IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Keira Sanders
03/06/2022

Outline

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

Executive Summary

In the course of this project, historical SpaceX launch and booster data was analyzed to determine the most effective launch parameters over time. A predictive model was built which can be fed launch parameters from a competing rocket company to determine if the first rocket stage will land successfully.

As expected, SpaceX's landing outcomes improved over time as they tuned booster versions, launch site locations, payload mass, and mission orbit. KSC (Kennedy space center) became the most successful launch site and a payload mass between 2500 – 5000 kg resulted in the most successful landings, especially when launched to an SSO orbit target.

Recommend that SpaceY build upon SpaceX's improvements over time by starting launches within the most successful payload mass range, from Kennedy Space Center in Florida.

Introduction

SpaceX is a leader in the commercial space flight space. By returning the first stage successfully to the ground, they achieve significant savings over non-commercial launches. Utilizing past SpaceX launch data, the goal is to develop a predictive model that will predict if the first stage will land after a successful launch. This can be used to inform and guide development of launch techniques and maximum payload for a competing rocket company.

Section 1

Methodology

Methodology

Executive Summary

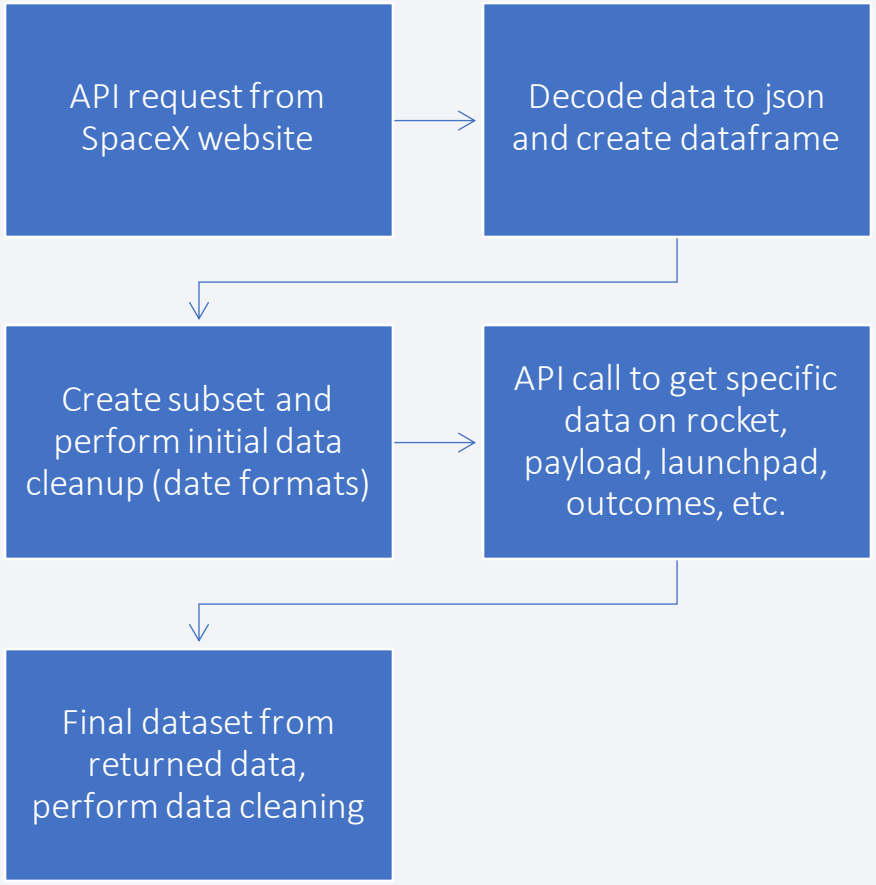
- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

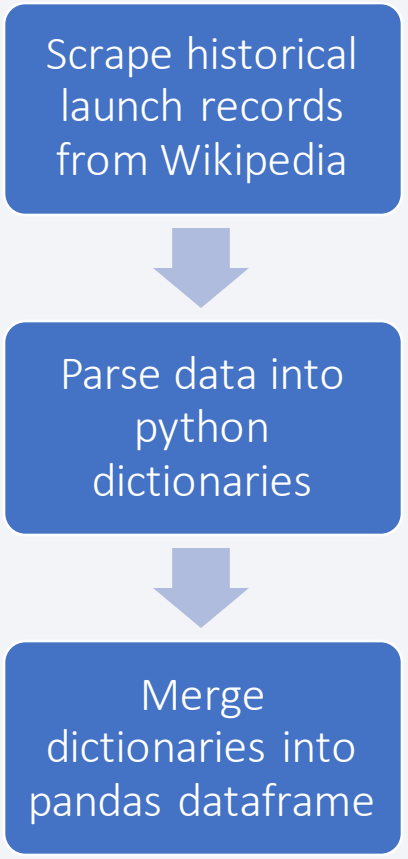
- Data sets were collected from 2 sources:
  - SpaceX flight data via API
  - Wikipedia data on historical launch outcomes, payload mass, etc.

Data Collection – SpaceX API

- Focusing on Falcon 9 launches, removed Falcon 1 data from subset
- Filled relevant missing values for payload mass
- https://github.com/keirasanders1/IBM_ds_capstone/blob/25dd8a916ed551ec15cca0c9a8891069543c1e09/jupyter-labs-spacex-data-collection-api.ipynb

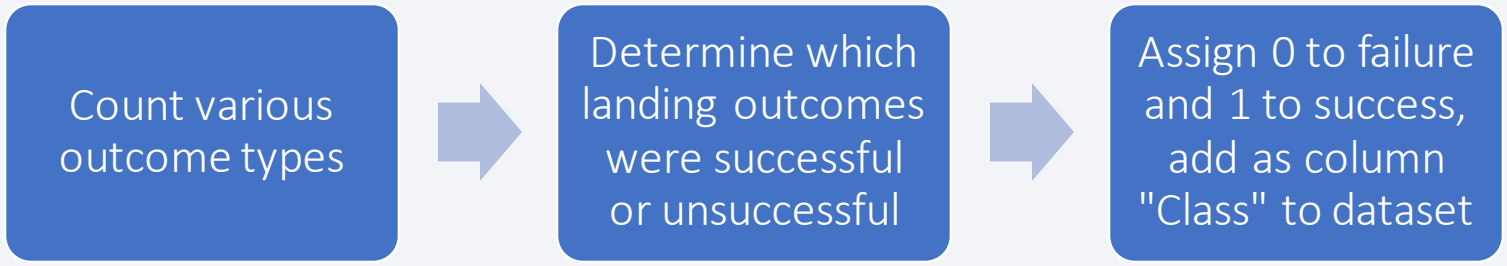


Data Collection - Scraping



- Scraped historical launch data from Wikipedia using Beautiful Soup
- Extract data into python dictionaries with table column names as keys
- Merge into pandas dataframe
- https://github.com/keirasanders1/IBM_ds_capstone/blob/25dd8a916ed551ec15cca0c9a8891069543c1e09/jupyter-labs-webscraping.ipynb

Data Wrangling



- https://github.com/keirasanders1/IBM_ds_capstone/blob/25dd8a916ed551ec15cca0c9a8891069543c1e09/labs-jupyter-spacex-Data_wrangling.ipynb

