Análisis del comportamiento y proyección del COVID 19 en Colombia y el Bogotá D.C

Diego Hernan Cadena Daza, Sebastián Valencia Cerón

*Bogotá D.C, Universidad Cooperativa de Colombia*

[diego.cadena@campusucc.edu.co](mailto:diego.cadena@campusucc.edu.co)

[sebastian.valenciac@campusucc.edu.co](mailto:sebastian.valenciac@campusucc.edu.co)

**La COVID-19 es la enfermedad causada por el nuevo coronavirus conocido como SARS-CoV-2. La OMS tuvo noticia por primera vez de la existencia de este nuevo virus el 31 de diciembre de 2019, al ser informada de un grupo de casos de «neumonía vírica» que se habían declarado en Wuhan (República Popular China).**

# I. INTRODUCCIÓN

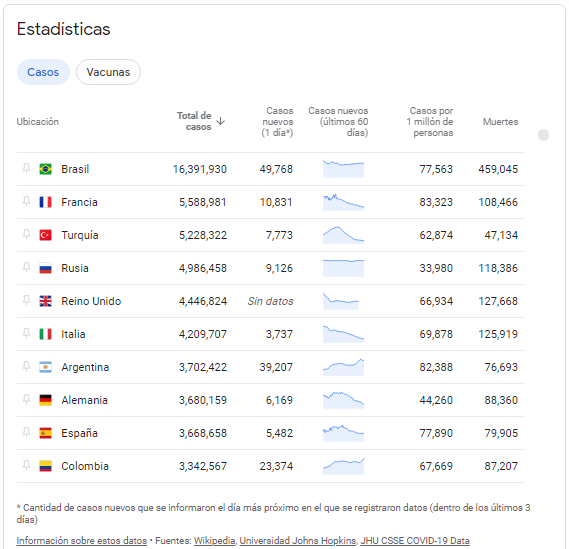
En el presente documento se realizara un análisis de los casos reportados desde sus inicios hasta la semana 74 correspondiente al mes de mayo del año 2021 para la ciudad de Bogotá D.C y nivel nacional (Colombia). Para el estudio de los datos se procesara semana a semana, donde se implementara un script elaborado en el lenguaje de programación phyton y para el tratamiento de información se utilizara el motor de Bases de Datos SQL Server junto a ello se utilizara el método de regresión polinomial de 3er grado. Con estas herramientas ya mencionadas se podrá identificar como ha sido la evolución de la enfermedad referente a los contagios y muertos por la misma y así mismo se podrá predecir comportamiento en los próximos meses.

# II. COVID 19

El COVID-19 se propaga cuando una persona infectada exhala gotitas y partículas respiratorias muy pequeñas que contienen el virus. Estas gotitas y partículas respiratorias pueden ser inhaladas por otras personas o depositarse sobre sus ojos, nariz o boca. En algunas circunstancias, pueden contaminar las superficies que tocan. Quienes están a menos de 6 pies de distancia de una persona infectada tienen mayor probabilidad de infectarse.

## A. Contagios en Colombia

A partir del reporte descargado de la página del gobierno de Colombia a corte de la semana 74 de mes de mayo del 2021. En Colombia alcanzo la cifra de 3.210.784 personas contagiadas de las cuales ya han fallecido 84.228 por la enfermedad de COVID-19, estos datos comparados con otros países del mundo son preocupantes ocupando el puesto número 12 en personas contagiadas a nivel mundial, donde estados unidos es el país con más contagios con un total de 33.238.422



## B. Contagios en Bogotá D.C

La situación en Bogotá D.C a corte de la semana 74 del año 2021, no es nada favorable. Debido a que es la ciudad con más contagios en Colombia con una cifra de 895.177 contagiados. Así mismo, es la ciudad con mayor número de muertes siendo este un valor de 17.870 fallecidos por COVID-19.

