IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Anderson PAIVA

Sep 19nd, 2023

Important Preliminary remarks for reviewers.
Stéphane Dedieu, June 21st, 2022

I have no IBM cloud account anymore.

- **Interactive dashboard:** I could not find any good solution for sharing the interactive Plotly Dashboard: MyBinder doesn't work. Google Colab is good but…

  -if the interactive dashboard works very well with Google Colab (click on Colab icon in GitHub PlotlyDash notebook), exploring the Dashboard **requires a GoogleDrive account** for running the code (option "run after"). Insist on cell [9] there maybe an issue with "import dash". Run twice, it works.

  -You can import the notebook and run it on your computer if you installed Anaconda or any similar tool. In the application, you may need to replace: *app = JupyterDash(__name__); JupyterDash.infer_jupyter_proxy_config()* (for Google Colab) with initial: *app = dash.Dash(__name__)*

- **Interactive launch site maps** notebook with Folium works very well with ***nbviewer***

- Other notebooks in GitHub properly display dataframes and results.

- **SQL queries:** I built my own MySQL server, imported the .csv file in a Table, established a connection with my MySQL server, and used pandas sql queries in the form: *pd.read_sql_query*… because this was the solution that worked best.

Regarding the report, SQL queries were disruptive and I moved most of them to Appendix. Some SQL queries that were useful to the report are contextualized in the Results sections.

- **Machine learning prediction:** with default train test split, results were basically all similar. And boring. I shuffled the data set using the option *random_state*. With random_state=3, results are more diverse. And they are reproducible, except Decision Tree.

  X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=3)

Outline

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

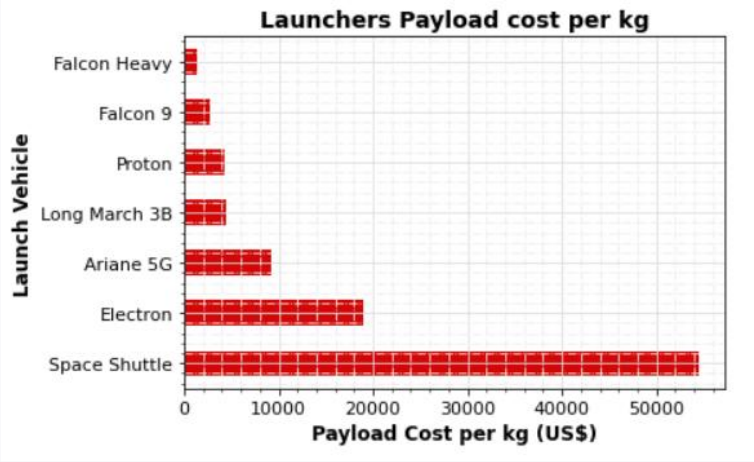
Executive Summary

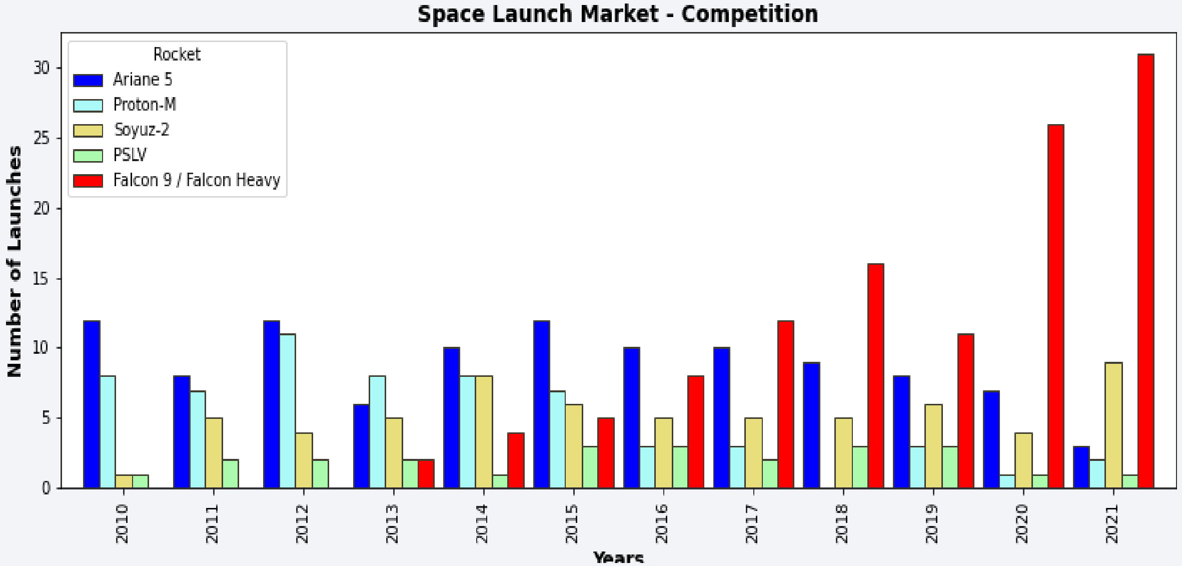
In a highly competitive business environment:

- SpaceX has revolutionized the launch of satellites with a concept of reusable launcher/booster: Falcon9/Falcon 9 Heavy
- The main advantage of this concept is the significant reduction in cost per kg.
- Reliability problems remain compared to classic launch vehicles like Soyuz or Ariane-5.
- For maintaining the “low cost” competitive advantage, compared with classic launchers, Falcon9 mission success is defined as the successful recovery or landing of the booster.
- Falcon9 booster successful recovery depends on features such as:
    - orbit
    - payload mass
    - booster versions
    - Launching sites…
- Based on these features, the best Machine Learning supervised classification model developed in this report, predicted booster recovery outcome with an accuracy close to 94%.

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

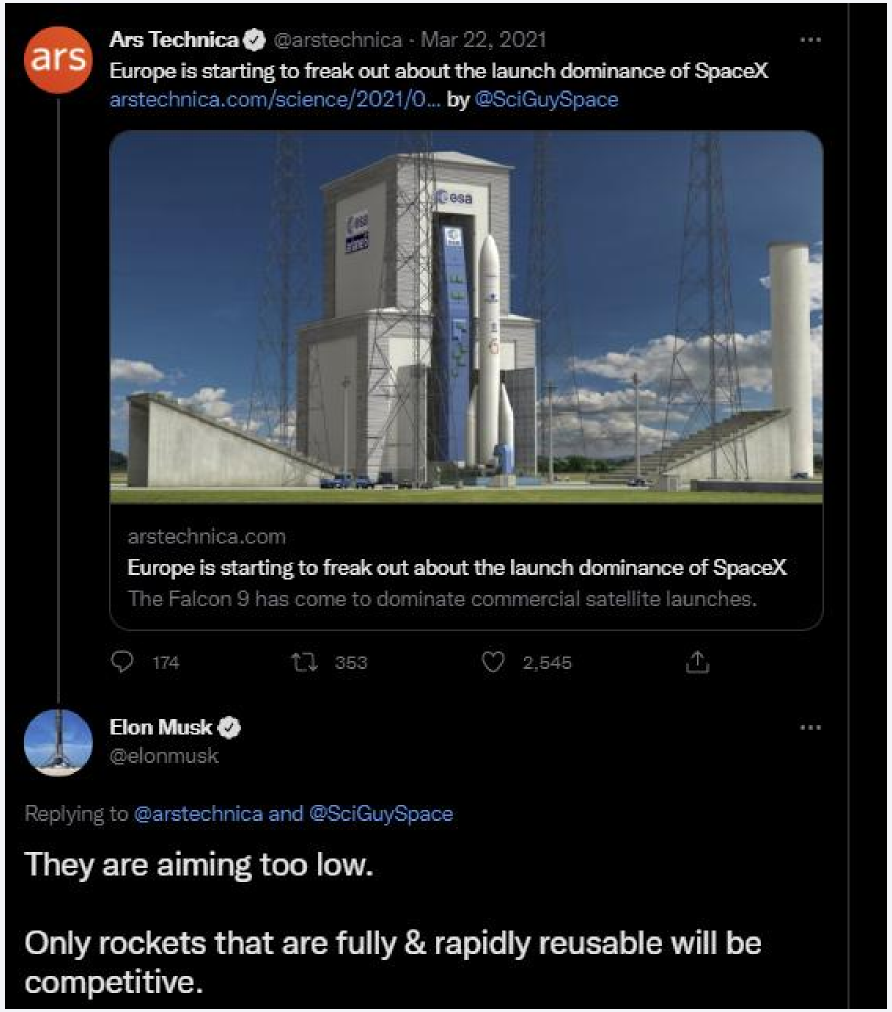
Introduction - Business understanding

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.**

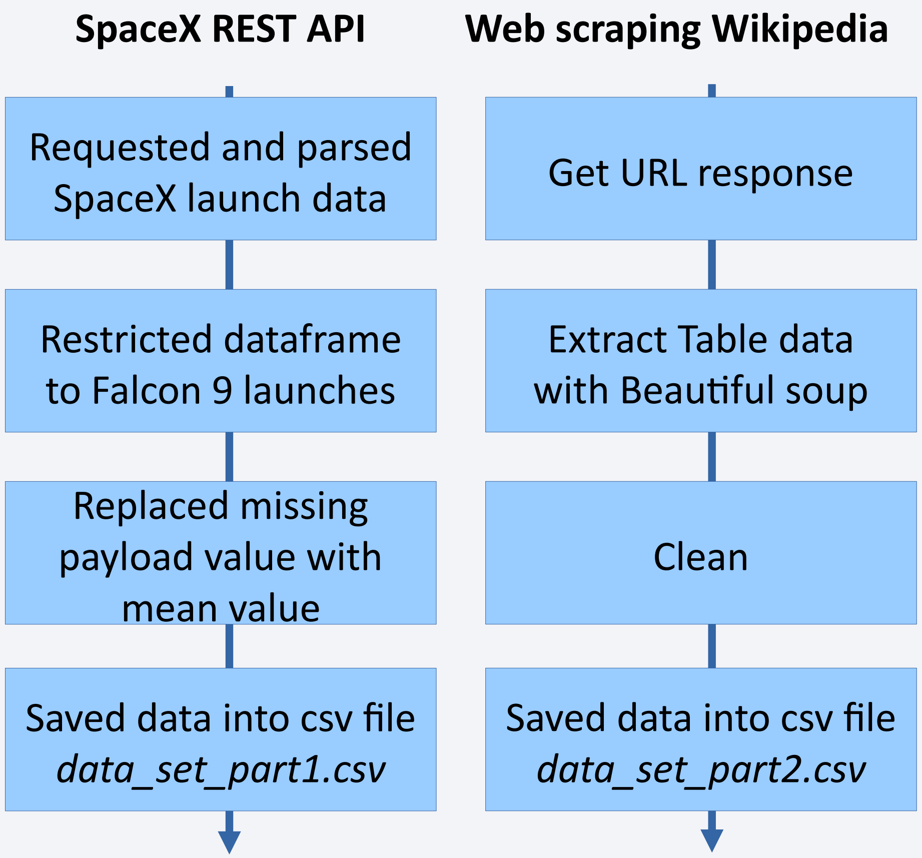
Data Collection

Data was collected from:

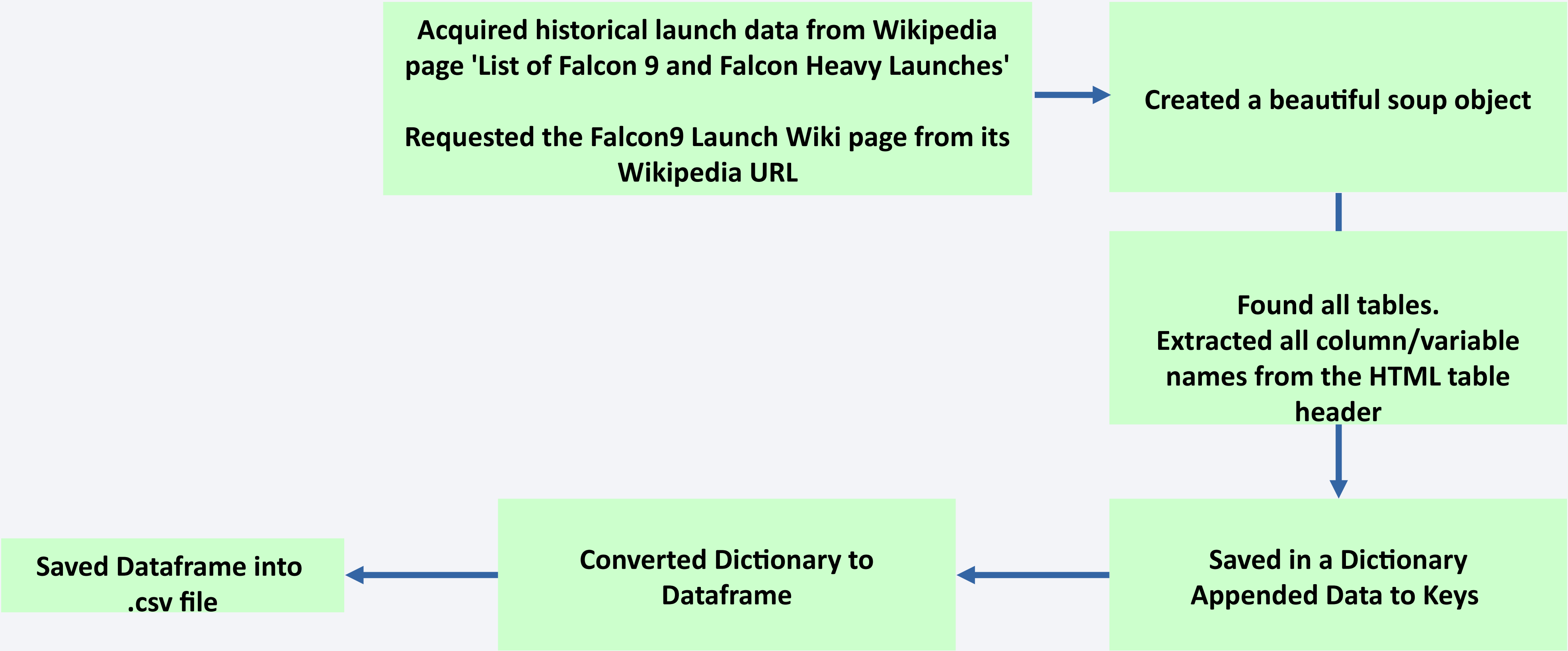
-open source SpaceX REST API

-webscraping Falcon9 launch data in Wikipedia

-webscraping Ariane5 launcher data in Wikipedia ***for reference graphs only.*** (not prepared for ML)



Data Collection - Scraping



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)

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

Requested and parsed
SpaceX data with GET request

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

Saved Dataframe into .csv file

Data Wrangling

Dataframe
From SpaceX API

Identified missing values

Replaced missing “PayloadMass” value with mean value

SpaceX dataset
Further Data wrangling

Identified:
- null values for each feature
- numerical and categorical features

Calculated:
- 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 1st 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

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Class
49	50	2018-05-11	Falcon 9	3750.00	GTO	KSC LC 39A	True ASDS	1
47	48	2018-04-02	Falcon 9	2760.00	ISS	CCAFS SLC 40	None None	0
50	51	2018-06-04	Falcon 9	5383.85	GTO	CCAFS SLC 40	None None	0
44	45	2018-01-31	Falcon 9	4230.00	GTO	CCAFS SLC 40	True Ocean	1
11	12	2015-01-10	Falcon 9	2395.00	ISS	CCAFS SLC 40	False ASDS	0

EDA with SQL

For the reviewer, for emulating SQL queries:

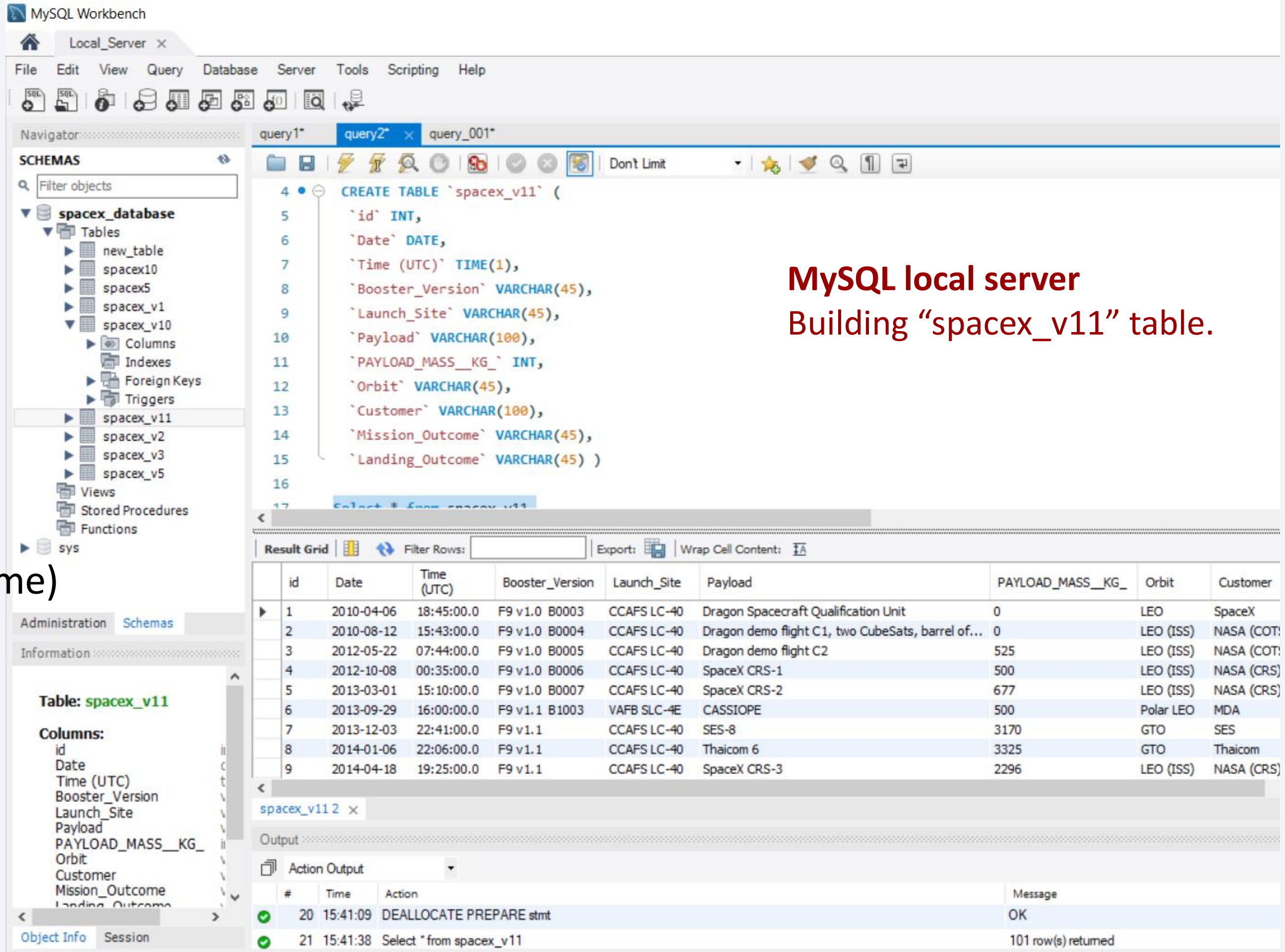
- **I created a MySQL local server**
- **I created a database: 'spacex_database'**
- **I saved SpaceX csv file in Table “spacex_v11” (see picture)**
- **I established a connection “conn” wih “spacex_database”.**