EDA with Data Visualization

- Various charts were plotted to determine relationships between variables, as well as an analysis of success rate.
- Scatter plots of:
  - Flight number vs. Payload mass
  - Flight number vs. Launch site
  - Payload mass vs. Launch site
  - Flight number vs. Orbit type
  - Payload mass vs. Orbit type
- Bar plot of success rate by Orbit
- Line graph of success rate by Year
- https://github.com/keirasanders1/IBM_ds_capstone/blob/25dd8a916ed551ec15cca0c9a8891069543c1e09/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- Queries performed:
  - List of unique launch sites
  - Data where launch site contains 'CCA'
  - Total payload carried by NASA launched boosters
  - Average payload mass carried by booster version F9 v1.1
  - Date of first successful landing outcome on a ground pad
  - Names of boosters successfully landed on drone ship with payload mass > 4000 < 6000
  - Total successful missions and total mission failures
  - Booster versions that have carried the max payload
  - Drone ship landing failures during 2015
  - Total count of unique landing outcomes between 2010-06-04 and 2017-03-20
- https://github.com/keirasanders1/IBM_ds_capstone/blob/25dd8a916ed551ec15cca0c9a8891069543c1e09/jupyter-labs-eda-sql-coursera.ipynb

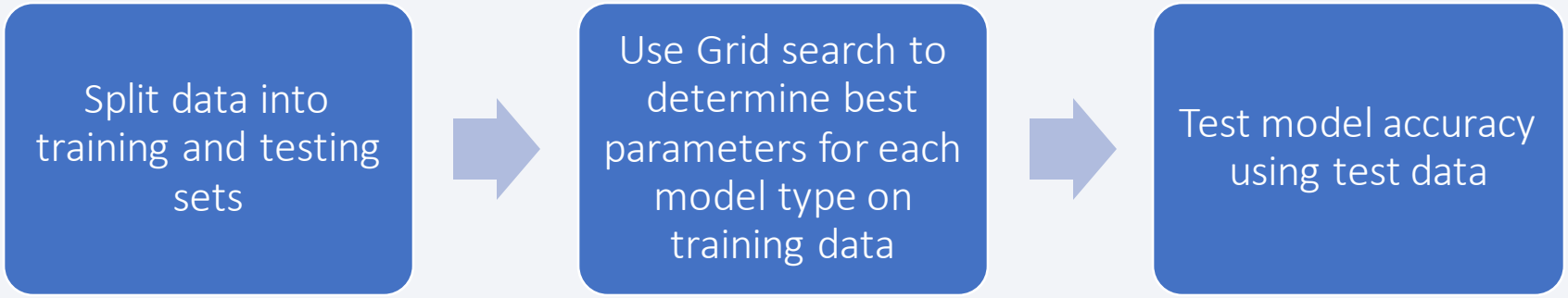
Build an Interactive Map with Folium

- Each SpaceX launch site was mapped, using landing outcome success/failure as markers, allowing for visualization of most successful launch site.
- Determined site distance from various infrastructure: cities, highways, coast, & rail to analyze launch site characteristics
- https://github.com/keirasanders1/IBM_ds_capstone/blob/25dd8a916ed551ec15cca0c9a8891069543c1e09/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Launch success rate by launch site to determine most successful launch site
- Launch outcome by payload mass for each booster type to determine what payload mass had most successful landings for each booster type
- https://github.com/keirasanders1/IBM_ds_capstone/blob/25dd8a916ed551ec15cca0c9a8891069543c1e09/spacex_dash_app.py

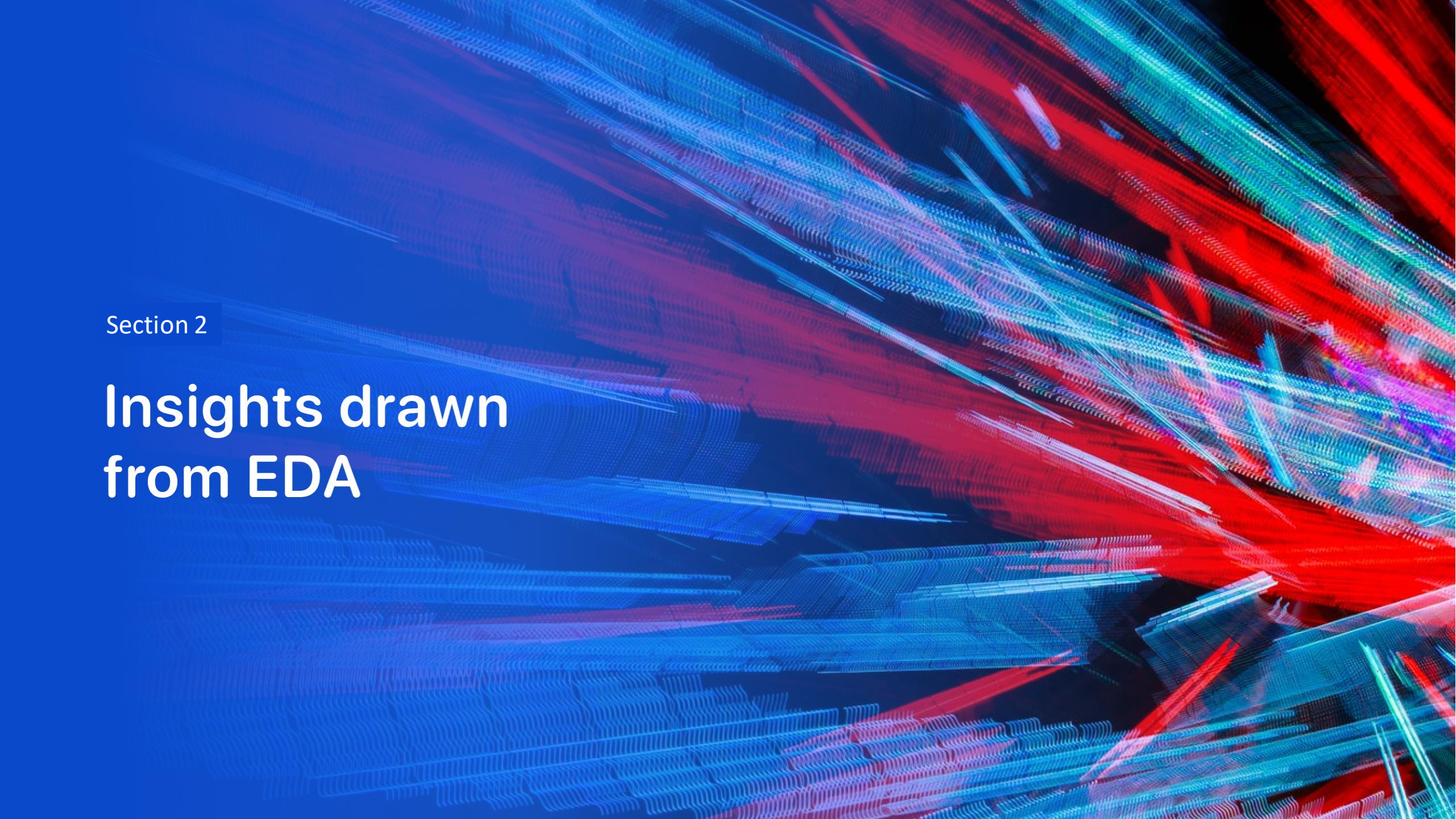
Predictive Analysis (Classification)



