IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Gina Vlassi
September 2021

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- Examine the characteristics of the SpaceY regarding the SpaceX
- Predict the outcomes of SpaceY's new launches
- Analyze and visualize through charts the results

Introduction

- Gather information about SpaceX and use them for our analysis
- Implement EDA with SQL commands and visualizations
- Use Folium and Plotly Dash for interactive analysis
- Use classification models for predictive analysis

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - A rest api is used to download the JSON records for our analysis
- Perform data wrangling
  - Count the number of occurrences to understand the distribution of values
  - Create a list with the landing outcome of each mission
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models like logistic regression, KNN, SVM, Decision Tree

Data Collection – SpaceX API

- Request the JSON records from SpaceX REST : https://api.spacexdata.com/v4/launches/past
- Convert them to static response object and normalize the JSON
- Create a dictionary with the data and convert it to a data frame
- https://github.com/GinaVl/IBM-Data-Science-Coursera/blob/main/Applied%20Data%20Science%20Capstone/Hands-on%20Lab/Data%20Collection/Working%20with%20Applied%20Data%20Science%20Capstone_1.ipynb

	static_fire_date_utc	static_fire_date_unix	net	window	rocket	success	failures	details	crew	ships	capsules	payloads
0	2006-03-17T00:00:00.000Z	1.142554e+09	False	0.0	5e9d0d95eda69955f709d1eb	False	[{'time': 33, 'altitude': None, 'reason': 'merlin engine failure'}]	Engine failure at 33 seconds and loss of vehicle	[]	[]	[]	[5eb0e4b5b
1	None	NaN	False	0.0	5e9d0d95eda69955f709d1eb	False	[{'time': 301, 'altitude': 289, 'reason': 'harmonic oscillation leading to premature engine shutdown'}]	Successful first stage burn and transition to second stage, maximum altitude 289 km, Premature engine shutdown at T+7 min 30 s, Failed to reach orbit, Failed to recover first stage	[]	[]	[]	[5eb0e4b6b

Data Collection - Scraping

- Get the HTML page and use BeautifulSoup to parse it
- Create empty dictionary and fill it by launching records extracted from table rows
- Convert the dictionary to a data frame
- https://github.com/GinaVI/IBM-Data-Science-Coursera/blob/main/Applied%20Data%20Science%20Capstone/Hands-on%20Lab/Data%20Collection/Working%20with%20Applied%20Data%20Science%20Capstone_2.ipynb

```
<tr>
<th scope="col">Flight No.
</th>
<th scope="col">Date and<br/>time (<a href="/wiki/Coordinated_Universal_Time" title="Coordinated Universal Time">UTC</a>)
</th>
<th scope="col"><a href="/wiki/List_of_Falcon_9_first-stage_boosters" title="List of Falcon 9 first-stage boosters">Version,<br/>B
ooster</a> <sup class="reference" id="cite_ref-booster_11-0"><a href="#cite_note-booster-11">[b]</a></sup>
</th>
<th scope="col">Launch site
</th>
<th scope="col">Payload<sup class="reference" id="cite_ref-Dragon_12-0"><a href="#cite_note-Dragon-12">[c]</a></sup>
</th>
<th scope="col">Payload mass
</th>
<th scope="col">Orbit
</th>
<th scope="col">Customer
</th>
<th scope="col">Launch<br/>outcome
</th>
<th scope="col"><a href="/wiki/Falcon_9_first-stage_landing_tests" title="Falcon 9 first-stage landing tests">Booster<br/>landing
</a>
</th></tr>
```

Data Wrangling

- Import the data from the Data Scarping lab
- Calculate the null values of each column
- Calculate the occurrences of each column
- Create new column with the landing outcome and calculate the average
- https://github.com/GinaVl/IBM-Data-Science-Coursera/blob/main/Applied%20Data%20Science%20Capstone/Hands-on%20Lab/Data%20Wrangling/Working%20with%20Applied%20Data%20Science%20Capstone%203.ipynb

EDA with Data Visualization

- The following plots constructed:
  - FlightNumber vs. PayloadMass
  - FlightNumber vs LaunchSite
  - PayloadMass vs LaunchSite
  - Orbit vs SuccessRate
  - Orbit vs FlightNumber
  - Orbit vs PayloadMass
  - Year vs Class
- https://github.com/GinaVl/IBM-Data-Science-Coursera/blob/main/Applied%20Data%20Science%20Capstone/Hands-on%20Lab/Explanatory%20Data%20Analysis/Working%20with%20Applied%20Data%20Science%20Capstone%205.ipynb

