IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Annie Lin>
<08/15/2024>

Outline

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

Executive Summary

- Summary of methodologies
  - SpaceX Data Collection with SpaceX API and Web Scraping
  - SpaceX Data Wrangling
  - SpaceX Data Analysis with SQL
  - SpaceX EDA Data Visualization
  - Analyze SpaceX Launching Sites with Folium
  - SpaceX Machine Learning Landing Predictions
- Summary of all results
  - Interactive Visual Analytics and Plotly Dash
  - EDA Results

Introduction

- We aim to predict if the first stage of SpaceX’s Falcon 9 will land successfully
- By determining whether the launch will land successfully we can determine the financial situation of the launch
- We will use data from the Falcon 9 launches from SpaceX’s website to predict the success of the first stage landing

Section
1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

- Requesting and analyzing data from SpaceX API
- By using data analysis tools we can load datasets and clean it to find helpful information
- Visualize data and extract meaningful patterns from it
- Build dashboard to analyze launch records interactively with Plotly Dash and use Folium to get interactive maps
- Build a predictive model from data

Data Collection – SpaceX API

- SpaceX REST API
  - Requesting and analyzing data from SpaceX API
- GitHub URL: https://github.com/nini2580/SpaceX_Falcon_9_Success_Landing/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	...	
89	86	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032
90	87	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032
91	88	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032
92	89	2020-10-24	Falcon 9	15600.0	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e3033
93	90	2020-11-05	Falcon 9	3681.0	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032

Data Collection - Scraping

- Web Scraping
  - By loading datasets and cleaning it we are able to find helpful information

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	[[SpaceX], \n]
1	2	CCAFS	Dragon	0	LEO	[[.mw-parser-output .plainlist ol,.mw-parser-o...
2	3	CCAFS	Dragon	525 kg	LEO	[[NASA], (, [COTS],)\n]
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	[[NASA], (, [CRS],)\n]
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	[[NASA], (, [CRS],)\n]

Launch outcome	Version Booster	Booster landing	Date	Time
Success\n	F9 v1.07B0003.18	Failure	4 June 2010	18:45
Success	F9 v1.07B0004.18	Failure	8 December 2010	15:43
Success	F9 v1.07B0005.18	No attempt\n	22 May 2012	07:44
Success\n	F9 v1.07B0006.18	No attempt	8 October 2012	00:35
Success\n	F9 v1.07B0007.18	No attempt\n	1 March 2013	15:10

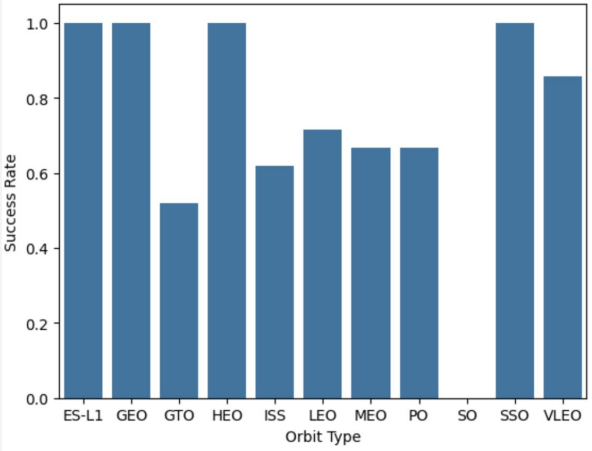
- GitHub URL: https://github.com/nini2580/SpaceX_Falcon_9_Success_Landing/blob/main/jupyter-labs-webscraping.ipynb

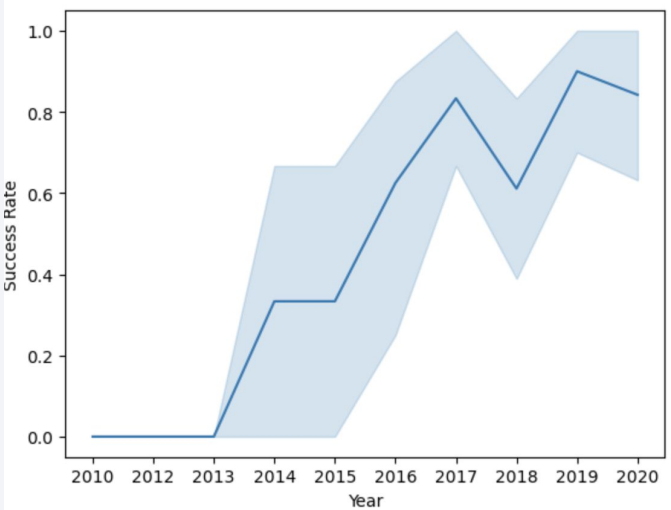
Data Wrangling

- Creating a data frame of tables and dealing with missing values using mean to replace the missing values
- Additionally, using EDA to look for informative patterns to help us come to a conclusion and determine training labels
- GitHub URL: https://github.com/nini2580/SpaceX_Falcon_9_Success_Landing/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Using scatter plots, bar plots, and line plots to visualize the data to determine the relationship between them
- GitHub URL: https://github.com/nini2580/SpaceX_Falcon_9_Success_Landing/blob/main/edadataviz.ipynb





