



PRESENTACIÓN DE

Proyecto final



Integrantes y objetivo

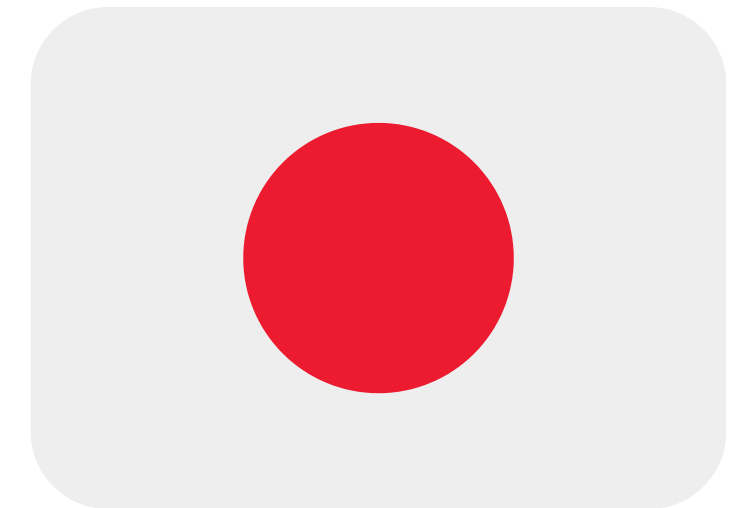
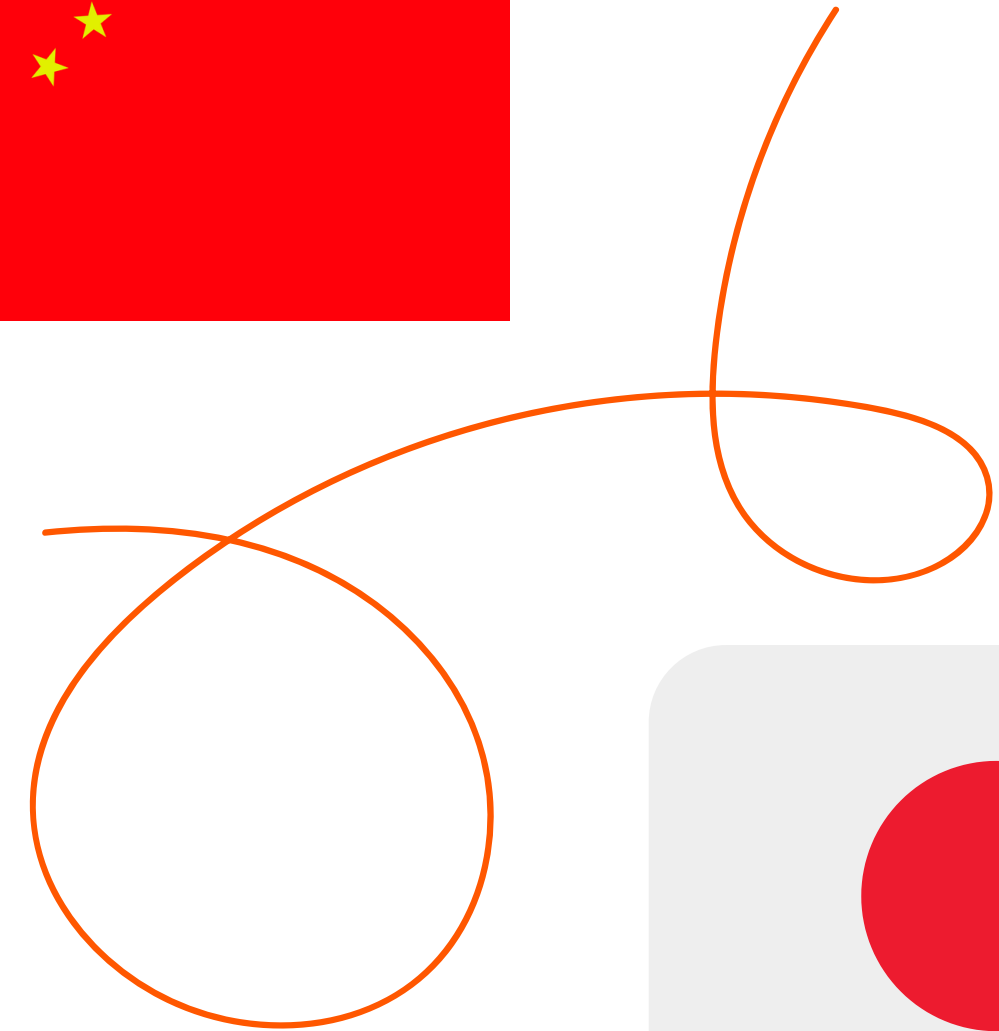
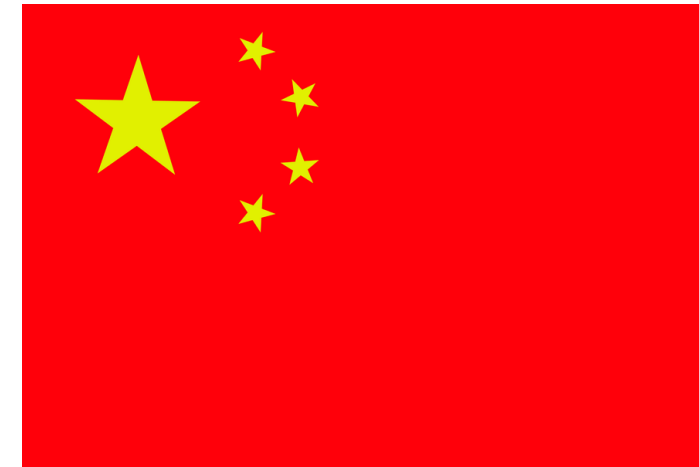


Demostrar las habilidades obtenidas en el diplomado de ciencia y análisis de datos con respecto a la búsqueda e identificación de bases de datos, utilización de software para la limpieza de estas mismas y analizar mediante gráficos los resultados obtenidos en dicha limpieza.

