

DATOS MUNDIALES



SQL + PYTHON

Mariana Carmona



DESCRIPCIÓN DEL PROBLEMA

- El problema que se busca resolver en este proyecto es la falta de integración y análisis de datos globales que puedan ser utilizados para obtener insights significativos en áreas como la demografía, la economía y la diversidad lingüística.
- La pregunta de investigación principal es cómo la información estructurada sobre ciudades, países e idiomas puede ser utilizada para extraer conclusiones valiosas que ayuden en la toma de decisiones informadas.
- El contexto relevante de este problema radica en la importancia de tener datos precisos y actualizados para entender mejor las dinámicas globales. Este tipo de análisis puede tener un impacto significativo en diversas áreas, como la planificación urbana, la formulación de políticas económicas y el entendimiento de la diversidad cultural.

OBJETIVOS DEL PROYECTO

1. Primer objetivo: Crear una base de datos que contenga información relevante sobre ciudades, países e idiomas. Esto incluirá la definición de los esquemas de las tablas y la importación de datos en un entorno de nube (Azure SQL Server) y un entorno local (MySQL Workbench)
 2. Segundo objetivo: Establecer una conexión eficiente con la base de datos utilizando Python y realizar consultas SQL que respondan a preguntas específicas sobre los datos. Esto incluirá la utilización de bibliotecas como `sqlalchemy` y `mysql-connector-python` para interactuar con las bases de datos, así como el uso de DataFrames de Pandas para manipular y analizar los datos obtenidos.
 3. Tercer objetivo: Crear visualizaciones efectivas de los datos utilizando bibliotecas como `matplotlib` y `seaborn`.
- La resolución de este problema puede beneficiar tanto a organizaciones o comunidades al proporcionar una herramienta valiosa para el análisis de datos globales. Esto permitirá tomar decisiones informadas basadas en datos, mejorando la planificación y la formulación de políticas en diversas áreas.

METODOLOGÍA DEL PROYECTO



1 - SQL

- Creación de la base de datos
- Creación de tablas
- Importar los datos

Azure
SQL Server

MySQL
Workbench



2 - Python

- Hacer conexión a la base de datos
- Realizar consultas desde Python para contestar preguntas



3 - Visualizaciones

- Hacer conexión a la base de datos
- Crear un visualizaciones en Python que responda las mismas preguntas.

DATOS UTILIZADOS

country

Columna	Tipo	Descripción
Code	char(3)	Código único del país.
Name	char(52)	Nombre del país.
Continent	enum	Continente al que pertenece el país (Asia, Europe, North America, etc.).
Region	char(26)	Región dentro del continente.
SurfaceArea	decimal(10,2)	Área de superficie del país.
IndepYear	smallint	Año de independencia.
Population	int	Población del país.
LifeExpectancy	decimal(3,1)	Esperanza de vida.
GNP	decimal(10,2)	Producto Nacional Bruto.
GNPOld	decimal(10,2)	Producto Nacional Bruto anterior.
LocalName	char(45)	Nombre local del país.
GovernmentForm	char(45)	Forma de gobierno.
HeadOfState	char(60)	Jefe de estado.
Capital	int	Identificador de la capital (relación con la tabla `city`).
Code2	char(2)	Código adicional del país.

countrylanguage

Columna	Tipo	Descripción
CountryCode	char(3)	Código del país (relación con la tabla `country`).
Language	char(30)	Idioma hablado en el país.
IsOfficial	enum	Indica si es un idioma oficial en el país ('T' para true, 'F' para false).
Percentage	decimal(4,1)	Porcentaje de personas que hablan el idioma en el país.

city

Columna	Tipo	Descripción
ID	int	Identificador único de la ciudad.
Name	char(35)	Nombre de la ciudad.
CountryCode	char(3)	Código del país (de la tabla `country`).
District	char(20)	Distrito o región a la que pertenece la ciudad.
Population	int	Población de la ciudad.



1.- SQL

AZURE SQL SERVER

- Creación del grupo de recursos
- Creación de una instancia de SQL Server
- Creación de la base de datos

ConexionAzure.ipynb

- Conexión a la base de datos creada en Azure
- Creación de las tablas
- Cargar datos

SQLServerAzure_CargaDatos.ipynb

MYSQL WORKBENCH

Se hizo uso de los scripts:

- world_city.sql
- world_country.sql
- world_countrylanguage.sql

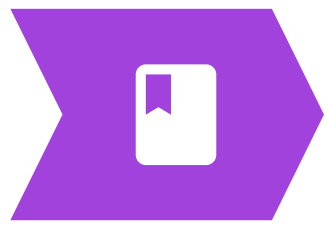
para importar la base de datos World en MySQL.

