IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

A presentation of the IBM DS0720EN
Data Science and Machine Learning Capstone Project
By Jan Strømmen
April 2024

Outline of this presentation

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

Executive Summary

- In this capstone project the SpaceX flight data have been analyzed using techniques such as data wrangling, EDA (Exploratory Data Analysis) and SQL data base queries for extracting and visualization of the data.
- Interactive maps using Folium have been created to better analyze the geographic location of launch sites with regards to suitability of each site with safety particularly in mind.
- An interactive Dashboard using Plotly Dash has been built to further improve the data analysis.
- Predictive analysis employing ML classification techniques were then used to find the best predictor of future SpaceX flights outcomes.
- A key result is that the results of SpaceX flights have been steadily improving over time and currently the chance of success is probably in the range of 80% to 90%.
- The analysis clearly demonstrates how the technology has matured into something which is now viable in terms of both reliability and commerciality.

Introduction

- SpaceX has developed a rocket system where the first stage of the rocket will return to earth and make a controlled landing on a designated site.
- The potential reuse of the first stage entails a significant saving of approximately $100 million but so far there has been a number of unsuccessful returns.
- It should be noted though, that the success rate has been steadily increasing over the years due to the maturation of this new technology.
- What is to be addressed here is to find a way to predict the chance of a successful return of the first stage for future launces.

Section 1

Methodology

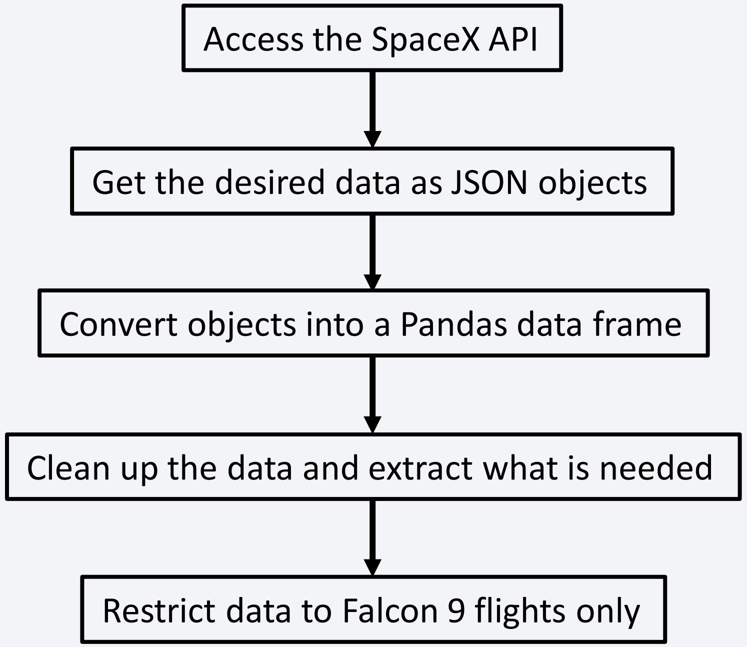
Methodology

Executive Summary

- Data collection methodology
  - Two methods were employed for getting hold of the data required; Request the data from the SpaceX API and secondly, web-scraping the data from a Wikipedia page.
- Perform data wrangling
  - Wrangling includes EDA, checking for missing data and deriving a training label for mission success or not.
- Perform EDA using visualization and SQL
  - This step includes displaying various plots and using SQL queries to understand the data better.
- Perform interactive visual analytics using Folium and Plotly Dash
  - Folium is used to create maps for launch site analysis and Dash to provide an interactive analysis tool of the data.
- Perform predictive analysis using classification models
  - Various ML models (Logistic regression, SVM, Decision Tree, KNN) were trained and tested on normalized data in order to find the best prediction method for future flight outcomes.

Data Collection

- Some of the key data required for this project are launch results such as rocket used, payload, launch and landing specifications and specifically the outcome of landing.
- In order to get hold of the required data, the SpaceX launch data were accessed from the SpaceX REST API which has this URL: https://api.spacexdata.com/v4/launches/past
- A get request followed by the .json() method provide a list of json objects which first is converted into a flat table and subsequently into a Pandas data frame.
- Further work involved cleaning up and extracting data for the features rocket, payloads, launchpad, cores, flight_number and date_utc.
- All data were then combined into a dictionary which were then converted to a Pandas data frame.
- The data frame were then restricted to contain only Falcon 9 flights.



Data Collection – SpaceX API

On the right is shown the major coding steps to get the desired data.

In addition, four functions have been used for calling the API to extract information using ID numbers in the launch data:

getBoosterVersion gets booster name from the rocket column

getLaunchSite gets launch site, longitude and latitude from the launchpad column

getPayloadData gets payload mass and orbit from the payload column

getCoreData gets landing outcome, type of the landing, number of flights with that core, whether gridfins were used, wheter the core is reused, wheter legs were used, the landing pad used, the block of the core which is a number used to separate version of cores, the number of times this specific core has been reused, and the serial of the core, from the cores column

Get the rocket launch data from the SpaceX API and store in the *response* variable

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

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

Convert the json *response* into the Pandas data frame *data*

```
# Use json_normalize meethod to convert the json result into a dataframe
data=pd.json_normalize(response.json())
```

Use the defined functions to get key data:
getBoosterVersion, getLaunchSite, getPayloadData, getCoreData

Combine the data into a dictionary and then into a Pandas data frame

```
launch_dict = {'FlightNumber': list(data['flight_number']),
'Date': list(data['date']),
'BoosterVersion':BoosterVersion,
'PayloadMass':PayloadMass,
'Orbit':Orbit,
'LaunchSite':LaunchSite,
'Outcome':Outcome,
'Flights':Flights,
'GridFins':GridFins,
'Reused':Reused,
'Legs':Legs,
'LandingPad':LandingPad,
'Block':Block,
'ReusedCount':ReusedCount,
'Serial':Serial,
'Longitude': Longitude,
'Latitude': Latitude}

# Create a data from launch_dict
data_df = pd.DataFrame.from_dict(launch_dict)
```

GitHub URL:
https://github.com/JanStrommen/CapstoneProject/blob/master/jupyter-labs-spacex-data-collection-api_JanS.ipynb

Data Collection - Scraping

Here are shown the major coding steps to get the desired data:

One function has been used to process a web scraped HTML table:

extract_column_from_header returns the column name

A lengthy code has been used to fill up the dictionary with data prior to conversion to a Pandas data frame.

Get the rocket launch data from the Falcon9 Launch HTML page and store in the *html_data* variable

```
html_data = requests.get(static_url)
```

Create a Beautiful Soup object

```
soup = BeautifulSoup(html_data.text)
```

Create list of all table elements from the BeautifulSoup object

```
html_tables = soup.find_all('table')
```

Continued on the right

Extract column names from the <th> elements

```
for element in first_launch_table.find_all('th'):
    name = extract_column_from_header(element)
    if name is not None and len(name) > 0:
        column_names.append(name)
```

Create dictionary with keys from *column_names*

```
launch_dict= dict.fromkeys(column_names)

# Remove an irrelvant column
del launch_dict['Date and time ( )']

# Let's initial the launch_dict with each value to be an empty list
launch_dict['Flight No.'] = []
launch_dict['Launch site'] = []
launch_dict['Payload'] = []
launch_dict['Payload mass'] = []
launch_dict['Orbit'] = []
launch_dict['Customer'] = []
launch_dict['Launch outcome'] = []
# Added some new columns
launch_dict['Version Booster']=[]
launch_dict['Booster landing']=[]
launch_dict['Date']=[]
launch_dict['Time']=[]
```

Fill dictionary with launch records from table rows and convert it to a Pandas frame

```
df=pd.DataFrame(launch_dict)
#data_df = pd.DataFrame.from_dict(launch_dict)
```

GitHub URL:
https://github.com/JanStrommen/CapstoneProject/blob/master/jupyter-labs-webscraping_JanS.ipynb

Data Wrangling

On the right is shown the major coding steps to obtain the desired data.

The data wrangling process:

Read the output data from the Data Collection step(s).

Identify missing values in attributes (28% missing for *LandingPad*).

Calculate number of launces on each *LaunchSite*.

Calculate the number and occurrence of each *orbit*.

Calculate the number and occurrence of mission outcome per *orbit* type.

Create a landing outcome label from the *Outcome* column and calculate the success rate for SpaceX (= 0.66).

