IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Dave C. LeFevre
April 14, 2022

Outline

Executive Summary

- Summary of methodologies
  - Collection of data about Space X missions from public sources
  - Wrangling the data gathered for exploratory analysis
  - Creation of visualizations to better explore the Space X data
  - Creation of a model to help determine optimal conditions for Space Y missions
- Summary of all results
  - Through public data we can create a model that will allow us to determine the best conditions for recovery allowing us to save money on rocket operations.

Introduction

- Project background and context:
  - At Space Y we aim to be the leading provider of space transport within the next 10 years. To do this we must have an idea of what ways we bring payload to orbit with less cost.
- Problem we want an answer to:
  - How can we predict the optimal conditions in which to launch payload that allow us to have the most successful missions and to recover costly first stage rockets.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology
  - Data was collected from public available sources, including the Space X data API and open sources such as Wikipedia.
  - Clean and wrangle the data from our sources into a useful data set for numerical analysis.
- Problem we want an answer to:
  - How can we predict the optimal conditions in which to launch payload that allow us to have the most successful missions and to recover costly first stage rockets.

Methodology

Executive Summary

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
- Using our data we created training sets to try with various methods of Logistical regression. After testing the accuracy of thesemethods with test data we chose a model that would help us findthe most economical conditions to launch rockets.

Data Collection

Data was collected from two sources:

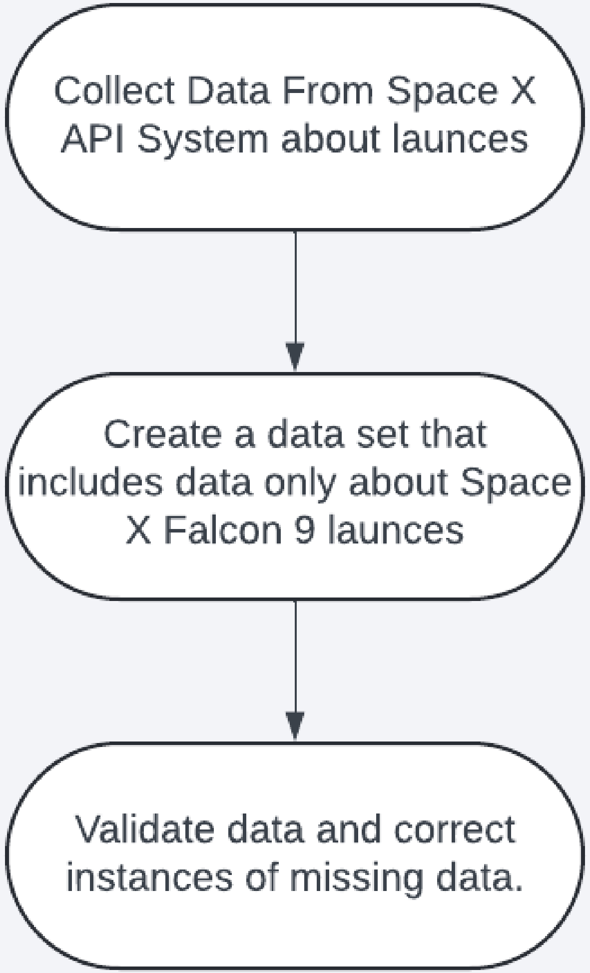
- The Space X API's interface for rocket data
  - https://api.spacexdata.com/v4/rockets/
- Wikipedia pages on Space X launches
  - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Data Collection – SpaceX API

- The Space X web site allows for download of publicly available data. We used their API to import this data into datasets.

- Notebook with SpaceX API data collection: https://github.com/dcl-coursera/data_science_capstone/blob/main/01-jupyter-labs-spacex-data-collection-api.ipynb

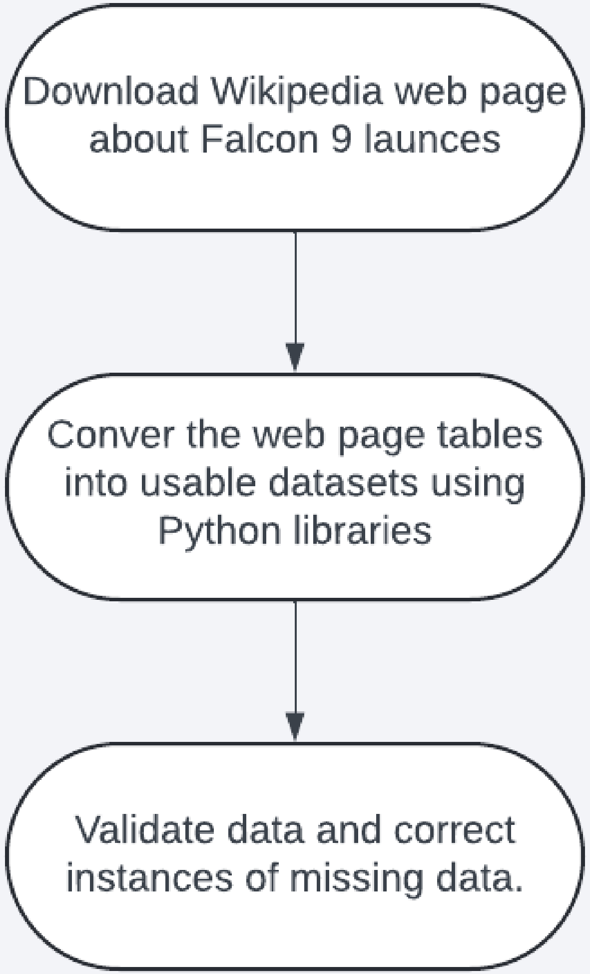
`



Data Collection - Scraping

- Wikipedia also collects all publicly known data about Falcon 9 launches. This data was downloaded an put into an form that can be analyzed.

- Notebook with web scraping data collection: https://github.com/dcl-coursera/data_science_capstone/blob/main/02-jupyter-labs-webscraping.ipynb



Data Wrangling

When combining and comparing the data from the two sources we:

- Calculated the number of launches on each site.
- Calculated the number and occurrence of each orbit
- Calculate the number and occurrence of mission outcome of the orbits
- Create a table of landing outcomes to allow for mathematical analysis

Notebook with Data Wrangling:
https://github.com/dcl-coursera/data_science_capstone/blob/main/03-labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

We explored the Data with the following plots:

- Flight number vs. Launch Site
- Payload vs. Launch Site
- Success Rate vs. Orbit Type
- Flight Number vs. Orbit Type
- Payload vs. Orbit Type
- Launch Success Yearly Trend

- Notebook with Data Visualization EDA: https://github.com/dcl-coursera/data_science_capstone/blob/main/05-jupyter-labs-eda-data viz.ipynb

EDA with SQL

Using SQL we found answers to some questions:

- Find the names of the unique launch sites
- Find 5 records where launch sites begin with `CCA`
- Calculate the total payload carried by boosters from NASA
- Calculate the average payload mass carried by booster version F9 v1.1
- Find the dates of the first successful landing outcome on ground pad
- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000
- Calculate the total number of successful and failure mission outcomes

EDA with SQL

Using SQL we found answers to some questions (continued):

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

Notebook with SQL-based EDA:
https://github.com/dcl-coursera/data_science_capstone/blob/main/04-jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Groups were added to a map to show all launch sites. Individual launch areas are indicated in these groups.

Individual success and failures are indicated on the map to visually see the record of each area.

Lines are added on the maps to show proximity to coasts and other map lines for an indication of what areas may work best for flights.

