



# UNIVERSIDAD POLITÉCNICA DE CHIAPAS

Reporte de proyecto:

## PERSONAS QUE HAYAN VISTO LA SAGA DE **STAR WARS**

### Probabilidad y Estadística

Alumno: Ada Isabel Gamboa Chanona

Docente: Doc. Carlos Alberto Díaz Hernández



## 1. CUADRO DE IDENTIFICACIÓN

<b>CUATRIMESTRE</b>	Mayo-Agosto
<b>CUATRIMESTRE Y GRUPO</b>	6to "A"
<b>ASIGNATURA</b>	Probabilidad y Estadística
<b>ACTIVIDAD</b>	PE.C1.A3 Reporte de proyecto
<b>FECHA DE ASIGNACIÓN</b>	Sábado, Mayo/29/2021
<b>FECHA DE ENTREGA</b>	Sábado, Mayo/29/2021
<b>MATRÍCULA</b>	193223
<b>NOMBRE</b>	Ada Isabel Gamboa Chanona

## CONTENIDO

1. CUADRO DE IDENTIFICACIÓN.....	2
2. LISTA DE FIGURAS .....	3
3. INTRODUCCIÓN .....	4
3.1 Descripción del problema .....	4
4. MÉTODOS.....	5
4.2 Selección de variables.....	5
4.2.1 Variable Cualitativa.....	5
4.2.2 Variable Cuantitativa Discreta .....	5
4.2.3 Variable cuantitativa Continua .....	5
4.3 Obtención de los datos .....	5
4.4 Análisis de datos .....	5
4.5 Presentación de datos .....	6

5. IMPLEMENTACIÓN .....	7
6. RESULTADOS .....	8
7. DISCUSIÓN.....	11
8. CONCLUSIÓN .....	12
9. BIBLIOGRAFÍA.....	12

## 2. LISTA DE FIGURAS

<b>Figura 1. Ejemplo de Polígono de Frecuencia .....</b>	<b>6</b>
<b>Figura 2. Ejemplo de Gráfica de Pastel .....</b>	<b>6</b>
<b>Figura 3. Ejemplo de Diagrama de Barras .....</b>	<b>6</b>
<b>Figura 4. Python .....</b>	<b>7</b>
<b>Figura 5. Numpy .....</b>	<b>7</b>
<b>Figura 6. Pandas .....</b>	<b>7</b>
<b>Figura 7. Matplotlib .....</b>	<b>7</b>
<b>Figura 8. Archivo csv .....</b>	<b>9</b>
<b>Figura 9. Datos .....</b>	<b>9</b>
<b>Figura 10. Diagrama Polígono.....</b>	<b>10</b>
<b>Figura 10. Diagrama Barra.....</b>	<b>10</b>
<b>Figura 10. Diagrama Pastel.....</b>	<b>11</b>

### 3. INTRODUCCIÓN

La estadística es una ciencia y una rama de las matemáticas a través de la cual se recolecta, analiza, describe y estudia una serie de datos a fin de establecer comparaciones o variabilidades que permitan comprender un fenómeno en particular.

Hay varios tipos de estadística, pero en este proyecto utilizaremos la estadística descriptiva, que nos permite presentar de manera resumida y organizada los datos numéricos obtenidos tras un estudio o análisis en particular. Su objetivo, por lo tanto, es describir las características principales de los datos reunidos y evitar generalizaciones.

En el proceso que seguiremos para llevar a cabo este tipo de estadística, será como primeros pasos la selección del tema, la obtención de los datos y el análisis de los mismo, presentándolos en diagramas prácticos como tablas y gráficas

El tema a tratar es “Las personas que han visto la saga de **STAR WARS**” donde averiguaremos datos como la película favorita de la persona, la cantidad de películas que ha visto y el tiempo que ha tomado viéndose las películas o la saga completa. Con ello podremos evaluar, por ejemplo, la película más popular, si las personas en general les gusta esta saga y el fanatismo que presentan conforme el tiempo que invierten viendo la franquicia.

