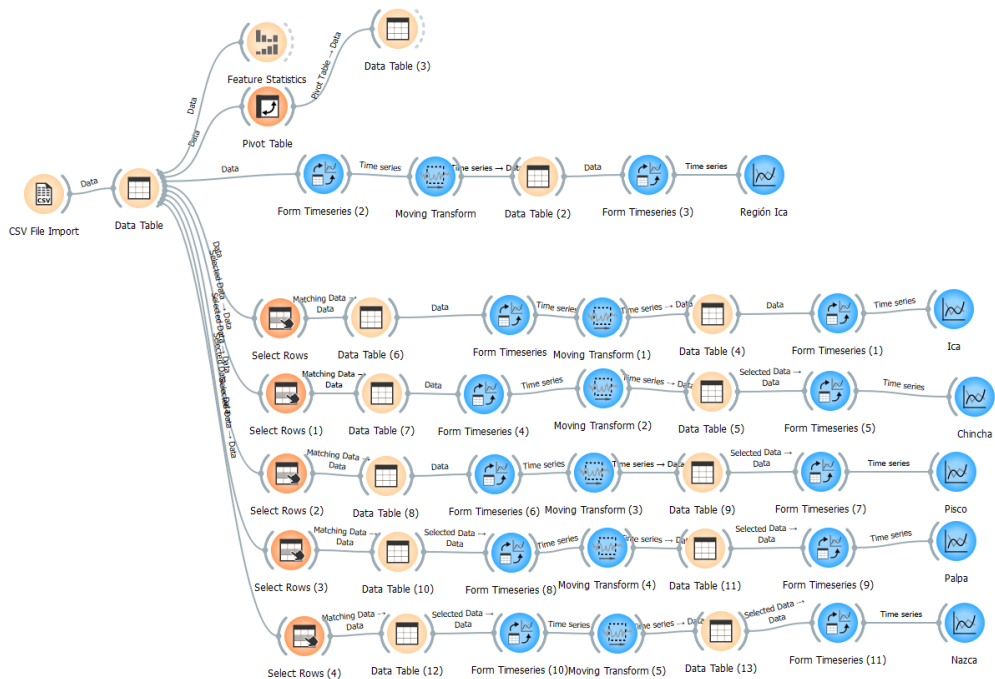


Fig. 42. Modelo de Widgets orange

El modelo presenta 4 líneas de análisis de minería de datos:

1. Línea de análisis descriptivo de los datos widget *Feature Statistics*
2. Línea de pivoteo de datos análisis bivariado widget *Pivote Table*
3. Línea Comportamiento de vacunación en la región Ica, *widgets Form Timeseries, Moving Transform y Line Chart.*
4. En el bloque inferior de las 5 líneas posterior, son las líneas de comportamiento de la vacunación en cada una de las provincias individualmente. Estas emplean los mismos widgets que la Línea 3.

En la Segunda Línea empleada para el análisis bivariado de las vacunaciones se realizó con el widget *Pivot Table* y los resultados se cargaban en el *widget Data Table* se muestra el uso de los *widgets*.

