Winning Space Race with Data Science

Rauf O. Giwa
06/16/2022

Outline

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

Executive Summary

- Data from public SpaceX API and SpaceX Wikipedia page.
- Quantitative data on successful and failed landing mission.
- Data exploration using SQL, visualization, folium maps, and dashboards.
- Optimized Standardized data and used GridSearchCV to find best parameters for machine learning prediction models. Visualize accuracy score of all models.
- Four models Logistic Regression, Support Vector Machine, Decision Tree Classifier, and K Nearest Neighbors. accuracy > 83.33%.

- Summary of methodolo
- Summary of all results

Introduction

- Project background and context

- The commercial space age is here, Virgin Galactic. Rocket Lab, Blue Origin.
- most successful is SpaceX
  - Price per launch : $62m compared to $165m.
  - Strategy: can reuse the first stage

Problem

- Allon Musk, SpaceY CEO, wish to compete with SpaceX
- Use machine learning to predict fails and successful landing of falcon 9 launches from the Wikipedia page

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data collected from SpaceX public API and SpaceX Wikipedia page
- Perform data wrangling
  - Quantify landings as successful of failed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - ML models were optimized to generate best prediction.

Data Collection

Collecting data involved use of many of APIs. E.g. http requests from SpaceX public API and scraping data from a tables in SpaceX's Wikipedia entry.

SpaceX data: FlightNumber, Date, BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights, GridFins,

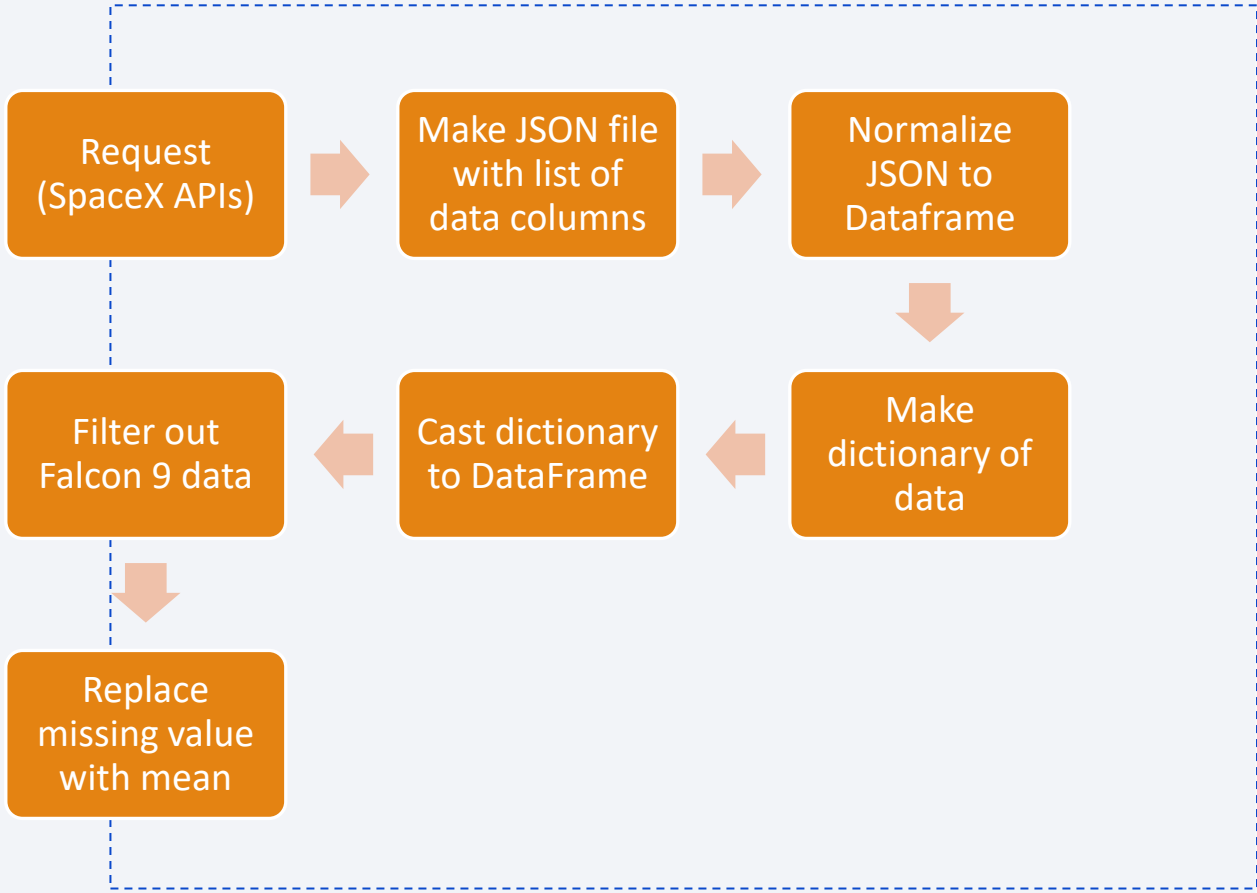
Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude

Wikipedia data: Flight No., Launch site, Payload, PayloadMass, Orbit, Customer, Launch outcome, Version Booster, Booster landing, Date, Time

Data Collection – SpaceX API

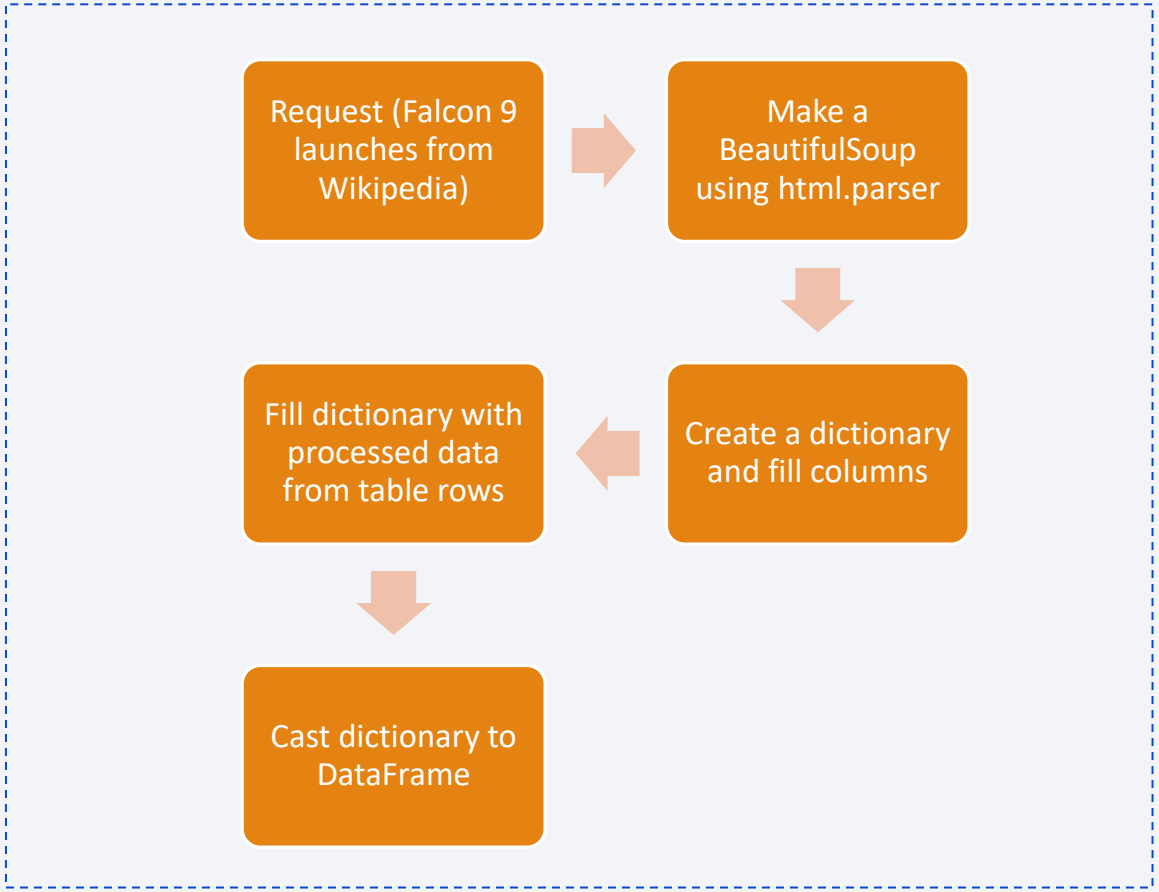
GitHub URL:

github.com/raufgiwa-scikits/IBM-Data-Science-Professional-Certificate/blob/main/Applied_Data_Science_Capstone/Week1-spacex-data-collection-api.ipynb



