



PROYECTO CAPSTONE

# PREDICCIÓN DE ATERRIZAJE SPACEX FALCON 9

Análisis predictivo para la optimización de costes en lanzamientos  
espaciales

---

Basado en la metodología de Ciencia de Datos

# EL DESAFÍO COMERCIAL

---

## LA REVOLUCIÓN DE LA REUTILIZACIÓN

SpaceX ha transformado la industria aeroespacial mediante la reutilización de la primera etapa del cohete Falcon 9.

- > **Coste SpaceX:** ~\$62 Millones por lanzamiento.
- > **Competencia:** >\$165 Millones por lanzamiento.
- > **Clave del Éxito:** Recuperar la primera etapa reduce drásticamente los costes.



# METODOLOGÍA DEL PROYECTO



## 1. RECOLECCIÓN

Obtención de datos mediante la API REST pública de SpaceX y Web Scraping de registros históricos en Wikipedia.



## 2. PROCESAMIENTO

Limpeza de datos (Data Wrangling), manejo de valores nulos y codificación One-Hot para variables categóricas.

## 3. ANÁLISIS (EDA)

Exploración visual y estadística (SQL) para identificar patrones de éxito según carga, órbita y sitio de lanzamiento.



## 4. PREDICCIÓN

Entrenamiento y evaluación de modelos de Machine Learning (Regresión Logística, SVM, Árboles, KNN).

# RECOLECCIÓN DE DATOS: API SPACE X

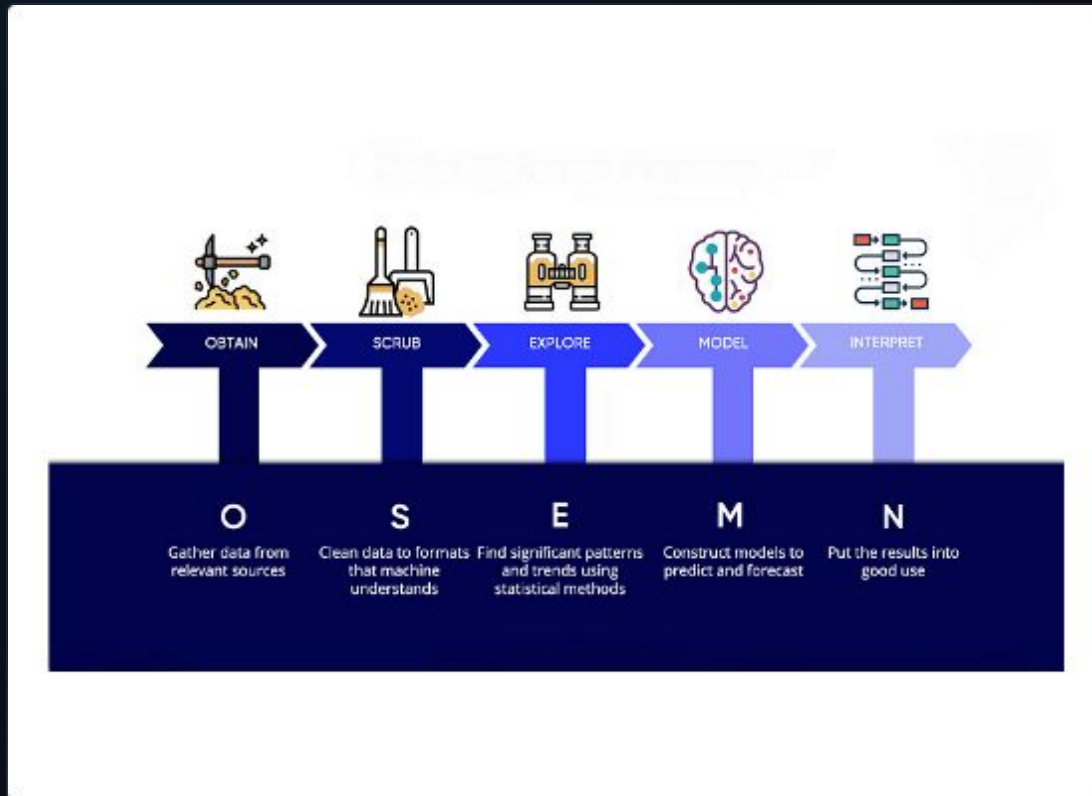
## CONEXIÓN A LA FUENTE

Utilizamos la librería requests en Python para interactuar con la API v4 de SpaceX.