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

We ran SQL queries with “pd.read_sql_queries”:

```
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



MySQL local server
Building “spacex_v11” table.

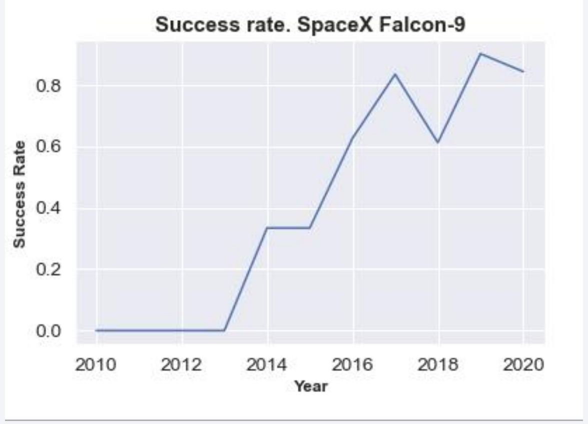
EDA with Data Visualization

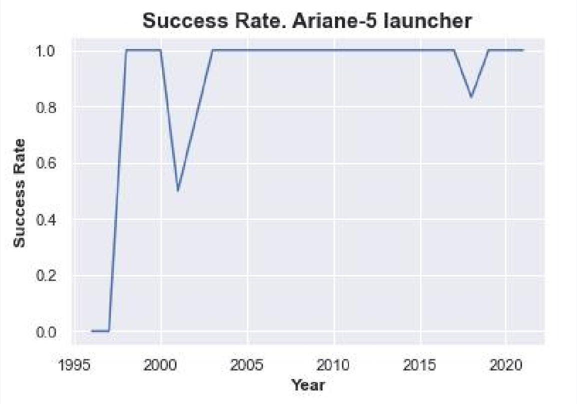
Exploring data in the Falcon 9 dataframe searching for factors and relations influencing launching success rate (booster recovery).

- 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

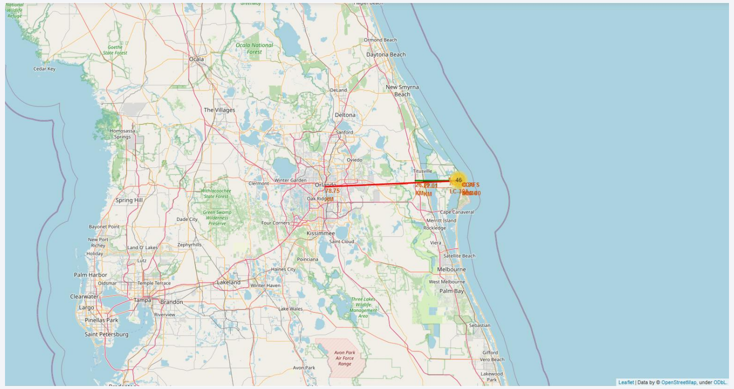
Launch success rate may depend on the location and proximity of a launch site. **Folium Interactive Map was used for visualizing and analyzing SpaceX Launch Sites.**

- 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

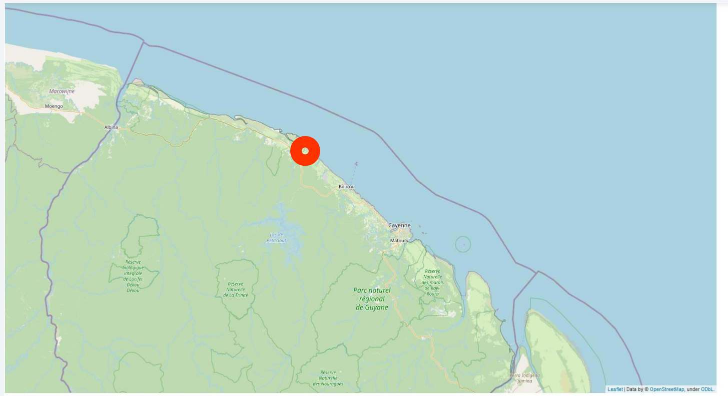
Calculated the distance between a launch site (CCAFS_SLC40 in Cape Canaveral, FL) and:

- 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)**

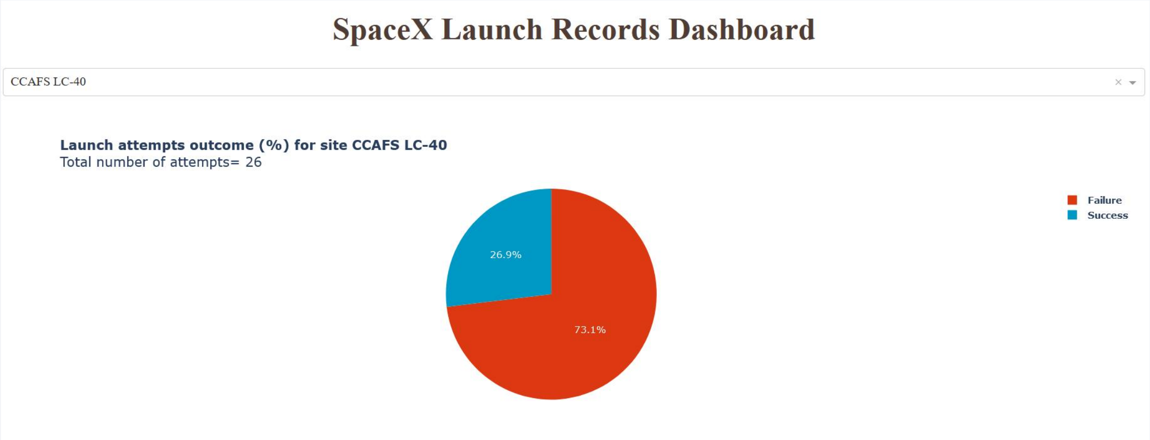
Build a Dashboard with Plotly Dash

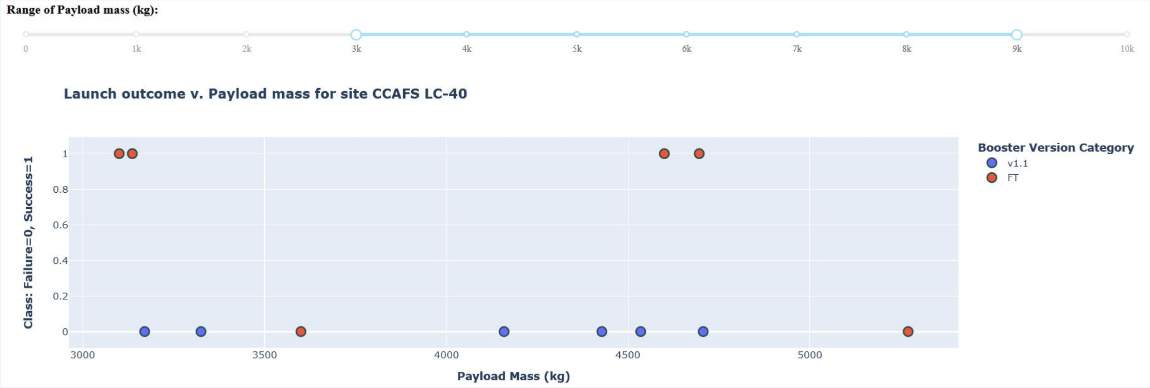
We built an interactive dashboard with Plotly including:

- 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)

Imported Falcon9 launching data from .csv to Pandas dataframe

Supervised Classification models
Launcher landing Labels

- Class 0: fail
- Class 1: success

Data normalization

Split Data: Training, Testing Sets
Shuffling option !

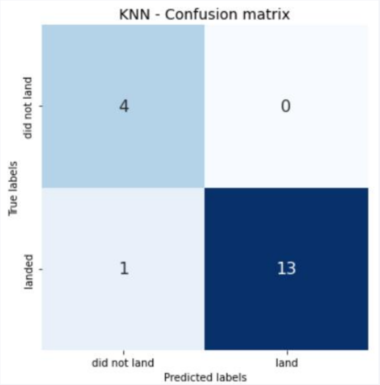
Model Training

Logistic Regression

Support Vector Machine (SVM)

