

Winning Space Race with Data Science

SANOGO Souleymane
2024



Outline

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

Executive Summary

Business Context:

The competitive space launch industry demands precise cost estimation and reliable booster recovery predictions. The success of Falcon 9 missions, hinging on the first stage landing, is integral to SpaceX's cost-effective approach, influencing competitiveness and pricing.

Objectives:

Our primary objective is to develop a robust predictive model for Falcon 9 first stage landings. This model aims to assist competing companies in making informed decisions during rocket launch contract bidding, contributing to cost efficiency and competitiveness.

Dataset Overview:

The dataset encompasses essential details from multiple Falcon 9 launches, providing a comprehensive foundation for analysis. Key features such as flight number, payload mass, and outcome are crucial

Modeling:

Implemented a Random Forest Classifier to predict Falcon 9 first stage landing success using features like orbit, payload mass, and others.

Evaluation:

1. Achieved a model accuracy of approximately 94%, signifying robust predictive capabilities.

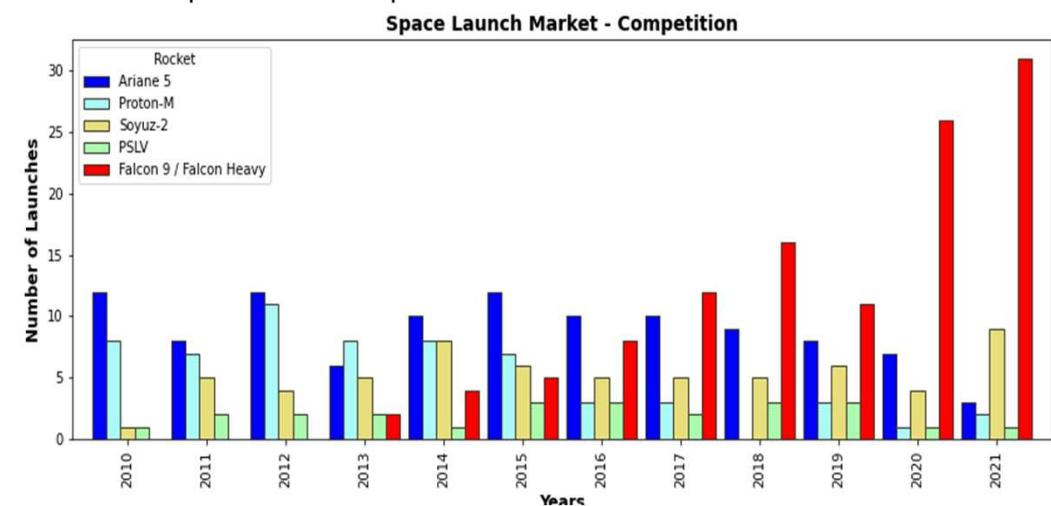
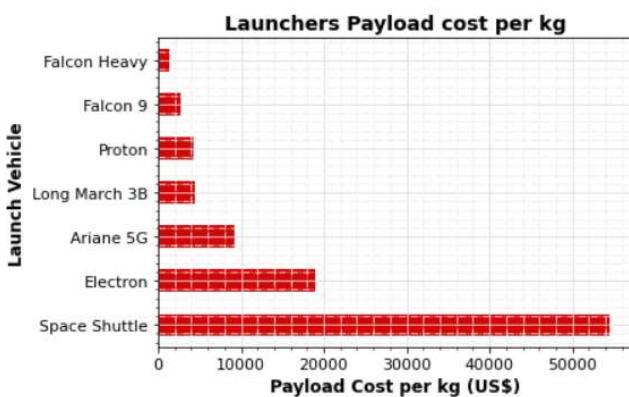
Introduction

- This report focuses on predicting Falcon 9 first stage landing success, impacting launch costs and competitive positioning. The dataset includes critical launch details, and our goal is to develop a robust predictive model.
- **Business Context**

SpaceX's reusable launcher concept has revolutionized satellite launches, offering cost advantages. The success of Falcon 9 missions, defined by booster recovery, depends on various features.

Introduction - Business understanding

- Commercial rocket launches is a very competitive market.
- From 2000 to 2017, Ariane 5 (ESA - Arianespace) dominated the market with: lower cost, high payload mass, **reliability**.
- In terms of cost, China, India, Russia are tough competitors with less reliable classic launchers like LongMarch, Proton, Soyuz-2, PSLV. New players like BlueHorizon are still in early stage.
- Since 2017, SpaceX Falcon9/Falcon9 Heavt 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.



Introduction - Business understanding

- **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
- SpaceX Starship will further crush the competition in both domains.
- Falcon9 and Starship costs are way lower due to reuse of boosters
- Ariane-6 will be a classic launcher for reliability purposes
- Falcon9 reliability can't compete with Ariane-5G record
- **Falcon9 booster recovery success rate is still an issue.**



Ariane 64 (4 boosters)

SpaceX Starship

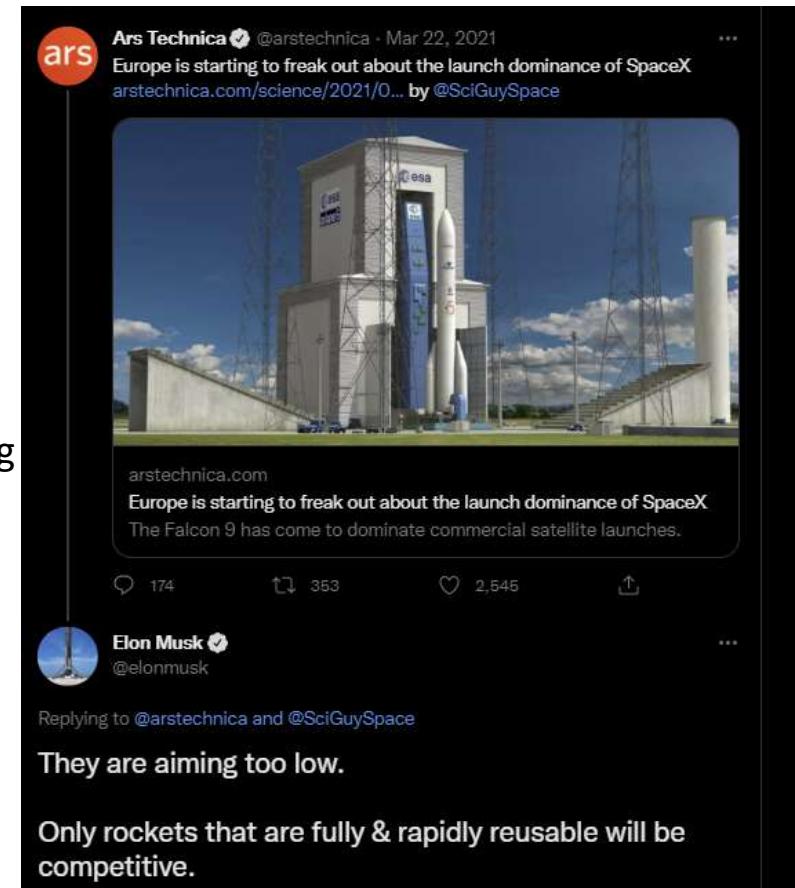
Launcher	Height	Weight	Payload LEO (kg)	Payload GTO-GEO (kg)	Mission to Mars
Ariane 5	20	20	6600	>10000	no
Ariane 6 (64)	21	21	21600	>11500	yes
Falcon9 / Falcon9 Heavy	19	19	22800 / 63800	8300	yes
Starship	18	18	100k	22000	yes

Introduction - Business understanding

SpaceX CEO Elon Musk has criticized European efforts in the emergent new space race. In March 2021, he wrote on Twitter:

"They are aiming too low. Only rockets that are fully & rapidly reusable will be competitive. Everything else will seem like a cloth biplane in the age of jets."

- Reliability v. Risky concept: “reusability” and Significant cost reduction
- SpaceX financial viability heavily depends on the successful landing of the booster (booster recovery) defined as “mission success”.
- 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.



Section 1

Methodology

Methodology

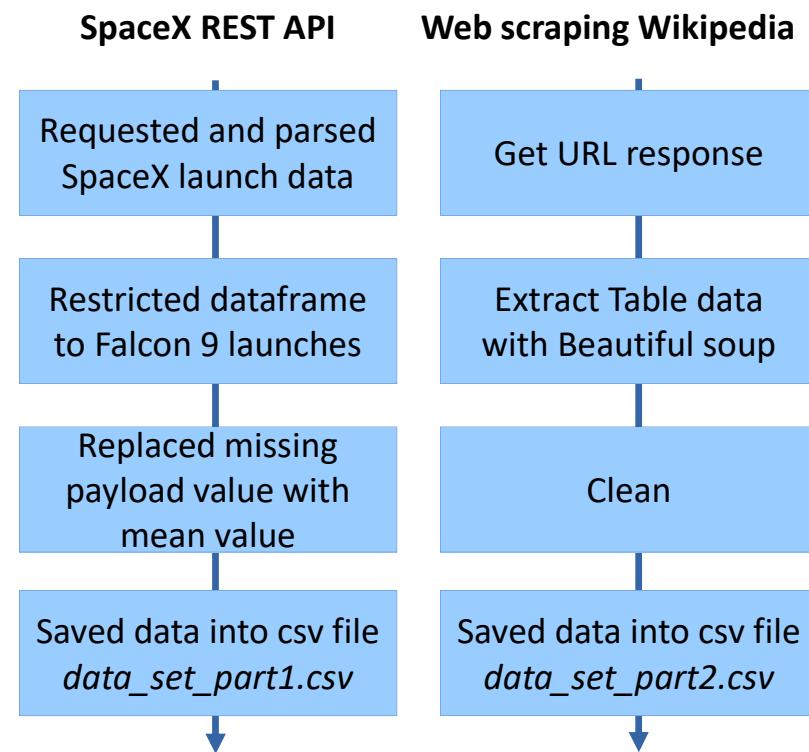
Methodology steps:

1. Data collection from open Data base and Wikipedia (Falcon9, Ariane-5).
2. Data wrangling.
3. Exploratory data analysis using SQL query and visualization of correlation between parameters.
4. Visual analytics: launch sites with Folium, success rates with Plotly Dash.
5. Classification Models development and validations. Selection of best predictive model.

