PRESENTATION

SpaceX Project

This project analyzes the SpaceX flight database to predict rocket booster recovery, enabling cost estimation and helping competitors assess if they can offer more cost-effective launch services.

By, Harsh Tripathi

Executive summary

- Analyzed SpaceX flight data to identify factors influencing rocket booster recovery.
- Built predictive models to estimate booster recovery outcomes in advance.
- Enabled early estimation of launch costs based on recovery likelihood.
- Provided insights for competitors to benchmark against SpaceX and explore cost advantages.

contents

Executive summary

Data collection

EDA

Predictive analysis

conclusion

Introduction

- Collected SpaceX flight data using APIs and web scraping techniques.
- Performed **Exploratory Data Analysis (EDA)** to uncover patterns and insights.
- Developed both static and **interactive visualizations** to enhance understanding.
- Applied **predictive methodologies** to forecast rocket booster recovery.
- Conducted **SQL-based EDA** for structured querying and analysis.
- Integrated **interactive Folium maps** for geospatial visualization of launch sites.

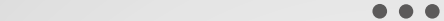
02

Data collection

SpaceX

Data Collection

Data collection through API and cleaning



	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Ser
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
89	86	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1
90	87	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	13	B1
91	88	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1
92	89	2020-10-24	Falcon 9	15600.0	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e3033383ecbb9e534e7cc	5.0	12	B1
93	90	2020-11-05	Falcon 9	3681.0	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb234e7ca	5.0	8	B1

90 rows × 17 columns

Data Collection

Data collection through Web Scraping and cleaning

```
[ ] df= pd.DataFrame({ key:pd.Series(value) for key, value in launch_dict.items() })
    display(df.head())
```

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.07B0003.18	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.07B0004.18	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.07B0005.18	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.07B0006.18	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.07B0007.18	No attempt\n	1 March 2013	15:10

```
[ ] df.to_csv('spacex_web_scraped.csv', index=False)
```

03

Data Wrangling

SpaceX

Data Wrangling

Number of launch sites

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

Number and occurrence of each orbit

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

Data Wrangling

Mission outcomes in different Orbits

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

cumulative mission success rate

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

```
np.float64(0.6666666666666666)
```

04

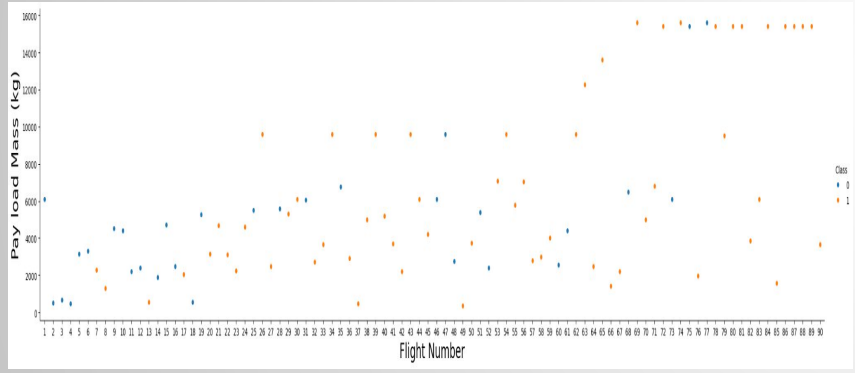
EDA visuals

SpaceX

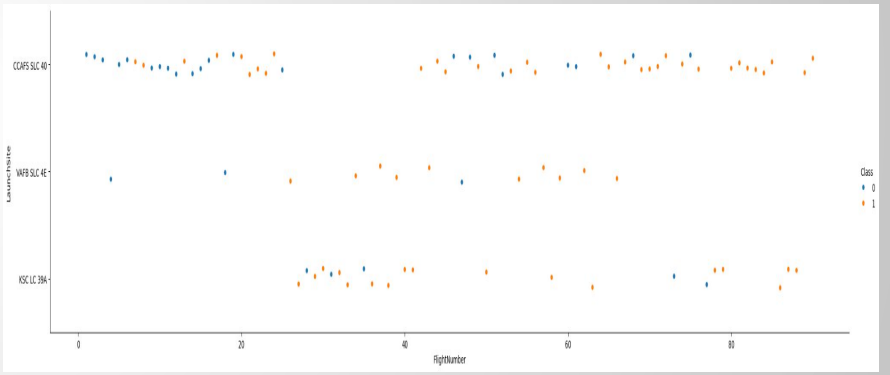
EDA

EDA visulisation

Launch outcome affected by flight number and Payload



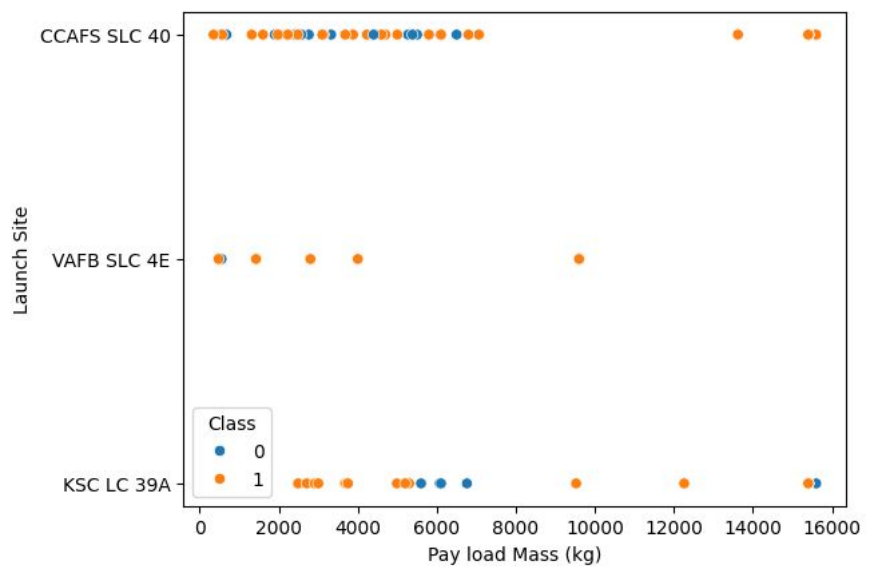
Launch outcome affected by flight number and Launch Site



