IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Felix Gerber
10.01.2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection
  - Data wrangling
  - Explanatory data analysis with data visualization and SQL
  - Interactive map with Folium
  - Dashboard with Plotly Dash
  - Predictive analysis
- Summary of all results
  - Successful identification of features predicting best success of launching
  - Decision Trees are best choice and result in highest accuracy in MLP

Introduction

- Project background and context
  - SpaceX advertises the Falcon 9 rocket launches with a cost of 62 million dollars – way cheaper than other companies. This is due to the reuse of the first stage of the rocket. It is necessary to determine the successful landing of the first stage to determine launch costs. In this project the reusing of the first stage will be predicted
- Problems you want to find answers
  - Best way to estimate costs for launches
  - Determine best place for launches

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX API & Web Scraping from Wikipedia
- Perform data wrangling
  - Filtering Data, Missing Values were dealt with, One Hot Encoding → Binary classification
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Building, tuning and evaluation of different classification models → choose the one with the best accuracy

Data Collection

- Data sets were collected from SpaceX API
- Web scraping from Wikipedia

→ combination of both to get complete information and perform detailed analysis

Data Collection – SpaceX API

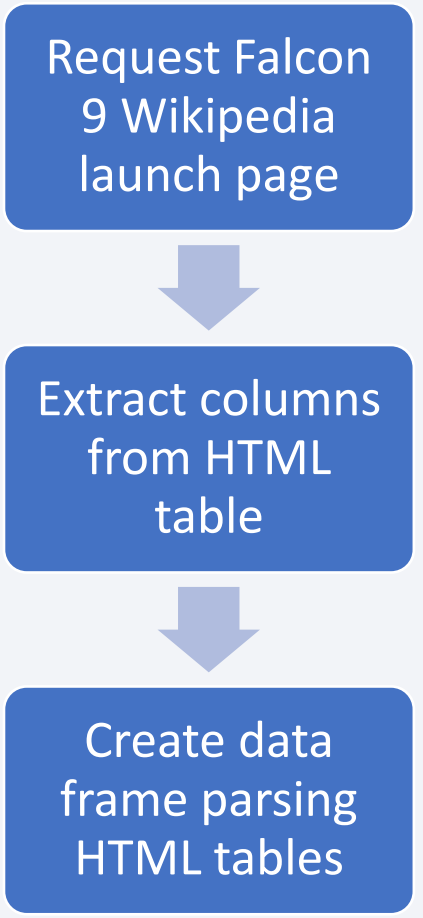
- Data were obtained from public API of SpaceX

- Spacex Data Collection API

Data Collection - Scraping

- Data was obtained from Wikipedia

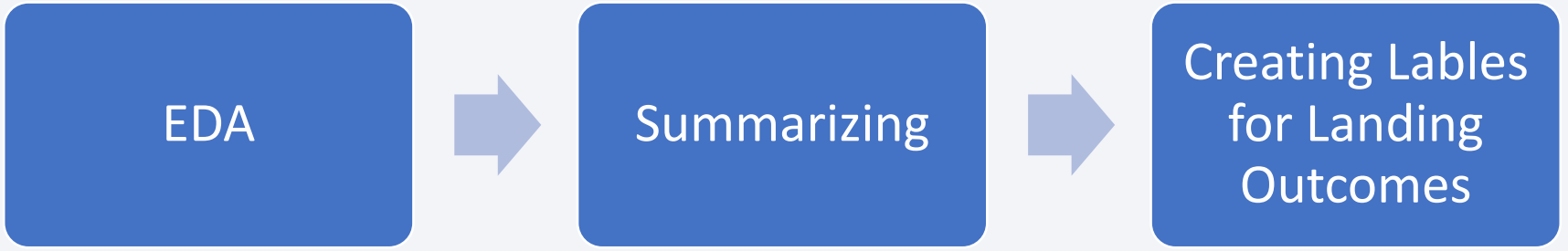
- Webscraping



Data Wrangling

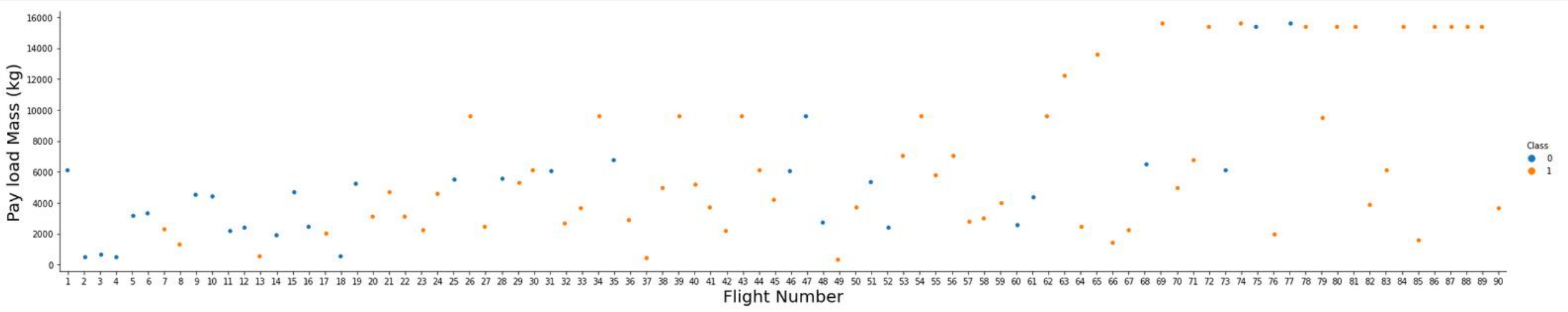
- Exploratory Data Analysis (EDA) was performed
- Summaries were calculated: launch per site, occurrences of each orbit, occurrences of mission outcome per orbit type
- Labeling of Outcome

- Data Wrangling



EDA with Data Visualization

- Scatterplots & Barplots were used to visualize relationship between different feature pairs
  - Payload Mass & Flight Number, Launch Site & Flight Number, Launch Site & Payload Mass, Orbit & Flight Number, Payload & Orbit
- EDA with Data Visualization



EDA with SQL

- Performed SQL queries:
  - Names of unique launch sites
  - Top 5 launch sites where name begins with string ‘CCA’
  - Total payload mass carried by boosters launched by NASA
  - Average payload mass carried by booster version F99 v1.1
  - Date of first successful landing outcome in ground pad
  - Names of boosters which have success in drone ship and payload mass between 4000 and 6000 kg
  - Total number of successful an failed mission outcomes
  - Names of booster version which have carried maximum payload mass
  - Failed landing outcomes in drone ship, booster versions and launch site names for the months in 2015
  - Ranking count of landing outcomes (success or failure) between 4th June 2010 and 20th March 2017 (ordered descending)
- EDA with SQL

Build an Interactive Map with Folium

- Markers, circles, lines and clusters were used with Folium to visualize data in an interactive map
  - Markers: indicate launch sites
  - Circles: indicate highlighted areas around specific coordinates
  - Lines: indicate distances between coordinates
  - Marker cluster: groups events in coordinate (like launches in one launch site)
- Interactive Map with Folium