- Utilized logistic regression, SVM, decision tree, and KNN classification models to predict landing outcomes.
- https://github.com/keirasanders1/IBM_ds_capstone/blob/7308912ac39649601099867d4864643dee63e6f1/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

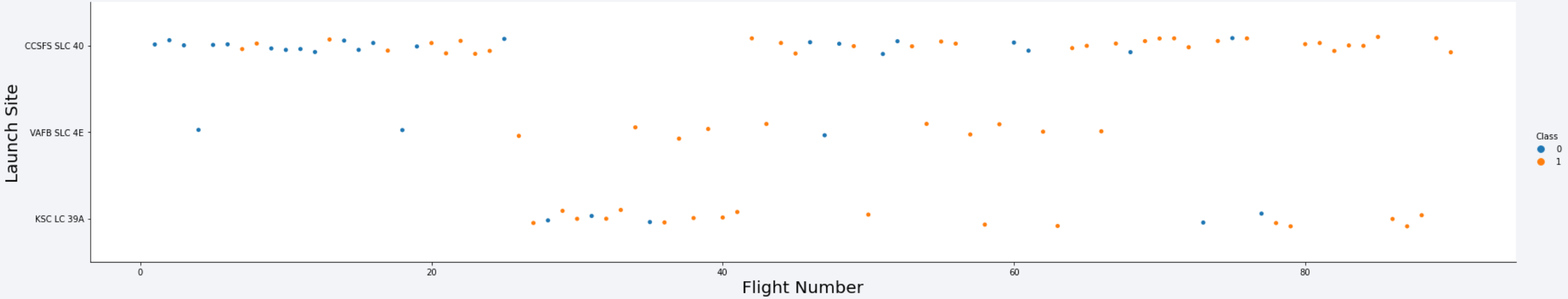
Section 2

Insights drawn from EDA



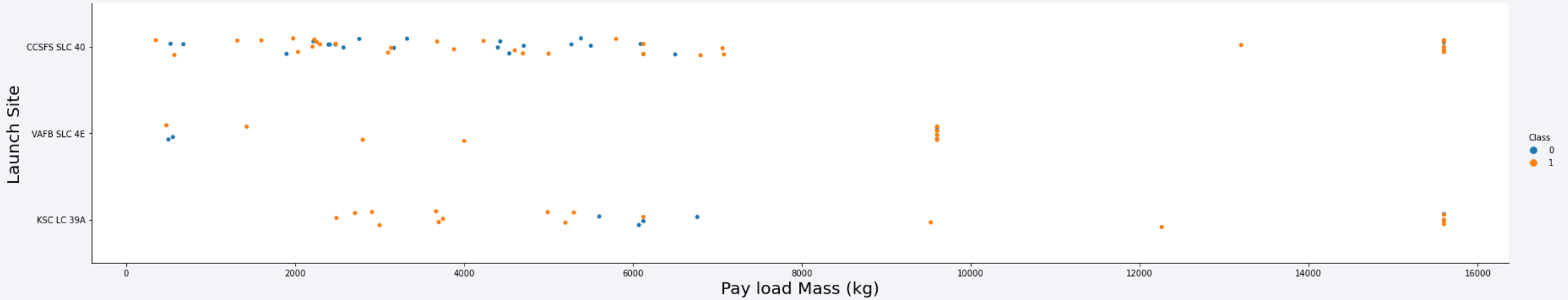
Flight Number vs. Launch Site

- Improved performance over time at launch site CCSFS.
- SpaceX focused on using CCSFS and KSC launch sites past launch #70



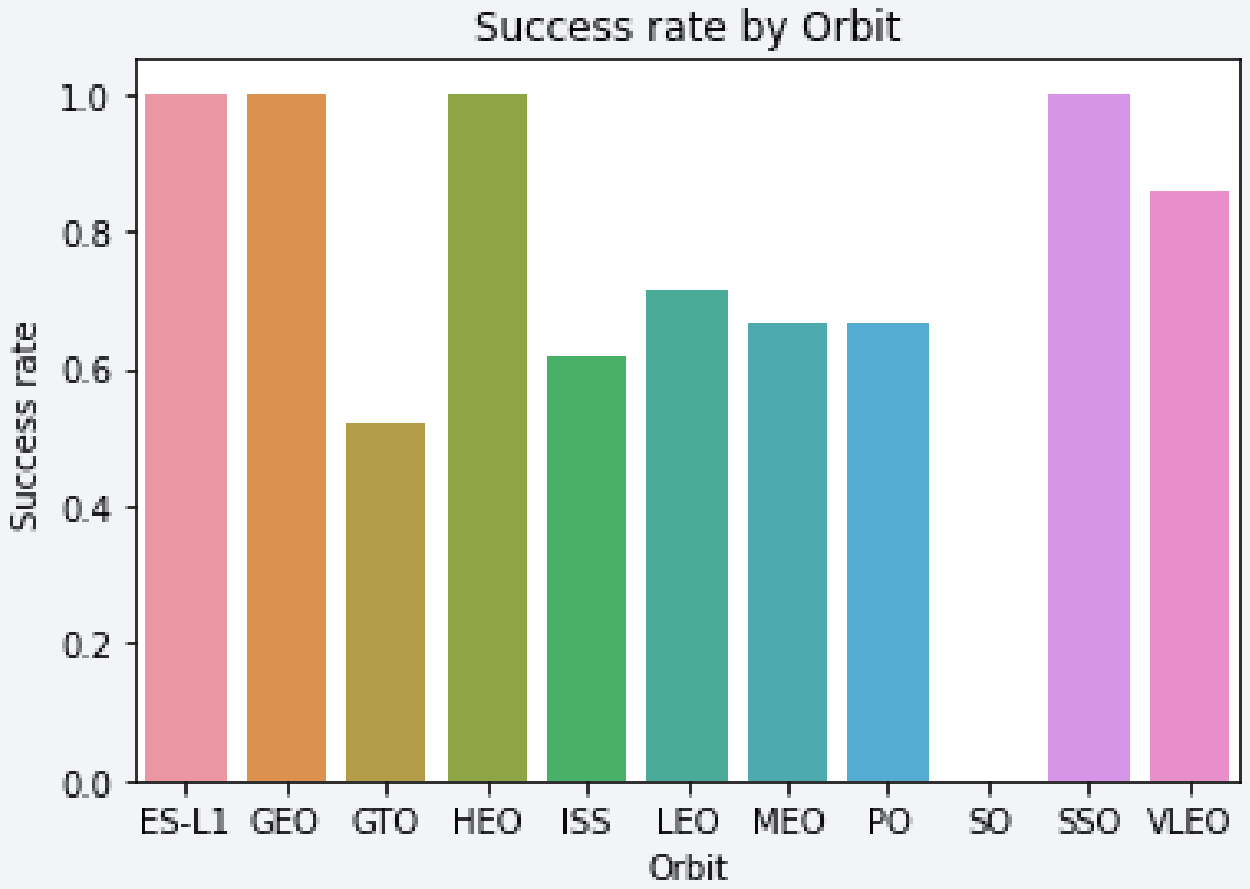
Payload vs. Launch Site

- No launches at site VAFB with payload > 10,000 kg.
- Payload mass > 8000 kg all launches have been successful

