

Fecha de publicación: 2 de noviembre de 2024.

# Análisis Predictivo de Supervivencia en Startups Basado en Financiamiento, Sector e Indicadores Clave

**Godina Ramos, David<sup>1</sup>, Guevara Rodríguez, Pedro David<sup>1</sup>, y López Ruiz, Brandon Alan<sup>1</sup>**

<sup>1</sup>Licenciatura en Inteligencia de Negocios, Facultad de Ciencias Administrativas y Sociales, Universidad Autónoma de Baja California, Unidad Ensenada, Campus Valle Dorado, 22890

Autor correspondiente: Guevara Rodríguez, Pedro David (e-mail: pedro.guevara.rodriguez@uabc.edu.mx).

Este proyecto fue hecho como parte de la evaluación final de la materia “Ciencia de Datos” de la Licenciatura en Inteligencia de Negocios, impartida en UABC.

**ABSTRACTO** Este estudio analiza la relación entre el financiamiento, el sector y otros indicadores clave con la probabilidad de supervivencia de startups. Usando datos de Crunchbase, se realizaron análisis estadísticos inferenciales y técnicas de machine learning, como regresión logística y K-Means. Los resultados muestran que el financiamiento es un factor determinante, identificándose un umbral mínimo de ~\$64 millones USD para garantizar una probabilidad de supervivencia del 99%. Además, se evidenciaron diferencias significativas en las tasas de éxito según el sector, con tecnología y comunicaciones destacando como sectores clave. También se encontró que startups con múltiples rondas de financiamiento tienen mayores probabilidades de supervivencia, resaltando la importancia de relaciones sólidas con inversores. Por otra parte, el clustering permitió segmentar startups en 7 grupos con características similares, destacando que las más exitosas suelen ser más antiguas (fundadas antes de 2011) y con acceso a altos niveles de capital. Este estudio proporciona insights estratégicos tanto para emprendedores como para inversores, identificando patrones clave que pueden optimizar decisiones en el ecosistema emprendedor.

**TÉRMINOS DE INDEXAMIENTO:** financiamiento, supervivencia, startups, machine learning, análisis estadístico.

## I. INTRODUCCIÓN

El éxito de las startups es fundamental para impulsar la innovación tecnológica y el crecimiento económico. Estas empresas emergentes desempeñan un papel clave al desarrollar soluciones disruptivas, crear empleo y atraer inversiones. Sin embargo, su supervivencia está lejos de estar garantizada, ya que enfrentan un entorno altamente competitivo y dinámico. La financiación inicial es uno de los factores más críticos que determina si una startup puede escalar y sostenerse en el tiempo. Comprender cómo se relaciona el financiamiento con la supervivencia puede proporcionar insights valiosos tanto para emprendedores como para inversores.

### A. PLANTEAMIENTO DEL PROBLEMA

A pesar de su importancia, muchas startups no logran superar los primeros años de operación. Estudios recientes indican que más del 60% de las startups fracasan dentro de los primeros cinco años, citando la falta de financiamiento y la incapacidad para adaptarse a

los cambios del mercado como causas principales [6][7]. Además, factores como el sector industrial y el número de rondas de financiamiento parecen desempeñar un papel crucial en la probabilidad de éxito, pero su impacto específico no ha sido completamente cuantificado. La necesidad de identificar umbrales de financiamiento, analizar diferencias entre sectores e investigar patrones de inversión es crucial para optimizar la toma de decisiones en el ecosistema de startups.

El crecimiento económico y la innovación tecnológica dependen en gran medida de la capacidad de las startups para escalar y prosperar en mercados competitivos. Sin embargo, las estadísticas muestran que más del 60% de las startups fracasan dentro de sus primeros cinco años de operación, a menudo debido a insuficiencia de financiamiento y estrategias ineficaces [1][6].

Para abordar estas brechas de conocimiento, este estudio busca responder las siguientes preguntas clave:

- ¿Existe un umbral de financiamiento que se asocie con la supervivencia de una startup?
- ¿Existe una relación entre la supervivencia de la startup y el sector de la industria donde se maneja?
- ¿Existe una relación con el número de rondas de financiamiento de una startup en su supervivencia?

Este trabajo se enfoca en proporcionar respuestas basadas en análisis estadísticos y técnicas de ciencia de datos, utilizando un dataset detallado de Crunchbase.

## II. TRABAJO RELACIONADO

El financiamiento inicial de las startups ha sido identificado como un factor crítico para su éxito. Según

## III. METODOLOGÍA

### A. ADQUISICIÓN DE DATOS

El dataset utilizado fue obtenido de Crunchbase, una base de datos confiable que documenta información sobre startups y sus actividades relacionadas. Incluye variables críticas para analizar las dinámicas de supervivencia de las startups:

- *permalink (string)*: Identificador único.
- *name (string)*: Nombre de la startup.
- *homepage\_url (string)*: URL de la página web.
- *category\_list (string)*: Categorías del negocio.
- *funding\_total\_usd (int)*: Monto total de financiamiento.
- *status (string)*: Estado operativo de la startup (operando, cerrado, etc.).
- *country\_code (string)*: Código de país.
- *region, city (string)*: Localización geográfica.
- *funding\_rounds (int)*: Número de rondas de financiamiento.
- *founded\_at, first\_funding\_at, last\_funding\_at (dates)*: Fechas relacionadas con el inicio y financiamiento.

