**Correlaciones entre exoplanetas y estrella anfitrionas descubiertas por el telescopio espacial Keppler.**

Edgardo Jara, Felipe Ovalle, Daniel Rojas , Sebastian Salgado

**Contexto y motivación:** En las últimas décadas se han descubierto miles de exoplanetas en distintos sistemas estelares, alcanzando casi 6.000 planetas confirmados a la fecha [[1]](https://exoplanetarchive.ipac.caltech.edu/docs/counts_detail.html#:~:text=Summary%20Counts). Esta abundancia de datos abre la oportunidad de analizar patrones en la formación de planetas de manera estadística. Un aspecto particularmente interesante es cómo las propiedades de la estrella anfitriona (como su tipo espectral, masa o temperatura) podrían influir en las características de sus planetas. Estudios previos sugieren, por ejemplo, que las estrellas pequeñas de tipo M (enanas rojas) prácticamente no albergan planetas de tamaño Neptuno, a diferencia de estrellas más masivas [[2]](https://arxiv.org/html/2404.11022v1#:~:text=Cloutier%20%26%20Menou%20,Neptunes%20at%20the%20faint%20end). Este tipo de hallazgo motiva profundizar en la relación entre la clase de estrella y los planetas que la orbitan. Comprender estas relaciones es valioso para la comunidad científica astronómica, ya que aporta información sobre los procesos de formación planetaria y puede guiar la búsqueda de nuevos mundos. La audiencia objetivo de este análisis son astrofísicos y astrónomos interesados en optimizar estrategias de búsqueda de exoplanetas (por ejemplo, decidir qué estrellas observar para encontrar cierto tipo de planeta) y en refinar modelos teóricos de formación de planetas en distintas condiciones estelares.

**Objetivos:** El objetivo principal del proyecto es investigar cómo las propiedades de las estrellas anfitrionas condicionan las características de sus exoplanetas. En particular, se busca cuantificar y describir patrones en masa y radio planetario según el tipo o clase de estrella. La pregunta científica de fondo es: ¿En qué medida el entorno estelar (por ejemplo, estrellas enanas rojas vs. estrellas tipo solar) afecta el tamaño, la masa u otras propiedades de los planetas que se forman? Para responderla, el proyecto apunta a: identificar correlaciones significativas entre variables estelares (masa, temperatura, tipo espectral) y variables planetarias (masa, radio, órbita), comprobar hipótesis específicas sobre diferencias en poblaciones de planetas según tipo de estrella (por ejemplo, ¿son menos comunes los planetas gigantes alrededor de estrellas de baja masa?), y si es posible, desarrollar modelos que estimen la probabilidad de cierto tipo de planeta en función de las características de la estrella.

**Datos:** Para llevar a cabo el análisis se utilizarán datos públicos del NASA Exoplanet Archive, una base consolidada de exoplanetas confirmados. En particular, trabajaremos con la tabla de parámetros compuestos (“Planetary Systems Composite Parameters”) del archivo, la cual compila en una sola fila por planeta tanto los parámetros estelares de su estrella anfitriona como los parámetros planetarios [[3]](https://exoplanetarchive.ipac.caltech.edu/docs/pscp_about.html#:~:text=About%20the%20Planetary%20Systems%20Composite,Planet%20Data%20Table). Este conjunto de datos incluye numerosas variables relevantes: por el lado de los planetas, se cuenta con masa, radio, período orbital, semieje mayor, densidad, entre otras propiedades; por el lado de las estrellas anfitrionas, se dispone de atributos como masa estelar, radio estelar, temperatura efectiva, luminosidad, metalicidad y

tipo espectral (cuando están disponibles). Todos estos datos son de tipo estructurado (formato tabular, por ejemplo, CSV), con valores numéricos (masas, radios, temperaturas) y categóricos (clasificación espectral, método de descubrimiento). El volumen de datos es moderado: en torno a 5.000–6.000 registros (cada registro corresponde a un planeta confirmado) junto con unas decenas de columnas descriptivas para cada uno. Dado que provienen de un repositorio público de la NASA, no existen restricciones de uso o publicación, y los datos se obtendrán mediante descarga directa desde el archivo (posiblemente aprovechando su API o un archivo CSV disponible). Antes de proceder al análisis, se realizará una depuración y revisión de calidad de los datos para manejar posibles valores faltantes o inconsistentes, ya que no todos los planetas tienen medidas completas en todas las variables.

