



IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Christian Contreras>

<20/05>



Descripción

- Resumen ejecutivo
- Introducción
- Metodología
- Resultados
- Conclusión
- Apéndice

Resumen ejecutivo

- Resumen de metodologías
 - Recopilación de datos
 - Gestión de datos
 - Análisis exploratorio de datos con visualización de datos
 - Análisis exploratorio de datos con SQL
 - Creación de un mapa interactivo con Plotly Dash
 - Análisis predictivo (clasificación)
- Resumen de todos los resultados
 - Resultados del análisis exploratorio de datos
 - Resultados del análisis predictivo

Introducción

- Antecedentes y contexto del proyecto.

SpaceX es la compañía más exitosa de la era espacial comercial, lo que facilita los viajes espaciales. La compañía anuncia el lanzamiento del cohete Falcon 9 en su sitio web, con un costo de 62 millones de dólares; otros proveedores cuestan más de 165 millones de dólares cada uno. Gran parte del ahorro se debe a que SpaceX puede reutilizar la primera etapa. Por lo tanto, si podemos determinar si la primera etapa aterrizará, podemos determinar el costo de un lanzamiento. Con base en información pública y modelos de aprendizaje automático, prediciremos si SpaceX reutilizará la primera etapa.

- Problemas a los que quieres encontrar respuestas.

¿Cómo influyen variables como la masa de la carga útil, el lugar de lanzamiento, el número de vuelos y las órbitas en el éxito del aterrizaje de la primera etapa?

- ¿Aumenta la tasa de aterrizajes exitosos con el tiempo?

- ¿Cuál es el mejor algoritmo para la clasificación binaria?

Section 1

Methodology

Metodología

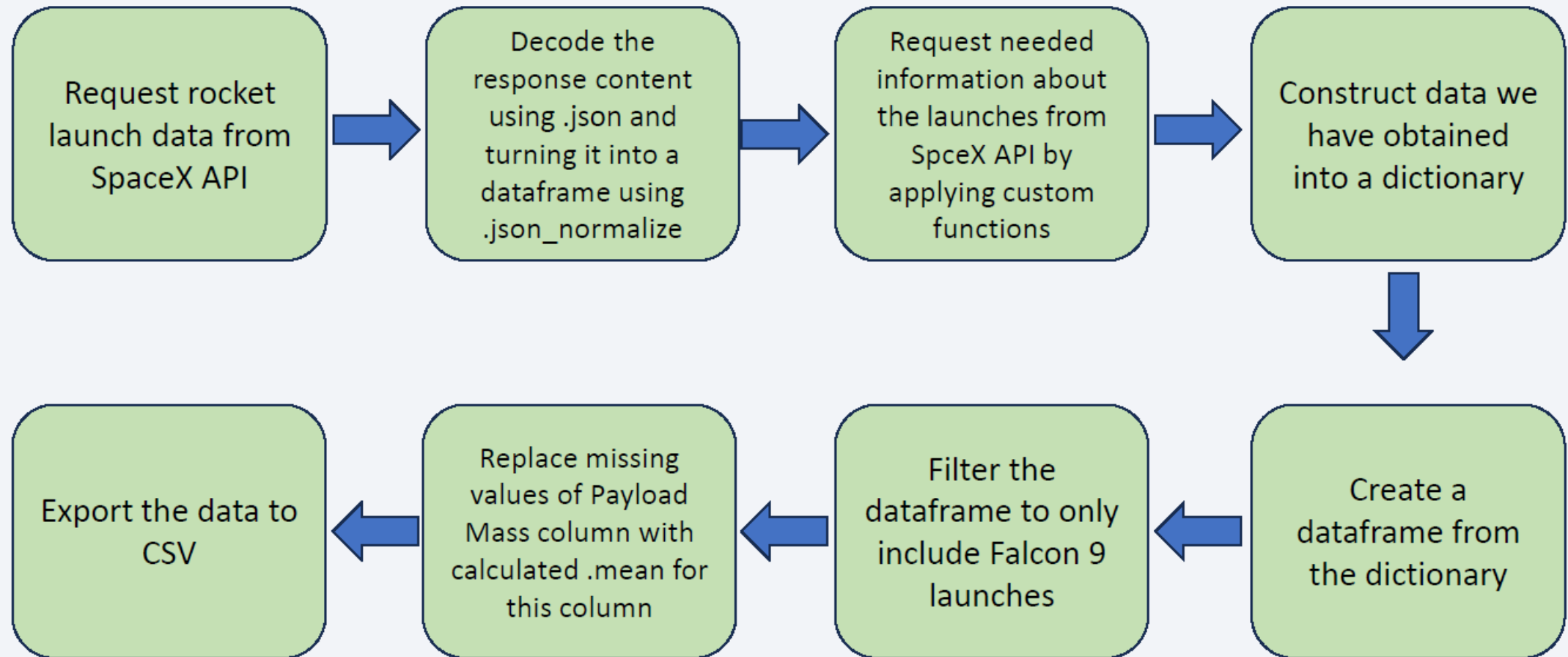
Reseumen ejecutivo

- Metodología de recopilación de datos:
 - Uso de la API Rest de SpaceX
 - Uso de Web Scraping de Wikipedia
- Realizar la manipulación de datos
 - Filtro de datos
 - Cómo lidiar con los valores faltantes
 - Uso de One Hot Encoding para preparar los datos para una clasificación binaria
- Realizar análisis exploratorios de datos (EDA) utilizando visualización y SQL
- Realizar análisis visuales interactivos utilizando Folium y PlotlyDash
- Realizar análisis predictivos utilizando modelos de clasificación
 - Realizar análisis predictivos utilizando modelos de clasificación

