IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Vitto Russo Munarriz
17th of October 2021

Outline

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

Executive Summary

In this report we explore the data collection process via web scrapping online repository, Wikipedia, and API integration directly with SpaceX's data source.

Feature engineering is used to discover the most valuable information that would benefit the creation of a model capable of predicting if a launch is successful or not before countdown.

Multiple algorithms were explored and optimized for the best results possible. Two algorithms gave a great results with 94% accuracy with out of sample data.

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.

To be able to competitively compete and bid against SpaceX a determining factor is being able to predict the proposed cost of SpaceX's bid. Since their cost is mayorly affected if the launch successful or not it is valuable to predict this before hand.

This information can be used if an alternate company wants to bid against SpaceX for a rocket launch. In this report we explore the methodology, exploration, and results to overcome this challenge.

Section 1

Methodology

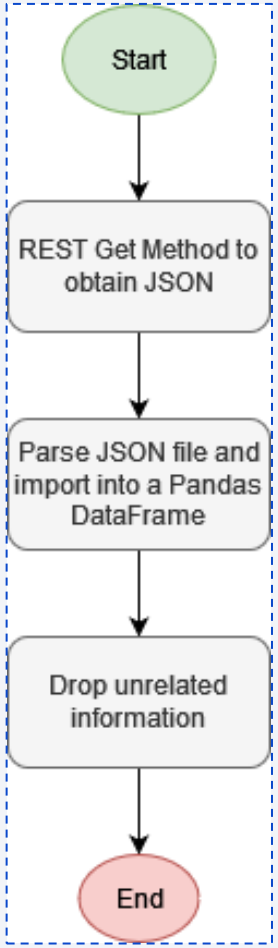
Methodology

Executive Summary

- Data collection methodology:
  - Two main methods were used to collect all the relevant data needed to initiate data exploration and modelling.
    - Web Scrapping SpaceX's historical data extracted from Wikipedia.
    - API integration directly with SpaceX's web service.
- Perform data wrangling
  - Basic data cleansing to remove missing, or incoherent data and data type correction.
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Feature engineering is used to find the best hyper parameter to optimize results on four different models that were evaluated.

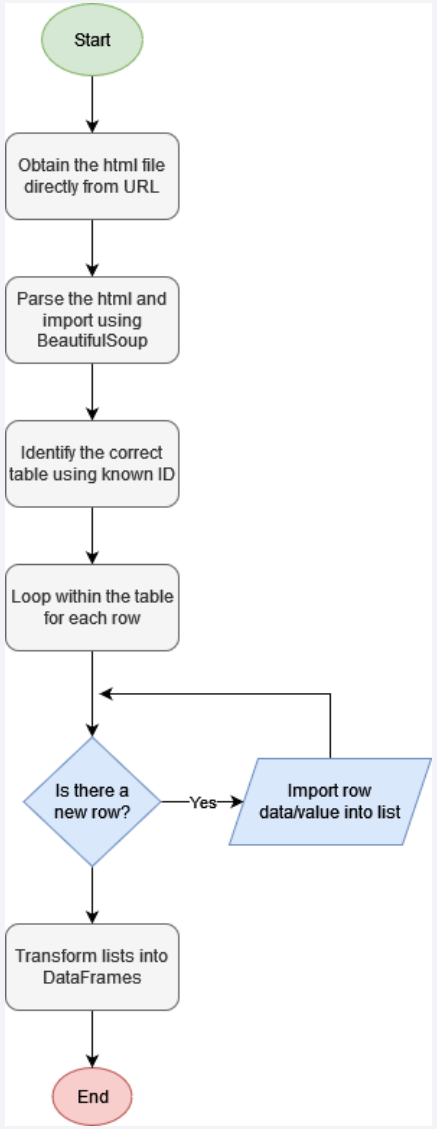
Data Collection – SpaceX API

- SpaceX has a **REST API** to obtain information about their launches that contains data about fairings, crew, rocket, capsules, payload, etc. Including **success status**.
- Using the GET method we extracted the information into a Pandas DataFrame and dropped any unrelated information.
- Github Link: https://bit.ly/2XoZwva



