IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

ALESIA KUCHARENKA

June 2025

Outline

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

Executive Summary

Today SPACEX is the most successful company in the field of commercial space.
What allows it to be so?

Summary of methodologies

- collecting relevant data from various sources
- improving the quality by performing data wrangling
- exploring the processed data
- applying some basic statistical analysis
- data visualization
- training a machine learning model to predict the result

Summary of all results

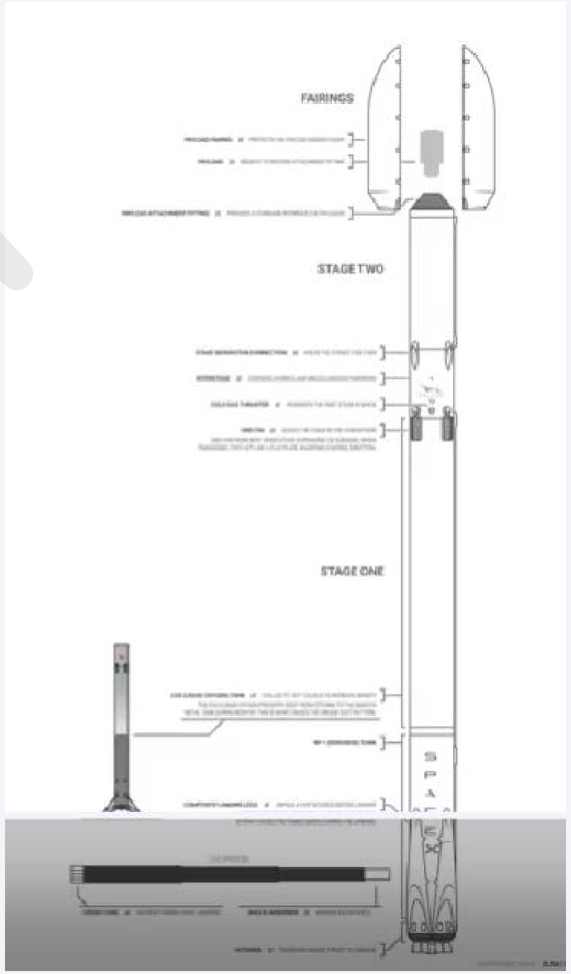
▲ All launch sites are in very close proximity to the ocean coast and to the Equator line, railways, highways and coastline and keep certain distance away from cities. ▲The total number of successful mission outcomes significantly exceed the number of failure missions. ▲KSC LC-39A launch site, rocket configuration Falcon 9 FT and payload mass between 2000 and 4000 kg have the highest success rate. ▲The first stage often returns successfully with more massive payloads, as the flight number increases, with heavy payloads are more for Polar, LEO and ISS. ▲ Types of orbits ES-L1, GEO, HEO and SSO have the highest success rates. ▲ All predictive models have the same classification accuracy.

Introduction

What is the reason for SPACEX success?

One reason SpaceX can do this is the rocket launches are relatively inexpensive. Unlike other rocket providers, SpaceX's Falcon 9 can recover the first stage.

This stage is large and it costs a lot.
Sometimes the first stage does not land. Sometimes it will crash. Other times, Space X will sacrifice the first stage due to the mission parameters like payload, orbit, and customer.



The goal is

- to determine the price of each launch
- to predict if SpaceX will reuse the first stage

Section 1

Methodology

Methodology

Executive Summary

- Data was gathered from the SpaceX REST API and obtained by web scraping related Wiki pages.

- The goal of wrangling was to perform EDA to find some patterns in the data and determine classes of landing labeled **1** for the booster landed successfully and **0** for unsuccessfully.

- An attempt was made to identify variables that directly affect the success of the first stage return.

- It were built an interactive map to find some geographical patterns about launch sites and a dashboard to perform interactive visual analytics on Spacex launch data in real-time.

- The goal of Classification was to predict the landing success of first stage. Specifically, to build a Machine Learning Pipeline for prediction and to determine Model with the Best Accuracy.

Data Collection

The data was collected in two ways:

- gathered from the SpaceX REST API (using the requests library to obtain the launch data and converting the resulting list of JSON objects to a dataframe)

```
url="https://api.spacexdata.com/v4/launches/past"
response =requests.get(url)
response.json()
```

- obtained by web scraping related Wiki pages (using the Python BeautifulSoup package to web scrape some HTML tables that contain valuable records and to parse the data from those tables and convert them into a Pandas data frame for further visualization and analysis)

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin1A	167.743129	9.047721
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2A	167.743129	9.047721
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2C	167.743129	9.047721
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin3C	167.743129	9.047721
4	6	2010-06-04	Falcon 9	NaN	LEO	CCAFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857

Data Collection – SpaceX API

- Rocket launch data was requested from SpaceX Application Programming Interface (API) with Uniform Resource Locator (URL)
- The response was viewed by calling the .json() method
- The structured json data was "normalized" into a flat table

Link

[to-try/jupyter-labs-spacex-data-collection-api (1).ipynb at main · Alesia-Koo/to-try]()

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

```
response = requests.get(spacex_url)
```

```
response.json()
data=pd.json_normalize(response.json())
```

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN
4	6	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.C

Data Collection - Scraping

The data was extracted from a Wikipedia page and was stored in a HTML table

It was

- used GET method to request the data HTML page
- and created a BeautufulSoup object from a response text content
- after filling in the parsed values into special object it was created a dataframe

Link

[to-try/jupyter-labs-webscraping (2).ipynb at main · Alesia-Koo/to-try]()

```
response=requests.get(static_url).text
```

```
BeautifulSoup=BeautifulSoup(response, "html.parser")
```

```
df= pd.DataFrame({ key:pd.Series(value) for key, value in launch_dict.items() })
```

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Launchoutcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	NaN	Success\n	F9 v1.07B0003.18	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA	NaN	Success	F9 v1.07B0004.18	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA	NaN	Success	F9 v1.07B0005.18	No attempt\n	22 May 2012	07:44

Data Wrangling

The goal was to perform some Exploratory Data Analysis (EDA) to find some patterns in the data and determine classes of landing labeled **1** for the booster landed successfully and **0** for unsuccessfully.

It was

- determined the number of landing outcomes on the column OUTCOMES

```
landing_outcomes=df['Outcome'].value_counts()
```

- created a set of outcomes where the first stage did not land successfully

```
bad_outcomes=set(landing_outcomes.keys()[[1,3,5,6,7]])
```

- created labels from column OUTCOME

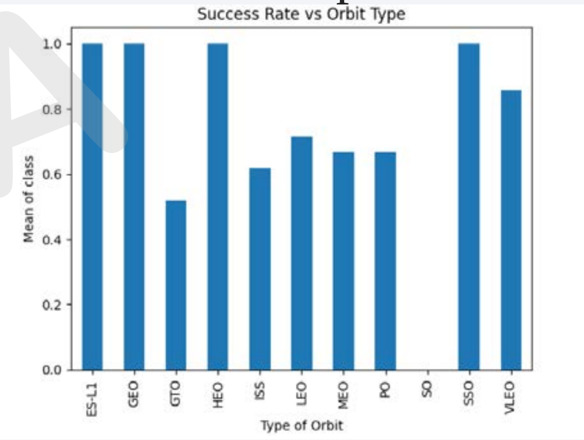
```
# landing_class = 0 if bad_outcome
# landing_class = 1 otherwise
landing_class=np.where(df['Outcome'].isin(bad_outcomes))
```

	Class
0	0
1	0
2	0
3	0

Link
[to-try/labs-jupyter-spacex-Data wrangling.ipynb at main · Alesia-Koo/to-try]

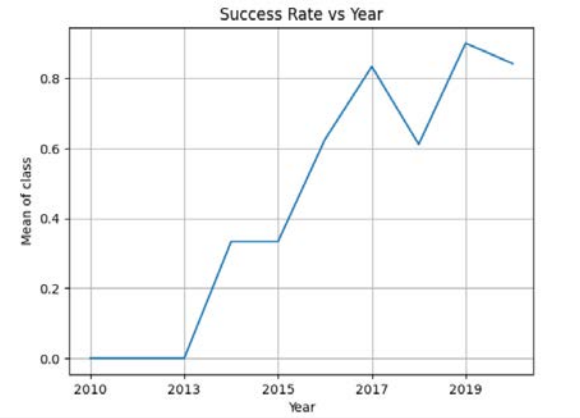
EDA with Data Visualization

An attempt was made to identify variables that directly affect the success of the first stage return. The most suitable for observing the relationship between two variables were scatter and line plots and bar chart.



