IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Grzegorz Szulik
26.07.2024

Outline

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

Executive Summary

- Summary of methodologies
  - Different classification techniques were used in order to dermine if the Falcon 9 first stage will land successfully.
- Summary of all results
  - the best model achieved – Tree Decision Classifier with an accuracy of 90%.

Introduction

- Project background and context

Spaces travels are really expensive that is why we need to be able to predict a successful landings.

- Problems you want to find answers

We want to achieve a model that will give us the answer if a space journey with a given set of parameters will end with a successful landing.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - api.spacexdata.com/v4/launches/past dataset was used, the dataset was converted first to JSON object and finally to Pandas DataFrame.
- Perform data wrangling
  - various calculations were performed, finally landing outcome column was created (Classes).
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - different models were tested: logistic regression, SVM, Decision Tree and KNN.

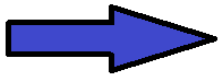
Data Collection

- A GET request was performer in order to achieve data from https://api.spacexdata.com/v4/launches/past
- A JSON object was created and then converted to Pandas DataFrame
- The DataFrame was filtered in order to include only Falcon 9 launches
- Missing values of PayloadMass were replaced with a mean value.

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004

Data Collection – SpaceX API

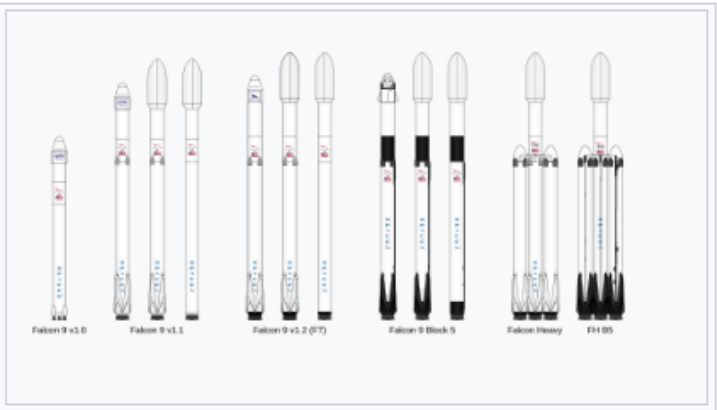

```
{
  "fairings": {
    "reused": false,
    "recovery_attempt": false,
    "recovered": false,
    "ships": []
  },
  "links": {
    "patch": {
      "small": "https://images2.imgbox.com/94/f2/NN6Ph45r_o.png",
      "large": "https://images2.imgbox.com/5b/02/QcxHUb5V_o.png"
    },
    "reddit": {
      "campaign": null,
      "launch": null,
      "media": null,
      "recovery": null
    },
    "flickr": {
      "small": [],
      "original": []
    },
    "presskit": null,
    "webcast": "https://www.youtube.com/watch?v=0a_00nJ_Y88",
    "youtube_id": "0a_00nJ_Y88",
    "article": "https://www.space.com/2196-spacex-inaugural-falcon-1-rocket-lost-launch.html",
    "wikipedia": "https://en.wikipedia.org/wiki/DemoSat"
  },
  "static_fire_date_utc": "2006-03-17T00:00:00.000Z",
  "static_fire_date_unix": 1142553600,
  "net": false,
  "window": 0,
  "rocket": "5e9d0d95eda69955f709d1eb",
  "success": false,
  "failures": [
    {
      "time": 33,
      "altitude": null,
      "reason": "merlin engine failure"
    }
  ],
```

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004

https://github.com/GrzegorzSzulik/Capstone/blob/main/SpaceX-data-collection-api.ipynb

Data Collection - Scraping



Left to right: Falcon 9 v1.0, v1.1, v1.2 "Full Thrust", Falcon 9 Block 5, Falcon Heavy, and Falcon Heavy Block 5.

BeautifulSoup

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n
1	2	CCAFS	Dragon	0	LEO	NASA	Success
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n

https://github.com/GrzegorzSzulik/Capstone/blob/main/Webscraping.ipynb

Data Wrangling

The following data (columns) were calculated:

- number of launches on each site
- number and occurence of each orbit
- landing outcome column (Class)

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0

https://github.com/GrzegorzSzulik/Capstone/blob/main/SpaceX-Data%20wrangling.ipynb

EDA with Data Visualization

In order to understand the data the following charts were created:

- FlightNumber vs. PayloadMass
- FlightNumber vs LaunchSite
- PayloadMass vs Launch Site
- FlightNumber vs Orbit type
- PayloadMass vs Orbit
- Success Rate vs Year

https://github.com/GrzegorzSzulik/Capstone/blob/main/EDA-dataviz.ipynb

EDA with SQL

In order to better understand data the following queries were performed:

- the names of the unique launch sites in the space mission
- 5 records where launch sites begin with the string 'CCA'
- the total payload mass carried by boosters launched by NASA (CRS)
- average payload mass carried by booster version F9 v1.1
- date when the first succesful landing outcome in ground pad was achieved
- names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- total number of successful and failure mission outcomes
- names of the booster_versions which have carried the maximum payload mass. Use a subquery
- records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015
- 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

https://github.com/GrzegorzSzulik/Capstone/blob/main/EDA-SQL-coursera_sqllite.ipynb

Build an Interactive Map with Folium

In order to better understand the data the following elements were added to a map:

- markers of all launch sites
- success/failed launches were marked
- distances between a launch site and its proximities were calculated and marked with lines

https://github.com/GrzegorzSzulik/Capstone/blob/main/Launch_site_location.ipynb

Build a Dashboard with Plotly Dash

In order to better understand success sites the following plots were created:

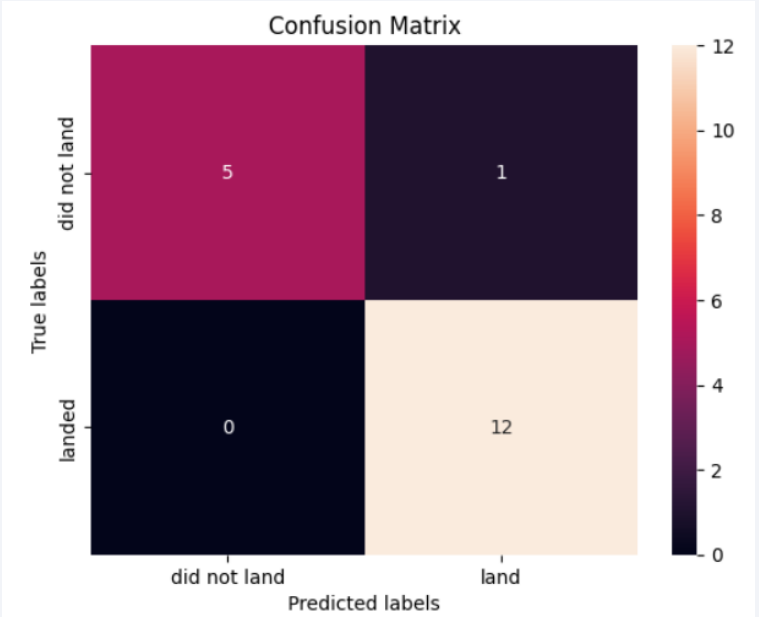
- launches status chart for all sites
- lunches status chart for each site
- correlation between PayloadMass and Success

https://github.com/GrzegorzSzulik/Capstone/blob/main/SpaceX_dash_app.py

Predictive Analysis (Classification)

The best performing classification model was obtained:

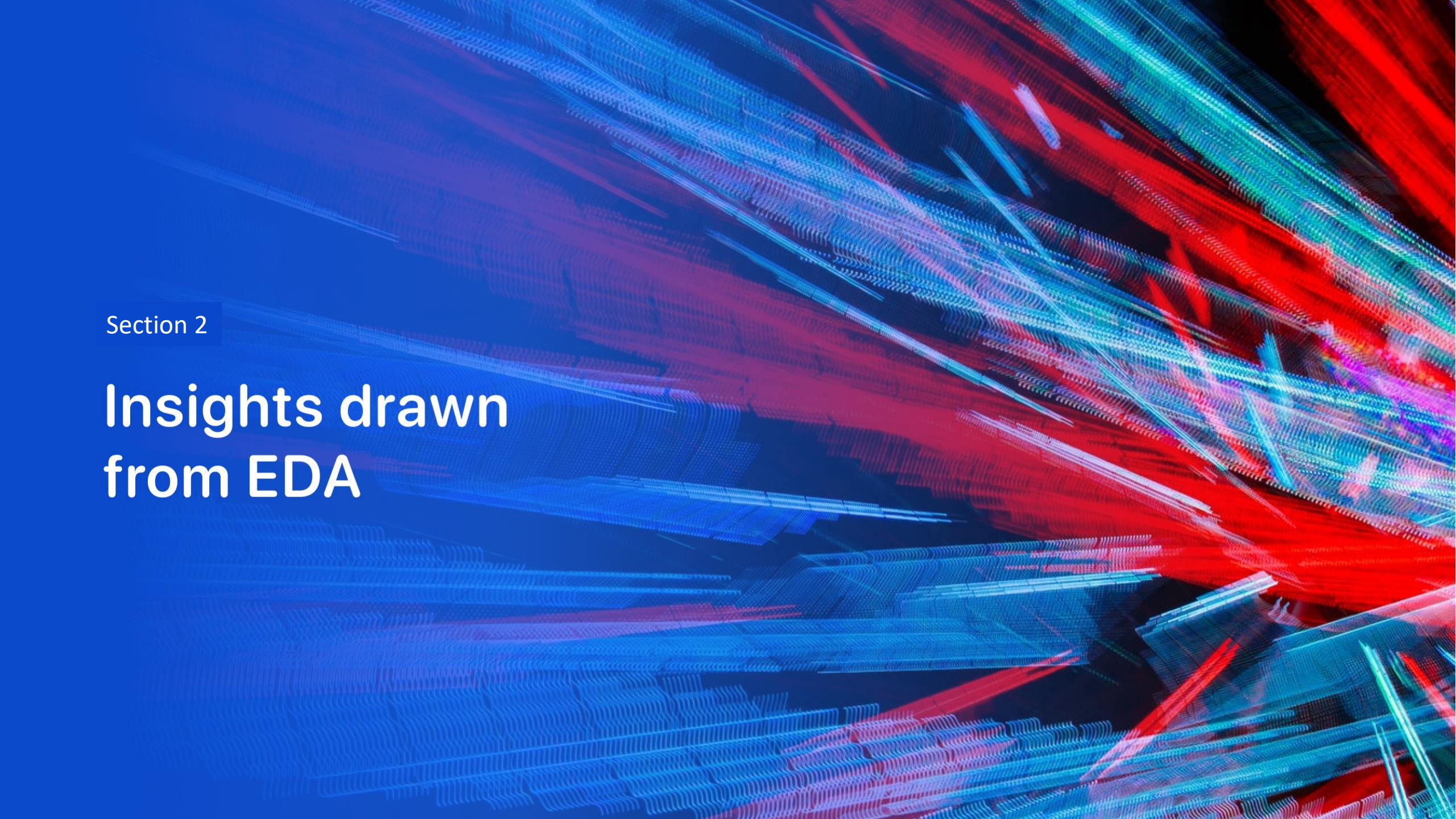
- Tree Decision Classifier
- accuracy: 0.9
- score: 0.94



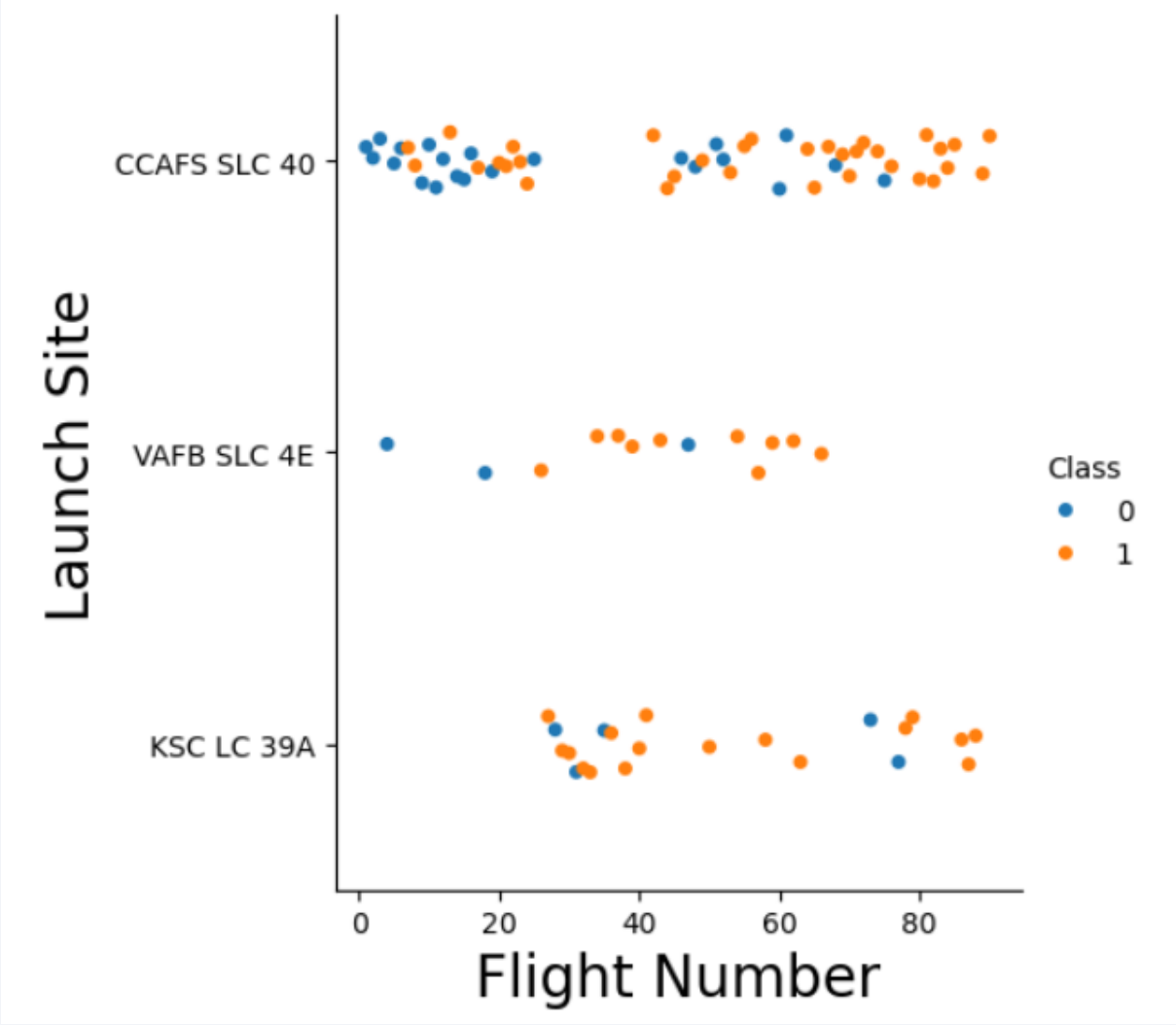
https://github.com/GrzegorzSzulik/Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

- Many data investigations were carried to ifentify the key features.
- Interactive analytics demo was created in order to understand success rates for different launch sites.
- Decision Tree Classifier was created with the best results with accuracy of 0.9 and score of 0.94.

