IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Murali Kalva
02JAN2023

Outline

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

Executive Summary

Summary of methodologies

- Data Collection through API
- Data Collection with Web Scraping
- Data Wrangling
- Exploratory Data Analysis with SQL
- Exploratory Data Analysis with Data Visualization
- Interactive Visual Analytics with Folium
- Interactive Dashboard with Ploty Dash
- Machine Learning Prediction

Summary of all results

- Exploratory Data Analysis result
- Interactive analytics in screenshots
- Predictive Analytics result

Introduction

Project background and context

SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars other providers cost upwards of 165 million dollars each much of the savings is because SpaceX can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. Instead of using rocket science to determine if the first stage will land successfully, we will train a machine learning model and use public information to predict if SpaceX will reuse the first stage.

Problems you want to find answers

Factors effecting successful landing and to predict if SpaceX will reuse the first stage

Section 1

Methodology

Methodology

Executive Summary

Data collection methodology:

SpaceX launch data that is gathered from an API, specifically the SpaceX REST API. This API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome. Another popular data source for obtaining Falcon 9 Launch data is web scraping related Wiki pages

Perform data wrangling

using the Python BeautifulSoup package to web scrape some HTML tables that contain valuable Falcon 9 launch records.

Dealing with Nulls

filter/sample the data to remove Falcon 1 launches

Perform exploratory data analysis (EDA) using visualization and SQL

Perform interactive visual analytics using Folium and Plotly Dash

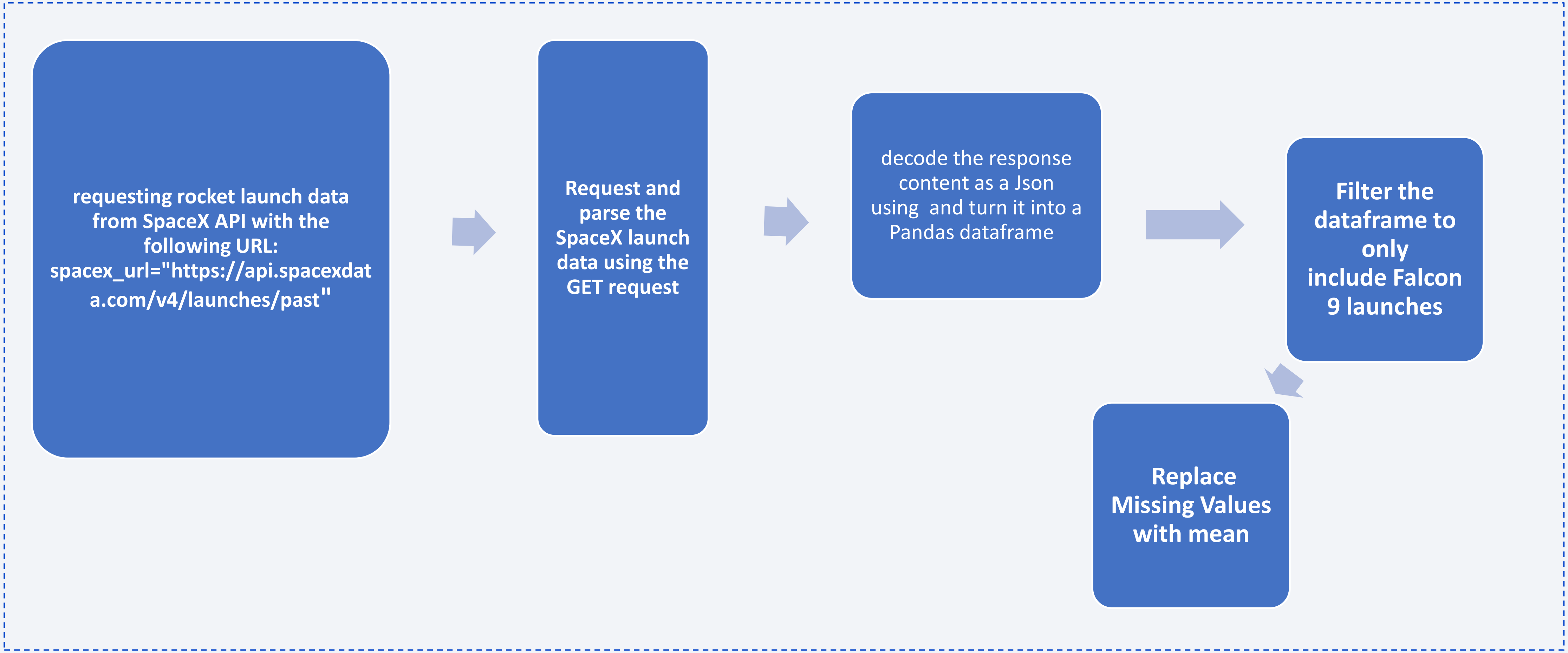
Perform predictive analysis using classification models

How to build, tune, evaluate classification models

Data Collection

1. Data Collection API
2. Data Collection with Web Scraping

Data Collection – SpaceX API



GitHub URL [testrepo/Data Collection API.ipynb at master · murali170717/testrepo (github.com)]