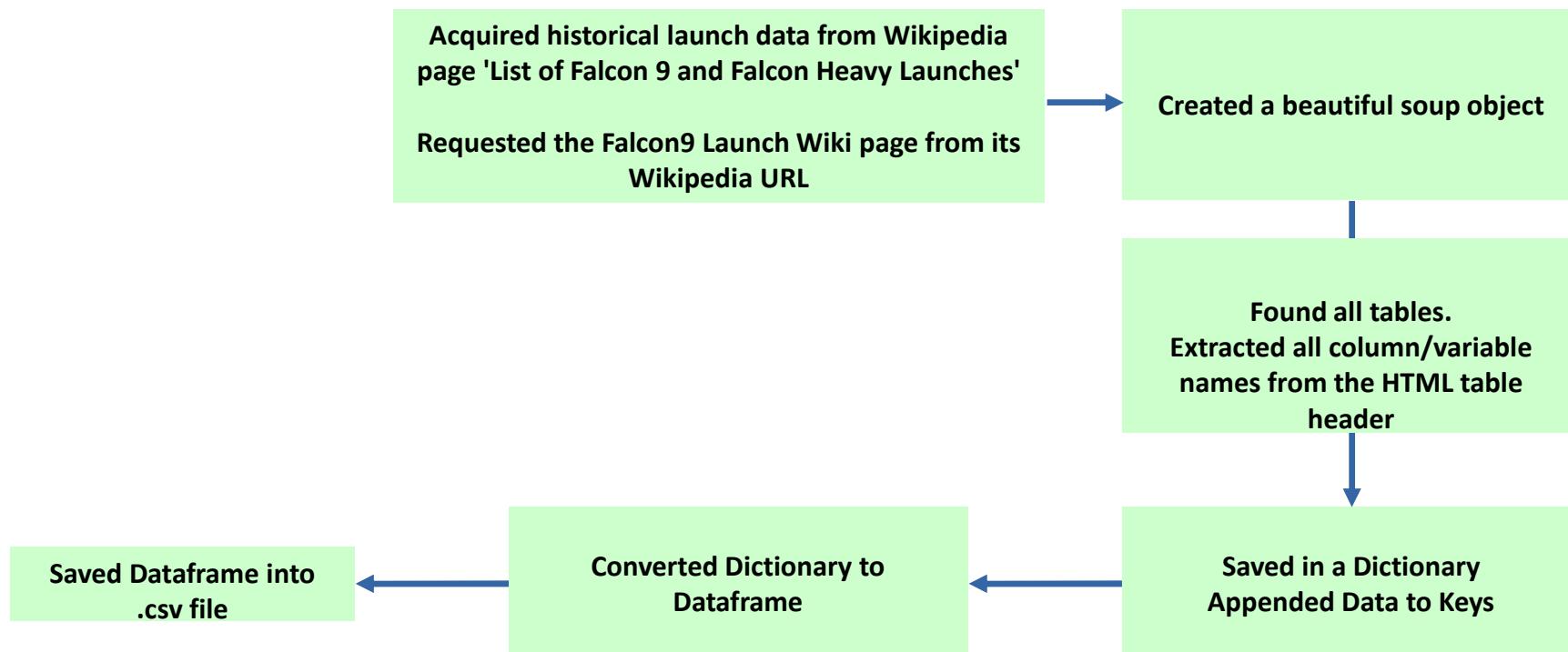
Data Collection

Data was collected from:

- open source SpaceX REST API
- webscraping Falcon9 launch data in Wikipedia
- webscraping Ariane5 launcher data in Wikipedia **for reference graphs only.** (not prepared for ML)



Data Collection - Scraping



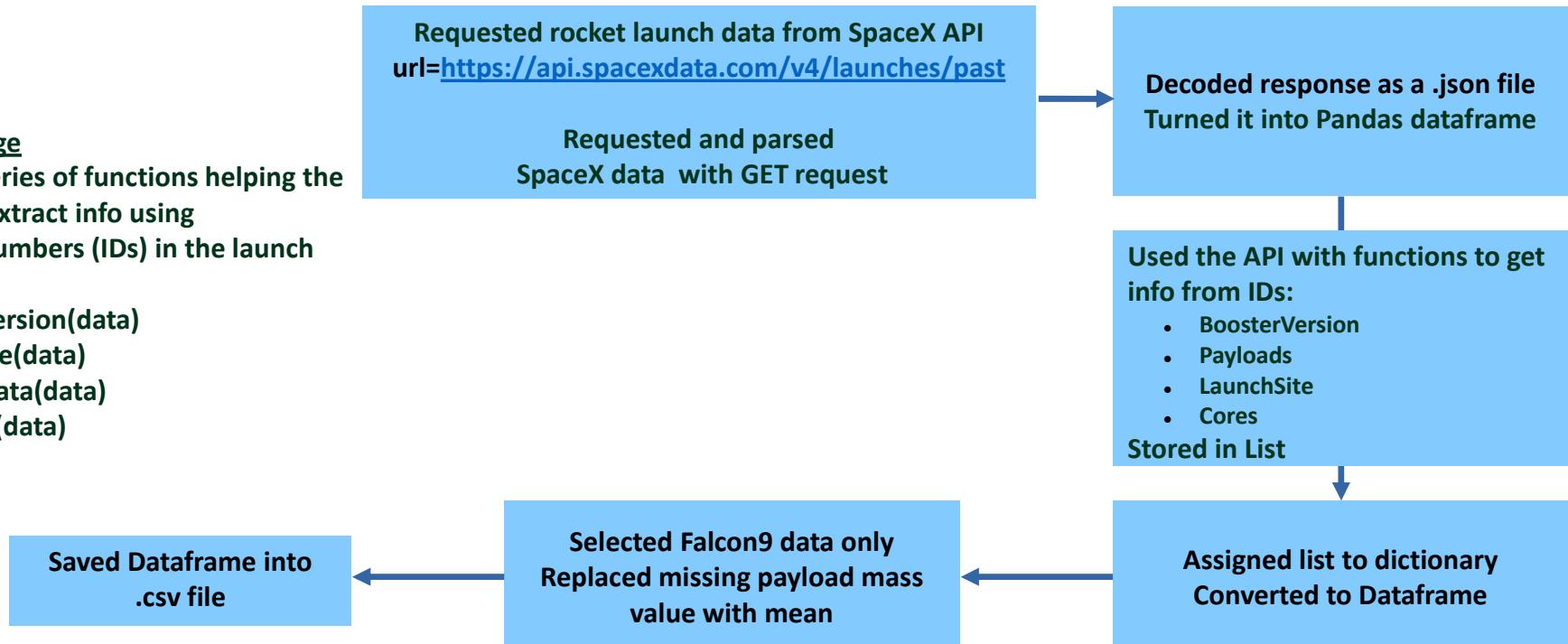
https://github.com/DrStef/Applied_Data_Science_Capstone/blob/main/GH_jupyter-labs-webscraping_v4.ipynb

Data Collection – SpaceX API

Preliminary Stage

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

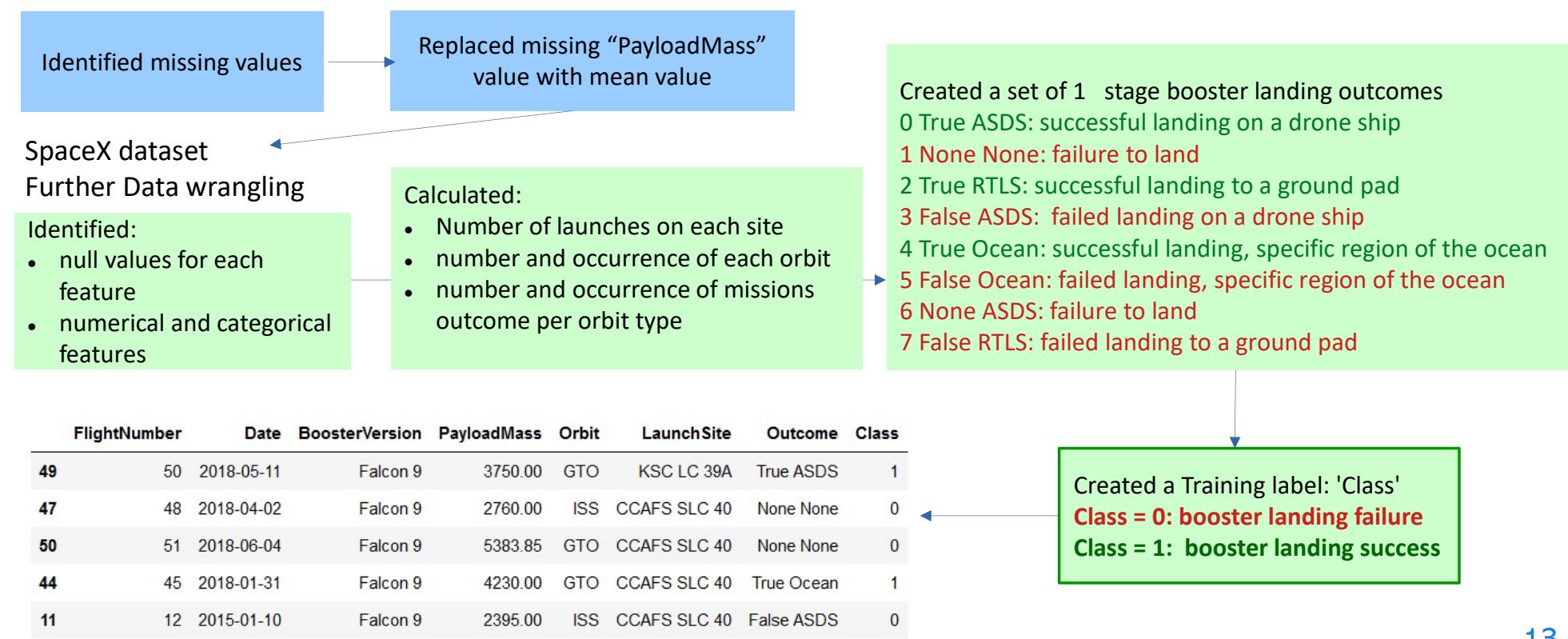
- `getBoosterVersion(data)`
- `getLaunchSite(data)`
- `getPayloadData(data)`
- `getCoreData(data)`



Data Wrangling

Dataframe

From SpaceX API



https://github.com/DrStef/Applied_Data_Science_Capstone/blob/main/GH_labs-jupyter-spacex-Data%20wrangling_v3.ipynb

EDA with SQL

For the reviewer, for emulating SQL queries:

- I created a MySQL local server
- I created a database: 'spacex_database'
- I saved SpaceX csv file in Table "spacex_v11" (see picture)
- I established a connection "conn" with "spacex_database".

```
conn=pymysql.connect(host='localhost',
                      port=int(3306),
                      user='root',
                      passwd="##########",
                      db='spacex_database' )
```

We ran SQL queries with "pd.read_sql_query":

```
df=pd.read_sql_query("Select * from spacex_v11 ",conn)
df.head(5)
```

1. to download data in a Jupyter Notebook (dataframe)
2. to display information about:
 - Launch sites
 - Payload mass
 - Booster versions
 - Mission outcomes
 - Booster landings