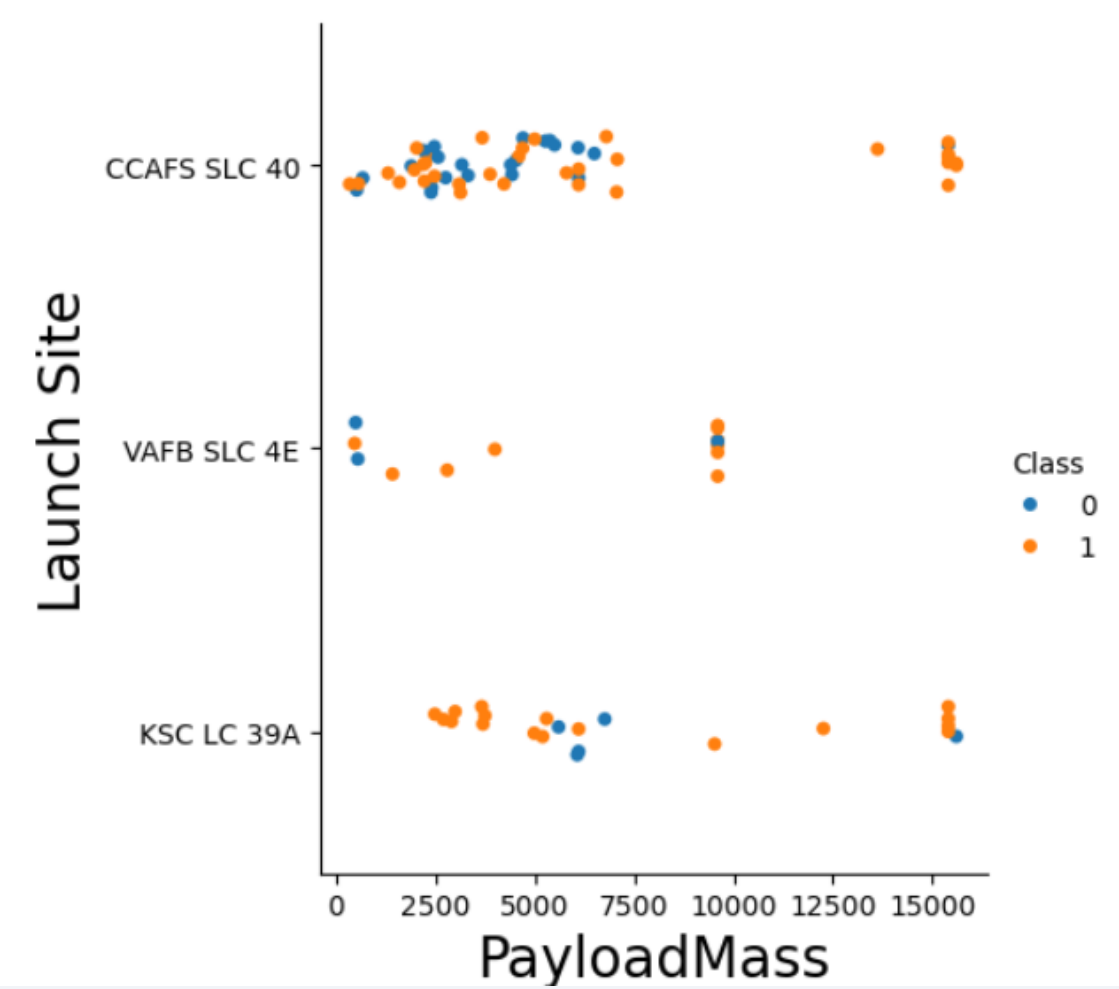

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



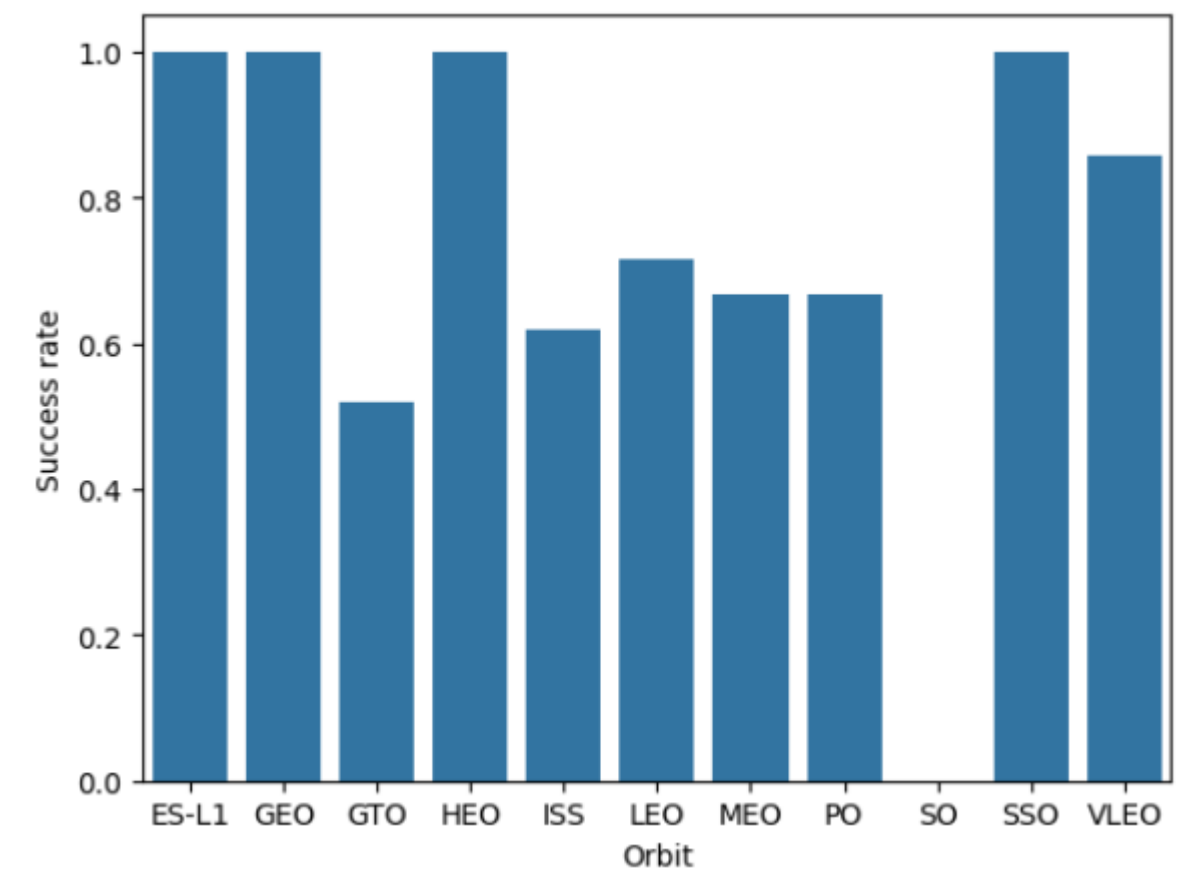
The success is more probable when the flight number ishigher.

Payload vs. Launch Site



Greated PayloadMass gives a higher chance for success.

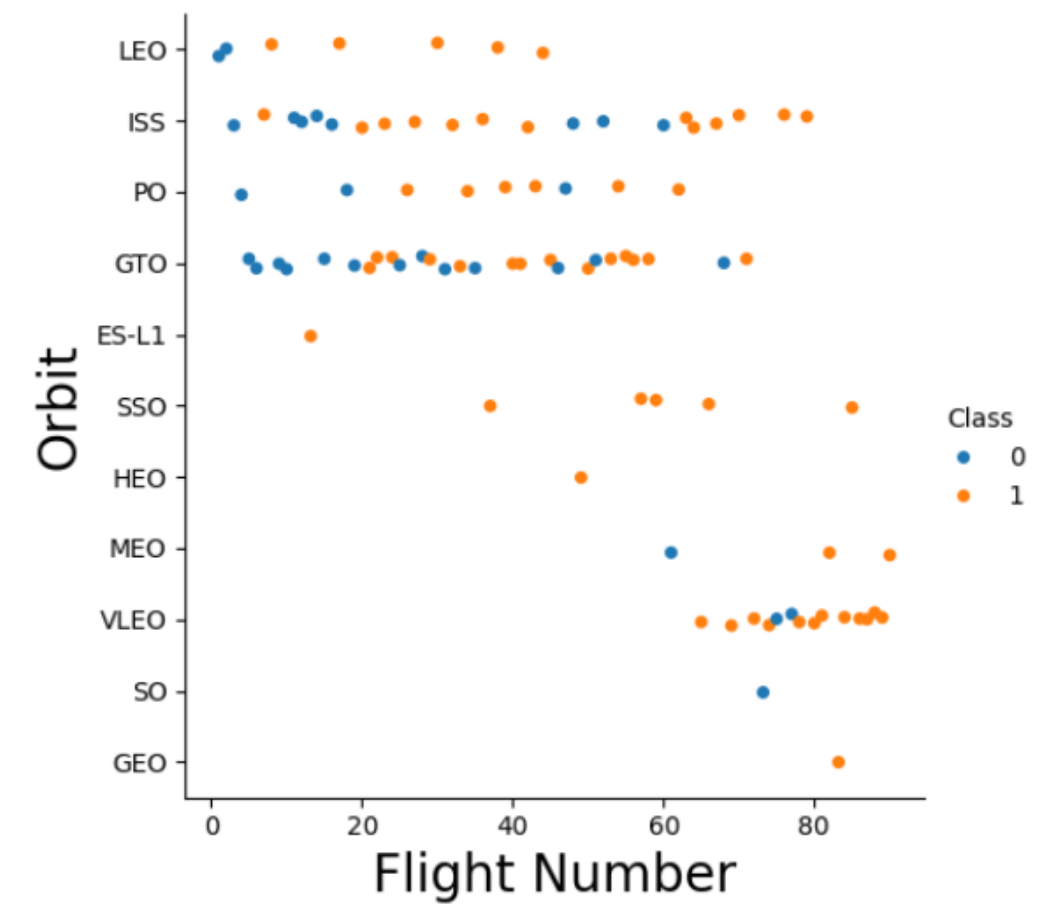
Success Rate vs. Orbit Type



Orbits with highest success rates:

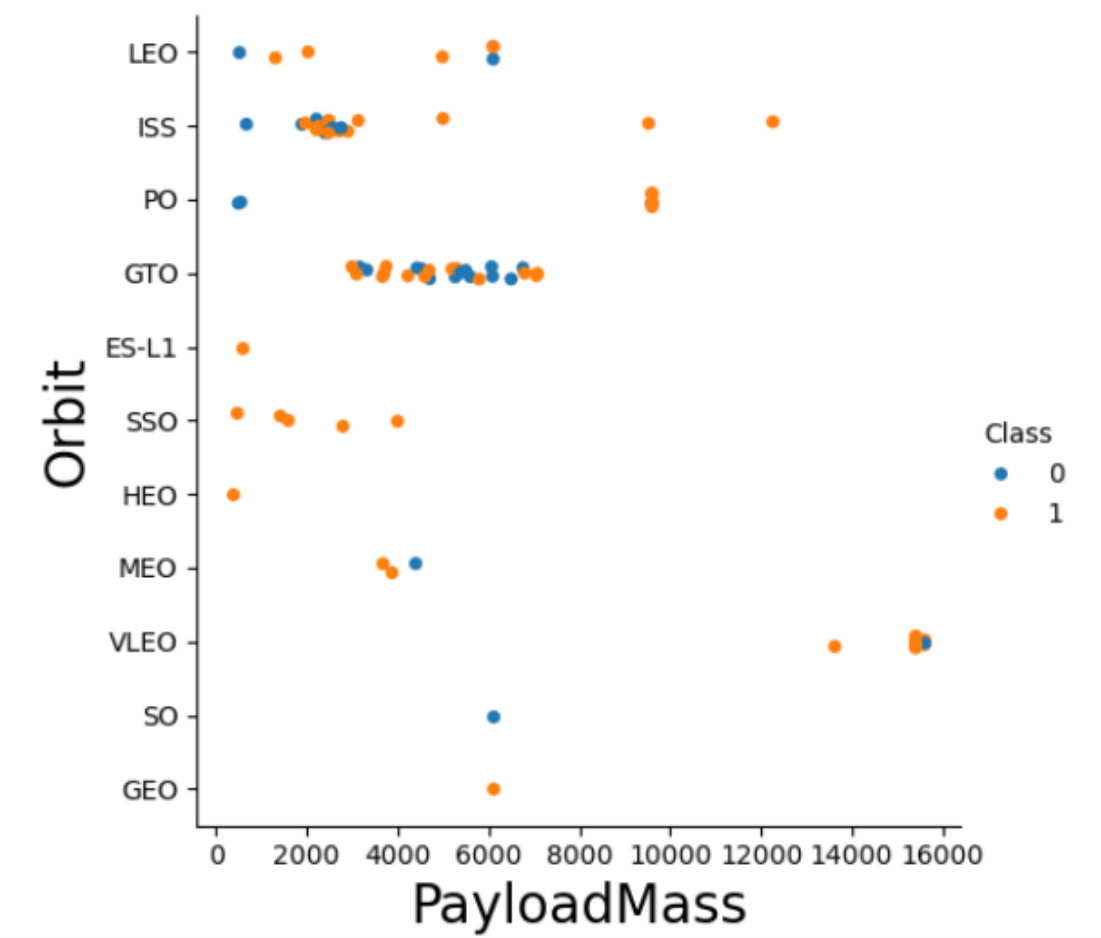
- ES-L1
- GEO
- HEO
- SSO

Flight Number vs. Orbit Type



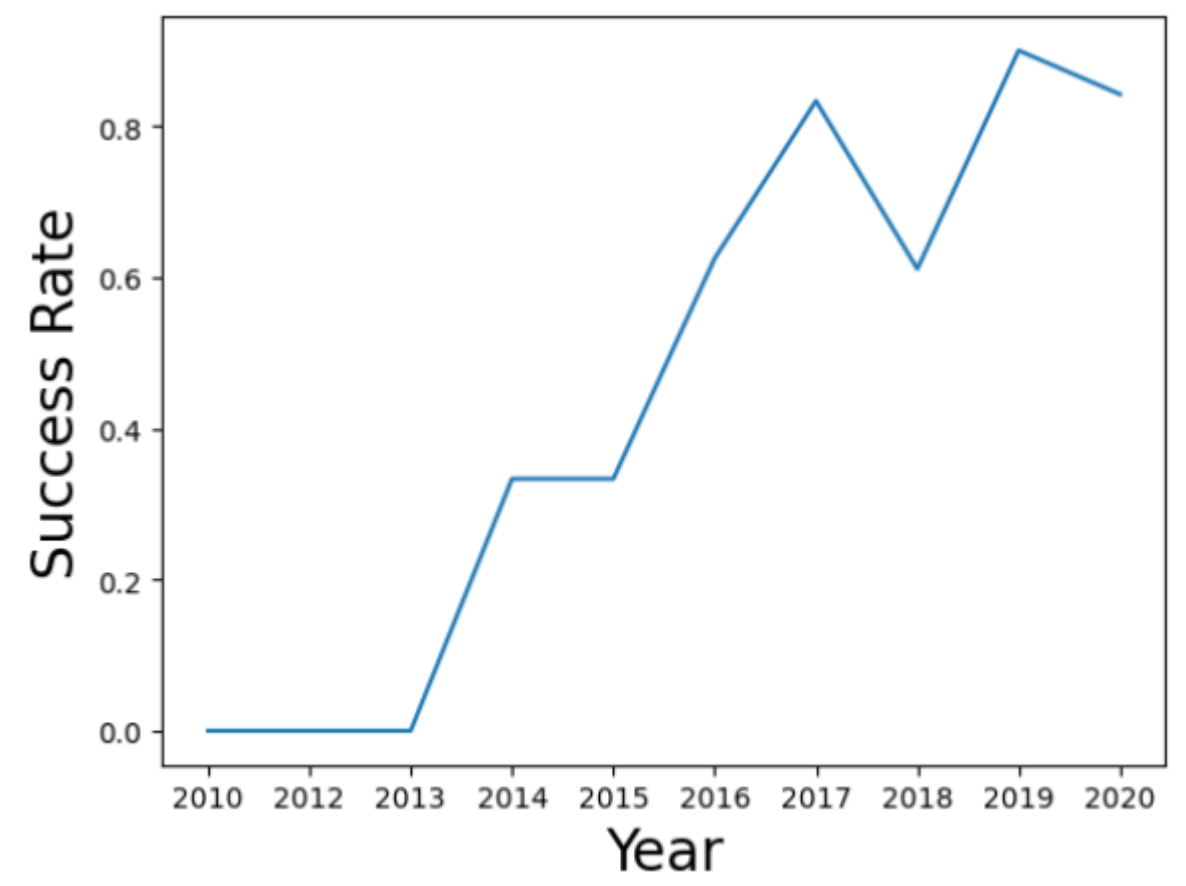
Higher Flight Number gives higher probability of success for all orbits.

Payload vs. Orbit Type



Higher PayloadMass increases chances for success to the most of orbits.

Launch Success Yearly Trend



Success rates in general is increasing since 2013.

All Launch Site Names

The names of the unique launch sites:

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

There are 4 of them.

