

2020

Proyecto Final Gestión de Datos

Programa Profesional

Analista Científico de Datos.

DAT002 – Gestión de Datos

para la Toma de Decisiones.

Prof.: José A. Cabezas Jaikel, MTBD.

Autor: Ing. Diego Araya Chaves

Versión: 1.3

Tabla de contenidos

[Tabla de Ilustraciones 3](#_Toc44094192)

[Introducción 4](#_Toc44094193)

[Descripción de las fuentes de datos 5](#_Toc44094194)

[Descripción destino de datos 7](#_Toc44094195)

[Descripción de las transformaciones realizadas 8](#_Toc44094196)

[Transformación del archivo CRI\_CSV\_GENERAL.csv 8](#_Toc44094197)

[Paso uno: obtener la dirección del archivo 8](#_Toc44094198)

[Paso uno: obtener la dirección del archivo 8](#_Toc44094199)

[Paso tres: transformación de los datos 9](#_Toc44094200)

[Transformación de los archivos WHO-COVID-19-global-data.csv y covid-19-sitrep-114.pdf 10](#_Toc44094201)

[Paso uno: obtener la dirección de los archivos 10](#_Toc44094202)

[Paso dos: establecer conexión con el archivo 11](#_Toc44094203)

[Paso tres: transformación de los datos 12](#_Toc44094204)

[Paso cuatro: Combinación de datos de los archivos WHO-COVID-19-global-data.csv y covid-19-sitrep-114.pdf 14](#_Toc44094205)

[Transformación del reportaje “Capacidad América Latina covid-19 - BBC News Mundo” 16](#_Toc44094206)

[Paso uno: obtener la dirección del archivo 16](#_Toc44094207)

[Paso dos: establecer conexión con el archivo 16](#_Toc44094208)

[Paso tres: transformación de los datos 17](#_Toc44094209)

[Paso cuatro: Combinación de datos de los archivos Tabla 1, Tabla 2, Tabla 3 21](#_Toc44094210)

[Descripción de limpieza realizada 22](#_Toc44094211)

[Limpieza del archivo CRI\_CSV\_GENERAL.csv 22](#_Toc44094212)

[Limpieza del archivo WHO-COVID-19-global-data.csv 22](#_Toc44094213)

[Limpieza del archivo covid-19-sitrep-114.pdf 22](#_Toc44094214)

[Limpieza de los archivos Tabla 1, Tabla2 y Tabla3 23](#_Toc44094215)

[Documentación de la herramienta de visualización elegida. 23](#_Toc44094216)

[Nombre de la herramienta. 23](#_Toc44094217)

[Licenciamiento / costo. 24](#_Toc44094218)

[Historia. 25](#_Toc44094219)

[Versiones. 25](#_Toc44094220)

[Ventajas de la herramienta. 26](#_Toc44094221)

[Desventajas con herramientas similares. 26](#_Toc44094222)

[Información de reportes 27](#_Toc44094223)

[Presentación del reporte: General Covid-19 CR 27](#_Toc44094224)

[Descripción del reporte 27](#_Toc44094225)

[Orígenes de datos del reporte 27](#_Toc44094226)

[Presentación del reporte: Datos de contagio según rango de edad 28](#_Toc44094227)

[Descripción del reporte 28](#_Toc44094228)

[Orígenes de datos del reporte 28](#_Toc44094229)

[Presentación del reporte: Datos de contagio según género 28](#_Toc44094230)

[Descripción del reporte 28](#_Toc44094231)

[Orígenes de datos del reporte 29](#_Toc44094232)

[Presentación del reporte: Datos de contagio según nacionalidad 29](#_Toc44094233)

[Descripción del reporte 29](#_Toc44094234)

[Orígenes de datos del reporte 29](#_Toc44094235)

[Presentación del reporte: Promedio de ocupación hospitalaria, salón vs UCI 30](#_Toc44094236)

[Descripción del reporte 30](#_Toc44094237)

[Orígenes de datos del reporte 30](#_Toc44094238)

[Presentación del reporte: Distribución de contagio global 31](#_Toc44094239)

[Descripción del reporte 31](#_Toc44094240)

[Orígenes de datos del reporte 31](#_Toc44094241)

[Presentación del reporte: Distribución de casos de contagio según país 32](#_Toc44094242)

[Descripción del reporte 32](#_Toc44094243)

[Orígenes de datos del reporte 32](#_Toc44094244)

[Presentación del reporte: Distribución de casos según región WHO 33](#_Toc44094245)

[Descripción del reporte 33](#_Toc44094246)

[Orígenes de datos del reporte 33](#_Toc44094247)

[Presentación del reporte: Casos en la región Centro Americana 34](#_Toc44094248)

[Descripción del reporte 34](#_Toc44094249)

[Orígenes de datos del reporte 34](#_Toc44094250)

[Presentación del reporte: Situación de los sistemas de salud de américa latina para enfrentar la pandemia del Covid19. 35](#_Toc44094251)

[Descripción del reporte 35](#_Toc44094252)

[Orígenes de datos del reporte 35](#_Toc44094253)

[Conclusiones 36](#_Toc44094254)

[Bibliografía. 37](#_Toc44094255)

# Tabla de Ilustraciones

[Ilustración 1:Diccionario de datos de CRI\_CSV\_GENERAL 5](#_Toc43052006)

[Ilustración 2: Diccionario de datos de WHO-COVID-19-global-data 6](#_Toc43052007)

[Ilustración 3: Diccionario de datos de covid-19-sitrep-114 6](#_Toc43052008)

[Ilustración 4: Diccionario de datos de Capacidad América Latina covid-19 - BBC News Mundo 7](#_Toc43052009)

[Tabla 1: Tipos de lincenciamientos de Power BI 22](#_Toc43052046)

# Introducción

El siguiente documento presenta el proyecto final del curso de Gestión de Datos para la Toma de Decisiones, el cual consiste en la elección de un tema en el que se quiera profundizar y realizar un análisis con las herramientas y metodologías que se aprendieron el curso, para posterior a eso presentar un documento formal de los resultados obtenidos.

Para la elaboración de este proyecto se procede a la obtención de datos de varias fuentes con un mismo tema relacionado al covid19, estos datos fueron obtenidos de diferentes fuentes públicas oficiales de internet y luego condensados en un solo repositorio para el su respectivo análisis, transformación, limpieza y presentación con aplicaciones de visualización de datos que se han visto en el curso, de esta forma desarrollar las habilidades técnicas que se necesitaran en el futuro en el análisis de datos.

Se eligió el tema de la pandemia del COVID-19 (SARS-CoV-2) porque se trata de un tema actual que ha afectado al mundo entero, por lo tanto, existe mucha información y de múltiples fuentes de donde obtener información, además de ser un tema de interés y salud pública.

# Descripción de las fuentes de datos

**Nombre de la fuente de datos:** CRI\_CSV\_GENERAL.csv

**Descripción del dataset:** Se trata de un archivo csv que contiene los datos generales del covid-19 en Costa Rica, elaborado por geovisión de las Universidad Estatal a Distancia (UNED), esta fuente de datos se actualiza cada día, por lo que para la presentación del proyecto es probable que los datos estén desactualizados, pero se pueden actualizar desde el link que se presenta.

**Link de la fuente:** <http://geovision.uned.ac.cr/oges/archivos_covid/05_19/05_19_CSV_GENERAL.csv>

**Diccionario de datos:**

[](https://github.com/diegoaraya/ProyectoGentionDatos/blob/master/Meta%20datos%20proyecto%20final.xlsx)

Ilustración 1:Diccionario de datos de CRI\_CSV\_GENERAL

Para poder tener acceso al diccionario de datos completo lo puede encontrar en:[**https://github.com/diegoaraya/ProyectoGestionDatos/blob/master/Meta%20datos%20proyecto%20final.xlsx**](https://github.com/diegoaraya/ProyectoGestionDatos/blob/master/Meta%20datos%20proyecto%20final.xlsx)

**Nombre de la fuente de datos:** WHO-COVID-19-global-data.csv

**Descripción del dataset:** Es un archivo csv que contiene los datos mundiales sobre los casos generales de contagio y defunción por corona virus, en el cual se presentan agrupaciones por región según la Organización mundial de la salud (WHO), los datos además se presentan por país. Este archivo es tomado de la WHO.

**Link de la fuente:** <https://covid19.who.int/>

**Diccionario de datos:**



Ilustración 2: Diccionario de datos de WHO-COVID-19-global-data

Para poder tener acceso al diccionario de datos completo lo puede encontrar en:[**https://github.com/diegoaraya/ProyectoGestionDatos/blob/master/Meta%20datos%20proyecto%20final.xlsx**](https://github.com/diegoaraya/ProyectoGestionDatos/blob/master/Meta%20datos%20proyecto%20final.xlsx)

**Nombre de la fuente de datos:** covid-19-sitrep-114.pdf

**Descripción del dataset:** Es un archivo pdf que contiene los datos mundiales actuales de casos de contagio de corona virus, así como de las defunciones. Se presentan también los datos totalizados acumulados de estos, pero lo que más destaca en el archivo, es que se prestan por clasificación de contagio, la fuente de este archivo es la organización mundial de la salud.

**Link de la fuente:** <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>

**Diccionario de datos:**



Ilustración 3: Diccionario de datos de covid-19-sitrep-114

Para poder tener acceso al diccionario de datos completo lo puede encontrar en:[**https://github.com/diegoaraya/ProyectoGestionDatos/blob/master/Meta%20datos%20proyecto%20final.xlsx**](https://github.com/diegoaraya/ProyectoGestionDatos/blob/master/Meta%20datos%20proyecto%20final.xlsx)

**Nombre de la fuente de datos:** Capacidad América Latina covid-19 - BBC News Mundo

**Descripción del dataset:** Se trata de un reportaje hecho por la cadena de noticias BBC, el cual expone datos de la capacidad de Latinoamérica para combatir al covid-19 de acuerdo a las cifras que se tienen de inversión en salud, equidad en el acceso a la salud y capacidad instalada en los diferentes países de la región

**Link de la fuente:** <https://www.bbc.com/mundo/noticias-america-latina-51916767>

**Diccionario de datos:**



Ilustración 4: Diccionario de datos de Capacidad América Latina covid-19 - BBC News Mundo

Para poder tener acceso al diccionario de datos completo lo puede encontrar en:[**https://github.com/diegoaraya/ProyectoGestionDatos/blob/master/Meta%20datos%20proyecto%20final.xlsx**](https://github.com/diegoaraya/ProyectoGestionDatos/blob/master/Meta%20datos%20proyecto%20final.xlsx)

# Descripción destino de datos

El repositorio de datos para este proyecto se encuentra en GitHub, la cual es una compañía sin fines de lucro que ofrece un servicio de hosting de repositorios almacenados en la nube que ayuda a los desarrolladores a almacenar y administrar su código, al igual que llevar un registro y control de cualquier cambio sobre este código.

