IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Summary of methodologies

In this study, we used information from the API provided by Space X and information that could be collected from Wikipedia.

Based on these information, we utilized several machine learning models to build a predictive model of Falcon 9 first stage successful landing and tested their effectiveness.

- Summary of all results

4-predictive models were built using four machine learning models: Logistic regression, SVM, Decision tree, and KNN.

The validation results showed that the decision tree model was able to predict the successful landing of Falcon 9 first stage with more than 90% prediction accuracy.

Introduction

- Project background and context

  SpaceX 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 SpaceX 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 SpaceX for a rocket launch.

- Problems you want to find answers

  In this study, we evaluated if we can predict if the Falcon 9 first stage will land successfully.

Section 1

Methodology

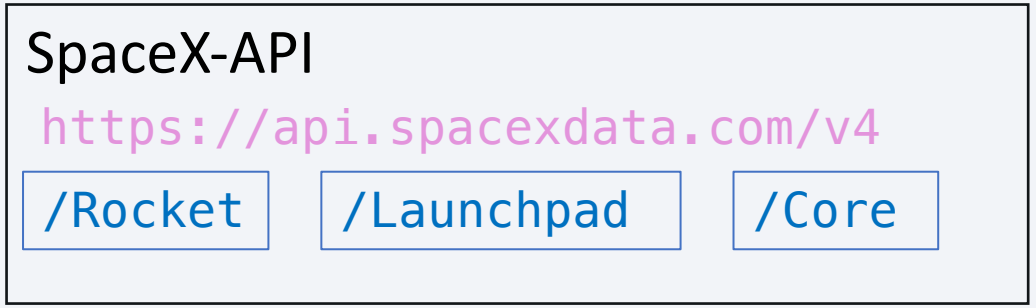
Methodology

Executive Summary

- Data collection methodology:
  - We used information from the API provided by Space X and information that could be collected from Wikipedia.
- Perform data wrangling
  - We tried some Exploratory Data Analysis (EDA) to find some patterns in the data, such as “Flinght number”, “BoosterVersion”, “PayloadMass”, etc., and determine what would be the label for training supervised models,
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - 4 models were built using four machine learning models: Logistic regression, SVM, Decision tree, and KNN.
  - Grid search was performed for each model to determine the optimal parameters.
  - 20 percent of all data was used as test data for accuracy verification.

Data Collection – SpaceX API

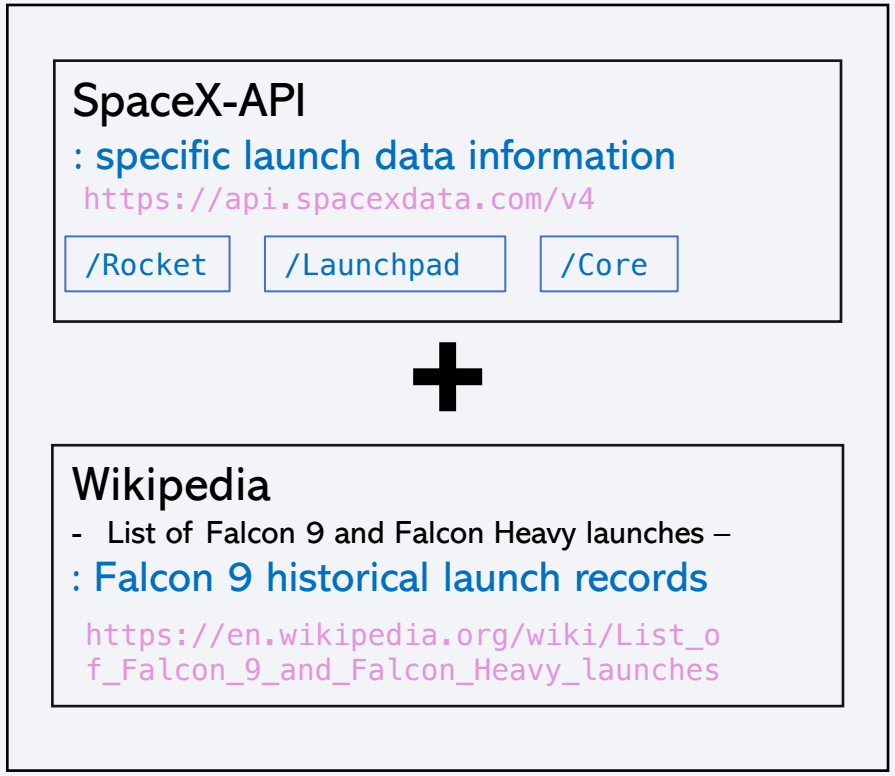
- SpaceX REST calls (https://api.spacexdata.com/v4) was used to collect specific launch data information. Several data was collected via "/Rocket", "/Launchpad", and "/Core"
- "/Rocket" was used to collect the information about the booster.
- "/Launchpad" was used to collect the information about the launch sites.
- "/Core" was used to collect other information, such as the outcome of the landing, the type of the landing, number of flights with that core, whether gridfins were used, wheter the core is reused, wheter legs were used, the landing pad used, etc.



- GitHub URL of the completed SpaceX API calls is:
https://github.com/OzasaHiro/Coursera_IBM_Datascience/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

- Web scraping was conducted to collect Falcon 9 historical launch records from a Wikipedia page titled `List of Falcon 9 and Falcon Heavy launches`



- GitHub URL of the completed web scraping notebook is: https://github.com/OzasaHiro/Coursera_IBM_Datascience/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

- We tried some Exploratory Data Analysis (EDA) to find some patterns in the data, such as “Flinght number”, “BoosterVersion”, “PayloadMass”, etc., and determine what would be the label for training supervised models.
- The number of launches on each site was caucurated.
- The number and occurrence of each orbit was caucurated.
- The number and occurence of mission outcome of the orbits was caucurated.
- In addition, “Class”-column was created with 1 for a successful landing and 0 for a failed landing to create a predictive model.

- GitHub URL of the completed data wrangling related notebooks is:

https://github.com/OzasaHiro/Coursera_IBM_Datascience/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Scatter point charts were created to check the relationship between:
  - FlightNumber and Payload Mass
  - FlightNumber and Launch sites
  - Launch sites and Payload Mass
  - FlightNumber and Orbit type
  - Payload and Orbit type
- Bar chart was created to see the relationship between success rate and orbit typ
- Line chart was created to see the success rate per year

- GitHub URL of the completed EDA with data visualization notebook is:
https://github.com/OzasaHiro/Coursera_IBM_Datascience/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- Check the names of the unique launch sites in the space mission.
- Check 5 records where launch sites begin with the string 'CCA' .
- Check the total payload mass carried by boosters launched by NASA (CRS).
- Check average payload mass carried by booster version F9 v1.1.
- List the date when the first succesful landing outcome in ground pad was acheived.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000.
- List the total number of successful and failure mission outcomes.
- List the names of the booster_versions which have carried the maximum payload mass.
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- 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.
- GitHub URL of the completed EDA with SQL notebook is:
https://github.com/OzasaHiro/Coursera_IBM_Datascience/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- The location of the Launch site was confirmed on the map.
- The number of launches from each launch site was also displayed.
- In addition, the distance from each launch site to nearby city, railroad, and coastline was also displayed.

- GitHub URL of the completed interactive map with Folium map is:
  https://github.com/OzasaHiro/Coursera_IBM_Datascience/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Pie chart was created to check the Launch success count for all/each launch site.
- A slider was added to select payload range.
- Scatter chart was created to show the correlation between payload and launch success.

- GitHub URL of the completed Plotly Dash lab is:
  https://github.com/OzasaHiro/Coursera_IBM_Datascience/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

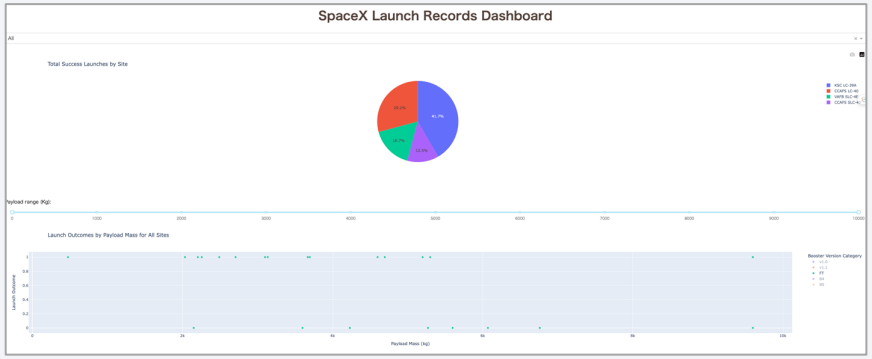
- 4 models were built using four machine learning models: Logistic regression, SVM, Decision tree, and KNN.
- Grid search was performed for each model to determine the optimal parameters.
- 20 percent of all data was used as test data for accuracy verification.
- The best model was determined by comparing the results of verification with test data from these four models.