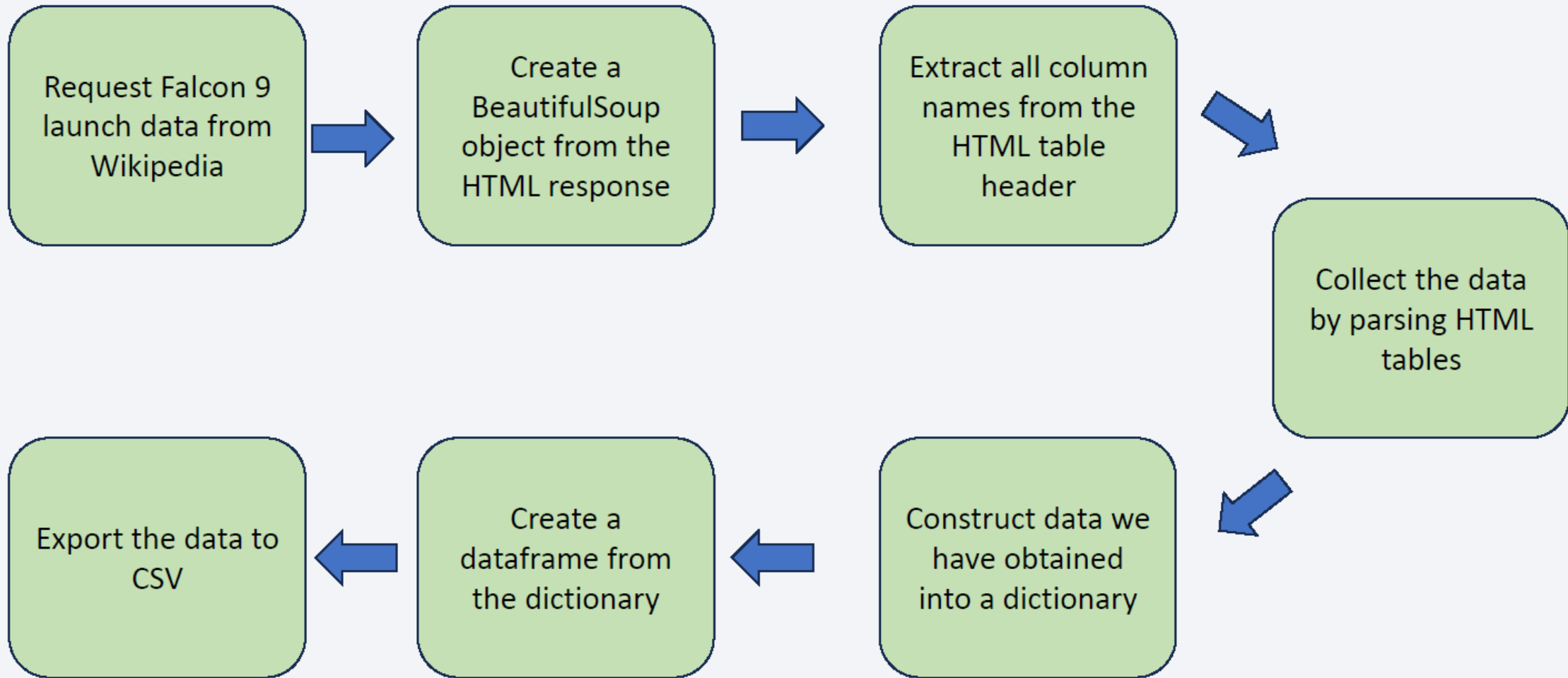
Recopilación de datos

- Solicitudes de API de la API REST de SpaceX
- Número de vuelo, fecha, versión del amplificador, masa de la carga útil, órbita, lugar de lanzamiento, resultado, etc.
- Extracción web de datos de una tabla en la página de Wikipedia de SpaceX
- Número de vuelo, lugar de lanzamiento, masa de la carga útil, órbita, resultado, etc.

Recopilación de datos – API de SpaceX



Recopilación de datos - Scraping



Disputa de datos

- El conjunto de datos muestra varios casos diferentes en los que el cohete no aterrizó con éxito. En ocasiones, se intentó un aterrizaje y falló debido a un accidente.
 - "Océano Verdadero" significa que la misión aterrizó con éxito en una región específica del océano, mientras que "Océano Falso" significa que la misión no aterrizó con éxito en una región específica del océano.
 - "RTLS Verdadero" significa que la misión aterrizó con éxito en una plataforma de aterrizaje.
 - "RTLS Falso" significa que la misión aterrizó sin éxito en una plataforma de aterrizaje.
 - "ASDS Verdadero" significa que la misión aterrizó con éxito en una nave completa. "ASDS Falso" significa que la misión aterrizó sin éxito en una nave no tripulada.
 - Estas etiquetas de entrenamiento se convierten en "1" o "0", donde "1" significa un aterrizaje exitoso y "0" significa un aterrizaje fallido.
1. Realizar análisis exploratorios de datos y determinar las etiquetas de entrenamiento.
 2. Calcular el número de lanzamientos en cada sitio.
 3. Calcular el número y las ocurrencias de cada órbita.
 4. Calcular el número de ocurrencias del resultado de la misión por tipo de órbita.
 5. Crear una etiqueta de resultado de aterrizaje a partir de la columna "Resultado".
 6. Exportar los datos a CSV.

EDA con visualización de datos

- Gráficos trazados:

Número de vuelo vs. masa de la carga útil

Payload Mass vs Launch Site

Masa de la carga útil vs. tipo de órbita

Número de vuelo vs. lugar de lanzamiento

Orbit Type vs Success Rate

Número de vuelo vs. tipo de órbita

<https://github.com/HobbieRS4/Capstone/blob/main/edadataviz.ipynb>

EDA con SQL

- Consultas SQL realizadas:

- Mostrar los nombres de los sitios de lanzamiento únicos en la misión espacial.
- Mostrar 5 registros donde los sitios de lanzamiento comiencen con la cadena "CCA".
- Mostrar la masa total de la carga útil transportada por los propulsores lanzados por la NASA (CRS).
- Mostrar la masa promedio de la carga útil transportada por la versión F9 v1.1 del propulsor.
- Indicar la fecha del primer aterrizaje exitoso en una plataforma terrestre.
- Indicar los nombres de los propulsores que aterrizaron con éxito en una nave no tripulada con una masa de carga útil superior a 4000 pero inferior a 6000.
- Indicar el número total de misiones exitosas y fallidas.
- Indicar los nombres de las versiones de propulsores que transportaron la máxima masa de carga útil.
- Indicar los aterrizajes fallidos en naves no tripuladas, junto con sus versiones de propulsor y los nombres del sitio de lanzamiento, en cada mes de 2015.
- Clasificar el número de aterrizajes (como "Fracaso" (nave no tripulada) o "Éxito" (plataforma terrestre)) entre el 4 de junio de 2010. y el 20 de marzo de 2017 en orden descendente

https://github.com/HobbieRS4/Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Construye un mapa interactivo con Folium

Marcadores de todos los sitios de lanzamiento Explain why you added those objects

- Se agregó un marcador con un círculo, una etiqueta emergente y una etiqueta de texto del Centro Espacial Johnson de la NASA que utiliza sus coordenadas de latitud y longitud como ubicación de inicio.
- Se agregaron marcadores con círculo, etiqueta emergente y etiqueta de texto de todos los sitios de lanzamiento utilizando sus coordenadas de latitud y longitud para mostrar sus ubicaciones geográficas y proximidad al Ecuador y las costas.

Marcadores de colores de los resultados del lanzamiento para cada sitio de aterrizaje

- Se agregaron marcadores de colores de lanzamientos exitosos (verde) y fallidos (rojos) usando Marker Cluster para identificar qué sitios de lanzamiento tienen tasas de éxito relativamente altas

Distances Between A Launch Site to its Proximities

- Se agregaron líneas de colores para mostrar las distancias entre el sitio de lanzamiento KSC LC-39A (como ejemplo) y sus proximidades, como ferrocarril, autopista, costa y ciudad más cercana.

Cree un panel de control con Plotly Dash

Lista desplegable de sitios de lanzamiento:

- Se agregó una lista desplegable para habilitar la selección del sitio de lanzamiento

Gráfico circular que muestra los lanzamientos exitosos (todos los sitios/ciertos sitios)

- Se agregó un gráfico circular para mostrar el recuento total de lanzamientos exitosos para todos los sitios y el recuento de éxitos vs. fallidos para el sitio, si se seleccionó un sitio de lanzamiento específico