El para acceder al repositorio es el siguiente:

<https://github.com/diegoaraya/ProyectoGestionDatos>

# Descripción de las transformaciones realizadas

En esta sección lo que se trata de explicar paso por paso cuales fueron las transformaciones que se le hicieron a los datos que están contenidos en el depositario elegido, como tenemos diversas fuentes de datos lo que haré es abordarlas una a una, de esa maneja dejar más claro lo que se hice con cada una.

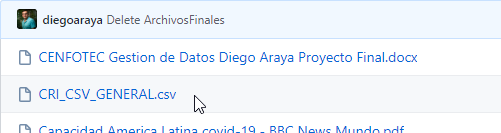
## Transformación del archivo CRI\_CSV\_GENERAL.csv

La primera transformación las voy hacer sobre el archivo CRI\_CSV\_GENERAL.csv, la descripción de esta fuente se encuentra detalla de la sección de fuentes de datos en los anteriores apartados.

Para la transformación de este archivo voy a utilizar Power Bi de Microsoft, ya me parece que se ajusta a lo que necesitamos extraer de este archivo. A continuación, presento paso a paso como se va a transformar este archivo:

Paso uno: obtener la dirección del archivo

Para esto me conecto a mi cuenta de githut he ingresamos a la misma, luego de lo cual escojo el archivo con el que voy a trabajar, y doy clic en el archivo para que se habrá.

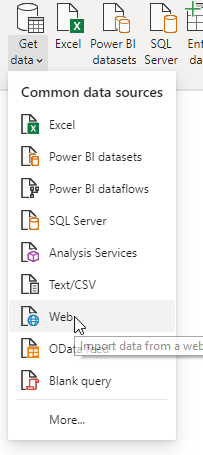
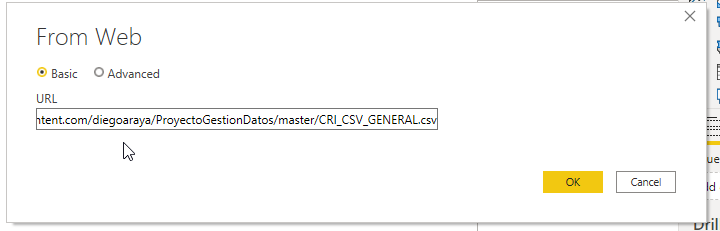


Una vez dentro del archivo procedemos a dar clic el botón que dice “RAW”, esto nos permite obtener la información del enlace donde se encuentra el archivo, para mi caso está dirección es:

<https://raw.githubusercontent.com/diegoaraya/ProyectoGestionDatos/master/CRI_CSV_GENERAL.csv>

Paso uno: obtener la dirección del archivo

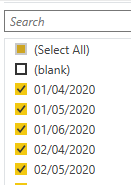
Para este Paso abrimos el programa Power BI, y voy a la sección de conexiones y hago clic en get data, y uso una que dice WEB, en caso que no este listado, se puede dar clic en “***more”*** y la busco, cuando le demos clic se abre una pantalla que solicita la dirección url donde se encuentra el archivo, en la cual copio la dirección que obtuve en el paso uno, por ultimo voy clic en ok.

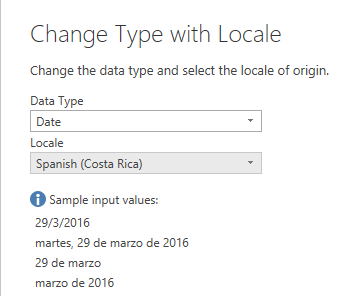


Paso tres: transformación de los datos

En este caso se me abre una pantalla que permite las transformaciones de los todos, para los cuales procedo con los siguientes pasos

1. Establecer los encabezados, para esto escojo el botón de “***use first row as hearders***” que está en la parte superior en la sección de “***transform***”
2. Borro las columnas que no voy a utilizar, lo que hago que selecciono las columnas que voy a borrar, para este caso son investig\_posi, edmin\_posi, edmax\_posi, remin\_fall, remax\_fall, remin\_UCI, remax\_UC, EDMIN\_RECUP, EDMAX\_RECUP, Muestras, MUEST\_NEW, DIA\_COVID19, afec\_posi, afec\_recu, afec\_fall, letalidad, por\_recup, transpor.
3. Filtro los espacios en blanco, dando clic sobre el botón de opciones de filtro de la columna de fecha, y quitando el check que dice blank.

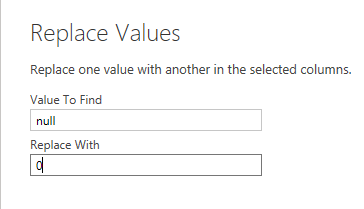


1. Establecer el tipo de dato para la columna llamada fecha, en este caso doy clic derecho sobre la columna fecha, selecciono “***Change type***”, luego selecciono el apartado de “***Using locale***”, esto lo hago por que el formato de la fecha viene dado en el formato costarricense, y en el caso de mi maquila la fecha está establecida en formato estadounidense, por lo que crea problemas a la hora de interpretar los días y los meses ya que en Costa Rica el formato es DD/MM/YYYY y el formato de Estado Unidos está establecido como MM/DD/YYYY lo que causa problemas en el manejo posterior. Cuando se abre la pantalla escojo el tipo de dato como fecha y en la locación selecciono Costa Rica.

1. Estables el tipo de datos para el resto de las columnas, para lo cual solo procedo a seleccionar el resto de las columnas, doy clic derecho y escojo el tipo de dato numérico.



1. Ahora hay algunos datos que están en Null, en cuyo caso los sustituyo por cero, para que todo tenga un poco más de coherencia, para lograr esto voy a la parte de “***transform***” y selecciono la opción de “Replace Values” que se representa de esta manera , y cuando me aparece la caja de dialogo para la sustitución de datos escojo los siguientes valores:



1. Creo otra nueva columna que se llamará *Letalidad*, para la cual escojo en la opción de “***custom column***” y me presenta un cuadro en el cual la formula sería: = [fallecidos] / [positivos], luego de lo cual procedo a cambiar el tipo de datos por el tipo porcentaje.
2. Creo otra nueva columna que se llamará *Tasa\_Recuperacion*, para la cual escojo en la opción de “***custom column***” y me presenta un cuadro en el cual la formula sería: = [RECUPERADOS] / [positivos], luego de lo cual procedo a cambiar el tipo de datos por el tipo porcentaje.

Luego de terminar las transformaciones que me parecen relevantes dentro del dataset procedo a aplicar los cambios y después a guardar el paquete de transformación en el depositario de datos que tengo disponible para estos fines, en mi caso en una carpeta almacenada en Dropbox.

## Transformación de los archivos WHO-COVID-19-global-data.csv y covid-19-sitrep-114.pdf

La segunda transformación se va a basar en los archivos WHO-COVID-19-global-data.csv y covid-19-sitrep-114.pdf, la descripción de estas fuentes se encuentra detalla de la sección de fuentes de datos en los anteriores apartados.

Para la transformación de estos archivos voy a utilizar Power Bi de Microsoft. A continuación, presento paso a paso como se van a transformar estos archivos:

Paso uno: obtener la dirección de los archivos

Para esto ingreso a mi cuenta de GitHub, luego de lo cual escojo los archivos con los que voy a trabajar para obtener la dirección RAW del archivo .csv y la dirección de descarga del .pdf, en la anterior transformación explique cómo se obtiene dicha dirección, por lo que no lo voy a mencionar de nuevo, pero las direcciones son

<https://raw.githubusercontent.com/diegoaraya/ProyectoGestionDatos/master/WHO-COVID-19-global-data.csv>

<https://github.com/diegoaraya/ProyectoGestionDatos/blob/master/covid-19-sitrep-114.pdf>

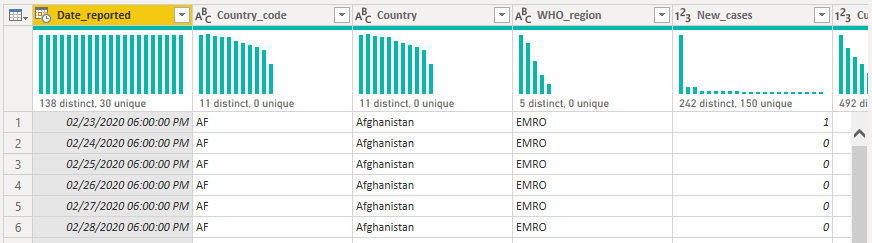
Paso dos: establecer conexión con el archivo

#### Archivo WHO-COVID-19-global-data.csv

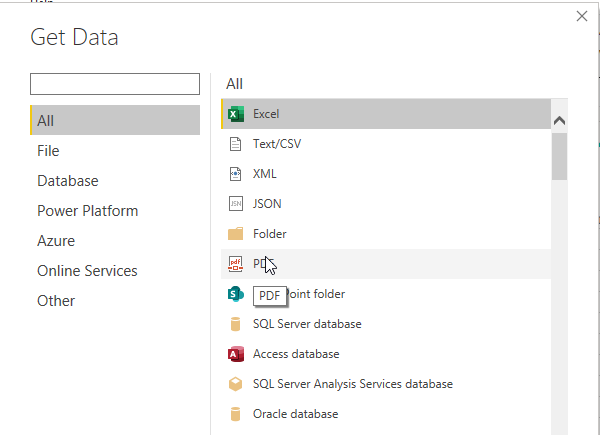
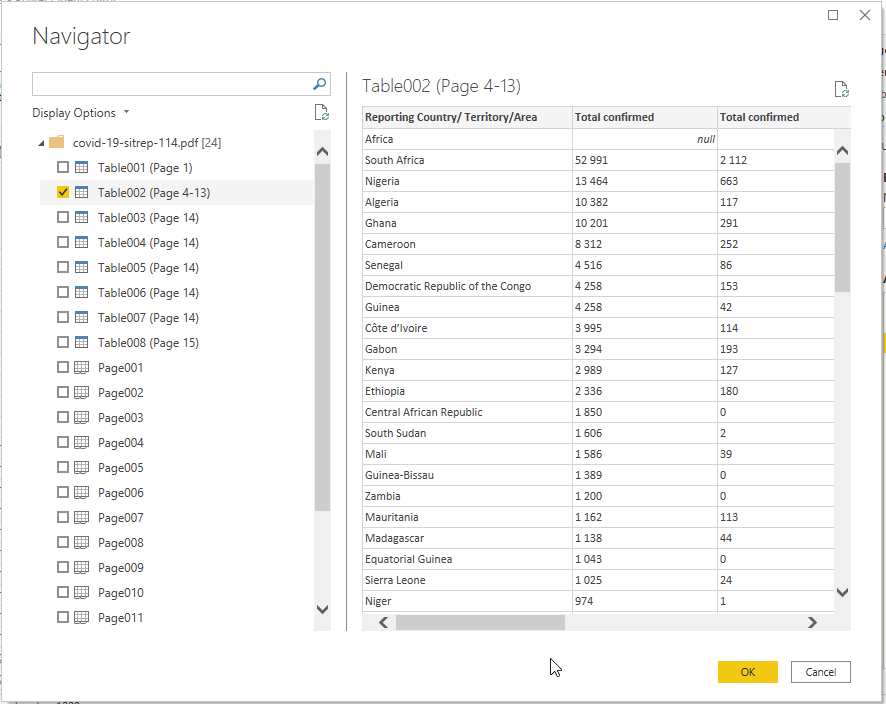
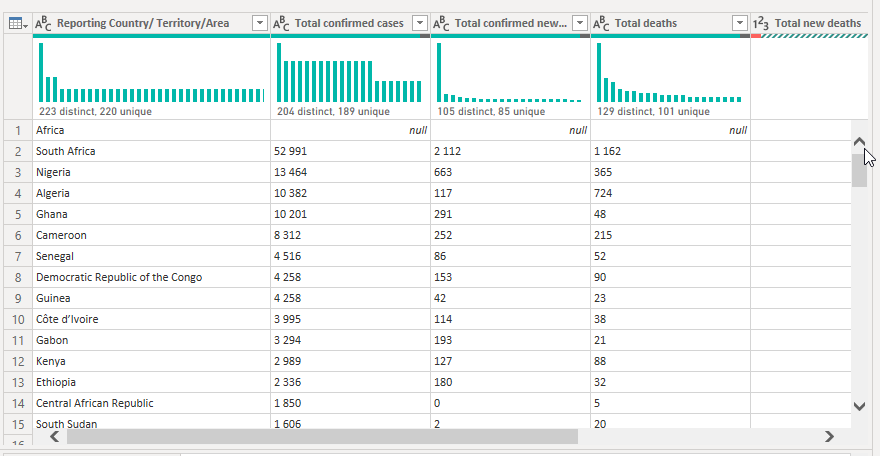
Para este Paso abrimos el programa Power BI, y voy a la sección de conexiones y hago clic en “***get data”***, y uso la conexión que dice WEB, cuando hago clic se abre una pantalla que solicita la dirección ***URL*** donde se encuentra el archivo, en la cual copio la dirección que obtuve en el paso uno, por último, doy clic en ok. Esto se hace para el archivo archivos WHO-COVID-19-global-data.csv



