IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Peter Leupold
November 17, 2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection with SpaceX API and Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis using SQL
  - EDA through Data Visualization
  - Interactive Visual Analytics (Folium and Dashboard)
  - Predictive Analysis through Machine Learning

- Summary of all results
  - Exploratory data analysis results
  - Interactive analytics demo in screenshots
  - Predictive analysis results

Introduction

- Project background and context

  Reusing the first stage of a space rocket can save a considerable amount of money.

  SpaceX is able to offer launches for $62 million, while competitors charge the triple or more.

  But the first stage can only be reused if it lands successfully.

  If we can identify the key factors that determine whether a landing will be successful or not, we can help increase the rate of success and thus reduce cost.

- Problems you want to find answers to
  - Which are the determining factors for a successful landing?
  - Predict the price of a planned launch from a set of these parameters.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data about launches was downloaded via SpaceX API and scraped from wikipedia
- Perform data wrangling
  - Missing values were replaced by mean values
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Four ML models were used to predict mission outcome
  - The four accuracies were compared

Data Collection

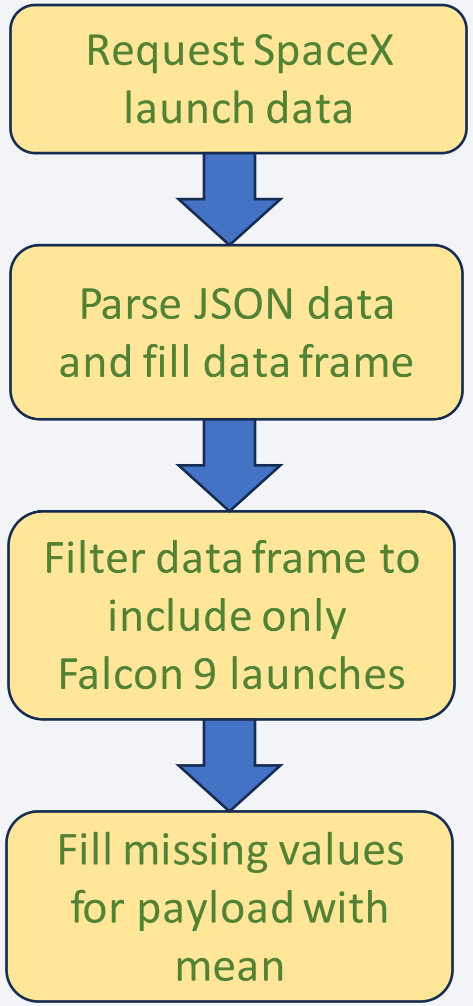
- Collect openly available launch data via
  - Download through SpaceX API
  - Scraping Wikipedia page
- Parse collected data
- Fill data frame
- Deal with missing data

Data Collection – SpaceX API

- SpaceX REST API offers access to launch data
- Make GET request
- Retrieve format is JSON
- Parse data into data frame
- Filter out only Falcon 9 launches
- Fill missing values with mean

- Notebook for API:

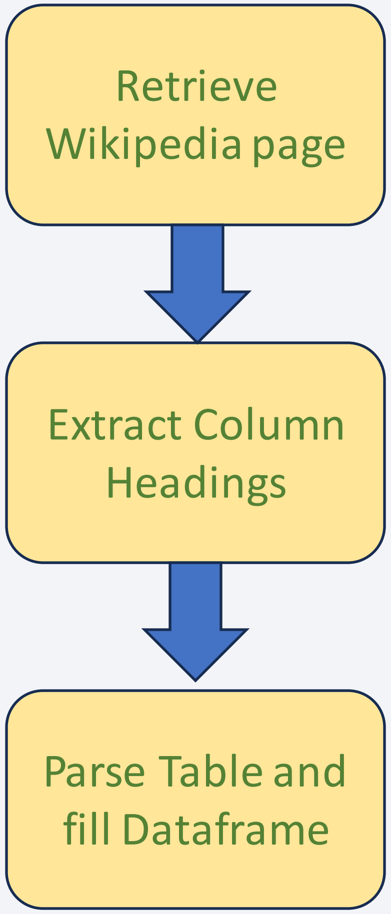
  [Github SpaceX API]



Data Collection - Scraping

- We retrieve Falcon 9 launch data from Wikipedia
- Via BeautyfulSoup we
  - Extract the table
  - Obtain the column headings
  - Parse the table data
- We store the data in a data frame for further processing
- Notebook for web scraping at:

  Github Webscraping

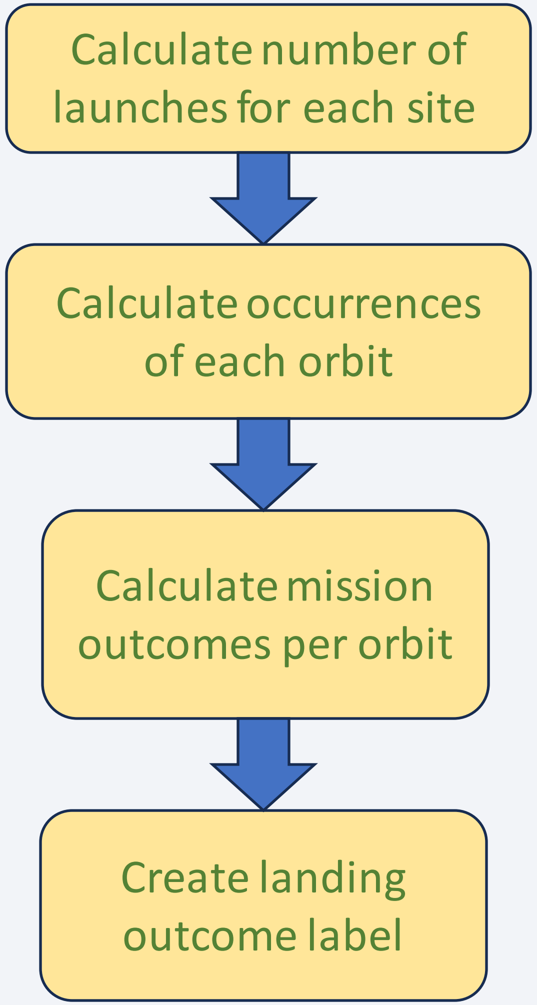


Data Wrangling

- Detect missing data
  - Only in column LandingPad

- Convert landing outcome into categorial data
  - Find values for good and bad outcomes
  - Introduce new column for good/bad outcome

- Notebook for data wrangling at:
  - [Github Data Wrangling]()



EDA with Data Visualization

- Scatter plotsFlightNumber
  - FlightNumber vs. PayloadMass
  - FlightNumber vs. LaunchSite
  - PayloadMass vs. LaunchSite
  - FlightNumber vs. Orbit
  - PayloadMass vs. Orbit
- Bar charts plotted
  - Success Rate vs. Orbit
- Line chart
  - Success Rate by year

Explore connection between launch site and success

Explore connection between orbit and success

Explore connection between Year of launch and success

- Notebook for visual EDA at:
- Github Visual EDA

EDA with SQL

- Through SQL queries we explore connections between success rate and
  - Launch Site
  - Payload masses
  - Launch Dates
  - Booster types

- Notebook for EDA with SQL:
  - [Github for EDA with SQL]

Build an Interactive Map with Folium

- Add objects to visualize mission outcomes on map
  - Circles for launch sites
  - Marker clusters indicating launches and success or not
  - Lines indicating distance to coast, rail, and road
- Allows visual analysis of geographic influence factors
- Notebook for Folium Map:
  - [Github for Folium Map]