Data Collection - Scraping

Request the Falcon9 Launch Wiki page from its URL: "https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"

Create a object from the HTML

Extract all column/variable names from the HTML table header

Create a data frame by parsing the launch HTML tables

GitHub URL [testrepo/Data Collection with Web Scraping.ipynb at master · murali170717/testrepo (github.com)]

Data Wrangling

Use the method on the column to determine the number of launches on each site

Use the method to determine the number and occurrence of each orbit in the column

Use the method on the column to determine the number of

Using the , create a list where the element is zero if the corresponding row in is in the set ; otherwise, it's one

GitHub URL [testrepo/Data Wrangling.ipynb at master · murali170717/testrepo (github.com)]

EDA with Data Visualization

charts were plotted

Flight Number vs. Payload Mass, Flight Number vs. Launch Site, Payload Mass vs. Launch Site, Orbit Type vs. Success Rate, Flight Number vs. Orbit Type, Payload Mass vs Orbit Type and Success Rate Yearly Trend

Scatter plots show the relationship between continuous variables and Bar charts show comparisons among categorical variables.

GitHub URL [testrepo/EDA with Data Visualization.ipynb at master · murali170717/testrepo (github.com)]

EDA with SQL

- names of the unique launch sites in the space mission
- 5 records where launch sites begin with the string 'CCA'
- total payload mass carried by boosters launched by NASA (CRS)
- average payload mass carried by booster version F9 v1.1
- date when the first successful landing outcome in ground pad was acheived.
- 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
- names of the booster_versions which have carried the maximum payload mass. Use a subquery
- failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Ranking 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

GitHub URL testrepo/EDA with SQL.ipynb at master · murali170717/testrepo (github.com)

Build an Interactive Map with Folium

- Added Marker with Circle, Popup Label and Text Label of NASA Johnson Space Center using its latitude and longitude coordinates as a start location. - Added Markers with Circle, Popup Label and Text Label of all Launch Sites using their latitude and longitude coordinates to show their geographical locations and proximity to Equator and coasts.
- Added coloured Markers of success (Green) and failed (Red) launches using Marker Cluster to identify which launch sites have relatively high success rates.
- Added coloured Lines to show distances between the Launch Site KSC LC-39A (as an example) and its proximities like Railway, Highway, Coastline and Closest City.

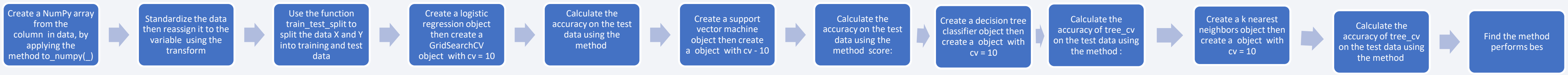
- GitHub URL: testrepo/Interactive Visual Analytics with Folium.ipynb at master · murali170717/testrepo (github.com)

Build a Dashboard with Plotly Dash

- Launch Site Drop-down Input Component: to select launch site
- Callback function to render success-pie-chart based on selected site dropdown: to know site has the largest successful launches and site has the highest launch success rate
- Range slider to select payload: to know Which payload range(s) has the highest and lowest launch success rate
- Callback function to render success-payload-scatter-chart scatter plot: to know which F9 Booster version has the highest launch success rate

GitHub URL testrepo/Dashboard_with_Ploty_Dash.py at master · murali170717/testrepo (github.com)

Predictive Analysis (Classification)



GitHub URL [testrepo/Machine Learning Prediction.ipynb at master · murali170717/testrepo (github.com)]

