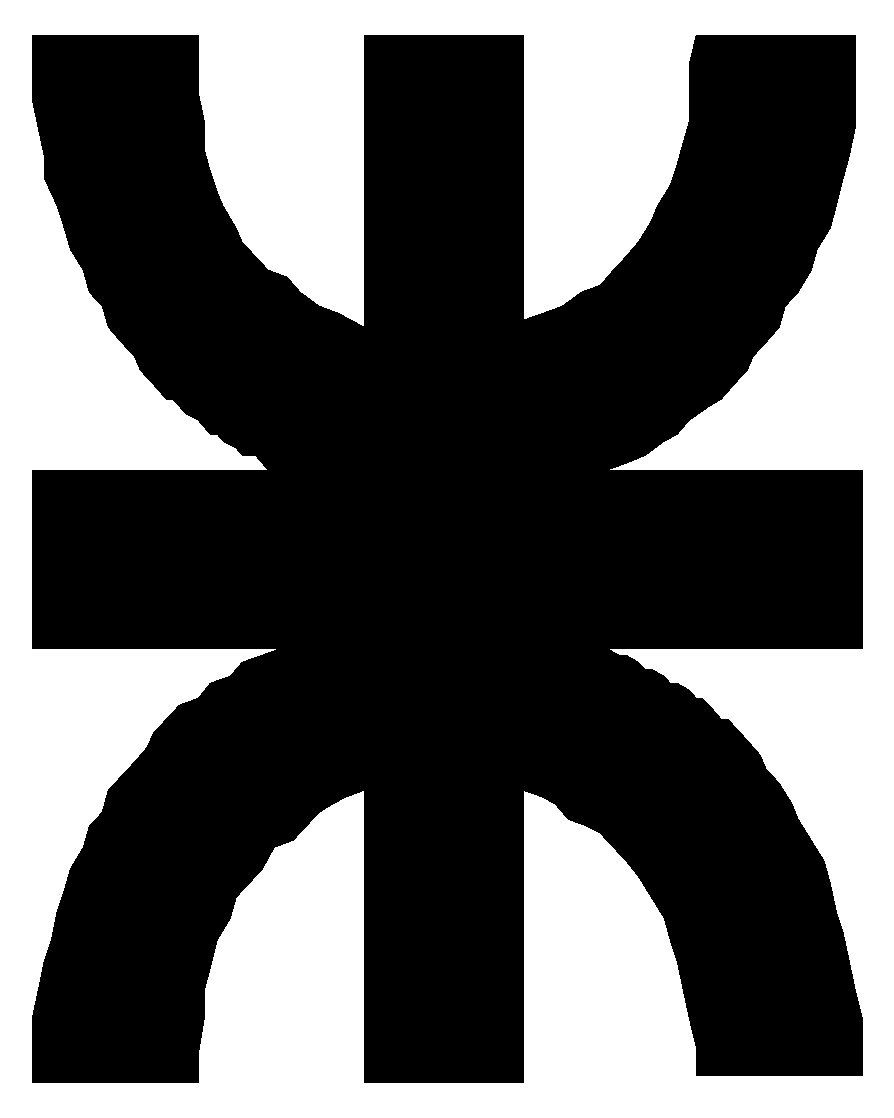
**Universidad Tecnológica Nacional**

**Facultad Regional Córdoba**

**Ingeniería en Sistemas de Información**

**Cátedra: Big Data**

**5K4\_**

**Grupo**

**Integrantes:**

* Buchaillot, Emilio 65530
* Moll, Luciana 64585
* Solovey, Melina 65023
* Tita, Romina 64587

**Docentes:**

* Maldonado, Calixto Alejandro
* Mana, Franco
* Vaca, Pablo Andrés

**Índice**

[Introducción 3](#_Toc486604335)

[Planificación 4](#_Toc486604336)

[Análisis y limpieza de datos 6](#_Toc486604337)

[Meteoritos Caídos 11](#_Toc486604338)

[Cráteres 12](#_Toc486604339)

[Simulación 14](#_Toc486604340)

[Probabilidades 17](#_Toc486604341)

[Visualizaciones 19](#_Toc486604342)

[Meteoritos.pag1 20](#_Toc486604343)

[Conclusiones 21](#_Toc486604344)

[Meteoritos.pag2 24](#_Toc486604345)

[Conclusiones 25](#_Toc486604346)

[Cráteres 28](#_Toc486604347)

[Conclusiones 29](#_Toc486604348)

[Asteroides 32](#_Toc486604349)

[Conclusiones 33](#_Toc486604350)

[Simulación 36](#_Toc486604351)

[Probabilidades 37](#_Toc486604352)

# Introducción

¿Como surgió esta idea? Comenzamos a buscar dataset en la web para darnos una idea de que tipo de información hay disponible, y así definir qué hacer. De esa manera nos topamos varios dataset de la NASA, uno de los más ricos, en cuanto a cantidad de información, correspondía a meteoritos. Lo analizamos muy por arriba, probamos Watson sobre 1000 filas que extrae, y encontramos que tiene información variada: clasificación (según materiales y el desgaste), tamaño, ubicación, entre otros. Nos fascinó la idea de poder analizar y sacar conclusiones de toda esa información.

Ahora bien, nuestra gran pregunta fue ¿qué historia les podemos contar con esos datos?, y a partir de esa interrogación nos planteamos otra ¿qué son y de donde vienen los meteoritos? Esta última nos llevó a relacionarlos con los asteroides, así que buscamos dataset de los mismos, y nos topamos con uno bastante grande, con información muy diversa. Ese fue el mecanismo para agregar más fuente de datos al proyecto: una simple pregunta, que establece una nueva relación.

Para poder contarles esta historia tuvimos que indagar en la astronomía, comprender órbitas, clasificaciones de cuerpos celestes, la formación química de los mismos, refrescar conceptos de física y profundizarlos un poco más.

Nuestra historia tratara de ubicarnos dentro del espacio y de ponernos al tanto de los riesgos que pueden venir de él. Queremos que con nuestra historia ustedes sean capaces de entender el pasado, y a partir de allí, que puedan viajar al espacio, para comprender los peligros a los que todos estamos expuestos.

# Planificación

**Objetivos:**

El principal objetivo de este proyecto es la concientización de la sociedad respecto de los riesgos potenciales que se encuentran en el espacio. Por tal motivo, buscamos transmitir los resultados de manera entendible, generando información que despierte el interés de los espectadores.

Además, también consideramos importante el hecho de poder embeber de conocimiento astronómico a un ámbito diferente, como lo es la ingeniería en sistemas, mediante el análisis de datos.

Por otro lado, como grupo, tenemos como objetivo propio el formar conocimientos de Big Data.

**Motivación:**

La motivación principal que encontramos es el desconocimiento general de la sociedad respecto de los riesgos potenciales que nos rodean en el espacio. Además, al grupo le resultó muy interesante este tema y todos mostramos muchas ganas para llevarlo adelante.

**Fuentes de Datos:**

* Meteoritos: <https://data.nasa.gov/Space-Science/Meteorite-Landings/gh4g-9sfh/data>
* Probabilidades de Impacto: <https://cneos.jpl.nasa.gov/sentry/details.html#?des=2010%20GZ60>
* Asteroides: <https://ssd.jpl.nasa.gov/sbdb_query.cgi#x>
* Cráteres: <http://www.passc.net/EarthImpactDatabase/>
* Datos de la simulación realizada por el grupo

**Requerimientos:**

El proyecto está formado por tres grandes partes:

* ***Aprender del pasado***: a partir del análisis de los dataset de meteoritos y cráteres se obtendrán conclusiones sobre el daño y las características de los objetos.
* ***Visualizar el presente***: analizando el dataset de los asteroides comprenderemos que es lo que hoy nos rodea, y si representa algún peligro.
* ***Comprende el futuro y las probabilidades***: en base al dataset de las probabilidades de impacto obtenido de la NASA y del resultado de las simulaciones llevadas a delante por el grupo, se sacarán conclusiones del impacto y el daño que podrían causar los asteroides al chocar en la tierra.

**Actividades:**

1. Definir los Data Set.
2. Limpiar, agrupar, unificar conceptos.
3. Definir Objetivo.
4. Definir Motivación.
5. Realizar procesamiento de información.
6. Realizar análisis de los resultados obtenidos.
7. Elaborar informe sobre los resultados.
8. Definir forma gráfica en que se presentará la información.
9. Elaborar presentación para otros individuos (Prezi, PowerPoint).
10. Preparar un discurso explicando las cifras obtenidas, desarrollo de la Historia de Datos.

A cada actividad se le asigno 1 punto.

**Herramientas Utilizadas:**

* Hadoop
* Hive
* Power Bi
* Lenguaje R
* RStudio
* Jupyter

**Links de interés e investigación:**

1. Fórmulas de impacto: <http://www.purdue.edu/impactearth/Content/pdf/Documentation.pdf>
2. Simulador de impacto de la universidad de Purdue, EEUU: <http://www.purdue.edu/impactearth>
3. Dia del asteroide (organización): <https://asteroidday.org/>
4. Clasificación de meteoritos: <https://es.wikipedia.org/wiki/Clasificaci%C3%B3n_de_meteoritos>

**Períodos de liderazgo:**

* Cada alumno será líder del equipo por 3 semanas consecutivas.
* Se realizó un sorteo para determinar el liderazgo:
  + Del 07/04 al 28/04 : Solovey
  + Del 28/04 al 19/05 : Tita
  + Del 19/05 al 09/06 : Moll
  + Del 09/06 al 30/06 : Buchaillot

**Stakeholders posibles:**

* Miembros del equipo
* Integrantes del Curso Big Data
* Comunidad educativa
* Organizaciones astronómicas
* Comunidad científica
* La sociedad en general

# Análisis y limpieza de datos

Se cuentan con 3 dataset a los cuales se le hará un análisis profundo y la correspondiente limpieza de datos: **Asteroides y Cometas**, **Meteoritos** y **Cráteres**. El resto de las fuentes de datos no se analizarán en este apartado. A continuación, se hará un análisis del contenido de los mismos, seccionado por cada uno de ellos, es decir, que tendremos tres apartados, y en cada uno desarrollaremos los principales puntos a analizar:

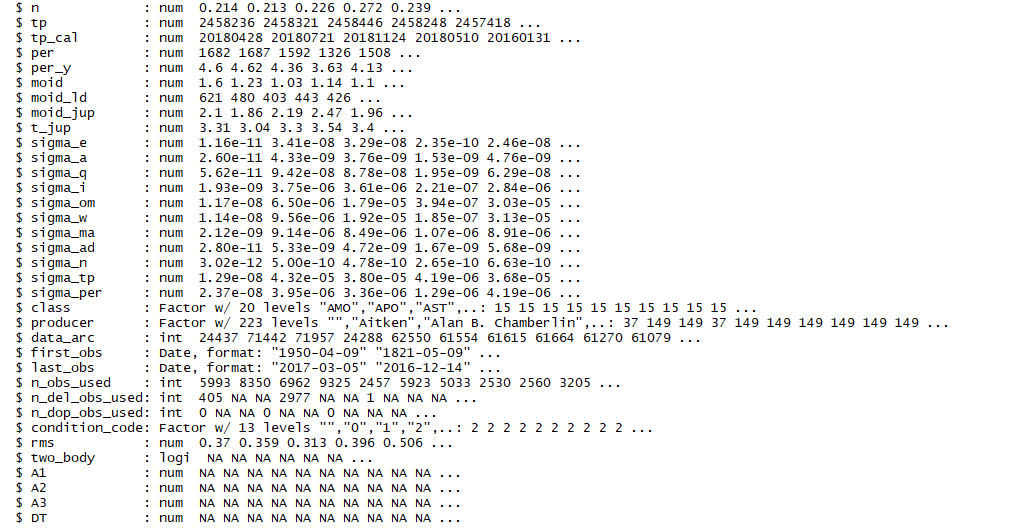
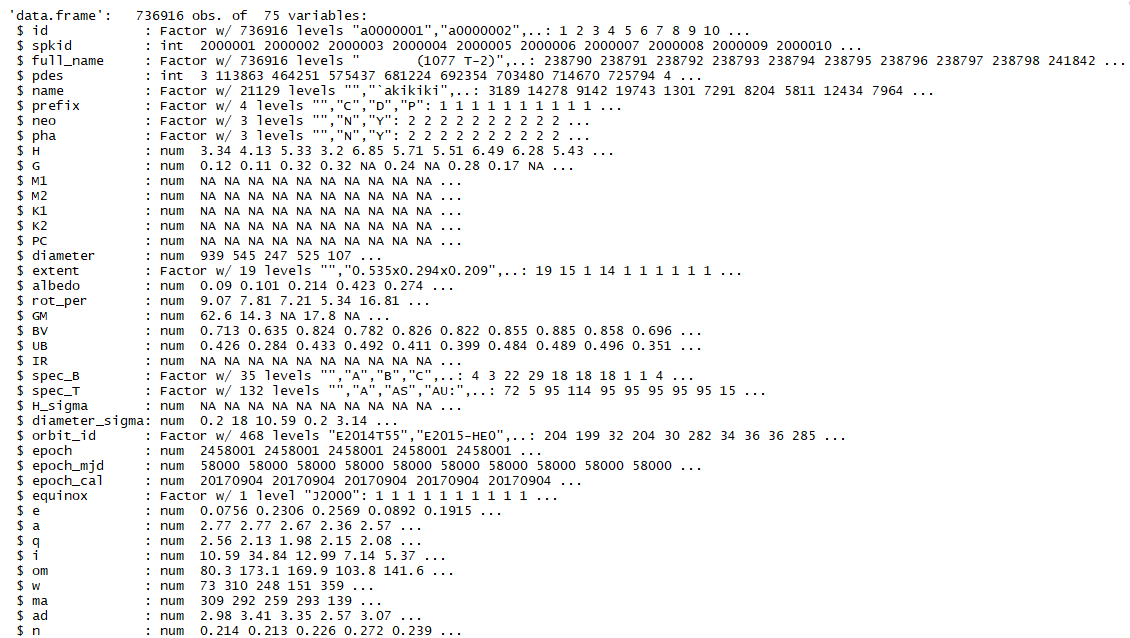
* Completitud de los datos.
* Formato de los Datos.
* Estadística básica de cada columna:
  + Para Variables Continuas:
    - Valor Mínimo
    - Percentil 25%
    - Valor Medio o Percentil 50%
    - Percentil 75%
    - Valor Máximo
    - Valores Nulos (si los hubiese)
  + Para Variables Categóricas:
    - 6 primeras categorías con más cantidad de datos.
    - Valores Nulos (si los hubiese)
* Valores atípicos.
* Estrategias de Limpieza.

