



IBM Developer  
SKILLS NETWORK

# Winning Space Race with Data Science

Tey Ming Chuan  
30/11/2022



# Outline

---

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

# Executive Summary

---

## Summary of methodologies

- API Calls
- Webscraping

## Summary of all results

- Missions with higher payloads at CCAFS SLC 40 and Orbit Type of LEO, ISS and PO are very likely to succeed
- The first stage of Falcon 9 is very likely to land successfully

# Introduction

---

- Taking a role of a data scientist working for a new rocket for Space Y
- To gather useful insights to compete against Space X
- To determine the price of each launch
- To determine if the first stage of Falcon 9 will land successfully



Section 1

# Methodology

# Methodology

---

## Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- 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 to build, tune, evaluate classification models

# Data Collection

---

## SpaceX launch data will be gathered from the SpaceX REST API

- Contains data on launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
- Used to predict whether SpaceX will attempt to land a rocket or not

## Webscraping HTML table

- Contains information on Falcon 9 launch records
- To address: Wrangling Data using an API, Sampling Data and Dealing with Nulls

# Data Collection

Github URL:

[https://github.com/mariotey/IBM\\_Coursera/blob/master/Week 1/jupyter-labs-spacex-data-collection-api.ipynb](https://github.com/mariotey/IBM_Coursera/blob/master/Week%201/jupyter-labs-spacex-data-collection-api.ipynb)



SpaceX REST API

Open Source REST API for launch, rocket, core, capsule, starlink, launchpad, and landing pad data.

Build [\[link\]](#) Docker build [\[link\]](#) Issues [\[link\]](#) Interface [\[link\]](#)

We are not affiliated, associated, authorized, endorsed by, or in any way officially connected with Space Exploration Technologies Corp (SpaceX), or any of its subsidiaries or its affiliates. The names SpaceX as well as related names, marks, emblems and images are registered trademarks of their respective owners.

<https://api.spacexdata.com/v4/>

[api.spacexdata.com/v4/capsules](https://api.spacexdata.com/v4/capsules)

```
{
  "reuse_count": 0,
  "water_landings": 1,
  "land_landings": 0,
  "last_update": "Hanging in atrium at SpaceX HQ in Hawthorne",
  "launches": [
    {
      "id": "5eb87cdeffd86e000604b330",
      "serial": "C101",
      "status": "retired",
      "type": "Dragon 1.0"
    }
  ]
}
```

[api.spacexdata.com/v4/cores](https://api.spacexdata.com/v4/cores)

```
{
  "block": null,
  "reuse_count": 0,
  "rtls_attempts": 0,
  "rtls_landings": 0,
  "asds_attempts": 0,
  "asds_landings": 0,
  "last_update": "Engine failure at T+33 seconds resulted in loss of vehicle",
  "launches": [
    {
      "id": "5eb87cd9ffd86e000604b32a",
      "serial": "Merlin1A",
      "status": "lost"
    }
  ]
}
```

[api.spacexdata.com/v4/launches/past](https://api.spacexdata.com/v4/launches/past)

```
url = "https://api.spacexdata.com/v4/launches/past"
response = request.get(url)
response.json()
```



# Data Collection

Github URL:

[https://github.com/mariotey/IBM\\_Coursera/blob/master/Week 1/jupyter-labs-webscraping.ipynb](https://github.com/mariotey/IBM_Coursera/blob/master/Week%201/jupyter-labs-webscraping.ipynb)



2020 [edit]

In late 2019, *Gregory Stohrer* stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,<sup>[*link*]</sup> 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.<sup>[*link*]</sup>

[edit] Flight No.	Date and time (UTC)	Version, Booster <sup>[1]</sup>	Launch site	Payload <sup>[2]</sup>	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:23 <sup>[link]</sup>	FA 001.0 B1040.4	CCAFS, SLC-40	Starlink 2 v1.0 (90 satellites)	15,800 kg (34,400 lb) <sup>[3]</sup>	LEO	SpaceX	Success	Success (crowd step)
Third large batch and second operational flight of Starlink constellation. One of the 90 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations. <sup>[link]</sup>									
79	19 January 2020, 15:33 <sup>[link]</sup>	FA 001.0 B1040.4	KSC, LC-35A	Crew Dragon in-flight abort test <sup>[link]</sup> (Dragon C206.1)	12,000 kg (26,500 lb)	Sub-orbital <sup>[link]</sup>	NASA (COTS) <sup>[link]</sup>	Success	No attempt
An atmospheric test of the Dragon 2 abort system after <i>Starlink</i> . The capsule fired its SuperDraco engines, reached an apogee of 42 km (26 mi), deployed parachutes after reentry, and <i>splashed down</i> in the ocean 31 km (19 mi) downrange from the launch site. The test was previously stated to be accomplished with the <i>Crew Dragon Demo 1</i> capsule <sup>[link]</sup> but that test article exploded during a ground test of SuperDraco engines on 20 April 2019. <sup>[link]</sup> The abort test used the capsule originally intended for the first crewed flight. <sup>[link]</sup> As expected, the booster was destroyed by aerodynamic forces after the capsule aborted. <sup>[link]</sup> First flight of a Falcon 9 with only one functional stage — the second stage had a <i>mass simulator</i> in place of its engine.									
80	29 January 2020, 14:07 <sup>[link]</sup>	FA 001.0 B1001.2	CCAFS, SLC-40	Starlink 3 v1.0 (90 satellites)	15,800 kg (34,400 lb) <sup>[3]</sup>	LEO	SpaceX	Success	Success (crowd step)
Third operational and fourth large batch of Starlink satellites, deployed in a circular 200 km (130 mi) orbit. One of the failing Satlites was caught, while the other was fished out of the ocean. <sup>[link]</sup>									
81	17 February 2020, 15:05 <sup>[link]</sup>	FA 001.0 B1005.4	CCAFS, SLC-40	Starlink 4 v1.0 (90 satellites)	15,800 kg (34,400 lb) <sup>[3]</sup>	LEO	SpaceX	Success	Failure (crowd step)
Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km x 380 km (132 mi x 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 <sup>[link]</sup> due to incorrect wind data. <sup>[link]</sup> This was the first time a flight proven booster failed to land.									
82	7 March 2020, 04:33 <sup>[link]</sup>	FA 001.0 B1009.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3.2)	1,977 kg (4,350 lb) <sup>[link]</sup>	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
Last launch of phase 1 of the CRS (contract). Carried <i>Starliner</i> , an <i>ESA</i> platform for testing external payloads onto <i>ISS</i> . <sup>[link]</sup> Originally scheduled to launch on 2 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to scrap out the second stage instead of replacing the faulty part. <sup>[link]</sup> It was SpaceX's 50th successful launch <sup>[edit section: 50th]</sup> booster, the third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.									
83	18 March 2020, 12:09 <sup>[link]</sup>	FA 001.0 B1040.5	KSC, LC-35A	Starlink 5 v1.0 (90 satellites)	15,800 kg (34,400 lb) <sup>[3]</sup>	LEO	SpaceX	Success	Failure (crowd step)

Web scraping with BeautifulSoup

FlightNumber		Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin1A	167.743129	9.047721
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2A	167.743129	9.047721
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2C	167.743129	9.047721
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin3C	167.743129	9.047721
4	6	2010-06-04	Falcon 9	NaN	LEO	CCAFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857

