IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Sandino Jardim
2022-03-19

Outline

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

Executive Summary

- Summary of methodologies
  - Data source:
    - SPACEX REST API
    - SPACEX Wikipedia page (w/ web scrapping)
  - Data Analysis:
    - Visualization tools (Folium, Plotly)
    - SQL
    - Machine Learning
- Summary of all results
  - Effect of payload mass and orbit to the successful landing
  - Grasp of the role of payload mass and booster version category in the landing outcomes

Introduction

- Project background and context
  - Collection, preparation and analysis of data to find the effect of each feature on the outcome of landing
- Problems you want to find answers
  - What are the promising first stages of a new launch that will let to successfully lands
  - We need to know what are the factors the affects the success of first stage landing

Section 1

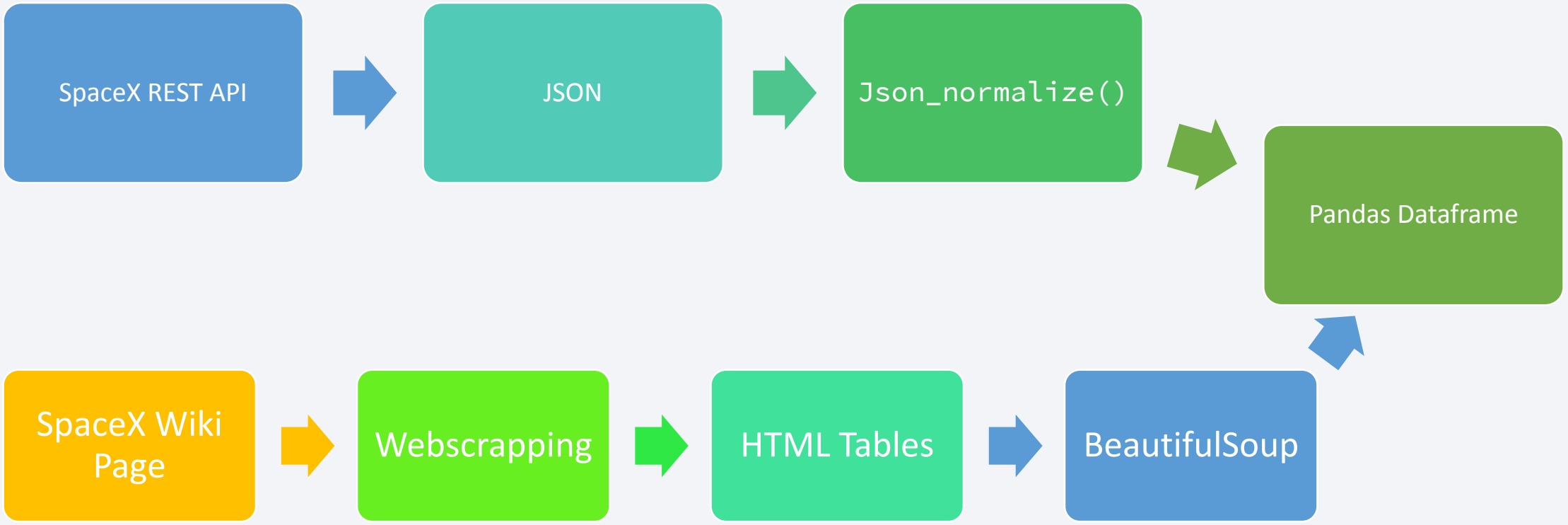
Methodology

Methodology

Executive Summary

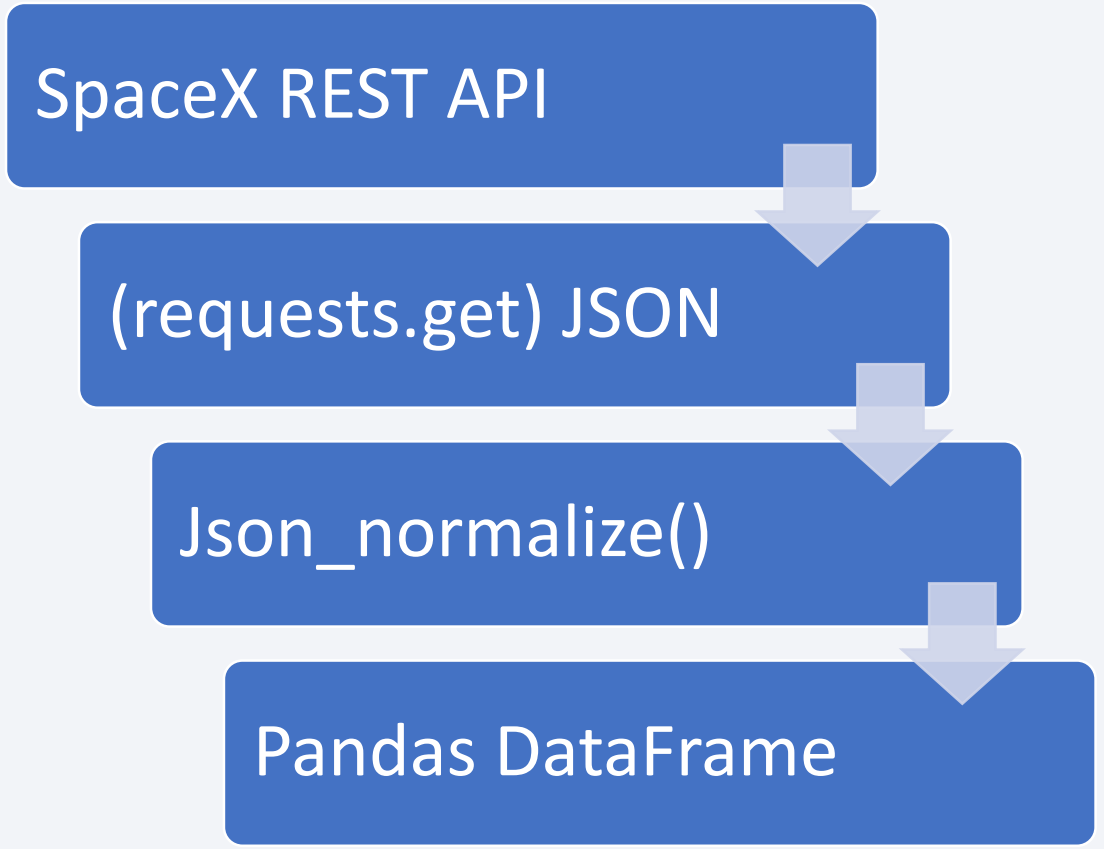
- Data collection methodology:
  - Data is collected in JSON format from SPACEX REST API and from HTML tables from Wikipedia Page of SPACEX
- Perform data wrangling
  - Filtering Falcon 9 launches, replacing NULL values with payload mass mean and conversion of categorical columns into numerical values
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Standardization, split between training and testing set, GridSearch to find best hyperparameters, testing of models and output of confusion matrix