Notebook with Folium Map:
https://github.com/dcl-coursera/data_science_capstone/blob/main/06-lab_jupyter_launch_site_location%20.ipynb

Build a Dashboard with Plotly Dash

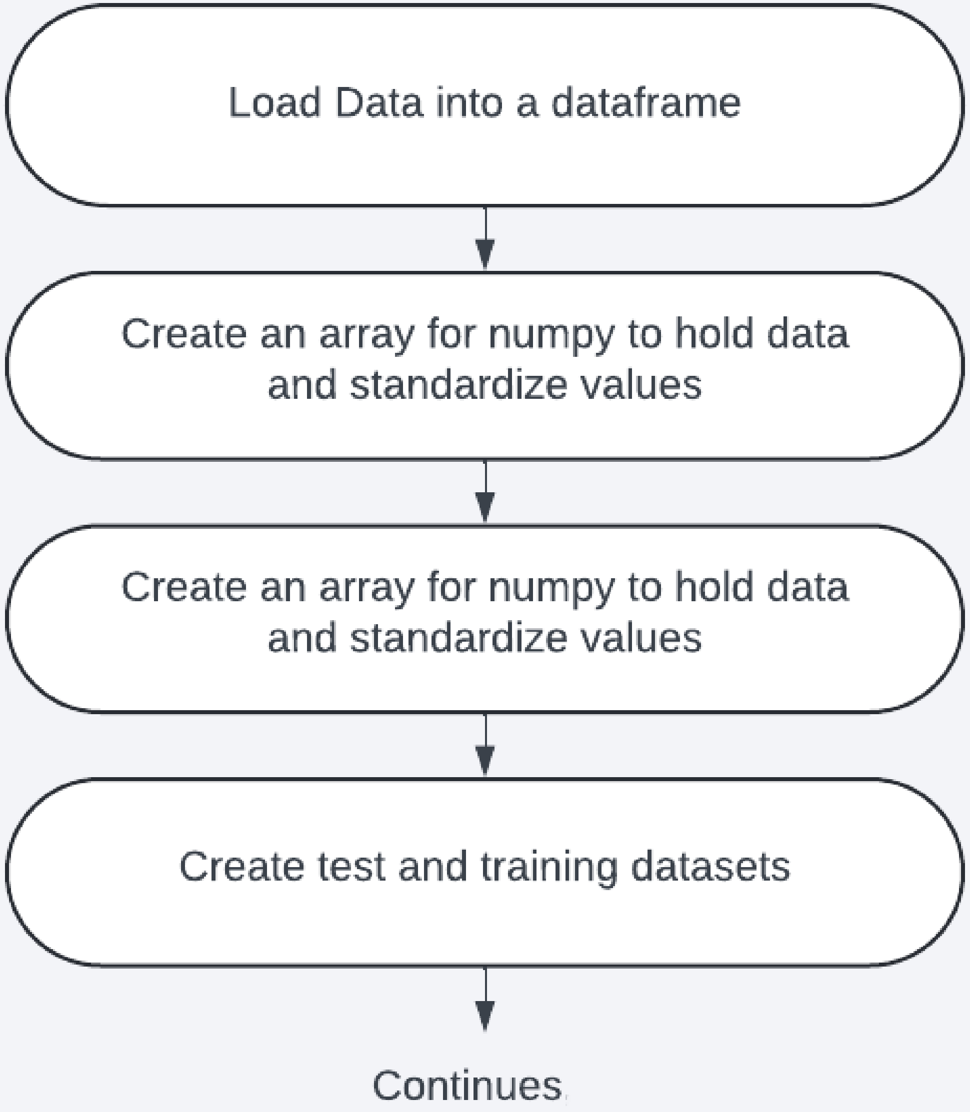
A dashboard was built to show the percentages of successful launches per all four sites as well as the percentages of successful vs. unsuccessful launches at each site.

To show the relationship between payload weights and mission success an interactive scatter plot is included. The plot has a slider to see how the plot changes vs. different ranges of payload weights and be viewed for each selected launch site as well as for all launches as a whole.

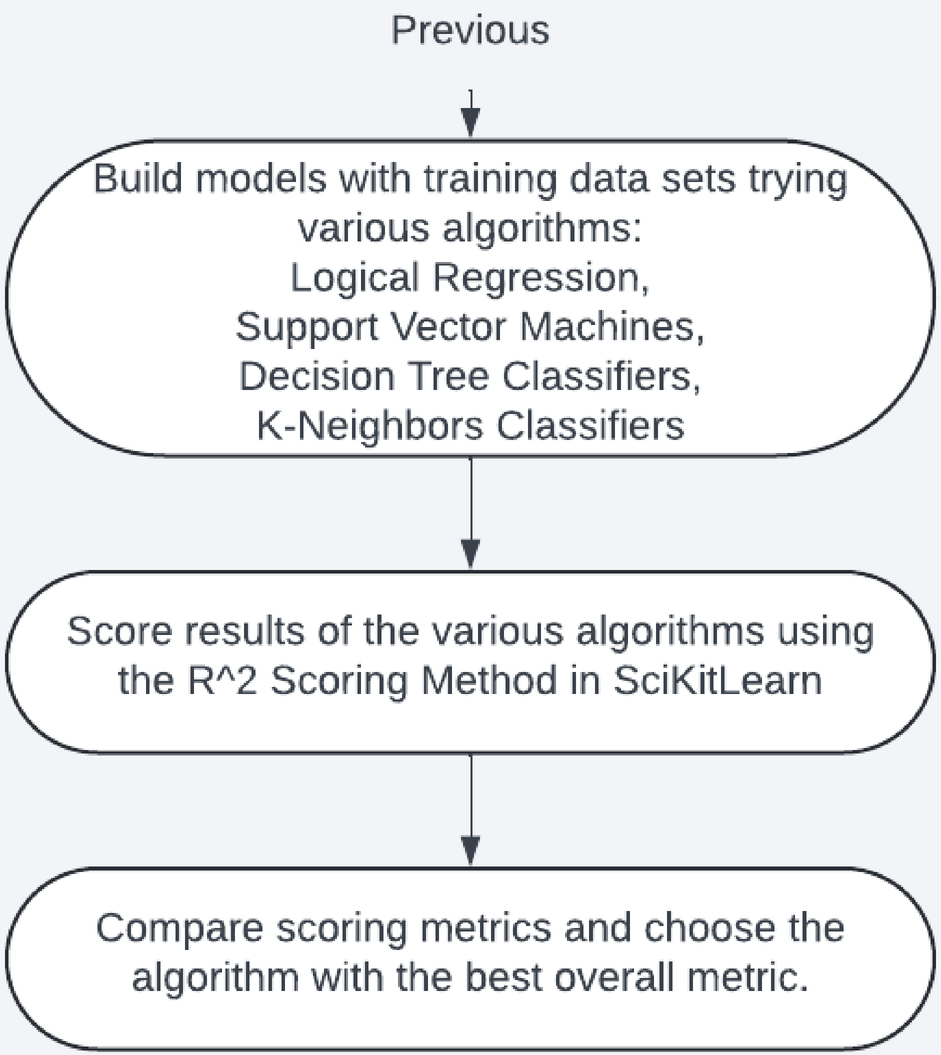
This combination allows for checks to see what sites and payload weight combinations work best for mission success.