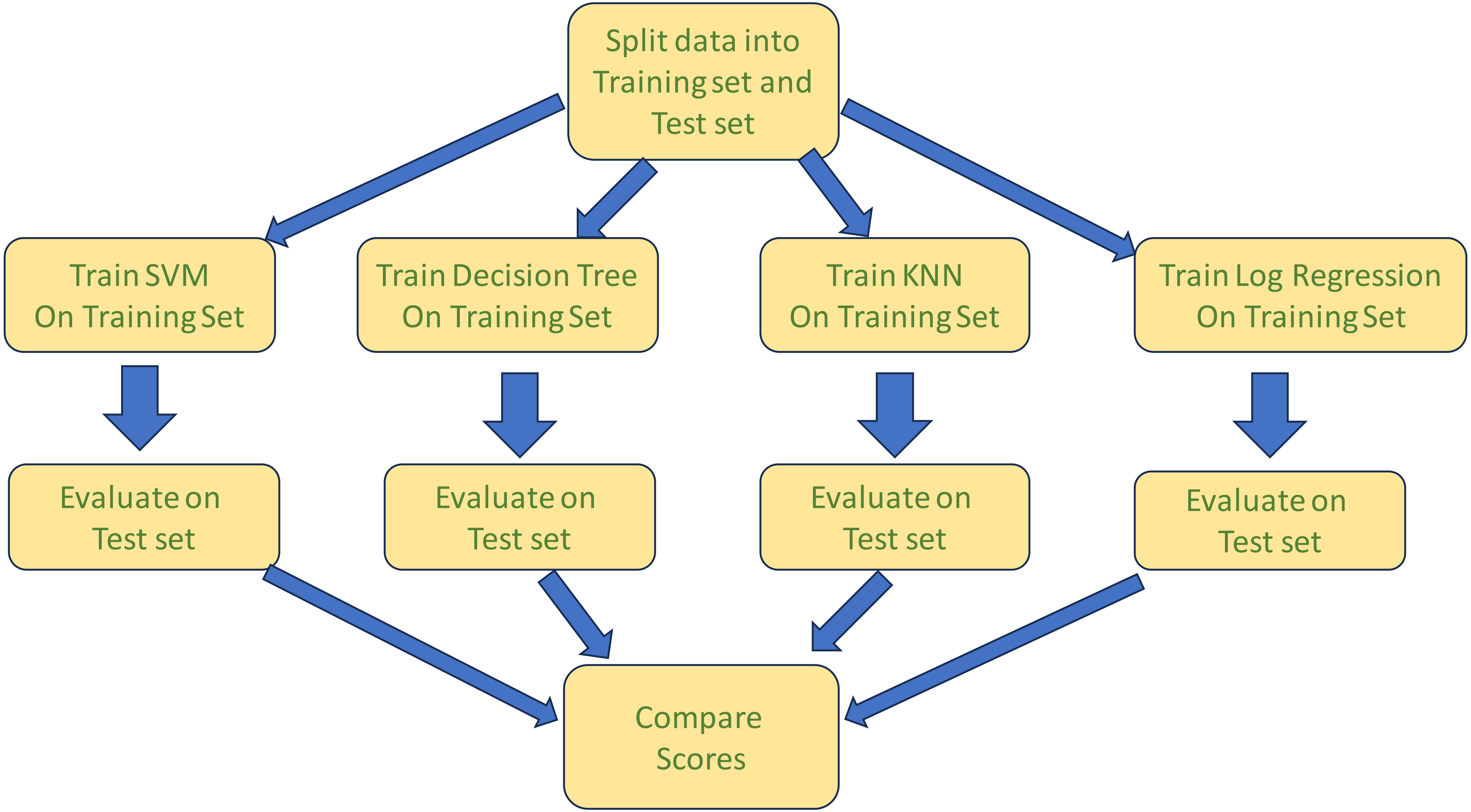
Build a Dashboard with Plotly Dash

- A pie chart displays
  - The distribution of successful landings between launch sites
  - The distribution of failures and successes for a given site
- The view of one or all sites is selected in a dropdown menu
- Payload masses can be restricted to a range via an input slider
- A Scatterplot groups missions by booster version, outcome, and payload mass
- The dashboard allows closer analysis of influence of launch site on outcome
- Python code for dashboard
  - Github for Dashboard

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Notebook for ML prediction:
  - Github for ML Prediction

Predictive Analysis (Classification)



Results

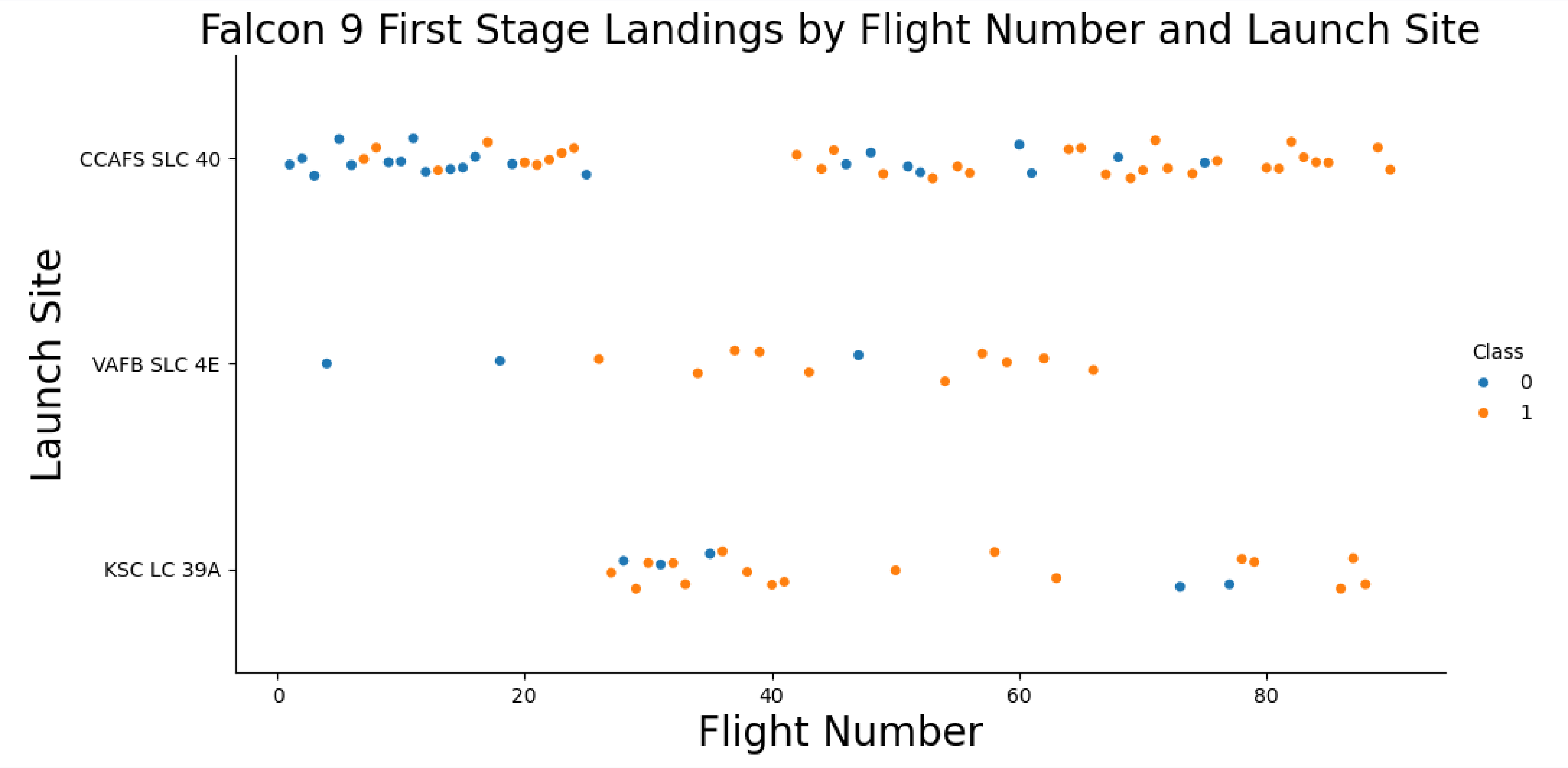
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

