Winning Space Race with Data Science

Andrew Jansen
10 Mar, 2024

Outline

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

Executive Summary

- Methodologies: Data on Space X launches collected and wrangled using pandas, visualized using plotly and folium, explored with SQL, then used to fit models using sklearn.
- Classifiers provide decent predictive power, with a decision tree model having a slight edge, on this small dataset. More launch data in the future should improve the model.

Introduction

- Space X's Falcon 9 rockets have a reusable first stage, but the landing of those first stages is not guaranteed and therefore not always reusable. The cost varies as a result.
- We want to know what factors influence the probability of a successful landing for the first stage and build a model that can predict that probability and estimate cost for future launches.

Section 1

Methodology

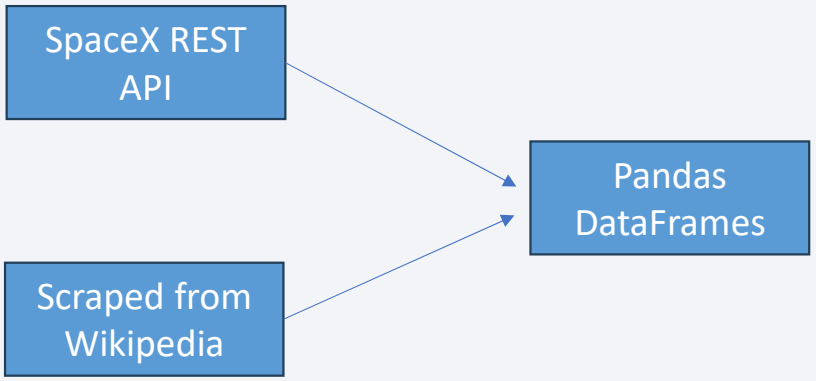
Methodology

Executive Summary

- Data collection methodology:
  - Data was collected Via the public SpaceX API and scraped from a Wikipedia on launch history
- Perform data wrangling
  - Data was cleaned and transformed/enriched using Pandas
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Classifiers were fit to training data, tuned via Grid Search, and scored on test data using sklearn.

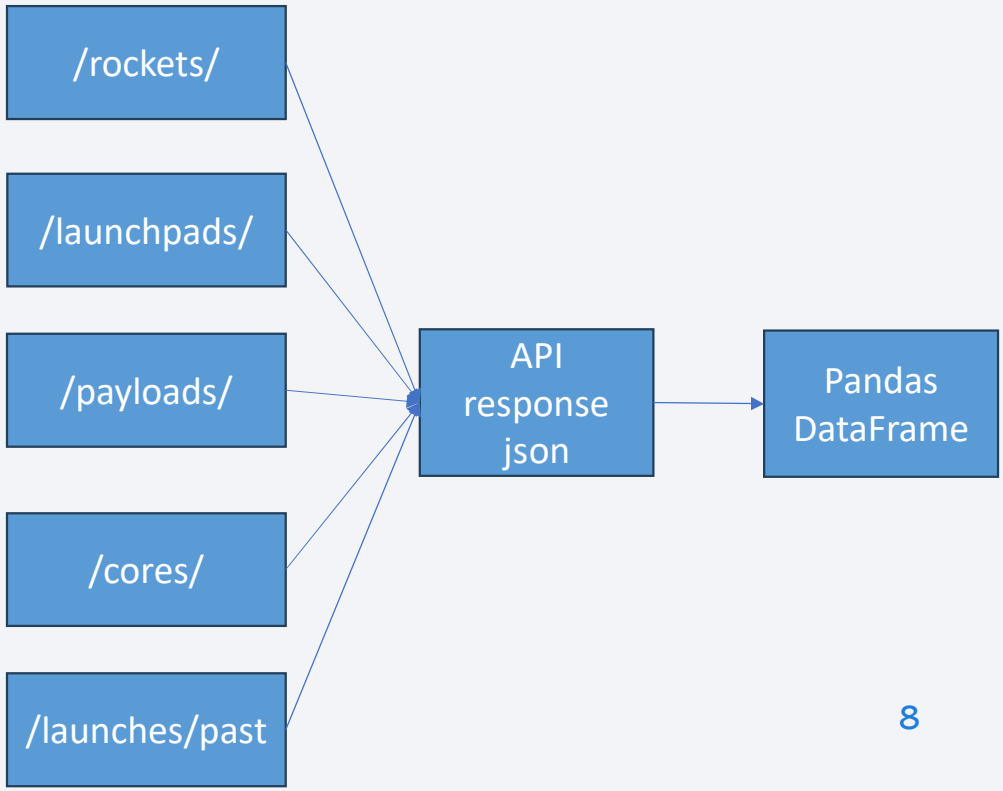
Data Collection

- Detailed data on 90 SpaceX Falcon 9 launches were collected via two sources:
  - The SpaceX REST API using python requests package
  - The List of Falcon 9 and Falcon Heavy launches WikiPedia page using python BeautifulSoup package



Data Collection – SpaceX API

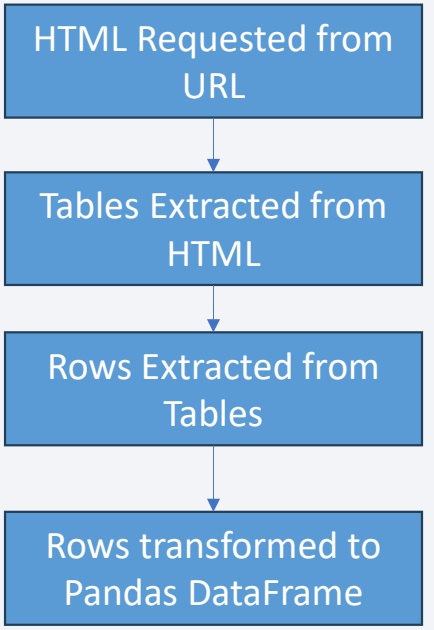
- The following endpoints of the API at api.spacexdata.com were used to gather detailed data about each launch using the Falcon 9:
  - /v4/rockets/
  - /v4/launchpads/
  - /v4/payloads/
  - /v4/cores/
  - /v4/launches/past
- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose



Data Collection - Scraping

- Data was scraped from the Wiki tables on the List of Falcon 9 and Heavy Launches Wikipedia page using BeautifulSoup.

- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose



Data Wrangling

Cleaning

- Null values for payload mass were imputed with the average payload mass
- Launches with irrelevant boosters were removed
- Column data types and value counts were checked

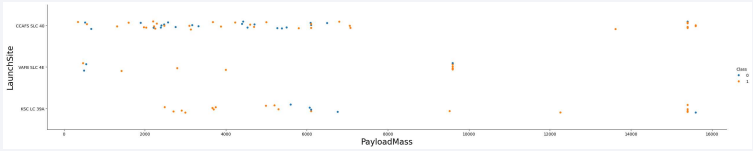
Transforming

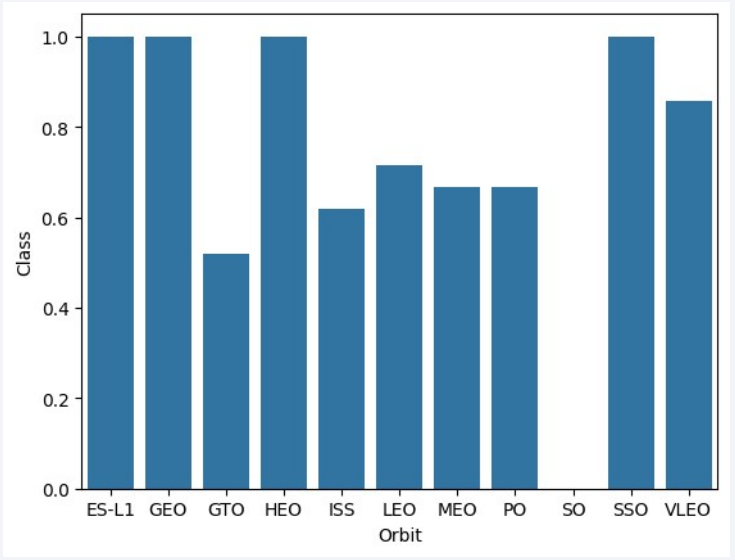
- Landing outcomes were coded into a success/failure category

- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose

EDA with Data Visualization

- Payload mass, orbit type, and launch success rate were plotted against flight number / launch date to understand trends over time
- Success rate was also plotted by orbit type to better understand the success rate trend
- Payload mass was plotted against launch site to understand site utilization
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and





EDA with SQL

To explore this data, it was exported to a SQLite DB and queried for information like:

- Unique launch sites and landing outcomes
- Total payload mass for boosters launched by NASA (CRS), avg. payload mass for Falcon 9 v1.1 boosters, and all boosters that have carried the highest payload
- Date of the first successful landing in ground pad, and a list of drone ship failures from 2015
- Counts of launches by landing outcome from 2010-06-04 to 2017-03-20 and counts of launches by mission outcome
- Boosters which have success in drone ship at payload mass between 4000 and 6000

- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

Build an Interactive Map with Folium

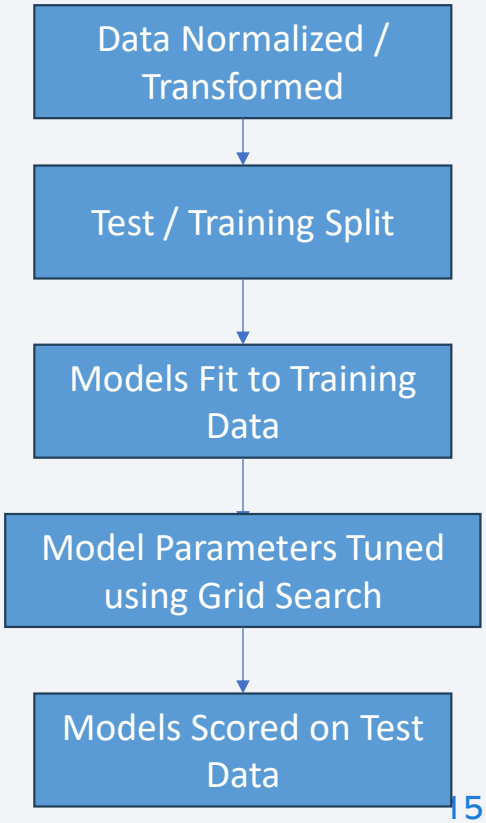
- Added a circle around Johnson Space Center, circles/markers for launches at the launch site coordinates clustered by location and color coded by success/failure to a Folium map of the U.S.
- Also added measurement lines between select launch sites and coasts and major roads
- These objects were added to make an easily navigable map to visualize success rate, total utilization, and geographic features of launch sites (e.g., proximity to coasts, highways)
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose

Build a Dashboard with Plotly Dash

- The Plotly Dash includes:
  - A pie chart of total successes by launch site
  - A scatter plot of successes and failures against payload mass and grouped by booster version
  - Filters for launch site and payload mass
- These interactive visuals allow exploration of total launch site usage and success rate by booster version at each site
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

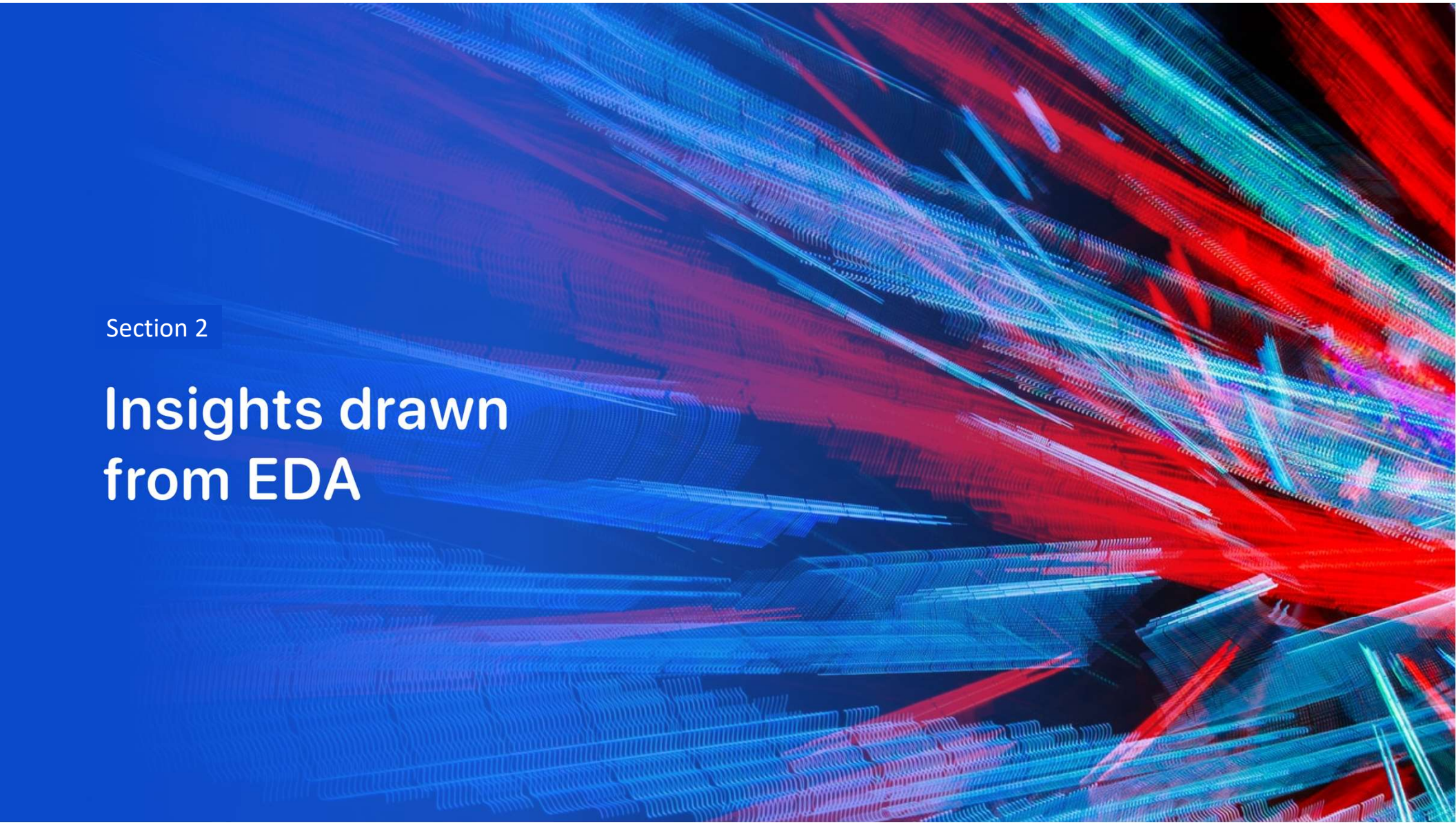
Predictive Analysis (Classification)

