IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Alevtina Zabegaeva
26.02.2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection
  - Data wrangling
  - EDA with data visualization
  - EDA with SQL
  - Interactive map with Folium
  - Dashboard with Plotly dash
  - Predictive Analysis (Classification)
- Summary of all results
  - Exploratory data analysis results
  - Interactive analytics demonstration
  - Predictive analysis results

Introduction

- Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage.
- Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against space X for a rocket launch.

Section 1

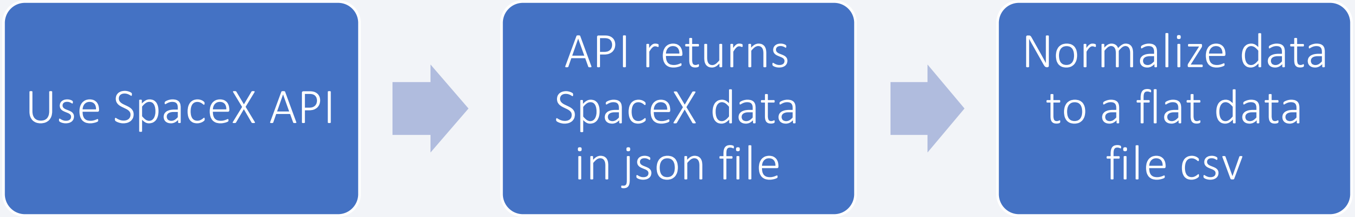
Methodology

Methodology

- Data collection methodology:
  - SpaceX REST API
  - Web scraping from Wiki
- Perform data wrangling
  - Removing irrelevant columns
  - One hot encoding data field
- 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

- Following steps are made for collecting the data:
  - Request to the SpaceX API
  - Clean the requested data



  - Extract a Falcon 9 launch records HTML table from Wikipedia
  - Parse the table and convert it into a Pandas data frame

Data Collection – SpaceX API

- Request and parse rocket launch data from SpaceX API using the GET request
- Filter the dataframe to only include Falcon 9 launches
- Dealing with Missing Values
- Export dataset to a csv file
- SpaceX API calls notebook: https://github.com/alevtina-zabegaeva/Capstone-project-coursera-SpaceX/blob/main/Data%20Collection%20API.ipynb

Data Collection - Scraping

- Request the Falcon9 Launch Wiki page from its URL
- Extract all column/variable names from the HTML table header
- Create a data frame by parsing the launch HTML tables
- Export dataset to a csv file
- Web scraping notebook: https://github.com/alevtina-zabegaeva/Capstone-project-coursera-SpaceX/blob/main/Web_scraping.ipynb

Data Wrangling

- Exploratory Data Analysis was performed:
  - the number of launches on each site are calculated
  - the number and occurrence of each orbit are calculated
  - the number and occurrence of mission outcome per orbit type are calculated
- Training labels were determined:
  - Landing outcome are converted to classes, either 0 or 1
  - 0 is bad outcome, i.e. the booster did not land
  - 1 is good outcome, i.e. the booster did land
- Data wrangling notebook: https://github.com/alevtina-zabegaeva/Capstone-project-coursera-SpaceX/blob/main/Data%20Wrangling.ipynb

EDA with Data Visualization

- Following charts were plotted:
  - scatter point chart to visualize the relationship between Flight Number and Launch Site
  - scatter point chart to visualize the relationship between Payload and Launch Site
  - bar chart to visualize the relationship between success rate of each orbit type
  - scatter point chart to visualize the relationship between Flight Number and Orbit type
  - scatter point chart to visualize the relationship between Payload and Orbit type
  - line chart to visualize the launch success yearly trend
- EDA with data visualization notebook: https://github.com/alevtina-zabegaeva/Capstone-project-coursera-SpaceX/blob/main/EDA_with_Visualization.ipynb

EDA with SQL

- Following SQL queries are performed:
  - select distinct LAUNCH_SITE from SPACEXTBL;
  - select * from SPACEXTBL where LAUNCH_SITE like 'CCA%' limit 5;
  - select sum(payload_mass__kg_) from SPACEXTBL where customer = 'NASA (CRS)';
  - select avg(payload_mass__kg_) from SPACEXTBL where booster_version like 'F9 v1.1%';
  - select min(DATE) from SPACEXTBL where landing__outcome = 'Success (ground pad)';
  - select booster_version from SPACEXTBL where landing__outcome = 'Success (drone ship)' and (payload_mass__kg_ between 4000 and 6000);
  - select mission_outcome, count(mission_outcome) from SPACEXTBL group by mission_outcome;
  - select booster_version from SPACEXTBL where payload_mass__kg_=(select max(payload_mass__kg_) from SPACEXTBL);
  - select substr(DATE, 6, 2) as month, landing__outcome, booster_version, launch_site from SPACEXTBL where landing__outcome='Failure (drone ship)' and substr(DATE,1,4)='2015';
  - select landing__outcome, count(landing__outcome) count from SPACEXTBL where (DATE between '2010-06-04' and '2017-03-20') and landing__outcome like 'Succes%' group by landing__outcome;
- EDA with SQL notebook: https://github.com/alevtina-zabegaeva/Capstone-project-coursera-SpaceX/blob/main/EDA_with_SQL.ipynb

Build an Interactive Map with Folium

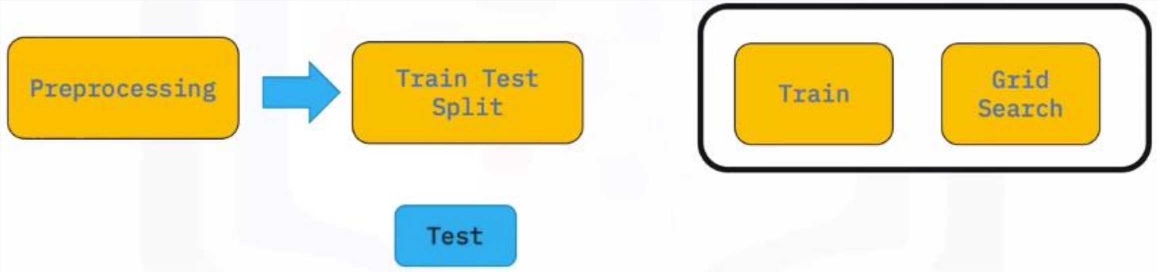
- With Folium more interactive visual analytics are performed, using following map objects:
  - circles with text labels and markers – to mark all launch sites on a map
  - clustered markers with different colors – to mark successful und unsuccessful launches
  - lines with markers - to show distances to proximities,
- Interactive map with Folium: https://github.com/alevtina-zabegaeva/Capstone-project-coursera-SpaceX/blob/main/IVA%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