- > **Endpoint:** <https://api.spacexdata.com/v4/launches/past>
- > **Datos Extraídos:** Detalles del vuelo, carga útil, núcleo (core) del cohete, sitio de lanzamiento y resultado del aterrizaje.
- > **Filtrado:** Se excluyeron lanzamientos del Falcon 1 para centrarse exclusivamente en la arquitectura Falcon 9.

```
# Ejemplo de petición a la API response =  
requests.get(spacex_url) data =  
pd.json_normalize(response.json()) # Extracción de  
características clave data['BoosterVersion'] =  
getBoosterVersion(data) data['LaunchSite'] =  
getLaunchSite(data) data['PayloadMass'] =  
getPayloadData(data)
```

# LIMPIEZA Y PREPARACIÓN (WRANGLING)



## TRANSFORMACIÓN DE DATOS

Los datos crudos requieren un procesamiento intensivo para ser útiles en los modelos predictivos.

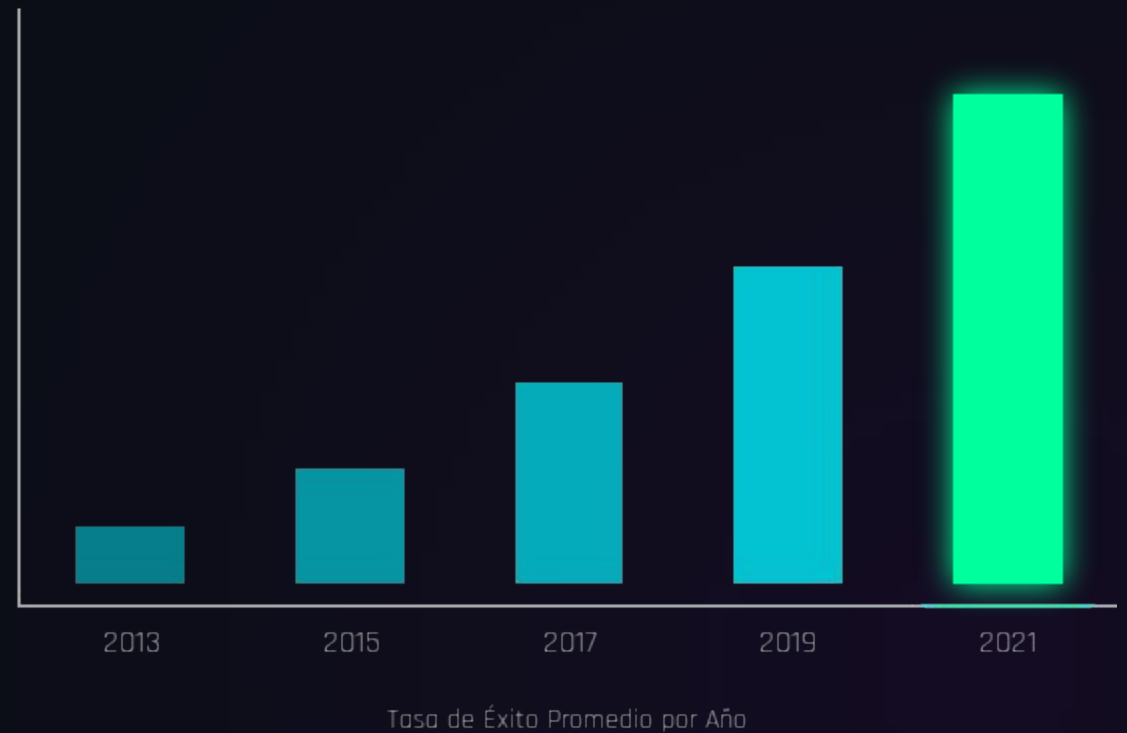
- > **Valores Nulos:** Imputación de la media para la masa de la carga útil (`PayloadMass`).
- > **Etiquetado de Clases:** Creación de la variable objetivo `Class`.
  - 1 = Éxito (Aterrizaje confirmado)
  - 0 = Fallo (Estrellado o perdido)
- > **Resultado:** Dataset limpio con 90 lanzamientos y atributos estandarizados.

# ANÁLISIS EXPLORATORIO: TENDENCIA ANUAL

## EVOLUCIÓN DEL ÉXITO

El análisis temporal revela una clara curva de aprendizaje en la tecnología de SpaceX.

