

Startups que han llegado a convertirse en unicornio

Memoria EDA

Pablo Hortal Llera



PRIMER ANÁLISIS DE LOS DATOS

- Primero de todo importamos las librerías que vamos a usar seguro.
- En segundo lugar, importamos el CSV con el que vamos a trabajar.
- A continuación, usamos `"unicornios.head(10)"` para mostrar las primeras 10 filas del DataFrame `"unicornios"`. Es una forma rápida de obtener una vista previa de los datos y ver cómo están estructurados.
- Después usamos `"unicornios.shape"` para obtener la forma (número de filas y columnas) de una matriz o dataframe. En el caso específico de `unicornios.shape`, devuelve una tupla con dos valores: el número de filas y columnas en el dataframe `unicornios`. El primer valor corresponde al número de filas y el segundo al número de columnas.
- El método `"info()"` se utiliza para obtener información sobre un DataFrame, incluyendo el número de filas y columnas, el tipo de datos de cada columna, el uso de memoria, entre otros. En el caso específico de `unicornios.info()`, se mostrará información sobre el DataFrame `unicornios`, como el número de filas, el número de columnas, los nombres de las columnas, la cantidad de valores no nulos en cada columna y su tipo de dato correspondiente
- `"sns.pairplot()"` es una función de Seaborn que permite crear gráficos de dispersión (scatter plots) e histogramas para todas las combinaciones de variables numéricas en un DataFrame. Es útil para visualizar la relación entre pares de variables y detectar patrones o correlaciones.

Análisis univariante:

1 - Comenzamos con un código se utiliza para crear un histograma de la variable `"Valuation"` en el dataset de empresas unicornio para ver si sigue una distribución normal o no. Al utilizar el parámetro `"kde=True"`, también se agrega una estimación de la densidad de la distribución de la variable. Al ajustar el parámetro `"binwidth=5"`, se establece el ancho de los intervalos del histograma en 5 unidades, lo que ayuda a que los datos se visualicen de manera más clara y con una mejor distribución en el eje x. Podemos ver que no sigue una distribución normal, ya que la mayoría de los unicornios aparecen reflejados en la primera sección del histograma.

Más abajo, vemos un código realiza que una prueba estadística llamada "Prueba D'Agostino-Pearson". La función `normaltest()` de `scipy.stats` se utiliza para realizar dicha

prueba y devuelve dos valores: el estadístico de la prueba y el valor p.

Aunque el histograma anterior deja bastante claro que no, hago la prueba D'Agostino-Pearson. Esta muestra que hay evidencia suficiente para rechazar la hipótesis nula de que los datos siguen una distribución normal. Esto significa que los datos no se ajustan a una distribución gaussiana (como se puede ver en el histograma de arriba) y se debe tener precaución al aplicar técnicas estadísticas que asumen una distribución normal.

2- Para la variable "Valuation", hago un histograma y calculo su media y desviación estándar para tener una idea de la distribución de los valores y la dispersión de los mismos.

Para ello genero un histograma de la columna "Valuation", que representa la valoración de las empresas unicornio. El histograma tiene 20 barras y los ejes x e y están etiquetados con "Valoración" y "Frecuencia", respectivamente. El título del gráfico es "Distribución de la valoración de las empresas unicornio".

Luego, se calcula la media y la desviación estándar de la columna "Valuation" utilizando las funciones "np.mean" y "np.std" de la biblioteca NumPy. La media es el valor promedio de las valoraciones de las empresas unicornio en el DataFrame, mientras que la desviación estándar mide cuánto varían las valoraciones con respecto a la media. Estos valores pueden ser útiles para entender mejor cómo se distribuyen los datos.

La media obtenida de 3.28 indica que el valor promedio de las valoraciones de las empresas unicornio en el dataset es de 3.28.