Launch Site Names Begin with 'CCA'

5 records where launch sites begin with `CCA`:

- CCAFS LC-40
- CCAFS LC-40
- CCAFS LC-40
- CCAFS LC-40
- CCAFS LC-40

Total Payload Mass

The total payload carried by boosters from NASA:

- 45 596 kg

Average Payload Mass by F9 v1.1

The average payload mass carried by booster version F9 v1.1:

- 2534.67 kg

First Successful Ground Landing Date

The date of the first successful landing outcome on ground pad:

- 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

The names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000:

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

Total Number of Successful and Failure Mission Outcomes

The total number of:

- successful – 100
- failure – 1

mission outcomes.

Boosters Carried Maximum Payload

The names of the booster which have carried the maximum payload mass:

- 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

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

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

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

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

Landing_Outcome	Count
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

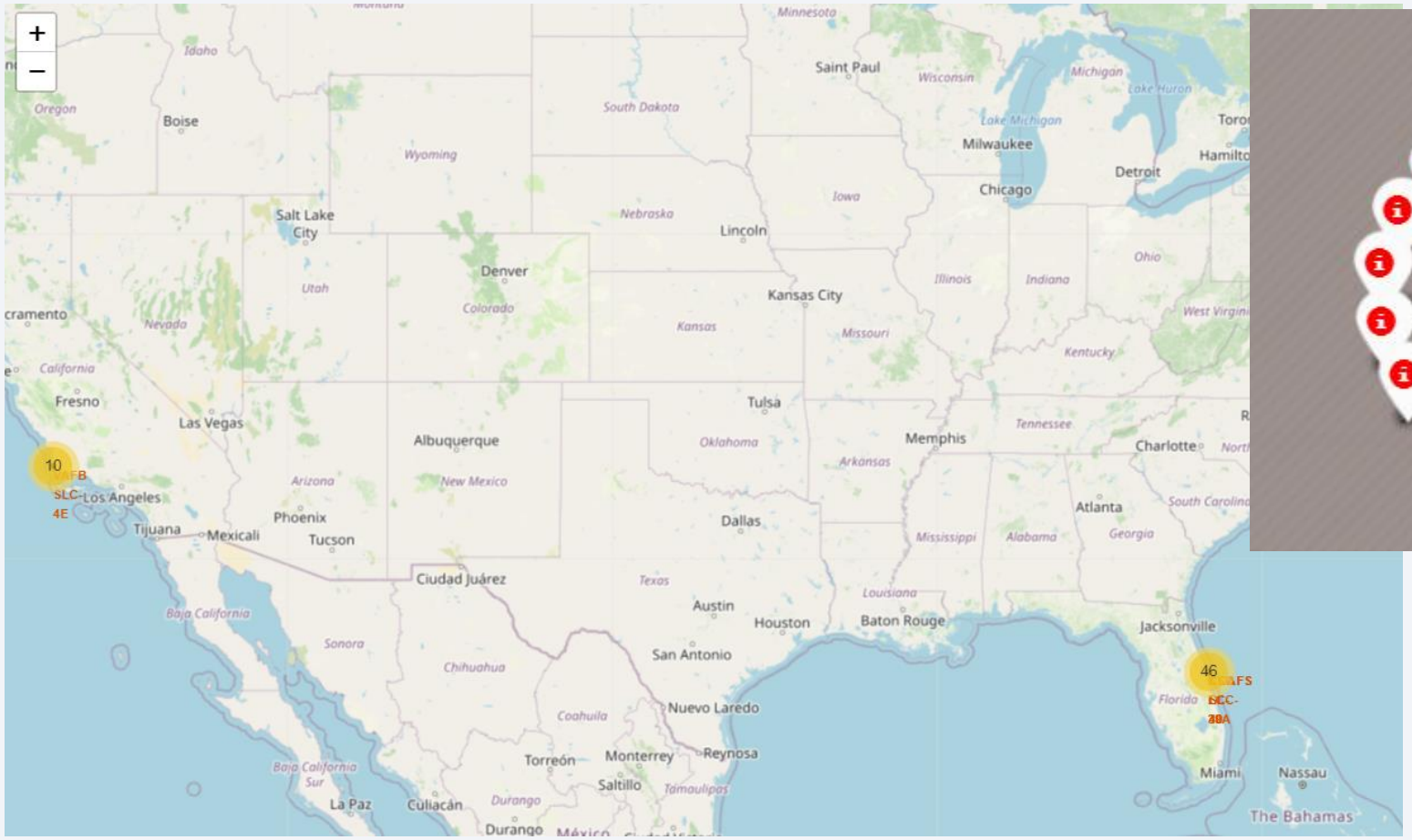
Section 3

Launch Sites Proximities Analysis

Site locations

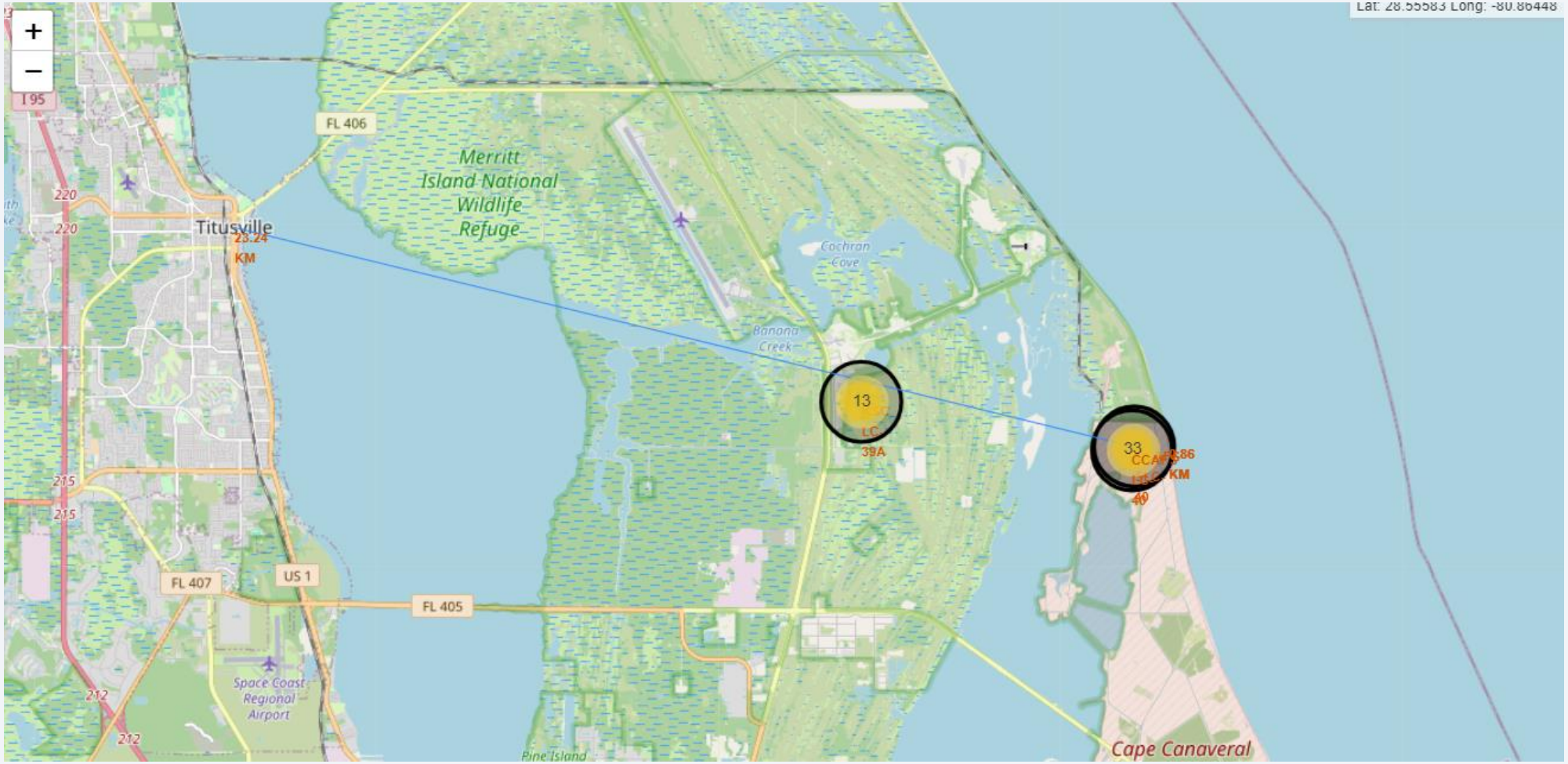
All sites marked on the map.

