
# From Data Science to Space

Mario Mompean Herrero

Marzo 2023

# OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

# EXECUTIVE SUMMARY

- Data Collection:
  - With API
  - With Web Scraping
- Data Wrangling
- Data Analysis
  - With SQL
  - With Data Visualization
- Folium for Interactive Visual Analytics
- Predictions with ML

# INTRODUCTION

- En este proyecto hablamos sobre el rol de un científico de datos en una compañia de lanzamientos de cohetes.
- Analizaremos el trabajo de SPACE X.
- A partir del analisis de SPACE X buscaremos las prestaciones más rentables.
- Estudiaremos si los cohetes pasan la primera fase y utilizaremos modelos de ML para predecir posibles perspectivas futuras

# METHODOLOGY

- Data Collection
- Data Wrangling
- EDA and interactive visual analytics methodology
- EDA with Data Visualization
- EDA with SQL
- Predictive Analysis

# Data Collection

## With API

1. Tomamos el correspondiente archivo de formato JSON
2. Lo convertimos a un DF con Pandas
3. Construimos el Dataset

Pinche aquí

## With Web Scrapping

1. Extraemos los datos del HTML
2. Lo convertimos a un diccionario
3. Lo devolvemos a un CSV

Pinche aquí

# Data Wrangling

1. Calculamos los lanzamientos en cada lugar
2. Calculamos el número de veces de una órbita dada
3. Creamos la etiqueta para cada salida

Pinche aquí

# EDA with Data Visualization

Gráficos usados:

1. Dispersión
2. Líneas
3. Barras

Pinche aquí

# EDA with SQL

1. Cargamos el paquete SQL en Python

2. Exploramos los datos con sentencia SQL

Pinche aquí

# EDA and interactive visual analytics

1. Utilizamos las marcas del paquete 'Folium' de Python

2. Rojo representa fallo en lanzamiento

3. Verde representa éxito en el lanzamiento

Pinche aquí

# Predictive Analytics

1. Utilizamos el paquete 'Scikit-learn' de Python para usar algoritmos de ML

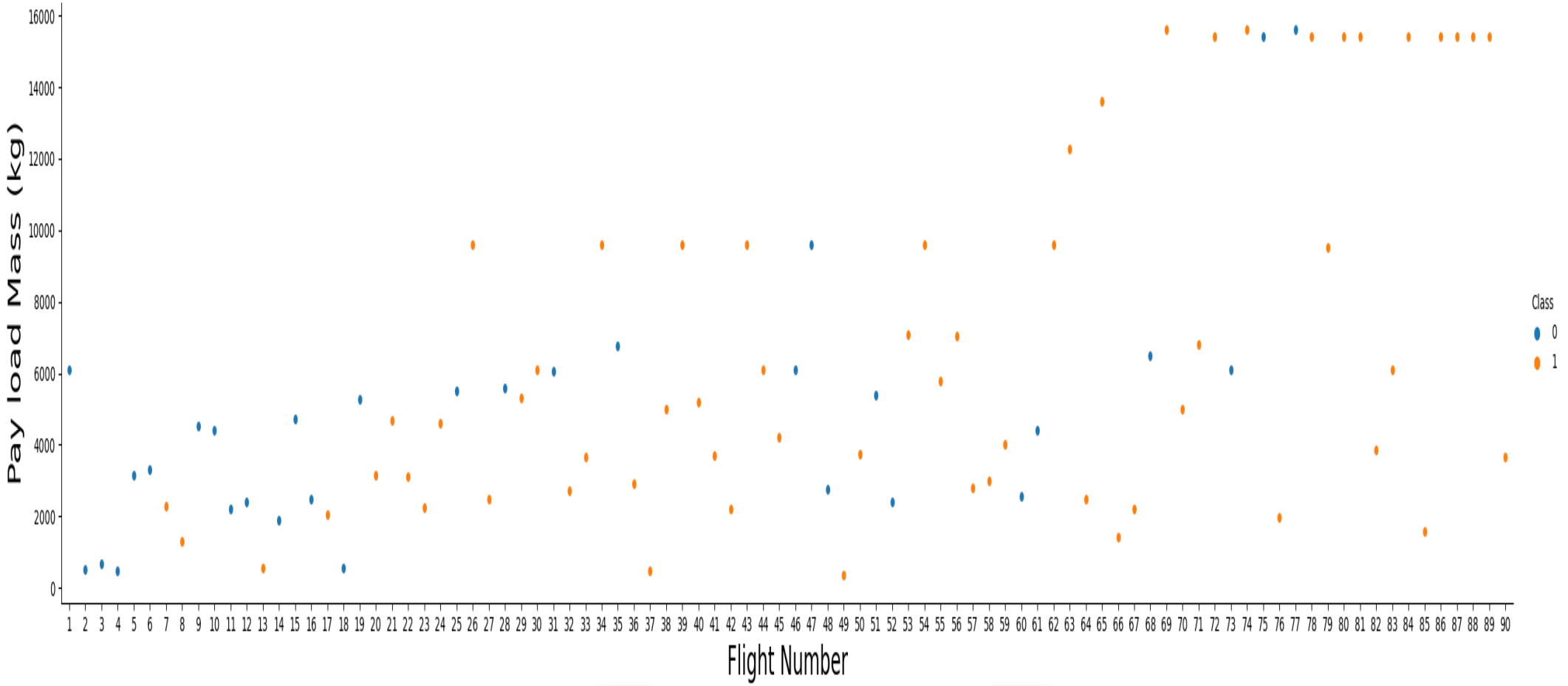
2. Comparamos la precisión de varios modelos de ML

