



# UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL CUSCO

FACULTAD DE INGENIERÍA ELÉCTRICA,  
ELECTRÓNICA, INFORMÁTICA Y MECÁNICA

ESCUELA PROFESIONAL DE  
INGENIERÍA INFORMÁTICA Y DE SISTEMAS

6 de agosto de 2022

## *ANÁLISIS DE DATOS DATOS MUESTRALES COVID-2019*

PRESENTA

Mollinedo Peña, Alvaro Sebastian  
182921

Pumacahua Apaza, Leonardo Chepper  
171524

ASIGNATURA

Análisis de Datos Empresariales

PROFESOR

Garcia Zanabria, Germain

# 1. Introducción

En el presente trabajo informativo abordaremos temas sobre el análisis de datos aplicados a cualquier sector público o privado y el estudio de estos mismos para alcanzar ciertos objetivos y resultados con respecto a los datos estudiados. Redactaremos todo el proceso el cual conllevo realizar este trabajo, dificultades, resultados, etc.

De esta manera podremos implementar varias herramientas que nos faciliten en el proceso para alcanzar el objetivo planteado, como claro ejemplo para el uso de datos, hojas de cálculo Excel y para un análisis mucho más profundizado, detallado y mejor implementado, Power BI, que nos facilitará este trabajo de analizar grandes cantidades de información y también, de implementar ilustraciones dinámicas como gráficos de barras, circulares, mapas cromáticos o estadísticos.

Para este caso, usamos un dataset de la página oficial de Perú en donde se publican diferentes estudios estadísticos realizados a tiempo real, con referencia <https://www.datosabiertos.gob.pe/>. El dataset que escogimos Pruebas Moleculares del Instituto Nacional de Salud para COVID-19, tomamos este tema debido a la coyuntura actual en la que vivimos.

Esperamos que el trabajo sea de su agrado.

## 2. Desarrollo

Como mencionamos anteriormente, tomamos un dataset del Instituto Nacional de Salud del Perú para COVID-19, dónde la variable de estudio fueron las pruebas moleculares que se realizaron.

Este dataset, inicialmente, contenía más de 1 millo de datos procesados, lo cual era una cantidad exorbitante para analizar, el procesador de datos Excel no admitía esta cantidad de filas, por lo tanto tuvimos que utilizar un procesador mucho más versátil, el cual fue Power BI.

De igual forma, la cantidad de datos era muy grande y el proceso para analizar, cuantificar y procesar estos tomaba un tiempo considerable lo cual sería algo perjudicial al momento de presentar ante la clase.

Adelantando un poco los datos que tenía este dataset, contenía información de pacientes, como íbamos mencionando, que se realizaron una prueba molecular, estos comprendían el lugar en dónde se hicieron las pruebas, teniendo como valor los diferentes departamentos del Perú, para lo cual, debido a la localidad en la que nos encontramos (nosotros los desarrolladores del trabajo informativo), usamos solamente los que se realizaron en el departamento del Cusco.

### 2.1. Diccionario de Datos

A continuación, presentaremos una tabla da una breve explicación de los datos o variables que contenía el dataset.