EDA with SQL

- The SQL queries performed:
  - Display the names of the unique launch sites in the space mission
  - Display 5 records where launch sites begin with the string 'CCA‘
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first successful landing outcome in ground pad was achieved
  - List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - List the total number of successful and failure mission outcomes
  - List the names of the booster_versions which have carried the maximum payload mass
  - List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
  - Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order
- https://github.com/GinaVI/IBM-Data-Science-Coursera/blob/main/Applied%20Data%20Science%20Capstone/Hands-on%20Lab/Explanatory%20Data%20Analysis/Working%20with%20Applied%20Data%20Science%20Capstone%204.ipynb

Build an Interactive Map with Folium

- The Folium package is used for the dashboard. The following interactions added:
  - Mark all launch sites on a map
  - Mark the success/failed launches for each site on the map
  - Calculate the distances between a launch site to its proximities
- https://github.com/GinaVl/IBM-Data-Science-Coursera/blob/main/Applied%20Data%20Science%20Capstone/Hands-on%20Lab/Interactive%20Visual%20Analytics%20and%20Dashboard/Working%20with%20Applied%20Data%20Science%20Capstone%206.ipynb

Predictive Analysis (Classification)

- The following classifications models constructed and the accuracy of each evaluated:
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - K Nearest Neighborhood
- https://github.com/GinaVl/IBM-Data-Science-Coursera/blob/main/Applied%20Data%20Science%20Capstone/Hands-on%20Lab/Predictive%20Analysis%20(Classification)/Working%20with%20Applied%20Data%20Science%20Capstone%207.ipynb

Results

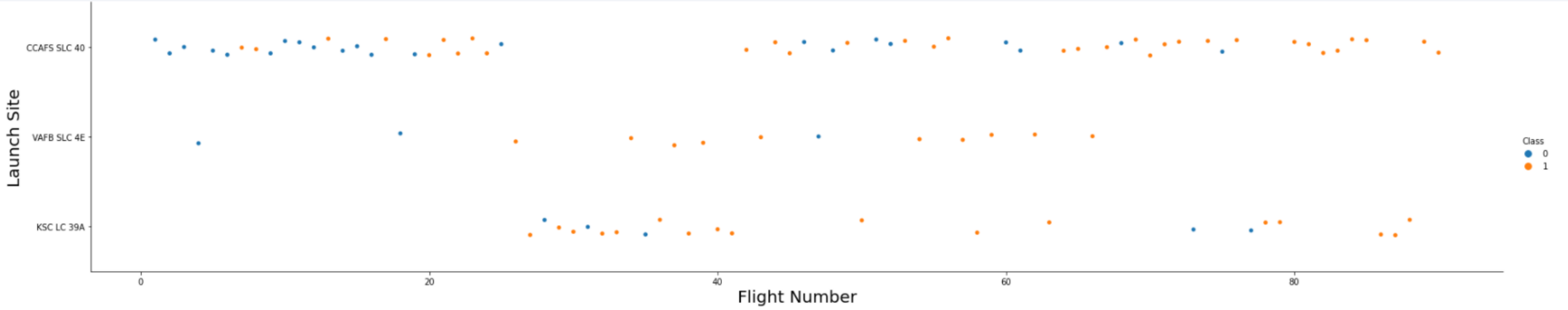
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