EDA with SQL

- Use SQL to subset, group, and look at parts of the SpaceX data sets to understand the data in relation to one another
- GitHub URL: https://github.com/nini2580/SpaceX_Falcon_9_Success_Landing/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)';
```

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

SUM(PAYLOAD_MASS__KG_)
45596

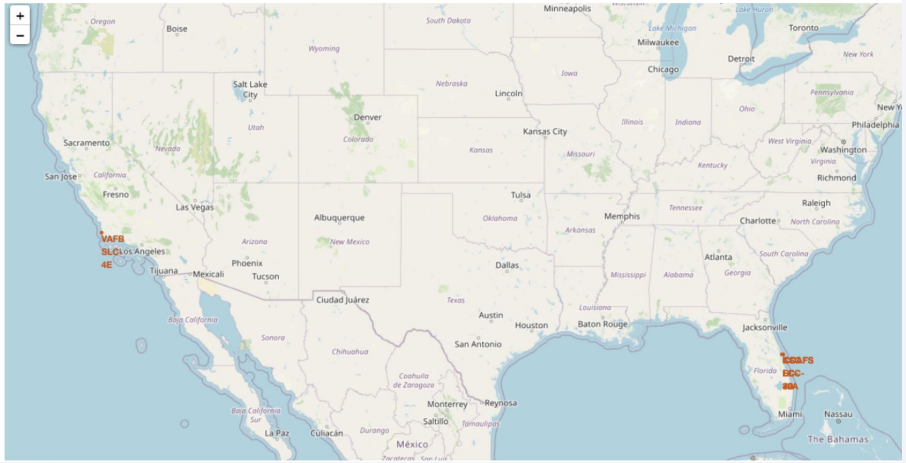
```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE BOOSTER_VERSION LIKE 'F9 v1.1%';
```

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

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

Build an Interactive Map with Folium

- Use Folium map and added markers, circles, lines, etc. to show where the launch sites were and marking their successes and failures