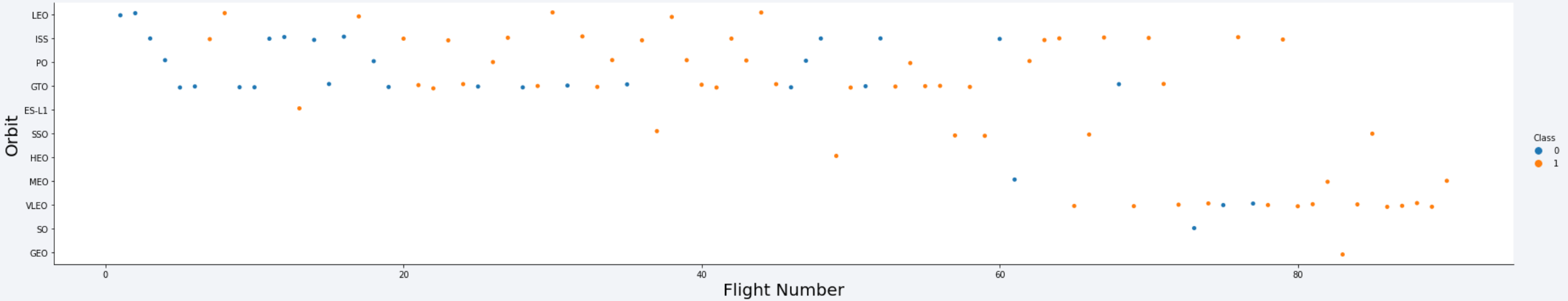


Success Rate vs. Orbit Type

- Orbit types ES-L1, GEO, HEO, and SSO most successful (100% launch success)
- VLEO also high success rate (>80%)
- GTO least successful (~50% successful launches)

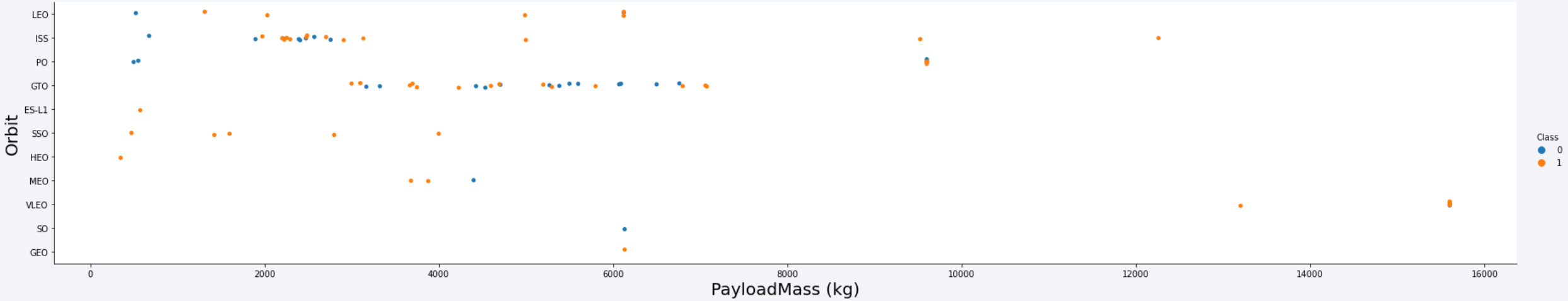


Flight Number vs. Orbit Type



- LEO performance improvement over time
- GTO does not appear to have approved over time
- VLEO more recent launches

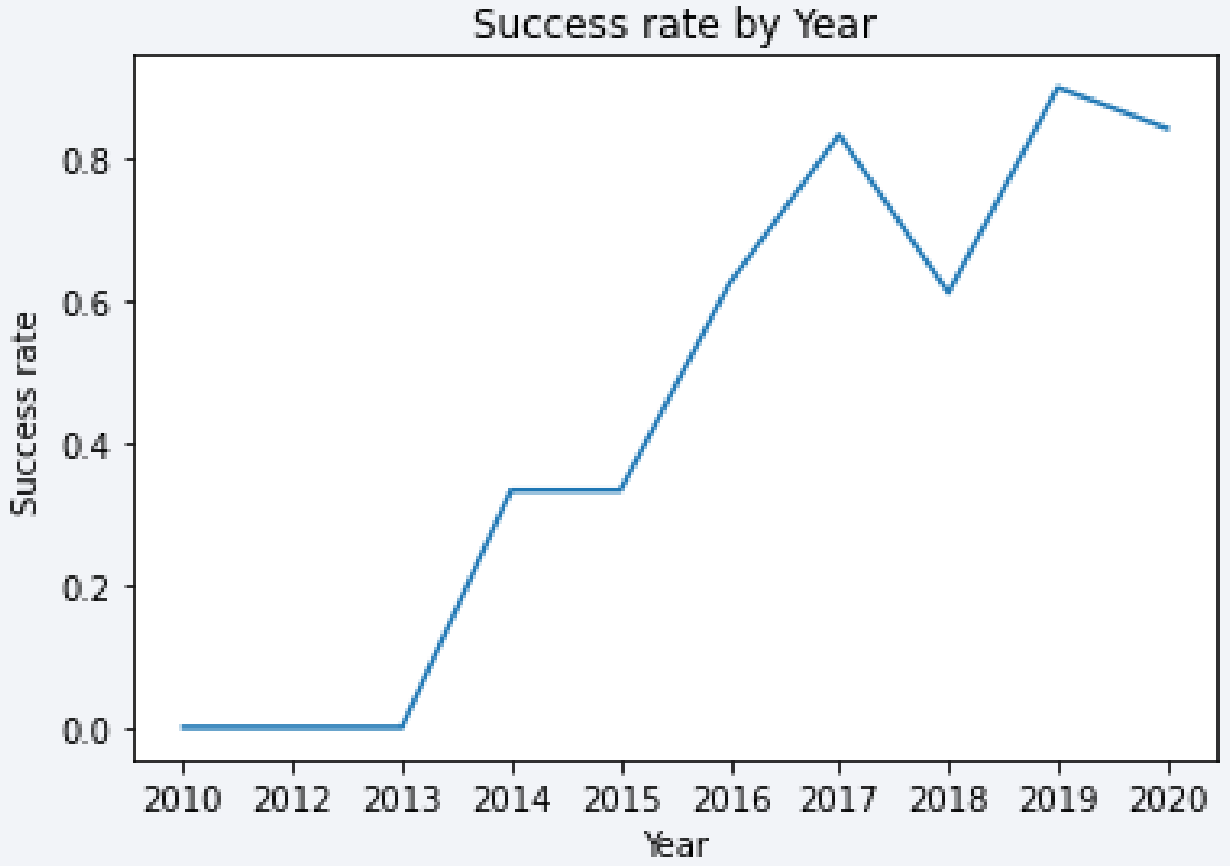
Payload vs. Orbit Type



- Launches to LEO and ISS orbits more successful with higher payloads
- Launches to SSO orbit all successful and all lower payloads (0 – 4000kg)
- Launches to VLEO heaviest payloads w/ good success rate.

Launch Success Yearly Trend

- Launch success rate increased steadily since 2013 until slight dip in 2018.
- As of 2020, launch success rate > 80%



All Launch Site Names

- All Space X launch sites utilized.

