IBM Developer
SKILLS NETWORK

CAPSTONE PROJECT: SpaceX launch

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusions
- Appendix

Executive Summary

- In the cutthroat world of space exploration, SpaceX has pioneered transformative approaches to satellite launches

- Falcon 9 and Falcon 9 Heavy have rewritten the playbook, offering a groundbreaking way to cut costs significantly, reducing the price per kilogram for payload launches.

- Despite these advances, challenges in terms of reliability persist, setting SpaceX's innovations apart from conventional launch vehicles like Soyuz or Ariane-5.

- For SpaceX, success is defined by one critical criterion: the triumphant recovery or landing of the booster.

- The outcome of Falcon 9 booster recovery hinges on multiple factors, including the unique **orbital path**, **payload weight**, **booster versions**, and the choice of **launch sites.**

- Remarkably, the model has achieved an accuracy rate approaching 94%, making it a robust tool for predicting Falcon 9 booster recovery outcomes.

Introduction

- Since 2017, SpaceX Falcon9/Falcon9 Heavy is increasingly dominating the market with better cost per kg thanks to the reusable booster concept.

- Competition may heat up again with the advent of Ariane 6 and SpaceX Starship.

- **Falcon 9/Falcon9 heavy v. Ariane 6**

  Falcon9 / Falcon 9 Heavy can easily compete with new Ariane 6 (2 or 4 boosters) in terms of:

    - Cost per kg
    - Payload mass

- Using SpaceX Falcon9 database, we develop AI tools:
  - for visualizing, analyzing features conditioning mission success
  - for predicting Falcon 9 success and SpaceX financial viability using supervised classification algorithms.

Methodology

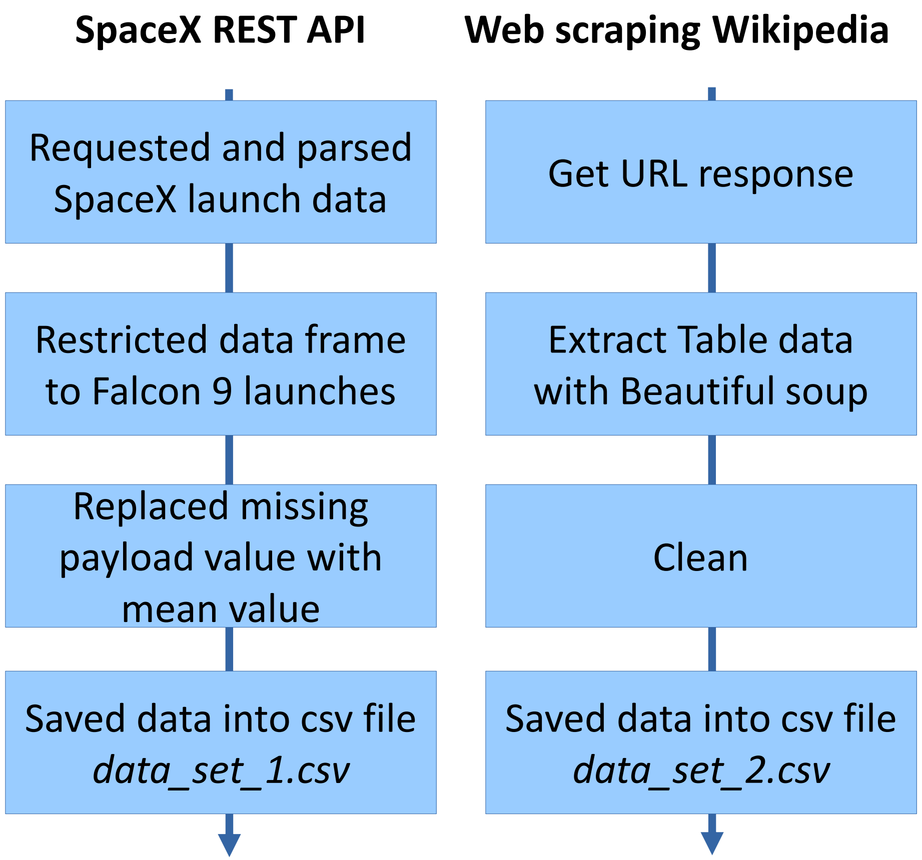
Steps:

- Data Collection from IBM skills network database and Wikipedia
- Data Wrangling
- EDA using SQL and visualization of correlation between parameters
- Visual analytics: Launch sites with Folium, success rate with Plotly dashboard.
- Classification model development and validations. Selection of best predictive model.

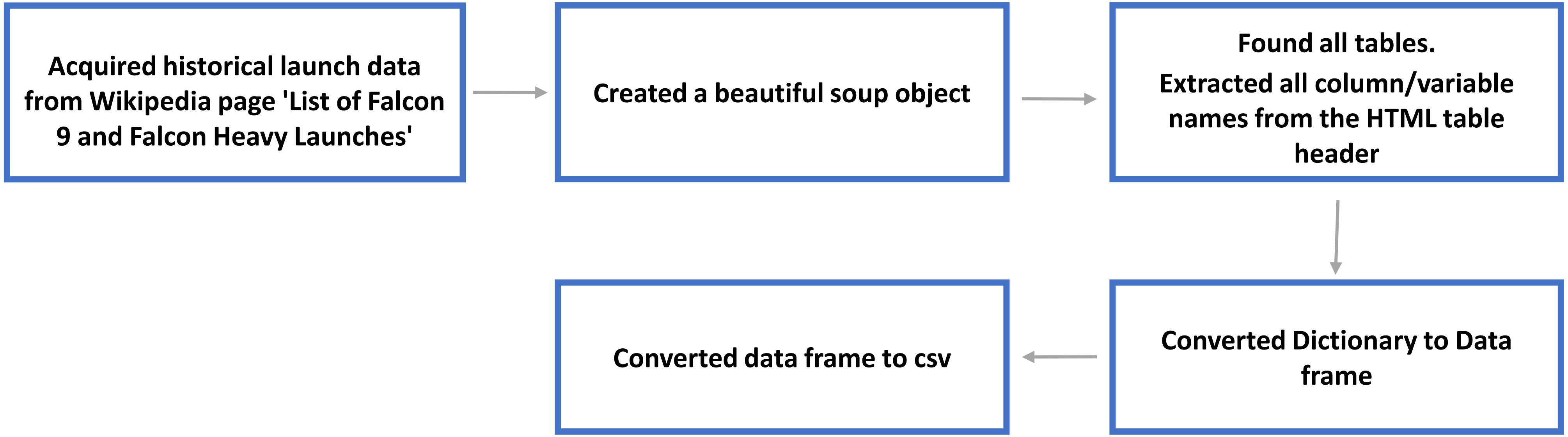
Data Collection

Data was collected from:

- Open source SpaceX REST API
- Webscraping Falcon9 launch data in Wikipedia



Data Collection - Scraping



Data Collection – SpaceX API

A series of functions helping the use the API to extract info using identification numbers (IDs) in the launch data.