- > Desde 2013 hasta 2020, la tasa de éxito ha incrementado constantemente.
- > Los primeros años muestran alta variabilidad experimental.
- > A partir de 2017, la tecnología "Block 5" estabilizó los aterrizajes, convirtiéndolos en rutina.



# ANÁLISIS POR SITIO DE LANZAMIENTO



## GEOGRAFÍA DEL ÉXITO

Distintas plataformas presentan diferentes tasas de éxito, influenciadas por el tipo de órbita y misión.

- > **CCAFS SLC-40 (Florida):** El sitio más activo, con una tasa de éxito mixta debido a la variedad de misiones.
- > **KSC LC-39A (Florida):** Histórico (Apollo), usado para misiones tripuladas y Falcon Heavy. Alta tasa de éxito.
- > **VAFB SLC-4E (California):** Usado para órbitas polares, con una tasa de éxito excepcionalmente alta.

# VISUALIZACIÓN GEOESPACIAL (FOLIUM)



## PROXIMIDAD Y SEGURIDAD

El análisis interactivo con Folium permite calcular distancias a infraestructuras críticas.

Se confirmó que todos los sitios de lanzamiento mantienen distancias de seguridad estrictas con áreas pobladas, pero están estratégicamente cerca de costas y vías férreas para logística.



## CLUSTERS DE ATERRIZAJE

Los marcadores en el mapa revelan clústeres de éxito (Verde) y fallo (Rojo).

KSC LC-39A muestra la mayor densidad de marcadores verdes, indicando una fiabilidad superior para misiones comerciales y gubernamentales.

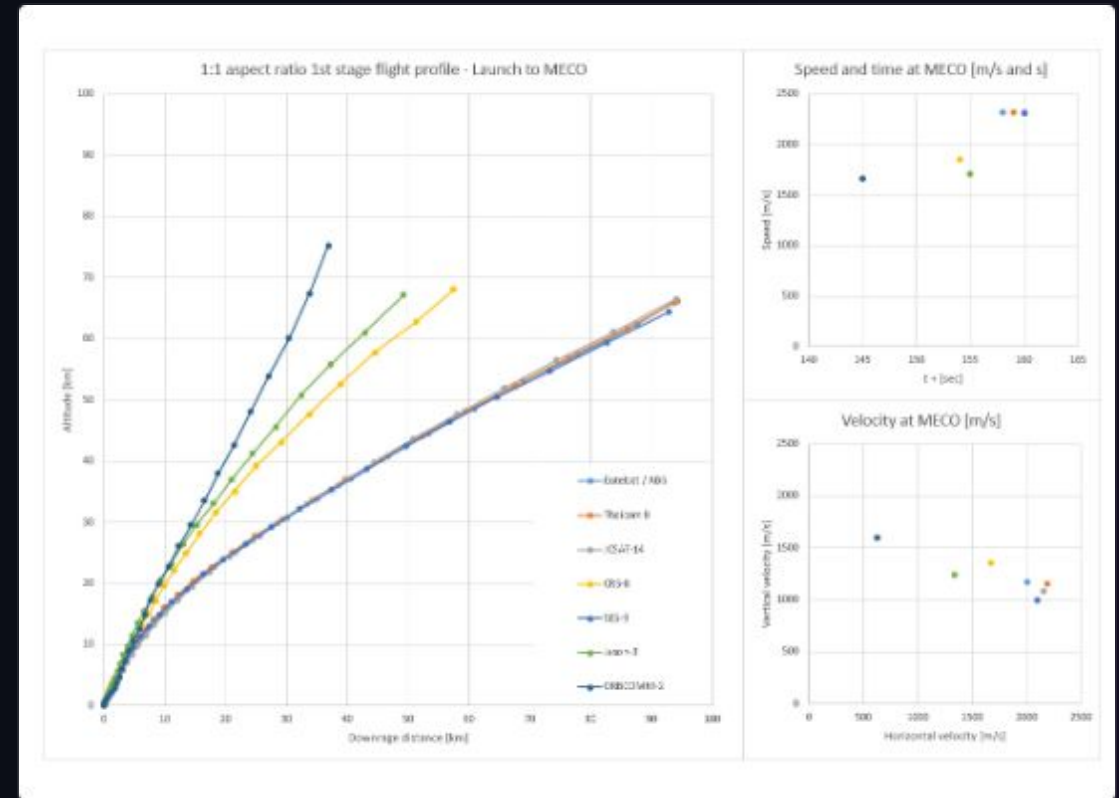


# CARGA ÚTIL VS. RESULTADO

## IMPACTO DEL PESO EN EL ATERRIZAJE

Analizamos si una carga más pesada dificulta el aterrizaje de la primera etapa.