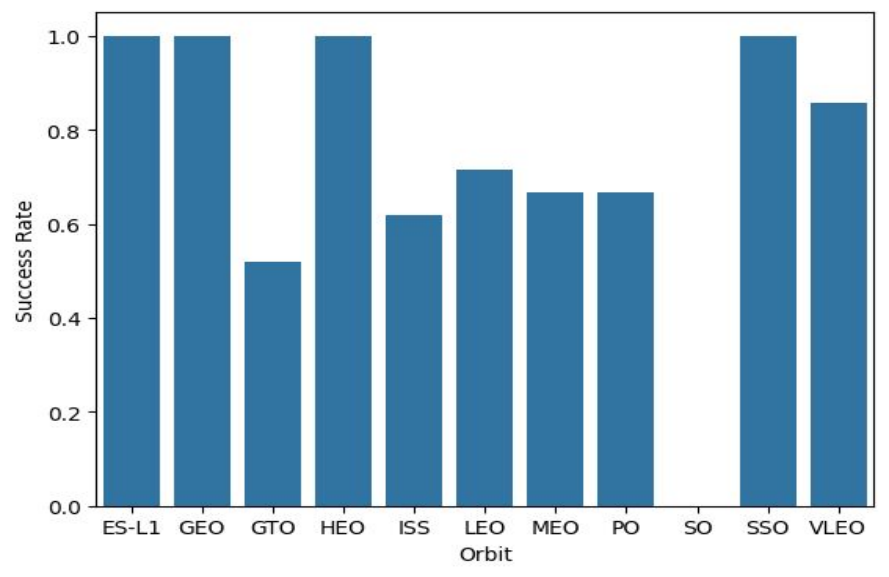
EDA

EDA visulisation

Launch outcome affected by Launch Site and Payload



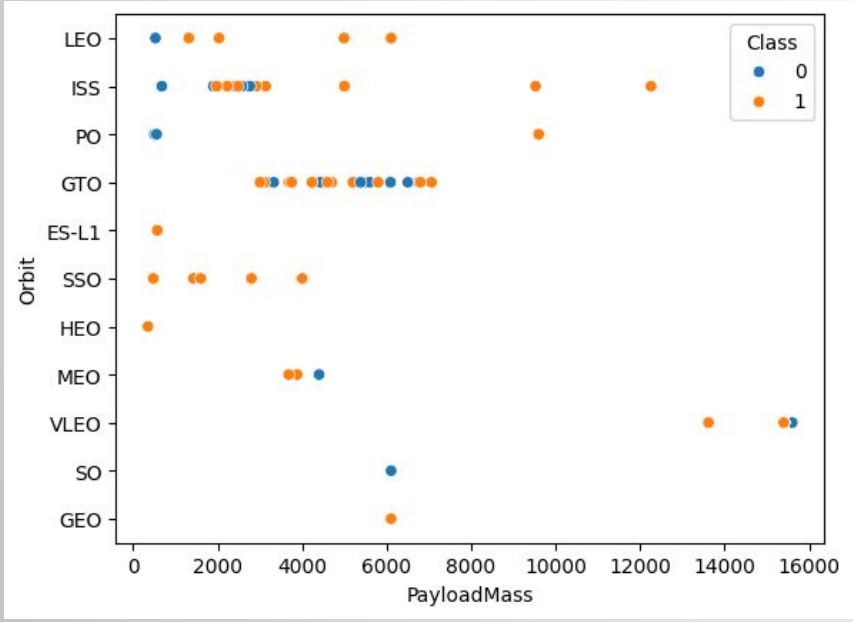
Launch outcome affected by orbit type



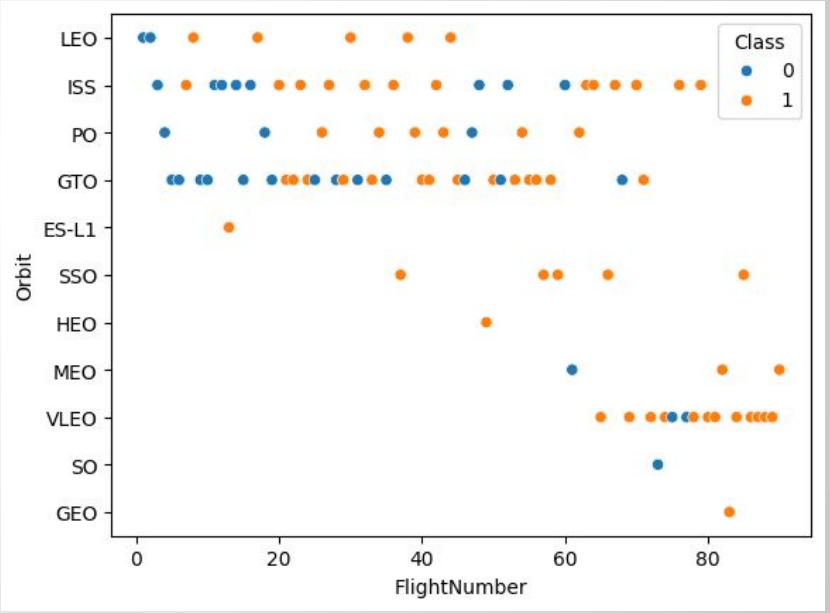
EDA

EDA visulisation

Launch outcome affected by Payload mass and Orbit type



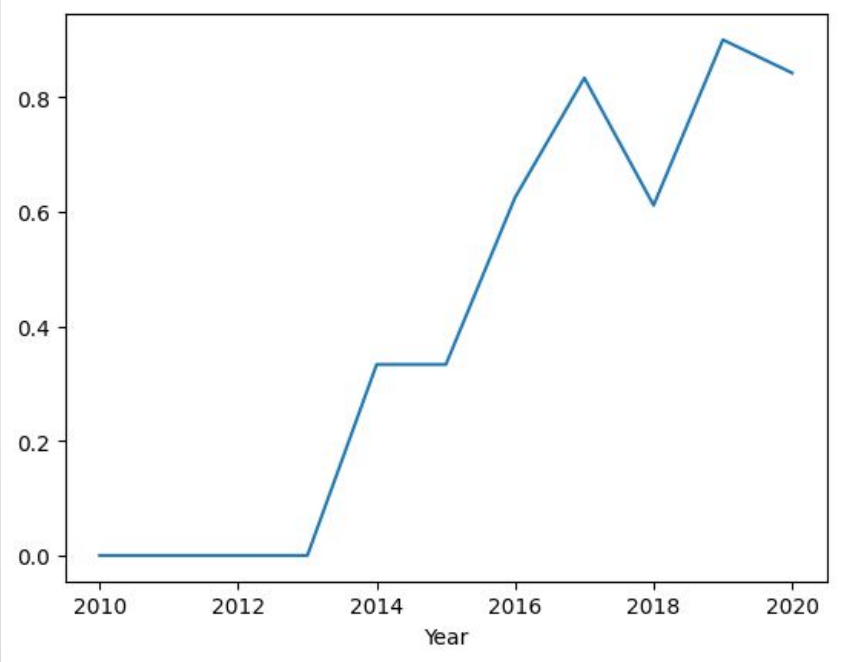
Launch outcome affected by flight number and orbit type



EDA

EDA visualization

Launch success trend over the years



05

EDA SQL

SpaceX

EDA
EDA SQL

Unique launch sites in the space mission

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

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

Launch Sites starting with CCA

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

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

EDA

EDA SQL

Total mass carried by boosters launched by NASA

```
 * sqlite:///my_data1.db
Done.
sum(PAYLOAD_MASS__KG_)
45596
```

Average mass carried by F9 V1.1

```
 * sqlite:///my_data1.db
Done.
avg(PAYLOAD_MASS__KG_)
2534.6666666666665
```

EDA

EDA SQL

First successfully landing on ground pad

```
 * sqlite:///my_data1.db
Done.
 min(DATE)
 2015-12-22
```

Boosters with success on drone ship and payload between 4000 and 6000

```
 * sqlite:///my_data1.db
Done.
 Booster_Version
 F9 FT B1022
 F9 FT B1026
 F9 FT B1021.2
 F9 FT B1031.2
```