Data Collection



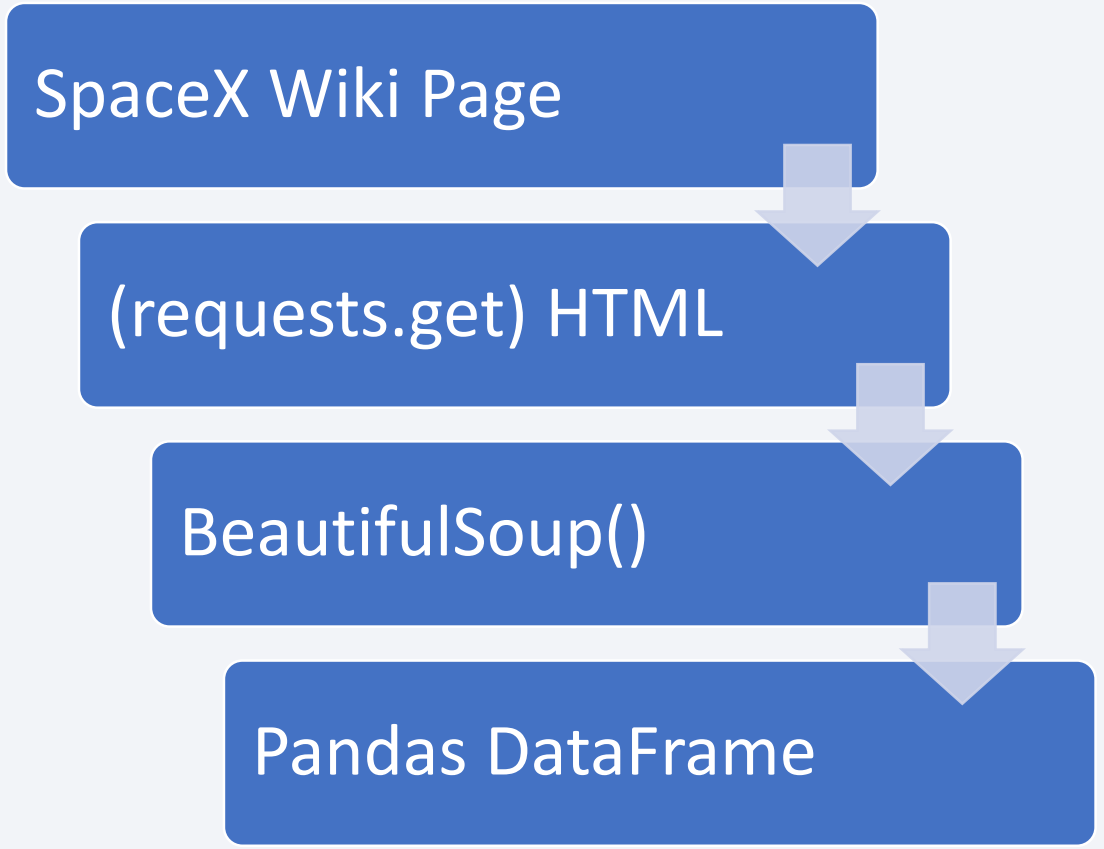
Data Collection – SpaceX API

- GitHub URL:
  - Data Collection API



Data Collection - Scraping

- GitHub URL:
  - Data Collection Web Scrapping



Data Wrangling

- GitHub URL:
  - Data Wrangling

Filtering

Cleaning unnecessary data (i.e., falcon9 launches)

Replacement

Checking for NULL values, replacing for mean values

Categorization

Conversion of categorical variables into numerical ones

EDA with Data Visualization

- GitHub URL:
  - EDA w/ Data Visualisation

Line plot

Year x Success Rate

Bar chart

Orbit x Success Rate

Scatter plot

To find relationship between variables

EDA with SQL

- Listing:
  - Launch sites, outcome counts, booster versions and first date of successful landing
- Calculating:
  - Total payload mass, average payload mass, max payloud mass
- Ranking
  - Counting of different outcomes
- GitHub URL:
  - EDA with SQL

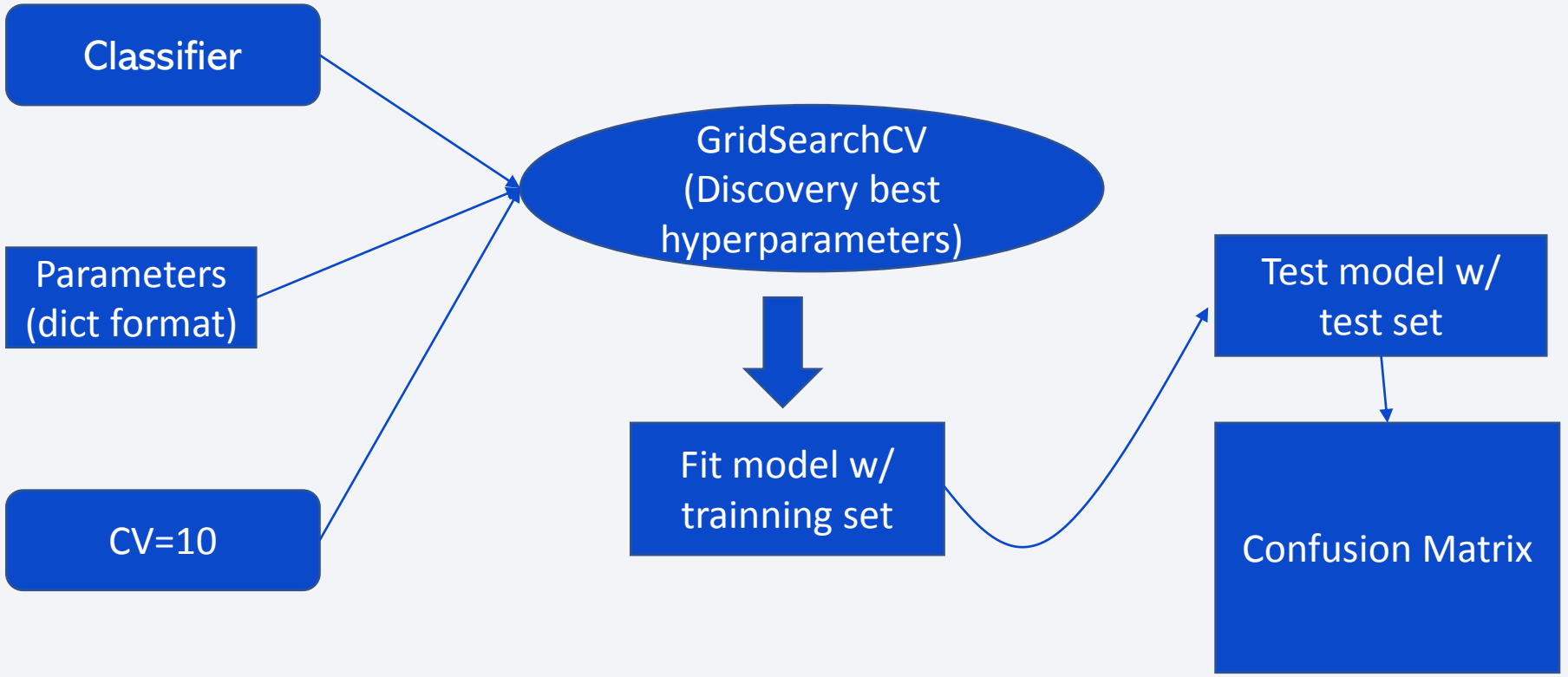
Build an Interactive Map with Folium

- Creation of map objects in a Folium map:
  - Circles: location with a popup
  - Markers: name of the location
  - Marker cluster: outcomes at the site
  - Mouse position: display lat/long points on the map
  - Distance marker: distance indication
  - Lines: connection of closing sites
- GitHub URL: Folium Map

