



UNIVERSIDAD TECNOLÓGICA

ESCUELA DE POSGRADOS “ESPOG”

MAESTRÍA EN BIG DATA Y CIENCIA DE DATOS

Resolución: RPC-SO-32-No.536-2023

PROYECTO DE TITULACIÓN EN OPCIÓN AL GRADO DE MAGÍSTER

Título del trabajo:

Análisis de parámetros físicos que clasifican las estrellas aplicado a una muestra del repositorio de observaciones espaciales Gaia

Línea de Investigación:

Ciencias de la ingeniería aplicadas a la producción, sociedad y desarrollo sustentable

Campo amplio de conocimiento:

Tecnologías de la Información y la Comunicación (TIC)

Autor/a:

Paul Estuardo Andrade Alban

Tutor/a:

Renato Mauricio Toasa Guachi

Mario Ruben Pérez Cargua

Quito – Ecuador

2025

APROBACIÓN DEL TUTOR



Yo, Mario Ruben Pérez Cargua portador de la C.I: 0603251984 en calidad de Tutores del trabajo de investigación titulado: Análisis de parámetros físicos que clasifican las estrellas aplicado a una muestra del repositorio de observaciones espaciales Gaia.

Elaborado por: Paul Estuardo Andrade Alban, C.I: 0501454250, estudiante de la Maestría: Big Data y Ciencia De Datos de la **UNIVERSIDAD TECNOLÓGICA ISRAEL (UISRAEL)**, para obtener el Título de Magister, me permito declarar que luego de haber orientado, estudiado y revisado el trabajo de titulación de posgrado, la apruebo en todas sus partes.

Quito D.M., Marzo de 2025

Firma

APROBACIÓN DEL TUTOR



Yo, Renato Mauricio Toasa Guachi portador de la C.I: 1804724167 en calidad de Tutores del trabajo de investigación titulado: Análisis de parámetros físicos que clasifican las estrellas aplicado a una muestra del repositorio de observaciones espaciales Gaia.

Elaborado por: Paul Estuardo Andrade Alban, C.I: 0501454250, estudiante de la Maestría: Big Data y Ciencia De Datos de la **UNIVERSIDAD TECNOLÓGICA ISRAEL (UISRAEL)**, para obtener el Título de Magister, me permito declarar que luego de haber orientado, estudiado y revisado el trabajo de titulación de posgrado, la apruebo en todas sus partes.

Quito D.M., Marzo de 2025

Firma

DEDICATORIA

El presente trabajo va dedicado al padre Dios, por la sabiduría y fortaleza al desarrollar mi tesis.

También a mi hermano José R. Andrade, por su consejo y motivación.

AGRADECIMIENTO

A Renato Mauricio Toasa Guachi y Mario Ruben Pérez Cargua en su ayuda, colaboración y orientación en el desarrollo de este trabajo.

A mi Hermano en su consejo brindado.

A mis Progenitores por sus consejos.

A mi perro Osito que ya no está conmigo, que en paz descance.

DECLARACIÓN DE AUTORIZACIÓN POR PARTE DEL ESTUDIANTE



Yo, Paul Estuardo Andrade Alban portador de la C.I: 0501454250, autor/a del trabajo de titulación denominado: Análisis de parámetros físicos que clasifican las estrellas aplicado a una muestra del repositorio de observaciones espaciales Gaia. Previo a la obtención del título de Maestría En Big Data y Ciencia De Datos.

1. Declaro tener pleno conocimiento de la obligación que tienen las instituciones de educación superior, de conformidad con el Artículo 144 de la Ley Orgánica de Educación Superior, de entregar el respectivo trabajo de titulación para que sea integrado al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.
2. Manifiesto mi voluntad de ceder a la Universidad Tecnológica Israel los derechos patrimoniales consagrados en la Ley de Propiedad Intelectual del Ecuador, artículos 4, 5 y 6, en calidad de autor@ del trabajo de titulación, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente. En concordancia suscribo este documento en el momento que hago entrega del trabajo final en formato impreso y digital como parte del acervo bibliográfico de la Universidad Tecnológica Israel.
3. Autorizo a la SENESCYT a tener una copia del referido trabajo de titulación, con el propósito de generar un repositorio que democratice la información, respetando las políticas de prosperidad intelectual vigentes.

Quito D.M., Marzo de 2025



Firmado electrónicamente por:
**PAUL ESTUARDO
ANDRADE ALBAN**

Firma

Tabla de contenidos

APROBACIÓN DEL TUTOR	ii
APROBACIÓN DEL TUTOR	iii
DEDICATORIA	iv
AGRADECIMIENTO	v
DECLARACIÓN DE AUTORIZACIÓN POR PARTE DEL ESTUDIANTE	vi
INFORMACIÓN GENERAL.....	10
Contextualización del tema	10
Problema de Investigación.....	11
Objetivo general	11
Objetivos específicos	11
Vinculación con la sociedad y beneficiarios directos.....	11
CAPÍTULO I: DESCRIPCIÓN DEL PROYECTO	13
1.1. Contextualización general del estado del arte	13
1.2. Proceso investigativo metodológico	15
1.3. Análisis de resultados.....	16
CAPÍTULO II: PROPUESTA.....	18
2.1 Fundamentos teóricos aplicados	18
2.2 Descripción de la Propuesta.....	23
2.3 Validación de la propuesta.....	38
2.4 Matriz de articulación	41
CONCLUSIONES.....	43
RECOMENDACIONES.....	44
BIBLIOGRAFÍA.....	45
ANEXOS.....	47

Índice de tablas

Tabla 1 <i>Resumen de Clases Espectrales</i>	20
Tabla 2 <i>Tratamiento de Valores Nulos</i>	25
Tabla 3 <i>Perfil de Especialistas</i>	39
Tabla 4 <i>Criterios de Valoración</i>	39
Tabla 5 <i>Análisis de Valoración de los Especialistas</i>	40
Tabla 6 <i>Matriz de Articulación</i>	41

Índice de figuras

Figura 1 <i>Repositorio de Gaia DR3</i>	17
Figura 2 <i>Diagrama del Ciclo de Investigación</i>	23
Figura 3 <i>Carga del Conjunto de Datos</i>	24
Figura 4 <i>Proceso de Limpieza</i>	26
Figura 5 <i>Selección de Columnas</i>	27
Figura 6 <i>Actualización de Nombres de Columnas</i>	28
Figura 7 <i>Cálculos Aplicados que Proporciona el Sentido a los Datos</i>	29
Figura 8 <i>Máximo Grado de Calor en el conjunto de Datos</i>	30
Figura 9 <i>Clasificación Espectral de la Estrella Asociado a un Color</i>	31
Figura 10 <i>Diagrama Hertzsprung-Russell</i>	34
Figura 11 <i>Comparativa de Edad Estelar del Segmento</i>	35
Figura 12 <i>Clasificación por Grupo de Color</i>	36

INFORMACIÓN GENERAL

Contextualización del tema

La Agencia Espacial Europea (ESA) es una entidad que se desarrolla en cooperación entre varias entidades gubernamentales que son parte de la Unión Europea con el objetivo que realizar la exploración espacial, misma que trata sobre la observación de fenómenos cosmológicos donde se incluye la exploración de planetas, estrellas, galaxias y otros cuerpos celestiales que forman parte del Universo (European Space Agency, 2024).

La ESA realiza ejecuta misiones con la finalidad de explorar el espacio exterior mediante el envío de sondas (nave espacial que no lleva tripulación humana) diseñadas para explorar e investigar el exterior recolectando datos de naturaleza científica, cuya información capturada por los dispositivos de medición se envía de regreso al centro de operaciones en la Tierra, y muy diferente de los satélites que orbitan nuestro planeta, están destinadas a llevar a cabo viajes más extensos con el propósito de estudiar el espacio interestelar y los diversos cuerpos celestiales (European Space Agency, 2024).

La misión Gaia soportada por la ESA, ha incursionado en el campo de la Astrofísica, que forma parte de un área más grande las ciencias denominada Astronomía, la cual se enfoca a proporcionar información del conjunto de datos que recopila registros sobre las posiciones, movimiento y propiedades físicas de las estrellas que conforman nuestra galaxia que comúnmente recibe el nombre de Vía Láctea. El programa ha abierto las puertas facilitando la comprensión estructural de la galaxia (GAIERVERSE, s.f.).

Gaia ha permitido construir un catálogo estelar de los más completos hasta ahora, disponible al público y bajo la denominación Data Release 3 (DR3), donde se recopilan datos astrométricos, fotométricos y espectroscópicos. Muchas de las mediciones representan el punto de partida para realizar análisis sobre las características y propiedades físicas de las estrellas, así de esta manera se las puede clasificar acorde a los varios tipos espectrales y poder comprender su ciclo evolutivo (Gaia Archive, s.f.). A nivel académico al contar con datos de precisión, Gaia ha generado la apertura de nuevas posibilidades de investigación, enseñanza y aprendizaje en astronomía que van más allá del impacto dentro de la comunidad científica, permitiendo que la persona común pueda acceder a la información con el afán de fomentar el papel en la ciencia y astronomía en la sociedad.

Problema de Investigación

En el ámbito astronómico, existe una brecha en la comprensión de la evolución estelar de ciertos tipos de estrellas, especialmente aquellas con características atípicas, debido a las limitaciones de los métodos tradicionales de análisis. El desconocimiento afecta a la comunidad que incluye profesionales, estudiantes y aficionados que no disponen de medios accesibles en este campo de estudio para comprender la evolución estelar. Actualmente no se identifican medios ni herramientas de fácil entendimiento que permitan incursionar en el campo astronomía, ocasionando desconocimiento y en consecuencia la falta de interés de todos aquellos quienes desean aprender de este campo del conocimiento.

Objetivo general

Realizar el proceso de ciencia de datos para realizar la clasificación y medición de edad de las estrellas mediante la información del repositorio de observaciones estelares Gaia Data Release 3 establecido como sujeto de análisis.

Objetivos específicos

- Contextualizar los fundamentos teóricos relacionados con los parámetros físicos que clasifican y señalan la vida de las estrellas.
- Analizar las variables que intervienen en la medición de clasificación y datación estelar aplicado a una fracción del repositorio Gaia Data Release 3.
- Describir el comportamiento de clasificación y edad estelar por medio de un análisis de datos.
- Valorar mediante criterio objetivo la propuesta de investigación a través de la opinión de profesionales en el campo de manejo de datos.

Vinculación con la sociedad y beneficiarios directos

La información recopilada en este proyecto tiene como objetivo contribuir al conocimiento sobre nuevas tendencias e innovaciones en el campo del análisis de datos, destacando los avances más recientes relacionados con el uso de datos masivos en los procesos de investigación.

Este aporte busca facilitar a las personas el acceso a los beneficios y avances tecnológicos en el análisis de datos, especialmente en el ámbito de la astronomía. Gracias a estas herramientas, incluso

quienes poseen conocimientos básicos pueden utilizar plataformas de procesamiento de información para realizar sus propios estudios, consultas y observaciones, sin depender exclusivamente de expertos en ciencias.

Además, el acceso a Internet permite que este conocimiento se encuentre disponible, no solamente como una contribución científica a la comunidad, sino también como una fuente de comprensión que inspire la creación de nuevos trabajos de investigación. El proyecto se presenta como recurso de consulta y referencia, al consolidar y recopilar información de diversas fuentes de datos.

CAPÍTULO I: DESCRIPCIÓN DEL PROYECTO

1.1. Contextualización general del estado del arte

La Internet se ha convertido en el avance más significativo de la edad moderna, equiparable a la invención de la rueda y en consecuencia ha facilitado el acceso al conocimiento, su única limitación podría ser la restricción del pensamiento.

Este hito viene acompañado de varias características como la capacidad de guardar la información presentada como texto, imagen, audio y video ocupando el mínimo espacio de almacenamiento en conjunto con técnicas avanzadas de procesamiento de información que han permitido el desarrollo modelos matemáticos, estadísticos o predictivos permitiendo comprender los diversos aspectos relacionados con el comportamiento sociales, económico, financiero, etc. En la nueva era de la evolución tecnológica varias instituciones han transformado sus modelos de trabajo, adaptando sus modelos de operación al reducir la intervención humana.

El tema de investigación intenta realizar el estudio de los datos recolectados con respecto a las observaciones astronómicas para detectar posibles sistemas de estelares en relación al cosmos, haciendo uso de valores numéricos recolectados como parte datos masivos, mismos que deben pasar por un proceso de depuración hasta alcanzar un formato que permita realizar un análisis adecuado sobre el que se intenta comprender ciertos fenómenos con la posibilidad de realizar la representación gráfica que forma parte de las herramientas de gestión analítica con la finalidad de incorporar información visual para su fácil interpretación.

El estudio forma parte de comprender la estructura evolutiva de las estrellas haciendo uso de técnicas de procesamiento sencillas y no basado en grandes infraestructuras de hardware y software, donde se intenta obtener una clasificación con base en las características y propiedades que las definen, siendo un buen ejemplo de este análisis el tipo espectral y luminosidad con el propósito de categorizar lo sistemas estelares, además de la medición de la edad de una estrella de tal manera que permita entender la varias etapas de su formación.

Desde los diagramas simples hasta alcanzar el hito de las misiones espaciales actuales, los procesos de medición han evolucionado de manera muy convincente mediante la base teórica que ayuda a clasificar la evolución y por otra parte las misiones tales como Gaia han generado una revolución mediante la obtención de información por medio de la obtención de datos altamente precisa.

En tal sentido cuando hablamos del proyecto constituido como Gaia, mismo que ha llegado a construir los catálogos que contienen toda la información relacionada con la posición y movimiento de las estrellas se ha logrado posicionar como unos de los recursos más invaluable dentro del campo de las ciencias. La información recopilada permite desarrollar modelos de análisis estrellas con la finalidad de explorar el entorno de la galaxia.

La clasificación basada en el espectro representa un punto clave cuando se instituyó el sistema de tipo espectral O, B, A, F, G, K, M que agrupa a los sistemas basados en sus espectros, absorción y emisión implícitamente relacionados con la temperatura, química y gravedad. La ciencia que se encarga de analizar el espectro de luz denominada Fotometría, permite realizar la evaluación de luz que emiten los cuerpos celestes en sus longitudes de onda, lo que las personas perciben como un color, así de esta manera se ha conseguido optimizar la recopilación de información relacionada a la evolución de una estrella (womenshistory, s.f.).

Las nuevas mejoras para obtener información vienen con nuevos retos, como la variabilidad de los datos y limitaciones de los instrumentos conduce a generar incertidumbres de las mediciones, también hay que agregar la presencia de polvo y otros fenómenos físicos que se indiquen en las observaciones, alterando de alguna forma los datos al momento de catalogar un sistema. Sin embargo, las nuevas tecnologías abren la posibilidad de capturar datos detallados, de esa manera se han extendido las posibilidades de estudiar el cosmos en regiones densamente pobladas de estrellas y en galaxias muy lejanas.

Actualmente las ciencias están atravesando un buen momento debido al avance de herramientas informáticas y algoritmos de inteligencia artificial y aprendizaje de máquina que han permitido manejar grandes volúmenes de datos para extraer patrones y conocimientos muy difíciles imposibles de identificar manualmente. Dicha integración ha facilitado la apertura de nuevos caminos para el análisis de datos, como, por ejemplo, la habilidad de reconocer características específicas que facilitan clasificaciones.

La disponibilidad de un amplio conjunto de herramientas de trabajo para realizar prácticas de análisis permite que la persona promedio se encuentre en la capacidad de crear sus propias implementaciones de una forma insospechada que integran sus propios ecosistemas de procesamiento y análisis de datos astronómicos.

Las librerías permiten manejar datos espectroscópicos y fotométricos para realizar análisis estadísticos, por ejemplo, utilizando Pandas, las personas pueden limpiar y estructurar catálogos a partir de las observaciones espaciales como Gaia y posteriormente podría aplicarse un algoritmo de interés para clasificar estrellas por tipo espectral o luminosidad.

Los instrumentos se respaldan en repositorios y bases de datos abiertas con acceso a espectros, imágenes y catálogos estelares. Aunque las tecnologías están transformando la astronomía, también nace la necesidad de estandarizar formatos de datos, mejorar la interoperabilidad entre plataformas y garantizar la calidad de los datos son algunas de las principales áreas de enfoque. Sin embargo, la colaboración internacional y el uso de datos abiertos están allanando el camino para un progreso sostenido.

El lenguaje de software Python es uno de los más populares para desarrollar el análisis de datos debido a sus características como la facilidad de aprendizaje, simplicidad, flexibilidad y la disponibilidad de un conjunto de librerías en varias disciplinas científicas que integran el procesamiento de volúmenes de información, análisis estadístico, visualización avanzada, etc. Sus capacidades permiten manipular y analizar datos tabulares y series temporales con facilidad perfectos para organizar información de catálogos astronómicos, y así de esta manera concretar la meta de desarrollar un proceso de análisis práctico que puede ser replicado en otros ámbitos.

1.2. Proceso investigativo metodológico

El proyecto tiene como objetivo realizar una investigación basada en la observación, recolección y análisis de información soportada por el método cualitativo (Sampieri, 2014).

Además, la investigación corresponde al tipo descriptivo debido a que se intenta caracterizar y clasificar las estrellas según sus parámetros físicos y también correlacional, porque analiza la relación entre variables como la edad de las estrellas, la clasificación estelar. Se aplica un enfoque exploratorio, ya que se aplican técnicas de ciencia de datos orientado a la astronomía.

