

Grado en Estadística

Título: Open data y su impacto económico y social.
Aplicación con Shiny y Amazon Quick Sight.

Autor: Núria Núñez Rubio

Director: Salvador Torra Porras

Departamento: Econometría, Estadística y Economía Española

Convocatoria: 2020-2021



RESUMEN DEL TRABAJO

En este trabajo se muestra un análisis detallado de la influencia de los datos abiertos en la sociedad y en la economía actual. Para recoger la información necesaria se han consultado fuentes de organizaciones e instituciones que trabajan con datos abiertos, de este modo se ha conocido la evolución, innovación y mejora de la economía que están causando.

Este estudio tiene también como objetivo la implementación de dos aplicaciones de análisis y visualización de datos abiertos, concretamente de comercio exterior de España. En estas aplicaciones se reutilizan datos con ayuda de técnicas estadísticas para sacar de ellos información de valor. Cada una de las aplicaciones se realizan con una plataforma diferente (Shiny y Quick Sight), de esta manera se ha podido ver cuál de ellas da mejores resultados y cuál se usaría para llevar a cabo futuros proyectos.

SUMMARY OF THE WORK

This paper shows a detailed analysis of the influence of open data on society and the economy today. To collect the necessary information, sources from organizations and institutions that work with open data have been consulted, in this way the evolution, innovation and improvement of the economy that they are causing have been known.

This study also aims at the implementation of two applications for analysis and visualization of open data, specifically foreign trade in Spain. In these applications, data is reused with the help of statistical techniques to extract valuable information from it. Each of the applications are made with a different platform (Shiny and Quick Sight), in this way we have been able to see which of them gives better results and which would be used to carry out future projects.

PALABRAS CLAVE

Datos abiertos, Economía de los datos, Reutilización de datos, Inteligencia empresarial, Software analítico, API, Análisis de datos, Visualización de datos.

KEYWORDS

Open data, Data economy, Data reuse, Business intelligence, Analytical software, API, Data analysis, Data visualization.

CLASIFICACIÓN AMS

62-XX STATISTICS

62-02 Research exposition

62-04 Explicit machine computation and programs (not the theory of computation or programming)

62-07 Data analysis

62Pxx Applications

62P20 Applications to economics

Índice

I. INTRODUCCIÓN	7
II. METODOLOGÍA.....	8
III. DATOS ABIERTOS.....	9
3.1. ¿Qué son los datos abiertos?.....	9
3.2. Importancia y actualidad de los datos abiertos.....	11
3.3. Regulación y protección de datos.....	14
3.4. Clasificación de datos abiertos	17
3.5. Beneficios de los datos abiertos	18
3.6. Reutilización de datos abiertos.....	21
3.6.1. Proceso de reutilización.....	24
3.6.2. Importancia de la reutilización para empresas y negocios	26
3.6.3. Negocios y aplicaciones basadas en datos	29
3.6.4. Los datos abiertos en las administraciones públicas.....	30
IV. PLATAFORMAS DE VISUALIZACIÓN DE DATOS	32
4.1. Descripción e importancia	32
4.2. Descripción y posibilidades de Shiny y Amazon QuickSight	33
4.3. Comparación de plataformas.....	35
V. PROYECTO DE VISUALIZACIÓN DE DATOS ABIERTOS SOBRE COMERCIO EXTERIOR DE ESPAÑA.....	37
5.1. Descripción e interés real del proyecto	37
5.2. Descripción de los datos: datos de UN Comtrade Database	37
5.3. Maqueta inicial del proyecto	41
5.4. Proyecto en Shiny	42
5.4.1. Explicación de cálculos, indicadores y otros objetos de Dashboard	42
5.4.2. Presentación del proyecto.....	44
5.5. Proyecto en QuickSight.....	48
5.5.1. Explicación de cálculos, indicadores, y otros objetos del Dashboard	48
5.5.2. Presentación del proyecto.....	50
5.6. Comparación de resultados	53
VI. CONCLUSIONES	57
BIBLIOGRAFÍA	59
WEBGRAFÍA	60
ANEXO	63
Descarga e instalación de Shiny.....	63
Uso de Quick Sight	64
Código de la aplicación en Shiny.....	67

I. INTRODUCCIÓN

En este trabajo se quiere conocer la función que tienen los datos y la estadística en el día a día de la sociedad y de la economía. En concreto se investigará sobre cómo se saca cada vez más valor de los datos abiertos, tanto valor económico para empresas y organizaciones públicas, como información para progresar y optimizar infinidad de aspectos de la vida cotidiana.

Se estudiará el avance y la extensión de la estadística en los últimos años en los diferentes sectores de la industria y economía española y europea, así como las diferentes maneras de usarla para obtener beneficios a través de la reutilización de los datos abiertos.

La manera más eficiente y efectiva de reutilizar información para una empresa u organización es con la creación de aplicaciones y *dashboards* que analicen visualmente estos datos. Las aplicaciones al ser interactivas permiten al usuario entender mejor el análisis y la situación actual de los datos respecto un periodo pasado o de referencia.

Por tanto, en este trabajo se quieren comparar dos plataformas que permiten crear aplicaciones de visualización y análisis de datos para ver las diferentes capacidades y posibilidades que ofrecen. Estas dos plataformas serán Shiny de RStudio, que es un software libre y gratuito, y Quick Sight de Amazon, que es un software de análisis empresarial de pago. Para la comparación de las plataformas se plantea un proyecto de interés económico: **Análisis del comercio exterior de España**, en el que se analizarán los datos de UN Comtrade¹, que es una plataforma de datos abiertos comerciales creada por las Naciones Unidas. El objetivo es crear una aplicación con cada una de las plataformas que refleje de manera muy visual un análisis de los datos de interés y, finalmente poder concluir que ventajas e inconvenientes tienen cada una, cuál de ellas acaba siendo mejor para un análisis de datos públicos, y cuál usaríamos para futuros análisis en otros entornos de trabajo y/o con otros objetivos.

En cuanto a la estructura del trabajo, este se compone de dos grandes bloques: uno referente a los datos abiertos y otro enfocado a las plataformas de visualización. El primer bloque es de tipo teórico y se compone de diferentes temas de actualidad que se han considerado importantes e interesantes para responder a los objetivos de investigación del trabajo con respecto a la importancia de la *open data*. El segundo bloque contiene la parte empírica, en este se explicará el funcionamiento de las dos plataformas, se hará un seguimiento del proyecto de análisis y visualización de datos, y por último se compararán las dos plataformas en cuanto a posibilidades y teniendo en cuenta los resultados obtenidos del proyecto en cada una.

¹ <https://comtrade.un.org/>

II. METODOLOGÍA

Para la elaboración de la parte teórica del trabajo se ha investigado sobre el tema de los datos abiertos, su actualidad, y su influencia en la sociedad y en la economía. Buscando información en artículos web y leyendo *papers* de la mano de instituciones europeas y organizaciones del ámbito de la estadística y los datos abiertos. La parte más importante de la búsqueda de información es sintetizarla correctamente para poder enlazarla con datos de otras fuentes y responder a las preguntas de interés planteadas inicialmente.

En el segundo bloque de las plataformas de visualización, se ha indagado (a priori de realizar el proyecto) en las posibilidades que ofrece cada una de ellas, que limitaciones tienen y cuáles son las principales diferencias.

En cuanto al proyecto de análisis y visualización de datos, se elaboran las maquetas de las aplicaciones y se explica detalladamente el objetivo de cada elemento que se quiere construir en ellas. Primero para Shiny² y después para Quick Sight³, se lleva a cabo el proyecto realizando un análisis de los datos de UN Comtrade y plasmándolo con elementos visuales como gráficos y tablas. En el trabajo se explica detalladamente como se ha construido cada uno de los elementos de las aplicaciones en cada plataforma.

Una vez acabadas las dos aplicaciones, se comparan las plataformas teniendo en cuenta los resultados obtenidos con el proyecto. De esta manera se puede ver qué ventajas e inconvenientes se han tenido con cada una y cuál ha sido mejor para llevar a cabo un análisis y visualización de datos reales de comercio exterior de España.

² <https://shiny.rstudio.com/>

³ <https://aws.amazon.com/es/quicksight/>

III. DATOS ABIERTOS

3.1. ¿Qué son los datos abiertos?

Los datos abiertos (*open data*) en inglés son datos a los que se puede acceder de forma libre, gratuita e ilimitada. Todo el mundo puede acceder a ellos sin restricciones, se pueden reutilizar y redistribuir libremente.

La mayor parte de los datos abiertos son generados por el gobierno y las administraciones públicas, ya que reúnen y controlan una gran cantidad de datos que, además, deben de ser públicos por ley. De estos datos se benefician muchos colectivos y empresas, además del mismo gobierno⁴.

Podrán ser abiertos todos aquellos datos que cumplan las condiciones de privacidad y confidencialidad (que no haya restricciones legales, financieras o de derechos de autor que impidan su libre difusión), además deben tener las siguientes propiedades:

- Los datos deben ser publicados sin procesar, es decir en su forma original, con el mayor detalle posible.
- Tienen que estar actualizados, deben publicarse de manera frecuente para que estos datos representen la realidad.
- Deben estar disponibles en un formato que pueda ser procesado directamente por un ordenador. Por ejemplo, un formato PDF no sería válido, pero un CSV o XML sí, ya que se pueden leer con un lenguaje de programación.
- Siguiendo con el punto anterior, los formatos deben ser de un software libre. Es decir, no serían válidos formatos como Word o Excel ya que dependen de un software propietario para poderse leer.

Los datos abiertos están creando un campo de investigación social y económico del cual se esperan extraer nuevos conocimientos e ideas para encontrar nuevas aplicaciones en un futuro. Se espera en los próximos años una explosión importante de la cantidad de datos, ya sean públicos como privados disponibles de forma abierta. La apertura de datos fomenta tanto la transparencia como la construcción de nuevos modelos de negocio basados en el análisis de estos datos.

Los datos generados por ciudadanos, empresas, Administraciones públicas, y más recientemente por los objetos (como sensores, cámaras, ordenadores, etc.) han aumentado exponencialmente en los últimos años y la tendencia es muy fuerte. De este crecimiento

⁴ <https://www.data.gov/>

exponencial del volumen, de la variedad y la velocidad de generación de datos, además de la mayor capacidad de captura, almacenamiento, procesamiento y análisis, es de donde surge la denominada **Economía de Datos**.

Cada vez más, todo lo que sucede en la vida cotidiana de un ciudadano deja un rastro en forma de dato. Por ejemplo, cada vez que un usuario realiza una compra a través de un comercio online, se deja una huella en alguna base de datos: qué compró, cuánto pagó, cómo pagó, que productos compró, dónde estaba cuando realizó la compra, etc. De la misma manera que en el ejemplo, infinidad de información se almacena en forma de dato.

No solo tenemos estos datos que somos capaces de generar y captar a través de la transformación digital, sino que existen organizaciones y negocios que permiten utilizar sus datos para construir información o aplicaciones sobre ellos. De esta manera muchas empresas pueden obtener datos sobre el comportamiento de usuarios en plataformas digitales, por ejemplo, empresas financieras, aseguradoras, grandes distribuidores, supermercados o marcas de moda aprovechan estos datos para conocer mejor los gustos e interés y adecuar sus productos y ofertas.

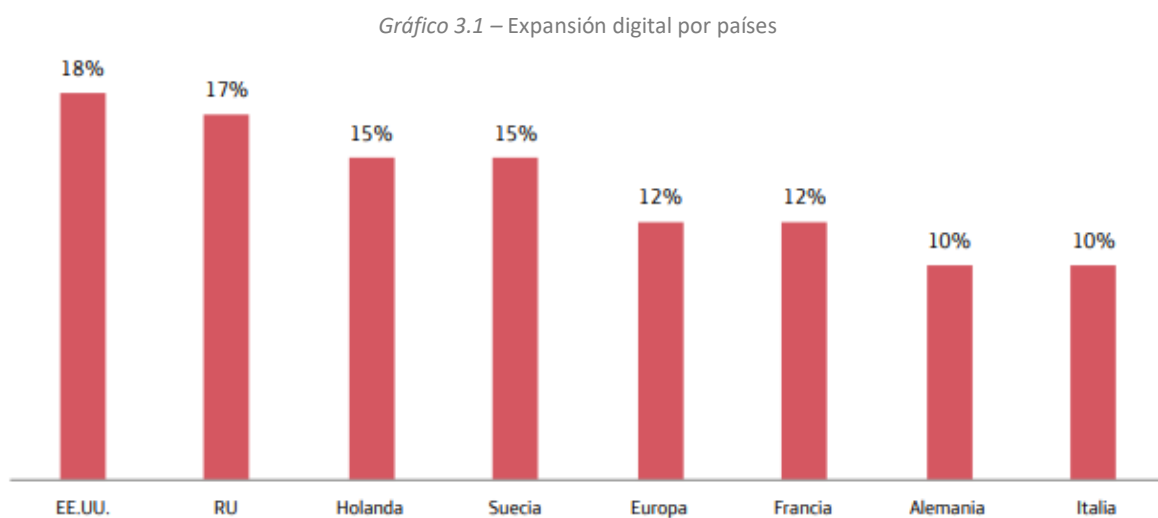
Los datos son valores que reflejan situaciones o hechos que se producen y permiten recrear o conocer una realidad. Un dato puede surgir de distintas fuentes y presentarse de diversas formas como un dato alternativo: una imagen de una cámara, una fotografía, un mensaje, una voz en una conversación, la temperatura, el grado de humedad, el número de unidades vendidas, la hora en que se entrega un pedido, etc. Estos datos pueden ser utilizados por un número ilimitado de agentes para un infinito número de aplicaciones para producir bienes y servicios, se consideran el equipamiento y estructura básica necesaria para que un país, región u organización funciones de forma adecuada.

En el ámbito de la economía, los datos se convierten en esenciales para el desarrollo de esta, en la medida que habilitan la innovación, la eficiencia de los procesos y la sofisticación de los bienes y servicios que se producen. Y es que una de las características de la época digital es la producción y el almacenamiento de datos e información a bajo coste, lo que alimenta aún más el círculo de la innovación.

Según José María Álvarez, presidente ejecutivo de Telefónica “los recursos más valiosos en el futuro serán los datos”. Según la empresa reseñada anteriormente, el volumen de datos en los próximos años se multiplicará por once, la velocidad de transferencia de información será cien veces la actual y la capacidad se multiplicará por cien.

El motivo principal de todos estos avances ha sido el abaratamiento de la computación, que ha sido exponencial en los últimos años, con el que la expansión de la Economía de datos se ha reforzado. Otro elemento clave ha sido el incremento de los dispositivos conectados a Internet, como teléfonos móviles, tabletas, sensores, relojes inteligentes, coches o incluso hogares conectados. Finalmente, el tercer elemento que favorece el desarrollo es la interconectividad social. Las redes sociales son un fenómeno que ha vivido una expansión global en los últimos años, no solo en número sino también en capacidad.

En el siguiente gráfico vemos el grado de aprovechamiento de esta expansión digital en diferentes países:



Fuente: McKinsey Global Institute Analysis.

Europa representa el promedio de los seis países que se muestran en el gráfico. Se puede observar que el grado de aprovechamiento es mayor en Estados Unidos que en Europa, y que en países como Holanda o Suecia hay un 50% más de aprovechamiento que en países como Alemania o Italia.

3.2. Importancia y actualidad de los datos abiertos

Como se ha comentado anteriormente, el dato se ha convertido en un recurso fundamental para cualquier proceso económico. De donde se extrae el valor de los datos es de su tratamiento, procesamiento y análisis, ya que el dato por si solo (en su estado bruto) carece de valor. Gracias a las innovaciones en Big Data y Ciencia de Datos se han conseguido leer los datos cada vez de mejor manera. Incorporando nuevas herramientas como lenguajes de programación o métodos científicos de datos, nuevas capacidades como el *Internet de las*

cosas y el *Cloud Computing*, y las nuevas destrezas profesionales en el ámbito digital son las que han hecho posible, que se generen modelos de negocio en torno al dato. El propósito de todos estos avances en la economía de los datos es extender la frontera del conocimiento y de las capacidades innovadoras tanto de las personas como de las empresas y Administraciones públicas, a partir de la explotación inteligente del dato.

La *economía de los datos* está favoreciendo el surgimiento de nuevos modelos de negocio que están reinventando la estructura de muchos mercados y sectores, aumentando la eficiencia en la producción y en la distribución de bienes y servicios. La innovación y la transparencia están rompiendo las barreras de ciertos mercados que se basaban en la información, fomentando una nueva organización de los negocios al aparecer nuevas empresas que impulsan una mayor competencia en el mercado.

Junto a estos nuevos modelos de negocio surgen nuevos retos alrededor del gran volumen de datos que hace falta gestionar. Ya que mientras un solo dato puede ser observado, analizado, comprendido y utilizado, estos hechos dejan de ser posibles cuando se habla de cientos, miles o millones de datos, a no ser que se utilicen nuevas disciplinas y herramientas. Entre estas herramientas hace falta distinguir principalmente estas dos:

- *Business Intelligence* o inteligencia de negocio: permite explicar el pasado, crear cuadros de mando con visualizaciones de la información o informes de errores.
- *Data science* o ciencia de datos: se basa en técnicas que, aparte de realizar un análisis descriptivo, permiten anticipar o predecir sucesos futuros.

El *open data* se ha convertido en una parte vital para la toma de decisiones y la investigación en Europa. Contabilizar los beneficios que produce es difícil ya que los más importantes y significativos son indirectos, lo que sí sabemos es que el hecho de abrir los datos de una institución u organización es un hecho de contribución vital para permitir una economía creciente, innovadora y ética como pilar para la democracia, seguridad y transparencia de la sociedad.

El tamaño del mercado de datos abiertos en 2019 era de 172.300 millones de euros en Europa (último dato oficial del que disponemos, según el portal de datos europeo), este dato nos lleva a pensar que el impacto de los datos abiertos representa una parte significativa de la economía global de los datos. Alrededor del 12% del valor de la economía de datos se crea directamente a partir de datos abiertos, y alrededor del 45% se ve afectado indirectamente por datos abiertos.

Se prevé que el crecimiento del tamaño del mercado de datos hasta 2025 será de una tasa de entre el -5.5% y el 4.6% desde 2000 y que en los último 5 años se mantendrá más estable entre

el 1,4% y el 2,9%. Si nos situamos en un escenario optimista de crecimiento podemos decir que el valor de los datos abiertos aumentará debido a la rápida expansión de nuevas aplicaciones y oportunidades de reutilización por parte del usuario como resultado de una mayor integración de la digitalización de la sociedad.

Dentro del crecimiento anterior, se prevé que el mercado de datos abiertos crezca entre un 4 y un 16% en los próximos años, el amplio rango de porcentaje de variación es debido a que se plantean varios escenarios de crecimiento.

Podemos hablar de tres posibles escenarios de crecimiento, según el portal de datos europeo⁵:

- **Escenario de referencia** (tasa de crecimiento del 7,5%): donde las tendencias de crecimiento seguirán de manera similar a la evolución actual.
- **Escenario de desafío** (tasa de crecimiento del 4,3%): en este escenario el mercado de datos abiertos crece de manera más pausada que en el escenario anterior. Nos situamos en un contexto económico menos positivo o en un bajo nivel de innovación de datos.
- **Escenario de alto crecimiento** (tasa de crecimiento del 15,7%): en esta última situación el mercado entraría en una trayectoria de crecimiento más rápida, donde habría un alto nivel de innovación de datos y un modelo de gobierno de datos abiertos y transparente.

No todos los países tienen el mismo potencial de crecimiento de datos abiertos, por este motivo es necesario diferenciar entre varios grupos:

- Países de alto crecimiento: con un crecimiento medio del 15.7%, en este grupo encontramos países como República Checa, Estonia y España.
- Países de crecimiento medio: con un crecimiento medio del 7.5%, se encuentran países como Austria y Noruega.
- Países de crecimiento bajo: con un crecimiento medio del 4.3%, en este grupo están países como Hungría o Islandia.

El crecimiento medio para la Unión Europea es de 10,413%, y una previsión de tamaño de mercado de datos abiertos de 334.200 millones de euros en 2025.

En la siguiente tabla podemos ver el crecimiento estimado del mercado de los datos abiertos para los próximos años bajo un escenario optimista:

⁵ <https://data.europa.eu/es>

Tabla 3.2 – Evolución del tamaño del mercado

Año	2021	2022	2023	2024	2025
<i>Tamaño del mercado en billones de euros</i>	224.86	248.28	274.13	302.68	334.20

Fuente: Portal de datos europeo

3.3. Regulación y protección de datos

La esencia de la regulación de los datos es la protección de la privacidad de información, generalmente mediante la prohibición o limitación de divulgación, o del uso indebido de información sobre personas físicas.

