IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Akhmadaliev Yakhyokhon
July 6, 2022

Outline

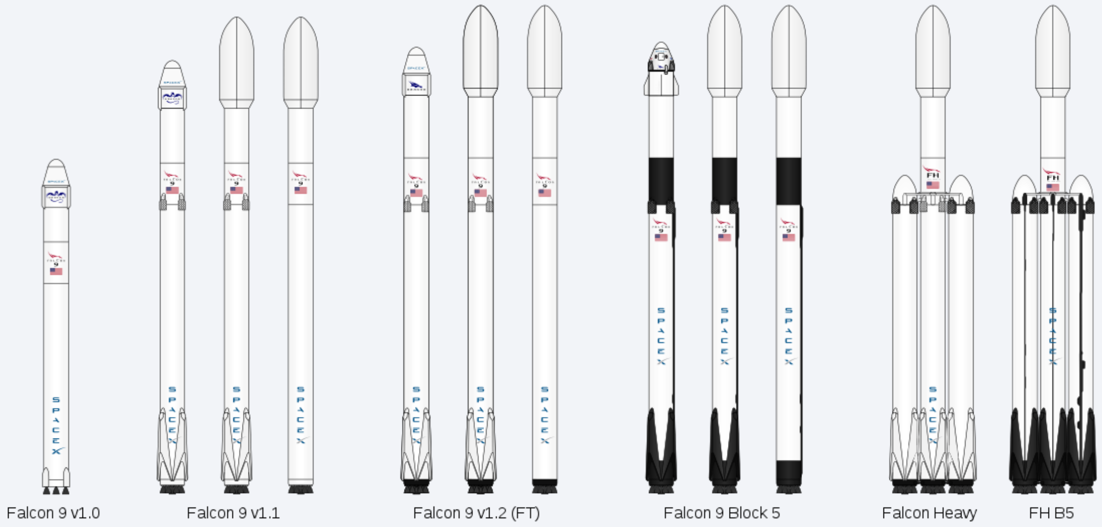
- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- Summary of methodologies
    - Data collection
    - Data wrangling
    - EDA with data visualization
    - EDA with SQL
    - Building interactive map with Folium
    - Building a dashboard app with Plotly
    - Predictive analysis(Classification)
- Summary of all results
    - Exploratory data analysis results
    - Interactive analytics demo in screenshots
    - Predictive analysis results

Introduction

- In this project we will predict if the Falcon 9 first stage will land successfully or not. SpaceX advertises its Falcon 9 rockets at cost $62 million which costs 165 million in other providers. Savings are because SpaceX reuses its first stage. This analysis is used if another company wants to compete with SpaceX.
- Problems to solve:
  - What influences to success of launches?
  - What are the important factors to consider when launching rockets?



1200px-Falcon9_rocket_family.svg.png

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Collect from [SpaceX REST API] and from Wikipedia
- Perform data wrangling
  - Perform EDA to find some patterns in data, determine labels
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Preprocess data, split data train and test sets, perform Grid Search, test different models

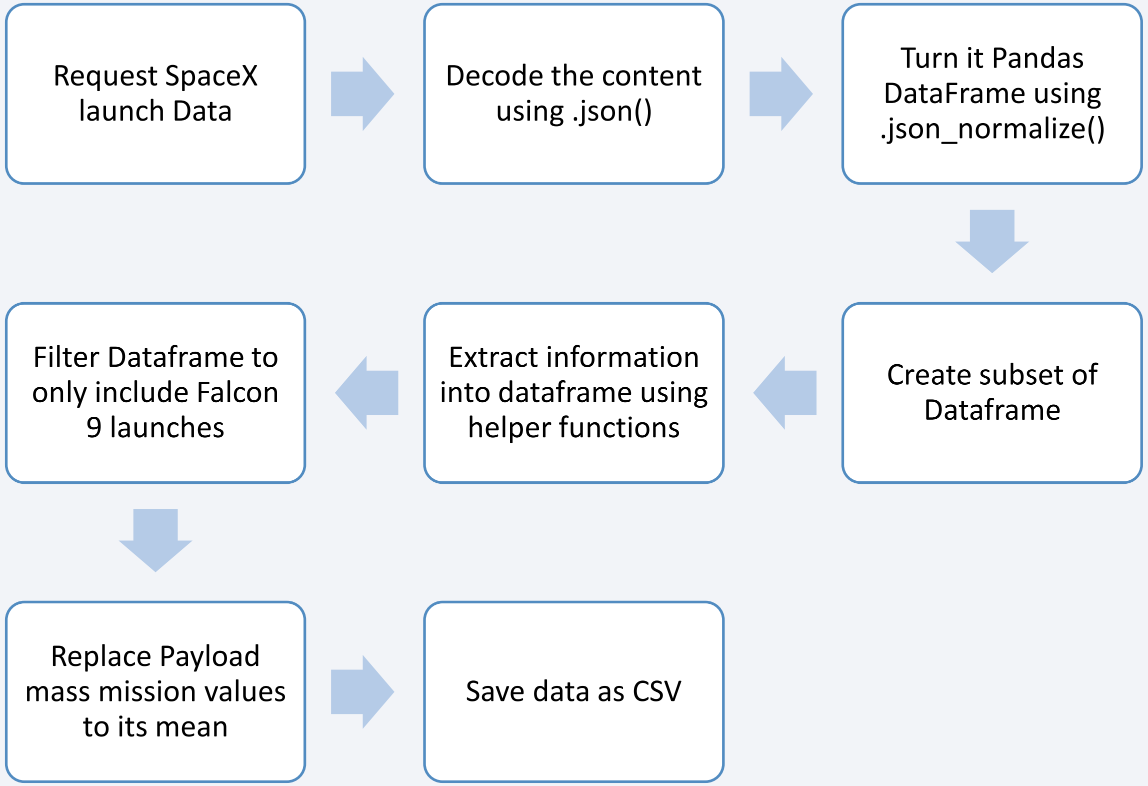
Data Collection

- Data will be collected using SpaceX RESR API and using Web Scraping from Wikipedia.
- Some data cleaning will be performed during data collection.

Request Data from Rest API → Web scraping from Wikipedia → Perform some cleaning → Save datasets as CSV file