Source code of Plotly Dash dashboard:
https://github.com/dcl-coursera/data_science_capstone/blob/main/07-spacex_dash_app.py

Predictive Analysis (Classification)



Predictive Analysis (Classification)



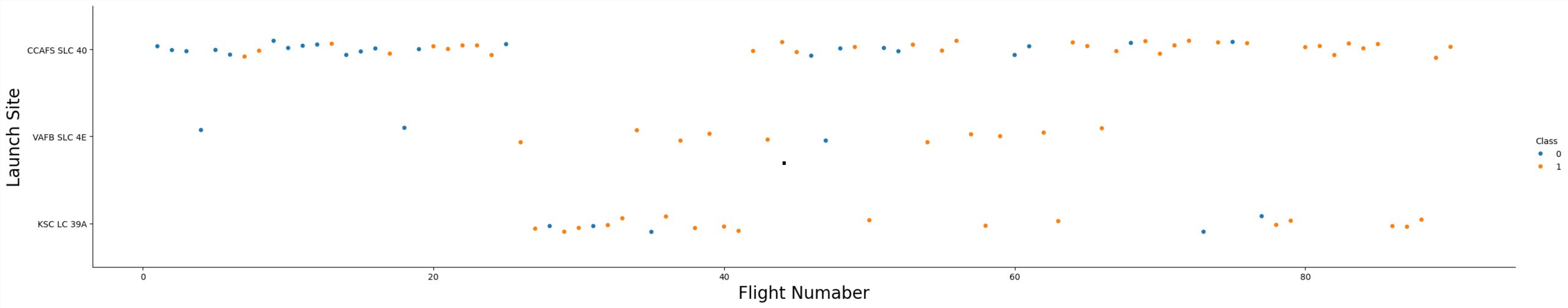
Predictive Analysis (Classification)

Notebook with Predictive Analysis:
https://github.com/dcl-coursera/data_science_capstone/blob/main/08-SpaceX_Machine%20Learning%20Prediction_Part_5%20(copy).ipynb

Section 2

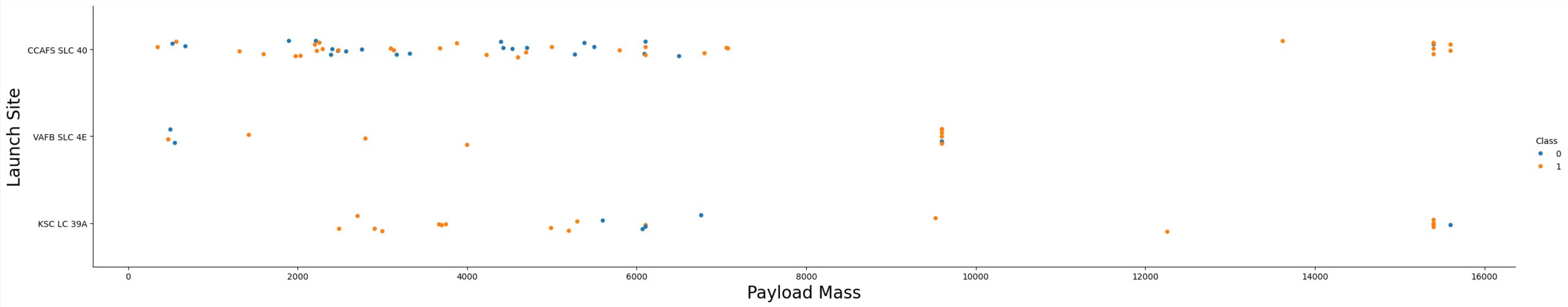
Insights drawn from EDA

Flight Number vs. Launch Site



We see that different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.

