IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

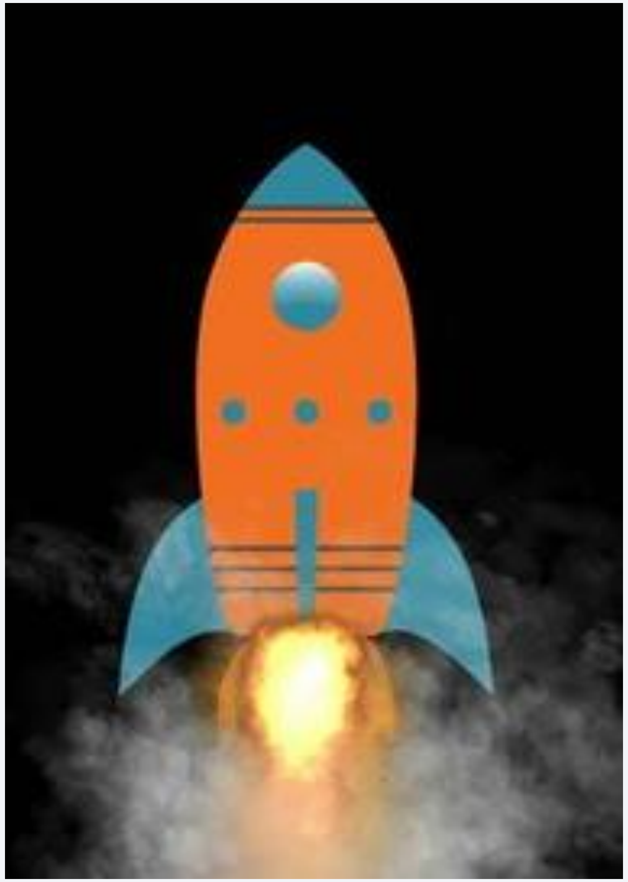
Abel Lucido
15 April 2024

Outline

Executive Summary

- Data Collection using get request
- Data Wrangling
- Exploratory Data Analysis using SQL
- Data Visualization
- Building Interactive Visual Analytics and Dashboards
- Predictive Analysis using Machine Learning algorithms

Introduction



Space exploration has unlocked the mysteries of the cosmos and expanded our understanding of the universe but it remains costly. 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. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Using requests.get then converting to json then to a dataframa using pandas
- Perform data wrangling
  - Removing nulls and data preparation by converting categorical variables to numerical variables
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Using machine learning algorithms such as Logistic regression, SVC, Decision Tree, and KNN

Data Collection

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

↓

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

↓

```
response.json()
```

↓

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

Data Collection – SpaceX API

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

⬇

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

⬇

```
response.json()
```

⬇

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

Data Collection - Scraping

```
static_url =
"https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"
```

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

```
soup = BeautifulSoup(response.content, 'html.parser')
```

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

Data Wrangling

Removing null data

Nan
Nan
Nan

Delete

Converting True/False to 1/0

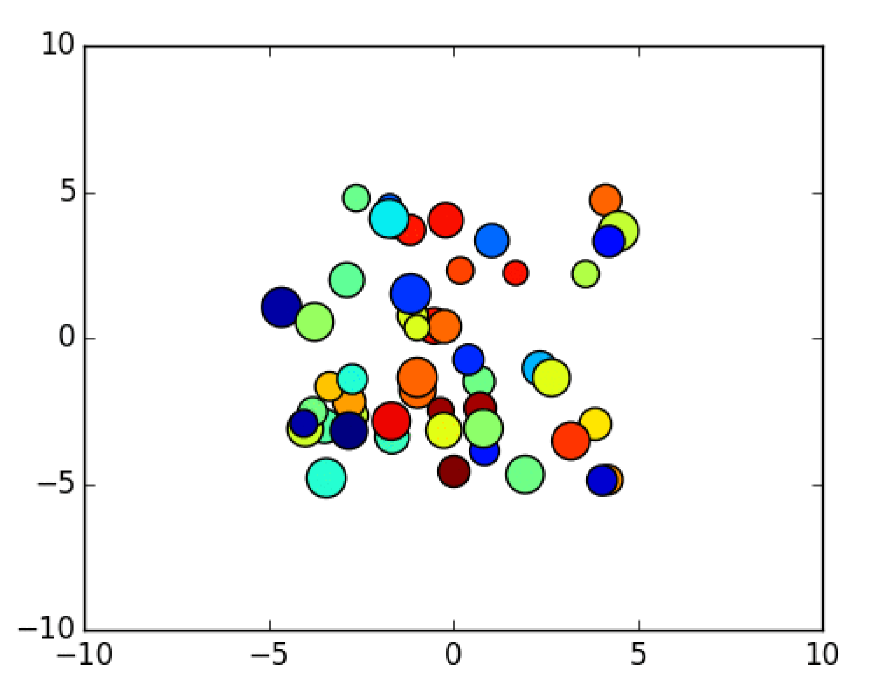
True
False
False

1
0
0

Github

https://github.com/abellucido/ibm_data_science_course/blob/main/Course10_Data_Wrangling.ipynb

EDA with Data Visualization



Github

https://github.com/abellucido/ibm_data_science_course/blob/main/Course10_EDA_with_Data_Visualization.ipynb

