Predictive Analysis on Space X Falcon 9 , 1st Stage

Presented by Bishwas Sherchan

Executive Summary

The purpose of this report is to conduct predictive analysis on successful landing of the first stage booster from Falcon 9, Space X. Based on our predictive model created using historical data, we have concluded that there is **~84%** chance that the first stage of falcon 9 will land back on earth.

Outline

- Introduction
- Methodology
- Results
- Discussion
- Conclusion
- Appendix

Introduction

Background Information:

- 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.

Goals:

- Conduct data analysis and build a predictive model to report likelihood of first stage landing

Data Collection

- Data was primarily collected by making a get request to the SPACEX API.
- The response from get request was in JSON format. Hence, it was turned into Pandas data frame to make data preprocessing simpler.
- In addition, data was also collected by web scraping Falcon 9 wiki page. HTTP get method was used to request the HTML page and by creating a Beautiful Soup object, relevant data were extracted and turned into a Pandas data frame for preprocessing.

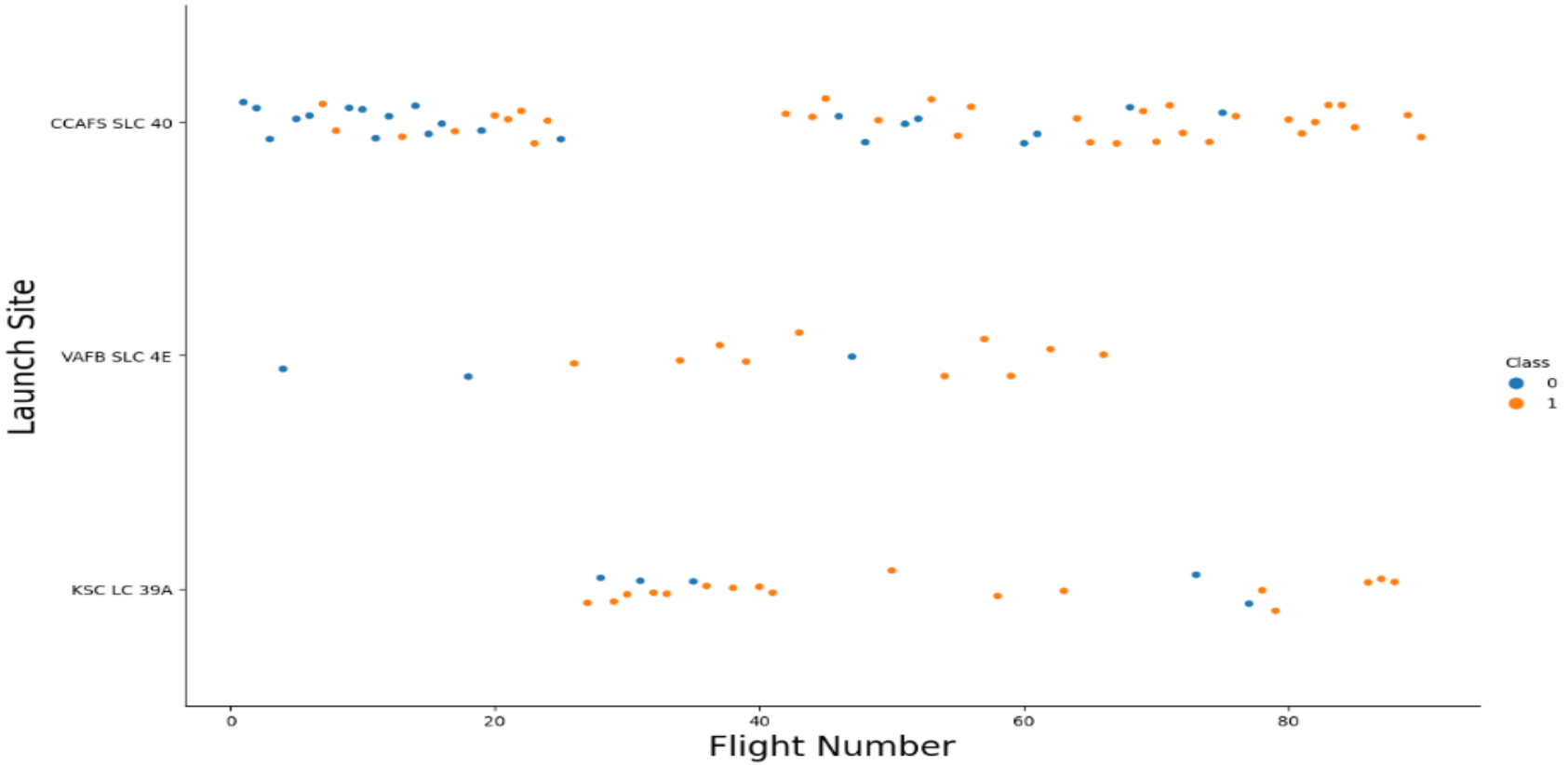
Data Wrangling and Feature Engineering

- Two columns (payload mass and landing pad) in the data frame consisted of null values.
- Null values in payload mass were replaced with mean value whereas no action was taken for landing pad since it mean it landed on the ground.
- Data were also filtered to ensure it only included falcon 9 data
- One hot encode was also performed to turn categorical variables to numerical.
- Boolean values were also replaced by the numerical values.

```
FlightNumber      0
Date              0
BoosterVersion    0
PayloadMass       5
Orbit             0
LaunchSite        0
Outcome           0
Flights           0
GridFins          0
Reused            0
Legs              0
LandingPad       26
Block             0
ReusedCount       0
Serial            0
Longitude         0
Latitude          0
```

