Fuentes de datos:

Se obtuvo una base de datos públicos construida sobre datos del ranking internacional de Forbes a través de [Kaggle](https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://www.kaggle.com/&ved=2ahUKEwiCv6yn4_mGAxUpVPEDHZUeCEQQFnoECCUQAQ&usg=AOvVaw358aJVdRF5ENauJCrosrX1) y dos bases de datos de 2010 y 2020 para hacer una comparación general.

La base de datos incluye datos exclusivamente relativos al ranking Forbes (puesto, valor total, nombre de la fortuna), datos personales (edad, sexo, fecha de nacimiento) e información geopolítica (país, datos económicos y sociales del país, etc.) que nos permitirán crear un contexto socio económico alrededor de cada fortuna.

Hipótesis**:**

Originalmente se pretendía demostrar que la distribución de edad de los millonarios estaba decreciendo, pero al tomar como referencia los promedios de edad de 2010 (61,57 años) y 2020 (64,18 años) se observó que la tendencia era en aumento. Por lo que comprobaremos la continuidad de esta tendencia y analizaremos factores que nos permitan una visión más amplia ya que esta base es más completa que las ediciones anteriores.

Se pretende desarrollar un análisis demográfico, geográfico y comercial de los billonarios.

Demostrando lo siguiente:

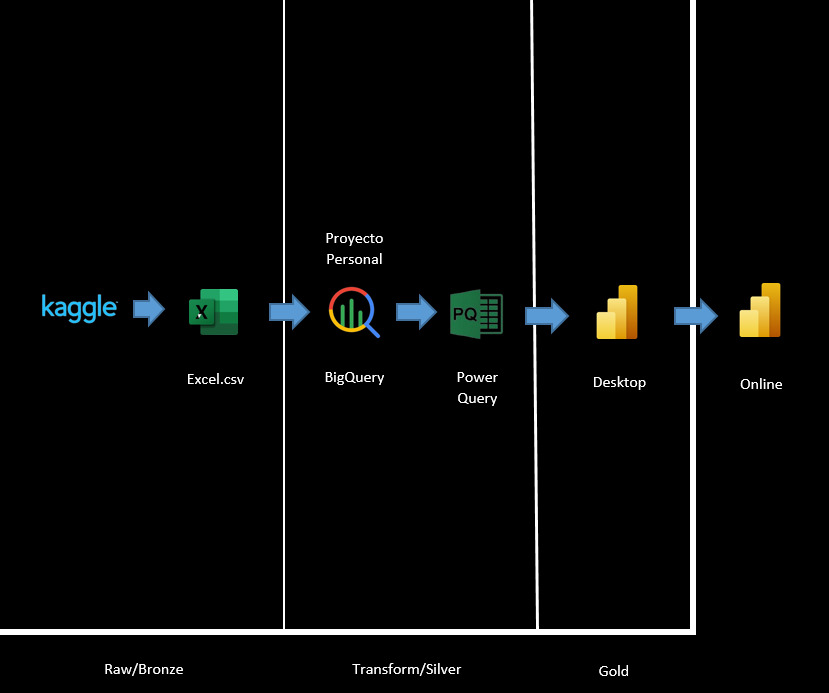
A- La edad promedio de los millonarios continúa en aumento.

B- Hay una mayoría de hombres por sobre las mujeres en el ranking, si se comprueba estudiar la distribución por edad y el coeficiente selfmade.

C- La distribución geografía de los millonarios se centra en los países con mayor nivel de desarrollo

D- Tecnología y Finanzas son las industrias con más millonarios

**Preparación del dataset:**

**Flujo de datos:**  
  


Antes de prepararlo subí una versión del Raw a BigQuery (capa bronce) para tenerla de respaldo.

Para preparar los datos para consumo realice los siguientes pasos.   
  
1- Hice una carga previa en PowerQuery para visualizar la calidad de los datos y detecte que había datos faltantes que eran accesibles, valiéndome información pública de algunos países que no se encontraba correctamente completa y era fácilmente accesible.