RESULTS:

- with more massive payloads, the first stage often returns successfully;
- as the flight number increases, the first stage is more likely to land successfully;
- for the VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000);
- with heavy payloads the successful landing are more for Polar, LEO and ISS;
- in the LEO orbit, success seems to be related to the number of flights; in the GTO orbit, there appears to be no relationship between flight number and success.



It was observed that the success rate since 2013 kept increasing till 2020 and which orbits have the highest success rates.

Link

[to-try/edadataviz.ipynb at main · Alesia-Koo/to-try]

EDA with SQL

Using SQL queries it was checked information about

- the names of the unique launch sites in the space mission;
- the total and average payload masses carried by different booster versions;
- booster versions which have carried the maximum payload mass;
- the date when the first successful landing outcome in ground pad was achieved;
- the total number of successful and failure mission outcomes;
- various information about the success of the first stage landing at different time intervals.

Link

to-try/jupyter-labs-eda-sql-coursera_sqllite.ipynb at main · Alesia-Koo/to-try

Build an Interactive Map with Folium

May the launch success rate depend on the location and its proximities?

It was built an interactive map to find some geographical patterns about launch sites.

Using circles, markers, lines and special coordinates on the world map it was marked all SpaceX launch sites, the success/failed launches for each site and calculated the distances between a launch site to its proximities (such as closest cities, railways, highways).

If a launch was successful (class=1), then it was used a green marker and if a launch was failed, it was used a red marker (class=0)
A launch only happened in one of the four launch sites, which means many launch records have the exact same coordinate. To solve this problem it was used marker clusters.

Link

to-try/lab_jupyter_launch_site_location (3).ipynb at main · Alesia-Koo/to-try

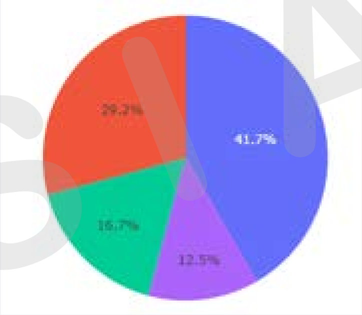
Build a Dashboard with Plotly Dash

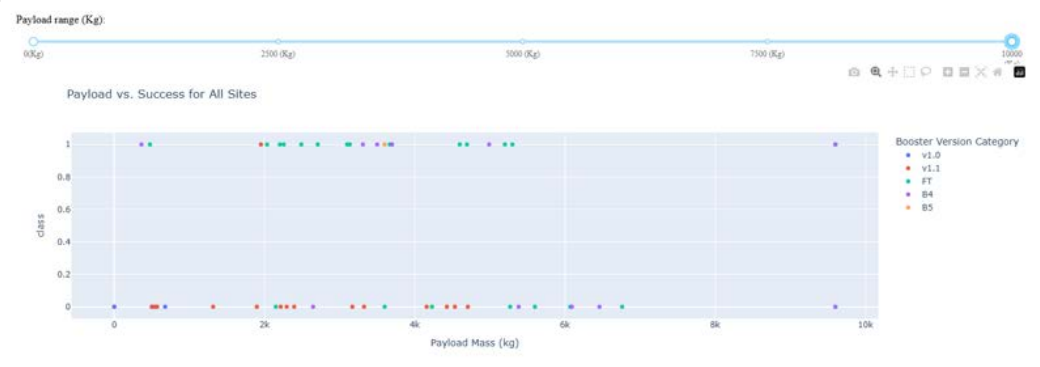
A Dashboard was built to perform interactive visual analytics on Spacex launch data in real-time.
It contains a dropdown list and a range slider to interact with charts:

- to get the selected launch site and render a pie chart visualizing launch success counts;

- to find if variable payload is correlated to mission outcome;

- to visually observe how payload may be correlated with mission outcomes for selected site(s) and with different boosters.





Link
[to-try/spacex-dash-app (3).py at main · Alesia-Koo/to-try]

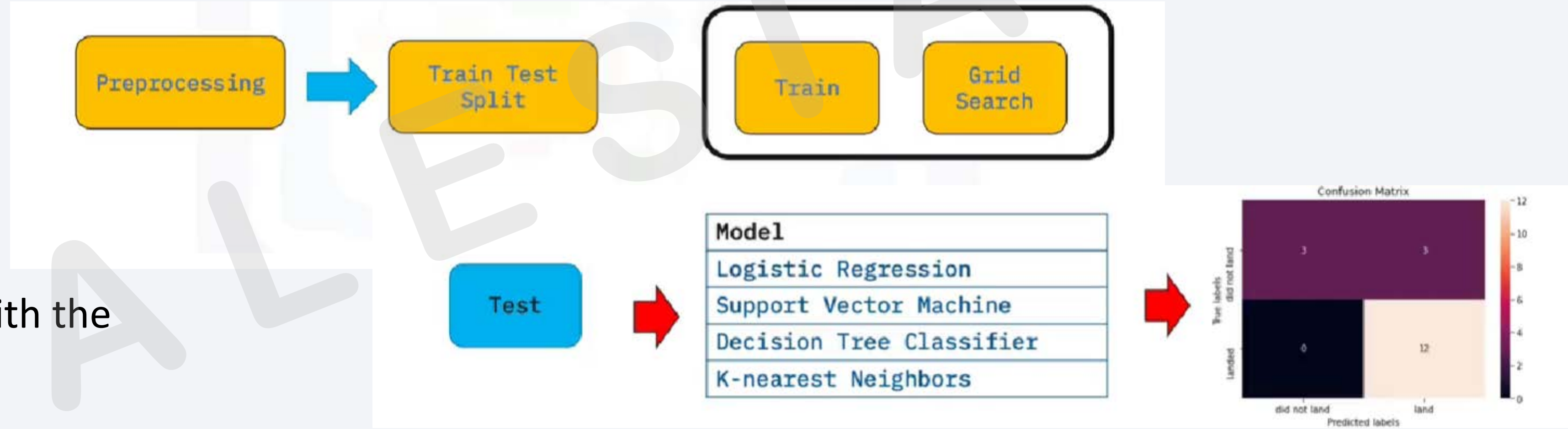
Predictive Analysis (Classification)

The goal of Classification was to predict the landing success of first stage.