- GitHub URL of the completed predictive analysis lab is:

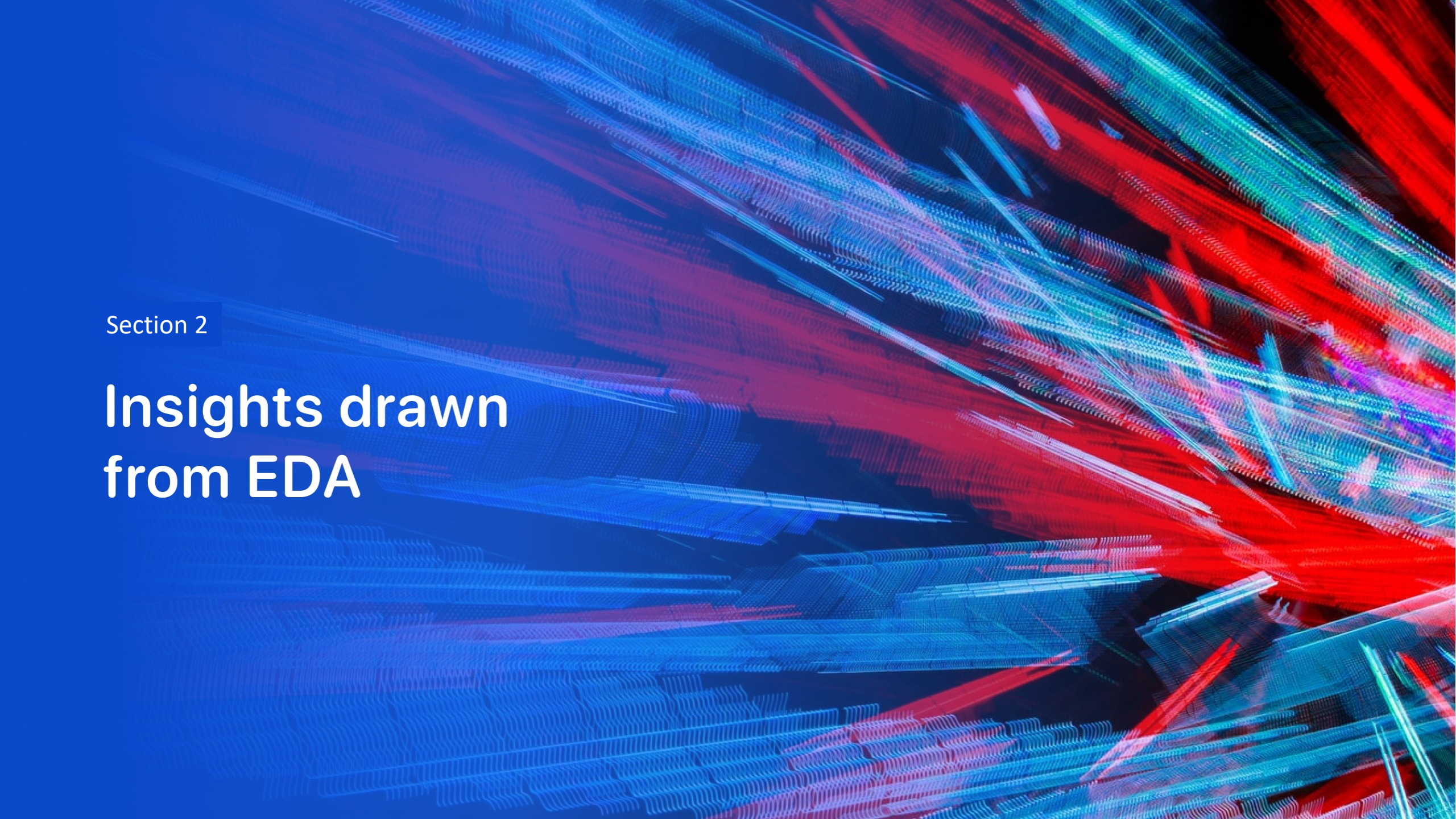
https://github.com/OzasaHiro/Coursera_IBM_Datascience/blob/main/SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Results

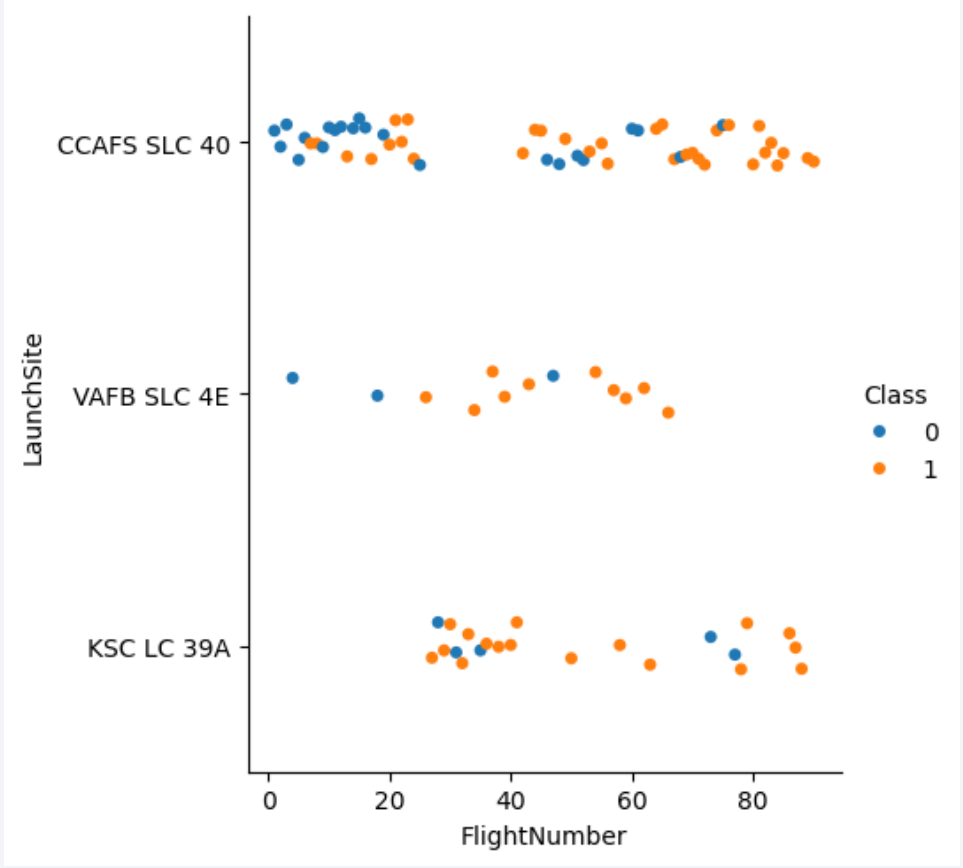
- Exploratory data analysis results
  - The sucess rate since 2013 kept increasing till 2020.
  - Success rates vary by Orbit type. ES-L1, GEO and SSO have a 100% success rate. On the other hand, SO has never been successful.
- Interactive analytics demo in screenshots



- Predictive analysis results
  - The validation results showed that the decision tree model was able to predict the successful landing of Falcon 9 1st-stage with more than 90% prediction accuracy.