Map with clusters and markers of landings

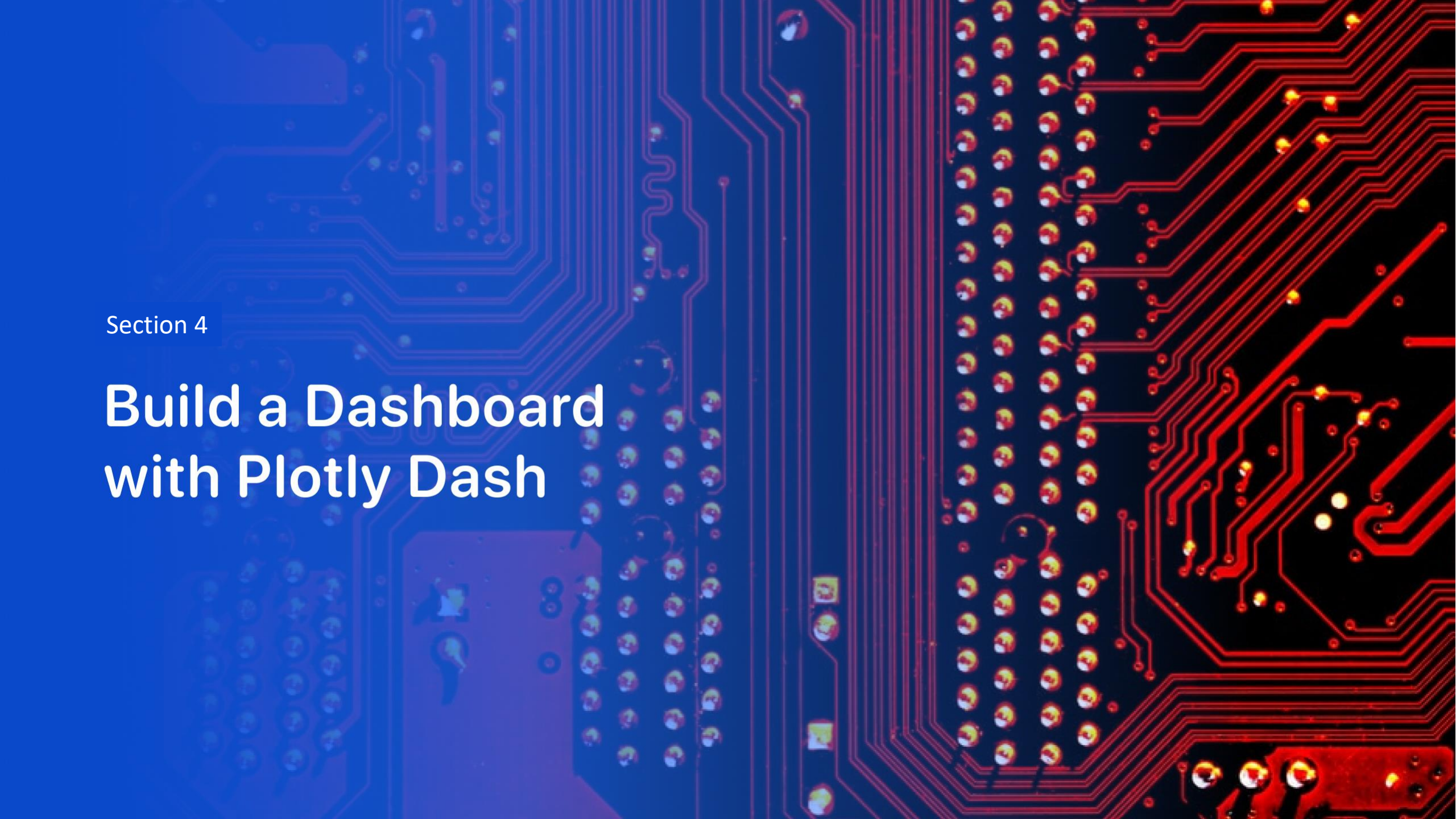


There were more landings on the east coast.

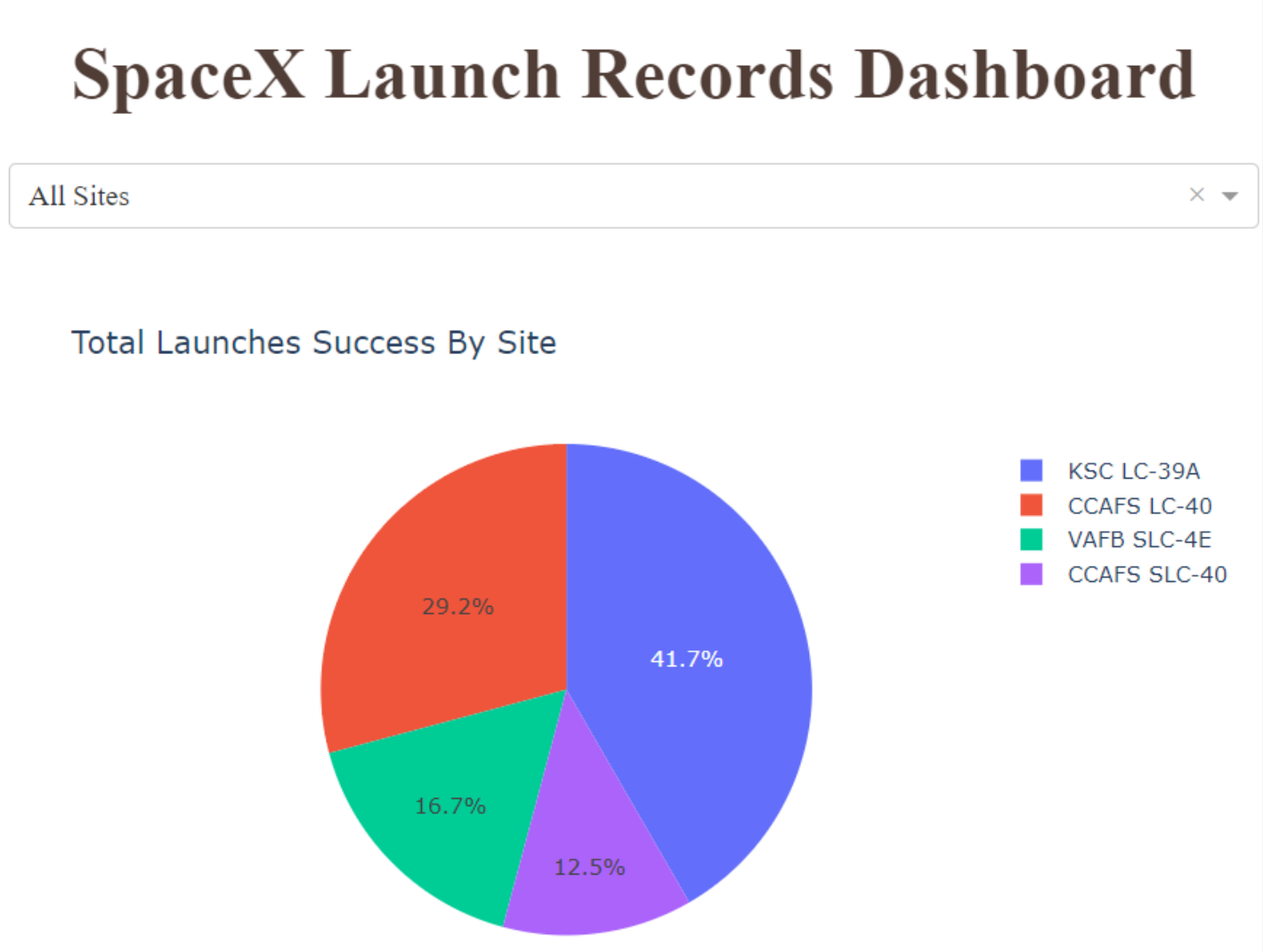
Map with distances to proximities



One of the cities is distanced more than 20 km from the chosen site.