#### 3.1 Descripción del problema

La saga de películas de STAR WARS es una franquicia que se presentó en pantalla desde 1977, siendo su primera película ‘STAR WARS: Una nueva esperanza’ después del éxito de la película empezó a salir una saga completa y en la actualidad la franquicia tiene 9 películas principales sin contar las series y demás películas que cultivan la historia, debido a la cantidad de películas, resulta algo tedioso y muy tardado ver toda la saga si en realidad no es de tu gusto, del mismo modo, a las personas pueden resultarles ligeras y entretenidas si son verdaderos fans. Por ello, esta investigación busca obtener a las personas que han visto la saga y determinar el gusto por la franquicia o una de las películas, el tiempo que invierten en verla y la cantidad de películas que han visto, utilizando preguntas que al final serán presentadas en gráficas para analizar los resultados.

## **4. MÉTODOS**

### **4.2 Selección de variables**

Para seleccionar las variables que se evaluarán se han establecido criterios sobre los requerimientos del proyecto es por ello, que se establecieron tres variables principales para abarcar el tema

#### **4.2.1 Variable Cualitativa**

La variable cualitativa que se establecerá será '*¿Cuál es tu película favorita de la saga?*' tomando como respuesta cada una de las películas principales de la franquicia, la variable se identifica como cualitativa porque no representan a un número o valor contable.

#### **4.2.2 Variable Cuantitativa Discreta**

La variable cuantitativa discreta se representará con la cantidad de películas que ha visto el encuestado, es discreta debido a que la cantidad de películas se reduce a nueve, por lo tanto, es un valor entero y definido.

#### **4.2.3 Variable cuantitativa Continua**

Para la variable cuantitativa continua se preguntará el tiempo que ha invertido viendo las películas, siendo continua debido a que el individuo puede haber invertido 0 horas (en el caso de que no haya visto las películas o 1 hora, 1 hora y media, etc. En caso de que haya visto la mitad de una película, una película completa o haya hecho un maratón de la saga.

### **4.3 Obtención de los datos**

Para obtener los datos necesarios se usará la herramienta Google Forms, en donde se realizará la encuesta para posteriormente compartirla a 200 personas en redes sociales (ya sea WhatsApp o Facebook) para que sea contestada y así poder analizar los datos obtenidos, adquiriendo así un tipo de muestra no probabilística, que, en este caso, es la manera más rápida y efectiva para recabar la información.

### **4.4 Análisis de datos**

Para el análisis de datos se recogerán los obtenidos en la encuesta y se analizarán cada uno de ellos utilizando las medidas de tendencia central. En los datos cuantitativos, se sacará la media aritmética para obtener el promedio de los datos, la media geométrica y la media

truncada con un porcentaje de 10% con la finalidad de analizar cada uno de los resultados para cálculos posteriores.

Del mismo modo se calculará la mediana de los datos cuantitativos, con la finalidad de obtener el dato central de la serie de datos, y se realizará el cálculo de la moda, para corroborar la frecuencia más alta extraída.

También recabaremos otras características con el análisis de los datos, tales como el sesgo de la gráfica que se presentará, la varianza muestral, la desviación estándar y del mismo modo la distribución de frecuencia de los datos cuantitativos continuos, tomando en cuenta conceptos como el rango, el número de clases, ancho de la clase, limites, marca de clase, y la distribución de frecuencia absoluta y relativa.

## 4.5 Presentación de datos

Para presentar los datos del análisis se usará la herramienta Excel donde se almacenarán los datos obtenidos en diferentes tablas.