- **getBoosterVersion(data)**
- **getLaunchSite(data)**
- **getPayloadData(data)**
- **getCoreData(data)**

Requested rocket launch data from SpaceX API
url=https://api.spacexdata.com/v4/launches/past

Decoded response as a .json file
Turned it into Pandas dataframe

Used the API with functions to get info from IDs:
- **BoosterVersion**
- **Payloads**
- **LaunchSite**
- **Cores**

Stored in List

Assigned list to dictionary
Converted to Dataframe

Selected Falcon9 data only
Replaced missing payload mass value with mean

Converted the data frame to csv file

Data Wrangling

Identified missing values

Replaced missing "PayloadMass" value with mean value

- null values for each feature
- numerical and categorical features

Further Calculations

- Number of launches on each site
- number and occurrence of each orbit
- number and occurrence of missions outcome per orbit type

Created a set of 1   stage booster landing outcomes

0 True ASDS: successful landing on a drone ship

1 None None: failure to land

2 True RTLS: successful landing to a ground pad

3 False ASDS: failed landing on a drone ship

4 True Ocean: successful landing, specific region of the ocean

5 False Ocean: failed landing, specific region of the ocean

6 None ASDS: failure to land

7 False RTLS: failed landing to a ground pad

Created a Training label: 'Class'

Class = 0: booster landing failure

Class = 1: booster landing success

PayloadMass	Orbit	LaunchSite	Outcome	Class
3750.00	GTO	KSC LC 39A	True ASDS	1
2760.00	ISS	CCAFS SLC 40	None None	0
5383.85	GTO	CCAFS SLC 40	None None	0
4230.00	GTO	CCAFS SLC 40	True Ocean	1
2395.00	ISS	CCAFS SLC 40	False ASDS	0

EDA with SQL

Following information was displayed using SQL.

- Launch sites
- Payload mass
- Booster versions
- Mission outcomes
- Booster landings

```
%%sql SELECT "Landing Outcome", COUNT(*) AS Count_Landing_Outcomes
FROM SPACEXTABLE
WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY "Landing Outcome"
ORDER BY Count_Landing_Outcomes DESC;
```

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

"Landing Outcome"	Count_Landing_Outcomes
Landing Outcome	32

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

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

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

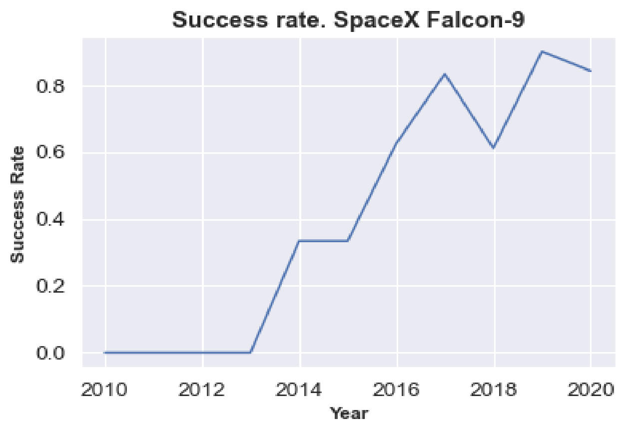
EDA with data Visualization

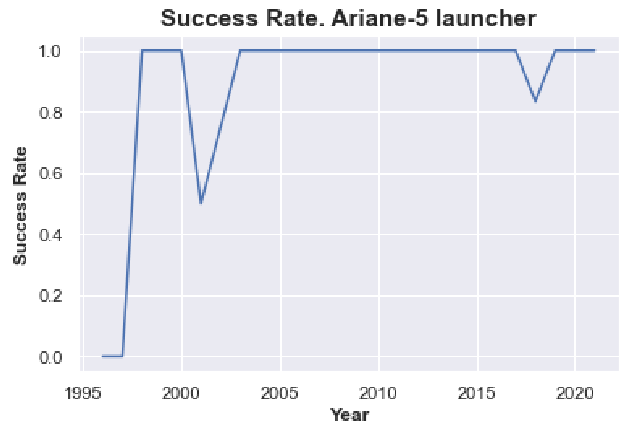
Exploring data in the Falcon 9 dataframe searching for factors and relations influencing launching success rate (booster recovery).

- Payload mass
- Orbit type
- Launch site

Graphs and scatter charts with Matplotlib – Seaborn and Analysis. Results with Scatter charts are labeled: class 0-1 (failure/success).

- Payload mass v. Flight Number
- Launch Site v. Flight number
- Launch Site v. Payload mass
- Orbit v. Flight number
- Orbit v. Payload mass
- Histogram: success rate for each orbit

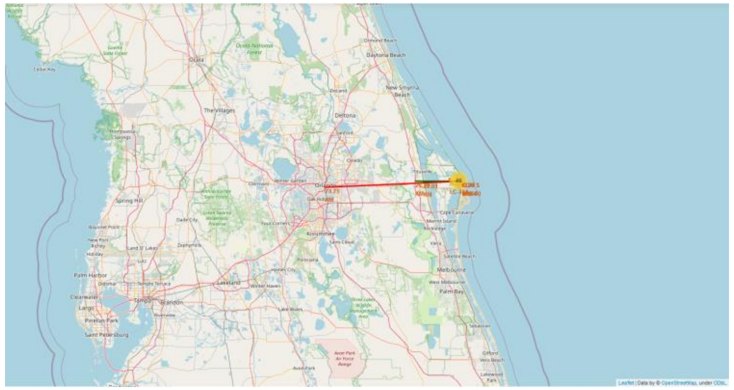




Interactive map with Folium

Folium Interactive Map was used for visualizing and analyzing SpaceX Launch Sites.

- Used Interactive mapping library called Folium
- Identified all SpaceX launch sites on a map: Florida, California
- Included longitude and latitude info.
- Identified successful/failed launches for each site on map