Build a Dashboard with Plotly Dash

- Dashboard with different graphs and plots to visualize data
  - Pie chart to visualize percentage of launches by site
  - Scatterplot to visualize correlation between payload mass and launch success
- Combination of those graphs and plots helps to analyze relation between payload mass and launch sites and helps to identify the best place to launch according to payloads.
- Dashboard with Plotly Dash

Predictive Analysis (Classification)

- Comparing logistic regression, support vector machine, decision tree and k nearest neighbor algorithms for classification

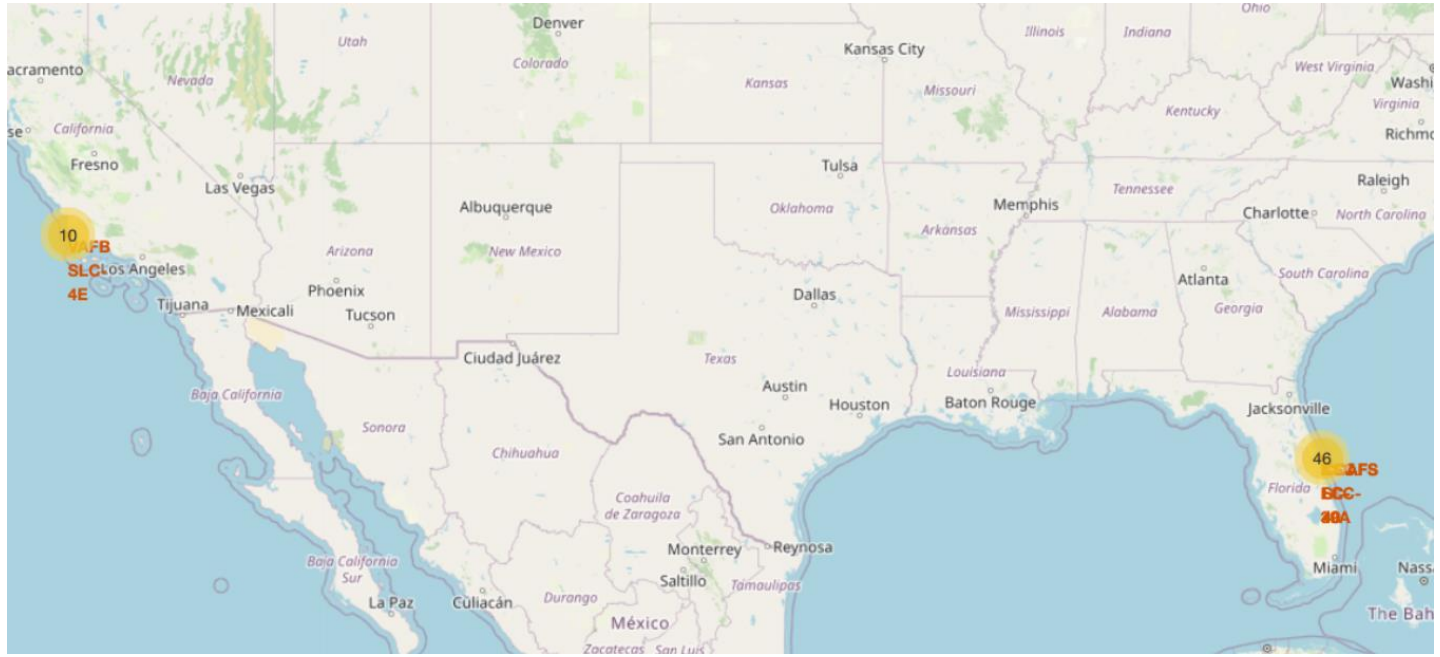
- Predictive Analysis - Classification

Results

- Exploratory data analysis results
  - 4 different launch sites
  - Average payload of F9 v1.1 booster is 2,928 kg
  - First successful landing outcome in 2015 (one year after first launch)
  - Many successful landing of Falcon 9 booster versions with payload mass above average
  - Almost every mission outcome was successful
  - Landing outcomes improved over years
- Interactive analytics demo in screenshots
- Predictive analysis results

Results

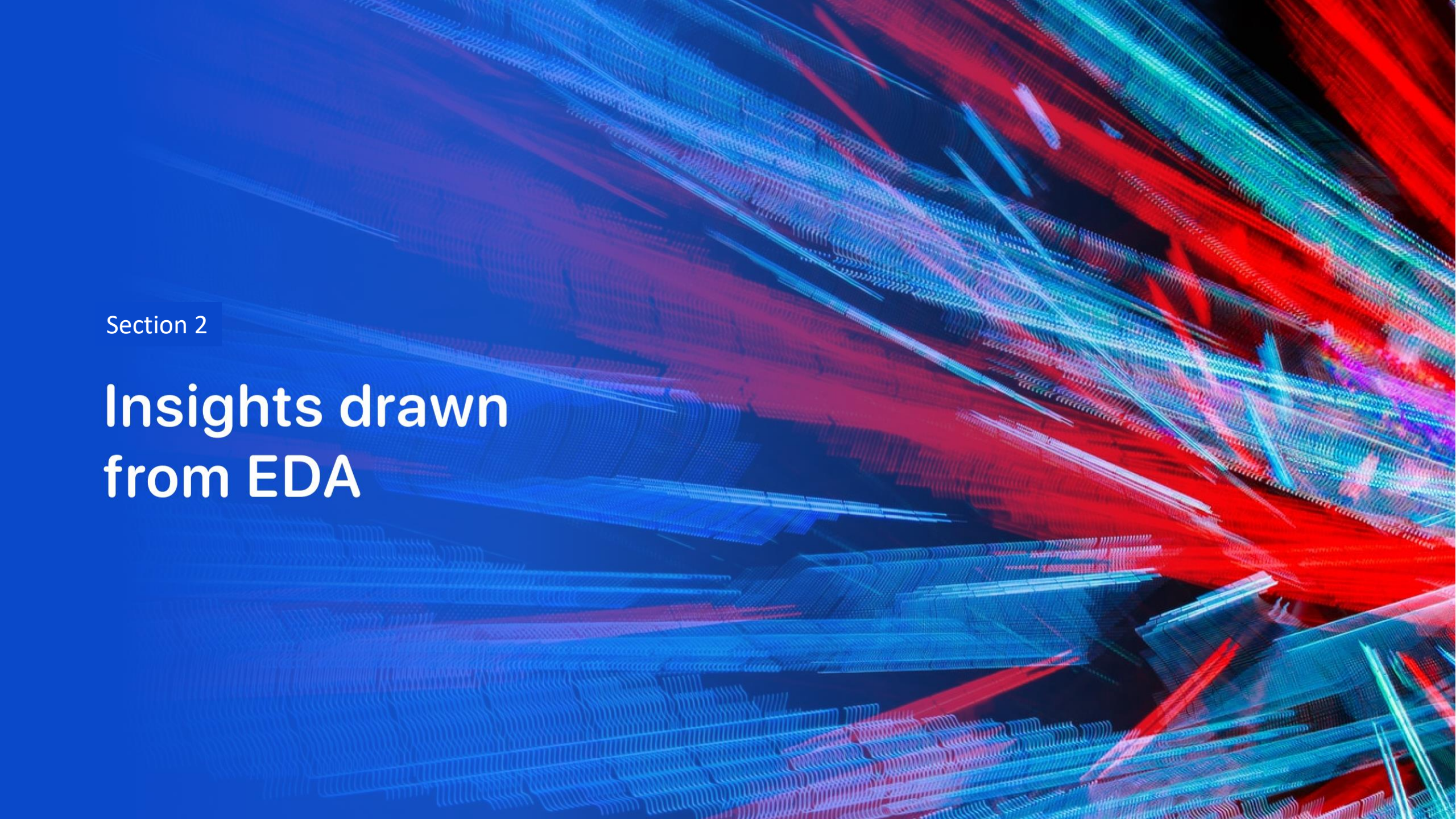
- Interactive analytics demo in screenshots
  - Launch sites usually near sea in safe places with good infrastructure
  - Most launches at east coast