Build a Dashboard with Plotly Dash

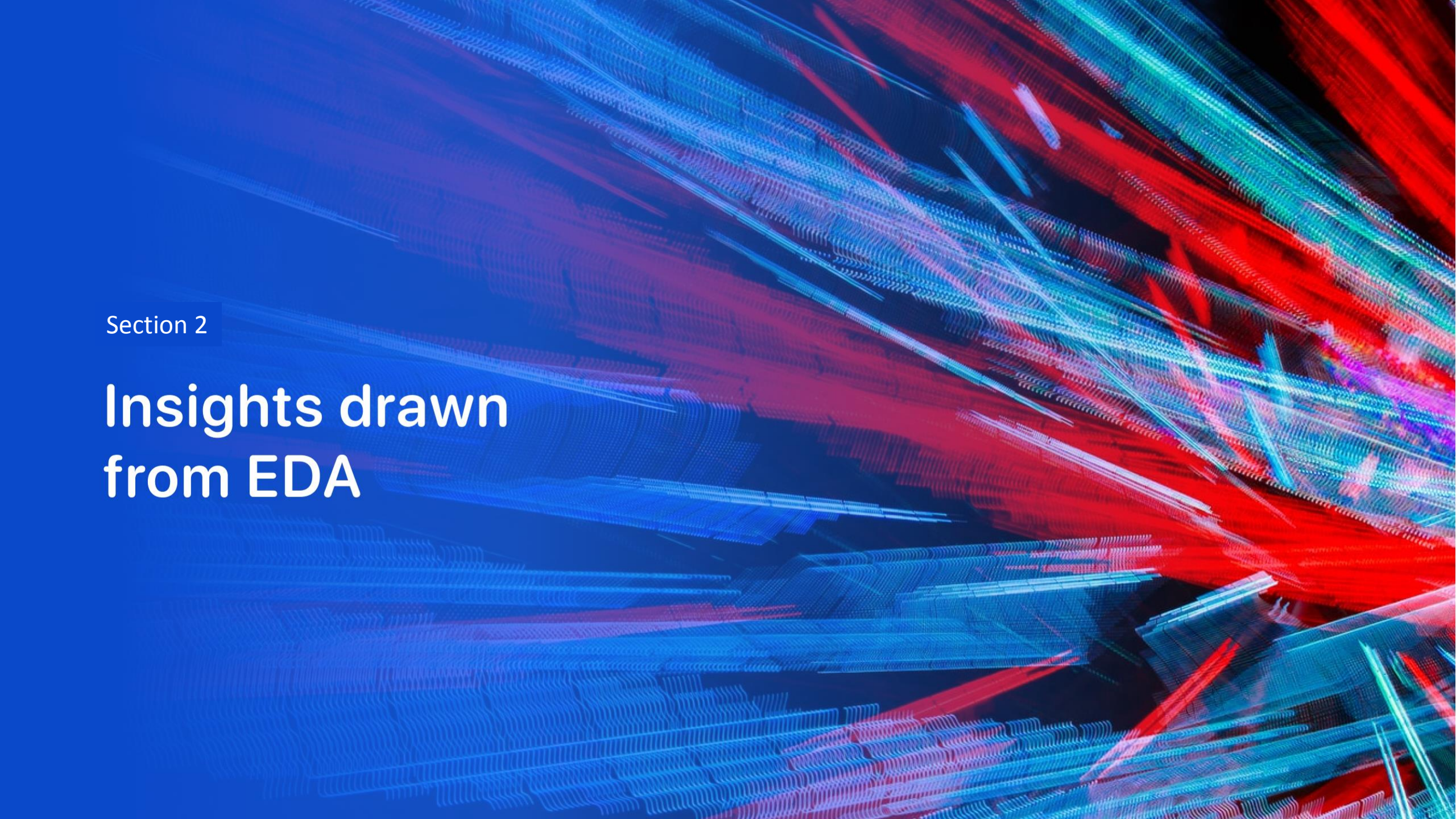
- Graphs and charts used:
  - Scatter plot – payload mass X success rate X booster version categories
  - Pie chart – success rate X launch site
- Interactions:
  - Dropdown – choose launch sites
  - Range slider – set range of payload mass
- GitHub URL: Dashboard

Predictive Analysis (Classification)

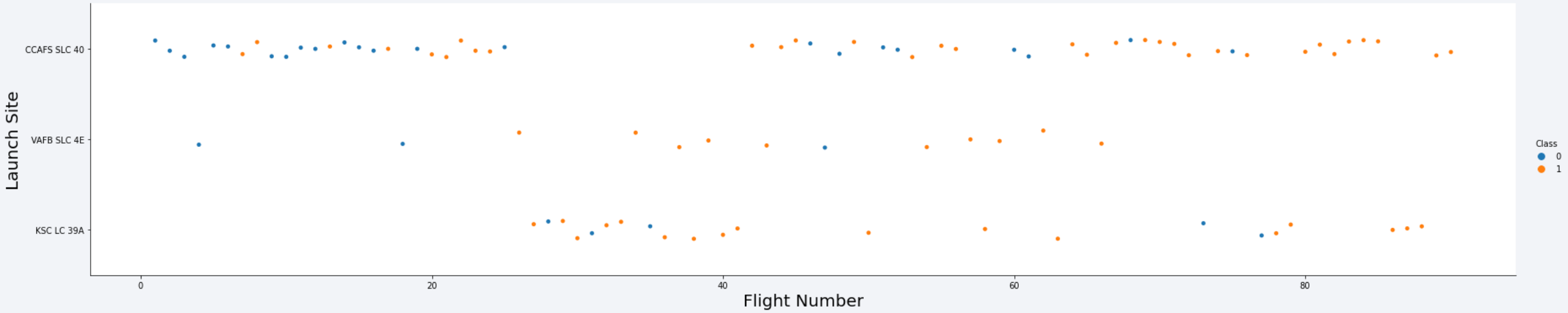


Results

- Payload mass has impact in the landing outcome
  - With the orbit and booster versions
- Folium map displays proximities to coastlines and high/railways
- Predictive analysis presents accuracy of 83.3% to predict the outcome

