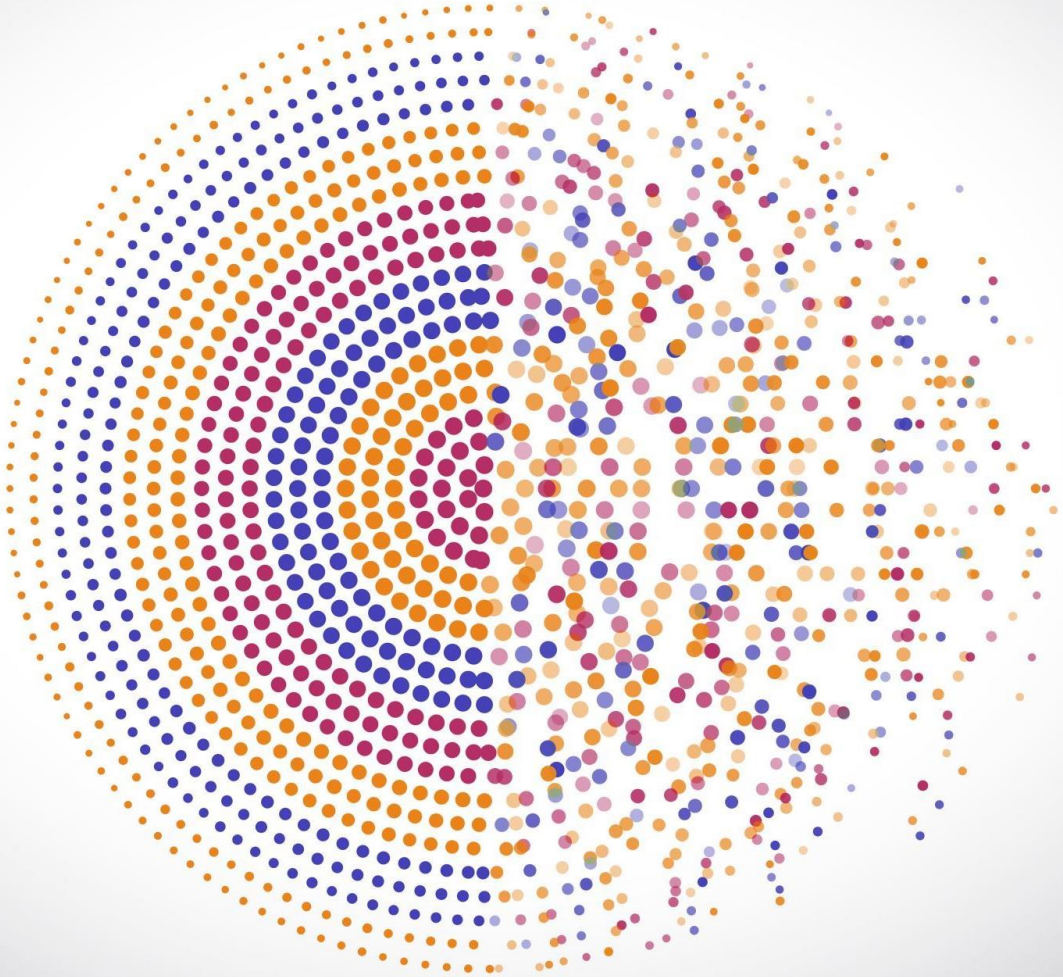
IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Yue Liu
July 30, 2023

Outline

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



Executive Summary

- Summary of methodologies
  - Collect data
  - Data wrangling
  - Exploratory Data Analysis
  - Data Visualization
  - Machine Learning Prediction
- Summary of all results
  - Exploratory Data Analysis results
  - Data Visualizaton results
  - Prediction Analysis results

Introduction

Project background and context

The Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

Problems you want to find answers

- What factors contribute to successful landing of the rocket?
- What's the successful landing rate for each site?

Section 1

Methodology

Methodology

- Executive Summary
- Data collection methodology:
  - Using API and web scarping to collect the data
- Perform data wrangling
  - Understand the data by statistical calculating and enriching data
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Splitting the data into training set and test set
  - Applying the model and evaluating the model performances