Decision Tree Classifier

K Nearest Neighbors Classifier



Testing set

Predictive analysis
Accuracy
Confusion matrix

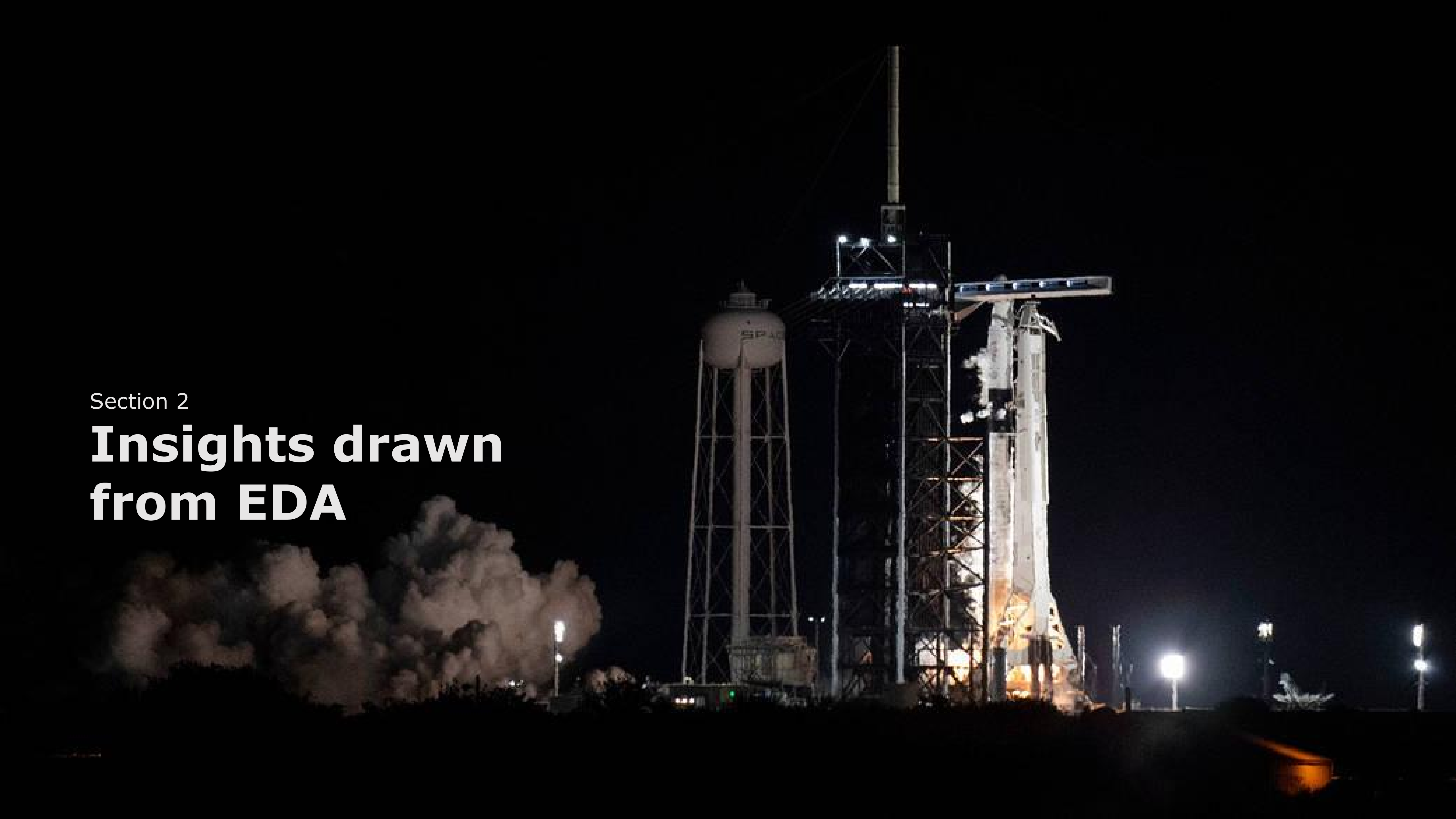
Selection of best model

Note to the reviewer:
Data was shuffled in the split data process. For making results a bit more interesting
"train_test_split(X, Y, test_size=0.2, random_state=3)"

Optimization of models hyper-parameters with Scikit-learn GridSearchCV
Refined Optimization when possible (LR, SVM) for achieving highest accuracy with training set.

Results

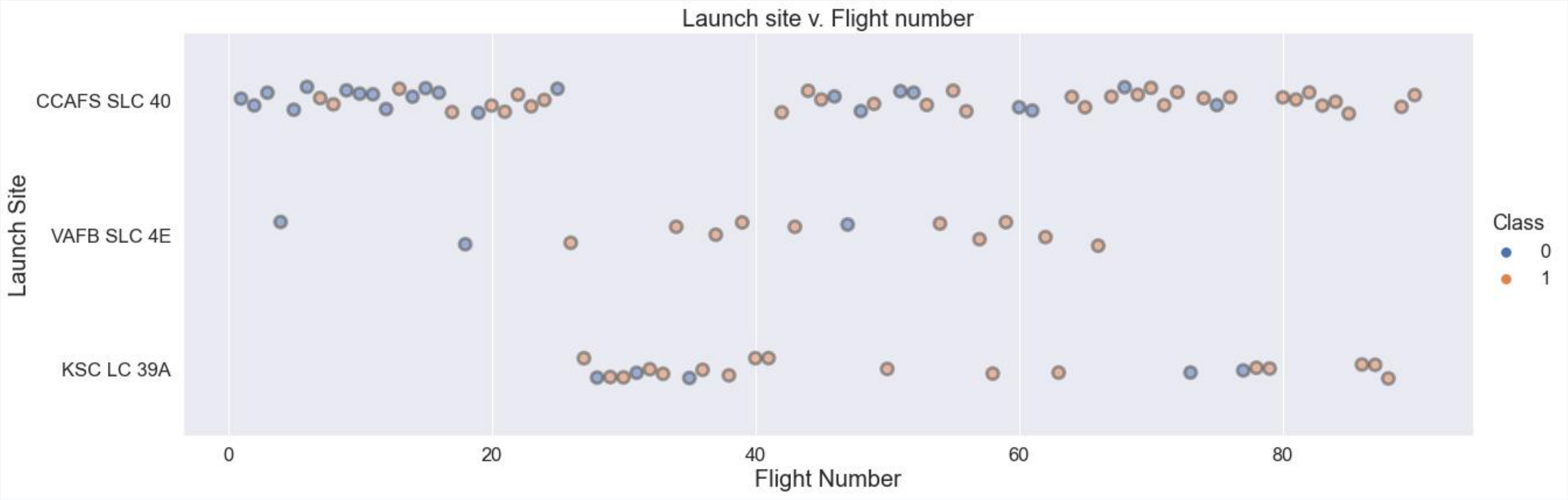
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results



Section 2

Insights drawn from EDA

Launch Site v. Flight Number

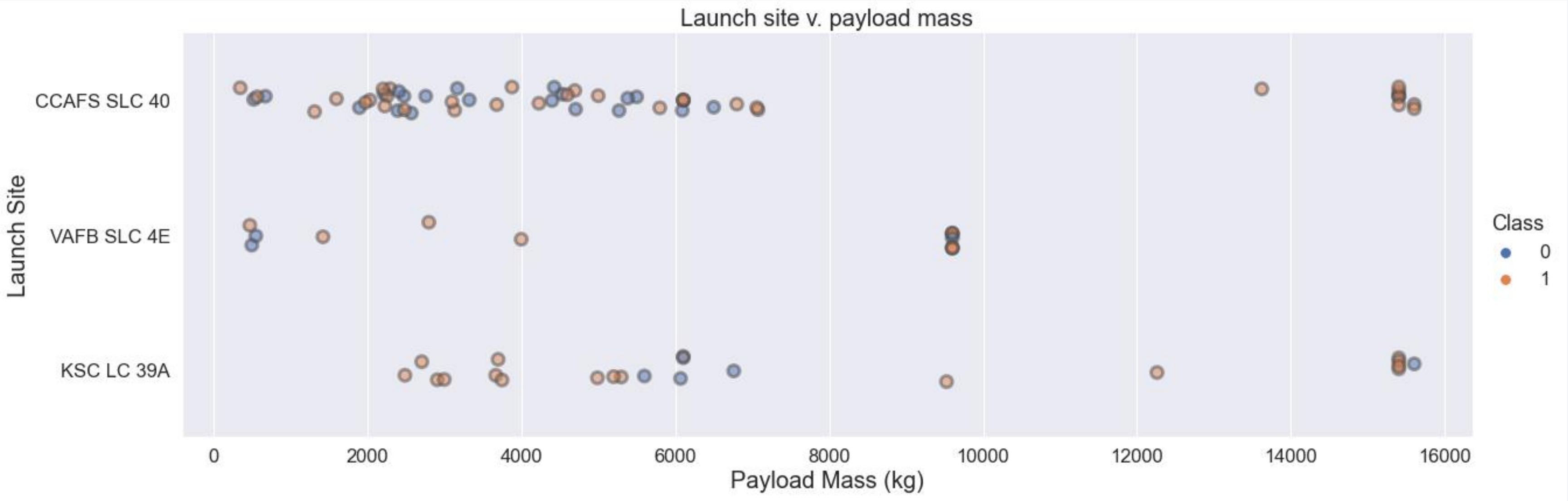


The chart displays valuable info about:

- 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

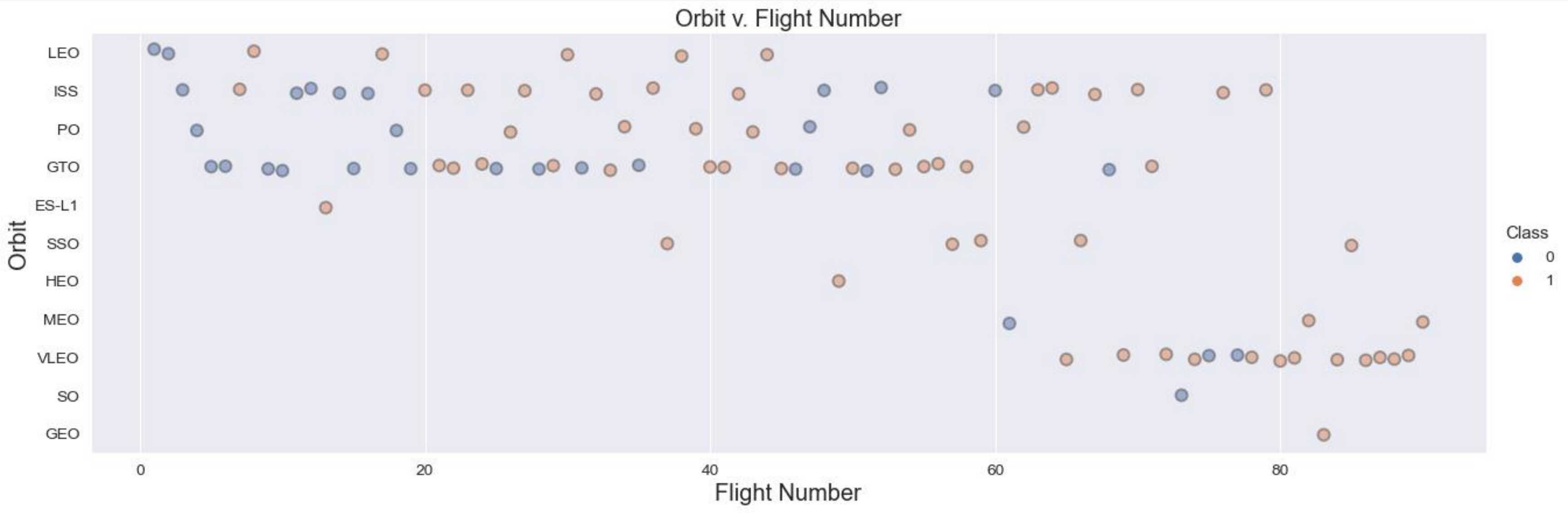


The chart brings additional info:

- 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:

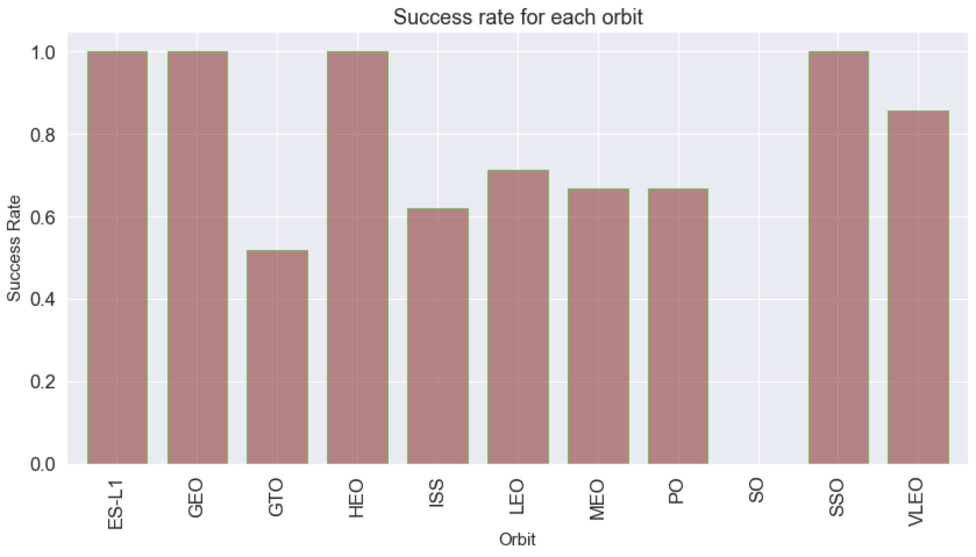
- GTO is a transfer orbit to GEO. Low thrust engines of the payload (satellite) complete the orbiting phase.
- We ignore results: GEO, SO, HEO, ESL-1, MEO. The number of flights is not significant.

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

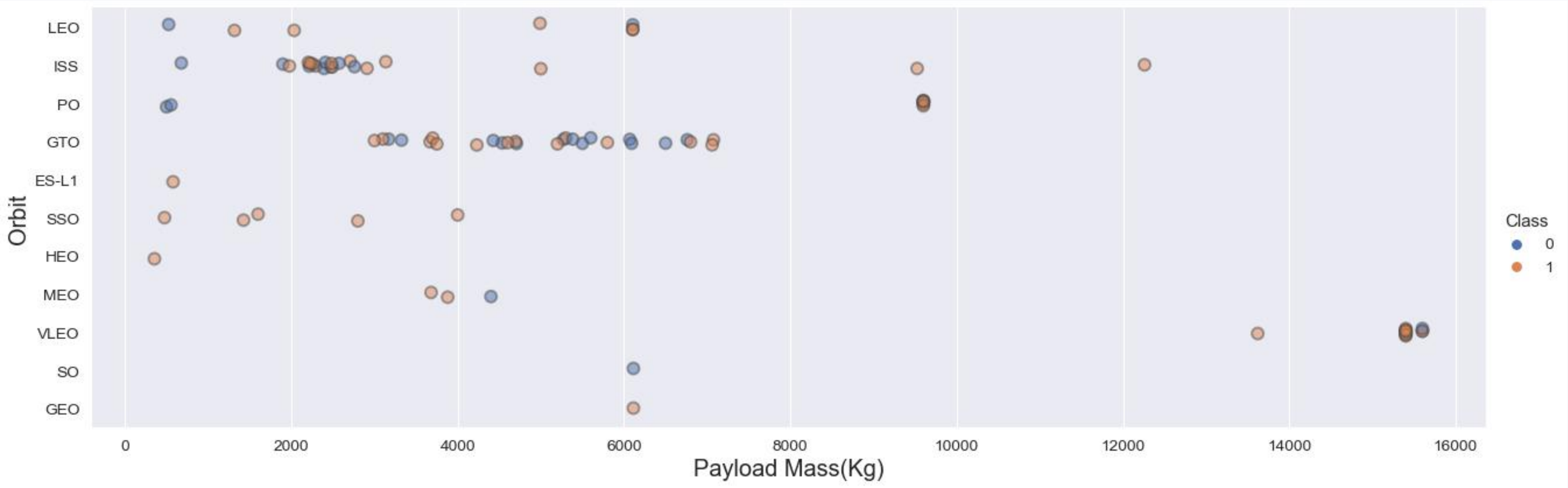
Success rate may strongly depend on both:

- 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.**

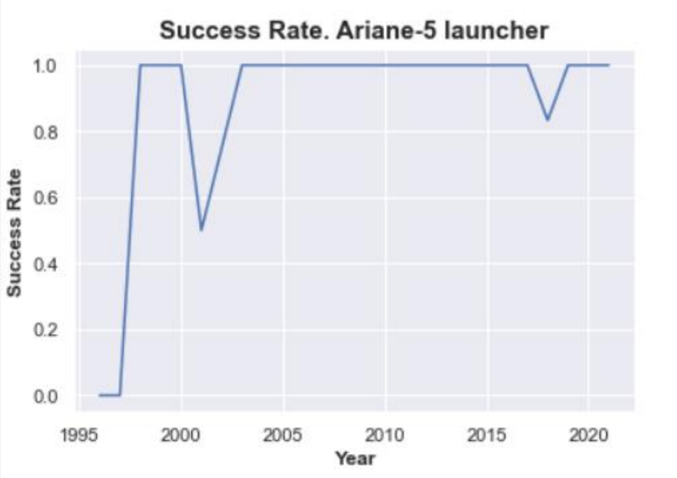
Therefore Falcon9 “success rate” in this report is defined after successful booster recovery (landing).