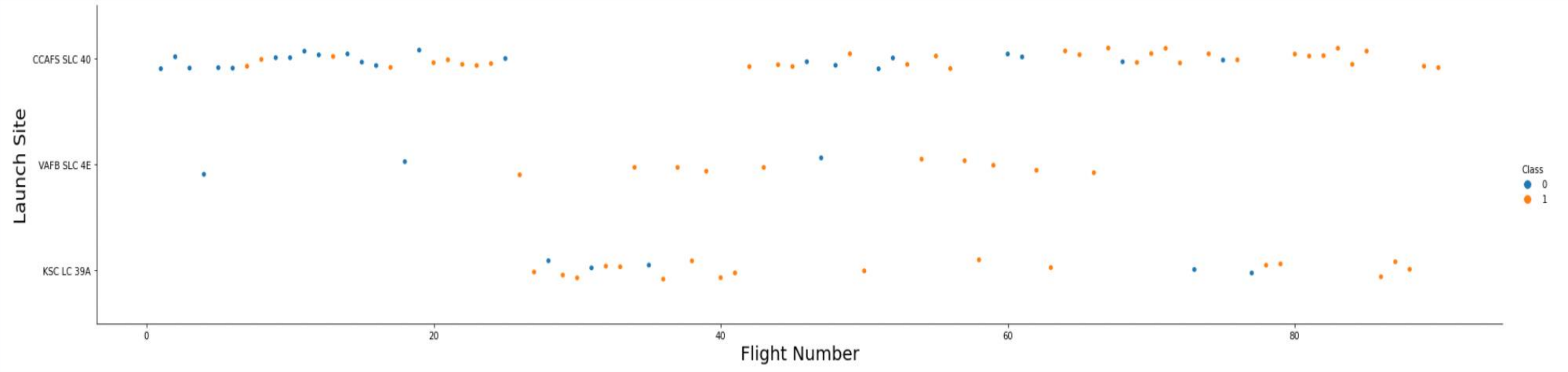
Results

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

Section 2

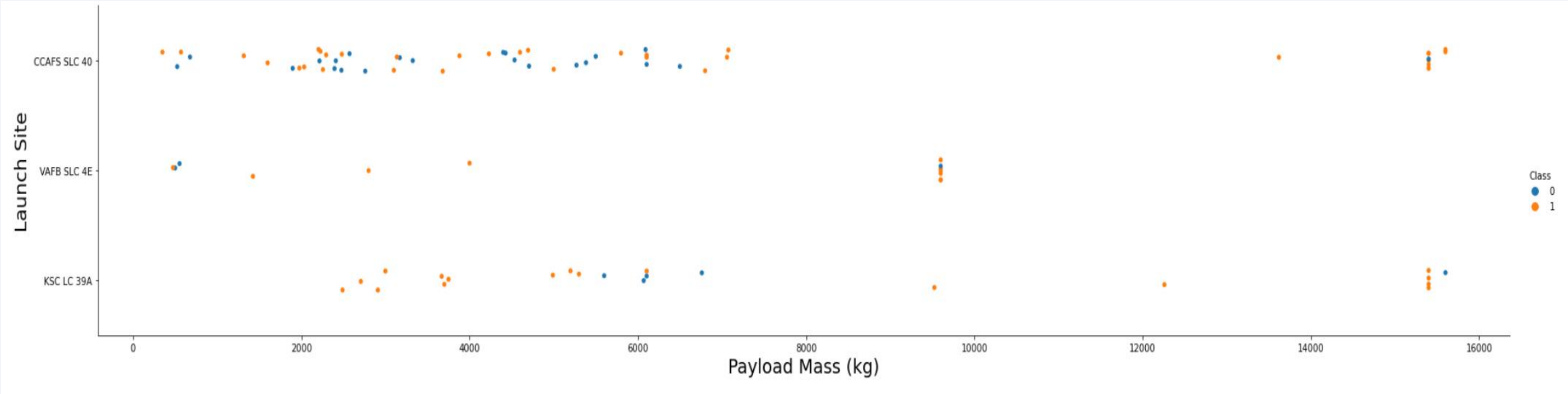
Insights drawn from EDA

Flight Number vs. Launch Site



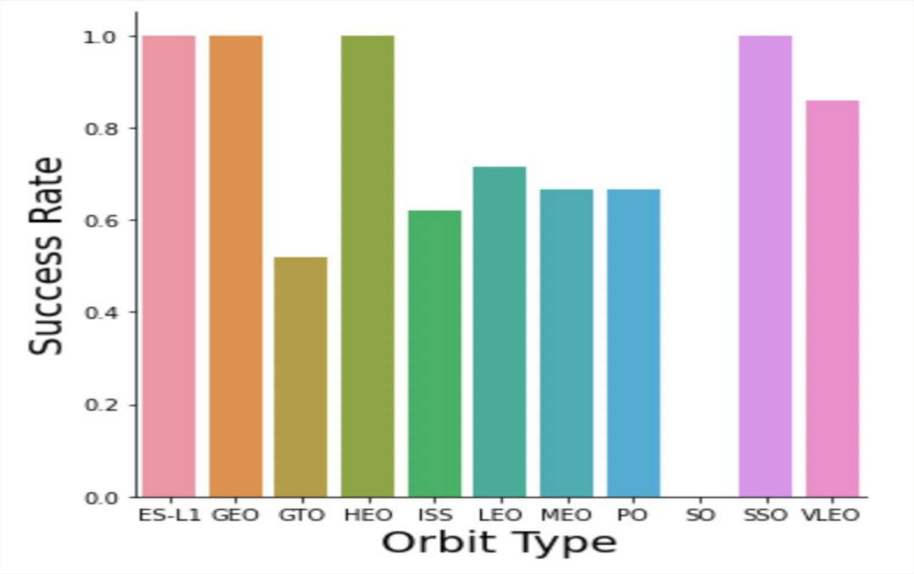
larger the flight numbers at a launch site, the greater the success rate at a launch site