- GitHub URL: https://github.com/nini2580/SpaceX_Falcon_9_Success_Landing/blob/main/lab_jupyter_launch_site_location.ipynb



Build a Dashboard with Plotly Dash

- Build dashboards to analyze data interactively with Plotly Dash
    - Added dropdowns, range sliders, pie charts, scatter plots, etc.
- GitHub URL: https://github.com/nini2580/SpaceX_Falcon_9_Success_Landing/blob/main/spacex_dash_app.py

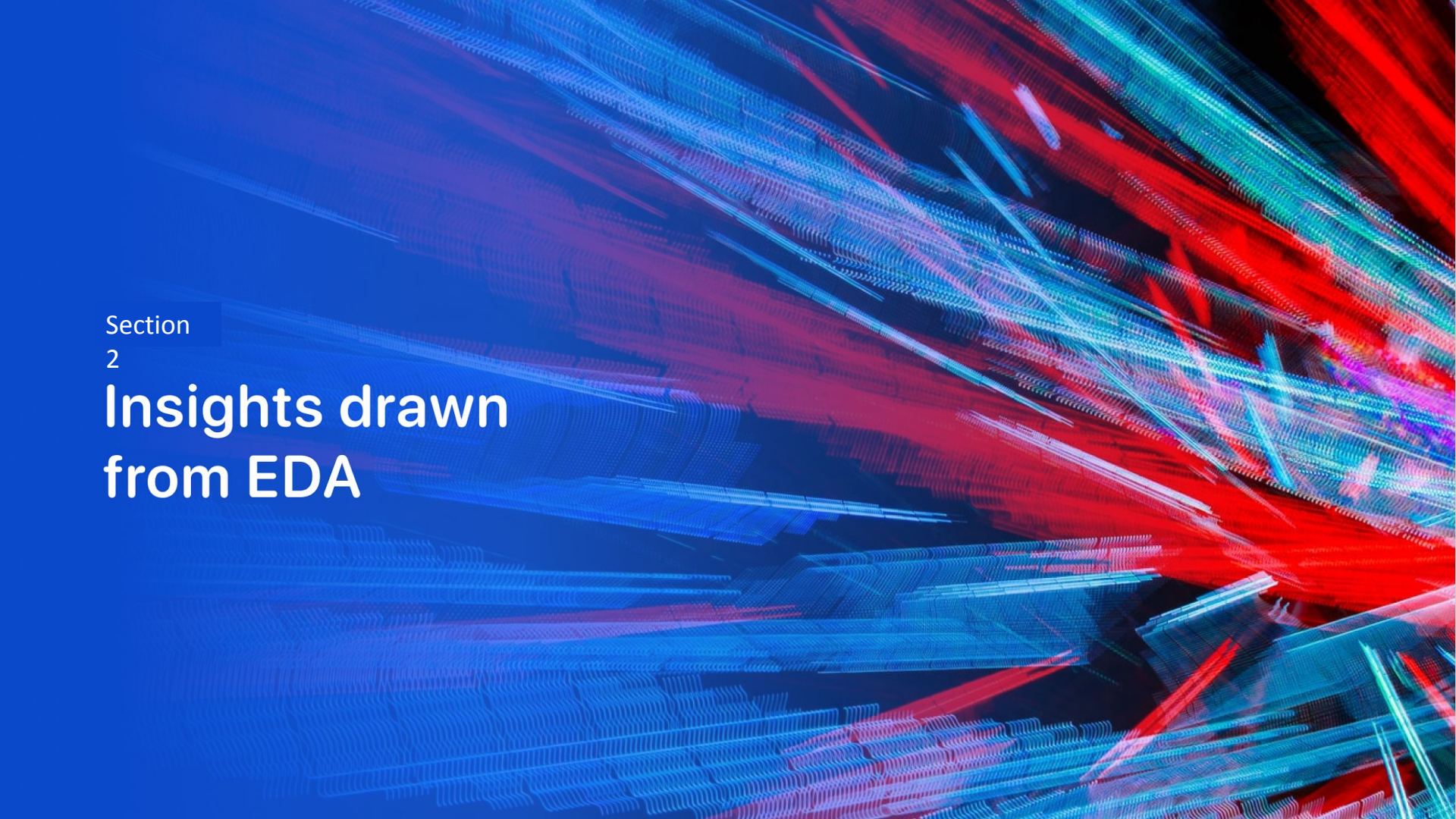
Predictive Analysis (Classification)

- Created columns, standardize data, and split them into training and testing sets
- Determine the best method to use from the data to get the best results