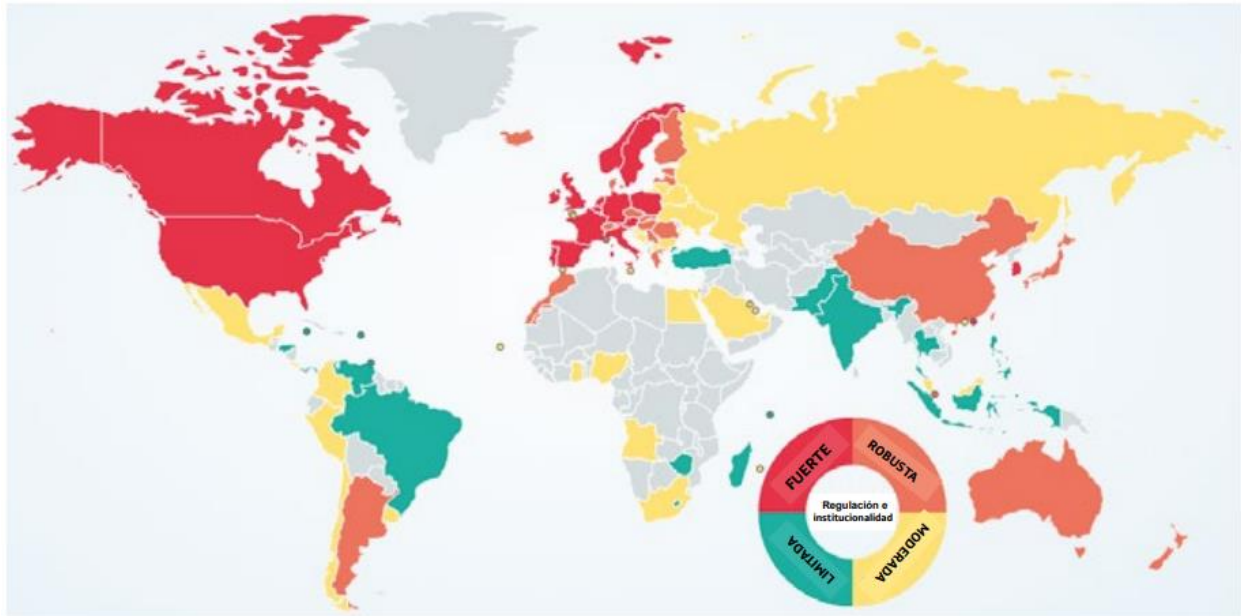
La recopilación y el almacenamiento de información personal por parte de autoridades (como el censo, identificaciones, huellas o declaraciones de impuestos), por parte de servicios de salud (como datos médicos o historiales clínicos, o de organizaciones privadas como entidades financieras (como gastos personales o transacciones de pago) contribuye a la necesidad de crear reglas de juego al tratamiento de datos personales.

Europa es la región del mundo en la que la protección de datos ha alcanzado un nivel más elevado, mientras que en Norteamérica lo que ha alcanzado un grado más alto de desarrollo es la privacidad. Un creciente número de países ha adoptado (o está en proceso) nuevas leyes de protección de datos y privacidad. Estas nuevas legislaciones contemplan en mayor o menor medida un conjunto de principios comunes en la regulación de los datos como:

1. El reconocimiento de la protección de datos como un derecho fundamental
2. La existencia de derechos individuales y exigibles en cuanto a la privacidad
3. La creación de una autoridad de control independiente

En el siguiente mapa se puede ver en color rojo los países con una mayor regulación de datos, y en verde los países con menos leyes y regulación de datos.

Mapa 3.3 – Regulación de datos en el mundo



Fuente: 2017 DLA Piper.

Se observa que en las regiones donde la revolución tecnológica avanza con más fuerza también lo hace la necesidad de regularizar los datos, así como Norte América o Europa, seguidas de regiones como Australia, China y Argentina.

En el ámbito europeo, la protección de datos está configurada como un derecho fundamental de los ciudadanos. Los principios y derechos básicos quedan recogidos en el Reglamento General de Protección de Datos (RGPD) de la Unión Europea, en vigor a partir de mayo de 2018 como normativa única, cuya elaboración empezó en 2012.

El RGPD aplica importantes sanciones económicas por el incumplimiento de la norma a las empresas, Administraciones públicas y organizaciones, pudiendo llegar a los veinte millones de euros o el 4% de la facturación global anual.

Las sanciones aplicadas en la Unión Europea pueden distinguirse en:

Figura 3.3



Fuente: Portal de datos europeo

Por otro lado, existe una normativa que prohíbe la transferencia de datos a países que no cuenten con un nivel adecuado de protección de datos, y establece un procedimiento para determinar formalmente si un país cumple con dicho nivel de protección.

La principal consecuencia de que un país sea declarado adecuado es que se podrán transferir datos desde los estados de la Unión Europea sin necesidad de ningún tipo de trámite o autorización especial. De momento han sido aceptados países como Suiza, Andorra, Argentina, Canadá, Estados Unidos, Nueva Zelanda o Uruguay.

En España, las agencias autonómicas de Cataluña y del País Vasco, ejercen las funciones de control respecto los ficheros de datos de carácter personal creados o gestionados por las comunidades autónomas y por la Administración local de su ámbito territorial. Sin embargo, los ficheros privados de estas comunidades autónomas son competencia de la Agencia Española de Protección de Datos.

Otro término importante que comentar es el de la soberanía de datos, esta se basa en el concepto de que la información creada o almacenada de forma digital está sujeta a las leyes y regulaciones del país en el que se encuentra dicha información.

El diferente desarrollo normativo de cada país ha creado un mapa global de soberanía de datos, dividido entre los países neutrales (actitud abierta) y los no neutrales, aquellos que consideran que la información de sus ciudadanos no debe ser almacenada ni procesada fuera de sus fronteras, estableciendo así una soberanía de datos. Esta mayor o menor neutralidad significa tomar la decisión de ubicar los centros de datos (infraestructuras y servicios) y trasladarlos al ámbito interno del perímetro de seguridad nacional.

El exponencial crecimiento del uso de servicios de Internet obliga a un crecimiento del tamaño de las infraestructuras de los centros de datos que soportan estos servicios. El almacenamiento en la nube ofrece un potencial simplificado que elimina las fronteras geográficas y físicas. Este *Cloud Computing*, es una de las tecnologías digitales consideradas más importantes en cuanto al aumento de la productividad y la mejora de la Economía de los Datos. Permite optimizar el uso de los recursos, así como crear modelos de negocio y estrategias de mercado en forma de Economía Digital. Así el almacenamiento en la nube presenta tres formas de contratación:

1. Infraestructura como servicio, que proporciona acceso a redes, a equipos y software dedicado al espacio y almacenamiento de datos.
2. Plataforma como servicio, permite centrarse en la implementación y administración de aplicaciones
3. Software como servicio o aplicaciones de uso final

3.4. Clasificación de datos abiertos

Los datos abiertos se suelen organizar en catálogos de datos, donde es más fácil encontrar una determinada base de datos en función de sus características. Estas características pueden ser la frecuencia de actualización de los datos, el formato, el agente que lo publica, el ámbito geográfico y otras etiquetas que puedan definir los datos.

Como hemos comentado en el punto anterior, una propiedad importante de los datos abiertos es que deben darse en un formato válido. Podemos clasificarlos en los siguientes formatos:

- CSV: Valores Separados por Comas, es un formato de texto plano donde los campos de valores vienen separados por comas o punto y coma. Podemos decir que es el formato más sencillo de usar, y además es una de las maneras en las que se puede exportar un documento de Excel (recordamos que este último no sería un formato válido).
- XML: Permite la interpretación de los datos para diferentes lenguajes, lo cual le hace ser el favorito para el intercambio de información estructurada entre diferentes plataformas. Hace falta añadir que muchas bases de datos permiten su exportación en este formato.
- RDF: Macro de Descripción de Recursos, está basado en XML y facilita la reutilización y modelización de los datos en recursos Web.
- KML: Esta basado también es XML, pero este se usa para representar datos geográficos en tres dimensiones. Muchos programas de análisis de datos permiten su lectura directa para la representación de regiones o localizaciones en un mapa.

Los formatos comentados son los más frecuentes de cada estilo, aunque hay mucha variedad de formatos parecidos.

Dejando de lado los formatos nos centramos otra manera de clasificar los conjuntos de datos. La escala más conocida de clasificación de datos abiertos es: Las 5 estrellas de Tim Berners-Lee. Esta escala clasifica los datos en 5 categorías en función de su grado de reutilización, obteniendo una estrella los de menor facilidad de reutilización.

- La primera estrella se alcanza con la simple acción de publicar los datos en cualquier formato. En este nivel se encuentran datos no estructurados o difíciles de leer y manipular.
- El nivel de dos estrellas lo tienen los datos que se encuentran en un formato con estructura pero que pertenecen a un software propietario, como Word o Excel.

- Se alcanzaría la tercera estrella cuando el formato de los datos estructurados use un software libre, como CSV y XML.
- El nivel de cuatro estrellas se alcanza cuando se usan URI's (URL's fijas) para identificar los datos de manera directa e individualizada.
- El máximo nivel de cinco estrellas lo tienen los datos que están relacionados con conjuntos de datos de otros usuarios u organizaciones, proporcionando así el contexto de estos. De esta manera se facilita la reutilización y creación de nueva información.

La mayoría de los datos abiertos publicados son conjuntos que pertenecen al nivel de tres estrellas, de tal manera que permiten a los usuarios disponer de la información en un formato de libre y sin limitación de uso ni difusión.

Si hablamos en concreto de España, en los últimos años se han creado iniciativas para ofrecer datos abiertos de mayor nivel, con lo que se han creado portales de datos que disponen hasta de cinco estrellas. Entre estas iniciativas podemos encontrar portales autonómicos, estatales y universitarios.

3.5. Beneficios de los datos abiertos

Los datos abiertos son una herramienta que permite reutilizar su información para usos muy distintos, de ellos se benefician tanto los ciudadanos, como las empresas y la administración. El agente que los publica como, por ejemplo, el gobierno, puede beneficiarse demostrando más transparencia frente a la ciudadanía proporcionando datos abiertos de las diferentes instituciones gubernamentales.

Por otro lado, en las plataformas de datos abiertos se puede crear una red de participación e interacción ciudadana, de clientes de empresas públicas o privadas. De esta manera se minimiza la desinformación y se mejora la eficiencia gubernamental, ya que se reducen las consultas y preguntas.

Hace falta destacar que los mayores beneficios de los datos abiertos se obtienen cuando estos se reutilizan para generar nueva información, crear modelos de negocios o nuevos servicios basándonos en los datos de las administraciones.

Uno de los mayores beneficios que producen los datos abiertos es la creación de empleo. Las empresas emergentes que utilizan los datos abiertos para su negocio están creando

directamente puestos de trabajo de datos abiertos, como analistas de datos o comercializadores del producto. Si estos negocios también necesitan un gerente de ventas para su empresa debido al crecimiento de esta, este puesto de gerente sería un trabajo creado indirectamente gracias a los datos abiertos.

En total, el mercado laboral directo e indirecto de datos abiertos se estima que en 2020 ha sido de entre 312.600 – 377.000 trabajadores, significativamente mayor al tamaño que tenía en 2016 de entre 235.700 – 284.900, según el portal de datos europeo.

Como hemos visto el potencial de creación de empleo mediante la publicación y reutilización de datos abiertos es muy significativo, y dentro de este empleo hay diferencias entre los empleados que trabajan directamente con los datos abiertos y los que trabajan indirectamente debido a ello. Los empleados directos de datos abiertos son persona que están generando, proporcionando, agregando y reutilizando datos abiertos. Estos empleados son empleados directos de datos abiertos ya que no habrían existido sin datos abiertos. En cambio, los empleados indirectamente debido a los datos abiertos son personas que están tratando datos abiertos en organizaciones que no basan su negocio o enfoque central en los datos abiertos.

En 2017 había 20.229 empleados de datos abiertos en España, 229.000 empleados directos en Europa y 2.019.833 mil personas indirectamente empleadas. Es decir, había un total de 1,06 millones de empleados de datos abiertos en Europa.

En la siguiente tabla se puede ver la evolución prevista del número empleados de datos abiertos para los próximos años:

Tabla 3.5 – Previsión de la evolución de trabajadores de datos abiertos.

Año	2021	2022	2023	2024	2025
<i>Número de trabajadores de datos abiertos (en miles)</i>	1 100	1 105	1 111	1 116	1 122

Fuente: Portal de datos europeo

En todos los países de la Unión Europea existe un portal nacional de datos abiertos. Estos portales están configurados y mantenidos por un equipo de trabajadores de datos abiertos. Estos equipos además de ser responsables del portal también llevan a cabo la coordinación de iniciativas de datos abiertos, la creación de materiales para publicar y reutilizar datos abiertos, la provisión de formación a funcionarios públicos y la promoción de la publicación y

reutilización de datos abiertos. En promedio hay 4 empleados por equipo nacional trabajando directamente debido a los datos abiertos.

Otro beneficio que destacar de los datos abiertos es la optimización del transporte, así como del tiempo de espera de los ciudadanos. La información que se proporciona a través de dispositivos personales disminuye el tiempo real de espera y el tiempo que las personas perciben que han estado esperando. Gracias a la información en tiempo real disponible se reduce el tiempo de espera aproximadamente 2 minutos de media por persona y un 30% menos de tiempo de espera percibido. Los datos abiertos pueden ahorrar potencialmente aproximadamente 27 millones de horas anuales para los usuarios de tren en Europa, lo que vienen a ser 1.600 millones de minutos ahorrados por todos los usuarios.

A parte del transporte público se optimizan los desplazamientos en transporte privado y se reducen los flujos de tráfico. Los datos de tráfico en tiempo real pueden ayudar a encontrar rutas alternativas menos congestionadas, tomar estas rutas alternativas no solo ayudan a llegar de manera más rápida al destino, sino que podría ayudar a disolver el atasco a medida que los coches se redistribuyen a través de carreteras alternativas. Los datos de tráfico abiertos en tiempo real pueden reducir el tiempo que un conductor pasa en el tráfico entre 4,9 y 7,1 horas al año en las ciudades de Europa. Entre el 60% y el 70% de los desplazamientos desde y hacia el trabajo se realizan en coche, de esta manera en total son aproximadamente 500-730 millones de horas potencialmente ahorradas cada año por conductores europeos que van y vuelven del trabajo en las zonas urbanas.

Por otro lado, las aplicaciones abiertas basadas en datos pueden servir como una herramienta para ayudar a los hogares a reducir su consumo de energía, proporcionándoles informes y sugerencias sobre cómo disminuir el consumo de energía. La reducción media en 2019 fue del 2% o de 0,62 KWh al día. Las herramientas basadas en datos abiertos podrían ahorrar potencialmente 5,8 millones de toneladas de petróleo al año al ayudar a reducir el consumo de energía de los hogares.

Por último, hace falta destacar que los datos abiertos pueden ayudar a salvar vidas, cada año aproximadamente se realizan 293 millones de llamadas de emergencia en toda Europa, de las cuales 141 millones se hacen al número europeo de emergencias "112". Los datos abiertos pueden ayudar a los servicios de emergencia a conseguir la ubicación del incidente más rápido. Para optimizar los servicios de emergencia, se pueden utilizar varios tipos de datos abiertos como: abrir datos de los números de emergencia, abrir datos de mapas y abrir datos de tráfico a tiempo real. Se estima que en 2020 ha habido 7.000 vidas potencialmente

salvadas, de las cuales 1.425 has sido salvadas debido a la disminución de muertes en carreteras.

Como se ha comentado anteriormente, la revolución digital que supone el desarrollo de la Economía de Datos ha ejercido un gran impacto sobre el empleo. Esta revolución digital abre un importante y variado abanico de oportunidades laborales que es conveniente aprovechar. Uno de estos empleos emergentes es el de director de datos (*data manager*), el máximo responsable de los equipos especialistas en Big Data dentro de las empresas lidera la gestión y el análisis de los datos, también es su función asegurar la calidad de los datos y la generación de valor en los distintos niveles ejecutivos.

El segundo empleo potencial es el de ingeniero de datos (*data engineer*), es la ocupación que más está aumentando recientemente. El ingeniero de datos es la persona que se encarga de la identificación, la limpieza y la clasificación de los datos que serán analizados e interpretados por la figura de los científicos de datos. Su misión es preparar la información para su posterior procesamiento y puesta en valor.

El científico de datos (*data scientist*) es otro empleo muy buscado actualmente, este es el encargado de convertir importantes volúmenes de datos en información útil para la empresa, mediante técnicas de análisis avanzadas que permiten detectar y predecir el comportamiento de los consumidores. Su misión es analizar datos para extraer información de valor que ayude a la empresa a tomar decisiones.

Por último, hace falta destacar el puesto que ocupa el director de seguridad de la información o el encargado de la protección de datos (*data protection officer*), ya que son unos empleos de los que se prevé que más van a crecer en los últimos años. El director de seguridad es el responsable de proteger la información digital de los sistemas, su misión consiste en diseñar y ejecutar el plan de ciberseguridad para proteger la información procesada, almacenada y transportada por los sistemas, así como el tratamiento de amenazas.

3.6. Reutilización de datos abiertos

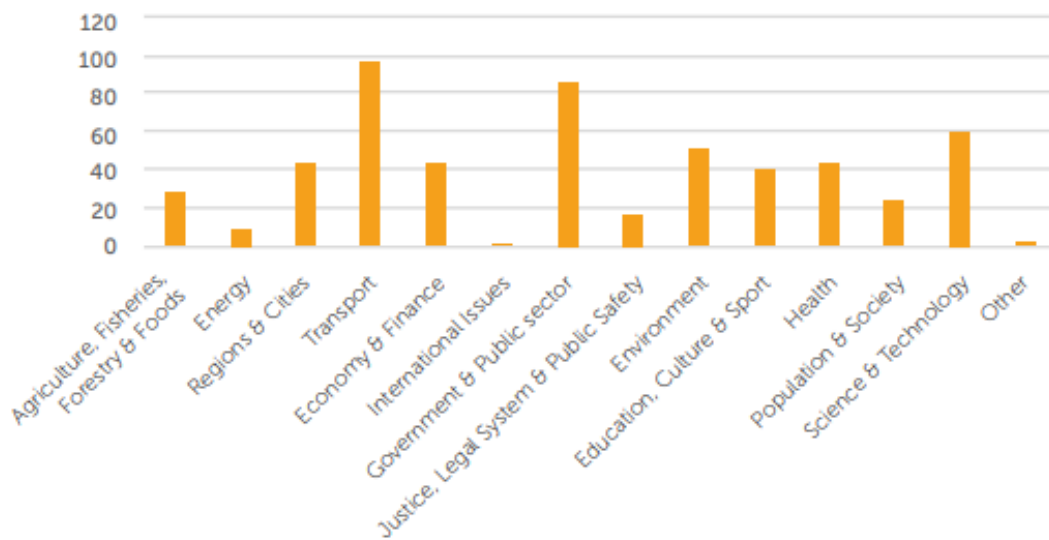
La reutilización de los datos abiertos consiste en utilizar información publicada por la administración, el gobierno o empresas del sector público con el fin de generar nueva información de interés, ya sea comercial, económico, educativo, etc.

El principal objetivo de los datos abiertos es la Reutilización de la Información del Sector Público (RISP), permitiendo su reutilización tanto a particulares como a empresas.

Entre los sectores empresariales y económicos se distingue un grupo de sectores que está situado en un escenario de mayor contribución y crecimiento en cuanto a la reutilización de datos (alrededor del 15,7%). En este grupo anterior se distinguen dos subgrupos, los “sectores de alto impacto” con creación probada y exitosa de impacto de datos abiertos y los “sectores de alto potencial”, que tienen un alto potencial para aprovechar datos abiertos y así crear más valor. Dentro de los sectores de alto impacto encontramos sectores como la Administración Pública, actividades profesionales, científicas y técnicas, información y comunicación, o transporte y almacenamiento. Por otro lado, en los sectores de alto potencial, se sitúan el sector de la agricultura, servicios financieros y aseguranzas, salud, educación, comercio o actividades inmobiliarias.

En el siguiente gráfico podemos ver el número de casos de estudio de reutilización de datos disponibles en los portales de datos abiertos europeos, distinguimos por sectores económicos:

Gráfico 3.6.1 – casos de estudio de reutilización de datos abierto

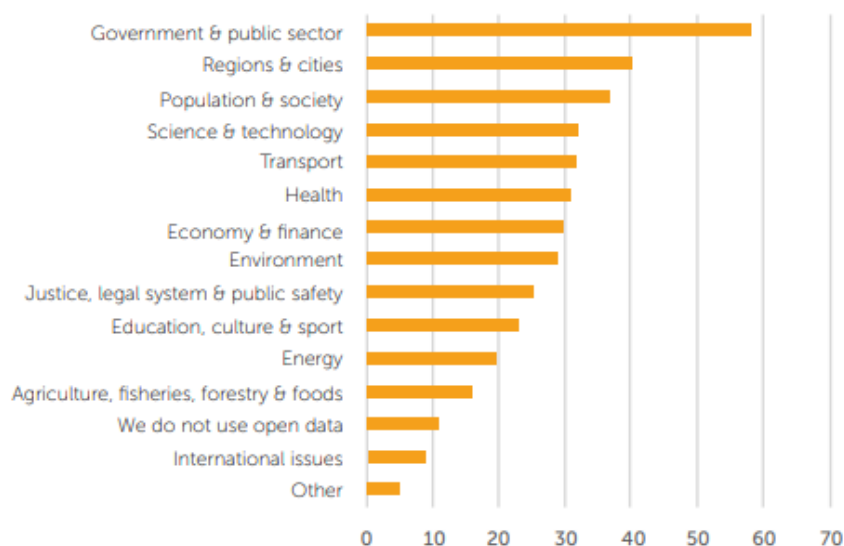


Fuente: Portal de datos europeo

Se observa que los sectores con más casos de reutilización de datos son con gran diferencia: el sector del transporte y el sector público y de gobierno. Por otro lado se aprecia que hay una reutilización de datos abiertos inapreciable en los sectores de la energía, cuestiones internacionales o en la justicia y seguridad pública.