Para las gráficas se utilizarán diferente presentación para cada variable, se dispondrá de una *gráfica de polígono de frecuencia* para los datos continuos como la que aparece en la Figura 1. Para los datos discretos, se especificarán en un *diagrama de barras* (Figura 3) en donde se sacará la frecuencia relativa de los datos para poder realizarla, y para la variable cualitativa se representarán los datos en una *gráfica de pastel* (Figura 2)



Figura 1. Ejemplo de Polígono de Frecuencia

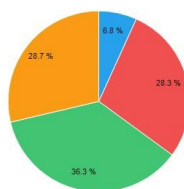


Figura 2. Ejemplo de Gráfica de Pastel

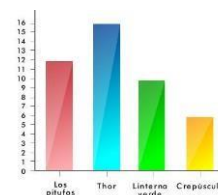


Figura 3. Ejemplo de Diagrama de Barras



## 5. IMPLEMENTACIÓN

Para la implementación del proyecto y enfocarlo en el desarrollo de software, se realizará un programa en Python utilizando Ciencia de Datos para llevar a cabo todas las operaciones estadísticas del proyecto y del mismo modo que sea capaz de realizar las gráficas con solo meter un archivo .csv. El programa tendrá como característica que dispondrá de una interfaz visual para facilitar el uso de la aplicación. Para realizarlo, se usarán diferentes métodos y librerías de Python que ayudarán a automatizar el proceso de programación, así como también nos brindarán todos los recursos necesarios, ya que estas librerías contienen características específicas que se pueden aplicar a los cálculos estadísticos, por ejemplo:

- Numpy: Es una biblioteca para el lenguaje de programación Python que da soporte para crear vectores y matrices grandes multidimensionales, junto con una gran colección de funciones matemáticas de alto nivel para operar con ellas.
- Pandas: Es una biblioteca de Python que sirve para tener un objeto DataFrame rápido y eficiente para la manipulación de datos con indexación integrada. Del mismo modo es una excelente herramienta para leer y escribir datos entre estructuras de datos en memoria y diferentes formatos: archivos de texto y CSV, Microsoft Excel, bases de datos SQL y el rápido formato HDF5;
- Matplotlib: Es una biblioteca completa para crear visualizaciones estáticas, animadas e interactivas en Python. Sirve para desarrollar gráficos de calidad y figuras interactivas que puedan hacer zoom, desplazarse, actualizar etc.

En general, estas son las herramientas que se utilizarán para la elaboración del programa y con las cuales harán que funcione correctamente, del mismo modo para la interfaz visual se utilizará la librería TKINTER.



Figura 4. Python



Figura 5. Numpy



Figura 6. Pandas



Figura 7. Matplotlib

El proceso que llevamos a cabo es sumamente sencillo y la lógica del programa se basa en utilizar los métodos de las librerías correspondientes para obtener lo que se especificó párrafos antes.

En primer lugar, realizamos la encuesta y sacamos el archivo CSV, que será importado hacia el programa por funciones de TKINTER, a continuación, dicho archivo se convertirá en respectivos DATAFRAMES y ARRAYS para separar los tres tipos de variables. El arreglo de cada una de las variables pasara por diferentes funciones para calcular las medidas de tendencia central y dispersión tanto para datos agrupados como para datos no agrupados.

Al obtener todos los datos correspondientes y tenerlos debidamente guardados se procederá a organizar la información, en la implementación se hicieron cuatro tablas, tres representativas de cada variable, donde definen, en caso de los agrupados, el número de clases, los límites superior e inferior, la marca de clase, y los dos tipos de frecuencia, en cuestión de la organización de los datos de las siguientes dos variables se distribuyeron en sus respectivas tablas definiendo las frecuencias. En la cuarta tabla se agregaron las medidas de tendencia central para datos agrupados como para no agrupados (media aritmética, media Geométrica, media truncada, mediana, moda, varianza, desviación estándar y sesgo)

