IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Guilherme Baptista
31/12/2021

Outline

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

Executive Summary

- Summary of methodologies
  - Methodologies for data colletion applied: WebScrapping, SpaceX API
  - Data Analysis: Scatter chart, Bar charts, Maps, Dashboard
  - Machine Learning: Trained and tested data sets with LogReg, SVM, Decision Tree, KNN
- Summary of all results
  - Flight outcome is predictable with 80%+ accuracy
  - Flight outcomes improve with each flight
  - Landing back on land is more reliable than ship pad

Introduction

- Business Scenario

  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

- Problems you want to find answers

  If we can determine if the first stage will land, we can determine the cost of launch. This information can be used if an alternate company wants to bid against Space X for a rocket launch.

  We are going patterns in successful and failed outcomes, as well as attempt to build a prediction model to estimate if first stage will land back successfully.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data was collected leveraging Public Wiki information regarding the SpaceX program
- Perform data wrangling
  - Data was processed using Pandas and Beautiful Soup and classifying landing data as successful/unsuccessful , dealing the missing values, removing irrelevant data columns and visualizing
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

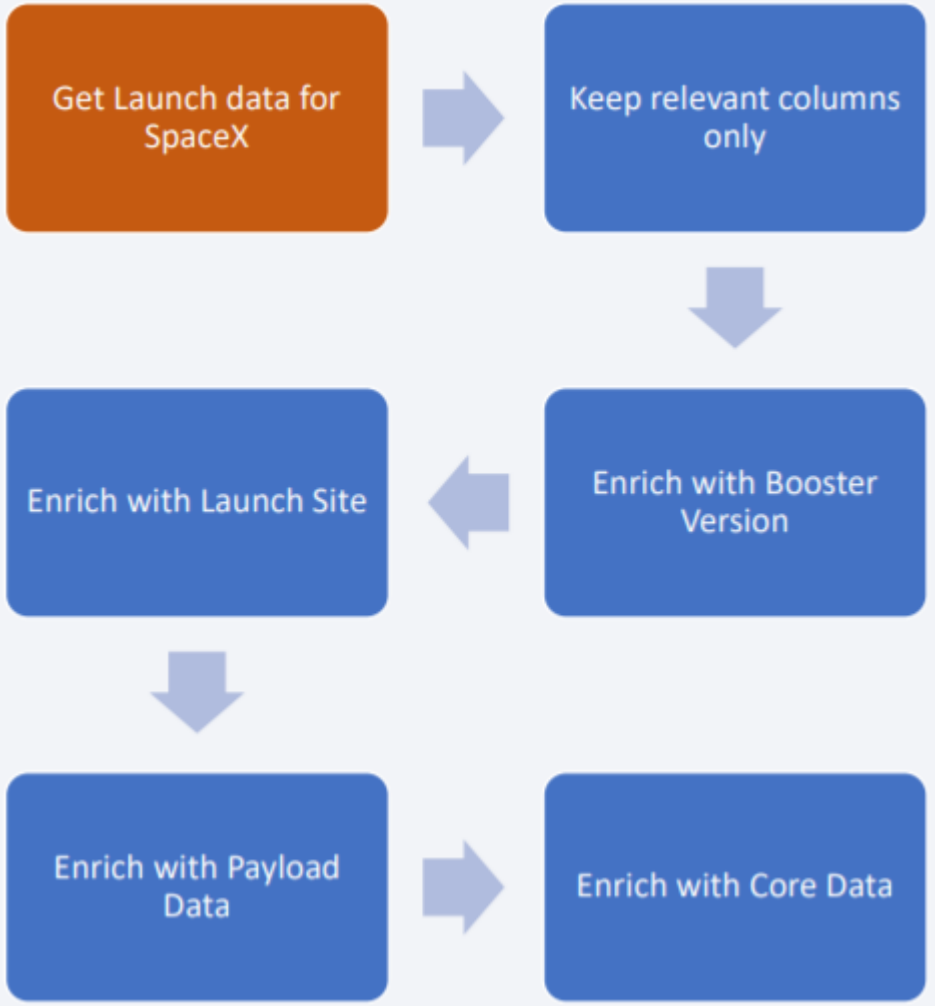
Data Collection

- Data is collected from SpaceX REST API which are about launch and landing specifications, rocket usage.
- Data was further enriched via SpaceX API calls with additional information
  - Rocket
  - Launchpad
  - Payload Mass
  - Orbit

Data Collection – SpaceX API

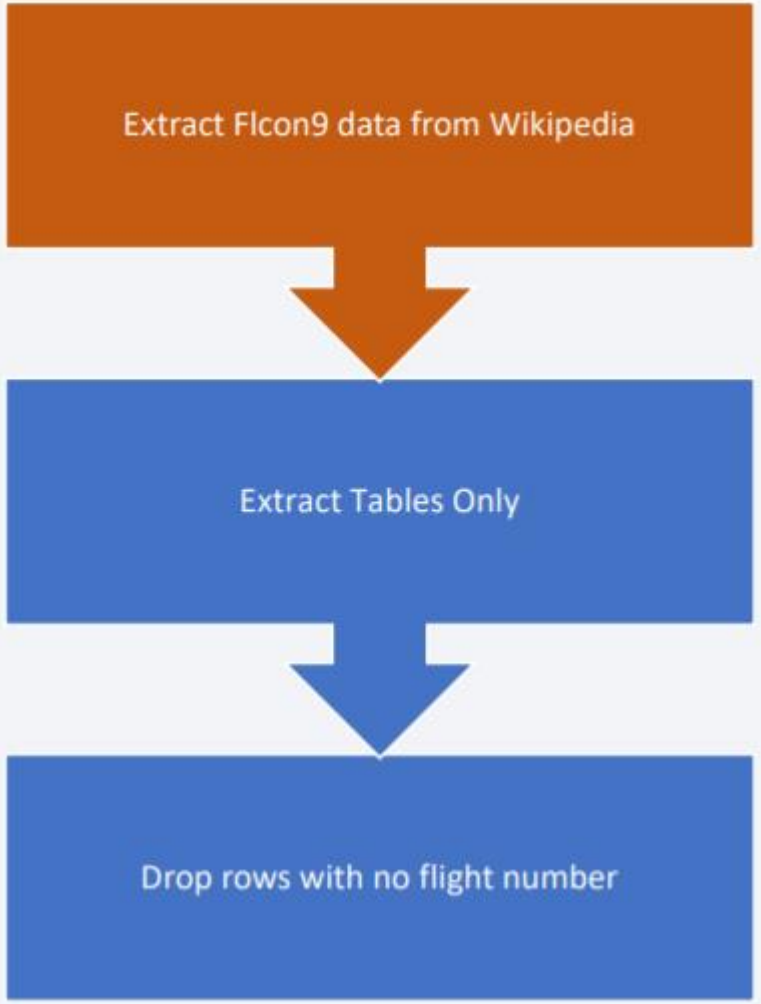
- Starting point: download data via SpaceX API https://api.spacexdata.com/v4/launches/past

- Collecting the data [github]



