IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Vu Minh Duy
2-Feb-2022

Outline

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

Executive Summary

- Summary of methodologies
  - ✓ Data sets were collected from SpaceX API and Wikipedia web-scraping
  - ✓ EDA was done with visualization and SQL
  - ✓ Interactive visualization was done with Folium and Plotly Dash
  - ✓ Construct and evaluate 4 models for predictive analysis, including logistic regression, decision tree, support vector machine and k-nearest neighbour
- Summary of all results
  - ✓ Overall, success rate of rocket launch increases since 2013 till 2020
  - ✓ All launch sites locate near coastal lines; among them, the best launch site was KSC LC-39A, which had the highest success count
  - ✓ Important predictors include payload mass and orbit type
  - ✓ Predictive analyses have been done, which show all classification models have similar accuracy, with accuracy score >80%

Introduction

- In this capstone, we will model the prediction of Falcon 9 first stage landing success possibility
- Based on the landing outcome prediction, the cost of a launch can be determined. This may give bidder against SpaceX and edge in the competition.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

- The data sets were collected via SpaceX API and Wikipedia web scraping using BeautifulSoup package.

✓ **SpaceX API**:

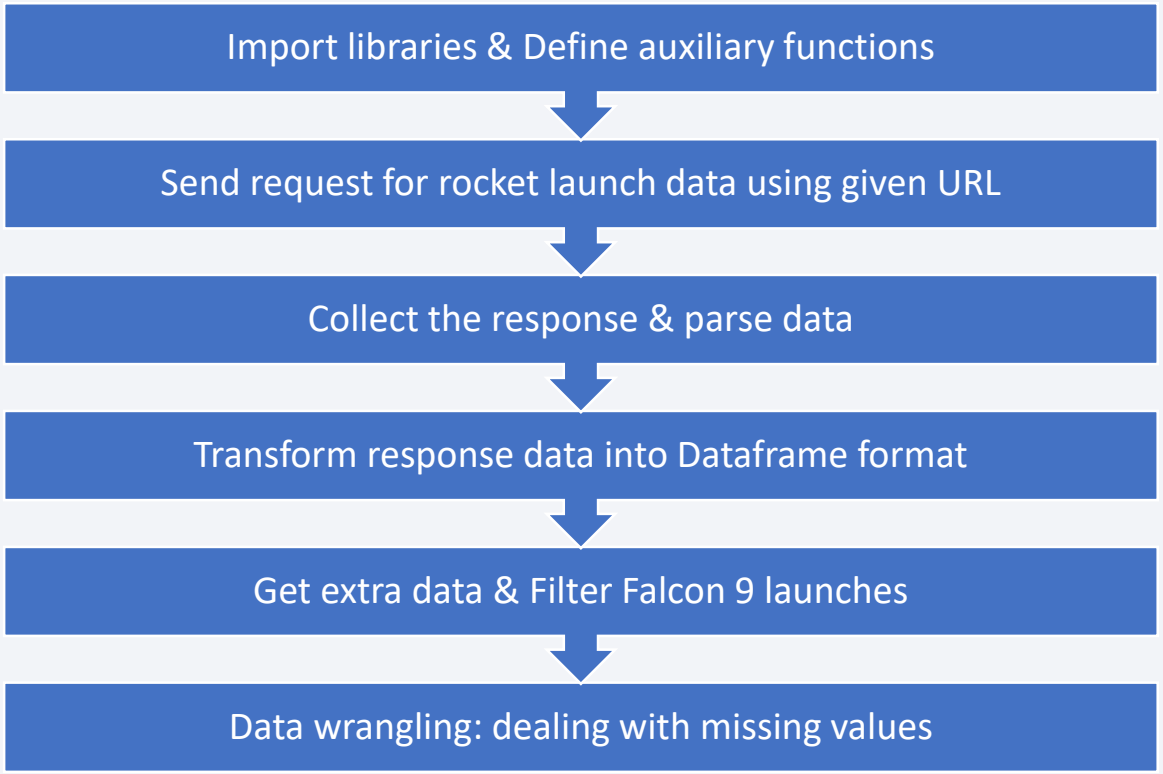
✓ **Wikipedia**:

Data Collection – SpaceX API

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

- GitHub URL of the completed SpaceX API calls notebook as an external reference and peer-review purpose:

https://github.com/victor-vu/IBM-Watson-Studio/blob/817ad5981c4fd383486738045ebc525e594cdd28/Week%201%20-%20API%20Data%20Collection.ipynb

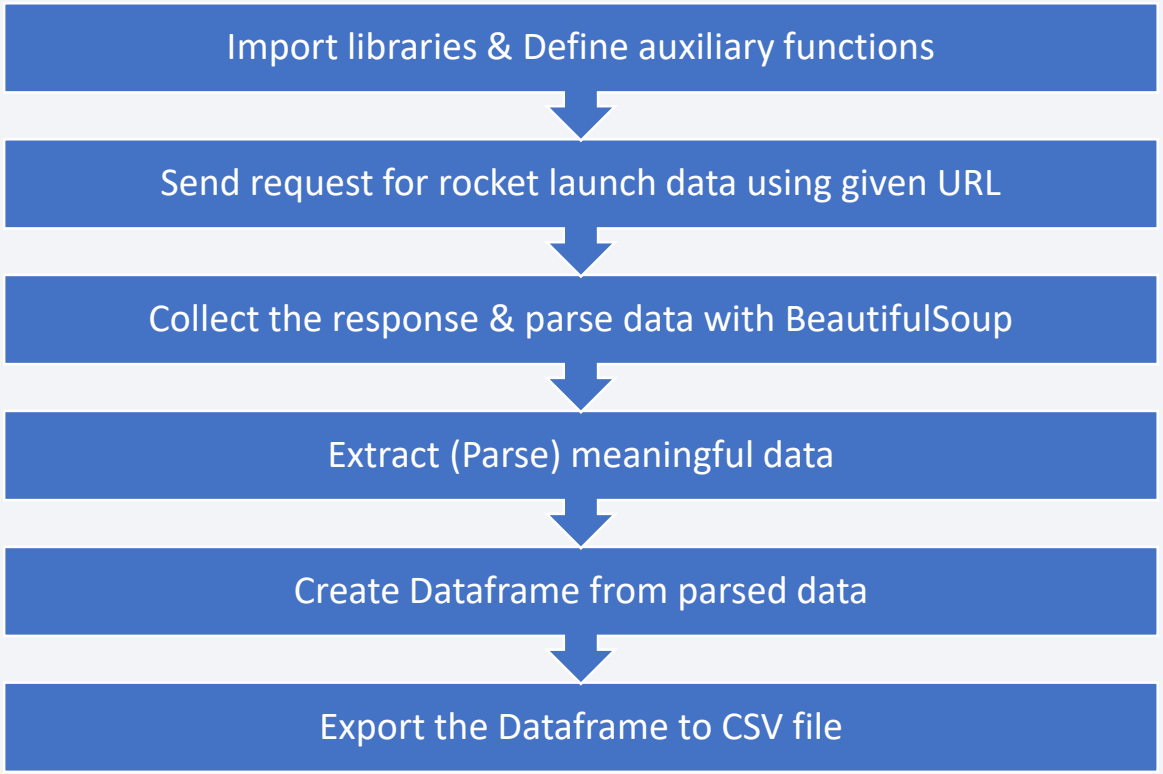


Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose:

https://github.com/victor-vu/IBM-Watson-Studio/blob/817ad5981c4fd383486738045ebc525e594cdd28/Week%201%20-%20Data%20Collection%20from%20Web%20Scraping.ipynb