Podemos comparar el gráfico anterior con el siguiente en el que se visualizan las categorías de datos abiertos más utilizadas. Este gráfico se basa en los datos de una encuesta realizada en 2017:

Gráfico 3.6.2 – Datos abiertos más utilizados



Fuente: Portal de datos europeo

Podemos ver que la categoría de datos abiertos más usada es la del sector público y gubernamental, esta también es una de las categorías con más casos de estudios de reutilización en los portales europeos. En cambio, vemos que el sector con más casos de reutilización pasa a ser la quinta categoría más usada, y se coloca por detrás de categorías como la población y sociedad, o como la ciencia y la tecnología.

Las organizaciones a la hora de llevar a cabo sus proyectos de reutilización de datos deben definir el problema que se quiere solucionar con los datos. Esta tarea requiere el suficiente conocimiento de negocio como para establecer la organización necesaria y estimar las ganancias que se van a obtener. Por otro lado, también se exige un conocimiento técnico para saber transformar un problema de negocio en un caso de trabajo, es decir en un problema científico viable, donde basándonos en la evidencia de los datos y utilizando herramientas analíticas, estadísticas y matemáticas, poder extraer conclusiones sólidas, proponer acciones y predecir comportamientos futuros.

Estas tareas comentadas, con un componente de conocimiento del negocio y ciencia, es lo que se conoce como ciencia de datos (Data science). Los científicos de datos, uno de los perfiles profesionales más demandados hoy en día en el mercado laboral, son los encargados de llevar a cabo estas funciones. Con un conocimiento que mezcla aptitudes matemáticas, estadísticas, informáticas y de negocio, ellos son los responsables de abordar proyectos y desarrollos que tienen partes muy diferenciadas:

- **Identificación del problema:** planteamiento de la pregunta que necesita ser contestada para resolver un problema concreto del negocio.
- **Obtención de la información:** estudiar qué fuentes de datos son las relevantes para solucionar el problema.
- **Explotación y análisis de la información:** mediante técnicas de análisis estadístico y de visualización de datos, buscar respuestas desde un punto de vista descriptivo del problema.
- **Modelización de los datos:** Crear modelos matemáticos que basados en millones de registros que permitan descubrir patrones de comportamiento, predecir cómo actuarán nuestros clientes y usuarios, o desarrollar sistemas de recomendación, entre otros.
- **Comunicación de resultados:** Traducir los resultados al lenguaje de negocio, de manera que sea comprensible por los que tomarán las decisiones en las organizaciones.

En definitiva, la *Data Science* es el conjunto de técnicas que nos permiten pasar de un almacenamiento de datos a unas aplicaciones de valor con los mismos. Estos modelos ya están presentes en nuestras vidas, como cuando encendemos el navegador y estima el tiempo que tardaremos en llegar a nuestro destino basándose en el pasado y en datos de tráfico, existe un modelo matemático detrás de esto. Estamos rodeados de algoritmos y modelos que nos facilitan la toma de decisiones, y que nos conocen gracias a la huella de datos que dejamos, tanto en el mundo virtual como en el físico.

3.6.1. Proceso de reutilización

Para llegar a sacar valor de la reutilización de datos hace falta que estos pasen previamente por muchas fases de almacenamiento, limpieza y tratamiento, entre otros. Las fases por las que pasa un dato desde que se recoge hasta que se saca valor de él son las siguientes:

1. Fase de provisión del dato: se divide en dos subfases

- 1.1. **Aprovisionamiento y abastecimiento:** en esta etapa se identifican aquellos conjuntos de datos que serán necesarios para obtener la información que queremos. Para ello haría falta realizar un inventario de la información que genera la empresa y que se está almacenando, y de la misma manera hacer inventario para aquella información que se genera, pero no se está almacenando y que pudiera ser

potencialmente útil. Finalmente se completaría el inventario con información externa a la organización, a través de terceros o de fuentes abiertas de datos.

- 1.2. **Captura y almacenamiento:** en esta etapa se realizan las tareas necesarias para extraer la información que se ha seleccionado en el paso anterior y para integrarlos en el sistema, listos para ser procesados y almacenados. Existe una gran diversidad de maneras de captar la información dependiendo del tipo de datos, de la frecuencia con la que se tienen que capturar o la exactitud y calidad de estos.
2. Fase de transformación: Una vez que los datos están almacenados, se procede a limpiarla y enriquecerla con el objetivo de tener información precisa y de calidad, de forma que las decisiones basadas en dicha información se sustenten en información proveniente de fuentes fiables. El proceso para mejorar la calidad del dato incluye las siguientes tareas:
 - 2.1. **Perfilado:** consiste en realizar un análisis general de los datos después de cada transformación, con el objetivo de determinar posibles fallos respecto a la información ya cargada previamente, de tal forma que se puedan detectar posibles anomalías en los datos.
 - 2.2. **Limpieza:** en este paso se garantiza que cada registro proporciona un valor válido en cada campo. Para ello se llevan a cabo las siguientes tareas:
 - Estandarización: consiste en asegurar que los datos del mismo tipo (de cada campo) tienen el mismo formato, a fin de que luego puedan ser comparados o cruzados. Por ejemplo, que todas las fechas estén en formato dd/mm/yyyy.
 - Normalización: consiste en asignar un valor único que se pueda asignar a diferentes valores que representen la misma idea. Por ejemplo, que todos los campos que determinen el sexo de un individuo, usen H para hombre y M para mujer.
 - Matching: consiste en la eliminación de duplicados con el objetivo de mejorar la exactitud de la información.
 - 2.3. **Estructuración:** Si se considera necesario, se reestructuran los conjuntos de datos, como la división de un campo que recoge información del nombre y apellidos en 3 campos diferentes.
 - 2.4. **Enriquecido:** tarea de enriquecer el conjunto de datos con nuevos campos que aporten información. Por ejemplo, a la tabla de clientes podríamos añadirle campos de información relativa obtenida en modelos analíticos, como la probabilidad de compra de un cierto producto.

3. Fase de descubrimiento y modelado: se realiza el proceso de análisis por parte de los analistas de negocio y los científicos de datos con el objetivo de extraer el máximo valor de los datos en información.
4. Fase de exposición: Es la fase final de la cadena, se realizan todas las tareas para poner el nuevo modelo en producción, puesto que no se genera valor sino se comparte el conocimiento extraído con los usuarios del negocio. Hay que tener en cuenta que en las fases anteriores se ha trabajado de tal forma que se han invertido los recursos mínimos necesarios para que en el caso de que no se extraiga valor, la pérdida sea mínima también. Por lo tanto, en esta fase es donde se realizan las tareas de validación, de preparación del espacio para albergar el modelo de salida, la documentación del proceso, etc. En resumen, se industrializa el proceso hecho anteriormente.

3.6.2. Importancia de la reutilización para empresas y negocios

Dependiendo del sector de trabajo las empresas pueden beneficiarse de diferentes tipos de datos abiertos. Por ejemplo, en el sector de la agricultura interesan datos meteorológicos y de la tierra, precios del mercado, control de plagas, demanda del consumidor, etc. Por otro lado, un negocio que se dedica a la venta de un determinado producto puede beneficiarse de datos que definan la necesidad de dicho producto en las diferentes regiones del país (por ejemplo, si el producto es una bebida interesa ver donde se consumen más bebidas parecidas a esta y donde hay menos oferta), interesaría estudiar las características del producto que son más importantes para los clientes, los precios del mercado en cuestión, etc. Estos datos proporcionan información vital con la que se puede conseguir una mejora de la productividad, economía y desarrollo de la empresa.

En definitiva, un negocio que dispone de información de su sector de actividad como, los precios de venta del producto, demanda del consumidor, precios de materias primas, tasa de empleo en el sector, etc. Puede mejorar y crear una imagen del negocio adaptada a la realidad, es decir introducirse en el mercado con unos precios competentes y un producto de calidad que cumpla las exigencias del cliente; así como de cara a la contratación de empleados, colocarse en buen lugar en cuanto a condiciones de trabajo siendo conocedor de la oferta laboral actual.

Los proyectos de reutilización de datos que se abordan en las empresas pueden tener diferentes objetivos, pero los principales son los siguientes:

- **Conocimiento de los clientes y usuarios:** analizando los datos de los clientes, podemos dirigir estrategias, aumentar ventas y fidelizar dichos clientes. Dependiendo del negocio, esto toma diversas formas: si el negocio es digital, observar la navegación de los usuarios o los contenidos que visitan, entre otros; si se dispone de tiendas físicas y hay capacidad de medir por donde pasan los clientes y qué producto acaban comprando, se podrá cambiar la estrategia y aumentar las ventas.
- **Reducción de costes:** los datos internos de las empresas permiten descubrir ineficiencias en los procesos y la organización. Un caso que ayuda a optimizar los costes es la predicción de la demanda que va a tener una empresa en diferentes plazos de tiempo. Si la empresa conoce anticipadamente la demanda que va a tener, podrá ajustar los servicios logísticos y de mantenimiento, ahorrando costes innecesarios y aumentando el grado de cumplimiento del servicio que otorga a sus clientes.
- **Creación de nuevos productos y servicios de información:** la huella que dejan los usuarios en las bases de datos de las empresas puede permitirles dar recomendaciones de comportamiento a sus clientes con el objetivo de optimizar sus gastos o mejorar sus condiciones. Un ejemplo claro son los portales inmobiliarios, que pueden recoger información de los intereses de los usuarios y así saber qué tipo de vivienda está más demandada. Juntando esta información con el precio de muchos inmuebles de diferentes zonas, pueden ofrecer un servicio de tasación automática según sus características y la zona donde se encuentran.
- **Nuevos negocios:** los datos de las empresas son útiles para terceras empresas con negocios completamente diferentes, por lo que pueden crear nuevas líneas que sean una fuente de ingresos basados en los datos.

En el siguiente cuadro podemos ver los problemas más frecuentes abordados con análisis de datos en cada sector o industria:

Tabla 3.6 – soluciones con análisis de datos según sector

	Manufacturas y recursos naturales	Medios de comunicación	Servicios	Gobierno	Educación	Retail	Banca	Seguros	Salud	Transporte	Servicios públicos
Mejorar la experiencia de usuario	52%	78%	66%	43%	76%	83%	77%	77%	73%	69%	44%
Eficiencia de procesos	45%	33%	35%	49%	65%	43%	41%	50%	73%	69%	78%
Marketing	43%	89%	53%	17%	41%	78%	66%	58%		38%	17%
Reducción de costes	42%	33%	35%	37%	35%	30%	41%	31%	45%	56%	61%
Mejora en gestión de riesgos	14%	22%	29%	29%	35%	22%	52%	58%	55%	31%	61%
Nuevos productos	23%	67%	37%	14%	24%	35%	27%	50%		19%	33%
Desarrollo de información de productos	26%	33%	44%	31%	12%	22%	23%	19%	9%	19%	11%
Mejora de las capacidades de seguridad	17%	22%	21%	34%	29%	13%	27%	27%	9%	19%	28%
Cumplimiento normativo	11%	22%	18%	23%	18%	9%	25%	23%	27%	31%	44%

Fuente: Gartner, «State of big data adoption», 2015, disponible en [/www.slideshare.net/denisreimer/big-data-industry-insights-2015](http://www.slideshare.net/denisreimer/big-data-industry-insights-2015)

Como se puede ver en el cuadro, el problema más abordado es el mejorar el Marketing de la empresa en el sector de los medios de comunicación, seguido de otros como mejorar la eficiencia de procesos en los servicios públicos o mejorar la experiencia de los usuarios en el sector de la banca, los seguros, la salud, el transporte y la educación.

Según distintos informes, en España, el 84,4% de las compañías está realizando proyectos de análisis de datos o tiene planes inminentes de hacerlo. El 70% de las empresas financieras declara que ya han hecho este tipo de proyectos, enfocados principalmente en el conocimiento profundo de los clientes, la mejora de sus productos, el análisis de riesgo o la detección de fraude de diferentes tipos. En el caso del sector de telecomunicaciones, aproximadamente el 60% de las organizaciones está realizando proyectos de análisis de datos para analizar la movilidad de sus usuarios, entender como están conectados, mejor sus campañas de marketing o realizar comunicaciones personalizadas. En otros sectores como la logística, la adopción del análisis de datos supera ampliamente el 50%.

Con el aumento de la reutilización de datos abiertos resultan muchos beneficios como hemos visto anteriormente. Para una empresa el hecho de incorporar productos y servicios junto a una mejor calidad de datos abiertos puede mejorar la satisfacción del usuario, proyectar mayor confianza y tener una mejor imagen como proveedor de datos abiertos.

Hace falta destacar que solo el 12% de las empresas utilizan Open Data como principal fuente de información para desarrollar aplicaciones. El beneficio más extendido e importante para las empresas que reutilizan datos abiertos es sin duda la innovación que pueden conseguir con la información que se extrae. Otros beneficios también muy destacados son la reducción de costes y el aumento de la eficiencia en las gestiones y tareas. Se estima que se han ahorrado

potencialmente 739,8 millones de euros debido al tiempo ahorrado, entre 13.700 y 20.000 millones de euros reducidos en costes laborales, por último, se han ahorrado 79.600 millones de euros en facturas de electricidad.

El desarrollo de la economía de datos ha proporcionado también una motivación para la creación de empresas que analizan y procesan información y productos de valor añadido para terceros o para el público.

En 2017 se publicó un informe que implica el análisis de 636 empresas cuya actividad se basa en la reutilización de información pública y privada para desarrollar productos para terceros y ciudadanos. Estas empresas aportan una facturación total de 1.700 millones de euros anuales y emplean a unas 19.362 personas. El mayor crecimiento se registró en los sectores económicos y financieros, con 219 empleados más que en 2014, seguido por la industria de la información geográfica, con 215 empleados más. El estudio mostró que la facturación media por empresa era de 2,68 millones de euros.

3.6.3. Negocios y aplicaciones basadas en datos

Las industrias data-driven son empresas basadas en la información, donde los datos son el epicentro de los procesos y de la toma de decisiones. Las características de estas empresas son las siguientes:

- Basadas en una política de analítica de datos
- Intensivas en uso de herramientas innovadoras de análisis de datos y de seguimiento y control de indicadores clave de la empresa.
- Desaparición de las jerarquías tradicionales, y fortalecimiento de la comunicación directa entre dirección y empleados para una toma de decisiones ágil.
- Evaluación por objetivos.

El sector financiero y de seguros es un claro ejemplo de la aplicación intensiva de los principios básicos de la Economía de Datos. La tecnología está impulsando rápidamente la transformación del sector de los servicios financieros, con potencial de aumentar la competencia, la innovación y la eficiencia.

Gracias a los datos abiertos se han desarrollado muchas aplicaciones de interés social y económico. Su objetivo es proyectar una mejor visualización de la información y permiten al usuario interactuar con los datos para una visión más detallada de los mismos.

Un sector importante y en el que la población está interesada es el meteorológico, existen infinidad de aplicaciones que permiten ver el tiempo o una previsión de este a tiempo real basándose en datos abiertos.

Otro ejemplo muy conocido es Google Maps, que ofrece mapas de todos los países vía satélite, puede marcar la ruta para llegar a un determinado destino, conociendo el tráfico o los horarios de tren y autobús entre otros.

Un último modelo de aplicación muy usado es de las aplicaciones de destinos turísticos, como la búsqueda y comparación de hoteles y apartamentos, los cuales se basan en la opinión de personas que se han alojado en ellos y han valorado sus servicios y características.

3.6.4. Los datos abiertos en las administraciones públicas

Las administraciones públicas tienen la gran labor de publicar una enorme cantidad de datos abiertos, lo cual beneficia a muchas entidades e individuos, pero también a ellas mismas ya que pueden reutilizar dicha información.

Estas administraciones usan los datos abiertos para mejorar la eficiencia de sus procesos y gestiones, así como para la toma de decisiones. Basándose en la información se puede llevar a cabo una gestión más eficaz de tareas y recursos públicos (ayudas, subvenciones) e incluso prever hechos y catástrofes como incendios o inundaciones. Un ejemplo es la modelación del turismo en cierta zona para optimizar el funcionamiento del transporte público.

La reutilización de información por parte de las Administraciones públicas ha hecho que más ciudadanos muestren confianza hacia el uso de servicios públicos basados en datos. Se demuestra que la reutilización permite crear servicios de mayor calidad y adaptados a los usuarios. Un ejemplo para el transporte público es la creación de aplicaciones en las que se puede consultar la llegada de los autobuses y trenes a tiempo real.

En los últimos años, a medida que se ha visto la posibilidad de mejorar los servicios utilizando información de datos abiertos, se ha promovido la importancia de actuar dentro del ámbito público basándose en datos. Todavía esta forma de optimizar las labores de una empresa está

mucho más extendida en el ámbito privado que en público, en este último aún falta mucho para tener los datos como base de la mejora y creación de servicios para población.

En el siguiente gráfico donde vemos el tamaño del mercado de datos abiertos en función del sector económico, es necesario destacar el tamaño del sector de la administración pública:

Gráfico 3.6.3 – Tamaño del mercado por sector económico



Fuente: Portal de datos europeo

La administración pública abarca un mercado de 22.111 millones de euros. Este valor es más del doble del tamaño de mercado del segundo sector con más influencia de datos abiertos: el sector industrial.

Se estima que en el 2020 el sector público gracias a la apertura de datos ha ganado un valor de 22.000 millones de euros. Un beneficio importante de la apertura de datos por parte de las administraciones públicas es el ahorro de costes que se puede generar abriendo datos, se estima un ahorro de costes de 1.700 millones de euros para 2020 en Europa.

Otro principal beneficio a la hora de abrir datos es el potencial de Open Data para permitir una mejor toma de decisiones políticas.

IV. PLATAFORMAS DE VISUALIZACIÓN DE DATOS

4.1. Descripción e importancia

Las plataformas de visualización de datos son una herramienta de inteligencia empresarial cada vez más usada en muchas empresas. Estas plataformas permiten analizar datos para generar información de interés, crear informes y de esta manera poder valorar la situación de la empresa y tomar decisiones en función de los resultados.

Para la directiva de una empresa es muy importante poder crear aplicaciones donde visualizar de manera clara la información, como por ejemplo la producción de los diferentes departamentos, poder compararla con los meses o años anteriores, ver los clientes más potentes en cada momento, etc. También es relevante el hecho de poder interactuar con estos datos, es decir, poder filtrar la variable temporal (si queremos ver los datos del último mes, el último año o los últimos días) así como seleccionar clientes, regiones o servicios, dependiendo de la información que se visualice.

Estas plataformas son la herramienta idónea para construir aplicaciones basadas en datos abiertos. Nos permiten crear los gráficos, informes, indicadores, etcétera, más adecuados para representar los datos de interés. De esta manera tendremos una aplicación interactiva que analice un conjunto de datos abiertos, así será mucho más rápido y sencillo acceder y comprender la información que nos proporcionan.

Además de las dos plataformas de análisis y visualización de datos que se verán al detalle en este trabajo, existen otras potencialmente útiles y bien valoradas en el sector de la estadística, como:

- **Tableau:** Plataforma estadounidense creada en 2003, es una de las más conocidas y usadas en todo el mundo. (<https://www.tableau.com/>)
- **Power BI:** Herramienta de Microsoft creada en 2011, esta ofrece una total compatibilidad con otras plataformas de la compañía como Excel, SQL Server, etc. (<https://powerbi.microsoft.com/>)
- **Google Data Studio:** Plataforma muy parecida a las dos anteriores con el beneficio que se ejecuta totalmente en la nube. (<https://datastudio.google.com/>)
- **Infogram:** Nacida en Letonia en 2012, plataforma muy sencilla y completamente gratuita. (<https://infogram.com/>)

4.2. Descripción y posibilidades de Shiny y Amazon QuickSight

En este trabajo vamos a tratar con 2 plataformas de visualización de datos. La primera de ellas es Shiny, una librería del conocido programa R Studio que permite crear aplicaciones interactivas de análisis de datos. Esta primera plataforma es de software libre i totalmente gratuito, simplemente hace falta instalar la librería “Shiny” en el mismo R Studio para poder usarla. Un impedimento para crear una aplicación con Shiny puede ser el desconocimiento de la programación con R, ya que las aplicaciones se crean mediante programas que debe crear el propio usuario del software, de esta manera si no se tiene un cierto nivel de programación sería imposible crear una aplicación de esta manera.

Vamos a ver un ejemplo de programa de Shiny para ejecutar una aplicación, este ejemplo es propio de Shiny y se puede ver ejecutando en R este código:

```
library(shiny)
runExample("01_hello")
```

El código de R para crear una App con Shiny se compone de dos partes, la primera es una secuencia de comandos de interfaz de usuario (archivo ui.R) que controla el diseño y aspecto de la aplicación, y la segunda es una secuencia de comandos de servidor (server.R) que contiene las instrucciones necesarias para construir la aplicación.

