IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Matt Wilke
Data Analyst Student
data@data.com
10/09/2022

Outline

Executive Summary – Slide 3

Introduction – Slide 4

Methodology – Slide 5

Results – Slides 6-44

Conclusion - Slide 45

Appendix – Slide 46

Executive Summary

Summary

SpaceX is a private space company who is providing a cheap alternative to space accommodations.

SpaceX Goal

To be able to provide a cheaper and more efficient option for everything space.

SpaceX Claim

SpaceX claims to be able to save a ton of money per launch because they can re-use the first phase of the rocket by reclaiming it/safely returning the first phase back to Earth after the launch.

Our Approach:

Created dashboards to show success/failure at launch sites

Used SpaceX's own data to back up our claims

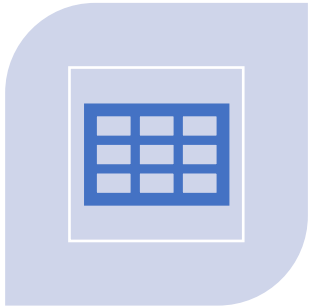
Created a map to show where the launch sites are and capture their results

Used Machine learning algorithms to show success rates up to 84% accuracy

Introduction

Main Goal

THE MAIN GOAL OF THIS PROJECT IS TO DETERMINE IF SPACEX CAN DELIVER ON ITS PROMISE TO MAKE SPACE TRAVEL CHEAPER/MORE AFFORDABLE.



WE USED THE DATA WE GATHERED FROM SPACEX, AS WELL AS GATHERING DATA FROM THE SPACEX WIKI PAGE.

What do we want to answer?

Does SpaceX have a reliable rocket system in place to be able to claim cheaper space travel?

How will we answer this question?

Through exploring the data collected from both SpaceX and the Wiki page.

Using charts and graphs to display our findings.

Using machine learning techniques for the accuracy of the first-stage rocket landing successfully.

Section 1

Methodology

Methodology

Executive Summary

Data collection methodology:

Data was collected from the SpaceX website along with the SpaceX Wiki page.

Perform data wrangling

Looked at where the Rockets were launched from, and where the data was located as to success or failure.

Perform exploratory data analysis (EDA) using visualization and SQL

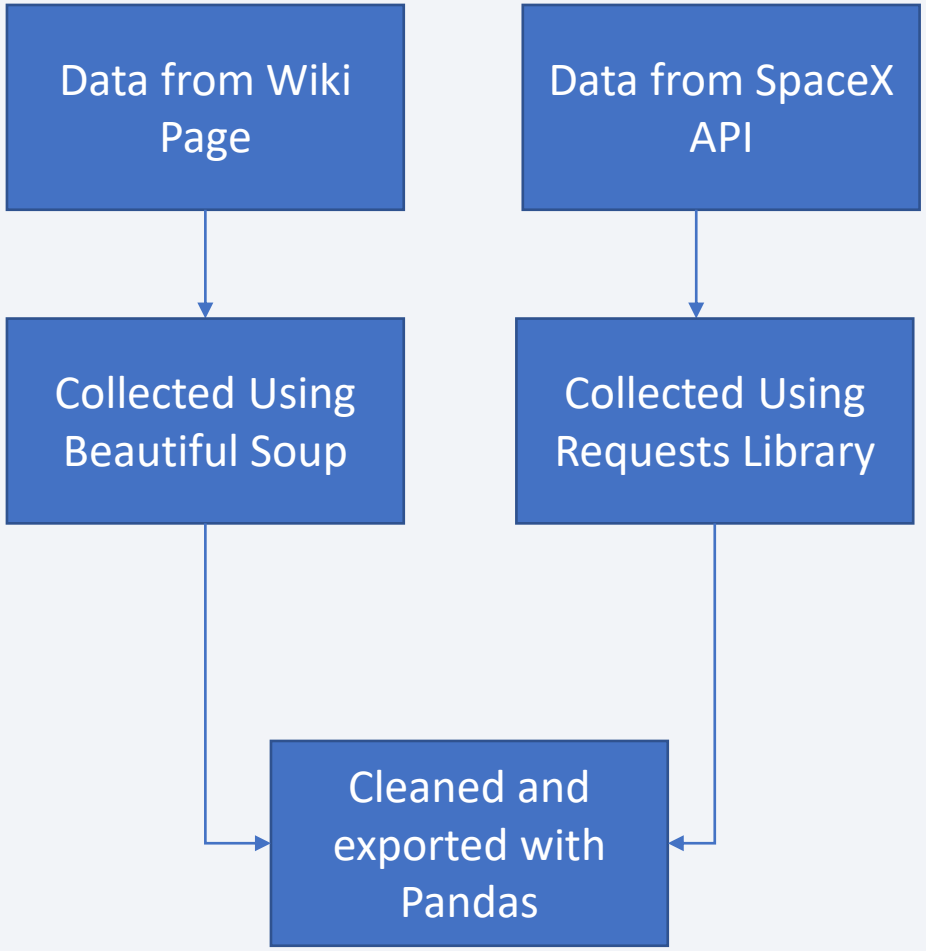
Perform interactive visual analytics using Folium and Plotly Dash

Perform predictive analysis using classification models

We built, tuned, and found the best parameters for the best model to use to predict whether the rocket will land successfully or not with up to 84% accuracy.

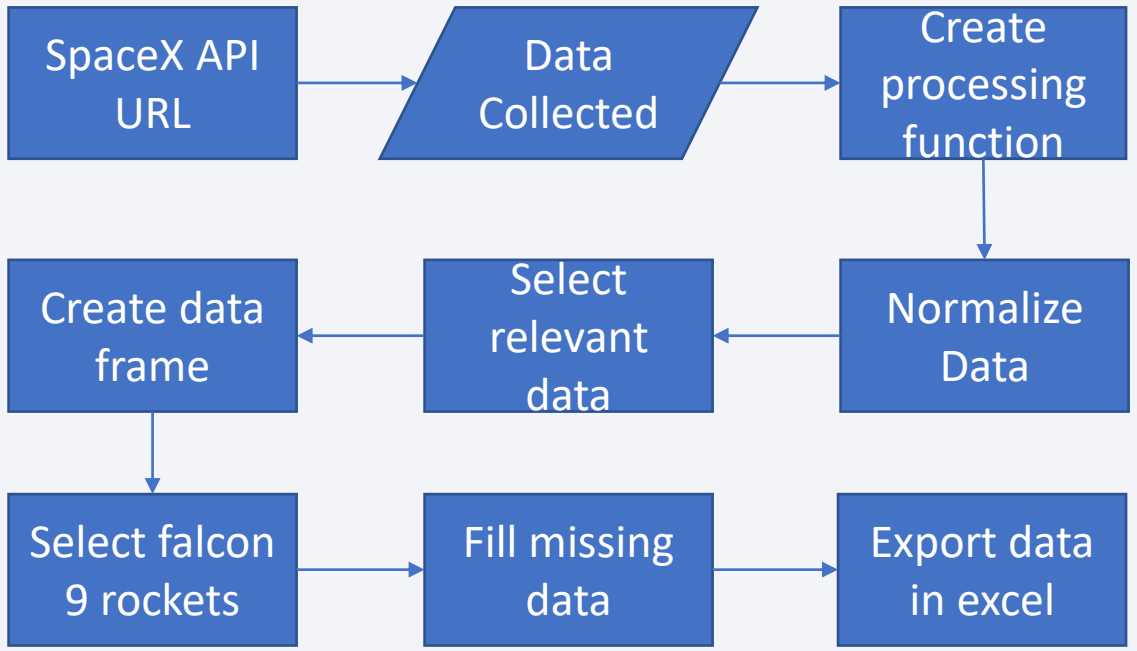
Data Collection

- Data Sets
  - Data sets were collected using the available excel document on SpaceX's API, along with web scraping for additional data on success/failure percentages for each launch.
  - We used the Pandas library and NumPy to collect and analyze the data to get it to be easily interpreted.
  - Since we are only interested in the first phase of the rocket launch, we narrowed the results down to the Falcon 9 rocket which is the Phase 1 rocket.
    - We filled the missing payload information in the dataset collected with the average of the payloads launched.

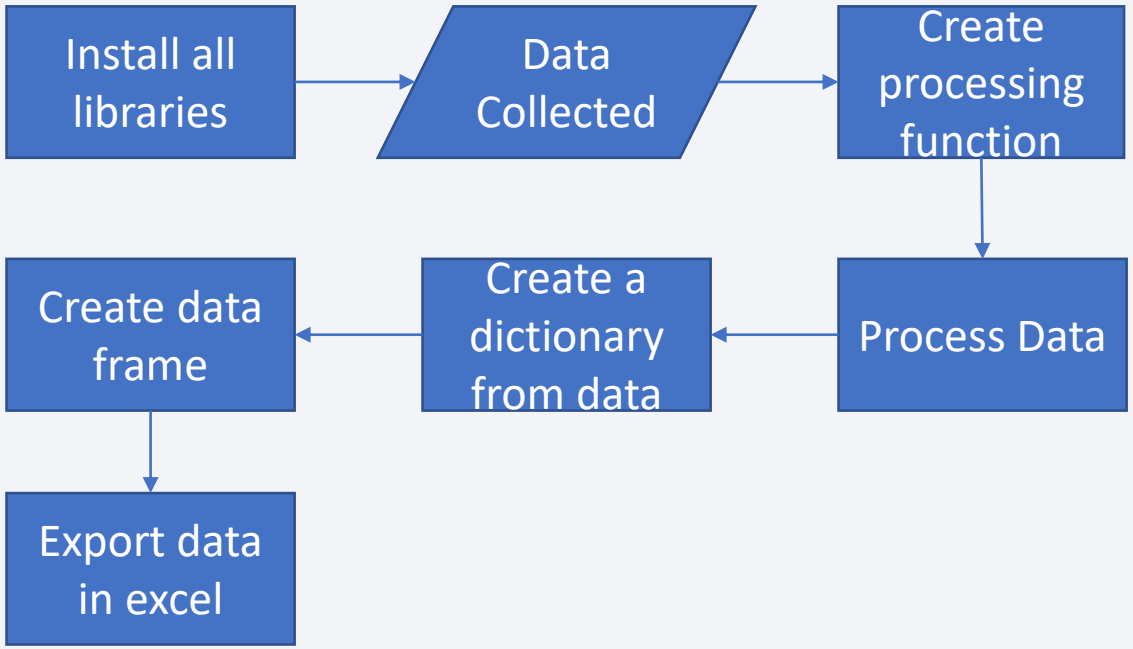


Data Collection – SpaceX API

- Gathered the API call URL.
- Created all get functions to gather: Booster Version, Launch Site, Payload, Cores, Rocket, Success/Failed, Lat/Long, Landing Pad, etc.
- GitHub Page