Minerva Nuit Pimentel Cobian
Luis Guillermo Briceño Ochoa
René Miguel Rico Moctezuma
Kassandra Larios
Briand Alfredo Sánchez Granada

Selección y Justificación de la Base de Datos

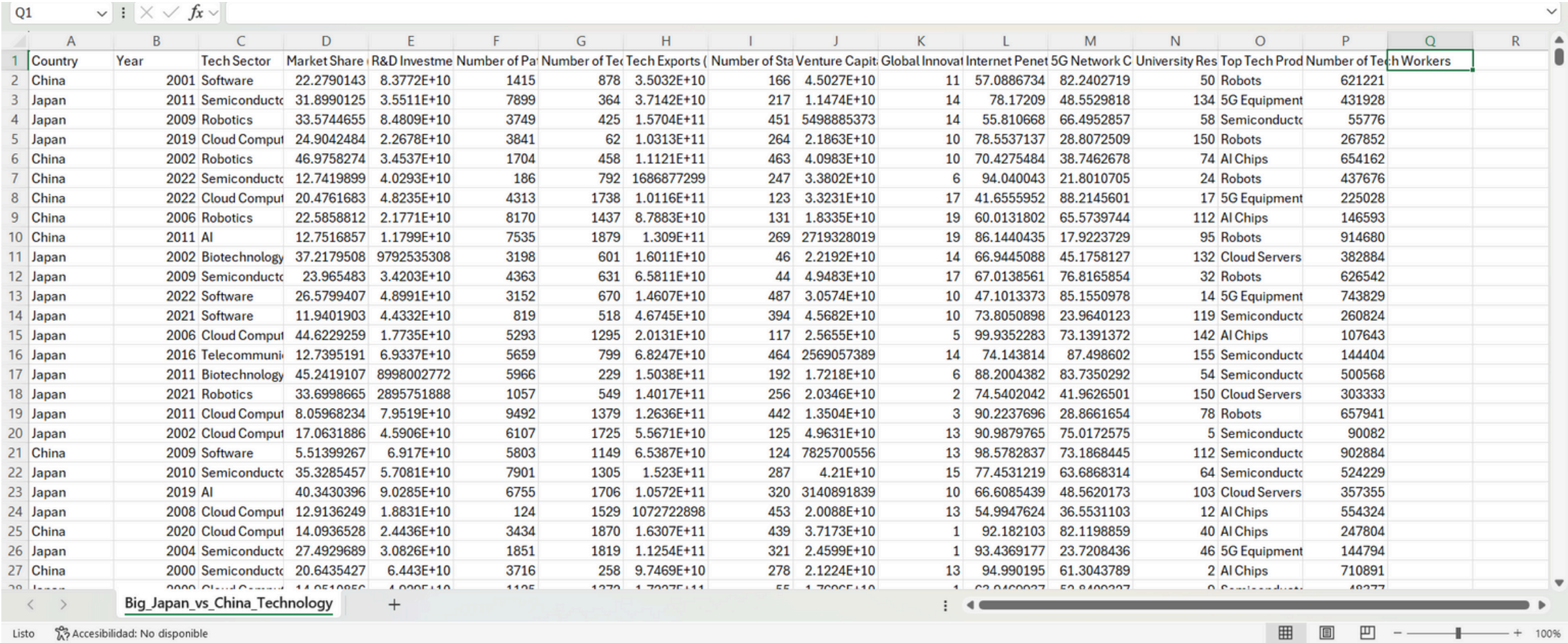
Optamos por un análisis directo hacia la comparativa del sector tecnológico de dos grandes países, Japón y China. Los cuales mediante un Ranking Mundial de Innovación de entre el año 2000 al 2023, se sometieron a un sondeo con respecto al área informática la cual se analiza desde el procesamiento y levantamiento de una nube hasta algo más físico como capacitores y/o componentes eléctricos para la elaboración del armado de PCB's



Limpieza de datos

1er vista

Como se puede observar en la ilustración 1, La base de datos está acomodada en un formato el cual no es muy cómodo de analizar por lo que se prosiguió a dar una limpieza en el doc CSV acomodando de una forma más óptima la información, dándo un orden y un ID a cada respuesta para así identificar de forma más eficiente cada valor.