# III. *Plataformas manejadas (SQL SERVER - PHYTON)*

SQL SERVER es un sistema de gestión de base de datos relacional desarrollado como un producto de software con la función principal de almacenar y recuperar datos según lo solicitado por otras aplicaciones. Con esta herramienta descripta se utilizara refinamiento de la información debido a la volumetría. Para que el análisis sea más conciso se realizara la extracción de la información por semanas desde sus inicios hasta la semana 74 correspondiente al mes de mayo del 2021, seguido a ello se ejecuta el siguiente script para la generación de la información requerida.

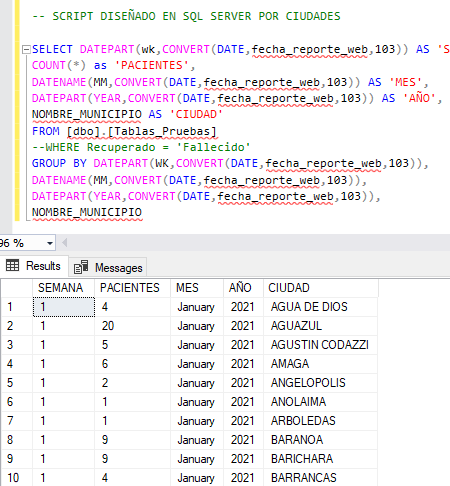
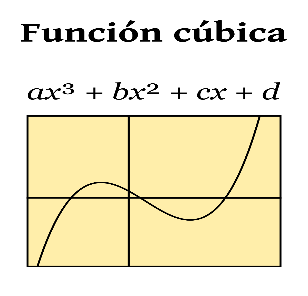


Fig. 1. Script en SQL Server

Ya con la información obtenida del script anterior, se realiza la descarga en formato .CSV para que luego sea el insumo que tomara el software realizado en PYTHON.

PYTHON es un lenguaje de programación interpretado cuya filosofía hace hincapié en la legibilidad de su código.2 Se trata de un lenguaje de programación multiparadigma, ya que soporta parcialmente la orientación a objetos, programación imperativa y, en menor medida, programación funcional. Es un lenguaje interpretado, dinámico y multiplataforma. Este lenguaje se caracteriza por ser usado a nivel mundial, como una herramienta de análisis de datos e inteligencia artificial.

## Fórmula utilizada para hallar método de regresión polinomial de tercer grado



Para una mayor facilidad del manejo de la información, se hizo uso de las siguientes librerías las cuales tiene una funcionalidad descrita al frente de cada uno:

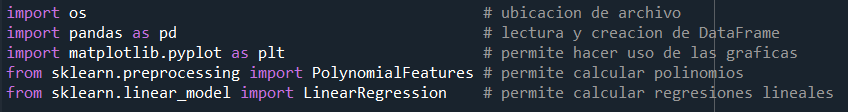


Fig. 2. Librerías de PYTHON

Realizado el llamado a las funciones, procedemos a ubicar la ruta y extracción de la información obtenida del script de SQL Server

Fig. 3. Dirección URL cargue archivo

Analizado el problema, nos dimos cuenta que habían procesos que estábamos repitiendo en diferentes ocasiones, por lo cual se definieron las siguientes funciones las cuales permiten graficar con mayor facilidad y hacer el cálculo de las predicciones requeridas.

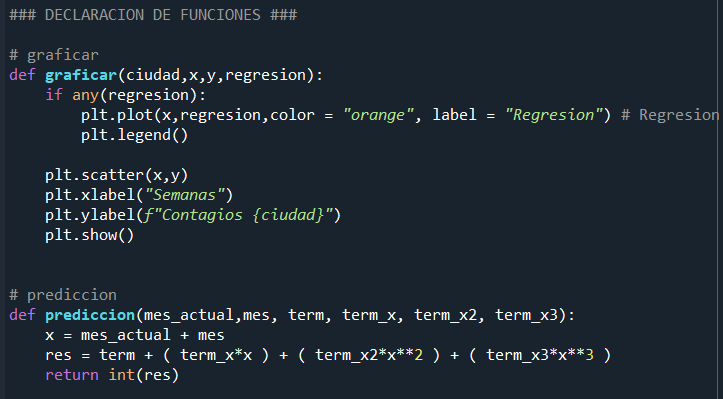


Fig. 4. Funciones

Los procedimientos descritos de este punto en adelante aplican tanto para el estudio de contagios a nivel nacional como para Bogotá.