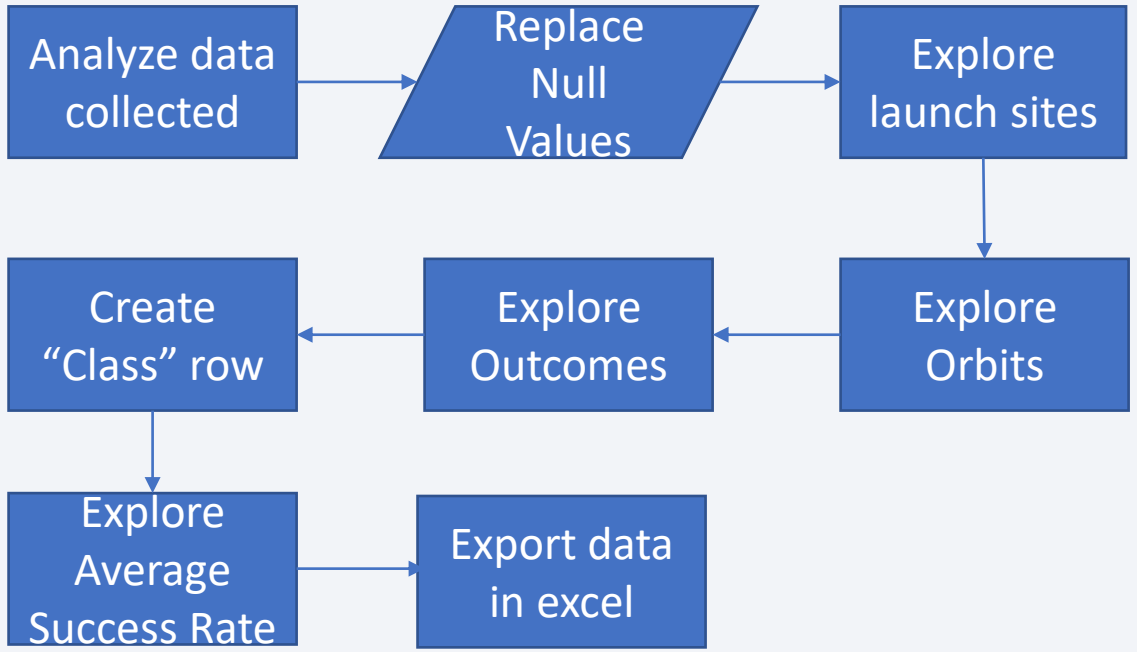
Data Collection - Scraping



- We are using Beautiful Soup to scrape our data. We gather our data from the SpaceX wiki page. We are looking for the Falcon 9 launches. We clean all our data and create a pandas data frame from the data and finally export the data
- GitHub Page

Data Wrangling

- The data we collected was analyzed using pandas to determine the most relevant aspects of the data.
- We removed all the null values from the payload section and filled them with the average payload from the rest of the values.
- We looked at all the launches that occurred at each of the launch sites.
- We explored the orbits the rocket sent the payload to.
- We explored the success rate of the rocket and created a class row to tell whether it was a successful or failed attempt.
- GitHub Page



EDA with Data Visualization

- Charts Created:
  - **Payload Mass VS. Flight Number**: to show which ones were successful/fail. We did this to see if there was a comparison between the payload mass and how many flights.
  - **Flight Number VS. Launch Site**: to show if success/failure was based on how many flights were taken.
  - **Success/Failure VS. Orbit**: to show if the orbit has anything to do with the success of the mission.
  - **Success Rate by Year**: to show what the success rate looks like from 2010-2020.
- GitHub Page

EDA with SQL

- SQL Queries:
  - Show the different launch sites
  - First successful ground pad landing
  - Successful drone landing with payload mass between 4000 and 6000
  - Number of successful and failed missions
  - Failed drone ship attempts in 2015
  - Landing outcome between 06-04-2010 and 03-20-2017 in descending order
- GitHub Page

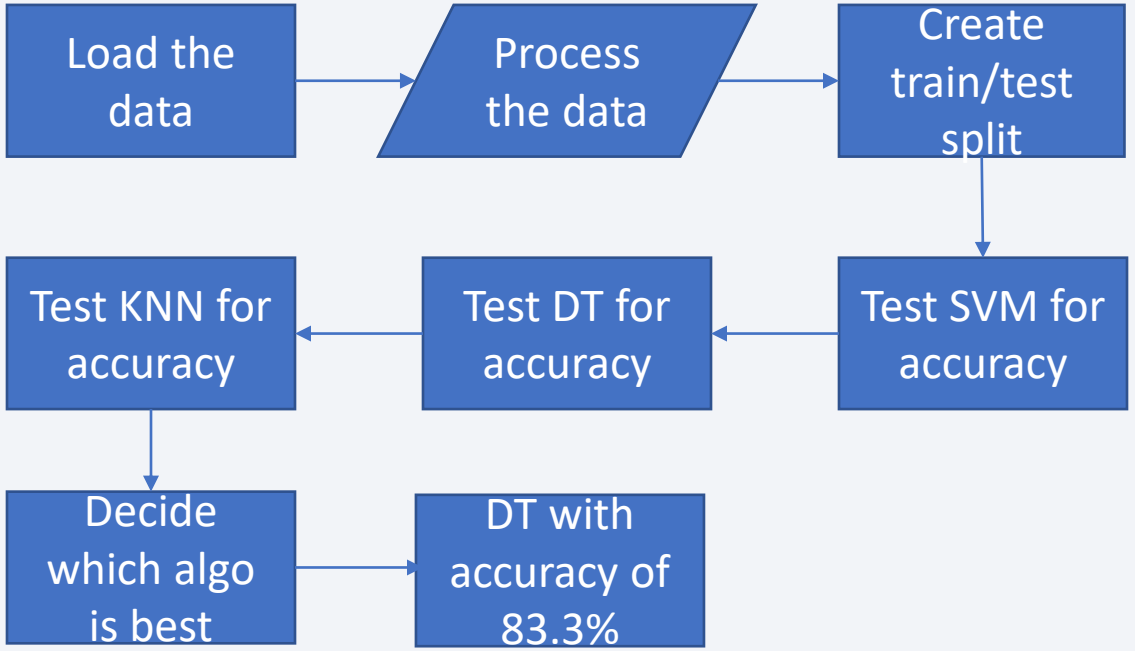
Build an Interactive Map with Folium

- Folium Objects Used:
  - Marker
    - Used to mark where the Launch Sites are located
  - Circle
    - Used to plot the highlighting circle for the markers
  - Marker Cluster
    - Clusters the Markers to show how many Markers are in that location
  - PolyLine
    - Plots a line on the map to show the distance from the launch site to the coastline and to the closest airport
- GitHub Page

Build a Dashboard with Plotly Dash

- Created a drop-down list to select the Launch Site you are interested in; the default selection is all of them
- When you select a Launch Site it will show a pie graph with the percentage of successful/failed missions
- There is also a graph that compares the success/fail rate against the Payload Mass
    - There is a range bar you can change (0-10000 kg) that shows the booster version used as the color coding.
- GitHub Page

Predictive Analysis (Classification)



We created the best predictor by using multiple different algorithms: Support Vector Machine, Decision Tree, and KNN.

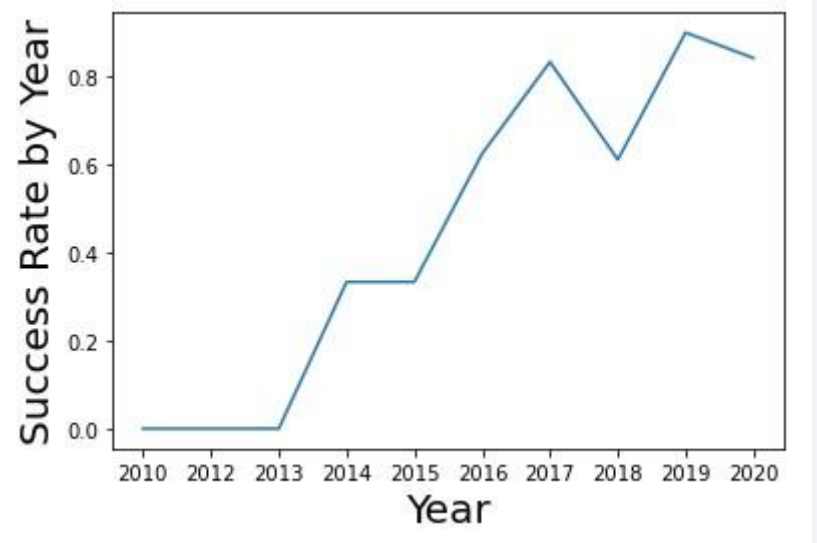
We used Grid Search to find the best parameters to use for the highest accuracy

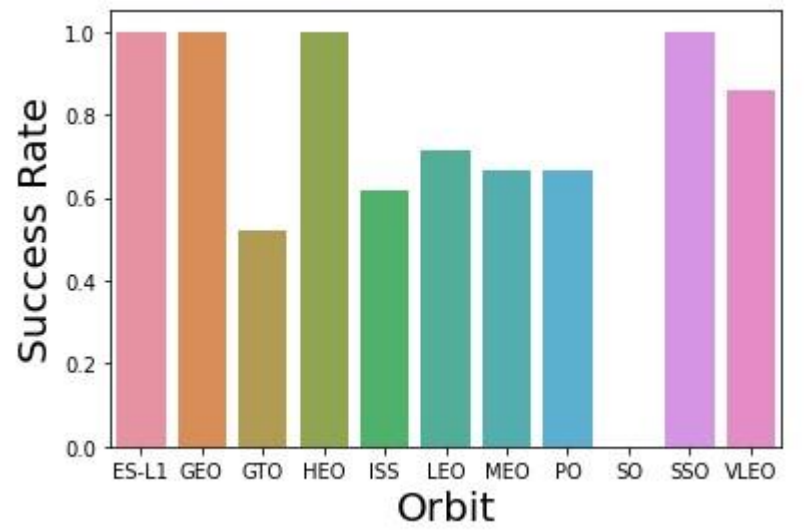
We found the best algorithm to use was the Decision Tree algorithm with an accuracy of 83.3%

