IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Data Analyses from Space X launch history and booster landing. A Visualization whit a supervised learning model

A Three Model whits a predict of 0.91 %

Introduction

- Safing a Rocking spends 62 million Dollars. When we find out what we can do to make the Rocket better and find a Solution to fix the Problem
- What details of the launch are predictive of booster recovery?
- With model can we best predict booster recovery?
- With what accuracy can we predict booster recovery?

Section 1

Methodology

Executive Summary

- Data collection methodology:
  - Use the SpaceX REST API and webscraping to collect a list of launches and launch parameters including booster landing outcomes
- Perform data wrangling
  - Missing values were replaced with average, only Falcon 9 launches are considered
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Use scitkit-learn to tune a Logistic Regression, Support Vector Maschine, Decition Tree and K-Nearest Neighbors models.

Data Collection

Two datasets were collected, using different collection techniques:

Direct request to the SpaceX REST API

Webscraping on the Wikipedia page

Collected data consists of:

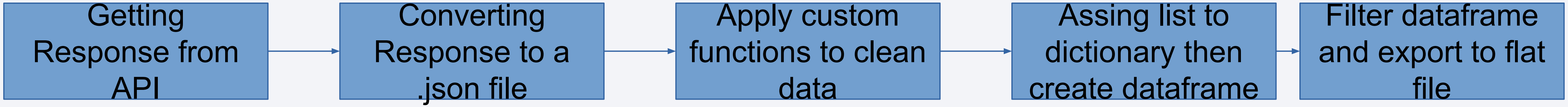
A list of rocket flights and relevant flight

Single Falcon 9 launches

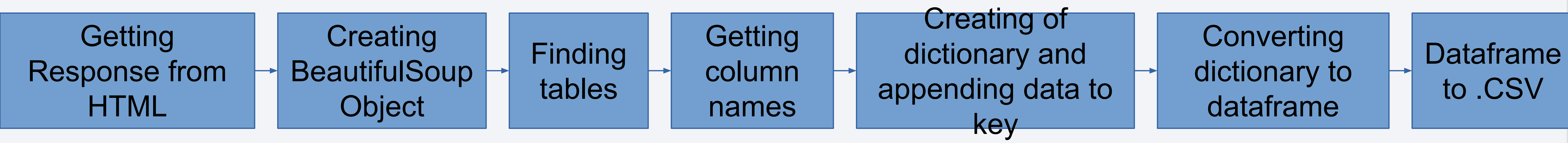
Data Collection – SpaceX API

- Used requests to get data from SpaceX Rest API.

[Link]

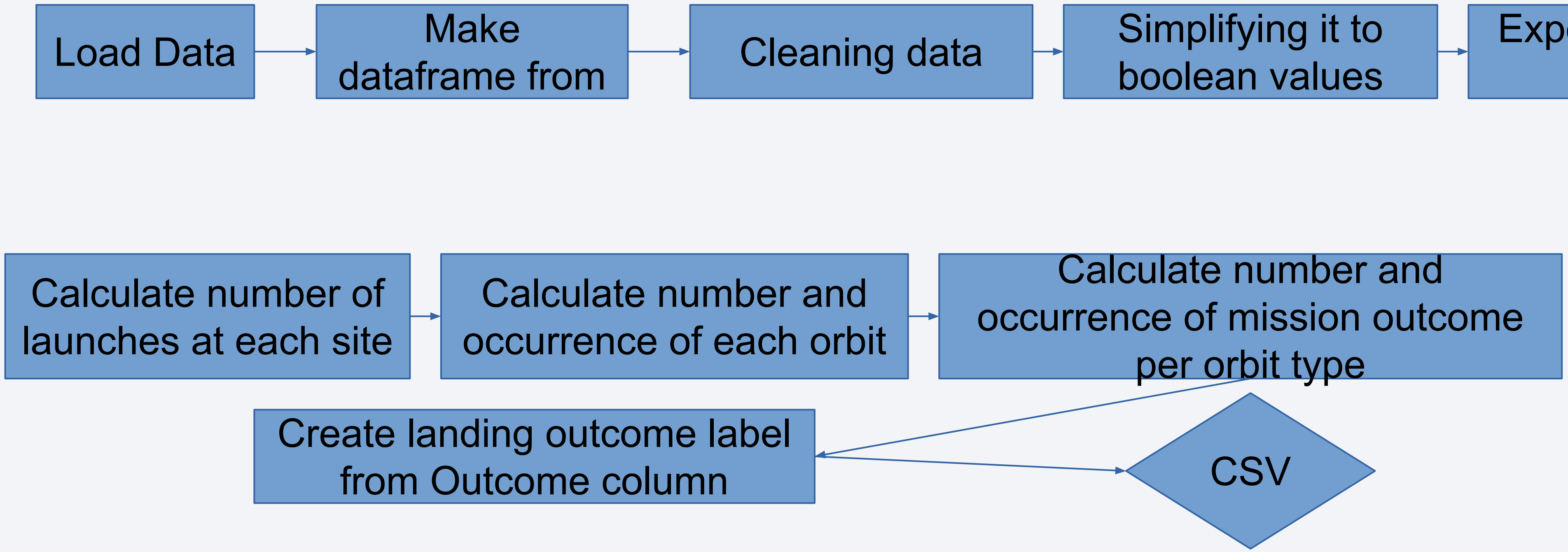


Data Collection - Scraping



[jupyter-labs-webscraping.ipynb]

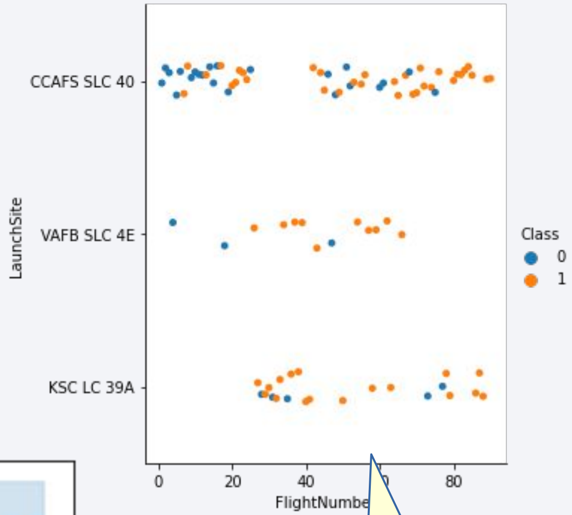
Data Wrangling



[labs-jupyter-spacex-Data wrangling]

EDA with Data Visualization

Exploratory data analysis is analyzing data and find there Charakter with using graphs

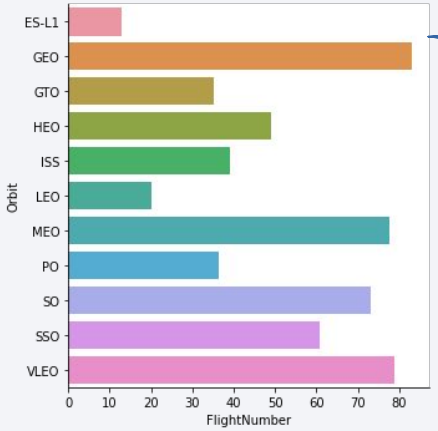


Load Data

Make dataframe fromit

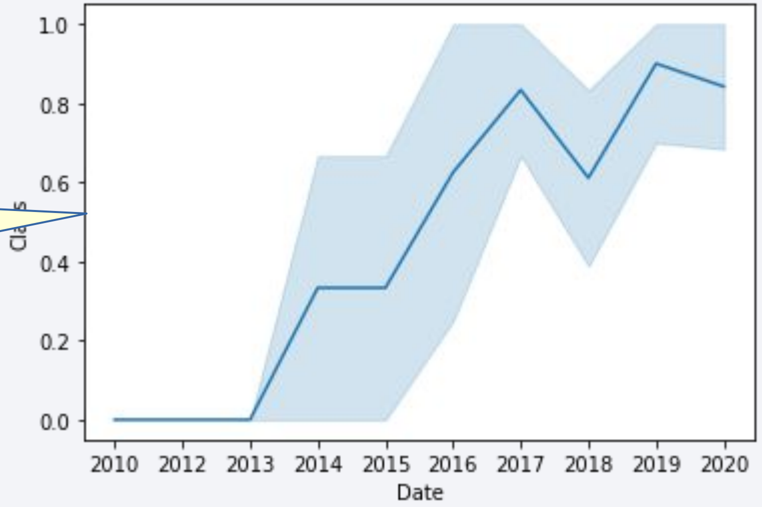
Create Visualization

Collect Insights



Bar Graph shows relationship

Line Grapg shows trends



Scatter plots show dependency of attributes

[Assignment: Exploring and Preparing Data]