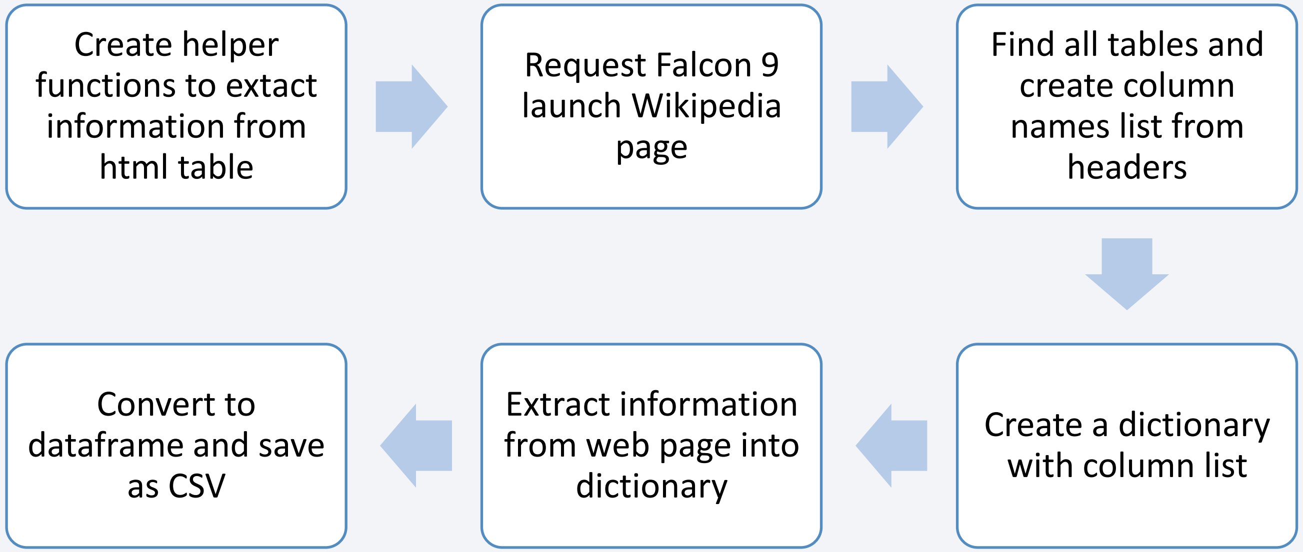
Data Collection – SpaceX API

- Create helper functions to extract information about Booster Version, Launch Site, Payload Mass, Orbit etc.
- GitHub link for [Jupyter Notebook]



Data Collection - Scraping

- Perform web scraping to collect Falcon 9 historical data from Wikipedia page – “List of Falcon 9 and Falcon Heavy launches”
- GitHub link for [Jupyter Notebook]



Data Wrangling

- Perform some EDA and determine labels for supervised learning
  - check for null values
  - Identify column data types(numerical, categorical)
  - Create landing outcome label and assign values 1, 0 for success and fail
  - Export data as csv
- GitHub link for [Jupyter Notebook]

EDA with SQL

- Display the names of the unique launch sites
- Display records where launch site name starts with 'CCA'
- Display the total payload mass of boosters launched by NASA (CRS)
- Display the average payload mass of booster version F9 v1.1
- List the data when first successful landing outcome in ground pad
- List the names of boosters which have success in drone ship and payload mass between 4000 and 6000
- List the total number of successful and failure launches
- List the names of booster versions which have carried maximum payload mass
- List failure landing outcomes in drone ship for year 2015
- Rank the successful landing outcomes
- GitHub link for [Jupyter Notebook]

EDA with Data Visualization

- Plot Flight Number vs Payload Mass to check its importance on successful launch outcome
- Plot launch site and Payload Mass if there is a relationship
- Bar chart to visualize the relationship between success rate and orbit type
- Scatter plot to explore relationship between flight number and orbit type
- Scatter plot to explore relationship between payload mass and orbit type
- Visualize the yearly success trend
- GitHub link for [Jupyter Notebook]()

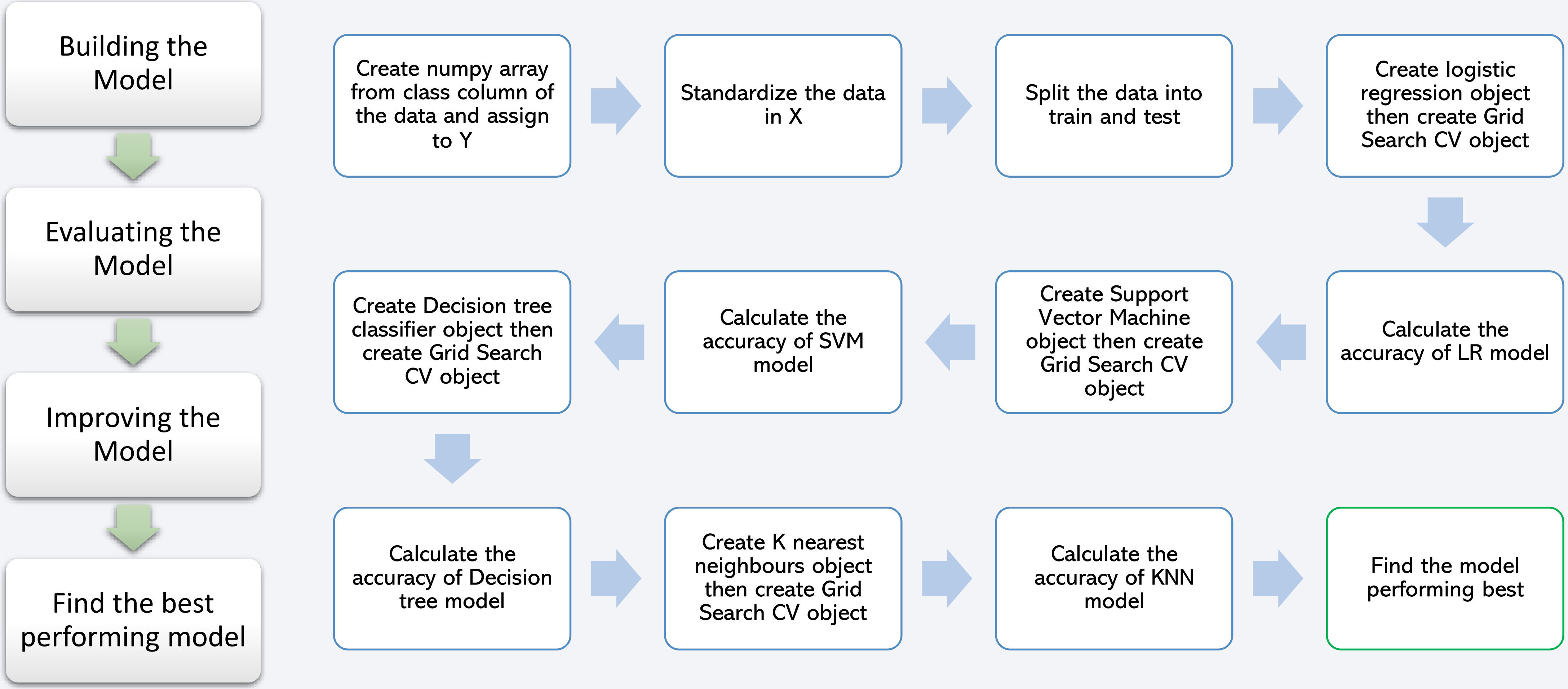
Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Mark all launch sites on Map with circles
- Mark successful and failure launches in each site on Map
- Calculate the distance between a launch site and its proximities and mark with lines
- GitHub link for [Jupyter Notebook]

Build a Dashboard with Plotly Dash

- Add launch site dropdown input component to check each site's success rate
- Add callback function to render pie chart for selected site
- Add range slider to select payload mass
- Add callback function to render scatter chart for selected site and payload mass
- GitHub link for Plotly dash application