Milestones:

- to build a Machine Learning Pipeline for prediction

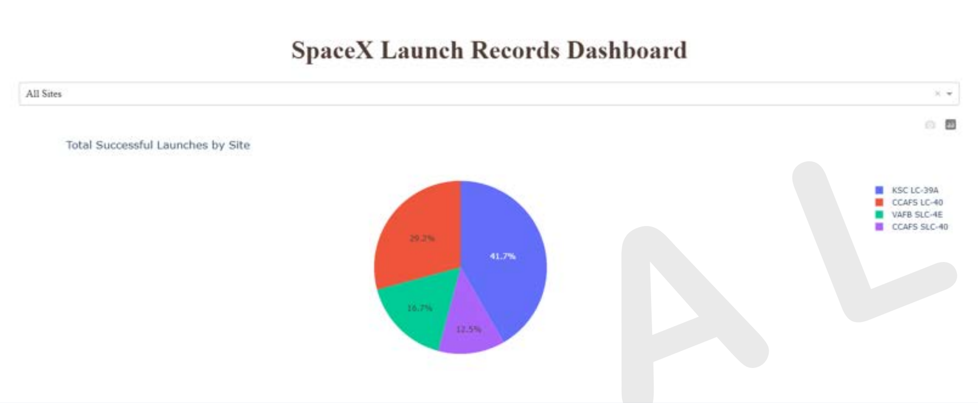
- to determine Model with the Best Accuracy

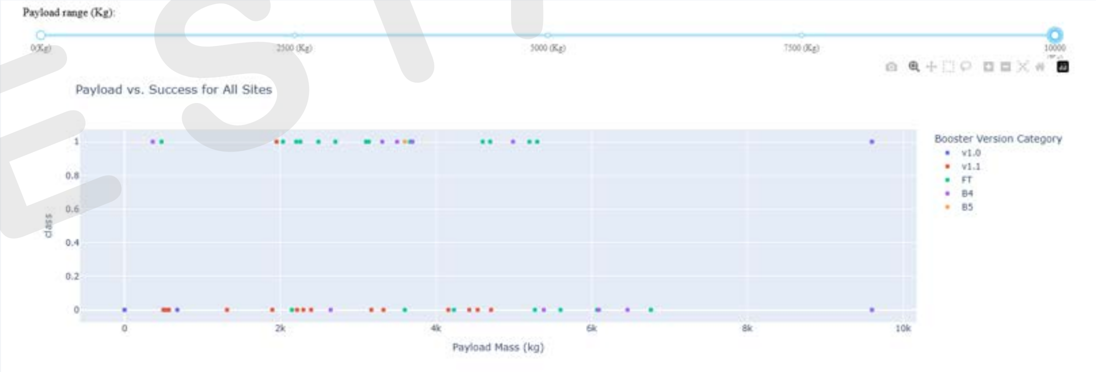


Link
[to-try/SpaceX_Machine Learning Prediction_Part_5 (2).ipynb at main · Alesia-Koo/to-try]

Results

The first stage often returns successfully with more massive payloads, as the flight number increases, with heavy payloads are more for Polar, LEO and ISS. The success rate since 2013 kept increasing till 2020. ES-L1, GEO, HEO and SSO types of orbits have the highest success rates. The total number of successful mission outcomes significantly exceed the number of failure missions.



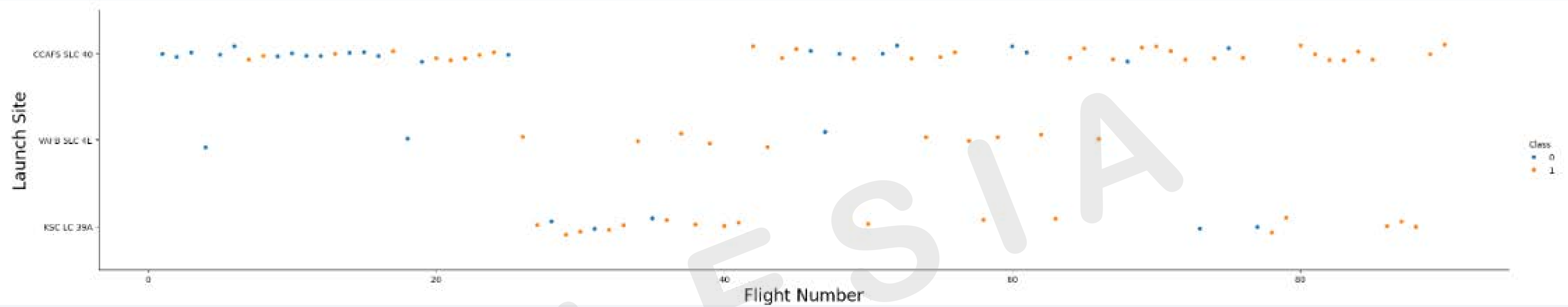


The accuracy of all models from Machine Learning (Logistic Regression, Support Vector machines, Decision Tree Classifier, and K-nearest neighbors) on the test data has approximately the same result (0.833-0.875)

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site



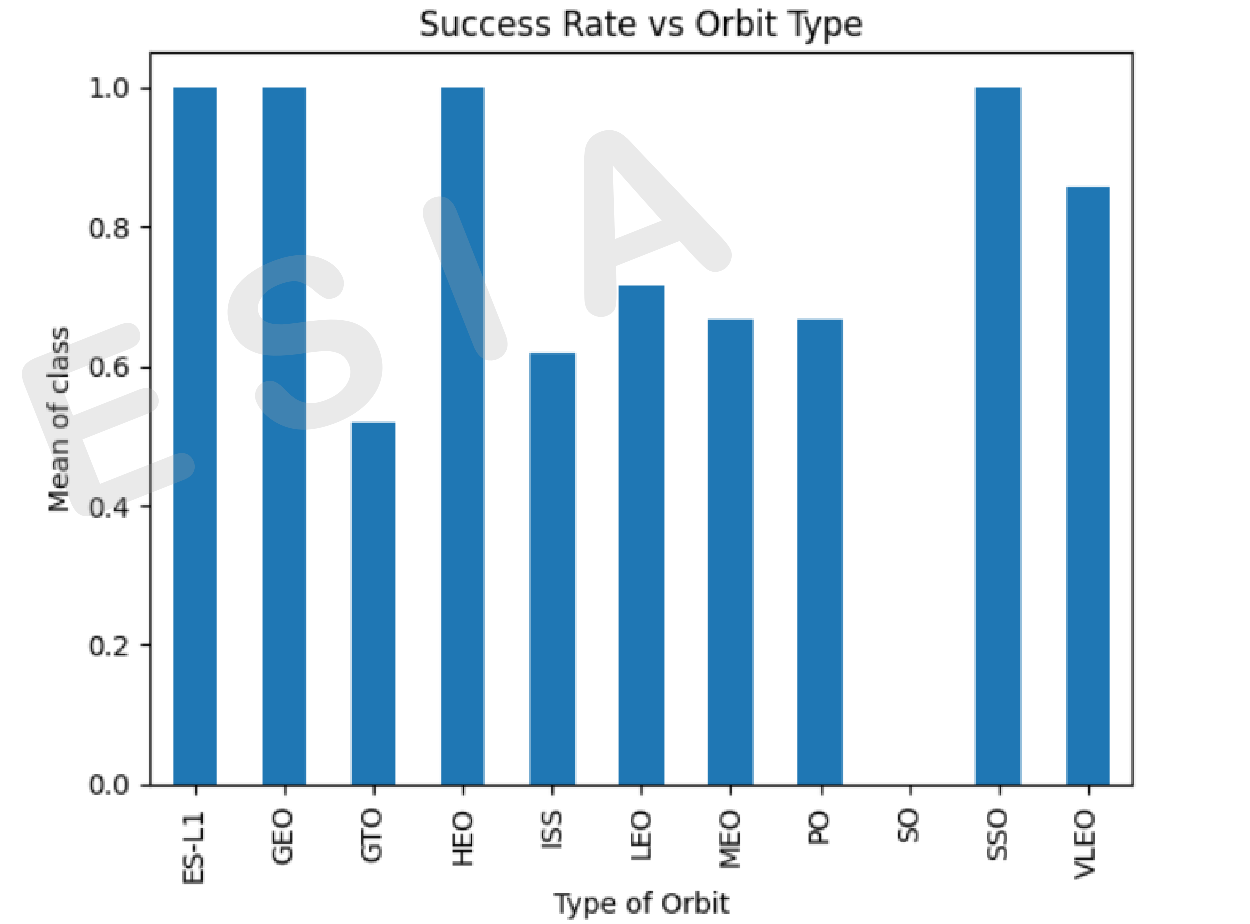
- As the flight number increases, the first stage is more likely to land successfully.
- The largest number of launches were made from the site CCAFS SLC-40. There have been many successful and many failed missions here.
- Least number of launches were made from the site VAFB-SLC. Most missions were successful.