- Add the GitHub URL: https://github.com/nini2580/SpaceX_Falcon_9_Success_Landing/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

	0
Method	Test Data Accuracy
Logistic_Reg	0.833333
SVM	0.833333
Decesision Tree	0.833333
KNN	0.833333

Results

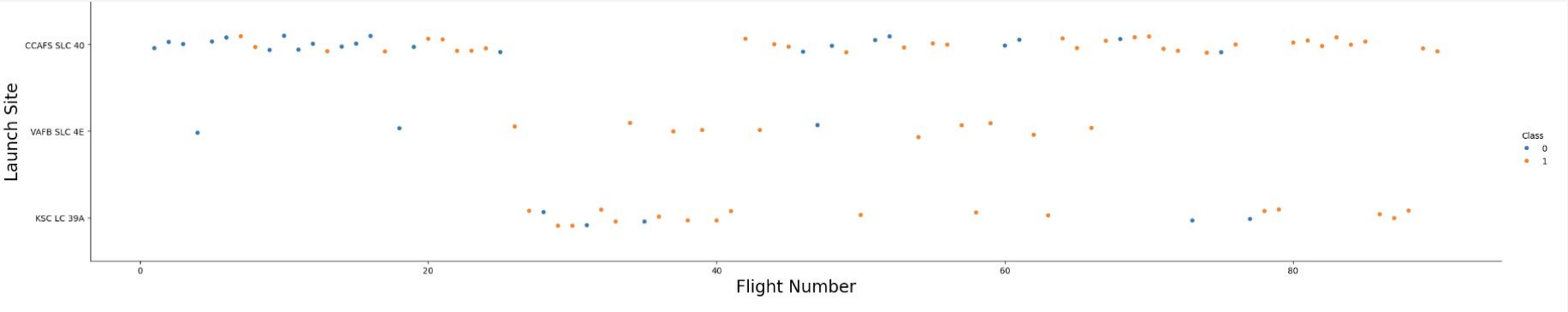
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results



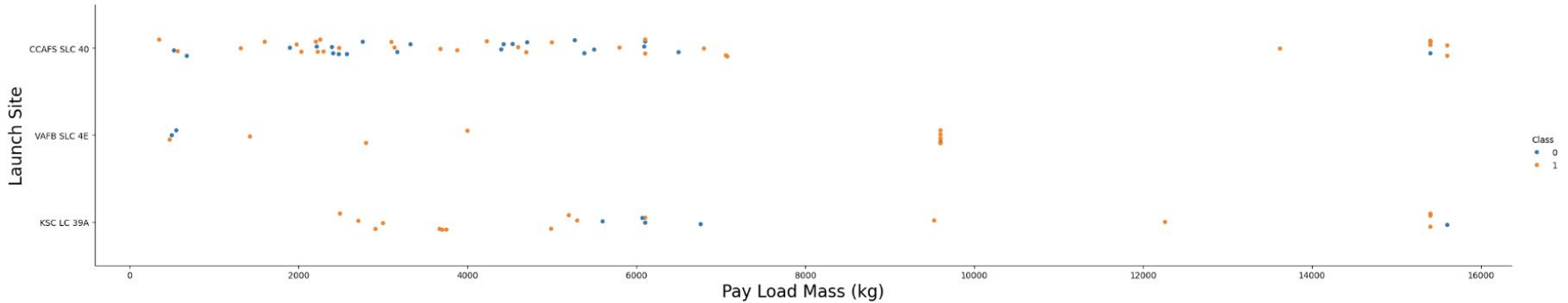
Section 2

Insights drawn from EDA

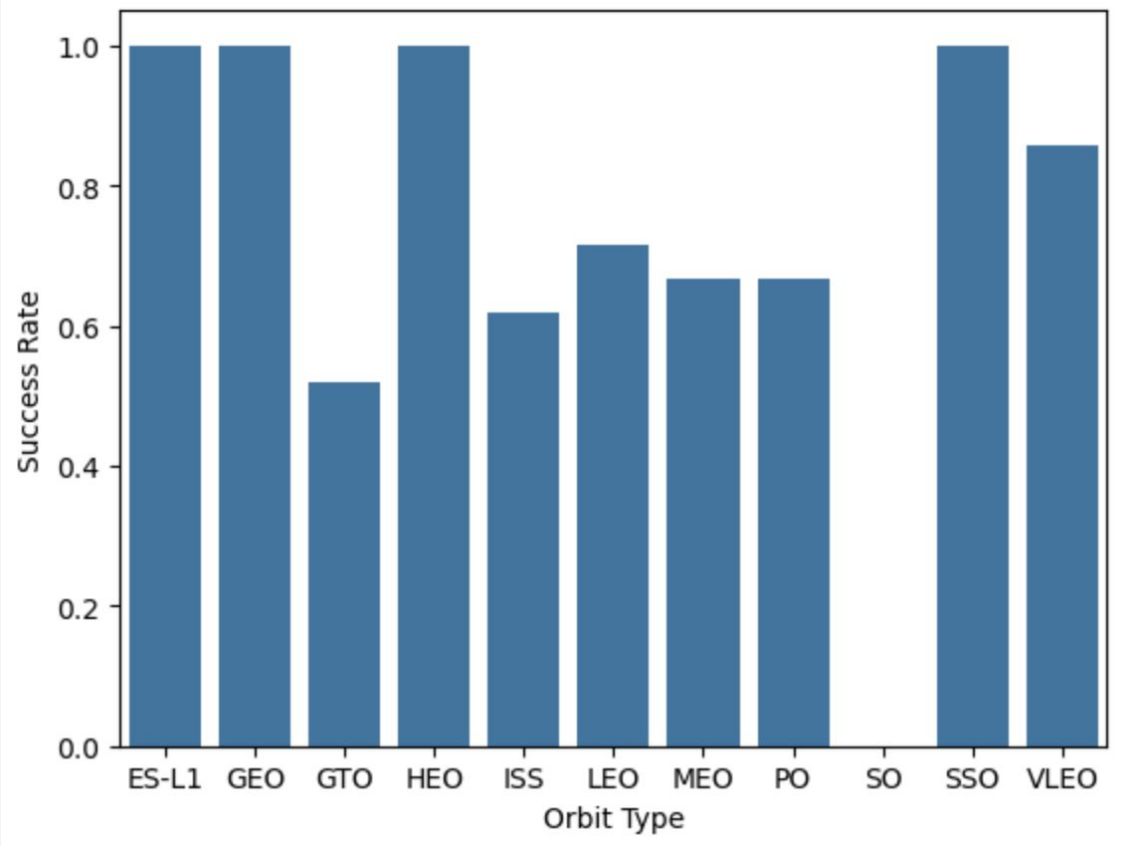
Flight Number vs. Launch Site



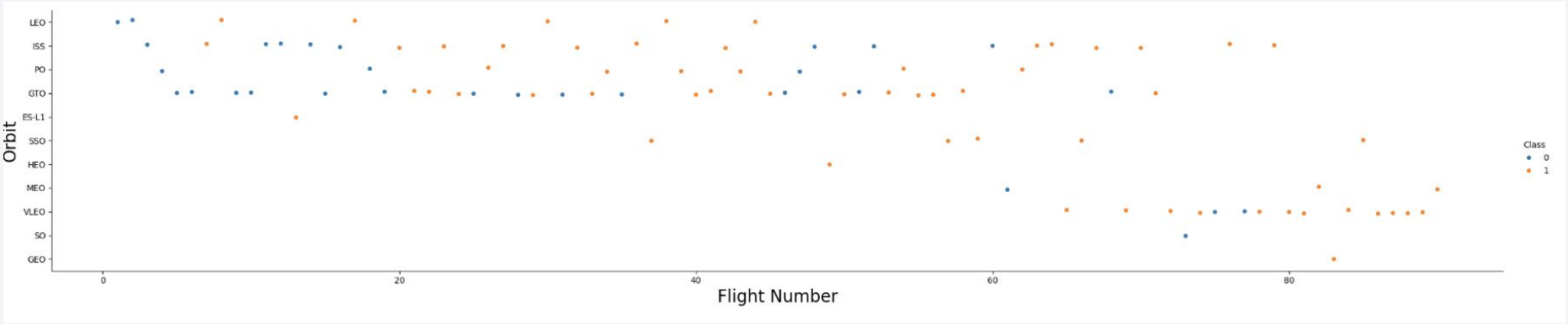
Payload vs. Launch Site



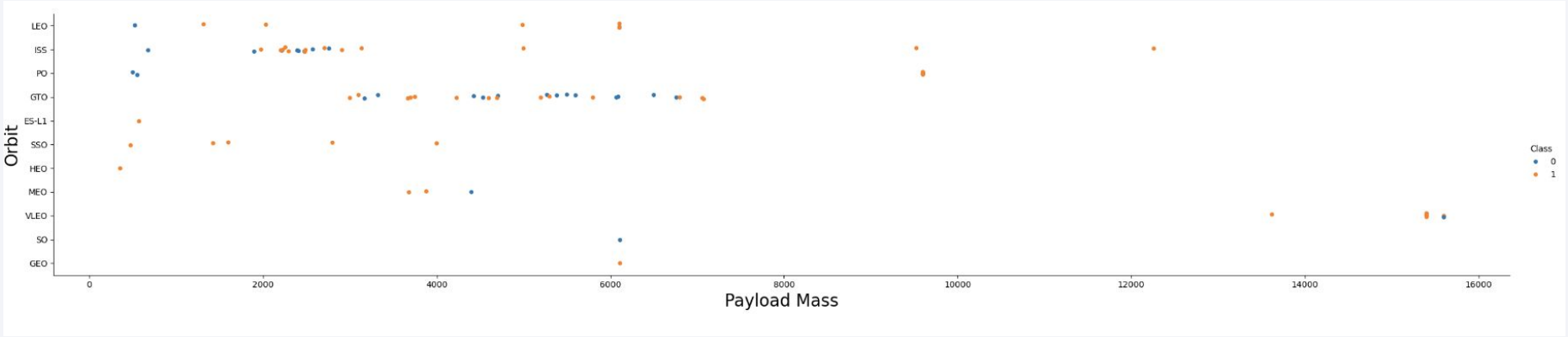
Success Rate vs. Orbit Type



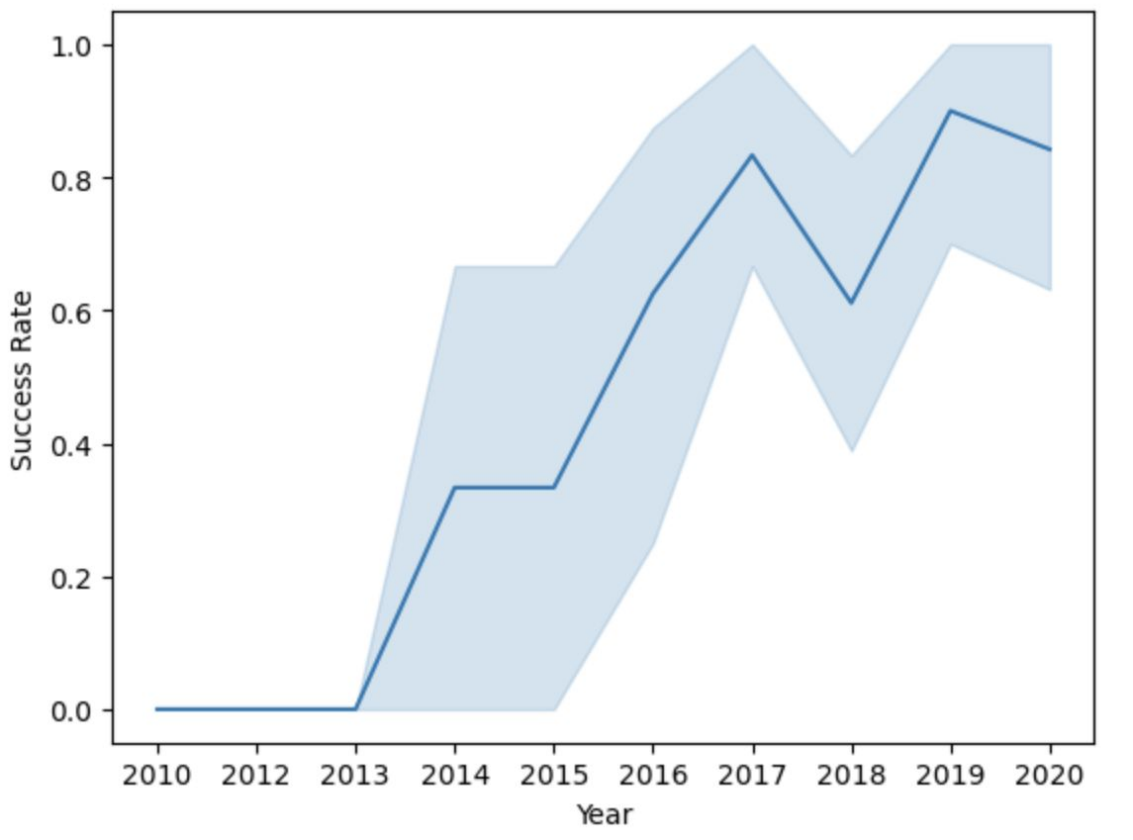
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

