Winning Space Race with Data Science

Jose Aguilera
November 2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection through API
  - Data Collection with Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Data Visualization
  - Interactive Visual Analytics with Folium
  - Machine Learning Prediction
- Summary of all results
  - Exploratory Data Analysis result
  - Interactive analytics in screenshots
  - Predictive Analytics result

Introduction

Project background and context

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 Space X 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 space X for a rocket launch.

This goal of the project is to create a machine learning pipeline to predict if the first stage will land successfully.

Problems you want to find answers

- What factors determine if the rocket will land successfully?
- The interaction amongst various features that determine the success rate of a successful landing.
- What operating conditions needs to be in place to ensure a successful landing program.

Section 1

Methodology

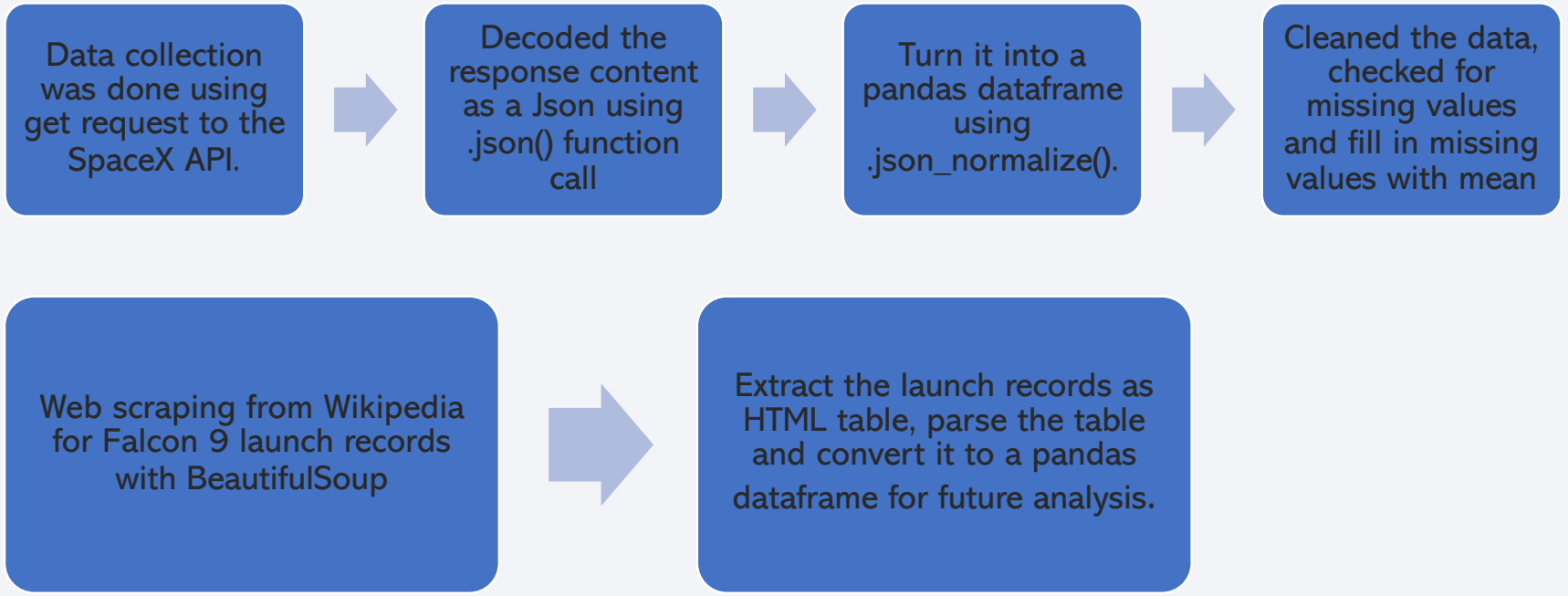
Methodology

Executive Summary

- Step 1: Data collection was Web scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia
- Step 2: Data wrangling where it was created a column with the landing outcome (bad or good)
- Step 3: Exploratory data analysis where it was used SQL to run queries to get insights of the data
- Step 4: Interactive visual analytics where it was used python, folium and dash to create graphical information
- Step 5: Prepare the data to machine learning converting categorical variables into dummy variables to perform machine learning.
- Step 6: Predictive analysis using machine learning to predict the outcome of the launching

Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts.



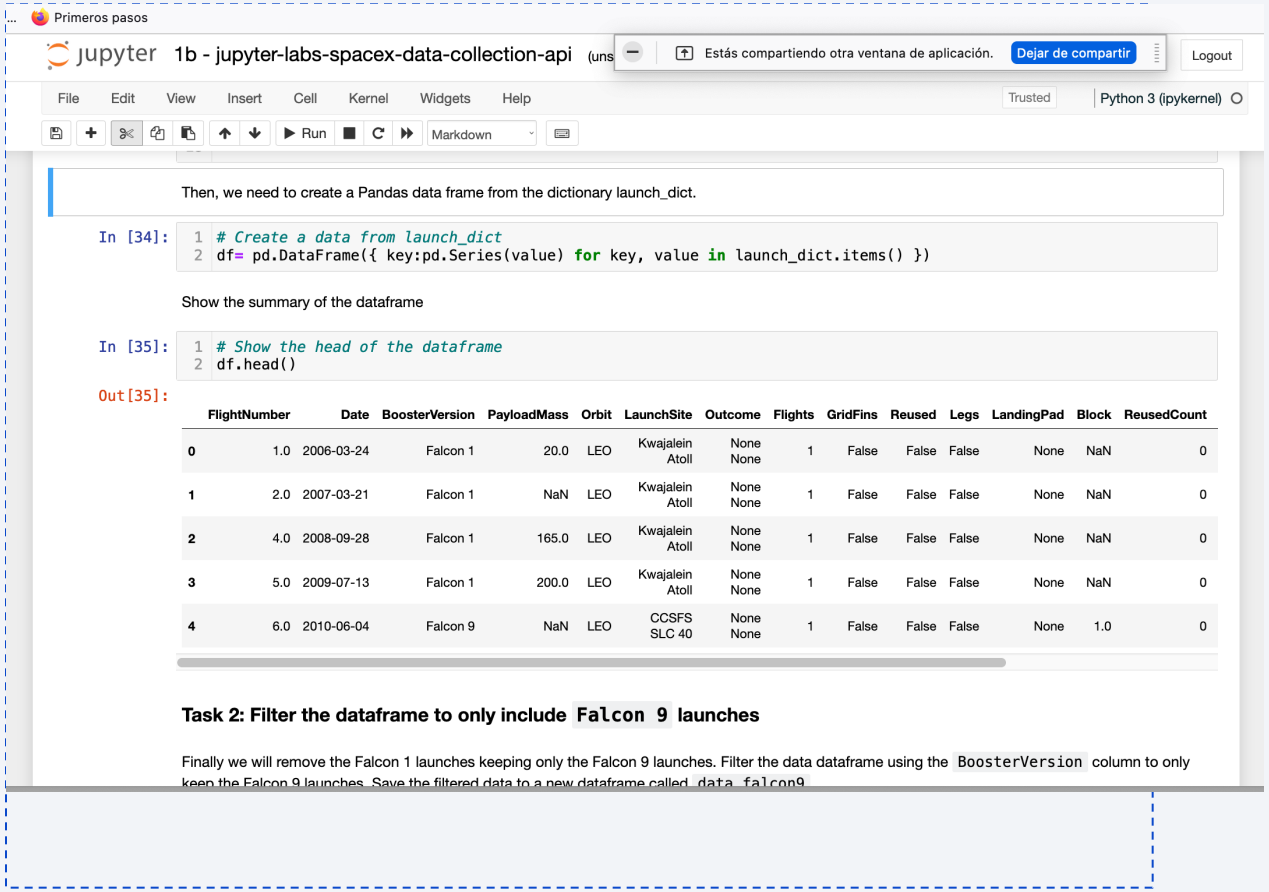
Data Collection – SpaceX API

- **Present your data collection with SpaceX REST calls:**

- Get request to the SpaceX API to collect data
- clean the requested data
- Data wrangling and formatting.