Asteroides y Cometas

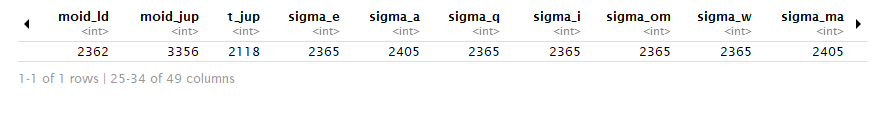
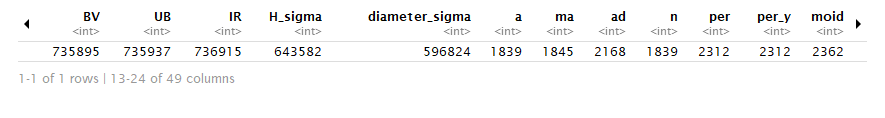
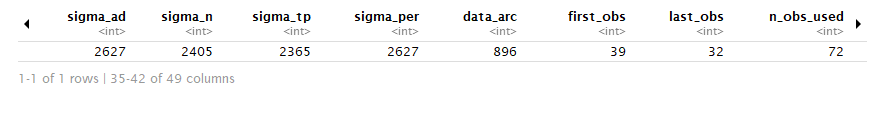
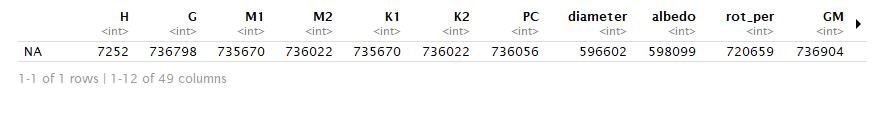
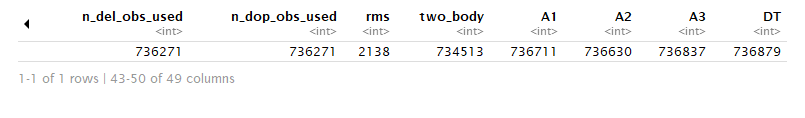
Datos generales del dataset

* 736916 registros(filas).
* 75 variables(columnas).
* Fuente: NASA
* URL: <https://ssd.jpl.nasa.gov/sbdb_query.cgi#xç>

**Estructura de los datos**



**Cantidad de valores nulos(NA) de cada columna**

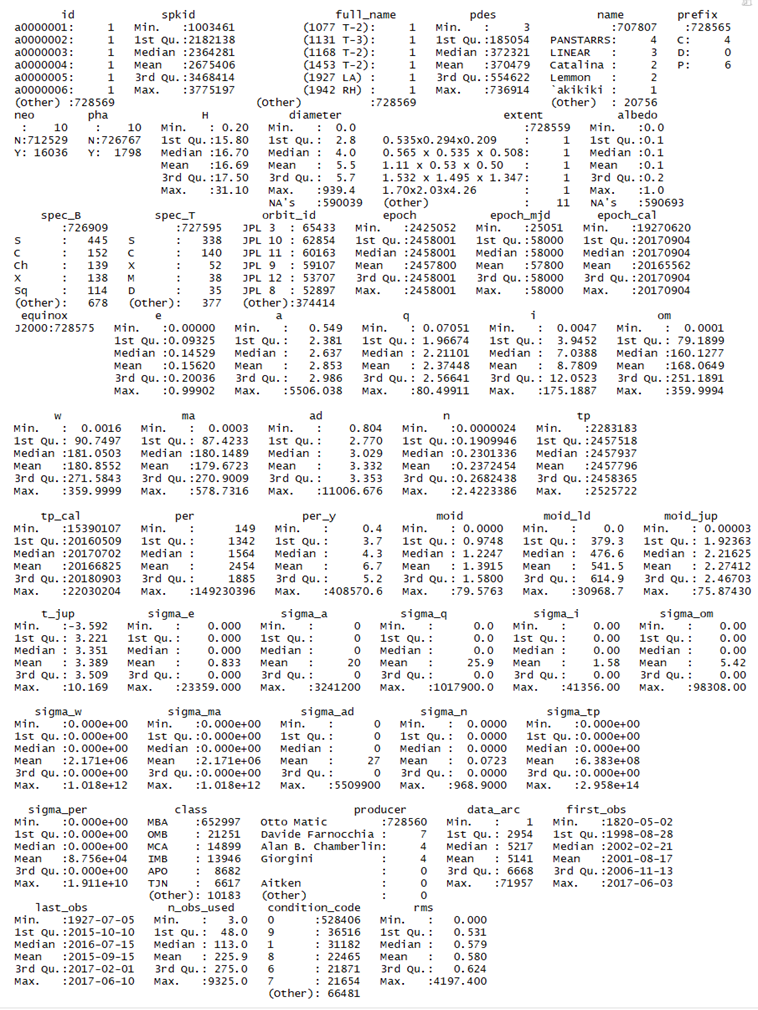


Podemos observar que muchas columnas tienen la mayoría de valores nulos, por tal motivo se decide eliminar las siguientes: **G**, **M1**, **M2**, **K1**, **K2**, **PC**, **rot\_per**, **GM**, **BV**, **UB**, **IR**, **H\_sigma**, **diameter\_sigma**, **n\_del\_obs\_used**, **n\_dop\_obs\_used**, **two\_body**, **A1**, **A2**, **A3**, **DT**.

En cuanto a las columnas que tienen menor cantidad de valores nulos se decide eliminar las correspondientes filas.

Una vez que se aplicaron los pasos anteriores, las únicas columnas que siguen con valores nulos son **diameter** y **albedo**, que se deciden conservar para obtener una información minuciosa de los datos que contienen.

**Estadísticas básicas de las columnas obtenidas en la reducción anterior**



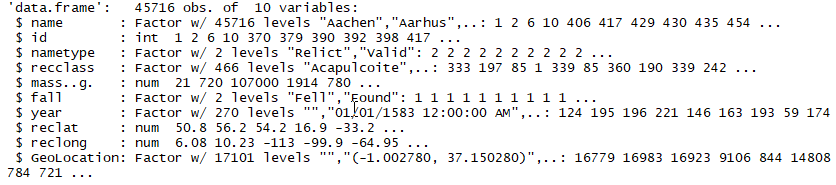
Si observamos con detenimiento las estadísticas, podremos notar que algunas tienen valores vacíos asignados y predominan en las mismas, como es el caso de **extent**, **prefix**, **spec\_B**, **spec\_T**. Por tal motivo, las mismas también serán excluidas de la tabla, al igual que **producer** y **equinox**, ya que tienen un solo valor o uno que predomina ampliamente sobre el resto. Además, se removerán 10 filas en las que, las columnas **neo** y **pha,** no tienen un valor booleano.

## Meteoritos Caídos

Datos generales del dataset

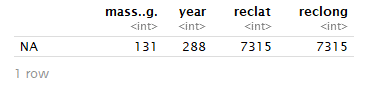
* 45716 registros(filas).
* 10 variables(columnas).
* Fuente: NASA
* URL: <https://data.nasa.gov/Space-Science/Meteorite-Landings/gh4g-9sfh>

**Estructura de los datos**

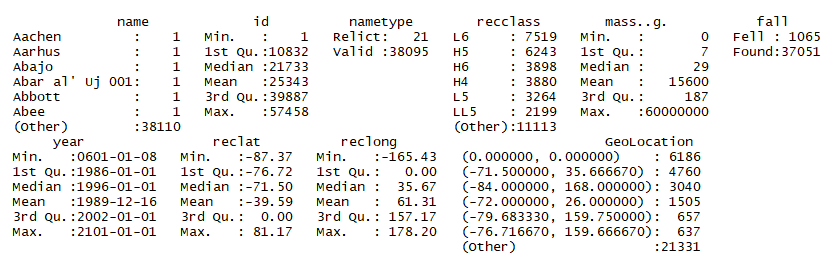


Como pude observarse, toma como tipo Factor a variables categóricas, incluyendo a **year**, que es una fecha. Por eso, debemos realizar la correspondiente conversión para poder manipular correctamente este campo.

**Cantidad de valores nulos(NA) de cada columna**

Observando la tabla y los valores, se tomó la de decisión de eliminar las filas que cumplan con la condición de ser nulas en las columnas **mass..g.**, **year**, **reclat** y **reclong**.

**Estadísticas básicas de las columnas obtenidas en la reducción anterior**



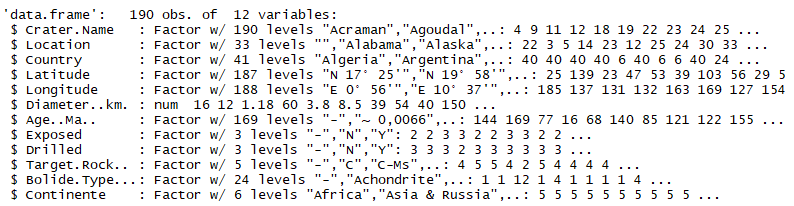
Analizando este resumen estadístico, parecería estar todo en orden, pero si nos centramos sobre la columna **GeoLocation**, podemos observar que tiene muchos valores de coordenadas en (0.0 , 0.0). Sin embargo, no eliminaremos tales filas a fin de analizar el resto de los campos, pero no se tendrá en cuenta en el caso de plasmar la distribución de ubicaciones en un mapa.

## Cráteres

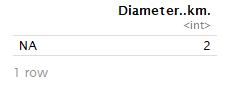
Datos generales del dataset

* 190 registros(filas).
* 12 variables(columnas).
* Fuente: Planetary and Space Science Centre, University of New Brunswick, Fredericton, New Brunswick, Canadá.
* URL: <http://www.passc.net/EarthImpactDatabase>

**Estructura de los datos**

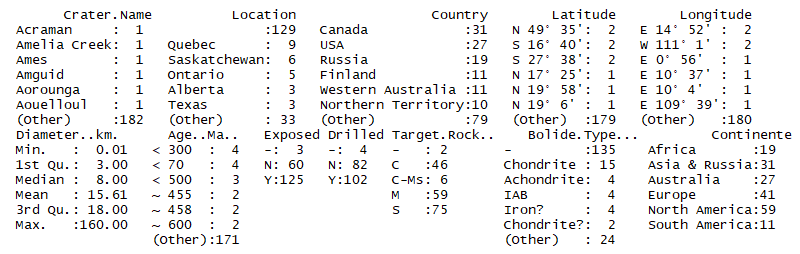


**Cantidad de valores nulos(NA) de cada columna**



En este caso solo tenemos 2 valores nulos, por tal motivo se eliminan las correspondientes filas.

**Estadísticas básicas de las columnas obtenidas en la reducción anterior**



Si bien existen valores en blanco en **Location**, no se eliminarán las filas que se contengan estos valores. Eso se debe a que, en primer lugar, es un dataset muy pequeño, y por otro, tenemos las coordenadas, de manera que podemos acceder a la ubicación de manera precisa. En este caso, la medida que se toma es eliminar directamente esta columna.

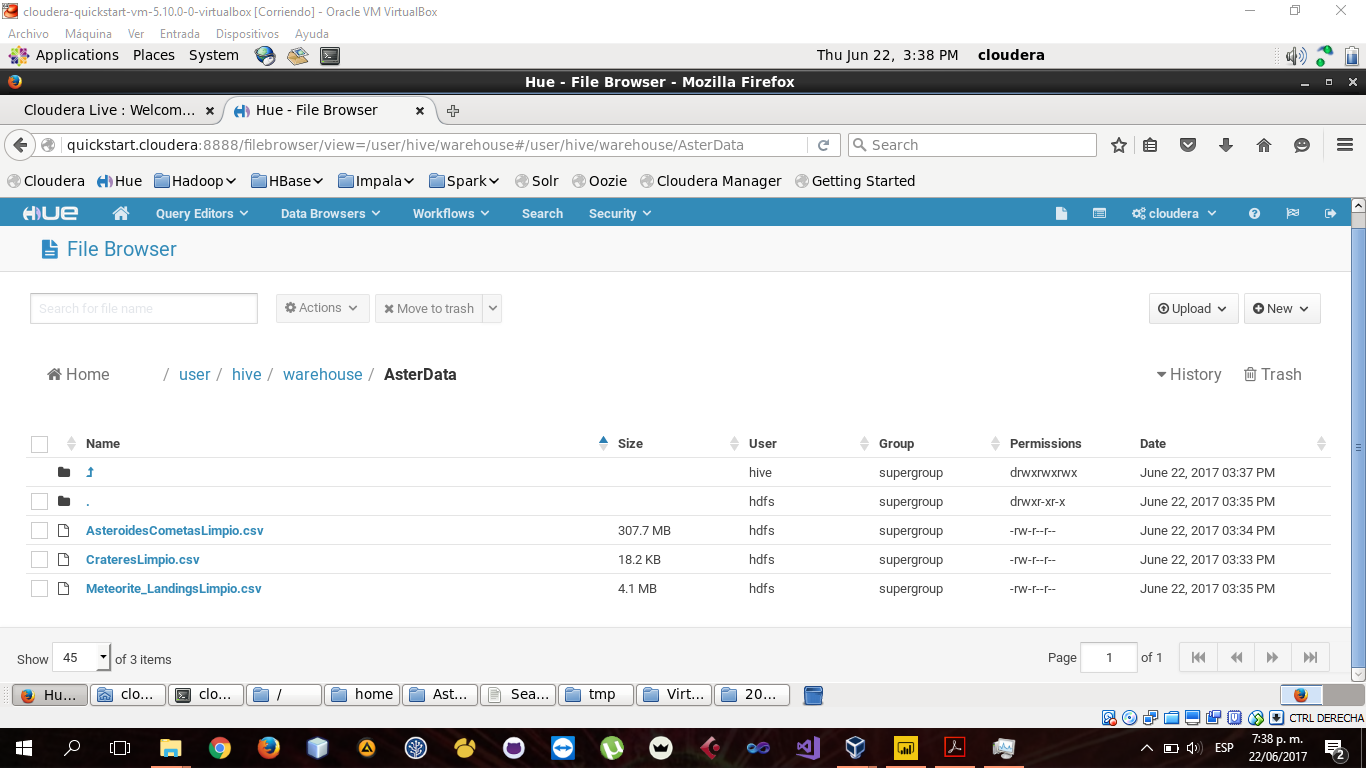
Algo que tendremos que tener en cuenta a la hora de la visualización, es el formato y la correspondiente conversión de **Age..Ma..**, dado que no son SOLO valores, sino que están acompañados de simbología. Por tal motivo, se formateará tal columna teniendo en cuenta solo el valor, y dejando de lado los símbolos que lo acompañan.