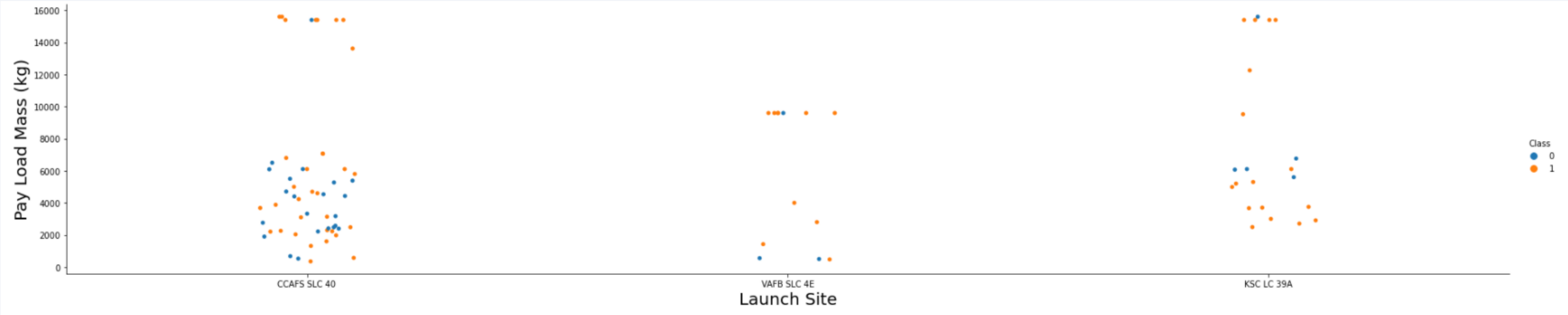
Flight Number vs. Launch Site

- Observing the data we understand that as the number of the flight increases, is more likely to land successfully on the launch site



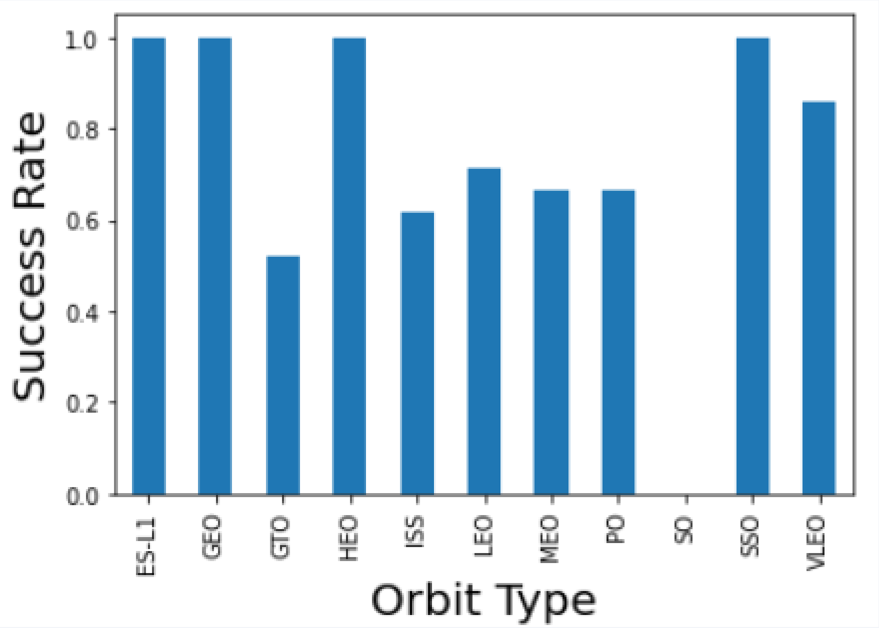
Payload vs. Launch Site

- Observing the data we understand that sites CCAFS SLC 40 and KSC LC 39A have the the successful landing of bigger pay load mass in KG.



