IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

James Knopp
5/29/2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - EDA with data visualization
  - EDA with SQL
  - Map building with Folium
  - Building a dashboard with Plotly Dash
  - Predictive analysis (Classification)
- Summary of all results
  - EDA Results
  - Interactive analytics
  - Predictive analytics

Introduction

- This project focuses on predicting if the Falcon 9 first stage will land successfully.
- Goal: Using previous launch data, are we able to build a model that can predict whether a first stage will land successfully?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected using a REST API and web scraping from Wikipedia
- Perform data wrangling
  - Data was one-hot encoded for machine learning classification models
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - LRR, KNN, SVM, and Decision Tree have been built and evaluated for best performance

Data Collection

- Space X launch data is gathered from the SpaceX rest API
- API data includes: launches, rocket information, payload mass, launch specifications, landing specifications, and launch outcome
- Falcon 9 launch data was scraped from the below table from Wikipedia

[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)

Data Collection – SpaceX API

- 1) Get URL of API
- 2) Create response object
- 3) Static Response
- 4) Normalize json to data frame

- https://github.com/jimknopp2/Space-X-IBM-Capstone/blob/master/Space%20X%20Data%20Collection%20-%20API.ipynb

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

```
response = requests.get(spacex_url)
```

```
static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json'
```

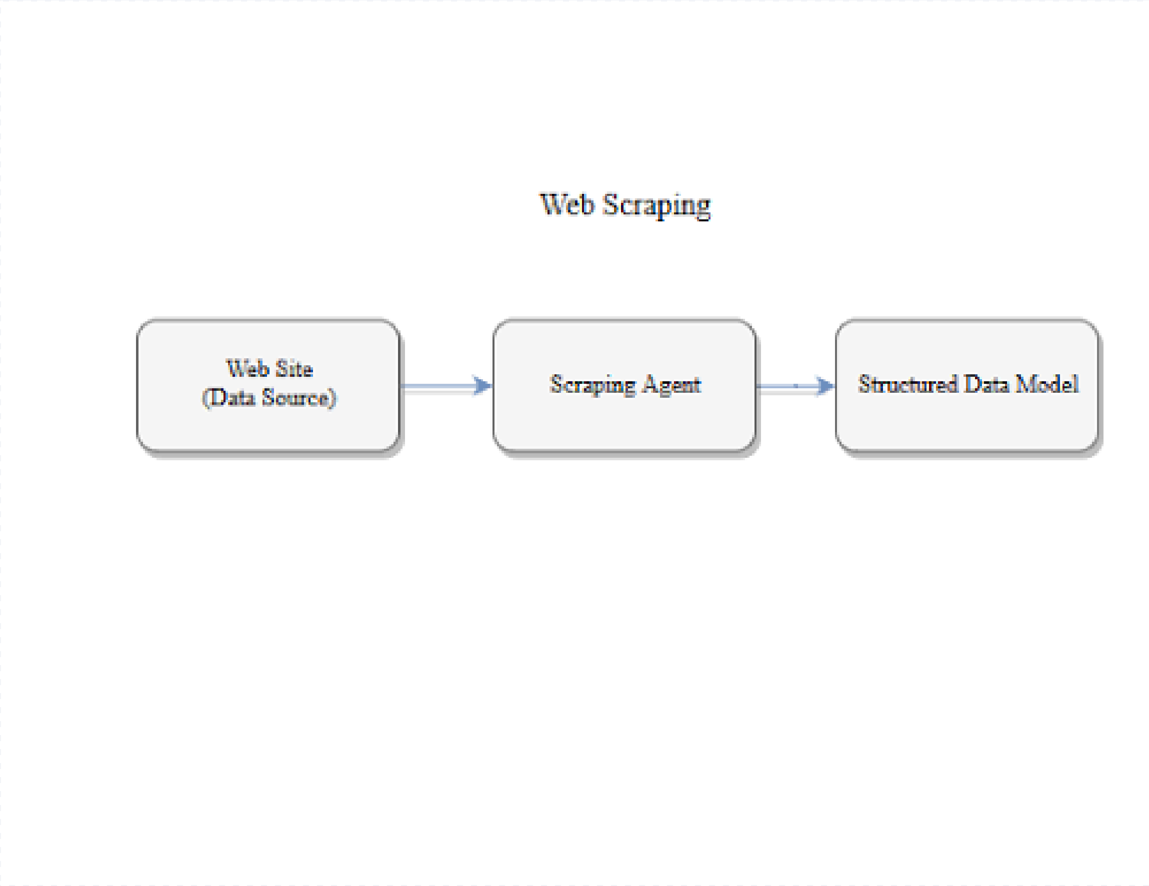
```
# Use json_normalize meethod to convert the json result into a dataframe
data = pd.json_normalize(response.json())
```

```
# Get the head of the dataframe
data.head()
```