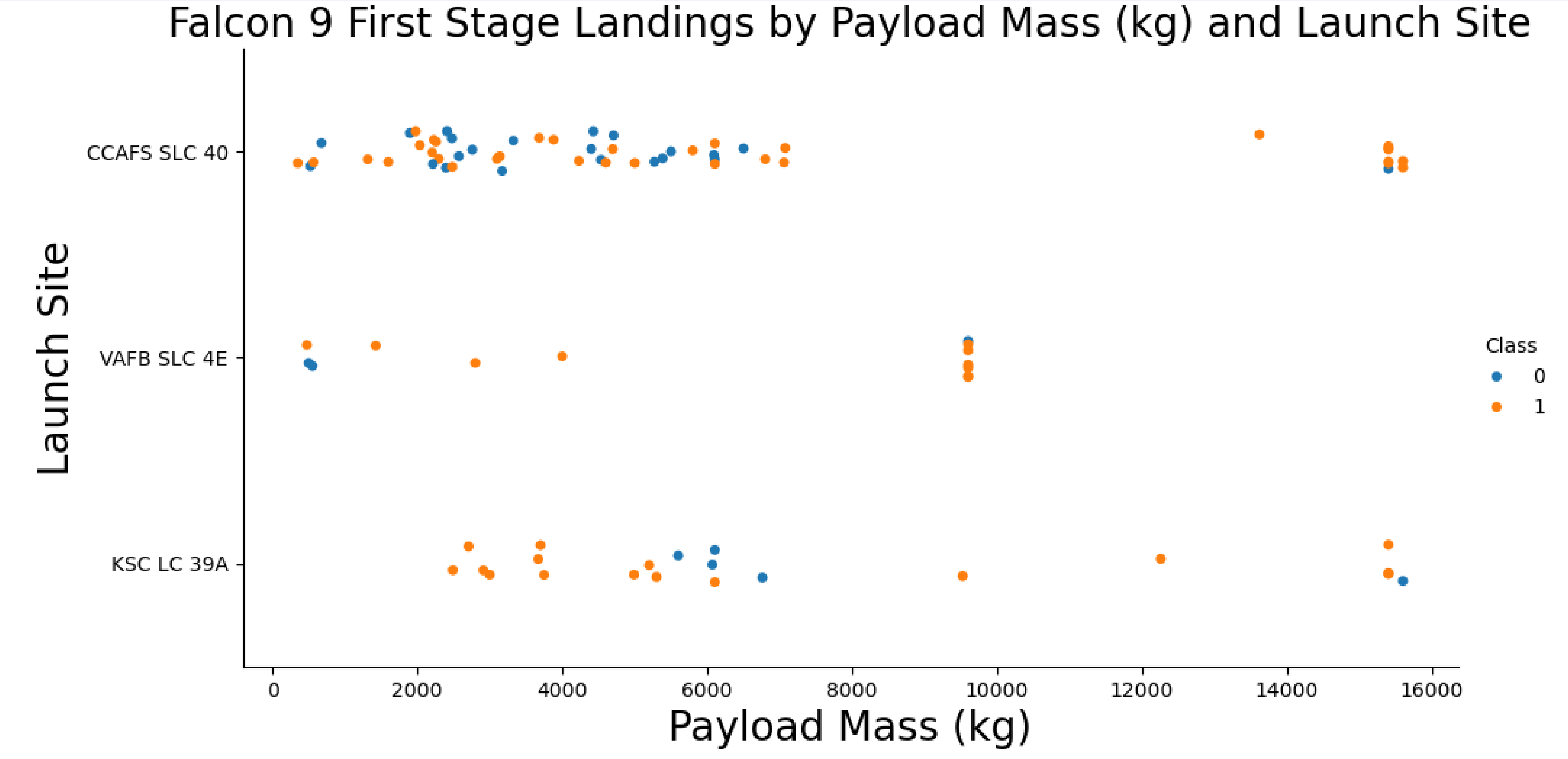
Flight Number vs. Launch Site

Blue dots successes, orange dots failures



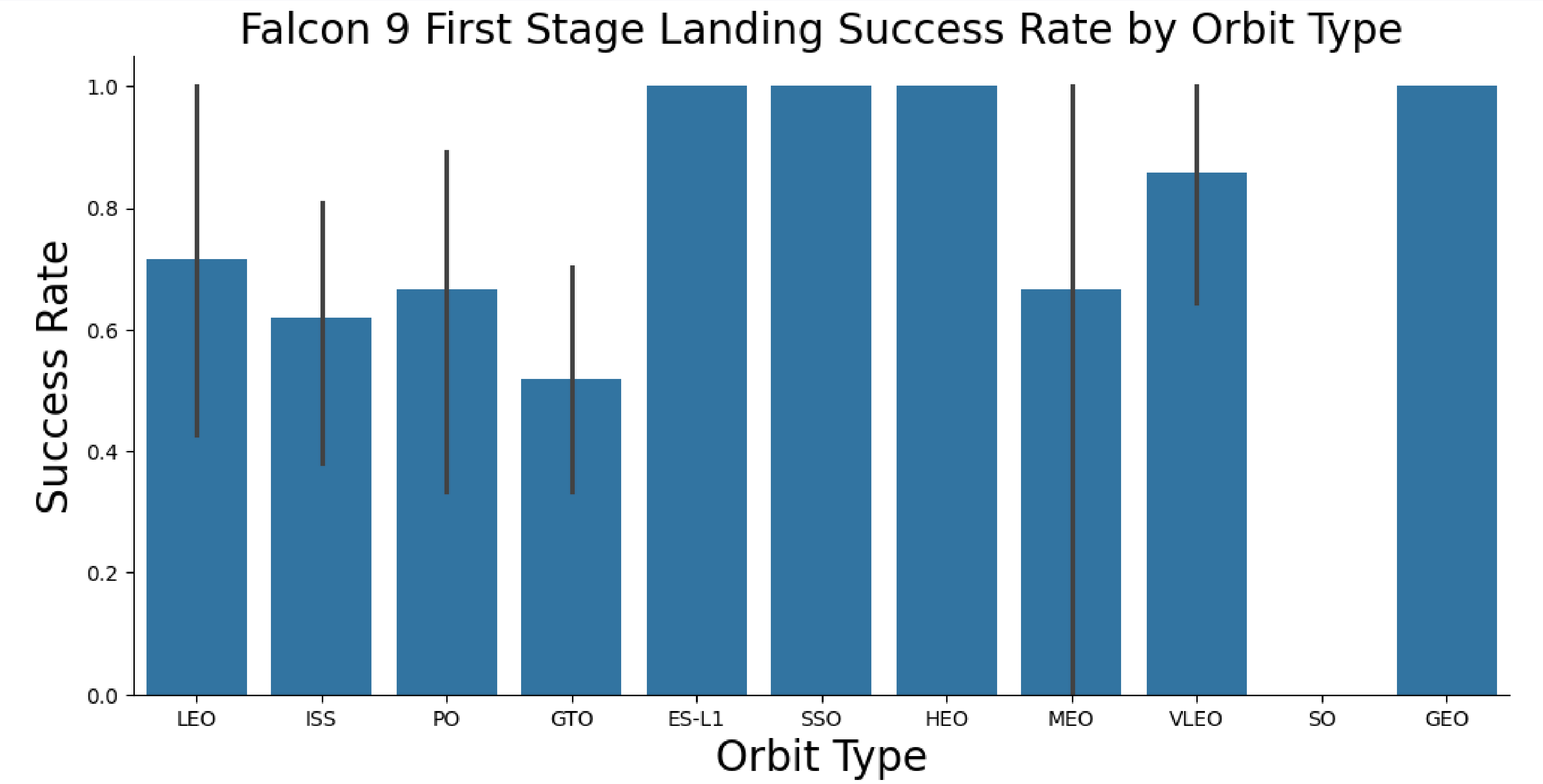