```
#  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.

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
```

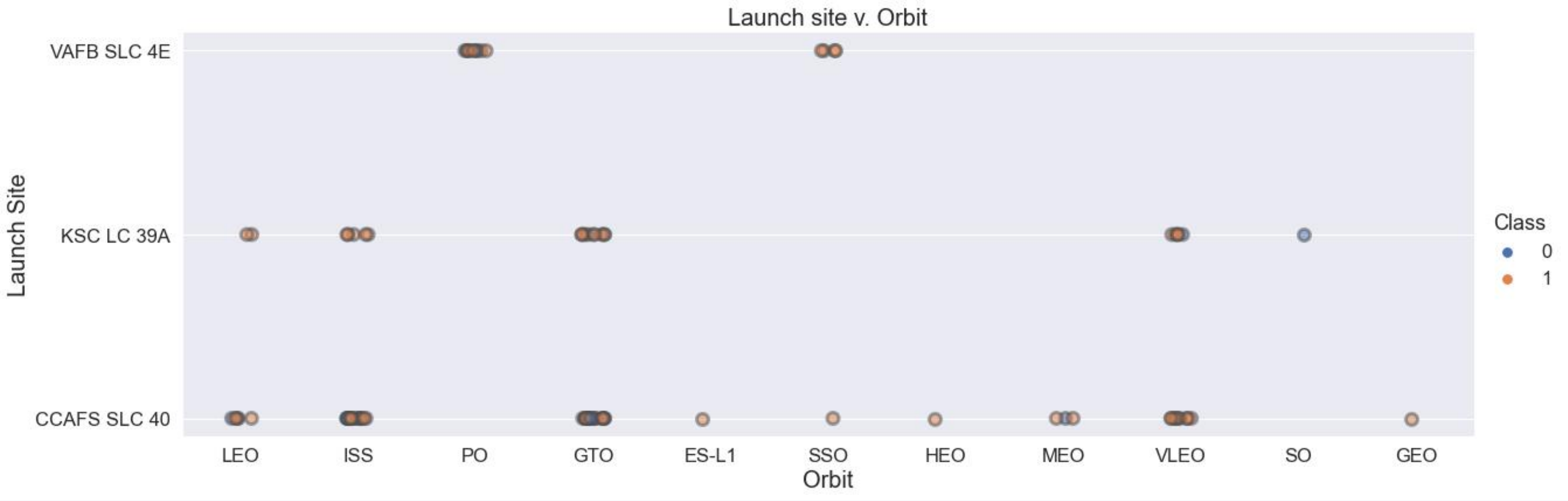
There are 4 distinct launch sites

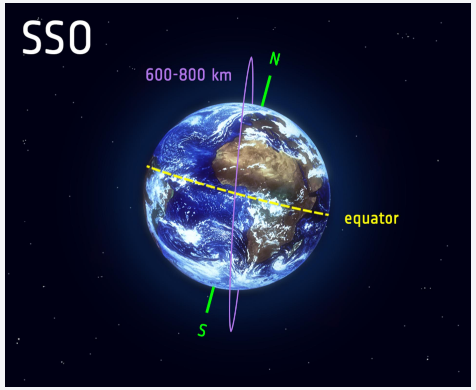
5 records where launch sites begin with `CCA`

```
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

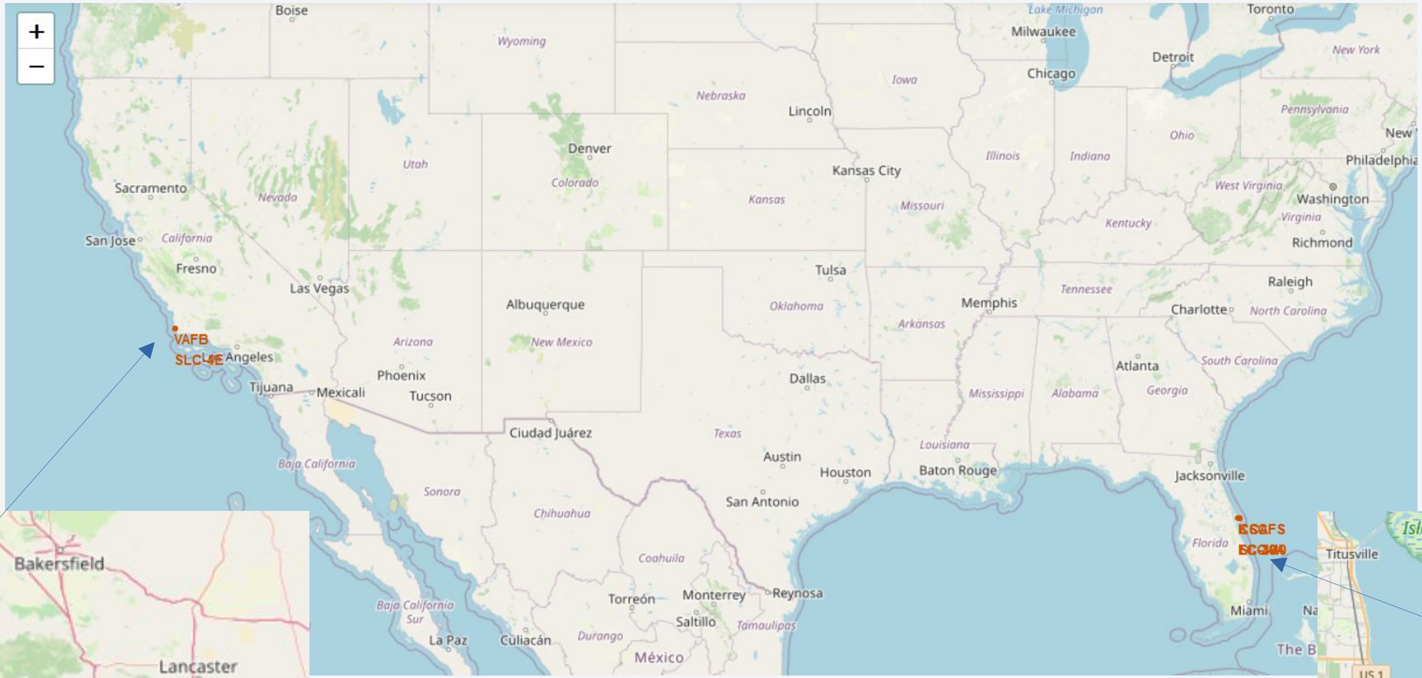


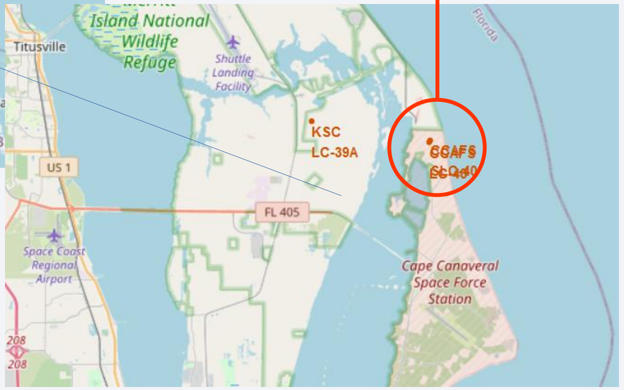


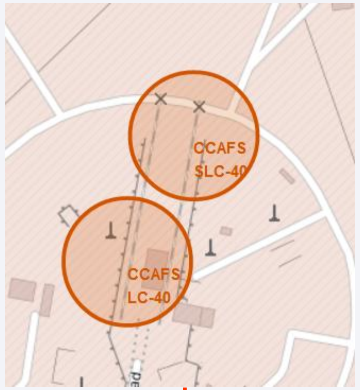
We added that scattered chart, which has implications in the analysis of launch sites.

- 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



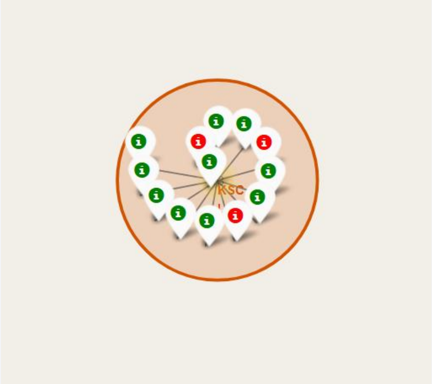




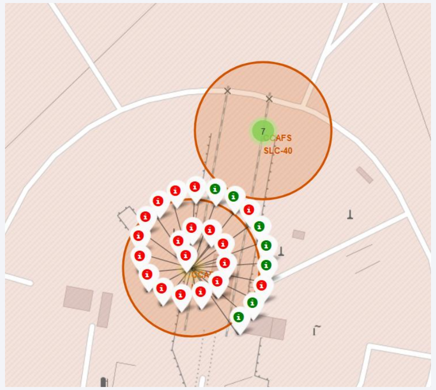
- **VAFB SLC-4E: Vandenberg Space Launch Complex 4 (CA)**
- **KSC-LC29A: Kennedy Space Center - Merritt Island (FL)**
- **CCAFS-LC40: Cape Canaveral Launch Complex 40 (FL)**
- **CCAF-SLC40: Cape Canaveral Space Launch Complex 40(FL)**

Falcon 9 Success/Failed launches for each site

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



Kennedy Space Center (FL)
KSC LC 39A



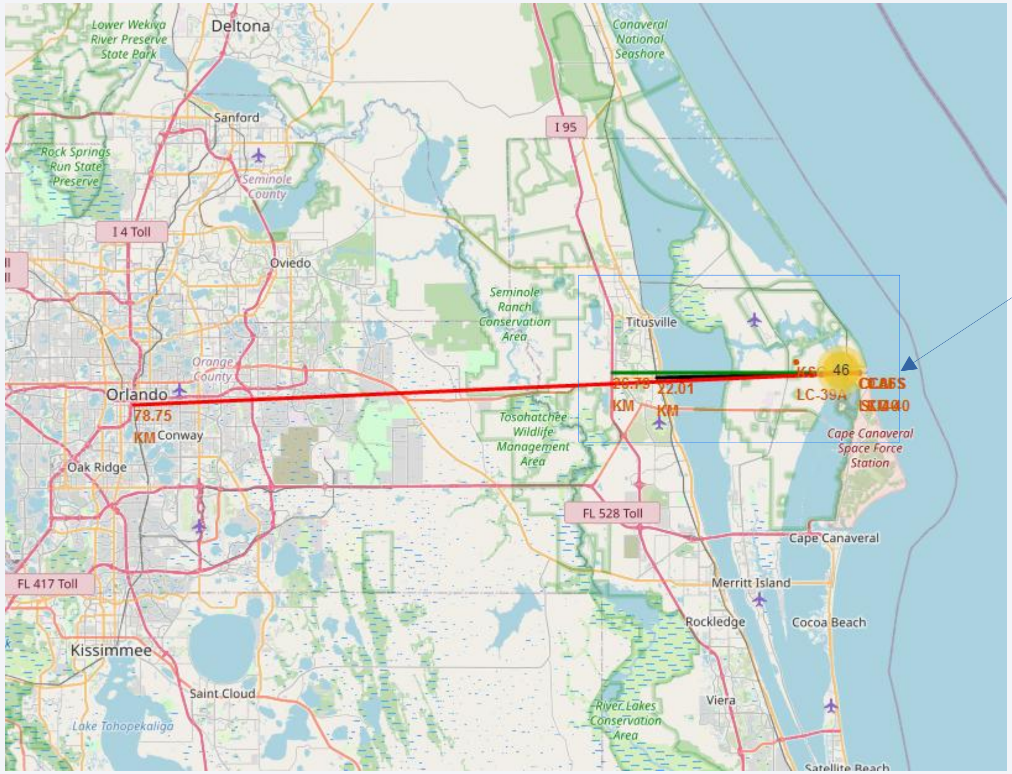
Cape Canaveral (FL)
CCAFS-LC40

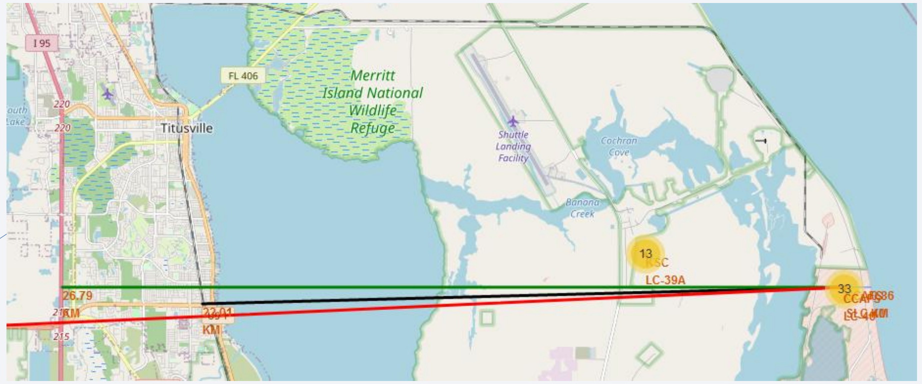
Cape Canaveral (FL)
CCAFS-SLC40

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.