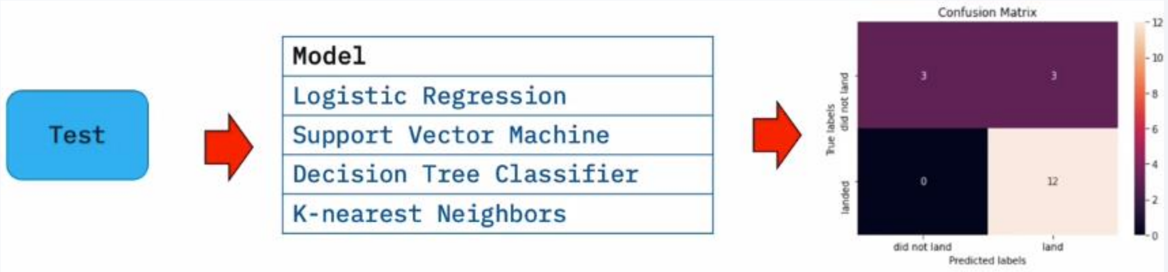
- The dashboard application contains following components:
  - Input component - dropdown list with four different launch sites to interact with a pie chart
  - Input component - range slider with payloads to interact with a scatter point chart
  - Pie chart which shows the total success launches for each site and percentage of successful launches
  - Scatter point chart which shows payload vs. launch outcome with color-label the Booster-version on each scatter point
- Those plots and interactions were added to find insights visually, i.e. to analyze SpaceX launch data, and answer the following questions:
  - Which site has the largest successful launches?
  - Which site has the highest launch success rate?
  - Which payload range(s) has the highest launch success rate?
  - Which payload range(s) has the lowest launch success rate?
  - Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest launch success rate?
- Link to Plotly Dash: https://github.com/alevtina-zabegaeva/Capstone-project-coursera-SpaceX/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Perform exploratory Data Analysis and determine Training Labels
  - Create a column for the class
  - Standardize the data
  - Split into training data and test data



- Find best Hyperparameter for SVM, Classification Trees, KNN and Logistic Regression
  - Find the method performs best using test data



- Link to predictive analysis: https://github.com/alevtina-zabegaeva/Capstone-project-coursera-SpaceX/blob/main/ML%20Prediction%20lab.ipynb

