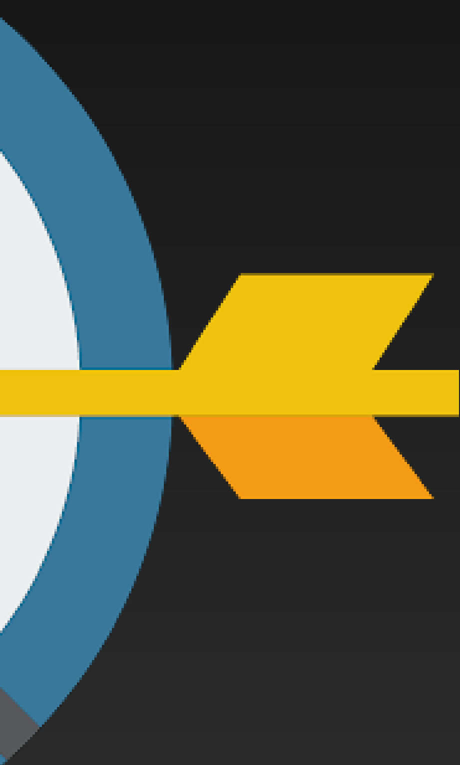
Science Capstone P

- Executive Su
- Introduction
- Methodolog
- Results
- Conclusion



Summary of methodologies

- Data collection
- Data wrangling
- Exploratory Data Analysis with Dat
- Exploratory Data Analysis with SQ
- Building an interactive map with F
- Building a Dashboard with Plotly D
- Predictive analysis (Classification)

Summary of all results

- Exploratory Data Analysis results

Project background and context

SpaceX is the most successful company of th
age, making space travel affordable. The com
9 rocket launches on its website, with a cost
other providers cost upward of 165 million do
savings is because SpaceX can reuse the firs
can determine if the first stage will land, we c
of a launch. Based on public information and
models, we are going to predict if SpaceX wi

Questions to be answered

- How do variables such as payload mass, lau
flights, and orbits affect the success of the fi

- Does the rate of successful landings increas

Data collection methodology:

- Using SpaceX Rest API
- Using Web Scrapping from Wikipedia

Performed data wrangling

- Filtering the data
- Dealing with missing values
- Using One Hot Encoding to prepare the data to

Performed exploratory data analysis (EDA) usin

Performed interactive visual analytics using Fo

Performed predictive analysis using classificati

- Building, tuning and evaluation of classification

Methodology

ocess involved a combination of API requests from
ping data from a table in SpaceX’s Wikipedia entry

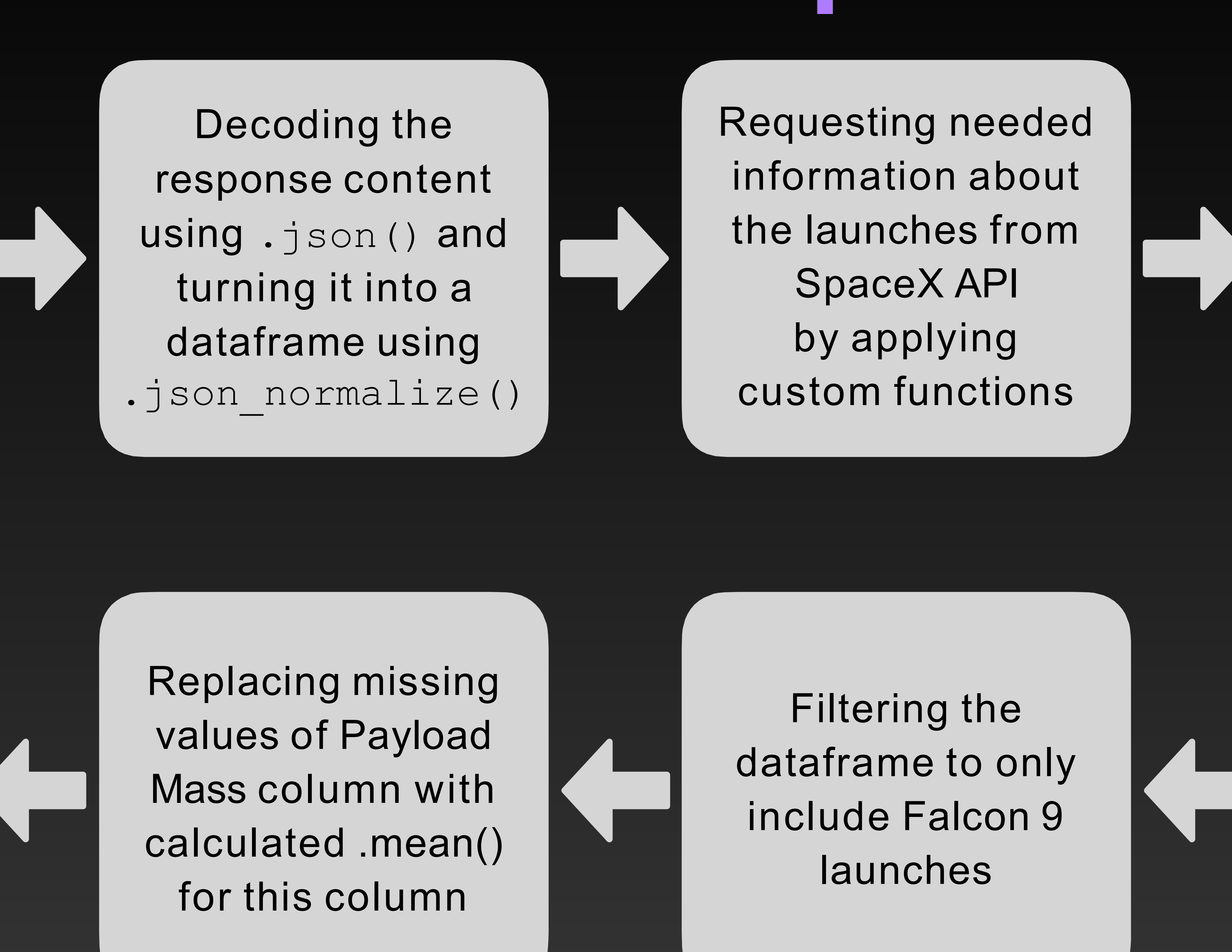
h of these data collection methods in order to get
the launches for a more detailed analysis.