EDA
EDA SQL

Total failure vs success

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

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

Booster that have carried maximum payload mass

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

EDA
EDA SQL

Records of landing outcome failures in 2015

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

Ranking of landing outcomes between 04-06-2014 and 20-03-2017

Landing_Outcome	Count
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

EDA
SQL EDA

Records of landing outcome failures in 2015

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

Ranking of landing outcomes between 04-06-2014 and 20-03-2017

Landing_Outcome	Count
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

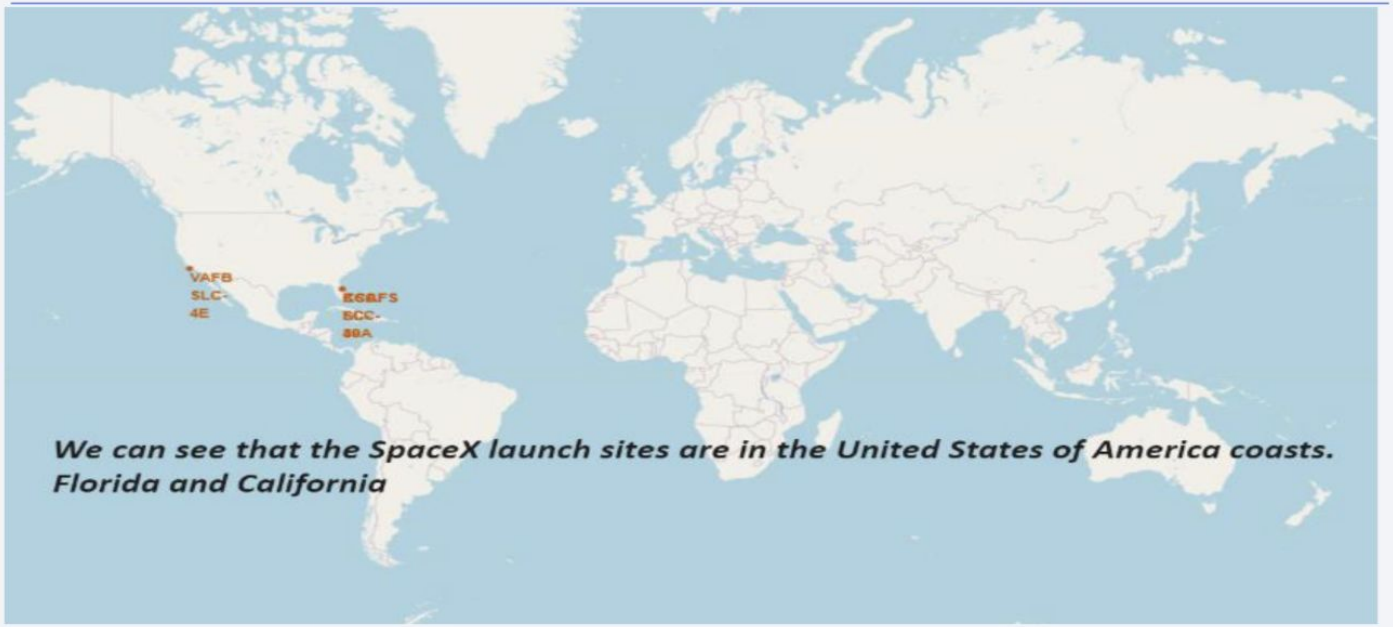
06

Folium

SpaceX

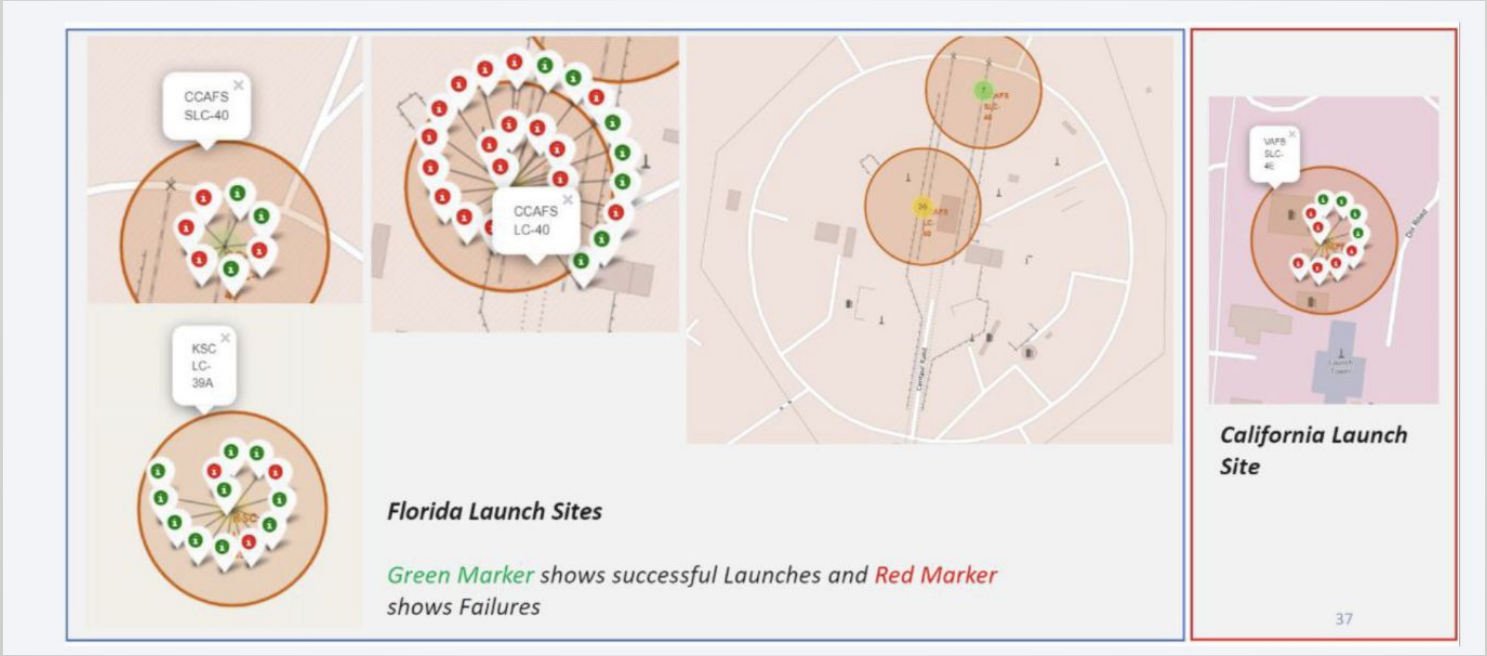
Launch Sites Proximities Analysis

All launch sites global map markers



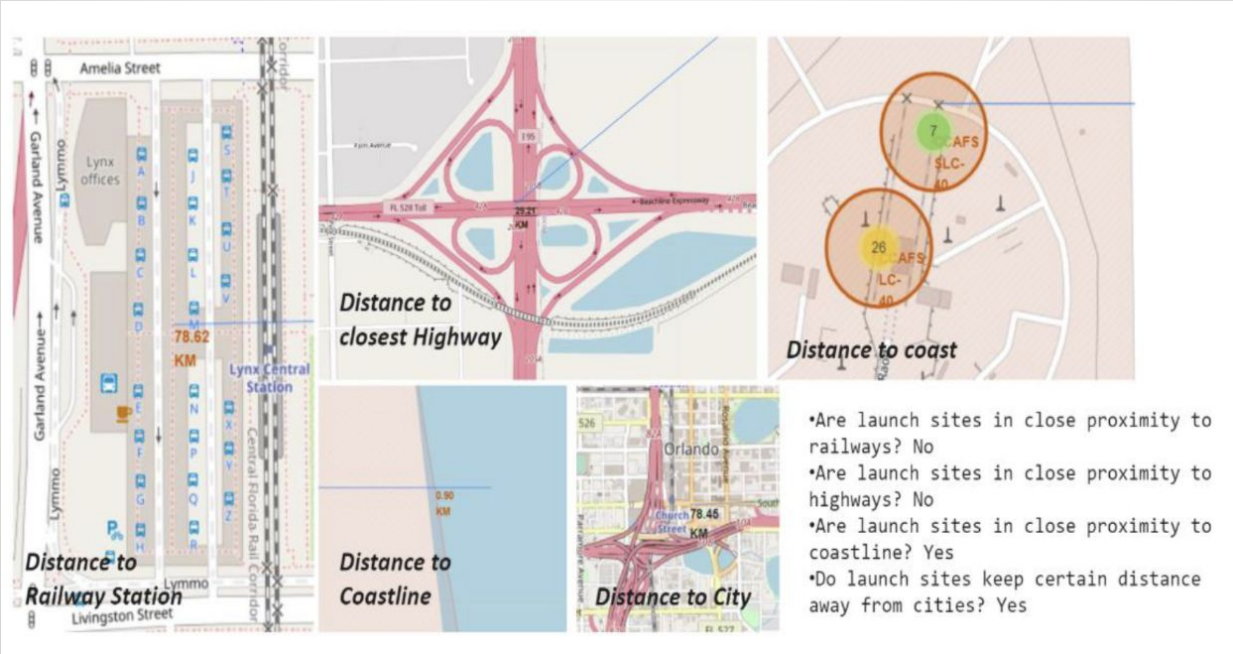
Launch Sites Proximities Analysis

Markers showing launch sites with color labels



Launch Sites Proximities Analysis