- Logistic Regression, Support Vector Machine, Decision Tree, and K-Nearest Neighbors classifiers were fit to 80% of the data for training, tuned using grid searches to find the best hyper-parameters, and scored on 20% of the data for testing using sklearn.
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose



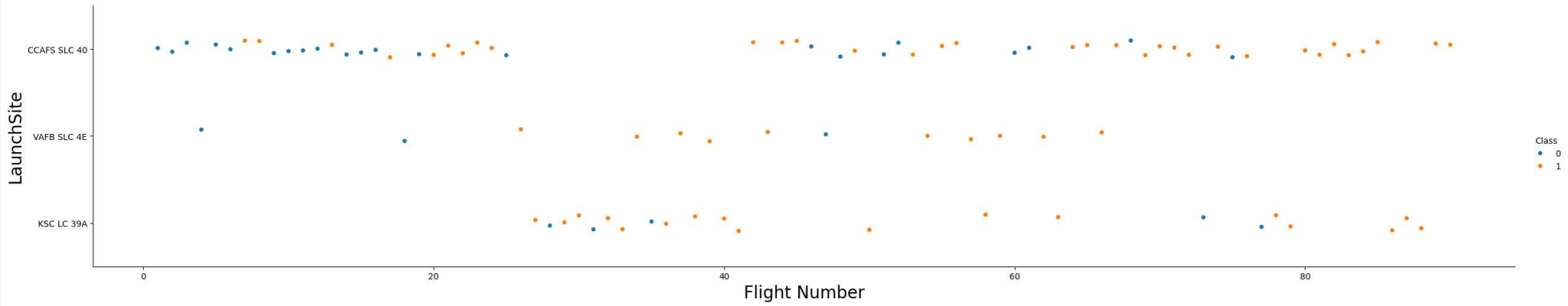
Results

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


Section 2
Insights drawn
from EDA

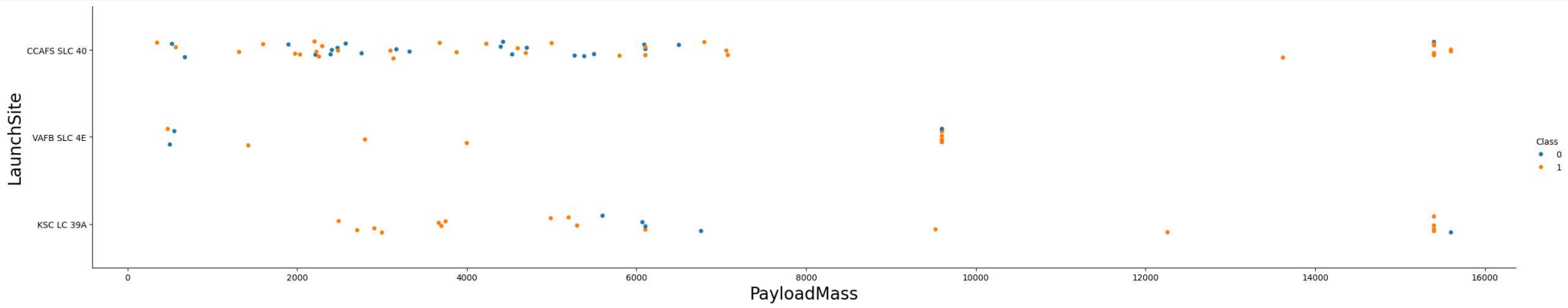
Flight Number vs. Launch Site

- This visual gives a sense for how often each launch site has been used and the trend over time of success/failure at those sites.



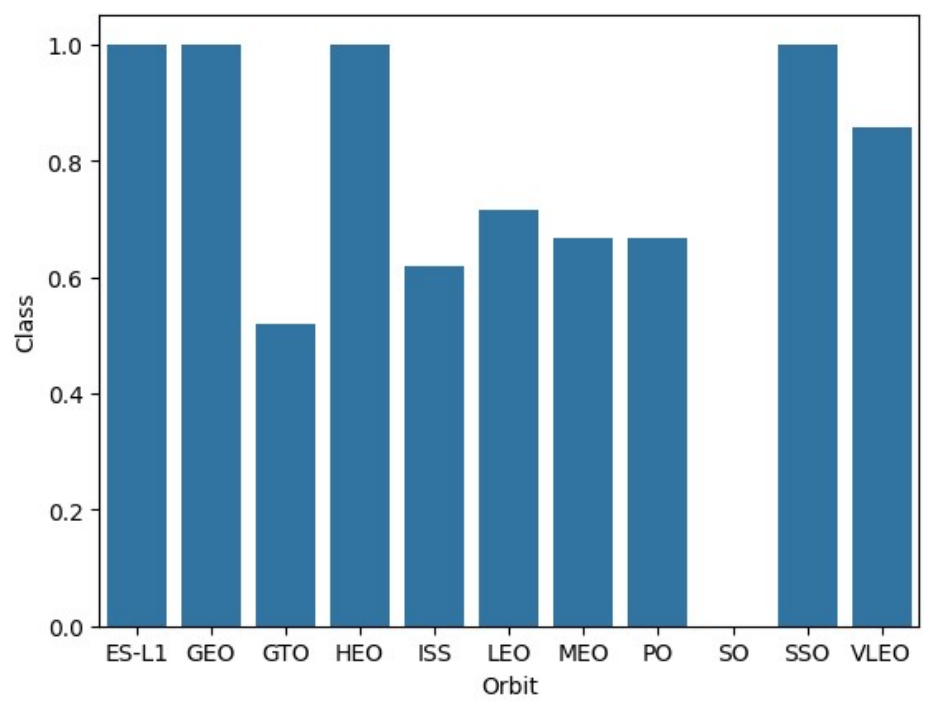
Payload vs. Launch Site

- This visual gives a sense of success rate for different payload masses and payload masses by launch site



