IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Didac Marti Pinto
December 2022

Outline

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

Executive Summary

- Methodology
  - First, collect data
  - Second, analyze data with:
    - Visualize data statistics
    - Interactive dashboard
    - Maps
  - Third, build a machine learning (ML) model to predict a mission success
- Summary of all results
  - Exploratory data analysis to see what variables are related to success
  - 4 different ML models to predict success

Introduction

- SpaceX is an American company founded in 2002 by Elon Musk who in words of Donald J Trump ‘is doing the rocket, he likes rockets, and does good at rockets too’. Click here.
- The objective of this project is to understand what leads to the success or failure of a rocket mission, and be able to predict if a mission will be successful or not

Section 1

Methodology

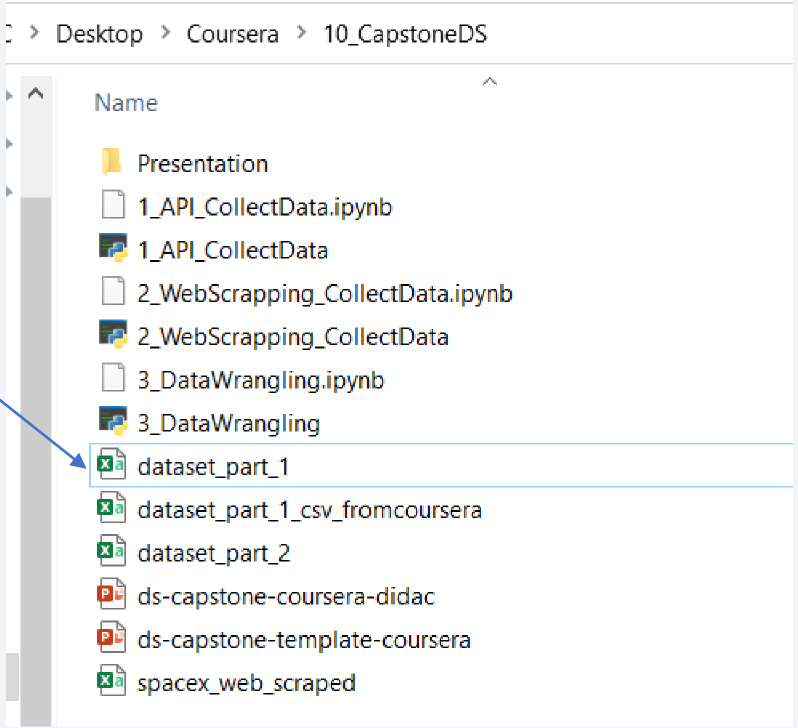
Methodology

1. Data collection:
    - SpaceX API
    - WebScrapping
2. Data wrangling
    - Prepare data for analysis
    - Calculate landing sites, orbits, and outcomes
3. Data visualization
    - Scatterplots, bar charts, and time line trends
    - Calculate landing sites, orbits, and outcomes
4. SQL queries
5. Folium maps
    - Show location of launch sites
    - Show successes/failures at each launch site
6. Plotly Dashboard
    - Success Pie chart and correlation chart between payload and sucess
    - Interactive
7. Predictive analysis using ML classification models
    - 4 models: Logistic regression, Support Vector machines, Decision trees, and k-nearest neighbors

Data Collection

- 2 alternative methods to collect data:
  1. SpaceX **API**
  2. **WebScrapping** Wikipedia page “List of Falcon 9 and Falcon Heavy launches”
     - (use a static version of the link)

Save in .csv format



Data Collection – SpaceX API

1. Request the SpaceX launches data by calling the SpaceX API. Normalize json.

```
# Better use this static link than the one above for this course
static_json_url = 'https://cf-courses-data.s3.us.cloud-object-stor
response = requests.get(static_json_url)
data = pd.json_normalize(response.json())
```

2. Convert it into a readable pandas dataframe (see code for details)

3. Filter the dataframe to include only Falcon 9 Launches

```
data_falcon9 = df[df['BoosterVersion']!='Falcon 1']
```

4. Replace missing 'PayLoadMass' values by the mean of the column

```
# Calculate the mean value of PayloadMass column
PayloadMass_mean = data_falcon9['PayloadMass'].mean()
# Replace the np.nan values with its mean value
data_falcon9['PayloadMass'] = data_falcon9['PayloadMass'].fillna(PayloadMass_mean)
```

Code link in GitHub

Data Collection - WebScraping

1. Get data from the url using 'requests.get'

```
# To keep the lab tasks consistent, you will be asked to scrape the data from
# a snapshot of the List of Falcon 9 and Falcon Heavy launches Wikipage updated on 9th June 2021
static_url = "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launche

# Use requests.get() method with the provided static_url and assign the response to a object
response = requests.get(static_url).text
```

2. Turn it into a 'Beautiful Soup object'

```
# Use BeautifulSoup() to create a BeautifulSoup object from a response text content
soup = BeautifulSoup(response, "html5lib")
```

3. Turn html code into pandas df (see code)

[Code link in GitHub]

Data Wrangling

1. Calculate landings on each site

```
df.value_counts('LaunchSite')
'''
LaunchSite
CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
'''
```

2. Calculate the number and occurrence of each orbit

```
df.value_counts('Orbit')
'''
Orbit
GTO      27
ISS      21
VLEO     14
PO        9
LEO       7
SSO       5
MEO       3
ES-L1     1
GEO       1
HEO       1
SO        1
'''
```

3. Calculate the number of occurrences of mission outcome per orbit type

```
landing_outcomes = df.value_counts('Outcome')
'''
Outcome
True ASDS      41  -- success
None None      19  -- fail
True RTLS      14  -- success
False ASDS      6  -- fail
True Ocean      5  -- success
False Ocean     2  -- fail
None ASDS       2  -- fail
False RTLS      1  -- fail
```

4. Create a 0-1 landing outcome label ‘Class’. And calculate success rate with it

```
# Success rate
df["Class"].mean()
'''
0.6666
'''
```

[Code link in GitHub]()

EDA with Data Visualization

- Categorical scatterplots: To see the relation between success and
  - Launch Site and flight number
  - Launch site and payload
  - Flight number and orbit type
  - Payload and orbit type
- Bar charts
  - Success rate by orbit type
- Time trends
  - Success rate change over time

Code link in GitHub

EDA with SQL

1. Display names launch sites
2. Display 5 records where launch sites begin by ‘CAA’
3. Display total payload mass by boosters launched by NASA
4. Display average payload mass carried by booster version F9 v1.1
5. List date when the first successful landing outcome in a ground pad was achieved

6. List names successful boosters where payload mass (4000,6000)

7. List number of success/fail missions

8. List Names of booster versions that have carried the maximum payload mass

9. List full records in year 2015

10. Rank the count of successful landing outcomes between the date 04-06-2010 and 20-03-2017.

[Code link in GitHub]

Build an Interactive Map with Folium

- Map 1: Markers showing location of launch sites in the US
- Map 2: Markers also show the number of success missions at each site
- Map 3: Show lines that measure proximity to objects

Code link in Github

Build a Dashboard with Plotly Dash

- Pie chart showing success/failure.
  - Interactivity: select site
- Correlation between payload and success.
  - Interactivity: Choose payload range and site.