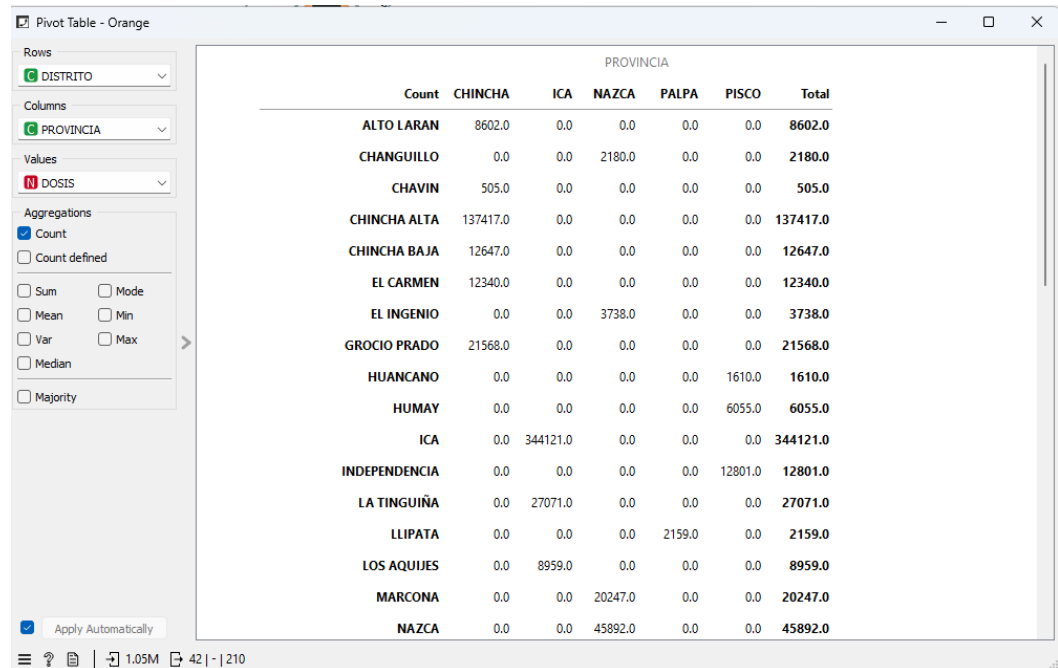


Fig. 43. Estructura del *widget Pivote Table*

La estructura del *widget Pivote Table* nos permite realizar el pivoteo de las variables a conveniencia. Como se muestra en la Fig. 38, en la parte izquierda de la ventana se puede seleccionar las variables que se deseen combinar para la columna y la fina del reporte. En la opción *values* debe seleccionarse la variable que se va a computar. Nuestro caso son las dosis.

En la parte inferior se tiene una variedad de posibles cálculos que se pueden realizar con la variable. Tomar en cuenta si la variable es numérica se podrá seleccionar cualquiera de los cálculos estadísticos descriptivos. Si la variable es categórica solo se podrá realizar conteos.

En la Tercera Línea ´del modelo se tiene el uso de los widgets para la creación del comportamiento de la vacunación considerándola como una serie de tiempos por ello se emplea el *widget Form Timeseries* para determinar la variable de fecha que se va a utilizar.

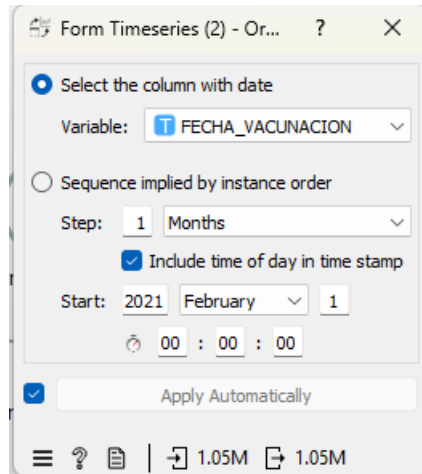


Fig. 44. Configuración de la variable de tiempo

En la Fig. 39, se pueden utilizar dos posibles variables de tiempo, por un lado, la variable tal cual como la fecha de vacunación o puede seleccionarse la secuencia de tiempos en la segunda opción.

En el siguiente *widget del Moving Transform*, para determinar el tipo de transformación que se van a dar los datos.

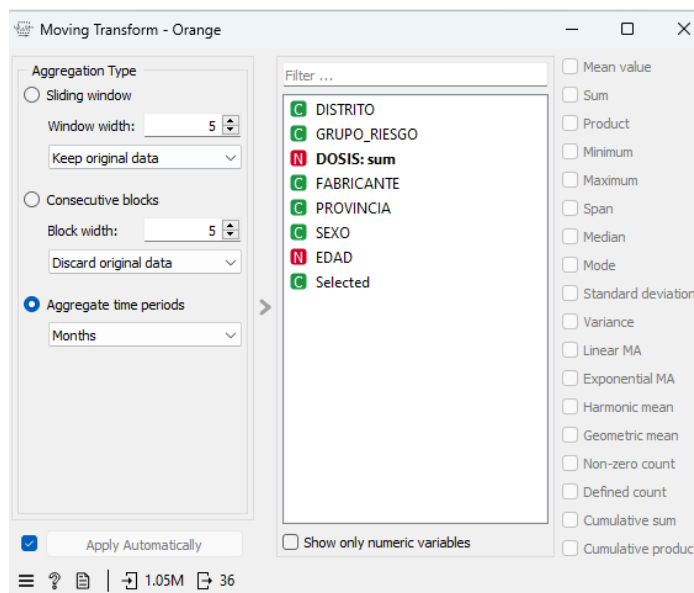


Fig. 45. Conteo de los datos de las dosis

En la Cuarta Línea del bloque de comportamiento de la vacunación para cada una de las provincias de la Región, el diseño es similar a la Tercera Línea, con la salvedad de que, de la base de datos cargada, esta se debe seleccionar solo los datos para cada provincia,

motivo por el cual se utiliza el widget *Select Rows* para seleccionar las columnas que corresponden solo con los datos de la provincia como se muestra en la Fig. 41.