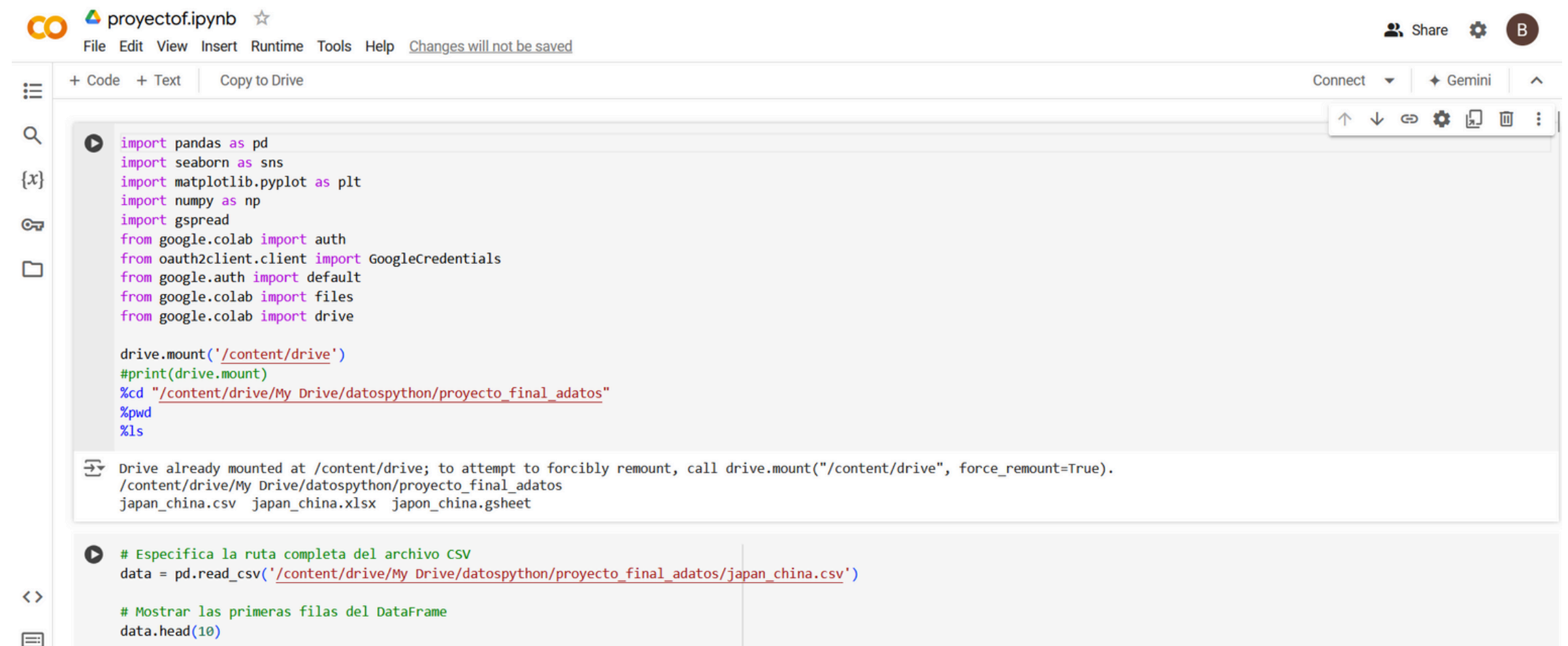
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Country	Year	Tech Sector	Market Share	R&D Investme	Number of Pa	Number of Tec	Tech Exports (Number of Sta	Venture Capit	Global Innovat	Internet Penet	5G Network C	University Res	Top Tech Prod	Number of Tech	Workers	
2	China	2001	Software	22.2790143	8.3772E+10	1415	878	3.5032E+10	166	4.5027E+10	11	57.0886734	82.2402719	50	Robots	621221		
3	Japan	2011	Semiconduct	31.8990125	3.5511E+10	7899	364	3.7142E+10	217	1.1474E+10	14	78.17209	48.5529818	134	5G Equipment	431928		
4	Japan	2009	Robotics	33.5744655	8.4809E+10	3749	425	1.5704E+11	451	5498885373	14	55.810668	66.4952857	58	Semiconduct	55776		
5	Japan	2019	Cloud Comput	24.9042484	2.2678E+10	3841	62	1.0313E+11	264	2.1863E+10	10	78.5537137	28.8072509	150	Robots	267852		
6	China	2002	Robotics	46.9758274	3.4537E+10	1704	458	1.1121E+11	463	4.0983E+10	10	70.4275484	38.7462678	74	AI Chips	654162		
7	China	2022	Semiconduct	12.7419899	4.0293E+10	186	792	1686877299	247	3.3802E+10	6	94.040043	21.8010705	24	Robots	437676		
8	China	2022	Cloud Comput	20.4761683	4.8235E+10	4313	1738	1.0116E+11	123	3.3231E+10	17	41.6555952	88.2145601	17	5G Equipment	225028		
9	China	2006	Robotics	22.5858812	2.1771E+10	8170	1437	8.7883E+10	131	1.8335E+10	19	60.0131802	65.5739744	112	AI Chips	146593		
10	China	2011	AI	12.7516857	1.1799E+10	7535	1879	1.309E+11	269	2719328019	19	86.1440435	17.9223729	95	Robots	914680		
11	Japan	2002	Biotechnology	37.2179508	9792535308	3198	601	1.6011E+10	46	2.2192E+10	14	66.9445088	45.1758127	132	Cloud Servers	382884		
12	Japan	2009	Semiconduct	23.965483	3.4203E+10	4363	631	6.5811E+10	44	4.9483E+10	17	67.0138561	76.8165854	32	Robots	626542		
13	Japan	2022	Software	26.5799407	4.8991E+10	3152	670	1.4607E+10	487	3.0574E+10	10	47.1013373	85.1550978	14	5G Equipment	743829		
14	Japan	2021	Software	11.9401903	4.4332E+10	819	518	4.6745E+10	394	4.5682E+10	10	73.8050898	23.9640123	119	Semiconduct	260824		
15	Japan	2006	Cloud Comput	44.6229259	1.7735E+10	5293	1295	2.0131E+10	117	2.5655E+10	5	99.9352283	73.1391372	142	AI Chips	107643		
16	Japan	2016	Telecommuni	12.7395191	6.9337E+10	5659	799	6.8247E+10	464	2569057389	14	74.143814	87.498602	155	Semiconduct	144404		
17	Japan	2011	Biotechnology	45.2419107	8998002772	5966	229	1.5038E+11	192	1.7218E+10	6	88.2004382	83.7350292	54	Semiconduct	500568		
18	Japan	2021	Robotics	33.6998665	2895751888	1057	549	1.4017E+11	256	2.0346E+10	2	74.5402042	41.9626501	150	Cloud Servers	303333		
19	Japan	2011	Cloud Comput	8.05968234	7.9519E+10	9492	1379	1.2636E+11	442	1.3504E+10	3	90.2237696	28.8661654	78	Robots	657941		
20	Japan	2002	Cloud Comput	17.0631886	4.5906E+10	6107	1725	5.5671E+10	125	4.9631E+10	13	90.9879765	75.0172575	5	Semiconduct	90082		
21	China	2009	Software	5.51399267	6.917E+10	5803	1149	6.5387E+10	124	7825700556	13	98.5782837	73.1868445	112	Semiconduct	902884		
22	Japan	2010	Semiconduct	35.3285457	5.7081E+10	7901	1305	1.523E+11	287	4.21E+10	15	77.4531219	63.6868314	64	Semiconduct	524229		
23	Japan	2019	AI	40.3430396	9.0285E+10	6755	1706	1.0572E+11	320	3140891839	10	66.6085439	48.5620173	103	Cloud Servers	357355		
24	Japan	2008	Cloud Comput	12.9136249	1.8831E+10	124	1529	1072722898	453	2.0088E+10	13	54.9947624	36.5531103	12	AI Chips	554324		
25	China	2020	Cloud Comput	14.0936528	2.4436E+10	3434	1870	1.6307E+11	439	3.7173E+10	1	92.182103	82.1198859	40	AI Chips	247804		
26	Japan	2004	Semiconduct	27.4929689	3.0826E+10	1851	1819	1.1254E+11	321	2.4599E+10	1	93.4369177	23.7208436	46	5G Equipment	144794		
27	China	2000	Semiconduct	20.6435427	6.443E+10	3716	258	9.7469E+10	278	1.2224E+10	13	94.990195	61.3043789	2	AI Chips	710891		
28	Japan	2000	Cloud Comput	14.0510056	4.000E+10	1105	1270	1.700E+11	55	1.700E+11	1	62.0400007	50.8400007	0	Semiconduct	400000		

