IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Greg Sorg
December 2021

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

SpaceX missions have had greater success through time, launching exclusively in California and Florida, USA. These launch sites are typically located in close proximity to coasts, highways, and railroads while keeing some distance from cities. A variety of orbit types are utilized by SpaceX, although VLEO is used for the greatest payload masses. The B4 boosters are used for the heaviest payloads. By utilizing a trained decision tree, we were able to predict launch outcomes correctly ~94% of the time. This high accuracy could be helpful for new companies attempting to break into the space race.

Introduction

- 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
- We will look into SpaceX launch locations
- We will examine various influences on SpaceX launch success
- We will try to predict the success or failure of a SpaceX launch using machine learning techniques.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was gathered via SpaceX API
- Perform data wrangling
  - Data cleaned from missing values, narrowed to needed data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Use SciKitLearn to build and evaluate classification models

Data Collection – SpaceX API

- *Used 'requests' library to make HTTP requests from the SpaceX API*
- *Defined helper functions to create dataframe of necessary information*
- *Reduced dataframe to only Falcon9 launches*
- *Replace missing payload masses with the mean*
- Link to Data Collection API Lab:
  https://github.com/SorgGD/CourseraFinalProject/blob/master/Data%20Collection%20API%20Lab.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 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
- Assign a landing outcome label (0, 1)

- https://github.com/SorgGD/CourseraFinalProject/blob/master/Data%20Wrangling%20Lab%20Coursera%20Capstone.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
- https://github.com/SorgGD/CourseraFinalProject/blob/master/EDA%20with%20Visualization%20(week2%20cont.).ipynb

EDA with SQL

- Connected to DB2 instance on IBM Cloud
- Selected all records from SpaceXTable
- Retrieve 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
- Find first date of successful landing on the ground pad
- Find booster names that successfully landed on drone ship with payload mass between 4000 and 6000 kg
- Find total successful and failed missions
- Find booster names that carried max payload mass
- https://github.com/SorgGD/CourseraFinalProject/blob/master/Exploratory%20Data%20Analysis.ipynb

Build an Interactive Map with Folium

- Show all launch sites
- Display success/failures at each site (find spatial patterns)
- Add lat/long indicator for mouse position (in order to know 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)
- https://github.com/SorgGD/CourseraFinalProject/blob/master/Visual%20Analytics%20with%20Folium.ipynb

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
- https://github.com/SorgGD/CourseraFinalProject/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

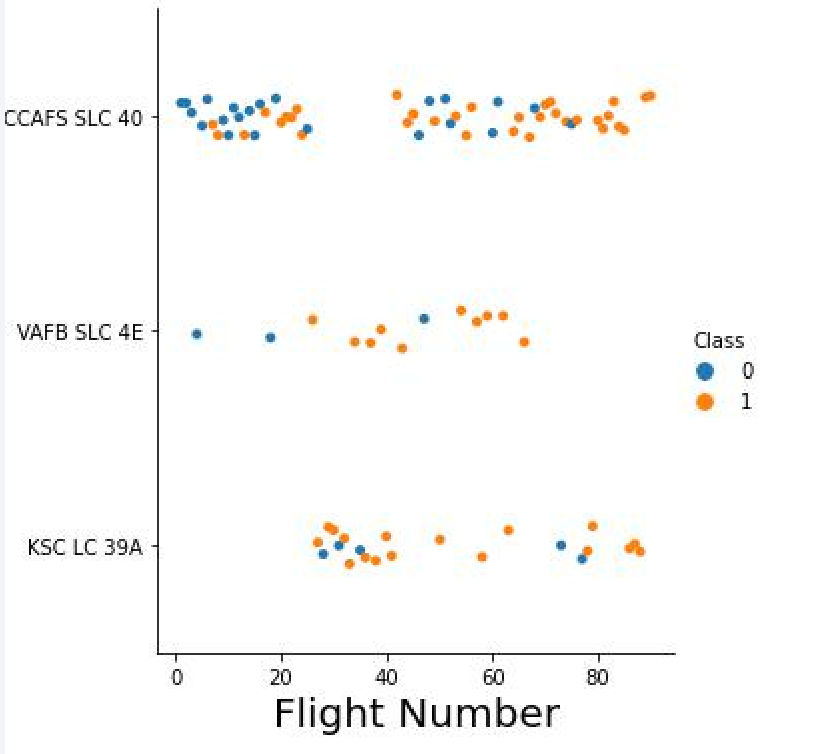
- Split data into train/test sets
- Run 4 classification models
  - Logistic regression
  - Support vector machine
  - Decision tree
  - K-nearest neighbors
- Determine best performing model by testing on test set, and determine accuracy score
- Use confusion matrix to further evaluate model performance
- https://github.com/SorgGD/CourseraFinalProject/blob/master/Machine%20Learning%20Prediction%20Lab.ipynb

