IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

N Simons
17thMarch 2025

Outline

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

Executive Summary

SpaceX 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 SpaceX can reuse the first stage. The cost of a launch can be determined if the first stage will land, . Our goal was to use this data to predict whether SpaceX will attempt to land a rocket or not.

Data collection

Data collection was done by webscrapping from the SpaceX REST API and the Falcon 9 Wiki pages. This gave data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome. This helped to train a machine learning model .

Exploratory Data analysis was done using visualization and SQL. Interactive visual analytics was doneusing Folium and Plotly Dash.Predictive analysis was done using classification models

Results

- All the launch sites are in proximity to the Equator line
- All the launch sites are in very close proximity to the coast
- Launch site KSC LC-39A had the largest number of succesful launches
- Launch site CCAFS SLC-40 had the smallest number of succesful launches
- As the flight number increases, the first stage is more likely to land successfully
- The most successful orbit types were ES-L1, GEO,HEO and SSO
- The launch sucess rate kept increasing from 2013 to 2020
- The date when the first succesful landing outcome in ground pad was achieved was 22nd July 2018
- The total number of successful mission outcomes was 100 with 1 failed mission outcome.
- The average payload mass carried by booster version F9 v1.1 was 2928.4 kg
- The payload with highest succesful launches is between 3000-4000kg

Introduction

Project background and context

SpaceX advertises Falcon 9 rocket launches are far less expensive than other providers. Much of the savings is because SpaceX can reuse the first stage.

The first stage does most of the work and is much larger than the second stage.

This stage is quite large and expensive.

Unlike other rocket providers, SpaceX's Falcon 9 can recover the first stage

If we can determine if the first stage will land, we can determine the cost of a launch,

Goals

- Gather information about Space X and create dashboards
- Determine if SpaceX will reuse the first stage.
- Train a machine learning model and use public information to predict if SpaceX
- will reuse the first stage.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

how data sets were collected.

Data Collection – SpaceX API

REQUESTED AND PARSED THE SPACEX LAUNCH DATA USING THE GET REQUEST

FILTERED THE DATAFRAME TO ONLY INCLUDE FALCON 9 LAUNCHES

USED THE MEAN AND .REPLACE() THE FUNCTION TO REPLACE NP.NAN VALUES IN THE DATA WITH THE MEAN

https://github.com/NanaSimons/IBM-Data-ScienceWork/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

Requested the Falcon9 Launch Wiki page from its URL

Extracted all column/variable names from the HTML table header

Created a data frame by parsing the launch HTML tables

https://github.com/NanaSimons/IBM-Data-ScienceWork/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

- performed Exploratory Data Analysis (EDA) to find some patterns in the data and determine what would be the label for training supervised models.
- Converted outcomes into Training Labels with 1 meaning the booster successfully landed and 0 meaning it was unsuccessful.
- Used the method value_counts() on the column LaunchSite to determine the number of launches on each site:
- Used the method .value_counts() to determine the number and occurrence of each orbit in the column Orbit
- Used the method .value_counts() on the column Outcome to determine the number of landing_outcomes.Then assigned it toa variable(landing_outcomes).
- Using the Outcome, create a list where the element is zero if the corresponding row in Outcome is in the set bad_outcome; otherwise, it's one. Then assign it to the variable landing_class

https://github.com/NanaSimons/IBM-Data-ScienceWork/blob/main/labs-jupyter-spacex-Data wrangling (1).ipynb

EDA with Data Visualization

- Visualized the relationship between Flight Number and Launch Site using a scatterplot
- Visualized the relationship between Payload Mass and Launch Site using a scatterplot
- Visualized the relationship between success rate of each orbit type using a barchart
- Visualized the relationship between FlightNumber and Orbit type using a scatterplot
- Visualized the relationship between Payload Mass and Orbit type using a scatterplot
- Visualized the launch success yearly trend using a line chart to get the average launch success trend

- https://github.com/NanaSimons/IBM-Data-ScienceWork/blob/main/edadataviz.ipynb

EDA with SQL

Used SQL to:

- 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 succesful 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 records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months 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.

- https://github.com/NanaSimons/IBM-Data-ScienceWork/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- added each site's location on a map using site's latitude and longitude coordinates by pinning them on a map using folium Map object
- used folium.Circle to add a highlighted circle area with a text label on a specific coordinate
- created and added folium.Circle and folium.Marker for each launch site on the site map
- marked the successful/failed launches for each site on the map using MarkerCluster object
- Used MousePosition on the map to get coordinate for a mouse over a point on the map .
- Used folium.PolyLine` object draw a line between a launch site and its closest city, railway, highway,coastine

Reasons objects were added

- Marker clusters can be a good way to simplify a map containing many markers having the same coordinate.
- Colour-labeled markers in marker clusters helped to easily identify which launch sites have relatively high success rates.
- MousePosition helped to easily find the coordinates of any points of interest

https://github.com/NanaSimons/IBM-Data-

Build a Dashboard with Plotly Dash

Used Plotly Dash to

- dropdown list to enable Launch Site selection
- pie chart to show the total successful launches count for all sites
- a callback function to render success-pie-chart based on selected site dropdown
- a range Slider to Select Payload
- scatter chart to show the correlation between payload and launch success
- a callback function to render the success-payload-scatter-chart scatter plot

Reasons why the above plots and interactions were added

- dropdown menu to let us select different launch sites
- callback function is to get the selected launch site from site-dropdown and render a pie chart visualizing launch success counts.
- range Slider to easily select different payload ranges to identify visual patterns.
- a callback function to render the success-payload-scatter-chart scatter plot to visually observe how payload may be correlated with mission outcomes for selected site(s).

- https://github.com/NanaSimons/IBM-Data-ScienceWork/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Created a NumPy array from the column Class in data, by applying the method to_numpy() then assign it to the variable Y
- Standardized the data in X then reassigned it to the variable X using the transform provided (refer to appendix)
- Used the function **train_test_split** to split the data X and Y into training and test data
- Created a logistic regression object then created a **GridSearchCV object logreg_cv** and fitted the object to find the best parameters from the dictionary parameters.
- Calculated the accuracy on the test data using the **method score**
- Created a support vector machine object then created a GridSearchCV object svm_cv and fitted the object to find the best parameters from the dictionary parameters.
- Created a decision tree classifier object then created a GridSearchCV object tree_cv and fitted the object to find the best parameters from the dictionary parameters.
- Created a k nearest neighbours object then create a **GridSearchCV object knn_cv** and fitted the object to find the best parameters from the dictionary parameters
- Identified the model that performed best by calculating the accuracy of the test data using the method score

https://github.com/NanaSimons/IBM-Data-ScienceWork/blob/main/SpaceX_Machine Learning Prediction_Part_5.ipynb