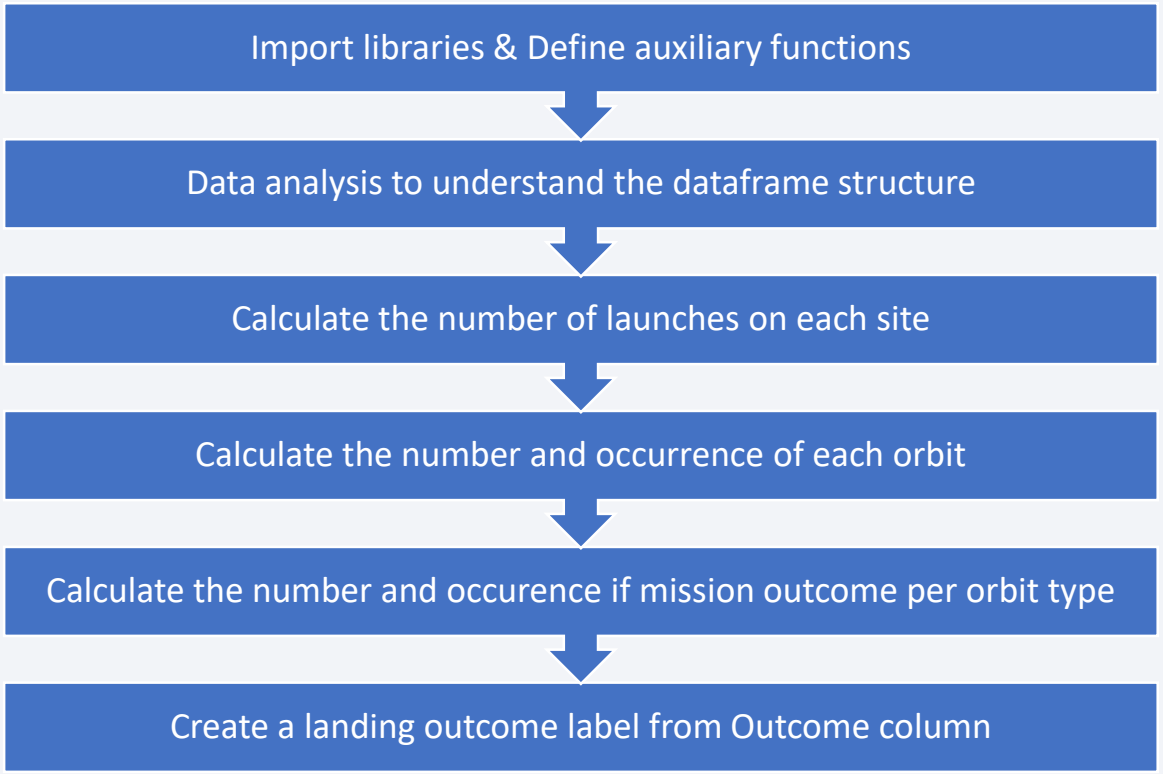


Data Wrangling

- Present your data wrangling process using key phrases and flowcharts

- GitHub URL of completed data wrangling related notebooks, as an external reference and peer-review purpose:

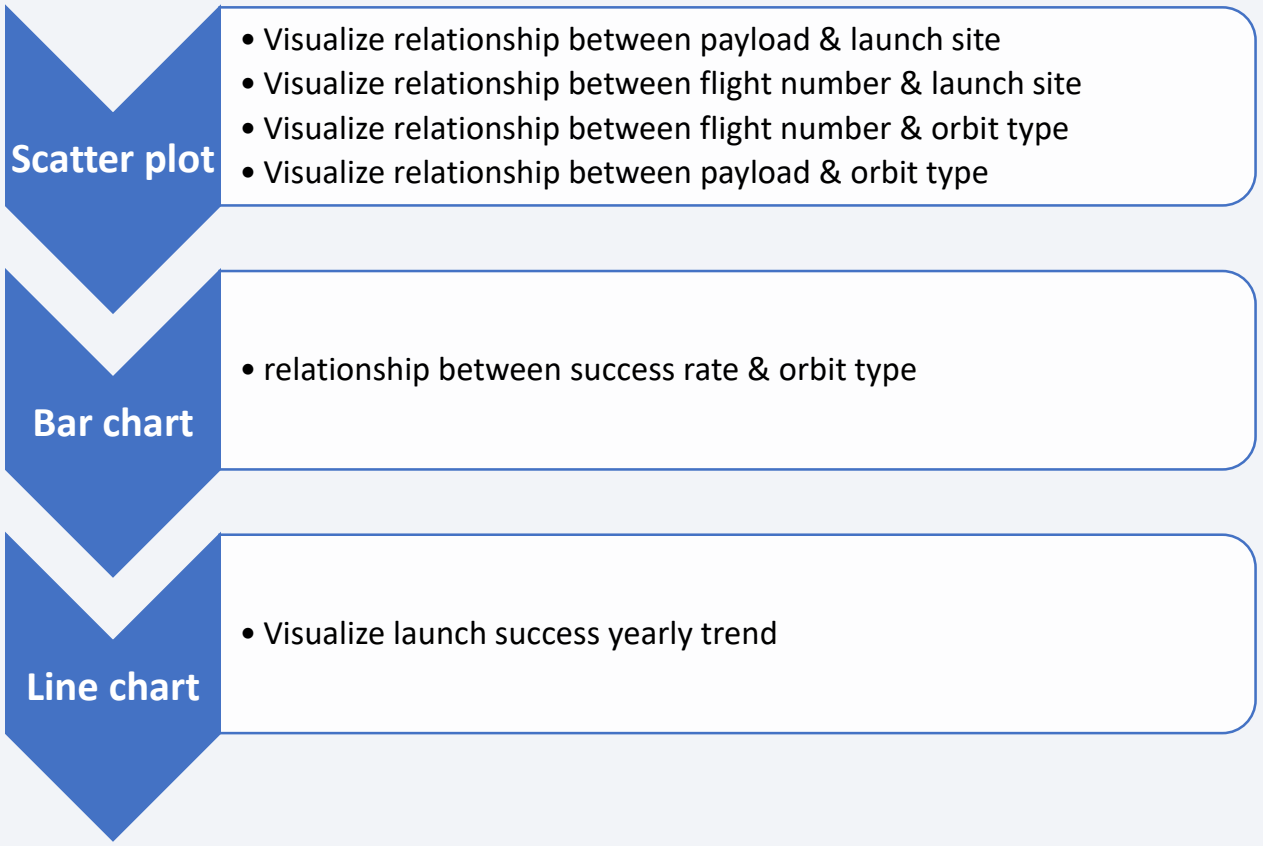
https://github.com/victor-vu/IBM-Watson-Studio/blob/817ad5981c4fd383486738045ebc525e594cdd28/Week%201%20-%20Data%20Wrangling%20&%20EDA.ipynb



EDA with Data Visualization

- Summarize charts were plotted
- GitHub URL of completed EDA with data visualization notebook, as an external reference and peer-review purpose:

https://github.com/victor-vu/IBM-Watson-Studio/blob/817ad5981c4fd383486738045ebc525e594cdd28/Week%202%20-%20EDA%20with%20Visualization.ipynb



EDA with SQL

- Summarize the SQL queries performed
- GitHub URL of completed EDA with SQL notebook, as an external reference and peer-review purpose:

https://github.com/victor-vu/IBM-Watson-Studio/blob/817ad5981c4fd383486738045ebc525e594cdd28/Week%202%20-%20EDA%20with%20SQL.ipynb

- **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 successful landing outcome in ground pad was acheived**
- **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**
- **List the failed landing_outcomes in drone ship, their booster versions, and launch site names for 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**

Build an Interactive Map with Folium

- Summarize map objects created and added to a folium map:
  - ✓ Marker for launch sites with color-labeled representing the launch outcomes
  - ✓ Cluster of markers for those sharing the same location
  - ✓ Distance lines and distance markers to proximities (eg. railway, coast line)
- GitHub URL of completed interactive map with Folium map, as an external reference and peer-review purpose:

https://github.com/victor-vu/IBM-Watson-Studio/blob/817ad5981c4fd383486738045ebc525e594cdd28/Week%203%20-%20Interactive%20Visual%20Analytics%20(Folium).ipynb