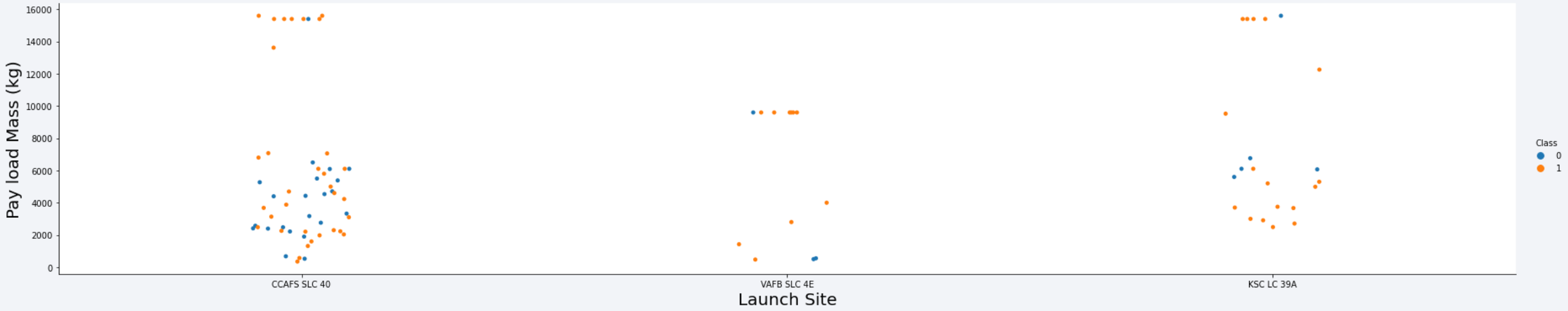

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



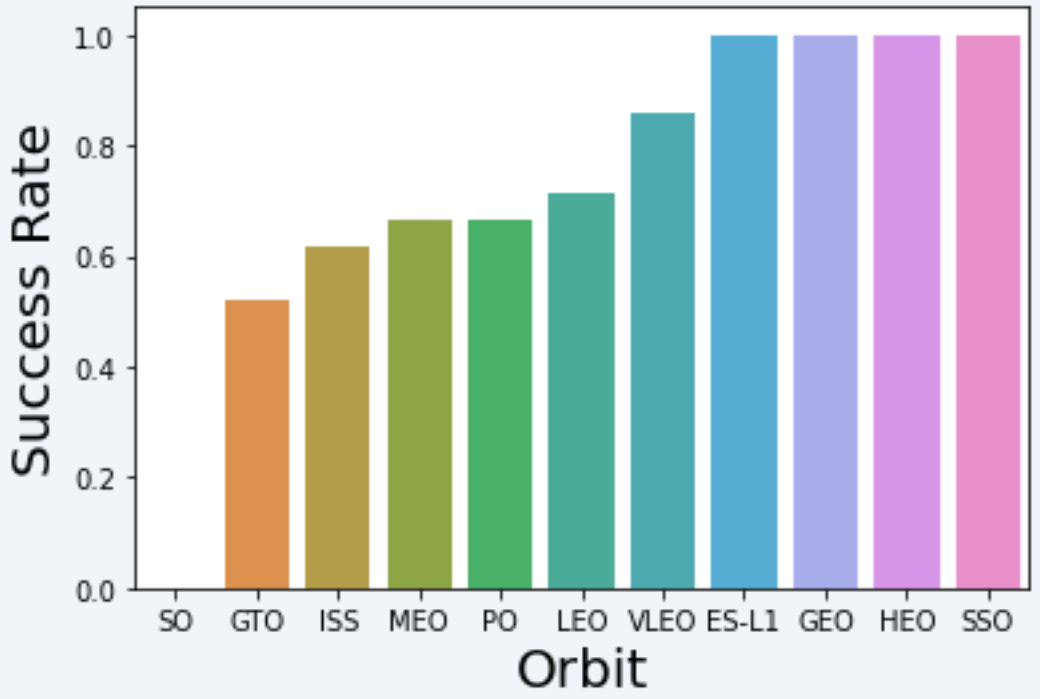
- CCAFS SLC 40 – Successful launches increase after FN #40
- VAFB – Less number of insuccess
- KSC – Best success rate

Payload vs. Launch Site



- CCAFS SLC 40 – Payload mass > 12000 = high success rate
- VAFB – Payload mass was irrelevant on success rate
- KSC – Payload mass > 8000 = high success rate

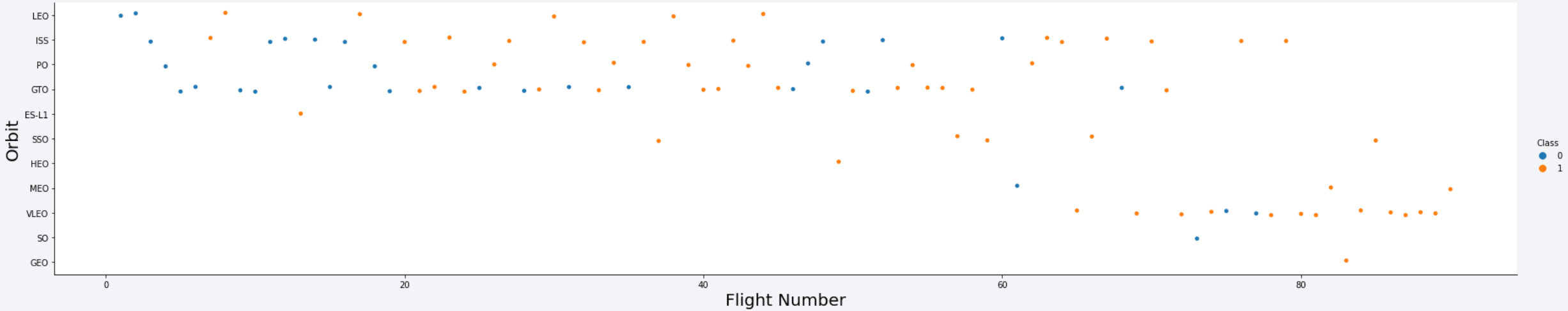
Success Rate vs. Orbit Type



- ES-L1
- GEO
- HEO
- SSO

100% of success rate

Flight Number vs. Orbit Type

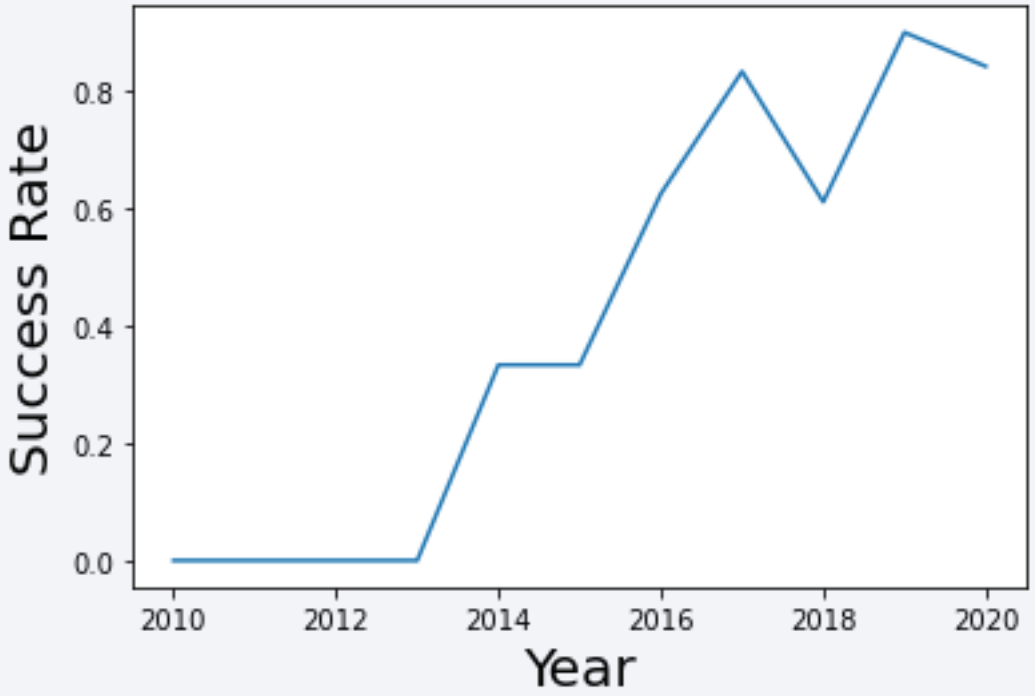


- Above 80, no failures in any orbit type

Payload vs. Orbit Type

- Above 7000 kg, only 1 launch failed in any orbit type
- Zero failures in SSO for any PL