Seguido a esto se grafica los datos obtenidos de la base de datos, se puede evidenciar los picos de contagios los cuales concuerdan con la situación vivida en los meses anteriores, esto para corroborar el respectivo funcionamiento del software en cuanto a su graficación. Dichas graficas fueron de dispersión, así que se procesa la información para obtención de datos acumulativos de contagios según la semana correspondiente. Se puede observar en los siguientes gráficos:

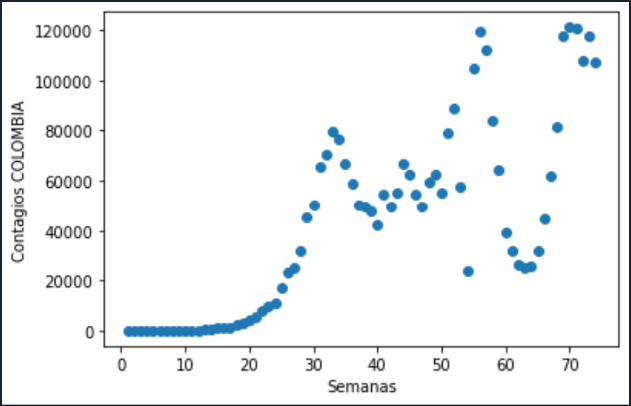


Fig. 4. Gráfico de dispersión nivel Colombia

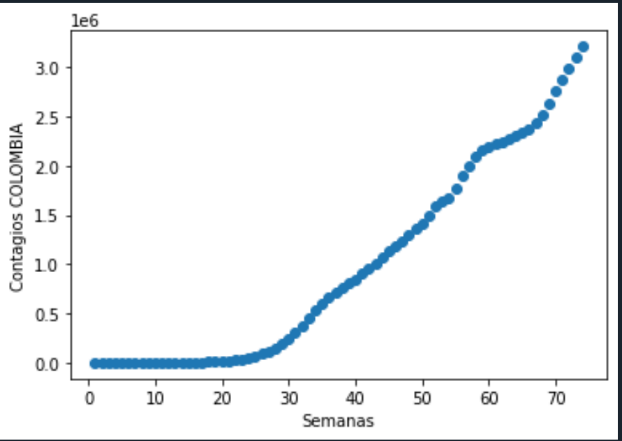
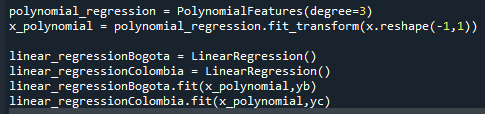


Fig. 5. Gráfico de casos acumulativos

La grafica por acumulación de datos, nos permite realizar un mejor análisis a nivel visual, ya que uno de los fines de este informe es predecir la cantidad de contagios para diferentes meses del presente año.

## Regresión polinomial cubica

En este punto se planteó el tipo de regresión a manejar, en nuestro caso fue una regresión polinomial de tercer orden, esta decisión se tomó con ayuda del lenguaje PYTHON y algunas de las librerías ya mencionadas como “PolynomialFeatures” y “LinearRegression”, esta última suele ser confusa por su nombre, ya que nos hace referencia a regresiones lineales y así es. Aunque una de las librerías nos provee de funciones especiales para regresiones lineales, la cual no vamos a implementar en este caso. También nos permite entrenar el algoritmo para que este estudie los datos y nos pueda generar los parámetros a manejar en nuestra regresión. Lo anteriormente mencionado se puede ver con más facilidad en la siguiente imagen seguida de una breve explicación.

Fig. 6. Obtención de la regresión

En la anterior imagen se muestra el procedimiento de obtención del polinomio, donde se creó una variable de tipo ‘PolynomialFeatures’ y se le asignó el parámetro ‘degree=3’ que hace referencia al grado que se manejó en el polinomio. Los datos almacenados en la variable ‘x’, que corresponden a las semanas en las que se registraron contagios, se le transforma a un polinomio, siguiente a esto, se crea una instancia de la clase ‘LinearRegresion’ y en esta enviaremos como parámetro los datos de los contagios acumulativos por semana y las semanas estudiadas. De esta forma entrenamos el algoritmo para que identifique por medio de un polinomio y los datos que debe tomar para cada Y (total contagios por semana), cuáles son los coeficientes más apropiados a usar en este caso en específico.