Read data

```
df=pd.read_csv(URL)
df.isnull().sum()/df.shape[0]*100
```

```
# Apply value_counts() on column LaunchSite
df['LaunchSite'].value_counts()
```

```
CCAFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
```

```
# Apply value_counts on Orbit column
df['Orbit'].value_counts()
```

```
GTO      27
ISS      21
VLEO     14
PO        9
LEO       7
SSO       5
MEO       3
ES-L1     1
HEO       1
SO        1
GEO       1
```

```
# landing_outcomes = values on Outcome column
landing_outcomes = df['Outcome'].value_counts()
landing_outcomes
```

```
True ASDS      41
None None      19
True RTLS      14
False ASDS      6
True Ocean      5
False Ocean     2
None ASDS       2
False RTLS      1
```

```
for i,outcome in enumerate(landing_outcomes.keys()):
    print(i,outcome)
```

```
0 True ASDS
1 None None
2 True RTLS
3 False ASDS
4 True Ocean
5 False Ocean
6 None ASDS
7 False RTLS
```

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

```
# landing_class = 0 if bad_outcome
# landing_class = 1 otherwise
landing_class = [0 if landing in bad_outcomes else 1 for landing in df['Outcome']]
```

```
df['Class']=landing_class
```

```
df["Class"].mean()
```

```
0.6666666666666666
```

GitHub URL:
https://github.com/JanStrommen/CapstoneProject/blob/master/labs-jupyter-spacex-data_wrangling_jupyterlite_JanS.ipynb

EDA with Data Visualization

Charts used and purpose

- Scatter chart to show flight number vs launch site and success/failure result. This to display site usage and developments over time.
- Scatter chart to show payload vs launch site and success/failure result. This to display site usage and any effects related to payload size.
- Bar chart to show success rates of the numerous orbit types used by SpaceX. Conclusions relating to orbit type of interest.
- Scatter chart to show flight number vs orbit type and success/failure result. This to display orbit type and developments over time.
- Scatter chart to show payload vs orbit type and success/failure result. This to display orbit type and any effects related to orbit type.
- Line chart to show annual success rates development over time.

GitHub URL:
https://github.com/JanStrommen/CapstoneProject/blob/master/jupyter-labs-eda-dataviz_JanS.ipynb

EDA with SQL

Recurring issues with trying to use IBM Cloud for the SQL processing resulted in the decision to employ Sqlite3 in the normal and much preferred 'Skills Network Labs' instead.
The following Sqlite queries were performed:

- Create list of all the launch sites used by SpaceX
- Create list of first five launch sites with name beginning with 'KSC'
- Calculate total payload mass of boosters working for customer 'NASA (CRS)'
- Calculate total payload mass of boosters working for all NASA customers
- Calculate average payload mass for F9 v1.1 boosters
- Calculate the date of first successful booster landing on drone ship
- Create list of boosters with successful ground pad landing and payloads between 4000kg and 6000kg
- Calculate number of successful and failure mission outcomes
- Create list of booster versions to have carried the maximum payload
- Create list of records displaying month, successful ground pad landing and launch site for year 2017
- Create list in descending order of successful landings between 2010-06-04 and 2017-03-20

GitHub URL:
https://github.com/JanStrommen/CapstoneProject/blob/master/jupyter-labs-eda-sql-edx_sqllite-JanS.ipynb

Build an Interactive Map with Folium

Maps made and purpose

- Two maps made to display locations of the launch sites used.
- Four maps to display the launch outcomes for each launch site over time. Excellent maps giving a brilliant overview.
- For the California launch site a map is made with markers and lines to highlight distances from site to nearest railroad, major road and closest population center.
- This to highlight geographic land location and proximity to ocean and infrastructure in potential danger in case of failing launches.

GitHub URL:
https://github.com/JanStrommen/CapstoneProject/blob/master/lab_jupyter_launch_site_location.jupyterlite_JanS.ipynb

Build a Dashboard with Plotly Dash

Plots made and purpose

- The dash application is capable of displaying a pie chart with the distribution of successful launches across all sites, alternatively the split of success and failure for one individual site.
- In addition, a scatter chart will show site launches as a function of payload with the booster version employed.
- The payload range to use in the plots can be controlled by an interactive range slider.
- This makes it easy to compare results for various launch sites and how payloads and booster versions used correlates with the outcomes.

GitHub URL:
https://github.com/JanStrommen/CapstoneProject/blob/master/spacex_dash_app_JanS.py

Predictive Analysis (Classification)

Classification models

- Predictive analysis classification has been attempted with four different methodologies.
- Logistic Regression, Support Vector Machine, Decision Tree and K Nearest Neighbours.
- 80% of data was used for training and 20% for testing.
- The Decision tree model gave a high training score but the lowest score on testing.
- The other three models performed equally well on training and testing with a testing score of 0.833.

GitHub URL:
https://github.com/JanStrommen/CapstoneProject/blob/master/SpaceX_Machine_Learning_Prediction_Part_5_JanS.ipynb

Results

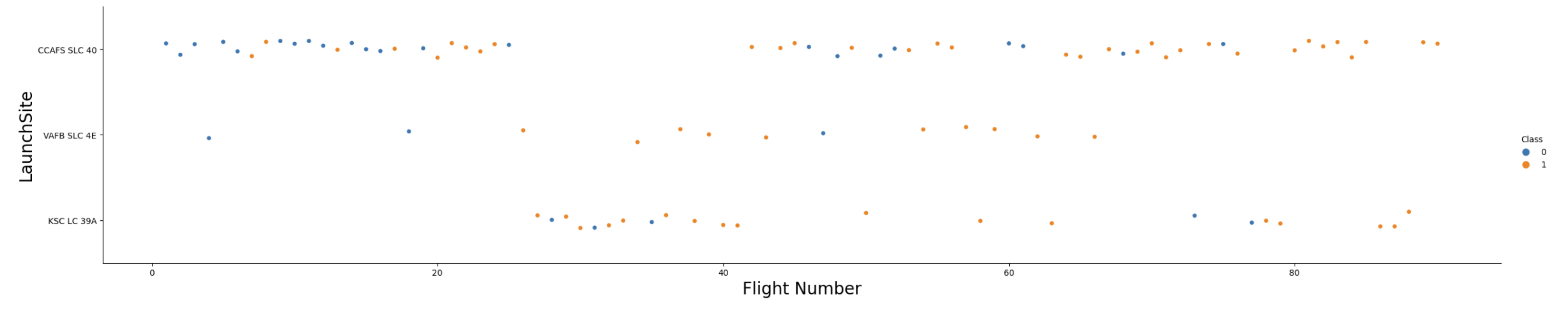
Summary of major results

- A key result is that the results of SpaceX flights have been steadily improving over time and currently the chance of success is probably in the range of 80% to 90%.
- The SSO and VLEO orbits are supported by statistics for this range.
- It is questionable to assign one specific launch site as superior to the others but site KSC LC-39A do stand out with a success score of 76.9%.
- The Logistic Regression, Support Vector Machine and K Nearest Neighbours classifications seem to give better results than the Decision Tree classification but this could be influenced by chosen parameters and size of data set.