```
%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'

```
%sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success
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
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success

Total Payload Mass

```
%sql SELECT SUM(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)';
```

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

SUM(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

```
%sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE BOOSTER_VERSION LIKE 'F9 v1.1%';
```

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

AVG(PAYLOAD_MASS__KG_)
2534.6666666666665

First Successful Ground Landing Date

```
%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

%sql SELECT BOOSTER_VERSION, PAYLOAD FROM SPACEXTBL WHERE LANDING_OUTCOME = "Success (drone ship)" AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_< 6000

Booster_Version	Payload
F9 FT B1022	JCSAT-14
F9 FT B1026	JCSAT-16
F9 FT B1021.2	SES-10
F9 FT B1031.2	SES-11 / EchoStar 105

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT MISSION_OUTCOME, COUNT(MISSION_OUTCOME) FROM SPACEXTBL GROUP BY MISSION_OUTCOME;
```

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

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

Boosters Carried Maximum Payload

%sql SELECT BOOSTER_VERSION, PAYLOAD FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);

Booster_Version	Payload
F9 B5 B1048.4	Starlink 1 v1.0, SpaceX CRS-19
F9 B5 B1049.4	Starlink 2 v1.0, Crew Dragon in-flight abort test
F9 B5 B1051.3	Starlink 3 v1.0, Starlink 4 v1.0
F9 B5 B1056.4	Starlink 4 v1.0, SpaceX CRS-20
F9 B5 B1048.5	Starlink 5 v1.0, Starlink 6 v1.0
F9 B5 B1051.4	Starlink 6 v1.0, Crew Dragon Demo-2
F9 B5 B1049.5	Starlink 7 v1.0, Starlink 8 v1.0
F9 B5 B1060.2	Starlink 11 v1.0, Starlink 12 v1.0
F9 B5 B1058.3	Starlink 12 v1.0, Starlink 13 v1.0
F9 B5 B1051.6	Starlink 13 v1.0, Starlink 14 v1.0
F9 B5 B1060.3	Starlink 14 v1.0, GPS III-04
F9 B5 B1049.7	Starlink 15 v1.0, SpaceX CRS-21

2015 Launch Records

%sql SELECT substr(Date,6,2), BOOSTER_VERSION, LAUNCH_SITE, LANDING_OUTCOME FROM SPACEXTBL WHERE substr(Date,0,5) = '2015' AND LANDING_OUTCOME = 'Failure (drone ship)';

substr(Date,6,2)	Booster_Version	Launch_Site	Landing_Outcome
01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

```
%sql SELECT * FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' ORDER BY DATE DESC;
```

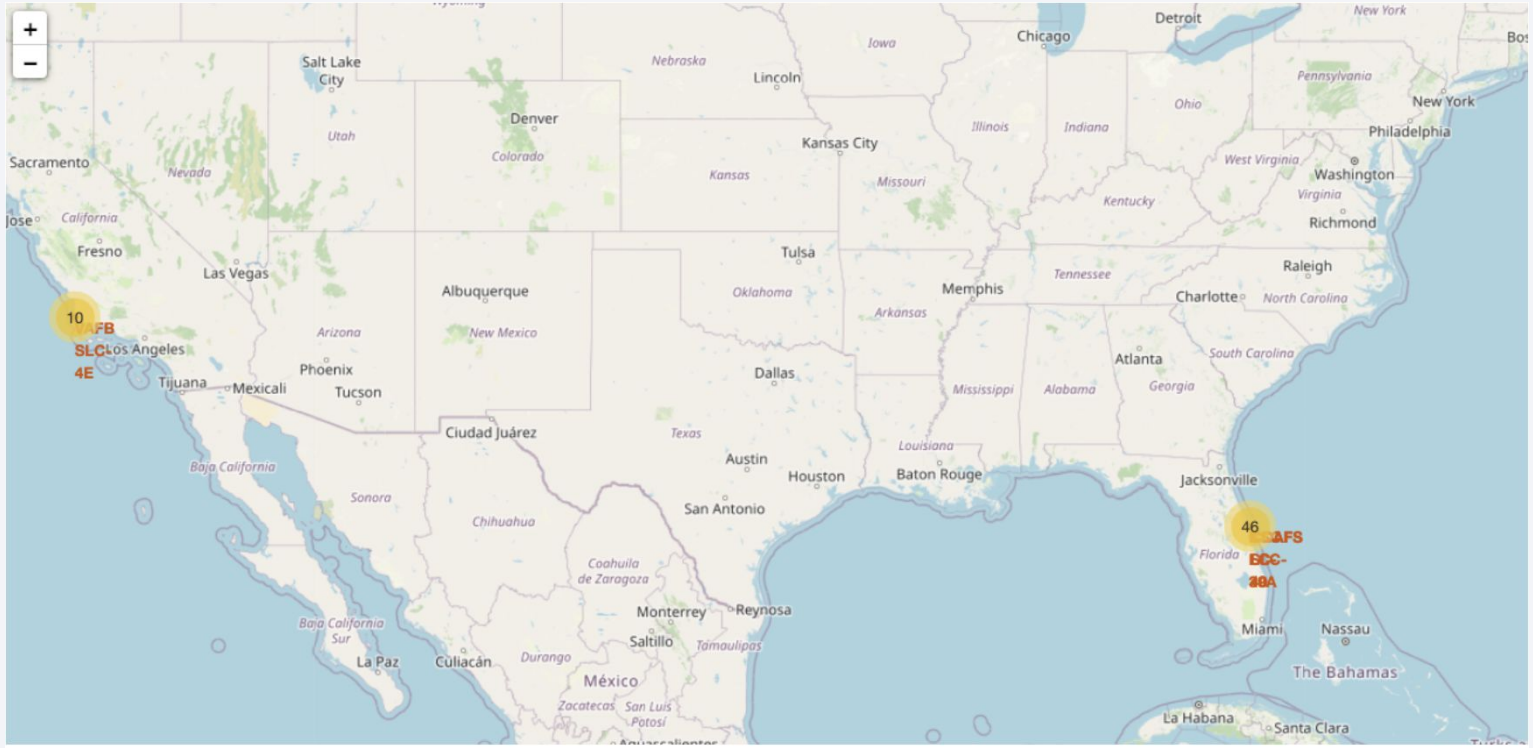
```
 * sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome
2017-03-16	6:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success
2017-01-14	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success
2016-08-14	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success
2016-07-18	4:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success
2016-06-15	14:29:00	F9 FT B1024	CCAFS LC-40	ABS-2A Eutelsat 117 West B	3600	GTO	ABS Eutelsat	Success
2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success