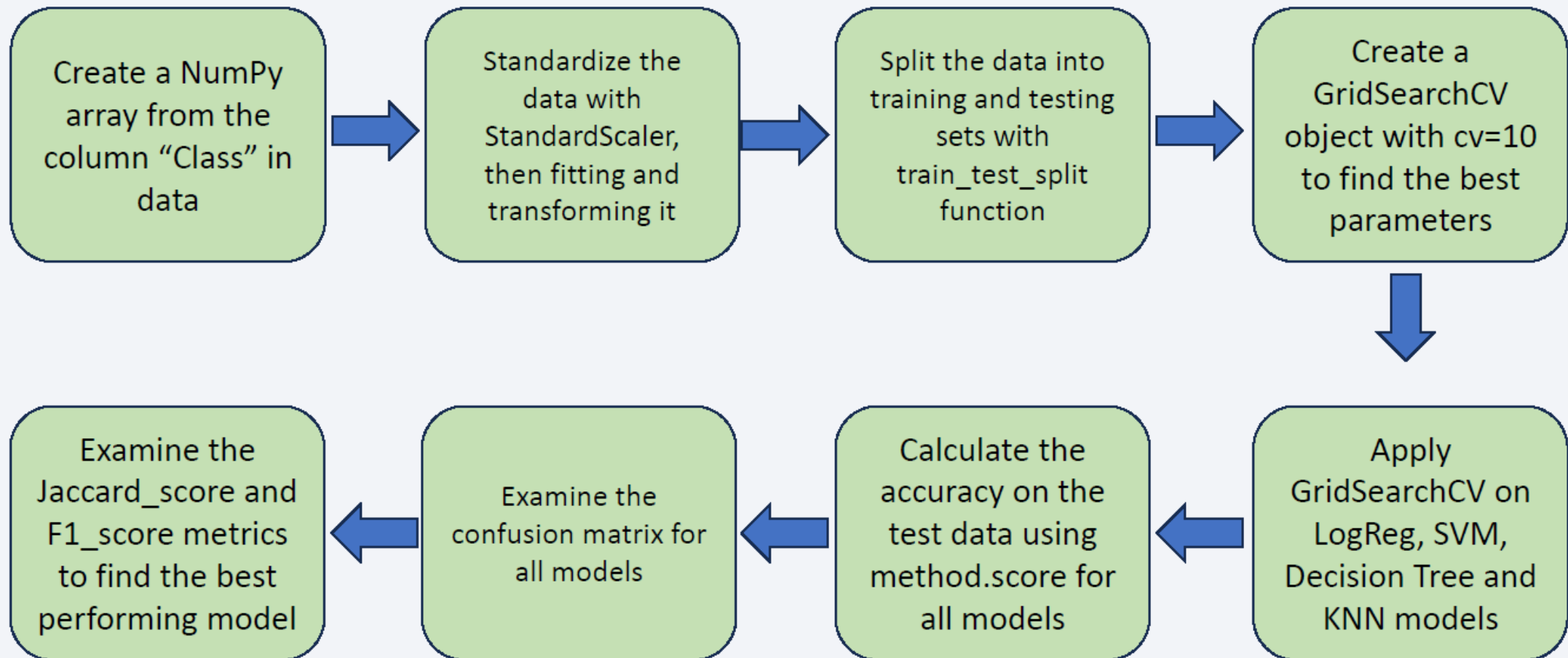
Control deslizable del rango de masa de carga útil

- Se agregó un control deslizable para seleccionar el rango de carga útil.

Diagrama de dispersión de la masa de la carga útil frente a la tasa de éxito para diferentes versiones del propulsor

- Se agregó un gráfico de dispersión para mostrar la correlación entre la carga útil y el éxito del lanzamiento.

Predictive Analysis (Classification)



Resultados

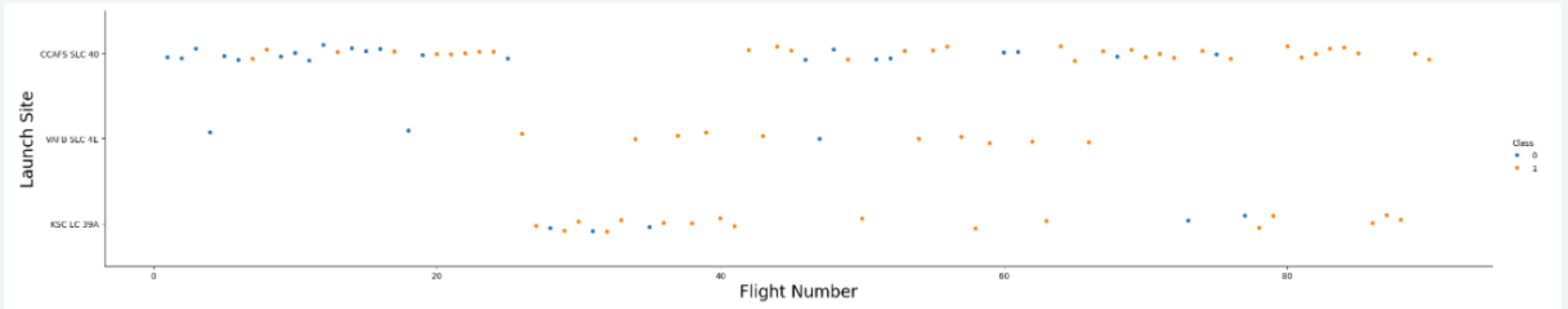
- Resultados del análisis exploratorio de datos
- Demostración de análisis interactivo en capturas de pantalla
- Resultados del análisis predictivo

The background of the slide is an abstract composition. It features a dark blue base color. Overlaid on this are numerous diagonal streaks in shades of red and cyan. A faint, light blue grid pattern is also visible, particularly in the lower half of the image. The overall effect is dynamic and technological.

Section 2

Insights drawn from EDA

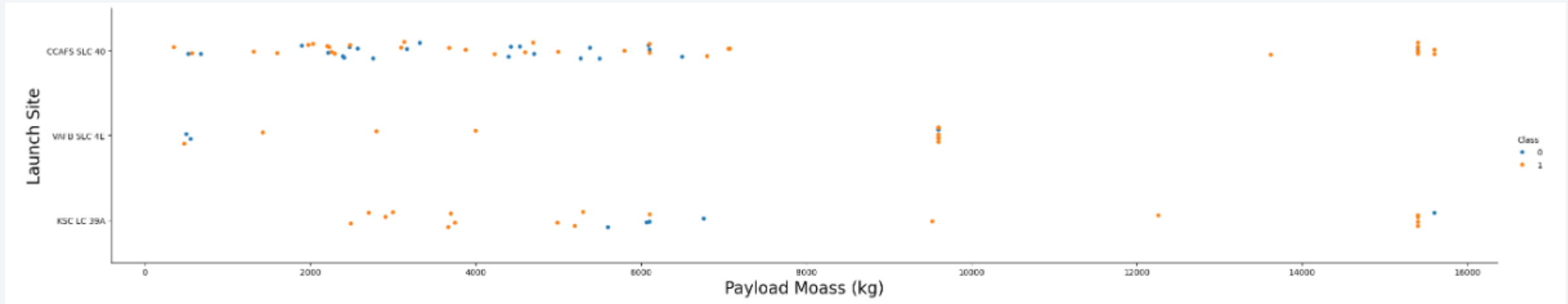
Número de vuelo vs. Lugar de lanzamiento



Resultados

- Los primeros vuelos fracasaron y los posteriores tuvieron éxito.
- El sitio de lanzamiento CCAFS SLC 40 registra aproximadamente la mitad de todos los lanzamientos.
- VAFB SLC 4E y KSC LC 39A tienen tasas de éxito más altas.