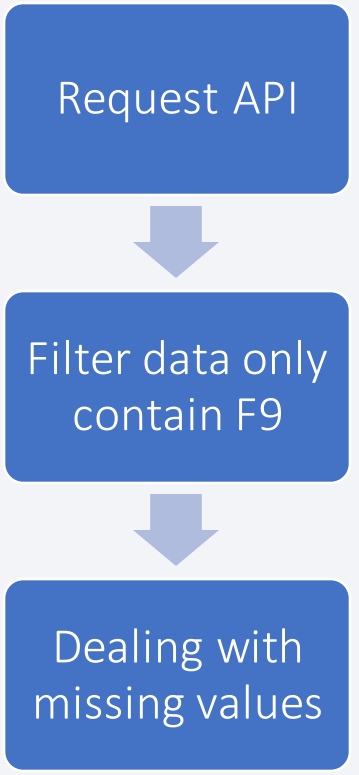
Data Collection

- Sources:
  - SpaceX API data collection
  - SpaceX web scaping data collection

Data Collection – SpaceX API

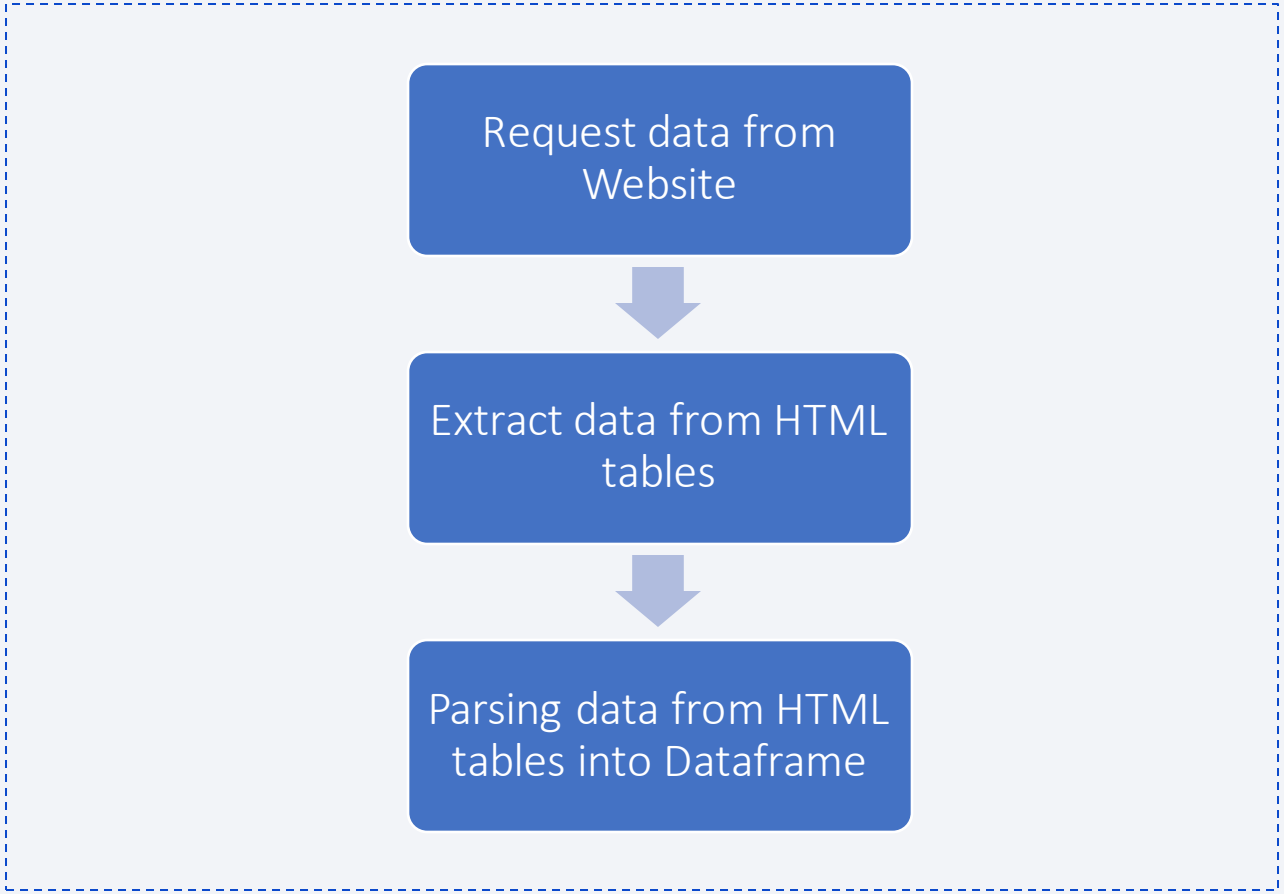
- Present your data collection with SpaceX REST calls using key phrases and flowcharts

https://github.com/yueliu-github/IBM_DS_CapstoneProject/blob/main/Data%20Collection%20API%20Lab.ipynb



Data Collection - Scraping

https://github.com/yueliu-github/IBM_DS_CapstoneProject/blob/main/Data%20Collection%20with%20Web%20Scraping.ipynb



Data Wrangling

- Collect data then explorary the data
- data wrangling process using key phrases and flowcharts

- https://github.com/yueliu-github/IBM_DS_CapstoneProject/blob/main/Data%20Wrangling.ipynb

EDA with Data Visualization

- Using scatter plot to visualize the relationship between two variables
- Using line chart to visualize the success rate changes during a time period
- Using bar chart to visualize the success rate among different orbit types

https://github.com/yueliu-github/IBM_DS_CapstoneProject/blob/main/Exploratory%20Data%20Analysis%20for%20Data%20Visualization.ipynb

EDA with SQL

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first succesful landing outcome in ground pad was achieved
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

https://github.com/yueliu-github/IBM_DS_CapstoneProject/blob/main/Exploratory%20Data%20Analysis%20with%20SQL.ipynb

Build an Interactive Map with Folium

The folium map objects contains markers, circles, lines, etc.

- Markers indicates the target objectives for launch sites
- Circles means the area where the launch sites located
- Lines shows the distance between launch sites and most closed city, railway or coastline

https://github.com/yueliu-github/IBM_DS_CapstoneProject/blob/main/Interactive%20Visual%20Analytics%20with%20Folium%20lab.ipynb

Build a Dashboard with Plotly Dash

The dashboard is an interactive dashboard with all launch sites, and it contains pie chart and scatter plot.

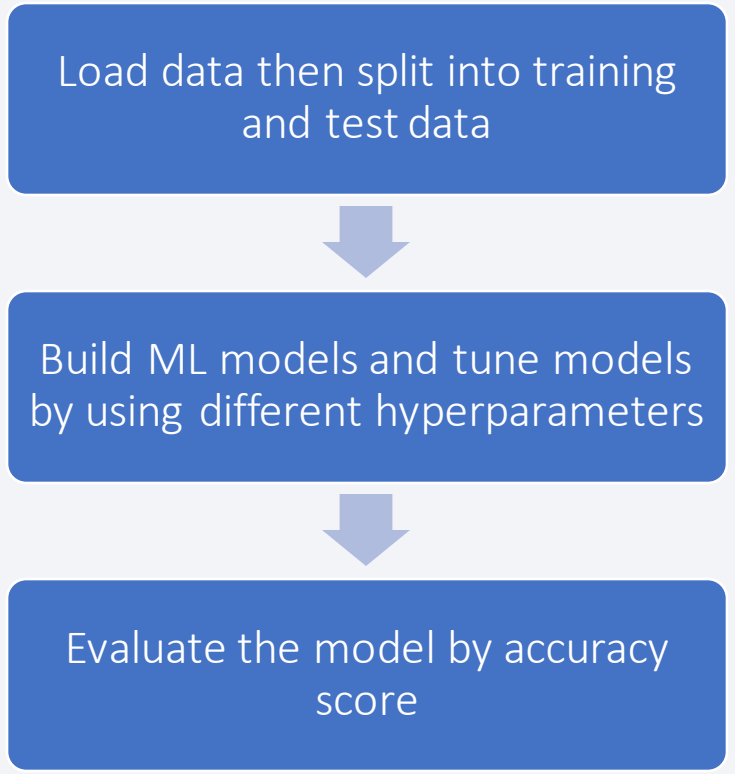
- Pie chart shows the proportion of the launches for each site
- Scatter plot indicates the relationship between the outcome and payload mass for different booster version

https://github.com/yueliu-github/IBM_DS_CapstoneProject/blob/main/dash_interactivity.py

Predictive Analysis (Classification)

- Load data into Dataframe, then split data into training set and test set
- Use Logistic regression, SVM, Decision Tree, KNN models to train the data
- Evaluate the model performance by using accuracy score and confusion matrix

https://github.com/yueliu-github/IBM_DS_CapstoneProject/blob/main/Machine%20Learning%20Prediction%20LAB.ipynb