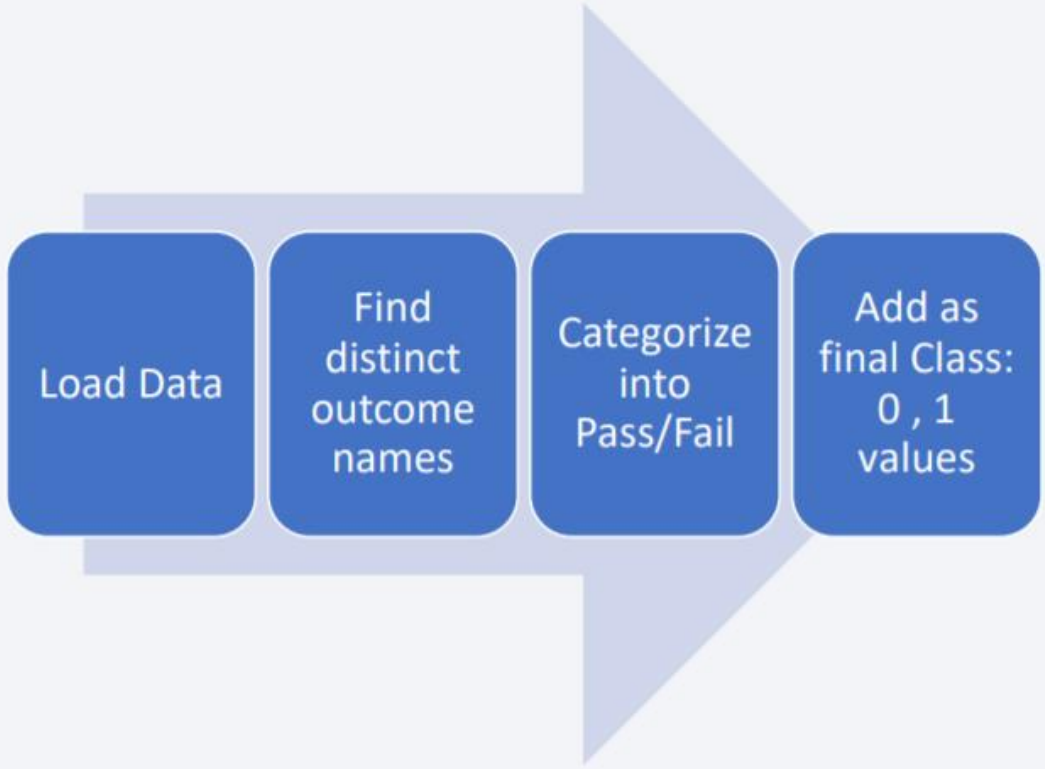
Data Collection - Scraping

- Download public Wikipedia page for Falcon9 launches
- Web Scrapping [github]



Data Wrangling

- Previously acquired data was loaded into a data frame
- Text value for the mission outcome was converted into fixed “Class” category and added as a new column, signifying: (0 – Fail and 1 – Success)
- Data processed using Pandas in 3 steps: Data Calculation of launches / Calculation the number of orbits / Labeling Replace Missing data for critical columns replaced with the average of exist data
- Data Wrangling [github]



EDA with Data Visualization

1) Flight Number and Launch Site (Scatter Chart)

- Plotted scatter chart color coding by mission outcome
- Helps identify better what launch sites were more likely to lead to successful mission
- Helps identify if outcome improved over time

2) Payload and Launch site (Scatter Chart)

- Plotted scatter chart color coding by mission outcome
- Helps identify if specific launch sites were required based on payload weight

3) Orbit vs Success Rate (Bar Chart)

- Plotted % of success full launches per orbit
- Helps identify if specific orbits are more likely to lead to a successful mission outcome

4) EDA with Data Visualization [github]

EDA with Data Visualization

4) Flight Number and Orbit type (Scatter Chart)

- Plotted scatter chart color coding by mission outcome
- Helps identify if the outcome by Orbit improved over time

5) Payload and Orbit type (Scatter Chart)

- Plotted scatter chart color coding by mission outcome
- Helps identify if heavy vs light payloads may benefit from certain orbit

6) Launch success trend (Line Chart)

- Plotted yearly trend of success average
- Helps to identify if the technology was improving over years

- EDA with Data Visualization [github]

EDA with SQL

Find

- Distinct Launch Site names
- Total Payload Mass carried out for NASA
- Average payload for Falcon 9 booster
- How early was first successful to land on Ground pad was achieved
- Successful drone ships booster names for payload 4 ton to 6 ton
- Discover total number of successful and failure outcomes
- Find boosters capable of carrying the maximum payload
- Discover all drone ship outcomes for certain date range and rank in order

EDA With SQL [github]

Build an Interactive Map with Folium

- Created folium map to plot Space X Launch sites
- Circle with label represents a Launch site and Name
- Added color coded icons near the launch site to represent individual launches and outcomes
- Plotted additional line to measure proximity to water, and to highlight the fact that most sites are adjacent to the ocean (likely for safety reasons)

Visual Analytics with Folium [github]

Build a Dashboard with Plotly Dash

- Combo box to select All sites vs Specific Launch Site
    - Show pie chart of mission outcomes for selected Site/All Sites
- Slider to chose maximum payload to consider for charts
    - Show scatter chart to plot outcomes by payload weight and Booster

Plotty Dash [github]

Predictive Analysis (Classification)

1) Load the available data for Flights, Parameters and Outcome

2) Normalize the data values to prepare for machine learning algorithms

3) Split the data into Train vs Test data sets

4) Attempt multiple types of prediction Algorithms : Logistic Regression, Support Vector Machine, Decision Tree, K-Nearest neighbors)

5) For each algorithm run Grid Search CV to fine tune parameters for best outcome

6) Plot Confusion matrix per each tuned model to find best

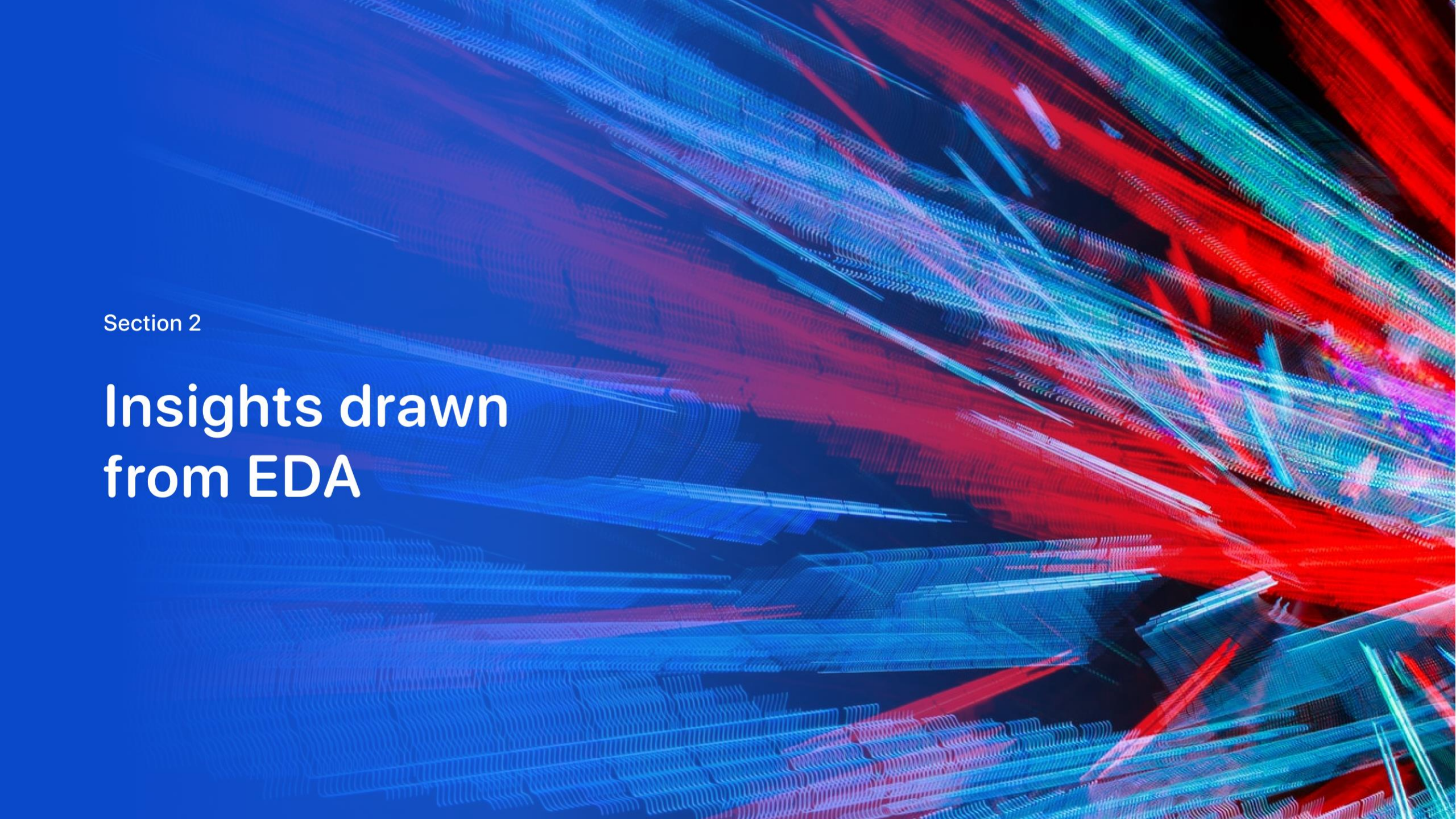
Predictive Analysis [github]

