IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

E. Willmott
24 March 2022

Outline

Please note:

1) Internal hyperlinks do not work properly when reading a pdf file in a Web browser. They link to the subsequent page. Everything else works as it should, so please navigate using the page numbers provided. On the other hand, Adobe Acrobat Reader may not load all graphs due to memory management issues, although this seems to be resolved when using either Full Screen mode, Ctrl+L or Read mode, Ctrl+H

2) This report was created with Microsoft PowerPoint 2010 desktop version, as I was unable to create a Microsoft account to use PowerPoint Online.

Executive Summary (1/3): Methodology

- Methodologies:
  - Data Collection with SpaceX API[1] and Web scraping[2]
  - Data Wrangling
  - Exploratory Data Analysis -- EDA with SQL and Data Visualization
  - Launch Sites Locations Analysis with Folium
  - Building a Dashboard with Plotly Dash
  - Predictive Analysis -- Classification with Logistic Regression, SVM, Decision Tree, and KNN

Executive Summary (2/3): Results

* Results:
  * SpaceX’s overall success rate (first stage landed successfully): 66.7%
  * Success rate improved over time (determined by flight number)
  * Success rate depended on launch site, with higher success rates for sites KSC LC-39A and VAFB SLC-4E
  * For site CCAFS SLC-40, success rate was higher for heavier payload masses
  * Success rate depended on orbit type, with 100% success for orbit types ES-L1 (1/1), GEO (1/1), HEO (1/1) and SSO (5/5)
  * For some other orbit types, success improved for heavier payload masses or over time (determined by flight number)
  * Success rate continually improved from 2013 to 2020

Executive Summary (3/3): Results continued

* Results continued…
  * SQL was used to query the data and gain informative insights (detailed elsewhere (p.66))
  * All launch sites were in close proximity to the coast line and the Equator
  * One launch site was in close proximity to railway and highway, while not too close to cities
  * Launch site KSC LC-39A had the largest number of successful launches and the highest success rate
  * Some payload ranges had higher success rates: 1952-3696.65 and 4600-5300 kgs
  * FT booster version had the highest success rate
  * All classification models performed the same: 83.3% accuracy on the test data

Table of Contents (1/5)

Table of Contents (2/5)

Table of Contents (3/5)

Table of Contents (4/5)

Table of Contents (5/5)

* Appendices continued…

Introduction (1/2): Background

* I assume the role of a Data Scientist who works for SpaceY, a start-up intending to compete with SpaceX.
* "SpaceX designs, manufactures and launches advanced rockets and spacecraft."[3]
* SpaceX Falcon 9 rocket launches cost $62M each, while launches by other providers cost >$165M each.[4]
* Much of the savings are because SpaceX can reuse its first stage.[4]
* If we predict whether the first stage will land successfully, we could determine launch cost.
* This will help SpaceY make more informed bids against SpaceX for rocket launches.

Introduction (2/2): Questions

1. Which features affect a successful landing and how?

2. Can we predict if the first stage of Falcon 9 will land successfully?

Methodology

Methodology (1/2)

Executive Summary

* Data collection methodology:
  * SpaceX API[1]
  * Web scraping from Wikipedia List of Falcon 9 and Falcon Heavy launches[2]

* Perform data wrangling:
  * Replace missing values, create training labels to represent launch outcome.

Methodology (2/2)

Executive Summary continued…

- Perform exploratory data analysis (EDA) with visualization and SQL
- Perform interactive visual analytics with Folium and Plotly Dash
- Perform predictive analysis with classification models:
  - Data split to training and test data.
  - Logistic Regression, SVM, Decision Tree and KNN models built and tuned with GridSearchCV().
  - Evaluated model accuracy on the test data using .score().

Data Collection

Data sets were collected with SpaceX API[1] and through Web scraping from Wikipedia[2]:

- SpaceX API[1]
  - Collected launches, rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
    - GET request.
    - Filter the dataframe to only include Falcon 9 launches.
    - Replace missing values for PayloadMass with its mean.
- Wikipedia: List of Falcon 9 and Falcon Heavy launches (a snapshot of Wikipedia page updated on 9 June 2021)[2].
  - Web scraping with BeautifulSoup.

Data Collection: SpaceX API

1. GET request to obtain SpaceX launch data.
2. Decode with .json() and convert to Pandas dataframe **data** with .json.normalize().
3. Use API again to get more data using the IDs for each launch with helper functions (see flowchart →).
4. Create a dictionary and a Pandas dataframe **launch_df**.
5. Filter **launch_df** to only include Falcon 9 launches.
6. Replace missing values of PayloadMass with its mean.

Flowchart 1: SpaceX API calls

GET request

↓

.json() and .json.normalize()
convert to pd.dataframe

↓

getBoosterVersion(data)

↓

getLaunchSite(data)

↓

getPayloadData(data)

↓

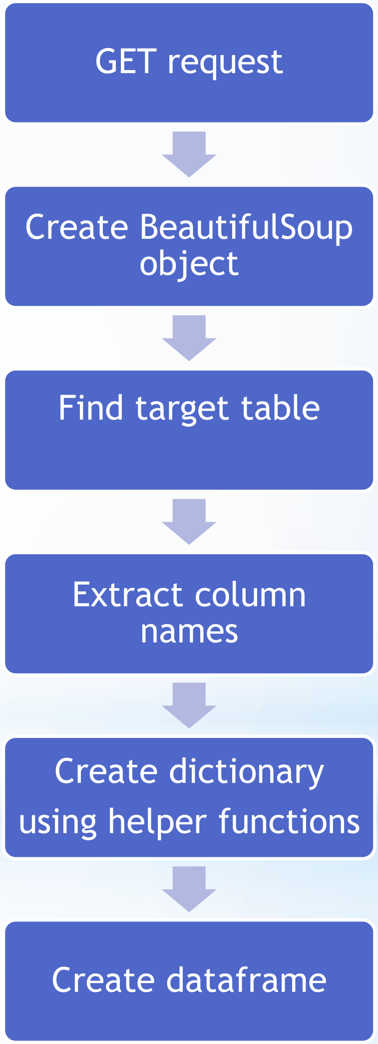
getCoreData(data)

GitHub URL of completed SpaceX API calls notebook

Data Collection: Web Scraping (Wikipedia)

- From Wikipedia List of Falcon 9 and Falcon Heavy launches (9^{th} June 2021)[2]
- Web scraping with BeautifulSoup (see flowchart →)
- Convert to a pandas dataframe

GitHub URL of completed Web scraping notebook



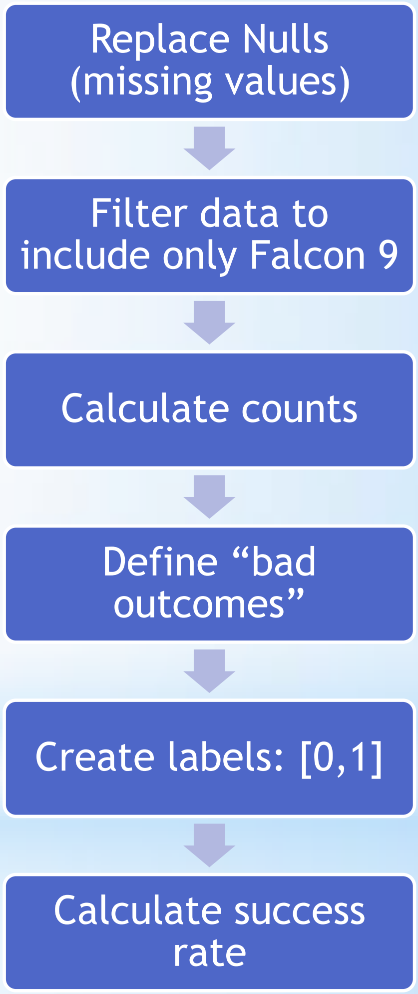
Flowchart 2: Web scraping

Data Wrangling

* Done in previous steps:
  * Filtered dataset to only include Falcon 9.
  * Replaced missing values of PayloadMass by its mean.
* Calculate:
  * # of launches per site.
  * # and occurrence of each orbit.
  * # and occurrence of mission outcome per orbit type.
* Convert landing outcome to training labels:
  * **1**: successful landing
  * **0**: unsuccessful landing
* Determine success rate.
* In the next steps:
  * Create dummy variables and cast variables to float.

[GitHub URL of completed data wrangling notebook]

Flowchart 3: Data wrangling



EDA with Data Visualization

Charts were plotted to visualize the effects of flight number, payload, launch site, orbit type and year on launch outcome and success rate:

1. Effect of flight number (continuous flight attempts) and launch site on launch outcome.
2. Effect of payload and launch site on launch outcome.
3. Success rate of each orbit type.
4. Effect of flight number and orbit type on launch outcome.
5. Effect of payload and orbit type on launch outcome.
6. Launch success yearly trend.

GitHub URL of completed EDA with data visualization notebook

EDA with SQL (1/2)

The following SQL queries were performed:

* Names of unique launch sites in the space mission.
* Five records where launch sites begin with the string ‘CCA’.
* Total payload mass carried by boosters launched by NASA (CRS).
* Average payload mass carried by booster version F9 v1.1.
* Date when the first successful landing outcome in ground pad was achieved.

[GitHub URL of completed EDA with SQL notebook]

EDA with SQL (2/2)

…the following SQL queries were performed:

* Names of boosters that had a successful landing in drone ship and had payload mass between 4,000 and 6,000 kg.
* Total number of successful and failed mission outcomes.
* Names of booster versions that carried the maximum payload mass.
* Failed landing outcomes in drone ship, their booster versions, and launch site names in 2015.
* Count of landing outcomes between 2010-06-04 and 2017-03-20.

GitHub URL of completed EDA with SQL notebook

Build an Interactive Map with Folium

The map objects below were added to a folium map:

* A circle and a marker for each launch site to mark its location.
* A marker cluster marking each launch outcome (success/failure) with a color to visualize launch outcomes.
* Lines to mark the proximities of a launch site (CCAFS SLC-40), including the coastline, a city (Orlando), a railway and a highway to understand the distances and proximities of a launch site.

GitHub URL of completed interactive map with Folium

nbviewer URL of completed interactive map with Folium

Please use the above nbviewer URL to view interactive map.

Build a Dashboard with Plotly Dash

The plots/graphs and interactions below were added to a dashboard:

1. Dropdown list to enable launch site selection.
2. Pie chart to show total successful launches count for all sites;
   or show success vs failed counts for a specific launch site.
3. Range slider to select payload range.
4. Scatter chart to show correlation between payload and launch success.