Data Collection - Scraping

- 1) Create response request from URL
- 2) Create BeautifulSoup object
- 3) Create a dataframe by parsing the HTML tables

- https://github.com/jimknopp2/Space-X-IBM-Capstone/blob/master/Space%20X%20Data%20Collection%20-%20Web%20Scraping.ipynb



Data Wrangling

- 1) Identify missing (NULL) values
- 2) Identify values of landing outcomes
- 3) Group landing outcomes into "good" and "bad" categories
- 4) Create landing outcome column using a loop and then append column to dataframe
- https://github.com/jimknopp2/Space-X-IBM-Capstone/blob/master/SpaceX%20Data%20Wrangling.ipynb

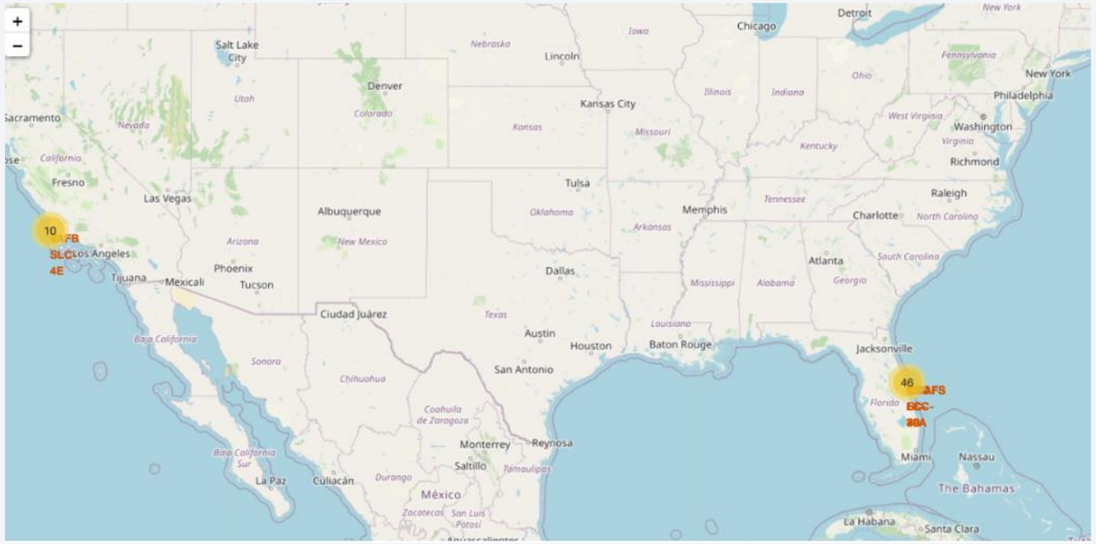
EDA with Data Visualization

- Scatterplot of flight number vs payload mass to see if there is any relationship between flight number and payload mass
- Scatterplot of flight number vs launch site to see if there is any relationship between flight number vs launch site
- Scatterplot of payload mass vs launch site to see if there is any relationship between payload mass and launch site
- Bar chart of success rate for orbit type to see what orbits were most successful
- Line chart plotting success rate per year
- https://github.com/jimknopp2/Space-X-IBM-Capstone/blob/master/Space%20X%20_Visualization%20with%20Pandas%20and%20Matplotlib.ipynb

EDA with SQL

- Queries were run on the Space X dataset to analyze:
- - Unique launch sites
- - Total payload mass carried by boosters
- - Dates of successful landings
- - Number of successful and failed landings
- https://github.com/jimknopp2/Space-X-IBM-Capstone