Code link in GitHub

Predictive Analysis (Classification)

- Build 4 Machine Learning Models:
  - Logistic Regression
  - Support Vector Machines (SVM
  - K-Nearest Neighbors (KNN)
  - Decision Trees
- Evaluate performance:
  - Cross validation: Split sample into train and test. Measure accuracy in the test (and train) sample
- Choice:
  - Almost identical performance
  - Decision trees marginally better

Code link in GitHub

Results

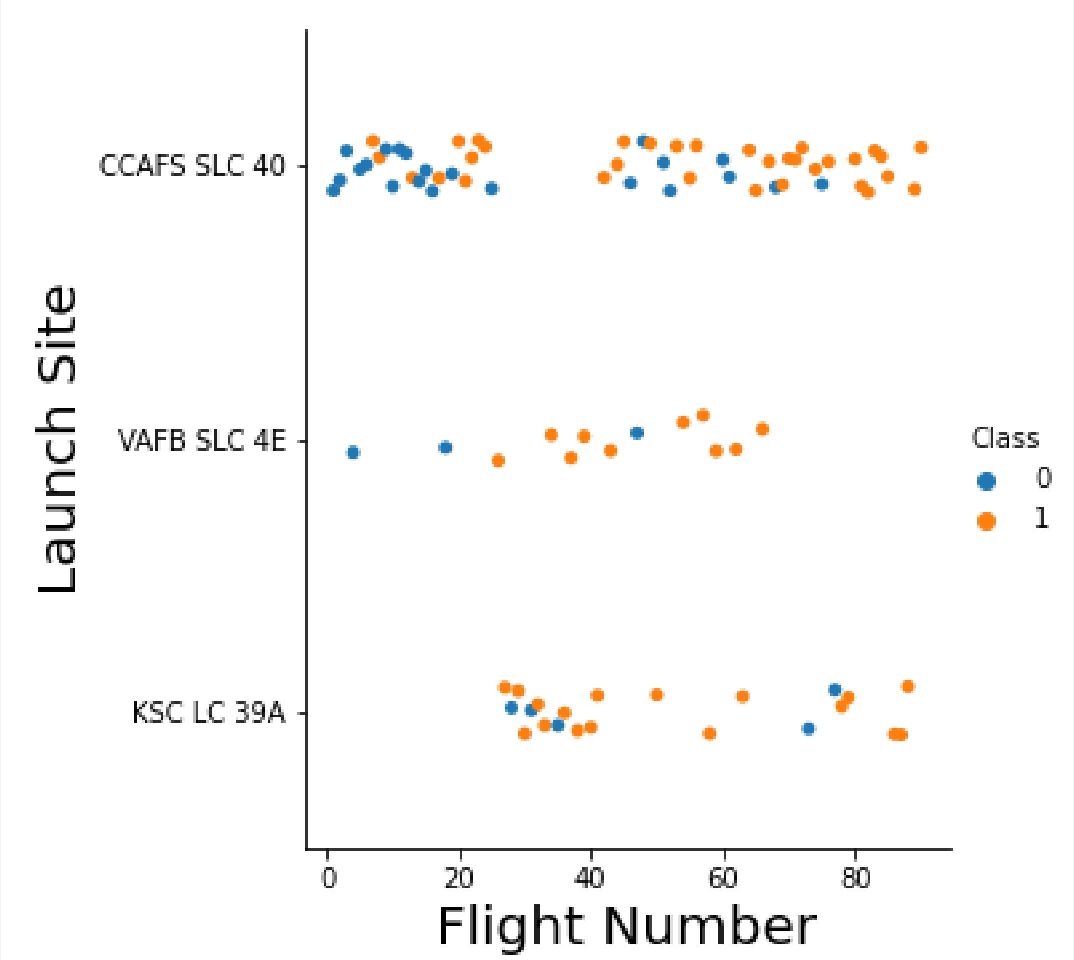
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