Si todo carga bien se presenta esta pantalla



#### Archivo covid-19-sitrep-114.pdf

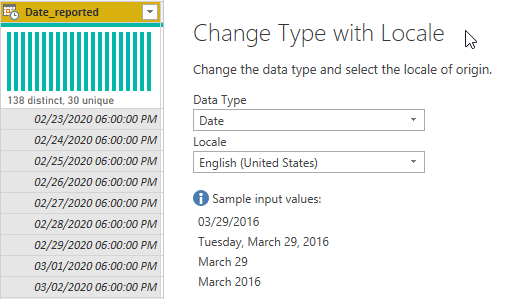
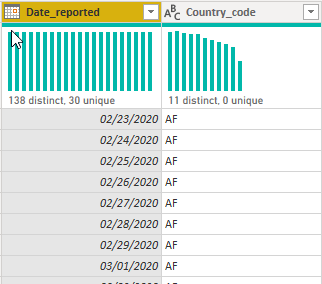
En el caso del archivo covid-19-sitrep-114.pdf, la conexión en un poco diferente, antes de establecer la conexión debo descargar el archivo a mi máquina, esto porque se trata de una archivo pdf y solo se puede establecer una conexión si se encuentra en un archivo local, igual que en el anterior voy a la sección de conexiones, pero cuando abre las opciones disponibles doy directamente en “***more***” ya que la conexión que necesito para este archivo no está disponible entre las más comunes, para esto caso en particular estoy buscando la conexión con archivos pdf, voy clic en “***connect***”, esto abre un dialogo el cual permite seleccionar el archivo desde la direcciones donde descargue el pdf. Posterior a esto se abre una ventana que permite la exploración del contenido del archivo y la visualización de los datos que contiene, que este caso selecciono solo las secciones que contienen los datos que estoy buscando y con los que voy a trabajar en la transformación de datos, en mi caso identifico que la sección que contiene los datos que deseo se llama Table002 (Page 4-13 ), la cual es la tabla con la que voy a trabajar.

Paso tres: transformación de los datos

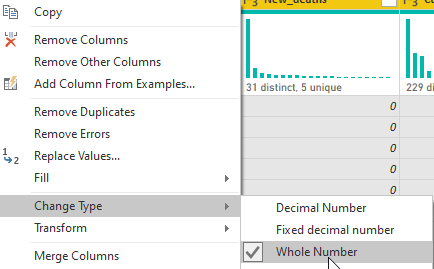
Para los dos casos se abre una pantalla con la información de los datos contenidos en los archivos, a continuación, describo los pasos que seguí para las transformaciones de los datos de los dos archivos, ya que mi objetivo es unirlos en un solo dataset.

#### Archivo WHO-COVID-19-global-data.csv

1. Establecer los encabezados, para esto escojo el botón de “***use first row as hearders***” que está en la parte superior en la sección de “***transform***”
2. Establecer el tipo de dato para la columna llamada Date\_reported, en este caso doy clic derecho sobre la columna Date\_reported, selecciono “***Change type***”, luego selecciono el apartado de “***Using locale***”, esto lo hago porque el formato de la fecha viene en formato “***data/time”*** y necesito que solo se exprese “***date***”



Posterior a eso la fecha queda en el formato que necesito, como se aprecia en los cuadros.

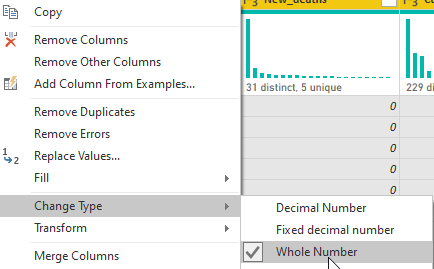
1. Cambiar las columnas que contienes valores numéricos pero que están en formato texto y los paso formato numérico, para selección las columnas New\_cases, Cumulative\_cases, New\_deaths, Cumulative\_deaths luego doy clic derecho sobre ellas y selecciono “***Change Type***” y luego “***Whole Number***”

Para este archivo son todas las transformaciones que necesita hacer ya que las columnas con pocas y los datos vienen muy limpios, por lo que no veo la necesidad de seguir transformando para los pasos que siguen.

#### Archivo covid-19-sitrep-114.pdf

1. Renombrar los nombres de las columnas, para esto es damos clic en la columna que le voy a cambiar el nombre, doy clic derecho y escojo la opción de “Rename” el detalle de los cambios de nombre es el siguiente:

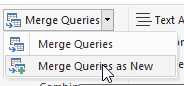
* ***Reporting Country/ Territory/Area*** se cambia por ***Country***
* ***Total confirmed cases*** se cambia por ***Total\_cases***
* ***Total confirmed new cases*** se cambia por ***Total\_new\_cases***
* ***Total deaths*** se cambia por ***Total\_deaths***
* ***Total new deaths*** se cambia por ***Total\_new\_deaths***
* ***Transmission classification^{i}*** se cambia por ***Transmission\_classification***
* ***Days since last reported case*** se cambiar por ***Days\_last\_reported***

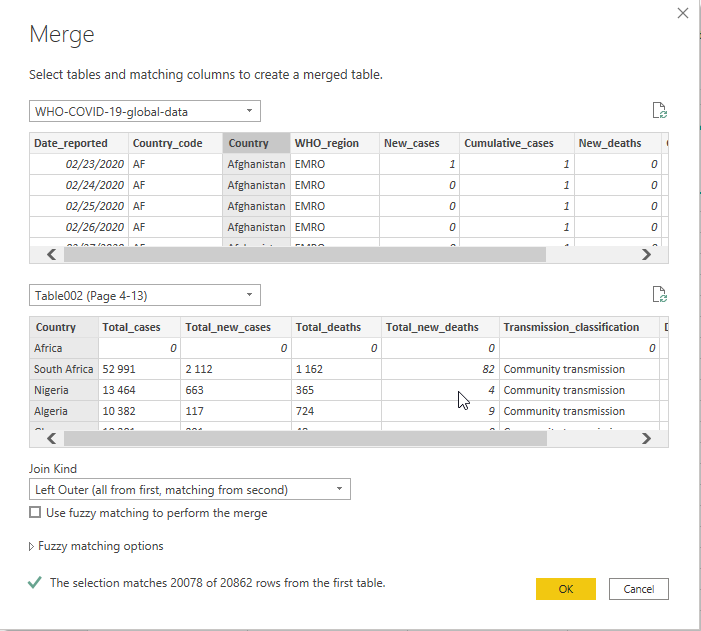
1. Ahora hay algunos datos que están en Null, en cuyo caso los sustituyo por cero, para que todo tenga un poco más de coherencia, para lograr esto voy a la parte de “***transform***” y selecciono la opción de “Replace Values” que se representa de esta manera , y cuando me aparece la caja de dialogo para la sustitución de datos escojo los siguientes valores.
2. Cambiar las columnas que contienes valores numéricos pero que están en formato texto y los paso formato numérico, para selección las columnas Total\_cases, Total\_new\_cases, Total deaths, Total\_deaths, Total\_new\_deaths, Days\_last\_reported luego doy clic derecho sobre ellas y selecciono “***Change Type***” y luego “***Whole Number***”

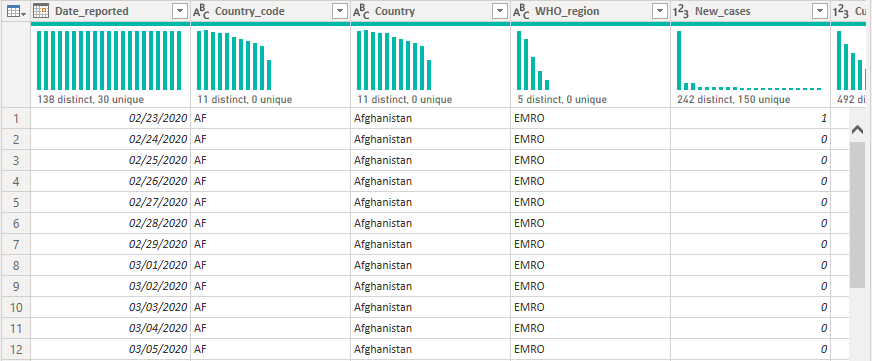
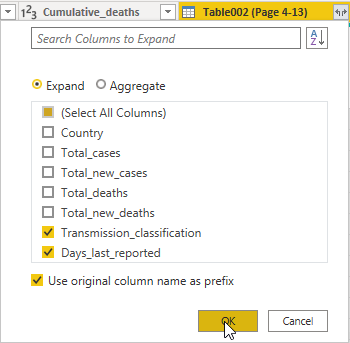
Paso cuatro: Combinación de datos de los archivos WHO-COVID-19-global-data.csv y covid-19-sitrep-114.pdf

Se procede con la combinación de los archivos ya que necesito datos que están uno y en el otro no, aunque presentan una estructura similar no contienen exactamente los mismos datos, a continuación, presento la explicación de los procesos que utilice para la combinación de estos:

1. En la ventana de transformación de datos de Power BI, en la sección de “***Home***” elijo la opción de “***Merge Queries***” y luego doy clic en la opción de “***Merge Queries as New***”, esto para tener una nueva combinación y no alterar los archivos originales.



