

UNIVERSIDAD ARGENTINA DE LA EMPRESA - UADE

Ciencia de Datos – Trabajo Práctico Obligatorio

1er. Cuatrimestre del 2025

Docente: Francisco, Mariano Daniel

**COVID-19 en Argentina: Casos, Vacunación y Defunciones – Análisis Exploratorio y Predictivo**

Grupo 1

Alumnos:

Martinez Blanco, Abril - LU: 1156495

Cardozo, Ivana Elizabeth - LU: 1071244

Cahuana, Kevin - LU: 1061135

Vignolo, Nahir - LU: 1137587

1. Descripción

Este proyecto de Ciencia de Datos tiene como objetivo analizar el impacto de la pandemia de COVID-19 en la República Argentina durante los años 2020 y 2021. Utilizando datos oficiales abiertos del Ministerio de Salud de la Nación, se estudian los casos confirmados, las defunciones generales y las dosis de vacunas aplicadas.

A través de gráficos, análisis estadísticos y agrupaciones por edad, provincia y causa de muerte, buscamos identificar patrones y validar hipótesis clave: exceso de mortalidad, subregistro de muertes por COVID, impacto desigual por provincias y análisis demográfico del impacto.

El resultado de este análisis permitirá generar conocimientos aplicables a la gestión de futuras pandemias, así como también sentar bases para un modelo predictivo de evolución sanitaria.

1. Roles

### **Analista del Negocio y de la Problemática**

Responsable: Abril Martinez Blanco  
**Tareas:**

* Presentar el contexto del proyecto y la importancia del análisis
* Definir las preguntas que se buscaban responder (hipótesis)
* Explicar a quién va dirigido y por qué es útil para la salud pública

### **Responsable del Análisis de Datos y Limpieza**

Responsable: Nahir Vignolo  
**Tareas:**

* Contar cómo se cargaron los datos
* Explicar cómo se limpiaron: nulos, columnas, filtros
* Mostrar cómo se organizaron los DataFrames

### **Responsable del Análisis Exploratorio y Visualizaciones**

Responsable: Ivana Cardozo  
**Tareas:**

* Mostrar gráficos hechos en Colab
* Explicar lo que se ve en cada visualización
* Relacionar visualizaciones con hipótesis

### **Responsable de Visualización Interactiva y Conclusiones**

Responsable: Kevin Cahuana  
**Tareas:**

* Mostrar cómo se pasó de Colab a Google Sheets
* Enseñar el dashboard en Power BI
* Resumir hallazgos y propuesta de valor

1. Tubería de Datos

3.1 Ingesta de Datos (Fuentes)

**Datasets utilizados:**

1. Casos de COVID-19

*Fuente: Ministerio de Salud de la Nación*

<https://datos.gob.ar/dataset/salud-covid-19-casos-registrados-republica-argentina>

**Detalle:** Casos confirmados por fecha, edad, sexo, jurisdicción, método de diagnóstico, etc.

1. Vacunación (Dosis aplicadas)

*Fuente: Ministerio de Salud de la Nación*

<https://datos.gob.ar/dataset/salud-vacunas-contra-covid-19-dosis-aplicadas-republica-argentina>

**Detalle:** Registro diario de dosis aplicadas por tipo, grupo etario, sexo y provincia.

1. Defunciones Generales

*Fuente: Dirección Nacional de Estadísticas de Salud*

http://datos.salud.gob.ar/dataset/27c588e8-43d0-411a-a40c-7ecc563c2c9f/archivo/fab9e990-865c-43c4-a643-3dbc3b70a934

**Detalle:** Registro anual de defunciones por provincia, sexo, edad y causa básica (según CIE-10). Incluye todas las muertes, no sólo por COVID.

3.2 Procesamiento y Limpieza de Datos

* **Herramientas:** Google Colab (Python)
* **Librerías:** *pandas*, *seaborn*, *matplotlib*
* **Tareas realizadas:**
  + Eliminación de valores nulos
  + Filtrado de columnas relevantes
  + Agrupación por provincia, edad, año y causa
  + Generación de nuevos DataFrames limpios

3.3 Exportación de Datos Limpios

Una vez finalizado el procesamiento y la limpieza de los datasets en Google Colab, exportamos los DataFrames en formato .csv, utilizando funciones propias de la librería pandas. Esta etapa fue fundamental para generar archivos intermedios con los datos estructurados y filtrados, listos para ser utilizados en herramientas de visualización.

Los archivos .csv se descargaron localmente y luego fueron subidos manualmente a Google Drive. Desde allí, los abrimos con Google Sheets para transformarlos en fuentes de datos compatibles con Power BI, donde creamos visualizaciones interactivas y dashboards dinámicos. Esta integración permitió que el trabajo de análisis realizado en Python pudiera ser presentado de manera más accesible, intuitiva y comprensible para una audiencia técnica y no técnica.

