IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Gianni Bisante
August 5, 2024

Outline

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

Executive Summary

- The overarching methodology for the all analyses of the project was the EDA method. Which, was carried out using the Python and SQL languages. In sequence of execution: data preparation, calculation of targeted values, data visualization, and predictive analytics.
- Results are presented with screenshots of both code and visualizations.

Introduction

- The primary purpose of the project is to demonstrate the skills learned through this IBM Data Science Professional Certificate Course.
- The project's target for application is SPACEX rocket data.
- Areas of interest:
  - ❑ What are the contributing factors of successful mission take-offs and landings at specified launch sites?
  - ❑ Which launch sites had the highest percentage of successful missions?
  - ❑ Can a model be constructed to predict correct mission outcomes accurately?

Section 1

Methodology

Methodology

Executive Summary

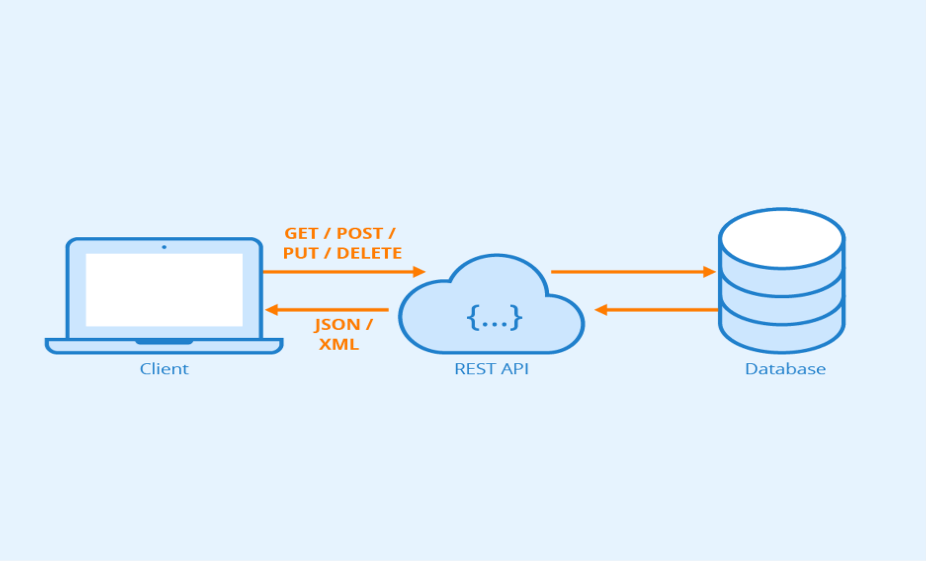
- Data collection methodology:

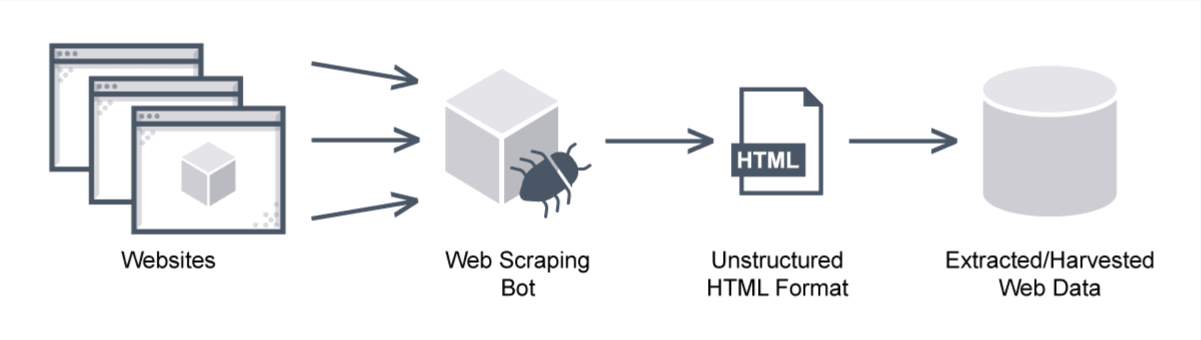
  SpaceX REST API, Web Scraping

- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

- Data sets were collected from the SpaceX API, which included: launches, rocket model, cargo delivery, launch information, landing information, mission results.





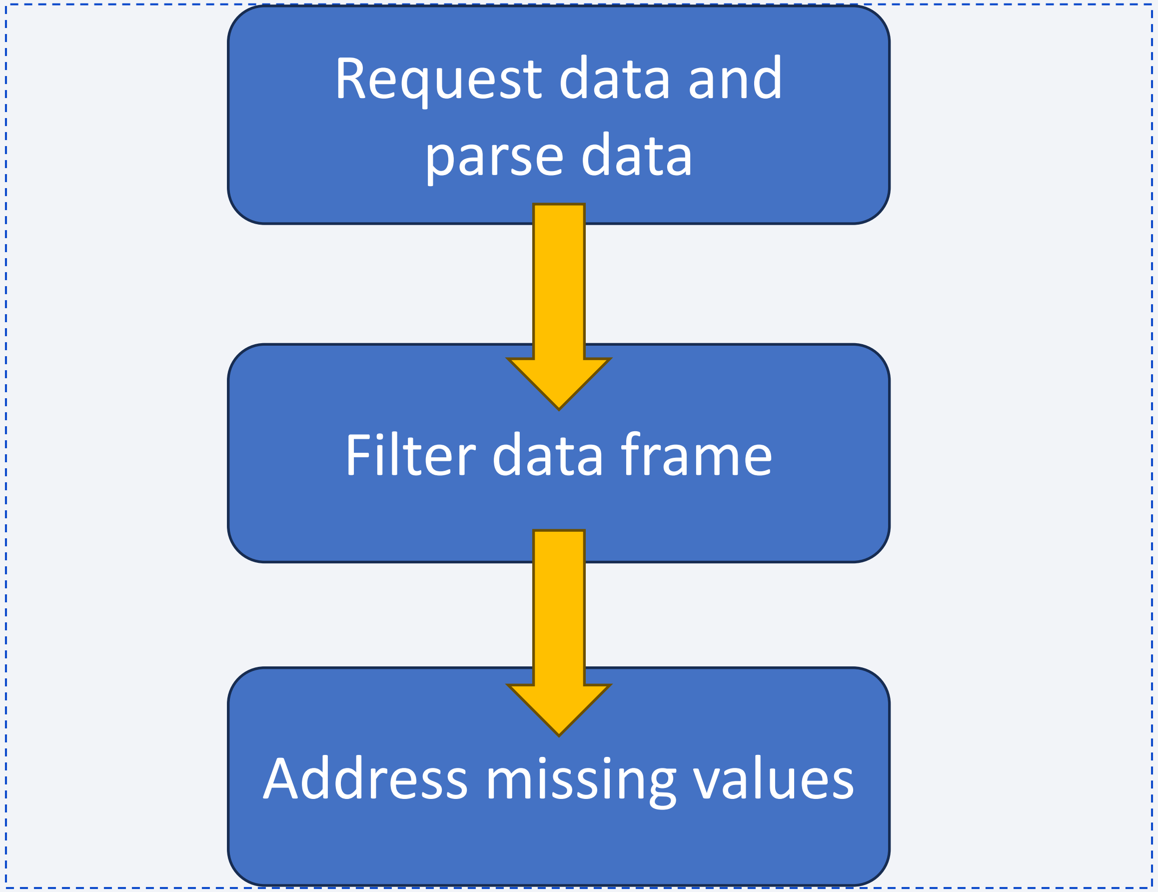
- Data sets were collected from the SpaceX API, which included: launches, rocket model, cargo delivery, launch information, landing information, mission results.