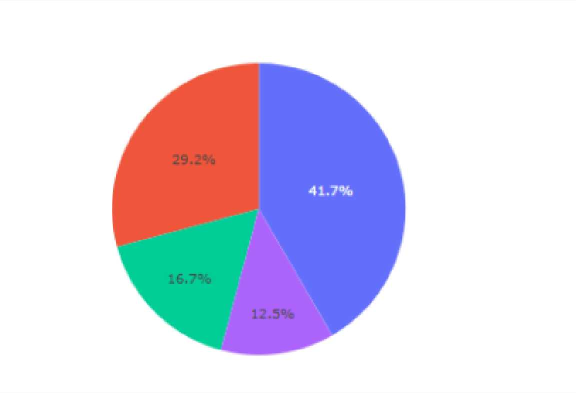
Build an Interactive Map with Folium



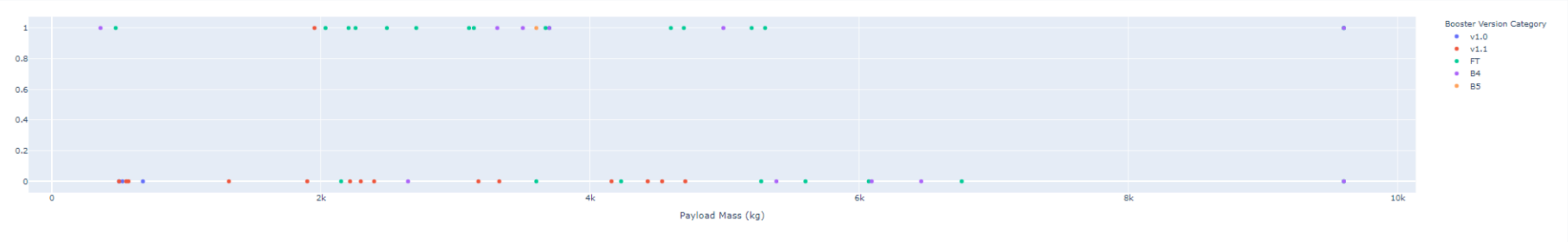
- Map markers have been added to the map to find an optimal location to build a launch site
- https://github.com/jimknopp2/Space-X-IBM-Capstone/blob/master/SpaceX%20Folium.ipynb

Build a Dashboard with Plotly Dash

Pie Chart showing successful launches by site



Scatterplot of payload mass vs success for all sites



- https://github.com/jimknopp2/Space-X-IBM-Capstone/blob/master/SPACEX.PY

Predictive Analysis (Classification)

1. Standardize X variable set
2. Split data into training and testing datasets
3. Create logistic regression, SVM, KNN, and decision tree models
4. Evaluate models using score method and confusion matrix

- https://github.com/jimknopp2/Space-X-IBM-Capstone/blob/master/SpaceX%20-%20Machine%20Learning%20Classification.ipynb

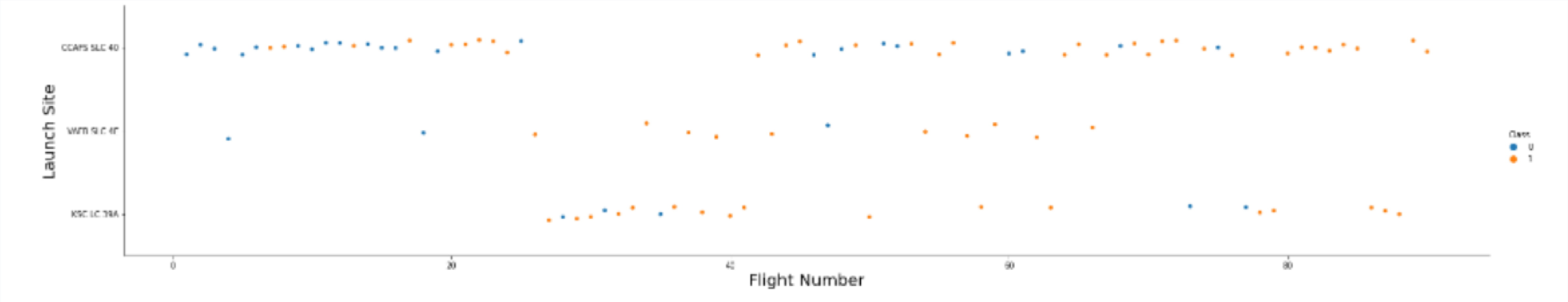
Results

- The SVM, KNN, and Logistic Regression models have the best accuracy
- Low-weighted payloads perform better than high-weighted payloads
- KSC LC 39A was the most successful launch sited
- Orbit GEO, HEO, SSO, ES L1 has the best success rate