Payload vs. Launch Site

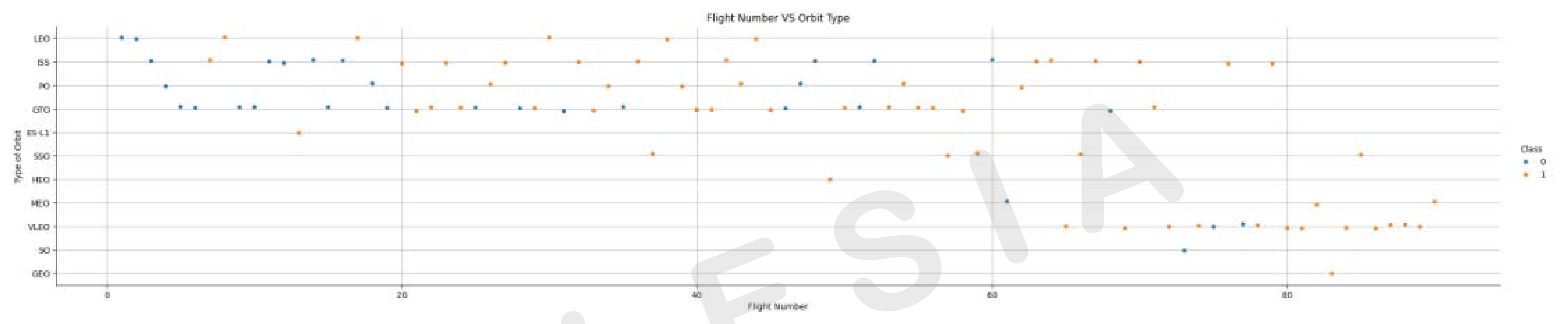
- For the VAFB-SLC launch site there are no rockets launched for heavy payload mass (greater than 10000).
- The most successful launches for the site KSC LC-39A were those with smaller payload mass.

Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO types of orbits have the highest success rates.
- There were no successful launches into the SO orbit type.

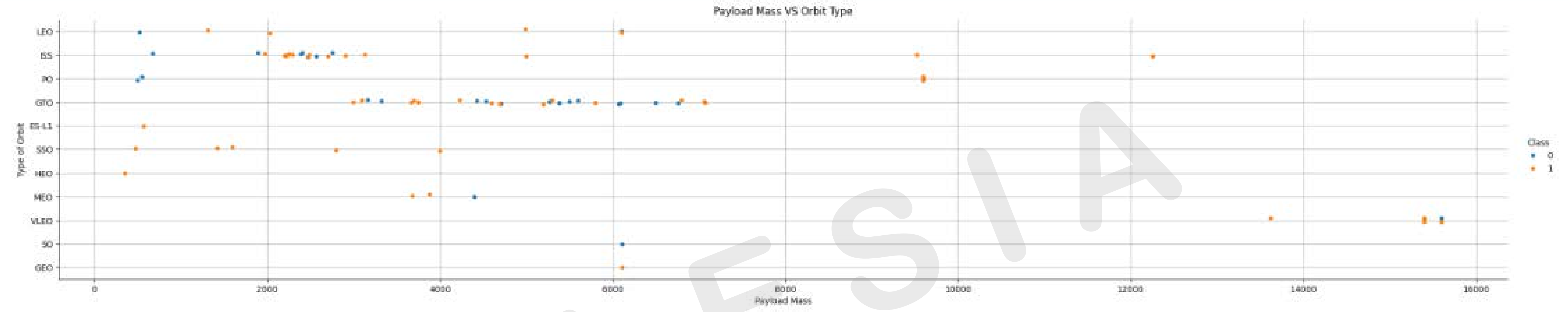


Flight Number vs. Orbit Type



- In the LEO orbit, success seems to be related to the number of flights.
- In the GTO orbit, there appears to be no relationship between flight number and success.
- There are only successful launches in the orbit SSO.

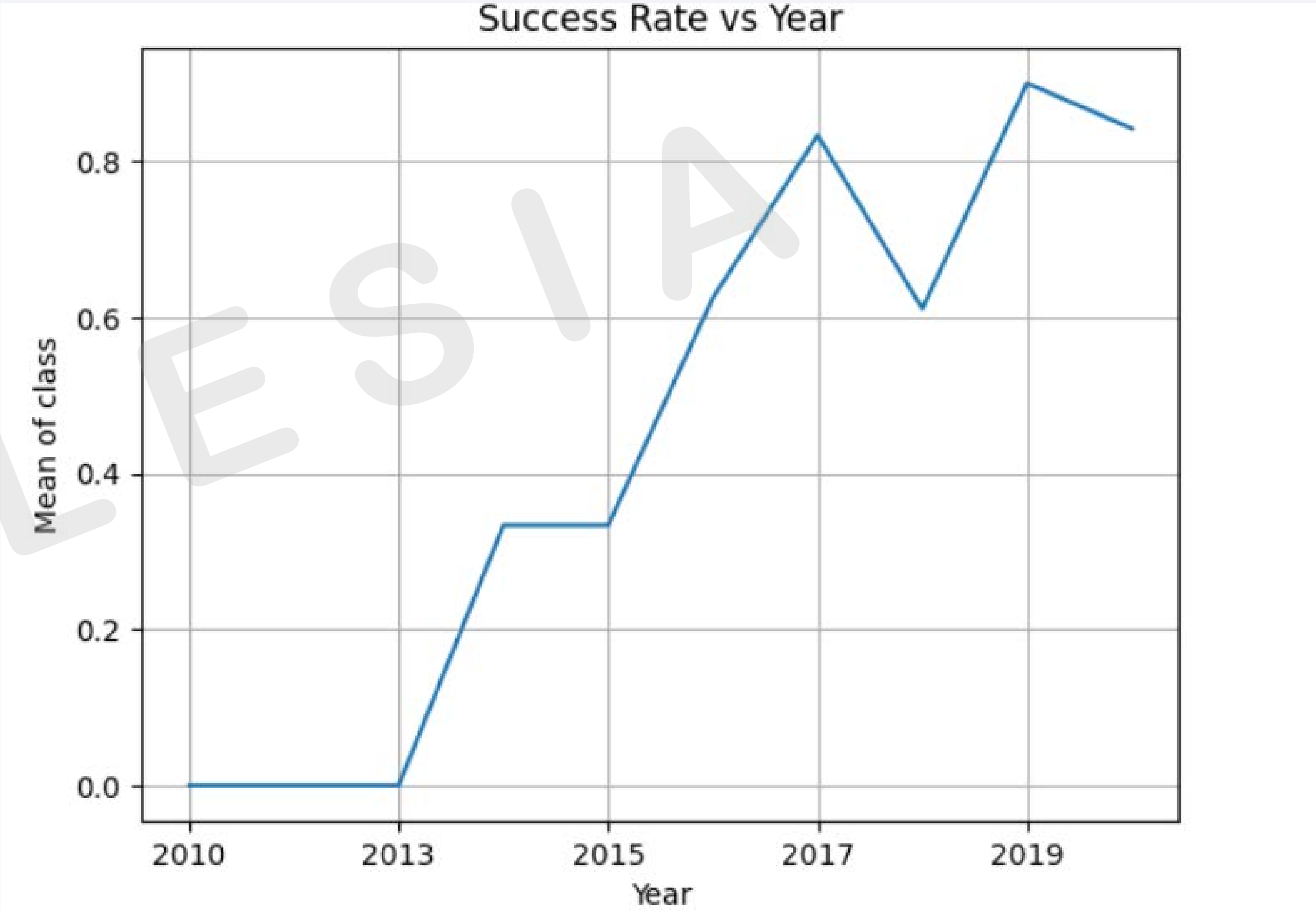
Payload vs. Orbit Type



- With heavy payloads the successful landing are more for Polar, LEO and ISS orbits.
- For the GTO orbit, it's difficult to distinguish between successful and unsuccessful landings as both outcomes are present.