La desviación estándar obtenida de 7.47 indica que los valores de las valoraciones están dispersos en un rango de 7.47 alrededor de la media. Una desviación estándar alta indica que hay una gran variabilidad en los datos. En este caso, una desviación estándar de 7.47 sugiere que hay una amplia gama de valoraciones en las empresas unicornio del dataset, lo que puede ser útil para identificar las empresas más valiosas, por ejemplo.

3- Ahora vamos a calcular las frecuencias de cada sector de industria. Ya que nos permitirá tener una idea de cuáles son los sectores más comunes en las empresas unicornio y, por lo tanto, identificar las áreas de negocio más exitosas en este tipo de empresas. Esta información puede ser de utilidad para inversores, emprendedores y otras partes interesadas en la industria de startups y tecnología. Además, también nos permite tener una mejor comprensión del panorama empresarial actual en términos de sectores y tendencias emergentes.

Primero se ha utilizado la función `.replace()` para reemplazar ciertas cadenas de caracteres en la columna 'Industry'. Luego se ha agrupado el DataFrame por sector de industria usando la función `groupby()` y se ha contado la frecuencia de cada sector usando la función `size()`. El resultado se ha almacenado en un nuevo DataFrame llamado `frecuencia_industrias`. Después se ha ordenado este DataFrame de mayor a menor frecuencia usando la función `.sort_values()`.

Finalmente, se ha graficado un histograma de barras usando la función `plt.bar()`, especificando las columnas del DataFrame como los datos para las barras y añadiendo etiquetas para los ejes y el título usando las funciones `plt.xlabel()`, `plt.ylabel()` y `plt.title()`. También se ha girado el eje x para que las etiquetas de los sectores de industria se puedan leer más fácilmente con la función `plt.xticks()`.

A golpe de vista vemos que Fintech, Ecommerce e Internet Software Services son los sectores de industria más comunes entre las empresas unicornio, por lo que podemos intuir que estos sectores tienen una alta probabilidad de éxito y de generar altas valoraciones en el mercado actual. Esto podría deberse a varios factores, como la creciente importancia de la tecnología en el mundo de los negocios y la creciente demanda de soluciones financieras y de comercio electrónico en línea. Además, estos sectores podrían estar atrayendo una mayor inversión de capital de riesgo debido a su potencial de crecimiento y rentabilidad.

Análisis Bivariante:

1- Primero uso El método “`info()`” se utiliza para obtener información sobre un DataFrame, incluyendo el número de filas y columnas, el tipo de datos de cada columna, el uso de memoria, entre otros. En el caso específico de `unicornios.info()`, se mostrará información sobre el DataFrame `unicornios`, como el número de filas, el número de columnas, los nombres de las columnas, la cantidad de valores no nulos en cada columna y su tipo de dato correspondiente.

2 - Para analizar la relación entre la valoración de las empresas unicornio y el sector de industria al que pertenecen, he agrupado los datos por sector de industria utilizando el método `groupby` de Pandas y he utilizado la función `agg` para calcular la cuenta de empresas por sector, la media de valoración por sector y la suma de valoración por sector. Luego he ordenado los resultados por valor medio en orden descendente utilizando el método `sort_values`.

El análisis bivariante muestra el conteo, la media y la suma de la columna 'Valuation' agrupada por la columna 'Industry'. Los resultados se ordenan en función de la media de las valoraciones agrupadas por la industria, en orden descendente.

El resultado, como digo, es una tabla que muestra el conteo de empresas en cada sector de la industria, la media de la valoración y la suma total de la valoración en cada sector.

Al analizar los resultados, podemos sacar varias conclusiones:

Los sectores "Other" y "Artificial Intelligence" tienen los valores más altos de media y suma de valuación.

El sector "Fintech" tiene el mayor número de empresas unicornio, con un total de 79.

Los sectores con menor media y suma de valuación son "Mobile & telecommunications" y "Cybersecurity", respectivamente.

Los sectores con menor cantidad de empresas unicornio son "Travel" y "Edtech", con solo 6 y 9 respectivamente.

En general, podemos ver que los sectores tecnológicos son los que tienen más presencia en la lista de empresas unicornio, y los sectores financieros y de inteligencia artificial son los que tienen los valores más altos.