Results

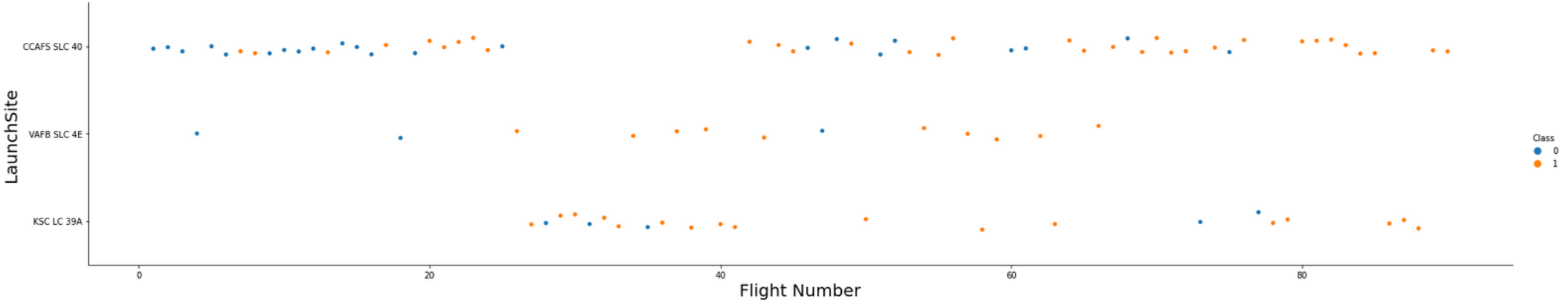
- Predictive analysis results
  - Decision tree was best classifier when it comes to predicting successful landings
  - Accuracy: 90%

	LogReg	SVM	Tree	KNN
F1_Score	0.909091	0.916031	0.929134	0.900763
Accuracy	0.866667	0.877778	0.900000	0.855556


Section 2
Insights drawn from EDA

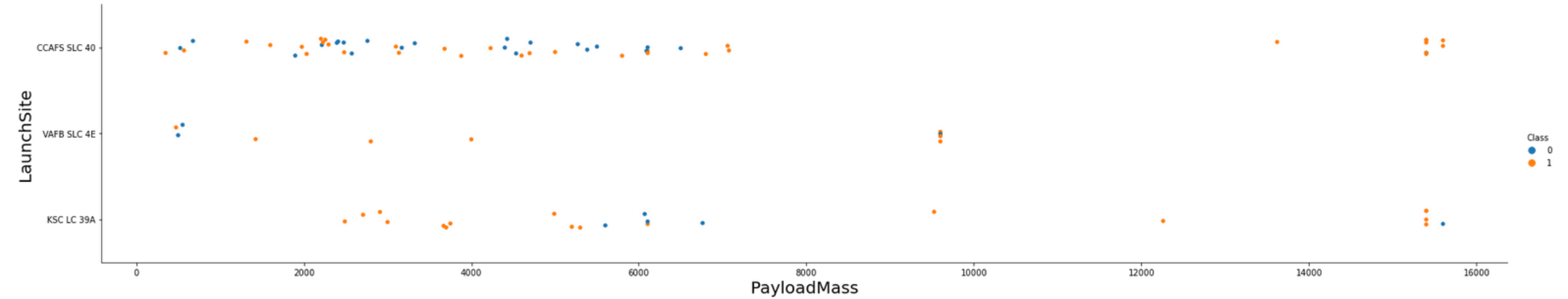
Flight Number vs. Launch Site

- Best launch site is CCAF5 SLC 40 → most launches were successful
- Worst launch site is KSC LC 39A
- General success rate improved over time



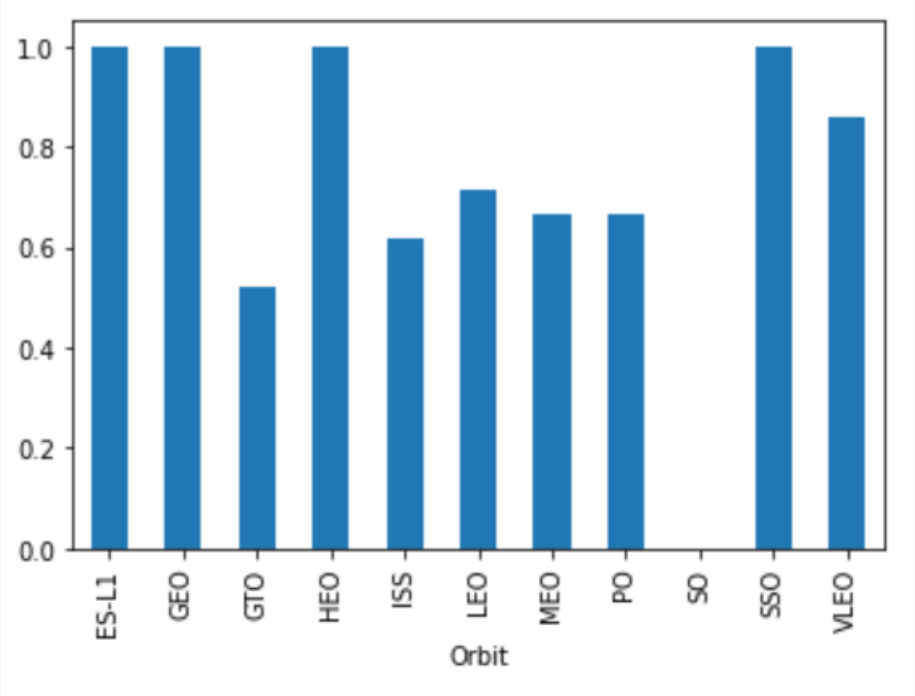
Payload vs. Launch Site

- Payloads over 9000 kg have great success rate
- Payloads over 12000 kg seem to be impossible for VAFB SLC 4E launch site