Payload vs. Launch Site

Blue dots successes, orange dots failures



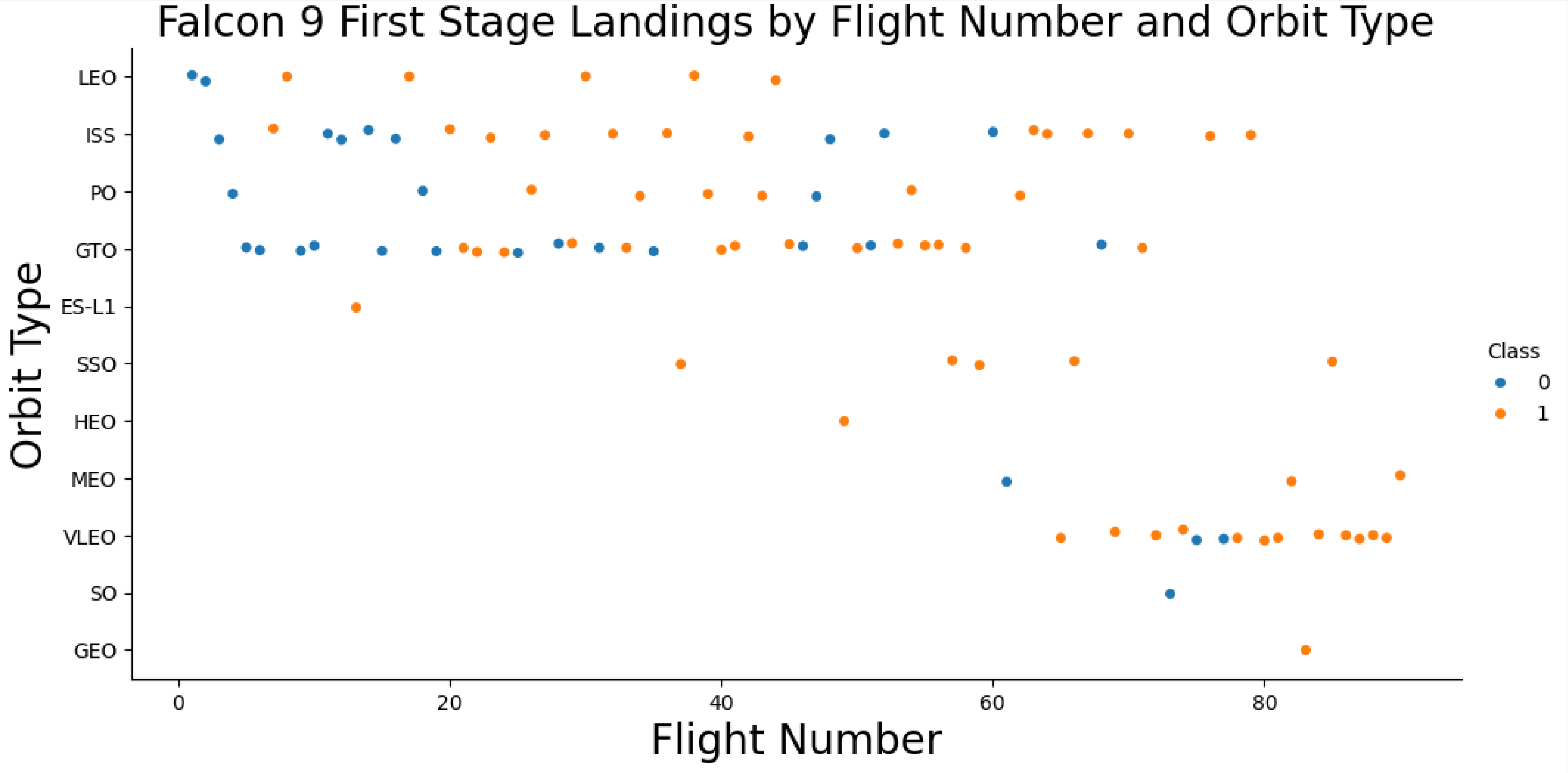
Success Rate vs. Orbit Type

No success for the SO orbit.