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
  - ✓ Drop-down box for selection of launch sites
  - ✓ Pie chart representing success count & success rate in all sites and individual site
  - ✓ Payload range slider to select the range of payload mass for display
  - ✓ Scatter plot representing the relationship between Payload and success count
- GitHub URL of completed Plotly Dash lab, as an external reference and peer-review purpose:

https://github.com/victor-vu/IBM-Watson-Studio/blob/2de1113316880c7b4c812d7bf164b1bf932978f7/Week%203%20-%20Dash%20&%20Plotly.ipynb

Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- GitHub URL of completed predictive analysis lab, as an external reference and peer-review purpose:

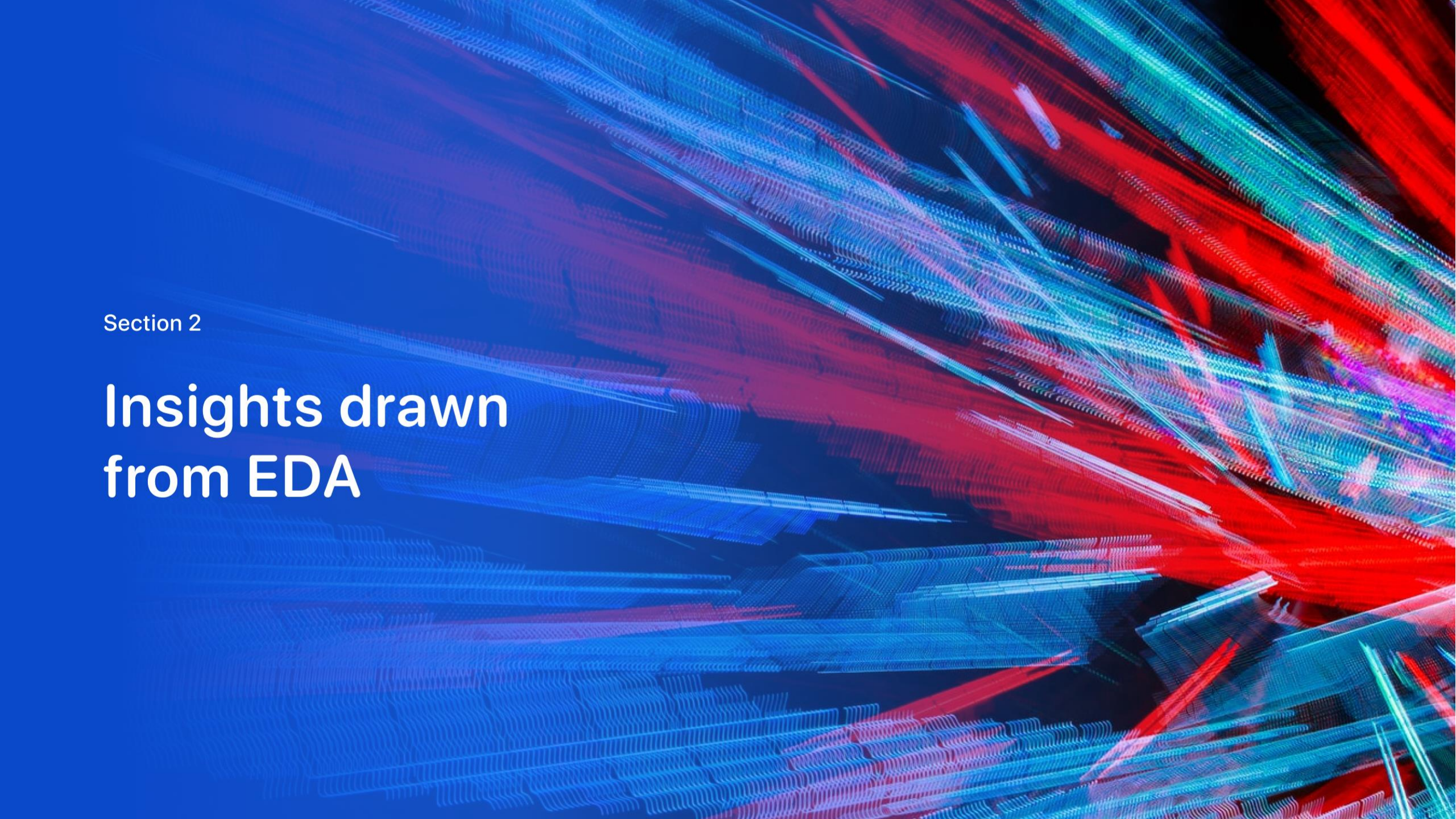
https://github.com/victor-vu/IBM-Watson-Studio/blob/2bab5863970fac770e418b612339d12994c9009c/Week%204%20-%20Machine%20Learning.ipynb