CCAFS_SLC40 in Cape Canaveral FL
Coordinates: -80.577º, 28.563º

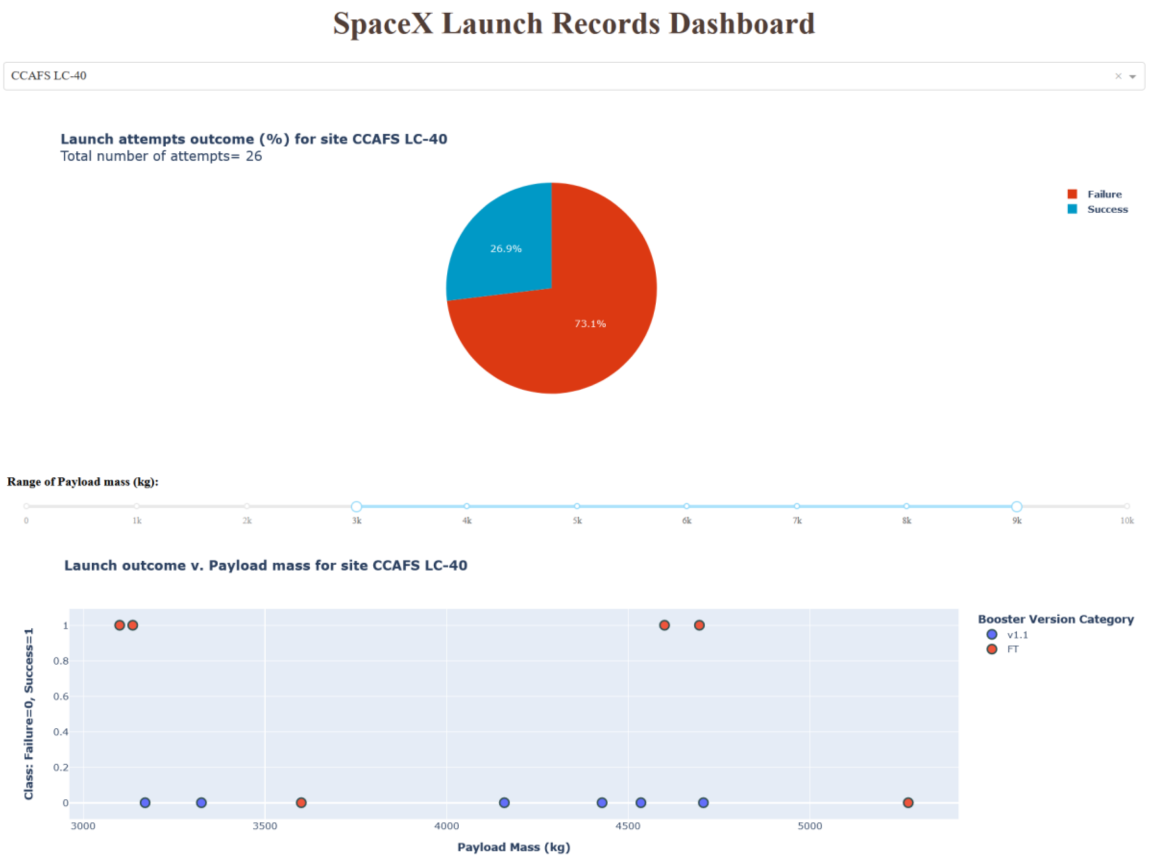
Dashboard with Plotly Dash

an interactive dashboard with Plotly was built which includes:

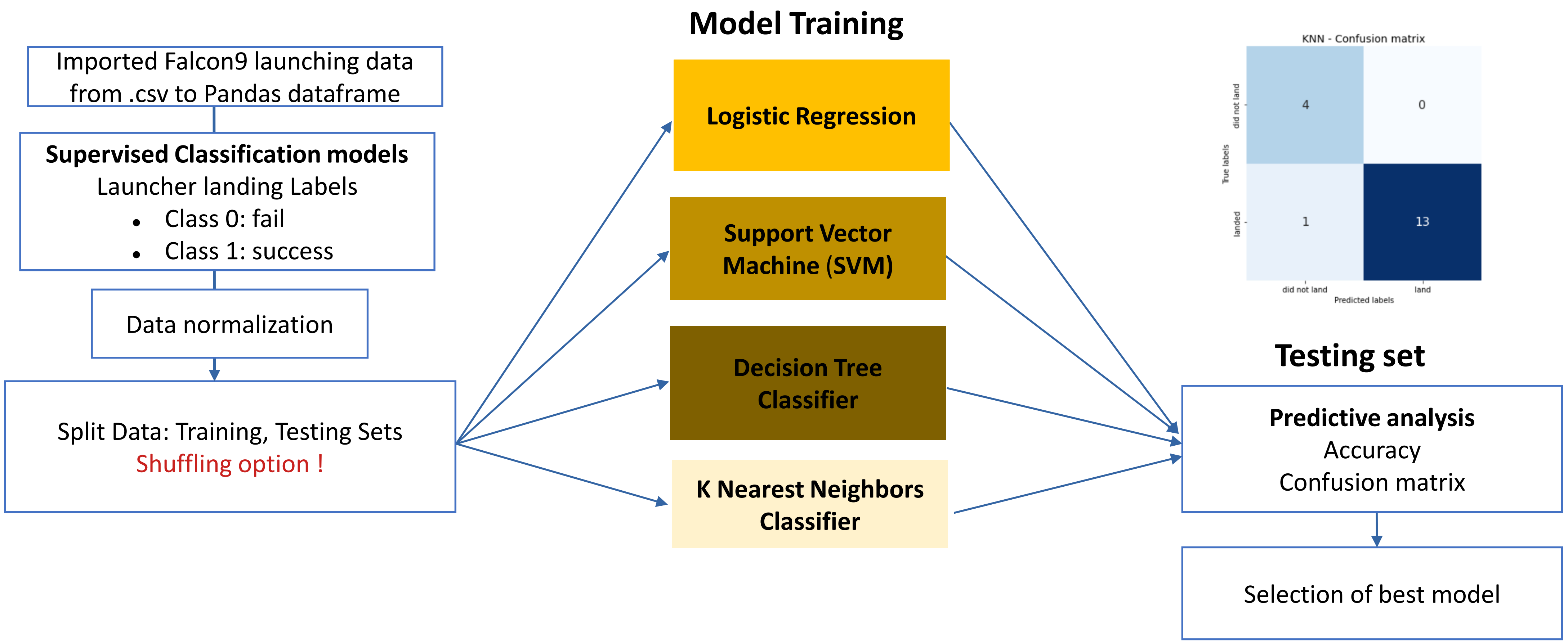
- Dropdown menu for selecting launch sites
- Pie charts displaying success rate.
- Scatter chart displaying launch site, payload mass, success/failure
- Range slider for selecting range of payload mass (kg).

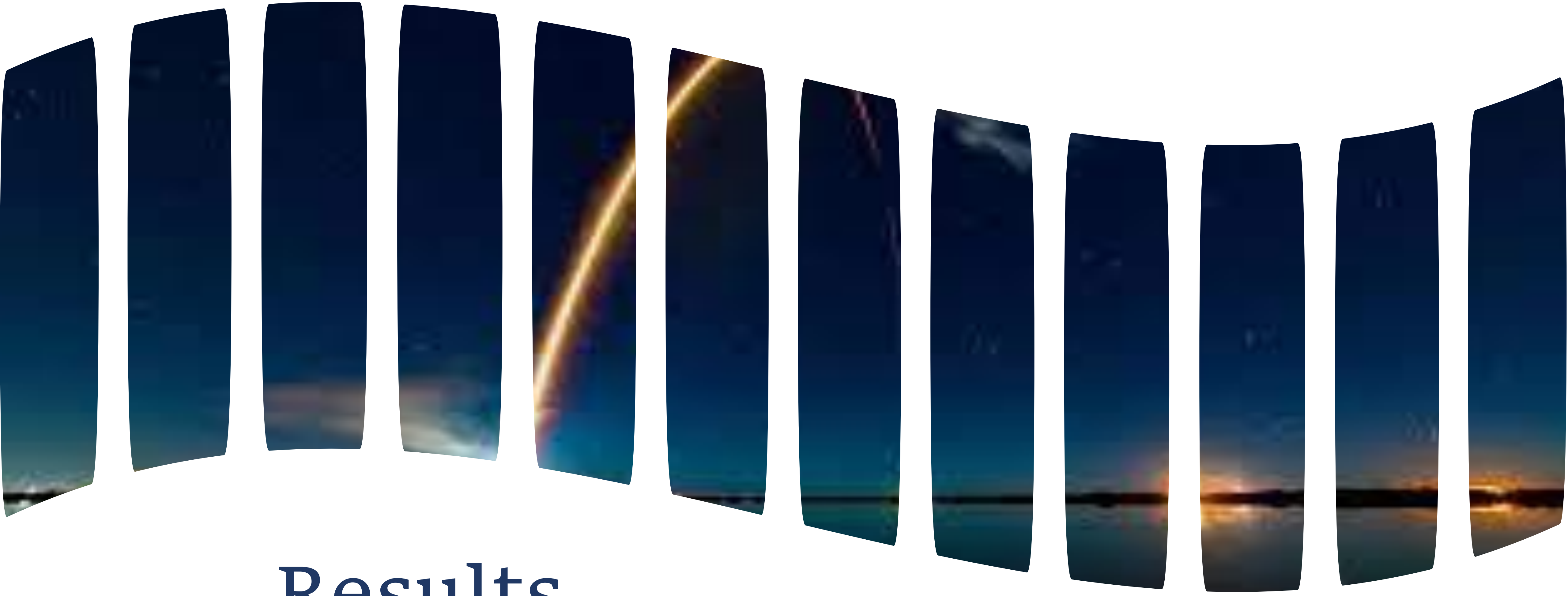
for analyzing SpaceX launch records features:

- site with largest successful launches.
- site with highest launch success rate
- payload range(s) with highest launch success rate
- payload range(s) with lowest launch success rate
- F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) with highest launch success rate.