- > **Cargas Ligeras:** Altas tasas de éxito en órbitas bajas (LEO).
- > **Cargas Pesadas:** Mayor dificultad en órbitas de transferencia geoestacionaria (GTO) debido al mayor consumo de combustible necesario para la inserción orbital, dejando menos reserva para el aterrizaje.



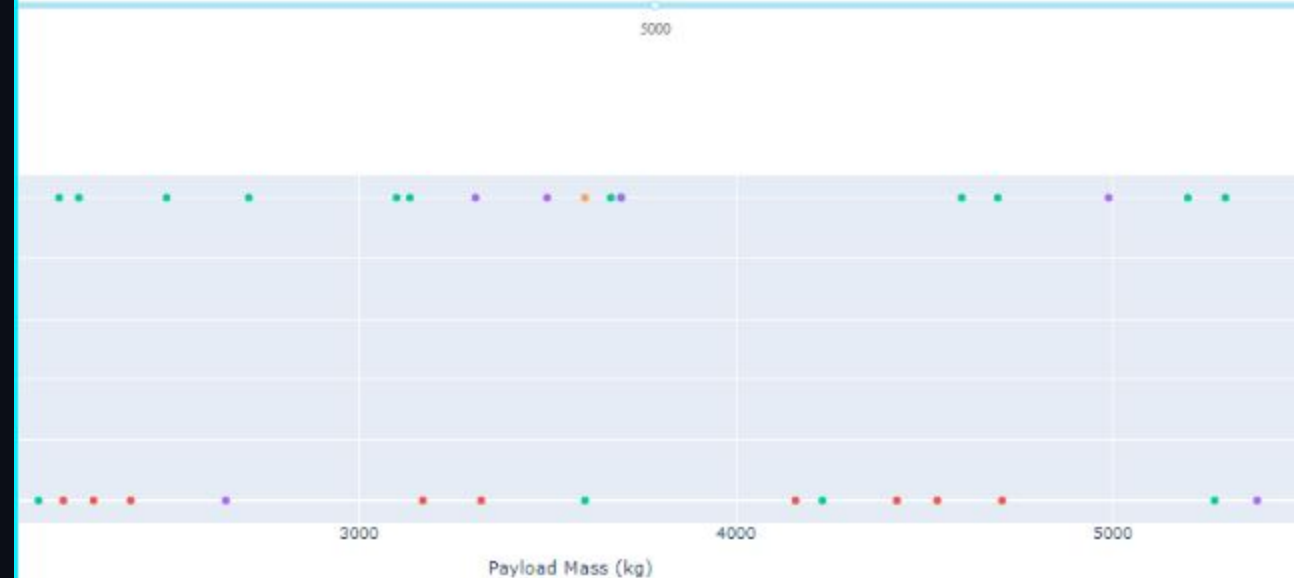
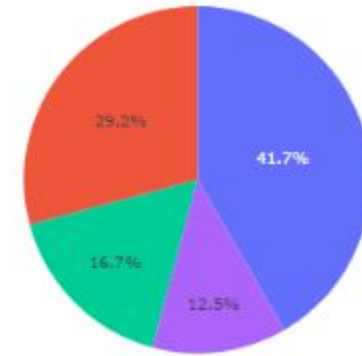
# DASHBOARD INTERACTIVO

## EXPLORACIÓN EN TIEMPO REAL

Se desarrolló una aplicación web interactiva utilizando **Plotly Dash** para permitir a los stakeholders explorar los datos dinámicamente.

- > **Filtro por Sitio:** Dropdown para seleccionar plataformas específicas.
- > **Slider de Carga:** Ajuste del rango de peso (Kg) para ver su correlación con el éxito.
- > **Gráficos Dinámicos:** Gráfico de pastel (Pie Chart) de éxito y Scatter Plot de correlación que se

## SpaceX Launch Records Dashboard



# PREPARACIÓN PARA MACHINE LEARNING



## ESTANDARIZACIÓN

Las variables numéricas (como Masa de Carga) tienen rangos muy distintos a las categóricas.