In [5]:
```
%%sql
select distinct launch_site from SPACEXDATASET
```

```
 * ibm_db_sa://jts26188:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databas
es.appdomain.cloud:32536/bludb
Done.
```

Out[5]:

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

Launch Site Names Begin with 'CCA'

In [9]:
```
%%sql
select * from SPACEXDATASET where launch_site like 'CCA%' limit 5
```

```
 * ibm_db_sa://jts26188:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

Out[9]:

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landi
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failu
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	Failu
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	

- Top 5 dataset records from the launch sites CCAFS LC-40 and CCAFS SLC-40

Total Payload Mass

- SpaceX boosters have carried a total of 48,213 kg for NASA

In [10]:
```
%%sql
select SUM(payload_mass__kg_) as total_payload_mass from SPACEXDATASET
where customer like '%NASA (CRS)%'
```

```
 * ibm_db_sa://jts26188:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databas
es.appdomain.cloud:32536/bludb
Done.
```

Out[10]:

total_payload_mass
48213

Average Payload Mass by F9 v1.1

- Average payload mass using the F9 v1.1 booster was 2,928 kg

In [11]:
```
%%sql
select AVG(payload_mass__kg_) as avg_payload_mass from SPACEXDATASET
where booster_version like '%F9 v1.1'
```

```
 * ibm_db_sa://jts26188:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databas
es.appdomain.cloud:32536/bludb
Done.
```

Out[11]:

avg_payload_mass
2928

First Successful Ground Landing Date

- SpaceX first successful ground landing was on 2015-12-22

In [12]:
```
%%sql
select MIN(DATE) as first_success_date from SPACEXDATASET
where mission_outcome = 'Success' AND landing__outcome like '%ground pad%'
```

```
 * ibm_db_sa://jts26188:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databas
es.appdomain.cloud:32536/bludb
Done.
```

Out[12]:

first_success_date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- F9 Boosters FT - B1020, 1022, 1026 and their variants successfully landed on drone ships when carrying payload mass between 4000 and 6000 kg.

In [14]:
```
%%sql
select B.booster_version from
(select * from SPACEXDATASET where mission_outcome = 'Success' AND landing__outcome like '%drone
WHERE B.payload_mass__kg_ between 4000 and 6000
```

```
 * ibm_db_sa://jts26188:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databas
es.appdomain.cloud:32536/bludb
Done.
```

Out[14]:

booster_version
F9 FT B1020
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Only 1 mission out of 100 was a failure

In [16]:
```
%%sql
select count(mission_outcome) as total_success from SPACEXDATASET
where mission_outcome like '%Success'
```

```
 * ibm_db_sa://jts26188:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databas
es.appdomain.cloud:32536/bludb
Done.
```

Out[16]:

total_success
99

In [17]:
```
%%sql
select count(mission_outcome) as total_failures from SPACEXDATASET
where mission_outcome like '%Failure%'
```

```
 * ibm_db_sa://jts26188:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databas
es.appdomain.cloud:32536/bludb
Done.
```

Out[17]:

total_failures
1

Boosters Carried Maximum Payload

In [18]:
```
%%sql
select booster_version from SPACEXDATASET
Where payload_mass__kg_ = (select MAX(payload_mass__kg_) as max_payload from SPACEXDATASET)
```

```
 * ibm_db_sa://jts26188:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

Out[18]:

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

- F9 Booster versions B5 1048 – 1060+ have all carried the maximum payload mass (kg).

2015 Launch Records

- 2 failed landings on the drone ship occurred in 2015, both from launch site CCAFS LC-40

In [23]:
```
%%sql
select DATE, landing__outcome, booster_version, launch_site from SPACEXDATASET
where Year(DATE) = 2015 and landing__outcome like '%Failure (drone ship)%'
```

```
 * ibm_db_sa://jts26188:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databas
es.appdomain.cloud:32536/bludb
Done.
```

Out[23]:

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

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

- Between June 2010 and March 2017 10 landings were not attempted, there were 7 confirmed failures and 8 confirmed successes.

In [34]:
```
%%sql
select B.landing__outcome, count(B.landing__outcome) as outcome_count from
(select landing__outcome FROM SPACEXDATASET Where DATE between '2010-06-04' And '2017-03-20') B
GROUP BY landing__outcome order by 2 desc
```

```
 * ibm_db_sa://jts26188:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databas
es.appdomain.cloud:32536/bludb
Done.
```

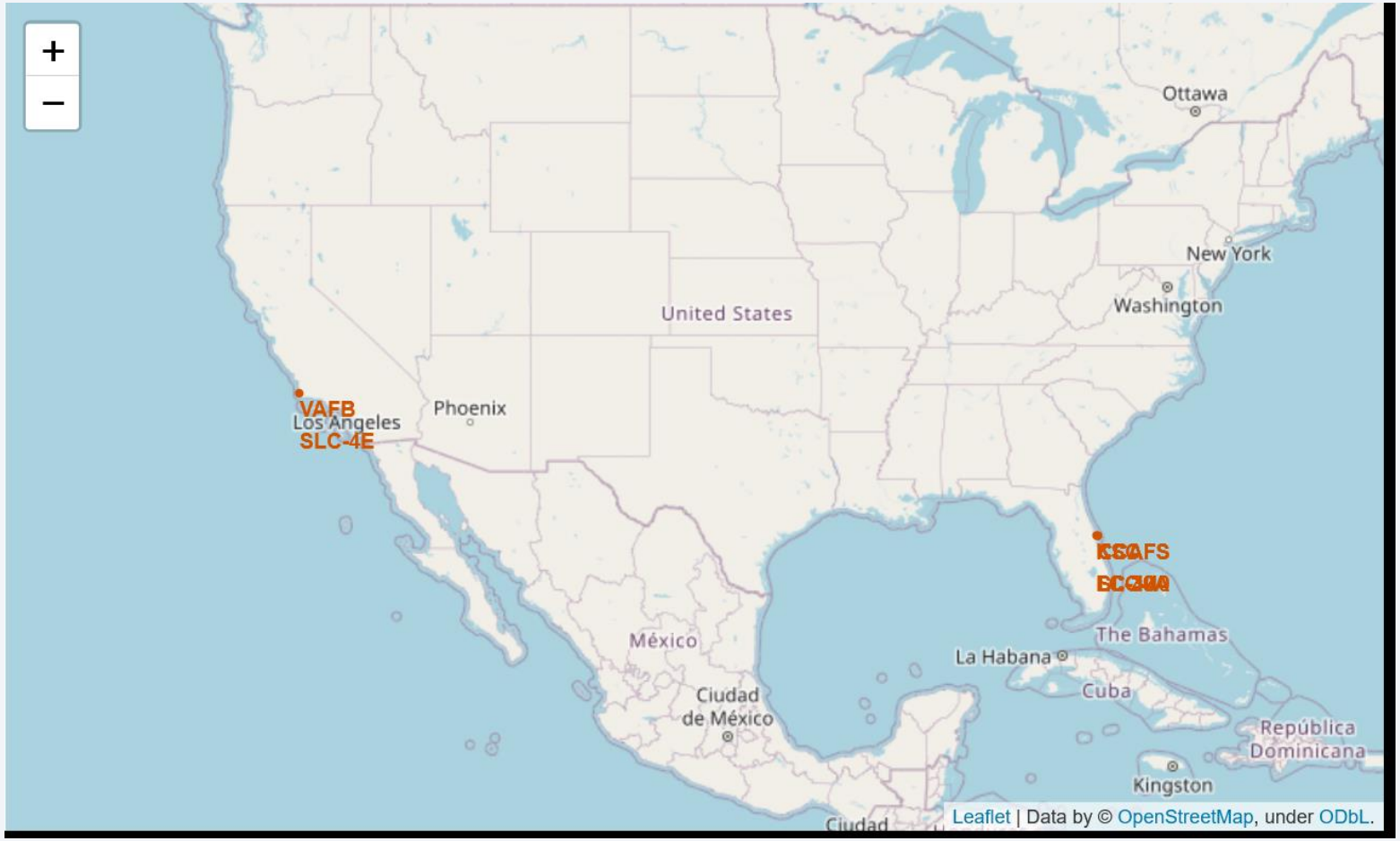
Out[34]:

landing__outcome	outcome_count
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