ILUSTRACIÓN 1. Documento bruto de la información de la BD.

Limpieza de datos

2da vista

Mediante Google Colab se realizó un pequeño programa para efectuar nuestra segunda y última limpieza mediante el lenguaje de programación Python, siendo esta la herramienta más útil y rápida para poder identificar mediante los pandas, las variables que nosotros deseemos extraer.



```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
import gspread
from google.colab import auth
from oauth2client.client import GoogleCredentials
from google.auth import default
from google.colab import files
from google.colab import drive

drive.mount('/content/drive')
#print(drive.mount)
%cd "/content/drive/My Drive/datospython/proyecto_final_adatos"
%pwd
%ls

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
/content/drive/My Drive/datospython/proyecto_final_adatos
japan_china.csv  japan_china.xlsx  japon_china.gsheets

# Especifica la ruta completa del archivo CSV
data = pd.read_csv('/content/drive/My Drive/datospython/proyecto_final_adatos/japan_china.csv')

# Mostrar las primeras filas del DataFrame
data.head(10)
```

ILUSTRACIÓN 2. Programa realizado en Google Colab

Limpieza de datos

Analizando la base de datos al inicio, pudimos percatarnos del tipo de información que manejaba el sondeo, por lo que el utilizar números y texto podría generarnos un conflicto por falta de valores, pero en este caso, siendo una base de datos más específica fue mucho más sencillo reorganizar toda la información mediante Python, por lo que proseguiríamos al manejo de nuestros gráficos y demostrar mejor los objetivos principales que identificamos en la base de datos.



	Country	Year	Tech Sector	Market Share (%)	R&D Investment (in USD)	of Patents Filed (Annual)	Number of Tech Companies	Tech Exports (in USD)	Number of Startups	Venture Capital Funding (in USD)	Global Innovation Ranking	Internet Penetration (%)	Network Coverage (%)	University Research Collaborations	Top Tech Products Exported	Number of Tech Workers
0	China	2001	Software	22	83772077459	1415	878	35031547673	166	45026942421	11	57	82	50	Robots	621221
1	Japan	2011	Semiconductor	32	35511335905	7899	364	37142090222	217	11473812659	14	78	49	134	5G Equipment	431928
2	Japan	2009	Robotics	34	84809484676	3749	425	157040631146	451	5498885373	14	56	66	58	Semiconductors	55776
3	Japan	2019	Cloud Computing	25	22678211745	3841	62	103128355995	264	21862782400	10	79	29	150	Robots	267852
4	China	2002	Robotics	47	34536550748	1704	458	111205589580	463	40982820657	10	70	39	74	AI Chips	654162
5	China	2022	Semiconductor	13	40293285924	186	792	1686877299	247	33802354648	6	94	22	24	Robots	437676

ILUSTRACIÓN 3. Datos arrojados una vez introducidos en el programa de Google Colab.