Pinche aquí

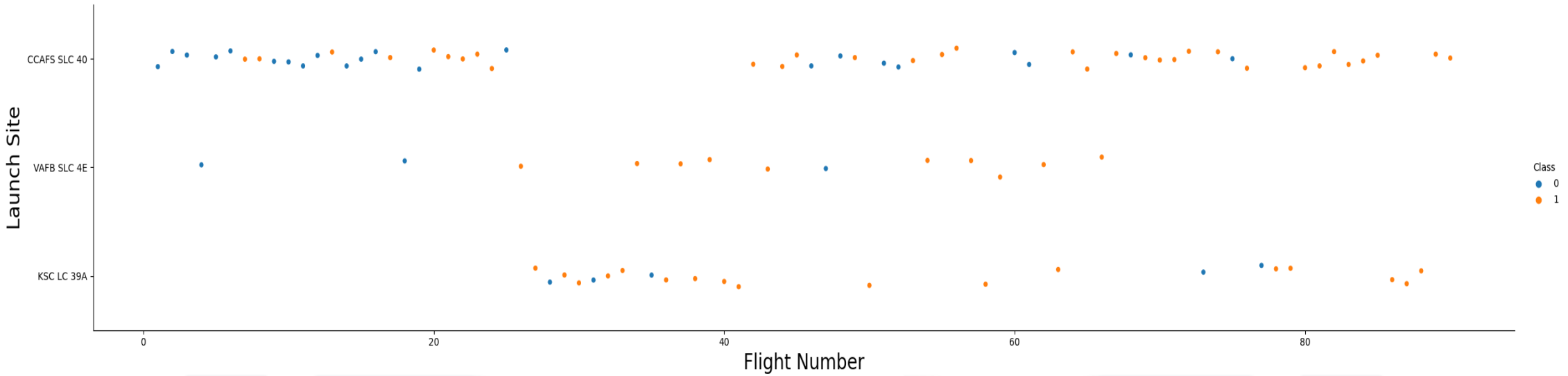
# RESULTS

- En esta sección analizamos los resultados de los datos tras aplicarles nuestras técnicas explicadas en la metodología