Results

Exploratory data analysis results

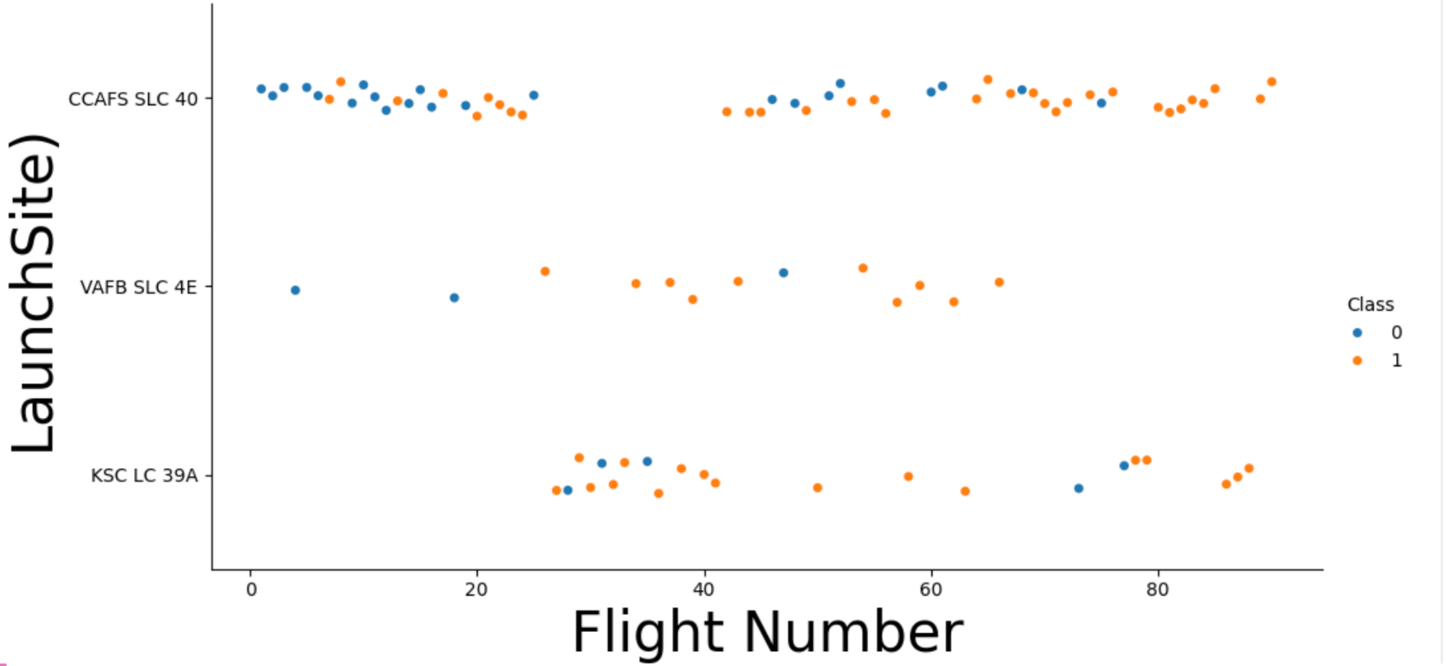
Interactive analytics demo in screenshots

Predictive analysis results

Section 2

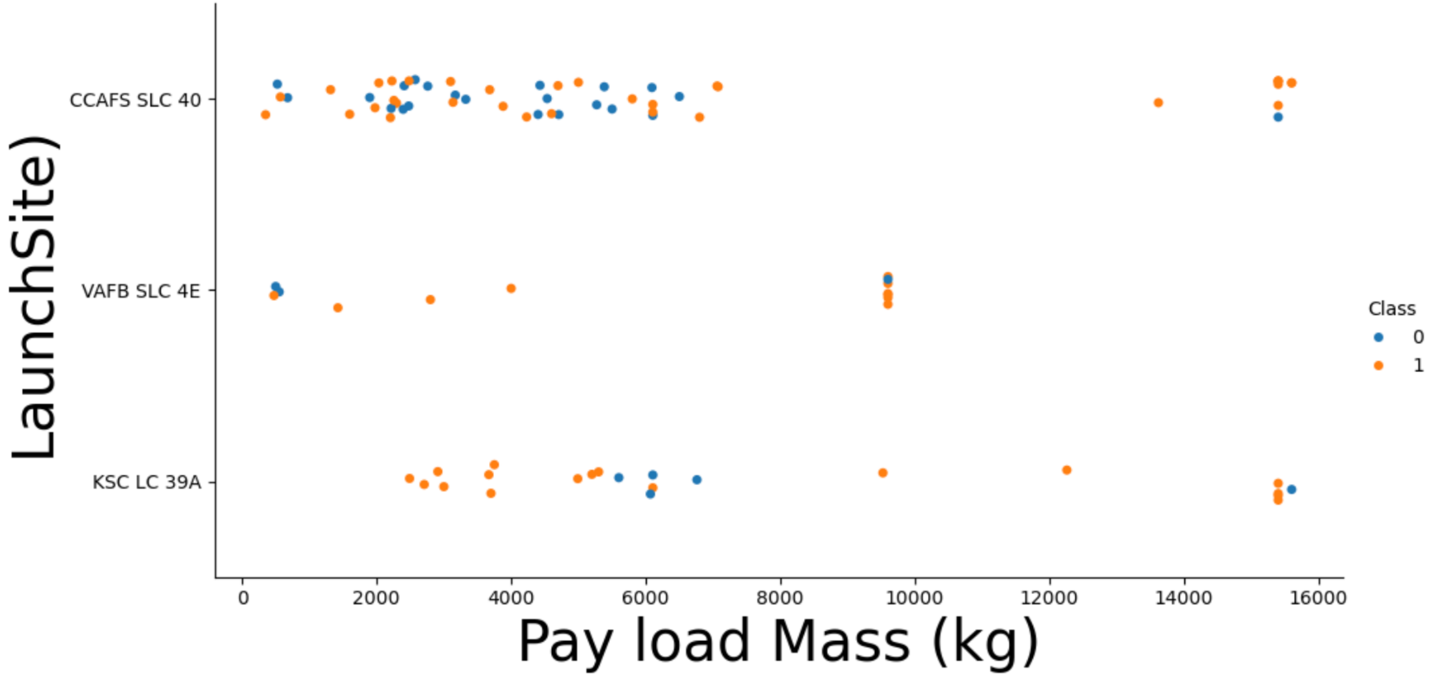
Insights drawn from EDA

Flight Number vs. Launch Site



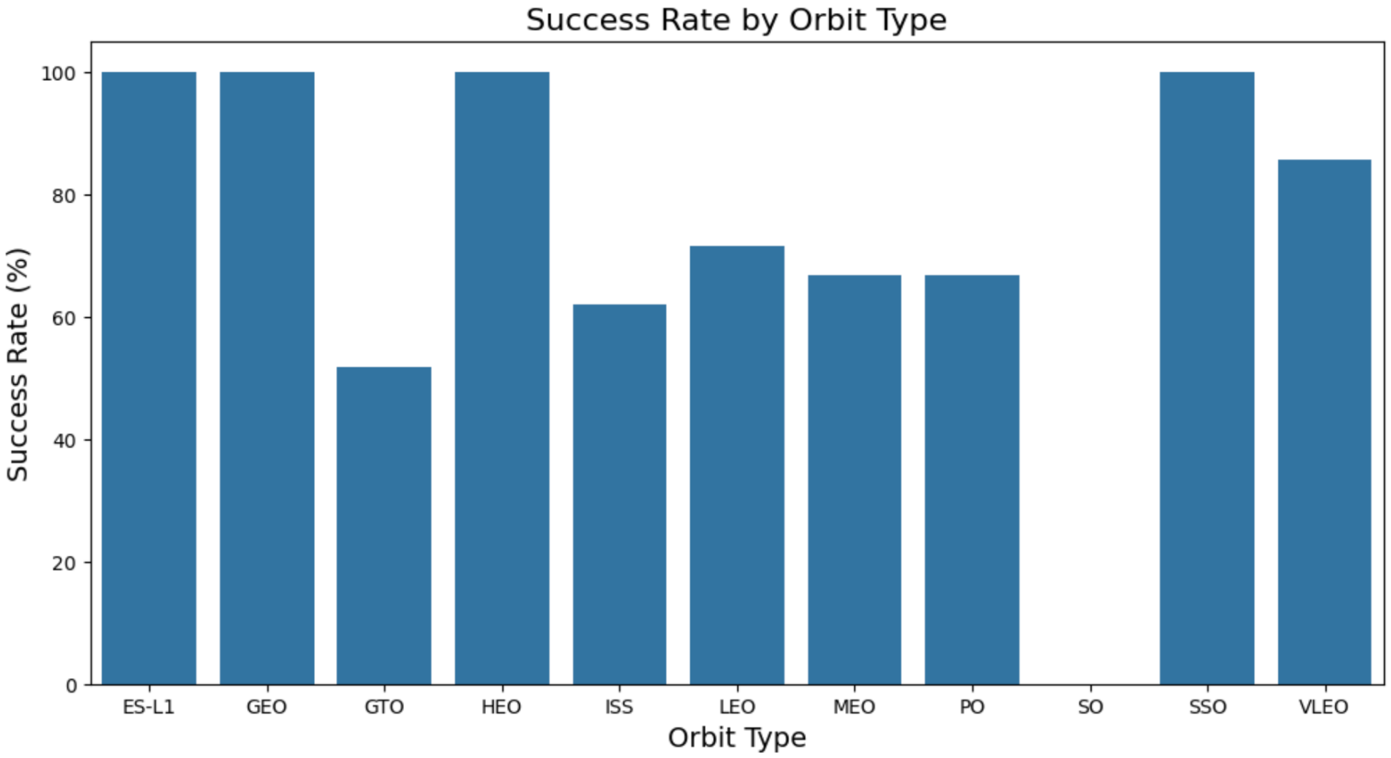
- The FlightNumber indicates the continuous launch attempts
- 1= successful launch outcome, 0=unsuccessful launch outcome
- As the flight number increases, the first stage is more likely to land successfully.