Para el desarrollo inicial de este proyecto, y debido a restricciones en la descarga directa de la tabla oficial de la NASA, se optó por utilizar una copia íntegra del catálogo público de la *Exoplanet Encyclopaedia* (https://exoplanet.eu/catalog/csv/) y su espejo alojado en Île‑de‑France Smart Services. Este catálogo contiene una fila por planeta confirmado e incluye **98 columnas** con parámetros del planeta (masa, radio, densidad, período orbital, semieje mayor, excentricidad) y de la estrella anfitriona (masa, radio, temperatura efectiva, metalicidad, luminosidad, tipo espectral), así como metadatos del descubrimiento (año y método). El archivo descargado —que denominamos exoplaneteu\_catalog-1.csv— reúne aproximadamente **7 782** registros. Dado que este conjunto mantiene la misma estructura general que la tabla PSCompPars, permite explorar las relaciones estrella–planeta de forma análoga.

**Procedimiento de recolección y preparación.** Los datos se descargaron en formato CSV y se cargaron en un *notebook* de Python utilizando la librería *pandas*. Se revisaron los nombres de los planetas para descartar duplicados (no se encontraron) y se calcularon los porcentajes de valores faltantes por columna, observando, por ejemplo, que la masa planetaria está ausente en alrededor del **53 %** de los registros y el tipo espectral de la estrella en cerca del **48 %**. Se investigaron rangos físicos para detectar valores atípicos o inconsistentes y se documentó todo el proceso de limpieza.

Además, se derivaron variables auxiliares que facilitan el análisis estadístico:

* **spec\_class**: clase espectral abreviada obtenida a partir de la primera letra de star\_sp\_type, agrupando en categorías O, B, A, F, G, K, M y “Other”.
* **is\_giant**: indicador de planeta gigante gaseoso, definido como verdadero si la masa planetaria es ≥ 0,3 masas de Júpiter o si el radio planetario es ≥ 0,6 radios de Júpiter.
* **Categorías de planeta** (rocoso, sub‑neptuno, gigante) según el radio planetario, empleando umbrales habituales en la literatura (por ejemplo, < 1,6 radios terrestres para rocosos; 1,6–3,9 para sub‑neptunos; > 3,9 para gigantes).

Todas estas transformaciones y procedimientos quedaron documentados en el análisis exploratorio de datos (EDA) incluido en el repositorio. Una vez que se disponga de acceso completo a la tabla PSCompPars, se aplicarán los mismos pasos de limpieza y derivación para validar y ampliar los resultados preliminares.

Preguntas de investigación: A partir de estos datos, se proponen al menos 5 preguntas específicas que guiarán el análisis y permitirán explorar distintas facetas del problema:

1. Distribución de tamaños según tipo de estrella: ¿Cómo varía el radio medio de los exoplanetas en sistemas de enanas rojas tipo M en comparación con sistemas de estrellas tipo G (similares al Sol) u otros tipos estelares? Esta pregunta explorará si las estrellas más frías y pequeñas tienden a tener planetas significativamente más pequeños (más supertierras y menos gigantes gaseosos) que las estrellas más grandes.
2. Relación masa-masa (estrella-planeta): ¿Existe una correlación clara entre la masa de la estrella y la masa de sus planetas? Por ejemplo, se investigará si las estrellas más masivas tienden a albergar planetas más masivos (y viceversa), o si por el contrario hay un límite en masa planetaria independiente de la masa estelar. Esta relación ayuda a entender cómo la masa de la estrella podría influir en el proceso de acumulación de material planetario.
3. Frecuencia de planetas gigantes en distintas estrellas: ¿Cuál es la frecuencia de planetas tipo Júpiter (gigantes gaseosos de gran masa) alrededor de estrellas de diferente tipo espectral? Se comparará, por ejemplo, la proporción de estrellas tipo M que tienen Júpiteres calientes o fríos versus la proporción en estrellas tipo solar. Esta pregunta busca confirmar hipótesis como “las enanas rojas poseen menos planetas gigantes” y cuantificar esas diferencias.
4. Clasificación de planetas según contexto estelar: ¿Es posible clasificar el tipo de exoplaneta (por ejemplo, diferenciar si es de composición de estado rocoso o gaseoso) analizando las características de su estrella anfitriona? Aquí se intentará clasificar, según dado las características de la estrella. ¿podemos concluir información a partir de su estrella?
5. Cuántos planetas de cierto tipo se descubren por año: Por último, buscamos reordenar la información y estudiar qué exoplanetas son más comunes en este estudio local, donde pueden ser tantos gigantes gaseosos como supertierras.

Estas preguntas fueron refinándose a medida que avanzaba el análisis. Inicialmente, el foco estaba en la distribución de tamaños y en la relación entre la masa de la estrella y la de sus planetas; sin embargo, al explorar el catálogo y revisar la bibliografía se añadieron preguntas sobre la frecuencia de planetas gigantes por tipo estelar, la posible clasificación de planetas a partir de las características de la estrella y la evolución temporal de los descubrimientos por tipo de planeta.

**Diseño tentativo:** Para llevar a cabo este estudio se hará un análisis de distintos archivos CSV a través de Python para extraer la información, usaremos herramientas de diferentes librerías como Pandas y Numpy para filtrar, agrupar y comparar datos de las distintas fuentes e ir respondiendo las preguntas. Estos pasos serán fundamentales para centrar nuestro estudio en un nicho especifico, y sacar conclusiones en base a los parámetros que tomemos

**Análisis Exploratorio de Datos de Exoplanetas**

**Pregunta a responder**

El proyecto se propone investigar cómo las propiedades de las estrellas anfitrionas condicionan las características de sus exoplanetas. En particular se busca responder si el tipo espectral, la masa o el tamaño de la estrella influyen en el tamaño y la masa de los planetas, y con qué frecuencia aparecen planetas gigantes en diferentes clases estelares.

**Generalidades del dataset**

1. **Exoplanet Encyclopaedia — catálogo completo (CSV)**

**Enlace:** <https://exoplanet.eu/catalog/csv/>

**Qué contiene:** descarga íntegra del catálogo (una fila por planeta confirmado; opción “Download as CSV”).

**Para qué nos sirve:** base con parámetros planetarios y estelares para explorar relaciones estrella–planeta.

1. **NASA Exoplanet Archive — Planetary Systems Composite Parameters (PSCompPars)**

**Enlace (CSV vía TAP):** [https://exoplanetarchive.ipac.caltech.edu/TAP/sync?query=select%20\*%20from%20pscomppars&format=csv](https://exoplanetarchive.ipac.caltech.edu/TAP/sync?query=select%20*%20from%20pscomppars&format=csv)

**Qué contiene:** tabla compuesta (una fila por planeta) con parámetros del sistema, estrella y planeta.

**Para qué nos sirve:** referencia oficial para validar/extender resultados y futuras integraciones.

1. **Île‑de‑France Smart Services — mirror del catálogo Exoplanet.eu**

**Enlace:** <https://data.smartidf.services/explore/dataset/exoplaneteu_catalog-1/download/?format=csv>

**Qué contiene:** copia sincronizada del catálogo de la Exoplanet Encyclopaedia (mismas columnas/registros).

**Para qué nos sirve:** alternativa estable cuando la descarga directa falla; con este mirror generamos el CSV del EDA.

**Fuente principal:** Usaremos el NASA Exoplanet Archive – Planetary Systems Composite Parameters (PSCompPars), una tabla compuesta que reúne parámetros del sistema, de la estrella y del planeta en una fila por planeta confirmado, pensada para análisis estadísticos de poblaciones. Cuando falta un parámetro en la referencia por defecto, la tabla selecciona automáticamente valores de otras referencias disponibles (lo que maximiza cobertura, aunque no garantiza coherencia interna a nivel de sistema).