EDA with SQL

SQL is an indispensable tool for Data Scientist and analysts. Is a powerful tool for analyzing data and drawing useful insights from it.

booster_version	payload_mass__kg_	landing__outcome
F9 v1.1	4535	No attempt
F9 v1.1 B1011	4428	No attempt
F9 v1.1 B1014	4159	No attempt
F9 v1.1 B1016	4707	No attempt
F9 FT B1020	5271	Failure (drone ship)
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1030	5600	No attempt
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1032.1	5300	Success (ground pad)
F9 B4 B1040.1	4990	Success (ground pad)
F9 FT B1031.2	5200	Success (drone ship)
F9 B4 B1043.1	5000	Success (ground pad)
F9 FT B1032.2	4230	Controlled (ocean)
F9 B4 B1040.2	5384	No attempt
F9 B5 B1046.2	5800	Success
F9 B5 B1047.2	5300	Success
F9 B5 B1046.3	4000	Success
F9 B5B1054	4400	No attempt
F9 B5 B1048.3	4850	Success
F9 B5 B1051.2	4200	Success
F9 B5B1060.1	4311	Success
F9 B5 B1058.2	5500	Success
F9 B5B1062.1	4311	Success

[Applied-Data-Science-Capstone]

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

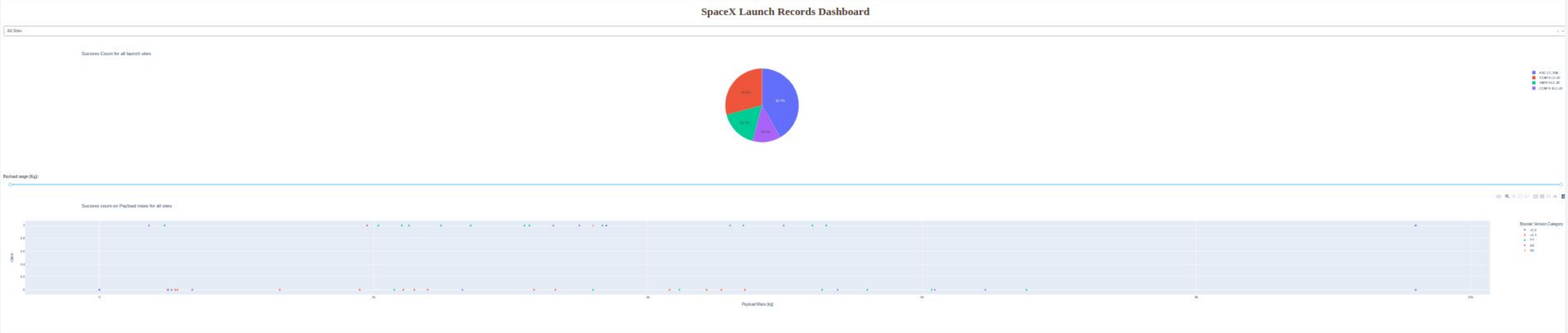
Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

lab_jupyter_launch_site_location

Build a Dashboard with Plotly Dash

Pi Chart showing total success for all sites or by certain launch site

- Percentoage of success in relation to launche site



[dash]

Predictive Analysis (Classification)

Building a Model

- Load the data into a dataframe
- Transform it into a Numpy dataframe
- Standardize the Dataframe
- Split the data into train and test data
- List the maschine algorithmen down
- Set out the Parameters
- Fit out your Dataset

Evaluation Model

- check accuracy for every Model
- Plot Confusion Matrix

Find the best Modell

https://github.com/Mathis0490/Applied-Data-Science-Capstone/blob/F%C3%BCr-meine-Projekte/SpaceX_Machine_Learning_Prediction_Part_5.ipynb

Results

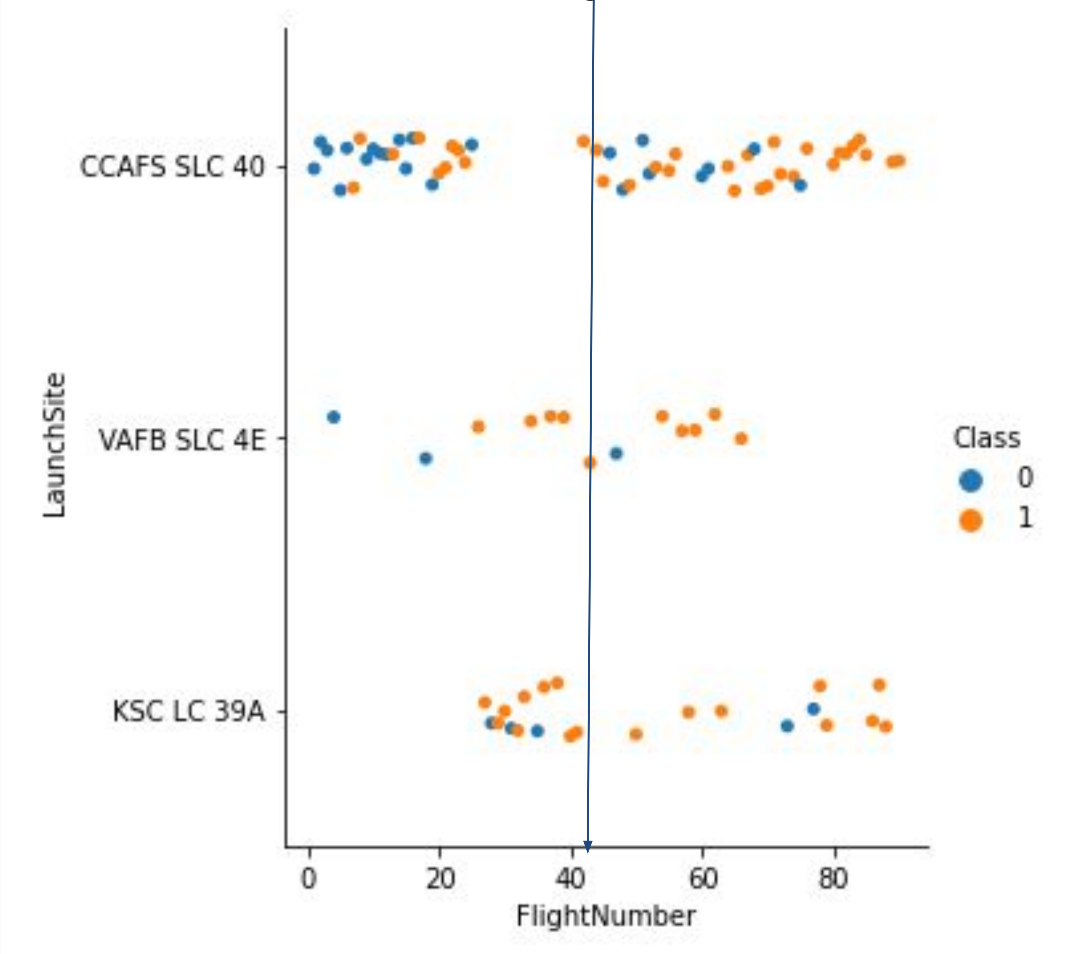
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