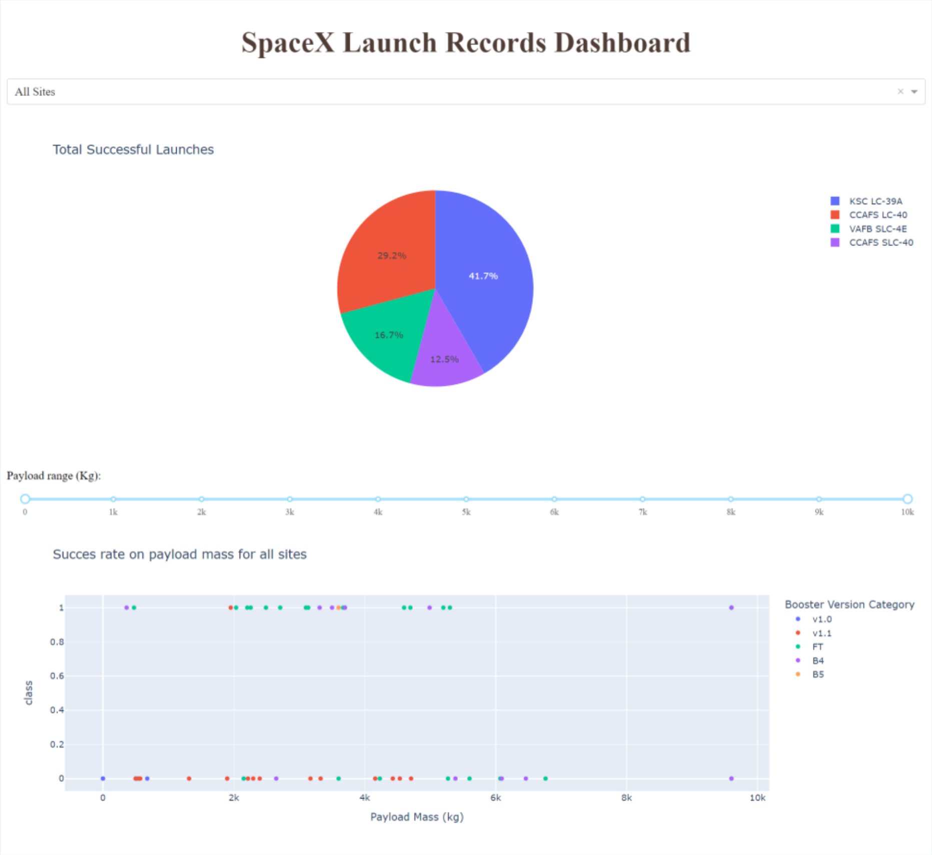
Predictive Analysis (Classification)



GitHub link for [Jupyter Notebook]

Results

- Selected labels: Flight number, Date, Booster Version, Payload Mass, Orbit, Launch Site, Outcome etc.
- Best performing model: Support vector machine and Decision tree

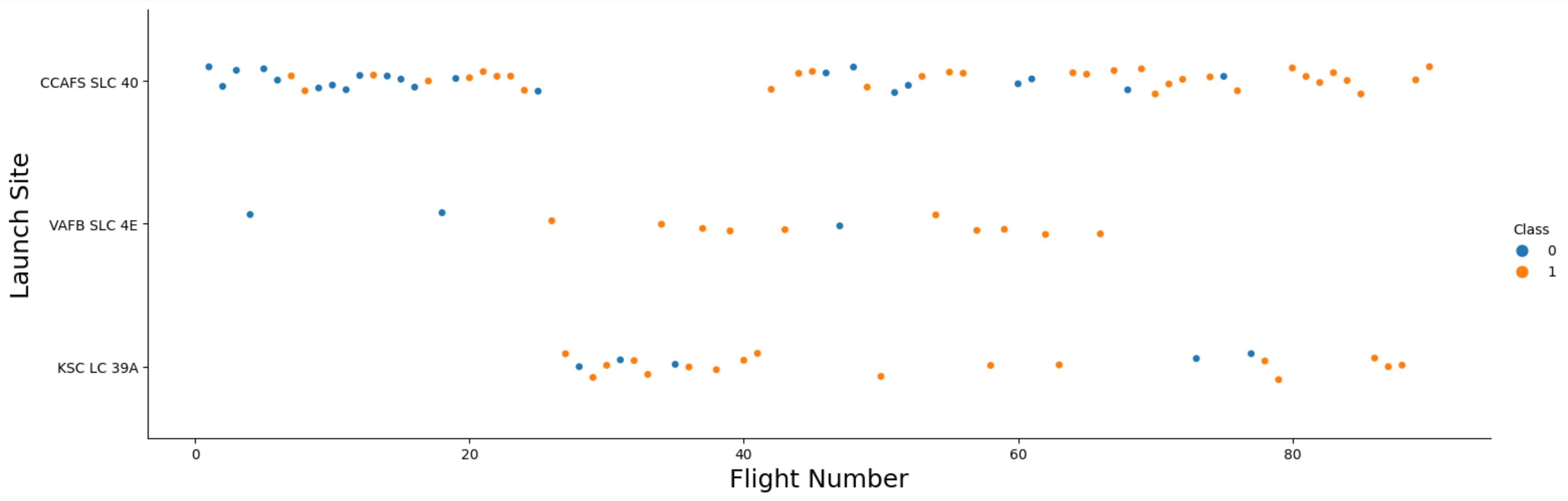


Section 2

Insights drawn from EDA

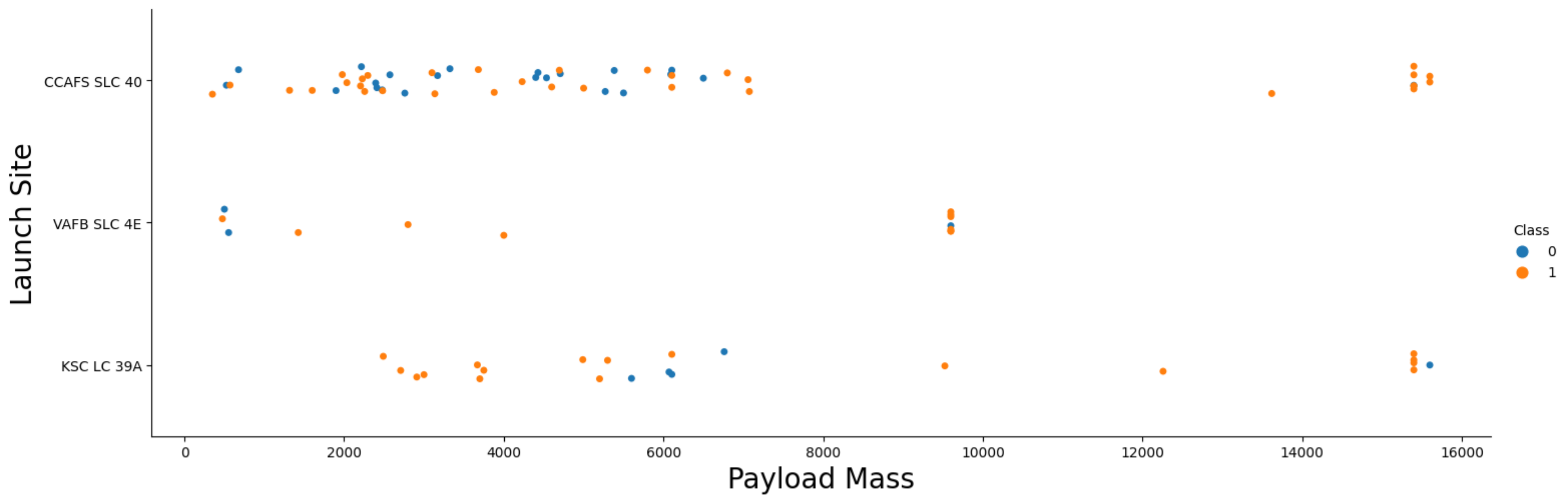
Flight Number vs. Launch Site

- As flight number increases, number of successful launches also increases in each launch site.
- The most of the launched occurred in CCAFS SLC 40