Section 2

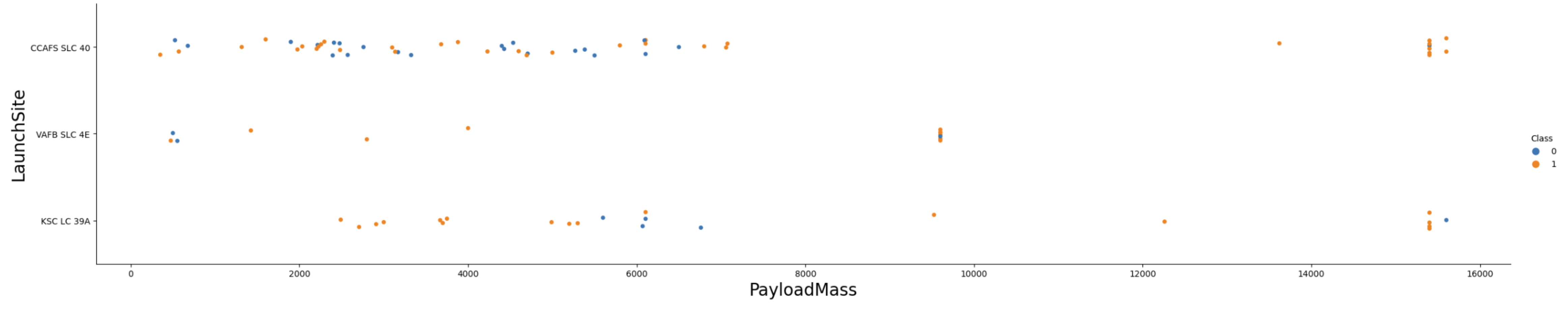
Insights drawn from EDA

Flight Number vs. Launch Site



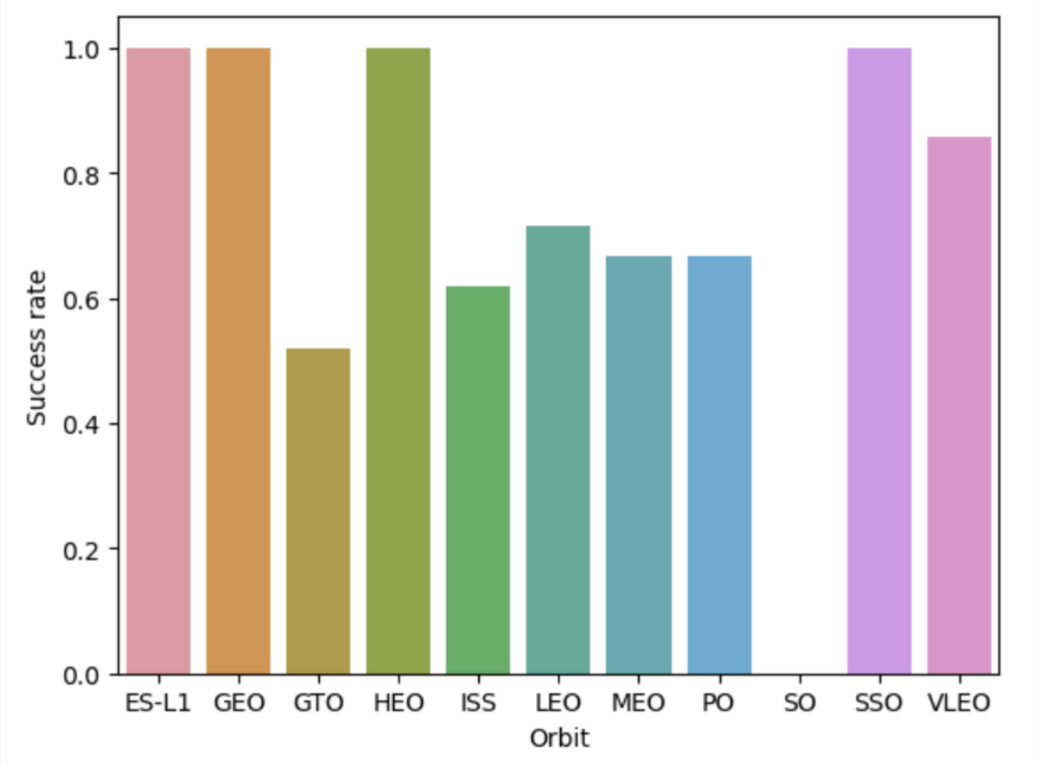
- The most used launch site is CCAFS SLC 40 with 55 out of the 90 launches.
- Note though the pause between flights 25 and 42.
- Overall, there is a clear trend of more successful booster landings with increasing flight numbers.
- The success rates appear higher for the launch sites VAFB SLC 4E and KSC LC 39A, however, if the first 25 Flight numbers are excluded the differences in success seem small.
- An indication of a learning enterprise which improves over time.

Payload vs. Launch Site



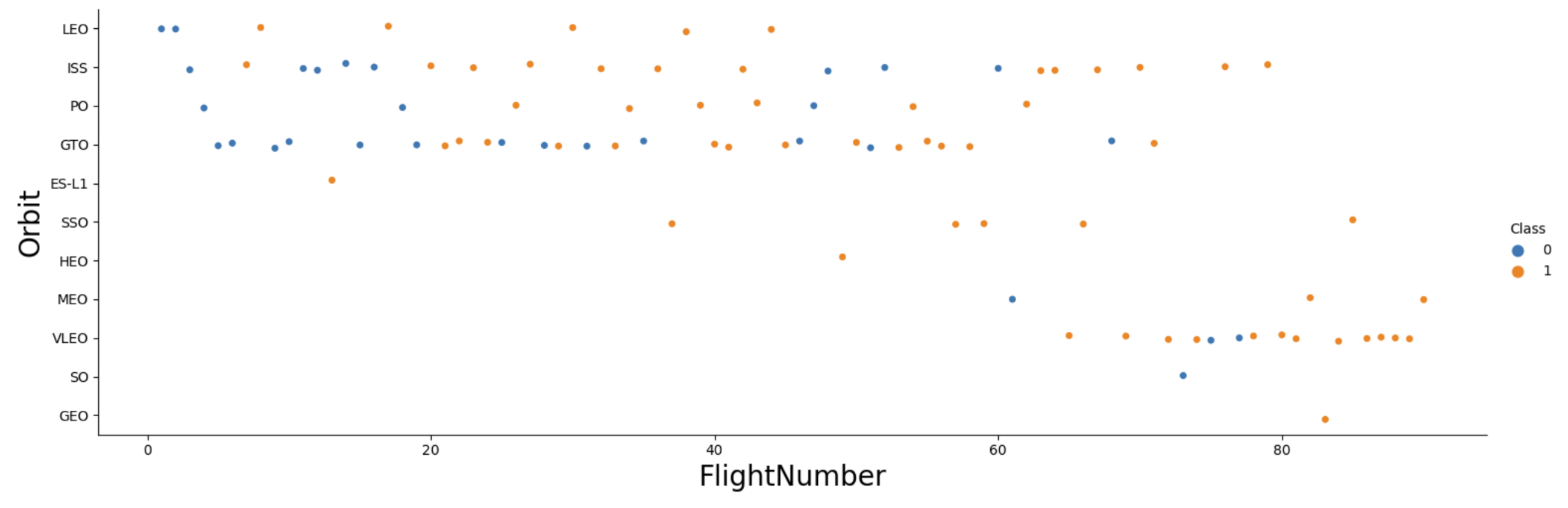
- Launch site VAFB SLC 4E appears to have an upper limit on payload mass of 9600kg and a high success rate.
- The other two launch sites have a maximum payload mass of 15600kg.
- Launch site CCAFS SLC 40 has a mixed bag of success and failures, regardless of payload mass.
- This last point is actually caused by the high number of failures during the first 25 launces.
- A plot of Payload mass vs flight number shows that for flight numbers 1-16, 12 failed possibly due to increased payload.
- Maybe a low payload flight was easier to control in the early days of SpaceX?

Success Rate vs. Orbit Type



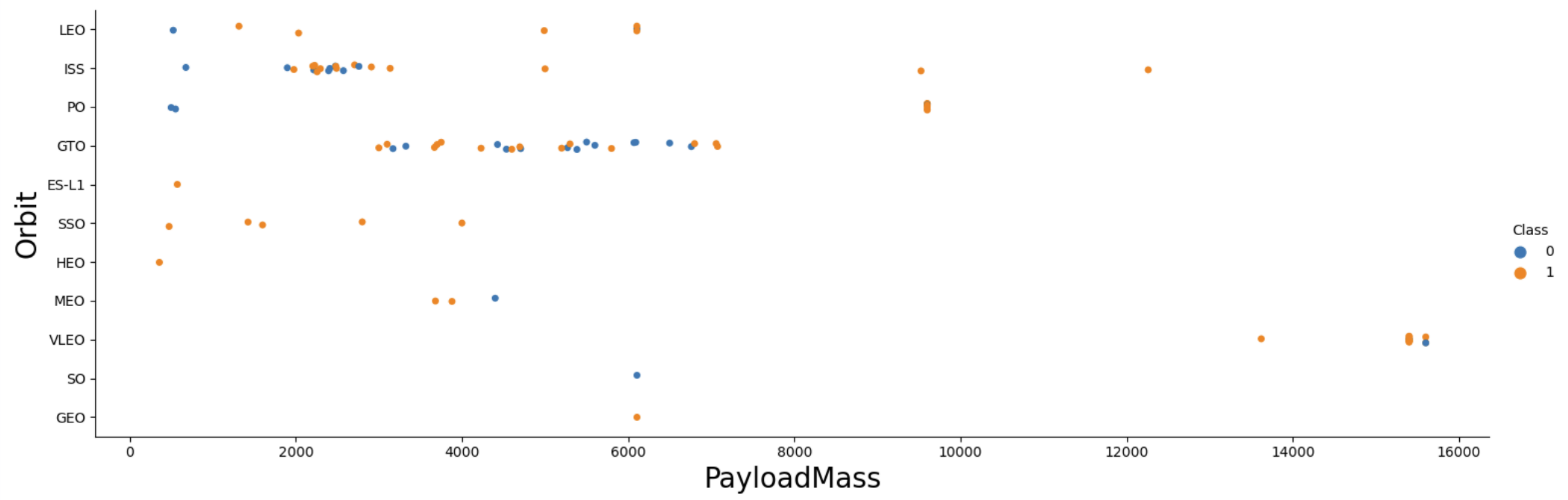
- Three of the orbit types with 100% success rate (ES-L1, GEO and HEO) are high altitude orbits, however, each of them has only one single occurrence.
- Orbit type SO has just one single occurrence, a failure.
- Orbit type SSO has also 100% success rate and with a total of 5 flights.
- The other orbit types are more statistically meaningful with a higher number of occurrences but also with many low flight numbers.
- The combined SSO and the VLEO orbit types have a total of 19 occurrences over the period 2017-2020.
- I therefore believe the SSO and VLEO orbit types are best representing today's chance of success for a SpaceX project.

