

Análisis exploratorio de los casos confirmados de COVID-19 en Bogotá, Colombia

Alfonso Navas

Research Software Engineering in Python

Módulo de análisis de datos

Marzo de 2021



Latin American alliance for
Capacity building in Advanced physics

LA-CoNGA physics

Introducción y datos utilizados

Dada la coyuntura actual causada por la pandemia por COVID-19, se aplicaron las herramientas adquiridas en el curso *Research Software Engineering in Python* para realizar un análisis exploratorio sobre los datos de los casos confirmados de esta enfermedad en Bogotá D.C., Colombia.

The screenshot shows the 'Datos Abiertos Bogotá' website. The header includes the logo, a search bar, and navigation links: 'Inicio de Sesión para Entidades', 'Datos', 'Entidades', 'Temáticas', 'Acerca de', and 'Ayuda'. The breadcrumb trail reads: 'HOME / DATOS / CASOS CONFIRMADOS DE COVID-19 EN BOGOTÁ D.C.'. The main title is 'Casos confirmados de COVID-19 en Bogotá D.C.'. Below it, a description states: 'Permite conocer la evolución del número de casos confirmados con nuevo coronavirus COVID-19 en el contexto de la pandemia.' The source is 'Secretaría Distrital de Salud'. There are icons for Creative Commons BY license, a download icon with '449 Descargas', and a visit icon with '56105 Visitas'. The data table has two rows: 1) 'Casos confirmados de COVID-19 en Bogotá D.C.' with a CSV icon and buttons for 'Previsualización' and 'Descargar'. 2) 'Metadatos Número de casos confirmados por el ...' with a CSV icon, a description 'Metadatos: Descripción de los datos relacionados en el conjunto de datos "Número de casos confirmados por el laboratorio de COVID- 19"', and buttons for 'Previsualización' and 'Descargar'.

Los datos se descargaron el 23 de febrero de 2021 directamente de la plataforma distrital de datos abiertos de Bogotá D.C.

Enlace de la fuente:
<https://datosabiertos.bogota.gov.co/dataset/numero-de-casos-confirmados-por-el-laboratorio-de-covid-19-bogota-d-c>

Herramientas utilizadas



- Este análisis fue realizado en su totalidad dentro de la *13 TeV ATLAS Open Data virtual machine (VM)*, usando el ambiente computacional *Jupyter Notebook* y el lenguaje de programación *Python*.

La información necesaria para descargar e instalar esta VM puede encontrarse en:

<http://opendata.atlas.cern/release/2020/documentation/vm/>

- Las bibliotecas que usadas fueron: **Numpy**, **Pandas**, **Matplotlib**

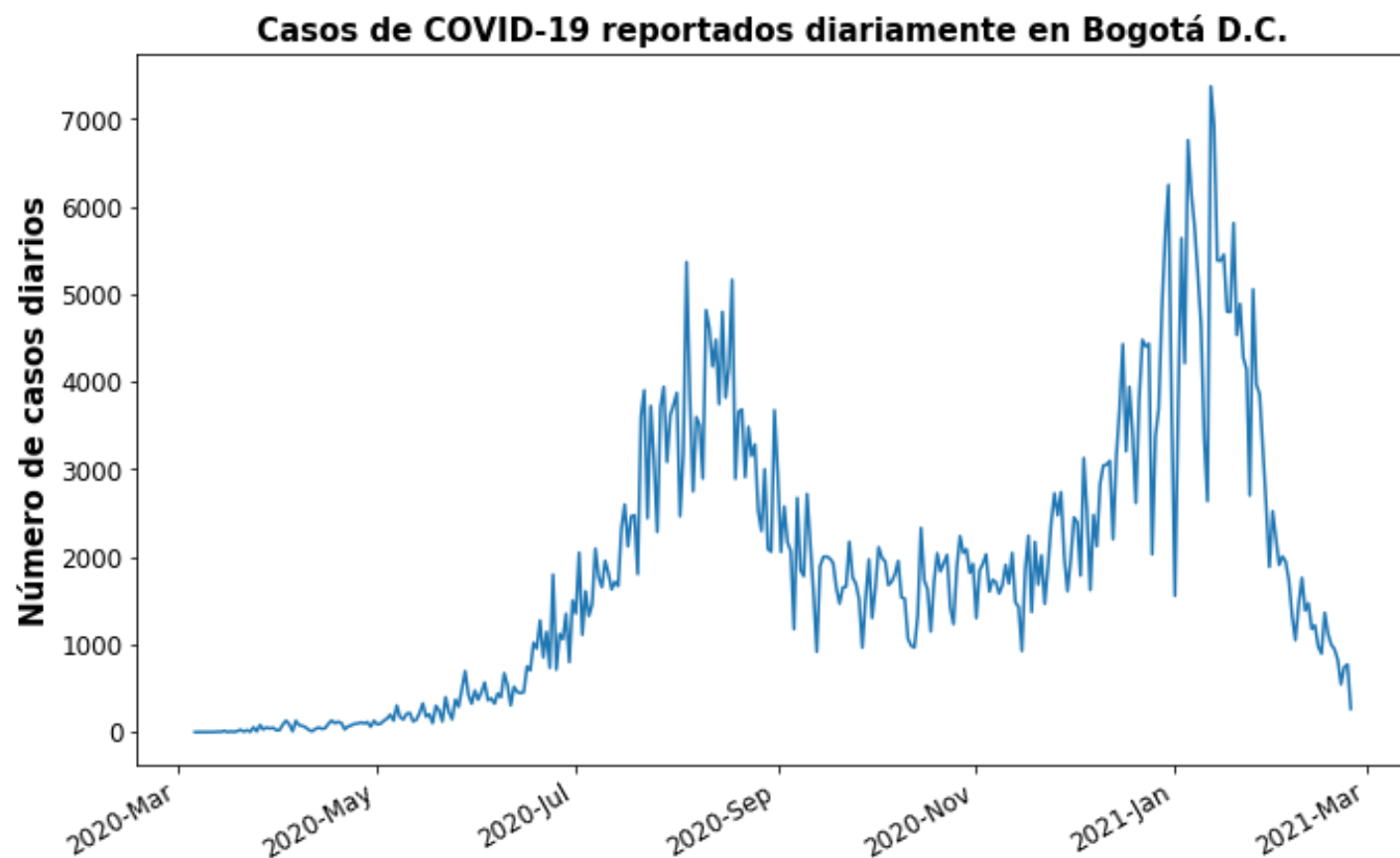
- Los datos estaban en formato `.csv` y consistían en una tabla de 655.252 filas \times 10 columnas. En cada fila se reportaba un caso y las columnas contenían información sobre:

La fecha de inicio de síntomas, la fecha de diagnóstico, ciudad, localidad, edad, sexo, fuente o tipo de contagio, ubicación y estado.

- Ninguna columna presentó espacios vacíos excepto la de **la fecha de inicio de síntomas**. Para evitar ambigüedades, esta información fue descartada de nuestro análisis.

[illegible]

Número de casos confirmados diariamente en Bogotá D.C.



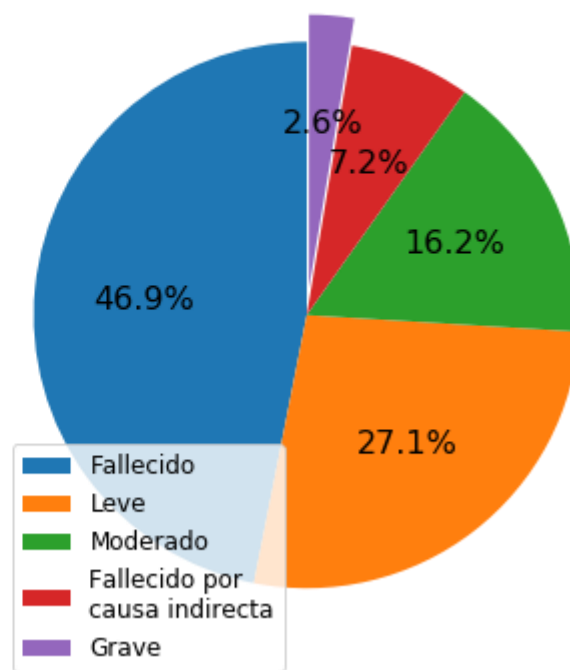
- Se construyó una serie de tiempo contando el número de casos reportados cada día.
- En esta gráfica se verifica que la ciudad ha atravesado por **dos picos epidemiológicos**. El primero ocurrió en **agosto del 2020** y el segundo en **enero de 2021**.