Payload vs. Launch Site



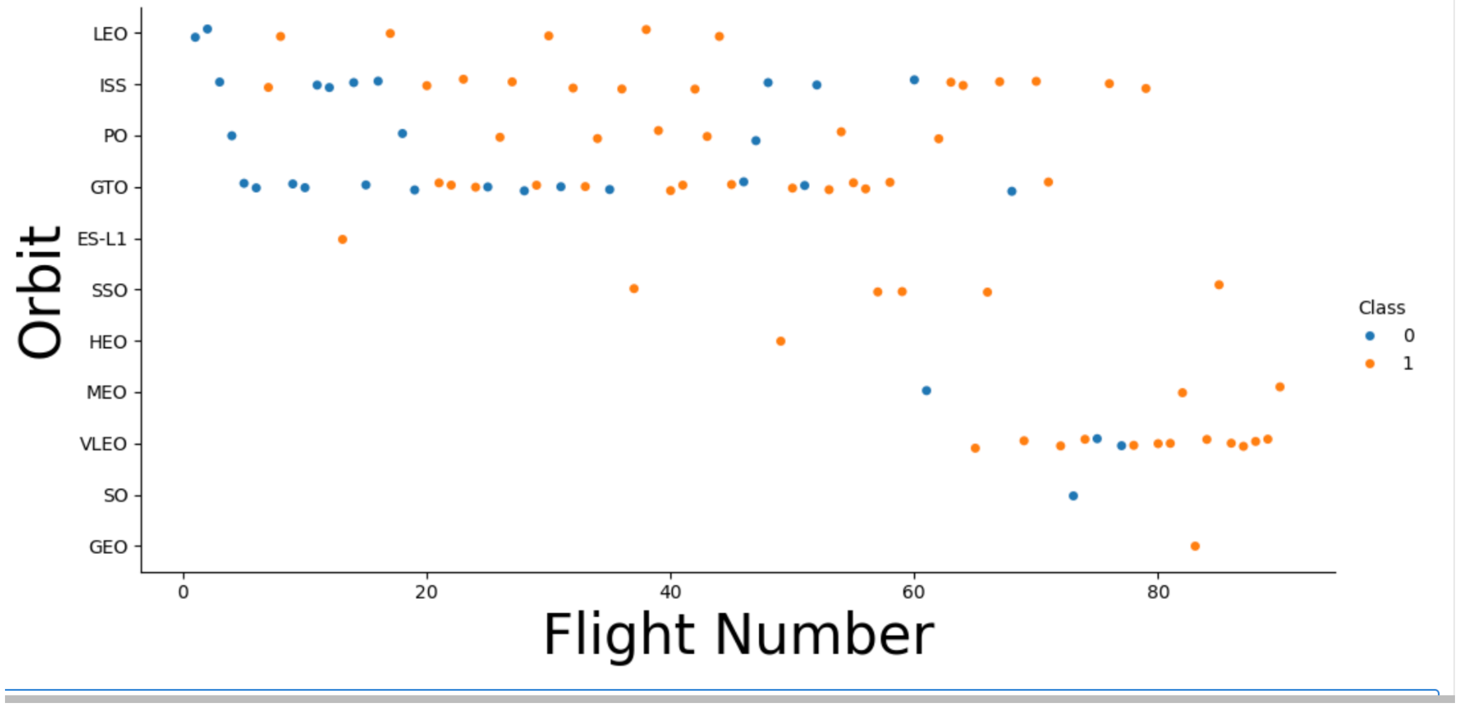
- for the VAFB-SLC launchsite there are no rockets launched for payload mass greater than 10000kg
- for the CCAFS SLC-40 launchsite, there appears to be no relationship payload mass and success for masses below 6000kg
- for the KSC LC-39A launchsite there were most successful launch outcomes were for payload masses between 2000kg and 6000kg

Success Rate vs. Orbit Type



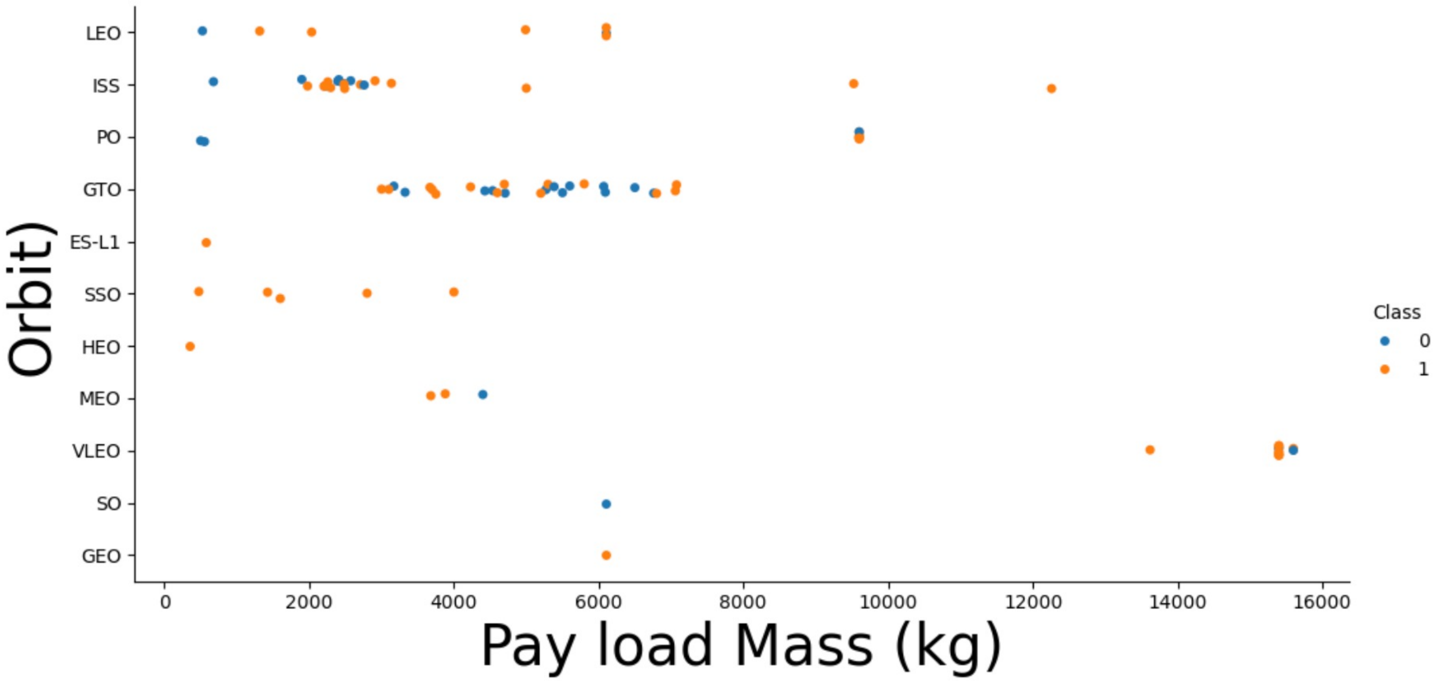
- The most successful orbit types were ES-L1, GEO,HEO and SSO
- The least succesful orbit type was GTO
- There were no successful launches from the SO orbit type