Section 2

Insights drawn from EDA

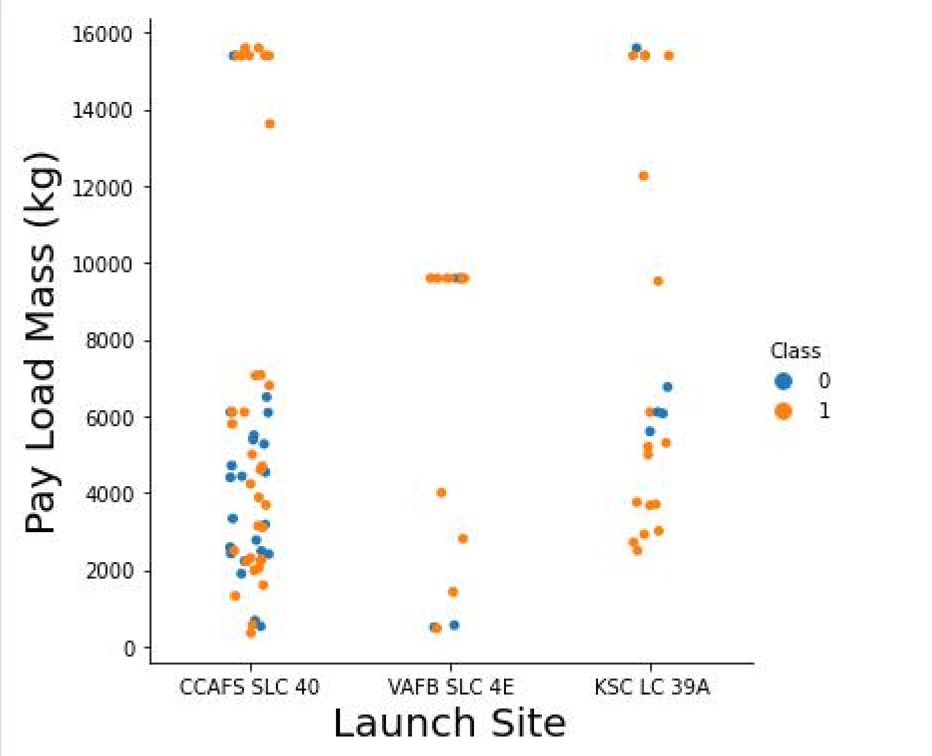
Flight Number vs. Launch Site

- 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



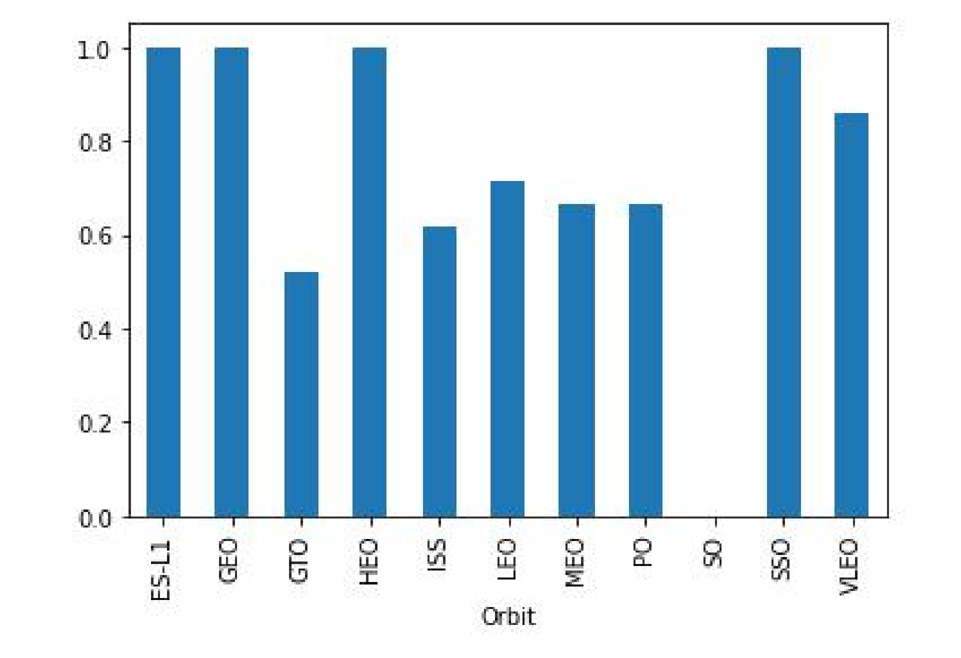
Payload vs. Launch Site

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



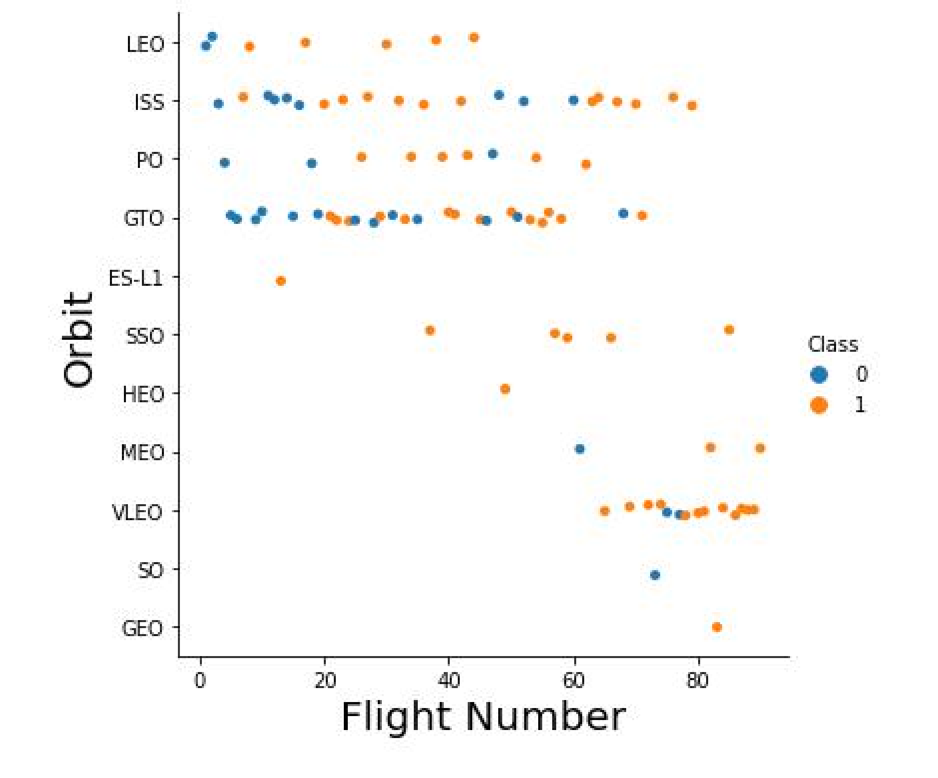
Success Rate vs. Orbit Type

- Orbits with a 100% success rate:
  - ES-L1
  - GEO
  - HEO
  - SSO
- Orbit of type "SO" had a 0% success rate



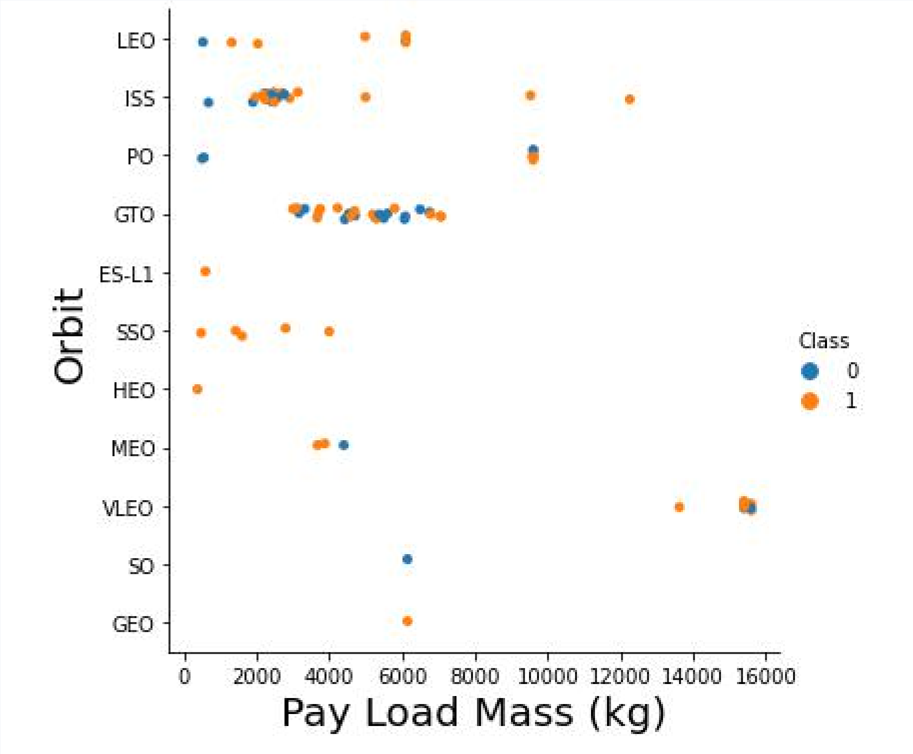
Flight Number vs. Orbit Type

- *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*



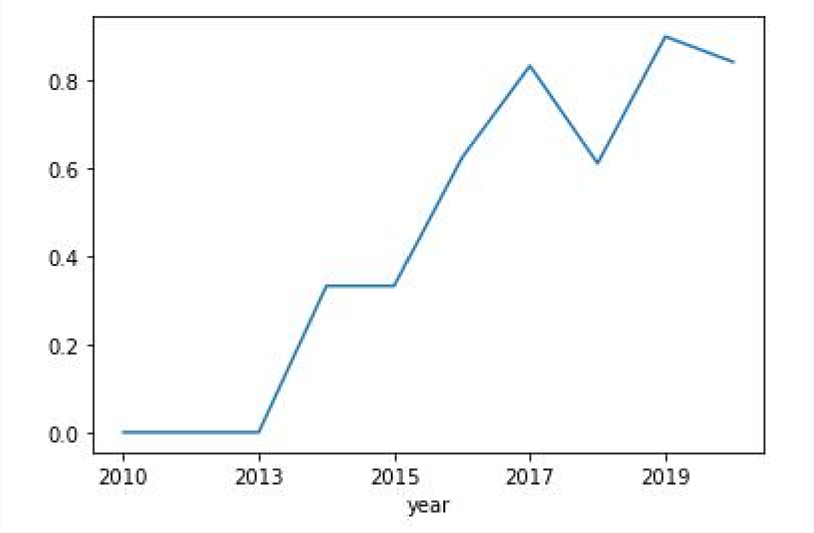
Payload vs. Orbit Type

- 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
```

```
 * ibm_db_sa://xvv36624:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

Out[11]:

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

Launch Site Names Begin with 'CCA'

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

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

```
 * ibm_db_sa://xvv36624:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

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

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

Total Payload Mass

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)%'
```

```
 * ibm_db_sa://xvv36624:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

Out[19]:

1
48213

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

Average Payload Mass by F9 v1.1

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%'

 * ibm_db_sa://xvv36624:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.

Out[22]:
```

1
2534

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

First Successful Ground Landing Date

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)%'

 * ibm_db_sa://xvv36624:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

Out[26]:

1
2015-12-22

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

Successful Drone Ship Landing with Payload between 4000 and 6000

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
```

```
 * ibm_db_sa://xvv36624:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

Out[32]:

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

- 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

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

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

```
 * ibm_db_sa://xvv36624:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

Out[37]:

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

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

Boosters Carried Maximum Payload

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)

 * ibm_db_sa://xvv36624:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

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

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

2015 Launch Records-Failed Landings: Drone Ship

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

 * ibm_db_sa://xvv36624:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

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

- 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

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

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
```

```
 * ibm_db_sa://xvv36624:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

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

- 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)

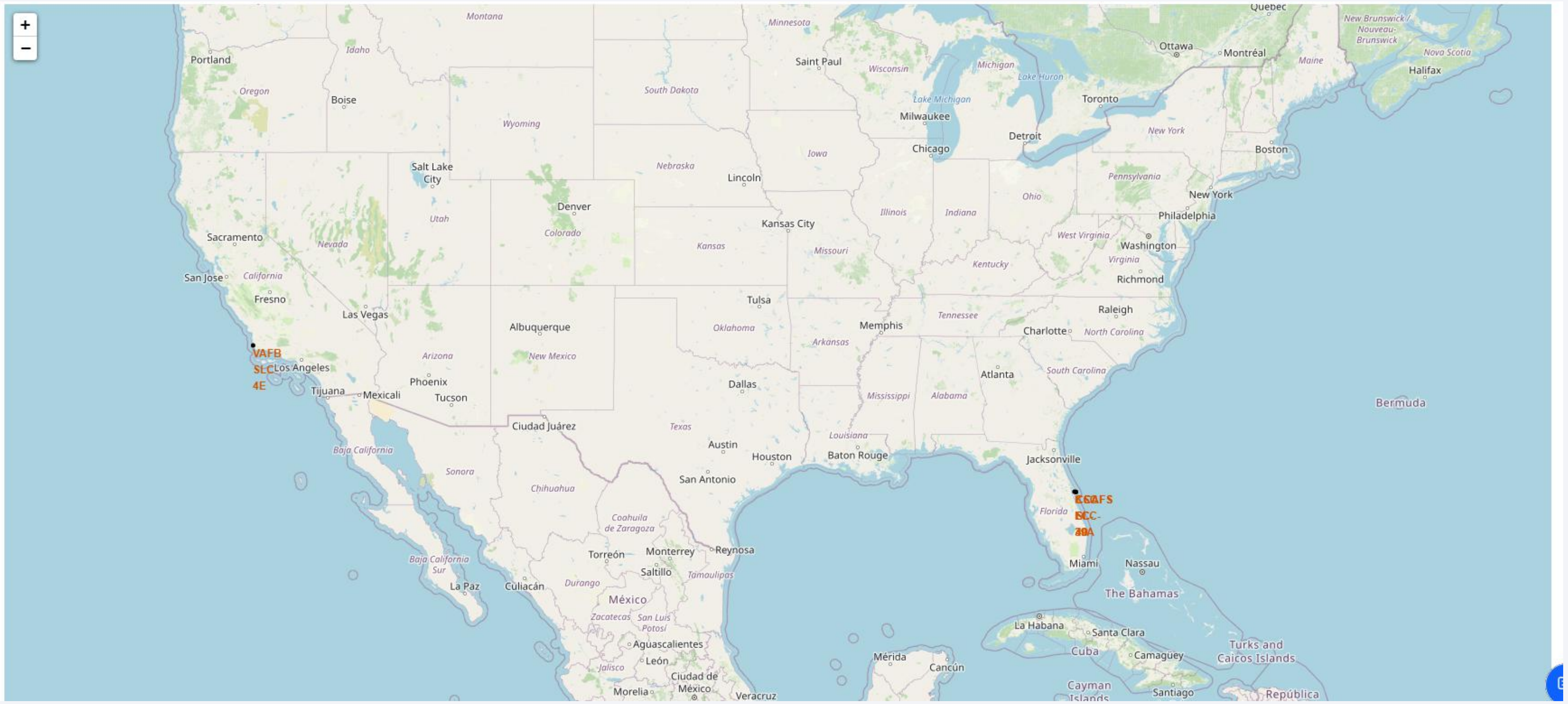
Section 4

Launch Sites Proximities Analysis

SpaceX Launch Sites (Folium)

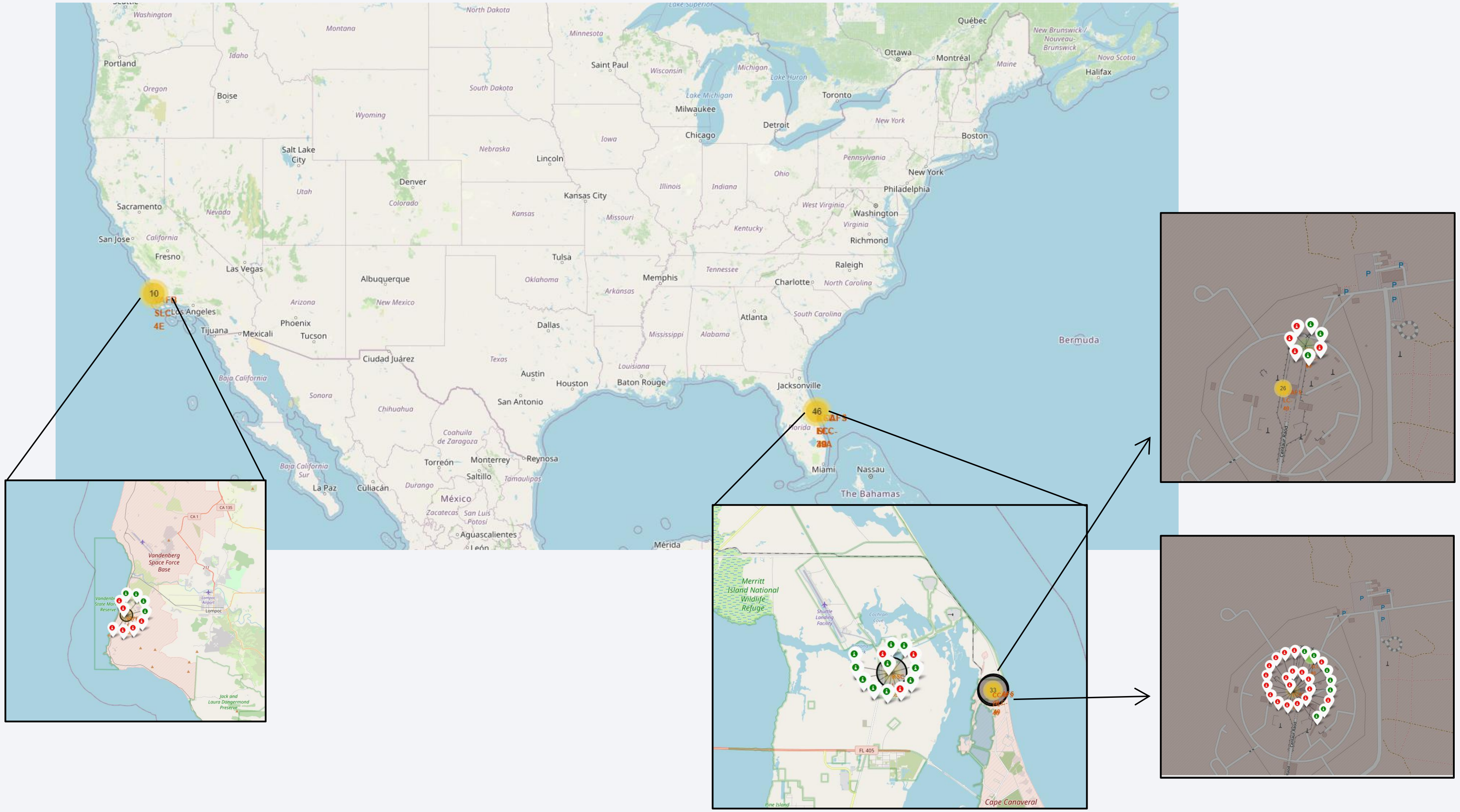
- All Launch Sites are in California or Florida, USA

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



Launch Outcomes (Folium)

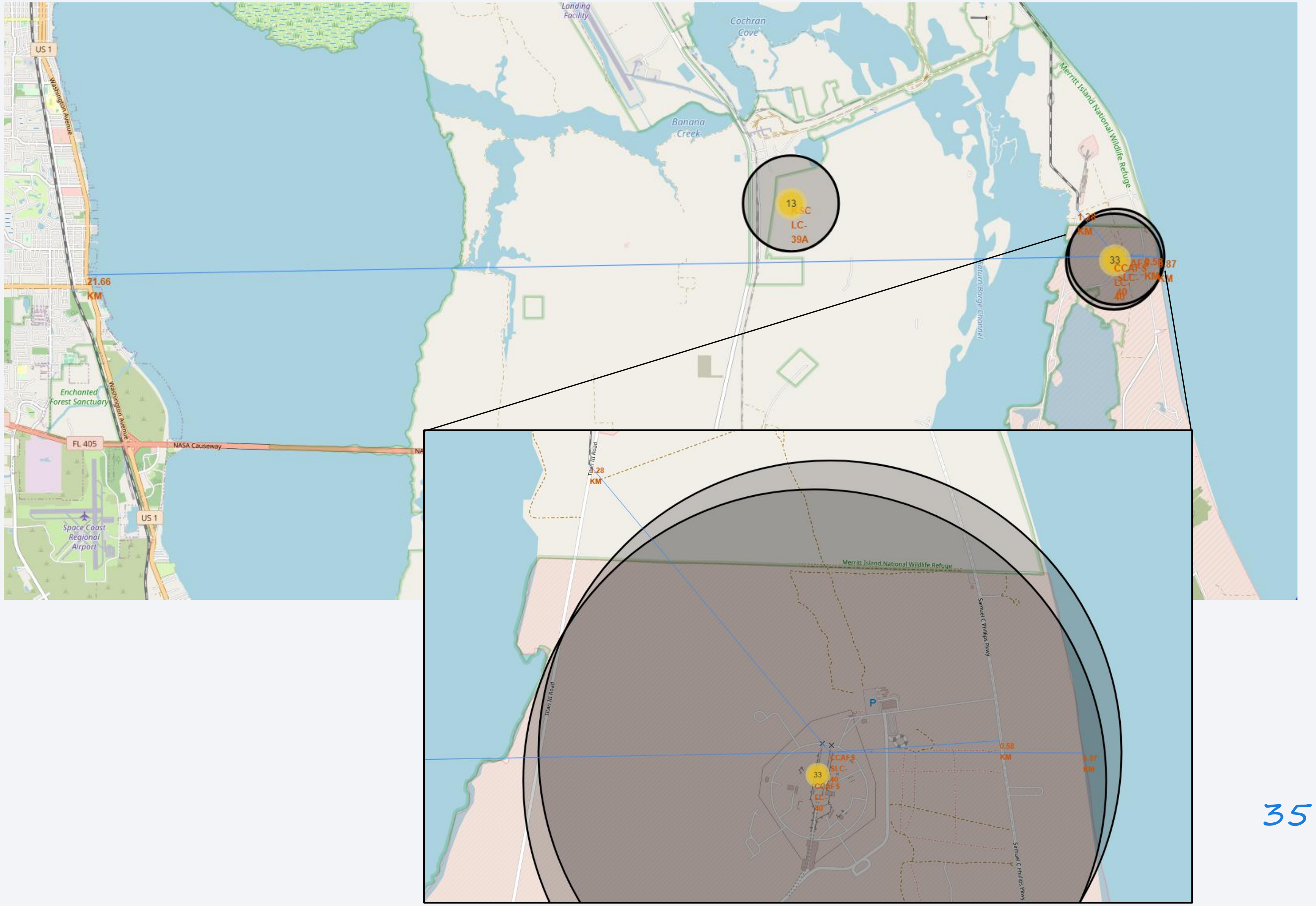
- There are more launch sites in Florida than California
- Site CCAFS SLC-40 had the most launches
- KSC LC-39A had the highest success rate



CCAFS SLC-40 Distance to Selected Features (Folium)

- CCAFS SLC-40 Launch Site is Somwhat 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)

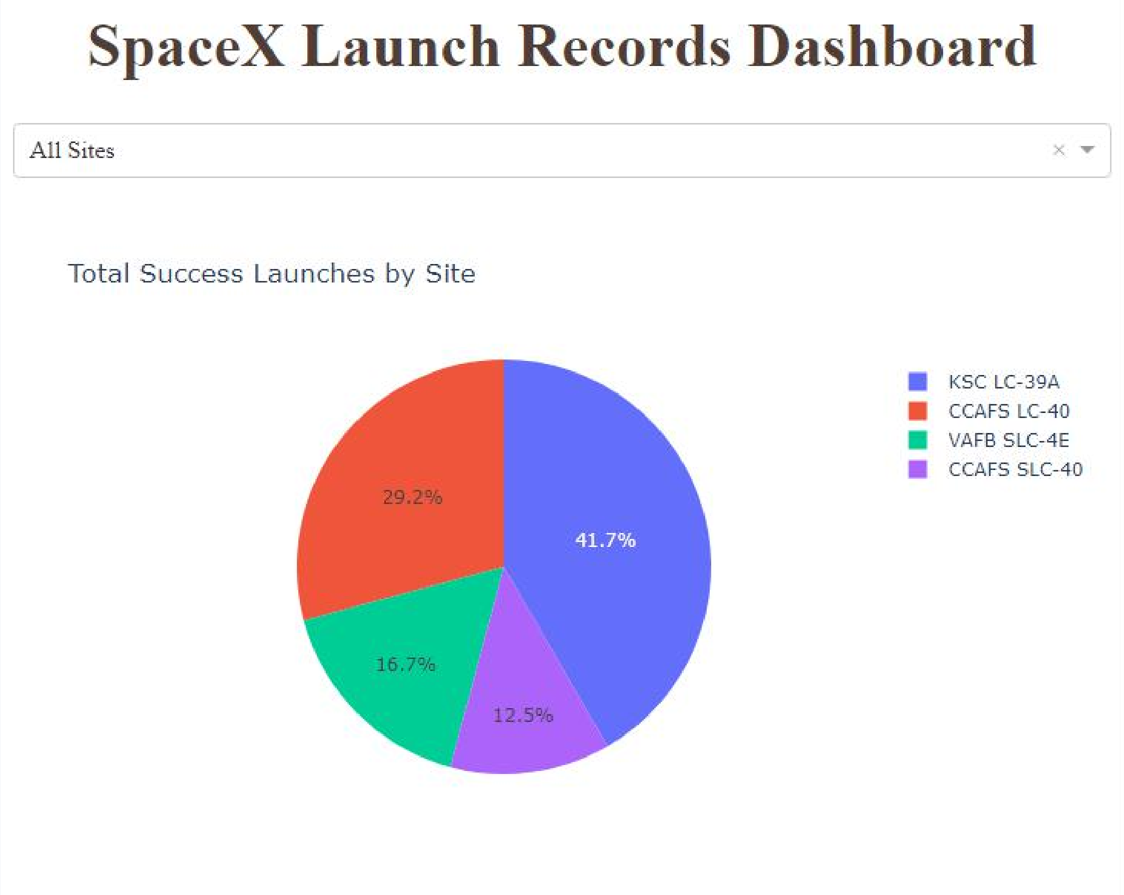


Section 5