Data Collection - Scraping

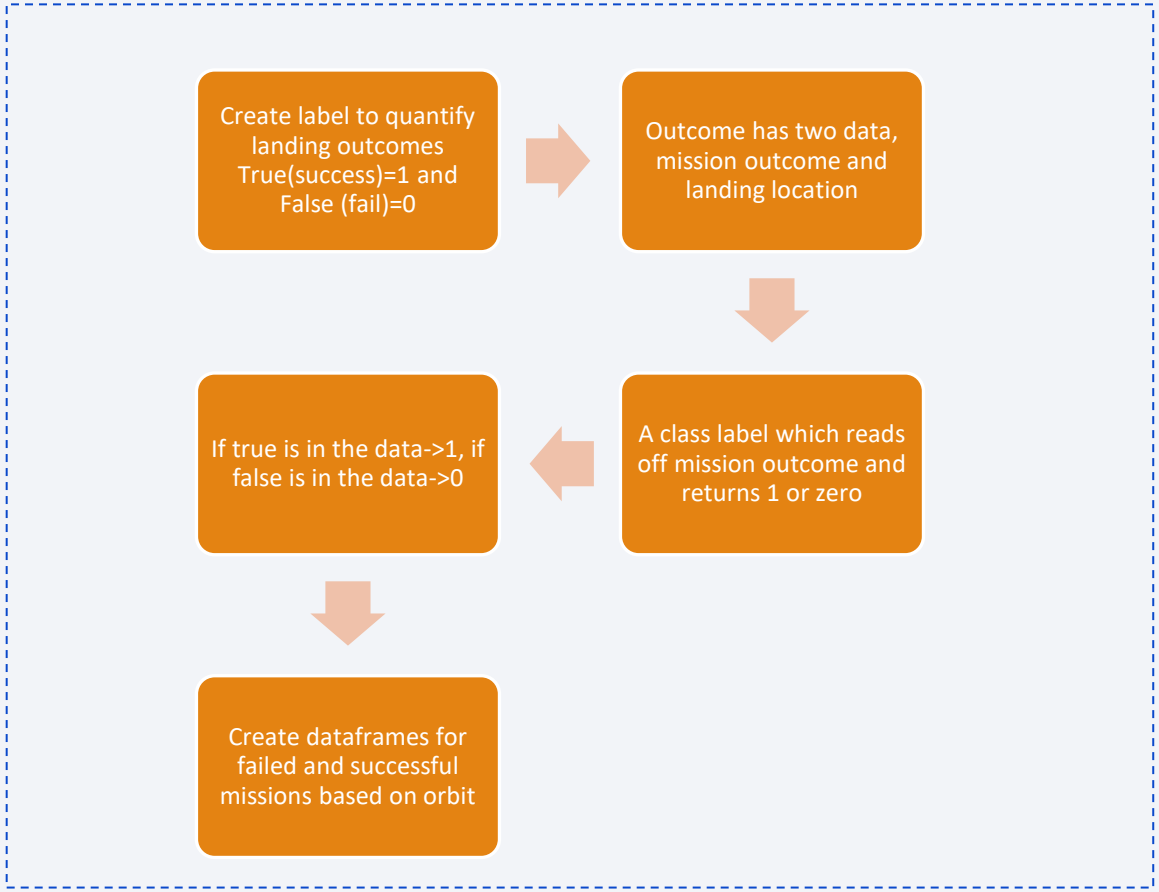
Add the GitHub URL :
github.com/raufgiwa-scikits/IBM-Data-Science-Professional-Certificate/blob/main/Applied_Data_Science_Capstone/Week1-DataCollection%20with%20webscraping.ipynb



Data Wrangling

Describe how data were processed

- Add the GitHub URL : github.com/raufgiwa-scikits/IBM-Data-Science-Professional-Certificate/blob/main/Applied_Data_Science_Capstone/Week1-Data%20wrangling.ipynb



EDA with Data Visualization

Exploratory Data Analysis performed on variables Flight Number, Payload Mass, Launch Site, Orbit, Class and Year.

Plots Used:

Scatter: Flight Number vs. Payload Mass, Flight Number vs. Launch Site, Payload Mass vs. Launch Site, Flight Number vs. Orbit, Payload vs Orbit.

Bar chart: Orbit vs. Success Rate.

Line chart: Success Yearly Trend

Scatter plots, line charts, and bar plots were used to visually observe relationships between variables to see if an relationship exists.

GitHub url: github.com/raufgiwa-scikits/IBM-Data-Science-Professional-Certificate/blob/main/Applied_Data_Science_Capstone/WEEK2-eda-dataviz.ipynb

EDA with SQL

Data set uploaded into IBM DB2 Database.

SQL performed using Python-SQL commands.

Queries gave deeper insight into the dataset.

Queries gave information about mission outcomes, dates , launch sites, pay load masses, and booster versions, and landing outcomes

GitHub url: https://github.com/raufgiwa-scikits/IBM-Data-Science-Professional-Certificate/blob/main/Applied_Data_Science_Capstone/Week2-eda-sql-coursera_sqllite.ipynb

```
%sql select DISTINCT launch_site from SPACEXTBL
```

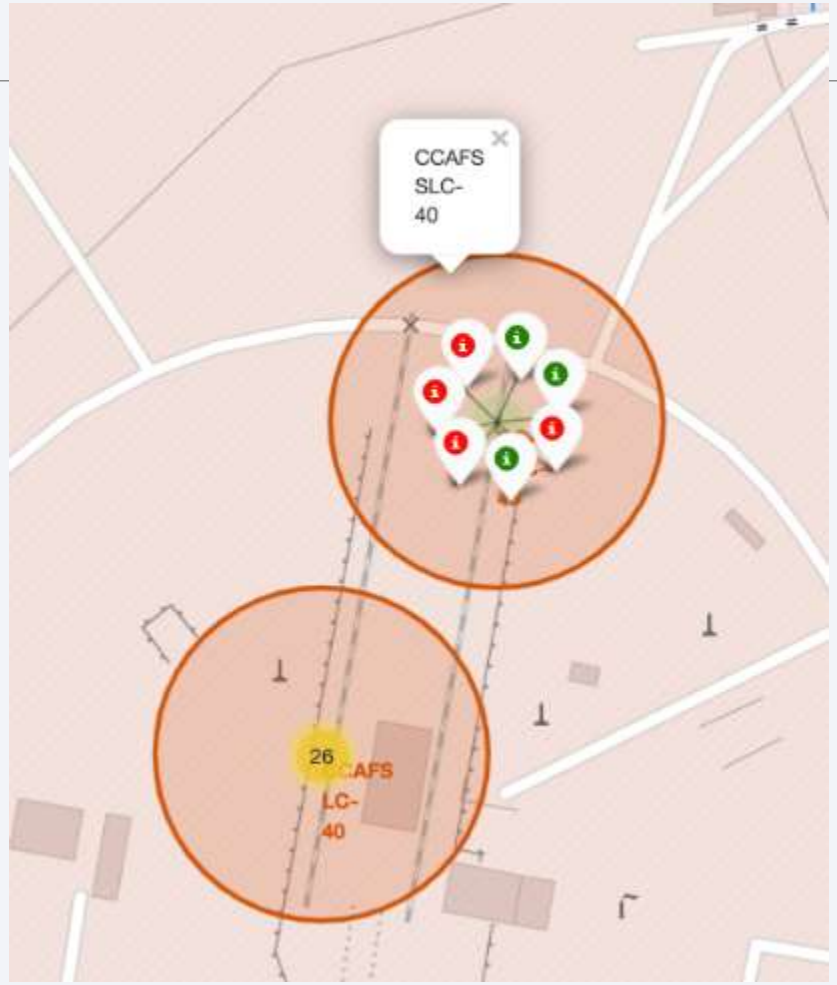
```
 * ibm_db_sa://fcd44147:***@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32459/bludb
Done.
```

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

Build an Interactive Map with Folium

Folium maps with markers representing location of launch site. Green/red for successful/failed missions. Circles are area of launch site Launch Sites, successful and unsuccessful landings, and a proximity example to key locations: Railway, Highway, Coast, and City.

