IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Christian Baumgartner
04.05.2022

Outline

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

Executive Summary

- **Summary of methodologies**
  - Data Collection API / Webscraping
  - Data Wrangling
  - Exploartory Data Analysis with
    - SQL
    - Visualization (Matplotlib/Seaborn, Folium)
  - Predictive Analysis using Machine Learning
- **Summary of all results**

Introduction

- **Project background and context**
  - SpaceX advertises cheap rocket launches (62 million dollars)
  - cheap compared to other providers (> 165 million dollars)
  - Saving come from reusable first rocket stages

- **Problems you want to find answers**

  Determine whether the first stage will land to determine the cost of a launch based on mission parameters, e.g. payload, orbit, launch site or customer

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data requested from SpaceX API
  - Webscraping from Wikipedia Falcon 9 launches
- Perform data wrangling
  - Dealing with missing values (e.g. payload NaN replaced with mean payload)
  - One-Hot encoding of labeled 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
  - Logistic regression, SVM, Decision Tree, kNN

Data Collection

Data Sources:

- SpaceX API

  Direct request from SpaceX data via Json files

- Wikipedia: List of Falcon 9 and Falcon Heavy launches

  https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

  webscraping using BeautifulSoup

Data Collection – SpaceX API

- Data collection with SpaceX REST calls

- GitHub: https://github.com/DrBChris/IBM-Data-Science-Capstone-Project/blob/main/DataCollection.ipynb

Flowchart:

- define URL
- request data
- convert json into a Pandas dataframe
- normalize data

Data Collection - Scraping

- Webscraping with BeautifulSoup
- From Wikipedia: List of Falcon 9 and Falcon Heavy launches
- GitHub: https://github.com/DrBChris/IBM-Data-Science-Capstone-Project/blob/main/webscraping.ipynb

Flowchart:

- define URL
- request data
- create BeautifulSoup object from parsed data
- extract tables
- fill Pandas dataframe with parsed data

Data Wrangling

- Identifying of missing values for each attribute
- Identifying the data types
- Creating a landing outcome label (either 1 = success or 0 = failure)

- GitHub: https://github.com/DrBChris/IBM-Data-Science-Capstone-Project/blob/main/DataWrangling.ipynb

EDA with Data Visualization

- Used scatterplots to visualize
  - Flight Number vs Launch Site
  - Payload vs Launch Site
  - Flight Number vs Orbit
  - Payload vs Orbit
- Bar Chart
  - Success vs Orbit
- Line Chart
  - Yearly success trend
- GitHub: https://github.com/DrBChris/IBM-Data-Science-Capstone-Project/blob/main/DataViz.ipynb

EDA with SQL

- List of SQL Queries:
  - Unique launch sites
  - 5 records that begin with ‘CCA’
  - Total payload carried by boosters launched by NASA (CRS)
  - Average payload mass carried by booster version F9 v1.1
  - Date when the first successful landing outcome in ground pad was achieved
  - Name of boosters with success (4000 < mass < 6000)
  - Total number of successful and failure missions
  - Name of booster versions which carried the maximum payload
  - Failed laning outcomes in drone ship, their booster version and launch sites for 2015
  - Ranking the number of landing outcomes between 2010-06-04 and 2017-03-20
- GitHub: https://github.com/DrBChris/IBM-Data-Science-Capstone-Project/blob/main/SQL.ipynb

Build an Interactive Map with Folium

- Visualizing all launch sites with circles
- Visualizing launch outcome with colored markers at each launch site
- Calcuating the distance between launch site and closest
    - coast line
    - highway
    - railway
    - city
- GitHub: https://github.com/DrBChris/IBM-Data-Science-Capstone-Project/blob/main/Folium.ipynb

Build a Dashboard with Plotly Dash

- Piechart to visualize successrates of all and individual launch sites
- Scatterplot to visualize the relation between outcome and payload including a slider to display individual payload ranges
- GitHub: https://github.com/DrBChris/IBM-Data-Science-Capstone-Project/blob/main/dash.py

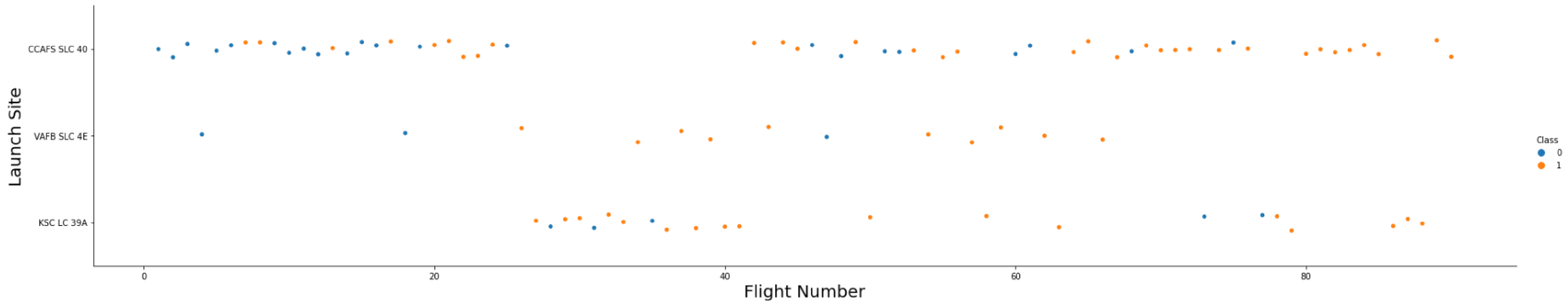
Predictive Analysis (Classification)

- We use different machine learning approaches to predict the outcome of a mission:
  - Logistic Regression
  - Support Vector Machines
  - Decision Trees
  - K Nearest Neighbor
- GridSearchCV was used to fine-tune hyperparameters
- GitHuB: https://github.com/DrBChris/IBM-Data-Science-Capstone-Project/blob/main/ML.ipynb

