IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Albertina Popp
3/3/2022

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- The commercial space age is here. One of the most successful is SpaceX, and It is relatively inexpensive, because can reuse the first stage. If we can determine if the first stage will land, we can determine the cost of a launch.

- The goal of this project is to help SpaceY to achieve better results by gathering public information about Space X.

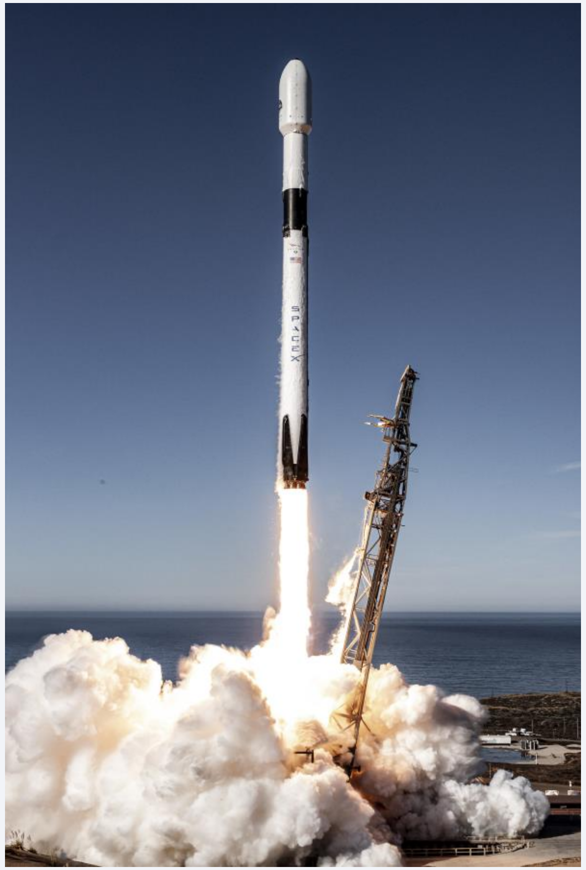
- Summary of methodologies
  - Data Collection through API and Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL and Data Visualization
  - Interactive Visual Analytics with Folium
  - Machine Learning Prediction
- Summary of all results
  - There are differences between the launch sites, orbit and payload mass (KSC LC-39A with light payload appears to be the most successful combination).
  - It is possible to predict the success of a landing with 83,33% accuracy

Introduction

The commercial space age is here, and companies are making space travel affordable for everyone.

One of the most successful is SpaceX. It is relatively inexpensive (62 M versus 165 M), because can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch.

The goal of this project is to determine the price of each launch, by gathering public information about Space X. This information will help SpaceY to achieve better results.



QUESTIONS

- Which launch sites have the highest success rate?
- What factors determine if the rocket will land successfully?
- What kind of infrastructures (cities, highways, streets) are close/far away?

Section 1

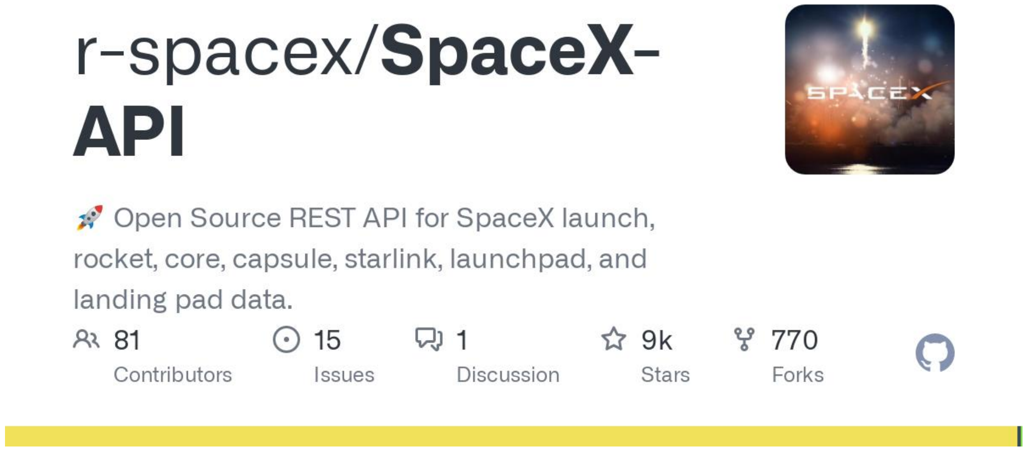
Methodology

Methodology

- Data collection methodology
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

- SpaceX launch data was gathered from an API, specifically the SpaceX REST API. This API provided data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
- Web scraping was performed to collect Falcon 9 historical launch records from a Wikipedia page



Data Collection – SpaceX API

SpaceX API

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

Data Collection

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

Json to Dataframe

```
data=pd.json_normalize(response.json())
```

New Dataframe

```
# Show the head of the dataframe
new_data.head(5)
```

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reus
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False
4	6	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False

https://github.com/Alberpopp/Coursera_Capstone/blob/main/Lab%201.1:%20Data%20Collection%20API.ipynb

Data Collection - Scraping

Wikipedia

```
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&
oldid=1027686922"
```

Data Collection

```
# use requests.get() method with the provided static_url
# assign the response to a object
html_data = requests.get(static_url)
html_data.status_code
```

The data was parsed and converted it into a Pandas data frame

New Dataframe

```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
soup = BeautifulSoup(html_data.text, 'html.parser')
```

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10

https://github.com/Alberpopp/Coursera_Capstone/blob/master/Lab%201.2:%20Data%20%20Webscraping.ipynb

Data Wrangling

An Exploratory Data Analysis (EDA) was performed to find some patterns in the data and determine what would be the label for training supervised models.

In the data set, there are several different kinds of outcomes (True Ocean, False Ocean, True RTLS, False RTLS, True ASDS and False ASDS).