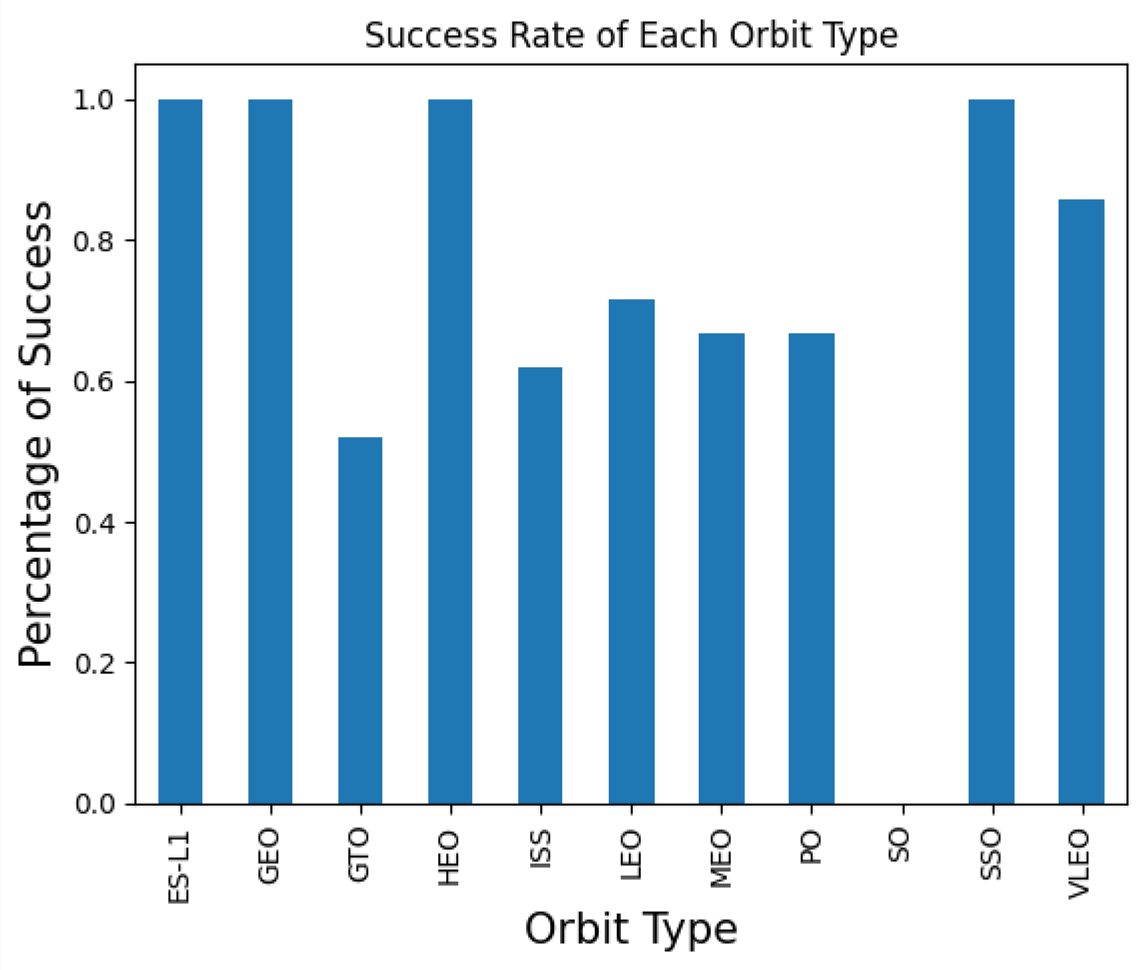
Payload vs. Launch Site



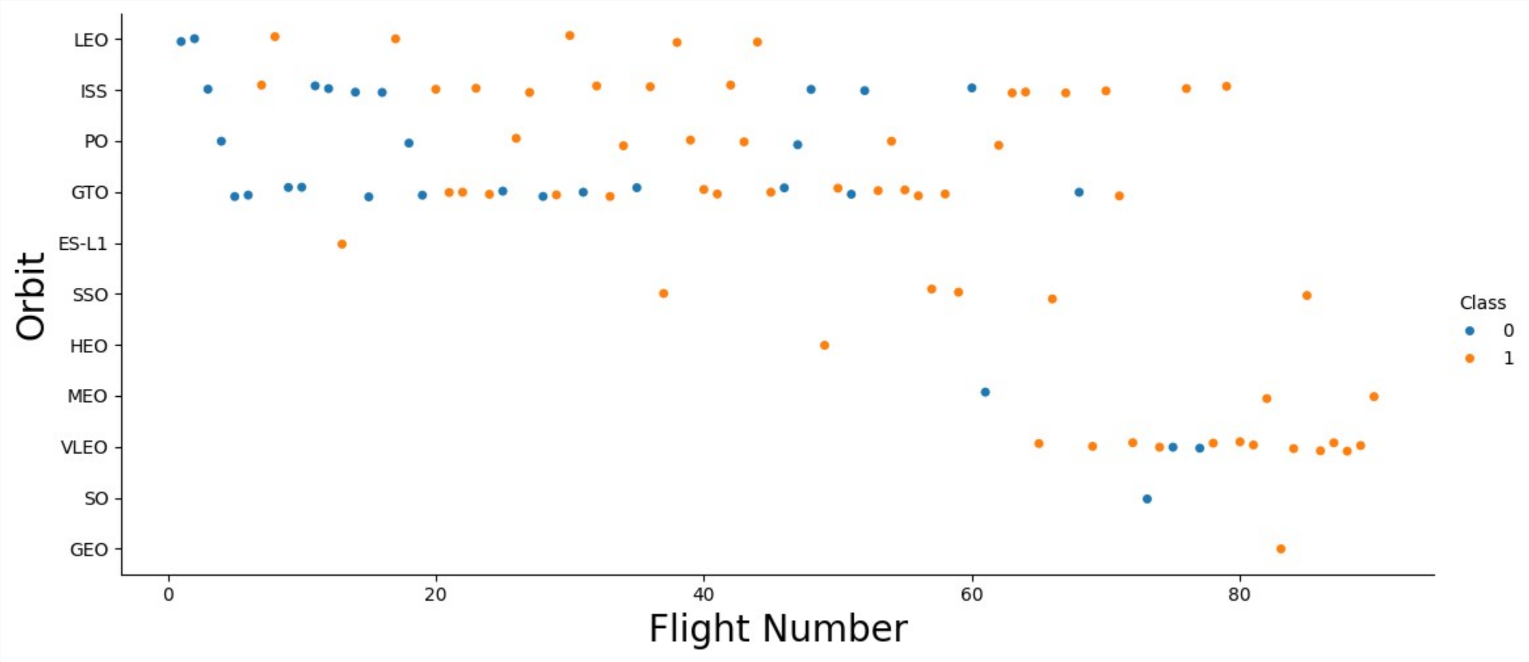
For the VAFB-SLC launchsite there are no rockets launched for heavy payload mass(greater than 10000).

Success Rate vs. Orbit Type

Several orbits have perfect success rates. These are ES-L1, GEO, HEO, and SSO. VLEO also have very high success rates.

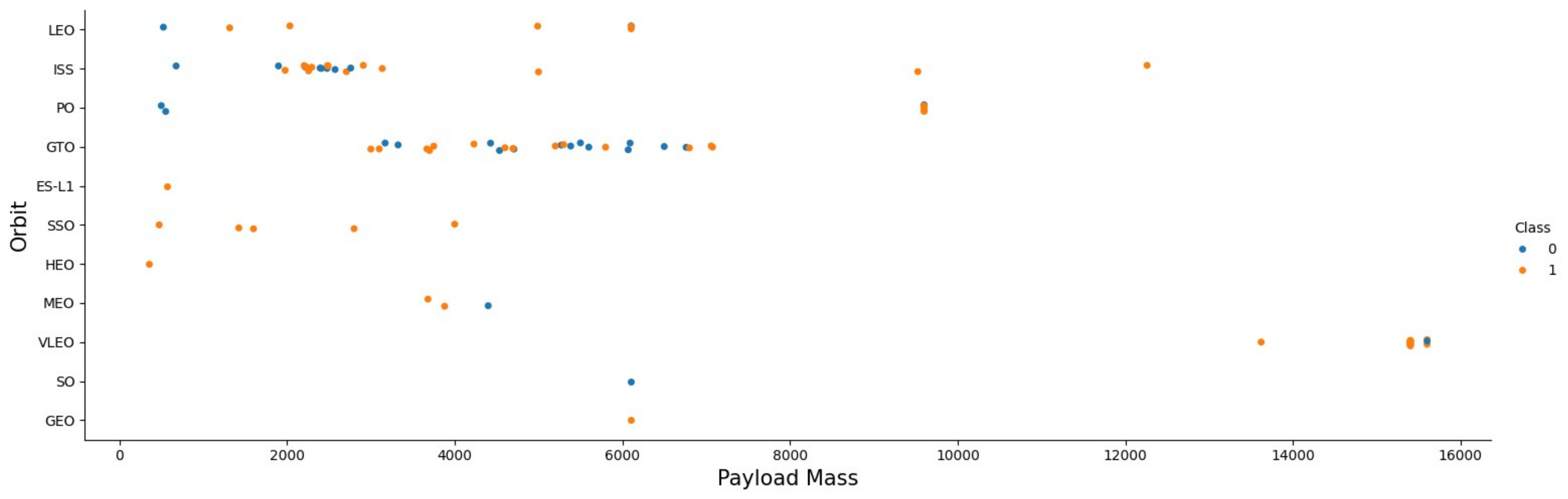


Flight Number vs. Orbit Type



In the LEO orbit the Success appears related to the number of flights; however there seems to be no relationship between flight number when in GTO orbit.

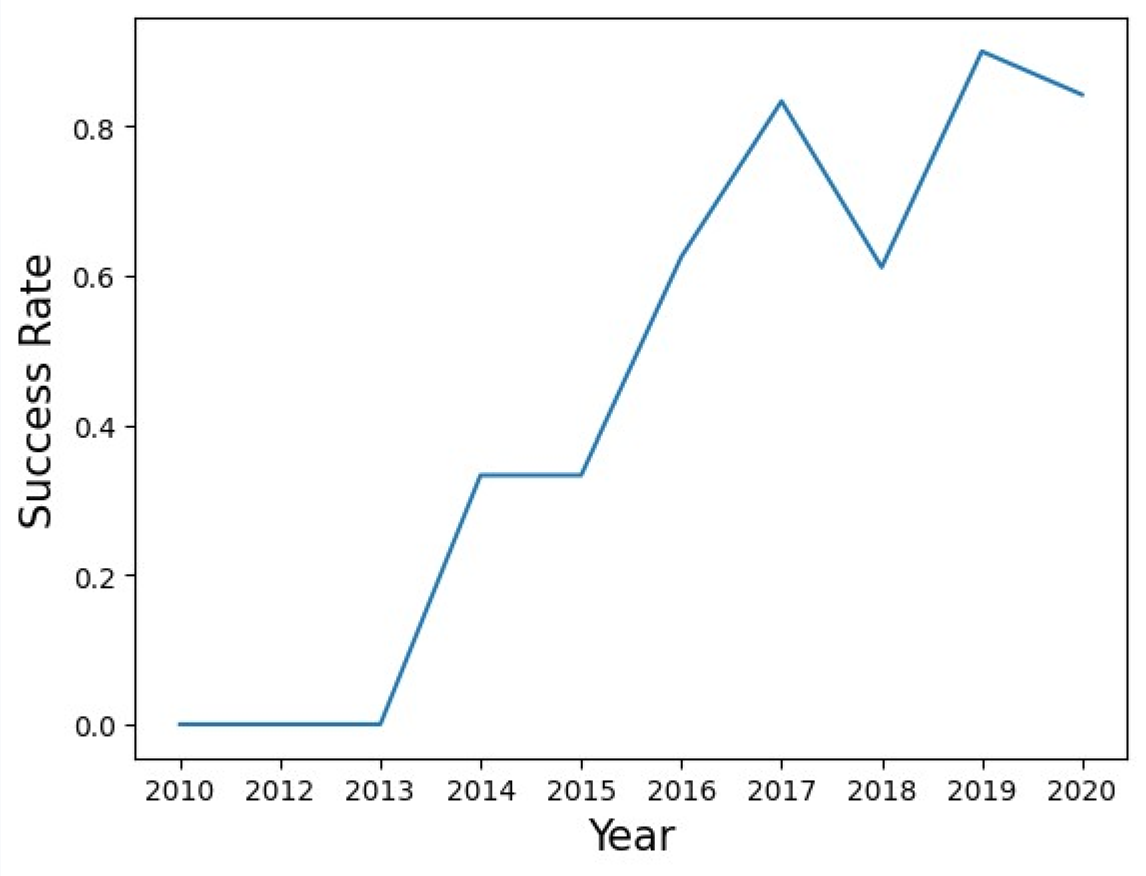
Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS. However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend

The success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing. The percentage of successful missions increases with the experience of the crew and the company.



All Launch Site Names

Find the names of the unique launch sites

SELECT DISTINCT("Launch_Site") FROM SPACEXTABLE

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

Launch Site Names Begin with 'CCA'

Find 5 records where launch sites begin with `CCA`

SELECT * FROM SPACEXTABLE 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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

SELECT sum("PAYLOAD_MASS__KG_") FROM SPACEXTABLE WHERE "Customer" like "NASA%"

sum("PAYLOAD_MASS__KG_")
99980

Average Payload Mass by F9 v1.1

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

- SELECT avg(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE "Booster_Version" like "F9 v1.1%"

avg(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

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

SELECT Date FROM SPACEXTABLE 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

SELECT DISTINCT(Booster_Version) FROM SPACEXTABLE WHERE Mission_Outcome="Success" and "PAYLOAD_MASS__KG_" > 4000 and "PAYLOAD_MASS__KG_" < 6000

These boosters: F9 v1.1, F9 v1.1 B1011, F9 v1.1 B1014, F9 v1.1 B1016, F9 FT B1020, F9 FT B1022, F9 FT B1026, F9 FT B1030, F9 FT B1021.2, F9 FT B1032.1, F9 B4 B1040.1, F9 FT B1031.2, F9 FT B1032.2, F9 B4 B1040.2, F9 B5 B1046.2, F9 B5 B1047.2, F9 B5 B1048.3, F9 B5 B1051.2, F9 B5B1060.1, F9 B5 B1058.2, F9 B5B1062.1

Total Number of Successful and Failure Mission Outcomes

Calculate the total number of successful and failure mission outcomes

SELECT COUNT(Date) FROM SPACEXTABLE WHERE Mission_Outcome="Success"

Success: 98

SELECT COUNT(Date) FROM SPACEXTABLE WHERE Mission_Outcome
!="Success"

Fail: 3

Boosters Carried Maximum Payload

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

SELECT Booster_Version, "PAYLOAD_MASS__KG_" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE)

Booster_Version	PAYLOAD_MASS__KG_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

SELECT substr ("--JanFebMarAprMayJunJulAugSepOctNovDec", strftime ("%m", Date) * 3, 3) as month, * FROM SPACEXTABLE WHERE substr(Date,0,5)='2015'

month	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
Jan	2015-01-10	9:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
Feb	2015-02-11	23:03:00	F9 v1.1 B1013	CCAFS LC-40	DSCOVR	570	HEO	U.S. Air Force NASA NOAA	Success	Controlled (ocean)
Mar	2015-03-02	3:50:00	F9 v1.1 B1014	CCAFS LC-40	ABS-3A Eutelsat 115 West B	4159	GTO	ABS Eutelsat	Success	No attempt
Apr	2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
Apr	2015-04-27	23:03:00	F9 v1.1 B1016	CCAFS LC-40	Turkmen 52 / MonacoSAT	4707	GTO	Turkmenistan National Space Agency	Success	No attempt
Jun	2015-06-28	14:21:00	F9 v1.1 B1018	CCAFS LC-40	SpaceX CRS-7	1952	LEO (ISS)	NASA (CRS)	Failure (in flight)	Precluded (drone ship)
Dec	2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

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

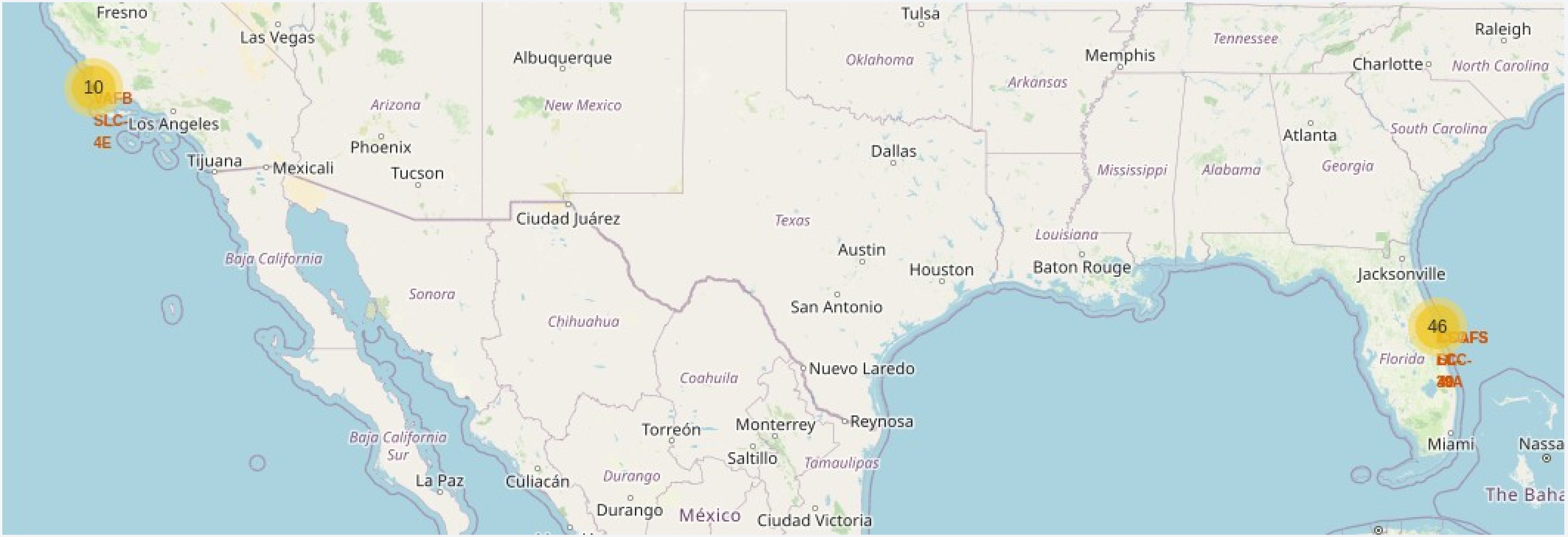
SELECT Date, Landing_Outcome, Count(Landing_Outcome) Num from SPACEXTABLE WHERE Date > "2010-06-04" AND Date < "2017-03-20" GROUP BY Landing_Outcome ORDER by Num Desc

Date	Landing_Outcome	Num
2012-05-22	No attempt	10
2016-04-08	Success (drone ship)	5
2015-01-10	Failure (drone ship)	5
2015-12-22	Success (ground pad)	3
2014-04-18	Controlled (ocean)	3
2013-09-29	Uncontrolled (ocean)	2
2015-06-28	Precluded (drone ship)	1
2010-12-08	Failure (parachute)	1