Visualización de datos

Objetivo

Determinar la inversión de compra de startups

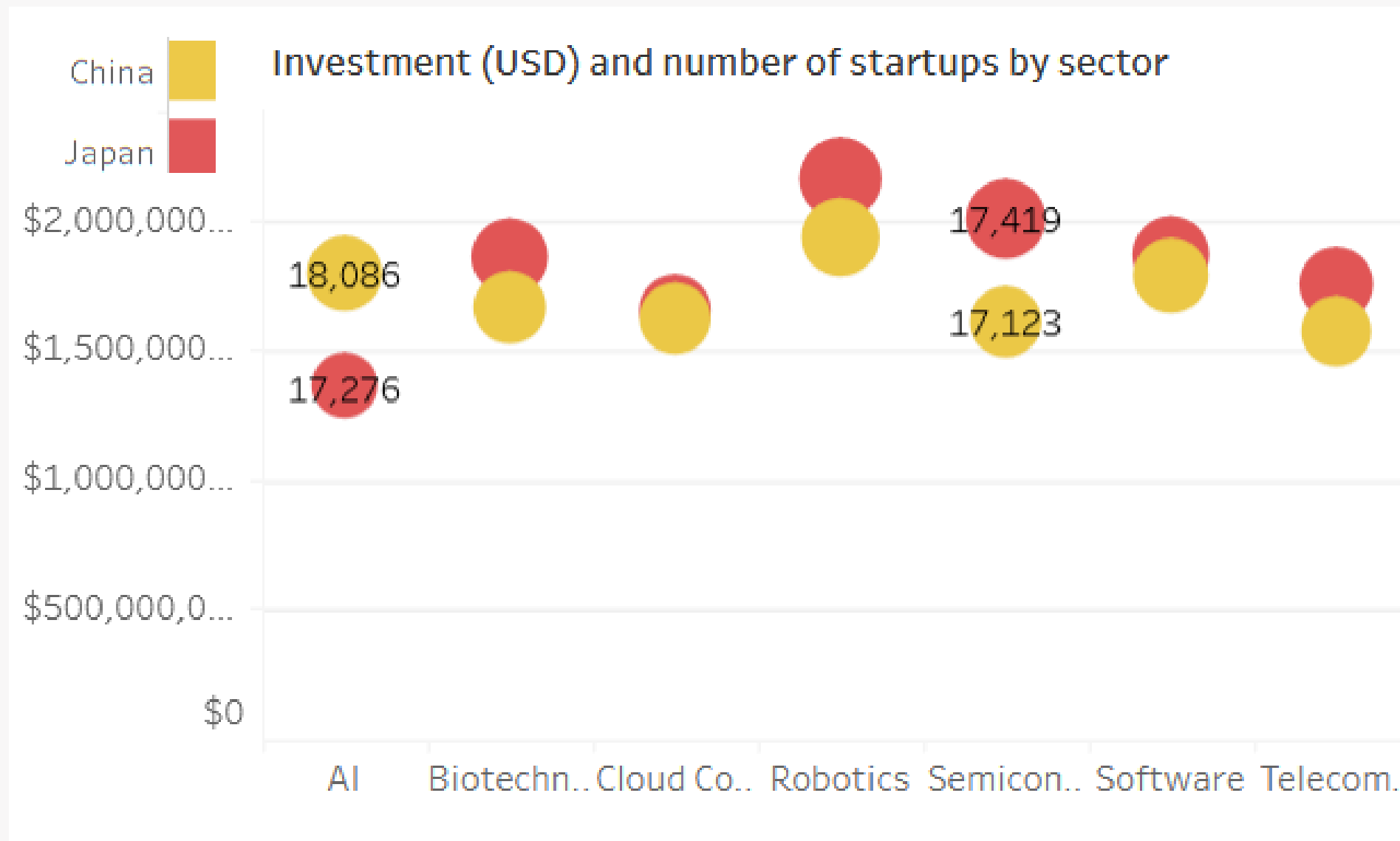


GRÁFICO 1. Identificación de la inversión

Los sectores en los que ambos países se encuentran empatados en cuanto a la inversión en empresas con potencial desarrollo son: Robótica, Software, biotecnología. Mientras que el área en el que ambos países menos han realizado compras de empresas es en el de Inteligencia artificial, esto puede deberse a que esta tecnología se encuentra aún en expansión.