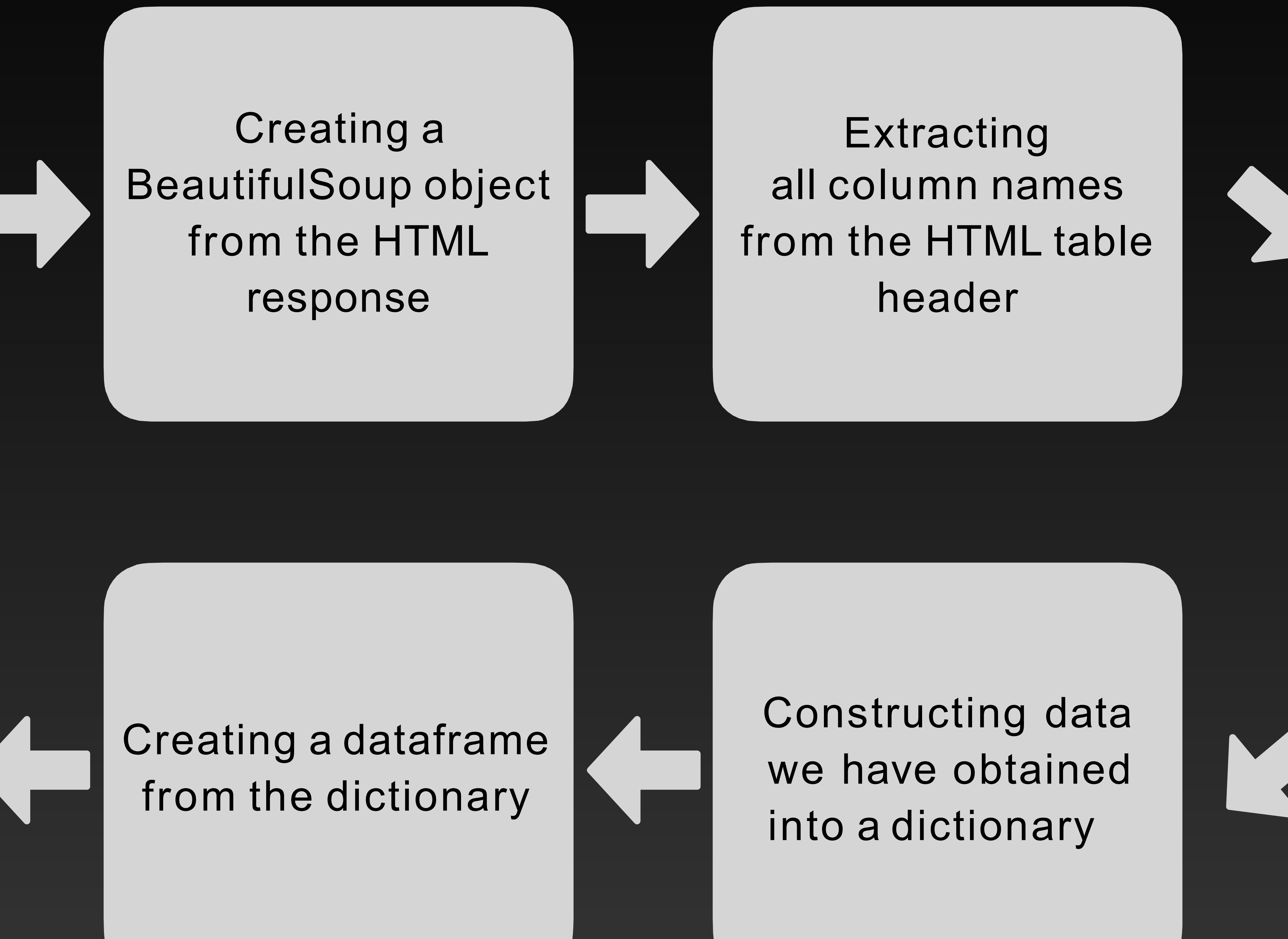
s are obtained by using SpaceX REST API:

ber, Date, BoosterVersion, PayloadMass, Orbit, Lau
Flights, GridFins, Reused, Legs, LandingPad, Block
gitude, Latitude

s are obtained by using Wikipedia Web Scra

Launch site, Payload, PayloadMass, Orbit, Custom


Decoding the response content using .json() and turning it into a dataframe using .json_normalize()
Requesting needed information about the launches from SpaceX API by applying custom functions
Filtering the dataframe to only include Falcon 9 launches
Replacing missing values of Payload Mass column with calculated .mean() for this column


Creating a BeautifulSoup object from the HTML response
Extracting all column names from the HTML table header
Constructing data we have obtained into a dictionary
Creating a dataframe from the dictionary

re several different cases where the
ccessfully. Sometimes a landing was
ue to an accident; for example, True
ion outcome was successfully landed
he ocean while False Ocean means
as unsuccessfully landed to a specific
ue RTLS means the mission outcome
d to a ground pad False RTLS means
as unsuccessfully landed to a ground
he mission outcome was successfully
False ASDS means the mission
ssfully landed on a drone ship.

se outcomes into Training Labels with
successfully landed, "0" means it was

Perform expl
and determ

Calculate the
o

Calculate the n
of

Calculate the n
of mission ou

Create a lan
from O

Expo

otted:

r vs. Payload Mass, Flight Number vs. Launch Si

e, Orbit Type vs. Success Rate, Flight Number v

vs Orbit Type and Success Rate Yearly Trend

w the relationship between variables. If a relatio

ed in machine learning model.

comparisons among discrete categories. The go

een the specific categories being compared an

trends in data over time (time series)

s of the unique launch sites in the space mission

where launch sites begin with the string ‘CCA'

payload mass carried by boosters launched by NASA (CRS)

payload mass carried by booster version F9 v1.1

n the first successful landing outcome in ground pad was achieved

the boosters which have success in drone ship and have payload ma

nber of successful and failure mission outcomes

the booster versions which have carried the maximum payload mass

ding outcomes in drone ship, their booster versions and launch site

f landing outcomes (such as Failure (drone ship) or Success (ground

7-03-20 in descending order

ch Sites:

with Circle, Popup Label and Text Label of NASA Johnson S
ongitude coordinates as a start location.

with Circle, Popup Label and Text Label of all Launch Sites
ordinates to show their geographical locations and proxim

of the launch outcomes for each Launch Site:

d Markers of success (Green) and failed (Red) launches usin
unch sites have relatively high success rates.

n a Launch Site to its proximities:

d Lines to show distances between the Launch Site KSC LC

odown List:

lown list to enable Launch Site selection.

g Success Launches (All Sites/Certain Site):

hart to show the total successful launches count for all
ed counts for the site, if a specific Launch Site was sele

Mass Range:

to select Payload range.

Payload Mass vs. Success Rate for the different B

er chart to show the correlation between Payload and L

Standardizing the data with StandardScaler, then fitting and transforming it

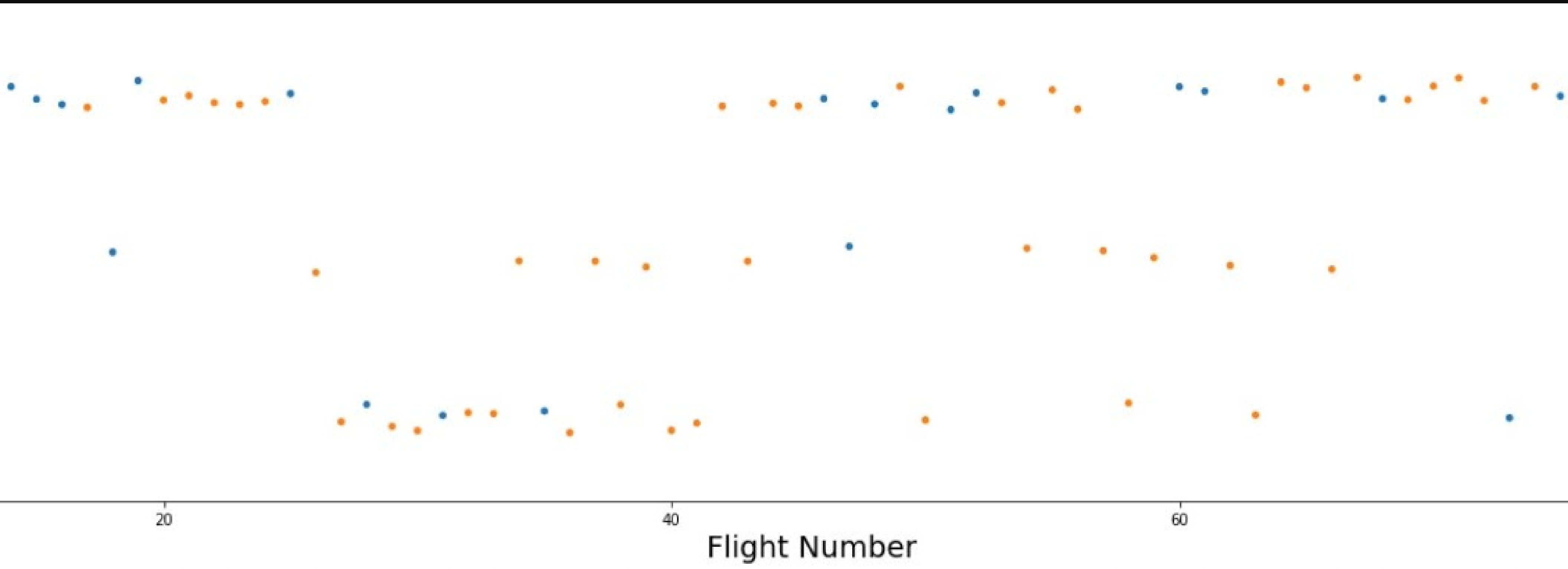
Splitting the data into training and testing sets with train_test_split function