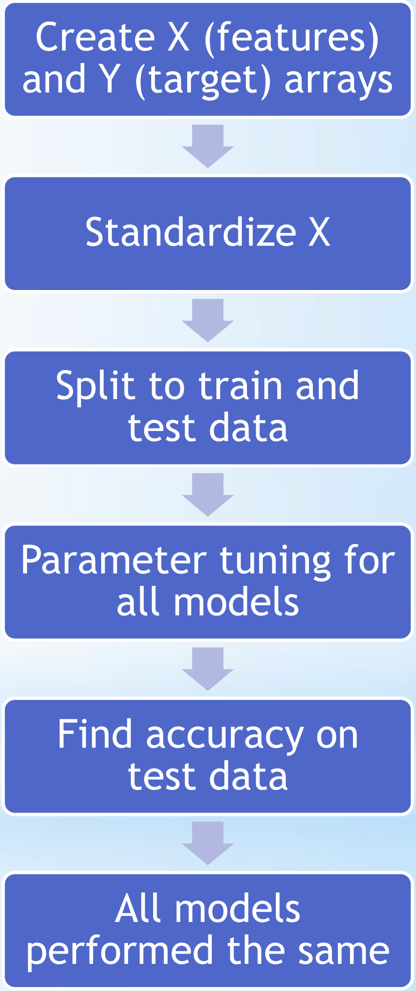
GitHub URL of completed Plotly Dash lab (python file)

Predictive Analysis (Classification)

Built, evaluated, improved, and found the best performing classification model:

1. Create X (features) & Y (target) arrays based on Spacex dataset.
2. Standardize X with StandardScaler().
3. Split X and Y to training and test data.
4. Models used: Logistic regression, SVM, Decision tree, and KNN.
5. Best parameters found for each model using GridSearchCV().
6. Evaluated accuracy on test data with .score() & confusion matrix.
7. All models performed practically the same.

GitHub URL of completed predictive analysis lab



Flowchart 4: Prediction

Results

Insights drawn from EDA

Flight Number vs Launch Site

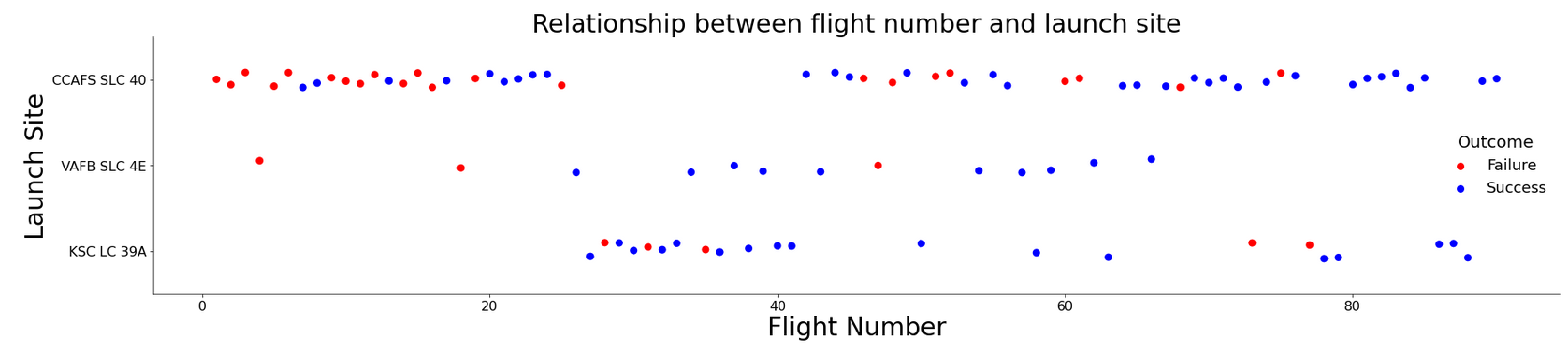


Figure 1

- Increasing flight number was associated with higher success rate.
- VAFB SLC-4E and KSC LC-39A had higher success rates.

Payload vs Launch Site

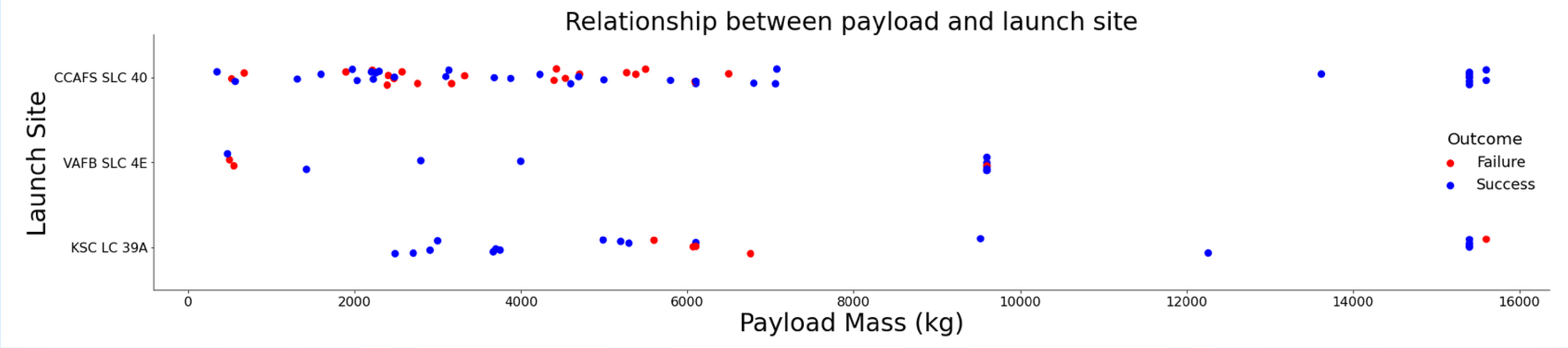


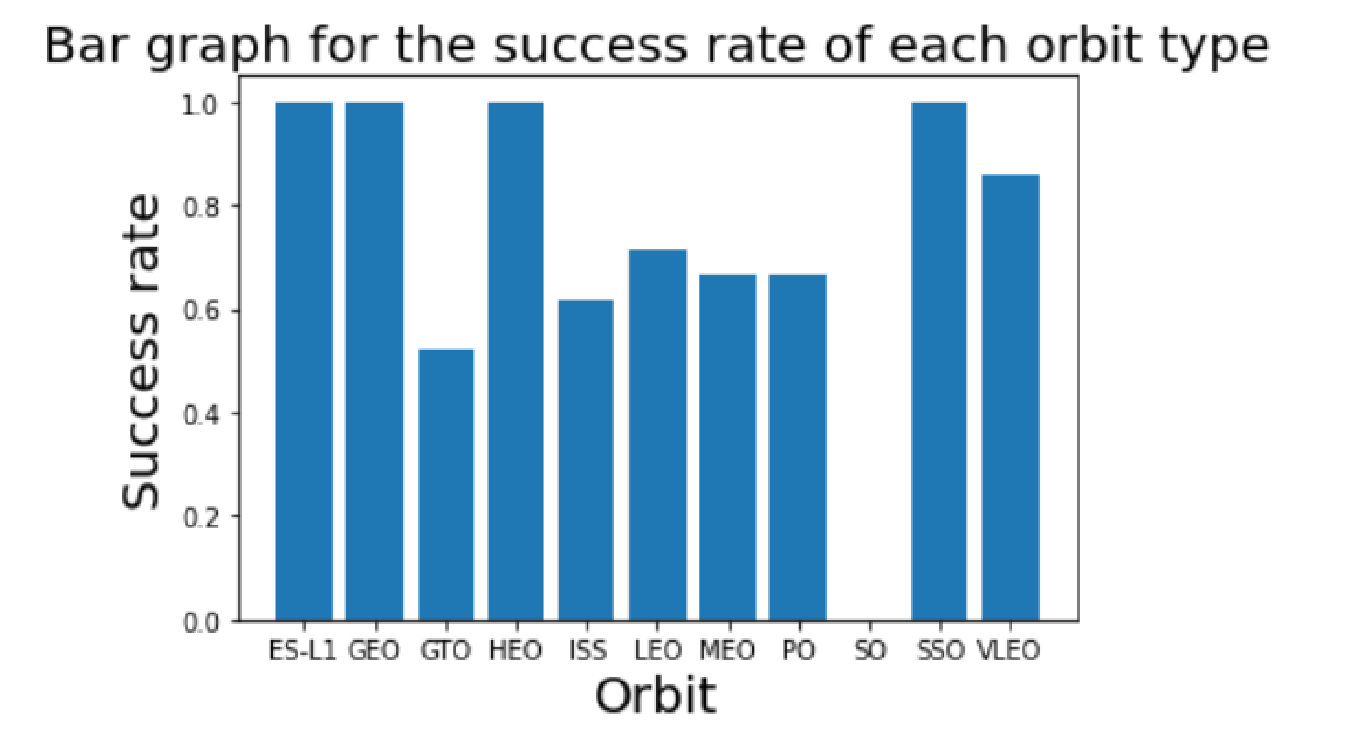
Figure 2

- CCAFS SLC-40 had higher success rates for heavier payload masses.
- VAFB SLC-4E had no heavy payload mass rockets (> 10K kg).
- KSC LC-39A had high success rate for low and heavy payload masses, but not for payload mass of approximately 6K kg.

Success Rate by Orbit Type

Figure 3

* ES-L1, GEO, HEO and SSO:
  Success rate 100%
* VLEO:
  Success rate > 80%
* Other orbits:
  Success rate < 70%



Flight Number vs Orbit Type

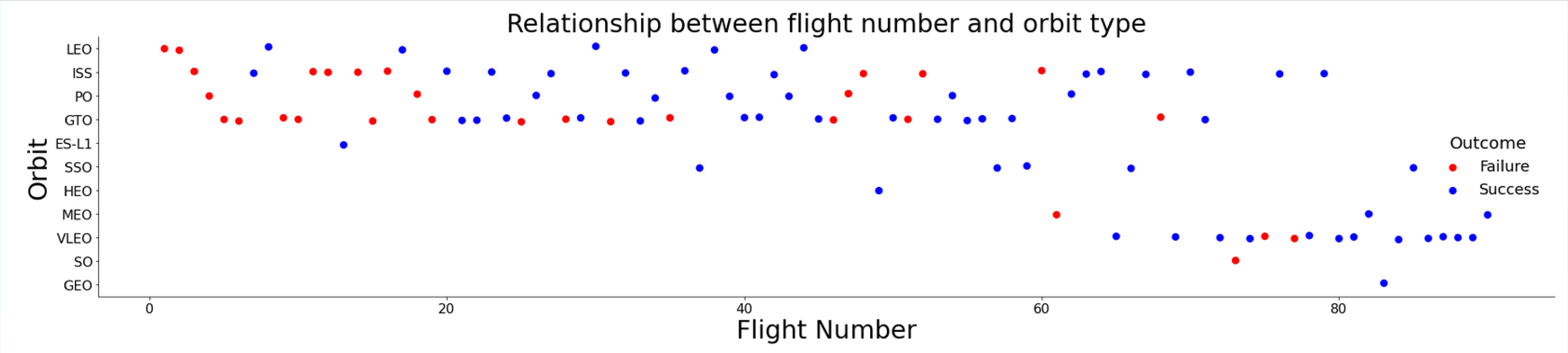


Figure 4

- LEO: Higher success rate for higher flight numbers.
- GTO: Success rate is independent of flight number.
- SSO, HEO, VLEO, SO and GEO: Higher flight numbers only (later launches).