Launch Success Yearly Trend

- The success rate since 2013 kept increasing till 2020.



All Launch Site Names

```
%%sql
SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
```

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

- There are only four unique names of launch sites of Space X

Launch Site Names Begin with 'CCA'

```
%%sql
SELECT * from SPACEXTABLE where Launch_Site like 'CCA%' limit 5;
```

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

- All of this launches were in the same orbit and had successful mission outcomes.

Total Payload Mass

```
%%sql
select  sum(PAYLOAD_MASS__KG_) as Total_mass_NASA_CRS , Customer from  SPACEXTABLE where Customer='NASA (CRS)' ;
```

Total_mass_NASA_CRS	Customer
45596	NASA (CRS)

- This result is for NASA (CRS) customer only.

Average Payload Mass by F9 v1.1

```
%%sql
select avg(PAYLOAD_MASS__KG_) as Average_of_F9_v_1_1 , Booster_Version from SPACEXTABLE where Booster_Version='F9 v1.1';
```

Average_of_F9_v_1_1	Booster_Version
2928.4	F9 v1.1

- This result is for Booster version F9 v1.1 only.

First Successful Ground Landing Date

```
%%sql
select min(Date), Mission_Outcome, Landing_Outcome from SPACEXTABLE where Landing_Outcome='Success (ground pad)';
```

min(Date)	Mission_Outcome	Landing_Outcome
2015-12-22	Success	Success (ground pad)

- The first successful landing outcome in ground pad was achieved on December 22, 2015.

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql
select Booster_Version, PAYLOAD_MASS__KG_,  Landing_Outcome from SPACEXTABLE where Landing_Outcome='Success (drone ship)' and PAYLOAD_MASS__KG_ between 4000 and 6000;
```

Booster_Version	PAYLOAD_MASS__KG_	Landing_Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

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

Total Number of Successful and Failure Mission Outcomes

```
%%sql
select count(Mission_Outcome) as Success_mission,Mission_Outcome from SPACEXTABLE where Mission_Outcome='Success' ;
```

Success_mission	Mission_Outcome
98	Success

```
%%sql
select count(Mission_Outcome)  as Failure_mission, Mission_Outcome from SPACEXTABLE where Mission_Outcome!='Success' ;
```

Failure_mission	Mission_Outcome
3	Failure (in flight)

- The vast majority of missions were successful.

Boosters Carried Maximum Payload

```
%%sql
select  distinct Booster_Version, PAYLOAD_MASS__KG_ from SPACEXTABLE where PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG_) from SPACEXTABLE ) ;
```

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

- The names of the booster versions which have carried the maximum payload mass.

2015 Launch Records

```
%%sql
select substr(Date,6,2) as Month_names, Date, Booster_Version, Launch_Site, Landing_Outcome from  SPACEXTABLE
```

```
where Landing_Outcome='Failure (drone ship)' and substr(Date,0,5)='2015';
```

Month_names	Date	Booster_Version	Launch_Site	Landing_Outcome
01	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- The failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

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

```
%%sql
select Landing_Outcome, count(Landing_Outcome) as Count, Date from SPACEXTABLE where Date between '2010-06-04' and
```