Flight Number vs. Orbit Type



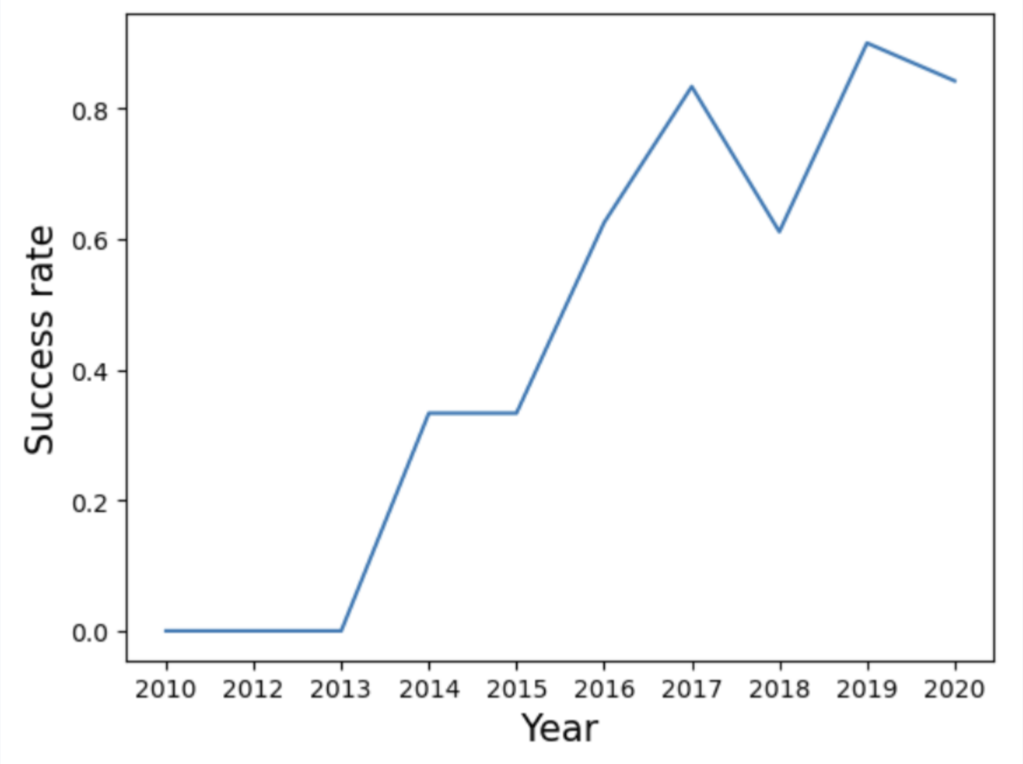
- All orbit types seems to demonstrate that success rate goes up with number of flights
- The exception is orbit type GTO where successes and failures are evenly distributed over time
- There is also a clear shift in orbit types towards VLEO type over the last few years
- This could be related to the deployment of thousands of Starlink satellites

Payload vs. Orbit Type



- ISS orbits have most flights within a narrow payload range.
- To a lesser extent that also applies to GTO orbits.
- VLEO orbits all have high payloads.

Launch Success Yearly Trend



- The launch success rate shows a steady increase from 2013
- Exception is 2017 where the success rate decreased.
- It is reasonable to expect that the technology now has matured so much that the near future success rates will be between 0.8 and 0.9.

All Launch Site Names

Task 1

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

```
%sql select distinct Launch_Site from SPACEXTBL
```

```
 * sqlite:///my_data1.db
Done.
```

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

This select distinct statement returns only distinct (different) Launch_Site values.

Launch Site Names Begin with 'KSC'

Task 2

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

```
%sql select * from SPACEXTBL where Launch_Site like 'KSC%' LIMIT 5
```

```
 * sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-03-16	6:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attempt
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-05-01	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)
2017-05-15	23:21:00	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070	GTO	Inmarsat	Success	No attempt

This select ….. where statement returns records only with launch site values beginning with KSC and up to a total of 5 records.

Total Payload Mass

Task 3

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

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where Customer= 'NASA (CRS)'
```

```
 * sqlite:///my_data1.db
Done.
```

sum(PAYLOAD_MASS__KG_)
45596

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where Customer like 'NASA%'
```

```
 * sqlite:///my_data1.db
Done.
```

sum(PAYLOAD_MASS__KG_)
99980

The select sum first statement returns the total payload mass carried for customer 'NASA (CRS)'.
The select sum second statement returns the total payload mass carried for all NASA customers.

Average Payload Mass by F9 v1.1

Task 4

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

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where Booster_Version like 'F9 v1.1%'
```

```
 * sqlite:///my_data1.db
Done.
```

avg(PAYLOAD_MASS__KG_)
2534.6666666666665

The select avg statement returns the average payload mass for booster 'F9 v1.1%'.

First Successful Ground Landing Date

Task 5

List the date where the succesful landing outcome in drone ship was acheived.

Hint:Use min function

```
%sql select min(Date) from SPACEXTBL where "Landing_Outcome" = "Success (drone ship)"
```

```
 * sqlite:///my_data1.db
Done.
```

min(Date)
2016-04-08

The select min(Date) statement returns the date of the first successful landing on a drone ship.

Successful Drone Ship Landing with Payload between 4000 and 6000

Task 6

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

```
%%sql

select Booster_Version from SPACEXTBL
where "Landing_Outcome" = "Success (ground pad)"
    and PAYLOAD_MASS__KG_ > 4000
    and PAYLOAD_MASS__KG_ < 6000
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version
F9 FT B1032.1
F9 B4 B1040.1
F9 B4 B1043.1

This select … where statement creates a list of boosters with a successful ground pad landing and a set paylod mass range.

Total Number of Successful and Failure Mission Outcomes

Task 7

List the total number of successful and failure mission outcomes

```
%%sql
select count(*) from SPACEXTBL
where "Mission_Outcome" like "Success%"
```

```
 * sqlite:///my_data1.db
Done.
```

count(*)
100

```
%%sql
select count(*) from SPACEXTBL
where "Mission_Outcome" like "Failure%"
```

```
 * sqlite:///my_data1.db
Done.
```

count(*)
1

These select… where statements return the total number of successful and failure missions respectively.

Boosters Carried Maximum Payload

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
%%sql

select Booster_Version from SPACEXTBL
where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

```
 * sqlite:///my_data1.db
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

This select .... where statement creates a list of boosters that were carrying the maximum payload mass (of 15600 kg), hence the high number of identical payload values.

2015 Launch Records

Task 9

List the records which will display the month names, succesful landing_outcomes in ground pad ,booster versions, launch_site for the months in year 2017

Note: SQLLite does not support monthnames. So you need to use substr(Date,6,2) for month, substr(Date,9,2) for date, substr(Date,0,5),='2017' for year.

```
%%sql

select substr(Date, 6, 2) as Month, Landing_Outcome, Booster_Version, Launch_Site from SPACEXTBL
where substr(Date, 0, 5) = '2017' and "Landing_Outcome" = "Success (ground pad)"
```

```
 * sqlite:///my_data1.db
Done.
```

Month	Landing_Outcome	Booster_Version	Launch_Site
02	Success (ground pad)	F9 FT B1031.1	KSC LC-39A
05	Success (ground pad)	F9 FT B1032.1	KSC LC-39A
06	Success (ground pad)	F9 FT B1035.1	KSC LC-39A
08	Success (ground pad)	F9 B4 B1039.1	KSC LC-39A
09	Success (ground pad)	F9 B4 B1040.1	KSC LC-39A
12	Success (ground pad)	F9 FT B1035.2	CCAFS SLC-40

This select… where statement returns the month, landing outcome, booster version and launch site for launches with a successful ground pad landing in 2017.