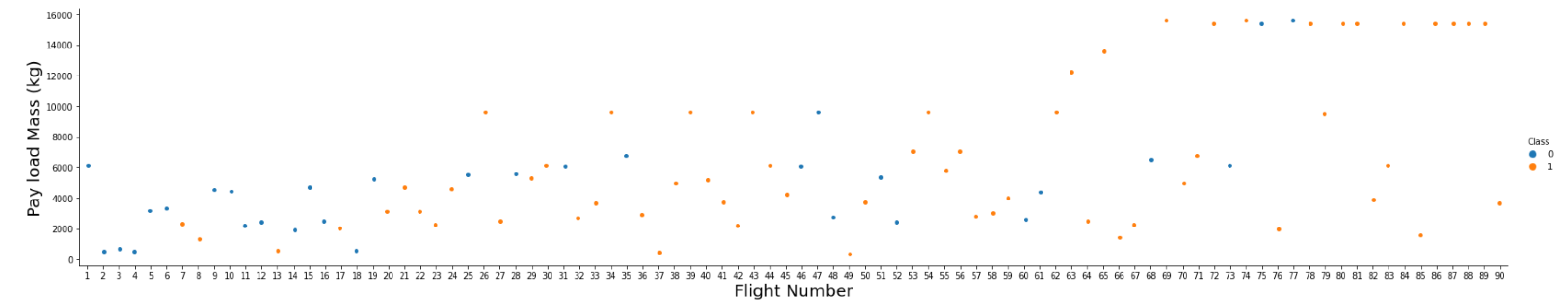
Section 2

Insights drawn from EDA

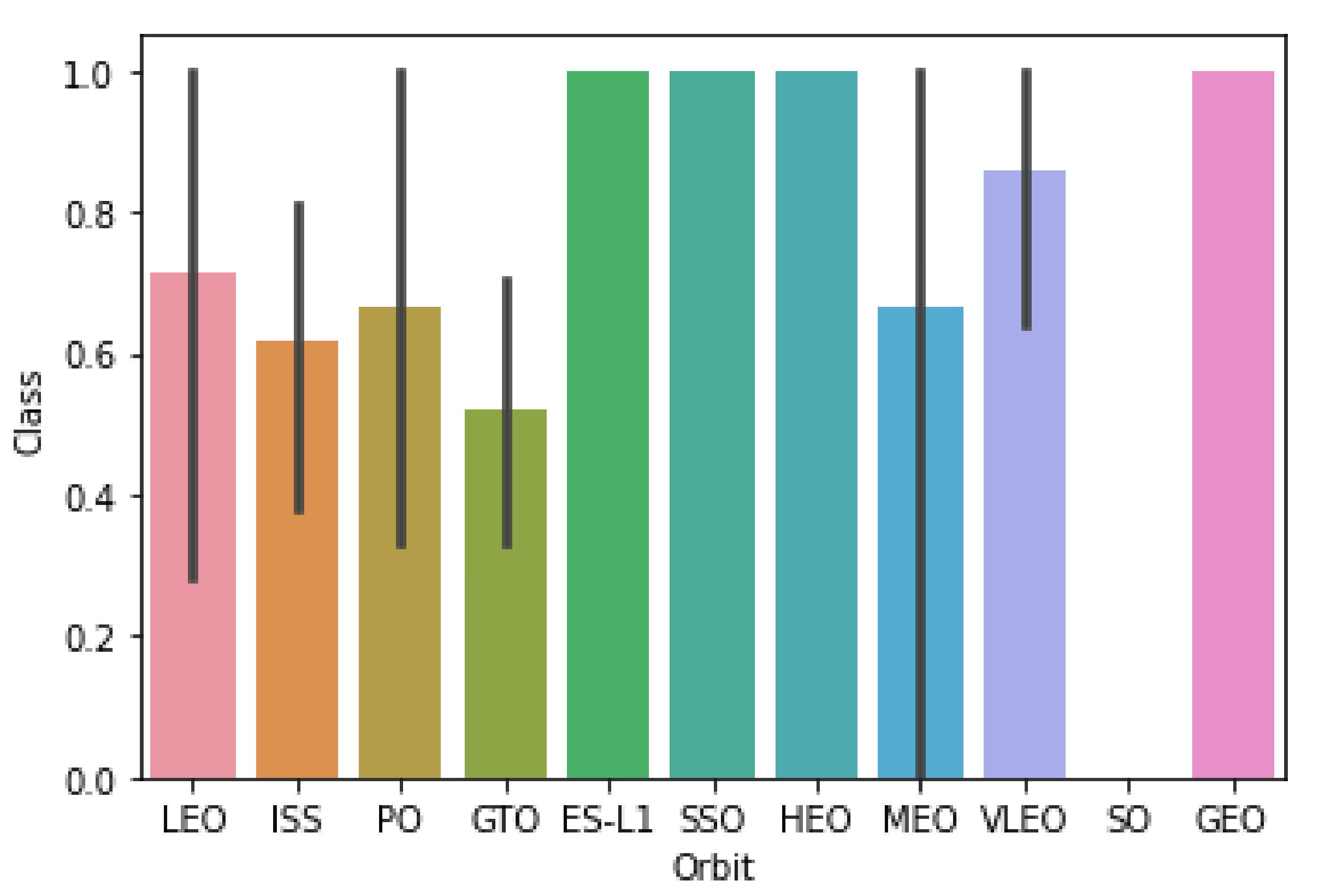
Flight Number vs. Launch Site



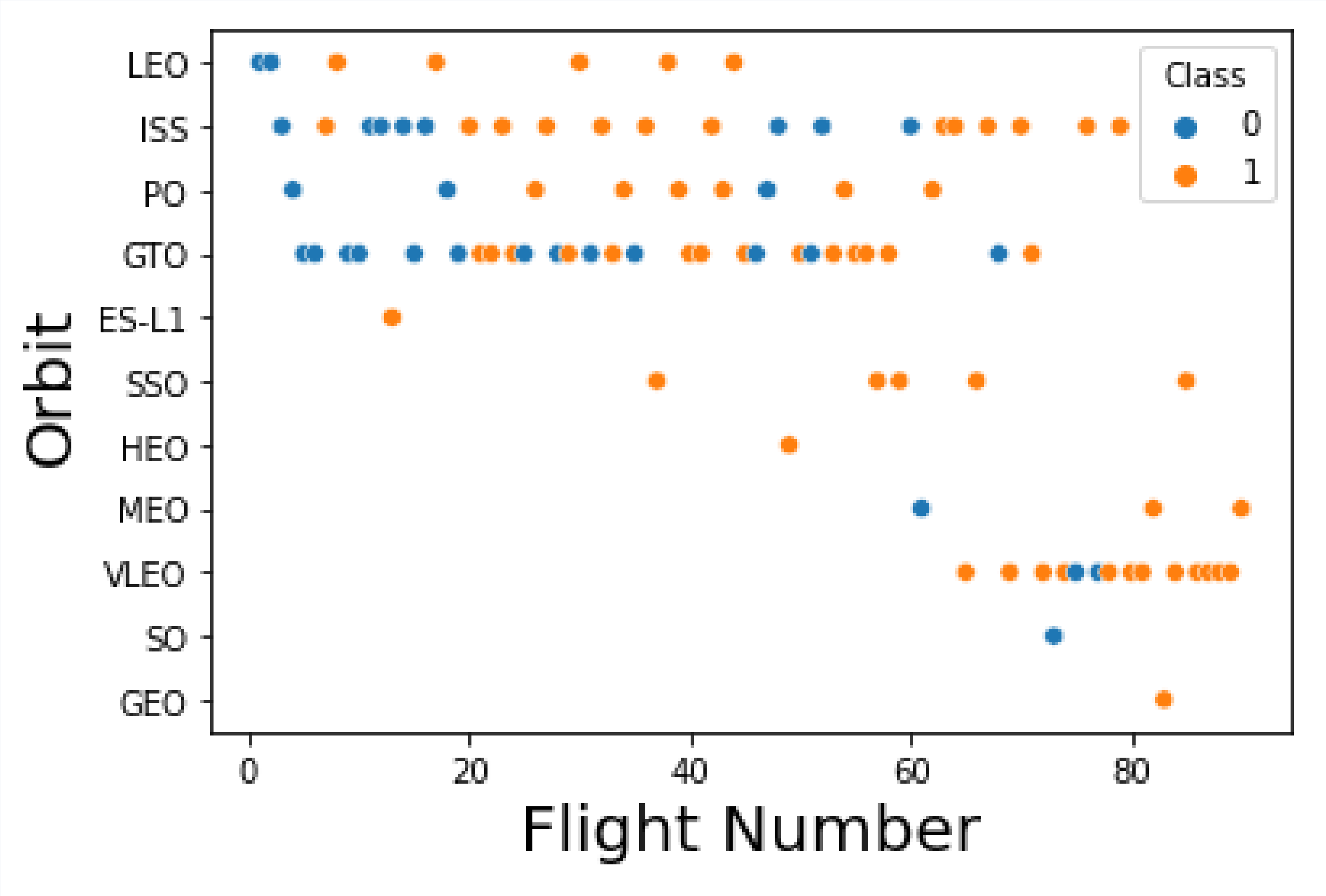
Payload vs. Launch Site



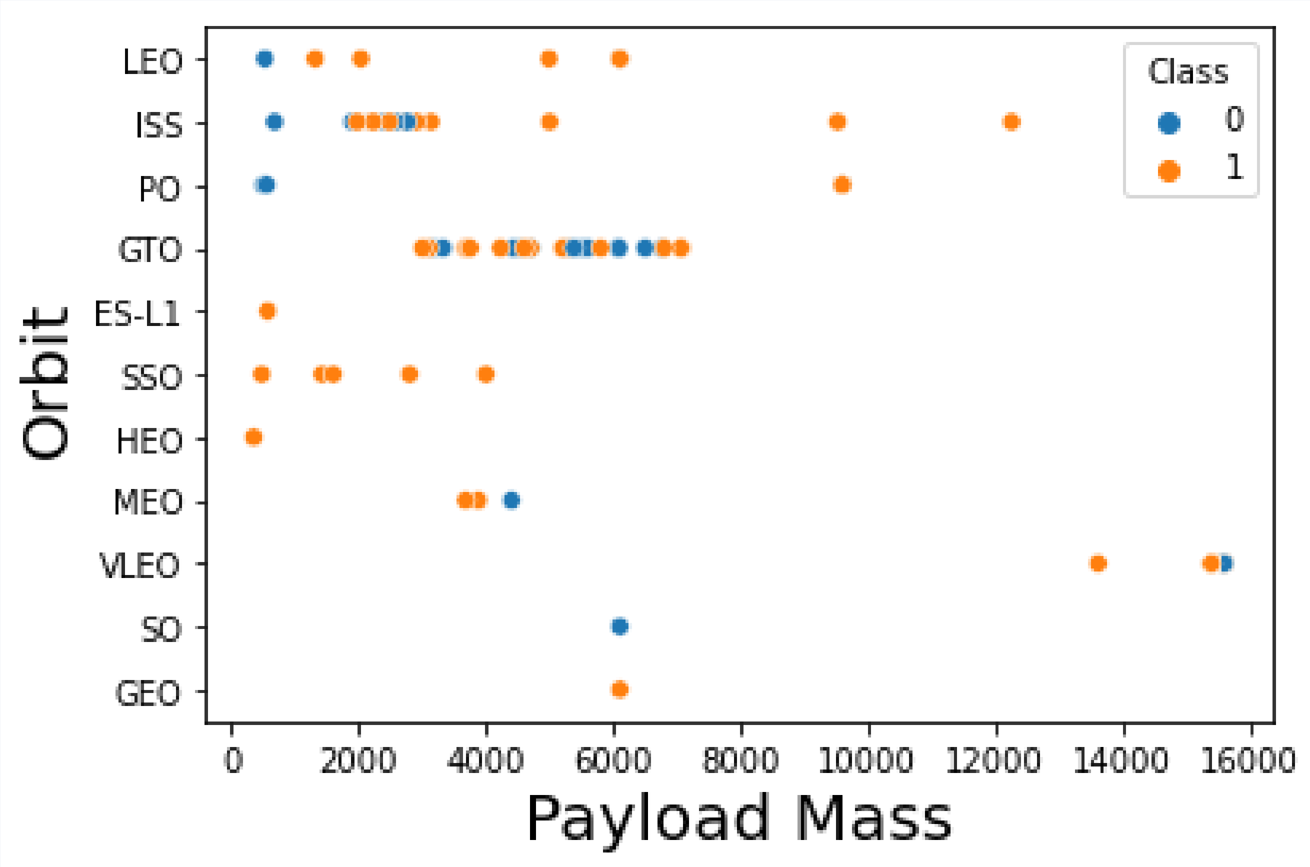
Success Rate vs. Orbit Type



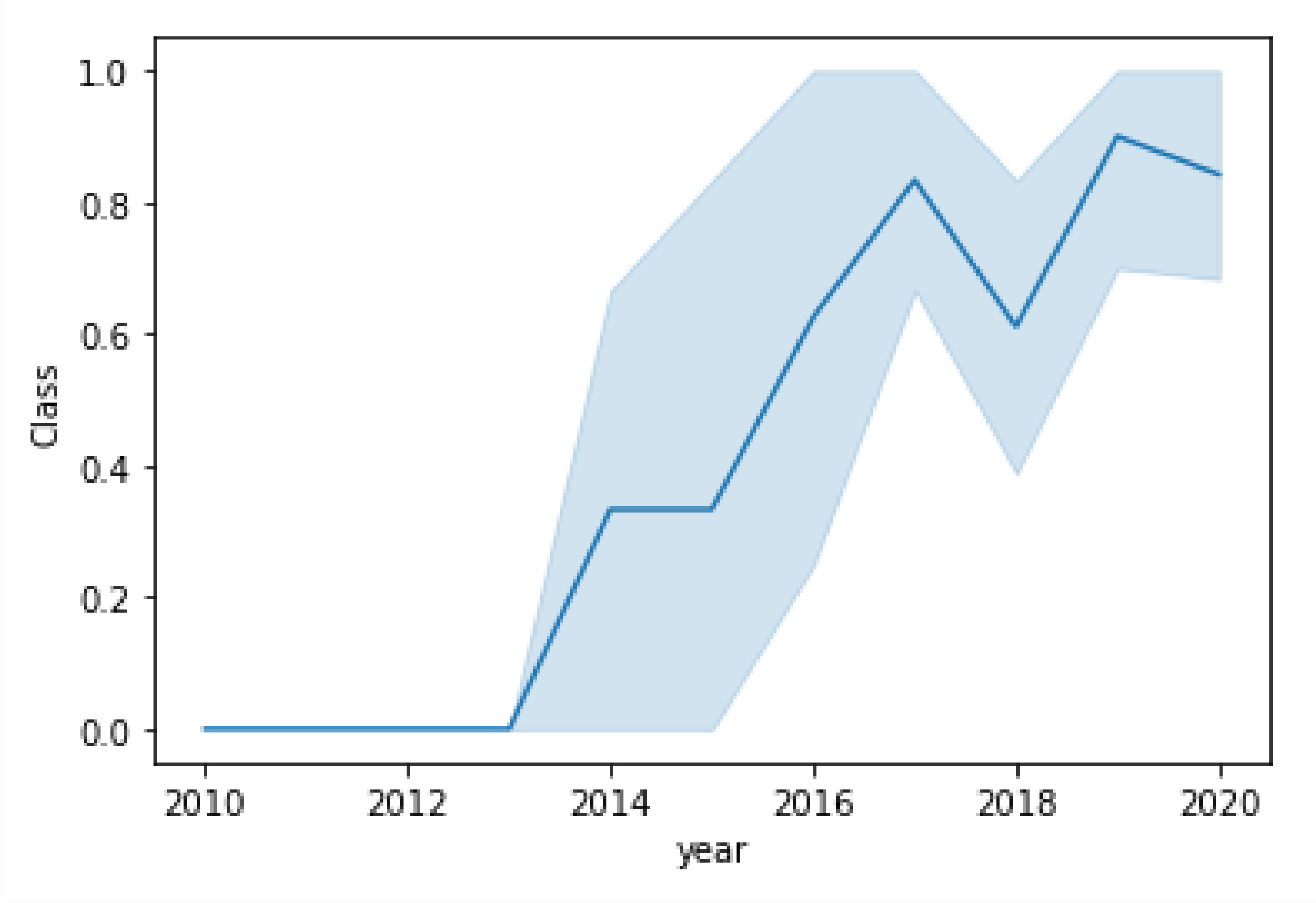
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

```
%%sql
SELECT DISTINCT launch_site
FROM SPACEXDATASET
```

```
 * ibm_db_sa://qsn36287:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

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

Launch Site Names Begin with 'CCA'

```
%%sql
SELECT *
FROM SPACEXDATASET
WHERE launch_site like 'CCA%'
LIMIT 5
```

```
 * ibm_db_sa://qsn36287:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
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 (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

```
%%sql
SELECT SUM(payload_mass__kg_)
FROM SPACEXDATASET
WHERE customer = 'NASA (CRS)'
```

```