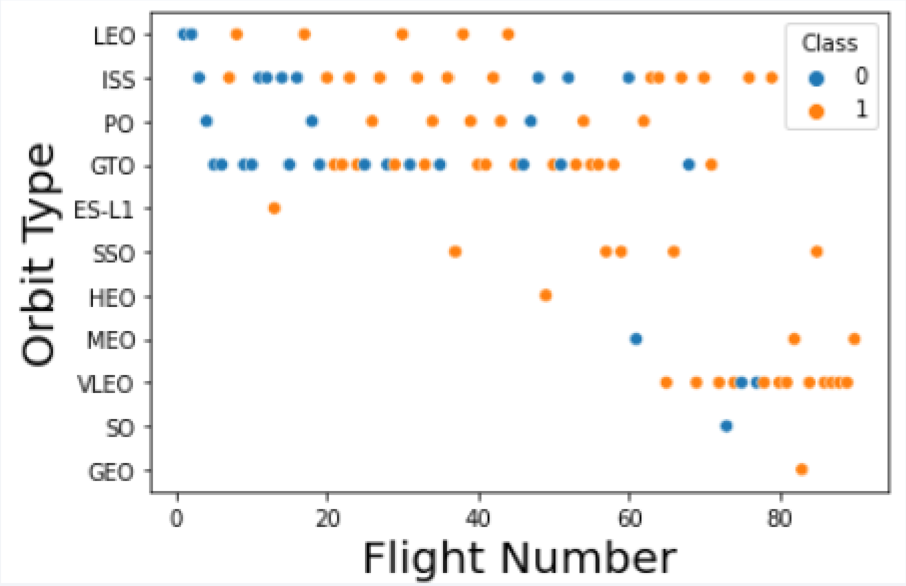
Success Rate vs. Orbit Type

- Orbit types ES-L1, GEO, HEO and SSO have the absolute landing success



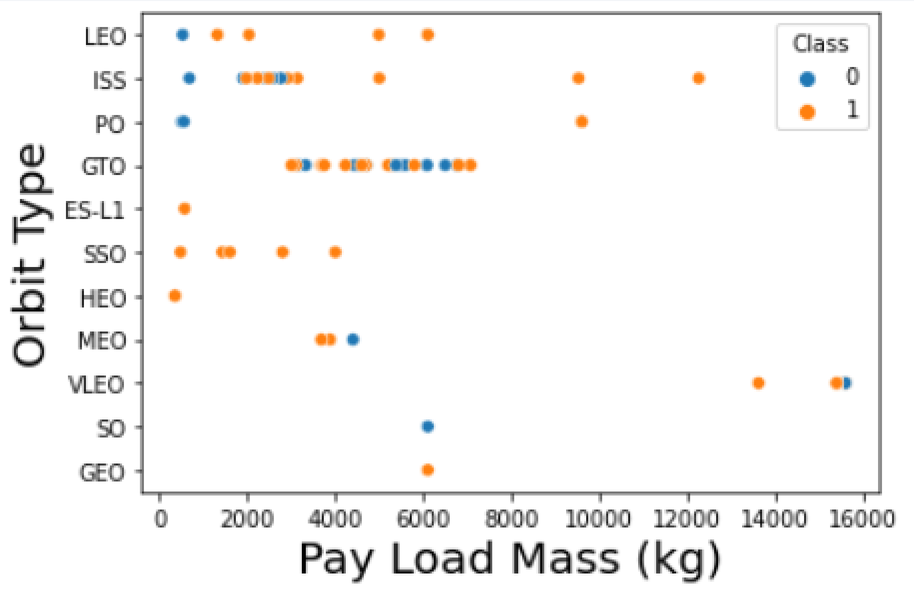
Flight Number vs. Orbit Type

- Observing the data we understand that as the flight number increases in Orbit types LEO, SSO, VLEO, we have the more successful landing outcome



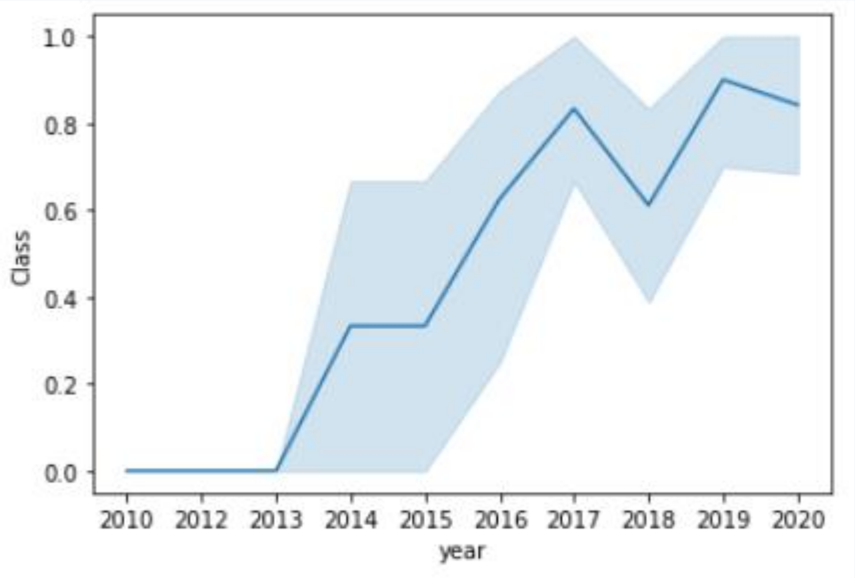
Payload vs. Orbit Type

- Observing the data we understand that the most successful lands need from 2000 to 6000 KG of pay load