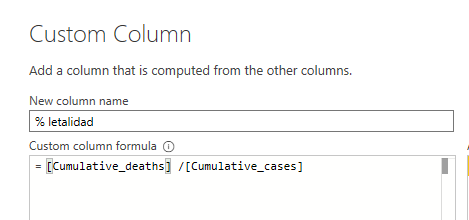
1. Esta opción abre una pantalla con un asistente para creación de la combinación, lo primero es elegir las fuentes de datos que se van a combinar, el asistente da una nuestra del contenido de ambas tablas, para este caso selección la columna Country para usar como llave primaria de indexación, esta será utilizada por el asisten para hacer la combinación entre ambas tablas, después de tener listos los datos que nos solicita el asistente doy clic en ok para continuar.
2. Esto abre una tabla con la combinación de las dos fuentes, sin embargo, el proceso no está completo, para completarlo se debe ir al final de las columnas a la derecha, aquí se nos presenta la segunda tabla que se combinó, y hay que darle clic en el botón que tiene dos flechas para extender los datos combinados, esto abre una pantalla que nos permite, seleccionar las columnas que vamos a visualizar, para este caso las columnas que estoy buscando combinar son la de transmission\_classification y la de Days\_last\_reported, mantengo seleccionada la opción de “***Expand***” y la opción de “***Use original colunm name as prefix***” y doy clic en OK.

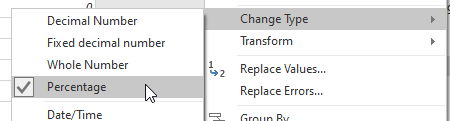


1. Doy un chequeo a los datos para saber que todos los datos están correctos y se encuentran en orden.
2. Cambio el nombre las columnas anexadas de la siguiente manera:

* ***Table002 (Page 4-13).Transmission\_classification*** por ***Transmission\_classification***
* ***Table002 (Page 4-fgt13).Days\_last\_reported*** por ***Table002 (Page 4-13).Days\_last\_reported***

1. Renombrar la tabla del resultado de la combinación por un nombre más adecuado, en mi caso Covid19\_world\_data.
2. Agrego una columna nueva, para esto selección en el menú la cejilla de “***Add Column***”, y en esta opción selección “***Custom Column***”
3. Esta opción abre una pantalla, la cual nombre “***% letalidad***”, y el código que se debe digitar en ella es ***“=[Cumulative\_deaths] /[Cumulative\_cases]***”



1. Luego doy clic en “***ok***”, esto lo que refleja es el porcentaje de letalidad del virus***.***
2. Luego doy clic derecho sobre esta columna y procedo a cambiar el tipo de dato a porcentual. 

Luego de terminar con las transformaciones de este archivo, lo que procedo hacer es a guardar el modelo y aplicar los cambios.

## Transformación del reportaje “Capacidad América Latina covid-19 - BBC News Mundo”

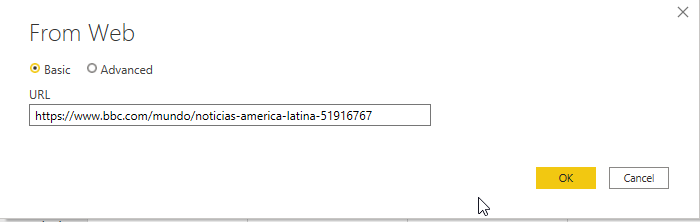
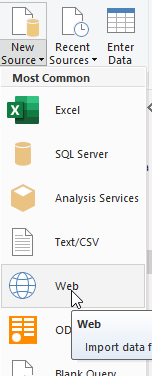
Esta fuente de datos se trata de un reportaje de BBC, que está contenido en su página web, este reportaje presenta datos interesantes sobre la capacidad instalada en los hospitales para hacer frente a la pandemia, inversión de los países en salud, y la cobertura sanitaria que tiene las personas en los diferentes países de Latinoamérica, por lo que procedo a extraer los datos que se encuentra en el mismo, ya que me parecen interesantes para mostrar.

Paso uno: obtener la dirección del archivo

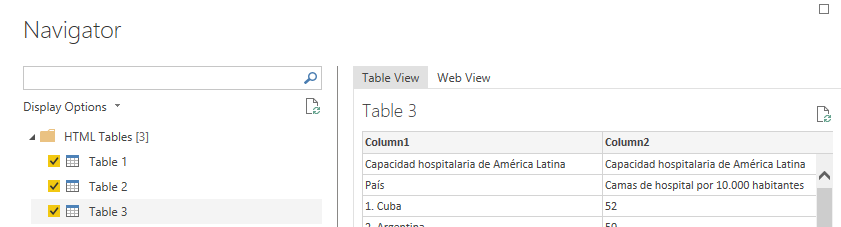
Como mencione en el apartado anterior el archivo se encuentra contenido en una página web de la BBC News, la dirección de dicho archivo es: <https://www.bbc.com/mundo/noticias-america-latina-51916767>

Paso dos: establecer conexión con el archivo

Abrimos el programa Power BI, y voy a la sección de conexiones y hago clic en get data, y uso la conexión que dice WEB, ya que este es el que se utiliza para este tipo de contenido, cuando hago clic se abre una pantalla que solicita la dirección **URL** de referencia al reportaje, en la cual copio la dirección que obtuve en el paso uno, por ultimo voy clic en ok.

Posterior a esto se abre una ventana que permite la exploración del contenido del reportaje y la visualización de los datos que contiene, que este caso selecciono solo las secciones que contienen los datos que estoy buscando y con los que voy a trabajar en la transformación de datos, en mi caso identifico que la secciones que contiene los datos que deseo se llaman **tabla 1**, **tabla 2** y **tabla 3**, la cual es la tabla con la que voy a trabajar. Este proceso nos importa las tres tablas manera individual.



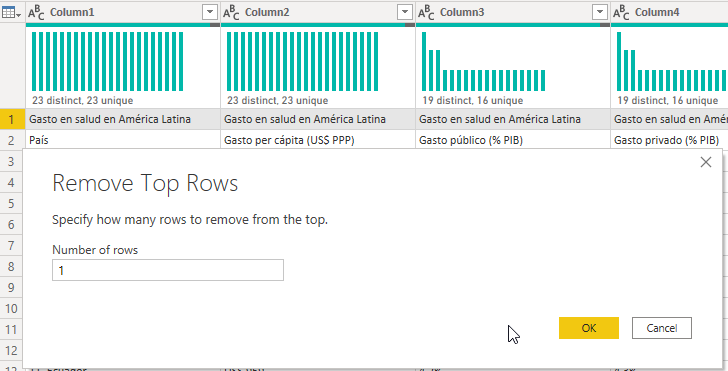
Paso tres: transformación de los datos

Una vez que tenemos las tres tablas importadas en el gestor de transformación de datos de Power BI, procedo a las transformaciones de cada una de ellas.

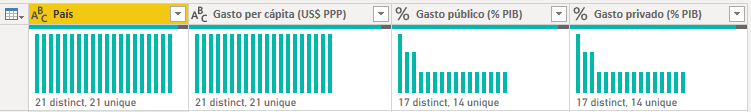
#### Tabla 1

Al importa la tabla 1 me doy cuenta que no tiene encabezados, además que la primera fila de datos es un dato referente al nombre de la tabla, por lo que lo primero que hay que hacer es empezar a limpiar la tabla y darle formato.

1. ******Primero que todo es eliminar la primera fila, ya que el dato que contiene no corresponde a el encabezado, para esto doy clic en apartado de “***remove rows”***  esta opción abre una pantalla que nos pregunta el número de la fila que deseamos eliminar, por lo digitamos el número 1, para que sea la primera fila la que quite.



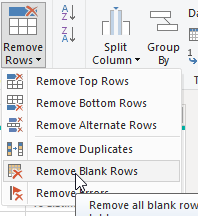
1. Establecer los encabezados, para esto escogemos la opción “Use first row as headers”, esto permite que la tabla tenga los encabezados que viene desde la fuente



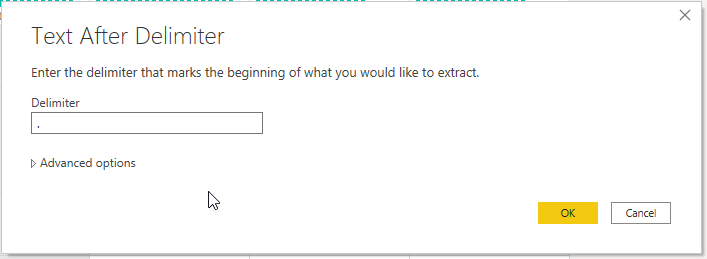
1. Cambiar el nombre de las columnas, ya que los nombres contienen caracteres especiales, para esto uso la opción de “***rename”***

* ***Gasto per cápita (US$ PPP***) por ***Gasto per cápita US$***
* ***Gasto público (% PIB)*** por ***Gasto público (PIB)***
* ***Gasto privado (% PIB)*** por ***Gasto privado (PIB)***

1. Elimino los datos que están en null que están en la última fila, para esto selecciono la opción de “***Remove Rows***” y escojo la opción de “***Remove blanck rows***”



1. Se debe quitar el valor de número de la columna de ***país***, para esto utilizo la opción de “***Extract***” y en ese menú la opción de “***Text after delimiter***”, esto abre una pantalla donde pongo lo que estoy buscando quitar, para este caso un ***punto y un espacio en blanco***.



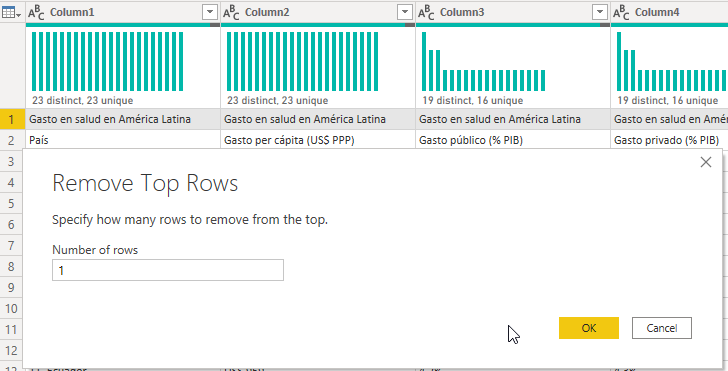
1. Aplico el mismo procedimiento, pero esta vez en la columna de ***gasto per cápita***, y busco cambiar US$
2. Le aplico a la columna de ***Gasto per cápita*** el tipo de dato de para decimal***.***
3. Procedo a darle el tipo correspondiente a cada una de las columnas de la tabla, por ejemplo, el ***gasto público*** y ***gasto privado*** son valores de porcentaje.