Payload vs. Launch Site



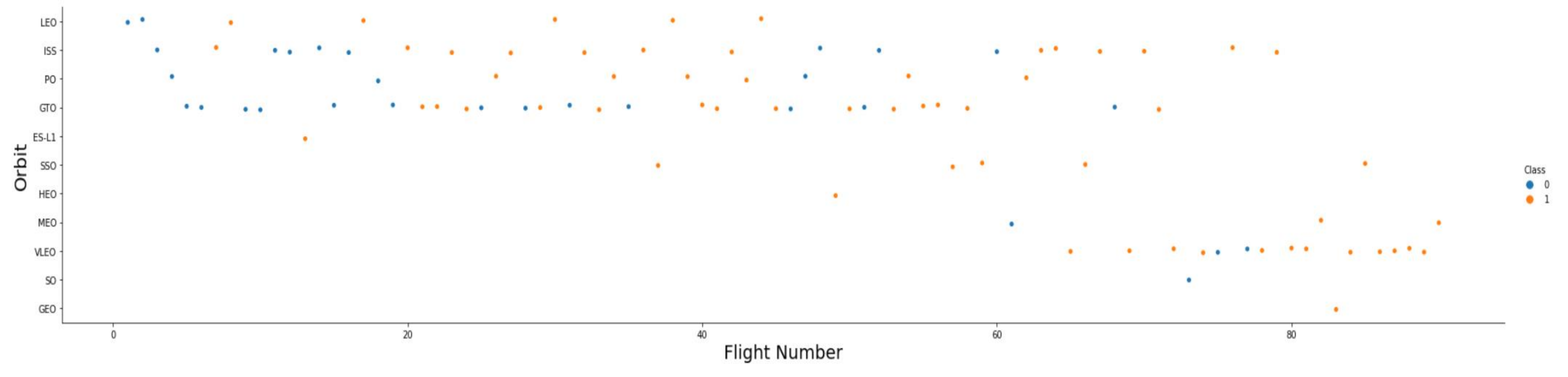
higher the payload mass, the higher the success rate.

Success Rate vs. Orbit Type



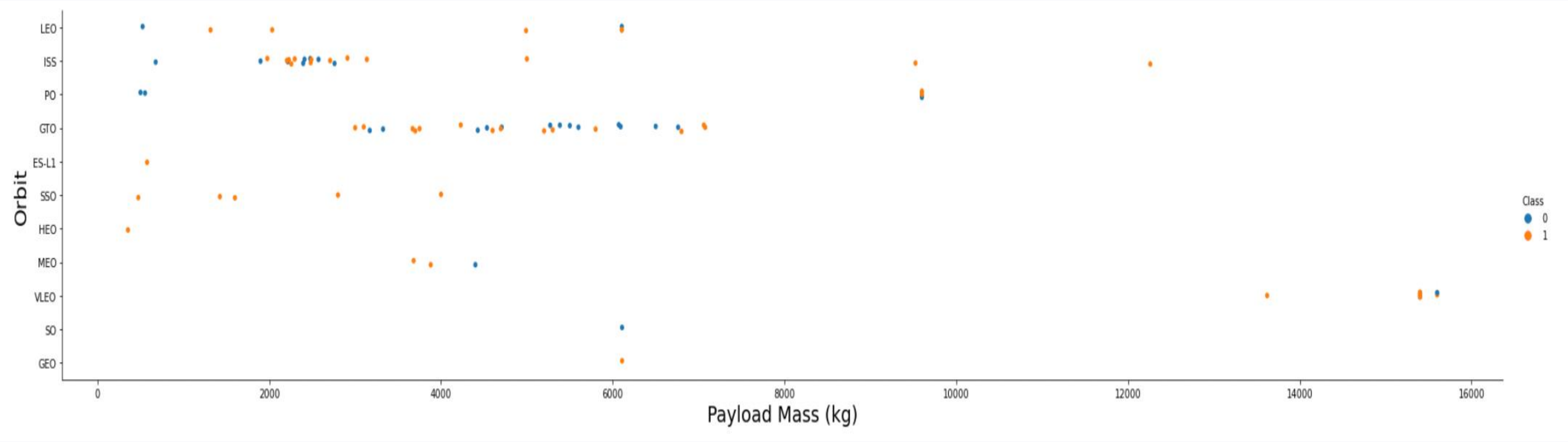
Orbits ES-L1, GEO, HEO, SSO and 100% success rate and orbit SO has 0% success rate

Flight Number vs. Orbit Type



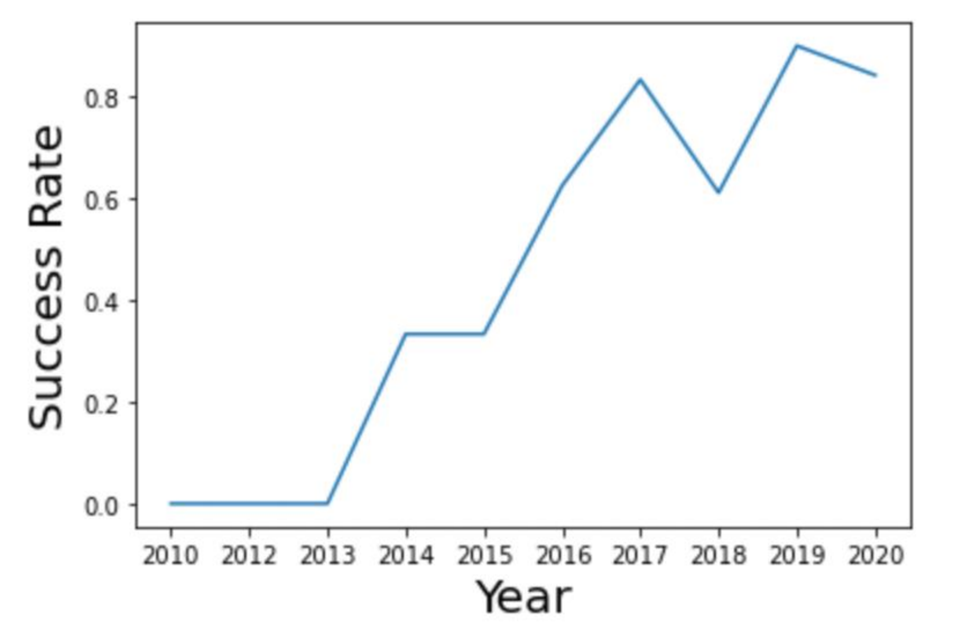
in the LEO orbit, success is related to the number of flights whereas in the GTO orbit, there is no relationship between flight number and the orbit

