

Winning Space Race with Data Science

Rick Schneider Apr 2024



Outline

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

Executive Summary

Summary of methodologies

- Used SpaceX API and Scraped data from Falcon 9 Wiki Page
- Cleaned the data replaced missing payload data with the Payload size mean
- Reviewed the data and created various plots to show relationships that influenced the launch outcome
- Created interactive dashboard to investigate launch success rates and see how payload size influenced the success of the launch
- Created an interactive map to show the launch successes and failures visually, aw well as the proximity to the coastline, highways, railways and citys
- Used 4 models: Logistic Regression, K Nearest Neighbors, Support Vector Machine and Decision Tree. Each were optimized and assessed for accuracy of the prediction. The Decision Tree was the best model with an accuracy of 90%.

Executive Summary continued

Summary of questions

- •Can we predict the success of Falcon 9 Launches? Yes, with 90% accuracy
- •Where should we locate our launch sites? Isolated areas, close to the coastline, close to railway, with reasonable highway access.
 - •Are launch sites in close proximity to railways? Yes, need Railway access for large heavy parts
 - •Are launch sites in close proximity to highways? Somewhat. Need to limit commute time for workers
 - •Are launch sites in close proximity to coastline? Yes, for launch recovery at sea.
 - •Do launch sites keep certain distance away from cities? Yes, to limit damage to population / city in event of launch failure.
- •Are there factors that influence the success rate? Yes, Payload size, Orbit, Booster Version influence the success of the launch.

Introduction

SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch.

Client Questions

- •Can we predict the success of Falcon 9 Launches?
- •Where should we locate our launch sites?
 - •Are launch sites in close proximity to railways?
 - •Are launch sites in close proximity to highways?
 - •Are launch sites in close proximity to coastline?
 - •Do launch sites keep certain distance away from cities?
- •Are there factors that influence the success rate?



Methodology

Executive Summary

- Data collection methodology:
 - Initial launch data was collected through the SpaceX API
 - Falcon 9 Launch data was collected through web scraping
- Performed data wrangling
 - Identified gaps in the payload data and replaced missing values with the Payload Mean
 - Created Class column in the data to show Outcome Success (1) and Failure (0)
- Performed exploratory data analysis (EDA) using visualization and SQL
- Performed interactive visual analytics using Folium and Plotly Dash
- Performed predictive analysis using classification models
 - Looked at Logistic Regression, k Nearest Neighbor (KNN), Support Vector Machine (SVM) and Decision Tree models to see which model was the most accurate at predicting successful missions.

Data Collection

Detailed flowcharts shown in subsequent slides

Initial Launch Data was gathered via the SpaceX API

Specific Falcon 9 Launch information was gathered via Scraping SpaceX web Wiki page

Data Collection – SpaceX API

- Get the Launch Data via the API Calls
- Normalize the JSON response
- Clean the data (replace missing values) Payload with Mean
- Extract Falcon 9 Launch Data
- Reset the Flight Number

 Notebook GitHub URL - Data Collection Get the Booster Version from the Rockets
Column API Call

https://api.spacexdata.com/v4/rockets/

Get the Launch Site and GPS Coordinates from the Launchpads Column API Call

https://api.spacexdata.com/v4/launchpads/

Get the payload mass and orbit from the payloads Column API Call

https://api.spacexdata.com/v4/payloads/

Get the landing outcome, Type of landing, number of flights with core, is gridfins were used, whether the core was reused, and core serial number from the Cores Column API Call

https://api.spacexdata.com/v4/cores/

Data Collection - Scraping

Tool used - BeautifulSoup

Notebook GitHub URL - Web
 Scraping

Get Falcon 9 Launch Data from the Wiki Page https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_a nd_Falcon_Heavy_launches&oldid=1027686922 Extract column / variable names from the table headers Get the Launch details form the 3rd table on the page Create the Data Frame from the Dictionary List

Data Wrangling

Explore data for Missing Values Identify Numerical and categorical Data Columns Count Launches for each site Orbit types Mission outcomes per Orbit type Create list of outcomes (Class Field) - Failure (0) - Success (1) Calculate Falcon 9 success rate = 67%

Notebook GitHub URL - Data Wrangling

EDA with Data Visualization

Plot Type	Selected Visual	Reason
Scatter Plot	Flight Number vs Launch Site with respect to Success	Look at Success / Failure for flights over time at each launch site Show volume of flights per site too.
Scatter plot	Pay Load Mass vs Launch Site	Look at relationship of payload size per launch site Shows distribution of payload size per launch site
Bar chart	Orbit va Success Rate	Look at relationship between Orbit and success rate of mission
Scatter plot	Flight Number vs Orbit with respect to Success	Look at Orbits over time with respect to mission success
Line Chart	Year vs Success Rate	Look at success rate over time