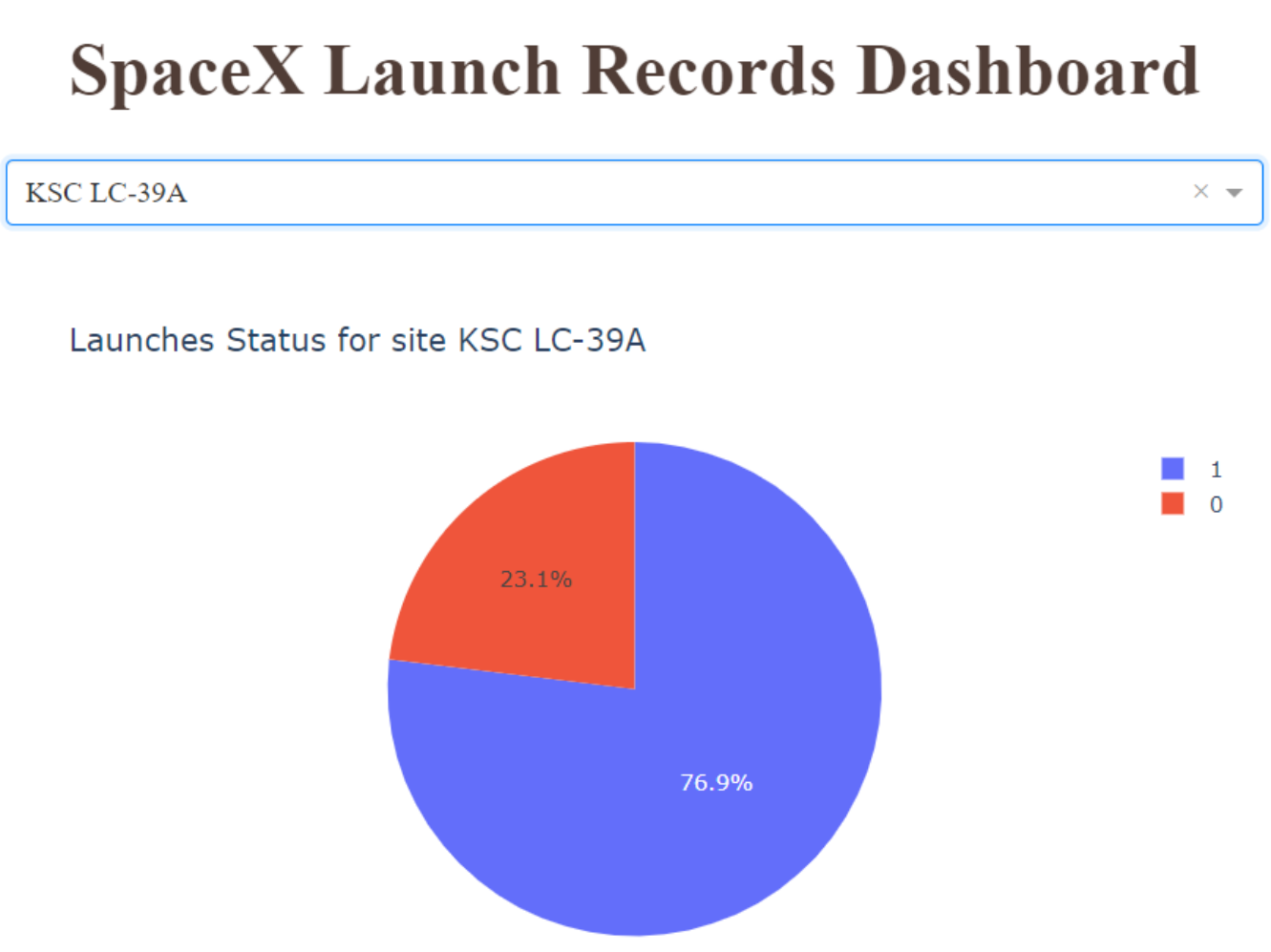

Section 4
Build a Dashboard with Plotly Dash

Total Launches Success by Site



More than 40% of success landing took place in **KSC LC-39A**.

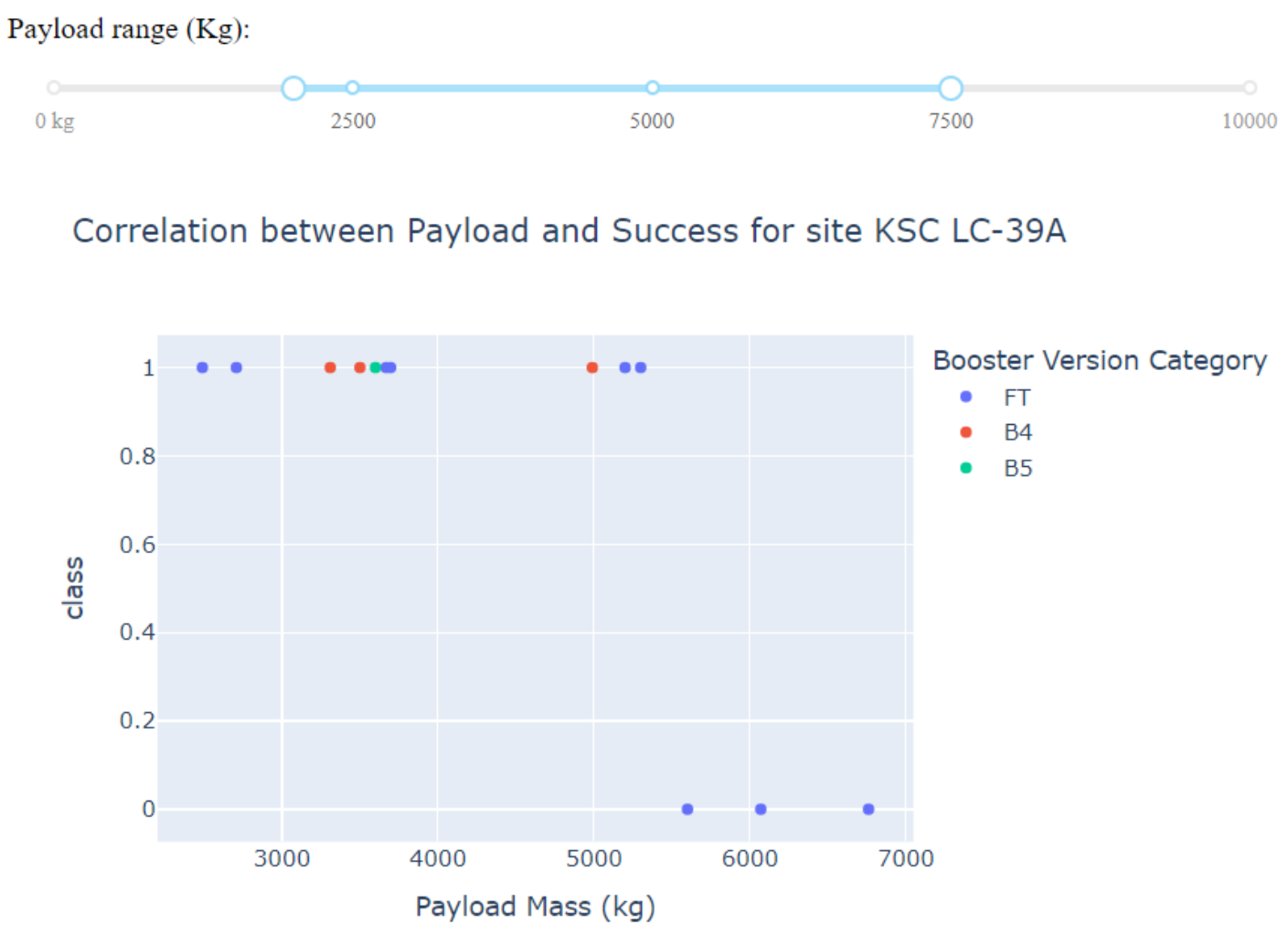
Launches Status for site KSC LC-39A



KSC LC-39A

has also the highest success rate.