Payload vs. Orbit Type



As payload Mass increases, the successful landing are more for PO, LEO and ISS orbits

Launch Success Yearly Trend



- From 2013 to 2020 there is increase in success rate

All Launch Site Names

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

```
%sql select distinct launch_site from spacex;
```

```
 * ibm_db_sa://knq10819:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

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

Keyword distinct used to select unique launch sites

Launch Site Names Begin with 'CCA'

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

```
%sql select * from spacex where launch_site like 'CCA%' limit 5;
```

```
 * ibm_db_sa://knq10819:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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)
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-12	22:41:00	F9 v1.1	CCAFS LC-40	SES-8	3170	GTO	SES	Success	No attempt

% is used along with limit 5 to display launch site Names that bedign with ‘CCA’

Total Payload Mass

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

```
%sql select sum(payload_mass__kg_) as total_payload_mass from spacex where customer = 'NASA (CRS)';
```

```
 * ibm_db_sa://knq10819:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

total_payload_mass
22007

Sum function along with where is used to to display payload mass by NASA(CRS)

Average Payload Mass by F9 v1.1

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

```
%sql select avg(payload_mass__kg_) as average_payload_mass from spacex where booster_version like '%F9 v1.1%';
```

```
 * ibm_db_sa://knq10819:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

average_payload_mass
3226

Avg function along with % used to display average payload mass by F9 v1.1

First Successful Ground Landing Date

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

Hint:Use min function

```
%sql select min(date) as first_successful_landing from spacex where landing__outcome = 'Success (ground pad)';
```

```
 * ibm_db_sa://knq10819:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

first_successful_landing
2017-01-05

Min function is used to display first successful ground landing date

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
%sql select booster_version from spacex where landing__outcome = 'Success (drone ship)' and payload_mass__kg_ between 4000 and 6000;
```

```
 * ibm_db_sa://knq10819:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

booster_version
F9 FT B1022
F9 FT B1031.2

Where and between are used to display successful drone ship landing with payload between 4000 and 6000

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

```
%sql select mission_outcome, count(*) as total_number from spacex group by mission_outcome;
```

```
 * ibm_db_sa://knq10819:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

mission_outcome	total_number
Success	44
Success (payload status unclear)	1

Count is used to display total no.of successful and failure mission outsomes

Boosters Carried Maximum Payload

```
%sql select booster_version from spacex where payload_mass__kg_ = (select max(payload_mass__kg_) from spacex);
```

```
 * ibm_db_sa://knq10819:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

0]:

booster_version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3

Max function is used to display boosters carried maximum payload

2015 Launch Records

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

```
%%sql select monthname(date) as month, date, booster_version, launch_site, landing__outcome from spacex
    where landing__outcome = 'Failure (drone ship)' and year(date)=2015;
```

```
 * ibm_db_sa://knq10819:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

MONTH	DATE	booster_version	launch_site	landing__outcome
October	2015-10-01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)

Where and year function used to failed 2015 launch records

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(*) as count_outcomes from spacex
    where date between '2010-06-04' and '2017-03-20'
    group by landing__outcome
    order by count_outcomes desc;
```

```
 * ibm_db_sa://knq10819:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

22]:

landing__outcome	count_outcomes
No attempt	7
Failure (drone ship)	2
Success (drone ship)	2
Success (ground pad)	2
Controlled (ocean)	1
Failure (parachute)	1

Between and desc in order is used to rank landing outcomes is descending order between 2010-06-04 and 2017-03-02

