

Universidad Nacional de San Agustín de Arequipa Escuela Profesional de Ciencia de la Computación Curso: Tópicos en Ciencia de Datos



Informe Dashboard:

Docente: Ana Maria Cuadros Valdivia

Informe del Dashboard de Propagación y Olas Pandémicas del COVID-19 Introducción

El propósito de este proyecto es analizar y visualizar los datos de la pandemia de COVID-19 a nivel mundial. El dashboard permite a los usuarios explorar datos sobre casos, muertes, vacunaciones y otros indicadores relevantes, facilitando la comprensión de la propagación y el impacto de la pandemia en diferentes regiones y países.

Proceso de Limpieza de Datos

Descripción de los Pasos

El proceso de limpieza de datos se realizó en dos scripts principales: limpieza.py y limpieza2.py.

1. Selección de Columnas:

- Se seleccionaron columnas relevantes para el análisis, como iso_code, continent, location, date, total_cases, new_cases, total_deaths, new_deaths, population, entre otras.
- Las columnas seleccionadas permiten un análisis completo de la situación de la pandemia, incluyendo datos demográficos y económicos.

2. Manejo de Valores Faltantes:

 Se rellenaron los valores faltantes en las columnas numéricas con ceros para asegurar la integridad de los datos.

3. Eliminación de Duplicados:

Se eliminaron las filas duplicadas para evitar redundancias en el análisis.

4. Cálculo de Nuevas Columnas:

 Se calcularon nuevas columnas como total_cases_per_million y total_deaths_per_million para normalizar los datos por población y facilitar comparaciones entre países.

Scripts de Limpieza

 limpieza.py: Se encarga de la selección inicial de columnas y la creación de un archivo CSV limpio.

```
import pandas as pd

# Lista de columnas a conservar

columnas_conservar = [
    'iso_code', 'continent', 'location', 'date', 'total_cases',
    'new_cases', 'new_cases_smoothed', 'total_deaths', 'new_deaths',
    'new_deaths_smoothed', 'population', 'median_age', 'aged_65_older',
    'aged_70_older', 'gdp_per_capita', 'extreme_poverty',
    'cardiovasc_death_rate', 'diabetes_prevalence', 'total_vaccinations',
    'people_vaccinated', 'people_fully_vaccinated', 'stringency_index'

df = pd.read_csv('owid-covid-data.csv')

df_limpio = df[columnas_conservar]

df_limpio.to_csv('owid-covid-data-limpio.csv', index=False)

df_limpio.to_csv('owid-covid-data-limpio.csv', index=False)
```

 limpieza2.py: Realiza transformaciones adicionales, como la conversión de fechas y el cálculo de nuevas métricas.

```
import pandas as pd

# Paso 1: Cargar los datos desde un archivo CSV

df = pd.read_csv('owid-covid-data-limpio.csv')

# Paso 2: Convertir la columna 'date' a tipo datetime para facilitar el manejo de fechas

df['date'] = pd.to_datetime(df['date'])

# Paso 3: Rellenar valores faltantes

numeric_cols = df.select_dtypes(include=['float64', 'int64']).columns

df[numeric_cols] = df[numeric_cols].fillna(0)

# Paso 4: Verificar y eliminar duplicados

df = df.drop_duplicates()

# Paso 5 Agregar columnas adicionales

df['total_cases_per_million'] = (df['total_cases'] / df['population']) * 1e6

df['total_deaths_per_million'] = (df['total_deaths'] / df['population']) * 1e6

df.to_csv('owid-covid-data-procesado.csv', index=False)

print(df.head())
print(df.info())
```