### B. PROCESAMIENTO DE DATOS

Siguiendo buenas prácticas en la preparación de datos, se realizaron las siguientes acciones:

- Conversión de columnas: Las fechas se transformaron a formato *datetime* y los valores de *funding\_total\_usd* a numéricos para facilitar su análisis.
- Manejo de valores nulos: Los valores nulos en las categorías como *category\_list* se rellenaron con "unknown". Además, los valores faltantes en *founded\_at* y *last\_funding\_at* se completaron

## B. PREGUNTAS CLAVE

estudios recientes, las startups con financiamiento inicial adecuado tienen una mayor probabilidad de sobrevivir y escalar, ya que pueden invertir en recursos estratégicos y responder más eficazmente a los cambios en el mercado [2][3]. Además, investigaciones en aprendizaje automático han demostrado que la combinación de variables como el número de rondas de financiamiento y el monto total recibido son predictores clave del desempeño a largo plazo de estas empresas [4]. Por ejemplo, Khavul y Deeds (2016) resaltan que una mayor inversión en las primeras etapas se correlaciona con una probabilidad más alta de éxito operativo en los primeros cinco años [5]. Asimismo, Hannigan, Briggs y Sundararajan (2021) destacan la importancia de sectores específicos y la diversificación del capital como factores adicionales en la sostenibilidad de las startups [6].

con datos de *first\_funding\_at*. Se implementaron estrategias para completar valores nulos en las columnas *city* y *country\_code*. Usando mapas de agrupación (*groupby*) de *country\_code* a *city* y de *city* a *country\_code*, se imputaron valores faltantes en estas columnas, asignando un valor relacionado cuando era posible o marcándolos como "unknown" en caso de no encontrar coincidencias. Esto asegura una mayor consistencia en los datos geográficos sin introducir sesgos significativos.

- Fechas previas a 1960 se ajustaron a 1 de enero de 1960; las posteriores al día actual fueron ajustadas a la fecha de análisis.
- Solamente para el análisis estadístico inferencial y la implementación de modelos de Machine Learning, se realizaron dos acciones clave en el procesamiento de los datos. Primero, en la variable *status*, se unificaron las categorías *acquired* e *ipo* bajo el estado *operating*, ya que ambas representan startups que alcanzaron metas financieras significativas, permitiendo simplificar el análisis entre startups exitosas (*operating*) y aquellas que han cerrado (*closed*). Segundo, se eliminaron registros con valores en 0 en atributos esenciales como *funding\_total\_usd*, dado que esta variable es fundamental para los análisis y modelos predictivos, y su ausencia podría introducir sesgos o reducir la precisión de los resultados. Estas acciones aseguran datos depurados y estructurados para un análisis robusto y confiable.

## C. ANÁLISIS DE DATOS

### 1) ANÁLISIS EXPLORATORIO DE DATOS (EDA)

El análisis exploratorio de datos (EDA) se utilizó como una etapa inicial para comprender las características principales del dataset, identificar patrones generales y detectar posibles valores atípicos. Se realizaron las siguientes actividades:

- Se utilizó el método describe() para calcular estadísticas resumen de las variables numéricas y temporales del dataset. Este análisis permitió observar el rango, la media, la mediana y la desviación estándar de las variables clave, proporcionando una visión inicial de la dispersión y características generales del dataset.
- Gráficas de pastel: Se exploró la variable categórica status para visualizar la proporción de startups en diferentes estados operativos. A su vez, también se exploró mediante la misma manera la variable categórica region y la variable categórica category\_list.
- Diagramas de dispersión: Se generó un diagrama de dispersión que muestra la distribución del financiamiento total recibido por las startups y su distribución del número de rondas de financiamiento que recibieron. Esto permitirá identificar cómo se distribuye la cantidad de rondas dentro del dataset, destacando posibles sesgos en el acceso a capital.
- Gráficas de barras: Mediante gráficas de ese estilo, se permitió visualizar a las startups con un mayor número de rondas de financiamiento, así como la región con mayor financiamiento, entre otras. Esto genera información útil para poder observar una posible correlación entre dichas variables.

### 2) ANÁLISIS ESTADÍSTICO INFERENCIAL

Para evaluar la relación entre el financiamiento total recibido y la probabilidad de supervivencia de las startups, se empleó un modelo de regresión logística binaria. Este método es adecuado porque la variable dependiente es categórica y binaria (supervivencia: 1, no supervivencia: 0), permitiendo modelar probabilidades en función del financiamiento. Se formuló una hipótesis nula donde no existe una relación significativa entre ambas variables, contrastada con una hipótesis alternativa que plantea una correlación significativa. Este análisis permite identificar la magnitud, dirección y significancia estadística de la relación, apoyando conclusiones sobre los niveles de financiamiento críticos para la supervivencia [7].

Para analizar las diferencias en los montos de financiamiento entre sectores industriales, se empleó el Análisis de Varianza (ANOVA), una herramienta estadística diseñada para comparar las medias de múltiples grupos. La hipótesis nula estableció que no hay diferencias significativas en los montos promedio de financiamiento entre sectores, mientras que la hipótesis

alternativa planteó que al menos un sector presenta diferencias significativas. Previo al análisis, los datos se filtraron para considerar los 10 sectores con mayor frecuencia en el dataset y se eliminaron los valores nulos en la columna *funding\_total\_usd*. Posteriormente, los datos se agruparon por sector industrial, y se calculó el estadístico F y el valor p mediante ANOVA. Al obtener un valor p menor a 0.05, se concluyó que existen diferencias significativas entre los montos de financiamiento de distintos sectores.

Para identificar los sectores específicos con diferencias significativas, se aplicó la prueba Tukey HSD, que evalúa comparaciones múltiples entre pares de sectores. Esta prueba identificó qué combinaciones de sectores presentan discrepancias estadísticamente relevantes en sus montos promedio de financiamiento. El uso combinado de ANOVA y Tukey HSD es adecuado debido a la naturaleza categórica de los sectores industriales y la variable continua del financiamiento. Este enfoque permitió detectar tendencias clave en el financiamiento de startups y destacar las diferencias relevantes entre sectores, proporcionando información valiosa para entender las dinámicas de inversión.