Estado general de los casos confirmados con COVID-19 en Bogotá D.C.

Porcentaje de recuperados frente al total de casos reportados



Otros casos



De los **655.252** casos confirmados de COVID-19 se tiene que:

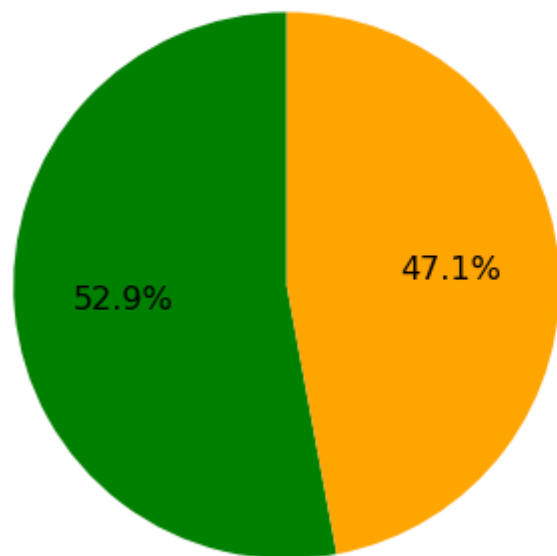
- El **95.5%** de pacientes se ha recuperado.

El otro **4.5%** está compuesto por:

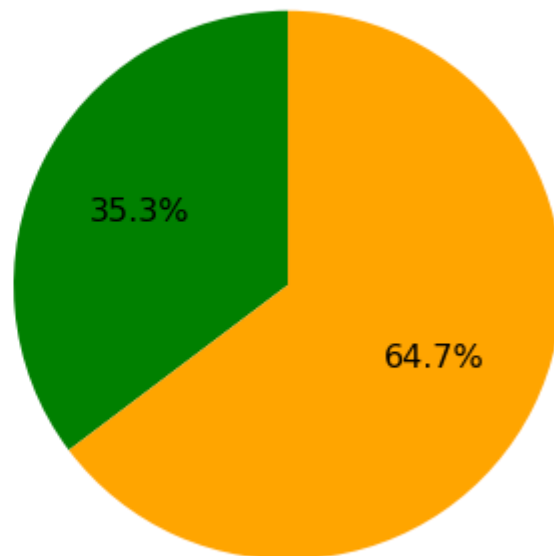
- **Fallecidos por consecuencias directas** asociadas a COVID-19 (**46.9%, 2.1% del total**)
- **Fallecidos por causa indirecta** de la COVID-19 (**7.2%**) y pacientes con **síntomas**
- **leves** (**27.1%**)
- **moderados** (**16.2%**)
- **graves** (**2.6%**)

