IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Rafael Alonso Lazo
Tananta
27/11/2022

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- Project background and context

Nowadays, Space X company is reducing costs of rockets by reusing its first stages. And also if an alternate company wants to bid against SpaceX for a rocket launch, Space X could give them a more exactly price or budget. In order to do that we can use historical information of landing first stages success and find some patterns or variables that have an influence on this.

- Problems or questions we want to answer:

1. What are the characteristics or the variables that influence in the success Stage landing?

2. What should be the correct AI Model that we have to use in order to predict if the Falcon 9 first stage will land successfully?

Section 1

Methodology

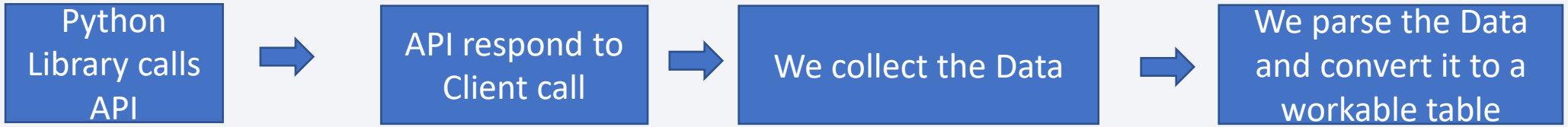
Methodology

Executive Summary

- Data collection methodology:
  - How data was collected is described in the following slides.
- Perform data wrangling
  - How data was processed is described in the following slides.
- 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 Classification Models were built, tuned and evaluated are described in the following slides.

Data Collection

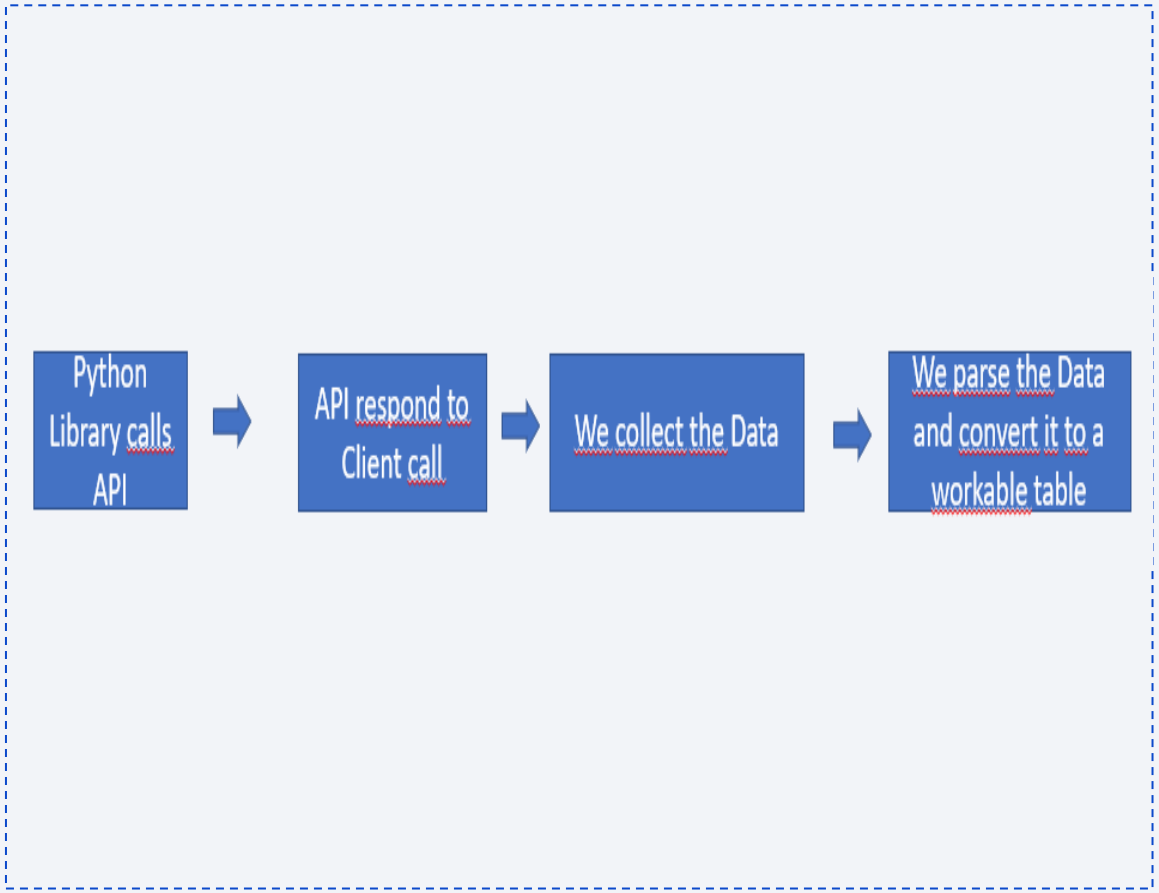
- Data sets where collected from API of SpaceX which is provided in the course.
- FlowChart of Data Collection



Data Collection – SpaceX API

- In the following Chart we can see the flow of Data Collection.

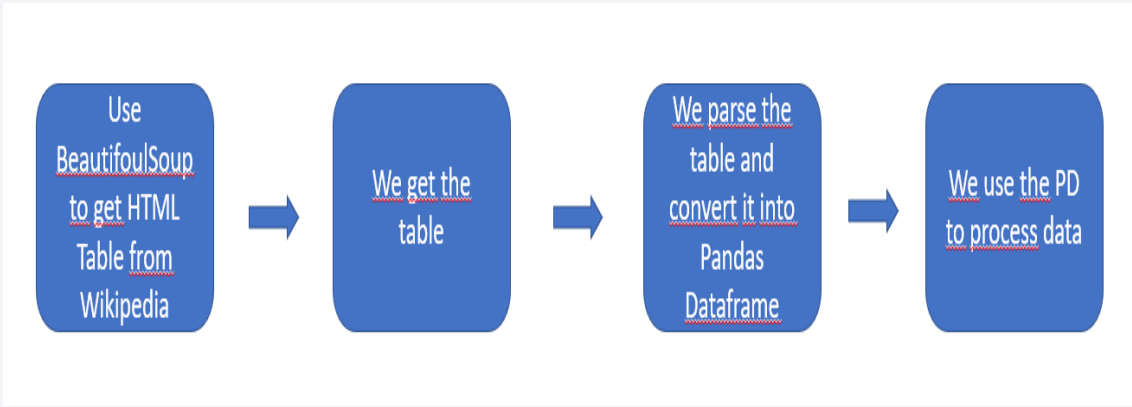
- GutHub URL is here: https://github.com/rlazot/TESTREPO/blob/master/FINAL_ASSIGMENT_DATA_COLLECTION.ipynb



Data Collection - Scraping

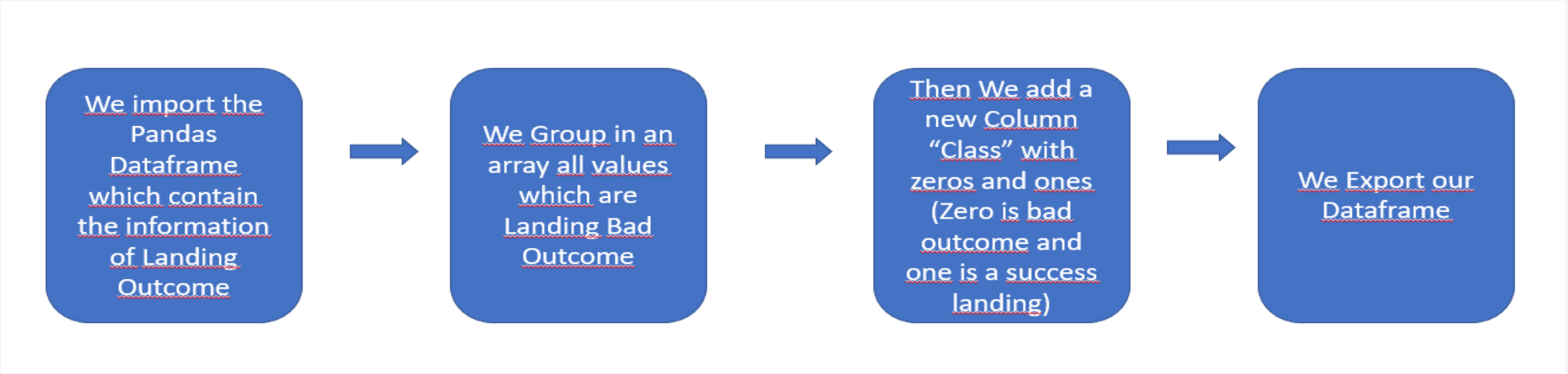
- In this case we are using BeautifulSoup to do WebScraping, and we are extracting a Falcon 9 launch records HTML table from Wikipedia
- Finally, we parse the table and convert it into a Pandas data frame, so that we can use it later.

- GitHub URL is here: https://github.com/rlazot/TESTREPO/blob/master/Data%20Collection%20with%20Web%20Scraping.ipynb



Data Wrangling

- Data were processed by transforming “Outcome” strings into label “Class”, the process is described in the following part:



- Github URL is here: https://github.com/rlazot/TESTREPO/blob/master/Complete%20the%20EDA%20lab.ipynb

EDA with Data Visualization

- We have been using the following charts:
- Scatter Point Chart, a) to plot out the FlightNumber vs. PayloadMassand overlay the outcome of the launch. b) to plot FlightNumber vs LaunchSite, set the parameter x parameter to FlightNumber,set the y to Launch Site and set the parameter hue to 'class'. c) To visualize the relationship between Payload and Launch Site. D) to visualize the relationship between FlightNumber and Orbit type, e) to visualize the relationship between Payload and Orbit type
- Bar Chart, to visualize the relationship between success rate of each orbit type.
- Line Chart, to visualize the launch success yearly trend.
- GitHub URL is here: https://github.com/rlazot/TESTREPO/blob/master/Complete%20the%20EDA%20with%20Visualization%20lab.ipynb

EDA with SQL

- We performed SQL queries to:
- 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 acheived.
- 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. Use a subquery
- 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
- GitHub URL is here: https://github.com/rlazot/TESTREPO/blob/master/Complete%20the%20EDA%20with%20SQL%20lab.ipynb