Se utilizó el asistente de importación de Workbench



2.- PYTHON

- Hacer conexión a la base de datos:
 - SQLAlchemy para Azure SQL Server - [SQLServer_Preguntas.ipynb](#)
 - MySQL Connector para MySQL - [MySQL_Preguntas.ipynb / MySQL_Preguntas_Pandas.ipynb](#)
- Responder consultas utilizando código SQL desde Python
 - **Mostrar el nombre y la población de todos los países del continente europeo.**
 - **Mostrar los nombres y las áreas de superficie de los cinco países más grandes del mundo (en términos de área de superficie).**
 - **Calcular la población total de todos los países de cada continente y mostrar el resultado junto con el nombre del continente.**
 - **Mostrar el nombre de las ciudades y la población de todos los países de Europa, ordenados por población de la ciudad de manera descendente.**
 - **Actualiza la población de China (código de país 'CHN') a 1500000000 (1.5 mil millones).**



2.- PYTHON

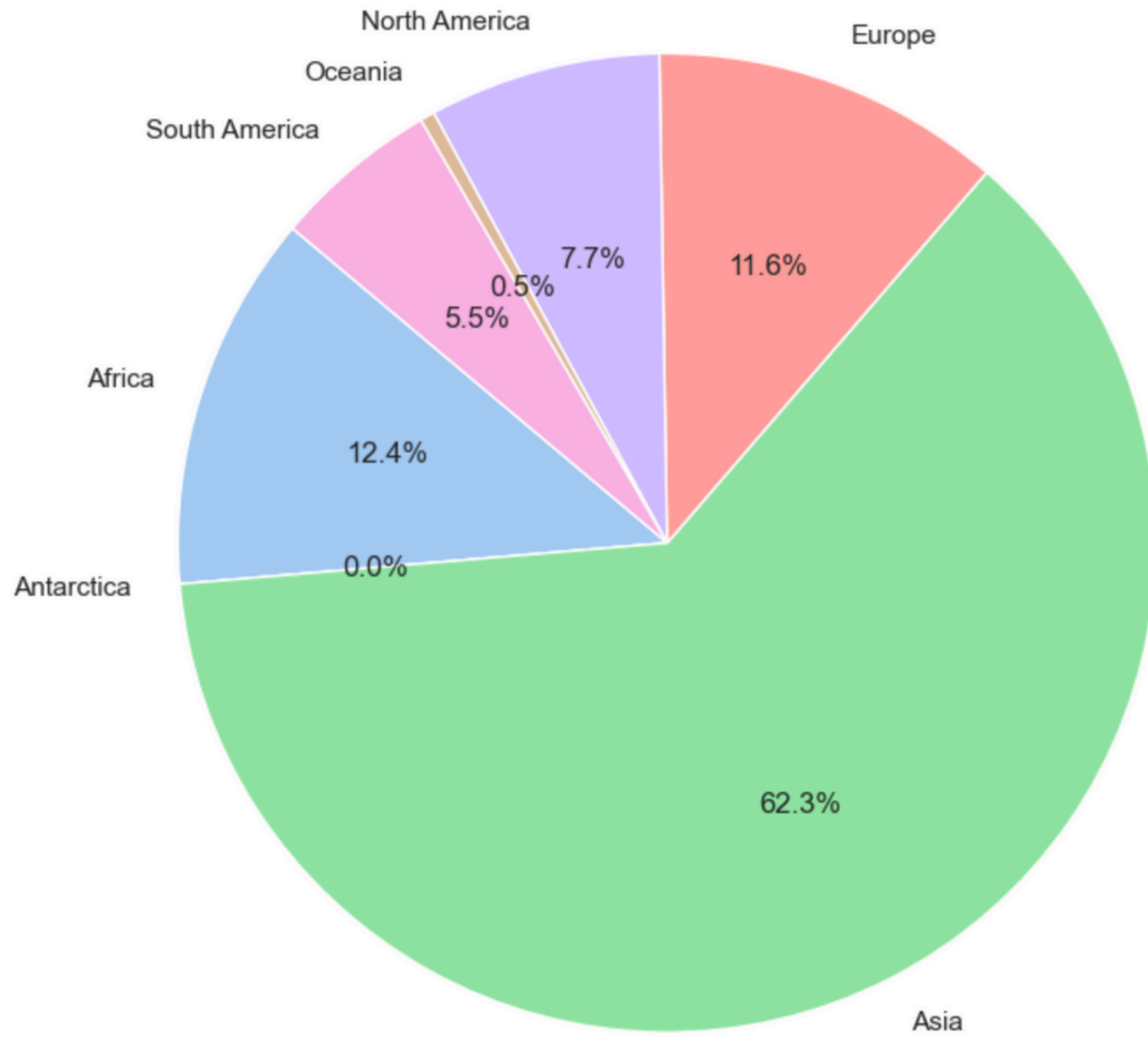
- Responder consultas utilizando código SQL desde Python
 - ¿Cuáles son los países con más ciudades en la base de datos?
 - ¿Qué país tiene la mayor diversidad lingüística en términos de cantidad de idiomas hablados?
 - ¿Cuáles son las 10 ciudades más pobladas del mundo?
 - ¿Qué países tienen una expectativa de vida mayor a 80 años?
 - ¿Cuáles son los continentes con la mayor cantidad de países?
 - ¿Qué continentes tienen la mayor y menor densidad poblacional, y cuál es esa densidad?
 - ¿Cuál son los 3 idiomas más hablados en el mundo (considerando solo los datos disponibles)?
 - ¿Cuáles son los 3 países que tienen el mayor Producto Interno Bruto (PIB)?
 - ¿Cuál es la proporción de habitantes que viven en la capital de cada país respecto a su población total?
 - ¿Cuáles son los países con la mayor superficie terrestre en cada continente?