Space X Launch site locations

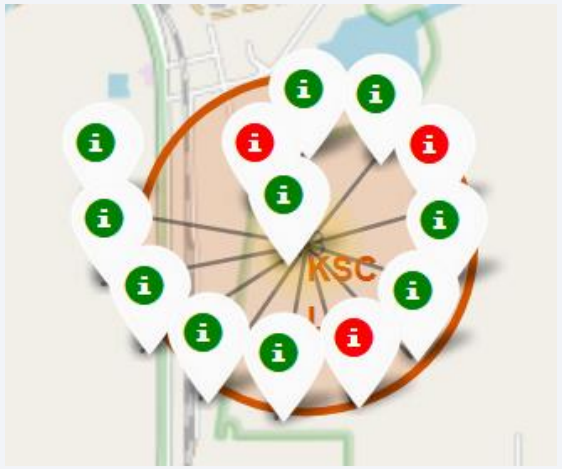
- All launch sites are coastal
- 1 in California
- 3 in Florida
- 2 within same complex in Florida (Cape Canaveral)



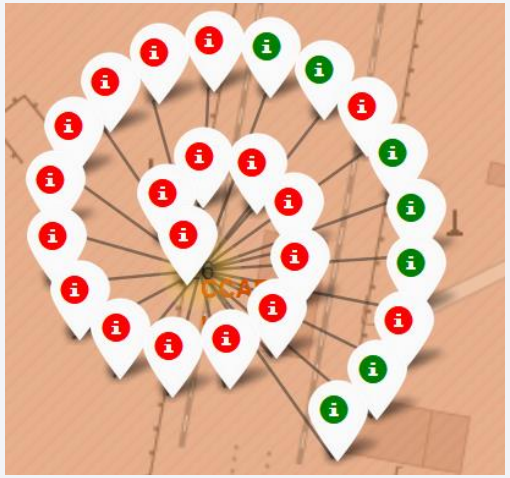
Launch Outcomes at each launch site

VAFB SLC-4E

KSC LC-39A



CCAFS LC-40

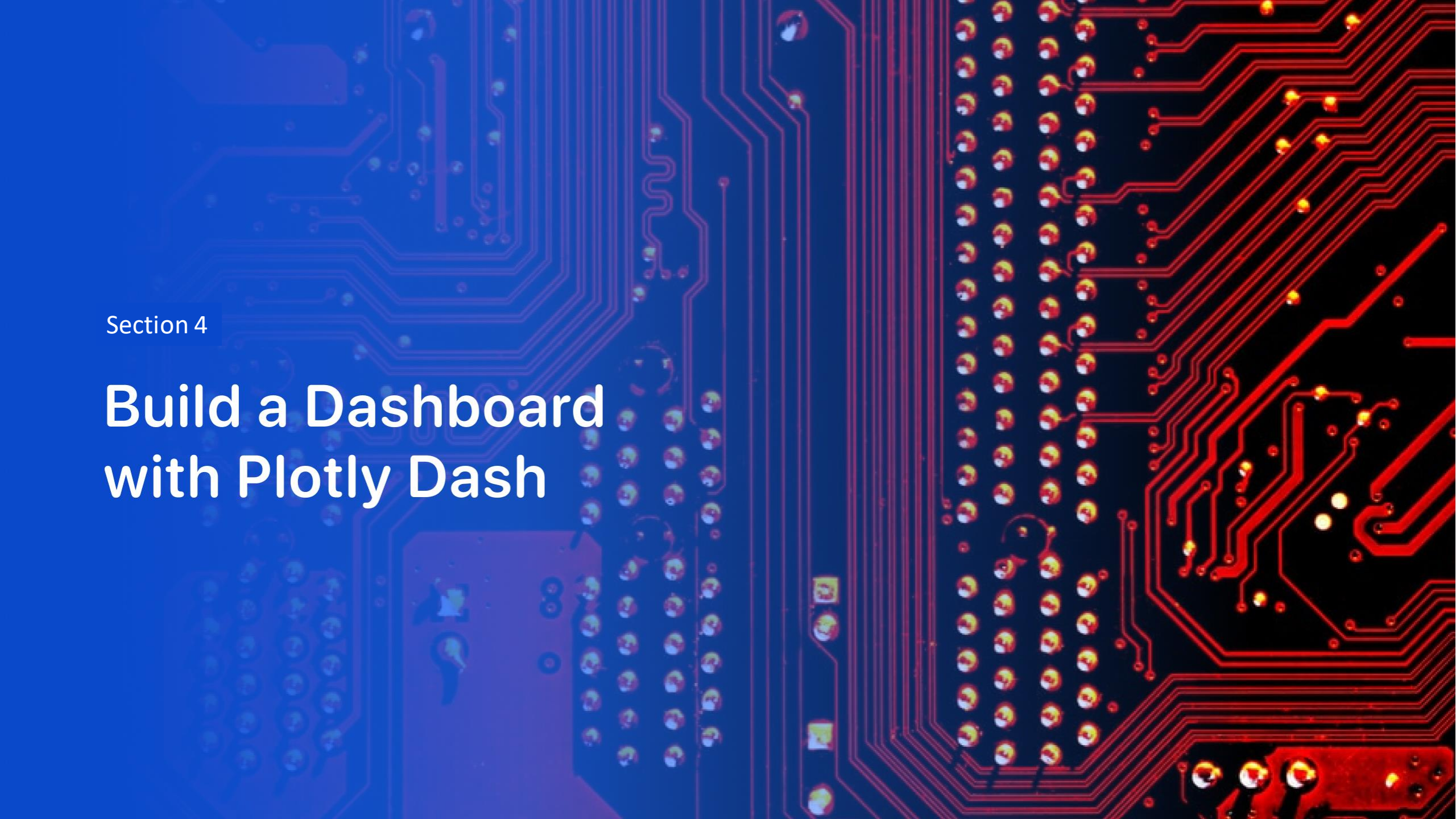


CCAFS SLC-40

- KSC (Kennedy) highest % of success launches
- CCAFS LC-40 (Cape Canaveral) most launches, but poor success rate.