EDA with SQL

INPUT

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

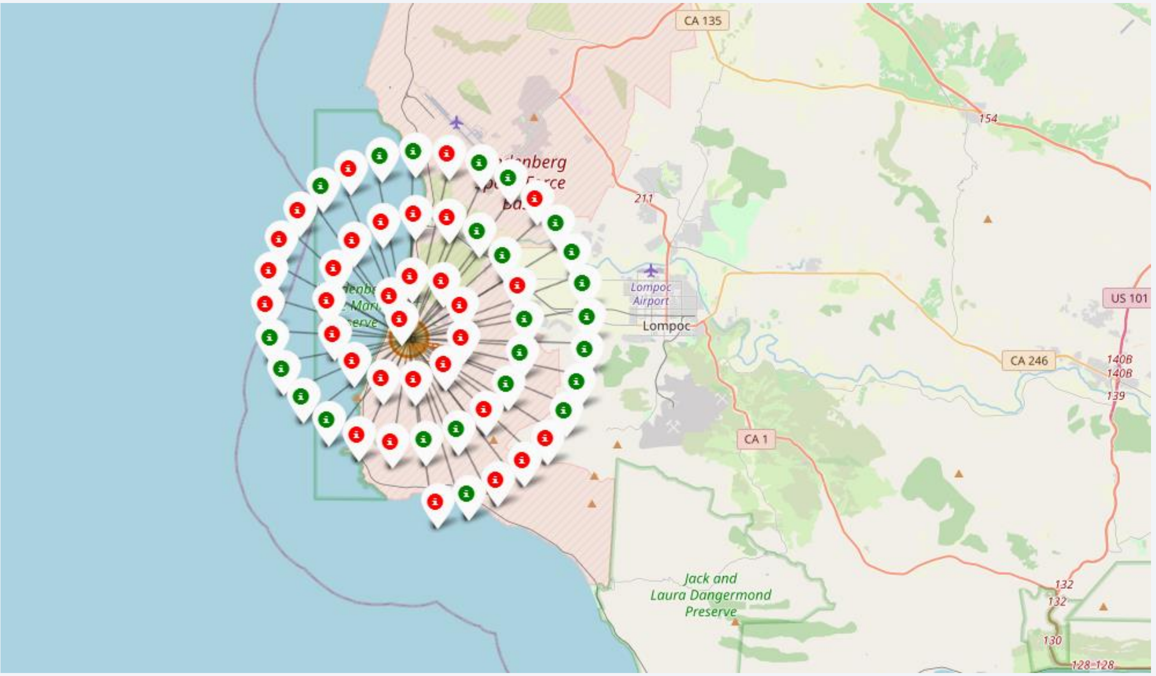
INPUT

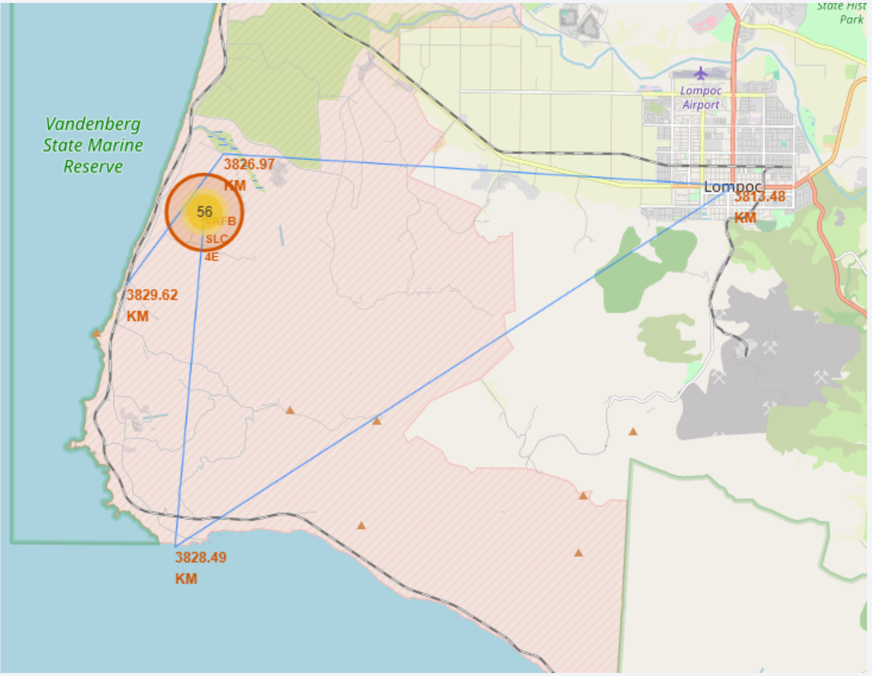
```
%sql SELECT
Booster_Version,
Landing_Outcome,
PAYLOAD_MASS__KG_ FROM
SPACEXTABLE WHERE
Landing_Outcome = "Success"
and PAYLOAD_MASS__KG_
BETWEEN 4000 AND 6000
```

Github

https://github.com/abellucido/ibm_data_science_course/blob/main/Course10_EDA_with_SQL.ipynb

Build an Interactive Map with Folium





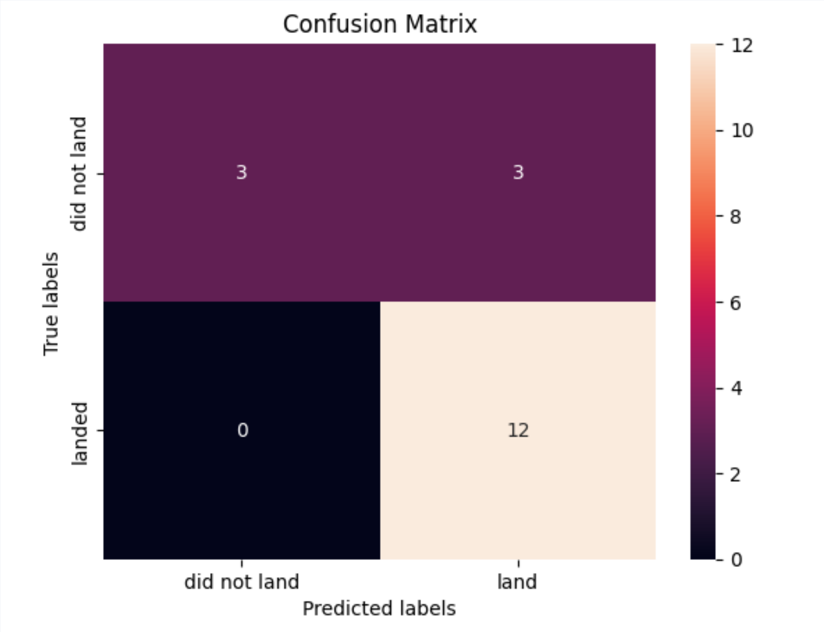
Github

https://github.com/abellucido/ibm_data_science_course/blob/main/Course10_Folium_interactive_map.ipynb

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

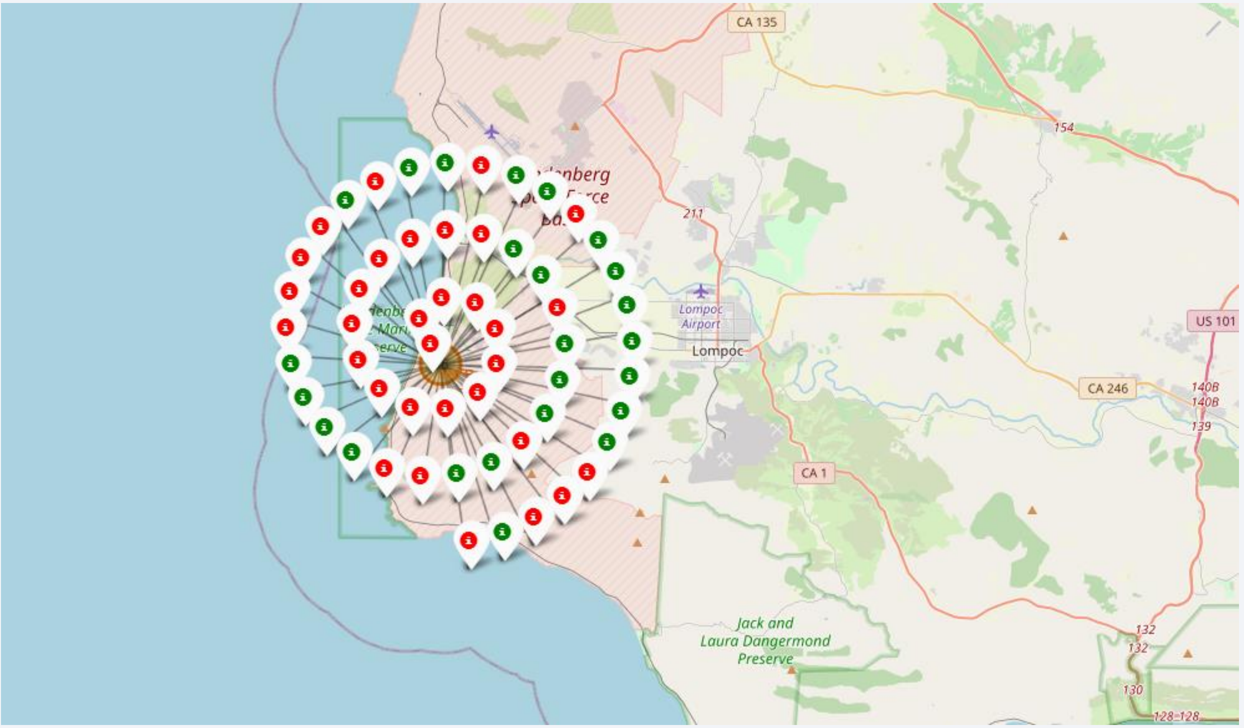
Predictive Analysis (Classification)