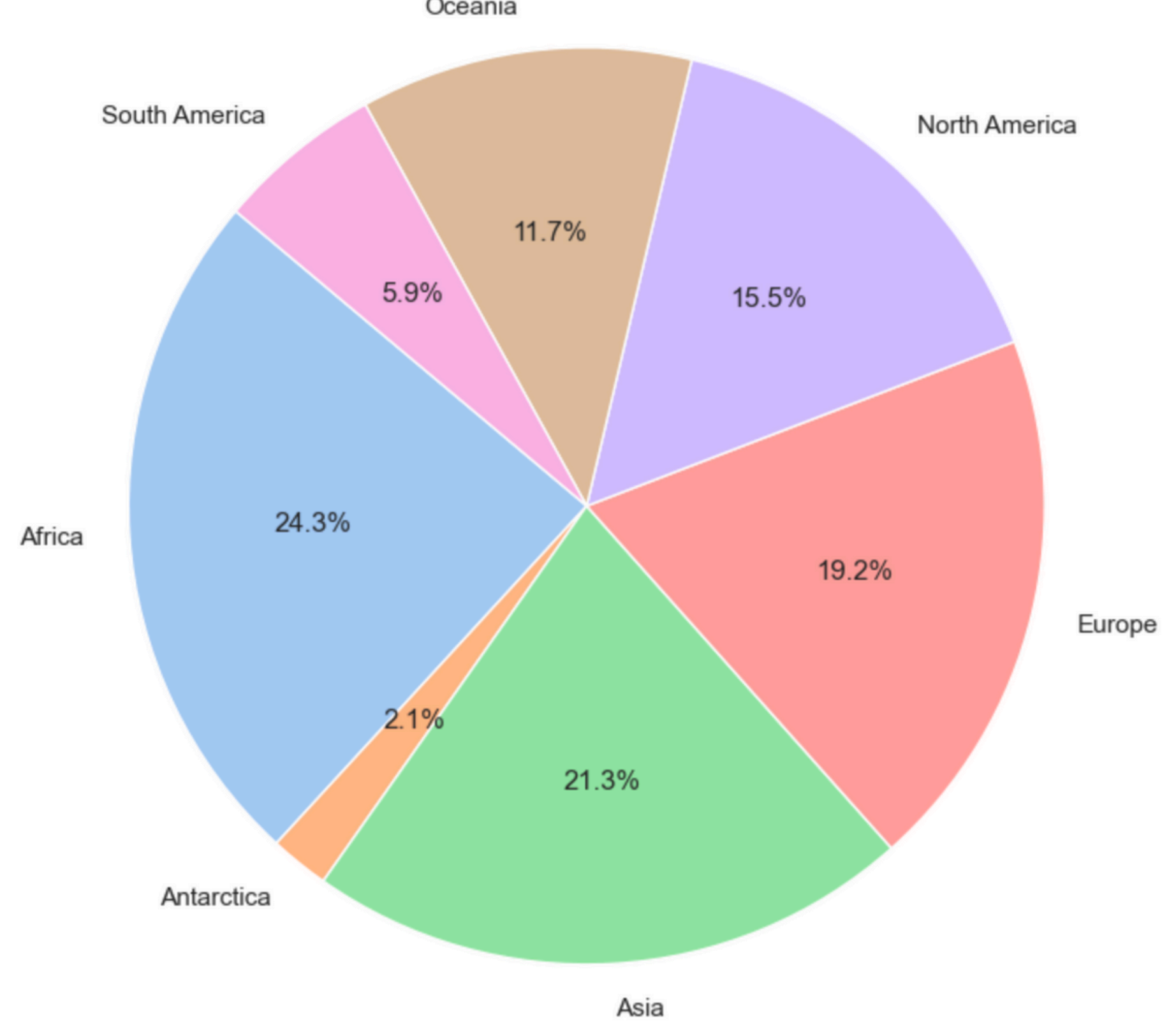
3.- VISUALIZACIONES / PYTHON

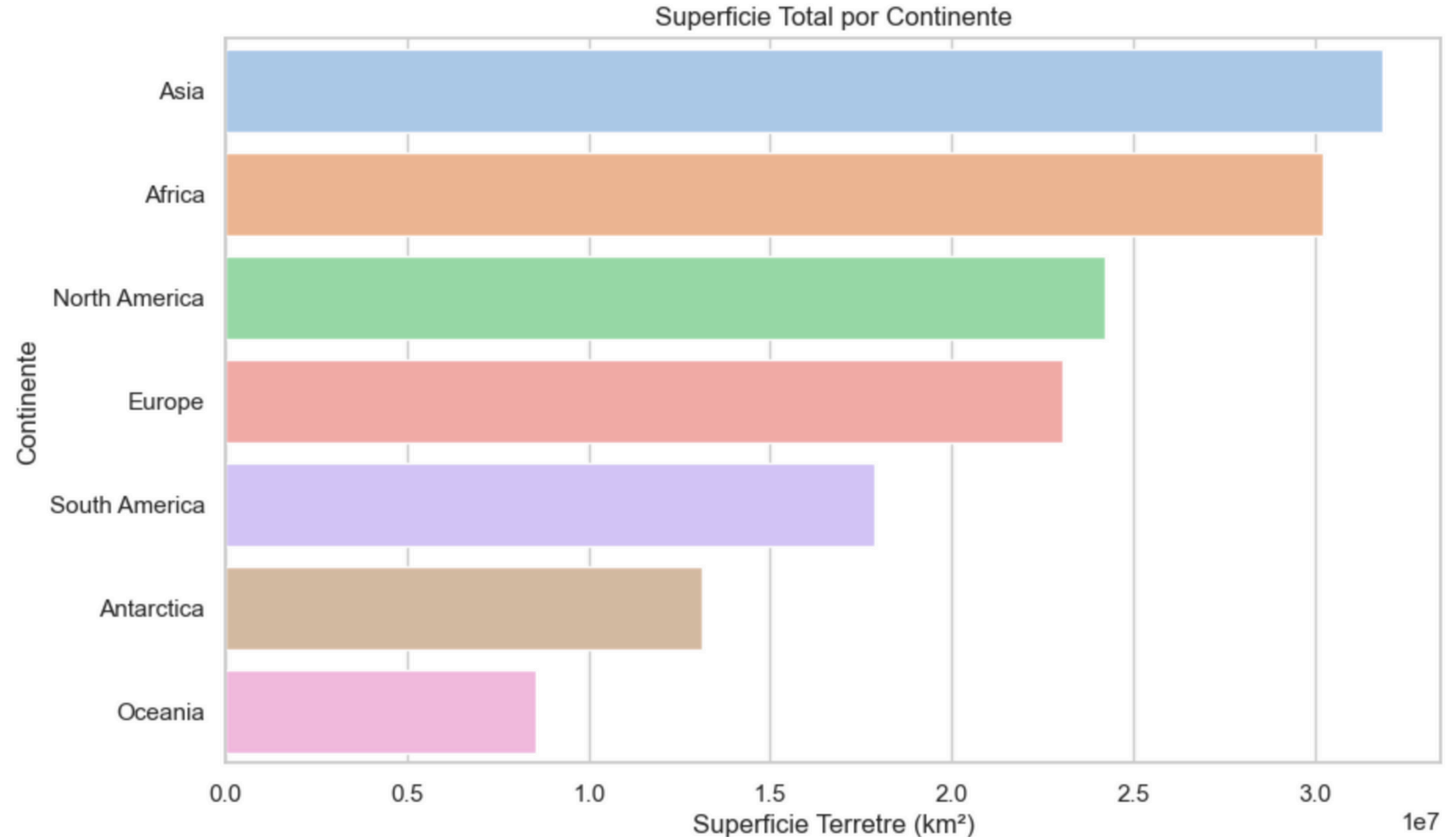
- Hacer conexión a la base de datos
- Hacer visualizaciones de la preguntas
MySQL_Visualizaciones.ipynb