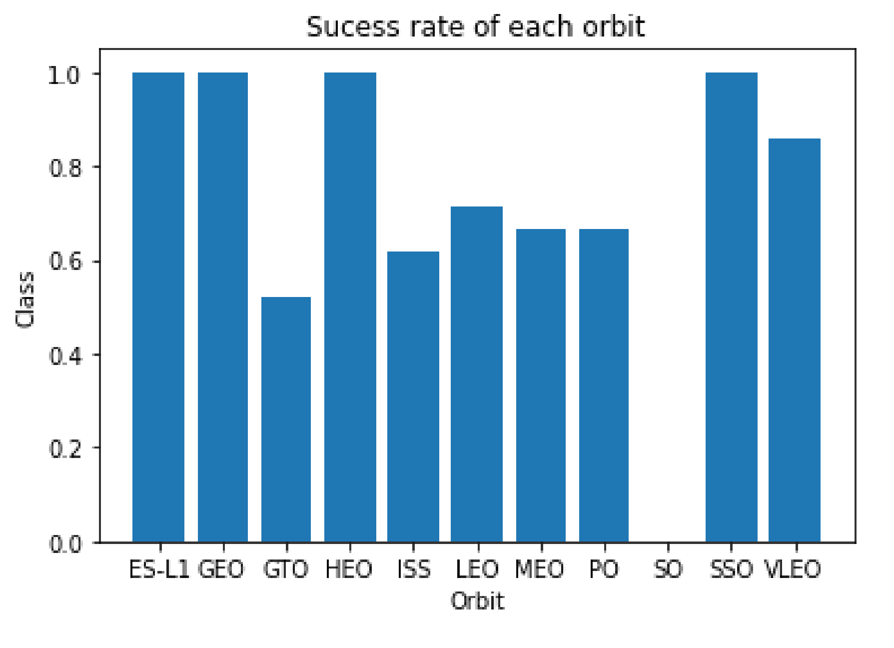
Those outcomes were converted into Training Labels with 1 (the booster successfully landed) , and 0 (it was unsuccessful).

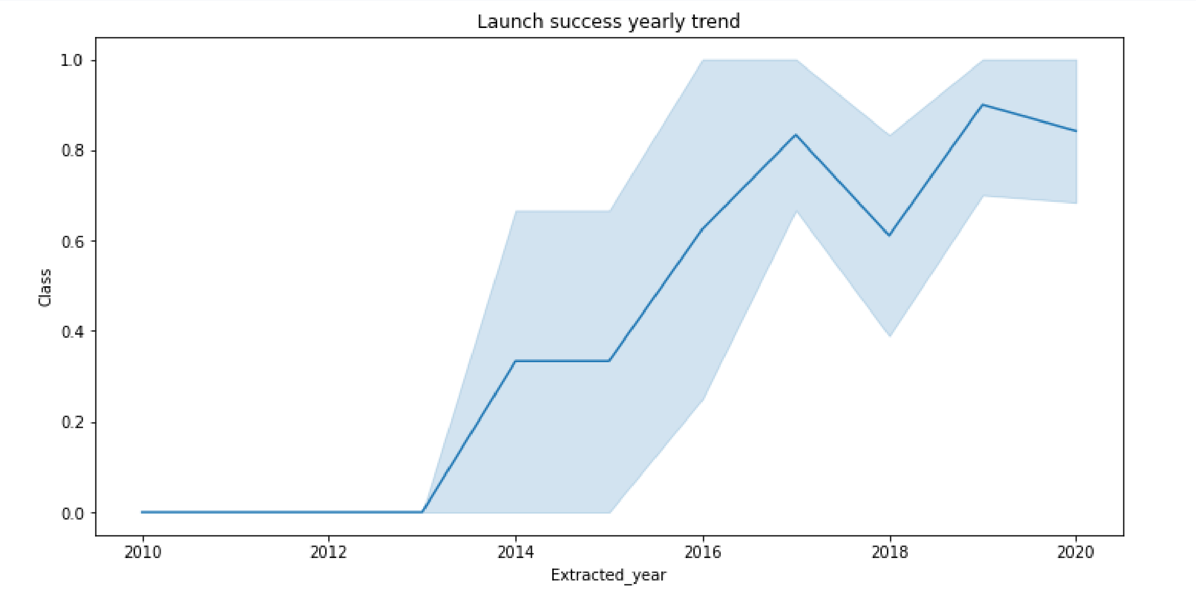
```
# landing_class = 0 if bad_outcome
# landing_class = 1 otherwise
landing_class = []
for outcome in df['Outcome']:
    if outcome in bad_outcomes:
        landing_class.append(0)
    else:
        landing_class.append(1)
```

https://github.com/Alberpopp/Coursera_Capstone/blob/master/Lab%202:%20Data%20wrangling.ipynb

EDA with Data Visualization

A bar chart was used to visually check if there were any relationship between success rate and orbit type.





To get the average launch success trend, a line chart was plotted

https://github.com/Alberpopp/Coursera_Capstone/blob/master/Lab%203.2:%20EDA%20with%20Visualization.ipynb

EDA with SQL

The names of unique launch sites in the space mission.

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

The total payload mass carried by boosters launched by NASA (CRS)

The average payload mass carried by booster version F9 v1.1

the date when the first successful landing outcome in ground pad was achieved

The total number of successful and failure mission outcomes

The failed landing outcomes in drone ship, their booster version and launch site names.

https://github.com/Alberpopp/Coursera_Capstone/blob/master/Lab%203.1:%20Exploratory%20Analysis%20Using%20SQL.ipynb

Build an Interactive Map with Folium

- In order to find some geographical patterns about launch sites, all the launch sites were marked as circles.
- Also, success/failed launches for each site on the map were marked as markers. If a launch was successful (class=1), a green marker was used, and if a launch was failed (class=0), a red marker was used.
- To explore and analyze the proximities of launch sites, distances between a launch site to its proximities were calculated and added to the map as lines.

https://github.com/Alberpopp/Coursera_Capstone/blob/master/Lab%204:%20Launch%20Sites%20Locations%20Analysis%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

A pie chart was added to see which of the four launch sites had the largest success count. Then, the detailed success rate (class=0 vs. class=1) of an specific site

A Range Slider to Select Payload was added, to find if variable payload is correlated to mission outcome.

A scatter plot with the x axis to be the payload and the y axis to be the launch outcome (i.e., class column) was added, to visually observe how payload may be correlated with mission outcomes for selected site(s). In addition, the Booster version was color-labeled to observe mission outcomes with different boosters.

https://github.com/Alberpopp/Coursera_Capstone/blob/master/Dashboard.py

Predictive Analysis (Classification)

The "Class" column was transformed to a NumPy array and assigned to variable Y.

The data was standardized and then reassigned it to the variable X using a transformation.

A training and a testing data were created by splitting the data using the function train_test_split. The training data was divided into validation data, a second set used for training data.

Then, the models (SVM, Classification Trees and Logistic Regression) were trained and hyperparameters were selected using the function GridSearchCV.