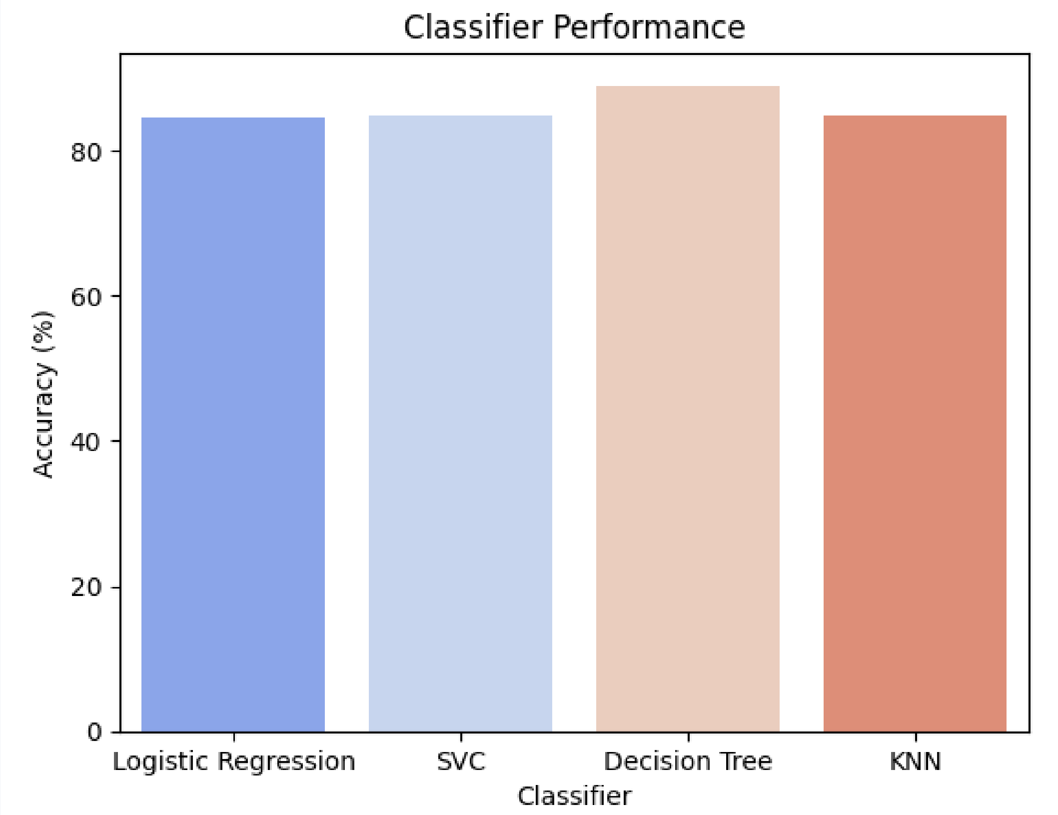


Github

https://github.com/abellucido/ibm_data_science_course/blob/main/Course10_Machine_Learning_Classification_Predictive_Analysis.ipynb

Results


CA 135
154
211
Lompoc Airport
Lompoc
US 101
CA 246
CA 1
Jack and Laura Dangermond Preserve


Classifier Performance
Accuracy (%)
0
20
40
60
80
Logistic Regression
SVC
Decision Tree
KNN
Classifier

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

Show a scatter plot of Flight Number vs. Launch Site

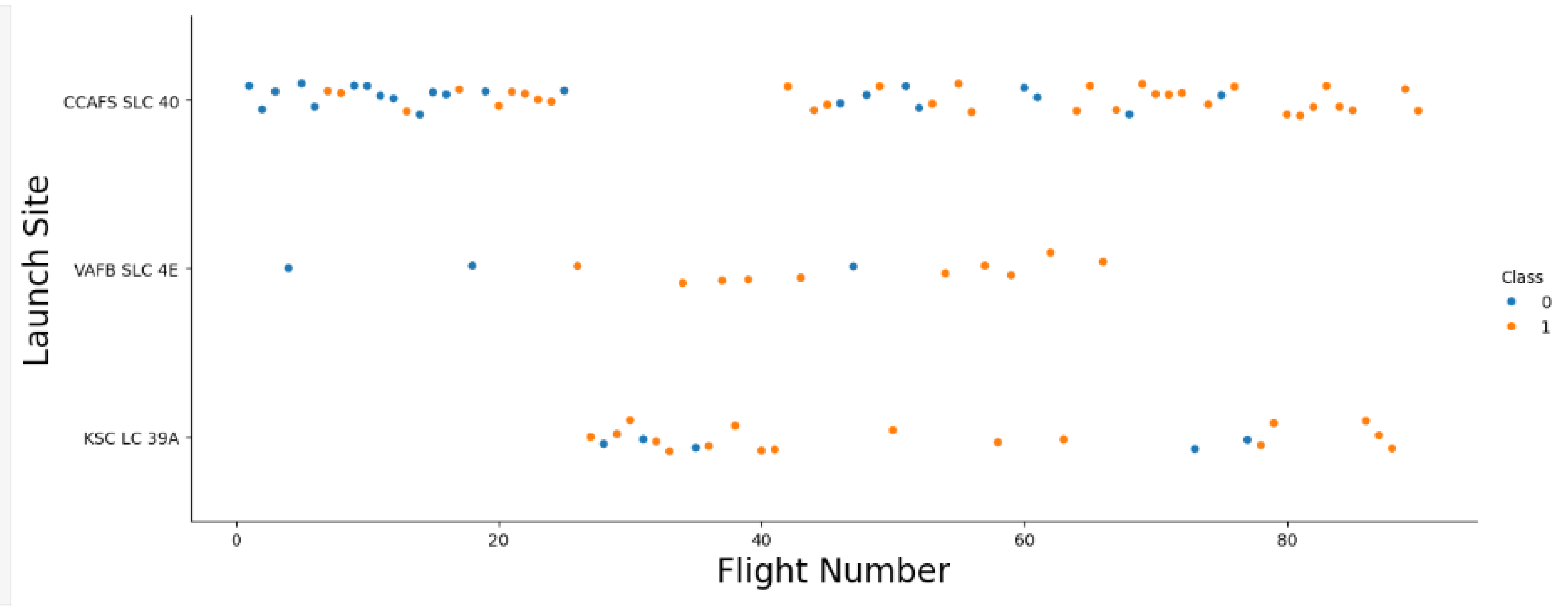


Figure . Scatter plot of Flight Number (x-axis) vs Launch Site (y –axis)