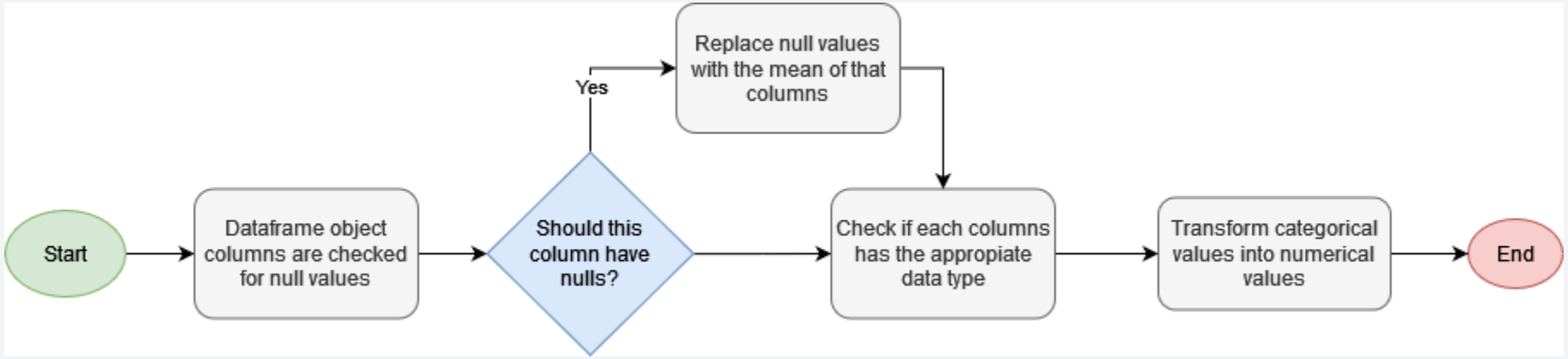
Data Collection - Scraping

- Pythons response library is used to download the **html file** into python to parse using BeautifulSoup.
- The data is looped through using the id html tags to filter information and import into **Pandas Dataframes**.
- GitHub Link: https://bit.ly/3BSYdU6



Data Wrangling

- Basic data wrangling technics are used to clear values from potencial problems that could show up later in the process of modelling of feature engineering. The main look outs that were filtered were: Null values, Correct data types, Convert categorical values, Simply values
- GitHub link: https://bit.ly/3vtx7kg



EDA with Data Visualization

- Using data visualization methods such as scatter plots, bar plots, and line graphs we could start to grasp a basic understanding of the correlation and importance of each feature.
- Some of the graphs we looked into include:
  - Payload vs Flight Number
  - Launch Site vs Flight Number
  - Launch Site vs Payload Mass
  - Success rate vs Orbit
  - Orbit vs Flight Number
  - Orbit vs Payload Mass
  - Success Rate vs Years
- The main conclusion of these exercise is the correlation and weight found between launch site, payload mass and orbit vs the success rate of the mission.
- GitHub Link: https://bit.ly/3n6GeDF

EDA with SQL

- Using the SQL engine we could obtain interesting analysis on the information we had extracted previously.
- Insights:
  - The first successful landing on a ground pad was achieved in 2015
  - Only 4 booster versions have successfully landed with an initial payload mas between 4000 and 6000 kgs
  - Since the historical data has been recollected there have been 38 successful launches
- Github Link: https://bit.ly/3BM5zJ7