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

Task 10

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 landings from SPACEXTBL
where Date >= "2010-06-04" and Date <= "2017-03-20"
group by Landing_Outcome
order by landings desc
```

```
 * sqlite:///my_data1.db
Done.
```

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

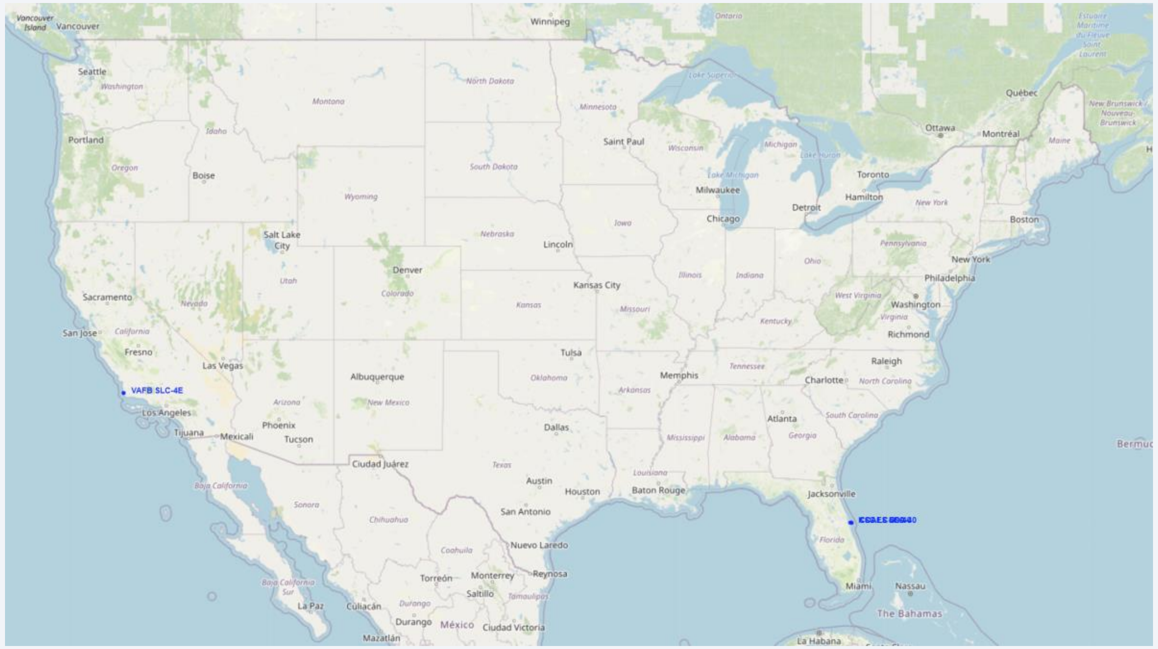
This select… where statement calculates the number of Landing_Outcomes over the specified time window and lists the landing outcomes and the numbers in descending order.

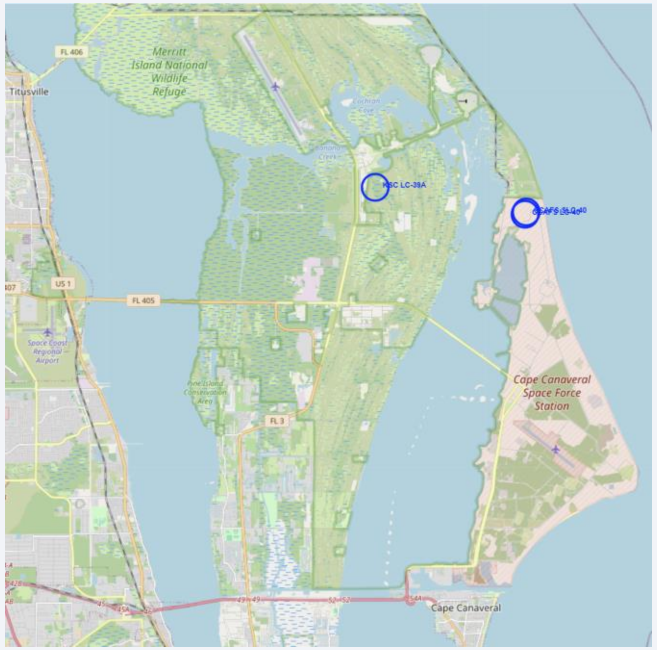
Section 3

Launch Sites Proximities Analysis

All SpaceX launch sites

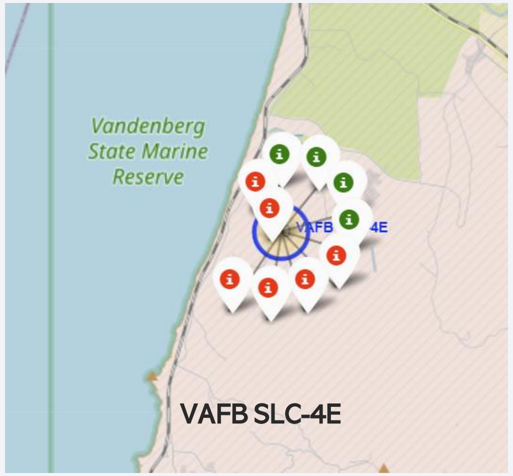
- Left map below shows one launch site is located in California whereas three sites are closely located in Florida.
- The California site , VAFB SLC-4E, is situated near the Vandenberg Air Force Base.
- The three Florida sites, KSC LC-39A, CCAFS LC-40 and CCAFS SLC-40 are situated at the Cape Canaveral Space Force Station.
- All sites are located close to oceans and therefore well suited for launches over open water rather than over populated land areas.
- The long distance to the equator is possibly a disadvantage for some orbit types.

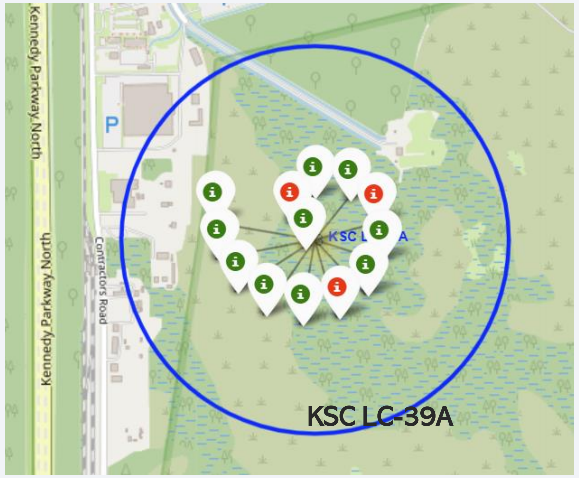


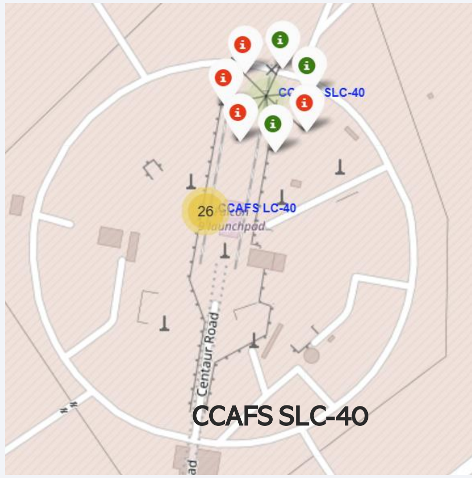


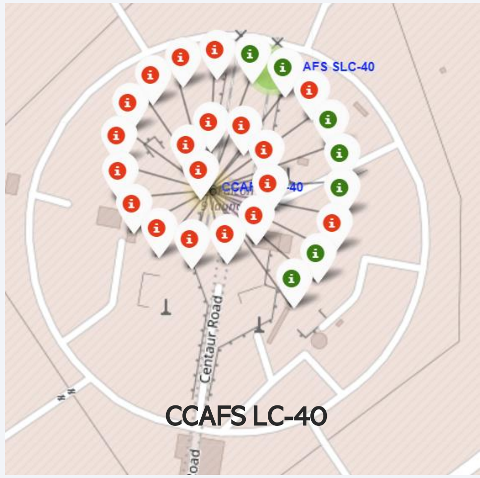
Visual display of all launch outcomes

- The four maps below show the launch outcomes for each of the four sites over the period April 2010 to June 2018.
- A green icon represents a success and a red icon a failure.
- KSC LC-39A launches appear to be the more successful with 10 out of 13 launches, however, all were done as recently as after 2016.