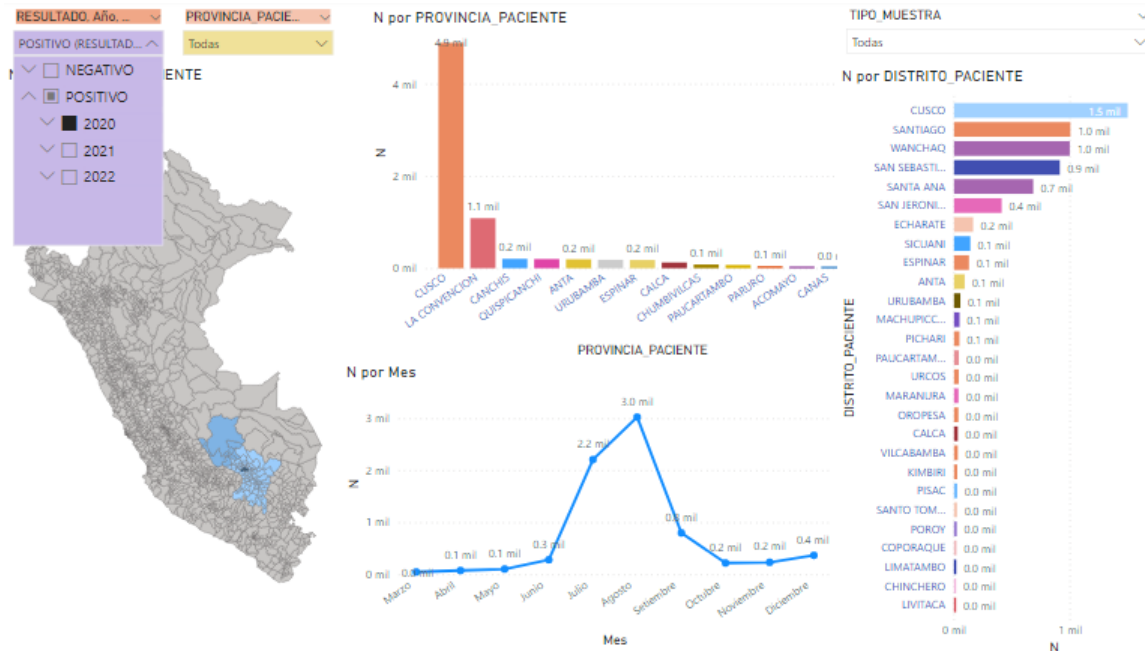
Variable	Tipo de Dato	Descripción
FECHA CORTE	DATE	FECHA DEL CORTE QUE SE REALIZÓ EL CORTE DE INFORMACIÓN
UUID	STRING	CODIFICACIÓN ASIGNADA A UN PACIENTE
FECHA MUESTRA	DATE	FECHA DE TOMA MUESTRA
EDAD	INT	EDAD DEL PACIENTE
SEXO	STRING	SEXO DEL PACIENTE
INSTITUCION	STRING	INSTITUCIÓN A LA QUE PERTENECE EL ESTABLECIMIENTO ORIGEN
UBIGEO PACIENTE	STRING	UBIGEO DEL PACIENTE, CORRESPONDIENTE AL UBIGEO DE RENIEC
DEPARTAMENTO PACIENTE	STRING	DEPARTAMENTO DEL PACIENTE
PROVINCIA PACIENTE	STRING	PROVINCIA DEL PACIENTE
DISTRITO PACIENTE	STRING	DISTRITO DEL PACIENTE
UBIGEO MUESTRA	STRING	UBIGEO DEL ESTABLECIMIENTO ORIGEN, LUGAR DONDE SE TOMÓ LA MUESTRA. UBIGEO CORRESPONDIENTE AL INEI
DEPARTAMENTO MUESTRA	STRING	DEPARTAMENTO DEL ESTABLECIMIENTO ORIGEN, LUGAR DONDE SE TOMÓ LA MUESTRA
PROVINCIA MUESTRA	STRING	PROVINCIA DEL ESTABLECIMIENTO ORIGEN, LUGAR DONDE SE TOMÓ LA MUESTRA
DISTRITO MUESTRA	STRING	DISTRITO DEL ESTABLECIMIENTO ORIGEN, LUGAR DONDE SE TOMÓ LA MUESTRA
TIPO MUESTRA	STRING	TIPO DE MUESTRA
RESULTADO	STRING	RESULTADO COVID POSITIVO O NEGATIVO