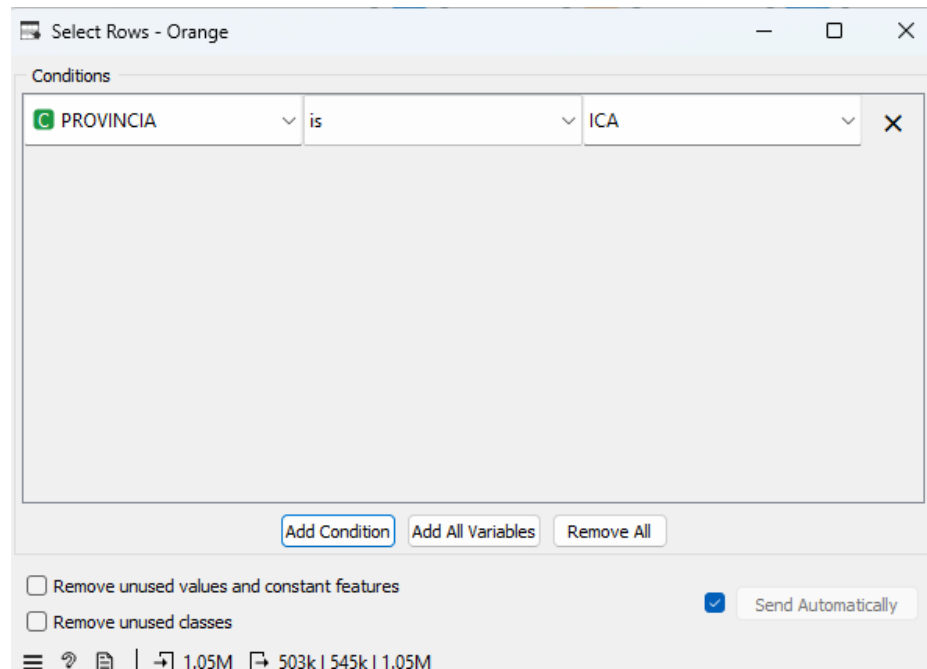


Fig. 46. Selección de las vacunaciones en la provincia de Ica.

Como se aprecia en la Fig. 41, se tiene que aplicar una condición para seleccionar los registros de la vacunación de la provincia correspondiente (en la Fig. 41 para el caso de provincia de Ica).

Lo importante es resaltar que solo se tiene que seleccionar la condición respectiva, no es necesario ningún tipo de código, de allí su facilidad de uso.

III. RESULTADOS

Siendo el estudio del nivel descriptivo, se realizará el análisis descriptivo de las variables en estudio. En el caso de las variables categóricas se realiza el análisis de frecuencias, y para el caso de las variables numéricas las medidas de tendencia central y las medidas de dispersión, con las cuales se debe obtener la tabla de los intervalos de confianza, el mismo que se realiza al 95%. Los resultados se muestran en las tablas siguientes.

3.1 Análisis de Frecuencias de las variables categóricas

TABLA III
FRECUENCIA DE VACUNAS SEGÚN FABRICANTE

Fabricante	Frecuencia	%
PFIZER	619,282	59,06
SINOPHARM	255,585	24,37
MODERNA	96,153	9,17
ASTRAZENECA	77,534	7,39
	1048,554	100,00

En la Tabla III, se puede apreciar que las vacunas predominantes en la vacunación son las de los fabricantes PFIZER y SINOPHARM con una 83.43% de los vacunados.

TABLA IV
FRECUENCIA DE VACUNAS SEGÚN SEXO

Sexo	Frecuencia	%
FEMENINO	528,537	50,41
MASCULINO	520,017	49,59
	1048,554	100,00

Con relación a los vacunado según el sexo, se puede ver que existen un equilibrio en las frecuencias de vacunas puestas a los del sexo Masculino y Femenino, similar al 50% para cada una.

TABLA V
FRECUENCIA DE VACUNAS POR PROVINCIAS

Provincia	Frecuencia	%
ICA	503,370	48,01
CHINCHA	269,850	25,74
PISCO	166,894	15,92
NAZCA	90,323	8,61
PALPA	18,117	1,73
	1048,554	100,00

En cuanto a las vacunaciones realizadas en las diversas provincias de la región, se puede ver que la provincia de ICA y la provincia de CHINCHA cuentan con la mayor cantidad de personas vacunadas, las mismas que representan el 73.75%.