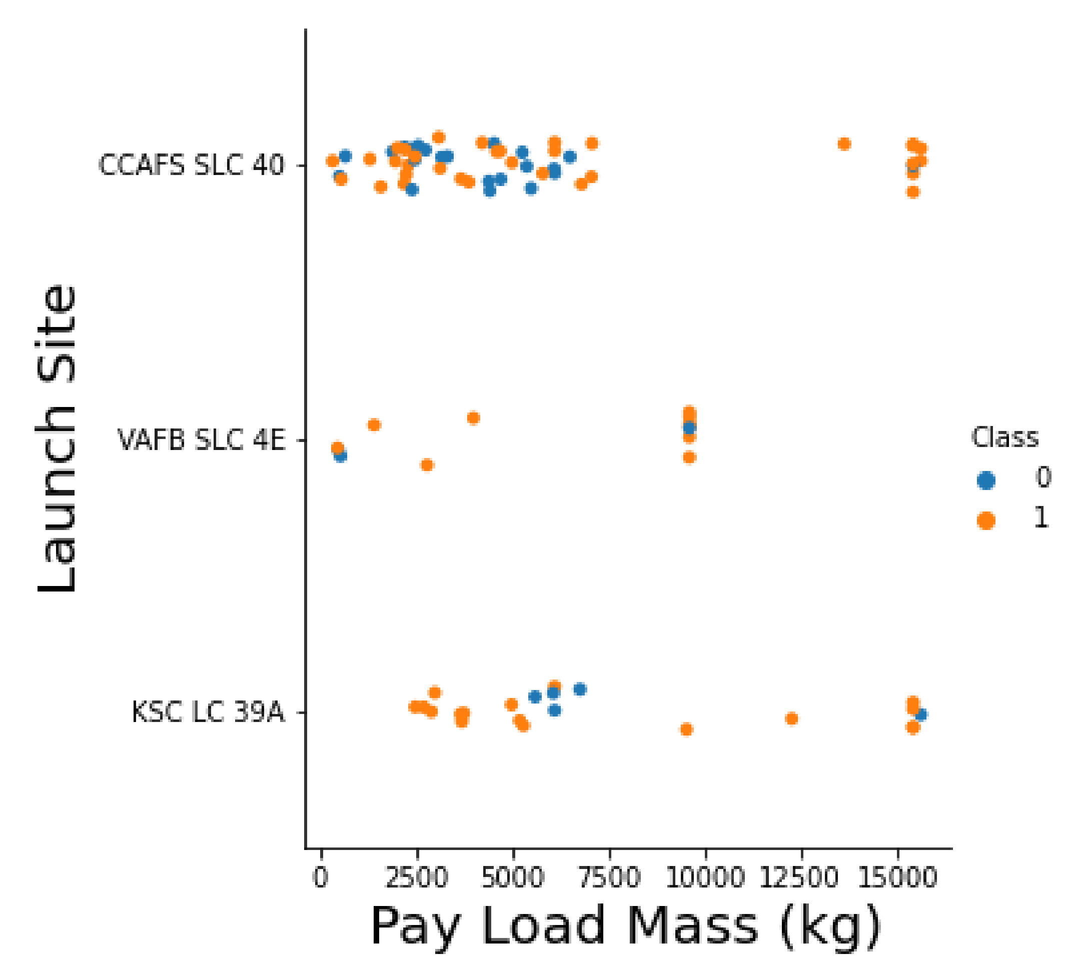
Flight Number vs. Launch Site

- CCAFS SLC 40 most used launch site
- VAFB SLC 4E least used launch site
- First flights fail more often
- Last flights succeed more
- Important success increase in CCAFS SLC 40



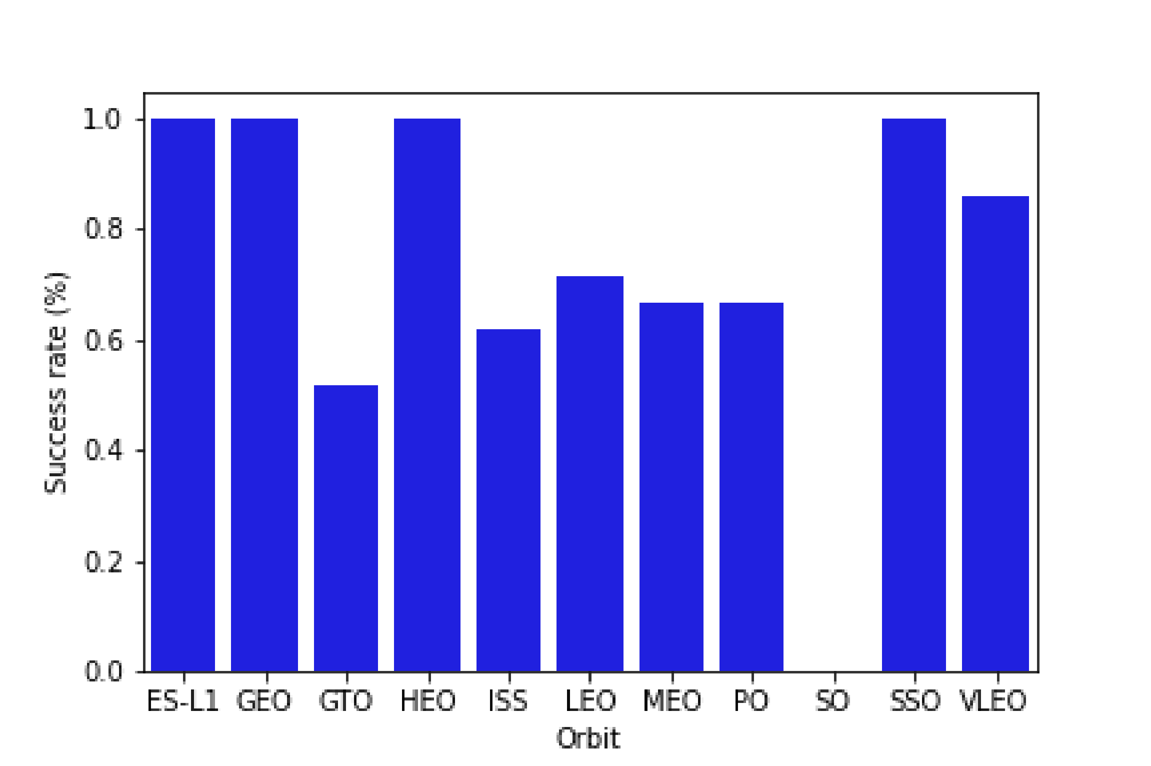
Payload vs. Launch Site

- CCAFS SLC 40 performs better with high payload mass
- VAFB SLC 4E never has a mass above 10000