3 - A partir del gráfico de dispersión que he creado usando la función scatter, es posible observar que la mayoría de las empresas unicornio tienen una valoración menor a 30 billones de dólares, y que la mayoría de ellas fueron fundadas en los últimos 10 años. También se puede ver que hay una cantidad significativa de empresas unicornio entre los años 2012-2016 y que tienen una valoración entre 0 y 10 billones de dólares. Además, se puede notar una dispersión mayor en las empresas fundadas hace más de 10 años, lo que puede ser indicativo de una mayor variabilidad en las valoraciones de las empresas más antiguas.

Análisis Multivariante:

1 - Lo primero que hice fue calcular la matriz de correlación entre las variables.

Primero, he creado una matriz de correlación para las variables 'Valuation', 'year', 'month' y 'day' utilizando el método 'corr()'.

Este resultado es una matriz de correlación que muestra la relación entre cada par de variables en el conjunto de datos de unicornios.

2- Primero, he eliminado la columna "mes_numero" del DataFrame "unicornios" usando el método "drop" con el argumento "axis=1". Luego, he creado una matriz de correlación para las variables restantes usando el método "corr" del DataFrame "unicornios_sin_mes". Finalmente, he utilizado la función "heatmap" de la biblioteca

Seaborn para visualizar la matriz de correlación en un mapa de calor con etiquetas de anotación activadas y una paleta de colores "coolwarm".

Para interpretar los resultados, debemos tener en cuenta los siguientes puntos:

Los coeficientes de correlación varían entre -1 y 1. Un valor de 1 indica una correlación positiva perfecta, un valor de -1 indica una correlación negativa perfecta y un valor de 0 indica que no hay correlación entre las variables.

Las correlaciones más fuertes (ya sean positivas o negativas) se observan en las celdas que tienen valores más cercanos a -1 o 1. Los valores cercanos a 0 indican una correlación débil o inexistente.

A partir de los resultados, se pueden obtener las siguientes conclusiones:

La correlación entre "Valuation" y "year" es negativa y moderada (-0.263822). Esto indica que a medida que aumentan los años, la valoración tiende a disminuir. Sin embargo, la correlación no es muy fuerte, por lo que no se puede establecer una relación clara entre estas dos variables.

Las correlaciones entre "Valuation" y "month" (-0.034605), "Valuation" y "day" (-0.023996), "year" y "month" (-0.009743), "year" y "day" (-0.024289), y "month" y "day" (-0.037771) son todas muy bajas y cercanas a 0. Esto sugiere que no hay relaciones lineales significativas entre estos pares de variables.

En resumen, el mapa de calor muestra una correlación negativa moderada entre la valoración y el año, pero las otras variables no parecen tener relaciones lineales significativas entre sí.

Hipótesis 1:

He utilizado ese código para crear un gráfico de línea que muestra la cantidad de startups unicornio por año. Primero, he agrupado los datos por año y he contado la cantidad de unicornios en cada año. Luego, he creado un gráfico con el eje X representando los años y el eje Y representando la cantidad de startups unicornio. La línea del gráfico muestra cómo ha ido variando el número de startups unicornio a lo largo de los años. La visualización me ha permitido confirmar que sí que ha habido una tendencia positiva en el número de unicornios desde el año 2020.

Más abajo, he creado un gráfico de barras que muestra la cantidad de startups unicornio

por año. Para esto, primero he agrupado los datos por año y he contado la cantidad de unicornios por año. Luego, he utilizado el método "bar" de matplotlib para crear el gráfico de barras, especificando el eje X como los años y el eje Y como la cantidad de startups unicornio. He añadido un título, etiquetas para los ejes X e Y y he mostrado el gráfico con el método "show" de matplotlib. De esta manera, también podemos visualizar fácilmente cómo ha evolucionado el número de startups unicornio a lo largo de los años.