TABLA VI
FRECUENCIA DE VACUNAS POR DISTRITOS

Distrito	Frecuencia	%
ICA	344,121	32,82
CHINCHA ALTA	137,417	13,11
PISCO	87,250	8,32
PUEBLO NUEVO	55,625	5,30
NAZCA	45,892	4,38
LA TINGUIÑA	27,071	2,58
PARCONA	23,321	2,22
SUNAMPE	23,158	2,21
GROCIO PRADO	21,568	2,06
SALAS	21,133	2,02
MARCONA	20,247	1,93
SANTIAGO	19,154	1,83
SAN CLEMENTE	18,533	1,77
VISTA ALEGRE	18,266	1,74
TUPAC AMARU	17,622	1,68
INCA		
SAN ANDRES	15,631	1,49
SUBTANJALLA	15,285	1,46
(Other)	137,26	13,09
	1048,554	100,00

En la tabla VI se presentan las vacunas aplicadas en los distritos de las diversas provincias, se puede destacar que las capitales de provincia son las que cuentan con la mayor cantidad de vacunados, como se puede comprobar en Ica, Chincha y Pisco. Es importante preciar que en la región Ica, hay dos distritos con el mismo nombre Pueblo Nuevo (en Ica y en Chincha, con una mayor población vacunada en Chincha con 47,589) motivo por el cual se ubica en la 4 posición.

TABLA VII
FRECUENCIA DE VACUNAS SEGÚN GRUPO DE RIESGO

Grupo de riesgo	Frecuencia	%
PERSONAS DE 20 A 29 AÑOS	187,284	17,86
PERSONAS DE 30 A 39 AÑOS	164,814	15,72
PERSONAS DE 40 A 49 AÑOS	137,423	13,11
60 AÑOS A MÁS, ADULTO MAYOR	136,516	13,02
PERSONAS DE 50 A 59 AÑOS	105,65	10,08
PERSONAS DE 12 A 17 AÑOS	98,025	9,35
PERSONAS DE 5 A 7 AÑOS	42,855	4,09
PERSONAS DE 18 A 19 AÑOS	37,458	3,57
PERSONAS DE 8 A 9 AÑOS	29,675	2,83
PERSONAS DE 10 A 11 AÑOS	29,173	2,78
PERSONAL DE SALUD	18,766	1,79
PERSONAS DE 70 AÑOS A MÁS	15,913	1,52
6 MESES A 4 AÑOS	9,343	0,89
OTROS GRUPOS DE RIESGO	8,019	0,76
GESTANTE	5,820	0,56
PERSONAS PRIVADAS DE SU LIBERTAD (INPE)	4,050	0,39
POLICIA NACIONAL DEL PERU	3,395	0,32
(Other)	14,375	1,37
	1048,554	100,00

En la Tabla VII se puede apreciar que la mayor población vacunada se encuentra en el intervalo de 20 a 59 años con aproximadamente el 70%. Se aprecia en la tabla igualmente que la policía represente un pequeño porcentaje de vacunados con solo el 0,32% de la población, además de la población en las cárceles representan similarmente 0,39% de lo vacunados.

3.2 Análisis descriptivo de las variables numéricas

TABLA VIII
MEDIDAS DE TENDENCIA CENTRAL

Estadístico	Edad	Dosis
Media =	36,02	2,2
Mediana =	33	2
Moda =	25	1

El análisis descriptivo de las variables Edad y Dosis, no indican que la media de la población vacunada es de 36,02 con una media de la Dosis de 2,2 vacunas.

TABLA IX
MEDIDAS DE DISPERSIÓN

Estadístico	Edad	Dosis
Desviación Estándar =	17,09	1,06
La varianza =	402,06	1,16
Error estándar =	0,0006	3,7E-05

TABLA X
INTERVALOS DE CONFIANZA PARA LOS DATOS (95%)

Estadístico	Media	Desviación estándar	Intervalos de confianza 95%	
			Límite inferior	Límite superior
Edad =	36,02	17,09	2,52	69,52
Dosis =	2,2	1,06	0,12	4,28

De la Tabla X, para el intervalo de confianza del 95%, la edad de las personas vacunadas se encuentra en un rango desde los 2 años hasta los 69 años, con un rango de dosis aplicadas entre 0,12 y 4,28.