Notebook GitHub URL - EDA SQL

EDA with SQL

SQL	Reason
sql SELECT DISTINCT `Launch_Site` from SPACEXTABLE	Find the Launch Sites in the Data
SELECT * from SPACEXTABLE where `Launch_Site` like 'CCA%' limit 5	Look at sample data to understand the types of data available
%sql SELECT sum(PAYLOAD_MASSKG_) from SPACEXTABLE where `Customer` like 'NASA (CRS)'	Look at the payload Total for a specific Customer - NASA (CRS)
%sql SELECT avg(PAYLOAD_MASSKG_) from SPACEXTABLE where `Booster_Version` like 'F9 v1.1%'	Look at the Average payload mass for Falcon 9 V1.1% type booster
%sql SELECT min(Date) from SPACEXTABLE where `Landing_Outcome` = 'Success (ground pad)'	Look at the first success of the ground pad landing
%sql SELECT Booster_Version, Landing_Outcome,PAYLOAD_MASSKG_ from SPACEXTABLE where `Landing_Outcome` = 'Success (drone ship)' and (PAYLOAD_MASSKG_ > 4000 and PAYLOAD_MASSKG_ < 6000)	Look at the Booster Version Success for Drone Ship Landings for Payloads between 4000 and 6000 KG
%sql SELECT `Mission_Outcome`, count(`Mission_Outcome`) from SPACEXTABLE GROUP BY `Mission_Outcome`	Look at Mission Outcomes
%sql SELECT Booster_Version from SPACEXTABLE where PAYLOAD_MASSKG_ = (Select max(PAYLOAD_MASSKG_) from SPACEXTABLE) order by Booster_Version desc	Which Booster Versions carried the maximum payload
%sql SELECT substr ("JanFebMarAprMayJunJulAugSepOctNovDec", substr(Date, 6,2) * 3, 3) as 'Month', Landing_Outcome, Booster_Version,Launch_Site from SPACEXTABLE where substr(Date,0,5)='2015' and Landing_Outcome = 'Failure (drone ship)'	Drone Ship landing Failures in 2015
%sql SELECT `Landing_Outcome`, count(Landing_Outcome) as 'Count' from SPACEXTABLE where ((Date < '2017-03-20') and (Date > '2010-06-04')) GROUP BY `Landing_Outcome` Order by count(Landing_Outcome) desc	Ranking the landing outcomes between 2017-03-20 and 2010-06-04

Interactive Map Visualizations with Folium

Folium Map Items	Reason
Labeled the launch Sites on the world Map	Easily locate the launch sites
Blue Circle for each site	Note the location on the map
Cluster marker with red and green icons (drill down) to note success and failed missions	Visually see the volume of launches and their status
Distance markers – City, Highway and railway	Note the distance in KM to nearest item to show positioning of launch site with respect to those items for planning purposes for competing sites

Plotly Dash Visualization

Plotly Interactive Visuals	Purpose
Pie Chart of successful launches per launch site	Visual to see the distribution of the launches across the various launch sites which is the dominate launch site
Pie Chart Drill down per Site	Look at the success rate for each launch site
Scatter plot of Payload par launch site with respect to success or failure of the mission	Look at the relationship between Payload Size and the success or failure of the mission per launch site
Payload Slider	Focus in on a particular range of payload sizes

Predictive Analysis (Classification)

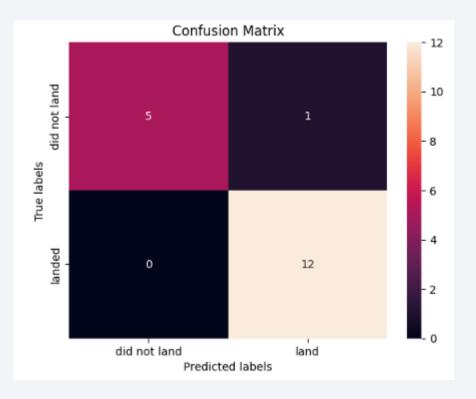
Decision Tree Parameters that made up the best Prediction accuracy.

```
{'criterion': 'gini',
 'max_depth': 14,
 'max_features': 'sqrt',
 'min_samples_leaf': 4,
 'min_samples_split': 5,
 'splitter': 'random'}
```

Create a Numpy Pandas series of the Class array (the info we want to predict). Standardize the data to be used in the model Split the data into test and training samples For each model (logistic regression, KNN, Decision Tree and Support Vector Machine) Create the model object Run the model with a GridSearch Fit the model - find the best params Check the model accuracy Check the confusion matrix predictions