# Data Wrangling

---

- Wrangling Data using API → use *json\_normalize* to convert the response object of the API call to a dataframe
- Incomplete information → utilize API to target other endpoints to gather specific data
- Filter only for data specific to Falcon 9
- Treatment of NULL values → replace NULL values with the mean of attribute / one hot encoding
- “*Outcome*” of the dataset needs to be converted to Classes y (either 0 or 1)

# EDA with Data Visualization

---

- Scatter Point Chart
- Bar Chart
- Line Chart

## Github URL:

*[https://github.com/mariotey/IBM\\_Coursera/blob/master/Week2/IBM-DS0321EN-SkillsNetwork\\_labs\\_module\\_2\\_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb](https://github.com/mariotey/IBM_Coursera/blob/master/Week2/IBM-DS0321EN-SkillsNetwork_labs_module_2_jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb)*

# EDA with SQL

---

- Names of Unique Launch Sites in the Space Mission
- Records where Launch Sites begin with “CCA”
- Total Payload Mass carried by boosters launched by NASA (CRS)
- Average Payload Mass carried by booster ver F9 v1.1
- Date when the first successful landing outcome in the ground pad was achieved
- Names of boosters that have success in drone ships with  $4000 < \text{payload mass} < 6000$
- Total number of successful and failure mission outcomes
- Names of booster\_versions that have carried the maximum payload mass
- Failure landing\_outcomes in drone ship, booster versions, launch\_site for the months in the year 2015.

**Github URL:** [https://github.com/mariotey/IBM\\_Coursera/blob/master/Week2/jupyter-labs-eda-sql-coursera\\_sqllite.ipynb](https://github.com/mariotey/IBM_Coursera/blob/master/Week2/jupyter-labs-eda-sql-coursera_sqllite.ipynb)

# Build an Interactive Map with Folium

---

- Folium.Circle → To mark the successful / failed launches for each site on the map
- Folium.Marker → To label the launch site names
- Folium.PolyLine → To label the calculated distances between a launch site to its proximities

**Github URL:** [https://github.com/mariotey/IBM\\_Coursera/blob/master/Week3/IBM-DS0321EN-SkillsNetwork\\_labs\\_module\\_3\\_lab\\_jupyter\\_launch\\_site\\_location.jupyterlite.ipynb](https://github.com/mariotey/IBM_Coursera/blob/master/Week3/IBM-DS0321EN-SkillsNetwork_labs_module_3_lab_jupyter_launch_site_location.jupyterlite.ipynb)



# Build a Dashboard with Plotly Dash

---

## Input

- Dropdown List → To select Launch Sites
- Range Slider → To select the interested range of Payload Mass

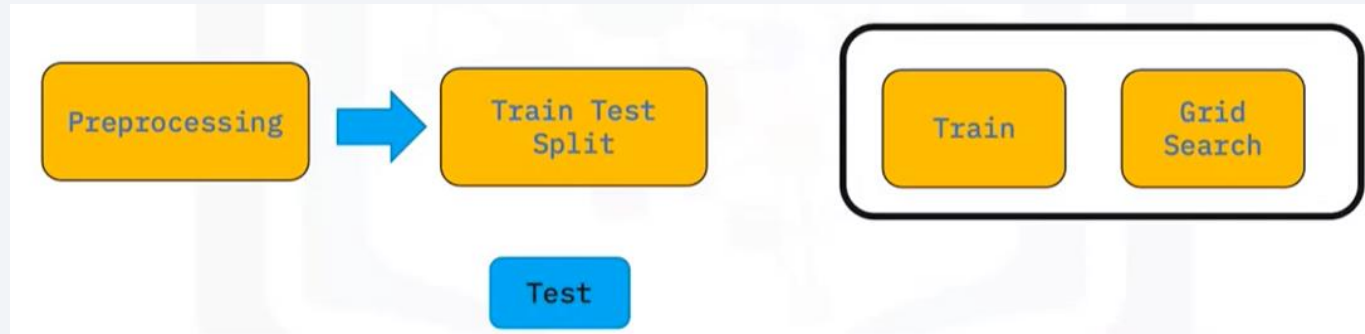
## Output

- Piechart → To show the distribution of successful and failed missions at interested launch sites
- Scatterplot → To show the correlation between the interested range of Payload Mass and the Success at interested launch sites

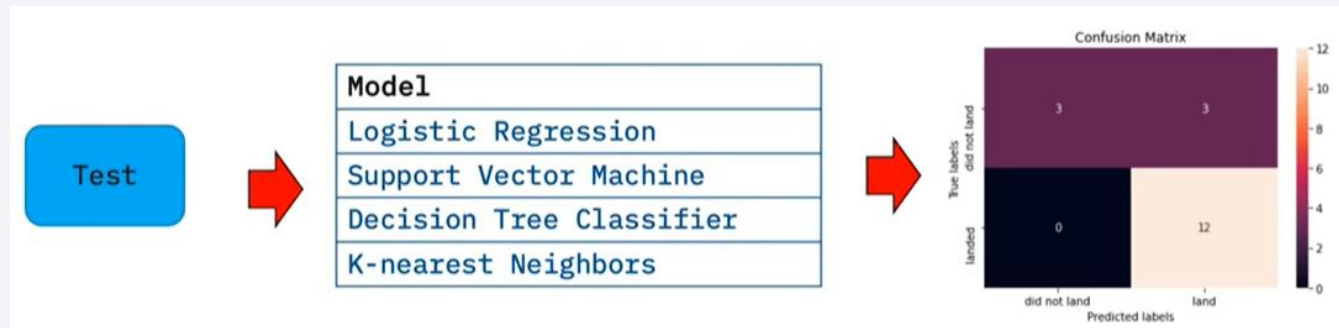
**Github URL:** [https://github.com/mariotey/IBM\\_Coursera/blob/master/Week3/spacex\\_dash\\_app.py](https://github.com/mariotey/IBM_Coursera/blob/master/Week3/spacex_dash_app.py)

# Predictive Analysis (Classification)

- Predict whether first stage of Falcon 9 will land successfully



- Determine the model with the best accuracy



**Github URL:** [https://github.com/mariotey/IBM\\_Coursera/blob/master/Week4/IBM-DS0321EN-SkillsNetwork\\_labs\\_module\\_4\\_SpaceX\\_Machine\\_Learning\\_Prediction\\_Part\\_5.jupyterlite.ipynb](https://github.com/mariotey/IBM_Coursera/blob/master/Week4/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb)

# Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results



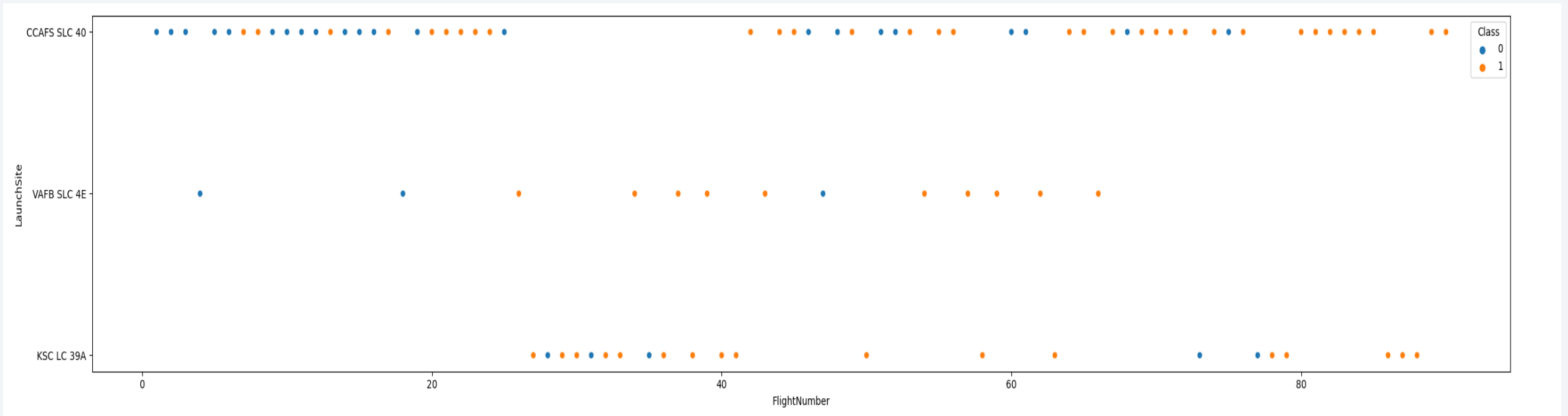
The background of the slide is an abstract composition. It features a dark blue base color. Overlaid on this are numerous diagonal streaks in shades of red and cyan. A faint, light blue grid pattern is also visible, particularly in the lower-left quadrant. The overall effect is dynamic and technological.

Section 2

# Insights drawn from EDA



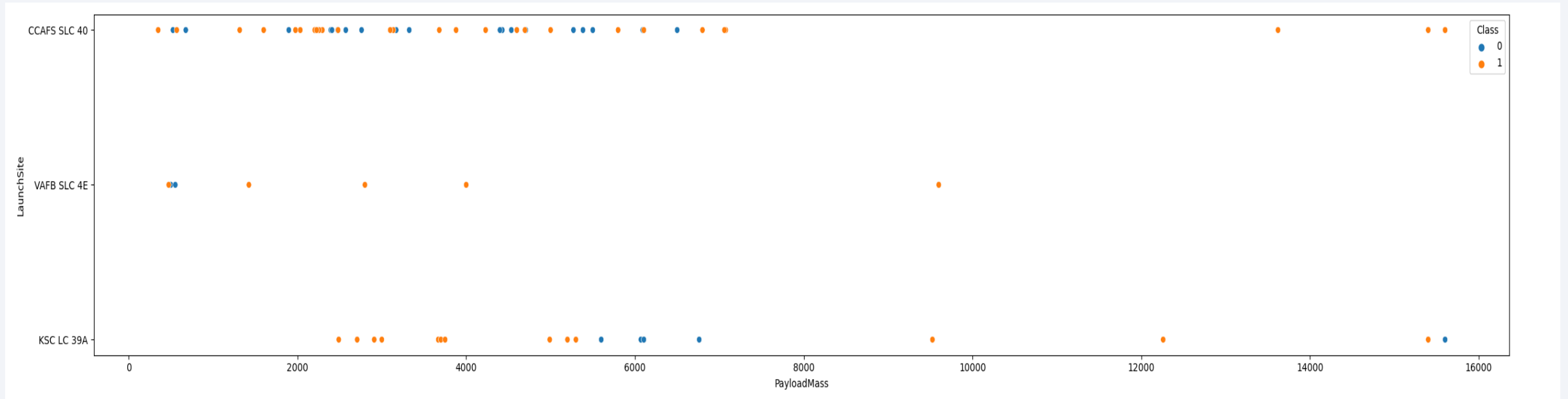
# Flight Number vs. Launch Site



- For CCAFS SLC 40, flight numbers above 25 seem to be more successful than flight numbers below 25
- There are more missions conducted at CCAFS SLC 40 than VAFB SLC 4E and KSC LC39A
- There are too few missions conducted at VAFB SLC 4E and KSC LC39A to determine a trend