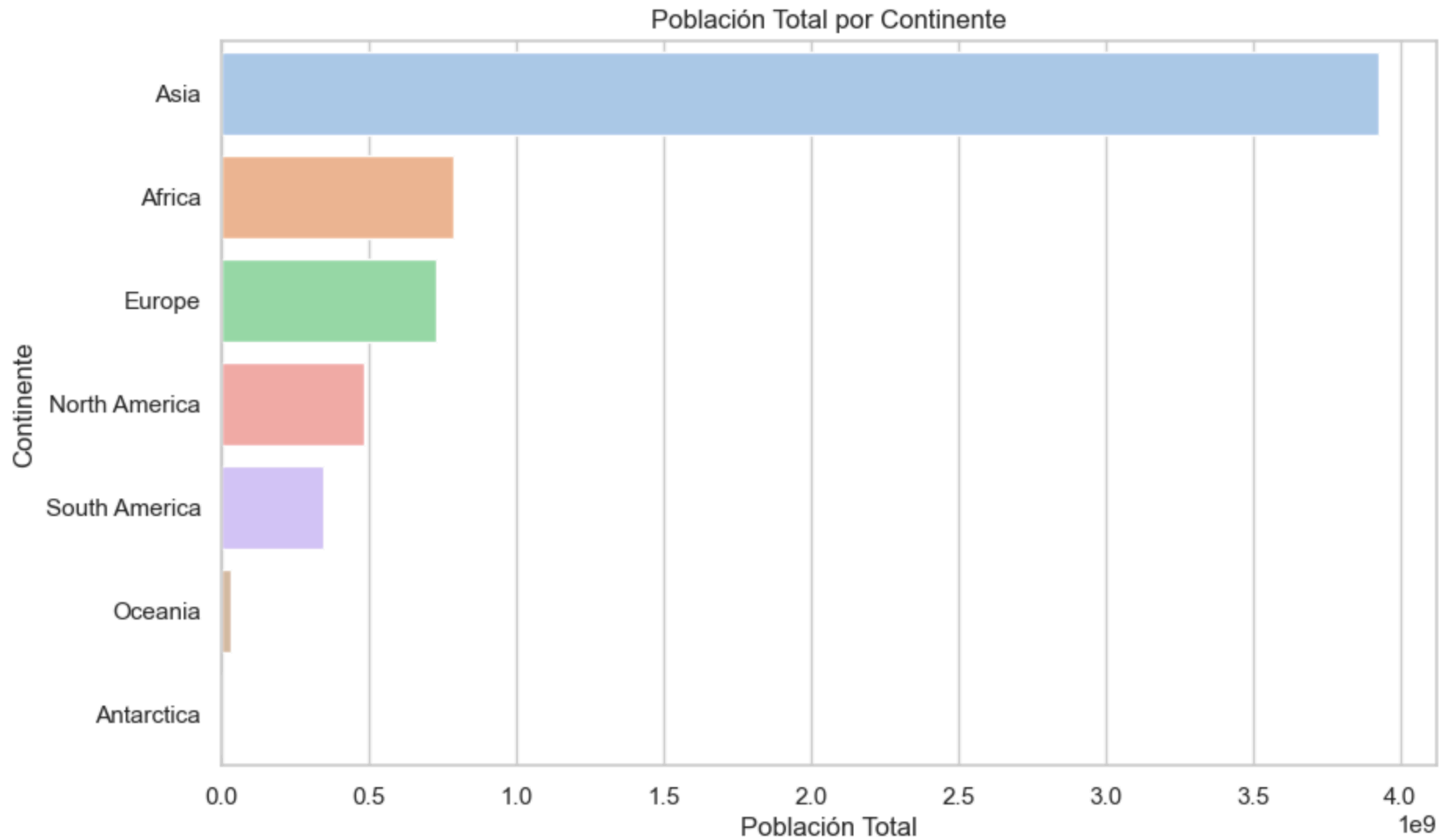
Distribución de la Población Total por Continente