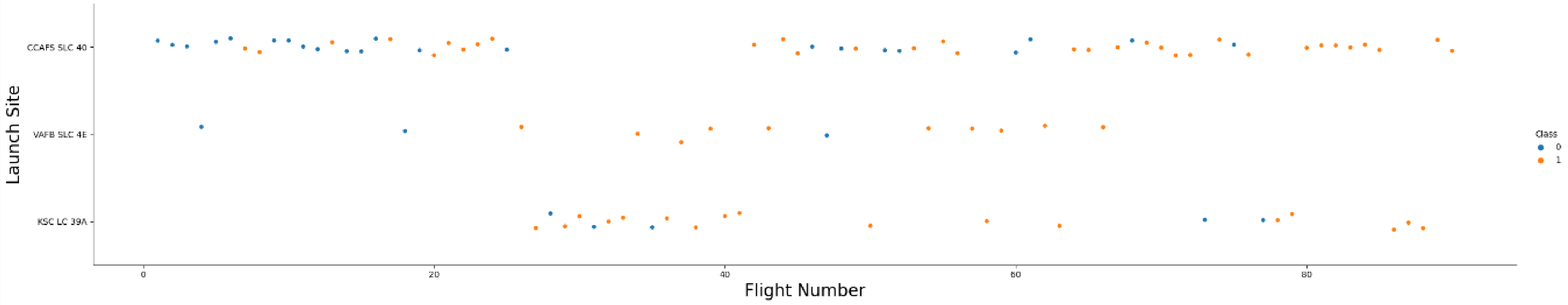
Results

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

Section 2

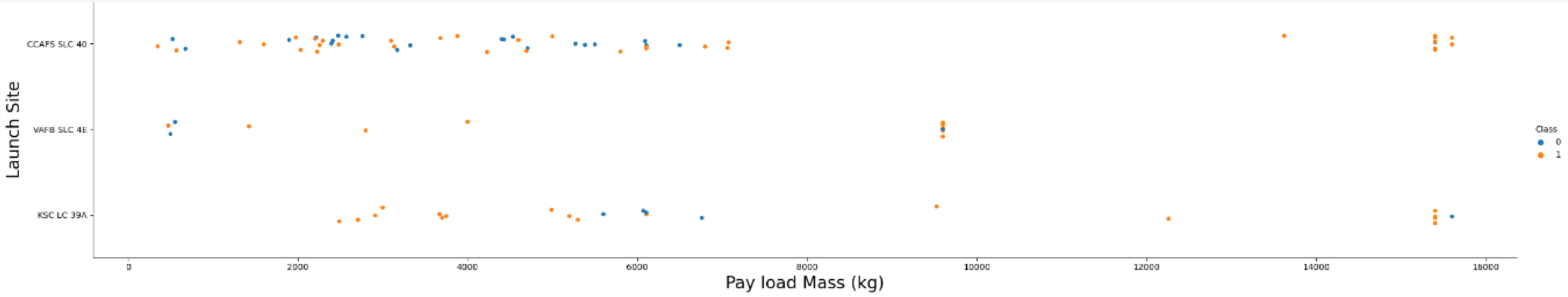
Insights drawn from EDA

Flight Number vs. Launch Site



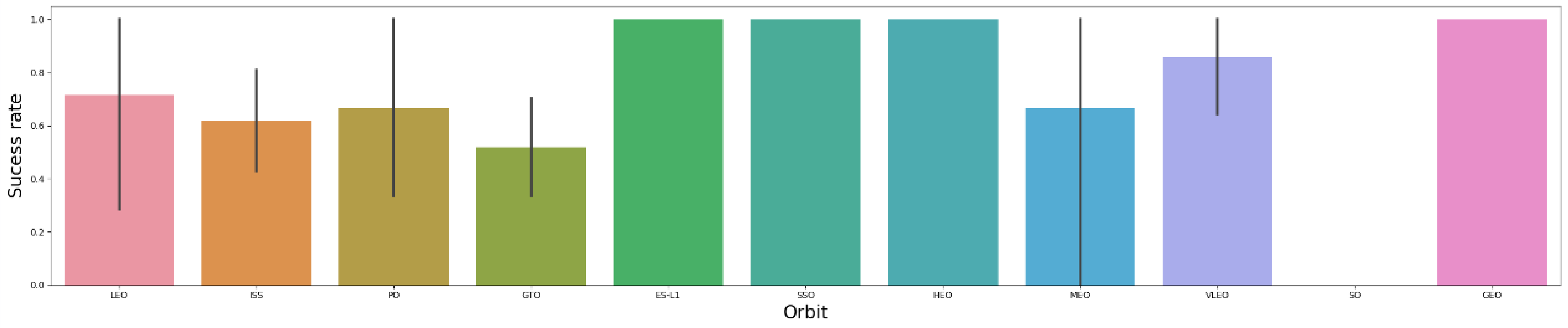
- At each launch site, subsequent launches were more often successful than previous ones
- There were especially many failures before the 20th launch, and almost all of them were at site CCAFS SLC-40

Payload vs. Launch Site



- At VAFB-SLC 4E launch site no rockets launched for heavy payload mass (greater than 10000 kg)
- At KSC LC-39A launch site no rockets launched for light payload mass (less than 2000 kg)

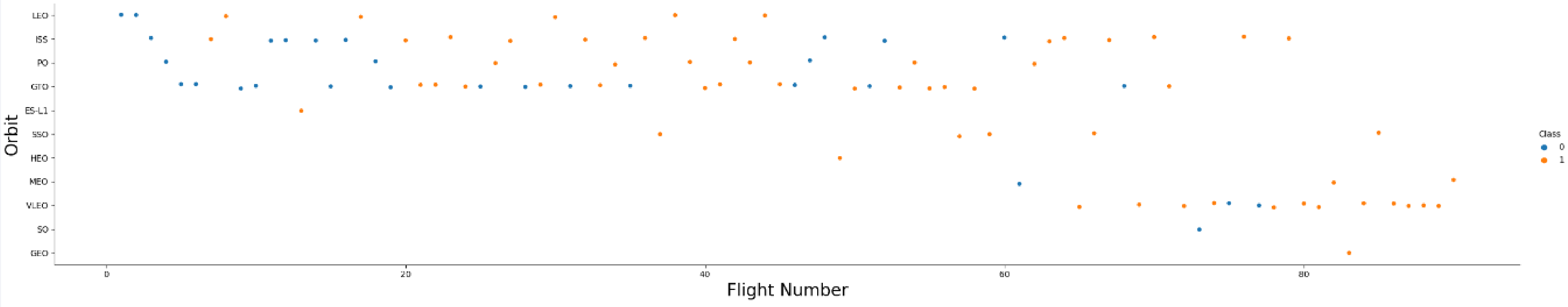
Success Rate vs. Orbit Type



Following orbits have high success rate:

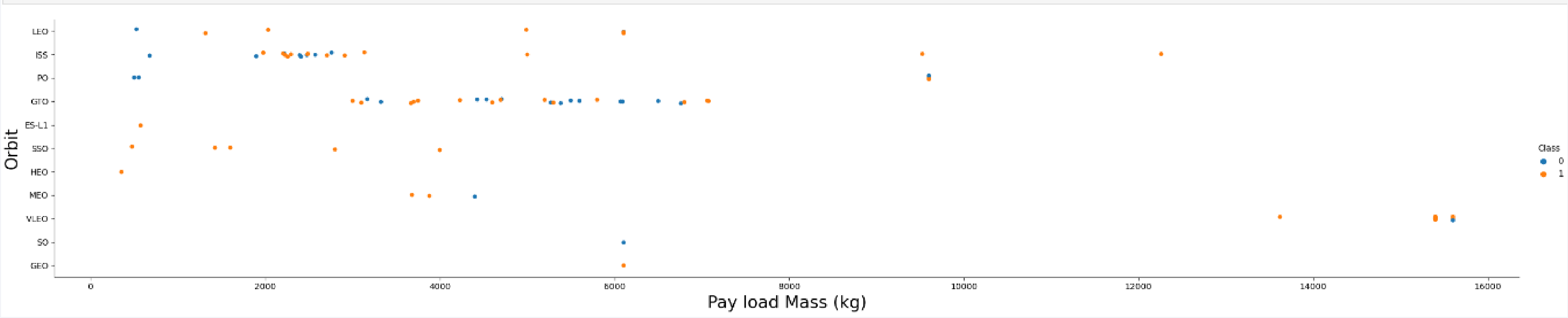
- ES-L1: Lagrange point L1 between the Sun and the Earth
- SSO: Sun-synchronous orbit
- HEO: Highly elliptical orbit
- VLEO: Very low Earth orbit
- GEO: Geosynchronous orbit 35,786 km above Earth's equator

Flight Number vs. Orbit Type



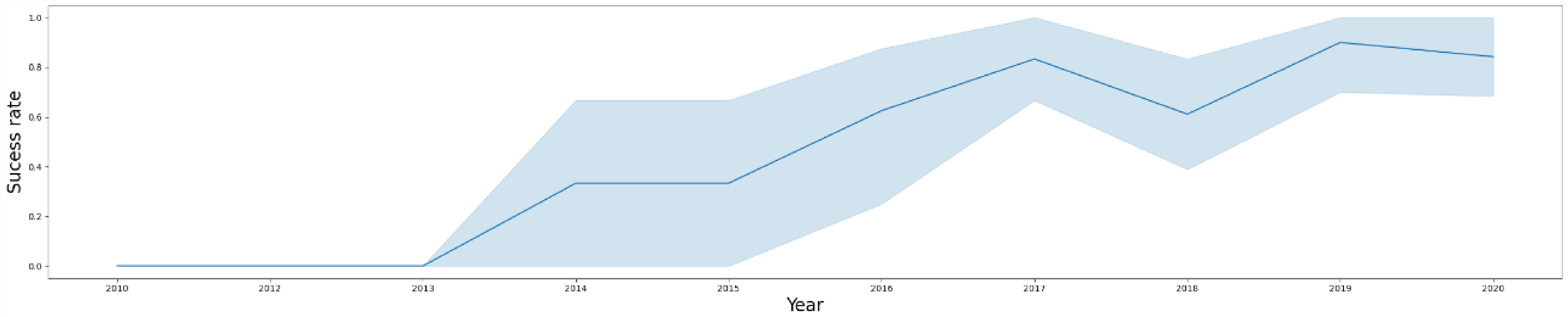
- In the LEO orbit the Success appears related to the number of flights
- On the other hand, there seems to be no relationship between flight number and success in GTO and ISS orbits