## **6. RESULTADOS**

Los resultados de la implementación se podrán visualizar a la hora de ejecutar el programa, observarán una ventana emergente como la que se muestra en la figura 8. Se consiguió poder importar el archivo CSV (figura 8), en cuanto el sistema haya recabado el documento procederá a abrir la interfaz mostrando todas las tablas de las diferentes variables (figura 9), del mismo modo se podrá acceder a sus respectivos diagramas con los botones colocados bajo las tablas (figura 10). Con este diseño se podrá visualizar de manera más accesible la información que se recabo en el proyecto



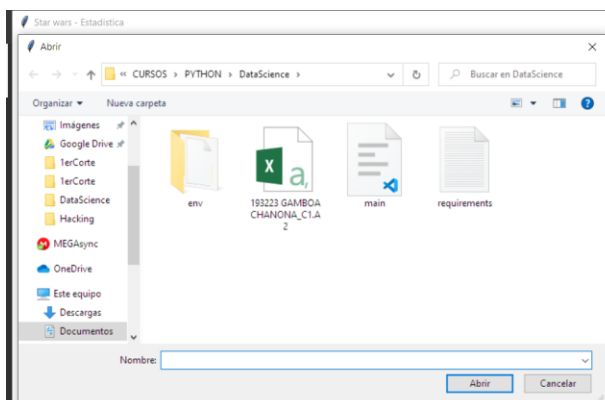


Figura 8. Archivo csv

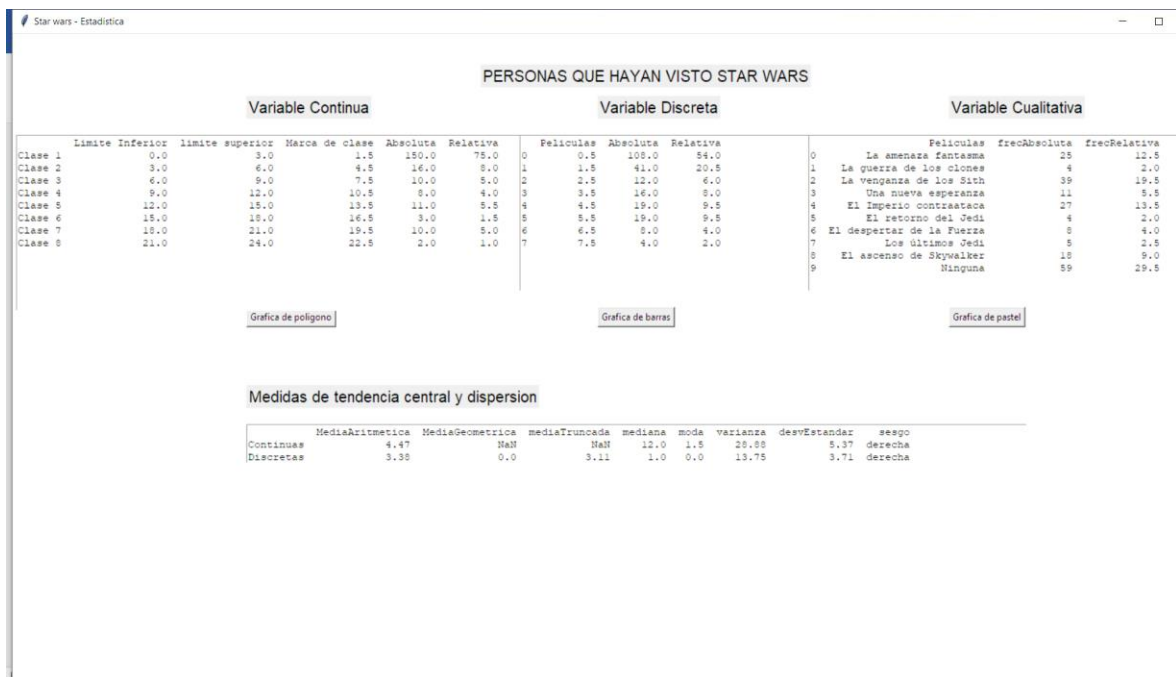


Figura 9 .Datos

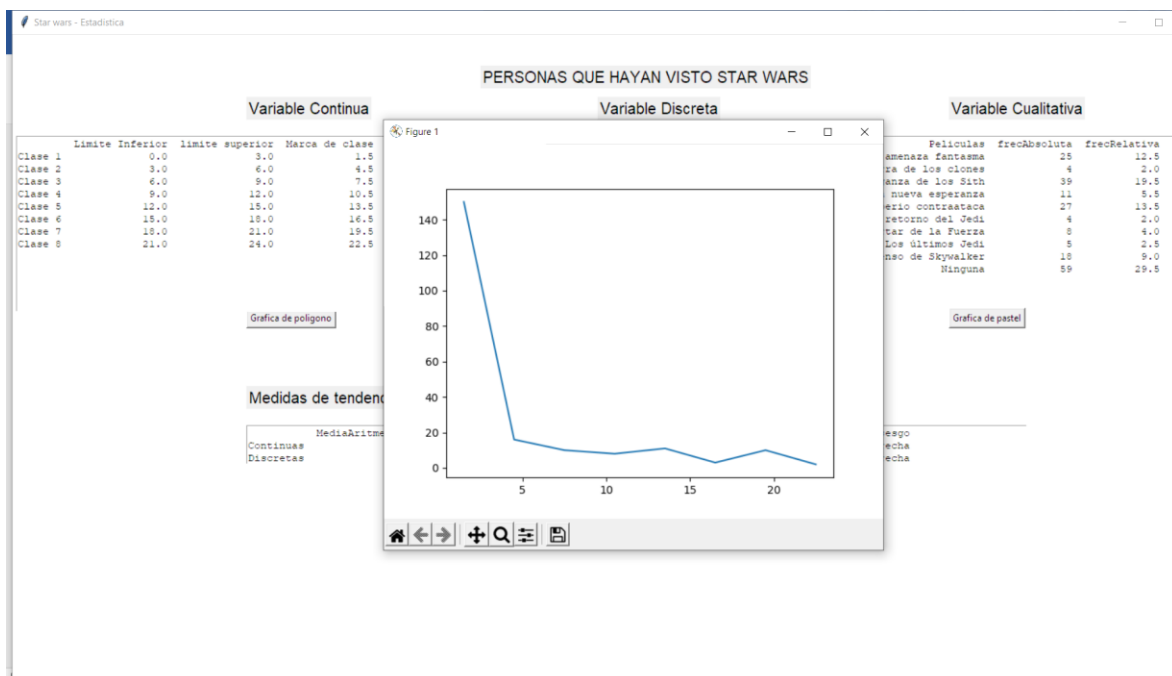