GitHub Page

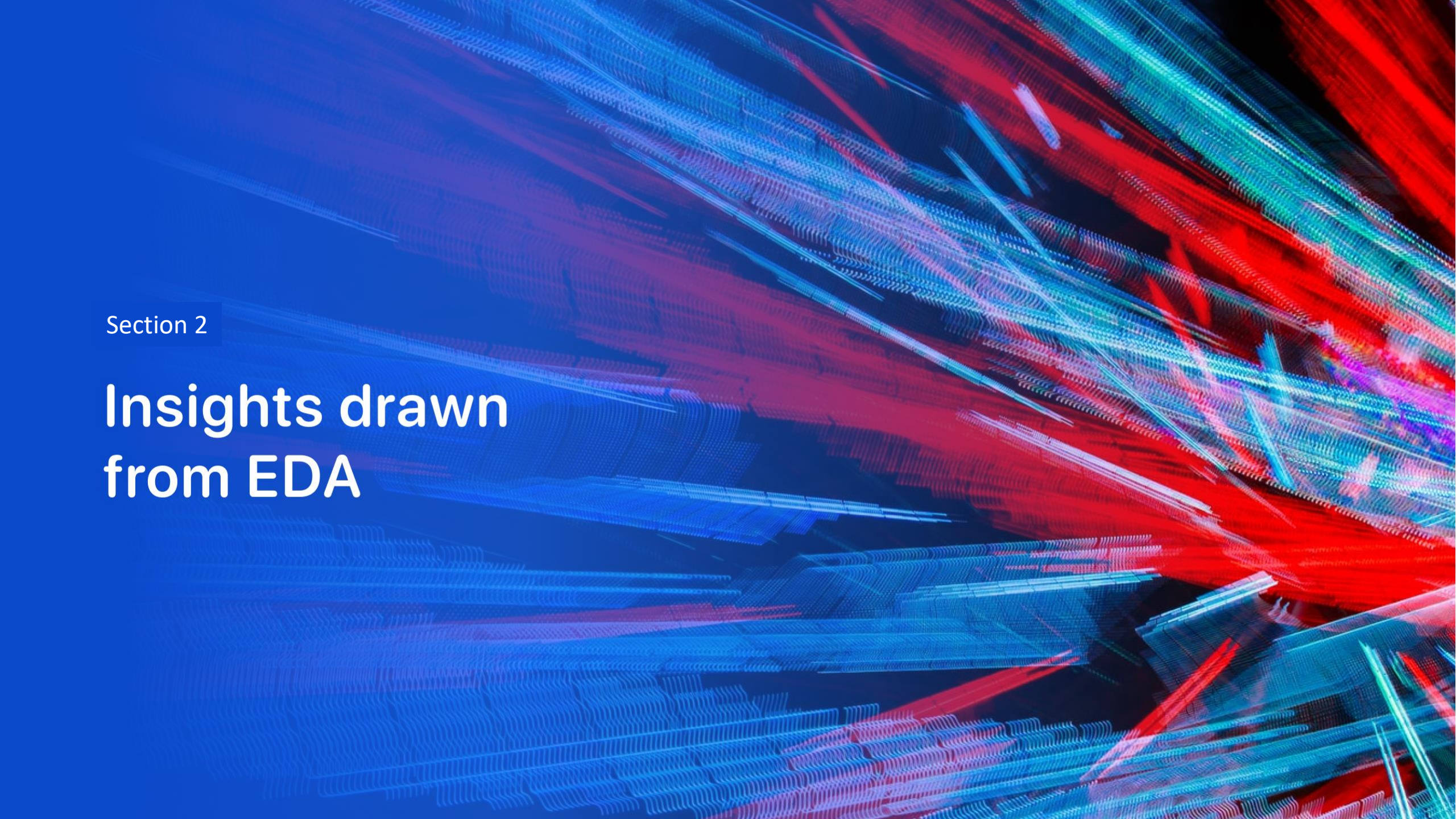
Results

- Our exploratory data results concluded that SpaceX's launch success rate had been improving year over year as shown by the graph to the right.



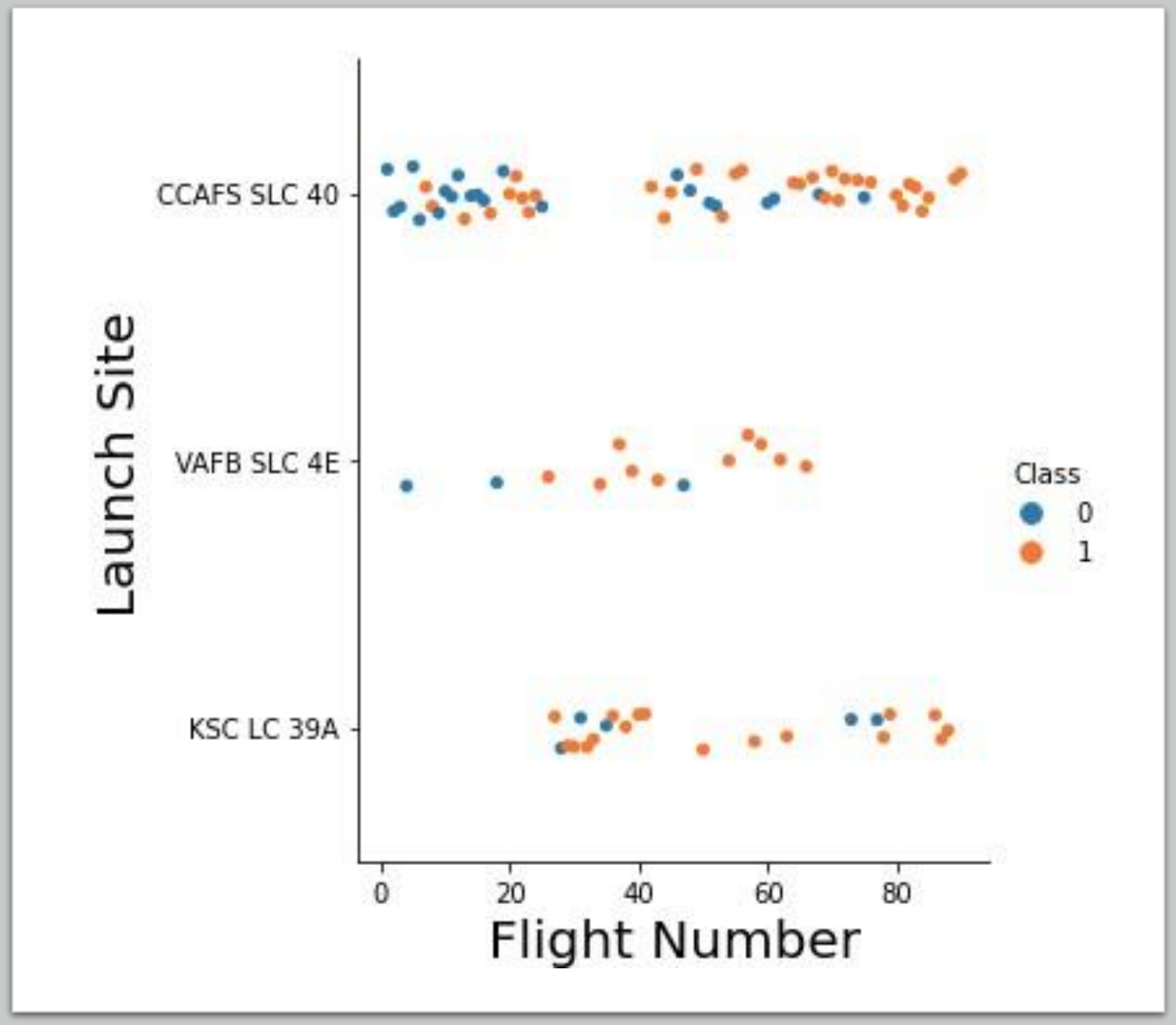


- They have a very high success rate across different orbits as well.
- All of this concludes that SpaceX has a very high chance of a successful mission decreasing the cost of space travel


Section 2
Insights drawn from EDA

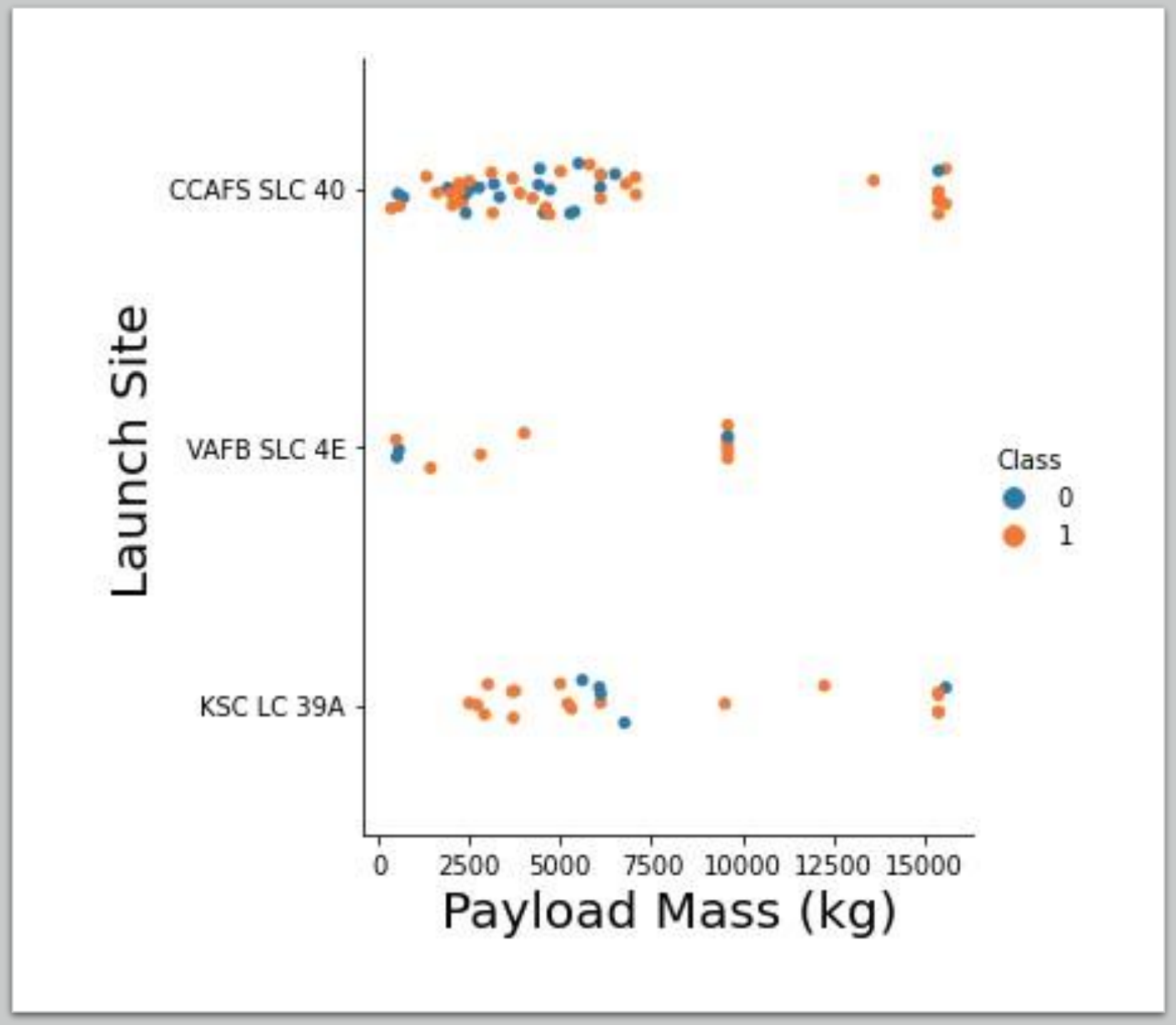
Flight Number vs. Launch Site

- This plot shows the number of flights on the X axis and the launch site on the Y axis.
- The color coding represents success (1) and failed (0) mission outcomes.
- You can see that no missions were attempted at KSC launch site until there were at least 20 launches with that rocket.
- In general the more launches the rockets have, the more success rate of the mission.