The screenshot shows the MySQL Workbench interface. In the top-left pane, the 'Schemas' tree is expanded to show the 'spacex_database' schema, which contains several tables like 'new_table', 'spacex10', 'spacex_v1', 'spacex_v10', and 'spacex_v11'. The 'spacex_v11' table is selected. In the top-right pane, the 'query1*' tab displays the SQL code for creating the 'spacex_v11' table:

```
CREATE TABLE `spacex_v11` (
  `id` INT,
  `Date` DATE,
  `Time (UTC)` TIME(1),
  `Booster_Version` VARCHAR(45),
  `Launch_Site` VARCHAR(45),
  `Payload` VARCHAR(100),
  `PAYLOAD_MASS__KG_` INT,
  `Orbit` VARCHAR(45),
  `Customer` VARCHAR(100),
  `Mission_Outcome` VARCHAR(45),
  `Landing_Outcome` VARCHAR(45)
)
```

Below this, the 'Result Grid' pane shows the first 10 rows of data from the 'spacex_v11' table:

	id	Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer
1	2010-04-06	18:45:00.0	F9 v1.0	B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX
2	2010-08-12	15:43:00.0	F9 v1.0	B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of...	0	LEO (ISS)	NASA (COT)
3	2012-05-22	07:44:00.0	F9 v1.0	B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COT)
4	2012-10-08	00:35:00.0	F9 v1.0	B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)
5	2013-03-01	15:10:00.0	F9 v1.0	B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)
6	2013-09-29	16:00:00.0	F9 v1.1	B1003	VAFB SLC-4E	CASSIOPE	500	Polar LEO	MDA
7	2013-12-03	22:41:00.0	F9 v1.1		CCAFS LC-40	SES-8	3170	GTO	SES
8	2014-01-06	22:06:00.0	F9 v1.1		CCAFS LC-40	Thaicom 6	3325	GTO	Thaicom
9	2014-04-18	19:25:00.0	F9 v1.1		CCAFS LC-40	SpaceX CRS-3	2296	LEO (ISS)	NASA (CRS)

The bottom pane shows the 'Object Info' and 'Session' tabs.

MySQL local server
Building "spacex_v11" table.

https://github.com/DrStef/Applied_Data_Science_Capstone/blob/main/GH_jupyter-labs-eda-sql-coursera_v10.ipynb

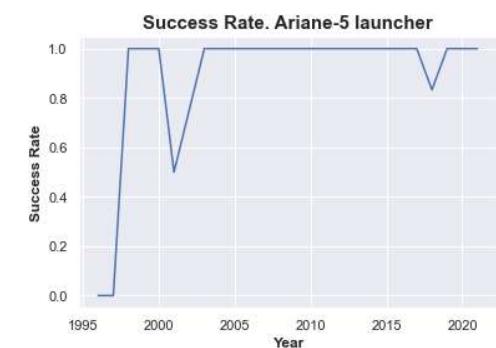
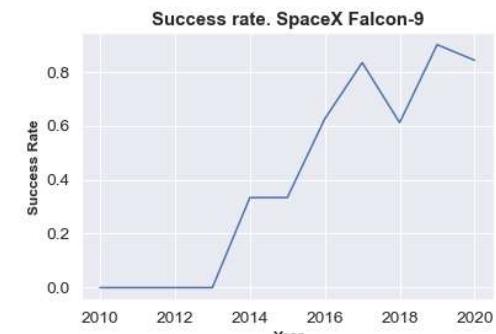
EDA with Data Visualization

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
- Falcon 9 & Ariane-5 launch success yearly trend.



Build an Interactive Map with Folium

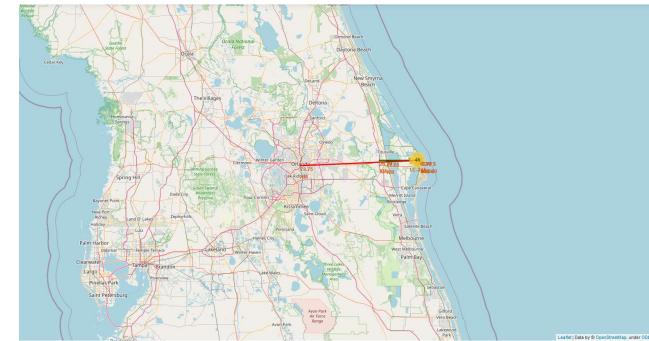
Launch success rate may depend on the location and proximity of a launch site. **Folium Interactive Map was used for visualizing and analyzing SpaceX Launch Sites.**

- 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

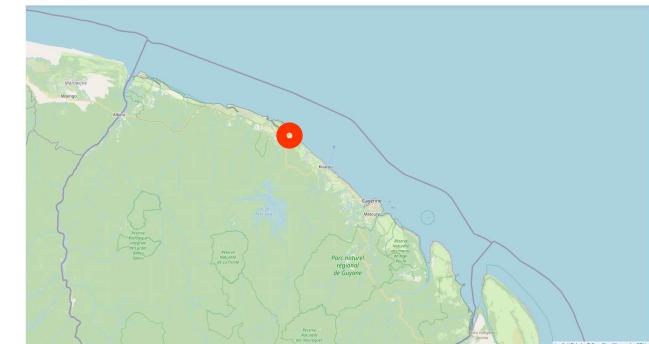
Calculated the distance between a launch site (CCAFS_SLC40 in Cape Canaveral, FL) and:

- Closest coastline
- Closest high traffic density railway: Florida East Coast Railway
- Closest high traffic density highway: Interstate I95
- Closest high density urban area: Orlando (FL)

For reference, we added the localization of European Space Agency (ESA) /ArianeEspace Ariane 5 and Soyuz launch pads in Kourou, French Guiana.



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



Ariane launch pad - Kourou in French Guiana
Coordinates: -52.792°, 5.265° (~ Equator)

https://nbviewer.org/github/DrStef/Applied_Data_Science_Capstone/blob/main/GH_lab_jupyter_launch_site_location_v4.ipynb

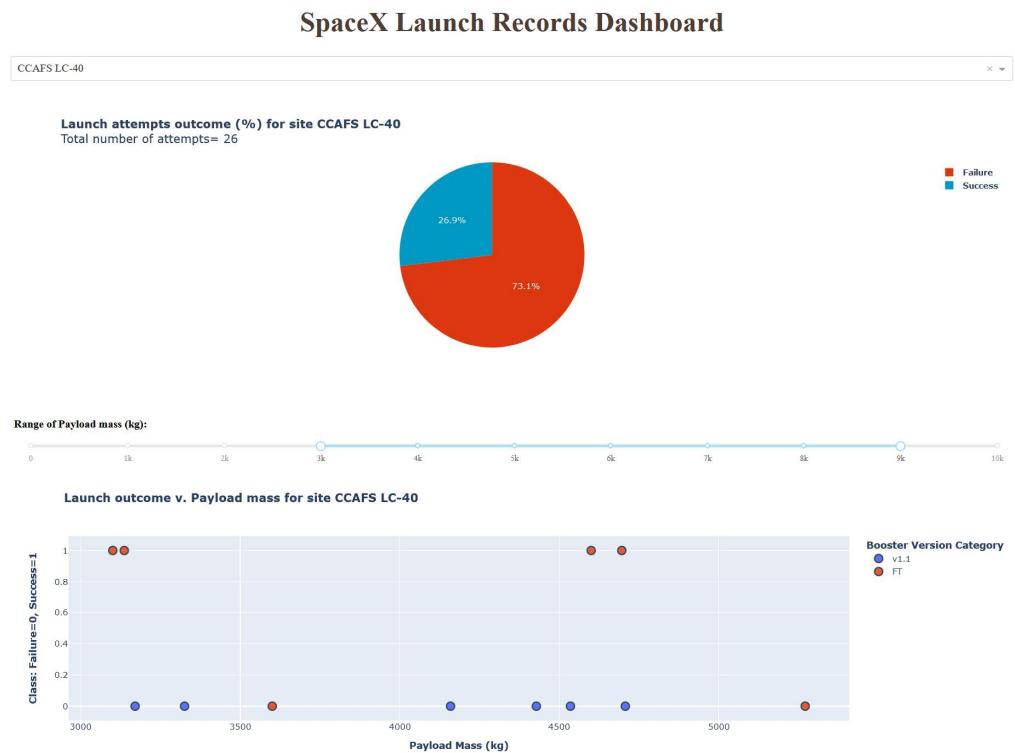
Build a Dashboard with Plotly Dash

We built an interactive dashboard with Plotly including:

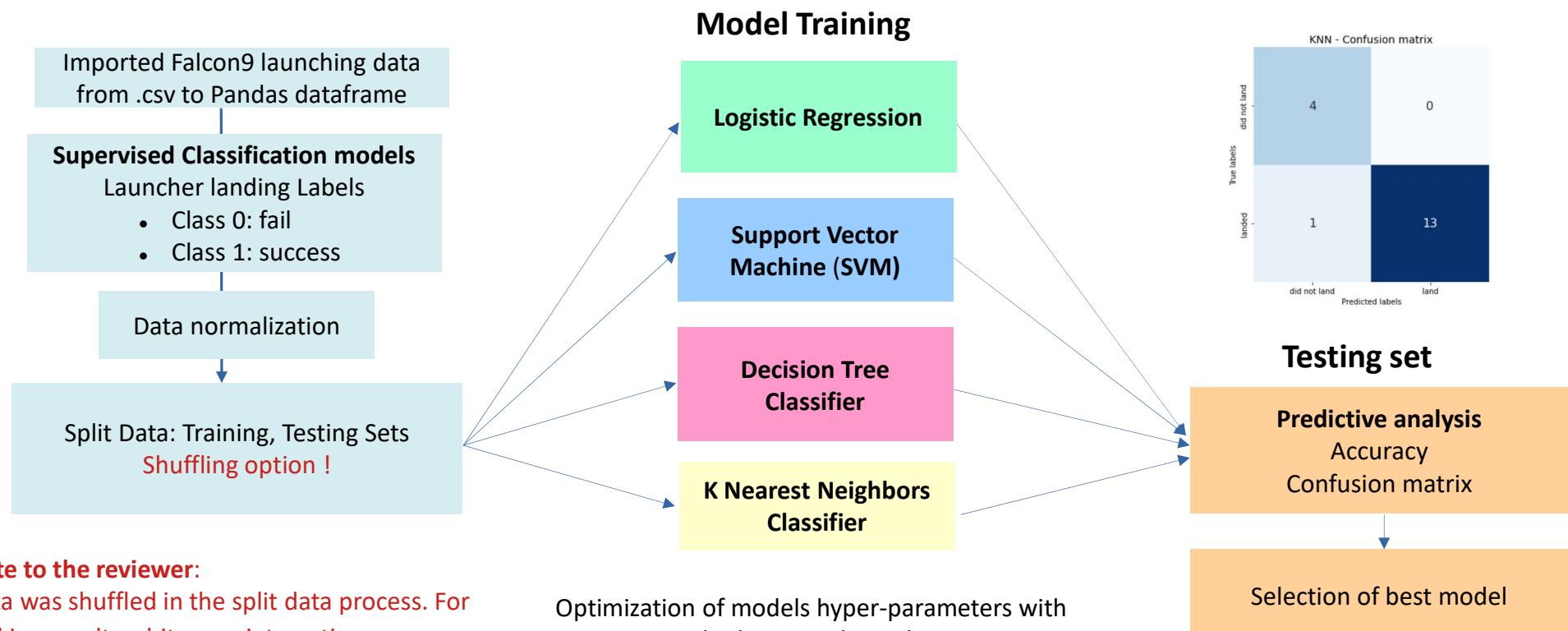
- Dropdown menu for selecting launch sites
- Pie charts displaying success rate.
- Scatter chart displaying launch site, payload mass, success/failure
- Range slider for selecting range of payload mass (kg).