The accuracy on the test data was calculated using the method score

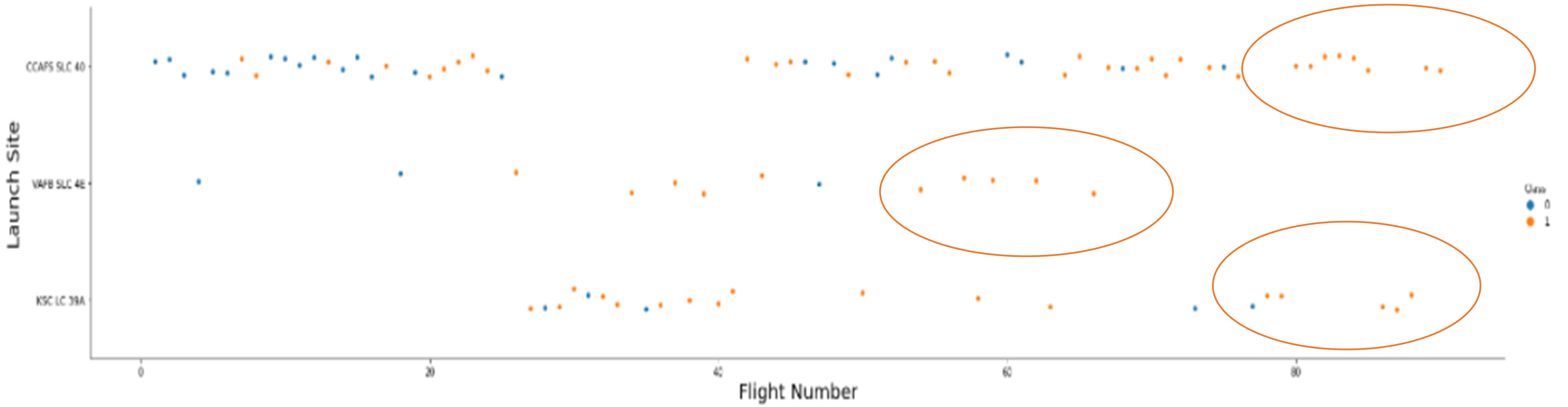
https://github.com/Alberpopp/Coursera_Capstone/blob/master/Lab%205:%20Machine%20Learning%20Prediction.ipynb

Results

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

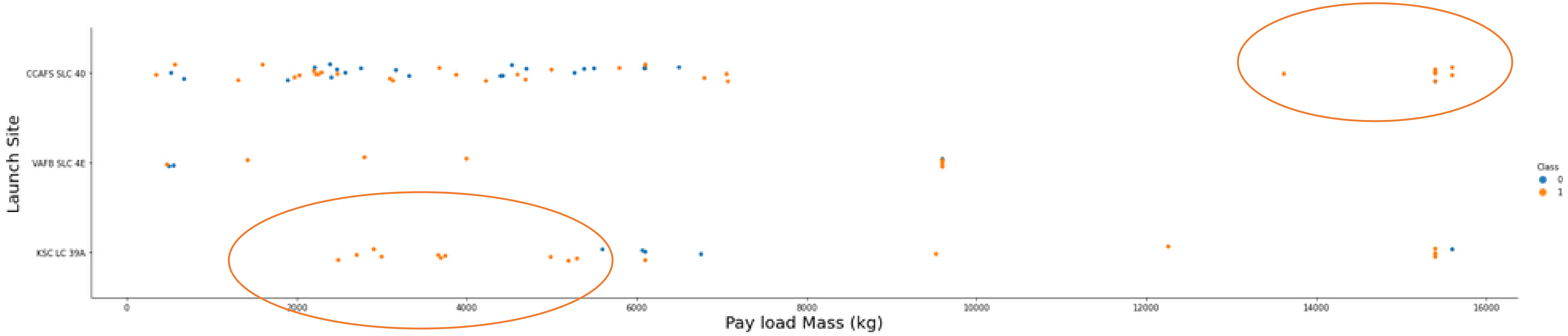
Section 2

Insights drawn from EDA



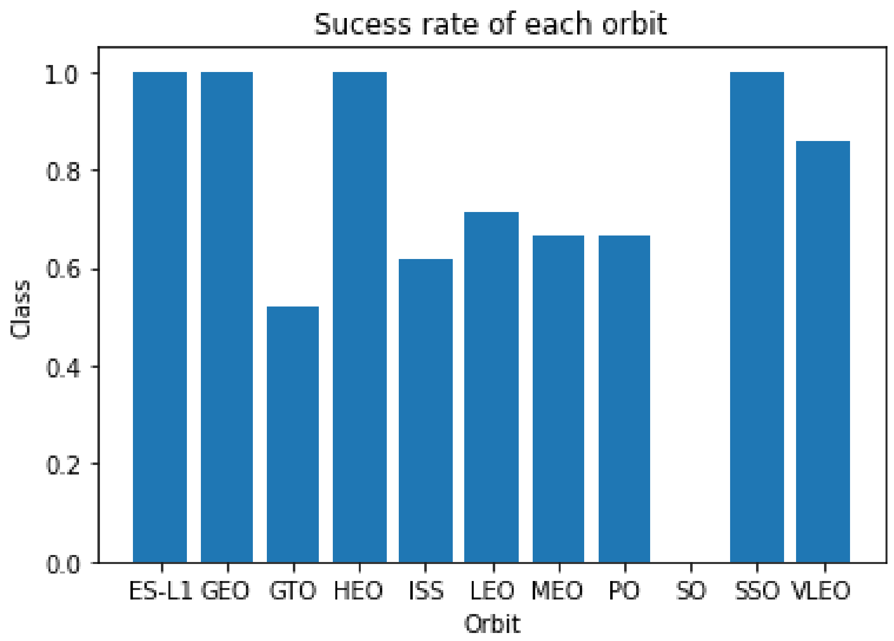
Flight Number vs. Launch Site

The larger the flight amount at a launch site is, the greater the success rate



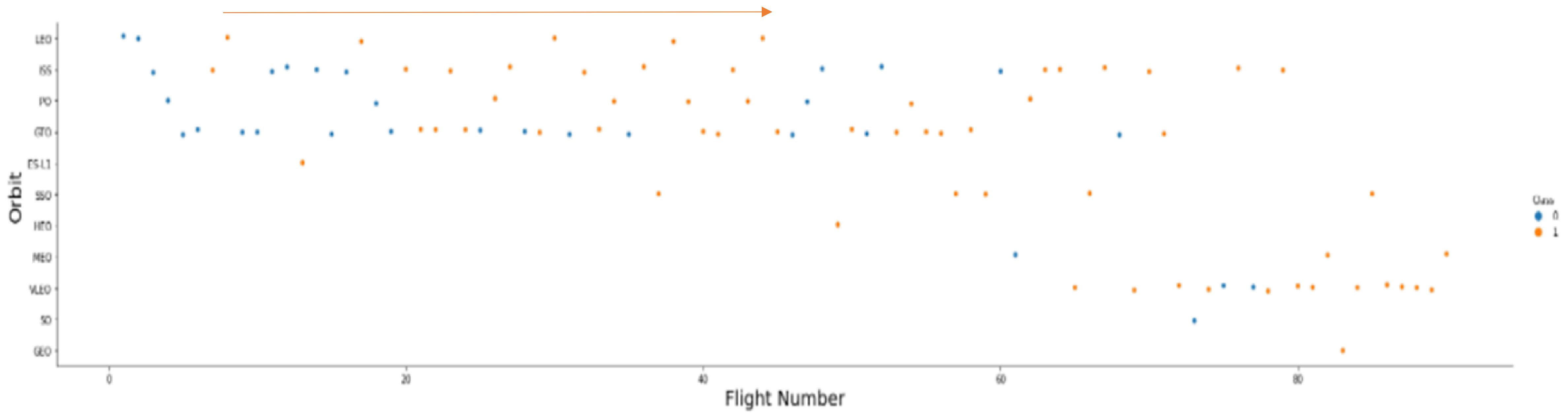
Payload vs. Launch Site

- For the VAFB-SLC launch site there are no rockets launched for heavy payload mass (greater than 10000 kg).
- For KSC LC 39A, the lighter the payload mass, the greater the success rate.