Figura 10. Diagrama polígono

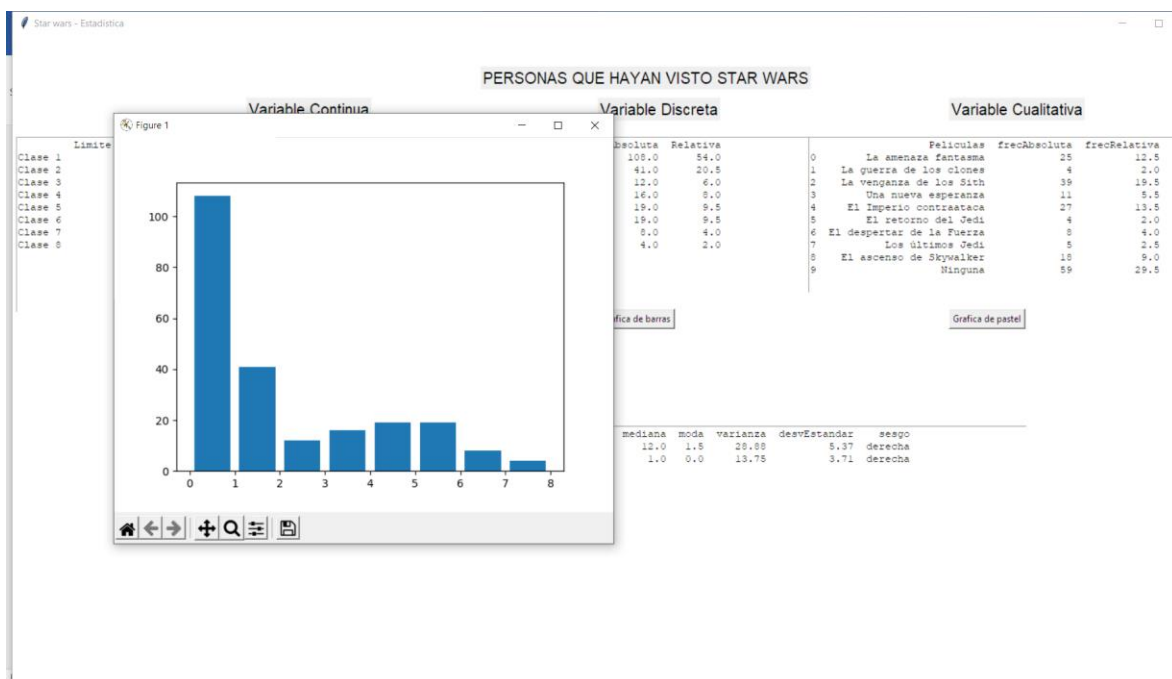


Figura 10. Diagrama de barras

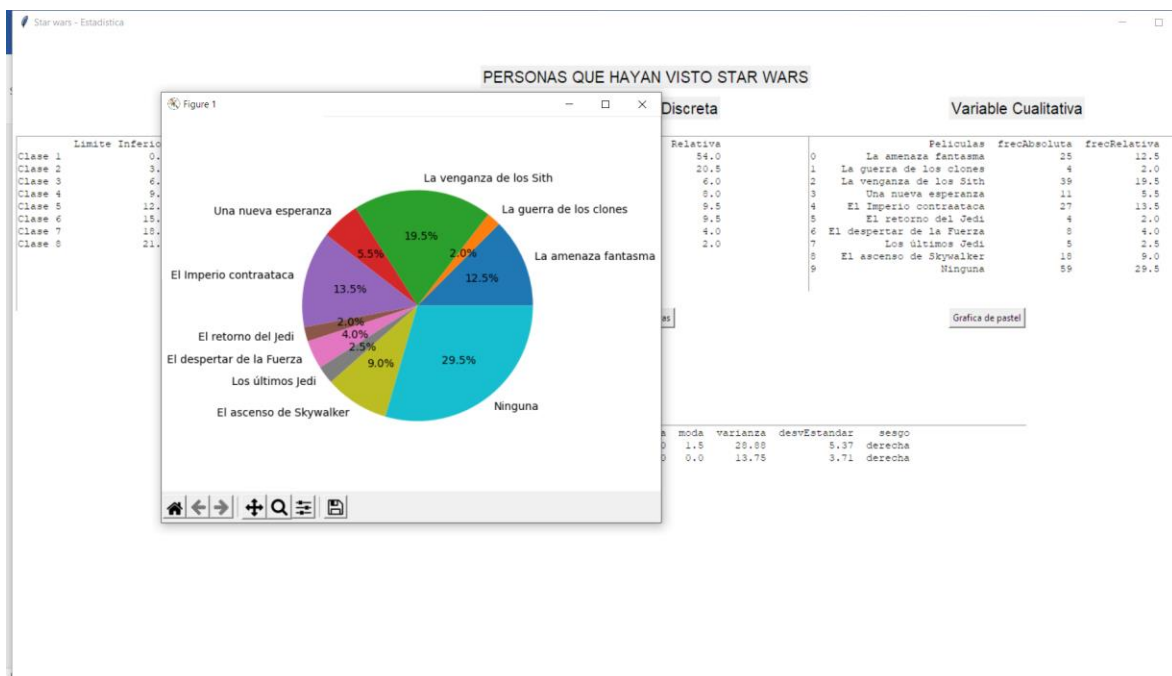


Figura 10. Diagrama de pastel

## 7. DISCUSIÓN

Los datos de una encuesta siempre tienen información que dar y este no es la excepción, como se puede apreciar en cada uno de los diagramas, estos representan los resultados obtenidos del proyecto.

