

# **INFORME DE ANÁLISIS DE DATOS: "ANÁLISIS DESCRIPTIVO DE SUPERHÉROES: REVELANDO SECRETOS DE MARVEL, DC Y GEORGE LUCAS MEDIANTE PYTHON Y VISUALIZACIÓN DE DATOS"**

*Realizado por:*

*Jesús David Sánchez*

*Maturín- Monagas- Venezuela*

*Junio del 2024*

## INDICE DE CONTENIDO

<b>INTRODUCCIÓN .....</b>	<b>3</b>
<b>DESARROLLO.....</b>	<b>4</b>
<i>Formulación de las Preguntas.....</i>	<i>4</i>
<i>Objetivo general:.....</i>	<i>5</i>
<i>Objetivo específico: .....</i>	<i>5</i>
<i>Obtención de los datos.....</i>	<i>5</i>
<i>Investigación de los datos .....</i>	<i>5</i>
<i>Limpieza de datos .....</i>	<i>6</i>
<i>Análisis de datos .....</i>	<i>6</i>
<i>Identificar los 5 héroes más poderosos del conjunto de datos. ....</i>	<i>7</i>
<i>Puntualizar el número de héroes con los ojos azules, marrones y verdes. ....</i>	<i>8</i>
<i>Precisar el porcentaje de héroes hombres y mujeres del conjunto de datos. ....</i>	<i>8</i>
<i>Presentación de los resultados .....</i>	<i>9</i>
<b>CONCLUSIÓN.....</b>	<b>10</b>

## INTRODUCCIÓN

*En el presente informe se presenta un análisis de datos descriptivo realizado sobre un conjunto de datos que contiene información detallada sobre superhéroes de Marvel, DC y George Lucas. El objetivo principal de este análisis es obtener una comprensión detallada de estos icónicos personajes a través de la utilización del lenguaje de programación Python y las librerías Pandas y Matplotlib para el procesamiento y visualización de datos.*

*El análisis se llevó a cabo siguiendo un ciclo de trabajo que abarcó desde la formulación de preguntas clave hasta la presentación de los resultados obtenidos. Se plantearon objetivos generales y específicos para guiar el análisis, incluyendo la determinación del número de héroes de cada universo, la identificación de los héroes más poderosos, el conteo de héroes con diferentes colores de ojos, y la precisión del porcentaje de héroes hombres y mujeres en el conjunto de datos.*

*A través de este informe, se busca proporcionar una visión detallada y significativa de los superhéroes analizados, presentando los resultados de manera clara y práctica a través de representaciones gráficas. El análisis descriptivo realizado permitirá obtener una comprensión más profunda de estos personajes icónicos y su relevancia en el mundo del entretenimiento.*

## **DESARROLLO**

*Se realizó un análisis de datos de tipo descriptivo sobre un conjunto de datos que albergaba información de superhéroes de Marvel, DC y George Lucas para obtener información detallada de estos héroes. Para ello se utilizó el lenguaje de programación Python para realizar todo el trabajo, haciendo uso de las librerías que dispone este lenguaje para esto: “Pandas” y “Matplotlib”. Con este trabajo se busca analizar el conjunto de datos y presentar los resultados en forma gráfica con su explicación de la manera más sencilla y práctica. En este análisis de datos se usó el siguiente ciclo de trabajo:*

- 1. Formulación de las preguntas*
- 2. Obtención de los datos*
- 3. Investigación de los datos*
- 4. Preparación los datos (Limpieza)*
- 5. Análisis de los datos*
- 6. Presentación de los resultados*

### **Formulación de las Preguntas**

*Los objetivos a seguir en este análisis, parten del conjunto de preguntas que hay que responder, por lo tanto, se da inicio desde una perspectiva o pregunta general (Objetivo general) hacia el desglose de preguntas mucho más específicas (Objetivos específicos).*

*Los objetivos de este análisis fueron:*

**Objetivo general:**

*Realizar un análisis descriptivo exhaustivo de un conjunto de datos que contiene información sobre héroes de Marvel, DC y George Lucas, con el fin de describir y presentar de manera detallada datos relevantes y significativos sobre estos personajes icónicos.*