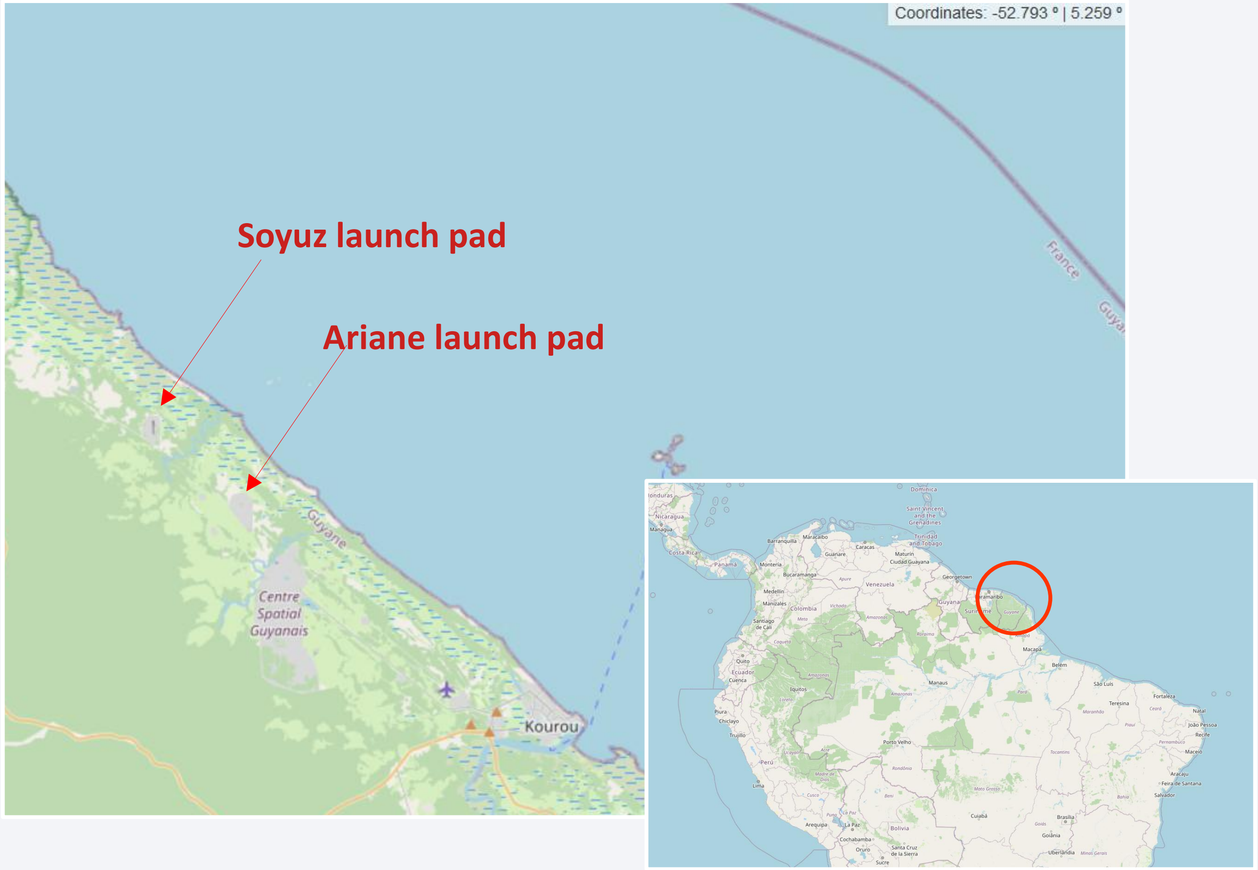
European Space Agency (ESA)/ArianeEspace launch sites in Kourou, French Guiana (France).

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.

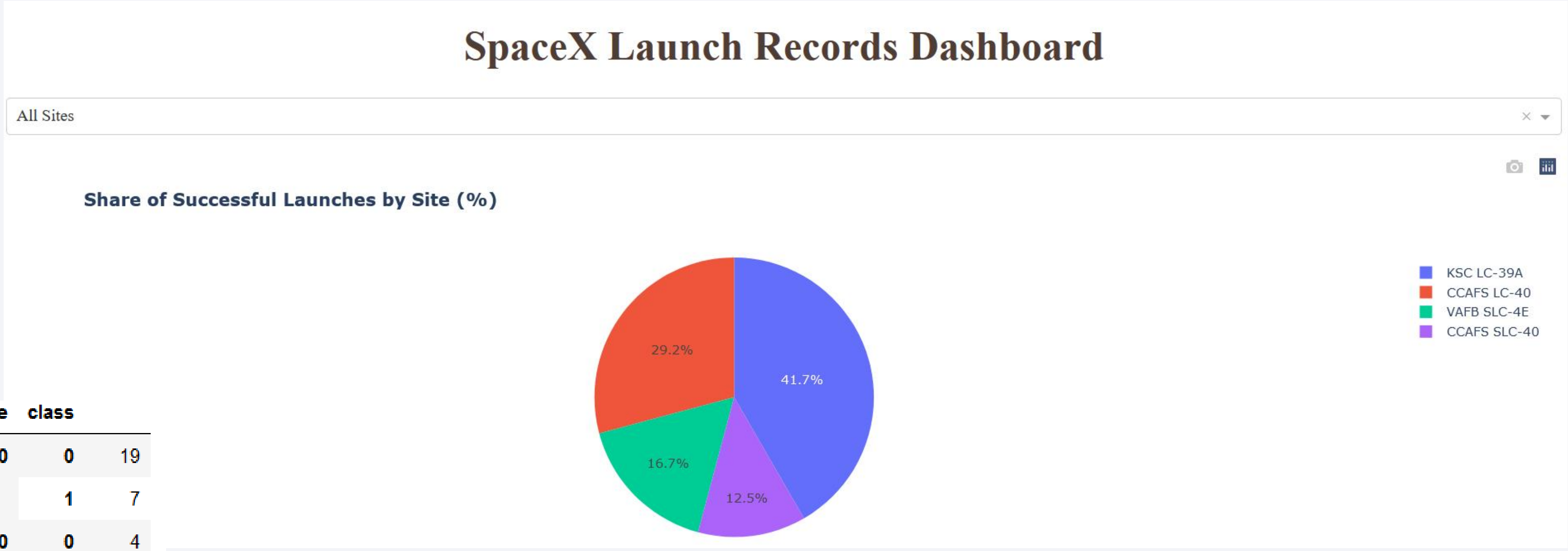
SpaceX could compete and reduce energy at lift-off for GTO-GEO flights by introducing a concept like "Sea Launch".



Section 3

Build a Dashboard With Plotly Dash

SpaceX Falcon 9: Launch success count for all sites



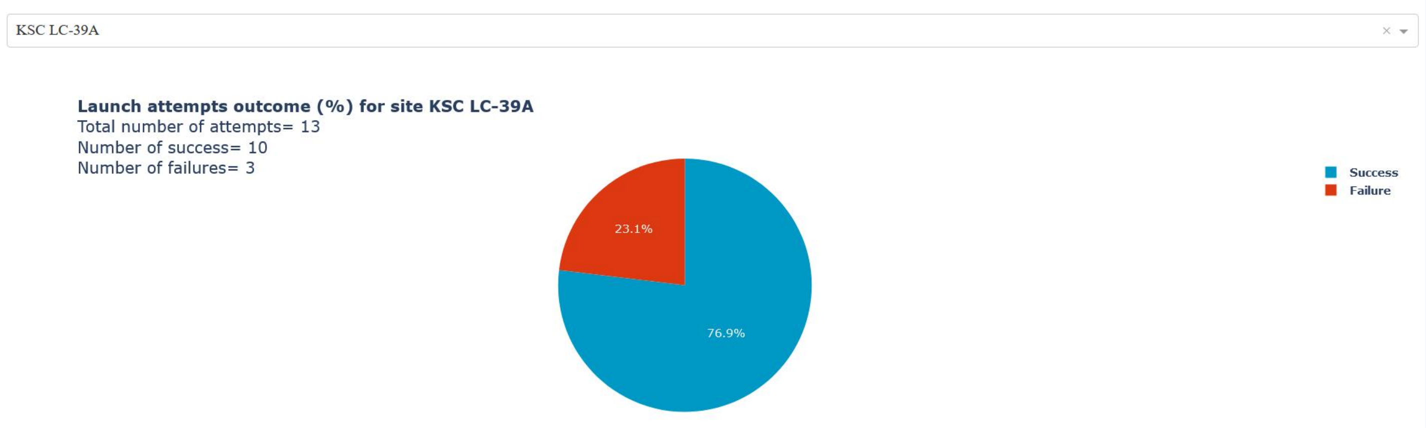
Launch Site	class	
CCAFS LC-40	0	19
	1	7
CCAFS SLC-40	0	4
	1	3
KSC LC-39A	0	3
	1	10
VAFB SLC-4E	0	6
	1	4

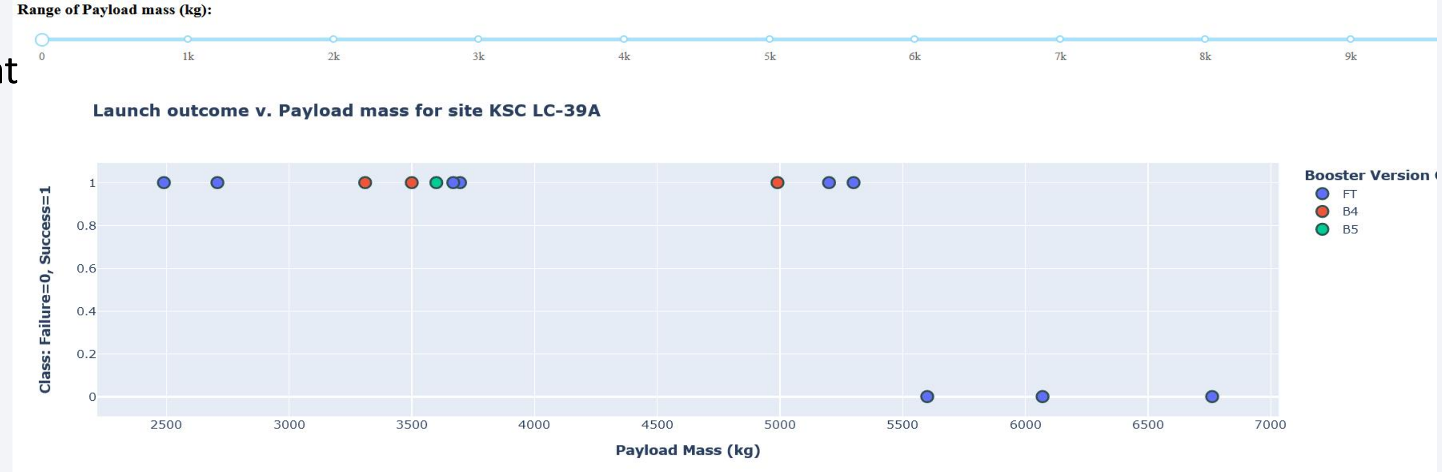
The dahsboard allows an interactive visualization and analysis of Falcon successful launches.
It completes scattered charts.
Here for all sites, we can identify the launch site with highest success rate.
Kennedy Space Center in Florida.