```
'2017-03-20' group by Landing_Outcome order by Count desc ;
```

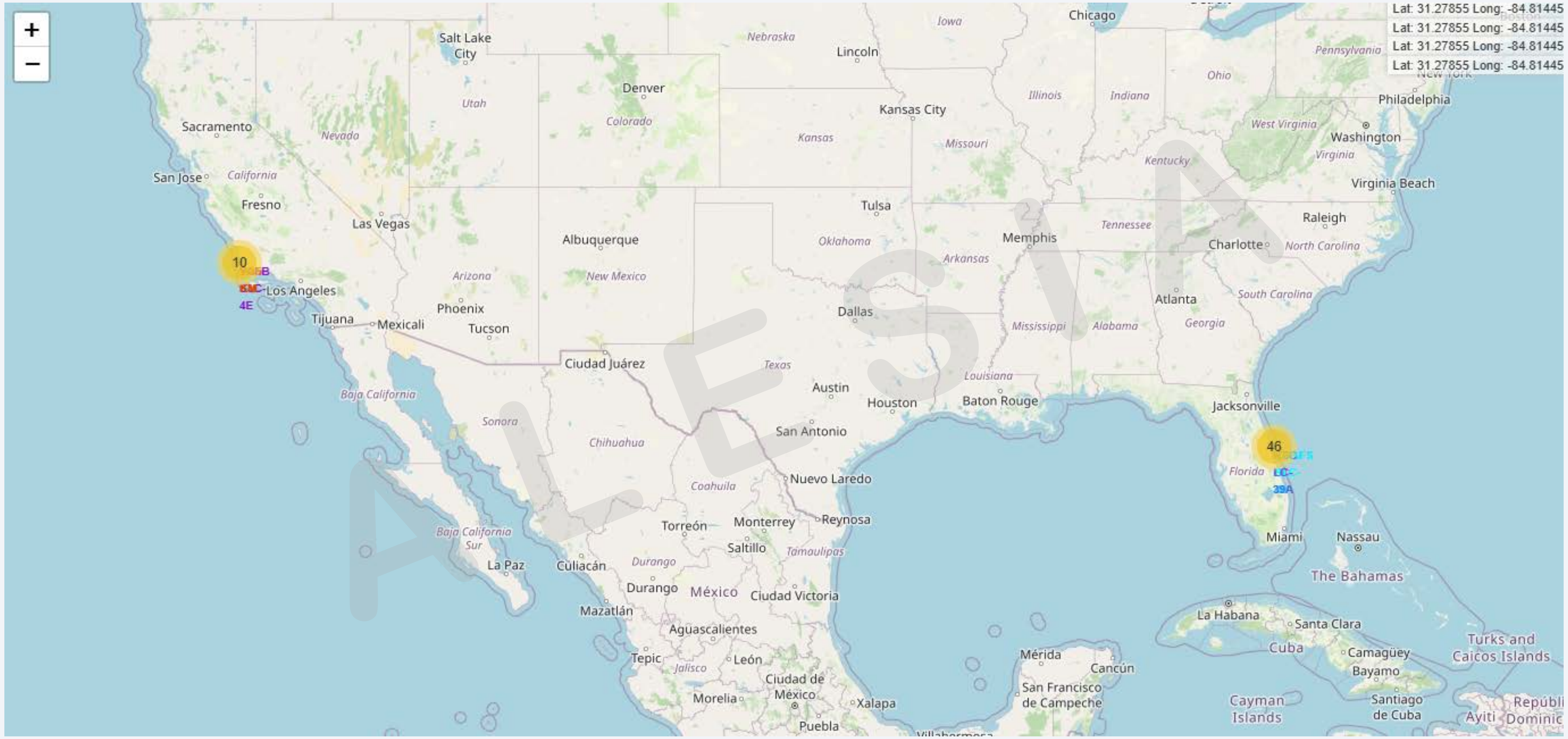
Landing_Outcome	Count	Date
No attempt	10	2012-05-22
Success (drone ship)	5	2016-04-08
Failure (drone ship)	5	2015-01-10
Success (ground pad)	3	2015-12-22
Controlled (ocean)	3	2014-04-18
Uncontrolled (ocean)	2	2013-09-29
Failure (parachute)	2	2010-06-04
Precluded (drone ship)	1	2015-06-28

- The count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order

Section 3

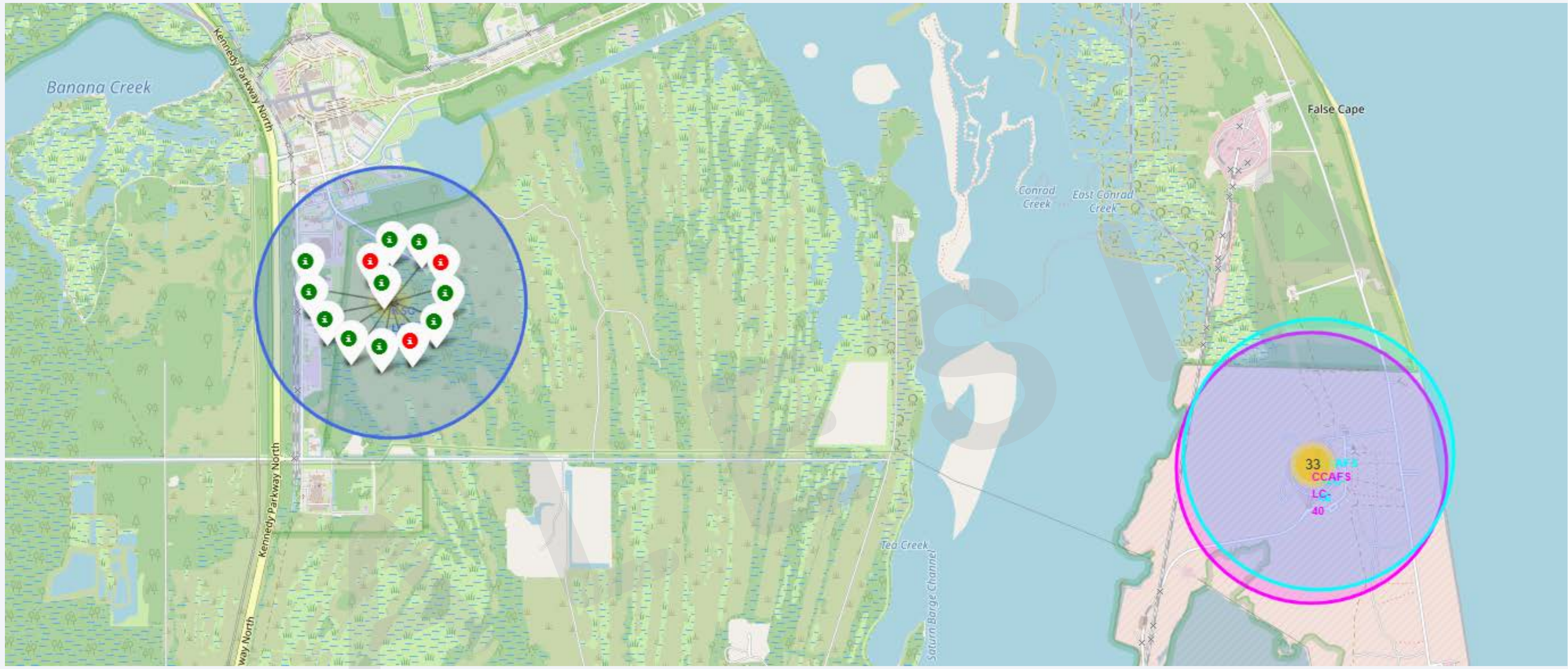
Launch Sites Proximities Analysis

Space X launch sites on a map



- All launch sites are in very close proximity to the ocean coast and to the Equator line.

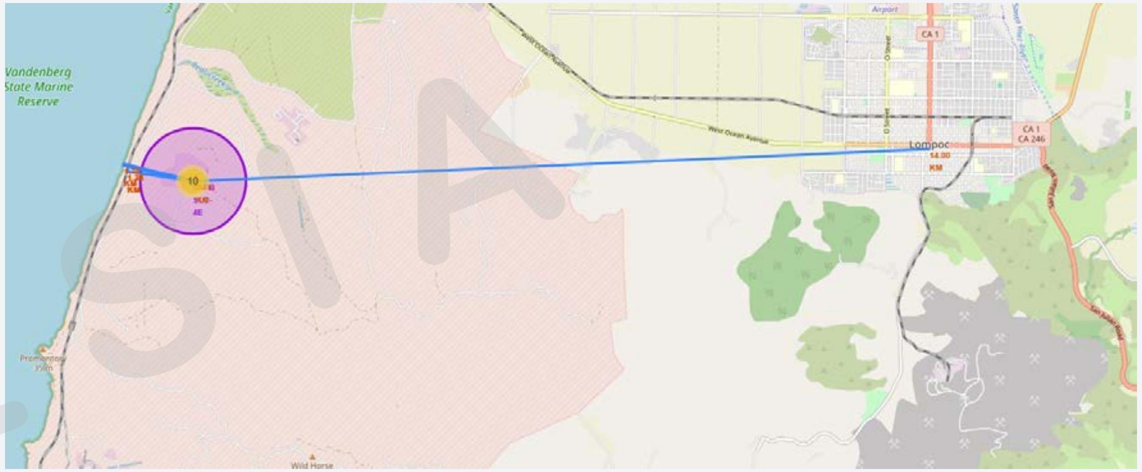
Launch outcomes for each site



- It was created markers for all launch records.

If a launch was successful, it was used a green marker and if a launch was failed, it was used a red marker.

The proximities of launch sites



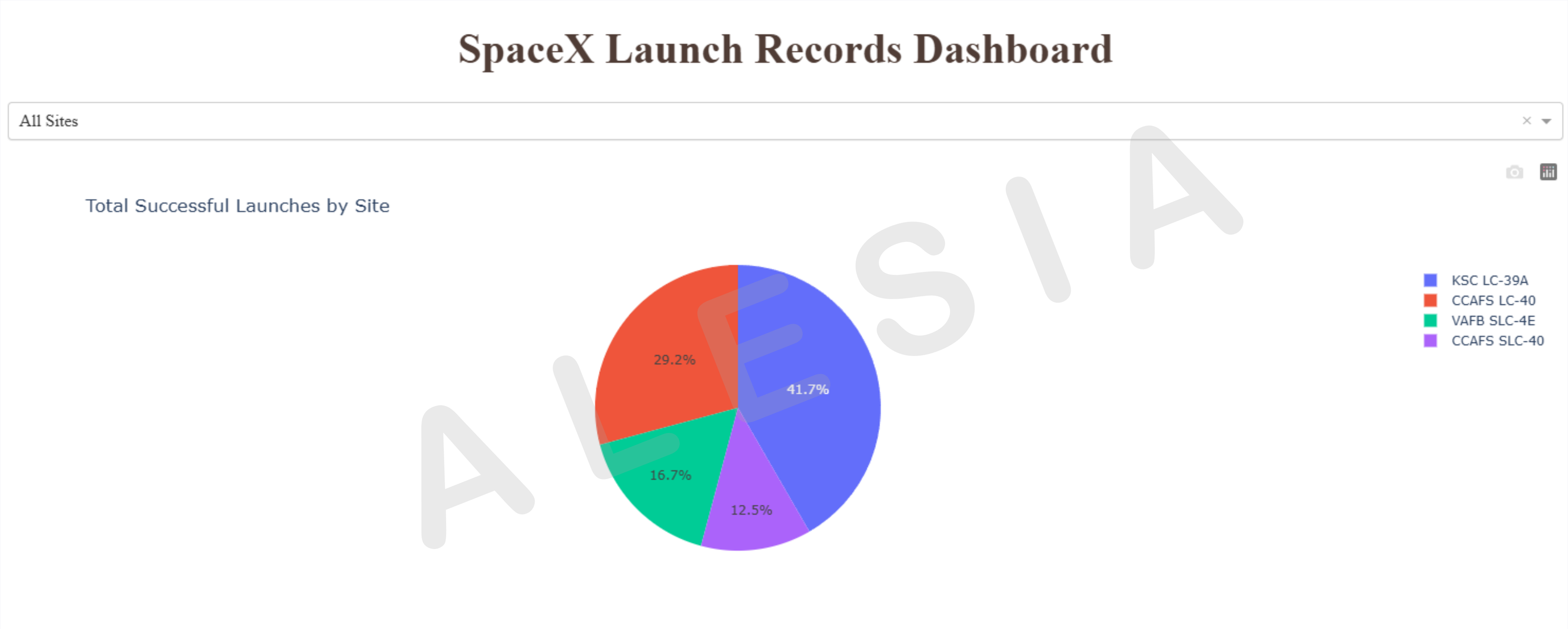
- Launch sites are in close proximity to railways, highways and coastline.

- Launch sites keep certain distance away from cities.

Section 4

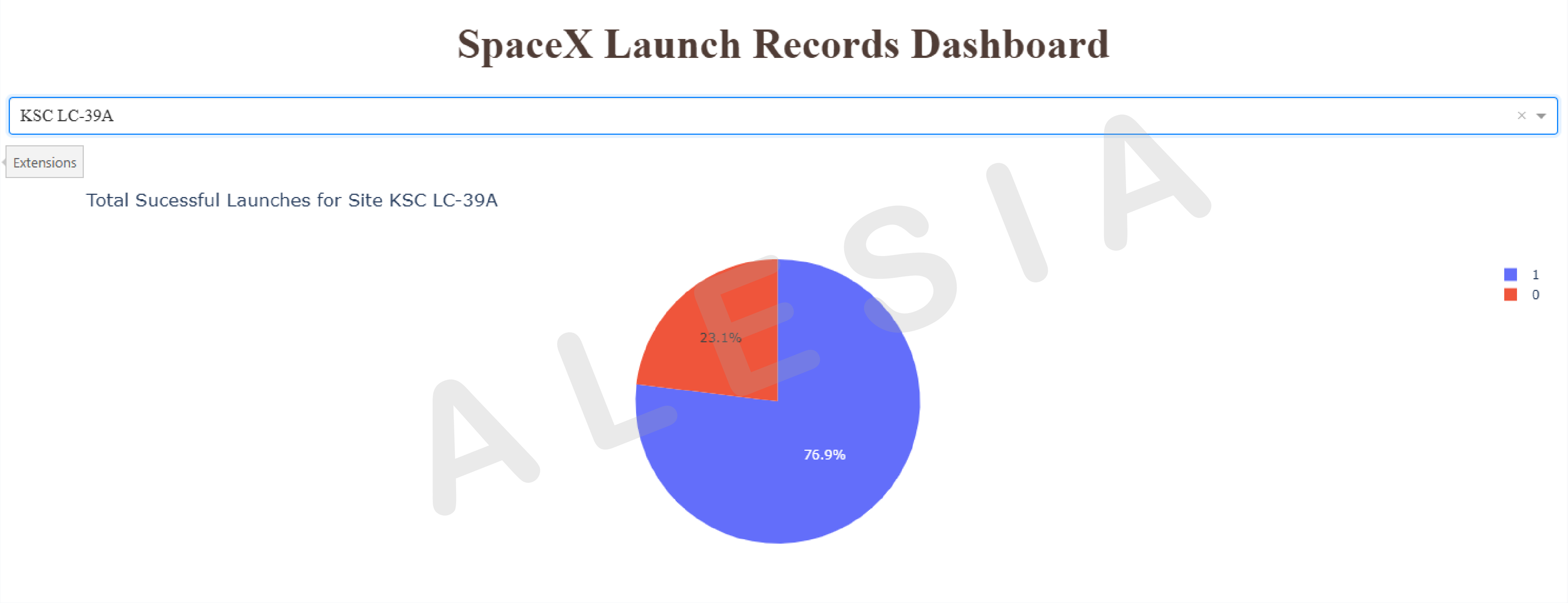
Build a Dashboard with Plotly Dash

Total successful launches by site



- KSC LC-39A launch site has the highest success rate.

Launch site with the highest success rate



- KSC LC-39A launch site has the highest launch success rate.

Payload vs. Launch outcome

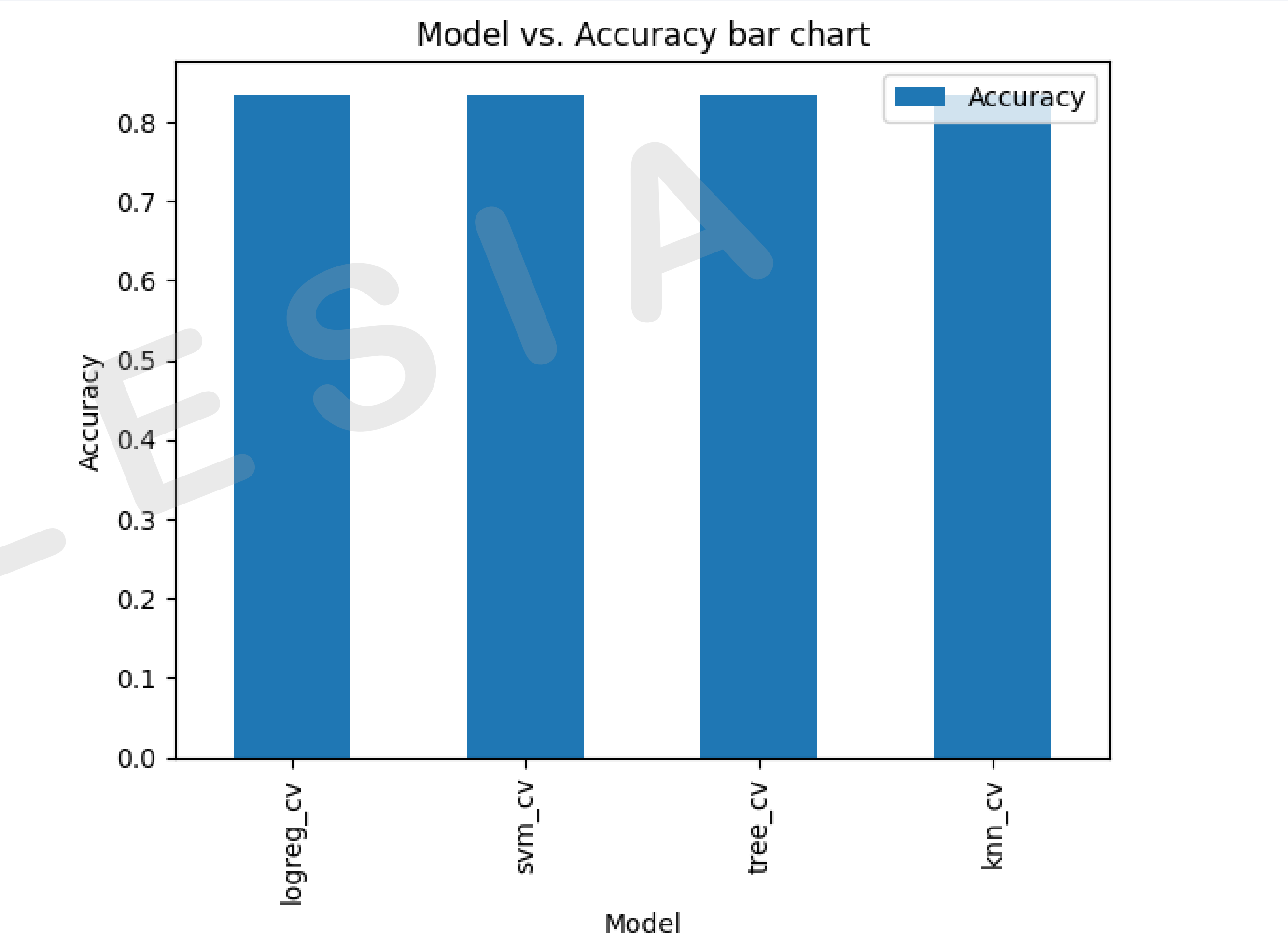
- FT booster version and payload mass between 2000 and 4000 kg have the largest success rate.