Payload vs. Launch Site

Show a scatter plot of Payload vs. Launch Site

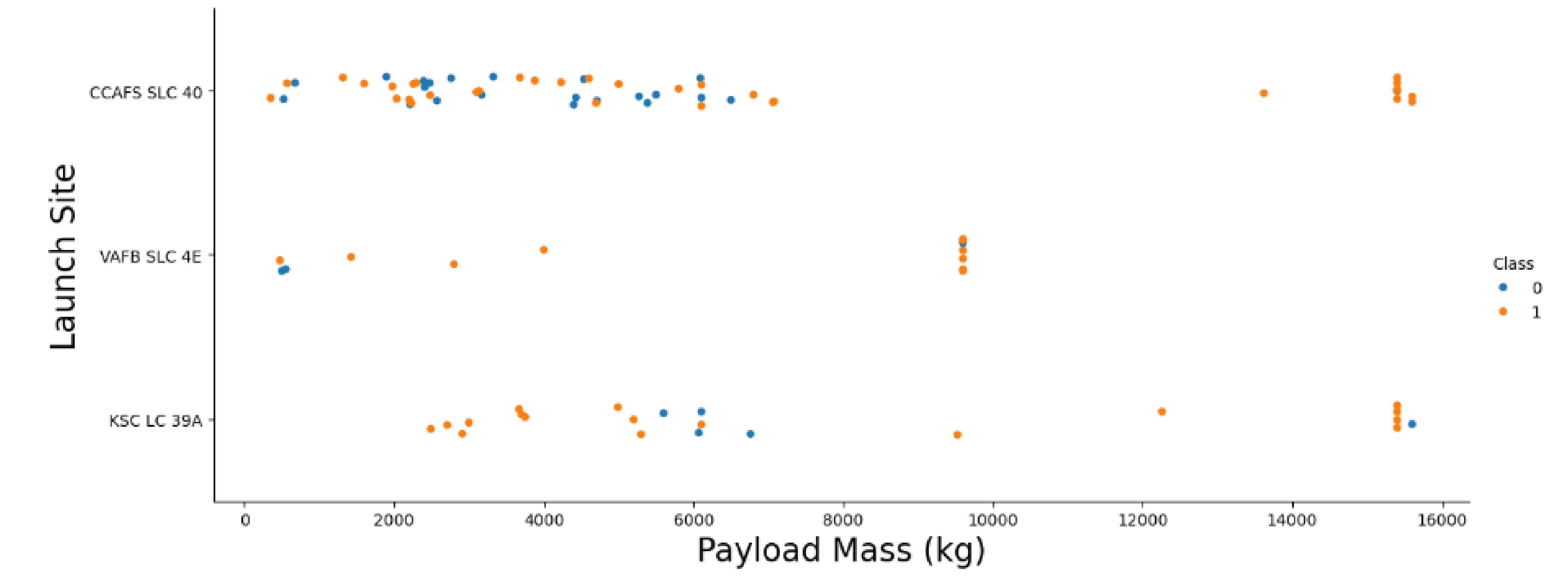
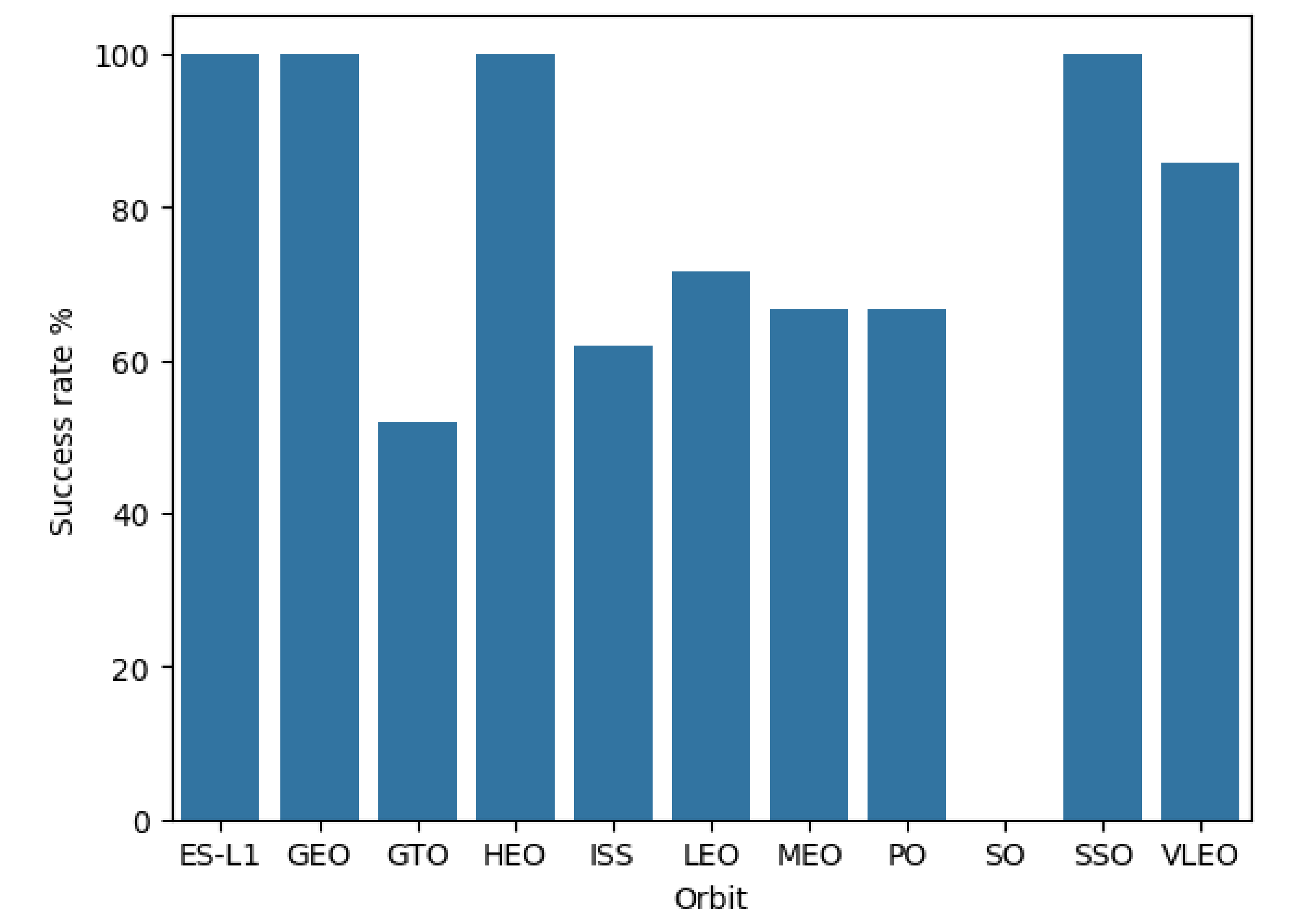


Figure . Scatter plot of Payload Mass (x-axis) vs Launch Site (y –axis)

Success Rate vs. Orbit Type



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

Figure . Bar plot of success rate per Orbit

Flight Number vs. Orbit Type

Show a scatter point of Flight number vs. Orbit type

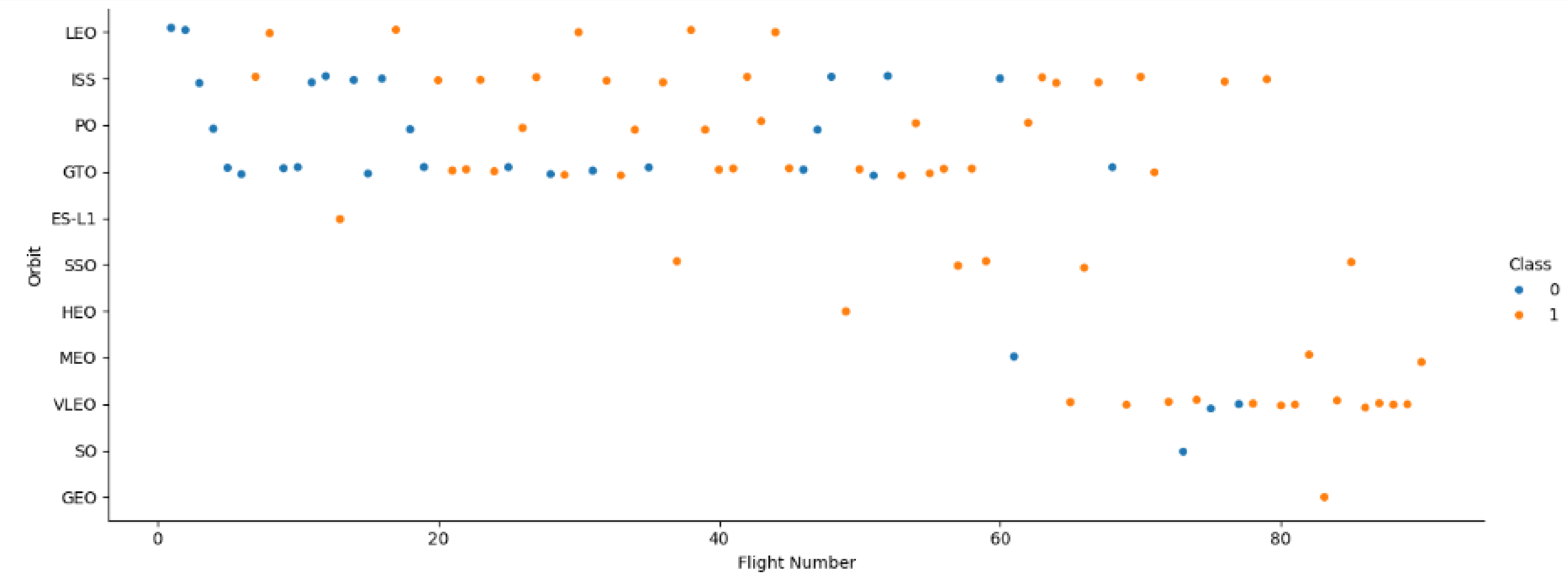


Figure . Scatter plot of Flight Number (x-axis) vs Orbit Type (y –axis)