Hallada la respectiva regresión se muestra el resultado de esta, en la consola para conocer sus coeficientes





Fig. 7. Impresión de coeficiente por consola

Sabiendo que los polinomios de grado tres son de la forma tenemos, para el caso de Colombia los coeficientes son:

a = 137889,67

b = -36268,08

c = 1697,14

d = -9,09

Obtenidos los valores de los coeficientes se procede a calcular los valores que tomara en Y en dicho polinomio para poder graficar esta regresión sobre los datos que ya teníamos y así ver con más claridad que tanta desviación puede llegar a tener el método usado

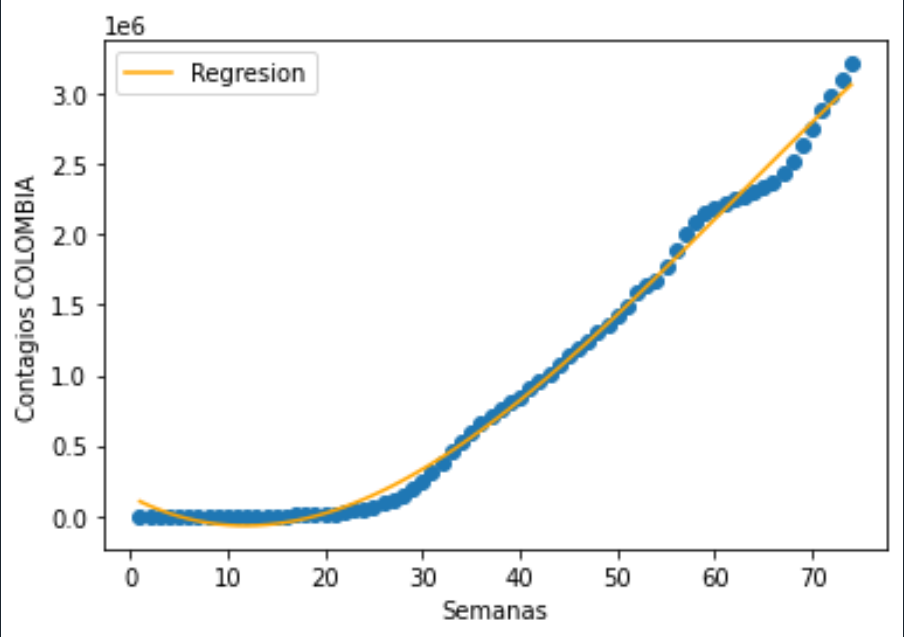


Fig. 8. Grafica de regresión

Vemos que nuestra regresión se dibuja sobre la gráfica de datos y esta parece ser muy acertada, y para confirmar que tan exacto puede llegar a ser este método podemos obtener el valor R cuadrado

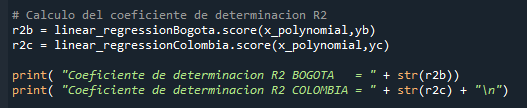




Fig. 9. Resultado de R2

El cual en Python es considerado como un puntaje, cuanta más cerca se encuentre del 1 más exacto será. En nuestro caso, para Colombia vemos que tenemos un puntaje de 0,99 lo cual nos representa que vamos por muy buen camino. Conociendo nuestro polinomio y sabiendo que este método tiene un 99% de confiabilidad se procede con la predicción de contagios para el mes de agosto y diciembre del año 2021

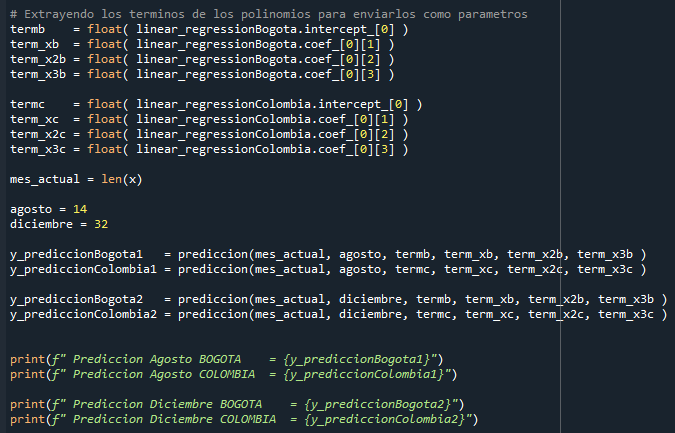


Fig. 10. Predicción de datos.