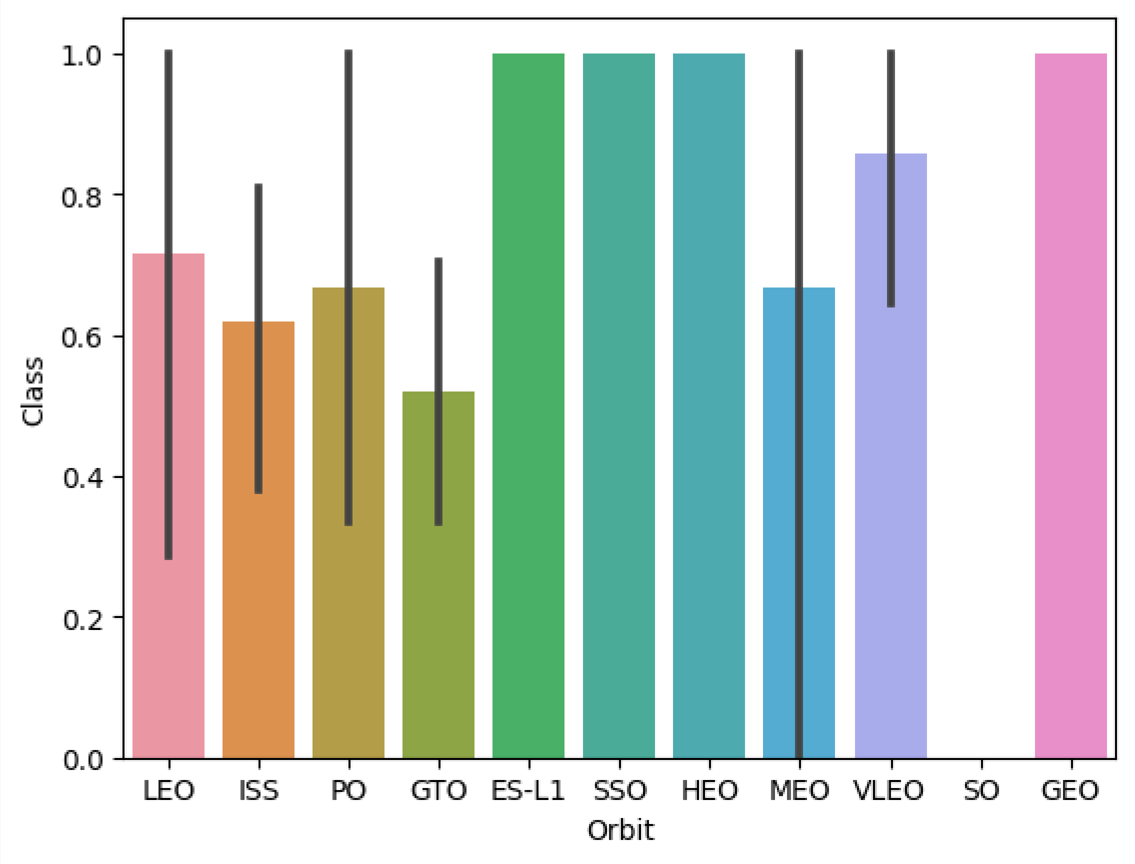
Payload vs. Launch Site

- More launches in CCAFS SLC 40 with payload mass between 1000 and 7000, and launches with around 15000 payload mass was all successful
- In KSC LC 39A launch site, number of successful launches decreases with around 6000 kg payload mass
- For VAFB SLC 4E, there were not launches after 10000 kg payload mass



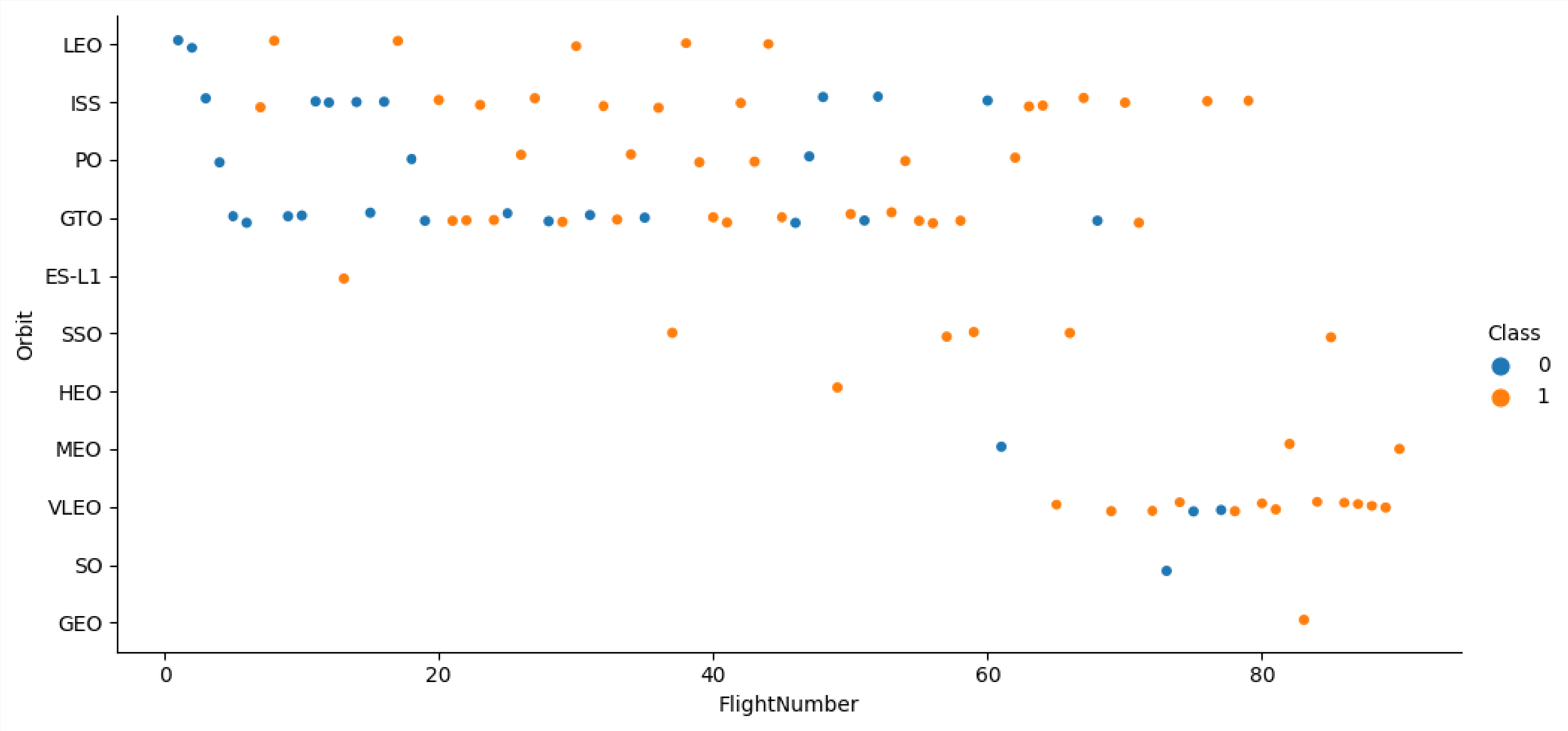
Success Rate vs. Orbit Type

- Success rate is 100% for ES-L1, SSO, HEO and GEO
- Lowest success rate is launches for GTO, about 50%