Section 2

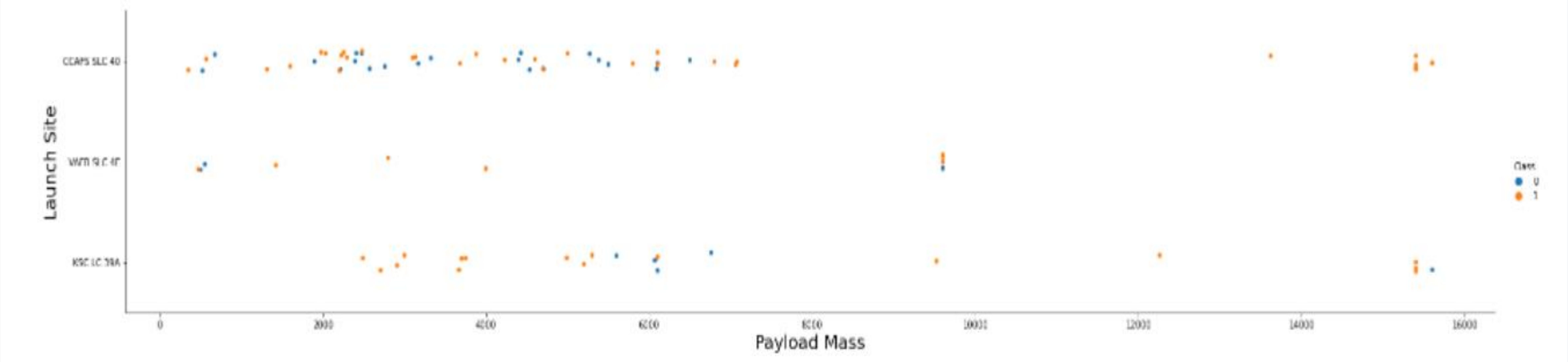
Insights drawn from EDA

Flight Number vs. Launch Site



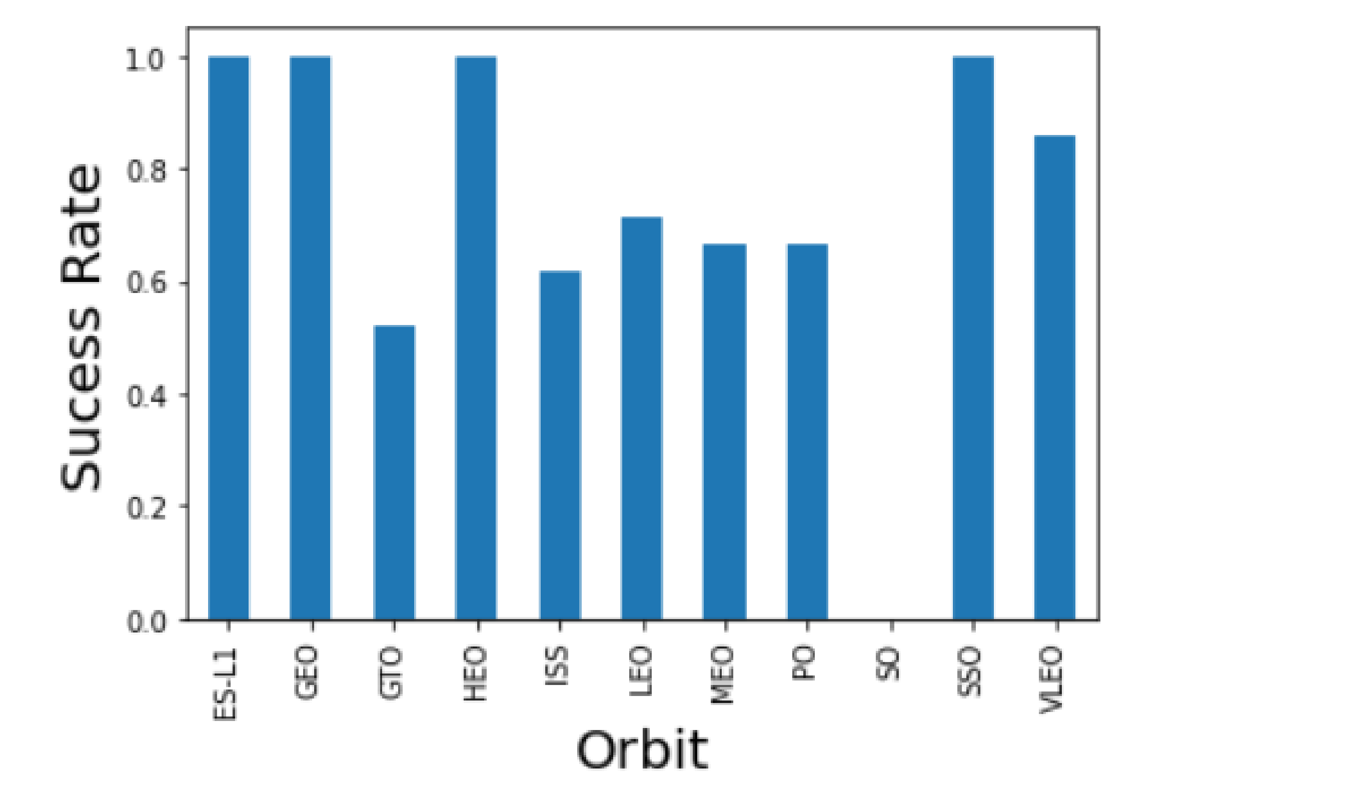
- Launches from site CCAFS SLC 40 have more successful landings than other launch sites