Correlation between PayloadMass and Success

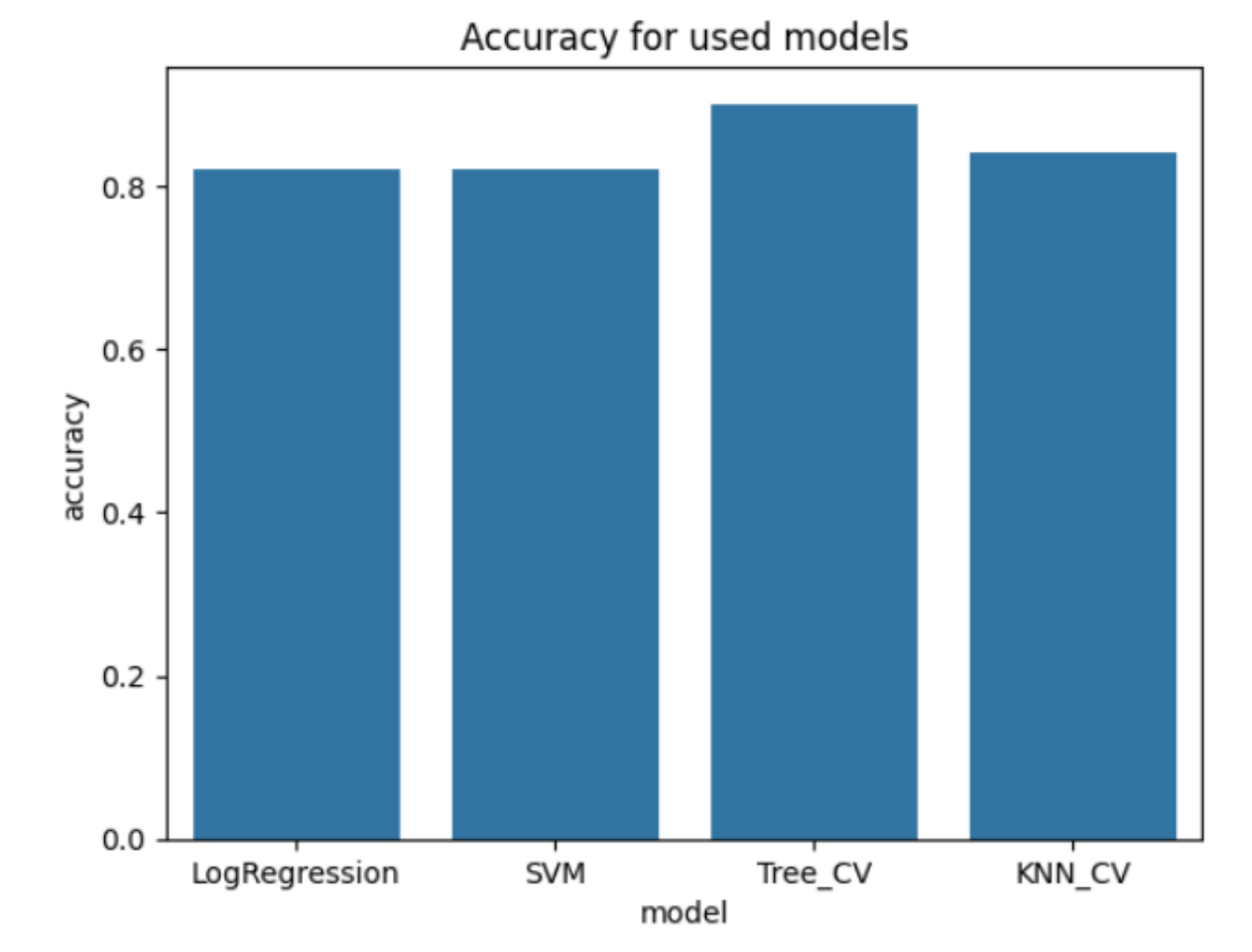


For **KSC LC-39A** payloadMass below 5500 kg give the highest chance for success.

Section 5

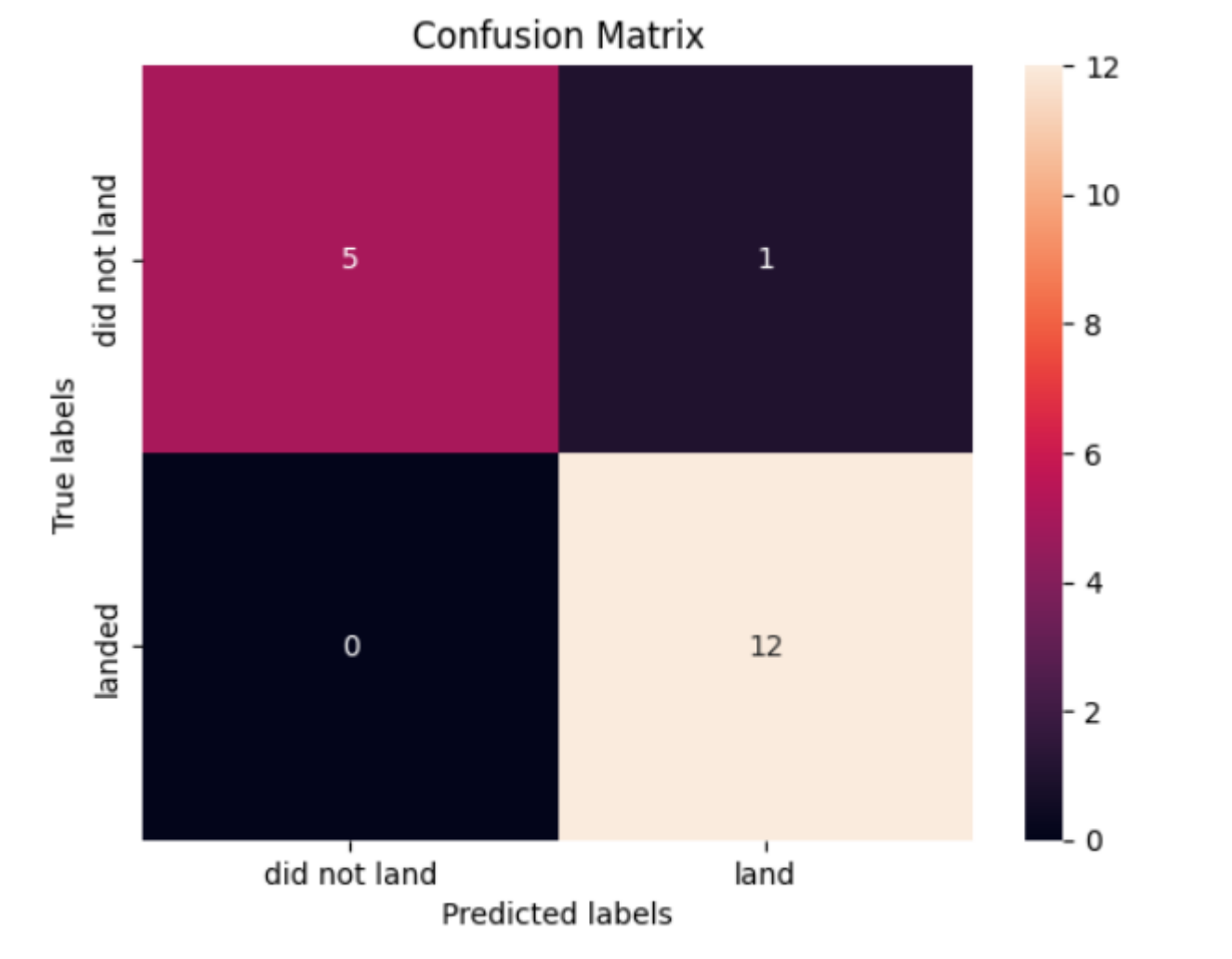
Predictive Analysis (Classification)

Classification Accuracy



The highest classification accuracy equal 0.9 was achieved by Decision Tree Classification.

Confusion Matrix for Decision Tree



Only one case was not predicted correctly by Decision Tree Classifier.

Conclusions

- 4 different models were used to get an appropriate classifier.
- The best method was Decision Tree Classifier with the accuracy of 90%.
- The worst method were Logistic Regression and SVM with accuracy of 82%.
- Using the created Decision Tree Classifier we have a high probability of success landing.

Appendix

Examplary SQL queries used within the project:

- SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
- SELECT * FROM SPACEXTABLE WHERE Launch_Site LIKE 'CCA%' LIMIT 5;
- SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Customer = 'NASA (CRS)';
- SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTABLE WHERE Booster_Version LIKE 'F9 v1.1%';
- SELECT Date FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)' ORDER BY Date LIMIT 1;
- SELECT COUNT(Mission_Outcome) FROM SPACEXTABLE WHERE Mission_Outcome LIKE 'Failure%';

Thank you!