Build an Interactive Map with Folium

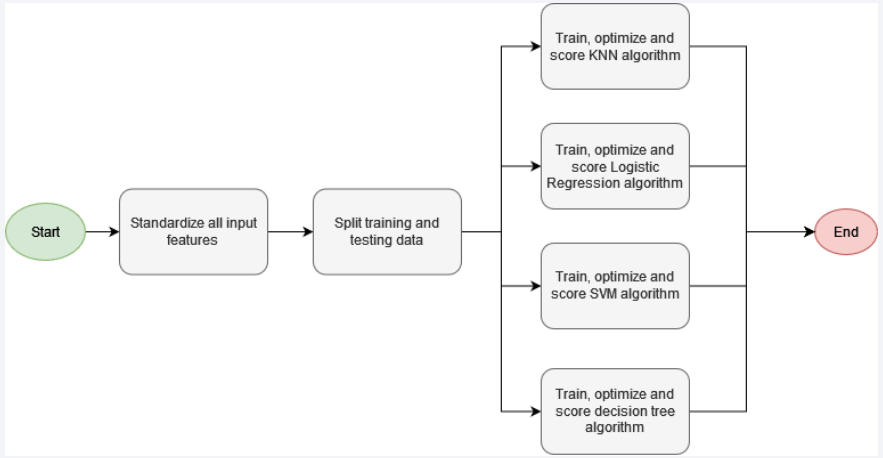
- Using the Folium plugins we built an interactive map including markers, lines and circles to provide an easier understanding of the information.
- To have a context of the area we mapped all the launches to their geographical location. An interesting insight is that both KSC LC-39A and CCAFS SLC-40 are relatively close together.
- Also is the distance to both the closest mayor city and the shore which are approximate 71km and 8km of distance away.
- GitHub Link: https://bit.ly/3pcABpW

Build a Dashboard with Plotly Dash

- The dashboard built displayed 2 main plots, a pie chart and a scatter plot.
- Both plots could be changed interactively with a dropdown menu listing all the launch sites and an option with all launches displayed and secondly a range slicer to change the payload mass range.
- A dashboard, and specifically this dashboard, is a easy and intuitive way to discover cleaver insights like the range in which launches have a higher success rate and which launches are better than other.
- GitHub Link: https://bit.ly/2Z7o2Sd

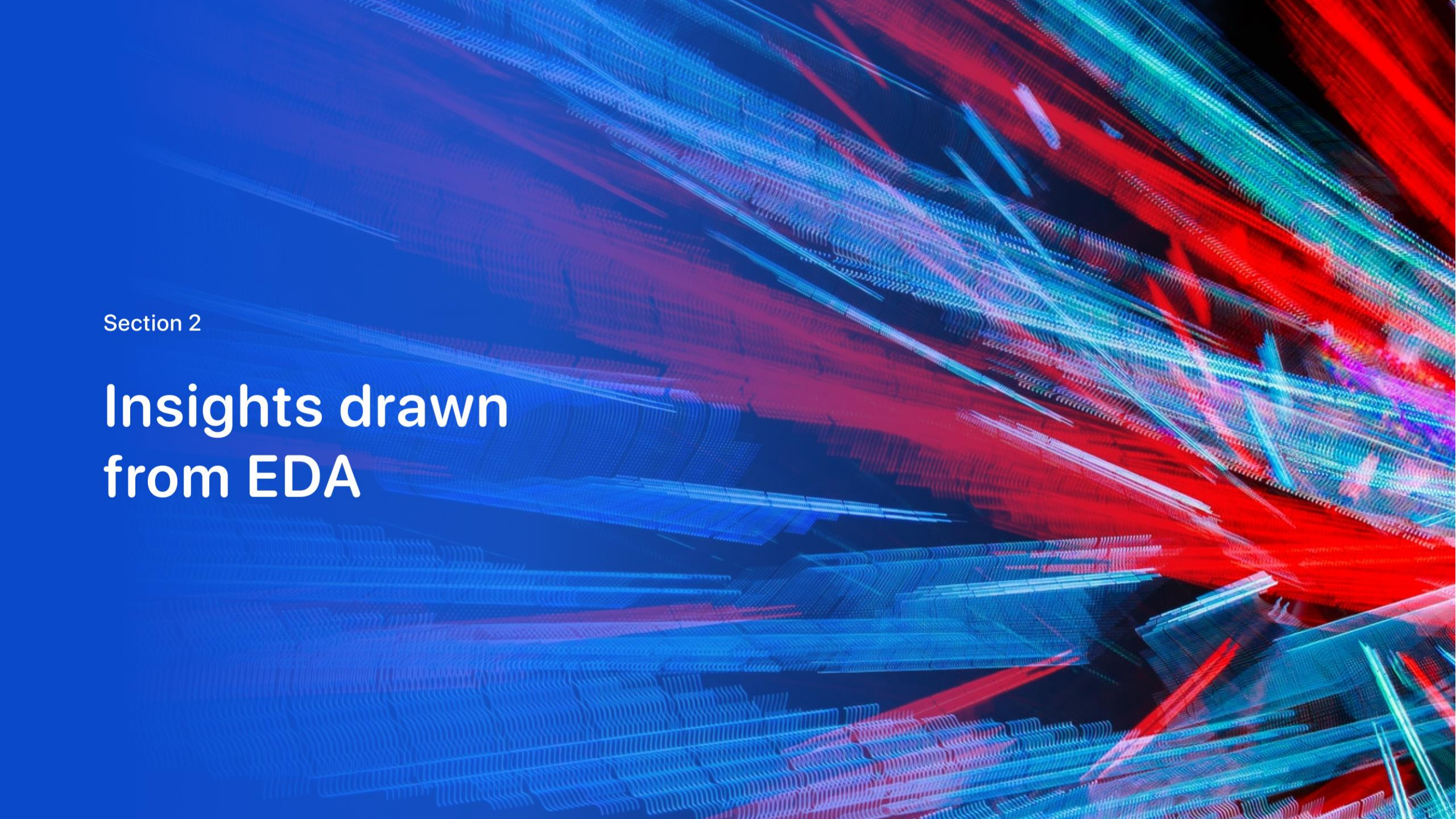
Predictive Analysis (Classification)