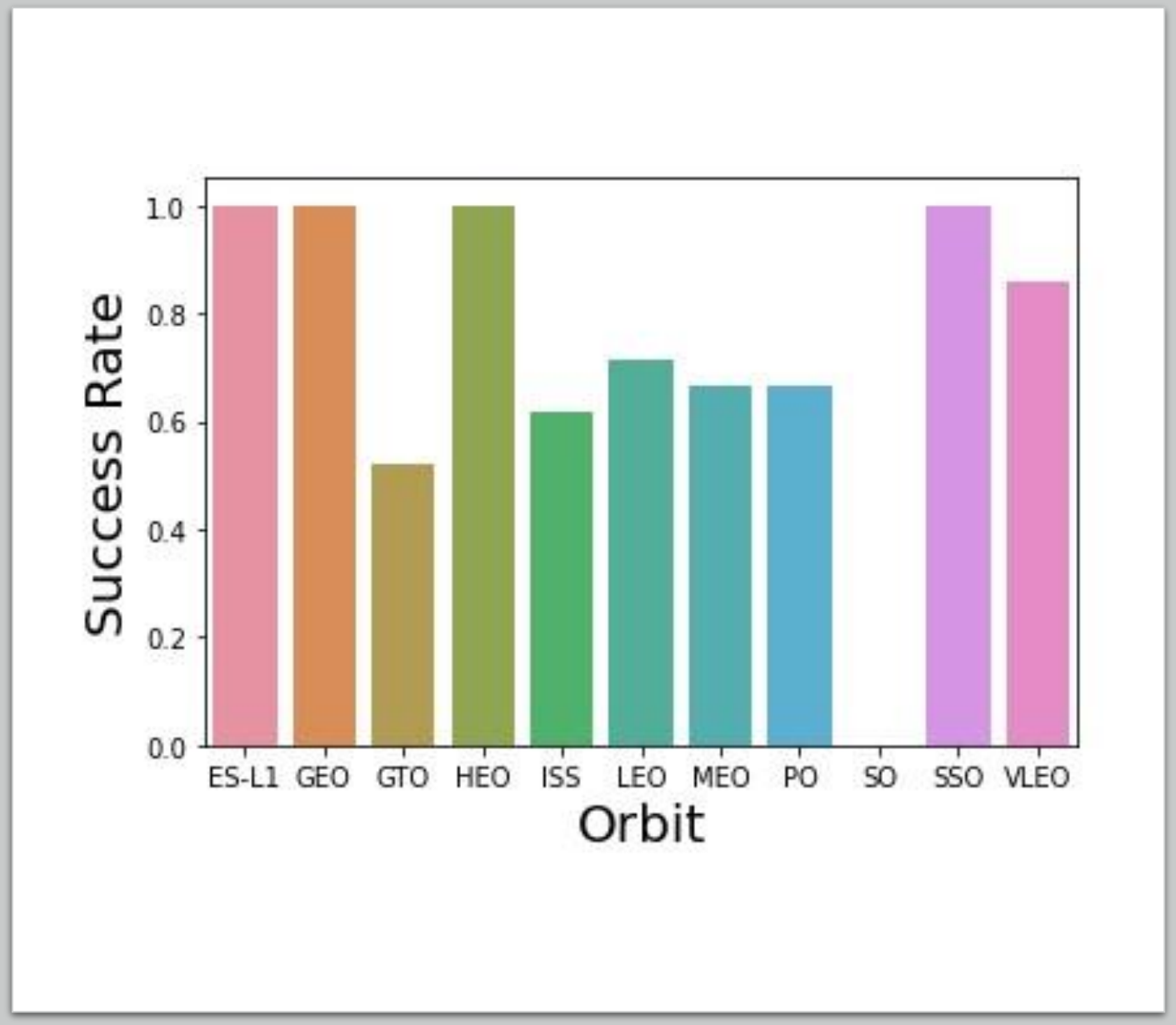
Payload vs. Launch Site

- Here you can see the payload at each launch site.
- CCAFS launch site has a big mix of successful/failed missions with lower payloads. With higher payloads, they are more successful.
- KSC has a wide range of payloads and does pretty good with low payload and high.
- VAFB either launches low payloads or 10k payloads only.



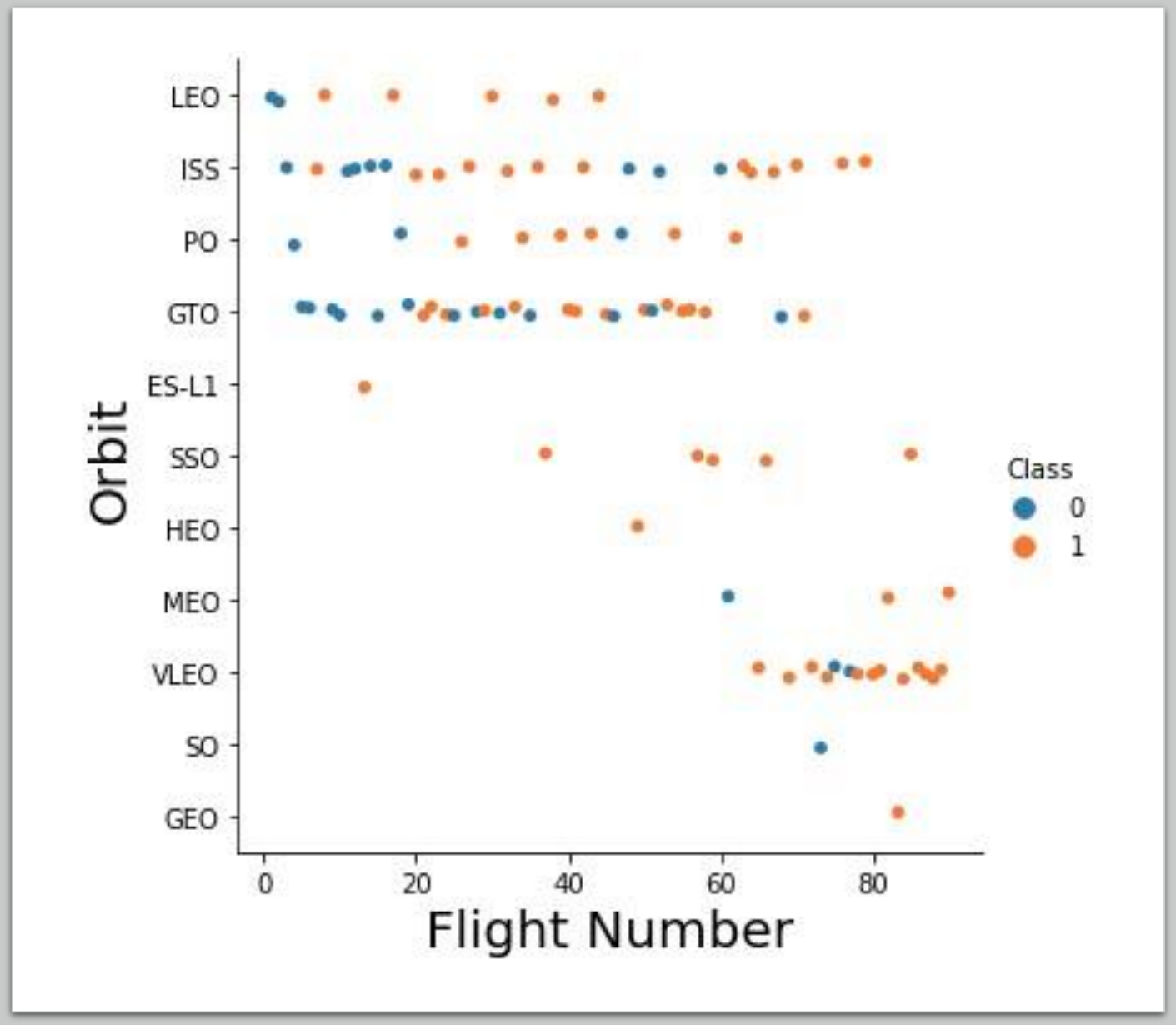
Success Rate vs. Orbit Type

- Here we can see where the rocket's end destination was going.
- L1, GEO, HEO, and SSO all have 100% success rates!
- SO has the lowest at a 0% success rate.



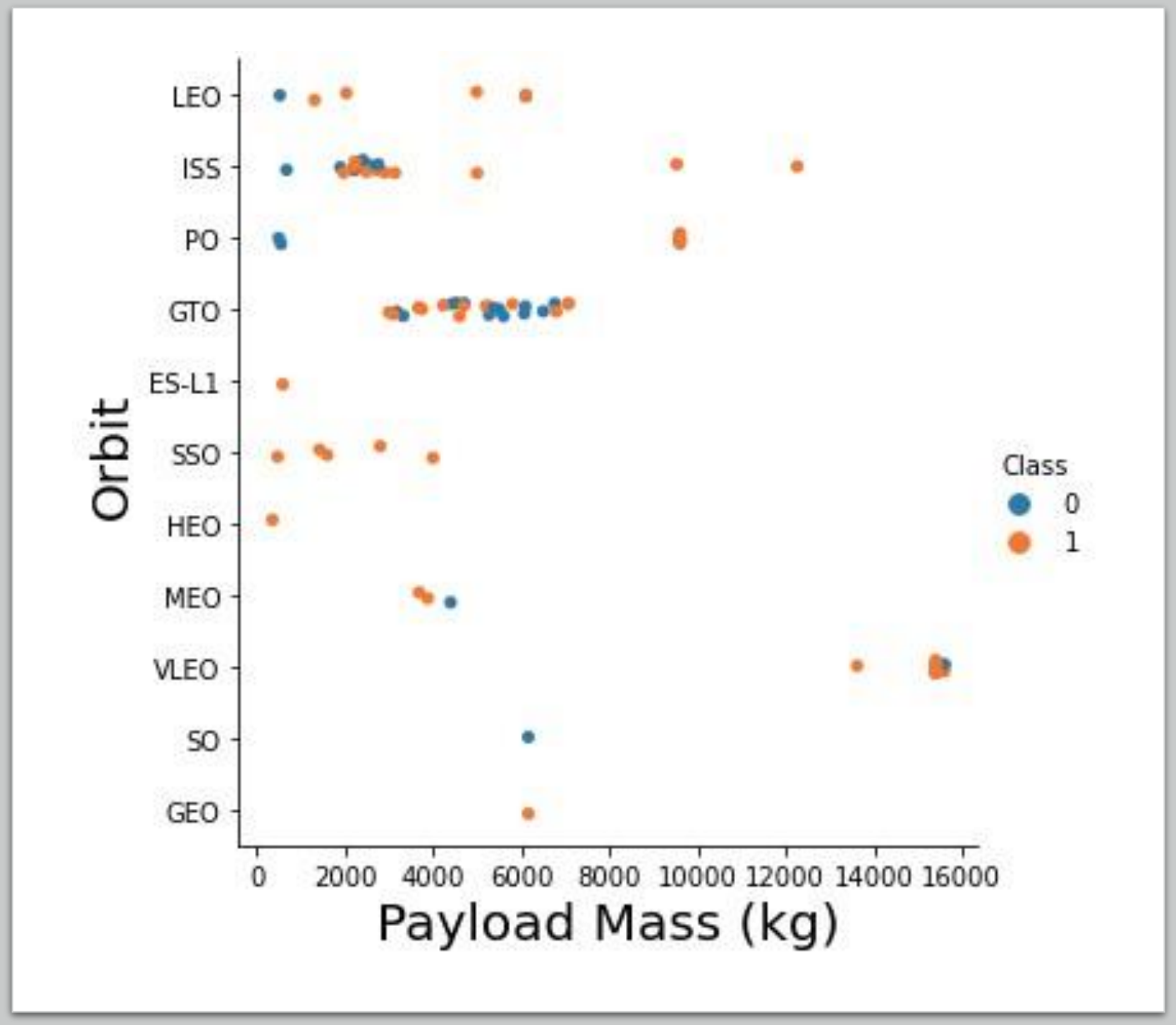
Flight Number vs. Orbit Type

- This plot shows the success rate for the Orbit vs Flight Number.
- This chart outlines what the previous graph was representing with the success rates with orbits.
- You can see with most orbits the higher the flight number the higher the success rate with orbits besides GTO.