* Granularidad: 1 fila = 1 planeta confirmado.
* Número de filas y columnas: 7 782 planetas confirmados y 98 variables numéricas y categóricas.
* Principales variables: mass (masa del planeta, en masas de Júpiter), radius (radio del planeta, en radios de Júpiter), star\_mass (masa estelar, en masas solares), star\_radius (radio estelar, en radios solares), semi\_major\_axis (semieje mayor orbital).
* Fuente: catálogo público de la Exoplanet Encyclopaedia, similar a la tabla PSCompPars del NASA Exoplanet Archive.
* Valores faltantes: alrededor del 53 % de los planetas carece de masa, el 34 % carece de radio, el 47,8 % no tiene tipo espectral; no se detectaron nombres duplicados.
* Columnas (ejemplos):
  + Identificadores y sistema: pl\_name, hostname, número de planetas y de estrellas.
  + Parámetros del planeta: masas y radios en distintas unidades, período orbital, semieje mayor, excentricidad, inclinación, densidad, temperatura de equilibrio, indicadores de detección (tránsito, VR, etc.).
  + Parámetros de la estrella: masa, radio, temperatura efectiva, metalicidad, luminosidad, tipo espectral; además de astrometría (distancia, paralaje, movimiento propio).
  + Metadatos de descubrimiento: método, año, referencia, instalación/instrumento.
  + Tipos de dato: numéricos continuos (masas, radios, periodos), categóricos (tipo espectral, método), fechas/años y banderas. [Archivo de Exoplanetas de NASA](https://exoplanetarchive.ipac.caltech.edu/docs/API_PS_columns.html?utm_source=chatgpt.com)

**Origen y recolección.**

* Descarga reproducible vía TAP/API (ADQL o endpoints REST), p. ej., select \* from pscomppars en CSV/JSON; el archivo mantiene recuentos oficiales y documentación pública de columnas.

**Tipos de dato**

El conjunto incluye **84 columnas numéricas** y **14 categóricas**. Las variables numéricas abarcan masas, radios, periodos, semiejes y otros parámetros físicos; las categóricas incluyen nombres (name), tipo espectral de la estrella (star\_sp\_type) y método de descubrimiento.

Se derivaron dos variables auxiliares para este análisis:

* **spec\_class**: primera letra del tipo espectral agrupada en clases A, F, G, K, M y O; otros valores se agrupan en “Other”.
* **is\_giant**: indicador binario que marca un planeta como gigante si su masa ≥ 0,3 M\_J o su radio ≥ 0,6 R\_J.

**Estadística descriptiva**

Las principales variables numéricas presentan las siguientes estadísticas resumidas:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Count | Mínimo | Mediana | Media | Máximo |
| Masa del planeta (mass) | 3 663 | 1.42e‑09 M\_J | 2.50 M\_J | 14.94 M\_J | 78.27 M\_J |
| Radio del planeta (radius) | 5 153 | 0.00107 R\_J | 0.236 R\_J | 0.542 R\_J | 18.80 R\_J |
| Masa estelar (star\_mass) | 6 488 | 0.011 M\_⊙ | 0.94 M\_⊙ | 0.932 M\_⊙ | 23.47 M\_⊙ |
| Radio estelar (star\_radius) | 6 044 | 1.43e‑05 R\_⊙ | 0.95 R\_⊙ | 1.49 R\_⊙ | 91.40 R\_⊙ |
| Semieje mayor (semi\_major\_axis) | 5 427 | 0.00087 AU | 0.1201 AU | 196.10 AU | 80 679 AU |

Estas cifras revelan la gran dispersión de las masas planetarias (de microplanetas hasta gigantes de 78 M\_J) y el amplio rango de órbitas, desde planetas ultra‑cercanos (0,00087 AU) hasta cuerpos extremadamente distantes.