Finalmente, se empleó un modelo de regresión logística binaria para analizar si el número de rondas de financiamiento (*funding\_rounds*) influye significativamente en la probabilidad de supervivencia de las startups (*status*). Este análisis complementa el uso de regresión logística mencionado previamente, enfocándose ahora en cómo las rondas de financiamiento afectan la supervivencia. La hipótesis nula establece que no hay relación significativa entre estas variables, mientras que la hipótesis alternativa sugiere lo contrario. El modelo ajusta probabilidades de supervivencia en función del número de rondas, evaluando la dirección y magnitud del impacto mediante el coeficiente de la variable independiente y la significancia estadística a través del valor p. Este enfoque permite identificar si un mayor acceso a rondas de financiamiento incrementa la probabilidad de éxito de las startups, proporcionando información valiosa para decisiones estratégicas en la búsqueda de capital.

### 3) ALGORITMOS DE MACHINE LEARNING

Se utilizó el algoritmo K-Means para agrupar startups en 7 clusters basados en su financiamiento total (en millones de USD) y el número de rondas de financiamiento. Se preprocesaron los datos eliminando outliers y escalando las variables para garantizar una correcta agrupación. Los clusters identificaron patrones clave, calculando características promedio como financiamiento, rondas, región más frecuente, fecha promedio de fundación y categoría más frecuente. Este análisis permite segmentar startups en grupos homogéneos, facilitando la interpretación de tendencias y decisiones estratégicas [8].

Se utilizó un modelo de regresión logística para determinar el umbral mínimo de financiamiento necesario

para alcanzar una probabilidad de supervivencia del 99% en startups. Este modelo asocia la probabilidad de supervivencia (*operating*) con el financiamiento total (*funding\_total\_usd*). Para ello, se filtraron los datos eliminando las startups con financiamiento igual a cero y se dividieron en conjuntos de entrenamiento y prueba para construir el modelo. A partir del modelo ajustado, que considera los coeficientes de regresión, se calculó el nivel crítico de financiamiento necesario para alcanzar el 99% de probabilidad de supervivencia. Este umbral proporciona un valor de referencia clave para emprendedores e inversores, permitiendo identificar el monto mínimo que maximiza la probabilidad de éxito en startups.

Además, se realizó un testeo del modelo con una muestra de 10 empresas seleccionadas aleatoriamente. Se generaron predicciones de supervivencia para estas empresas, mostrando el financiamiento total, el estado real (*operating* o *closed*) y la probabilidad de supervivencia predicha. Este análisis refuerza la utilidad práctica del modelo al proporcionar un marco estratégico para la toma de decisiones financieras basadas en datos, optimizando las probabilidades de éxito empresarial.

#### D. VISUALIZACIÓN DE DATOS

La visualización de datos facilitó la interpretación de patrones clave y tendencias dentro del dataset, utilizando diversas técnicas gráficas:

- Gráficas de pastel: Representaron proporciones de variables categóricas como *status*, *category\_list* y *region*, ayudando a identificar sectores y regiones dominantes, así como la distribución de estados operativos.
- Diagramas de barras: Destacaron startups con mayor número de rondas de financiamiento, sectores con mayor financiamiento promedio y diferencias entre estados operativos, proporcionando insights claros sobre áreas de inversión clave.
- Diagramas de dispersión: Mostraron relaciones entre financiamiento total y rondas de financiamiento, identificando posibles correlaciones y startups con valores extremos.
- Clusters (K-Means): Visualizaron grupos homogéneos basados en financiamiento y rondas, resaltando patrones en startups similares.
- Gráficas comparativas: Compararon valores nulos en el dataset original frente al procesado, evidenciando la efectividad del preprocesamiento.

Estas visualizaciones complementaron el análisis estadístico, facilitando una comunicación clara y decisiones estratégicas basadas en datos.

## IV. RESULTADOS

### A. ANÁLISIS EXPLORATORIO DE DATOS (EDA)

Primeramente, se realizó una tabla comparativa donde se mostró los resultados del procesamiento de datos, comparando la cantidad de datos nulos en el dataset original y en el dataset procesado, obteniendo el siguiente resultado:

Comparación del porcentaje de valores nulos entre el dataset original y el actual:

	Column	Original Nulls (%)	Current Nulls (%)
11	founded_at	22.93	0.00
7	state_code	12.88	0.00
8	region	12.10	0.01
9	city	12.10	0.01
6	country_code	10.48	0.00
2	homepage_url	7.62	0.00
3	category_list	4.74	0.00
12	first_funding_at	0.04	0.00
1	name	0.00	0.00
0	permalink	0.00	0.00
4	funding_total_usd	0.00	0.00
5	status	0.00	0.00
10	funding_rounds	0.00	0.00
13	last_funding_at	0.00	0.00

FIGURA 1. Comparación del porcentaje de valores nulos

Se puede concluir que el procesamiento del dataset fue efectivo en contrarrestar la mayoría de valores nulos encontrados en el dataset.