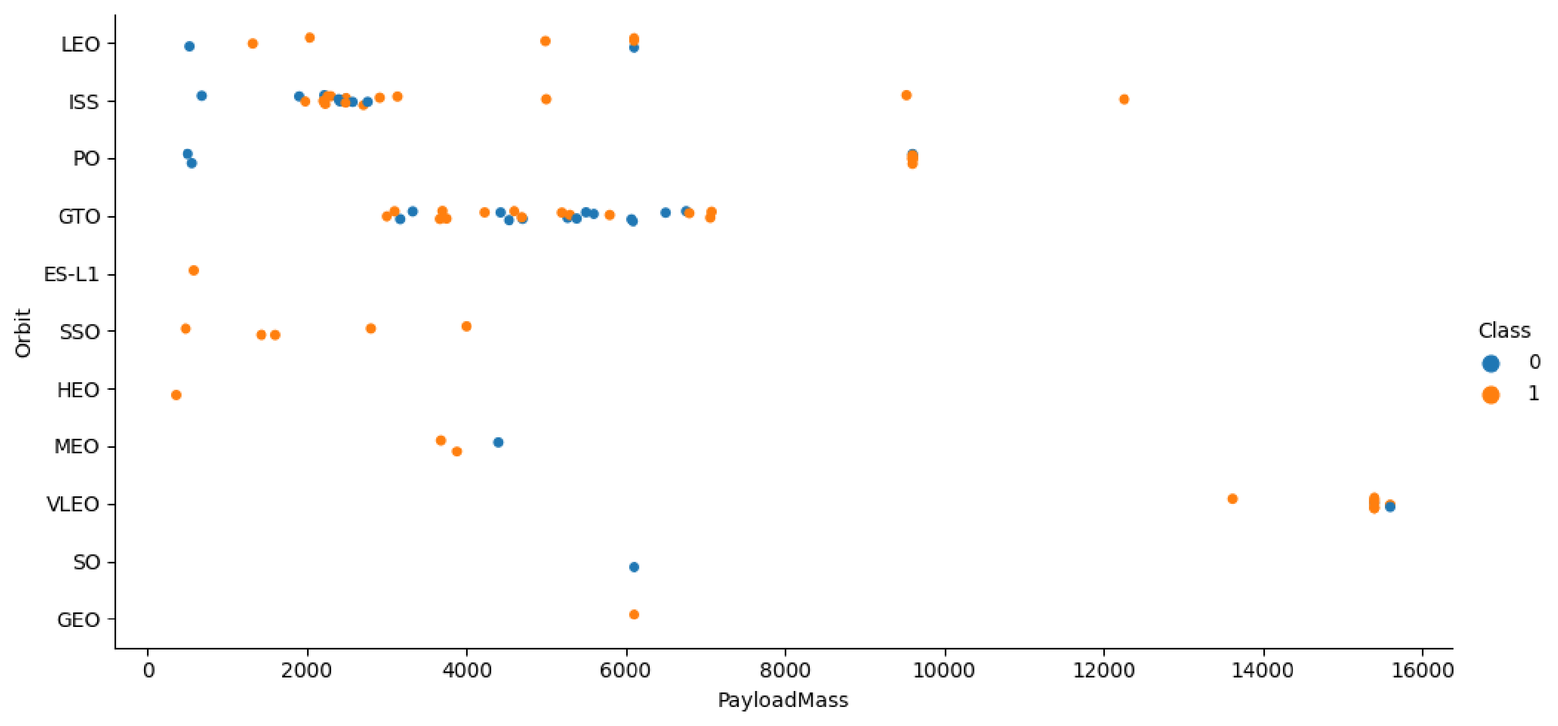
Flight Number vs. Orbit Type

- For GTO orbit, there is no relationship with flight number
- LEO orbit success rate is related to flight number



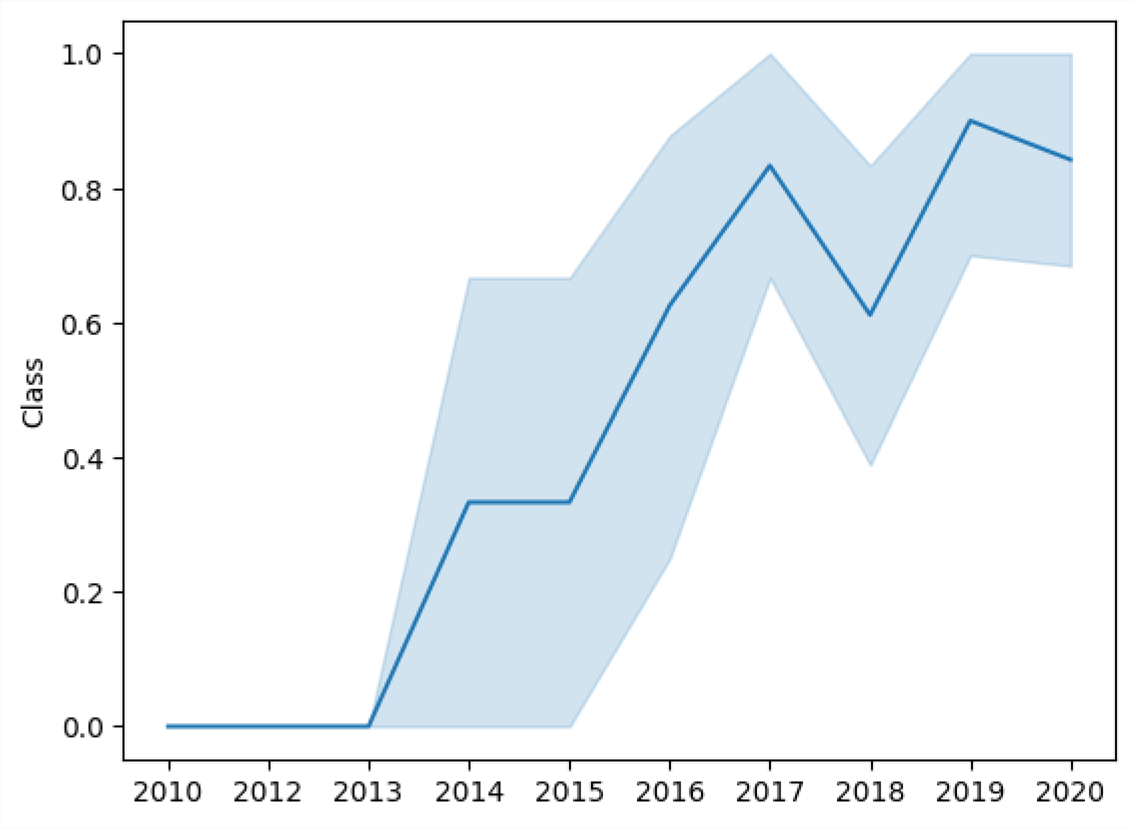
Payload vs. Orbit Type

- For VLEO, LEO, ISS, SSO, success rate is higher with heavy payload masses
- For GTO, it is difficult to distinguish because it has both positive and negative launches together



Launch Success Yearly Trend

- Success rate of launches increased gradually by the year of 2017
- Decreased sharply in 2018 and increased again to about 90% in 2019(highest success rate)



All Launch Site Names

- There are four launch sites:
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
  - VAFB SLC-4E

```
[4]: %%sql
        select DISTINCT launch_site from SPACEXTBL

      * ibm_db_sa://nsr36690:***@2d46b6b4-cbf6-40eb-bbce
     32328/bludb
     Done.
```

[4]:

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://nsr36690:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/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

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

```
%%sql
select sum(payload_mass__kg_) as total_payload_mass from SPACEXTBL
where customer = 'NASA (CRS)'
```

```
 * ibm_db_sa://nsr36690:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs
Done.
```

total_payload_mass
45596

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1: 2928kg

```
%%sql
select avg(payload_mass__kg_) as average_payload_mass from SPACEXTBL
where booster_version = 'F9 v1.1'
```

```
 * ibm_db_sa://nsr36690:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io
Done.
```

average_payload_mass
2928

First Successful Ground Landing Date

- The first successful landing outcome on ground pad was in December 22, 2015

```
%%sql
select DATE, landing__outcome from SPACEXTBL
where landing__outcome = 'Success (ground pad)'
order by DATE
limit 1
```

```
 * ibm_db_sa://nsr36690:***@2d46b6b4-cbf6-40eb-bbce-6251
Done.
```