# Procesamiento

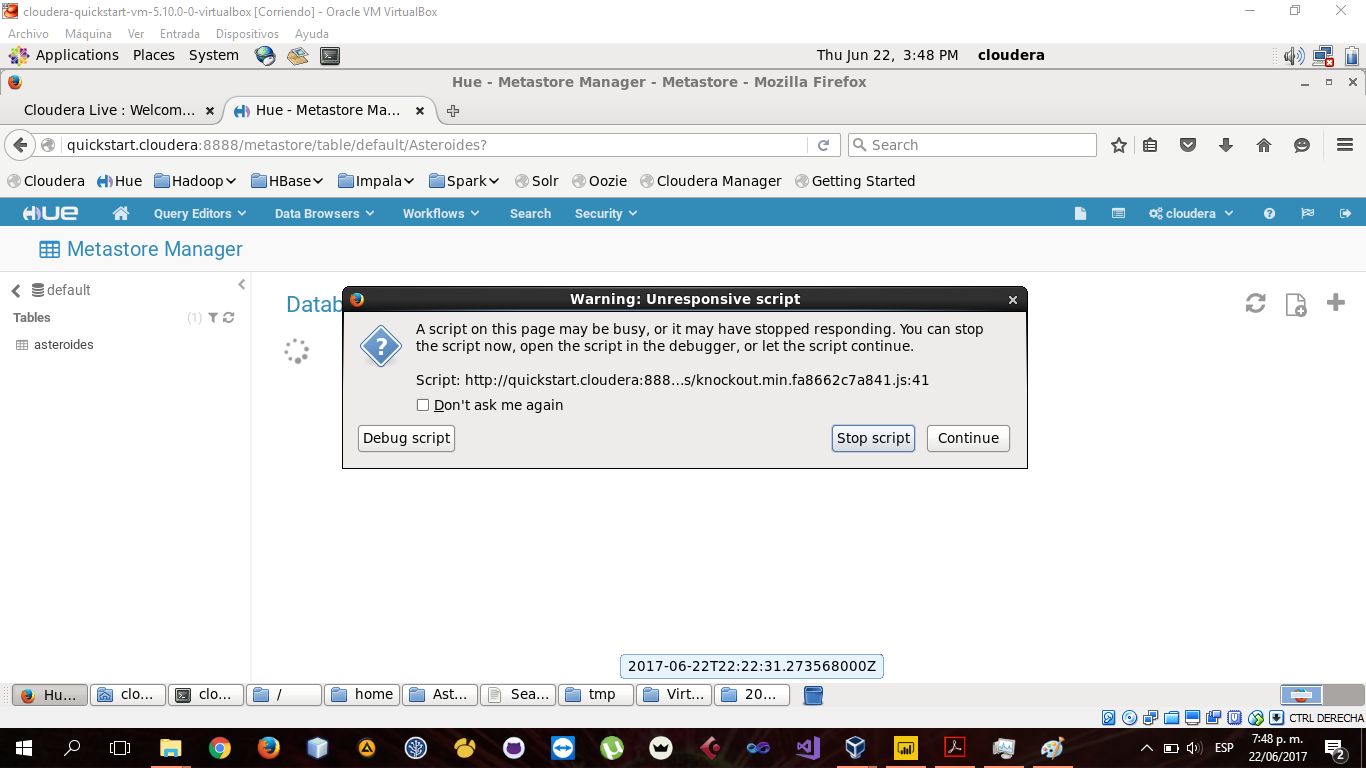
Para poder hacer el procesamiento de toda esta información y armado de “tablas” para luego poder visualizarlo con Power BI, hemos decido utilizar el Framework Hadoop siendo complementado con Hive para establecer la infraestructura de almacenamiento de datos y poder hacer MapReduce con las consultas HiveSQL.

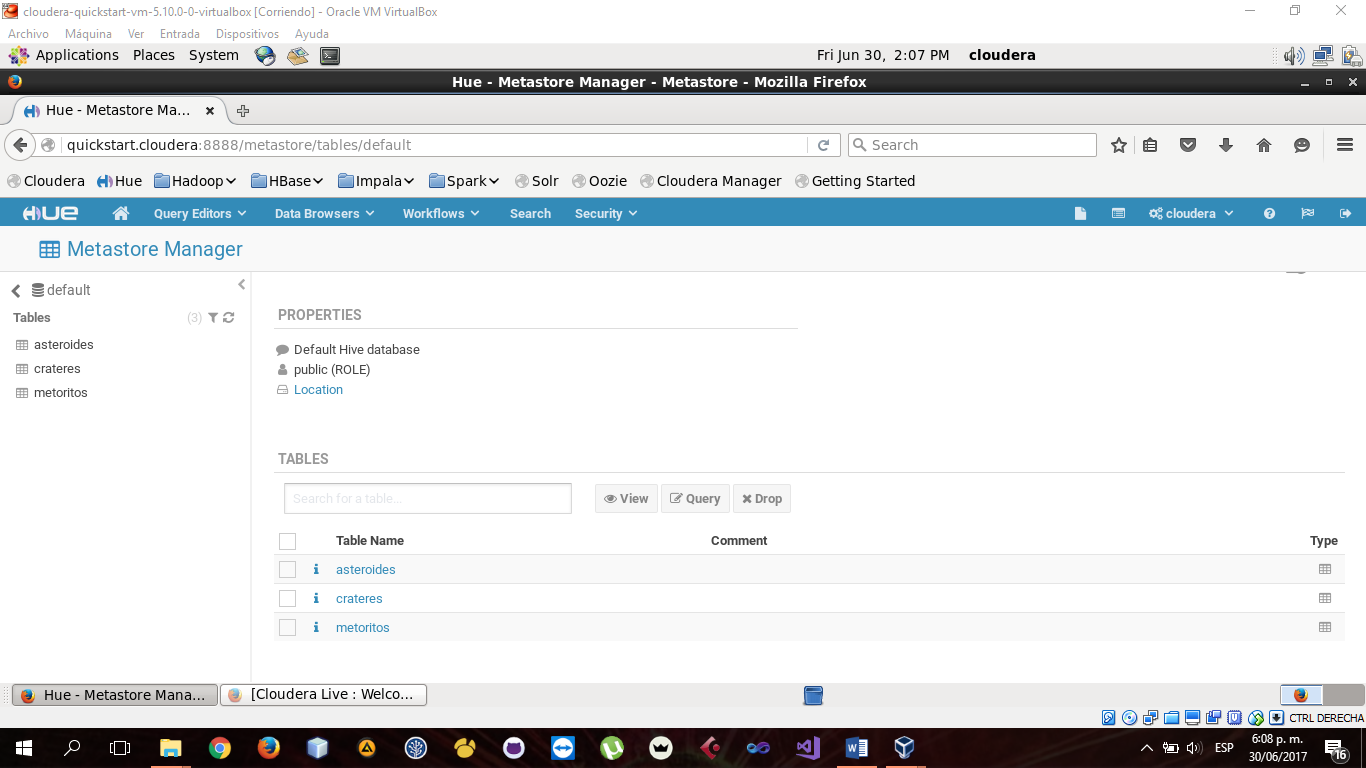
La máquina virtual de Cloudera nos sirvió en gran medida para poder generar con un solo nodo esto que queríamos y luego poder conectarlo a los softwares de visualización.

Realizando las subidas correspondientes de los archivos .csv ya limpios al nuevo sistema de archivos se crearon las tablas referidas a estos datasets. Utilizamos la herramienta wizard de creación de tablas desde los archivos que facilitó un poco el panorama y llegar a cumplir el objetivo.

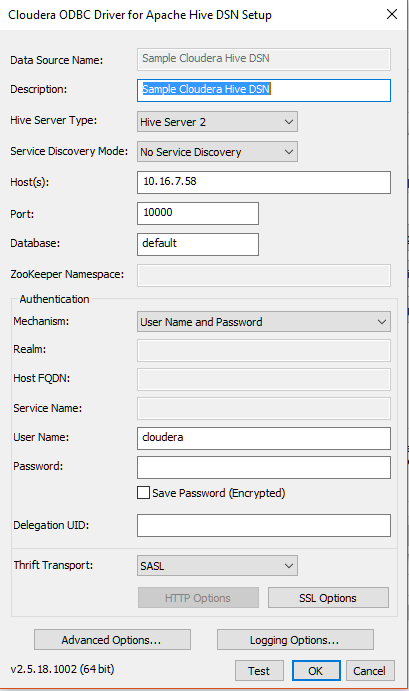


En este primer proceso ocurrieron varios inconvenientes entre ellos la configuración de VirtuaBox para poder recibirlos causo problemas. Luego en la ejecución de los diferentes scripts HiveSQL y la creación de tablas desde archivos ocurrían errores por desconocimieto del uso de la herramienta y problemas en el procesamiento por los recursos de las computadoras.



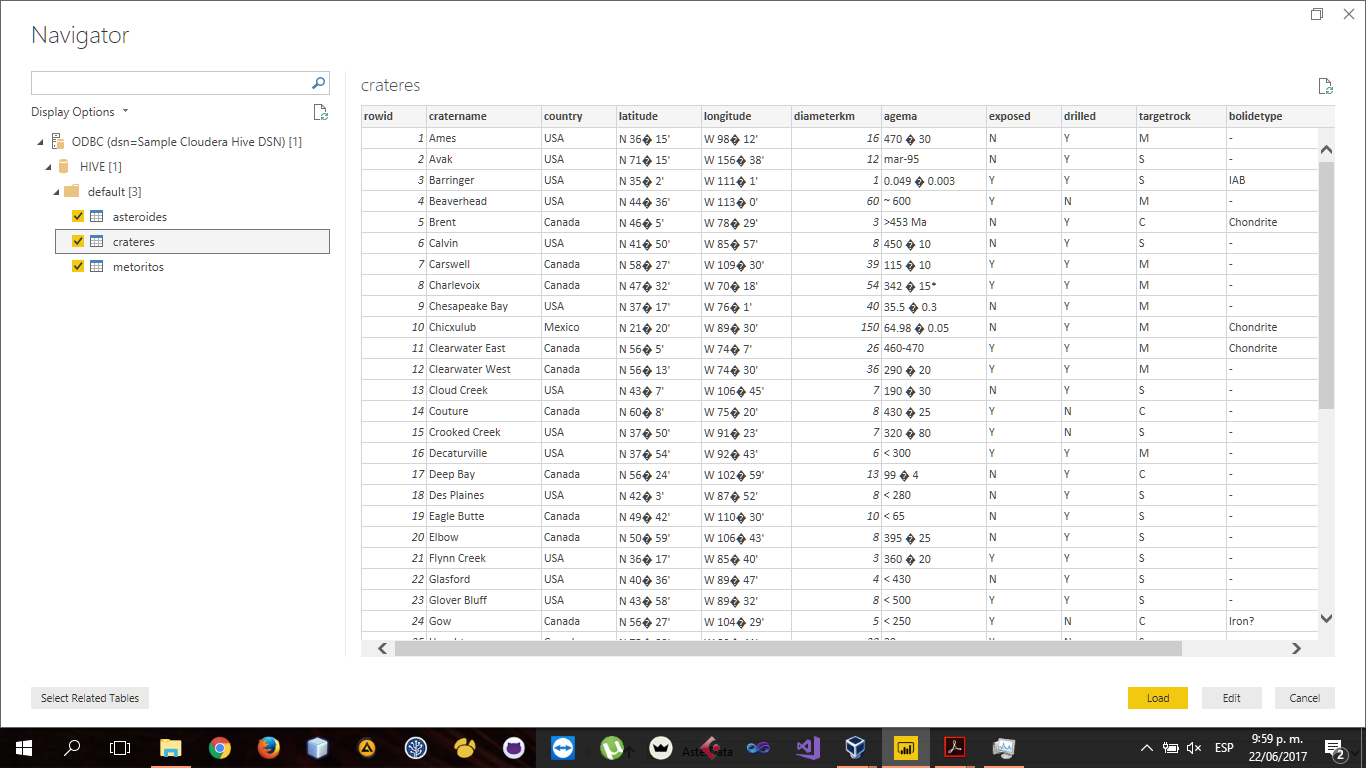


Luego de haber creado todas las tablas en la máquina virtual, se configuró la conexión ODBC mediante el driver correspondiente de Cloudera, indicando todos los parámetros requeridos, como se indica en la siguiente imagen.



Para luego a partir de Power BI obtener esa conexión y generar las nuevas tablas virtuales datos que nos facilitaron el trabajo de visualización y la utilización de diferentes configuraciones en esta última herramienta para que los datos sean leídos de manera correcta y los resultados sean fehacientes.

Estas configuraciones intentamos en varias computadoras, ya que en varias de estas no nos fue posible establecer la conexión hasta que dimos con los parámetros correctos y entendimos nuestro problema.