Results

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

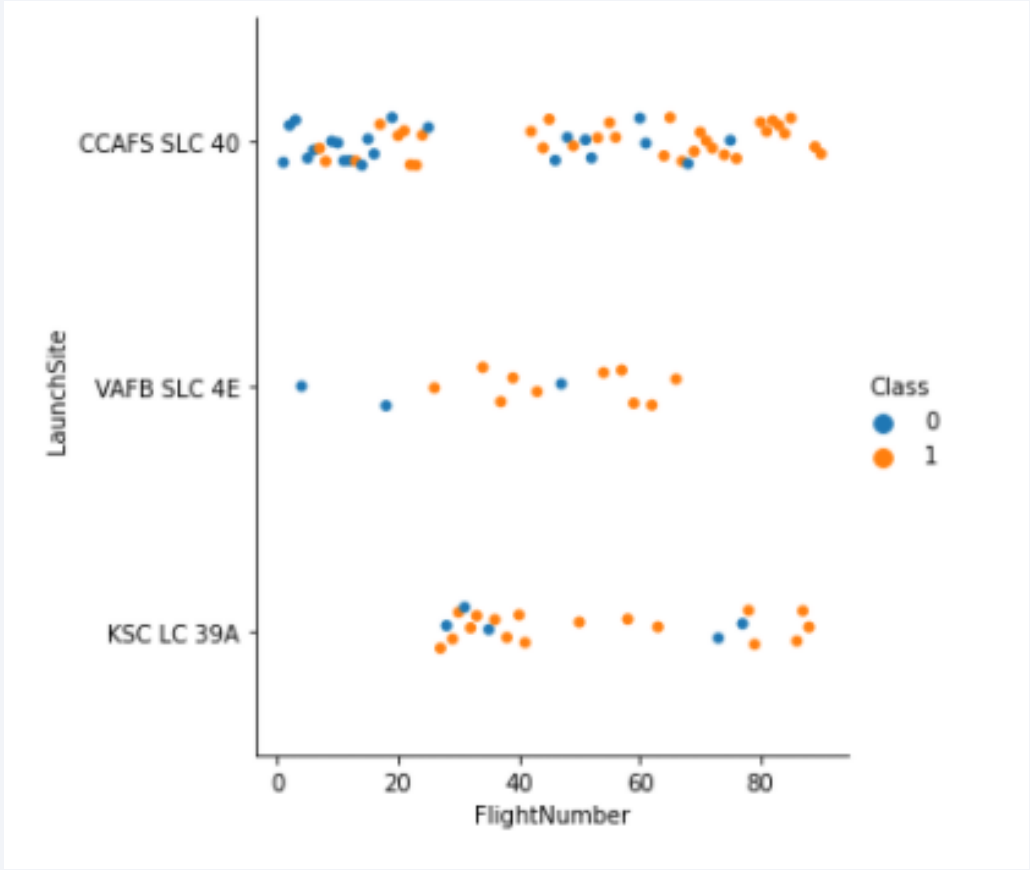
Section 2

Insights drawn from EDA



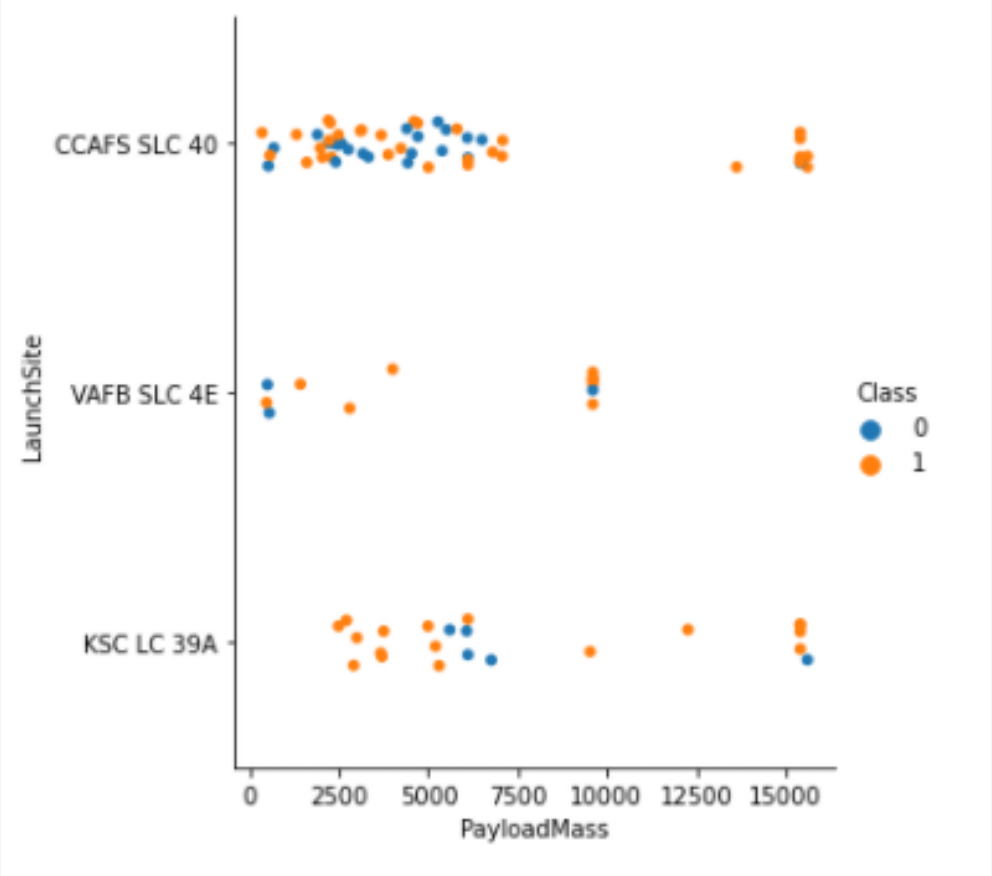
Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site
- Launch site CCAFS SLC 40 has the most number of flight while VAFB SLC 4E has the least one



Payload vs. Launch Site

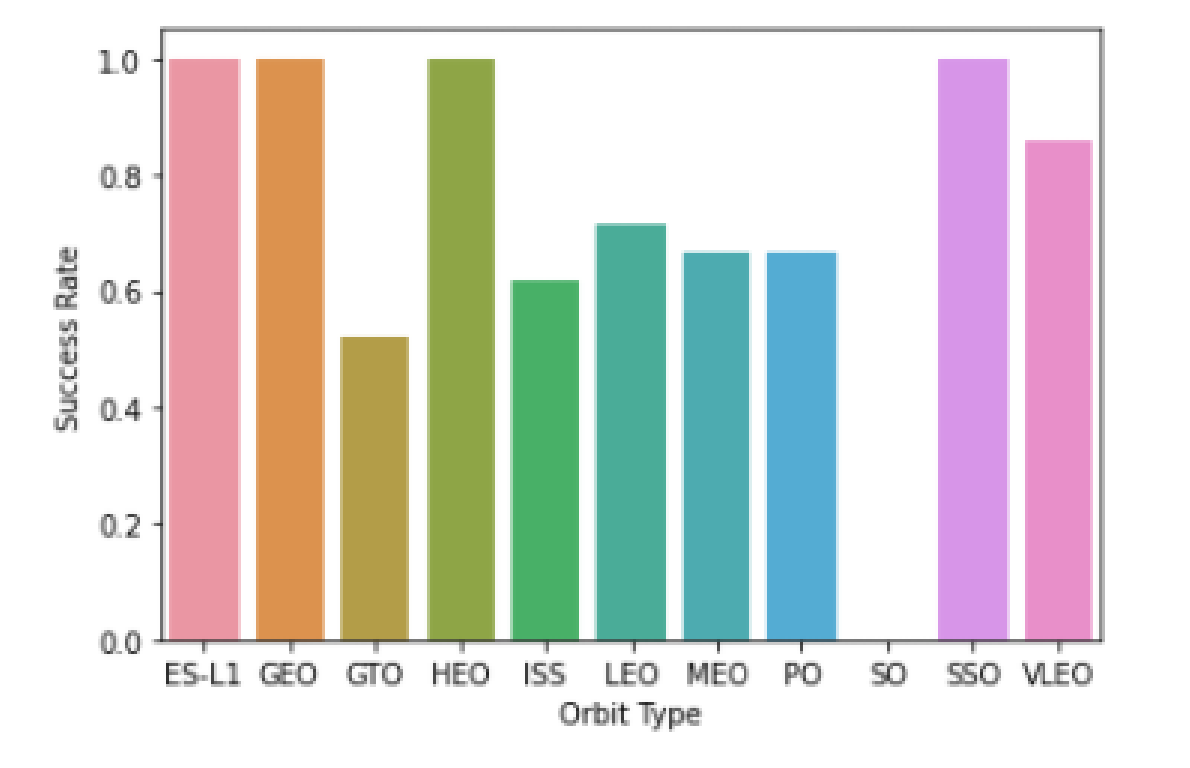
- Show a scatter plot of Payload vs. Launch Site
- Launch site VAFB SLC 4E does not operate high payload missles.



Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type

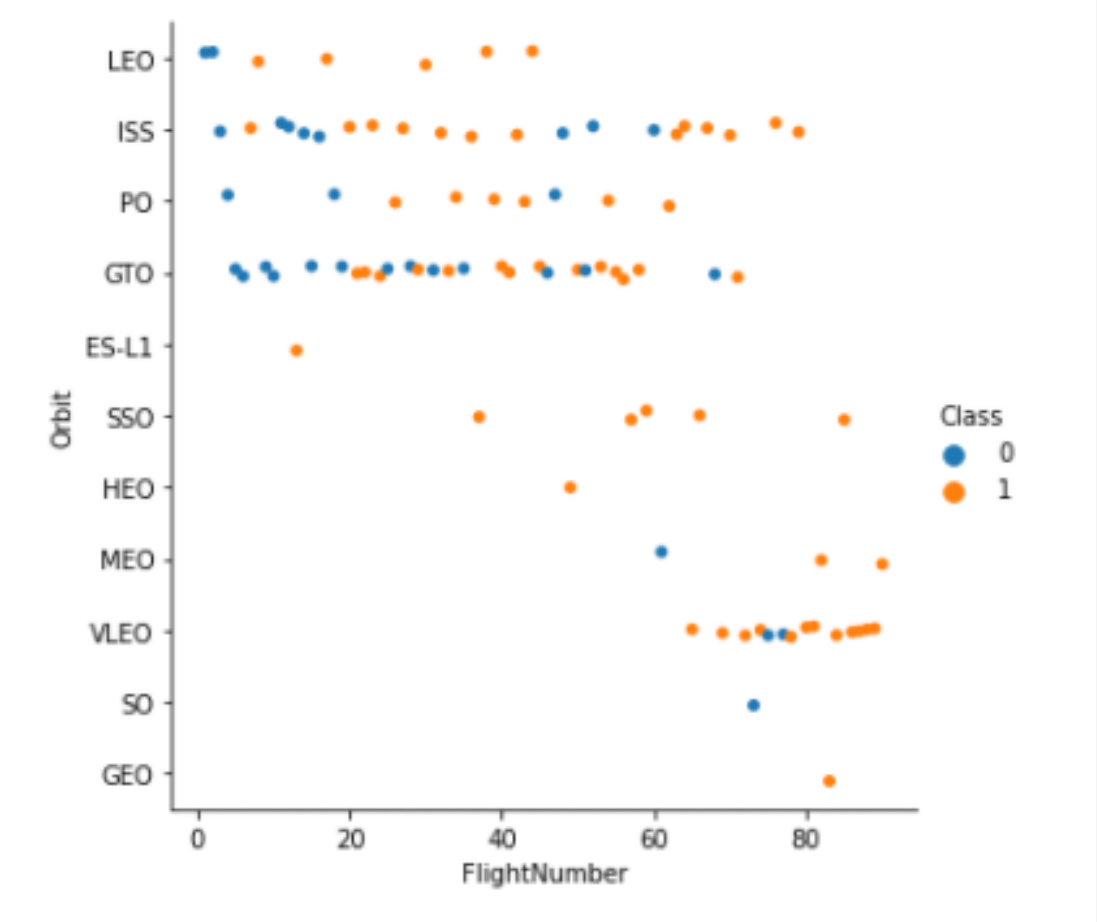
- The highest success rates were achieved with orbit type ES-L1, GEO, HEO and SSO