Classification – Predictive Analysis





Results

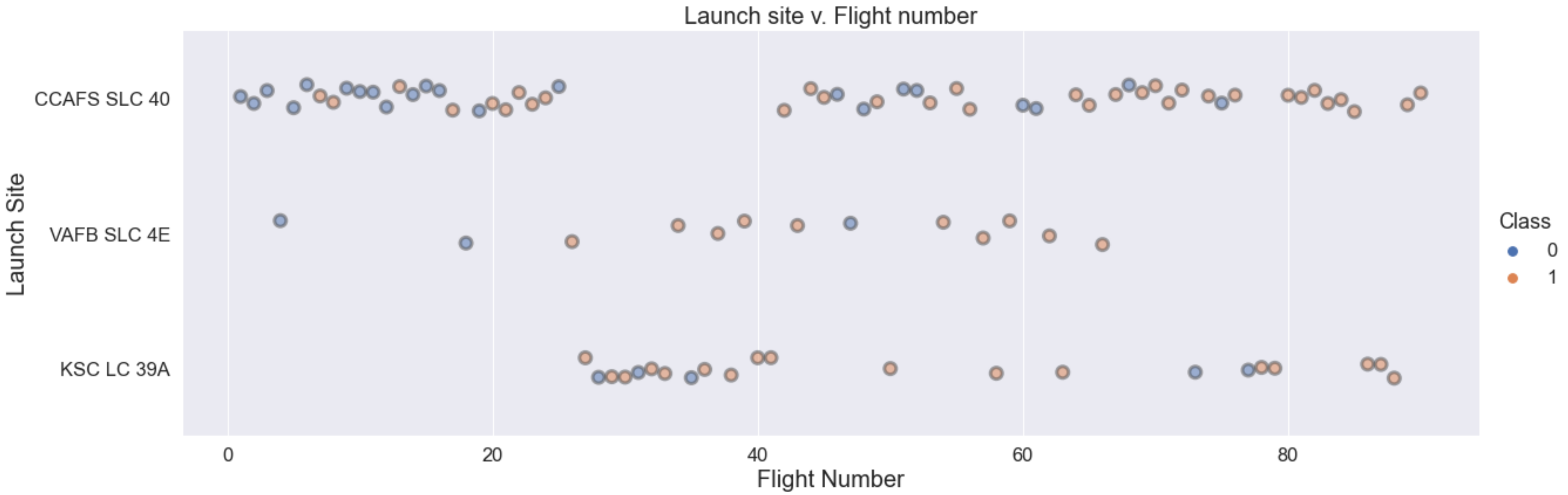
- EDA results
- Interactive analytics
- Predictive analysis results

Launch Site vs. Flight Number

The chart displays valuable info about:

- Chronology: flight numbers
- Number of flights per launch site
- Success/Failure per launch site

- Cape Canaveral CCAFS-SLC 40 is the most used launch site.
- CCAFS-SLC 40 concentrates most of failures , particularly **in the early stage of Falcon9 project**.
- Given CCAFS-SLC 40 southern location, most “risky” GTO and GEO launches may take place there.
- Additional info needed: orbit, payload mass

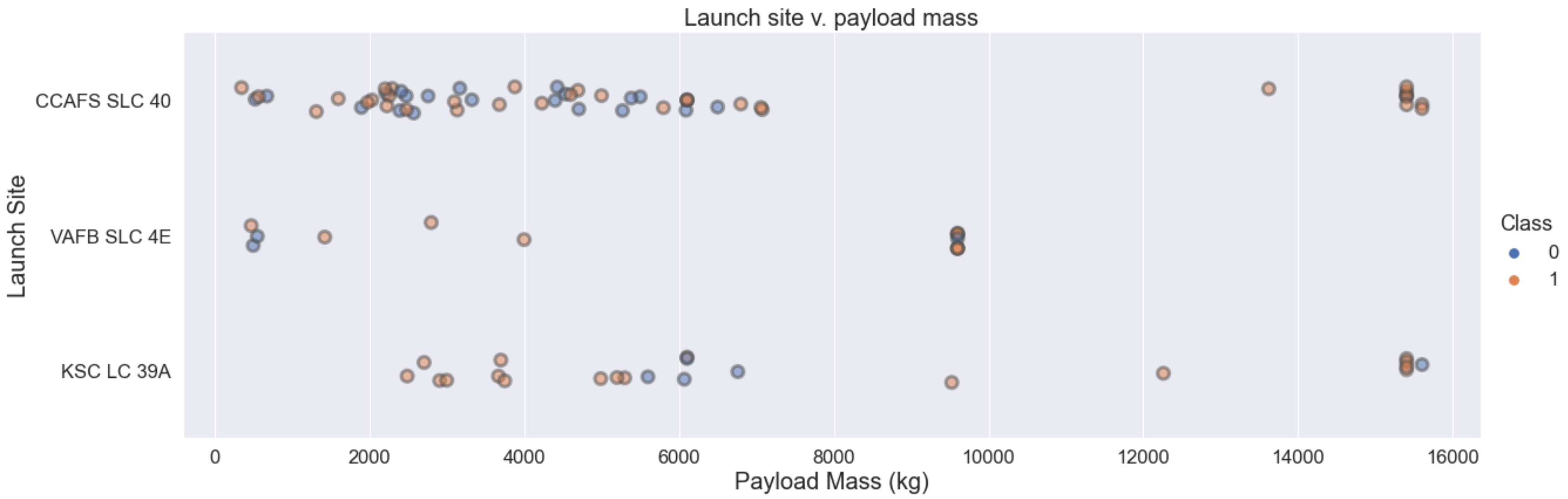


Launch Site vs. Payload

The chart brings additional info:

- Payload mass per launch site
- Success/Failure per payload mass

- Given Falcon9 specifications, heavy payloads > 10000 kg are sent to low/medium orbits LEO/MEO only.
- It looks like the percentage of failures is lower for heavy payload. Which would indicate that low orbits are less risky to the success of the mission (recovery of booster).
- Light payloads are not necessarily all sent to GTO/GEO.
- More information is needed for extracting some correlation: success rate v. payload/orbit