Exploratory Data Analysis with Visualizations

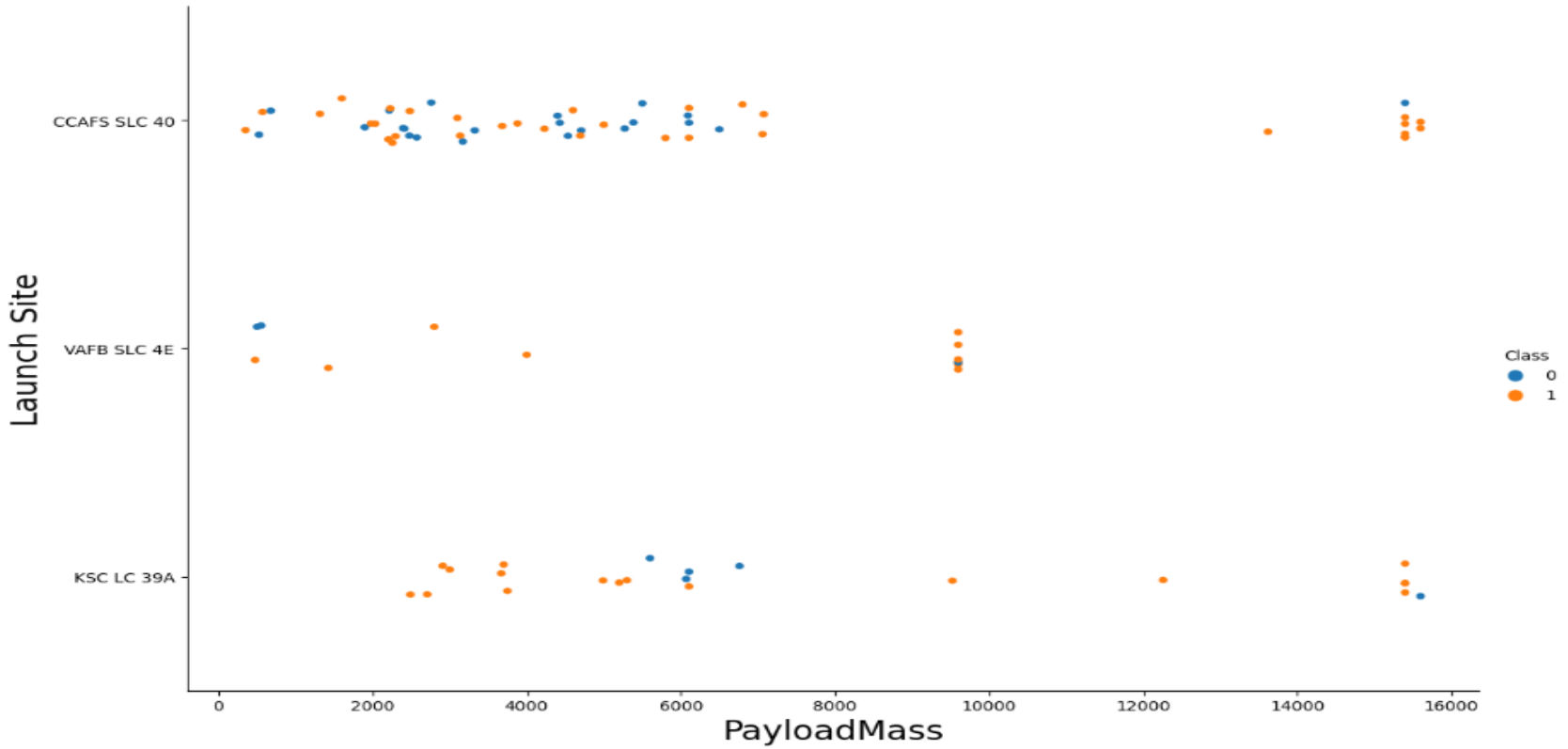
Relationship Between Flight Number and Launch Site



- In this categorical plot, you can see that as the flight number increases, and successful landing is generally increasing as well

Exploratory Data Analysis with Visualizations

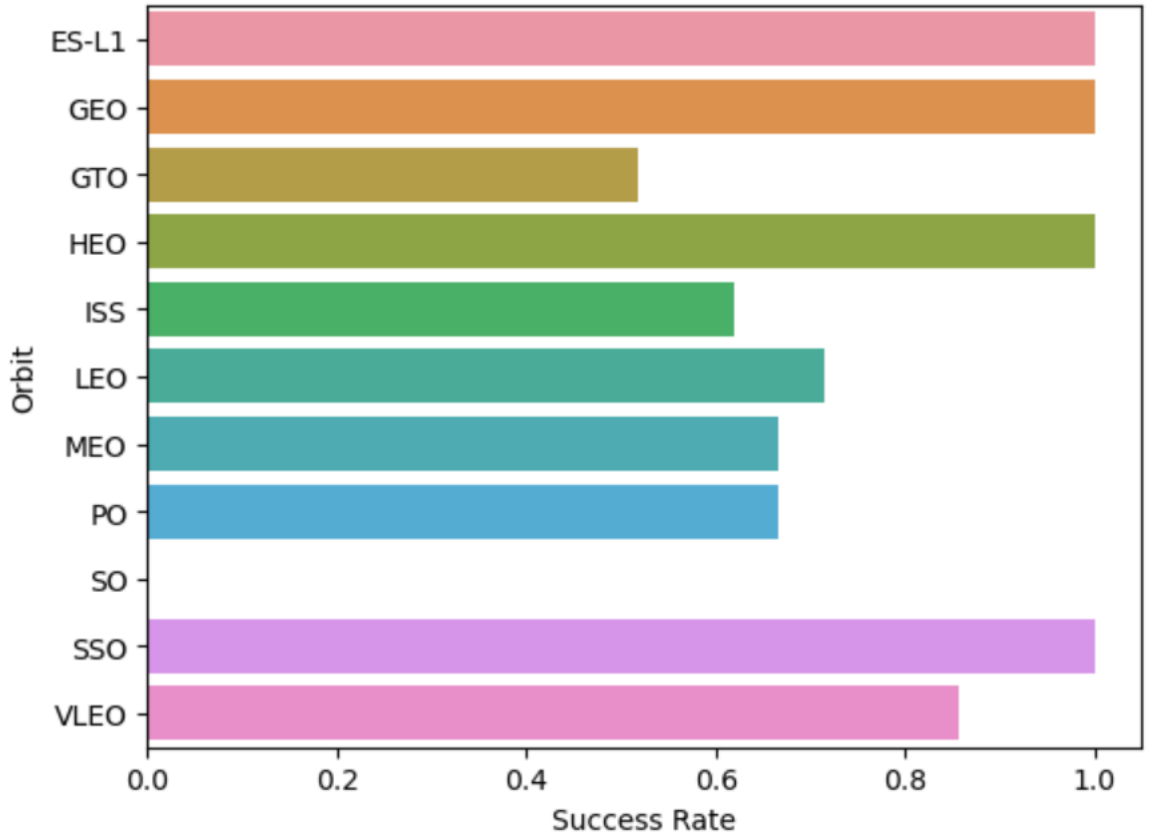
Relationship Between Pay load and Launch Site