Data Collection – SpaceX API

- Data was gathered using the “requests” Python library from a given URL.
- API requests come in the form of a json file.

➢ Notebook reference

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN
4	6	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0

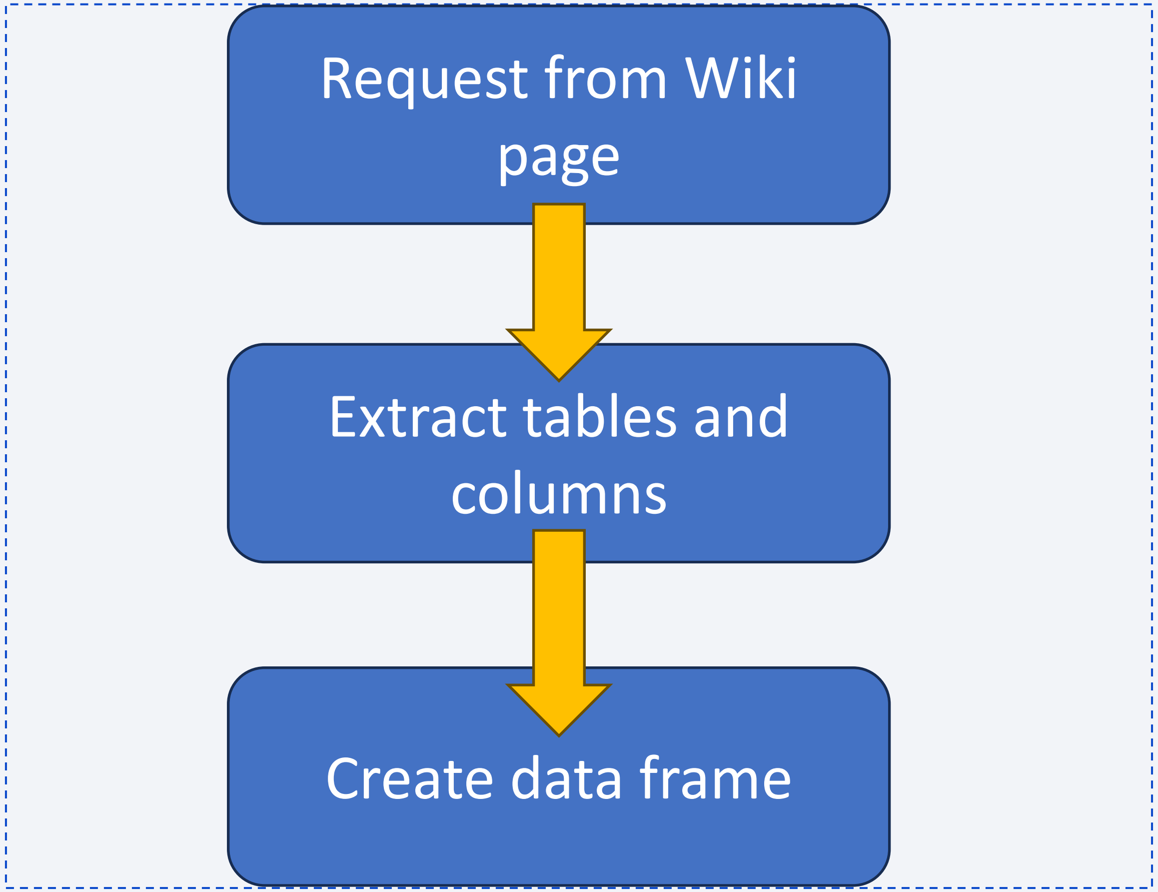


Data Collection – Scraping

- Data was gathered using the “requests” Python library from a given URL.
- Use the BeautifulSoup library to parse html data.

➢ Notebook reference

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	4 June 2010
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0004.1	Failure	8 December 2010	8 December 2010
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	22 May 2012
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	8 October 2012
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	1 March 2013
...	...	...	...	...	...	...	...	...	...	...	...
116	117	CCSFS	Starlink	15,600 kg	LEO	SpaceX	Success\n	F9 B5B1051.10	Success	9 May 2021	9 May 2021
117	118	KSC	Starlink	~14,000 kg	LEO	SpaceX	Success\n	F9 B5B1058.8	Success	15 May 2021	15 May 2021
118	119	CCSFS	Starlink	15,600 kg	LEO	SpaceX	Success\n	F9 B5B1063.2	Success	26 May 2021	26 May 2021
119	120	KSC	SpaceX CRS-22	3,328 kg	LEO	NASA	Success\n	F9 B5B1067.1	Success	3 June 2021	3 June 2021
120	121	CCSFS	SXM-8	7,000 kg	GTO	Sirius XM	Success\n	F9 B5	Success	6 June 2021	6 June 2021



Data Wrangling

- Using the Pandas library the data frame is created and using the Numpy library to gather statistics from the data.
- The prepared data was analyzed checking for insights on specific variables.

➢ Notebook reference

Sort and count unique launch sites

✓ 3 Launch sites

Calculate the number and frequency of orbits

✓ 11 orbits

Calculate number and occurrence of mission outcomes

✓ 8 different outcomes

EDA with Data Visualization

- Visualizations were made using the Matplotlib and Seaborn libraries.
- Using variables such as payload mass, launch site, orbit type, and year affects the landing success of the first stage of the Falcon 9 rocket.

➢ Notebook reference

EDA with SQL

- Using SQL queries in a python environment by creating a local SQLite database.
- With the queries, analysis was conducted on launch sites, payload mass, and mission landing outcomes.

➢ Notebook reference

Build an Interactive Map with Folium

- The Folium library allows the creation of interactive maps. On the maps markers were created to indicate launch sites along with their corresponding success rates. Other factors including distance between railways, highways, coastlines or cities were calculated and plotted.

- Notebook reference

Build a Dashboard with Plotly Dash

- Following the Dash framework, an interactive dashboard was created with the assistance of the Plotly library.
- The dashboard presents the potential effectiveness of a mission depending on its starting point.
- Starting with the launch site, a slider was also made to set a payload mass for the mission, and how it affects mission status.