for analyzing SpaceX launch records features:

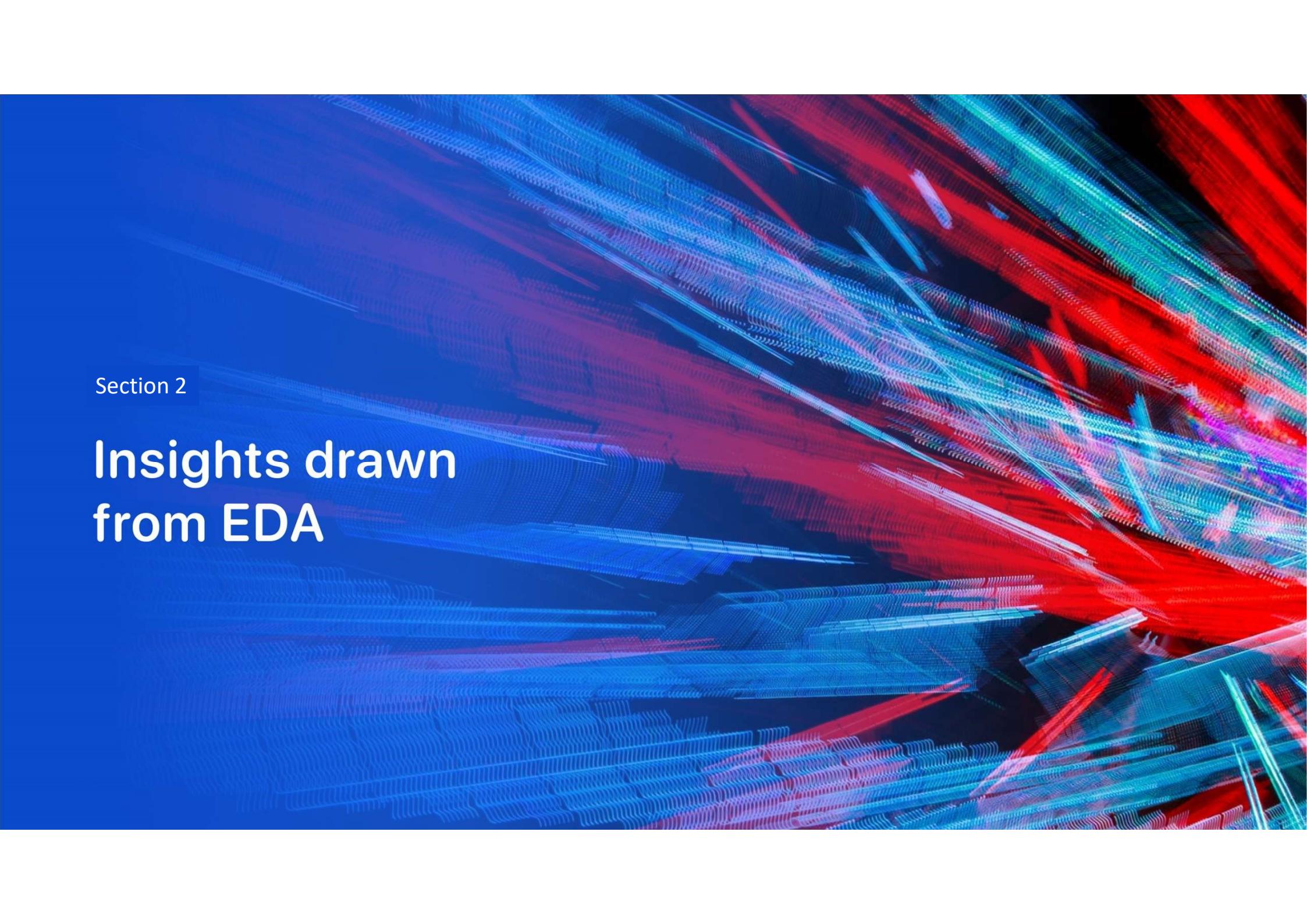
- site with largest successful launches.
- site with highest launch success rate
- payload range(s) with highest launch success rate
- payload range(s) with lowest launch success rate
- F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) with highest launch success rate.



Predictive Analysis (Classification)



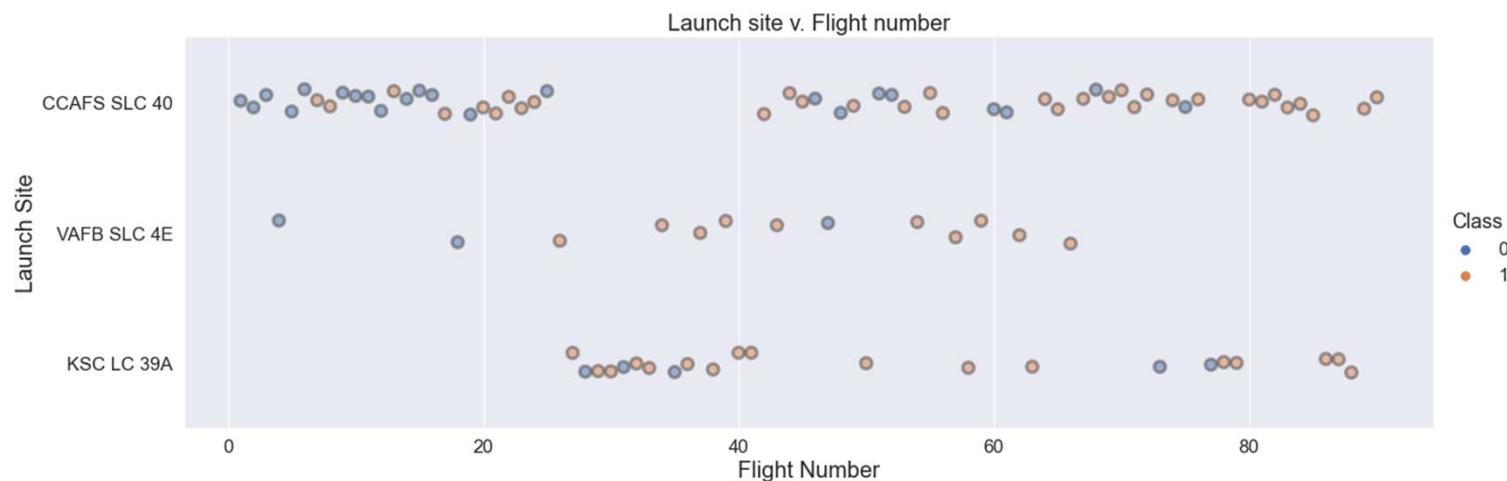
https://github.com/DrStef/Applied_Data_Science_Capstone/blob/main/GH_SpaceX_Machine%20Learning%20Prediction_Part_5_v8_shuffling.ipynb

The background of the slide features a complex, abstract pattern of glowing lines. These lines are primarily blue and red, with some green and white highlights. They form a dense, three-dimensional grid-like structure that curves and twists across the frame. The lines have a slight glow and appear to be composed of many small, individual points or pixels. The overall effect is one of motion, depth, and digital complexity.