Launch Success Yearly Trend



- > 60% of Success Rate after 2016
- Decrease in 2018
- Recovered so on

All Launch Site Names

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

```
In [11]: %sql select distinct(LAUNCH_SITE) from SPACEXDATASET
```

Out[11]:

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

Launch Site Names Begin with 'CCA'

```
%sql select * from spacex where launch_site like 'CCA%'
```

```
 * ibm_db_sa://cpv48249:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

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
2013-12-03	22:41:00	F9 v1.1	CCAFS LC-40	SES-8	3170	GTO	SES	Success	No attempt

- List of Launch Site names beginning with ‘CCA’

Total Payload Mass

In [13]:
```
%%sql select sum(payload_mass__kg_) as total_payload_mass from SPACEXDATASET
where customer = 'NASA (CRS)'
```

Out[13]:

total_payload_mass
45596

Total Payload Mass of Rockets whose Customer is NASA = 45596 kg

Average Payload Mass by F9 v1.1

In [14]:
```
%%sql select avg(payload_mass__kg_) as avg_payload_mass from SPACEXDATASET
where booster_version = 'F9 v1.1'
```

Out[14]:

avg_payload_mass
2928

Average Payload Mass of Rockets whose Booster version is F9 v1.1 = 2928 kg

First Successful Ground Landing Date

In [15]:
```
%%sql select min(DATE) as first_successful_landing from SPACEXDATASET
where landing__outcome = 'Success (ground pad)'
```

Out[15]:

first_successful_landing
2015-12-22

First successful ground landing date was 2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

In [16]:
```
%%sql select booster_version,landing__outcome from SPACEXDATASET
where (landing__outcome = 'Success (drone ship)') and (payload_mass__kg_ between 4000 and 6000)
```

Out[16]:

booster_version	landing__outcome
F9 FT B1022	Success (drone ship)
F9 FT B1026	Success (drone ship)
F9 FT B1021.2	Success (drone ship)
F9 FT B1031.2	Success (drone ship)

Four boosters had successful landed ond drone ship with a payload mass between 4000 and 6000

Total Number of Successful and Failure Mission Outcomes

In [17]:

```
%%sql select mission_outcome,count(*) as count from SPACEXDATASET
group by mission_outcome
```

Out[17]:

mission_outcome	COUNT
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Total Number of Successful and Failure Mission Outcomes

Boosters Carried Maximum Payload

```
%%sql select booster_version,payload_mass__kg_ from SPACEXDATASET
where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXDATASET)
```

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

Names of the boosters which have carried max payload

2015 Launch Records