IV. DISCUSIÓN

El estudio refleja las vacunaciones realizadas en nuestro estudio iniciadas en el 2021 con poca cobertura por la carencia de vacunas para atender a la población. En ese mismo camino en la que se investiga la disponibilidad de las vacunas a nivel internacional donde esta cobertura varía entre el 1% y más del 70% [1], en la investigación con relación a Malasia. En el caso de nuestra región en enero del 2021 solo se pudo vacunar al 5.09 de la población de la región y en el pico más alto en septiembre del mismo año se alcanzó el 15.25%. a nivel de la vacunación igualmente es importante lo encontrado por “un grupo de investigación de vacunas para COVID de la Universidad de Oxford considera que debe ser efectiva con una o dos dosis” [16], y esto tiene cierta concordancia con nuestros resultados donde la mayor cantidad de vacunados se da con esta cantidad de vacunas, 32% con una vacuna y 30% con dos vacunas, este porcentaje representa más del 50% de la población de nuestra región y que a la actualidad como se desprende del anexo 01 la cantidad de muertes ha sido drásticamente reducida. En contraposición de nuestro análisis en donde la vacuna Sinopharm es la segunda más utilizada con un 24,39%, estudio sobre los tipos de vacunación aplicados en Corea del Sur para aumentar la tasa de vacunación, la vacuna Sinopharm no figura dentro de las posibilidades [4].

V. CONCLUSIONES

Culminada nuestro trabajo de tesis, se ha llegado a las siguientes conclusiones de los resultados y objetivos planteado.

1. La analítica de los datos es una técnica adecuada que permite evaluar los datos, especialmente lo relacionado con la vacunación del COVID-19.
2. Se logró conocer la evolución de la vacunación a nivel de la población de la región de Ica, mediante el uso de plataformas y herramientas tecnológicas como Google Colaboratory, Python y diversas librerías son eficientes en la obtención de los resultados.
3. Utilizar software de minería de datos como Orange Data Mining es una excelente opción para poder obtener conocimiento de los datos sobre la vacunación del COVID-19.
4. Conocer los fabricantes de las vacunas que mayormente se han utilizado es sumamente importante, ya que gracias a ellas se ha podido volver a la normalidad en este año 2023.
5. Importante igualmente conocer que la cantidad de vacunas puestas a los ciudadanos ha habido casos muy aislados en las que se han puesto hasta 8 o 9 vacunas. Se hace notar que lo decretado por el gobierno para la 4ta vacuna como obligatoria y a pesar de las restricciones impuestas solo el 10% de la población vacunada se pusieron la 4ta vacuna.

VI. RECOMENDACIONES

Si bien es cierto que la evaluación de la vacunación en la región de Ica es importante para conocimiento de las características de la vacunación, como la cantidad de vacunas puestas, el tipo de vacuna utilizada y los distritos donde se han aplicado estas vacunas y que han sido abordados en el presente estudio. Queda aún en el aire la posibilidad de otros estudios que se pueden realizar con la data que está disponible de forma libre en el MINSA, se conoce que aún existen personas que han fallecido a causa de esta enfermedad entre junio y agosto del 2023, por lo que se abre una importante posibilidad de investigación en conocer si estas personas estaban vacunadas o no estaban vacunadas, si estaban vacunadas que vacunas se habían puesto y que cantidad, como también sería importante conocer la edad, sexo y en que distrito del país se han reportado estas (ver anexo 01) y que no han sido abordadas en el presente estudio de tesis.