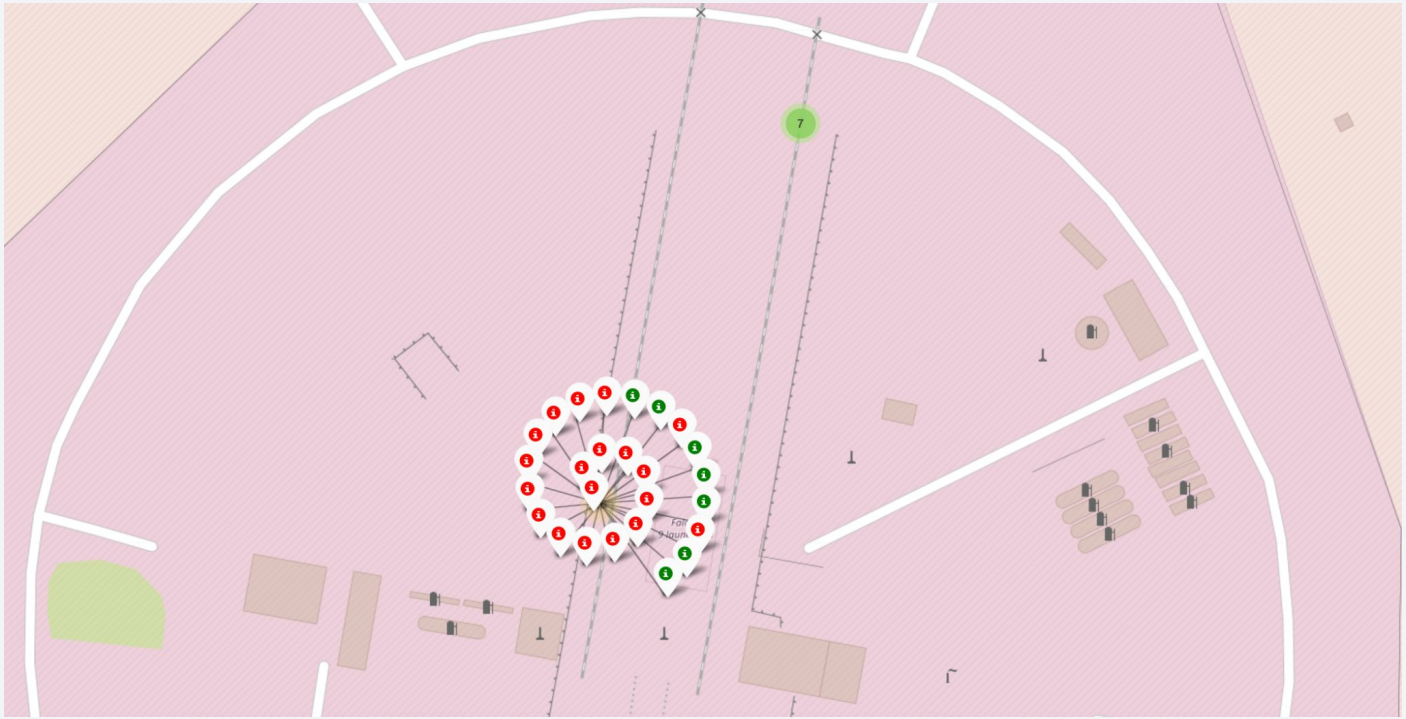
Section 3

Launch Sites Proximities Analysis

Areas of Launch Sites

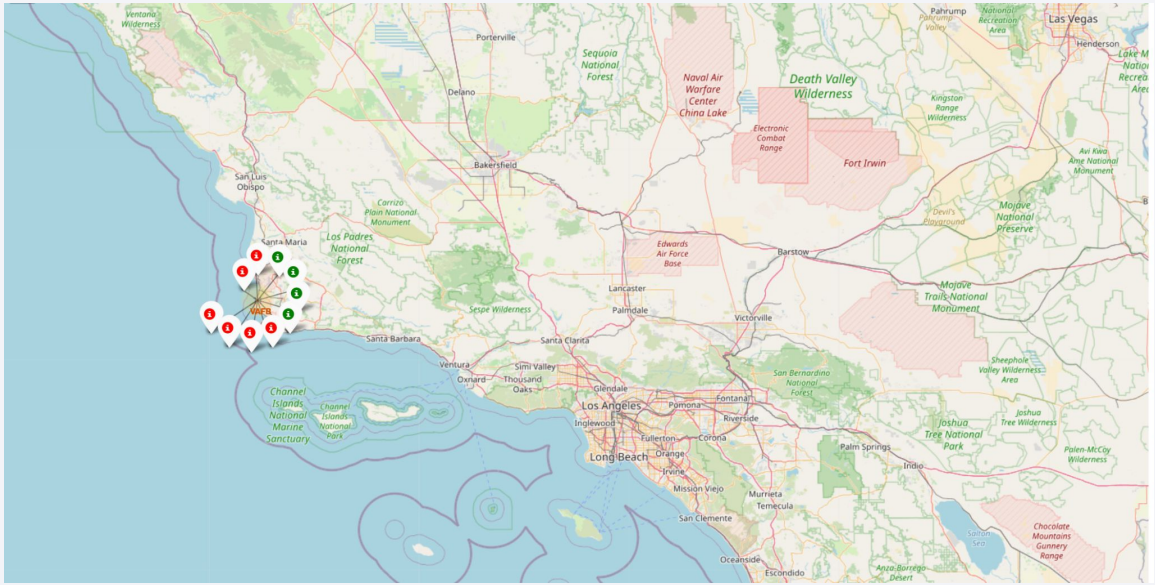


Launch Sites Success Rates (Florida)