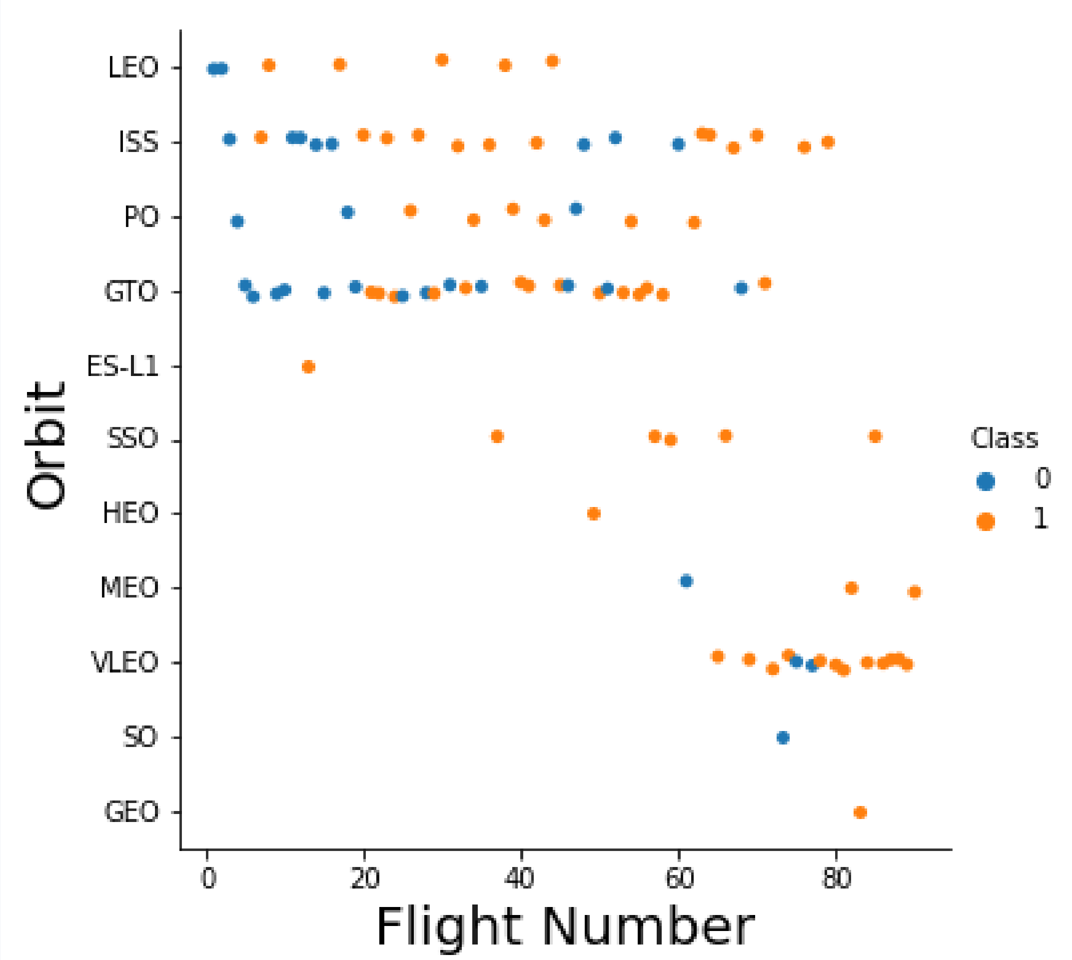
Success Rate vs. Orbit Type

- Several Orbit types have 100% success rate
- SO 0% has 0% success rate
- Small sample size in some Orbits may lead to the high/low success rates



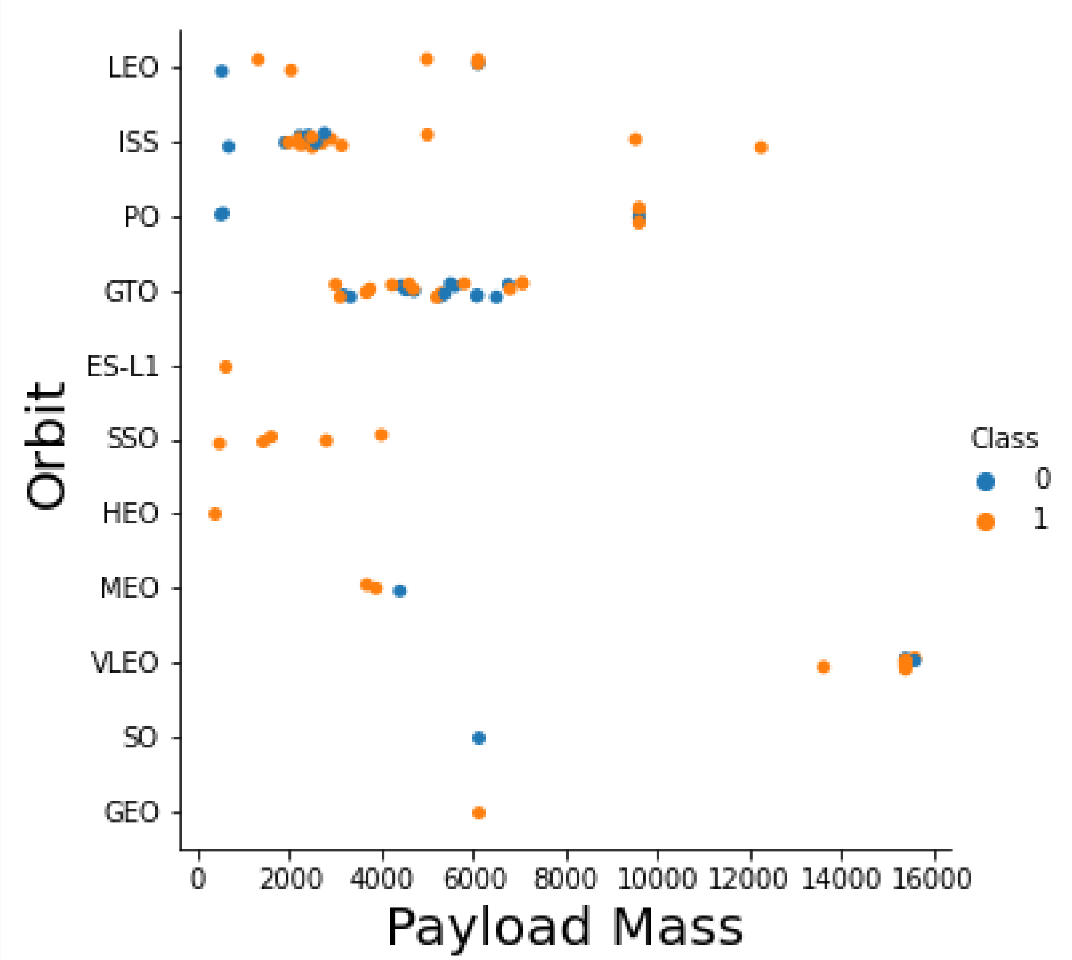
Flight Number vs. Orbit Type

- Over time there is a shift from some orbit types to others
- Again, first flights perform worse