Payload vs. Orbit Type



- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS orbits
- However for GTO we cannot distinguish this well as positive landing rate and negative landing (unsuccessful mission) are both there here

Launch Success Yearly Trend



Success rate since 2013 kept increasing till 2020

All Launch Site Names

```
%sql select DISTINCT LAUNCH_SITE from SPACEXTBL;
```

```
 * ibm_db_sa://ckm19829:***@55fbc997-9266-4331-afd3-888b05e734c0.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31929/bludb
Done.
```

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

There are four unique launch sites:

- 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://ckm19829:***@55fbc997-9266-4331-afd3-888b05e734c0.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31929/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

Tere are 5 records where launch sites begin with `CCA`, they all have CCAFS LC-40 name and booster version F9 v1.0

Total Payload Mass

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

```
 * ibm_db_sa://ckm19829:***@55fbc997-9266-4331-afd3-888b05e734c0.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31929/bludb
Done.
```

1
45596

Total payload carried by boosters from NASA equal to 45596 kg

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://ckm19829:***@55fbc997-9266-4331-afd3-888b05e734c0.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31929/bludb
Done.
```

1
2534

Average payload mass carried by booster version F9 v1.1 equal to 2534 kg

First Successful Ground Landing Date

```
%sql select min(DATE) from SPACEXTBL where landing__outcome = 'Success (ground pad)';
```

```
 * ibm_db_sa://ckm19829:***@55fbc997-9266-4331-afd3-888b05e734c0.bs2io90l08kqblod8lcg.databases.appdomain.cloud:31929/bludb
Done.
```

1
2015-12-22

First successful landing outcome on ground pad took place on 2015.12.22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
 * ibm_db_sa://ckm19829:***@55fbc997-9266-4331-afd3-888b05e734c0.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31929/bludb
Done.
```

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

There are four boosters F9 FT which have successfully landed on drone ship and had payload mass greater than 4000 kg but less than 6000 kg

Total Number of Successful and Failure Mission Outcomes

```
%sql select mission_outcome, count(mission_outcome) from SPACEXTBL group by mission_outcome;
```

```
 * ibm_db_sa://ckm19829:***@55fbc997-9266-4331-afd3-888b05e734c0.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31929/bludb
Done.
```

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Total number of successful and failure mission outcomes equal to 100 and 1 corresponding

Boosters Carried Maximum Payload

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

```
 * ibm_db_sa://ckm19829:***@55fbc997-9266-4331-afd3-888b05e734c0.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31929/bludb
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

The booster F9 B5 have carried the maximum payload mass

2015 Launch Records

```
%sql select substr(DATE, 6, 2) as month, landing__outcome, booster_version, launch_site from SPACEXTBL where landing__outcome='Failure (drone ship)' and substr(DATE,1,4)='2015';
```

```
 * ibm_db_sa://ckm19829:***@55fbc997-9266-4331-afd3-888b05e734c0.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31929/bludb
Done.
```

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

In 2015 there were only two failed landing_outcomes in drone ship, their booster version is F9 v1.1 and launch site name is CCAFS LC-40

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

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