Payload vs. Launch Site



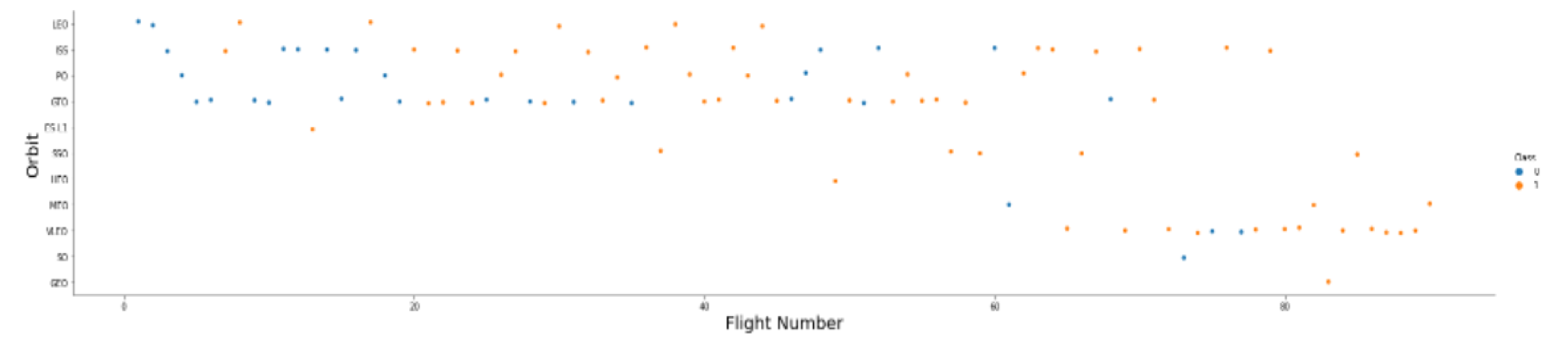
- Most lower mass payloads are launched from site CCAFS SLC 40

Success Rate vs. Orbit Type



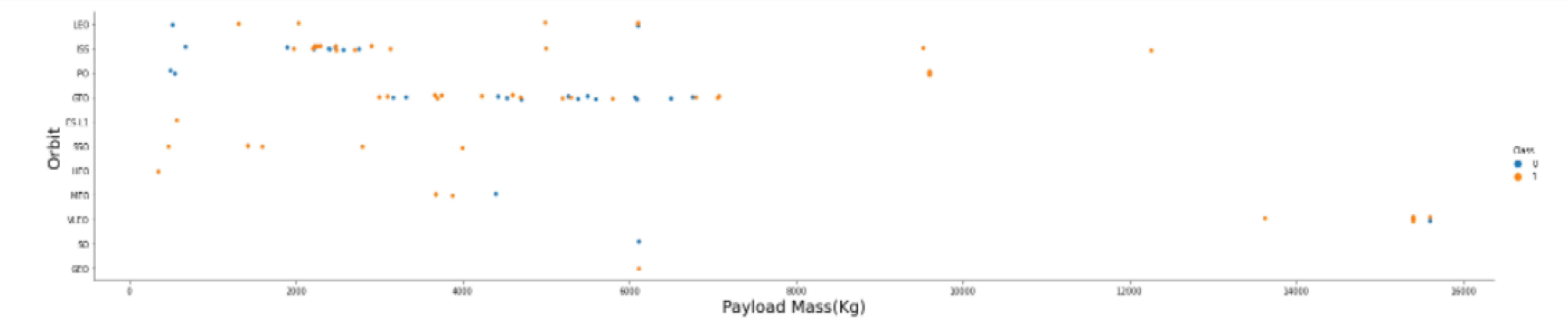
- Orbits ES-L1, GEO, HEO, and SSO have highest success rate

Flight Number vs. Orbit Type



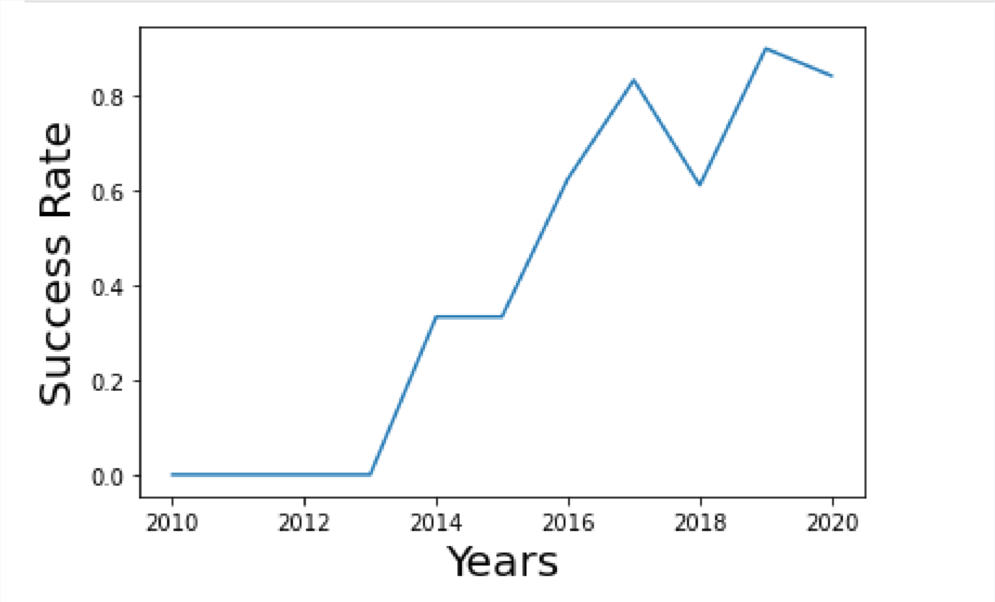
- Most of the latest flight launches are orbit VLEO