Visualización de datos

Identificar el Ranking Mundial de Innovación por sector tecnológico

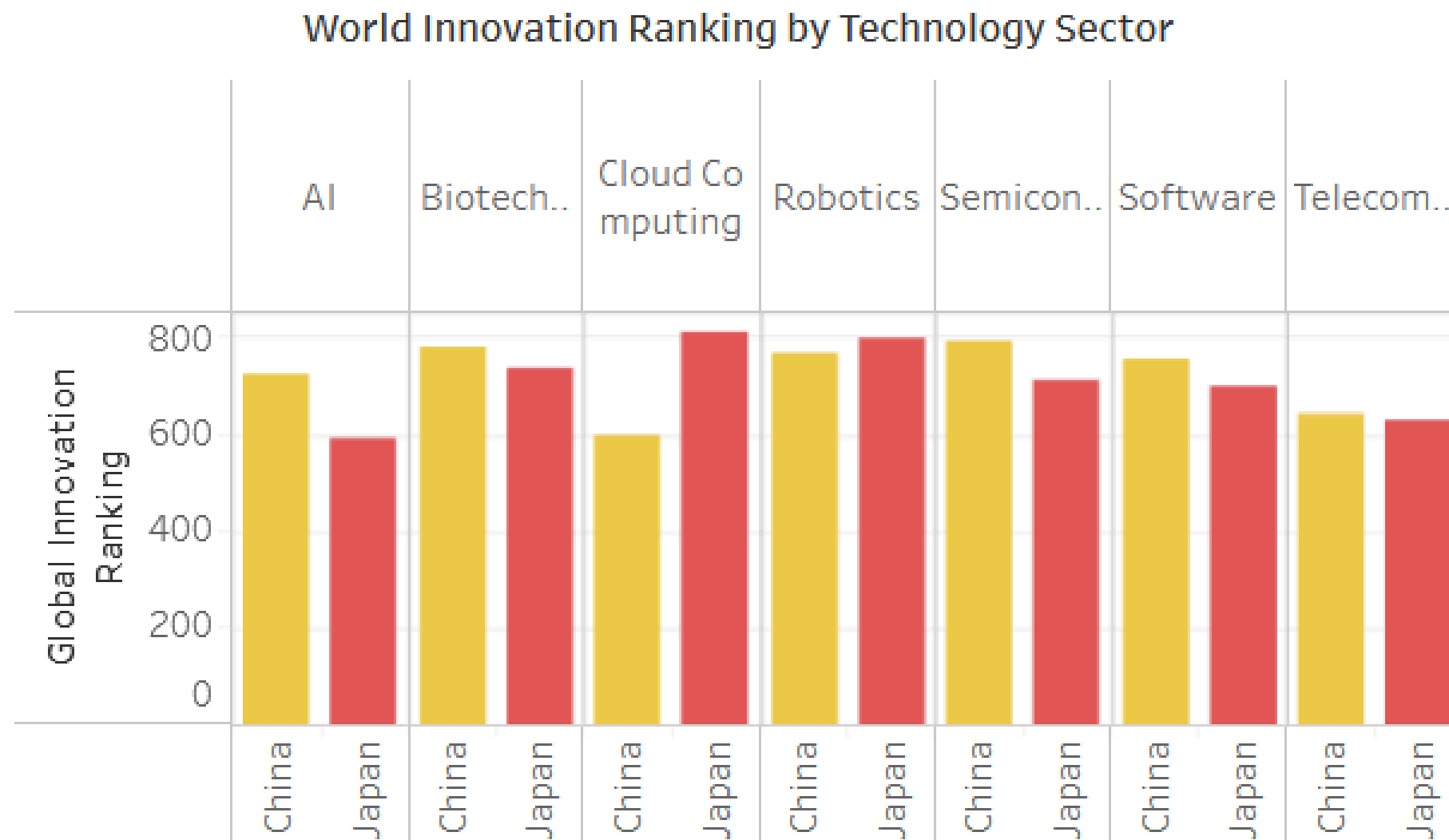


GRÁFICO 2. Ranking de innovación mundial

Objetivo

Observamos que ambos países se encuentran en competencia en la mayoría de los sectores, sin embargo, si existe una mayor estabilidad de números para China, para mejorar tendría que aumentar números en el cómputo en la nube, mientras que Japón si tendría que mejorar sus números por ejemplo en la inteligencia artificial y la industria de los semiconductores.