Flight Number vs. Orbit Type



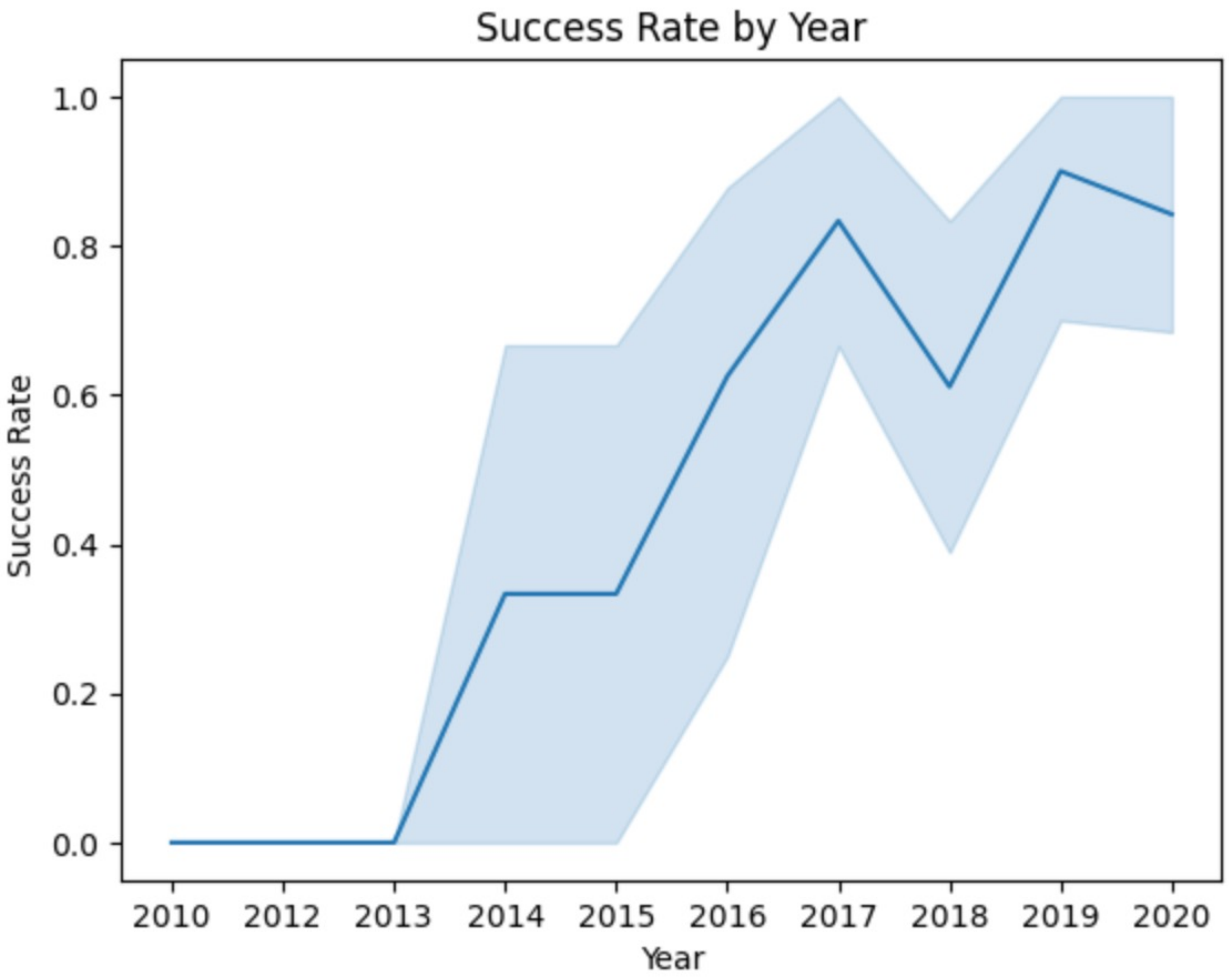
- For the LEO orbit, success seems to be related to the number of flights. Conversely, in the GTO orbit, there appears to be no relationship between flight number and success.

Payload Mass vs. Orbit Type



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

Launch Success Yearly Trend



- The launch sucess rate kept increasing from 2013 to 2020
- Launch success rate remained the same from 2014 to 2015
- Launch success rate gradually increased fromn 2015 to 2017
- Launch success rate decreased between 2017 and 2018, and then increased again

Launch Site Names

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

The names of the unique launch sites in the space mission

5 Records Where Launch Site Name Begins with 'CCA

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (p
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 (p
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	N
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	N
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	N

Total Payload Mass

sum (PAYLOAD_MASS__KG_)
45596

The total payload mass carried by boosters launched by NASA (CRS) was 45596Kg

Average Payload Mass by F9 v1.1

AVG(PAYLOAD_MASS__KG_)
2928.4

The average payload mass carried by booster version F9 v1.1 was

2928.4 kg

First Successful Ground Landing Date

min (Date)
2018-07-22

- The date when the first succesful landing outcome in ground pad was achieved was 22nd July 2018

Successful Drone Ship Landing with Payload between 4000 and 6000

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

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

- F9FT B1022
- F9FT B1026
- F9FT B1021.2
- F9FT B1031.2

Total Number of Successful and Failure Mission Outcomes

Mission_Outcome	Total
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

- The total number of successful mission outcomes was 100 with 1 failed mission outcome.

Boosters Carried Maximum Payload

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

- The names of the booster_versions which have carried the maximum payload mass

2015 Launch Records

Landing_Outcome	Booster_Version	Launch_Site	month	year
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	01	2015
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	04	2015

The records showing the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

- In 2015,failure in landing outcomes occurred in January and April only.
- In 2015,the CCAFS LC-40 launch site was the only launch site to record failures in landing outcome

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

Date	Landing_Outcome	count(*)
2016-04-08	Success (drone ship)	14
2015-12-22	Success (ground pad)	9
2015-06-28	Precluded (drone ship)	1
2015-01-10	Failure (drone ship)	5
2014-04-18	Controlled (ocean)	5
2013-09-29	Uncontrolled (ocean)	2
2012-05-22	No attempt	21
2010-06-04	Failure (parachute)	2

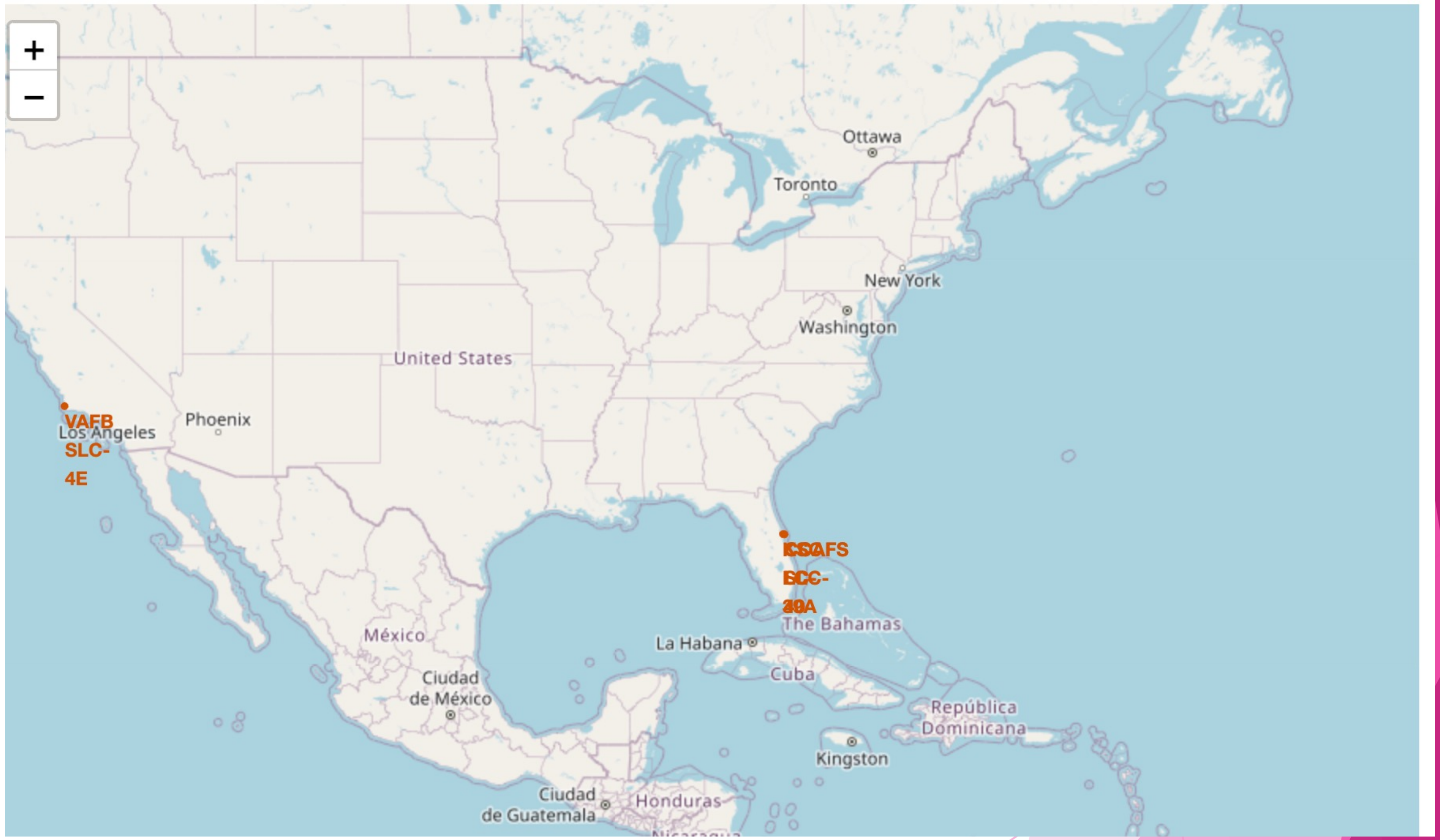
- There were 14 Successful droneship landings between 4th June 2010 and 20th March 2017
- There were 9 Successful ground pad landings between 4th June 2010 and 20th March 2017
- There were 5 failed drone ship landings between 4th June 2010 and 20th March 2017
- There were controlled landings in the ocean between 4th June 2010 and 20th March 2017
- There were 2 uncontrolled ocean landings between 4th June 2010 and 20th March 2017