Payload vs. Orbit Type



- Strong correlation between ISS orbit and payload mass around 2,000

Launch Success Yearly Trend



- The average success rate increases over time

All Launch Site Names

Query:

```
select Unique(LAUNCH_SITE) from SPACEXTBL;
```

Result:

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

Launch Site Names Begin with 'CCA'

Query:

```
select LAUNCH_SITE from SPACEXTBL
where (LAUNCH_SITE) like 'CCA%' limit 5;
```

Result:

LAUNCH_SITE
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

Total Payload Mass

Query:

```
select sum(PAYLOAD_MASS__KG_) as payloadmass from SPACEXTBL
where (CUSTOMER)='NASA (CRS)';
```

Result:

PAYLOADMASS
45596

Average Payload Mass by F9 v1.1

Query:

```sql
select sum(PAYLOAD_MASS__KG_) as payloadmass from SPACEXTBL
where (BOOSTER_VERSION)='F9 v1.1';

```

Result:

PAYLOADMASS
14642

First Successful Ground Landing Date

Query:

```
select min(DATE) from SPACEXTBL
where (MISSION_OUTCOME)='Success';
```

Result:

1
2010-06-04

Successful Drone Ship Landing with Payload between 4000 and 6000

Query:

```
select BOOSTER_VERSION from SPACEXTBL
where (LANDING__OUTCOME)='Success (drone ship)'
and (PAYLOAD_MASS__KG_) > 4000 and (PAYLOAD_MASS__KG_) < 6000;
```

Result:

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

Total Number of Successful and Failure Mission Outcomes

Query:

```
select count(MISSION_OUTCOME) from SPACEXTBL
where MISSION_OUTCOME='Success'
or MISSION_OUTCOME='Failure (in flight)';
```

Result:

Result set 1

Find

1
100

Boosters Carried Maximum Payload

Query:

```
select BOOSTER_VERSION from SPACEXTBL
where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL);
```

Result:

BOOSTER_VERSION
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5

2015 Launch Records

Query:

```
select LANDING__OUTCOME, BOOSTER_VERSION, LAUNCH_SITE from SPACEXTBL
where (LANDING__OUTCOME)='Failure (drone ship)' and DATE like '2015%';
```

Result:

Result set 1

LANDING__OUTCOME	BOOSTER_VERSION
Failure (drone ship)	F9 v1.1 B1012
Failure (drone ship)	F9 v1.1 B1015

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

Query:

```sql
SELECT LANDING__OUTCOME FROM SPACEXTBL
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' ORDER BY DATE DESC;
```

Result:

LANDING__OUTCOME
No attempt
Success (ground pad)
Success (drone ship)
Success (drone ship)
Success (ground pad)
Failure (drone ship)
Success (drone ship)
Success (drone ship)
Success (drone ship)
Failure (drone ship)