El método principal es el análisis de datos aplicado a la clasificación y datación estelar, complementado con aprendizaje de máquina orientado a la clasificación, soportados por la herramienta Python y sus bibliotecas especializadas en procesamiento de volúmenes de datos, visualizaciones y construcción de modelos predictivos.

El proyecto parte de la exploración de posibles orígenes de información del tema sujeto de investigación, es importante mencionar la existencia de mucha información disponible donde es importante reconocer la confiabilidad de la fuente, en este caso la fuente de información proveniente de la Agencia Espacial Europea garantiza la confiabilidad de la información, posteriormente se intenta desechar datos no esenciales y de escaso valor. La extracción de información relevante es fundamental.

La información recopilada en formato de valores separados por coma (CSV) se clasifica y categoriza según su relevancia, en esta etapa se elimina información con campos vacíos (valores nulos) y mediante criterios de selección se establecen algunos campos de interés, mismos que se utilizan como base para realizar otros cálculos tales como espectro, luminosidad, longevidad y a continuación obtener información estadística que permita determinar y clasificar los grupos de estrellas, además de otra información relevante. También se puede incluir la presentación de diagramas de los cuales se espera faciliten la comprensión del estudio.

Por lo tanto, con este desarrollo se estima aplicar un proceso que comienza desde la obtención de datos crudos hasta alcanzar una visualización que permita comprender el verdadero poder de los datos.

1.3. Análisis de resultados

El procesamiento se realiza en un entorno de nube Google (Colaboratory) que incorpora características de ejecución de análisis de Datos Masivos (Big Data) con capacidades de manejar volúmenes de datos, almacenamiento y acceso a herramientas de análisis, además de su óptima integración con otras herramientas de datos tales como Tableau y Kaggle. Su operación no requiere inversión significativa, ni configuraciones complejas, por lo tanto, resulta una opción eficiente en términos de costo-beneficio. Además, al aprovechar la compatibilidad del ecosistema, el análisis puede publicarse en internet como un medio para compartir el conocimiento hacia la comunidad y colaboración al mundo de la ciencia de datos.

El presente trabajo se desarrolla como un prototipo, es decir, una prueba de concepto que valida el enfoque y las metodologías empleadas. Tomando en consideración lo flexible y adaptable, se permite su extensión a más conjuntos de datos del repositorio con características similares. Se utiliza un conjunto de datos que proviene del catálogo astronómico Gaia Data Release 3 que recopila

información detallada sobre varias características que definen varios cuerpos celestes como las estrellas dentro de nuestra galaxia.

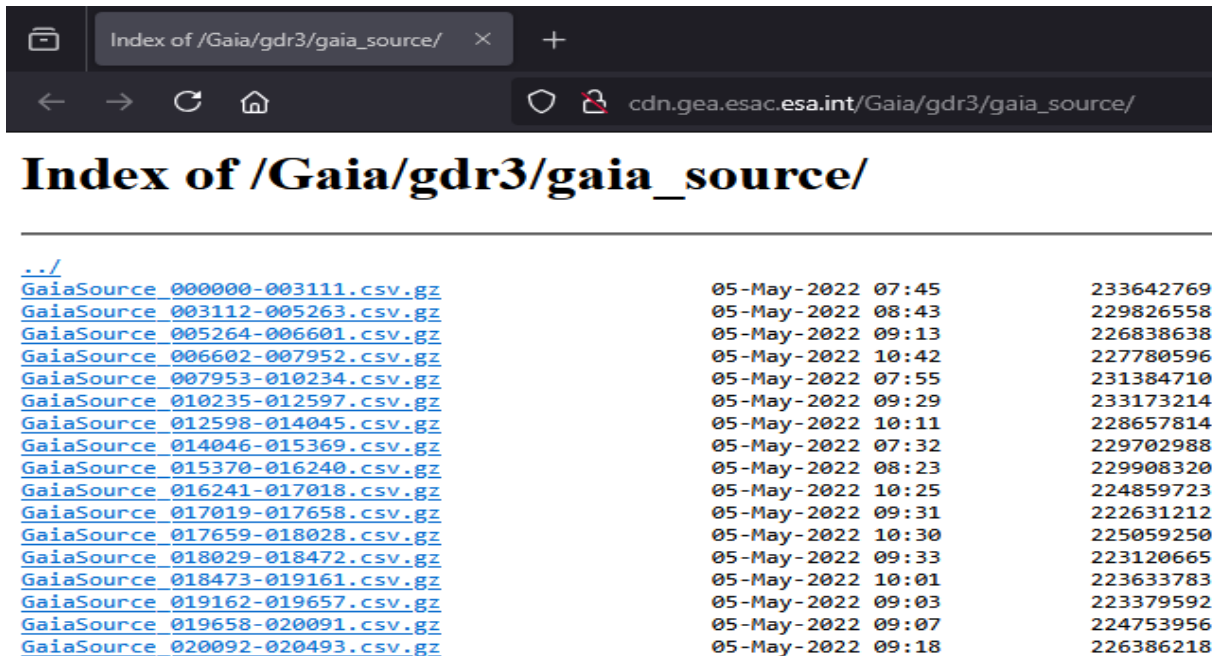
Se utilizan únicamente dos sub-catálogos de todo el conjunto de datos que representan una pequeña porción del cielo observable que se encuentran conformados del formato de archivos separados por comas (CSV) que comprende la región de interés definida, donde se incluye la información necesaria para cumplir con los objetivos del estudio. Procesar el conjunto completo de datos podría requerir un alto costo computacional sin aportar beneficios significativos al proyecto en cuestión. Además, dichos catálogos contienen los datos clave para la clasificación y datación de estrellas en la región estudiada y responder a criterios de relevancia, eficiencia y alcance del proyecto. A continuación, se detallan los archivos:

- GaiaSource_000000-003111.csv.gz
- GaiaSource_003112-005263.csv.gz

En la Figura 1 se muestra el sitio web que almacena el repositorio de la Misión Gaia, fuente del origen de datos indicado.

Figura 1

Repositorio de Gaia DR3



File Name	Size	Modified
../		
GaiaSource_000000-003111.csv.gz	233642769	05-May-2022 07:45
GaiaSource_003112-005263.csv.gz	229826558	05-May-2022 08:43
GaiaSource_005264-006601.csv.gz	226838638	05-May-2022 09:13
GaiaSource_006602-007952.csv.gz	227780596	05-May-2022 10:42
GaiaSource_007953-010234.csv.gz	231384710	05-May-2022 07:55
GaiaSource_010235-012597.csv.gz	233173214	05-May-2022 09:29
GaiaSource_012598-014045.csv.gz	228657814	05-May-2022 10:11
GaiaSource_014046-015369.csv.gz	229702988	05-May-2022 07:32
GaiaSource_015370-016240.csv.gz	229908320	05-May-2022 08:23
GaiaSource_016241-017018.csv.gz	224859723	05-May-2022 10:25
GaiaSource_017019-017658.csv.gz	222631212	05-May-2022 09:31
GaiaSource_017659-018028.csv.gz	225059250	05-May-2022 10:30
GaiaSource_018029-018472.csv.gz	223120665	05-May-2022 09:33
GaiaSource_018473-019161.csv.gz	223633783	05-May-2022 10:01
GaiaSource_019162-019657.csv.gz	223379592	05-May-2022 09:03
GaiaSource_019658-020091.csv.gz	224753956	05-May-2022 09:07
GaiaSource_020092-020493.csv.gz	226386218	05-May-2022 09:18

Nota. La Figura 1 describe el repositorio Gaia DR3. Fuente Agencia Espacial Europea.

CAPÍTULO II: PROPUESTA

2.1 Fundamentos teóricos aplicados

Big Data – Ciencia de Datos

El concepto de manejar datos masivos, comúnmente denominado **Big Data**, actualmente se ha establecido como la manera de manejar y analizar grandes volúmenes de información que resultan demasiado complejos y extensos para ser procesados y analizados a través de métodos y procedimientos comunes o tradicionales. En este sentido se introducen nuevas formas de procesamiento en conjunto con la **Ciencia de los Datos** que combina algunas disciplinas como estadística, análisis y aprendizaje, afín de obtener conocimiento de dichos conjuntos de datos. En el ámbito de la astronomía están colaborando a reducir los tiempos de respuesta en la obtención de resultados provenientes en este caso de los datos de telemetría generados por las observaciones y mediciones astronómicas (OpenAI, 2024).

Repositorio de Observaciones Gaia

Gaia es la misión patrocinada por la ESA y su objetivo consiste en mapear los sistemas estelares en la galaxia denomina Vía Láctea (término que se debe a que se parece a un río lechoso mirado desde nuestro planeta). Además, el repositorio no solamente contiene conjuntos de datos, también dispone de guías, manuales, indicaciones de cómo manejar la información, incluyendo acceso a otras fuentes de datos de otros proyectos relacionados.

La información recopilada permite generar un catálogo estelar que incluyen muchos detalles que incluyen posiciones, movimientos, y características que se pueden procesar y analizar de manera eficiente, facilitando la comprensión de la estructura y evolución de la Vía Láctea. Dicha evolución representa un proceso que cambia a lo largo de la vida de la estrella, desde el momento de su formación hasta el ciclo que marca su muerte (Gaia Archive, s.f.).

Herramientas y Aplicaciones

Entre las herramientas de análisis se dispone de Visual Studio Code, un editor de código flexible y extensible desarrollado por Microsoft que permite añadir extensiones que potencian su uso y proporcionan soporte para Python, Jupyter, etc. (Microsoft, s.f.). Jupyter Notebooks es la herramienta que admite código en tiempo real como ecuaciones, visualizaciones y texto explicativo para ejecutar código en celdas independientes, facilitando la prueba y depuración del código (Proyect Jupyter, s.f.).

Las técnicas empleadas en el procesamiento de la información se basan en la adquisición y transformación de un Dataset (conjunto de datos), asistido por un lenguaje de programación muy utilizado en la actualidad como Python, el cual corresponde a un lenguaje de programación de alto nivel, interpretado y muy fácil de aprender, conocido por su sintaxis clara y su amplio soporte para diversas aplicaciones que dispone de una vasta cantidad de bibliotecas diseñadas para manejo de Big Data, de esta manera se propone realizar un procesamiento aceptable para la manipulación de información (Datacamp, s.f.).

Librerías como pandas, matplotlib entre otras, facilitan las transformaciones y análisis que pueden terminar en gráficos para su mejor entendimiento, además de algún modelo de aprendizaje. Los cálculos realizados se basan en cálculo empírico cuyo enfoque trata de realizar operaciones sustentadas en métodos o técnicas que se articulan en la observación y experimentación en lugar de teorías estrictamente matemáticas o modelos abstractos, y en general dicho cálculo se utiliza para obtener resultados prácticos y aproximados tomando como punto de partida los datos experimentales y observaciones directas (Interactive Chaos, s.f.).

Posteriormente toda la información obtenida finaliza en una presentación por medio de la herramienta Tableau que consiste de una aplicación enfocada a la visualización de datos que transforma la información en visualizaciones dinámicas, facilita el análisis por medio de tableros interactivos lo que permite identificar e interpretar la información. Se integra con varios orígenes de datos de donde se puede mencionar bases de datos y hojas de cálculo. Incluye las historias como el medio de organizar las visualizaciones de forma secuencial para narrar un análisis de manera ordenada, clara y concisa (Tableau, s.f.).

Técnicas

Análisis de Datos

El proceso se divide en fases en el instante que se forman a partir de nubes de gas y polvo y que por acción de la gravedad colapsan y forman una estrella denominada proto-estrella, posteriormente en la fase de secuencia principal éstas pasan la mayor parte de su vida fusionando hidrógeno en helio, hasta que finalmente éste se agota para funcionar helio en componentes más pesados, hasta llegar al punto donde se alcanza el hierro, la estrella se vuelve inestable y finalmente termina como lo que se denominan enanas blancas, estrellas de neutrones o agujeros negros, dependiendo de sus características (Futurism, s.f.).

Se pueden analizar rasgos tales como a temperatura, magnitud, distancia, composición y muchas otras más, mismas que se pueden obtener del repositorio público Gaia, de tal manera que mediante técnicas de análisis que son fundamentales se puede comprender la evolución y clasificación de las estrellas y elaborar correlaciones por medio de metodologías de análisis de datos astronómicos (Cosmos, 2024).

Además, se puede conducir un análisis que clasifique las estrellas según sus características observacionales entre los que se indican a continuación (European Space Agency, s.f.):

- Temperatura (teff_gsppho): temperatura de la estrella definido por el espectro de radiación que es usado para clasificar los tipos espectrales basados en su calor. El espectro está codificado en letras del alfabeto con un rango que oscila desde los más caliente (O) hasta lo más frío (M). La Tabla 1 muestra la clasificación del espectro (Gaia_source, 2023).

Tabla 1

Resumen de Clases Espectrales

Clase Espectral	Temperatura Efectiva (K)	Color
O	28,000 - 50,000	Azul
B	10,000 - 28,000	Azul-blanco
A	7,500 - 10,000	Blanco
F	6,000 - 7,500	Blanco-amarillo
G	4,900 - 6,000	Amarillo
K	3,500 - 4,900	Naranja
M	2,000 - 3,500	Rojo

Nota. Clases Espectrales, 2024. Fuente Australia Telescope National Facility.

- Gravedad (logg_gspphot): es la aceleración debida a la gravedad en la superficie de la estrella, medida en cm/s^2 . La gravedad superficial ayuda a distinguir entre estrellas de secuencia

principal, gigantes y supergigantes. Las estrellas de secuencia principal tienen una gravedad más alta que las gigantes y supergigantes (Gaia_source, 2023).

- Metalicidad (mh_gspphot): indica la abundancia de elementos más pesados que el helio en una estrella (Gaia_source, 2023).
- Magnitud (phot_g_mean_mag): es una medida del brillo de una estrella. El brillo absoluto permite obtener el tamaño y luminosidad de la estrella (Gaia_source, 2023).
- Distancia (distance_gspphot): La distancia a las estrellas es muy grande y en este sentido se utilizan unidades de medida tales como (Gaia_source, 2023):
 - Unidad Astronómica (AU): distancia promedio entre la Tierra y el Sol, un aproximado de 150 millones de kilómetros, utilizado en la medición de distancias en el Sistema Solar.
 - Parsec (pc): distancia a la estrella en ángulo de paralaje, equivalente a 3.26 años luz, utilizado en la medición de trayectos a estrellas y galaxias.
 - Año Luz (ly): distancia recorrida por la luz dentro del año terrestre, un aproximado de 9.46 billones de kilómetros, utilizado en la medición a objetos dentro y fuera de la galaxia.

La distancia a la estrella desde la Tierra, permite calcular el brillo absoluto y estimar el tipo de estrella al comparar el brillo observado con el esperado. La distancia mínima posible es 1 pc y la máxima es 100 kpc(kilopcs).

- Sistema Múltiple (non_single_star): indicador que categoriza estrellas como un sistema de estrella única o un sistema de algunas estrellas que se orbitan entre ellas, lo cual evita conclusiones erróneas en el estudio (Gaia_source, 2023).
- Luminosidad: cantidad total de energía que la estrella emite en forma de luz. Existen algunos puntos de vista, con respecto a la medición, por ejemplo, su unidad de medida puede mostrarse en vatios y también se puede comparar con base a la luminosidad del Sol (unidad L_{\odot}). El cálculo para obtener luminosidad se desprende de las fórmulas de magnitud absoluta y luminosidad (eBlanc, 2010):

Fórmula 1 para realizar el cálculo de Magnitud Absoluta

$$M = m - 5\log_{10}(d) + 5$$

Donde:

m es la magnitud aparente

d es la distancia en parsecs

Fórmula 2 para realizar el cálculo Luminosidad

$$L/L_{\odot} = 10^{0.4 \cdot (M_{\odot} - M)}$$

$$L/L_{\odot} = 10^{0.4 \cdot (4.83 - M)}$$

Donde:

L es la luminosidad de la estrella

L_{\odot} es la luminosidad del sol (1×10^{26} watts)

M_{\odot} es la magnitud absoluta del Sol (4.83 en la banda G).

M es la magnitud absoluta calculada.

Si $M < M_{\odot}$, la estrella es más luminosa que el Sol.

Si $M > M_{\odot}$, la estrella es menos luminosa que el Sol.

Si $M = M_{\odot}$, la estrella tiene la misma luminosidad que el Sol.

Fórmula 3 para realizar el cálculo de la Edad Estelar, mediante fórmula empírica simplificada que establece un estimado (OpenAI, 2024).

$$\text{Edad} \approx 1 + 10^{(mh - \gamma / 10 \cdot \text{logg})}$$

Donde:

logg es la gravedad superficial en la escala logarítmica (logg_gspphot)

mh es la metalicidad ([Fe/H] o mh_gspphot)

γ es la constante (-0.5) empírica ajustada según modelos estelares.

Representación Gráfica

En el contexto de visualización se intenta utilizar diagramas tales como de Dispersión, Barras, Histogramas para visualizar la relación entre las variables a fin de clasificar las estrellas (The Data Visualisation Project, 2024). También se emplea el conocido diagrama de Hertzsprung-Russell para realizar la evaluación de las variables involucradas (eBlanc, 2010).