El análisis descriptivo inicial muestra que el financiamiento total promedio de las startups es de \$18.48 millones USD, pero la mediana es significativamente menor (\$2 millones USD), lo que refleja una distribución sesgada con outliers que alcanzan hasta \$30.08 mil millones USD. La mayoría de las startups reciben solo una ronda de financiamiento (mediana de 1 ronda), aunque algunas excepcionales alcanzan hasta 19 rondas. En promedio, las startups se fundaron en 2008, con fechas de fundación que van desde 1960 hasta 2024, lo que evidencia la coexistencia de startups históricas y emergentes. Además, las rondas iniciales de financiamiento suelen ocurrir poco después de la fundación (promedio en 2011), mostrando una rápida búsqueda de capital en el ecosistema emprendedor. A continuación, se mostrarán gráficas descriptivas de diferentes hallazgos encontrados en el dataset utilizando estadística descriptiva:

Distribución de las 10 Categorías más Populares

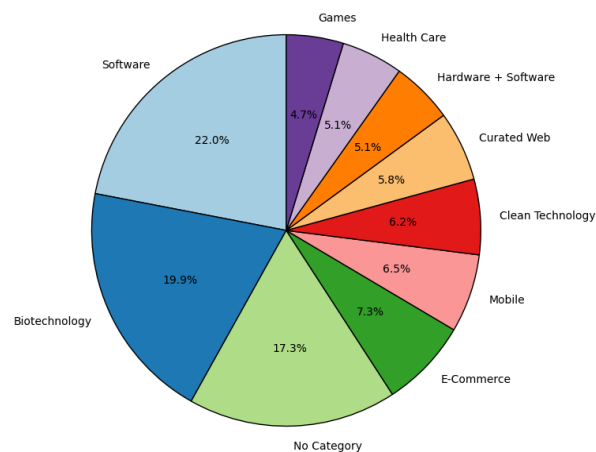
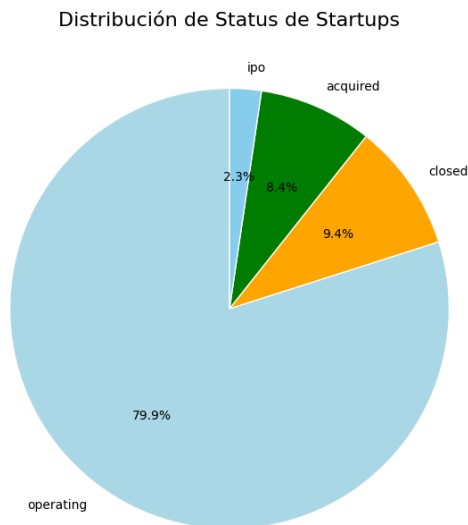


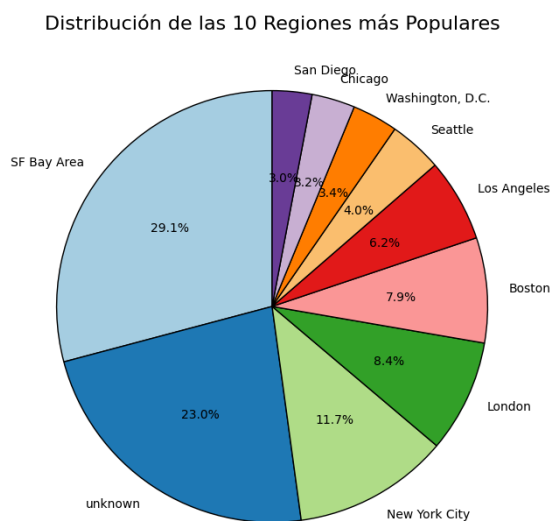
FIGURA 2. Distribución de las 10 Categorías más Populares

El gráfico de pastel muestra la distribución porcentual de las startups agrupadas en las 10 categorías más populares. "Software" representa la mayor proporción con el 22%, seguida de "Biotechnology" (19.9%) y "No Category" (17.3%), mientras que categorías como "Games" y "Health Care" tienen menor representación con 4.7% y 5.1%, respectivamente.



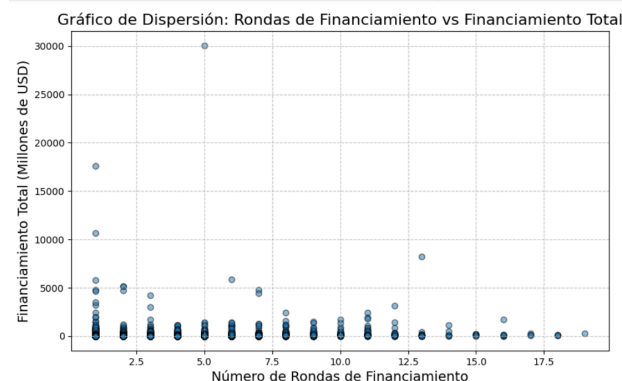
**FIGURA 3.** Distribución de Status de Startups

El gráfico de pastel presenta la distribución de startups según su estado operativo. La mayoría (79.9%) se encuentran activas (*operating*), mientras que un 9.4% han cerrado (*closed*), un 8.4% han sido adquiridas (*acquired*), y un 2.3% han salido a bolsa (*ipo*).



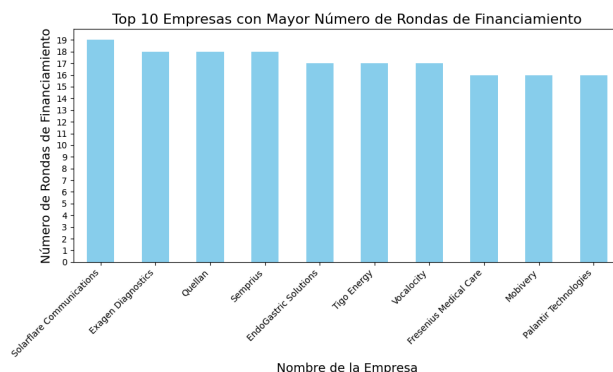
**FIGURA 4.** Distribución de las 10 Regiones más Populares

El gráfico de pastel muestra la distribución de las startups en las 10 regiones más populares. El área de la Bahía de San Francisco (*SF Bay Area*) domina con un 29.1%, seguido por startups de ubicación desconocida (*unknown*) con un 23%. Regiones como Nueva York (11.7%) y Londres (8.4%) también presentan una alta concentración.