 * ibm_db_sa://qsn36287:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

1
45596

Average Payload Mass by F9 v1.1

```
%%sql
SELECT AVG(payload_mass__kg_)
FROM SPACEXDATASET
WHERE booster_version = 'F9 v1.1'
```

```
 * ibm_db_sa://qsn36287:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

1
2928

First Successful Ground Landing Date

```
%%sql
SELECT MIN(Date)
FROM SPACEXDATASET
WHERE landing__outcome like 'Success%'
```

```
 * ibm_db_sa://qsn36287:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql
SELECT booster_version
FROM SPACEXDATASET
WHERE landing__outcome = 'Success (drone ship)'
    AND payload_mass__kg_ > 4000
    AND payload_mass__kg_ < 6000
```

```
 * ibm_db_sa://qsn36287:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
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

```
%%sql
SELECT COUNT(mission_outcome)
FROM SPACEXDATASET
WHERE mission_outcome like '%Success%'
```

```
 * ibm_db_sa://qsn36287:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

1
100

```
%%sql
SELECT COUNT(mission_outcome)
FROM SPACEXDATASET
WHERE mission_outcome like '%Failure%'
```

```
 * ibm_db_sa://qsn36287:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

1
1

Boosters Carried Maximum Payload

```
%%sql
SELECT booster_version, payload_mass__kg_
FROM SPACEXDATASET
WHERE payload_mass__kg_ = (
    SELECT MAX(payload_mass__kg_)
    FROM SPACEXDATASET)
```

```
 * ibm_db_sa://qsn36287:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

```
%%sql
SELECT booster_version, launch_site, landing__outcome
FROM SPACEXDATASET
WHERE landing__outcome like 'Failure (drone ship)'
AND DATE BETWEEN '2015-01-01' AND '2015-12-31'
```

```
 * ibm_db_sa://qsn36287:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

booster_version	launch_site	landing__outcome
F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

```
%%sql
SELECT landing__outcome, COUNT(landing__outcome)
FROM SPACEXDATASET
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY landing__outcome
ORDER BY COUNT(landing__outcome) DESC
```