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

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 are F9 FT B1022, F9 FT B1026, F9 FT B1021.2, F9 FT B1031.2

```
%%sql
select booster_version, landing__outcome, payload_mass__kg_ from SPACEXTBL
where landing__outcome = 'Success (drone ship)' and payload_mass__kg_ between 4000 and 6000
```

```
 * ibm_db_sa://nsr36690:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.database
Done.
```

booster_version	landing__outcome	payload_mass__kg_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes: 71

```
%%sql
select count(landing__outcome) as total_number from SPACEXTBL
where landing__outcome like 'Failure%' or landing__outcome like 'Success%'
```

```
 * ibm_db_sa://nsr36690:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08k
Done.
```

total_number
71

Boosters Carried Maximum Payload

- There are 12 boosters which have carried the maximum payload mass

```
%%sql
select booster_version, payload_mass__kg_ from SPACEXTBL
where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXTBL)
```

```
 * ibm_db_sa://nsr36690:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l
Done.
```

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

2015 Launch Records

- There was two failed landing in drone ship in year 2015

```
%%sql
select DATE,booster_version, launch_site, landing__outcome from SPACEXTBL
where landing__outcome = 'Failure (drone ship)' and DATE like '2015%'
```

```
 * ibm_db_sa://nsr36690:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08k
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

- Number of Failures and successes in drone ship are same

```
%%sql
select landing__outcome, count(landing__outcome) as total_number from SPACEXTBL
where DATE between '2010-06-04' and '2017-03-20'
group by landing__outcome
order by total_number desc
```