Build an Interactive Map with Folium

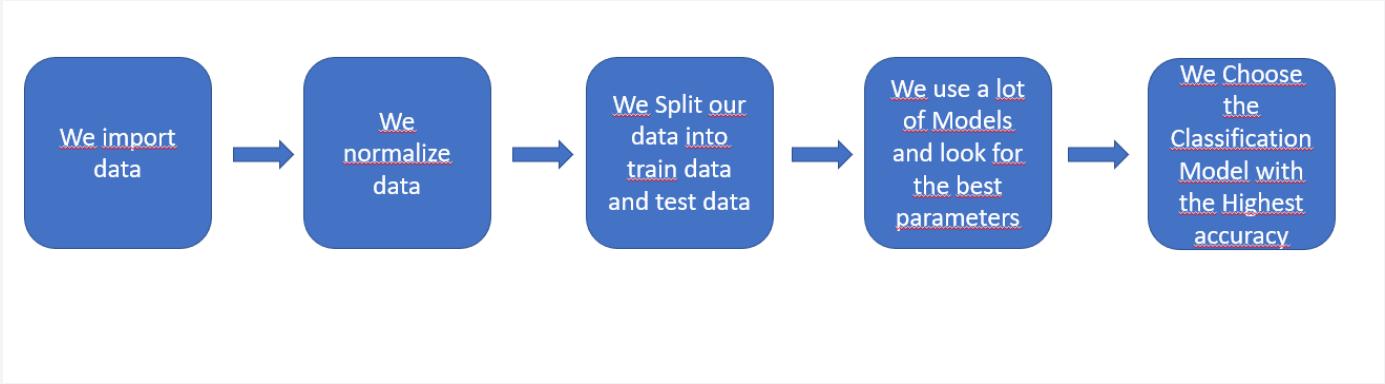
- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- We add map objects such as:
- Folium.circle, to add a highlighted circle area with a text label on a specific coordinate.
- Folium.Marker, for each launch site on the site map.
- Colors, to identify success outcome.
- PolyLine, between a launch site to the selected coastline point, to indentify distance.
- GitHub URL here: https://github.com/rlazot/TESTREPO/blob/master/Complete%20the%20Interactive%20Visual%20Analytics%20with%20Fol.ipynb

Build a Dashboard with Plotly Dash

- Plots/graphs and interactions we use:
- pie chart to show the total successful launches count for all sites
- scatter chart to show the correlation between payload and launch success
- site-dropdown, to give user facility to select site and filter by it.
- payload-slider, to give user facility to filter between ranges of payload.
- GitHub URL is here:
  https://github.com/rlazot/TESTREPO/blob/master/spacex_dash_app.py

Predictive Analysis (Classification)

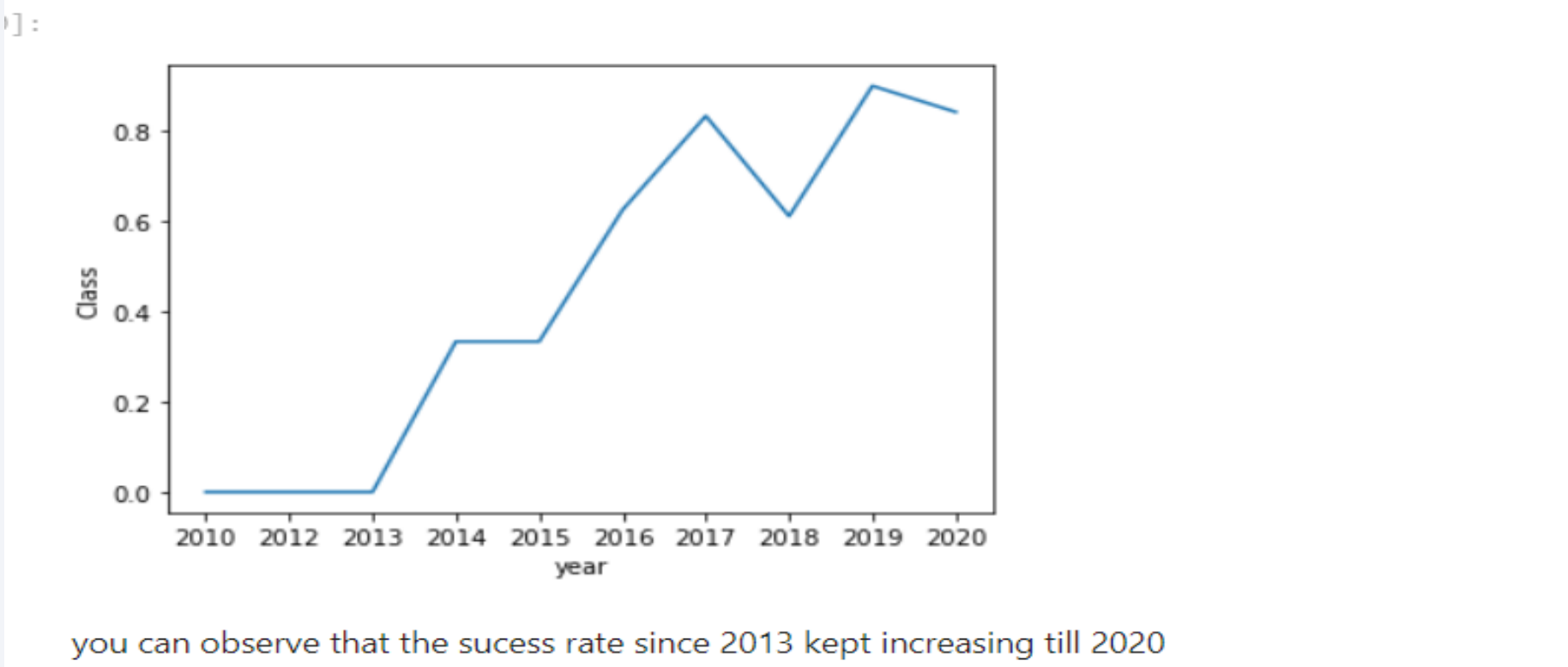
- How did we do it?
- We build our classification model by using historical data and after we normalize it.
- We evaluated our model by using testing data.
- We improve our classification model by using Grid Cross Validation and it find our best parameters, so our best performance for our model.
- We found the best performing classification model by using the best of the models, in other word, the best model with the best parameters which is the one that have the highest value in the accuracy of MODEL on the trained data and in the accuracy of MODEL on the test data.



- GitHub URL is here: https://github.com/rlazot/TESTREPO/blob/master/Complete%20the%20Machine%20Learning%20Prediction%20lab.ipynb

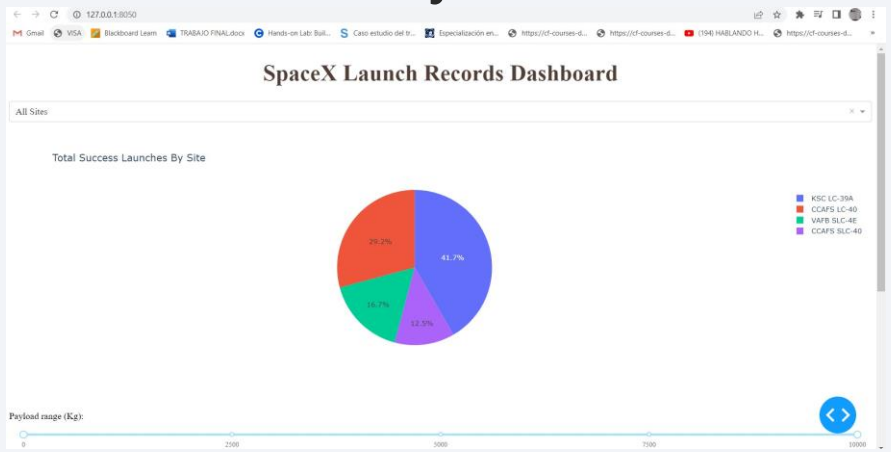
Results

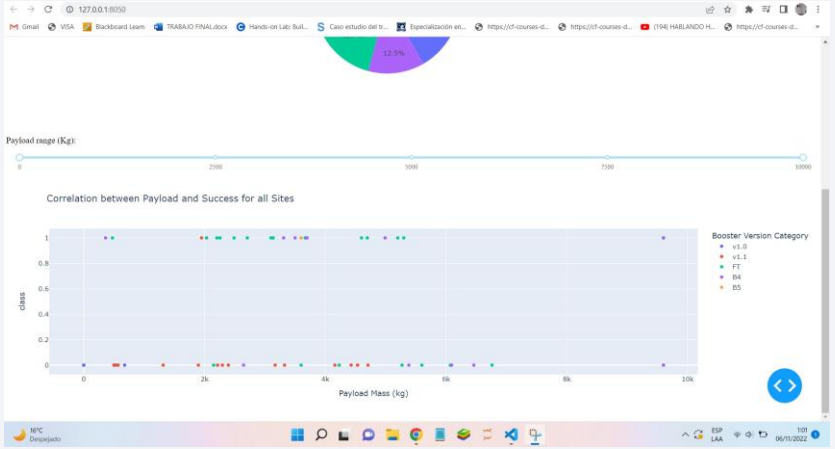
- Exploratory data analysis results



Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots





- Predictive analysis results

Results

Predictive analysis results

We can see that decision tree classifier model is the best model with the highest accuracy.

TASK 9

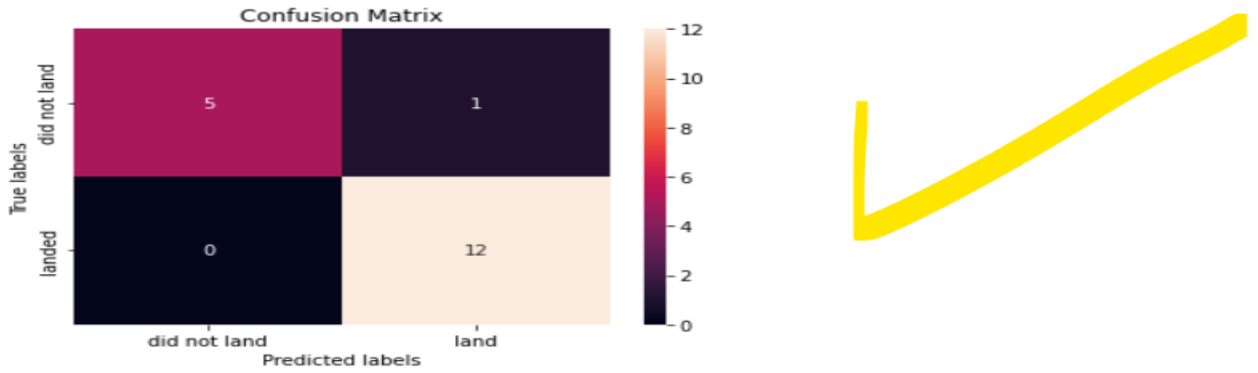
Calculate the accuracy of tree_cv on the test data using the method `score` :