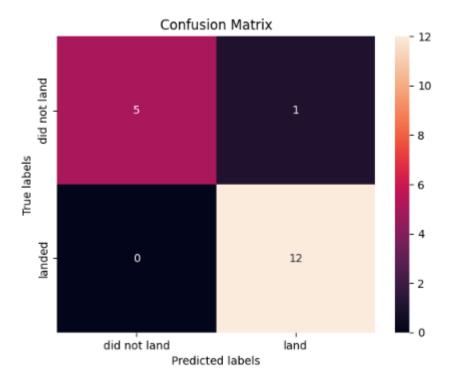
Logistic Regression Results

• Model Accuracy - 0.82222222222222



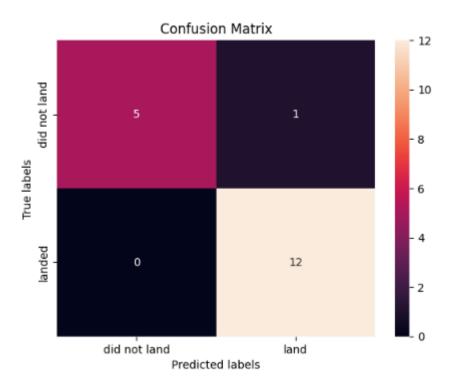
Support Vector Machine Results

• Model Accuracy - 0.8222222222223

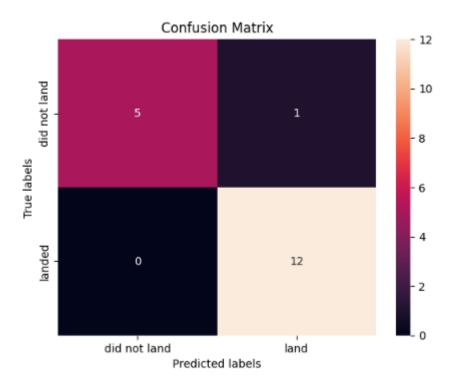


Decision Tree Results

• Model Accuracy - 0.9



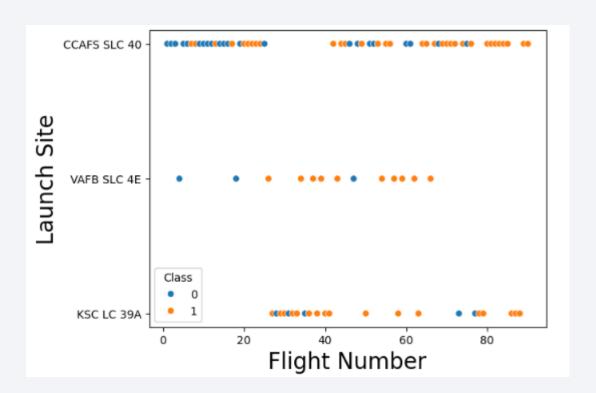
K Nearest Neighbors Results





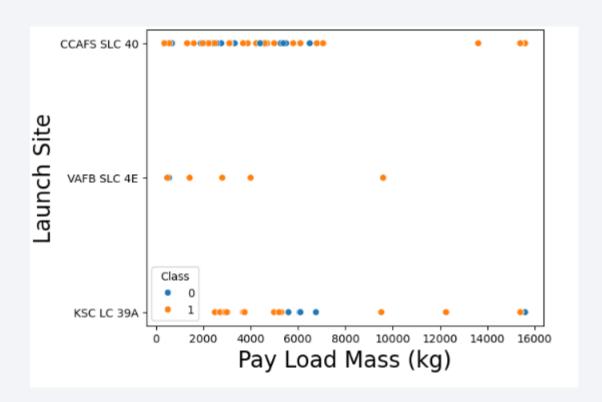
Flight Number vs. Launch Site

- Success (1 Orange)
- Failure (O Blue)
- Higher Flight numbers are more recent
- Higher success rate over time



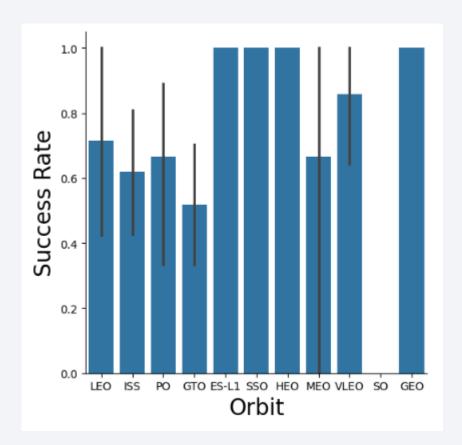
Payload vs. Launch Site

- CCAFS SLC-40 has the most launches, VAFB SLC-4E has the least launches
- Least number of Failures at VAFB SLC 4E