## 2.2. Análisis de Datos

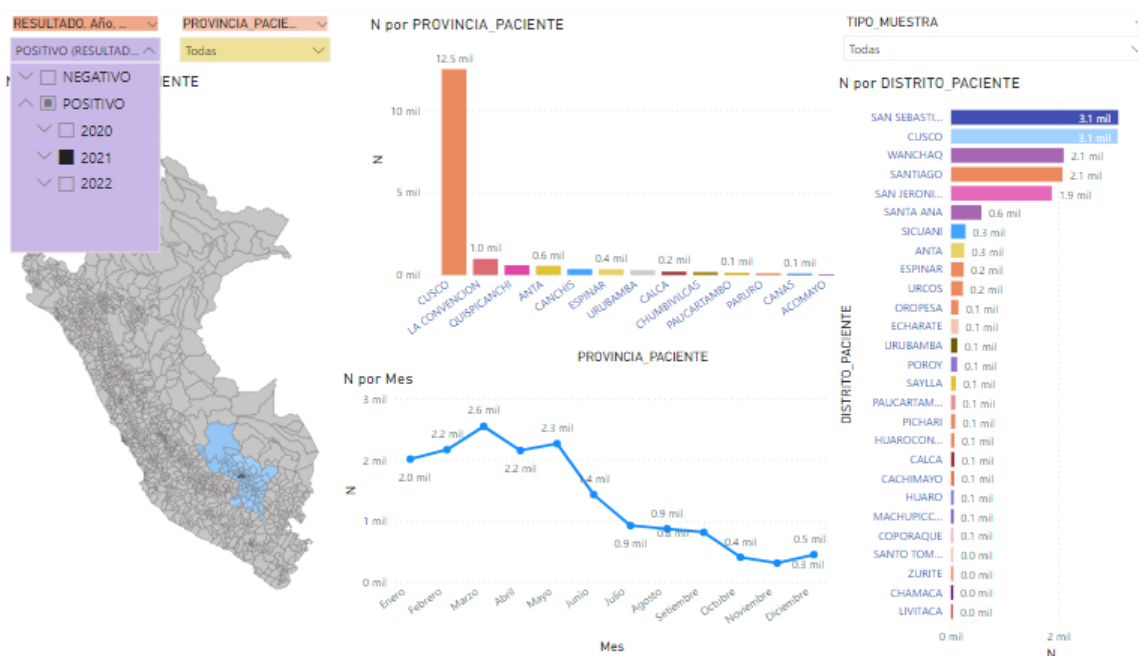
### 2.2.1. Primera Tarea Analítica

Como primera tarea analítica compararemos la cantidad de casos positivos en el departamento de Cusco, agrupando mes a mes y así conocer la variación de contagios, veremos los picos de contagios y los relacionamos con acontecimientos que provocan el alza de casos positivos.

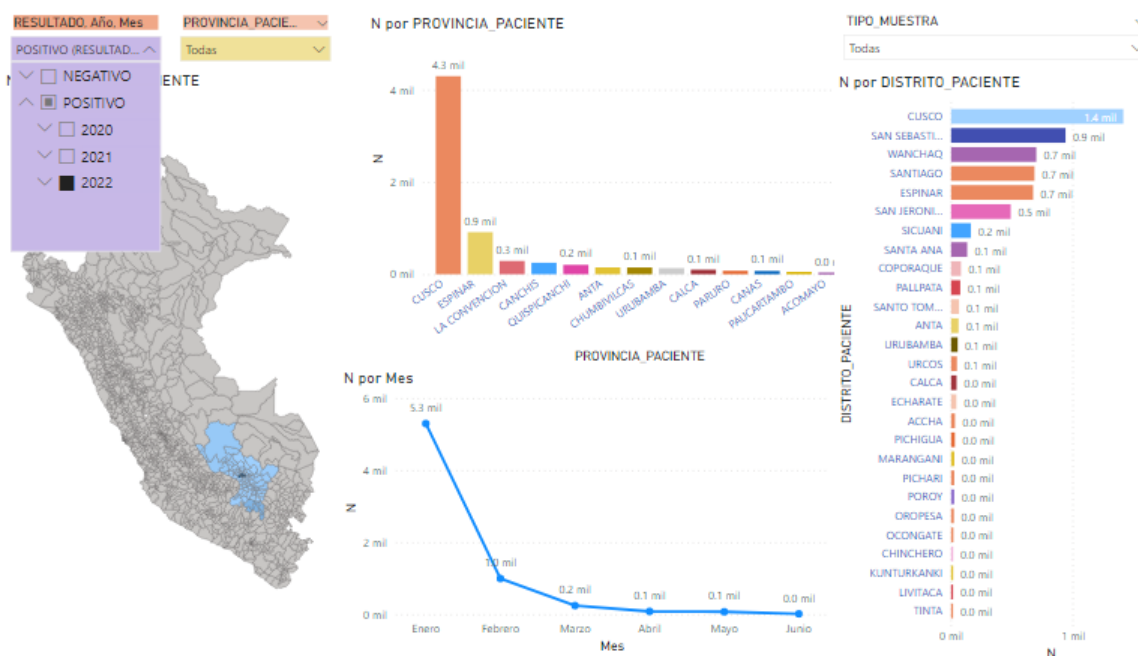
**2020.** El crecimiento de casos positivos desde el mes de julio hasta llegar al pico de todo el año en el mes de agosto se debe principalmente al inicio de la primera ola de contagios que se venía registrando en todo el Perú que fue provocado por el levantamiento de la cuarentena. Esta fue decayendo paulatinamente porque la población tomó las medidas sanitarias para evitar más contagios y así reduciendo el número de casos positivos Aunque la caída se debe también ante la pequeña caída de pruebas que se tomó la población.