**Objetivo específico:**

- *Determinar cuántos Héroes de Marvel, DC y George Lucas hay en el conjunto de datos.*
- *Identificar los 5 héroes más poderosos del conjunto de datos.*
- *Puntualizar el número de héroes con los ojos azules, marrones y verdes.*
- *Precisar el porcentaje de héroes hombres y mujeres del conjunto de datos.*

**Obtención de los datos**

*El conjunto de datos fue descargado de la plataforma comunitaria Kaggle en formato .CSV, el cual contenía 28118 filas y 39 columnas, pesando más de 46 MB.*

**Investigación de los datos**

*Los datos del conjunto de datos, fueron revisados por medio del dataframe que proporciona la librería de “Pandas”, para constatar que, el contenido del mismo estaba relacionado con las preguntas establecidas y contenía la información necesaria para dar respuesta a las preguntas. Gracias a la visualización general que proporciona los Dataframe se pudo comprobar que efectivamente, todo se encontraba en el conjunto de datos, sin embargo este contenía muchos valores basura, por lo que había que limpiar los datos.*

*Cabe resaltar en este punto que se pudo utilizar una herramienta de análisis de datos como Excel para realizar la investigación, pero el conjunto de datos tenía mucho registros como para revisarlos a cabalidad, así que se optó por explotar las potencialidades que ofrece Pandas.*

## **Limpieza de datos**

*Para la limpieza de datos se siguió usando las bondades de Pandas. Se verifico la presencia de datos Nulos, que desde el primer vistazo al Dataframe se había visualizado algunos. Como se esperaba había muchos valores nulos, los cuales se eliminaron. De primera instancia se reconoce que no es una buena práctica, sin embargo, en este caso se buscaba reducir lo máximo posible los datos.*

*Luego se eliminaron un conjunto de columnas que no aportaban nada al estudio, se dejaron otras porque se quería guardar el dataframe para estudios próximos y estas columnas, sí proporcionan información relevante para otros estudios de este estilo.*

*También, Se verificó la presencia de valores duplicados, los cuales si había, por lo que se eliminaron. Para ese momento los datos se redujeron a solo 51 filas. Se hizo otra verificación y había presencia de valores duplicados, por lo que se aplicó algo de lógica a la situación, y se estableció la siguiente premisa: “no pueden haber más de dos personajes con el mismo seudónimo”. Por lo que, se procedió a eliminar los valores duplicados de la columna “Character”.*

*Posterior a ello, para constatar que el nivel de poder de los héroes (columna “Power”) no presentaba una anomalía se hizo un diagrama de dispersión, el cual mostraba pequeñas agrupaciones en rangos que no revelaban anomalías. Con esta verificación se “cerró” la etapa de limpieza de datos, ya que los valores que eran necesarios para dar respuesta se encontraban hasta ese momento limpio.*

## **Análisis de datos**

*Determinar cuántos Héroes de Marvel, DC y George Lucas hay en el conjunto de datos.*

*Para analizar y dar respuesta a esta pregunta, se obtuvo un conteo de cada instancia de la columna "Creator", la cual tiene tres: Marvel, DC y George Lucas. Entonces, obteniendo un contador de cada instancia se puede determinar el número de ocurrencias que hay en cada fila de estas tres instancias. Por lo tanto, se pudo representar en un gráfico de barras que revela que hay 16 héroes de Marvel, 10 de DC y 1 de George Lucas en el conjunto de datos. Estos valores concuerdan con la información obtenida del conjunto de datos ya limpios que para ese punto contaba con 27 registros totalmente limpios. Cada columna asignada a un correspondiente creador, se le colocó un color único: Marvel azul, DC rojo y George Lucas azul marino; para hacer más visual el gráfico.*

#### **Identificar los 5 héroes más poderosos del conjunto de datos.**

*Para identificar a los 5 héroes más poderoso, hay dos opciones considerar única y exclusivamente a los héroes cuyo nivel de poder llegue a 100 o establecer un mínimo, lo cual fue lo que se realizó. El mínimo establecido, fue de un nivel de 80, por lo que se tomó los 5 valores que correspondieran a esa premisa y se ordenaron los resultados en forma ascendente. Como los valores de los datos de la columna "Power" era de tipo objeto, se cambió al tipo de datos enteros, para poder operar sobre valores numéricos que permitiesen determinar cuáles son mayores o menores a otros. El resultado obtenido es:*