- **Link**

https://github.com/joseaguilera85/IBMDataScience/blob/main/1b%20-%20jupyter-labs-spacex-data-collection-api.ipynb



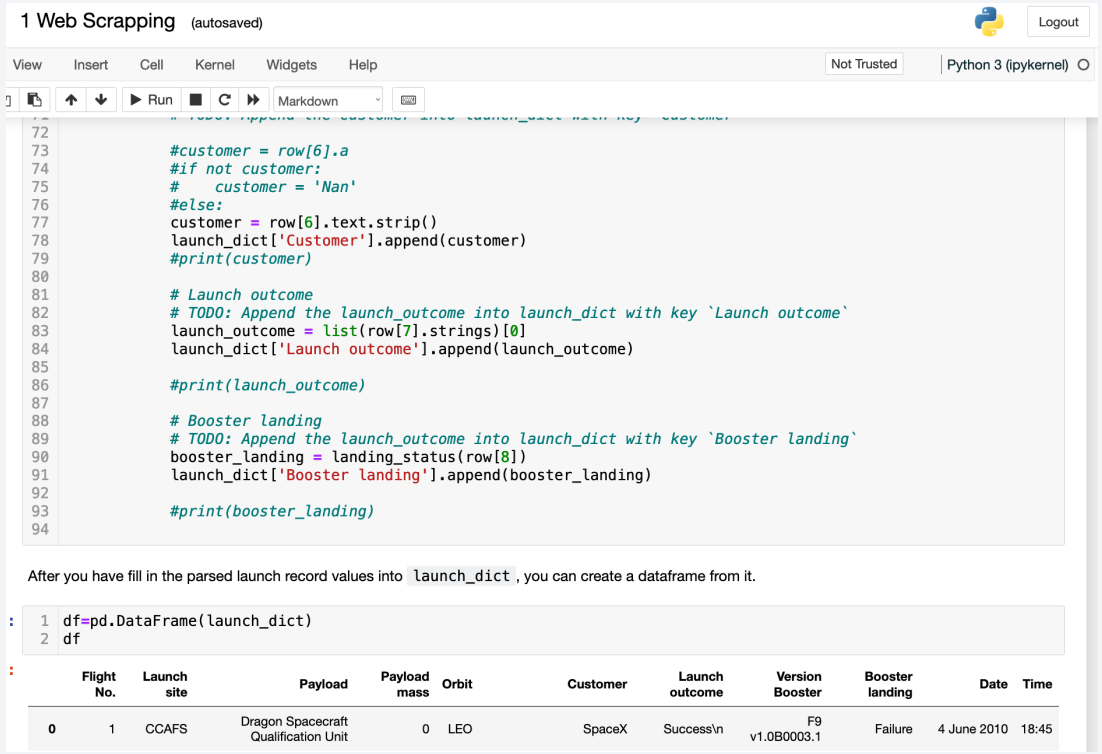
Data Collection - Scraping

- **Present your web scraping process using key phrases and flowcharts**

1. Web scrapping to webscrap Falcon 9 launch records with BeautifulSoup
2. Parsed the table and converted it into a pandas dataframe.

- **The link to the notebook is**

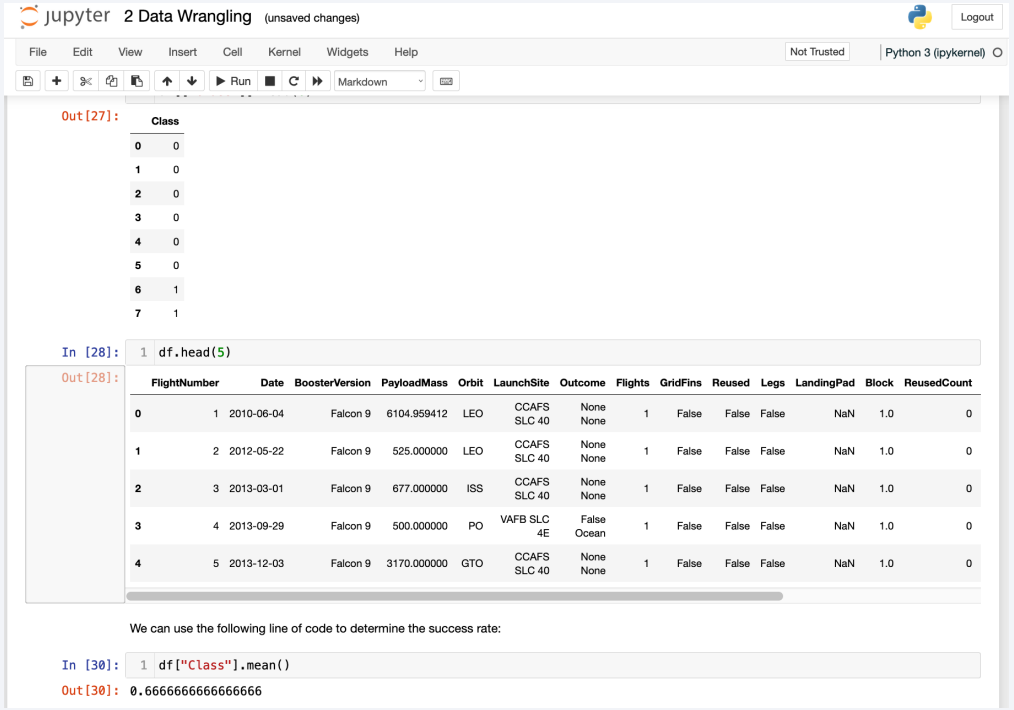
https://github.com/joseaguilera85/IBMDataScience/blob/main/1%20Web%20Scrapping.ipynb



Data Wrangling

Describe how data were processed

- Exploratory data analysis and determined the training labels.
- Calculated the number of launches at each site, and the number and occurrence of each orbits
- Landing outcome label from outcome column and exported the results to csv.

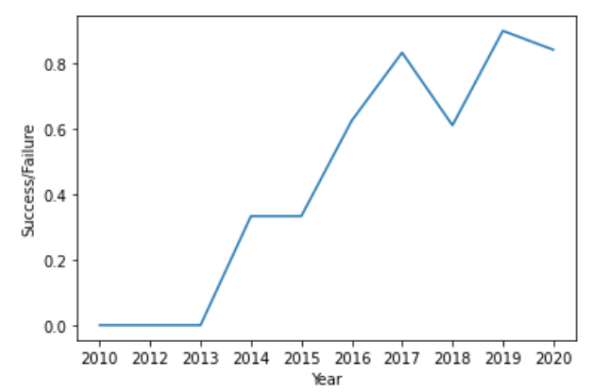


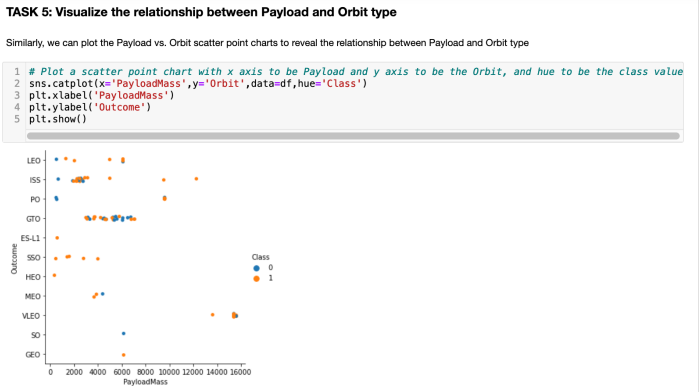
- **Link**

https://github.com/joseaguilera85/IBMDataScience/blob/main/1%20Web%20Scrapping.ipynb

EDA with Data Visualization

- **Summarize what charts were plotted and why you used those charts**
- 1st - The launch success yearly and conclude that the trend is positive
- 2nd – The relation between payload, orbit type and success





- **Link**

https://github.com/joseaguilera85/IBMDataScience/blob/main/3%20Analysis%20using%20pandas%20and%20mathplotlib.ipynb

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- Queries:
  - Display the names of the unique launch sites in the space missionThe total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - The total number of successful and failure mission outcomes
  - List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

- **Link**

https://github.com/joseaguilera85/IBMDataScience/blob/main/4%20Exploratory%20Analysis%20with%20SQL%20lab.ipynb

Build an Interactive Map with Folium

- Marked all launch sites, and added map objects such as markers, circles, lines to mark the success or failure of launches for each site on the folium map.
- Used the color-labeled marker clusters, we identified which launch sites have relatively high success rate.
- Calculated the distances between a launch site to its proximities like the Orlando City

- **Link**

https://github.com/joseaguilera85/IBMDataScience/blob/main/5%20Visual%20Analytics.ipynb

Build a Dashboard with Plotly Dash

Summarize what plots/graphs and interactions you have added to a dashboard / Explain why you added those plots and interactions

- Built an interactive dashboard with Plotly dash
- Plotted pie charts showing the total launches
- Plotted scatter graph showing the relationship of Outcome and Payload Mass (Kg) for the different booster version.