Section 2

Insights drawn from EDA

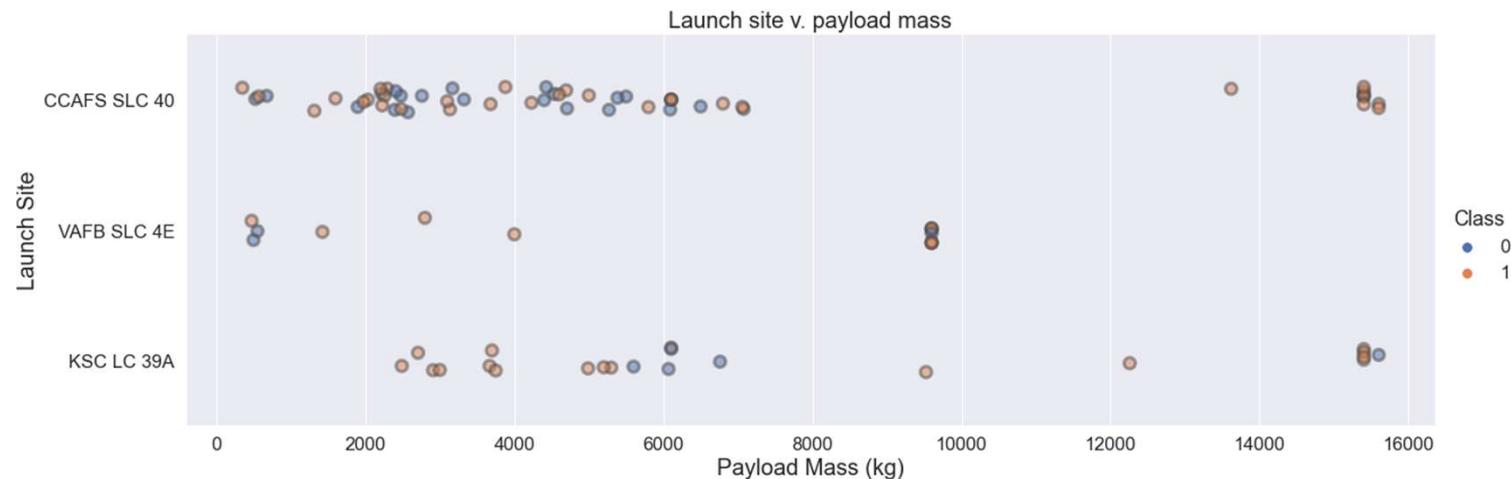
Launch Site v. Flight Number



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

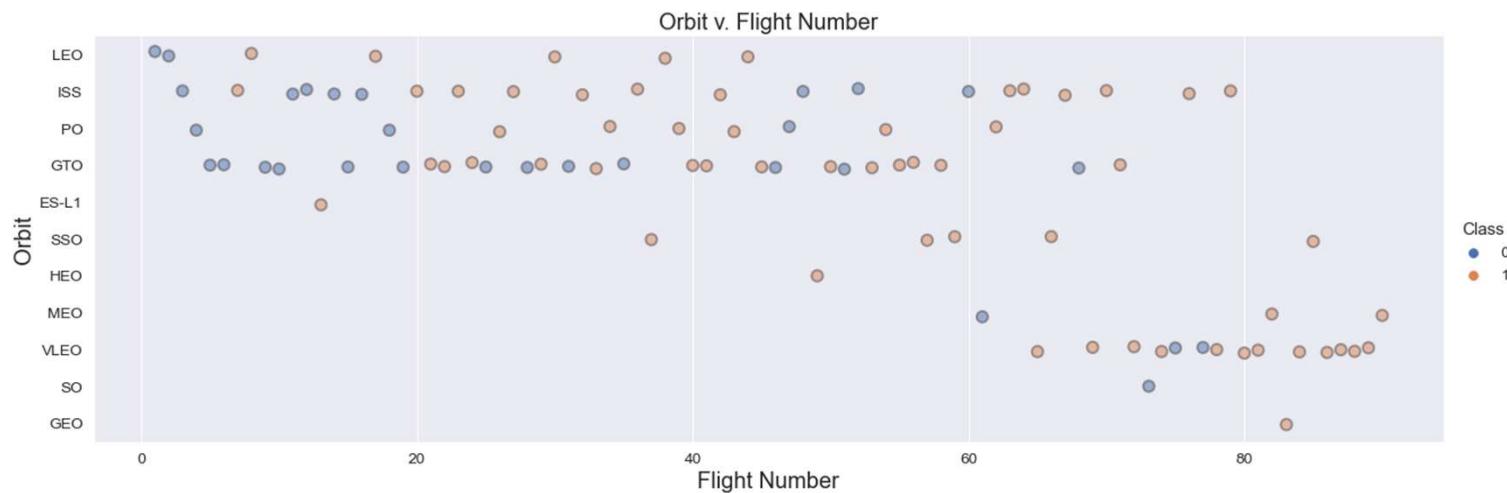
Launch Site v. Payload



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 v. 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

Remarks:

- GTO is a transfer orbit to GEO. Low thrust engines of the payload (satellite) complete the orbiting phase.
- We ignore results: GEO, SO, HEO, ESL-1, MEO. The number of flights is not significant.

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:

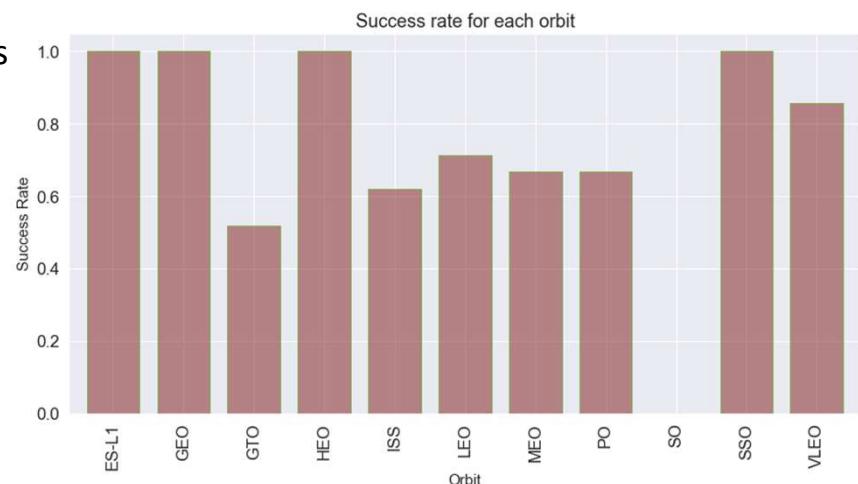
- payload mass
- orbit.

meaning the amount of energy deployed at lift-off, that may induce **strong noise/vibrations that are known to damage satellites***.

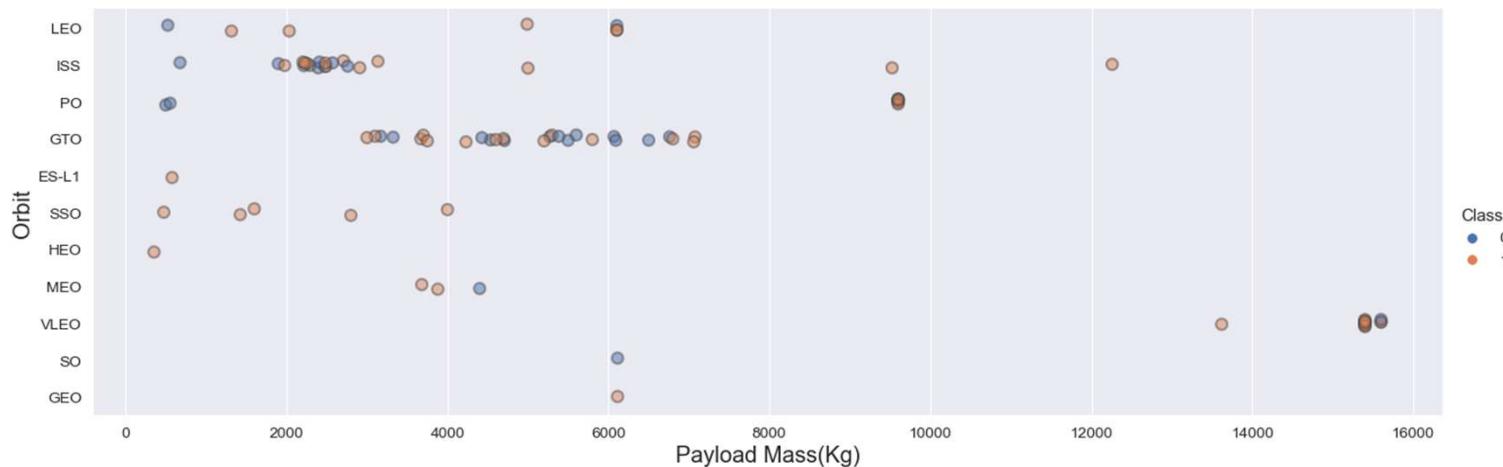
Vibrations could damage some of the booster electronics, inertial guidance systems... and cause booster recovery/landing failure.

We also need additional info about payload mass/orbit. Fortunately it is available.

* <https://adsabs.harvard.edu/full/1996ESASP.386..237F>



Orbit Type v. 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.
- Between 2000 and 7500 kg, success rate seems to be evenly distributed for GTO.
- Independently of payload mass, GTO is a risky “orbit” affecting missions success rate. Falcon 9 reliability improves over time, but there are still recent failed booster recovery after GTO launches.

Total Number of Successful and Failure Mission Outcomes

Total number of successful and failure mission outcomes (from SQL queries).

Here success is defined based on properly launching/orbiting payload. Success rate is very high: ~99% like Ariane-5.

Nevertheless, Falcon9 maintains a competitive advantage in terms of cost per kg compared with classic launchers like Ariane-5, **only if the reusable booster is recovered.**

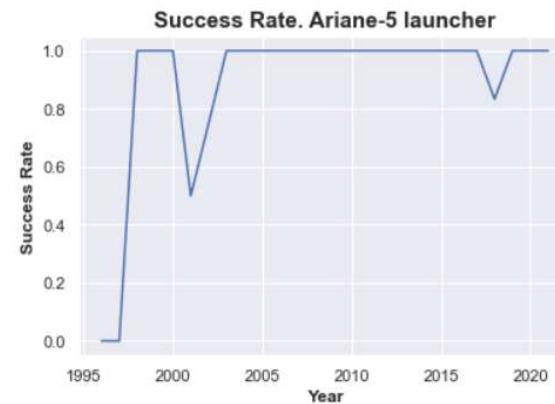
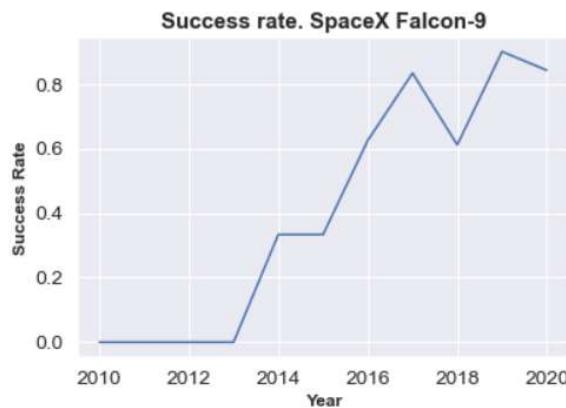
Therefore Falcon9 “success rate” in this report is defined after successful booster recovery (landing).

```
# sql query
qsf= """Select (Select Count(Mission_Outcome) from spacex_v11 where Mission_Outcome like '%Success%')
       as Successful_Missions,
       (Select Count(Mission_Outcome) from spacex_v11
        where Mission_Outcome like '%Failure%') as Failed_Missions """

success_failure= pd.read_sql_query(qsf,conn)
print(success_failure)

   Successful_Missions  Failed_Missions
0                  100                 1
```

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 we investigate next
- Independently of payload mass, orbits, **Ariane 5 has a close to 100% success rate** for 82 flights since 2003.
- Falcon9 average booster recovery success rate si 66%.
- Success rate currently sufficient for SpaceX financial viability.

The background of the slide is a nighttime satellite photograph of Earth. The curvature of the planet is visible against the dark void of space. City lights are scattered across continents as glowing yellow and white dots, with larger clusters appearing over major urban centers. In the upper right quadrant, a bright green aurora borealis or aurora australis is visible, appearing as a horizontal band of light.