Calculating the accuracy on the test data using the method .score() for all models

Examining the confusion matrix for all models

- Exploratory data ana
- Interactive analytics screenshots
- Predictive analysis re

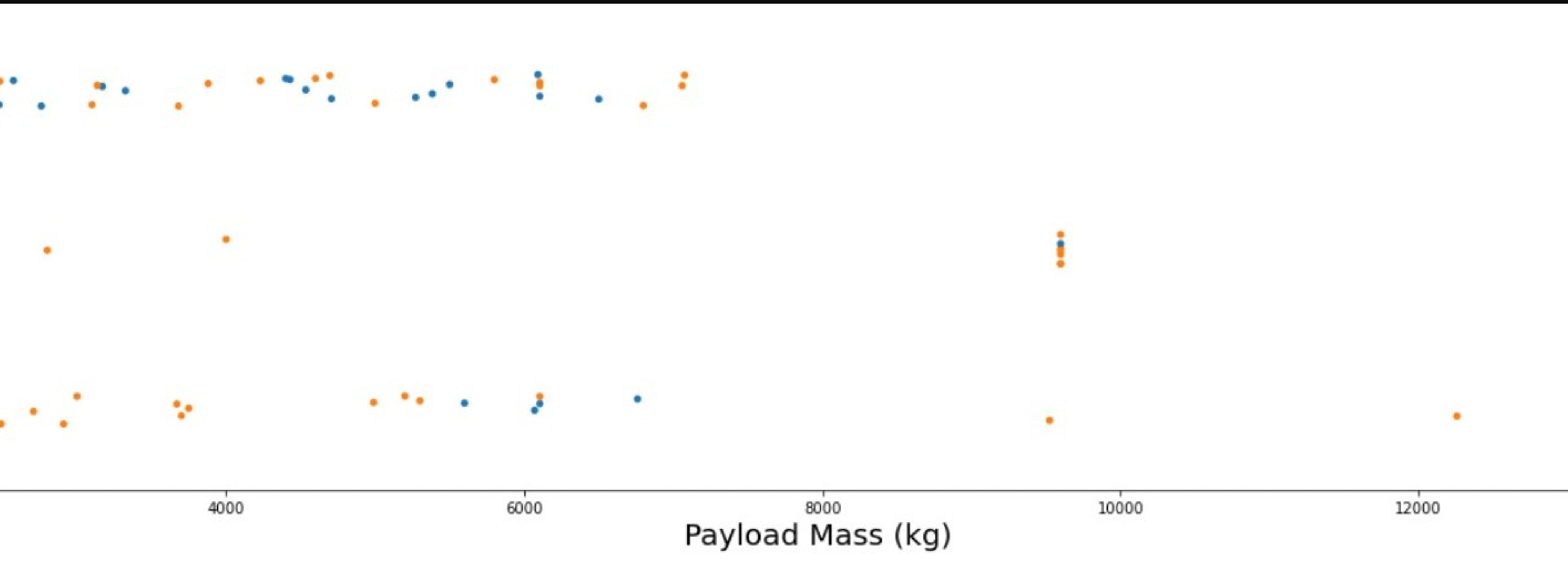
DA with Visualizatio



flights all failed while the latest flights all succeed

SLC 40 launch site has about a half of all launches

E and KSC LC 39A have higher success rates.



unch site the higher the payload mass, the hig

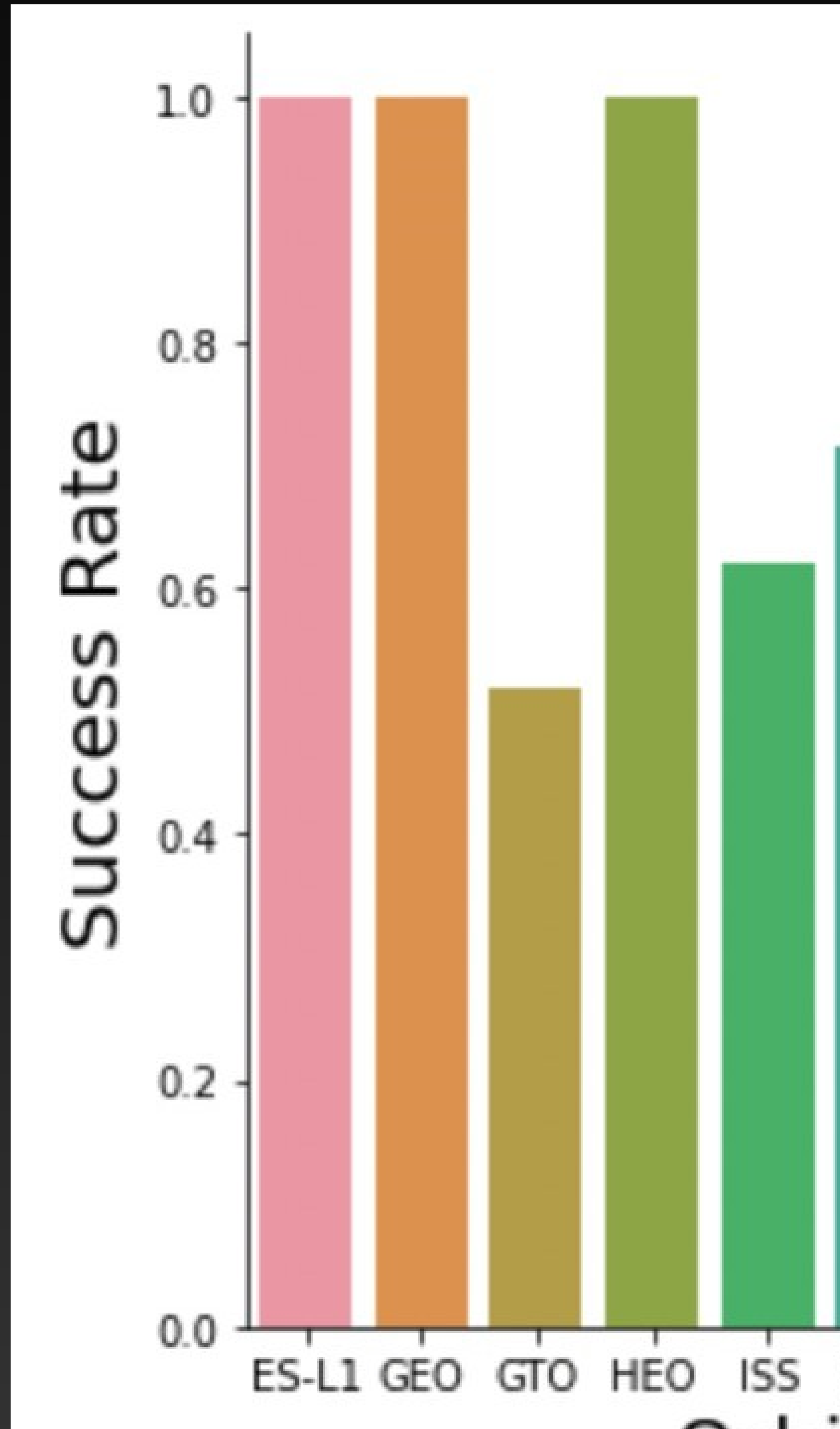
launches with payload mass over 7000 kg wer n 100% success rate:

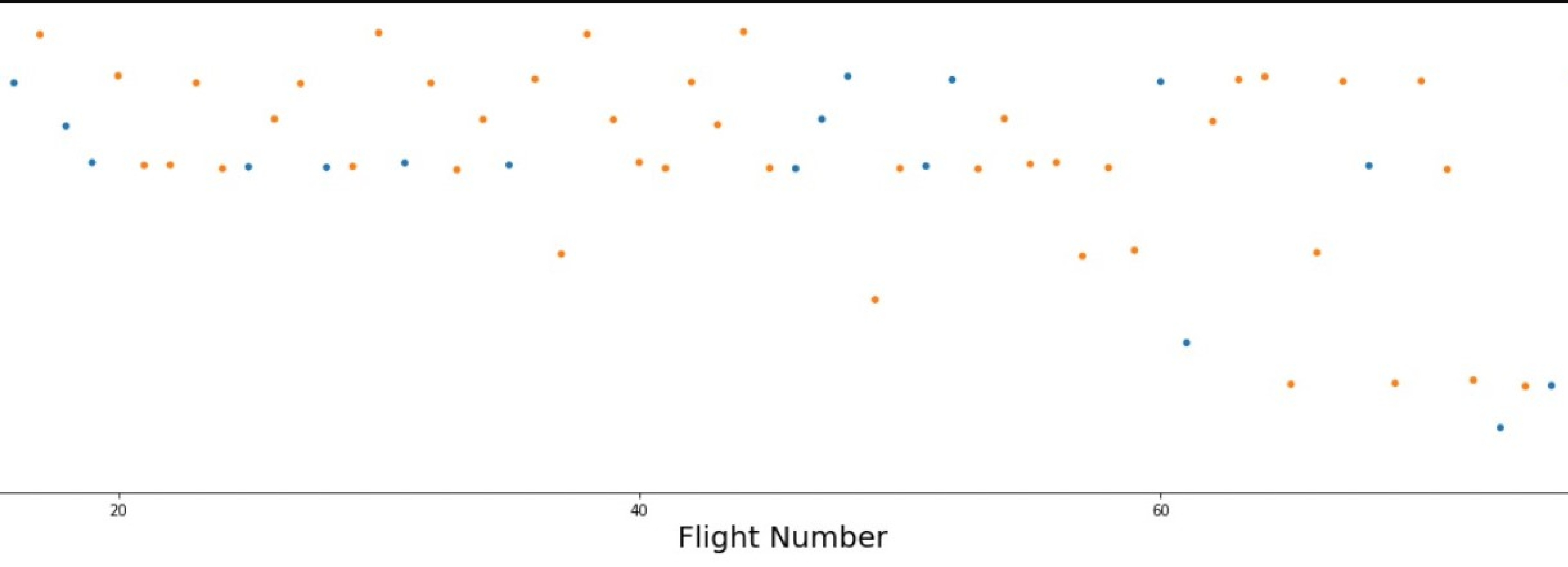
EO, HEO, SSO

n 0% success rate:

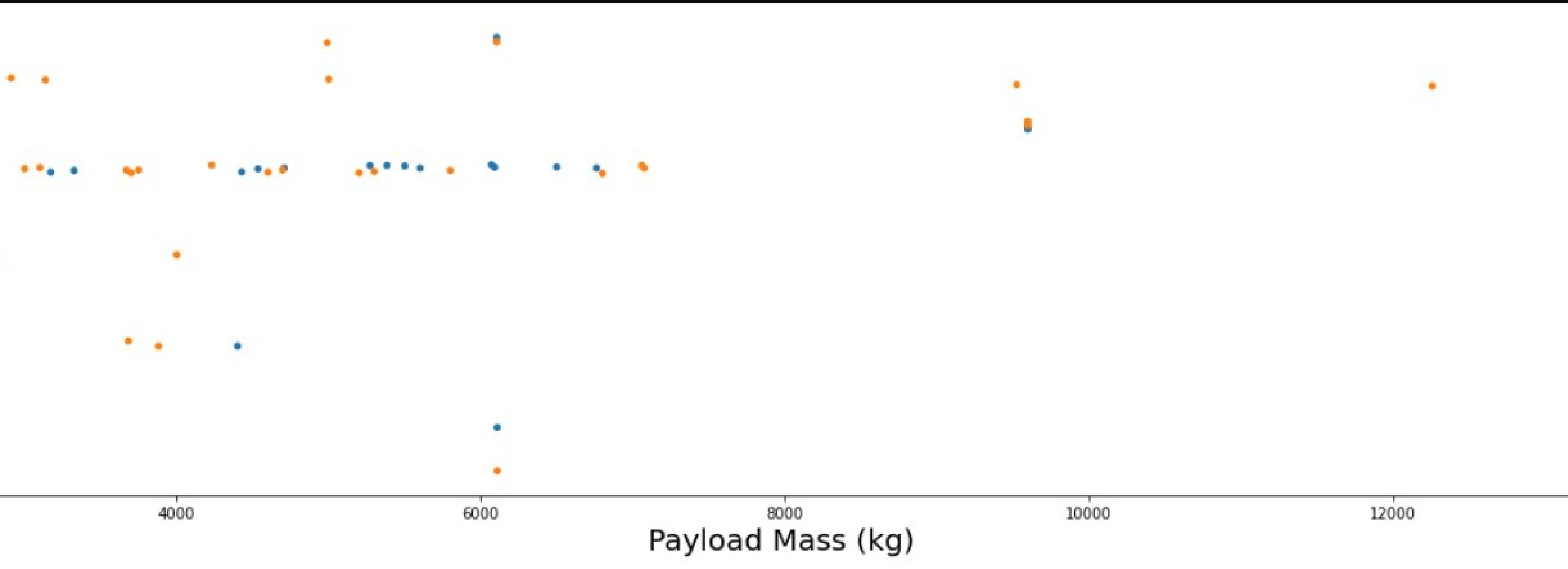
n success rate
0% and 85%:

S, LEO, MEO, PO



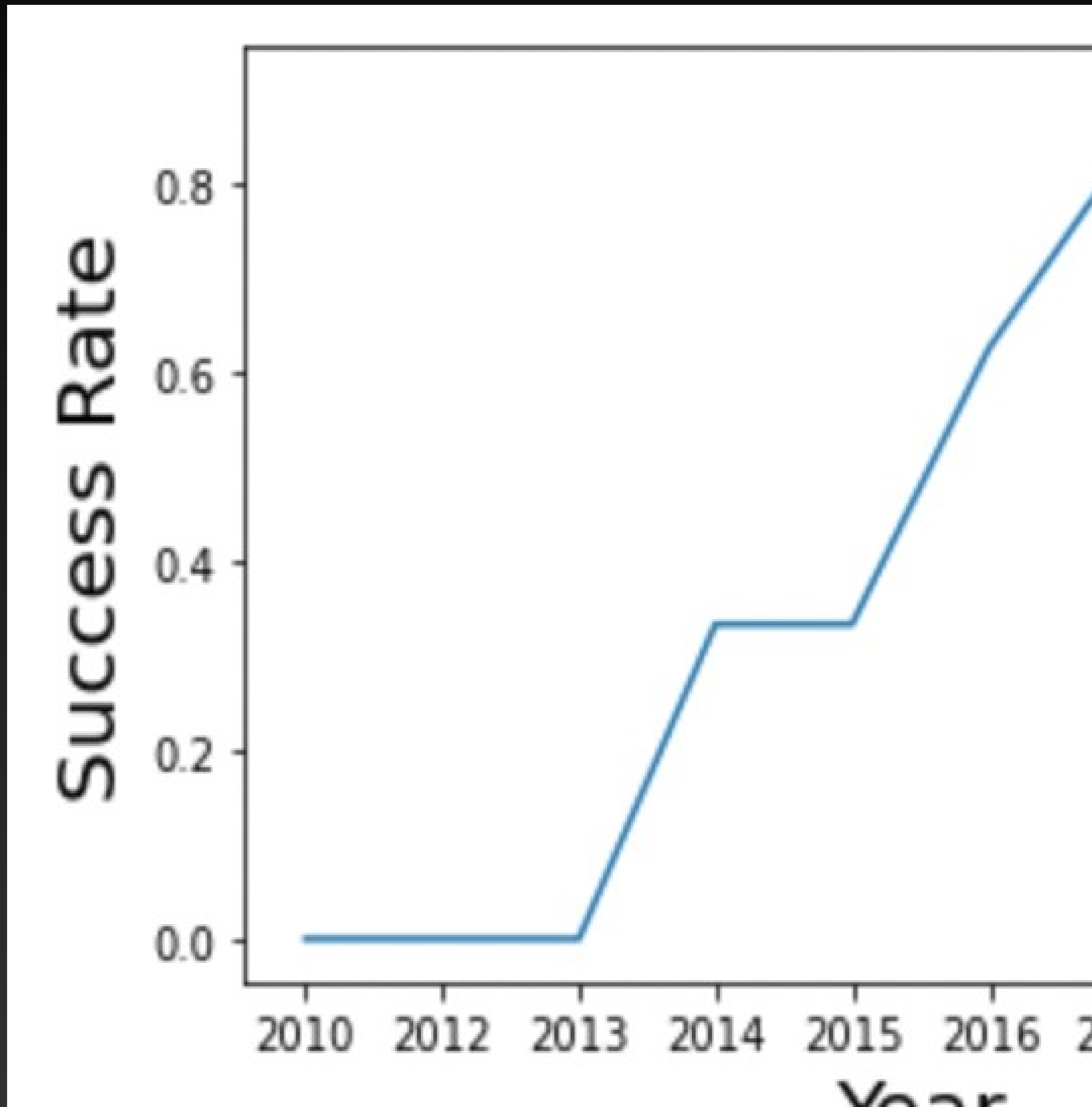


orbit the Success appears related to the nu
er hand, there seems to be no relationship



loads have a negative influence on GTO orb
d Polar LEO (ISS) orbits.

ss rate
kept
till 2020.



EDA with SQL

```
launch_site from SPACEXDATASET;
```

```
322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.app
```

the names of the unique launch sites in the

```
PACEXDATASET where launch_site like 'CCA%' limit 5;
```

```
322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqblod8lcg.databases.app
```

ster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mi
l.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Su
l.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Su
l.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Su
l.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Su
l.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Su

```
oad_mass__kg_) as total_payload_mass from SPACEXDATASET where customer = 'NASA
```

```
322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90108kqblod8lcg.databases.app
```

the total payload mass carried by boosters
S).

```
oad_mass__kg_) as average_payload_mass from SPACEXDATASET where booster_versio
```

```
322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqblod8lcg.databases.app
```

average payload mass carried by booster v

```
) as first_successful_landing from SPACEXDATASET where landing__outcome = 'Succ
322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqblod8lcg.databases.app
```

date when the first successful landing outco

chieved.

between 4000 and 6000

```
version from SPACEXDATASET where landing__outcome = 'Success (drone ship)' and
```

```
322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqblod8lcg.databases.app
```

names of the boosters which have success

mission outcomes

```
outcome, count(*) as total_number from SPACEXDATASET group by mission_outcome;
```

```
322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.app
```

	total_number
	1
	99
nclear)	1

total number of successful and failure miss

```
version from SPACEXDATASET where payload_mass__kg_ = (select max(payload_mass_
322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.app
```

```
me(date) as month, date, booster_version, launch_site, landing__outcome from S
_outcome = 'Failure (drone ship)' and year(date)=2015;
```

322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.app

oster_version	launch_site	landing__outcome
v1.1 B1012	CCAFS LC-40	Failure (drone ship)
v1.1 B1015	CCAFS LC-40	Failure (drone ship)

failed landing outcomes in drone ship, thei
nd launch site names for the months in year

```
__outcome, count(*) as count_outcomes from SPACEXDATASET
ween '2010-06-04' and '2017-03-20'
ng__outcome
_outcomes desc;
```

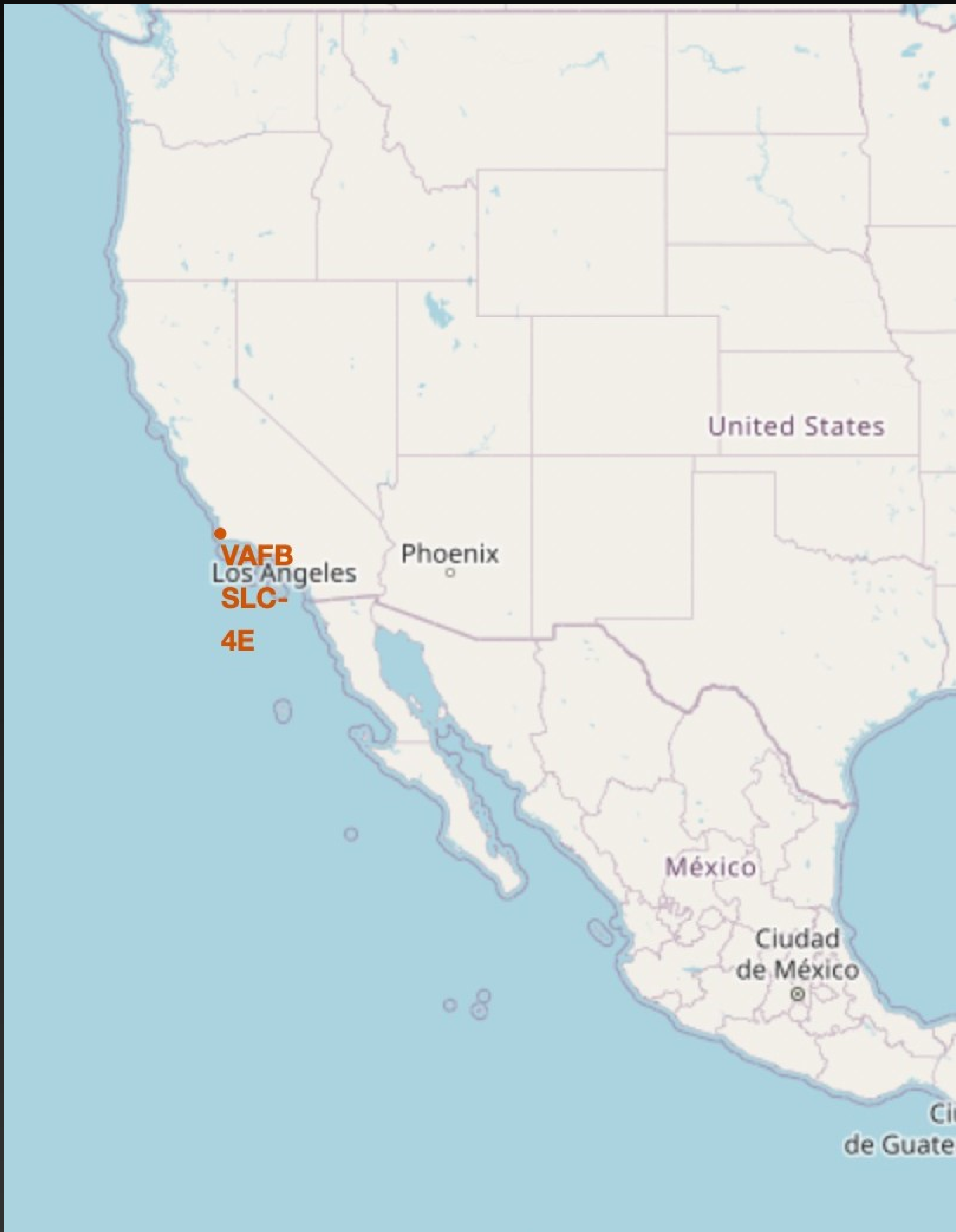
```
322:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqblod8lcg.databases.ap
```