All launch sites global map markers



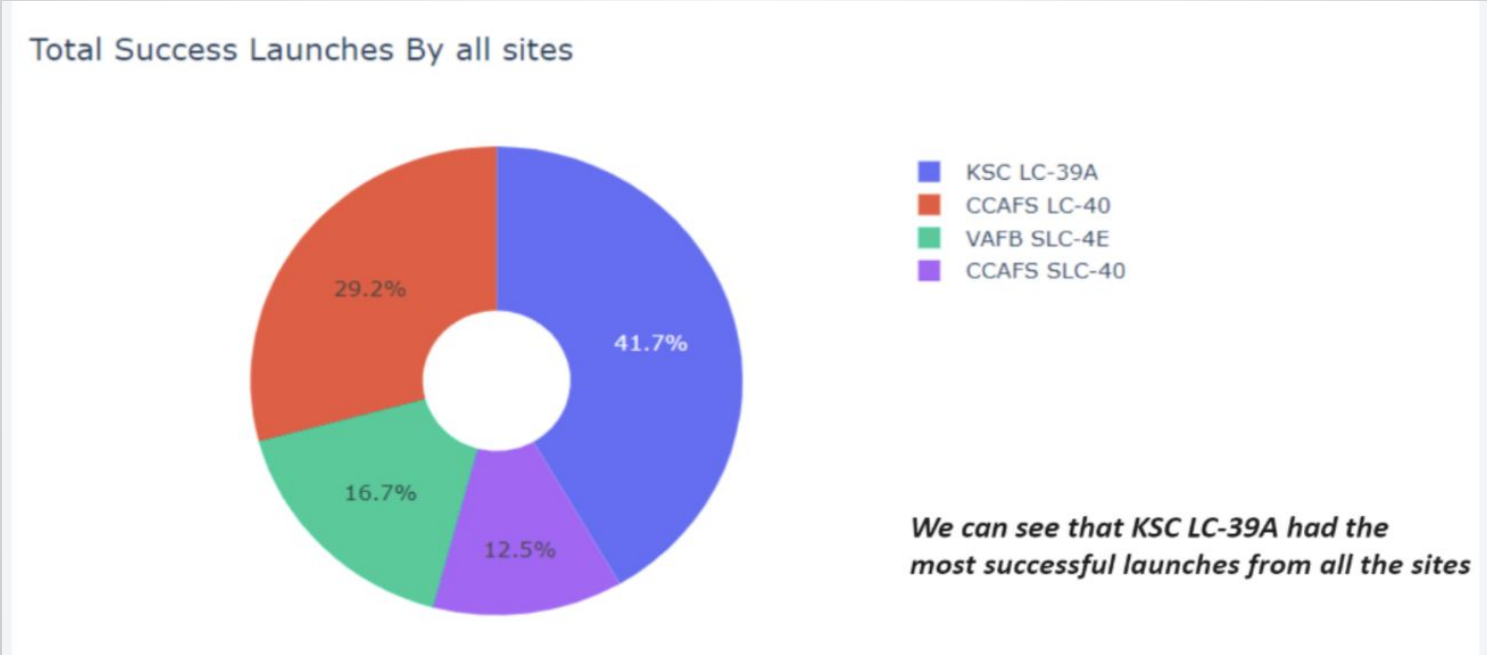
07

Plotly Dash

SpaceX

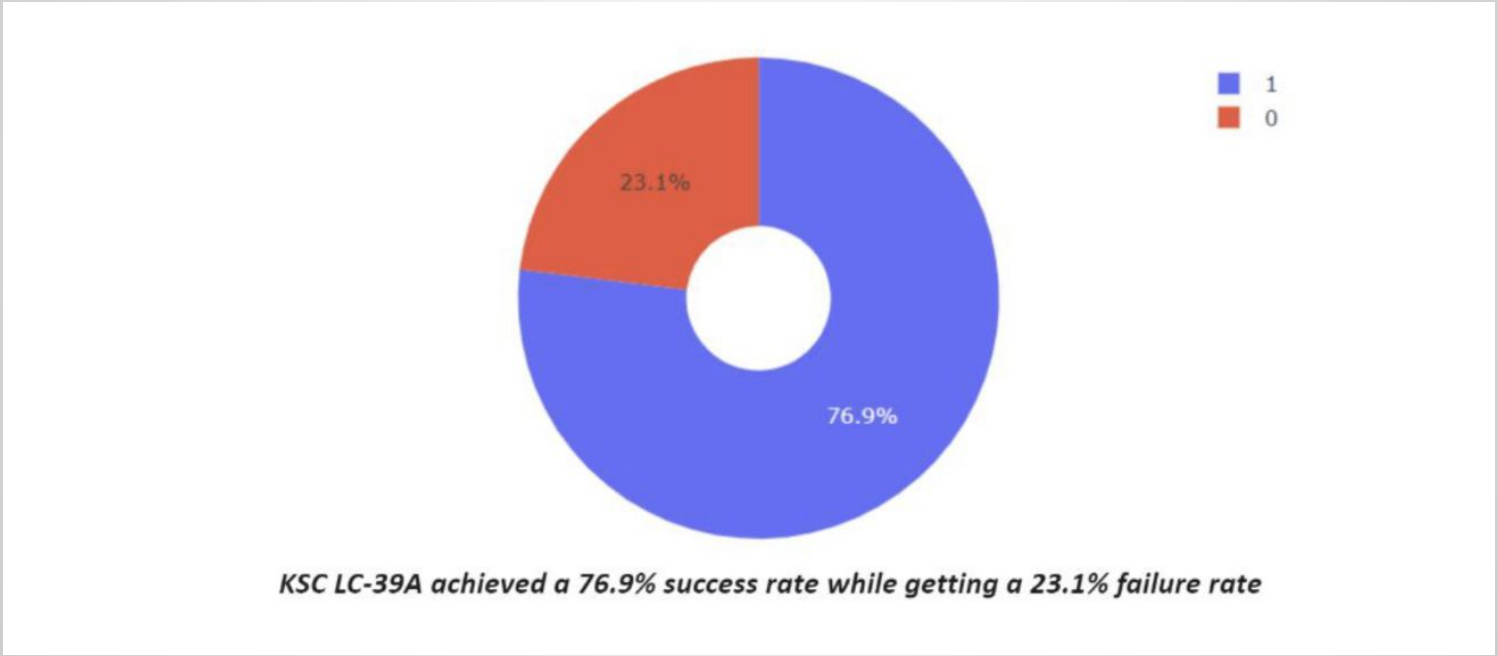
Dashboard with Plotly Dash

Pie chart showing success percentage achieved by each Launch site



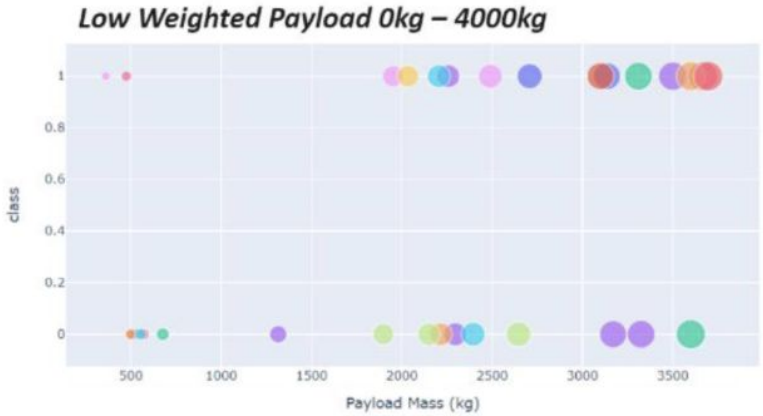
Dashboard with Plotly Dash

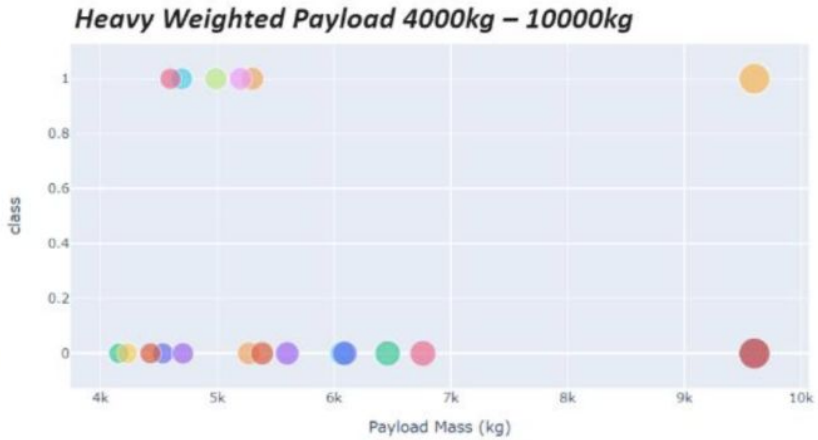
Pie chart showing Launch Site with highest Launch success ratio



Dashboard with Plotly Dash

Scatter Plot og Payload vs Launch Outcome for all sites, with different payload selected in the range slider