Results

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

Section 2

Insights drawn from EDA

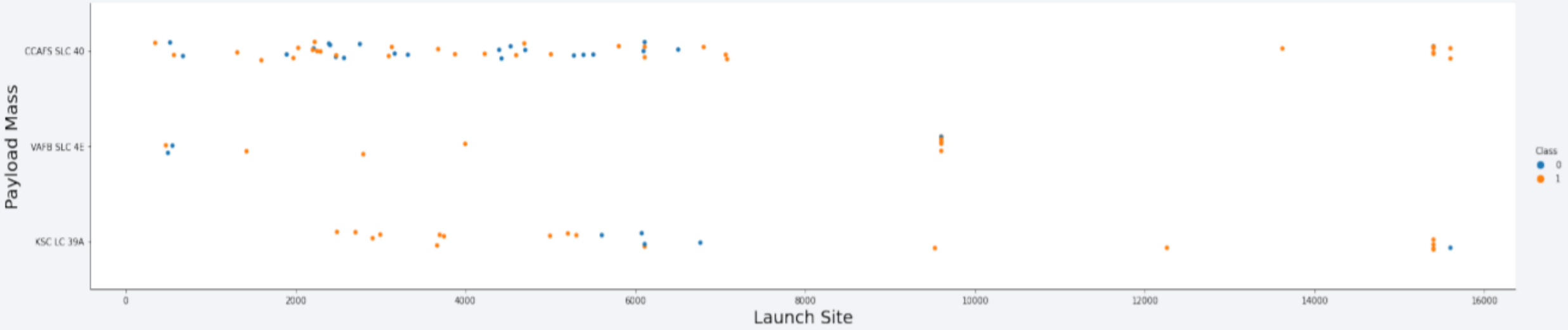


Flight Number vs. Launch Site

- In this scatter plot we can see that the outcome is improving with the number of flight attempts
- This is true for each launch site, and is especially evident over the first 30 flights

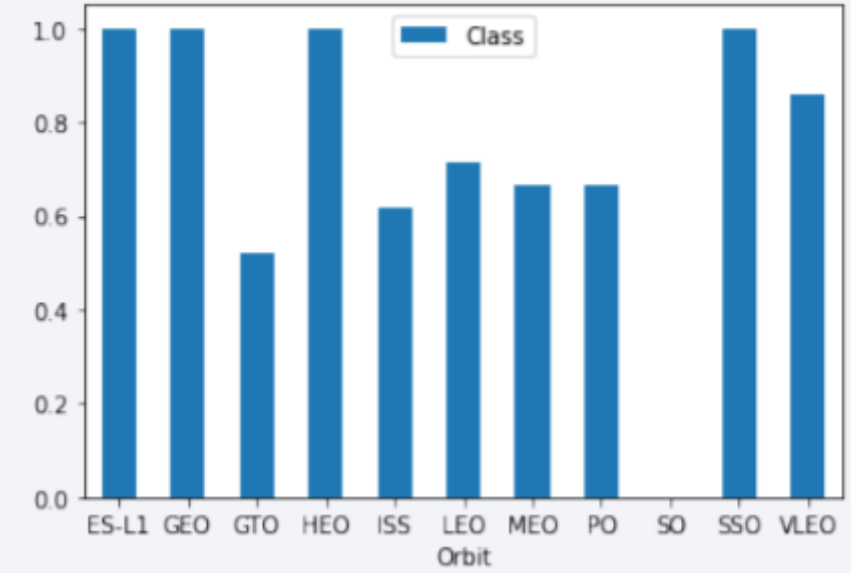
Payload vs. Launch Site

- Here we can see that certain Launch sites had max weight capacity (VAFB-SLC max 10,000)
- However, there is no obvious correlation between weight to success



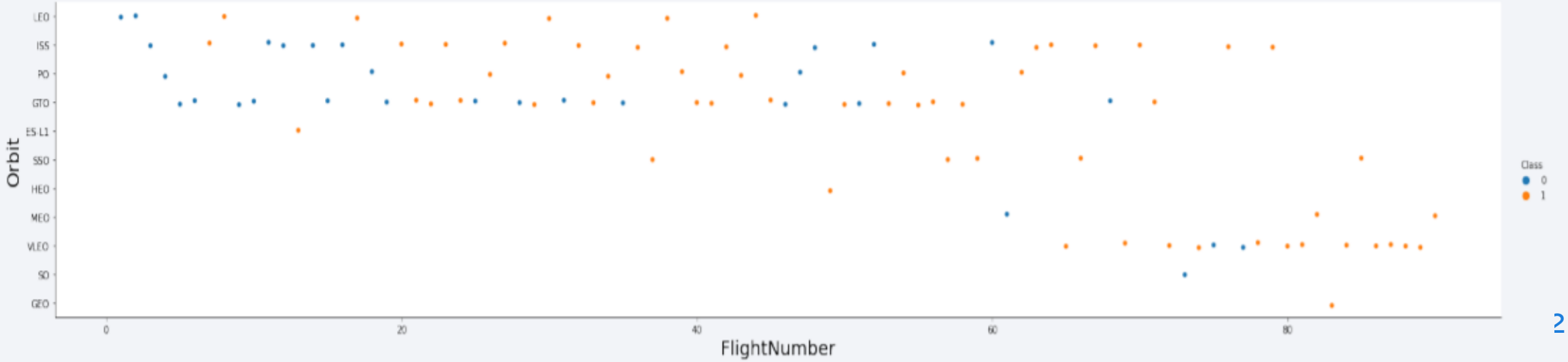
Success Rate vs. Orbit Type

- Certain Orbits have much higher success probability
- Especially the following 4 have clean track record
  - ES-L1
  - GEO
  - HEO
  - SSO