ount_outcomes
0

ractive map with Fo

es are in proximity to the
and is moving faster at
ny other place on the
h. Anything on the
h at the equator is
1670 km/hour. If a ship is
equator it goes up into
o moving around the
speed it was moving
This is because of inertia.
p the spacecraft keep up
eed to stay in orbit.

in very close proximity
launching rockets
it minimises the risk of
dropping or exploding

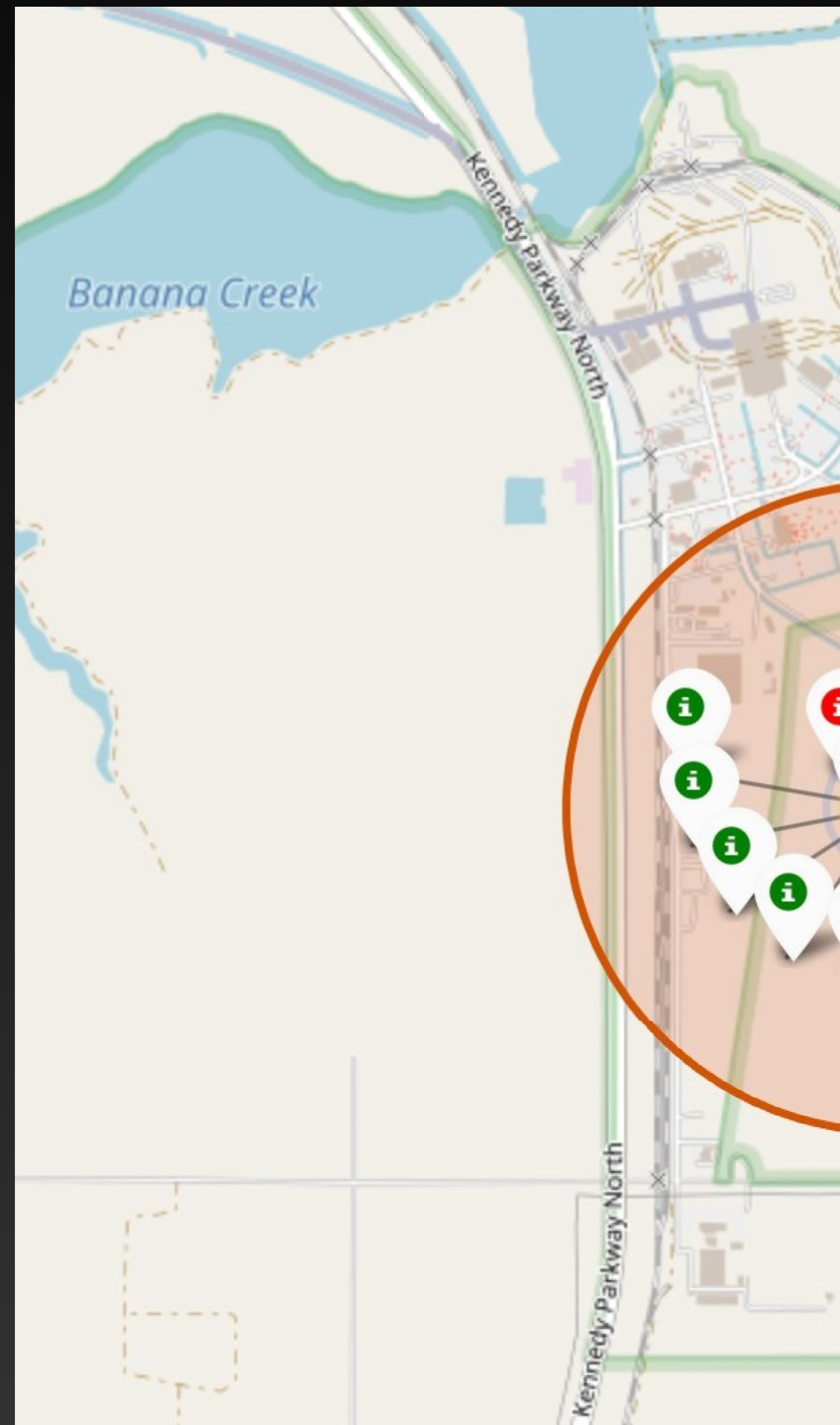


olour-labeled markers
be able to easily
hich launch sites have
igh success rates.

arker = Successful

ker = Failed Launch

e KSC LC-39A has a



SC LC-39A to its proximiti

lysis of the launch
can clearly see that

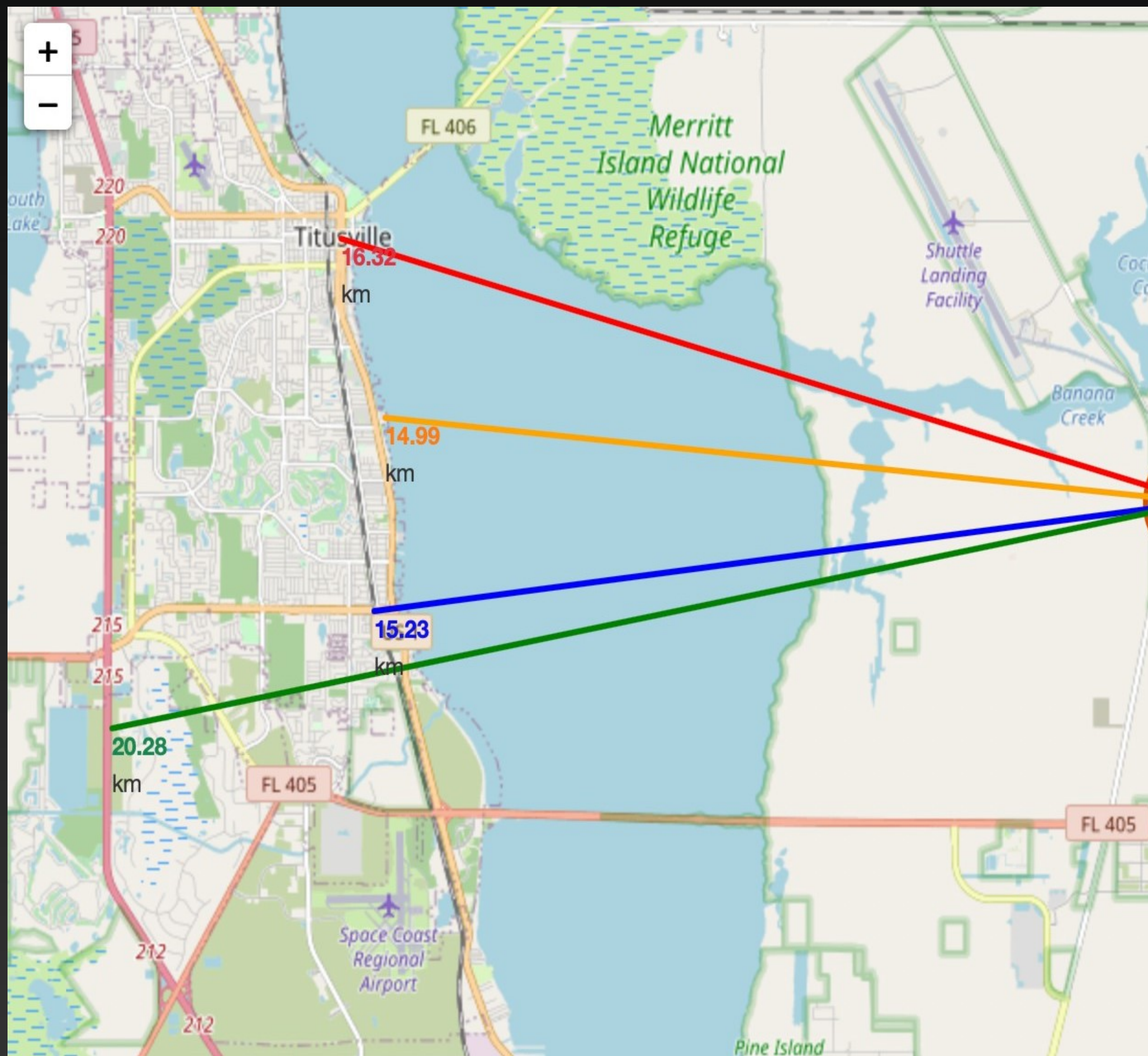
railway (15.23 km)