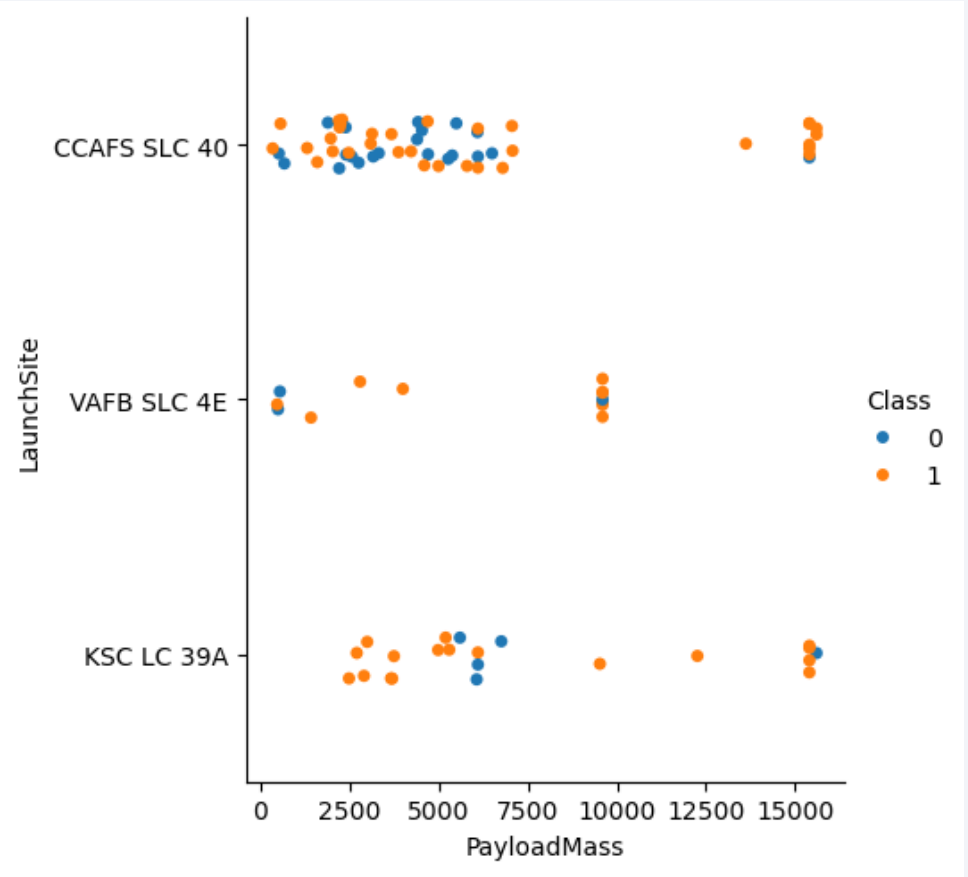

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



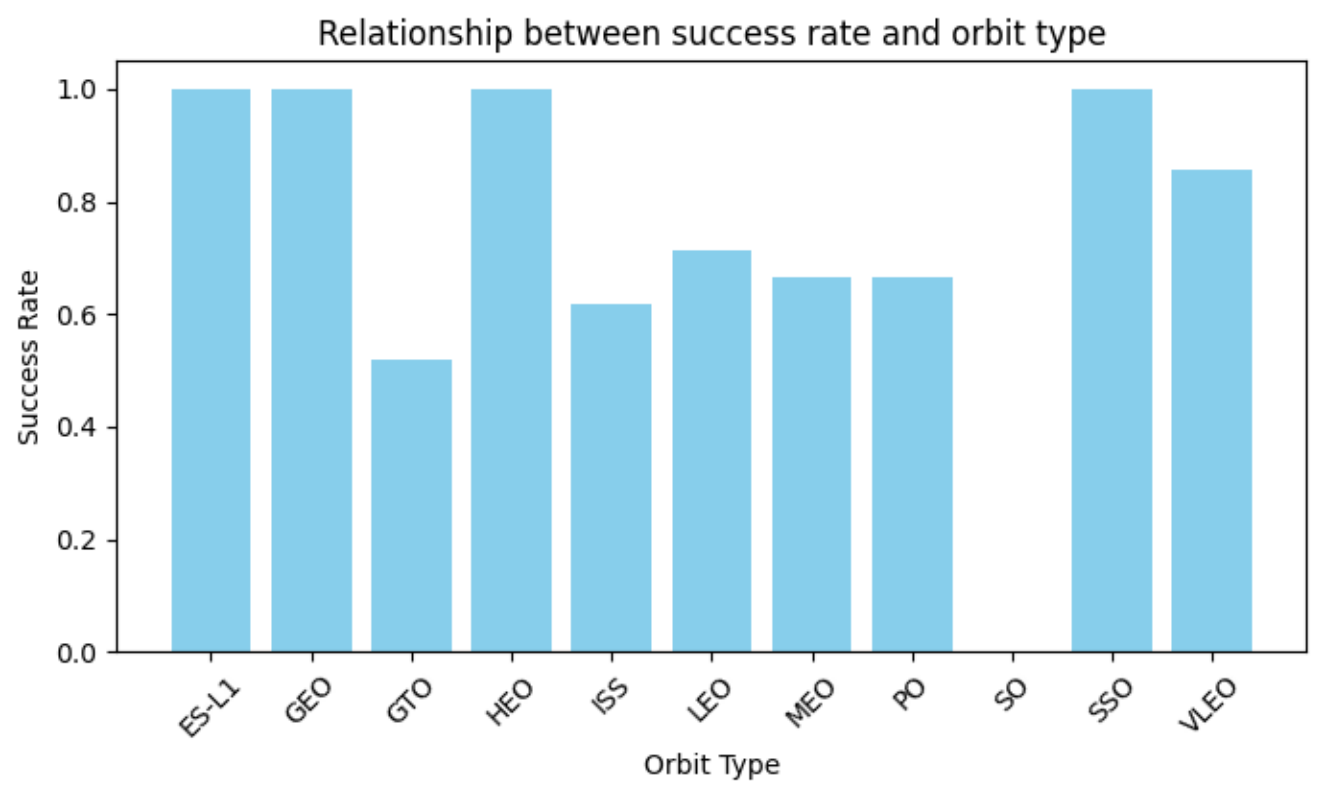
- VAFB-Launch site is not used after Flight No.66.

Payload vs. Launch Site



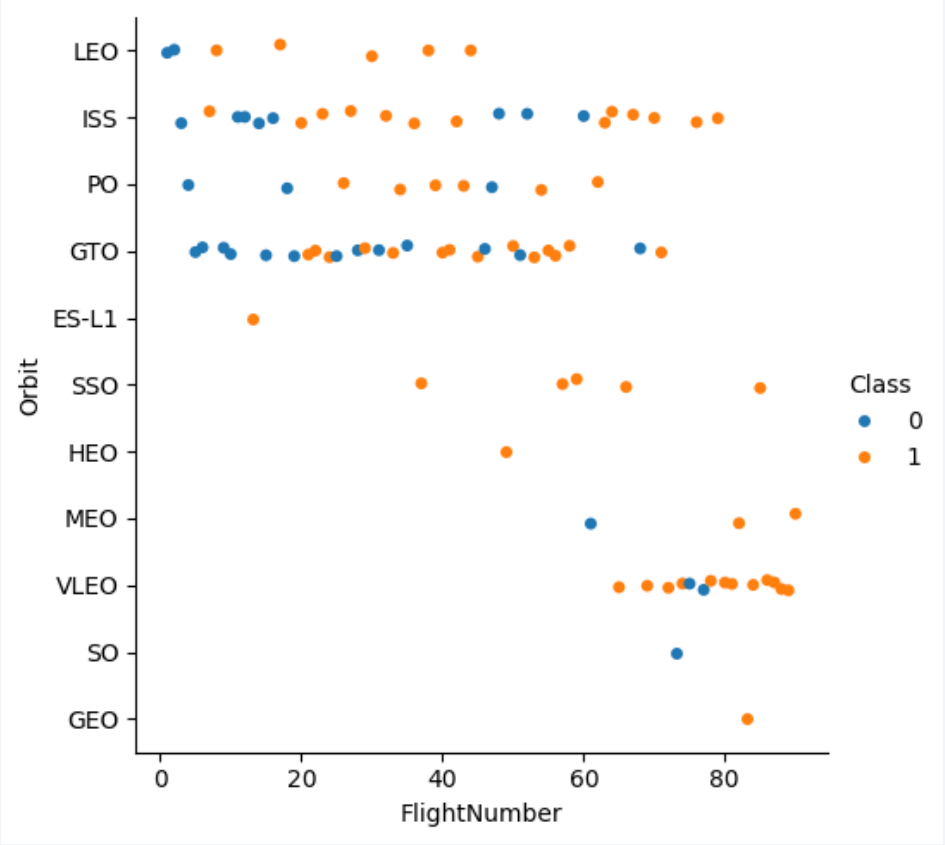
- For the VAFB-SLC launch-site there are no rockets launched for heavypayload mass(greater than 10,000).