- CCAFS LC-40 launches spans the years 2010-2016 and displays clearly the increasing success rate over time.
- Given the above it may be stated that the launch site KSC LC-39A has performed better than the others.



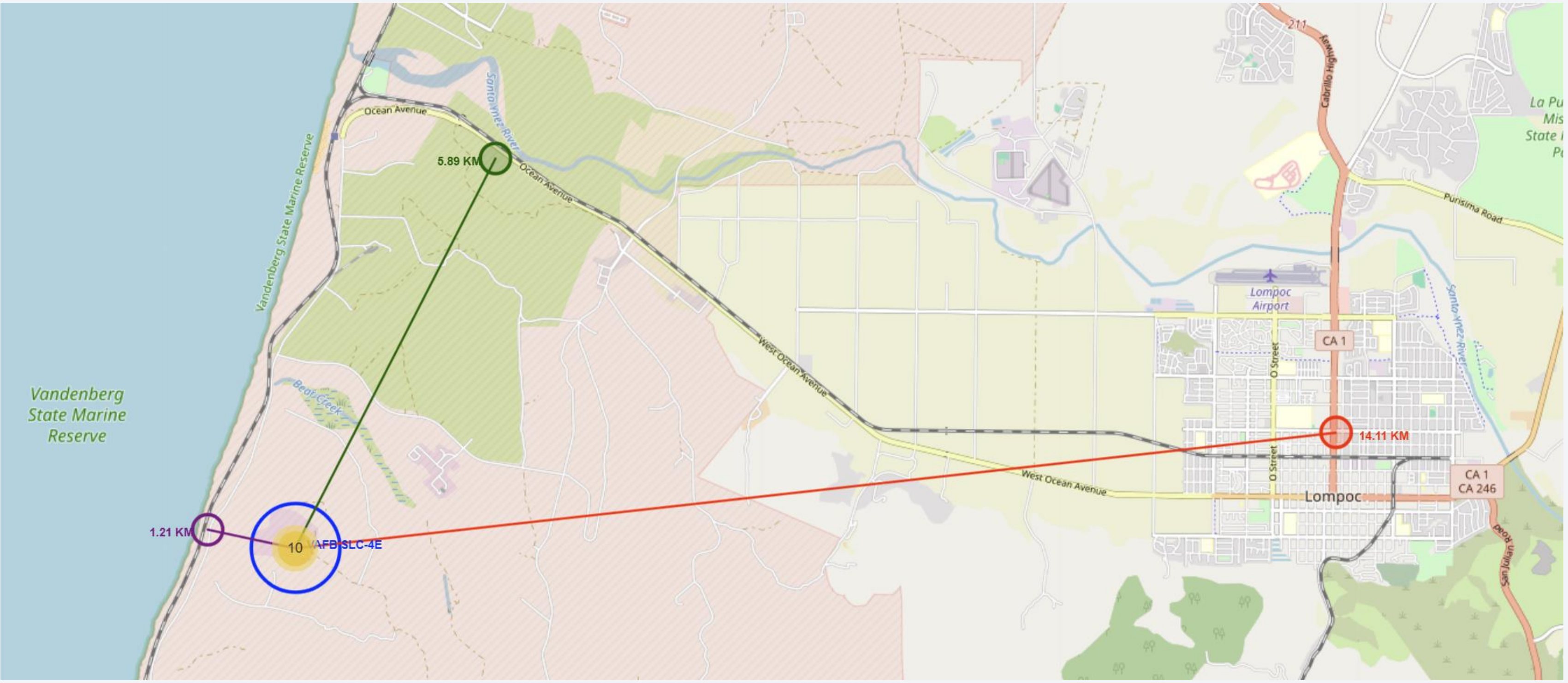






California site with surroundings

- The map below shows distances from the VAFB SLC-4E launch site and to the nearest coastline/railroad, highway and city.
- Distance to the coastline and railroad is about 1.2 km.
- Distance to the nearest highway is about 5.9 km.
- Distance to the nearest city (Lompoc) is about 14.1 km.
- The launch site has adequate safety distances to highways and populated areas and traffic on a railroad is easy to control.

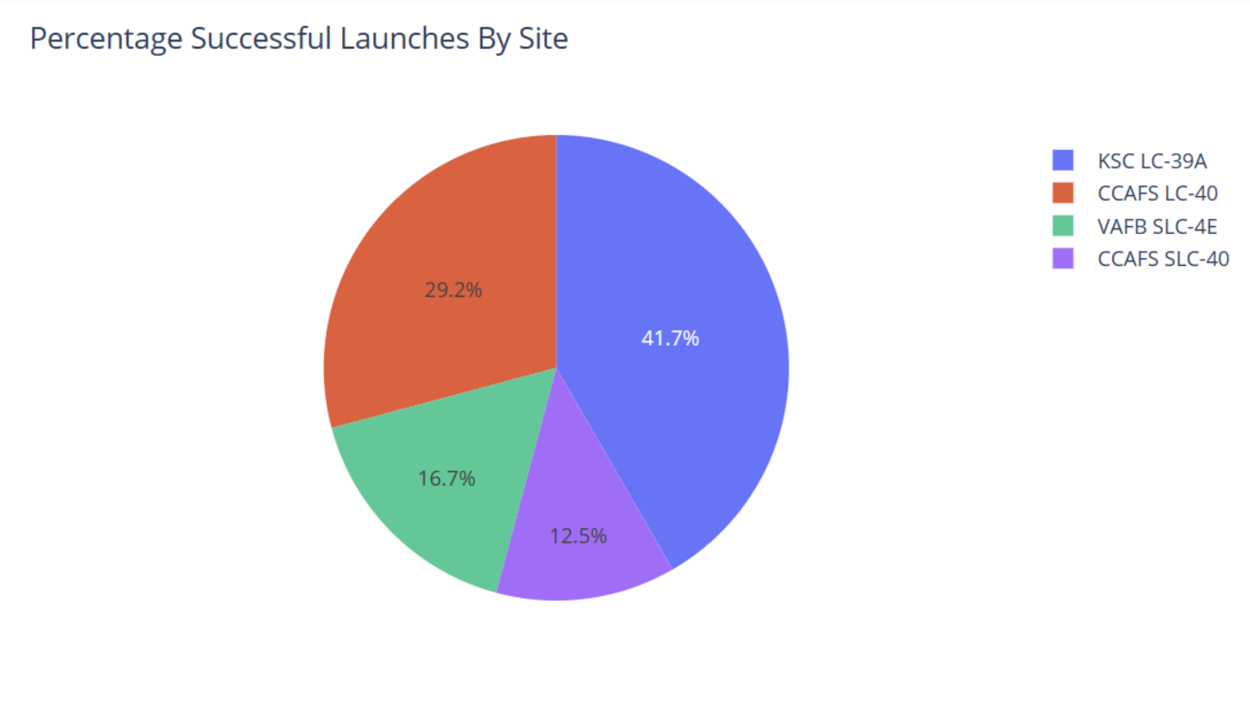


Section 4

Build a Dashboard with Plotly Dash

Successful launches across sites

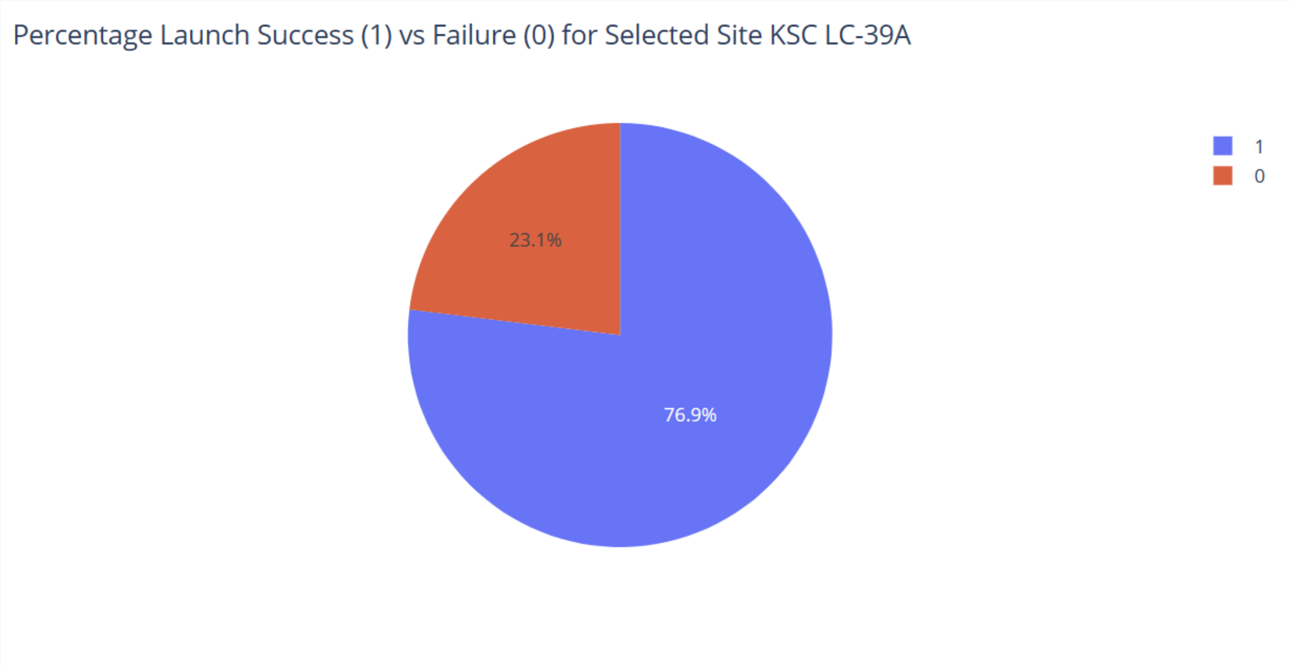
- The data set used for this task contains data for a total of 56 SpaceX launces during the period April 2010 to June 2018.
- 24 of the launches are classified as 'success' and 32 as 'failures'.
- Site KSC LC-39A launches are the most successful with 10 out of the total of 24 and achieved with just a total of 13 launces from the site.
- However, it is important to also keep in mind factors such as when launces occurred and how many launces were involved for each site.



Successful launches for individual site(s)

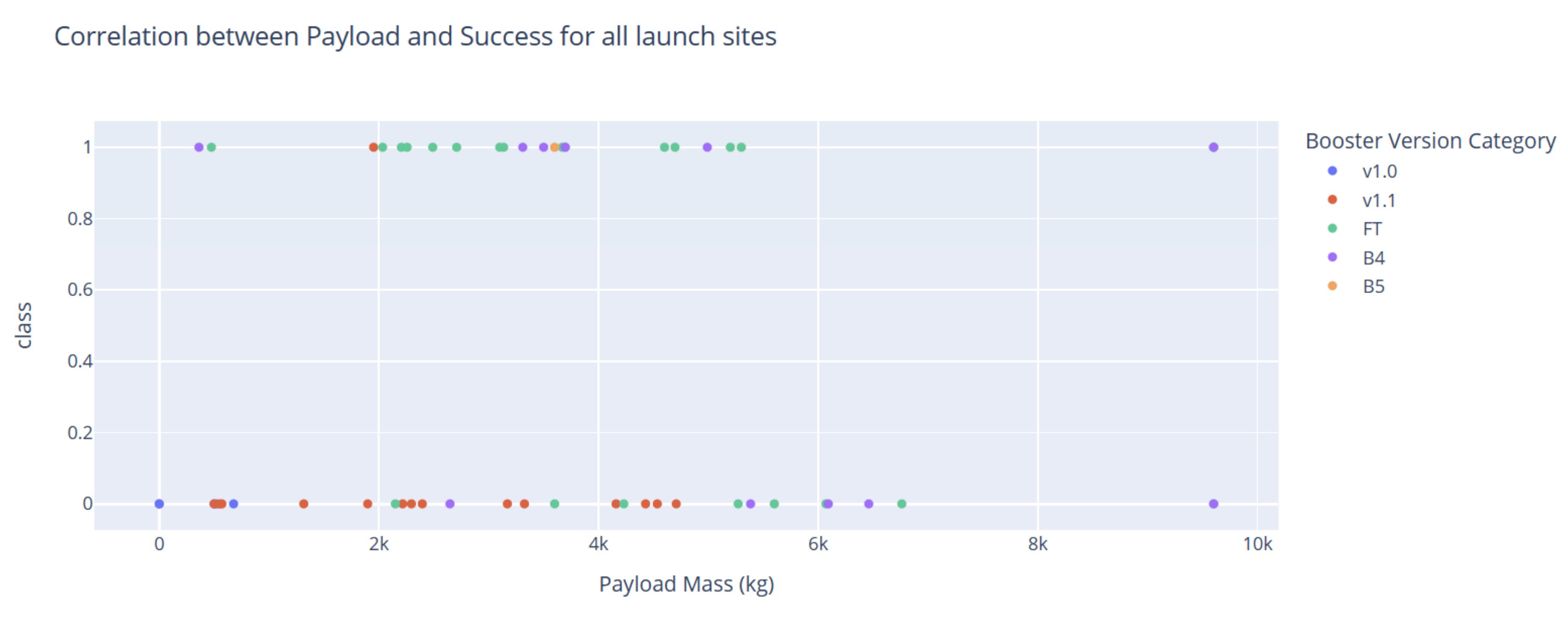
- The data set used for this task contains data for a total of 56 SpaceX launces during the period April 2010 to June 2018.
- As stated on previous page site KSC LC-39A has the best performance with 10 out of 13 launches being a success.