Payload vs Orbit Type

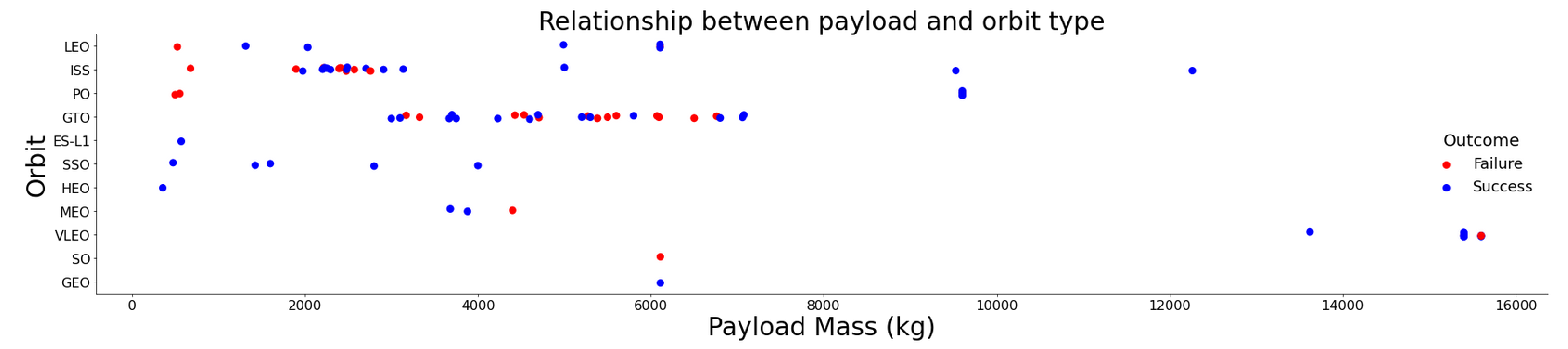


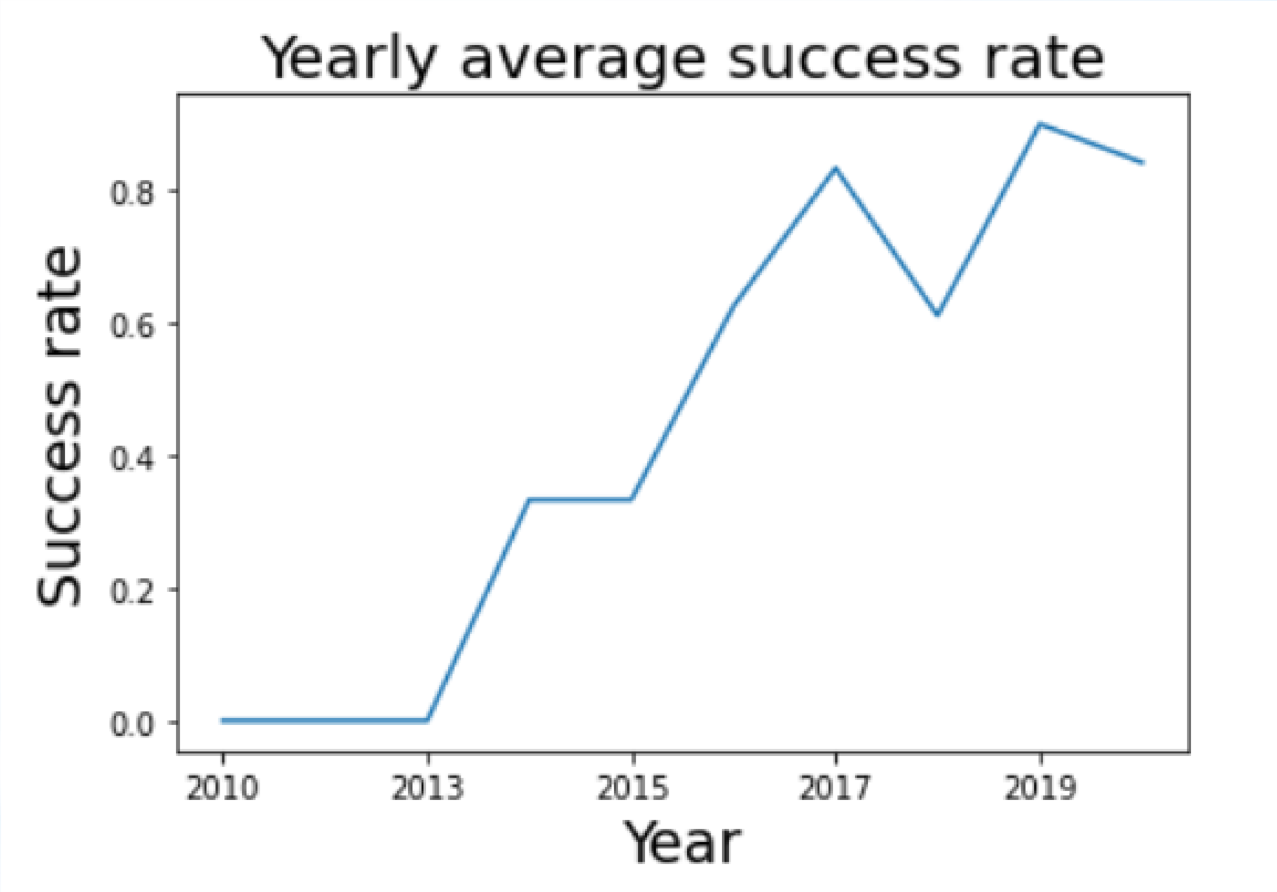
Figure 5

* LEO, ISS and Polar: Higher success rate for heavier payload masses.
* GTO: Success rate is independent of payload mass.
* ES-L1, SSO, HEO, MEO, SO, GEO: No heavy payload masses.
* VLEO: Only heavy payload masses.

Launch Success: Yearly Trend

Figure 6

* Success rate continually increased from 2013 to 2020.



All Launch Sites (Query 1)

* Find the names of the unique launch sites:

Display the names of the unique launch sites in the space mission

In [4]:
```
%sql select distinct Launch_site as "Launch Site" from SPACEXDATASET;
```

Out[4]:

Launch Site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

* There are 4 unique launch sites in the SpaceX dataset:
  * CCAFS LC-40
  * CCAFS SLC-40
  * KSC LC-39A
  * VAFB SLC-4E

Launch Site Names Beginning with ‘CCA’ (Query 2)

* Find 5 records where launch site names begin with ‘CCA’:

Display 5 records where launch sites begin with the string 'CCA'

In [5]:

```
%sql select * from SPACEXDATASET where launch_site like 'CCA%' limit 5;
```

Out[5]:

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

* 5 records are shown above where the launch site names begin with ‘CCA’, e.g. CCAFS LC-40.
* Each record has: date, time (UTC), booster version, launch site, payload, payload mass (kg), orbit, customer, mission outcome, and landing outcome.

Total Payload Mass (Query 3)

* Calculate the total payload mass carried by boosters of NASA (CRS):

Display the total payload mass carried by boosters launched by NASA (CRS)

In [6]:
```
%sql select sum(payload_mass__kg_) as "Total payload mass (kg)" from SPACEXDATASET where customer = 'NASA (CRS)';
```

Out[6]:

Total payload mass (kg)
45596

* The total payload mass carried by boosters of NASA (CRS): 45,596 kg.

Average Payload Mass by F9 v1.1 (Query 4)

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

Display average payload mass carried by booster version F9 v1.1

In [7]:
```
%sql select avg(payload_mass__kg_) as "Average payload mass (kg)" from SPACEXDATASET where booster_version like 'F9 v1.1%';
```

Out[7]:

Average payload mass (kg)
2534

* The average payload mass carried by booster version F9 v1.1: 2,534 kg.

First Successful Ground Landing Date (Query 5)

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

List the date when the first successful landing outcome in ground pad was acheived.

Hint:Use min function

In [8]:

```
%sql select min(DATE)as "Date of the first successful landing in ground pad" from SPACEXDATASET where landing__outcome = 'Success (ground
```

Out[8]:

Date of the first successful landing in ground pad
2015-12-22

* The date of the first successful landing outcome on ground pad:

22 December 2015

Successful Drone Ship Landing with Payload between 4,000 & 6,000 kg (Query 6)

* List the names of boosters which have successfully landed on drone ships and had a payload mass greater than 4,000 kg but less than 6,000 kg:

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

In [9]:

```
%sql select distinct(booster_version) as "Booster version names" from SPACEXDATASET where landing__outcome = 'Success (drone ship)'\
and payload_mass__kg_ between 4000 and 6000;
```

Out[9]:

Booster version names
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

* Names of boosters which have successfully landed on drone ships and had a payload mass between 4,000-6,000 kg are displayed above.
* All were FT booster versions.

Total Number of Successful and Failed Mission Outcomes (Query 7)

* Calculate the total number of successful and failed mission outcomes:

List the total number of successful and failure mission outcomes

In [10]:
```
%sql select mission_outcome as "Mission outcome", count(mission_outcome) as "Count" from SPACEXDATASET group by mission_outcome\
order by Count desc;
```

Out[10]:

Mission outcome	Count
Success	99
Failure (in flight)	1
Success (payload status unclear)	1

* The total number of successful and failed mission outcomes:

99 Successes + 1 Success (payload status unclear), 1 Failure (in flight).

Boosters Carrying Maximum Payload (Query 8)

* List the names of the boosters that carried the maximum payload mass:

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

In [11]:
```
%sql select distinct(booster_version) as "Booster version names"\
from SPACEXDATASET where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXDATASET);
```

Out[11]:

Booster version names
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

* The boosters that carried the maximum payload mass were all version B5.

2015 Launch Records (Query 9)

- List the failed landing outcomes on drone ships, their booster versions, and launch site names for the year 2015:

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

In [12]:
```
%sql select booster_version as "Booster version names", launch_site as "Launch site",\
landing__outcome as "Landing outcome" from SPACEXDATASET where landing__outcome = 'Failure (drone ship)' and year(DATE) = '2015';
```

Out[12]:

Booster version names	Launch site	Landing outcome
F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

- The failed landing outcomes on drone ships in 2015, with their booster versions and launch site names, are shown above.
- There were 2 failed landing outcomes on drone ships in 2015, both were F9 v1.1 booster versions and both launched from CCAFS LC-40.