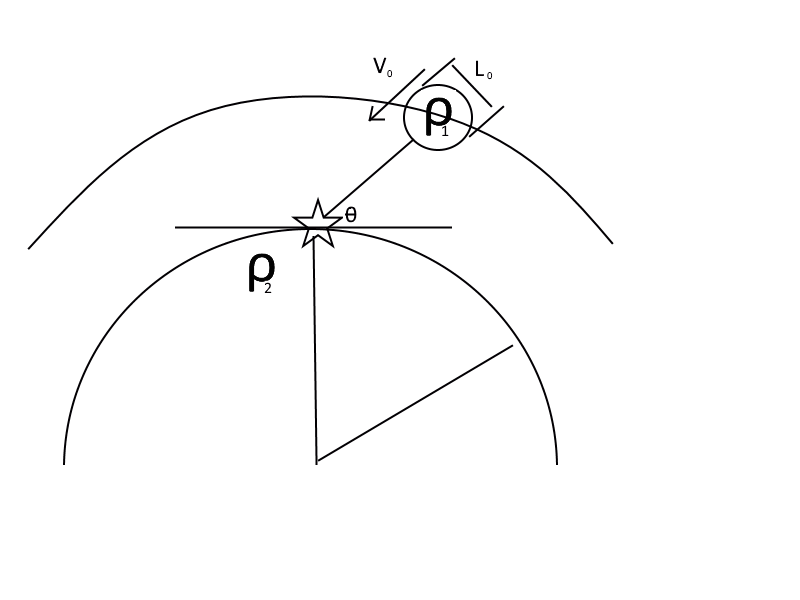
# Simulación

Para simular la caída de un asteroide, se tuvo que limitar solamente al impacto del mismo en la tierra, y el posterior análisis de las dimensiones del cráter generado.

Para ello nos basamos en un paper y documentaciones presentada por la Universidad Purdue, Indiana, Estados Unidos, quienes realizaron un simulador del impacto de asteroides, tomando diferentes parámetros como ser:

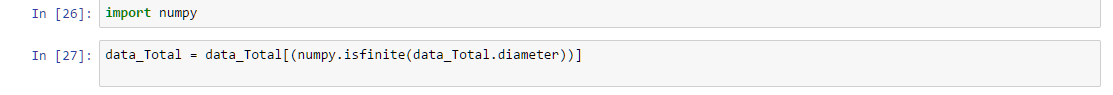
* Densidad
* Diámetro
* Tipo y material
* Densidad del lugar de choque
* Distancia de la persona

Con los cuales se calcula el tamaño del cráter, el impacto ambiental y energético que causaría el choque. Con bastante desarrollo analítico llegaron a respuestas fehacientes, pero para el desarrollo puntual de nuestra simulación acotamos las variables y la cantidad de posibles casos quedándonos solamente con el cálculo final cráter.



Este desarrollo comenzó con una nueva limpieza de datos, a partir del activo que ya habíamos obtenido anteriormente, obteniendo solamente las columnas y filas que contengan los valores que requerimos para el procedimiento, es decir eliminamos filas que no contengan el diámetro del asteroide en cuestión.



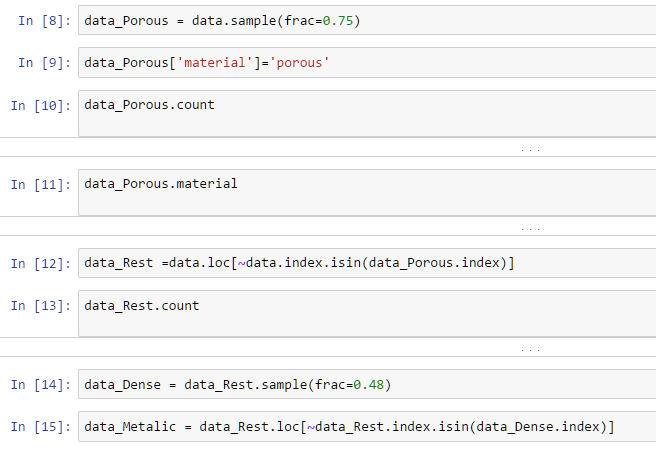


Luego ya teniendo solamente las filas que contienen un diámetro, continuamos con la inferencia de los tipos de Asteroides, ya una clasificación de su material concisa no se posee. La suposición que se hizo se basó en la clasificación que se tiene de una pequeña muestra de Asteroides, llamada Clasificación Tholen, la cual, analizando el espectro de luz de los mismos, se deduce el material por el cual están compuestos y por lo tanto la densidad del mismo.

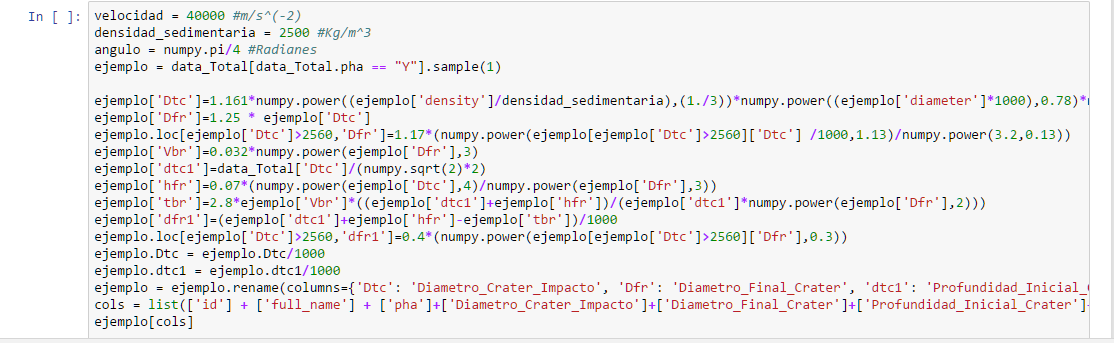
Esta pequeña muestra está basada en el estudio de 978 asteroides obteniendo porcentajes generales que sirvieron para hacer una clasificación aleatoria de todo el dataset para tener el tamaño del cráter que generarían todos estos cuerpos celestes. Los resultados de esta clasificación que se utilizaron son:

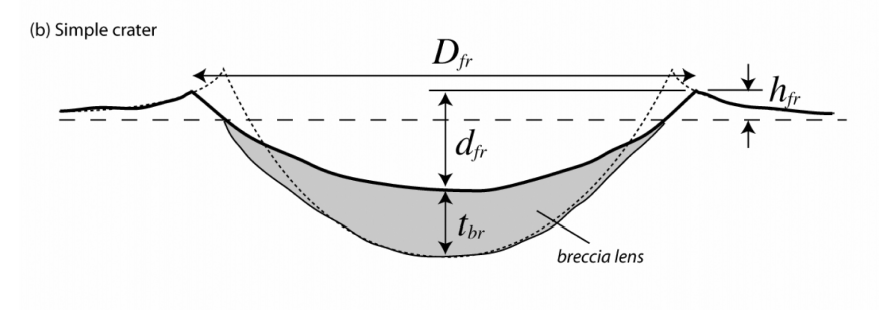
* 75 % de los asteroides de la muestra presentaban características carbonaceas, es decir rocas porosas, las cuales tienen una densidad aproximada de 1500 kg/m3
* 17 % presentan características mas densas, siliceos rocosos, con una densidad de 3000 kg/m3
* Y el resto está muy repartido en varios tipos (7 aproximadamente) pero son casos puntuales de ciertos asteroides y de similares características que podían ser englobados en un 8 % de características metálicas con una densidad aproximada de 8000 kg/m3

Tenemos en claro que el error en estas estimaciones es muy grande, aun así habiendo consultado con personas conocedoras del tema y habiendo analizado otro tipo de estimaciones, nos convencimos que era suficiente para alcanzar nuestro objetivo.



Ya teniendo el dataset preparado para comenzar a realizar los cálculos, comenzamos a estudiar la documentación del simulador de la Universidad de Purdue, donde nos topamos con una gran cantidad de fórmulas, las cuales no son difíciles de comprender, se propuso acotar el amplio espectro de efectos que íbamos a analizar, y concluimos con las fórmulas del cráter.





Estos cálculos se basan en dos etapas del cráter:

1. Cuando ocurre el impacto propiamente dicho, la compresión del impacto genera una mayor profundidad y un menor diámetro que el final.
   * Las variables son:
     + Dtc: Diámetro inicial del cráter.
     + dtc1: Profundidad inicial del cráter.
2. Cuando ya la Tierra, o la zona de impacto, reacciona por la tercera ley de Newton, dispersa la energía generando un mayor diámetro al original y una menor profundidad por las diferentes fuerzas normales.
   * Las variables son:
     + Dfr: Diámetro inicial del cráter.
     + dfr1: Profundidad inicial del cráter.
     + tbr: Representa esta diferencia de profundidades.
     + hbr: Altura máxima sobre el nivel de la tierra.

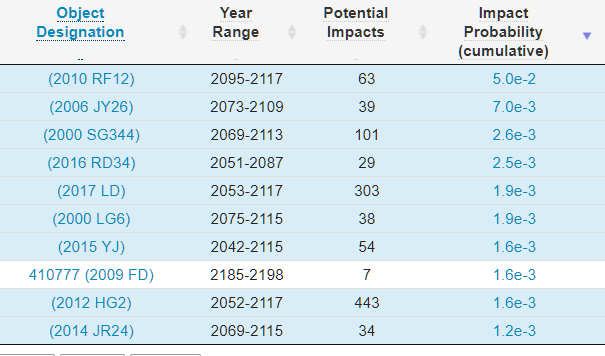
Los cálculos de los cráteres y de la mayoría de los efectos causados por los asteroides están divididos en dos grupos, los pequeños cuerpos celestes que actúan de igual forma en la gran mayoría de los casos y por otro lado están los cuerpos más grandes que actúan muchas veces impredeciblemente. El dato que divide a estos grupos es el diámetro de choque del asteroide a un valor de 2,5 km.

Todos estos datos fueron exportados y subidos nuevamente a HDFS para su posterior análisis en con Power BI y generar las diferentes visualizaciones.

C:\Users\Emilio\AppData\Local\Microsoft\Windows\INetCache\Content.Word\jup5Exportado.png

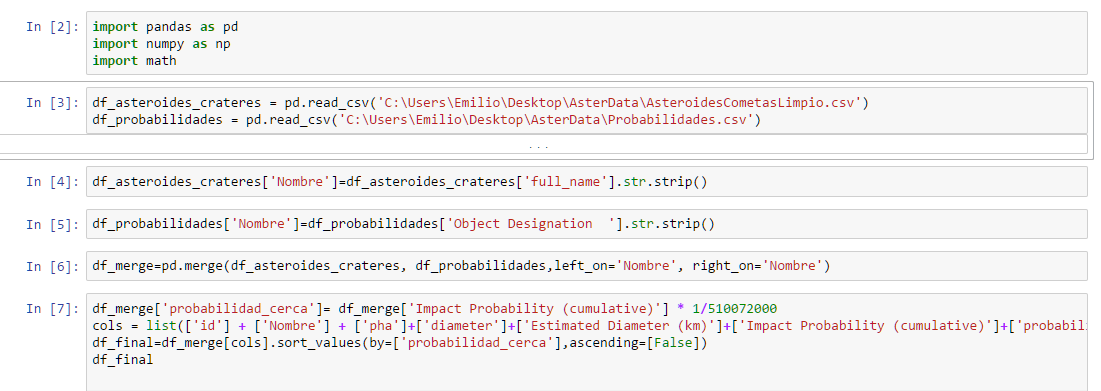
# Probabilidades

El último dataset analizado para llegar a cumplir con nuestro objetivo es proveniente de un sistema de monitoreo constante de la NASA, que calcula las probabilidades de colisión de muchos de los cuerpos celestes que ya hemos analizado. Lo importante y curioso de este dataset, que si bien es chico (700 asteroides aproximadamente), se encuentran solamente los que actualmente registran una probabilidad de que caigan en la Tierra y si esta tiende a 0, aproximadamente a 10-12 el registro es “eliminado”.



En nuestro plan en descubrir y responder el interrogante de “¿Qué probabilidad hay que ese asteroide caiga cerca de mí, en mi km2?”se redujeron la cantidad de variables a tener en cuenta para calcular dicha probabilidad, quedando solamente la probabilidad de impacto indicado por el dataset y la equiprobabilidad de en qué lugar de la Tierra caiga (1 en los 510.072.000 km2 de superficie).

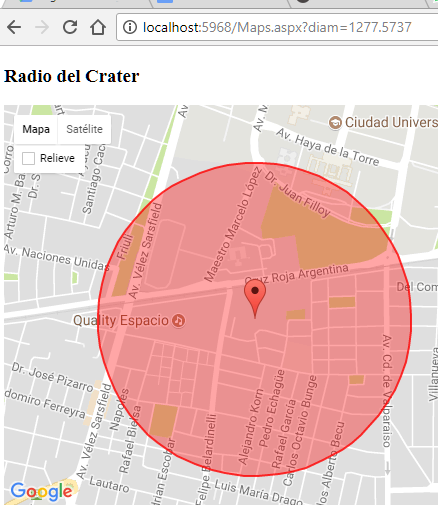
Sumado a este interrogante se analizó también qué tamaño de cráter generaría ese impacto probable. Por lo tanto, pudimos realizar un join entre los dataset para tener datos más ricos, ya que se unieron mediante el nombre, el cual es un identificador único y utilizado de la misma manera en todos los datasets de la NASA.



Por otro lado, también desarrollamos una aplicación web, utilizando la API de Google Maps, para mostrar el tamaño del cráter si ese asteroide en nuestra posición

En conclusión, las chances de caída de un Asteroide en la Tierra en promedio, es muy baja, lo que puede llevar a un desinterés general por ello, pero la probabilidad es muy fluctuante y “poco precisa”, por lo tanto puede pasar cualquier cosa en la inmensidad del espacio y el impacto sea inminente. Actualmente los datos obtenidos son para un periodo de tiempo lejano para el tiempo en que un ser humano vive, siendo la mas grande a 5x10-2.

Y más chica aun es la probabilidad si uno piensa en que el centro del impacto es uno y para tomar dimensión del tamaño del cráter, desarrollamos una aplicación web, utilizando la API de Google Maps marcando la UTN como centro y el diámetro que va a tener el cráter.