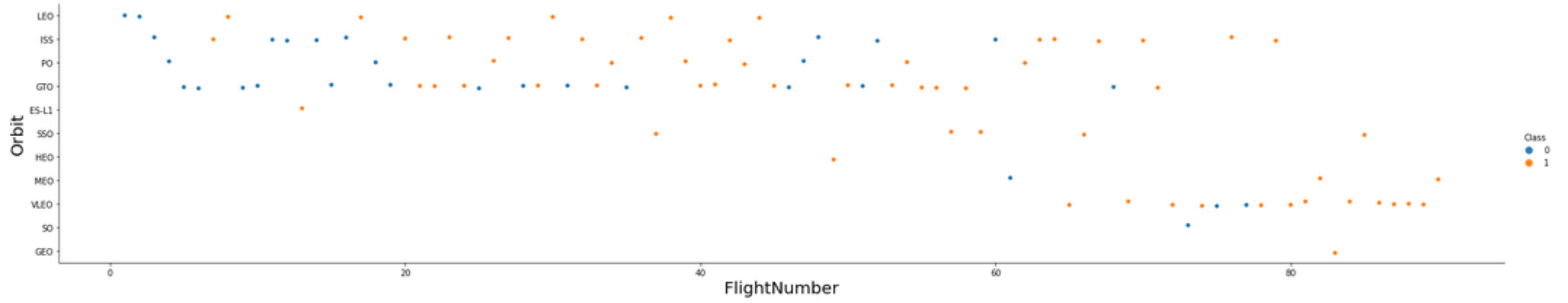
Success Rate vs. Orbit Type

- Best success rates for ES-L1, GEO, HEO and SSO orbits
- SO has worst success rate with 0%



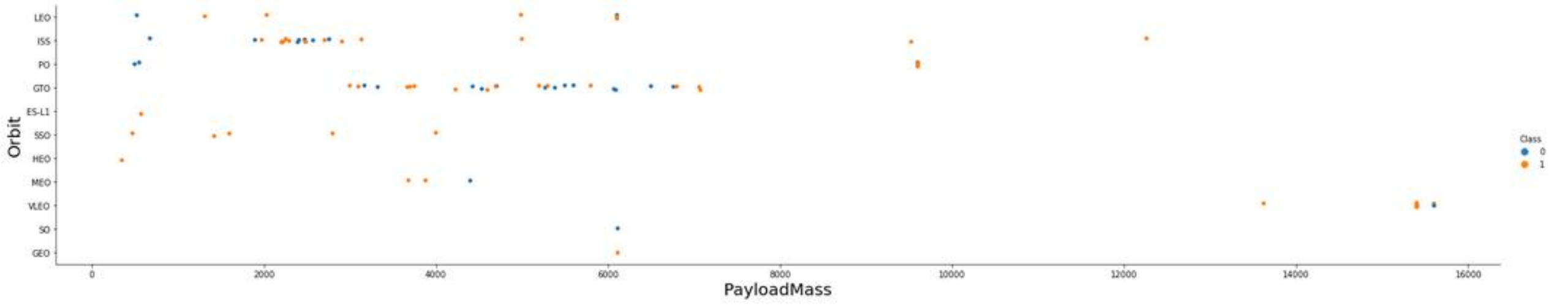
Flight Number vs. Orbit Type

- Success rate improved over time for all orbits
- Frequency of VLEO orbit increases recently



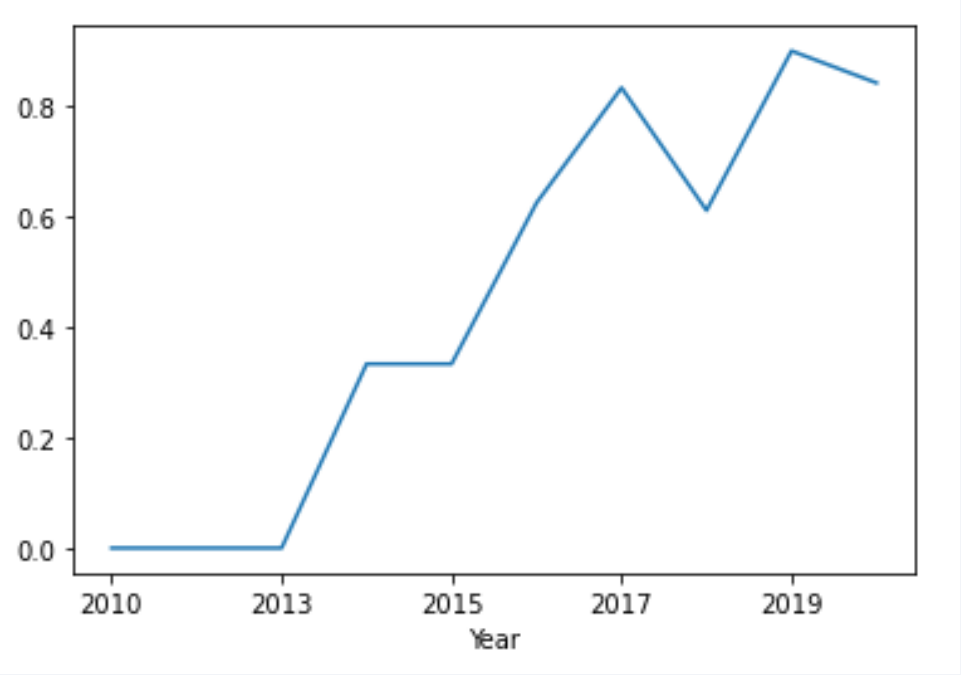
Payload vs. Orbit Type

- Few launches for SO and GEO
- ISS has wide range of payload and good success rate



Launch Success Yearly Trend

- Success rate increased since 2013 and kept increasing until 2020
- Strong increase from 2013 to 2017
- Before 2013 no success



All Launch Site Names

- 4 launch sites were found in data set

Launch_Site
None
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

Launch Site Names Begin with 'CCA'

- 5 records where launch site begins with ‘CCA’
- These are 5 Cape Canaveral launches

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	SpaceX	Success	Failure (parachute)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	NASA (CRS)	Success	No attempt
03-12-2013	22:41:00	F9 v1.1	CCAFS LC-40	SES-8	3170.0	GTO	SES	Success	No attempt

Total Payload Mass

- Total payload calculated by summing all payloads whose codes contain ‘NASA (CRS)’

```
sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)'
```

TOTAL_PAYLOAD
38856.0

Average Payload Mass by F9 v1.1

- Filtering data by booster version and calculating average from this

```
sql SELECT AVG(PAYLOAD_MASS__KG_) AS AVG_PAYLOAD FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1';
```

```
 * sqlite:///my_data1.db
Done.
```

AVG_PAYLOAD
2928.4

First Successful Ground Landing Date

- Filtering Data by successful landings on ground pad
- Determining the minimum date leads to the first successful landing

```
sql SELECT MIN(DATE) AS FIRST_SUCCESS_LANDING FROM SPACEXTBL WHERE Landing_Outcome = 'Success (ground pad)';
```

```
 * sqlite:///my_data1.db
Done.
```