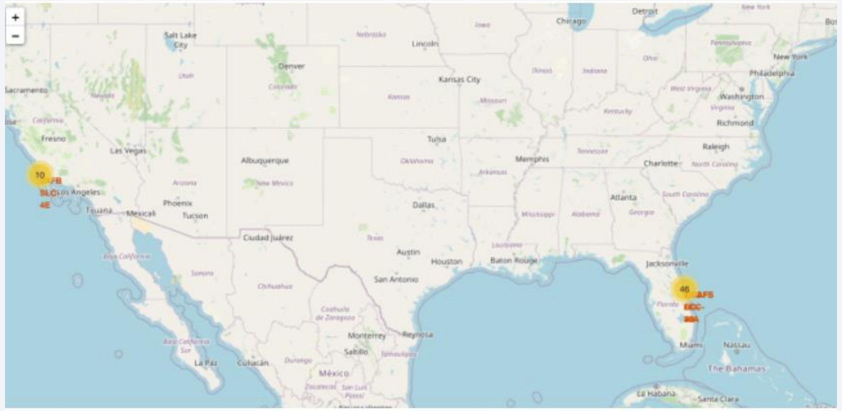
Section 3

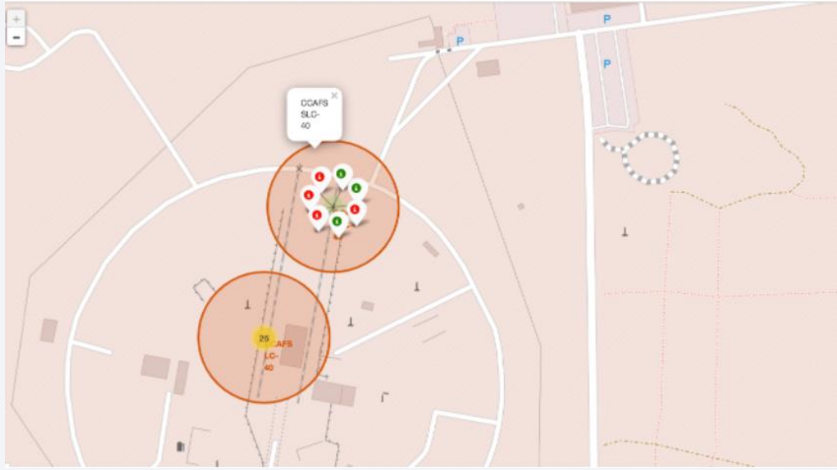
Launch Sites Proximities Analysis

Folium – All Launch Sites

- As seen above, the launch sites are in proximity to the equator and the coast. This makes sense as it takes less fuel to get into space from the equator due to the physics of Earth's rotation. The launch sites in close proximity to the coast are also logical for safety reasons

Folium – success/failed launches for all launch sites





- From the color-labeled markers in marker clusters, you should be able to easily identify which launch sites have relatively high success rates

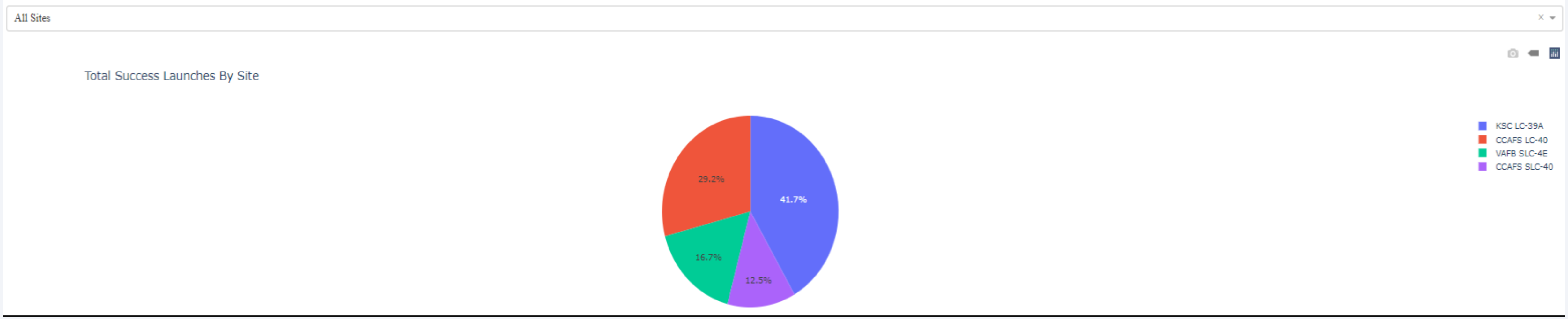
Folium – distances between launch sites

- Above is the distance between launch sites

Section 4

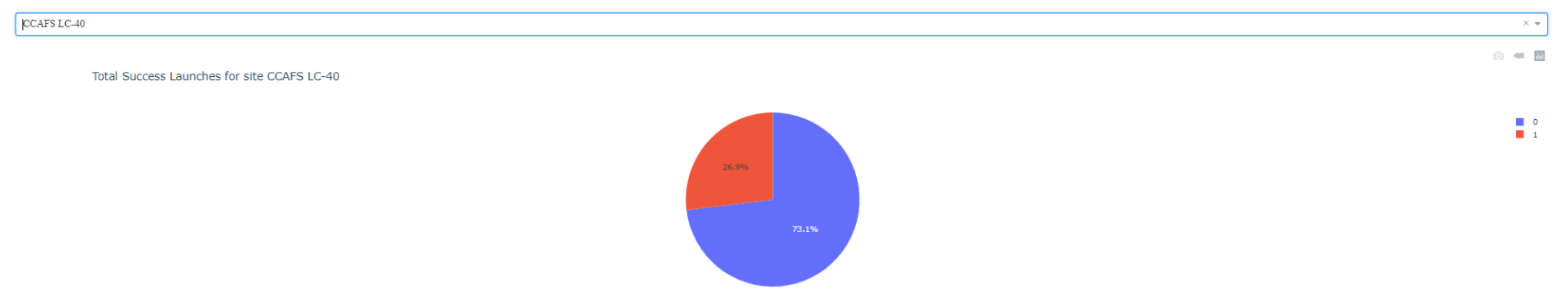
Build a Dashboard with Plotly Dash

Launch Success For All Launch Sites



KSC LC-39A had the highest number of successful launches out of all launch sites

Launch Site With Highest Launch Success Ratio



Launch site CCAFS LC-40 has a launch success rate of 73.1% and a launch failure rate of 26.9%

