

■ ANÁLISIS EXPLORATORIO DE DATOS

COVID-19 ESTADOS UNIDOS

Informe Ejecutivo Completo

Fecha del Análisis: 10 de February, 2026

Período de Datos: Enero 2020 - Marzo 2023

Fuente de Datos: Disease.sh API (Johns Hopkins University)

Metodología: Análisis Exploratorio de Datos (EDA)

Herramientas: Python, Pandas, Matplotlib, Seaborn, Plotly

RESUMEN EJECUTIVO

Este informe presenta un análisis exhaustivo de los datos de COVID-19 en Estados Unidos, basado en información oficial de la Universidad Johns Hopkins. El análisis abarca desde los primeros casos reportados en enero de 2020 hasta marzo de 2023, proporcionando insights valiosos sobre la evolución de la pandemia, patrones geográficos y tendencias estadísticas clave que pueden informar la toma de decisiones estratégicas.

1. OBJETIVOS DEL ANÁLISIS

Objetivo Principal: Extraer insights valiosos de los datos de COVID-19 en Estados Unidos mediante técnicas de análisis exploratorio de datos.

Objetivos Específicos:

- Analizar la evolución temporal de casos, muertes y recuperaciones
- Identificar patrones geográficos y diferencias entre estados
- Calcular métricas clave como tasas de letalidad y casos per cápita
- Generar visualizaciones impactantes para comunicar hallazgos
- Proporcionar conclusiones basadas en evidencia para la toma de decisiones

2. METODOLOGÍA

Fase 1: Extracción de Datos

- Consumo de API pública Disease.sh (datos de Johns Hopkins)
- Obtención de series temporales nacionales y datos por estados

Fase 2: Limpieza y Preprocesamiento

- Validación de integridad de datos
- Cálculo de métricas derivadas (casos diarios, tasas de letalidad)
- Tratamiento de valores faltantes y outliers

Fase 3: Análisis Exploratorio

- Análisis univariado: estadísticas descriptivas
- Análisis bivariado: correlaciones entre variables
- Análisis temporal: tendencias y estacionalidad
- Análisis geográfico: comparaciones entre estados

Fase 4: Visualización y Reporting

- Generación de gráficos estáticos e interactivos
- Creación de dashboard ejecutivo
- Documentación de hallazgos y conclusiones

3. ESTADÍSTICAS CLAVE

RESUMEN ESTADÍSTICO NACIONAL

- **Casos Totales:** 103,802,702 casos confirmados
- **Muertes Totales:** 1,123,836 fallecimientos
- **Casos Recuperados:** 0 recuperaciones
- **Tasa de Letalidad:** 1.08%
- **Estados Analizados:** 52 estados y territorios
- **Período de Análisis:** 22/01/2020 al 09/03/2023