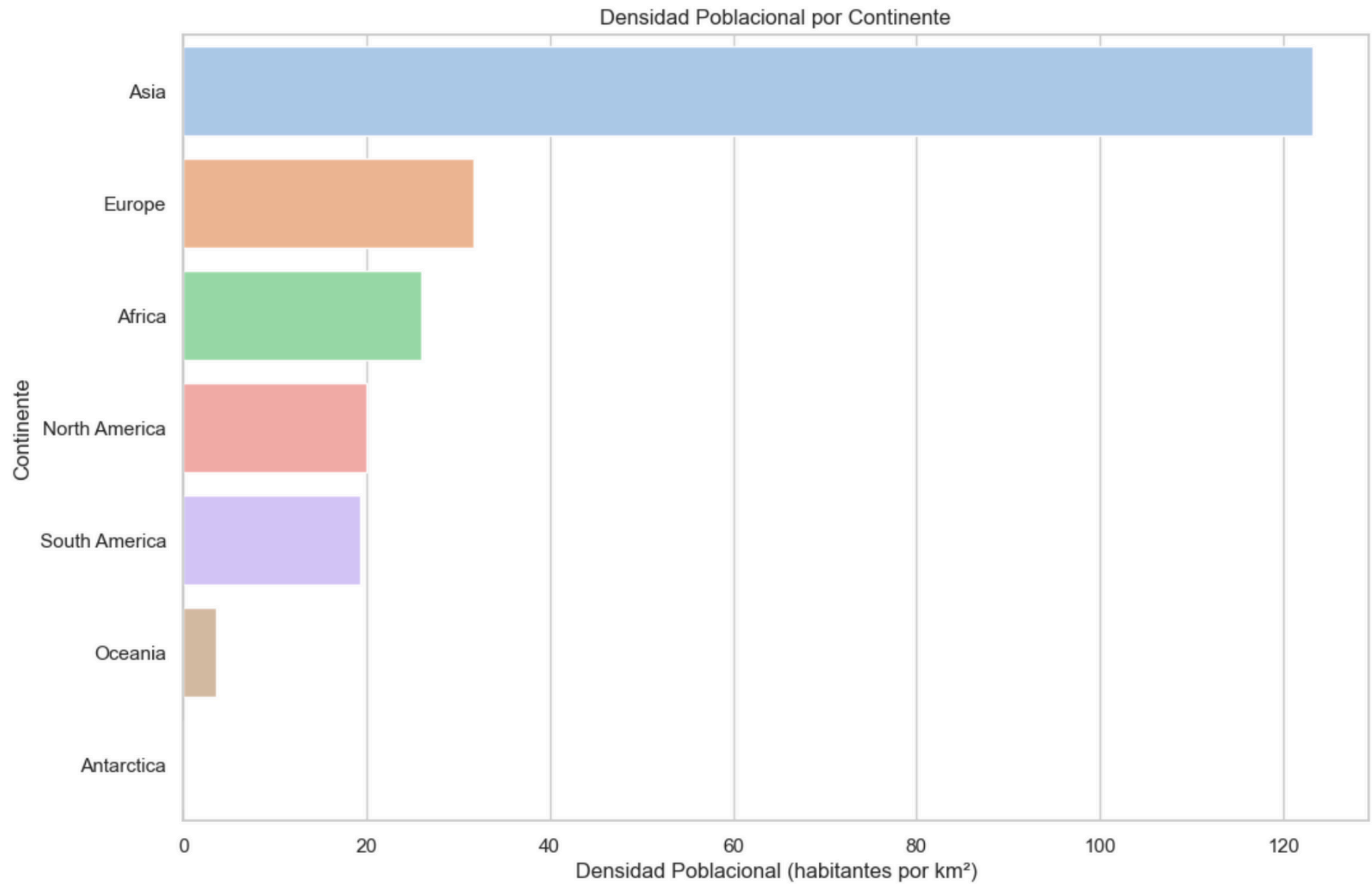


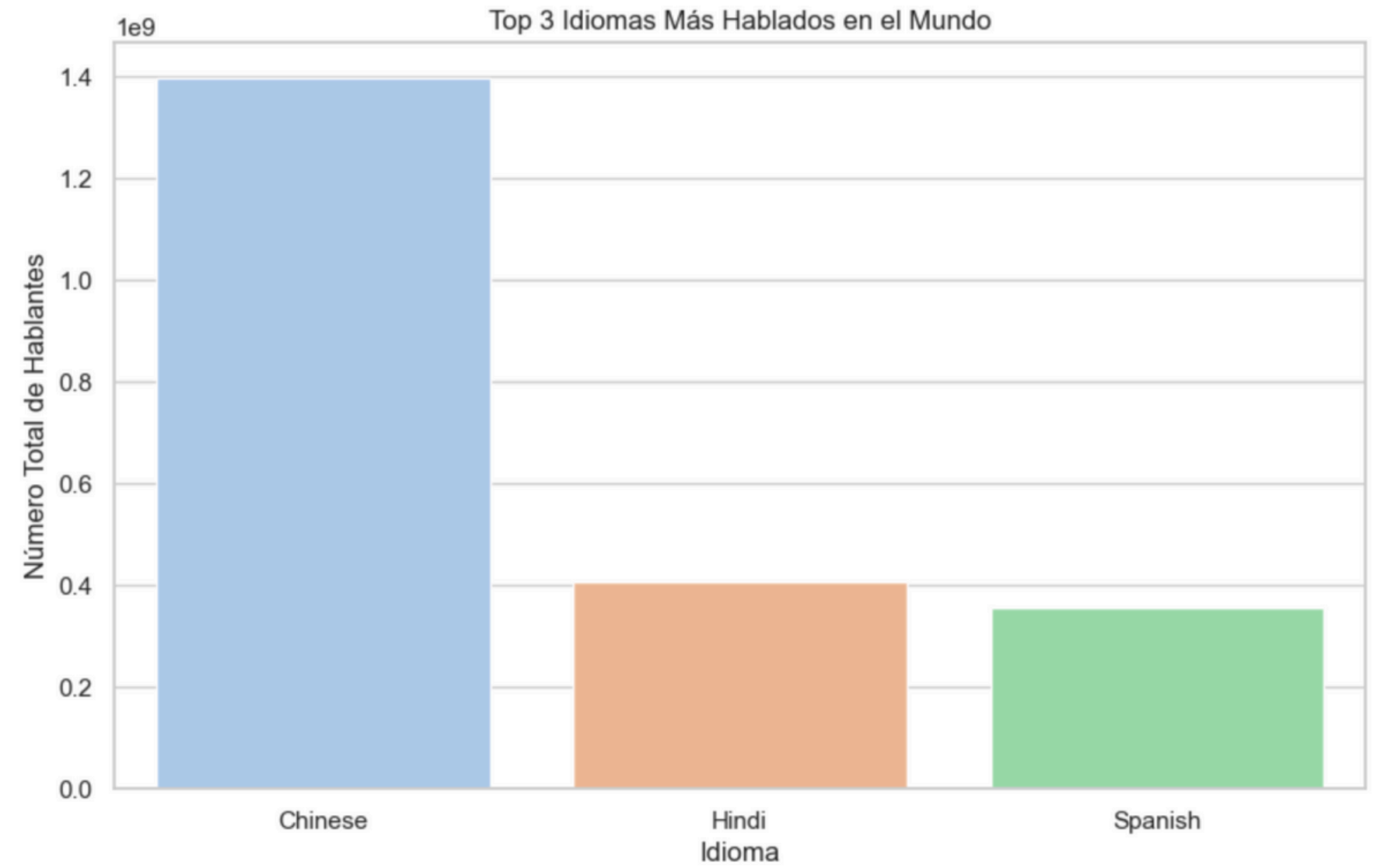