➢ File reference

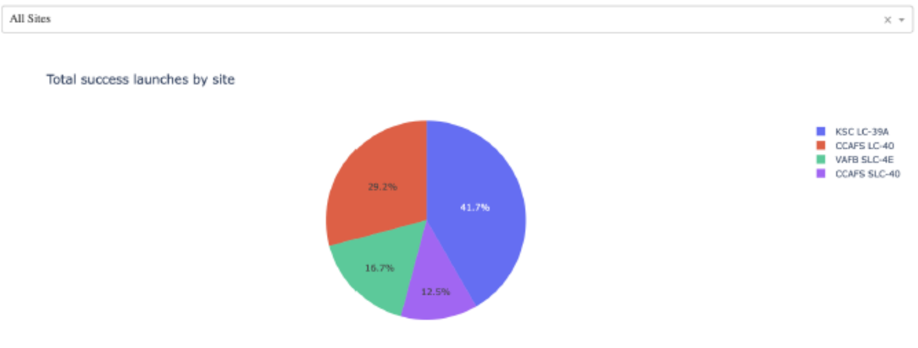
Predictive Analysis (Classification)

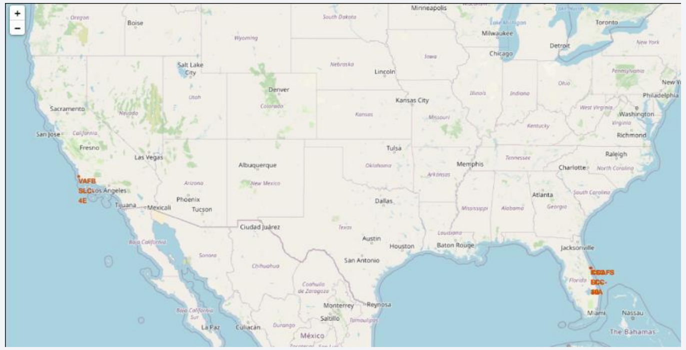
- The Scikit-learn library was used for the predictive analytics, while matplotlib and seaborn was used for visualizations.
- The main issue to be addressed was the proper classification of the data.
- Data was loaded and processed before being divided into training and test sets.
- Calculations were executed with a grid search, continuously adjusting for the best parameters for each model.
- Finally, all models were compared for accuracy.

➢ Notebook reference

Results

- Exploratory analysis determined the relationship between multiple variables and the mission outcomes of the Falcon 9 rocket's landings. Python libraries Numpy, Pandas, in conjunction with SQL was used for data analysis and manipulation. While, Python libraries Matplotlib and Seaborn visualized the analysis.
- Interactive analytics images:



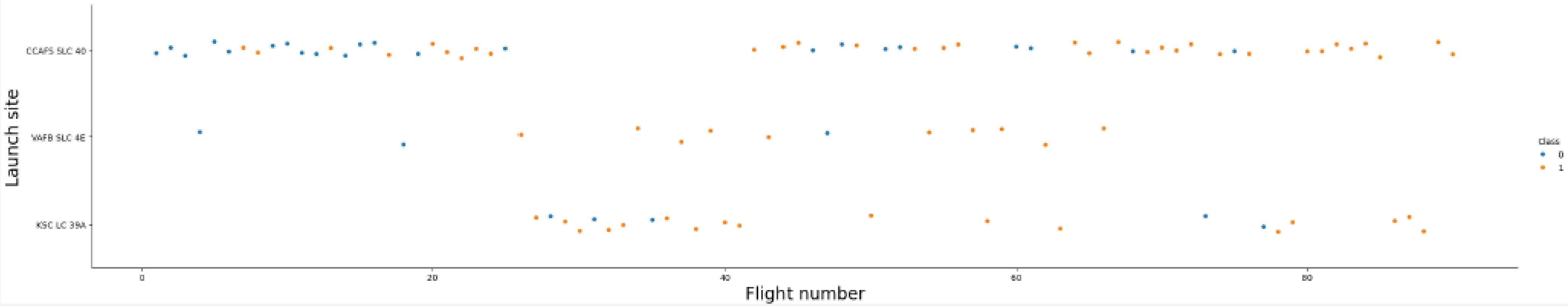


- The scikit-learn libraries machine learning algorithms laid the foundations of the predictive models, which classify and predict the success of the Falcon 9 landings.

Section 2

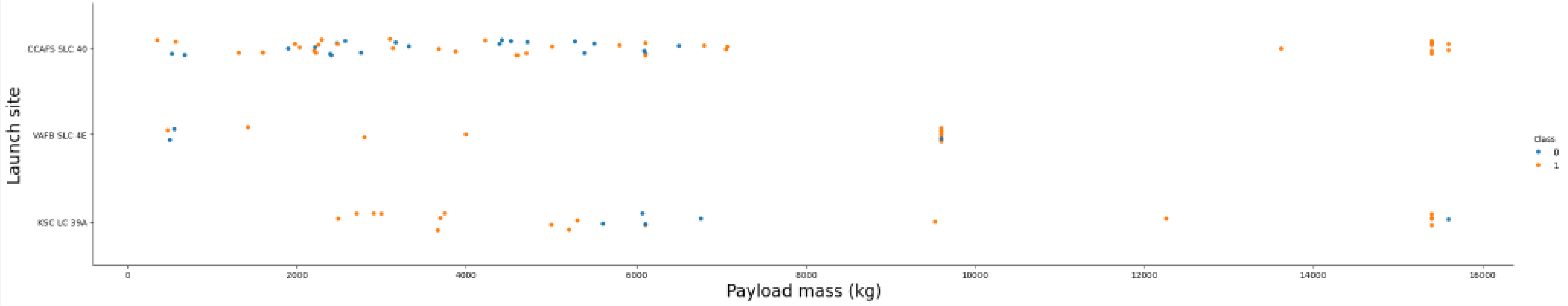
Insights drawn from EDA

Flight Number vs. Launch Site



- ❖Most missions emerged from launch site CCAFS SLC 40, launch site VAFB SLC 4E was the opposite.
- ❖Launch site CCAFS SLC 40 shows the higher flight numbers had higher probabilities of success.
- ❖While the former cannot be concluded from the other launch sites due to less observations, most missions were successful.

Payload vs. Launch Site



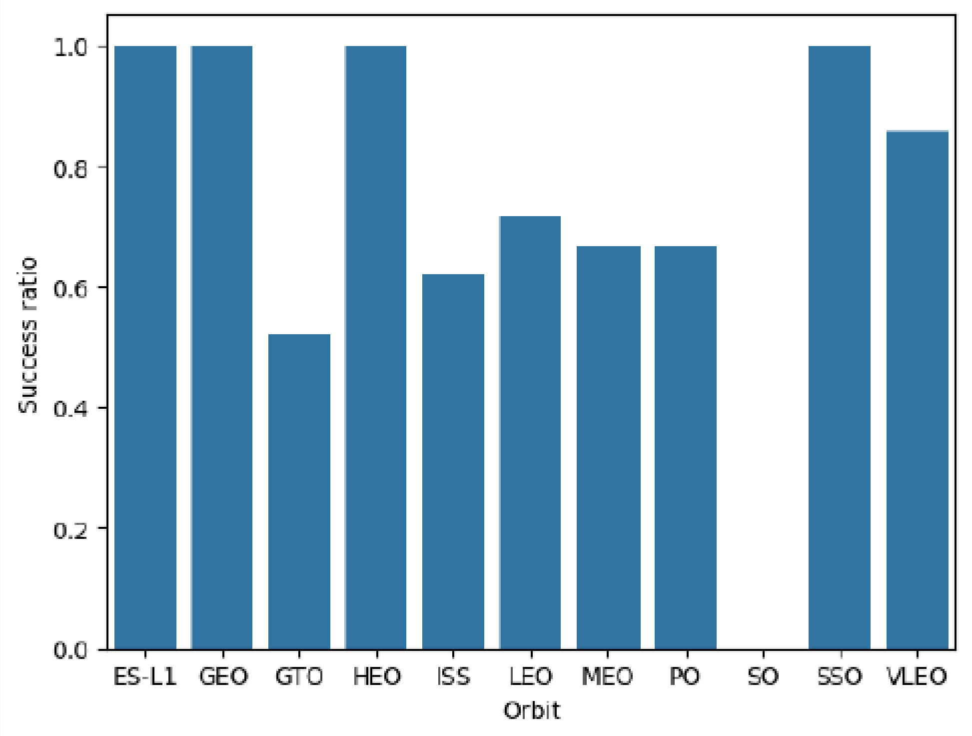
❖Most missions have a payload mass less than or equal to 7,000kg.