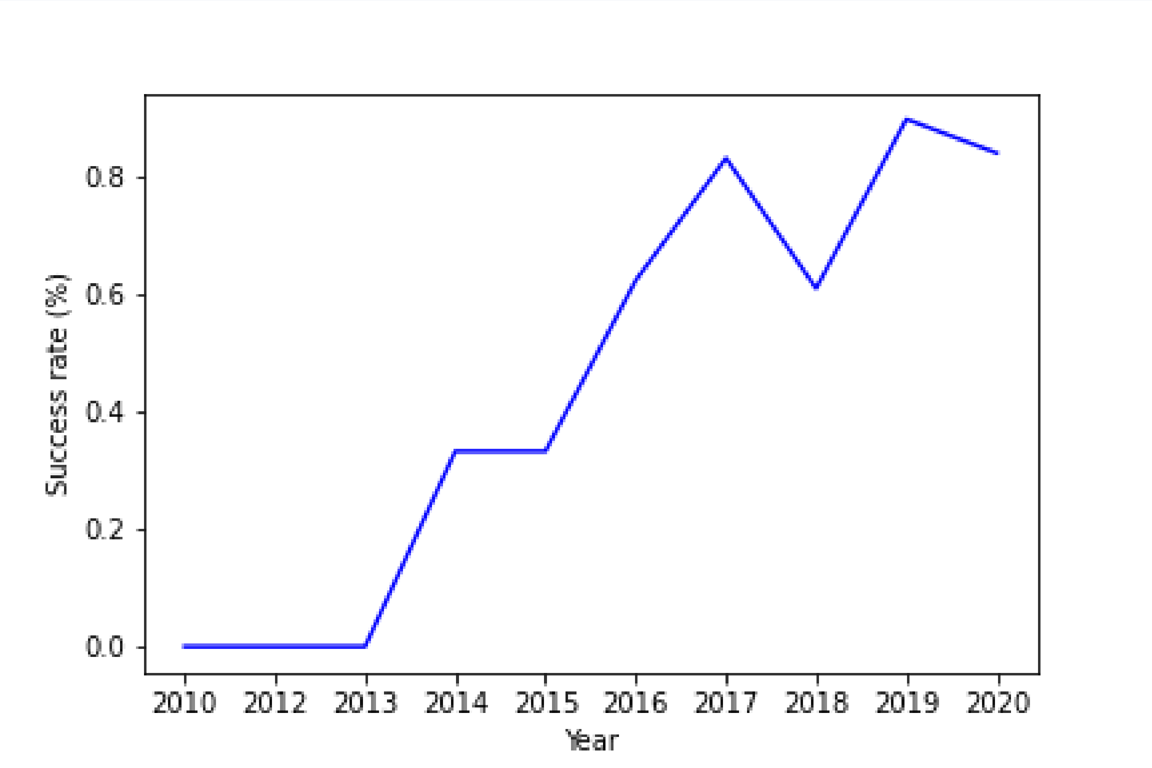
Payload vs. Orbit Type

- Some orbit types have more often high payloads than others



Launch Success Yearly Trend

- Big increase in the success rate from 2010 to 2017
- After that the success rate did not increase significantly and in 2018 was lower than in 2017



All Launch Site Names

- Find the names of the unique launch sites

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

- 4 sites, already seen

Launch Site Names Begin with 'CCA'

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

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	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

- This is an example of 5 records launches from CCAFS LC-40
- There are more than 5 records

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

SUM("PAYLOAD_MASS__KG_")
45596

- That’s a lot of Kg!

Average Payload Mass by F9 v1.1

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

AVG("PAYLOAD_MASS__KG_")
2534.6666666666665

- The average is around of 2500Kg

First Successful Ground Landing Date

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

MIN("DATE")
01-05-2017

- Until 2017 there was no successful landing outcome on ground pad

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

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

- Four different types

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

SUCCESS	FAILURE
100	1

- Very high success rate!

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

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

- Many different boosters can carry the maximum payload mass

2015 Launch Records

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

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

- These failed

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

- 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, in descending order

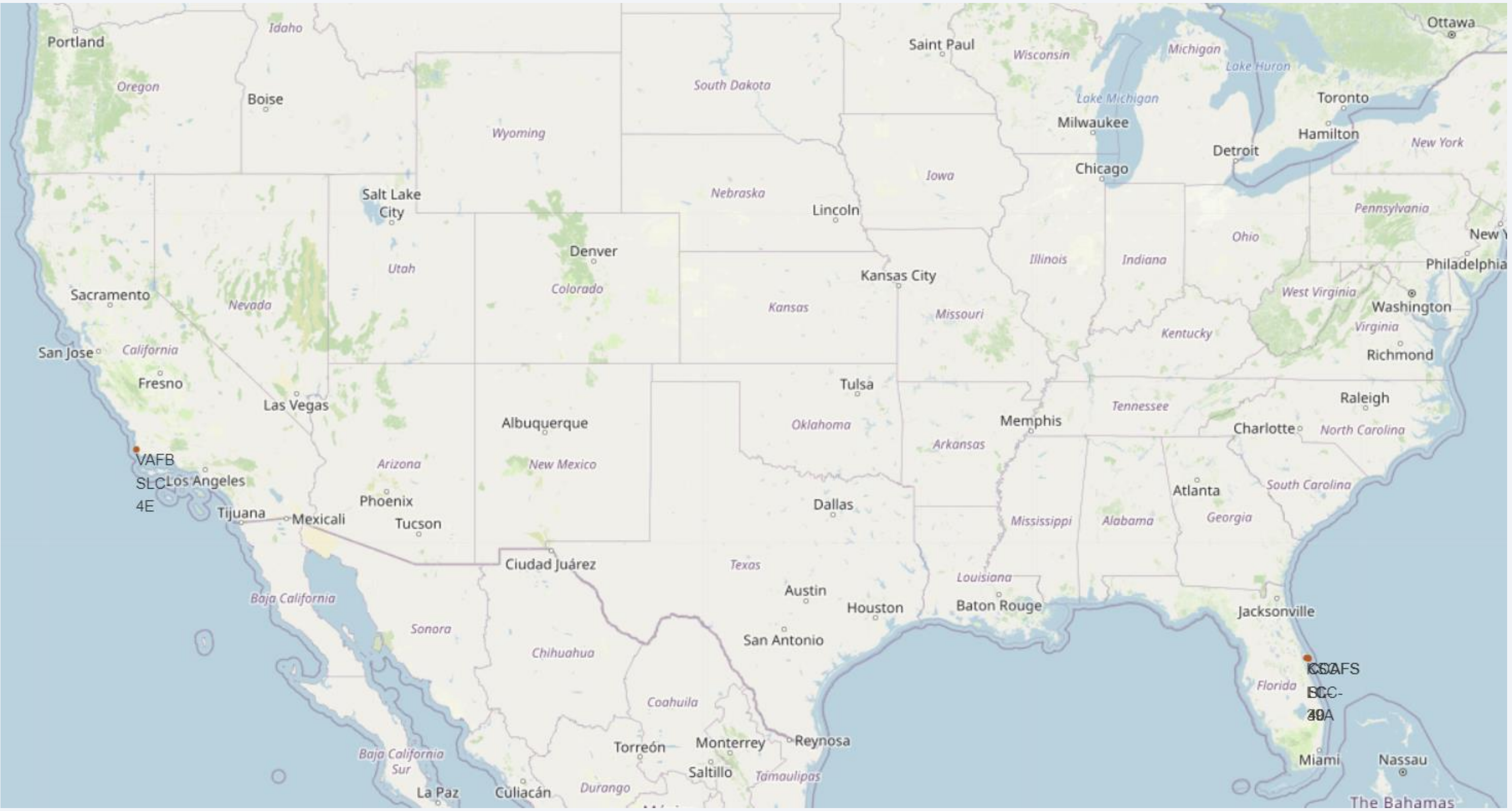
Landing _Outcome	COUNT("LANDING _OUTCOME")
Success	20
Success (drone ship)	8
Success (ground pad)	6

Section 3

Launch Sites Proximities Analysis

Location Launch sites

- All launch sites are in California or Florida
- Which are closer to the equator and closer to the sea