2.2 Descripción de la Propuesta

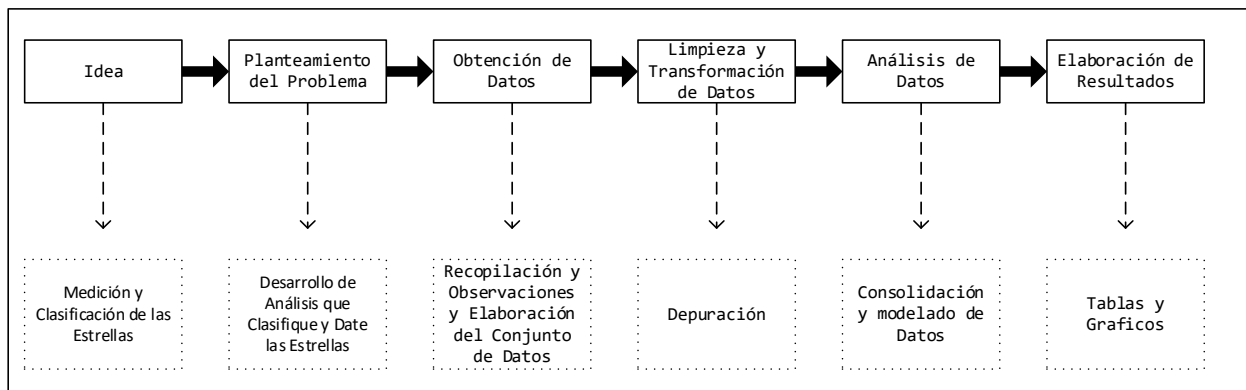
El presente trabajo que trata sobre la estructura de la galaxia, ha permitido recopilar información sobre características de definen las estrellas mediante el uso de técnicas de datos masivos, donde se utilizan algunos utilitarios para explorar y analizar datos, el enfoque está orientado a describir comportamientos clave sobre la evolución estelar, además de aplicar el manejo de muchos registros para extraer conclusiones significativas que contribuyan por medio de los datos.

a) Estructura general

El proceso de análisis de datos tiene una estructura que articula las siguientes estrategias. La Figura 2 describe el proceso del ciclo del proyecto.

Figura 2

Diagrama del Ciclo de Investigación



Nota. En la Figura 2 se presentan las fases del ciclo del proyecto que permitieron conseguir el objetivo.

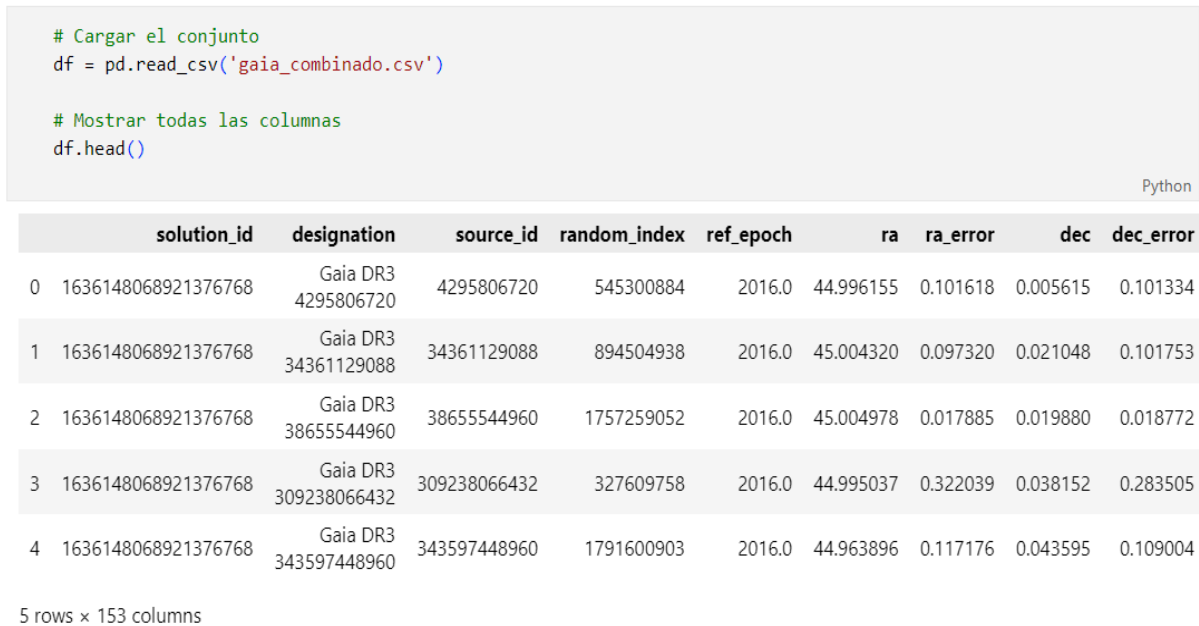
El procesamiento de datos de la Misión Gaia se encuentra dentro de los preceptos que abarcan el **Big Data** por cuanto el volumen de la información incluye millones de registros procedentes de las observaciones astronómicas y constituye un tamaño que oscila en el orden de los gigabytes. La variedad hace referencia a una gran cantidad de variables por cuanto los archivos contienen una estructura de 153 columnas y n-filas que representan una observación estelar, a pesar que los datos no se representan en tiempo real se pueden utilizar estrategias de administración de rendimiento y

memoria en caso de ser requerido. Toda la información procede de la misión Gaia, aunque pueden existir errores que necesiten una validación. La extracción de información significativa permite ayudar a realizar la clasificación y evolución de estrellas (Netacad, s.f.).

Durante el proceso de extracción de datos primero se obtienen los datos de interés, para posteriormente unificar estos catálogos en un único archivo. Posteriormente el nuevo archivo se carga a la aplicación, el comando de Pandas “df.head()” muestra las primeras filas del conjunto para observar su estructura y validar la importación de los datos, en este punto solo se dispone de datos de naturaleza numérica. En la Figura 3 se presenta la muestra cargada en la herramienta de Visual Code mediante un libro de Jupyter comúnmente conocido como notebook. Dicho proceso se ampara bajo enunciados expuestos dentro de la fase de recopilación (Hevodata, s.f.).

Figura 3

Carga del Conjunto de Datos



Nota. La Figura 3 describe la carga de datos dentro de la aplicación Visual Studio Code.

Ahora que se dispone de todos los datos se realiza el proceso de transformación con la finalidad de establecer el formato requerido, eliminar inconsistencias, valores nulos para garantizar su calidad (Hevodata, s.f.).

El problema que se tiene con los datos al momento, radica en que muchas observaciones tienen campos vacíos. Existen algunas estrategias que permiten afrontar el tratamiento de nulos, en la Tabla 2 se presenta el detalle de algunas estrategias.

Tabla 2

Tratamiento de Valores Nulos

Estrategia	Descripción	Ventajas	Desventajas
Eliminar filas con campos vacíos	Elimina filas con datos faltantes	Fácil de aplicar, no introduce sesgos	Se pierde información importante
Rellenar con un valor específico	Sustituye valores faltantes con la media, mediana o valores fijos	Mantiene el tamaño del conjunto de datos	Puede introducir sesgos si no es representativo
Reemplazo con algoritmos	Predice valores faltantes usando modelos estadísticos	Más preciso, ideal para valores críticos	Requiere más trabajo y modelos complejos
Mantener campos vacíos	Deja los valores vacíos si son relevantes para el análisis	Útil si la ausencia de datos tiene significado	Complica el análisis si no se maneja adecuadamente

Nota. La Tabla 3 las estrategias de manejo de valores faltantes. Fuente Datasource.

De la tabla descrita con anterioridad el trabajo con datos espaciales, rellenar esos valores con estimaciones o mantener los vacíos no es útil, ya que puede introducir errores o sesgos en el análisis, además de existir una cantidad significativa de columnas sin valores en cada registro. En este caso cuando se manejan datos espaciales, eliminar las filas con valores nulos resulta ser la opción más conveniente al asegurar que el conjunto de datos sea más preciso y consistente y permita mantener la calidad del análisis evitando resultados distorsionados, donde cada valor tiene un impacto significativo en los resultados. En la Figura 4 se presenta una muestra del proceso de limpieza de los datos.

Figura 4

Proceso de Limpieza

```
# Eliminar entradas con valores nulos
df2 = filtrar.dropna()

# Comprobar entradas con valores nulos
df2.isna().sum()
```

Python

```
dataset_name          0
solution_id           0
source_id             0
teff_gspphot         0
logg_gspphot         0
mh_gspphot           0
phot_g_mean_mag      0
distance_gspphot     0
non_single_star      0
classprob_dsc_combmod_star  0
classprob_dsc_combmod_quasar 0
classprob_dsc_combmod_galaxy 0
dtype: int64
```

Nota. La Figura 4 describe la limpieza de los datos.

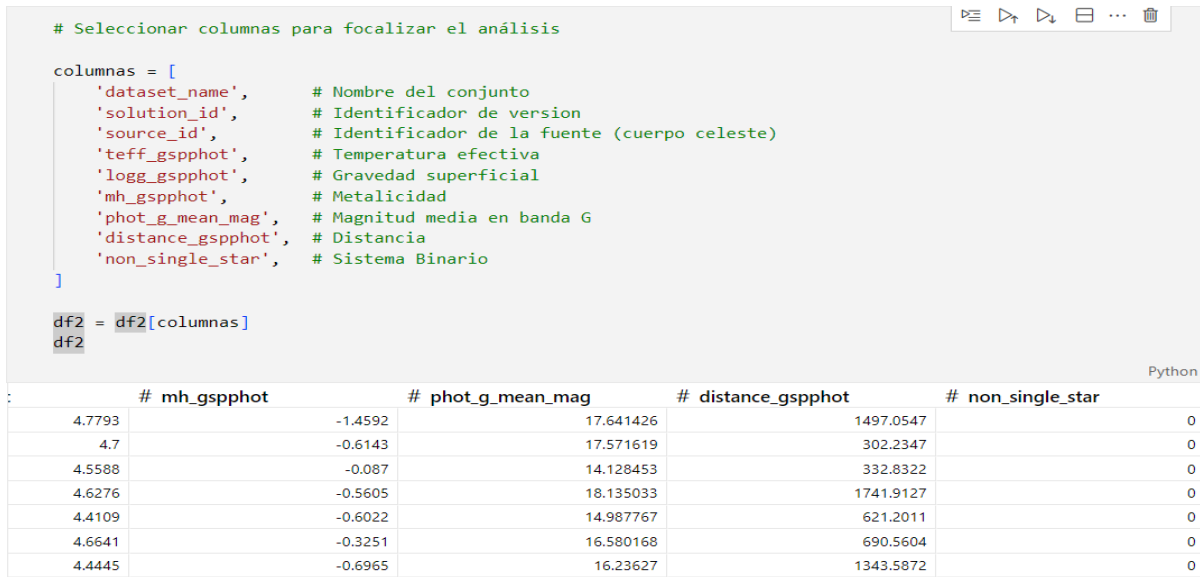
Tomando en cuenta que el repositorio incluye información de muchos cuerpos celestes, se aplican los correspondientes filtros para únicamente seleccionar estrellas. El conjunto incluye tres campos importantes relacionados a la probabilidad de ser una estrella, fuente de radio casi estrellar (quasar) o un agujero negro, mientras en factor de 1 (100%) se acerque a este valor, más probable en recaer en cualquiera de estas categorías será la posibilidad (Gaia_source, 2023).

- classprob_dsc_combmod_star → % de Estrella
- classprob_dsc_combmod_quasar → % de Quasar
- classprob_dsc_combmod_galaxy → % de Galaxia

En ese caso el criterio que se establece corresponde que un factor de 0.9 a 1 pertenece a una fuente de luz que se acerca a representar una estrella. Cuando se ha alcanzado la coherencia y normalizaron de los formatos de los datos para su correcta integración con herramientas de análisis, se deben aplicar un sentido a todas esas observaciones de tipo numérico, se intenta categorizar los datos como base a las indicaciones que se encuentran disponibles en la enciclopedia en línea de la documentación Gaia (gaia data release 3 documentation, s.f.). Además, se seleccionan únicamente los campos de interés para este presente estudio. La Figura 5 presenta la selección de los campos.

Figura 5

Selección de Columnas



Nota. La Figura 5 describe la selección de los campos de interés.

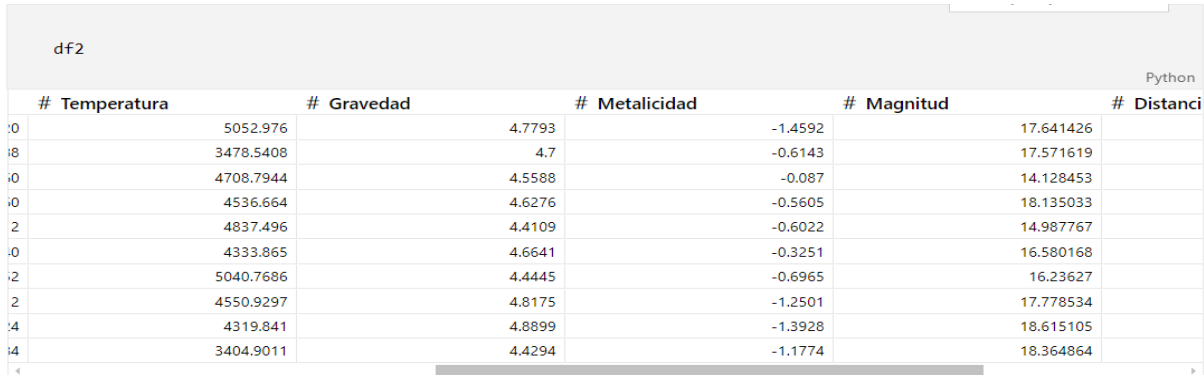
En esta etapa se asignan nombres más intuitivos en concordancia con el fenómeno sujeto de estudio, el detalle de los nombres se define a continuación:

- Nombre del conjunto
- Identificador de versión
- Identificador de la fuente
- Temperatura efectiva
- Gravedad superficial
- Metalicidad
- Magnitud media en banda G
- Distancia
- Sistema Binario

La Figura 6 muestra los nombres de las columnas que son reemplazados por textos más simples de entender.

Figura 6

Actualización de Nombres de Columnas



	# Temperatura	# Gravedad	# Metalicidad	# Magnitud	# Distancia
0	5052.976	4.7793	-1.4592	17.641426	
1	3478.5408	4.7	-0.6143	17.571619	
2	4708.7944	4.5588	-0.087	14.128453	
3	4536.664	4.6276	-0.5605	18.135033	
4	4837.496	4.4109	-0.6022	14.987767	
5	4333.865	4.6641	-0.3251	16.580168	
6	5040.7686	4.4445	-0.6965	16.23627	
7	4550.9297	4.8175	-1.2501	17.778534	
8	4319.841	4.8899	-1.3928	18.615105	
9	3404.9011	4.4294	-1.1774	18.364864	

Nota. La Figura 6 describe la actualización de los nombres de las columnas.

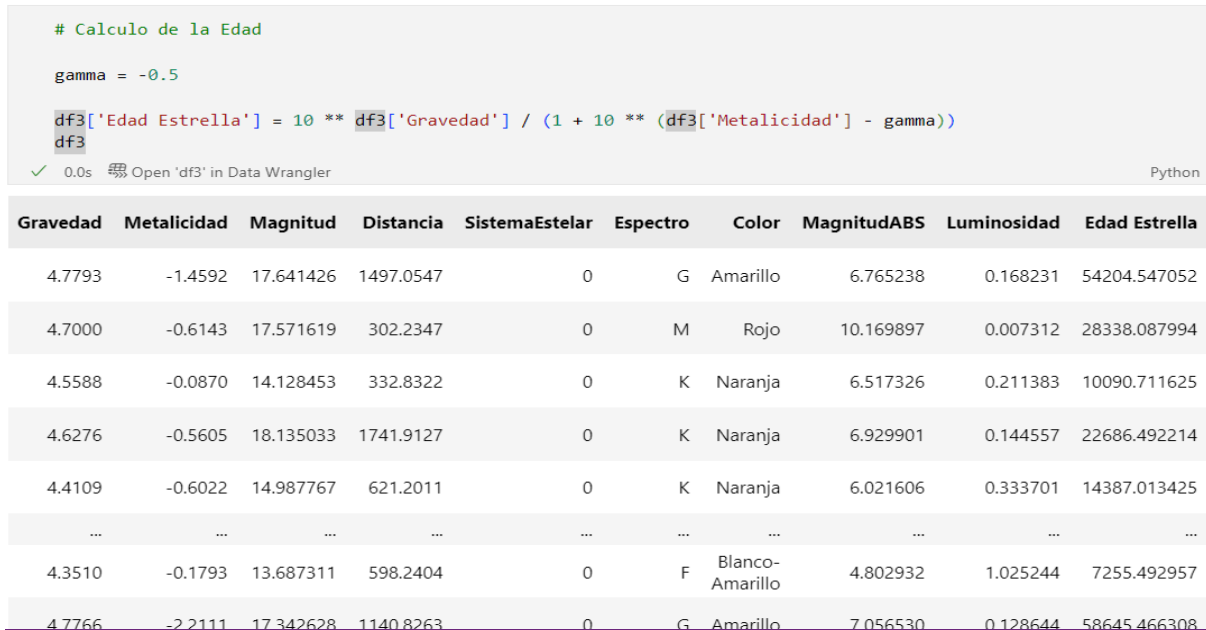
En esta fase que se aplica al conjunto de datos, se realiza una serie de operaciones matemáticas que van a dar sentido a los datos observacionales, se define la clasificación de las estrellas con base al espectro de luz visible que podría traducirse como el color del cuerpo celeste. Además, se realizan los cálculos para determinar la luminosidad de las estrellas tomando como referencia el resplandor de nuestra estrella "El sol". Por último, se determina el cálculo aproximado de la edad de una estrella, así de esta manera se constituyen todos los parámetros requeridos para llevar a cabo el análisis.