VII. REFERENCIAS BIBLIOGRÁFICAS

- [1] G. Krishnan y M. R. B. Khan, «Data Analytics and Prediction Model for Malaysian Covid 19 Vaccination Progress», *Malays. J. Sci. Adv. Technol.*, pp. 84-89, jul. 2022, doi: 10.56532/mjsat.v2i3.61.
- [2] O. ONU, «Enfermedad del coronavirus (COVID-19) | Naciones Unidas», United Nations. Accedido: 26 de septiembre de 2023. [En línea]. Disponible en: <https://www.un.org/es/coronavirus>
- [3] Stedman, «¿Qué es Análisis o analítica de datos? - Definición en WhatIs.com», *ComputerWeekly.es*. Accedido: 4 de febrero de 2023. [En línea]. Disponible en: <https://www.computerweekly.com/es/definicion/Analisis-o-analitica-de-datos>
- [4] Y.-J. Choi, J. Lee, y S. Y. Paek, «Public Awareness and Sentiment toward COVID-19 Vaccination in South Korea: Findings from Big Data Analytics», *Int. J. Environ. Res. Public Health*, vol. 19, n.º 16, Art. n.º 16, ene. 2022, doi: 10.3390/ijerph19169914.
- [5] J. Alcocer-Carranza *et al.*, «Actitud ante la vacuna contra COVID-19 de los habitantes del estado de Guerrero, México», *Univ. Salud*, vol. 24, n.º 2, pp. 108-116, ago. 2022, doi: 10.22267/rus.222402.264.
- [6] B. Lusich y I. C. P. Salazar, «Trombosis de eje porto-espleno-mesentérico por síndrome de trombosis con trombocitopenia inducida por la vacuna COVID-19 AstraZeneca ChAdOx1nCov-19: reporte de caso», *Rev. Hematol.*, vol. 26, n.º 3, Art. n.º 3, dic. 2022, doi: 10.48057/hematologa.v26i3.435.
- [7] J. A. Solís Jara y A. P. Torres Boni, «“Reacciones adversas post vacuna covid-19 en la población adulta”», *Univ. Católica Cuenca*, 2022, Accedido: 31 de enero de 2023. [En línea]. Disponible en: <https://dspace.ucacue.edu.ec/handle/ucacue/13020>
- [8] D. A. M. Suarez *et al.*, «IMPACTO DE LA VACUNACION COVID-19 EN LA INTERPRETACION DE LA IMAGEN METABOLICA PET-TC DEL PACIENTE ONCOLOGICO», *Seram*, vol. 1, n.º 1, Art. n.º 1, may 2022, Accedido: 31 de enero de 2023. [En línea]. Disponible en: <https://piper.espacio-seram.com/index.php/seram/article/view/8486>
- [9] K. J. Chuqui Castro y M. C. Landin Gutama, «“Factores asociados a la aceptación de la vacuna COVID-19 en el cantón Azogues “», *Univ. Católica Cuenca*, 2022, Accedido: 31 de enero de 2023. [En línea]. Disponible en: <https://dspace.ucacue.edu.ec/handle/ucacue/12964>
- [10] P. Almada, R. E. Buffa, H. Pignata, y Y. Vernetti, «Encefalopatía aguda posterior a vacuna covid 19. Reporte de casos», *Rev. Argent. Ter. Intensiva*, vol. 39, sep. 2022, Accedido: 31 de enero de 2023. [En línea]. Disponible en: <https://revista.sati.org.ar/index.php/MI/article/view/840>

- [11] V. López-Sandoval, «Distribución de vacuna COVID-19: Combinando SEIR y Machine Learning», *Uniciencia*, vol. 36, n.º 1, pp. 208-222, dic. 2022, doi: 10.15359/ru.36/1.12.
- [12] C. M. Ríos-González, «Predisposición para recibir la vacuna contra el COVID-19 en Paraguay: estudio exploratorio online», *Med. Clínica Soc.*, vol. 5, n.º 2, pp. 119-120, ago. 2021, doi: 10.52379/mcs.v5i2.189.
- [13] S. Cris, L. de Armentia, M. José, y A. Laura, «EVALUACION DE LA RESPUESTA HUMORAL FRENTE A LA VACUNACION COVID-19 DEL PERSONAL DE SALUD DEL HOSPITAL SBARRA», vol. 3, 2021.
- [14] A. Pérez Soto y D. Berríos Viana, «Determinantes sobre la aceptación de la vacuna COVID-19, en el Distrito Nacional, República Dominicana, durante el período de febrero a marzo de 2021», Thesis, Santo Domingo: Universidad Iberoamericana (UNIBE), 2021. Accedido: 31 de enero de 2023. [En línea]. Disponible en:
<https://repositorio.unibe.edu.do/jspui/handle/123456789/546>
- [15] M. C. Piqueras, J. H. Carmona, y J. P. Bernáldez, «“Vísteme despacio que tengo prisa”. Un análisis ético de la vacuna contra la COVID-19: fabricación, distribución y reticencia», *Enrahonar Int. J. Theor. Pract. Reason*, vol. 65, pp. 57-73, oct. 2020, doi: 10.5565/rev/enrahonar.1307.
- [16] J. E. Gomez Marín *et al.*, «Una hoja de ruta para la Vacuna COVID 19 en Colombia, un reto posible», *Infectio*, vol. 25, n.º 1, p. 7, sep. 2020, doi: 10.22354/in.v25i1.901.
- [17] J. M. A. Corrales Chire, «Percepciones de la aceptación de la vacuna contra el Covid-19 en personas que acuden a un mercado popular en Arequipa 2021», *Univ. Nac. San Agustín Arequipa*, 2021, Accedido: 31 de enero de 2023. [En línea]. Disponible en:
<http://repositorio.unsa.edu.pe/handle/20.500.12773/12315>
- [18] L. L. Aranda Chiclayo, «Conocimiento y aceptación frente a la vacuna Covid-19 en el personal de salud de Hospitales nivel II-1 Red Trujillo, 2021», *Repos. Inst. - UCV*, 2021, Accedido: 31 de enero de 2023. [En línea]. Disponible en:
<https://repositorio.ucv.edu.pe/handle/20.500.12692/73839>
- [19] L. E. Paico Linares, «Reflexiones de los usuarios acerca de la efectividad de la vacuna Covid-19, de un establecimiento de salud, Ventanilla, 2022», *Repos. Inst. - UCV*, 2022, Accedido: 31 de enero de 2023. [En línea]. Disponible en:
<https://repositorio.ucv.edu.pe/handle/20.500.12692/95375>
- [20] J. Supo, «Tipos, Niveles y Diseños en Investigación | PODCAST», Youtube. [En línea]. Disponible en: <https://www.youtube.com/watch?v=0ftt84ppXWE&t=67s>