**FIGURA 5.** Rondas de Financiamiento vs Financiamiento Total

El gráfico de dispersión muestra la relación entre el número de rondas de financiamiento y el financiamiento total recibido por las startups (en millones de USD). Se observan outliers significativos en términos de financiamiento total, con startups que superan los 10,000 millones de USD, independientemente del número de rondas. Estos valores extremos reflejan casos atípicos que pueden influir en el análisis y requieren consideración en interpretaciones posteriores.



**FIGURA 6.** Top 10 Empresas con Mayor Número de Rondas de Financiamiento

El gráfico de barras muestra las 10 empresas con el mayor número de rondas de financiamiento en el dataset. Solarflare Communications lidera con 19 rondas, seguida de otras como Exagen Diagnostics y Quellan con 18 rondas. Estas startups representan casos excepcionales de acceso recurrente a capital, indicando posibles ventajas estratégicas o de mercado en sus modelos de negocio.



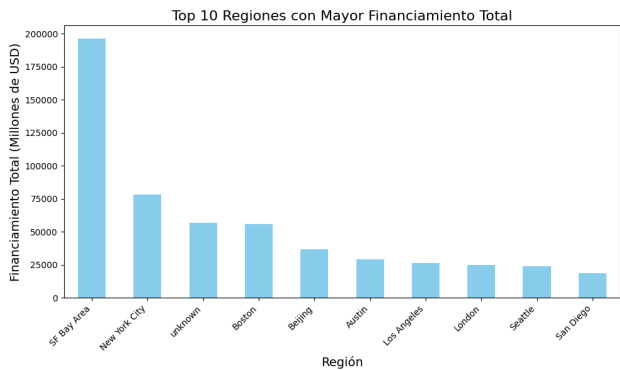


FIGURA 7. Top 10 Regiones con Mayor Financiamiento Total

El gráfico de barras muestra las regiones con mayor financiamiento total recibido (en millones de USD). La región del Área de la Bahía de San Francisco (*SF Bay Area*) sobresale significativamente con más de 200,000 millones de USD, seguida por Nueva York con aproximadamente 75,000 millones de USD. Otras regiones como Boston, Beijing y Austin presentan montos notables, destacando su importancia como centros de actividad emprendedora.

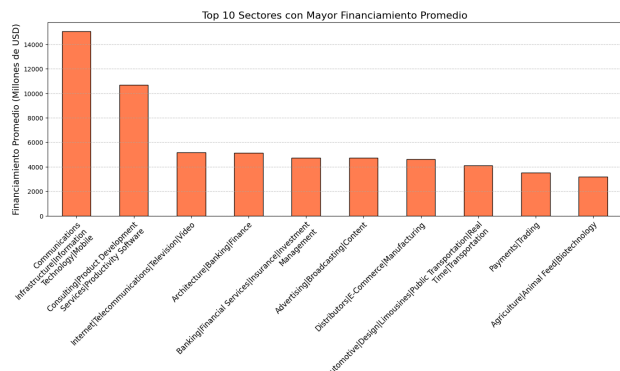


FIGURA 8. Top 10 Sectores con Mayor Financiamiento Promedio

El gráfico de barras destaca los 10 sectores con mayor financiamiento promedio (en millones de USD). El sector de Communications lidera con más de 14,000 millones de USD en promedio, seguido por Infraestructure/Information Technology/Mobile y Consulting/Product Development Services/Productivity Software. Estos sectores reflejan una alta concentración de capital, lo que sugiere su relevancia estratégica en el mercado global de startups.

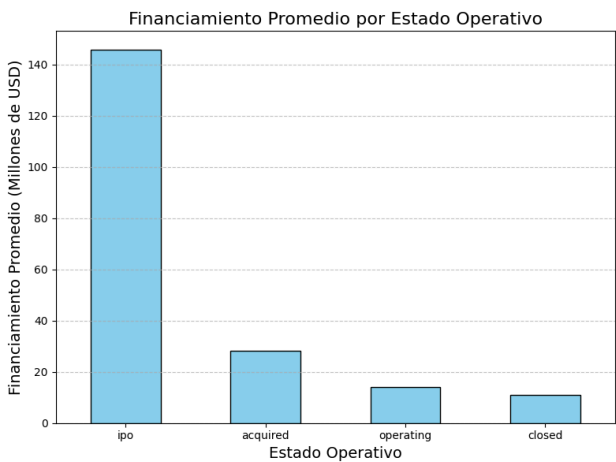


FIGURA 9. Financiamiento Promedio por Estado Operativo

El gráfico de barras presenta el financiamiento promedio recibido por startups según su estado operativo. Las startups que han salido a bolsa (*ipo*) lideran con un promedio superior a 140 millones de USD, seguidas por las adquiridas (*acquired*) con aproximadamente 40 millones de USD. Las startups activas (*operating*) y cerradas (*closed*) muestran promedios considerablemente menores, reflejando el impacto del estado operativo en la captación de recursos.