En el ejemplo “01_hello” el ui.R es el siguiente (en los comentarios “#” se explica que hace cada opción):

```
library(shiny)

shinyUI(fluidPage(

  titlePanel("Hello shiny!"), #define el título del panel

  sidebarLayout(
    #se crea un diseño lateral
    sidebarPanel(
      #con el diseño lateral se crea un panel lateral
      sliderInput("bins",
        "Number of bins:",
        min = 1,
        max = 50,
        value = 30) #en el panel lateral un selector deslizable
    ),

    mainPanel(
      #panel principal (fuera del panel lateral)
      plotOutput("distPlot") #grafico output que se llama "distPlot"
    )
  )
))
```

Y el server.R es:

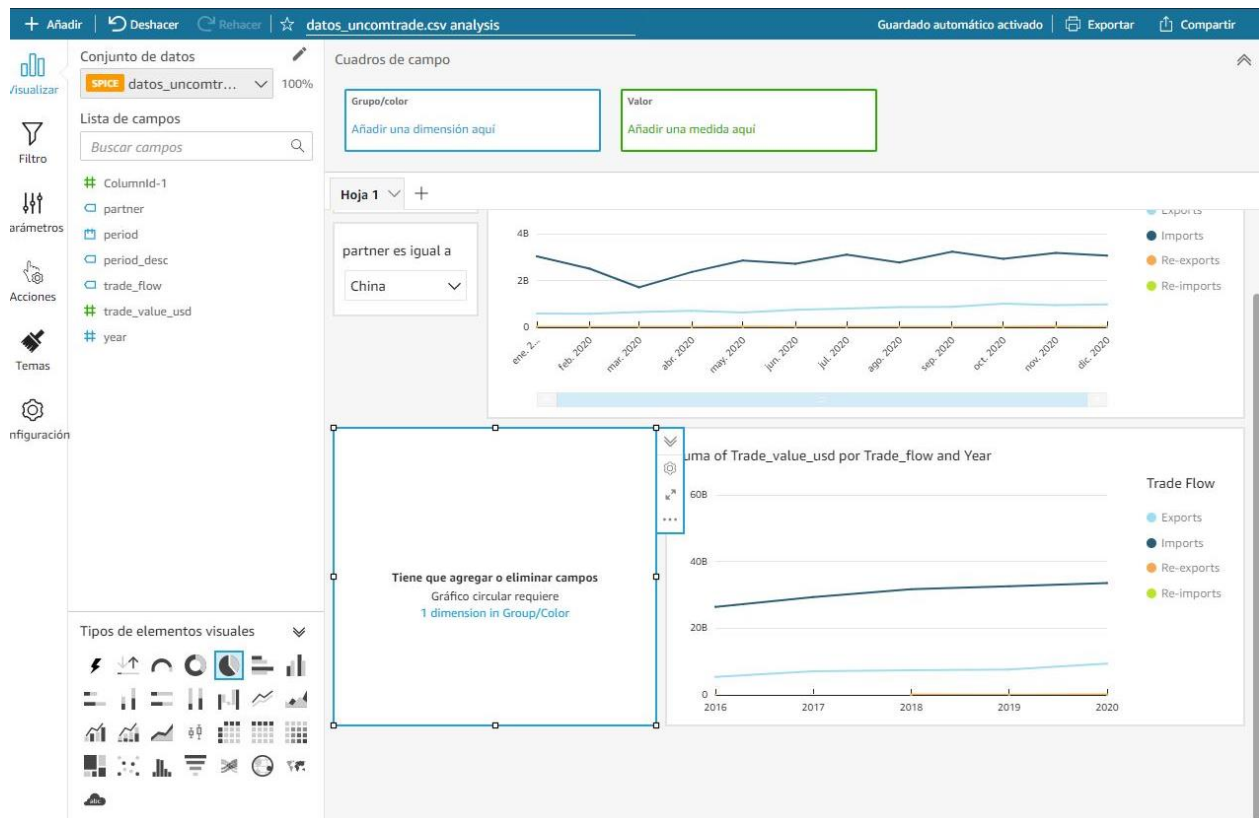
```
shinyServer(function(input, output) {  
  #se define el gráfico de output "distPlot" que se ha visto en el ui.R  
  output$distPlot <- renderPlot({  
    x <- faithful[, 2]  
    bins <- seq(min(x), max(x), length.out = input$bins + 1)  
  
    #en este caso es un histograma:  
    hist(x, breaks = bins, col = 'darkgray', border = 'white')  
  })  
})
```

Como se ha visto antes, solamente se necesitan conocimientos de programación de R para crear una aplicación con Shiny.

La segunda plataforma que se verá es Amazon QuickSight, esta pertenece a Amazon como su nombre indica y es de pago. Hace falta registrarte en Amazon Web Services y formalizar una cuenta para poder usar la aplicación. Esta aplicación parece mucho más sencilla e intuitiva de usar, ya que no hace falta saber ningún lenguaje de programación, además se pueden ver todos los objetos de visualización disponibles (gráficos, tablas, etc.) en el mismo espacio donde se creará la aplicación. De esta manera es más sencillo encontrar el gráfico más adecuado a los datos y comparar entre distintos tipos de gráficos.

En la siguiente imagen se puede ver el aspecto que tiene QuickSight a la hora de crear un análisis de datos:

Figura 4.2 – Amazon Quick Sight



En la parte inferior izquierda se pueden ver todos los tipos de visualizaciones que se pueden crear, solamente hace falta seleccionar una de ellas e introducirle los campos que se desean ver en cada eje. Estos campos se ven en la barra lateral, los que tienen un símbolo “#” son los campos numéricos, los que tienen un símbolo de calendario son los campos de fechas y los que tienen un símbolo que se asemeja a una etiqueta son los atributos. Cada tipo de gráfico tiene sus propios ajustes para poder modificar lo que se quiere ver y de qué manera (colores, orden de los valores, totales, subtotales, líneas de cuadrícula, etc.).

4.3. Comparación de plataformas

Analizando las posibilidades que presenta cada plataforma, se han observado varias diferencias en cuanto a la accesibilidad, la variedad de visualizaciones, el grado de dificultad de uso, los conocimientos previos necesarios, etc. En este apartado del trabajo se comparan al detalle los aspectos mencionados de cada una de las plataformas.

La principal diferencia entre las dos plataformas que se van a usar es el tipo de software que presentan cada una: Shiny es libre y gratis para cualquier usuario, solo hace falta instalarse el programa R-Studio y el paquete de Shiny, así como cualquier paquete que se necesite para

crear la aplicación. En cambio, Amazon Quick Sight es de pago y hace falta crear una cuenta en Amazon AWS de empresa o de usuario y abonar el pago mensual o anualmente para poder usar la plataforma.

Un beneficio de que la plataforma Quick Sight pertenezca a Amazon es que se puede acceder a ella desde cualquier sitio ya que está “en la nube” y se puede seguir trabajando sobre la aplicación y hacer modificaciones en cualquier lugar con acceso a internet. Esto no ocurre con Shiny ya que para modificar y trabajar sobre la aplicación es necesario tener instalado el R-Studio y poseer los archivos de código que componen la aplicación. Una vez se publica la aplicación de Shiny “en la nube” solo se puede acceder a ella como visitante, es decir, solo se puede ver el resultado y no es posible realizar ninguna modificación desde allí.

Relacionado con la accesibilidad a las plataformas, una vez acabada la aplicación en Shiny esta se publica y se crea un enlace con el cual se puede acceder a la misma, de esta manera se puede compartir libremente el enlace y cualquier persona puede ver el resultado. En cambio, cuando se acaba una aplicación en Quick Sight, esta se publica en la misma plataforma y solo se puede compartir con usuarios de Amazon AWS, no cualquier persona puede acceder a la aplicación.

A la hora de crear la aplicación en Shiny, son necesarios unos conocimientos de programación en lenguaje de R avanzados, así como aprender el código de Shiny que tiene comandos y funciones propias. Es decir, hace falta un aprendizaje importante antes de programar la aplicación, ya que como se ha comentado antes, el programa que se ejecuta debe seguir una estructura y formato determinados para que el R lo lea correctamente como una aplicación y no como un programa cualquiera.

En cambio, para crear una aplicación en Quick Sight solo hace falta tener conocimientos básicos de estadística y de análisis de datos, como saber que formato debe tener cada tipo de datos, qué medidas son las adecuadas para hacer cada análisis, los posibles tipos de gráficos y visualizaciones que se pueden usar en cada ocasión, etc. Quick Sight es mucho más intuitivo y fácil de comprender y usar, ya que tiene un formato más visual para el usuario.

Por último, hace falta comentar que en Shiny, al programar la aplicación desde 0, esta se puede personalizar al gusto del usuario a unos niveles más altos que en Quick Sight. La plataforma de Amazon contiene unas opciones de modificación y personalización limitadas para cada gráfico (como colores, formas, ejes, etc.), en cambio en Shiny hay infinidad de colores, opciones de gráfico, leyendas, y muchos más objetos que se pueden añadir a una visualización. No hay que olvidarse de que para conocer todas las opciones que puede ofrecer Shiny hace falta investigar y aprender nuevos comandos y códigos para implementarlas, ya que por sí solo R no te muestra todas estas posibilidades, sino que es como un folio en blanco.

V. PROYECTO DE VISUALIZACIÓN DE DATOS ABIERTOS SOBRE COMERCIO EXTERIOR DE ESPAÑA

5.1. Descripción e interés real del proyecto

La parte más práctica de este trabajo consiste en crear una aplicación de análisis y visualización de datos abiertos, la cual será una fuente de información para particulares y empresas del sector económico que tengan interés en el comercio exterior español (como empresas del sector de la banca).

El proyecto que se va a llevar a cabo se centra en analizar datos de comercio exterior para poder mostrar gráficos, tablas y otros elementos visuales para comprender la información de manera clara. Se quiere poder comparar de manera fácil el número y el valor en dólares de importaciones y exportaciones de diferentes periodos de tiempo, así como diferenciar los datos por países de origen o destino.

Como hemos mencionado varias veces durante el trabajo, una parte muy importante es que se podrá interactuar con los datos. El usuario podrá escoger que periodos de tiempo quiere ver, si quiere visualizar el conjunto de los países asociados o solo alguno/s de ellos, o si interesa ver el total de operaciones o solo las importaciones o exportaciones del territorio. De esta manera se crearán diferentes visualizaciones de los datos que ayudarán a comprender la situación actual de los valores y su evolución temporal.

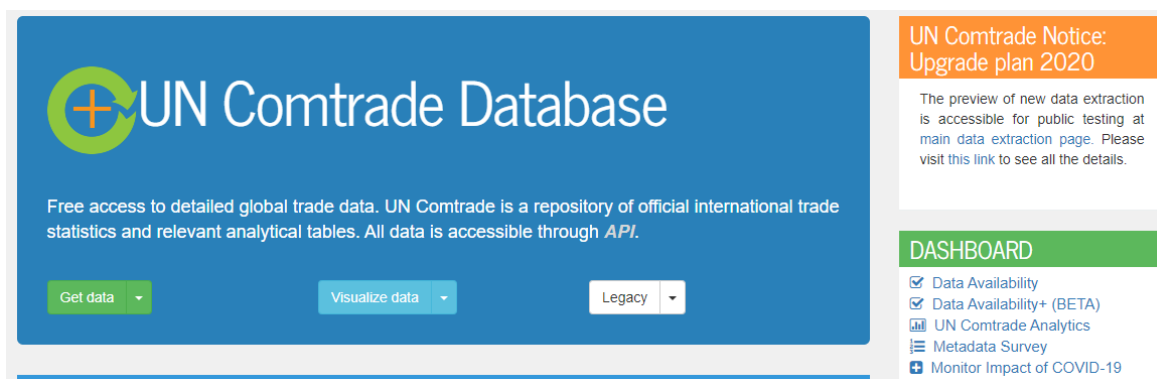
Este proyecto está basado en un proyecto real de Inteligencia Empresarial actual, es decir, las empresas españolas del sector bancario están interesadas en tener una aplicación donde se visualicen los datos de comercio exterior de cada una de ellas. Como nosotros no podemos disponer de datos propios de empresas lo haremos con datos abiertos que nos proporciona el gobierno español, de esta manera veremos qué información podemos sacar de estos datos y crearemos una aplicación sin restricciones de uso ni de privacidad.

5.2. Descripción de los datos: datos de UN Comtrade Database

Los datos que vamos a analizar para crear la aplicación son el número y el valor en dólares de importaciones y exportaciones que se realizan en España. Los datos los extraeremos del portal de datos UN Comtrade, este portal ha sido creado por las Naciones Unidas y en él se pueden encontrar datos comerciales a tiempo real de 170 países.

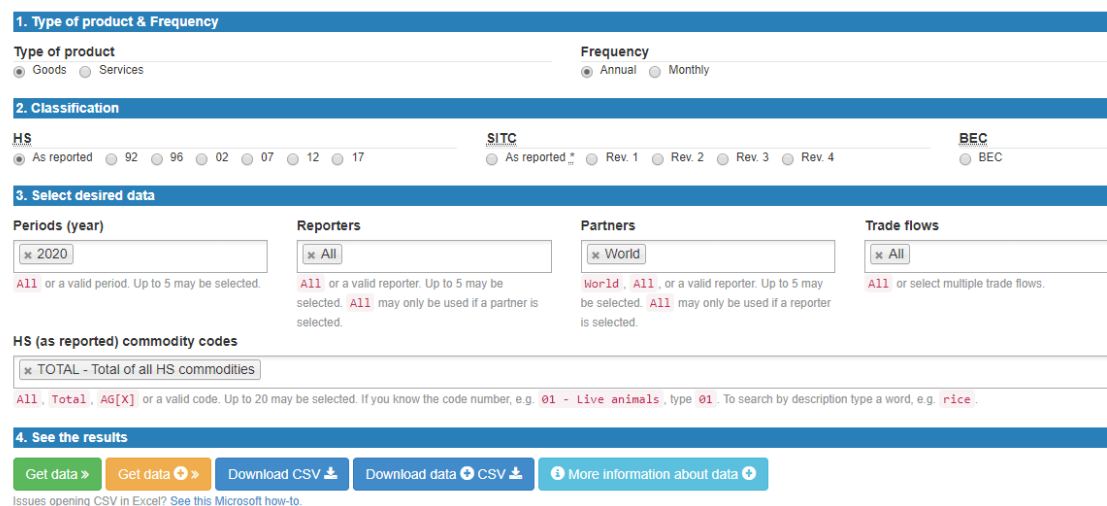
La pantalla de inicio de la web de UN Comtrade encontramos los siguientes apartados:

Figura 5.2.1 – UN Comtrade



En el panel principal de la imagen se encuentra un acceso directo “Get data” que lleva a la búsqueda de datos para su descarga en diferentes formatos, si se accede al botón se llega a la siguiente pantalla de filtrado y búsqueda de datos:

Figura 5.2.2 – UN Comtrade



Este es uno de los métodos mediante los cuales se pueden extraer datos de UN Comtrade.

Si se accede al botón “Visualize data” se llega a una página donde se ofrecen diferentes análisis de datos mediante gráficos y elementos visuales de diferentes proveedores y creados de aplicaciones. Por otro lado, en el lado derecho de la pantalla de inicio se encuentra información sobre la disponibilidad de datos en la plataforma y la planificación de las próximas subidas y actualizaciones de datos.

Otro apartado de interés que se encuentra en la página principal es el que se ve en la siguiente imagen, se llama “Analytical Tables” y en él se encuentran diferentes tablas y análisis de datos de la plataforma. En cada una de las publicaciones se puede ver la frecuencia de los datos (anuales, mensuales, trimestrales, etc.) i la última publicación o actualización de los datos.

Figura 5.2.3 – UN Comtrade

ANALYTICAL TABLES

World, Region and Country Profile
PUBLISHED IN JUNE
Overview (in tables, charts and analytical text) of the latest trends of trade in goods and services of World, selected region and trade groups, and most countries and areas in the world.
[Download T35-36 data series](#)

Monthly tables of imports and exports
PUBLISHED AT THE END OF THE MONTH
Total imports and exports by regions and countries or areas.
[Download T35-36 data series](#)

Product Profile
PUBLISHED IN DECEMBER
Overview (in tables, charts and analytical text) of the latest trends of trade in goods by commodity classes (SITC Rev.3 three-digit level).
[Download T38 data series](#)

Quarterly table of external trade conversion factors
PUBLISHED IN MARCH, JUNE, SEPTEMBER, DECEMBER
External trade conversion factors in US dollars per national currency.
[Download T38 data series](#)

Standard Unit Value (SUV)
PUBLISHED IN JANUARY
This indicator is calculated from original datasets to determine global unit values including their acceptable range unit value for each HS sub-heading.
[Download SUV data series](#)

Quarterly tables of import and export indices
PUBLISHED IN JAN, FEB, APR, MAY, JUL, AUG, OCT, NOV
External trade indices of total of total exports and imports.
[Download T37,39 data series](#)

Quarterly tables of manufactured goods exports
PUBLISHED IN MARCH, JUNE, SEPTEMBER, DECEMBER
External trade indices and indicators of manufactured goods exports.
[Download T40 data series](#)

Quarterly tables of fuel imports
PUBLISHED IN MARCH, JUNE, SEPTEMBER, DECEMBER
External trade indices and indicators of fuel imports.
[Download T18-19 data series](#)

Para obtener los datos que se analizarán en las aplicaciones en este trabajo, se extraen mediante una API entre R y UN Comtrade, de esta manera se dispone de los datos directamente en un formato apto para R ya que se importan en el mismo programa sin pasar por un formato externo.

Queremos obtener los datos de manera que sean mensuales y que estén lo más actualizados posibles. En la siguiente imagen vemos el formato que tienen los datos una vez extraídos:

Tabla 5.2 – Datos UN Comtrade

year	period	period_desc	trade_flow_code	trade_flow	reporter_code	reporter	partner_code	partner	trade_value_usd
2020	202005	May 2020	1	Imports	724	Spain	4	Afghanistan	44195
2020	202005	May 2020	1	Imports	724	Spain	8	Albania	5293321
2020	202005	May 2020	1	Imports	724	Spain	12	Algeria	148484076
2020	202005	May 2020	1	Imports	724	Spain	24	Angola	22006428
2020	202005	May 2020	1	Imports	724	Spain	32	Argentina	74284677
2020	202005	May 2020	1	Imports	724	Spain	36	Australia	11865183
2020	202005	May 2020	1	Imports	724	Spain	44	Bahamas	37017
2020	202005	May 2020	1	Imports	724	Spain	48	Bahrain	6292169
2020	202005	May 2020	1	Imports	724	Spain	51	Armenia	215046
2020	202005	May 2020	1	Imports	724	Spain	52	Barbados	233358
2020	202005	May 2020	1	Imports	724	Spain	60	Bermuda	4641
2020	202005	May 2020	1	Imports	724	Spain	70	Bosnia Herzegovina	1701847
2020	202005	May 2020	1	Imports	724	Spain	68	Bolivia	2017153

Los campos de interés son los siguientes:

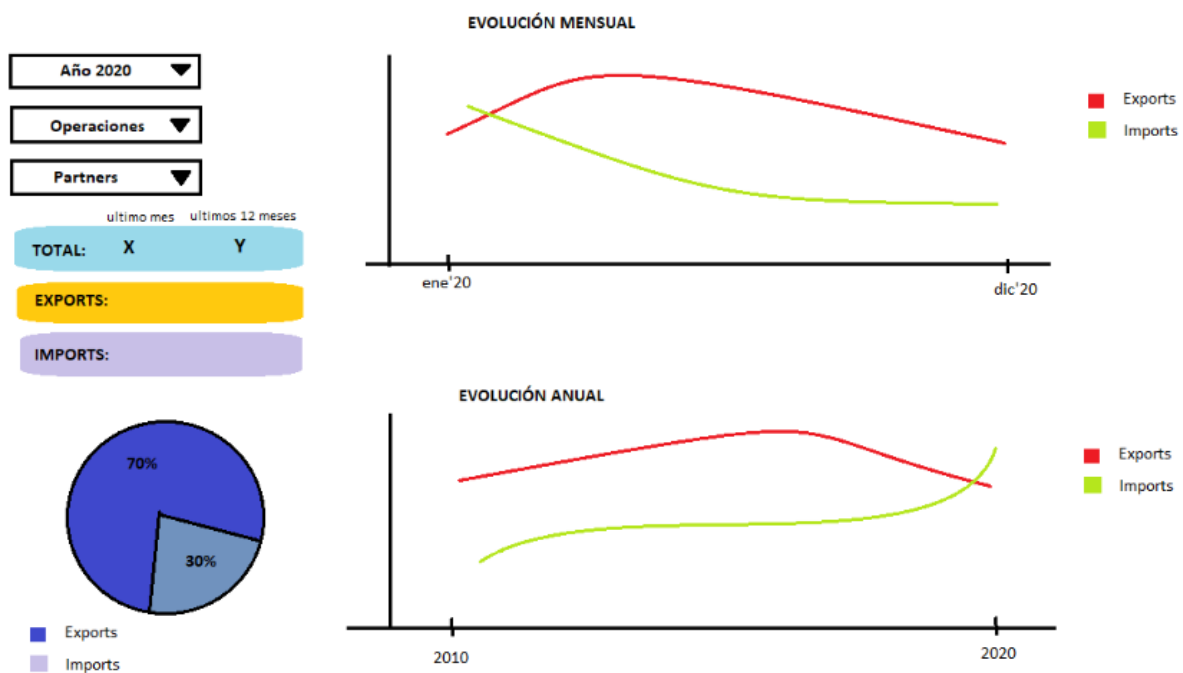
- Year: año de la operación.
- Period: campo formado por el año y el mes de la operación en formato “yyyymm”.
- Period_desc: descripción del campo period.
- Trade_flow_code: es un identificador de el sentido de la operación, es 1 si la operación es una importación, 2 si es una exportación, 3 si es una re-exportación y 4 si es una re-importación.
- Trade_flow: descripción del campo trade_flow_code, tiene dos niveles, Imports y Exports.
- Reporter_code: identificador del país de origen de la operación en caso de exportaciones (y reexportaciones), o del país de destino en caso de importaciones (y reimportaciones), (en nuestro caso siempre es 724).
- Reporter: país de origen de la operación en caso de exportaciones (y reexportaciones), o país de destino en caso de importaciones (y reimportaciones), (en nuestro caso siempre es España).
- Partner_code: identificador del país de origen de la operación en caso de importaciones (y reimportaciones), o del país de destino en caso de exportaciones (y reexportaciones).
- Partner: país de origen de la operación en caso de importaciones (y reimportaciones), o país de destino en caso de exportaciones (y reexportaciones).
- Trade_value_usd: valor de la operación en dólares.