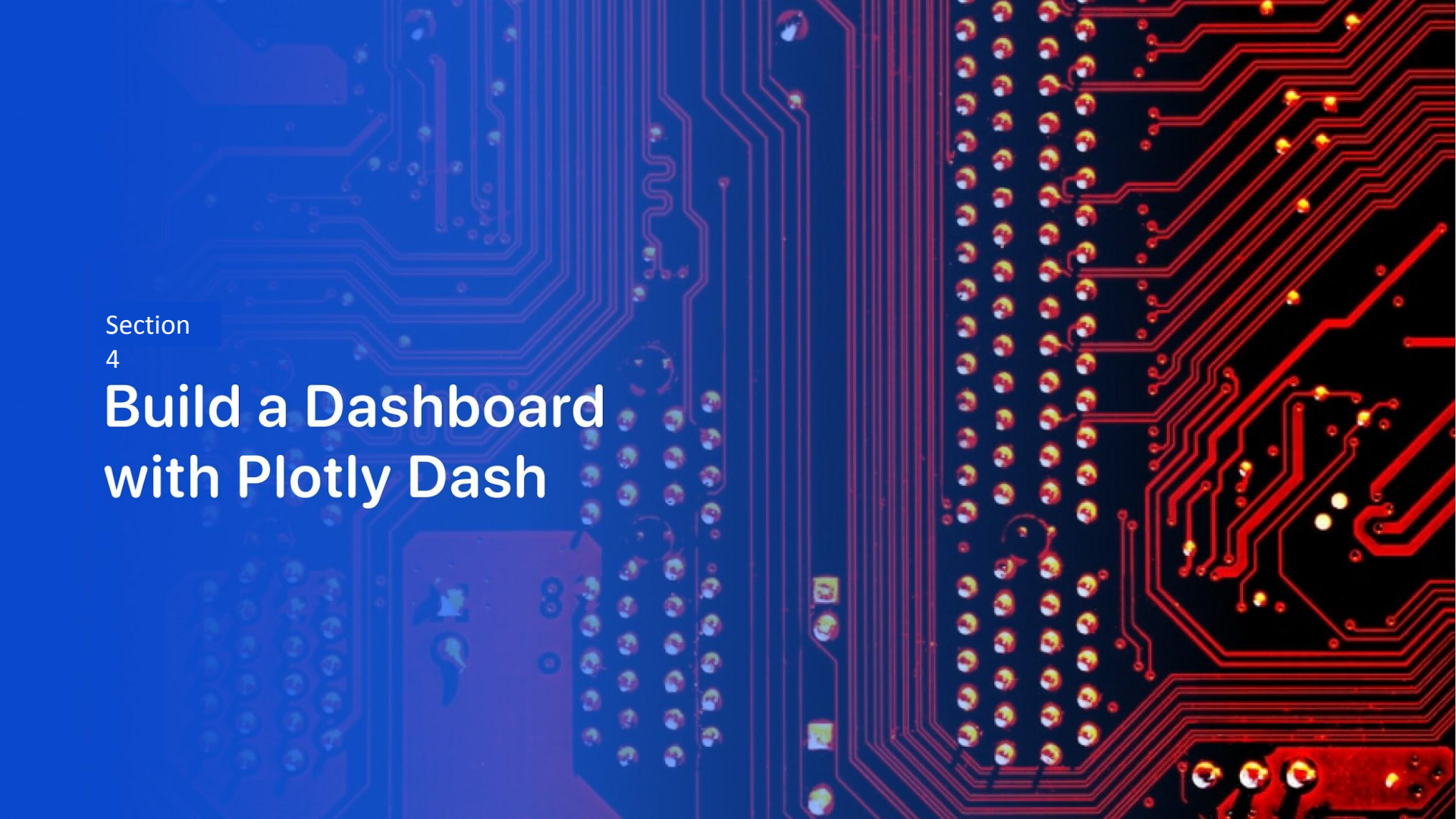


Red represents Failure and Green represents Success

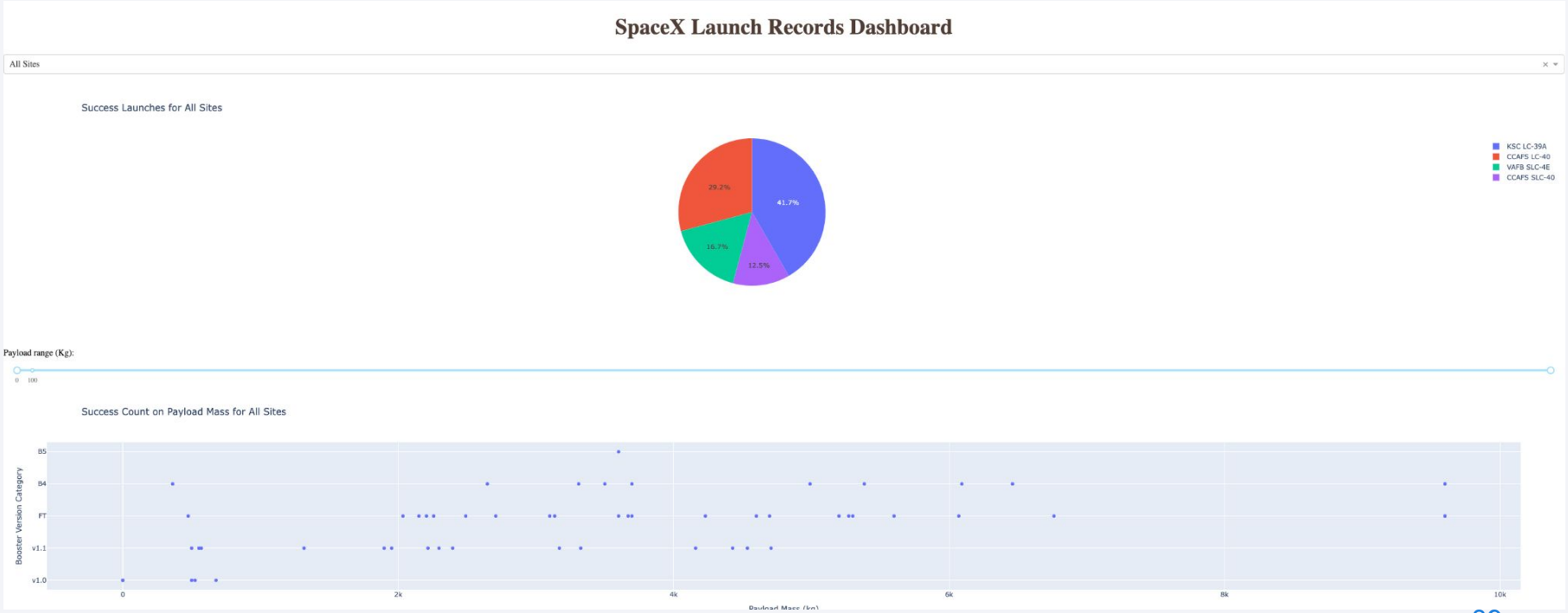
Launch Site Success Rates (California)