- A success rate of 76.9% which is standout performance.
- However, performance should also be considered based on number of launces with special awareness to low numbers.
- The table below indicates a success rate around 40% for two sites but with low number of launces.
- As stated before site CCAFS LC-40 shows poor performance but were involved in the very early launces.

Site	Launches	Successes	Success (%)
CCAFS LC-40	26	7	26.9%
VAFB SLC-4E	10	4	40.0%
KSC LC-39A	13	10	76.9%
CCAFS SLC-40	7	3	42.8%



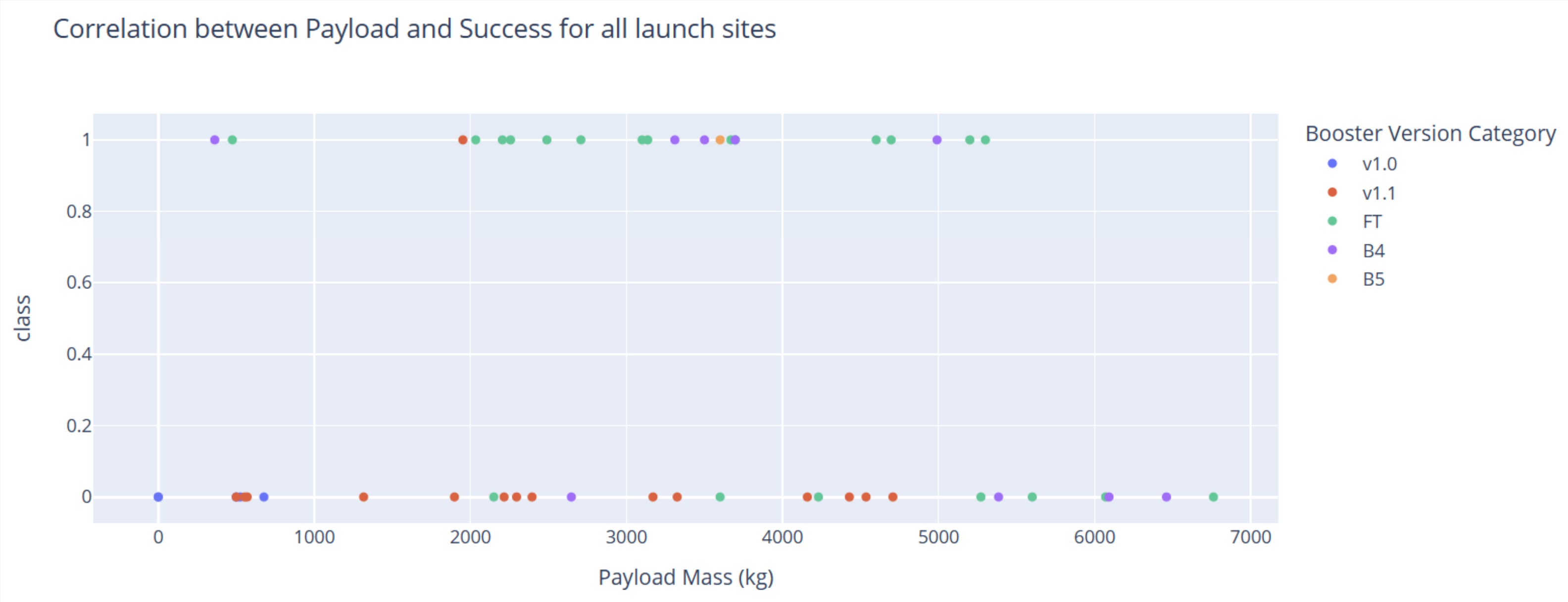
Booster version performance

- The data set used for this task contains data for a total of 56 SpaceX launces during the period April 2010 to June 2018.
- A total of 16 out of the 24 successful launches were equipped with the FT booster.
- 6 other successful launches were equipped with the B4 booster.
- The FT and the B4 boosters are dominantly successful.



Payloads vs Launch outcomes

- The data set used for this task contains data for a total of 56 SpaceX launces during the period April 2010 to June 2018.
- Ignoring the five high payload launches of 9600kg of which 3 were successful we can say:
- Payload range 2000kg – 5200kg seems to be the most successful.
- Payloads above 5200kg has had a lot of failures.
- Payloads below 2000kg has many failures but probably related to the early launces.

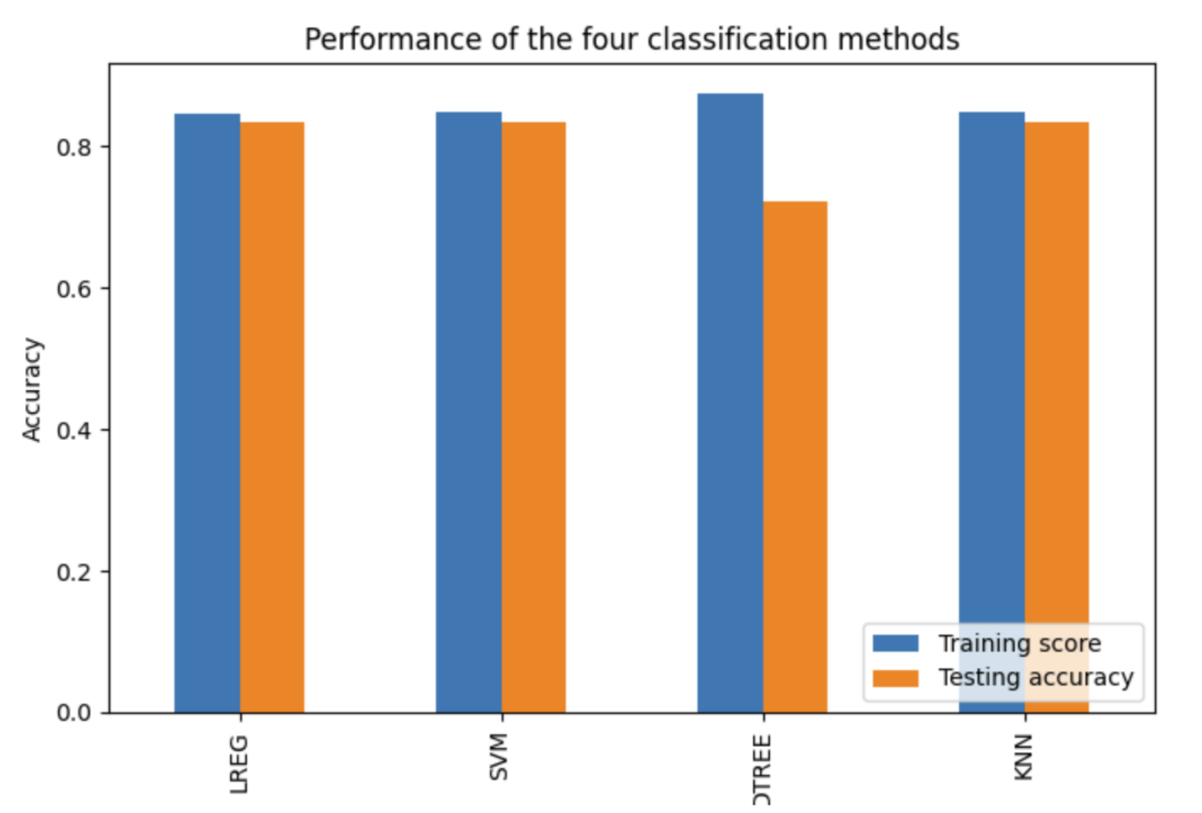


Section 5

Predictive Analysis (Classification)

Classification Accuracy

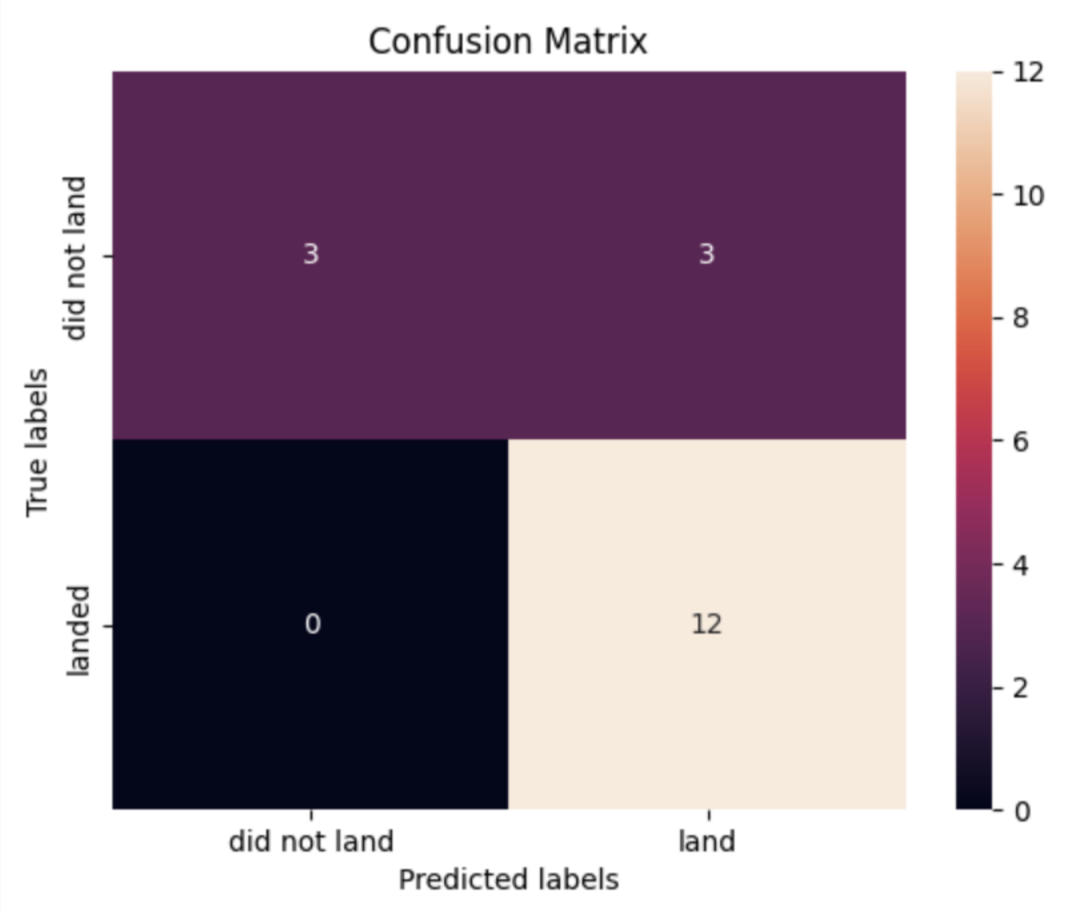
- Predictive analysis classification has been attempted with four different methodologies:
- Logistic Regression, Support Vector Machine, Decision Tree and K Nearest Neighbours.