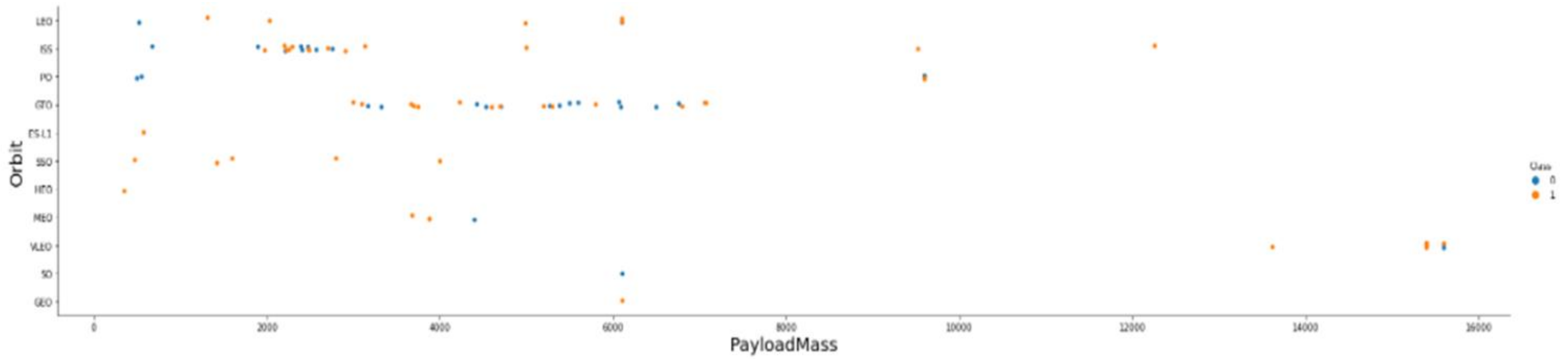
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO, SSO, VLEO had the most success rate



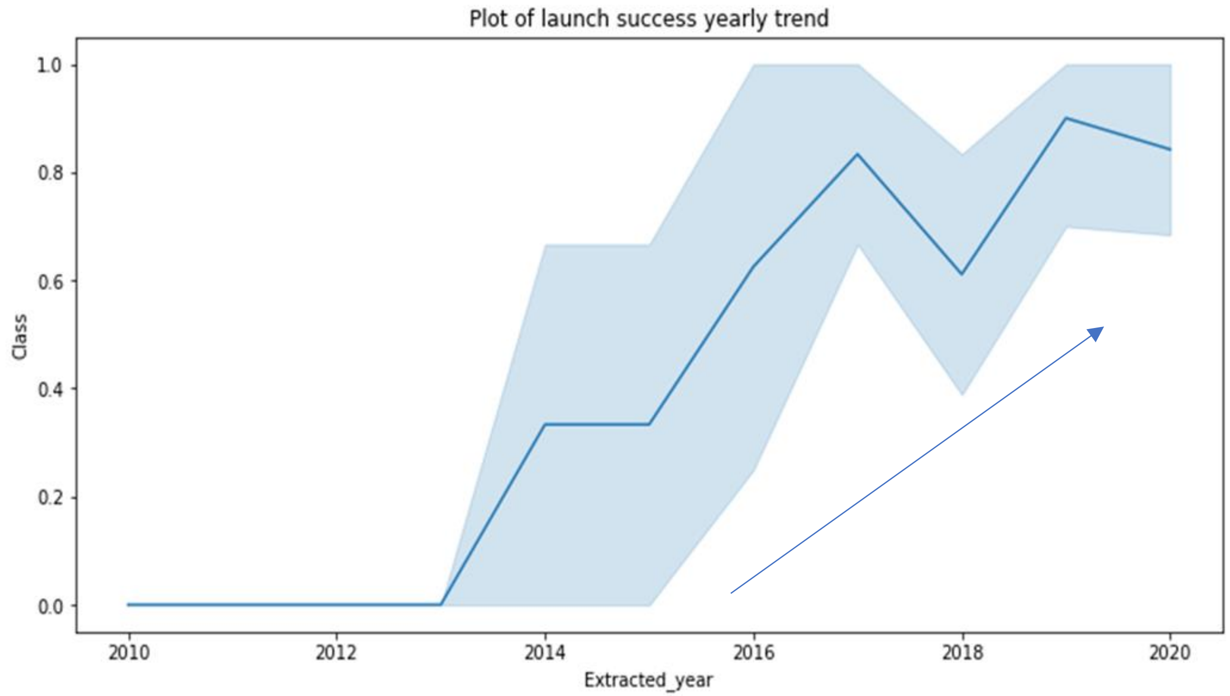
Flight Number vs. Orbit Type

- 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 the orbit is GTO.



Payload vs. Orbit Type

- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- However, for GTO is not possible to distinguish this well as both positive landing rate and negative landing (unsuccessful mission) are mixed.



Launch Success Yearly Trend

- The sucess rate since 2013 kept increasing till 2020

All Launch Site Names

- There are four launch sites: 1 in California (VAFB SLC-4E) and the others in Florida.
- CCAFS LC-40 and CCAFS SLC-40 are close to each other.

```
%sql select distinct launch_site from SPACEXTBL
```

```
 * ibm_db_sa://cqd41143:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

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

Launch Site Names Begin with 'CCA'

```
%sql SELECT * FROM SPACEXTBL WHERE launch_site LIKE 'CCA%'
LIMIT 5
```

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

Total Payload Mass

- The total payload carried by boosters from NASA is 45596 kg

```
%sql SELECT SUM(payload_mass__kg_) AS Total_PayloadMass FROM
SPACEXTBL WHERE customer LIKE 'NASA (CRS)'
```

```
 * ibm_db_sa://cqd41143:***@fbd88901-ebdb-4a4f-a32e-9822b9
fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:3273
1/bludb
Done.
```

total_payloadmass
45596

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1 is 2928 kg.