Visualización de datos

Revisar la ganancia de exportaciones por sector

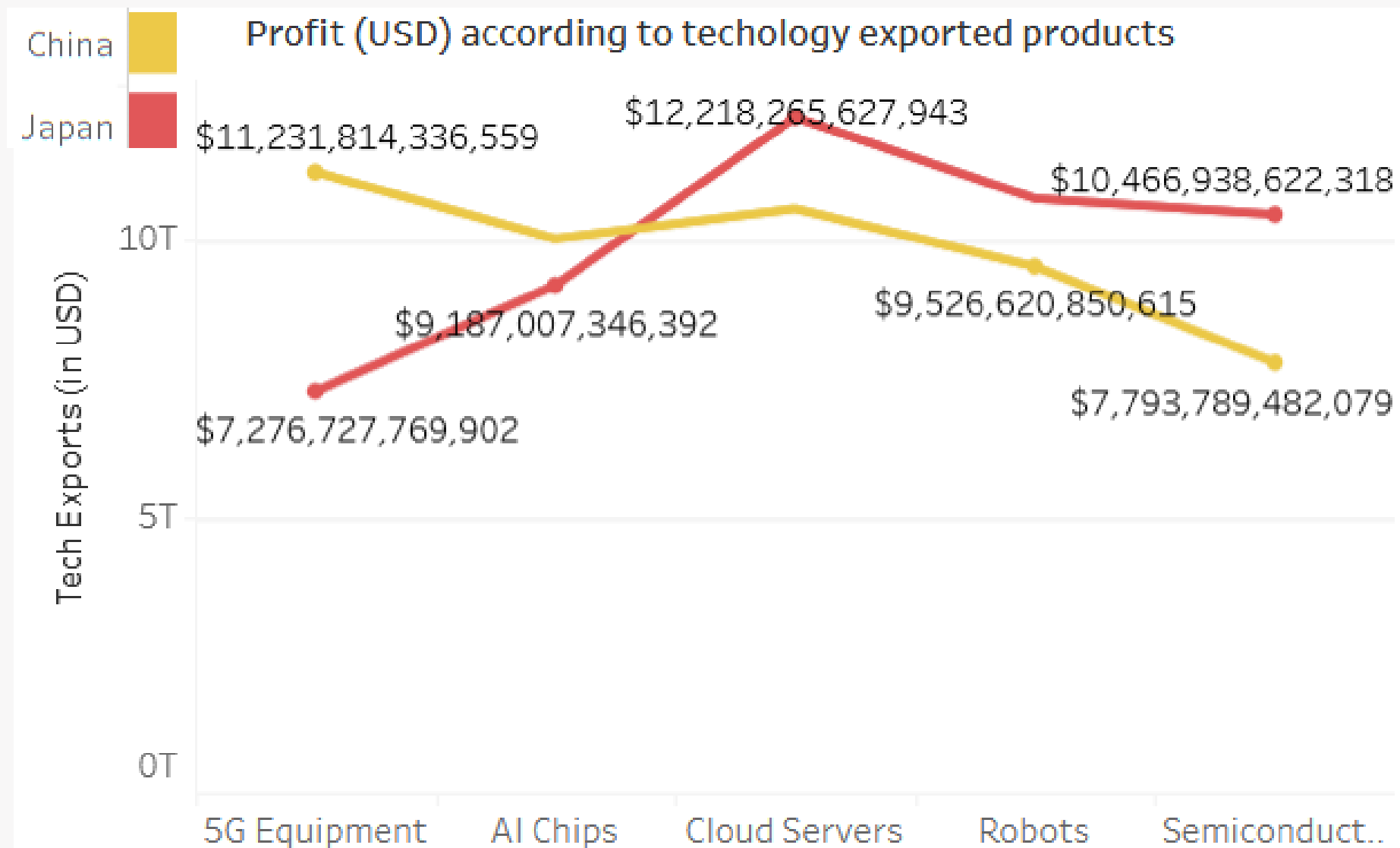


GRÁFICO 3. Productos exportados de ambos países.

Objetivo

Se puede observar que para Japón las mayores ganancias han sido obtenidas en la industria del cómputo en la nube, mientras que para China es la tecnología 5G,

Visualización de datos

Objetivo

Market Share es el indicador en porcentaje que indica las ventas que ha tenido un negocio, en comparación con las ventas totales de un mercado.

Observamos que para china los porcentajes mas altos se encuentran en la robótica y los semiconductores, mientras que para Japón tienen valores altos en la tecnología en la nube, la Inteligencia artificial y los dispositivos con tecnología 5G.

Pero hay que resaltar la importancia de que china maneja más componentes electrónicos por la mano de obra que maneja,