Section 5

Predictive Analysis (Classification)

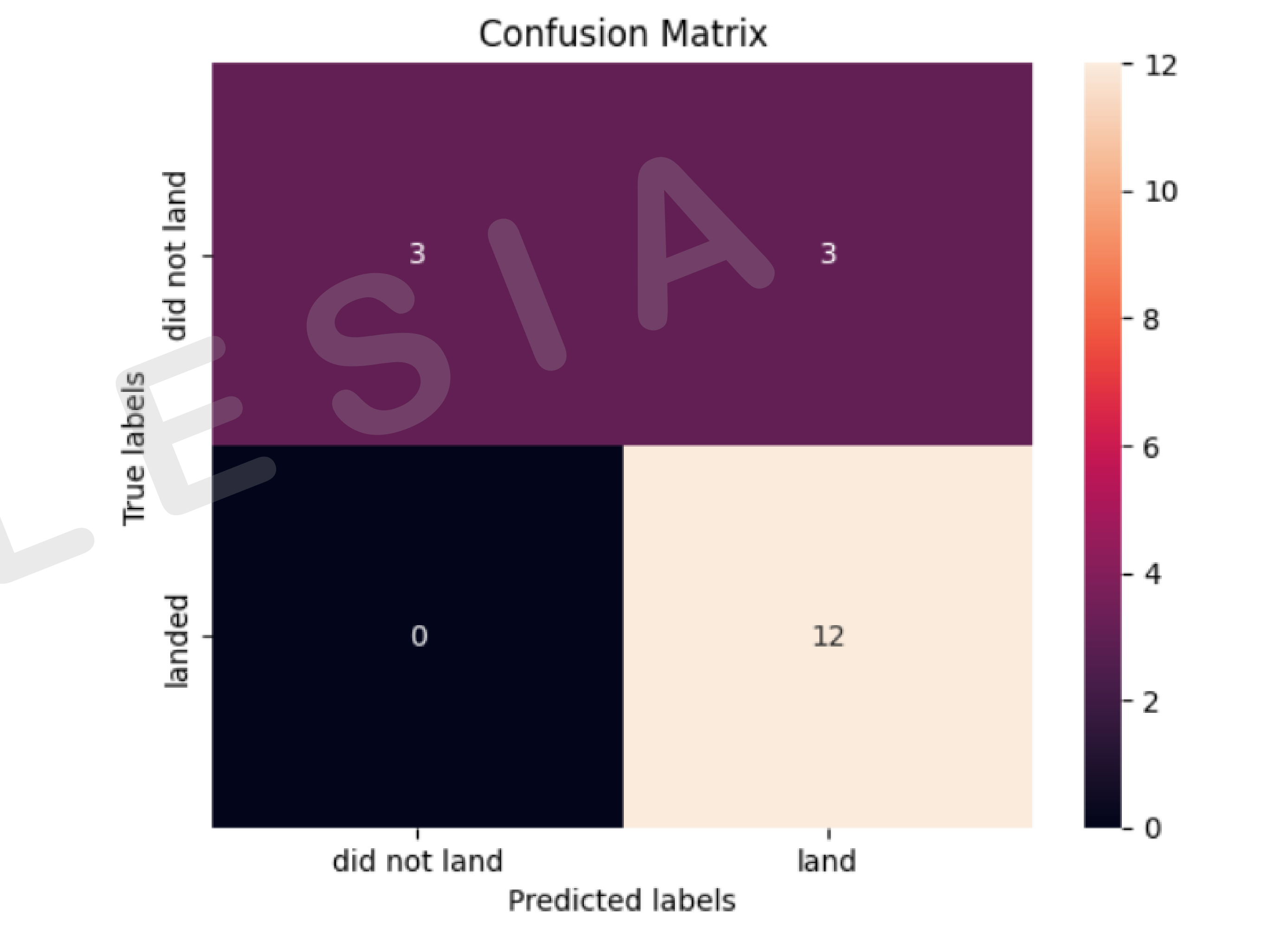
Classification Accuracy

- All models have the same classification accuracy



Confusion Matrix

- All models have the same classification accuracy.
- Therefore, the confusion matrices are the same.



Conclusions

- There are only four unique names of launch sites of Space X.
- All launch sites are in very close proximity to the ocean coast and to the Equator line, railways, highways and coastline and keep certain distance away from cities.
- The total number of successful mission outcomes significantly exceed the number of failure missions.
- KSC LC-39A launch site, rocket configuration Falcon 9 FT and payload mass between 2000 and 4000 kg have the highest success rate.
- .The largest number of launches were made from the site CCAFS SLC-40. There have been many successful and many failed missions here.
- Least number of launches were made from the site VAFB-SLC. Most missions were successful.
- The first stage often returns successfully with more massive payloads, as the flight number increases, with heavy payloads are more for Polar, LEO and ISS.
- ES-L1, GEO, HEO and SSO types of orbits have the highest success rates.
- All predictive models have the same classification accuracy.

Appendix

Links

- https://api.spacexdata.com/v4/launches/past
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json
- https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_1.csv
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_2.csv