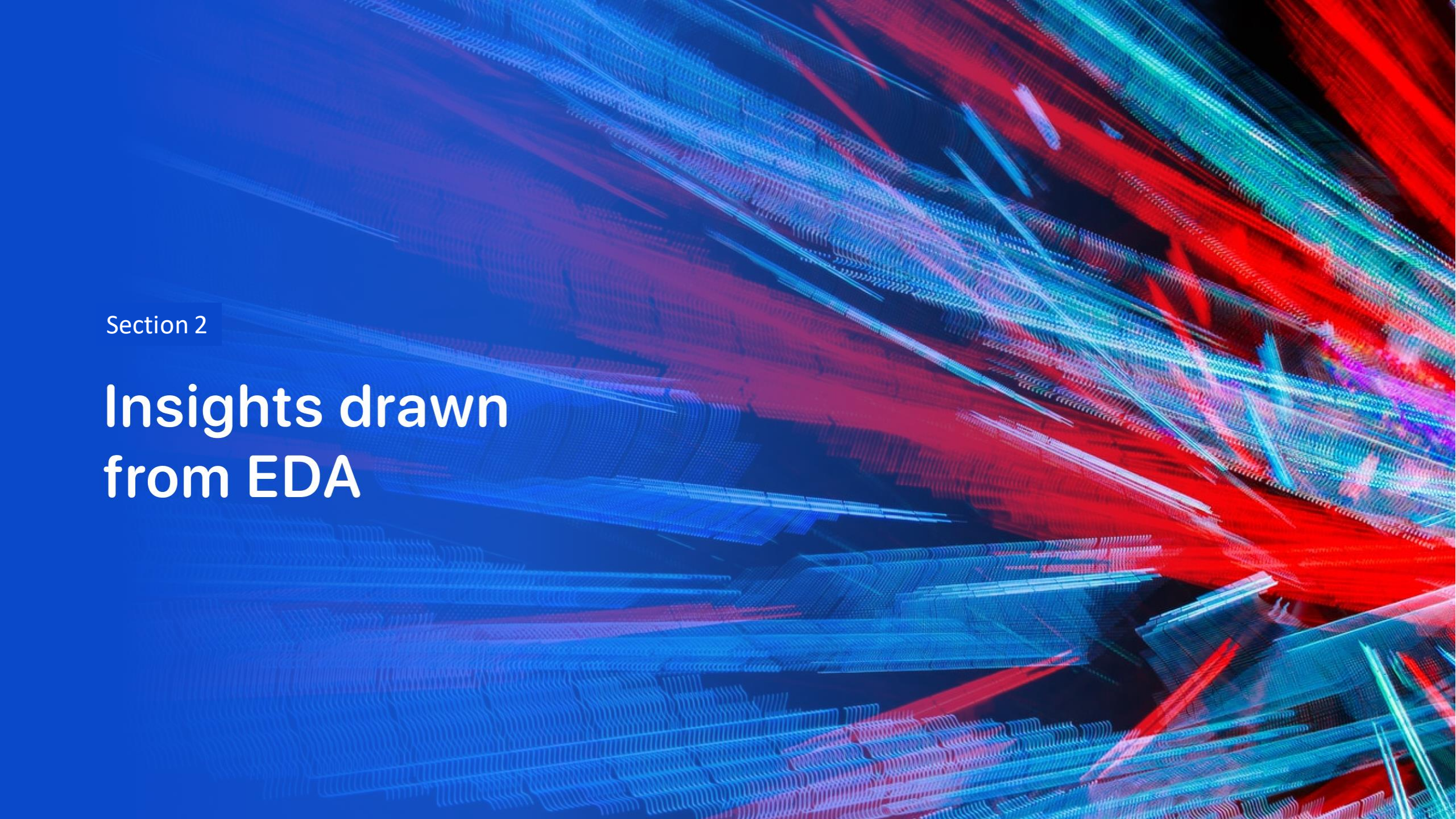


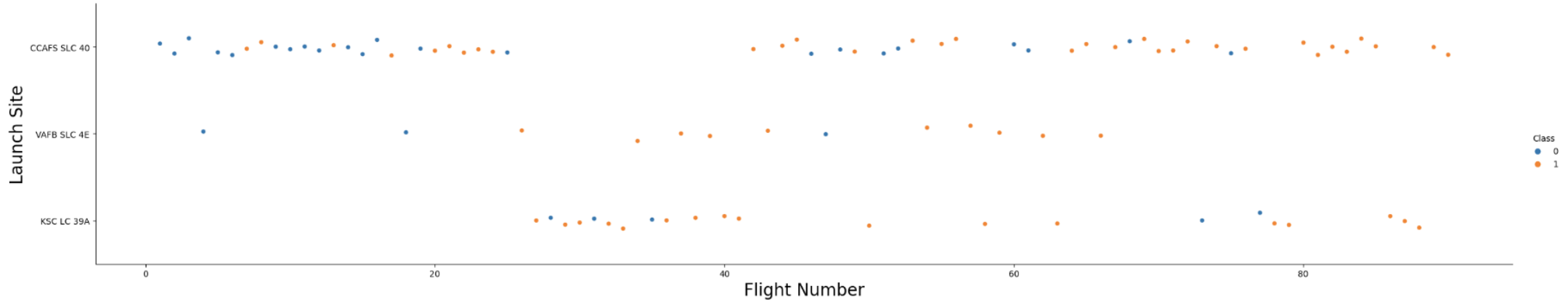
Results

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

Section 2

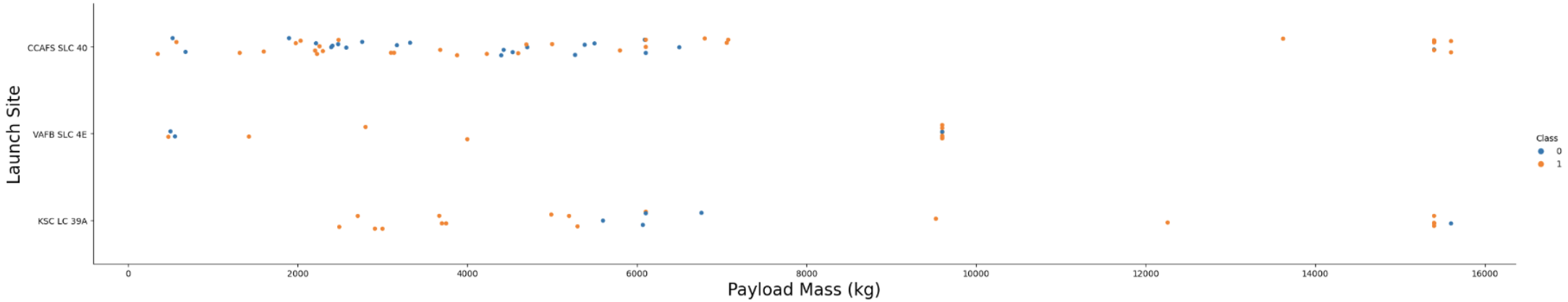
Insights drawn from EDA





Flight Number vs. Launch Site

- With increasing of the flight number, the launch site has more success landing