Section 3

Launch Sites Proximities Analysis

Interactive Visual Analytics to Identify Geographical Patterns About Launch Sites using Folium .

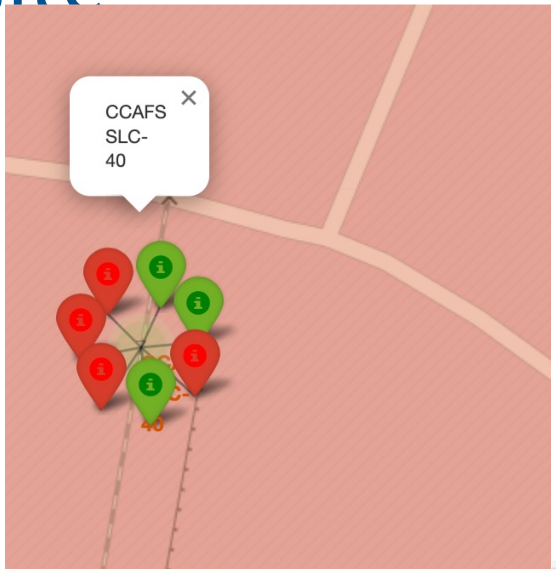
- All the launch sites are in proximity to the Equator line
- All the launch sites are in very close proximity to the coast .This is because Space X performs controlled landings in the oceans.

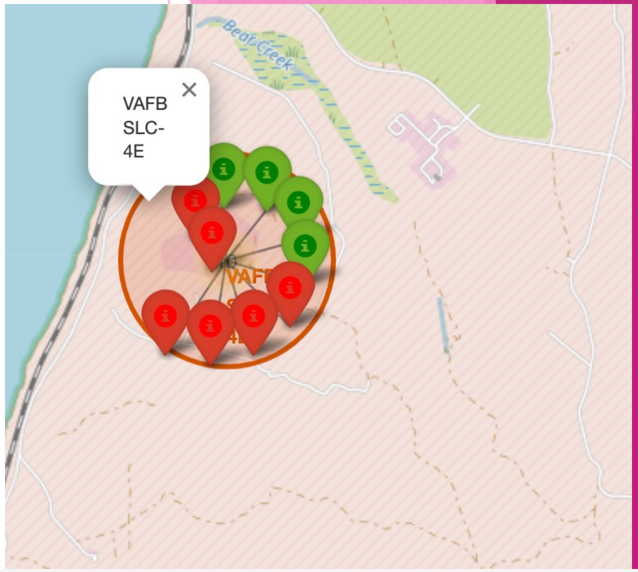


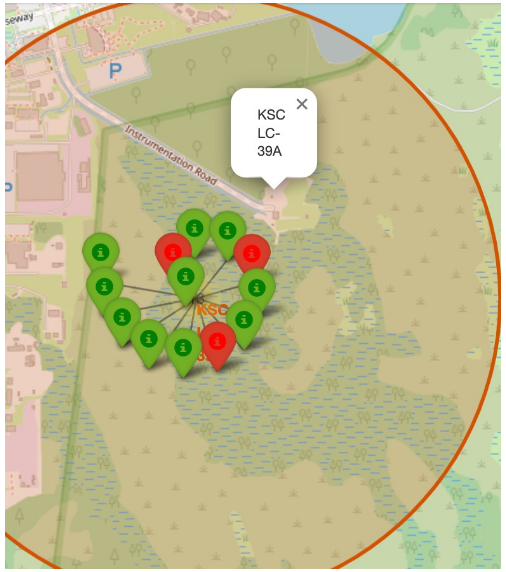
Interactive Visual Analytics with Folium showing The launch Outcomes For each site

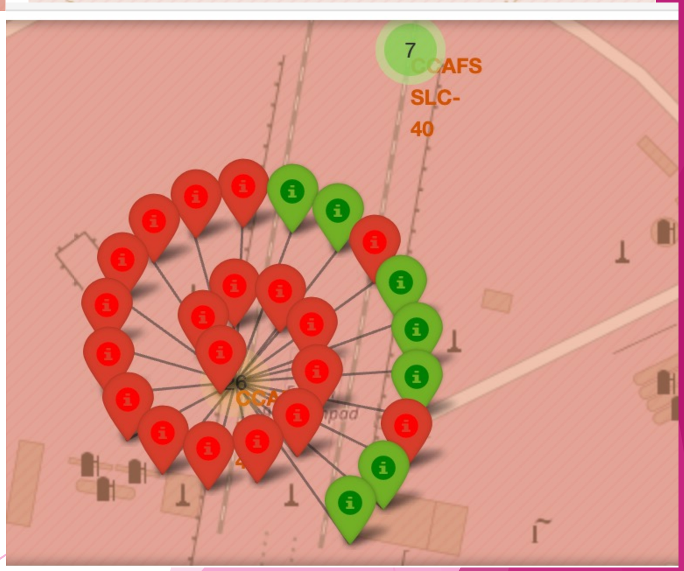
Green marker= successful launch,Red marker= unsuccessful launch

- Launchsite KSC LC-39A had the highest successful launch ratio
- Launchsite CCAFS LC-40 had the highest total number of launches
- Launchsite CCAFS SLC-40 had the lowest total number of launches