Rank Landing Outcomes between 2010-06-04 and 2017-03-20 (Query 10)

* Rank the counts of landing outcomes (such as, Failure (drone ship) or Success (ground pad)) between 2010-06-04 & 2017-03-20 in descending order:

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

In [13]:

```
%sql select landing__outcome as "Landing outcome", count(landing__outcome) as "Count"\
from SPACEXDATASET where DATE between '2010-06-04' and '2017-03-20'\
group by landing__outcome order by count(landing__outcome) desc;
```

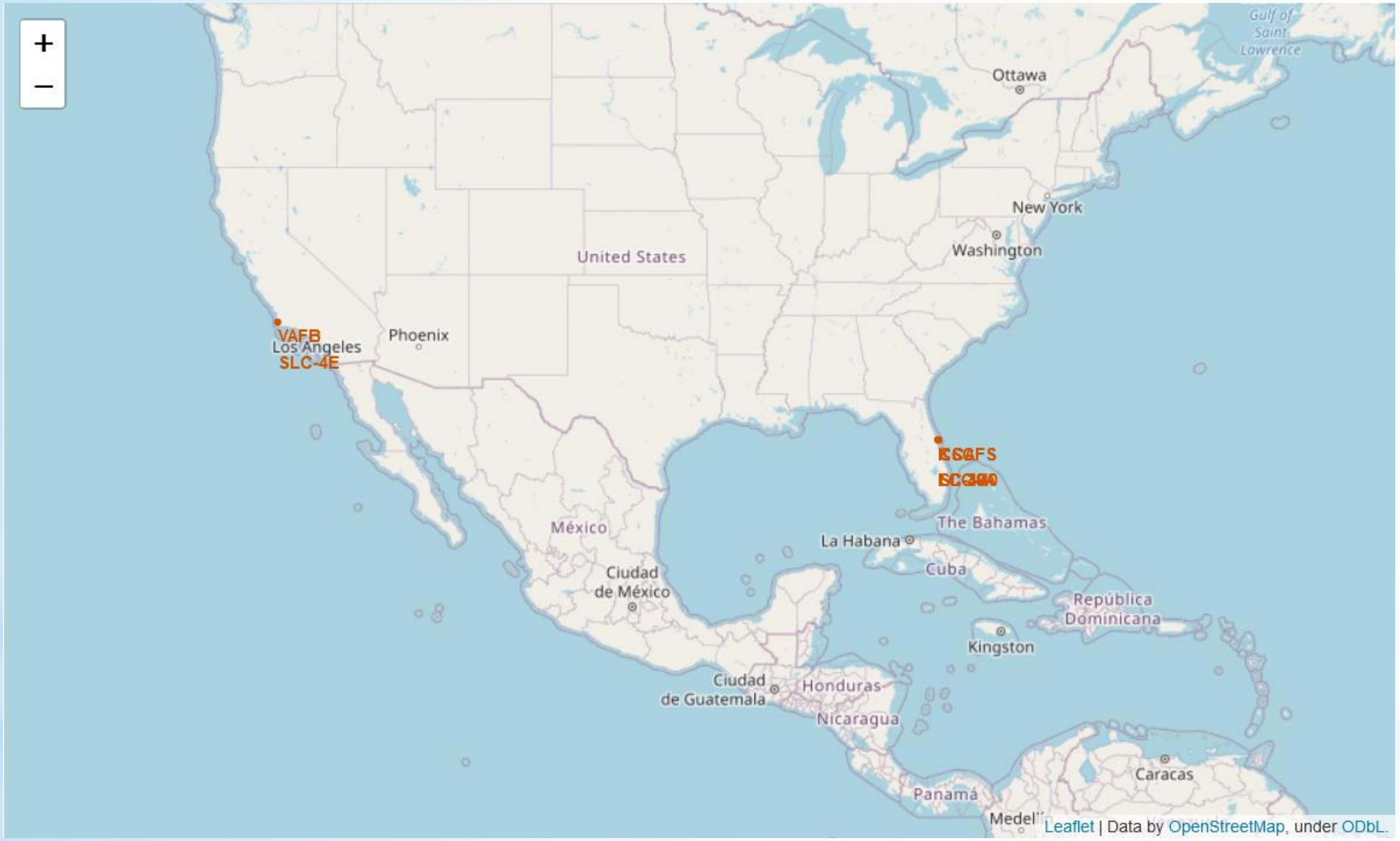
Out[13]:

Landing outcome	Count
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

* The counts (left) notably include:
  * 5 successful landings on a drone ship.
  * 3 successful landings on a ground pad.

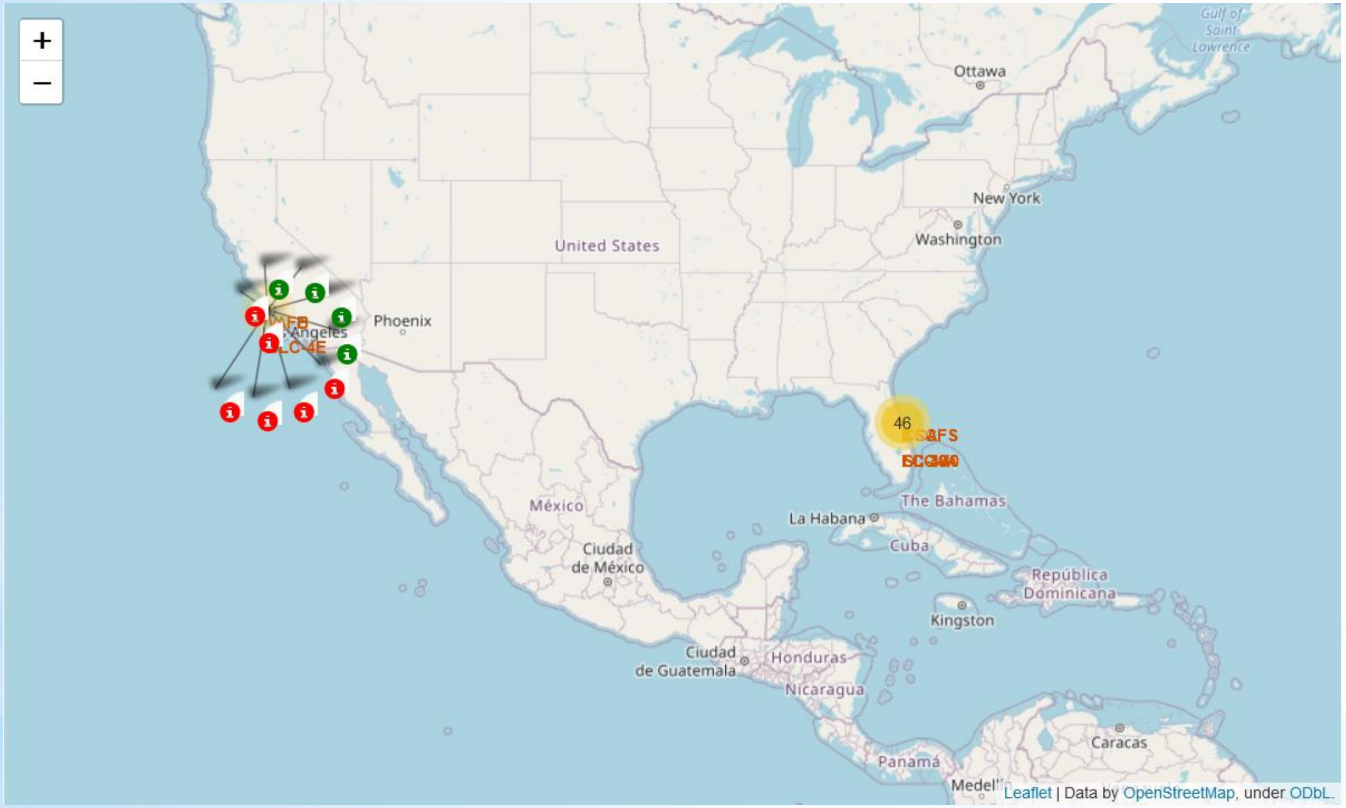
Launch Sites Proximities Analysis

Map of Launch Site Locations



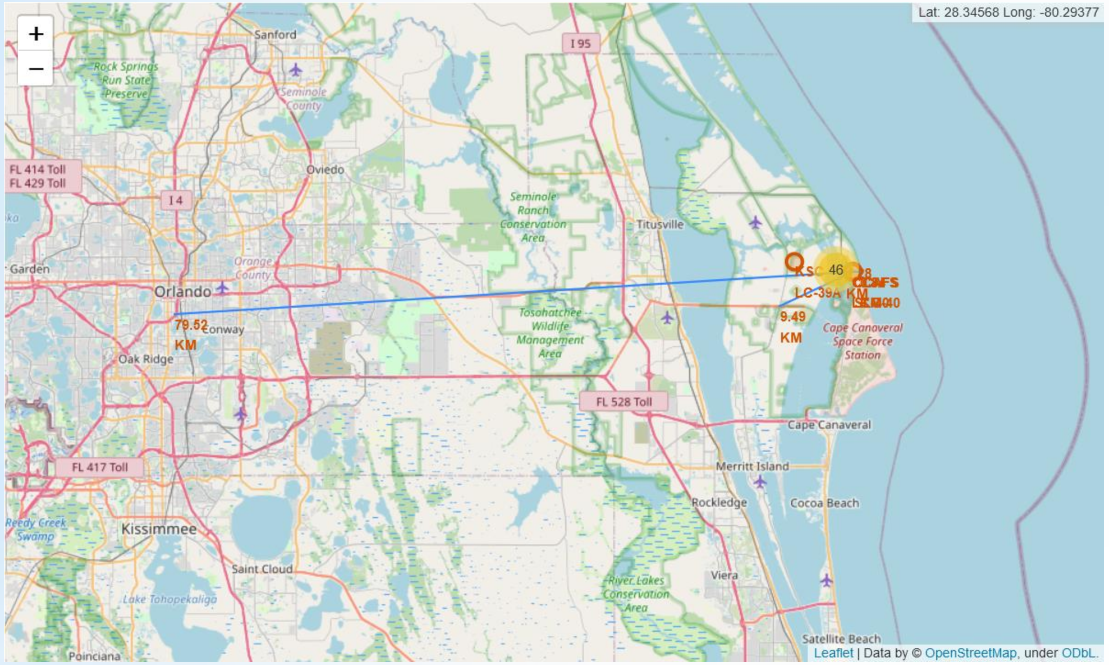
- This folium map presents all launch site locations on a 'global' map.
- All launch site locations are in very close proximity to a coastline. This is important to prevent danger to human life in case of failure.[5]
- All launch sites are in proximity to the Equator, supporting the rocket in its launch speed, due to the Earth's rotational speed being at its maximum on the equatorial plane.[5]