Descripción del Dashboard

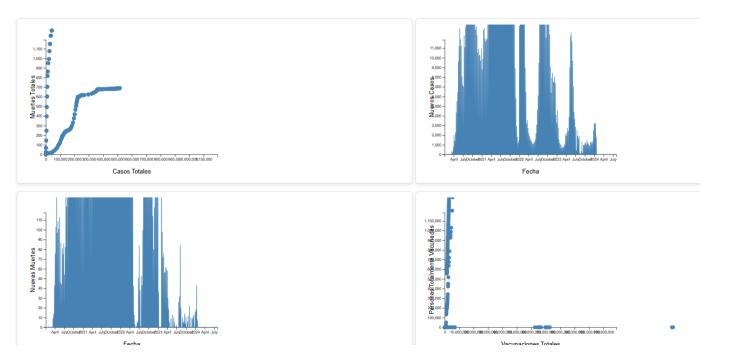
Estructura General

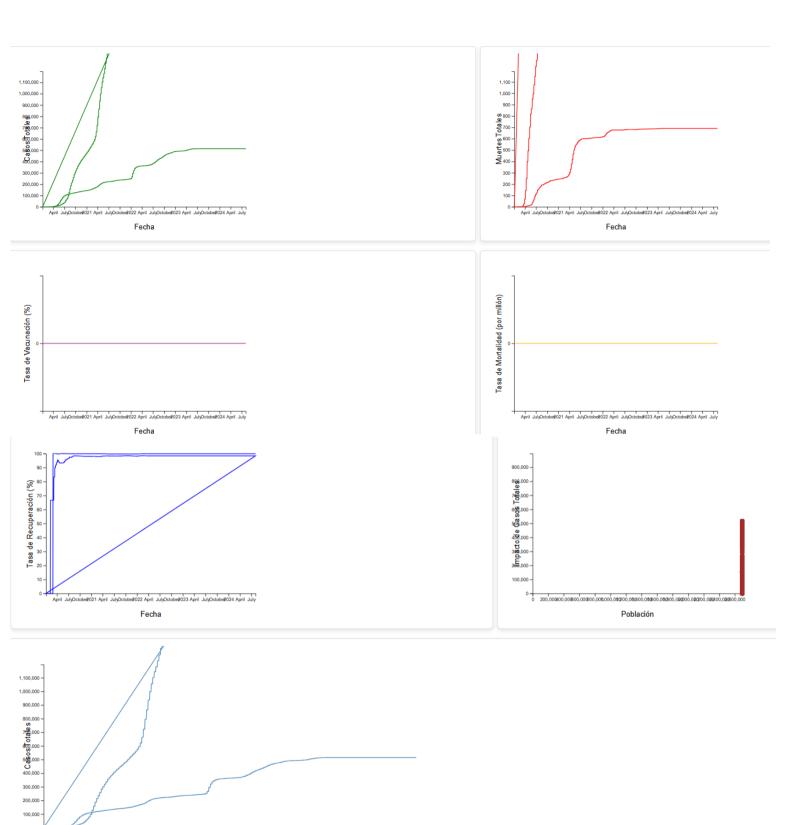
El dashboard está compuesto por una serie de gráficos interactivos que permiten visualizar diferentes aspectos de la pandemia. Los usuarios pueden seleccionar continentes y países específicos para analizar los datos.

Gráficos y Visualizaciones

- Casos vs Muertes: Gráfico de dispersión que muestra la relación entre casos totales y muertes totales.
- **Nuevos Casos y Nuevas Muertes**: Gráficos de líneas que muestran la evolución temporal de nuevos casos y muertes.
- **Vacunaciones**: Gráfico de dispersión que muestra la relación entre vacunaciones totales y personas totalmente vacunadas.
- Casos Totales y Muertes Totales: Gráficos de líneas que muestran la evolución temporal de casos y muertes totales.
- Tasa de Vacunación y Tasa de Mortalidad: Gráficos de líneas que muestran la evolución temporal de estas tasas.
- Tasa de Recuperación: Gráfico de líneas que muestra la evolución temporal de la tasa de recuperación.
- **Impacto en la Población**: Gráfico de dispersión que muestra la relación entre la población y el impacto de casos totales.

Países seleccionados: Philippines, Qatar Enero 2020





Funcionalidades Interactivas

• Selección de Continentes y Países: Los usuarios pueden seleccionar continentes y países específicos para analizar los datos.



• Línea de Tiempo Ajustable: Los usuarios pueden ajustar la línea de tiempo para analizar los datos en diferentes rangos de fechas.

Enero 2022 Abril 2023

Implementación Técnica Tecnologías Utilizadas

 D3.js: Biblioteca de JavaScript para producir visualizaciones de datos dinámicas e interactivas.

```
document.addEventListener('DOMContentLoaded', function () {
         let allData = [];
         let selectedCountries = [];
         let selectedContinent = null;
         d3.csv('owid-covid-data-procesado.csv').then(function (data) {
             allData = data.map(d => {
                 d.date = new Date(d.date);
                 return d;
             const continents = [...new Set(data.map(d => d.continent))].filter(c => c);
             const continentTags = d3.select('#continent-tags');
             continentTags.selectAll('span')
                 .data(continents)
                 .enter()
                 .append('span')
                 .attr('class', 'tag')
.on('click', function (event, d) {
                     const isSelected = !d3.select(this).classed('selected');
                     continentTags.selectAll('span').classed('selected', false);
                      if (isSelected) {
                         d3.select(this).classed('selected', true);
                         selectedContinent = d;
                         updateCountryTags(selectedContinent);
                     } else {
29
                          selectedContinent = null;
                          updateCountryTags(null);
             function updateCountryTags(continent) {
                 const countryTags = d3.select('#country-tags');
                 countryTags.selectAll('*').remove();
                 let countries = continent
                     ? [...new Set(data.filter(d => d.continent === continent).map(d => d.location))]
                 countryTags.selectAll('span')
                     .data(countries)
                     .enter()
                      .append('span')
                      .text(d => d)
```