Payload vs. Orbit Type

Show a scatter point of Flight number vs. Orbit type

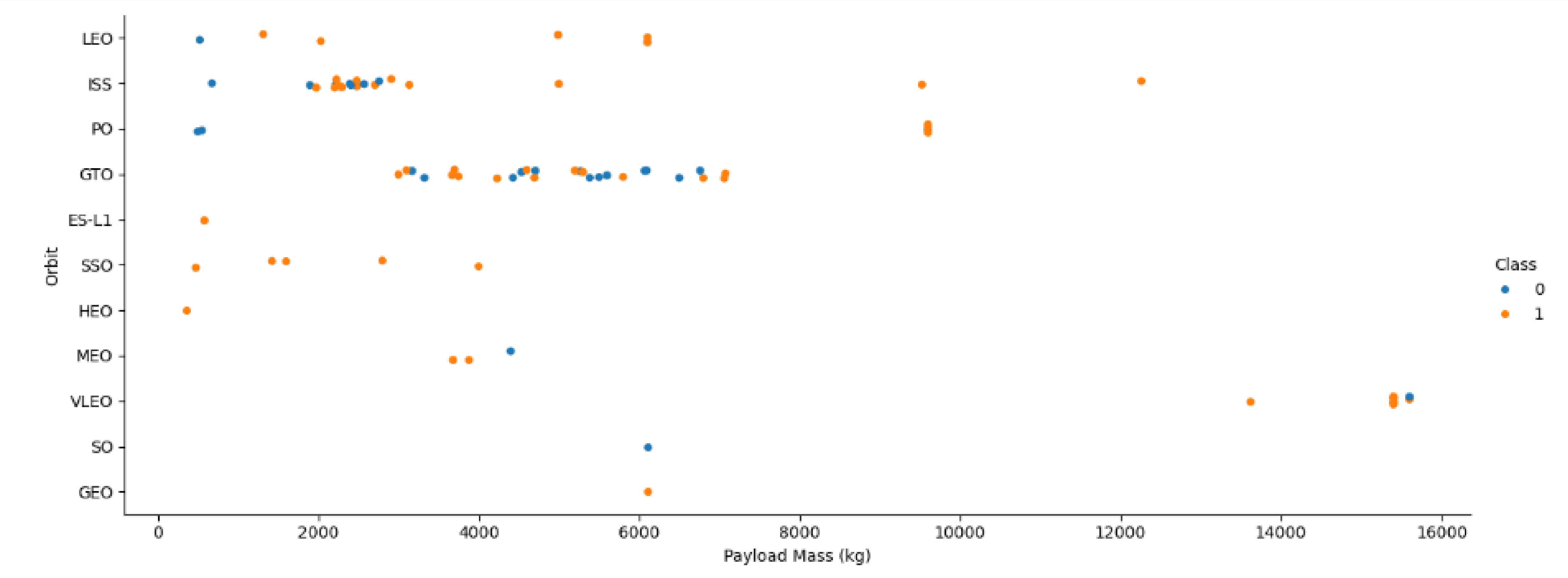
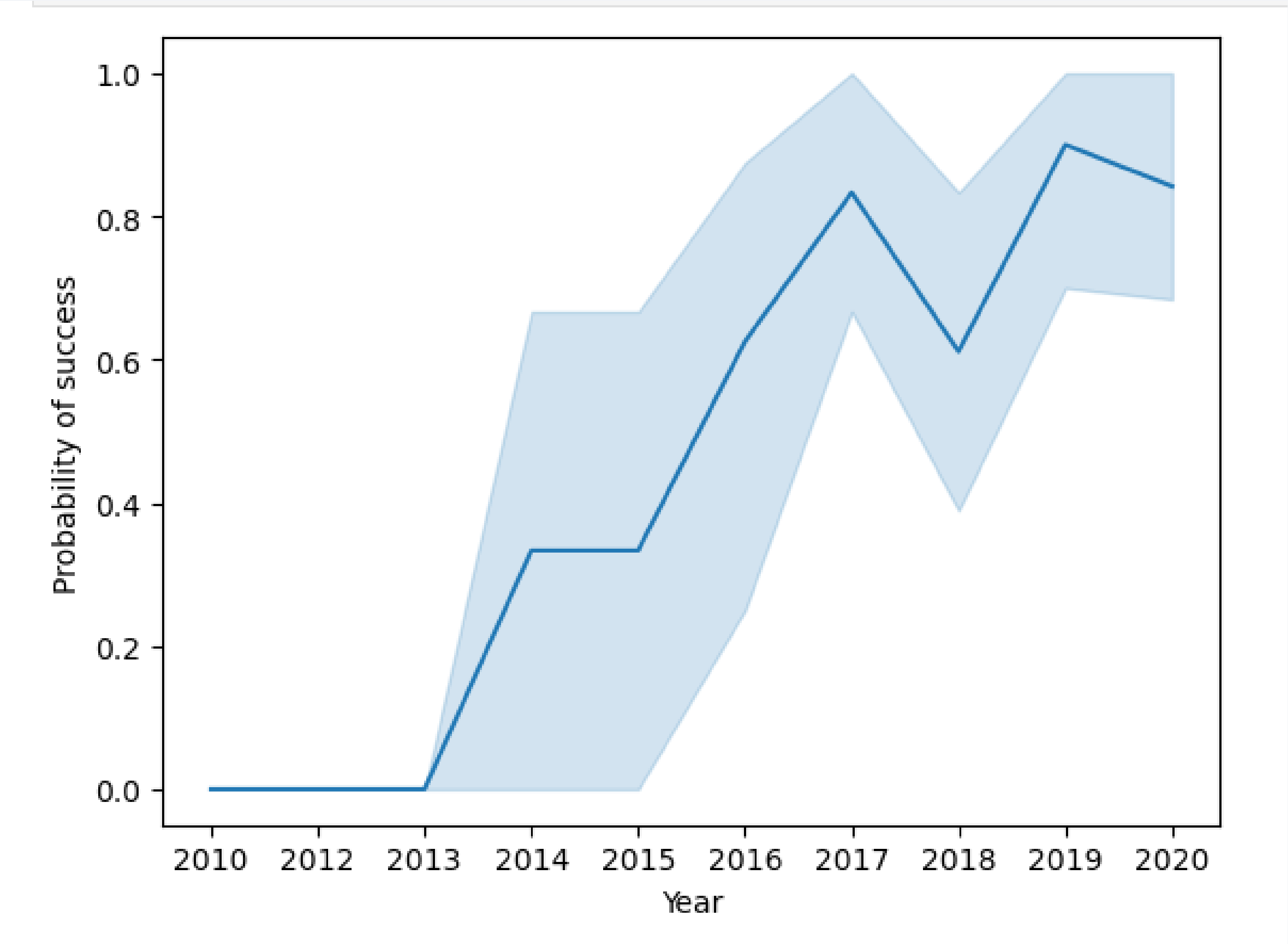


Figure . Scatter plot of Payload Mass (x-axis) vs Orbit Type (y –axis)

Launch Success Yearly Trend



Show a line chart of yearly average success rate

Figure . Line chart of yearly average success rate

All Launch Site Names

Find the names of the unique launch sites

INPUT

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

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

Launch Site Names Begin with 'CCA'

Find 5 records where launch sites begin with `CCA`

INPUT

```
%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

Total Payload Mass

Calculate the total payload carried by boosters from NASA

INPUT

```
%sql SELECT COUNT(*)
FROM  SPACEXTABLE WHERE
Customer = "NASA (CRS)"
```

COUNT(*)
20

Average Payload Mass by F9 v1.1

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

INPUT

```
%sql SELECT
AVG(PAYLOAD_MASS__KG_) FROM
SPACEXTABLE WHERE
Booster_Version Like "F9
v1.1"
```

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

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

INPUT

```
%sql SELECT MIN(Date)
FROM SPACEXTABLE WHERE
Landing_Outcome =
"Success"
```

MIN(Date)
2018-07-22

Successful Drone Ship Landing with Payload between 4000 and 6000

List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

INPUT