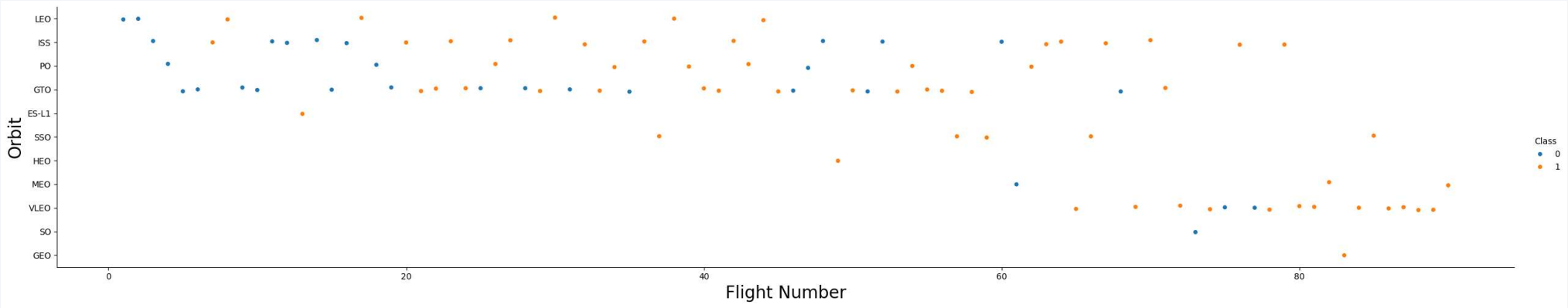
Success Rate vs. Orbit Type

- This visual shows success rate by orbit type, but requires some context in terms of total launches by orbit type over time.



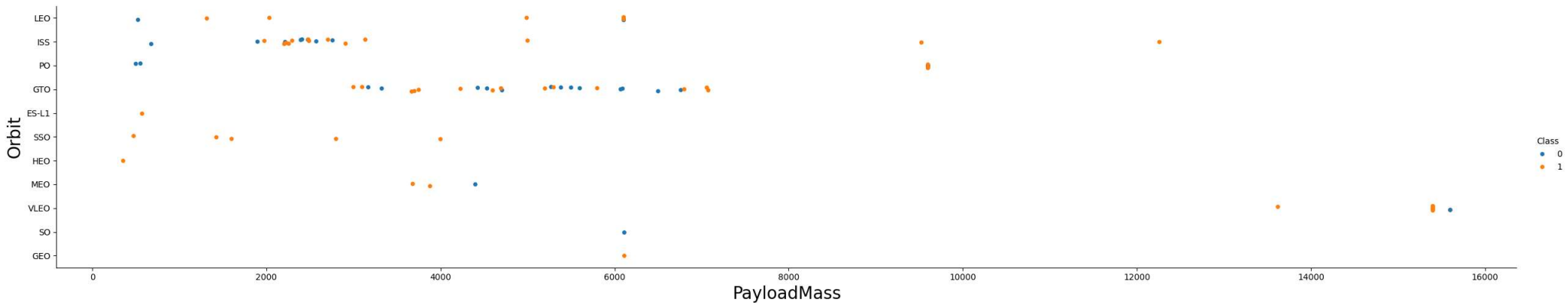
Flight Number vs. Orbit Type

- This visual provides additional context of total orbit launches over time.



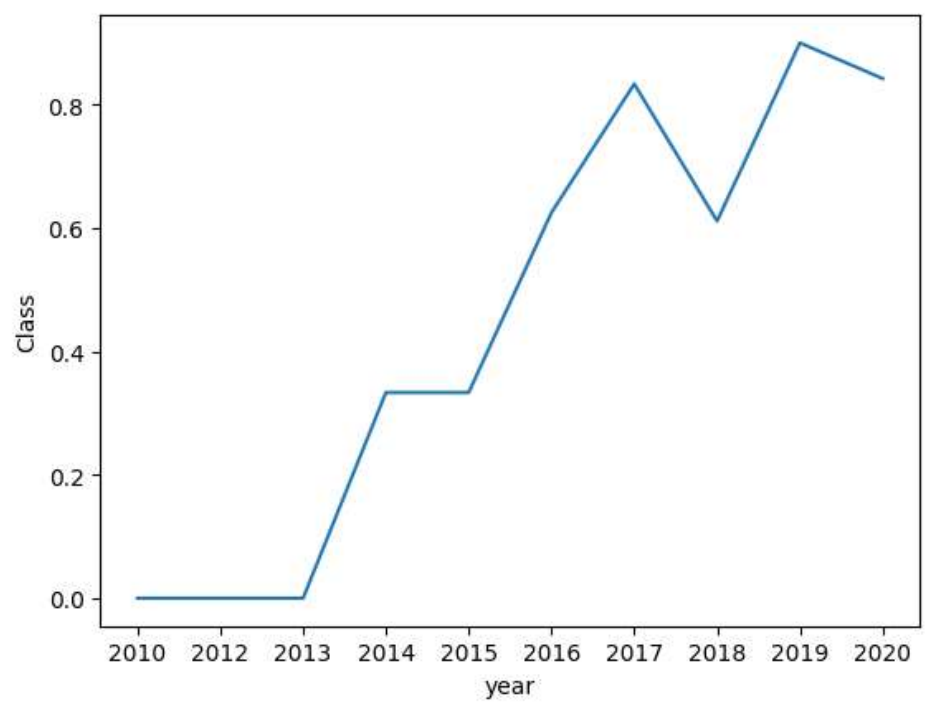
Payload vs. Orbit Type

- This provides insight into how the different orbit types relate to payload mass, providing context to the previous payload mass success rates.



Launch Success Yearly Trend

- This visual shows the role of time and technological advancement in determining the success probability of Falcon 9 booster launches.



All Launch Site Names

- This query lists the 4 unique launch sites.

```
%sql select distinct Launch_Site from spacextable
```

```
 * sqlite:///my_data1.db
Done.
```

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

Launch Site Names Begin with 'CCA'

- This query shows a sample of records for launch sites beginning with ‘CCA’.

```
%sql select * from spacextable where Launch_Site like 'CCA%' limit 5
```

Python

```
* sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outc
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (paracl
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 (paracl
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No atte
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No atte
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No atte

Total Payload Mass

- Displays the total of roughly 45.6k kg in payloads launched for NASA (CRS).

```
%sql select sum(PAYLOAD_MASS__KG_   ) total_mass_kg\
from spacextable where Customer = 'NASA (CRS)'
```

```
 * sqlite:///my_data1.db
Done.
```

total_mass_kg
45596

Average Payload Mass by F9 v1.1

- This query returns the roughly 2.5k kg avg payload mass for booster version F9 v1.1

```
%sql select avg(PAYLOAD_MASS__KG_   ) avg_mass_kg\
from spacextable where Booster_Version like 'F9 v1.1%'
```

```
 * sqlite:///my_data1.db
Done.
```

avg_mass_kg
2534.6666666666665

First Successful Ground Landing Date

- This query returns the first (min) date of a successful ground pad launch.