También detecte que había mas registros únicos de los esperados en el campo country, por lo que al mirar en detalle encontré que territorios no independientes del reino unido (Islas vírgenes, Guernesey, etc.) que al carecer de datos acerca de las variables económicas territoriales decidí unificar a Reino unido.

Complete con Excel la información faltante en el dataset y cambie los valores antes mencionados afectando a los siguientes campos:

* country
* cpi\_country
* cpi\_change\_country
* gdp\_country
* gross\_tertiary\_education\_enrollment
* gross\_primary\_education\_enrollment\_country
* life\_expectancy\_country
* tax\_revenue\_country\_country
* total\_tax\_rate\_country

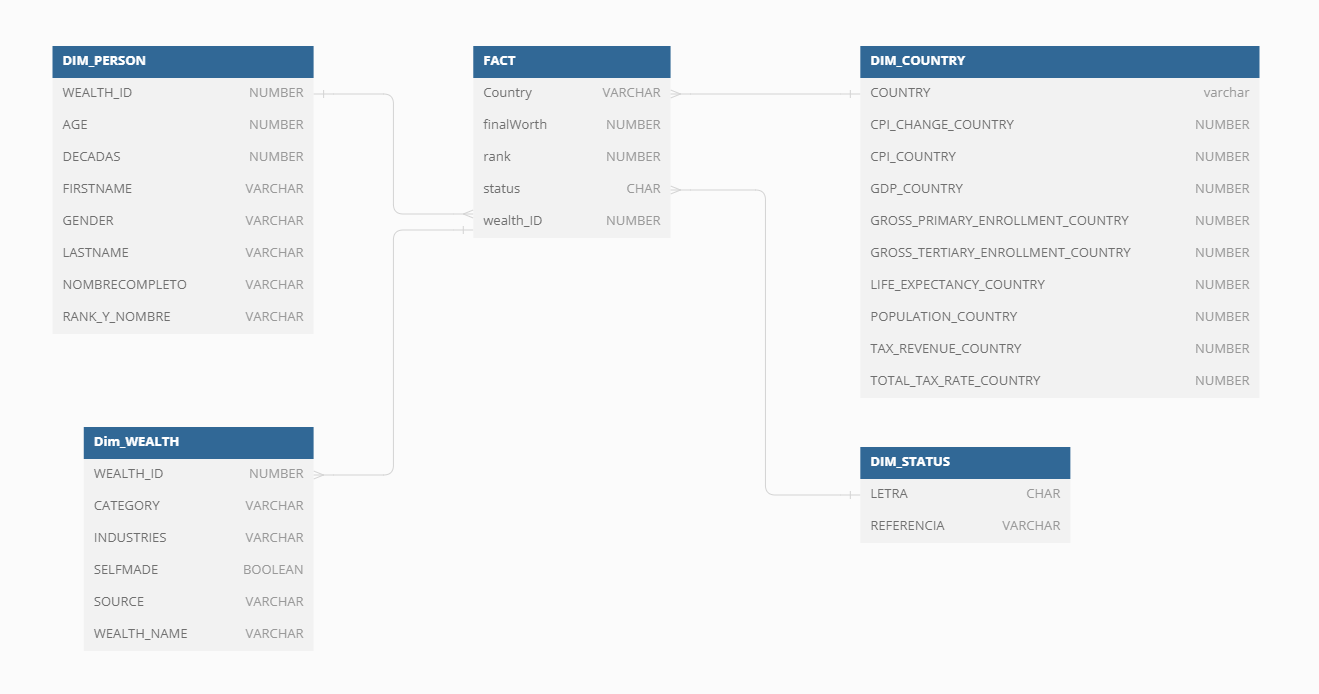
2 – Cargue del csv actualizado a BigQuery.

3 – Limpie campos que no se utilizaran y cree una tabla adicional de referencias para el campo Status.