```
 * ibm_db_sa://qsn36287:***@2d46b6b4-cbf6-40eb-bbce-6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32328/bludb
Done.
```

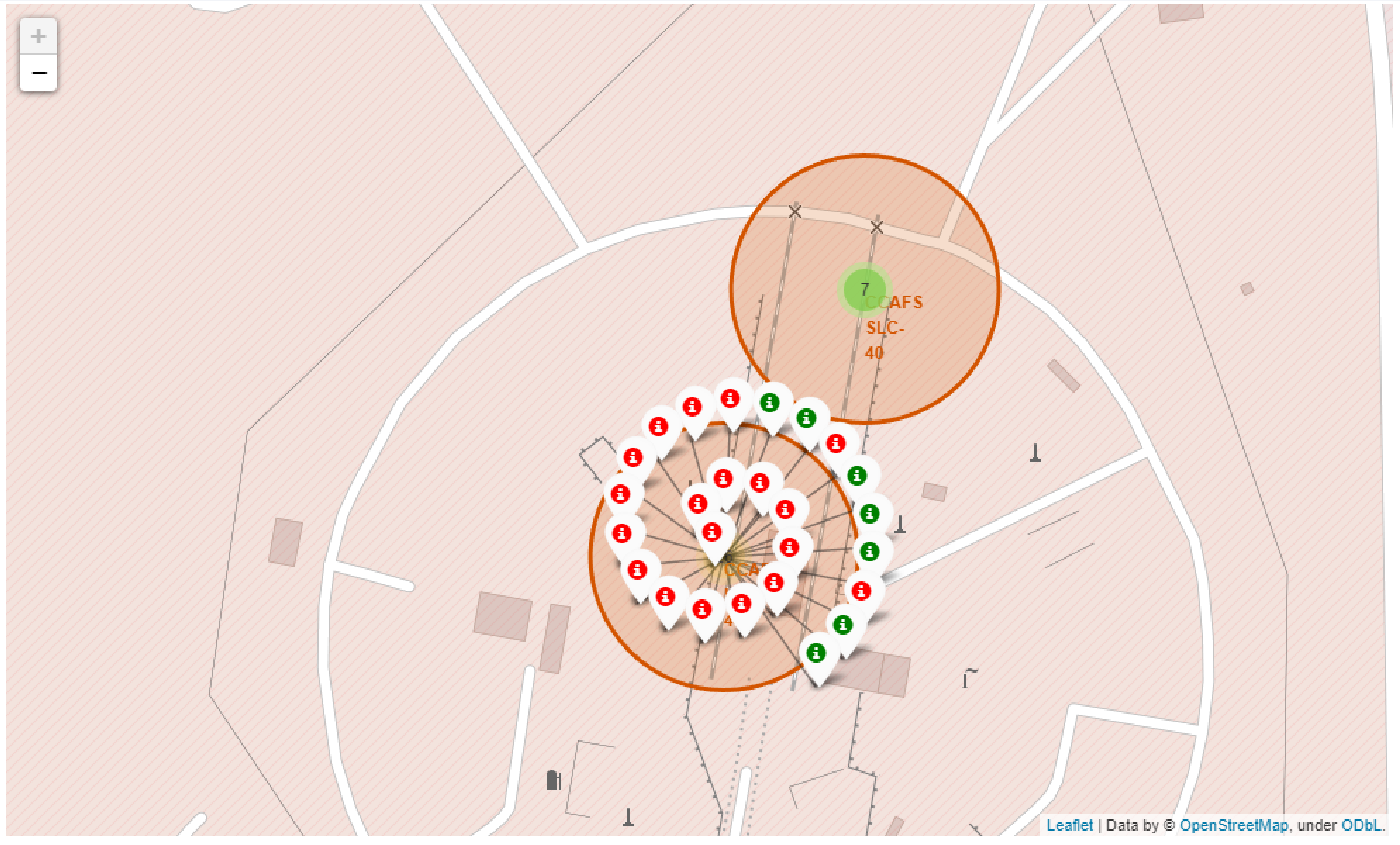
landing__outcome	2
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

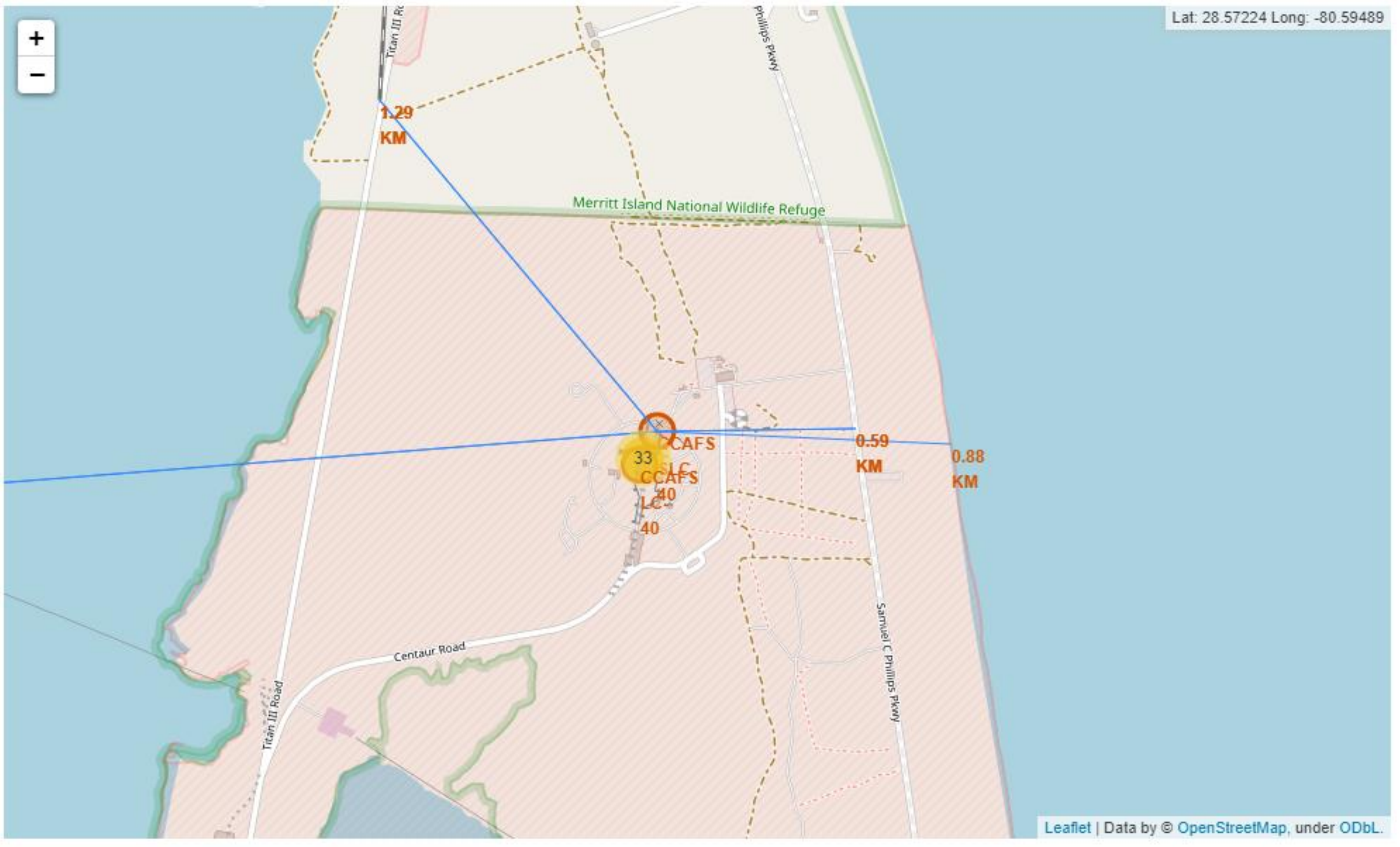
Launch Sites Proximities Analysis

SpaceX Launch Sites

Example of Color-Labeled Launch Outcomes