```
 * ibm_db_sa://nsr36690:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od
Done.
```

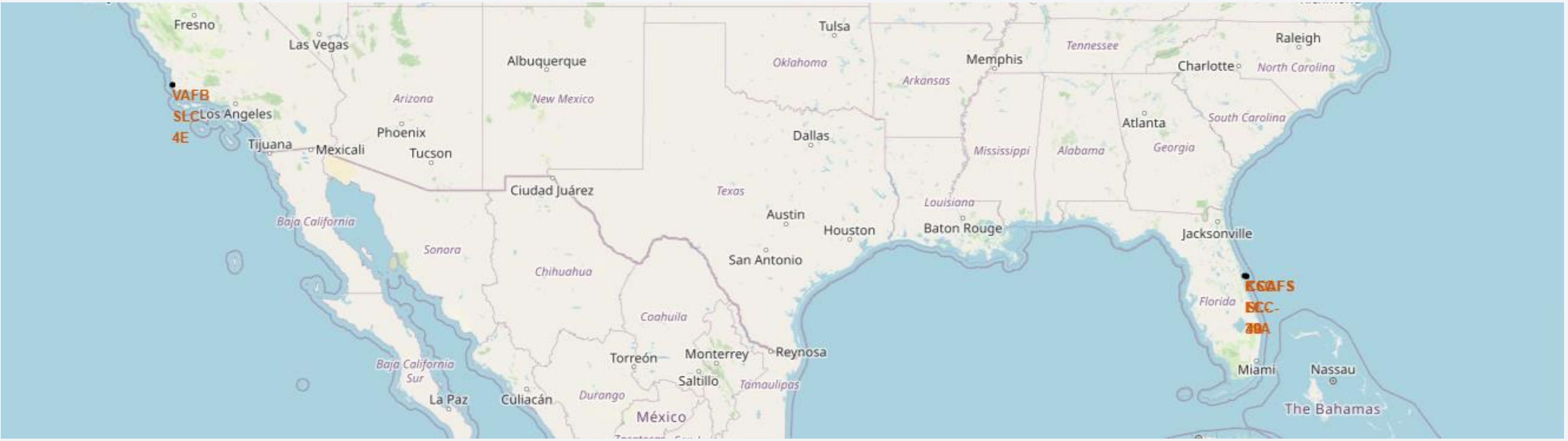
landing__outcome	total_number
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

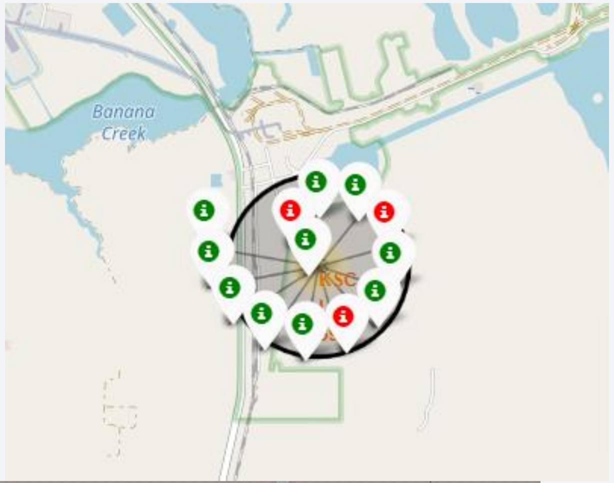
Locations of all launch sites

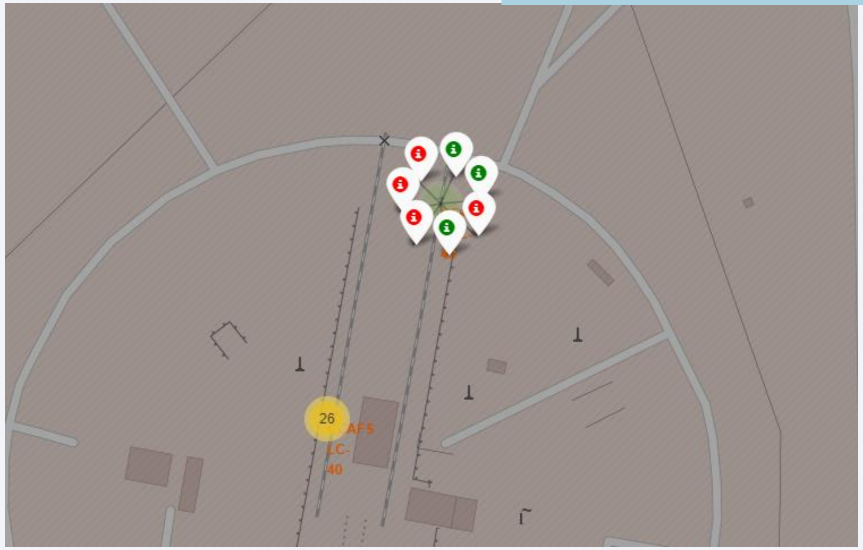
- There 4 launch sites - CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E
- CCAFS LC-40, CCAFS SLC-40, KSC LC-39A are located close to each other in Florida
- VAFB SLC_4E is located in California

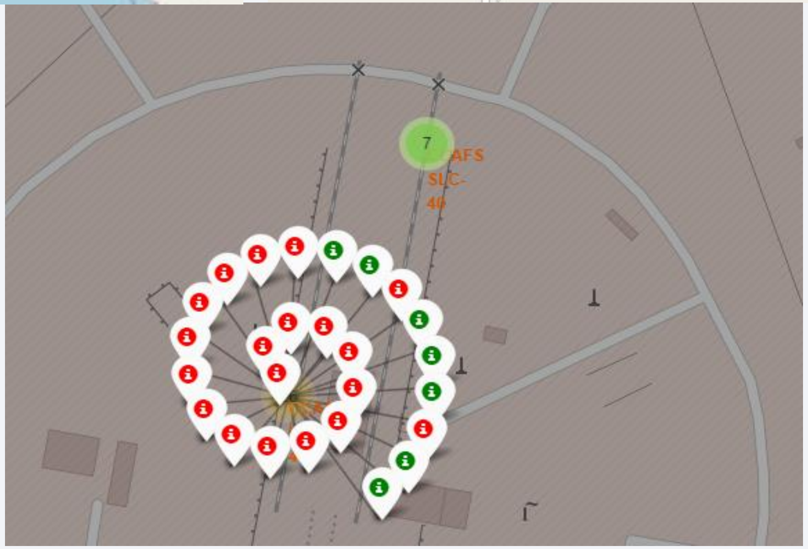


Launch outcomes on Map

- The most launches and failed launches was in CCAFS LC-40
- The most successful launches was in KSC LC-39A







Launch site proximities

KSC LC-39A launch site and its proximities

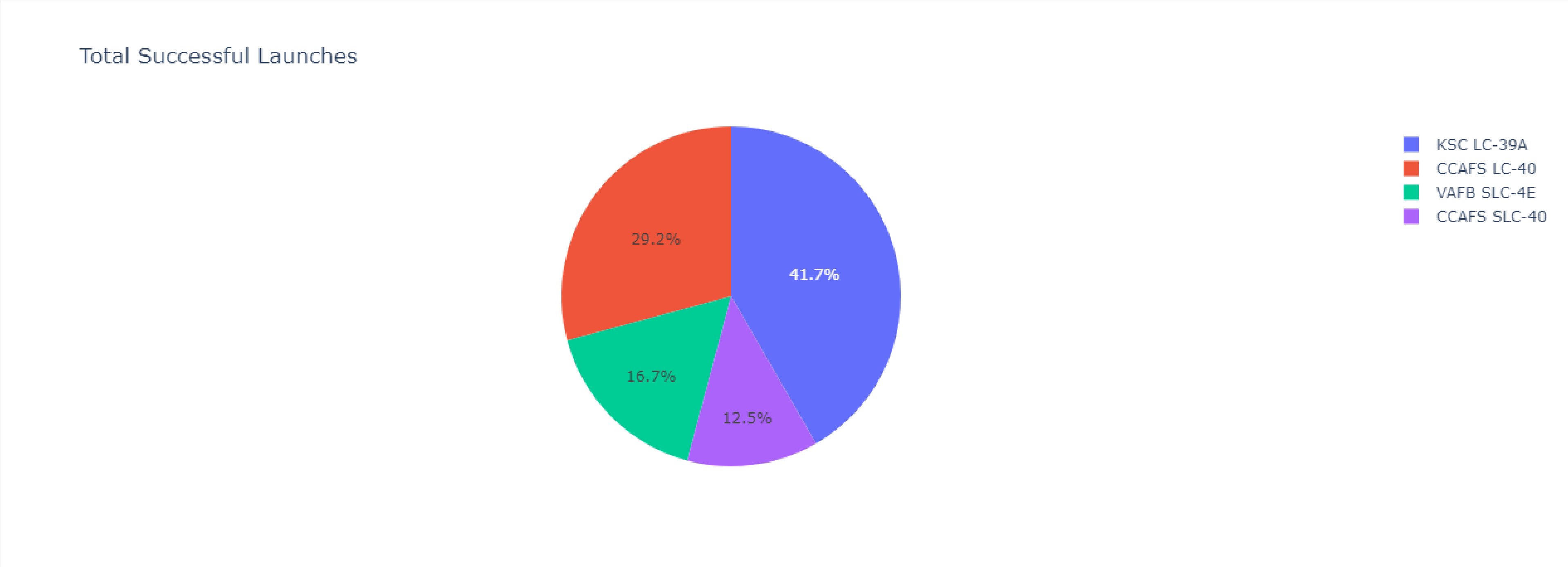
Highway, railway and coastline is very close to its location

Section 4

Build a Dashboard with Plotly Dash

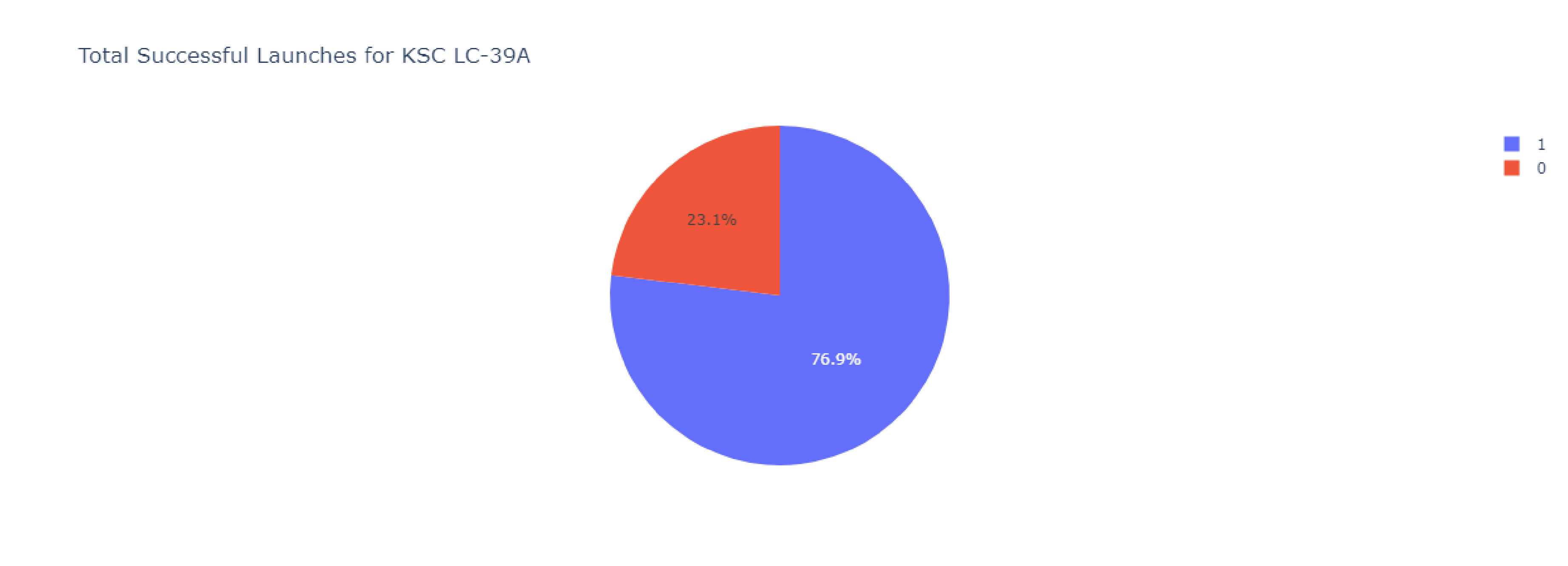
Launch Success rate

- The most successful launches occurred in KSC LC-39A, about 41.7%
- Launches in CCAFS SLC-40 had the lowest success rate, about 12.5%



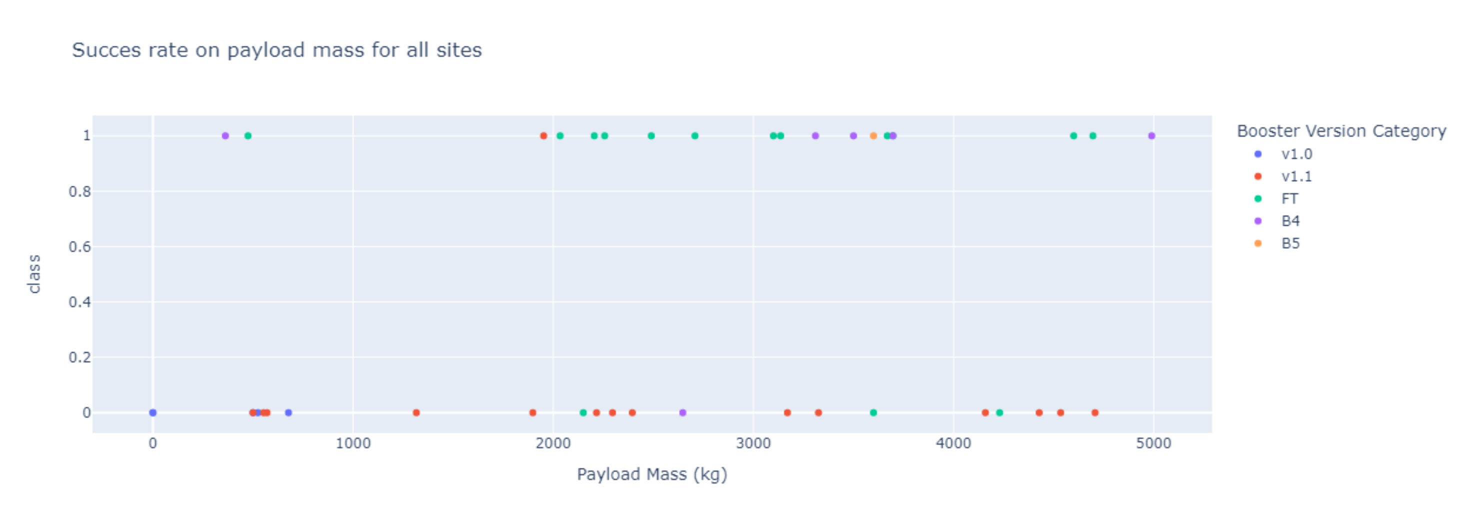
Launch site with highest success rate

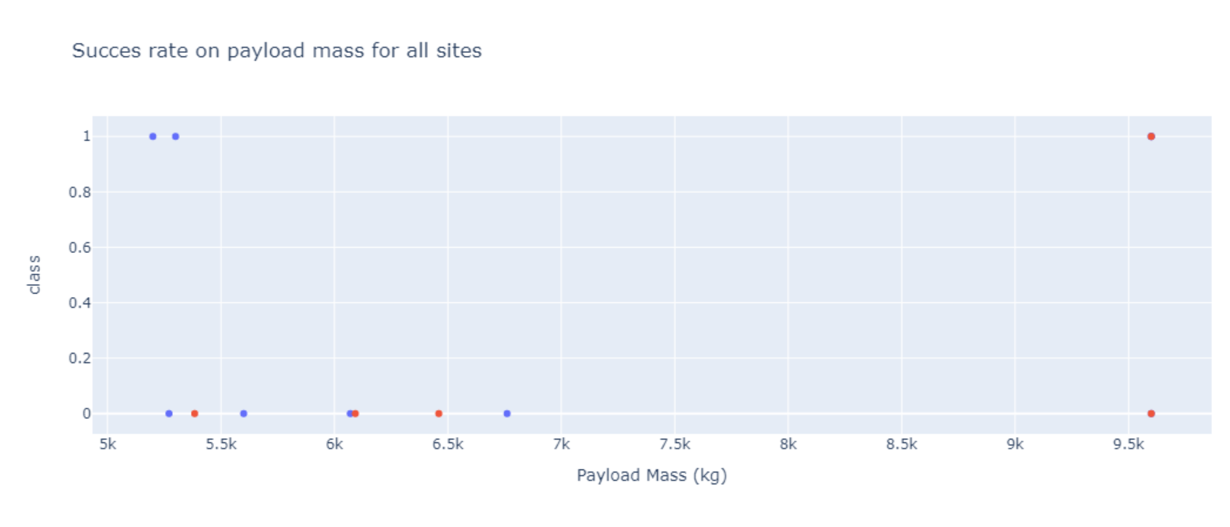
- Success rate: 76.9%
- Failure rate: 23.1%



Success rate vs Payload Mass