Success Rate vs. Orbit Type



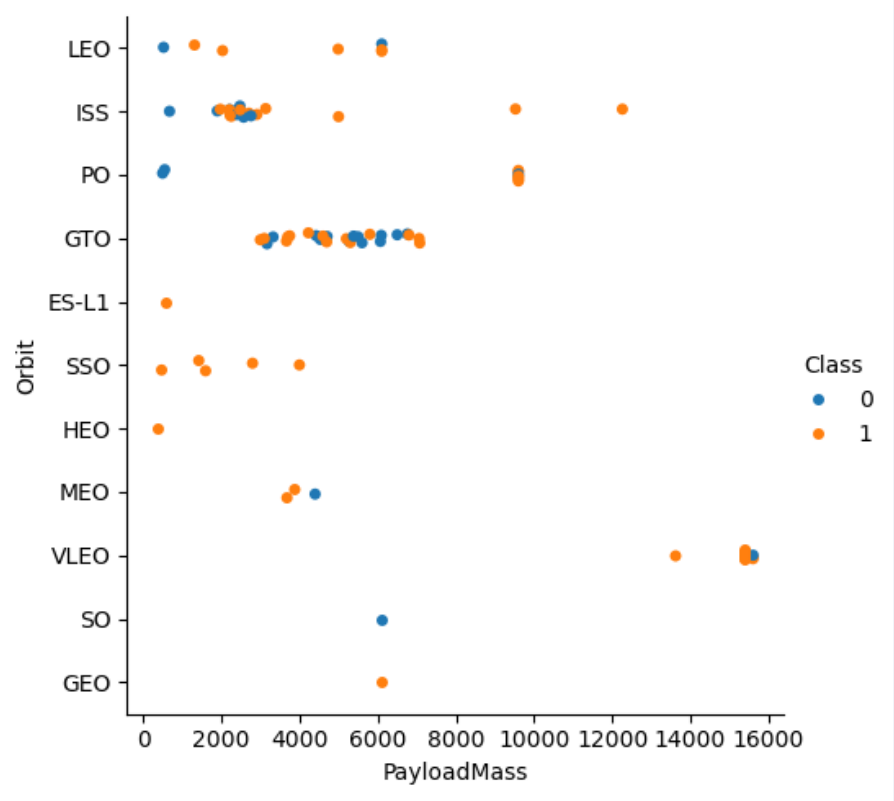
- ES-L1, GEO and SSO have a 100% success rate. On the other hand, SO has never been successful.

Flight Number vs. Orbit Type



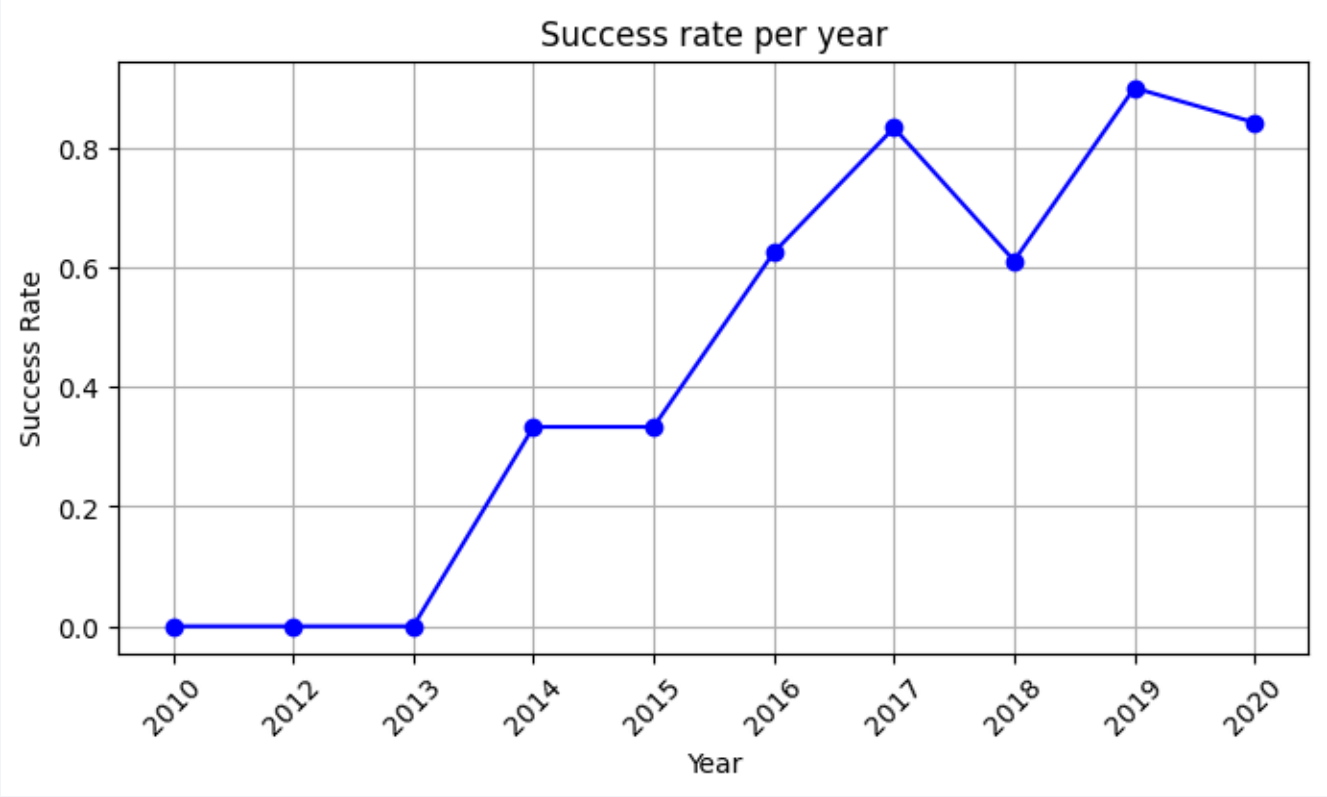
- In the LEO orbit, the success landing appears related to the number of flights.
- On the other hand, there seems to be no relationship between flight number when in GTO orbit.

Payload vs. Orbit Type



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

Launch Success Yearly Trend



- The sucess rate since 2013 kept increasing till 2020.

All Launch Site Names

Launch_Site	Note
VAFB SLC-4E	Vandenberg Air Force Base Space Launch Complex 4E
KSC LC-39A	Kennedy Space Center Launch Complex 39A
CCAFS SLC-40, CCAFS LC-40	Cape Canaveral Space Launch Complex 40

- There are a total of four launch sites at Vandenberg Air Force Base and Kennedy Space Center.

Launch Site Names Begin with 'CCA'

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

- The only booster-version used is F9 v1.0.
- Orbit is only for LEO and often for ISS.

Total Payload Mass

TotalPayloadMass
45596

- The total payload carried by boosters from NASA is 45,596 kg.

Average Payload Mass by F9 v1.1

AveragePayloadMass
2928.4

- The average payload mass carried by booster version F9 v1.1 is 2,928.4 kg.

First Successful Ground Landing Date

FirstSuccessfulLandingDate
2015-12-22

- The first successful landing on the Ground pad was in 12/22/2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

- Boosters which had payload mass greater than 4000 but less than 6000, and have successfully landed on drone ship is only F9 FT.

Total Number of Successful and Failure Mission Outcomes

TotalSuccess	TotalFailure
100	1

- The mission itself has had 100 successes and only one failure.

Boosters Carried Maximum Payload

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 which have carried the maximum payload mass is F9 B5.

2015 Launch Records

Month	Booster_Version	Launch_Site	Landing_Outcome
01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- There were two failed landings in drone-ship in 2015.
- In both cases, the booster version was F9 v1.1, and the launch site was CCAFS LC-40.