Flight Number vs. Orbit Type

- Show a scatter point of Flight number vs. Orbit type

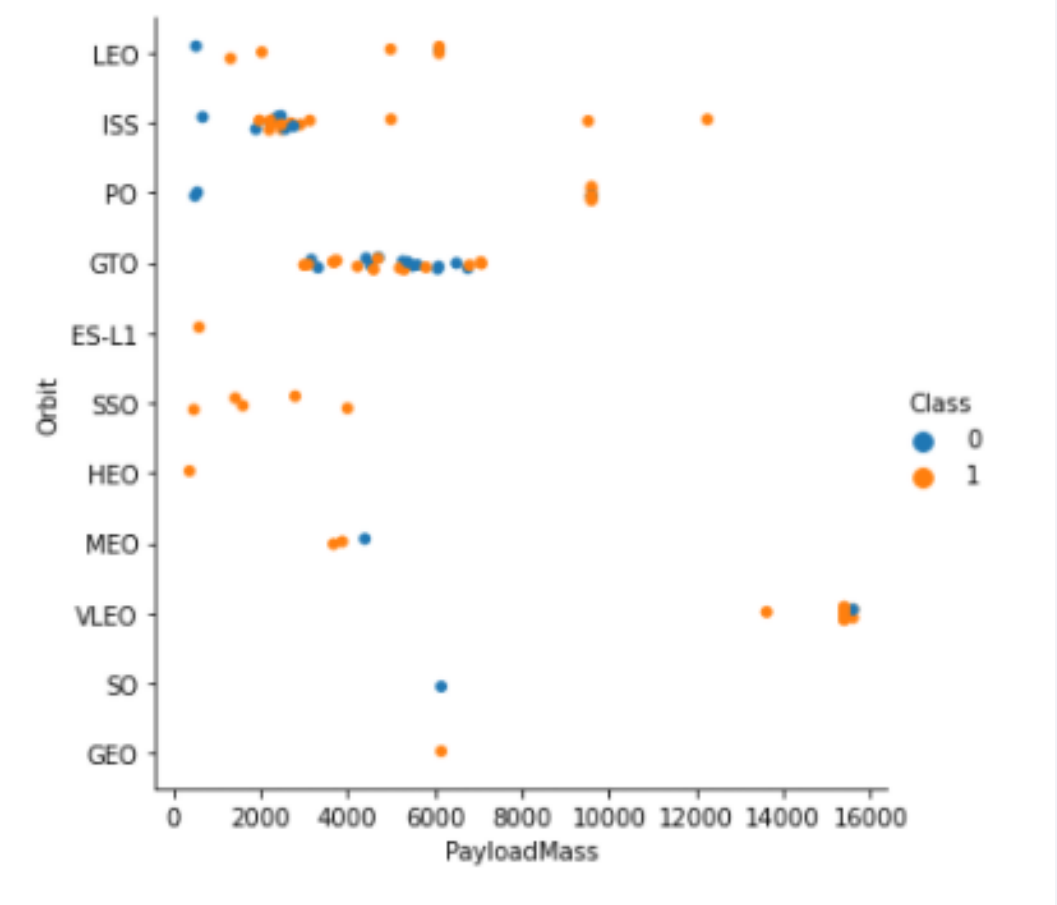
- In the LEO orbit, the success rate appears related to the number of flights. In contrast, there is likely no such relationship in the GTO orbit.



Payload vs. Orbit Type

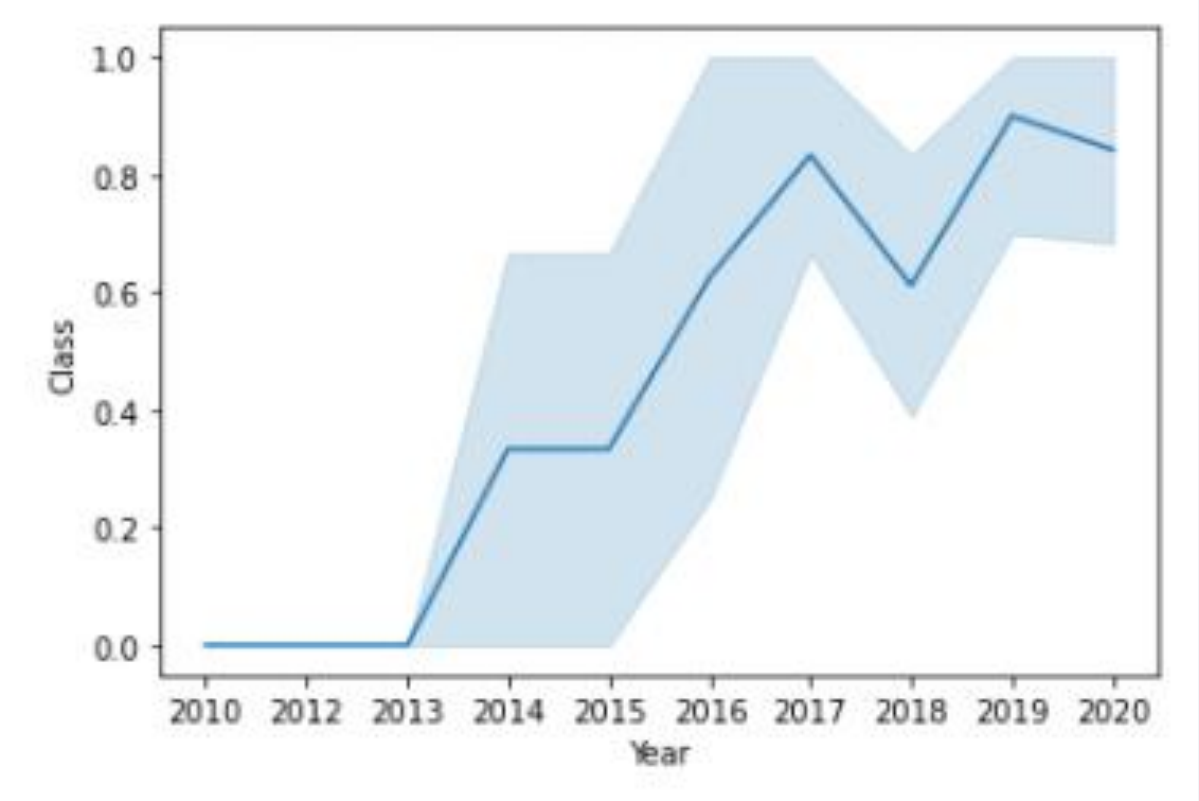
- Show a scatter point of payload vs. orbit type

- With heavy payloads, the successful landing rate are more for Polar, LEO and ISS. For GTO, we it is indistinguishable since both outcomes present in the range of payload mass



Launch Success Yearly Trend

- Show a line chart of yearly average success rate
- Success rate increases since 2013 till 2020



All Launch Site Names

- There are 4 unique launch sites, namely: CCAFS LC-40, CCAFS SLC-40, KSC LC-39A and VAFB SLC-4E
- Present query result:

```
%sql select distinct LAUNCH_SITE from SPACEXDATASET
```

```
 * ibm_db_sa://gjj23124:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/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`

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