FIRST_SUCCESS_LANDING
01-05-2017

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
- According to these filters four drone ships have landed successfully

```
sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000 AND LANDING_OUTCOME = 'Succ
```

```
 * sqlite:///my_data1.db
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

- Most of the mission outcomes are successful just one failure occurred, where no defined outcome exists for 35 missions

Mission_Outcome	QTY
None	35
Failure (in flight)	1
Success	63
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- Those are the boosters which have carried the maximum payload mass

Booster_Version
F9 FT B1029.1
F9 FT B1036.1
F9 B4 B1041.1
F9 FT B1036.2
F9 B4 B1041.2
F9 B5B1048.1
F9 B5 B1049.2

2015 Launch Records

- In 2015 two landing outcomes in drone ship, boosted versions and launch sites were viewed

MONTH	Booster_Version	Launch_Site
01	F9 v1.1 B1012	CCAFS LC-40
04	F9 v1.1 B1015	CCAFS LC-40

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

- Successful landing outcomes between 2010-06-04 and 2017-03-20, in descending order

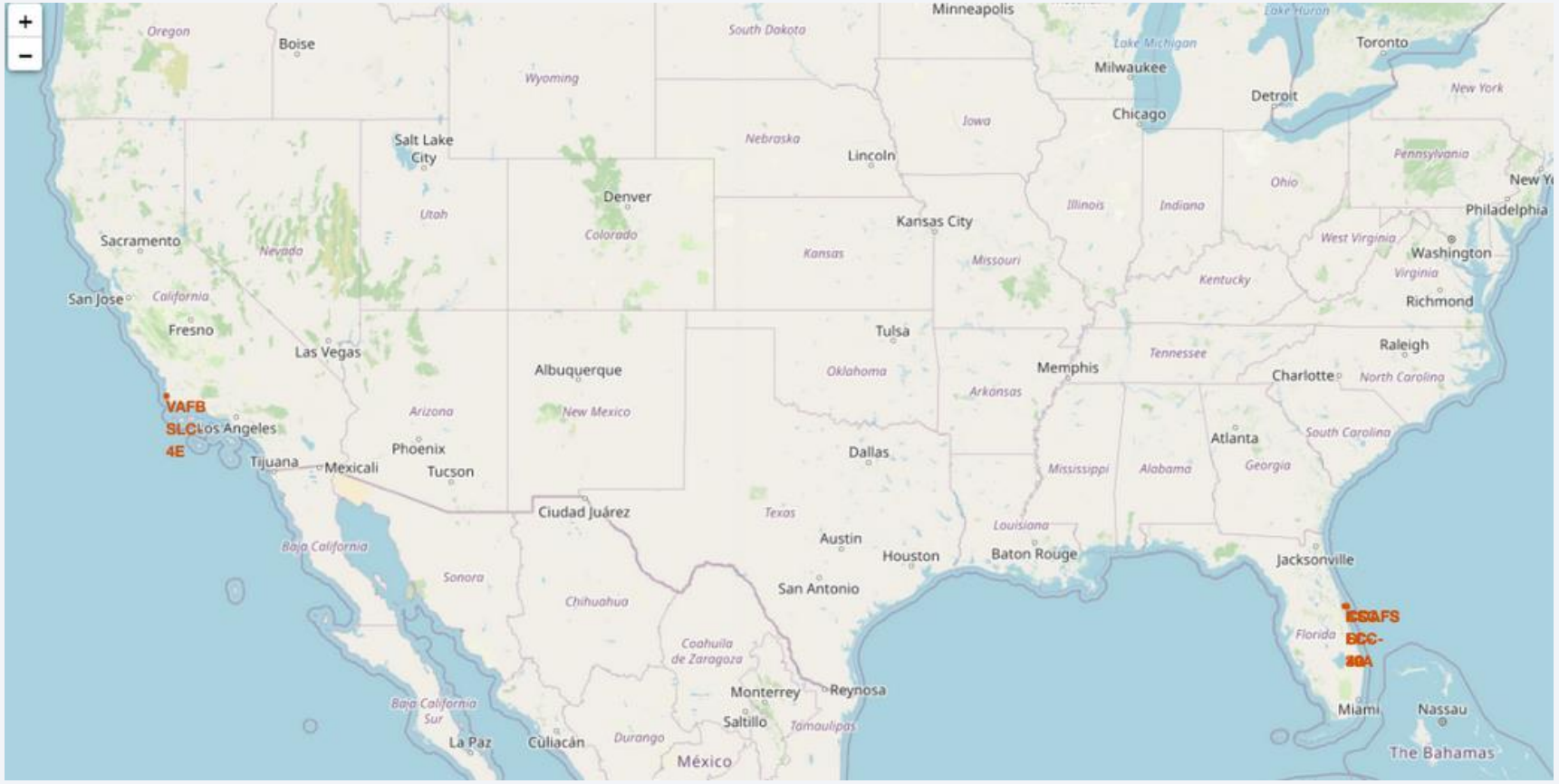
Landing_Outcome	COUNT(LANDING_OUTCOME)
Success (drone ship)	8
Success (ground pad)	6
Success	6

Section 3

Launch Sites Proximities Analysis

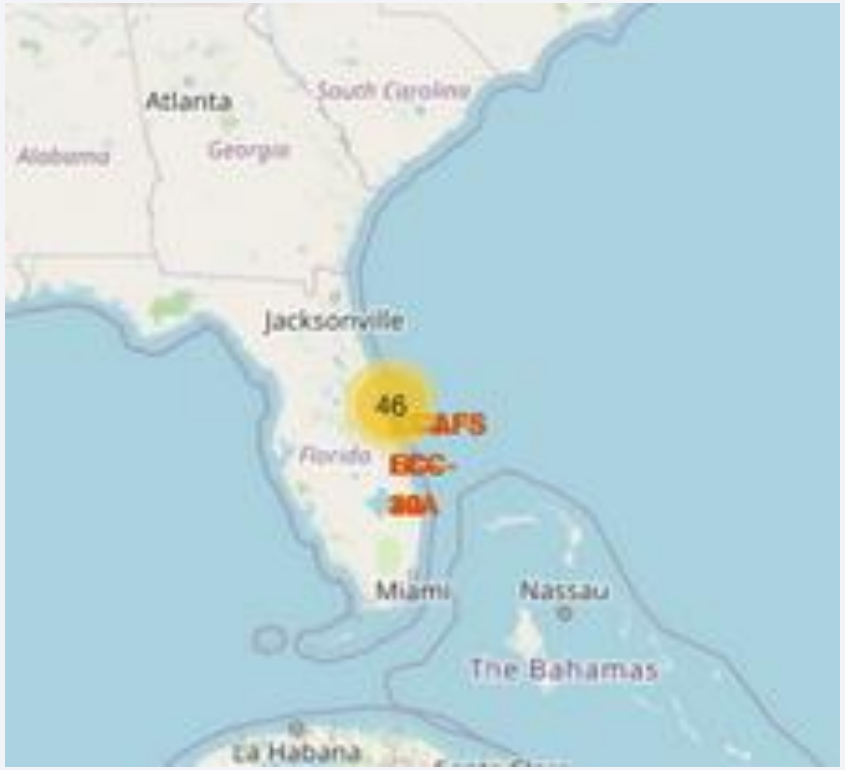
All launch sites

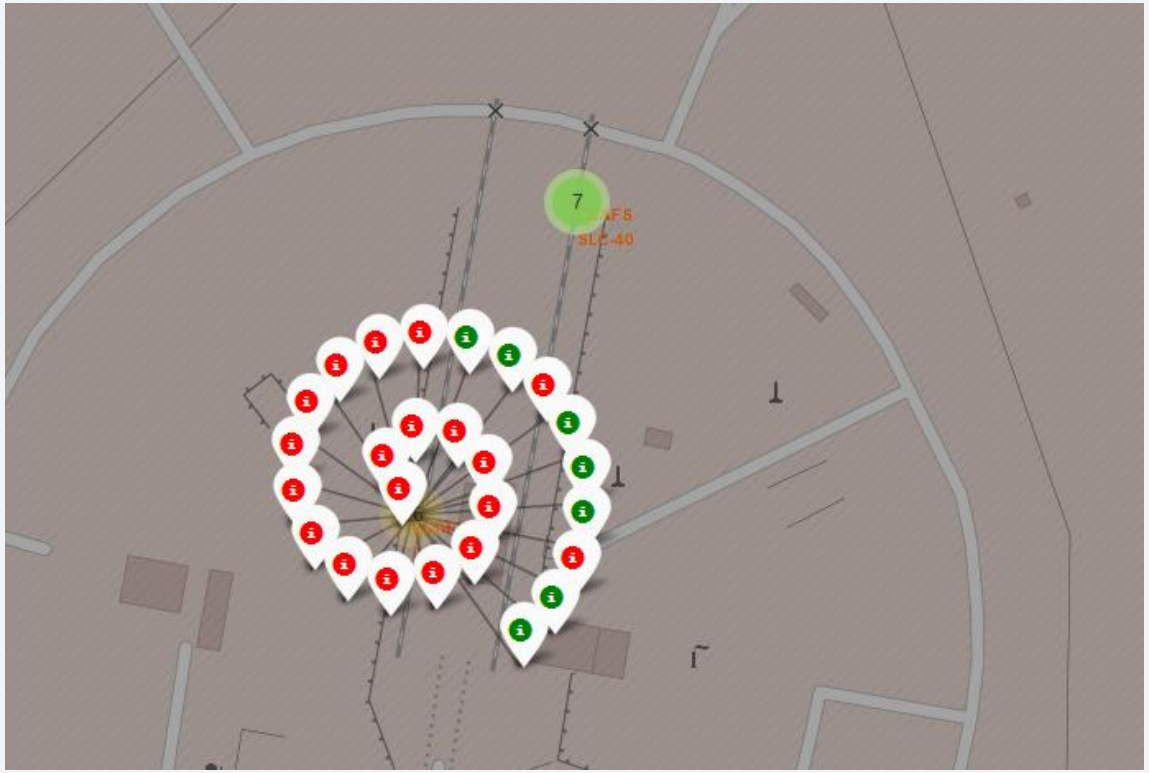
- Launch sites on east and west coast near sea



Launch Outcomes

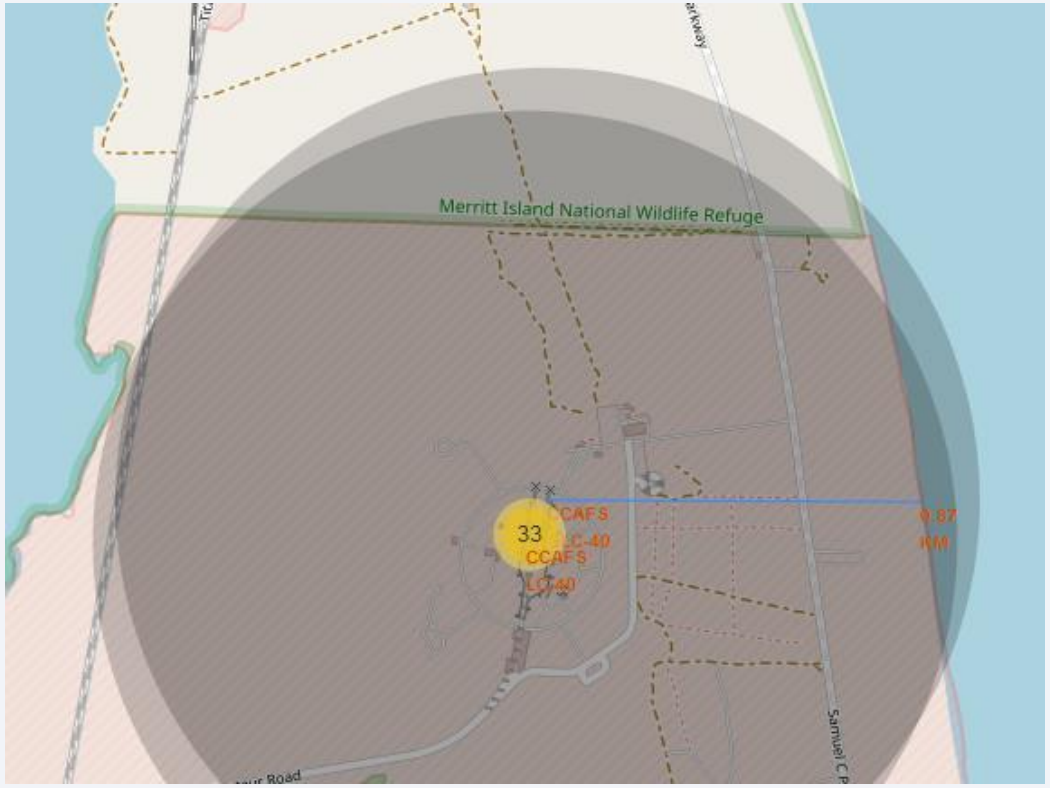
- Green markers = successful / red markers = failed launches