M⊙: masa solar (unidad para comparar masas estelares); ≈ 1,988 × 10³⁰ kg.

R⊙: radio solar (unidad para tamaños estelares); la IAU adoptó valores nominales de referencia.

L⊙: luminosidad solar (unidad para potencias radiadas); con valor nominal recomendado por la IAU.

AU (au): unidad astronómica (longitud); exactamente 149 597 870 700 m (IAU 2012, Res. B2).

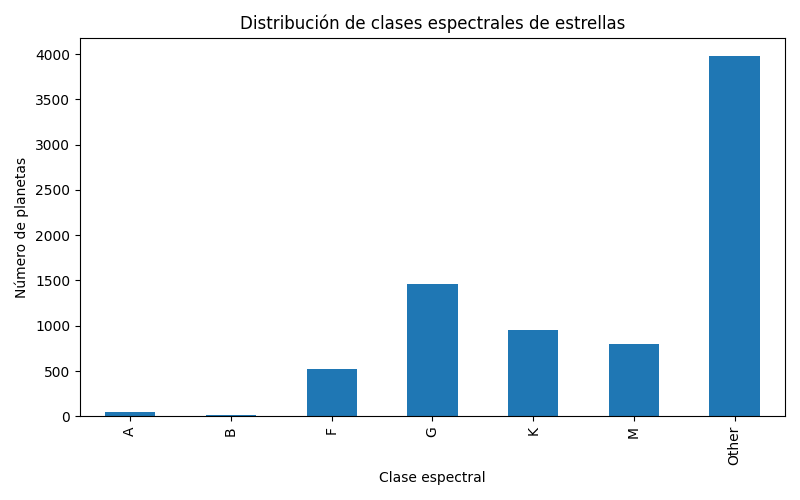
R\_J: radio de Júpiter (unidad para radios planetarios); referencia típica ≈ 69 911 km (radio medio).

M\_J: masa de Júpiter (unidad para masas planetarias); ≈ 1,898 × 10²⁷ kg (JPL/SSD).

**Visualización**

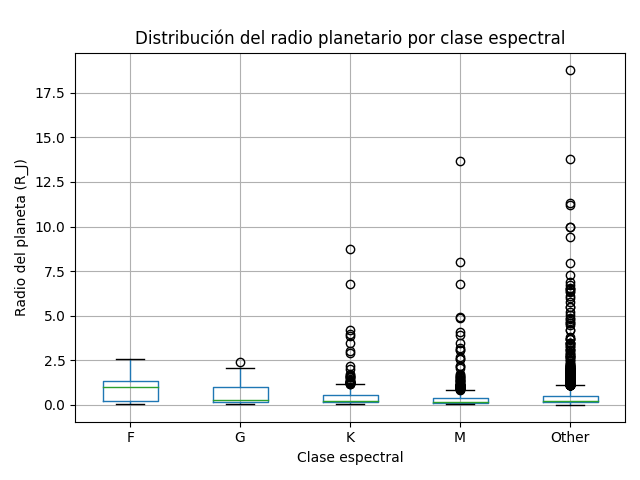
**Distribución de clases espectrales**

Las clases estelares **G** y **K** son las más comunes, seguidas de **M** y **F**. Casi la mitad de los registros corresponden a estrellas sin clasificación específica o poco usual (“Other”).