```
 * ibm_db_sa://gjj23124:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/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

- Calculate the total payload carried by boosters from NASA

```
%sql select sum(PAYLOAD_MASS__KG_) as Total_Payload from SPACEXDATASET where CUSTOMER = 'NASA (CRS)'
```

```
 * ibm_db_sa://gjj23124:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

total_payload
45596

Average Payload Mass by F9 v1.1

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

```
%sql select avg(PAYLOAD_MASS__KG_) as Average_Payload from SPACEXDATASET where BOOSTER_VERSION = 'F9 v1.1'
```

```
 * ibm_db_sa://gjj23124:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

average_payload
2928

First Successful Ground Landing Date

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

```
%sql select min(DATE) as First_successful_landing from SPACEXDATASET
```

```
 * ibm_db_sa://gjj23124:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

first_successful_landing
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

```
%sql select distinct BOOSTER_VERSION, LANDING__OUTCOME, PAYLOAD_MASS__KG_ from SPACEXDATASET where LANDING__OUTCOME = 'Success
(drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000
```

```
 * ibm_db_sa://gjj23124:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

booster_version	landing__outcome	payload_mass__kg_
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

```
%sql select MISSION_OUTCOME, count(MISSION_OUTCOME) from SPACEXDATASET group by MISSION_OUTCOME
```

```
 * ibm_db_sa://gjj23124:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/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 mas

```
%sql select distinct BOOSTER_VERSION from SPACEXDATASET where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXDATA
SET)
```

```
 * ibm_db_sa://gjj23124:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