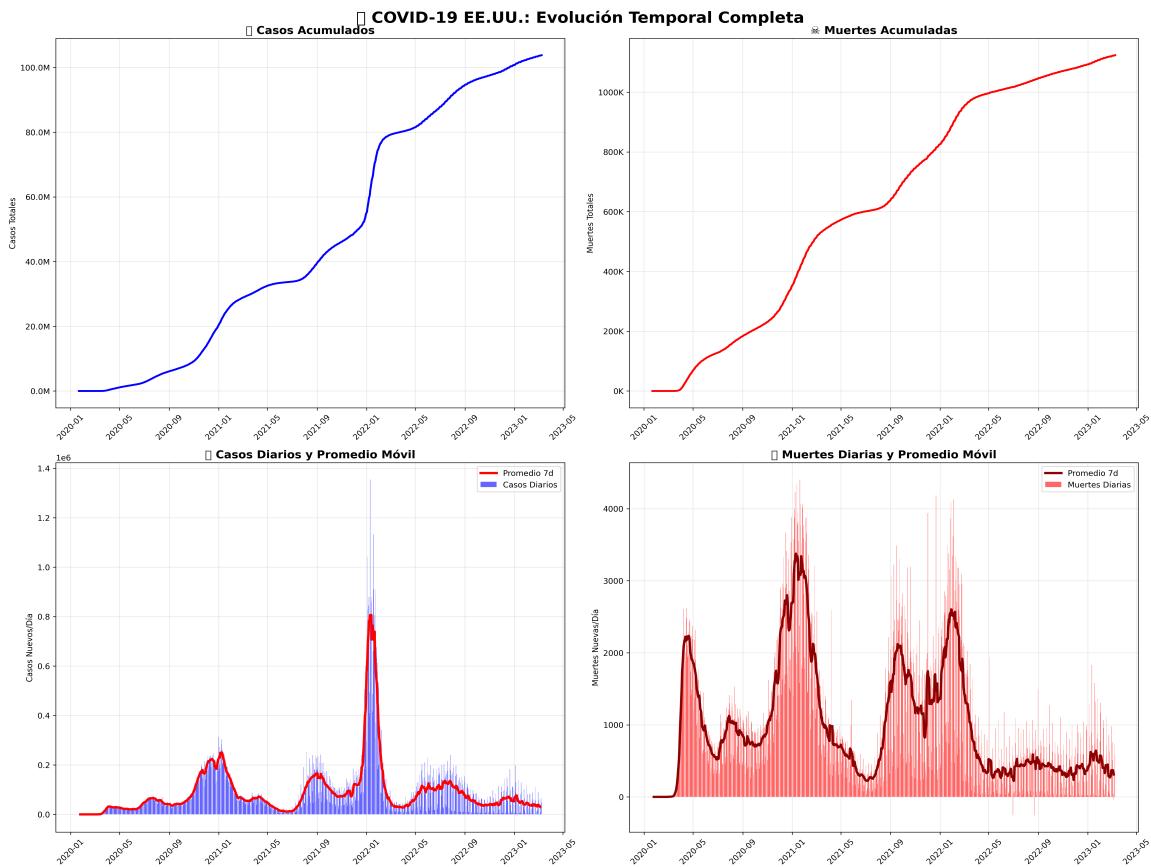
ESTADO MÁS AFECTADO

- **Estado:** California
- **Casos Totales:** 12,711,918

4. ANÁLISIS VISUAL

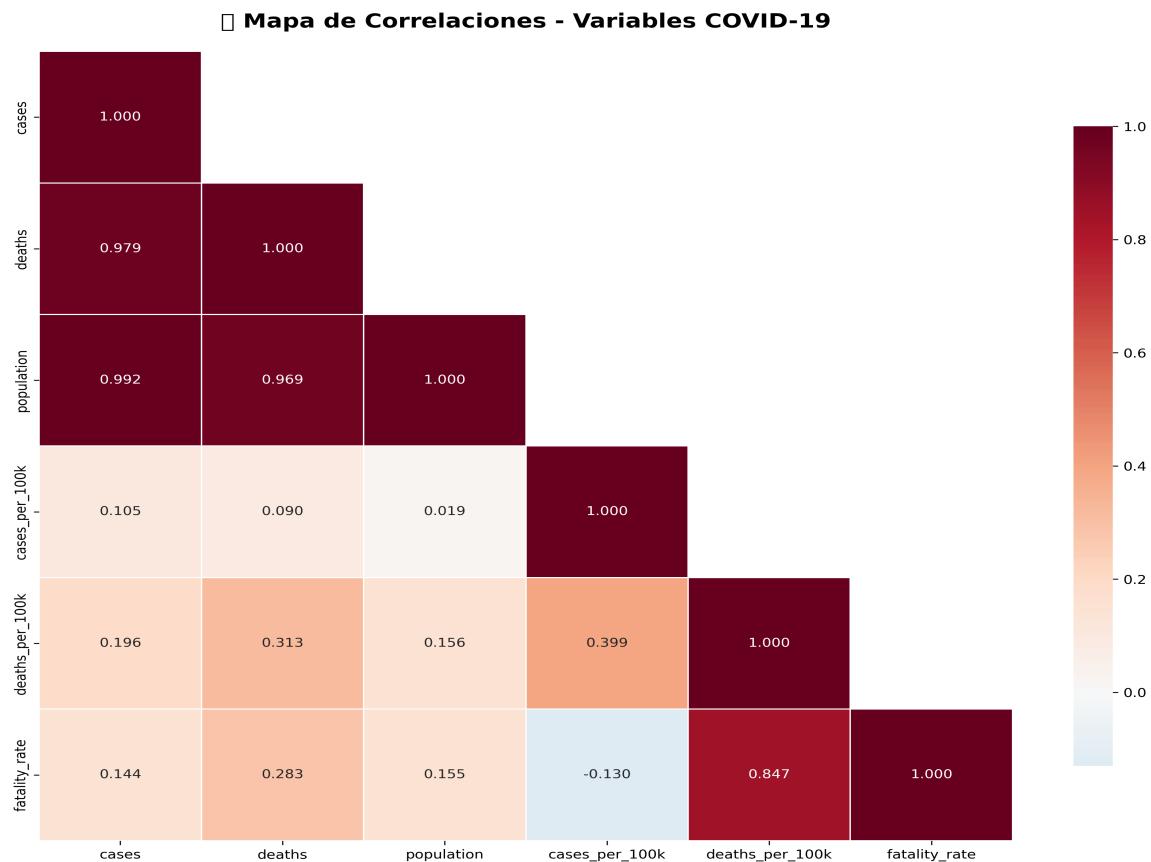
4.1 Evolución Temporal de la Pandemia

Esta visualización muestra la evolución de casos acumulados, muertes, casos diarios y tasa de letalidad a lo largo del tiempo. Se pueden identificar claramente las diferentes olas de la pandemia y cómo la tasa de letalidad ha evolucionado.