Link

https://github.com/joseaguilera85/IBMDataScience/blob/main/6%20Dashboard%20with%20Plotly%20Dash.ipynb

Predictive Analysis (Classification)

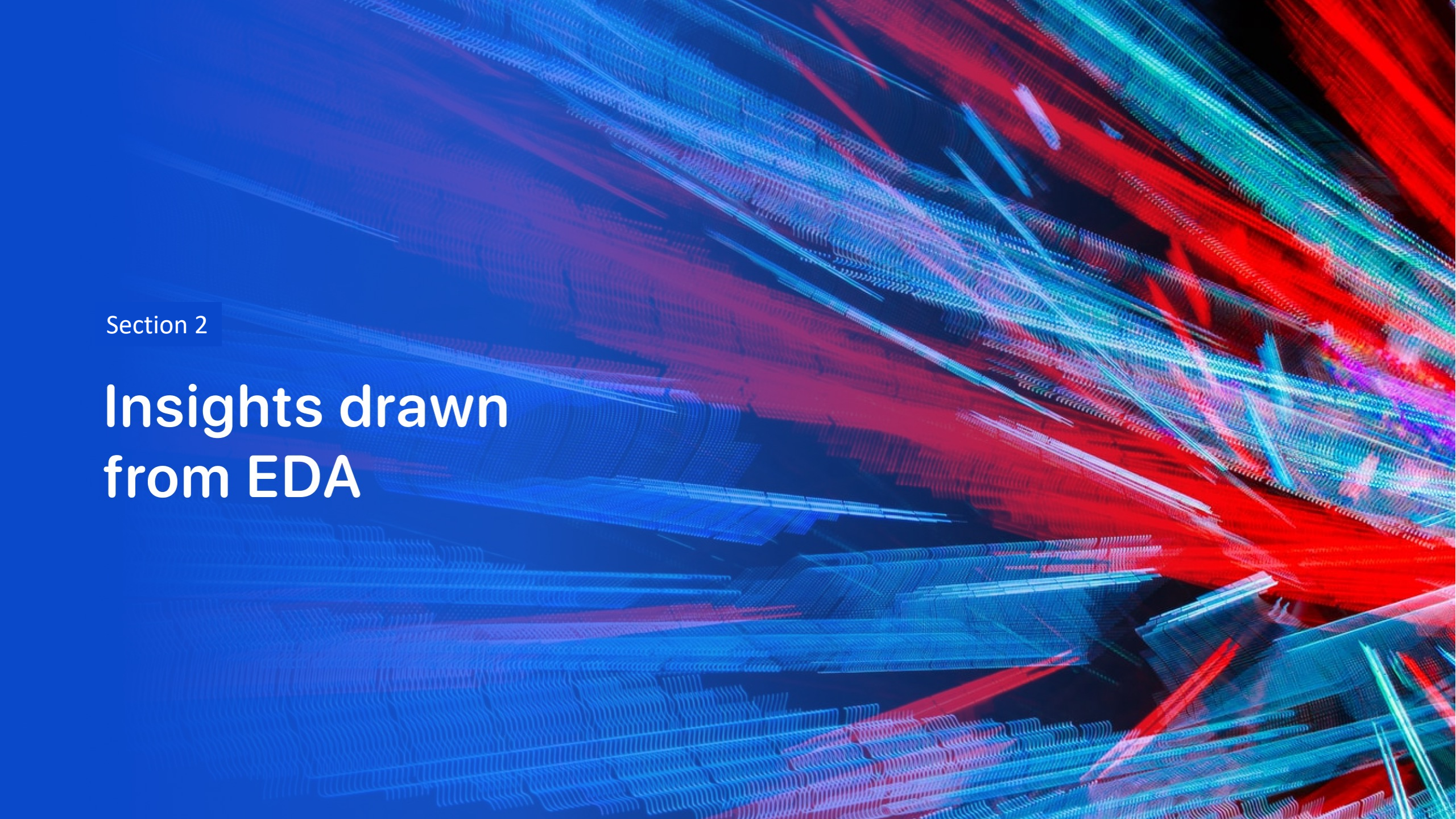
- **Summarize how you built, evaluated, improved, and found the best performing classification model**
  - We loaded the data
  - Transformed the data, split our data into training and testing.
  - Built different machine learning models and tune different hyperparameters using GridSearchCV.
  - Found the best performing classification model.

- **Link**

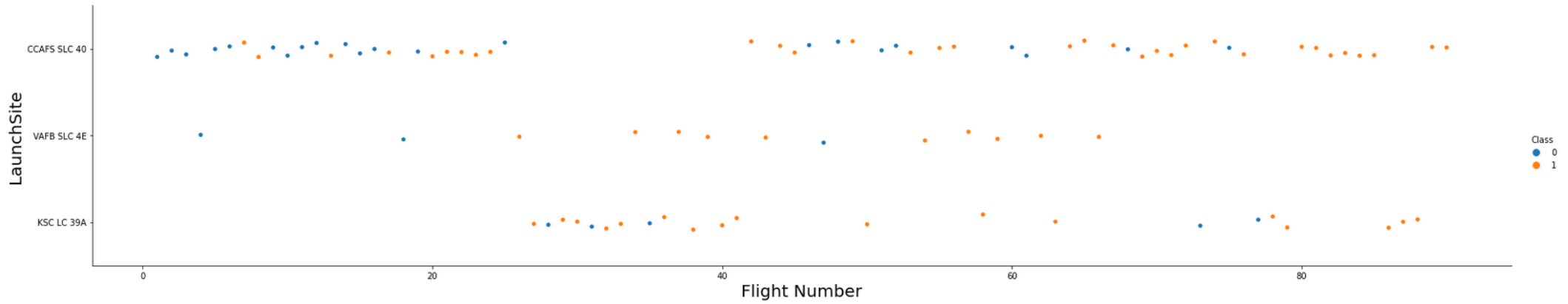
https://github.com/joseaguilera85/IBMDataScience/blob/main/7%20SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

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


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



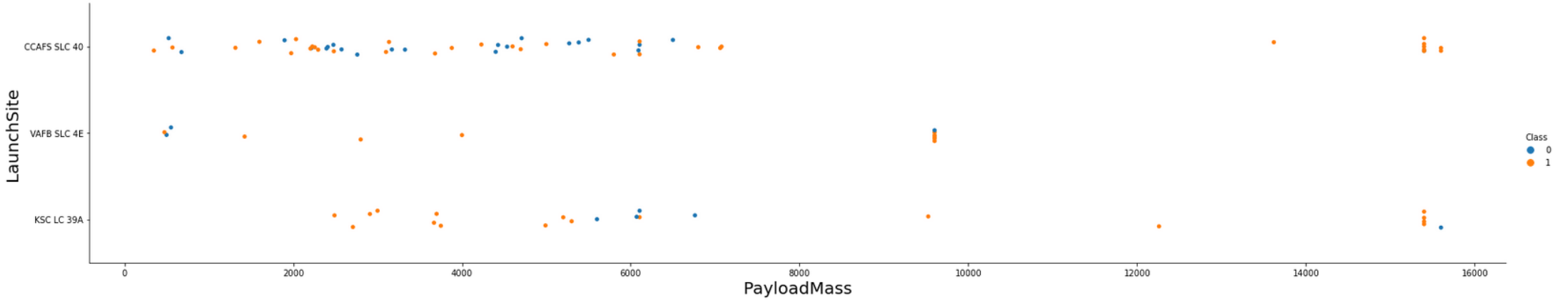
Now try to explain the patterns you found in the Flight Number vs. Launch Site scatter point plots.