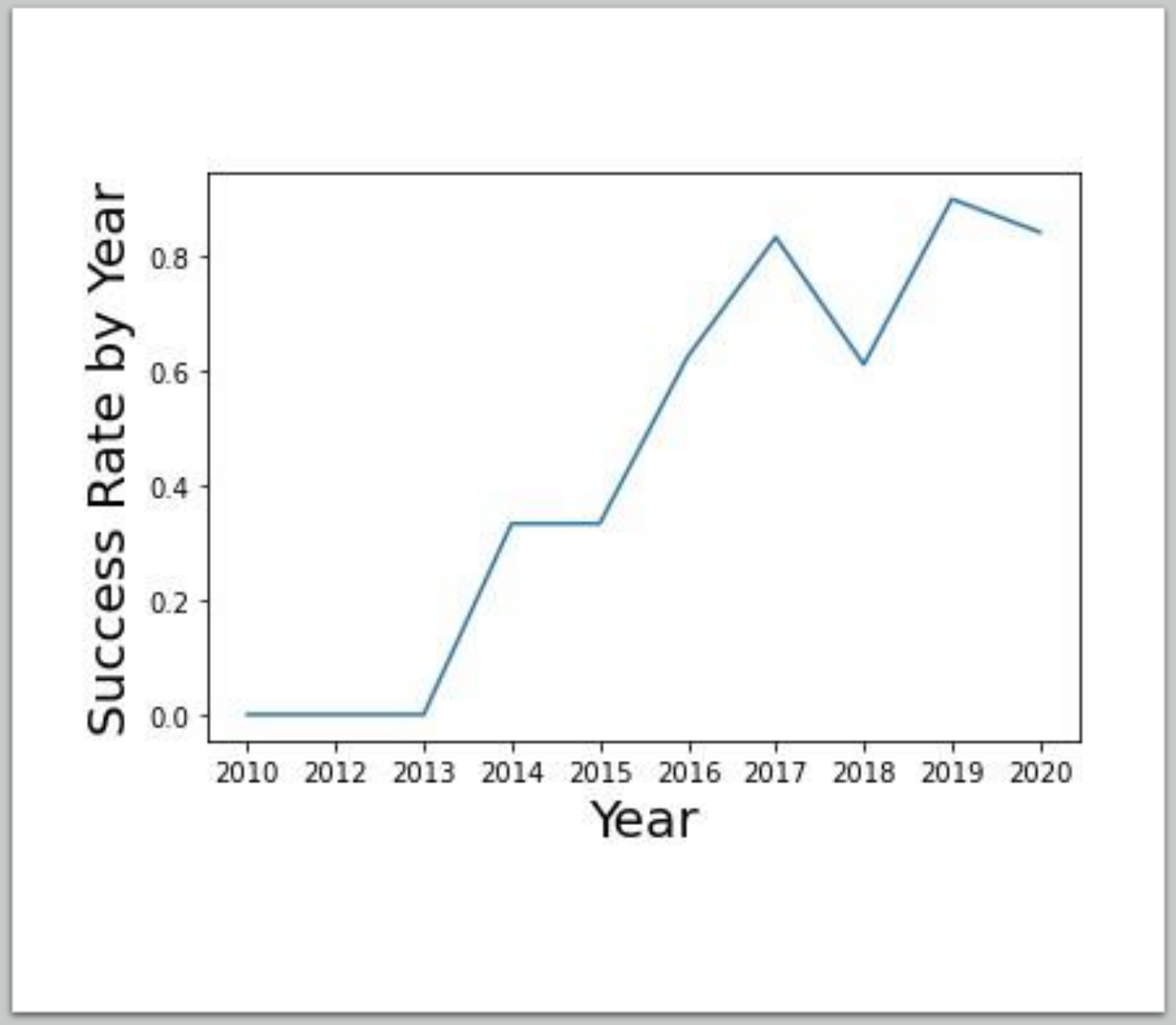
Payload vs. Orbit Type

- This chart shows the weight of the payload that was sent to each orbit.
- Again, GTO has mixed results for success/failed outcomes and only launches payloads between 3000 and 8000.
- Payload amounts greater than 6000 are not sent to GEO, SO, MEO, HEO, SEO, L1 orbits.
- VLEO orbit only has 12000+ payloads.



Launch Success Yearly Trend

- Here we can see the success rate of the mission year over year.
- The lowest success rate is 0% with their success rate rising to 38% in 2014.
- They continued to improve their mission success rate all the way to 80% in 2020.



```
%sql select UNIQUE(launch_site) FROM SPACEX
```

```
 * ibm_db_sa://dsh69876:***@b1bc1829-6f45-4cd
main.cloud:32304/bludb
Done.
```

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

All Launch Site Names

- Here we query the SPACEX database to display the list of unique launch sites in the database.
- There is a typo in one of the launch sites which is why it shows 4 launch sites. In our ED we fix this.

```
%sql select * from SPACEX \
WHERE launch_site LIKE 'CCA%'\
LIMIT 5;
```

```
 * ibm_db_sa://dsh69876:***@b1bc1829-6f45-4cd4-bef4-10cf081900bf.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32304/bludb
Done.
```

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (p
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 (p
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	N
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	N
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	N

Launch Site Names Begin with 'CCA'

- In this query we wanted to select the first 5 records with the launch sites that start with CCA.
- Here you can see all these Orbits go to LEO which has about a 78% success rate.
- All the booster versions for these launches are using the F9 v1.0.

```
%sql select sum(payload_mass__kg_) from SPACEX \
WHERE customer = 'NASA (CRS)';
```

```
 * ibm_db_sa://dsh69876:***@b1bc1829-6f45-4cd4-bef4
main.cloud:32304/bludb
Done.
```

1
45596

Total Payload Mass

- Here we queried to find the total payload that was shipped by NASA.
- The total payload mass is 45,596 kg.

```
%sql select avg(payload_mass__kg_) from SPACEX \
WHERE booster_version = 'F9 v1.1';
```

```
 * ibm_db_sa://dsh69876:***@b1bc1829-6f45-4cd4-bef
main.cloud:32304/bludb
Done.
```

1
2928

Average Payload Mass by F9 v1.1

- We wanted to find the average payload sent to orbit by the F9 v1.1 rocket.
- The average mass is 2,928 kg.

```
%sql select date from SPACEX \
WHERE landing__outcome = 'Success (ground pad)' \
LIMIT 1;
```

```
 * ibm_db_sa://dsh69876:***@b1bc1829-6f45-4cd4-bef
main.cloud:32304/bludb
Done.
```

DATE
2015-12-22

First Successful Ground Landing Date

- Here we queried the database to find the first successful ground lading by SpaceX.
- The first successful ground landing is 12-22-2015.

```
%sql select booster_version, payload_mass__kg_ from SPACEX \
WHERE landing__outcome = 'Success (drone ship)' AND payload_mass__kg_ > 4000 AND payload_mass__kg_ < 60
```

```
 * ibm_db_sa://dsh69876:***@b1bc1829-6f45-4cd4-bef4-10cf081900bf.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32304/bludb
Done.
```

booster_version	payload_mass__kg_
F9 FT B1022	4696
F9 FT B1026	4600
F9 FT B1021.2	5300
F9 FT B1031.2	5200

Successful Drone Ship Landing with Payload between 4000 and 6000

- Here we queried to find a booster version for the successful drone ship landing missions with payloads between 4000 and 6000.
- The 4 booster versions are: F9 FT B1022, F9 FT B1026, F9 FT B1021.2, and F9 FT B1031.2
- With payloads 4696 kg, 4600 kg, 5300 kg, and 5200 kg, respectively.

```
%sql select mission_outcome, count(mission_outcome) as outcome from SPACEX \
group by mission_outcome;
```

```
 * ibm_db_sa://dsh69876:***@b1bc1829-6f45-4cd4-bef4-10cf081900bf.c1ogj3sd0tgt
main.cloud:32304/bludb
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

- Here we wanted to see the successful and failed mission outcomes in the database.
- There is 1 Failure in flight, 99 Successful, and 1 Success (payload status unclear)

```
%sql select booster_version as booster_max from SPACEX \
where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEX);
```

```
 * ibm_db_sa://dsh69876:***@b1bc1829-6f45-4cd4-bef4-10cf081900bf.c1ogj
main.cloud:32304/bludb
Done.
```

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

Boosters Carried Maximum Payload

- In this query we wanted to see the booster versions that carried the max payload.
- There are 12 booster versions that have carried the max payload.

```
%sql select booster_version, launch_site, landing__outcome from SPACEX \
where landing__outcome = 'Failure (drone ship)' and YEAR(date) = 2015;
```

```
 * ibm_db_sa://dsh69876:***@b1bc1829-6f45-4cd4-bef4-10cf081900bf.c1ogj3sd0t
main.cloud:32304/bludb
Done.
```

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

2015 Launch Records

- These are the failed launches in 2015.
- There are 2 failed launches.
- Both failed launches were launched from the CCAFS LC-40 launch site.