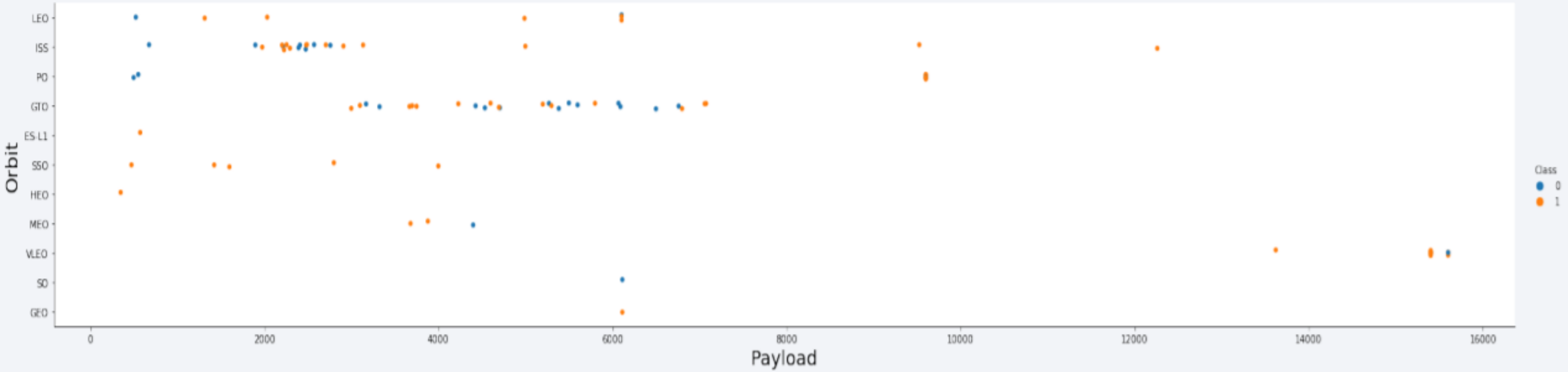
Flight Number vs. Orbit Type

- Looking at this chart we can see if the Orbits with failures have improved success after multiple attempts and learned lessons
- For example, LEO,ISS and GFO – have better outcomes after the initial flights
- We can also see that some of the orbits were introduced only at a later stage and are relatively new



Payload vs. Orbit Type

- This chart helps us identify what orbits are best suited for what payload
- GFO , SSO, LEO are used exclusively for load < 10 ton
- VLEO – exclusively for higher loads > 10 ton



Launch Success Yearly Trend

- This chart shows us the overall yearly trend of successful launches
- As the trend is overall rapidly improving, we can infer that the SpaceX scientists are learning and fine tuning the technology successfully

All Launch Site Names

- From the dataset we can discover unique Launch Site names

```
: %sql SELECT DISTINCT LAUNCH_SITE  FROM SPACEX
```

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

Launch Site Names Begin with 'CCA'

- We can identify the first outcomes for all CCA launch sites

```
%sql SELECT * from SPACEX where LAUNCH_SITE LIKE 'CCA%' LIMIT 5
```

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

Total Payload Mass

- Total payload carried by boosters from NASA

```
%sql SELECT SUM(payload_mass__kg_) FROM SPACEX WHERE  customer like 'NASA%'
```

99980

Average Payload Mass by F9 v1.1

```
%sql SELECT AVG(payload_mass__kg_) FROM SPACEX WHERE  booster_version = 'F9 v1.1'
```

2928

First Successful Ground Landing Date

- From here we can see that landing back to a Ground pad is a mature technology with first successful outcome dating back to 2015

```
%sql SELECT MIN(DATE) FROM SPACEX where LANDING__OUTCOME = 'Success (ground pad)'
```

2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

```
%sql SELECT DISTINCT(BOOSTER_VERSION) FROM SPACEX where PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000 and LANDING__OUTCOME = 'Success (drone ship)'
```

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

Total Number of Successful and Failure Mission Outcomes

- From this total count by outcome, we can see, that vast majority of the missions has resulted in successful outcom

1	landing__outcome
3	Failure
5	Failure (drone ship)
2	Failure (parachute)
38	Success
14	Success (drone ship)
9	Success (ground pad)

Boosters Carried Maximum Payload

- From this table we can see maximum Payload carried out at 15600 kg
- Multiple Boosters are able of carrying out such payload

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1048.5	15600
F9 B5 B1049.4	15600
F9 B5 B1049.5	15600
F9 B5 B1049.7	15600
F9 B5 B1051.3	15600
F9 B5 B1051.4	15600
F9 B5 B1051.6	15600
F9 B5 B1056.4	15600
F9 B5 B1058.3	15600
F9 B5 B1060.2	15600
F9 B5 B1060.3	15600

2015 Launch Records

- Here we can see that only 2 failed drone ships happened in 2015 and all of them were from CCAFS LC-40 launch site

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

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

- From this outcome table we can see, that most complex/unpredictable landings back o a drone ship with a 50% success ratio