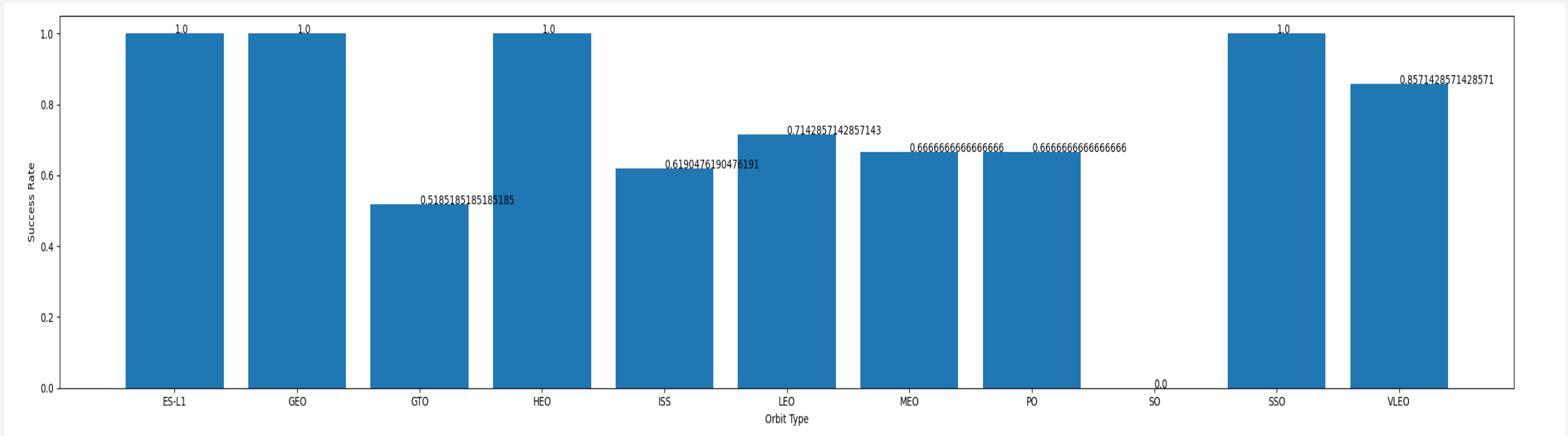


# Payload vs. Launch Site



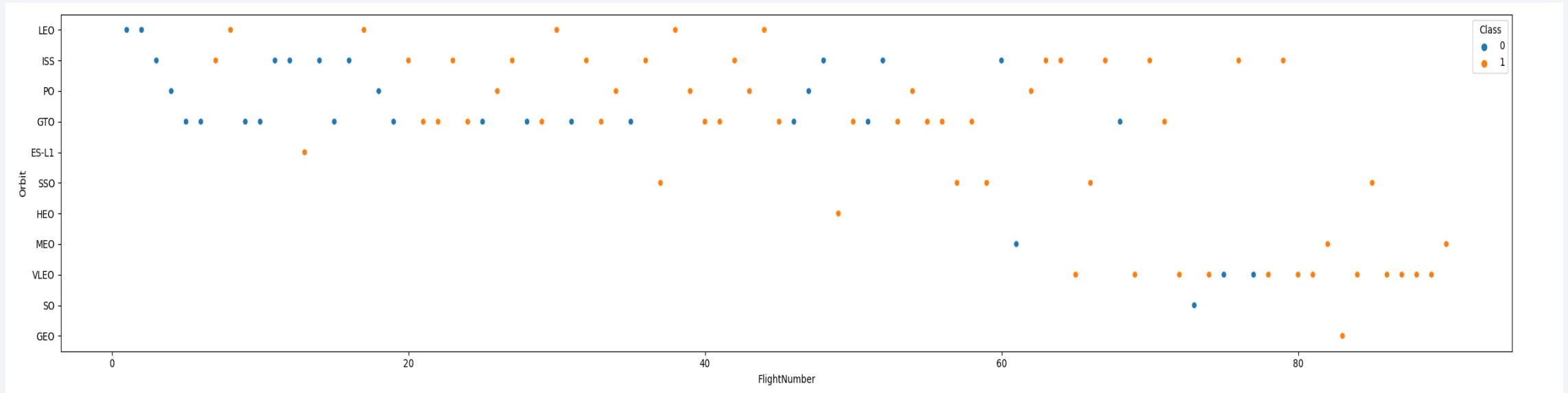
- Missions with payload below 8000kg were mostly held at CCAFS SLC 40. There does not seem to be any trend
- Missions are most successful when held at VAFB SLC 4E
- Missions held at KSC LC39A were mostly successful with the exception of payloads ranging between 5500 to 7000kg

# Success Rate vs. Orbit Type



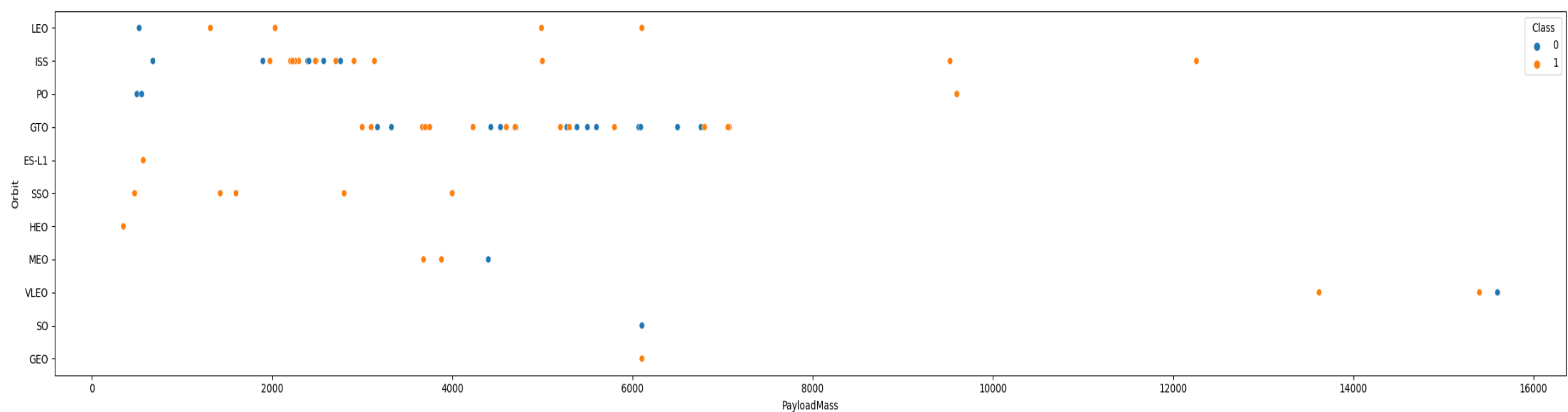
- Orbit Type of ES-L1, GEO, HEO and SSO had a success rate of 100%
- Orbit Type of SO had a success rate of 0%. Either all missions done in this orbit type failed or there are no missions done in this orbit type.

# Flight Number vs. Orbit Type



- Orbit Type LEO, ISS, PO, GTO and ES-L1 mostly handle flight numbers below 40
- Orbit Type SSO, HEO, MEO, VLEO, SO and GEO handle flight numbers above 40
- Orbit Type LEO successes appear to be related to the number of flights while GTO does not seem to exhibit any trend.

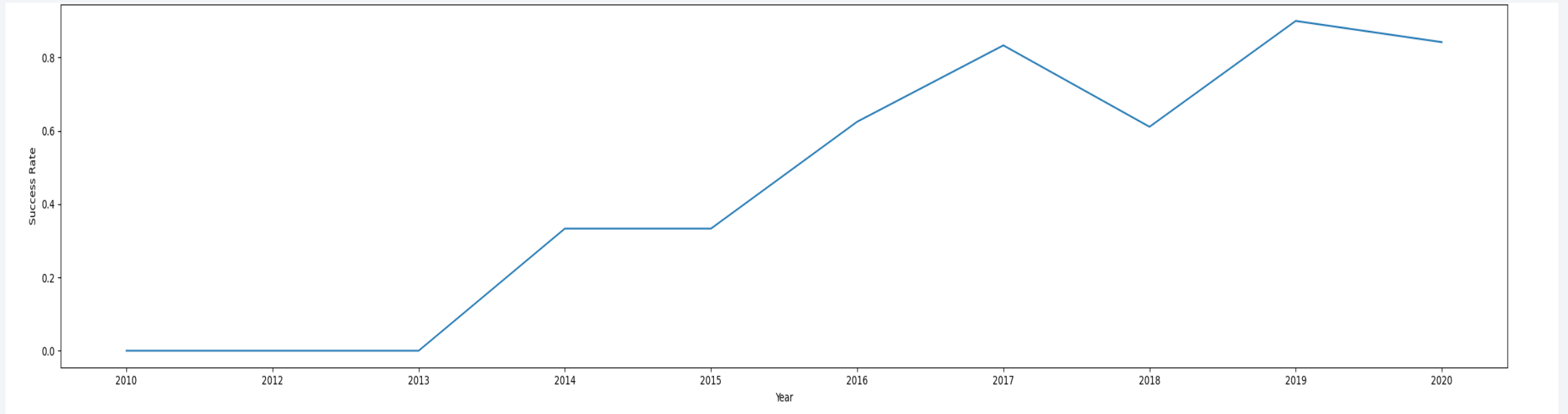
# Payload vs. Orbit Type



- For Orbit type LEO, ISS and PO, they seem to have more successful missions when handling heavier payloads
- For Orbit type GTO, it has a mixture of successful and failed missions which makes the identification of a trend difficult.
- For the rest of the Orbit types, they seemed rather consistent throughout various payloads

# Launch Success Yearly Trend

---



- Launch missions from 2013 to 2020 mostly exhibit an increasing trend of success rate.
- In 2017 and 2019, the success rate of Launch Missions declined.



# All Launch Site Names

---

In [14]: `%sql SELECT DISTINCT(Launch_Site) FROM SPACEXTBL`

`* sqlite:///my_data1.db`

Done.

Out[14]: **Launch\_Site**

CCAFS LC-40

VAFB SLC-4E

KSC LC-39A

CCAFS SLC-40

- Names of unique launches in the space mission

# Launch Site Names Begin with 'CCA'

In [30]:

```
%sql SELECT * FROM SPACEXTBL \
      WHERE Launch_Site like "CCA%" LIMIT 5
```

```
* sqlite:///my_data1.db
```

Done.