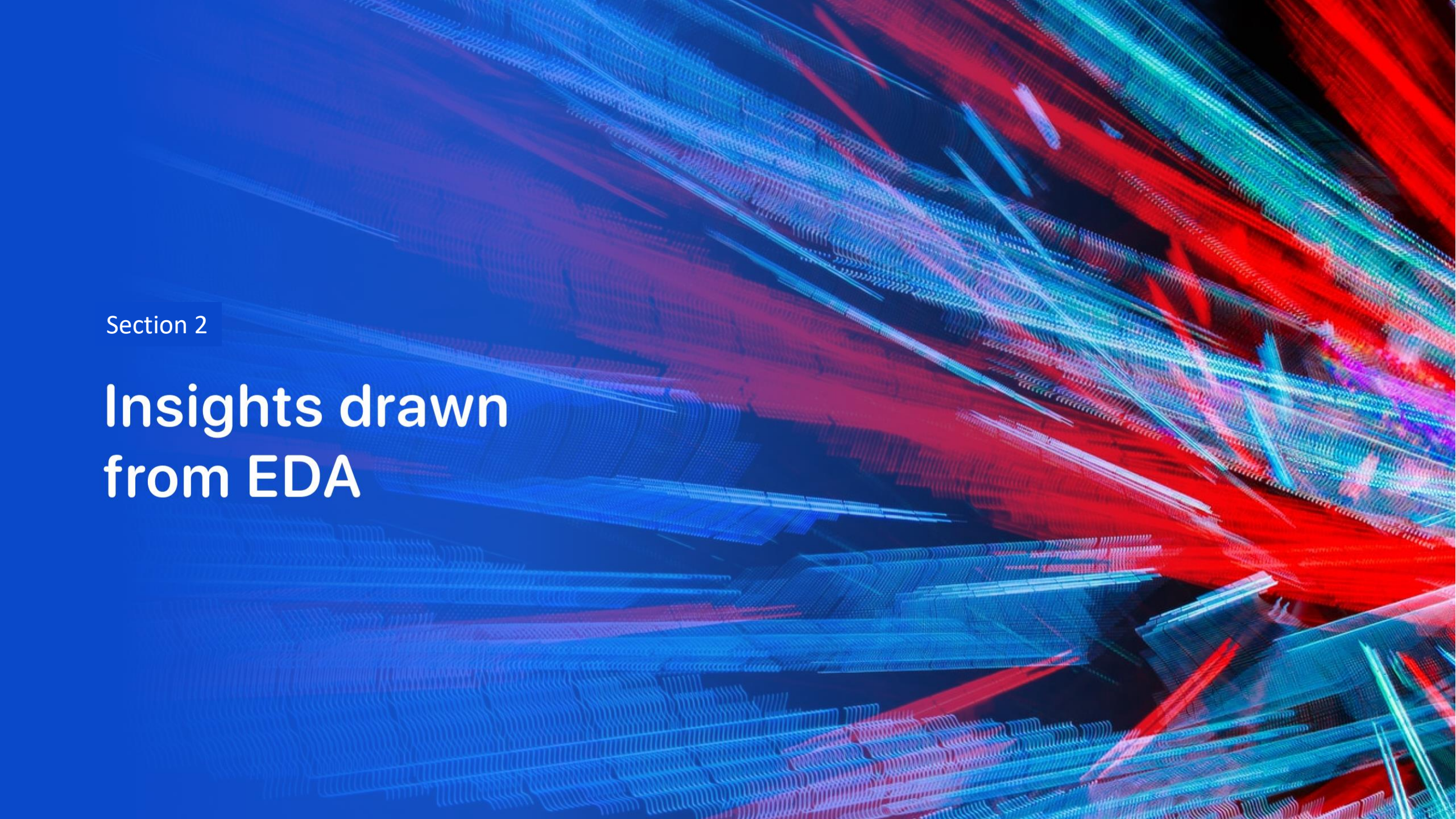
In [23]:
```
tree_cv.score(X,Y)
```

Out[23]: 0.8888888888888888

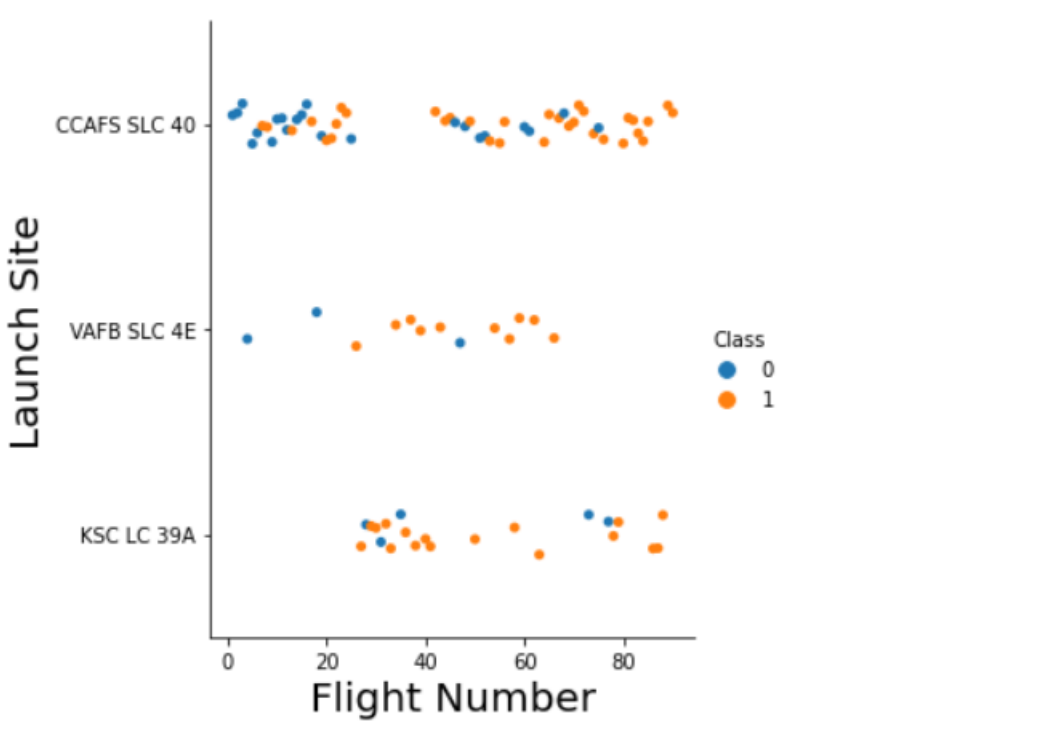
We can plot the confusion matrix

In [24]:
```
yhat = tree_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```



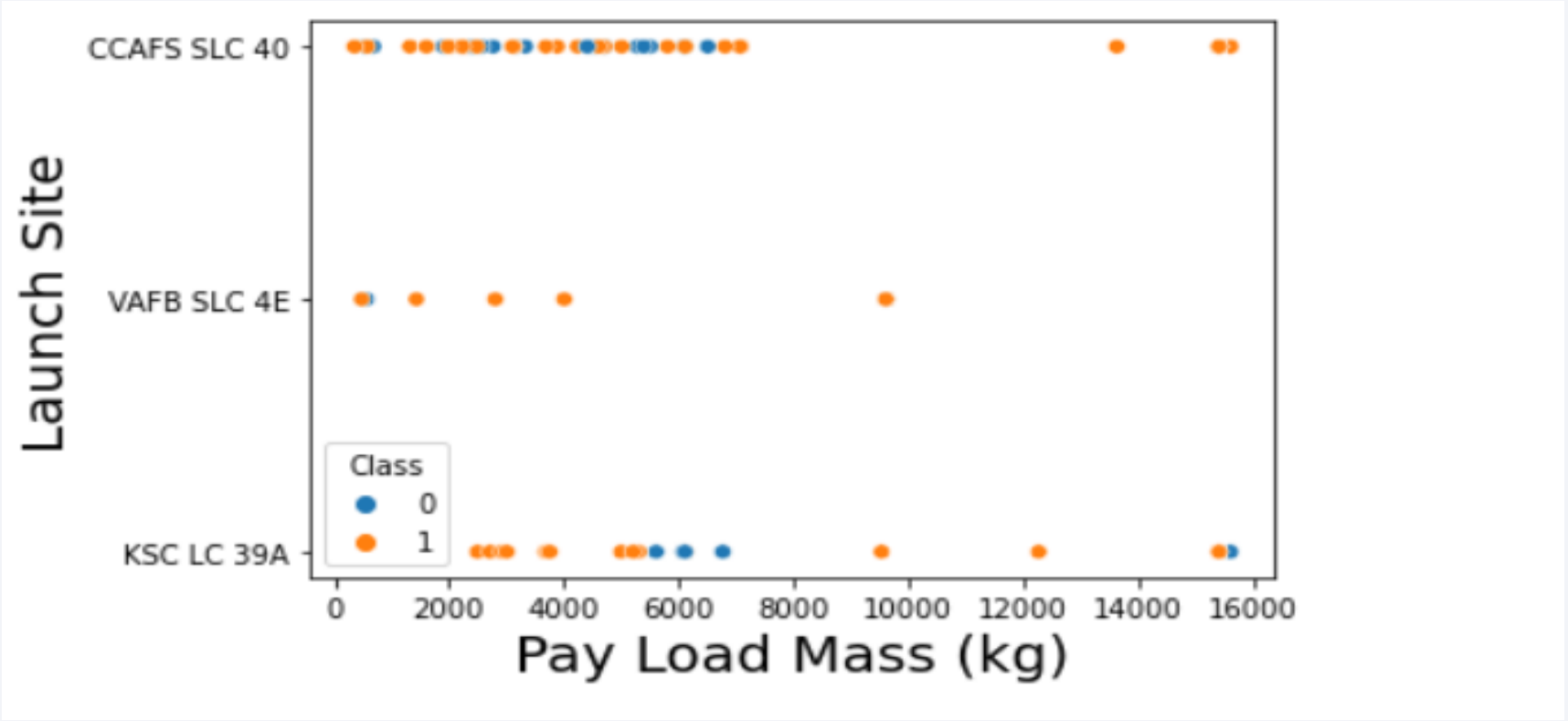

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



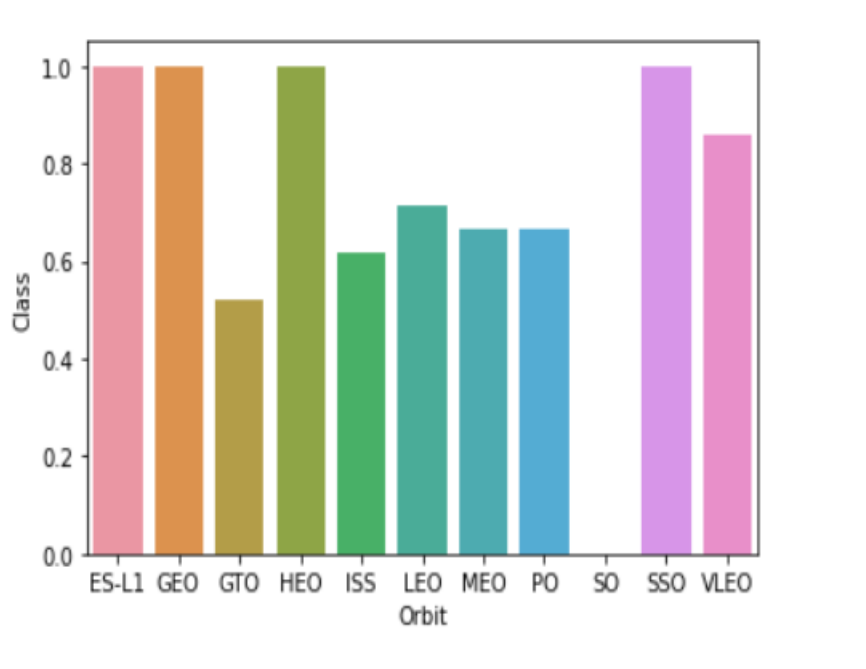
- Explanation: We see that the more Flight Number, the more Landing Success there is for all Launch Sites, but we can see that We don´t have enough information for lower Flight Number for Two of Lauch Site.

Payload vs. Launch Site



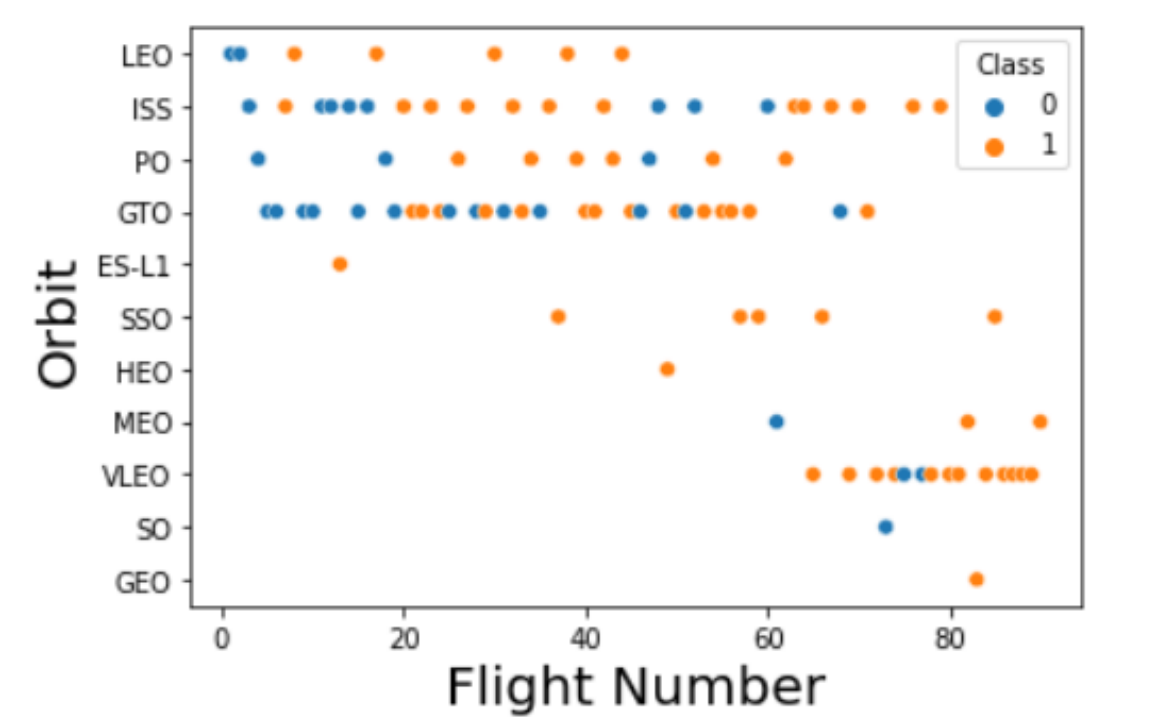
- Explanation: We observe Payload Vs. Launch Site scatter point chart that for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000), but at the same time that there is not a relationship between Launch Site for CCAFS SLC 40 and Payload Mass.

Success Rate vs. Orbit Type



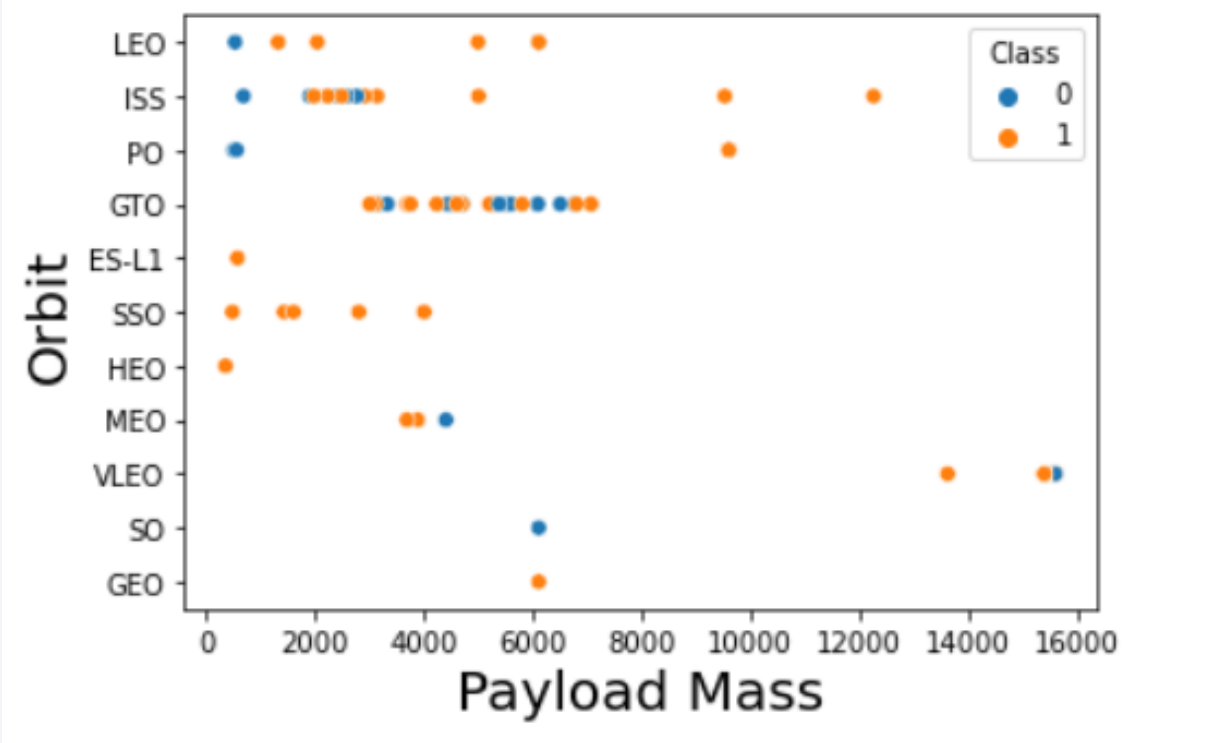
Explanation: The orbits with the highest success rate is ES-LI, GEO, HEO AND SSO.

Flight Number vs. Orbit Type



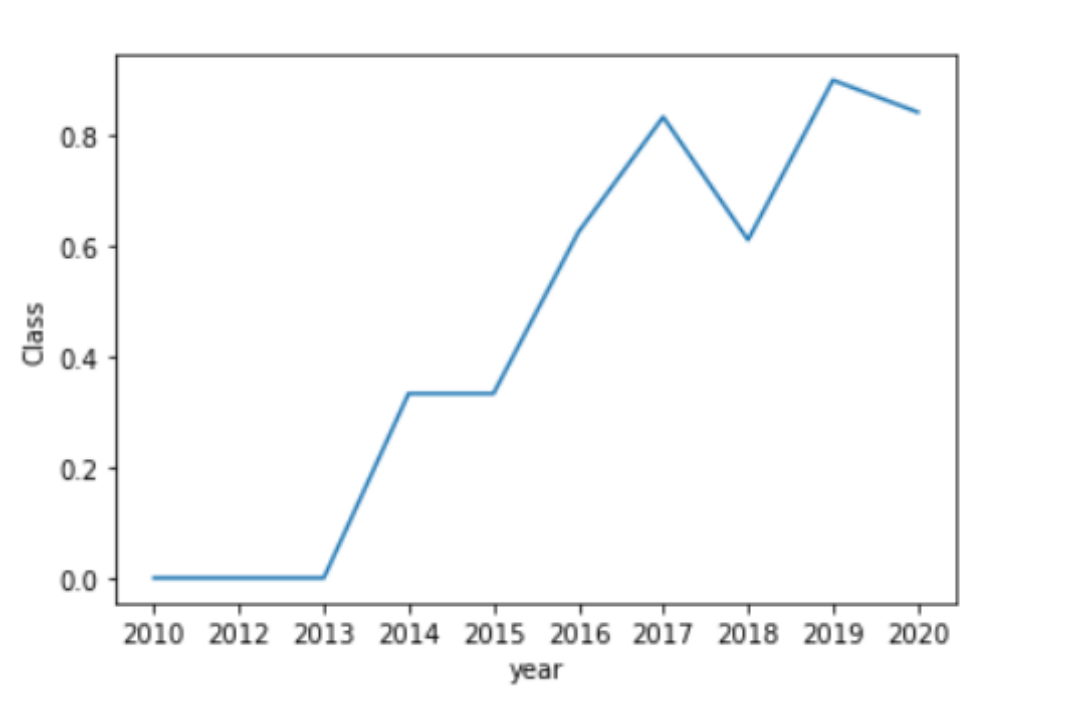
Explanation: We see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit. And also in PO, could be a relationship with Flight Number because we have only 1 Unsuccess landing.

Payload vs. Orbit Type



Explanation: 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. Also that for ES-L1, HEO AND GEO we don´t have enough information.

Launch Success Yearly Trend



Explanation: We can observe that the success rate since 2013 kept increasing till 2020 and we can see the increase never follows an absolute line, but it decrease a little after 1 or 2 years, so we hope that for 2021 it increase and for 2022 or 2023 it decrease a little.

All Launch Site Names

Task 1

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

```
9]: %sql select COUNT(launch_site) AS COUNTER,launch_site from SPACEX GROUP BY launch_site
```

```
 * ibm_db_sa://lgp68344:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

9]:

counter	launch_site
26	CCAFS LC-40
34	CCAFS SLC-40
25	KSC LC-39A
16	VAFB SLC-4E

Explanation: We are displaying the names of the unique launch sites in the space mission and counting how many values of each of then we have.