5.3. Maqueta inicial del proyecto

Antes de empezar a crear las aplicaciones se ha diseñado una maqueta del proyecto, esta maqueta refleja el contenido que se quiere ver en las aplicaciones. Construyendo esta maqueta conseguimos tener un objetivo claro de lo que queremos hacer con los datos y de esta manera será más claro ver a posteriori si hemos tenido alguna complicación o si no se ha podido llevar a cabo alguna parte de la aplicación. Además, podremos comparar más fácilmente las dos aplicaciones resultantes de Shiny y de Quick Sight teniendo la maqueta inicial de referencia.

La aplicación constará de dos páginas o *dashboards*, en la primera página se quiere ver una visión general de los datos:

Figura 5.3.1 – Maqueta del proyecto de visualización de datos 1



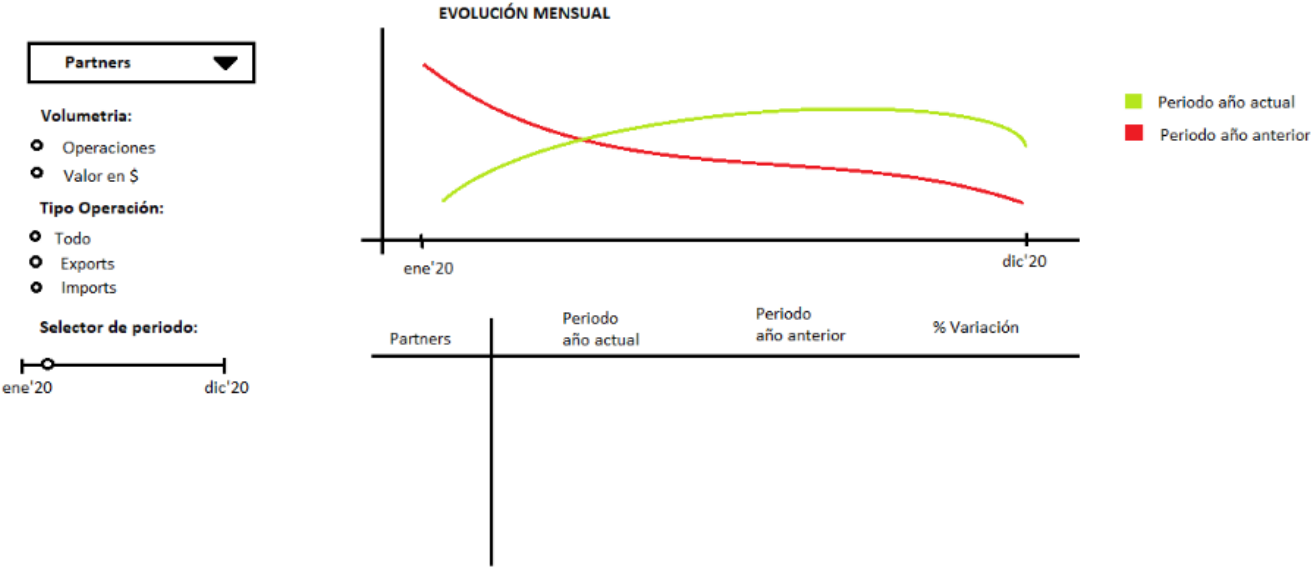
Esta página contiene tres selectores para filtrar los datos, en el primer selector se escogerá el año que se quiere visualizar en el gráfico de evolución mensual y en los indicadores, en el segundo se podrá seleccionar si se quiere ver el número de operaciones o el valor de estas en dólares (este selector afecta a todo el dashboard), y en el tercer selector se podrán seleccionar los países que nos interesan ver.

Debajo de los selectores se encuentran tres indicadores donde se verán el total de operaciones (o dólares), las exportaciones e importaciones del último mes del año seleccionado y de los últimos doce meses.

En el gráfico de tarta se verán las proporciones de operaciones o dólares del año seleccionado. En el gráfico de evolución mensual se verán las importaciones y exportaciones por meses, y en el gráfico de evolución anual se verán los por años hasta el año actual.

En la segunda página se quiere ver una comparación anual más al detalle:

Figura 5.3.2 – Maqueta del proyecto de visualización de datos 2



En esta página hay cuatro selectores para filtrar los datos que se visualizan, en el primero se seleccionarán los países de interés (igual que en la primera página), en el segundo se escogerá si se quieren ver el número de operaciones o el valor en dólares de estas, en el tercero se seleccionará si se quieren ver las exportaciones, importaciones o el total, y por último se escogerá el periodo de tiempo que se quiere visualizar.

En el gráfico mensual esta vez se verán los datos del año actual (en la maqueta 2020) y los datos del mismo periodo para el año anterior. En la tabla de abajo se verán los datos del año actual y del año anterior distinguiendo entre los países seleccionados, además se verá la variación entre los dos periodos.

5.4. Proyecto en Shiny

5.4.1. Explicación de cálculos, indicadores y otros objetos de Dashboard

En este apartado se explican al detalle los cálculos y modificaciones que se han hecho sobre los datos para crear cada uno de los elementos del *dashboard* en Shiny. Como hemos

comentado anteriormente la creación de una aplicación en Shiny se basa en un código de R elaborado de manera que, para poder crear un gráfico, por ejemplo, hay que tratar los datos de forma que solo haga falta ponerlos dentro de la función de “gráfico” o “tabla” y esta función lea los datos correctamente (esta función es *ggplot()* en el caso de los gráficos de líneas, *pie3D()* en el gráfico de sectores y *datatable()* en el caso de la tabla). Una vez se comprende la estructura de la aplicación, las características de cada elemento y la parte estética de la aplicación, lo más difícil es programar el código que modifique los datos para cada elemento y que tenga en cuenta los filtros que se aplicarán con los selectores.

Empezando por la primera página del *dashboard*, para el gráfico de evolución mensual se inicializan los datos de la siguiente manera: Se cogen los datos que tengan como *Partner* a China i como año el último año con datos que es 2020. Por otro lado, se ordenan estos datos por el periodo (“202001” es enero, “202002” es febrero, etc.) para que el eje X en el gráfico se haga correctamente de enero a diciembre. Por último, se suma el campo de valor en dólares por cada mes, y cada tipo de operación para poder distinguir entre exportaciones, importaciones, reexportaciones y reimportaciones. Para hacer la suma agrupando por los diferentes campos se usa la función de R *aggregate()*, la cual suma el campo que se le indique agrupándolo por un listado de variables que actúan como factor. De esta manera queda: la descripción del periodo (mes en formato: “January 2020”) para el eje X, el valor en dólares para el eje Y, y el tipo de operación para los subgrupos. Para que cambien los datos con los selectores lo que se hace es que cada vez que se le da al botón de “plot” se filtren los datos con los años y los *Partners* seleccionados (se cambia 2020 y China que eran los valores por defecto). No le afecta el selector de tipo de operación porque en el gráfico se ven todos los tipos siempre.

Para el gráfico de evolución anual también inicializamos cogiendo los datos que tengan como *Partner* a China. Después sumamos el campo de valor en dólares por cada año y tipo de operación. De esta manera queda: el año para el eje X, el valor en dólares para el eje Y, y el tipo de operación para los subgrupos. Para que le afecte el selector de *Partner* lo que se hace es cada vez que se le da al botón “plot” se filtran de nuevo los datos con los *Partners* seleccionados. No le afecta el selector de año ni el de tipo de operación ya que en el gráfico se ven todos los años y todos los tipos de operación siempre.

El tercer elemento de esta página es el gráfico de sectores, para crear este elemento se inicializan los datos cogiendo las observaciones que tengan como *Partner* a China y como año el último año con datos que es 2020. Por último, se suma el campo del valor en dólares distinguiendo por el tipo de operación y se calcula la proporción de dólares de cada tipo de operación sobre el total. De esta manera queda la proporción de dólares como volumetría y

el tipo de operación como grupo. Cada vez que se haga una selección se filtrarán de nuevo los datos con el año y los *Partners* seleccionados. A este gráfico no le afecta el selector de tipo de operación porque se ven siempre todos los tipos.

En la segunda página del *dashboard* se hace una comparación de datos anuales como se ha explicado en el apartado de la maqueta del proyecto. En esta página hay dos elementos, el primero de ellos es un gráfico de líneas en el que se ve la evolución mensual de el año seleccionado y del año anterior. Para este gráfico se inicializan los datos de la siguiente manera: se cogen los datos que tengan como *Partner* a China, que el tipo de operación sea Exportación, y que el año sea el último año con datos o el último año menos uno (en nuestro caso será 2020 o 2019). Posteriormente se suma el valor en dólares por cada periodo, es decir por cada mes distinguiendo entre los dos años. De esta manera quedará el valor en dólares para el eje Y, los meses en el eje X, y el año como grupo. A este gráfico le afectan los tres selectores, y cada vez que se hace una nueva selección se filtran los datos de nuevo cogiendo el año seleccionado y el anterior, los tipos de operaciones y los *Partners*.

Por último, en la segunda página se encuentra una tabla donde se ven los valores de cada *Partner* para el año actual y el anterior, y también se ve la variación anual. Para esta tabla los datos se inicializan cogiendo los datos del año actual y el anterior, que tengan como tipo de operación una Exportación. Una vez filtrados los datos se suma el valor en dólares por cada *Partner* y por cada año, también se calcula la variación de un año a otro en porcentaje creando una nueva variable llamada “% Variación” que se ha calculado de la siguiente forma:

$$\% \text{ Variación} = \frac{\text{Valor año actual} - \text{Valor año anterior}}{\text{Valor año anterior}} \cdot 100$$

Hace falta destacar que los subconjuntos de datos creados para cada uno de los gráficos y tablas no han sido creados de forma separada, sino que para los elementos que necesitaban los mismos filtros se han aprovechado esos datos para optimizar el código del programa y no repetir filtros y funciones redundantes.

5.4.2. Presentación del proyecto

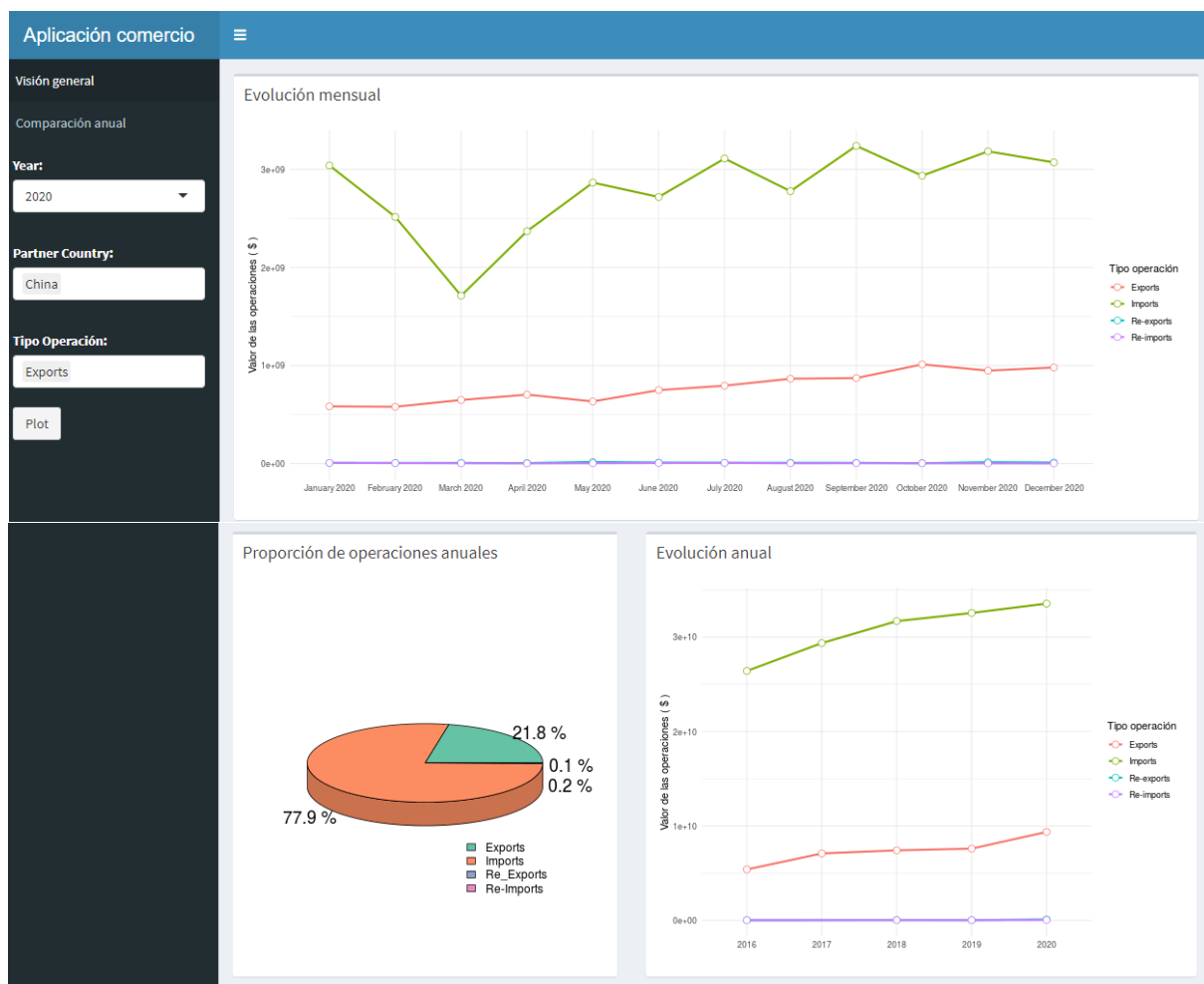
La aplicación hecha con Shiny una vez acabada da muy buenos resultados, aunque no ha sido del todo fácil su creación, ya que como se ha comentado varias veces el programa que se ha

necesitado escribir no es intuitivo, requiere mucha dedicación y consecuentemente más tiempo invertido que el que se desearía.

El enlace de la aplicación final es el siguiente: <https://nurianunezrubio13.shinyapps.io/shiny/>. En este apartado se explica detalladamente el resultado de la aplicación y sus diferentes apartados, así como los inconvenientes y dificultades que se han podido encontrar para llevarla a cabo.

Como se ha visto en la maqueta, la aplicación contiene dos páginas, la primera se llama “Visión general” y en la siguiente imagen se puede ver su resultado con los valores por defecto:

Figura 5.4.1 – Resultado proyecto en Shiny 1



Se ha creado una barra lateral que se oculta dándole a las tres rallas que aparecen arriba al lado del título de la aplicación. En esta barra lateral se encuentran arriba las dos pestañas que nos permiten navegar entre las páginas de la aplicación, en esta imagen nos encontramos en la primera. En la misma barra lateral se han situado finalmente los selectores para dejar más espacio para los gráficos y visualizaciones. En cuanto a los selectores, el selector de año solo

permite seleccionar un año, en cambio el selector de *Partner* y el de Tipo de Operación permiten hacer selecciones múltiples. Los selectores son tipo lista (se extiende la lista de opciones de selecciones al hacer clic en ellos) y el de *Partner* y Tipo de Operación se puede escribir la opción que se desea escoger para encontrarla mejor en la lista.

En la pantalla vemos en la parte superior el gráfico de evolución mensual del valor de las operaciones del año seleccionado, y en su parte derecha la leyenda del tipo de operación. En este gráfico si se seleccionan varios países, se verá reflejada la suma de estos países.

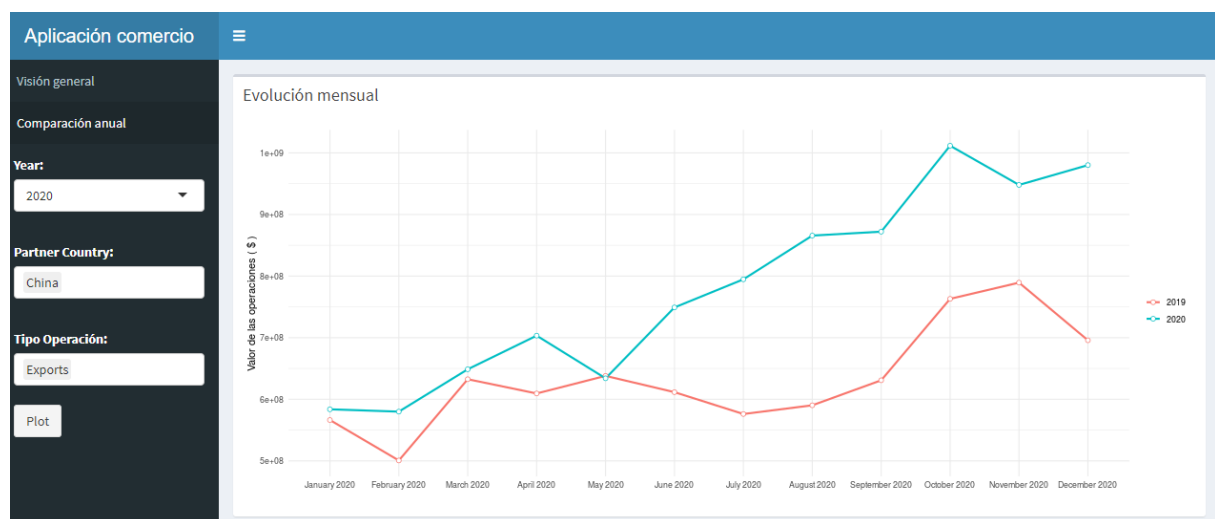
Abajo a la derecha se ve el gráfico de evolución anual, en él se ve la evolución de todos los años que existen en los datos cargados. De la misma manera que en el gráfico de evolución mensual, se ve a leyenda en la parte derecha y, si se selecciona más de un país se verá reflejada la suma de estos.

Por último, abajo a la izquierda se ve el gráfico de sectores donde se ven los porcentajes que ocupan cada tipo de operación para el año y los países seleccionados. Si se selecciona más de un país se suman los valores y se calculan de nuevo los porcentajes.

En esta página hace falta destacar que no se usa el selector de Tipo de Operación en ninguno de los elementos, ya que en las tres visualizaciones que hay se ven todos los tipos de operación siempre.

En la imagen de debajo se puede ver ahora la segunda página de la aplicación que se llama “Comparación anual”.

Figura 5.4.2 – Resultado proyecto en Shiny 2



Partners	Año actual	Año anterior	% Variación
China	9.371.245.146	7.603.625.340	23.25 %

Al pasar a la segunda pantalla, se observa que la barra lateral es siempre la misma para las dos páginas y esto incluye que haya los mismos selectores también. Ahora se están mostrando los valores por defecto en los selectores de la aplicación.

En el gráfico superior de evolución mensual se ve el valor de las exportaciones para cada mes del año actual (2020) y del año anterior, en la parte derecha aparece la leyenda para poder ver qué línea corresponde a cada año. En este gráfico se comparan de manera muy clara y visual el valor de las operaciones de cada año seleccionado con el año anterior. En este caso si se selecciona más de un país o tipo de operación se verá reflejada la suma de estos valores en el gráfico.