```
%%sql select DATE,booster_version,launch_site,landing__outcome from SPACEXDATASET
where (landing__outcome = 'Failure (drone ship)') and (DATE like '2015%')
```

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

All launch tries in 2015 from drone ship had failed

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

In [20]:
```
%%sql select count(landing__outcome) as count,landing__outcome from SPACEXDATASET
where DATE>='2010-06-04' and DATE<='2017-03-20'
group by landing__outcome
order by count desc
```

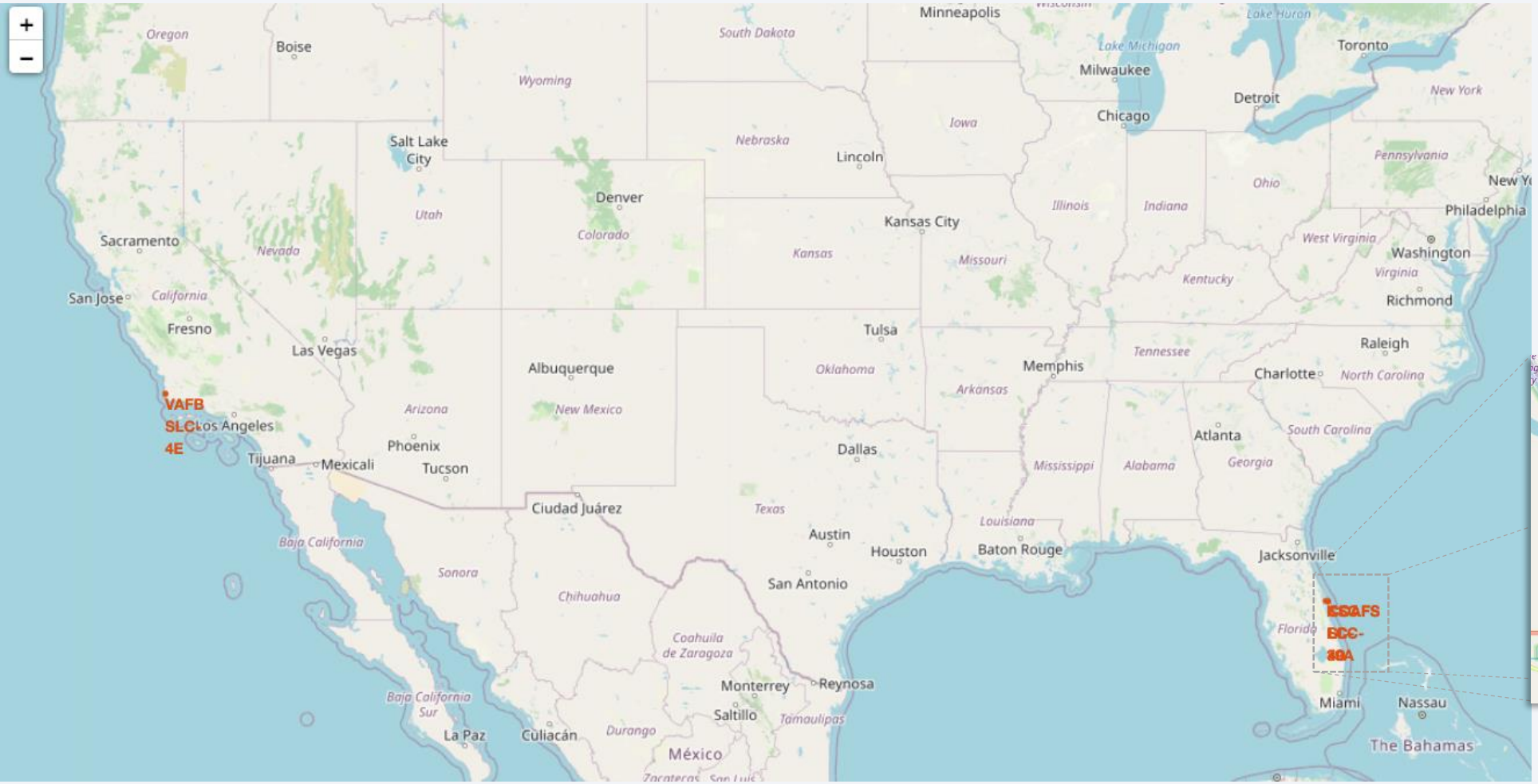
Out[20]:

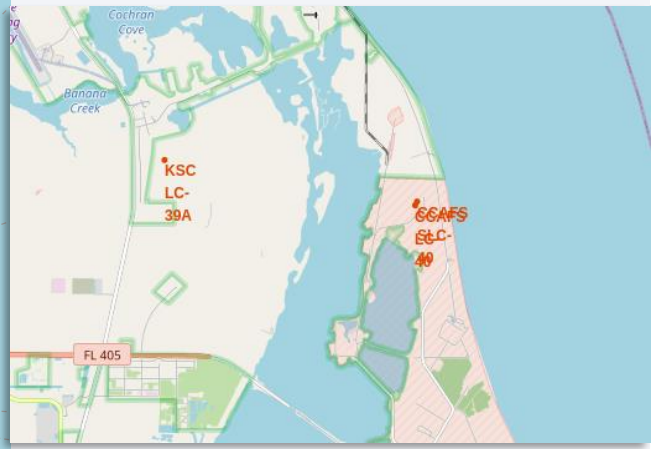
COUNT	landing__outcome
10	No attempt
5	Failure (drone ship)
5	Success (drone ship)
3	Controlled (ocean)
3	Success (ground pad)
2	Failure (parachute)
2	Uncontrolled (ocean)
1	Precluded (drone ship)

Only 8 success launches in this period