GitHub url: https://github.com/raufgiwa-scikits/IBM-Data-Science-Professional-Certificate/blob/main/Applied_Data_Science_Capstone/Week3_%20Interactive_Visua_Analytics_and_Dashboard.ipynb



Build a Dashboard with Plotly Dash

Dashboard includes pie chart and scatter plot.

Pie chart can select all launch and show individual launch site success rates or for each launch sites and show success and fail rate.

Scatter plot takes two inputs for All sites or individual site and payload mass on a slider between 0 and 10000 kg.

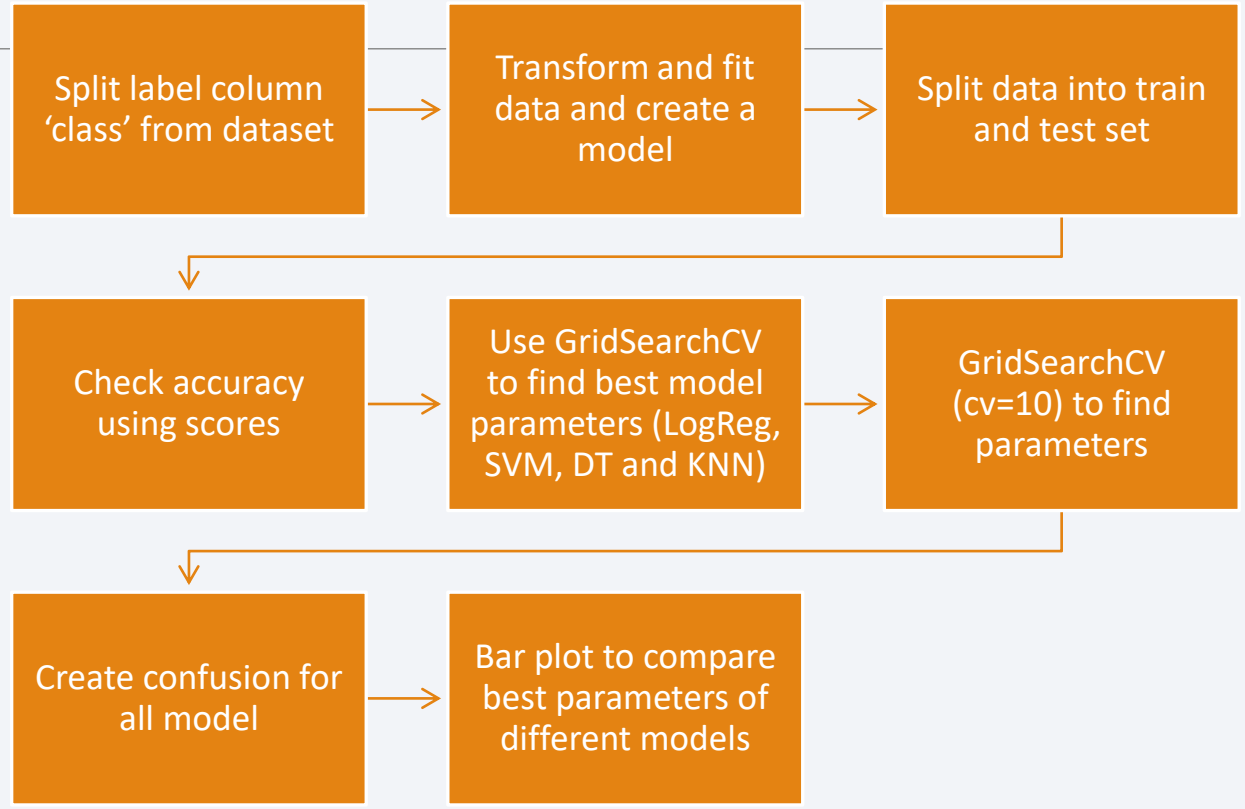
The scatter plot can help us see how success varies across launch sites, payload mass, and

booster version category.

Add the GitHub URL: https://github.com/raufgiwa-scikits/IBM-Data-Science-Professional-Certificate/blob/main/Applied_Data_Science_Capstone/spacex_dash_app.py

Predictive Analysis (Classification)

GitHub URL:
https://github.com/raufgiwa-scikits/IBM-Data-Science-Professional-Certificate/blob/main/Applied_Data_Science_Capstone/Week4%20Machine%20Learning%20Prediction.ipynb

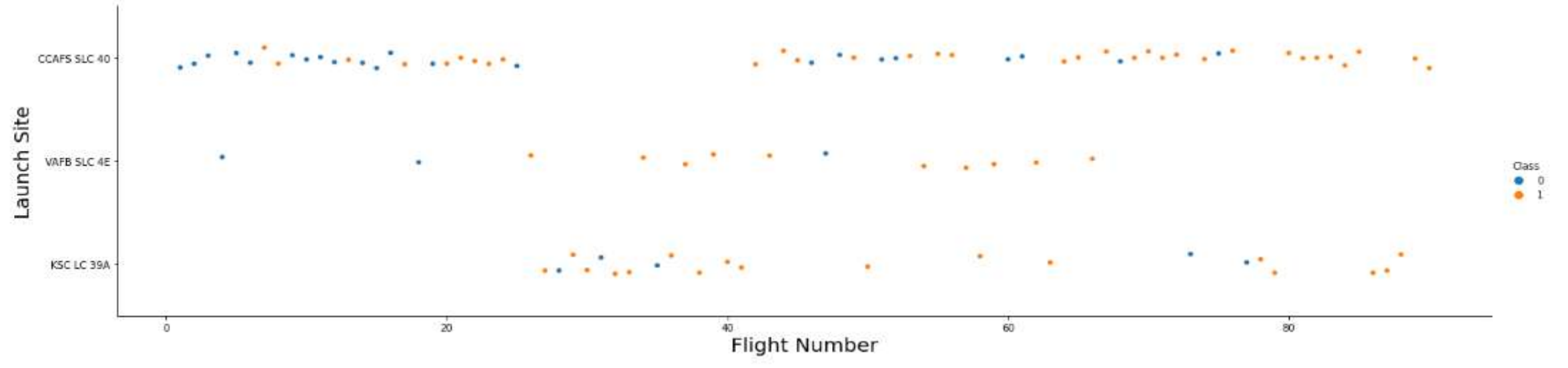


Results

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


Section 2
Insights drawn from EDA

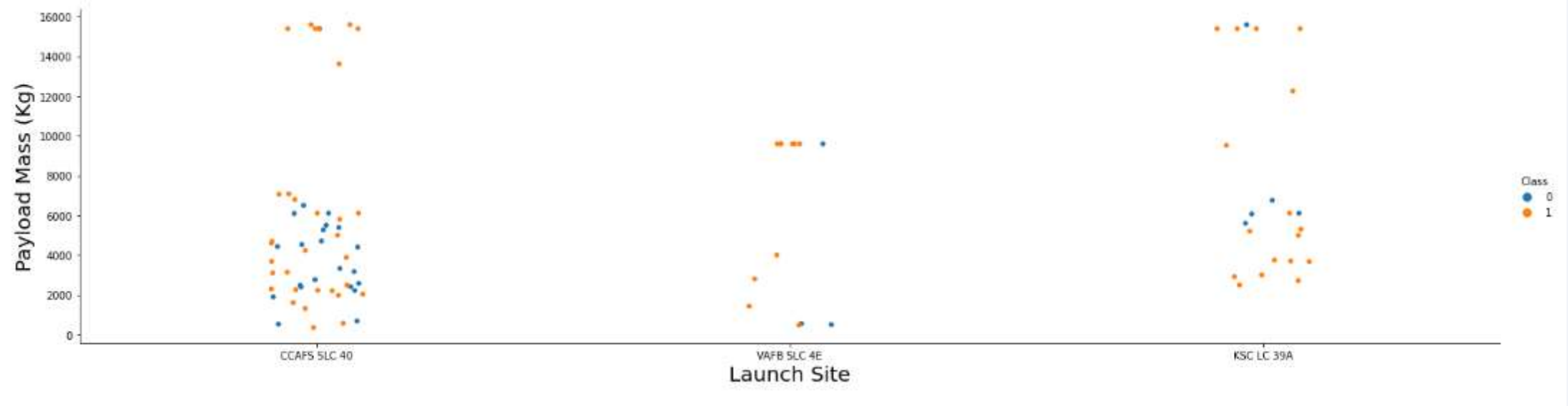
Flight Number vs. Launch Site



Blue indicates successful launch; Orange indicates unsuccessful launch.