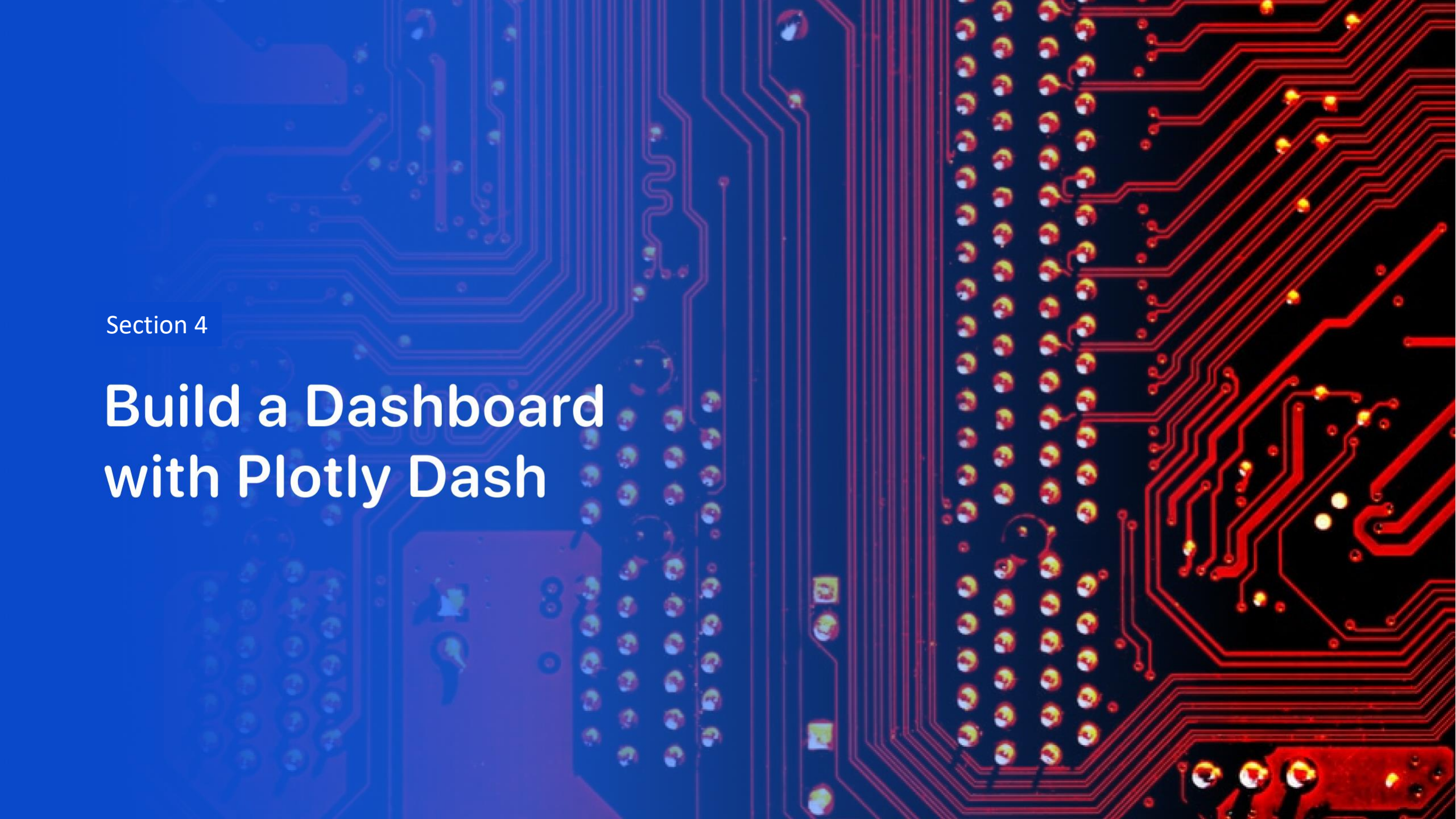




Logistics and Safety of Launch sites

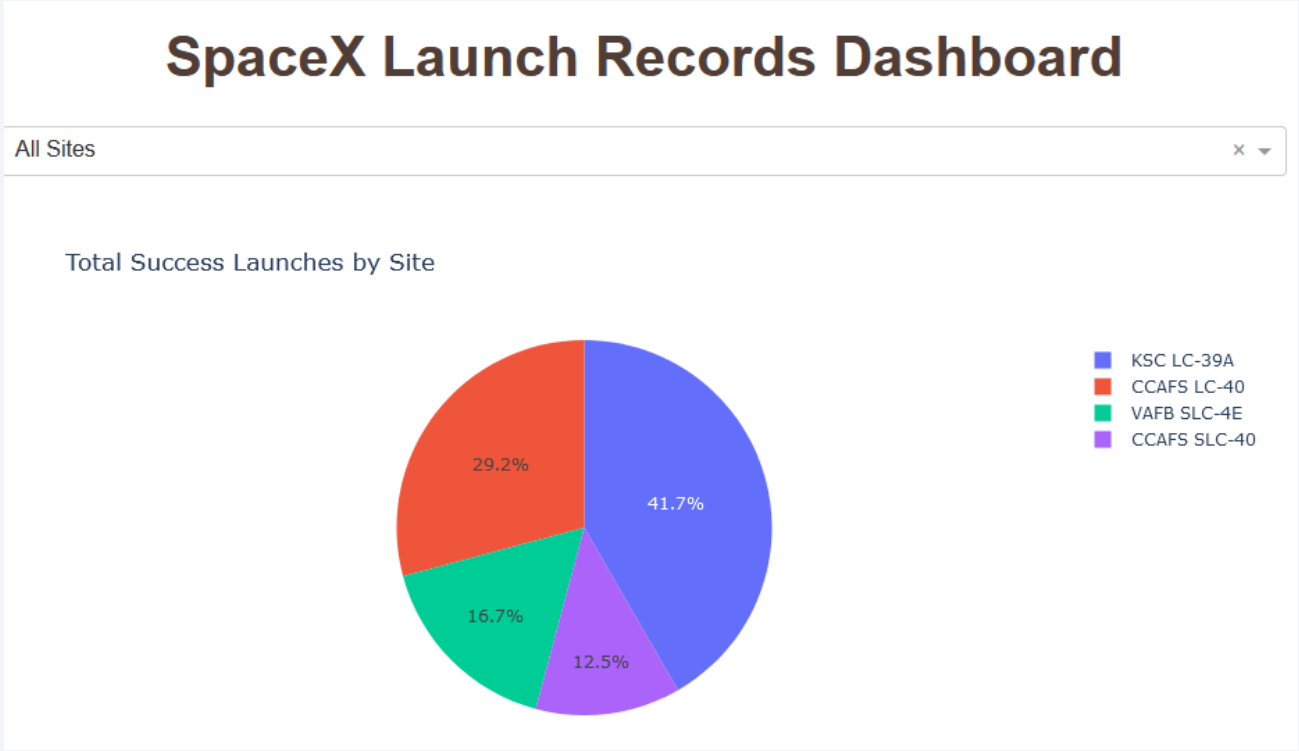
- Good logistic location (near railroads, roads)
- Safe place (no civil inhabited areas nearby and close to coast)




Section 4
Build a Dashboard with Plotly Dash

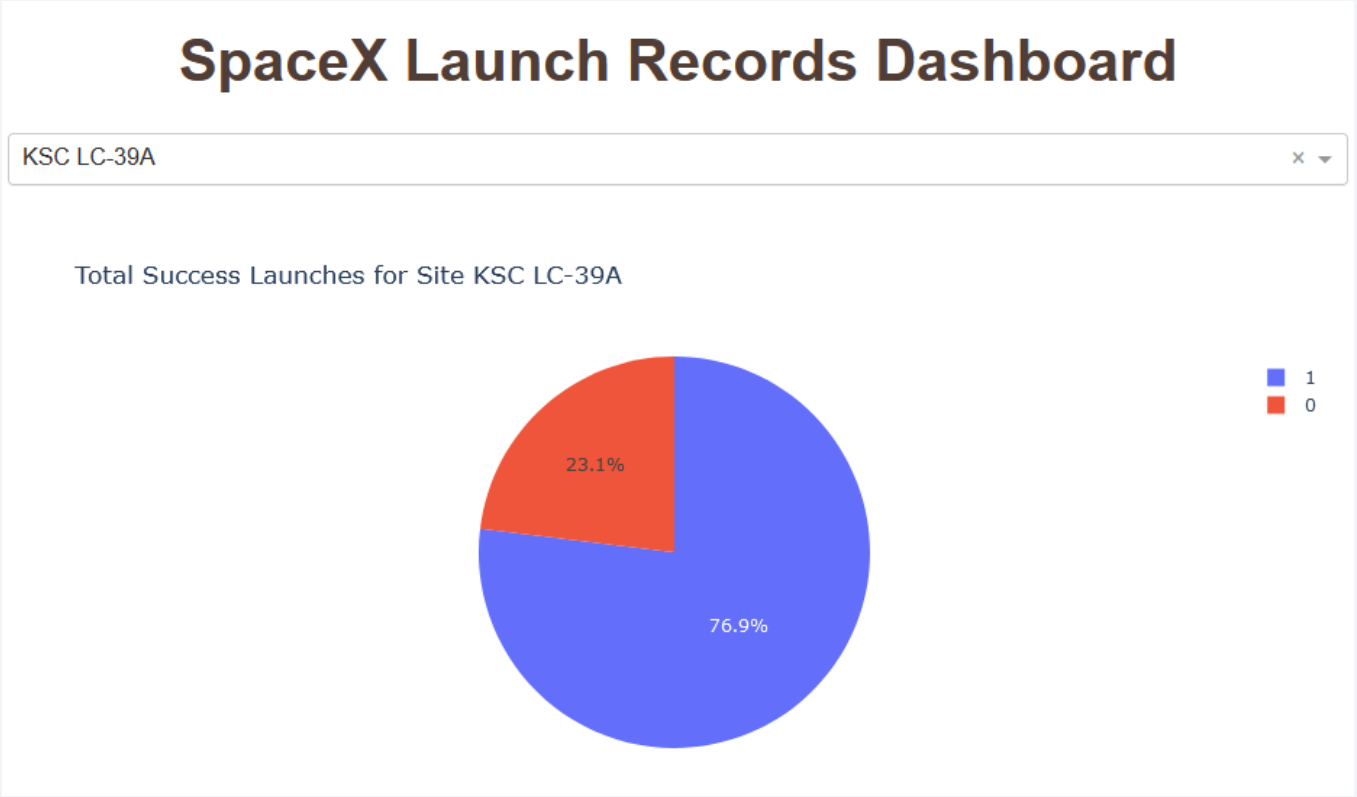
Launch success by site

- Location of launch has huge influence on success and seems to be key factor for successful missions



Launch Success Ratio for KSC LC-39A

- Launch site with highest launch success ratio (76.9 %)



Payload vs. Launch Outcome

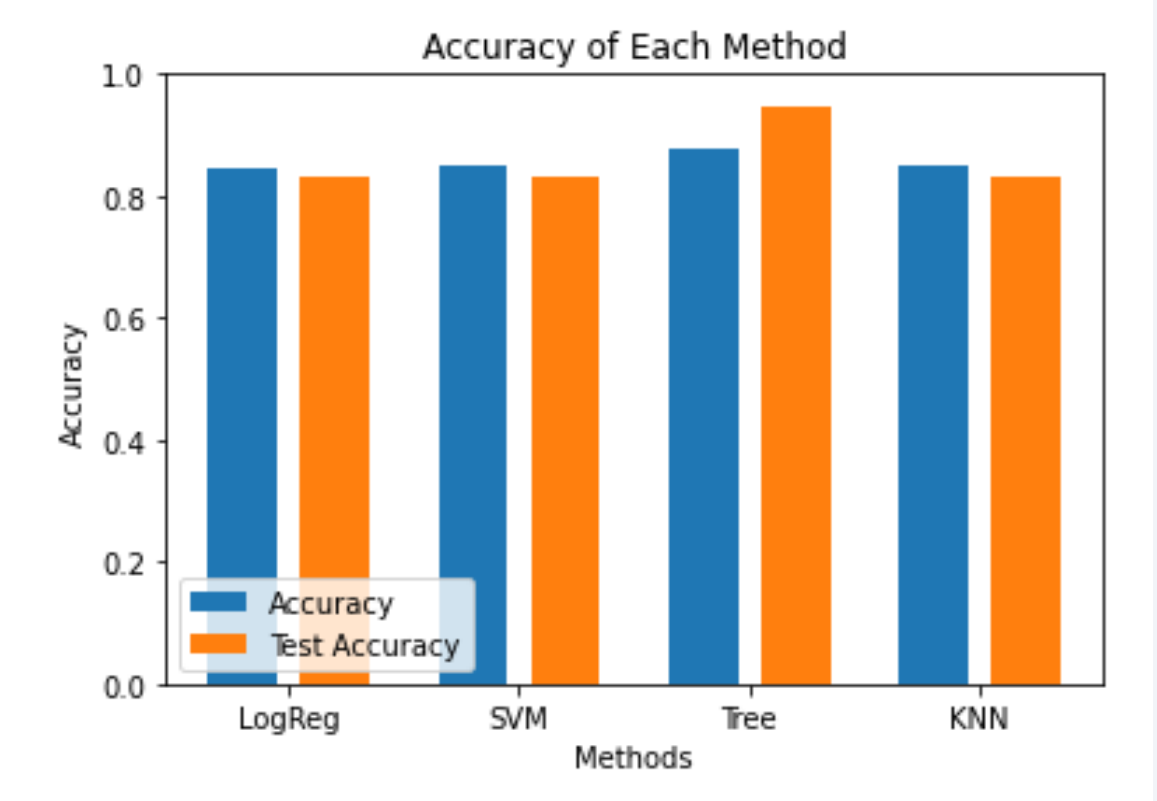
- Payloads under 6,000 kg result in great success
- FT boosters result in great success
- Combination of both is best

Section 5

Predictive Analysis (Classification)