En la primera variable, determinada como variable discreta se buscaba obtener el número de películas de la saga de Star Wars que las personas encuestadas habían visto, en los resultados, podemos ver que en promedio las personas apenas ven de cero a una película de la saga, se llegó a esta conclusión ya que ese rango presentó la mayor frecuencia

En la segunda variable, que fue descrita como variable cualitativa se proponía encontrar la película más vista o de mayor favoritismo entre los encuestados, sin embargo, debido a que el mayor número de encuestados apenas ha visto una o dos películas el resultado con mayor proporción oscila en ninguna, teniendo en segundo lugar Star Wars III La venganza de los sith, siendo la película mayor votada por los individuos.

En la tercera variable, se planteaba obtener la cantidad de horas que los entrevistados gastaban en ver las películas o la saga completa, tomándola como variable continua se presentó en una gráfica de polígono donde se puede apreciar que el mayor número de personas dedica cero horas en ver la saga, sin embargo, también se encontraron resultados de personas que llegan a gastar de entre 15 a 20 horas para ver las películas sin embargo su frecuencia es muy baja.

## 8. CONCLUSIÓN

Se puede concluir que a pesar del gran éxito que ha tenido la saga en épocas de antaño en el presente apenas hay personas que han disfrutado de ella, debido a circunstancias propias que he de suponer oscilan entre la poca atención que tienen hacia las películas a la gran cantidad de películas que hay, lo cual puede resultar exasperante para algunos individuos, es por ello que en los datos obtenidos la mayoría de los resultados fueron nulos y negativos, se distinguió que a las personas encuestadas no le gusta la saga y fueron pocos los que en verdad vieron alguna película y cada vez menos los que apreciaron todas las películas y se consideran fans de la serie. Con los datos obtenidos podemos llegar a pensar que el auge que tenía antes esta serie de películas está llegando a su fin dejando solamente a las personas que son entusiastas del género.

## 9. BIBLIOGRAFÍA

Numpy.org. 2021. What is NumPy? — NumPy v1.20 Manual. [online] Available at: <<https://numpy.org/doc/stable/user/whatisnumpy.html>> [Accessed 13 May 2021].

Significados. 2021. Significado de Estadística. [online] Available at: <<https://www.significados.com/estadistica/>> [Accessed 13 May 2021].

Fandom, S., 2021. Star Wars. [online] Star Wars Wiki. Available at: <[https://starwars.fandom.com/es/wiki/Star\\_Wars#:~:text=Logotipo%20de%20Star%20Wars.&text=Star%20Wars%2C%20traducido%20al%20espa%C3%B1ol,ficci%C3%B3n%20creado%20por%20George%20Lucas.](https://starwars.fandom.com/es/wiki/Star_Wars#:~:text=Logotipo%20de%20Star%20Wars.&text=Star%20Wars%2C%20traducido%20al%20espa%C3%B1ol,ficci%C3%B3n%20creado%20por%20George%20Lucas.)> [Accessed 13 May 2021].

Pandas.pydata.org. 2021. pandas - Python Data Analysis Library. [online] Available at: <<https://pandas.pydata.org/about/index.html>> [Accessed 13 May 2021].

Matplotlib.org. 2021. Matplotlib: Python plotting — Matplotlib 3.4.2 documentation. [online] Available at: <<https://matplotlib.org/>> [Accessed 13 May 2021].