- In range between 0kg and 5000kg payload mass, the highest success rate is for booster version FT
- There are only launches with booster version FT and B4 between 5000 and 10000 kg payload mass.



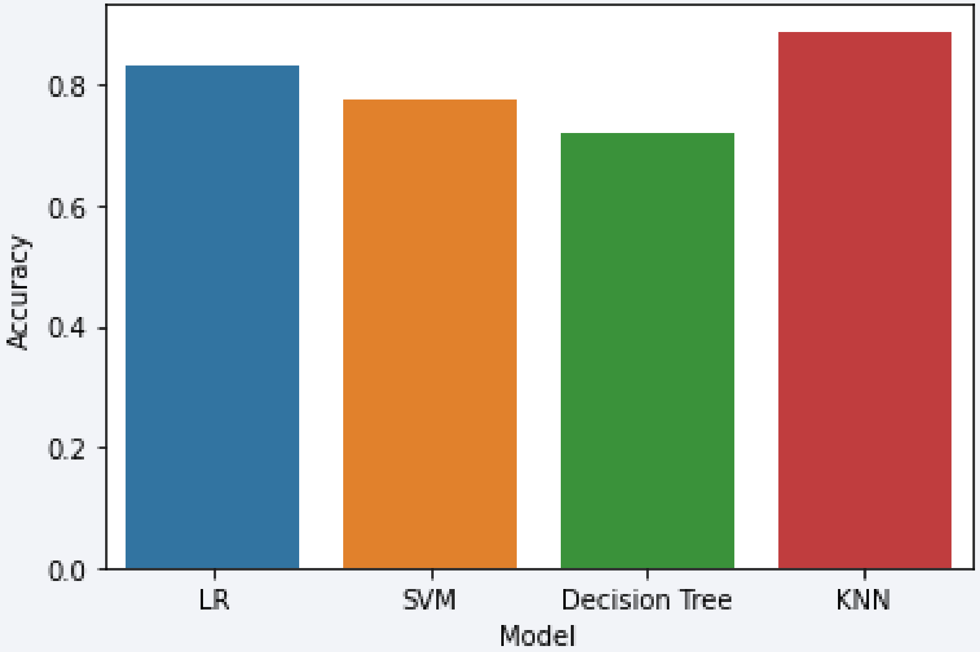


Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Highest classification accuracy is about 89% in K-nearest neighbors model



Confusion Matrix

- True positive: 7 successful launch outcomes predicted correctly as ‘landed’
- False positive: 1 failed launch outcome predicted wrongly as landed
- True negative: 4 successful launch outcomes predicted wrongly as did not land
- False negative: 6 failed launch outcomes predicted correctly as did not land

Conclusions

- Success rate has relationship with payload mass
- Most successful launches was for ES-L1, SSO, HEO, GEO Orbits
- LEO Orbit success rate has relationship with flight number
- Launches with heavy payloads(5000~) mostly launched to VLEO, PO and ISS
- Success rate of launches increased until 2017 and decreased in 2018
- Number of total successful and failure launches: 71
- Maximum payload mass: 15600kg
- Most launch sites located in Florida
- KSC LC-39A has the highest success rate, about 41.7%. 76.9% of its launches was successful
- Best classification model to predict launch outcome is KNN with about 89% accuracy
- Full project GitHub

Thank you!