- Graphic suggests an increased success rate over time. A plausible reason is an improvement in flight conditions . CCAFS appears to be the main launch site as it has the most volume.

Payload vs. Launch Site



Blue indicates successful launch; Orange indicates unsuccessful launch.

Payload mass appears to fall mostly between 0-6000 kg. Different launch sites also seem to use different payload mass

Success Rate vs. Orbit Type

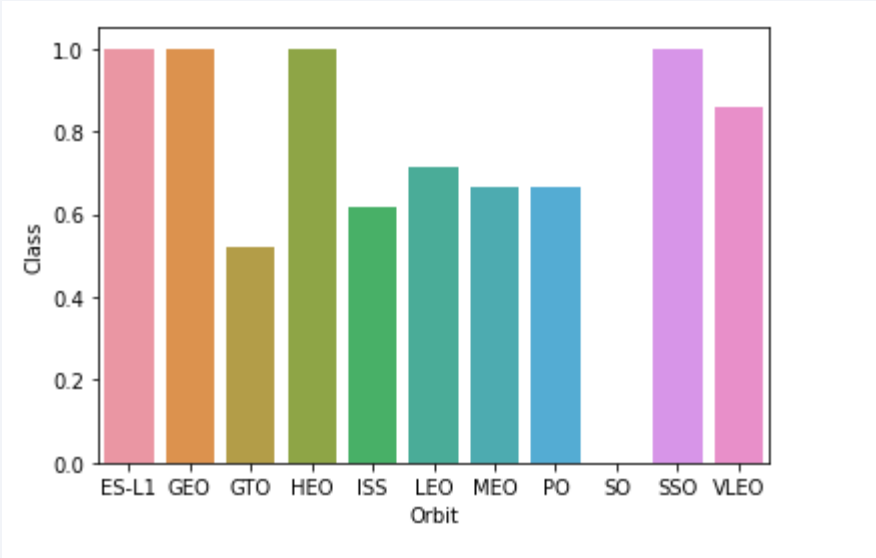
Success Rate Scale with 0 as fail and 1 as success

ES-L1 (1), GEO (1), HEO (1) have 100% success rate (sample sizes in parenthesis) SSO (5) has 100% success rate

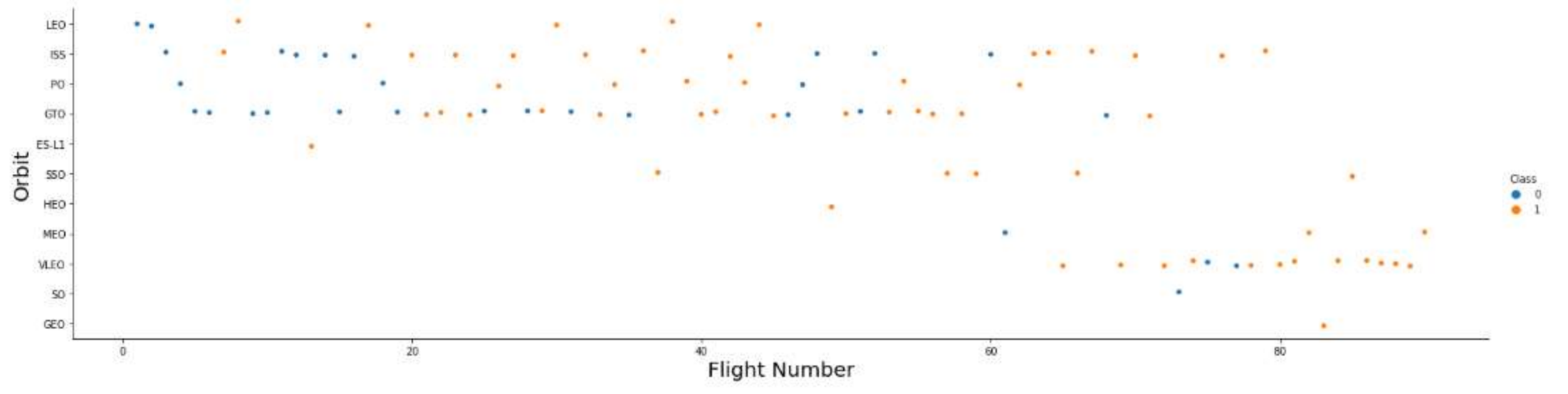
VLEO (14) has decent success rate and attempts

SO (1) has 0% success rate

GTO (27) has the around 50% success rate but largest sample



Flight Number vs. Orbit Type

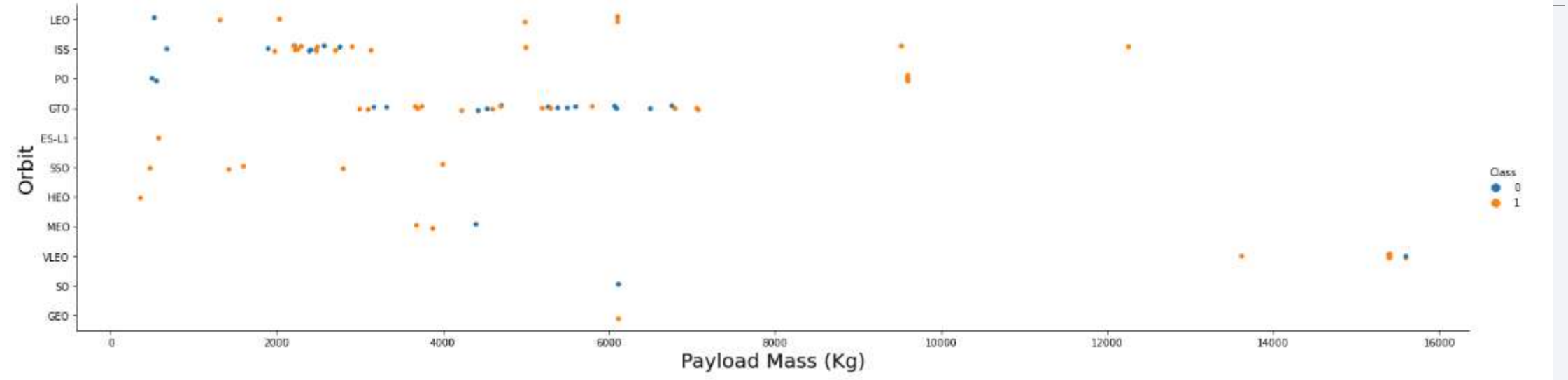


Blue indicates successful launch; Orange indicates unsuccessful launch.

Launch Orbit preferences changed over Flight Number. Launch Outcome seems to correlate with this preference.

SpaceX started with LEO orbits which saw moderate success LEO and returned to VLEO in recent launches SpaceX appears to perform better in lower orbits or Sun-synchronous orbits

Payload vs. Orbit Type



Blue indicates successful launch; Orange indicates unsuccessful launch.