```
%sql select landing__outcome, date from SPACEX \
where date between '2010-06-04' and '2017-03-20' \
order by date desc;
```

```
 * ibm_db_sa://dsh69876:***@b1bc1829-6f45-4cd4-bef4-1
main.cloud:32304/bludb
Done.
```

landing__outcome	DATE
No attempt	2017-03-16
Success (ground pad)	2017-03-06
Success (ground pad)	2017-02-19
Success (drone ship)	2017-01-14
Success (ground pad)	2017-01-05
Success (drone ship)	2016-08-14
Success (drone ship)	2016-08-04
Success (ground pad)	2016-07-18
Failure (drone ship)	2016-06-15
Success (drone ship)	2016-06-05
Success (drone ship)	2016-05-27
Failure (drone ship)	2016-04-03
Failure (drone ship)	2016-01-17
Success (ground pad)	2015-12-22
Controlled (ocean)	2015-11-02
Failure (drone ship)	2015-10-01
Precluded (drone ship)	2015-06-28
No attempt	2015-04-27
Failure (drone ship)	2015-04-14
No attempt	2015-02-03
Uncontrolled (ocean)	2014-09-21
Controlled (ocean)	2014-07-14
No attempt	2014-07-09
No attempt	2014-05-08
Controlled (ocean)	2014-04-18
No attempt	2014-01-06
No attempt	2013-12-03
Uncontrolled (ocean)	2013-09-29
No attempt	2013-03-01
No attempt	2012-10-08
No attempt	2012-05-22
Failure (parachute)	2010-12-08

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

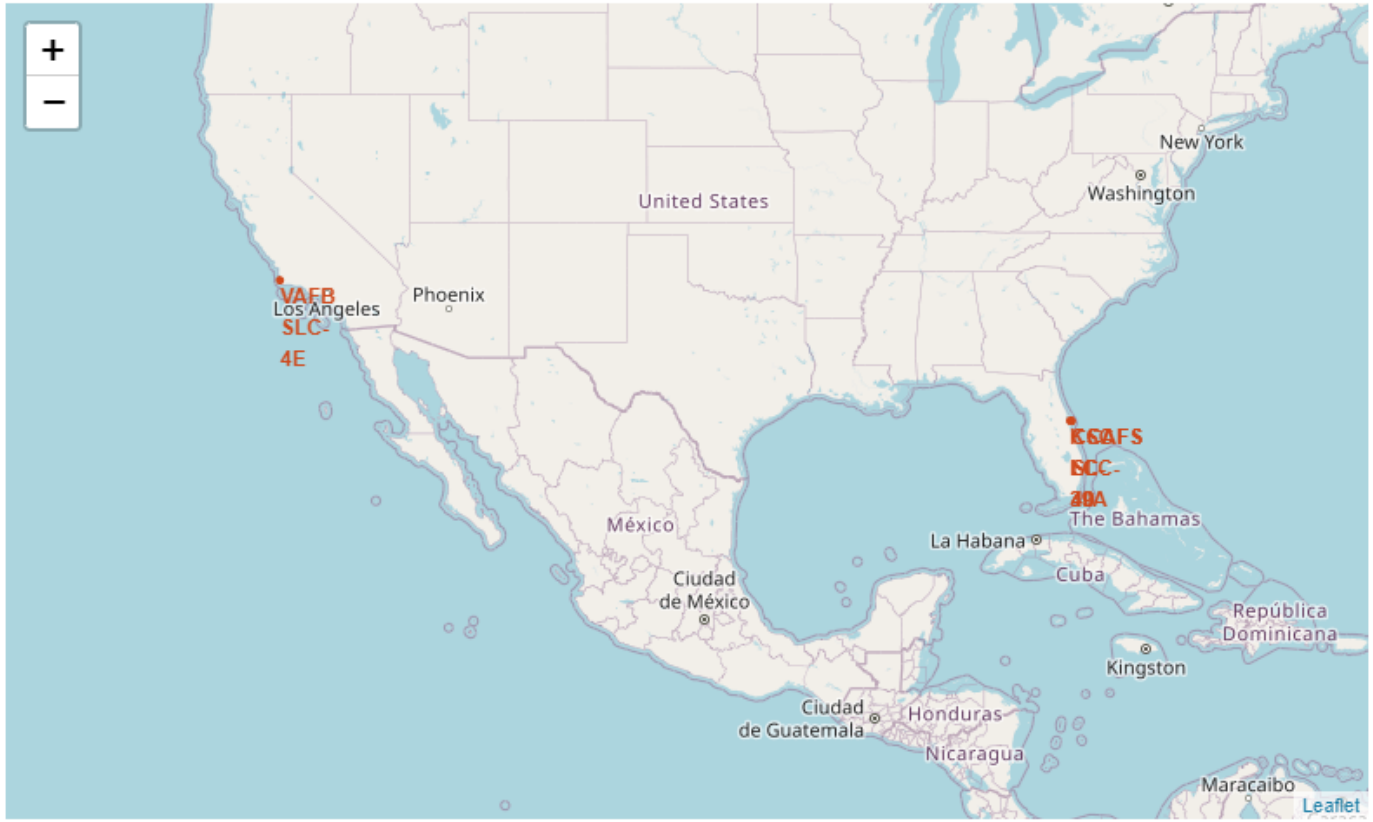
- These are the landing outcomes listed in descending order from 06-04-2010 to 03-20-2017