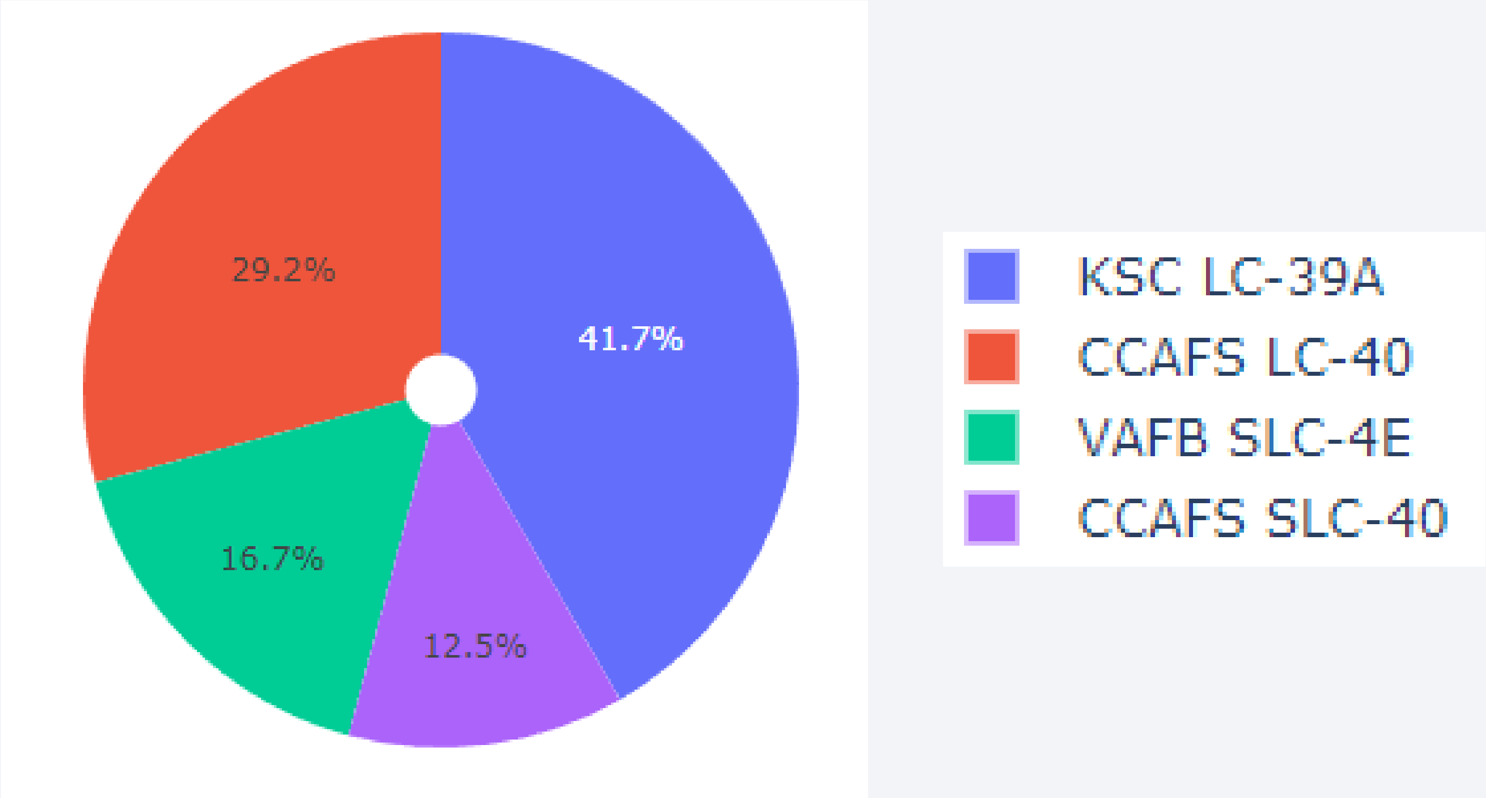
Distances from Launch Site



Section 4

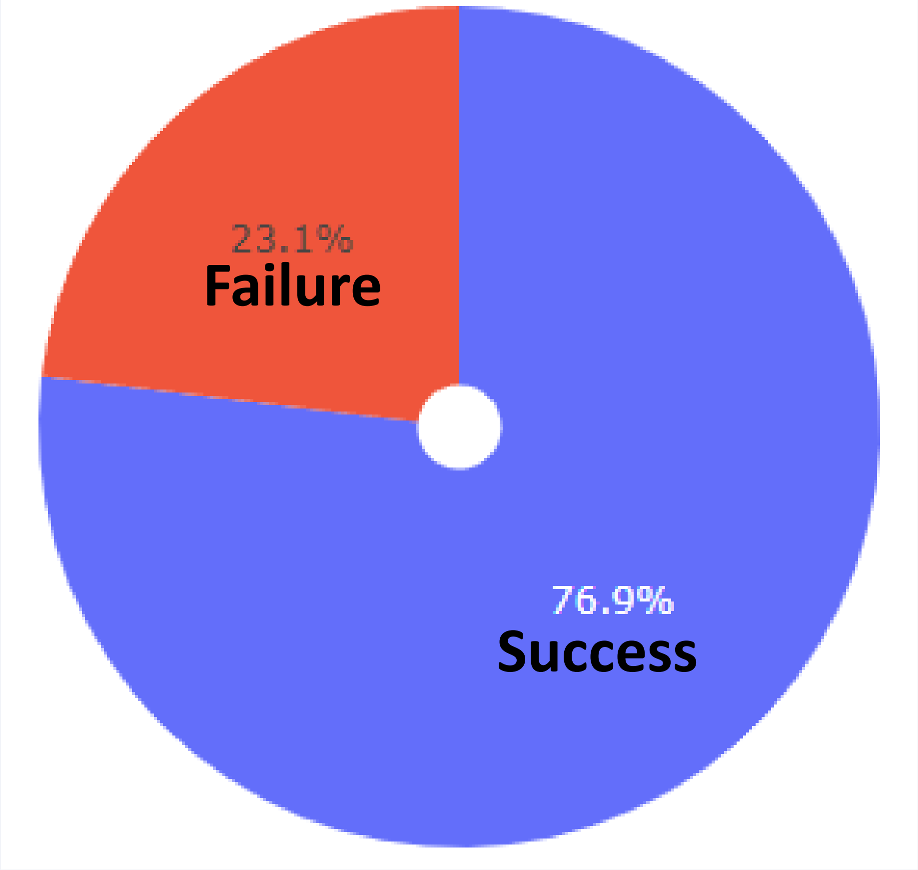
Build a Dashboard with Plotly Dash

Launch Success for all launch sites



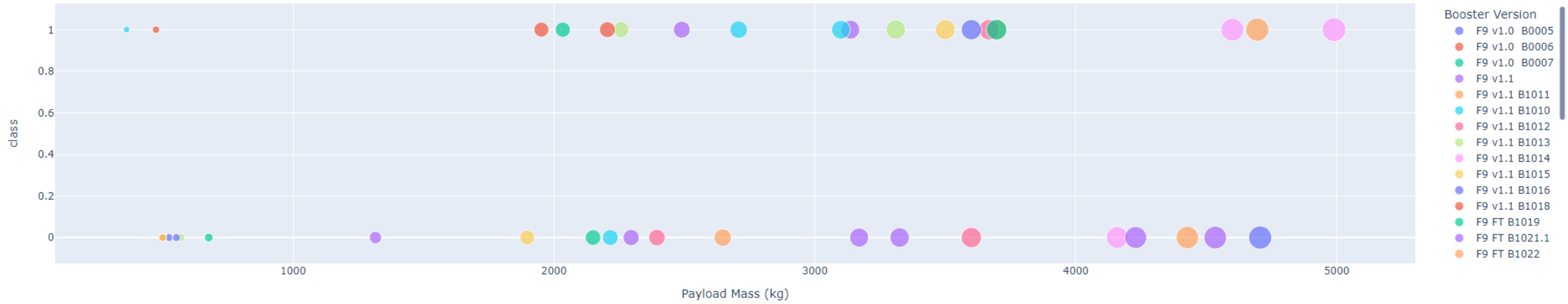
KSC LC-39A has to most success launching Falcon 9 rockets

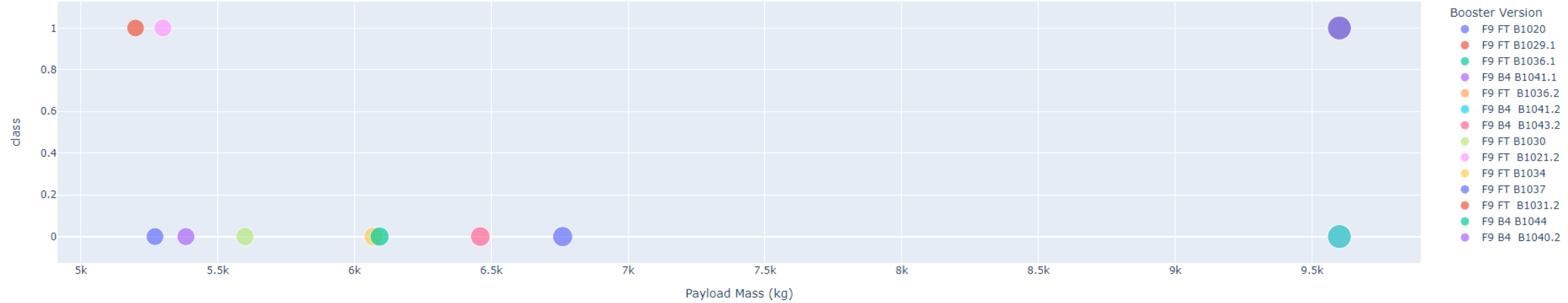
Total Success Launches for KSC LC-39A