Payload mass correlates with orbit

Relatively low payload mass were sent into LEO and SSO

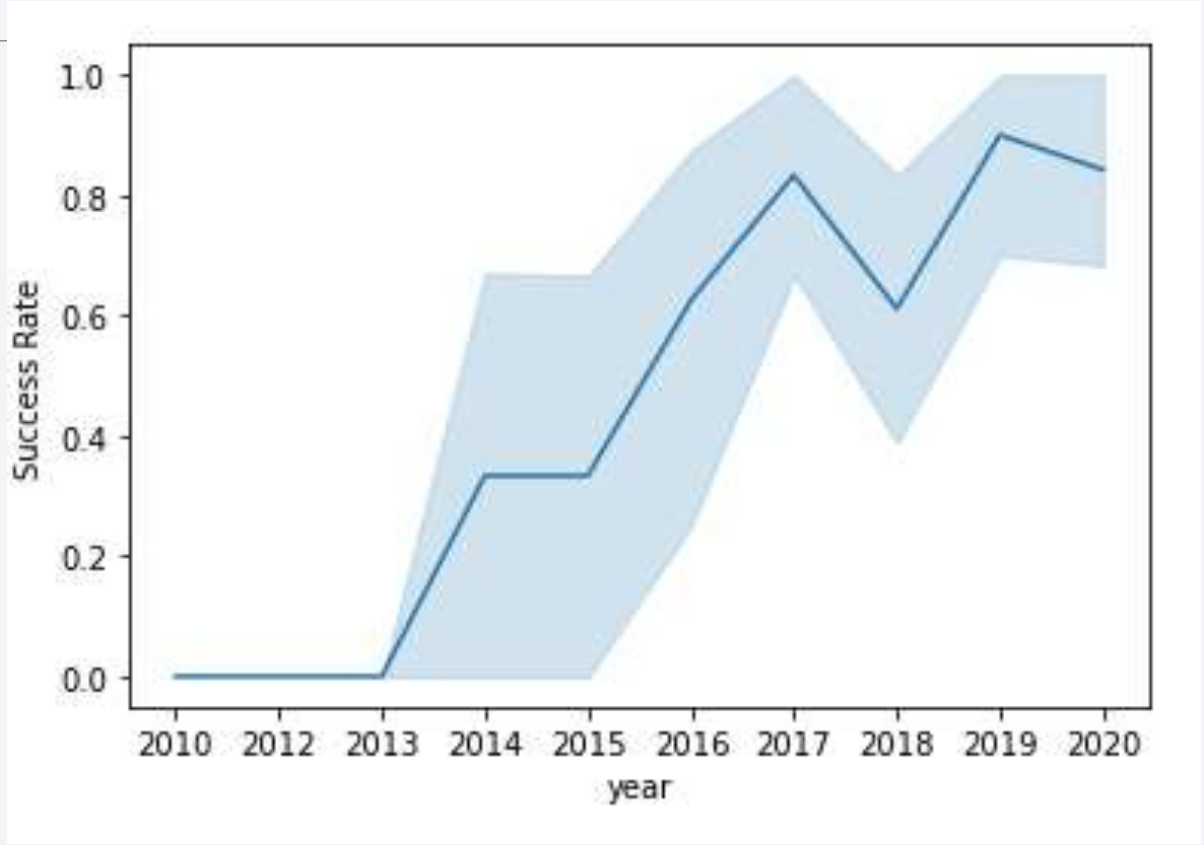
Launch Success Yearly Trend

Line : success rate:

light blue : 95% confidence

A generally increase in success over time since 2013 with a slight dip in 2018

Success since 2016 is around 0.8



All Launch Site Names

Query unique launch site names from database.

CCAFS LC-40 and CCAFS SLC-40 likely all represent the same site.

```
%sql select DISTINCT launch_site from SPACEXTBL
```

```
 * ibm_db_sa://fcd44147:***@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32459/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 SPACEXTBL  WHERE launch_site LIKE 'CCA%' FETCH FIRST 5 ROWS ONLY
```

```
 * ibm_db_sa://fcd44147:***@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32459/
ludb
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

First 5 records where launch sites begin with `CCA`

Total Payload Mass

```
%sql select sum(payload_mass__kg_)  AS total_payload_mass from SPACEXTBL  WHERE customer LIKE '%NASA%(CRS)%'
```

```
 * ibm_db_sa://fcd44147:***@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32459/bl
udb
Done.
```

total_payload_mass
48213

This query sums the total payload mass in kg where NASA was the customer.

CRS stands for Commercial Resupply Services which indicates that these payloads were sent to the International Space Station (ISS).

Average Payload Mass by F9 v1.1

```
%sql select AVG(payload_mass__kg_) AS average_payload_mass from SPACEXTBL  WHERE booster_version LIKE '%F9 v1.1%'
```

```
 * ibm_db_sa://fcd44147:***@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32459/bl
udb
Done.
```

average_payload_mass
2534

This query calculates average payload mass or launches with booster version F9 v1.1

Average payload mass of F9 1.1 is on the lowed end of payload mass range

First Successful Ground Landing Date

```
%sql select MIN(DATE) from SPACEXTBL  WHERE mission_outcome LIKE '%Success%'
```

```
 * ibm_db_sa://fcd44147:***@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0tgtu
0lqde00.databases.appdomain.cloud:32459/bludb
Done.
```

1
2010-06-04

This query returns the first successful ground pad landing date.

First ground pad landing was in 2010.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql select booster_version from SPACEXTBL WHERE (mission_outcome like '%Success%')
AND (payload_mass__kg_ BETWEEN 4000 AND 6000) AND (landing__outcome like '%Success%(drone ship)%')
```

```
 * ibm_db_sa://fcd44147:***@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0tgtu0lqde00.databases.appdomain.cl
d:32459/bludb
Done.
```

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

This query returns booster versions that had successful drone ship landings and a payload mass between 4000 and 6000 kg.

Total Number of Successful and Failure Mission Outcomes

```
%sql SELECT mission_outcome, count(*) as total FROM SPACEXTBL GROUP by mission_outcome ORDER BY mission_outcome
```

```
 * ibm_db_sa://fcd44147:***@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32459/bludb
Done.
```

mission_outcome	total
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Query return the total number of successful and failure mission outcomes

Most SpaceX missions were successful.

Boosters Carried Maximum Payload

This query lists the names of the booster which have carried the maximum payload mass

Boosters used for max payload are in the same family of boosters

```
%%sql select booster_version from SPACEXTBL where
payload_mass__kg_=(select max(payload_mass__kg_) from SPACEXTBL)
```

```
 * ibm_db_sa://fcd44147:***@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0
tgtu0lqde00.databases.appdomain.cloud:32459/bludb
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

This query returns the Month, Landing Outcome, Booster Version, Payload Mass (kg), and Launch site of the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

In [44]:
```
%%sql select  landing__outcome, booster_version, launch_site
from SPACEXTBL where DATE like '%2015%'
AND landing__outcome like '%Failure%drone ship%'
```

```
 * ibm_db_sa://fcd44147:***@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32459/bludb
Done.
```

Out[44]:

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

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

This query ranks 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

There are two types of successful landing outcomes: drone ship and ground pad landings.

There were 8/7 successful/failure landings in total during this time period

```
%%sql select landing__outcome, count(*) as count from SPACEXTBL
where Date between '2010-06-04' AND '2017-03-20'
GROUP by landing__outcome ORDER BY count Desc
```