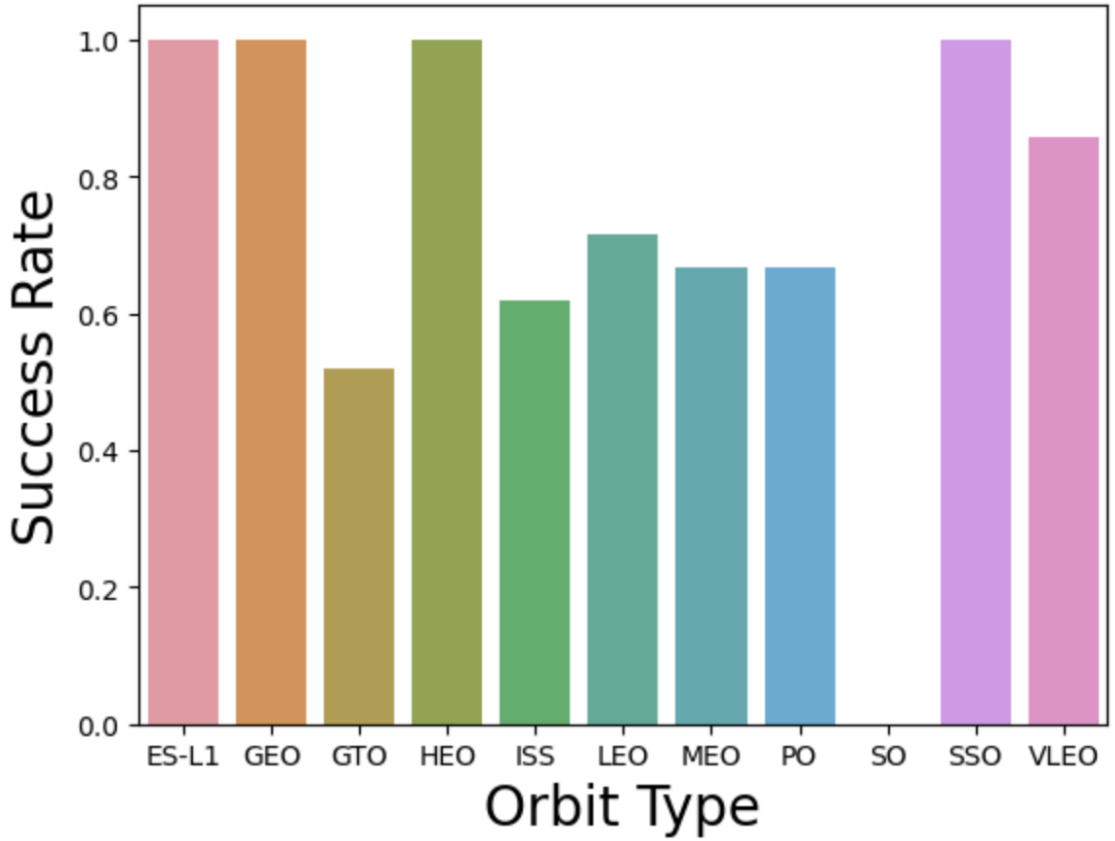


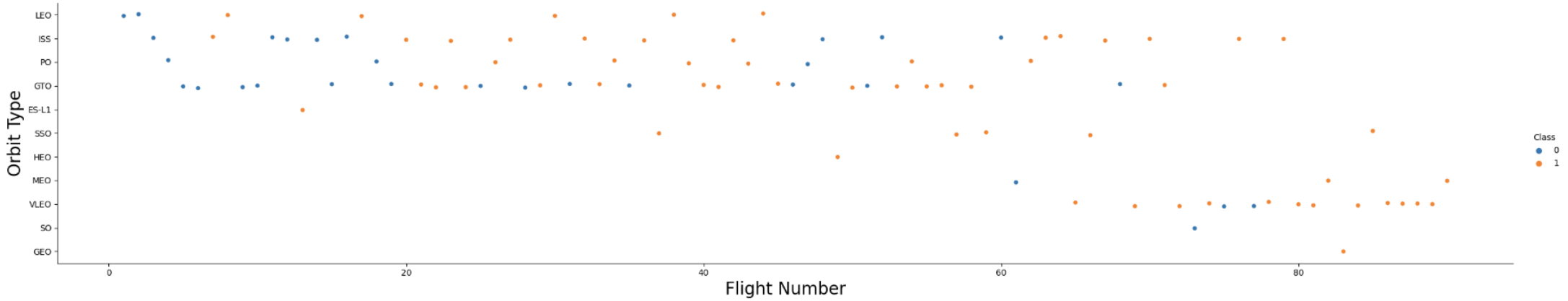
Payload vs. Launch Site

- For CCAFS SLC 40 launch site, the higher payload mass gets higher success rate

Success Rate vs. Orbit Type

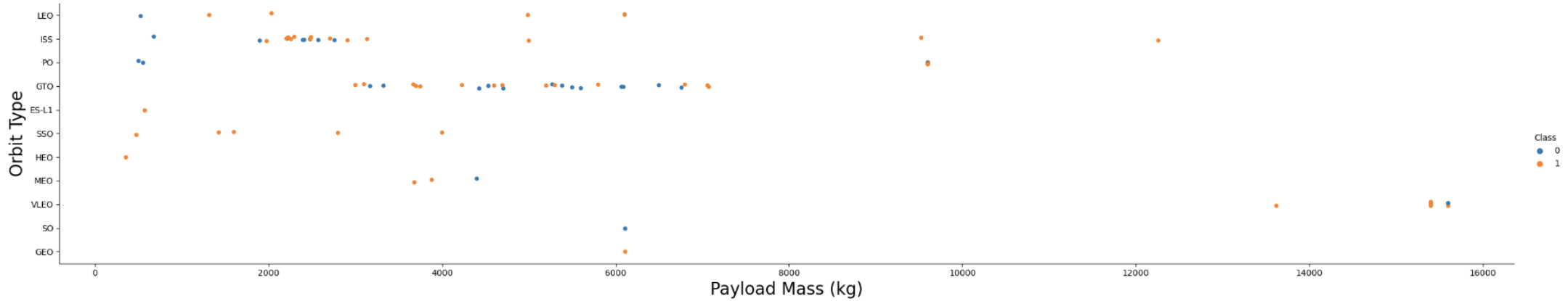
- ES-L1, GEO, HEO, SSO have the highest success rate.