Con estas transformaciones la tabla ya lista para trabajar en ella.

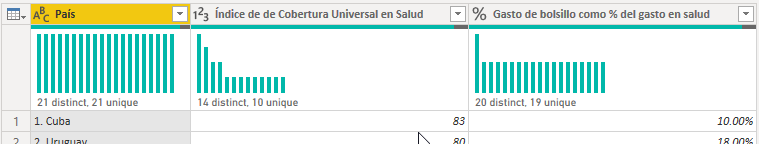
#### Tabla 2

Para la tabla 2 me doy voy a utilizar el mismo procedimiento que utilice en la primera ya que se encuentra en las mismas condiciones.

1. ******Primero que todo es eliminar la primera fila, ya que el dato que contiene no corresponde a el encabezado, para esto doy clic en apartado de “***remove rows”***  esta opción abre una pantalla que nos pregunta el número de la fila que deseamos eliminar, por lo digitamos el número 1, para que sea la primera fila la que quite.



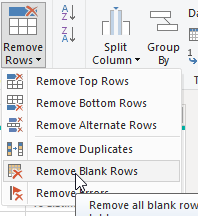
1. Establecer los encabezados, para esto escogemos la opción “***Use first row as headers***”, esto permite que la tabla tenga los encabezados que viene desde la fuente



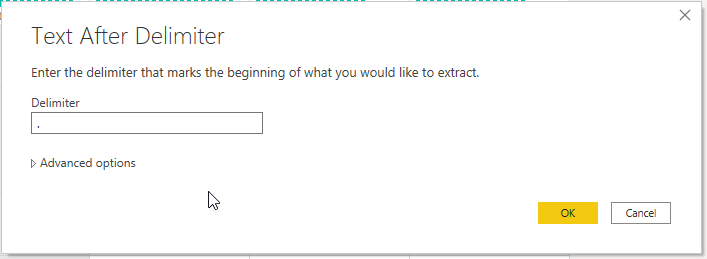
1. Cambiar el nombre de las columnas, ya que los nombres contienen caracteres especiales, para esto uso la opción de “***rename”***

* ***Índice de Cobertura Universal en Salud*** por ***Cobertura de Salud***
* ***Gasto de bolsillo como % del gasto en salud*** por ***% Gasto personal en salud***

1. Elimino los datos que están en **null** que están en la última fila, para esto selecciono la opción de “***Remove Rows***” y escojo la opción de “***Remove blanck rows***”



1. Se debe quitar el valor de número de la columna de **país**, para esto utilizo la opción de “***Extract***” y en ese menú la opción de “***Text after delimiter***”, esto abre una pantalla donde pongo lo que estoy buscando quitar, para este caso un ***punto y un espacio en blanco***.



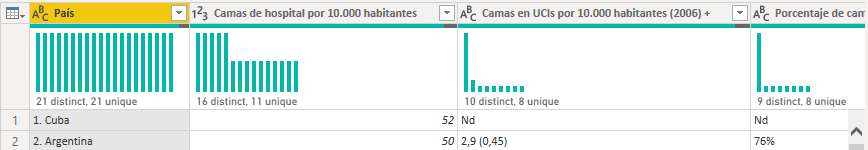
1. Procedo a darle el tipo correspondiente a cada una de las columnas de la tabla, para ***Cobertura de Salud*** es decimal y para es ***% Gasto personal en salud*** porcentaje.

Con estas transformaciones la tabla ya lista para trabajar en ella.

#### Tabla 3

Para la tabla 3 me doy voy a utilizar el mismo procedimiento que utilice en la primera ya que se encuentra en las mismas condiciones.

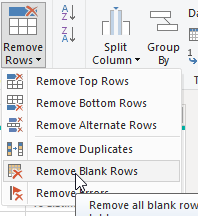
1. Primero que todo es eliminar la primera fila, ya que el dato que contiene no corresponde a el encabezado, para esto doy clic en apartado de “***remove rows”***, esta opción abre una pantalla que nos pregunta el número de la fila que deseamos eliminar, por lo digitamos el número 1, para que sea la primera fila la que quite.
2. Establecer los encabezados, para esto escogemos la opción “***Use first row as headers***”, esto permite que la tabla tenga los encabezados que viene desde la fuente



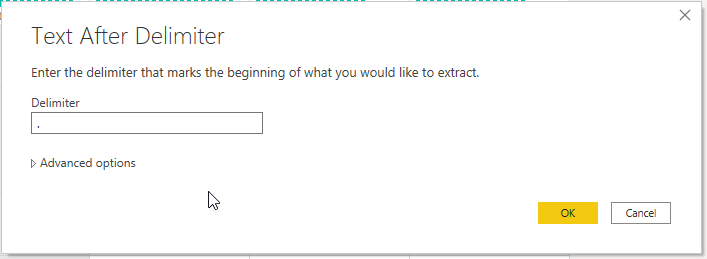
1. Cambiar el nombre de las columnas, ya que los nombres contienen caracteres especiales, para esto uso la opción de “***rename”***

* ***Camas de hospital por 10.000 habitantes*** por ***Camas hospital 10k***
* ***Camas en UCIs por 10.000 habitantes (2006) +*** por ***Camas UCIs 10.000k***
* ***Porcentaje de camas en UCIs en hospitales públicos (2006) +*** por ***Porcentaje camas UCIs H públicos***

1. Elimino los datos que están en **null** que están en la última fila, para esto selecciono la opción de “***Remove Rows***” y escojo la opción de “***Remove blanck rows***”



1. Se debe quitar el valor de número de la columna de **país**, para esto utilizo la opción de “***Extract***” y en ese menú la opción de “***Text after delimiter***”, esto abre una pantalla donde pongo lo que estoy buscando quitar, para este caso un punto y un espacio en blanco.



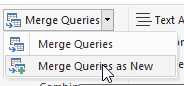
1. Quito las columnas de ***Camas UCIs 10.000k y Porcentaje camas UCIs H públicos*** ya que los datos están incompletos.

Con estas transformaciones la tabla ya lista para trabajar en ella.

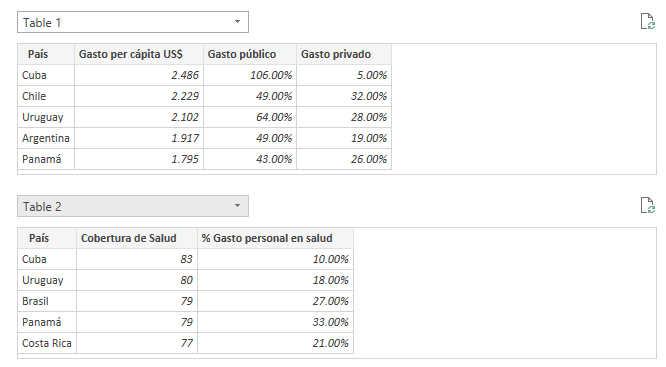
Paso cuatro: Combinación de datos de los archivos Tabla 1, Tabla 2, Tabla 3

Después de la transformación de los datos de las tablas obtenidas del artículo “**Capacidad América Latina covid-19 - BBC News Mundo**” procedo a unir estas tres tablas en una sola, para el mejor manejo de los datos.

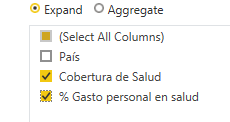
1. En la ventana de transformación de datos de Power BI, en la sección de “***Home***” elijo la opción de “***Merge Queries***” y luego doy clic en la opción de “***Merge Queries as New***”, esto para tener una nueva combinación y no alterar los archivos originales.



1. Esta opción abre una pantalla con un asistente para creación de la combinación, lo primero es elegir las fuentes de datos que se van a combinar para mi caso es la tabla 1 y tabla 2, el asistente da una nuestra del contenido de ambas tablas, para este caso selección la columna País para usar como llave primaria de indexación, esta será utilizada por el asisten para hacer la combinación entre ambas tablas, después de tener listos los datos que nos solicita el asistente doy clic en ok para continuar.



1. Esto abre una tabla con la combinación de las dos fuentes, sin embargo, el proceso no está completo, para completarlo se debe ir al final de las columnas a la derecha, aquí se nos presenta la segunda tabla que se combinó, y hay que darle clic en el botón que tiene dos flechas para extender los datos combinados, esto abre una pantalla que nos permite, seleccionar las columnas que vamos a visualizar, para este caso las columnas que estoy buscando combinar son la de **Cobertura de Salud** y la de **% Gasto personal en Salud**, mantengo seleccionada la opción de “***Expand***” y quitamos la opción de “***Use original colunm name as prefix***” y doy clic en OK.



1. Renombramos las nuevas columnas de la siguiente manera:
   1. **Sum of Cobertura de Salud** por **Cobertura de Salud**
   2. **Sum of % Gasto personal en salud** por **% Gasto personal en salud**
2. Asigno los tipos de datos para cada una de las nuevas columnas, **Cobertura de Salud** para decimal **y Gasto personal en salud** en porcentaje.
3. Repito los pasos para combinar la tabla 3 con la tabla resultante de la tabla 1 y la tabla 2.
4. La columna que voy a incorporar es este caso es **Camas hospital 10k**,la cual después voy a renombrar por Camas hospital 10k.
5. Renombro la nueva tabla como **Capadidad LA.**

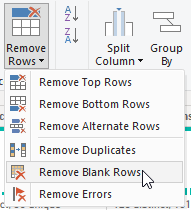
# Descripción de limpieza realizada

En cuanto a la limpieza de los archivos que se importaron en realidad estos estaban bastante limpios, sin embargo, si fue necesario implementar cierto nivel de limpieza en unos más que otros, algunas de estas limpiezas esta detalladas en la parte de transformación, sin embargo, para tener el detalle de cada una de las limpiezas procedo a mencionarlas a continuación:

## Limpieza del archivo CRI\_CSV\_GENERAL.csv

Este archivo en realidad no necesita una limpieza muy profunda, ya que los datos se encuentran bastante limpios, sin embargo, se le aplican el siguiente proceso:

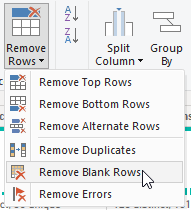
1. Se eliminan los datos en blanco o que están en null, de las últimas filas del archivo, para esto se utiliza la opción de “***Remove Rows***” y luego escojo la opción de “***Remove Black Rows***”



## Limpieza del archivo WHO-COVID-19-global-data.csv

Igual que el archivo anterior, este se encuentra bastante limpio, por lo que igual que el otro se le aplica un proceso.