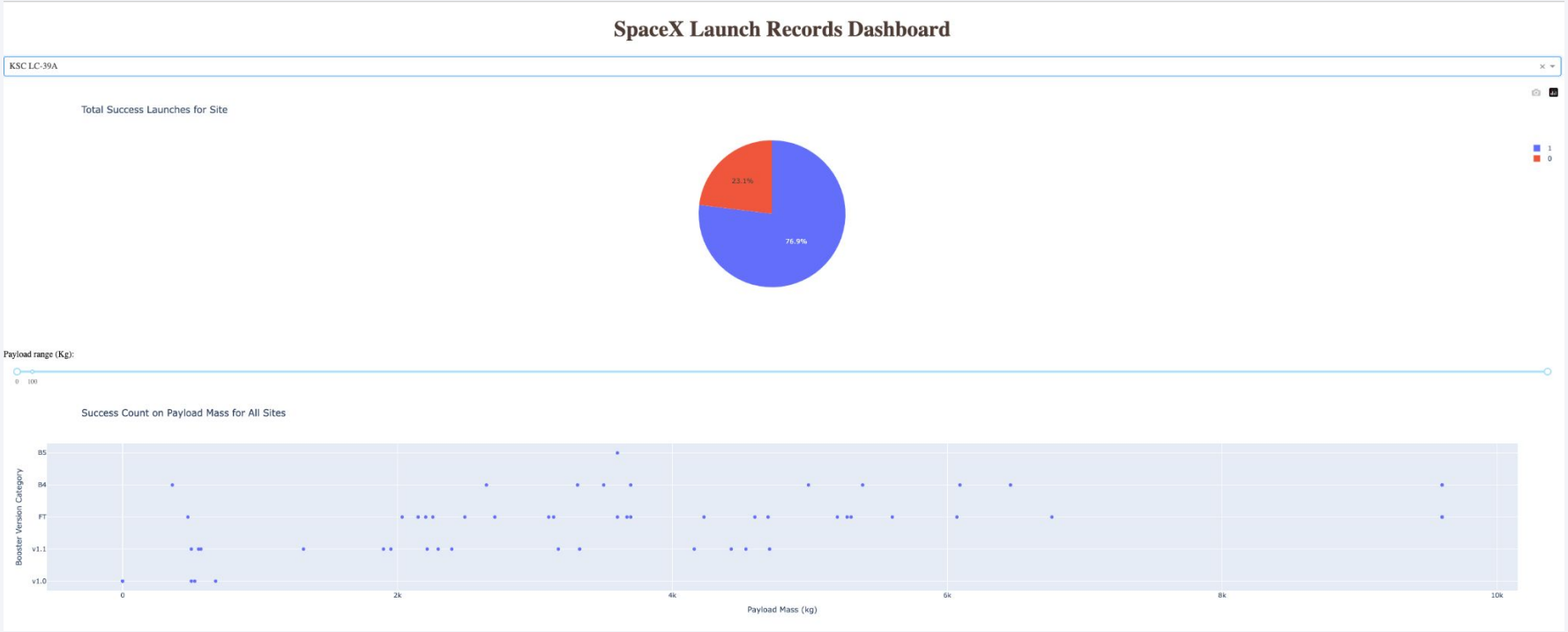
Red represents Failure and Green represents Success


Section
4
Build a Dashboard
with Plotly Dash

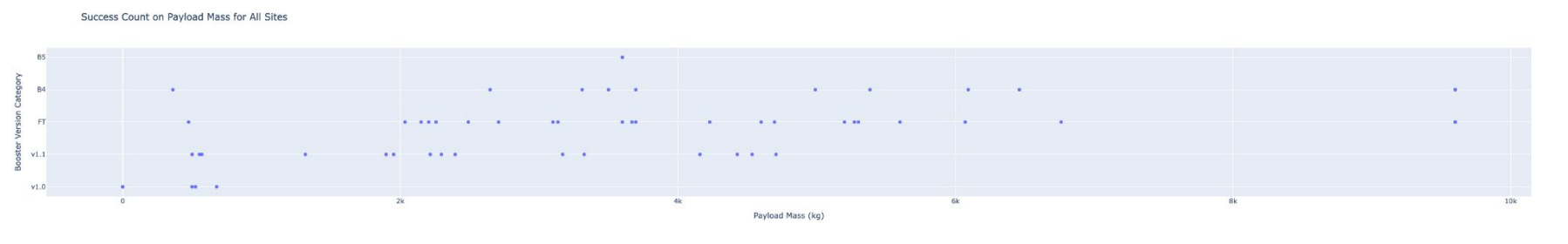
Launch Success For all Sites Pie Chart



Highest Launch Success Ratio Pie Chart



Payload v.s. Launch Outcome Scatter Plot



Section
5

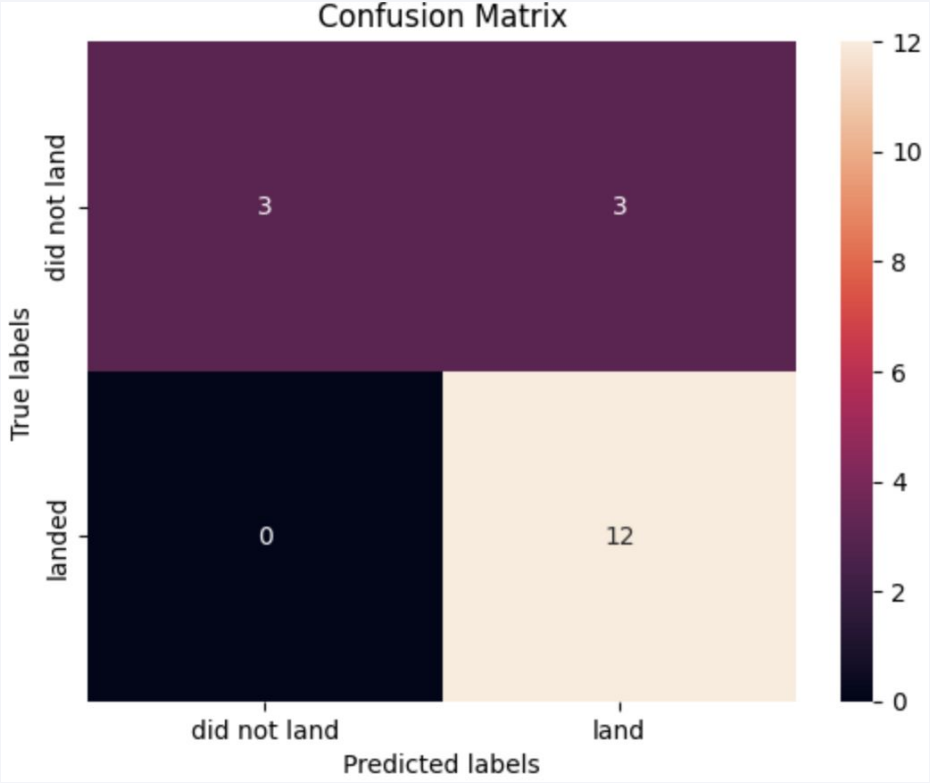
Predictive Analysis (Classification)

Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

- Find which model has the highest classification accuracy

Confusion Matrix



Conclusions

- Depending on the launch site you may get different rates of successes and failures
- Different methods performed similarly for predictions
- The launch site KSC LC-39A has the highest success rate compared to other sites
- There doesn’t seem to be a relationship between GTO orbit and flight number

Appendix

- https://github.com/nini2580/SpaceX_Falcon_9_Success_Landing

Thank you!