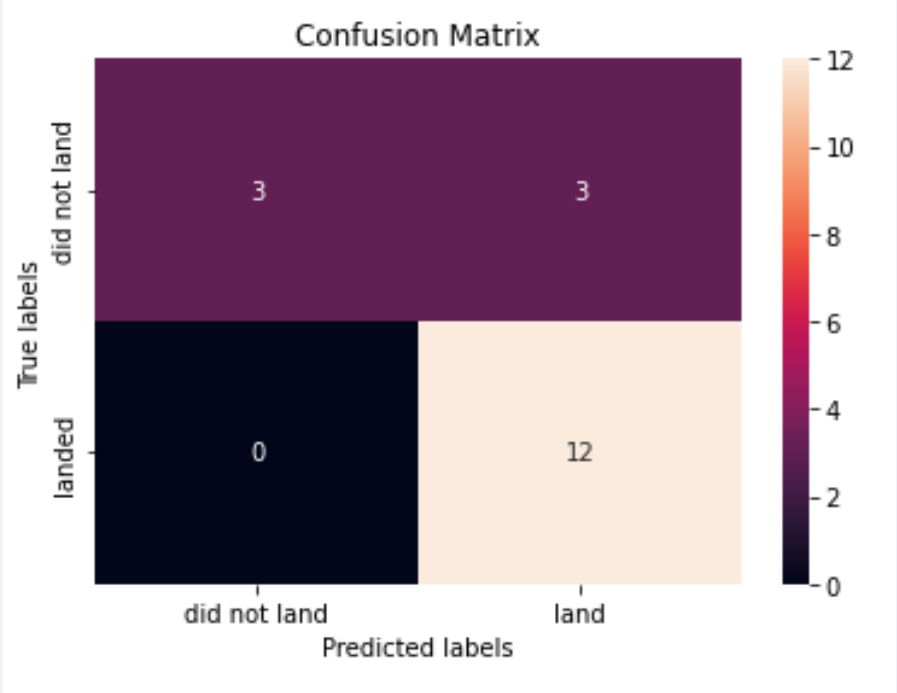
Classification Accuracy

- Comparison of logistic regression, support vector machine, decision tree and k-nearest-neighbors
- Decision tree performs best with accuracy of 87.68%



Confusion Matrix

- Confusion matrix shows that accuracy is high because of high number of true positives and true negatves



Conclusions

- Analysis of different data sources
- Best launch site is KSC LC-39A
- Launch with payload mass above 7,000 kg are best
- Success rate increases over time
- Decision tree is best to use prediction of successful landings

→ In the future launches can be predicted more accurate and in general launches will be more successful which leads to an increase of profits

Appendix

- Github Repository

Thank you!