Este paso fue clave para vincular el análisis exploratorio realizado con Python y la propuesta de valor final del proyecto, facilitando una presentación visual clara y filtrable por variables como provincia, año, grupo etario o causa de muerte.

3.4 Visualización Interactiva

* **Herramienta:** Power BI
* **Conexion:** A traves de Google Sheets
* **Funcionalidad:**
  + Visualización de defunciones por año, provincia y grupo etario
  + Gráficos comparativos con dosis aplicadas
  + Filtros interactivos por región y fecha

3.5 Análisis y Conclusión

* Validación de hipótesis planteadas (exceso de mortalidad, impacto geográfico, causas de muerte)
* Generación de modelos exploratorios para posibles predicciones
* Insights accionables para autoridades sanitarias

1. Información del Negocio

El análisis de datos de la pandemia de COVID-19 en Argentina tiene un alto valor estratégico para organismos de salud pública. A través de este proyecto, buscamos comprender el impacto real de la pandemia en términos de contagios, defunciones y vacunación, analizando tanto la información oficial como posibles indicadores indirectos como el exceso de mortalidad. Esta información permite generar aprendizajes clave que podrían aplicarse en caso de un nuevo brote o variante.

1. Desarrollo de la Hipótesis

**¿Qué es lo que intentamos proponer o resolver?**

El objetivo principal del trabajo es identificar patrones, correlaciones y posibles inconsistencias entre los casos confirmados de COVID-19, las defunciones generales y las dosis de vacunación aplicadas en la República Argentina durante el período 2020–2022. Queremos evaluar el impacto real de la pandemia, más allá de los números reportados, y entender su evolución desde múltiples dimensiones (temporal, geográfica, demográfica y clínica).

**Hipótesis de análisis**

1. **Aumento significativo de defunciones durante la pandemia**

Hipótesis:

Durante 2020 y 2021, las defunciones aumentaron significativamente respecto a años anteriores debido a la pandemia de COVID-19.

1. **Subregistro o discrepancia entre defunciones y muertes por COVID registradas oficialmente**

Hipótesis:

Existe una diferencia importante entre la cantidad de muertes por COVID-19 registradas oficialmente y el exceso de defunciones totales, lo que podría indicar subregistro o clasificación incorrecta.

1. **Impacto geográfico diferencial de la pandemia**

Hipótesis:

Algunas provincias muestran un mayor exceso de defunciones relativas en comparación con su tasa de contagios, lo cual podría reflejar desigualdades en el acceso a servicios de salud o en la infraestructura sanitaria.

1. **Análisis demográfico del impacto**

Hipótesis:

Los grupos etarios mayores fueron los más afectados tanto en contagios como en defunciones, y se observan cambios en las principales causas de muerte durante el período pandémico (por ejemplo, aumento de causas respiratorias y reducción en otras como accidentes o infartos).

**Audiencia**

Este análisis está dirigido a la Gerencia General y Técnica, con el objetivo de ofrecer insumos clave para la planificación estratégica en situaciones de emergencia sanitaria. También se espera que los hallazgos puedan ser utilizados por equipos de epidemiología, salud pública y analistas de riesgo.

**¿Cómo se resuelve la hipótesis para la entrega de valor?**

Dado que existe la posibilidad de que surja una nueva variante del COVID-19 o una pandemia similar en el futuro, es fundamental anticipar cómo podrían evolucionar los indicadores sanitarios clave, especialmente los contagios y las defunciones.

Para esto, proponemos el desarrollo de un modelo predictivo basado en el análisis histórico de:

* Tendencias de contagios y defunciones
* Efecto de las campañas de vacunación
* Factores demográficos (edad, sexo)
* Medidas sanitarias implementadas (cuarentenas, restricciones, etc.)

Este modelo permitirá:

* Simular escenarios futuros bajo diferentes condiciones.
* Apoyar decisiones de política pública y gestión sanitaria.
* Optimizar recursos en momentos críticos.
* Evaluar la eficacia de estrategias anteriores ante posibles rebrotes.

1. Propuesta de Valor

El análisis realizado permite detectar desigualdades estructurales en el impacto de la pandemia y evaluar la efectividad de las políticas sanitarias implementadas.

Al observar la tasa de letalidad, la distribución de vacunas y el exceso de mortalidad por provincia y grupo etario, se identifican posibles brechas en el acceso al sistema de salud. Esto proporciona insumos clave para mejorar la distribución de recursos y tomar decisiones más equitativas en futuras emergencias.

Además, los resultados permiten revisar los cambios en las causas de defunción más frecuentes durante los años críticos de la pandemia. Esta información es fundamental para rediseñar estrategias de prevención y mitigación, generando valor para la salud pública y aportando herramientas a los responsables de políticas sanitarias.