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

Section 4

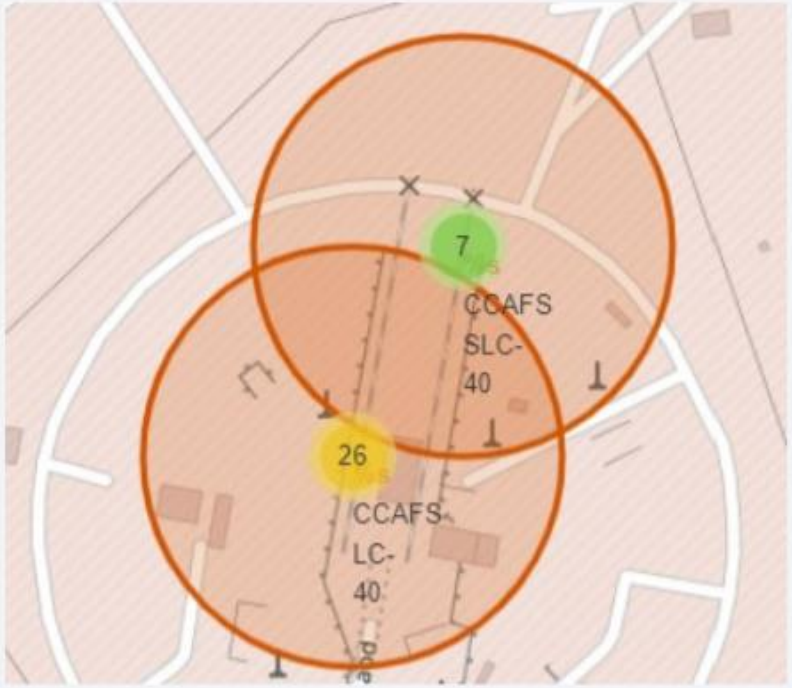
Launch Sites Proximities Analysis

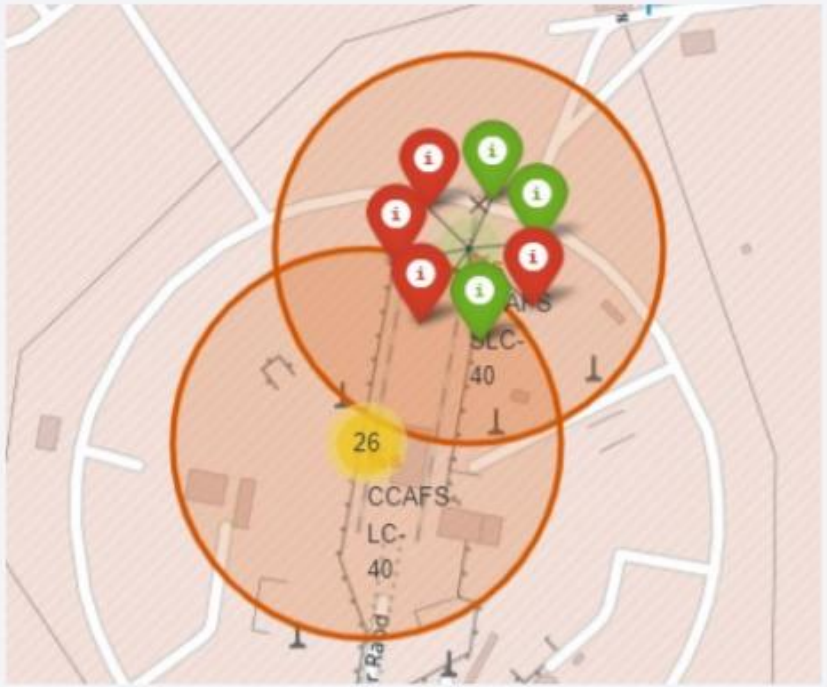
Mark all launchs sites on a map

• Mapping out all sites on the map we can see they are concentrated on the East and West coast of United States, closer to the equator.

Success/failed launches for each site

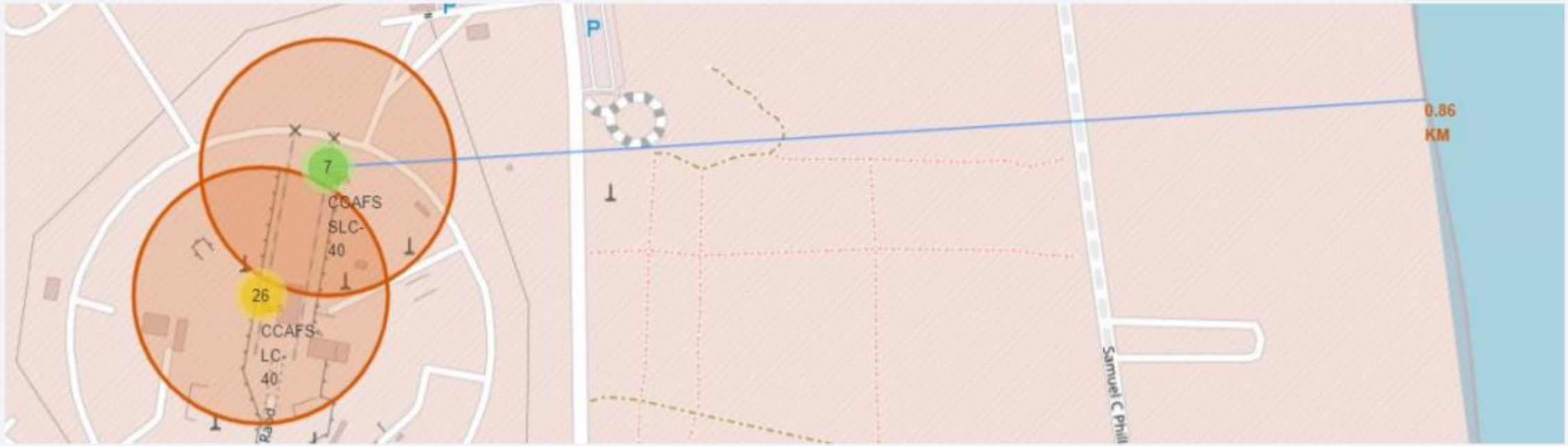
- On this representation, its easier to see ratio of successful/failed launches on each site





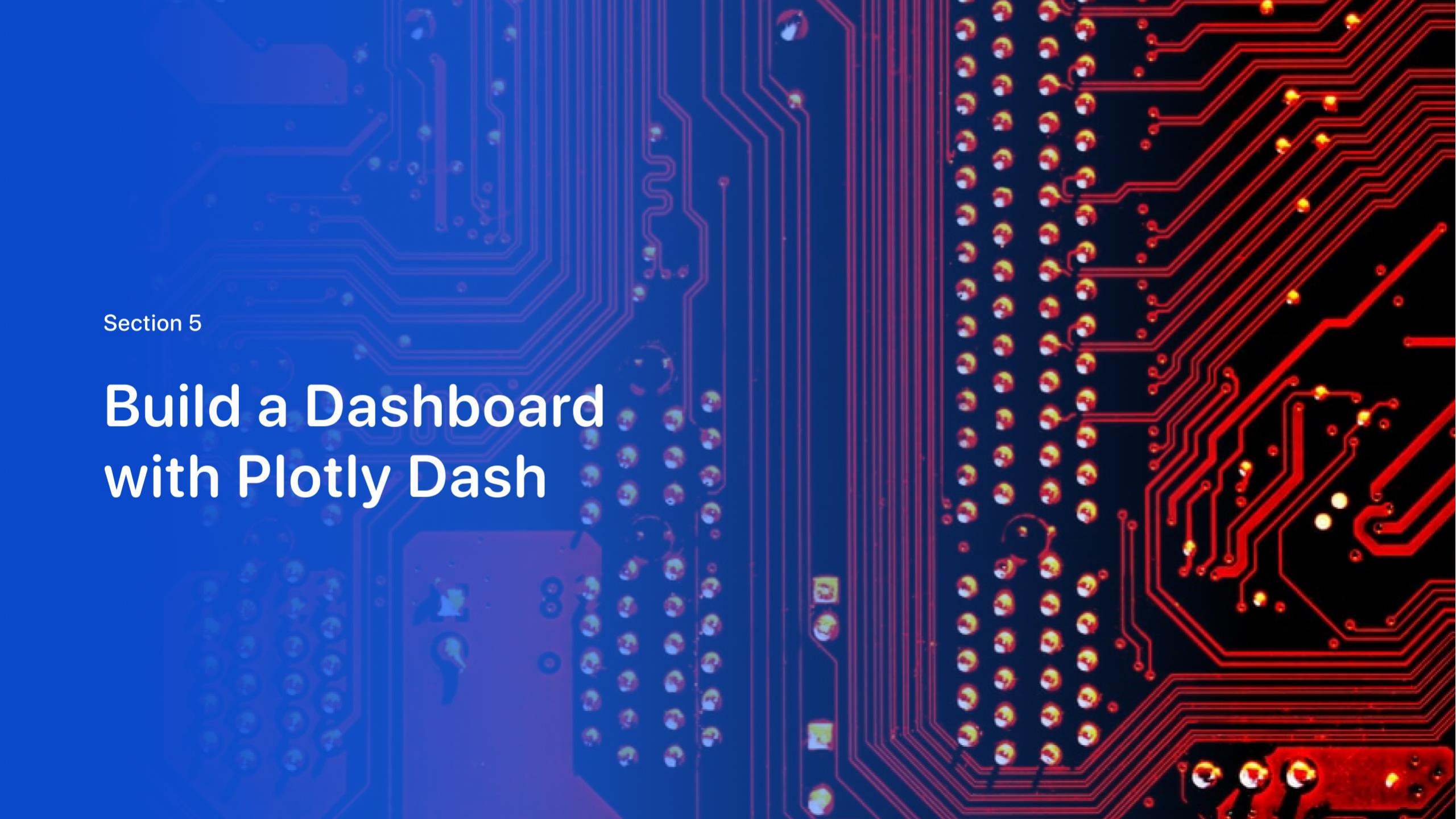
Proximity of launch sites

• In this image we can see that the sites are located in close proximity to the ocean < 1 km