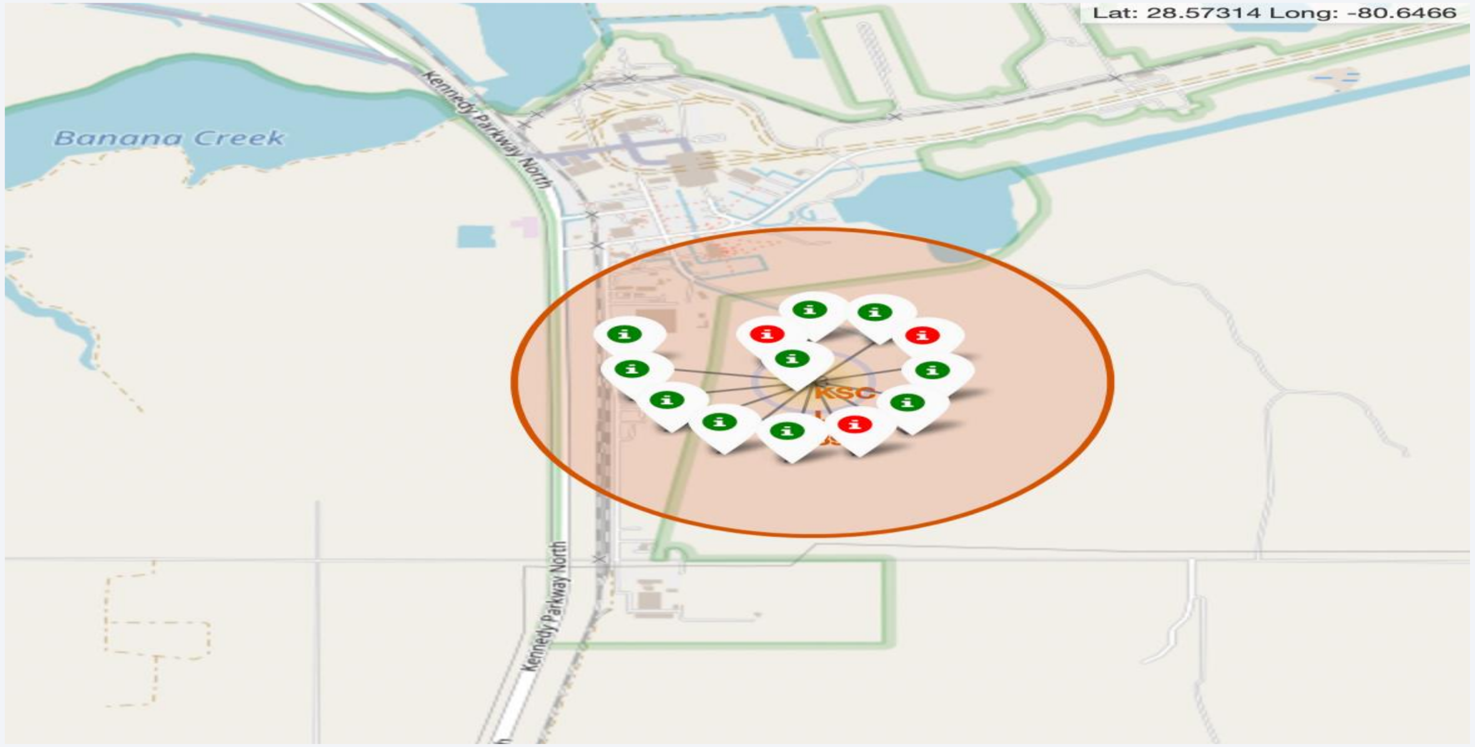
Section 3

Launch Sites Proximities Analysis

All launch sites

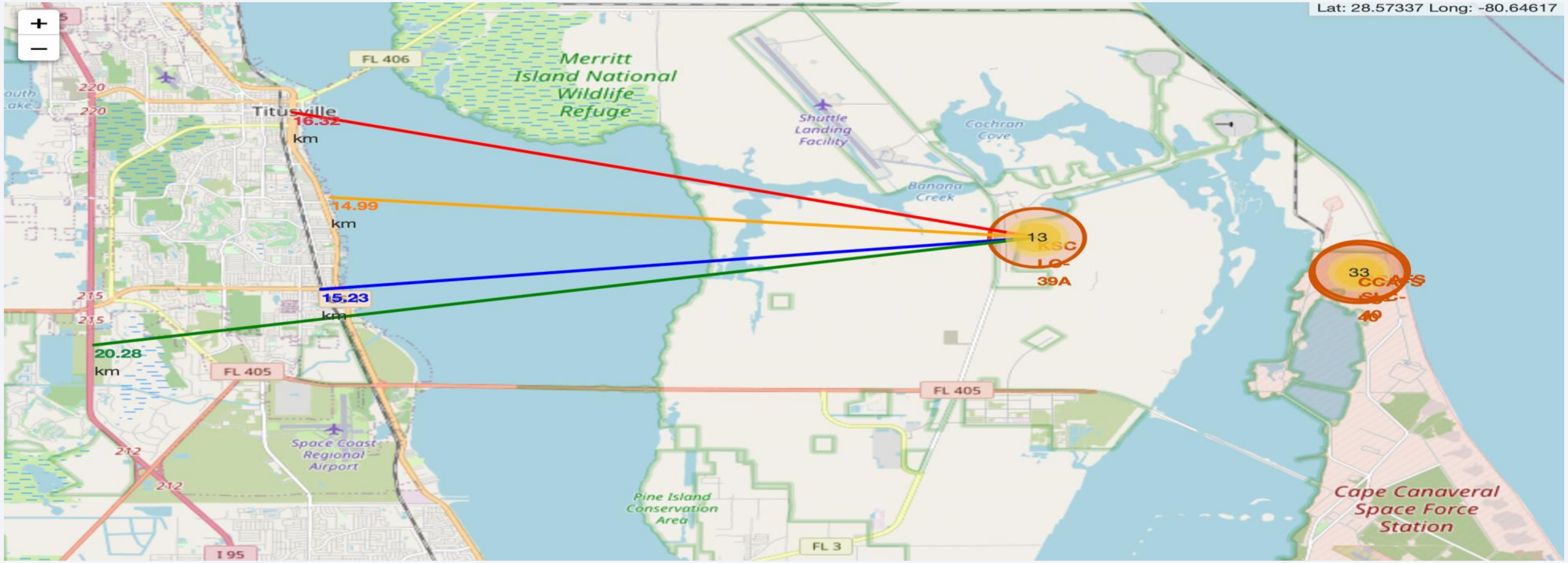
We can see the launch sites are in USA coasts, California and Florida minimizes the risk of debris as they will fall in ocean

