**Memoria EDA**

En el presente proyecto se plasma un poco la idea de que la astronomía, aunque tiene temáticas bastante interesantes, no es muy conocida por el público en general más allá de tópicos básicos. Nos basamos en esta premisa para la hipótesis o el contexto general del EDA “¿Astronomía o astrología?”.

Las hipótesis abarcadas en el proyecto son:

* Mientras las estrellas son más grandes, son más brillantes.
* No existen muchos planetas habitables.
* Se puede saber si una estrella se está acercando o alejando, observándola a través del tiempo.

Durante el desarrollo del trabajo al buscar documentación había mucha, sin embargo, no toda la documentación era útil para probar las distintas hipótesis planteadas.

En sitios como “Kaggle” existen distintos datos acerca de estrellas, sin embargo, son datos fotométricos que se utilizan para mediciones distintas a los cálculos físicos necesarios para comprobar las hipótesis. En el caso de sitios como “Ninja API”, ocurre algo similar.

Se extrajeron inicialmente datos desde “Rapidapi”, “Ninja API” y “Sloan” (desde Kaggle), pero las bases de datos más extensas, como era de esperarse acerca de las estrellas, se encuentran en el catálogo de exo-planetas de la NASA.

Inicialmente se intentó hacer web-scrapping del catálogo, pero finalmente no fue necesario, ya que la NASA cuenta con un portal para descargar los datos en distintos formatos, como el formato .csv.

Luego de cargar los datos y un análisis previo, se observan muchos datos nulos debido a que no todas las estrellas cuentan con información de sus atributos, debido a la naturaleza complicada de la obtención de este tipo de datos.

Al proceder a extraer la información necesaria para analizar la primera hipótesis debemos filtrar para eliminar valores nulos y evitar repetición de datos, ya que algunos de los datos se repiten al haber hecho la NASA o alguna otra entidad respectiva al tema varias mediciones en años distintos.

Luego de limpiar los datos podemos comparar las distintas características de las estrellas y realizar los respectivos análisis, esto en el caso de las primeras dos hipótesis, ya que en el caso de la tercera hipótesis utilizamos otros datos que representan una recopilación de datos de espectros de estrellas.

Alguna complicación pudo surgir a la hora de limpiar el dataset y observar los datos en gráficas, siendo necesario el uso de escalas logarítmicas y normalización para poder llevar a cabo dicha tarea.

Una de las principales tareas fue entender el porqué de cada uno de los fenómenos que experimentan las estrellas, por lo que la lectura de documentación en sitios como Wikipedia u otros fue fundamental.

Finalmente, pudieron ser analizadas las distintas hipótesis. La descripción del desarrollo de cada uno de los procedimientos se encuentra detallada en el archivo .ipynb del presente EDA.