Launch Success Yearly Trend

- During the years we observe that the average success rate increases



All Launch Site Names

- Find the names of the unique launch sites

```
%sql select distinct(launch_site) from SPACEXTBL
```

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

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

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

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

```
%sql select sum(payload_mass__kg_) from SPACEXTBL where customer = 'NASA (CRS)'
```

1
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

```
%sql select avg(payload_mass__kg_) from SPACEXTBL where booster_version ='F9 v1.1'
```

1
2928

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

```
%sql select DATE from SPACEXTBL where mission_outcome = 'Success' order by DATE limit 1
```

DATE
2010-06-04

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
%sql select payload from SPACEXTBL where payload_mass__kg_>=4000 and payload_mass__kg_<=6000 and mission_outcome = 'Success'
```

payload
AsiaSat 8
AsiaSat 6
ABS-3A Eutelsat 115 West B
Turkmen 52 / MonacoSAT
SES-9
JCSAT-14
JCSAT-16
EchoStar 23
SES-10
NROL-76
Boeing X-37B OTV-5
SES-11 / EchoStar 105
GovSat-1 / SES-16
SES-12
Merah Putih
Es hail 2
SSO-A
GPS III-01
Nusantara Satu, Beresheet Moon lander, S5
RADARSAT Constellation, SpaceX CRS-18
GPS III-03, ANASIS-II
ANASIS-II, Starlink 9 v1.0
GPS III-04 , Crew-1

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

```
%sql select count(mission_outcome) from SPACEXTBL
```

1
101

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

```
%sql select booster_version from SPACEXTBL where payload_mass__kg_=(select max(payload_mass__kg_) from SPACEXTBL) LIMIT 1
```

booster_version
F9 B5 B1048.4

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

```
%sql select landing__outcome,booster_version,launch_site from SPACEXTBL where
landing__outcome = 'Failure (drone ship)' and year(DATE) = '2015'
```

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