Section 3
Launch Sites
Proximities Analysis

Launch site | Locations



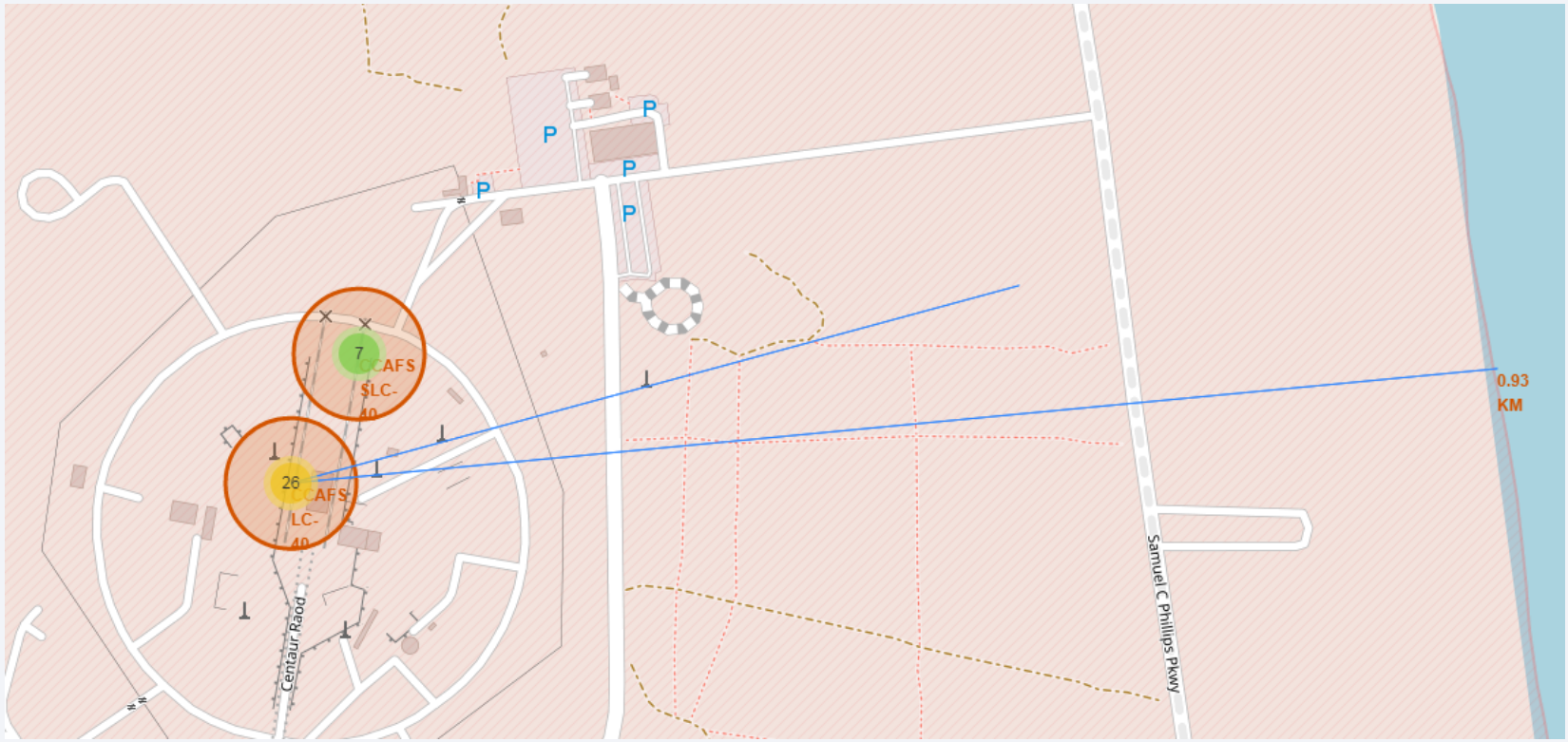


VAFB on western coast while CCAFS/KSC in eastern coast

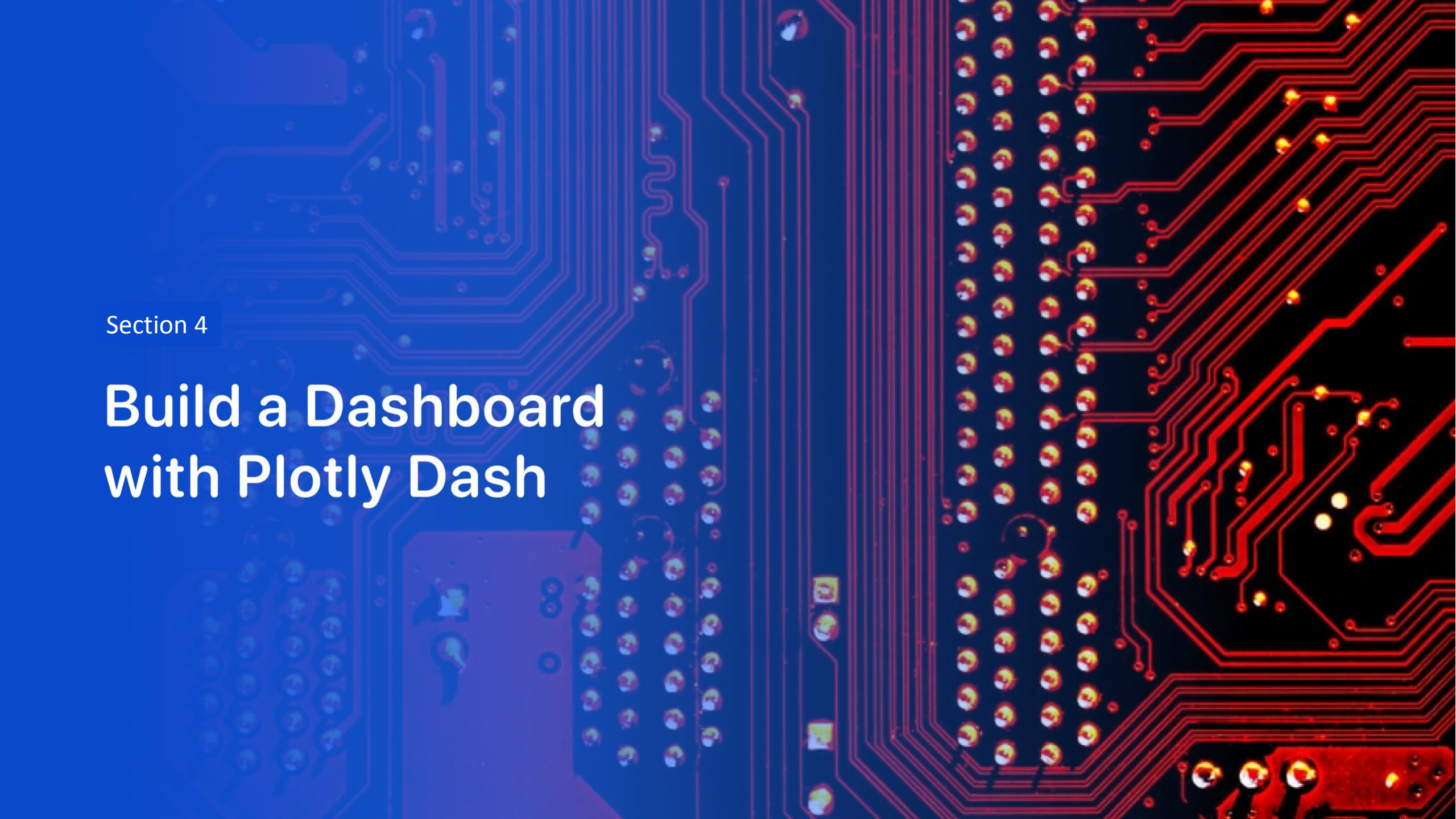
Success/failed launches for each site

- Cape Canaveral Launch Site (CCAFS LC-40) w/ most failed than successful launches

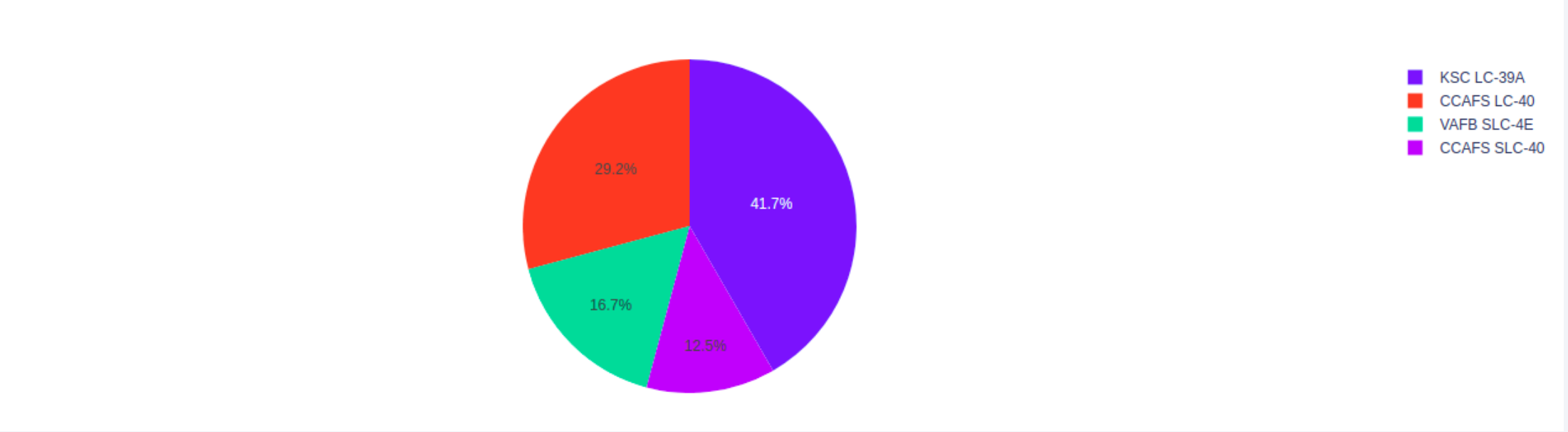
Distance from launch site to coast



- Less than 1km of distance to east coast

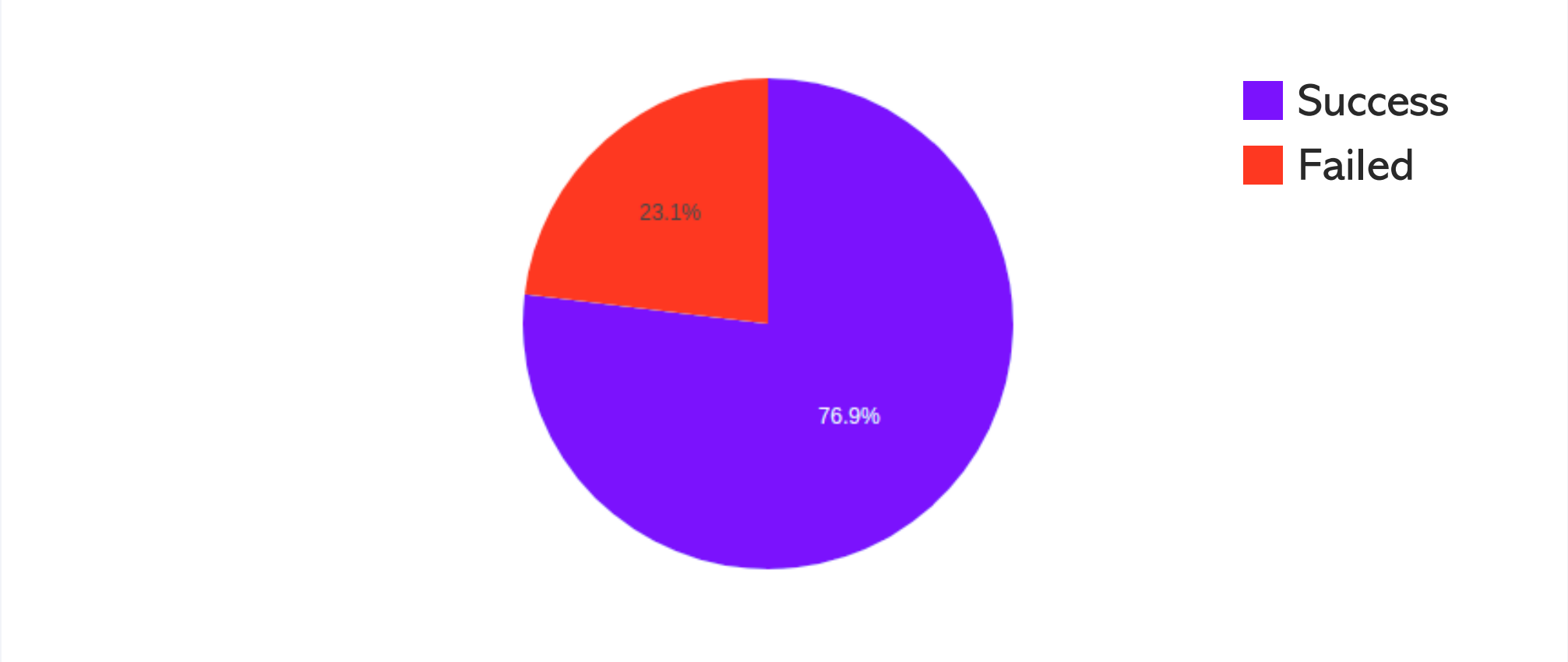

Section 4
Build a Dashboard with Plotly Dash

Launch Site Success Rate



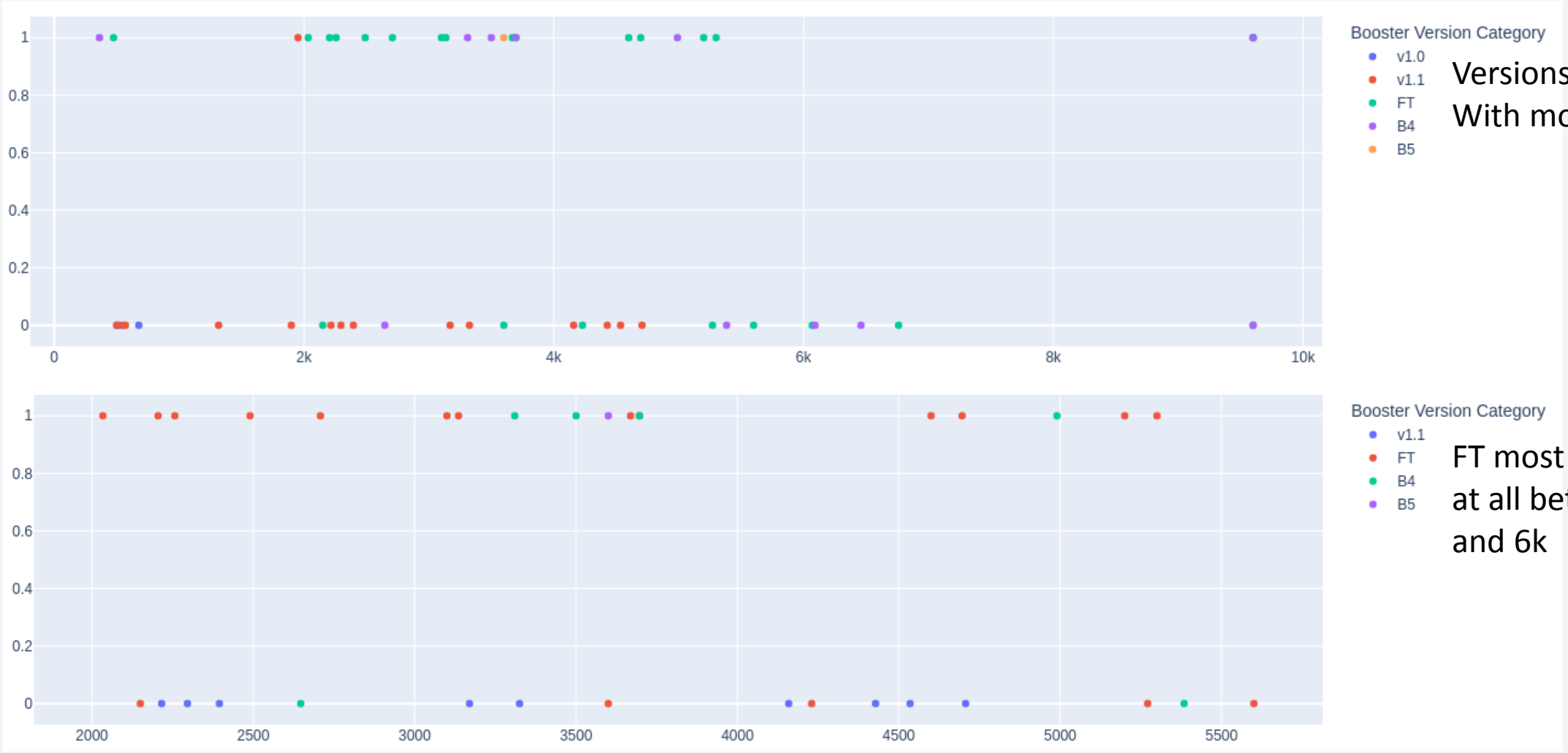
- KSC most successful / Cape Canaveral has the lowest success rate

KSC LC-39A Success Rate



- KSC obtained 10 of 13 successful launch tries

Correlation between payload mass vs success rate



Versions 1.0 and 1.1
With most failures