# Visualizaciones

A partir de las distintas fuentes de datos hemos generado un total de 6 informes:

* Meteoritos.pag1
* Meteoritos.pag2
* Cráteres
* Asteroides
* Simulación
* Probabilidades

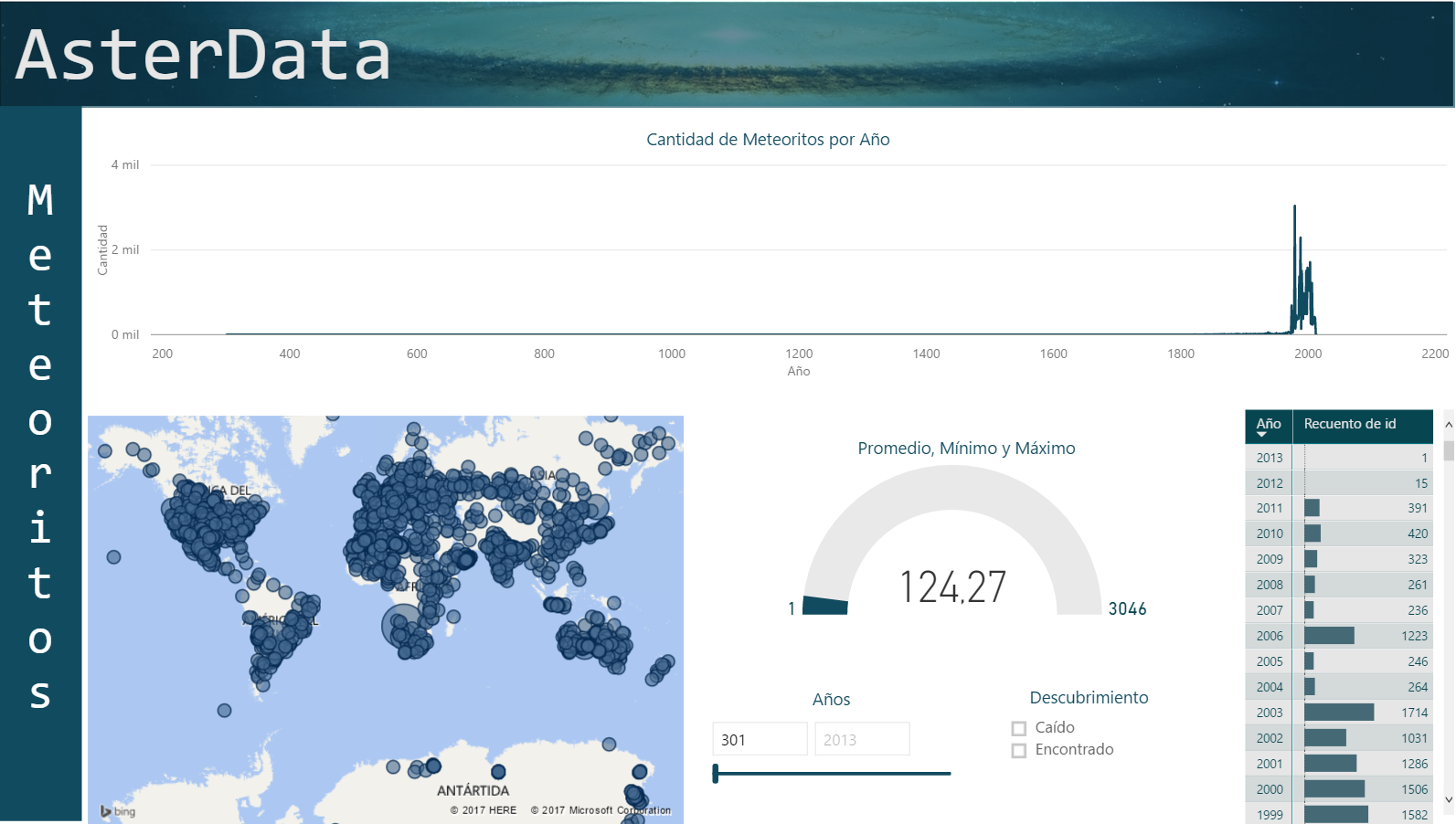
Cuando hablamos de informes, no nos referimos a una simple visualización, sino a un conjunto de gráficos acompañados por tablas q amplían la información, y un set de filtros que permiten un manejo dinámico de los datos.

Se puede acceder a estos informes ingresando al siguiente link:

<https://app.powerbi.com/view?r=eyJrIjoiNTBjNTIzZjctODlmMS00YWExLWE3YTQtMWJmYmM4NTY2Yzg5IiwidCI6ImRjZmI2MzJhLWI4OTYtNDI4OC04NDEzLWVjOGQ5NTQxMDZlNiIsImMiOjR9>

A continuación, haremos un análisis de cada uno de estos informes, mostrando en primera instancia una imagen general, explicando su contenido, y luego conclusiones obtenidas a partir de filtrar los datos.

## Meteoritos.pag1



1

2

3

4

5

6

Este informe está compuesto por un gráfico principal llamado Cantidad de Meteoritos por Años(Fig.1), el cual muestra, como su nombre lo indica, la cantidad de meteoritos que fueron encontrados o cayeron por año. Se encuentra bastante ilegible dado que hay registros que denotan de años muy antiguos.

Se cuenta con tres visualizaciones de soporte, por un lado, el mapa (Fig.2) que refleja las ubicaciones donde han caído/encontrado los meteoritos, un medidor (Fig.3) que indica la cantidad mínima, promedio y máxima de meteoritos por año, y una tabla (Fig.4) con dos columnas, una indica el año y la otra el recuento o la cantidad de meteoritos caídos/encontrados por año.

Finalmente, se dispone de dos filtros, uno corresponde a los años, y el otro al estado en que se descubrió el objeto. El estado “Caído” se corresponde a los meteoritos que pudieron ser divisados al caer, mientras que “Encontrado” indica que el meteorito simplemente se encontró, no hay un testimonio de su caída.

### **Conclusiones**

1. Filtrando por año desde 1900 en adelante y por el estado “Caído” el grafico (*Ilustración 1*) nos muestra, con más precisión, el flujo de la caída de meteoritos. Además, nos da una predicción de la cantidad que van a caer en los próximos años basándose en los 100 años pasados, con un intervalo de confianza del 95%. En la tabla (*Tabla 1*) se pueden ver con más precisión los valores de las predicciones.



Ilustración 1

|  |  |
| --- | --- |
| AÑO | PREVISIÓN (cantidad de meteoritos que van a caer) |
| 2017 | 2 |
| 2018 | 5 |
| 2019 | 1 |
| 2020 | 10 |
| 2021 | 4 |
| 2022 | 3 |
| 2023 | 3 |

Tabla 1

1. Tomando en cuenta el total de los registros del set de datos se puede observar que:

* El promedio de los meteoritos caídos por año son 4.49, el mínimo es 1 y el máximo son 17.
* El promedio de los meteoritos encontrados por año son 154.68, el mínimo son 1 y el máximo son 3044.

1. Si analizamos la tabla en muestra los meteoritos según el año en que han caídos, se puede observar cuales son los años que registran mas caídas de los mismos. Vamos a analizarlo desde dos filtros distintos:

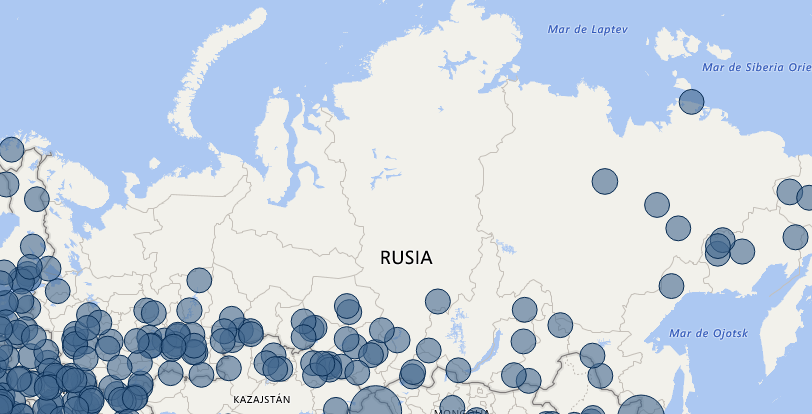
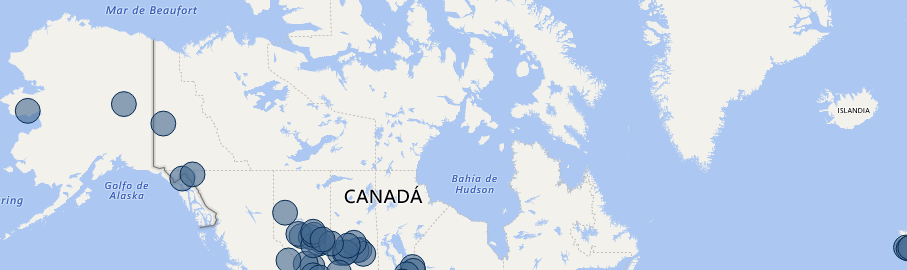
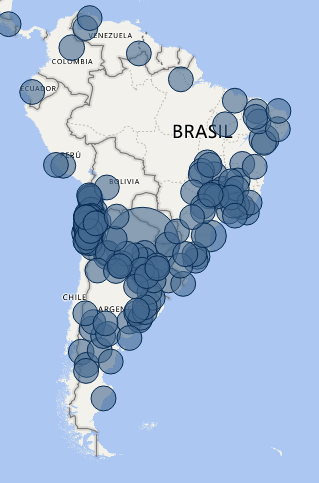
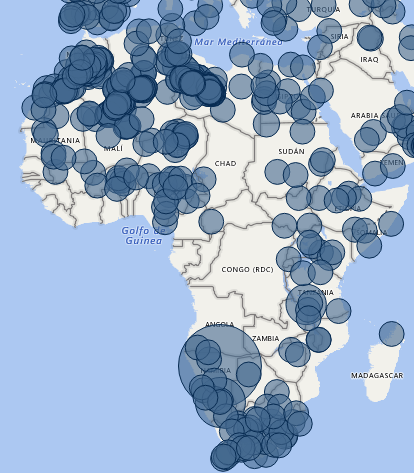
* Por un lado, si filtramos por “Caídos”, podemos ver que el año donde más meteoritos hay es en 1933, con un total de 17, seguido por 1949 con 13 y 1950 con 12 (*Tabla 2*).

Tabla 2

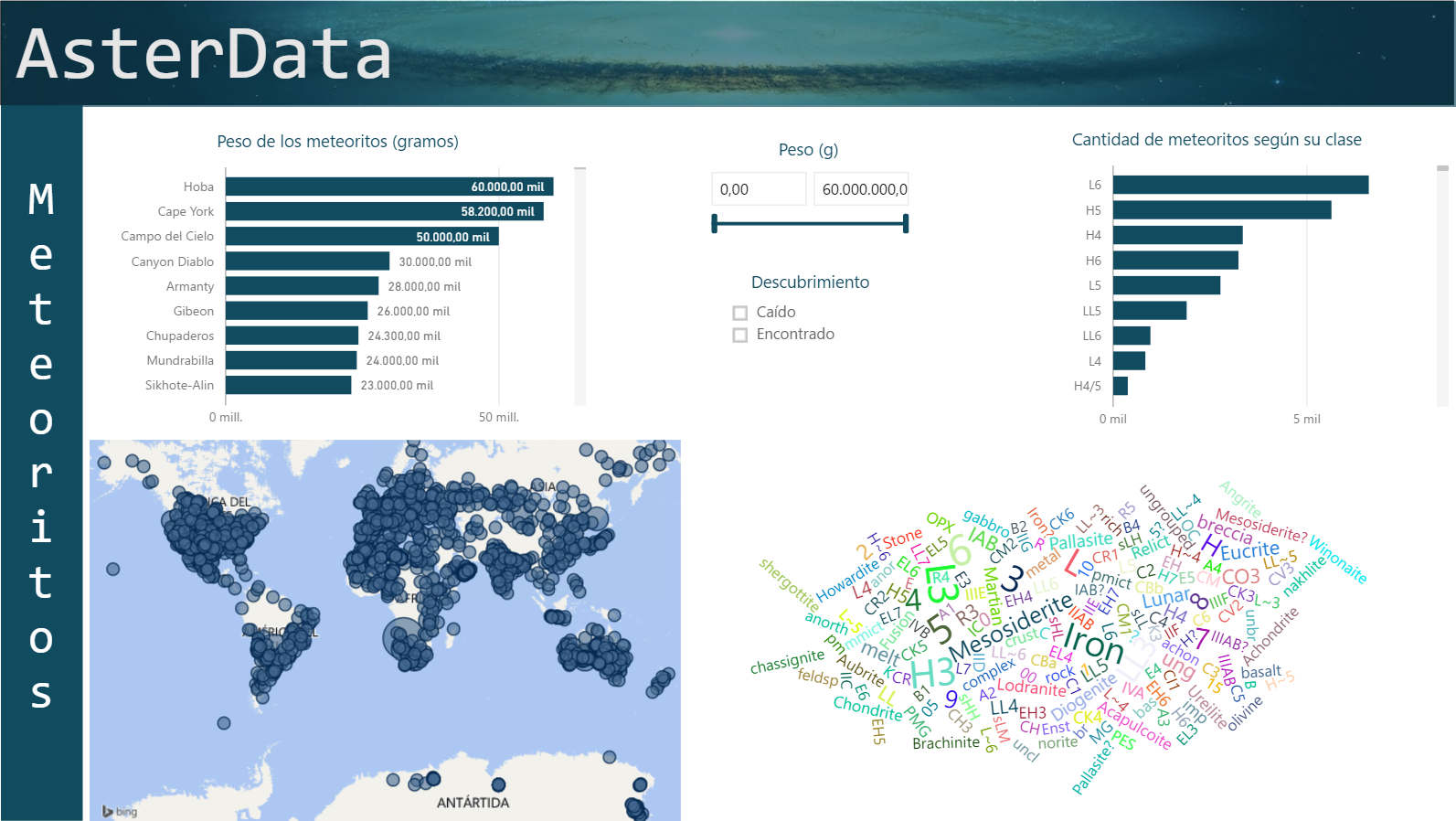
* Ahora bien, si filtramos por “Encontrados” resulta que el año donde más meteoritos han sido encontrados es en 1979, con un total de 3044, seguido por 1988 con 2289 y 2003 con 1705 (*Tabla 3*).

Tabla 3

1. También nos hemos puesto analizar en mas detalle el mapa. Allí, se observa que la zona amazónica, de Siberia, desiertos africanos y en Canadá no se registran prácticamente caídas o descubrimientos de meteoritos. Por lo tanto, nuestra hipótesis es que, tal situación se debe a la escasez de población en esas áreas geográficas.