* Date (fecha del ranking) era el mismo para todos.
* State, solo estaba completo para USA.
* Residence State reg, solo estaba completo para USA.
* Birth Date, no es relevante.
* Birth Month, no es relevante.
* Birth Day, no es relevante.
* Birth Year
* Latitude country, no lo utilizaremos.
* Longitude country, no lo utilizaremos.
* Title, demasiado incompleto.

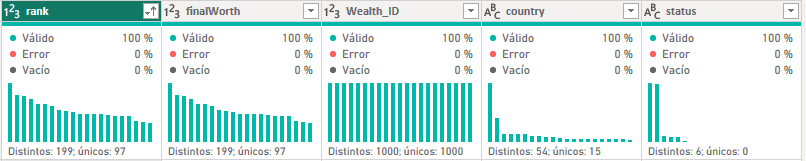


5 – Conecte BigQuery con PowerBI e importe el dataset.

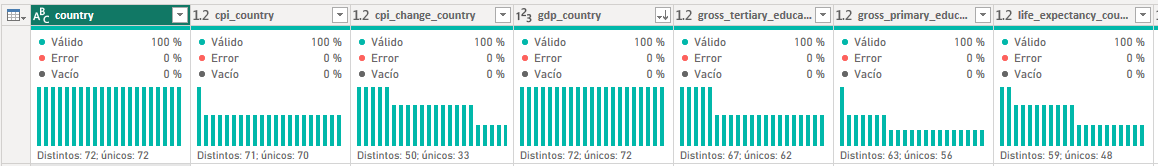
6- Modelo de datos:  
  


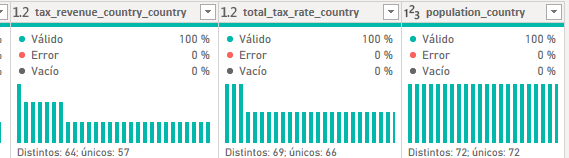
En este paso realice las adaptaciones para distribuir la información en tablas de hechos y dimensiones en PowerQuery, segmentación de la información y creación de relaciones, de acuerdo al modelo de datos.

* Facts

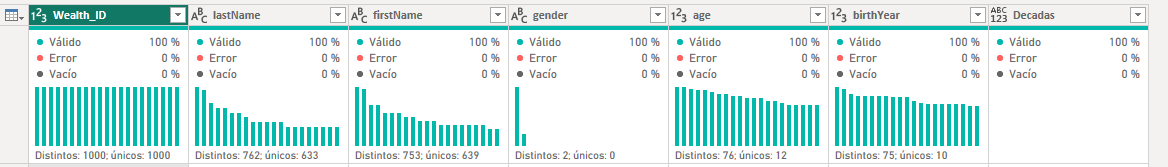


* Dim\_Country

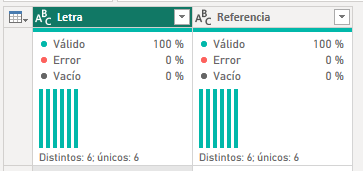




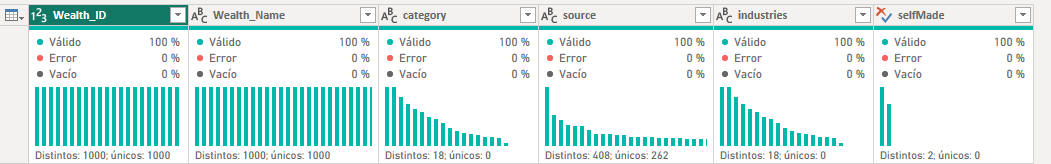
* Dim\_Person



* Dim\_Status



* Dim\_Wealth



7 Para finalizar la capa Gold:

Creación de columnas y ajuste de datos en PowerQuery

* índice Wealth\_ID (permite sortear que el ranking tiene posiciones compartidas)
* Décadas (valor auxiliar para facilitar legibilidad de gráfico)
* Cambio de valores en gdp\_country para facilitar la muestra de unidades.