- In this categorical plot, you can see that as the payload increases, and there is a high success rate for landing

Exploratory Data Analysis with Visualizations

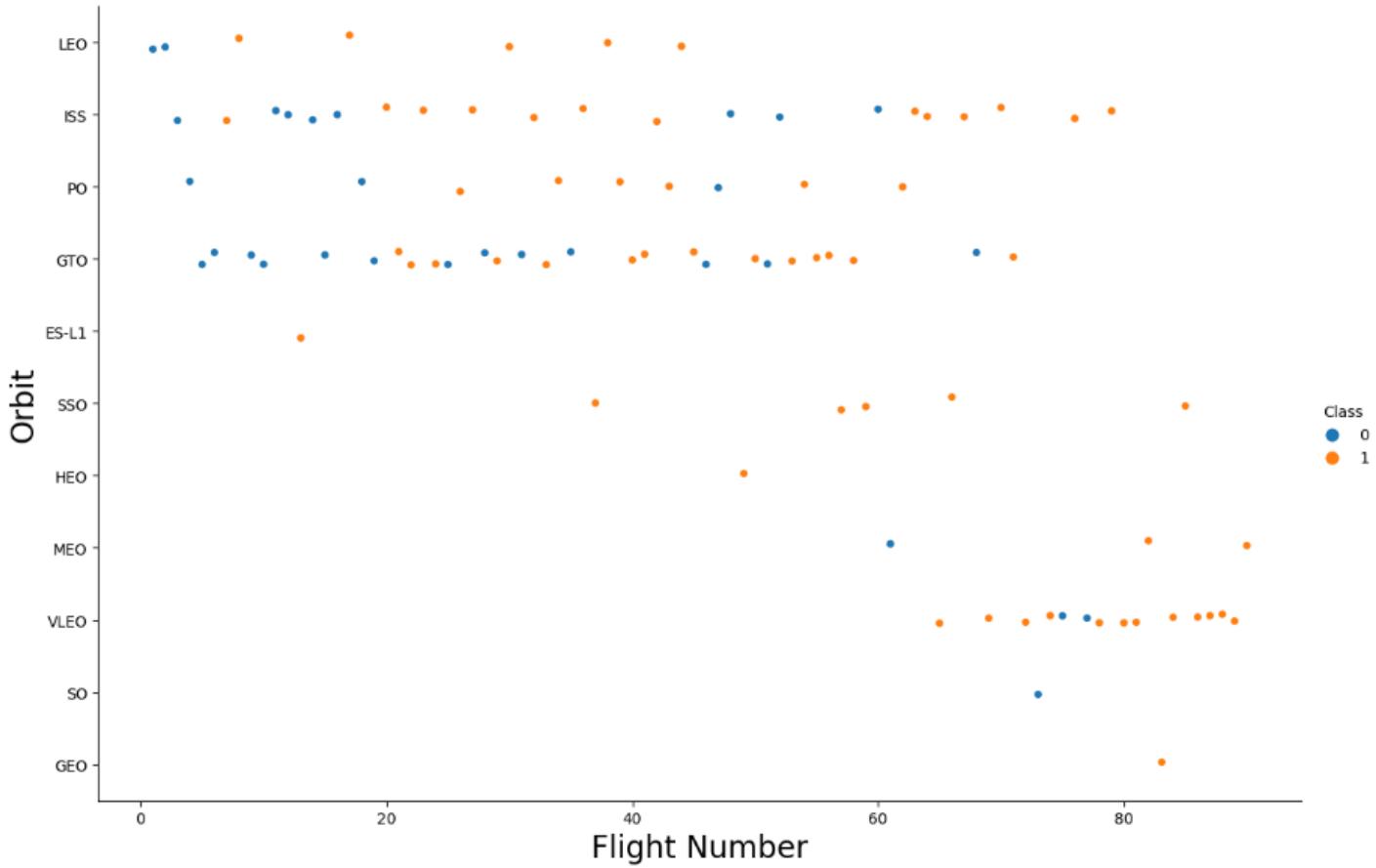
Relationship between Orbit and Success Rate



- In this bar plot, you can see success rate of landing first stage based on what orbit they launch too.