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

2015 Launch Records

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

```
%sql select BOOSTER_VERSION, LAUNCH_SITE, LANDING__OUTCOME from SPACEXDATASET where year(DATE) = 2015 and LANDING__OUTCOME = 'Fa
ilure (drone ship)'
```

```
 * ibm_db_sa://gjj23124:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/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

- 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(LANDING__OUTCOME) as COUNT from SPACEXDATASET where DATE between '2010-06-04' and '2017-03-2
0' group by LANDING__OUTCOME order by COUNT desc
```

```
 * ibm_db_sa://gjj23124:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

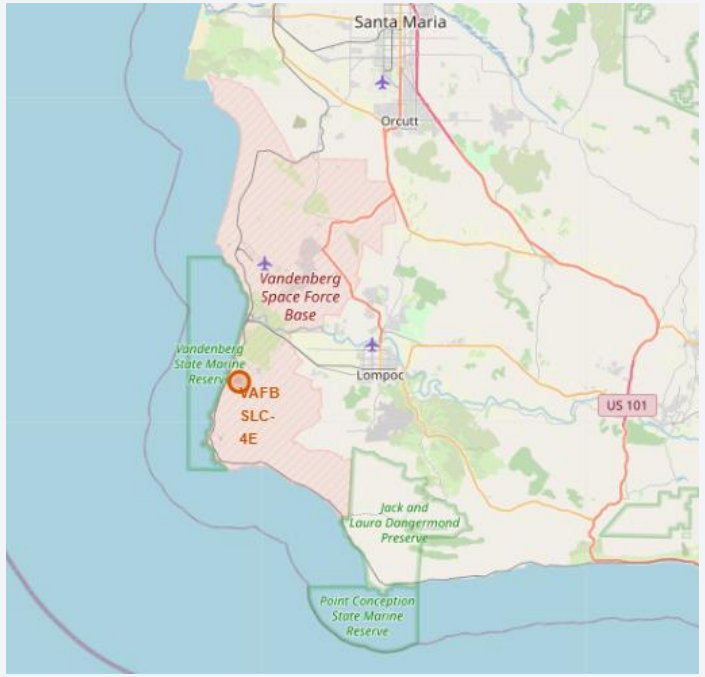
landing__outcome	COUNT
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 4

Launch Sites Proximities Analysis

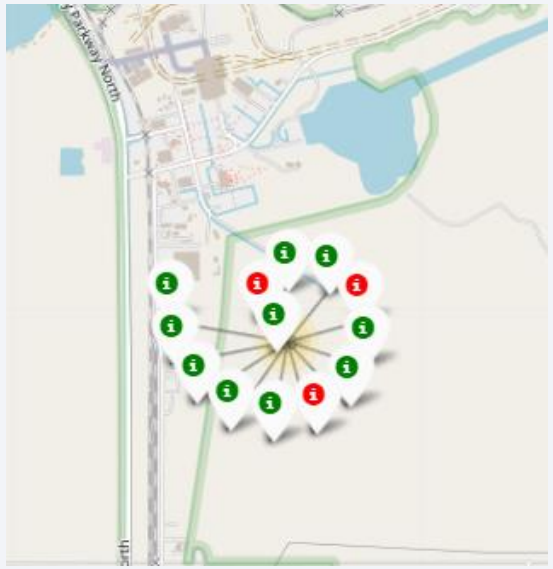
Mark all launch sites on a map

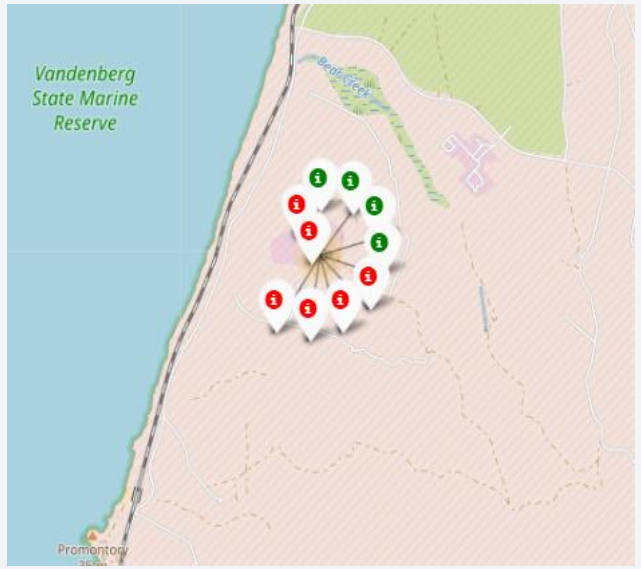
- All launch sites are in very close proximity to the coast, not to the Equator line

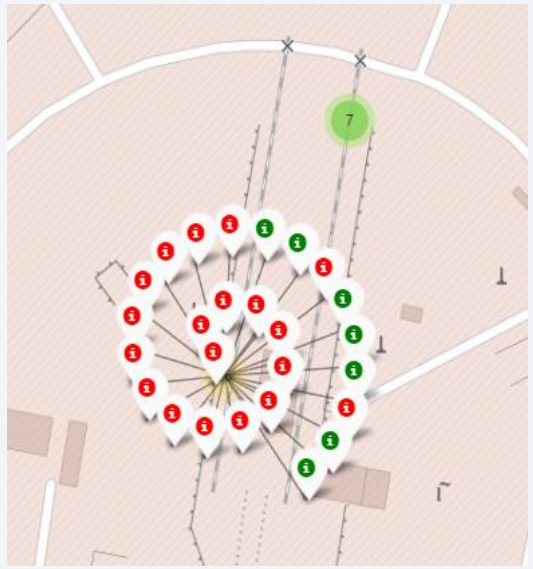


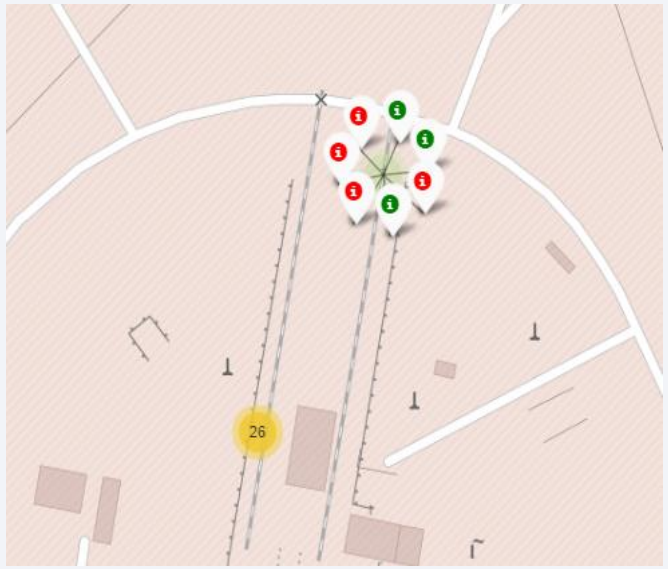
Mark the success/failed launches for each site on the map

- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map



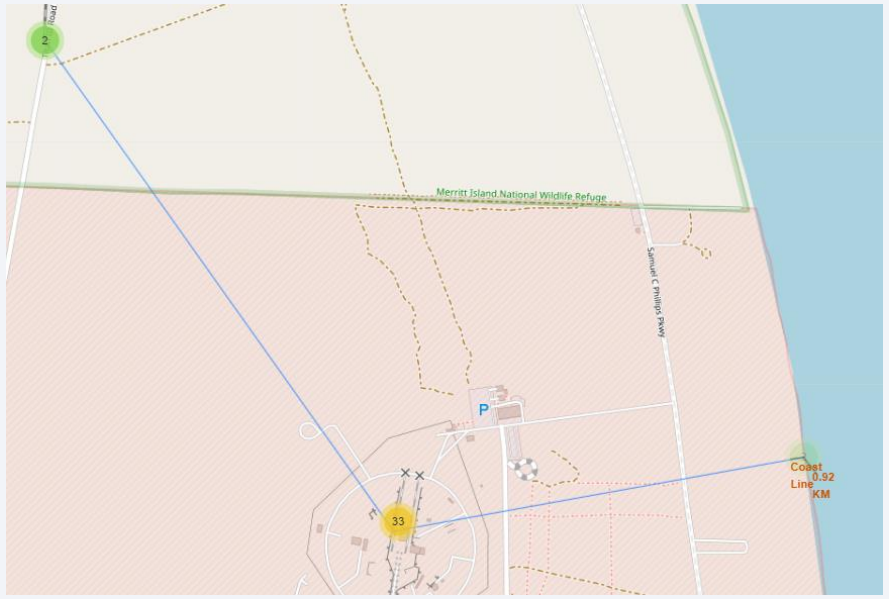


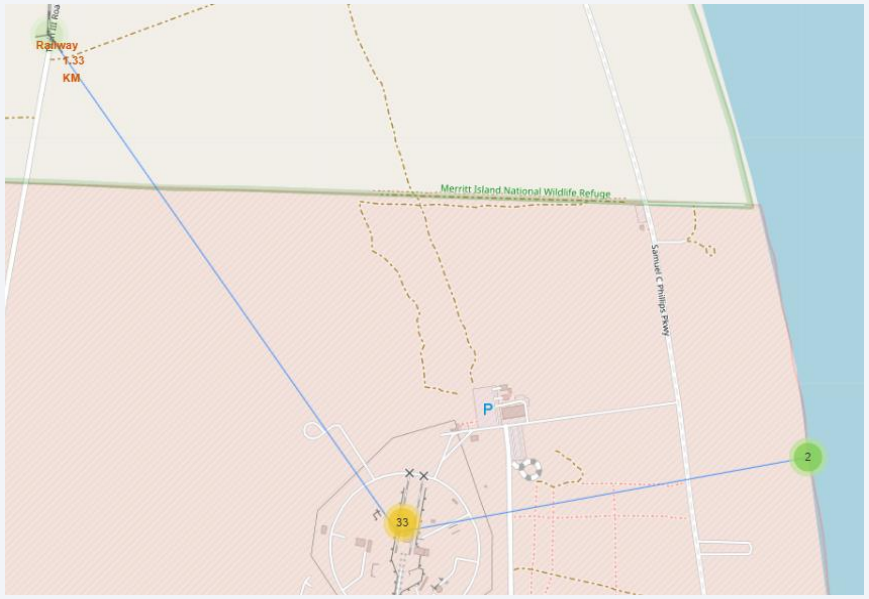




Calculate the distances between a launch site to its proximities

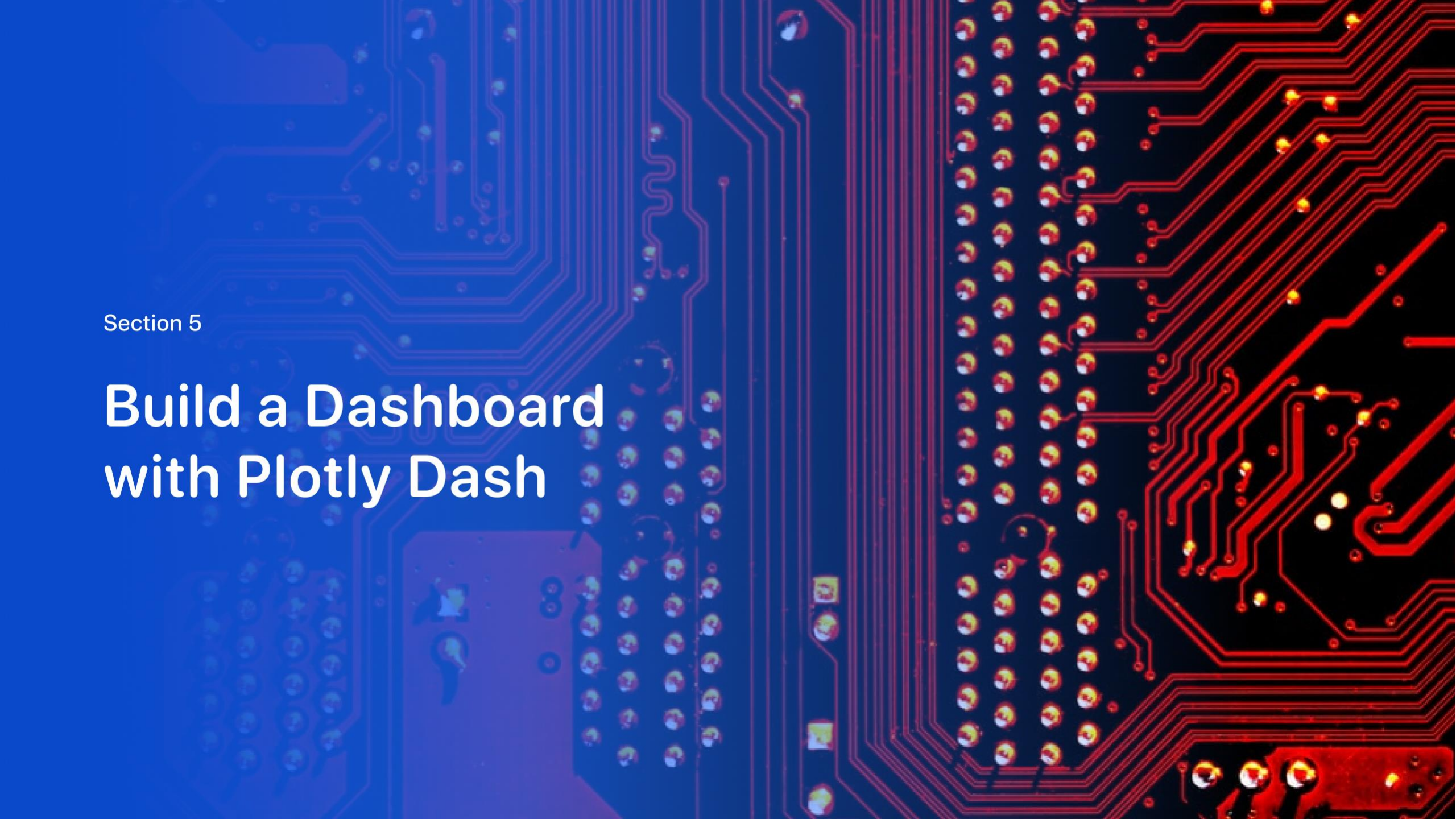
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed





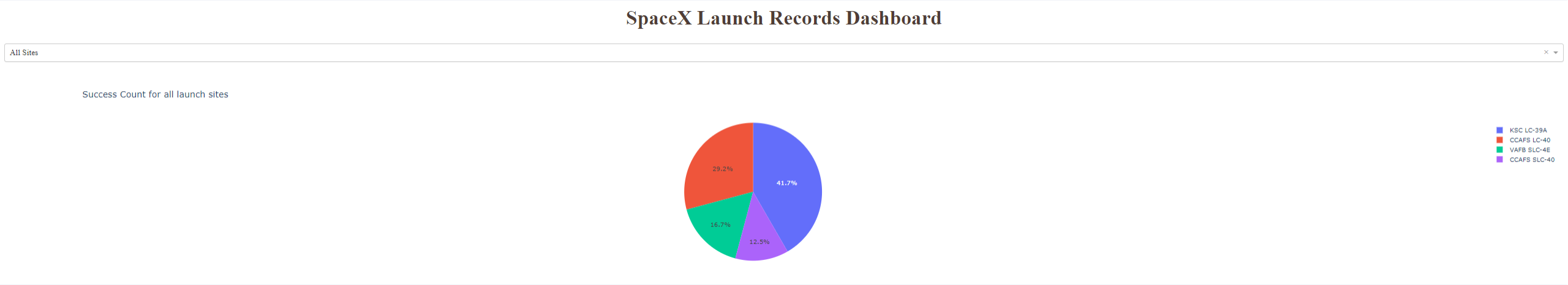
Section 5

Build a Dashboard with Plotly Dash