## B. ANÁLISIS ESTADÍSTICO INFERENCIAL

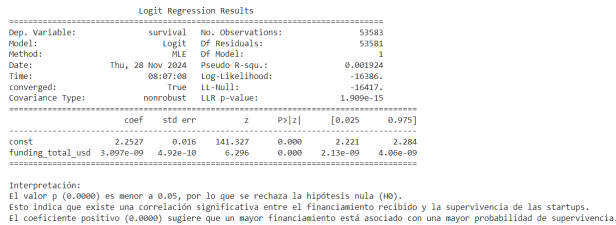


FIGURA 10. Resultados del Modelo de Regresión Logística: Financiamiento y Supervivencia

El cuadro muestra los resultados de un modelo de regresión logística que evalúa la relación entre el financiamiento total recibido (*funding\_total\_usd*) y la probabilidad de supervivencia (*status*) de las startups. El valor p ( $< 0.05$ ) indica que existe una correlación significativa entre ambas variables. El coeficiente positivo para *funding\_total\_usd* sugiere que un mayor financiamiento está asociado con una mayor probabilidad de supervivencia, destacando la importancia del capital como factor crítico para el éxito de las startups.

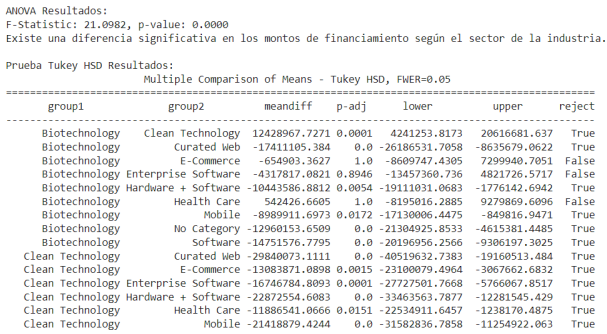


FIGURA 11. Resultados del ANOVA y Prueba Tukey HSD: Diferencias Significativas en Financiamiento entre Sectores Industriales.

El análisis de varianza (ANOVA) mostró un estadístico F de 21.0982 y un valor p de 0.0000, lo que indica que existen diferencias significativas en los montos de financiamiento promedio entre los sectores industriales. Esto nos lleva a rechazar la hipótesis nula (H0) que planteaba la inexistencia de diferencias, confirmando que el sector al que pertenece una startup influye significativamente en los niveles de financiamiento que recibe. Para identificar qué sectores presentan diferencias específicas, se utilizó la prueba Tukey HSD, que evalúa comparaciones entre pares de sectores. Los resultados muestran múltiples diferencias significativas entre sectores clave, como *Biotechnology* y *Software*, así como entre *Clean Technology* y *Curated Web*, entre otros. Estas diferencias subrayan la disparidad en la asignación de recursos financieros según el tipo de industria. Este análisis es relevante para emprendedores e inversores, ya

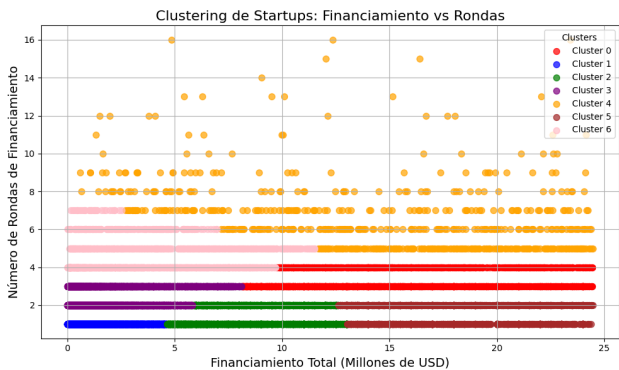


FIGURA 13. Clustering de Startups: Financiamiento Total vs Número de Rondas

El gráfico muestra los resultados del algoritmo de clustering K-Means, que agrupó las startups en 7 clusters con base en su financiamiento total (en millones de USD) y el número de rondas de financiamiento. Cada color representa un cluster, destacando patrones específicos entre startups con características similares. Los clusters permiten identificar startups con niveles altos de financiamiento y rondas recurrentes, frente a aquellas con menor actividad financiera, facilitando el análisis estratégico para inversiones y éxito.

que evidencia cómo el sector industrial puede influir en el acceso a financiamiento. Sectores como *Biotechnology* y *Clean Technology* destacan por sus niveles significativamente distintos de financiamiento en comparación con otras categorías, lo que puede ser crucial al tomar decisiones estratégicas.

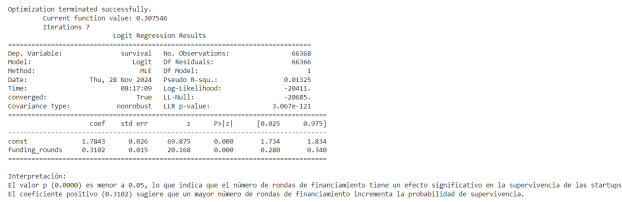


FIGURA 12. Resultados del Modelo de Regresión Logística: Número de Rondas de Financiamiento y Supervivencia

El cuadro presenta los resultados de un modelo de regresión logística que analiza el impacto del número de rondas de financiamiento (*funding\_rounds*) en la probabilidad de supervivencia (*status*) de las startups. El valor p ( $< 0.05$ ) confirma un efecto significativo, y el coeficiente positivo (0.3102) indica que un mayor número de rondas de financiamiento incrementa la probabilidad de supervivencia. Estos resultados resaltan la importancia de mantener acceso a múltiples rondas de financiamiento para mejorar las perspectivas de éxito.