1) Most of the succesfull came from the WAFB SLC 4E site

2) Lalety the WAFB SLC 4E site was not used

3) The last 13 launches was succesfull

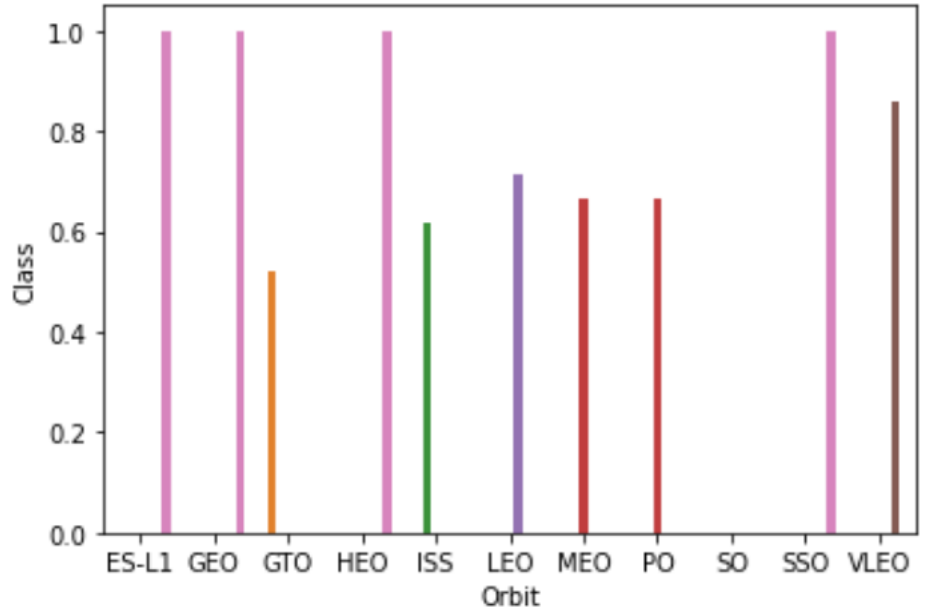
Payload vs. Launch Site



1) For the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

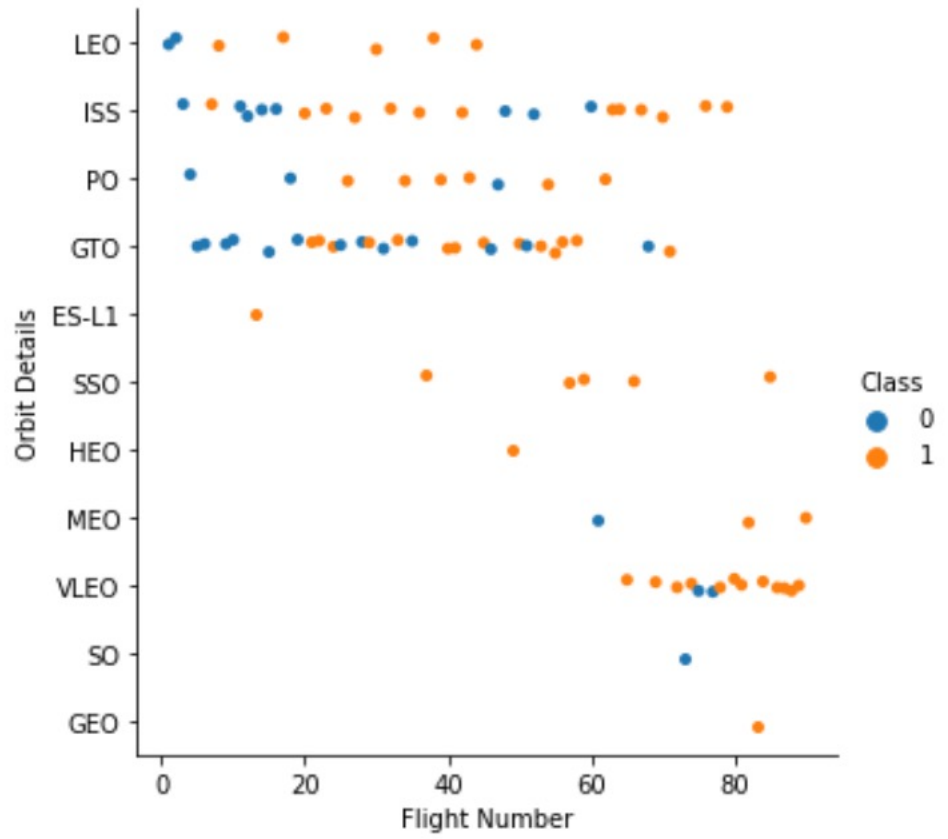
2) For the CCAFS and KSC launched for heavypayload mass(greater than 10000) all but one was succesfull

Success Rate vs. Orbit Type



1) ES-L1, HEO and SS0 has the greatest success rates

Flight Number vs. Orbit Type

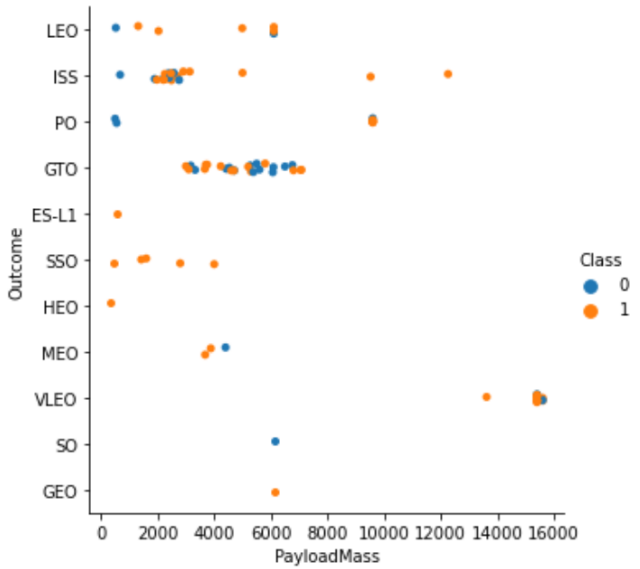


1) VLEO orbit the Success appears related to the number of flights

2) VLEO has most of the lately flights

3) There seems to be no relationship between flight number when in GTO orbit.

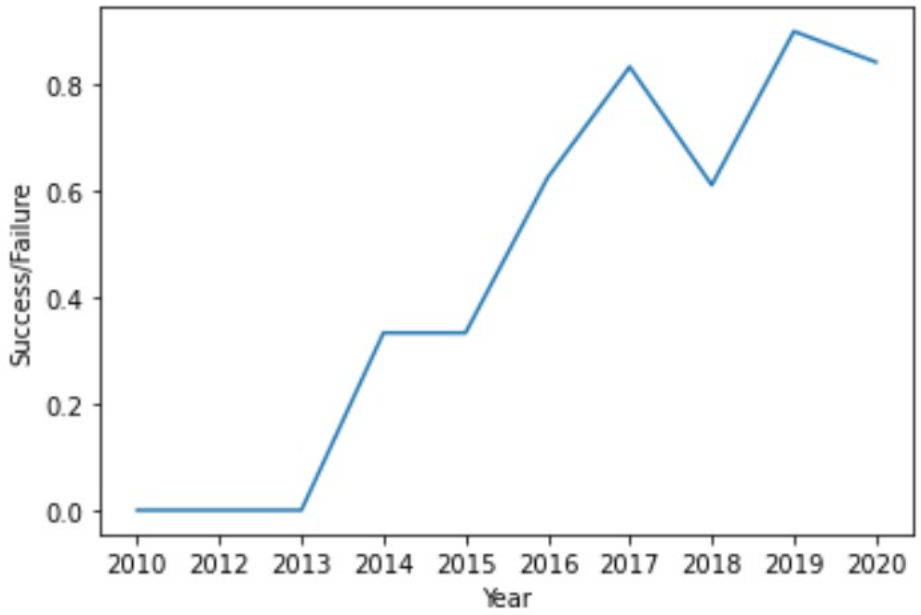
Payload vs. Orbit Type



1) With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

2) However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend



1) The sucess rate since 2013 kept increasing till 2020

2) From 2010 to 2013 the success rate was 0

All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here

There are only 4 launch sites

Task 1

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

```
%%sql
SELECT DISTINCT LAUNCH_SITE from SPACEX
```

```
 * ibm_db_sa://smt91818:***@19af6446-6171-4641-8aba-9dcff8e1b6ff.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30699/BLUDB
Done.
```

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Present your query result with a short explanation here

Here are the first 5 records of the lauch sites that begins with CCA

Task 2

Display 5 records where launch sites begin with the string 'CCA'

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

```
 * ibm_db_sa://smt91818:***@19af6446-6171-4641-8aba-9dcff8e1b6ff.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30699/BLUDB
Done.
```

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landingoutcome
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