**2021.** El año 2021 empezó nuevamente con un alza considerable esto debido a varios factores, el primero es que se vio una normalización de la pandemia, así como, las fiestas de fin de año como navidad y año nuevo en segundo lugar, la variante lambda se hizo presente en el Perú a inicios de Enero y desde Abril la mayoría de contagios fueron causadas por ella. Y es así como se refleja en el número de casos positivos de las pruebas. Y aunque el número de pruebas realizadas cada mes seguía en aumento el número de casos positivos fue disminuyendo a lo largo del año. Esto quiere decir que la población volvió a ser más precavida con los cuidados respectivos para evitar el contagio luego de la segunda ola.



**2022.** Para el año 2022 se volvió otra vez el alza de casos positivos una vez más las fiestas de fin de año, así como la llegada de nuevas variantes fueron la detonante del gran alza de casos positivos. Luego del mes de Enero se muestra una gran caída. Esto no es debido a que la población de un día para otro se cuidó adecuadamente sino a que los datos que aún no fueron verificados no se incluyen en el dataset lo cual provoca una gran caída de pruebas y en consecuencia una gran caída en los casos positivos llegando en Junio a ser mínimas y esto no se refleja en la realidad como sabemos que la tercera ola apenas está acabando.

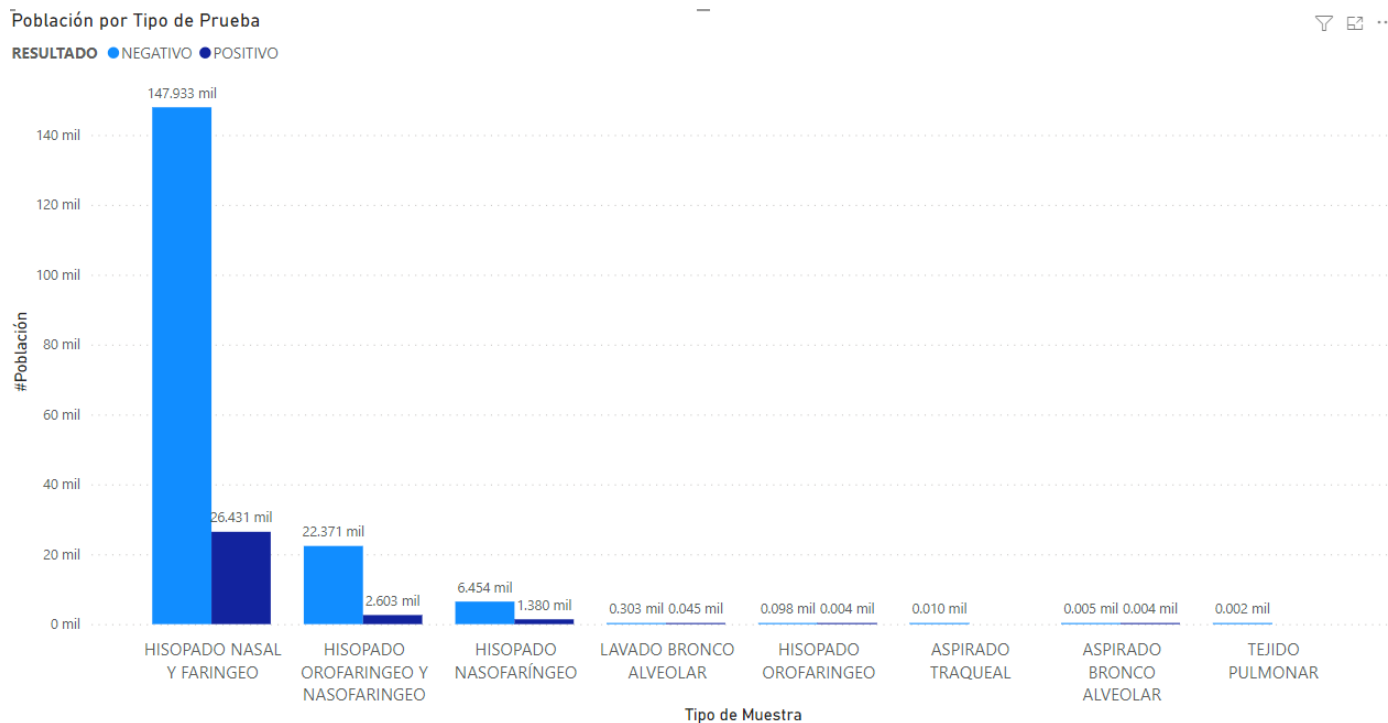


### 2.2.2. Segunda Tarea Analítica

La segunda tarea analítica, compararemos la cantidad de positivos según el tipo de prueba que la población realizó, tenemos como parámetro 8 tipos diferentes de tipos de muestra.