```
%sql select landing__outcome,count(landing__outcome)as LANDING_OUTCOME_COUNT from SPACEXTBL where DATE between '2010-06-04' and '2017-03-20' group by landing__outcome
```

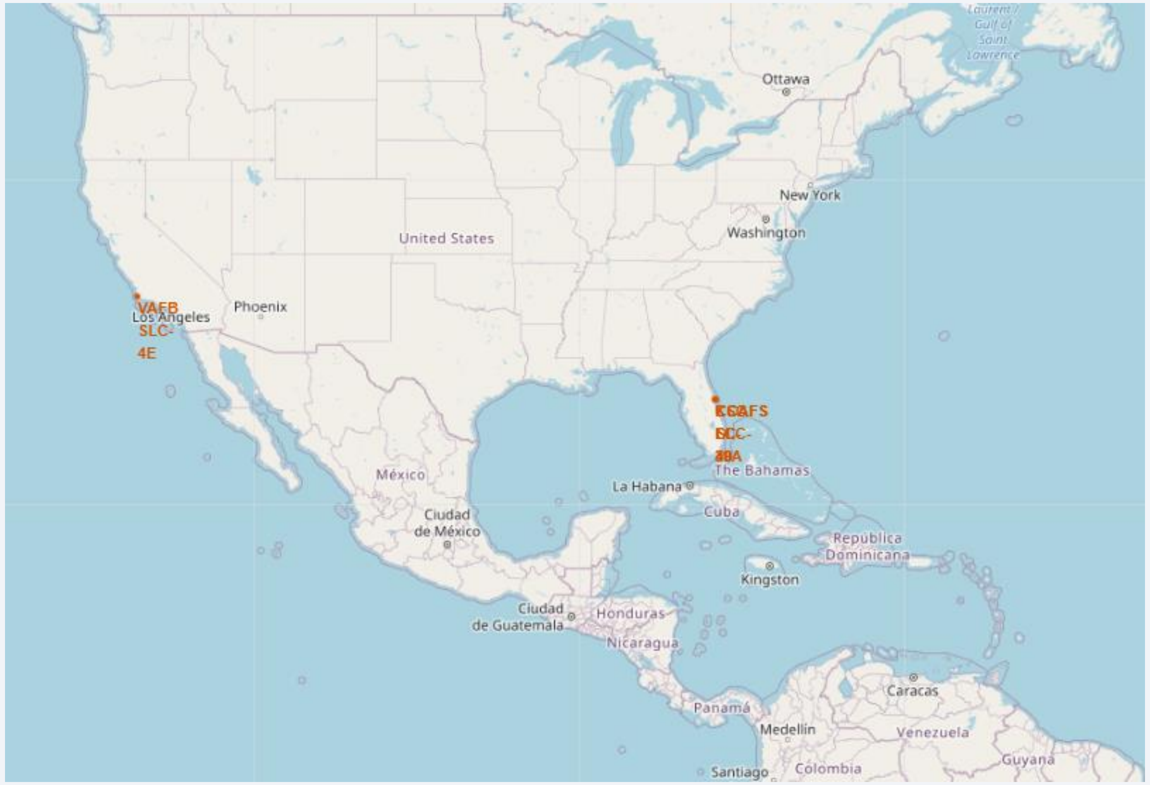
landing__outcome	landing_outcome_count
Controlled (ocean)	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	10
Precluded (drone ship)	1
Success (drone ship)	5
Success (ground pad)	3
Uncontrolled (ocean)	2

Section 4

Launch Sites Proximities Analysis

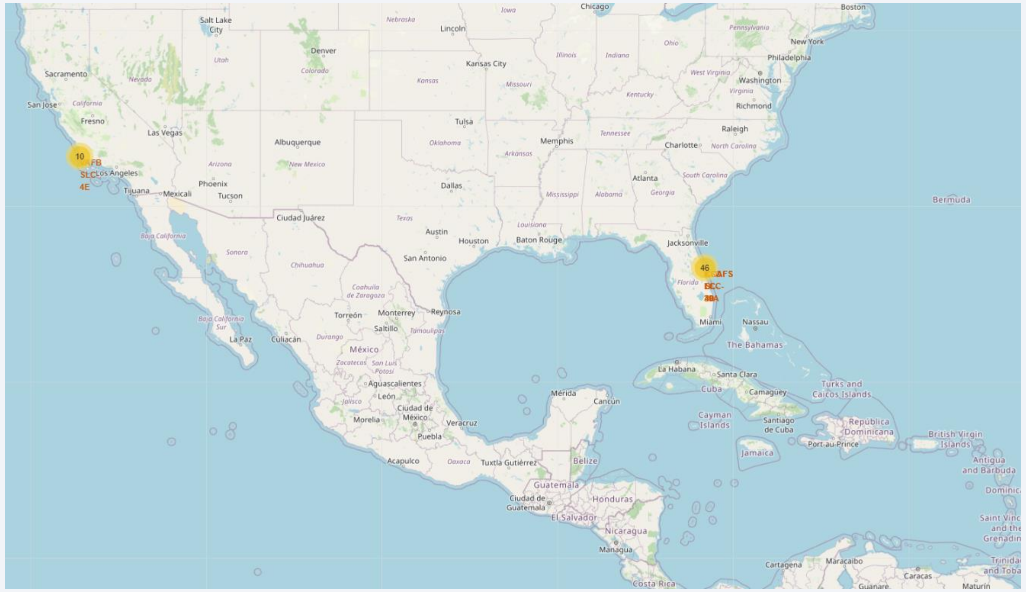
Launch all sites on a map

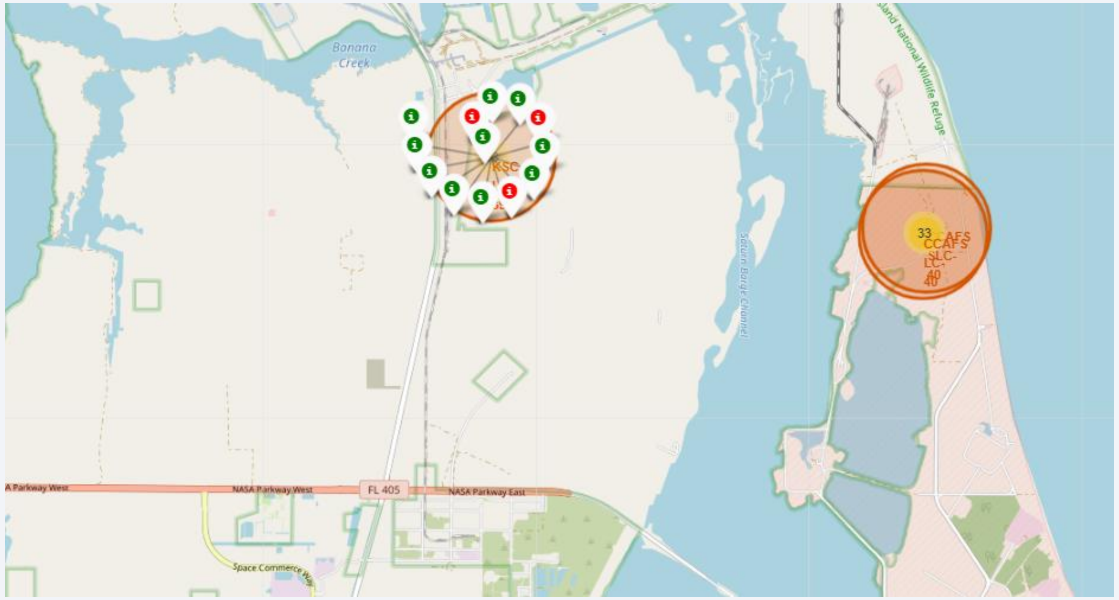