En la parte inferior de la página está situada la tabla de comparación, en ella se ve por cada país *Partner* la suma del valor para el año actual, para el año anterior y la variación de estos valores. Si seleccionamos varios países se pueden ordenar por la columna que se prefiera haciendo clic en el nombre, por defecto están ordenados alfabéticamente. En la siguiente imagen vemos una selección de seis países:

Figura 5.4.3 – Resultado proyecto en Shiny 3

Partners	Año actual	Año anterior	% Variación
Canada	2.155.555.752	2.256.386.120	-4.47 %
China	9.371.245.146	7.603.625.340	23.25 %
Cuba	676.711.710	1.017.182.805	-33.47 %
Finland	1.205.189.566	989.315.981	21.82 %
Iraq	164.038.800	202.347.010	-18.93 %
Sweden	2.802.704.968	2.694.311.779	4.02 %

Abajo a la izquierda nos dice el total de países seleccionados y cuantos está mostrando en esta página de la tabla. Arriba a la izquierda se puede escoger cuántas observaciones se quieren

ver a la vez, en este caso si el número de *Partners* supera los 10 se mostrarían en varias páginas. Arriba a la derecha se puede buscar un determinado país escribiendo su nombre (un país de los seleccionados). Por último, abajo a la derecha se puede pasar de una página a otra en el caso de que haya varias.

Como se puede ver, finalmente no se han incluido los indicadores que constaban en la maqueta ya que los selectores han ocupado más espacio del esperado y si se incluían en la pantalla principal dejaban demasiado poco espacio a los gráficos.

Otra cosa que ha variado respecto a la maqueta es que los selectores han sido los mismos para las dos páginas en vez de tener cada una los suyos distintos. Esto ha sido debido a que se han incluido en la barra lateral para optimizar el espacio y esta es común para todas las páginas de la aplicación. El tener los selectores en común ha acabado siendo una ventaja y una mejora respecto a la maqueta, ya que de esta manera si se hace una selección estando en una página concreta, al cambiar de página se mantienen las mismas selecciones.

5.5. Proyecto en QuickSight

5.5.1. Explicación de cálculos, indicadores, y otros objetos del Dashboard

En este apartado se explica cómo y de qué manera se han creado cada una de las visualizaciones de datos de la aplicación en QuickSight. Con esta plataforma no ha sido necesario crear subconjuntos de datos para cada gráfico o tabla, sino que en la misma plataforma se han podido modificar, calcular y añadir variables en el conjunto de datos para crear las visualizaciones de la manera deseada.

Empezando por la pantalla de “Visión general”, en el gráfico de evolución mensual se ha usado el gráfico de líneas, en el cual en el eje X se ha colocado el periodo, en el eje Y el valor en dólares y en el color el tipo de operación. La variable de periodo se ha tenido que modificar ya que estaba en formato “202012”, se ha cambiado usando el formato de fecha “yyyyMM” para tenerlo de la siguiente manera: “dic. 2020”, así las etiquetas del eje X son mucho más claras. Por otro lado, el valor en dólares se ha agrupado como la suma en cada punto del gráfico. A este gráfico le afecta tanto el selector de año como el de *partners*.

Para el gráfico de evolución anual se ha usado el mismo tipo de gráfico de líneas, en el cual se ha colocado el año en el eje X, el valor en dólares en el eje Y, y el tipo de operación en los colores. En este caso no ha hecho falta modificar ninguna variable. El valor en dólares también

se expresa como la suma en cada punto del gráfico. A este gráfico no le afecta el selector de año ya que queremos que se vean siempre todos los años, solo le afecta el selector de *partners*.

La tercera visualización en esta página es el gráfico de sectores, como volumetría se ha cogido el valor en dólares, y como dimensión para dividir los sectores se ha seleccionado el tipo de operación. El valor en dólares se ve reflejado como la suma por cada tipo de operación. A esta visualización le afecta tanto el filtro de año como el de *partners*.

En la segunda página del *Dashboard* de Comparación Anual se encuentra otro gráfico de evolución mensual, para este gráfico se ha tenido que crear una variable de “desc_mes” para poder ponerla en el eje X, esta se ha creado extrayendo el mes del campo de fecha “period” que se ha mencionado antes. El eje X que contiene la descripción de cada mes se ha ordenado con la variable “mes” que contiene el número de mes correspondiente, de esta manera queda el gráfico correctamente ordenado de enero a diciembre. En el eje Y se ha colocado el valor en dólares, y en el color se ha seleccionado la variable año. Igual que en los casos anteriores la variable del valor en dólares se expresa como la suma en cada punto del gráfico. A este objeto le afectan los tres selectores que encontramos en la pantalla, el de año, el de *partners*, y el de tipo de operación.

Otro objeto de la página de Comparación Anual es la tabla de *partners*, en la cual se han colocado como filas los *partners* y como columnas los años seleccionados. Como medidas A esta tabla le afectan los tres selectores también, tanto el de año como el de *partners* y el de tipo de operación. En QuickSight no ha sido posible calcular una columna que sea la variación anual de cada *partner* ya que no se pueden hacer cálculos complejos dentro de las visualizaciones. En lugar de la variación se ha añadido una columna que muestra el total del valor en dólares de cada *partner* para el total de años seleccionados. A esta tabla de datos le afectan también los tres selectores de la página.

Para tener un indicador de la variación anual del valor, se ha añadido a la página un indicador de rendimiento al cual se le ha colocado el valor en dólares como métrica y el año como grupo. Este objeto de visualización calcula la variación entre los años seleccionados de la volumetría dada. Le afectan los tres selectores, pero cuando se seleccionan más de dos años, solo se ven los primeros dos años seleccionados.

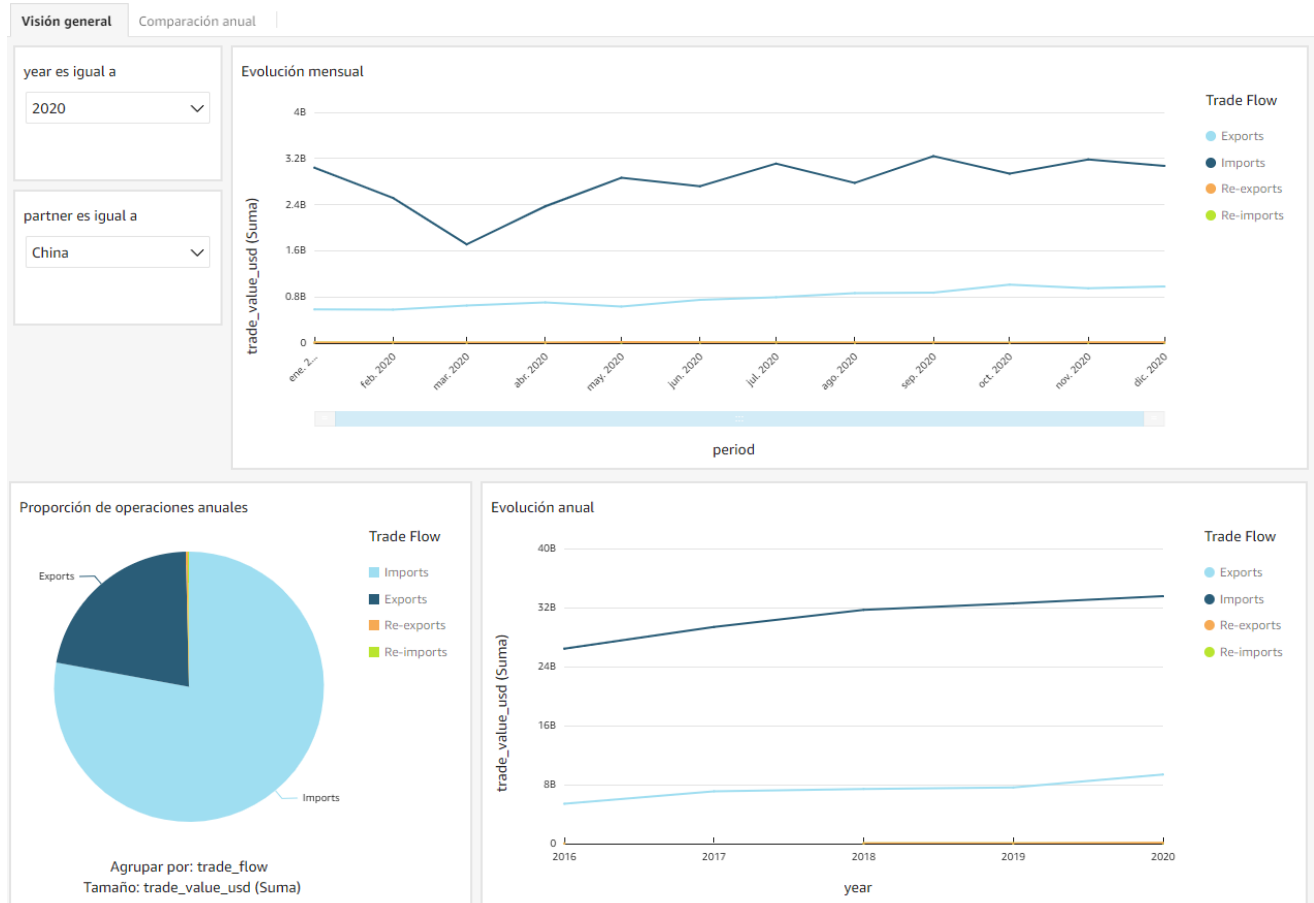
5.5.2. Presentación del proyecto

Una vez acabada la aplicación con QuickSight de Amazon, se puede decir que se han obtenido unos buenos resultados teniendo en cuenta las comodidades que aporta esta plataforma a la hora de crear visualizaciones.

QuickSight no permite crear un enlace directo a la aplicación para poderlo compartir libremente, sino que para poder acceder a ella se debe tener cuenta de Amazon Aws con acceso a QuickSight. De esta manera solo se puede compartir la aplicación internamente, entre usuarios de la misma plataforma.

Como ya se ha comentado varias veces, la aplicación consta de dos partes: la primera página de Visión general de los datos y la segunda de Comparación anual. En la siguiente imagen se puede ver el resultado final de la primera página del Dashboard con los valores por defecto:

Figura 5.5.1 – Resultado proyecto en Amazon 1



En esta imagen se observa arriba a la izquierda las pestañas por las que podemos cambiar de página e ir a la Comparación anual, hace falta destacar que al cambiar de página no se

guardarán las selecciones hechas en la página de origen ya que cada pantalla tiene sus propios selectores.

En la esquina superior izquierda vemos los dos selectores de esta página, el de año que afecta al gráfico de evolución mensual y al gráfico de sectores, y debajo se encuentra el de *Partners* que afecta a todos los gráficos de esta página. Ambos selectores permiten hacer selecciones múltiples sin límite de elementos escogidos. Cuando se escoge más de un año y/o más de un país, en los gráficos se refleja la suma de estos años y países.

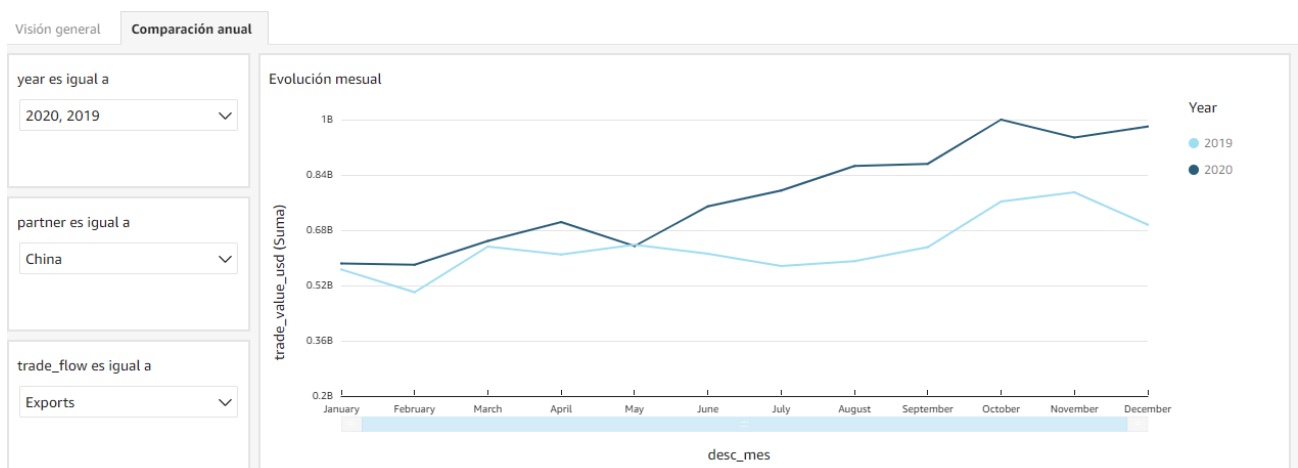
En el gráfico superior de Evolución mensual se puede ver el valor en dólares de las operaciones realizadas por los *partners* seleccionados en cada mes del año o años escogidos, distinguiendo entre los diferentes tipos de operación.

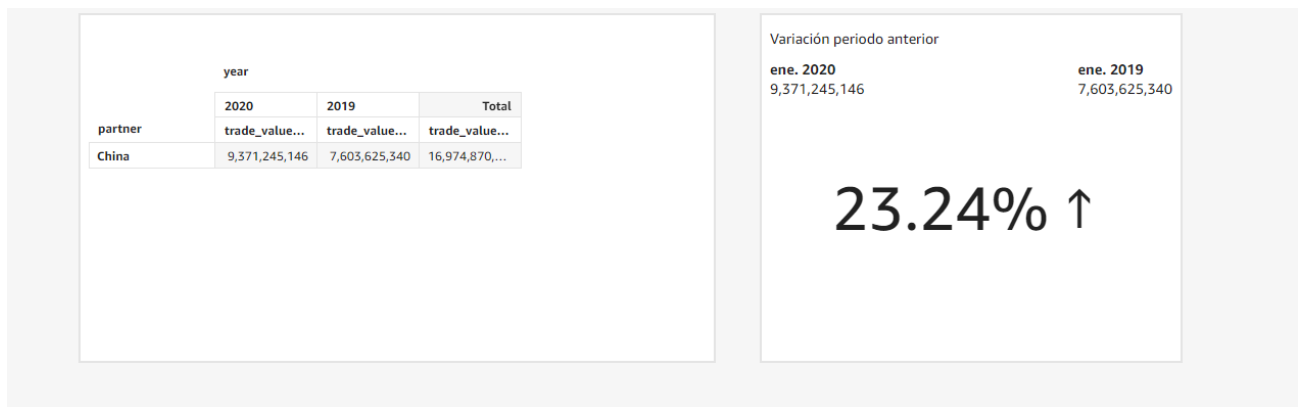
Abajo a la derecha se ve el gráfico de evolución anual, en el cual se ve el valor en dólares de las operaciones realizadas por los *partners* seleccionados en cada año, distinguiendo igual que antes por los diferentes tipos de operación.

Por último, la tercera visualización que se ve en esta página del Dashboard es el gráfico de sectores, en este gráfico se ve la suma del valor de las operaciones en dólares en el total de los años y los *partners* seleccionados, y solo se distingue por el tipo de operación.

En la siguiente imagen se puede ver el aspecto que tiene la segunda pantalla de Evolución anual con los valores por defecto con los que se abre la aplicación:

Figura 5.5.2 – Resultado proyecto en Amazon 2





En esta pantalla de la aplicación hay tres selectores, el de año y el de *partners* que se encuentran también en la pantalla que hemos visto anteriormente, y el de tipo de operaciones que solo lo contiene esta página. Los tres selectores afectan a todas las visualizaciones de la pantalla, y en los tres se permiten hacer selecciones múltiples.

El gráfico superior es el de evolución mensual, en él se ve el valor en dólares de las operaciones realizadas por los *partners* y el tipo de operación seleccionados para cada mes, distinguiendo entre los diferentes años. Si se selecciona un nuevo año, se añadirá una línea de diferente color al gráfico correspondiente a ese año. En el caso de que se seleccionen más de un *partner* o más de un tipo de operación, se verá reflejada la suma de todos los países y tipos de operaciones.

En la parte de abajo a la izquierda se encuentra la tabla de *partners*, en ella se muestra una fila por cada país seleccionado y una columna por cada año, de esta manera se puede comparar numéricamente el valor de los distintos años para cada uno de los países. También incluye una columna final que muestra el total del valor para cada país, es decir la suma de los años que se muestran en la tabla. Cuando se seleccionan más países la tabla se ve más completa, como en la siguiente imagen:

Figura 5.5.3 – Resultado proyecto en Amazon 3

partner	year		Total
	2020	2019	
Austria	2,525,585,643	2,610,403,384	5,135,989,027
Belgium	8,289,004,949	9,238,434,293	17,527,439,...
Brazil	2,588,089,404	2,904,545,271	5,492,634,675
Bulgaria	727,619,304	1,475,261,679	2,202,880,983
China	9,371,245,146	7,603,625,340	16,974,870,...
Ecuador	429,796,347	693,944,516	1,123,740,863
Greece	2,155,557,436	2,743,224,127	4,898,781,563
India	1,250,269,994	1,506,754,339	2,757,024,333

Como se puede ver, la tabla es una muy buena herramienta para comparar rápidamente el valor de las operaciones de los distintos países y la diferencia entre los años seleccionados.

Como se ha comentado anteriormente, no se ha podido incluir la columna del porcentaje de variación anual ya que QuickSight no deja hacer cálculos complejos dentro de las visualizaciones. Por este motivo se ha incluido un tercer elemento en el *dashboard*: un indicador KPI. Este indicador nos muestra la variación del valor de las operaciones entre dos años y el valor correspondiente a cada año, en la imagen se puede ver concretamente el porcentaje de variación entre el año 2021 y el 2020 de las exportaciones de China, en este caso ha habido un aumento del 23,24% del valor. En el caso de que se seleccionaran varios países o tipos de operaciones, se reflejaría como la variación total de la suma de los países y los tipos de operación. Si se escogiera un tercer año, este no se vería en el indicador, sino que sería necesario deselegionar alguno de los años que se muestran actualmente.

5.6. Comparación de resultados

Una vez hechas las dos aplicaciones con Shiny de R y Amazon Quick Sight, se comparan los resultados obtenidos con cada una de estas plataformas. Se ven los inconvenientes y ventajas de cada una de las aplicaciones, y de la misma manera se valora el proceso que se ha seguido para la creación de estas.

En primer lugar, a la hora de la lectura e importación de datos, es muy importante destacar el hecho de que R permite usar APIs para conectar directamente con la fuente de datos. De esta manera no hace falta exportar los datos de la fuente e importarlos después al programa de

Shiny, sino que con el comando que se explica a continuación se importa el conjunto de datos deseado de *UNcomtrade* al mismo R.

Instalando la librería “comtradr” en el R se cargan una serie de funciones que permiten manejar datos de *UNcomtrade*. En la aplicación se ha usado concretamente el siguiente comando:

```
ct_search(reporters = "Spain",  
          partners = "All",  
          start_date = (2019),  
          end_date = (2020),  
          freq="monthly",  
          type="goods",  
          trade_direction = "All")
```

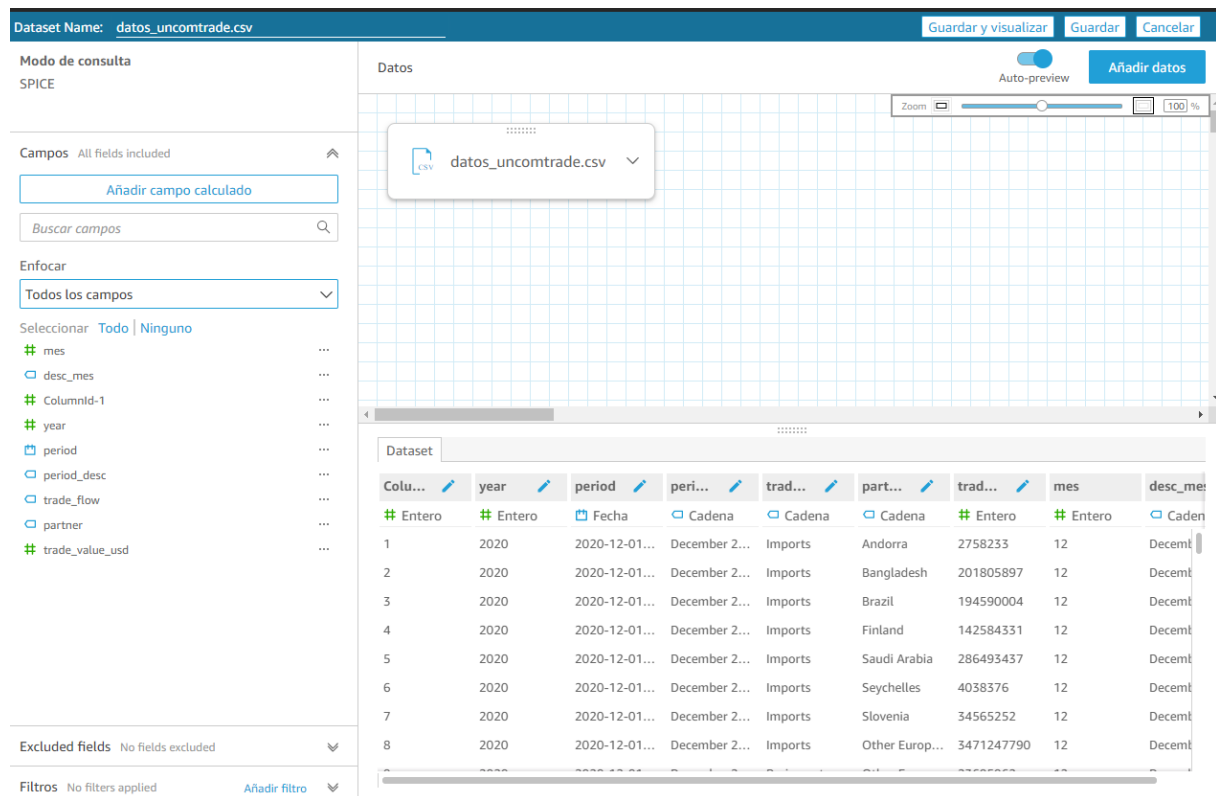
Con `ct_search()` estamos llamando a la importación de datos, dentro de esta función se indican unos parámetros que definen el conjunto concreto de datos que se quiere importar. En este caso se quiere que el *reporter* sea España, los *partners* se seleccionan todos, los datos serán de 2020 y la frecuencia de los datos será mensual. Ejecutando esta función en R se crea directamente un conjunto de datos con las variables y observaciones de 2020.