- There are a total of 32 launches on record. Some launches were not attempted, some are successful, and some are uncontrolled, controlled, and failures.

Section 3

Launch Sites Proximities Analysis

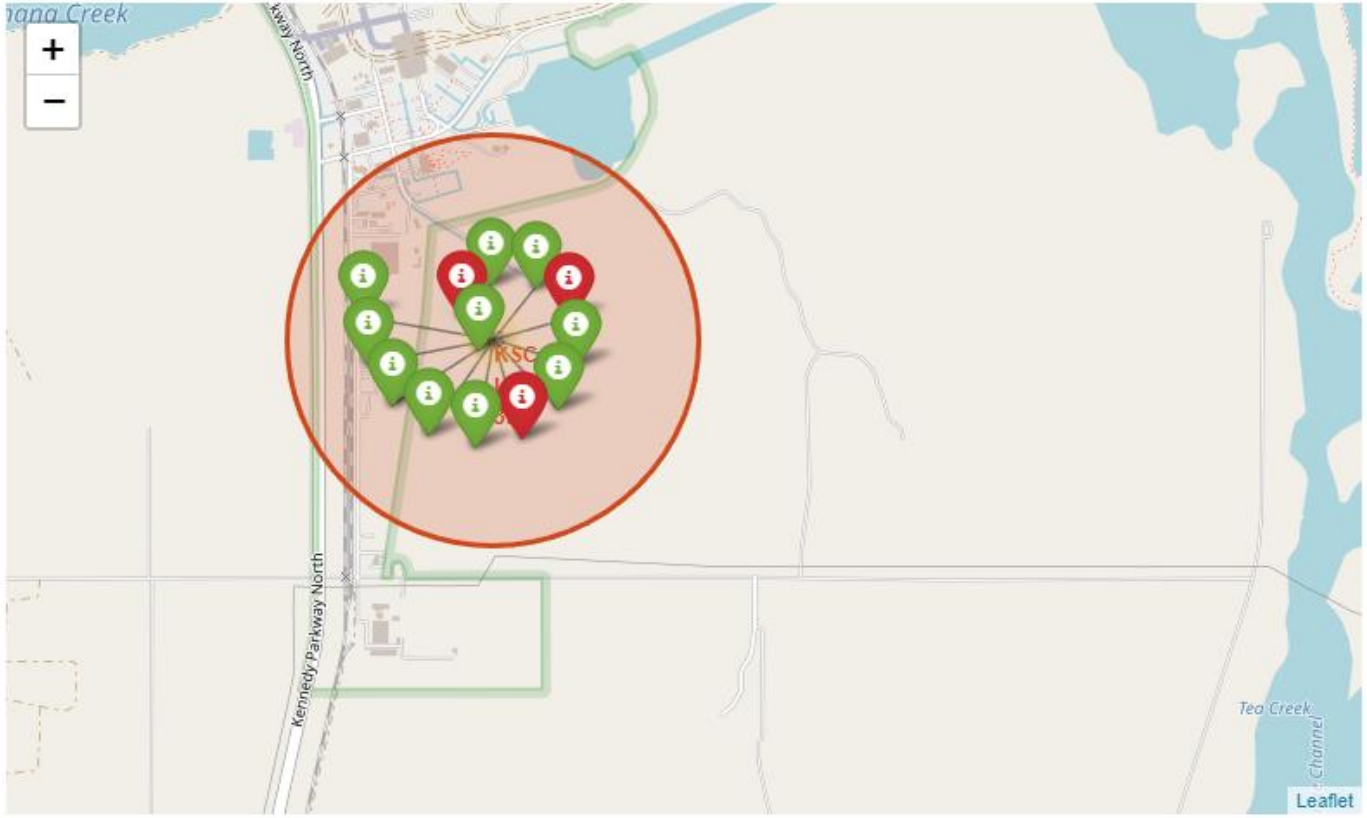
All SpaceX's Launch Sites

- This is a screenshot of the Folium map with all the Launch Sites marked.
- The launch sites are in California (1) and in Florida (2).



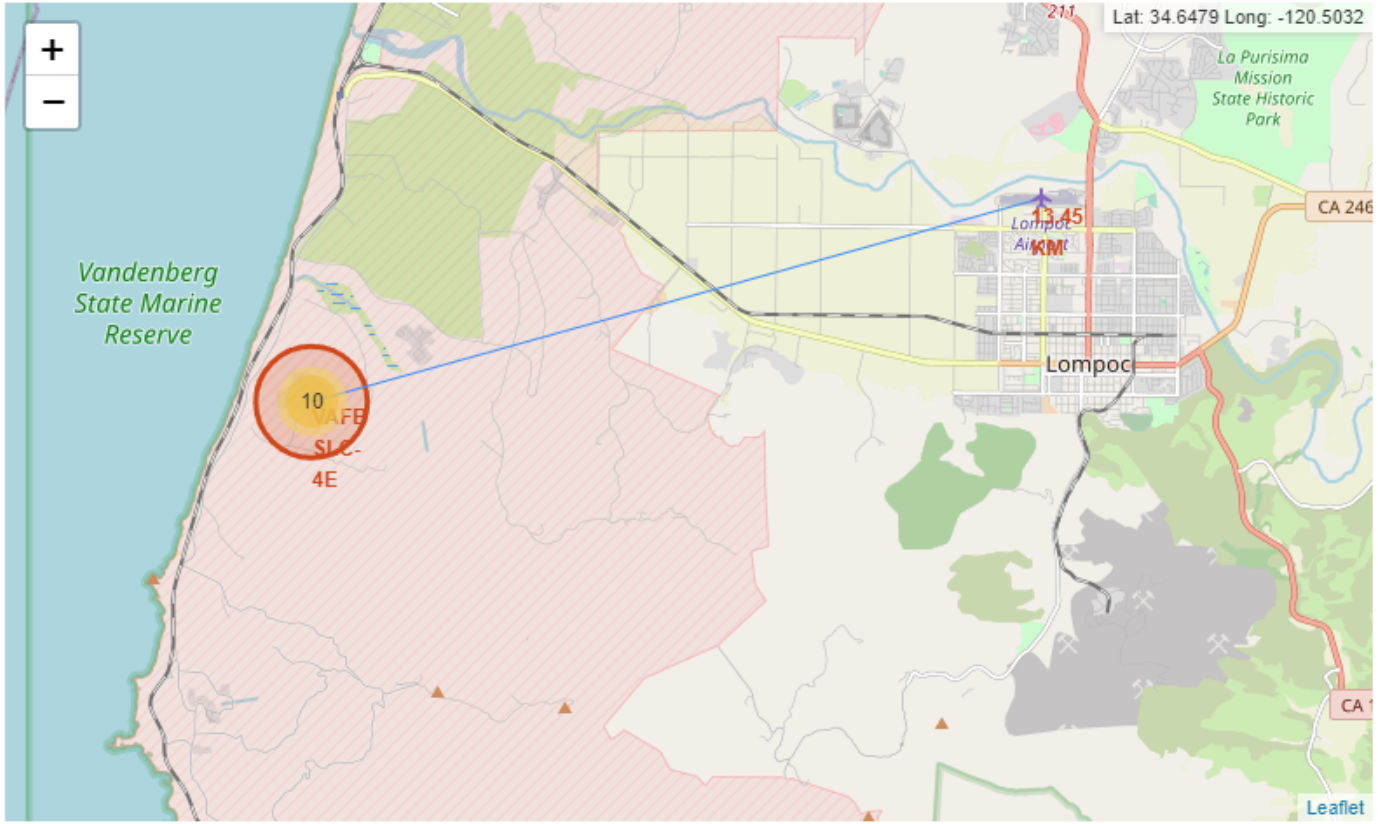
Success/Failed Markers

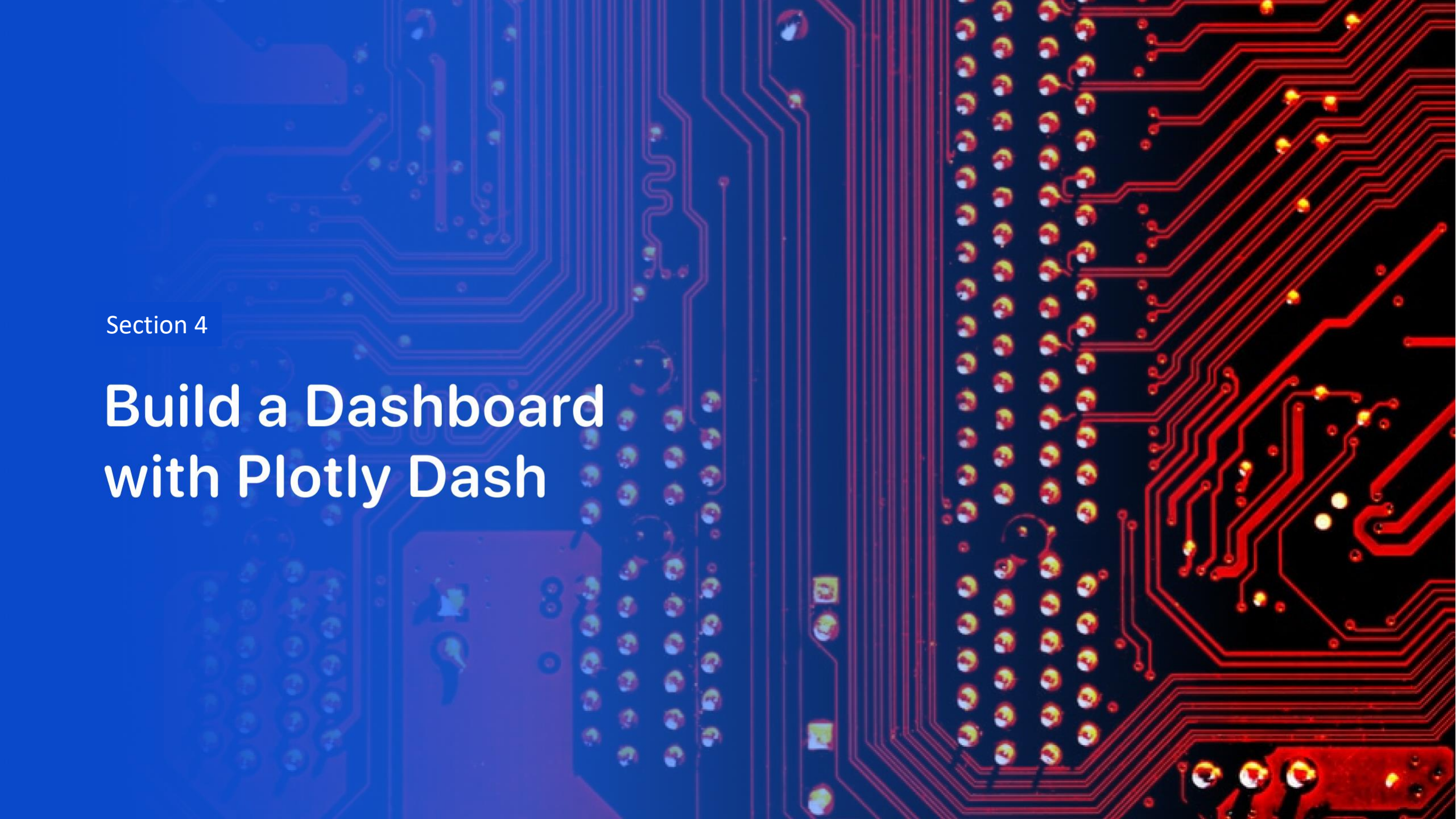
- This map shows the cluster launch sites with each launch being color coded.
- Green represents a successful mission.
- Red represents a failed mission.

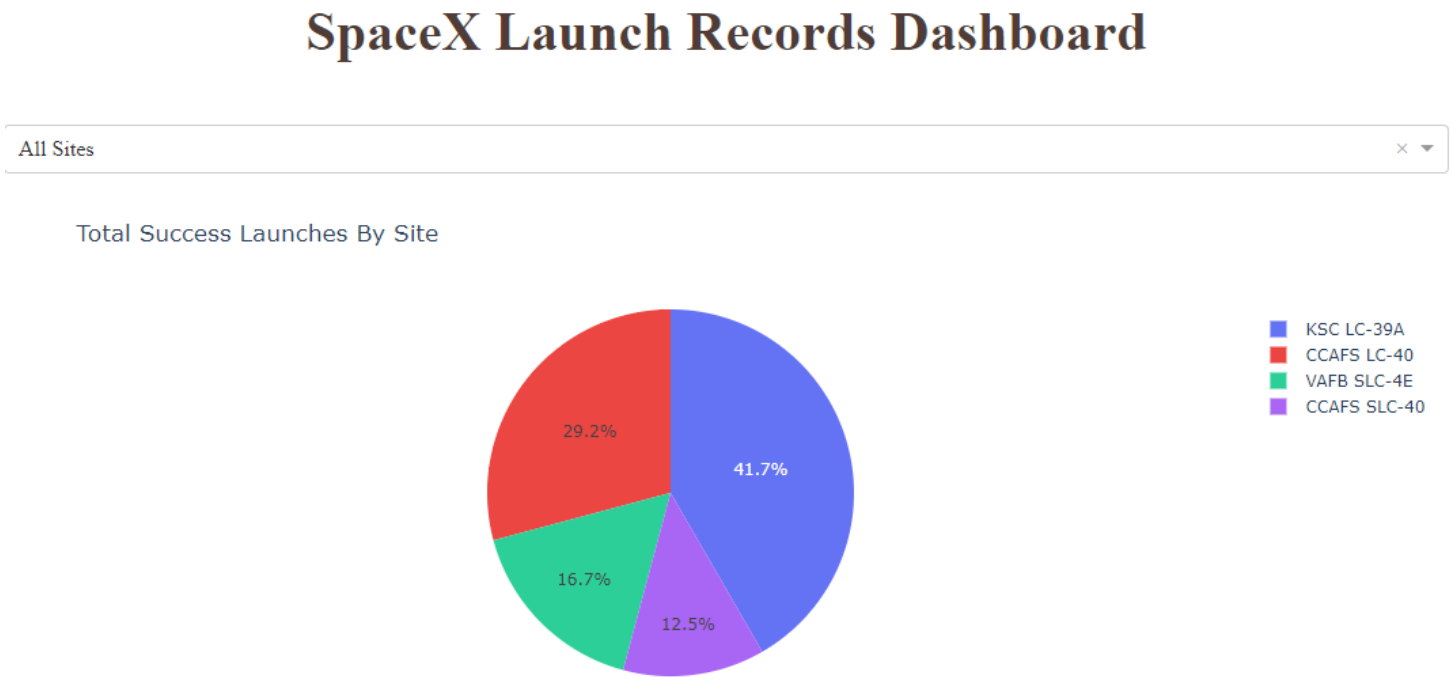


Distance between Launch Site and an Airport

- Here I chose the California Launch Site to show the distance between the launch site and the closest airport.
- The distance between the VAFB SLC-4E Launch Site and the Lompac Airport is 13.45 KM.

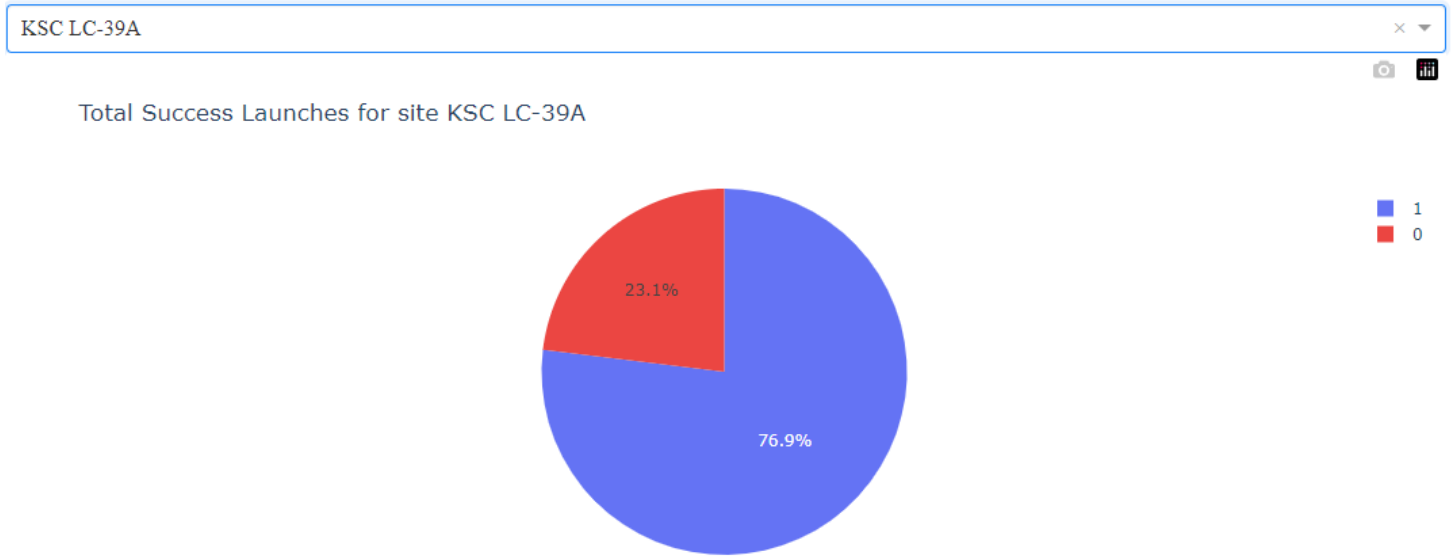



Section 4
Build a Dashboard with Plotly Dash



All Launch Site Success Pie Chart

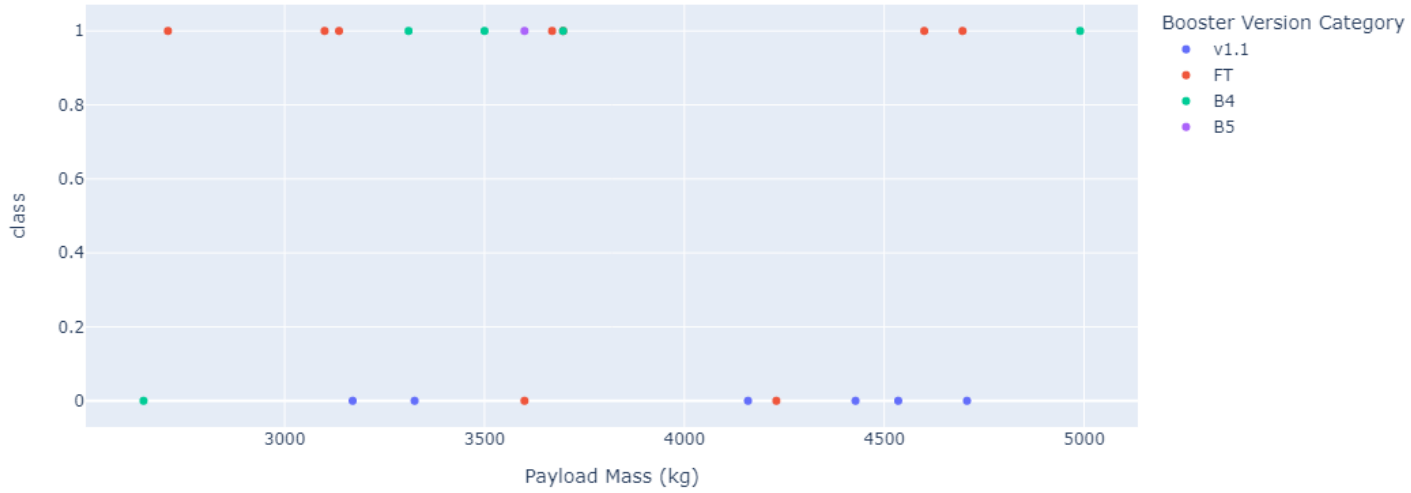
- Here is a dashboard using Plotly to show the total successful launches from each Launch Site.
- KSC LC 39A: 41.7%
- CCAFS LC-40: 29.2%
- VAFB SLC-4E: 16.7%
- CCAFS SLC-40: 12.5%



Highest Success Rate Launch Site

- This pie chart is showing the launch site with the highest success rate.
- KSC LC-39A has a success rate of 76.9%!