❖Overall, between all sites, larger payloads were more successful – greater than 7,000kg.

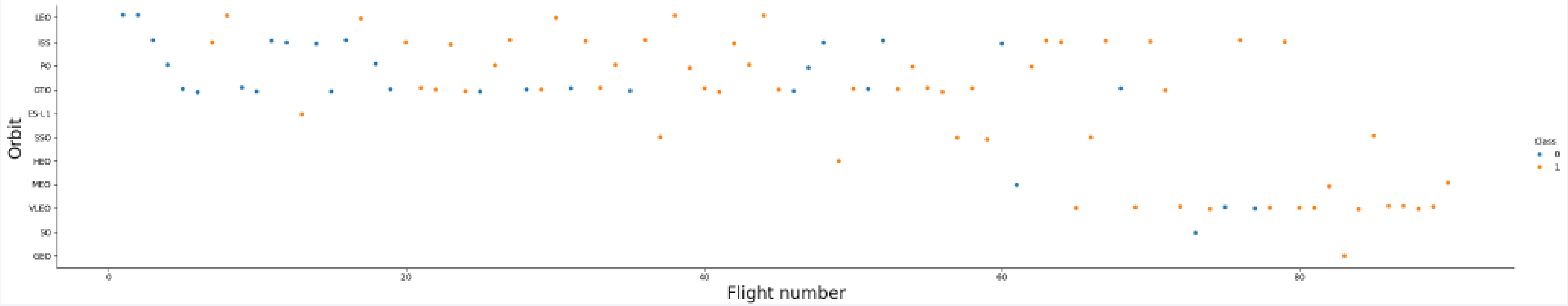
❖However, launch site KSC LC 39A also had high success rates among lighter payloads – sub-6,000kg.

Success Rate vs. Orbit Type

- ❖Orbits ES-L1, GEO, HEO, and SSO had observable 100% success ratios.
- ❖The least successful orbit was GTO hovering at 50%
- ❖ISS, LEO, MEO, PO sat between 60-70% success rates.
- ❖It is inconclusive whether orbital altitude affects mission success.

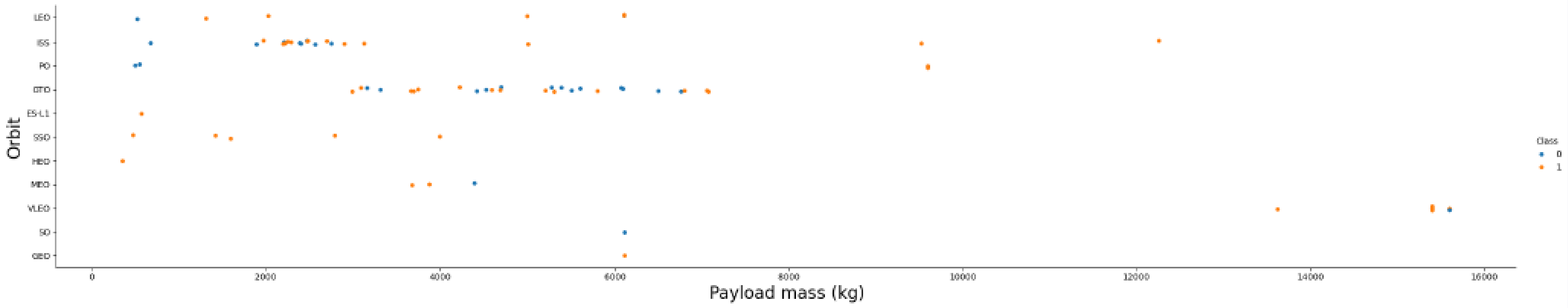


Flight Number vs. Orbit Type



- ❖Most missions occurred in orbits: GEO, SSO, HEO, MEO, VLEO and SO.
- ❖Among the most frequent orbits, mission numbers 20 and below seem to be the least successful.
- ❖The orbits with 100% success rates seem to be skewed due to low observation count.

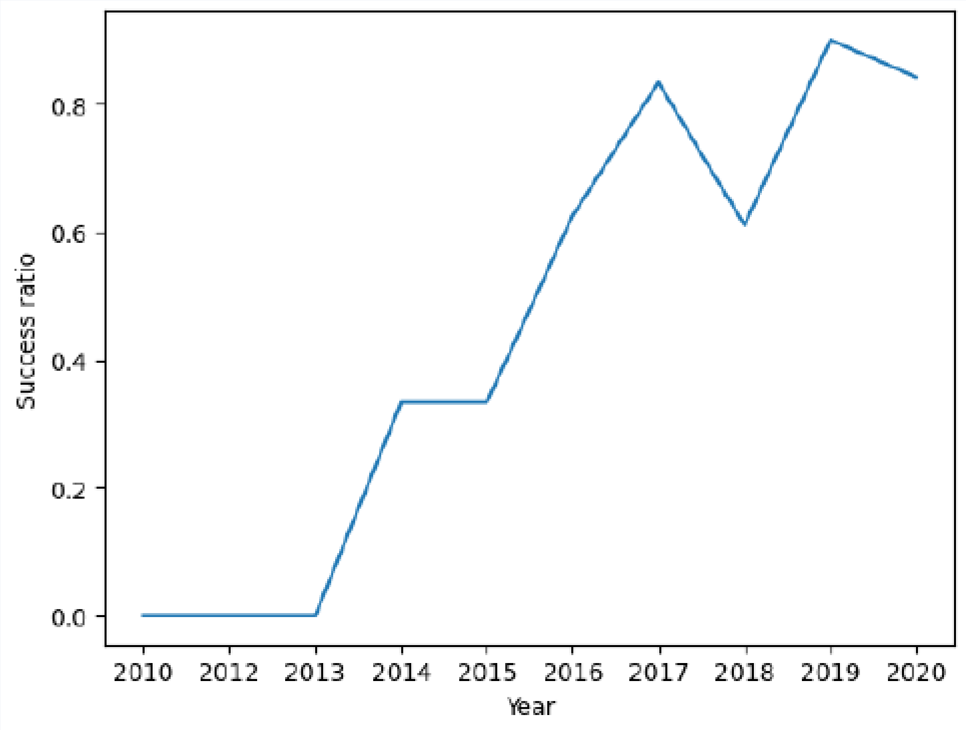
Payload vs. Orbit Type



❖Most orbit payloads were equal or less than 7,000kg.

❖While some orbits have observable typical ranges for payload mass, there seems to be no correlation between the variables with regards to mission success.

Launch Success Yearly Trend



- ❖Visible trend of increasing success rate throughout the years.
- ❖2013-2014 observed largest single year increase in success rate.
- ❖Approximate 20% dip in success rate from 2017-2018.
- ❖Peak success rate in 2019 at approximately 90%.