Build a Dashboard with Plotly Dash

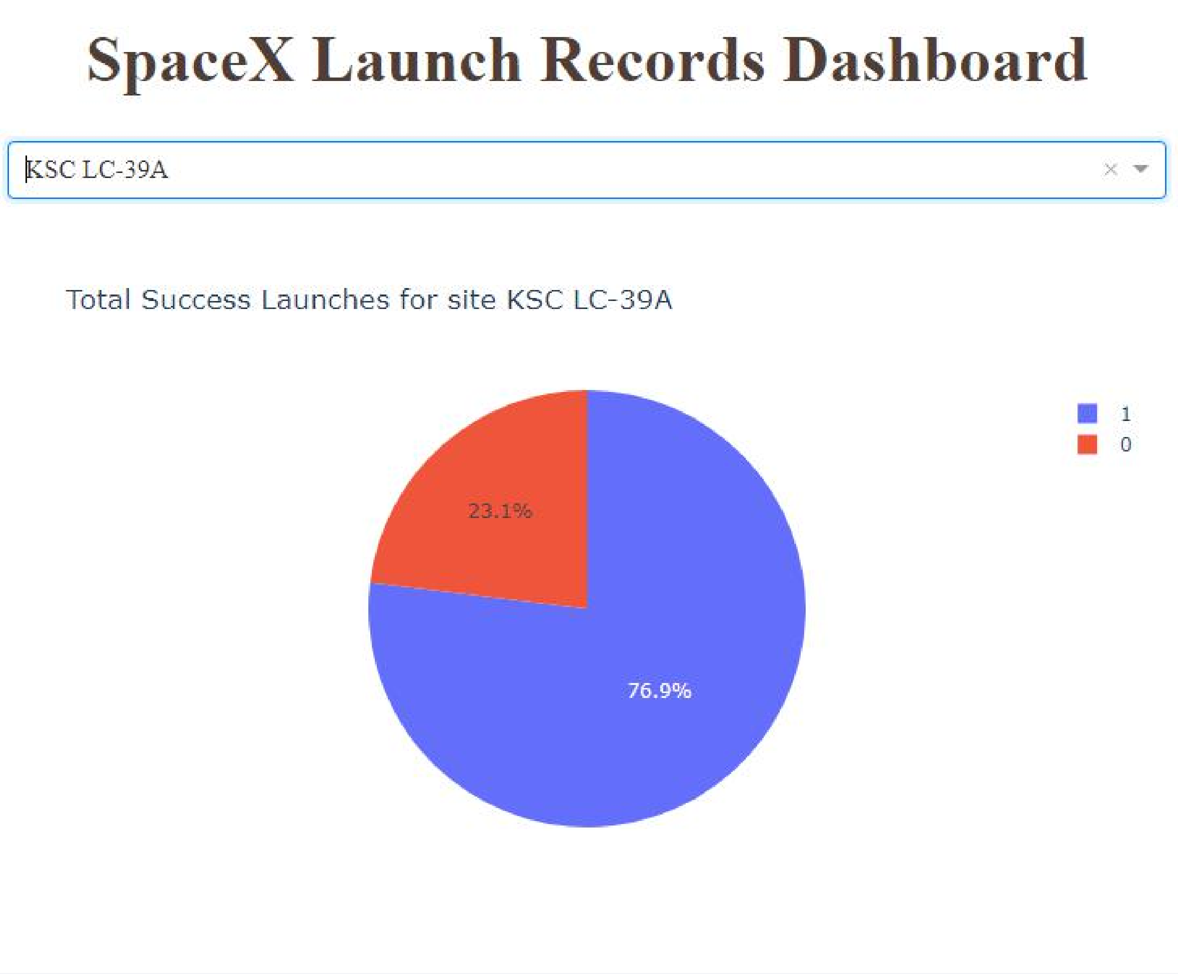
SpaceX Launch Records Dashboard: Launch Success by Site

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



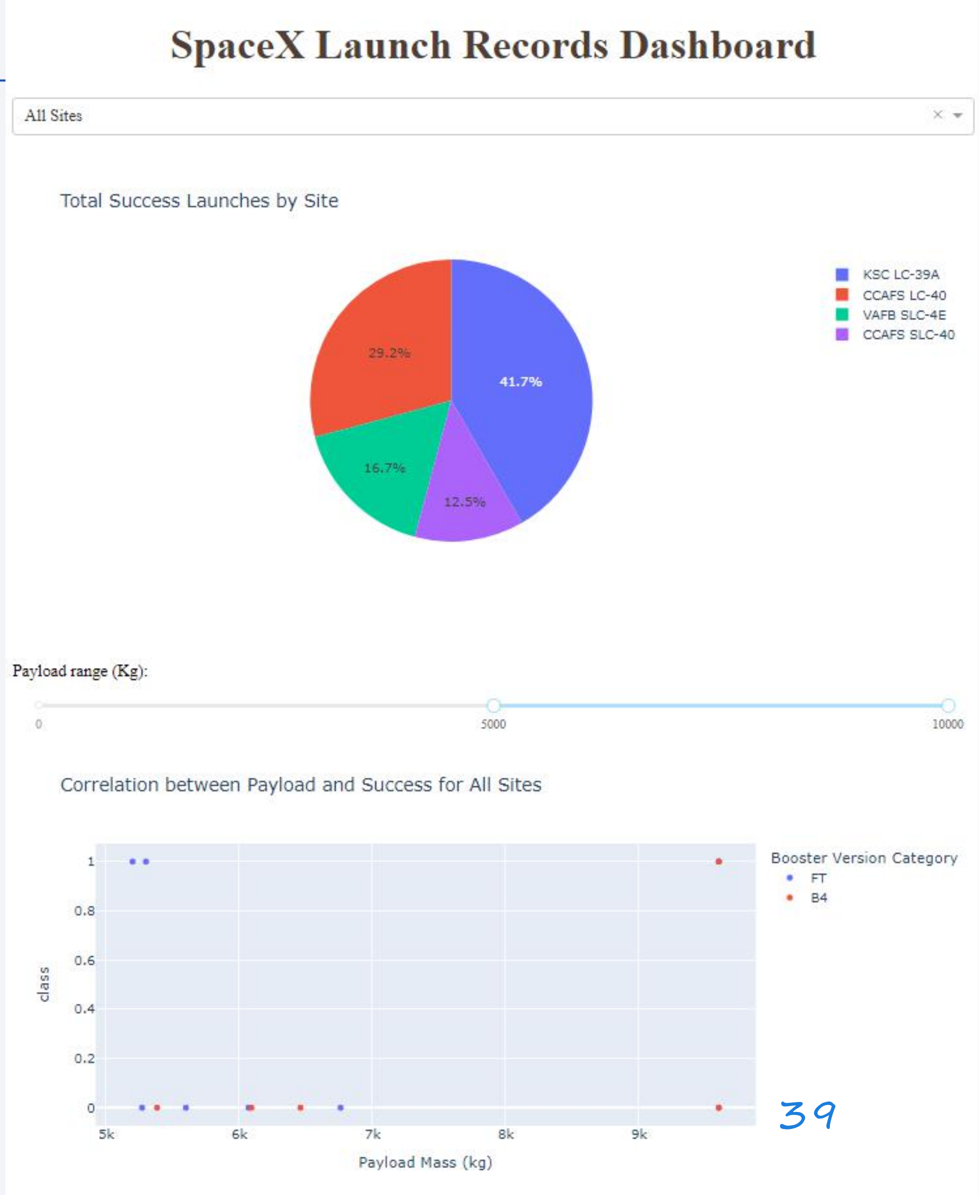
SpaceX Launch Records Dashboard: Launch Success at KSC LC-39A

- 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

- With the slider range of 5,000-10,000kg:
  - Only 2 boosters are represented
    - FT
    - B4
  - Both showed similar success rates
  - B4 Boosters carried the greatest payloads (10,000kg)

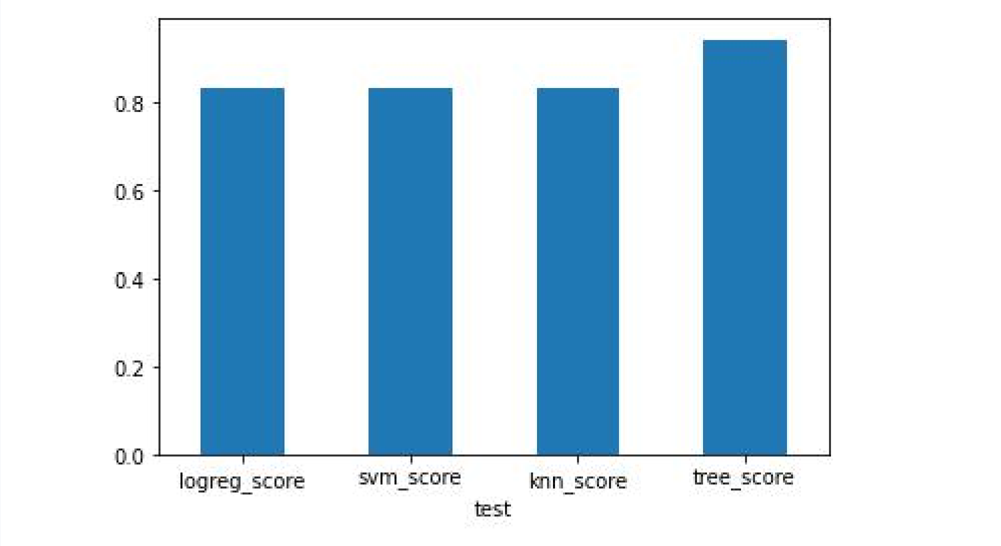


Section 6

Predictive Analysis (Classification)

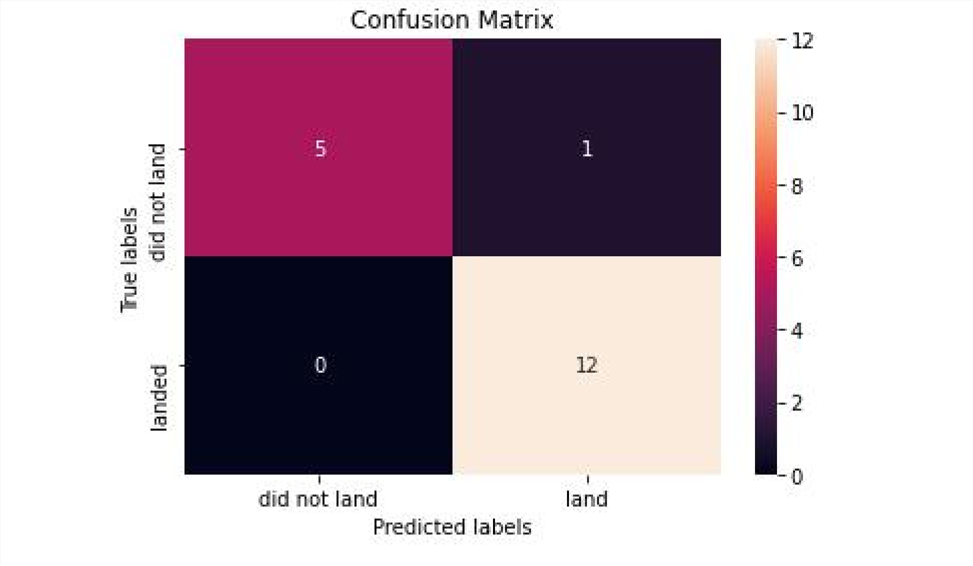
Classification Accuracy

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



Confusion Matrix - Decision Tree Model

- *The decision tree model had the greatest accuracy for the test data*
- *The model incorrectly assigned 1 value:*
  - *Model predicted 1 instance of '**land**' when truely it did not '**land**'*
- *Correctly assigned 17 values of the test set:*
  - *Correctly assigned 5 instances of '**did not land**'*
  - *Correctly assigned 12 instances of '**land**'*



Conclusions

- SpaceX Uses Launch Sites in Florida and California, exclusively
- 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 (94.4% accurate)

Thank you!