Flight Number vs. Orbit Type

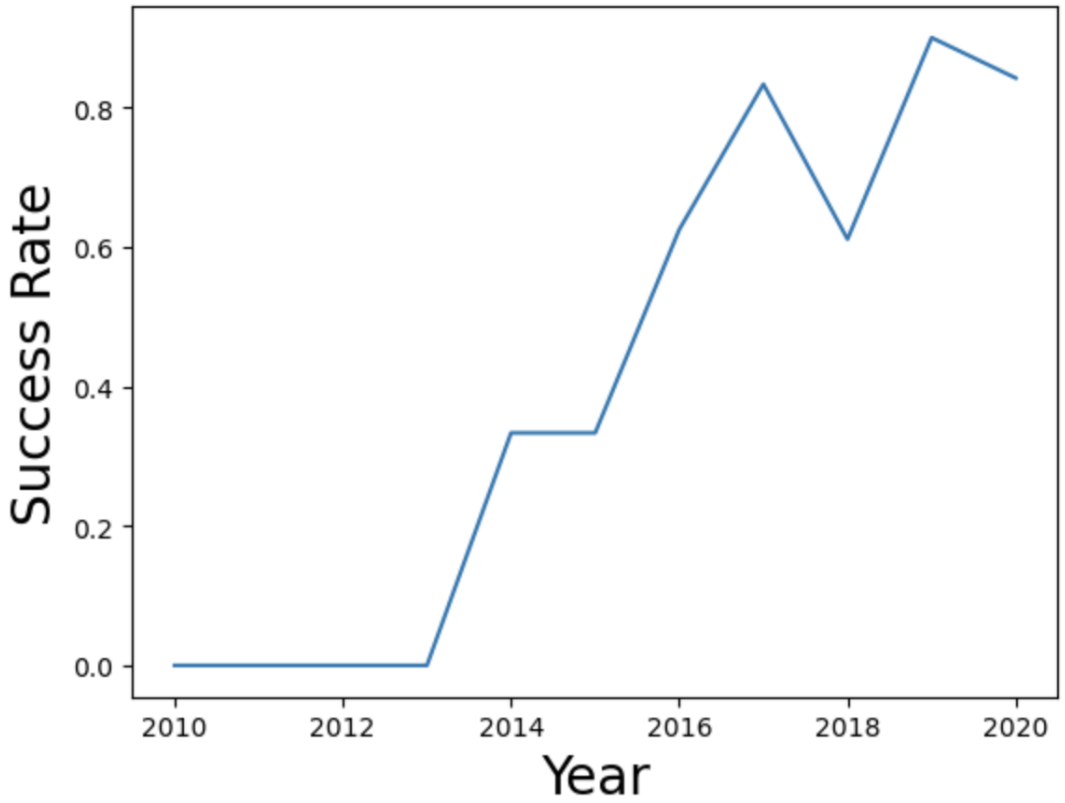
- there is no relationship between flight number and orbit type



Payload vs. Orbit Type

- The heavier payload mass shows more success landing for LEO, ISS, PO.

Launch Success Yearly Trend



- The success rate raises along with the years.

All Launch Site Names

- Find the names of the unique launch sites

Display the names of the unique launch sites in the space mission

```
%sql select distinct LAUNCH_SITE from SPACE_X;
```

```
 * ibm_db_sa://vjy26289:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
   sqlite:///my_data1.db
Done.
```

launch_site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E
None

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

```
%sql select * from SPACE_X where LAUNCH_SITE like 'CCA%' limit 5;
```

```
 * ibm_db_sa://vjy26289:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
   sqlite:///my_data1.db
Done.
```

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing_ou
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	(par
2010-08-12	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	(par
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No
2012-08-10	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql select sum(PAYLOAD_MASS__KG_) as total_payload_mass_kg from SPACE_X where customer = 'NASA (CRS)';
```

```
 * ibm_db_sa://vjy26289:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
   sqlite:///my_data1.db
Done.
```

total_payload_mass_kg
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

Display average payload mass carried by booster version F9 v1.1

```
%sql select avg(PAYLOAD_MASS__KG_) as avg__payload_mass_kg from SPACE_X where booster_version like 'F9 v1.1%';
```

```
 * ibm_db_sa://vjy26289:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
   sqlite:///my_data1.db
Done.
```

avg__payload_mass_kg
2534

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
%sql select min(DATE) as first_successful_landing from SPACE_X where landing_outcome = 'Success (ground pad)';
```

```
 * ibm_db_sa://vjy26289:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
   sqlite:///my_data1.db
Done.
```

first_successful_landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%sql select distinct booster_version from SPACE_X where landing_outcome = 'Success (drone ship)' and (PAYLOAD_MAS
```

```
 * ibm_db_sa://vjy26289:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
   sqlite:///my_data1.db
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

```
%sql select MISSION_OUTCOME, count(*) as total_nums from SPACE_X group by MISSION_OUTCOME;
```

```
 * ibm_db_sa://vjy26289:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
   sqlite:///my_data1.db
Done.
```

mission_outcome	total_nums
Failure (in flight)	1
Success	99
Success (payload status unclear)	1
None	898

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

```
%sql select booster_version from SPACE_X where PAYLOAD_MASS__KG_ in (select max(PAYLOAD_MASS__KG_) from SPACE_X);
```

```
 * ibm_db_sa://vjy26289:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
   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

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 4, 2) as month to get the months and substr(Date,7,4)='2015' for year.

```
%sql select MONTH(DATE) as month, landing_outcome, booster_version, launch_site from SPACE_X where landing_outcom
```