Orbit Type vs Flight Number

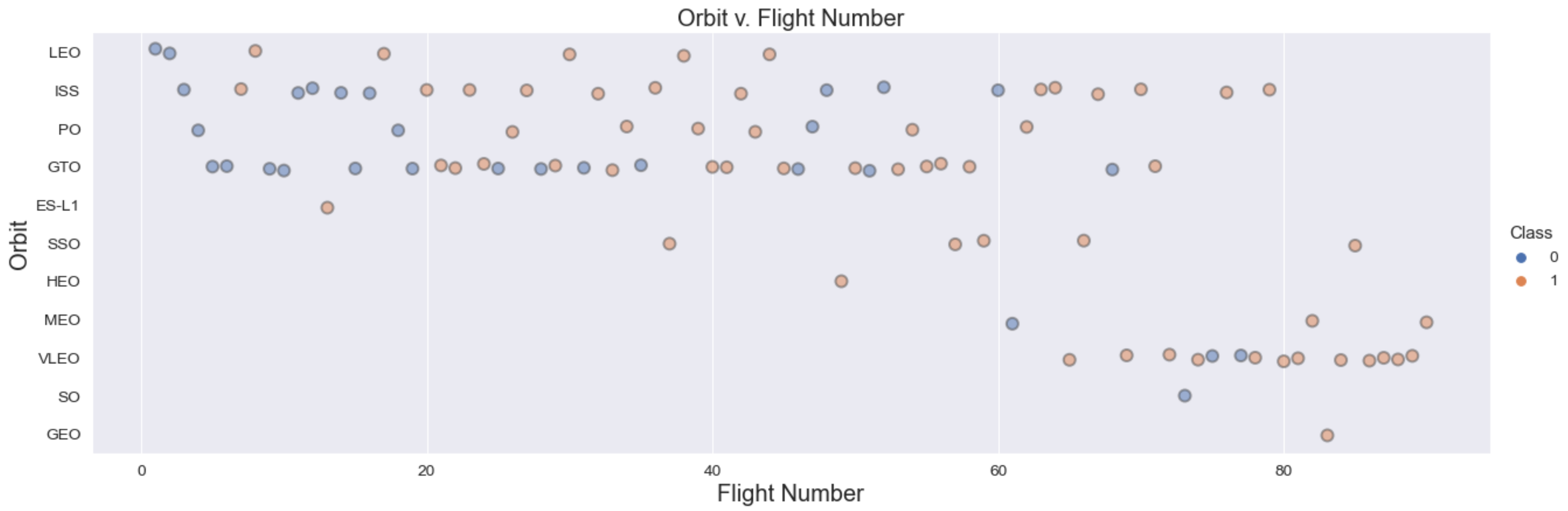
The chart brings additional info:

- Number of flights per Orbit.
- Success rate per orbit

- The number of flights for: GEO, SO, HEO, ESL-1, MEO is not significant for concluding about success rate.
- PO, SSO, ISS, VLEO are low orbits
- GTO is a transfer orbit to GEO.

It looks like GTO are higher risk missions, low orbits are lower risk.

We confirm with the following histogram.

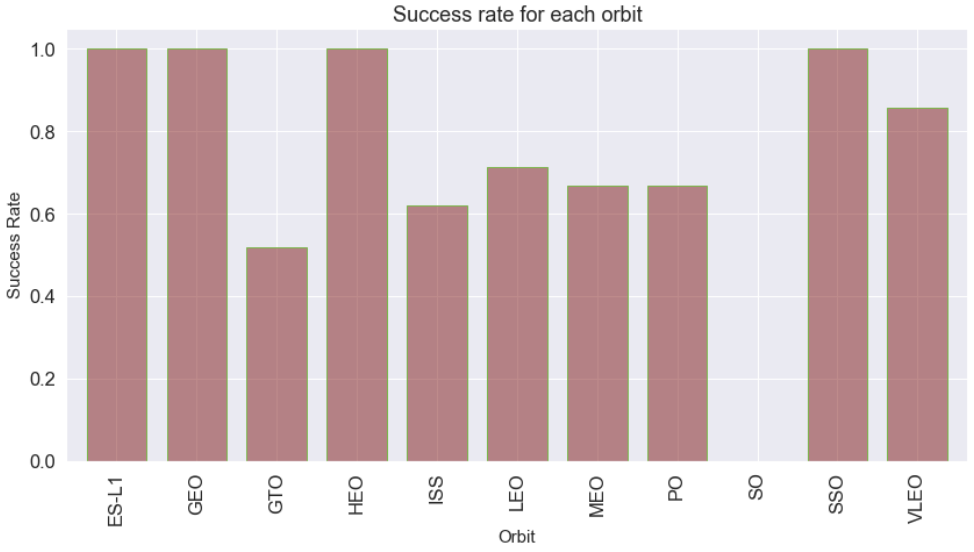


Success Rate vs orbit type

GTO sees the lowest success rate as suggested in previous slide.
SSO (polar low orbit) **the highest one.**

Success rate may strongly depend on both:

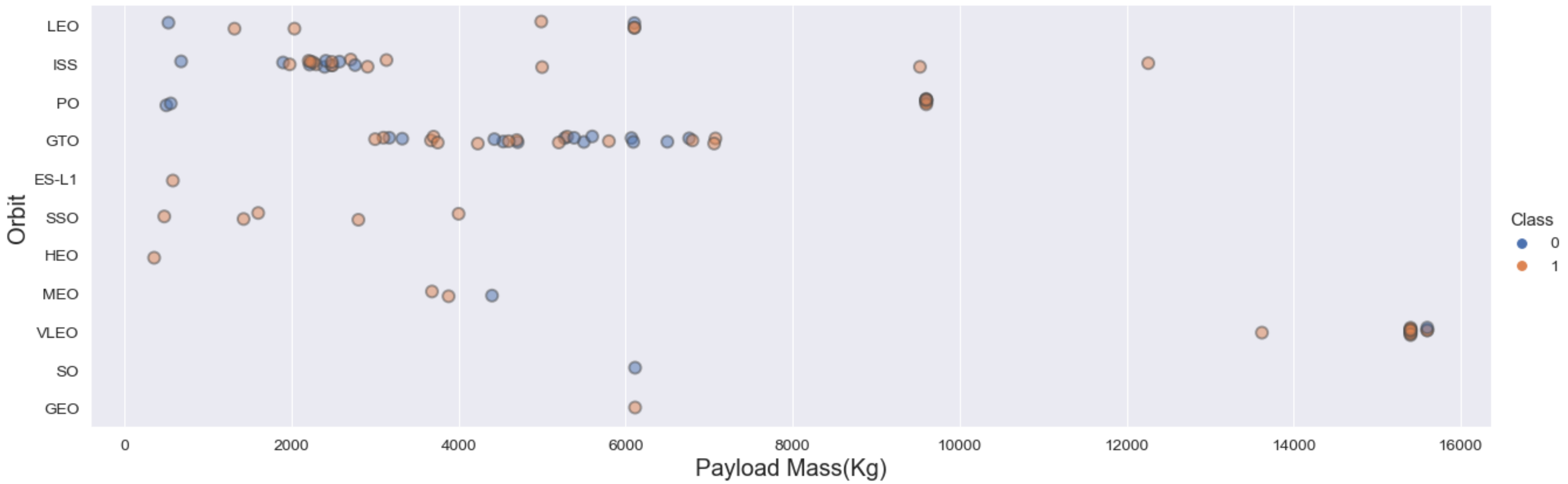
- payload mass
- orbit.