We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

08

Predictive analysis

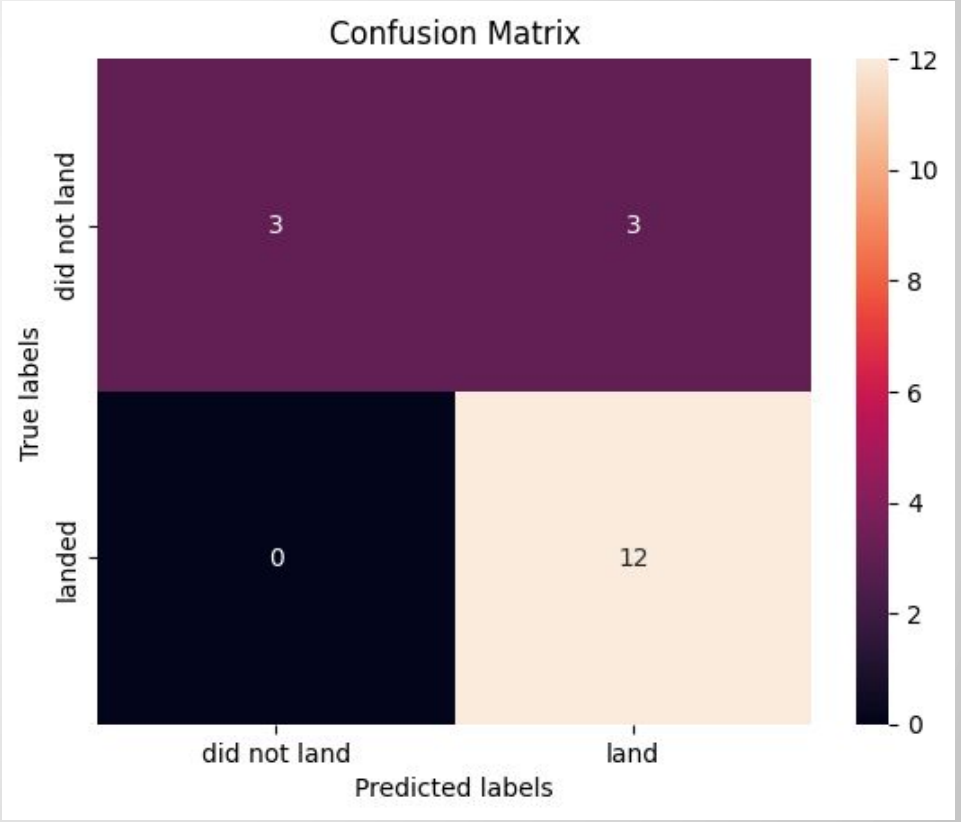
SpaceX

Predictive analysis Results
Logistic Regression

Accuracy score train data : 84.4%

Accuracy score test data : 83.3%

Confusion matrix :

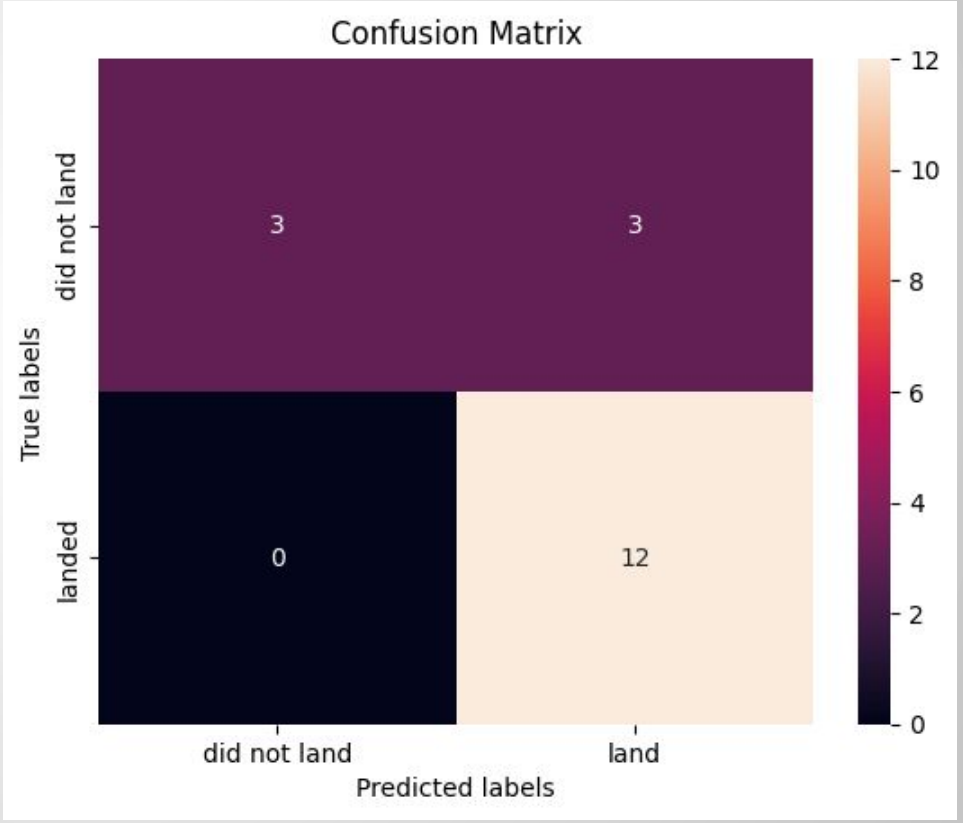


Predictive analysis Results
Vector machine object

Accuracy score train data : 84.8%

Accuracy score test data : 83.3%

Confusion matrix :