Exploratory Data Analysis with Visualizations

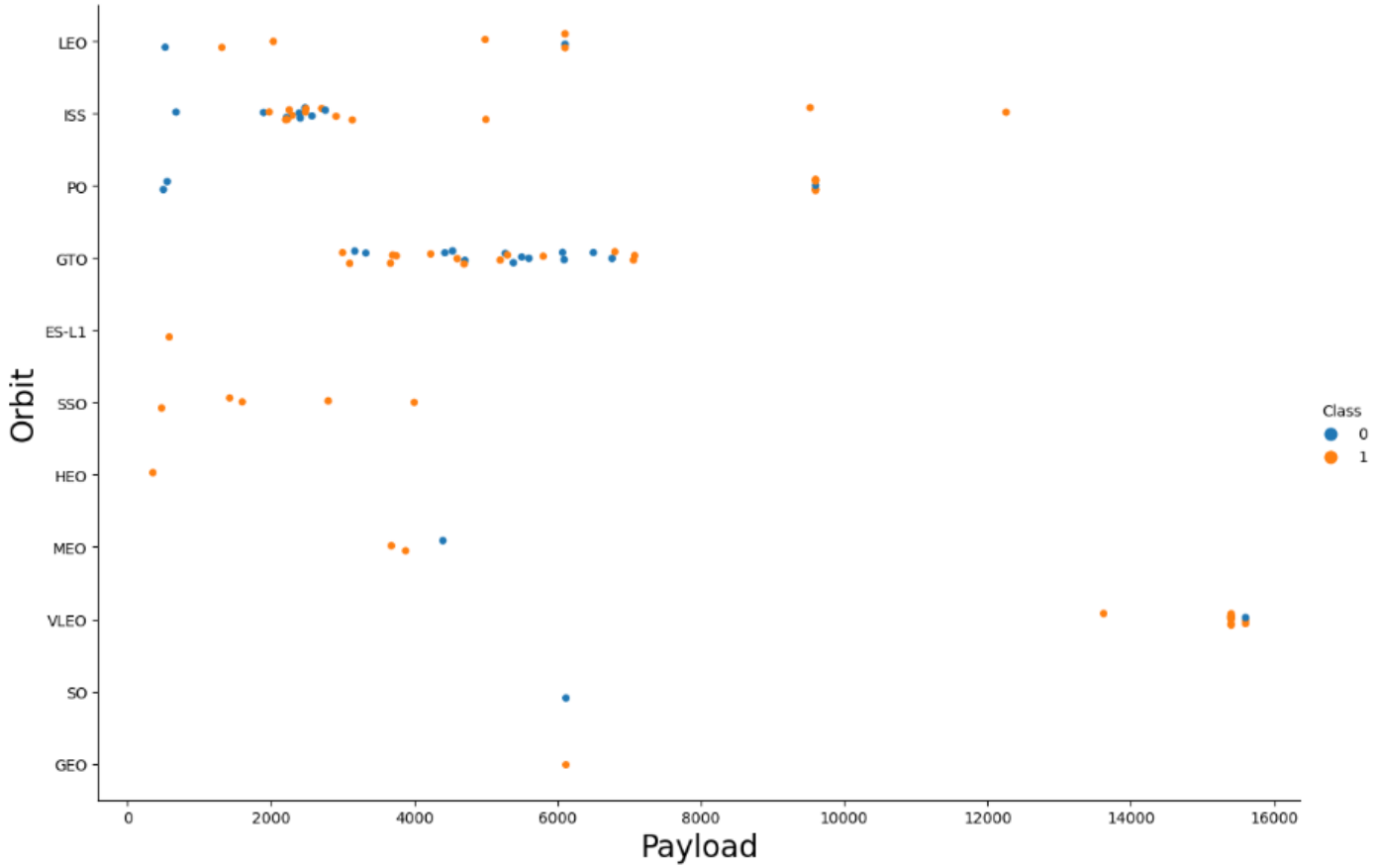
Relationship between flight number and Orbit



- In this categorical plot, you see generally that there seems to be no relation between flight number and orbit in terms of successful landing. However, it is worth pointing out that LEO orbit success appears to be related to the number of flight.