Map of Launch Outcomes



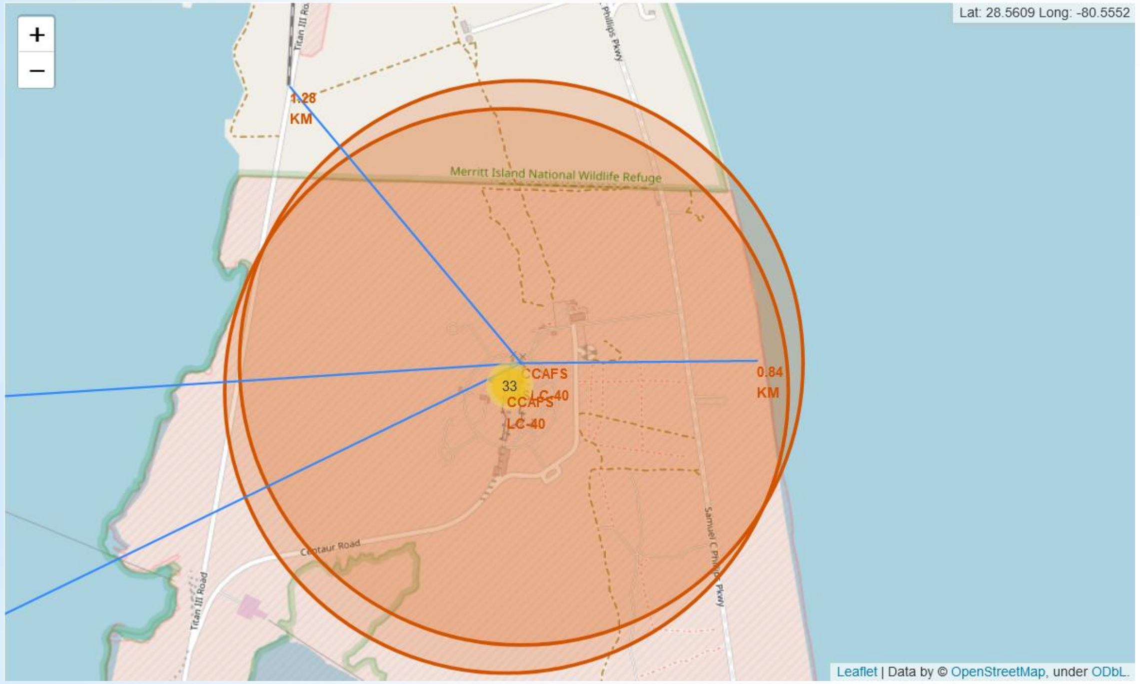
- This folium map shows color-labeled launch outcomes for VAFB SLC-4E.
- Successful outcomes are marked in green, with failures marked in red.
- The majority, 6 out of 10, launch outcomes shown, were a failure.
- A marker cluster in yellow marks a total of 46 launches for other launch sites without showing their actual launch outcomes.

Map of a Launch Site and its Proximities (1/2)



* This folium map shows CCAFS SLC-40 and its proximities, including a city (Orlando) at 79.52 km and a highway at 9.49 km.
* It displays close enough proximity to a highway allowing transportation, yet not too close so not to endanger traffic in case of failure.
* The city is not too close, also not to endanger the population in case of failure.

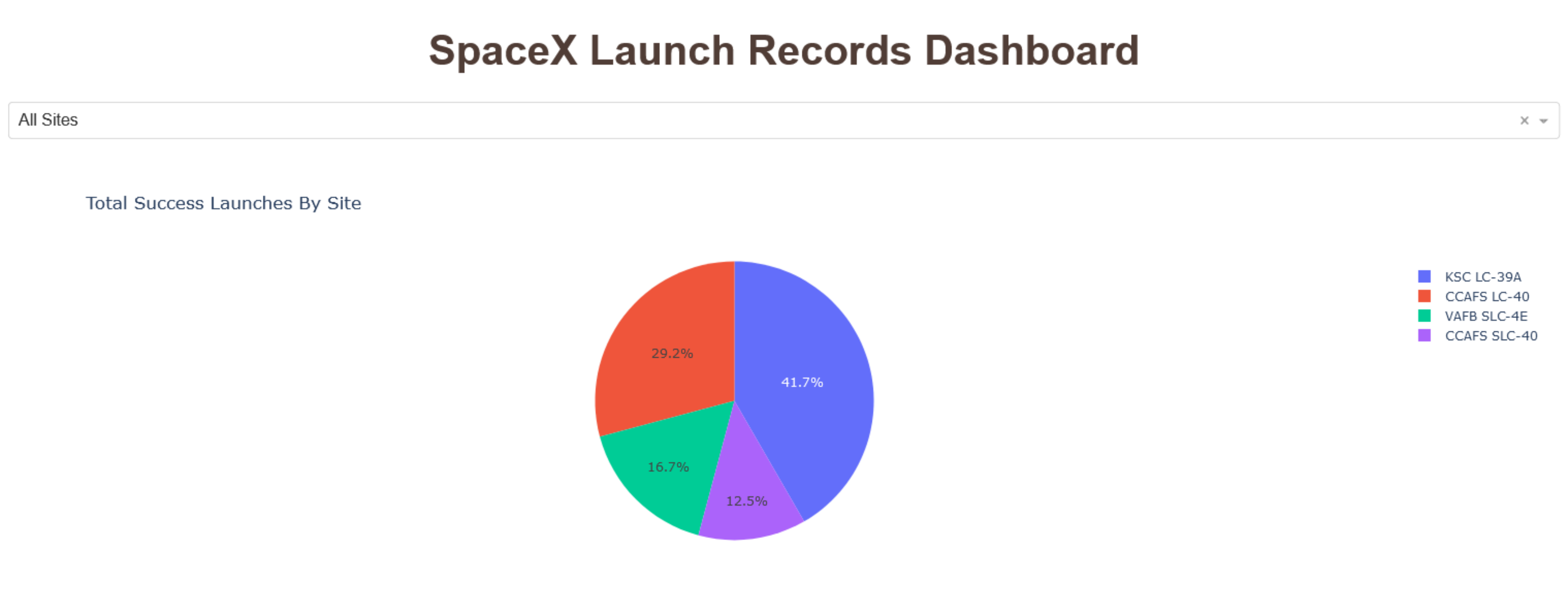
Map of a Launch Site and its Proximities (2/2)



- This folium map shows CCAFS SLC-40 and its proximities, including a railway at 1.28 km and a coastline at 0.84 km.
- Close proximity to a railway allows transportation.
- Close proximity to the coastline prevents danger to human life in case of failure, so that in such a case, parts will fall into the ocean.[5]

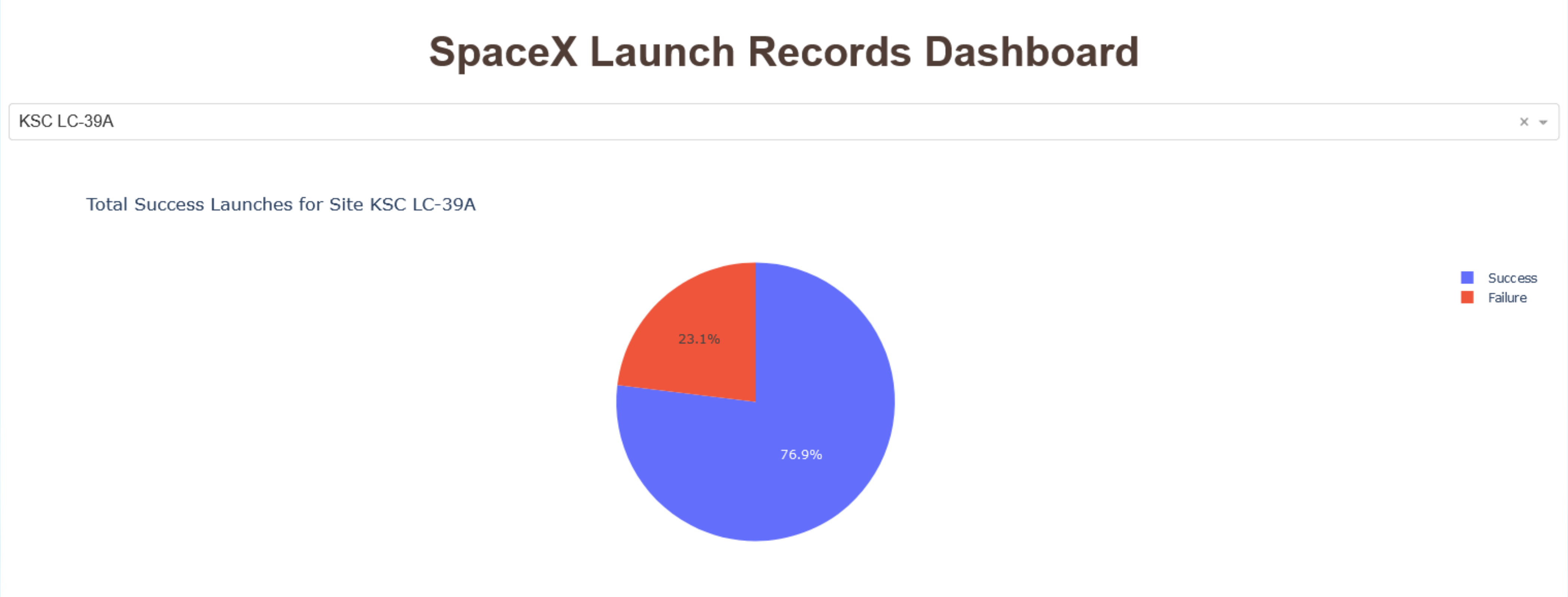
Build a Dashboard with Plotly Dash

SpaceX Launch Dashboard: Success Counts



* **Figure 7:** Pie chart showing launch success counts for all sites.
* **KSC LC-39A** had the largest number of successful launches: 41.7% of the total number of successful launches for all sites.