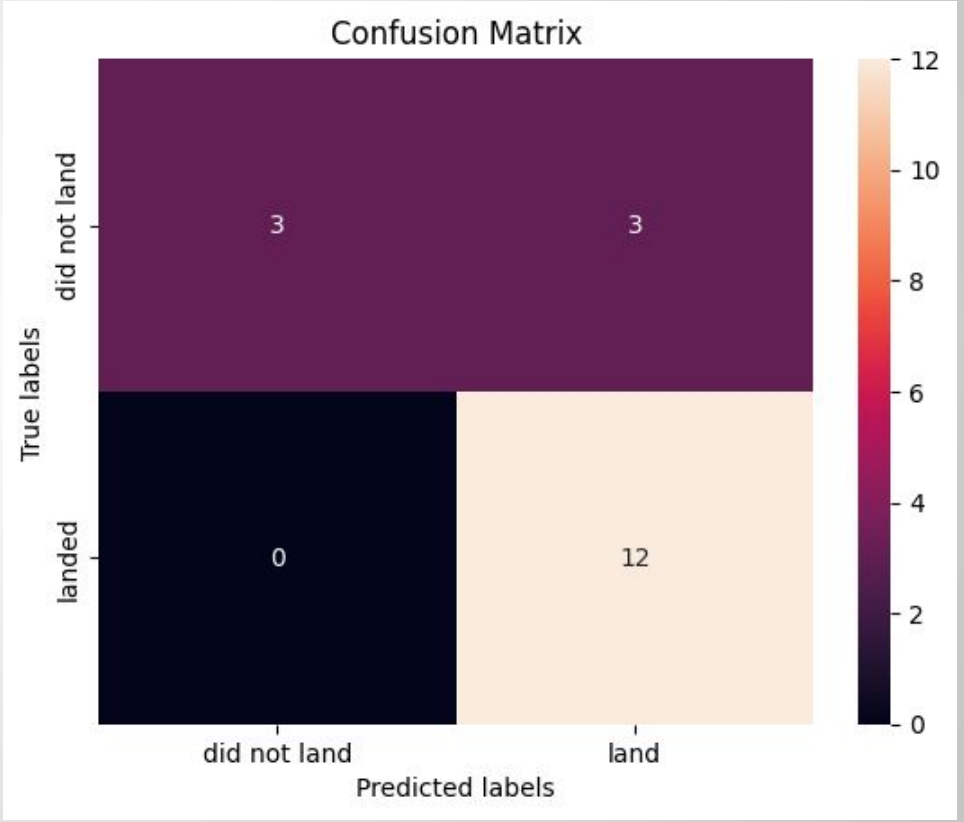


Predictive analysis Results
Decision Tree Classifier

Accuracy score train data : 87.7%

Accuracy score test data : 83.3%

Confusion matrix :

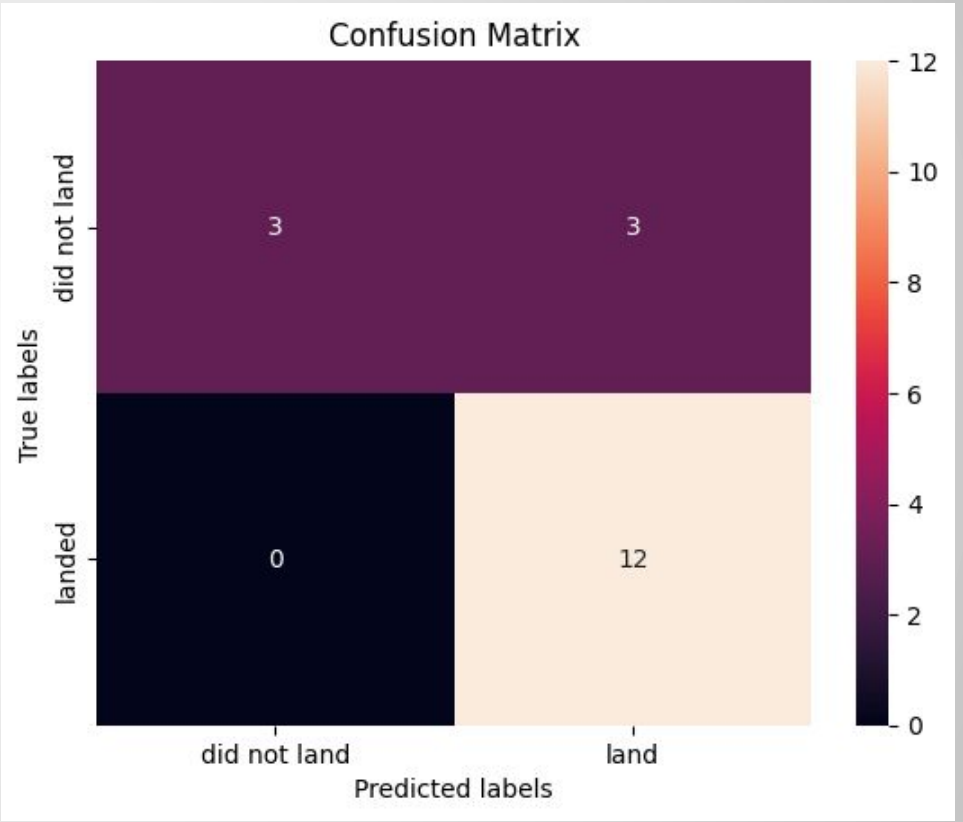


Predictive analysis Results
K nearest object classifier

Accuracy score train data : 84.8%

Accuracy score test data : 83.3%

Confusion matrix :



08

Conclusion

SpaceX

Conclusion

We can conclude that:

- The launch site that has more flights, have higher success rates.
- Launch site success rates has been increasing overtime
- Orbits ES-l1, GEO, HEO, SSO, VLEO had the highest success rate
- KSC LC-39A had the most successful launches of any sites
- The decision tree classifier is the predictive algorithm for this particular task

Best regards,

Thank you

Special thanks to all the people who made and released these awesome resources for free

SpaceX

Wikipedia