- We use the NASA coordinates to construct the map. With the red circles we highlight the all the launch sites.



Mark the success/failed launches for each site on the map

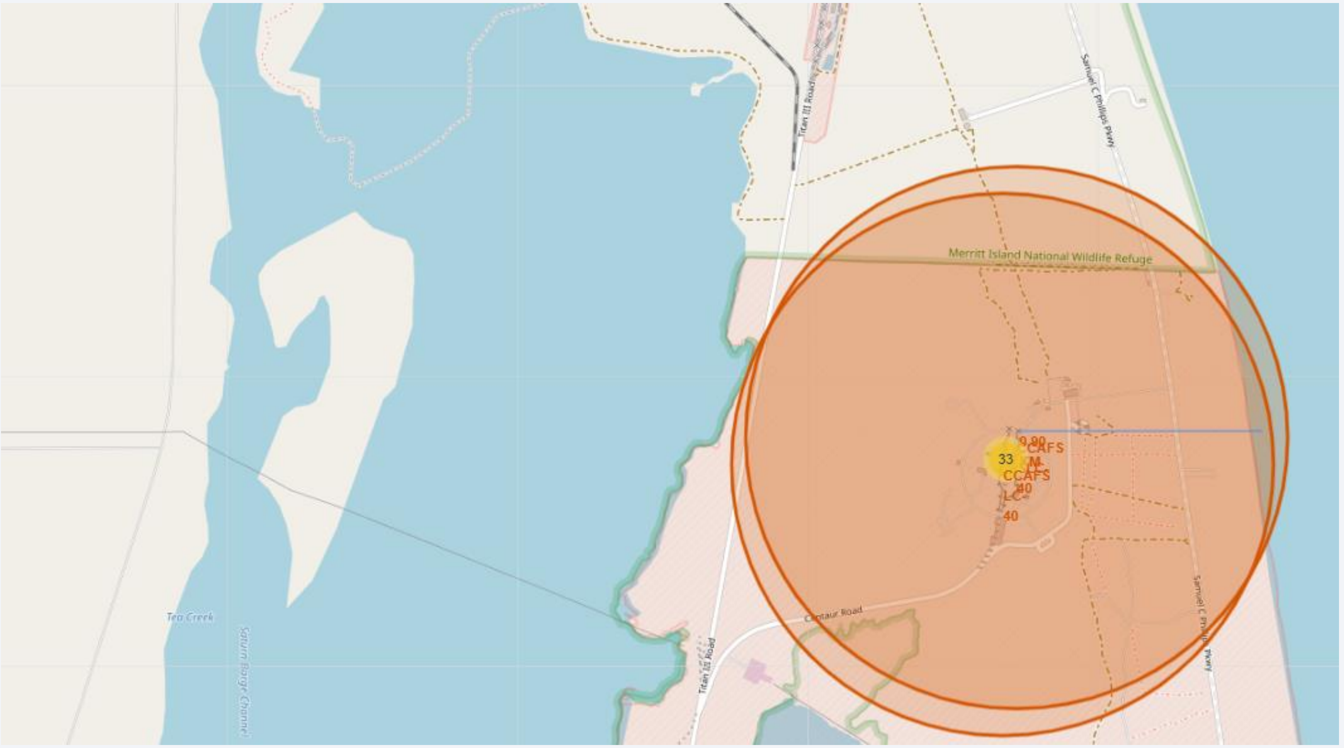
- With the yellow highlights we observe the total launches of each site
- The green spots specify successful launch in comparison with the red spots





Calculate the distances between a launch site to its proximities

- The blue line indicates the distance between the launch site and the closest coastline

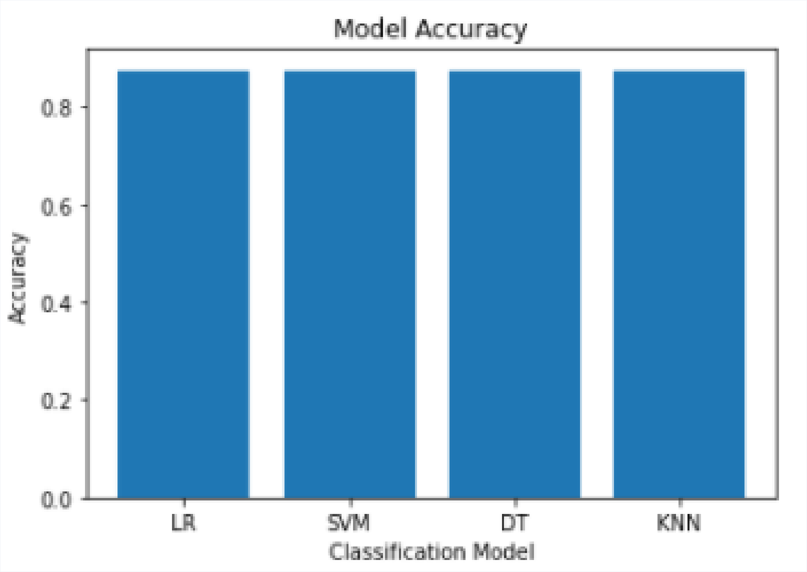


Section 6