Hipótesis 2:

Para saber si hay una relación entre la ubicación geográfica y el éxito de la startup he hecho lo siguiente: Primero, creamos una nueva columna llamada 'Successful' en el DataFrame 'unicornios'. En esta columna, asignamos el valor 1 a las filas donde el valor de la columna 'Valuation' es mayor que la mediana de la columna 'Valuation' del DataFrame, y el valor 0 en caso contrario.

Luego, creamos una tabla de contingencia con la función `pd.crosstab()` para mostrar la frecuencia de aparición de los valores de las columnas 'Country' y 'Successful' (Las tablas de contingencia se utilizan para analizar la relación entre dos variables categóricas y para evaluar la dependencia o independencia entre ellas mediante pruebas estadísticas como la prueba de chi cuadrado). Es decir, contamos cuántas startups tuvieron éxito (valor 1 en la columna 'Successful') y cuántas no lo tuvieron (valor 0 en la columna 'Successful') para cada país.

Importamos la función `chi2_contingency` de la librería `scipy.stats` para realizar el test de chi-cuadrado, el cual nos permite evaluar si hay una relación significativa entre las variables 'Country' y 'Successful'.

El resultado del test es una tupla con cuatro valores, pero solo nos interesa el segundo valor, que es el valor p. Asignamos los otros valores a las variables " _ " porque no los utilizaremos.

Finalmente, imprimimos el valor p obtenido en el paso anterior con la haciendo un `print()`.

Si este valor es muy pequeño (por ejemplo, menor que 0.05), concluimos que hay una relación significativa entre las variables 'Country' y 'Successful'.

Conclusión: Según el análisis realizado mediante la prueba de chi cuadrado, se obtuvo un valor p de 0.35. Esto indica que no hay suficiente evidencia para rechazar la hipótesis nula de que no hay relación significativa entre la ubicación geográfica y el éxito de una startup. Por lo tanto, no se puede afirmar que la ubicación geográfica sea un factor

determinante en el éxito de una startup. Sin embargo, cabe destacar que este resultado debe ser interpretado con precaución, ya que existen otros factores que pueden influir en el éxito de una startup y no han sido analizados en este caso (PIB, acceso a inversión, seguridad jurídica, etc).

A continuación, creí que podría ser interesante ver de una forma más visual lo anteriormente analizado, por eso vemos una tabla de contingencia creada con "pd.crosstab", en ella se cuenta la cantidad de startups exitosas y no exitosas para cada país. Por lo tanto, en el gráfico de barras apiladas, las barras de color azul corresponden a la cantidad de startups no exitosas y las barras de color naranja corresponden a la cantidad de startups exitosas para cada país.

Cada barra representa un país y está dividida en dos secciones: la sección inferior representa la cantidad de startups no exitosas y la sección superior representa la cantidad de startups exitosas. La altura total de cada barra representa la cantidad total de startups para ese país. En el código se crea la columna "Successful" que indica si la startup fue exitosa o no. Para esto, se compara el valor de la columna "Valuation" de cada startup con la mediana de los valores de "Valuation" de todas las startups. Si el valor de "Valuation" de una startup es mayor que la mediana, se considera que la startup fue exitosa y se asigna el valor 1 en la columna "Successful". En caso contrario, se asigna el valor 0.

"Exitosa" y "No Exitosa" la forma que se me ha ocurrido para poder ver de una forma más visual las startups más valiosas y menos valiosas dependiendo del país.

Conclusión: Volviendo al gráfico de barras apiladas, podemos observar que Estados Unidos tiene la mayor cantidad de startups exitosas y no exitosas en comparación con otros países, lo que sugiere que es un mercado importante para las startups. También podemos ver que algunos países tienen una proporción más alta de startups exitosas en comparación con otras, lo que puede ser interesante para futuros análisis en los que incluyamos más variables.