Analizar el “market share” de los productos exportados

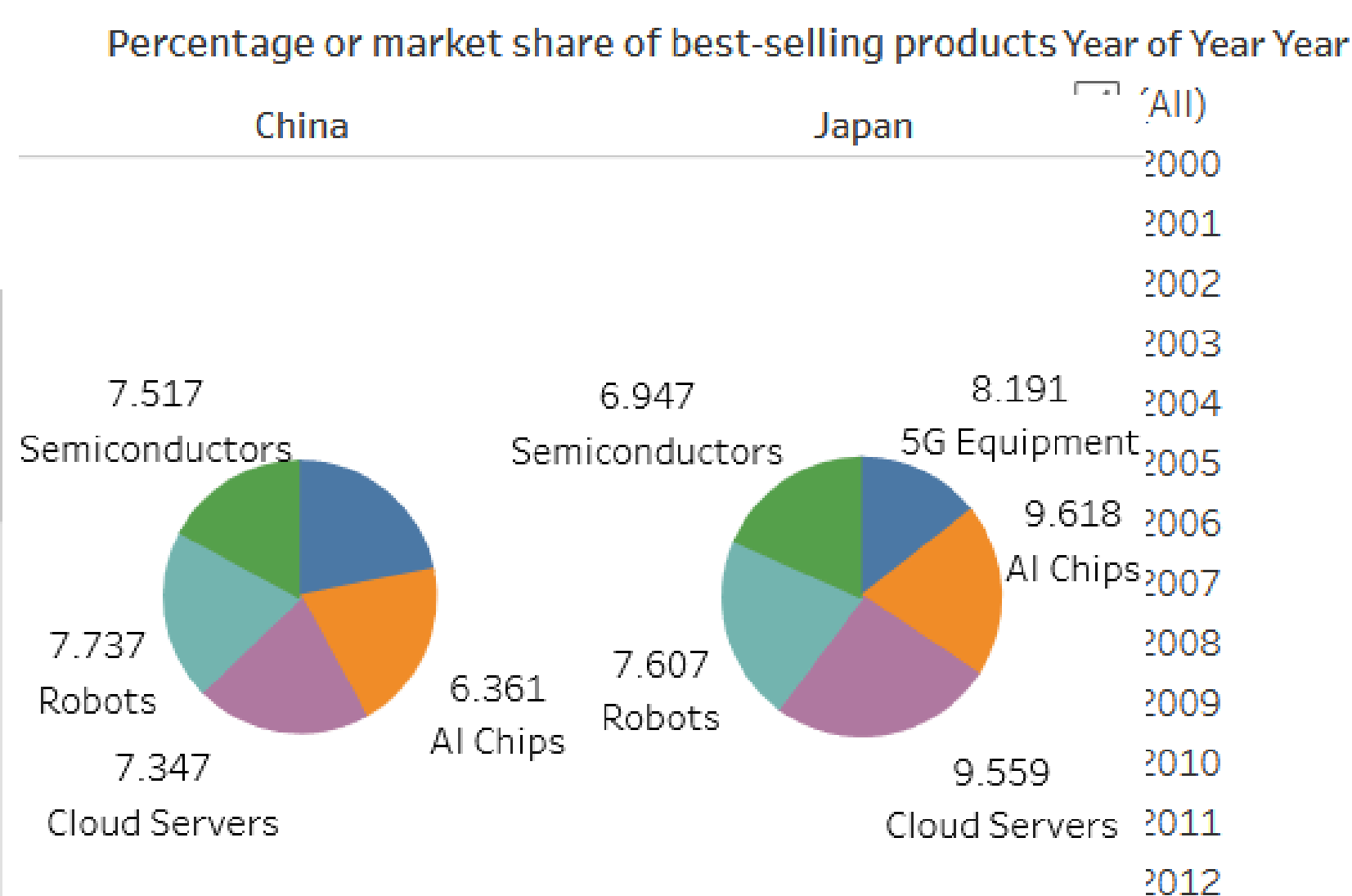
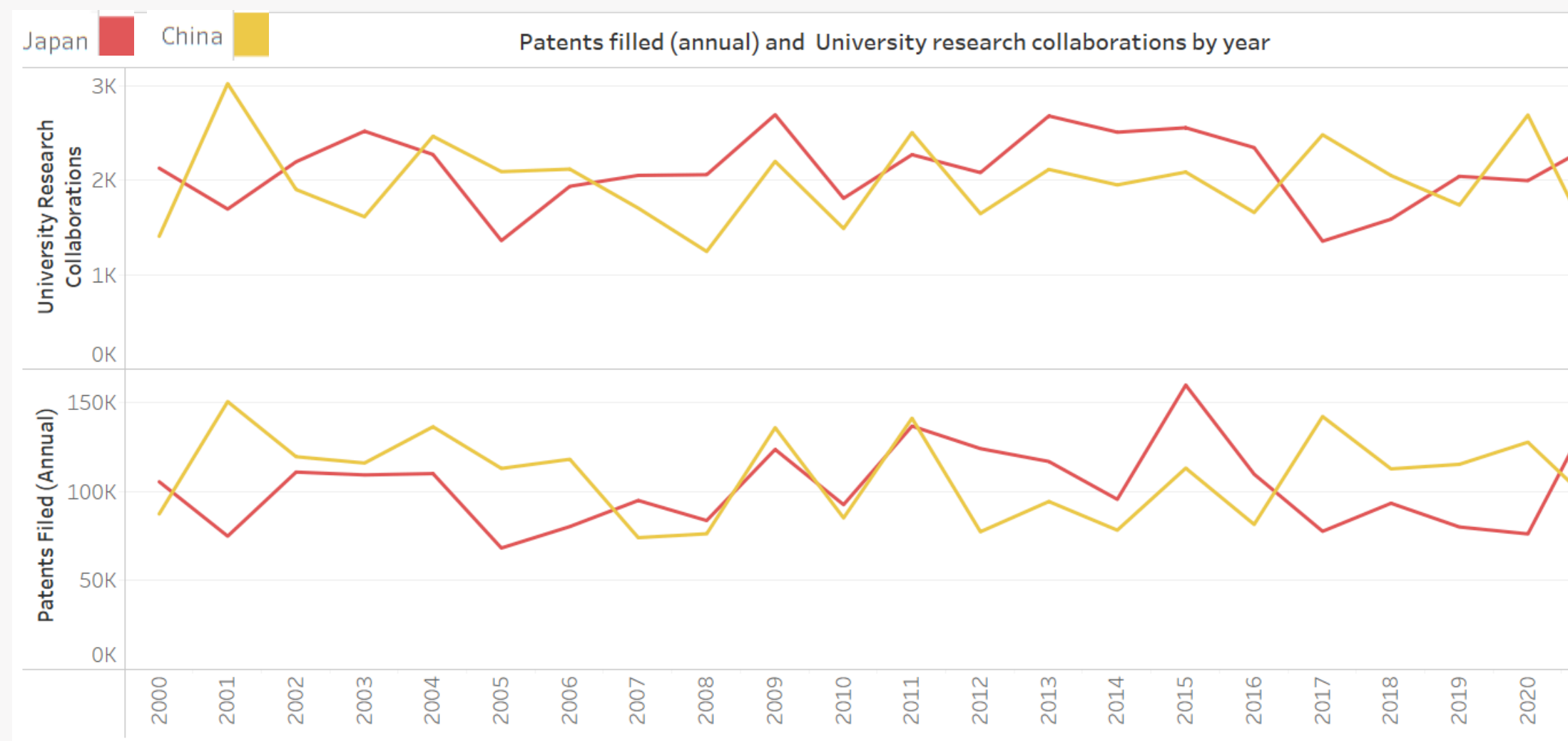


GRÁFICO 4. Market Share por producto

Visualización de datos

Objetivo

Revisar relación de numero de patentes con colaboraciones universitarias



Si es posible ver una correlación entre estos datos para ambos países, ya que se observa un fenómeno en el cual, en los años en los que incrementó el número de patentes se ve también un incremento en las Colaboraciones de investigación universitaria.

GRÁFICO 5. Patentes anuales y colaboraciones de investigación

CONCLUSIONES

Dos grandes potencias por años se han enfrentado ante la oferta y la demanda, se conoce que China siendo un País de gran producción en masas es famosa por la cantidad de productos tecnológicos que ofrece al resto de Países pero a ciencia cierta ¿qué tanto es lo que puede ofrecer china contra lo que ofrece Japón? Bueno gracias a estos análisis pudimos observar que por el ámbito de la innovación, Japón ha sabido aprovechar más las redes de información. Por lo que la implementación de la 5a generación de telefonía de redes móviles (5G) ha sido más favorable para el País del Sol naciente. A su vez, por avance tecnológico los CRM (Gestión de Relación con el Cliente), han ofrecido a todo tipo de empresa un servicio con el cual poder manejar de forma más eficiente tanto sus bases de datos, como cartera de clientes o incluso en los más avanzados manejan una inteligencia artificial que puede vincular todo lo relacionado con la empresa a una nube, siendo este el mayor avance tecnológico para la innovación del trabajo.

China por otra, sí maneja los mismos servicios que Japón ha ofrecido pero se demuestra que en cuestión de ventas no le ha favorecido a China, pero por el ámbito físico como los capacitores, microcontroladores inclusive en la creación de robots supera en número a Japón.