```
%sql SELECT AVG(payload_mass__kg_) AS Avg_PayloadMass FROM
SPACEXTBL WHERE booster_version = 'F9 v1.1'
```

```
 * ibm_db_sa://cqd41143:***@fbd88901-ebdb-4a4f-a32e-9822b
9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32
731/bludb
Done.
```

avg_payloadmass
2928

First Successful Ground Landing Date

- In 2015 was the first Successful Ground Landing

```
%sql SELECT MIN(DATE) AS FirstSuccessfull_landing_date FROM SPACEXTBL WHERE landing__outcome LIKE 'Success (ground pad)'
```

```
 * ibm_db_sa://cqd41143:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

firstsuccessfull_landing_date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- There are 4 boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 kg

```
%sql SELECT booster_version FROM SPACEXTBL WHERE landing__outcome = 'Success (drone ship)' AND payload_mass__kg_ > 4000 AND payload_mass__kg_ < 6000
```

```
 * ibm_db_sa://cqd41143:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

- There are 100 Successful Mission Outcomes and 1 Failure.

```
%sql SELECT COUNT(mission_outcome) AS SuccessOutcome FROM SPACEXTBL WHERE mission_outcome LIKE 'Success%'
```

```
 * ibm_db_sa://cqd41143:***@fbd88901-ebdb-4a4f-a32e-9822
b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:
32731/bludb
Done.
```

successoutcome
100

```
%sql SELECT COUNT(mission_outcome) AS FailureOutcome FROM SPACEXTBL WHERE mission_outcome LIKE 'Failure%'
```

```
 * ibm_db_sa://cqd41143:***@fbd88901-ebdb-4a4f-a32e-9822
b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:
32731/bludb
Done.
```

failureoutcome
1

Boosters Carried Maximum Payload

- The maximum payload mass was 15600 kg, and there were 12 Boosters.

```
%sql SELECT booster_version, payload_mass__kg_ FROM SPACEXTBL
WHERE payload_mass__kg_ = (SELECT MAX(payload_mass__kg_) FROM
SPACEXTBL ) ORDER BY booster_version
```

```
 * ibm_db_sa://cqd41143:***@fbd88901-ebdb-4a4f-a32e-9822b9f
b237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/
bludb
Done.
```

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600

2015 Launch Records

- In 2015, there were 2 failed landing outcomes in drone ships

```
%sql SELECT booster_version, launch_site, landing__outcome
FROM SPACEXTBL WHERE landing__outcome LIKE 'Failure (drone
ship)' AND DATE BETWEEN '2015-01-01' AND '2015-12-31'
```

```
 * ibm_db_sa://cqd41143:***@fbd88901-ebdb-4a4f-a32e-9822b
9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32
731/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)

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

```
%sql  SELECT landing__outcome, COUNT(landing__outcome) FROM
SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY landing__outcome ORDER BY COUNT(landing__outcome)
DESC
```

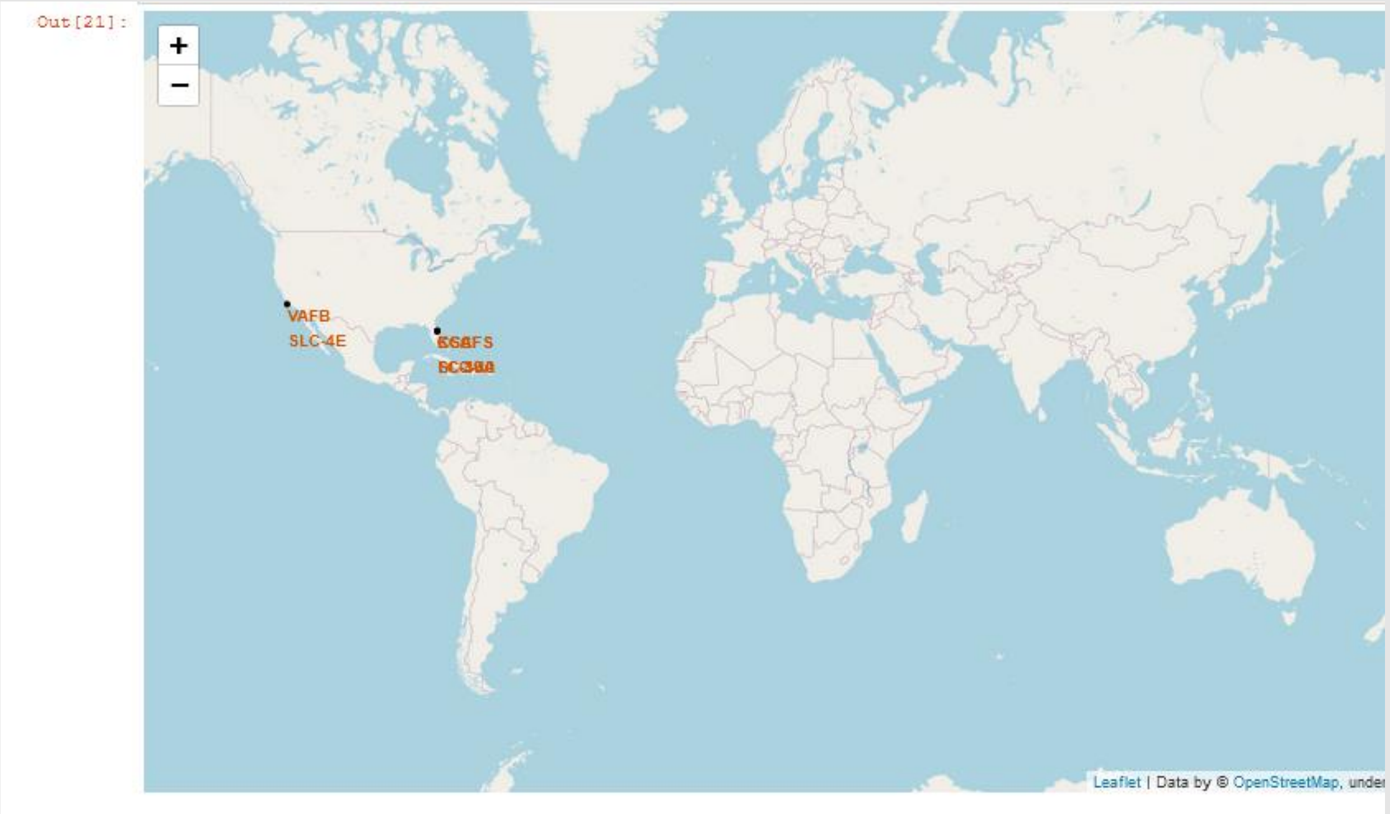
```
 * ibm_db_sa://cqd41143:***@fbd88901-ebdb-4a4f-a32e-9822b
9fb237b.c1ogj3sd0tgtu01qde00.databases.appdomain.cloud:32
731/bludb
Done.
```