En este punto del algoritmo se extraen los respectivos coeficientes del polinomio de la variable ‘linear\_regressionColombia’ ya que fueron almacenados como un arreglo, estos datos fueron enviados como parámetro a la función ‘predicción’, definida anteriormente, junto con el valor del número de semana actual y el número de semana que se quiere predecir los contagios, este último valor se toma en referencia a la semana actual, por ejemplo, en el caso de agosto seria la semana número catorce contando desde nuestra semana actual. Se realiza el mismo procedimiento para diciembre y se obtienen los siguientes datos.

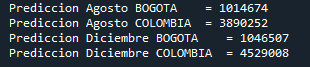


Fig. 10. Predicción

Para ver el comportamiento más detallado de estos datos de predicción, se realizó una gráfica en la cual se observa cual podría llegar a ser el futuro de los contagios.

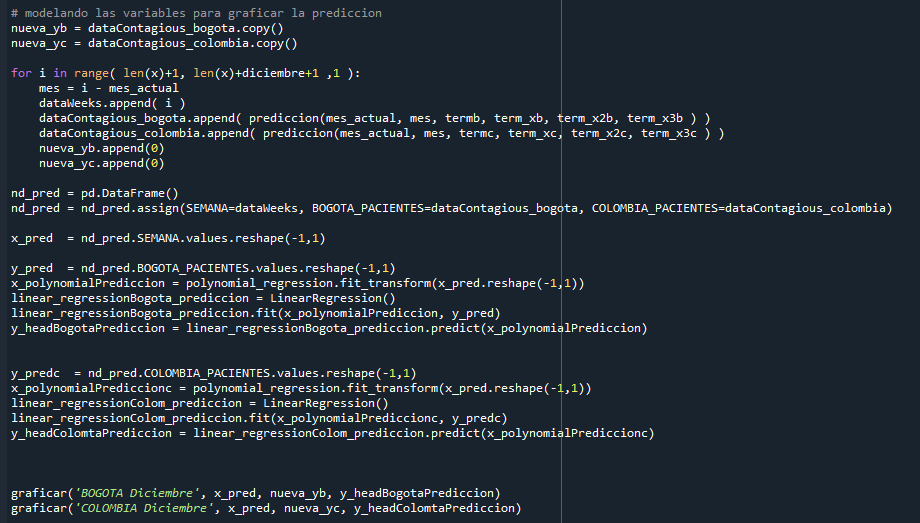


Fig. 11. Modelamiento de variables

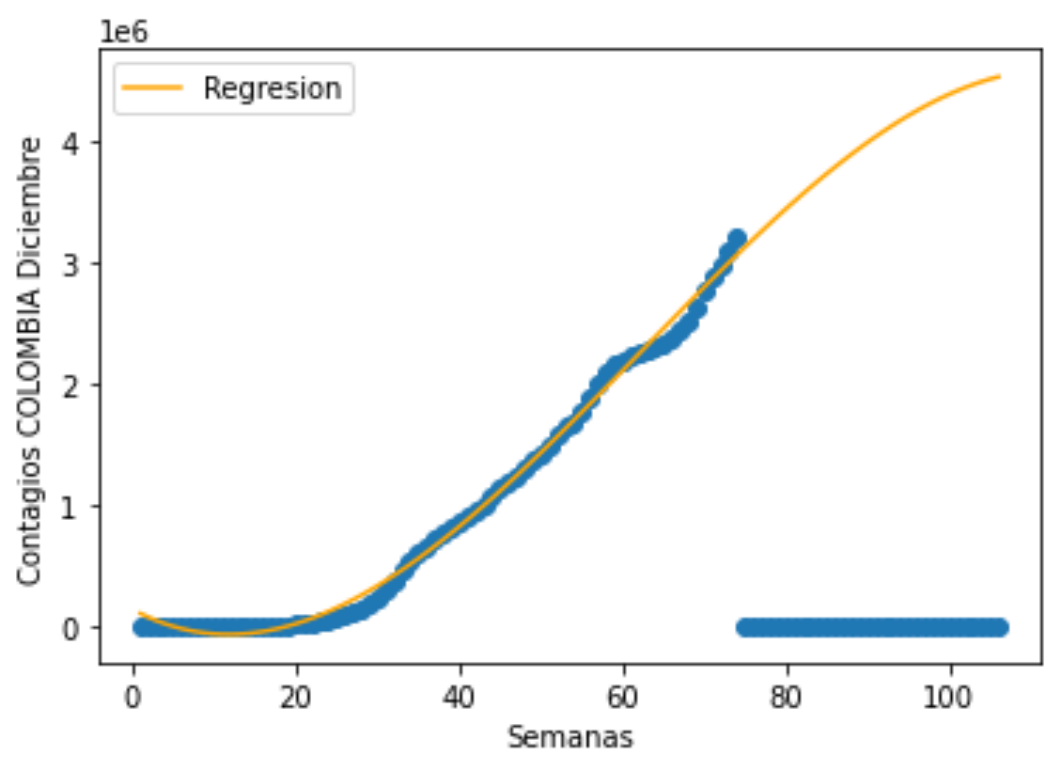
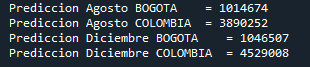


Fig. 12. Grafica de predicción en diciembre

Esta grafica se hizo con la predicción de diciembre para tener un panorama más amplio de lo que resta del año.

## ¿Cuentas personas se estima que habrán contagiadas de COVID-19 en Colombia y en Bogotá D.C para agosto y diciembre 2021?



### *Conclusiones*

A partir del análisis obtenido, se puede deducir que las herramientas utilizadas son las apropiadas para este tipo de escenarios. Debido a que da solución una necesidad de manera rápida y concisa, en donde la