```
%sql SELECT
Booster_Version,
Landing_Outcome,
PAYLOAD_MASS__KG_ FROM
SPACEXTABLE WHERE
Landing_Outcome = "Success"
and PAYLOAD_MASS__KG_
BETWEEN 4000 AND 6000
```

Booster_Version	Landing_Outcome	PAYLOAD_MASS__KG_
F9 B5 B1046.2	Success	5800
F9 B5 B1047.2	Success	5300
F9 B5 B1046.3	Success	4000
F9 B5 B1048.3	Success	4850
F9 B5 B1051.2	Success	4200
F9 B5B1060.1	Success	4311
F9 B5 B1058.2	Success	5500
F9 B5B1062.1	Success	4311

Total Number of Successful and Failure Mission Outcomes

Calculate the total number of successful and failure mission outcomes

INPUT

```
%sql SELECT COUNT(Mission_Outcome) FROM
SPACEXTABLE WHERE Mission_Outcome =
"Success"
```

COUNT(Mission_Outcome)
98

INPUT

```
%sql SELECT COUNT(Mission_Outcome) FROM
SPACEXTABLE WHERE Mission_Outcome !=
"Success"
```

COUNT(Mission_Outcome)
3

Boosters Carried Maximum Payload

List the names of the booster which have carried the maximum payload mass

INPUT

```
%sql SELECT 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

2015 Launch Records

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

INPUT

```
%sql SELECT substr(Date,
6,2), Landing_Outcome,
Booster_Version,
Launch_Site FROM
SPACEXTABLE WHERE
substr(Date,0,5)='2015'
```

substr(Date, 6,2)	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
02	Controlled (ocean)	F9 v1.1 B1013	CCAFS LC-40
03	No attempt	F9 v1.1 B1014	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
04	No attempt	F9 v1.1 B1016	CCAFS LC-40
06	Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40
12	Success (ground pad)	F9 FT B1019	CCAFS LC-40

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

INPUT

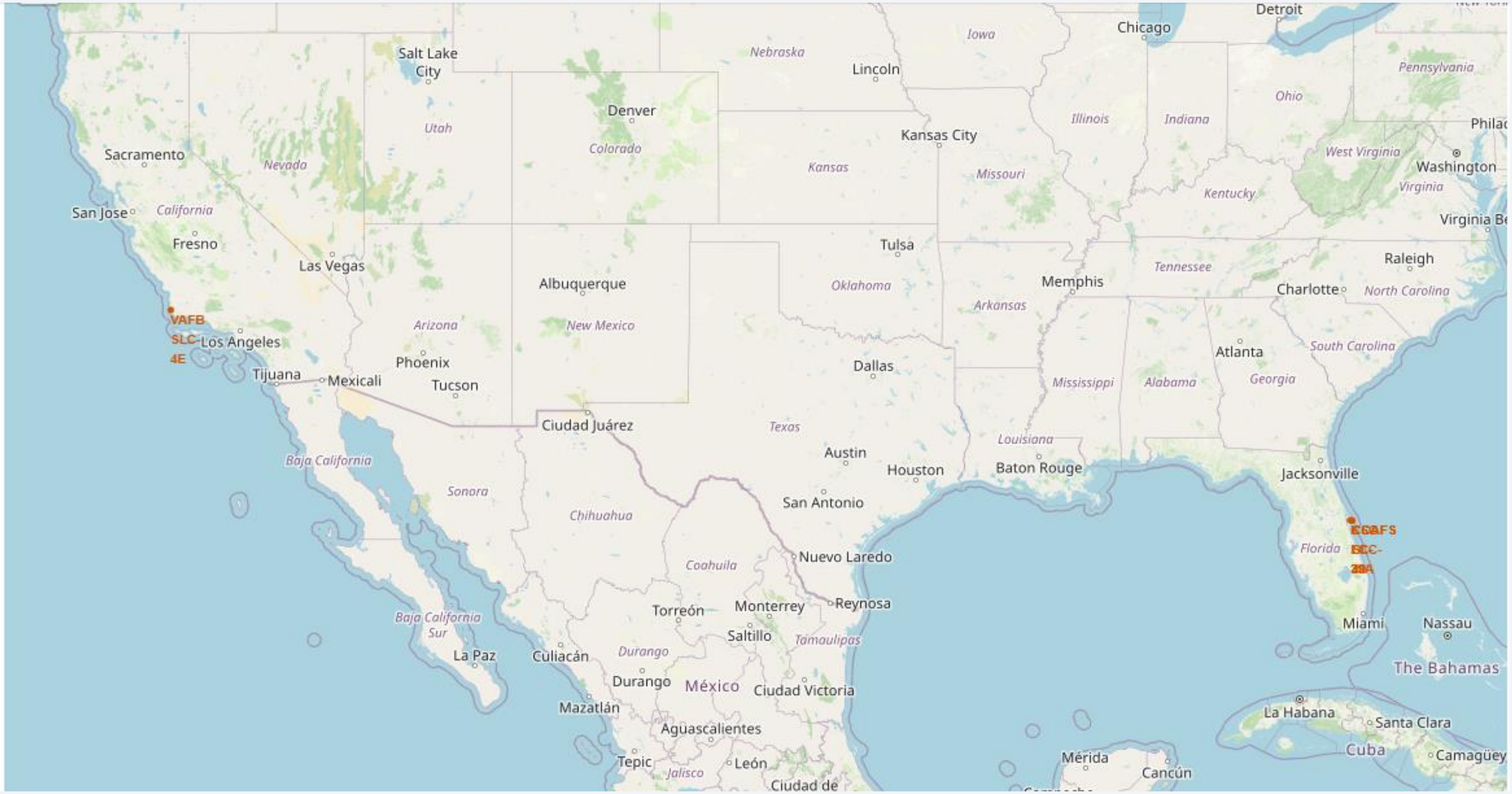
```
%sql SELECT * , count(Landing_Outcome) as CLO FROM SPACEXTABLE WHERE Date BETWEEN
'2010-06-04' AND '2017-03-20' GROUP BY Landing_Outcome Order BY CLO DESC
```

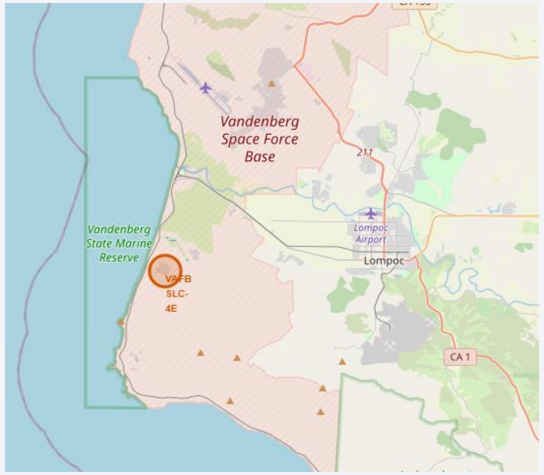
Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome	CLO
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt	10
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)	5
2015-01-10	9:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)	5
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)	3
2014-04-18	19:25:00	F9 v1.1	CCAFS LC-40	SpaceX CRS-3	2296	LEO (ISS)	NASA (CRS)	Success	Controlled (ocean)	3
2013-09-29	16:00:00	F9 v1.1 B1003	VAFB SLC-4E	CASSIOPE	500	Polar LEO	MDA	Success	Uncontrolled (ocean)	2
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)	2
2015-06-28	14:21:00	F9 v1.1 B1018	CCAFS LC-40	SpaceX CRS-7	1952	LEO (ISS)	NASA (CRS)	Failure (in flight)	Precluded (drone ship)	1