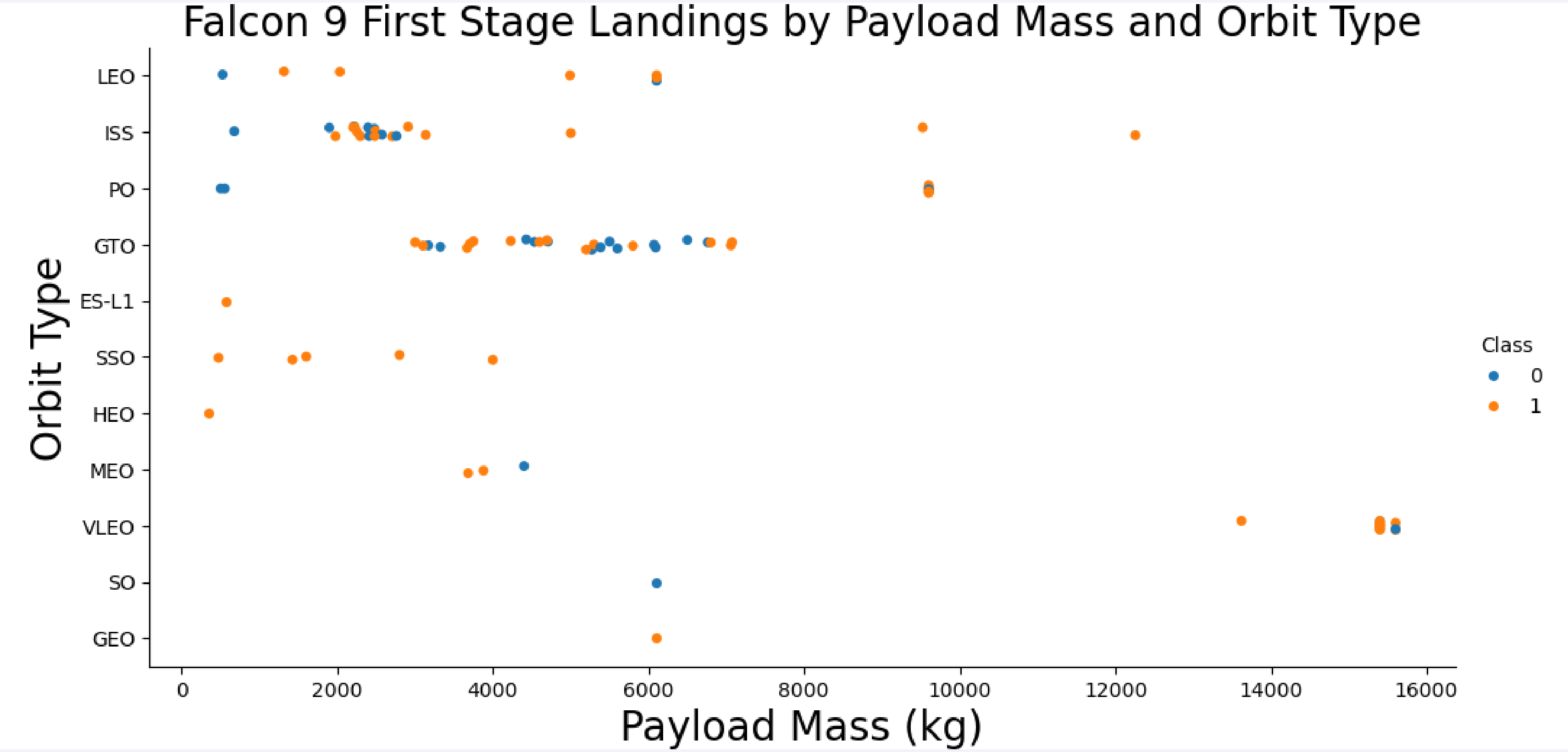
Flight Number vs. Orbit Type

As flight numbers rise, orbit types change



Payload vs. Orbit Type

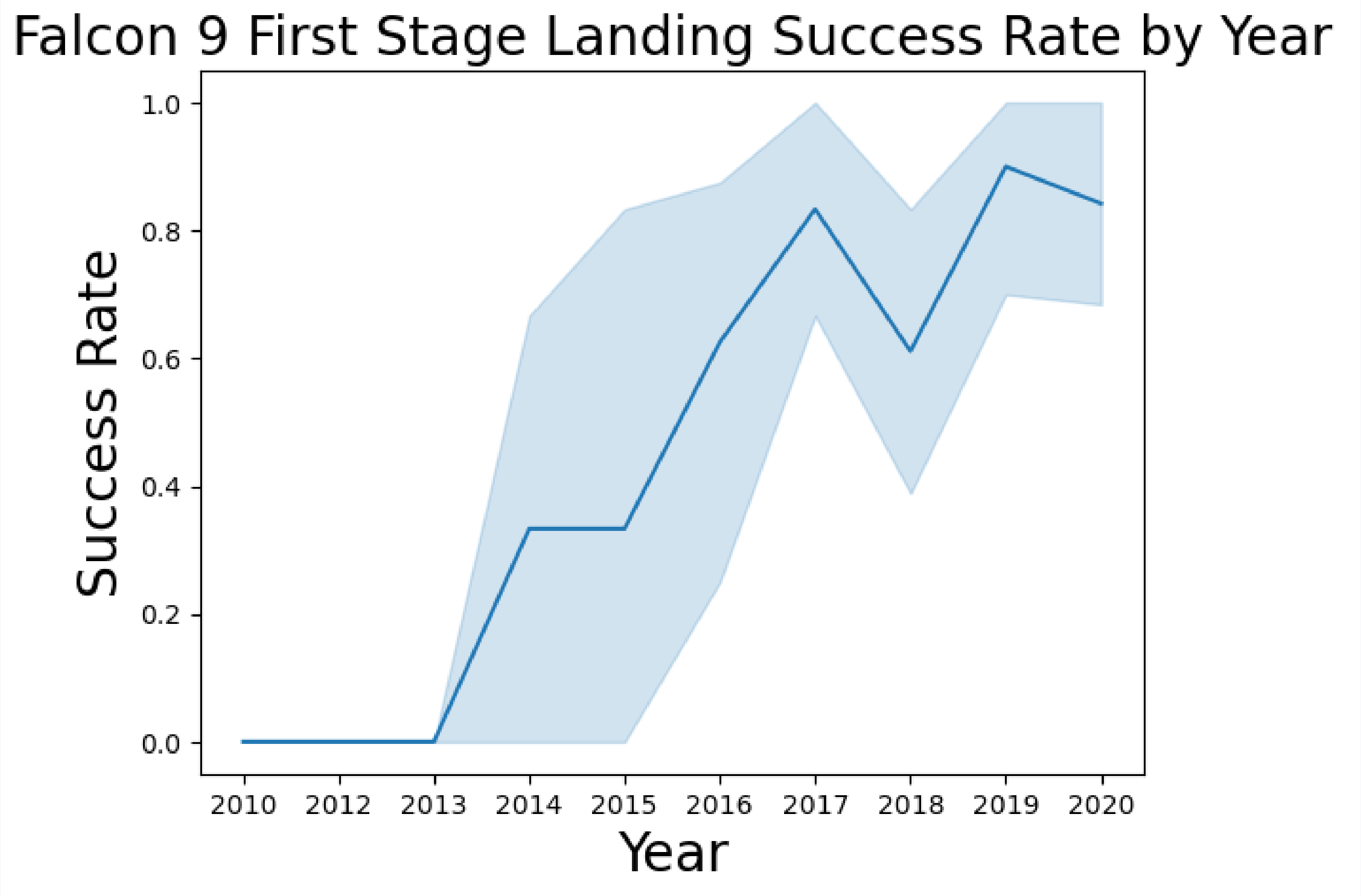
Different orbits have different typical payload masses



Launch Success Yearly Trend

- As time passes the success rate increases

- A stable high value might have been reached



All Launch Site Names

- Find the names of the unique launch sites

Query:

```
select DISTINCT "Launch_Site" from SPACEXTABLE
```

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

- There are four sites

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

Query:

```
select * from SPACEXTABLE
    where "Launch_Site" LIKE "CCA%" limit 5
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- Five examples of a large total number

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

Query:

```
select sum("PAYLOAD_MASS__KG_") from SPACEXTABLE
```

sum("PAYLOAD_MASS__KG_")
619967

- The total payload mass is 620 tons

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

Query:

```
select AVG("PAYLOAD_MASS__KG_") from SPACEXTABLE
   where "Booster_Version" = "F9 v1.1"
```

AVG("PAYLOAD_MASS__KG_")
2928.4

- The average payload had 2928 kg

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

Query:

```
select min(Date) from SPACEXTABLE
   where "Landing_Outcome" = "Success (ground pad)"
```

min(Date)
2015-12-22

- The first successful landing was on December 22, 2015

Successful Drone Ship Landing with Payload between 4000 and 6000

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

Query:

```
select distinct "Booster_Version" from SPACEXTABLE
   where "Landing_Outcome" = "Success (drone ship)"
      AND "PAYLOAD_MASS__KG_" > 4000
      AND "PAYLOAD_MASS__KG_"<6000
```

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

- Four boosters made successful landings in the specified payload range

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

Query:

```
Select(
   SELECT COUNT(*) As "Success" FROM SPACEXTABLE WHERE "Mission_Outcome" LIKE '%succe%')
      As "Success" ,
   COUNT(*) As "Failures" FROM SPACEXTABLE WHERE "Mission_Outcome" NOT LIKE '%succe%'
```

- A very high success rate in the sample

Success	Failures
100	1

Boosters Carried Maximum Payload

- We list the names of the boosters which have carried the maximum payload mass

Query:

```
select distinct "Booster_Version" from SPACEXTABLE
    where "PAYLOAD_MASS__KG_" = (select MAX("PAYLOAD_MASS__KG_")
         from SPACEXTABLE)
```

- Many different booster versions carried the 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

2015 Launch Records

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

Query:

```
SELECT substr(Date,6,2) as month, DATE,BOOSTER_VERSION, LAUNCH_SITE, "Landing_Outcome"
     FROM SPACEXTABLE where Landing_Outcome = 'Failure (drone ship)'
     AND substr(Date,0,5)='2015';
```

month	Date	Booster_Version	Launch_Site	Landing_Outcome
10	2015-10-01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

Two failures were found.

The months and year were extracted via substr from the date string

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

- Rank the count of landing outcomes between the dates 2010-06-04 and 2017-03-20, in descending order

Query:

```
SELECT Landing_Outcome, count(Landing_Outcome) AS "Count" FROM
SPACEXTABLE WHERE
    DATE BETWEEN '2010-06-04' and '2017-03-20')
    Group BY by Landing_Outcome
    ORDER by Count DESC;
```

- The possible outcomes are listed with their number of occurrences in descending order

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

Section 3

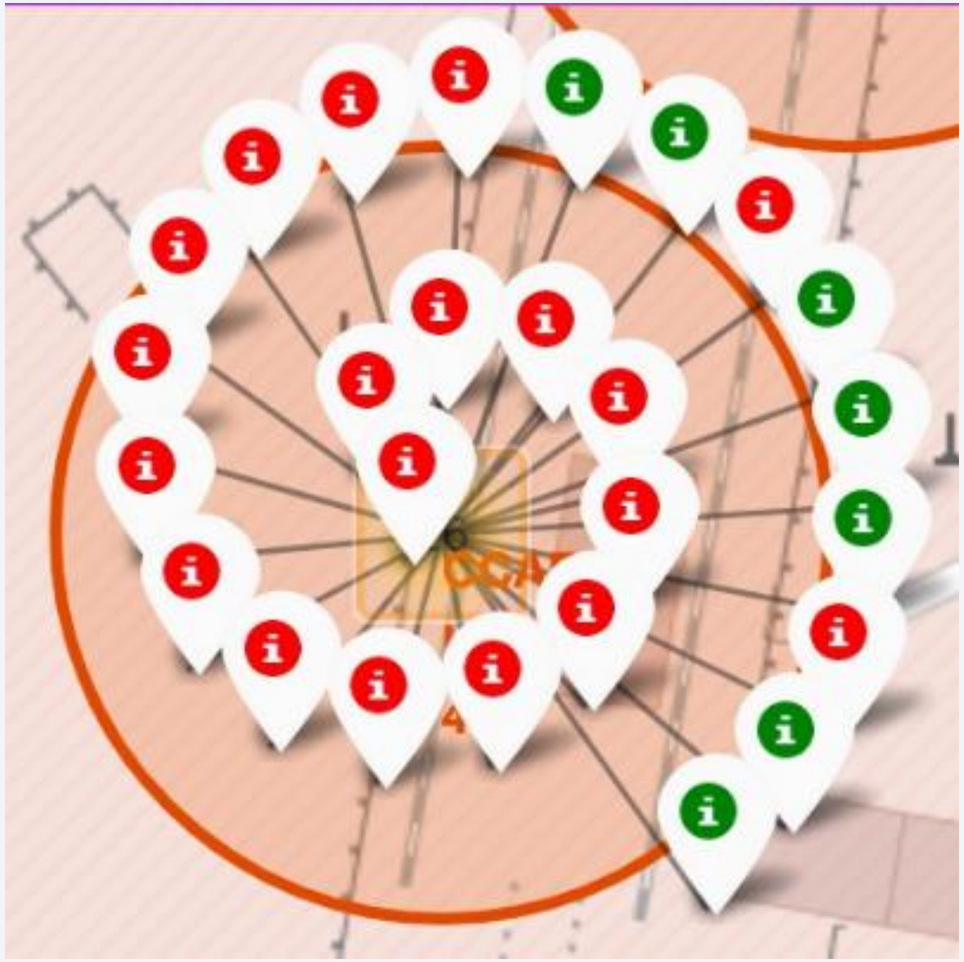
Launch Sites Proximities Analysis

SpaxeX Launch Sites

- All sites are in the Southern USA
- All sites are on the coast
- In California and Florida several sites are close to each other

Launch Markers around Sites

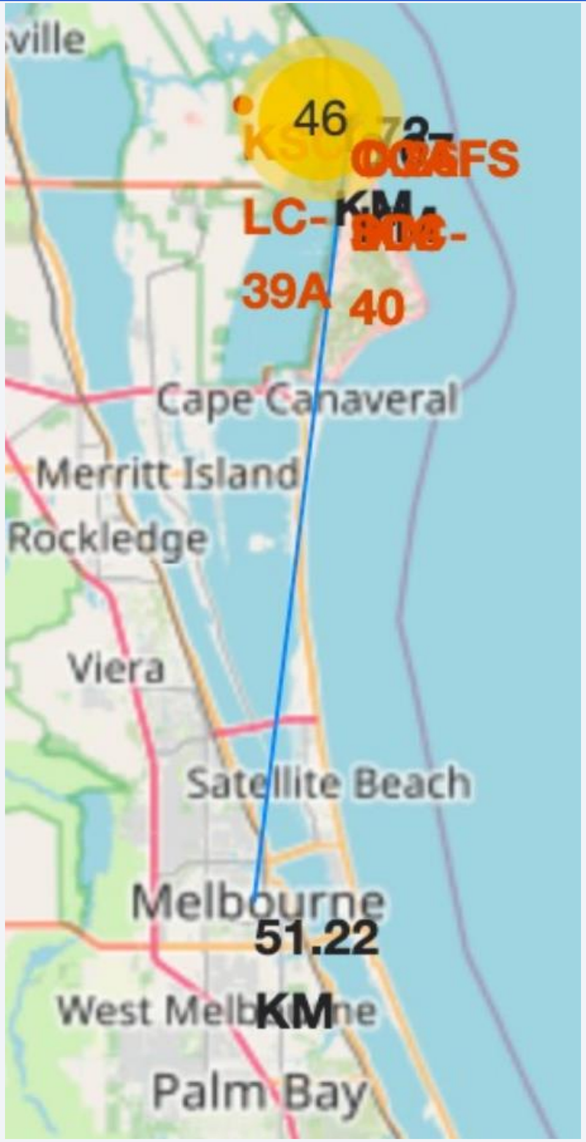
- For each launch from a site there is one marker
- Our example : CCAFS LC-40
- Red is for failure, green for success
- This allows a quick visual analysis of success rates for sites



Distances of Launch Sites to Landmarks

- There are patterns concerning the surroundings of launch sites