Out[30]:

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
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- 5 records where launch sites begin with the string “CCA”

# Total Payload Mass

---

```
In [32]: %sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL \
        WHERE Customer like "NASA (CRS)"

* sqlite:///my_data1.db
Done.

Out[32]: SUM(PAYLOAD_MASS__KG_)
        45596
```

- Total payload mass carried by boosters launched by NASA (CRS)

# Average Payload Mass by F9 v1.1

---

```
In [42]: # %sql SELECT * FROM SPACEXTBL \
#         WHERE Booster_Version like "F9 v1.1"

%sql SELECT AVG(PAYLOAD_MASS_KG_) FROM SPACEXTBL \
      WHERE Booster_Version like "F9 v1.1"

* sqlite:///my_data1.db
Done.

Out[42]: AVG(PAYLOAD_MASS_KG_)
          2928.4
```

- Average payload mass carried by booster version F9 v1.1

# First Successful Ground Landing Date

---

```
In [60]: %%sql SELECT min(Date) FROM SPACEXTBL
          WHERE "Landing_Outcome" like "Success%ground pad%"

* sqlite:///my_data1.db
Done.

Out[60]: min(Date)
          01-05-2017
```

- Date when the first successful landing outcome in the ground pad was achieved.
- The date shown in the output is in the format of DD-MM-YYYY



## Successful Drone Ship Landing with Payload between 4000 and 6000

---

```
In [67]: %%sql SELECT Booster_Version FROM SPACEXTBL
          WHERE "Landing_Outcome" like "Success%drone ship%"
          AND PAYLOAD_MASS__KG_ > 4000
          AND PAYLOAD_MASS__KG_ < 6000
```

```
* sqlite:///my_data1.db
```

```
Done.
```

```
Out[67]: Booster_Version
```

```
F9 FT B1022
```

```
F9 FT B1026
```

```
F9 FT B1021.2
```

```
F9 FT B1031.2
```

- Names of the boosters which have success in drone ships and have payload mass greater than 4000 but less than 6000

# Total Number of Successful and Failure Mission Outcomes

```
[14]: %%sql

SELECT COUNT(Mission_Outcome) FROM SPACEXTBL
      WHERE Mission_Outcome like "Success%"

-- SELECT COUNT(Mission_Outcome) FROM SPACEXTBL
--      WHERE Mission_Outcome like "Fail%"
```

```
* sqlite:///my_data1.db
Done.
```

```
[14]: COUNT(Mission_Outcome)
      _____
              100
```

```
[15]: %%sql

-- SELECT COUNT(Mission_Outcome) FROM SPACEXTBL
--      WHERE Mission_Outcome like "Success%"

SELECT COUNT(Mission_Outcome) FROM SPACEXTBL
      WHERE Mission_Outcome like "Fail%"
```

```
* sqlite:///my_data1.db
Done.
```

```
[15]: COUNT(Mission_Outcome)
      _____
              1
```

- 100 successful missions and 1 failed mission

# Boosters Carried Maximum Payload

---

```
In [80]: %sql SELECT Booster_Version, MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL
```

```
* sqlite:///my_data1.db  
Done.
```

```
Out[80]: Booster_Version  MAX(PAYLOAD_MASS__KG_)  
-----  
F9 B5 B1048.4                15600
```

- Name of the booster version that carried the maximum payload mass

# 2015 Launch Records

In [129...

```
%%sql
```

```
SELECT substr(Date, 4, 2) as Month, "Landing _Outcome", Booster_Version, Launch_Site FROM SPACEXTBL  
WHERE "Landing _Outcome" like "Fail%drone ship%"  
AND substr(Date, 7, 4) = '2015'
```

```
* sqlite:///my_data1.db
```

Done.

Out[129...

	Month	Landing _Outcome	Booster_Version	Launch_Site
	01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
	04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

- Records that display the months, booster version and launch sites of failed missions in the year 2015

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

---

[22]: %%sql

```
SELECT substr(Date, 7, 4) as Year, COUNT(*) as Count FROM SPACEXTBL
WHERE "Landing_Outcome" like "Success%"
AND CAST(substr(Date, 7, 4) as int) < 2018
AND CAST(substr(Date, 7, 4) as int) > 2010
GROUP By CAST(substr(Date, 7, 4) as int)
ORDER By COUNT(*) DESC
```

\* sqlite:///my\_data1.db

Done.

[22]:

Year	Count
2017	14
2016	5
2015	1

- Count of successful land outcomes between 2010-06-04 and 2017-03-20 in descending order

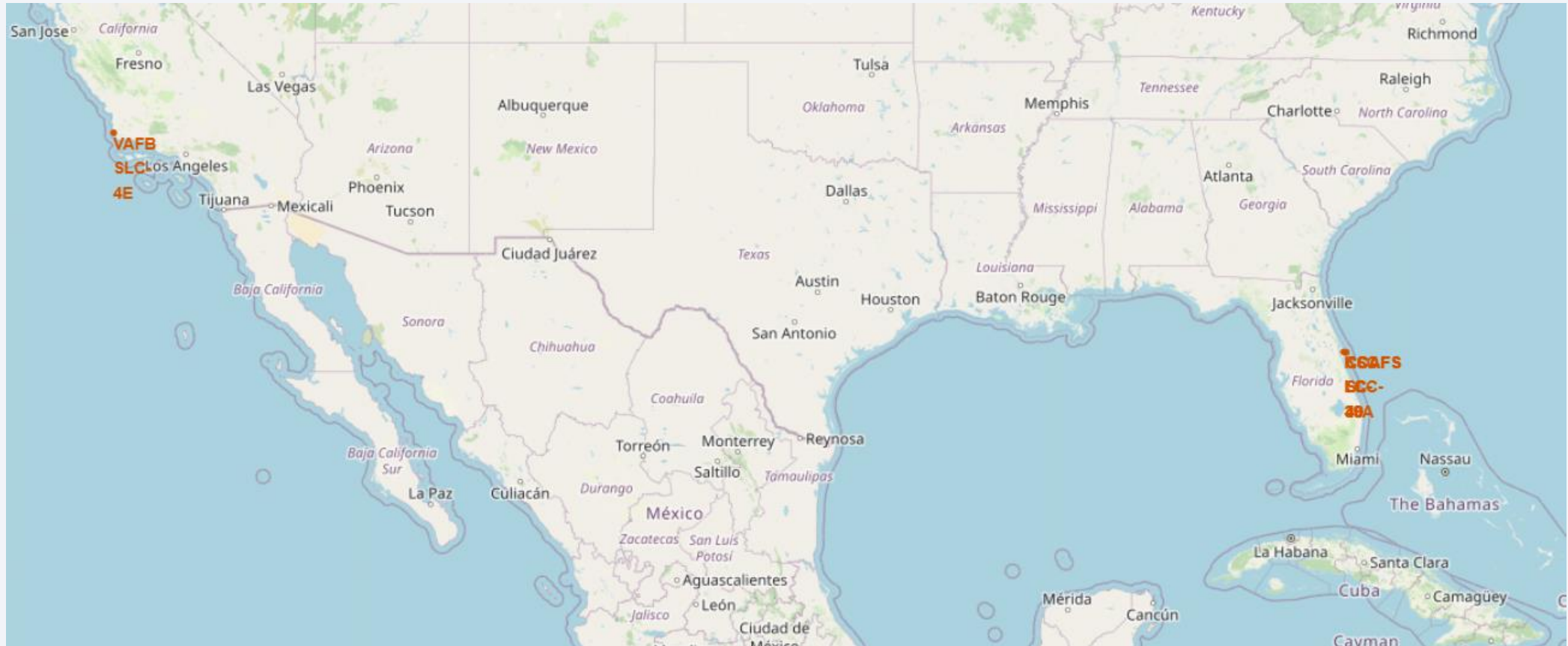
A satellite view of Earth from space, showing the curvature of the planet and the glow of city lights at night. The background is a deep blue gradient.