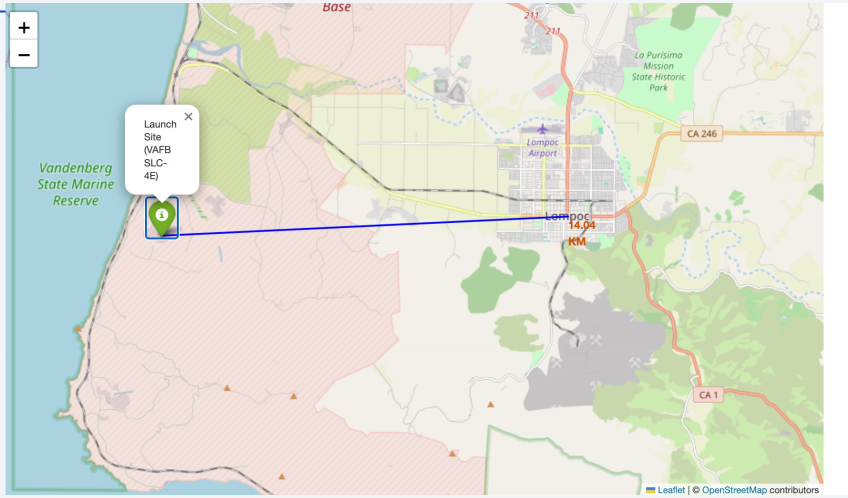


VAFB SLC-4E Launch Site Proximity to Railway, Highway, Coastline and City

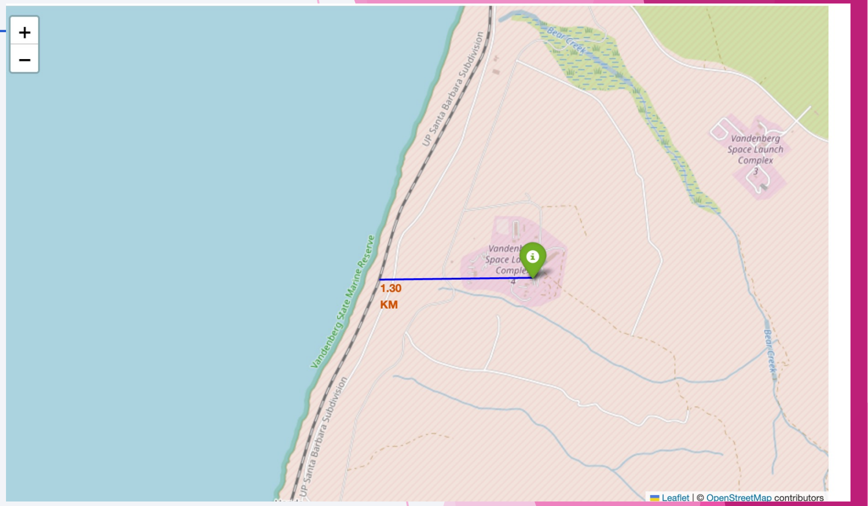
Findings

- Launchsite VAFB SLC-4E is 1.46Km from nearest coast line
- Launchsite VAFB SLC-4E is 1.30Km from the nearest railway line
- Launchsite VAFB SLC-4E is 1.17 Km from the nearest highway
- Launchsite VAFB SLC-4E is14.04Km from the nearest City