```
 * ibm_db_sa://ckm19829:***@55fbc997-9266-4331-afd3-888b05e734c0.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31929/bludb
Done.
```

landing__outcome	COUNT
Success (drone ship)	5
Success (ground pad)	3

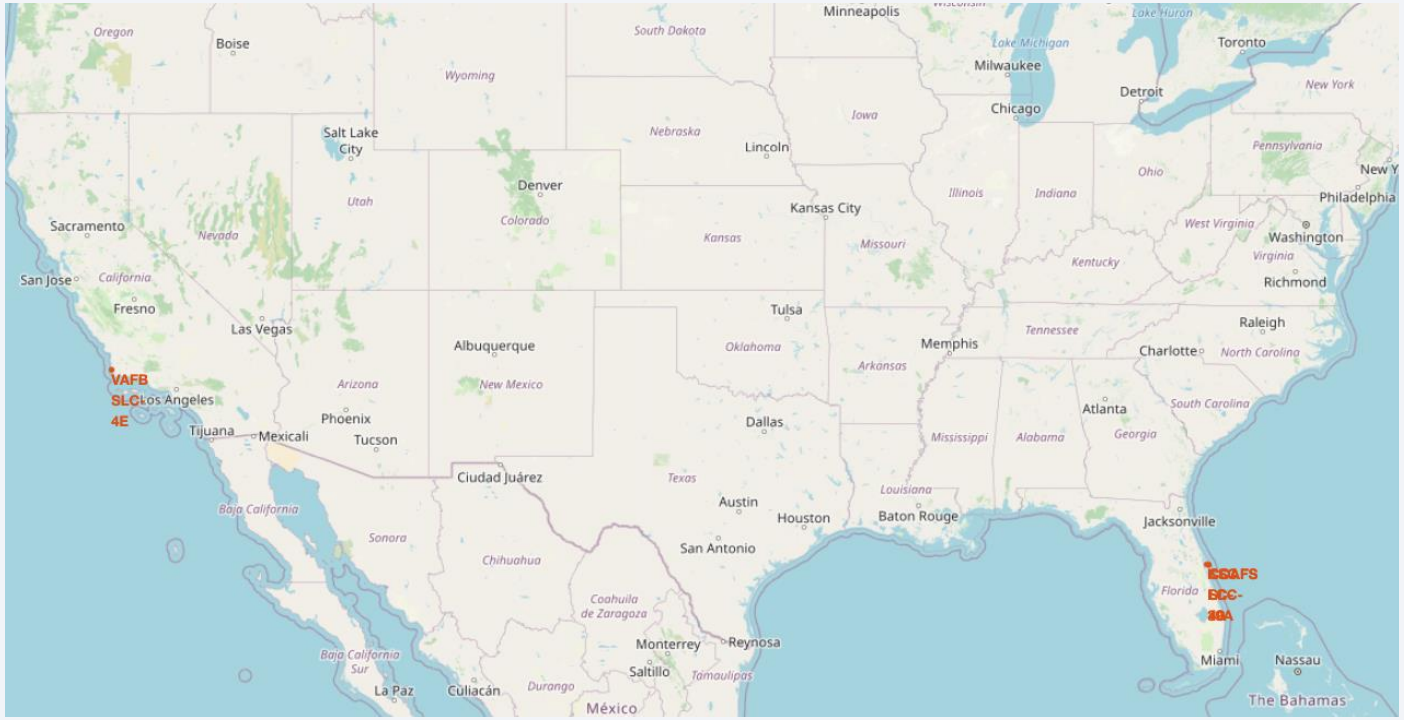
Count of landing outcomes between the date 2010-06-04 and 2017-03-20:

- Failure (drone ship): 5
- Success (ground pad): 3

Section 3

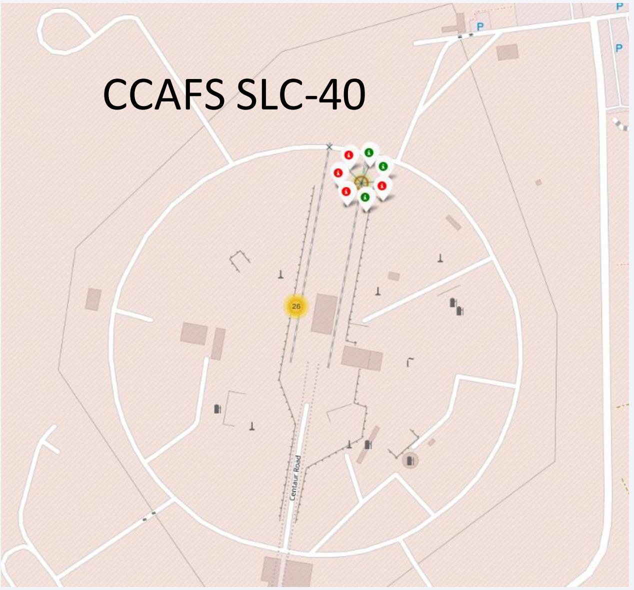
Launch Sites Proximities Analysis

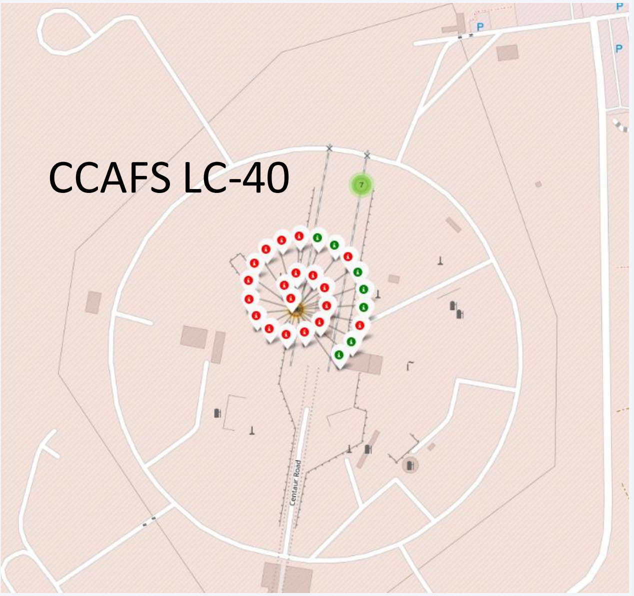
All Launch Sites on a Global Map



- Three of the four sites are very close to each other
- All launch sites are located not far from the Equator line (max latitude is 34.6°)
- All launch sites in very close proximity to the coast