Section 3

# Launch Sites Proximities Analysis

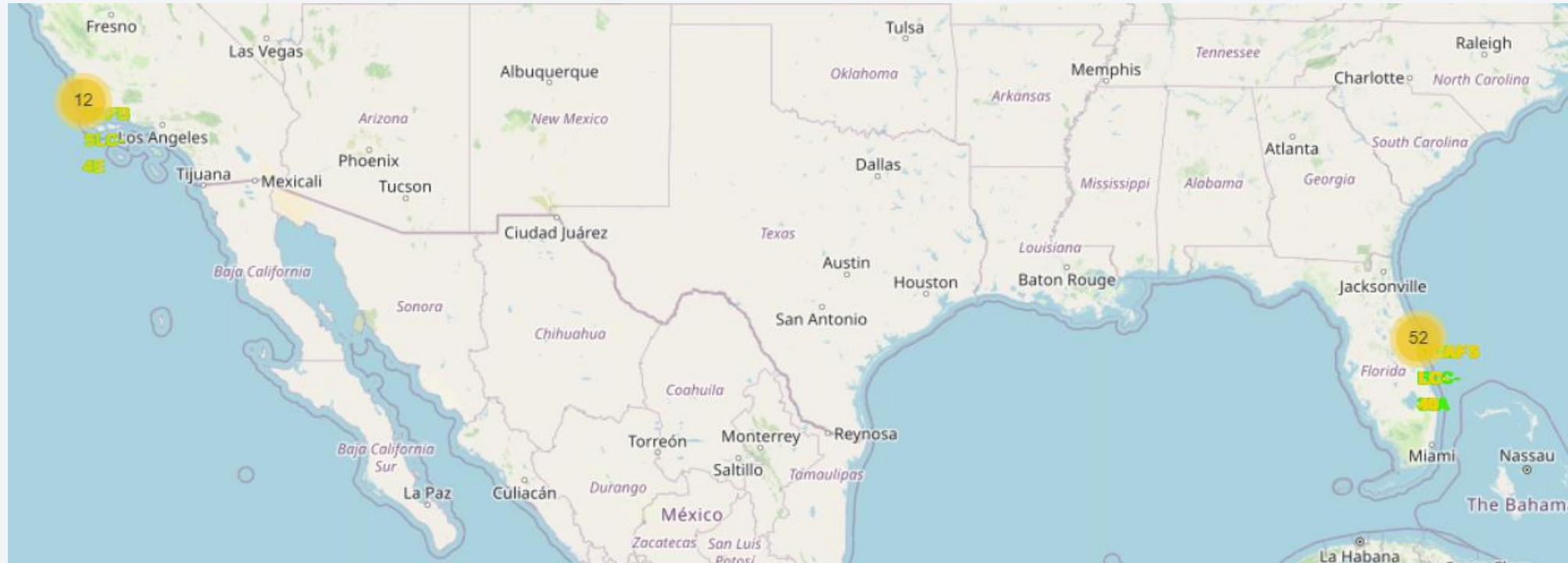


# Location of Launch Sites



- Locations are identified in the red dot and labelled with the names of the launch sites

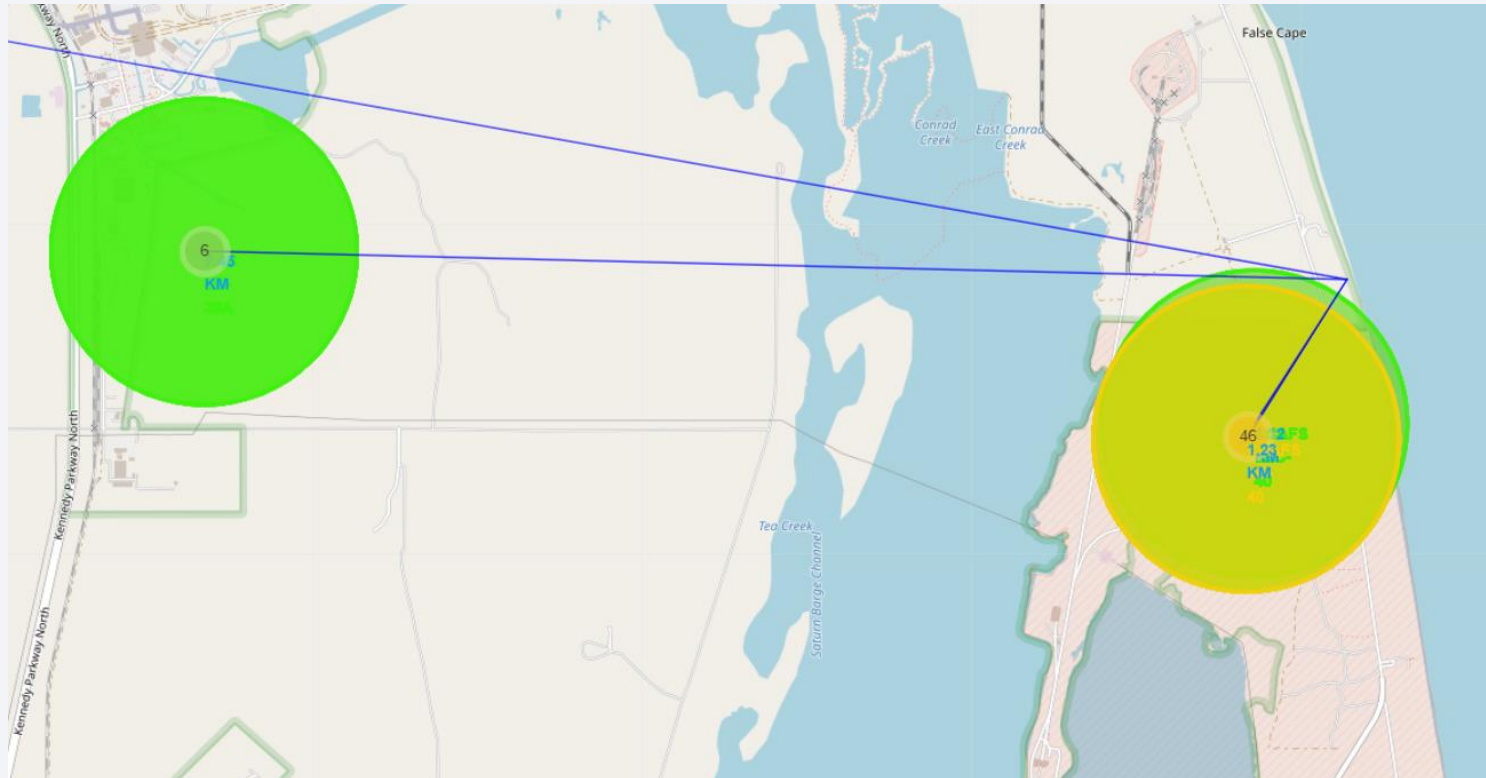
# Locations of Launch Sites with Successful / Failed Missions



