

# Análisis Exploratorio Caso Practico Tuberculosis

El análisis exploratorio se realizó en Python en el cual realizamos los siguientes gráficos para interpretar los datos:

Código Python:

## 1. Distribución por Genero

Código Python

```
by_gender = df.groupby("gender", as_index=False)["cases"].sum()  
by_gender["percent_of_total"] = 100 * by_gender["cases"] / by_gender["cases"].sum()
```

```
plt.figure(figsize=(8, 6))  
sns.barplot(data=by_gender, x="gender", y="percent_of_total")  
plt.title("Distribución de casos por género")  
plt.xlabel("Género")  
plt.ylabel("Porcentaje del total")  
plt.show()
```

## 2. Distribución por Grupo de Edad

Código Python

```
by_agegroup = df.groupby("agegroup", as_index=False)["cases"].sum()  
by_agegroup["percent_of_total"] = 100 * by_agegroup["cases"] / by_agegroup["cases"].sum()  
  
plt.figure(figsize=(10, 6))  
sns.barplot(data=by_agegroup, x="agegroup", y="percent_of_total", order=["0-14", "15-24", "25-34", "35-44", "45-54", "55-64", "65+"])  
plt.title("Distribución de casos por grupo de edad")  
plt.xlabel("Grupo de edad")  
plt.ylabel("Porcentaje del total")  
plt.show()
```

### 3. Tasa de Incidencia Tuberculosis

Código Python

```
plt.figure(figsize=(12, 6))
sns.lineplot(data=by_year, x="year", y="incidencia")
plt.title("Tasa de incidencia de tuberculosis (1995-2013)")
plt.xlabel("Año")
plt.ylabel("Casos por 100,000 habitantes")
plt.grid(True)
plt.show()
```

### 4. Países con mayor carga de Tuberculosis

Código Python

```
top_countries = df.groupby("country", as_index=False)["cases"].sum().sort_values("cases", ascending=False).head(10)

plt.figure(figsize=(12, 6))
sns.barplot(data=top_countries, x="cases", y="country")
plt.title("Top 10 países con mayor número de casos (1995-2013)")
plt.xlabel("Número de casos")
plt.ylabel("País")
plt.show()
```

### 5. Tasa de Incidencia Global por Año

Código Python

```
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
```

# Configuración de estilo para todos los gráficos

```
sns.set_style("whitegrid")
plt.rcParams['figure.dpi'] = 120
plt.rcParams['font.size'] = 10
```

# 1. Gráfico de Tasa de Incidencia Global por Año (Matplotlib/Seaborn)

```
plt.figure(figsize=(12, 6))
```

```
ax = sns.lineplot(data=by_year, x='year', y='incidencia',  
                  marker='o', linewidth=2.5, color='#1f77b4')
```

# Personalización

```
plt.title('Evolución de la Tasa de Incidencia Global de Tuberculosis (1995-2013)\nCasos por 100,000 habitantes',  
         pad=20, fontsize=12, fontweight='bold')
```

```
plt.xlabel('Año', labelpad=10)
```

```
plt.ylabel('Tasa de Incidencia', labelpad=10)
```

```
plt.xticks(by_year['year'][:,::2]) # Mostrar años alternados para mejor legibilidad
```

# Destacar puntos clave

```
max_incidencia = by_year.loc[by_year['incidencia'].idxmax()]
```

```
ax.annotate(f'Pico: {max_incidencia["year"]}\n({max_incidencia["incidencia"]:.1f} casos)',
```

```
          xy=(max_incidencia['year'], max_incidencia['incidencia']),
```

```
          xytext=(10, 30), textcoords='offset points',
```

```
          arrowprops=dict(arrowstyle="->", color='red'),
```

```
          bbox=dict(boxstyle="round", fc="w"))
```

# Guardar gráfico

```
plt.tight_layout()
```

```
plt.savefig('tasa_incidencia_global.png', bbox_inches='tight', dpi=300)
```

```
plt.show()
```

## 6. Casos por método de diagnóstico

Código Python

```
by_method_sorted = by_method.sort_values('percent_of_total', ascending=False)
```

```
fig = px.bar(by_method_sorted,
```

```
             x='method',
```

```
y='percent_of_total',
color='method',
color_discrete_sequence=px.colors.qualitative.Pastel,
text='percent_of_total',
title='<b>Distribución de Casos por Método de Diagnóstico</b><br><sup>Porcentaje del total de casos</sup>',
labels={'percent_of_total': 'Porcentaje (%)', 'method': 'Método de Diagnóstico'})
```

```
# Personalización
```

```
fig.update_traces(texttemplate='%{text:.1f}%', textposition='outside')
```

```
fig.update_layout(
    hovermode='x',
    showlegend=False,
    title_x=0.5,
    xaxis_tickangle=-45,
    yaxis_range=[0, by_method_sorted['percent_of_total'].max() + 10],
    plot_bgcolor='rgba(0,0,0,0)'
)
```

```
# Añadir anotación con insights
```

```
fig.add_annotation(
    x=0.5, y=-0.3,
    xref='paper', yref='paper',
    text="El diagnóstico por esputo pulmonar positivo representa más de la mitad de los casos identificados",
    showarrow=False,
    font=dict(size=10, color='grey')
)
```

```
# Guardar gráfico interactivo
```

```
fig.write_html('casos_por_metodo.html')
fig.show()
```

```
# Versión estática para PDF
```

```
plt.figure(figsize=(10, 6))
```

```

sns.barplot(data=by_method_sorted, x='method', y='percent_of_total',
            palette='pastel', edgecolor='black')
plt.title('Distribución de Casos por Método de Diagnóstico', pad=15)
plt.xlabel('Método de Diagnóstico')
plt.ylabel('Porcentaje del Total (%)')
plt.xticks(rotation=45)
plt.tight_layout()
plt.savefig('casos_por_metodo.png', dpi=300)
plt.show()

```

## Graficas:



