IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Raphael Graca

07/08/2022

Outline

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

Executive Summary

SpaceX is an American spacecraft manufacturer, space launch provider, and a satellite communications corporation. When it comes to spacecraft manufacturing, Falcon 9 is their most prominent rocket that can carry cargo and humans into Earth orbit (it is the first orbital class reusable rocket). Majority of launches take place in CA and FL, where there is a close vicinity of deep waters. The B4 boosters and Merlin engines are used for the heaviest payloads. Using various machine learning models, we have been able to predict launch outcomes correctly at the level of approx. 94.4% of the time. This high accuracy could be helpful for new companies attempting to break into the space race.

Introduction

- We will investigate SpaceX launch locations.
- We will define factors that affect SpaceX launch success.
- We will predict the success and failure of SpaceX launches using various machine learning methods.
- SpaceX can reuse its Falcon9 rockets, dramatically reducing the cost for space flight.
- If we can accurately predict the successful landing of the first stage, we can determine the costs of a launch.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected through a few different endpoint REST APIs and by means of web scrapping.
- Perform data wrangling
  - Data transformed, NaN values replaced with AVG, data scaled and standardized
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Used DataSpell by JetBrains to perform all classification models analysis and utilized Scikit-learn and Statsmodels to process machine learning activities.

Data Collection

- Used ‘requests’ library to make HTTP requests from the SpaceX API
- Defined helper functions to create DataFrame containing all necessary information
- Filter the data set so it could only include Falcon 9 launches
- Replaced values of the payload mass that contain nan with their mean
- Link to the lab at GitHub: *https://github.com/RapGAA/IBM-Final-Project/blob/main/Raphael%20Graca%20-%20Data%20Collection%20-%20API.ipynb*

Data Collection – SpaceX API

Import Libraries: requests, pandas, numpy, datetime

Define helper functions to extract necessary information:

- From Rocket: booster name
- From launchpad: launch site, position
- From payload: mass of payload, orbit
- From Cores: outcome of landing, type of landing, number of flights with core, version of core, number of core uses, serial number of core

Make request to static json url, convert response to pandas dataframe

Create new dataframe with response as well as outcomes from the beforementioned helper functions

Reduce dataframe to Falcon9 launches

Replace missing Payload mass values with mean

Data Collection – Web Scraping

- Requested Falcon9 Launch HTML Page
- Used Beautiful Soup and collected all relevant features
- Link to the lab at GitHub: https://github.com/RapGAA/IBM-Final-Project/blob/main/Raphael%20Graca%20-%20Web%20Scraping.ipynb

Data Wrangling

Determine data types
• Find missing values
• Calculate number of launches at each site
• Calculate number of each orbit type
• Calculate outcomes of each orbit type
• Apply a landing outcome binary encoding (0, 1)
• Link to the lab at GitHub: https://github.com/RapGAA/IBM-Final-Project/blob/main/Raphael%20Graca%20-%20Data%20wrangling.ipynb

EDA with Data Visualization

Flight number vs Launch Site: see how launch sites were utilized through time
• Payload vs Launch Site: find patterns in which sites may be used for different payload masses
• Success rates of orbit types (bar chart): check for relationships between success rate and orbit type
• Flight number and Orbit Type: see how orbit types have changed through time
• Payload vs orbit: which payload masses are destined for certain orbits
• Yearly trend of success: see changes in success rate through time
• Summarize what charts were plotted and why you used those charts
• Link to the lab at GitHub: https://github.com/RapGAA/IBM-Final-Project/blob/main/Raphael%20Graca%20-%20EDA%20-%20Data%20Visualisation.ipynb

EDA with SQL

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

- Selected all records from SpaceX Table.
- Retrieved all unique launch site names.
- Find 5 records with launch site beginning with ‘CCA’.
- Find total payload mass for boosters by NASA(CRS).
- Find average payload mass carried by booster F9 v1.1
- Link to the lab at GitHub: https://github.com/RapGAA/IBM-Final-Project/blob/main/Raphael%20Graca%20-%20EDA%20-%20SQL.ipynb

Build an Interactive Map with Folium

Show all launch sites

- Display success/failures at each site (find spatial patterns)
- Add latitude/longitude indicator for courser position (in order to find the coordinates)
- Add lines and corresponding distances from a selected launch site to coastline, highway, railroad, and city (to see how close/far each is from launch site)

Build a Dashboard with Plotly Dash

Pie Chart: successful launches by site
• Scatterplot: Payload vs success, with color indicating booster version
• Lets user see which boosters are most successful for various payload masses and
launch site(s)
• Drop-down menu to allow user to isolate data visualizations by launch site
• Slider bar to narrow the range of payloads displayed in scatterplot

Predictive Analysis (Classification)

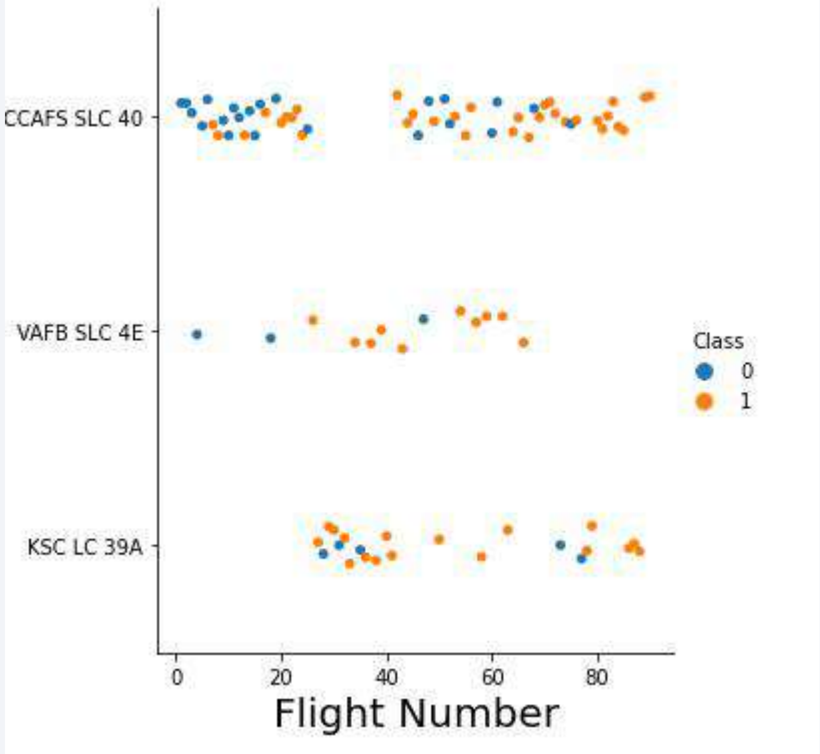
Split data into train/test sets
- Run 4 classification models
- Logistic regression
- Support vector machine
- Decision tree
- K-nearest neighbours
- Determine best performing model by testing on test set, and determine accuracy score
- Use confusion matrix to further evaluate model performance

Section 2

Insights drawn from EDA

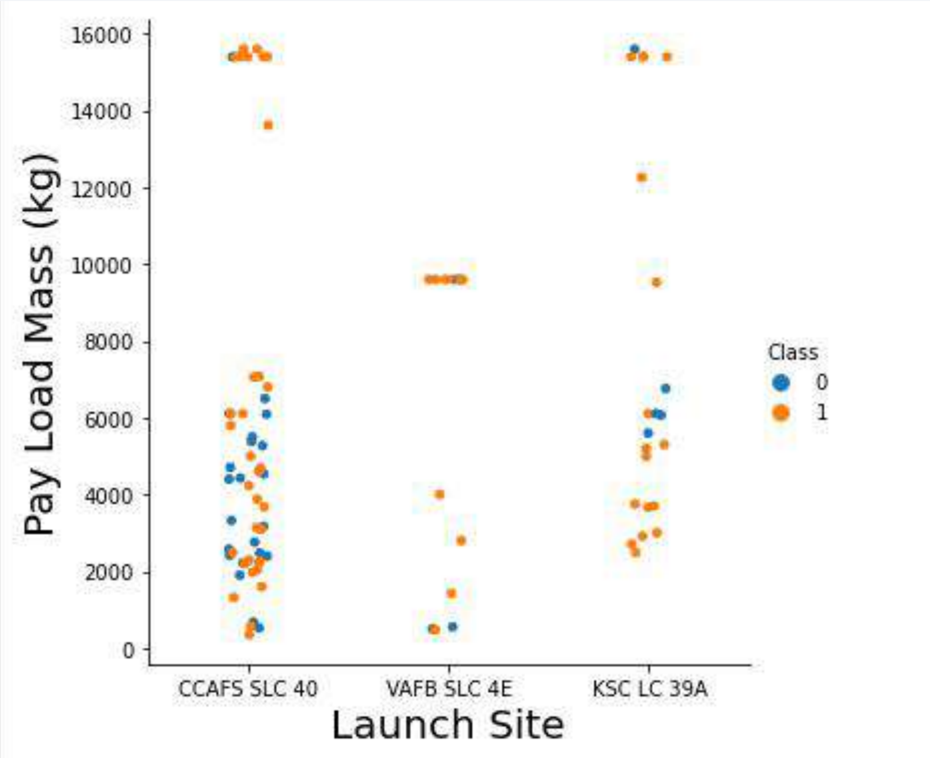
Flight Number vs. Launch Site

CCAFS SLC 40
- Most used
- Gap in use
- KSC LC 39A
- Most use for flight during gap at CCAFS SLS 40
- VAFB SLC 4E
- Least Used
- Has not been used Recently



Payload vs. Launch Site

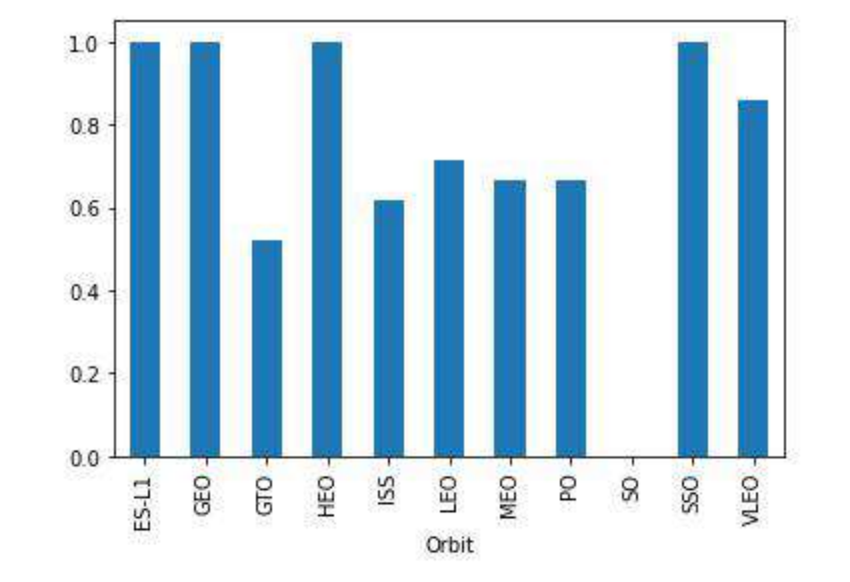
- There are no payloads greater than 10000kg at VAFB SLC 4E
- CCAFS SLC 40 has the greatest range of payloads



Success Rate vs. Orbit Type

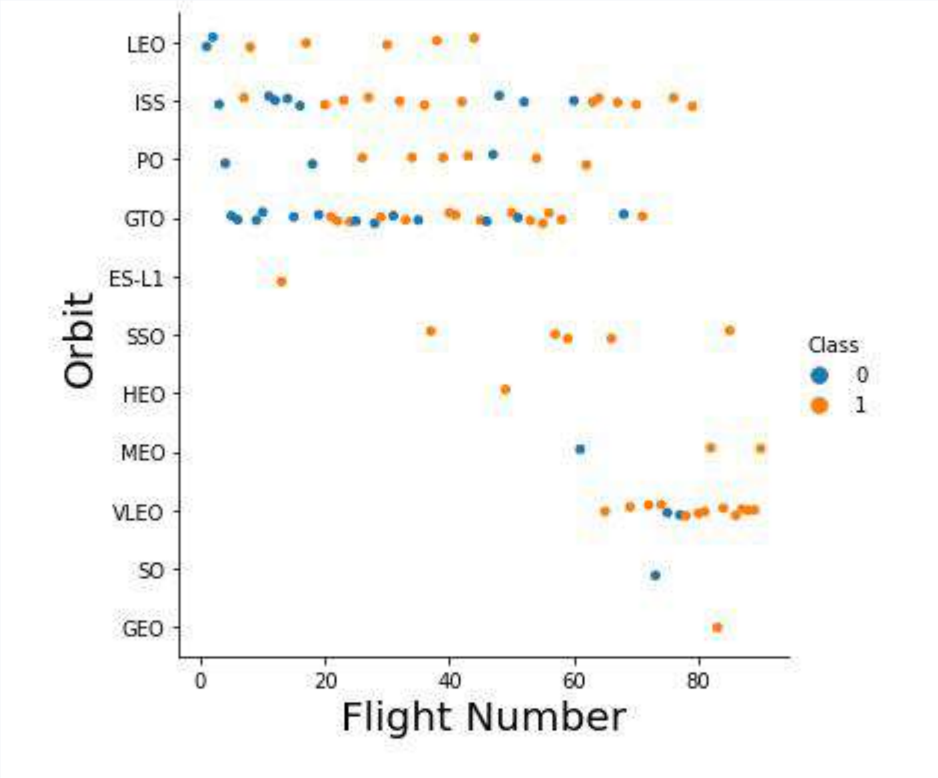
Orbits with a 100% success rate:

- ES-L1
- GEO
- HEO
- SSO
- Orbit of type “SO” had a 0% success rate



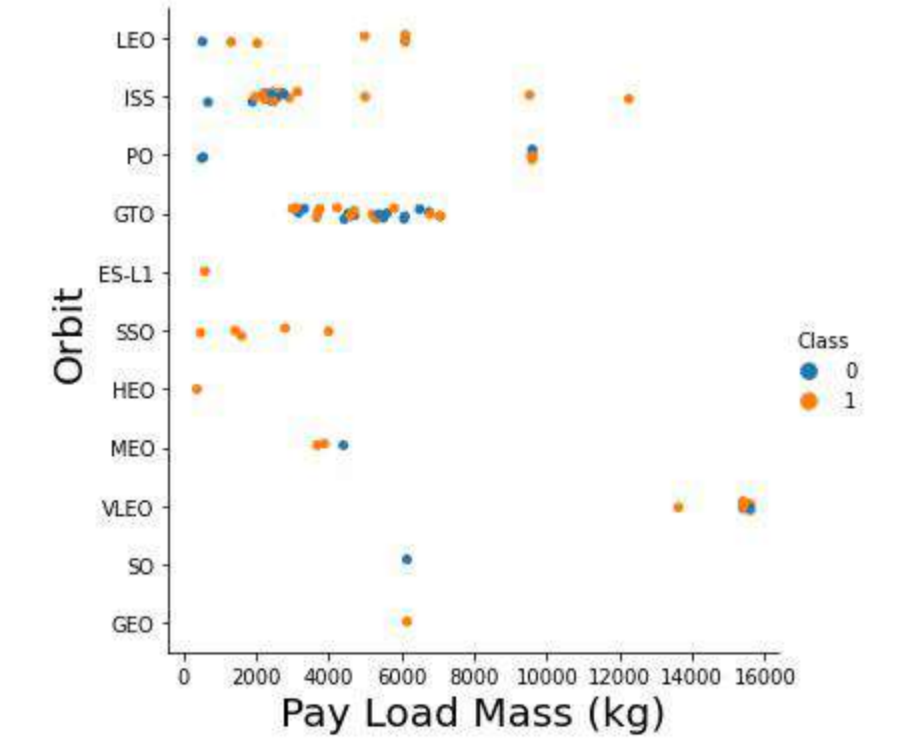
Flight Number vs. Orbit Type

- LEO orbits have been more successful as Flight Number progresses
- GTO orbits have experienced intermittent failures, even as Flight Number progresses
- ISS orbit has the largest range of Flight Number value



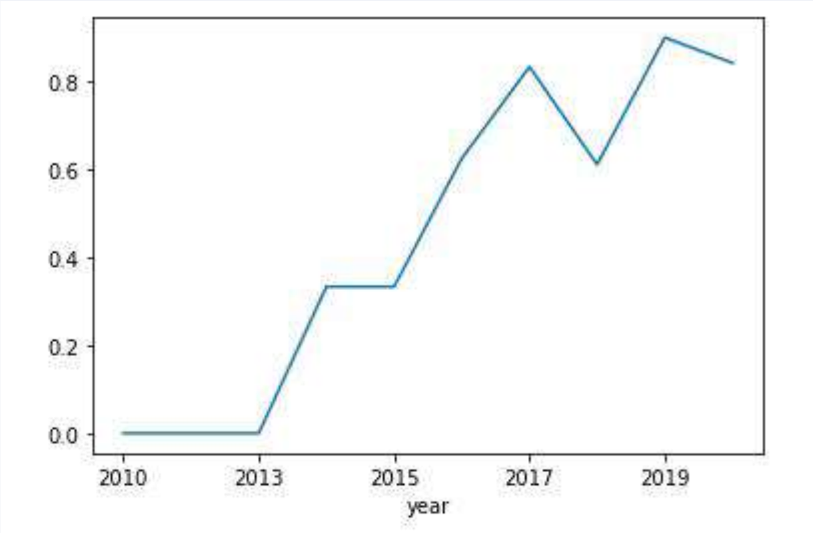
Payload vs. Orbit Type

- VLEO Orbit payloads are characterized by high mass
- GTO orbit payloads have had a mix of landing outcomes
- LEO and ISS have had increased success rate with increased payload



Launch Success Yearly Trend

- Launch success has generally increased over time (2013 to 2020)
- There was a slight dip in success during 2018



All Launch Site Names

- There are 4 Launch Sites:
- CCAFS LC-40
- CCAFS SLC-40
- KSC LC-39A
- VAFB SLC-4E

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

In [11]:
```
%sql select unique(launch_site) from SPACEXTBL
```

Out[11]:

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

Launch Site Names Begin with 'CCA'

- The first 5 records with launch site beginning with ‘CCA’ are above.
- The first record is from 2010

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

In [17]:
```
%sql select * from SPACEXTBL where launch_site like '%CCA%' limit 5
```

Out[17]:

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

- SPACEX boosters carried 48,213 kg payload for NASA (CRS)

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

In [19]:
```
%sql select sum(payload_mass__kg_) from SPACEXTBL where customer LIKE '%NASA (CRS)%'
```

Out[19]:

1
48213

Average Payload Mass by F9 v1.1

- Booster version F9 v1.1 has carried an average payload mass of 2,534kg

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

```
In [22]: %sql select avg(payload_mass__kg_) from SPACEXTBL where booster_version LIKE '%F9 v1.1%'
```

Out[22]:

1
2534

First Successful Ground Landing Date

- The first successful landing outcome on ground pad occurred on Dec 22 2015.

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

Hint:Use min function

In [26]:
```
%sql select min(DATE) from SPACEXTBL where landing__outcome LIKE '%Success (ground pad)%'
```

Out[26]:

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

The names of boosters which have successfully landed on drone ship and had a payload mass greater than 4000kg but less than 6000kg include:

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

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

In [32]:
```
%sql select unique(booster_version) from SPACEXTBL WHERE landing__outcome LIKE 'Success (drone ship)' AND payload_mass__kg_ > 4000 AND payload_mass__kg_ < 6000
```

Out[32]:

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

Total Number of Successful and Failure Mission Outcomes

- There was only 1 full failure outcome, which occurred in flight
- There was 1 success with an unknown outcome for the payload
- The rest of the missions were presented as successes (99)

List the total number of successful and failure mission outcomes

In [37]:
```
%sql select mission_outcome, count(*) from SPACEXTBL group by mission_outcome
```

Out[37]:

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- All of the booster versions which carried the maximum payload mass began with “F9 B5 B10XX.X”

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

```
In [39]: %sql select unique(booster_version) FROM SPACEXTBL WHERE payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXTBL)
```

Out[39]:

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

2015 Launch Records

In 2015, there were 2 failed landings on a drone ship:

- Both originated from CCAFS LC-40
- Booster F9 v1.1 B1012 on 2015-01-10
- Booster F9 v1.1 B1015 on 2015-04-14

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

In [42]:
```
%sql select booster_version, launch_site, DATE from SPACEXTBL where landing__outcome LIKE '%Failure (drone ship)%' AND year(DATE)=2015
```

Out[42]:

booster_version	launch_site	DATE
F9 v1.1 B1012	CCAFS LC-40	2015-01-10
F9 v1.1 B1015	CCAFS LC-40	2015-04-14

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

Between the specified dates, the most common landing outcome was “no attempt”
The next most common was a tie:
- Failure on drone ship (5)
- Success on drone ship (5)

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

```
In [55]: %sql select landing__outcome, COUNT(*) as count from SPACEXTBL where DATE > '2010-06-04' AND DATE < '2017-03-20' GROUP BY landing__outcome order by count DESC
```

Out[55]:

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

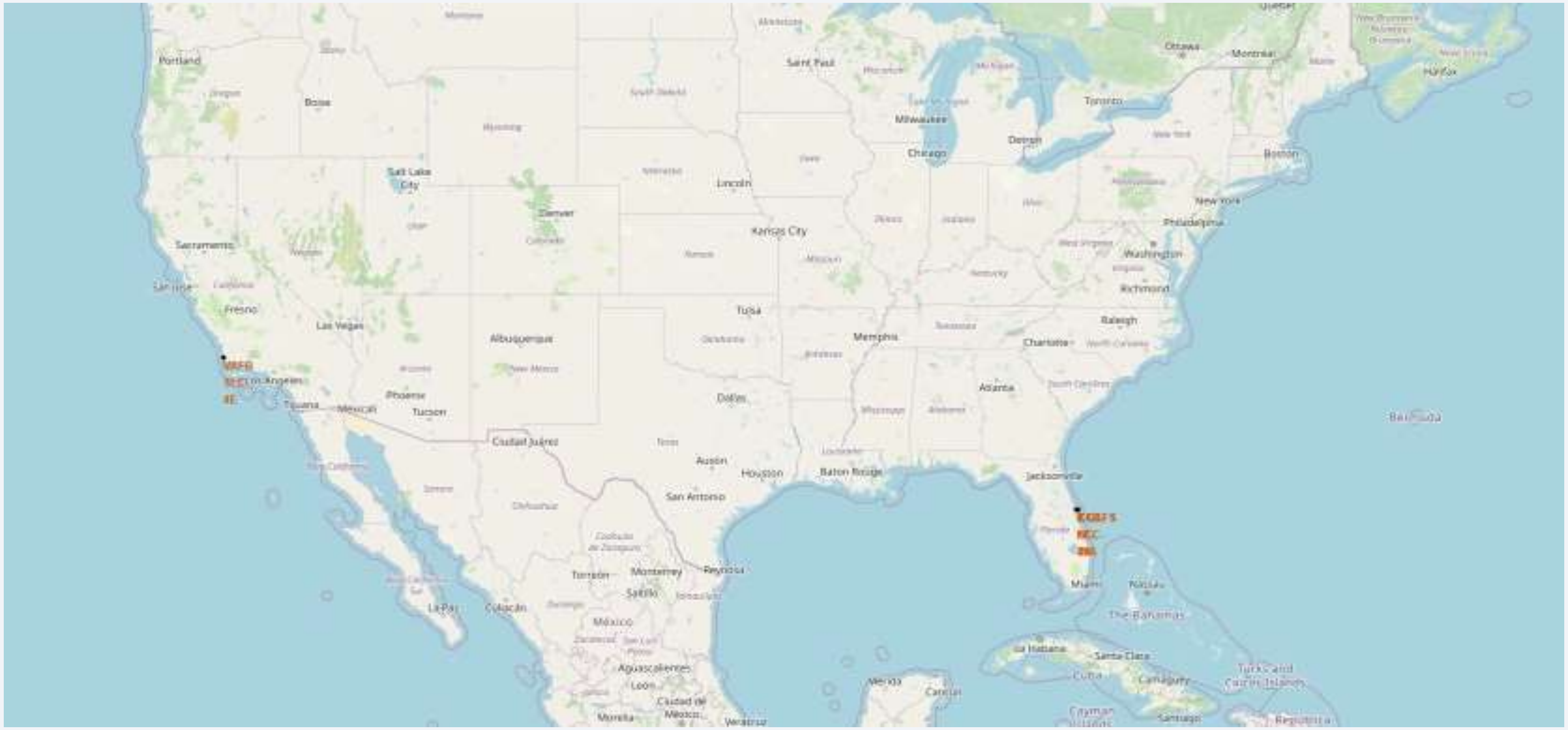
Section 3

Launch Sites Proximities Analysis

Folium Map – SpaceX Launch Sites (Screenshot 1)

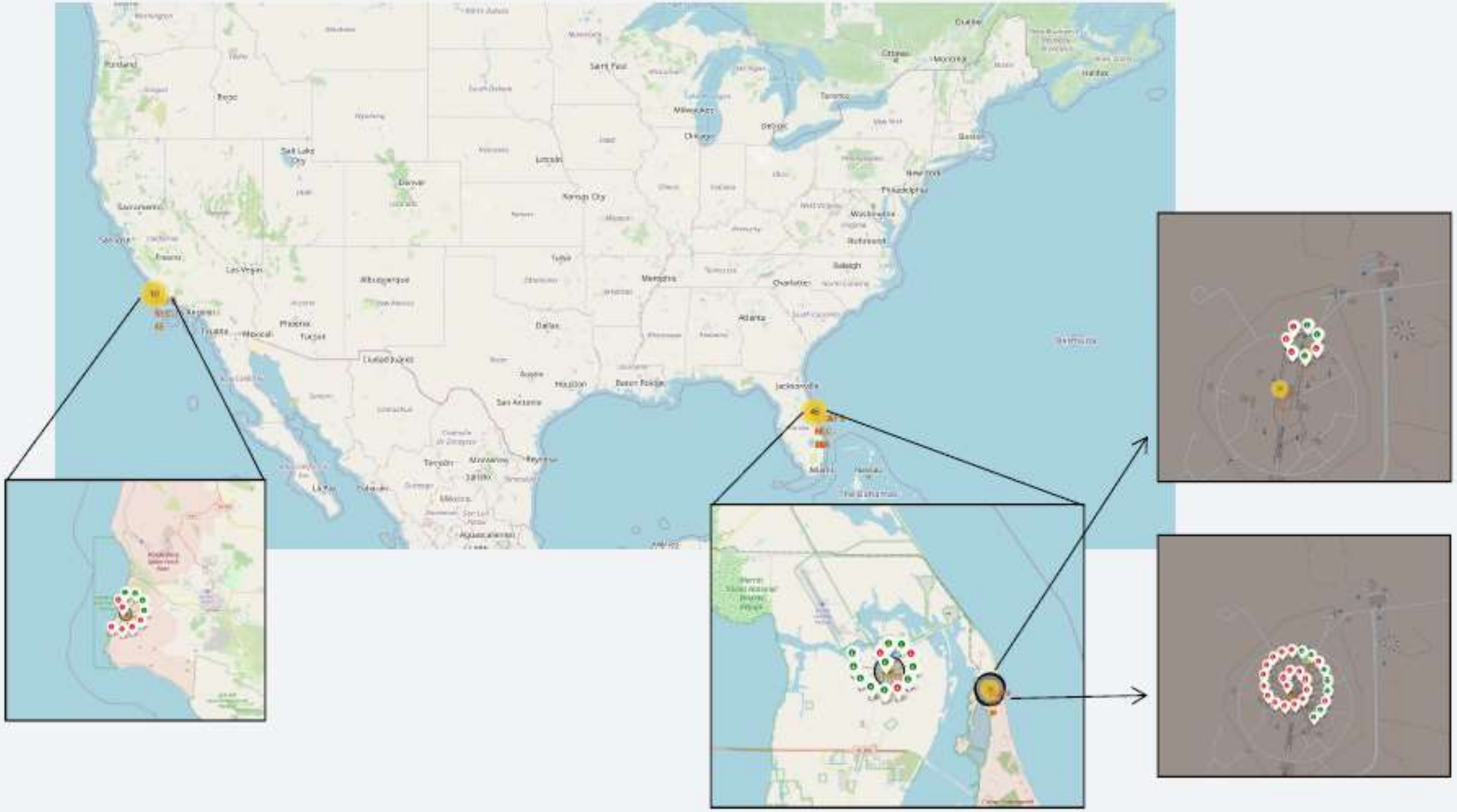
- All Launch Sites are in California or Florida, USA

- Launch Sites appear to be on Pacific and Atlantic coasts, not in the middle of the mainland



Specific Launch Sites Visualisation (Screenshot 2)

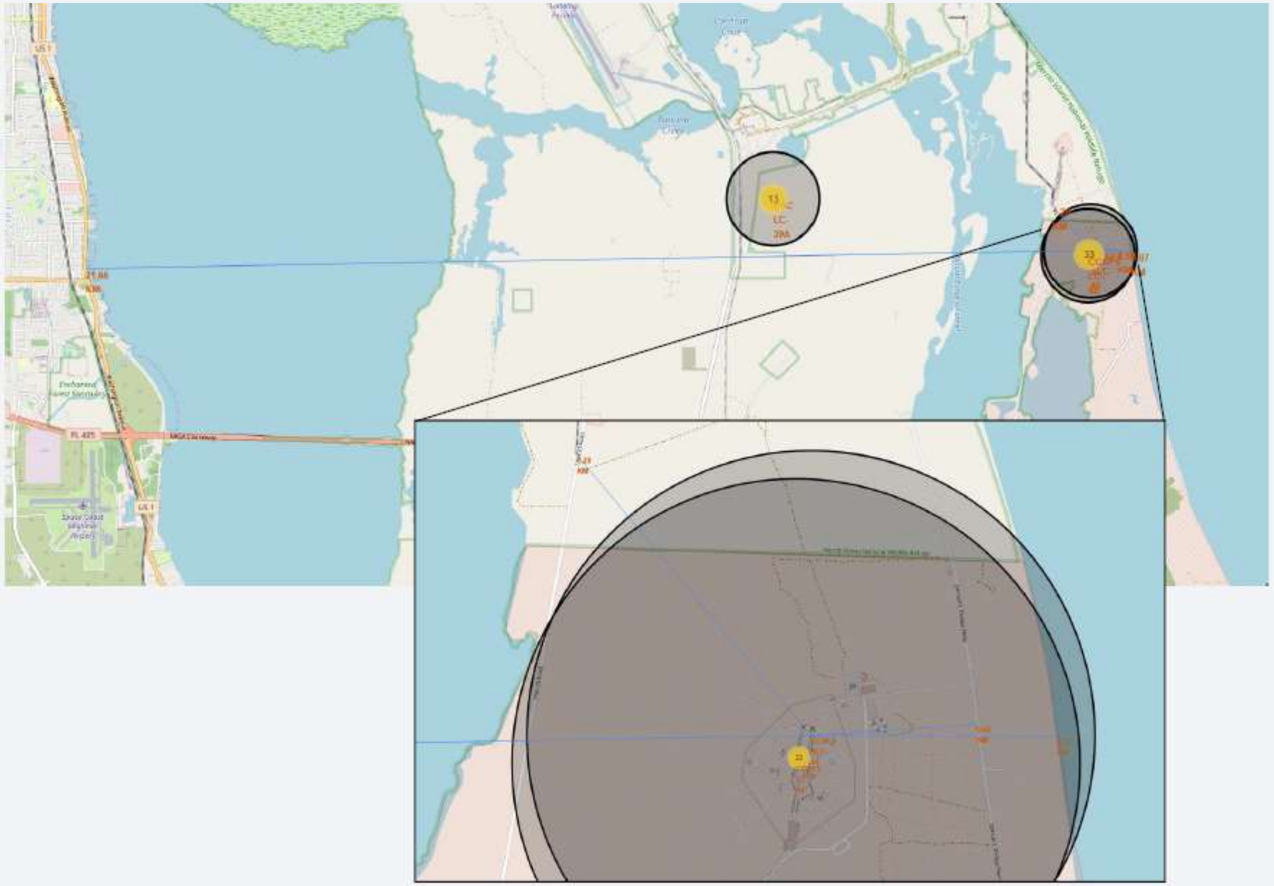
- There are more launch sites in Florida than California
- Site CCAFS SLC-40 had the most launches
- KSC LC-39A had the highest success rate



CCAFS SLC-40 Distance to Selected Features (Screenshot 3)

- CCAFS SLC-40 Launch Site is Somewhat far away from the closest city, Indian River City (~22km)

- This site is close to a highway (~0.6km), a coastline (0.87km), and a Railroad (~1.3km)

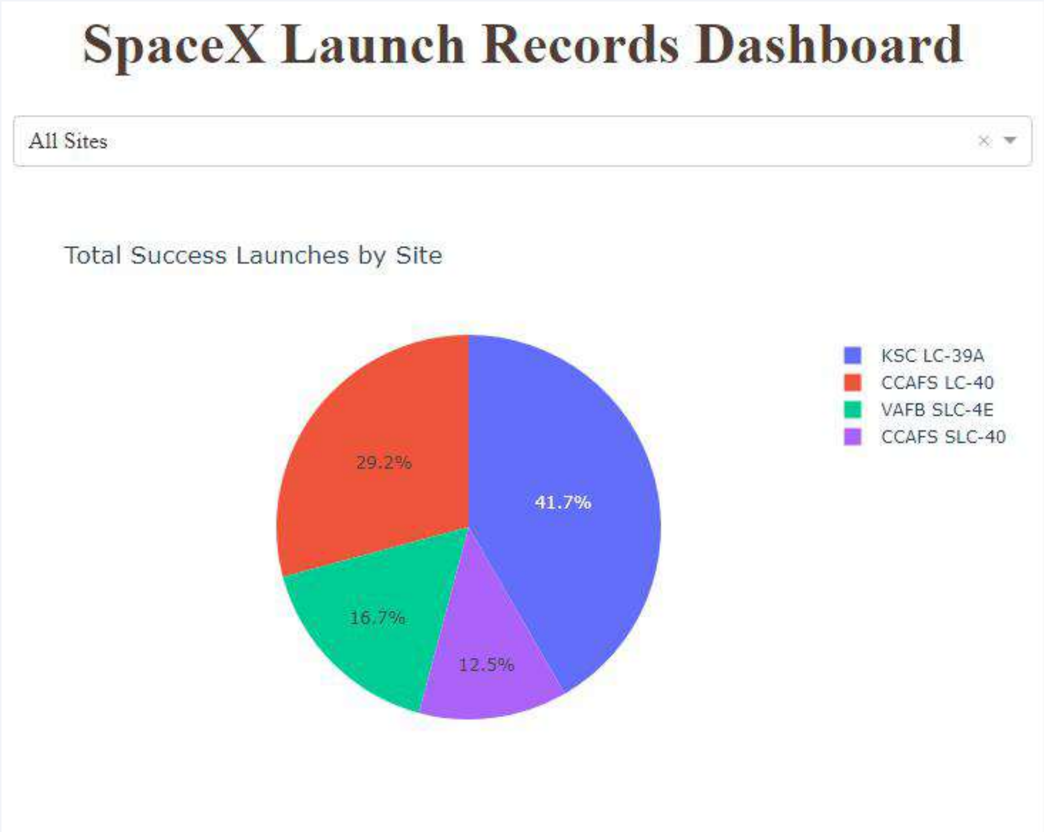


Section 4

Build a Dashboard with Plotly Dash

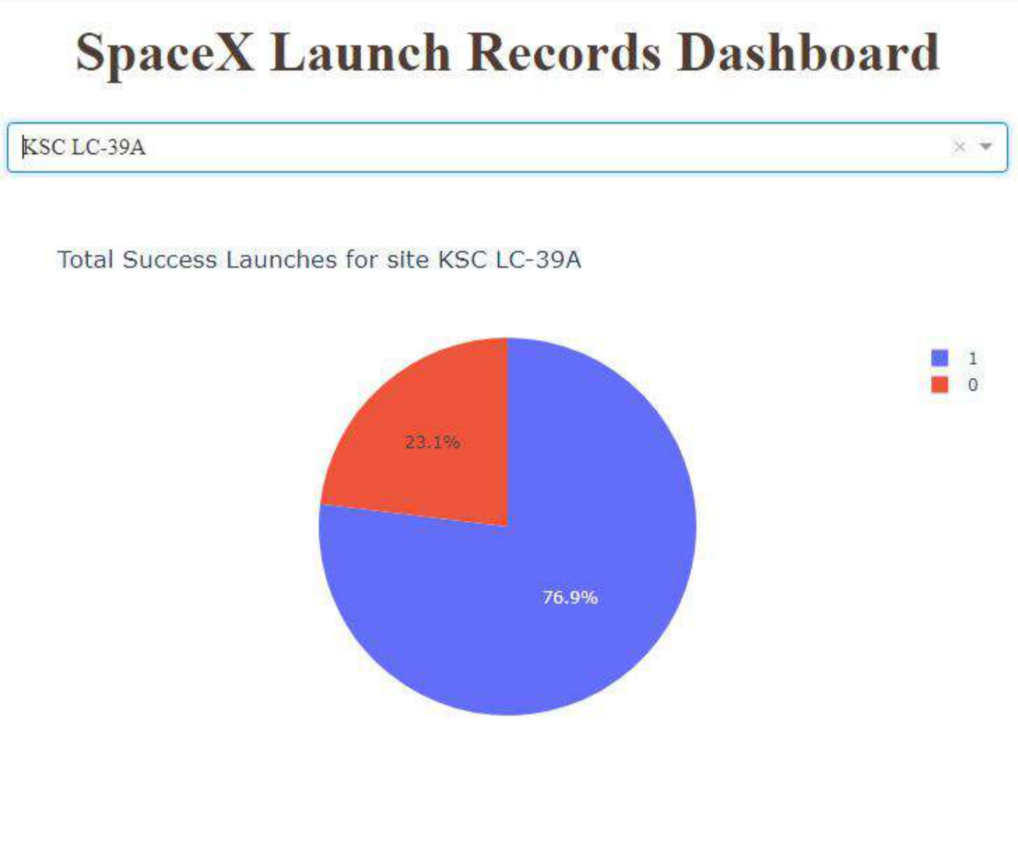
SpaceX Launch Records Dashboard: Launch Success by Site

- KSC LC-39A had the most successes
- CCAFS LC-40 had the second most successes
- CCAFS SLC-40 had the least number of successes



SpaceX Launch Records Dashboard: Launch Success at KSC LC-39A

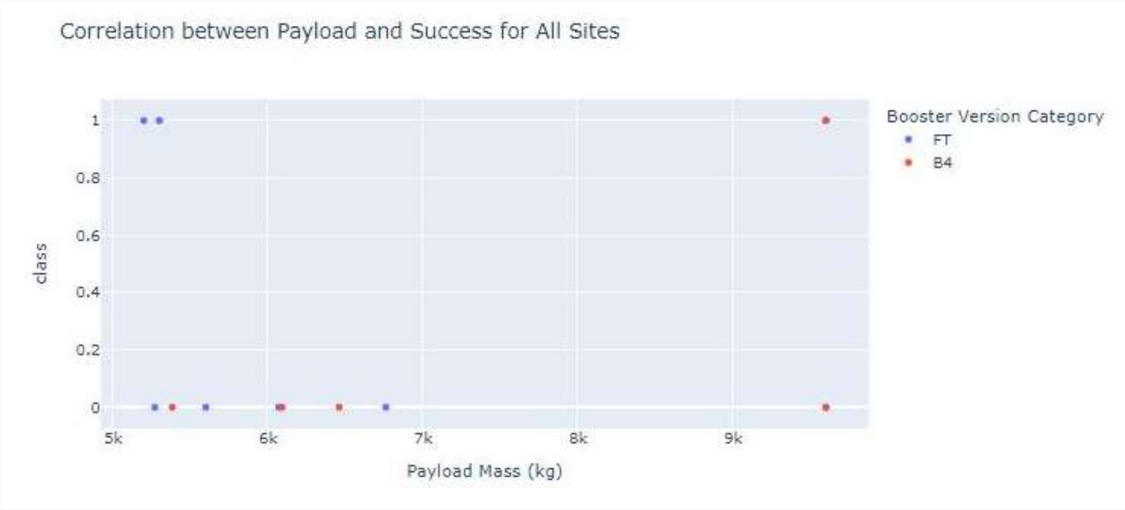
- KSC LC-39A had a 76.9% success rate
- KSC LC-39A had a 23.1% fail rate

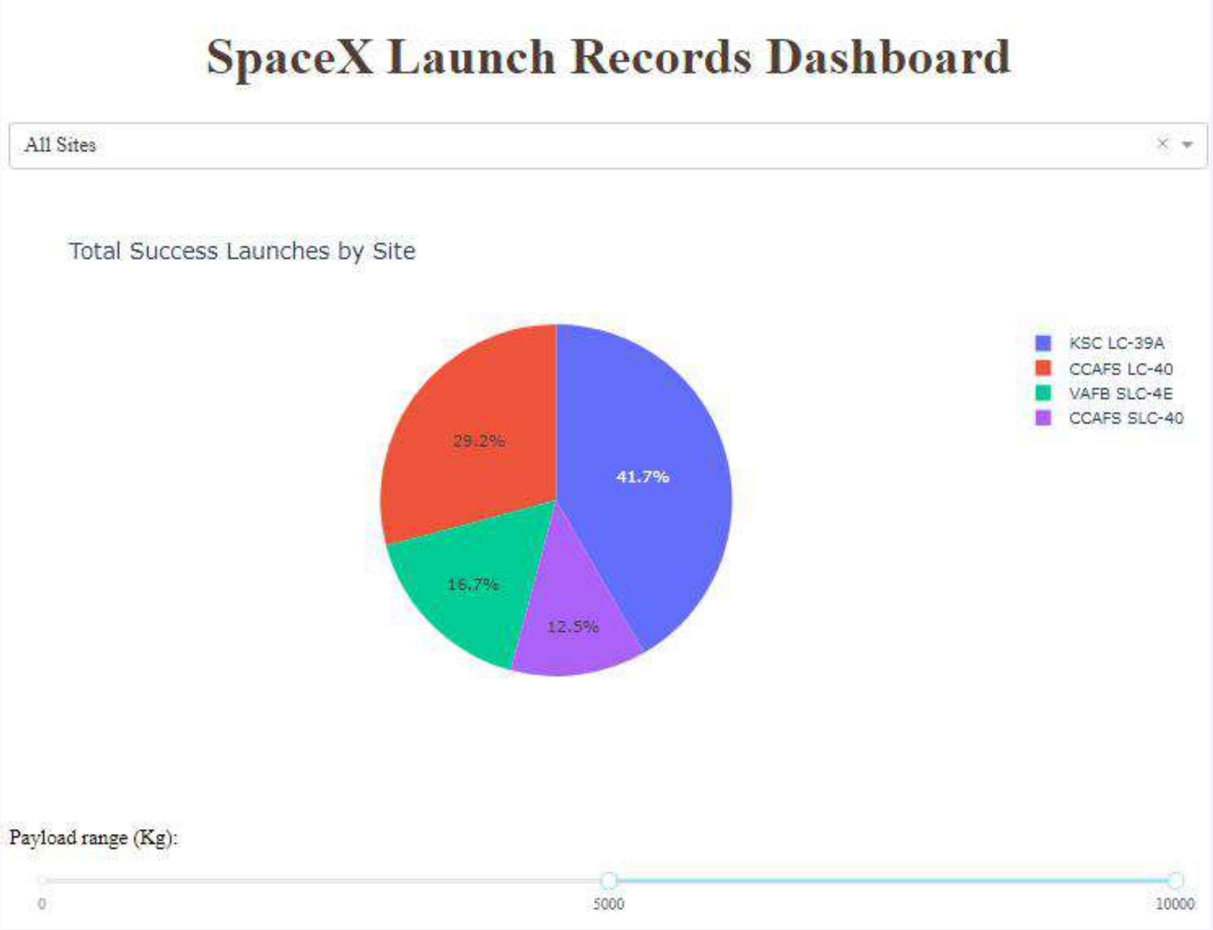


SpaceX Launch Records Dashboard: Payload vs Launch Outcome

The slider ranges from 5,000 to 10,000kg.
Only 2 boosters are represented:
- FT & B4

Both boosters show similar success rate.
B4 Boosters carry the greatest payloads.



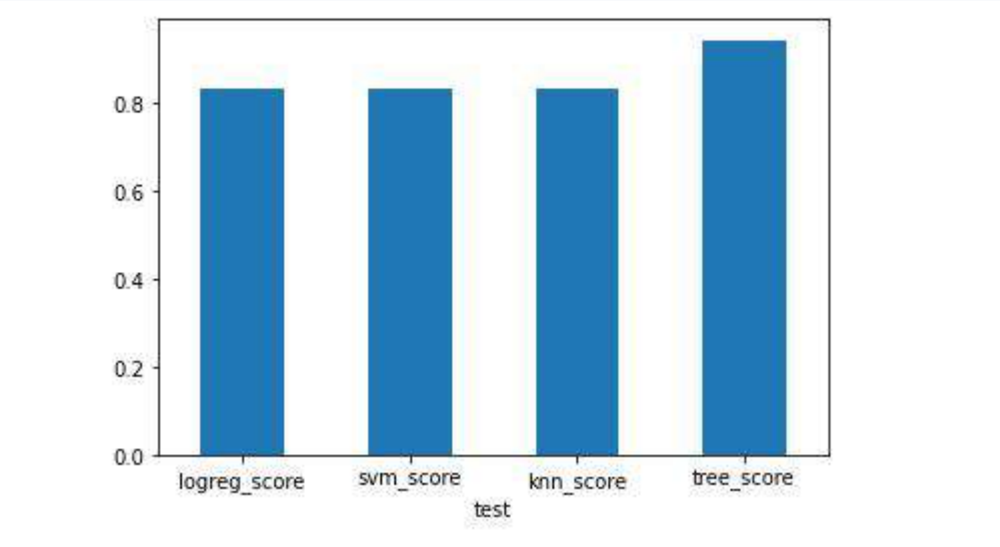


Section 5

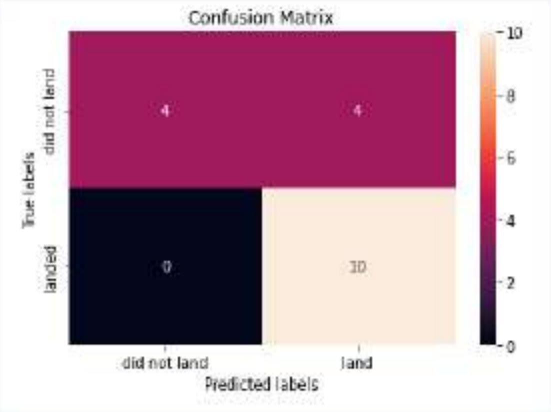
Predictive Analysis (Classification)

Classification Accuracy

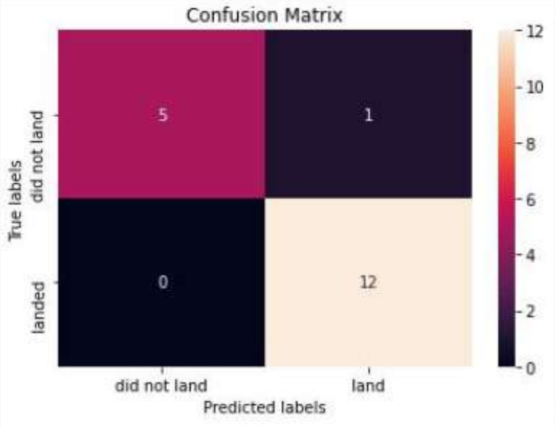
- All models performed above 80% accuracy
- The Decision Tree (tree_score) had the greatest accuracy on the test set (94.4% accurate)



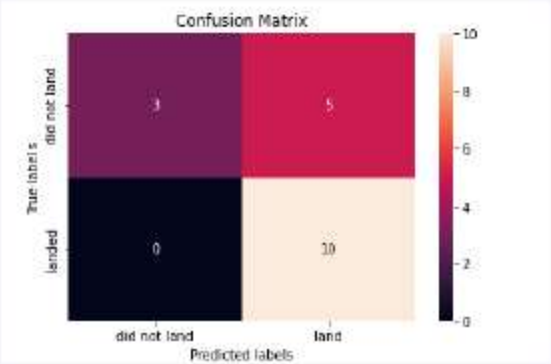
Confusion Matrices



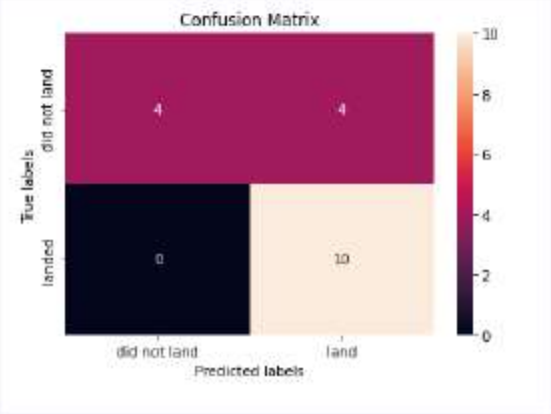
Logistic Regression



Decision Tree



SVM



KNN

Conclusions

- SpaceX Uses Launch Sites in Florida and California.
- Launch Sites are in close proximity to modes of transportation and the coast, while further away from cities.
- Most of SpaceX’s missions have been successful, especially as time has progressed.
- VLEO orbit types are used for the greatest payload masses.
- SpaceX typically uses B4 boosters for very heavy payloads.
- The decision tree predictor was highly accurate in predicting landings.
- ES-L1, GEO, HEO and SSO orbit types have high success rate.
- KSC LC-39A have the highest number of success launches.
- Payload range between 0kg to 5000kg performs better than the heavier payloads.

Thank you!