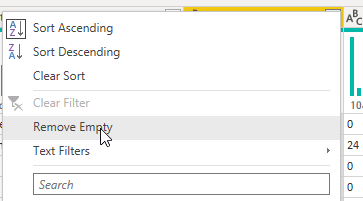
1. Se eliminan los datos en blanco o que están en null, de las últimas filas del archivo, para esto se utiliza la opción de “***Remove Rows***” y luego escojo la opción de “***Remove Black Rows***”



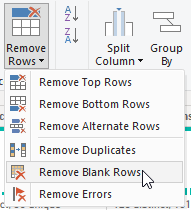
## Limpieza del archivo covid-19-sitrep-114.pdf

Para este archivo le aplico dos procesos ya que el presenta unos encabezados que agrupación que se importaron junto con los datos, además de datos en blanco, esto son los pasos que se realizaron:

1. Selección la columna de "***Confirmed cases***”, doy clic en la parte de filtros, selecciono la opción de “***Remove Empty***”, esto permite que los datos que estén en nulo se eliminen de los datos



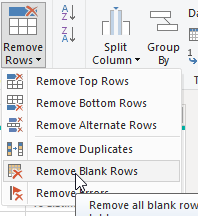
1. Se eliminan los datos en blanco o que están en null, de las últimas filas del archivo, para esto se utiliza la opción de “***Remove Rows***” y luego escojo la opción de “***Remove Black Rows***”



## Limpieza de los archivos Tabla 1, Tabla2 y Tabla3

Estos archivos presentaron pocas opciones de limpieza, al tener pocas columnas y pocas filas solo se procedió con la siguiente limpieza para los tres:

1. Elimino los datos que están en null que están en la última fila, para esto selecciono la opción de “***Remove Rows***” y escojo la opción de “***Remove blanck rows***”



1. Quito las columnas de ***Camas UCIs 10.000k y Porcentaje camas UCIs H públicos*** ya que los datos están incompletos.

# Documentación de la herramienta de visualización elegida.

Para la elección de la herramienta se evaluaron varias que existen en el mercado y después de un análisis, que tomo en cuenta, la facilidad de uso, la documentación existente, apoyo de la comunidad de usuarios, integración de la visualización con la gestión de los datos, el costo, manejo de múltiples fuentes de importación de datos, llego a la decisión de escoger la siguiente herramienta:

## Nombre de la herramienta.

La herramienta que se selecciona para la gestión de los datos y para la visualización de estos es Power BI de Microsoft, Power BI es plataforma de servicios de análisis empresarial basados en la nube, su objetivo es proporcionar visualizaciones interactivas y capacidades de inteligencia empresarial (en inglés: **business intelligence o BI**) con una interfaz lo suficientemente simple como para que los usuarios finales creen sus propios informes y paneles.

## Licenciamiento / costo.

Según se puede encontrar en la página de Power BI, existen tres tipos de licenciamientos, los cuales se describen a continuación:

|  |  |  |
| --- | --- | --- |
| Tipos de Licencia Power BI | | |
| Free | Pro | Premiun |
| 1. Incluye 1 GB de almacenamiento en Modo Import. | 1. Toda la funcionalidad de **Power BI** | 1. Tiene un costo base mínimo a cubrir |
| 2. No se pueden compartir con otros usuarios (reportes, dashboards y datasets). | 2. Costo Mensual de $ 10 USD por Usuario / Mes (más impuestos aplicables) | 2. Está basado en cantidad de recursos a utilizar |
| 3. Para compartir reportes solo es posible en modo público. | 3. Incluye 10 GB de Almacenamiento en Modo Import | 3. Se recomienda esta opción a partir de 500 usuarios (Costo-Beneficio) |
| 4. No incluye Dataflows | [4. Se requiere suscripción anual, más información en ¿Cuánto cuesta Power BI?](https://signup.microsoft.com/create-account/signup?products=CFQ7TTC0L3PB:0016&fSkuId=powerbi&fOfferId=03FB6305-216A-410F-856B-FD6142D94604&ru=https:%2F%2Fapp.powerbi.com%2F&pbi_source=webpurchase&scenarioID=Allotments&cmpid=pbiweb-pricing) | 4. Incluye adicionalmente licenciamiento de **Power BI Report Server** (On-Premise) |
|  | Tiene la opción de probar una versión de introducción por 60 días | 5. Parte de la suscripción de Office 365 |
|  |  | [6. Se requiere de suscripción mensual, más información en Calculadora de Power BI Premium](https://powerbi.microsoft.com/es-es/calculator/) |

Tabla 1: Tipos de lincenciamientos de Power BI

## Historia.

Esta aplicación fue concebida originalmente por Thierry D'Hers y Amir Netz del equipo de SQL Server Reporting Services en Microsoft. ​Fue diseñado originalmente por Ron George en el verano de 2010 y nombrado Proyecto Crescent. Project Crescent estaba disponible inicialmente para su descarga pública el 11 de julio de 2011 incluido con SQL Server Codename Denali. Más tarde renombrado a Power BI, Microsoft lo dio a conocer en septiembre de 2013 como Power BI para Office 365. La primera versión de Power BI se basó en complementos de Microsoft Excel: Power Query,Power Pivot y Power View. Con el tiempo, Microsoft también agregó muchas características adicionales como preguntas y respuestas, conectividad de datos de nivel empresarial y opciones de seguridad a través de las puertas de enlace de Power BI. Power BI fue lanzado por primera vez al público en general el 24 de julio de 2015.

## Versiones.

Las diferentes versiones claves del ecosistema de Power BI se encuentran las siguientes:

* **Power BI Desktop** La aplicación basada en escritorio de Windows para equipos y escritorios, principalmente para diseñar y publicar informes en el servicio.
* **Power BI Service** El servicio en línea basado en SaaS (software como servicio) (anteriormente conocido como Power BI para Office 365, ahora denominado PowerBI.com o simplemente power BI).
  + **Power BI Pro:** licencia pagada por usuario, necesaria para obtener acceso a funciones avanzadas y la capacidad de compartir informes.
  + **Power BI Premium:** licencias por escala, destinadas a grandes empresas.
* **Power BI Mobile Apps** Las aplicaciones de Power BI Mobile para dispositivos Android y iOS, así como para teléfonos y tabletas Windows.
* **Power BI Gateway** Puertas de enlace que se usan para sincronizar datos externos dentro y fuera de Power BI. En el modo empresarial, también pueden usar los flujos y PowerApps en Office 365.
* **Power BI Embedded** La REST API de Power BI se puede usar para crear paneles e informes en las aplicaciones personalizadas que sirven a los usuarios de Power BI, así como a los usuarios que no son de Power BI.
* **Power BI Report Server** Una solución de informes de Power BI local para empresas que no almacenan o no los datos en el servicio Power BI basado en la nube. Power BI Visuals Marketplace
* **Power BI Visuals Marketplace** Un mercado de objetos visuales personalizados y objetos visuales con tecnología R.

## Ventajas de la herramienta.

Cada día más las empresas necesitan más poder de generación de informes y fuerza analítica, Con Power BI las empresas pueden reunir, analizar y visualizar datos de toda la compañía, brindándoles una mejor perspectiva de sus operaciones y rendimiento, y permitiéndoles tomar decisiones más informadas basadas en datos reales. Estos son algunos de los beneficios clave de la herramienta son:

* Tiene una interfaz intuitiva que lo hace mucho más fácil de usar y fácil de navegar.
* Tiene integrado el gestor de datos ETL con la herramienta de visualización.
* La plataforma se integra con otras herramientas de gestión empresarial como SharePoint , Office 365y Dynamics 365.
* Tiene acceso a múltiples fuentes de datos como Salesforce, GitHub, QuickBooks, ZenDesk, Marketo, Facebook y Google Analytics, entre otras y además está integrado al ecosistema de herramientas de Microsoft.
* Se pueden gestionar grandes cantidades de datos.
* Gran cantidad de documentación de auto aprendizaje y una comunidad de usuarios muy activa, que crean gran cantidad de documentación
* Tiene soporte para R y para Python.
* Las funciones integradas de aprendizaje automático pueden analizar datos y ayudar a los usuarios a detectar tendencias valiosas y hacer predicciones informadas.
* La información se puede visualizar mediante poderosas plantillas para entender mejor los datos.
* Power BI está basado en la nube, por lo que los usuarios obtienen capacidades de inteligencia de vanguardia y algoritmos potentes que se actualizan periódicamente.
* Las potentes capacidades de personalización permiten a los usuarios crear paneles de control para que puedan acceder a los datos que necesitan rápidamente.
* Las alertas se pueden configurar en KPI para mantener a los usuarios actualizados métricas y mediciones importantes.
* Power BI garantiza que los datos estén seguros, ofreciendo controles de accesibilidad tanto interna como externa.

## Desventajas con herramientas similares.

* Es necesario subir el proyecto a la nube para poder compartirlo.
* Límite de tamaño de archivos 2 GB, septo para la versión no Premium.
* No maneja bien las relaciones complejas entre tablas, por ejemplo, si tiene más de una llave única.
* No proporciona muchas opciones para configurar a sus visualizaciones. lo que provoca que los usuarios tengan opciones limitadas para lo que pueden cambiar estas.
* Algunos usuarios opinan que la interfaz de usuario de Power BI a menudo se encuentra abarrotada y voluminosa. Es en el sentido de que hay muchos íconos de opciones que bloquean la vista del tablero o informe.
* Por lo general, Power BI es la herramienta de BI más fácil de usar si la está utilizando simplemente para importar datos y crear informes. Pero Power BI es un conjunto completo que tiene muchas otras herramientas interrelacionadas. Cuando el propósito de su uso es más que crear informes en Power BI Desktop, debe aprender y dominar otras herramientas como Gateways, Power BI Report Server, Power BI Services, etc.