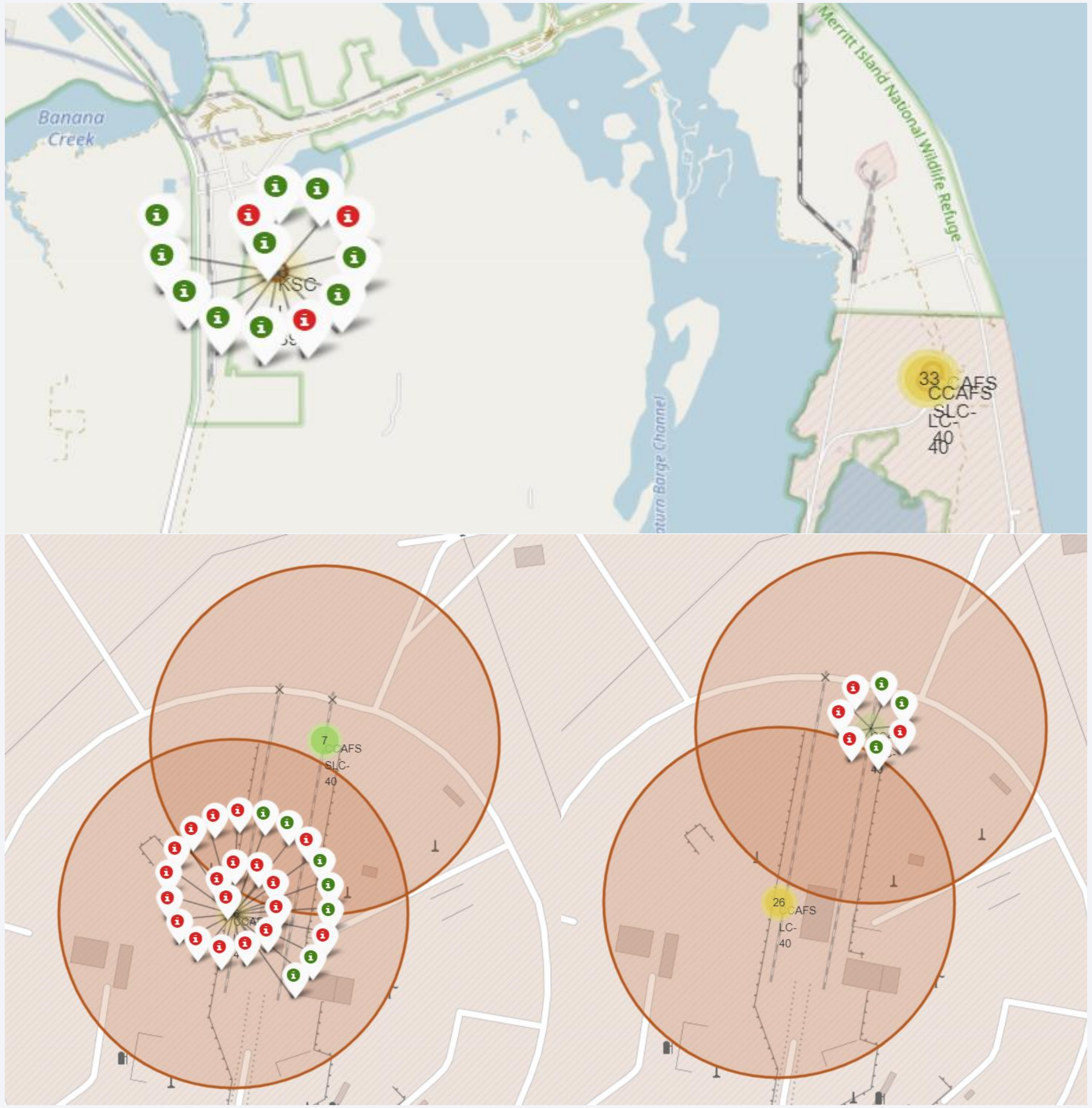


Launch sites success/failure - California

- Launch site in California
- High failure rate

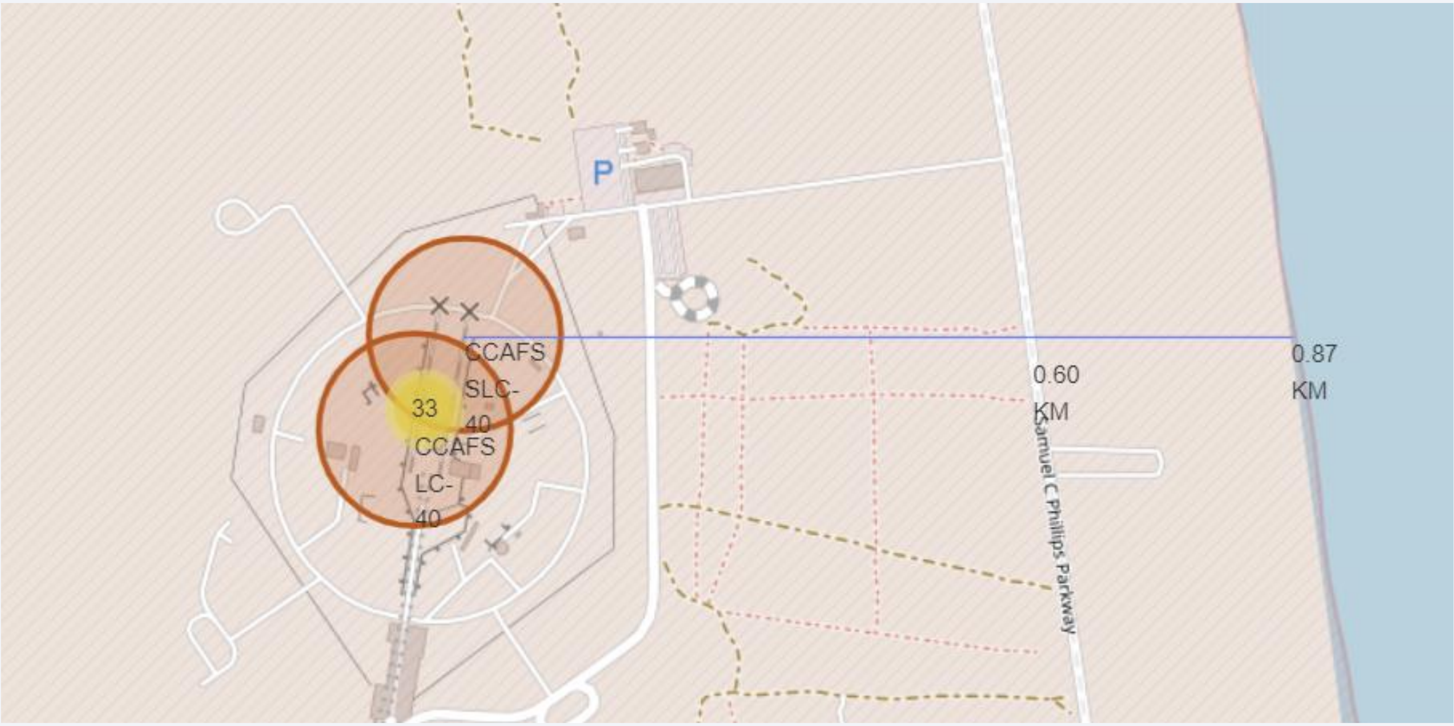
Launch sites success/failure - Florida

- The map on top shows the 3 launch sites in Florida. Plus the success/failures in one of the
- The bottom maps show the success/failures in the other 2 launch sites
- The launch site on the West has higher success rate than the others