SpaceX Falcon9 Launch site with highest launch success ratio

KSC LC-39A
Kennedy Space Center in Florida.
13 flights, 10 successful missions.

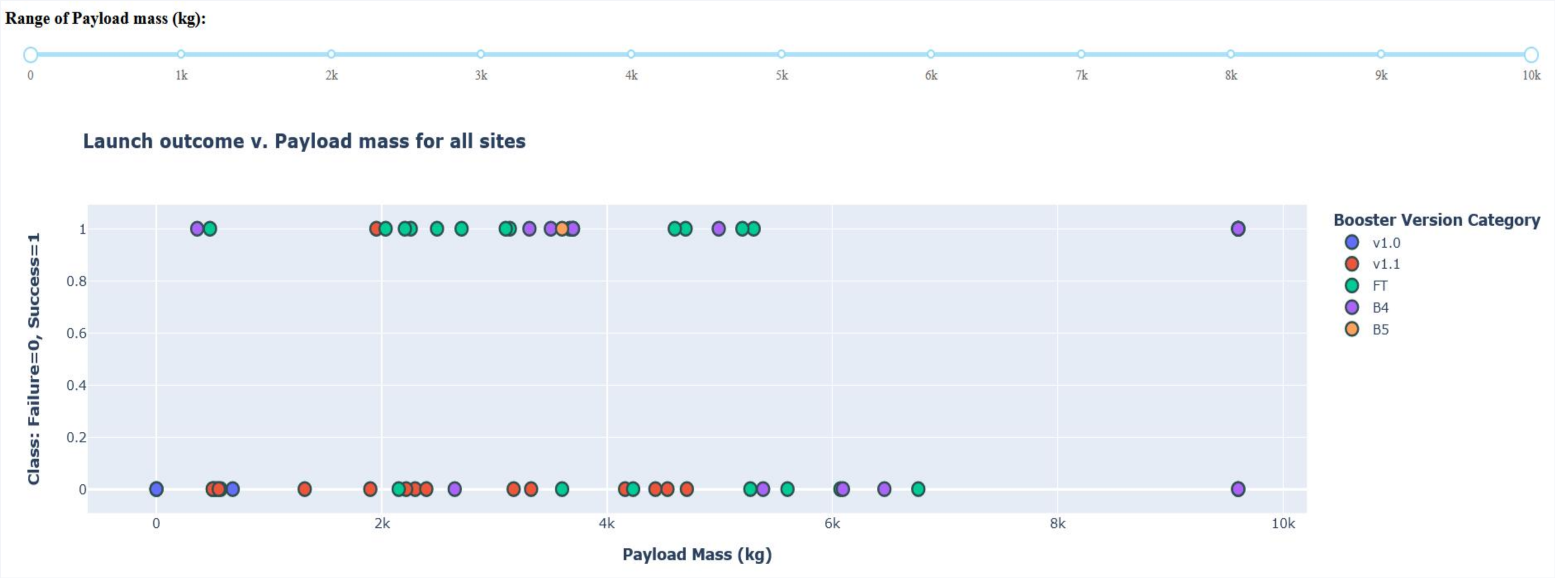
- Heavy payload are “high risk”
- Success does not seem to depend upon boosters versions with low mass payload <5500kg.
- B5 and FT are the most reused launchers. Data is not sufficient, but may indicates that they are as reliable as 1st time launchers.





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 """
```

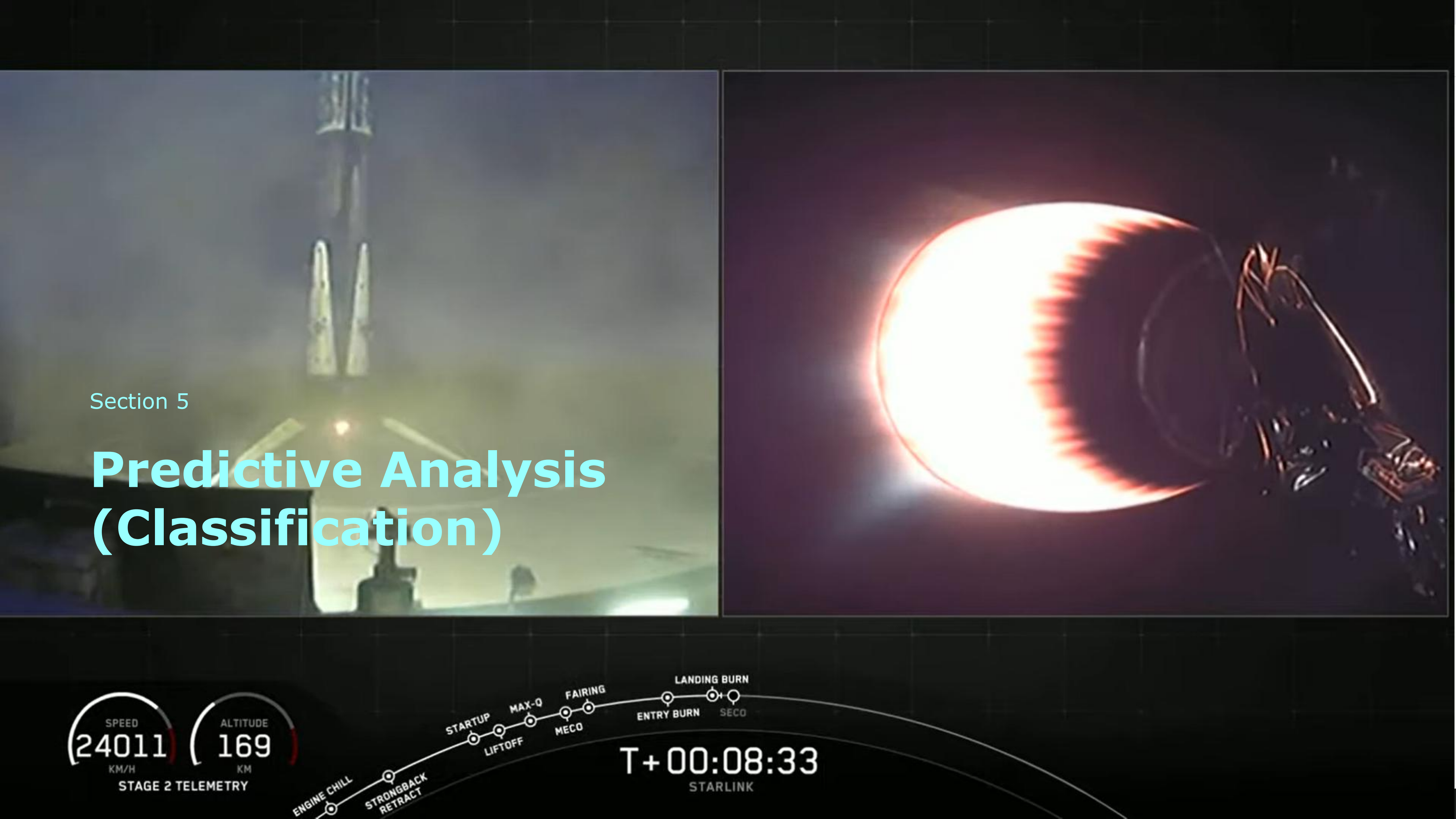
```
Booster_success_landing=pd.read_sql_query(q_boost_succ,conn)

print(Booster_success_landing)
```

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

Section 5

Predictive Analysis (Classification)



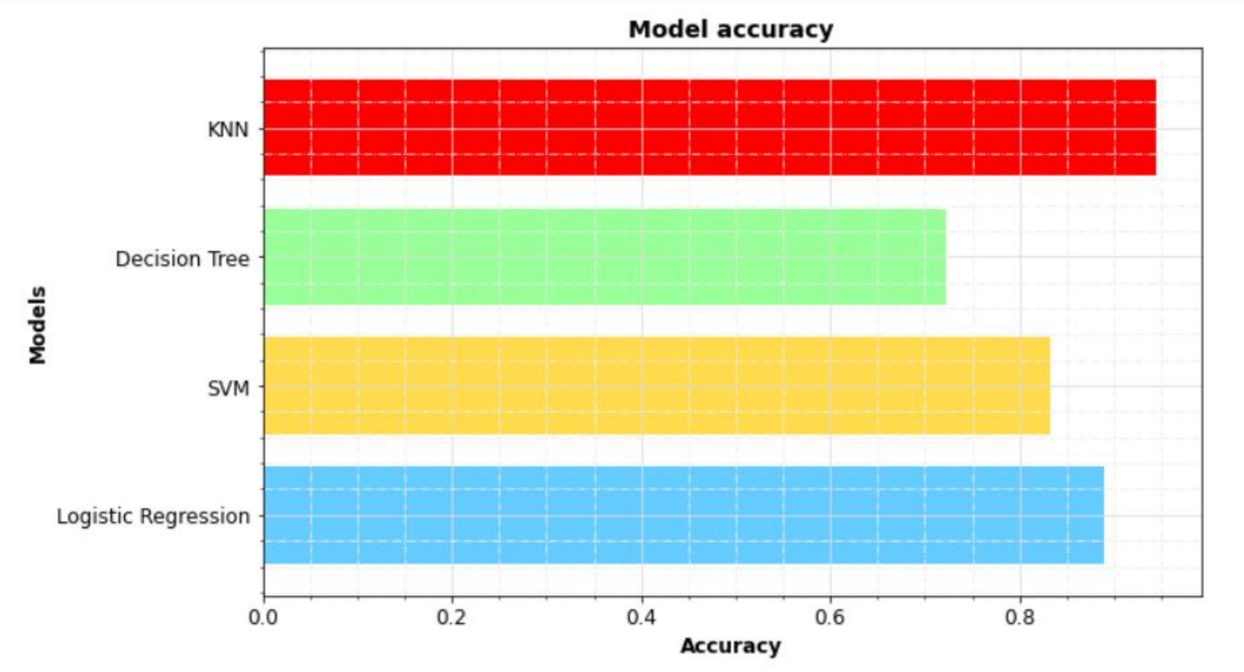
Classification Accuracy

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.

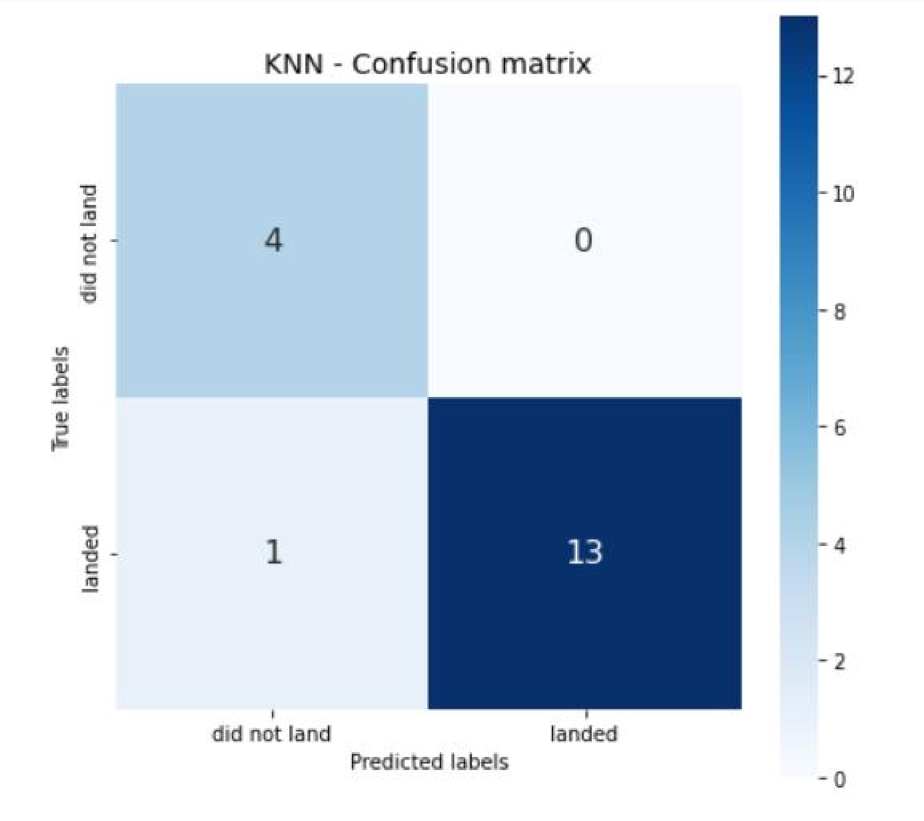
- Test set is too small.

- 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)



Conclusions

Conclusions

We defined success as a successful recovery process of Falcon9 booster.
From Falcon 9 and Wiki we imported data and developed tools for investigating factors driving successful booster recovery:
Payload, Orbit, Booster version, Launching sites…

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 compared with classic Ariane5 /Ariane 6.

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.

ESA/ArianeEspace, Soyuz have an advantage in terms of launch site, on the equator: less energy is required for GTO/GEO flights.

SpaceX could improve GEO/GTO flights success with a concept like “Sea Launch” close to the equator.
Recent successes with FT boosters landing on “drone ship”, bodes well for such a solution.

Even with recent recovery failure for GTO/GEO, SpaceX will maintain a sufficient lead in terms of cost per kg. vs Ariane5 and the new Ariane 6. Starship may well crush further the competition if similar success rate is achieved.

References and Jupyter Notebooks

[1] https://arstechnica.com/science/2021/03/european-leaders-say-an-immediate-response-needed-to-the-rise-of-spacex/
[2] https://www.arianespace.com/vehicle/ariane-5/
[3] https://www.arianespace.com/wp-content/uploads/2020/06/Arianespace_Brochure_Ariane5_Sept2019.pdf
[4] https://www.arianespace.com/ariane-6/
[5] https://www.arianespace.com/wp-content/uploads/2020/06/Arianespace_Brochure_Ariane6_Sept2019.pdf
[6] https://www.inverse.com/innovation/ariane-6-vs-spacex
[7] https://adsabs.harvard.edu/full/1996ESASP.386..237F

Appendix: SQL queries

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

Total Payload Mass

Calculate the total payload carried by boosters from NASA (2 methods)