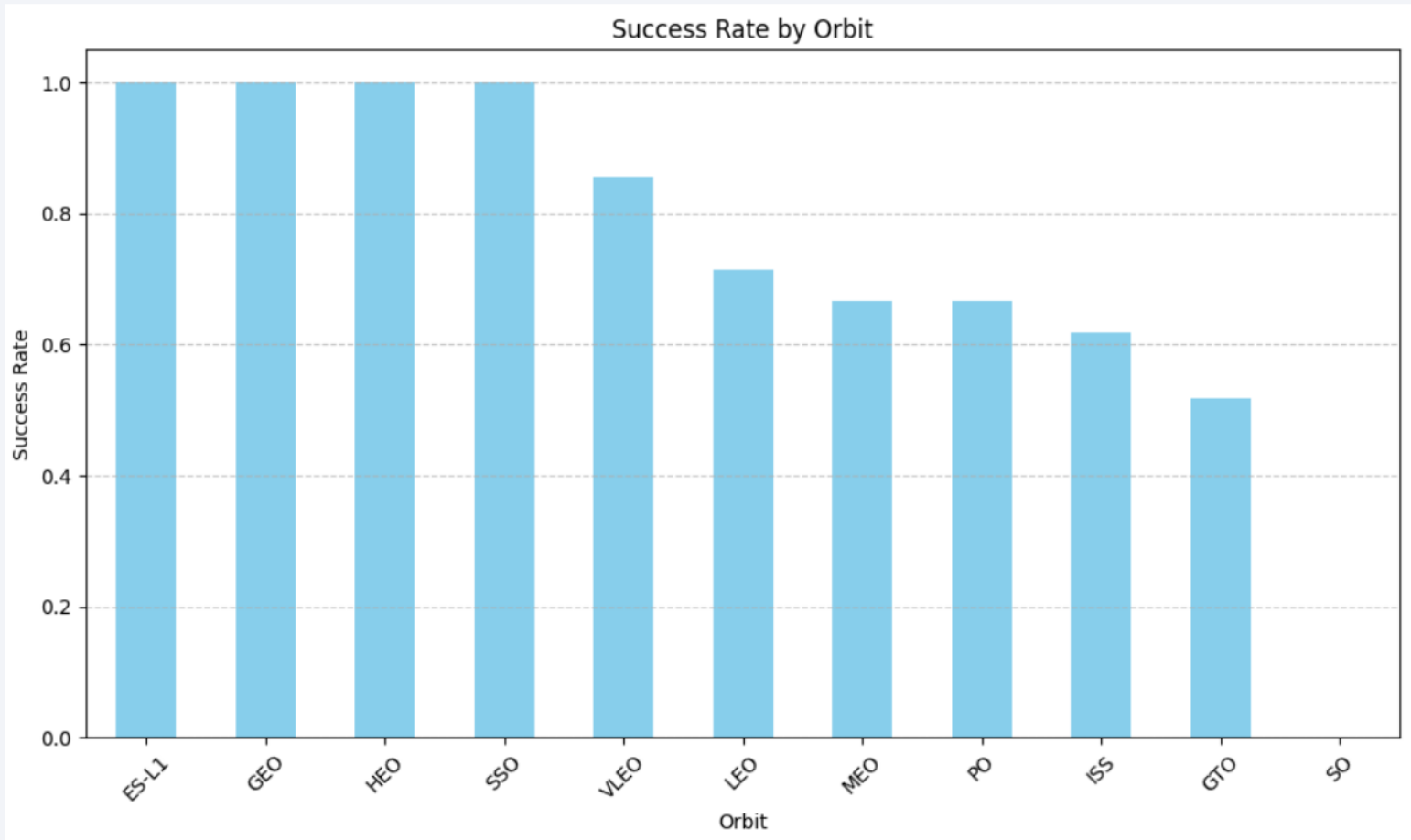
Carga útil vs. Lugar de lanzamiento



Resultados:

- Las cargas útiles más altas tienen una mayor tasa de éxito que las cargas más bajas en todos los sitios.
- El sitio de lanzamiento KSC LC 39A tiene una alta tasa de éxito para todas las cargas útiles superiores a 8000 kg.
- El sitio de lanzamiento CCAFS SLC 40 tiene una tasa de éxito inconsistente con cargas útiles inferiores a 9000 kg.

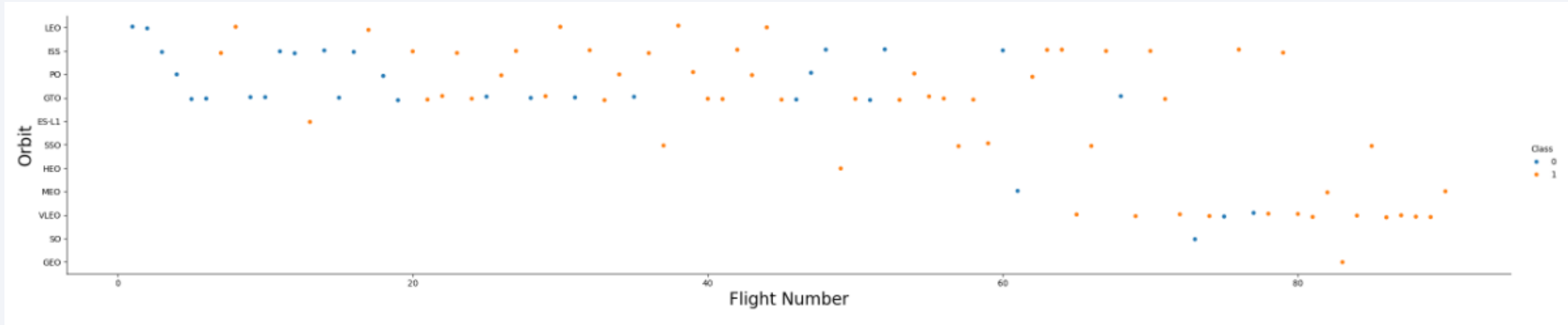
Tasa de éxito vs. tipo de órbita



Observaciones

- Los lanzamientos a órbitas ES-L1, GEO, HEO y SSO tienen una tasa de éxito del 100 %.
- Los lanzamientos a órbitas SO tienen una tasa de éxito del 0 %.
- Los lanzamientos a VLEO, LEO, MEO, PO, ISS y GTO tienen una tasa de éxito de entre el 50 % y el 85 %.

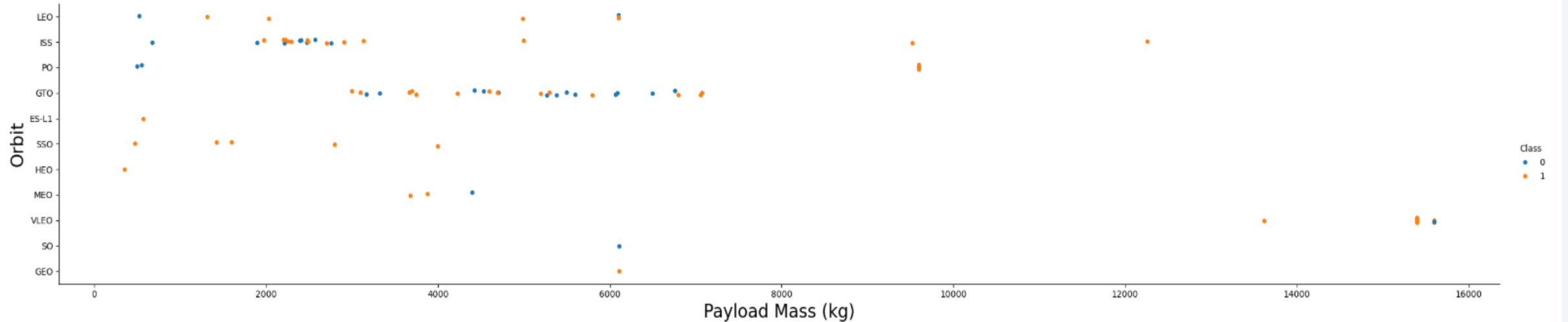
Número de vuelo vs. tipo de órbita



Observaciones

- No se intentó orbitar GEO, SO ni VLEO hasta bien entrada la fase de prueba.
- La tasa de éxito en órbitas LEO mejora con el tiempo.
- La tasa de éxito en órbitas de la ISS mejora, luego presenta fallos antes de tener más éxito. [21](#)