```
%sql select min(Date) first_ground_pad_success\
from spacextable where Landing_Outcome = 'Success (ground pad)'
```

```
 * sqlite:///my_data1.db
Done.
```

first_ground_pad_success
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- This query lists the four distinct booster versions that carried payloads of 4000-6000 kg.

```
%sql select distinct Booster_Version from spacextable\
where (PAYLOAD_MASS__KG_ between 4000 and 6000)\
and Landing_Outcome = 'Success (drone ship)'
```

```
 * sqlite:///my_data1.db
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

- This query shows the number of launches by mission outcome, which were overwhelmingly mission successes.

```
%sql select Mission_Outcome, count(1)\
from spacextable\
group by 1
```

```
 * sqlite:///my_data1.db
Done.
```

Mission_Outcome	count(1)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- This query shows the unique booster versions which carried the heaviest recorded payload(s).

```
%sql select distinct Booster_Version from spacextable\
where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from spacextable)\
order by 1;
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

2015 Launch Records

- This query presents both drone ship landing failures from 2015, along with their months.

```
%sql select substr(Date,6,2) month, Landing_Outcome, Booster_Version, Launch_Site from spacextable\
where substr(DATE,0,5) = '2015'\
and Landing_Outcome = 'Failure (drone ship)'
```

```
 * sqlite:///my_data1.db
Done.
```

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

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

- This query pulls the top landing outcomes in terms of total launches between the above dates.

```
%sql select Landing_Outcome, count(1) total from spacextable\
where Date between '2010-06-04' and '2017-03-20'\
group by Landing_Outcome order by 2 desc;
```