Launch Site Names Begin with 'CCA'

Task 2

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

In [33]:
```
%sql select * from SPACEX WHERE launch_site LIKE 'CCA%' --LIMIT 5
```

```
 * ibm_db_sa://lgp68344:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

Out[33]:

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

Explanation: We see that the customer more repeated is NASA for the 5 records we display.

Total Payload Mass

Task 3

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

In [25]:
```
%sql SELECT customer,SUM(payload_mass__kg_) AS SUM_payload_mass__kg_ from SPACEX GROUP BY customer HAVING customer = 'NASA (CRS)'
```

```
 * ibm_db_sa://lgp68344:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

Out[25]:

customer	sum_payload_mass__kg_
NASA (CRS)	45596

Explanation: We see that 45596 Kg is the total payload mass carried by boosters lauched by NASA

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here

Explanation: We can observe that the success rate since 2013 kept increasing till 2020 and we can see the increase never follows an absolute line, but it decrease a little after 1 or 2 years, so we hope that for 2021 it increase and for 2022 or 2023 it decrease a little.

First Successful Ground Landing Date

Task 5

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

Hint:Use min function

In [32]:
```
%sql SELECT MIN(date) as f_s_l_o,landing__outcome FROM SPACEX GROUP BY landing__outcome HAVING landing__outcome LIKE 'Success (ground pad)'
```

```
 * ibm_db_sa://lgp68344:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

Out[32]:

f_s_l_o	landing__outcome
2015-12-22	Success (ground pad)

Explanation: the first successful landing outcome in ground pad was achieved on 22/12/2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

Task 6

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

In [36]:
```
%sql SELECT booster_version,payload FROM SPACEX WHERE landing__outcome LIKE 'Success (drone ship)' AND (payload_mass__kg_ >4000 AND payload_mass__kg_
```

```
 * ibm_db_sa://lgp68344:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

Out[36]:

booster_version	payload
F9 FT B1022	JCSAT-14
F9 FT B1026	JCSAT-16
F9 FT B1021.2	SES-10
F9 FT B1031.2	SES-11 / EchoStar 105

Explanation: We see that this BOOSTER VERSIONS have success in drone ship and have payload mass greater than 4000 but less than 6000

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

In [39]:
```
%sql SELECT COUNT(*) as count FROM SPACEX  WHERE landing__outcome LIKE 'Success%' OR landing__outcome LIKE 'Failure%'
```

```
 * ibm_db_sa://lgp68344:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

Out[39]:

COUNT
71

Explanation: We that the total number of successful and failure mission outcomes is 71.

Boosters Carried Maximum Payload

Task 8

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

In [49]:
```
%sql SELECT DISTINCT booster_version FROM SPACEX WHERE payload_mass__kg_ = (SELECT MAX(payload_mass__kg_) AS MAX_payload_mass__kg_ FROM SPACEX)
```

```
 * ibm_db_sa://lgp68344:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

Out[49]:

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

Explanation: We see the list of the names of the booster versions which have carried the maximum payload mass.

2015 Launch Records

Task 9

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

In [55]:
```
%sql SELECT landing__outcome,booster_version,launch_site FROM SPACEX WHERE landing__outcome LIKE 'Failure (drone ship)' AND YEAR(DATE) = 2015
```

```
 * ibm_db_sa://lgp68344:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

Out[55]:

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

- Explanation: We see the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015. And the Launch_site is CCAFS LC-40 and its booster version F9 v1.1 B1012 and F9 v1.1 B1015.

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

Task 10

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 [58]:
```
%sql SELECT COUNT(*) AS COUNTER,landing__outcome FROM SPACEX WHERE DATE BETWEEN '2010-06-04' and '2017-03-20' GROUP BY landing__outcome ORDER BY COUNT
```

```
 * ibm_db_sa://lgp68344:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

Out[58]:

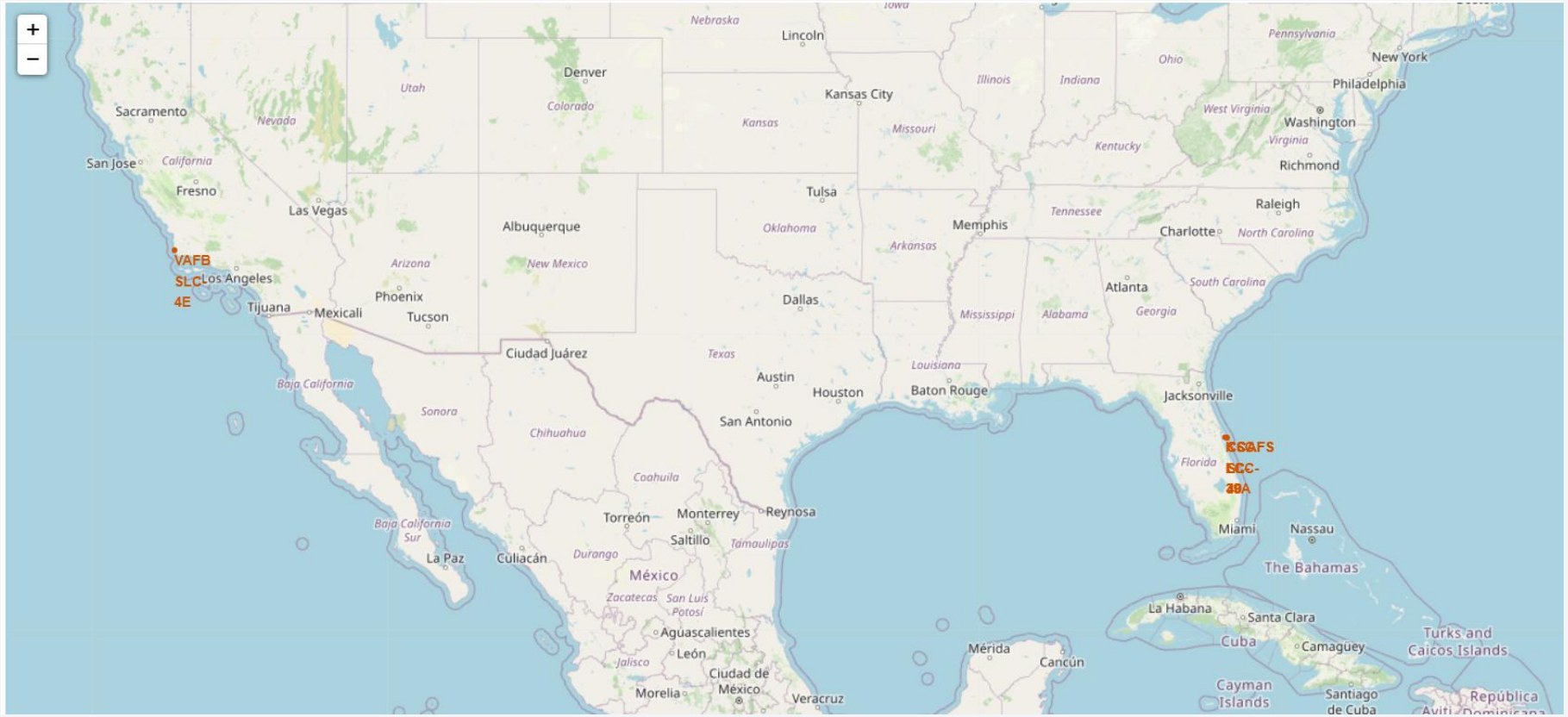
counter	landing__outcome
10	No attempt
5	Failure (drone ship)
5	Success (drone ship)
3	Controlled (ocean)
3	Success (ground pad)
2	Failure (parachute)
2	Uncontrolled (ocean)
1	Precluded (drone ship)

Explanation: We see that No attempt is the most common landing outcomes.

Section 3

Launch Sites Proximities Analysis

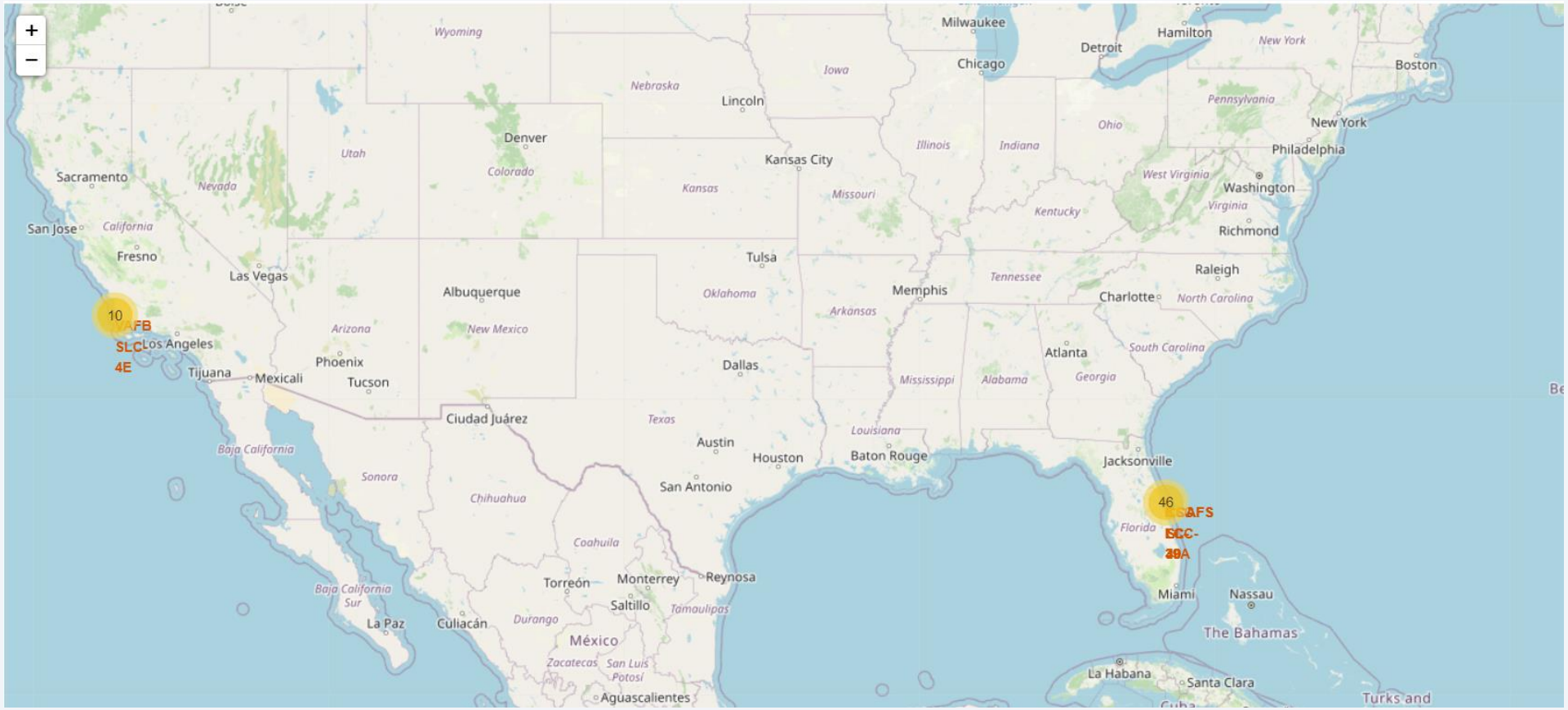
LAUNCH SITES LOCATIONS ON MAP



Explanation: All launch sites are not in proximity to the Equator line.