*Iron Man y Lex Luthor con un nivel de poder de 80*

*Capitan marvel con un nivel de 90*

*Y finalmente: Thor y Flash con un poder de 100*

*Para los conocedores del tema se ve que hay una combinación de héroes de Marvel y de DC... Esto si se quiere filtrar más, a solo un creador, o etiquetar en cada héroe su propio creador, también se puede hacer, no obstante para*

*efectos de este estudio solo se trabajó con el nombre del personaje y su nivel de poder.*

**Puntualizar el número de héroes con los ojos azules, marrones y verdes.**

*En este caso, se obtuvo como resultado el número de héroes cuyo color de ojos es azul, marrón y verde. La dinámica, de este análisis fue muy parecida al de la primera pregunta. Con la diferencia que se dispuso de un gráfico de barras horizontales para graficar la respuesta, esto se hizo con un propósito de añadirle algo de dinamismo a los gráficos para que no fueran exclusivamente, de barras tradicionales.*

*En los resultados se obtuvieron:*

- *3 héroes con los ojos verdes*
- *5 héroes con los ojos marrones*
- *17 héroes con los ojos azules.*

*Lo que deja a 2 héroes del conjunto de datos, que tienen un color de ojos diferentes a los establecidos, ya que la suma de este resultado es de 25 personajes, y se tienen un total de 27.*

**Precisar el porcentaje de héroes hombres y mujeres del conjunto de datos.**

*Finalmente, para precisar el porcentaje de hombres y mujeres del conjunto de datos se procedió primero a obtener el total del todo, del sexo de los personajes, es decir la suma de héroes femeninos y masculinos. Luego se obtuvo la sumatoria de los héroes hombres y de las héroes mujeres del conjunto de datos. Con estos resultados numéricos se procedió a obtener el porcentaje de cada instancia, y mediante un gráfico circular se presentaron los resultados correspondientes.*

*Los héroes masculinos corresponde a un 81,5 % del conjunto de datos, y las heroínas femeninas corresponden a un 18,5 %. Lo cual es un valor correcto,*



*ya que la suma de ambos porcentajes da como resultado el 100%. Es decir, que hay más héroes varones en el conjunto de datos que heroínas.*

### **Presentación de los resultados**

*Los resultados, representaciones gráficas, breves comentarios y cada paso práctico, dado en este estudio se presentará en un archivo .ipynb . El cual es un archivo de jupyter notebooks, que permite la inclusión de texto y de código Python, y permite una buena forma de visualizar y hacer un seguimiento cronológico de todos los pasos llevados en el análisis de datos.*

*Este archivo, junto el archivo fuente de datos .csv y un documento .pdf (que detallará todo lo realizado) serán subidos a un repositorio de Github.*

## CONCLUSIÓN

*En conclusión, el análisis de datos realizado proporcionó una visión detallada de los superhéroes incluidos en el conjunto de datos. Al identificar a los 5 héroes más poderosos, se encontró una combinación de personajes de Marvel y DC, con Iron Man, Lex Luthor, Capitán Marvel, Thor y Flash destacando por sus niveles de poder. Este análisis permitió comprender la diversidad y relevancia de estos icónicos personajes en el mundo del entretenimiento.*

*Además, al puntualizar el número de héroes con diferentes colores de ojos, se observó que la mayoría de los superhéroes tienen ojos azules, seguidos por aquellos con ojos marrones y verdes. Esta visualización a través de un gráfico de barras horizontales agregó dinamismo al análisis, permitiendo una comprensión clara de la distribución de colores de ojos entre los personajes.*

*Finalmente, al precisar el porcentaje de héroes hombres y mujeres en el conjunto de datos, se evidenció que los héroes masculinos representan la gran mayoría, con un 81.5%, mientras que las heroínas femeninas constituyen un 18.5%. Este hallazgo proporciona información relevante sobre la representación de género en el mundo de los superhéroes, destacando la predominancia de personajes masculinos en el conjunto de datos analizado.*

*En resumen, el análisis descriptivo llevado a cabo proporcionó una comprensión detallada y significativa de los superhéroes de Marvel, DC y George Lucas, ofreciendo información valiosa sobre su poder, características físicas y representación de género. Estos resultados contribuyen a una mejor comprensión de la diversidad y relevancia de estos personajes icónicos en la cultura popular.*