C. ALGORITMOS DE MACHINE LEARNING

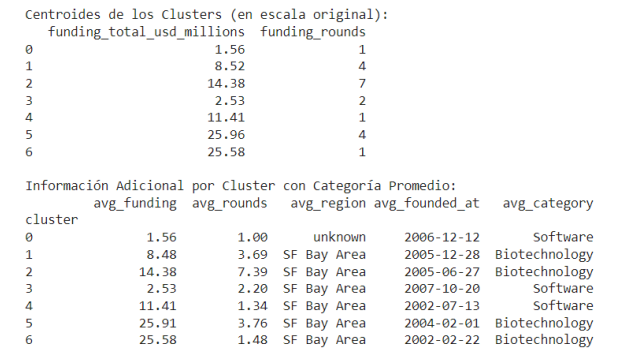


FIGURA 14. Centroides e Información Adicional de los Clusters

El cuadro superior presenta los centroides de los 7 clusters en escala original, indicando el financiamiento promedio (en millones de USD) y el número promedio de rondas de financiamiento. El cuadro inferior complementa con información adicional, como la región más frecuente (*avg\_region*), la categoría más frecuente (*category\_list*) y la fecha promedio de fundación (*avg\_founded\_at*) para cada cluster. La mayoría de los clusters tienen como región predominante el Área de la

Bahía de San Francisco (*SF Bay Area*) y las categorías que más dominan son Software y Biotecnología (*Biotechnology*). Además, se observa que, en promedio, las startups en estos clusters fueron fundadas antes del año 2011, destacando una tendencia hacia empresas más consolidadas en términos de tiempo.

Resultados del Testeo en 10 Empresas:			
Company Name	Funding Total (USD)	Actual Status (1=Operating)	\
62070	VIPAAR 850000.00	1	
38879	Neuropure 500000.00	0	
19110	eVeritas, Inc. 150000.00	1	
9479	Canas 403000.00	1	
847	Accelerator Centre 800000.00	1	
37813	myOrder 220000.00	1	
1766	Aerovance 120500000.00	0	
18153	enercast 2954475.00	1	
60870	User Replay 5624176.00	1	
34286	Mark Forged 1000000.00	1	
Predicted Survival Probability (%)			
62070	51.53		
38879	50.90		
19110	50.27		
9479	50.72		
847	63.98		
37813	50.39		
1766	99.58		
18153	55.28		
60870	59.96		
34286	51.79		

El umbral mínimo de financiamiento para una probabilidad de supervivencia del ~99% es: \$63,984,775.74

FIGURA 15. Resultados del Testeo del Modelo de Regresión Logística en 10 Empresas: Financiamiento Total y Probabilidad de Supervivencia Predicha.

El análisis del modelo de regresión logística binaria indica que el umbral mínimo de financiamiento necesario para alcanzar una probabilidad de supervivencia del ~99% es de \$63,984,775.74 USD. Este hallazgo subraya la relevancia de asegurar capital significativo para maximizar la probabilidad de éxito en startups. Las empresas con mayor financiamiento tienden a tener mayores probabilidades de supervivencia, como predice el modelo.

No obstante, en el testeo realizado a 10 empresas seleccionadas aleatoriamente, se observaron los siguientes patrones clave:

Startups con un financiamiento considerable, como Aerovance, presentan una alta probabilidad de supervivencia (99%), según el modelo. Sin embargo, en este caso, la empresa no está operando, lo que sugiere que factores adicionales fuera del financiamiento también impactan en la sostenibilidad de las startups. Empresas con financiamientos menores, como NeuroPure y eVeritas, Inc., muestran probabilidades de supervivencia más bajas, cercanas al 50%, lo que refleja el riesgo elevado asociado a recursos financieros limitados.

V. CONCLUSIONES Y RECOMENDACIONES

El análisis destaca que el financiamiento total es un factor crítico para la supervivencia de las startups. Se determinó que un financiamiento mínimo de ~\$63.9 millones USD es necesario para alcanzar una probabilidad de supervivencia del ~99%. Esto subraya la importancia de asegurar capital significativo desde las primeras etapas del emprendimiento. Además, se observó que las startups con más rondas de financiamiento tienen mayores probabilidades de éxito, lo que refuerza la necesidad de

establecer relaciones sólidas con inversores y garantizar un flujo constante de financiamiento a lo largo del tiempo. En términos sectoriales, el análisis ANOVA y la prueba Tukey HSD confirmaron diferencias significativas en los montos promedio de financiamiento entre sectores industriales. Sectores como Biotechnology y Clean Technology sobresalen, mientras que sectores como E-Commerce y Curated Web presentan menores niveles de financiamiento. Estas disparidades subrayan la necesidad de estrategias específicas según el sector para maximizar el acceso al capital y las probabilidades de éxito.

Por otro lado, el clustering permitió identificar que las startups más exitosas tienden a ser más antiguas (fundadas antes de 2011), con acceso a múltiples rondas de financiamiento y mayores cantidades de capital promedio. Esto sugiere que la experiencia y la capacidad de captación sostenida de recursos son factores determinantes en el éxito a largo plazo.

Un aspecto a considerar es que el dataset utilizado cuenta con una distribución altamente desbalanceada, con más del 80% de startups en estado operativo y menos del 20% cerradas. Este desbalance puede introducir ruido en los análisis predictivos y estadísticos, limitando la generalización de los resultados. Se recomienda que futuros análisis trabajen con un dataset balanceado (50% startups operativas y 50% cerradas) para mejorar la robustez y validez de las conclusiones.

- Con base en estos hallazgos, se hacen las siguientes recomendaciones:  
Estrategias de Captación Temprana: Las startups deben priorizar estrategias efectivas de captación de financiamiento desde las etapas iniciales, asegurando niveles de capital suficientes para su sostenibilidad y expansión.
- Diversificación Sectorial y Regional: Emprendedores e inversores deben considerar oportunidades en sectores emergentes y regiones menos saturadas, diversificando sus estrategias para maximizar el impacto.
- Apoyo Institucional: Instituciones y políticas públicas deben canalizar recursos hacia sectores estratégicos como tecnología y comunicaciones, identificados como áreas con altas tasas de éxito, para fortalecer el ecosistema emprendedor.
- Mejoras en los Datos: Futuros análisis deben enfocarse en datasets más balanceados y actualizados que reflejen las dinámicas actuales del ecosistema emprendedor, mejorando la precisión de los modelos y conclusiones.