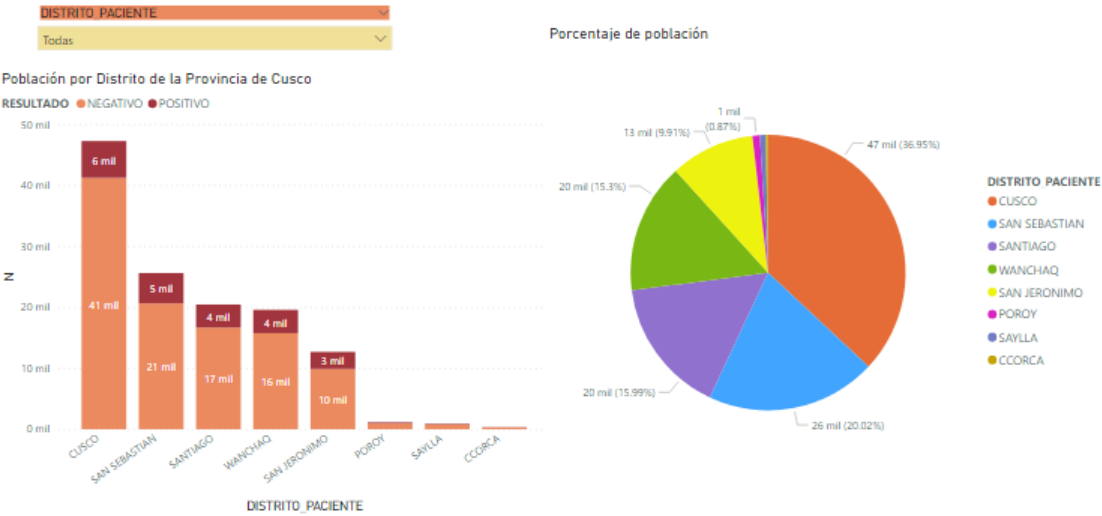
Podemos observar la cantidad de pruebas que se realizaron según el tipo de muestra, tenemos el dato más repetitivo que es el tipo de muestra de HISOPADO NASAL Y FARINGEO, y el menos constante, el del TEJIDO PULMONAR.

Mediante encuestas realizadas a la población, se tuvo la conclusión de que el tipo de muestra HISOPADO NASAL Y FARINGEO fue el tipo de muestra que la mayor parte de la población escogió, ya que este representaba un proceso “menos doloroso” y tomaba menos tiempo. Con respecto al usuario que tomaba este tipo de muestra como son los médicos o enfermeras, dijeron que este proceso representaba un menor costo en comparación con los diferentes tipos de muestras, ya que el instrumental que necesitaban era fácil de conseguir y el proceso a realizar era más sencillo para ellos.



### 2.2.3. Tercera Tarea Analítica

En la tercera tarea analítica compararemos los casos positivos agrupados por distrito, así como la población del distrito y de esta forma averiguar en qué distritos hubo mayor cantidad de contagios en función de su población. Los casos positivos en los distritos del Cusco están en un rango entre 12 % y 22 % del total de pruebas covid. Los distritos con mayor porcentaje de positivos son San Jeronimo, San Sebastian y Wanchaq el motivo es que en estos distritos existen bastantes lugares de reunión de personas como parques, mercados, bancos, etc. El primer lugar de estos es San Jeronimo con un 21.99 % de casos positivos del total de pruebas tomadas esto debido principalmente a que el mercado Vinocanchon es un gran foco de infección así como los varios lugares de reunión de personas que tiene el distrito. También cabe resaltar que los porcentajes no son muy dispares debido a que todos los distritos están cerca, y la mayoría de la población se mueve dentro de estos para sus tareas diarias, como ir a trabajar, ir a sus centros de estudio, etc. Esto contribuye a que el nivel de casos positivos sea parejo.



### 3. Conclusiones

El estudio analítico que hemos realizado nos ayuda a ver un panorama más completo de sobre las complicaciones y varios factores que conlleva la actual coyuntura en la que vivimos, de esta forma podemos conocer como fue que evolucionaron estos datos y las consecuencias que tuvo en varios sectores, tanto laborales como académicos, para personas en general.

De esta forma, se puede informar a la población y aceptar las consecuencias de la pandemia y empezar a tomar conciencia.