Markers showing launch sites



Green is successful launch and Red is failed launch

Distance from KSCLC-39A site to landmarks



From railway 15.23 km, from roadway 20.28 km, from coastline 14.99 km

Section 4

Build a Dashboard with Plotly Dash

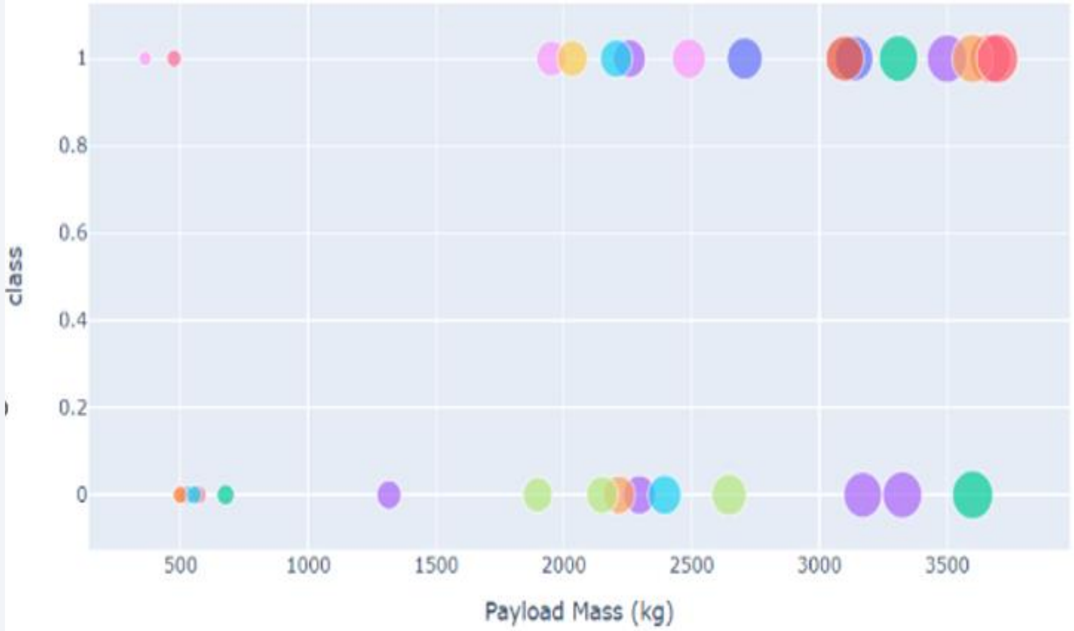
success by site using pie chart

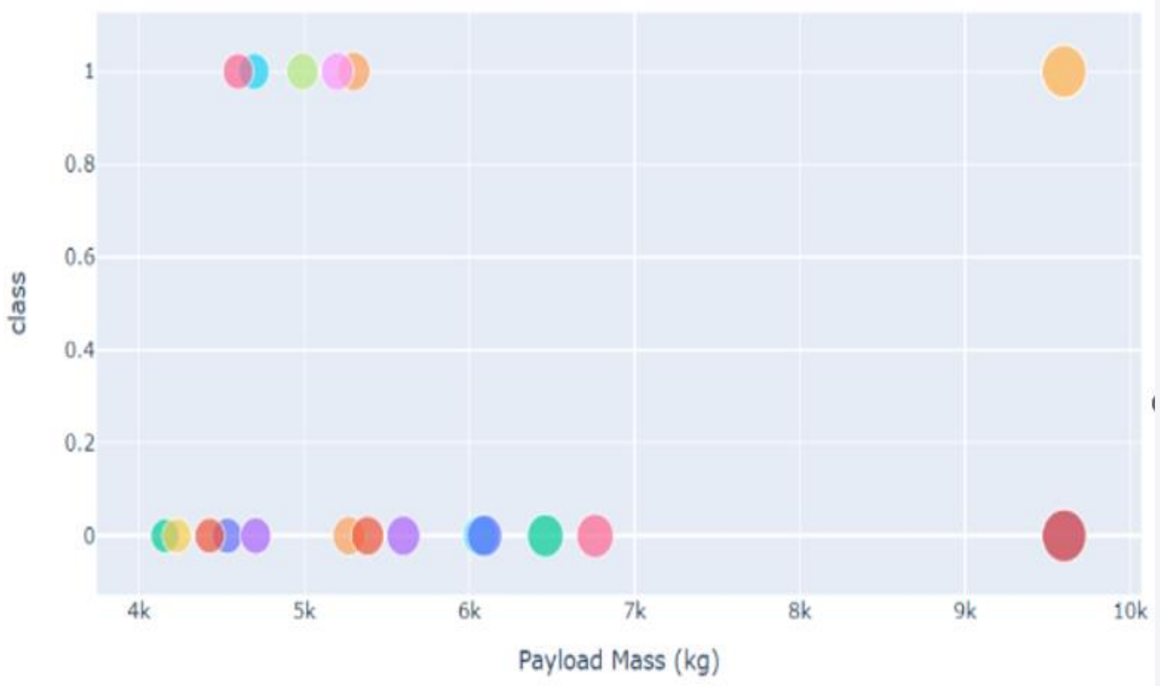
KSC LC-39A has highest success percentage