El conjunto de datos resultante tiene el valor científico que va a permitir estudiar la relación entre la temperatura y luminosidad, analizar la distribución de estrellas por tipo espectral, además, de estudiar su comportamiento conforme a la edad.

En la Figura 7 se indica una muestra de los cálculos ejecutados para completar esta etapa del análisis.

Figura 7

Cálculos Aplicados que Proporciona el Sentido a los Datos



Nota. La Figura 7 describe la parte final de los cálculos aplicados a los datos para interpretar los datos.

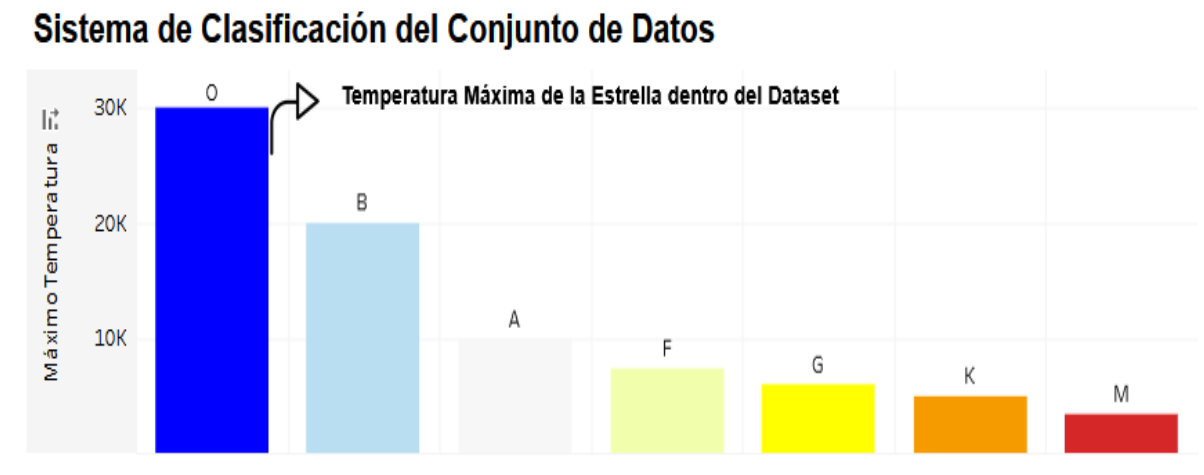
En la siguiente fase relacionado a la carga se enfoca en transformar y estructurar los datos antes de su visualización, sin necesidad de una base de datos. El archivo final generado se encuentra optimizado para el análisis haciendo uso de herramientas de visualización como Tableau. El proceso realizado con anterioridad asegura que los datos sean consistentes y listos para su uso, **en** este sentido los datos se encuentran en un formato adecuado para exploración y análisis visual (Hevodata, s.f.).

En la etapa de visualización, la información procesada se transmuta en gráficos y tablas que permitan facilitar el análisis. Tableau permite observar las relaciones en los datos de forma clara por medio de visualizaciones interactivas para un mejor entendimiento de la información sin necesidad de analizar números en forma directa. Los paneles interactivos permiten explorar la información de forma sencilla. Además, proporciona un ambiente perfecto que ofrece integración fácil con fuentes de datos compatibles y permite crear tableros accesibles y eficientes para proyectos con plazos cortos. Dispone de publicación web sin necesidad de infraestructura adicional, aprovechando que se dispone de licenciamiento de mano de la Universidad.

Con respecto al análisis visual el sistema Morgan-Keenan es el método utilizado para efectuar la clasificación. La Figura 8 establece que los datos obtenidos en este segmento del espacio incluyen estrellas que se ajustan en todos los espectros, las estrellas del tipo O tienen una temperatura que alcanza los 30.000 mil grados kelvin, siendo el grupo de estrellas con el mayor grado de calor dentro de este segmento.

Figura 8

Máximo Grado de Calor en el conjunto de Datos

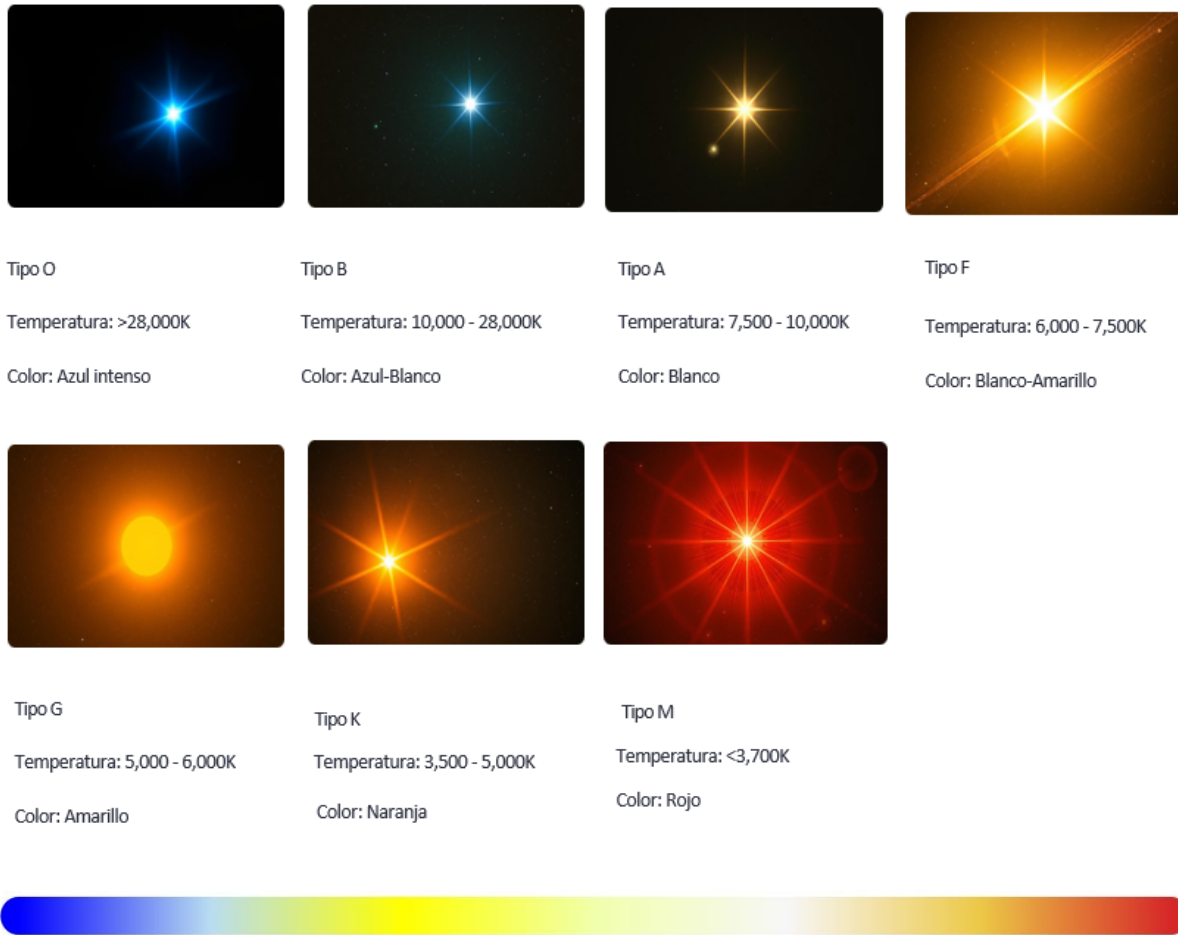


Nota. La Figura 8 describe la temperatura máxima de las estrellas dentro del conjunto de datos.

En la Figura 9 se puede observar un espectro de color visible asociado a la temperatura de la estrella que facilita la comprensión, aquí se puede observar que a mayor calor irradiado por el cuerpo más azulado será el color de la estrella, y en contraposición mientras menor sea el calor del cuerpo más fría será, esto dentro de los términos de las altas temperaturas. Existen otros cuerpos celestiales con menores escalas de calor, pero estos no se encuentran formando parte del estudio.

Figura 9

Clasificación Espectral de la Estrella Asociado a un Color



Nota. La Figura 9 describe el color de la estrella en base a su temperatura, junto con la escala de calor. Autoría Propia generado con IA.

La luminosidad de cada estrella especifica la cantidad de energía que se emite en forma de luz, y está asociado con la temperatura y tamaño del objeto. Significa que los cuerpos más calientes emiten una mayor cantidad de energía y en consecuencia son más brillantes con respecto a aquellas más frías.

La relación del factor de luz, luminosidad y temperatura de las estrellas, se toma como referencia nuestra estrella madre (Sol) que parte de una luminosidad estándar de $1 L_{\odot}$ (luminosidad solar) con

una temperatura de alrededor 5,700 K. Sin embargo, a medida que la luminosidad aumenta en factores de 10, 100, n-veces en función del Sol, la temperatura del cuerpo se incrementa.

En relación con la edad de las estrellas, el grupo de estrellas en sus fases más tempranas son más calientes y brillantes con un tiempo vida más corto, este es el caso de las estrellas de color más azuladas que se encuentran bajo los espectros O y B debido al consumo rápido de su combustible nuclear. Por otro lado, las estrellas más frías y pequeñas de colores con más tendencia al rojo, como el tipo espectral K y M tienen un tiempo de vida prolongado debido a un consumo bajo del combustible nuclear.

Cuando las estrellas envejecen, la tendencia es que la temperatura se reduzca, al igual que su luminosidad. Las características de temperatura, luminosidad y edad de una estrella están estrechamente relacionadas, dicha relación puede observarse mediante un diagrama de dispersión en forma de Diagrama Hertzsprung-Russell (H-R) para visualizar la relación entre las características mencionadas.

En el Eje X se encuentran los valores de Temperatura expresado en grados kelvin ($^{\circ}\text{K}$) en forma de escala inversa, donde el grado de calor disminuye de izquierda a derecha. El Eje Y de Luminosidad expresado en términos de la luminosidad del sol (L_{\odot}) expone en escala logarítmica que tan luminoso es el cuerpo con respecto con nuestro Sol. Cada punto corresponde a un cuerpo celeste dentro de un grupo específico asociado a su color o espectro.

En la Figura 10 el diagrama describe que la mayoría de estrellas ocupa el área diagonal partiendo desde la porción superior izquierda de altas temperatura y luminosidad hasta llegar al segmento inferior derecha de bajas temperaturas y luminosidad, esto significa que las estrellas se encuentran en su fase de secuencia principal porque están utilizando la fusión de hidrógeno en sus núcleos activos. La región superior por encima de la línea imaginaria de la secuencia principal contiene estrellas la evolucionadas de alta luminosidad, pero con índices de calor más bajos bajas, en contraste por debajo de la línea se ubican estrellas de menor tamaño que no poseen mucho combustible nuclear de alta temperatura, pero con reducida luminosidad.

En la Figura 11 el diagrama ayuda a identificar claramente que las estrellas que tienen más al color rojo tienen un mayor tiempo de vida, en contraste con las que tienen más azul, esto se debe a que las estrellas de mayor temperatura consumen más rápido su combustible principal de hidrógeno. Las estrellas **azules** son las más **jóvenes** y las **rojas y naranjas** son las más **antiguas**, alcanzando edades

muy elevadas, en este caso tenemos algunas que se acercan a los 200 mil millones de años. Las estrellas entre de la categorías **blancas - amarillo** tienen periodos de vida intermedias.

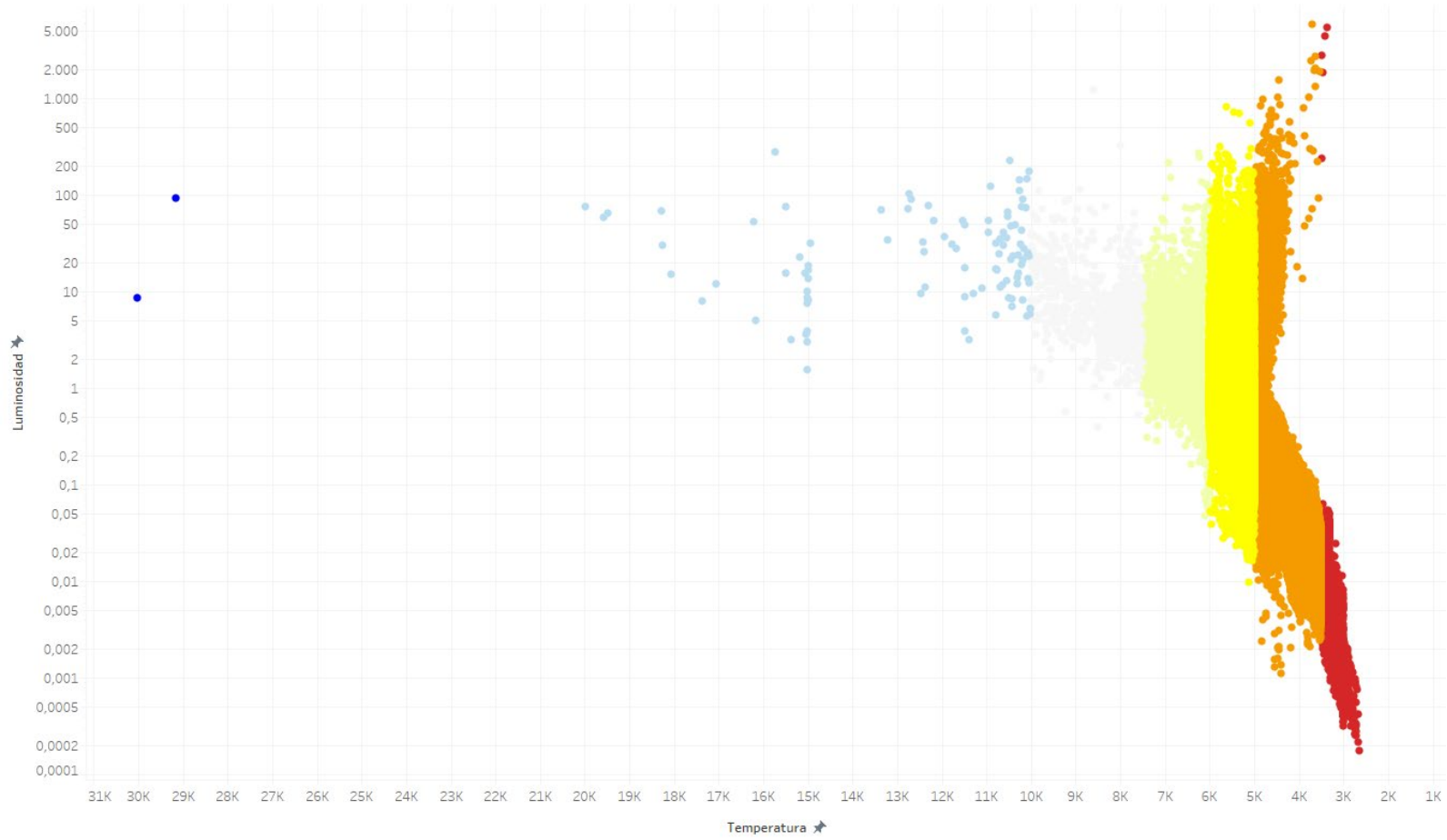
El gráfico de la Figura 12 muestra la distribución de estrellas según su tipo color espectral, fácilmente se identifica que una gran mayoría de las estrellas en este sector de la galaxia son naranjas (249 mil), amarillas (113K), esto revela que los cuerpos de espectro K y G son los tipos más comunes dentro del conjunto examinado, con periodos de vida extensos en la secuencia principal.

Además, los objetos celestes rojos (46 mil) y blanco-amarillos (15 mil) son de menor cantidad. Lo más notable en el diagrama es una casi ausencia de estrellas de color azul, posiblemente debido al consumo rápido de combustible y en consecuencia presentan vidas cortas, de ahí por qué son escasas en esta muestra del cielo.