Section 3

Launch Sites Proximities Analysis

Launch Site Names & Records

Before starting launch sites analysis, we list the names of all launch sites and some launch records (from SQL queries).

```
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
```

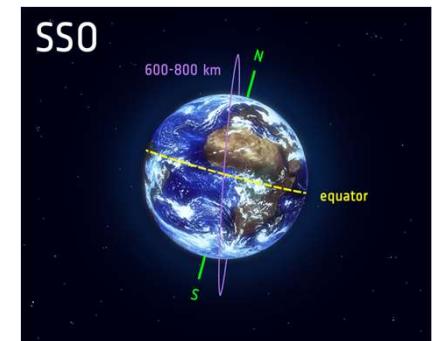
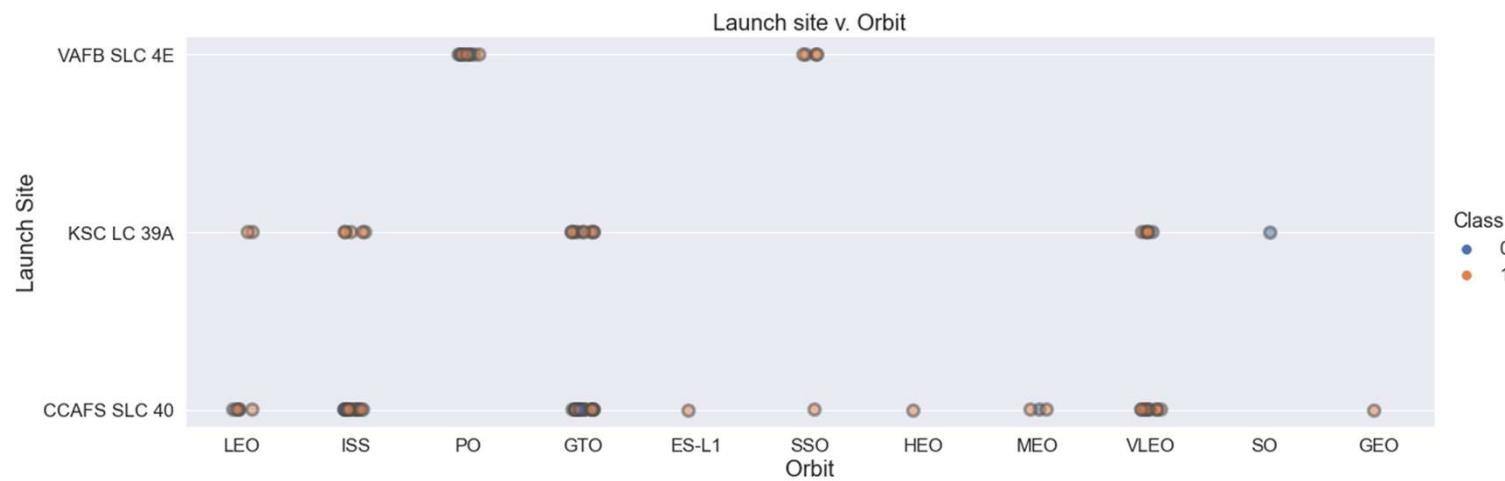
There are 4 distinct launch sites

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 v. Orbit type

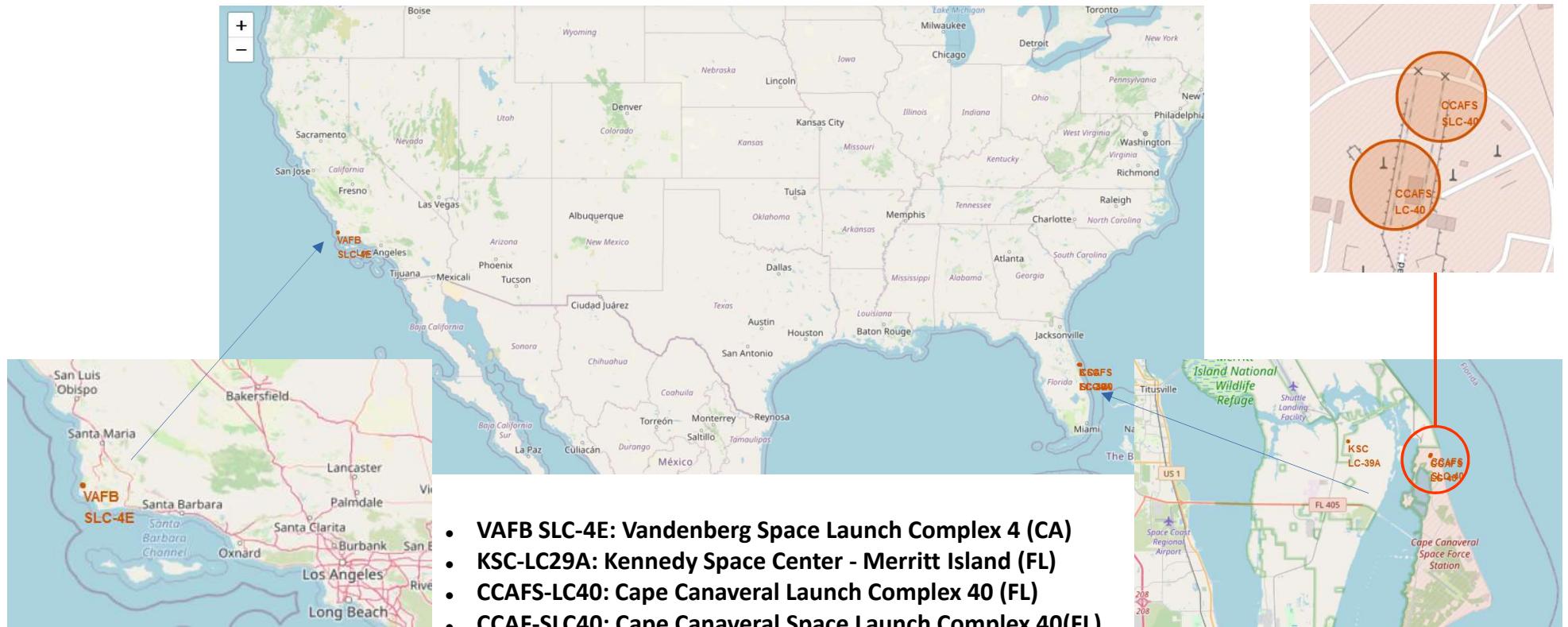


We added that scattered chart, which has implications in the analysis of launch sites.

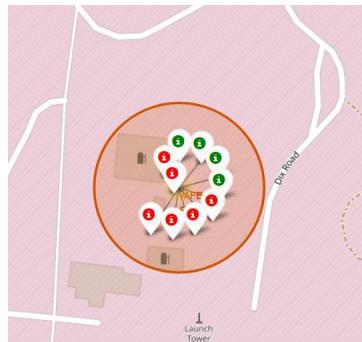
- VAFB SLC 4E launch site in California, is used only for PO and SSO orbits, **meaning low altitude polar orbits**. North or south? bound flights over the ocean.
- **All GTO/GEO launches take place in Florida launch sites.** As close as possible to the equator where earth tangential velocity is maximum, acting like a slingshot at lift off, from west to east over the ocean. (see gif). **This helps with orbiting GEO satellites with less energy.**



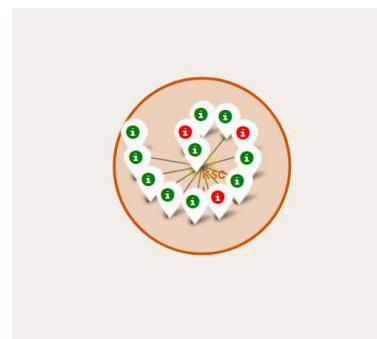
SpaceX: All launch sites



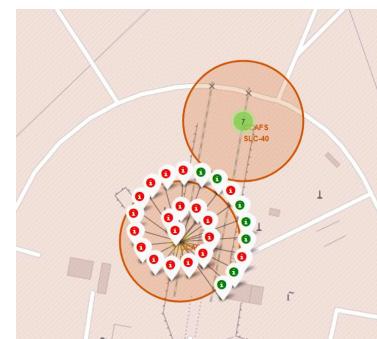
Falcon 9 Success/Failed launches for each site



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



Kennedy Space Center (FL)
KSC LC 39A



Cape Canaveral (FL)
CCAFS-LC40

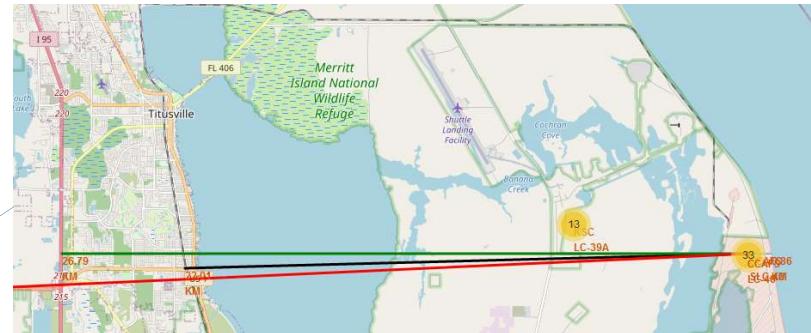
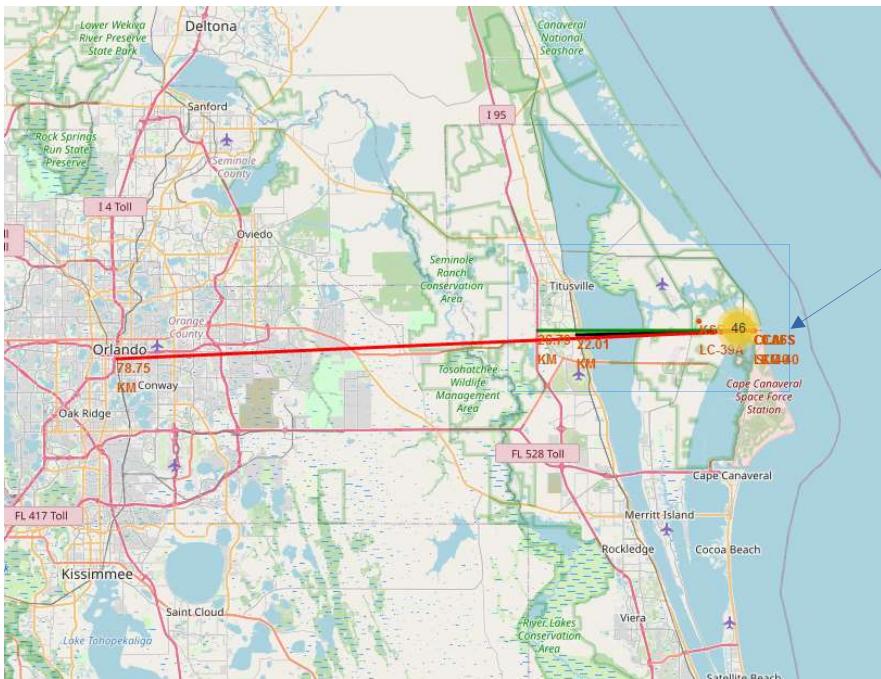


Cape Canaveral (FL)
CCAFS-SLC40

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

Launch sites are close to coasts. For safety issues if launcher is lost in the early stage of the flight.

Rockets are launched:

- From West to East over the ocean in Florida.
- North or South bound over the ocean in California. (Polar orbits only)

Launch sites are relatively far from populated areas for protecting population from serious incidents at lift off: explosion on the launch pad.

ESA Unique Launch site – Kourou, French Guiana.