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here

The total payload carried by boosters was 45596 kg

Task 3

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

```
1 %sql select sum(PAYLOAD_MASS__KG_) as payloadmass from SPACEX where (CUSTOMER) = 'NASA (CRS)';
```

```
 * ibm_db_sa://smt91818:***@19af6446-6171-4641-8aba-9dcff8e1b6ff.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:306
99/BLUDB
Done.
```

payloadmass
45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here

The average payload mass carried was 2928

Task 4

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

```
1 %sql select avg(PAYLOAD_MASS__KG_) as avgloadmass from SPACEX where (booster_version) = 'F9 v1.1';
```

```
 * ibm_db_sa://smt91818:***@19af6446-6171-4641-8aba-9dcff8e1b6ff.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30699/BLUDB
Done.
```

avgloadmass
2928

First Successful Ground Landing Date

- **Find the dates of the first successful landing outcome on ground pad**
- **Present your query result with a short explanation here**

The first successful landing outcome on ground pad was on 2010-06-04

Task 5

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

Hint:Use min function

```
%sql SELECT min(DATE) from SPACEX where (MISSION_OUTCOME) LIKE 'Success';
```

```
 * ibm_db_sa://smt91818:***@19af6446-6171-4641-8aba-9dcff8e1b6ff.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:30699/BLUDB
Done.
```

1
2010-06-04

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**
- **Present your query result with a short explanation here**
- There are 4 successful Drone Shop Landing with Payload between 4000 and 6000

Task 6

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

```
1 %sql select BOOSTER_VERSION from SPACEX where LandingOutcome ='Success (drone ship)' and PAYLOAD_MASS__KG_ BETWE
```

```
 * ibm_db_sa://smt91818:***@19af6446-6171-4641-8aba-9dcff8e1b6ff.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:306
99/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

- Calculate the total number of successful and failure mission outcomes
- Present your query result with a short explanation here

There are 99 successful missions

Task 7

List the total number of successful and failure mission outcomes

```
1 %sql select mission_outcome, count(mission_outcome) from SPACEX GROUP BY mission_outcome;
```