Figura 10

Diagrama Hertzsprung-Russell

Diagrama de Hertzsprung-Russell

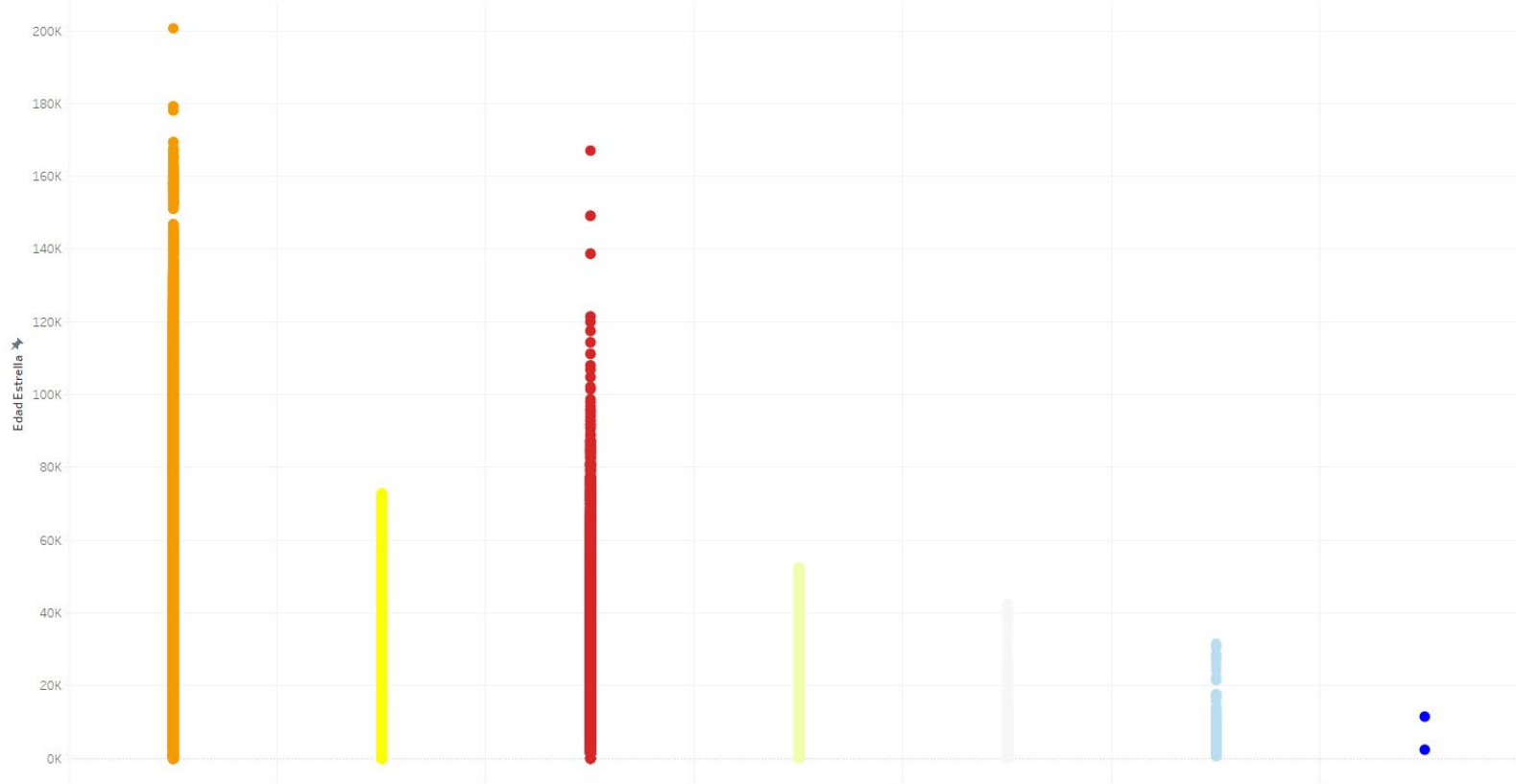


Nota. La Figura 10 describe la relación entre la Temperatura y Luminosidad.

Figura 11

Comparativa de Edad Estelar del Segmento

Comparativa Edad Estelar

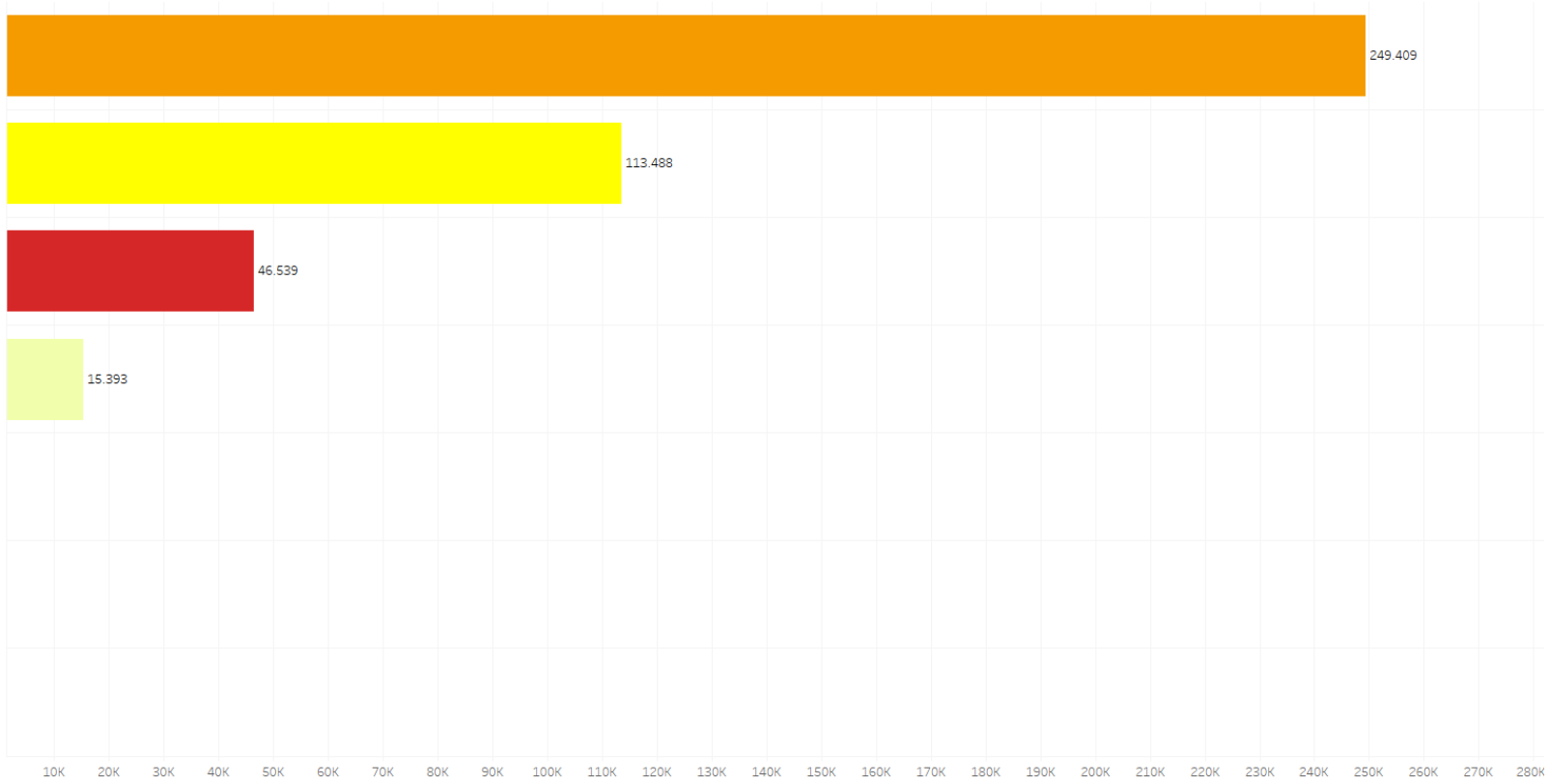


Nota. La Figura 11 describe el tiempo de vida que cada grupo de estrellas.

Figura 12

Clasificación por Grupo de Color

Distribución de Estrellas por Tipo Espectral



Nota. La Figura 12 describe la cantidad de estrellas por cada grupo de color.

b) Explicación del aporte

La adquisición de datos numéricos del repositorio de datos espaciales Gaia representa el inicio del trabajo, partiendo del conjunto de datos en formato CSV, mismo que será procesado dentro de la aplicación Visual Studio Code (VSC) y algunos complementos tales como Python (incluido librerías de análisis), Jupyter Notebooks que ayudarán a desarrollar el estudio paso a paso. En este sentido se han tomado en cuenta las siguientes etapas dentro del desarrollo del proyecto y que se detallan a continuación.

- Instalación y configuración del paquete de software comprendido por Visual Code, Python, Pandas, Datawangler, entre otros.
- Carga de datos en formato CSV del repositorio Gaia Data Release3 como fuentes de datos. Cada conjunto de datos representa un segmento escaneado del cosmos. Se unifican dos conjuntos en uno solo para este análisis.
- Exploración e inspección inicial de los datos para entender su estructura.
- Limpieza de datos necesaria para eliminar valores nulos, correcciones y seleccionar los tipos de datos adecuados.
- Análisis descriptivo para comprender los datos con la ayuda de técnicas básicas estadísticas.
- Visualizaciones para explorar relaciones en los datos con la finalidad de ayudar a interpretar los resultados.
- Documentación del proceso en Jupyter Notebooks utilizando celdas con el propósito de describir las acciones realizadas.
- Conclusiones de los resultados obtenidos.

c) Estrategias y/o técnicas

La ejecución del trabajo se fundamenta en consultas bibliográficas y en datos de observaciones astronómicas, complementados con información de diversos sitios web.

Estrategia:

Establecer la fuente de información y contestar al planteamiento.

Revisar el material bibliográfico para entender las técnicas y procesos de análisis de datos.

- Seleccionar y clasificar los datos recopilados.
- Colocar los datos de manera tabular y presentarla en forma gráfica si es necesario.

Objetivo:

- Recopilar, procesar y analizar información.
- Establecer las respectivas conclusiones de los datos y procesos analizados.

Acciones:

- Realizar una hoja de ruta, articular técnicas e implantar hitos.
- Recopilar información y guardarla dentro del repositorio local. Posteriormente se implementa el análisis.
- Categorizar y analizar los datos.
- Realizar cálculos que proporcionen sentido de los datos.
- Trasladar los datos numéricos en gráficos e interpretar los datos.
- Obtener conclusiones.

2.3 Validación de la propuesta

Se desarrolla la validación del trabajo por medio de expertos dentro del ámbito de Datos Masivos y Analistas de Datos con experiencia en el área de estudio, los cual cumplen el perfil acorde a con los criterios:

- Formación académica afín al proyecto.
- Experiencia enfocada al manejo de datos.
- Colaboración.

La valoración determina que los enfoques planteados se ajustan al nivel esperado considerándolo un aporte innovador aplicado al ámbito del manejo de datos. Las personas que realizaron la evaluación están de acuerdo con la propuesta acorde al contenido, normativas y teorías en la forma de efectuar el análisis.

En la Tabla 3 se presenta el perfil de los especialistas que participaron con su aporte sobre el trabajo realizado.

Tabla 3*Perfil de Especialistas*

Apellidos y Nombres	Años de experiencia	Titulación Académica	Cargo
Mg. Garcia Llanos Carmen Isabel	15 años	Magíster en Gestión De Sistemas De Información E Inteligencia De Negocios	Analista de Control
Mg. Andrade Alban Jose Rodolfo	20 años	Máster en Docencia	Analista de Facturación

Nota. Elaboración propia

En la Tabla 4 se especifican los criterios de valoración establecidos en esta encuesta.

Tabla 4*Criterios de Valoración*

Criterio	Descripción
Impacto	El alcance que tendrá la propuesta y su representatividad en la generación de valor
Aplicabilidad	La capacidad de implementación de la propuesta considerando que los contenidos sean aplicables
Conceptualización	La base de conceptos y teorías propias de la propuesta de manera sistémica y articulada
Actualidad	Los procedimientos actuales y los cambios científicos y tecnológicos considerados en la propuesta
Calidad Técnica	Los atributos cualitativos del contenido de la propuesta para satisfacer las expectativas de sus beneficiarios
Factibilidad	El nivel de utilización de la propuesta por parte de la organización acorde a los recursos disponibles
Pertinencia	La contundencia y conveniencia de la propuesta para solucionar el problema planteado.

Nota. Elaboración propia

En la Tabla 5 se exponen los resultados de los criterios de evaluación utilizando una escala ponderativa estandarizada que cuantifica la aceptación de los componentes que han sido evaluados. A continuación, se presentan los parámetros ponderados, sobre la validación de los expertos.

Tabla 5

Análisis de Valoración de los Especialistas

CRITERIOS	EXPERTO 1 (Jose Andrade)	EXPERTO 2 (Carmen García)	TOTAL	Porcentaje
Impacto	5	5	10	100%
Aplicabilidad	5	5	10	100%
Conceptualización	5	5	10	100%
Actualidad	5	5	10	100%
Calidad Técnica	5	5	10	100%
Factibilidad	5	5	10	100%
Pertinencia	5	5	10	100%
Total	35	35	70	100%

Nota. Elaboración propia

2.4 Matriz de articulación

En la Tabla 6 la matriz articula los sustentos teóricos, metodológicos, estratégicos-técnicos y tecnológicos empleados.

Tabla 6

Matriz de Articulación

Ejes o Partes Principales	Sustento Teórico	Sustento Metodológico	Estrategias / Técnicas	Descripción de Resultados	Instrumentos Aplicados
Obtención y Preparación de Datos	Repositorio astronómico, variables relevantes (Gaia Archive, s.f.)	Selección de datos y normalización	Pandas para filtrar, limpiar y estructurar datos	Datos listos para análisis	Python (Pandas) en Jupyter Notebooks
Limpieza y Transformación de Datos	Procesamiento de datos en astronomía (Datacamp, s.f.)	Aplicación de filtros y eliminación de valores atípicos	Normalización, eliminación de valores nulos o inconsistentes	Datos depurados con valores significativos	Python (Pandas) en Jupyter Notebooks
Generación de Nuevas Variables	Relación entre temperatura, color, espectro y luminosidad (Gaia Archive, s.f.)	Cálculo de luminosidad, clasificación espectral y estimaciones	Creación de nuevas columnas en base a ecuaciones astronómicas	Datos enriquecidos con información interpretativa	Python (Pandas) en Jupyter Notebooks
Estimación de Edad Estelar	Uso de variables como distancia, gravedad,	Cálculo de edad empírico	Operaciones matemáticas sobre variables clave	Estimaciones aproximadas de edad estelar	Python (Pandas) en Jupyter Notebooks

Ejes o Partes Principales	Sustento Teórico	Sustento Metodológico	Estrategias / Técnicas	Descripción de Resultados	Instrumentos Aplicados
	metalicidad (eBlanc, 2010)				
Visualización de Datos	Métodos de representación de datos astronómicos	Generación de gráficos de análisis visual	Gráficos de dispersión y diagramas de barras	Diagramas HR, tendencias de luminosidad y espectro	Tableau

Nota. Elaboración propia

CONCLUSIONES

Las herramientas de Big Data permiten procesar, analizar y visualizar grandes volúmenes de datos astronómicos, optimizando el estudio de los parámetros físicos de los cuerpos celestes. La accesibilidad a estas tecnologías facilita la caracterización, permitiendo la estimación de características a partir de atributos físicos de los cuerpos celestes.

En este trabajo se han evaluado las características de la temperatura, luminosidad y edad sobre una mínima parte del conjunto de datos base, se requieren de capacidades de mayor potencia de cómputo para procesar más información debido a la naturaleza y volumen de los datos.

El análisis gráfico de los datos arrojó que en el segmento del cosmos estudiado existe una alta concentración de estrellas de espectro Naranja y Rojo, y esto se debe cuando una nube de gas y polvo estelar en el espacio se colapsa para formar estrellas, en este sentido el material se reparte de forma desigual, haciendo más sencillo que se formen estrellas pequeñas porque estas necesitan menos material y en contraposición sólo se observan dos estrellas del tipo Azul debido a que necesitan de mucho más material para generar la fusión, por esta razón son las más raras.

La generación de gráficos y tablas facilita la interpretación de los resultados y contribuye a identificar patrones relevantes, demostrando ser una herramienta clave en la evaluación del desempeño del modelo.

RECOMENDACIONES

Se sugiere la implementación de plataformas de Big Data en la nube en caso que se necesiten mayores capacidades de procesamiento cuando se trata de datos astronómicos, así de esta manera se optimiza el tiempo del análisis.

Además de las estrellas, se sugiere investigar otros cuerpos celestes como planetas, asteroides, cometas y agujeros negros que también puede generar otros tipos de estudio de objetos que aporten información adicional para estudiar la interacción entre estos diferentes tipos de cuerpos celestes.

Se propone complementar el análisis con simulaciones computacionales para validar las observaciones y entender mejor la distribución de masas estelares en diferentes regiones del cosmos.

La valoración de los especialistas debe ser llevada a cabo por personas con el nivel de experiencia adecuado, a fin de garantizar una evaluación precisa y transmitir con claridad la viabilidad de la investigación.

BIBLIOGRAFÍA

- Cantat-Gaudin, T., Varga, J., y Hogerheijde, M. (2021). Dating Star Clusters in the Milky Way with Gaia EDR3 Data.
- Creevey, O., Sordo, R., Pailer, F., y Frémat, Y. (2022). Stellar Classification Using Machine Learning and Gaia DR3 Spectroscopy.
- Datacamp. (s.f.). *A Guide to Big Data Training*. <https://www.datacamp.com/es/blog/guide-to-big-data-training>
- eBlanc, F. (2010). *An Introduction to Stellar Astrophysics*. Canada: Astrophysics-Textbooks.
- European Space Agency. (2024). *European Space Agency*. <https://www.esa.int/>
- European Space Agency. (s.f.). *Getting Data*. <https://www.cosmos.esa.int/web/gaia-users/archive>
- Fouesneau, M., y Schultheis, M. (2023). Age Estimation of Stars Using Gaia DR3 Data.
- Futurism. (s.f.). *The Life Cycle Of A Star*. <https://futurism.com/the-life-cycle-of-a-star>
- Gaia Archive. (s.f.). *Gaia ESA Archive*. <https://gea.esac.esa.int/archive/>
- gaia data release 3 documentation. (s.f.). *gaia data release 3 documentation*. https://gea.esac.esa.int/archive/documentation/GDR3/Gaia_archive/chap_datamodel/sec_dm_astrophysical_parameter_tables/ssec_dm_astrophysical_parameters.html?utm_source=chatgpt.com
- Gaia_source. (2023). *20.1.1 gaia_source ▶ 20.1 Main source catalogue ▶ Chapter 20 Datamodel description ▶ Part V Gaia archive ▶ Gaia Data Release 3 Documentation release 1.3*. https://gea.esac.esa.int/archive/documentation/GDR3/Gaia_archive/chap_datamodel/sec_dm_main_source_catalogue/ssec_dm_gaia_source.html
- GAIERVERSE. (s.f.). <https://gaiaverse.eu/la-mision-gaia/>
- hevodata.com. (s.f.). *Understanding Big Data Processing: 2025's Ultimate Guide*. <https://hevodata.com/learn/big-data-processing/>
- <https://www.netacad.com>. (s.f.). *Introducción a la Ciencia de Datos*.