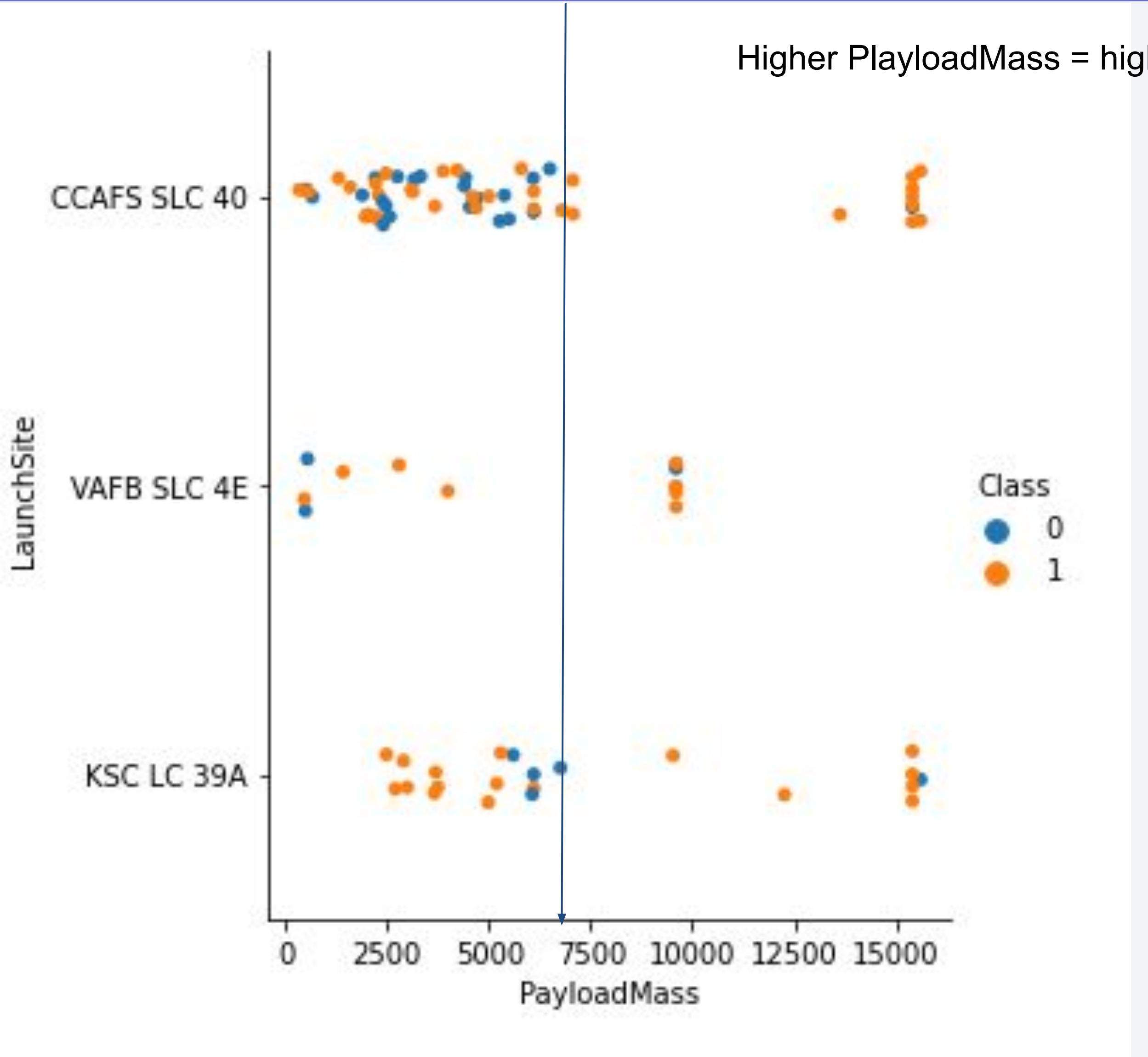
Insights drawn from EDA

Flight Number vs. Launch Site

After 40 Flight the succes rate is increasing

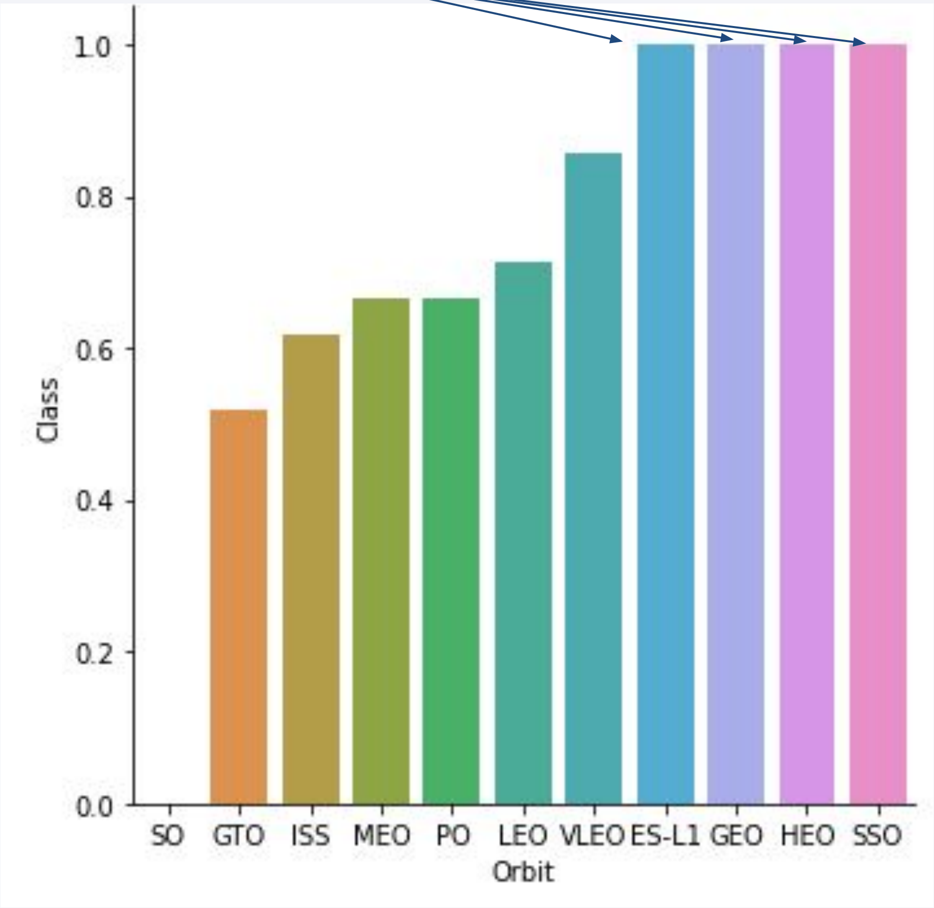


Payload vs. Launch Site



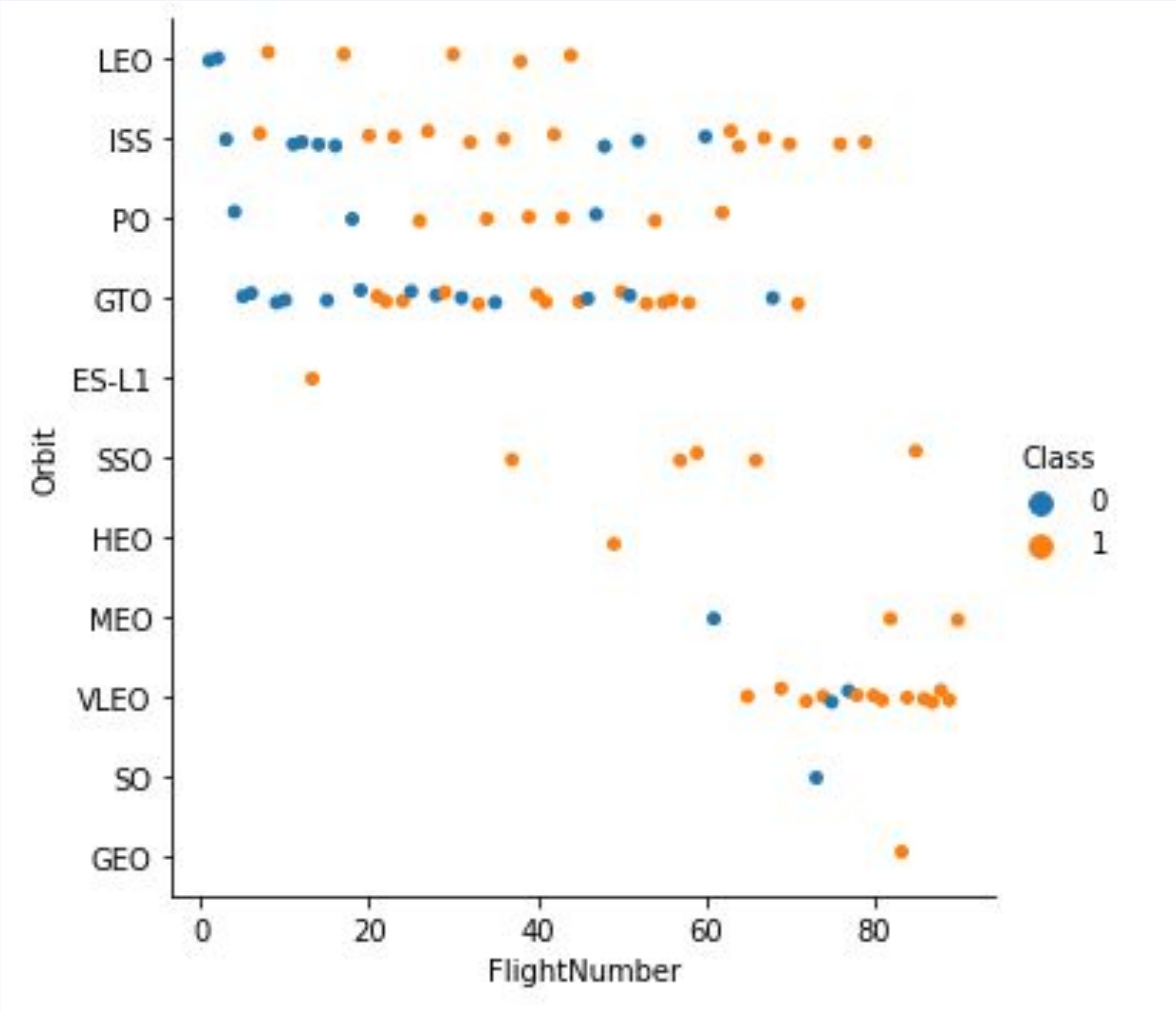
Success Rate vs. Orbit Type

High sucess rate

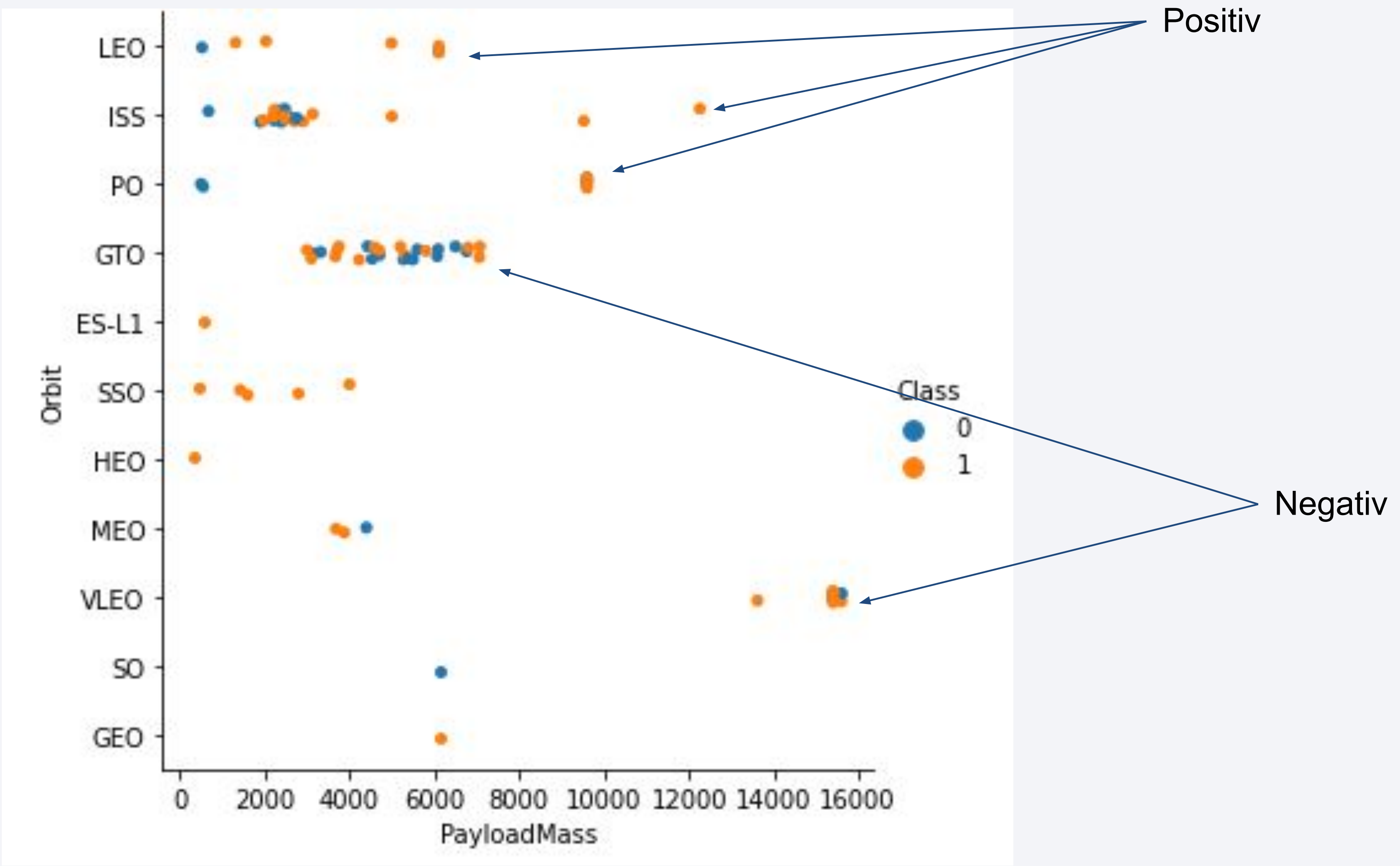


Flight Number vs. Orbit Type

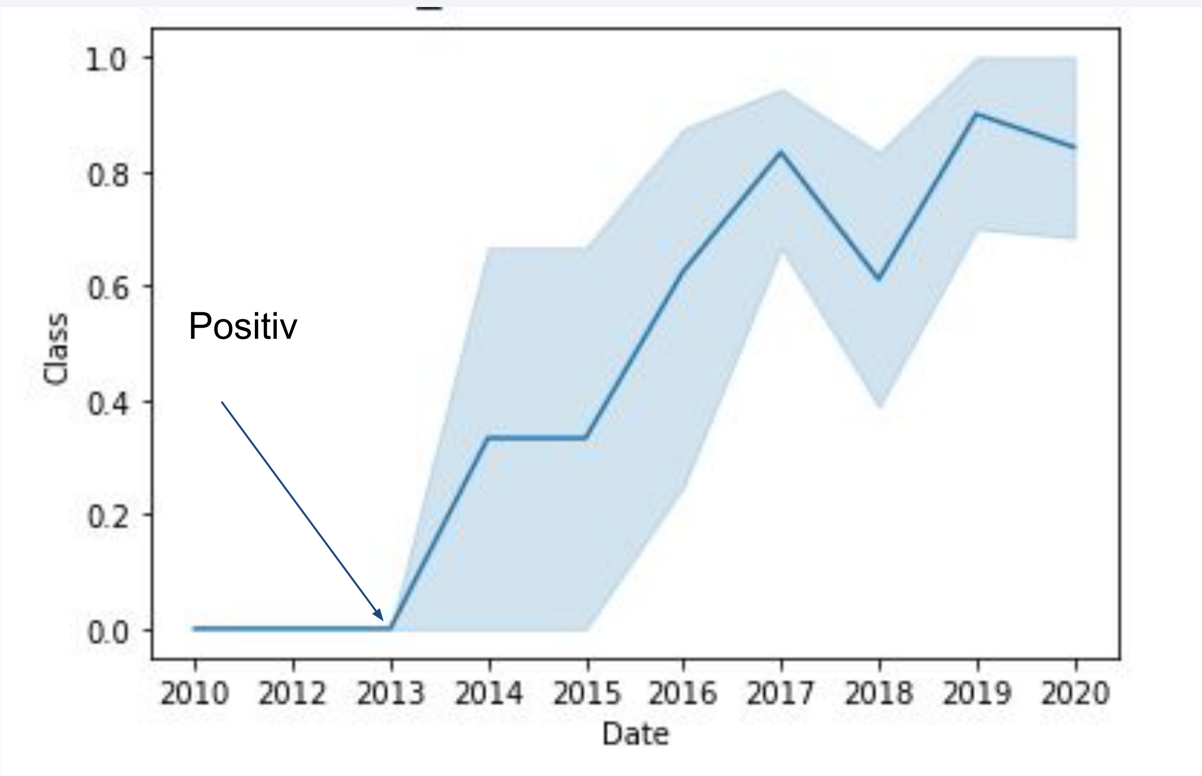
we see that higher the flight Number so higher the sucess rate



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

```
%sql select launch_site  from SPACEXTBL group by launch_site;
```

```
 * ibm_db_sa://hpj82237:***@764264db-9824-4b7c-82df-40d1b13897c2
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;
```

```
 * ibm_db_sa://hpj82237:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/BLUDB
Done.
```

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

```
%sql select SUM(payload_mass__kg_) from spacextbl where customer like '%NASA (CRS)%';
```

```
 * ibm_db_sa://hpj82237:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases
Done.
```

1
48213

Average Payload Mass by F9 v1.1

```