- Launch Sites with Successful missions are in Green
- Launches with Failed missions are in Yellow

# Proximity of Launch Sites to Specific Location

---



- Determining the proximity of launch sites to [28.57164, -80.57064]
- The longer the blue the further the launch sites are from the point of interest



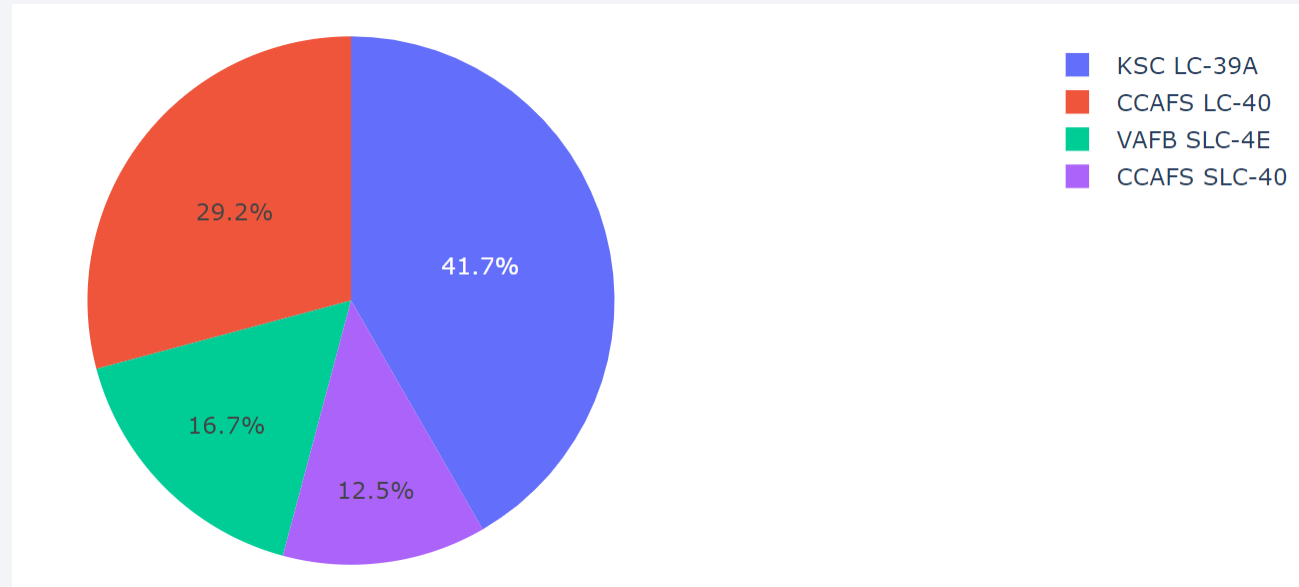
The background of the slide is a close-up, artistic photograph of a printed circuit board (PCB). The board is dark, and the intricate circuit traces are highlighted in a vibrant, glowing red. Numerous small, cylindrical components, likely capacitors or resistors, are visible, some of which also appear to be glowing. The lighting creates a sense of depth and technological sophistication.

Section 4

# Build a Dashboard with Plotly Dash

# Total Successful Launches by Launch Sites

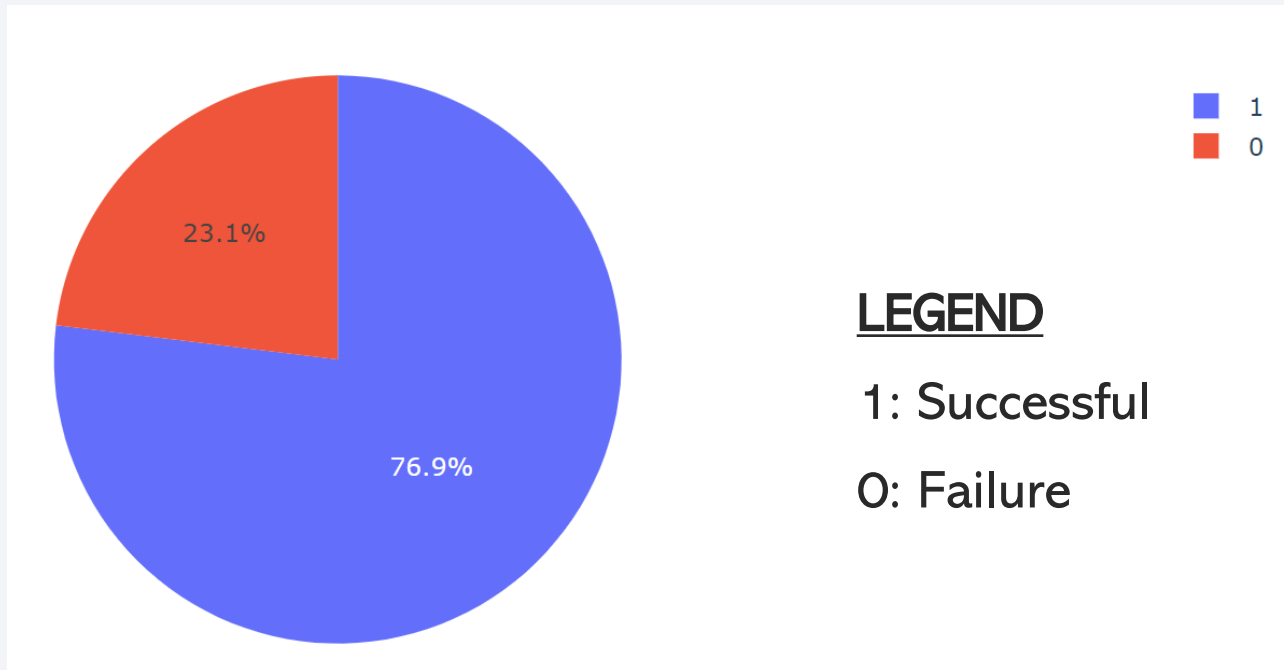
---



- KSC LC-39A contributes the most to successful launches while CCAFS SLC-40 contributes the least to successful launches