```
 * ibm_db_sa://vjy26289:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
   sqlite:///my_data1.db
Done.
```

MONTH	landing_outcome	booster_version	launch_site
10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
4	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

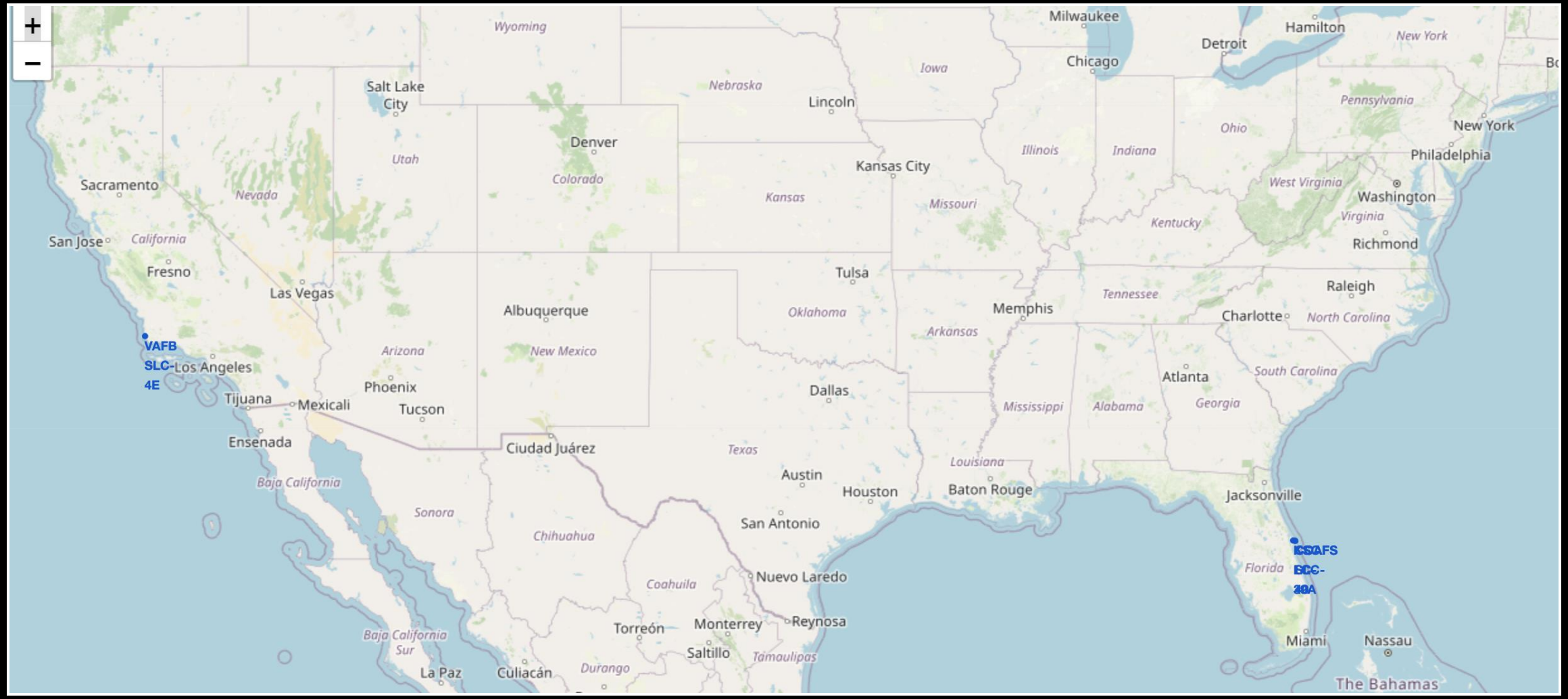
Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

```
%sql select landing_outcome, count(*) as total from SPACE_X where DATE between '2010-06-04' and '2017-03-20'group
```