Distribución de casos e incidencia de muertes por COVID-19 por sexo

Porcentaje de casos reportados por sexo

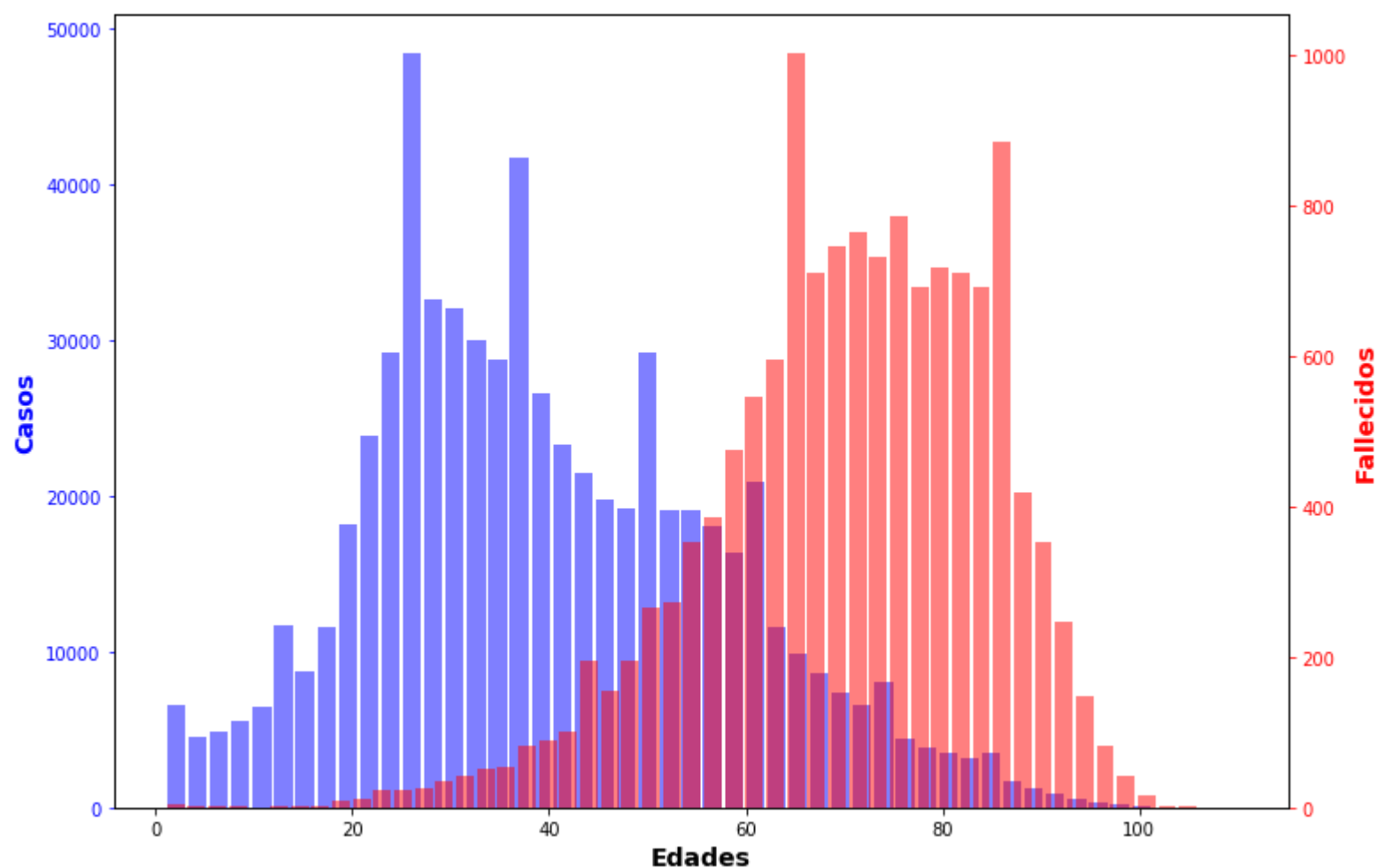


Incidencia de muertes por COVID-19 por sexo



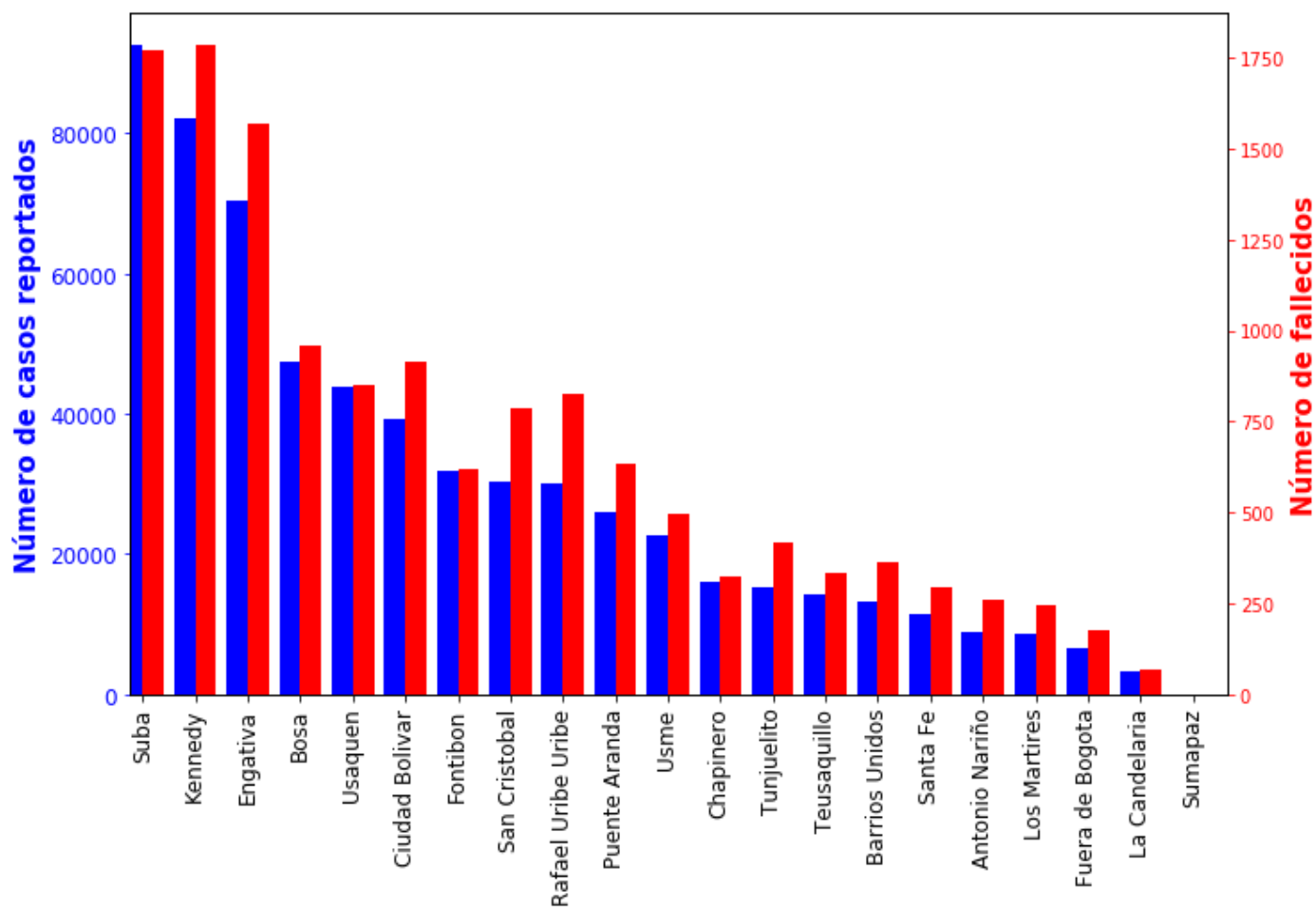
De los datos vemos que aunque son las personas del sexo **femenino** las que cuentan con un **mayor número de diagnósticos** de COVID-19, el **número de muertes** son mucho más elevadas entre personas del **sexo masculino**.

Distribución de casos e incidencia de muertes por COVID-19 por edad



Se observa que, aunque la **mayoría de casos** reportados fueron de personas entre los **25 y los 35 años**, hay un incremento desproporcional en el número de muertes al aumentar la edad, siendo **las más vulnerables las personas de la tercera edad**. Por lo tanto, se verifica que **la edad es definitivamente un factor de riesgo importante para la muerte por COVID-19**.

Distribución de casos e incidencia de muertes por COVID-19 por localidad



Al observar la gráfica nos preguntamos si **algunas localidades pudieron haber experimentado tasas de mortalidad más altas que las de otras.**

Para responder a esta pregunta:

1. Se calculó, a partir de los datos, **el número de fallecidos por cada 1000 casos confirmados de COVID-19** para cada localidad.
2. Se hizo una prueba de **aleatorización** Para determinar si la diferencia entre 2 localidades para esta variable es estadísticamente significativa.

Para este ejercicio, se compararon solamente dos localidades: **Barrios Unidos y Suba.**

Prueba de aleatorización (*Randomization Test*)