VIII. ANEXOS

Anexo 01: Incremento de muertes en la Perú

TABLA XI
INCREMENTO DE MUERTES EN EL PERU

Fecha	Incremento Muertos	Muertos	Muertos / millón	Incremento Confirmados	Confirmados	Confirmados / 100.000 - 14 días
21/09/2023	0	221.674	6.574,85	0	4.518.007	0
20/09/2023	0	221.674	6.574,85	0	4.518.007	0
19/09/2023	0	221.674	6.574,85	0	4.518.007	0
18/09/2023	0	221.674	6.574,85	0	4.518.007	0
17/09/2023	0	221.674	6.574,85	0	4.518.007	0
16/09/2023	0	221.674	6.574,85	0	4.518.007	0
15/09/2023	0	221.674	6.574,85	0	4.518.007	0
14/09/2023	0	221.674	6.574,85	0	4.518.007	0
13/09/2023	0	221.674	6.574,85	0	4.518.007	0
12/09/2023	0	221.674	6.574,85	0	4.518.007	0
11/09/2023	0	221.674	6.574,85	0	4.518.007	0
10/09/2023	0	221.674	6.574,85	0	4.518.007	0
09/09/2023	0	221.674	6.574,85	0	4.518.007	0
08/09/2023	0	221.674	6.574,85	0	4.518.007	0
07/09/2023	0	221.674	6.574,85	0	4.518.007	0
06/09/2023	0	221.674	6.574,85	0	4.518.007	0
05/09/2023	0	221.674	6.574,85	0	4.518.007	0
04/09/2023	0	221.674	6.574,85	0	4.518.007	0
03/09/2023	0	221.674	6.574,85	0	4.518.007	0,12
02/09/2023	0	221.674	6.574,85	0	4.518.007	0,12
01/09/2023	0	221.674	6.574,85	0	4.518.007	0,12
31/08/2023	0	221.674	6.574,85	0	4.518.007	0,12
30/08/2023	0	221.674	6.574,85	0	4.518.007	0,12
29/08/2023	0	221.674	6.574,85	0	4.518.007	0,12
28/08/2023	0	221.674	6.574,85	0	4.518.007	0,12
27/08/2023	0	221.674	6.574,85	0	4.518.007	0,63
26/08/2023	0	221.674	6.574,85	0	4.518.007	0,63
25/08/2023	0	221.674	6.574,85	0	4.518.007	0,63
24/08/2023	0	221.674	6.574,85	0	4.518.007	0,63
23/08/2023	0	221.674	6.574,85	0	4.518.007	0,63
22/08/2023	0	221.674	6.574,85	0	4.518.007	0,63
21/08/2023	51	221.674	6.574,85	42	4.518.007	0,63
20/08/2023	0	221.623	6.573,33	0	4.517.965	1,79
19/08/2023	0	221.623	6.573,33	0	4.517.965	1,79
18/08/2023	0	221.623	6.573,33	0	4.517.965	1,79
17/08/2023	0	221.623	6.573,33	0	4.517.965	1,79