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

Section 3

Launch Sites Proximities Analysis

Launch Sites locations

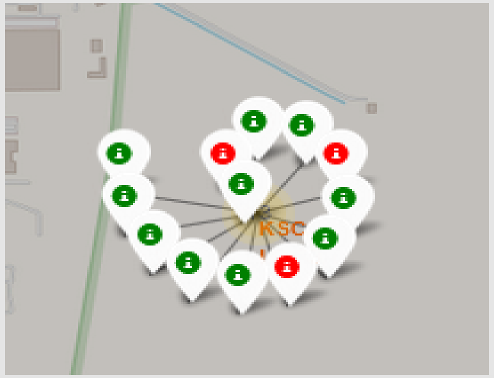


- All the launch sites are next to the Ocean and close to the equator

Lauch outcomes

VAFB SLC-4E

KSC LC-39A

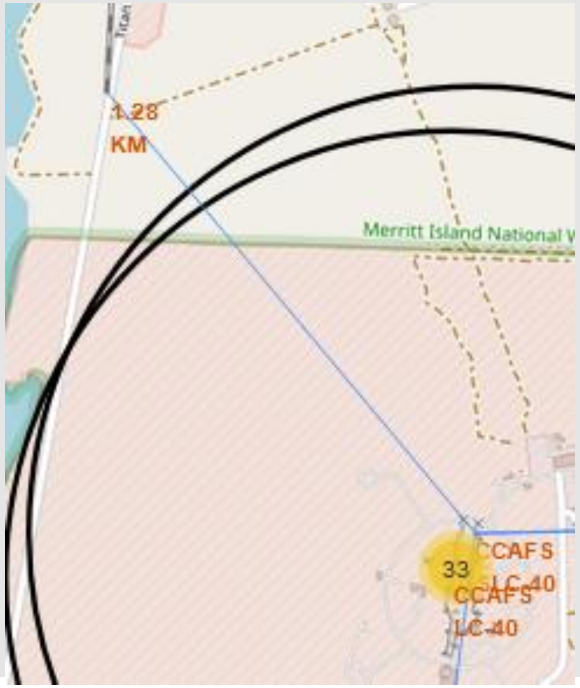


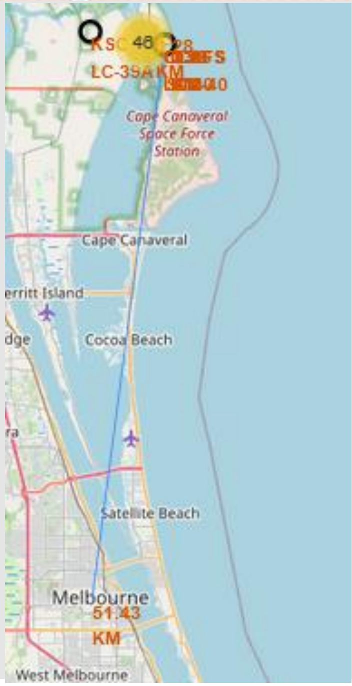
CCAFS LC-40

CCAFS SLC-40

- From the launch outcomes, is easy to see that KSC LC-39A has the highest possibility of success.

Launch sites surroundings



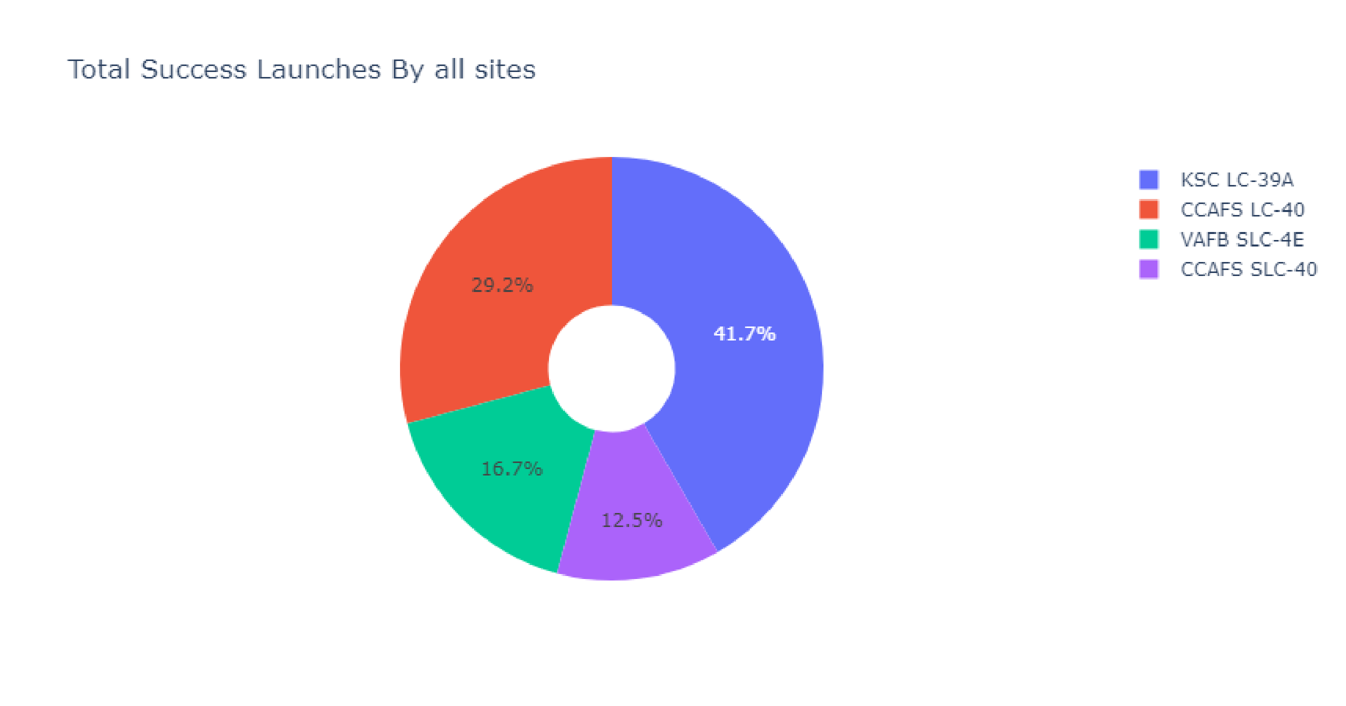


- As an example, **CCAFS SLC-40** is 0.86 km to the coast, 0.58 km to a highway, 1.28 km to a railway and 51.43 km to a city.