Interactive Chaos. (s.f.). *Librerías en Data Science*. <https://interactivechaos.com/es/manual/tutorial-de-python/librerias-en-data-science>

Microsoft. (s.f.). *Python in Visual Studio Code*. <https://code.visualstudio.com/docs/languages/python>

OpenAI.(2024).ChatGPT (versión del 15 de julio de julio)[Modelo de lenguaje de gran tamaño]. (s.f.). <https://chat.openai.com/chat>

Project Jupyter. (s.f.). *JupyterLab*. <https://jupyter.org/about>

Sampieri, R. H. (2014). *Metodología de la Investigación*. (S. D. McGRAW-HILL / INTERAMERICANA EDITORES, Ed.)

Tableau. (s.f.). *what-is-tableau*. <https://www.tableau.com/es-es/why-tableau/what-is-tableau>

The Data Visualisation Project. (2024). *Data Viz Project*. <https://datavizproject.com/>

www.cosmos.esa. (2024). *Data Release 3*. <https://www.cosmos.esa.int/web/gaia/data-release-3>

www.datasource.ai. (s.f.). *Todo Sobre El manejo De Datos Faltantes*. https://www.datasource.ai/es/data-science-articles/todo-sobre-el-manejo-de-datos-faltantes?utm_source=chatgpt.com

www.ute.edu.ec. (s.f.). *ODS 9 - Industria, innovación e infraestructura*. <https://www.ute.edu.ec/objetivos-de-desarrollo-sostenible/ods-9-industria-innovacion-e-infraestructura/>

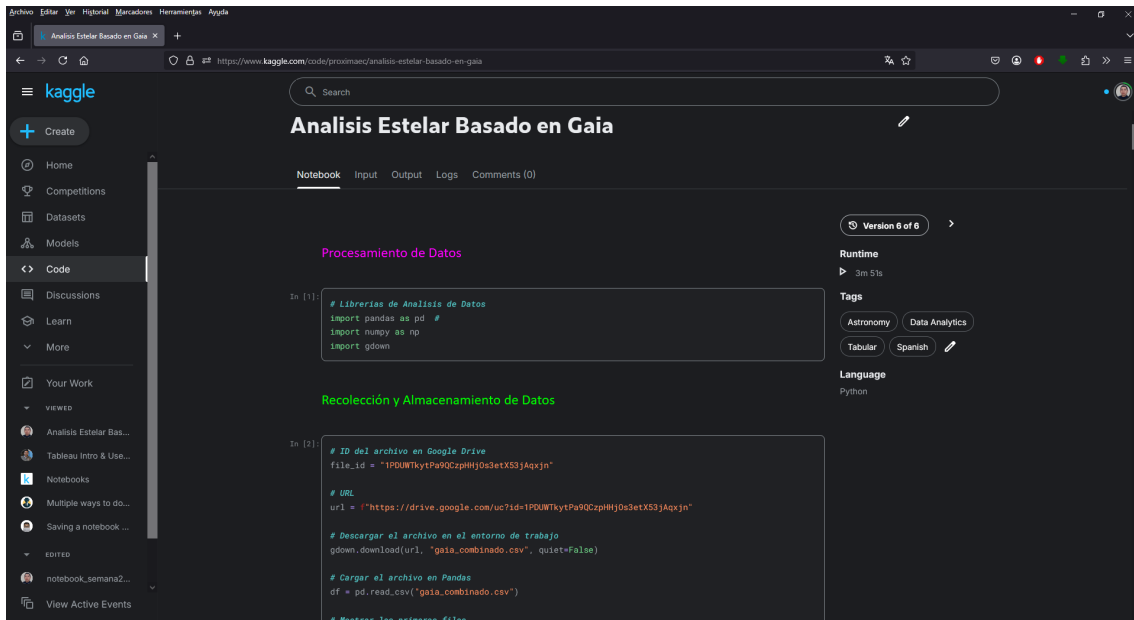
www.womenshistory.org. (s.f.). *annie-jump-cannon*. <https://www.womenshistory.org/education-resources/biographies/annie-jump-cannon>

ANEXOS

Publicación en la Web

Plataforma de Kaggle

<https://www.kaggle.com/code/proximaec/analisis-estelar-basado-en-gaia>



The screenshot shows a Kaggle notebook interface with the following content:

- Header:** "Analisis Estelar Basado en Gaia" with a search bar and navigation tabs (Notebook, Input, Output, Logs, Comments).
- Section 1:** "Procesamiento de Datos" (Data Processing) with code cell In [1]:

```
# Librerías de Análisis de Datos
import pandas as pd #
import numpy as np
import gdown
```
- Section 2:** "Recolección y Almacenamiento de Datos" (Collection and Storage) with code cell In [2]:

```
# ID del archivo en Google Drive
file_id = "1PDUWtKytPa9QCzpiHj0s3etX53jAqxjn"

# URL
url = f"https://drive.google.com/uc?id=1PDUWtKytPa9QCzpiHj0s3etX53jAqxjn"

# Descargar el archivo en el entorno de trabajo
gdown.download(url, "gaia_combinado.csv", quiet=False)

# Cargar el archivo en Pandas
df = pd.read_csv("gaia_combinado.csv")

# Mostrar los primeros fil...
```
- Runtime:** 3m 51s
- Tags:** Astronomy, Data Analytics, Tabular, Spanish
- Language:** Python

Plataforma de Tableau

<https://public.tableau.com/app/profile/paul.andrade10/viz/gaia/H1?publish=yes>

UNIVERSIDAD TECNOLÓGICA ISRAEL

ESCUELA DE POSGRADOS “ESPOG”

MAESTRÍA EN BIG DATA Y CIENCIA DE DATOS

INSTRUMENTO PARA VALIDACIÓN DE LA PROPUESTA

Descripción de la Propuesta

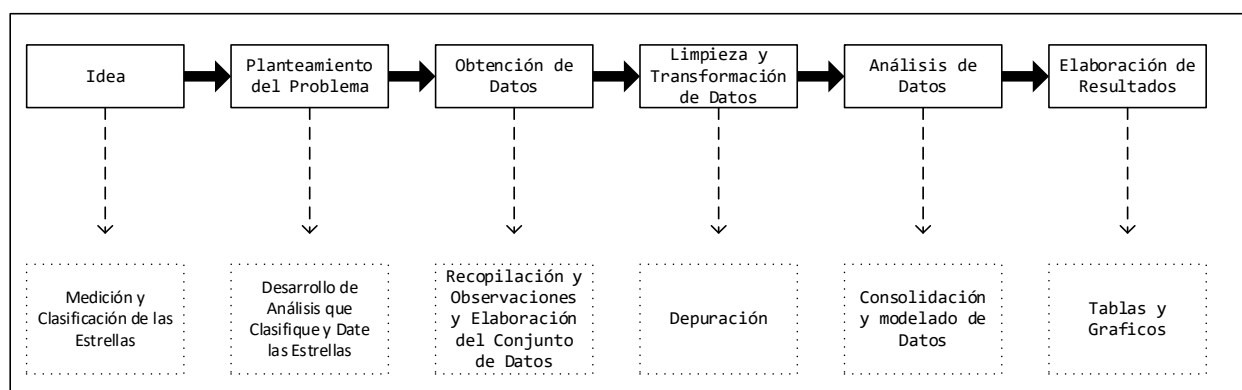
El presente trabajo que trata sobre la estructura de la galaxia, ha permitido recopilar información sobre características de definen las estrellas mediante el uso de técnicas de datos masivos, donde se utilizan algunos utilitarios para explorar y analizar datos, el enfoque está orientado a describir comportamientos clave sobre la evolución estelar, además de aplicar el manejo de muchos registros para extraer conclusiones significativas que contribuyan por medio de los datos.

a) Estructura general

El proceso de análisis de datos tiene una estructura que articula las siguientes estrategias. La Figura 2 describe el proceso del ciclo del proyecto.

Figura 1

Diagrama del Ciclo de Investigación



Nota. En la Figura 2 se presentan las fases del ciclo del proyecto que permitieron conseguir el objetivo.

El procesamiento de datos de la Misión Gaia se encuentra dentro de los preceptos que abarcan el **Big Data** por cuanto el volumen de la información incluye millones de registros procedentes de las observaciones astronómicas y constituye un tamaño que oscila en el orden de los gigabytes. La variedad hace referencia a una gran cantidad de variables por cuanto los archivos contienen una estructura de 153 columnas y n-filas que representan una observación estelar, a pesar que los datos no se representan en tiempo real se pueden utilizar estrategias de administración de rendimiento y memoria en caso de ser requerido. Toda la información procede de la misión Gaia, aunque pueden existir errores que necesiten una validación. La extracción de información significativa permite ayudar a realizar la clasificación y evolución de estrellas (Netacad, s.f.).

Durante el proceso de extracción de datos primero se obtienen los datos de interés, para posteriormente unificar estos catálogos en un único archivo. Posteriormente el nuevo archivo se carga a la aplicación, el comando de Pandas “df.head()” muestra las primeras filas del conjunto para observar su estructura y validar la importación de los datos, en este punto solo se dispone de datos de naturaleza numérica. En la Figura 3 se presenta la muestra cargada en la herramienta de Visual Code mediante un libro de Jupyter comúnmente conocido como notebook. Dicho proceso se ampara bajos enunciados expuestos dentro de la fase de recopilación (Hevodata, s.f.).

Figura 2

Carga del Conjunto de Datos

```
# Cargar el conjunto
df = pd.read_csv('gaia_combinado.csv')

# Mostrar todas las columnas
df.head()
```

Python

	solution_id	designation	source_id	random_index	ref_epoch	ra	ra_error	dec	dec_error
0	1636148068921376768	Gaia DR3 4295806720	4295806720	545300884	2016.0	44.996155	0.101618	0.005615	0.101334
1	1636148068921376768	Gaia DR3 34361129088	34361129088	894504938	2016.0	45.004320	0.097320	0.021048	0.101753
2	1636148068921376768	Gaia DR3 38655544960	38655544960	1757259052	2016.0	45.004978	0.017885	0.019880	0.018772
3	1636148068921376768	Gaia DR3 309238066432	309238066432	327609758	2016.0	44.995037	0.322039	0.038152	0.283505
4	1636148068921376768	Gaia DR3 343597448960	343597448960	1791600903	2016.0	44.963896	0.117176	0.043595	0.109004

5 rows x 153 columns

Nota. La Figura 3 describe la carga de datos dentro de la aplicación Visual Studio Code.

Ahora que se dispone de todos los datos se realiza el proceso de transformación con la finalidad de establecer el formato requerido, **eliminar inconsistencias, valores nulos** para garantizar su calidad (Hevodata, s.f.).

El problema que se tiene con los datos al momento, radica en que muchas observaciones tienen campos vacíos. Existen algunas estrategias que permiten afrontar el tratamiento de nulos, en la Tabla 2 se presenta el detalle de algunas estrategias.

Tabla 1

Tratamiento de Valores Nulos

Estrategia	Descripción	Ventajas	Desventajas
Eliminar filas con campos vacíos	Elimina filas con datos faltantes	Fácil de aplicar, no introduce sesgos	Se pierde información importante
Rellenar con un valor específico	Sustituye valores faltantes con la media, mediana o valores fijos	Mantiene el tamaño del conjunto de datos	Puede introducir sesgos si no es representativo

Estrategia	Descripción	Ventajas	Desventajas
Reemplazo con algoritmos	Predice valores faltantes usando modelos estadísticos	Más preciso, ideal para valores críticos	Requiere más trabajo y modelos complejos
Mantener campos vacíos	Deja los valores vacíos si son relevantes para el análisis	Útil si la ausencia de datos tiene significado	Complica el análisis si no se maneja adecuadamente

Nota. La Tabla 3 las estrategias de manejo de valores faltantes. Fuente Datasource.

De la tabla descrita con anterioridad el trabajo con datos espaciales, rellenar esos valores con estimaciones o mantener los vacíos no es útil, ya que puede introducir errores o sesgos en el análisis, además de existir una cantidad significativa de columnas sin valores en cada registro. En este caso cuando se manejan datos espaciales, eliminar las filas con valores nulos resulta ser la opción más conveniente al asegurar que el conjunto de datos sea más preciso y consistente y permita mantener la calidad del análisis evitando resultados distorsionados, donde cada valor tiene un impacto significativo en los resultados. En la Figura 4 se presenta una muestra del proceso de limpieza de los datos.

Figura 3

Proceso de Limpieza

```
# Eliminar entradas con valores nulos
df2 = filtrar.dropna()

# Comprobar entradas con valores nulos
df2.isna().sum()
```

Python

```
dataset_name          0
solution_id           0
source_id             0
teff_gspphot         0
logg_gspphot         0
mh_gspphot           0
phot_g_mean_mag      0
distance_gspphot     0
non_single_star      0
classprob_dsc_combmod_star  0
classprob_dsc_combmod_quasar 0
classprob_dsc_combmod_galaxy 0
dtype: int64
```

Nota. La Figura 4 describe la limpieza de los datos.

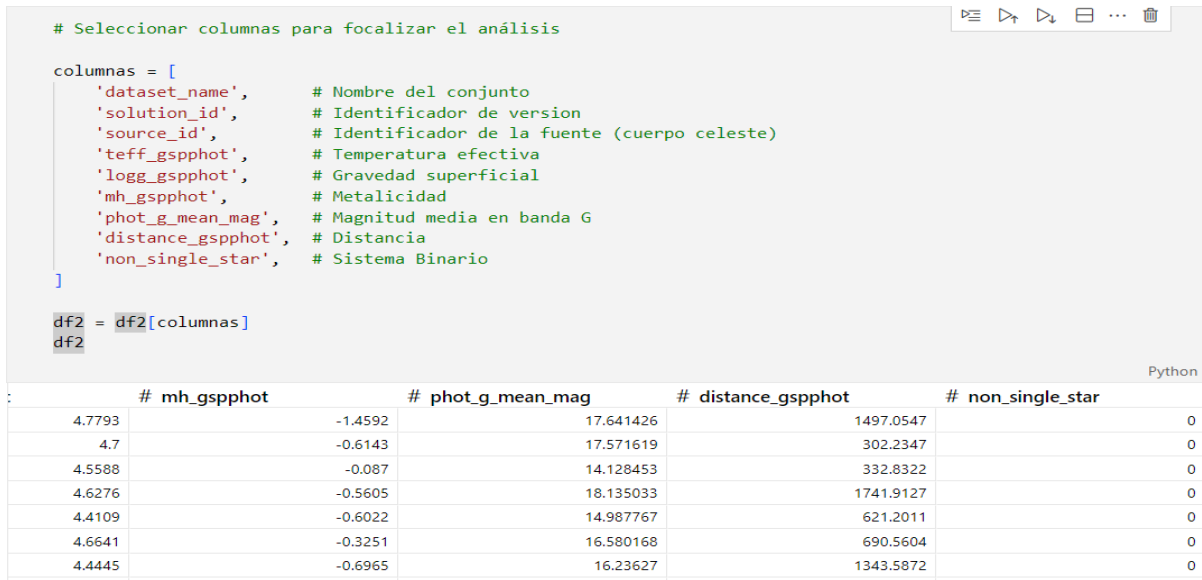
Tomando en cuenta que el repositorio incluye información de muchos cuerpos celestes, se aplican los correspondientes filtros para únicamente seleccionar estrellas. El conjunto incluye tres campos importantes relacionados a la probabilidad de ser una estrella, fuente de radio casi estrellar (quasar) o un agujero negro, mientras en factor de 1 (100%) se acerque a este valor, más probable en recaer en cualquiera de estas categorías será la posibilidad (Gaia_source, 2023).

- classprob_dsc_combmod_star → % de Estrella
- classprob_dsc_combmod_quasar → % de Quasar
- classprob_dsc_combmod_galaxy → % de Galaxia

En ese caso el criterio que se establece corresponde que un factor de 0.9 a 1 pertenece a una fuente de luz que se acerca a representar una estrella. Cuando se ha alcanzado la coherencia y normalizaron de los formatos de los datos para su correcta integración con herramientas de análisis, se deben aplicar un sentido a todas esas observaciones de tipo numérico, se intenta categorizar los datos como base a las indicaciones que se encuentran disponibles en la enciclopedia en línea de la documentación Gaia (gaia data release 3 documentation, s.f.). Además, se seleccionan únicamente los campos de interés para este presente estudio. La Figura 5 presenta la selección de los campos.

Figura 4

Selección de Columnas



Nota. La Figura 5 describe la selección de los campos de interés.