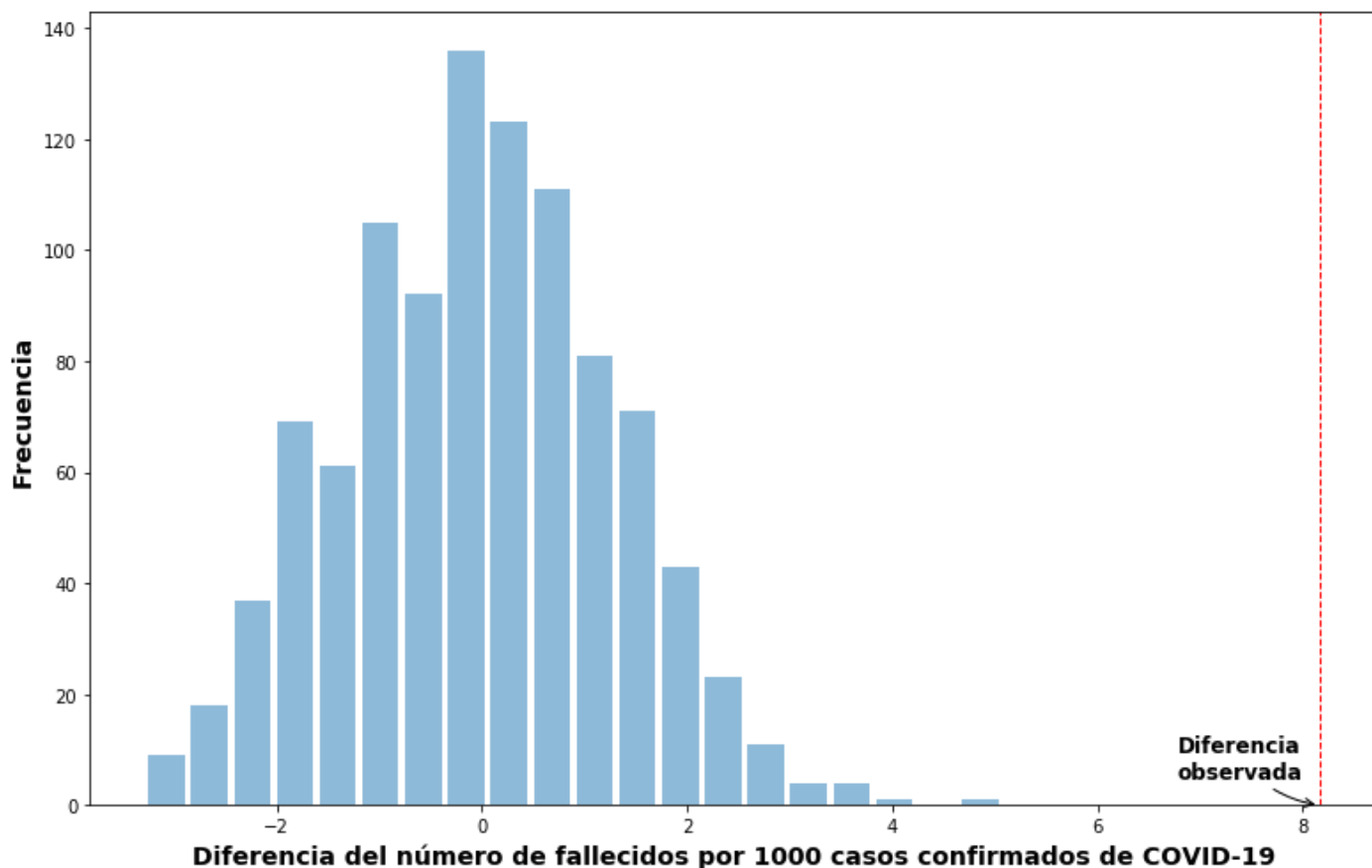
Partiendo de la **hipótesis nula**:

"Las diferencias en el número de fallecidos por 1000 casos confirmados de COVID-19 en los datos entre las localidades de Barrios Unidos y Suba se deben netamente al azar."

Se siguieron los siguientes pasos:

1. Se generaron 1000 parejas de vectores a partir de un muestreo aleatorio de los datos de ambas localidades.
2. Para cada pareja se calculó la diferencia de entre el número de fallecidos por 1000 casos confirmados de COVID-19, obteniendo así una "distribución" de probabilidad para esta variable aproximada a la que resultaría de ser cierta la hipótesis nula.
3. Por último, se comparará la diferencia del número de fallecidos por 1000 casos confirmados de COVID-19 entre Barrios Unidos y Suba calculada a partir de los datos con la distribución obtenida mediante el proceso de aleatorización.

Resultados de la prueba de aleatorización



- Verificamos de esta forma que **la diferencia del número de fallecidos por 1000 casos reportados de COVID-19 entre las localidades de Barrios Unidos y Suba es estadísticamente muy significativa.**
- Por lo tanto, podemos decir que **la pandemia en Bogotá no ha afectado de la misma manera a nivel de localidades.**
- Por otra parte, determinar los factores que generan esta discrepancia es un asunto complejo que requiere mucha más información que con la que se cuenta actualmente (que además se escapa a los límites de este proyecto).