Hipótesis 3:

Para probar si existen diferencias en la valoración promedio de las startups unicornio según el país en el que se encuentran, primero, agrupé los datos por país y calculé el promedio de la valoración para cada uno de ellos, seleccionando los 10 países con la mayor valoración promedio. Luego, usé el método "sns.barplot()" de Seaborn para crear una gráfica de barras que muestra la valoración promedio de las startups unicornio para los 10 países seleccionados. También cambié el nombre de la columna de valoración para que fuera más descriptiva.

Conclusión: La gráfica resultante muestra claramente las diferencias en la valoración promedio entre los 10 países seleccionados, lo que nos permite observar que existe una gran variabilidad en la valoración de estas empresas según su país de origen. Esto sugiere que hay diferencias significativas en la valoración de startups unicornio según su ubicación geográfica, lo que podría estar relacionado con factores como los que mencionamos en hipótesis anteriores y también por otros como el acceso al capital, el tamaño del mercado y la competencia en la industria.

Aclarar que la valoración promedio más alta se refiere a la media aritmética de las valoraciones de las startups unicornio en un determinado país. Es decir, si la valoración promedio de un país es mayor que la de otros países, significa que en promedio las startups unicornio en ese país tienen una valoración más alta que las de otros países.

Sabemos que Estados Unidos tiene muchas más empresas unicornio que Suecia, pero que en promedio las empresas de Suecia tienen una valoración más alta, esto puede ser porque:

- **Concentración de unicornios:** Estados Unidos podría tener un ecosistema empresarial y de inversión más desarrollado, lo que permite la creación de un mayor número de empresas unicornio. Esto puede deberse a factores como el tamaño del mercado, la disponibilidad de capital de riesgo, la infraestructura tecnológica y el acceso a talento.
- **Calidad vs. cantidad:** Aunque Suecia tenga menos empresas unicornio en comparación con Estados Unidos, el hecho de que en promedio tengan una valoración más alta podría indicar que estas empresas suecas se centran en la calidad, la innovación y el crecimiento sostenible. Esto podría ser un reflejo de las políticas gubernamentales, el apoyo a la investigación y el desarrollo, y una cultura empresarial que fomente la calidad en lugar de la cantidad.
- **Sectores específicos:** Hemos visto en el análisis univariante que Fintech e Internet software & services son unas de las industrias que aglutinan más unicornios, a su vez es el lugar en el que se localizan los unicornios suecos. Por este motivo podemos decir que están concentradas en sectores o industrias específicas que, en promedio, tienden a tener valoraciones más altas.
- **Diversidad del ecosistema empresarial:** La diferencia en la cantidad de unicornios y las valoraciones promedio entre Estados Unidos y Suecia refleja la diversidad en sus respectivos ecosistemas empresariales. Estados Unidos tiene una mayor variedad de industrias y tipos de empresas, lo que resulta en un mayor número de unicornios pero con valoraciones promedio más bajas.

Hipótesis 4:

En esta hipótesis, lo que quise probar es si existen ciudades específicas donde se concentra una cantidad significativa de startups unicornio respecto a otras. Para eso he hecho este código que muestre las 20 ciudades con más startups unicornio. Primero he contado la cantidad de unicornios por ciudad utilizando el método "groupby" y seleccionando las 20 ciudades con más unicornios (el método "count()") cuenta la cantidad de valores no nulos, y el método "sort_values()" los ordena en orden descendente. Luego, he creado un gráfico de barras utilizando el método bar y he personalizado el título, los ejes y la rotación de las etiquetas de las ciudades en el eje x.

Conclusión: El resultado es un gráfico que me muestra claramente las 20 ciudades con más startups unicornio y la cantidad de ellas en cada ciudad.

Podemos confirmar que sí hay ciudades en las que se concentran una cantidad significativa de startups unicornio. En concreto San Francisco es una ciudad que tiene un ecosistema emprendedor muy desarrollado y atractivo para la creación y crecimiento de startups. Es posible que esta ciudad ofrezca una serie de ventajas y recursos que no están disponibles en otras ciudades, lo que puede explicar por qué hay una mayor concentración de startups unicornio allí. Además, esta concentración también puede atraer a más inversores y talentos en el sector de tecnología y emprendimiento, lo que a su vez puede aumentar la probabilidad de que surjan más unicornios en el futuro en esa ciudad.