Exploratory Data Analysis with Visualizations

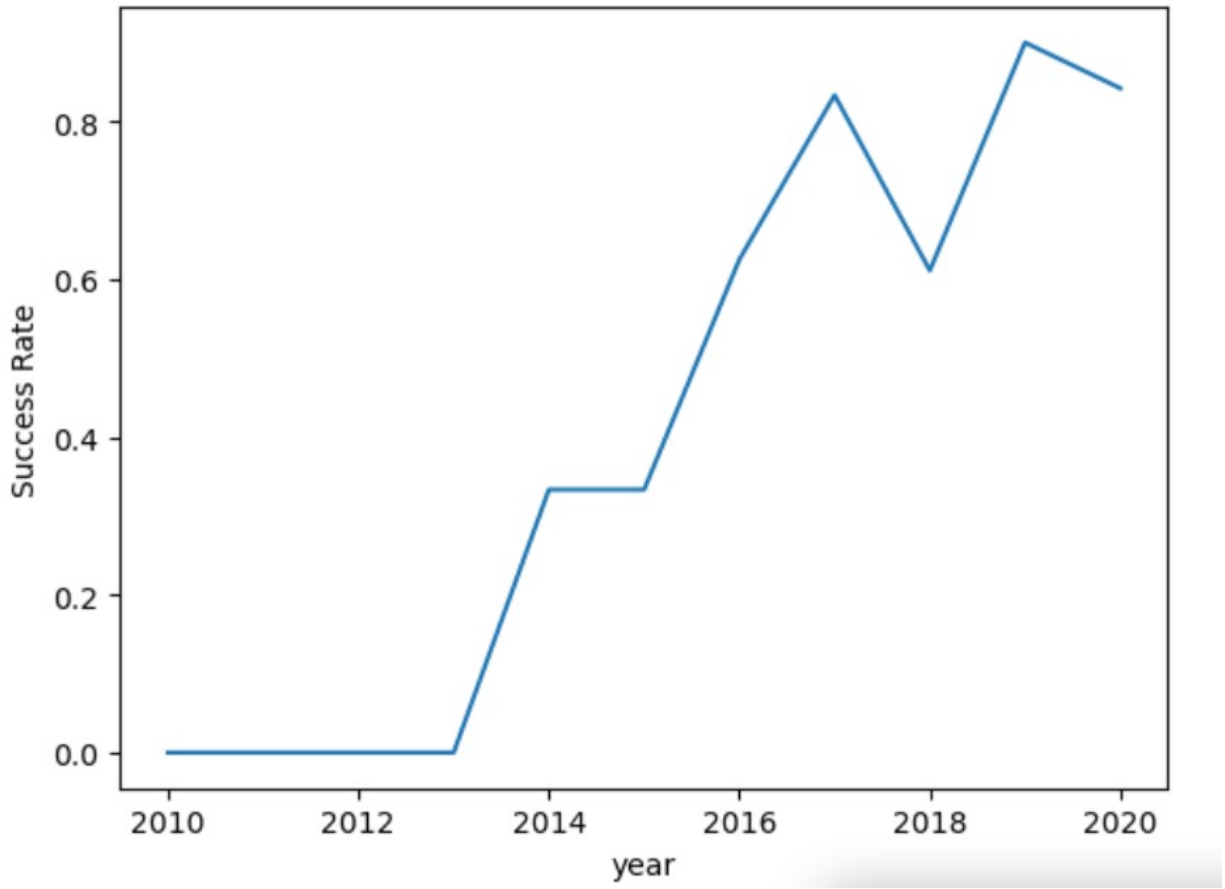
Relationship between payload and Orbit



- In this categorical plot, you see generally that there seems to be relation between payload and orbit in terms of successful landing. Higher the load , higher the success rate. However, it is worth pointing out that GTO orbit success appears to have no relation with payload.

Exploratory Data Analysis with Visualizations

Relationship between years and success rate



- After 2013, it is evident that as years progresses, there is higher success rate. This is probably because of improving and learning from their past mistakes.

Exploratory Data Analysis with SQL

Total Launch Sites

In [38]:
```
%%sql
SELECT Launch_Site, COUNT(*) FROM SPACEXTBL
GROUP BY Launch_Site
```

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

Out[38]:

Launch_Site	COUNT(*)
None	898
CCAFS LC-40	26
CCAFS SLC-40	34
KSC LC-39A	25
VAFB SLC-4E	16

- There are 4 launch sites. Disregard the None Values as it did not contain any informations.

Exploratory Data Analysis with SQL

Launch Site starting with CCA

```
%%sql
SELECT * FROM SPACEXTBL
WHERE Launch_Site LIKE 'CCA%'
LIMIT 5
```

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Custc
06/04/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0.0	LEO	Spa
12/08/2010	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0.0	LEO (ISS)	N (C
22/05/2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525.0	LEO (ISS)	N (C
10/08/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500.0	LEO (ISS)	N (
03/01/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677.0	LEO (ISS)	N (

Exploratory Data Analysis with SQL

Total payload mass carried by boosters launched by NASA (CRS)

```
%%sql
SELECT Customer,sum(PAYLOAD_MASS__KG_) FROM SPACEXTBL
GROUP BY Customer
HAVING Customer ='NASA (CRS)'
```

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

Customer	sum(PAYLOAD_MASS__KG_)
NASA (CRS)	45596.0

Exploratory Data Analysis with SQL

Average payload mass carried by booster version F9 v1.1

```
%%sql
SELECT Landing_Outcome FROM SPACEXTBL
```

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

Booster_Version	avg(PAYLOAD_MASS__KG_)
F9 v1.1	2928.4

Exploratory Data Analysis with SQL

First landing on ground pad

In [90]:
```
%%sql
SELECT Date, Landing_outcome FROM SPACEXTBL
WHERE Landing_outcome = 'Success (ground pad)'
```

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

Out[90]:

Date	Landing_Outcome
22/12/2015	Success (ground pad)
18/07/2016	Success (ground pad)
19/02/2017	Success (ground pad)
05/01/2017	Success (ground pad)
06/03/2017	Success (ground pad)
14/08/2017	Success (ground pad)
09/07/2017	Success (ground pad)
15/12/2017	Success (ground pad)
01/08/2018	Success (ground pad)

- First landing on ground pad happened in 22/12/2015.

Exploratory Data Analysis with SQL

Boosters successful in drone ship (payload greater than 4000 and less than 6000

```
%%sql
SELECT Booster_version FROM SPACEXTBL
WHERE Landing_outcome = 'Success (drone ship)'
AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_< 6000;
```

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

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

- 4 Boosters that had payload of more than 4000 and less than 6000 successfully landed in drone ship.

Exploratory Data Analysis with SQL

Total number of sucessfull and failure mission outcomes

```
%%sql
UPDATE SPACEXTBL
SET Mission_Outcome = 'Success'
WHERE Mission_outcome = 'Success '
```

```
 * sqlite:///my_data1.db
0 rows affected.

[]
```

```
%%sql
SELECT Mission_Outcome, COUNT(*) FROM SPACEXTBL
GROUP BY Mission_Outcome
HAVING Mission_Outcome IN ('Success','Failure (in flight)')
```

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

Mission_Outcome	COUNT(*)
Failure (in flight)	1
Success	99

- There 99 success and 1 failure mission outcomes.