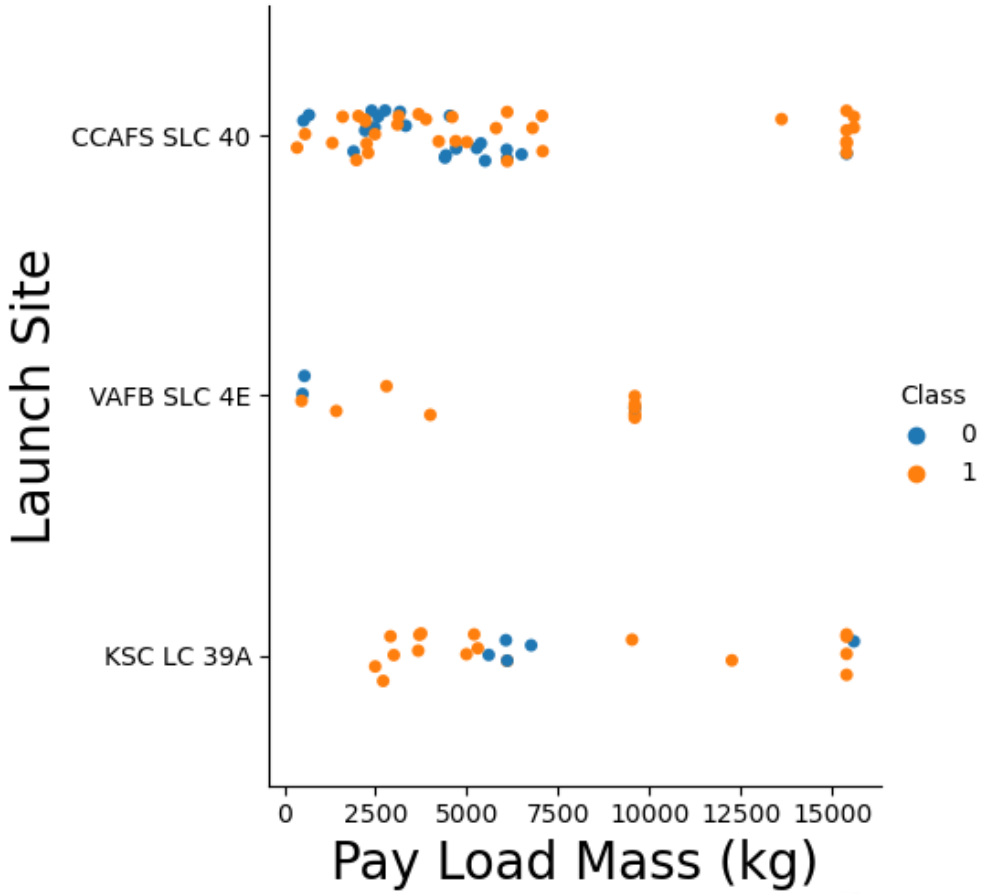
# EDA with Data Visualization

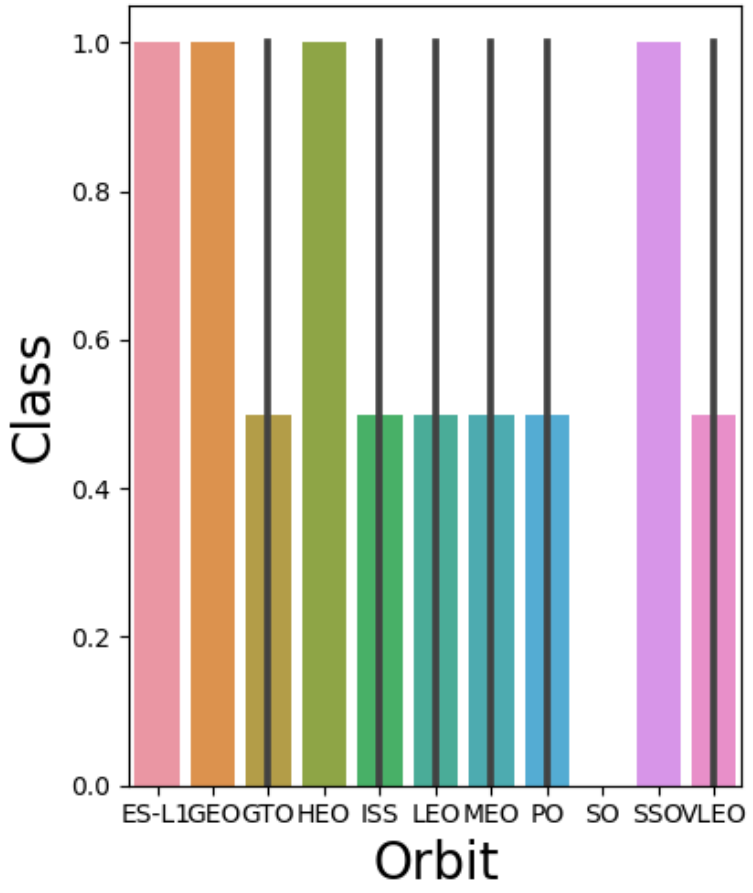


# EDA with Data Visualization

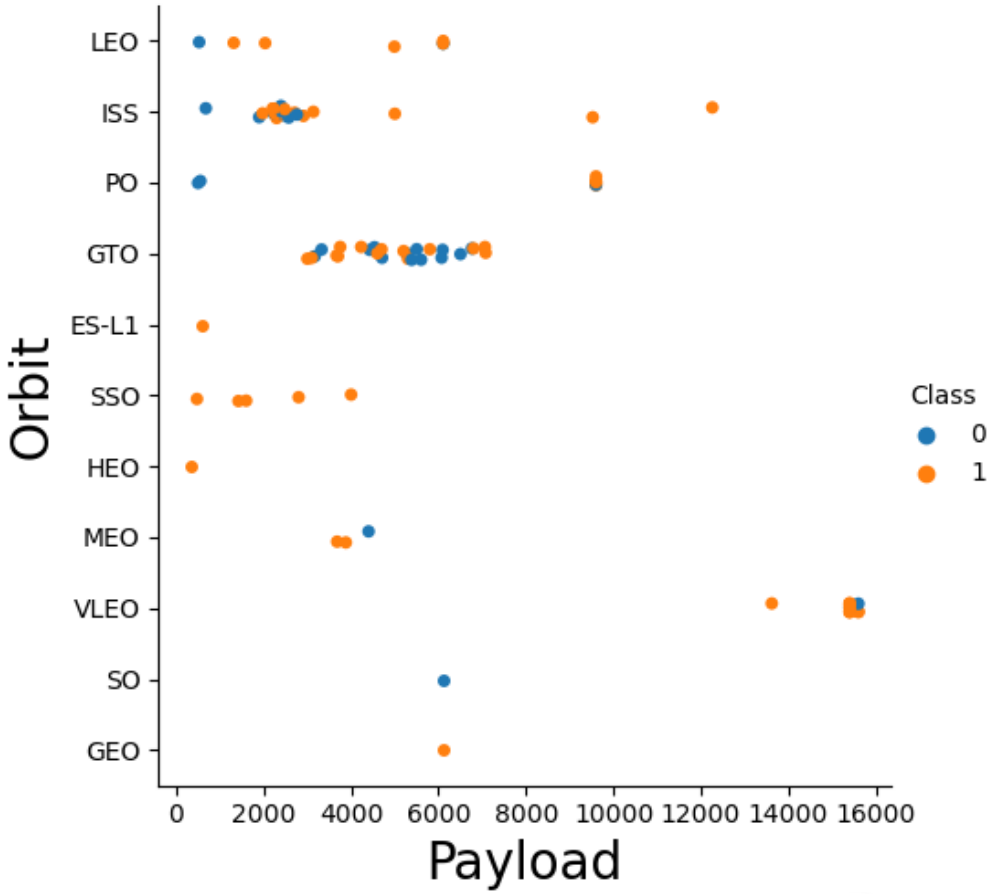


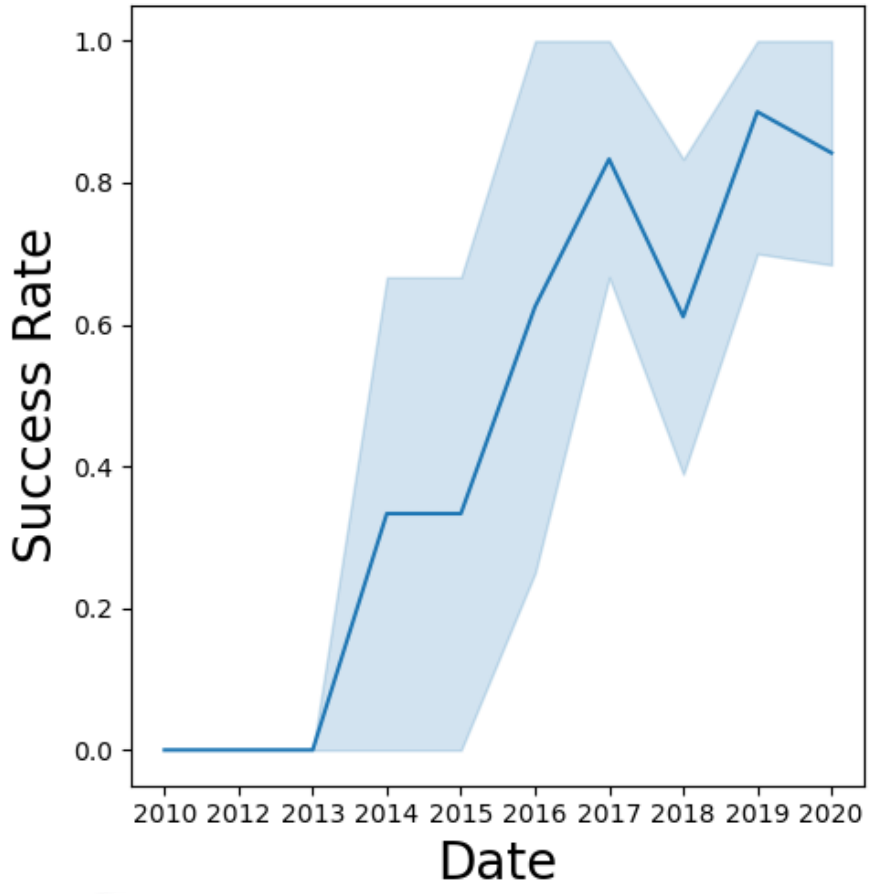
# EDA with Data Visualization



# EDA with Data Visualization



# EDA with SQL

### Task 1

Display the names of the unique launch sites in the space mission

In [11]:
```
%sql  SELECT DISTINCT LAUNCH_SITE  FROM SPACEXTBL;
```

```
 * sqlite:///my_data1.db
Done.
```

Out[11]:

| Launch_Site |
|---|
| CCAFS LC-40 |
| VAFB SLC-4E |
| KSC LC-39A |
| CCAFS SLC-40 |

### Task 2

Display 5 records where launch sites begin with the string 'CCA'

In [12]:
```
%sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;
```

```
 * sqlite:///my_data1.db
Done.
```

Out[12]:

| Date | Time (UTC) | Booster_Version | Launch_Site | Payload | PAYLOAD_MASS__KG_ | Orbit | Customer | Mission_Outcome | Landing _Outcome |
|---|---|---|---|---|---|---|---|---|---|
| 04-06-2010 | 18:45:00 | F9 v1.0 B0003 | CCAFS LC-40 | Dragon Spacecraft Qualification Unit | 0 | LEO | SpaceX | Success | Failure (parachute) |