Section 4

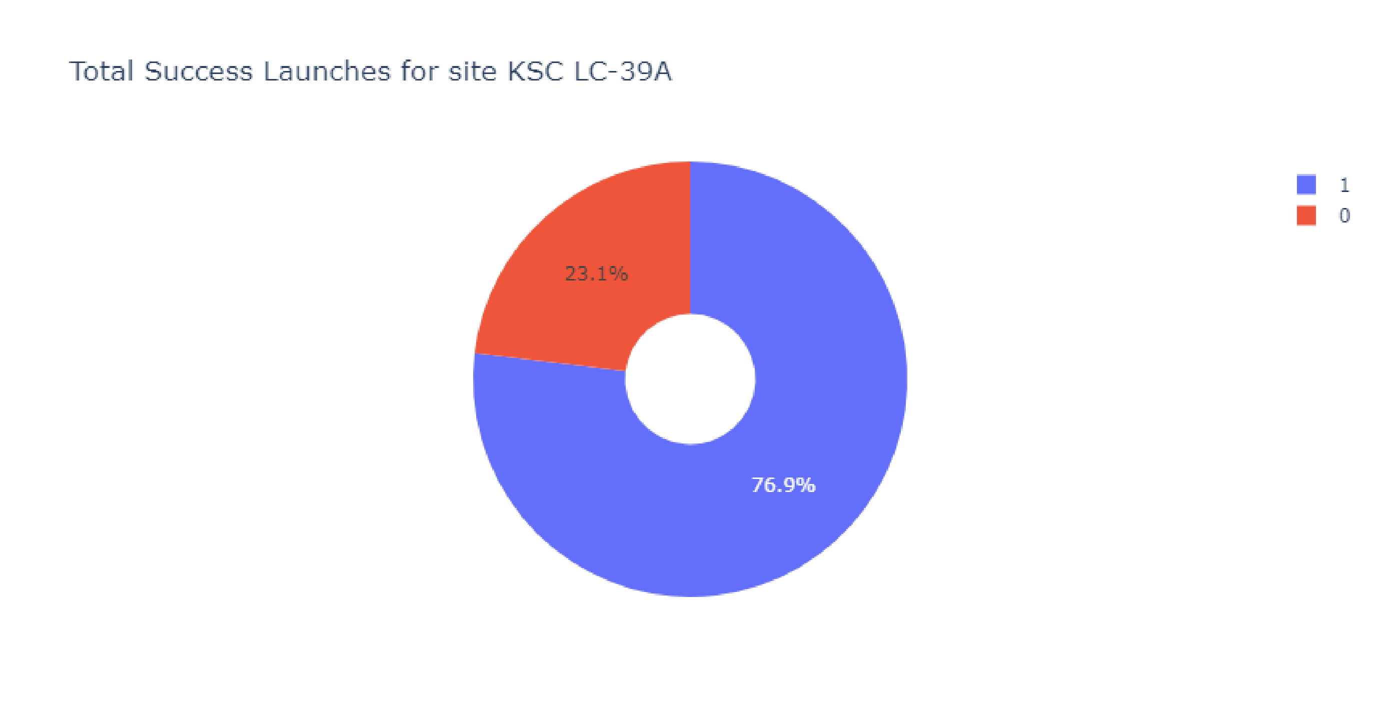
Build a Dashboard with Plotly Dash

Launch success count



- KSC LC-39A is the most successful launch site (41.7%)

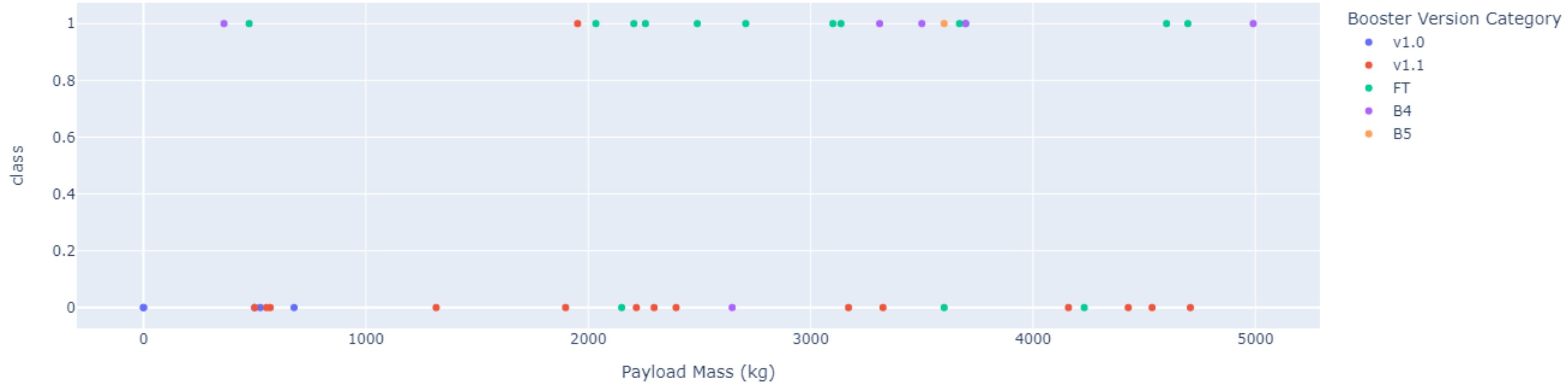
KSC LC-39A launches

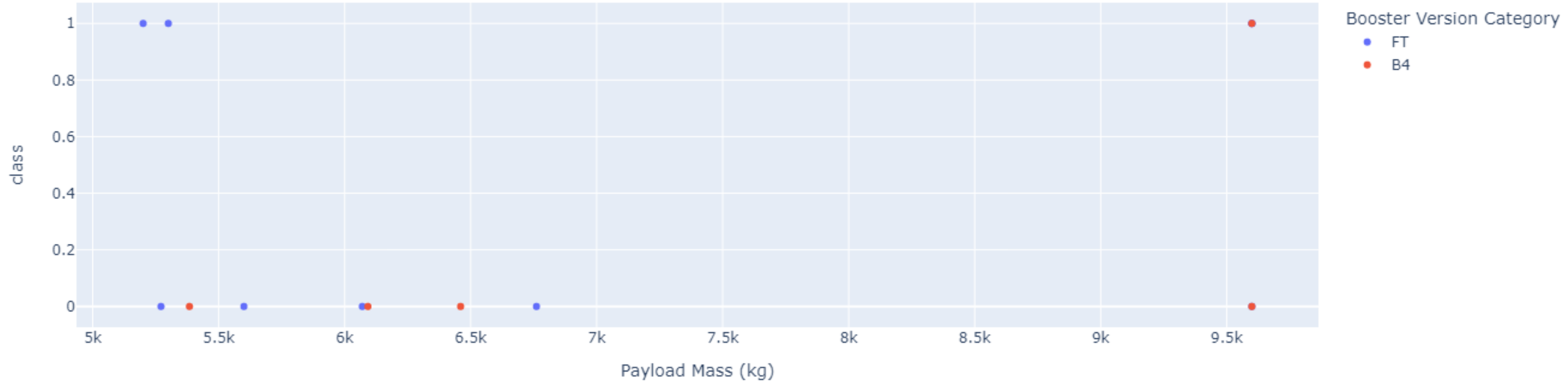


- 76.9% of the launches were successful

Payload impact

- Booster v1.1 had the worst success rate
- Booster FT appears to work very well between 2000 and 3000 kg
- With high payload mass, the success rate is worst