%sql select AVG(payload_mass__kg_) from spacextbl where booster_version like '%F9 v1.1%';

 * ibm_db_sa://hpj82237:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.database
Done.
```

1
2534

First Successful Ground Landing Date

```
%sql select min(DATE) from spacextbl where mission_outcome like 'Success';
```

```
 * ibm_db_sa://hpj82237:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kq
Done.
```

1
2010-06-04

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT BOOSTER_VERSION FROM SPACEX WHERE LANDING__OUTCOME = 'Success (drone ship)' \
AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;
```

```
 * ibm_db_sa://zpw86771:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.datak
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

```
%%sql
SELECT
  SUM(CASE WHEN mission_outcome LIKE '%Success%' THEN 1 ELSE 0 END) AS num_of_success,
  SUM(CASE WHEN mission_outcome LIKE '%Failure%' THEN 1 ELSE 0 END) AS num_of_failure
FROM
  SPACEXTBL;
```

```
 * ibm_db_sa://hpj82237:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.dat
Done.
```

num_of_success	num_of_failure
100	1

Boosters Carried Maximum Payload

```
%%sql
SELECT
  booster_version
FROM
  SPACEXTBL
WHERE
  payload_mass__kg_ = (SELECT MAX(payload_mass__kg_) FROM SPACEXTBL);
```

```
 * ibm_db_sa://hpj82237:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io9010
Done.
```

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

2015 Launch Records

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

```
%sql SELECT DATE, booster_version, launch_site, landing__outcome FROM SPACEXTBL WHERE YEAR(date) = 2015 AND landing__outcome = 'Failu
```

```
 * ibm_db_sa://hpj82237:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/BLUDB
Done.
```

DATE	booster_version	launch_site	landing__outcome
2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

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

```
%sql SELECT landing__outcome, COUNT(*) AS cnt_of_landing_outcome FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROU
```

```
 * ibm_db_sa://hpj82237:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90108kqb1od8lcg.databases.appdomain.cloud:32536/BLUDB
Done.
```

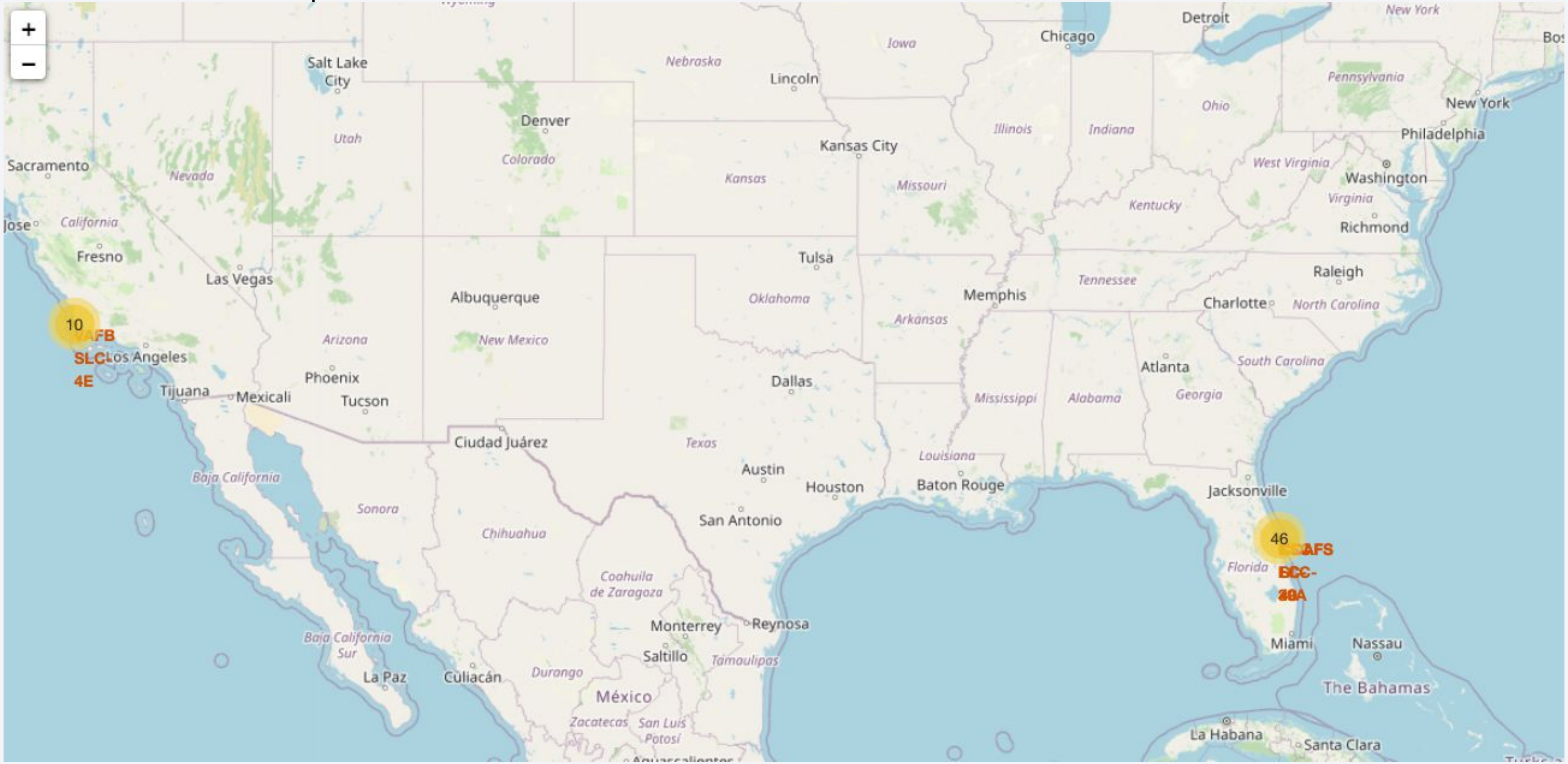
landing__outcome	cnt_of_landing_outcome
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 4