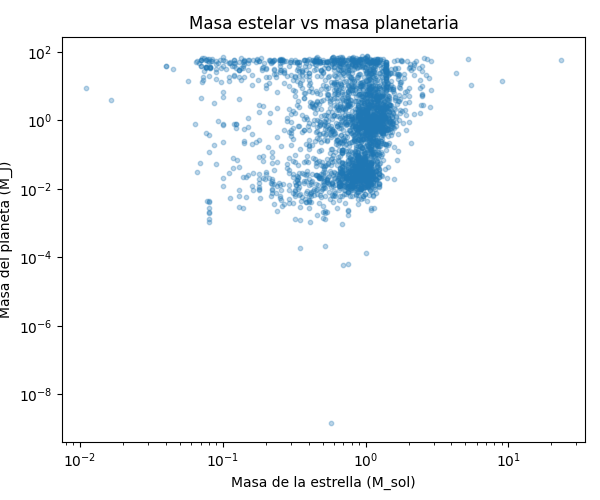
**Tamaño planetario según clase**

Un diagrama de caja muestra que las estrellas de tipo **F** y **G** albergan planetas más grandes (medianas ~1,0 y 0,26 R\_J), mientras que las de tipo **K** y **M** hospedan planetas más pequeños (~0,22 y 0,18 R\_J).



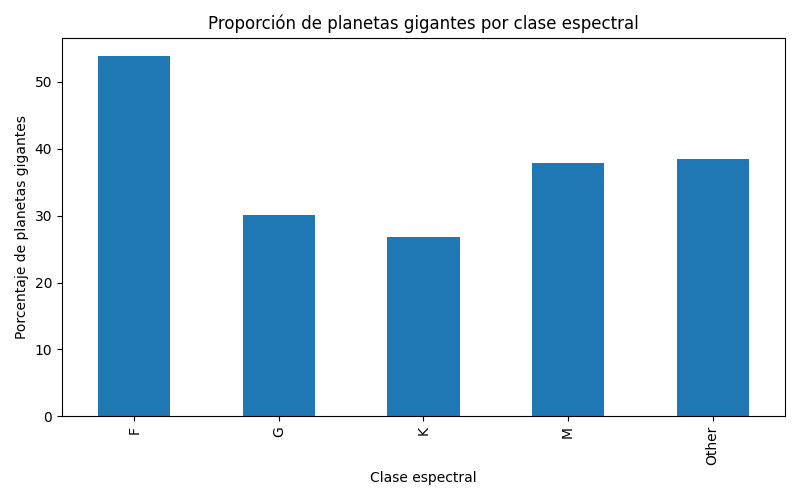
**Relación masa estelar–masa planetaria**

El diagrama de dispersión revela que la correlación entre masa estelar y masa planetaria es baja (coeficiente ~−0,05). Las estrellas de mayor masa no muestran tendencia clara a albergar planetas más masivos.



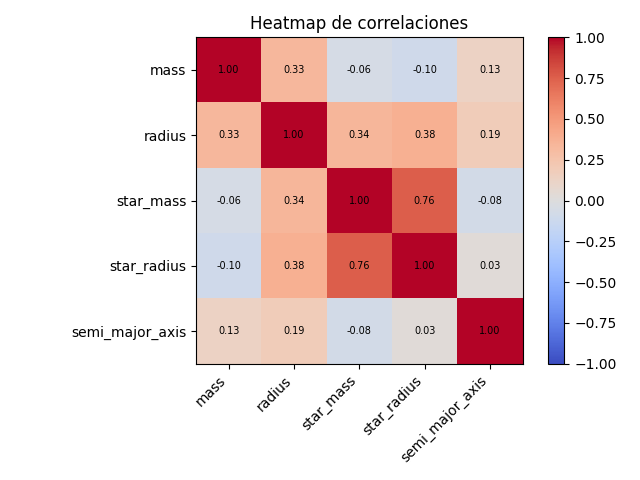
**Frecuencia de planetas gigantes**

Las estrellas **F** presentan la mayor fracción de planetas gigantes (~54 %), mientras que **G** y **K** tienen proporciones menores (~30 % y 27 %). Las enanas **M** también muestran una fracción notable de gigantes (~38 %).