```
 * ibm_db_sa://fcd44147:***@9938aec0-8105-433e-8bf9-0fbb7e483086.c1ogj3sd0
tgtu0lqde00.databases.appdomain.cloud:32459/bludb
Done.
```

landing__outcome	COUNT
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

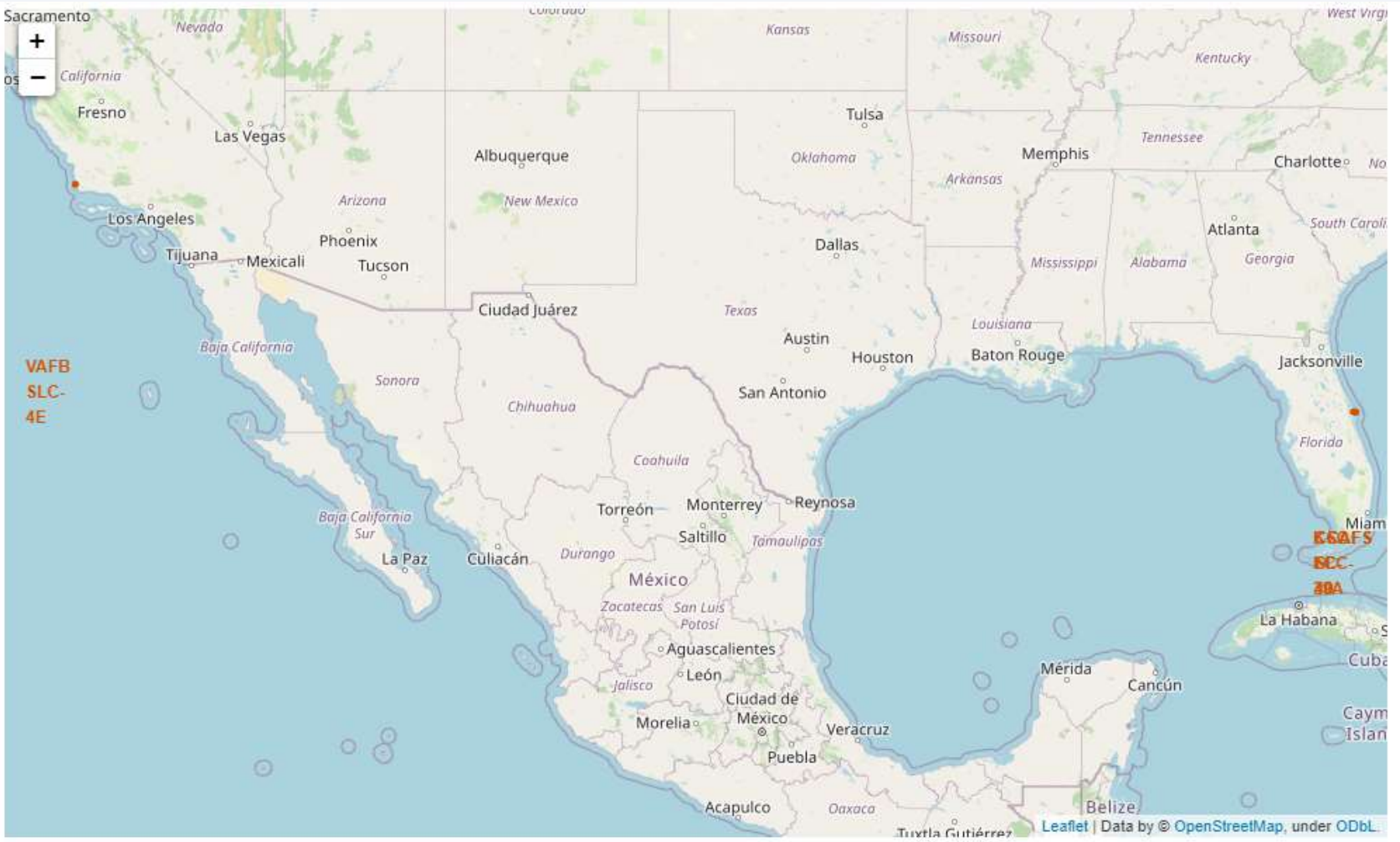
<Folium Map Screenshot 1>

Replace <Folium map screenshot 1> title with an appropriate title

Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map

Explain the important elements and findings on the screenshot

Launch site location

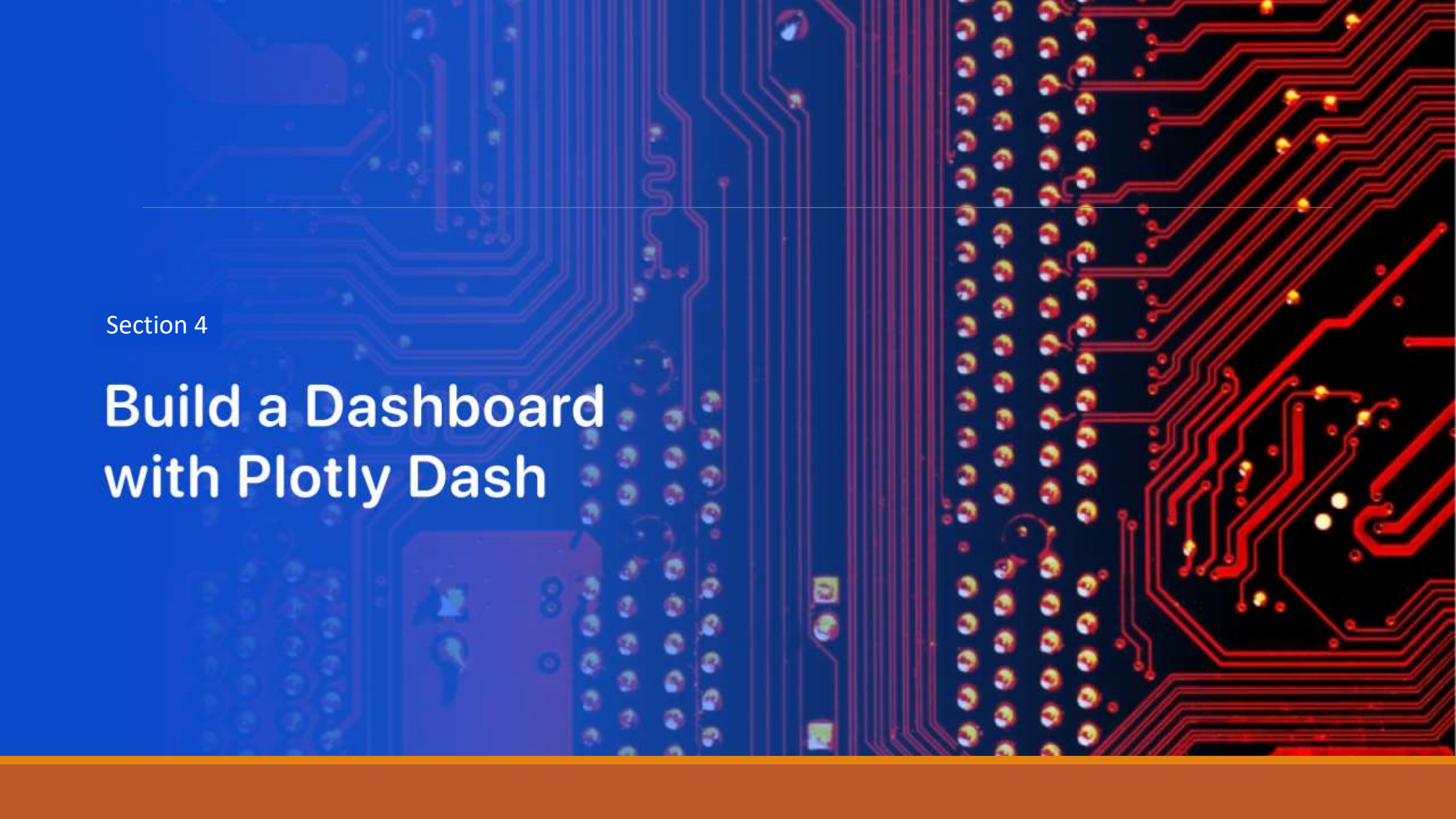


Folium map of launch site locations:

<Folium Map Screenshot 3>

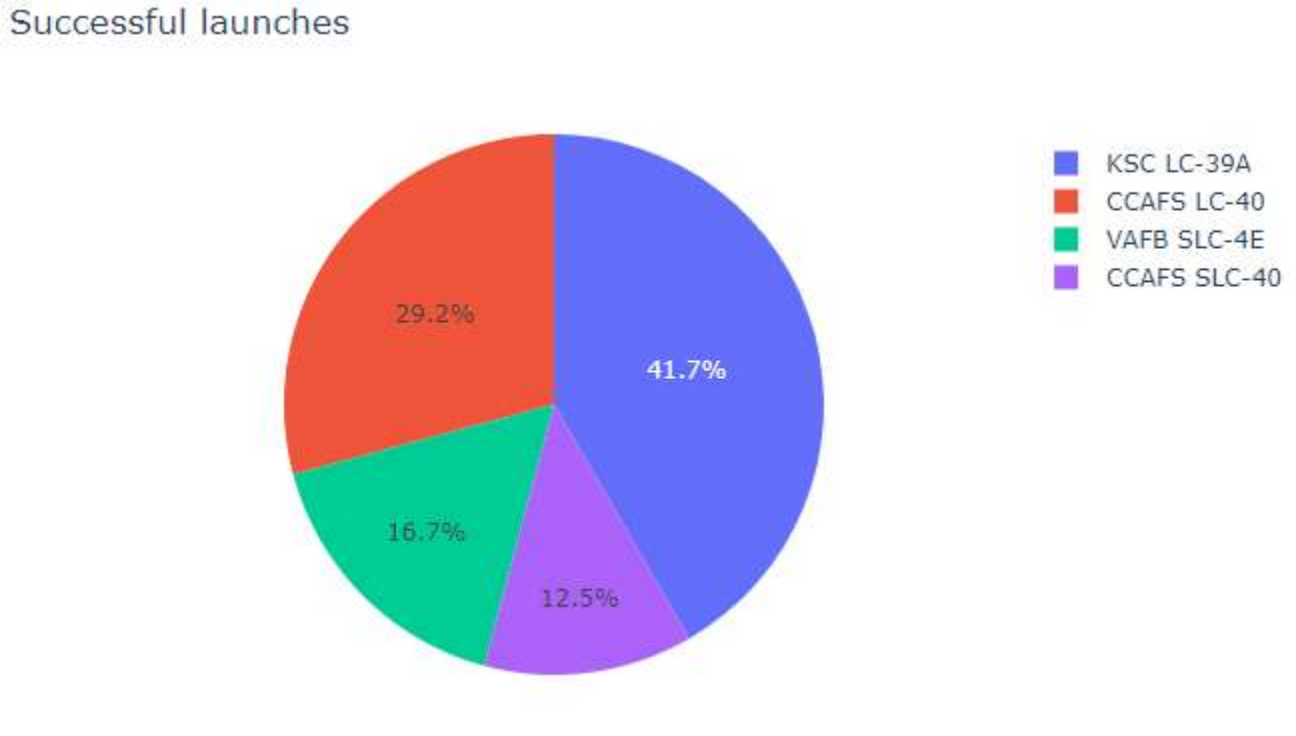
Interactive clusters on Folium maps to display each successful landing (green icon) and failed landing (red icon).