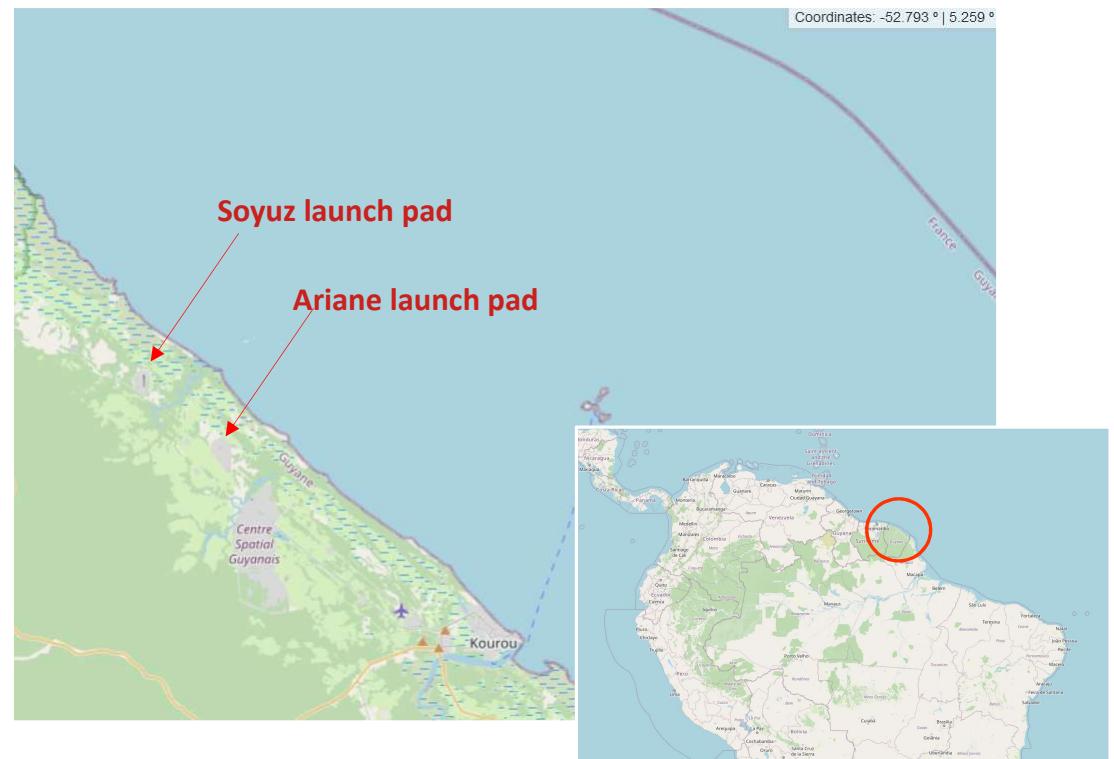
European Space Agency (ESA)/ArianeEspace launch sites in Kourou, French Guiana (France).

Kourou launch sites: Ariane, Soyuz

- are very close to the Equator: it's a strong advantage for GTO/GEO flights
- are in a remote area far from any high density inhabited area and high traffic infrastructures
- Are close to the coast

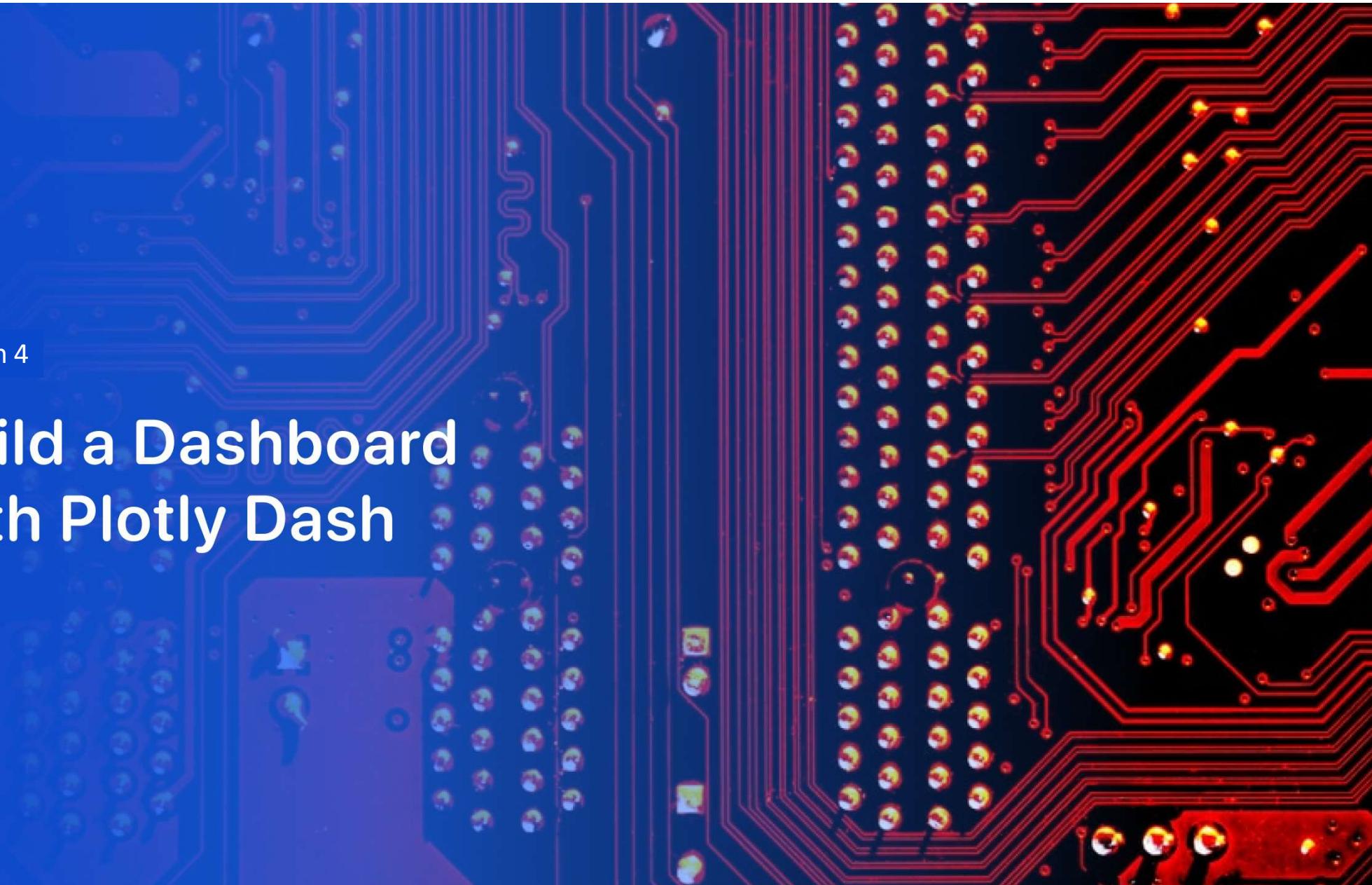
Strong Advantage over SpaceX, in terms of safety and GTO/GEO flights and energy required at lift-off.

SpaceX could compete and reduce energy at lift-off for GTO-GEO flights by introducing a concept like "Sea Launch".