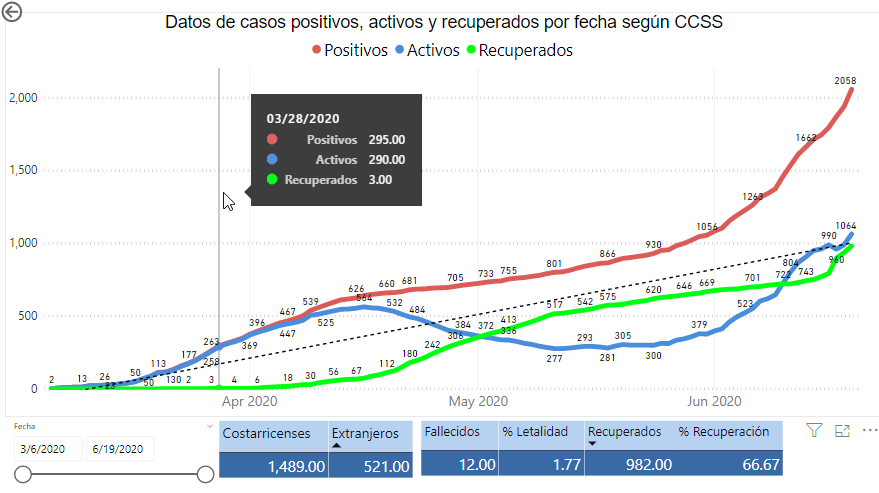
# Información de reportes

Unas ves que se ha hecho la carga, transformación y limpieza de los datos, así como la escogencia de la herramienta de visualización, procedo a presentar y explicar cada uno de los gráficos que se generaron:

## Presentación del reporte: General Covid-19 CR

### Descripción del reporte

Este reporte presenta un gráfico de líneas del total de los casos positivos, activos y recuperados a lo largo del tiempo desde que se presentó el primer caso en Costa Rica a la fecha actual. En él se presenta en cada línea los valores correspondientes a cada ítem, rojo para casos positivos, azul para los casos activos y verde para los casos recuperados, con una granularidad de 75 %, además de presentar el número de casos correspondiente a cada línea en la parte superior, presenta además una línea de tendencia, que para este gráfico indica que tiene una tendencia al alza.



Además, en la parte inferior presento un filtro, en el cual se puede seleccionar un rango de fechas para que se muestren los datos correspondientes en el gráfico, este por defecto esta de la fecha de inicio de la pandemia en Costa Rica, a la fecha actual, aparte de esto acompaño el gráfico con un resumen a manera de tabla de los casos confirmados correspondientes a costarricenses y a extranjeros, y otra tabla resumen del total de fallecidos, el porcentaje de letalidad del virus en Costa Rica (número de fallecidos entre el total de casos confirmados), el número total de recuperados a la fecha y el porcentaje de recuperación (número casos recuperados entre el total de casos confirmados).

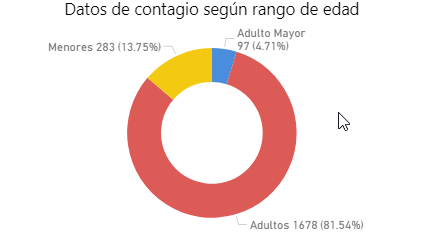
### Orígenes de datos del reporte

El reporte tiene como origen el dataset llamado CRI\_CSV\_GENERAL, el cual se explica a profundidad en la sección de *Descripción de las fuentes de datos*, y las columnas que se usaron para el mismo fueron, la Fecha, Positivos, Activos, Recuperados. Para el filtro se utiliza la columna fecha y para los datos de adicionales de las tablas se utiliza el máximo de las columnas de Costar\_posi, Extranj\_posi, fallecido, letalidad, recuperados y Tasa\_Recuperacion

## Presentación del reporte: Datos de contagio según rango de edad

### Descripción del reporte

Este reporte presenta una gráfica de dona, la cual representa los datos correspondientes a la agrupación según el rango edad, de acuerdo como los está agrupando el ministerio de salud, en el cual destacan tres grandes grupos, Adulto Mayor, Adultos y menores de edad, en la gráfica se presenta por cada grupo el número de casos positivos y el porcentaje respectivo. Escojo este reporte ya que para representar este tipo de datos es el que menor se acopla.



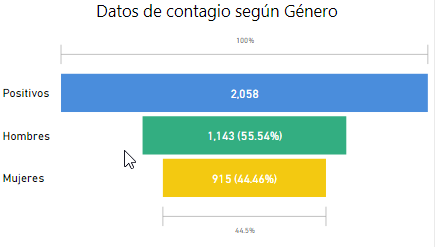
### Orígenes de datos del reporte

El reporte tiene como origen de datos el dataset llamado CRI\_CSV\_GENERAL, y las columnas que se usaron para el mismo fueron, adul\_posi, am\_posi, menor\_posi, la fuente de los mismo se describe en la sección de orígenes de datos.

## Presentación del reporte: Datos de contagio según género

### Descripción del reporte

Este gráfico se trata de una gráfica de Funnel o de embudo, el cual presenta el número de casos positivos por contagio del virus, según el género, en la parte superior del embudo está el total de número de casos de contagio, y después este se subdivide en el total de casos por hombres y por mujeres, presentando el número respectivo para cada género y el porcentaje de cada uno de ellos, si bien es cierto este tipo de gráficos se utiliza para representar procesos, me pareció interesante presentar los datos de esta forma, ya que para la clasificación de género deja muy claro las cantidades de cada uno.



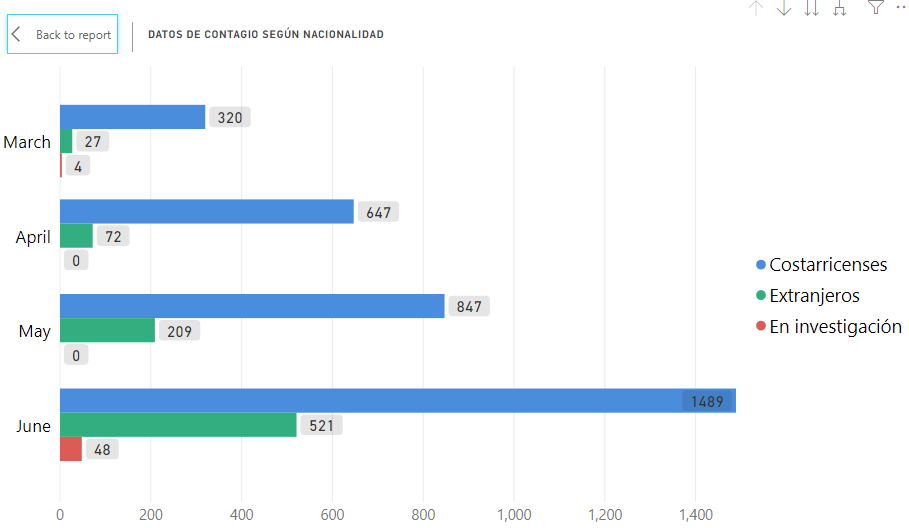
### Orígenes de datos del reporte

El reporte tiene como origen de datos el dataset llamado CRI\_CSV\_GENERAL, y las columnas que se utilizaron en este son el máximo de casos positivos(positivos), máximo de hombres positivos(hom\_posi) y por último el máximo de mujeres (muj\_posi).

## Presentación del reporte: Datos de contagio según nacionalidad

### Descripción del reporte

Este se trata de un gráfico de barras horizontales, el cual presenta datos resumen según procedencia de los casos de contagio, ya sean nacionales o extranjeros, además otra clasificación que denomina “en estudio”, en la cual esta agrupadas las personas de la cuales no se tiene información de la nacionalidad o está en proceso de averiguación de la misma. La información presentada se agrupa por mes, para tener una visión de cuál ha sido el comportamiento a lo largo de los meses que llevamos con el virus. En el reporte de se muestra el número de contagiados según su categoría por cada uno de los meses, este tipo de grafica permite ver de manera ordenada y simplificada la información.



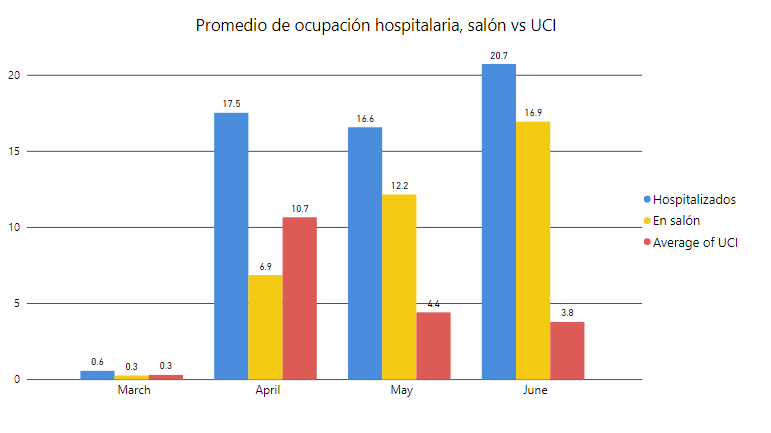
### Orígenes de datos del reporte

El reporte tiene como origen de datos el dataset llamado CRI\_CSV\_GENERAL, y las columnas que se utilizaron para este reporte son Máximo de Costarricenses positivos (costar\_posi), máximo de casos de extranjeros positivos (extranj\_posi), y el máximo de casos en investigación (investig\_posi).

## Presentación del reporte: Promedio de ocupación hospitalaria, salón vs UCI

### Descripción del reporte

Este se trata de una gráfico de barra verticales, el cual presenta el promedio de la ocupación hospitalaria desglosado por mes, esta ocupación es presentada como total de pacientes hospitalizados con el covid19, y desglosa cuántos de estos se encuentran en un salón especializado y cuales se encuentra en la unidad de cuidados intensivos (UCI), en la parte superior de cada barra se presenta el promedio de acuerdo al mes, por lo que se puede notar en los últimos meses el promedio de persona en la UCI va en disminución.



### Orígenes de datos del reporte

El reporte tiene como origen de datos el dataset llamado CRI\_CSV\_GENERAL, y las columnas que se utilizaron en este reporte son el promedio de personas hospitalizadas (hospital), promedio de personales en salón (salon), y el promedio de personas en la UCI (UCI). Casos positivos

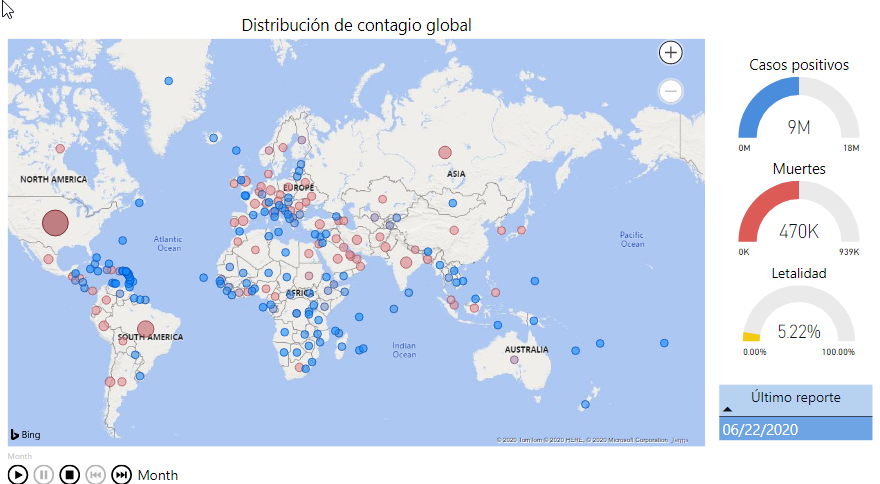
## Presentación del reporte: Distribución de contagio global

### Descripción del reporte

Este se trata de un gráfico de mapa, en el cual se muestra la distribución global del contagio por el virus, se muestra por cada país una esfera, la cual representa de acuerdo a su tamaño la cantidad de casos positivos, esta esfera además está representando de color azul donde hay menos casos y de se va degradando a color rojo más intenso con forme va creciendo la cantidad de positivos, también el grafico tiene un menú emergente con el resumen de positivos y muertes según se escoja en el mapa el país que se quiere ver. En el lado derecho se muestra un resumen con la totalidad de casos, las muertes y la taza de fatalidad a nivel mundial y un cuadro resumen con la fecha de la última actualización de los datos.