16/08/2023	0	221.623	6.573,33	0	4.517.965	1,79
15/08/2023	0	221.623	6.573,33	0	4.517.965	1,79
14/08/2023	83	221.623	6.573,33	170	4.517.965	1,79
13/08/2023	0	221.540	6.570,87	0	4.517.795	2,26
12/08/2023	0	221.540	6.570,87	0	4.517.795	6,48
11/08/2023	0	221.540	6.570,87	0	4.517.795	6,48
10/08/2023	0	221.540	6.570,87	0	4.517.795	6,48
09/08/2023	0	221.540	6.570,87	0	4.517.795	6,48
08/08/2023	0	221.540	6.570,87	0	4.517.795	6,48
07/08/2023	75	221.540	6.570,87	432	4.517.795	6,48
06/08/2023	0	221.465	6.568,65	0	4.517.363	5,20
05/08/2023	0	221.465	6.568,65	0	4.517.363	5,20
04/08/2023	0	221.465	6.568,65	0	4.517.363	5,20
03/08/2023	0	221.465	6.568,65	0	4.517.363	5,20
02/08/2023	0	221.465	6.568,65	0	4.517.363	5,20
01/08/2023	0	221.465	6.568,65	0	4.517.363	5,20
31/07/2023	101	221.465	6.568,65	329	4.517.363	5,20
30/07/2023	103	221.364	6.565,65	1.424	4.517.034	4,22
29/07/2023	0	221.261	6.562,60	0	4.515.610	2,75
28/07/2023	0	221.261	6.562,60	0	4.515.610	2,75
27/07/2023	0	221.261	6.562,60	0	4.515.610	2,75
26/07/2023	0	221.261	6.562,60	0	4.515.610	2,75
25/07/2023	0	221.261	6.562,60	0	4.515.610	2,75
24/07/2023	0	221.261	6.562,60	0	4.515.610	2,75
23/07/2023	0	221.261	6.562,60	0	4.515.610	2,75
22/07/2023	0	221.261	6.562,60	0	4.515.610	10,44
21/07/2023	0	221.261	6.562,60	0	4.515.610	10,44
20/07/2023	0	221.261	6.562,60	0	4.515.610	10,44
19/07/2023	0	221.261	6.562,60	0	4.515.610	10,44
18/07/2023	0	221.261	6.562,60	0	4.515.610	10,44
17/07/2023	0	221.261	6.562,60	0	4.515.610	10,44
16/07/2023	58	221.261	6.562,60	928	4.515.610	10,44
15/07/2023	0	221.203	6.560,88	0	4.514.682	7,68
14/07/2023	0	221.203	6.560,88	0	4.514.682	7,68
13/07/2023	0	221.203	6.560,88	0	4.514.682	7,68
12/07/2023	0	221.203	6.560,88	0	4.514.682	7,68
11/07/2023	0	221.203	6.560,88	0	4.514.682	7,68
10/07/2023	0	221.203	6.560,88	0	4.514.682	7,68
09/07/2023	160	221.203	6.560,88	2.591	4.514.682	7,68
08/07/2023	0	221.043	6.556,13	0	4.512.091	11,83
07/07/2023	0	221.043	6.556,13	0	4.512.091	11,83
06/07/2023	0	221.043	6.556,13	0	4.512.091	11,83
05/07/2023	0	221.043	6.556,13	0	4.512.091	11,83
04/07/2023	0	221.043	6.556,13	0	4.512.091	11,83
03/07/2023	0	221.043	6.556,13	0	4.512.091	11,83
02/07/2023	0	221.043	6.556,13	0	4.512.091	11,83
01/07/2023	0	221.043	6.556,13	0	4.512.091	11,83

30/06/2023	0	221.043	6.556,13	0	4.512.091	11,83
29/06/2023	0	221.043	6.556,13	0	4.512.091	11,83
28/06/2023	0	221.043	6.556,13	0	4.512.091	11,83
27/06/2023	0	221.043	6.556,13	0	4.512.091	11,83
26/06/2023	0	221.043	6.556,13	0	4.512.091	11,83
25/06/2023	250	221.043	6.556,13	3.988	4.512.091	11,83
24/06/2023	0	220.793	6.548,71	0	4.508.103	0
23/06/2023	0	220.793	6.548,71	0	4.508.103	0
22/06/2023	0	220.793	6.548,71	0	4.508.103	0
21/06/2023	0	220.793	6.548,71	0	4.508.103	0
20/06/2023	0	220.793	6.548,71	0	4.508.103	0
19/06/2023	0	220.793	6.548,71	0	4.508.103	0
18/06/2023	0	220.793	6.548,71	0	4.508.103	0
17/06/2023	0	220.793	6.548,71	0	4.508.103	0
16/06/2023	0	220.793	6.548,71	0	4.508.103	0
15/06/2023	0	220.793	6.548,71	0	4.508.103	0
14/06/2023	0	220.793	6.548,71	0	4.508.103	0

Fuente: Datosmacro - <https://datosmacro.expansion.com/otros/coronavirus/peru> 26/08/2023