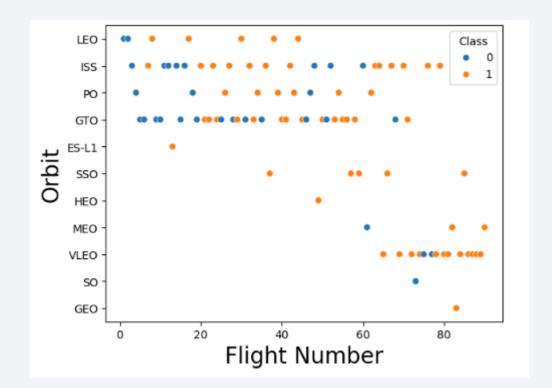
Success Rate vs. Orbit Type

- Success of 1 = 100%
- 4 Orbits with 100% Success
 Rate



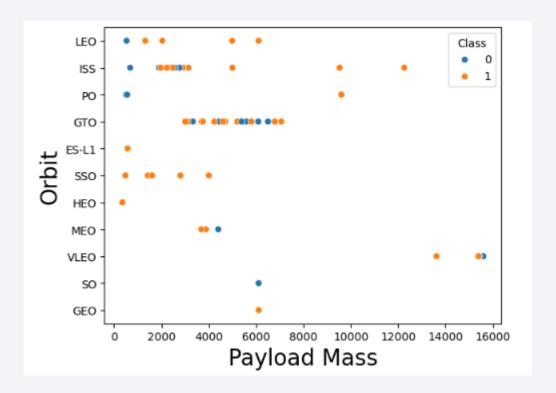
Flight Number vs. Orbit Type

- The success rate increases over time (Recent Flights have a higher Flight Number)
- High Failure rate early on



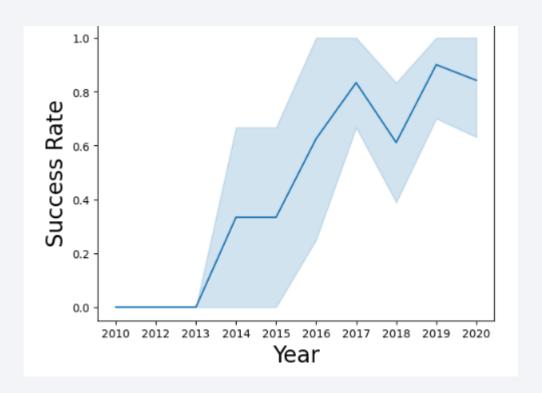
Payload vs. Orbit Type

- The scatter plot shows the Mission Success (1 – Orange) and Failure (0 – Blue) for each Orbit Type with respect to Payload Mass
- Orbits with Payloads over 6000KG were all Successful with one exception (VLEO Orbit)



Launch Success Yearly Trend

- The chart shows increased success over time, with the exception of a drop in the success rate in 2018.
- Could be explained by the number of landings not attempted... see appendix 3



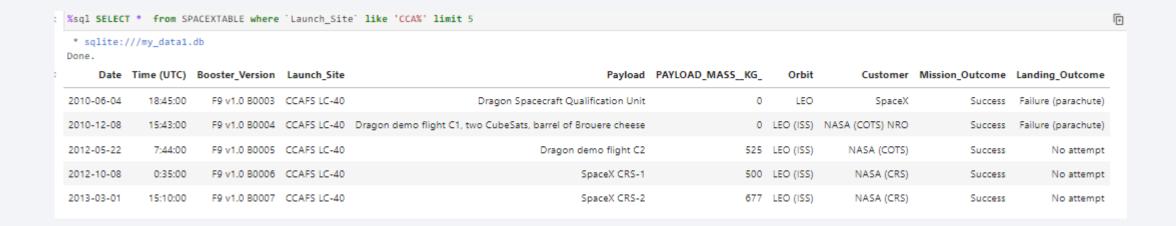
All Launch Site Names

• Find the unique launch sites in the data

```
%sql SELECT DISTINCT `Launch_Site` from SPACEXTABLE
 * sqlite:///my_data1.db
Done.
 Launch_Site
 CCAFS LC-40
 VAFB SLC-4E
  KSC LC-39A
CCAFS SLC-40
```

Launch Site Names Begin with 'CCA'

Quick look at the data for sample site(s)



Total Payload Mass

The total Payload Mass for NASA (CRS) is 45596KG

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

%sql SELECT sum(PAYLOAD_MASS__KG_) from SPACEXTABLE where 'Customer' like 'NASA (CRS)'

* sqlite://my_data1.db
Done.

sum(PAYLOAD_MASS__KG_)

45596
```

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1* (all Variations) is 2534.67KG
- The query for explicit F9 v1.1 Booster is commented out if that was what is needed.

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

"sql SELECT avg(PAYLOAD_MASS_KG_) from SPACEXTABLE where 'Booster_Version' like 'F9 v1.1%'

#%sql SELECT avg(PAYLOAD_MASS_KG_) from SPACEXTABLE where 'Booster_Version' Like 'F9 v1.1'

#%sql SELECT Booster_Version, PAYLOAD_MASS_KG_ from SPACEXTABLE where 'Booster_Version' Like 'F9 v1.1%'

* sqlite://my_datal.db
Done.

|: avg(PAYLOAD_MASS_KG_)

2534.666666666666665
```

First Successful Ground Landing Date

• The first successful ground pad landing outcome was Dec 22, 2015

```
%sql SELECT min(Date) from SPACEXTABLE where `Landing_Outcome` = 'Success (ground pad)'
  * sqlite://my_data1.db
Done.
  min(Date)
2015-12-22
```