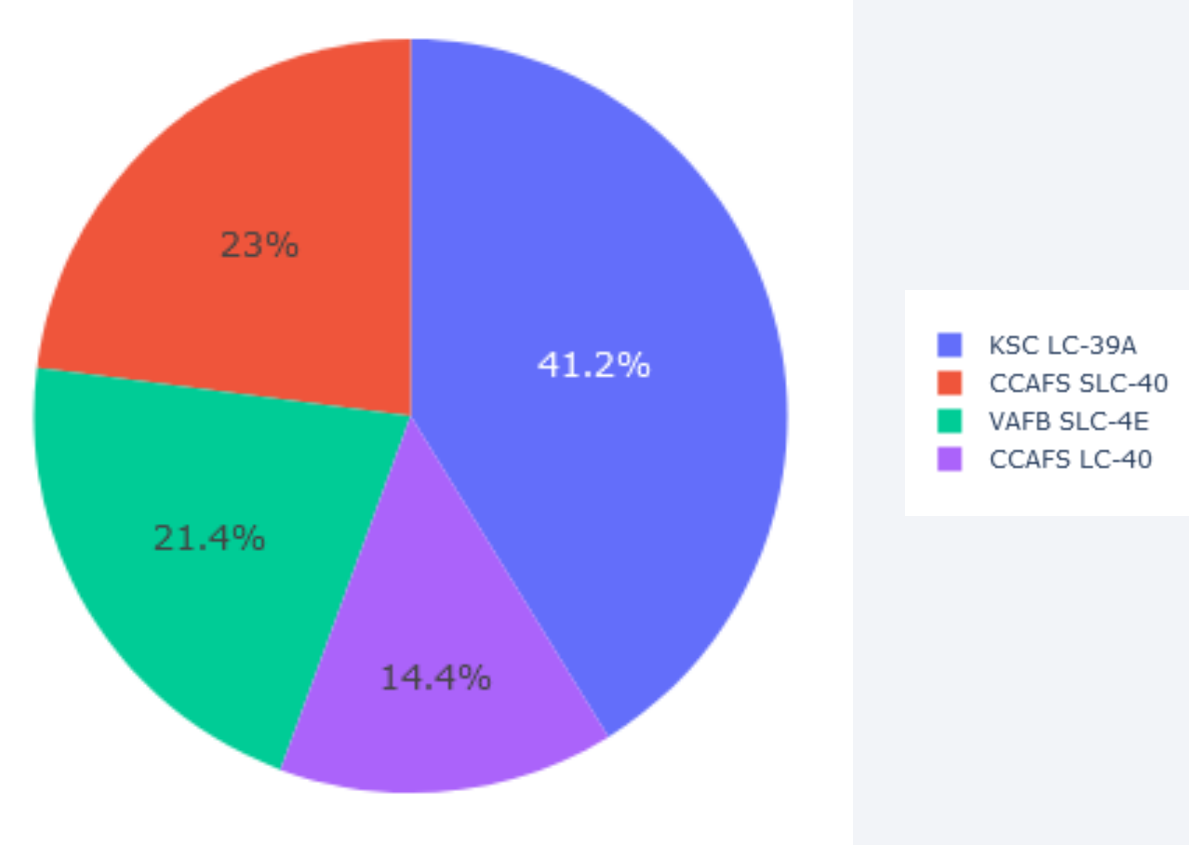
Launch site proximity to infrastructure

- Launch site is in close proximity to the coast, rail, and highways.
- Launch site is located far from city centers.


Section 4
Build a Dashboard with Plotly Dash

Launch success by launch site

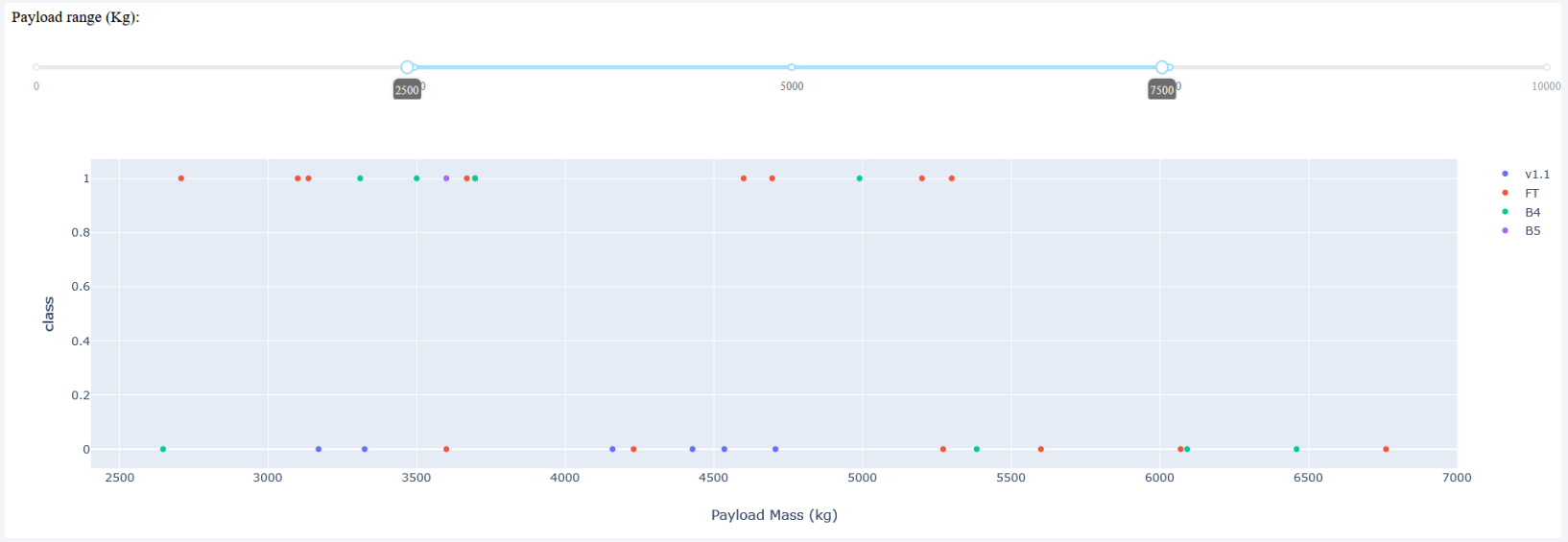
- KSC LC-39A (Florida - Kennedy) had highest % of successful launches of the 4 launch sites.