## Meteoritos.pag2



1

2

3

4

5

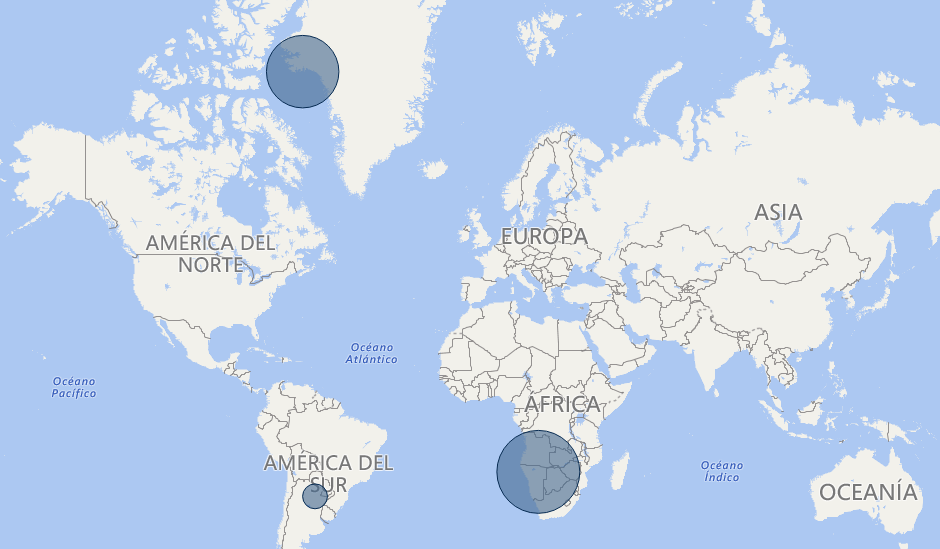
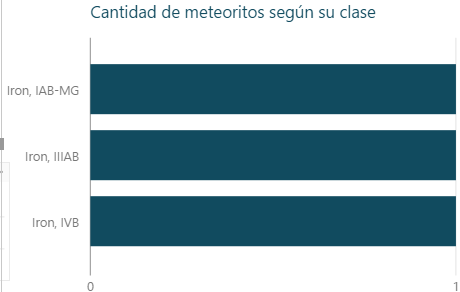
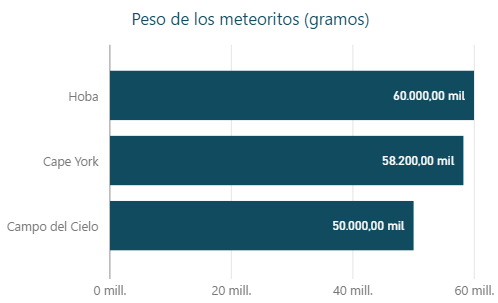
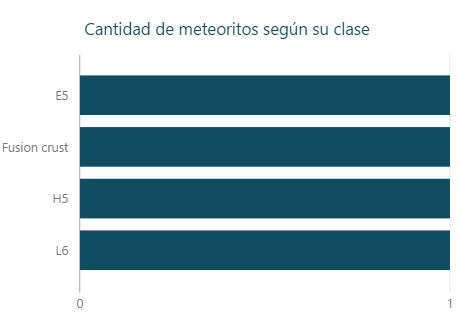
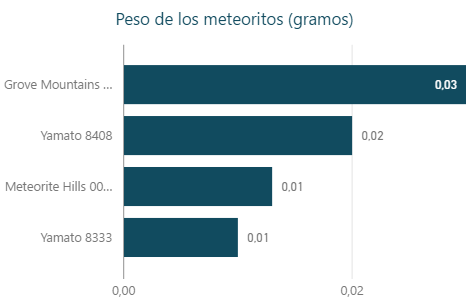
6

Este informe está compuesto por dos gráficos principales (Fig.1 y Fig.2). El primero muestra el peso de los meteoritos en gramos y esta ordenado desde el mayor al menor peso. El segundo refleja la cantidad de meteoritos según la clasificación a la cual corresponde. Si desea mayor información respecto a la clasificación puede dirigirse a <https://es.wikipedia.org/wiki/Clasificaci%C3%B3n_de_meteoritos> .

Como soporte encontramos nuevamente el mapa mostrando las ubicaciones de impacto (Fig.3) y además una nube de palabras con las diferentes clasificaciones de los meteoritos (Fig.4).

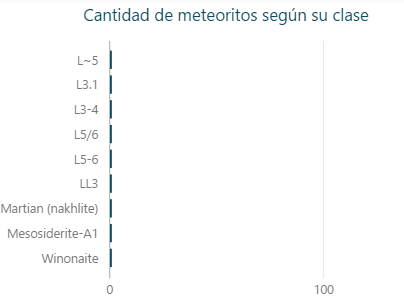
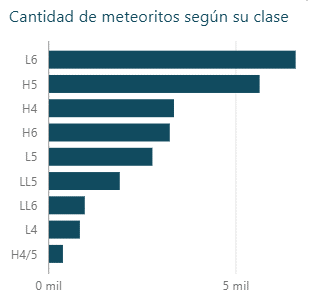
Como filtros disponemos del peso en gramos del objeto (Fig.5) y nuevamente el estado del Descubrimiento (Fig.6).

### **Conclusiones**

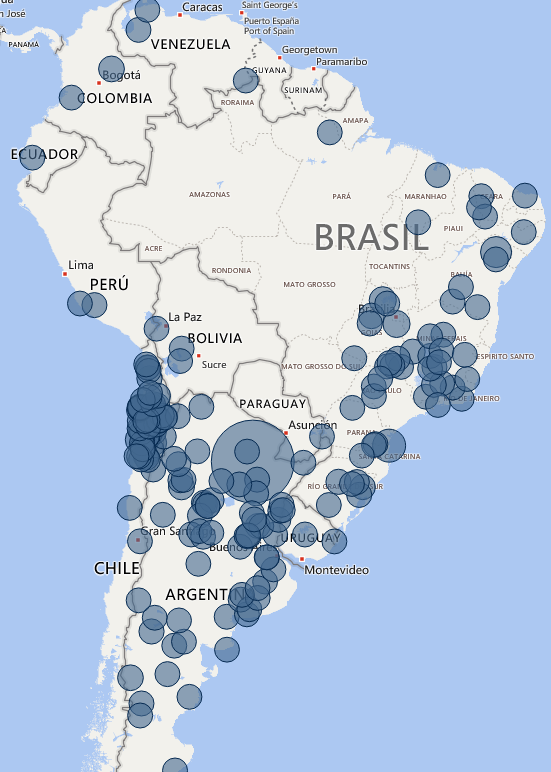
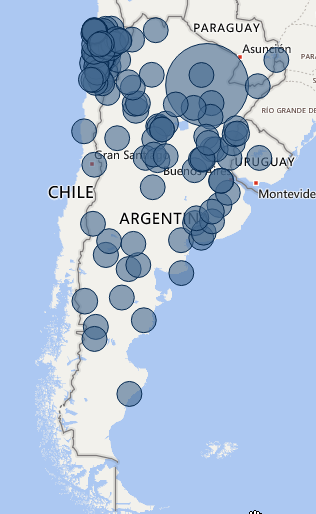
1. Si filtramos la información utilizando el peso, y movemos el límite inferior del mismo hasta los 40.000.000 gramos, podremos observar los tres meteoritos más grandes del mundo. Como se puede ver, el tercer meteorito más grande encontrado se encuentra en Argentina. Más precisamente en Chaco cerca de Villa Angela, San Bernardo y Samuhú. Y está compuesto de hierro perteneciente al grupo principal (MG) del complejo IAB.
2. Como se muestra en el gráfico 2, el meteorito más chico pesa 0.03 gramos formado mayormente por Carbono Orgánico Relicto (Relict OC). Los más chicos, como se ve en el mapa de la imagen fueron encontrados en la Antártida.

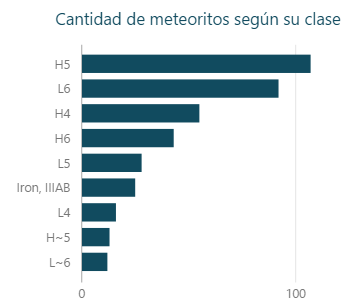


1. También hemos encontrado que la mayor cantidad de meteoritos son de clase L6, Condritis tipo Orgánico Metamorfismo térmico. Esto significa que son condritas (elementos rocosos) que han sufrido una gran metamorfosis térmica.

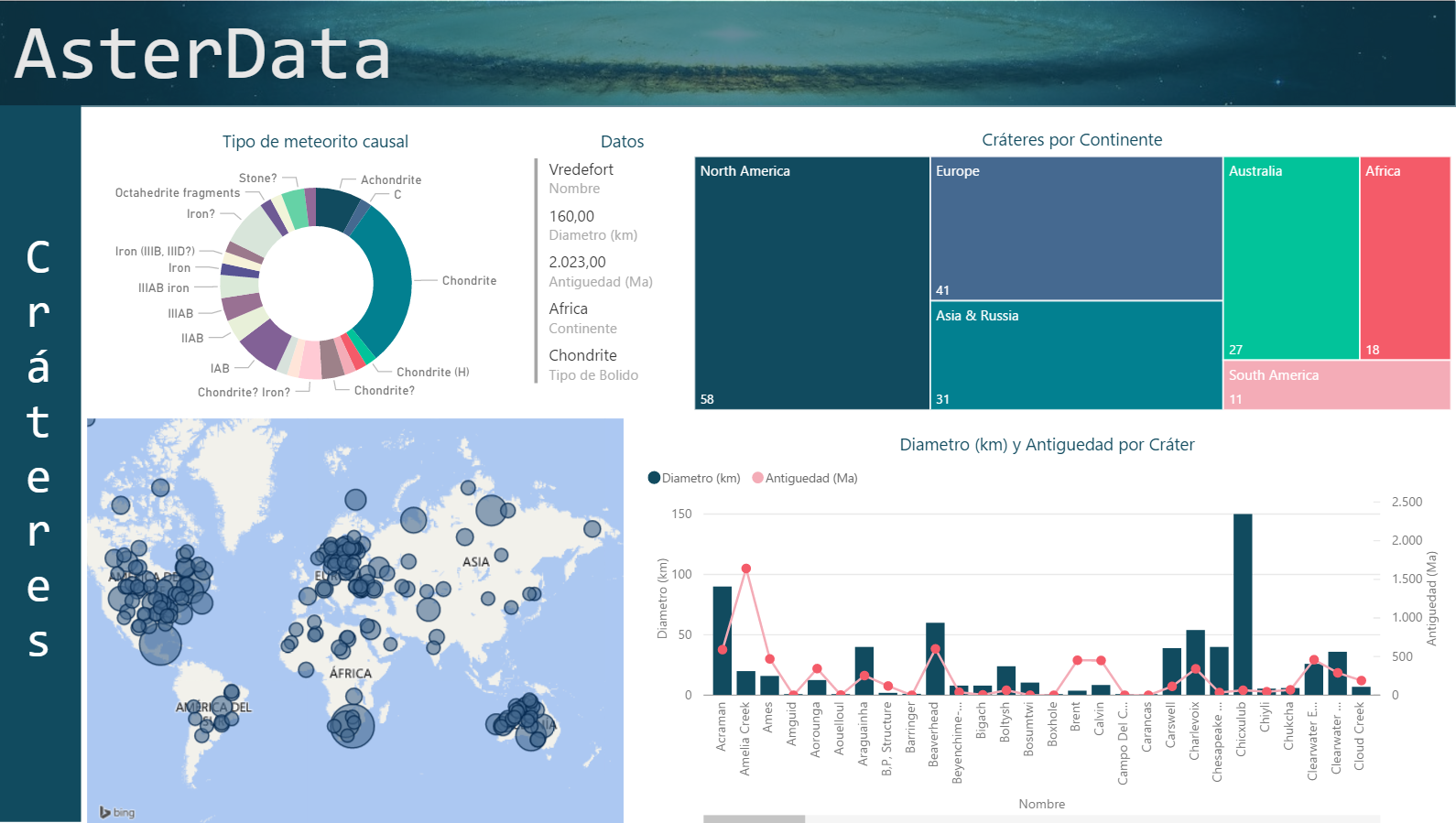


1. En este mapa podemos visualizar de mejor manera los meteoritos que se encuentran en América del Sur y en nuestro país, notando una mayor densidad de caída hacia el norte de Argentina y Chile.

## Cráteres



1

2

3

4

5

Este informe tiene como gráfico principal el diámetro y antigüedad por cráter (Fig.1), donde se especifica por cada uno de los cráteres existentes su diámetro (eje izquierdo), expresado en kilómetros, y su antigüedad (eje derecho), expresada en años. Además, se muestra un gráfico donde se indica el origen de los cráteres (Fig.2), es decir, el tipo de meteorito que lo generó.

También se pueden observar los datos (Fig.3) de cada uno de los cráteres, como el nombre, diámetro, antigüedad, continente y tipo de bólido, según los filtros seleccionados y ordenados según su tamaño.

En otra sección, podemos ver la cantidad de cráteres que hay por continente (Fig.4) y su distribución visualizada en un mapa (Fig.5)

### **Conclusiones**

1. Filtrando por los cráteres de mayor tamaño, observamos que éstos se encuentran en América del Norte y África, en la imagen (*Ilustración 1*) se muestra un mapa con sus ubicaciones, del lado izquierdo, y los datos de diámetro, antigüedad y nombre, del lado derecho.

A su vez, mostramos fotos de los 3 cráteres de mayor tamaño.