En la para inferior se muestra un secuenciador que funciona a manera de filtro, mostrando en el mapa y en los resúmenes, los datos por mes.

Se escoge este reporte ya sé que presenta de una manera fácil y bastante gráfica, cual es la situación de contagio a nivel mundial.



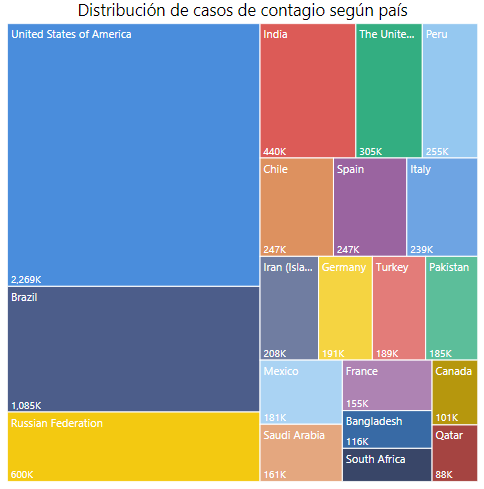
### Orígenes de datos del reporte

El reporte tiene como origen de datos el dataset llamado **Covid19\_world\_data**, y las columnas que se utilizaron en este reporte para el mapa son los casos acumulativos positivos(aculuative\_cases), casos acumulativos de muertes (acumulative\_deaths), el país (country), para el resumen de casos positivos se utiliza los datos de la suma de los nuevos casos (new\_cases), para el resumen de muertes la suma de las muertes (new\_deaths), para el resumen de la taza de letalidad se usa una variable es que la sum de la totalidad de muertes entre la totalidad de contagio. Y por último para el filtro se usa la fecha del reporte de casos (date\_reported)

## Presentación del reporte: Distribución de casos de contagio según país

### Descripción del reporte

Este es un gráfico de mapa árbol, el cual muestra un top 20 según los casos positivos alrededor del mundo, este grafico presenta el nombre del país y el número positivos, entre mayor sea el número de contagiados más grade será el cuadro que se presenta, teniendo como el mayor de todos a Estado Unidos de Norte América, seguido de Brasil y de tercero Rusia, este tipo de grafico permite ver de una manera muy clara la distribución de contagios.



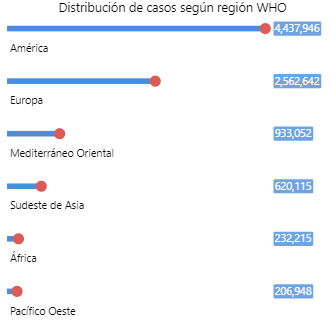
### Orígenes de datos del reporte

El reporte tiene como origen de datos el dataset llamado **Covid19\_world\_data**, y las columnas que se utilizaron en este reporte son la de país (country), el número de casos positivos acumulativos (acumulative\_Cases), y para presentar en la leyenda el número de cases de fallecidos (acumulative\_deaths).

## Presentación del reporte: Distribución de casos según región WHO

### Descripción del reporte

Este es un gráfico de barras horizontales, que presenta el total de casos de contagio según región, con base de las agrupaciones que tiene la organización mundial de la salud (WHO por sus siglas en ingles), como se aprecia en el grafico la región que mayor cantidad de contagios es América.



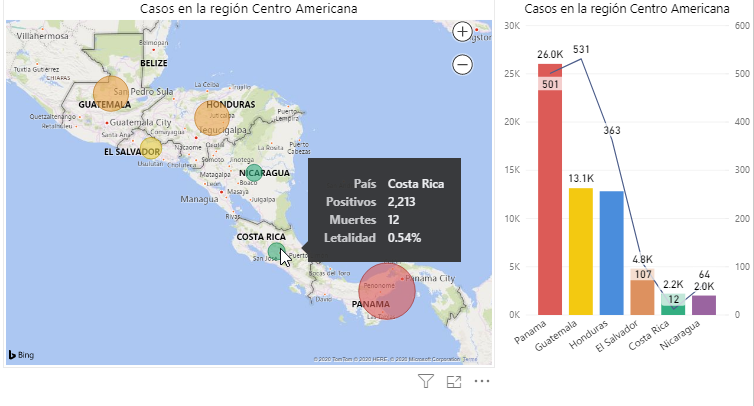
### Orígenes de datos del reporte

El reporte tiene como origen de datos el dataset llamado **Covid19\_world\_data**, y las columnas que se utilizaron en este reporte son la de país (country), el número de casos positivos acumulativos (acumulative\_Cases), y para presentar en la leyenda el número de cases de fallecidos (acumulative\_deaths).

## Presentación del reporte: Casos en la región Centro Americana

### Descripción del reporte

Este grafico es un gráfico de mapa acompañado de un gráfico de líneas y barras agrupadas, el cual presenta la información de casos positivos en la región desde Guatemala hasta Panamá, en el mapa de presenta en círculos los casos de contagio, aumentando su tamaño de acuerdo a la cantidad de casos, además se colorea en degradación siendo el verde el menor número y degradando hasta el rojo para representar la mayor cantidad, también se muestra un menú emergente con la información de cada país, mostrando el número de casos positivos, muertes y letalidad. En el caso del gráfico de barras de igual manera se muestra la cantidad de positivos y una línea que muestra la cantidad de muertes por país.



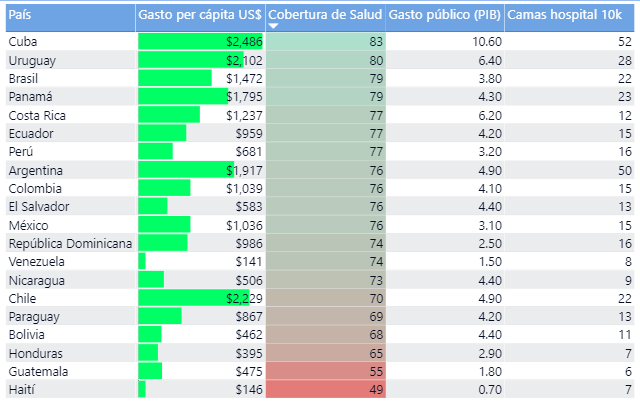
### Orígenes de datos del reporte

El reporte tiene como origen de datos el dataset llamado **Covid19\_world\_data**, y las columnas que se utilizaron en este reporte son la de país (country), la suma número de casos nuevos (new\_Cases), la suma número de casos nuevas muertes (new\_deaths), y para el resumen de la taza de letalidad se usa una variable es que la sum de la totalidad de muertes entre la totalidad de contagio.

## Presentación del reporte: Situación de los sistemas de salud de américa latina para enfrentar la pandemia del Covid19.

### Descripción del reporte

Este es una tabla, que presenta un resumen de algunos indicadores de la situación de américa latina con respecto a sus sistemas de salud antes de comenzar a la pandemia, la tabla presenta 20 países de la región, y muestra el gasto per cápita US$ (Dólares de Estado Unidos) de los países en salud, la cobertura del territorio, el porcentaje de gasto de acuerdo a producto interno bruto (PIB), y las camas disponibles en los hospitales por cada 10 mil habitantes (10K), además se presente en la columna de gasto per cápita, una gráfica de barras horizontales que muestra los valores de este gasto de acuerdo a el gasto de cada uno de los países, y en la cobertura de salud, se muestra una formato condicional siendo el verde el mayor y rojo el menor.



### Orígenes de datos del reporte

El reporte tiene como origen de datos el dataset llamado **Capacidad LA**, y las columnas que se utilizaron en este reporte son la de país, gasto per cápita US$, cobertura de salud, gasto público (PIB) y camas hospital 10K, este dataset es el resultado de las transformación de los datos de las tablas 1,2,3 obtenidas del artículo “Capacidad América Latina covid-19 - BBC News Mundo”

# Conclusiones

* La población en general a nivel mundial ha visto la importancia del análisis de los datos y como estos ayudan a entender menor la situación de Covid19, y cuál ha sido la importancia de estos para tomar acciones y la mitigación del virus.
* El sistema de salud a nivel mundial no se encuentra preparado para atender una pandemia en la que la mayoría de la población enferme al mismo tiempo.
* Con los datos que se tienen respecto a la cantidad de contagios a nivel mundial y la capacidad del sistema de salud de cada país (por lo menos a nivel de Latinoamérica), no se puede llegar a una conclusión determínate del porque en Costa Rica el virus se ha comportado de una manera afable, siendo que países con igual o mayor capacidad para atender la emergencia, ha visto el colapso de su sistema de salud.
* La curva de aprendizaje en la aplicación Power Bi es bastante baja, no se necesita tener conocimientos a nivel técnico para poder usar esta aplicación, sin embargo, si se necesita algún conocimiento de en visualizaciones y transformación de datos.

# Bibliografía.

Estrategia, F. i. (03 de Febrero de 2020). *Limpieza y Transformación de datos en Power BI*. Obtenido de YouTube: https://www.youtube.com/watch?v=LAEqnbM92oU

IDECA, U. A.-G. (Mayo de 2019). Instructivo para la creación de Diccionario de Datos. Bogota, Colombia.

Microsoft. (2020). *Documentación de Power BI*. Obtenido de https://docs.microsoft.com/es-es/power-bi/

Microsoft. (14 de Junio de 2020). *Precios de Power BI*. Obtenido de Power BI: https://powerbi.microsoft.com/es-es/pricing/

Organización Mundial de la Salud (WHO). (20 de Mayo de 2020). Obtenido de WHO Coronavirus Disease (COVID-19) Dashboard: https://covid19.who.int/

Organización Mundial de la Salud (WHO). (Mayo de 2020). *Coronavirus disease (COVID-2019) situation reports.* Recuperado el 20 de Mayo de 2020, de https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

Universidad Estatal a Distancia. (2015). *Situación Nacional COVID-19*. Obtenido de Geovisión: http://geovision.uned.ac.cr/oges/

Wallace, A. (23 de Marzo de 2020). *BBC News Mundo*. Obtenido de Coronavirus: qué capacidad tienen realmente los países de América Latina para hacer frente a la epidemia de covid-19.: https://www.bbc.com/mundo/noticias-america-latina-51916767

Wikipedia. (14 de Diciembre de 2018). *GitHub*. Obtenido de https://es.wikipedia.org/wiki/GitHub

Wikipedia. (05 de Abril de 2020). *Power BI*. Obtenido de https://es.wikipedia.org/wiki/Power\_BI