```
 * ibm_db_sa://vjy26289:***@b70af05b-76e4-4bca-a1f5-23dbb4c6a74e.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32716/bludb
   sqlite:///my_data1.db
Done.
```

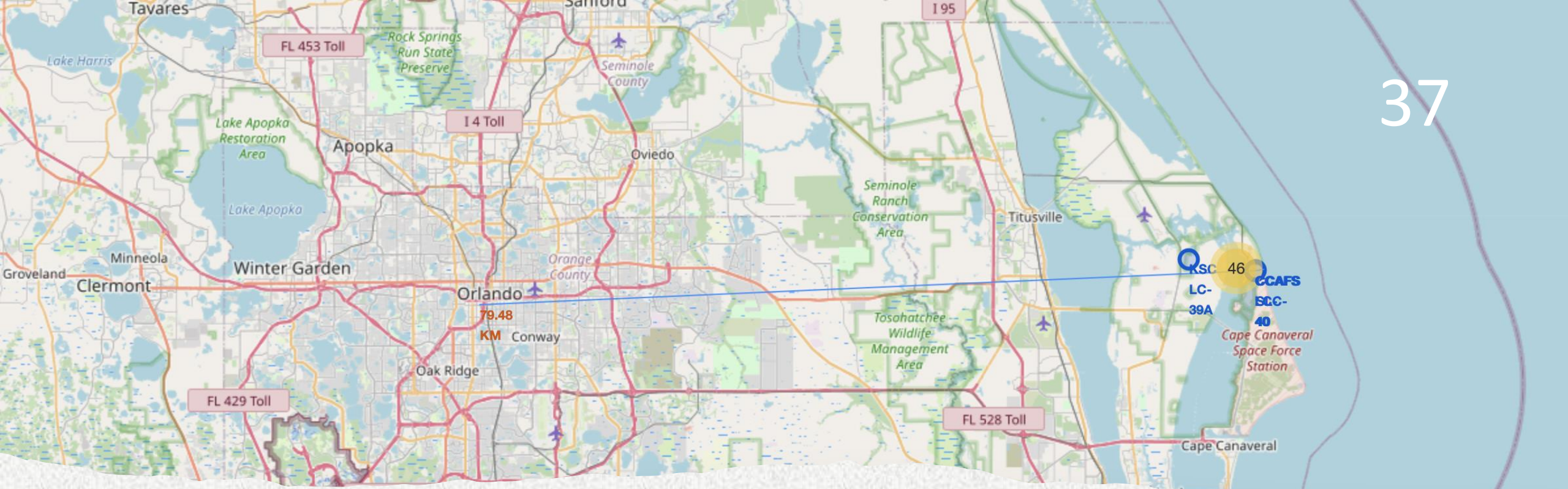
landing_outcome	total
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Success (ground pad)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	1
Precluded (drone ship)	1


Section 3
Launch Sites
Proximities Analysis

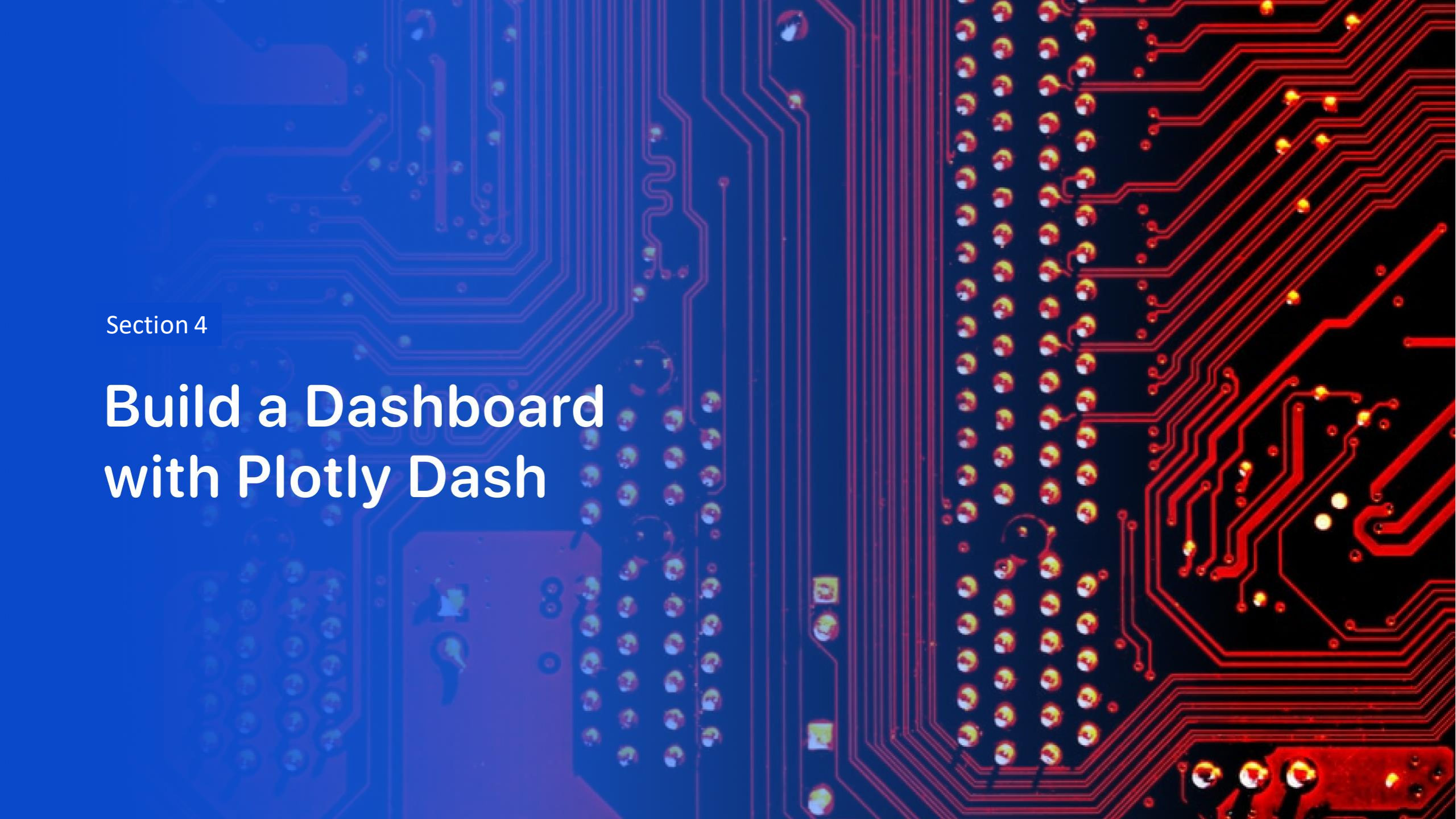


All launch sites map

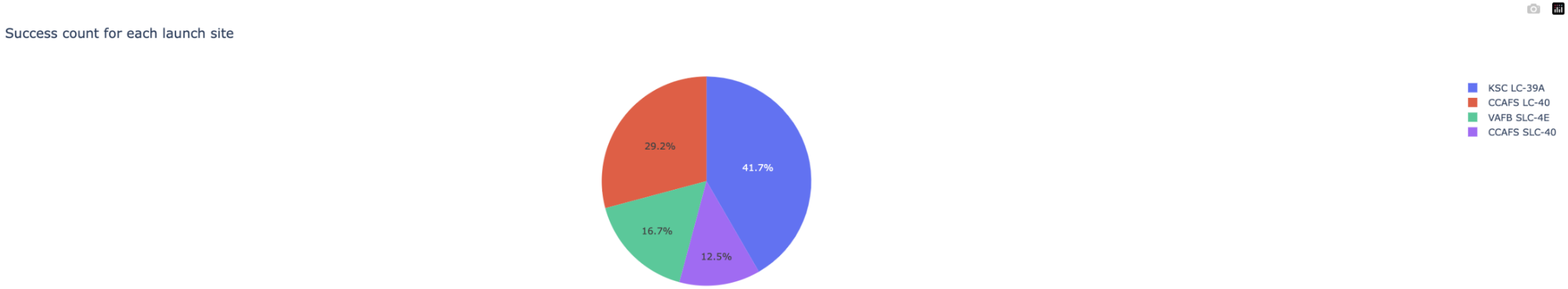
Launch outcomes for CCAFS SLC - 40



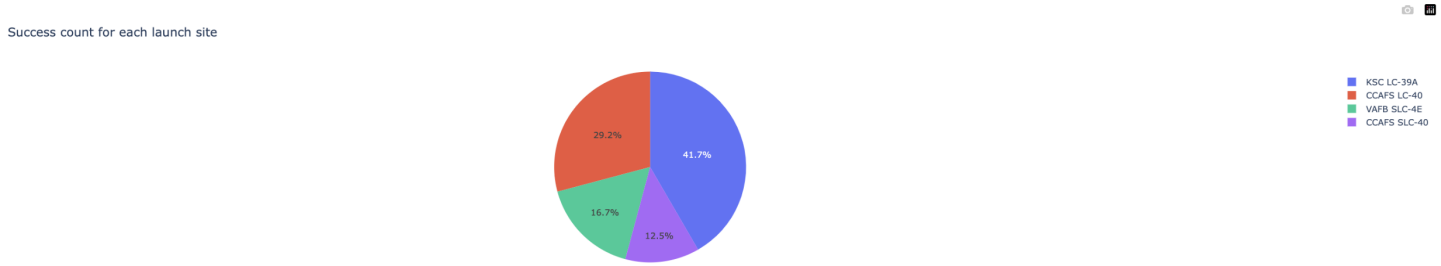
The distance between Orlando and CCAFS SLC 40


Section 4
Build a Dashboard with Plotly Dash

Pie chart: Success count for all launch sites



Pie chart: Highest success count for launch site



Scatter plot: Payload vs. Launch Outcome for all sites

Section 5

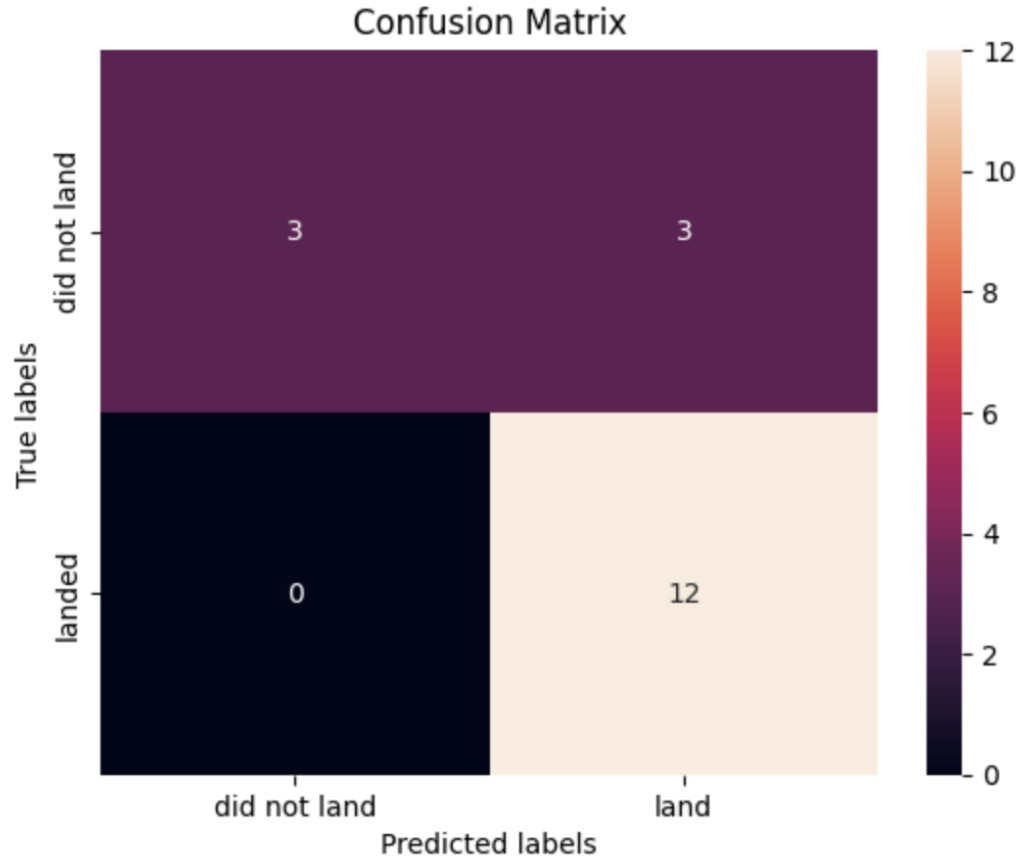
Predictive Analysis (Classification)

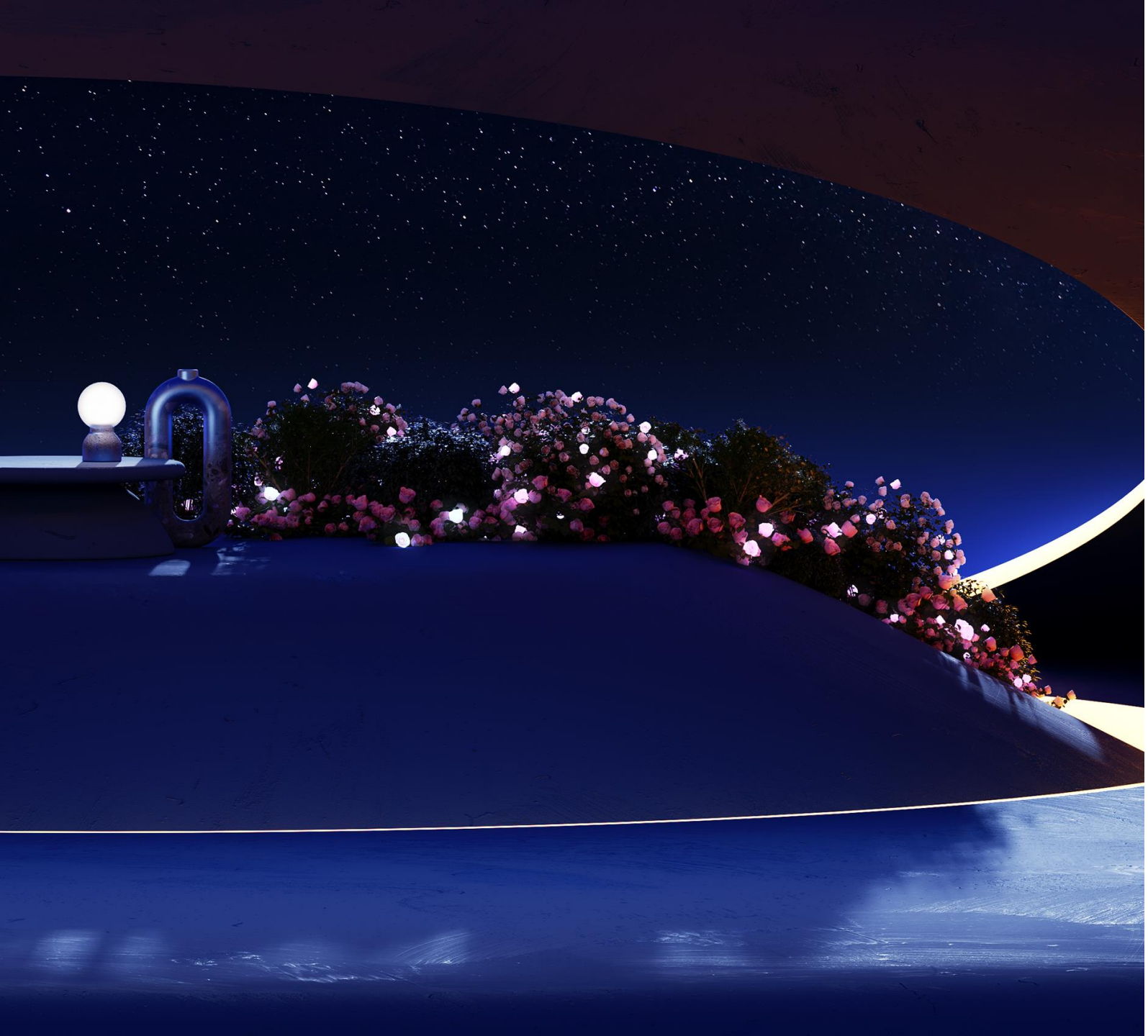
Classification Accuracy

- SVM has the highest score among these models

Confusion Matrix

- The confusion matrix of the SVM model.





Conclusions

- The payload mass and flight number play an important role in landing successfully.