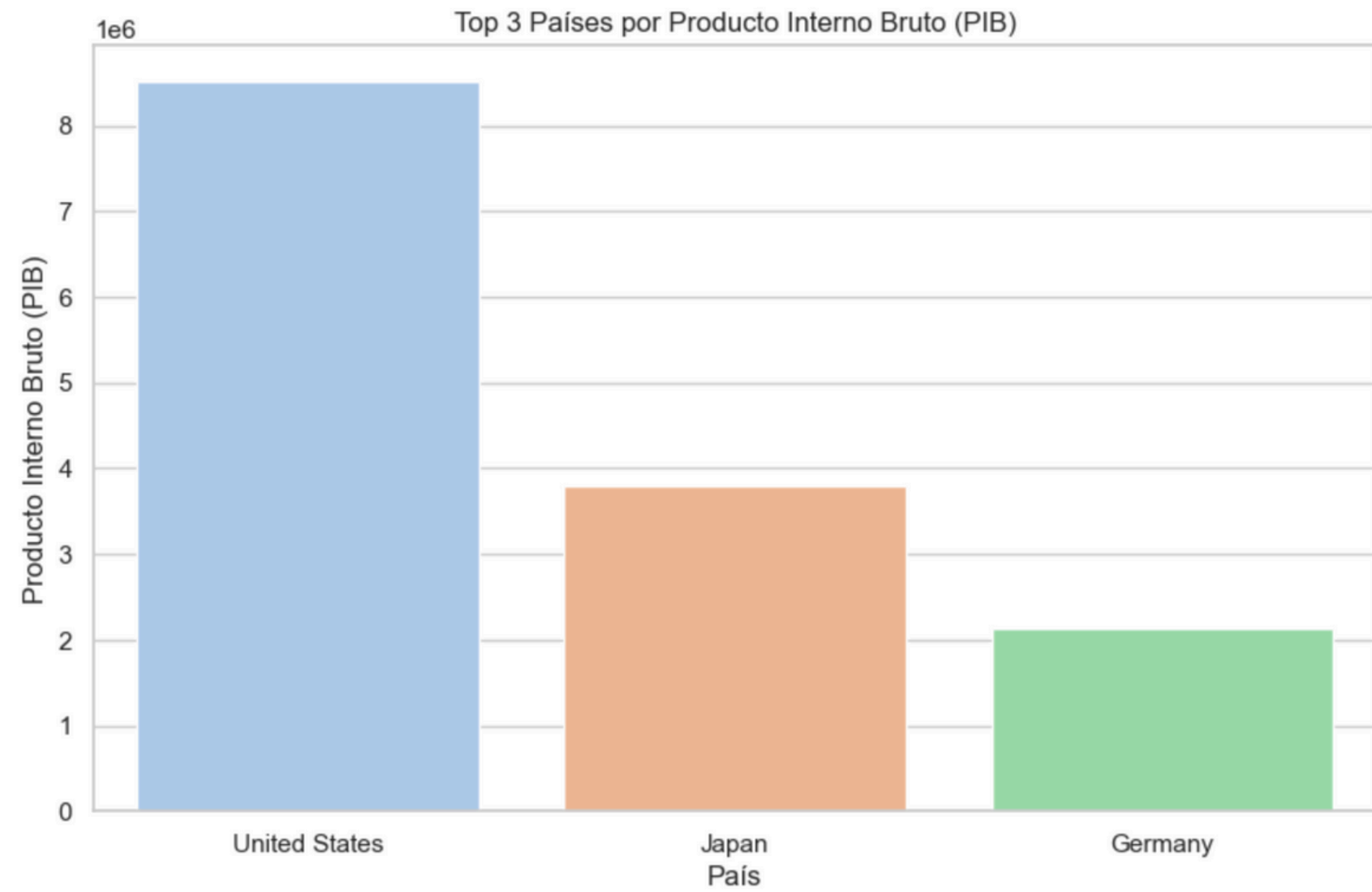
Número de Países por Continente