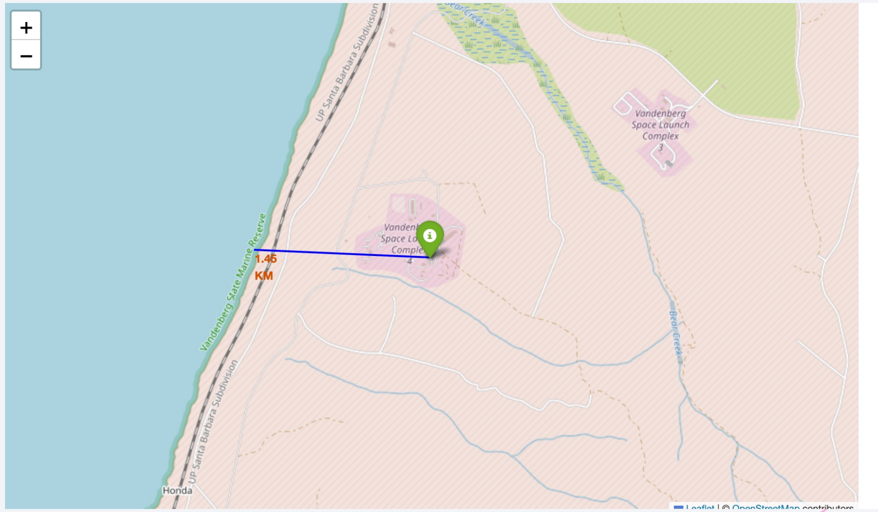
Distance to City



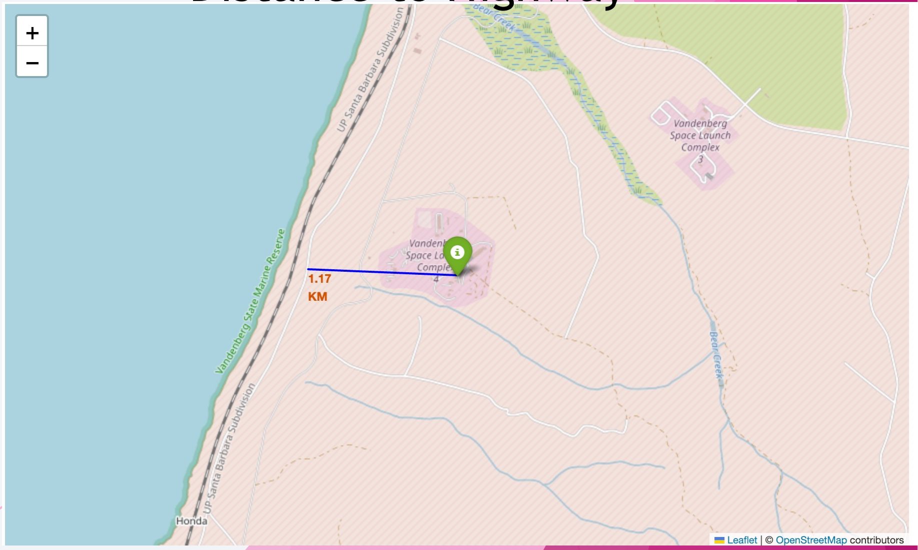
Distance to Railway line



Distance to Coastline



Distance to Highway

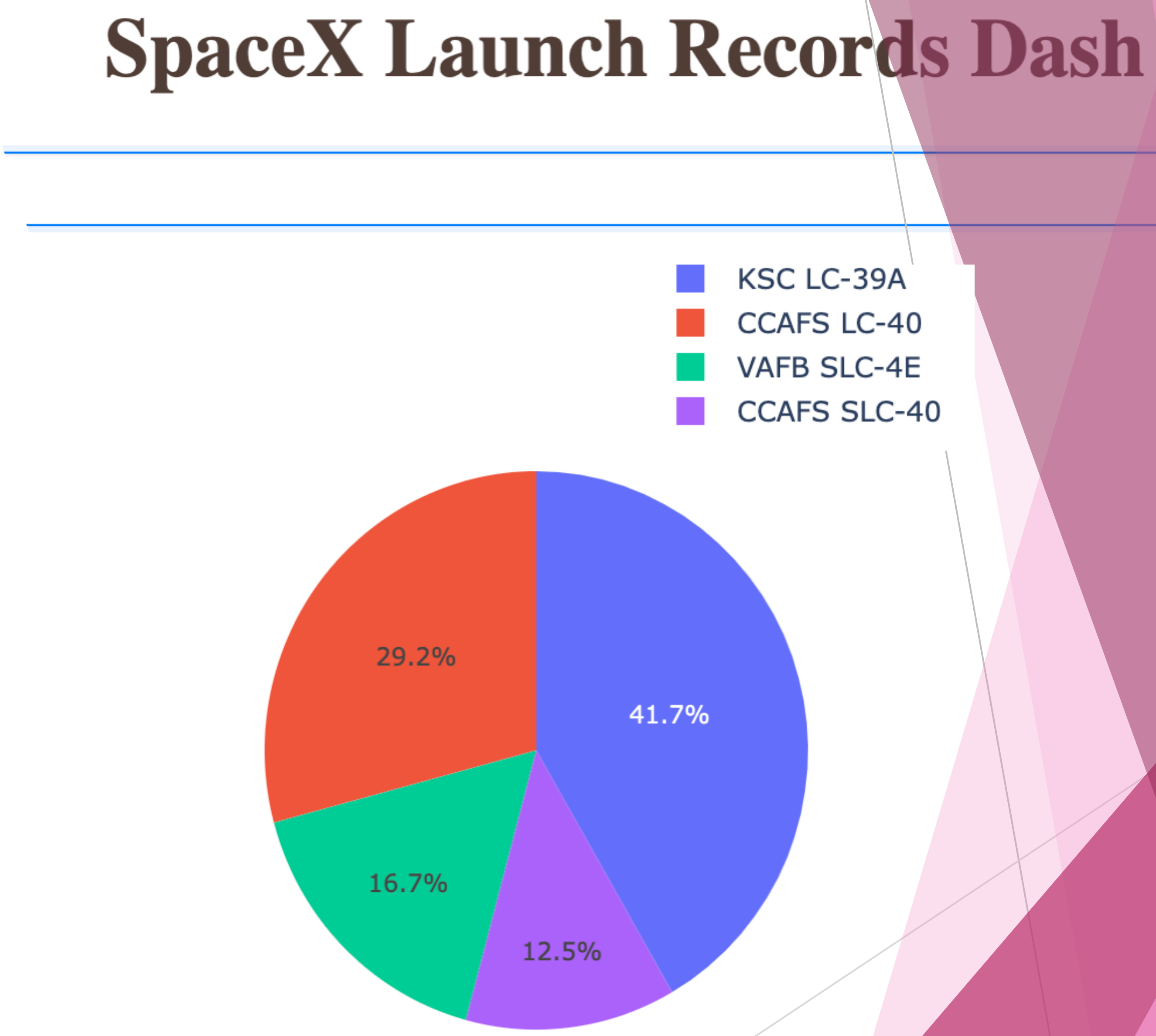


Section 4

Build a Dashboard with Plotly Dash

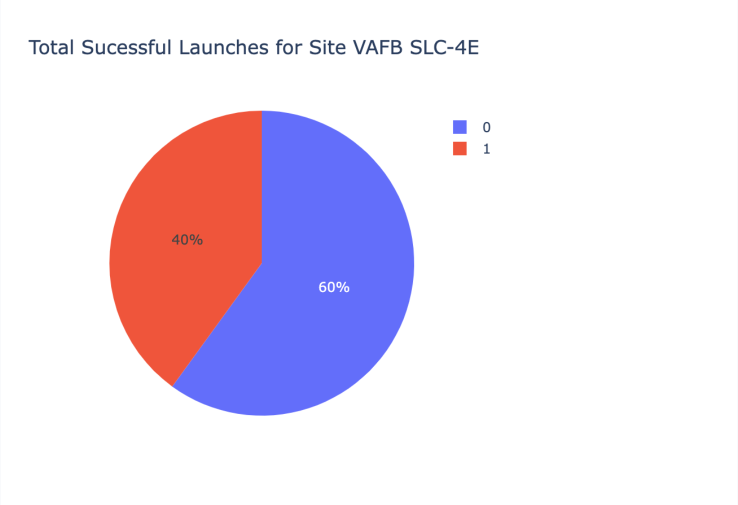
Total Successful Launches by Site

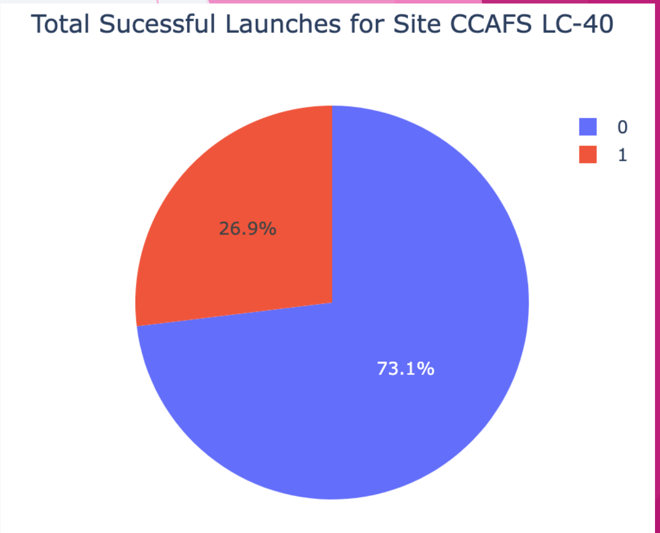
- Launch site KSC LC-39A had the largest number of succesful launches
- Launch site CCAFS SLC-40 had the smallest number of succesful launches



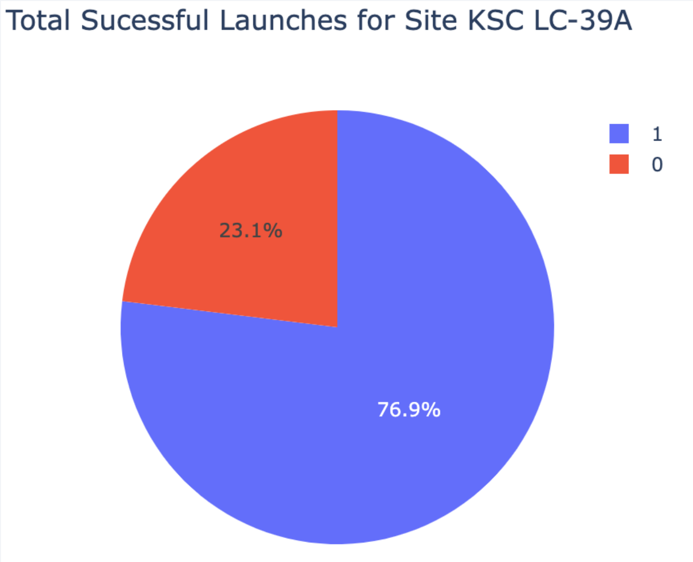
Launch Site with Highest Launch Success Ratio

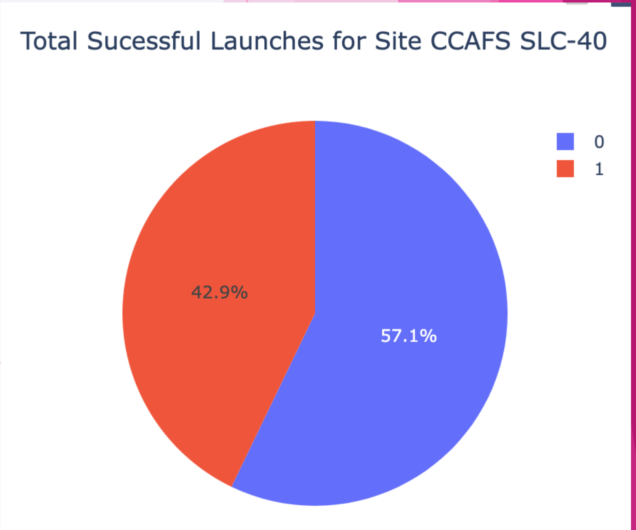
- 1 = successful launch whereas 0 = unsuccessful launch





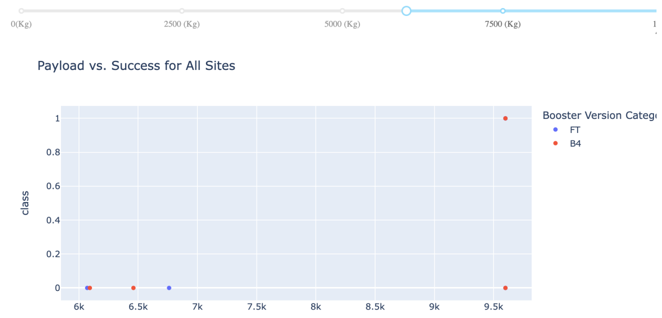
- Launch site with highest success ratio was KSC LC 39 with a success ratio of 76.9%
- Launch site with the lowest success ratio was CCAFS LC-40 with a success ratio of 26.9%

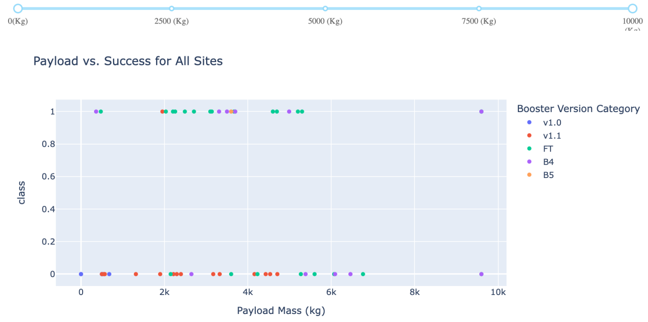




Successful launches by Payload mass

Payload range (Kg):





- 1 = successful launch whereas 0 = unsuccessful launch
- Payload with highest succesful launches is between 3000-4000kg
- Payload range with lowest succesful launches is between 0-2000kg
- Only booster version B4 launched successfully above a payload of 7500kg
- F9 Booster Version FT had the highest number of successful launches.