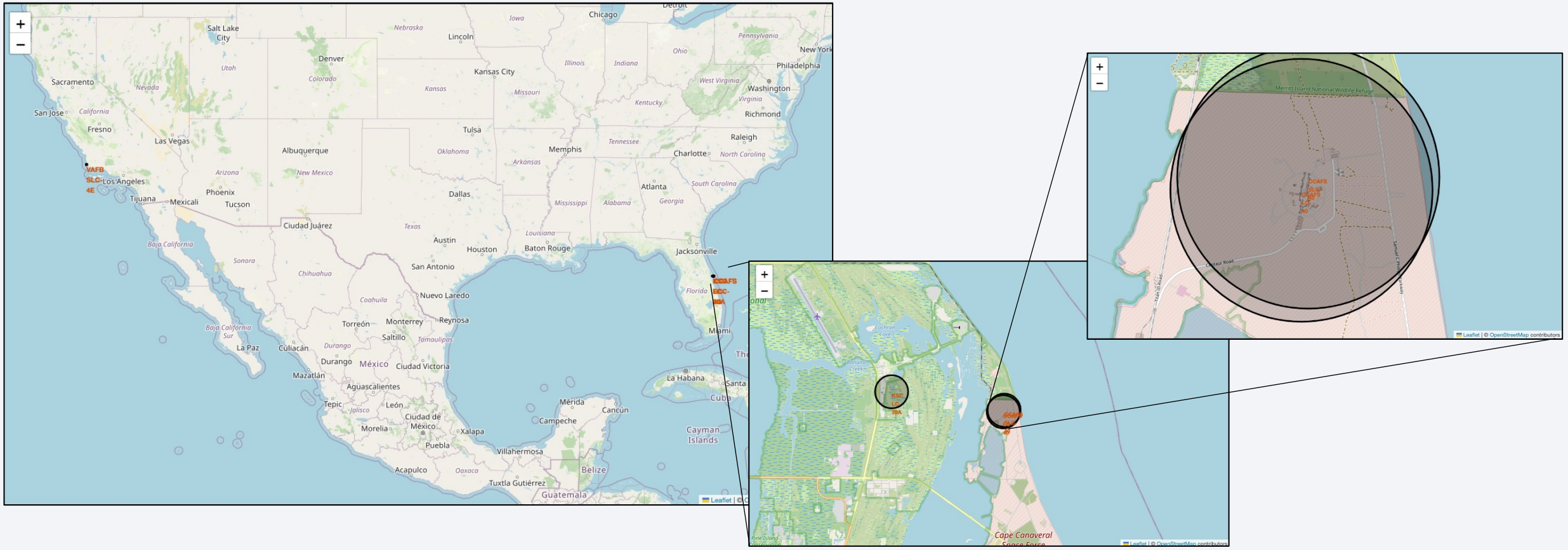
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

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

- Between 6/4/2010 and 3/20/2017, there were five successful and five unsuccessful landings on the drone-ship, respectively.

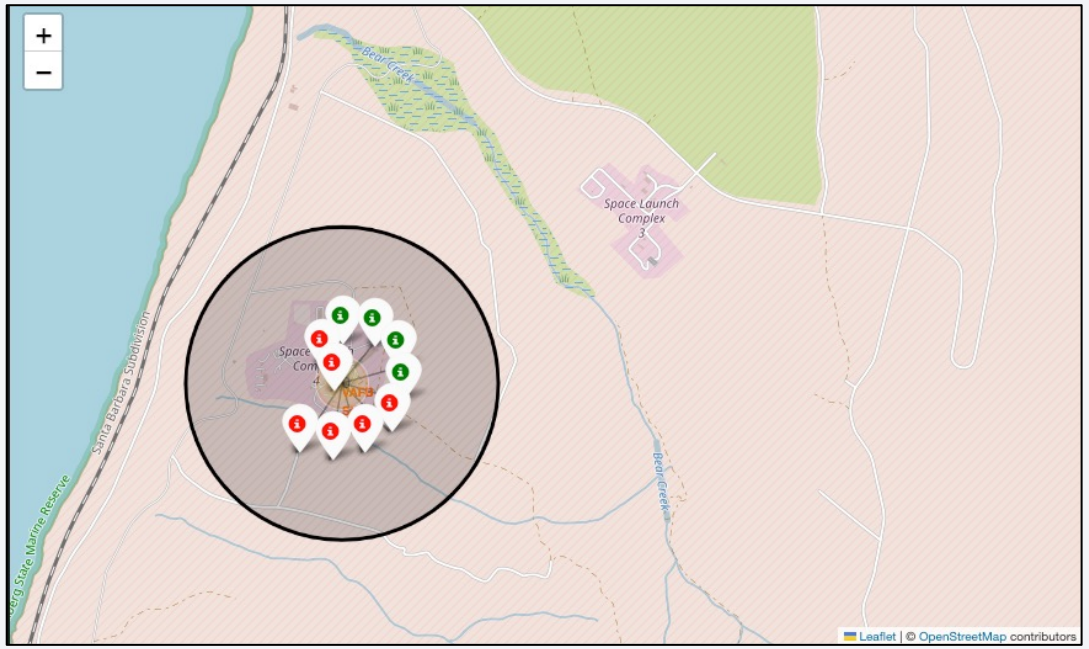

Section 3
Launch Sites
Proximities Analysis

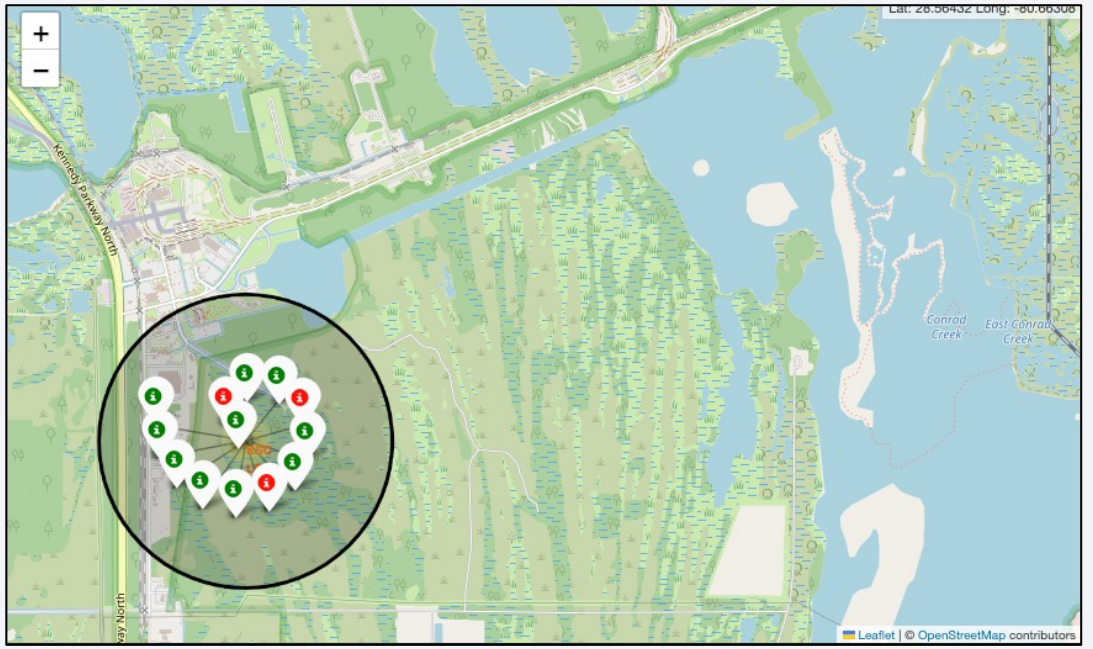
Launch site location



- There are one launch site at Vandenberg Air Force Base.
- There are three launch site at Kennedy Space Center.