Para indagar un poco más, he querido saber no solo en qué ciudades hay más unicornios, sino si se concentran dentro de las industrias más potentes.

El siguiente código filtra y analiza las tres ciudades con más unicornios en las industrias de Inteligencia Artificial, Automoción y Transporte, y Fintech:

`industrias_interes = ["Artificial Intelligence", "Auto & transportation", "Fintech"]`: Aquí, defino una lista llamada `industrias_interes` que contiene las tres industrias en las que estoy interesado.

`data_filtrada = unicornios[unicornios['Industry'].isin(industrias_interes)]`: Con este comando, filtro el conjunto de datos original `unicornios` para seleccionar solo aquellas filas donde la columna 'Industry' está dentro de la lista `industrias_interes`. Almaceno el resultado en una nueva variable llamada `data_filtrada`.

`conteo_ciudades = data_filtrada.groupby("City")["Company"].count()`: Utilizo la función `groupby` para agrupar el conjunto de datos filtrado `data_filtrada` por la columna 'City'. Luego, selecciono la columna 'Company' y cuento el número de empresas unicornio en cada ciudad utilizando la función `count()`. Guardo el resultado en una variable llamada `conteo_ciudades`.

`tres_principales_ciudades = conteo_ciudades.sort_values(ascending=False).head(3)`: Ordeno los valores de `conteo_ciudades` de forma descendente usando la función `sort_values()` con el argumento `ascending=False`. Luego, selecciono las tres primeras filas (las tres ciudades con más unicornios) usando la función `head(3)` y guardo el resultado en la variable `tres_principales_ciudades`.

Finalmente, imprimo las tres ciudades principales y su número de unicornios en las industrias seleccionadas:

En este bloque de código, utilizo un bucle `for` para iterar sobre los elementos de `tres_principales_ciudades` y su índice correspondiente. La función `enumerate()` me permite obtener el índice y el elemento (ciudad y conteo) al mismo tiempo. Imprimo el índice (empezando desde 1), la ciudad y el número de unicornios en cada una de las tres ciudades principales.

Conclusión: A partir de esta observación, podemos sacar varias conclusiones sobre la concentración de empresas en San Francisco, Nueva York y Londres:

- **Centros de innovación y tecnología:** Estas ciudades son reconocidas mundialmente como centros de innovación y tecnología, lo que atrae a empresas emergentes y consolidadas. San Francisco, en particular, es el hogar de Silicon Valley, mientras que Nueva York y Londres son importantes centros financieros y tecnológicos.
- **Acceso a capital:** La presencia de importantes inversores y fondos de capital riesgo en estas ciudades facilita a las empresas el acceso a financiamiento para crecer y expandirse.
- **Talento especializado:** La concentración de empresas en estas ciudades también atrae a profesionales altamente calificados, lo que crea un ecosistema de talento que beneficia a todas las compañías en la zona.
- **Redes de negocios y colaboración:** La cercanía geográfica de las empresas en estas ciudades fomenta la colaboración y la creación de redes de negocios, lo que puede acelerar el crecimiento y la innovación en diversos sectores.

- **Infraestructura y servicios:** Estas ciudades ofrecen infraestructuras y servicios de alta calidad, como transporte público eficiente, instalaciones de investigación y desarrollo, y una amplia gama de servicios profesionales que pueden ser útiles para las empresas.
- **Calidad de vida:** San Francisco, Nueva York y Londres son ciudades con una alta calidad de vida, lo que atrae a profesionales y emprendedores de todo el mundo.

En resumen, la concentración de empresas en San Francisco, Nueva York y Londres se debe a factores como la presencia de centros de innovación y tecnología, acceso a capital, talento especializado, redes de negocios, infraestructura y servicios, y calidad de vida. Estos factores crean un ambiente propicio para el crecimiento y éxito de las empresas en estas ciudades.