Orbit Type vs. Payload

The chart brings final info about "Orbit v. Payload". It describes the distribution "success rate v. (payload, orbit)"
Main trends:

- Maximum success rate with: low orbit except (ISS) and low payload mass
- ISS: based on "Orbit Type v. Flight Number" 5/8 failures occurred in the early stage of Falcon 9 project. When Falcon 9 reliability was low.

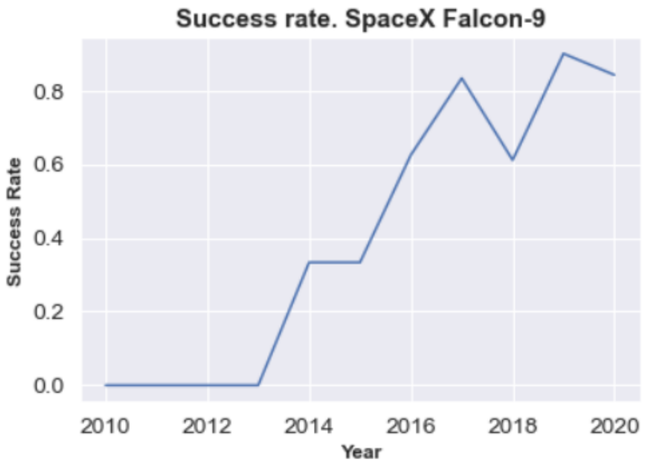


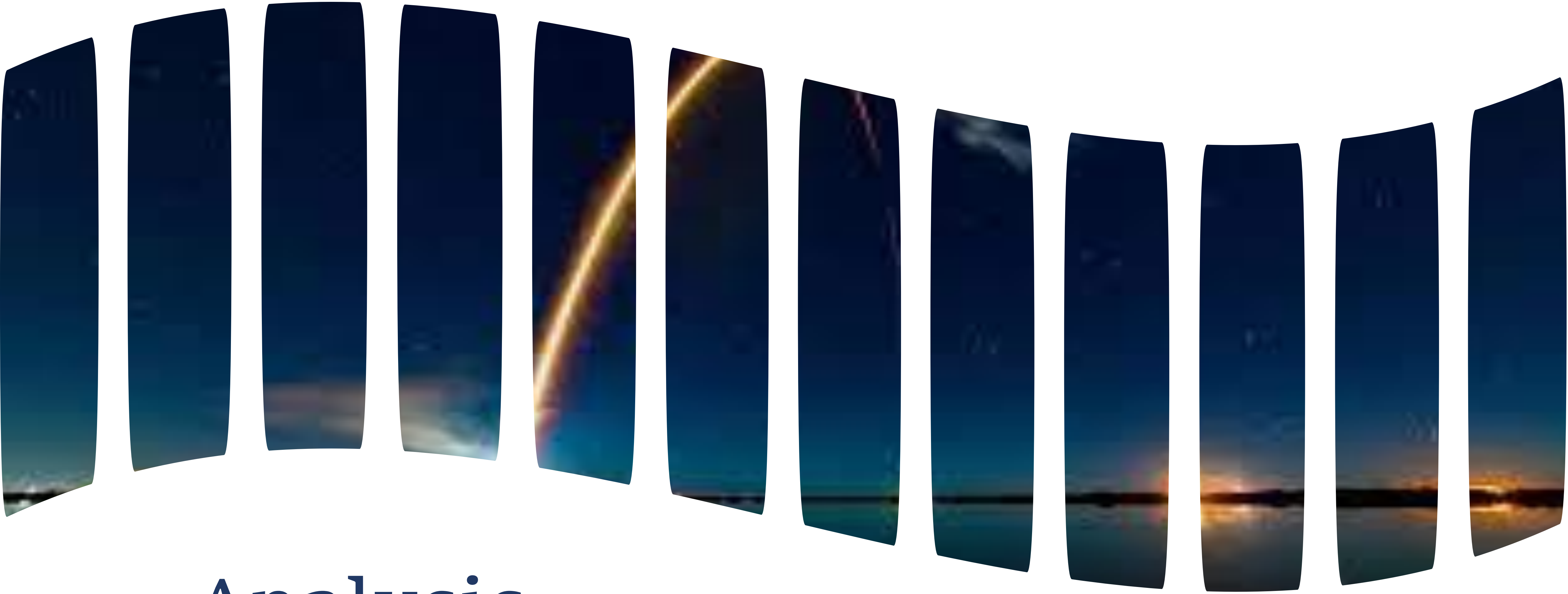
Launch Success Yearly Trend

Falcon 9 reliability significantly improves over time .
Success rate, **here defined after successful booster recovery for Falcon9,** depends on:

- Payload mass
- Orbit
- + other factors

- Falcon9 average booster recovery success rate si 66%.
- Success rate currently sufficient for SpaceX financial viability.





Analysis

Launch Site Names & Records

Launch Sites

```
df_unique_launchsites=pd.read_sql_query("Select distinct Launch_Site  from spacex_v11 ",conn)
print(df_unique_launchsites)
```

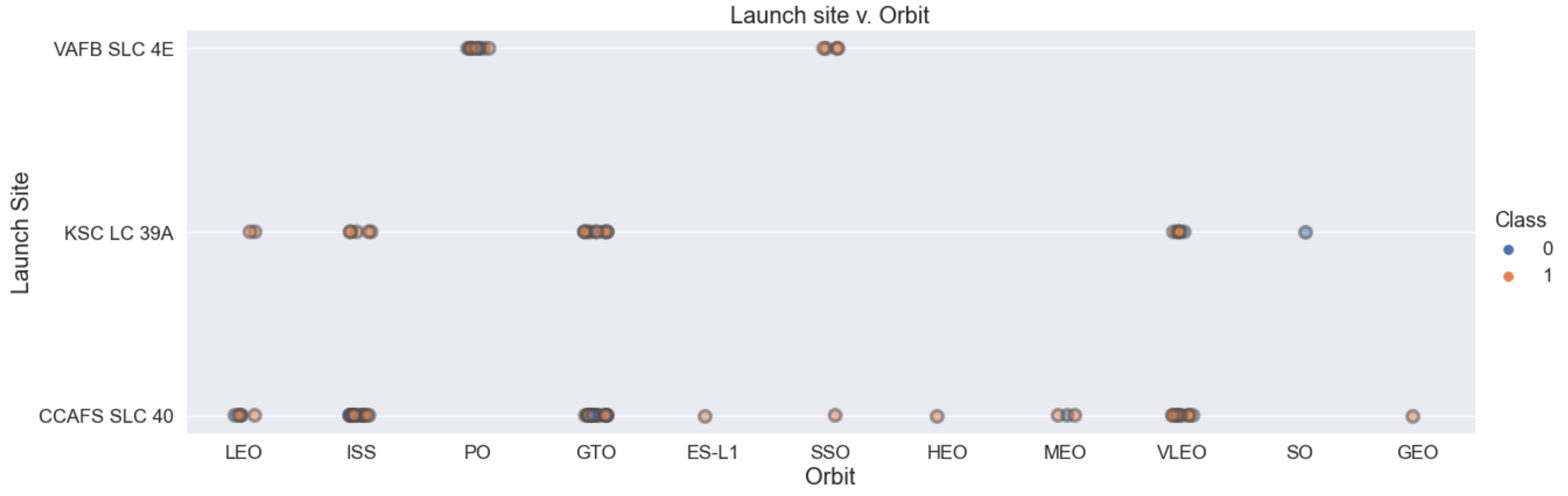
```
    Launch_Site
0   CCAFS LC-40
1   VAFB SLC-4E
2    KSC LC-39A
3  CCAFS SLC-40
```