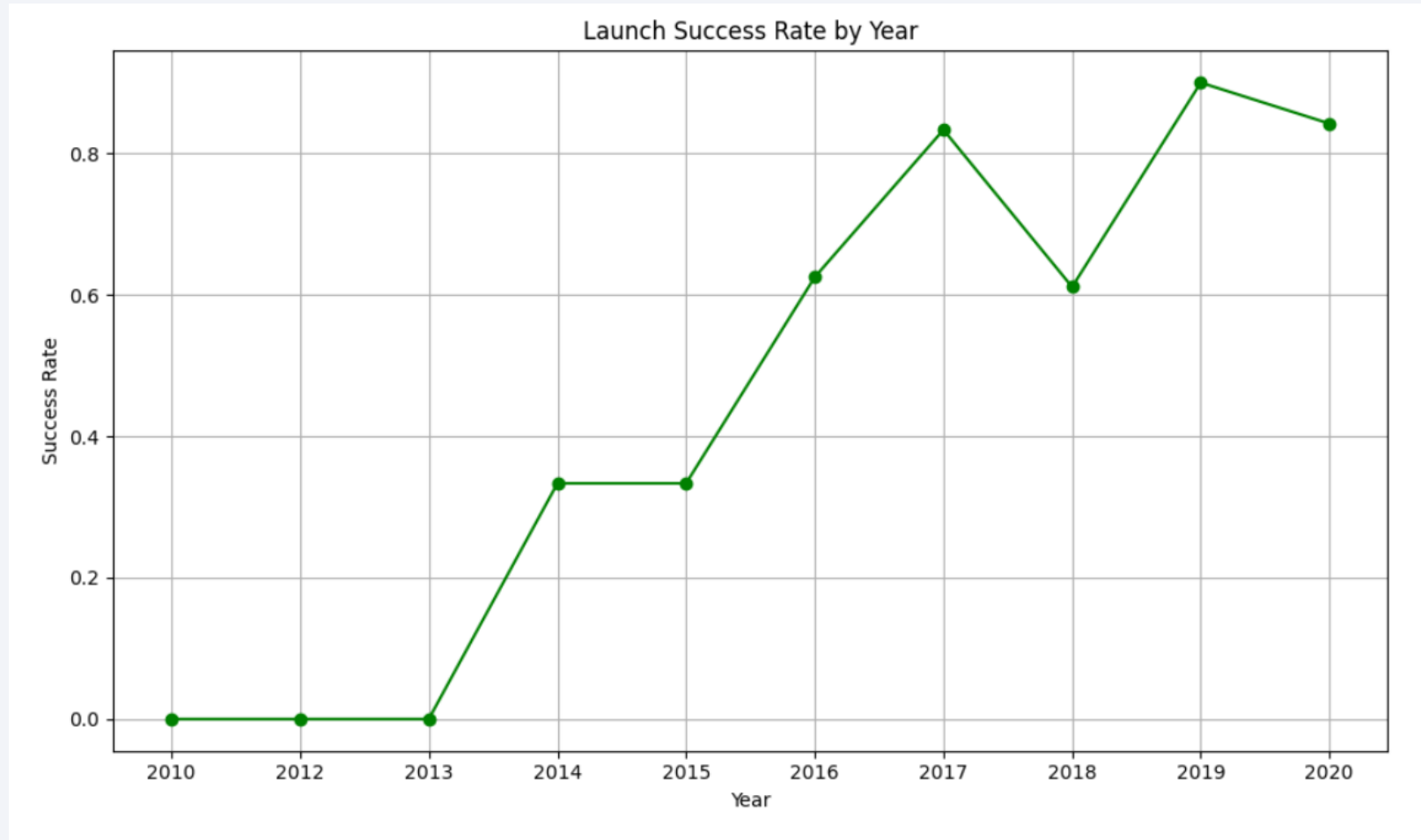
Carga útil vs. Tipo de órbita



Observaciones

- Las cargas útiles más pesadas tienen mayor éxito independientemente de la órbita.
- Las cargas útiles de aproximadamente 2000 kg generalmente tienen éxito independientemente de la órbita.
- Las cargas útiles de aproximadamente 4000 kg generalmente tienen éxito independientemente de la órbita.

Tendencia anual de éxito de lanzamiento



Observaciones

- El éxito de los lanzamientos ha mejorado con el tiempo.
- Hubo una caída de más del 20 % en la tasa de éxito entre 2017 y 2018, antes de recuperarse.

Todos los nombres de los sitios de lanzamiento

```
%sql select distinct launch_site from SPACEXDATASET;
```

```
* ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90108kqblod8lcg.databases.appdomain.cloud:31198/bludb  
Done.
```

launch_site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

- Nombres de sitios de lanzamiento
 - CCAFS LC-40
 - CCAFS SLC-40
 - KSC LC-39A
 - VAFB SLC-4E

Los nombres de los sitios de lanzamiento comienzan con 'CCA'

5 registros de lanzamientos desde sitios que empiezan con «CCA»

```
%sql select * from SPACEXDATASET where launch_site like 'CCA%' limit 5;
```

```
* ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90108kqblod8lclg.databases.appdomain.cloud:31198/bludb  
Done.
```

DATE	time_utc	booster_version	launch_site	payload	payload_mass_kg	orbit	customer	mission_outcome	landing_outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Masa total de carga útil

Masa total de carga útil transportada por los propulsores lanzados por la NASA

```
%sql select sum(payload_mass__kg_) as total_payload_mass from SPACEXDATASET where customer = 'NASA (CRS)';
```

```
* ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqblod8lcg.databases.appdomain.cloud:31198/bludb
```

Done.

total_payload_mass
45596

Masa de carga útil promedio por F9 v1.1

Masa promedio de carga útil transportada por el propulsor versión F9 v1.1

```
%sql select avg(payload_mass__kg_) as average_payload_mass from SPACEXDATASET where booster_version like '%F9 v1.1%';  
  
* ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90108kqblod8lcg.databases.appdomain.cloud:31198/bludb  
Done.
```

average_payload_mass
2534

Fecha del primer aterrizaje terrestre exitoso

Fecha del primer aterrizaje terrestre exitoso

```
%sql select min(date) as first_successful_landing from SPACEXDATASET where landing__outcome = 'Success (ground pad)';  
* ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90108kqblod8lcg.databases.appdomain.cloud:31198/bludb  
Done.
```

first_successful_landing
2015-12-22

Aterrizaje exitoso de un dron con una carga útil de entre 4000 y 6000

```
%sql select booster_version from SPACEXDATASET where landing__outcome = 'Success (drone ship)' and payload_mass__kg_ between 4000 and 6000;
```

```
* ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90108kqblod8lcg.databases.appdomain.cloud:31198/bludb  
Done.
```

booster_version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and FailureMission Outcomes

```
%sql select mission_outcome, count(*) as total_number from SPACEXDATASET group by mission_outcome;
```

```
* ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90108kqblod8lcg.databases.appdomain.cloud:31198/bludb  
Done.
```

mission_outcome	total_number
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Los propulsores transportaron la carga útil máxima

```
%sql select booster_version from SPACEXDATASET where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXDATASET);  
* ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90108kqblod8lcg.databases.appdomain.cloud:31198/bludb  
Done.
```

booster_version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

Récords de lanzamiento de 2015

```
%%sql select monthname(date) as month, date, booster_version, launch_site, landing__outcome from SPACEXDATASET
      where landing__outcome = 'Failure (drone ship)' and year(date)=2015;
```

```
* ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90108kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

MONTH	DATE	booster_version	launch_site	landing__outcome
January	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
April	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

Resultados de la clasificación entre el 04/06/2010 y el 20/03/2017

```
%%sql select landing__outcome, count(*) as count_outcomes from SPACEXDATASET
where date between '2010-06-04' and '2017-03-20'
group by landing__outcome
order by count_outcomes desc;
```

```
* ibm_db_sa://wzf08322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90108kqblod8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