| 08-12-2010 | 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) |
| 22-05-2012 | 07:44:00 | F9 v1.0 B0005 | CCAFS LC-40 | Dragon demo flight C2 | 525 | LEO (ISS) | NASA (COTS) | Success | No attempt |
| 08-10-2012 | 00:35:00 | F9 v1.0 B0006 | CCAFS LC-40 | SpaceX CRS-1 | 500 | LEO (ISS) | NASA (CRS) | Success | No attempt |
| 01-03-2013 | 15:10:00 | F9 v1.0 B0007 | CCAFS LC-40 | SpaceX CRS-2 | 677 | LEO (ISS) | NASA (CRS) | Success | No attempt |

### Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

In [31]:
```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)';
```

```
 * sqlite:///my_data1.db
Done.
```

Out[31]:

| SUM(PAYLOAD_MASS__KG_) |
|---|
| 45596 |

### Task 4

Display average payload mass carried by booster version F9 v1.1

In [14]:
```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1'
```

```
 * sqlite:///my_data1.db
Done.
```

Out[14]:

| AVG(PAYLOAD_MASS__KG_) |
|---|
| 2928.4 |

### Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

*Hint:Use min function*

In [51]:
```
%sql select min(DATE) from SPACEXTBL where "Landing__Outcome" = "Success (ground pad)";
```

```
 * sqlite:///my_data1.db
Done.
```