Exploratory Data Analysis with SQL

Total Failure landing in drone ship in 2015

```
%%sql
SELECT SUBSTR(Date,4,2) as month, Landing_outcome,Booster_Version, Launch_s
FROM SPACEXTBL
WHERE Year = '2015' and Landing_outcome = 'Failure (drone ship)'
```

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

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

- There were 2 failures landing in drone ship in 2015

Exploratory Data Analysis with SQL

Ranking successful landing between 04-06-2010 and 20-03-2017

```
%%sql
SELECT RANK() OVER (ORDER BY Date) as Rank, * FROM SPACEXTBL
WHERE Landing_Outcome LIKE 'Success%' AND Date BETWEEN '04-06-2010' AND '20
```

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

Rank	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orb
1	04/08/2016	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136.0	LE((ISS
2	05/01/2017	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300.0	LE(
3	05/04/2019	6:48:00	F9 B5B1056.1	CCAFS SLC-40	SpaceX CRS-17, Starlink v0.9	2495.0	LE((ISS
4	05/06/2016	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696.0	GT(
5	05/11/2018	20:14:00	F9 B5 B1046.1	KSC LC-39A	Bangabandhu-1	3600.0	GT(
					SpaceX CRS-		LE(

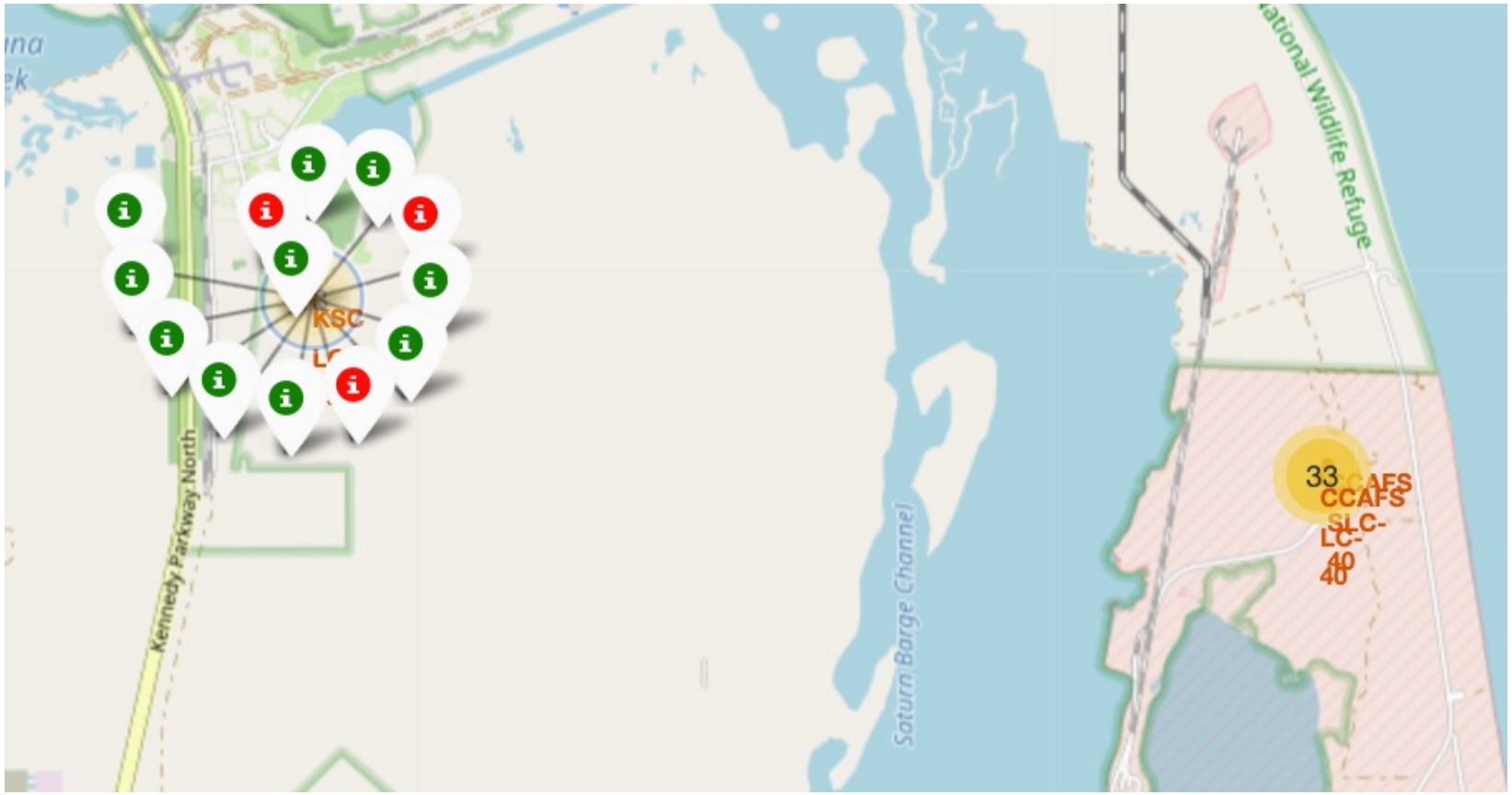
- There were 35 successful landings during the timeframe.