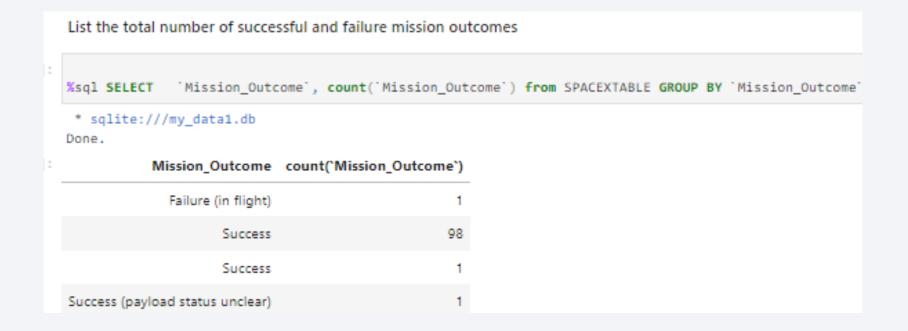
Successful Drone Ship Landing with Payload between 4000 and 6000

4 Successes with the right payload mass

				(drone ship)' and (PAYLOAD_M (drone ship)' and (PAYLOAD_MA	
* sqlite:///my_da	ta1.db				
Booster_Version La	anding_Outcome	PAYLOAD_MASSKG_			
F9 FT B1022 Suc	ccess (drone ship)	4696			
F9 FT B1026 Suc	ccess (drone ship)	4600			
F9 FT B1021.2 Suc	ccess (drone ship)	5300			
F9 FT B1031.2 Suc	ccess (drone ship)	5200			

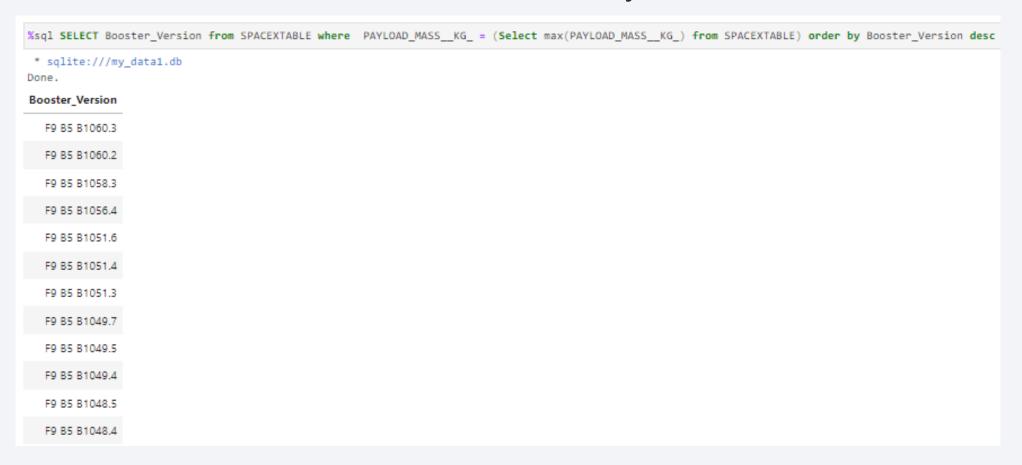
Total Number of Successful and Failure Mission Outcomes

- Table shows that there could be an issue with the Success Mission Outcome (special character in the field that makes the 1 Success stand out from the 98)
- 100 Successes vs 1 Failure



Boosters Carried Maximum Payload

12 Boosters Versions carried the Maximum Payload



2015 Launch Records

• 2 Failed Landings in 2015

%sql SELECT substr ("--JanFebMarAprMayJunJulAugSepOctNovDec", substr(Date, 6,2) * 3, 3) as 'Month', Landing_Outcome, Booster_Version,Launch_Site from SPACEXTABLE where substr(Date,0,5)='2015' and Landing Outcome = 'Failure (drone ship)'

* sqlite:///my_data1.db Done.