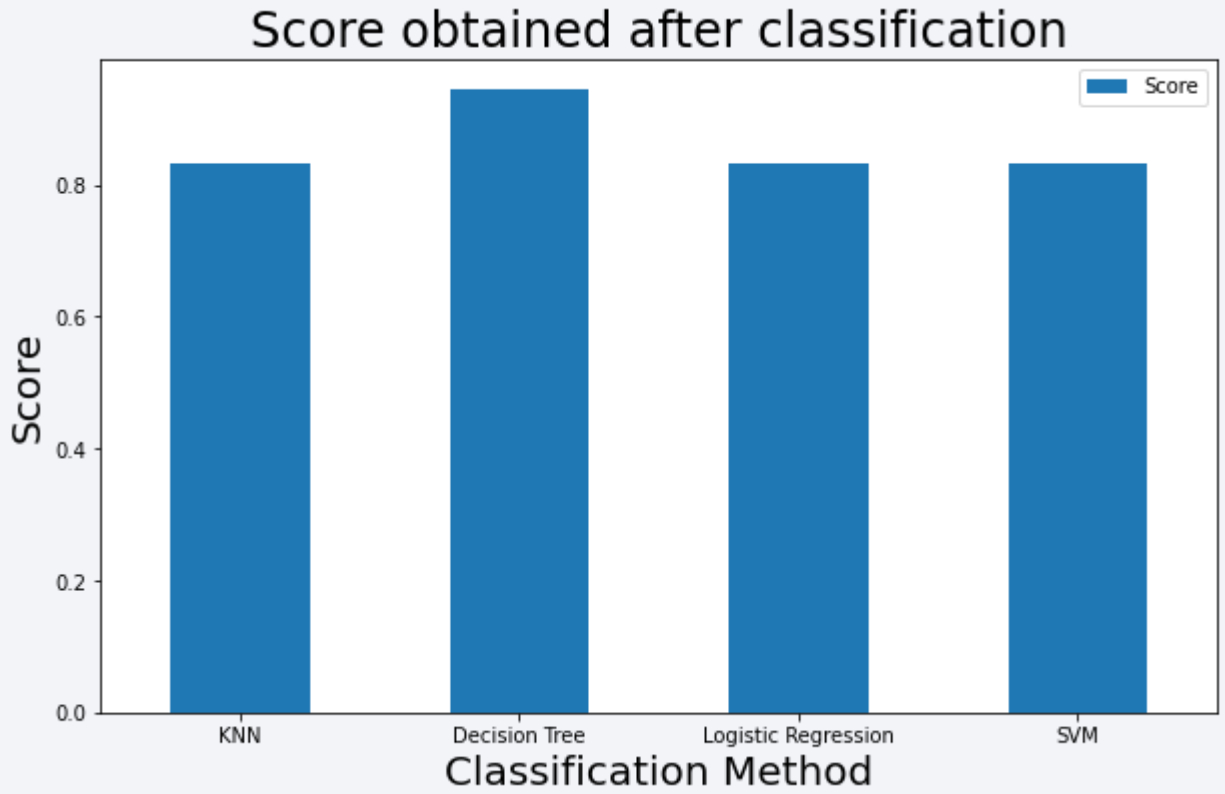
FT most succesful at all between 2k and 6k

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Four models with accuracy above 80%
- Decision Tree with best accuracy (94%)



Confusion Matrix | Decision Tree

- Successful landing predicted with 100% of accuracy
- Failed landings with low accuracy (50%)

Conclusions

- Success rate is proportional to time in years
- KSC LC-39A has the most successful launches
- CCAFS SLC-40 with best success ratio
- Orbits GEO, HEO, SSO, ES-L1 with best success rates
- Payload mass along with booster version and orbit has presents high probability to outcome of landing

Appendix

- Python notebooks:
  - GitHub Repo
- SQL data:
  - SpaceX DataSet
- Web scrapping
  - List of Falcon 9 and Falcon Heavy launches (Wikipedia page)
- SpaceX REST API

Thank you!