Creación de columnas dinámicas para apoyar a los gráficos

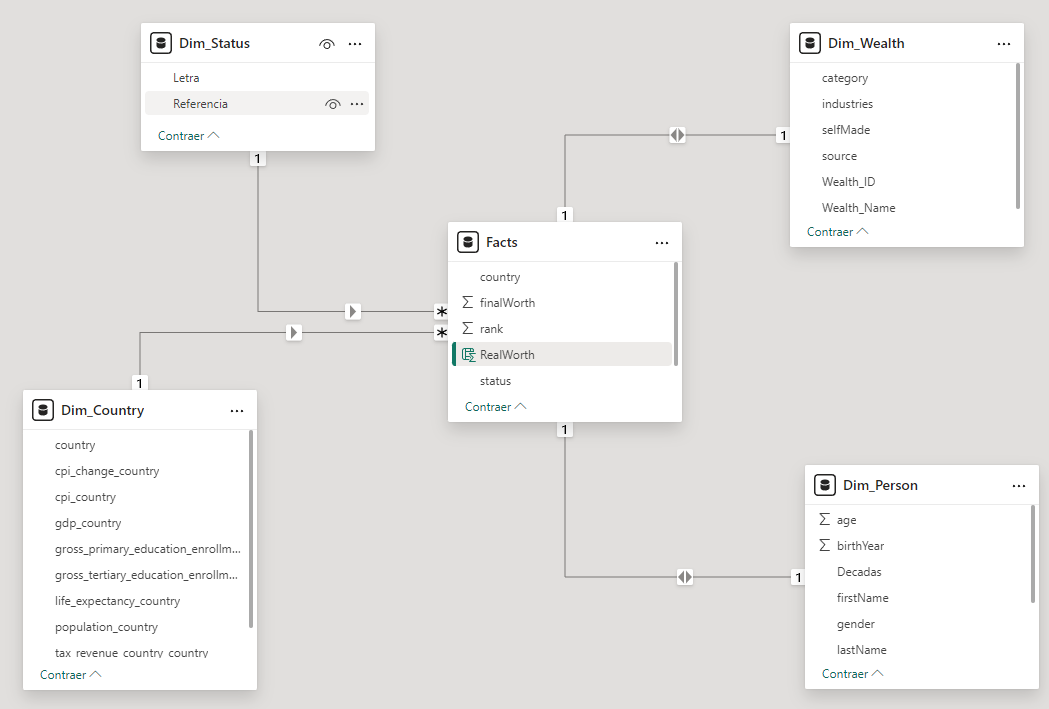
* NombreCompleto
* Rank\_NombreCompleto
* RealWorth para facilitar expresar unidades correctamente.

Finalmente genérelos gráficos, medidas y reportes definidos por el plan de métricas.