5 records where launch sites begin with `CCA

```
df_launchsites_CCA5=pd.read_sql_query("Select * from spacex_v11 where Launch_Site Like 'CCA%' Limit 5",conn)
df_launchsites_CCA5
```

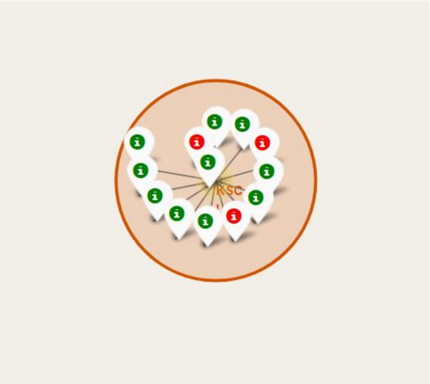
	id	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
0	1	2010-04-06	0 days 18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
1	2	2010-08-12	0 days 15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2	3	2012-05-22	0 days 07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
3	4	2012-10-08	0 days 00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
4	5	2013-03-01	0 days 15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Launch Site vs Orbit Type

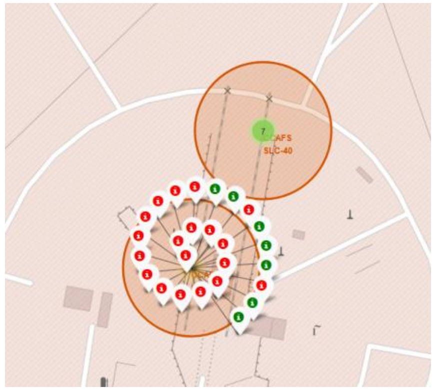


Falcon 9 Success/Failed launches

Vandenberg Space Launch Complex 4 (CA)
VAFB SLC-4E



Kennedy Space Center (FL)
KSC LC 39A



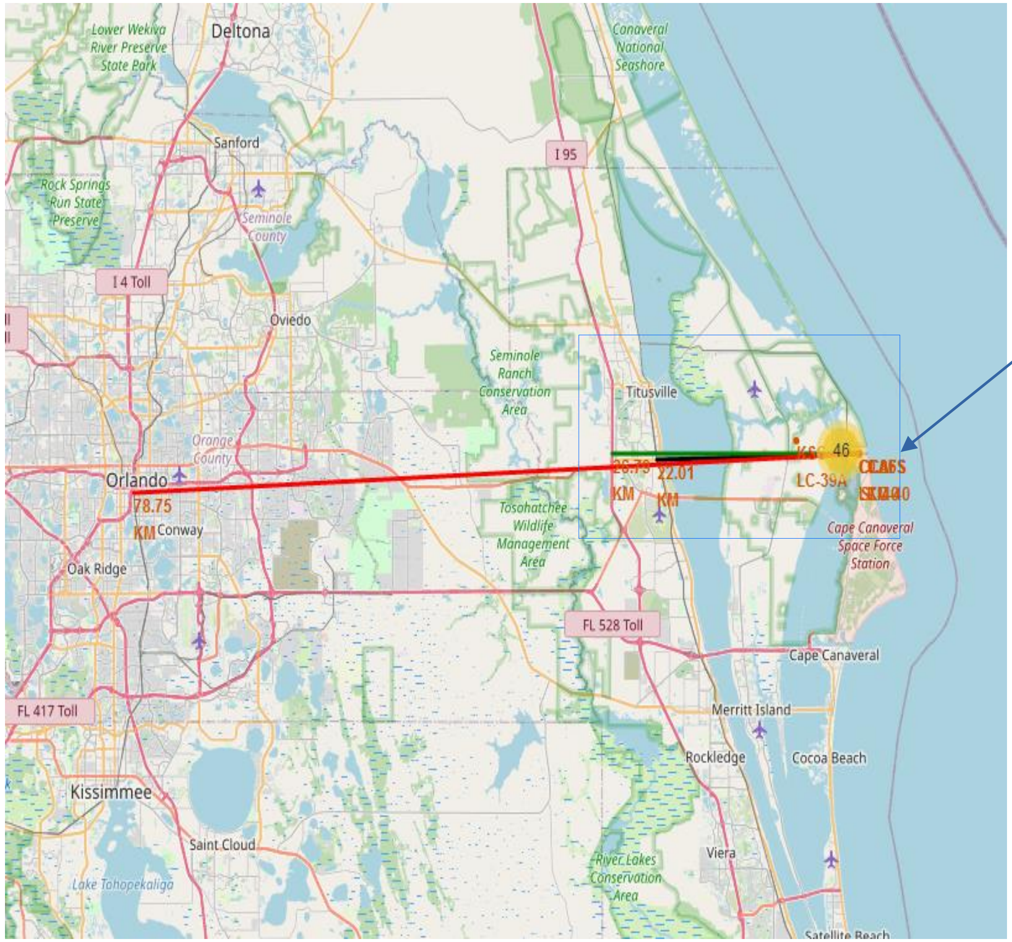
Cape Canaveral (FL)
CCAFS-LC40

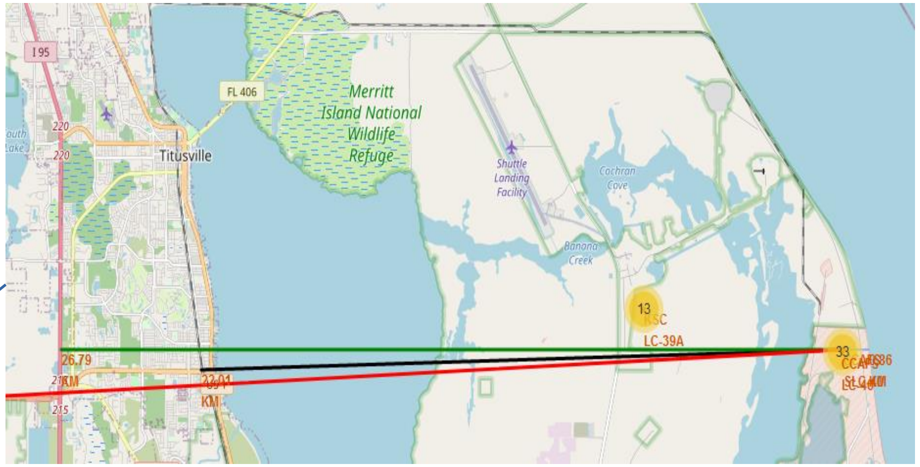
Cape Canaveral (FL)
CCAFS-SLC40

Launch Site	class	
CCAFS LC-40	0	19
	1	7
CCAFS SLC-40	0	4
	1	3
KSC LC-39A	0	3
	1	10
VAFB SLC-4E	0	6
	1	4

Table: Synthesis of launches outcomes
Class 0= failure
Class 1= success

Distances between a launch site to its proximities





Distance from CCAFS_SLC40 to:

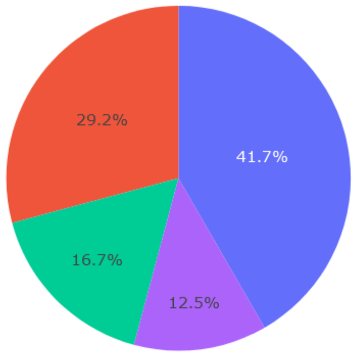
- **Closest coast: ~900 m**
- **Florida East Coast Railway: 22.0 km**
- **Highway I 95: 26.8 km**
- **Orlando: 78.75 km**

SpaceX Falcon 9: Launch success count for all sites

SpaceX Launch Records Dashboard

All Sites

Share of Successful Launches by Site (%)



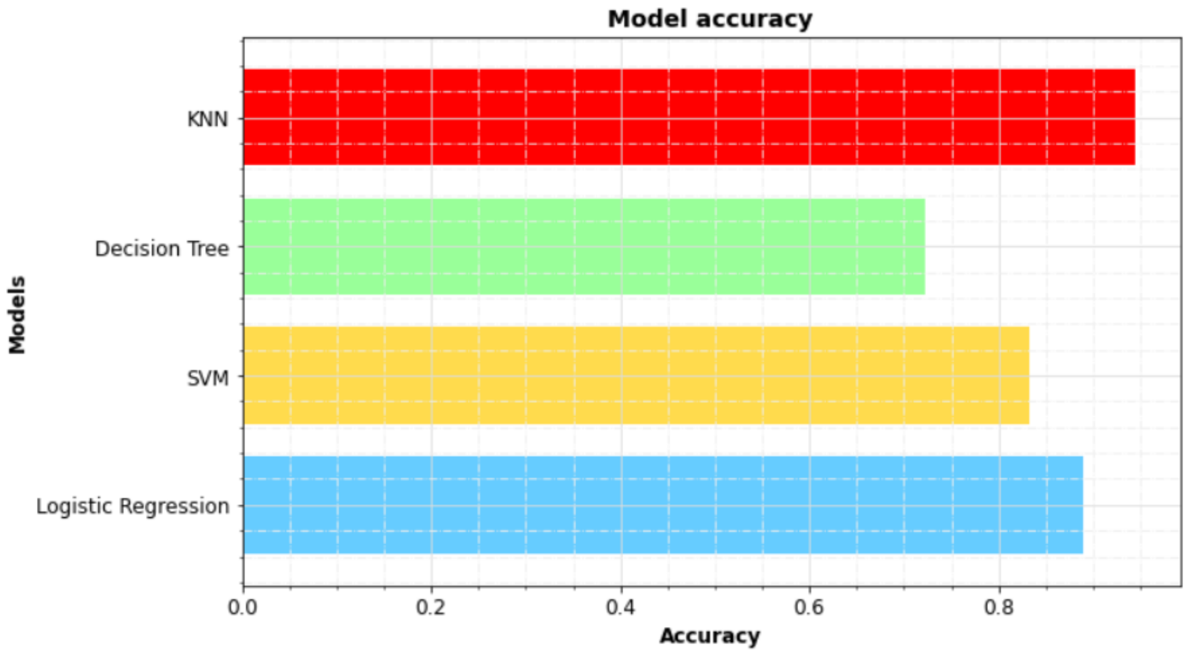
Launch Site	class	
CCAFS LC-40	0	19
	1	7
CCAFS SLC-40	0	4
	1	3
KSC LC-39A	0	3
	1	10
VAFB SLC-4E	0	6
	1	4

Classification Accuracy

Classification Accuracy with test set.

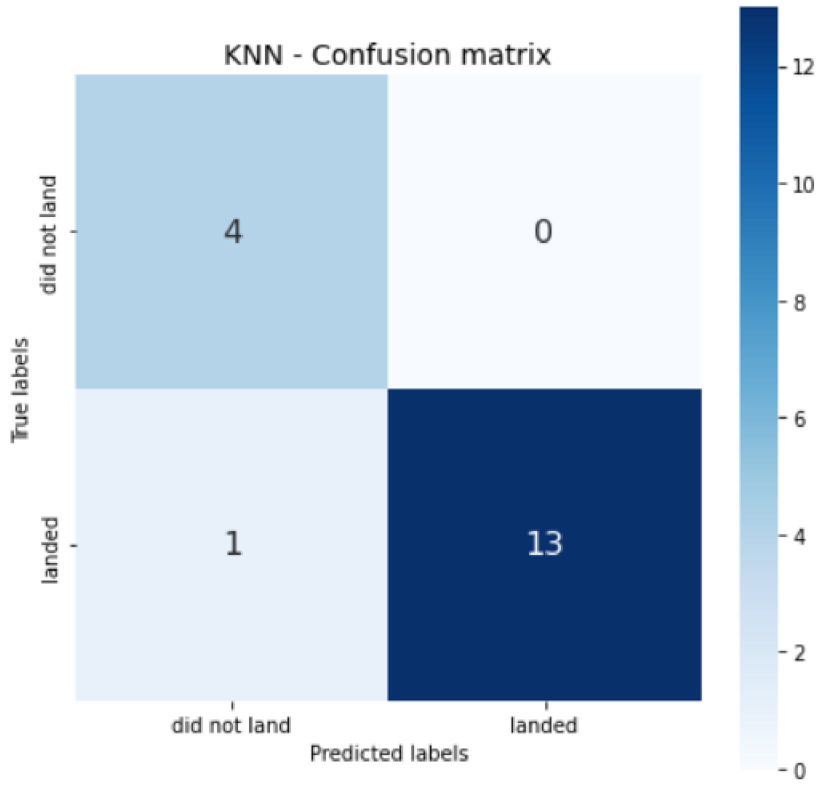
Results with “train test split” random_state=3

- Optimization of SVM and LR hyper-parameters was refined for increasing accuracy with train set.
- It did not necessarily improved accuracy with test set.
- Test set is too small.
- In our case, KNN exhibits the best accuracy: ~94%



Confusion Matrix

- k-nearest neighbors algorithm (k-NN) is the best “predictor”
- The model perfectly predicts mission failure
- 1 false negative for successful booster landing (recovery)



Conclusions

From Falcon 9 and Wiki we imported data and developed tools for investigating factors driving successful booster recovery:
Payload, Orbit, Booster version, Launching sites.

SpaceX GEO/GTO flights are more risky.
Failure could be linked to energy deployed at lift-off, vibrations damaging of electronics, control systems of the booster.

Even with recent recovery failure for GTO/GEO, SpaceX will maintain a sufficient lead in terms of cost per kg.

Based on these features, we selected a supervised classification model capable of predicting launch outcome with a 94% accuracy.
Falcon9 booster recovery process is more risky than a classic launcher, but its success rate:

-increases overtime

-at 66%, it is sufficient for maintaining a competitive cost per kg

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 query
q_count_landing= """ Select Landing_Outcome, count(*) as count_landings
                    from spacex_v11
                    where Date between '2010-06-04' and '2017-03-20'
                    group by Landing_Outcome
                    order by count_landings desc  """
```

```
count_landing= pd.read_sql_query(q_count_landing,conn)
count_landing.head(10)
```

	Landing_Outcome	count_landings
0	No attempt	10
1	Failure (drone ship)	5
2	Success (drone ship)	5
3	Controlled (ocean)	3
4	Success (ground pad)	3
5	Uncontrolled (ocean)	2
6	Failure (parachute)	1
7	Precluded (drone ship)	1

Thank You! !

Thank you !