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/labs/module_2/data/Spacex.csv
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/labs/module_2/data/Spacex.csv
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DV0101EN-SkillsNetwork/labs/v4/DV0101EN-Exercise-Generating-Maps-in-Python.ipynb
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/spacex_launch_geo.csv
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_2.csv
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_3.csv
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/spacex_launch_dash.csv
- https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/t4-Vy4iOU19i8y6E3Px_ww/spacex-dash-app.py
- https://dash.plotly.com/dash-core-components/dropdown
- https://dash.plotly.com/dash-core-components/rangeslider
- https://plotly.com/python/pie-charts/
- https://plotly.com/python/line-and-scatter/

DASH – Cloud IDE

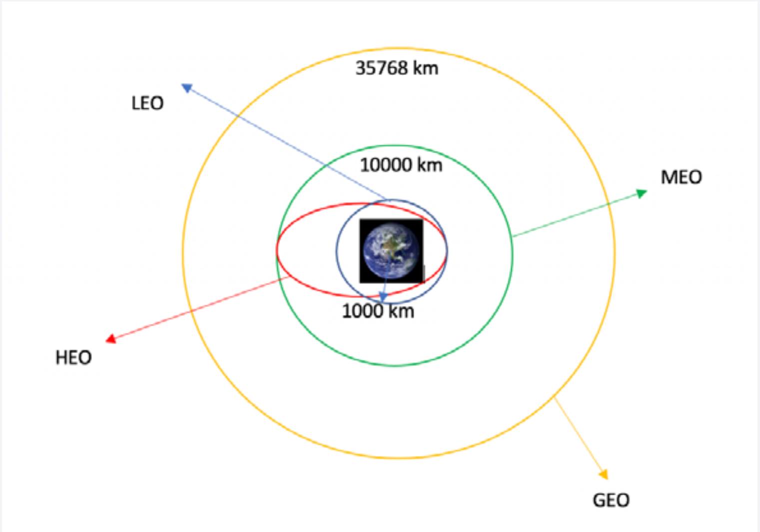
- https://theiadockernext-1-labs-prod-theiak8s-4-tor01.labs.cognitiveclass.ai/user/alesyakooche/#
- https://theiadockernext-1-labs-prod-theiak8s-4-tor01.labs.cognitiveclass.ai/user/alesyakooche/#
- https://theiadockernext-1-labs-prod-theiak8s-4-tor01.labs.cognitiveclass.ai/user/alesyakooche/#

Trust Notebook

View-Activate Command Palette-Trust Notebook

LABS

- Hands-on Lab : String Patterns, Sorting and Grouping
- Hands-on Lab: Built-in functions
- Hands-on Lab : Sub-queries and Nested SELECT Statements
- Hands-on Tutorial: Accessing Databases with SQL magic
- Hands-on Lab: Analyzing a real World Data Set



Thank you!