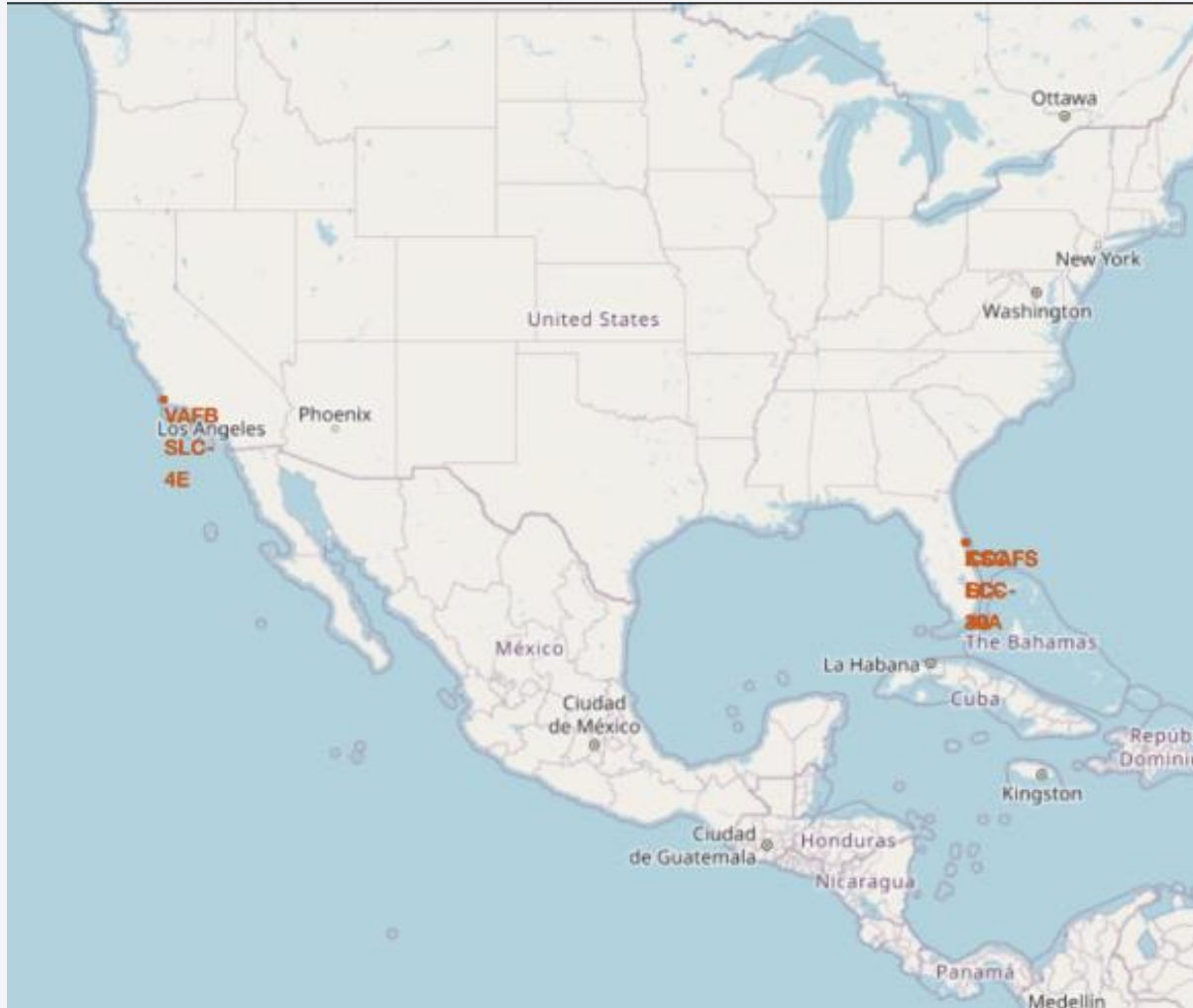
landing__outcome	count_outcomes
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

A satellite view of Earth from space, showing the curvature of the planet and city lights at night. The background is a deep blue gradient.