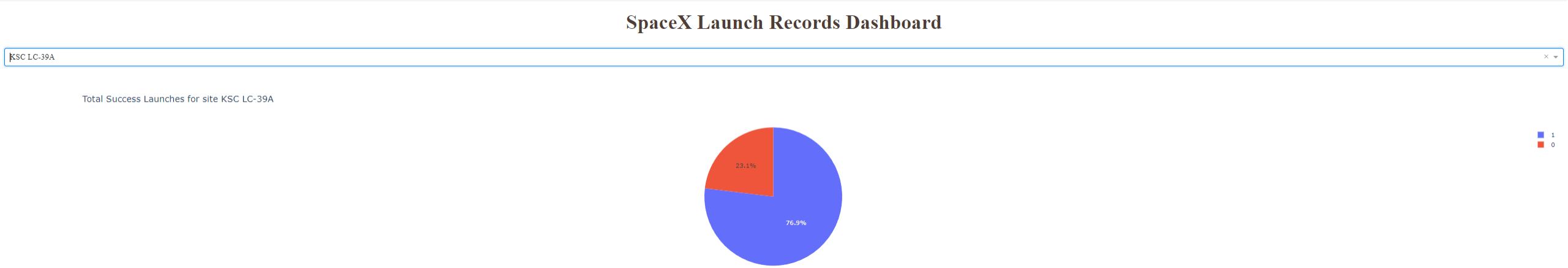
Success Count for All Launch Sites

- Show the screenshot of launch success count for all sites, in a piechart
- KSC LC-39A has the highest success count (41.7% of total all launch sites)



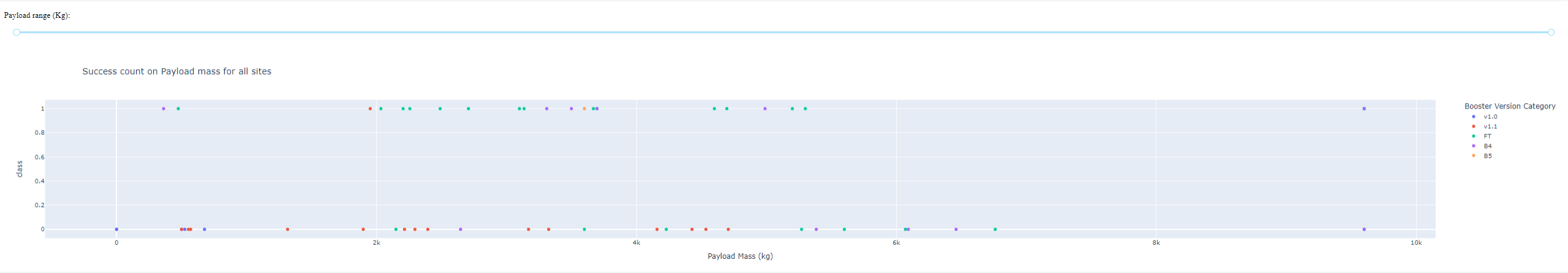
The Best Launch Site

- Success rate of the best launch site (KSC LC-39A) is 76.9%



Payload vs. Launch Outcome

- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Payload Mass below 6000 kg has the highest success rate



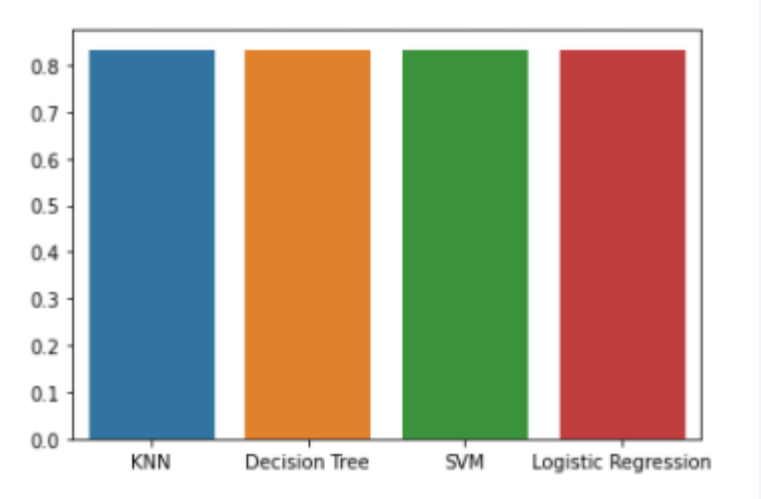
Section 6

Predictive Analysis (Classification)

Classification Accuracy

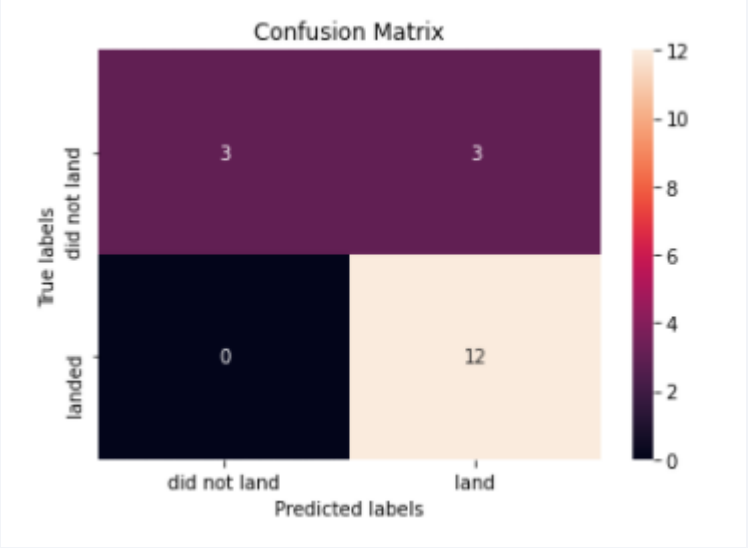
- Visualize the built model accuracy for all built classification models, in a bar chart

- All four selected models have similar classification accuracy



Confusion Matrix

- Since all models share the same accuracy score, their confusion matrices are similar.
- The below information can be seen from the confusion matrix:
  - 3 true negative in 18 predictions
  - 3 false negative in 18 predictions
  - 0 false positive in 18 predictions
  - 12 true positive in 18 predictions
  - Accuracy score = (3+12)/18 = 83.33%



Conclusions

- Overall, success rate of rocket launch increases since 2013 till 2020
- All launch sites locate near coastal lines; among them, the best launch site was KSC LC-39A, which had the highest success count
- Important predictors include payload mass and orbit type
- Predictive analyses have been done, which show all classification models have similar accuracy, with accuracy score >80%

Thank you!