```
 * sqlite:///my_data1.db
Done.
```

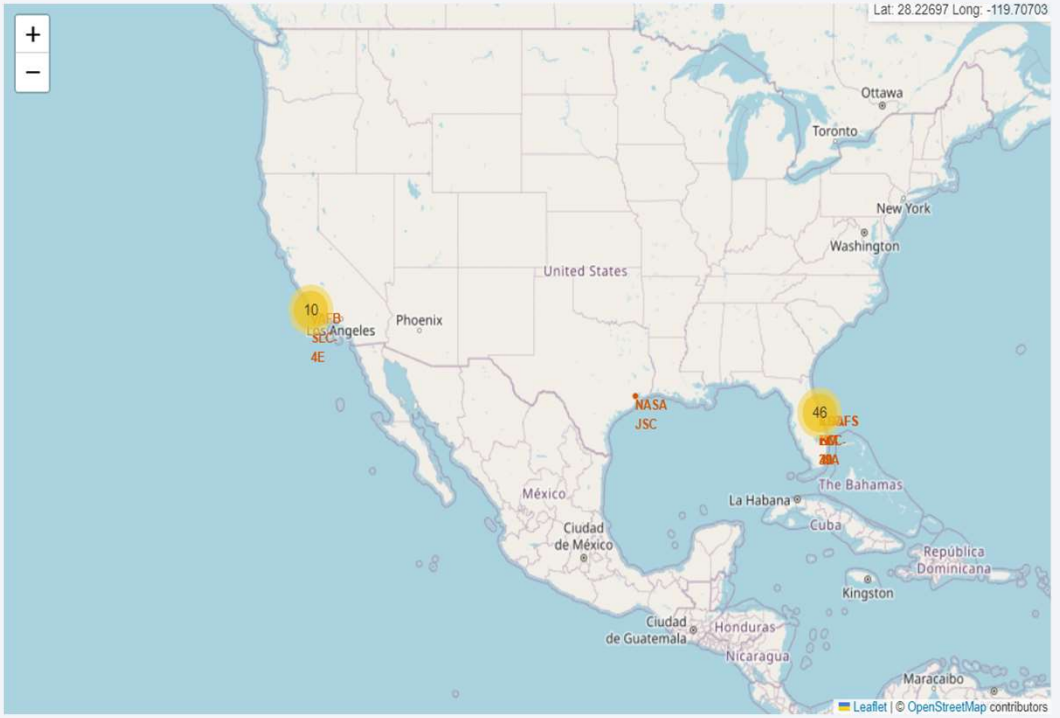
Landing_Outcome	total
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

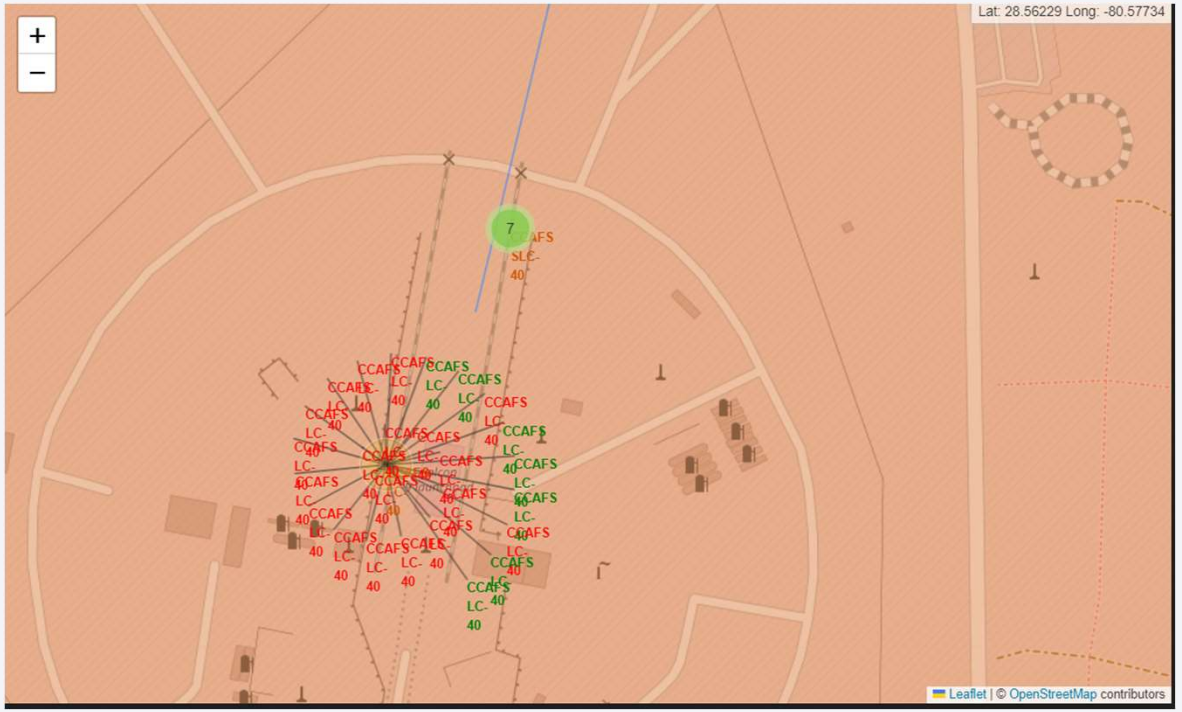
Folium Map Overview

- The Folium map features circles and text markers for each site + Johnson Space Center, with clusters for grouped points.



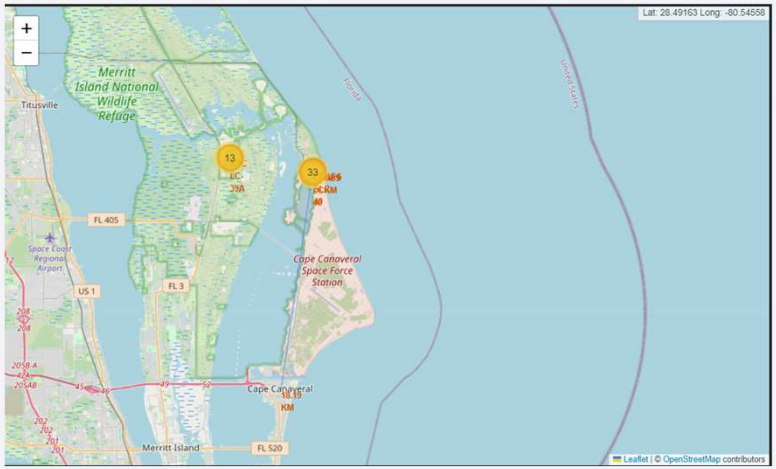
Folium Map Launch Outcomes

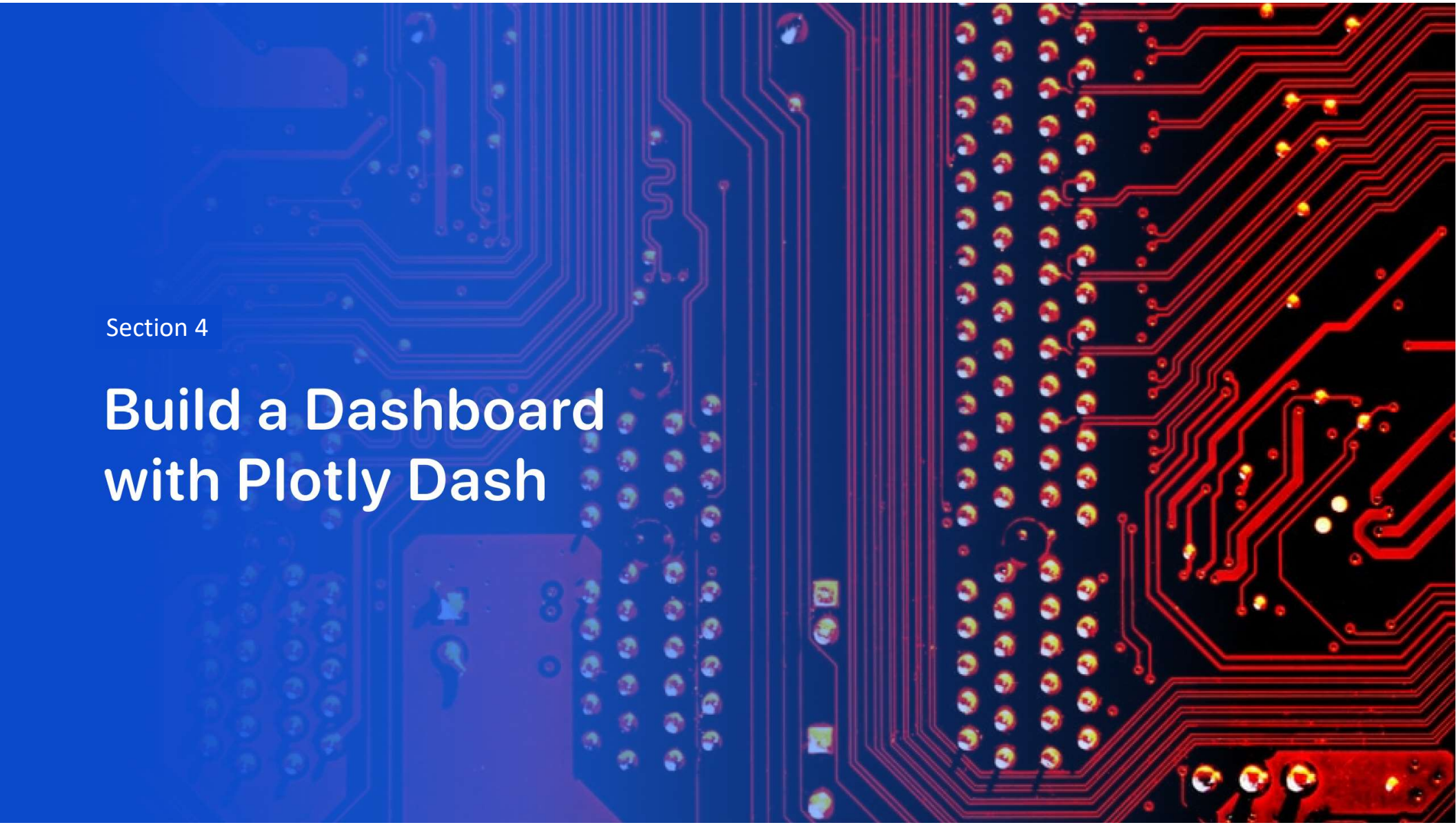
- The Folium map also has color-coded points for each launch at a site, shown here in a spiral grouped into a marker cluster.



Folium Map Proximities

- The Folium map also has polygon lines and distance markers from select launch sites to landmarks like the coast (0.9km) and the nearest city (18.2km).



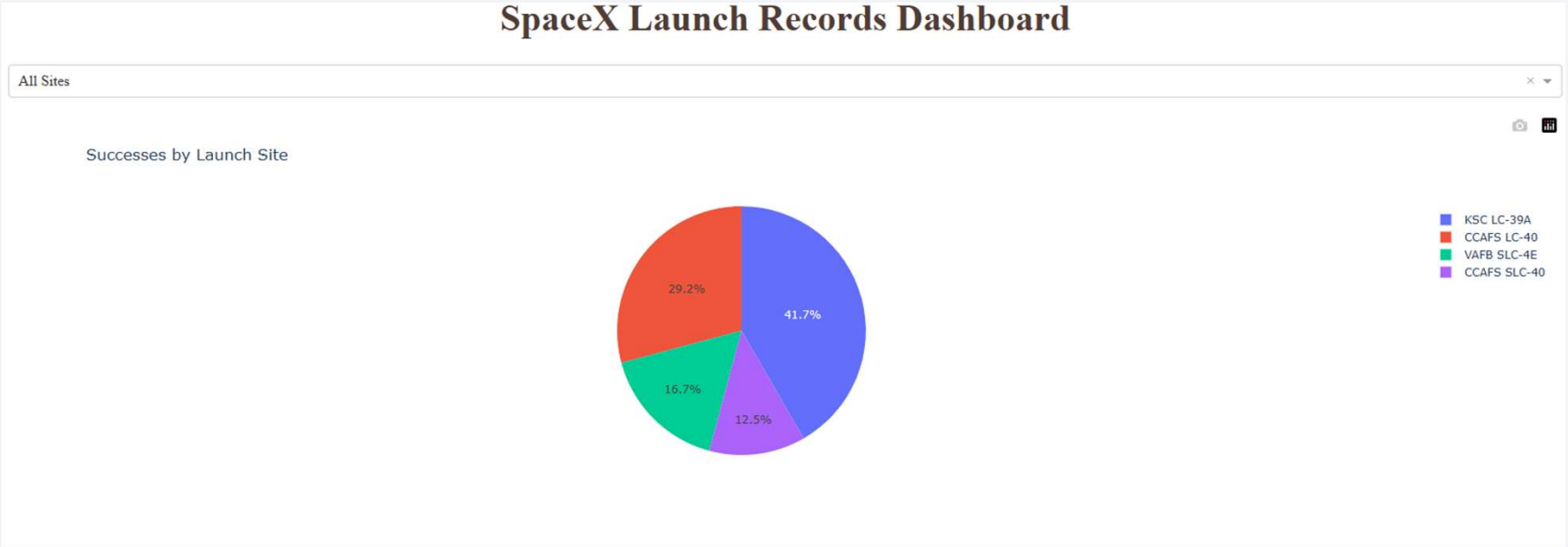