In this example VAFB SLC-4E shows 4 successful landings and 6 failed landings.


Section 4
Build a Dashboard with Plotly Dash

Launches with successful outcomes

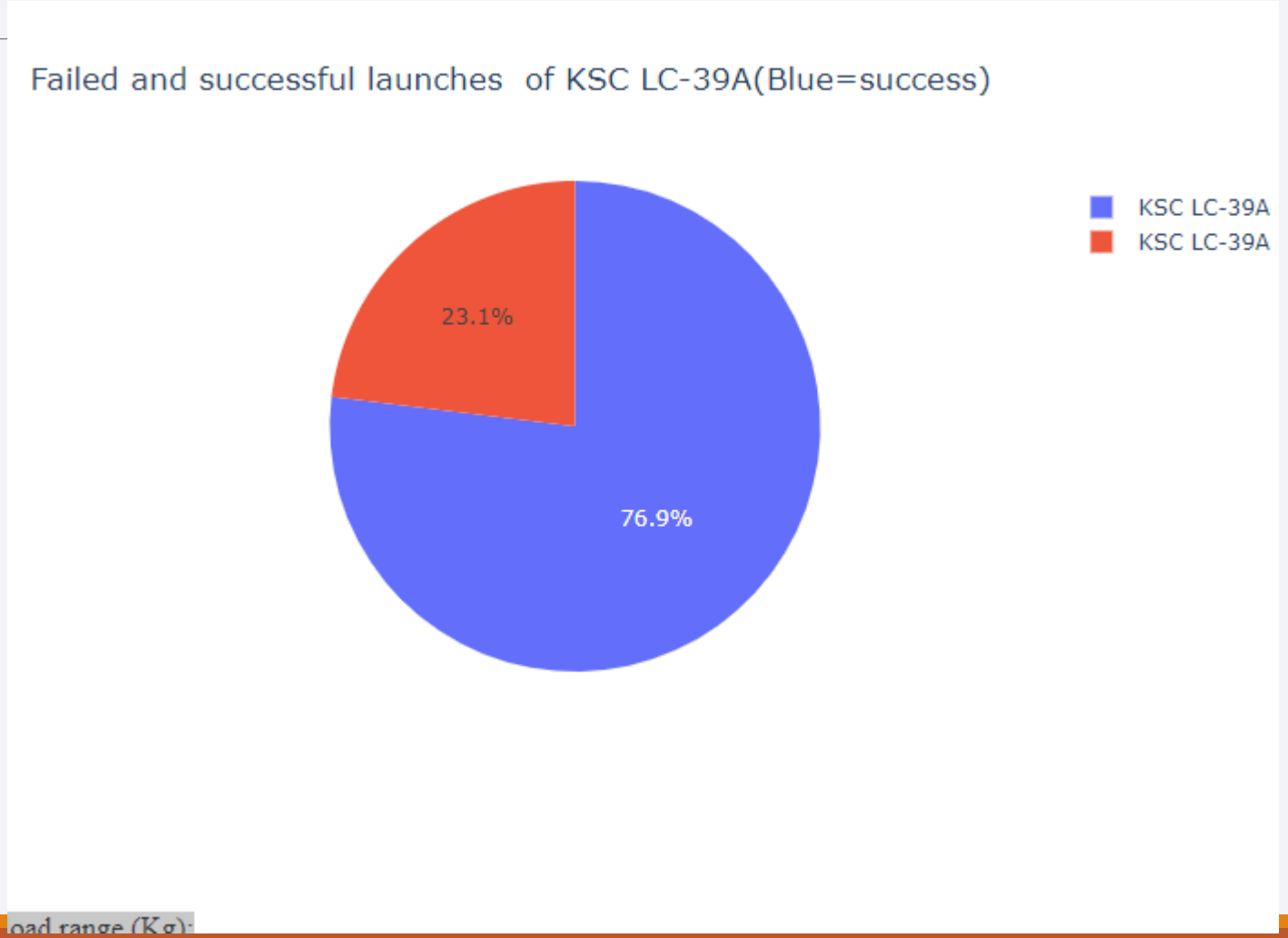
Diatribution of success rate across launch sites



Highest Success Rate Launch Site

Piechart for the launch site KSC LC-39A with highest launch success ratio

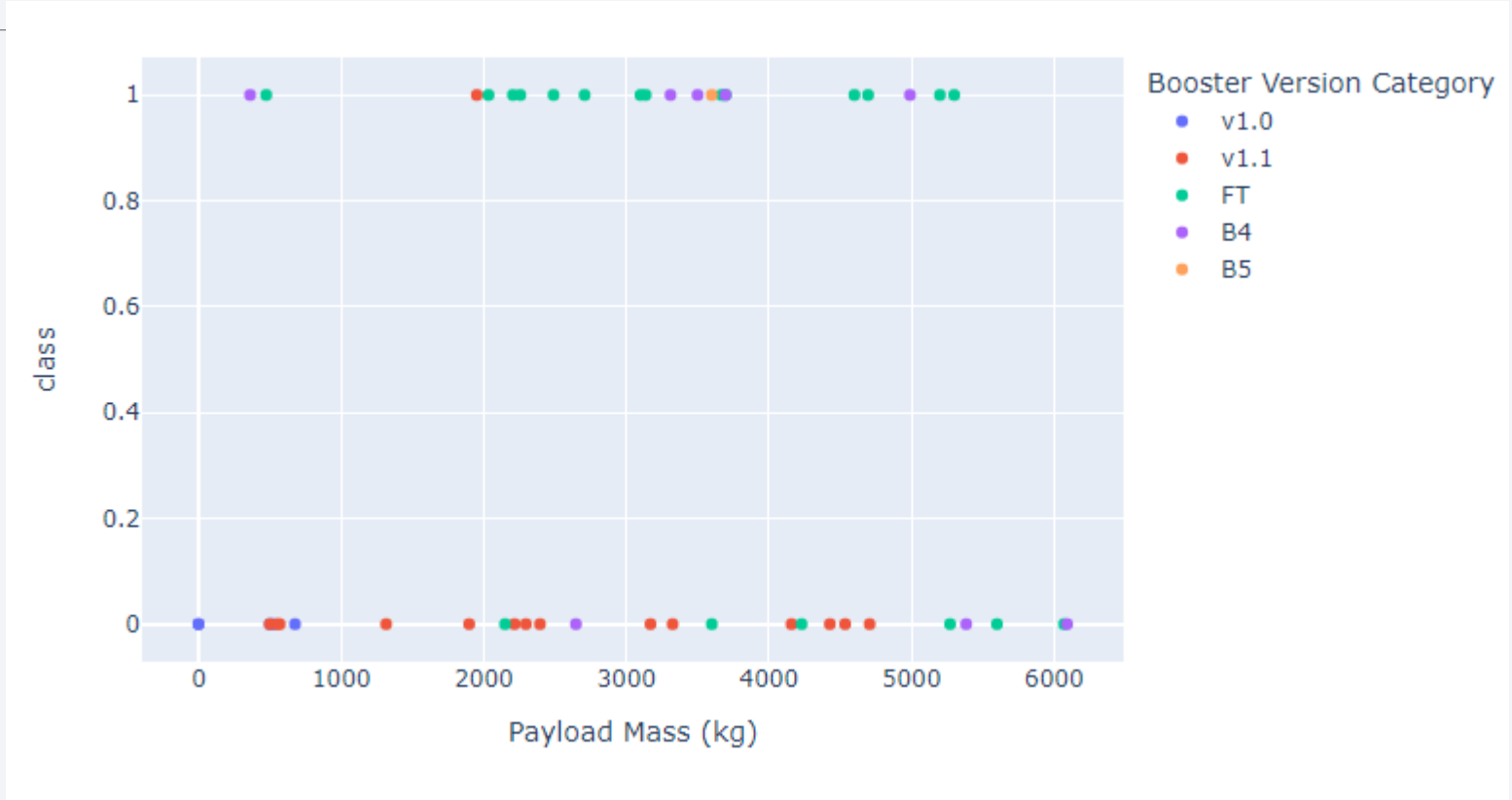
KSC LC-39A has the highest success rate with 10 successful landings and 3 failed landings.



Payload Mass vs. Success vs. Booster Version Category

Dashboard has a Payload range selector set from 0-10000. Class indicates 1 for successful landing and 0 for failure.

Scatter plot shows booster version category and number of launches in point size.