Utilizamos StandardScaler para normalizar los datos, asegurando que ninguna variable domine el entrenamiento del modelo por pura magnitud.



## TRAIN / TEST SPLIT

División del dataset para validación honesta.

```
X_train, X_test, Y_train, Y_test = train_test_split(
    X, Y, test_size=0.2, random_state=2 )
```

80% Entrenamiento / 20% Prueba

# MODELOS EVALUADOS

---

## REGRESIÓN LOGÍSTICA

Línea base robusta para problemas de clasificación binaria.



## SVM

Support Vector Machine. Eficaz en espacios dimensionales altos.



## ÁRBOL DE DECISIÓN

Modelo interpretable que simula la toma de decisiones humana.



## KNN

K-Nearest Neighbors. Clasificación basada en similitud de instancias cercanas.

Todos los modelos fueron optimizados usando GridSearchCV para encontrar los mejores hiperparámetros.

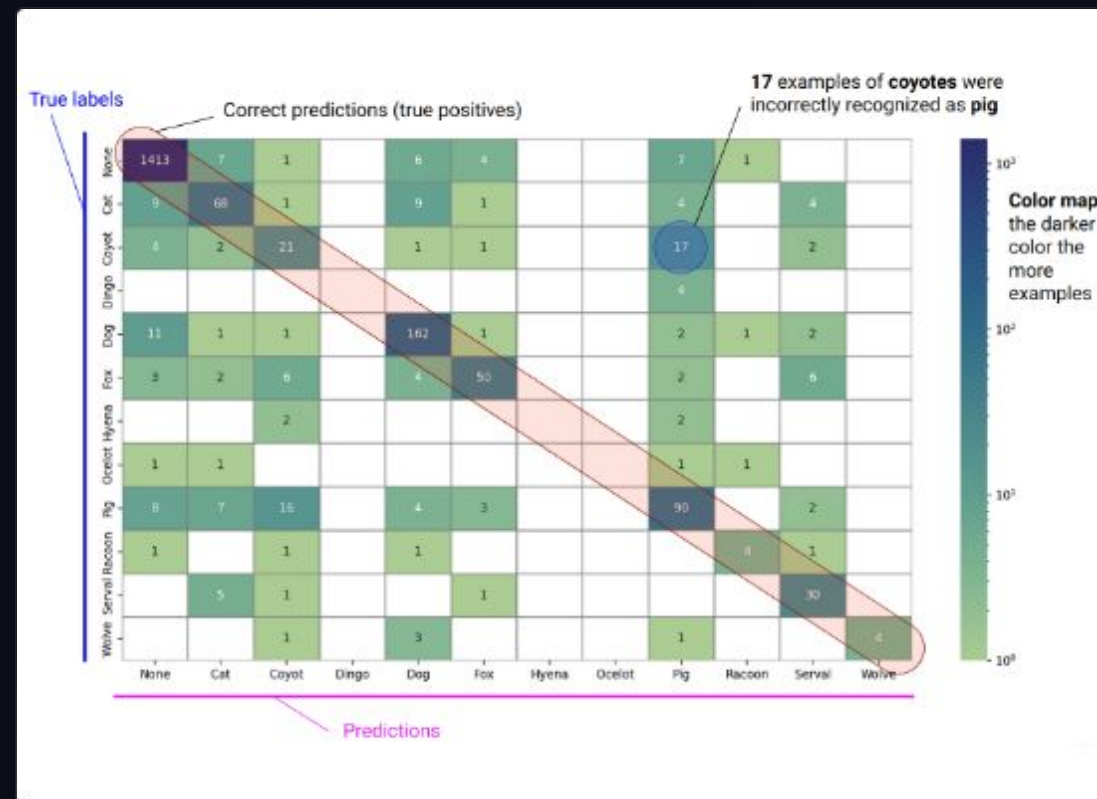
# EVALUACIÓN: MATRIZ DE CONFUSIÓN

## ENTENDIENDO LOS ERRORES

La precisión no lo es todo. La matriz de confusión nos muestra dónde se equivoca el modelo.

- > **Falsos Positivos:** Predecir que aterrizará, pero se estrella. (Riesgo alto de pérdida económica).
- > **Falsos Negativos:** Predecir que se estrellará, pero aterriza bien. (Oportunidad perdida).

Nuestros modelos mostraron una tendencia a minimizar los Falsos Positivos, lo cual es crucial para la seguridad financiera.