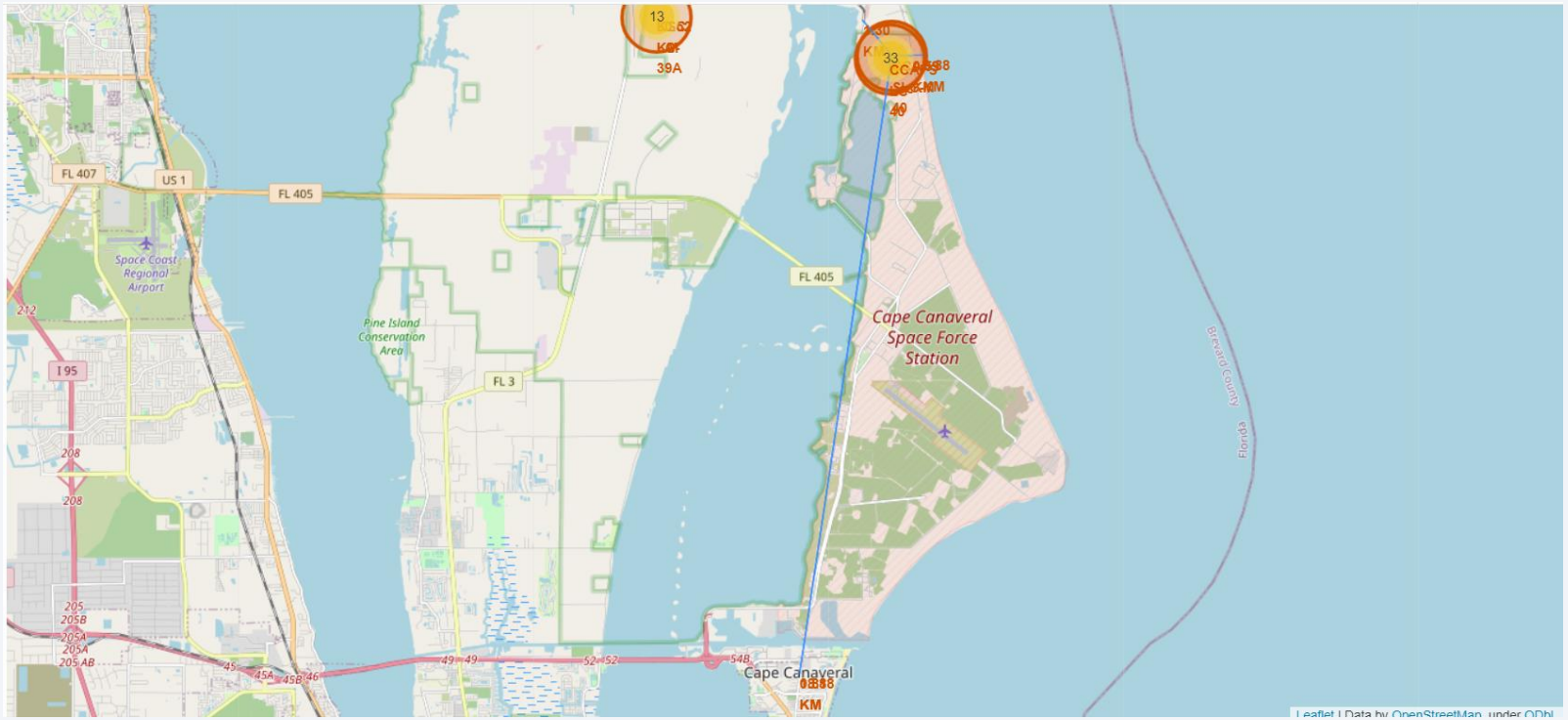
All launch sites are in very close proximity to the coast.

SUCCESS/FAILED LAUNCHES FOR EACH SITE ON THE MAP

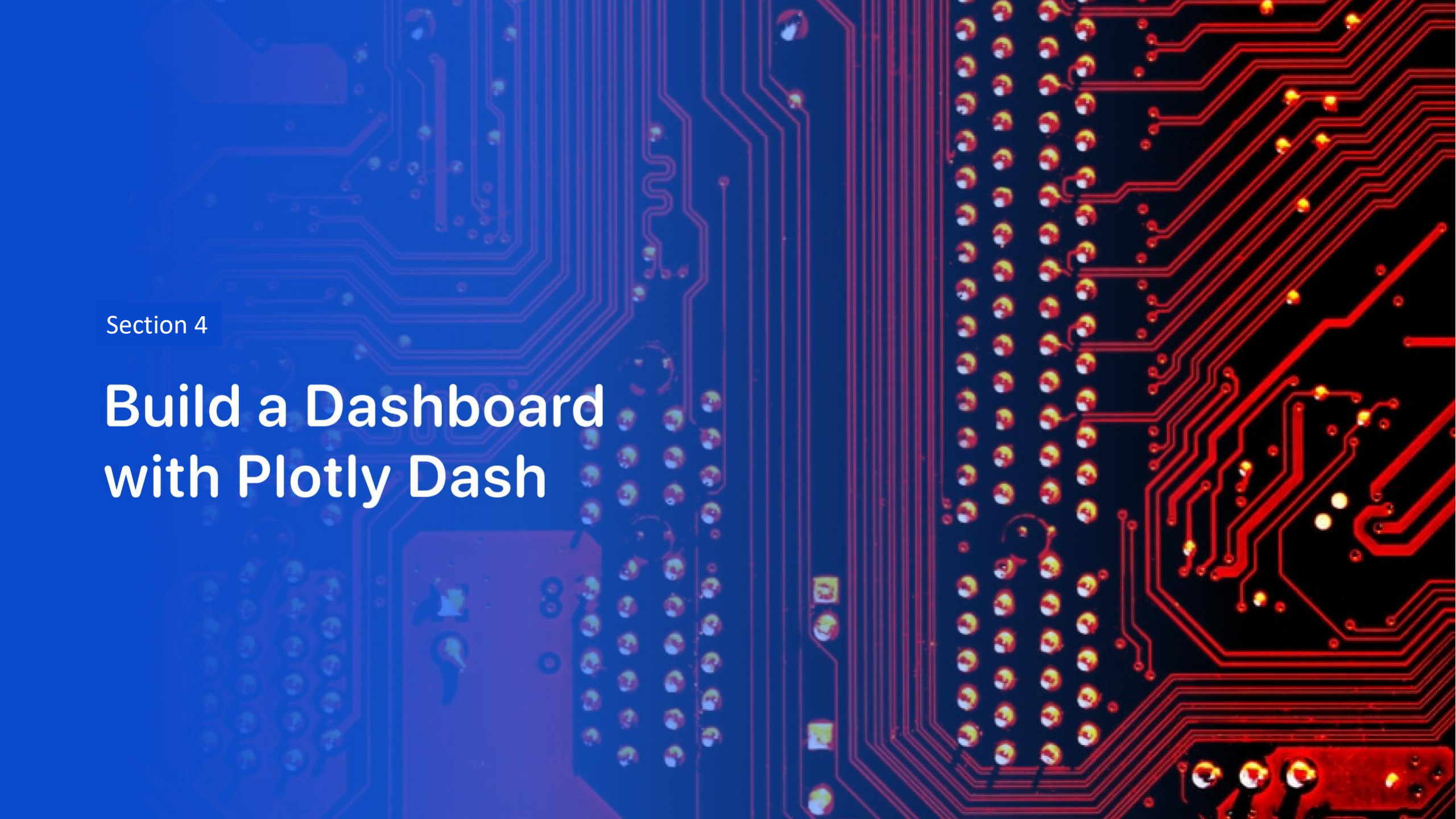


Explanation: We see that CCAFS SLC-40 is the Launch Site that have relatively high success rates. Also we can see that there are more attempts of Launches in CCAFS LC-40 Launch Site.

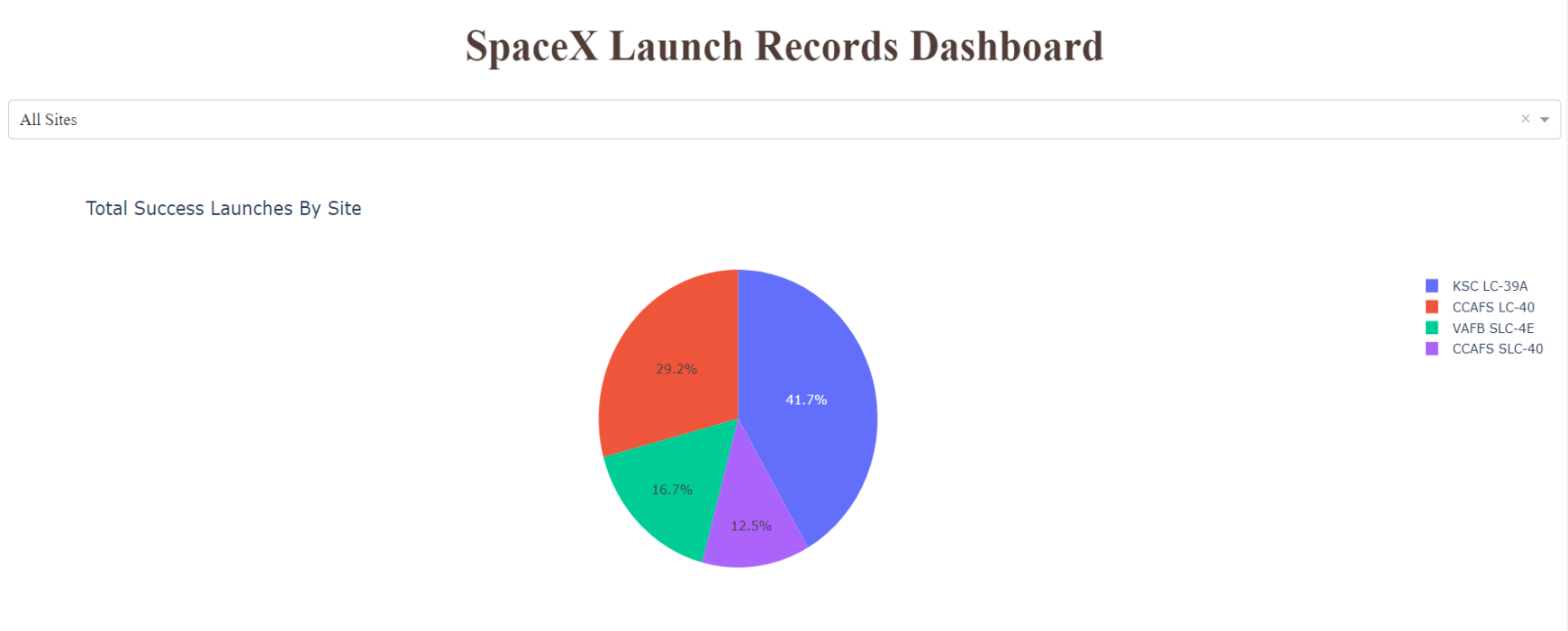
DISTANCES BETWEEN THE CCAFS SLC-40 LAUNCH SITE TO ITS PROXIMITIES



- Explanation: Launch Sites are in close proximity to railways. Launch Sites are in close proximity to highways. Launch Sites are in close proximity to coastline. Launch sites keep certain distance away from cities.