Success/failed Color-labeled Markers for each Site on a Map



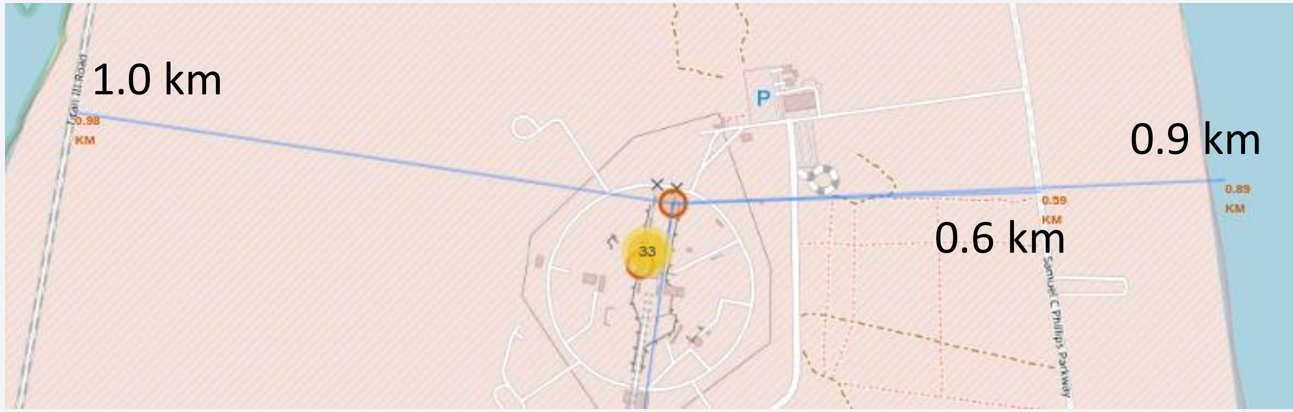


Launch sites KSC LC-39A and CCAFS LC-40 have relatively high success rates

Distances Between a Launch Site CCAFS SLC-40 to its Proximities

Launch Site CCAFS SLC-40 has a following distances to its proximities:

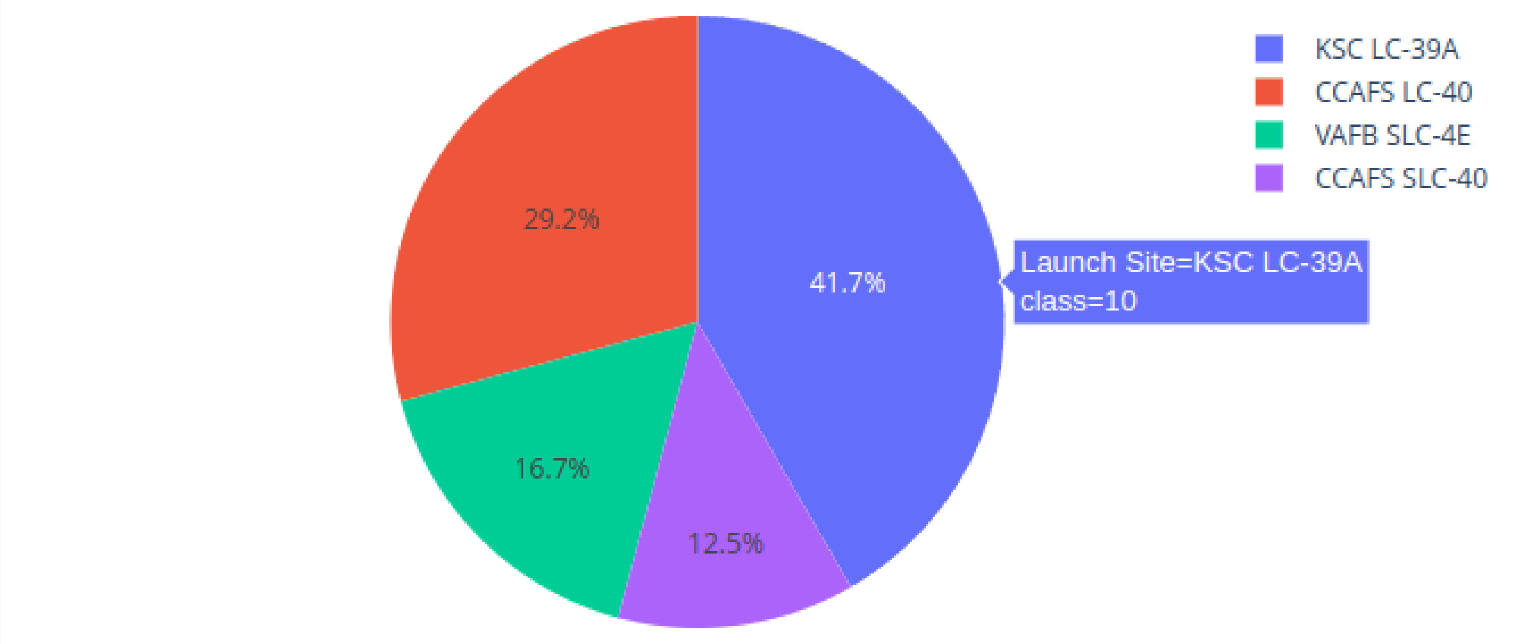
- 1.0 km and 0.6 km to railway and highway – a short distance for efficient logistics
- 0.9 km to coastline – a short distance to organize a floating platform to catch a first stage and minimize damage in case of failure
- 17.4 km to the nearest city – a long distance to minimize damage in case of failure



Section 4

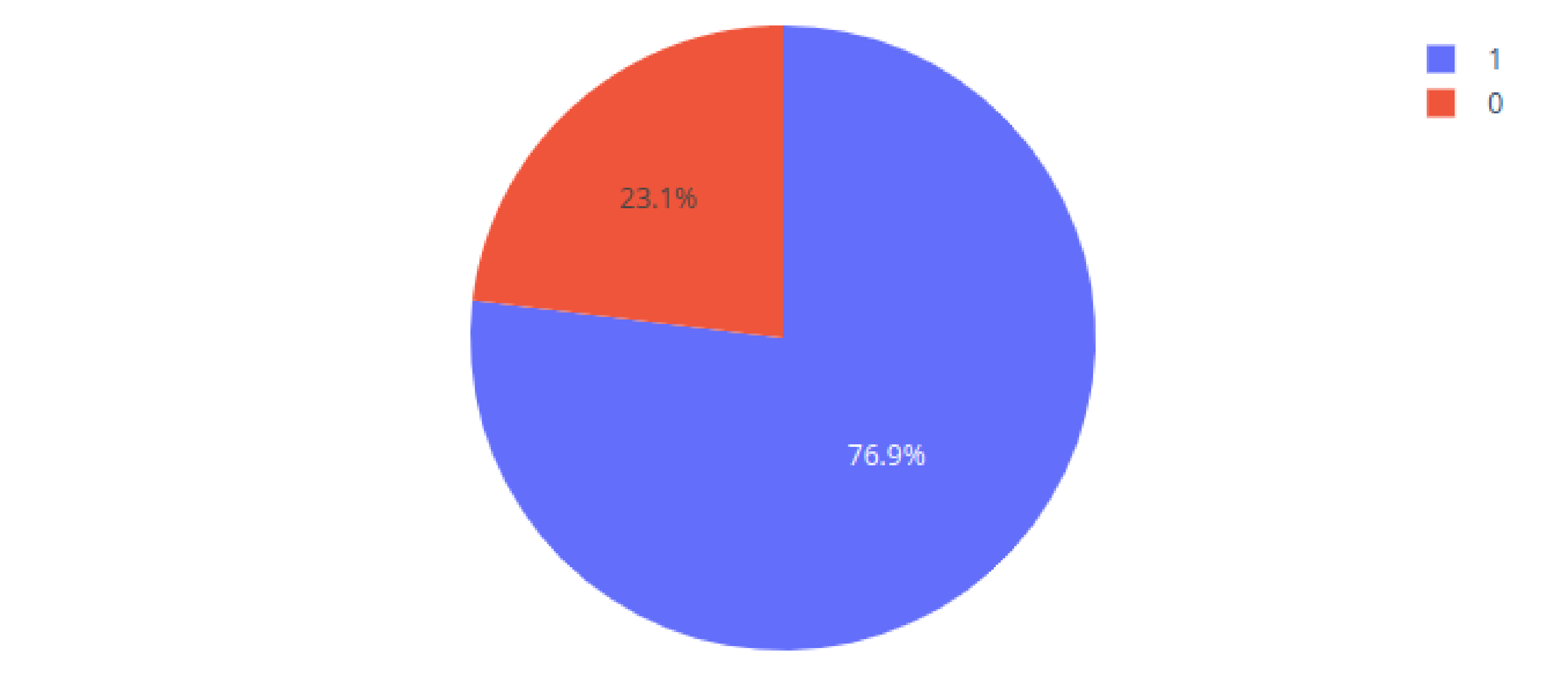
Build a Dashboard with Plotly Dash

Total Success Launches by Site