- The SVM has the best performance in predicting the landing success rate.

Appendix

- API URL: https://github.com/yueliu-github/IBM_DS_CapstoneProject/blob/main/Data%20Collection%20API%20Lab.ipynb
- Web scrab: https://github.com/yueliu-github/IBM_DS_CapstoneProject/blob/main/Data%20Collection%20with%20Web%20Scraping.ipynb
- Data wrangling: https://github.com/yueliu-github/IBM_DS_CapstoneProject/blob/main/Data%20Wrangling.ipynb
- EDA with data visualization: https://github.com/yueliu-github/IBM_DS_CapstoneProject/blob/main/Exploratory%20Data%20Analysis%20for%20Data%20Visualization.ipynb
- EDA with SQL: https://github.com/yueliu-github/IBM_DS_CapstoneProject/blob/main/Exploratory%20Data%20Analysis%20with%20SQL.ipynb
- Folium visualization: https://github.com/yueliu-github/IBM_DS_CapstoneProject/blob/main/dash_interactivity.py
- Dash Plotly: https://github.com/yueliu-github/IBM_DS_CapstoneProject/blob/main/dash_interactivity.py
- ML Prediction Analysis: https://github.com/yueliu-github/IBM_DS_CapstoneProject/blob/main/Machine%20Learning%20Prediction%20LAB.ipynb
- GitHub repo: https://github.com/yueliu-github/IBM_DS_CapstoneProject

Thank you!