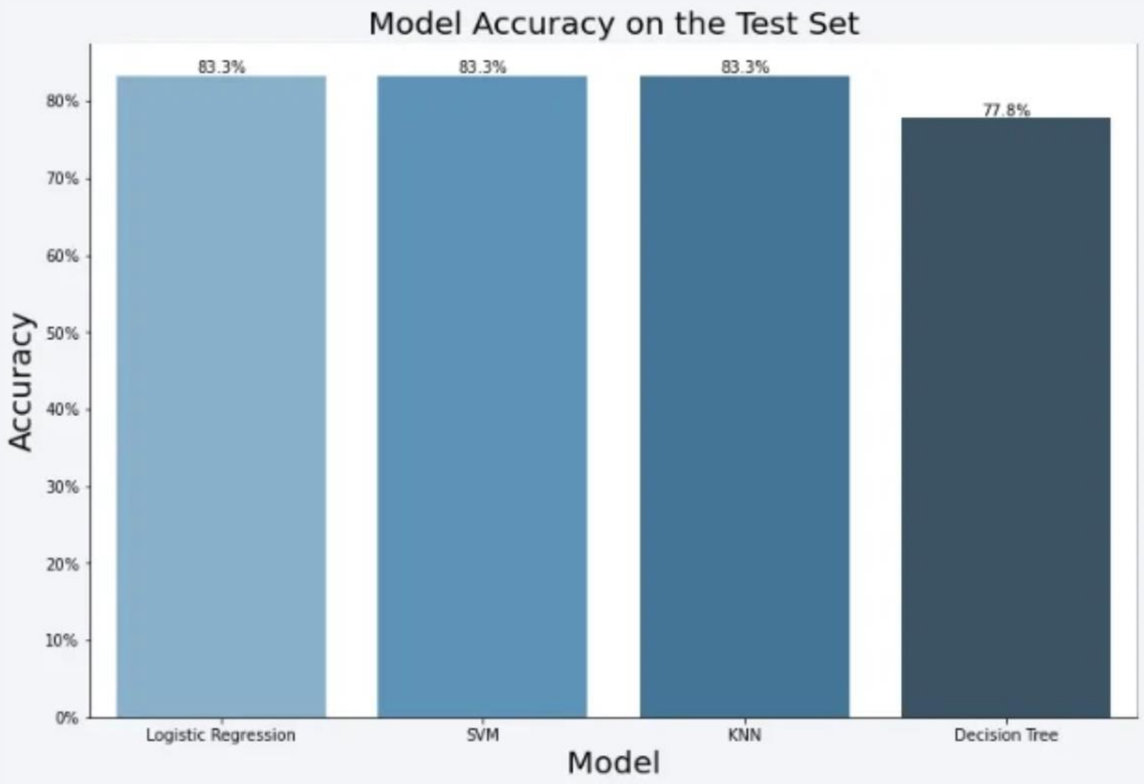
Payload Mass vs Launch Outcome

- Low weighted boosters of v1.1 do have low success rates
- Low weighted boosters of FT have high success rates

Section 5

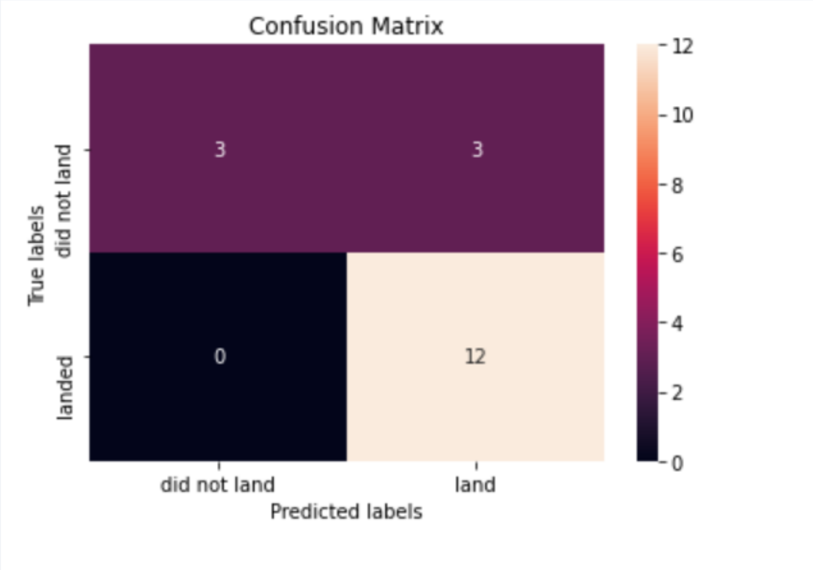
Predictive Analysis (Classification)

Classification Accuracy



Logistic Regression, SVM, and KNN have highest accuracy

Confusion Matrix



- Logistic Regression correctly predicted 50% of unsuccessful landings and 100% of successful landings

Conclusions

- The KNN, SVM, and LR models are the most effective for predicting landing outcomes
- KSC LC 39A had the most successful launches out of all the launch sites
- Low weighted payloads perform better than high weighted payloads
- The average launch success rate increases over time (year by year)

Thank you!