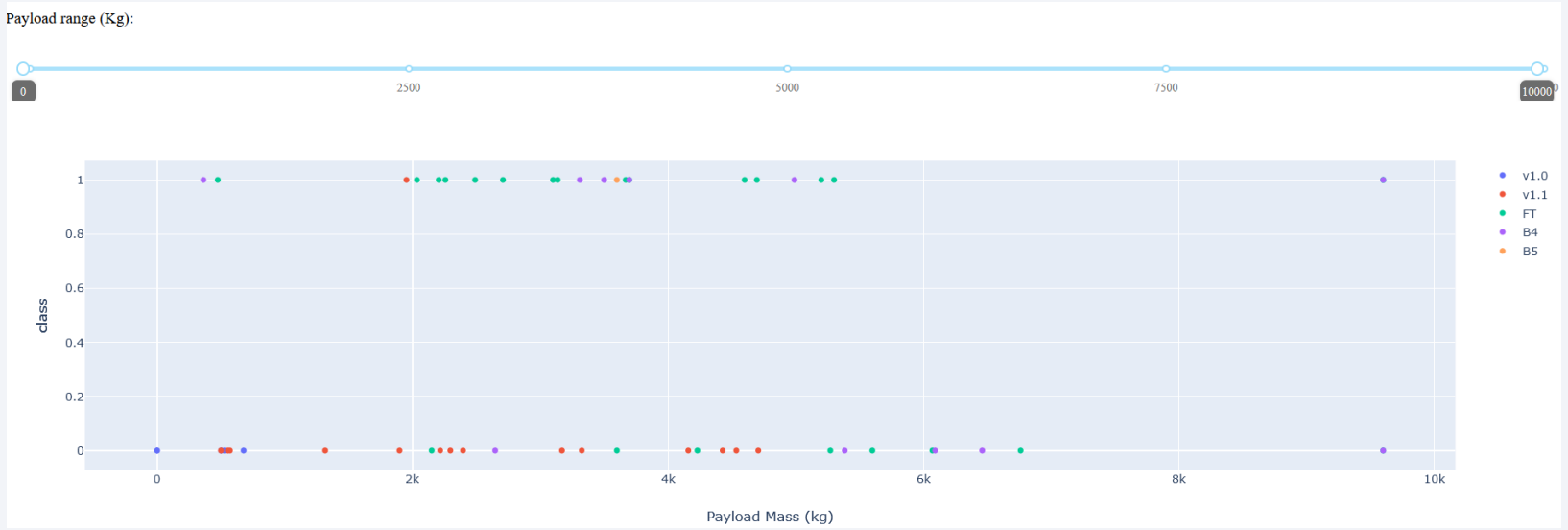


Launch Success rate of Kennedy

- KSC LC 39-A (Kennedy space center) successful launch rate of 76.9%

Launch Outcome by Payload Mass



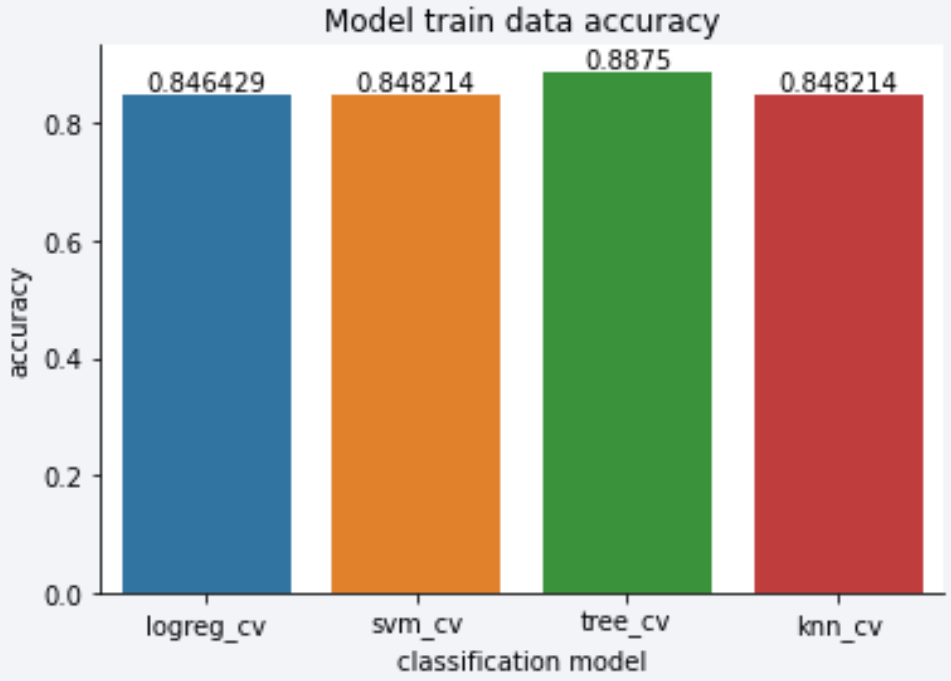


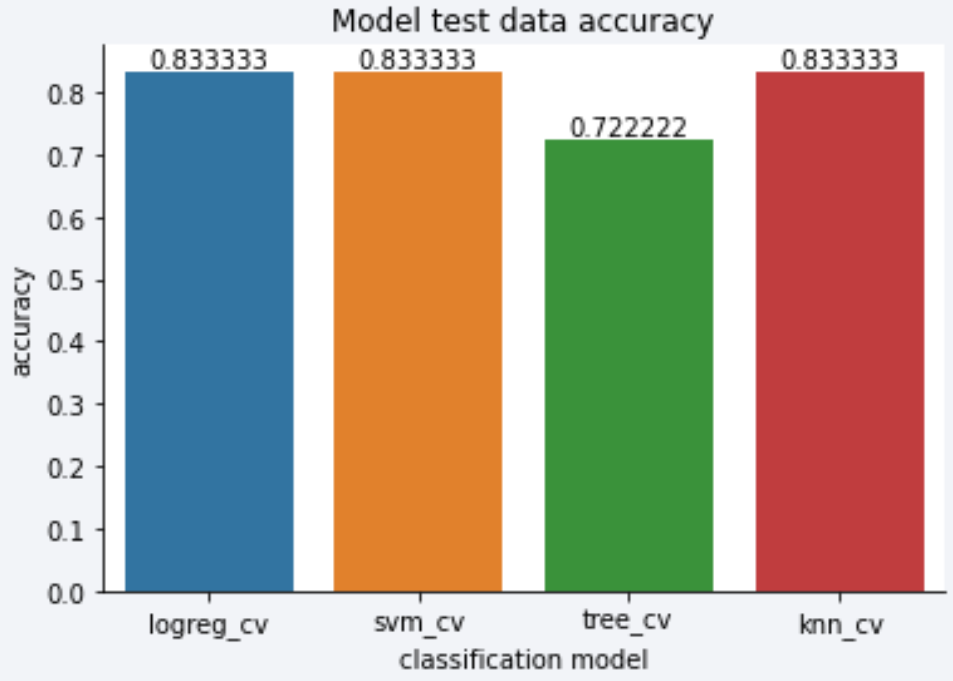
- Booster B4 only booster when payload mass > 7000 kg
- Most successful outcomes fall between payload mass of 2500 and 5000 kg

Section 5

Predictive Analysis (Classification)

Classification Accuracy





- Decision tree performed best on the training data, but worst using the test data.
- Using the limited amount of test data, the logistic regression, SVM, and KNN models all had similar performance.

Confusion Matrix

- The Logisitic regression, SVM, and KNN models all had the same performance on test data.
- Model(s) were very accurate predicting landings, but had issues with predicting when stage would not land.

Conclusions

- Launching to orbit types ES-L1, GEO, HEO, SSO, and VLEO (most recent launches were VLEO) resulted in high landing success rate
- SpaceX improved landing outcomes over time, and launches began to be focused on using KSC (Kennedy) and CCSFS (Cape Canaveral) launch sites in coastal Florida.
- Most successful landing outcomes launched with a payload between 2500 and 5000 kg
- While logistic regression, SVM, and KNN models all were able to predict successful landings very well, they did not perform well predicting when a landing would fail.
- More test data should be run to further refine modeling, but with current data, to give highest chance of initial mission & landing outcome: suggest launching payloads between 2500 – 5000 kg to an SSO orbit from Kennedy (KSC).

Thank you!