Launch Sites Proximities Analysis

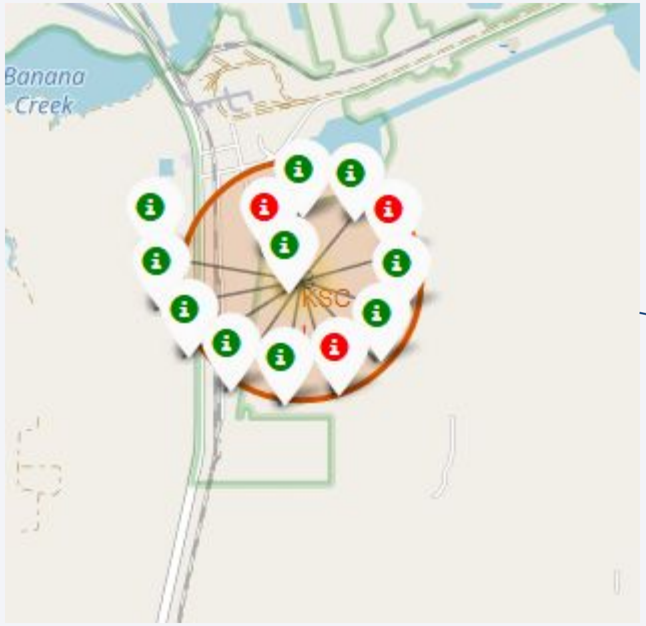
<Folium Map Screenshot 1>

We can see the SpaceX areas

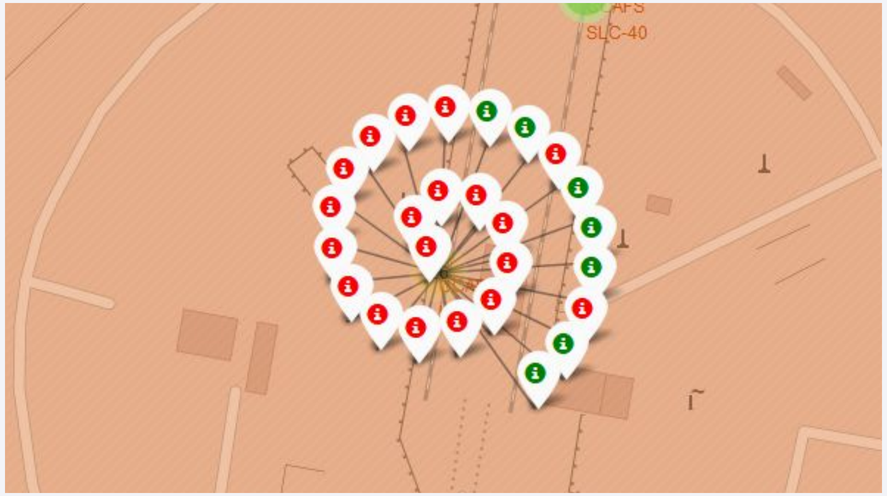


<Folium Map Screenshot 2>

Green are the succes full and red are the failed test

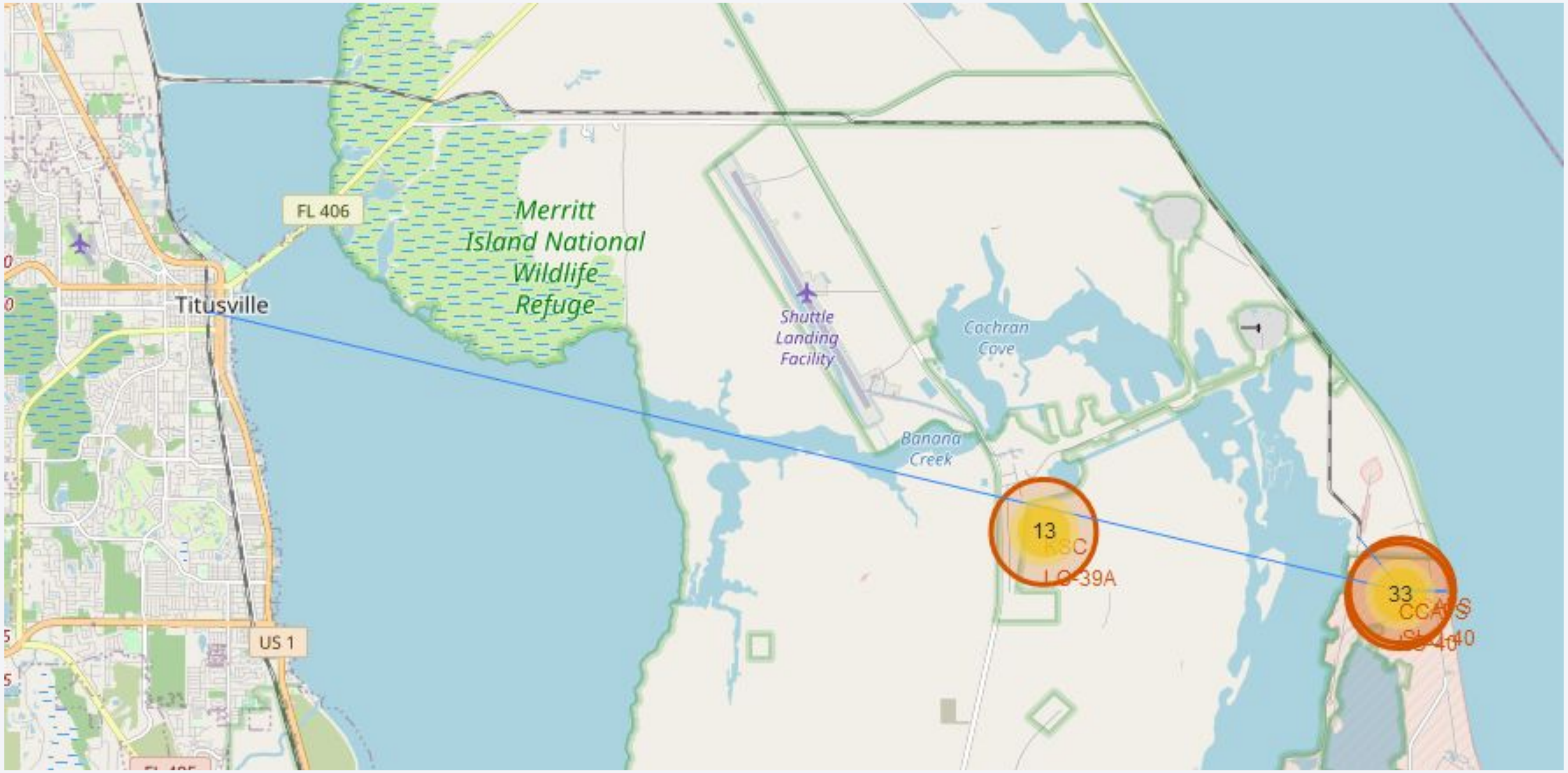