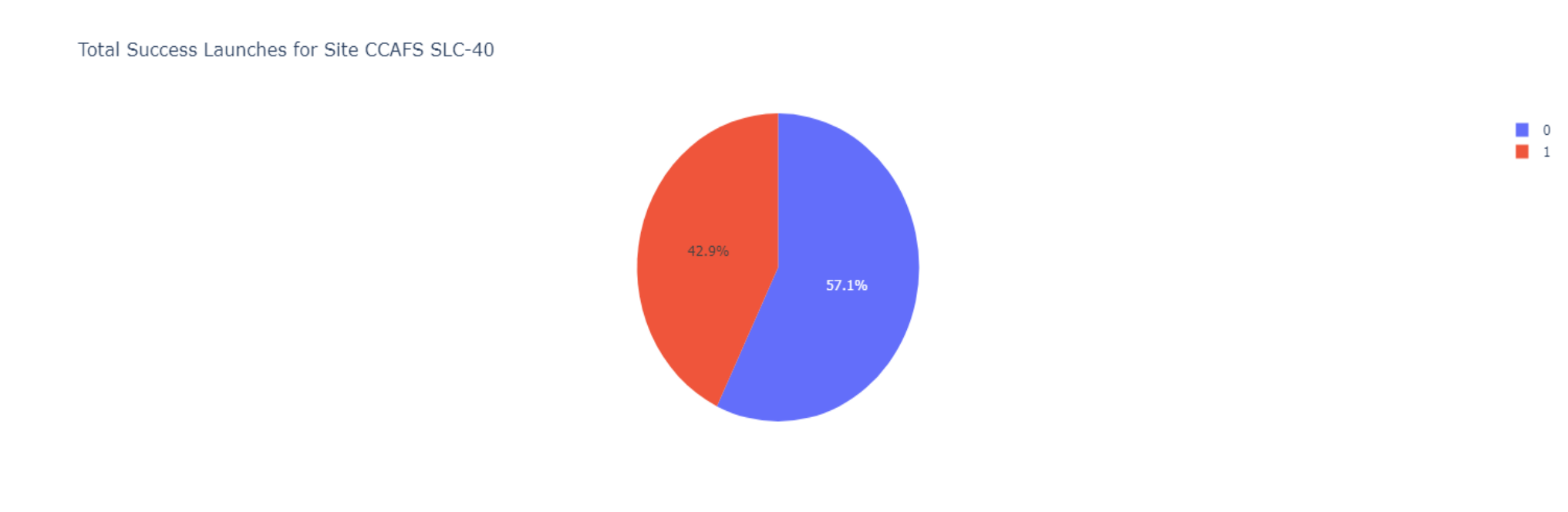

Section 4
Build a Dashboard with Plotly Dash

LAUNCH SUCCESS COUNT FOR ALL SITES



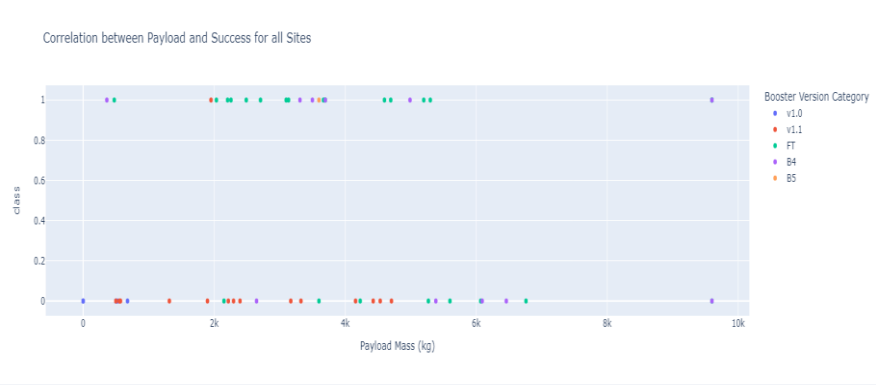
- Explanation: The Launch Site with the most success lunches is KSC LC-39A followed by CCAFS LC-40. The Launch Site with the fewest success lunches is CCAFS SLC-40.

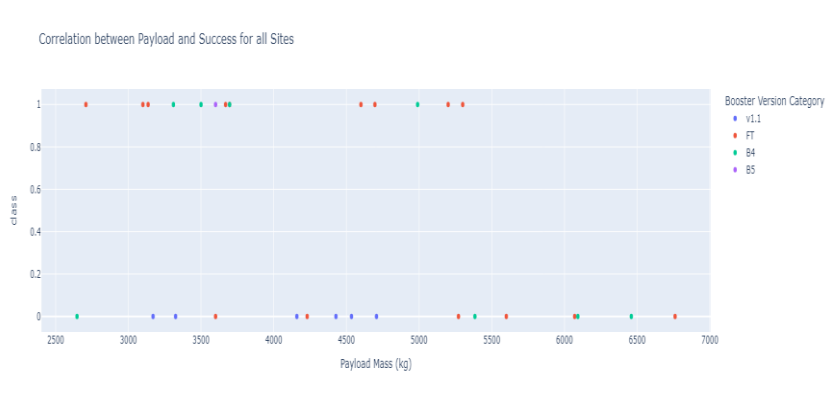
THE LAUNCH SITE WITH HIGHEST LAUNCH SUCCESS RATIO

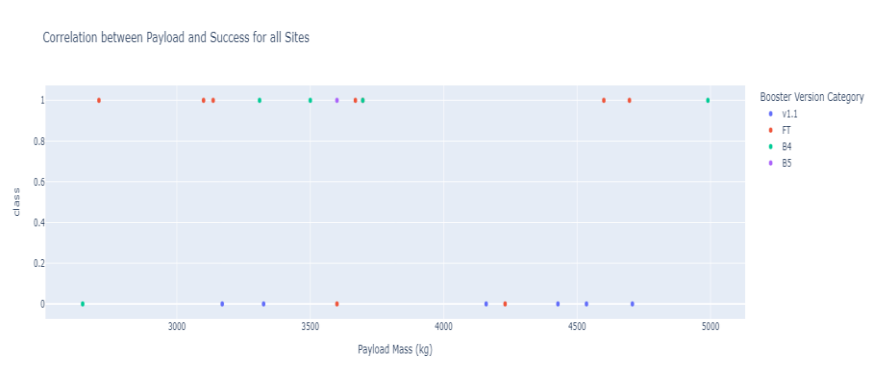


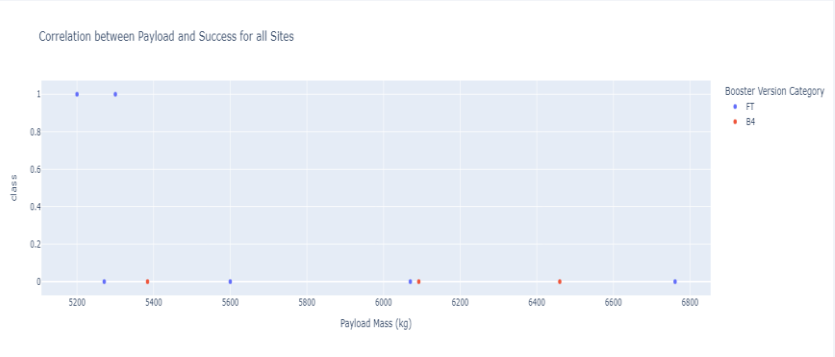
- Explanation: the launch site with highest launch success ratio is CCAFS SLC-40 which has 42,9% of success ratio.

CORRELATION BETWEEN PAYLOAD AND SUCCESS FOR ALL SITES









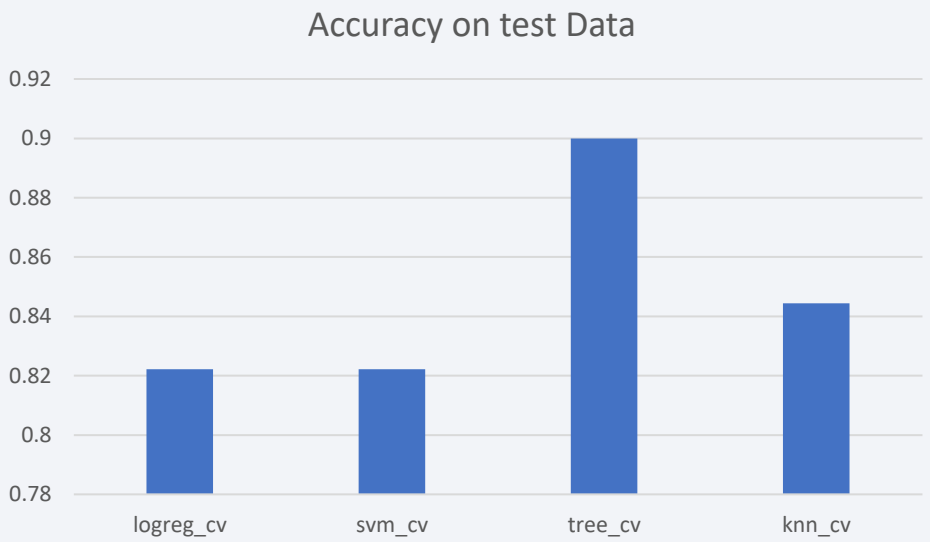
- Explanation: The Payload that has the largest success rate is from 2 000 to 4 000 kg. The Payload that has the lowest success rate is from 0 to 4 000 kg. The booster version that has the largest success rate is FT. The booster version that has the lowest success rate is v1,1.