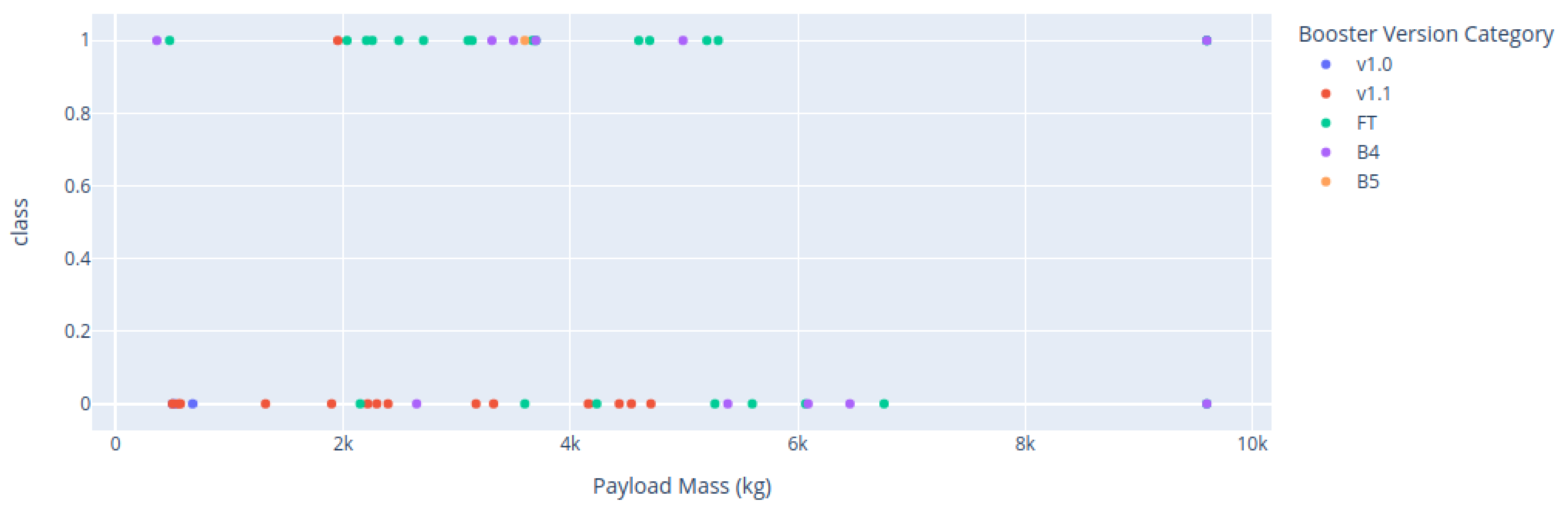
- the largest number of successful launches, namely 41,7%, occurred in site KSC LC-39A
- followed by site CCAFS LC-40 with 29,2%
- in sites VAFB SLC-4E and CCAFS SLC-40 there were significantly fewer successful launches

Success Launches for Site KSC LC-39A



- Percentage of successful launches in site with highest launch success ratio KSC LC-39A is 76,9%

Correlation between Payload and Success for all Sites



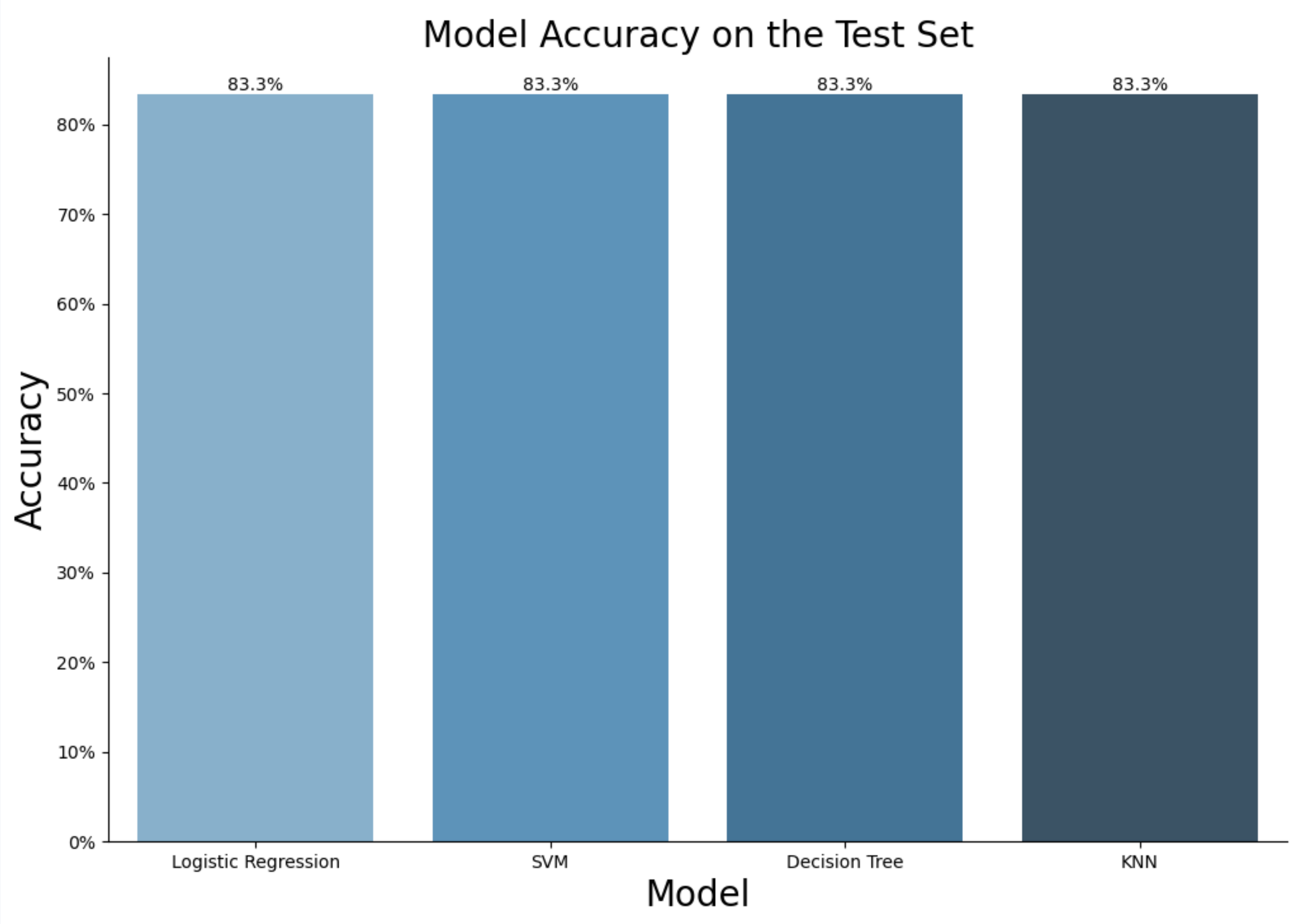
- Payload mass range between 1900 kg and 5300 kg has the largest success rate
- Launches with booster v1.0 and v1.1 were almost always unsuccessful
- On the contrary, launches with booster version FT, B4 and B5 have a much higher success rate, especially with a payload mass of up to 5300 kg