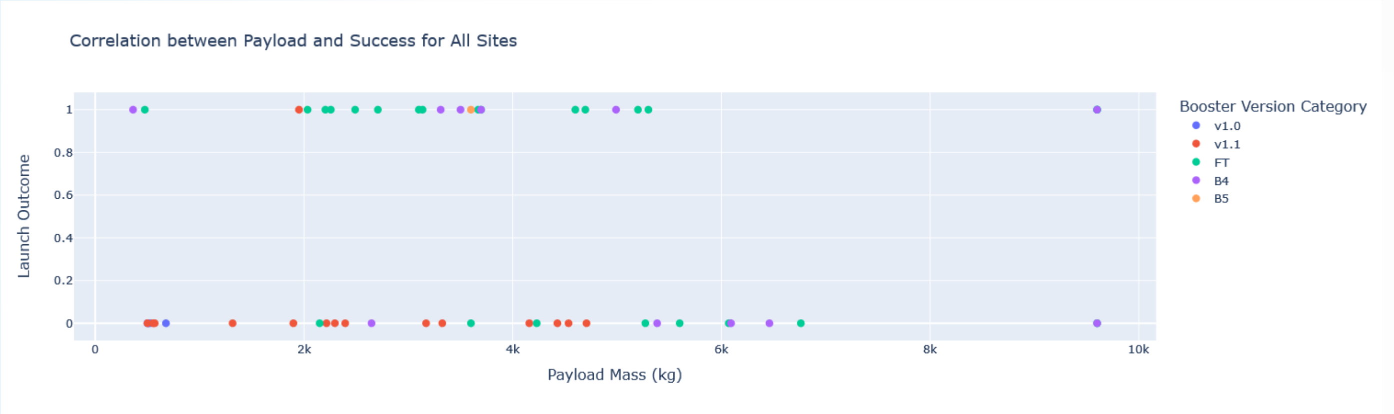
SpaceX Launch Dashboard: Highest Success Rate



* **Figure 8:** Pie chart for the launch site with the highest launch success rate.
* **KSC LC-39A** had the highest launch success rate: 76.9%.

SpaceX Launch Dashboard: Payload vs Outcome

Whole payload range



Payload range ~1000-5000 kg

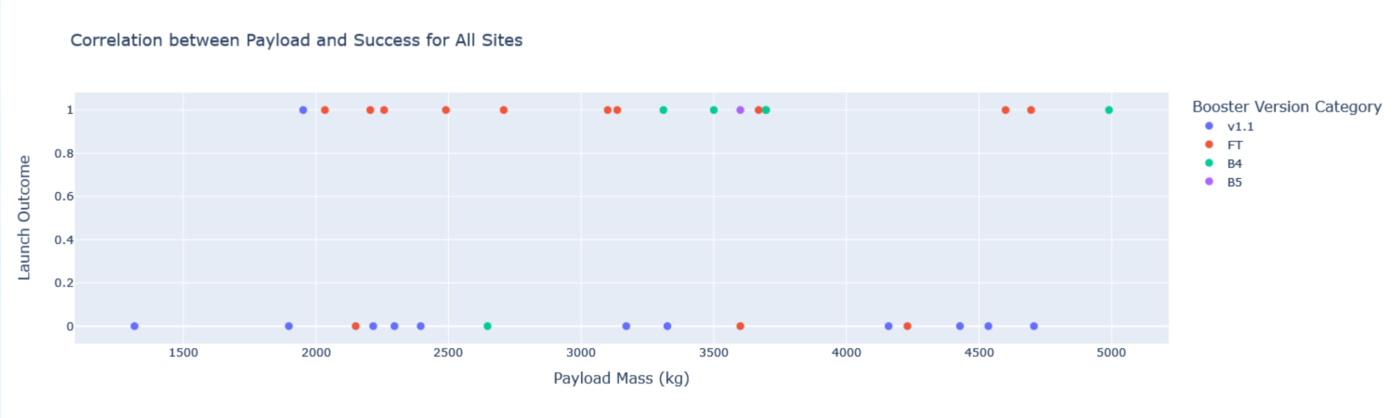


Figure 9

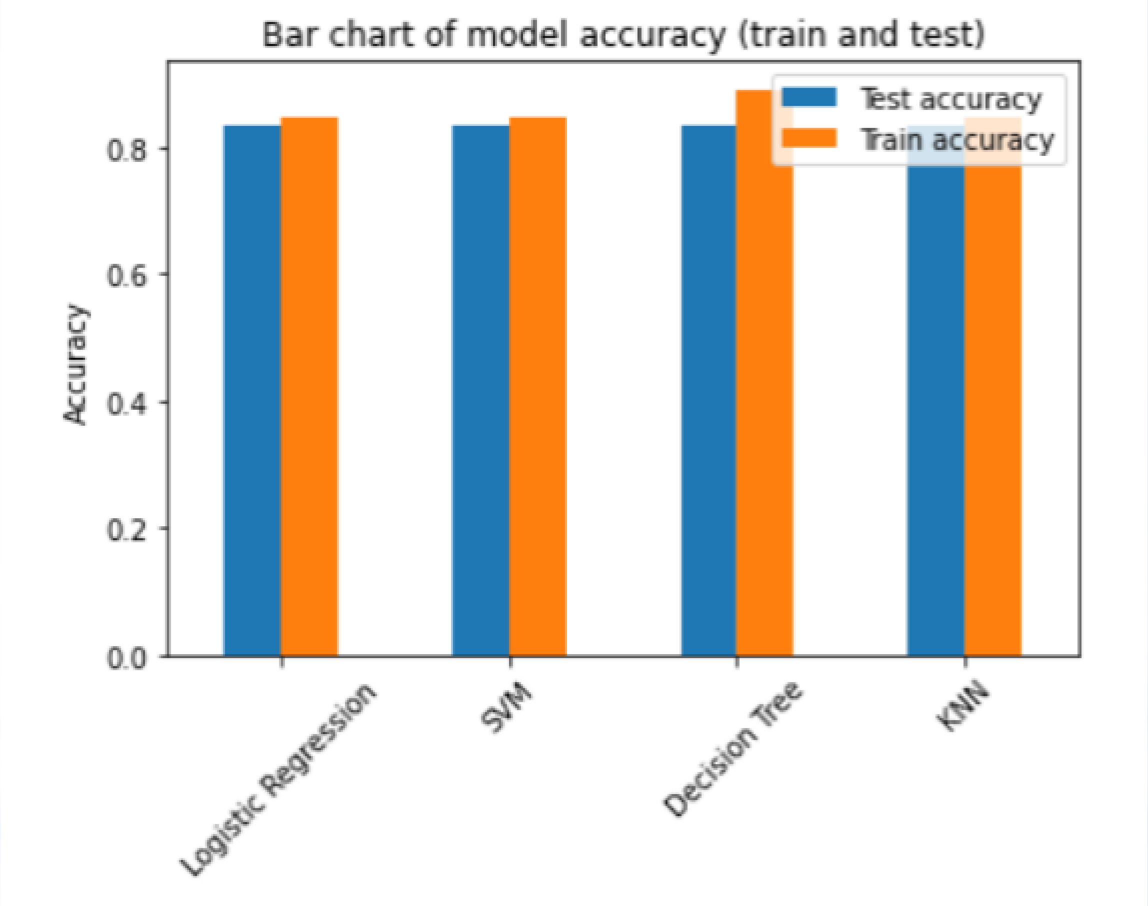
* Payload vs Launch Outcome scatter plots for all sites.
* Bottom plot shows payload ranges with highest success rates:
  1,952 - 3,696.65 kg
  and 4,600 - 5,300 kg.
* FT booster version had the highest success rate.

Predictive Analysis (Classification)

Classification Accuracy

Figure 10

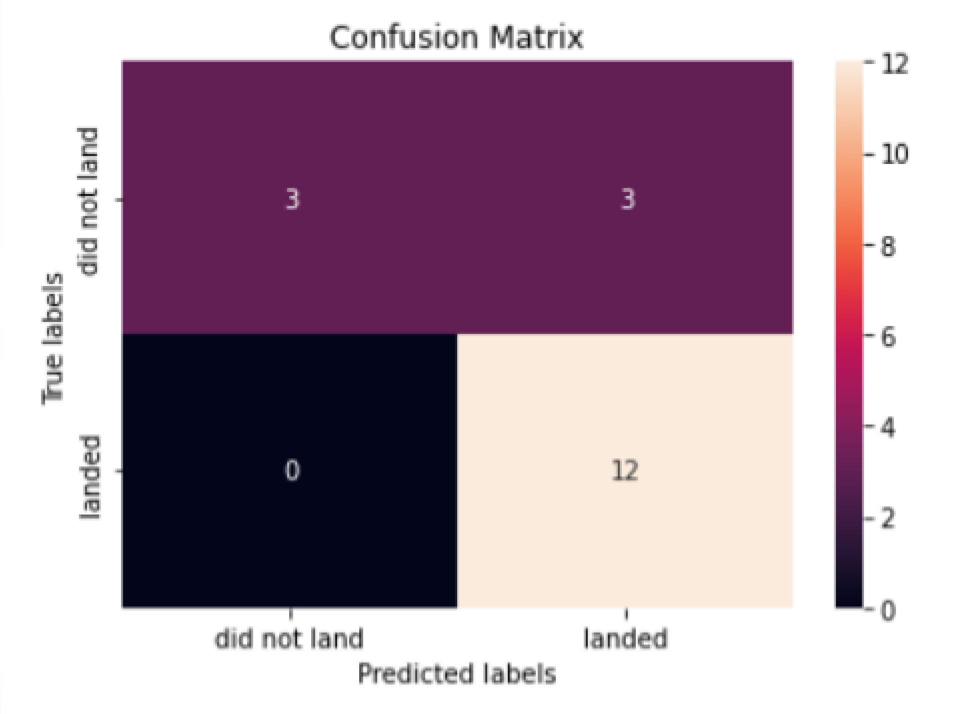
* Bar chart of model accuracy for all built classification models.
* **All models had the same classification accuracy on the test data: 83.3%.**



Confusion Matrix

Figure 11

* **The confusion matrix was identical for all models, as all models performed the same.**
* All models distinguished between the classes, with no false negatives → bottom left.
* All successful launches were predicted by the models to land successfully → bottom right.
  * Sensitivity TP/(TP + FN)[6]: 100%
* **The major problem with all models was false positives (wrongly predicted success) → top right.**
  * Specificity TN/(TN + FP)[6]: 50%.



Conclusions

Conclusions (1/4)

* **Overall launch success rate was 66.7%**
* The following features affect launch success rates:
  * Higher success rates with higher flight number and year, ie, with greater experience/technological advancements.
  * Launch site KSC LC-39A had the highest success rate.
  * Payload ranges 1,952-3,696.65 and 4,600-5,300 kgs had the highest success rates.
    * Launch site CCAFS SLC-40 had a higher success rate for heavier payloads.
  * Orbit type: 100% success rate for ES-L1, GEO, HEO and SSO. Mostly high orbits.
    * Polar, LEO and ISS: Higher success rate for heavier payloads. Mostly low orbits.
    * LEO: Higher success rate with increasing flight number. Low orbit.
  * FT Booster version had the highest success rate.

Conclusions (2/4)

* All launch sites are in close proximity to a coastline in order to prevent danger to human life in case of failure[5].
* All launch sites are on the Equator to support launch speed[5].
* At least one launch site was in close proximity to a railway and to a highway to allow transportation, but not too close to cities in order not to endanger the population.
* All classification models performed the same with 83.3% accuracy on the test data.
* **Future steps to improve model performance:** Add more data and features. Possibly use other classification models and/or ensembles of several models.
* **Further research:** Explore reasons for low success rate of certain launch sites, orbits and booster versions in order to enhance booster technology and improve launch/outcome success.