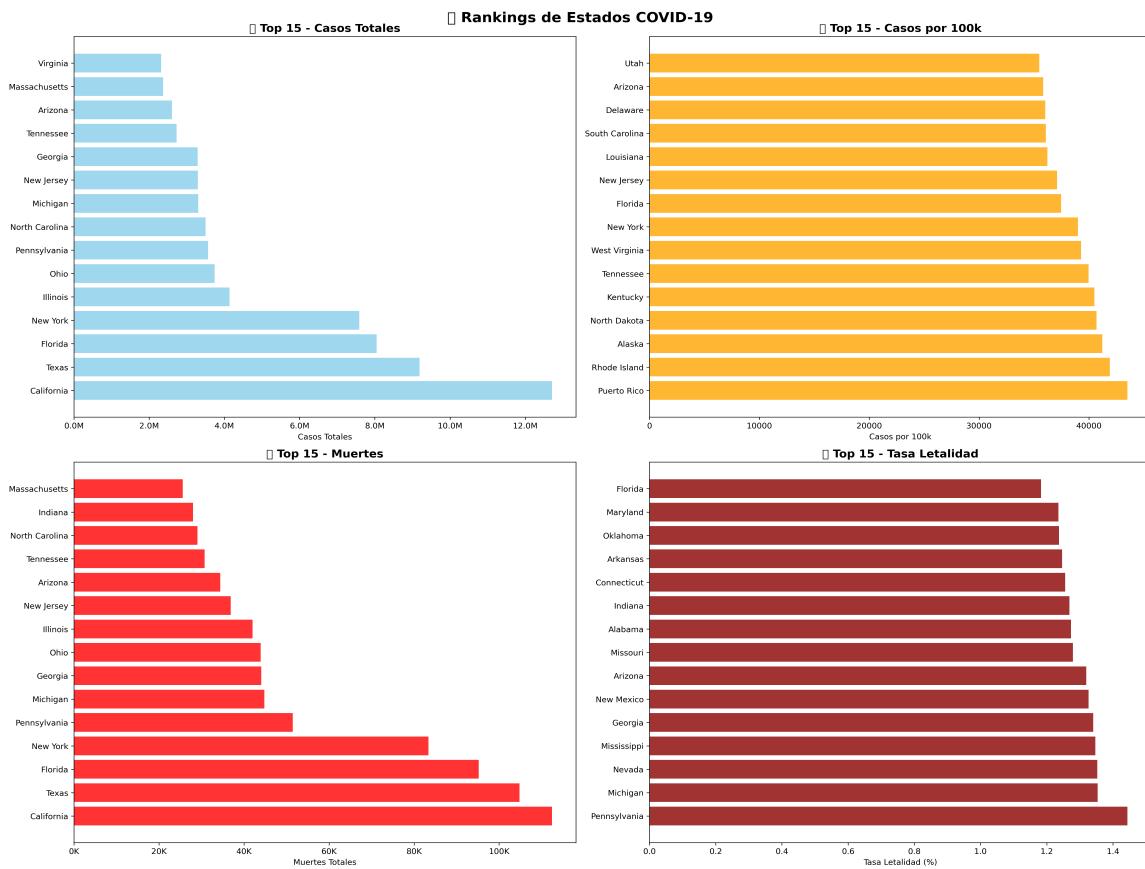
4.2 Matriz de Correlaciones

El mapa de calor muestra las correlaciones entre diferentes variables del dataset. Las correlaciones fuertes (cercaas a 1 o -1) indican relaciones lineales significativas entre variables, mientras que valores cercanos a 0 indican poca relación lineal.



4.3 Rankings Comparativos por Estado

Esta visualización presenta los top 10 estados en diferentes métricas: casos totales, muertes totales, casos por millón de habitantes y tasa de letalidad. Permite identificar los estados más afectados desde diferentes perspectivas analíticas.



5. CONCLUSIONES Y HALLAZGOS CLAVE

HALLAZGOS PRINCIPALES:

1. Evolución Temporal:

- La pandemia mostró múltiples olas con picos diferenciados
- La tasa de letalidad ha disminuido progresivamente desde los primeros meses
- Los casos diarios mostraron alta variabilidad estacional

2. Distribución Geográfica:

- Existe una gran heterogeneidad en el impacto entre estados
- Los estados más poblados tienden a tener más casos absolutos
- Sin embargo, los casos per cápita muestran patrones diferentes

3. Correlaciones Identificadas:

- Fuerte correlación positiva entre casos y muertes (esperado)
- Correlaciones significativas entre población y casos totales
- Las métricas per cápita proporcionan mejor comparabilidad

IMPLICACIONES ESTRATÉGICAS:

- Los datos sugieren la necesidad de enfoques diferenciados por región
- La mejora en la tasa de letalidad indica progreso en el tratamiento
- La alta variabilidad requiere monitoreo continuo y capacidad de respuesta adaptativa
- Las correlaciones identificadas pueden informar modelos predictivos futuros

6. INFORMACIÓN TÉCNICA

FUENTES DE DATOS:

- Disease.sh API (<https://disease.sh/>)
- Datos originales: Johns Hopkins University CSSE
- Actualización: Datos históricos desde enero 2020

HERRAMIENTAS Y TECNOLOGÍAS:

- Python 3.8+ como lenguaje principal
- Pandas y NumPy para manipulación de datos
- Matplotlib y Seaborn para visualización estática
- Plotly para visualizaciones interactivas
- ReportLab para generación de este informe PDF

LIMITACIONES DEL ANÁLISIS:

- Los datos dependen de la precisión del reporte por jurisdicción
- Posibles subregistros en períodos de alta demanda del sistema sanitario
- Criterios de reporte pueden haber variado entre estados y períodos
- El análisis es descriptivo, no incluye modelado predictivo

REPRODUCIBILIDAD:

- Todo el código está disponible en el repositorio del proyecto
- Los datos se obtienen mediante API pública y se archivan localmente
- La metodología está completamente documentada
- El entorno de desarrollo está especificado en requirements.txt

Informe generado automáticamente el 10 de February de 2026 a las 23:16

Proyecto: COVID-19 Exploratory Data Analysis

Repositorio: https://github.com/Pal-cloud/proyecto4_EDA_Pal

Metodología EDA siguiendo mejores prácticas de ciencia de datos