HTML y CSS: Para la estructura y el estilo del dashboard.

```
<!DOCTYPE html>
<html lang="es">
   <meta charset="UTF-8">
   <meta name="viewport" content="width=device-width, initial-scale=1.0">
   <title>Propagación y Olas Pandémicas del COVID</title>
   <script src="https://d3js.org/d3.v7.min.js"></script>
   <link rel="stylesheet" href="styles.css"</pre>
   <div class="header">
       <h1>Propagación y Olas Pandémicas del COVID</h1>
    <div class="container">
        <div class="sidebar">
               <label>Seleccionar Continente:</label>
                <div id="continent-tags"></div>
                <label>Seleccionar País:
                <div id="country-tags"></div>
        <div class="content">
            <div class="selection-box-container">
                <div id="selected-countries-box">
                    <span>Países seleccionados: </span>
                    <span id="selected-countries">Ninguno</span>
                <button id="reset-button">Restablecer</button>
            <div id="timeline-container">
               <div class="timeline">
                   <div class="horizontal-bar"></div>
                   <div class="vertical-bar" id="start-bar"></div>
                   <div class="vertical-bar" id="end-bar"></div>
                   <div class="date-label" id="start-label">Enero 2020</div>
                    <div class="date-label" id="end-label">Enero 2025</div>
            <div id="charts-container">
               <div id="cases-vs-deaths-chart" class="chart"></div>
               <div id="new-cases-chart" class="chart"></div>
               <div id="new-deaths-chart" class="chart"></div>
               <div id="vaccinations-chart" class="chart"></div>
                <div id="total-cases-chart" class="chart"></div>
```

Carga y Visualización de Datos

Los datos se cargan desde un archivo CSV procesado y se visualizan utilizando D3.js. Las funciones principales en el script se encargan de crear y actualizar los gráficos basados en los datos seleccionados.

Funciones Principales

- updateCharts(data): Actualiza todos los gráficos basados en los datos proporcionados.
- createCasesVsDeathsChart(data): Crea el gráfico de dispersión de casos vs muertes.
- createNewCasesChart(data): Crea el gráfico de líneas de nuevos casos.
- createNewDeathsChart(data): Crea el gráfico de líneas de nuevas muertes.
- createVaccinationsChart(data): Crea el gráfico de dispersión de vacunaciones.
- createTotalCasesChart(data): Crea el gráfico de líneas de casos totales.
- createTotalDeathsChart(data): Crea el gráfico de líneas de muertes totales.
- createVaccinationRateChart(data): Crea el gráfico de líneas de la tasa de vacunación.
- createDeathRateChart(data): Crea el gráfico de líneas de la tasa de mortalidad.
- createRecoveryRateChart(data): Crea el gráfico de líneas de la tasa de recuperación.
- createPopulationImpactChart(data): Crea el gráfico de dispersión del impacto en la población.
- createPropagationChart(data): Crea el gráfico de líneas de la propagación.

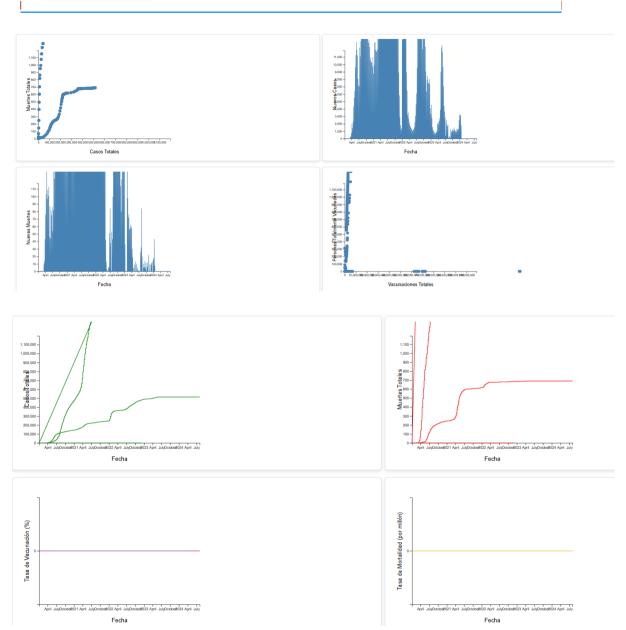
Conclusiones y Recomendaciones

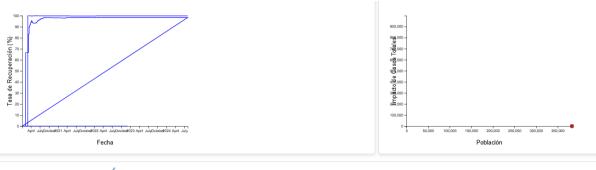
Resumen de Hallazgos

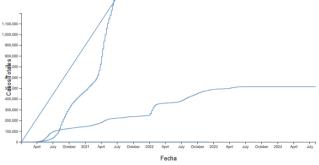
El dashboard proporciona una visión clara y detallada de la propagación y el impacto de la pandemia de COVID-19. Las visualizaciones permiten identificar patrones y tendencias en los datos, facilitando la toma de decisiones informadas

Resultados con los países:

Países seleccionados: Northern Cyprus, Philippines, Qatar







Posibles Mejoras y Extensiones Futuras

- **Inclusión de Más Datos**: Incorporar más indicadores y datos demográficos para un análisis más completo.
- **Mejoras en la Interactividad**: Añadir más funcionalidades interactivas, como la selección de múltiples países y continentes.
- Optimización del Rendimiento: Mejorar el rendimiento del dashboard para manejar grandes volúmenes de datos de manera eficiente.