All Launch Site Names

❖4 different launch sites in data set.

```
%sql SELECT DISTINCT(LAUNCH_SITE) from SPACEXTBL;
```

```
 * 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'

- ❖The first 5 records of launches beginning with 'CCA' are all related to site CCAFS LC-40.
- ❖All launches had target orbit LEO.
- ❖The first customer was SpaceX, the preceding four launches customer was NASA
- ❖All successfully launched, the first two failed at landing, the following three did not attempt.

```
%sql SELECT *from SPACEXTBL where (LAUNCH_SITE) LIKE 'CCA%' LIMIT 5;
```

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

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

Total Payload Mass

❖Total payload mass carried out by booster from NASA was 48,213kg.

```
%sql select sum(PAYLOAD_MASS__KG_) as 'Total payload mass' from SPACEXTBL where CUSTOMER like '%NASA (CRS)%';
```

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

Total payload mass
48213

Average Payload Mass by F9 v1.1

❖Average payload mass carried by booster version F9 v1.1 was 2928.4kg.

```
%sql select avg(PAYLOAD_MASS__KG_) as ' Average payload mass' from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1';
```

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

Average payload mass
2928.4

First Successful Ground Landing Date

❖First successful ground landing was on December 22, 2015.

```
%sql select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)';

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

min(DATE)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

❖Four different boosters successfully landed on the drone ship and had payload mass greater than 4000 but less than 6000.

```
%sql select BOOSTER_VERSION from SPACEXTBL where Landing_Outcome='Success (drone ship)' and PAYLOAD_MASS__KG_ BETWEEN 4000 a
```

```
 * 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

❖Only one mission failed to launch, the other one-hundred were successful.

```
%%sql
SELECT MISSION_OUTCOME,
        COUNT(*) as 'Total Numbner of Outcomes'
    FROM SPACEXTABLE
 GROUP BY MISSION_OUTCOME;
```

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

Mission_Outcome	Total Numbner of Outcomes
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

❖Twelve different versions of the booster carried the maximum payload.

```
%sql select BOOSTER_VERSION as 'Booster version' from SPACEXTBL where PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG_) from
```

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

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

- Only two missions failed the drone ship landing in 2015: January and April.
- Both launched from site CCAFS LC-40.
- They used different booster versions.

```
%%sql
SELECT CASE
            WHEN SUBSTR(`Date`, 6, 2) = '01'
            THEN 'January'
            WHEN SUBSTR(`Date`, 6, 2) = '02'
            THEN 'February'
            WHEN SUBSTR(`Date`, 6, 2) = '03'
            THEN 'March'
            WHEN SUBSTR(`Date`, 6, 2) = '04'
            THEN 'April'
            WHEN SUBSTR(`Date`, 6, 2) = '05'
            THEN 'May'
            WHEN SUBSTR(`Date`, 6, 2) = '06'
            THEN 'June'
            WHEN SUBSTR(`Date`, 6, 2) = '07'
            THEN 'July'
            WHEN SUBSTR(`Date`, 6, 2) = '08'
            THEN 'August'
            WHEN SUBSTR(`Date`, 6, 2) = '09'
            THEN 'September'
            WHEN SUBSTR(`Date`, 6, 2) = '10'
            THEN 'October'
            WHEN SUBSTR(`Date`, 6, 2) = '11'
            THEN 'November'
            WHEN SUBSTR(`Date`, 6, 2) = '12'
            THEN 'December'
        END AS `Month name`,
        `Landing_Outcome`,
        `Booster_Version`,
        `Launch_Site`
  FROM SPACEXTABLE
 WHERE `Landing_Outcome` LIKE '%Failure (drone ship)%'
   AND SUBSTR(Date, 0, 5) = '2015';
```

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

Month name	Landing_Outcome	Booster_Version	Launch_Site
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

❖Most missions during the time period made no attempt to land.

❖***note:::notebook refused to process the code correctly for some reason***

❖Correct order:

❖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.