En esta etapa se asignan nombres más intuitivos en concordancia con el fenómeno sujeto de estudio, el detalle de los nombres se define a continuación:

- Nombre del conjunto
- Identificador de versión
- Identificador de la fuente
- Temperatura efectiva
- Gravedad superficial
- Metalicidad
- Magnitud media en banda G
- Distancia
- Sistema Binario

La Figura 6 muestra los nombres de las columnas que son reemplazados por textos más simples de entender.

Figura 5

Actualización de Nombres de Columnas

df2						Python
	# Temperatura	# Gravedad	# Metalicidad	# Magnitud	# Distanci	
0	5052.976		4.7793	-1.4592	17.641426	
8	3478.5408		4.7	-0.6143	17.571619	
0	4708.7944		4.5588	-0.087	14.128453	
0	4536.664		4.6276	-0.5605	18.135033	
2	4837.496		4.4109	-0.6022	14.987767	
0	4333.865		4.6641	-0.3251	16.580168	
2	5040.7686		4.4445	-0.6965	16.23627	
2	4550.9297		4.8175	-1.2501	17.778534	
4	4319.841		4.8899	-1.3928	18.615105	
4	3404.9011		4.4294	-1.1774	18.364864	

Nota. La Figura 6 describe la actualización de los nombres de las columnas.

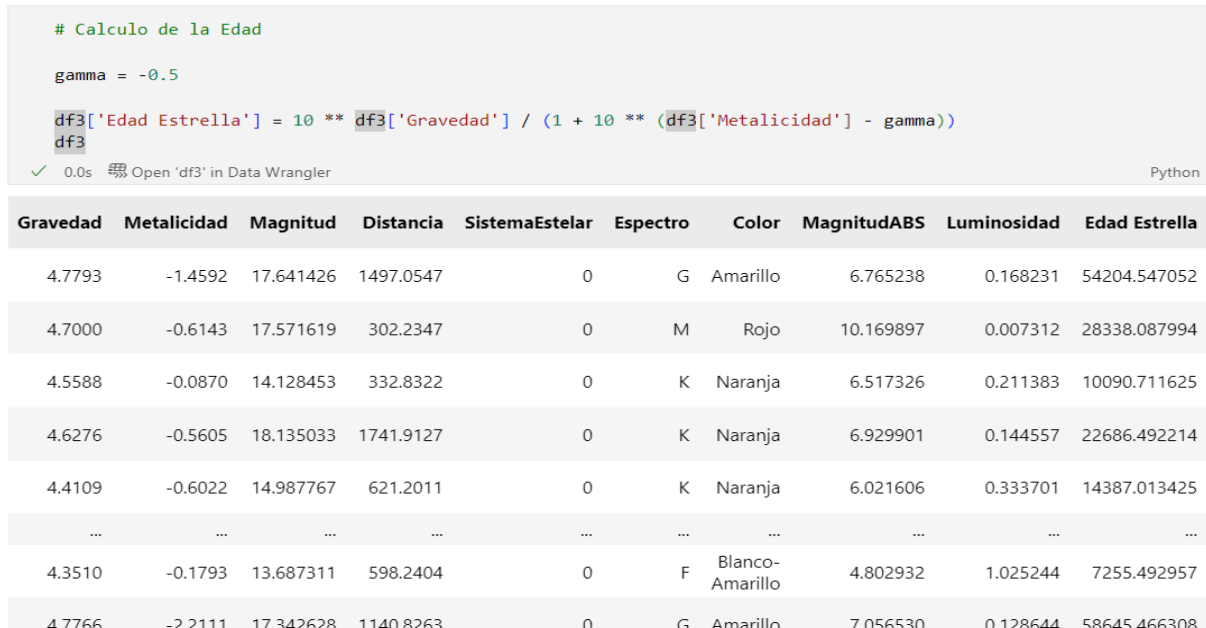
En esta fase que se aplica al conjunto de datos, se realiza una serie de operaciones matemáticas que van a dar sentido a los datos observacionales, se define la clasificación de las estrellas con base al espectro de luz visible que podría traducirse como el color del cuerpo celeste. Además, se realizan los cálculos para determinar la luminosidad de las estrellas tomando como referencia el resplandor de nuestra estrella "El sol". Por último, se determina el cálculo aproximado de la edad de una estrella, así de esta manera se constituyen todos los parámetros requeridos para llevar a cabo el análisis.

El conjunto de datos resultante tiene el valor científico que va a permitir estudiar la relación entre la temperatura y luminosidad, analizar la distribución de estrellas por tipo espectral, además, de estudiar su comportamiento conforme a la edad.

En la Figura 7 se indica una muestra de los cálculos ejecutados para completar esta etapa del análisis.

Figura 6

Cálculos Aplicados que Proporciona el Sentido a los Datos



Nota. La Figura 7 describe la parte final de los cálculos aplicados a los datos para interpretar los datos.

En la siguiente fase relacionado a la carga se enfoca en transformar y estructurar los datos antes de su visualización, sin necesidad de una base de datos. El archivo final generado se encuentra optimizado para el análisis haciendo uso de herramientas de visualización como **Tableau**. El proceso realizado con anterioridad asegura que los datos sean **consistentes y listos para su uso**, en este sentido los datos se encuentran en un formato adecuado para exploración y análisis visual (Hevodata, s.f.).

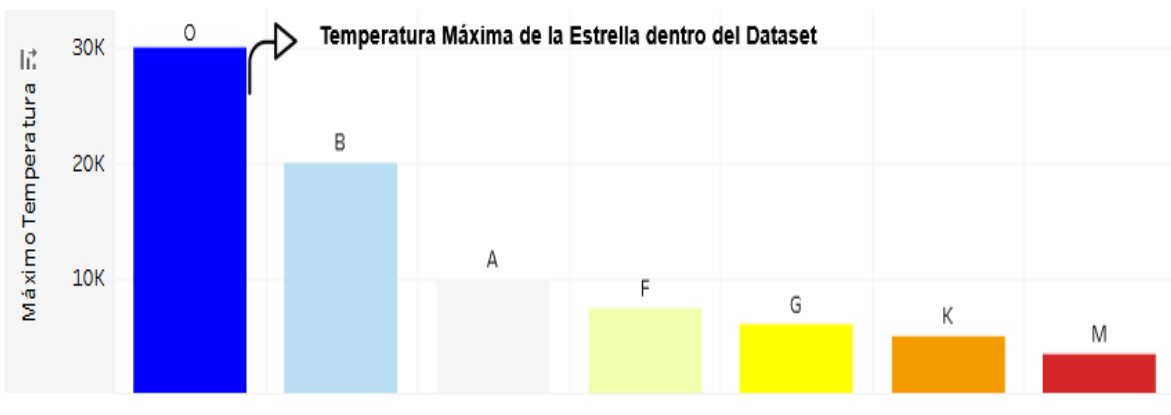
En la **etapa de visualización**, la información procesada se transmuta en gráficos y tablas que permitan facilitar el análisis. **Tableau** permite observar las relaciones en los datos de forma clara por medio de visualizaciones interactivas para un mejor entendimiento de la información sin necesidad de analizar números en forma directa. Los paneles interactivos permiten explorar la información de forma sencilla. Además, proporciona un ambiente perfecto que ofrece integración fácil con fuentes de datos compatibles y permite crear tableros accesibles y eficientes para proyectos con plazos cortos. Dispone de publicación web sin necesidad de infraestructura adicional, aprovechando que se dispone de licenciamiento de mano de la Universidad.

Con respecto al análisis visual el sistema Morgan-Keenan es el método utilizado para efectuar la clasificación. La Figura 8 establece que los datos obtenidos en este segmento del espacio incluyen estrellas que se ajustan en todos los espectros, las estrellas del tipo O tienen una temperatura que alcanza los 30.000 mil grados kelvin, siendo el grupo de estrellas con el mayor grado de calor dentro de este segmento.

Figura 7

Máximo Grado de Calor en el conjunto de Datos

Sistema de Clasificación del Conjunto de Datos

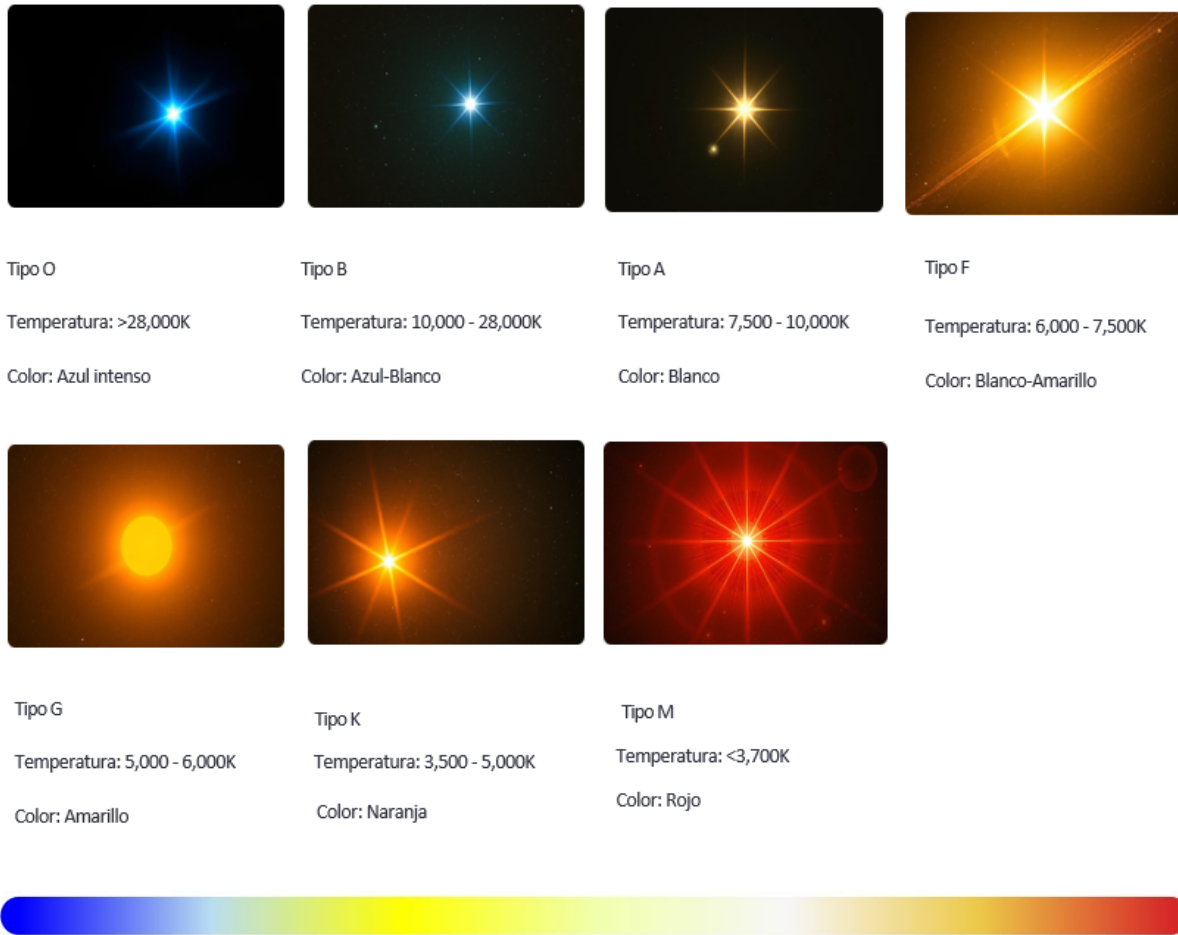


Nota. La Figura 8 describe la temperatura máxima de las estrellas dentro del conjunto de datos.

En la Figura 9 se puede observar un espectro de color visible asociado a la temperatura de la estrella que facilita la comprensión, aquí se puede observar que a mayor calor irradiado por el cuerpo más azulado será el color de la estrella, y en contraposición mientras menor sea el calor del cuerpo más fría será, esto dentro de los términos de las altas temperaturas. Existen otros cuerpos celestiales con menores escalas de calor, pero estos no se encuentran formando parte del estudio.

Figura 8

Clasificación Espectral de la Estrella Asociado a un Color



Nota. La Figura 9 describe el color de la estrella en base a su temperatura, junto con la escala de calor. Autoría Propia generado con IA.

La luminosidad de cada estrella especifica la cantidad de energía que se emite en forma de luz, y está asociado con la temperatura y tamaño del objeto. Significa que los cuerpos más calientes emiten una mayor cantidad de energía y en consecuencia son más brillantes con respecto a aquellas más frías.

La relación del factor de luz, luminosidad y temperatura de las estrellas, se toma como referencia nuestra estrella madre (Sol) que parte de una luminosidad estándar de $1 L_{\odot}$ (luminosidad solar) con

una temperatura de alrededor 5,700 K. Sin embargo, a medida que la luminosidad aumenta en factores de 10, 100, n-veces en función del Sol, la temperatura del cuerpo se incrementa.

En relación con la edad de las estrellas, el grupo de estrellas en sus fases más tempranas son más calientes y brillantes con un tiempo vida más corto, este es el caso de las estrellas de color más azuladas que se encuentran bajo los espectros O y B debido al consumo rápido de su combustible nuclear. Por otro lado, las estrellas más frías y pequeñas de colores con más tendencia al rojo, como el tipo espectral K y M tienen un tiempo de vida prolongado debido a un consumo bajo del combustible nuclear.

Cuando las estrellas envejecen, la tendencia es que la temperatura se reduzca, al igual que su luminosidad. Las características de temperatura, luminosidad y edad de una estrella están estrechamente relacionadas, dicha relación puede observarse mediante un diagrama de dispersión en forma de Diagrama Hertzsprung-Russell (H-R) para visualizar la relación entre las características mencionadas.

En el Eje X se encuentran los valores de Temperatura expresado en grados kelvin ($^{\circ}\text{K}$) en forma de escala inversa, donde el grado de calor disminuye de izquierda a derecha. El Eje Y de Luminosidad expresado en términos de la luminosidad del sol (L_{\odot}) expone en escala logarítmica que tan luminoso es el cuerpo con respecto con nuestro Sol. Cada punto corresponde a un cuerpo celeste dentro de un grupo específico asociado a su color o espectro.

En la Figura 10 el diagrama describe que la mayoría de estrellas ocupa el área diagonal partiendo desde la porción superior izquierda de altas temperatura y luminosidad hasta llegar al segmento inferior derecha de bajas temperaturas y luminosidad, esto significa que las estrellas se encuentran en su fase de secuencia principal porque están utilizando la fusión de hidrógeno en sus núcleos activos. La región superior por encima de la línea imaginaria de la secuencia principal contiene estrellas la evolucionadas de alta luminosidad, pero con índices de calor más bajos, en contraste por debajo de la línea se ubican estrellas de menor tamaño que no poseen mucho combustible nuclear de alta temperatura, pero con reducida luminosidad.

En la Figura 11 el diagrama ayuda a identificar claramente que las estrellas que tienen más al color rojo tienen un mayor tiempo de vida, en contraste con las que tienen más azul, esto se debe a que las estrellas de mayor temperatura consumen más rápido su combustible principal de hidrógeno. Las estrellas **azules** son las más **jóvenes** y las **rojas y naranjas** son las más **antiguas**, alcanzando edades

muy elevadas, en este caso tenemos algunas que se acercan a los 200 mil millones de años. Las estrellas entre de la categorías **blancas - amarillo** tienen periodos de vida intermedias.