Estos hallazgos proporcionan un marco estratégico para emprendedores, inversores e instituciones, optimizando las decisiones financieras y fortaleciendo el impacto del ecosistema de startups.



## REFERENCIAS

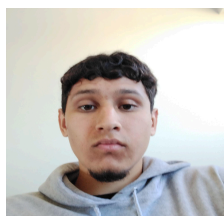
- [1] Cumming, D. J., & Fischer, H. M. (2014). Publicly funded business advisory services and entrepreneurial outcomes. *Research Policy*, 43(2), 404-418. <https://doi.org/10.1016/j.respol.2013.12.005>
- [2] Hannigan, T. J., Briggs, A. R., & Sundararajan, M. (2021). Venture capital and early-stage startups: Success factors in funding and scaling. *Journal of Business Venturing*, 36(5), 106033. <https://doi.org/10.1016/j.jbusvent.2021.106033>
- [3] Khavul, S., & Deeds, D. (2016). The influence of venture capital on innovation and startup success. *Strategic Entrepreneurship Journal*, 10(1), 90-104. <https://doi.org/10.1002/sej.1214>
- [4] Park, Y., Shin, J., & Lee, K. (2020). Startup funding patterns: A machine learning approach. *Technological Forecasting and Social Change*, 158, 120143. <https://doi.org/10.1016/j.techfore.2020.120143>
- [5] Colombo, M. G., & Dawid, H. (2016). Complementarities and success in high-tech startups: The role of business and financing. *Small Business Economics*, 47(3), 503-525. <https://doi.org/10.1007/s11187-016-9734-3>
- [6] Eesley, C. E., & Wu, L. (2016). Entrepreneurial financing: Decision-making and outcomes. *Management Science*, 62(9), 2776-2794. <https://doi.org/10.1287/mnsc.2015.2240>
- [7] Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. John Wiley & Sons.
- [8] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651-666. <https://doi.org/10.1016/j.patrec.2009.09.011>

Universidad Autónoma de Baja California ( UABC ) , campus valle dorado,, pero cuando salió del tronco común se dio cuenta que le apasionaba mas la programación y estar a la vanguardia de los avances tecnológicos , el soporte y el manejo de computadoras. Su objetivo es titularse y ejercer su profesión de programador en empresas de alto renombre mundial .



**David Godina Ramos.** Nacido el 29 de Febrero, en Ensenada B.C, desde una edad temprana, David mostró una gran curiosidad por aprender, lo que llevó a destacarse como estudiante en todas las etapas de su educación. En la secundaria, asistió a la Escuela Secundaria Técnica #19, donde descubrió su pasión por la tecnología aplicada y la gestión de información. Para la preparatoria, David ingresó al CBTis #41, donde se especializó en informática administrativa. Durante este periodo, adquirió conocimientos en software de oficina, diseño de bases de datos y automatización de procesos administrativos. Además, participó en un proyecto que optimizó la gestión de inventarios escolares mediante el uso de hojas de cálculo avanzadas. Con su enfoque claro en la tecnología y su aplicación práctica, David decidió continuar su formación en la Universidad Autónoma de Baja California (UABC), campus Valle Dorado, donde estudia la licenciatura en Inteligencia de negocios. Su objetivo es desarrollar soluciones ofimáticas innovadoras que permitan a las empresas locales optimizar sus procesos administrativos y mejorar su productividad.

**Pedro David Guevara Rodríguez** Nacido el 2 de enero de 2003 en Ensenada, B.C., Pedro mostró desde una edad temprana una pasión por la tecnología y el desarrollo de soluciones innovadoras. Egresado como Técnico en Programación del CECyTE BC Plantel Ensenada, decidió continuar su formación académica en la Universidad Autónoma de Baja California (UABC), donde actualmente es estudiante en etapa terminal de la Licenciatura en Inteligencia de Negocios y miembro activo del Tech4Good Research Lab. Desde abril de 2023, Pedro lidera el proyecto Emotions&Care, una iniciativa enfocada en apoyar la salud mental de jóvenes universitarios. Este proyecto ha sido presentado en múltiples congresos académicos internacionales de habla hispana y validado en sesiones tanto en Ensenada, Baja California, como en Madrid, España, consolidándose como un referente en innovación tecnológica y salud mental. Pedro también ha destacado como autor principal y coautor en dos artículos científicos indexados en revistas de prestigio internacional. Además, realizó una estancia científica de verano de siete semanas en la Universidad de Guadalajara, Jalisco, donde fortaleció sus habilidades de investigación y presentó los resultados de su trabajo en un congreso académico internacional celebrado en Nuevo Vallarta, Nayarit. Su objetivo es seguir desarrollando soluciones tecnológicas innovadoras que impacten positivamente en la sociedad y fomenten el bienestar en diversos sectores.



**Brandon Alan Lopez Ruiz** Nacido el 15 de mayo de 2003 , en Ensenada B.C, desde una temprana edad Alan mostró una gran pasión por la computación, Asistió a la escuela secundaria Héctor A. Migóni Fuentes , donde estudió diseño arquitectónico lo que no fue de su agrado , Posteriormente asistió a la preparatoria COBACH, donde se especializó en administración de empresas, decidió seguir el camino de la administración de empresas y usar todo su conocimiento adquirido para titularse como licenciado en administración de empresas en la