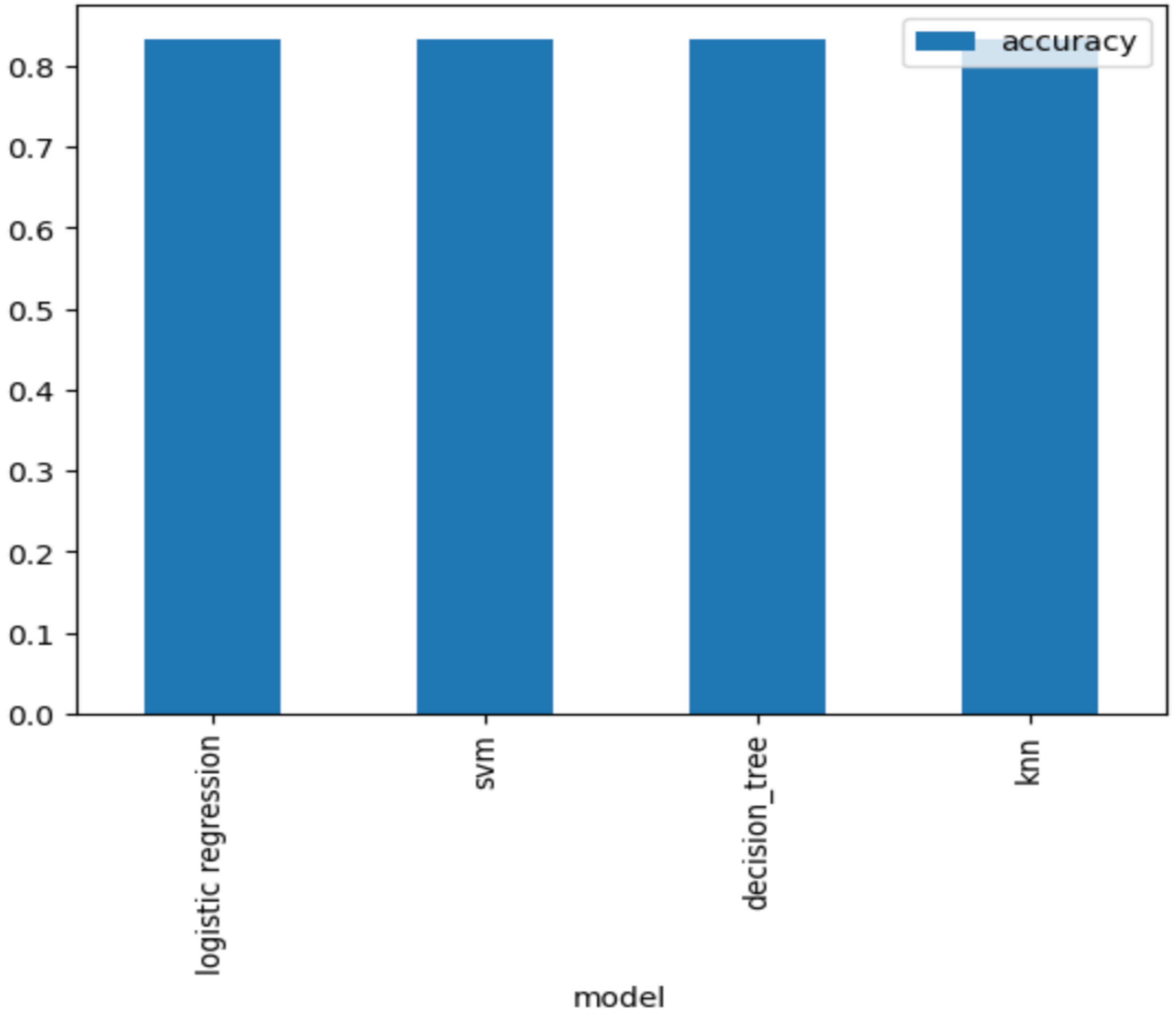
S. Simons

Section 5

Predictive Analysis (Classification)

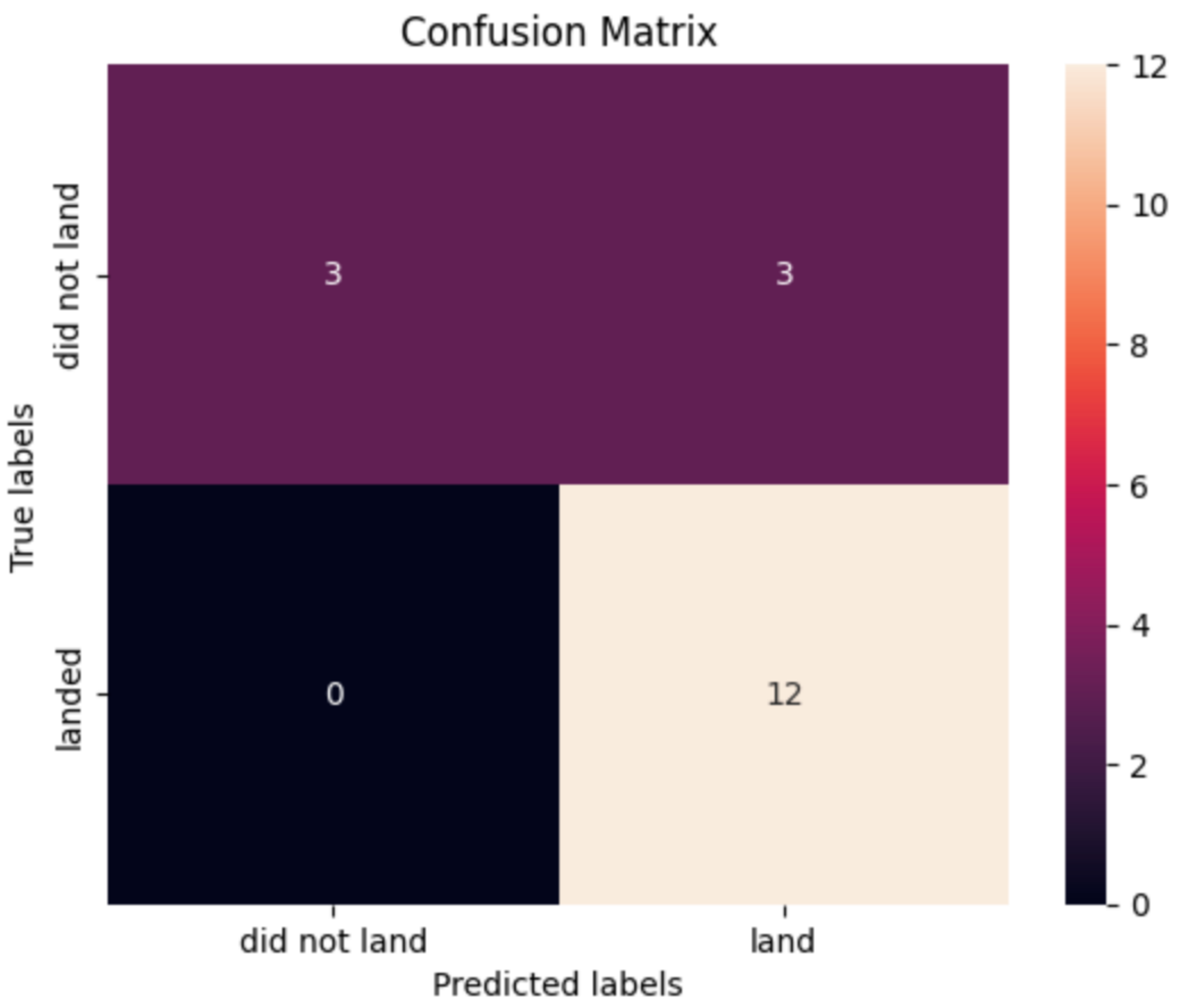
Classification Accuracy

model	accuracy
logistic regression	0.833333
svm	0.833333
decision_tree	0.833333
knn	0.833333



- All models had a classification accuracy of 83.3%
- This may have been due to the small dataset used.

Confusion Matrix



Conclusions

- All the Space X launch sites are in close proximity to the coast
- The launch success rate kept increasing from 2013 to 2020
- As the number of continuous launch attempts increases, the first stage is more likely to land successfully
- The most successful orbit types were ES-L1, GEO,HEO and SSO
- Payload mass with highest succesful launches is between 3000-4000kg

Appendix

Predictive Analysis

```
transform = preprocessing.StandardScaler()
scaler=preprocessing.StandardScaler().fit(X)
X=scaler.transform(X) type(X)
```

- SQL Queries

```
%sql select Distinct Launch_Site from SPACEXTABLE;
```

```
%sql select Booster_Version from SPACEXTABLE where Landing_Outcome='Success (drone ship)'and PAYLOAD_MASS__KG_ be
```

```
%sql select * from SPACEXTABLE Order by  Launch_Site  limit 5
```

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

```
%sql select min (Date)from SPACEXTABLE where  Landing_Outcome='Success'
```

N. Simons

Thank you!