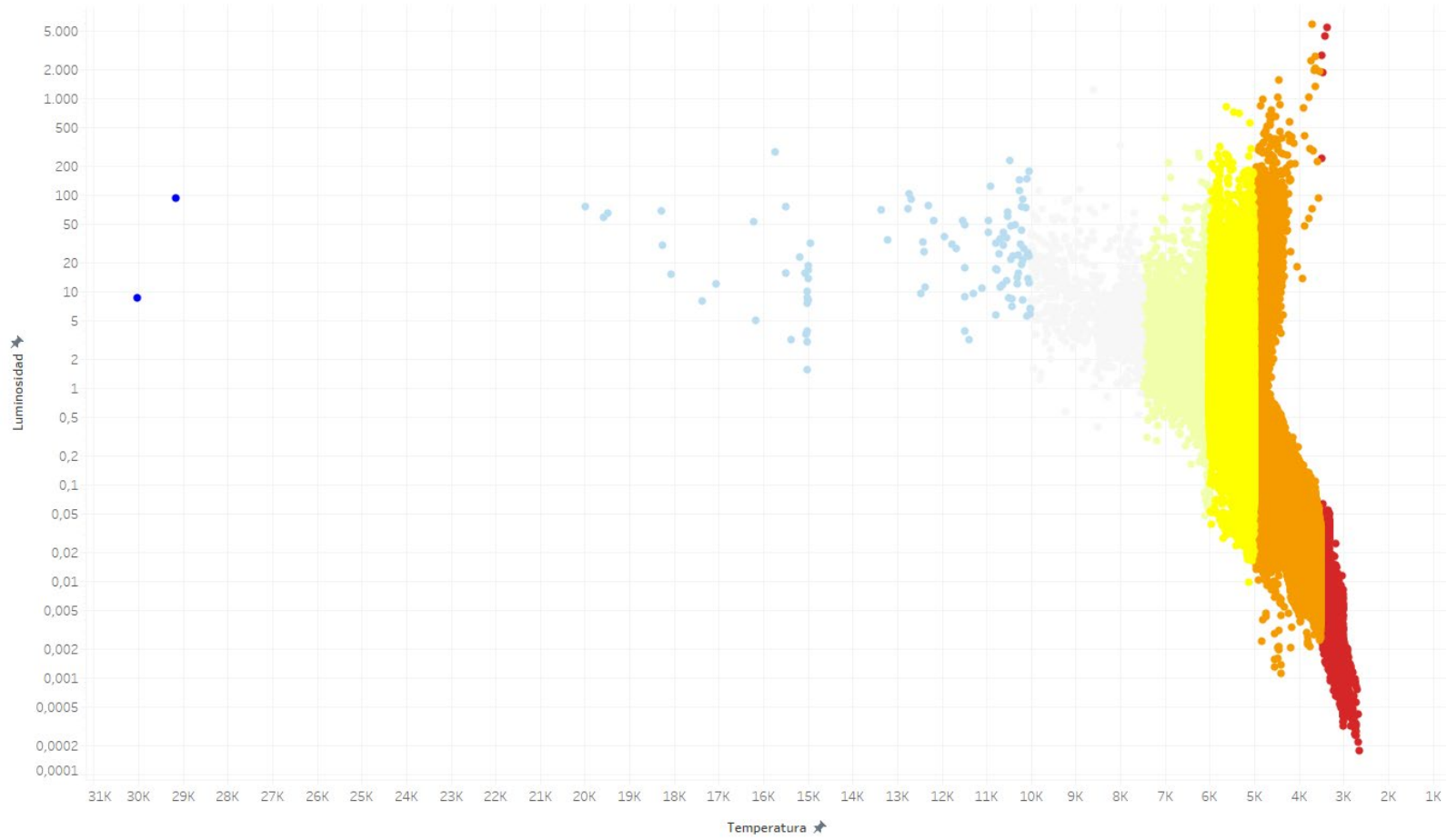
El **gráfico** de la Figura 12 **muestra la distribución de estrellas según su tipo color espectral**, fácilmente se identifica que una gran mayoría de las estrellas en este sector de la galaxia son naranjas (249 mil), amarillas (113K), esto revela que los cuerpos de espectro K y G son los tipos más comunes dentro del conjunto examinado, con periodos de vida extensos en la secuencia principal.

Además, los objetos celestes rojos (46 mil) y blanco-amarillos (15 mil) son de menor cantidad. Lo más notable en el diagrama es una casi ausencia de estrellas de color azul, posiblemente debido al consumo rápido de combustible y en consecuencia presentan vidas cortas, de ahí por qué son escasas en esta muestra del cielo.

Figura 9

Diagrama Hertzsprung-Russell

Diagrama de Hertzsprung-Russell

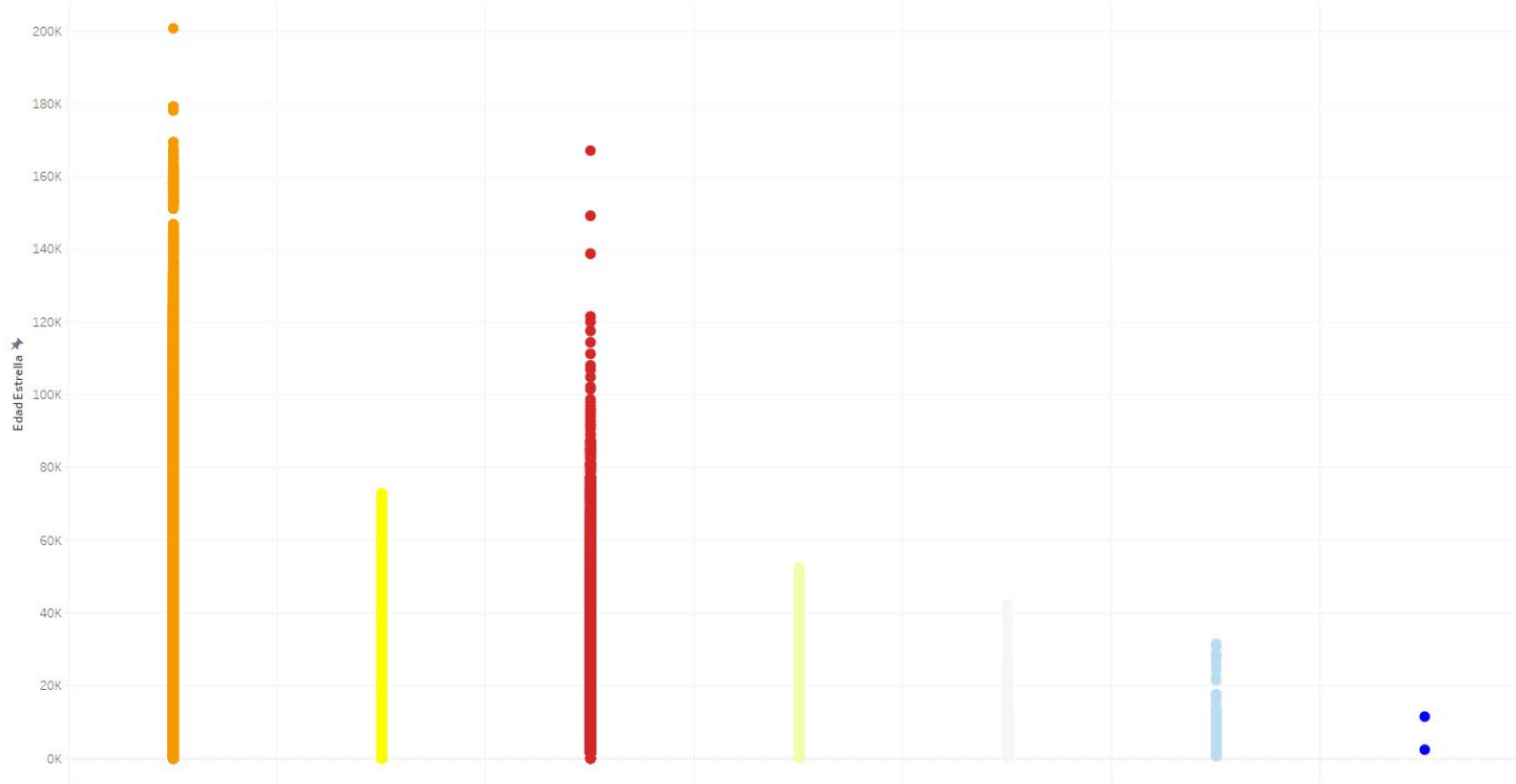


Nota. La Figura 10 describe la relación entre la Temperatura y Luminosidad.

Figura 10

Comparativa de Edad Estelar del Segmento

Comparativa Edad Estelar

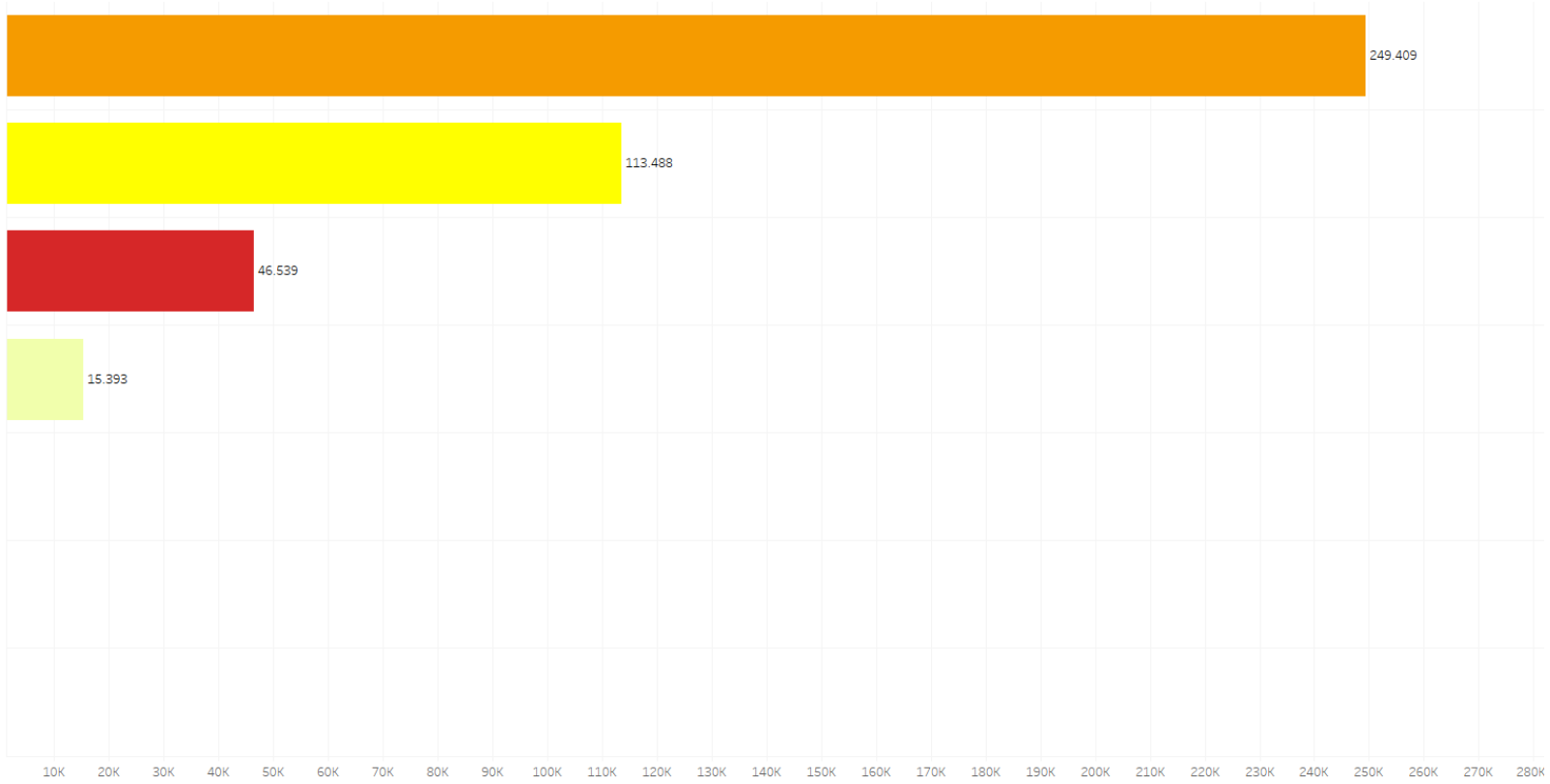


Nota. La Figura 11 describe el tiempo de vida que cada grupo de estrellas.

Figura 11

Clasificación por Grupo de Color

Distribución de Estrellas por Tipo Espectral



Nota. La Figura 12 describe la cantidad de estrellas por cada grupo de color.

b) Explicación del aporte

La adquisición de datos numéricos del repositorio de datos espaciales Gaia representa el inicio del trabajo, partiendo del conjunto de datos en formato CSV, mismo que será procesado dentro de la aplicación Visual Studio Code (VSC) y algunos complementos tales como Python (incluido librerías de análisis), Jupyter Notebooks que ayudarán a desarrollar el estudio paso a paso. En este sentido se han tomado en cuenta las siguientes etapas dentro del desarrollo del proyecto y que se detallan a continuación.

- Instalación y configuración del paquete de software comprendido por Visual Code, Python, Pandas, Datawangler, entre otros.
- Carga de datos en formato CSV del repositorio Gaia Data Release3 como fuentes de datos. Cada conjunto de datos representa un segmento escaneado del cosmos. Se unifican dos conjuntos en uno solo para este análisis.
- Exploración e inspección inicial de los datos para entender su estructura.
- Limpieza de datos necesaria para eliminar valores nulos, correcciones y seleccionar los tipos de datos adecuados.
- Análisis descriptivo para comprender los datos con la ayuda de técnicas básicas estadísticas.
- Visualizaciones para explorar relaciones en los datos con la finalidad de ayudar a interpretar los resultados.
- Documentación del proceso en Jupyter Notebooks utilizando celdas con el propósito de describir las acciones realizadas.
- Conclusiones de los resultados obtenidos.

c) Estrategias y/o técnicas

La ejecución del trabajo se fundamenta en consultas bibliográficas y en datos de observaciones astronómicas, complementados con información de diversos sitios web.

Estrategia:

Establecer la fuente de información y contestar al planteamiento.

Revisar el material bibliográfico para entender las técnicas y procesos de análisis de datos.

- Seleccionar y clasificar los datos recopilados.
- Colocar los datos de manera tabular y presentarla en forma gráfica si es necesario.

Objetivo:

- Recopilar, procesar y analizar información.
- Establecer las respectivas conclusiones de los datos y procesos analizados.

Acciones:

- Realizar una hoja de ruta, articular técnicas e implantar hitos.
- Recopilar información y guardarla dentro del repositorio local. Posteriormente se implementa el análisis.
- Categorizar y analizar los datos.
- Realizar cálculos que proporcionen sentido de los datos.
- Trasladar los datos numéricos en gráficos e interpretar los datos.
- Obtener conclusiones.

UNIVERSIDAD TECNOLÓGICA ISRAEL

ESCUELA DE POSGRADOS “ESPOG”

MAESTRÍA EN BIG DATA Y CIENCIA DE DATOS

INSTRUMENTO PARA VALIDACIÓN DE LA PROPUESTA

Estimado colega:

Se solicita su valiosa cooperación para evaluar la calidad del siguiente contenido digital “Análisis de parámetros físicos que clasifican las estrellas aplicado a una muestra del repositorio de observaciones espaciales Gaia”. Sus criterios son de suma importancia para la realización de este trabajo, por lo que se le pide que brinde su cooperación contestando las preguntas que se realizan a continuación.

Datos informativos

Validado por: GARCIA LLANOS CARMEN ISABEL

Título obtenido:

Ingeniera informática

Magister en Gestión de Sistemas de Información e Inteligencia de Negocios

C.I.: 0201340882

E-mail: cgarciall@hotmail.com

Institución de Trabajo: Contraloría General del Estado

Cargo: Administradora de Gestión de Proyectos

Años de experiencia en el área: 12



Instructivo:

- Responda cada criterio con la máxima sinceridad del caso.
- Revisar, observar y analizar la propuesta de la plataforma virtual, blog o sitio web.
- Coloque una X en cada indicador, tomando en cuenta que Muy adecuado equivale a 5, Bastante Adecuado equivale a 4, Adecuado equivale a 3, Poco Adecuado equivale a 2 e Inadecuado equivale a 1.

Tema: “Análisis de parámetros físicos que clasifican las estrellas aplicado a una muestra del repositorio de observaciones espaciales Gaia”

Indicadores	Muy adecuado	Bastante Adecuado	Adecuado	Poco adecuado	Inadecuado
Pertinencia	5				
Aplicabilidad	5				
Factibilidad	5				
Novedad	5				
Fundamentación pedagógica	5				
Fundamentación tecnológica	5				
Indicaciones para su uso	5				
TOTAL	35				

Observaciones: Ninguna

Recomendaciones: Ninguna

Lugar, fecha de validación: Quito 7 de marzo del 2025



Firmado electrónicamente por:
CARMEN ISABEL
GARCIA LLANOS

Firma del especialista
GARCIA LLANOS CARMEN ISABEL

UNIVERSIDAD TECNOLÓGICA ISRAEL

ESCUELA DE POSGRADOS “ESPOG”

MAESTRÍA EN BIG DATA Y CIENCIA DE DATOS

INSTRUMENTO PARA VALIDACIÓN DE LA PROPUESTA

Estimado colega:

Se solicita su valiosa cooperación para evaluar la calidad del siguiente contenido digital “Análisis de parámetros físicos que clasifican las estrellas aplicado a una muestra del repositorio de observaciones espaciales Gaia”. Sus criterios son de suma importancia para la realización de este trabajo, por lo que se le pide que brinde su cooperación contestando las preguntas que se realizan a continuación.

Datos informativos

Validado por: ANDRADE ALBAN JOSE RODOLFO

Título obtenido:

Ingeniero en Informática y Ciencias de La Computación

Master en Docencia

C.I.: 0502115587

E-mail: ocazo@hotmail.com

Institución de Trabajo: Corporación Nacional de Telecomunicaciones

Cargo: Analista de Facturación

Años de experiencia en el área: 20



Instructivo:

- Responda cada criterio con la máxima sinceridad del caso.
- Revisar, observar y analizar la propuesta de la plataforma virtual, blog o sitio web.
- Coloque una X en cada indicador, tomando en cuenta que Muy adecuado equivale a 5, Bastante Adecuado equivale a 4, Adecuado equivale a 3, Poco Adecuado equivale a 2 e Inadecuado equivale a 1.

Tema: “Análisis de parámetros físicos que clasifican las estrellas aplicado a una muestra del repositorio de observaciones espaciales Gaia”

Indicadores	Muy adecuado	Bastante Adecuado	Adecuado	Poco adecuado	Inadecuado
Pertinencia	5				
Aplicabilidad	5				
Factibilidad	5				
Novedad	5				
Fundamentación pedagógica	5				
Fundamentación tecnológica	5				
Indicaciones para su uso	5				
TOTAL	35				

Observaciones: Ninguna

Recomendaciones: Ninguna

Lugar, fecha de validación: Quito 7 de marzo del 2025



Firmado electrónicamente por:
**JOSE RODOLFO
ANDRADE ALBAN**

Firma del especialista
ANDRADE ALBAN JOSE RODOLFO