# RESULTADOS COMPARATIVOS

Precisión (Accuracy) en el conjunto de prueba



Todos los modelos alcanzaron un rendimiento similar, pero el **Árbol de Decisión** destacó por su interpretabilidad y velocidad de entrenamiento.

# EL MODELO GANADOR

## ÁRBOL DE DECISIÓN / SVM

Aunque numéricamente similares, recomendamos el Árbol de Decisión para explicar la lógica a los inversores.

- > **Precisión:** ~83.33% en datos no vistos.
- > **Variables Clave:**
  - > Masa de Carga (PayloadMass)
  - > Tipo de Órbita
  - > Número de Vuelo (Experiencia acumulada)



# CONCLUSIONES DE NEGOCIO

---

## VIABILIDAD PREDICTIVA

Es factible predecir el éxito del aterrizaje de la primera etapa con una precisión superior al 80% utilizando datos públicos.

## AHORRO DE COSTES

La capacidad de identificar lanzamientos de alto riesgo permite ajustar las ofertas (bids) y seguros, optimizando el margen de beneficio.

"La reutilización no es solo una ventaja técnica, es el modelo de negocio del futuro espacial."



# TRABAJO FUTURO

---



## DATOS METEOROLÓGICOS

Incorporar datos del clima en tiempo real (viento, oleaje) en la zona de aterrizaje para mejorar la precisión.



## TELEMETRÍA EN VIVO

Integrar datos de telemetría en tiempo real durante el lanzamiento para predicciones dinámicas.




## DEEP LEARNING

Utilizar Redes Neuronales para capturar patrones no lineales más complejos a medida que crece el dataset.

# ¿PREGUNTAS?

Gracias por su atención

 [github.com/Japura11/SpaceX-Falcon-9](https://github.com/Japura11/SpaceX-Falcon-9)

# IMAGE SOURCES

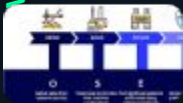
---



[https://i.ytimg.com/vi/TthLhqq4JUu/hq720.jpg?sqp=-oaymwEhCK4FEIIDSFryq4qpAxMIARUAAAAAGAEIAADIQj0AgKJD&rs=AOn4CLD8\\_SLlx60DHqaf\\_rLrz8NLCMZ6w](https://i.ytimg.com/vi/TthLhqq4JUu/hq720.jpg?sqp=-oaymwEhCK4FEIIDSFryq4qpAxMIARUAAAAAGAEIAADIQj0AgKJD&rs=AOn4CLD8_SLlx60DHqaf_rLrz8NLCMZ6w)

Source: [www.youtube.com](https://www.youtube.com)

---



[https://miro.medium.com/0\\*Lvb\\_AW6-EHer1Pvw.png](https://miro.medium.com/0*Lvb_AW6-EHer1Pvw.png)

Source: [blog.gopenai.com](https://blog.gopenai.com)

---



[https://irp.cdn-website.com/c70d7c4a/files/uploaded/Viewing\\_Guide\\_Map2025c.jpg](https://irp.cdn-website.com/c70d7c4a/files/uploaded/Viewing_Guide_Map2025c.jpg)

Source: [www.launchphotography.com](https://www.launchphotography.com)

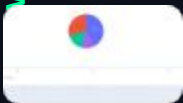
---



<http://i.imgur.com/wn47DXd.jpg>

Source: [www.reddit.com](https://www.reddit.com)

---



[https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/labs/module\\_3/images/dash\\_completed.PNG](https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/labs/module_3/images/dash_completed.PNG)

Source: [author-ide.skills.network](https://author-ide.skills.network)

---



<https://towardsdatascience.com/wp-content/uploads/2024/09/1qARgbCtsUIQwtk24bPecEg.png>

Source: [towardsdatascience.com](https://towardsdatascience.com)

# IMAGE SOURCES

---



[https://s3.amazonaws.com/marquee-test-akiaisur2rgicbmpehea/zl55VTniSsiPlvosPrHg\\_Reusable\\_Rockets.jpg](https://s3.amazonaws.com/marquee-test-akiaisur2rgicbmpehea/zl55VTniSsiPlvosPrHg_Reusable_Rockets.jpg)

Source: [interactive.satellitetoday.com](https://www.interactive.satellitetoday.com)



<https://www.nasa.gov/wp-content/uploads/2023/03/nhq202110270010.jpg>

Source: [www.nasa.gov](https://www.nasa.gov)