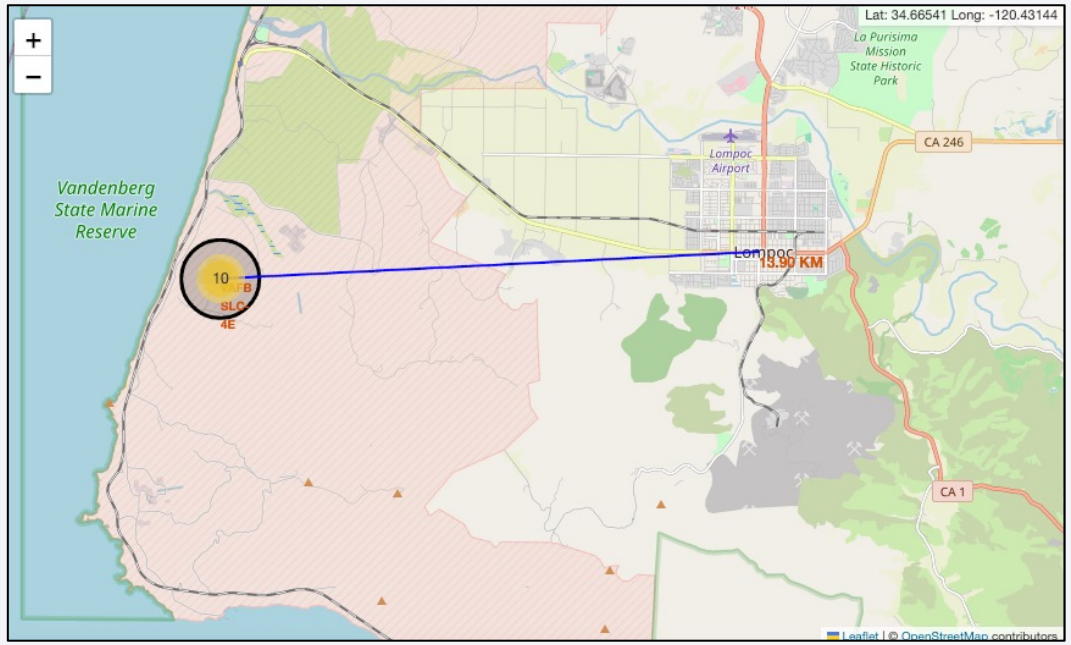
Launch outcomes

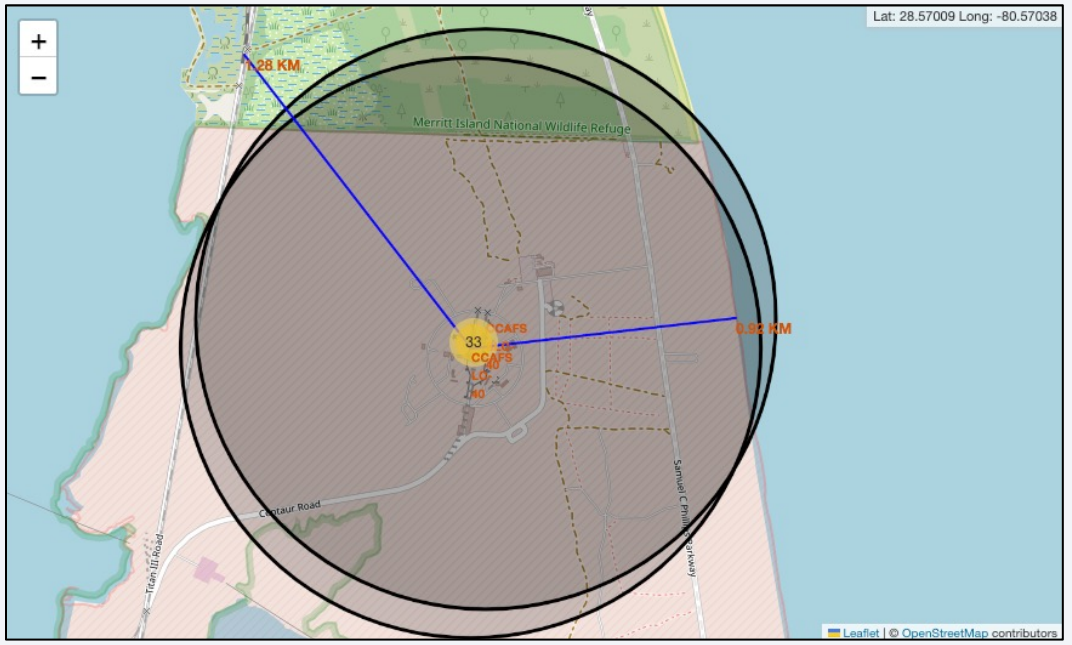




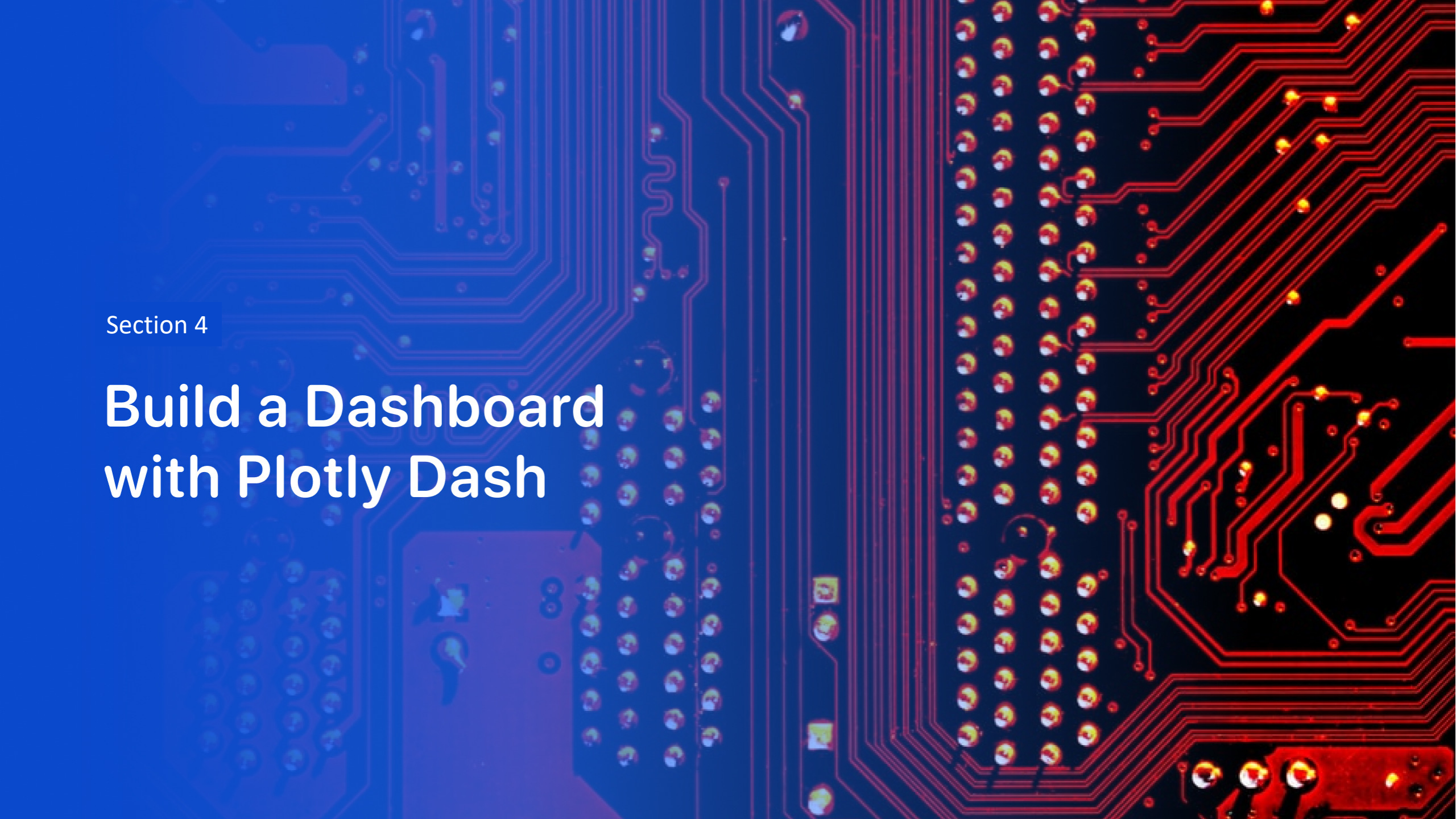
- Launch outcomes at VAFB SLC-4E(LEFT) and KSC LC-39A(RIGHT)
- Green markers indicate success launches and red markers indicate failed launches.
- KSC LC-39A has a higher launch count and success rate than VAFB SLC-4E.