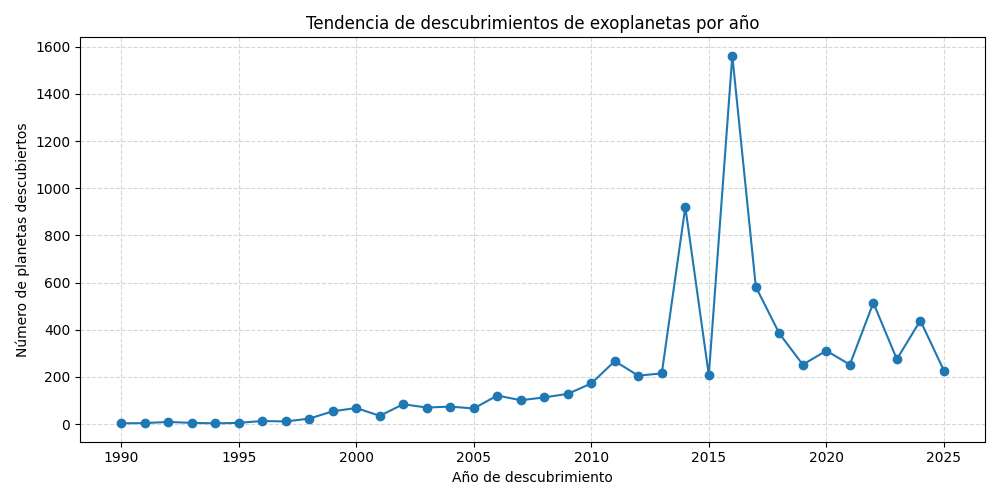
**Correlaciones**

El mapa de calor indica que el radio planetario se correlaciona moderadamente con el radio estelar (0,38) y la masa estelar (0,34). La masa del planeta se correlaciona débilmente con el radio y el semieje mayor.



**Tendencia temporal**

El número de descubrimientos de exoplanetas aumenta bruscamente entre 2014 y 2016 y se estabiliza alrededor de 400–500 descubrimientos anuales en los últimos años.



**Interacciones**

* **Masa y radio estelar vs. variables planetarias:** El radio del planeta muestra una relación moderada con el radio estelar y la masa estelar, indicando que estrellas más grandes tienden a albergar planetas de mayor tamaño. La correlación de masas es muy débil, sugiriendo que la masa de la estrella no determina directamente la masa del planeta.
* **Clase espectral y distribución de tamaños:** Las clases **F** y **G** destacan por tener mayor proporción de planetas gigantes y radios planetarios más altos. Las clases **K** y **M** concentran supertierras y mini‑neptunos, aunque todavía presentan algunos gigantes.
* **Frecuencia de gigantes:** La comparación de frecuencias confirma que las enanas rojas (tipo **M**) tienen menos gigantes gaseosos en porcentaje, pero no los excluyen, matizando la hipótesis de que estas estrellas no albergan Júpiteres calientes.

**Sumarización**

El análisis exploratorio de este catálogo de exoplanetas muestra que la distribución de tamaños y masas planetarias depende fuertemente del tipo espectral y del tamaño de la estrella anfitriona, mientras que la masa estelar se correlaciona poco con la masa planetaria. Las estrellas de tipo **F** y **G** albergan una mayor proporción de planetas gigantes, en contraste con las clases **K** y **M**. El alto porcentaje de valores faltantes sugiere la necesidad de imputación y depuración antes de modelar. Como siguientes pasos se recomienda profundizar en modelos predictivos que incorporen metalicidad y edad estelar, y comparar los resultados con la tabla **PSCompPars** del NASA Exoplanet Archive para validar la robustez de estos hallazgos.

**Referencias:**

* NASA Exoplanet Archive – Exoplanet and Candidate Statistics (consulta de estadísticas de exoplanetas confirmados) (<https://exoplanetarchive.ipac.caltech.edu/docs/counts_detail.html>) [[1]](https://exoplanetarchive.ipac.caltech.edu/docs/counts_detail.html#:~:text=Summary%20Counts).
* NASA Exoplanet Archive – Planetary Systems Composite Data Description (documentación de la tabla de parámetros compuestos utilizada) (<https://exoplanetarchive.ipac.caltech.edu/docs/pscp_about.html>) [[2]](https://exoplanetarchive.ipac.caltech.edu/docs/pscp_about.html#:~:text=About%20the%20Planetary%20Systems%20Composite,Planet%20Data%20Table).