Hipótesis 5:

Por último, para probar si hay patrones estacionales en el momento en que las startups alcanzan el estatus de unicornio. Primero, combiné las columnas 'year', 'month' y 'day' en una sola columna 'fecha' utilizando la función "`pd.to_datetime()`", que me permitió convertir los datos de esas columnas en un formato de fecha y hora. Además, corrijo errores con el parámetro 'errors' como 'coerce'.

Luego, eliminé las filas con fechas no válidas (NaN) utilizando la función "`dropna()`", y creé una nueva columna llamada 'mes_numero' que contiene solo el número del mes (1 para enero, 2 para febrero, etc.) a partir de la columna 'fecha' utilizando la función "`dt.month`".

Después, agrupé los datos por el número de mes utilizando la función `groupby()` y conté la cantidad de empresas unicornio que se crearon en cada mes utilizando la función `size()`. Este resultado lo guardé en 'unicorns_by_month_number'.

Como los números de los meses no son tan descriptivos, creé un diccionario que mapea los números de los meses a sus nombres correspondientes y utilicé el método "`map()`" para reemplazar los números de los meses por sus nombres en el índice de `unicorns_by_month_number`.

Finalmente, creé un gráfico de barras utilizando la función `plt.bar()` para visualizar los patrones estacionales en la cantidad de empresas unicornio. Configuré los ejes x e y y el título del gráfico utilizando las funciones `xlabel()`, `ylabel()` y `title()`. Además, giré los nombres de los meses utilizando la función `xticks(rotation=45)` para que se vean mejor.

Más abajo, realicé la prueba estadística Kruskal-Wallis para determinar si hay evidencia de estacionalidad en la cantidad de empresas unicornio registradas por mes en los datos que se tienen.

Primero, se utiliza la función `groupby()` para contar la cantidad de empresas unicornio por mes y se almacena en la variable `conteo_unicornios_por_mes`. Luego, se crea un diccionario llamado `datos_por_mes` para almacenar los conteos de unicornios por cada mes, utilizando un bucle para iterar sobre los elementos del diccionario `meses` que fue creado anteriormente.

A continuación, se utiliza una lista de comprensión para seleccionar los datos de cada mes del diccionario `datos_por_mes` y se pasan como argumentos a la prueba de Kruskal-Wallis utilizando el operador `*`. El resultado de esta prueba se almacena en las variables `stat` y `p_value`.

Después se establece un nivel de significancia (α) de 0.05 y se compara el valor p obtenido en la prueba de Kruskal-Wallis con este valor. Si el valor p es menor a α , se rechaza la hipótesis nula y se concluye que hay evidencia de estacionalidad en los datos. En caso contrario, no se rechaza la hipótesis nula y no hay evidencia de estacionalidad en los datos.

Finalmente, se imprime en pantalla si hay evidencia de estacionalidad o no basado en la comparación del valor p con el nivel de significancia establecido.

Para el análisis use la Kruskal-Wallis en lugar de otras pruebas, como la prueba ANOVA de un solo factor, debido a sus propiedades no paramétricas. (he asumido que los datos no siguen una distribución normal)

Conclusión: El resultado de la prueba Kruskal-Wallis sugiere que no hay diferencias significativas en la cantidad de empresas que se convierten en unicornios entre los diferentes meses del año. En otras palabras, no hay evidencia de estacionalidad en el momento en que las empresas alcanzan el estatus de unicornio.

Esta conclusión implica que, en términos generales, no hay un momento específico del año en el que sea más probable que las empresas se conviertan en unicornios. Por lo tanto, para los inversores interesados en invertir en unicornios, no hay una "temporada" particular para buscar oportunidades de inversión. En cambio, los inversores deben enfocarse en otros factores, como la calidad de la empresa, el equipo de gestión, el mercado objetivo y el potencial de crecimiento, en lugar de depender de patrones estacionales para tomar decisiones de inversión.