Distance from launch site to proximities

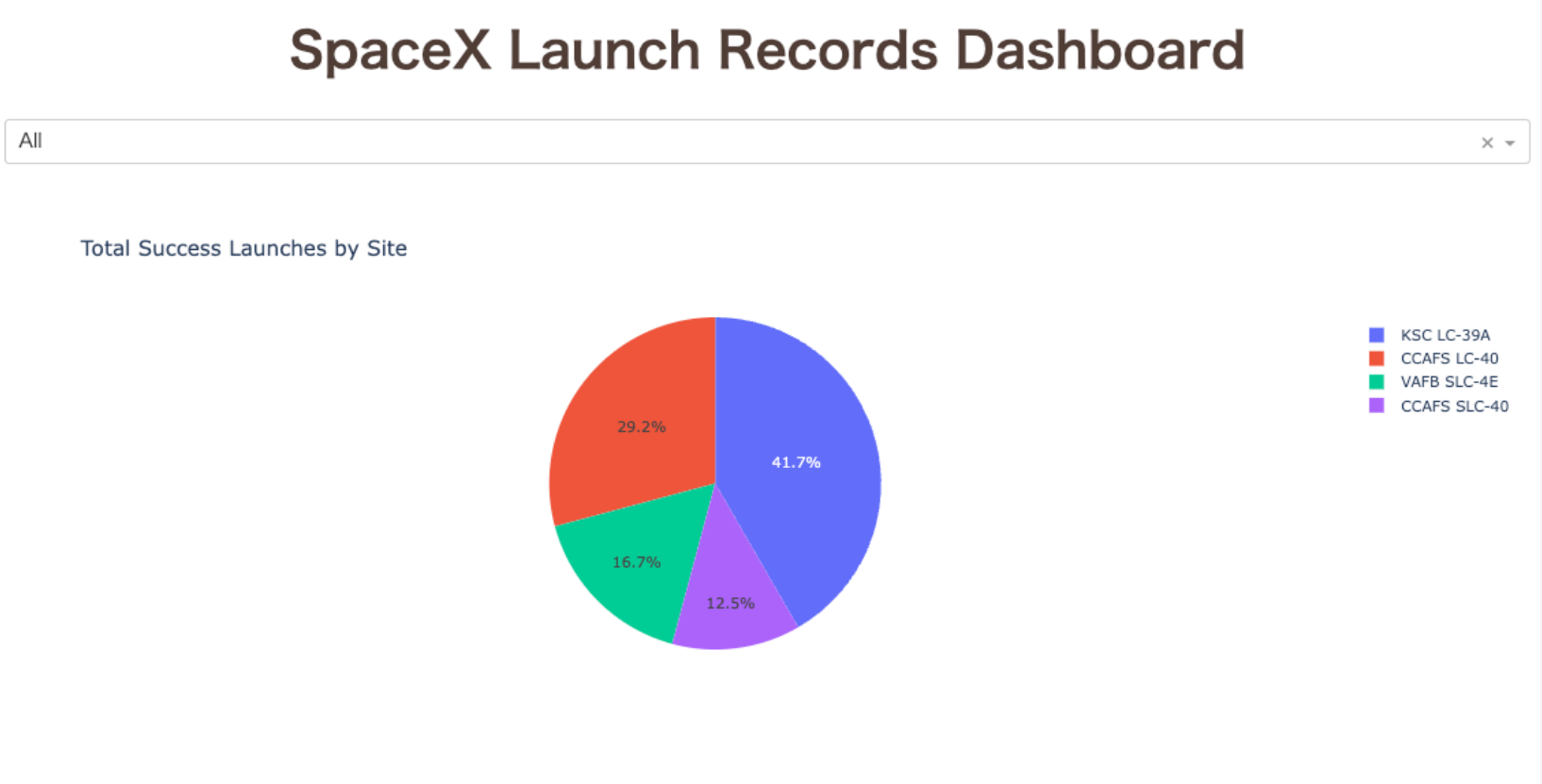




- Left shows the distance between VAFB SLC-4E to the nearest city, Lompoc. It is approximately 14 km.
- Right shows the distance from CCAFS SLC-40 to the nearest railway and the coastline. The distance to the railway is approx. 1.3 km, to the coastline is approx. 0.9 km.

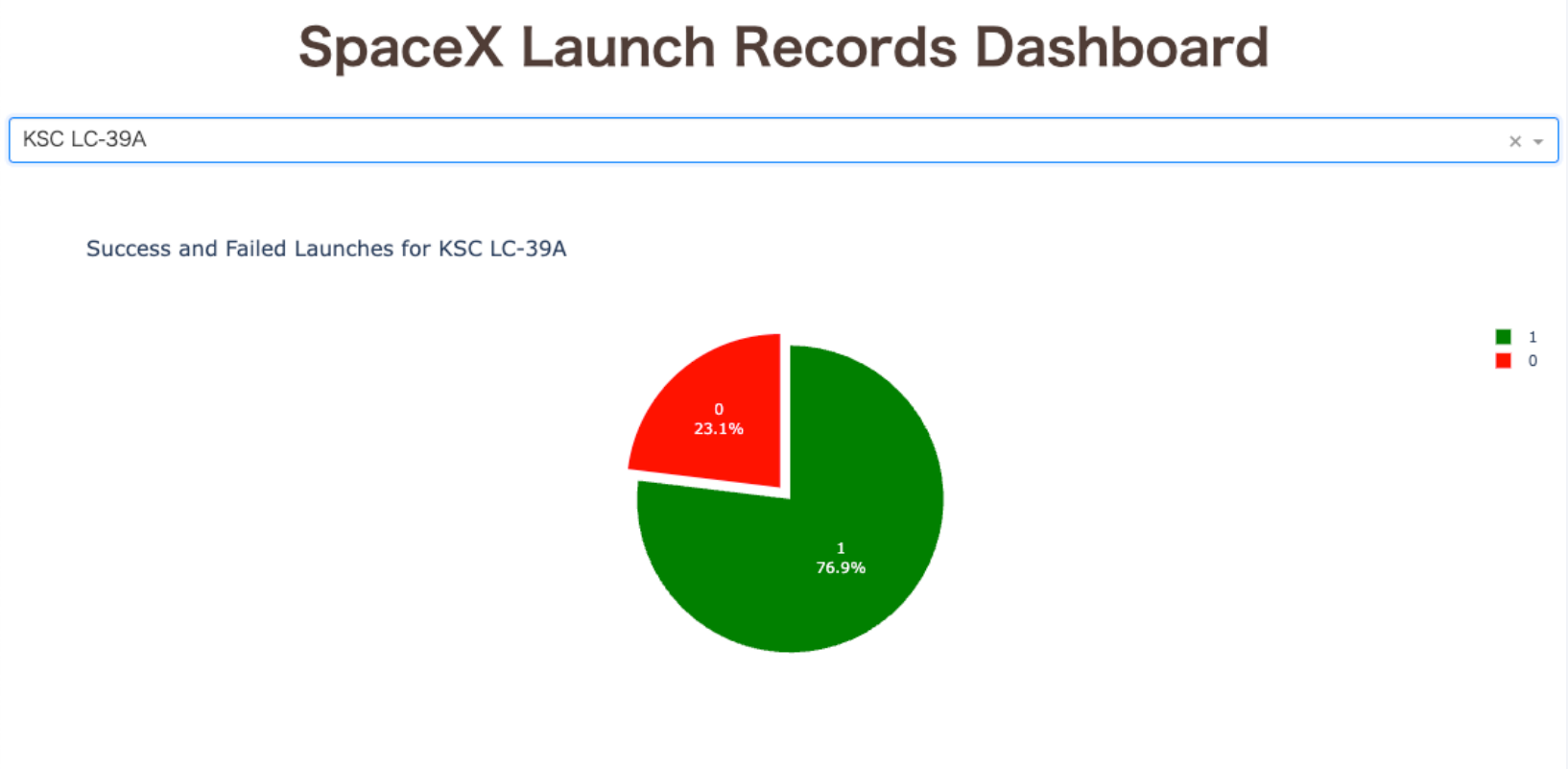

Section 4
Build a Dashboard with Plotly Dash

Launch success count for all launch sites



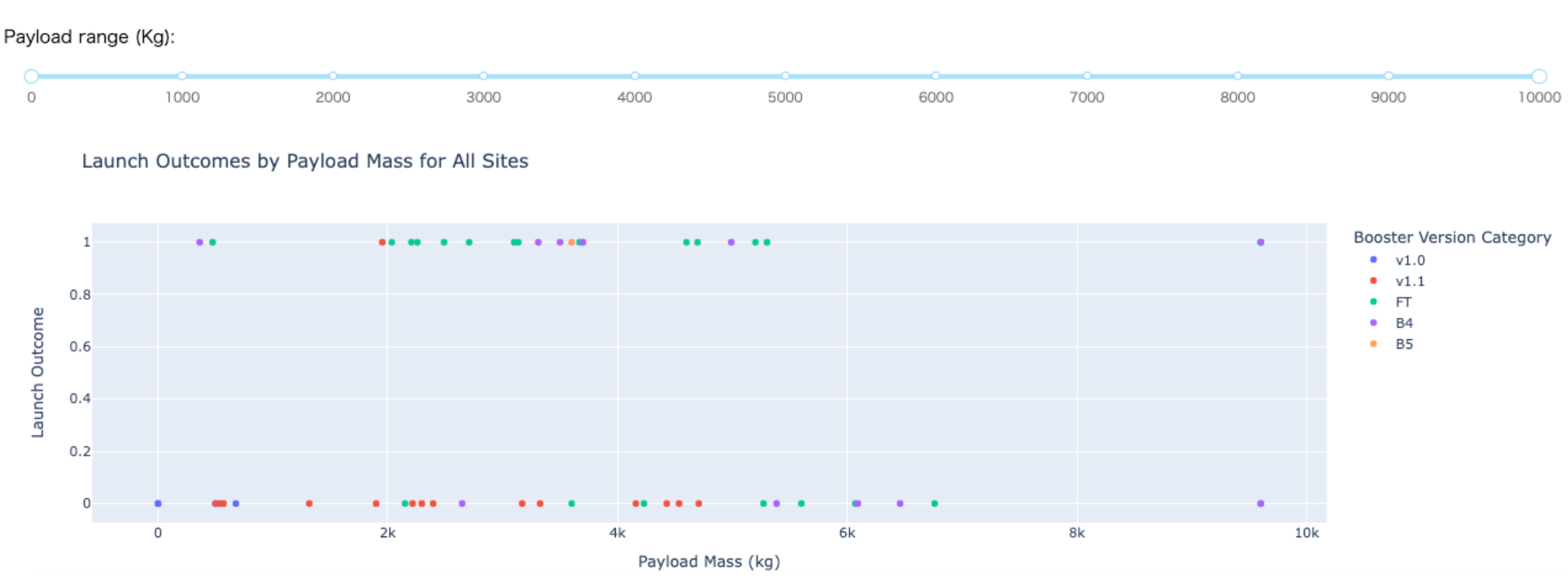
- The launch site with the highest number of successes is KSC LC-39A.