Launch Site Locations

4 Launch Sites: 1 X California and 3 X Florida

Most Successful Launch Site

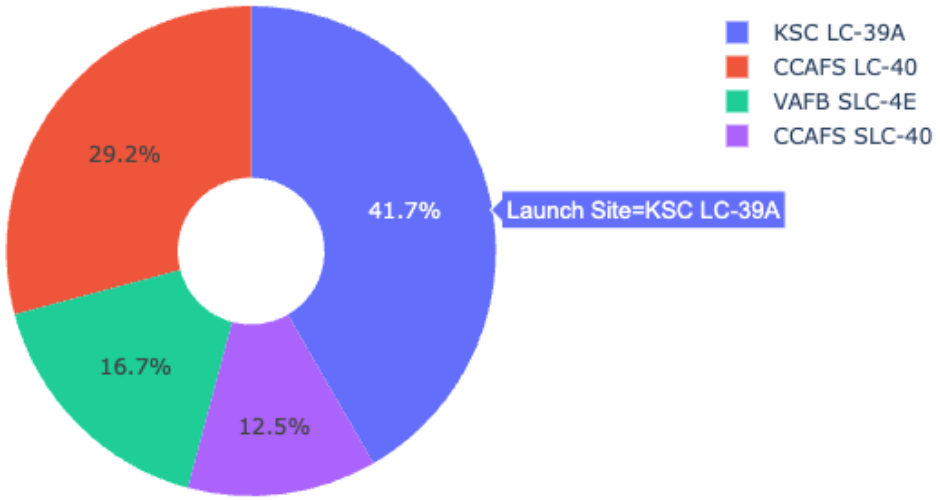
KSC LC -39A L has the most success



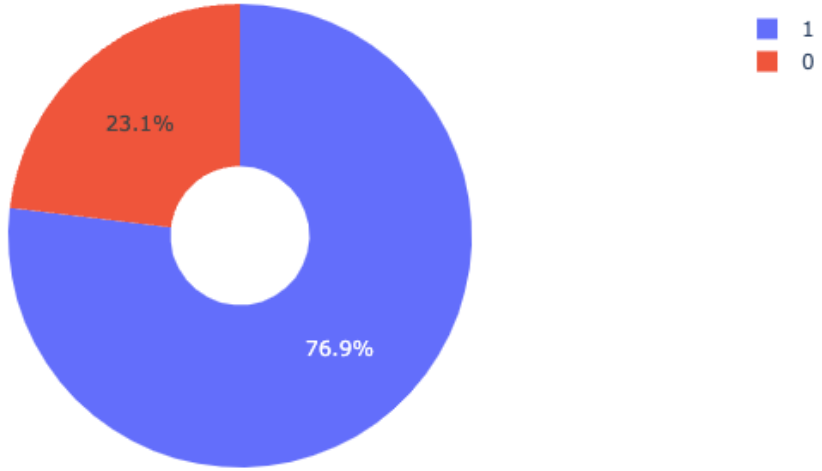
Total Launches

KSC LC 39A have the most launches and also most succes

Total Success Launches By all sites



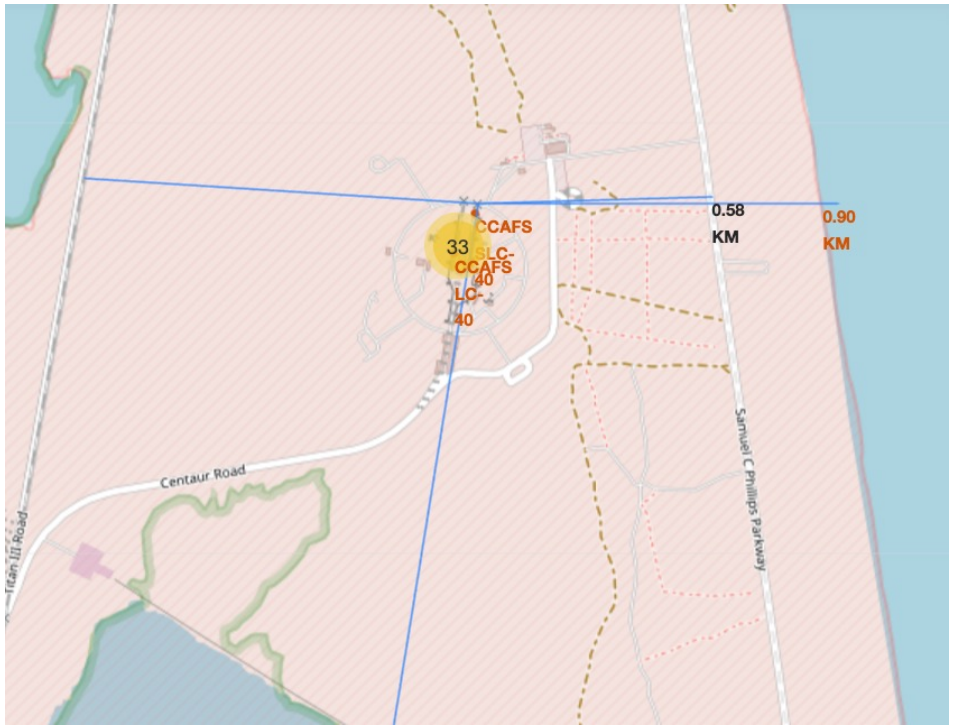
Total Success Launches for site KSC LC-39A