Section 5

Predictive Analysis (Classification)

Classification Accuracy

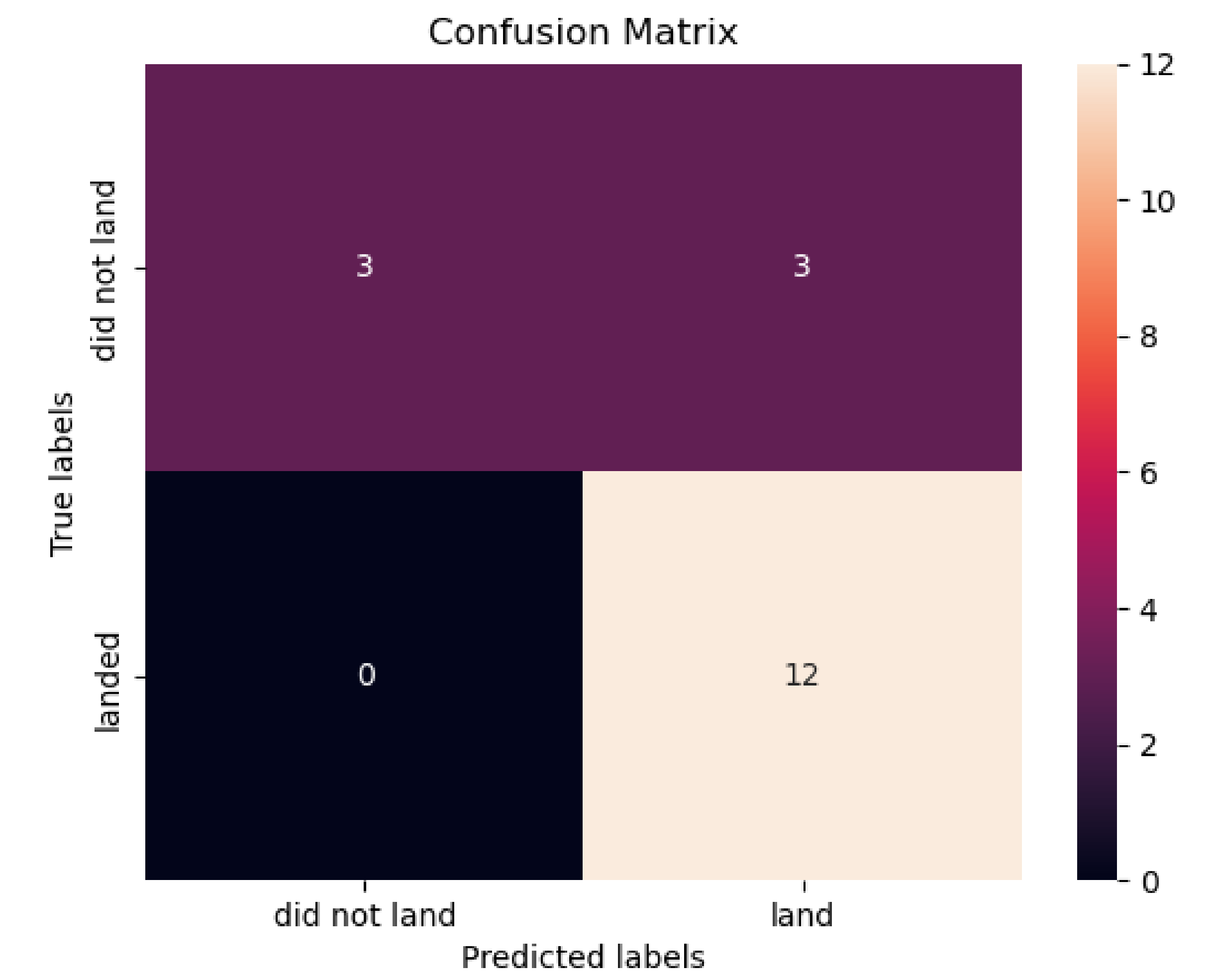
The results are practically the same. This is because the dataset is small and having lesser values.



Confusion Matrix

Confusion matrix is the same for each of four performing models.

Examining the confusion matrix, we see that classification models can distinguish between the different classes and the major problem is false positives.



Conclusions

- Payload mass range between 1900 kg and 5300 kg has the largest success rate
- Launch success rate started to increase in 2013 till 2020
- Orbits ES-L1, GEO, HEO, SSO, VLEO have the most success rate
- KSC LC-39A has the largest number of successful launches of any sites

Thank you!