Section 4

Build a Dashboard with Plotly Dash

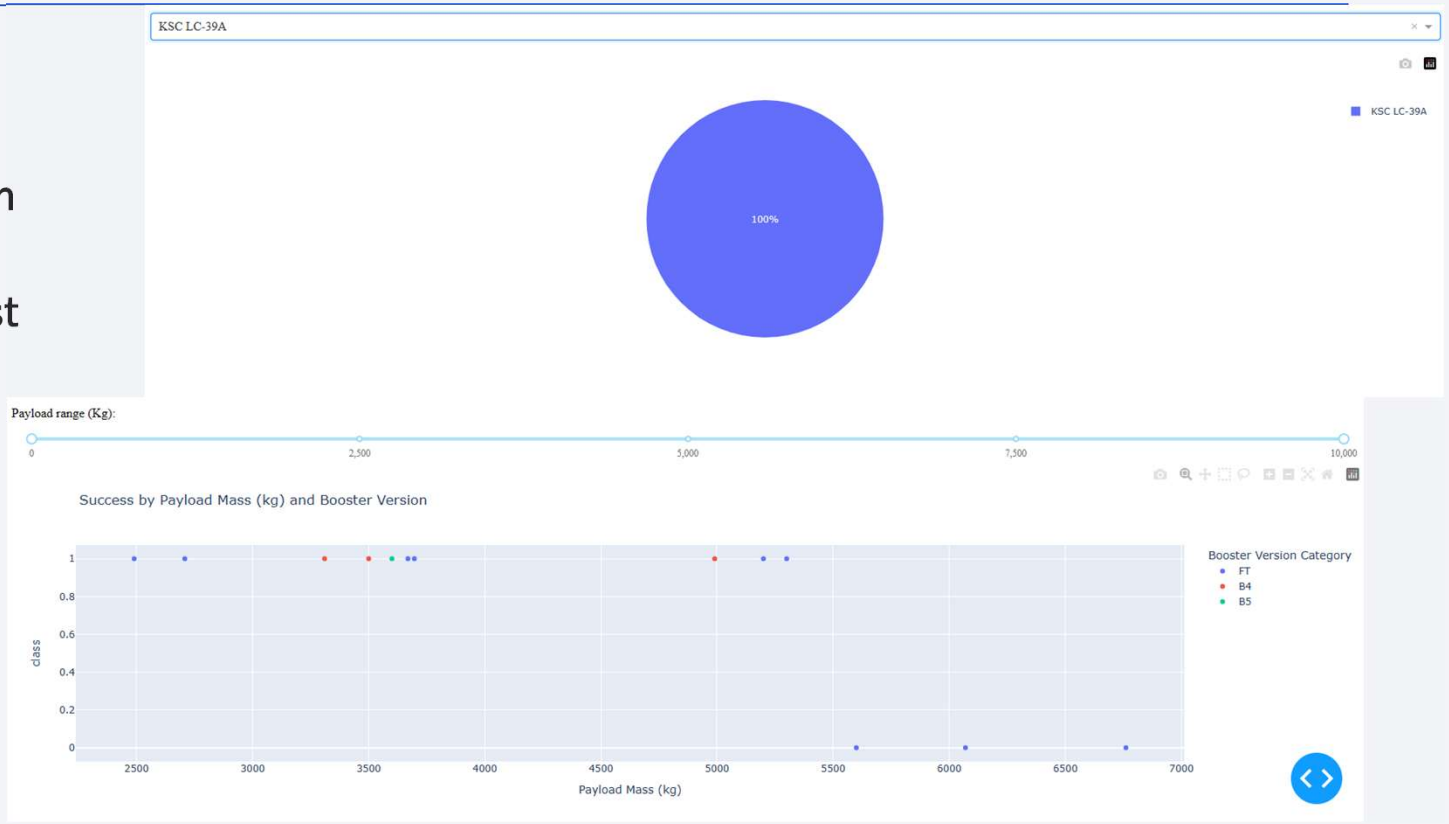
Successful Launches by Site – Pie Chart

- The Plotly dashboard contains a pie chart showing the total successful launches (as a share of all successes).



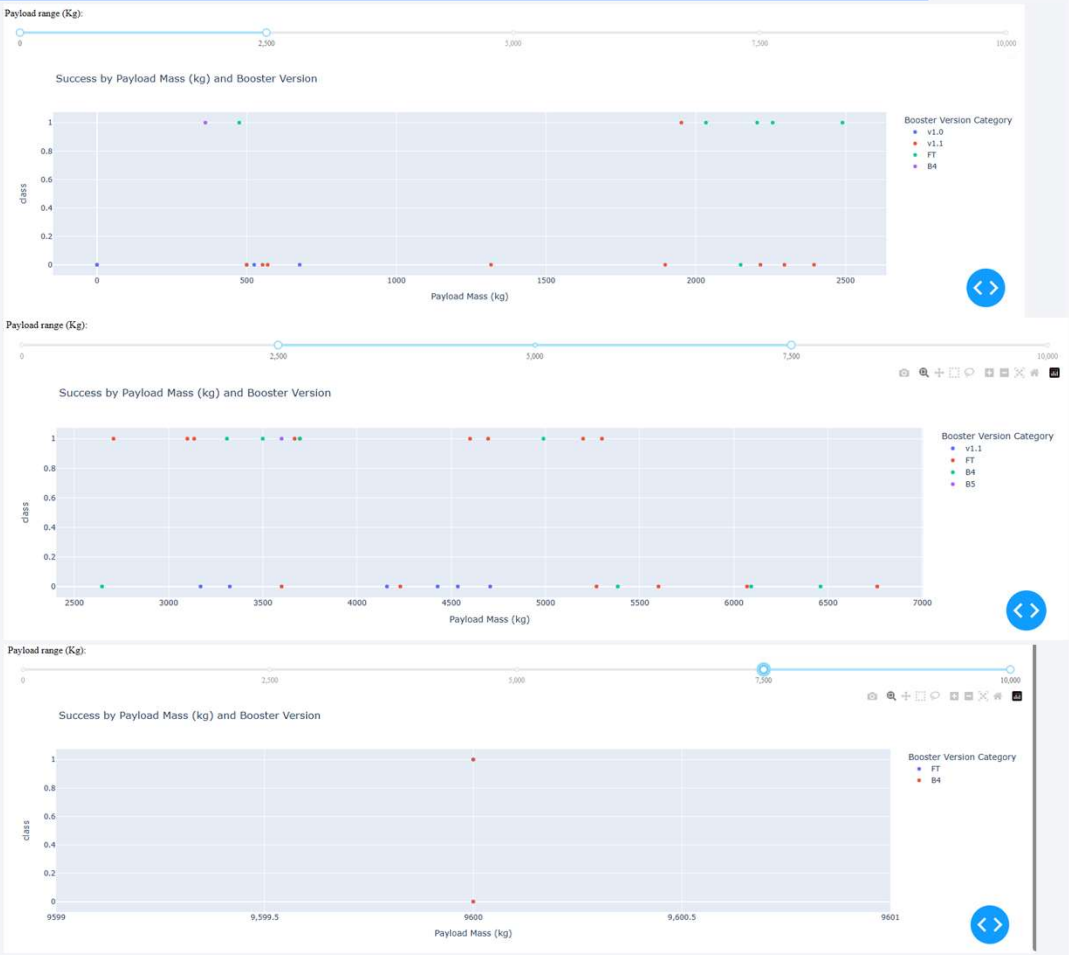
Launch Site with Most Successes

- The dash shows that KSC LC-39A had both the most total successes and highest success rate.



<Dashboard Screenshot 3>

- The dashboard shows most payloads in the 5-7.5 kg range and the FT Falcon 9 boosters having the highest success rate in that payload mass range.

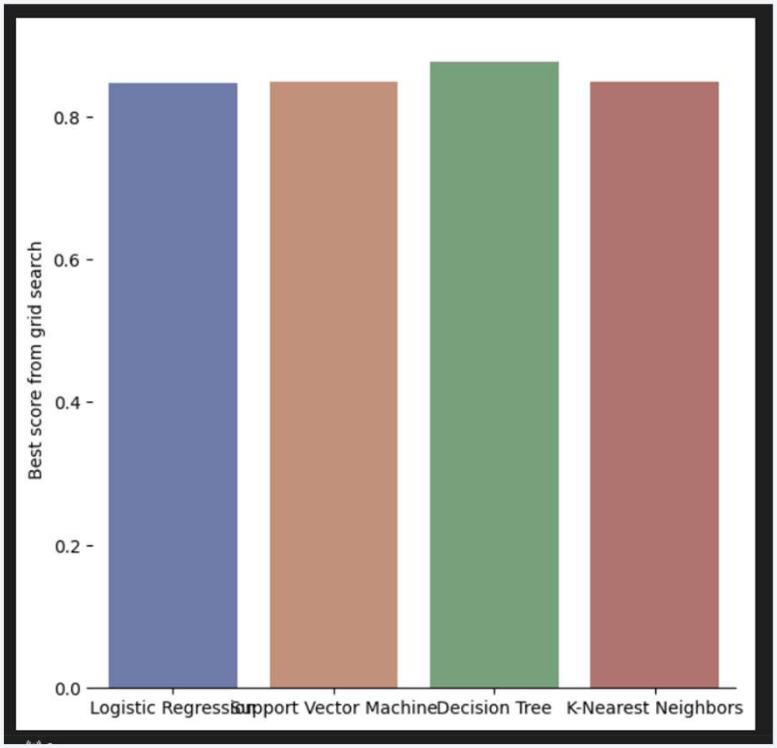


Section 5

Predictive Analysis (Classification)

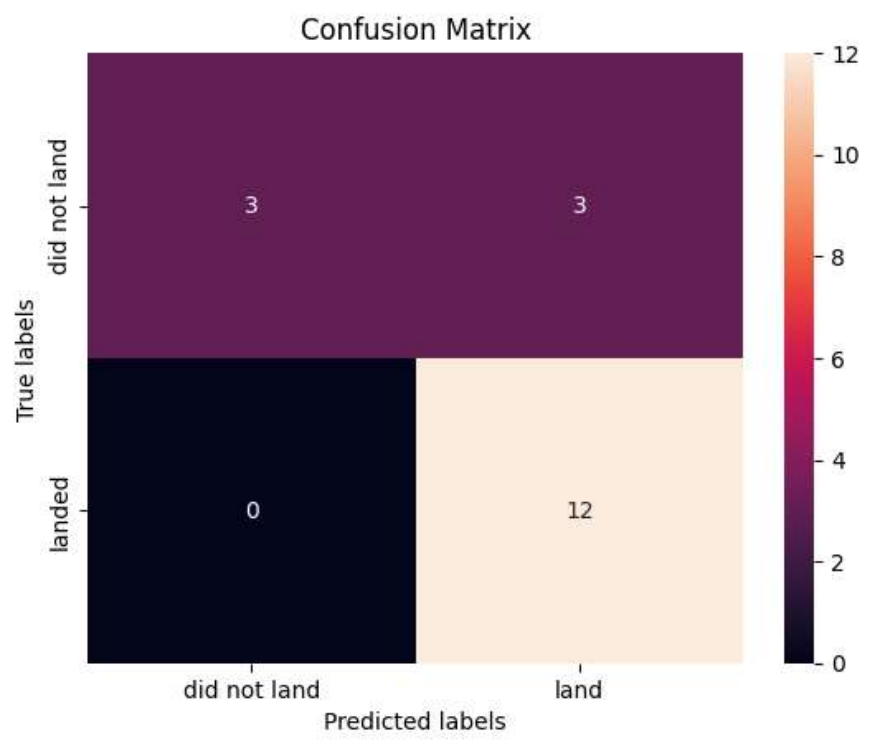
Classification Accuracy

- All models had identical confusion matrices with an accuracy of 83.33%, so instead the visual depicts the best score for a single cross-validation fold for each model, which gives a slight edge to the Decision Tree model.



Confusion Matrix

- All of the classifiers tested showed a recall of 100% (12/12 actual landings correctly classified), but only 80% precision (12/15 predicted landings actually landed).



Conclusions

- Important features include payload mass, orbit type, and date/flight number.
- The classifiers tested proved better at predicting actual landings (recall) than avoiding false positives (precision).
- All classifiers provide decent predictive power, with a decision tree model having a slight edge, on this small dataset. More launch data in the future should improve the model.

Thank you!