Month Landing_Outcome Booster_Version Launch_Site

Jan Failure (drone ship) F9 v1.1 B1012 CCAFS LC-40

Apr Failure (drone ship) F9 v1.1 B1015 CCAFS LC-40

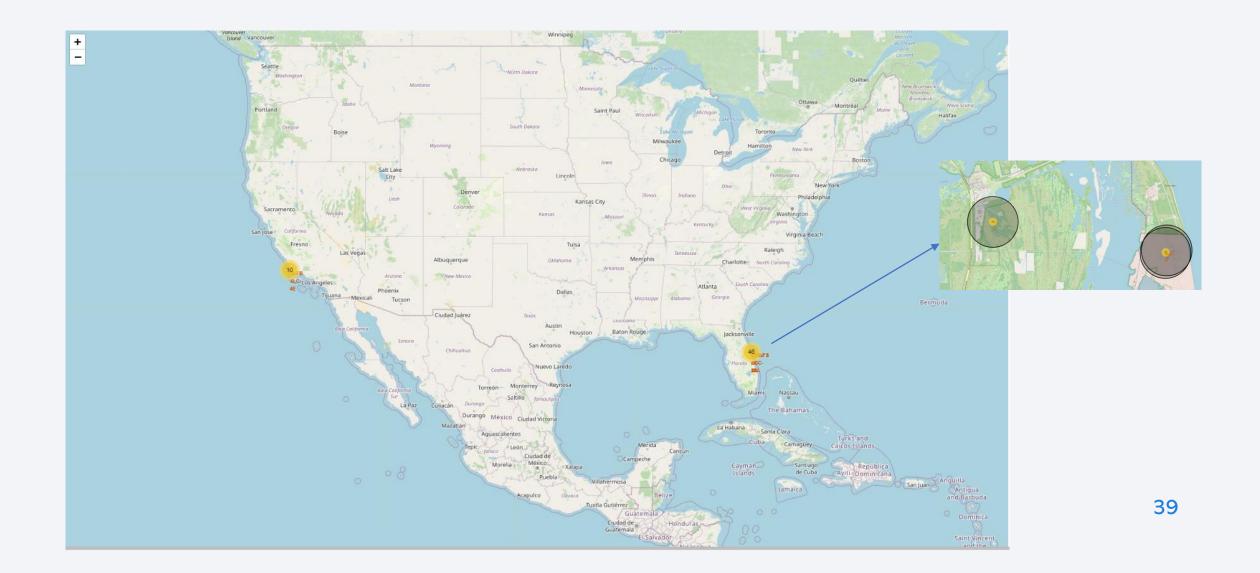
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

8 outcomes in the date range requested



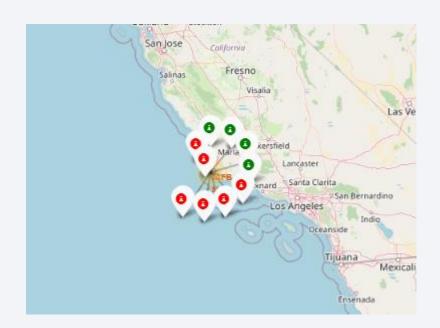


SpaceX Launch Sites – with expanded view for FL Sites



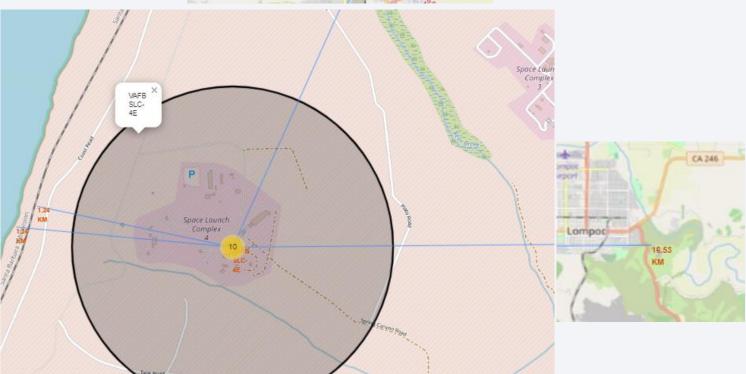
Launch Site Details

- Colored Icons show Successful Outcome (green), Failed outcome (Red)
- CA Launch site has 4 Green and 6 Red



Map with distance markers (exploded view)