KSC LC-39A Success Ratio: Pie chart

Success percentage is 76.9% and failure percentage is 23.1%

Payload vs Launch Outcome





Low weight payloads (left) has higher success than high weight payload (right)

Section 5

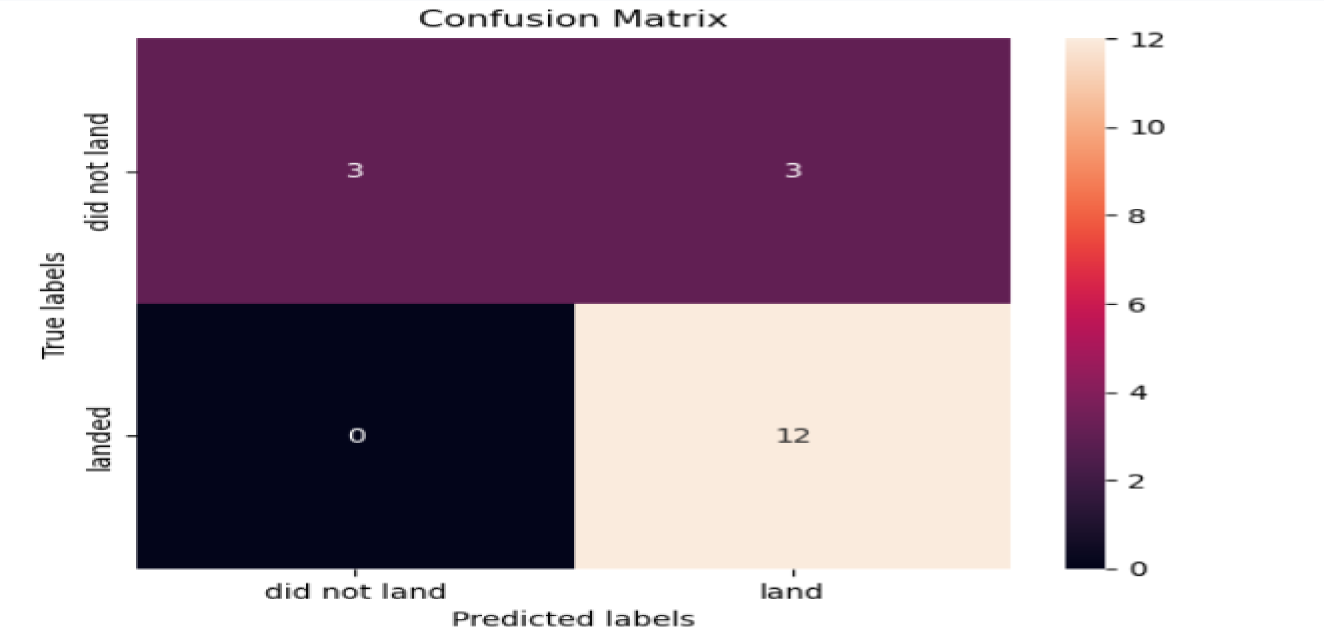
Predictive Analysis (Classification)

Classification Accuracy

	LogReg	SVM	Tree	KNN
Jaccard_Score	0.800000	0.800000	0.800000	0.800000
F1_Score	0.888889	0.888889	0.888889	0.888889
Accuracy	0.833333	0.833333	0.888889	0.833333

- From the above scores based on complete dataset, Decision model has highest accuracy and higher scores and is the best model

Confusion Matrix



From the above matrix we can see false positives

Conclusions

- larger the flight numbers at a launch site, the greater the success rate at a launch site
- higher the payload mass, the higher the success rate
- From 2013 to 2020 there is increase in success rate
- KSC LC-39A site has highest success percentage
- Low weight payloads has higher success than high weight payload
- Decision model has highest accuracy and higher scores and is the best model

Thank you!