Launch site proximites

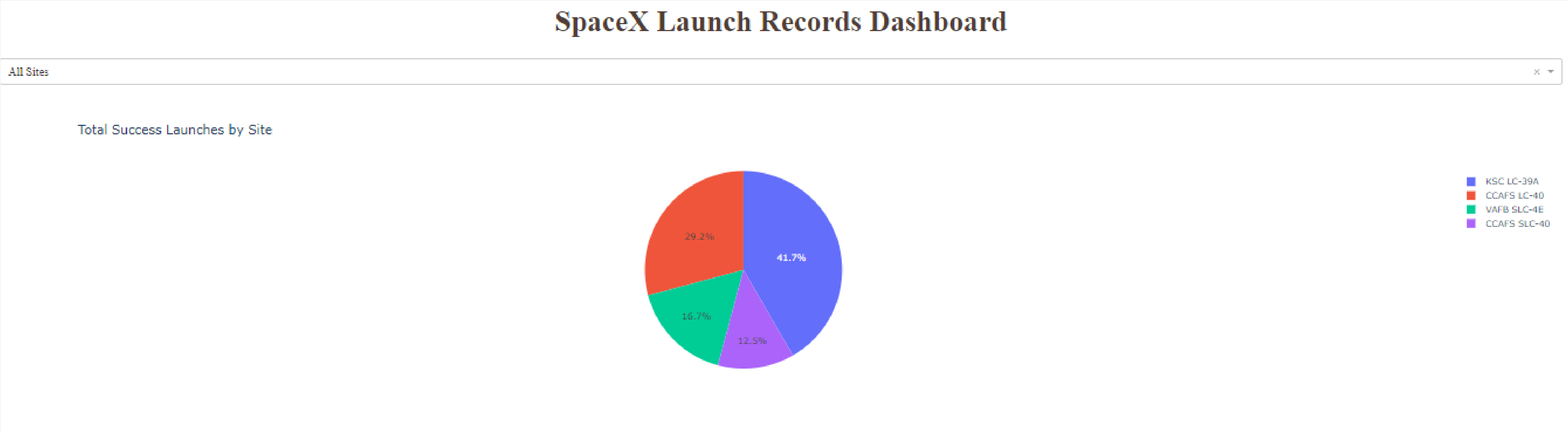
- The launch site CCAFS SLC-40 is at
  - 0.6km from a road
  - 0.87km from the sea