Section 3

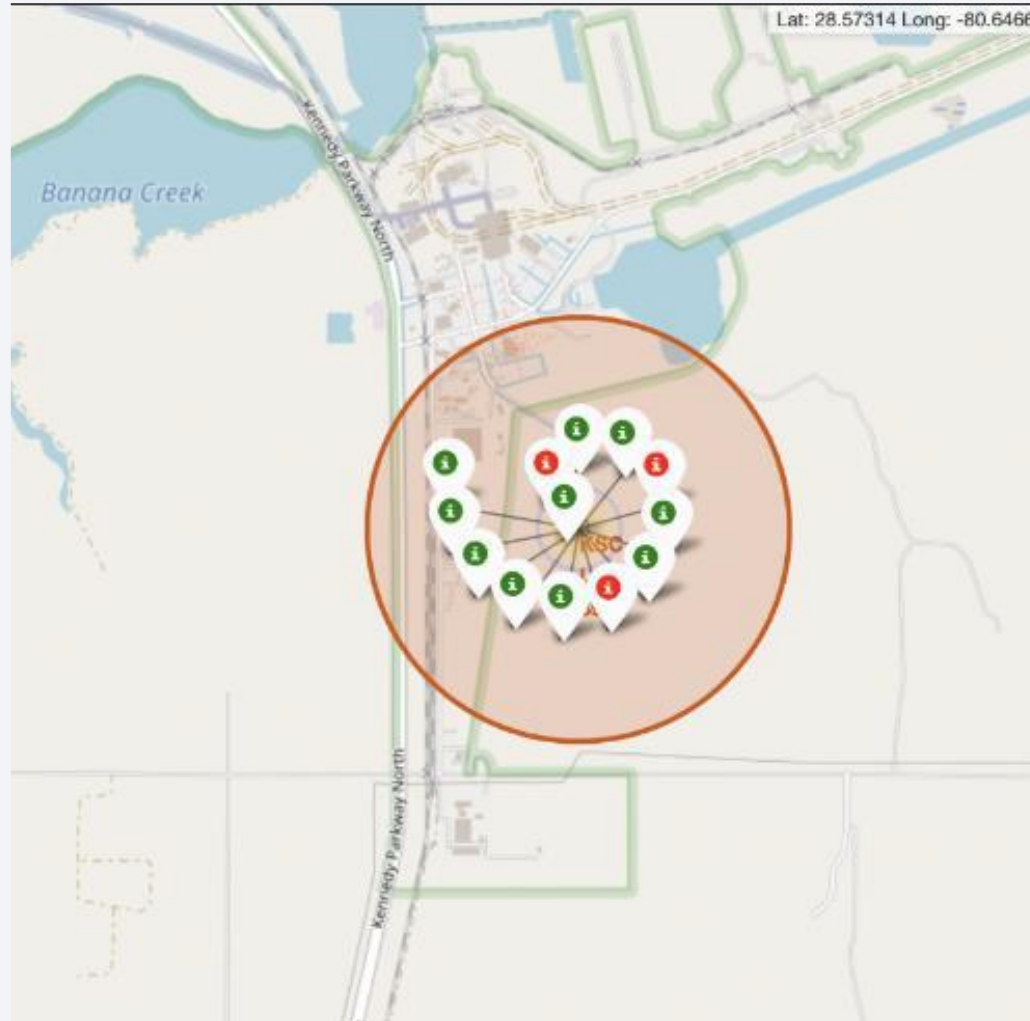
Launch Sites Proximities Analysis

Ubicación de los sitios de lanzamiento



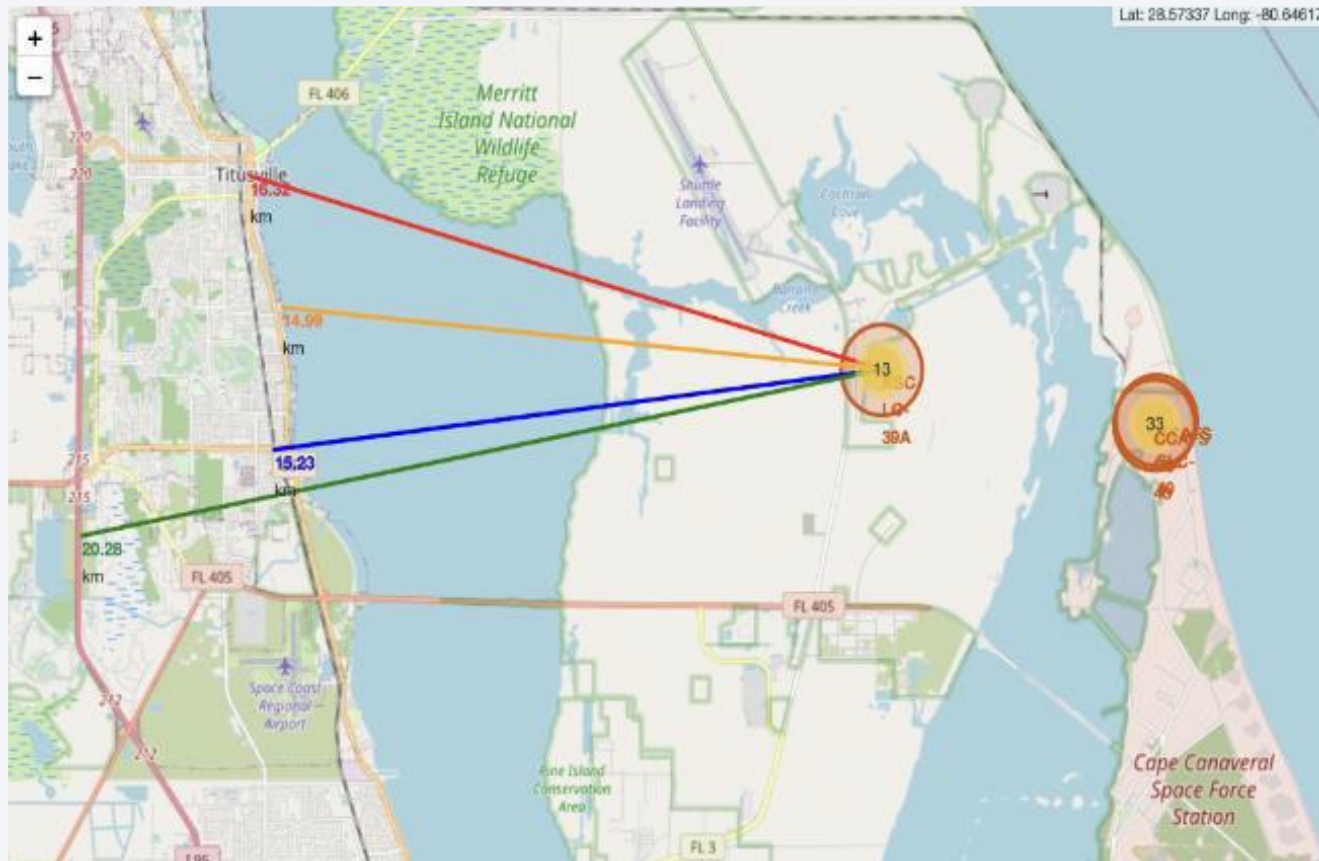
- Sitios de lanzamiento ubicados cerca del ecuador
 - El centro de la Tierra gira más rápido que cualquier otro lugar, aproximadamente a 1670 km/h.
 - Un objeto lanzado al espacio orbitará la Tierra a la misma velocidad que se movía antes del lanzamiento.
- Sitios de lanzamiento ubicados cerca de la costa
 - El lanzamiento de objetos hacia el océano minimiza el riesgo de caída o explosión de escombros cerca de las personas.

Lanzamientos desde el sitio KSC LC-39A



- El sitio de lanzamiento del KSC LC-39A tiene una alta tasa de éxito.
- Los marcadores verdes indican lanzamientos exitosos.
- Los marcadores rojos indican lanzamientos fallidos.

KSC LC-39A Proximidad a lugares de interés locales



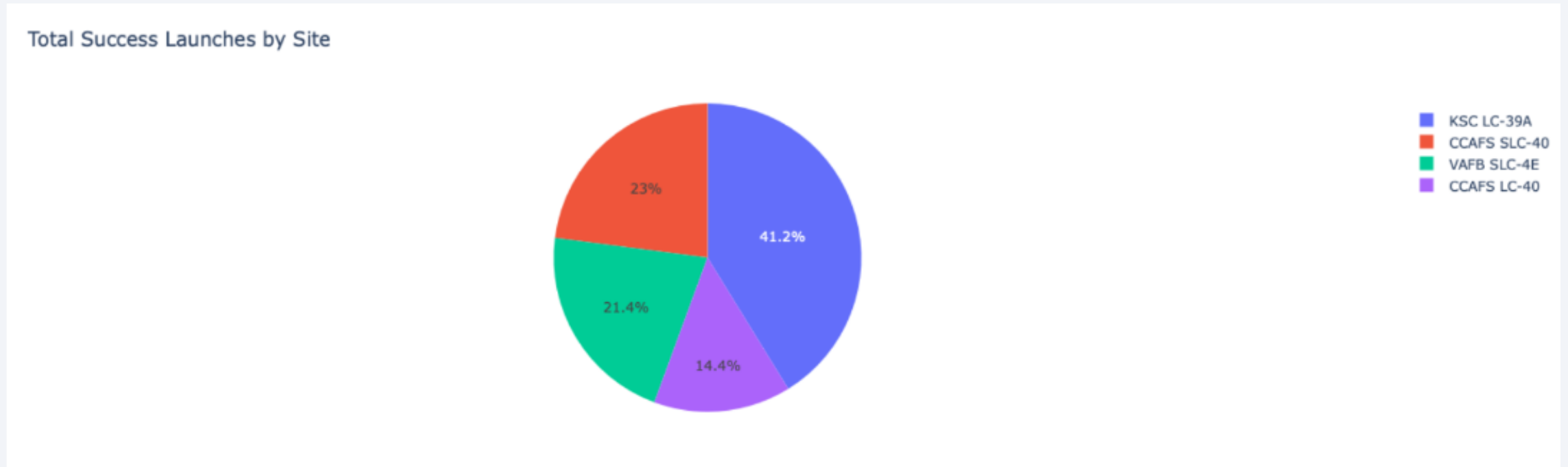
- 15,23 km hasta la vía férrea
- 20,28 km hasta la autopista
- 14,99 km hasta la costa
- 16,32 km hasta la ciudad más cercana:
Titusville



Section 4

Build a Dashboard with Plotly Dash

Tasa de éxito de lanzamiento para cada sitio



- Sitio con mayor éxito: KSC LC-39^a
- CCAFS SLC-40 y VAFB SCL-4E presentan tasas de éxito prácticamente iguales

Success Ratio for KSC LC-39A

Total Success Launches for Site KSC LC-39A



- La tasa de éxito más alta de todos los sitios
- 10 aterrizajes exitosos vs. 3 fracasos

Launch Outcome vs Payload Mass



➤ Las cargas útiles entre 2.000 y 5.500 kg tienen la mayor tasa de éxito.