En Amazon Quick Sight no existe una API que conecte la plataforma directamente con la fuente de datos. Lo que se ha hecho ha sido exportar los datos en Excel (.csv) e importarlos en Quck Sight. Este punto sería una ventaja trabajar en Shiny, ya que no depende de un tercer programa como Excel, sino que solo se necesita el R. También es mucho más sencillo importar datos con la API, y de esta manera poder hacer pruebas y nuevas importaciones de manera más rápida y ágil.

Una vez importado el conjunto de datos que se va a usar, es más fácil añadir y modificar datos con Amazon, ya que tiene un apartado de vista de datos donde se pueden crear campos calculados a partir de otros o cambiar el formato de estos de manera manual.

En la siguiente imagen se puede ver el apartado comentado en el párrafo anterior. En la parte inferior derecha hay una previsualización de los datos, de esta manera se ven reflejados los cambios que se hacen y se puede comprobar si los datos van quedando como deben. En la parte izquierda se ven los campos, donde apretando en los tres puntos que tienen cada uno se puede modificar el formato, el nombre, editar un cálculo o eliminar el campo. Como se puede ver, encima de los campos hay un botón de “añadir campo calculado” y allí se abre un panel donde puedes escoger funciones parecidas a las de Excel para calcular nuevos campos a partir de los ya existen.

Figura 5.6 – Visor de datos Amazon Quick Sight



En Shiny realizar cambios en los datos es más complicado, sobre todo si la aplicación esta avanzada (es decir, si se lleva buena parte del programa escrito), ya que los cambios que se hacen en el programa afectarán al resto de comandos porque los datos ya no serán los mismos. En cambio, en Amazon la aplicación se adapta a los cambios en los datos y es más fácil rectificar los gráficos y las visualizaciones que dependen de estos.

Otra cuestión que tiene que ver con el tratamiento de datos es que, en Amazon existe el impedimento de que no se pueden realizar cálculos en el mismo análisis de datos. Es decir, no se puede crear una tabla donde se vea la diferencia del año seleccionado con el año anterior, ya que no hay ningún comando que te permita filtrar por el año seleccionado y el año seleccionado menos uno (año anterior) y hacer el porcentaje de variación de estos valores. En cambio, en Shiny como se crea la aplicación con un programa hecho por el usuario, se pueden hacer cálculos de este tipo en cualquier momento del proceso (en los mismos gráficos o tablas incluso, para que sean sensibles a los selectores).

Otro aspecto que comentar son las diferencias entre las posibilidades de los selectores de cada una de las plataformas. En Shiny existe la posibilidad de tener los mismos selectores para todas las páginas de la aplicación, ya que se puede crear una barra lateral que no cambia al ir de una página a la otra. Esto puede ser una ventaja ya que las selecciones que se hagan en la

primera página se mantendrán al pasar a la segunda y viceversa. Por ejemplo, si se quieren ver las exportaciones de china del año 2020, se hace la selección de estos parámetros en una de las páginas y al cambiar de página no se tendrán que volver a seleccionar los parámetros de interés ya que se mantiene la selección. Esto no ocurre en Quick Sight, ya que no existe la posibilidad de crear un apartado común para todas las páginas y es necesario que cada página tenga sus propios selectores.

Siguiendo con los selectores, en Shiny existe la ventaja de poder escoger si se puede seleccionar más de un valor o no, por ejemplo, hay gráficos que si selecciona más de un año, se ven demasiados datos y es mejor que solo se pueda seleccionar un año a la vez. En Quick Sight no existe esta opción, pero tiene otras ventajas de las cuales Shiny carece, como poder escoger en cada uno de los selectores a qué gráficos y visualizaciones deben afectar, o incluir una opción en los selectores de “seleccionar todo” (para seleccionar todas las opciones en un selector de Shiny hay que seleccionarlas todas a mano).

Por último, hace falta comparar las posibilidades que tiene cada plataforma a la hora de compartir o difundir la aplicación. En el caso de Shiny, cuando se publica una aplicación se crea un enlace que se puede compartir libremente sin ninguna restricción. Es decir, este enlace se puede enviar por correo electrónico, por mensaje SMS, se puede poner en un documento como Word o PDF y acceder desde ahí, o incluso poner en una página web como hipervínculo. En cambio, cuando se publica una aplicación en Amazon Quick Sight, el creador de esta puede otorgar permisos a otros usuarios para que estos puedan acceder a la aplicación. Por tanto, no se puede difundir libremente una aplicación de Quick Sight, sino que debes ser usuario para poder acceder y, además te deben dar los permisos correspondientes. El sistema de difusión de la plataforma de Amazon puede funcionar bien para aplicaciones de uso interno en una empresa, ya que se pueden dar permisos por departamentos y enviar informes concretos, además de la seguridad que proporciona al no tener libre acceso a los datos.

VI. CONCLUSIONES

En este trabajo se han conseguido cumplir los objetivos planteados inicialmente, empezando por el marco teórico donde se quería investigar la actualidad e importancia de los datos abiertos. Recogiendo y comparando información de varios *papers* y artículos, se concluye que los datos abiertos son la base de la estructura para que un país u organización funcione de forma eficiente y adecuada, y que estos datos se han convertido en la esencia para el desarrollo de la economía y la innovación. Las afirmaciones anteriores se basan en todos los datos e indicadores obtenidos durante el trabajo, como el crecimiento del mercado de datos en los últimos años, el surgimiento de nuevos modelos de negocio o la mejora de la eficiencia en la producción.

La segunda meta que se planteó fue ver de qué manera afectan los datos abiertos al día a día de la sociedad. Durante el trabajo se ha visto como los datos abiertos han mejorado significativamente muchos aspectos de la vida cotidiana, por ejemplo, el aumento de puestos de trabajo generados por los datos, la reducción de tiempo de espera en infinidad de servicios como el transporte público, o el ahorro de costes para muchas empresas y organizaciones gracias a la información que les proporcionan estos datos.

El tercer objetivo que incluye el marco teórico del trabajo es ver cómo se consigue extraer valor de los datos abiertos, es decir, cómo a través de la reutilización de los datos estos se convierten en información valiosa. Contrastando información de organizaciones y entidades que trabajan con datos abiertos se ha visto qué es la *Data Science* al detalle, es decir, el conjunto de técnicas de almacenamiento, procesamiento, análisis y visualización de datos que permiten extraer este valor. Una vez visto el proceso, ha interesado ver también qué tipo de datos son los más reutilizados y qué sectores los utilizan. En España más del 80% de las entidades están realizando proyectos de análisis de datos o tienen planes de hacerlo.

La información recogida de organizaciones y empresas que reutilizan datos abiertos ha llevado a la conclusión de que un negocio que dispone de información de su sector de actividad puede mejorar i crear una imagen de negocio adaptada a la realidad, mantenerla actualizada, entrar en el mercado con unos precios competentes y un producto que cumpla las necesidades del cliente en cada momento.

El segundo gran bloque de este trabajo tiene como objetivo llevar a cabo un proyecto de reutilización de datos abiertos, concretamente elaborando dos aplicaciones de análisis y visualización de datos del comercio exterior de España. Una de las aplicaciones se ha creado con el programa Shiny de R y la otra con Quick Sight de Amazon para poder comparar estas

dos plataformas y ver qué posibilidades ofrece cada una. Las aplicaciones creadas han dado muy buenos resultados, por lo tanto, se puede decir que el grado de cumplimiento de la maqueta inicial del proyecto es alto. En cada una de las dos plataformas se han encontrado diferentes impedimentos para la creación de algunos elementos de la aplicación, pero de la misma manera se han conseguido solucionar con otras posibilidades que ofrecen. Hace falta comentar que crear la aplicación con Shiny ha sido más difícil y ha requerido más tiempo que crearla con Quick Sight, ya que en Shiny para crear una aplicación hace falta escribir un código complejo de programación en R, en cambio la plataforma de Amazon es mucho más intuitiva y no requiere aprender lenguajes o comandos específicos.

Una vez implementadas ambas aplicaciones se han visto las ventajas e inconvenientes de cada plataforma, valorando por encima Shiny porque permite importar los datos a través de una API directamente de la base de datos de UN Comtrade, la aplicación se puede difundir sin limitaciones compartiendo su enlace que se encuentra en este trabajo, y además ofrece muchas más posibilidades en cuanto a personalización de los objetos de la App. Valorando las plataformas desde un punto de vista empresarial, se concluye que sería preferible usar Quick Sight de Amazon, porque ofrece una mayor privacidad de datos, ya que para poder acceder a las Apps se deben distribuir permisos para cada una de las aplicaciones a los usuarios que interese, de esta manera se considera mejor que Shiny para un uso interno o privado de las aplicaciones.

Para acabar, considero que la realización de este trabajo ha sido muy enriquecedora, ya que se ha tratado a fondo un tema que está experimentando su mayor crecimiento y expansión hasta la actualidad, por este motivo es interesante indagar en él ya que todavía queda mucho por conocer. Añadiendo que se ha conseguido crear una aplicación con datos reales, en mi opinión es un paso muy importante para el conocimiento de la sociedad y de la economía poder sacar información y valor de los datos a través de la estadística.

BIBLIOGRAFÍA

Ministerio de Hacienda y Administraciones Públicas, Ministerio de Industria, Energía y Turismo, Gobierno de España. Reutilización de la información del sector público. Página de datos abiertos del Gobierno, 2013. [Consulta: 23 de febrero de 2021].

Disponible en:

<https://datos.gob.es/elearning/Unidades_Didacticas/Unidad_4/contenidos/descargas/unidad_imprimible.pdf>

ONTSI, Ministerio de Asuntos Económicos y Transformación digital. Las Administraciones Públicas ante la reutilización de la información pública. Página de datos abiertos del Gobierno, 2020. [Consulta: 23 de febrero de 2021].

Disponible en:

<<https://www.ontsi.red.es/sites/ontsi/files/2020-06/LasAAPPAnteLaReutilizaci%C3%B3nEdicion2020.pdf>>

Alberto Abella de DesideAtum, con colaboración de Universidad Rey Juan Carlos y Universidad Politécnica de Valencia. La reutilización de datos abiertos en España II. DesideAtum, 2019. [Consulta: 23 de febrero de 2021].

Disponible en:

<https://www.desidedatum.com/wp-content/uploads/2019/12/La_reutilizaci%C3%B3n_datos_abiertos_en_espa%C3%B1a_2019.pdf>

Jorn Berends, Wendy Carrara, Cosmina Radu. Analytical Report 9: The Economic Benefits of Open Data. Portal Europeo de Datos, 2020. [Consulta: 12 de marzo de 2021].

Disponible en:

<https://data.europa.eu/sites/default/files/analytical_report_n9_economic_benefits_of_open_data.pdf>

Wendy Carrara, Wae San Chan, Sander Discher, Eva van Steenberg. Creating Value through Open Data: Study on the Impact of Re-use of Public Data Resources. Portal Europeo de Datos, 2015. [Consulta: 16 de marzo de 2021].

Disponible en:

<https://www.capgemini.com/consulting-fr/wp-content/uploads/sites/31/2017/08/edp_creating_value_through_open_data_final_0.pdf>

Emilio Ontiveros, Verónica López Sabater. Economía de los Datos. Fundación Telefónica, 2020. [Consulta: 20 de marzo de 2021].

Disponible en:

<<https://www.fundacioncarolina.es/wp-content/uploads/2018/11/Libro-Economia-de-los-Datos-Ontiveros.pdf>>

Esther Huyer, Laura van knippenberg. The Economic Impact of Open Data: Opportunities for value creation in Europe. Unión Europea, 2020. [Consulta: 26 de marzo de 2021].

Disponible en:

<<https://op.europa.eu/en/publication-detail/-/publication/1021d8a7-5782-11ea-8b81-01aa75ed71a1/language-en/format-PDF>>

WEBGRAFÍA

Open Knowledge Foundation. ¿Qué son los datos abiertos? Open Data Handbook, 2009. [Consulta: 25 de febrero de 2021].

Disponible en:

<<http://opendatahandbook.org/guide/es/what-is-open-data/>>

Junta de Castilla y León. ¿Qué son los datos abiertos? Datos Abiertos de Castilla y León, 2012. [Consulta: 25 de febrero de 2021].

Disponible en:

<<https://datosabiertos.jcyl.es/web/es/iniciativa-datos-abiertos/datos-abiertos.html>>

Ministerio de Hacienda y Administraciones Públicas, Ministerio de Industria, Energía y Turismo, Gobierno de España. Los negocios basados en datos abiertos en España. Página de datos abiertos del Gobierno, 2018. [Consulta: 2 de marzo de 2021].

Disponible en:

<<https://datos.gob.es/es/noticia/los-negocios-basados-en-datos-abiertos-en-espana>>

Ministerio de Hacienda y Administraciones Públicas, Ministerio de Industria, Energía y Turismo, Gobierno de España. Las administraciones públicas ante la reutilización de la información pública. Página de datos abiertos del Gobierno, 2020. [Consulta: 4 de marzo de 2021].

Disponible en:

<<https://datos.gob.es/es/documentacion/las-administraciones-publicas-ante-la-reutilizacion-de-la-informacion-publica>>

Julián Valero, Universidad de Murcia. El uso de datos abiertos en la educación: riesgos jurídicos a tener en cuenta. Página de datos abiertos del Gobierno, 2020. [Consulta: 4 de marzo de 2021].

Disponible en:

<<https://datos.gob.es/es/blog/el-uso-de-datos-abiertos-en-la-educacion-riesgos-juridicos-tener-en-cuenta>>

Escuela de Organización Industrial. Las 5 estrellas de Kim-Berners Lee. Master de Propiedad Intelectual, 2011. [Consulta: 5 de marzo de 2021].

Disponible en:

<<https://www.eoi.es/blogs/eoipi/las-5-estrellas-de-tim-berners-lee/>>

Ministerio de asuntos Económicos y Transformación Digital, Gobierno de España. El nivel de madurez de los datos abiertos en España. Portal de Administración Electrónica, 2018. [Consulta: 5 de marzo de 2021].

Disponible en:

<https://www.administracionelectronica.gob.es/pae/Home/pae_Actualidad/pae_Noticias/Anio2018/Junio/Noticia-2018-06-13-Nivel-de-madurez-datos-abiertos-Espana.html>

Ministerio de Hacienda, Gobierno de España. Reutilización de la Información (Iniciativa Datos Abiertos). Sede Electrónica de la Agencia Tributaria, 2021. [Consulta: 7 de marzo de 2021].

Disponible en:

<https://www.agenciatributaria.es/AEAT.internet/Inicio/La_Agencia_Tributaria/Gobierno_abierto/Reutilizacion_de_la_informacion_Iniciativa_Datos_abiertos_/Reutilizacion_de_la_informacion_Iniciativa_Datos_abiertos_.shtml>

Gobierno de Navarra. Open Data y Reutilización de Información del Sector Público. Gobierno abierto de Navarra, 2021. [Consulta: 7 de marzo de 2021].

Disponible en:

<<https://gobiernoabierto.navarra.es/es/open-data/que-es-open-data/open-data-y-risp>>

Gobierno de Navarra. Reutilización: Aplicaciones con datos abiertos. Gobierno abierto de Navarra, 2021. [Consulta: 7 de marzo de 2021].

Disponible en:

<<https://gobiernoabierto.navarra.es/es/open-data/reutilizacion-aplicaciones-con-datos-abiertos>>

Ministerio de la Presidencia, Relaciones con las cortes y Memoria democrática. Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público. Agencia Estatal Boletín Oficial del Estado, 2007. [Consulta: 7 de marzo de 2021].

Disponible en:

<<https://www.boe.es/eli/es/l/2007/11/16/37/con>>

Montserrat García Alsina. Reutilización de la información: retos y oportunidades. Revista de los Estudios de Ciencias de la Información y de la Comunicación, 2017. [Consulta: 8 de marzo de 2021].

Disponible en:

<<http://comein.uoc.edu/divulgacio/comein/es/numero72/articles/reutilizacion-informacion-retos-opportunidades.html>>

Durga Prasad Acharya. Las 8 mejores plataformas de inteligencia empresarial para análisis y visualización de datos. Geekflare, 2021. [Consulta: 8 de marzo de 2021].

Disponible en:

<<https://geekflare.com/es/best-business-intelligence-tools/>>

Naciones Unidas. *UN Comtrade Database*. United Nations Publications Board, 2017. [Consulta: 28 marzo de 2021].

Disponible en:

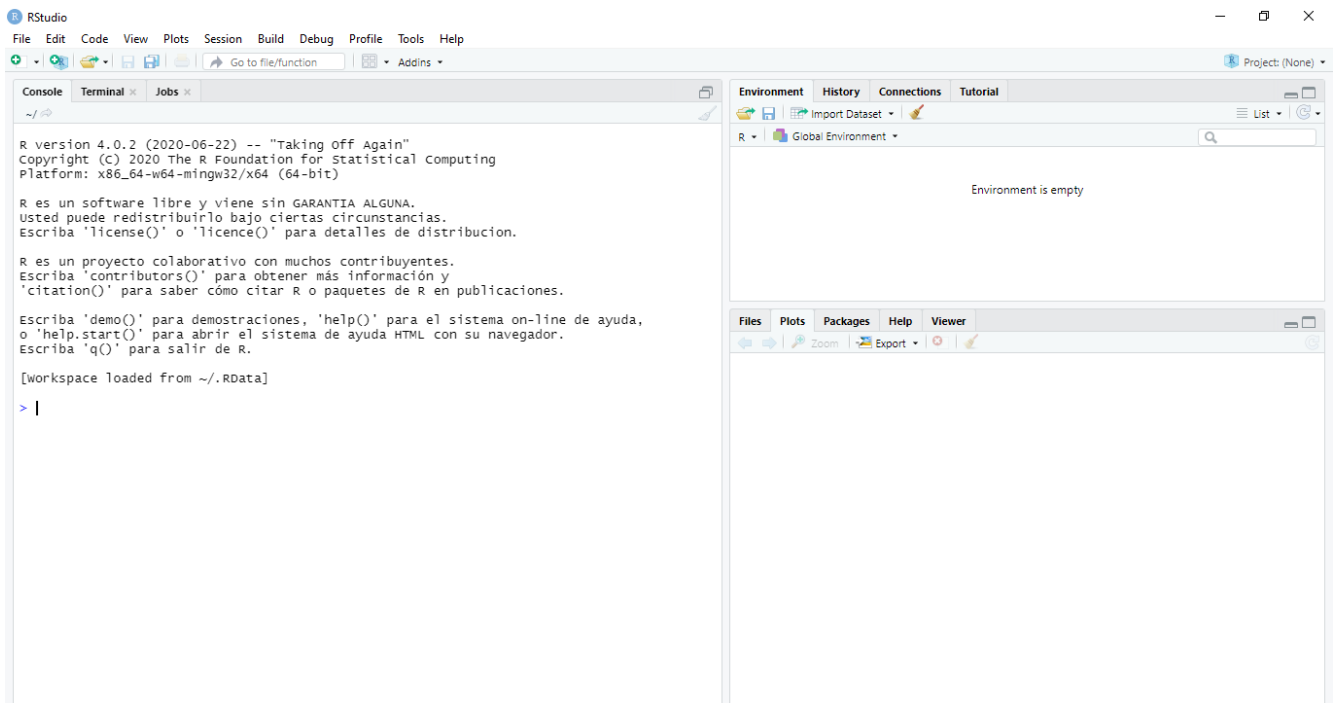
<<https://comtrade.un.org/>>

ANEXO

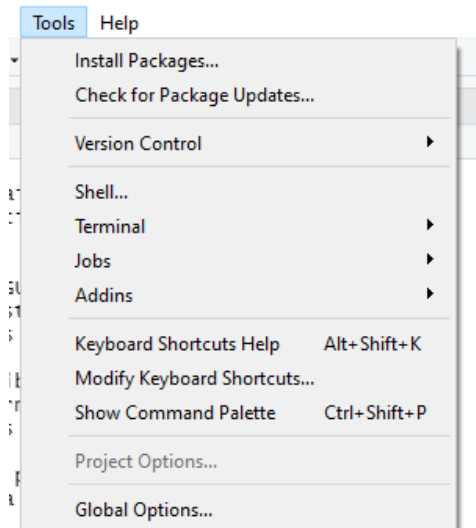
Descarga e instalación de Shiny

Dado que el programa R es un software libre, su instalación y la de sus librerías (como Shiny) son muy sencillas. Para instalar el programa de R Studio en un ordenador hace falta acceder a la siguiente página: <https://www.rstudio.com/products/rstudio/download/>. En la misma página existe un apartado donde se puede escoger que versión del programa se desea descargar, la versión que está disponible de manera gratuita y sin licencia es “R Studio Desktop”. Dependiendo del ordenador se puede seleccionar el archivo de R para Linux, Mac o Windows entre otros, a día 12/05/2021 se descargará la versión 4.0.1106 de R Studio. Cuando se complete la descarga, hace falta ejecutar el fichero .exe e instalar la aplicación siguiendo los pasos que indique la pantalla. Una vez terminada la instalación se creará un acceso directo al programa R en el escritorio del ordenador.