```
%%sql
SELECT Landing_Outcome,
    COUNT(*) AS 'outcome_count'
FROM SPACEXTABLE
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY Landing_Outcome
ORDER BY 'outcome_count' DESC;
```

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

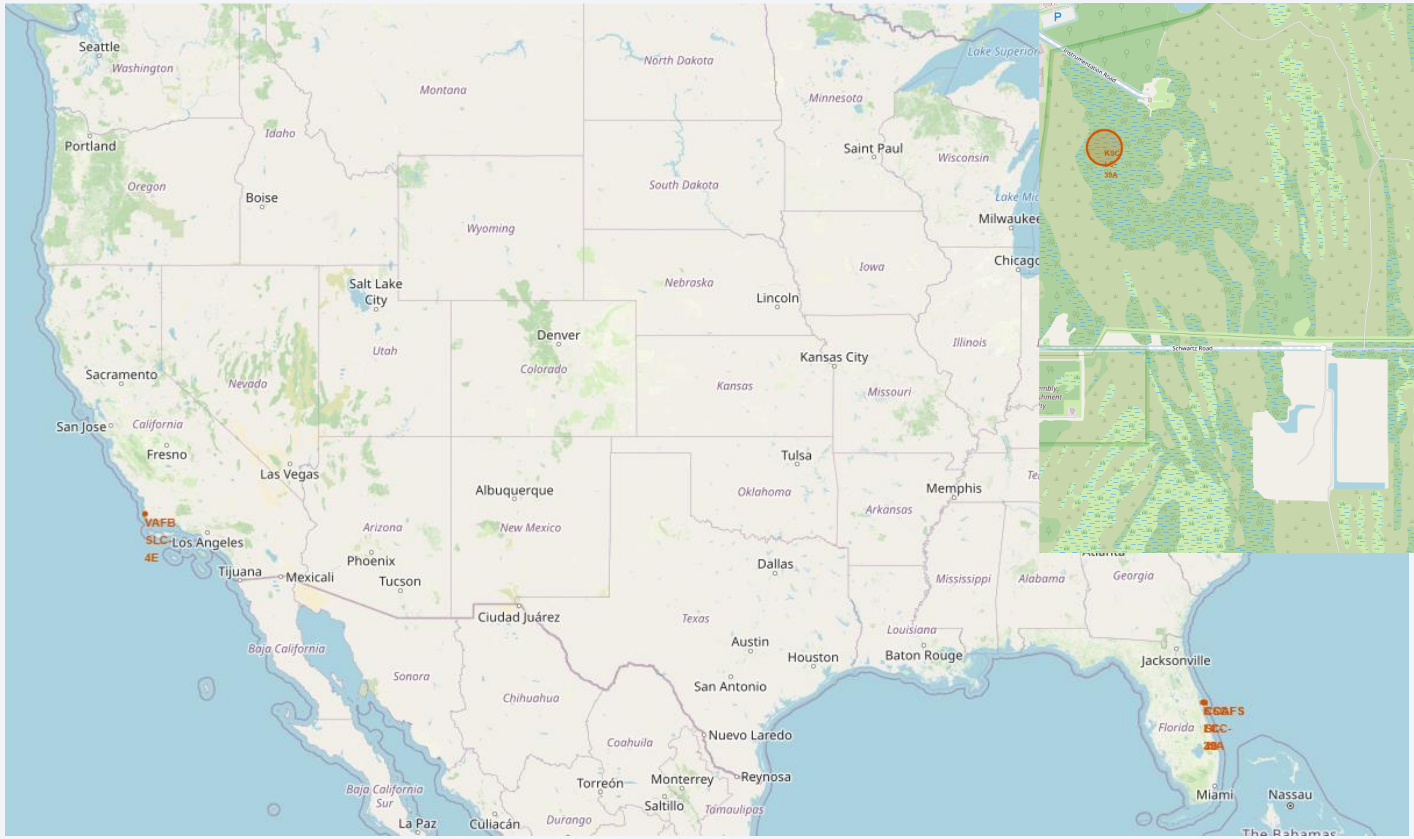
Landing_Outcome	outcome_count
Uncontrolled (ocean)	2
Success (ground pad)	3
Success (drone ship)	5
Precluded (drone ship)	1
No attempt	10
Failure (parachute)	2
Failure (drone ship)	5
Controlled (ocean)	3

Section 3

Launch Sites Proximities Analysis

All Launch Sites

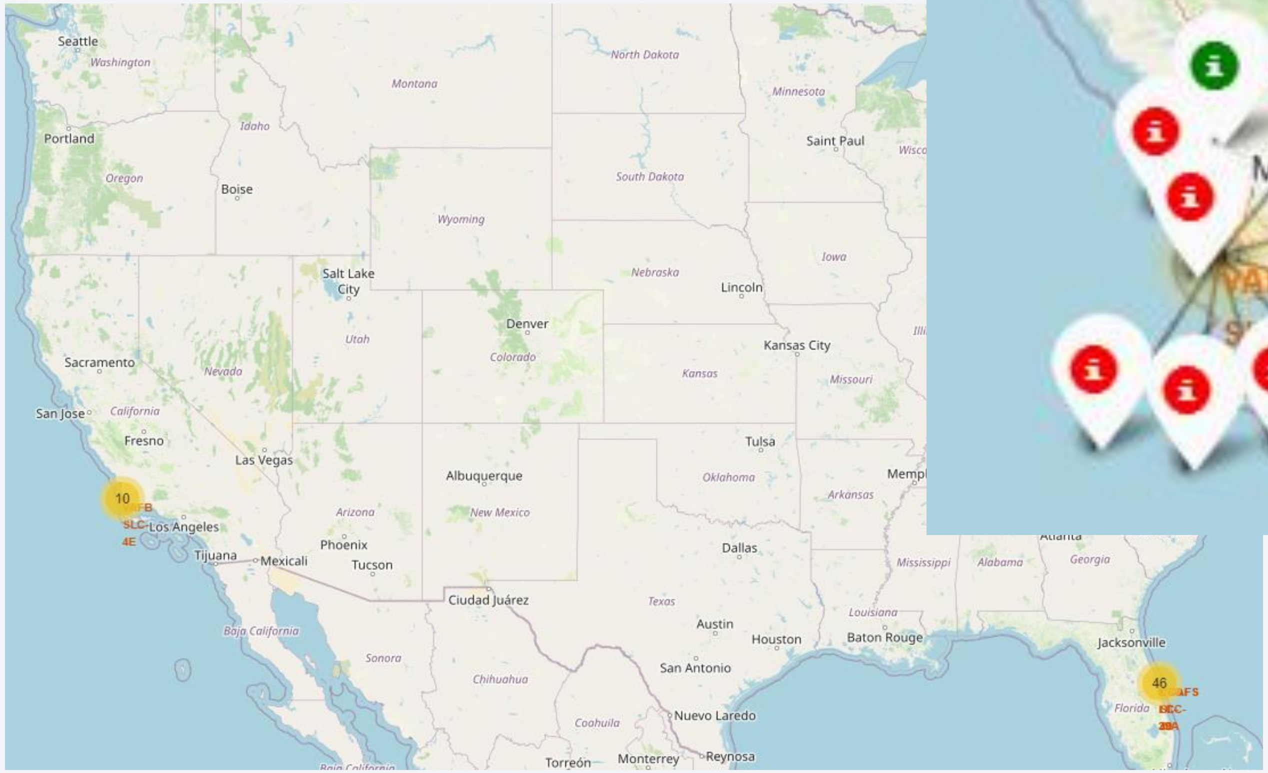
- ❖ Four launch sites: one in California and three in Florida.
- ❖*Second image is close-up of Florida launch sites.



Launch Outcomes

❖ Map is marked with marker clusters, which within them have other markers.

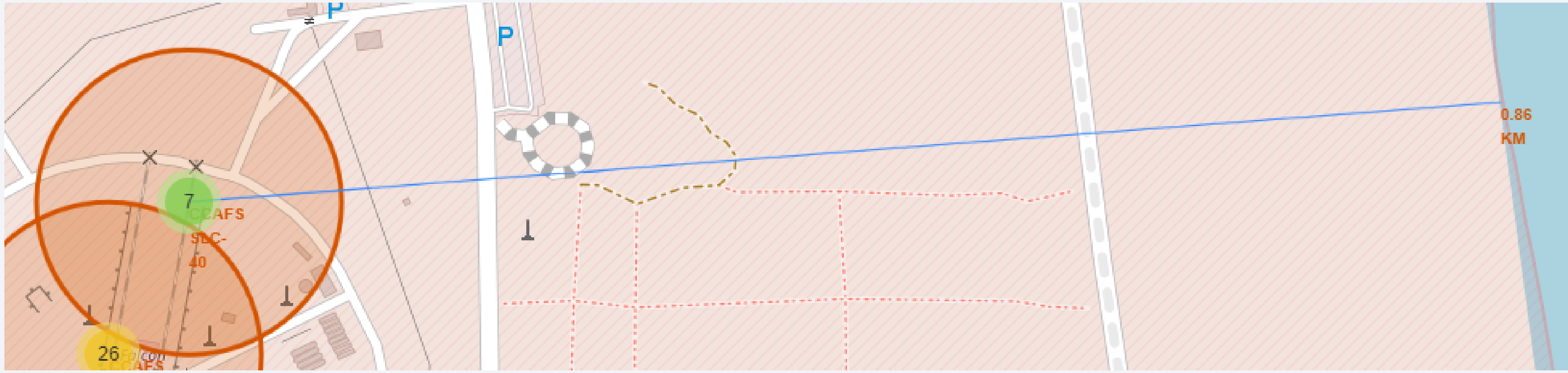
❖Green for successful and red for failed.



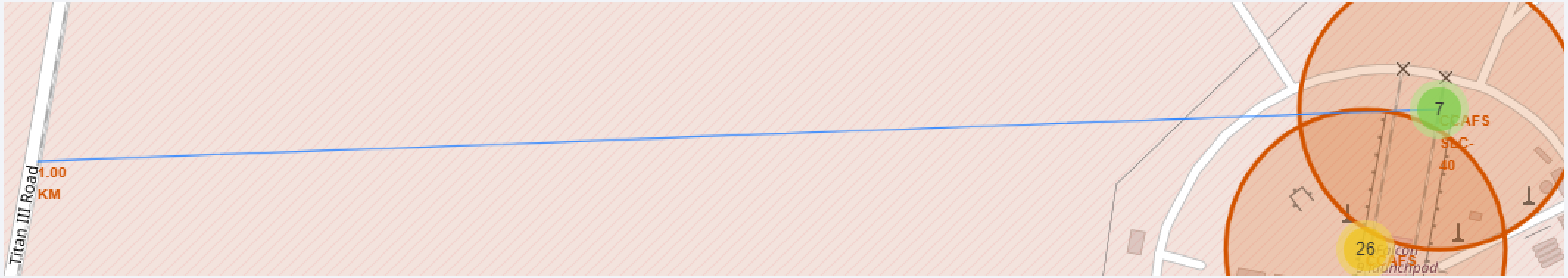
Distances to Other Points of Interest

❖These lines show distances to other points of interest.