Highest launch success ratio



- The highest success rate of all Launch sites is KSC LC-39A.
- Its success rate is 76.9 %.

Launch outcomes by Payload

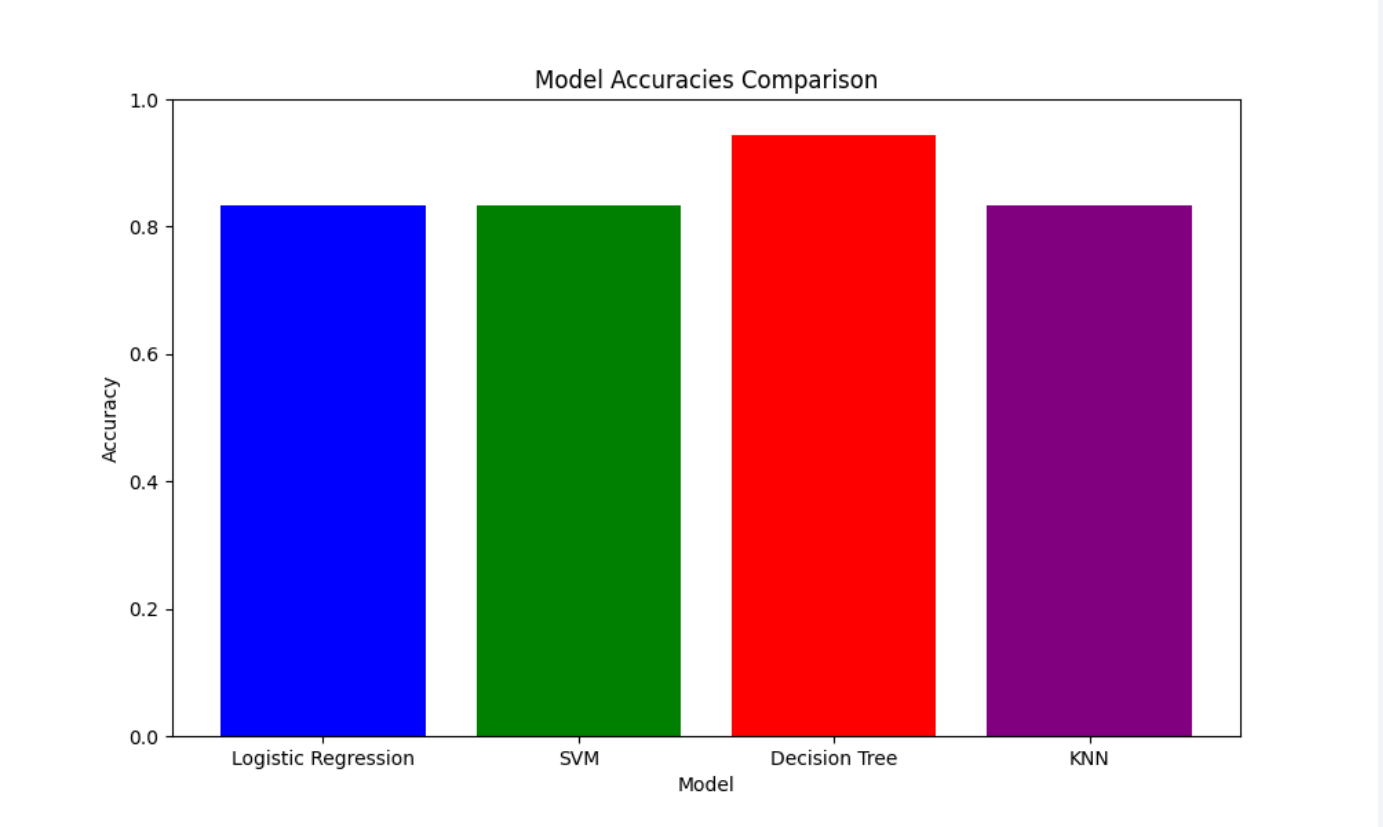


- The most successful booster-version is FT.
- The success rate is lower when the payload is above 6,000.

Section 5

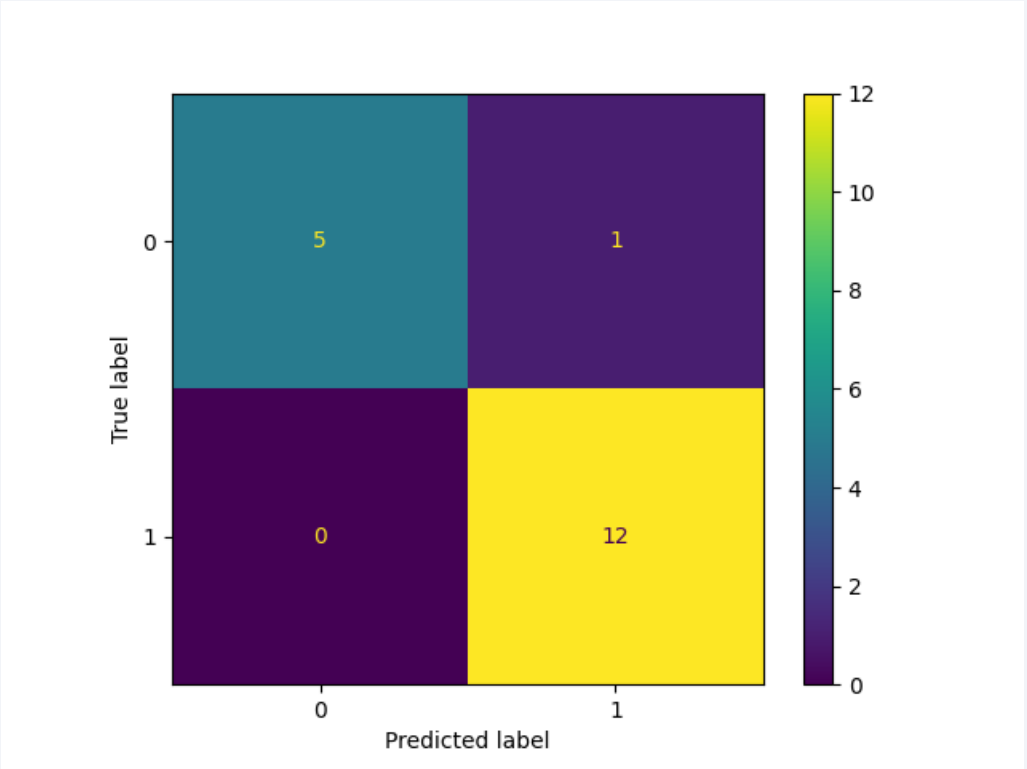
Predictive Analysis (Classification)

Classification Accuracy



- By applying the Decision Tree, a prediction accuracy of more than 90% was achieved.

Confusion Matrix



- In the decision tree model, there was only one False Positive (FP) out of 18 tests that were incorrectly predicted.

Conclusions

- Point 1

Decision tree model is able to predict the successful landing of Falcon 9 1st-stage with more than 90% prediction accuracy.

- Point 2

The sucess rate since 2013 kept increasing till 2020.

- Point 3

Success rates vary by Orbit type. ES-L1, GEO and SSO have a 100% success rate. On the other hand, SO has never been successful.

- Point 4

Since the number of SO and other orbit-types is likely to increase in the future, it would be advisable to periodically verify the performance of the prediction model.

Thank you!