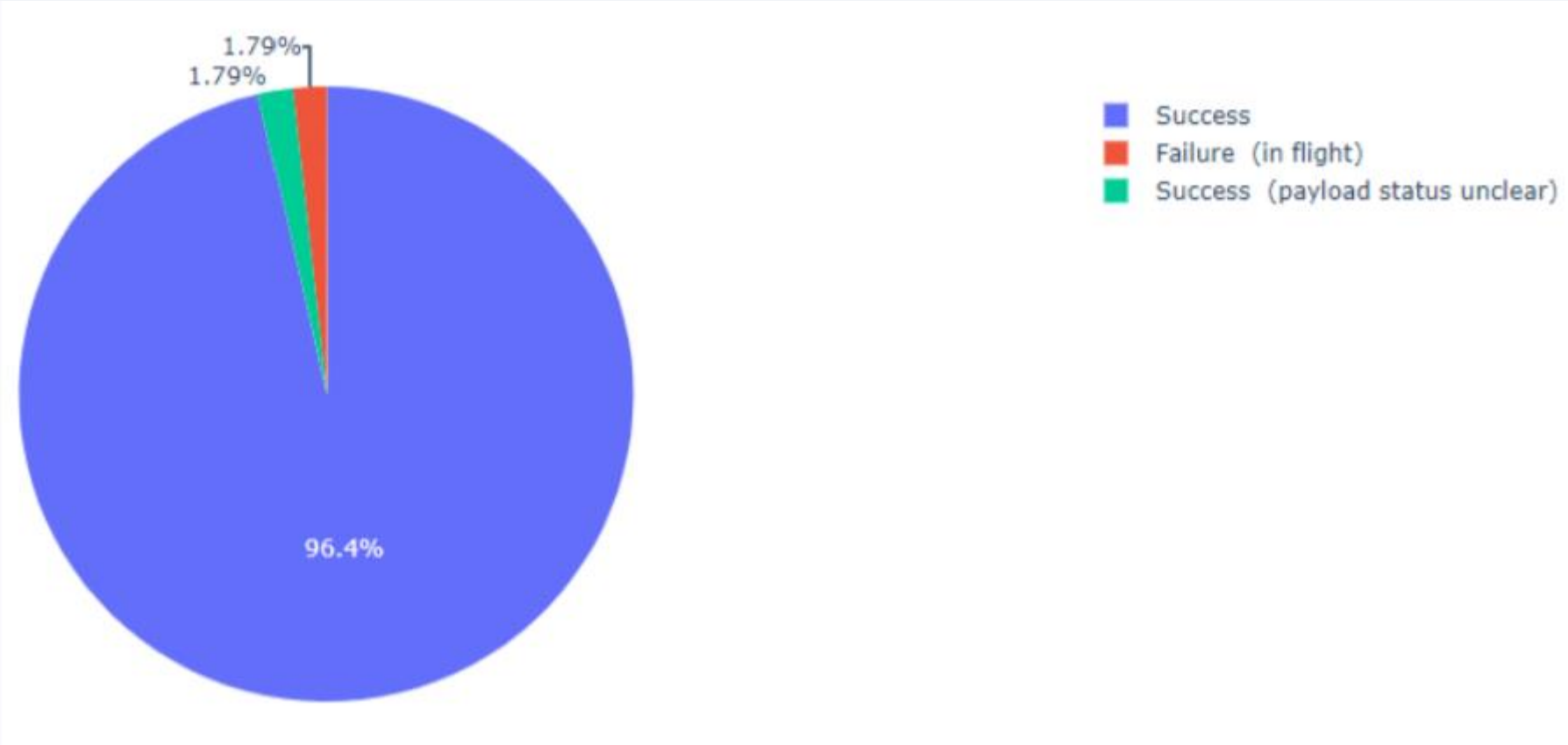
Section 5

Build a Dashboard with Plotly Dash



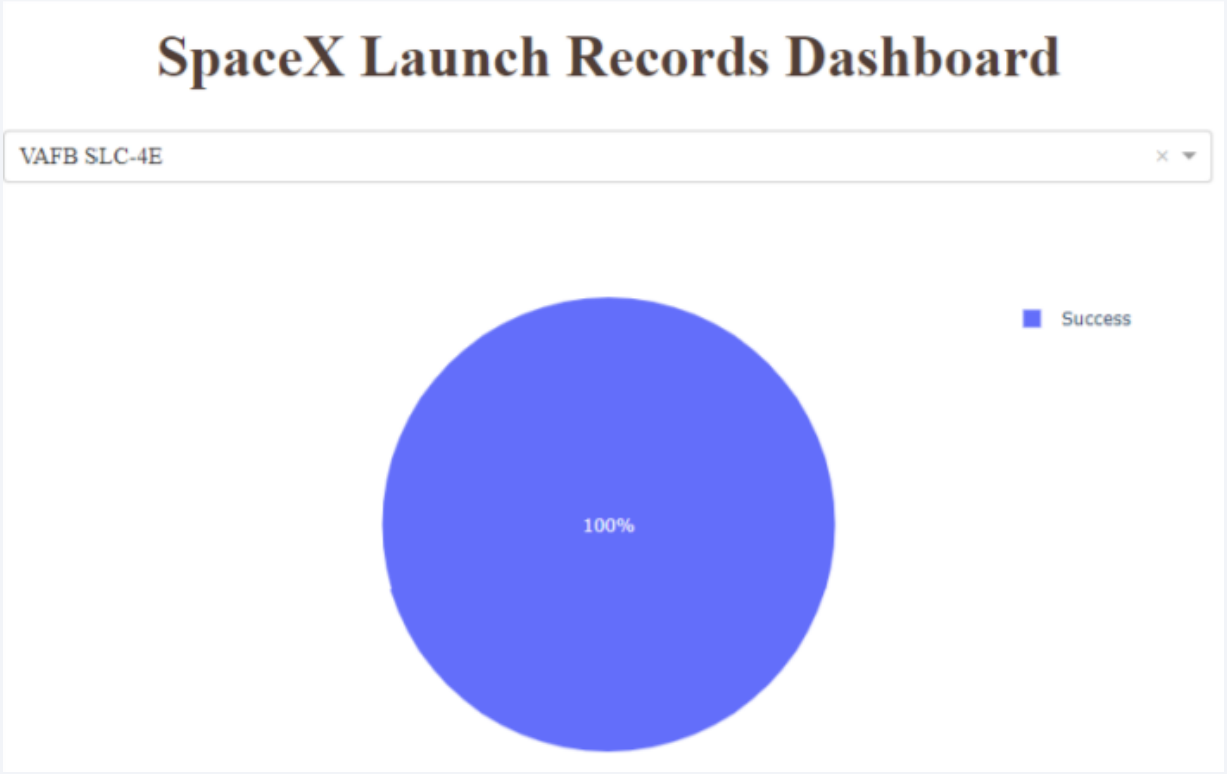
SpaceX Launch Records Dashboard

- This pie chart demonstrates that overall vast majority of the Falcon 9 outcomes were successful