Out[51]:

| min(DATE) |
|---|
| None |

# EDA with SQL

### Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

In [26]:
```
%sql SELECT BOOSTER_VERSION from SPACEXTBL WHERE 'LANDING__OUTCOME' = 'Success (drone ship)' and  PAYLOAD_MASS__KG_  >4000 and PAYLOAD_MASS__KG_ <6000
```

```
 * sqlite:///my_data1.db
Done.
```

Out[26]:

| Booster_Version |
|---|
| F9 B5 B1060.2 |
| F9 B5 B1058.3 |
| F9 B5 B1051.6 |
| F9 B5 B1060.3 |
| F9 B5 B1049.7 |

### Task 7

List the total number of successful and failure mission outcomes

In [19]:
```
%sql select count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure (in flight)'
```

```
 * sqlite:///my_data1.db
Done.
```

Out[19]:

| count(MISSION_OUTCOME) |
|---|
| 99 |

### Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

In [47]:
```
%sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT max(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

```
 * sqlite:///my_data1.db
Done.
```

Out[47]:

| Booster_Version |
|---|
| F9 B5 B1048.4 |
| F9 B5 B1049.4 |
| F9 B5 B1051.3 |
| F9 B5 B1056.4 |
| F9 B5 B1048.5 |

### Task 9

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

**Note: SQLLite does not support monthnames. So you need to use substr(Date, 4, 2) as month to get the months and substr(Date,7,4)='2015' for year.**

In [62]:
```
%sql SELECT BOOSTER_VERSION,LAUNCH_SITE,"Landing__Outcome" FROM SPACEXTBL WHERE "Landing__Outcome" = 'Failure (drone ship)' and substr(DATE,'2015')
```

```
 * sqlite:///my_data1.db
Done.
```

Out[62]:

| Booster_Version | Launch_Site | "Landing__Outcome" |
|---|---|---|

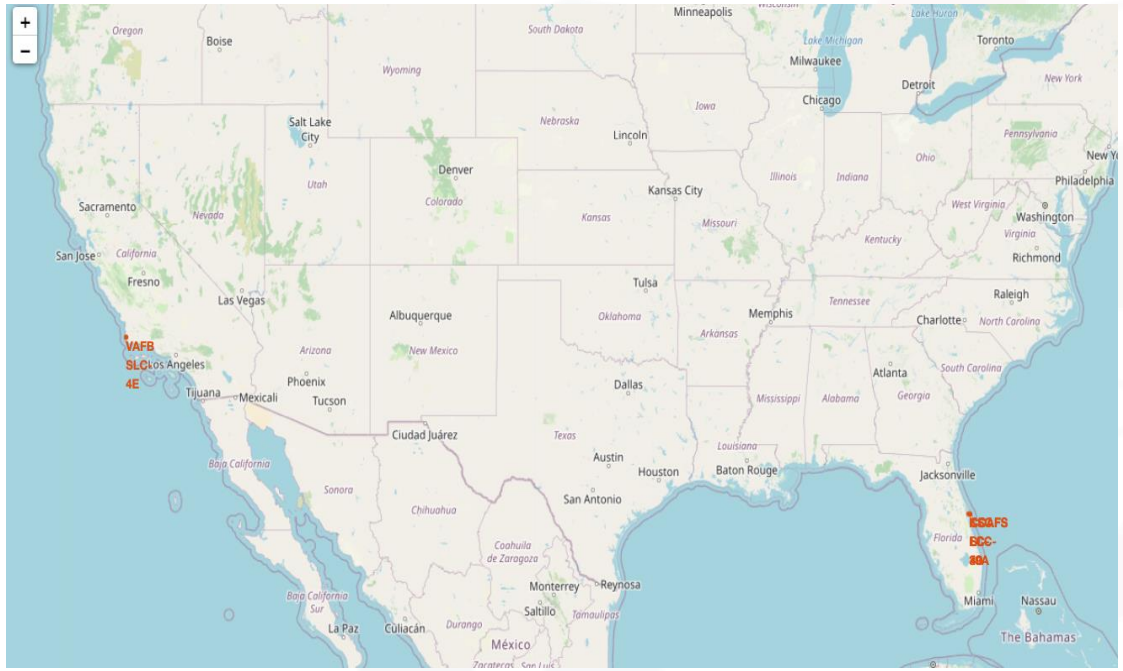
### Task 10

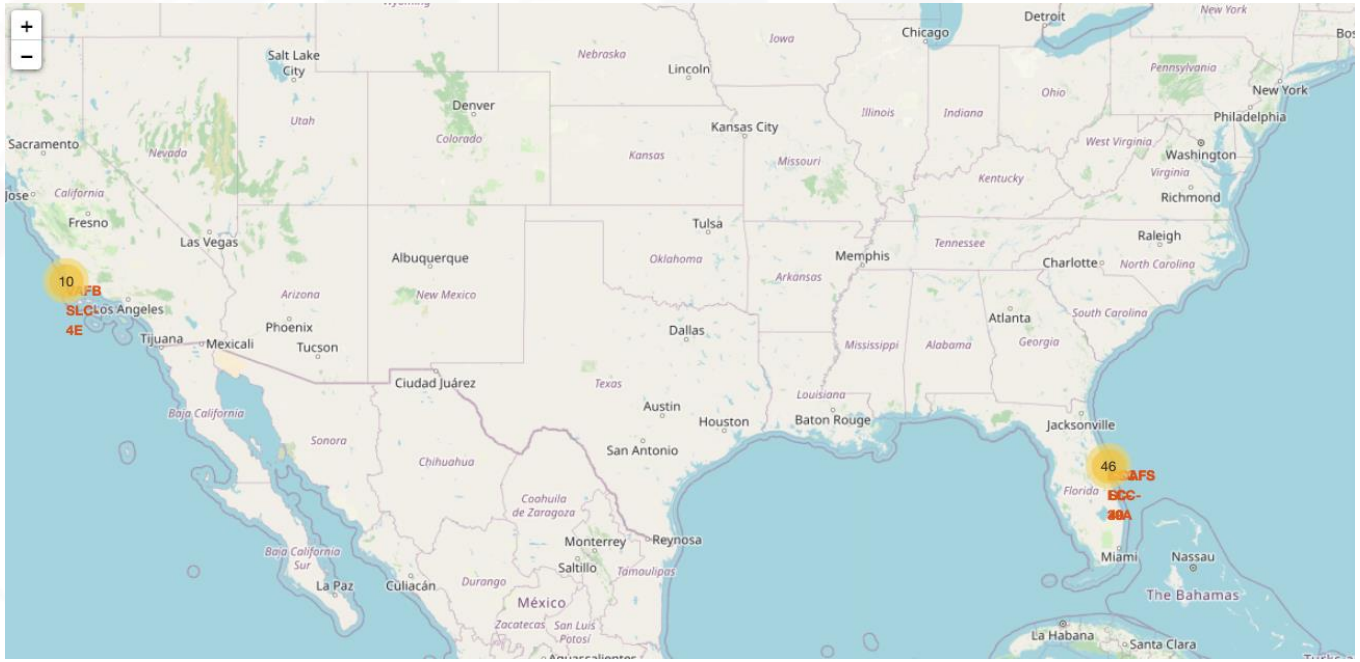
Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

In [59]:
```
%sql select * from SPACEXTBL where Landing__Outcome = 'Success (ground pad)' or and (DATE between '2010-06-04' and '2017-03-20') order by date desc
```