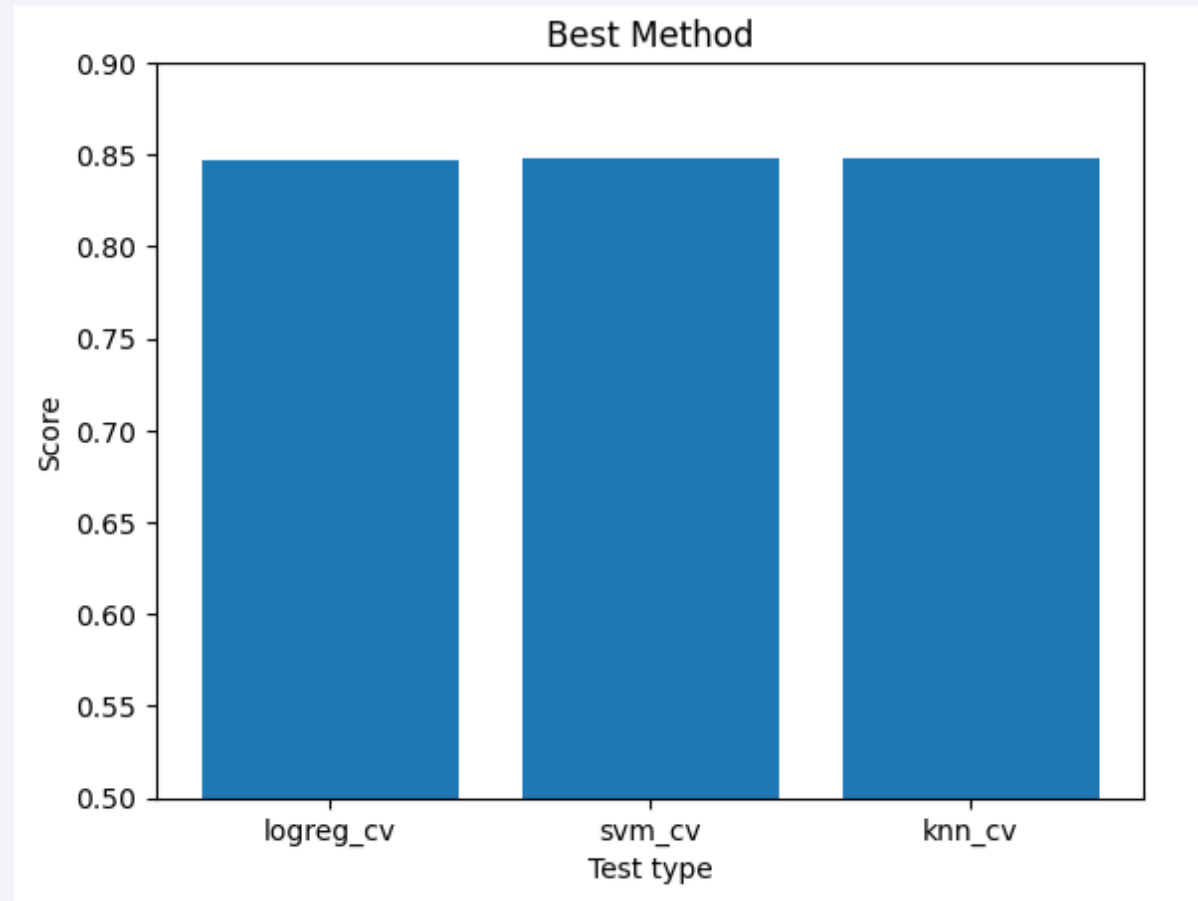


Section 5

Predictive Analysis (Classification)

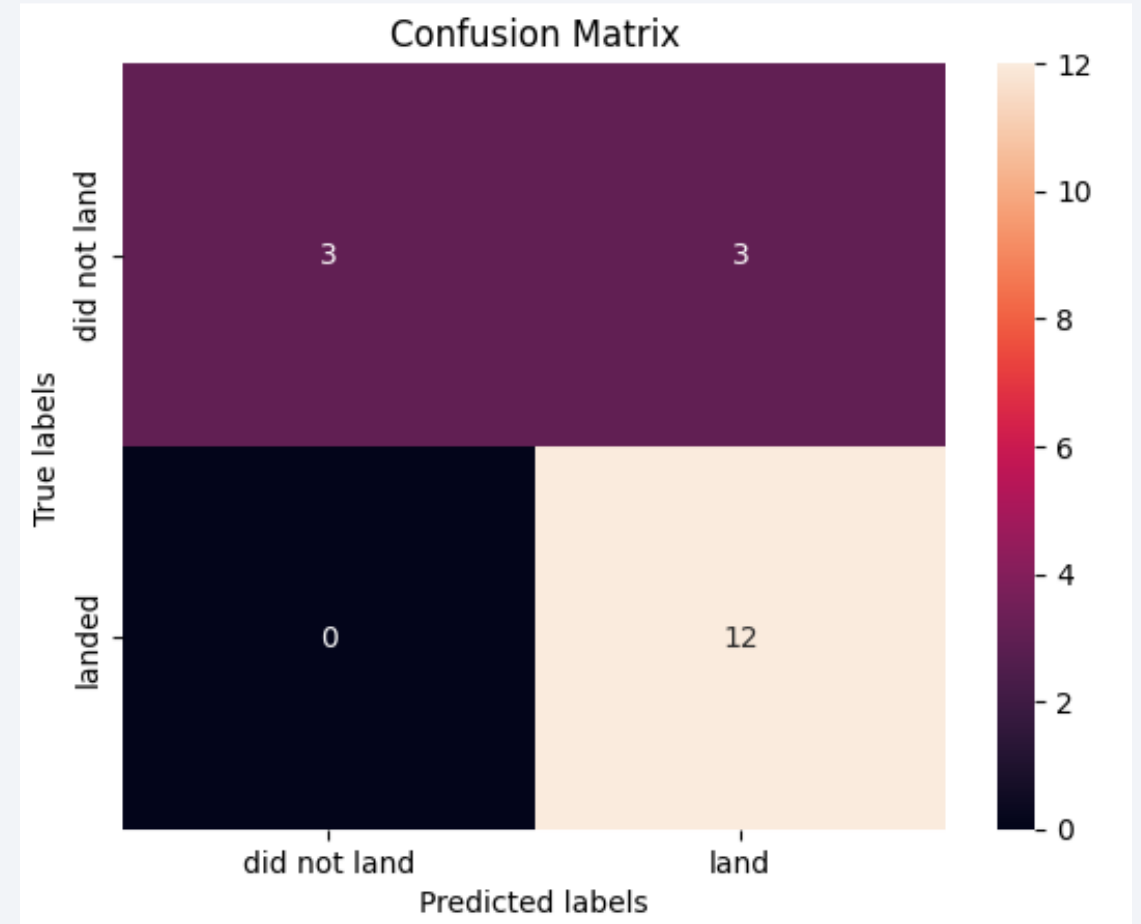
Precisión de la clasificación

- Visualice la precisión del modelo generado para todos los modelos de clasificación generados en un gráfico de barras.
- Encuentre el modelo con la mayor precisión de clasificación.



Matriz de confusión

- 12 Verdaderos positivos (la etiqueta verdadera es tierra, la etiqueta predicha es tierra)
- 3 Falsos positivos (la etiqueta verdadera no es tierra, la etiqueta predicha es tierra)
- 3 Verdaderos negativos (la etiqueta verdadera no es tierra, la etiqueta predicha no es tierra)
- 0 Falso negativo (la etiqueta verdadera no es tierra, la etiqueta predicha no es tierra)



Conclusiones

- La mayoría de los sitios de lanzamiento se encuentran cerca del ecuador y de una costa.
- La tasa de éxito de todos los lanzamientos ha aumentado con el tiempo.
- El sitio de lanzamiento KSC LC-39A tiene la mayor tasa de éxito.
- 100 % de éxito en lanzamientos a las órbitas ES-L1, GEO, HEO y SSO.

Apéndice

- Incluya cualquier activo relevante como fragmentos de código Python, consultas SQL, gráficos, resultados de Notebook o conjuntos de datos que pueda haber creado durante este proyecto.

Thank you!