```
#  For validation purposes... sum in df_NASA_CRS  'PAYLOAD_MASS__KG_'   column
df_NASA_CRS=pd.read_sql_query("Select * from spacex_v11  where Customer='NASA (CRS)'",conn)
print(df_NASA_CRS.head(2))
print('----')
print('Total payload mass, customer= NASA (CRS):', df_NASA_CRS['PAYLOAD_MASS__KG_'].sum(),' kg')
```

```
   id        Date      Time (UTC) Booster_Version  Launch_Site       Payload  \
0   4  2012-10-08 0 days 00:35:00  F9 v1.0  B0006  CCAFS LC-40  SpaceX CRS-1
1   5  2013-03-01 0 days 15:10:00  F9 v1.0  B0007  CCAFS LC-40  SpaceX CRS-2

   PAYLOAD_MASS__KG_      Orbit    Customer Mission_Outcome Landing_Outcome
0                500  LEO (ISS)  NASA (CRS)         Success      No attempt
1                677  LEO (ISS)  NASA (CRS)         Success      No attempt
----
Total payload mass, customer= NASA (CRS): 45596  kg
```

```
#  Based on SQL only...
sql_nasa_crs_mass= """ Select sum(PAYLOAD_MASS__KG_) as 'Total payload mass (kg) NASA CRS'
                    from spacex_v11
                    where Customer='NASA (CRS)' """
payload_NASA_CRS=pd.read_sql_query(sql_nasa_crs_mass,conn)
print(payload_NASA_CRS)
```

```
   Total payload mass (kg) NASA CRS
0                           45596.0
```

Average Payload Mass by F9 v1.1

Calculate the average payload mass carried by booster version F9 v1.1

```
payload_F9v11=pd.read_sql_query("Select avg(PAYLOAD_MASS__KG_) as 'avg mass (kg)' from spacex_v11 where Booster_Version='F9 v1.1'",conn)
print(payload_F9v11)
```

```
   avg mass (kg)
0         2928.4
```

First Successful Ground Landing Date

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

```
min_Date_success_landing=pd.read_sql_query("select min(Date) from spacex_v11 where Landing_Outcome = 'Success (ground pad)'",conn)
print(min_Date_success_landing)
```

```
    min(Date)
0  2015-12-22
```

Boosters Carried Maximum Payload

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

```
#  sql query
qboost= """Select distinct Booster_Version, max(PAYLOAD_MASS__KG_) as max_payload_mass
from spacex_v11
group by Booster_Version
order by max_payload_mass  desc"""
```

```
boost_max_load= pd.read_sql_query(qboost,conn)
boost_max_load.head(5)
```

	Booster_Version	max_payload_mass
0	F9 B5 B1049.4	15600
1	F9 B5 B1060.2	15600
2	F9 B5 B1048.4	15600
3	F9 B5 B1048.5	15600
4	F9 B5 B1056.4	15600

2015 Launch Records

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

```
#  sql query
q_failed_landing= """ Select Date, Booster_Version, Launch_Site, Landing_Outcome
                   from spacex_v11
                   where Landing_Outcome = 'Failure (drone ship)'
                   and Date like '%2015%'   """
```

```
fail_drone= pd.read_sql_query(q_failed_landing,conn)
fail_drone.head(5)
```

	Date	Booster_Version	Launch_Site	Landing_Outcome
0	2015-01-10	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
1	2015-04-14	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

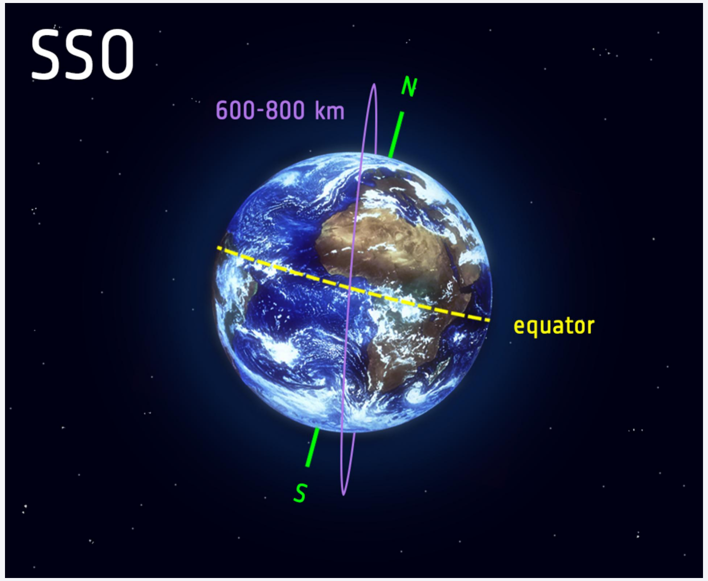
```
#  sql query
q_count_landing= """ Select Landing_Outcome, count(*) as count_landings
                    from spacex_v11
                    where Date between '2010-06-04' and '2017-03-20'
                    group by Landing_Outcome
                    order by count_landings desc  """
```

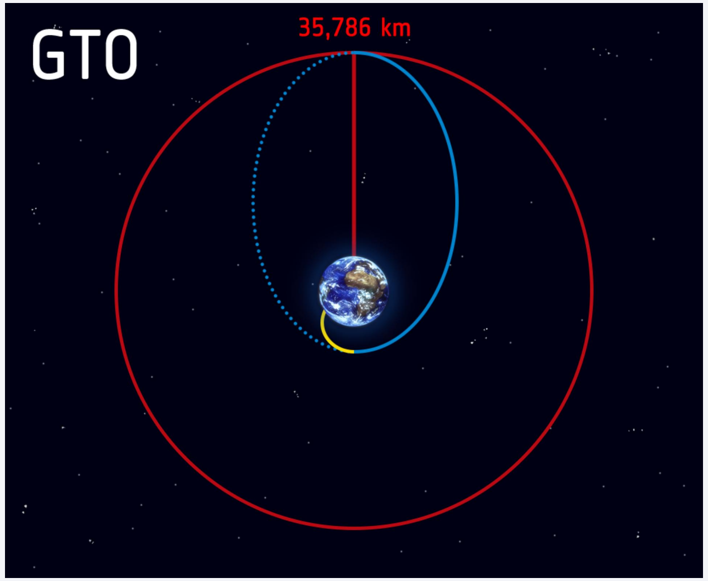
```
count_landing= pd.read_sql_query(q_count_landing,conn)
count_landing.head(10)
```

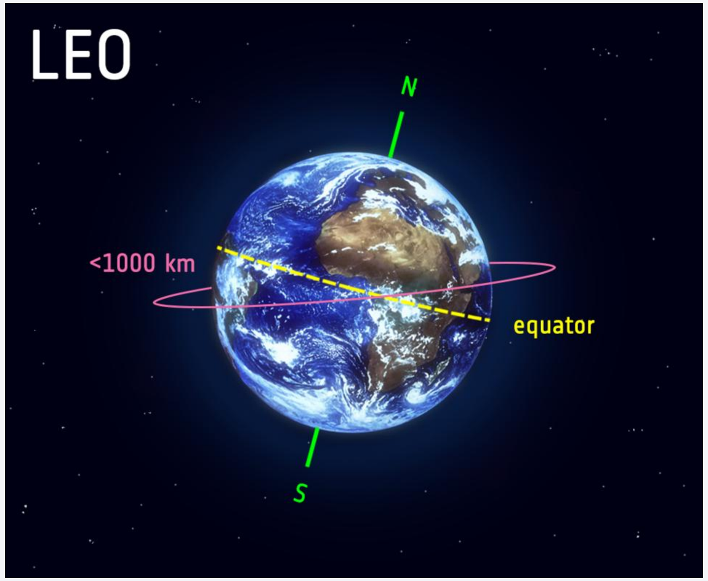
	Landing_Outcome	count_landings
0	No attempt	10
1	Failure (drone ship)	5
2	Success (drone ship)	5
3	Controlled (ocean)	3
4	Success (ground pad)	3
5	Uncontrolled (ocean)	2
6	Failure (parachute)	1
7	Precluded (drone ship)	1

Glossary

PO, SSO, GTO, GEO. From ESA website.








Thank you !