Predictive Analysis (Classification)

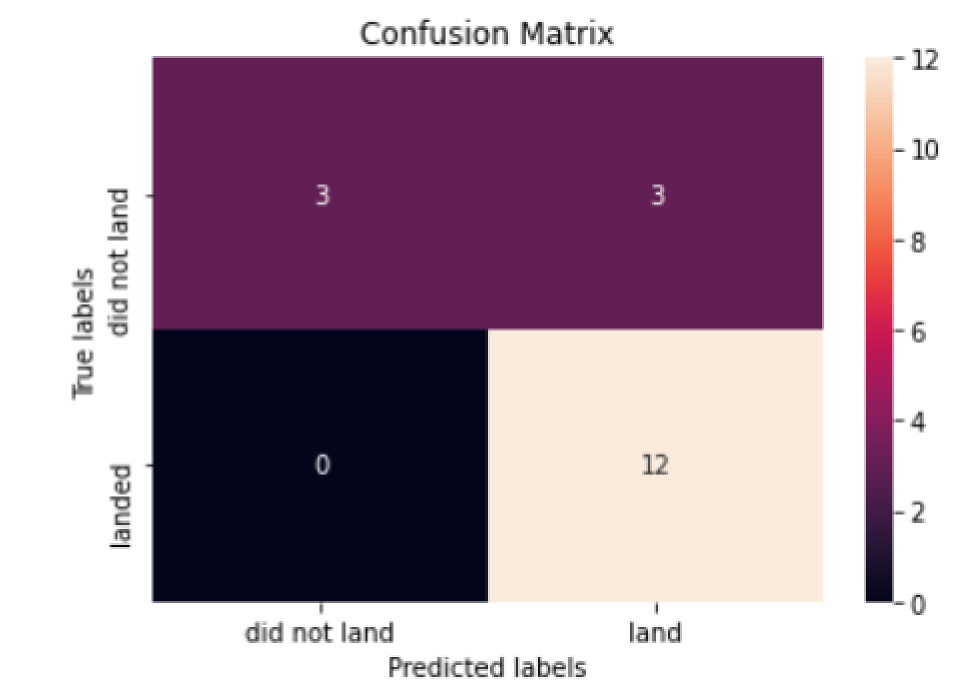
Classification Accuracy

- The accuracy results are practically the same of all the classification models



Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation



Conclusions

- Completing this analysis we achieved to:
  - Analyze real data from an api
  - Construct explanatory data analysis with simple SQL command and visualization charts
  - Use Folium package to draw a map with different interactions
  - Build different classification models and calculate the accuracy of the best model

Thank you!