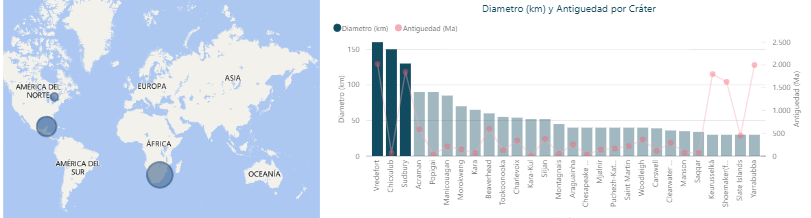
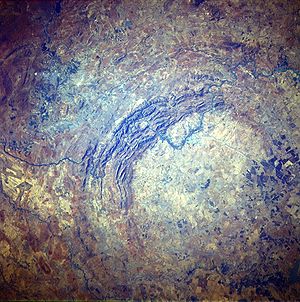
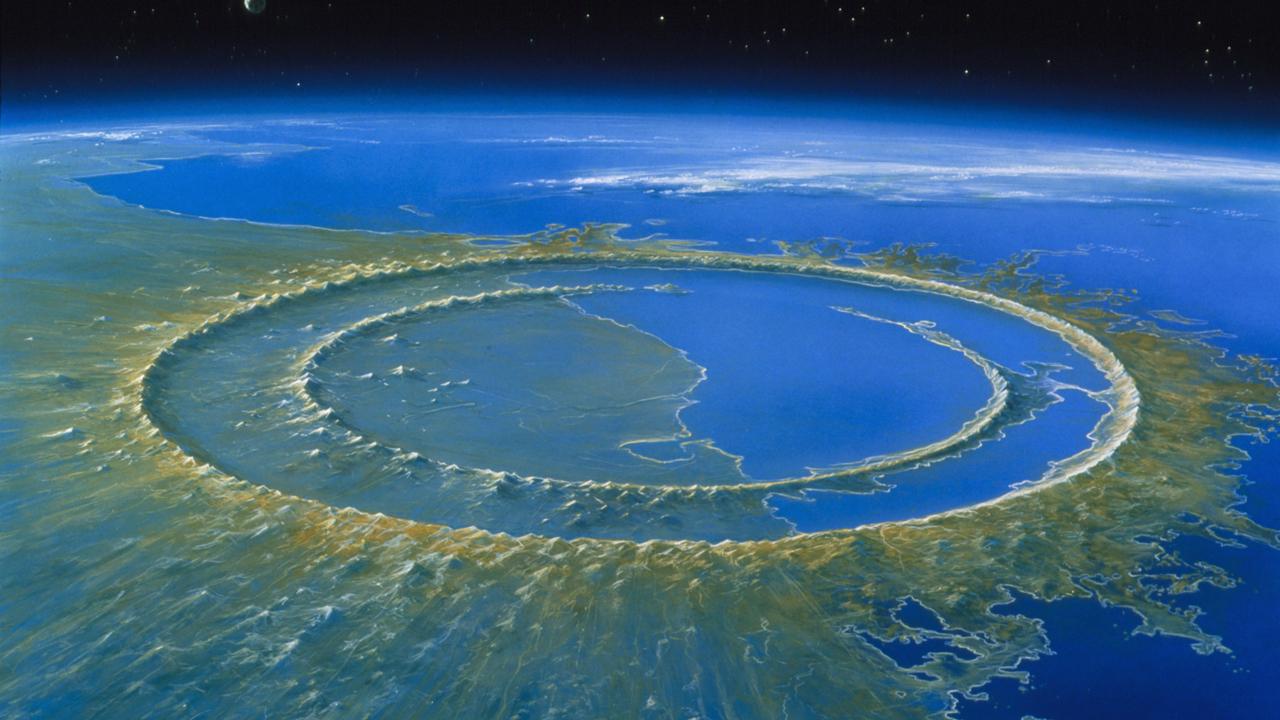
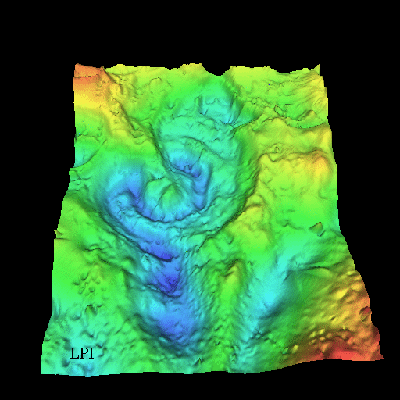


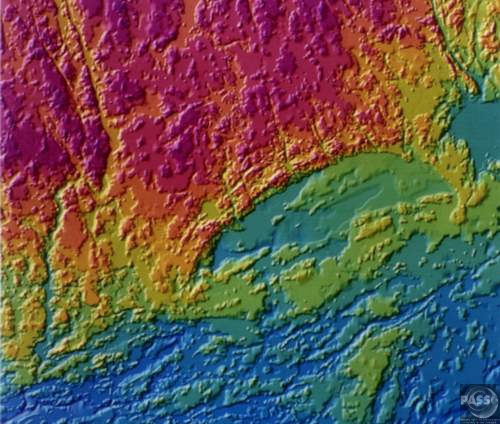
Ilustración 2

Vredefort

Chicxulub

Sudbury

1. Es sencillo apreciar que la mayoría de los cráteres fueron formados por meteoritos de tipo condrita y sus variaciones (*Ilustración 2*), esto significa que provienen de meteoritos no metálicos (rocosos) que no han sufrido procesos de fusión o de diferenciación en los asteroides de los que proceden.

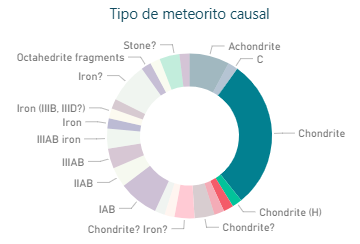


Ilustración 2



Chondrita

1. Nos pareció interesante mostrar un poco sobre los cráteres existentes en nuestro país (*Ilustración 3*). El primero se encuentra en la provincia de El Chaco, cercano a la localidad de Pte. Roque Saenz Peña, llamado Campo del Cielo y el otro está en nuestra provincia, cerca de la ciudad de Rio Cuarto.



Ilustración 3

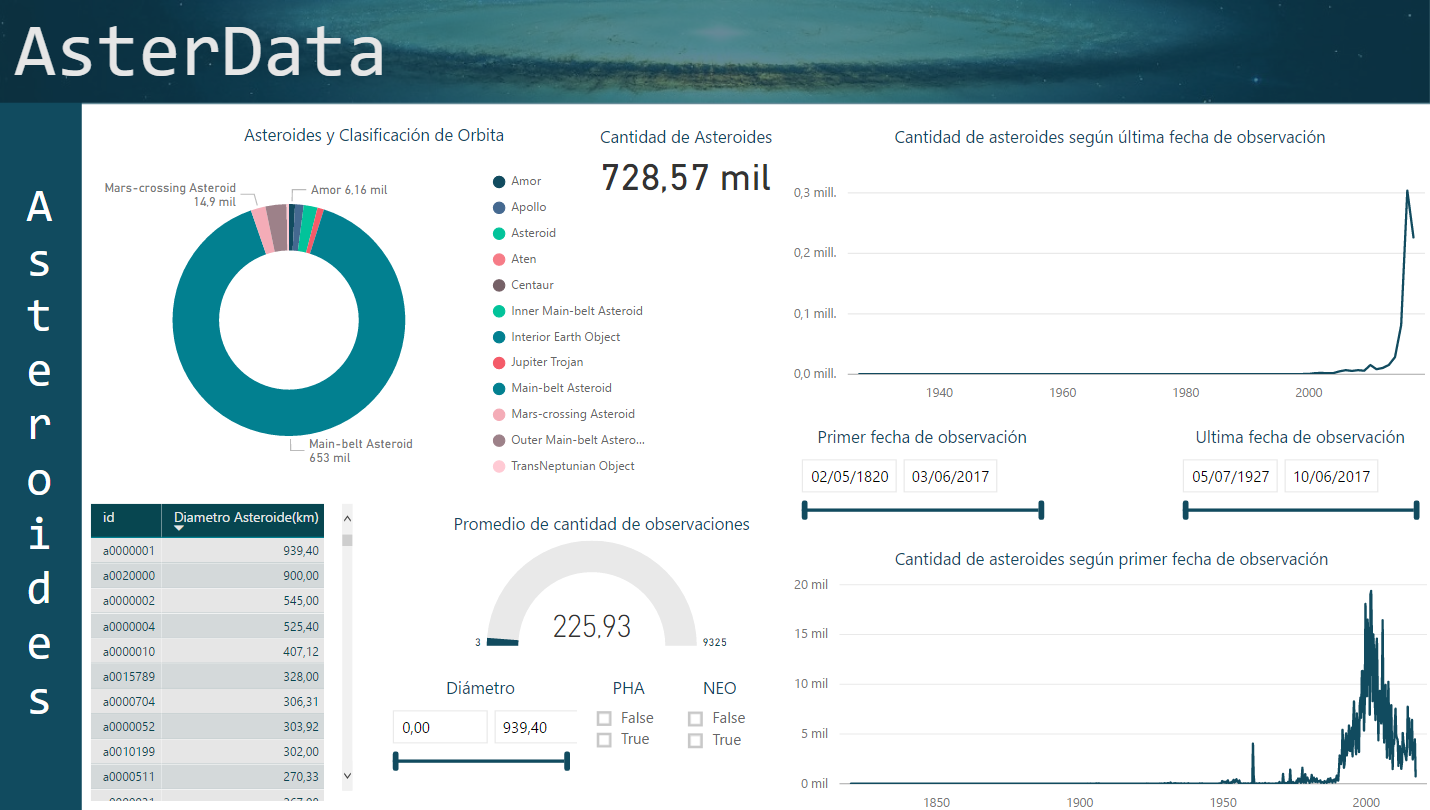


Campo del Cielo

Rio Cuarto

## Asteroides



1

2

3

4

5

6

7

8

9

10

11

En este informe podemos ver dos gráficos similares que muestran la cantidad de asteroides, en un caso según la última fecha de observación (Fig.1) y en el otro, según la primera fecha de observación (Fig.2). Luego se muestran la cantidad de asteroides que hay en cada una de las órbitas (Fig.3).

Para complementar podemos visualizar un listado (Fig.4) con el ID de los asteroides ordenados por su diámetro y un pequeño gráfico que muestra el promedio de cantidad de observaciones por año (Fig.5). Encima de todo, hay un contador del total de asteroides (Fig.6)

Dentro de los filtros, tenemos el diámetro máximo y mínimo (Fig.7), la clasificación, PHA (Asteroide Potencialmente Peligroso – Potentially Hazardous Asteroid) y NEO (Objeto Próximo a la Tierra – Near Earth Object) (Fig.8 y Fig.9). Por último, hay un filtro por primera y última fecha de observación (Fig.10 y Fig.11).

### **Conclusiones**

1. En primer lugar, ya que desde nuestro punto de vista es lo más significativo, analizaremos que ocurre con los asteroides clasificados como PHA. Éstos son los asteroides cuyas órbitas se cruzan con la órbita de la Tierra, es por esto que tienen más probabilidad de producir un impacto con nuestro planeta. En la siguiente figura (*Ilustración 1*), se puede ver que la cantidad de este tipo de asteroides es mucho menor que del total y sólo pertenecen a las órbitas Apollo, Aten, Amor y IEO (Objetos dentro de la Órbita Terrestre). A continuación, también se muestra una imagen (*Ilustración 2*) que describe estas órbitas y como se cruzan con la terrestre.

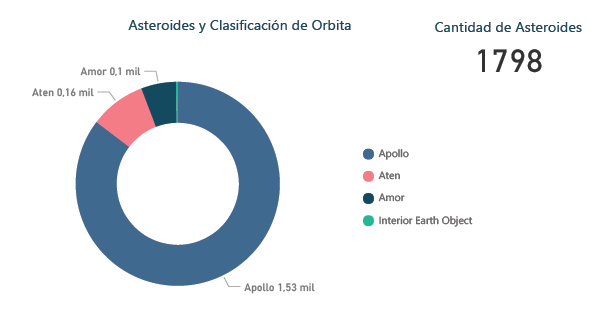
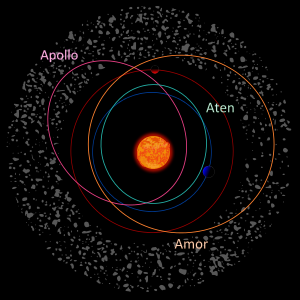
 

Ilustración 3

Ilustración 2

1. Además, los asteroides clasificados como potencialmente peligrosos no tienen un gran tamaño en comparación con los más alejados. El diámetro de los asteroides PHA oscila entre los 80 metros y 7 kilómetros (*Ilustración 3*), mientras que hay asteroides que miden hasta 940 kilómetros de diámetro.

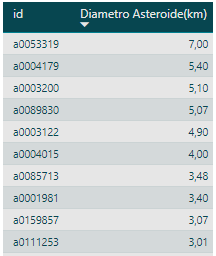


Ilustración 3

1. También se puede observar que la mayoría de los asteroides proviene del cinturón de asteroides (*Ilustración 4*) o cinturón principal (Main-Belt). Este cinturón está ubicado entre las orbitas de Marte y Júpiter. Si bien hay una gran cantidad de elementos en el cinturón, estos están muy dispersos y sería muy difícil chocar con uno de ellos.

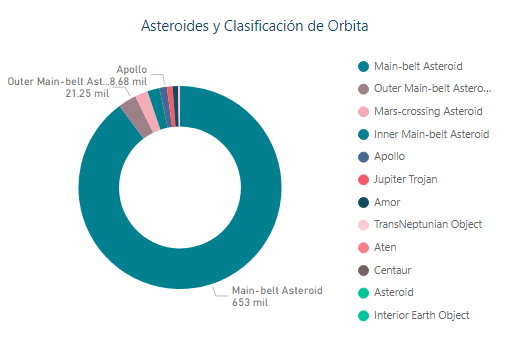


Ilustración 4

1. Analizando con mayor detalle las fechas de observación de los asteroides, podemos ver que a partir del año 2014 se produce un aumento exponencial en la cantidad de visualizaciones (*Ilustración 5*). Esto se puede producir debido al incremento del interés por parte de algunas organizaciones, como la ONU, que demuestran e informan a la sociedad del peligro que puede producir en la humanidad la caída de alguno de estos objetos.