- Roads are crucial infrastructure and are always close
- All sites are close to the coast so failed launches can crash into water
- Launch sites are not close to cities, in the example minimum 51 km away fpr security reasons



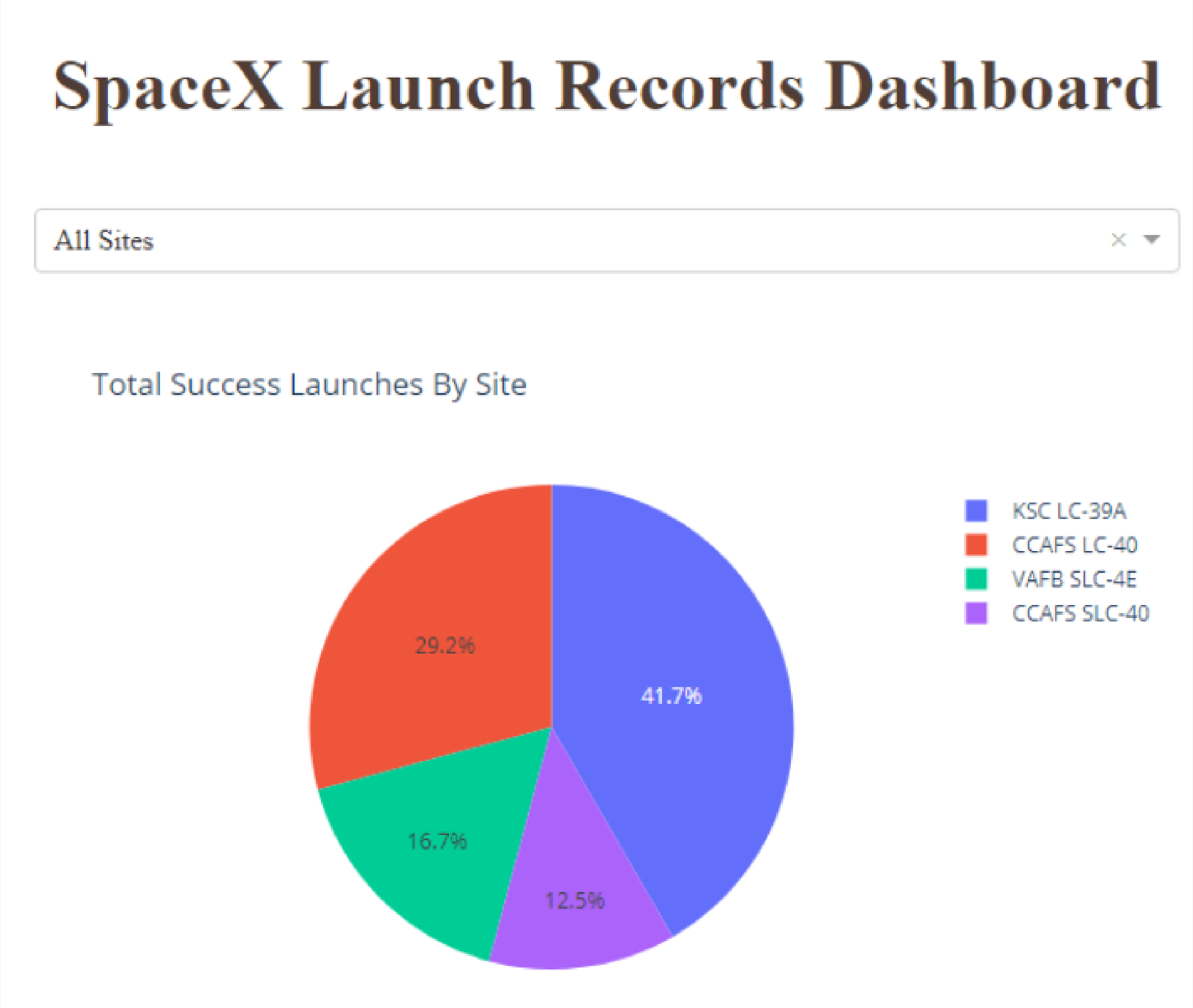
Section 4

Build a Dashboard with Plotly Dash

Successful Launches per Site

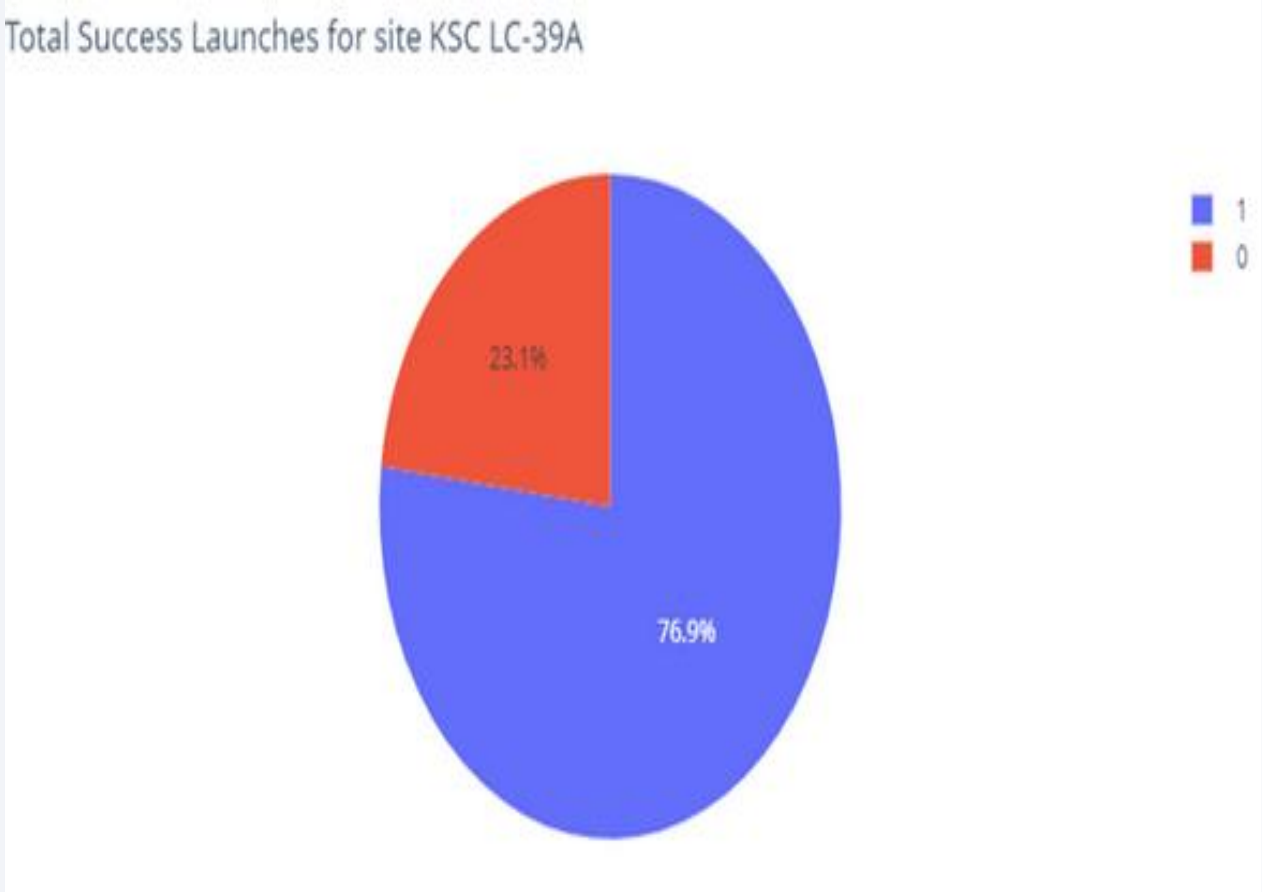
- Site KSC LC-39A has the biggest share of successful launches

- VAFE SLC-4E and CCAF5 LC-40 have low numbers of successful launches



Site with highest Success Ratio

- Red is the share of unsuccessful landings
- Blue is the share of successful landings
- Site KSC LC-39A has a good success rate of 76,9%
- Less than one out of four first stages do not land successfully



Payload vs. Launch Outcome per Site

- Payload has an influence on the success
- The highest success rate is in the range from two to five tons
- Among boosters, FT has the best performance