- Parametres for each method were fixed initially.
- Out of the 90 data samples 80% were used for the training and 20% for the testing.
- This bar chart shows both the training score and the testing accuracy for the four methods.
- Training score for the Decision Tree method comes out slightly better than for the tree other methods.
- However, testing accuracy is poorer for the Decision Tree model than for the other tree models, which comes out as equally good with a value of 0.833.
- Three of the methods therefore perform equally well.



	Training score	Testing accuracy
LREG	0.846429	0.833333
SVM	0.848214	0.833333
DTREE	0.873214	0.722222
KNN	0.848214	0.833333

Confusion Matrix

- Given the identical results for the best performing methods Logistic Regression, Support Vector Machine and K Nearest Neighbours, it is sufficient to display just one Confusion Matrix.
- Out of the 18 test cases 12 were correctly predicted as successful landings but 3 of the predicted landings did not land. The final 3 cases was correctly predicted as unsuccessful landings.
- It should be noted that the data here spans the entire lifetime of the SpaceX project, hence the scores should be better if the initial launces are left out of the classification process.



Conclusions

- A key conclusion is that the results of SpaceX flights have been steadily improving over time and currently the chance of success is probably in the range of 80% to 90%.

- The analysis clearly demonstrates how the technology has matured into something which is now viable in terms of both reliability and commerciality.

Appendix

- No appendices have been found required for this presentation.

Thank you!