```
 * sqlite:///my_data1.db
(sqlite3.OperationalError) near "and": syntax error
[SQL: select * from SPACEXTBL where Landing__Outcome = 'Success (ground pad)' or and (DATE between '2010-06-04' and '2017-03-20') order by date desc]
(Background on this error at: http://sqlalche.me/e/e3q8)
```

# Interactive map with Folium

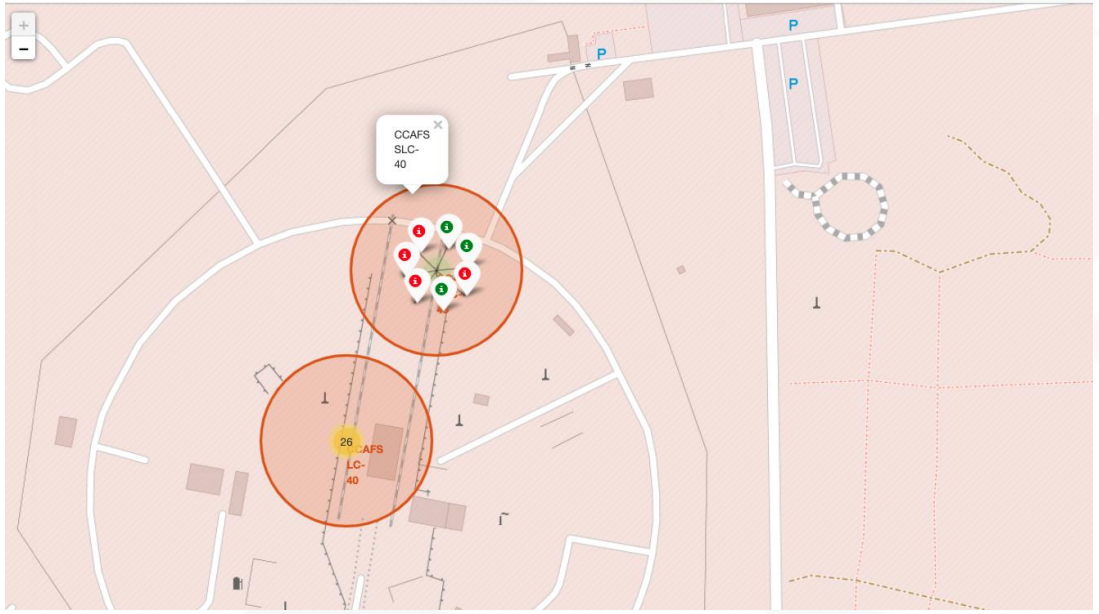


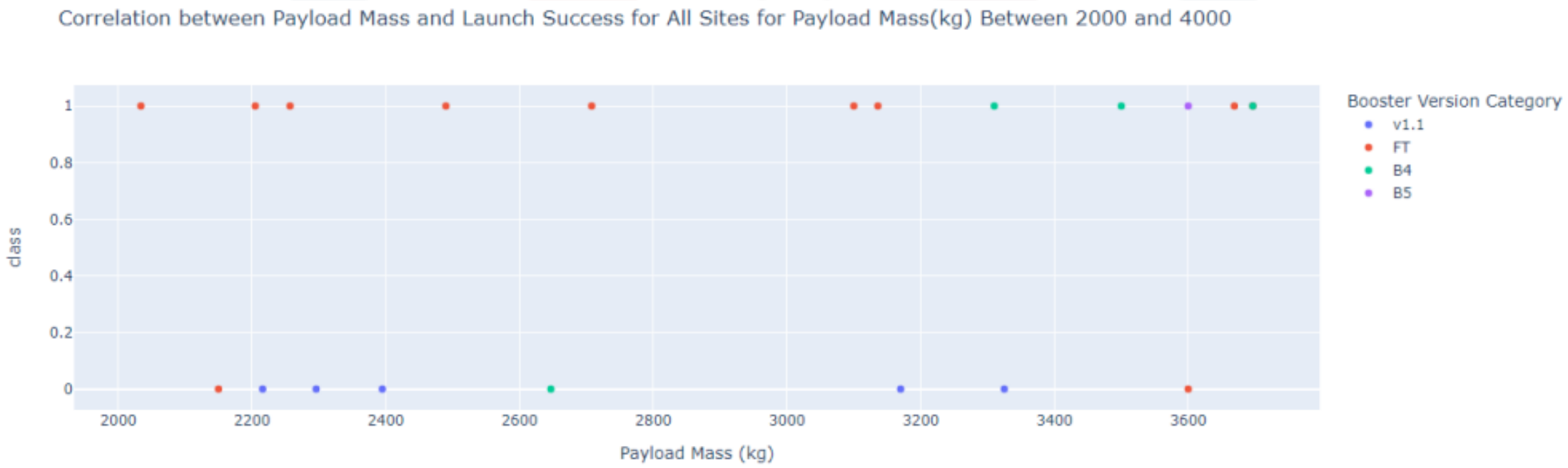

IBM Developer

SKILLS NETWORK

# Interactive map with Folium

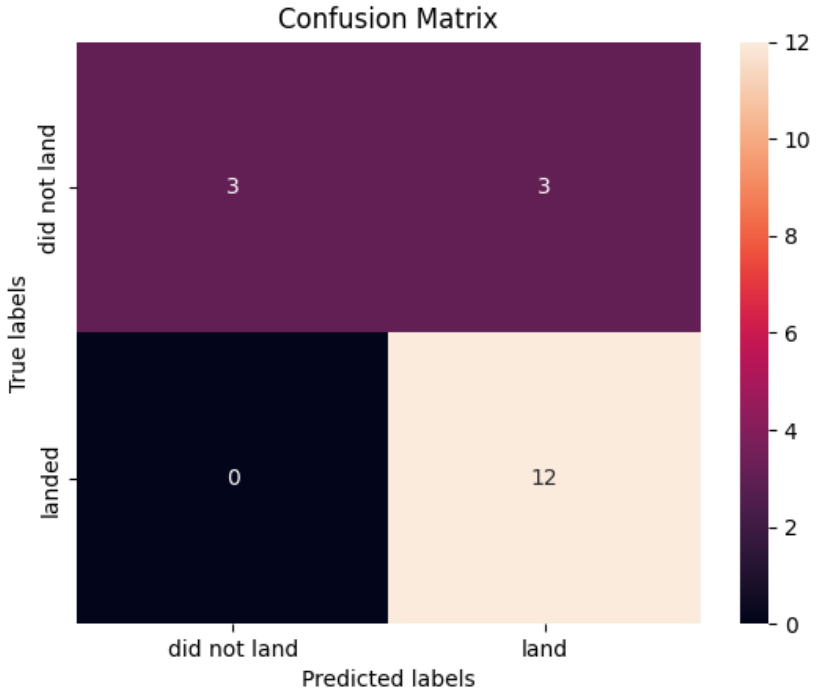