Conclusions (3/4)

* Success rate increased with time (experience/technological advancements).
* SpaceY could build on the experience/technological advancements of SpaceX in order to enhance its success rate.
* SpaceY could compete with SpaceX by decreasing its relative launch costs.
  * **SpaceY could enhance its success rate and thus decrease launch costs by:**
    * Using the identified payload ranges with the highest success rates.
    * Using only launch sites with high success rates, notably KSC LC-39A.
    * Launching largely to orbits with high success rates: ES-L1, GEO, HEO and SSO.
    * When launching to other orbits (Polar, LEO and ISS) use heavy payloads.
    * Basing SpaceY's booster on SpaceX's FT booster, which had the highest success rate.
    * Using the prediction models built herein to predict launch success before launching its rockets.

Conclusions (4/4)

* Higher success rates for higher payloads may be confounded by the fact that higher payloads (>10k kg) were used only on later flights (higher flight numbers), which are already associated with higher success rates (see Fig. 12 - p.69).

* **Please note:** Datasets (p.80) used for different stages of this project differ, which means the results obtained at different stages of the project differ based on the particular dataset used.

  * For example, the number of launches and the success rate for launch site VAFB SLC-4E differ based on the dataset used for EDA and data visualization (p.28) versus the dataset used for launch site locations analysis with Folium (p.46).

Appendices

Appendices (1/2)

Appendices (2/2)

References

1. "SpaceX API". https://api.spacexdata.com/v4/launches/past (accessed March 13, 2022).
2. Wikipedia, "List of Falcon 9 and Falcon Heavy launches", June 9, 2021. https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922 (accessed March 13, 2022).
3. SpaceX, 2022. https://www.spacex.com/ (accessed March 13, 2022).
4. IBM, "Applied Data Science Capstone", *Coursera*. https://www.coursera.org/learn/applied-data-science-capstone (accessed March 13, 2022).
5. A. Tiwari, "Why Are Rockets Launched From Areas Near The Equator?", Science ABC, Jan. 19, 2022. https://www.scienceabc.com/eyeopeners/why-are-rockets-launched-from-areas-near-the-equator.html (accessed March 13, 2022).
6. "False Positive Rate", *Split Software, Inc.*, 2022. https://www.split.io/glossary/false-positive-rate/ (accessed March 13, 2022).

SQL Queries Used

Results Summary: SQL (1/2)

1. The names of the launch sites are:
   * CCAFS LC-40
   * CCAFS SLC-40
   * KSC LC-39A
   * VAFB SLC-4E
2. Each record included: date, time (UTC), booster version, launch site, payload, payload mass (kg), orbit, customer, mission outcome, and landing outcome.
3. The total payload carried by boosters of NASA (CRS): 45,596 kg.
4. The average payload mass carried by booster version F9 v1.1: 2,534 kg.
5. The first successful landing on a ground pad was on 22 December 2015.

Results Summary: SQL (2/2)

6. The boosters that successfully landed on a drone ship and had payload masses between 4,000-6,000 kgs were all the FT version.
7. The total mission outcomes included 99 successful + 1 successful (payload status unclear), and 1 failure (in flight).
8. The boosters that carried the maximum payload mass were all the B5 version.
9. There were 2 failed landing outcomes on a drone ship in 2015. Both were F9 v1.1 booster versions and both were launched from CCAFS LC-40.
10. The count of landing outcomes between 2010-06-04 and 2017-03-20 was presented here (p.43).

Built Models: Best Parameters

* Logistic Regression:
  * 'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'
* SVM:
  * 'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'
* Decision Tree:
  * 'criterion': 'entropy', 'max_depth': 12, 'max_features': 'sqrt', 'min_samples_leaf': 4, 'min_samples_split': 2, 'splitter': 'random'
* KNN:
  * 'algorithm': 'auto', 'n_neighbors': 10, 'p': 1

Flight Number vs Payload

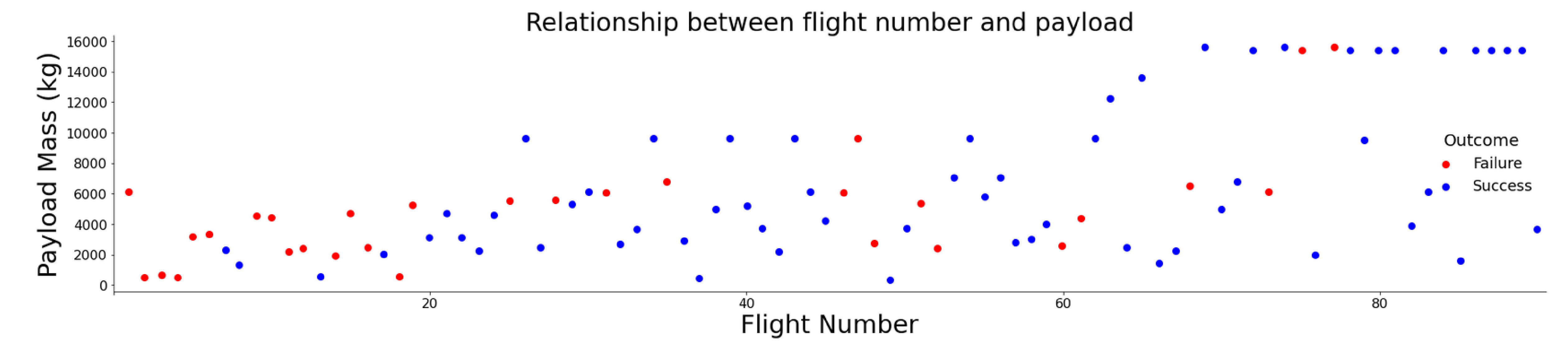


Figure 12

* Higher payloads were only used on later flights (higher flight numbers).
* Higher flight numbers and higher payloads were both associated with higher success rates.

Figures

Maps

Flowcharts

Images

Notebooks

1. GitHub URL of Completed SpaceX API Calls Notebook
2. GitHub URL of Completed Web Scraping Notebook
3. GitHub URL of Completed Data Wrangling Notebook
4. GitHub URL of Completed EDA with Data Visualization Notebook
5. GitHub URL of Completed EDA with SQL Notebook
6. GitHub URL of Completed Interactive Map with Folium
7. nbviewer URL of Completed Interactive Map with Folium
8. GitHub URL of Completed Plotly Dash Lab (Python File)
9. GitHub URL of Completed Predictive Analysis Lab

Helper Functions for Getting Data from SpaceX API

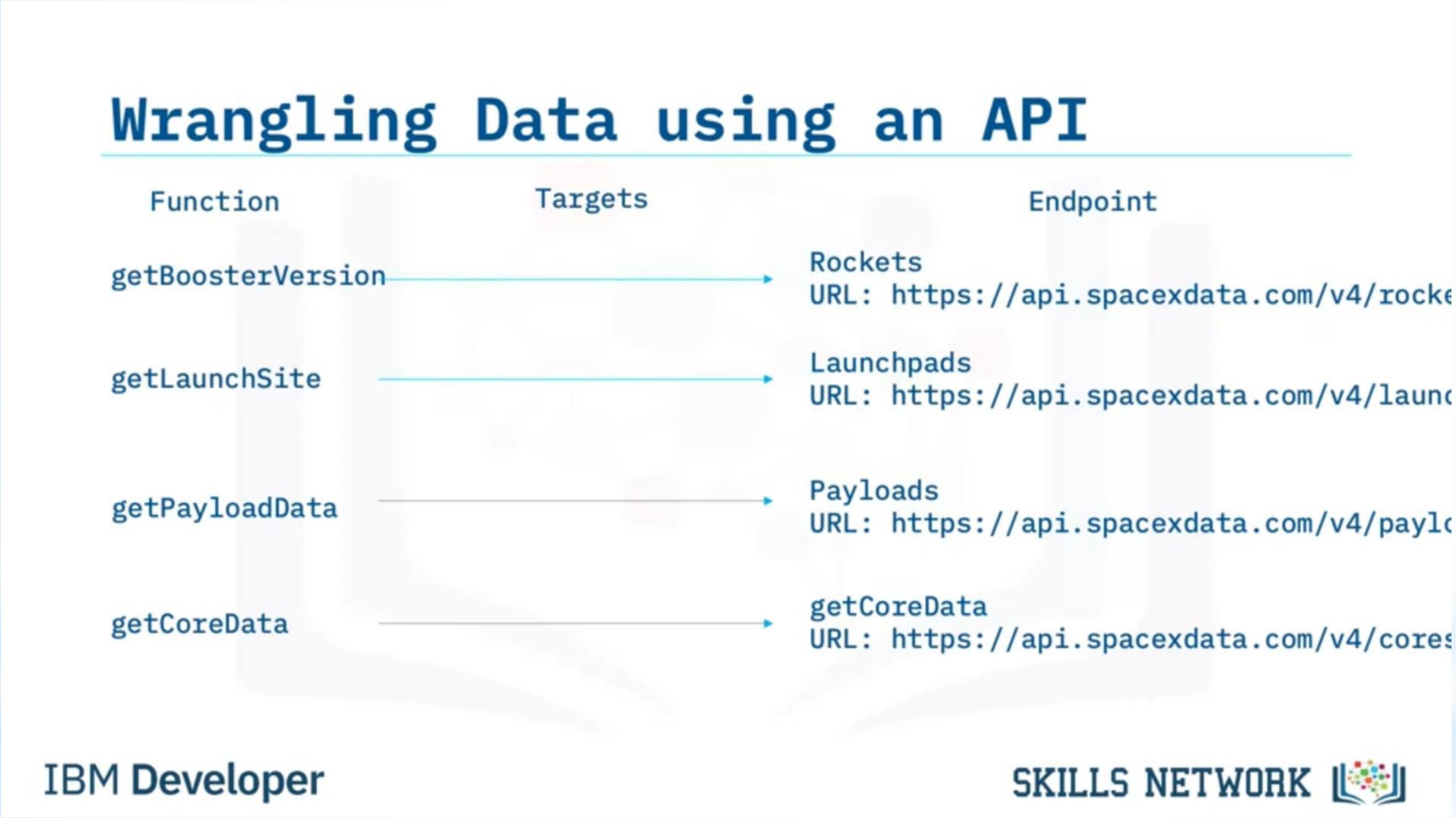


Image 1

Screenshot taken from Applied Data Science Capstone on Coursera[4].