```
 * ibm_db_sa://smt91818:***@19af6446-6171-4641-8aba-9dcff8e1b6ff.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:306
99/BLUDB
Done.
```

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here

2015 Launch Records

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

- Present your query result with a short explanation here

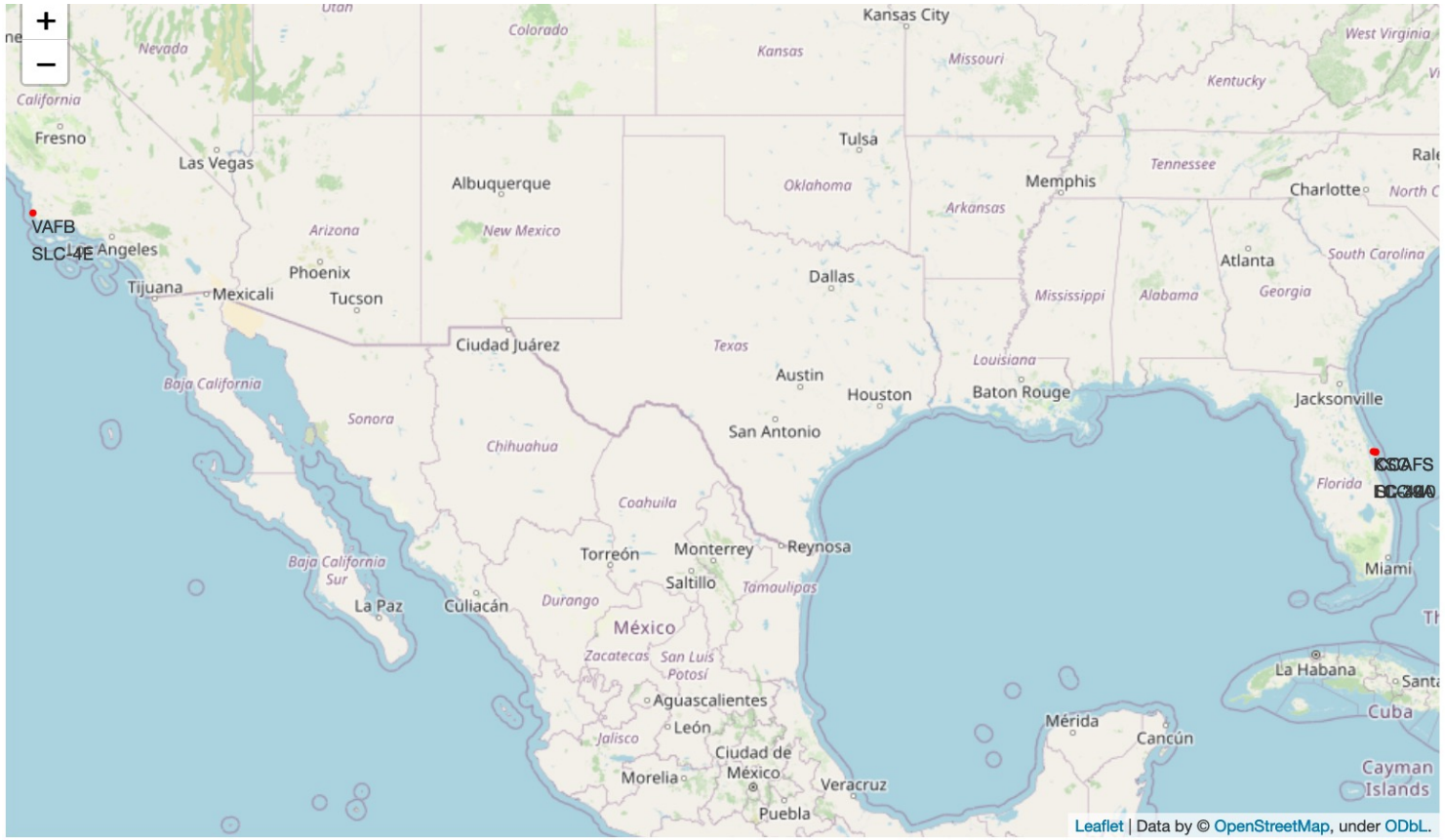
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- 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

- Present your query result with a short explanation here


Section 3
Launch Sites
Proximities Analysis

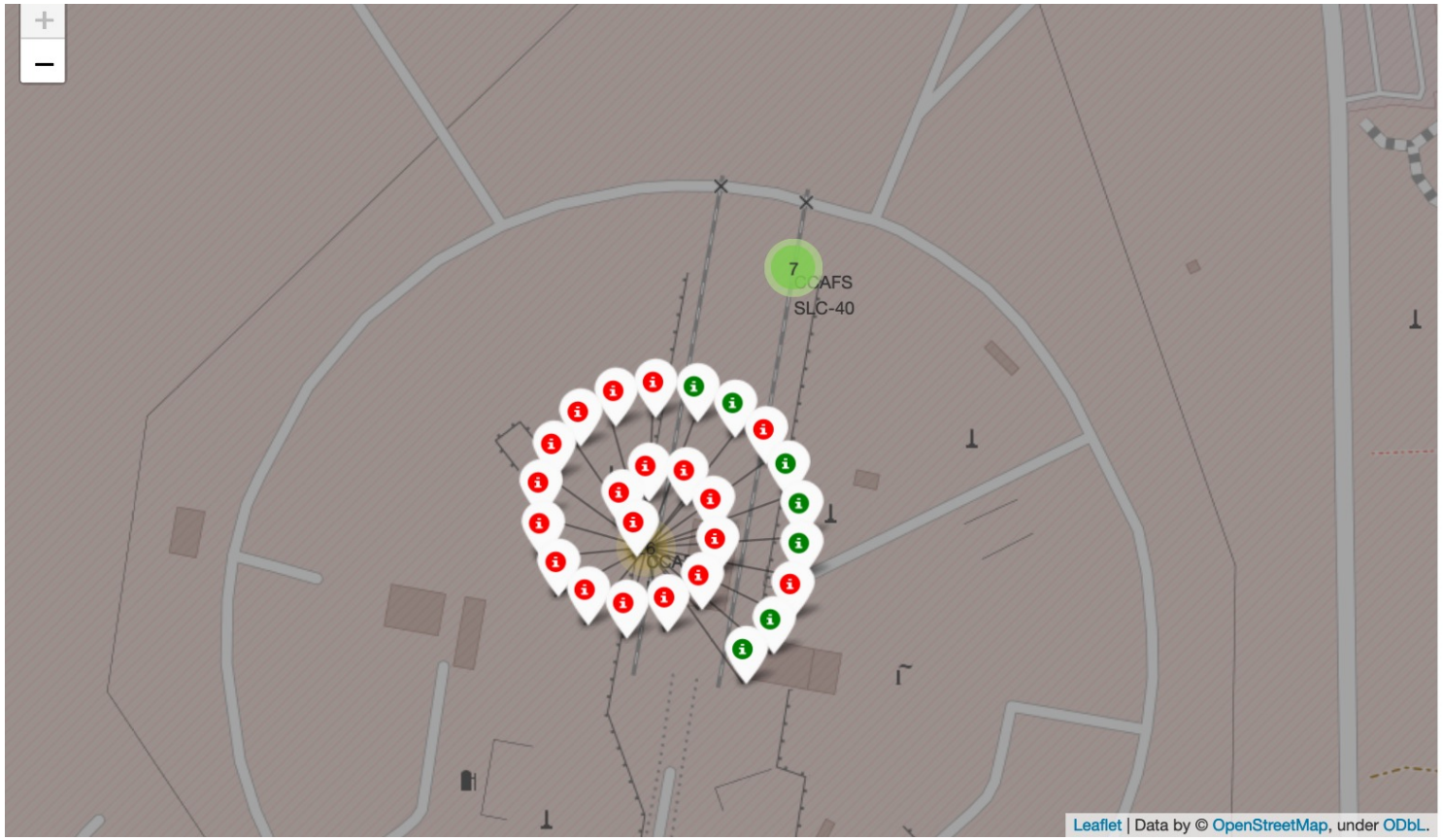
All launch sites on a map



- Explain the important elements and findings on the screenshot

1. There are 3 launch sites in Florida and 1 in California

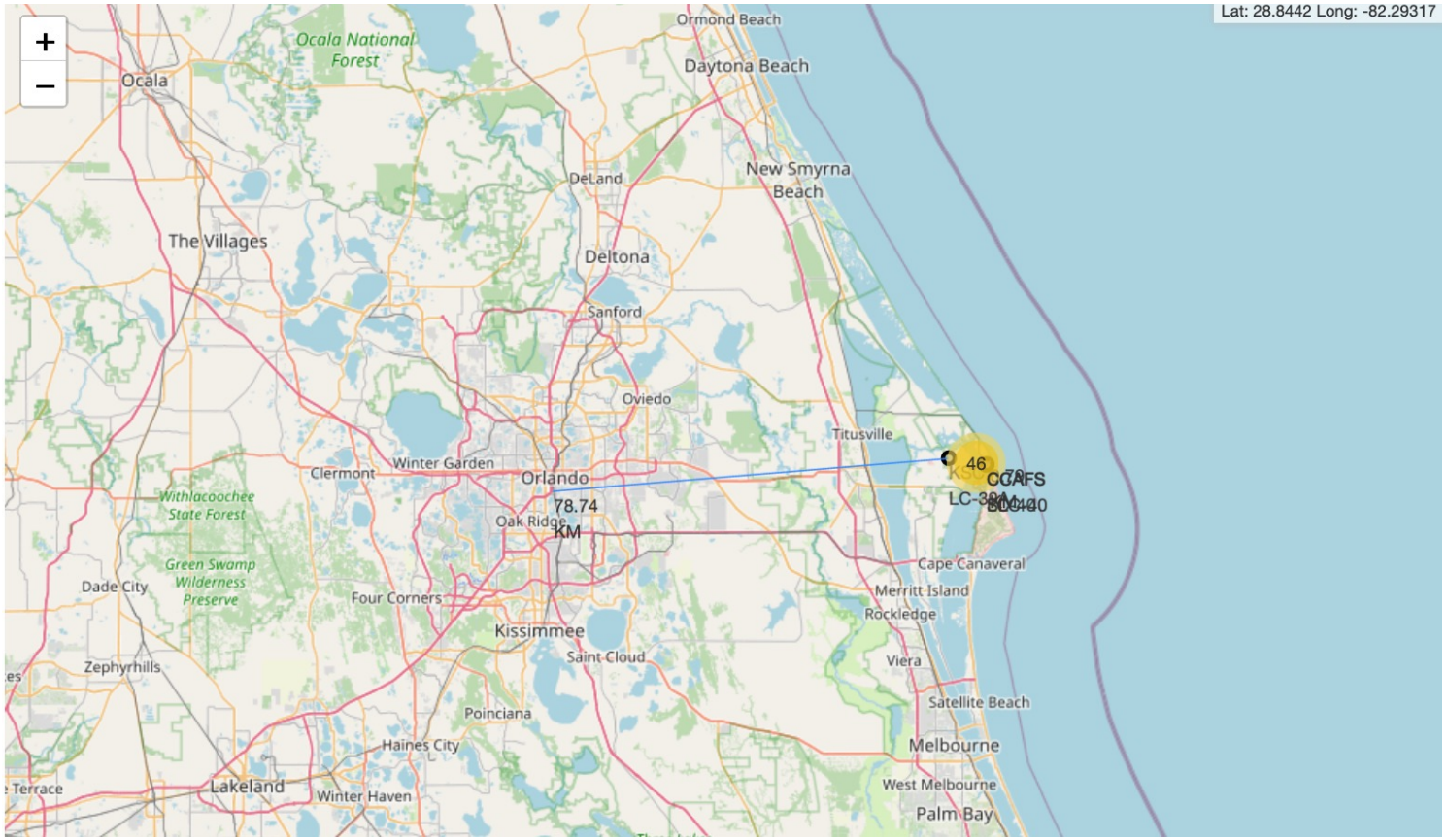
Success/failed launches for each site on the map