probabilidad de falla a la predicción es mínima frente a otros métodos. Las predicciones mencionadas para agosto y diciembre pueden variar, debido a la situación que el país está afrontando, donde en las aglomeraciones no se respeta el distanciamiento y esto puede llegar a ser un factor que altere muchos más las cifras en un menor tiempo.

### REFERENCIAS

1. <https://www.who.int/es/news-room/q-a-detail/coronavirus-disease-covid-19>, Que es la COVID 19.
2. <https://espanol.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html#:~:text=El%20COVID%2D19%20se%20propaga,sus%20ojos%2C%20nariz%20o%20boca>., Cómo se propaga el COVID-19
3. <https://www.datos.gov.co/Salud-y-Protecci-n-Social/Casos-positivos-de-COVID-19-en-Colombia/gt2j-8ykr/data>, Datos obtenidos de contagios en Colombia.
4. https://news.google.com/covid19/map?hl=es-419&gl=CO&ceid=CO%3Aes-419&mid=%2Fm%2F09c7w0, Datos de contagios a nivel mundial.
5. <https://github.com/CSSEGISandData/COVID-19/blob/master/README.md>, Repositorio Github de datos a nivel mundial
6. <https://www.arcgis.com/apps/dashboards/index.html#/bda7594740fd40299423467b48e9ecf6>, COVID-19 Dashboard.
7. https://openwebinars.net/blog/que-es-sql-server/, Que es SQL Server
8. <https://es.wikipedia.org/wiki/Python#Caracter%C3%ADsticas_y_paradigmas>, que es python
9. [1200px-Función\_cúbica.svg.png (1200×1597) (wikimedia.org)](https://upload.wikimedia.org/wikipedia/commons/thumb/8/88/Funci%C3%B3n_c%C3%BAbica.svg/1200px-Funci%C3%B3n_c%C3%BAbica.svg.png), Documentación de regresiones en Python