Section 3

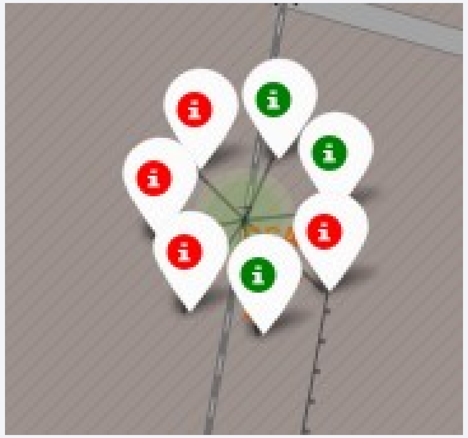
Launch Sites Proximities Analysis

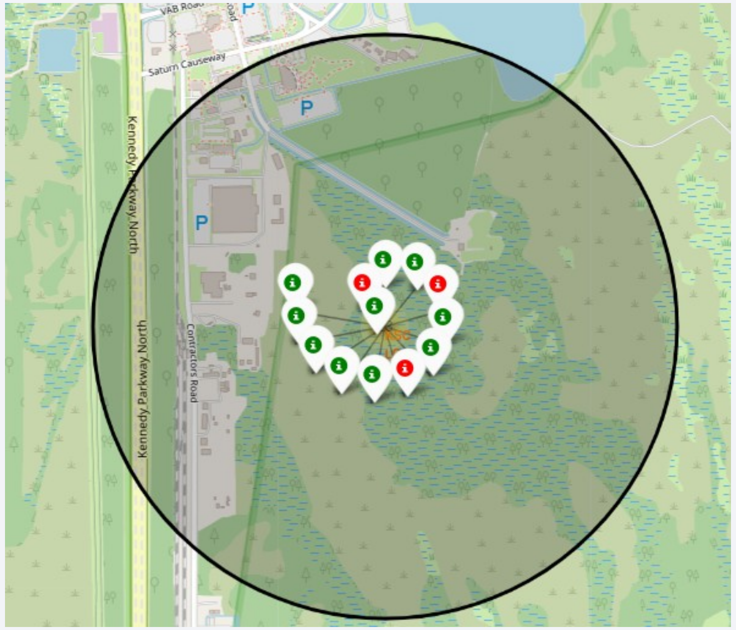
Folium map of launch site locations



Maps of Successful and failed launches

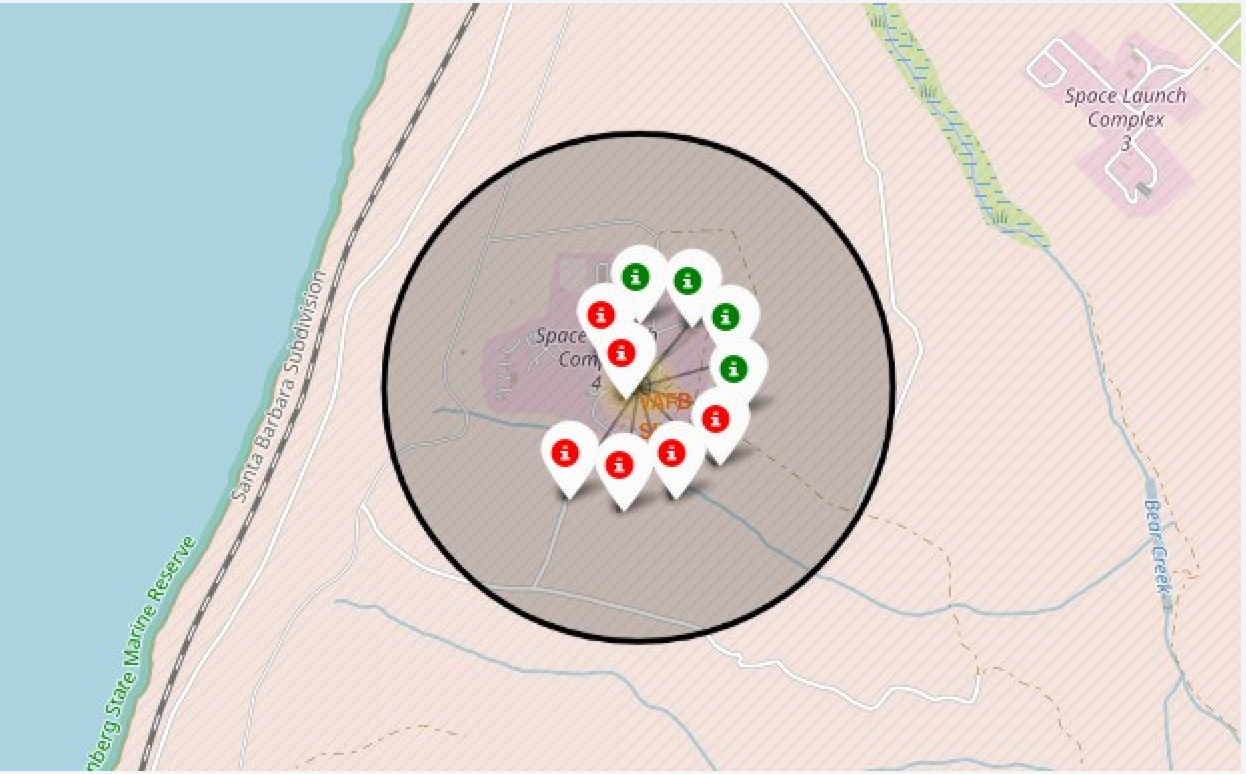
Successful and Failed Launches in Florida:



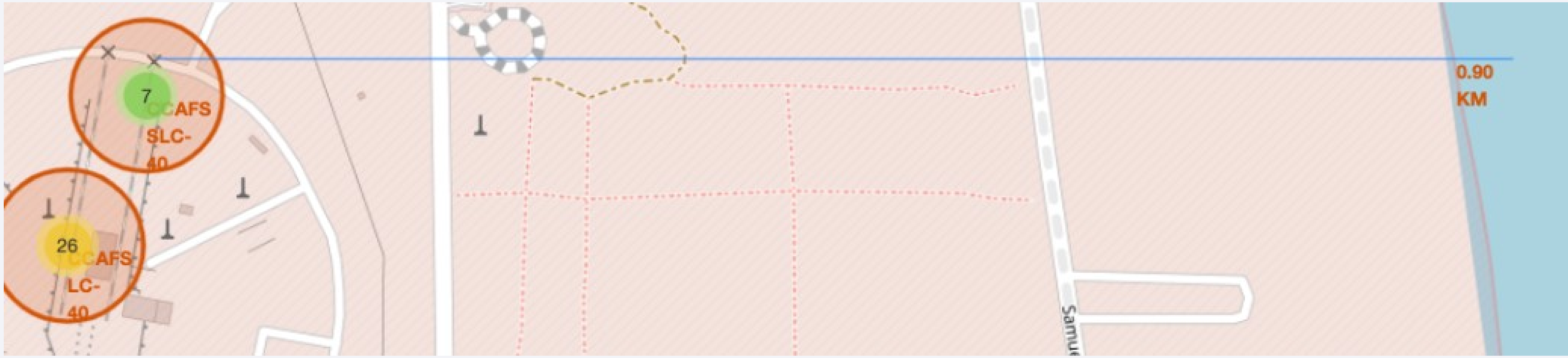


Maps of Successful and failed launches

Successful and Failed Launches in California:



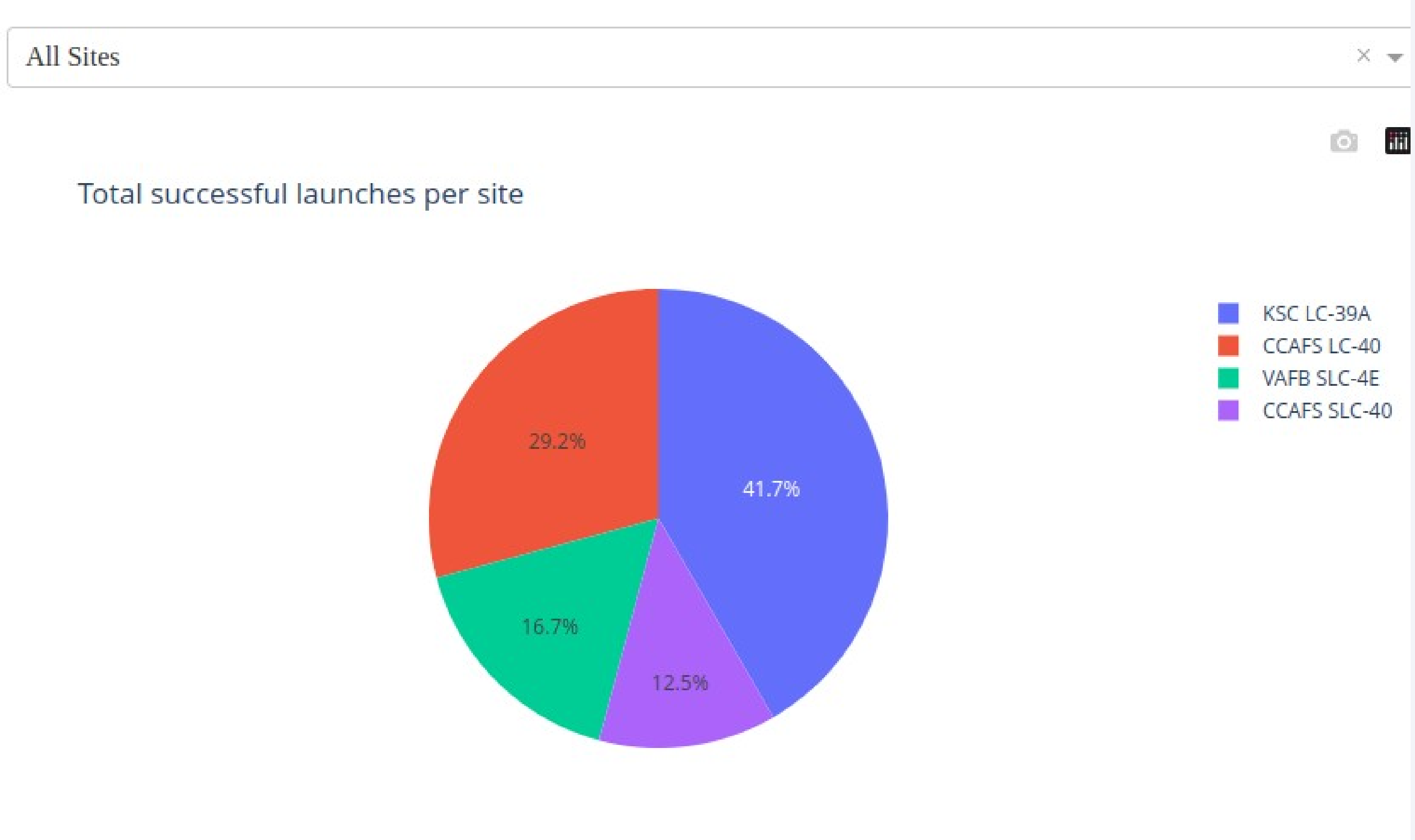
Folium map of proximity to the coast



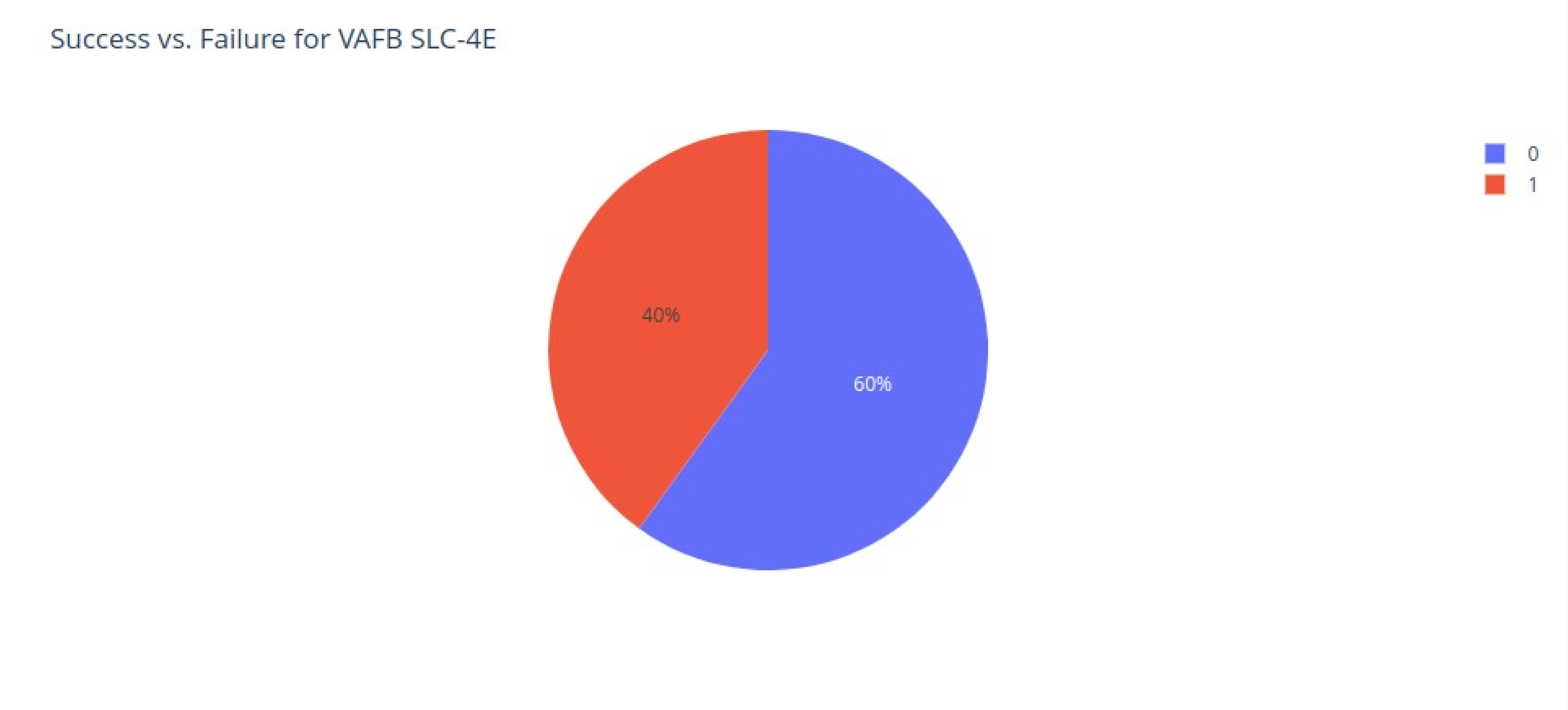
Section 4

Build a Dashboard with Plotly Dash

Total Successful Launches for All Sites



Highest Launch Success Ratio

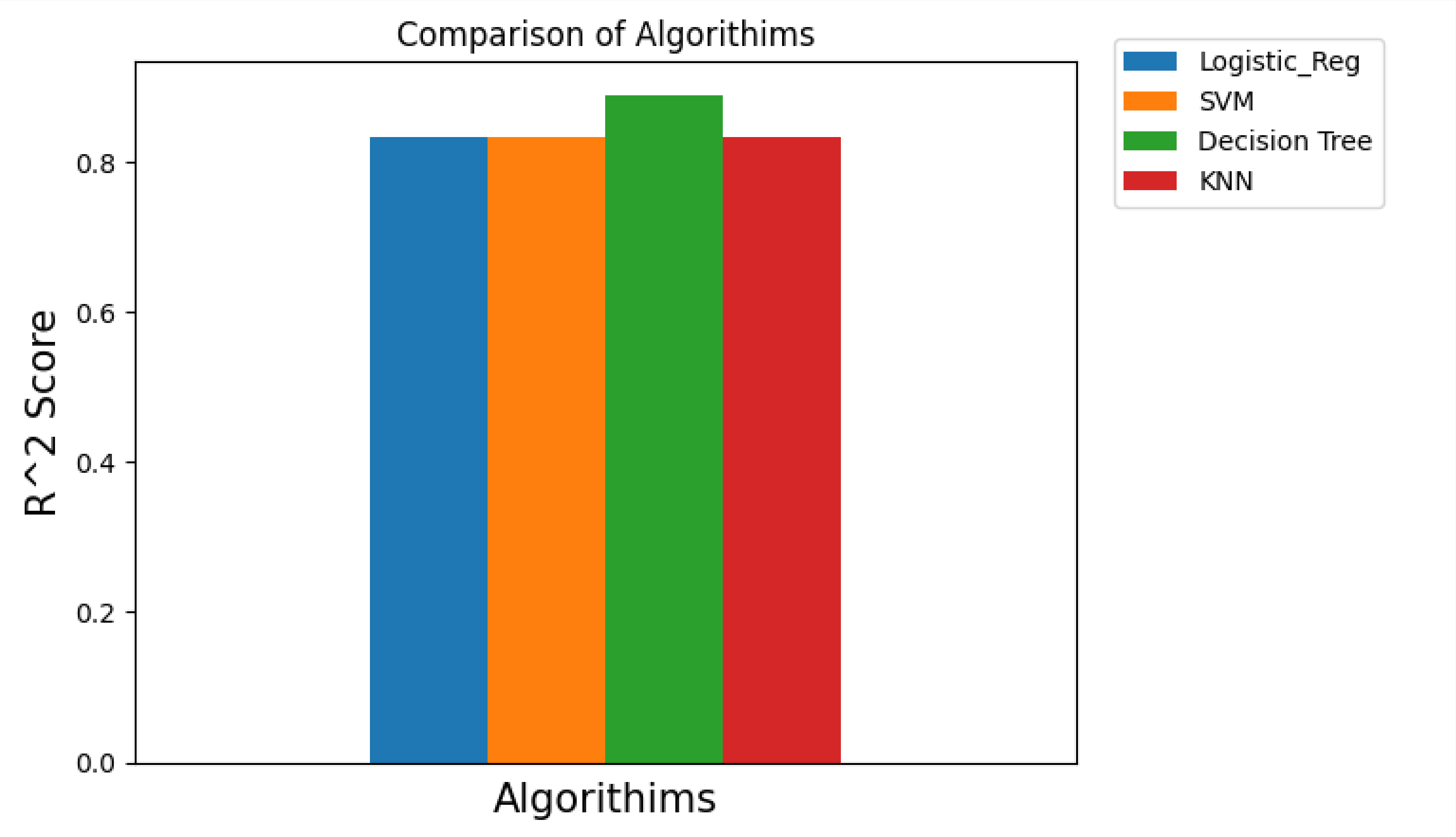


Payload vs. Launch for all sites

Section 5

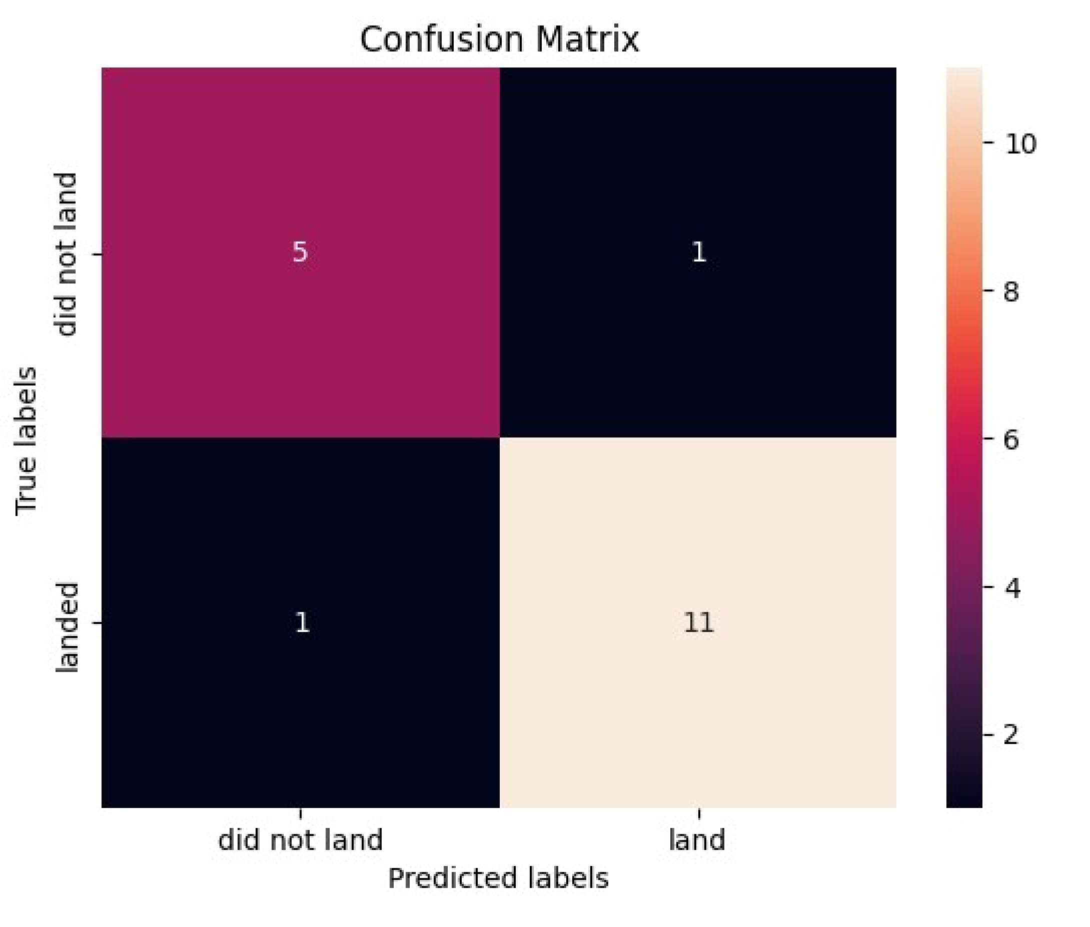
Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix

Confusion Matrix was best for the Decision Tree Algorithm, showing only 1 false positive and 1 false negative.



Conclusions

- The strongest algorithm for predictive analysis for mission success is the Decision Tree Algorithm.
- Successful launches and recoveries depend greatly on the size of the payload. Smaller payloads have greater success. Larger payloads work best a small number of launch sites.
- The success of missions improves as time passes and crews and companies have more experience.
- Site KSC-LC-39A has the best overall performance.
- ES11, GEO, HEO, and SSO orbit types have perfect performance records.

Thank you!