Section 5

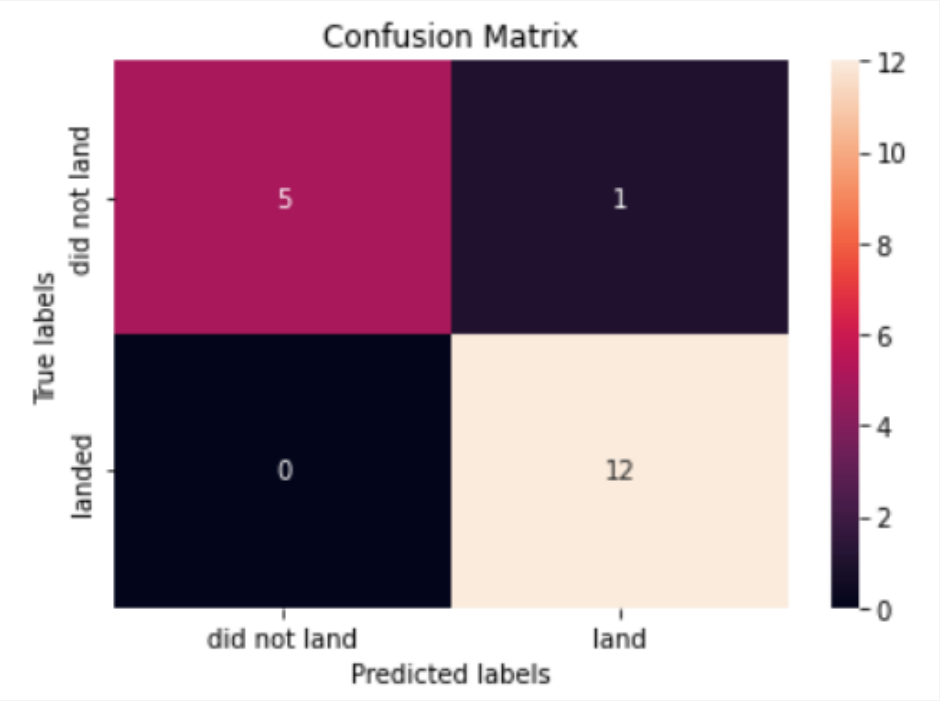
Predictive Analysis (Classification)

Classification Accuracy

- We have the Output for accuracy here
- The Decision Tree classification algorithm is the model that has the highest classification accuracy.



Confusion Matrix



- We can see here that for Decision Tree Classification Model there is just 1 wrong predicted value for 18 values, which is: The algorithm says that The Target Will Land but the true value is that it did’nt land.

Conclusions

- Decision tree classifier model is the best model with the highest accuracy.
- The Payload that has the largest success rate is from 2 000 to 4 000 kg.
- The booster version that has the largest success rate is FT.
- the launch site with highest launch success ratio is CCAFS SLC-40 which has 42,9% of success ratio.
- All launch sites are in very close proximity to the coast.
- We can observe that the success rate since 2013 kept increasing till 2020 and we can see the increase never follows an absolute line
- We see that the more Flight Number, the more Landing Success there is for all Launch Sites, but we can see that We don´t have enough information for lower Flight Number for Two of Lauch Site.

Appendix

- Data was collected from: https://api.spacexdata.com
- Data of List of Falcon 9 and Falcon Heavy launches was collected from https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

2020 [edit]

In late 2019, Gwynne Shotwell stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,[490] in addition to 14 or 15 non-Starlink launches. At 26 launches, 13 of which for Starlink satellites, Falcon 9 had its most prolific year, and Falcon rockets were second most prolific rocket family of 2020, only behind China's Long March rocket family.[491]

[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:21[492]	F9 B5 △ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[493]								
79	19 January 2020, 15:30[494]	F9 B5 △ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[495] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[496]	NASA (CTS)[497]	Success	No attempt
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes after reentry, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule,[498] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[419] The abort test used the capsule originally intended for the first crewed flight.[499] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[500] First flight of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its engine.								
80	29 January 2020, 14:07[501]	F9 B5 △ B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the ocean.[502]								
81	17 February 2020, 15:05[503]	F9 B5 △ B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 386 km (132 mi × 240 mi) elliptical orbit instead of launching into a circular orbit and firing the second stage engine twice. The first stage booster failed to land on the drone ship[504] due to incorrect wind data.[505] This was the first time a flight proven booster failed to land.								
82	7 March 2020, 04:50[506]	F9 B5 △ B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 △)	1,977 kg (4,359 lb)[507]	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
	Last launch of phase 1 of the CRS contract. Carries *Bartolomeo*, an ESA platform for hosting external payloads onto ISS.[508] Originally scheduled to launch on 2 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the faulty part.[509] It was SpaceX's 50th successful landing of a first stage booster, the third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.								
83	18 March 2020, 12:16[510]	F9 B5 △ B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second time the fairings were reused (Starlink flight in May 2019).[511] Towards the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-1 mission in October 2012. However, the payload still reached the targeted orbit.[512] This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual cleaning fluid trapped inside a sensor.[513]								
84	22 April 2020, 19:30[514]	F9 B5 △ B1051.4	KSC, LC-39A	Starlink 6 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)

Appendix

- Some SQL queries are here:

Task 7

List the total number of successful and failure mission outcomes

In [39]:
```
%sql SELECT COUNT(*) as count FROM SPACEX  WHERE landing__outcome LIKE 'Success%' OR landing__outcome LIKE 'Failure%'
```

```
 * ibm_db_sa://lgp68344:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

Out[39]:

COUNT
71

Task 8

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

In [49]:
```
%sql SELECT DISTINCT booster_version FROM SPACEX WHERE payload_mass__kg_ = (SELECT MAX(payload_mass__kg_) AS MAX_payload_mass__kg_ FROM SPACEX)
```

```
 * ibm_db_sa://lgp68344:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
Done.
```

Appendix

- Some Python Code Snippets are here:

TASK 1: Calculate the number of launches on each site

The data contains several Space X launch facilities: Cape Canaveral Space Launch Complex 40 **VAFB SLC 4E** , Vandenberg Air Force Base Space Launch Complex 4E **(SLC-4E)**, Kennedy Space Center Launch Complex 39A **KSC LC 39A** .The location of each Launch Is placed in the column `LaunchSite`

Next, let's see the number of launches for each site.

Use the method `value_counts()` on the column `LaunchSite` to determine the number of launches on each site:

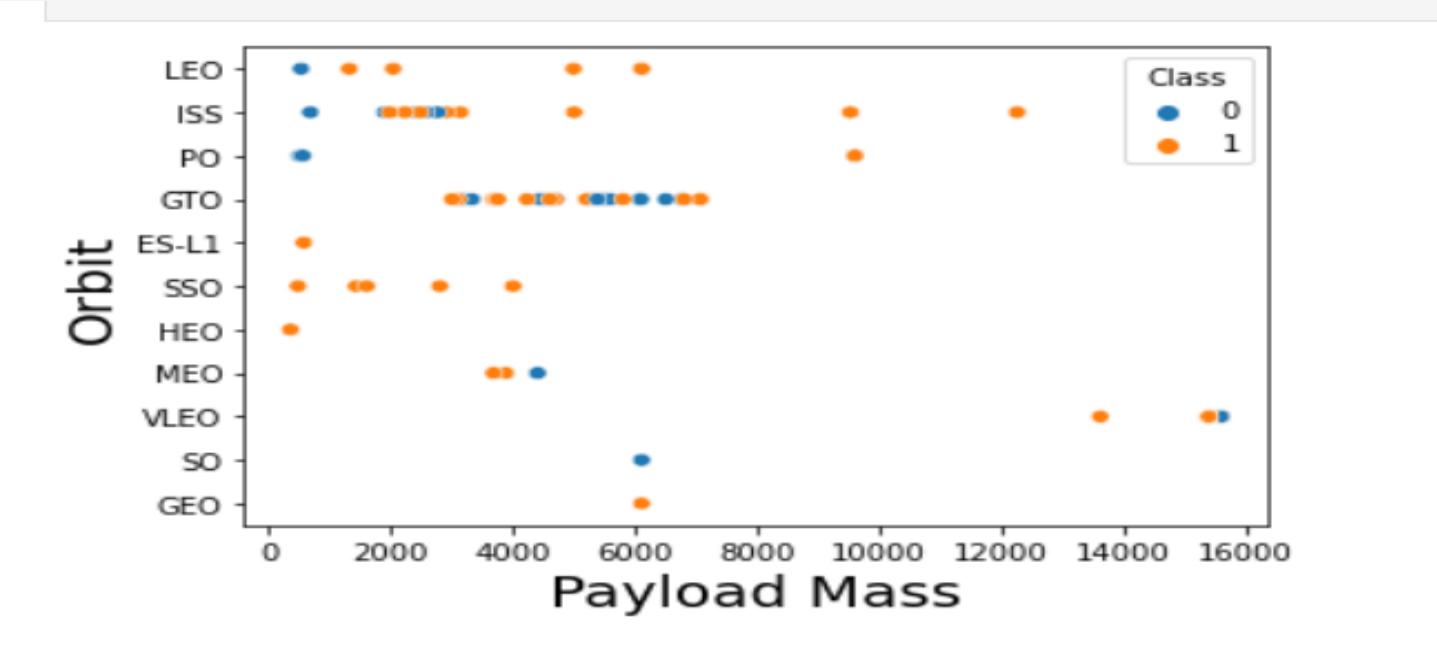
```
]:
# Apply value_counts() on column LaunchSite
df['LaunchSite'].value_counts()
```

```
]: CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
Name: LaunchSite, dtype: int64
```

Each launch aims to an dedicated orbit, and here are some common orbit types:

Appendix

- Some Charts are here:



Appendix

- Some Notebook outputs are here (top 10 records):

In [27]:
```
df.head(10)
```

Out[27]:

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitu
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561
5	6	2014-01-06	Falcon 9	3325.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1005	-80.577366	28.561
6	7	2014-04-18	Falcon 9	2296.000000	ISS	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1006	-80.577366	28.561
7	8	2014-07-14	Falcon 9	1316.000000	LEO	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1007	-80.577366	28.561
8	9	2014-08-05	Falcon 9	4535.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1008	-80.577366	28.561
9	10	2014-09-07	Falcon 9	4428.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1011	-80.577366	28.561

We can use the following line of code to determine the success rate:

Thank you!