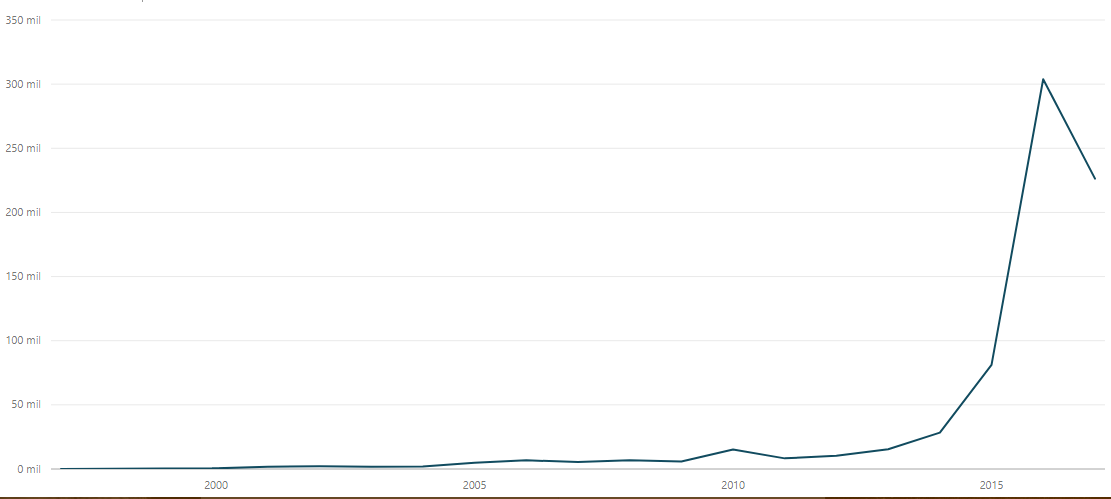


Ilustración 5

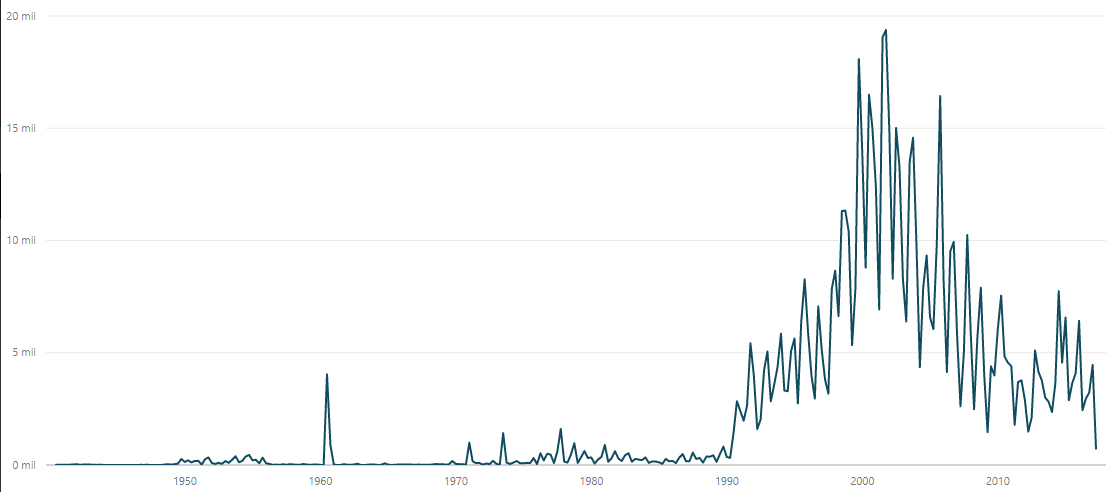
1. Por otra parte, se pueden analizar las fechas de la primera visualización (*Ilustración 6*), es notoria la forma en que van aumentando las observaciones a medido que avanza el tiempo. A partir de 1990, se van encontrando más 3000 asteroides nuevos por año, haciendo un pico alrededor del año 2000. Para esto tenemos algunas hipótesis, una de ellas es que, gracias al avance de la tecnología y la ciencia, se han podido adquirir mejores herramientas y elementos que perfeccionan el avistamiento de estos objetos. También se ha demostrado un creciente interés por parte de la comunidad científica sobre el origen, la forma y las consecuencias de los mismos. Esto puede haber sido principalmente incentivado por un fenómeno que ocurrió en 1989, donde un asteroide podría haber provocado una catástrofe de dimensiones similares a la de la desaparición de los dinosaurios o de miles bombas nucleares. ¡Esta situación no se dio sólo por 6 horas y media!

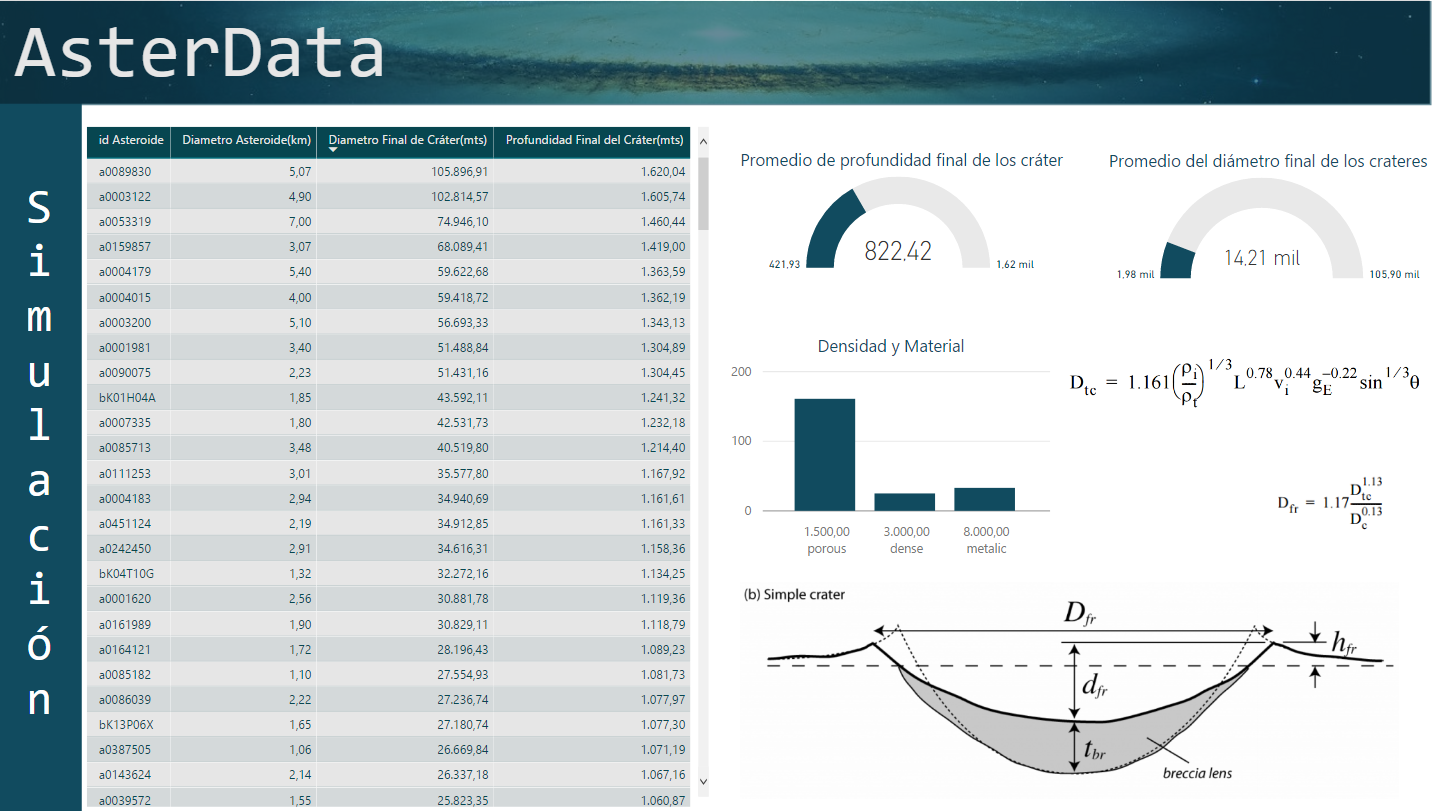
Ilustración 6

1. Pudimos observar, que durante el tercer trimestre se produce una disminución de la cantidad de observaciones en todos los años analizados (*Ilustración 7*). Esto coincide con el otoño/invierno del hemisferio norte (donde trabaja la NASA). No podemos afirmar precisamente porque se fue dando este fenómeno a través de los años pero suponemos que podría ser por la posición de la Tierra con respecto a las orbitas de asteroides observadas.

Ilustración 7

Ilustración 7

## Simulación



1

2

3

4

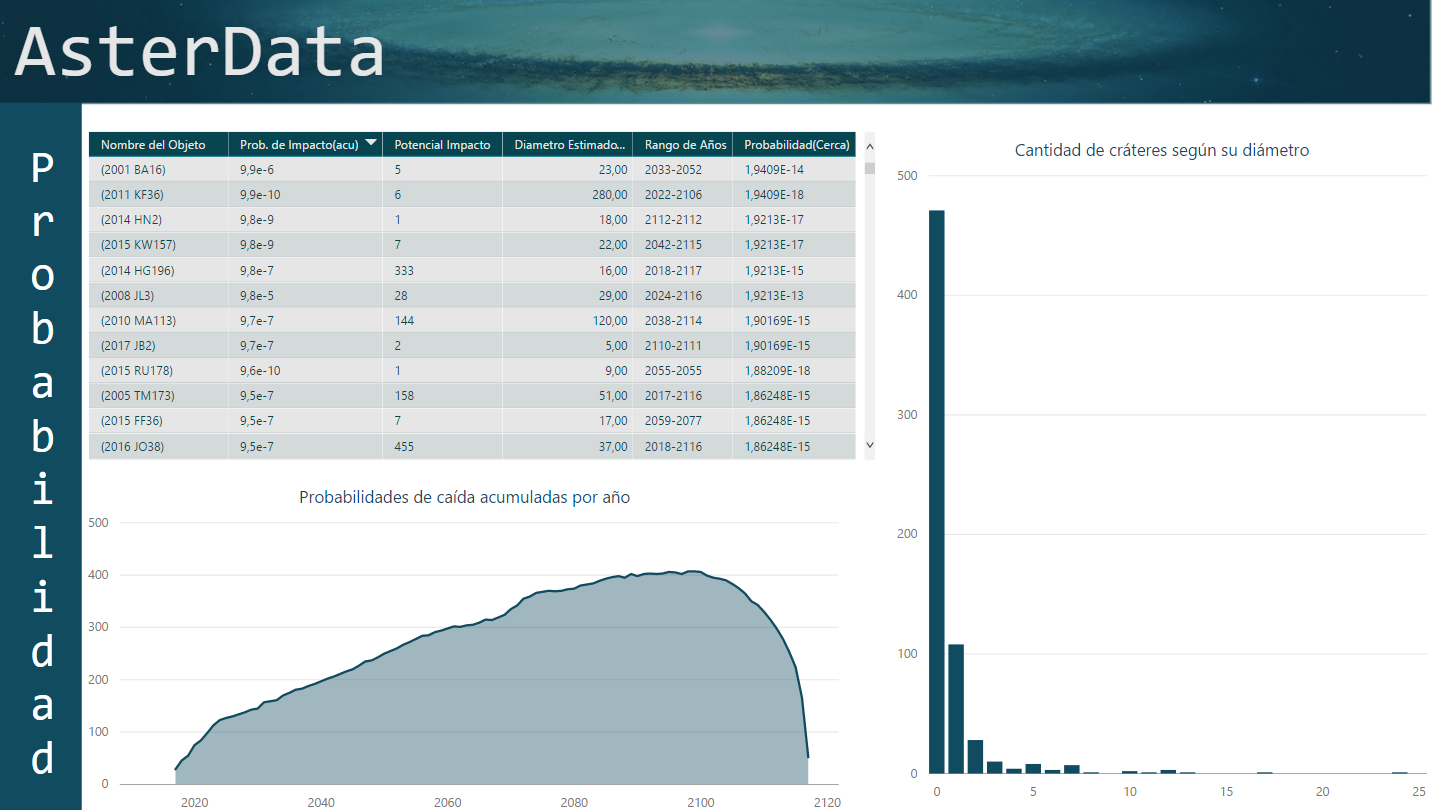
5

Este informe está compuesto por una tabla principal (Fig1) donde se muestran todos los asteroides con su posible cráter final, indicando allí de manera descendente el diámetro del mismo, siendo el mayor de 105 km, es decir aproximadamente 5 veces el diámetro de Circunvalación de Córdoba Capital.

Luego datos más estadísticos, donde la profundidad promedio (Fig2) supera los 800m y el diámetro promedio (Fig3) supera los 14 km. Cómo ya se ha explicado en la simulación, las probabilidades basadas en la clasificación Tholen indican que la mayoría de los asteroides, 75%, están formados por rocas porosas derivadas del carbón (Fig4).

Para completar la idea, se indican gráficos y fórmulas de referencia para los diferentes estados del Cráter al momento del impacto (Fig5)

## Probabilidades



2

1

3

En esta pantalla se muestran las Probabilidades de caída acumuladas por año (Fig1). La principal tendencia es hacia el año 2100, ya que los periodos en los cuales hay mayor cantidad de registros son de muy amplio rango y hacia un futuro que, ya a muchos les cuesta pensar. En promedio, las personas que nacen actualmente tienen la probabilidad de llegar a ver alguna de esas colisiones de efectuarse, pero por lo que pudimos simular no generarían un gran impacto, puesto que el diámetro es mucho menor al kilómetro (Fig3). Por lo tanto, no afectaría el rumbo general de la tierra, pero si de un territorio en particular.

Por otro lado, se muestra una tabla de probabilidades del dataset antes nombrado (Fig2), donde se puede analizar que la probabilidad más alta de caída de uno de esos asteroides es de 5x10-2. Donde también se informa la cantidad de posibles impactos, es decir uno de todos esos impactos es el que va a ocurrir, el diámetro estimado del asteroide, el rango de años y la probabilidad de que caiga en un punto de la tierra. Esta última al ser constante para todos los asteroides, sigue siendo el mismo asteroide el que tiene mayor probabilidad impactar, pero que justo caiga en un km2 a nuestra redonda, genera una probabilidad demasiado chica.

# Conclusión

La realización de este informe, nos llevó a conocer, buscar y aprender no solo las herramientas Power BI, Hadoop, Cloudera, Jupyter, R y R Studio, sino también introducirnos en el tema de meteoritos y asteroides. De esta forma, pudimos investigar sobre temáticas que no teníamos conocimientos previos.

Con este trabajo, nos dimos cuenta cuán importante es el análisis de los meteoritos y asteroides, ya que en cualquier momento puede impactar uno sobre nosotros sin enterarnos. Y gracias a la tecnología, podríamos salvar vidas, realizar papers científicos de gran interés en muchas áreas de investigación como la Geología, la Astronomía, la Matemática, Probabilidades y Estadísticas, etc.

Concluimos que es un tema demasiado amplio y que pueden realizarse muchos estudios y conclusiones, las cuales llevan a investigaciones científicas de gran envergadura. Con datos más actualizados, se podrían obtener predicciones más actuales.