❖As an example, the first image shows the launch site's distance to the coast.



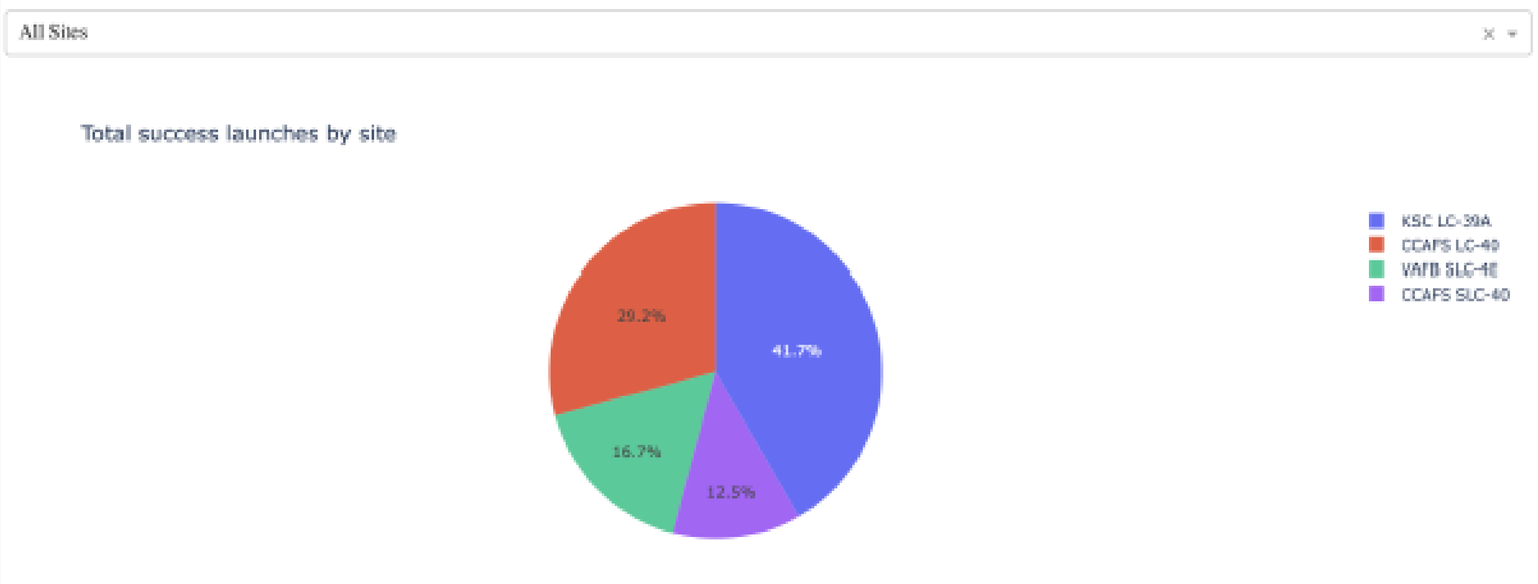
❖The second image shows the distance to a railroad.



Section 4

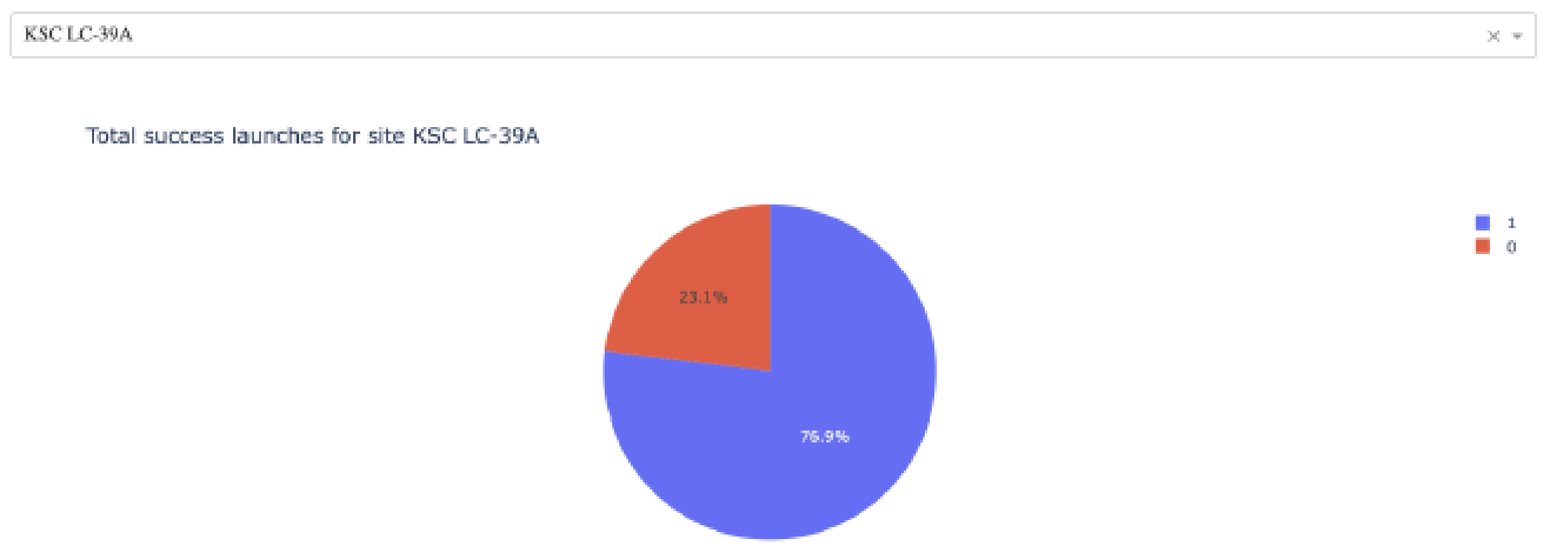
Build a Dashboard with Plotly Dash

Total Success of Launches by Site



- ❖Site KSC LC-39A has the most successful launches, accounting for 41.7% of successful launches.
- ❖While site CCAFS SLC-40 is the least accountable, at only 12.5%.

Launch Site with Highest Success Rate



❖KSC LC-39A also hast the highest success rate for launches at 76.9%.

Payload Launch Outcomes by Booster Versions for All Sites

- ❖Payload range set to 8,000kg due to observation limitations.
- ❖Most successful launches used an FT booster with a payload mass between 2,000-5,000kg.
- ❖Most failures used a v1.1 booster with a payload mass between 500-5,000kg.

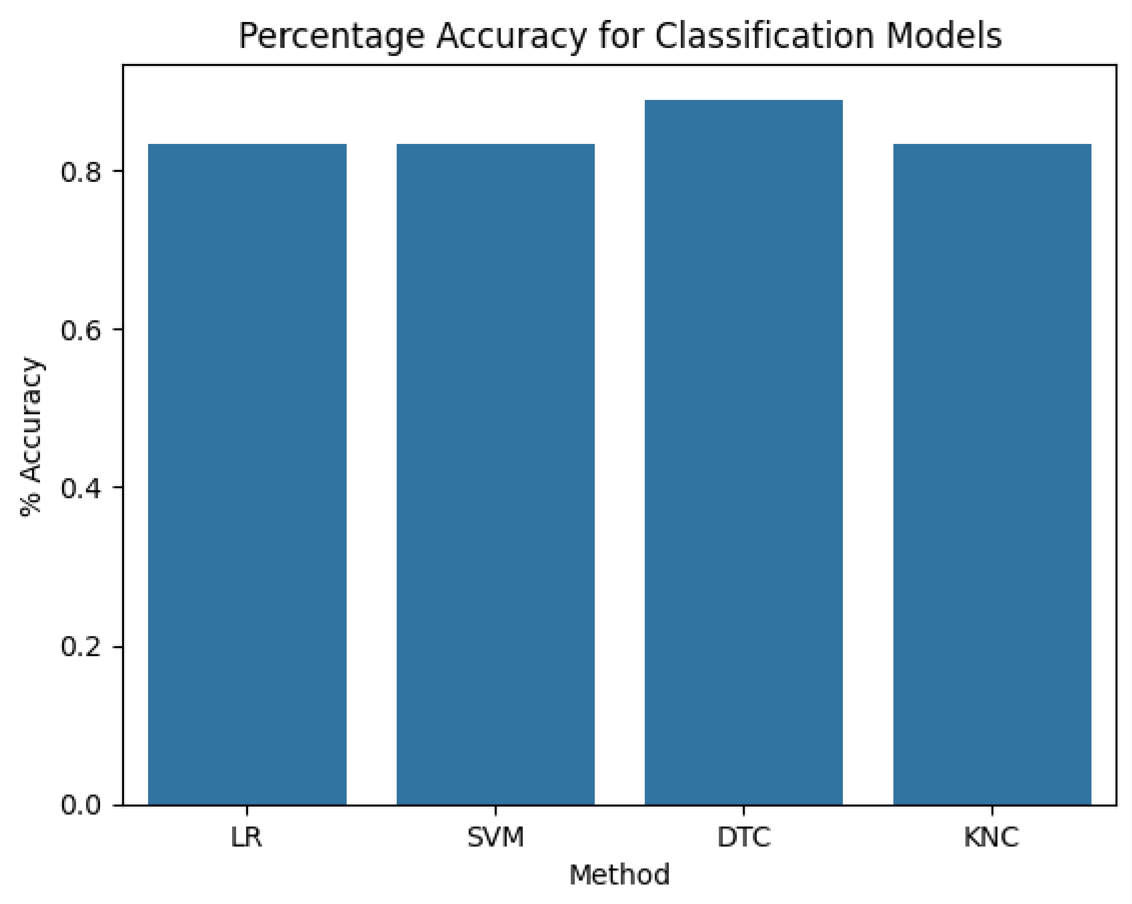
Section 5

Predictive Analysis (Classification)

Classification Accuracy