- Explain the important elements and findings on the screenshot

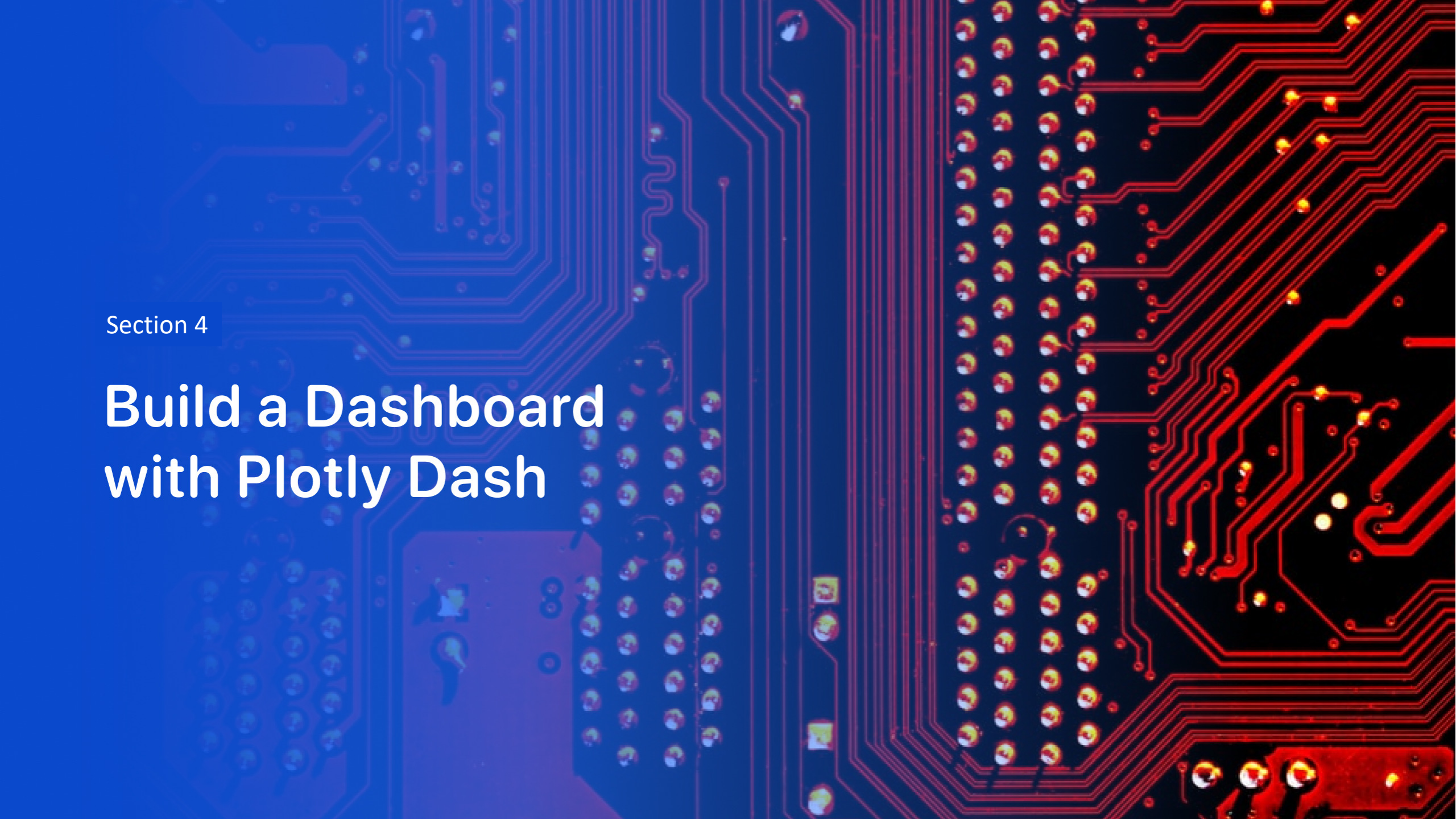
1. The launching site with most successful rate is KSC LC-39A with 10 of 13
2. The launching site with most launches if CCAFS LC-40

Distances between a launch site to Orlando



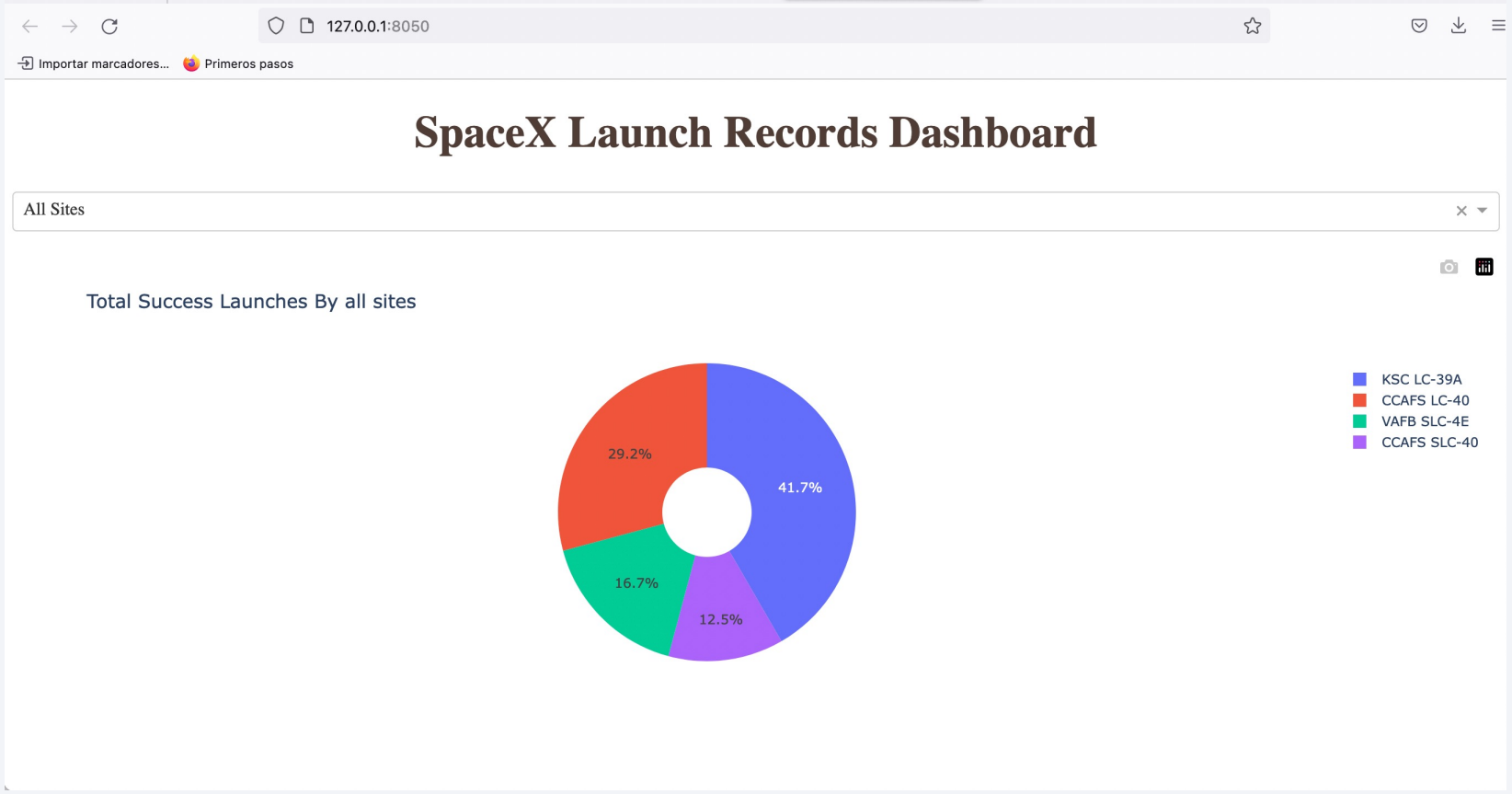
- Explain the important elements and findings on the screenshot

1. The most import city near the launching site is Orlando located 78.7km away from the launching site


Section 4
Build a Dashboard with Plotly Dash

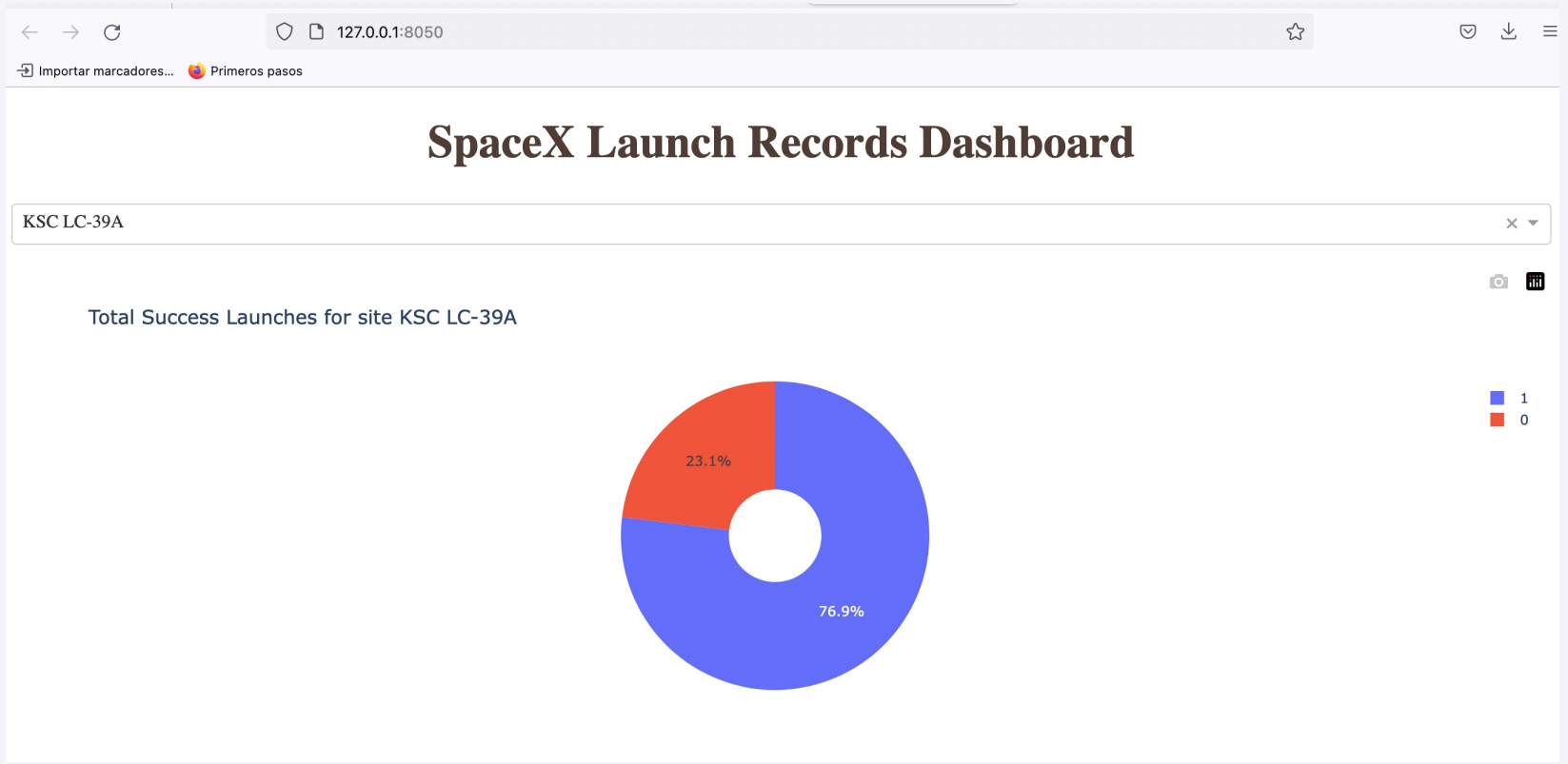
Success percentage achieved by each launch site

The KSC LC-39A has more success launches by all sites



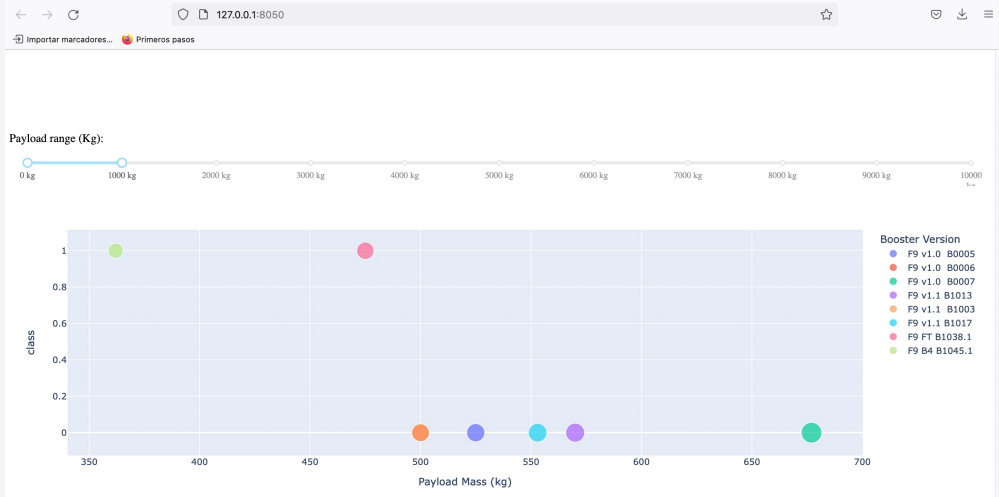
Launch site with the highest launch success ratio

- KSC LC-39A achieved a 76.9% success rate



Payload vs Launch Outcome for all sites, with different payload selected in the range slider