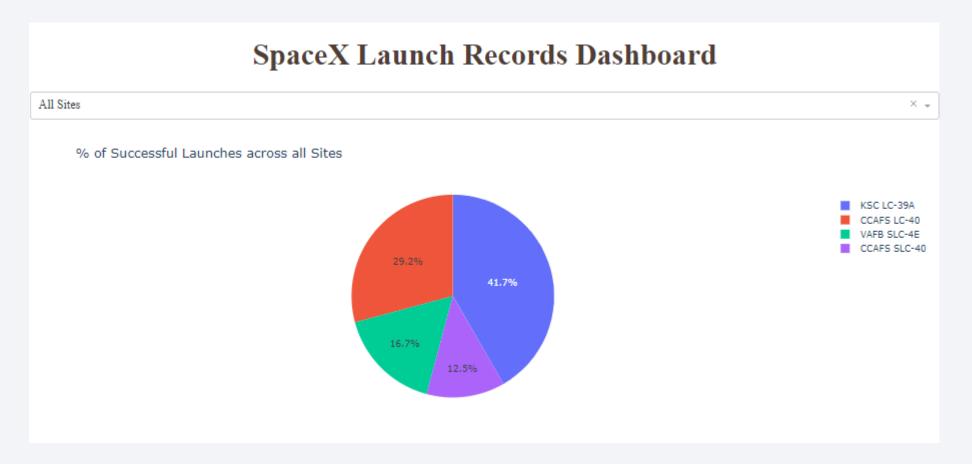


- Distance markers to
 - Coast
 - Highway
 - City
 - Railway



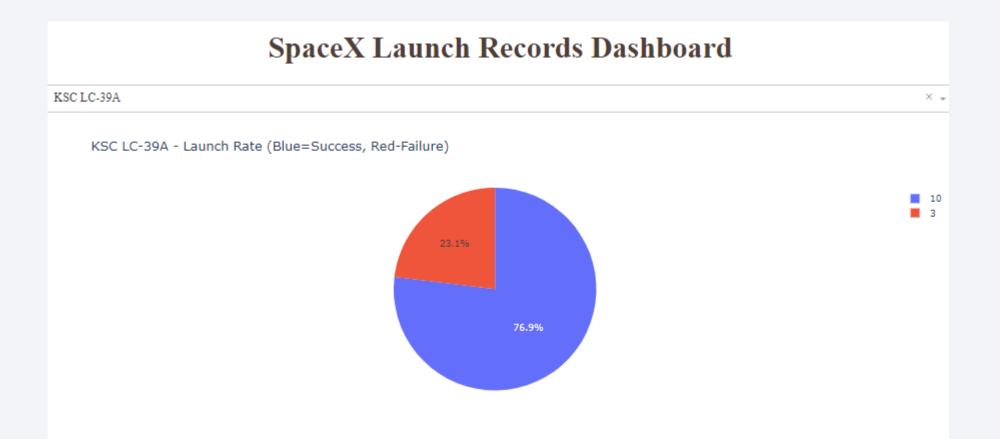
% of Successful launches across all launch site

• Easy view of which site has the most successful launches



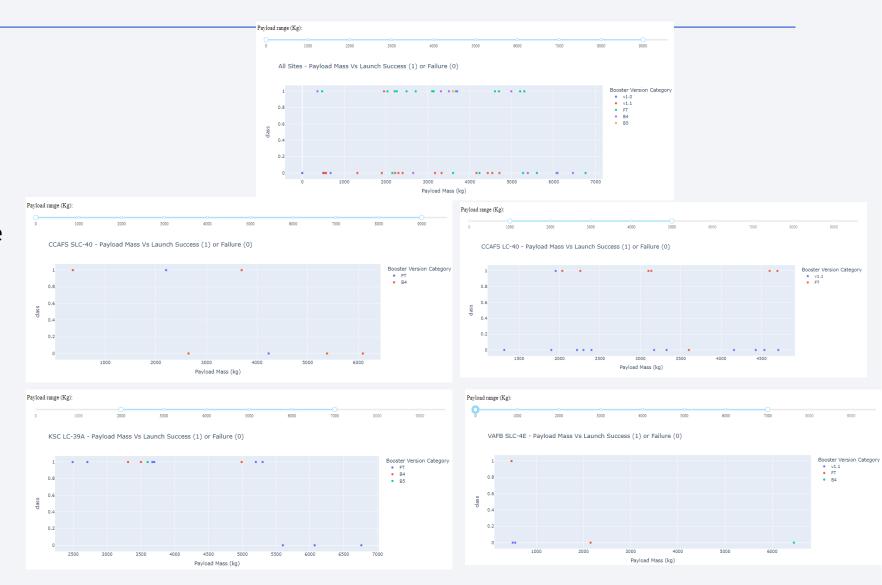
Site Success Rate Drilldown

- Drill down of the success rate at a selected site
- KSC LC-39A is the most successful Launch Site



Payload Mass vs Success or Failure by Booster Version

- VAFB SLC-40 has the least number of launches. Only one Success
- KSC LC-39A has failures above 5500 KG
- CCAFS SLC-40 has failures above 4000 KG
- CCAFS LC-40 has all but 1
 Successes with Booster V FT,
 while all but one Failure with
 Booster V1.1
- Payload range Slider will vary the view of the scatter plot for better focus





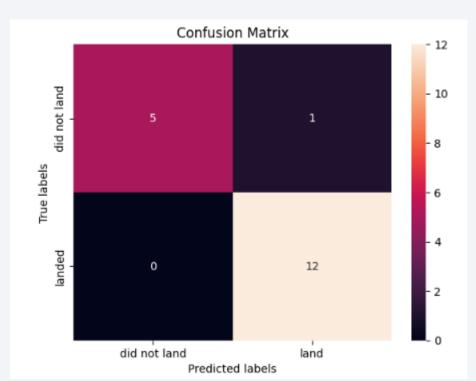
Classification Accuracy

 The Decision Tree Model has the best Model Accuracy of 90%



Confusion Matrix

- For the confusion matrix the Decision Tree Model:
 - Correctly identified 5 did not land and 12 land
 - Gave a false positive for 1 predicted landing
 - Did not predict any false negatives



Conclusions

- Using the Decision Tree model we can predict whether the Falcon 9 mission is successful with a 90% accuracy rate
- KSC LC-39A has the best Outcome Success Rate (76.9%). 100% for payloads under 5500KG... 0% for Payloads over 5500KG
- Launch site locations are best suited to be close to the coast (to recover landings at sea), isolated from populated areas (limit damages from failed launches), close to railway (to get parts), and somewhat near a highway (worker access).
- SpaceX has improved the success rate for launches over time. Most of the failures were early on (low Flight Numbers)
- 100% Success rate for Orbits: SSO, GEO, HEO, ES-L1 (see Appendix 2 for definitions)