highway (20.28 km)

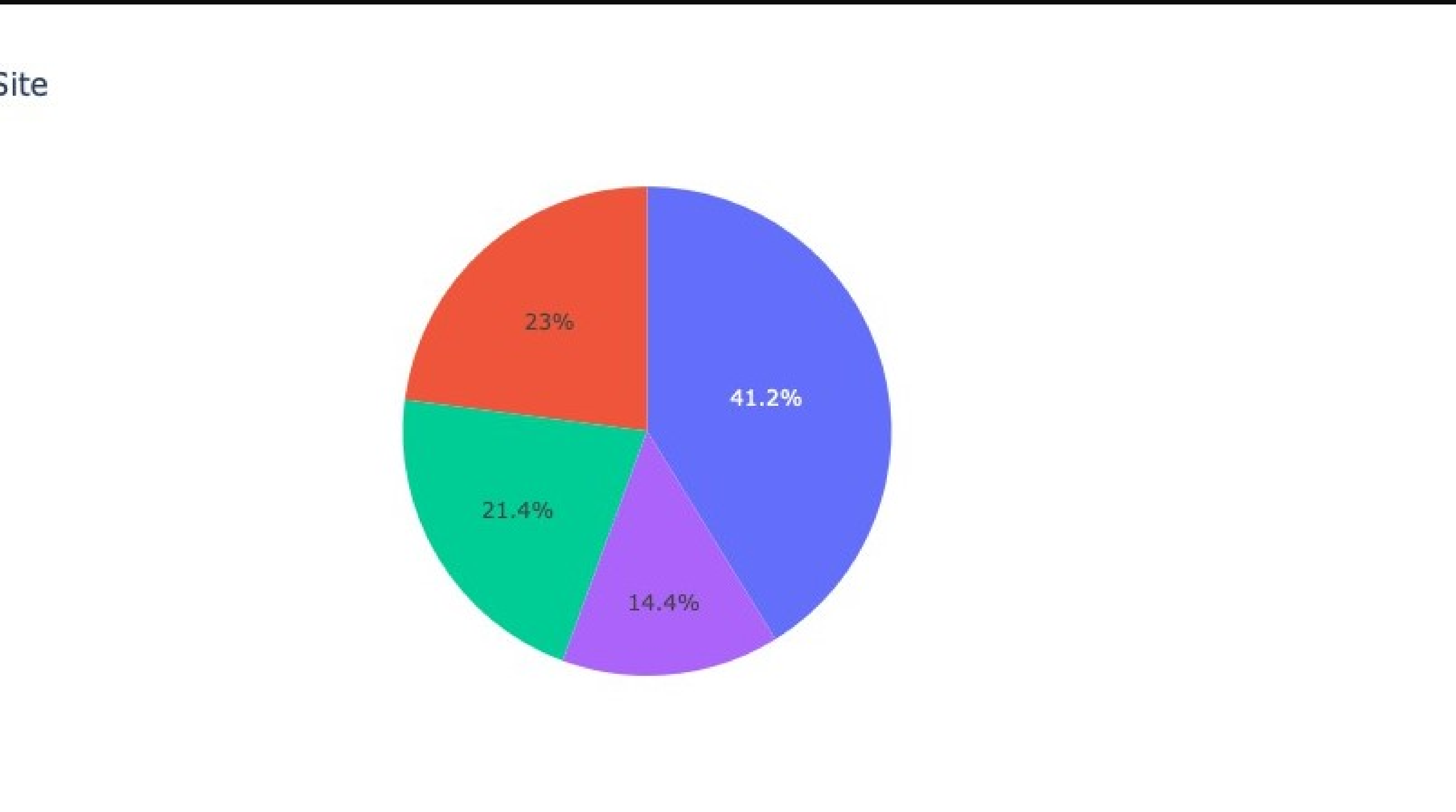
coastline (14.99 km)

KSC LC-39A is
closest city
.

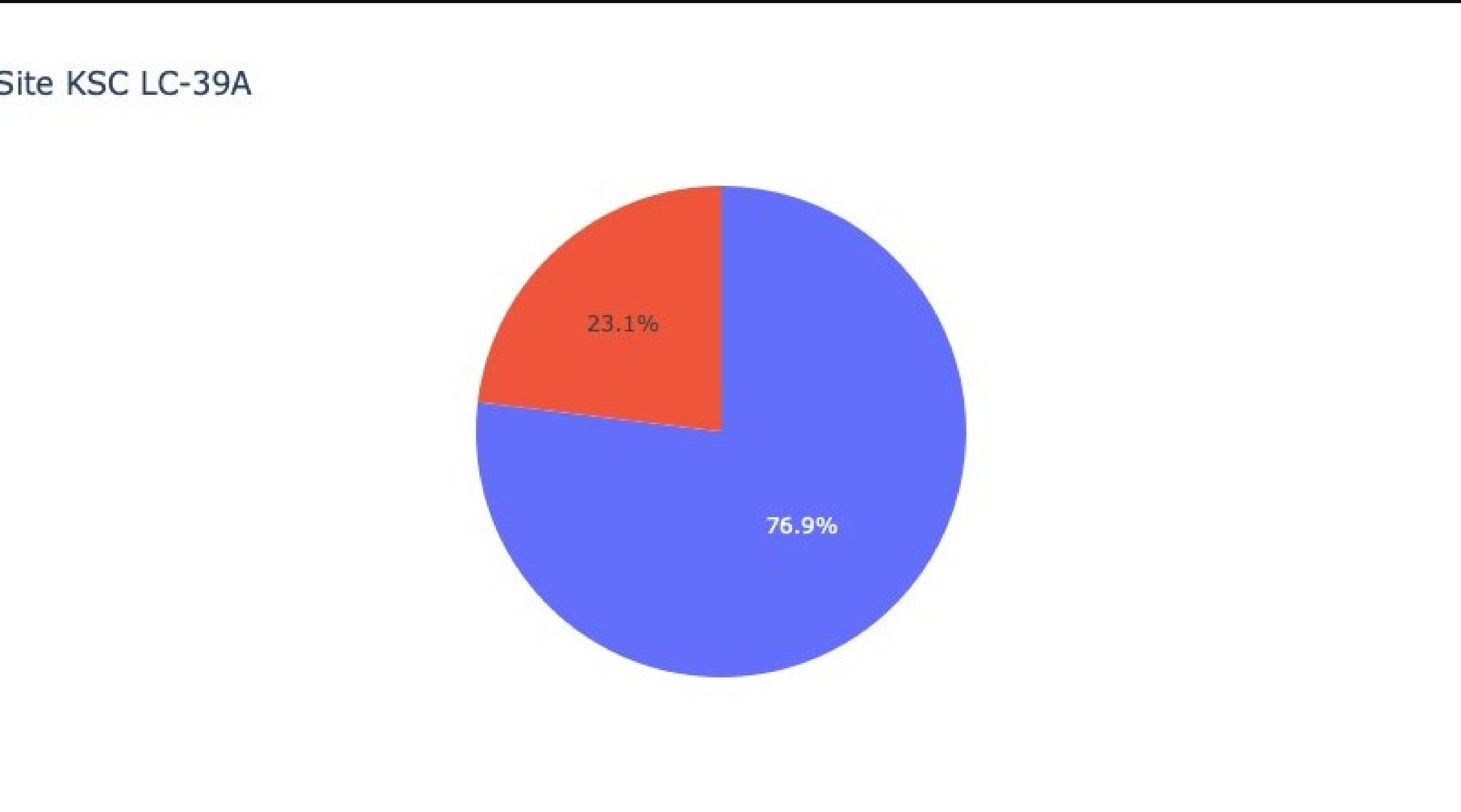
s high speed can
e 15-20 km in few
e potentially