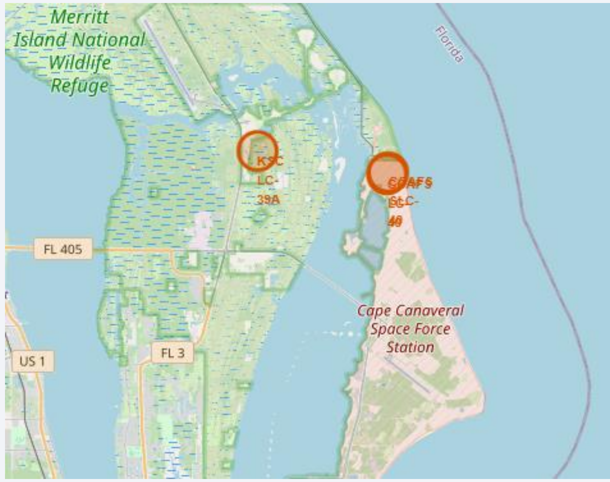
Section 3

Launch Sites Proximities Analysis

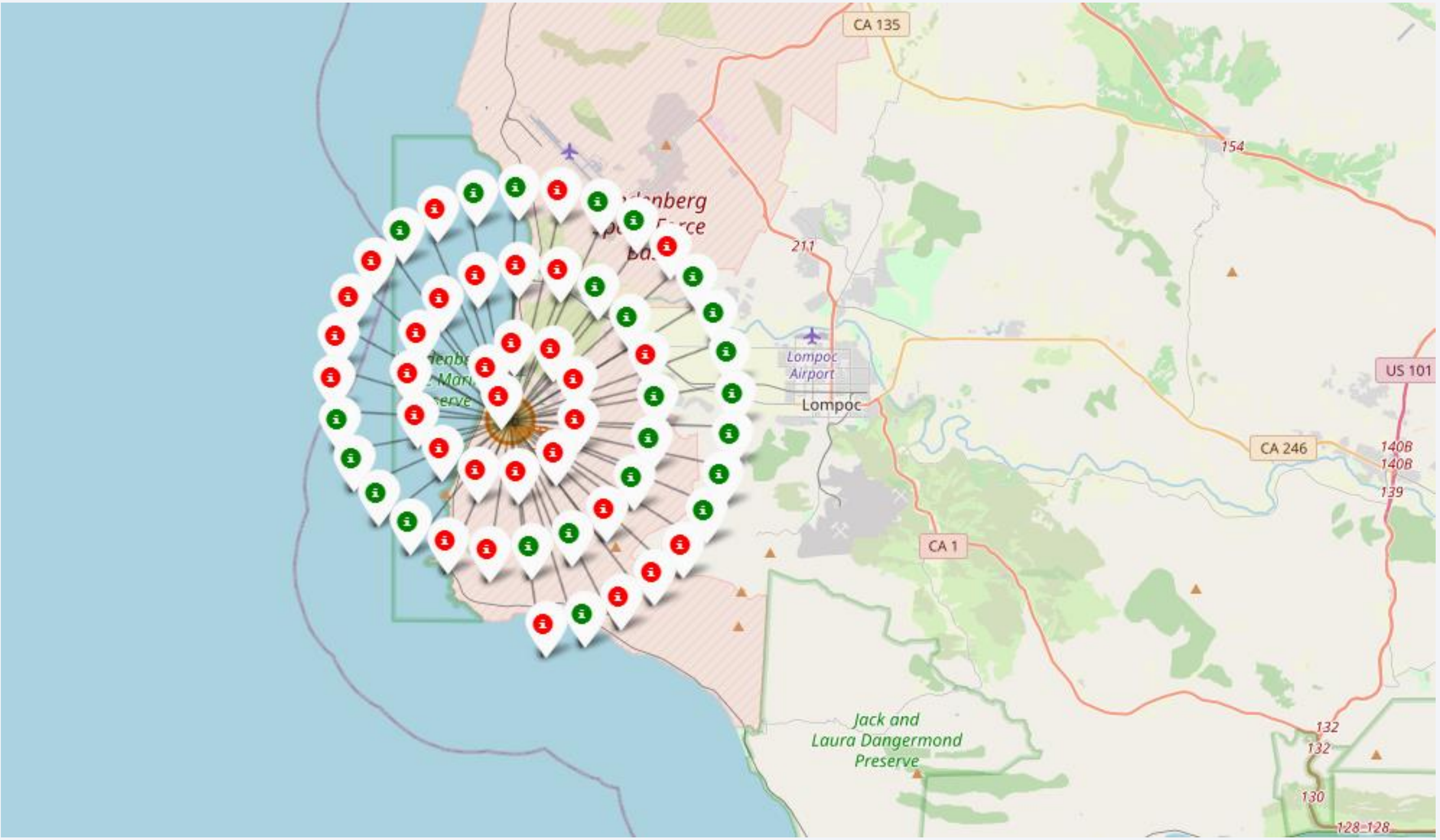
Space X Launch Sites



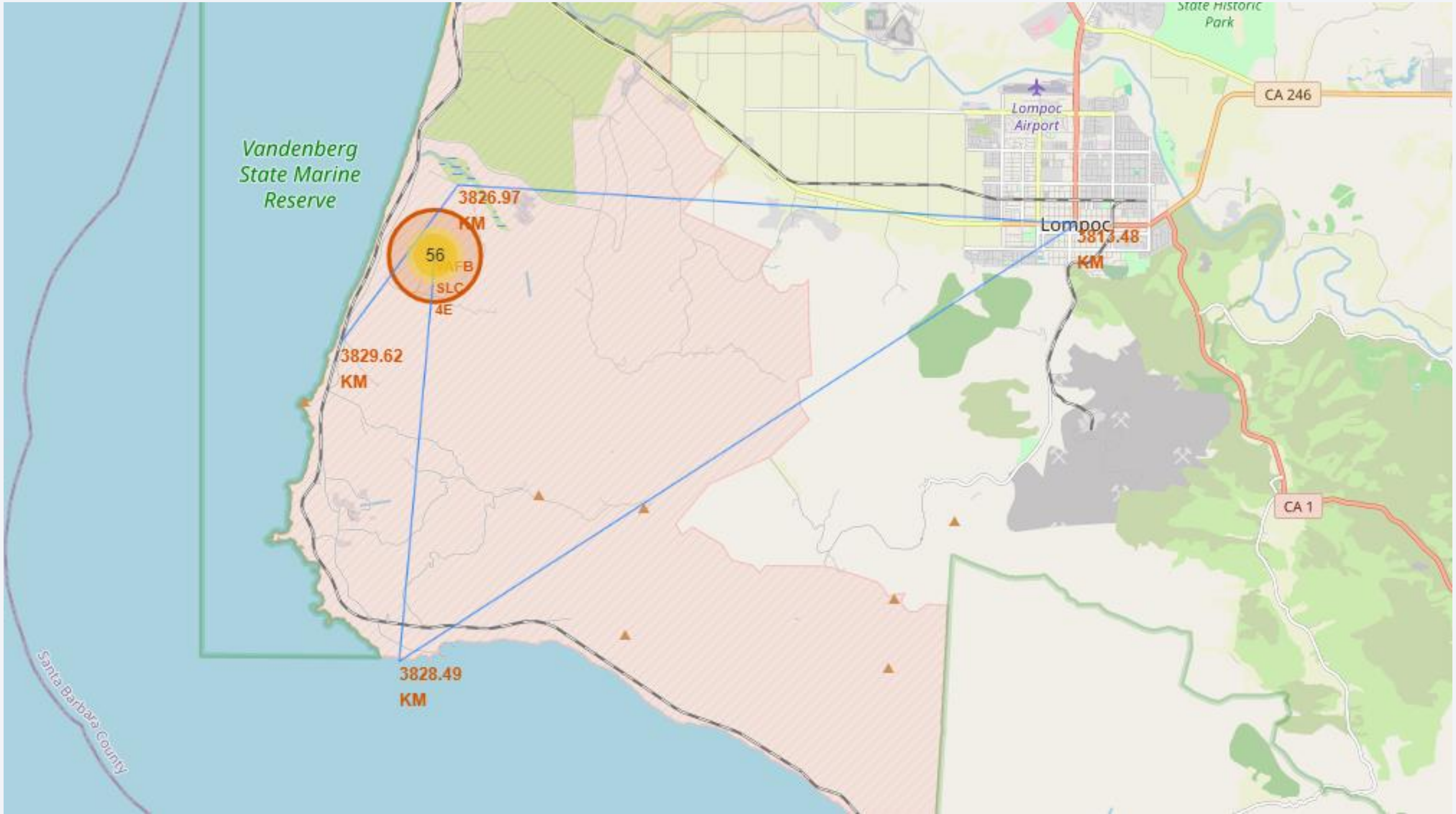




Success/Failed Launches



Launch Site closest infrastructure



Section 4

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title

- Show the screenshot of launch success count for all sites, in a piechart

- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title

- Show the screenshot of the piechart for the launch site with highest launch success ratio

- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title

- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider

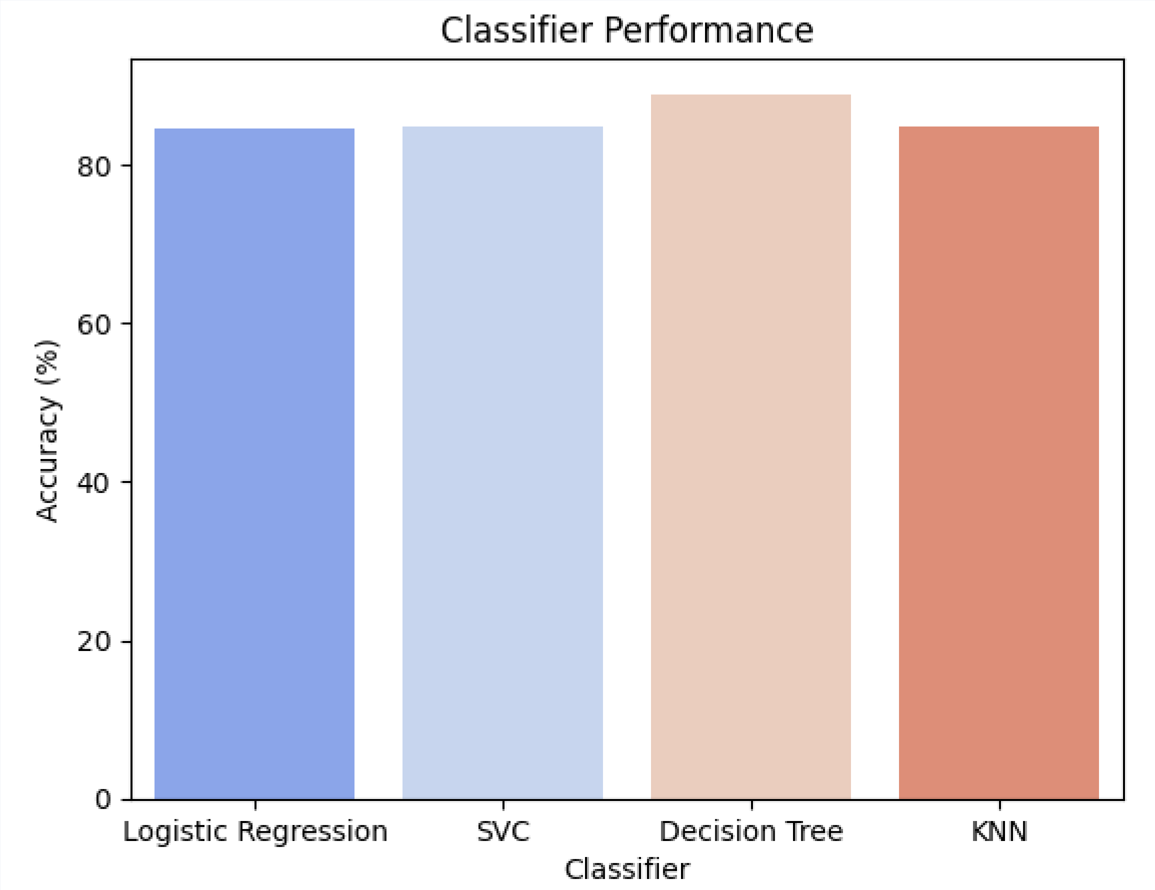
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

Predictive Analysis (Classification)

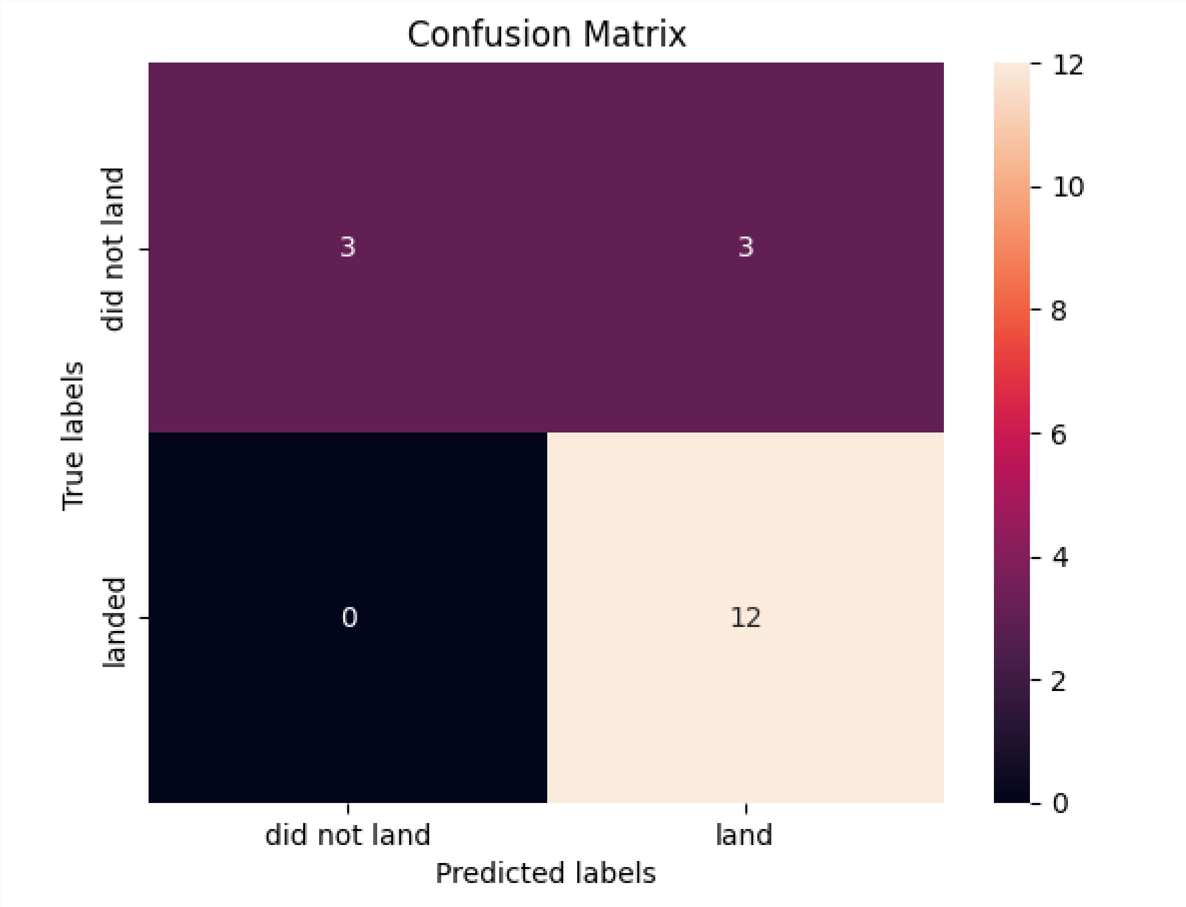
Classification Accuracy

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



Confusion Matrix

Decision Tree Classifier



Conclusions

We were able to identify the optimal launching site with the highest probability of success.

We were able to identify that Decision Tree Classifier has the highest accuracy of 88.93% among the machine learning algorithms in predicting if the landing outcome will be successful.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!