Section 5

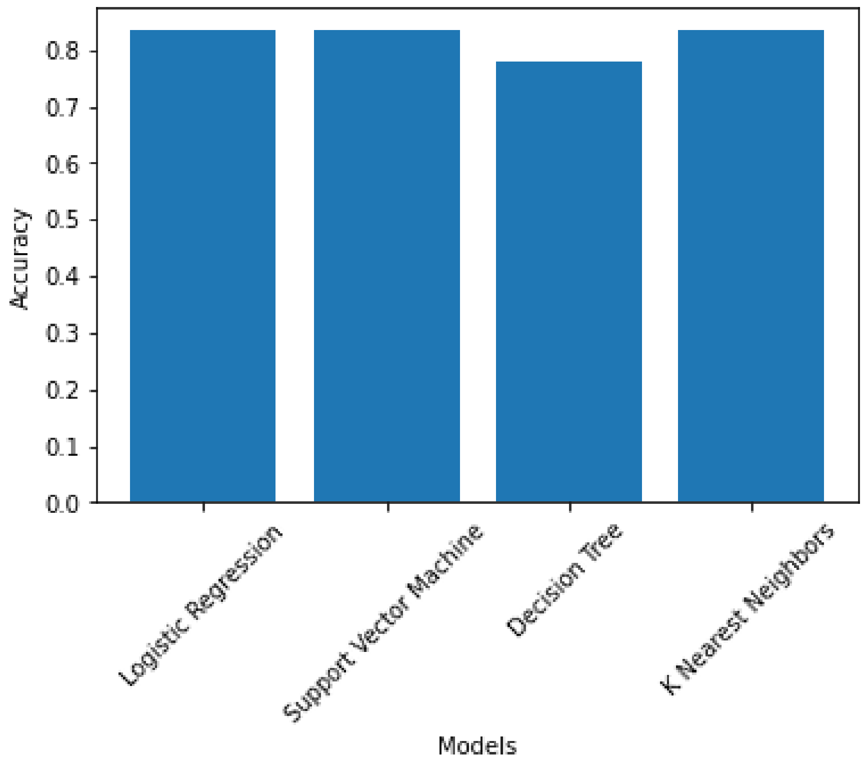
Predictive Analysis (Classification)

Classification Accuracy

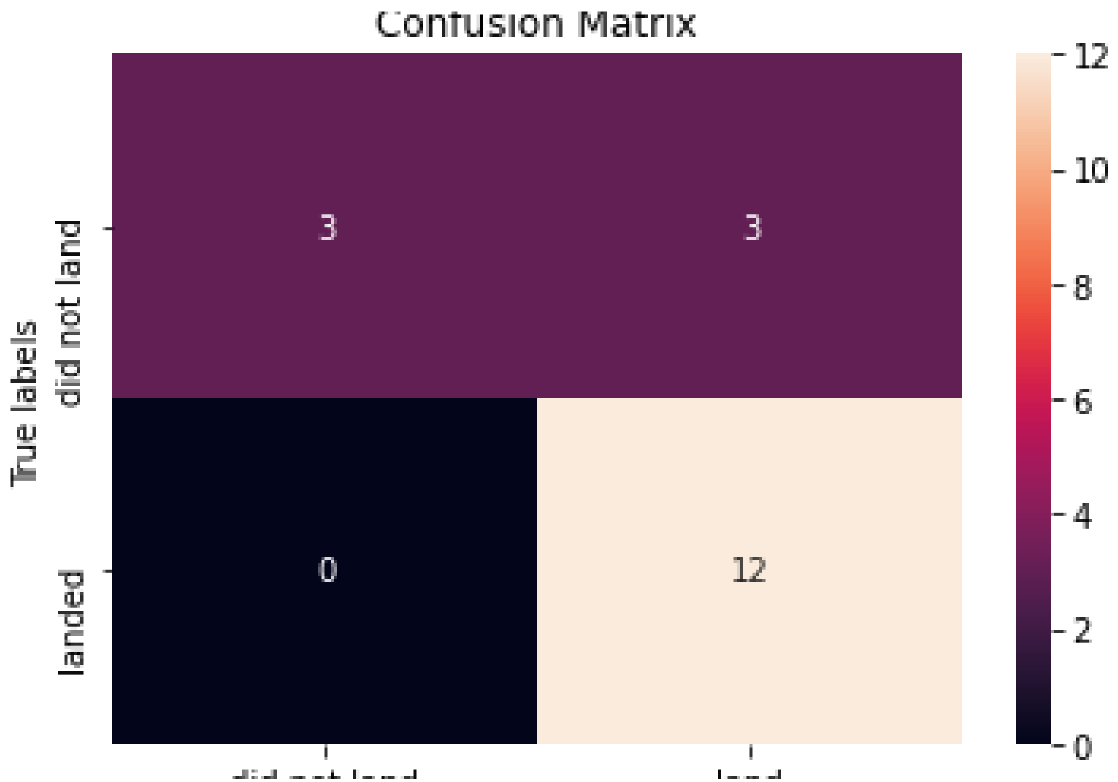
- Logistic Regression, Support Vector Machine and K Nearest Neighbors obtain the highest classification accuracy (0.83)

```
print('Accuracy for Logistics Regression method:', logreg_cv.score(X_test, Y_test))
print( 'Accuracy for Support Vector Machine method:', svm_cv.score(X_test, Y_test))
print('Accuracy for Decision tree method:', tree_cv.score(X_test, Y_test))
print('Accuracy for K nearest neighbors method:', knn_cv.score(X_test, Y_test))
```

```
Accuracy for Logistics Regression method: 0.8333333333333334
Accuracy for Support Vector Machine method: 0.8333333333333334
Accuracy for Decision tree method: 0.7777777777777778
Accuracy for K nearest neighbors method: 0.8333333333333334
```



Confusion Matrix



- The top 3 best performing models obtained the same Confusion Matrix, where all the failed landings were correctly classified. On the other hand, 3 false positive were obtained.

Conclusions

- The common factors related to a successful recovery are:
  - Launch date after 2017
  - Light payload (2000-4000 kg)
  - Launch site: KSC LC-39A
- With the models presented, it is possible to predict the outcome of a given recovery, with an accuracy of 83.33%
- Orbits ES L1, GEO, HEO and SSO have the highest success rates

Thank you!