# Visualization Dashboard

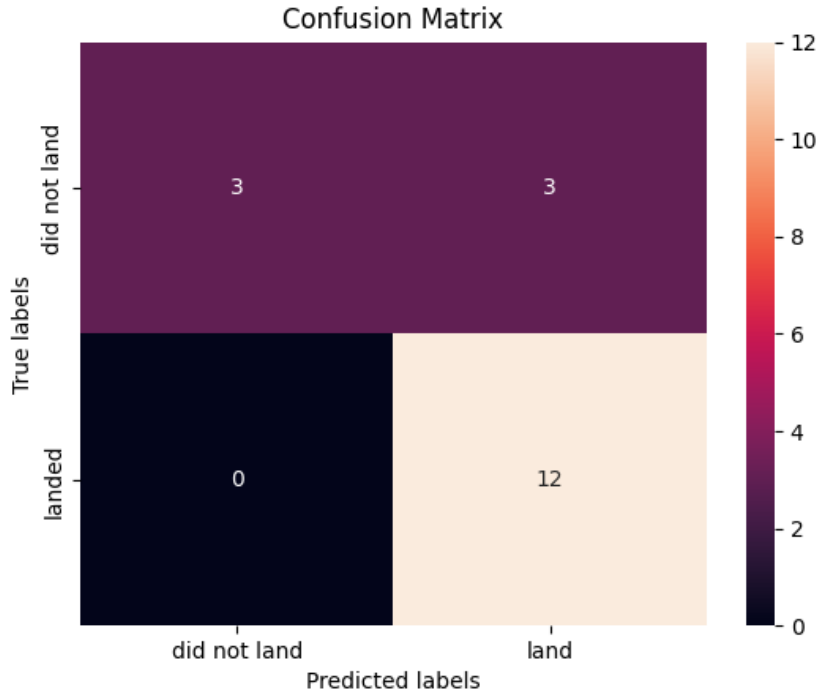


# Results for predictive analysis

- Regresión logística

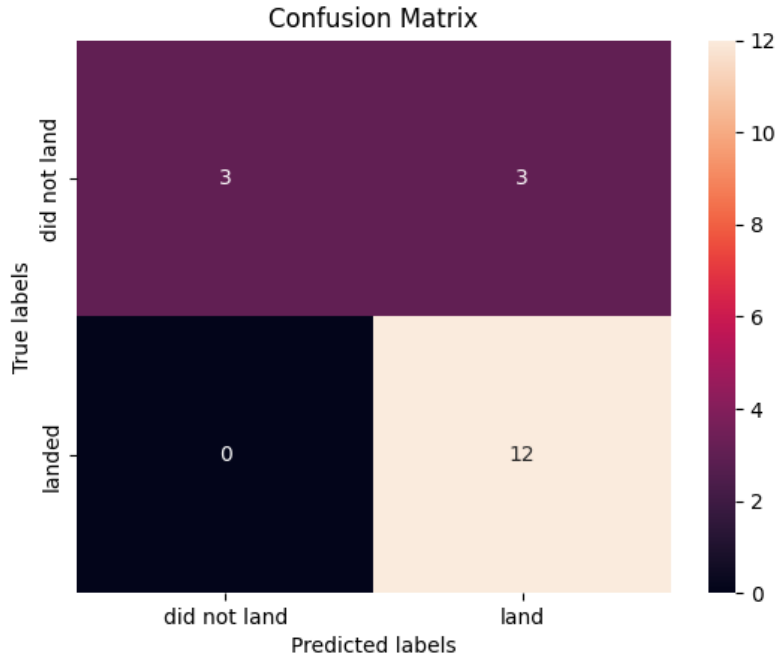



- SVM

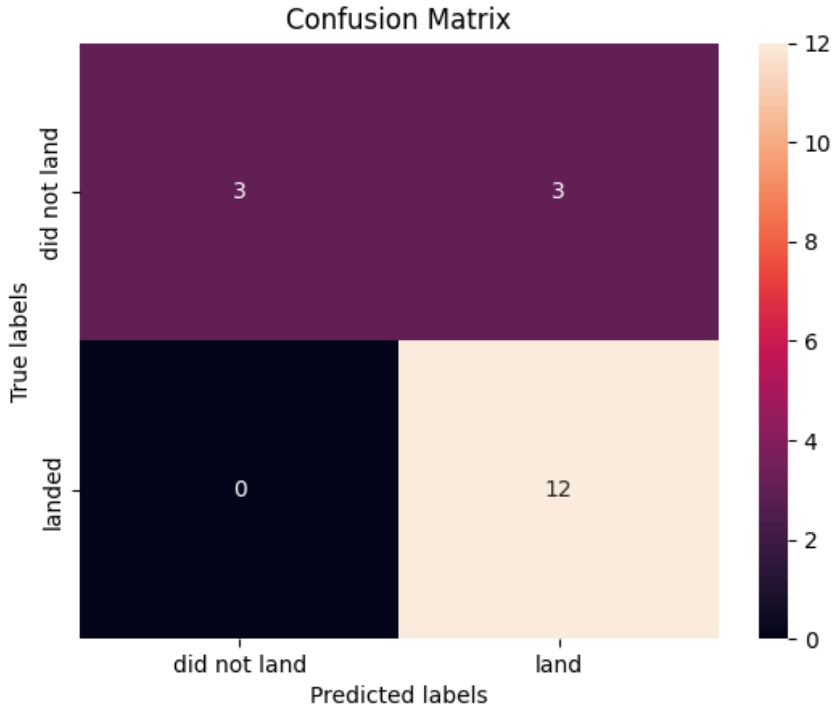


# Results for predictive analysis

- Decisión de árbol




- K-vecinos



# Accuracy for diferent models

| | Method | Accuracy Score (%) |
|---|---|---|
| 0 | Support Vector Machine | 83.333333 |
| 1 | Logistic Regression | 83.333333 |
| 2 | K Nearest Neighbour | 83.333333 |
| 3 | Decision Tree | 83.333333 |

# DISCUSSION

- Vemos que los diferentes métodos nos dan la misma precisión
- Elegiremos el modelo más simple computacionalmente
- En este caso nos podemos quedar con la regresión logística

# CONCLUSION

- Respecto a Space X, el punto de mayores lanzamientos con éxito es KSC-LC-39A
- A partir de 2015 el porcentaje de lanzamientos con éxito creció bastante
- Los lanzamientos se realizan casi siempre cerca de la costa ya que se pueden realizar pruebas de testeo de los propios cohetes con menos riesgo