	CountryName	CapitalName	CapitalPopulation	CountryPopulation	CapitalPopulationPercentage
0	Singapore	Singapore	4017733	3567000	112.6362
1	Gibraltar	Gibraltar	27025	25000	108.1000
2	Macao	Macao	437500	473000	92.4947
3	Pitcairn	Adamstown	42	50	84.0000
4	Saint Pierre and Miquelon	Saint-Pierre	5808	7000	82.9714
...
227	China	Peking	7472000	1500000000	0.4981
228	Pakistan	Islamabad	524500	156483000	0.3352
229	Nigeria	Abuja	350100	111506000	0.3140
230	United States	Washington	572059	278357000	0.2055
231	India	New Delhi	301297	1013662000	0.0297

	CountryName	CapitalName	CapitalPopulation	CountryPopulation	CapitalPopulationPercentage
0	Singapore	Singapore	4017733	3567000	112.6362
1	Gibraltar	Gibraltar	27025	25000	108.1000

CONCLUSIONES

- La distribución de la población por continente mostró que Asia es el continente con la mayor población, seguido por África y Europa, lo que refleja la gran densidad de población en regiones como Asia que alberga a países como China e India.
- Al listar las ciudades europeas por su población en orden descendente, se pudo observar que las ciudades más grandes, Moscow, London, St. Petersburg, Berlin, Madrid, son las que concentran la mayor cantidad de habitantes, lo que es crucial para entender la distribución urbana en Europa
- Los países con más ciudades en la base de datos fueron identificados, siendo China, India, Estados Unidos los más destacados, lo cual puede ser indicativo de la extensa urbanización y el desarrollo de infraestructuras en estos países
- Los países con una expectativa de vida mayor a 80 años, como Andorra y Japón, fueron destacados, y esto puede estar relacionado con altos estándares de vida, buenos sistemas de salud y calidad de vida en general
- Se determinó que Asia tiene la mayor densidad poblacional, mientras que Antártida/Oceanía tiene la menor, y esta información es crucial para estudios sobre recursos, urbanización y planificación regional
- Los tres países con el mayor PIB son Estados Unidos, Japón y Alemania, lo que es significativo para análisis económicos y estrategias de mercado

APLICACIONES Y FUTURAS DIRECCIONES

- Estos datos pueden ser utilizados por gobiernos y organismos internacionales para la planificación urbana y el desarrollo de infraestructuras.
- Las empresas pueden emplear los hallazgos económicos, como el análisis del PIB y la densidad poblacional, para tomar decisiones informadas sobre inversiones y estrategias de mercado.
- Los datos sobre diversidad lingüística pueden ser útiles para organizaciones educativas y culturales que buscan promover el aprendizaje de idiomas y la preservación de herencias culturales.
- Para futuras investigaciones, se sugiere ampliar la base de datos para incluir más variables.
- También sería beneficioso implementar técnicas avanzadas de análisis de datos, como el aprendizaje automático y la inteligencia artificial,