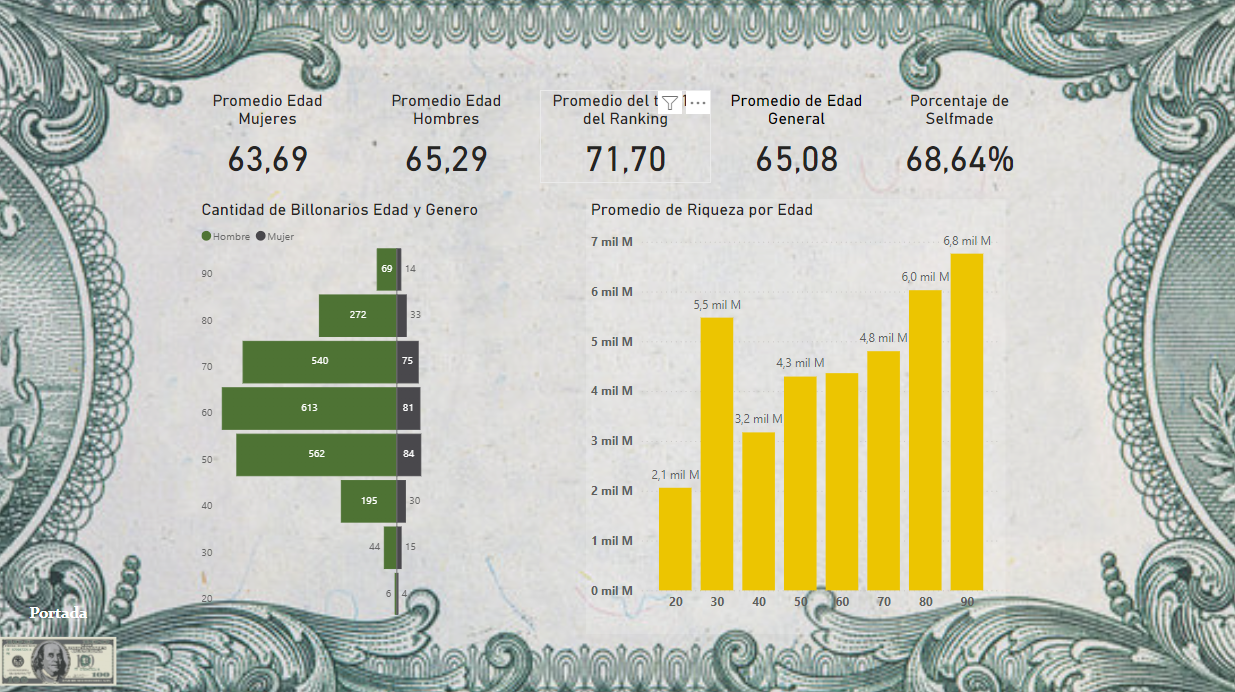


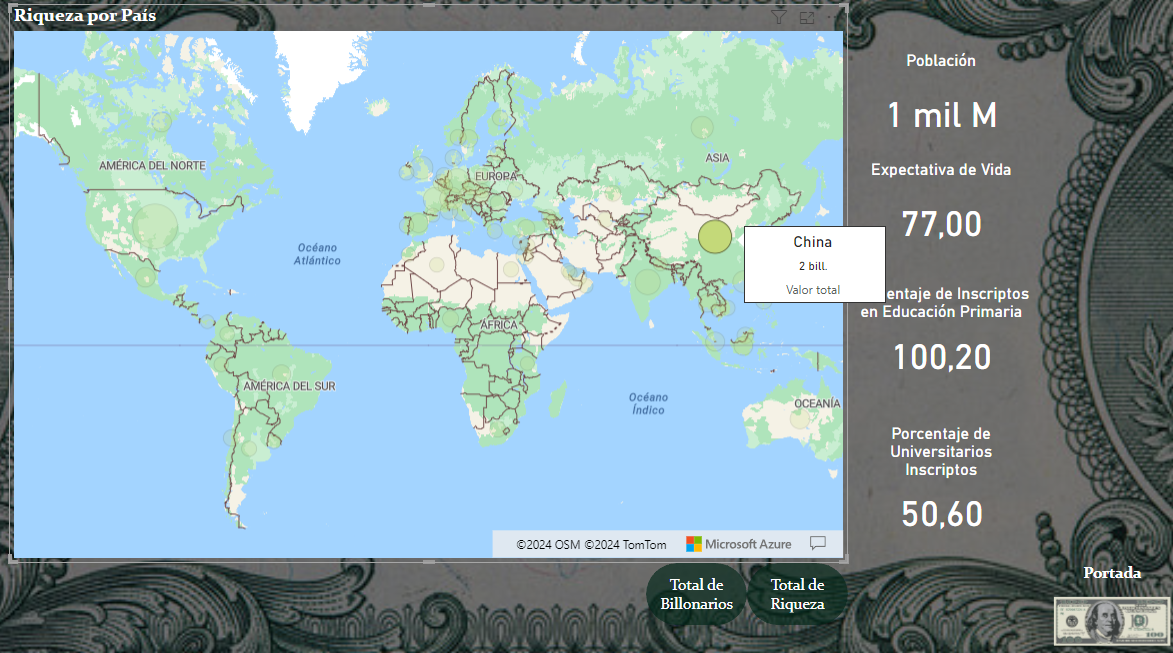
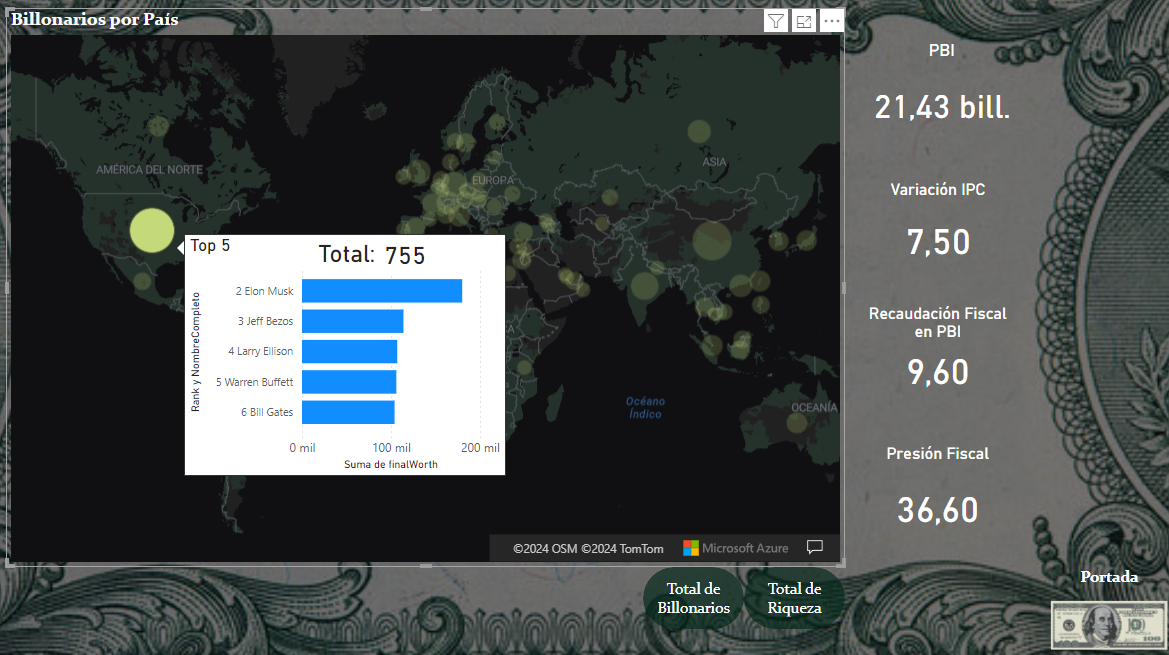
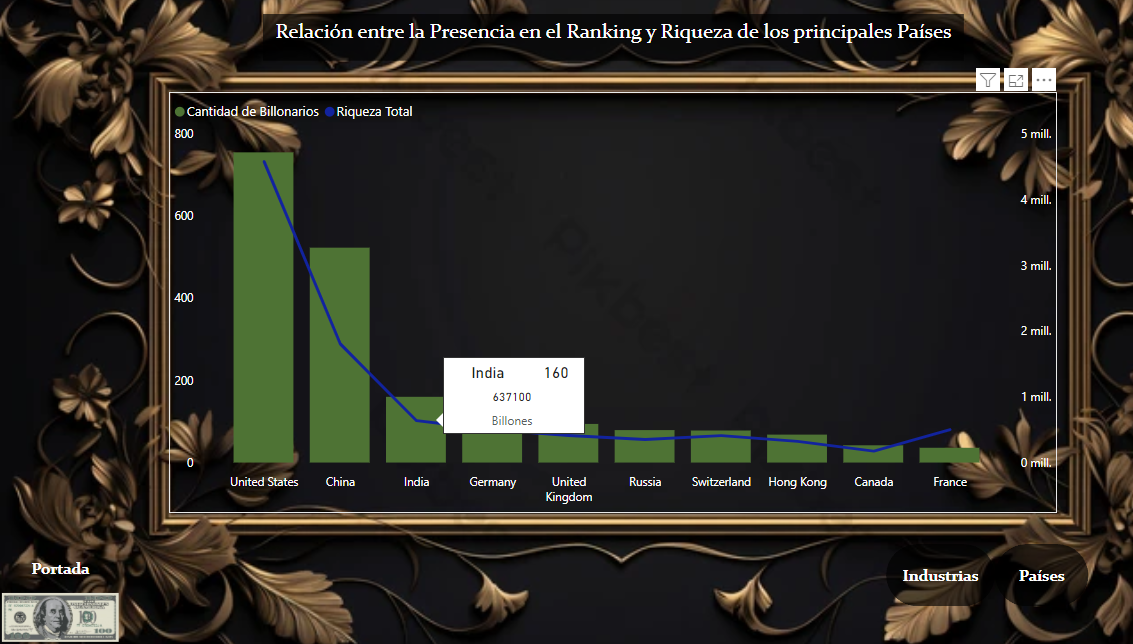


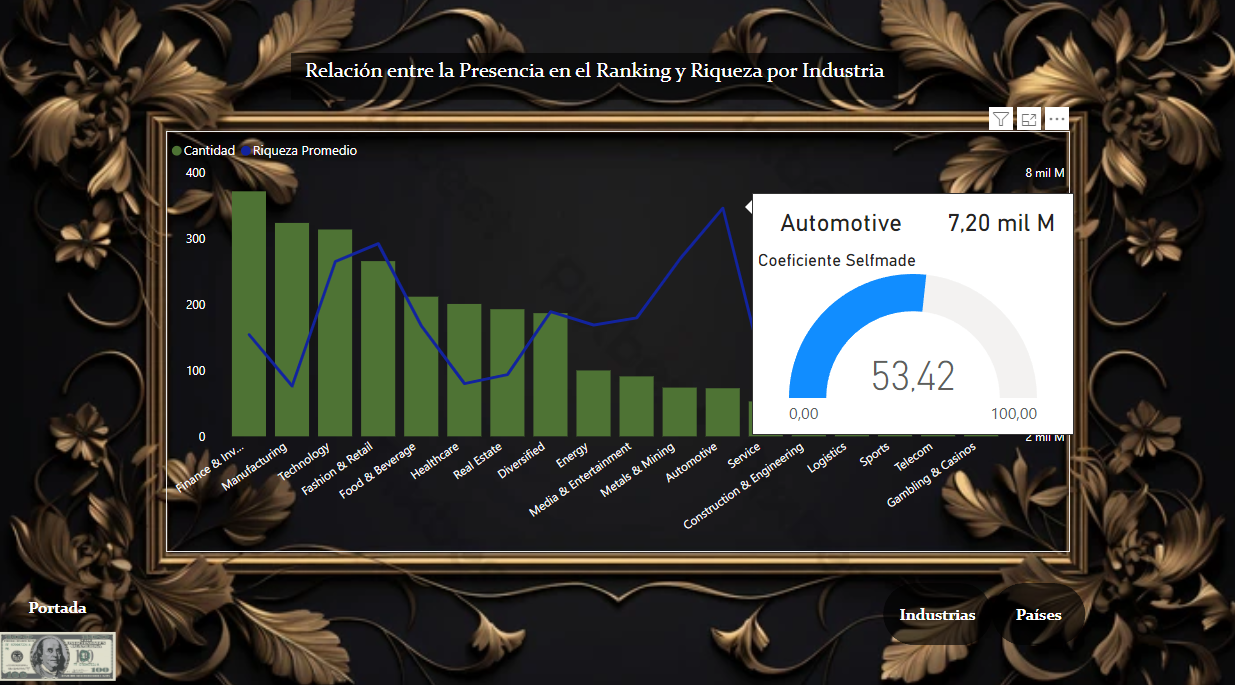
Capturas de Dashbord y relaciones de datos en Power BI