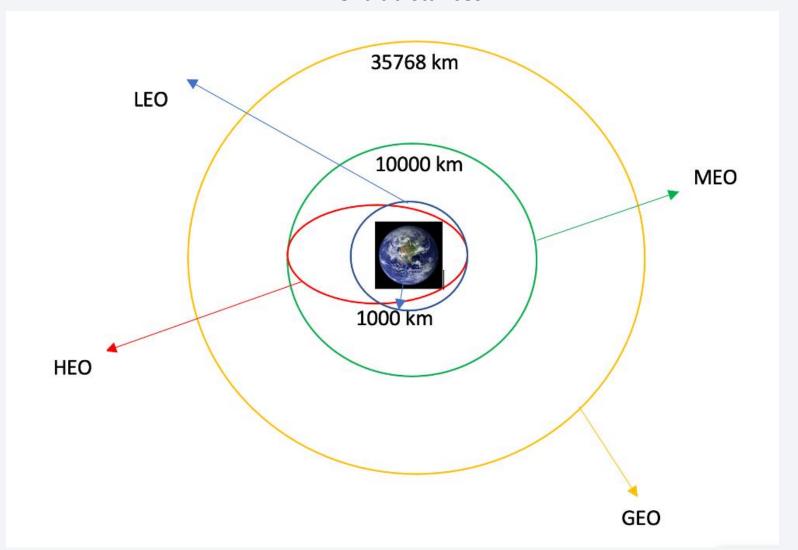
Conclusions - continued

- •Can we predict the success of Falcon 9 Launches? Yes, with 90% accuracy
- •Where should we locate our launch sites? Isolated areas, close to the coastline, close to railway, with reasonable highway access.
 - •Are launch sites in close proximity to railways? Yes, need Railway access for large heavy parts
 - •Are launch sites in close proximity to highways?

 Somewhat. Need to limit commute time for workers
 - •Are launch sites in close proximity to coastline? Yes, for launch recovery at sea.
 - •Do launch sites keep certain distance away from cities? Yes, to limit damage to population / city in event of launch failure.
- •Are there factors that influence the success rate? Yes, Payload size, Orbit, Booster Version influence the success of the launch.

Appendix

Orbit distances



Appendix 2

Orbit definitions

Each launch aims to an dedicated orbit, and here are some common orbit types:

- LEO: Low Earth orbit (LEO)is an Earth-centred orbit with an altitude of 2,000 km (1,200 mi) or less (approximately one-third of the radius of Earth),[1] or with at least 11.25 periods per day (an orbital period of 128 minutes or less) and an eccentricity less than 0.25.[2] Most of the manmade objects in outer space are in LEO [1].
- VLEO: Very Low Earth Orbits (VLEO) can be defined as the orbits 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[2].
- 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 [3].
- SSO (or SO): It is 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 [4].
- 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 [5].
- HEO A highly elliptical orbit, is an elliptic orbit with high eccentricity, usually referring to one around Earth [6].
- 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) [7]
- 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 [8]
- HEO Geocentric orbits above the altitude of geosynchronous orbit (35,786 km or 22,236 mi) [9]
- GEO It is a circular geosynchronous orbit 35,786 kilometres (22,236 miles) above Earth's equator and following the direction of Earth's rotation [10]
- . PO It is one type of satellites in which a satellite passes above or nearly above both poles of the body being orbited (usually a planet such as the Earth [11]

Appendix 3 – Landing Outcomes for 2018

Drop in Success Rate for 2018 could be attributed to the Launches where no landing attempt was made

Month	Landing_Outcome	Booster_Version	Launch_Site
Jan	Success (ground pad)	F9 B4 B1043.1	CCAFS SLC-40
Jan	Controlled (ocean)	F9 FT B1032.2	CCAFS SLC-40
Feb	No attempt	F9 FT B1038.2	VAFB SLC-4E
Mar	No attempt	F9 B4 B1044	CCAFS SLC-40
Mar	No attempt	F9 B4 B1041.2	VAFB SLC-4E
Apr	No attempt	F9 B4 B1039.2	CCAFS SLC-40
Apr	Success (drone ship)	F9 B4 B1045.1	CCAFS SLC-40
May	Success (drone ship)	F9 B5 B1046.1	KSC LC-39A
May	No attempt	F9 B4 B1043.2	VAFB SLC-4E
Jun	No attempt	F9 B4 B1040.2	CCAFS SLC-40
Jun	No attempt	F9 B4 B1045.2	CCAFS SLC-40
Jul	Success	F9 B5B1047.1	CCAFS SLC-40
Jul	Success	F9 B5B1048.1	VAFB SLC-4E
Aug	Success	F9 B5 B1046.2	CCAFS SLC-40
Sep	Success	F9 B5B1049.1	CCAFS SLC-40
Oct	Success	F9 B5 B1048.2	VAFB SLC-4E
Nov	Success	F9 B5 B1047.2	KSC LC-39A
Dec	Success	F9 B5 B1046.3	VAFB SLC-4E
Dec	Failure	F9 B5B1050	CCAFS SLC-40
Dec	No attempt	F9 B5B1054	CCAFS SLC-40