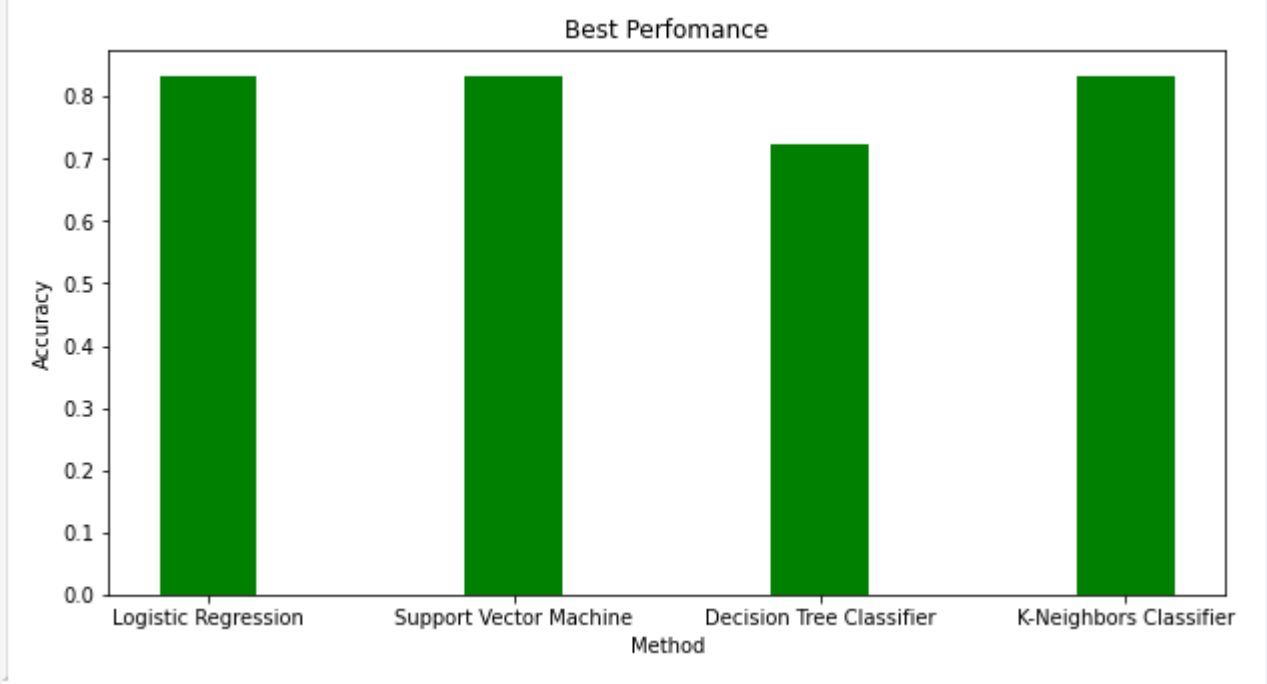
Section 5

Predictive Analysis (Classification)

Classification Accuracy

LR,SVM and KNN had accuracy of about 83%.

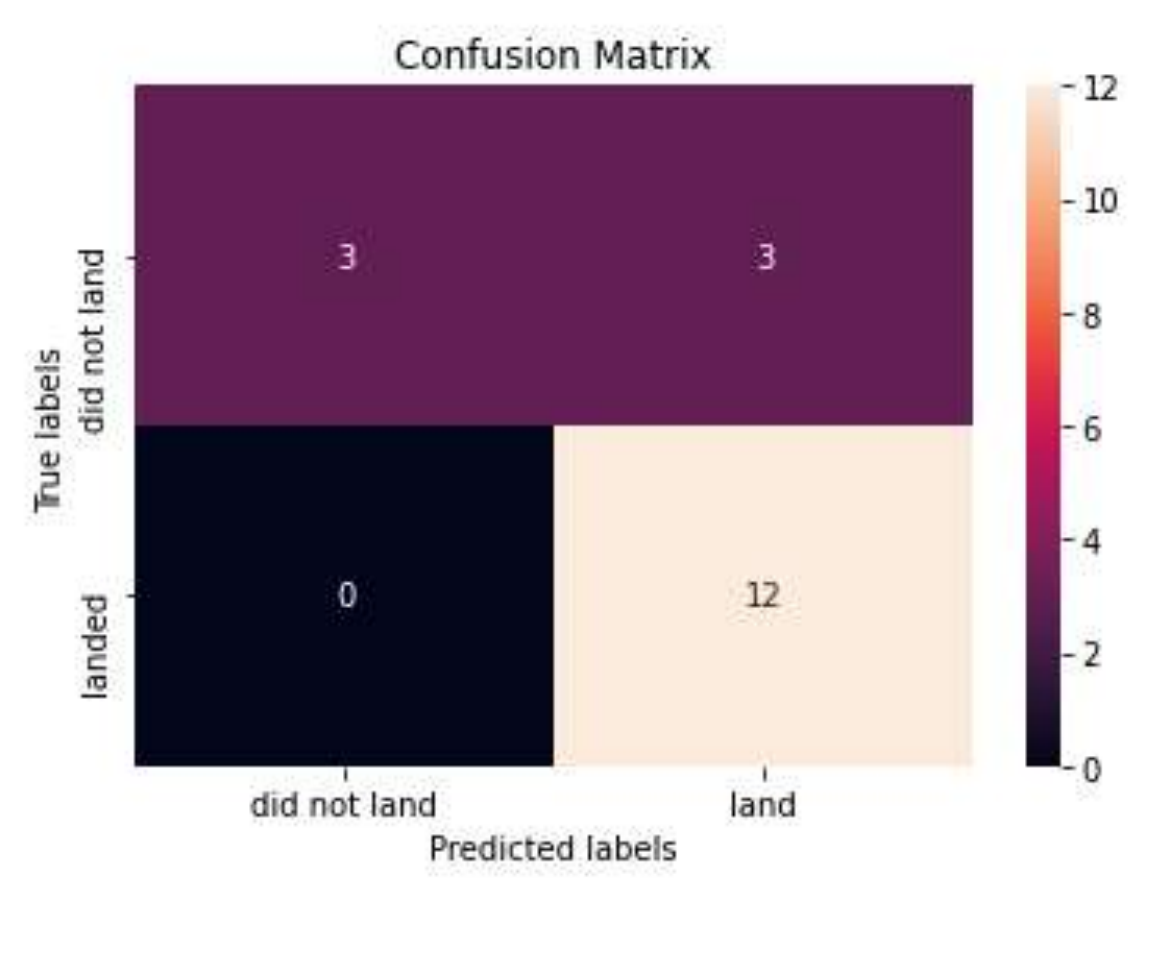
While DTC had an accuracy of 70%



Confusion Matrix

Confusion matrix for KNN classifier

reports accuracy of KNN



Conclusions

Developed a machine learning model for Space Y based on data from SpaceX

To find a model to predict best parameters to give successful landings. Stage 1 will successfully land to save ~$100 million USD

Used data from a public SpaceX API and web scraping SpaceX Wikipedia page

Data labels and stored data into a DB2 SQL database

Adashboard for visualization

machine learning model with an accuracy over 80%

Thank you!