- Graph 1: most of the launch outcome was failed (range 0 to 1000)
- Graph 2: most of the launch outcome was failed (range 5000 to 9000)

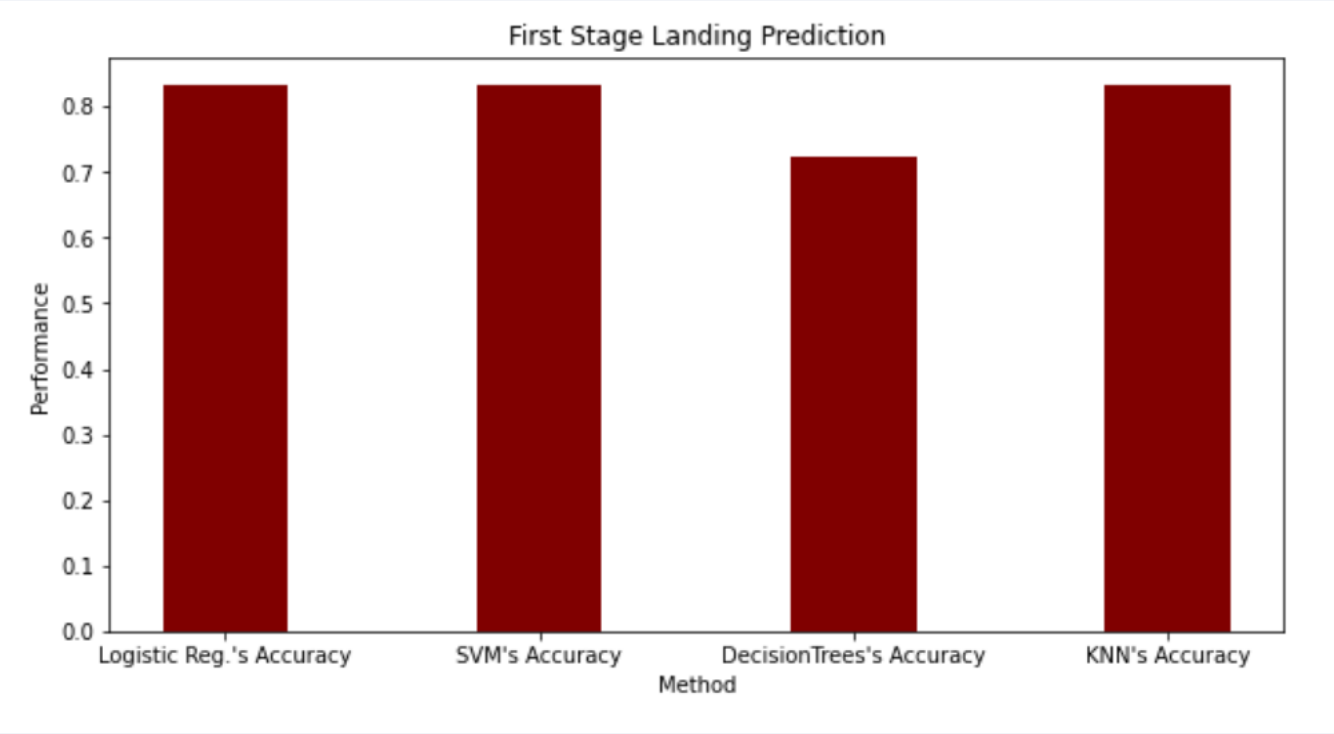


Section 5

Predictive Analysis (Classification)

Classification Accuracy

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

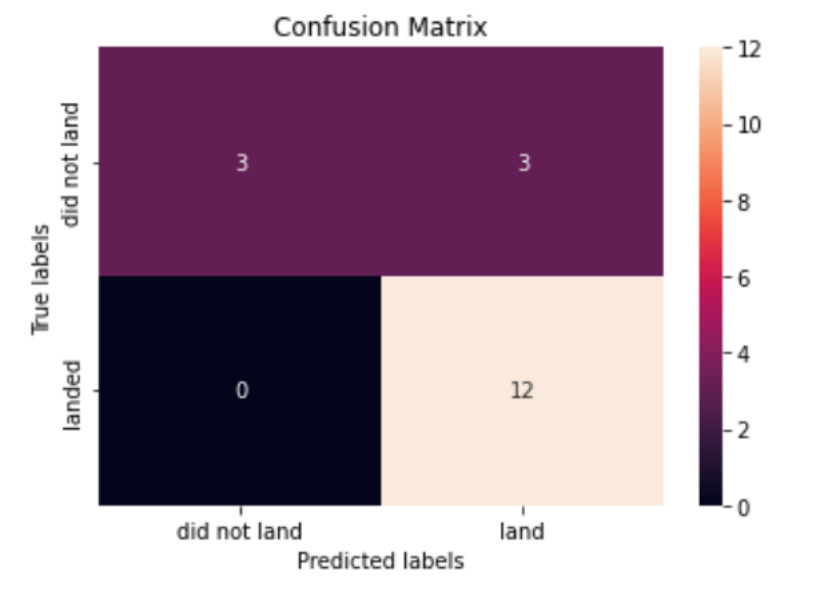


- Find which model has the highest classification accuracy

Logistic Reg.'s Accuracy: 0.8333333333333334
SVM's Accuracy: 0.8333333333333334
DecisionTrees's Accuracy: 0.7222222222222222
KNN's Accuracy: 0.8333333333333334

Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation → KNN Model



Conclusions

- Launch success rate started to increase in 2013 till 2020.
- KSC LC-39A had the most successful launches of any sites.
- Orbits ES-L1, GEO, HEO, SSO, VLEO had the most success rate.
- KNN Model is the best machine learning algorithm for this task.

Thank you!