# Distribution of Successful Launches at KSC LC-39A

---



- Majority of the launches are successful



# Payload vs. Launch Outcome for all Sites, with Different Payloads



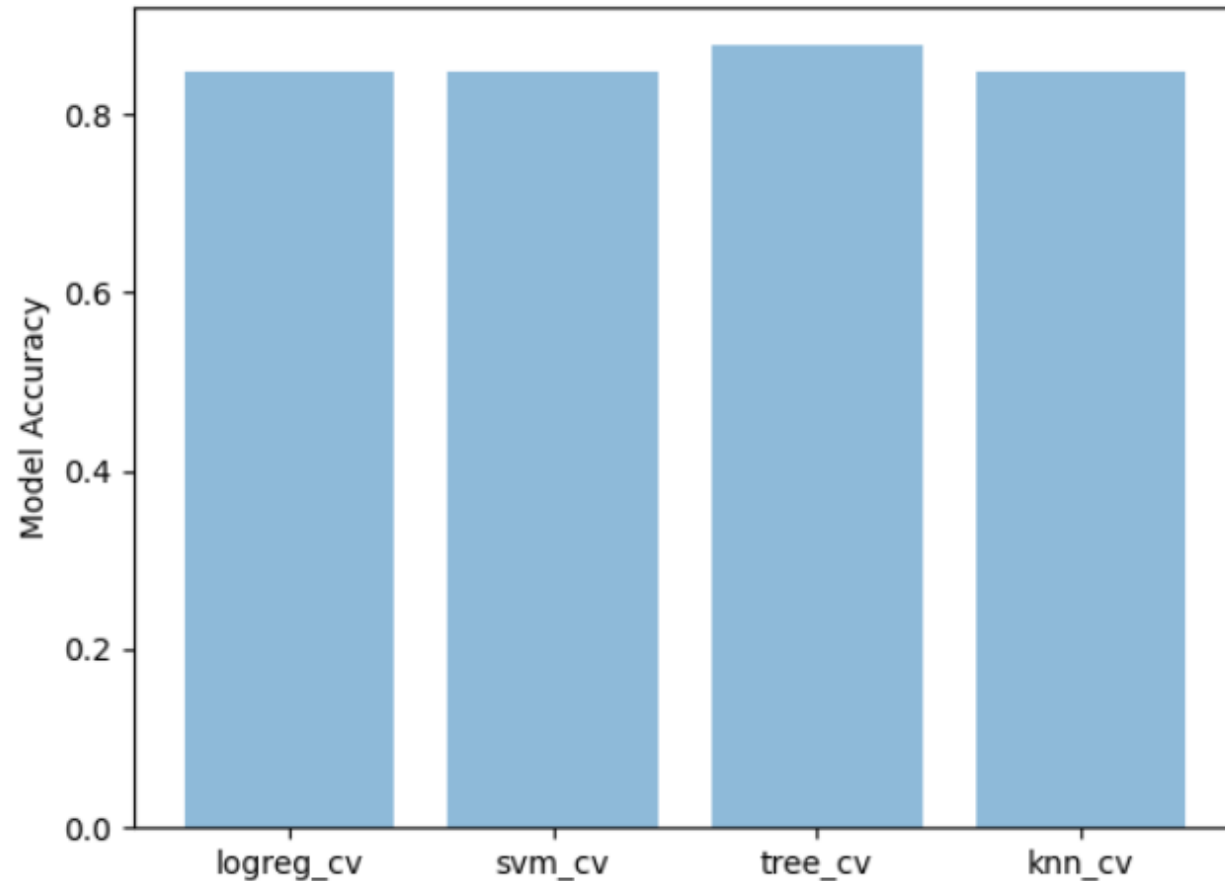
- FT Booster seems to be very reliable; being mostly successful across various payloads
- V1.1 seems to be unreliable with the least successful launches and contributing the most to failed missions across various payloads

Section 5

# Predictive Analysis (Classification)

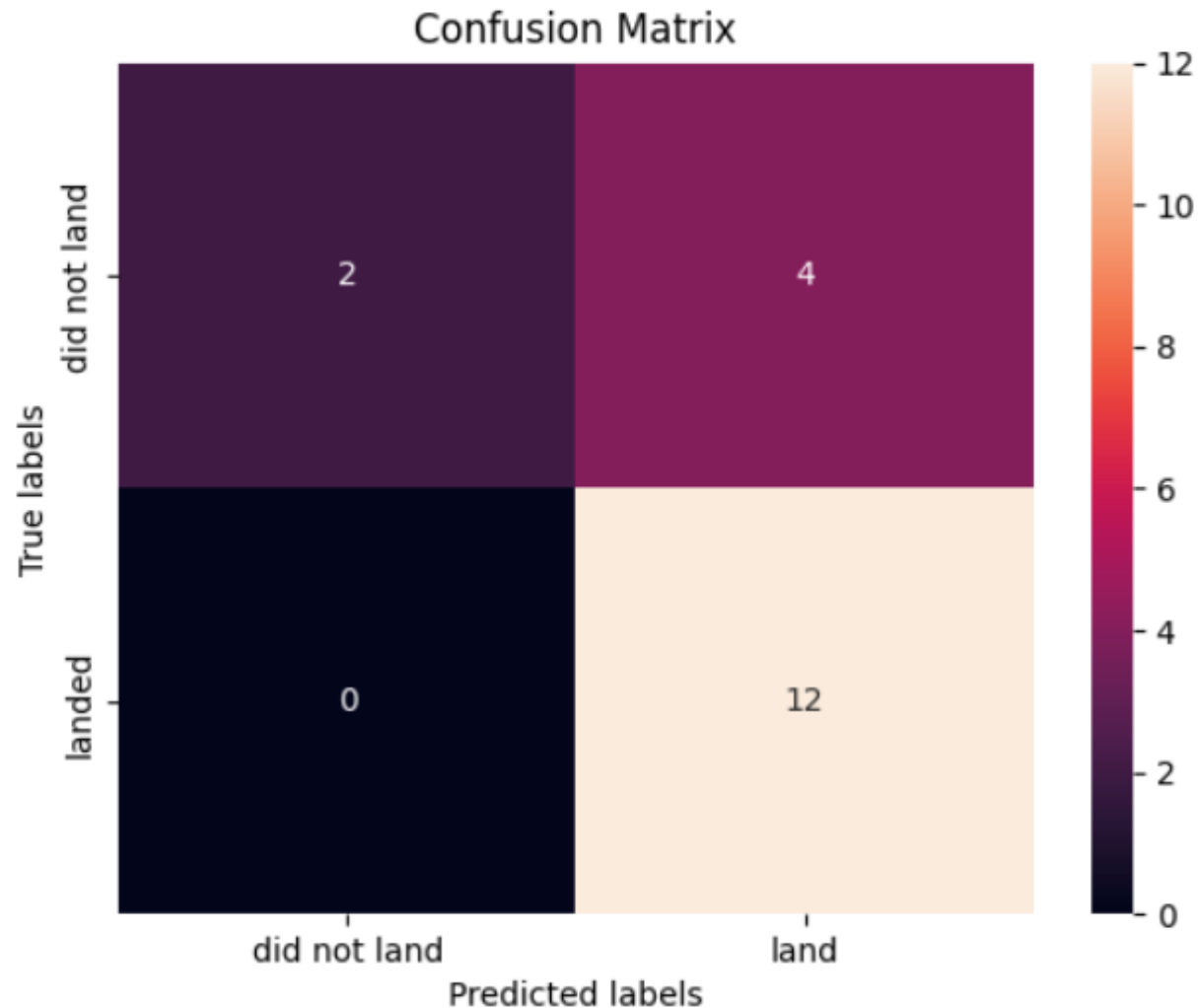
# Classification Accuracy

[0.8464285714285713, 0.8482142857142856, 0.8767857142857143, 0.8482142857142858]



- Decision tree classifier has the highest accuracy of 87.7%
- Log Regression classifier has the lowest accuracy of 84.6%

# Confusion Matrix



- Uses a decision tree classifier object
- Selected best model is relatively accurate in predicting and has a test error count of 4

# Conclusions

---

- Missions with higher payloads at CCAFS SLC 40 are very likely to succeed
- Missions with Orbit Type LEO, ISS and PO are very likely to succeed
- Majority of the missions are held near Florida
- Booster v1.1 needs to be reviewed for its reliability
- The first stage of Falcon 9 is very likely to land successfully

Thank you!