KSC



<Folium Map Screenshot 3>

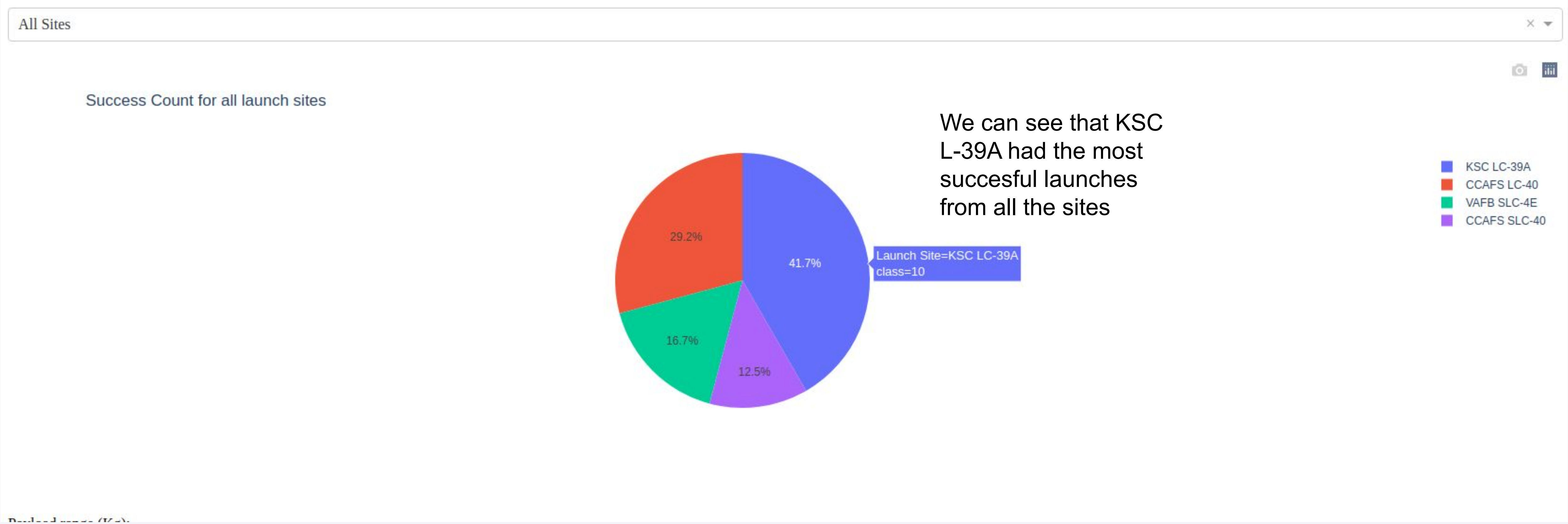
Show the distance to the next city



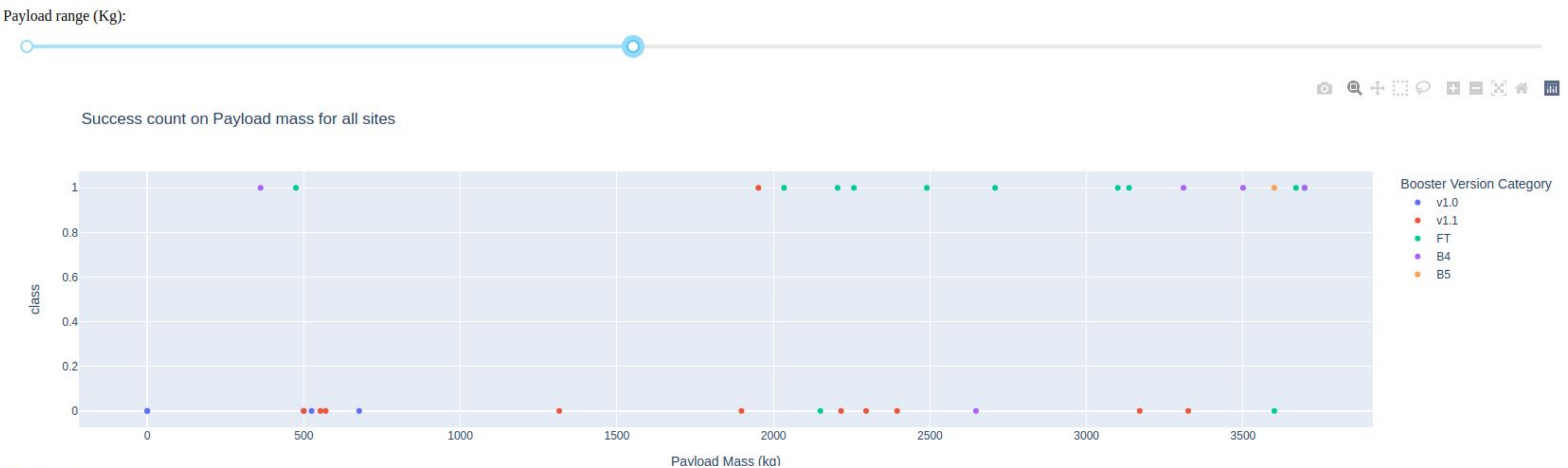
Section 5

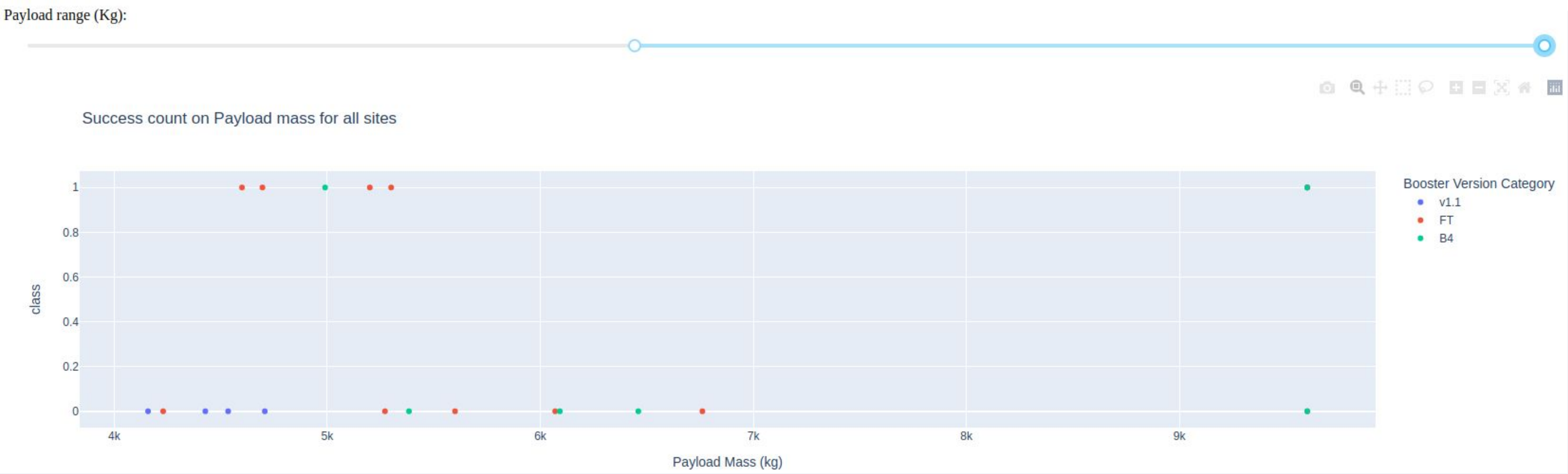
Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