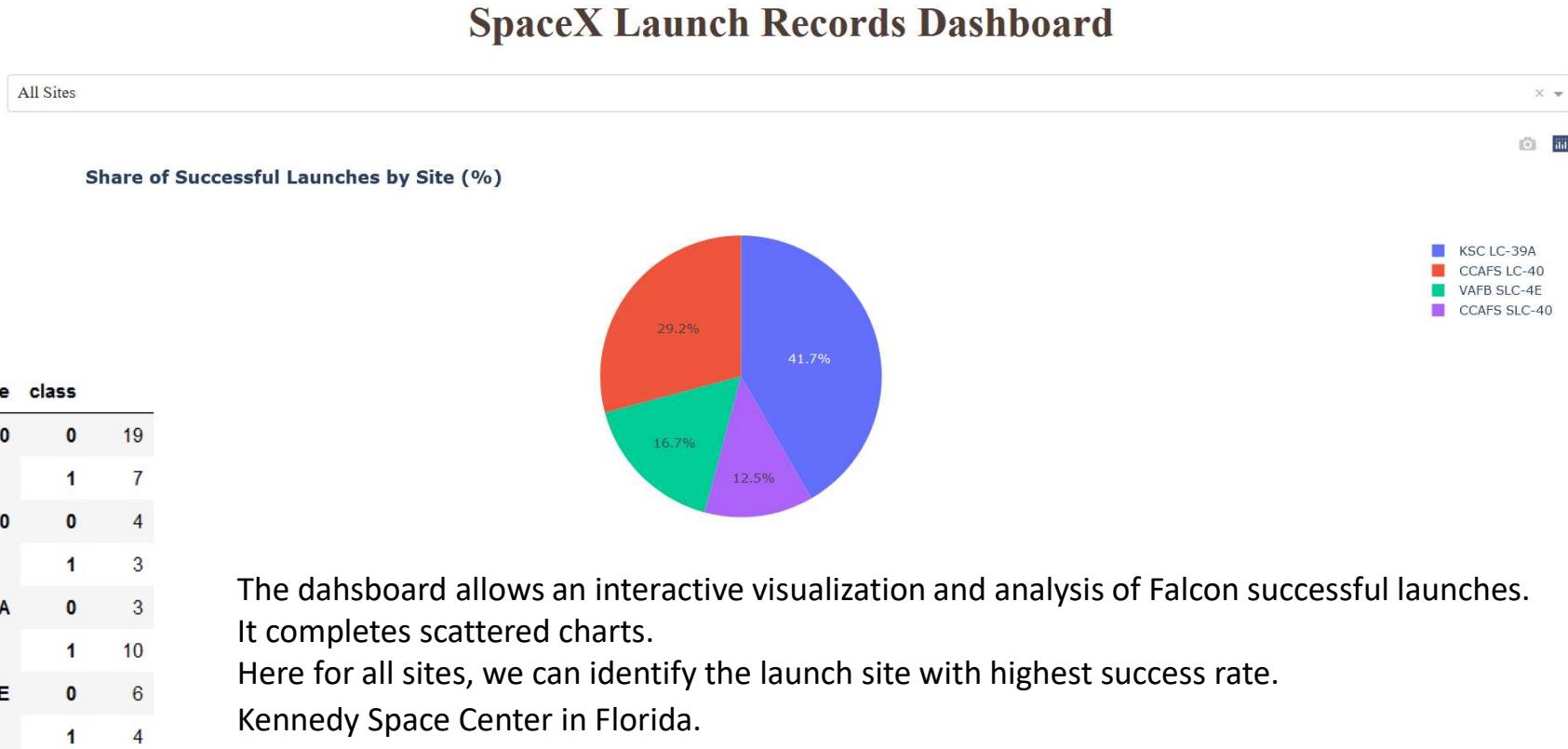


Section 4

Build a Dashboard with Plotly Dash



SpaceX Falcon 9: Launch success count for all sites

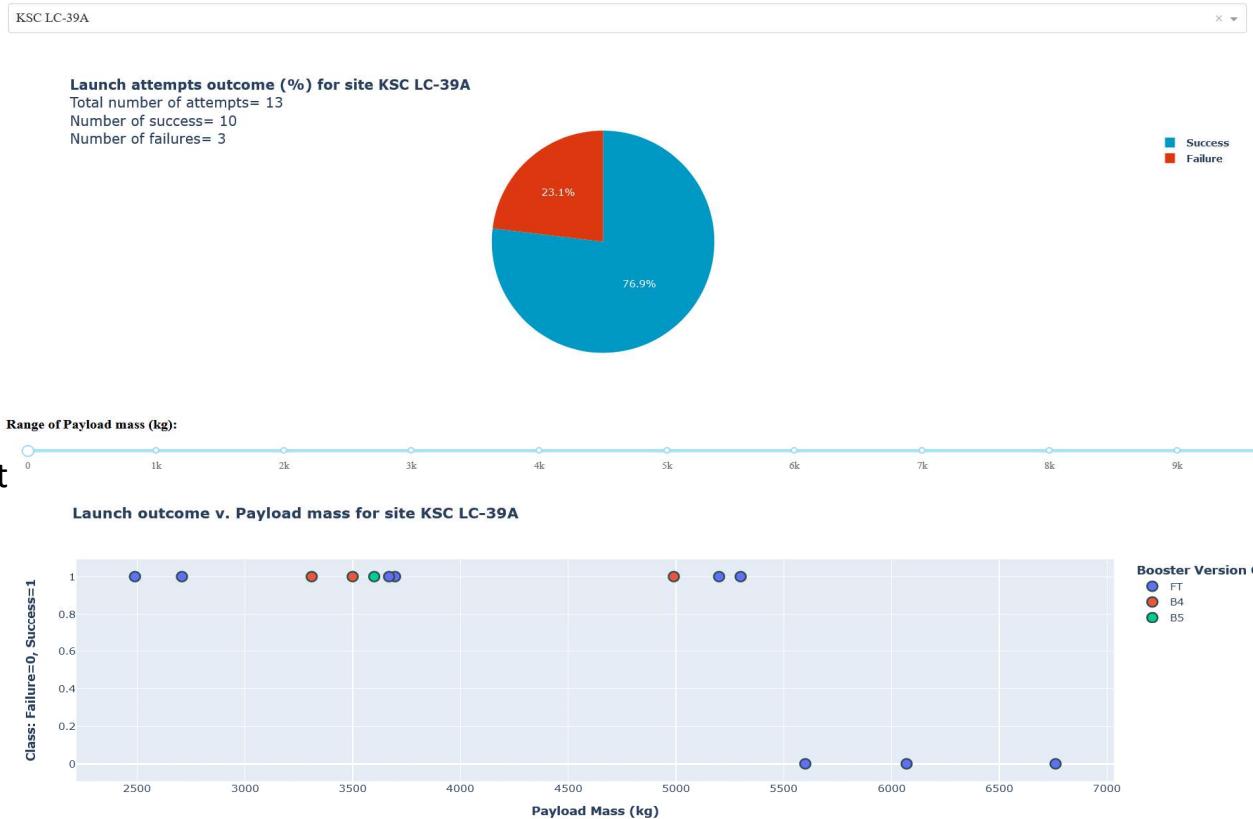


SpaceX Falcon9 Launch site with highest launch success ratio

KSC LC-39A

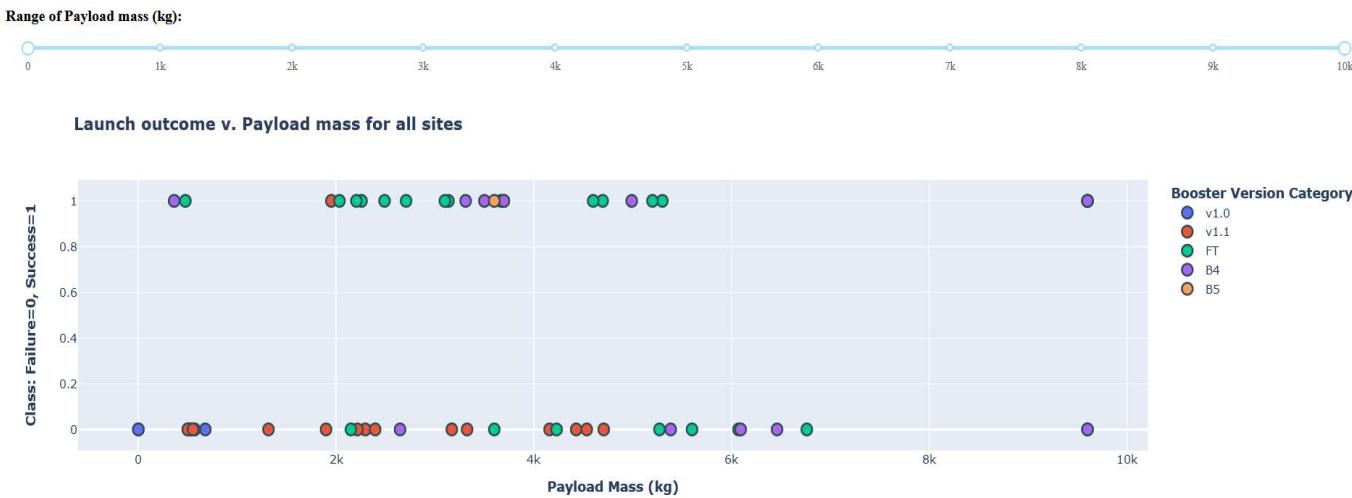
Kennedy Space Center in Florida.
13 flights, 10 successful missions.

- Heavy payload are “high risk”
- Success does not seem to depend upon boosters versions with low mass payload <5500kg.
- B5 and FT are the most reused launchers. Data is not sufficient, but may indicates that they are as reliable as 1 time launchers.



Launch outcome v. Payload mass (all sites)

- V1.0 and v1.1 are early launchers with low reliability.
Landing legs, were pioneered on the Falcon 9 v1.1 version, but that version never landed intact.
They were phased out in 2015.
- FT: “Full Thrust” is the next generation and has the highest success rate for payload mass under 6 tons. Including with “drone landing” (see details in next slide).
- Many FT flights are done with reused launchers. And show good reliability.
- Heavy payload are “high risk”.



Successful Drone Ship Landing with Payload mass between 4000 and 6000 kg

List the names of Falcon 9 boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000 kg

Recent Full Thrust (FT) boosters exhibit the highest success rate on drone ship landing. Including with GTO flights.

It bodes well if SpaceX introduces a concept like “Sea Launch” for GTO launches at sea, close to the equator.

```
# sql query
q_boost_succ="""
select Booster_Version from spacex_v11 where Landing_Outcome = 'Success (drone ship)'
and PAYLOAD_MASS_KG_ > 4000
and PAYLOAD_MASS_KG_ < 6000 """
```



```
Booster_success_landing=pd.read_sql_query(q_boost_succ,conn)
print(Booster_success_landing)
```

	Booster_Version
0	F9 FT B1022
1	F9 FT B1026
2	F9 FT B1021.2
3	F9 FT B1031.2

The background of the slide features a dynamic, abstract design. It consists of several curved, glowing lines in shades of blue and yellow, creating a sense of motion and depth. The lines are thicker in the center and taper off towards the edges, with some lines curving upwards and others downwards. The overall effect is reminiscent of a tunnel or a futuristic landscape.

Section 5

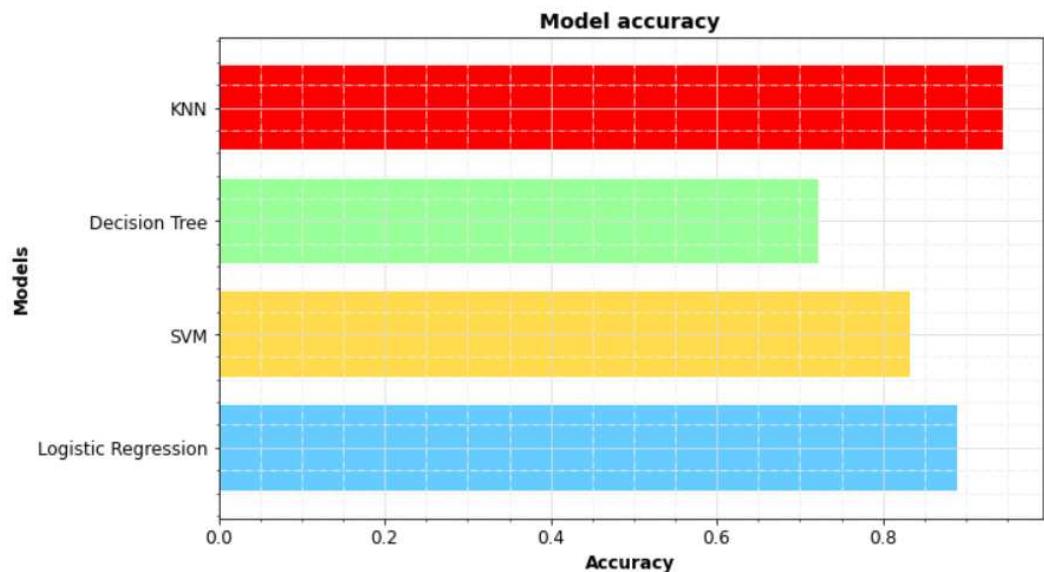
Predictive Analysis (Classification)

Classification Accuracy

Classification Accuracy with test set.

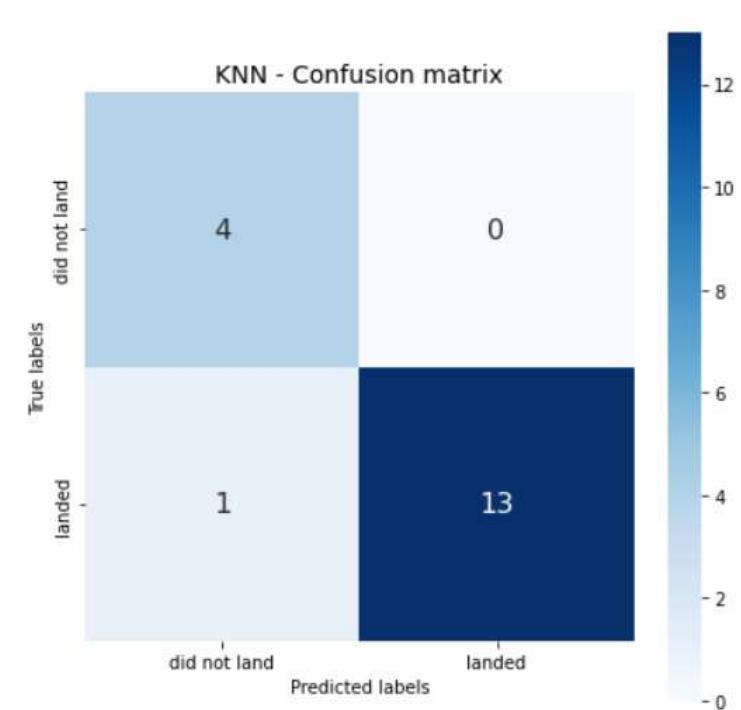
Results with “train test split” random_state=3

- 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)



Thank you!