Section 4

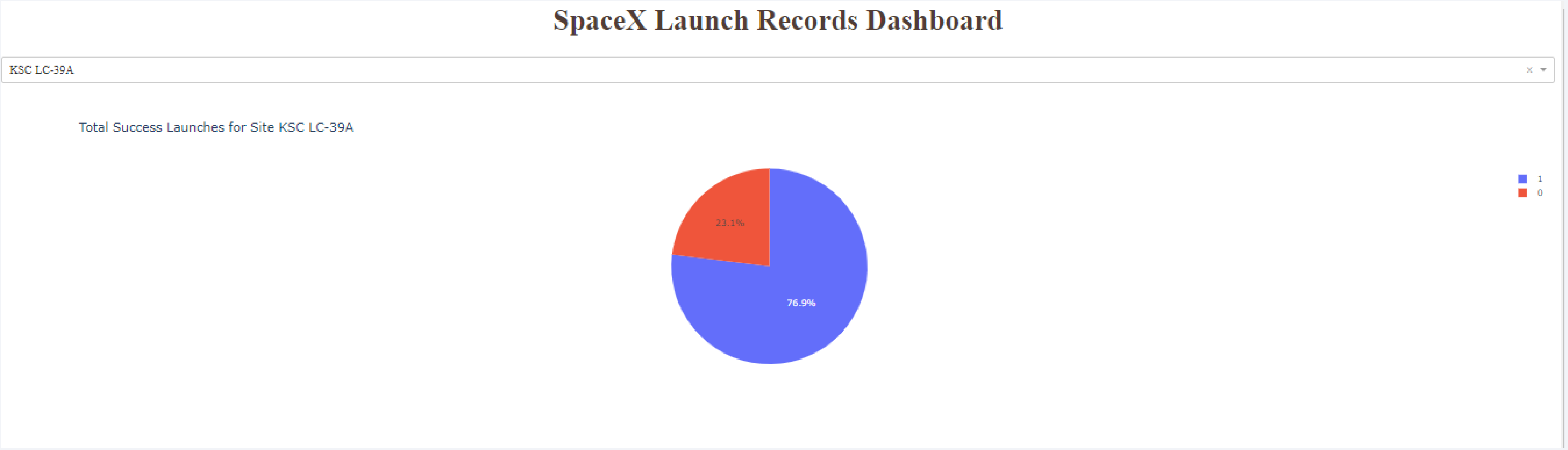
Build a Dashboard with Plotly Dash

Dashboard: Successful outcomes by site



- KSC-LC-39A has the highest number of successful missions 41.7%

Dashboard: Site with highest success rate



- KSC-LC-39A also has the highest success rate

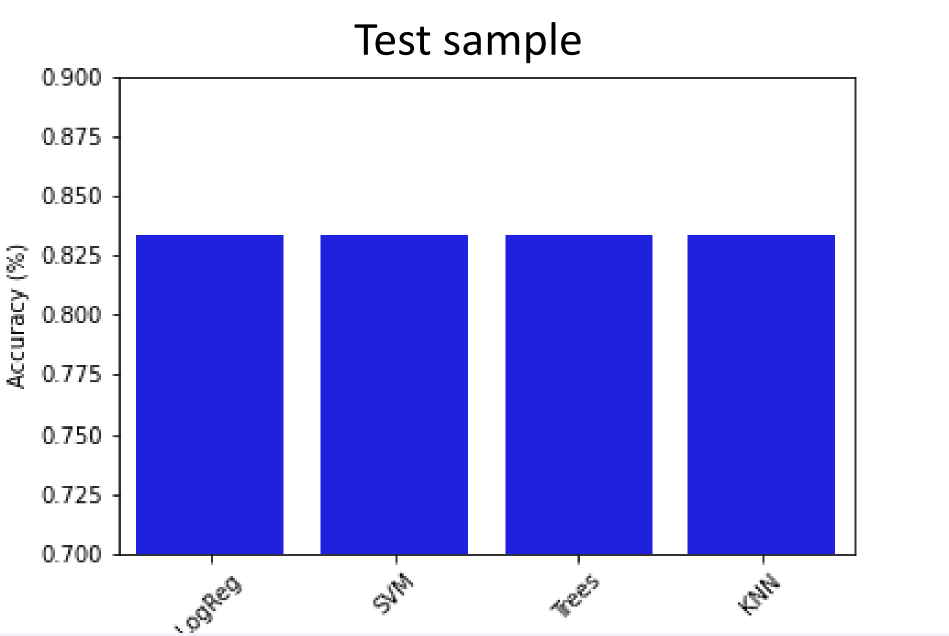
Dashboard: Correlation Payload and success

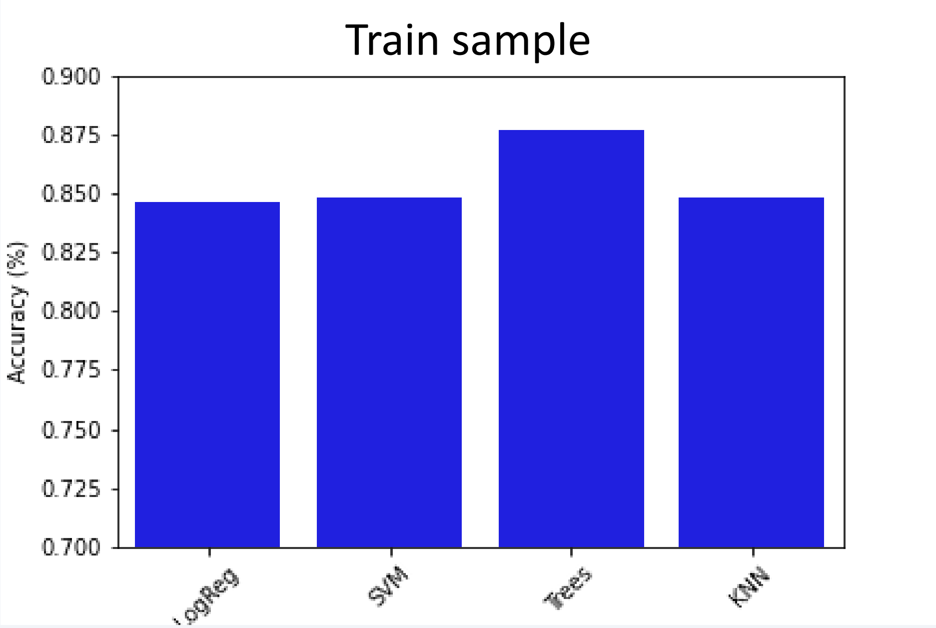
- Correlation between payload and success in KSC LC-39A
- Missions with higher payload are more likely to fail

Section 5

Predictive Analysis (Classification)

Classification Accuracy

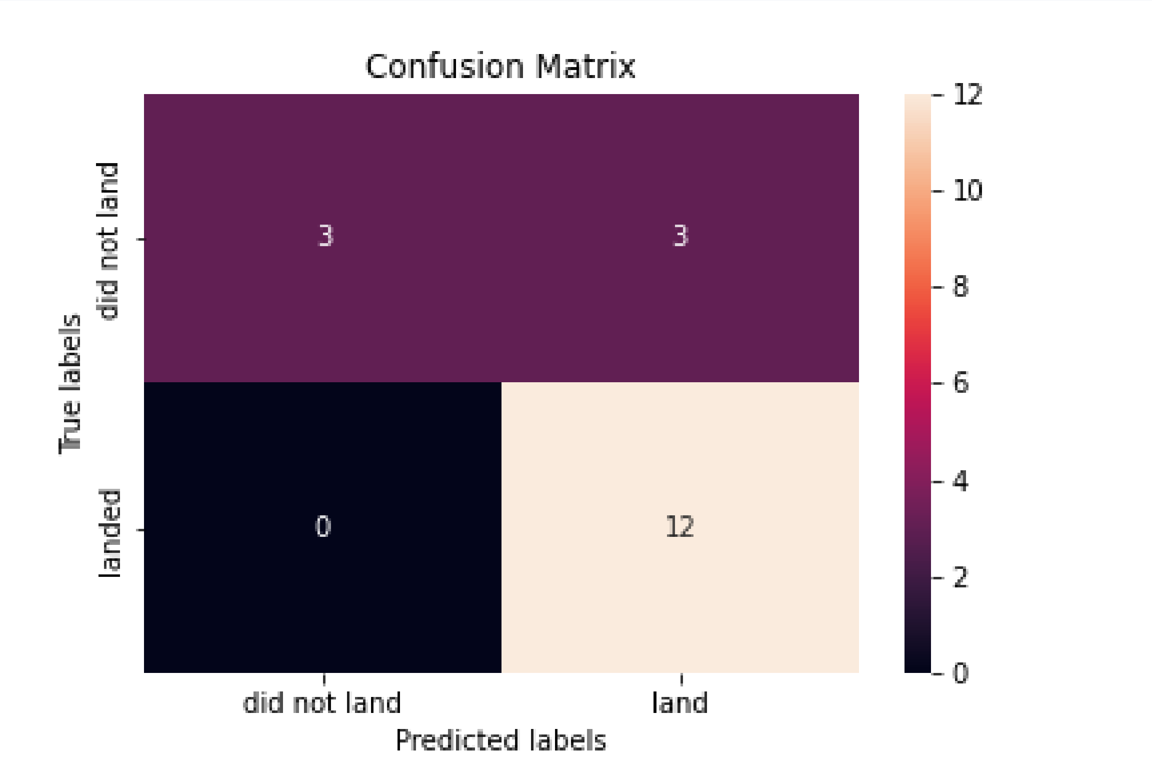




- Same accuracy in the test sample
- Decision trees have higher accuracy in the train sample

Confusion Matrix

- All models have the same confusion matrix



Conclusions

- The four ML models perform almost equally in predicting the success of a mission
- Decision Trees is slightly better at prediction

Thank you!