a Dashboard with
Dash



early shows that from all the sites, KSC LC 39A has



has the highest launch success rate (76.9%) with 1

show
ads
000
kg have
t
te.

Predictive analysis (Classification)

scores of the Test Set,
nfirm which method

scores may be due
st sample size (18
efore, we tested all
d on the whole

the whole Dataset
ne best model is the
Model. This model
igher scores, but also

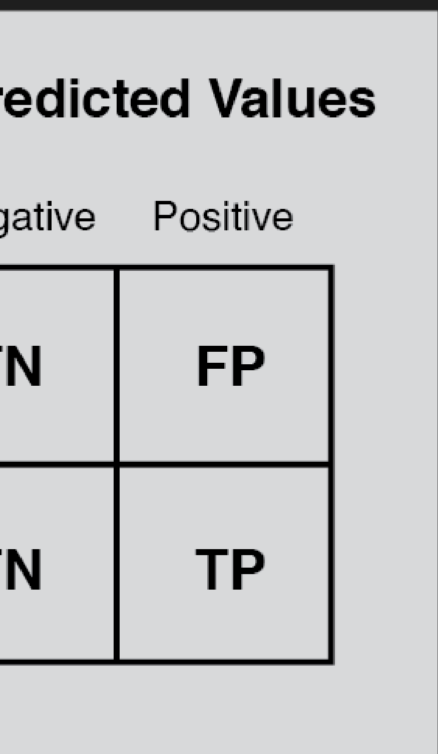
Scores and Accuracy o

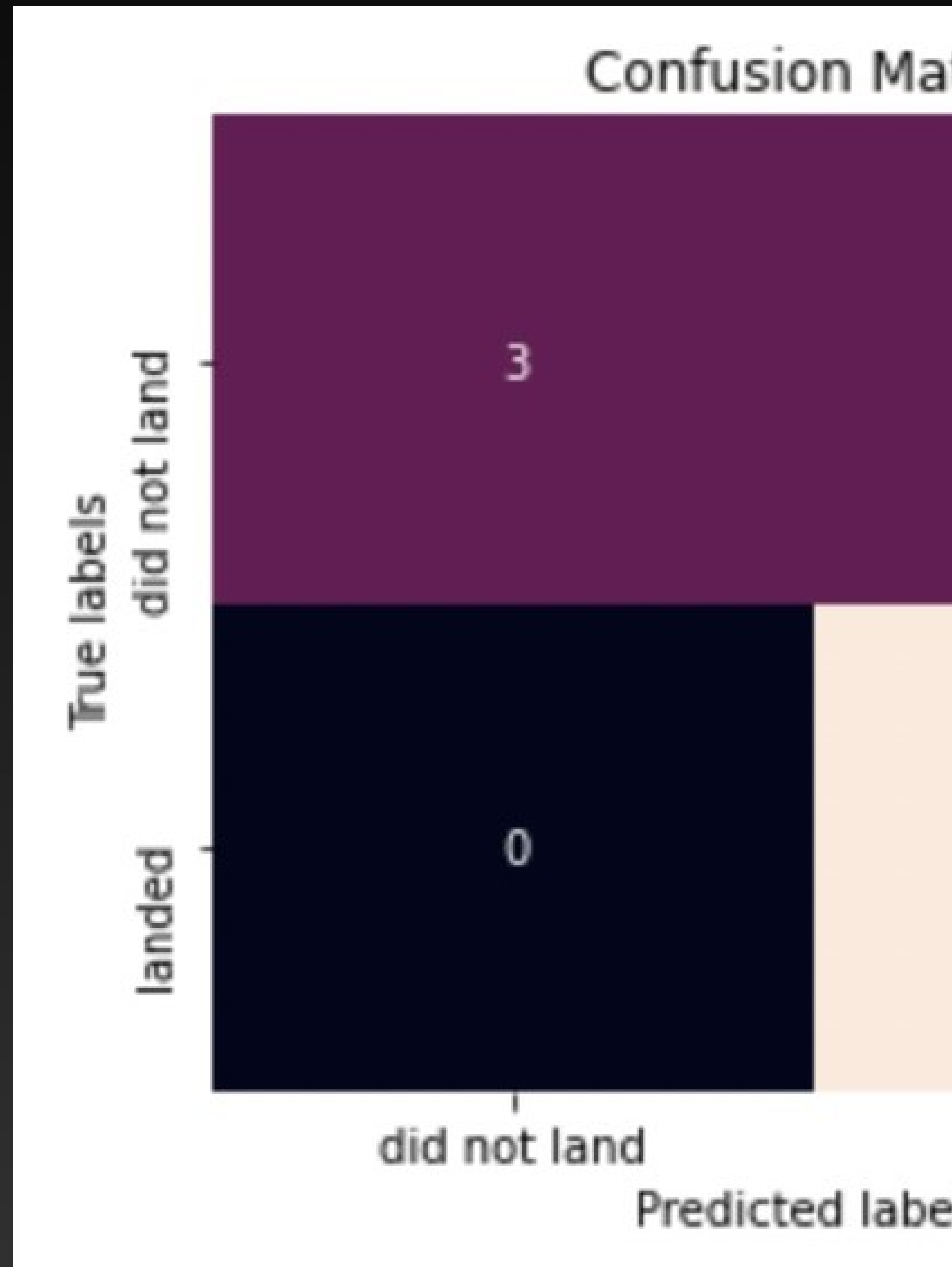
	LogReg	SVM
Jaccard_Score	0.800000	0.80000
F1_Score	0.888889	0.88888
Accuracy	0.833333	0.83333

Scores and Accuracy of th

	LogReg	SVM
Jaccard_Score	0.833333	0.84507
F1_Score	0.909091	0.91603

onfusion matrix, we see
ession can distinguish
erent classes. We see
roblem is false positives.





- Decision Tree Model is the best algori
- Launches with a low payload mass sh
  than launches with a larger payload m
- Most of launch sites are in proximity t
  and all the sites are in very close prox
- The success rate of launches increase
- KSC LC-39A has the highest success r
  from all the sites.
- Orbits ES-L1, GEO, HEO and SSO have

Special Thanks

[Instructors]

[Coursera]

[IBM]