CCAFS LC-40

CCAFS SLC-40

KSC LC-39A

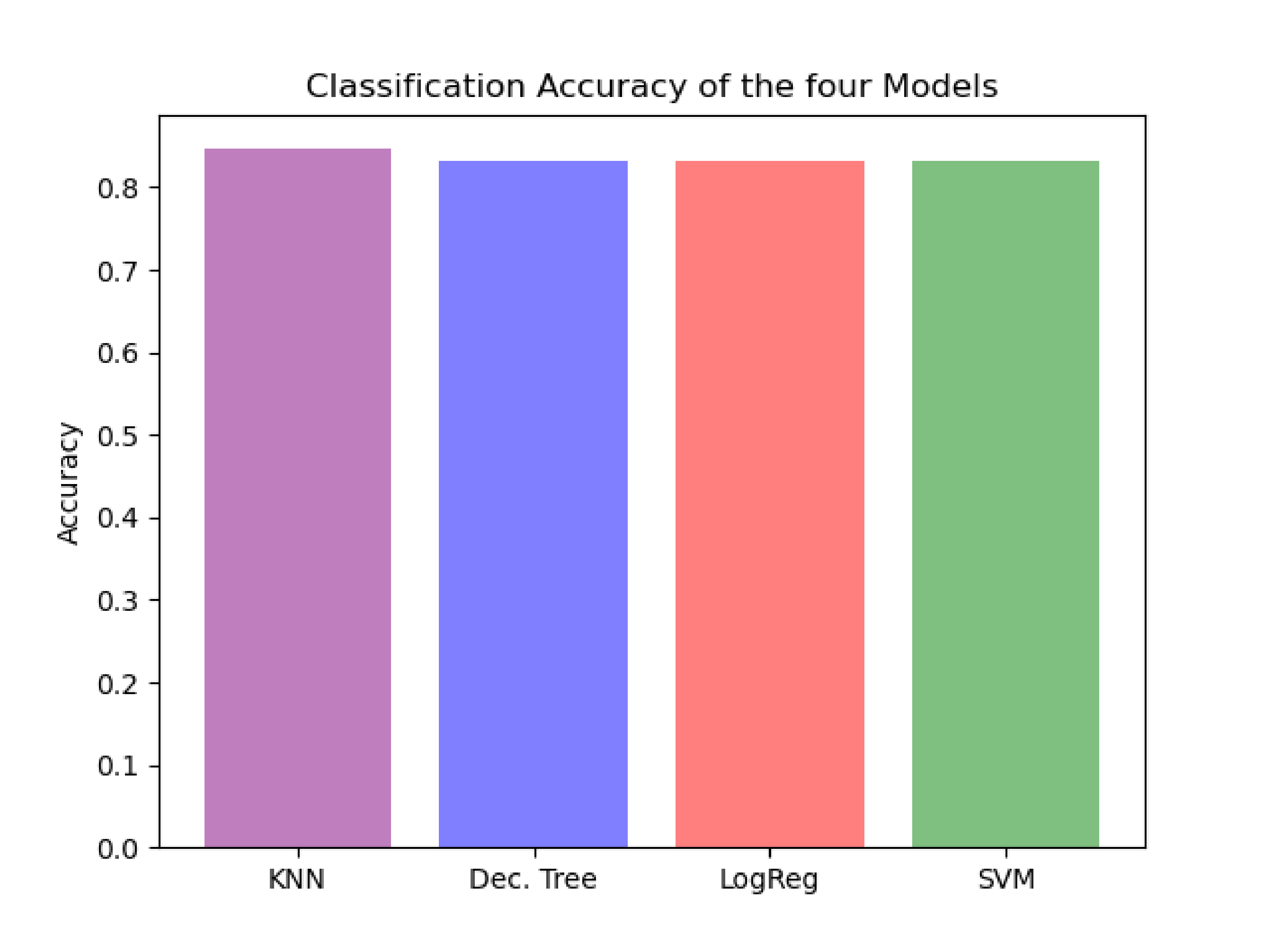
VAFB SLC-4E

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Performance is almost equal for all four models
- The nearest neighbor model has a very slight advantage



Confusion Matrix

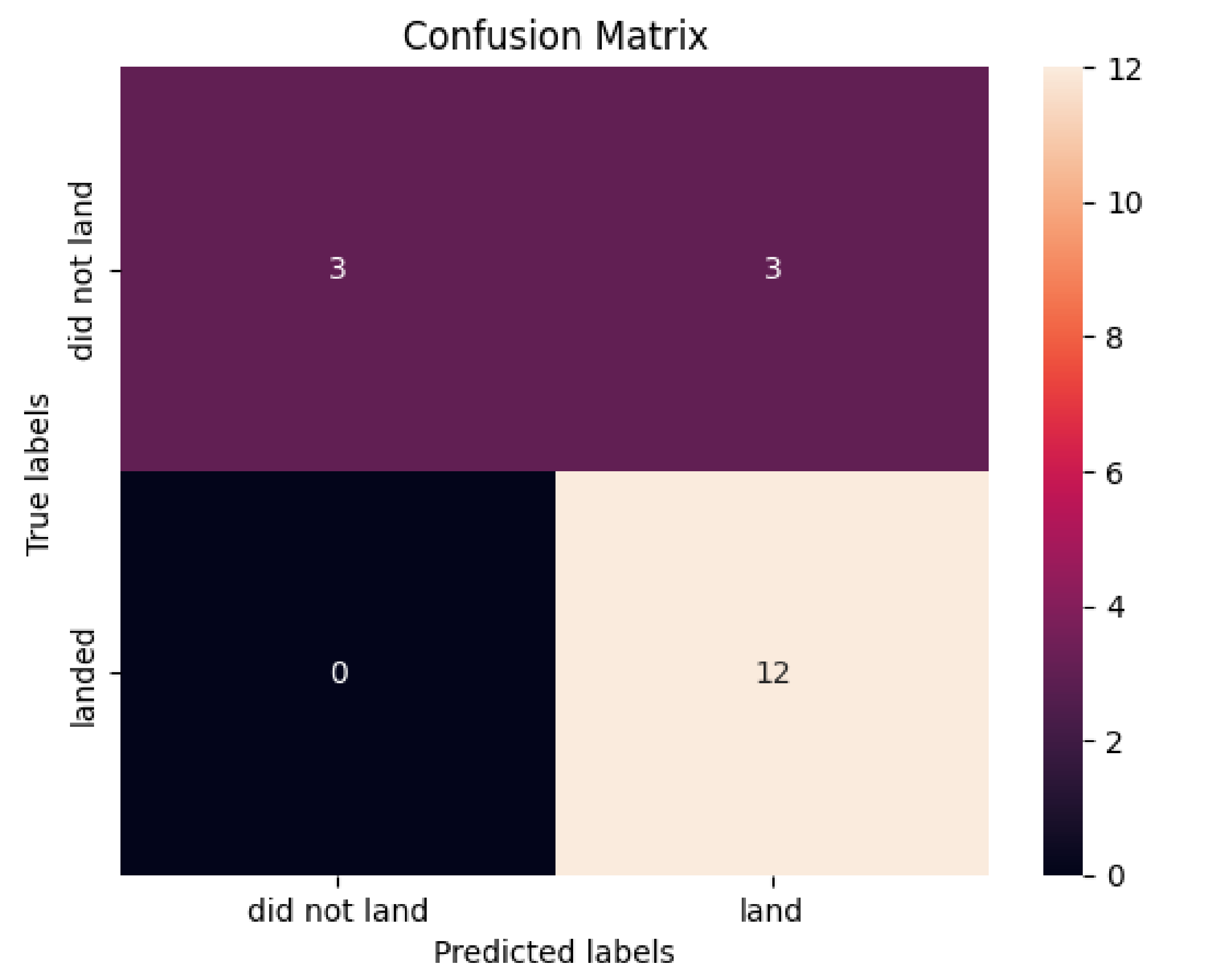
The Logistic Regression model classifies 15 cases correctly and fails in three cases.

The correct cases are

- Three predicted failures
- 12 predicted successes

Mistakes are

- Three failures were not predicted
- No success was falsely predicted as failure



Conclusions

- All four prediction models perform almost equally well
- They predict the mission outcome with an accuracy of about 84%
- Based on these predictions SpaceY could select more favorably launch sites, dates etc.
- The increased success rate could lead to a lower price than the one of SpaceX
- Our predictions promise a bright future for

SpaceY