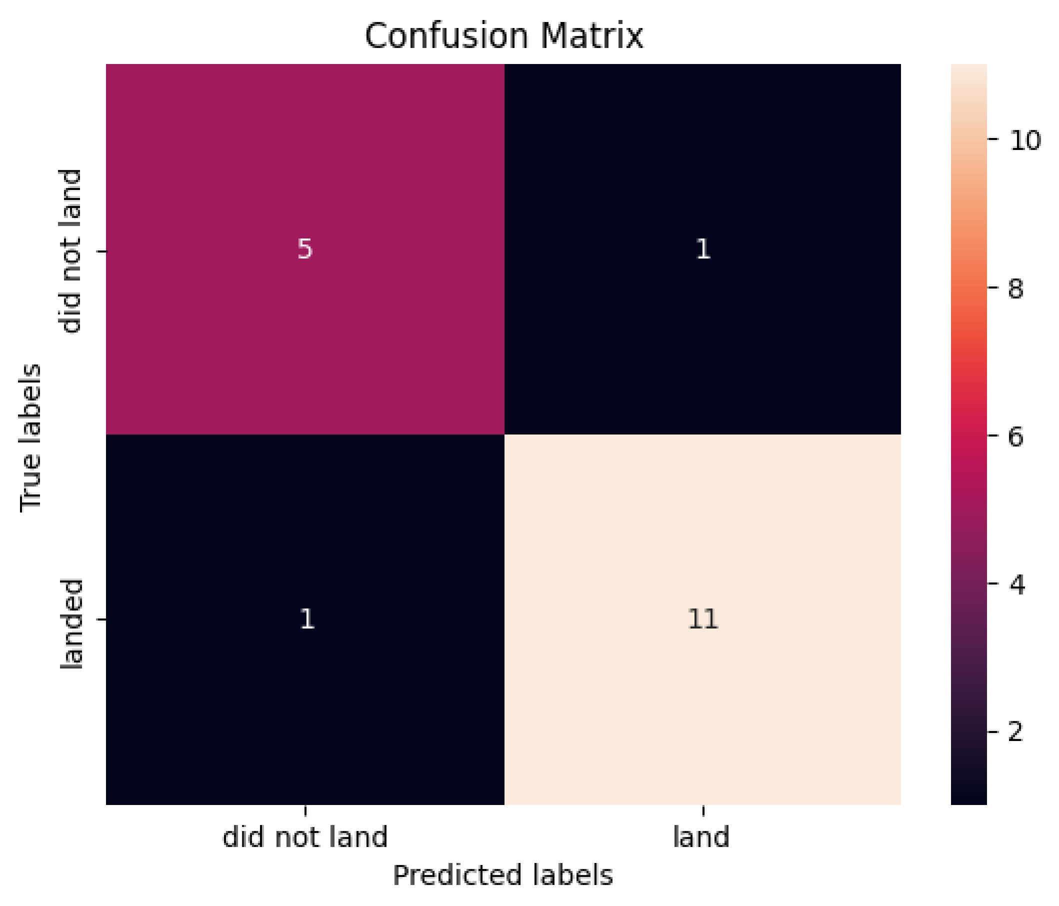
❖Most of the models had a prediction accuracy of approximately 83%.

❖The outlier, which is also the most accurate, was the decision tree model about 89%.



Confusion Matrix

- The decision tree model's confusion matrix shows most observations were predicted correctly using the model
- One observation was assigned as a false positive, top right box
- One observation was assigned as a false negative, bottom left box



Conclusions

- Larger payloads(> 8,000kg) found more success across all launch sites.
- Launch site KSC LC-39A was statistically the most successful launch site.
- Launch success rate improved as time passed.
- Orbit information was sometimes skewed, gradually specific orbits became more targeted.
- Launch sites are generally close to their relative coasts and far from cities.
- All prediction models were highly accurate and could be used as consistent predictors for launch outcomes.
- The decision tree model was the most accurate classifier.

Thank you!