Falcon 9: Stage 1

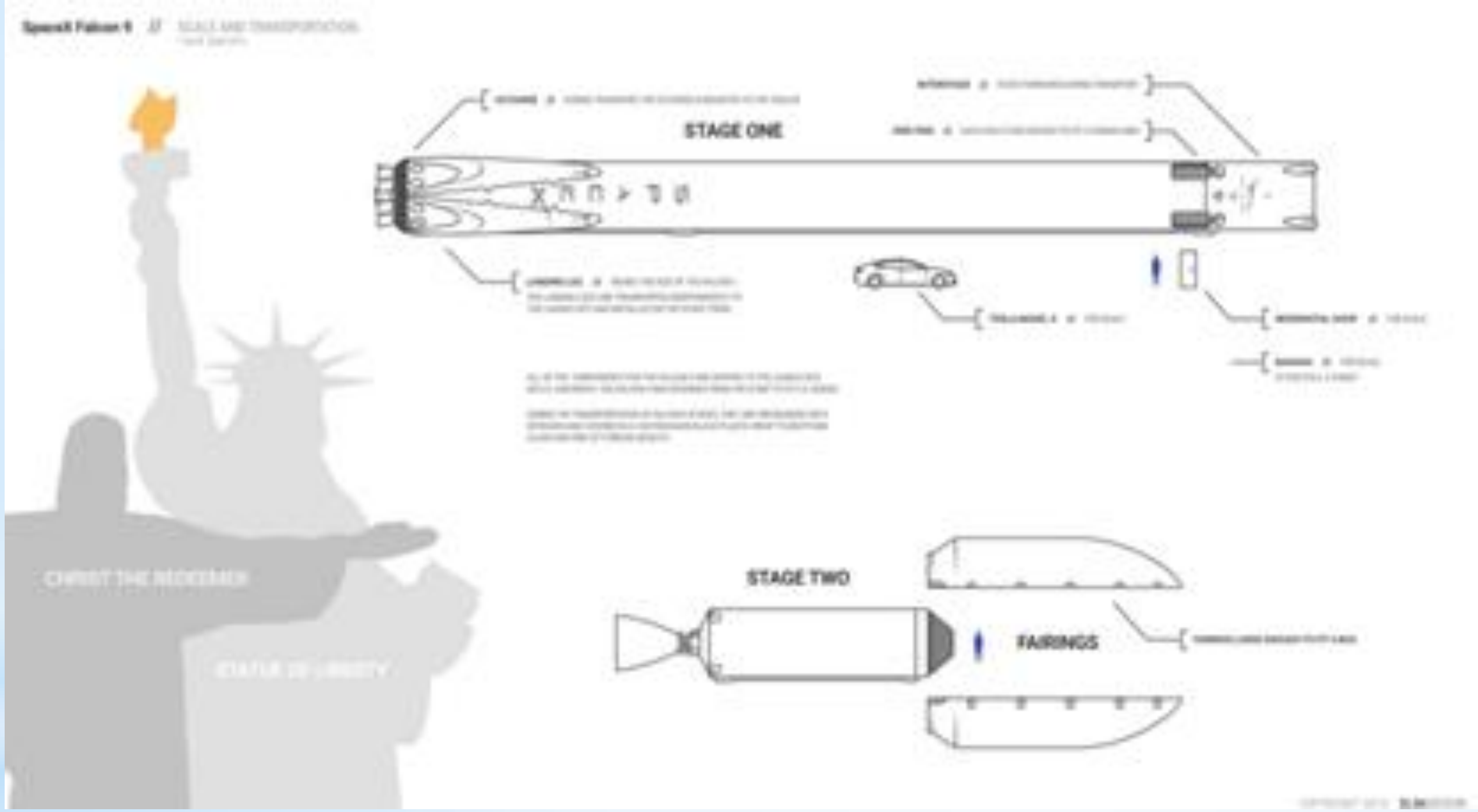


Image 2

* Screenshot illustrating Falcon 9 Stage 1[4].
* The goal of the project was to predict whether Falcon 9 Stage 1 will land successfully, thus reducing launch cost.

Orbits[4] (1/2)

- “**LEO**: Low Earth Orbit (LEO) is an Earth-centered orbit with an altitude of 2,000 km (1,200 mi) or less (approximately one-third of the radius of Earth), or with at least 11.25 periods per day (an orbital period of 128 minutes or less) and an eccentricity less than 0.25. Most of the manmade objects in outer space are in LEO.
- “**VLEO**: Very Low Earth Orbit (VLEO) can be defined as an orbit with a mean altitude below 450 km. Operating in these orbits can provide a number of benefits to Earth observation spacecraft as the spacecraft operates closer to the observation.
- “**GTO**: A Geosynchronous Orbit is a high Earth orbit that allows satellites to match Earth's rotation. Located at 22,236 miles (35,786 kilometers) above Earth's Equator, this position is a valuable spot for monitoring weather, communications and surveillance. Because the satellite orbits at the same speed that the Earth is turning, the satellite seems to stay in place over a single longitude, though it may drift north to south,” NASA wrote on its Earth Observatory website.
- “**SSO (or SO):** A Sun-Synchronous Orbit, also called a heliosynchronous orbit, is a nearly polar orbit around a planet, in which the satellite passes over any given point of the planet's surface at the same local mean solar time.”

Orbits[4] (2/2)

- “**ES-L1**: At the Lagrange points the gravitational forces of the two large bodies cancel out in such a way that a small object placed in orbit there is in equilibrium relative to the center of mass of the large bodies. L1 is one such point between the Sun and the Earth.
- “**HEO**: A highly elliptical orbit, is an elliptic orbit with high eccentricity, usually referring to one around Earth.
- “**ISS**: A modular space station (habitable artificial satellite) in low Earth orbit. It is a multinational collaborative project between five participating space agencies: NASA (United States), Roscosmos (Russia), JAXA (Japan), ESA (Europe), and CSA (Canada).
- “**MEO**: Geocentric orbits ranging in altitude from 2,000 km (1,200 mi) to just below geosynchronous orbit at 35,786 kilometers (22,236 mi). Also known as an intermediate circular orbit. These are most commonly at 20,200 kilometers (12,600 mi), or 20,650 kilometers (12,830 mi), with an orbital period of 12 hours.
- “**HEO**: Geocentric orbits above the altitude of geosynchronous orbit (35,786 km or 22,236 mi).
- “**GEO**: It is a circular geosynchronous orbit 35,786 kilometers (22,236 miles) above Earth's Equator and following the direction of Earth's rotation.
- “**PO**: It is one type of satellite, in which a satellite passes above or nearly above both poles of the body being orbited (usually a planet such as the Earth).”

Orbits Plot

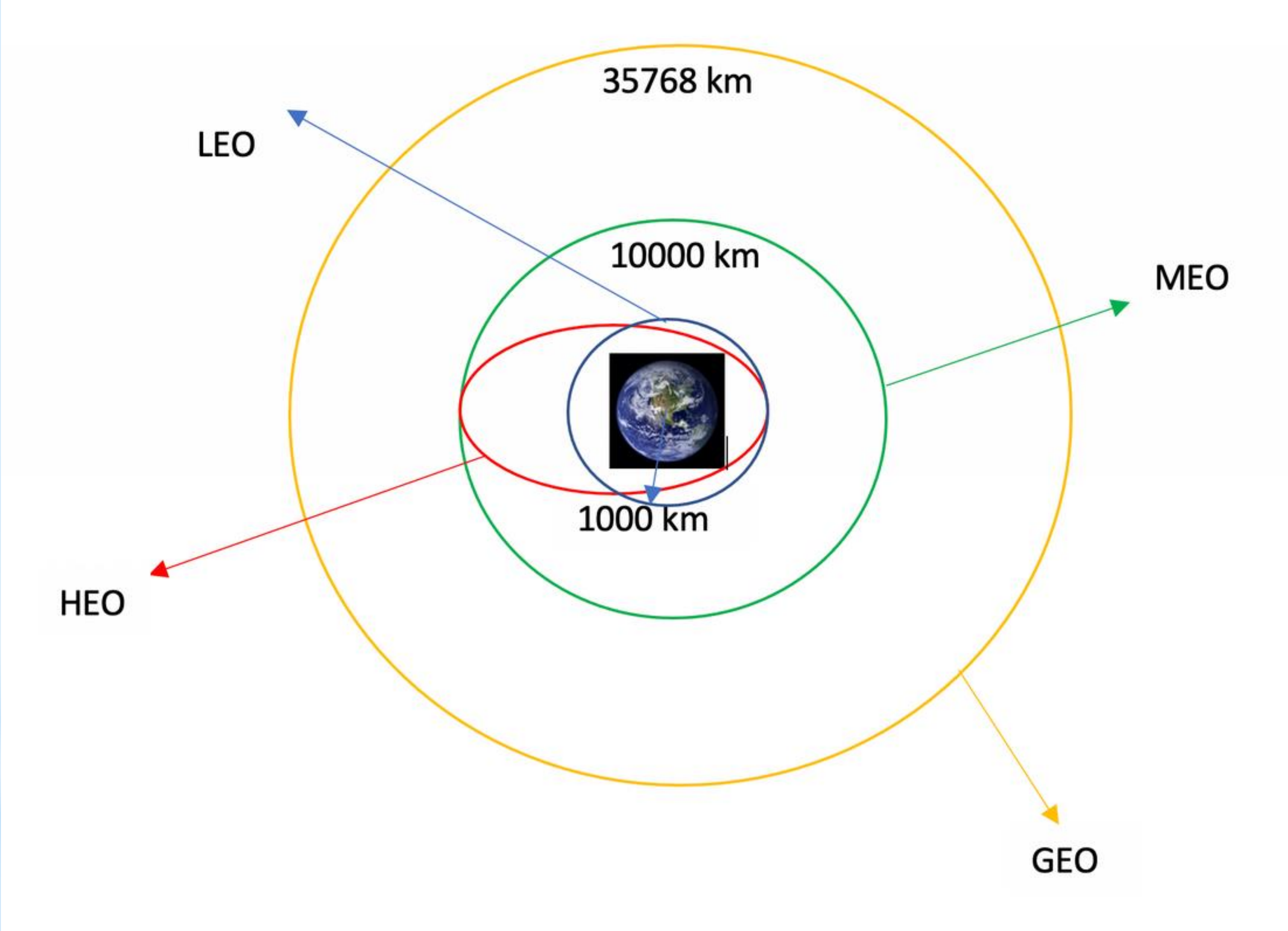


Image 3

Screenshot taken from Applied Data Science Capstone on Coursera[4].

Datasets Used

1. dataset_part_1_data_wrangling.csv (used for data wrangling)
2. Spacex_dataset_SQL.csv (used for SQL queries)
3. dataset_part_2_data_visualization.csv (used for EDA and data visualization)
4. spacex_launch_geo_folium.csv (used for launch site analysis with folium)
5. spacex_launch_dash.csv (used for launch dashboard application)
6. dataset_part_2_ML.csv (used for ML predictive analysis - classification)

Thank you!