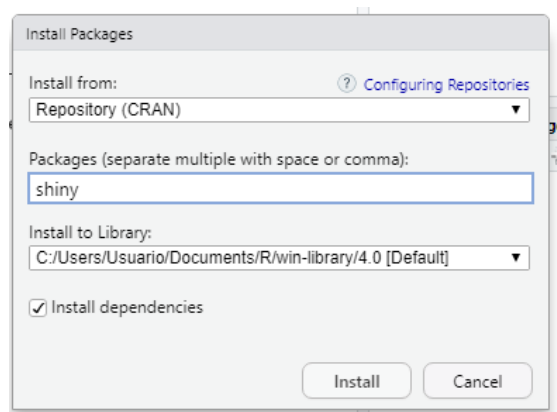
El aspecto de R Studio al abrirse es el siguiente:



El siguiente paso es descargar el paquete de Shiny, para ello hay que dirigirse a la parte superior izquierda de la pantalla en el apartado “tools” y se abrirá una ventana:



En la pestaña hace falta seleccionar “Install Packages” y se abrirá la siguiente ventana:



Se abrirá por defecto con el repositorio CRAN y una librería del ordenador donde se guardarán los ficheros. Solo hace falta escribir “Shiny” en el apartado “Packages” y darle al botón de instalar. Una vez instalado solo hará falta ejecutar el comando “library(Shiny)” antes de usar el paquete Shiny cada vez que se necesite.

Uso de Quick Sight

Para poder hacer uso de la plataforma Quick Sight de Amazon es necesario crear una cuenta de Amazon Aws. Con el siguiente enlace se accede a la página dónde el usuario puede escoger qué tipo de cuenta se quiere crear (edición de empresa o edición estándar de usuario), para llevar a cabo el trabajo se ha creado un perfil estándar. Dentro del enlace anterior, se accede a “edición estándar” y aplicando en “comenzar prueba gratuita” se abrirá una ventana para introducir los datos del usuario y crear la cuenta. Las próximas veces que se acceda a la plataforma se deberá aplicar en “inicio de sesión en la consola” (esquina superior derecha) para acceder con una cuenta ya existente.

aws

Contacte con nosotros Support Español Mi cuenta Inicie sesión en la consola

Productos Soluciones Precios Documentación Aprender Red de socios AWS Marketplace Habilitación para clientes Eventos Explorar más Q

Amazon QuickSight Información general QuickSight Q Análisis integrado Características Precios Introducción Recursos Clientes

Precios de Amazon QuickSight

Edición Enterprise [Detección de anomalías y alertas](#) **Edición Estándar** Comparación

Edición Estándar

Ideal para análisis y exploración de datos personales

Autores

Cree y publique paneles interactivos

9 USD

por usuario por mes
con suscripción anual

12 USD por usuario por mes si se paga mes a mes

10 GB de capacidad SPICE por usuario
0,25 USD por GB para agregar capacidad adicional

[Comenzar prueba gratuita](#)

Dentro de la plataforma hay tres partes importantes, la primera, la pestaña de “Conjuntos de datos”:

QuickSight 606542965768

Encuentre análisis y más

★ Favoritos

🕒 Reciente

📊 Paneles

📄 Análisis

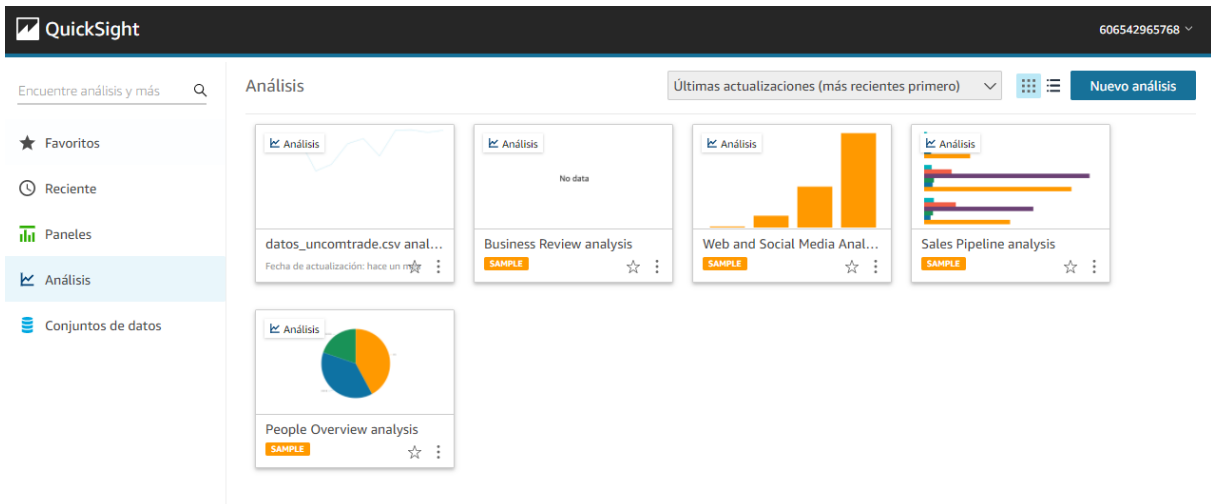
Conjuntos de datos

Nuevo conjunto de datos

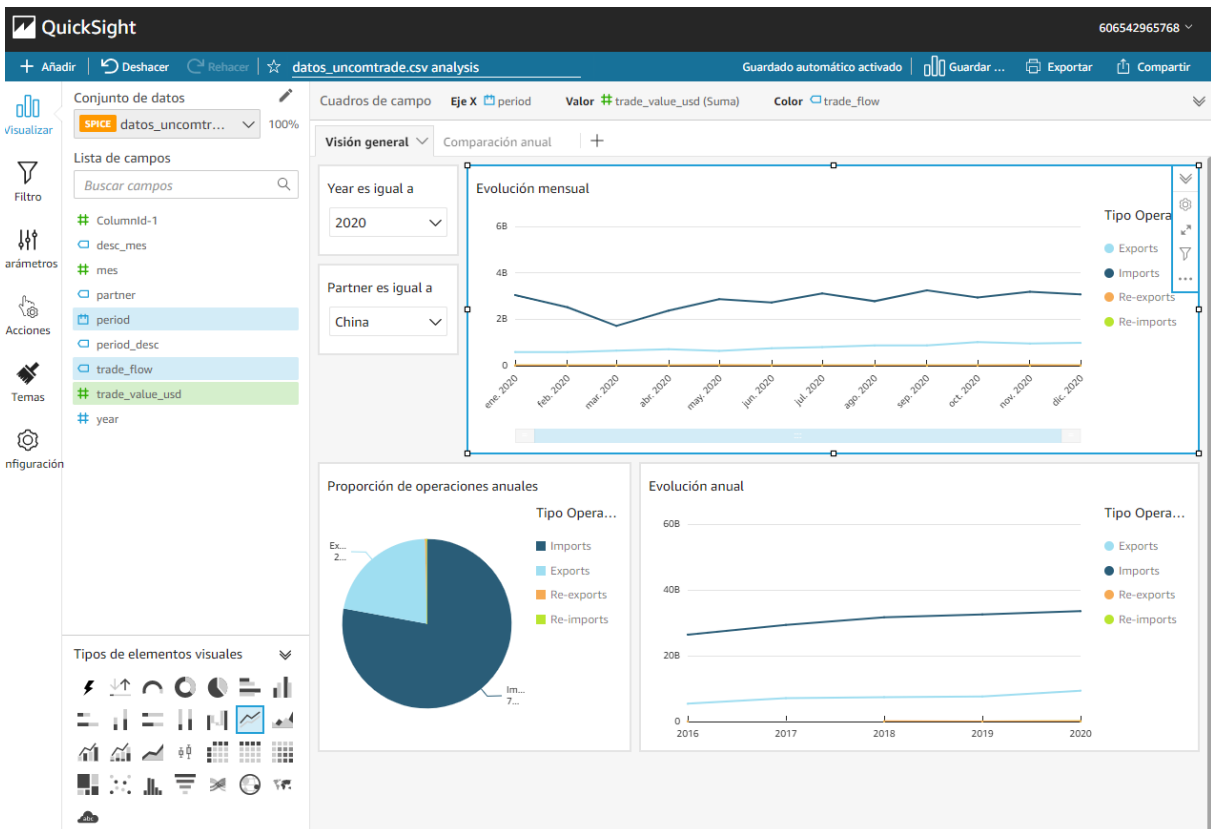
Nombre	Propietario	Última modificación
datos_uncomtrade.csv	Me	hace 2 meses
Web and Social Media Analytics	Me	hace 6 meses
Sales Pipeline	Me	hace 6 meses
People Overview	Me	hace 6 meses
Business Review	Me	hace 6 meses

En ella se encuentran todos los ficheros y bases de datos que se han importado. De los conjuntos de datos que se pueden ver en la imagen, el primero es el que hemos usado para crear la aplicación, y el resto son datos de prueba que contiene por defecto la plataforma.

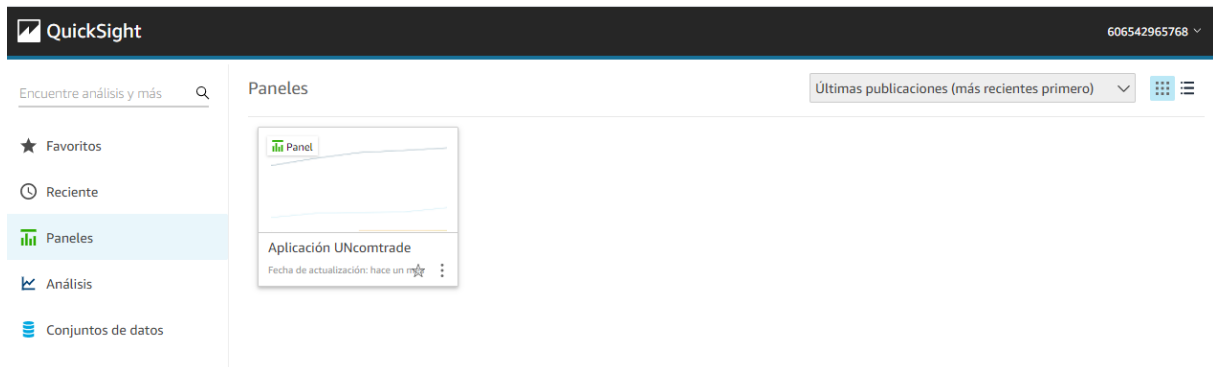
La parte más importante de la plataforma es la pestaña de “Análisis”, en ella es donde se crean las aplicaciones con los conjuntos de datos anteriores aplicando en “Nuevo análisis”:



Una vez seleccionado el conjunto de datos que se quiere analizar, se abre la pantalla de análisis, en ella se pueden añadir visualizaciones dándole a “+ añadir” (equina superior izquierda) y escoger que tipo de gráfico o indicador se quiere colocar de los que se encuentran en el recuadro “Tipos de elementos visuales”, entonces se colocan las volumetrías y dimensiones que se quieren ver en el gráfico. Una vez hecho este proceso, se pueden modificar los colores y diferentes aspectos de los gráficos aplicando en el icono de “ajustes” del gráfico que se desea editar.



Por último, cuando una aplicación ya está acabada se puede publicar en el apartado de “Paneles”, donde se ve realmente el resultado final de las aplicaciones ya que no se pueden editar, solamente visualizarla e interactuar con los datos:



Código de la aplicación en Shiny

```
library(shiny)
library(rjson)
library(RJSONIO)
library(shinydashboard)
library(comtradr)
library(shinycssloaders)
library(memoise)
library(dplyr)
library(ggplot2)
library(plotly)
library(plotrix)
library(RColorBrewer)
library(stringr)
library(DT)

partners <- comtradr:::get_country_db() %>%
  filter(partner) %>%
  .[["country_name"]] %>%
  .[, != "All"]

datos<-c()
```

```

for(i in 0:4){
  datos1<-ct_search(reporters = "Spain",
    partners = "All",
    start_date = (2020-i),
    end_date = (2020-i),
    freq="monthly",
    type="goods",
    trade_direction = "All")
  datos<-rbind(datos,datos1[,c(2,3,4,8,13,32)])
}
datos$year<-factor(datos$year)

datos1<-datos[datos$partner=="China",]
datosc<-datos1[datos1$year==2020,]
datosc<-datosc[order(datosc$period),]
datosc$period_desc<-factor(datosc$period_desc,levels=unique(datosc$period_desc))
datosc<-
aggregate(datosc$trade_value_usd,by=list(datosc$period_desc,datosc$period,datosc$trade
_flow),FUN=sum)
datosy<-
aggregate(datosc1$trade_value_usd,by=list(datosc1$year,datosc1$trade_flow),FUN=sum)
datosp<-aggregate(datosc$x,by=list(datosc$Group.3),FUN=sum)

datos4<-datosc1[datosc1$year%in%c((2020-1):2020),]
datost<-datos4[datos4$trade_flow==c("Exports"),]
datos4<-
aggregate(datost$trade_value_usd,by=list(datost$year,datost$period_desc,datost$period),F
UN=sum)
datos4<-datos4[order(datos4$Group.3),]
str_sub(datos4$Group.2[datos4$Group.1==(2020-1)],-4)<-"2020"
datos4$Group.2<-factor(datos4$Group.2,levels=unique(datos4$Group.2))

datost<-aggregate(datost$trade_value_usd,by=list(datost$partner,datost$year),FUN=sum)
datost<-
cbind(unique(datost[,1]),datost$x[datost$Group.2==(2020)],datost$x[datost$Group.2==(202
0-1)])
datost<-cbind(datost,rep(0,3-ncol(datost)),paste(round((as.numeric(datost[,2])-
as.numeric(datost[,ncol(datost)]))/as.numeric(datost[,ncol(datost)]),4)*100,"%"))

```

```

datost[,c(2,3)]<-format(as.numeric(datost[,c(2,3)]),big.mark = ".",decimal.mark = ",")
colnames(datost)<-c("Partners","Año actual","Año anterior","% Variación")

```

```

color <- brewer.pal(4, "Set2")

```

```

ui <- dashboardPage(
  dashboardHeader(title="Aplicación comercio exterior España"),
  dashboardSidebar(
    sidebarMenu(
      menuItem("Visión general", tabName = "dashboard1"),
      menuItem("Comparación anual", tabName = "dashboard2")
    ),
    selectInput("Year",
      label = "Year:",
      selected = 2020,
      choices = unique(datos$year),
      multiple = F),
    selectInput("Partner",
      label = "Partner Country:",
      selected = "China",
      choices = partners,
      multiple = TRUE),
    selectInput("Operacion",
      label = "Tipo Operación:",
      selected = "Exports",
      choices = unique(datos$trade_flow),
      multiple = T),
    actionButton("get_plot", "Plot")),
  dashboardBody(
    tabItems(
      tabItem("dashboard1",
        fluidRow( box(height=2, width = 12,
          title = "Evolución mensual",
          plotOutput("grafico1")
        )),
      fluidRow(
        box(height=2, width = 5,
          title = "Proporción de operaciones anuales",

```

```

        plotOutput("grafico3")
    ),

    box(height=2, width = 7,
        title = "Evolución anual",
        plotOutput("grafico2")
    )),
    tabItem("dashboard2",
        fluidRow( box(height=2, width = 12,
            title = "Evolución mensual",
            plotOutput("grafico4")
        ),
        fluidRow(box(height=2, width = 12,
            title = "",
            dataTableOutput("tabla")))))
    )
))

```

```

server <- function(input, output) {

  observeEvent(input$get_plot, {
    datosc1<-datos[datos$partner%in%input$Partner,]
    datosc<-datosc1[datosc1$year==input$Year,]
    datosc<-datosc[order(datosc$period),]
    datosc$period_desc<-factor(datosc$period_desc,levels=unique(datosc$period_desc))
    datosc<-
    aggregate(datosc$trade_value_usd,by=list(datosc$period_desc,datosc$period,datosc$trade
_flow),FUN=sum)
    datosy<-
    aggregate(datosc1$trade_value_usd,by=list(datosc1$year,datosc1$trade_flow),FUN=sum)
    datosp<-aggregate(datosc$x,by=list(datosc$Group.3),FUN=sum)
    datos4<-datosc1[datosc1$year%in%c((as.numeric(input$Year)-
1):as.numeric(input$Year)),]
    datost<-datos4[datos4$trade_flow%in%input$Operacion,]
    datos4<-
    aggregate(datost$trade_value_usd,by=list(datost$year,datost$period_desc,datost$period),F
UN=sum)
    datos4<-datos4[order(datos4$Group.3),]

```

```

str_sub(datos4$Group.2[datos4$Group.1==(as.numeric(input$Year)-1)],-4)<-input$Year
datos4$Group.2<-factor(datos4$Group.2,levels=unique(datos4$Group.2))
datost<-
aggregate(datost$trade_value_usd,by=list(datost$partner,datost$year),FUN=sum)
datost<-
cbind(unique(datost[,1]),datost$x[datost$Group.2==as.numeric(input$Year)],datost$x[datos
t$Group.2==(as.numeric(input$Year)-1)])
datost<-cbind(datost,rep(0,3-ncol(datost)),paste(round((as.numeric(datost[,2])-
as.numeric(datost[,ncol(datost)]))/as.numeric(datost[,ncol(datost)]),4)*100,"%"))
datost[,c(2,3)]<-format(as.numeric(datost[,c(2,3)]),big.mark = ".",decimal.mark = ",")
colnames(datost)<-c("Partners","Año actual","Año anterior","% Variación")

output$grafico1 <- renderPlot({
  ggplot(datosc, aes(x=datosc$Group.1, y=datosc$x, group =datosc$Group.3, colour
=datosc$Group.3 )) +
  labs(x="",y="Valor de las operaciones ( $ )",colour="Tipo operación")+
  geom_line(size=1) +
  geom_point( size=2, shape=21, fill="white") +
  theme_minimal())
output$grafico2 <- renderPlot({
  ggplot(datosy, aes(x=datosy$Group.1, y=datosy$x, group =datosy$Group.2, colour
=datosy$Group.2 )) +
  labs(x="",y="Valor de las operaciones ( $ )", colour="Tipo operación")+
  geom_line(size=1) +
  geom_point( size=3, shape=21, fill="white") +
  theme_minimal())
output$grafico3 <- renderPlot({
  perc<-round((datosp$x)/sum(datosp$x) *100,1)
  pie3D(datosp$x,labels=paste(perc,"%"),col=color)
  legend("bottomright", legend = c("Exports", "Imports", "Re_Exports","Re-Imports"),
  fill =color,bty="n"))
output$grafico4 <- renderPlot({
  ggplot(datos4, aes(x=datos4$Group.2, y=datos4$x, group =datos4$Group.1, colour
=datos4$Group.1 )) +
  labs(x="",y="Valor de las operaciones ( $ )",colour="")+
  geom_line(size=1) +
  geom_point( size=2, shape=21, fill="white") +
  theme_minimal())

```

```

output$tabla<-renderDataTable({
  datatable(datost)
})
})

output$grafico1 <- renderPlot({
  ggplot(datosc, aes(x=datosc$Group.1, y=datosc$x, group =datosc$Group.3, colour
=datosc$Group.3 )) +
  labs(x="",y="Valor de las operaciones ( $ )",colour="Tipo operación")+
  geom_line(size=1) +
  geom_point( size=3, shape=21, fill="white") +
  theme_minimal())
output$grafico2 <- renderPlot({
  ggplot(datosy, aes(x=datosy$Group.1, y=datosy$x, group =datosy$Group.2, colour
=datosy$Group.2 )) +
  labs(x="",y="Valor de las operaciones ( $ )",colour="Tipo operación")+
  geom_line(size=1) +
  geom_point( size=3, shape=21, fill="white") +
  theme_minimal())
output$grafico3 <- renderPlot({
  perc<-round((datosp$x)/sum(datosp$x) *100,1)
  pie3D(datosp$x,labels=paste(perc,"%"),col=color)
  legend("bottomright", legend = c("Exports", "Imports", "Re_Exports","Re-Imports"),
    fill =color,bty="n")
output$grafico4 <- renderPlot({
  ggplot(datos4, aes(x=datos4$Group.2, y=datos4$x, group =datos4$Group.1, colour
=datos4$Group.1 )) +
  labs(x="",y="Valor de las operaciones ( $ )",colour="")+
  geom_line(size=1) +
  geom_point( size=2, shape=21, fill="white") +
  theme_minimal())
})
output$tabla<-renderDataTable({
  datatable(datost)
})
})
}

```



```
shinyApp(ui = ui, server = server)
```