All Sites Payload between 2500 and 5000



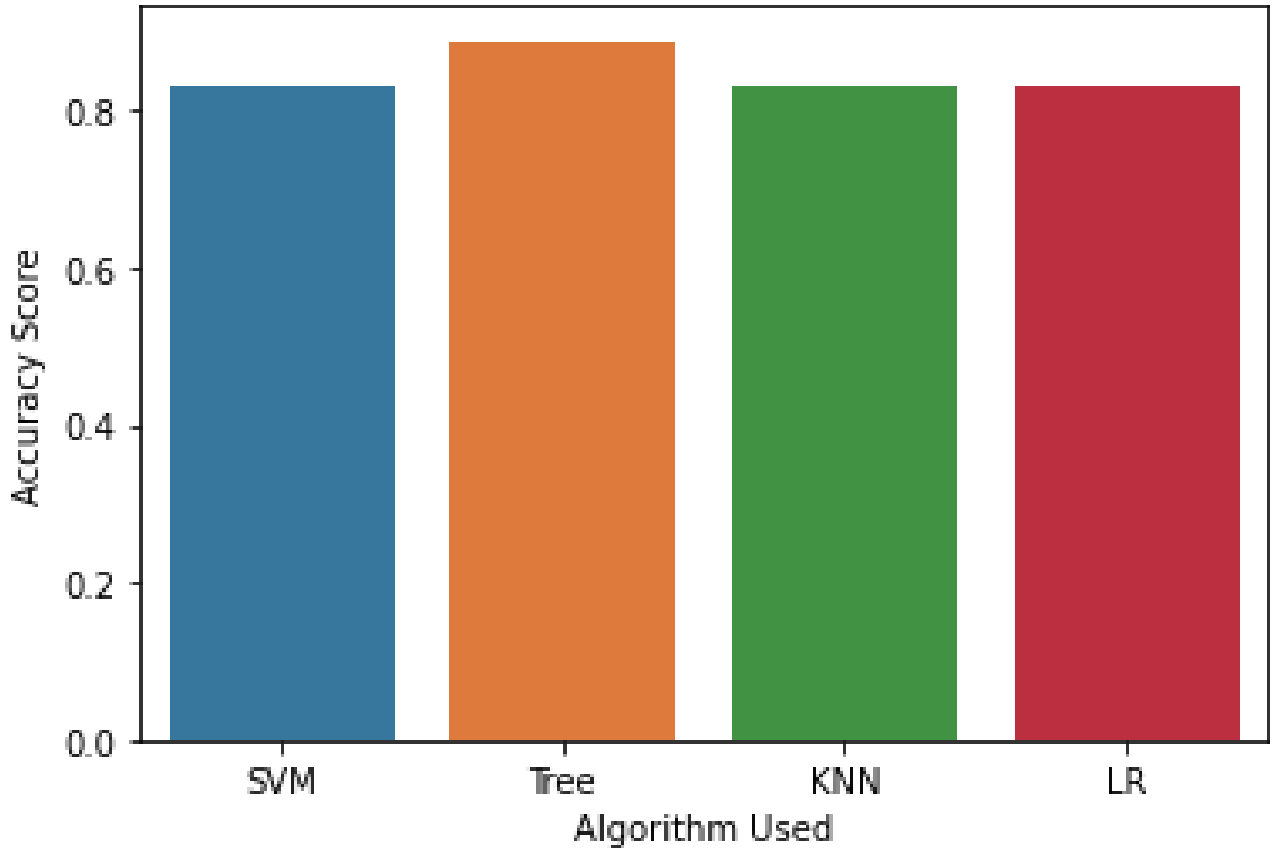
- This scatter plot is showing the Successful/Failed launches with payloads between 2500 and 5000 KG.
- There are multiple boosters used for these launches. F9 v1.1, FT, B4, and B5.

Section 5

Predictive Analysis (Classification)

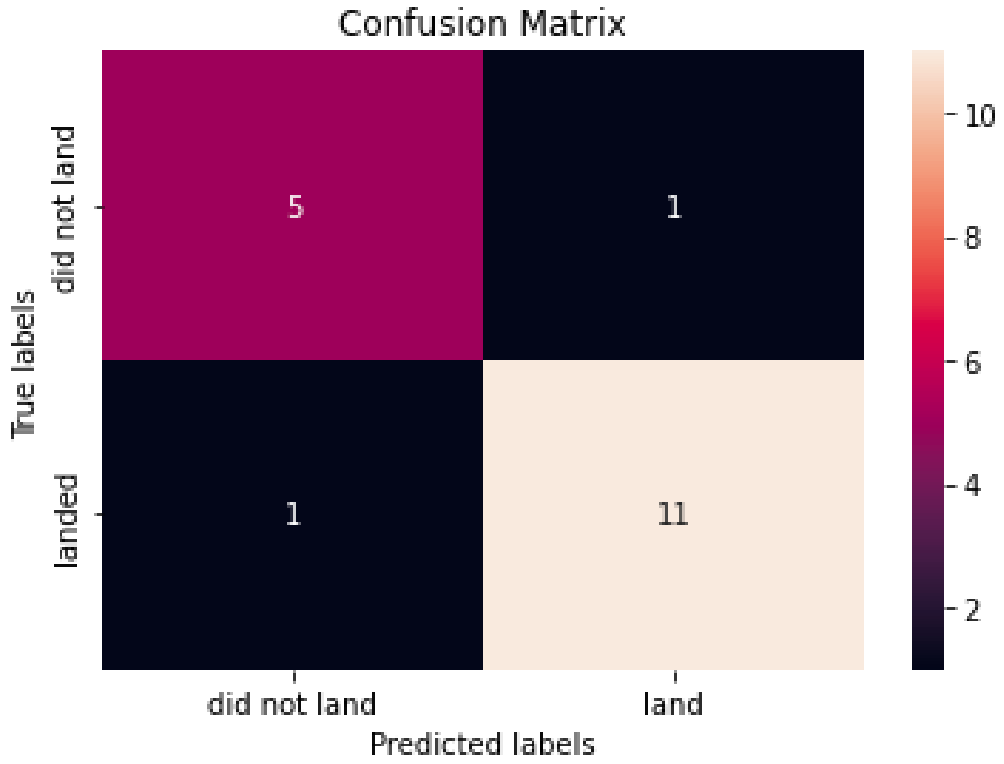
Classification Accuracy

- These are all the algorithms we used and their accuracy ratings.
  - SVM: 81%
  - KNN: 81%
  - LR: 81%
  - Decision Tree: 88%
- We will use the Decision Tree algorithm as it has an 88% Accuracy!



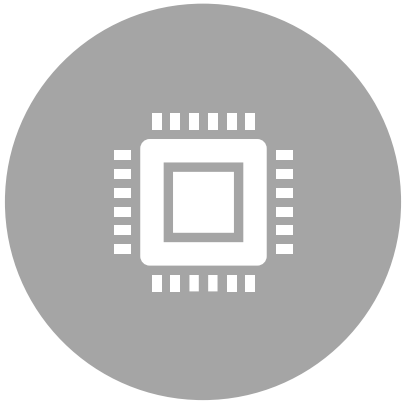
Confusion Matrix

- This is the confusion matrix for the Decision Tree Model.
- It classified 5 “did not land” and 11 “land” correctly.
- It has 1 false positive with “landed” and 1 false negative with “did not land”.



Conclusions

SPACEX HAD COME A LONG WAY IN THE YEARS WITH AN AVERAGE SUCCESSFUL MISSION RATING OF 83%.



WITH SPACEX AVERAGING A COST OF JUST 60 MILLION COMPARED TO 165 MILLION FROM ITS COMPETITORS; SPACEX IS A VERY GOOD ROUTE TO TAKE.

WITH OUR MODEL ABLE TO DETERMINE WITH 88% ACCURACY WHETHER THE MISSION WILL FAIL OR SUCCEED WE WILL HAVE A GOOD IDEA AS TO THE OUTCOME BEFORE THE ROCKET EVEN LAUNCHES.

Appendix

- We exported a few excel documents after doing our exploratory analysis.
- Dataset_part1.csv
- Dataset_part2.csv
- Dataset_part3.csv

Thank you!