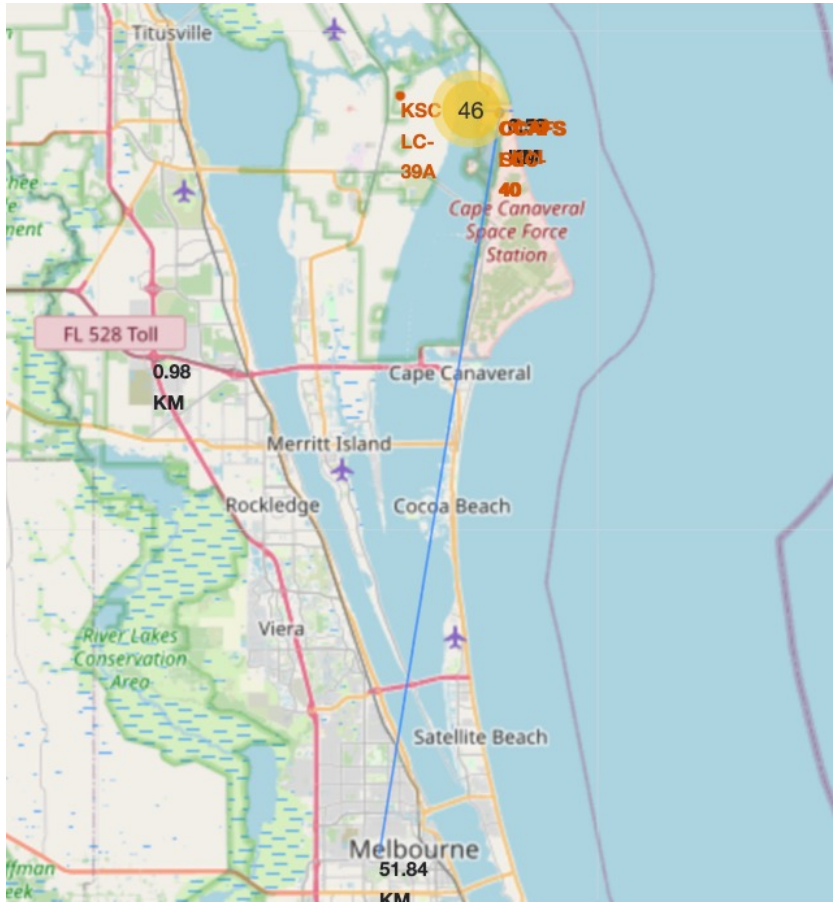


Launch Outcome VS PayLoad(kg)

Distances

- All launch are close to the coastline. This is basically to mitigate any harm to the population.\
- Thera are highways and railway tracks close to the site, but major population centers are quite further out.

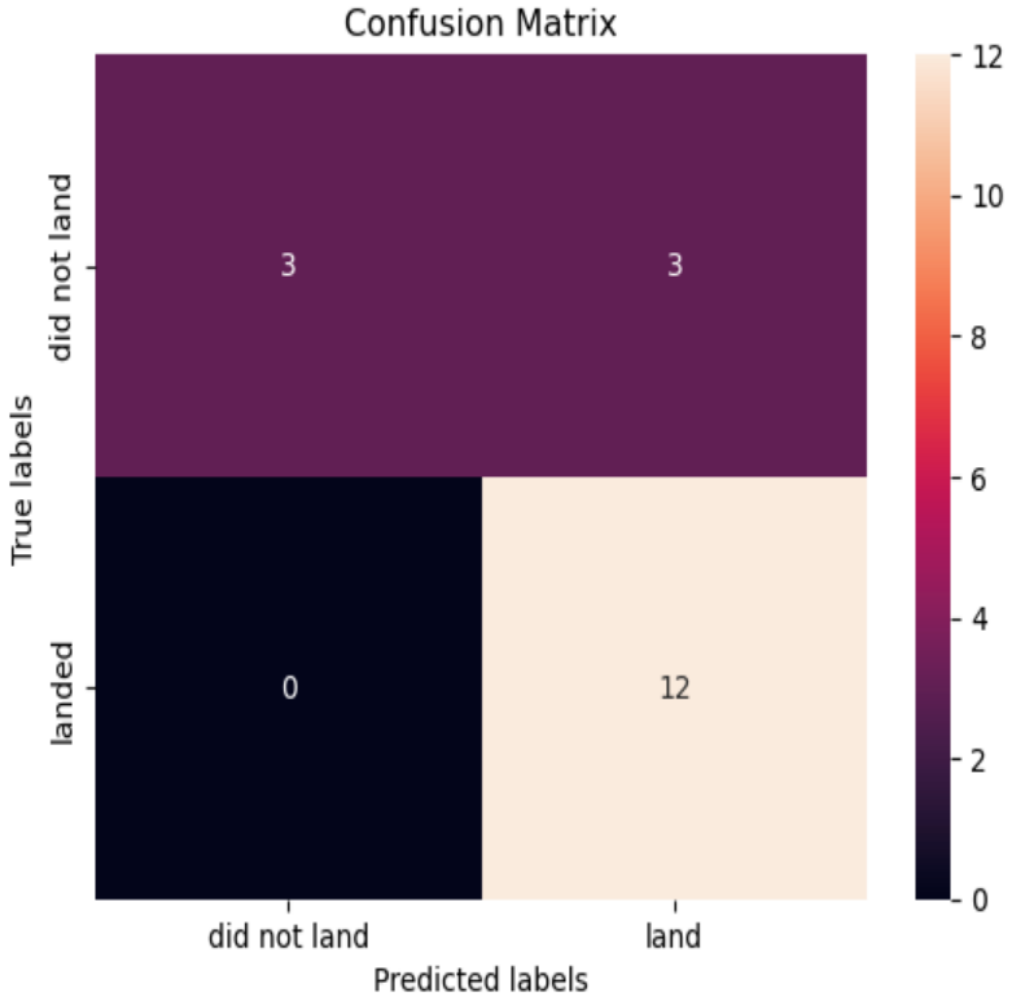




Predictive Analysis

- Trained the model with Logistic regression, SVM, Decision Tree and KNN algorithm and each model had the same accuracy of 83.33 % on test set with same confusion matrix. However, each model had different accuracy for their training set. Out of all 4 models, decision tree had the best training set accuracy of 89.9 % .
- Confusion matrix shows issue with False positive rate, where 3 of the unsuccessful landings were predicted as successful landing.
- **Metrics:**
  - Precision = TP/(TP + FP) = 12/ 15 = .80
  - Recall = TP/ (TP + FN) = 12 / 12 = 1
  - F1 Score = 2 * (Precision * Recall) / (precision + recall)

    = .89
  - Accuracy = (TP + TN)/(TP+TN+FP+FN) = .833



Conclusion

- Launch success increased over time.
- Higher payload mass, higher success rate across all launch sites.
- KSC LC-39A has the higher success rate (41%) among all launch sites and 76.9 success rate within itself
- Launch to ESL-L1, GEO, HEO, SSO orbits all have a 100% success rate on landings.
- All model performed equally on the test data set.

Consideration For Future

- Perform more/ better feature analysis and engineering to get better results.
- Try other models. Ada boosting may do better in this case.
- Adding regularization could help too.