KSC LC-39A has a 76.9% success rate

Low (<5000 kg) vs High (>5000 kg) Payload



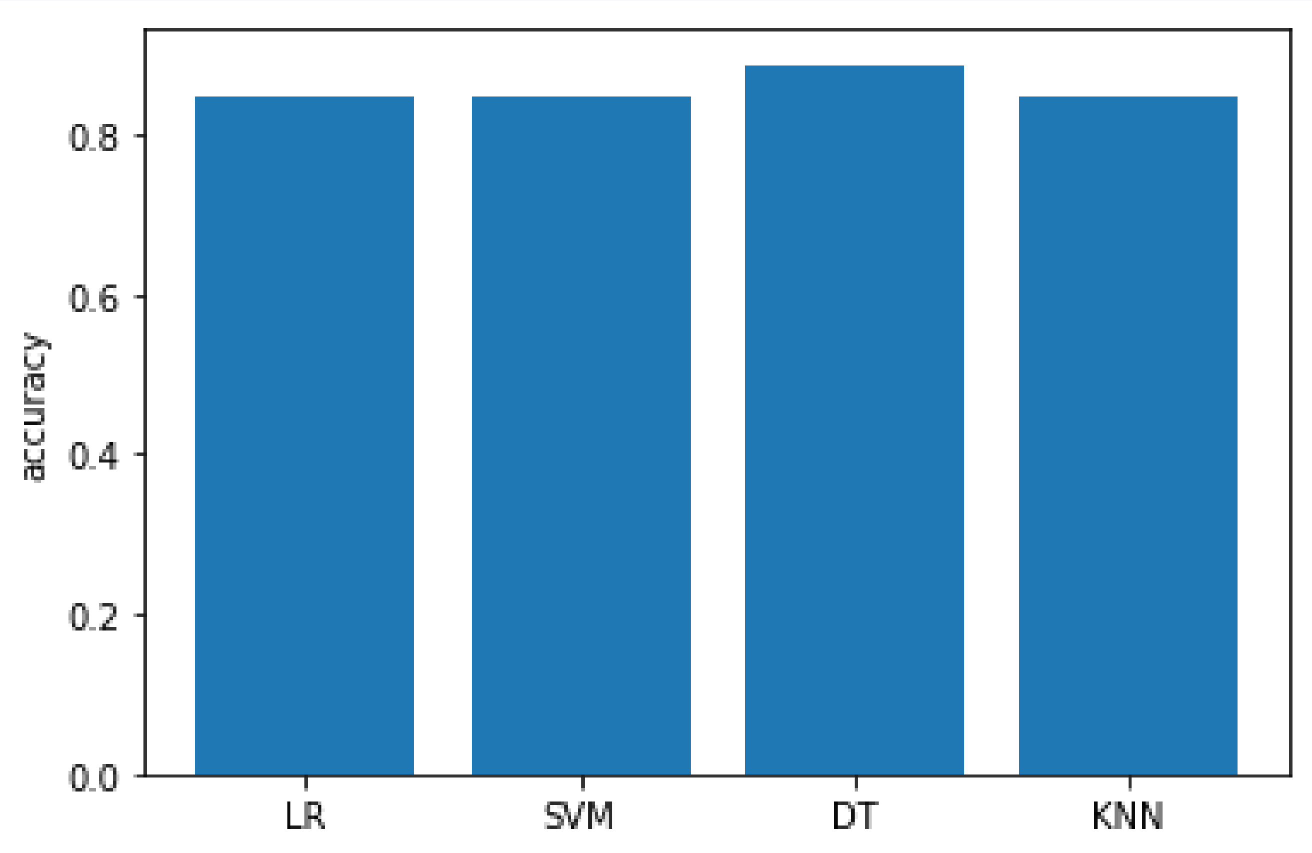


Lower payloads yield more successful missions

Section 5

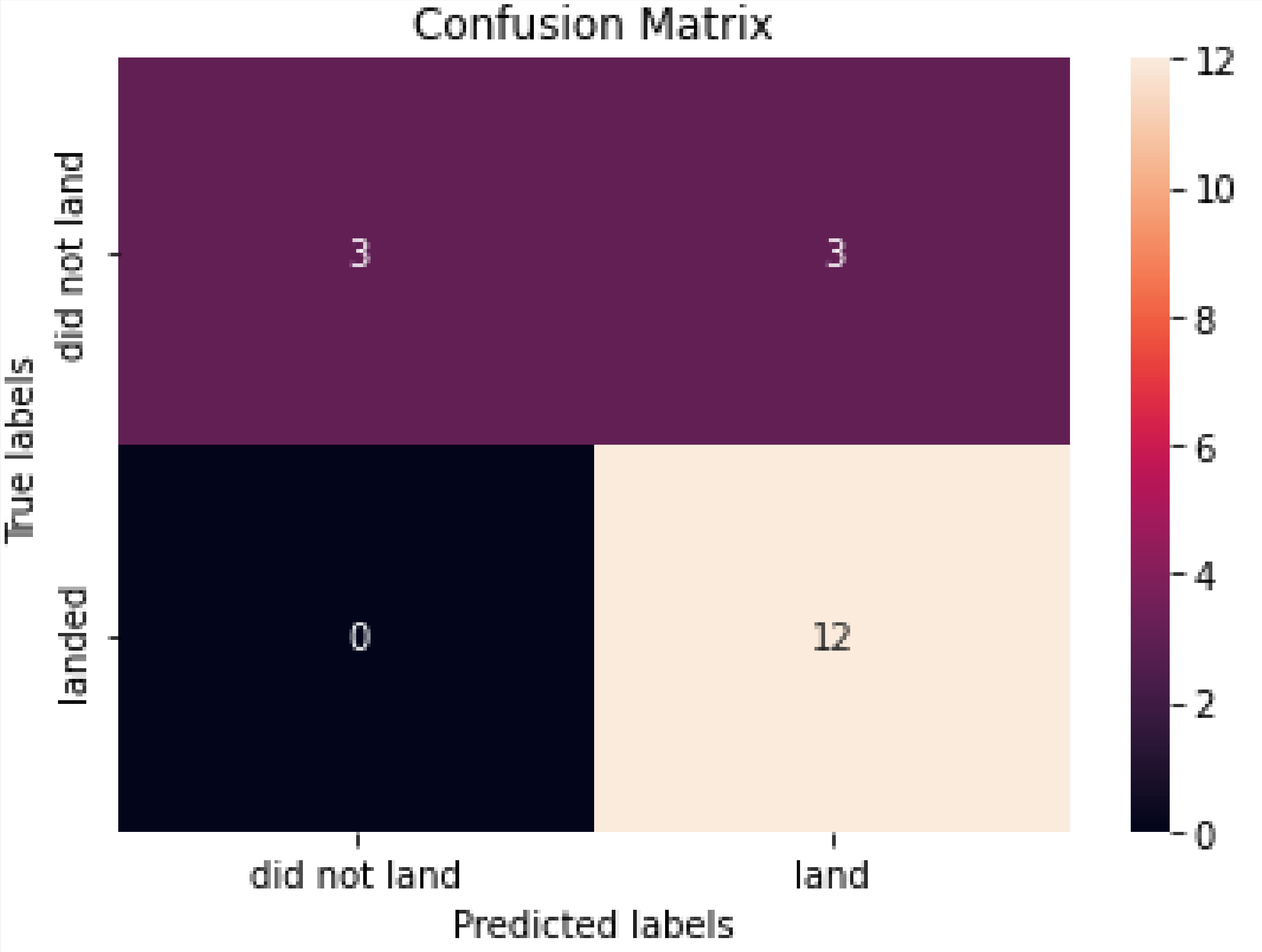
Predictive Analysis (Classification)

Classification Accuracy



Decision Trees (DT) achieved the highest accuracy

Confusion Matrix



Conclusions

- Since 2017 SpaceX has about 80% success rate over all missions
- The most successful launch site is KSC LC-39A
- ES-L1, SSO, HEO, GEO are the most successful orbits (100%), although some of these orbits have only been target once
- VLEO is a highly frequented orbit with a success rate of 87%
- Payloads below 5000 kg seem to be more successful
- Decision Trees are the best approach to predict the success of a mission

Thank you!