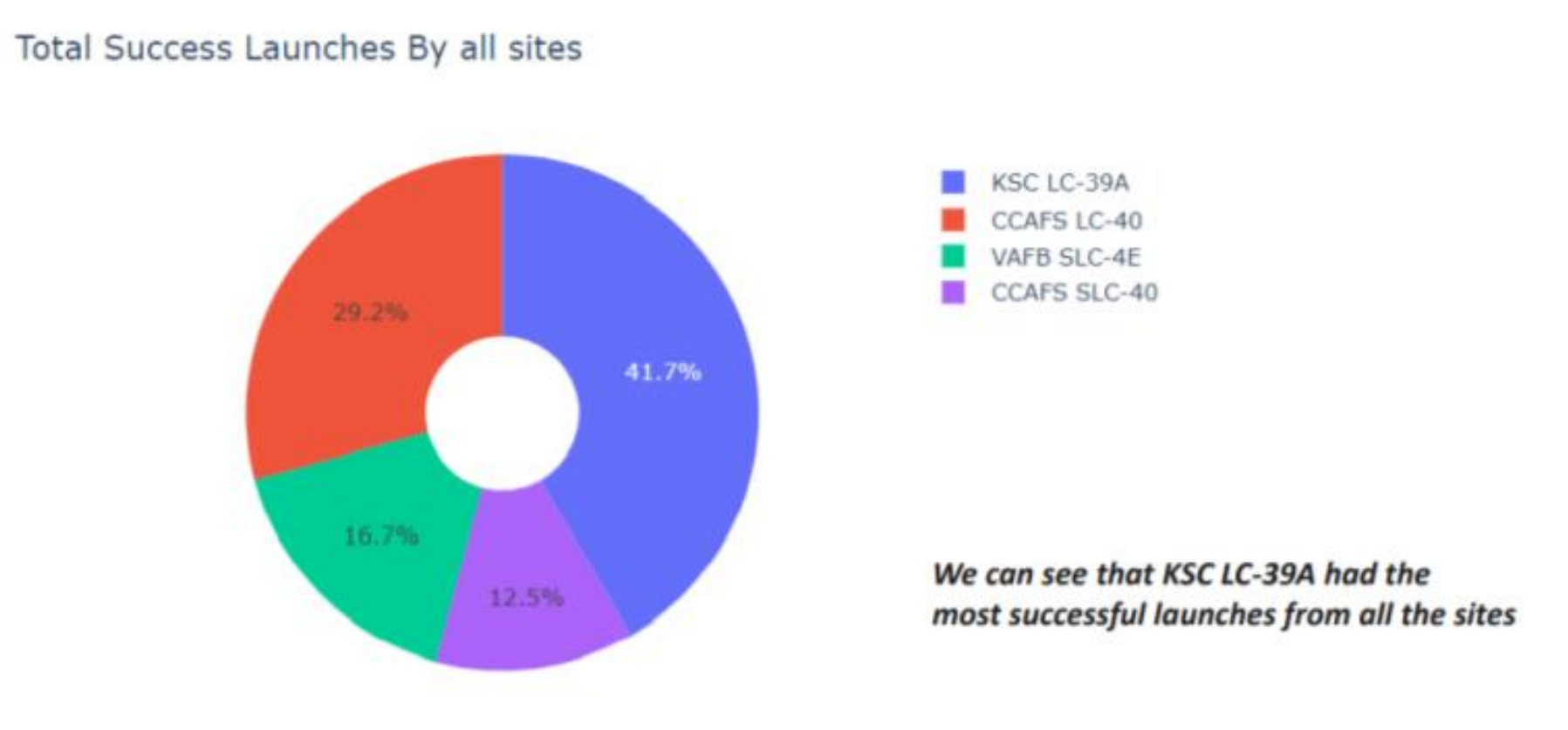


Space X – Most Successful Launch Site (for Falcon 9)

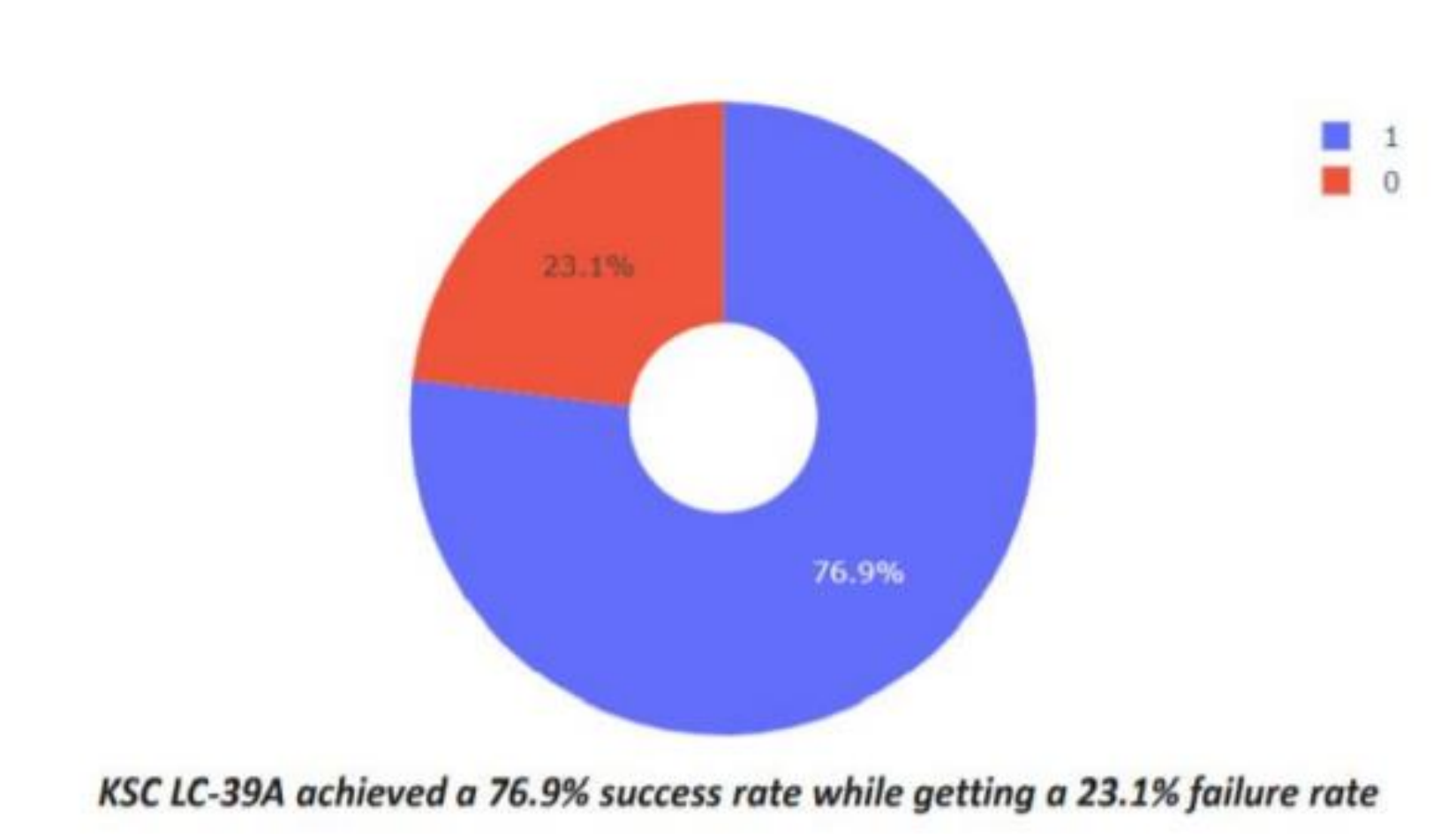
- Based on this chart we can see that VAFB SLC-4E launch site, had no failures for Falcon9 to date