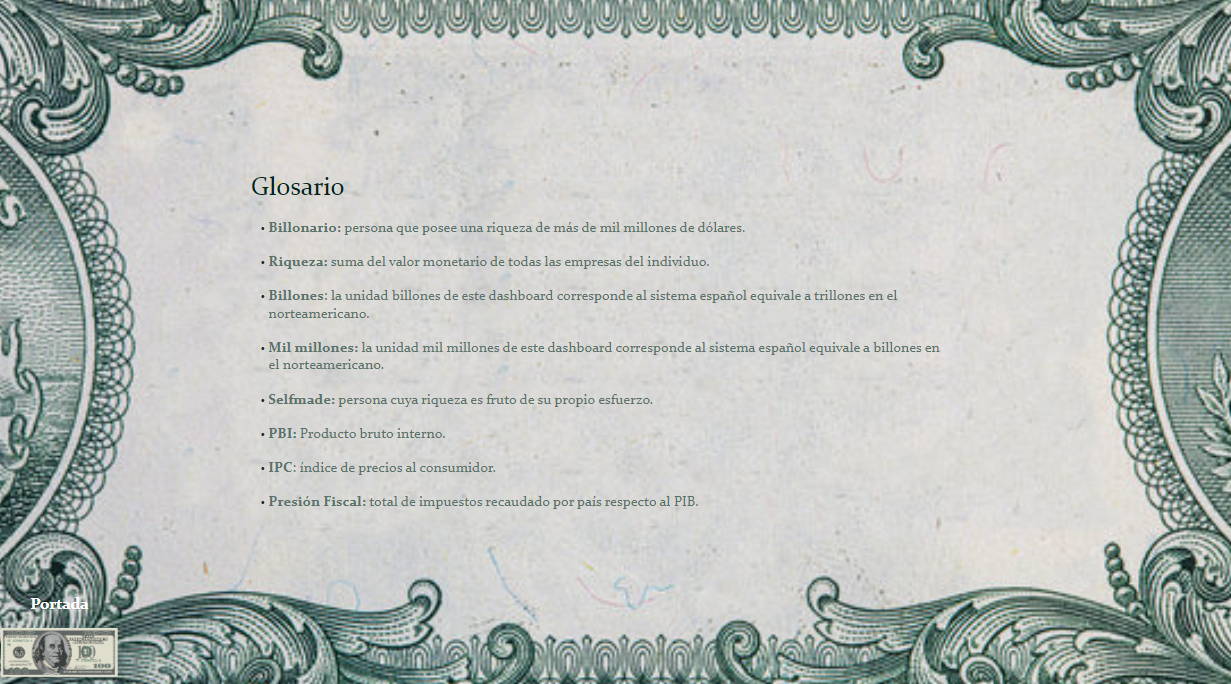










**Conclusiones Generales:**

Se comprobó la primera hipótesis con respecto a la edad, continua aumentando la edad promedio en el ranking de billonarios ya que mantiene una tendencia creciente en comparación con las medidas anteriores.

También se corroboró la segunda hipótesis la cual afirmaba que existen más hombres que mujeres en el ranking, pero no hay diferencia significativa en las edades promedio (menor al 5%), si la hay en el coeficiente de selfmade, se visualiza una gran cantidad de hombres en la categoría selfmade.

Los países que lideran en cantidad de millonarios son los mismos que lo hacen en valor total, con marcada diferencia a favor de Estados unidos, China, la tendencia general indica que los países con mayor nivel de desarrollo tienen más millonarios y un mayor valor.

China se presenta como un caso excepcional con un valor total bajo en relación con su cantidad de billonarios, mientras que Francia hace el caso inverso.

Se corrobora parcialmente la última hipótesis. Las principales industrias que concentran la mayor cantidad de millonarios incluyen a finanzas y tecnología, pero quedan relegadas en cuanto a valor total. Otras áreas como el Juego, la Industria Automotriz, Textil-Retail y la Logística, generan más riqueza (en promedio).