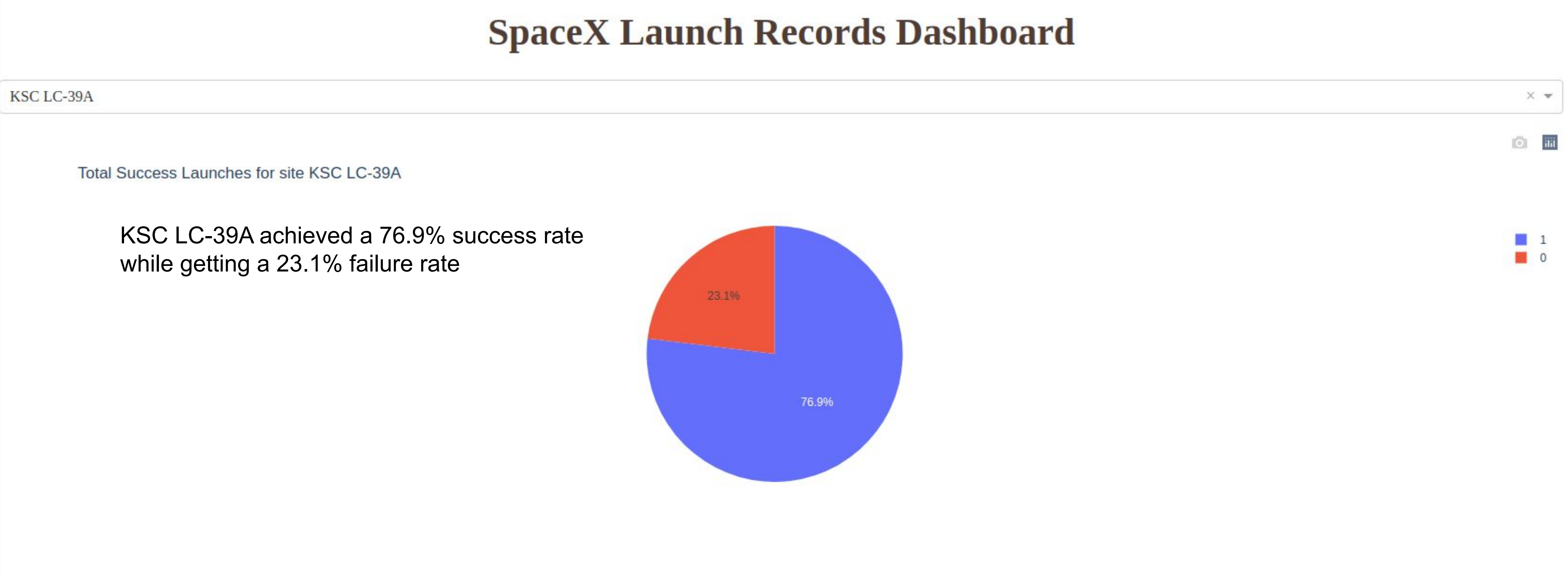


We can see that KSC L-39A had the most succesful launches from all the sites we can see the succes rates for los weightes payloads is higher than heavy weighted payloads





Launch Site with Highest Launch Success Ration

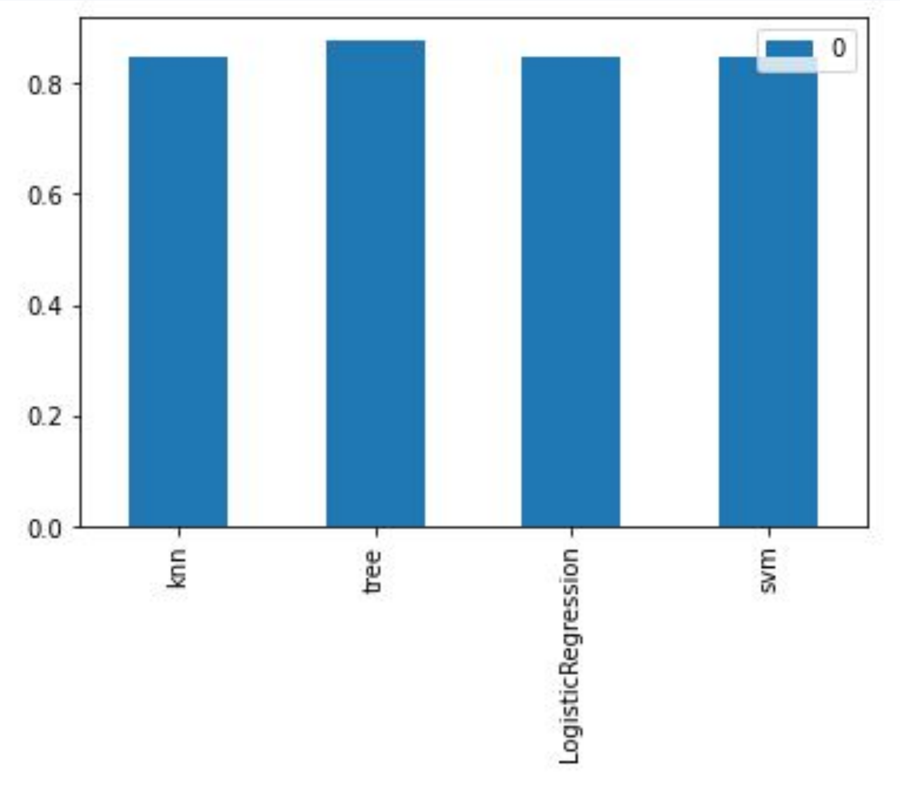


KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

Section 6

Predictive Analysis (Classification)

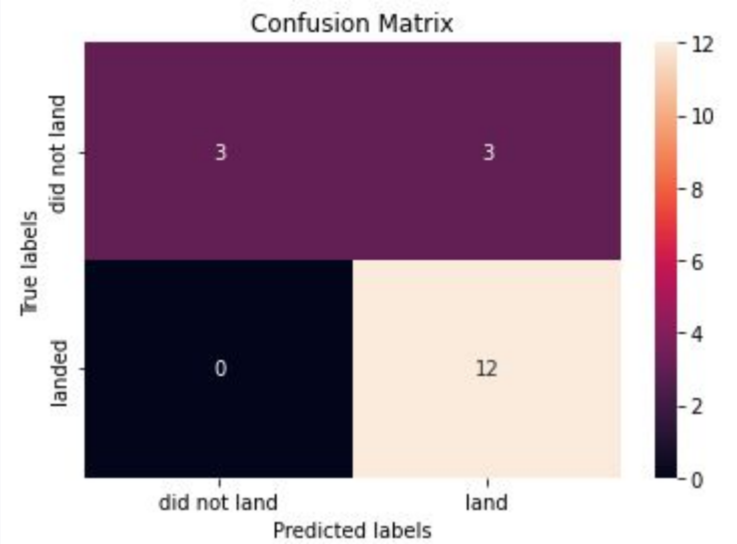
Classification Accuracy



knn	0.848214
tree	0.876786
LogisticRegression	0.846429
svm	0.848214

Confusion Matrix

		Predicted condition	
	Total population = P + N	Positive (PP)	Negative (PN)
Actual condition	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)



We see the Confusion matrix from the tree. We have only 3 False negativ answere.

Conclusions

1. Orbits ES-L1, GEO, HEO, SSo has highest Sucess rate
2. Success rates for SpaceX launches has been increasing relatively with time and it looks like soon they will reach the required target
3. KSC LC-39A had the most successful launches but inceasing payload mass seems to have negative impact on success
4. Decision Tree algorithm is the best for Maschine Learning Model for provided dataset

Appendix

Find the method performs best:

```
algorithms = {'KNN':knn_cv.best_score_,'Tree':tree_cv.best_score_,'LogisticRegression':logreg_cv.best_score_}
bestalgorithm = max(algorithms, key=algorithms.get)
print('Best Algorithm is',bestalgorithm,'with a score of',algorithms[bestalgorithm])
if bestalgorithm == 'Tree':
    print('Best Params is :',tree_cv.best_params_)
if bestalgorithm == 'KNN':
    print('Best Params is :',knn_cv.best_params_)
if bestalgorithm == 'LogisticRegression':
    print('Best Params is :',logreg_cv.best_params_)
```

```
Best Algorithm is Tree with a score of 0.875
Best Params is : {'criterion': 'entropy', 'max_depth': 16, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 10, 'splitter': 'best'}
```

Thank you!