DASHBOARD – Pie chart showing the success percentage achieved by each launch site



DASHBOARD – Pie chart for the launch site with highest launch success ratio

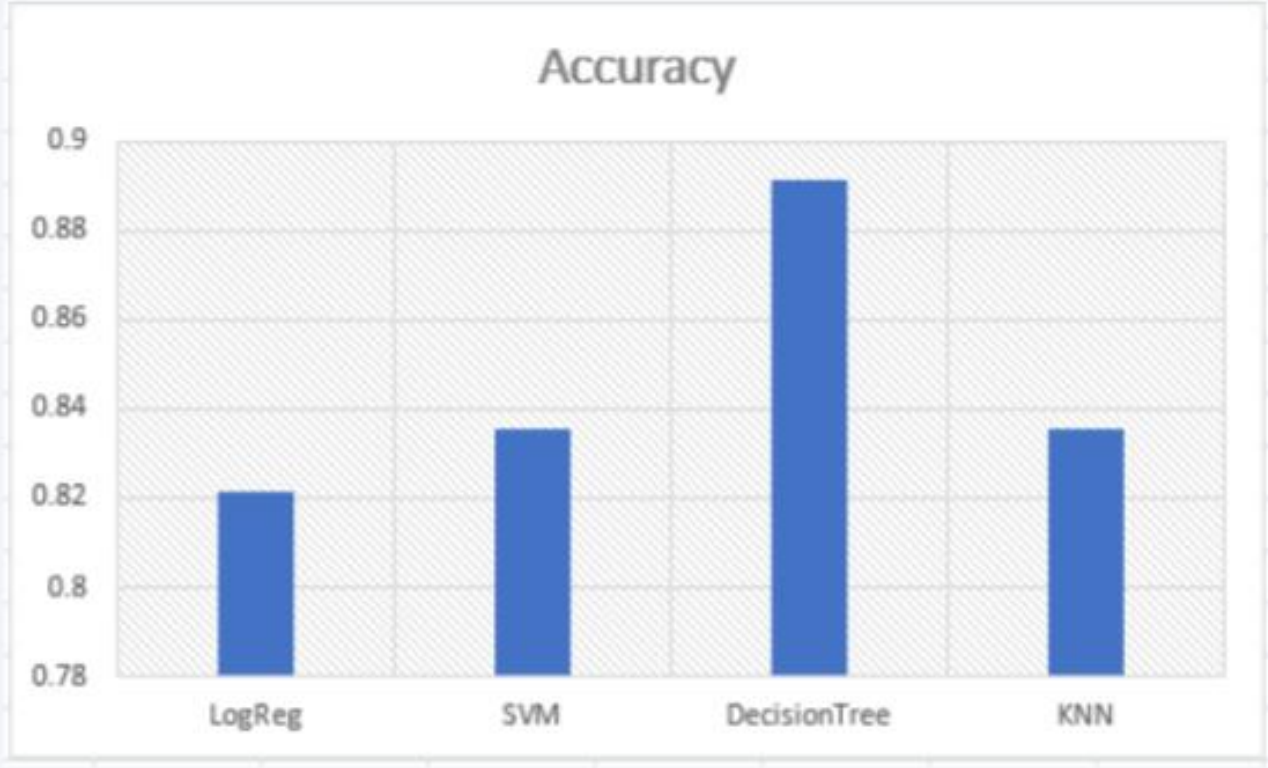


Section 6

Predictive Analysis (Classification)

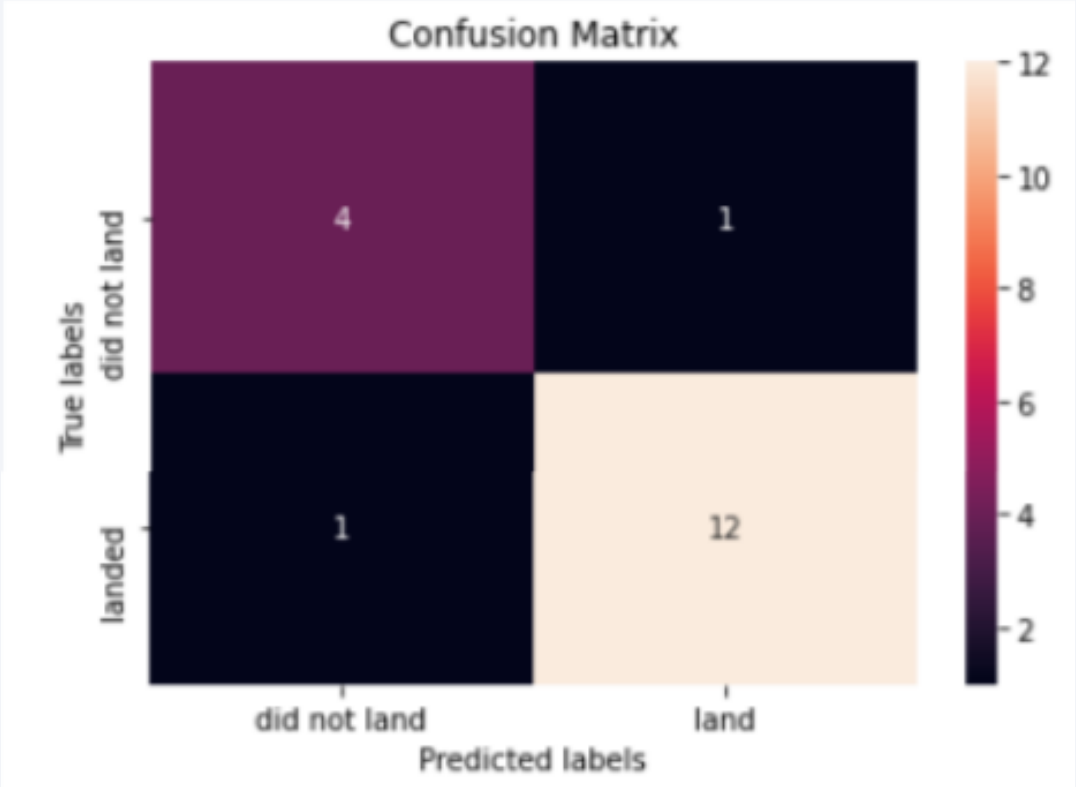
Classification Accuracy

Algo	Accuracy
LogReg	0.821429
SVM	0.835714
DecisionTree	0.891071
KNN	0.8357



Confusion Matrix

- Based on actual Score over test results LogReg and SVM had similar results
- This Confusion matrix highlights identical problem in the 2 black quadrants were 1 of each kinds of predictions was incorrect.



Conclusions

- Space X project reliability is improving over years
- Landings on ground platforms are more reliable than on water
- It is possible with 80%+ accuracy to predict the outcome of a mission based on Flight #, Payload, Booster and Orbit details

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!