- For some given algorithms, in specific classification algorithms, categorical input values are not accepted, for that reason categorical values are translated into numerical values. These translated values despair with numerical values like payload mass and cause bias in the algorithm weight.
- In this section we standardized all input values, trained 4 different models and obtained the most efficient values of the hyper parameters for each one.
- GitHub Link: https://bit.ly/3jcGlXx



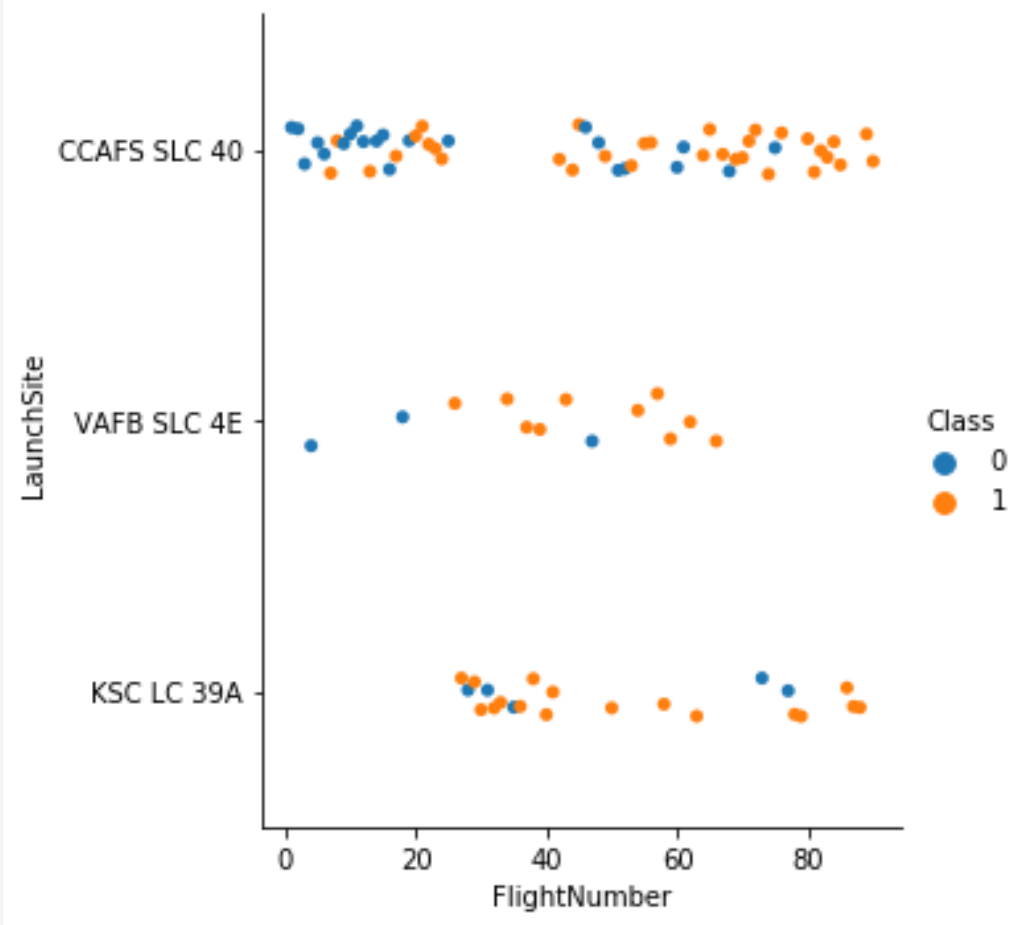
Section 2

Insights drawn from EDA



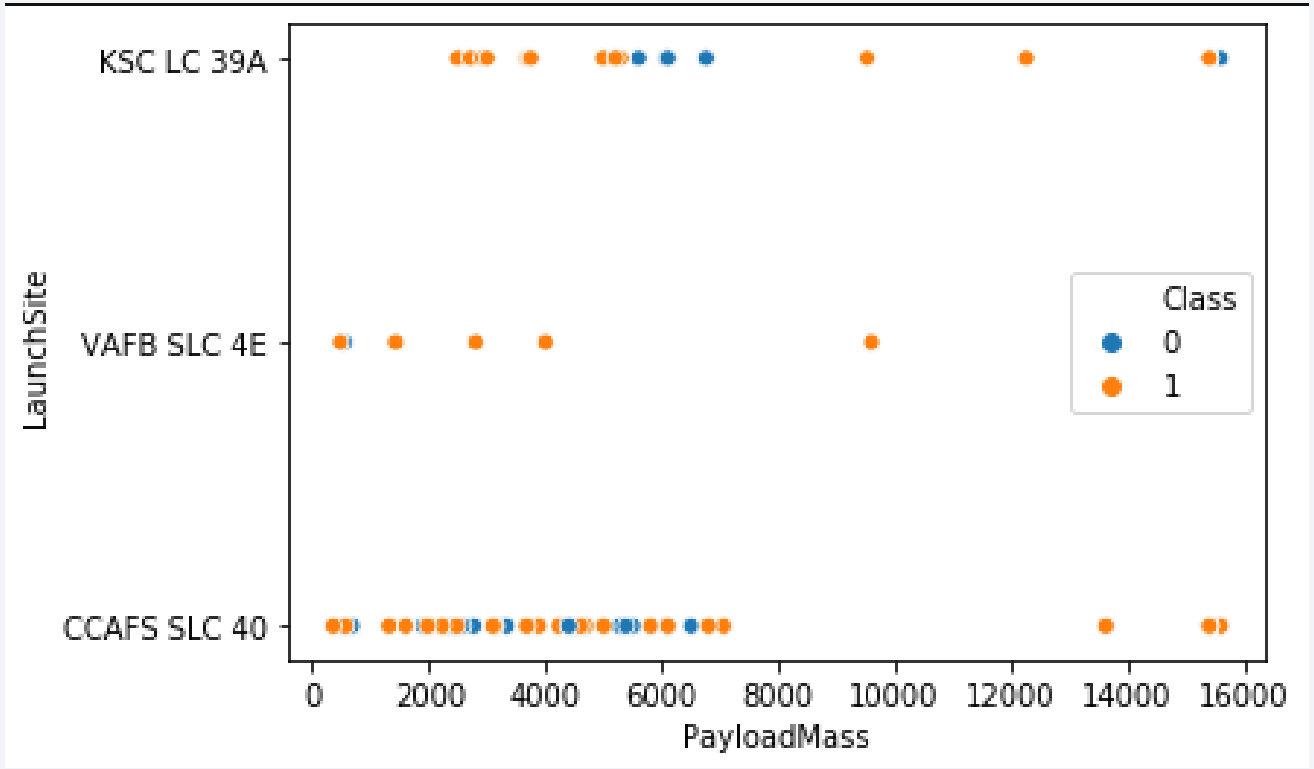
Flight Number vs. Launch Site

- During the EDA using data visualization Flight Number vs Launch Site resulted in the conclusion that “KSC LC 39A” and “VAFB SLC 4E” were the highest success rate of launches.



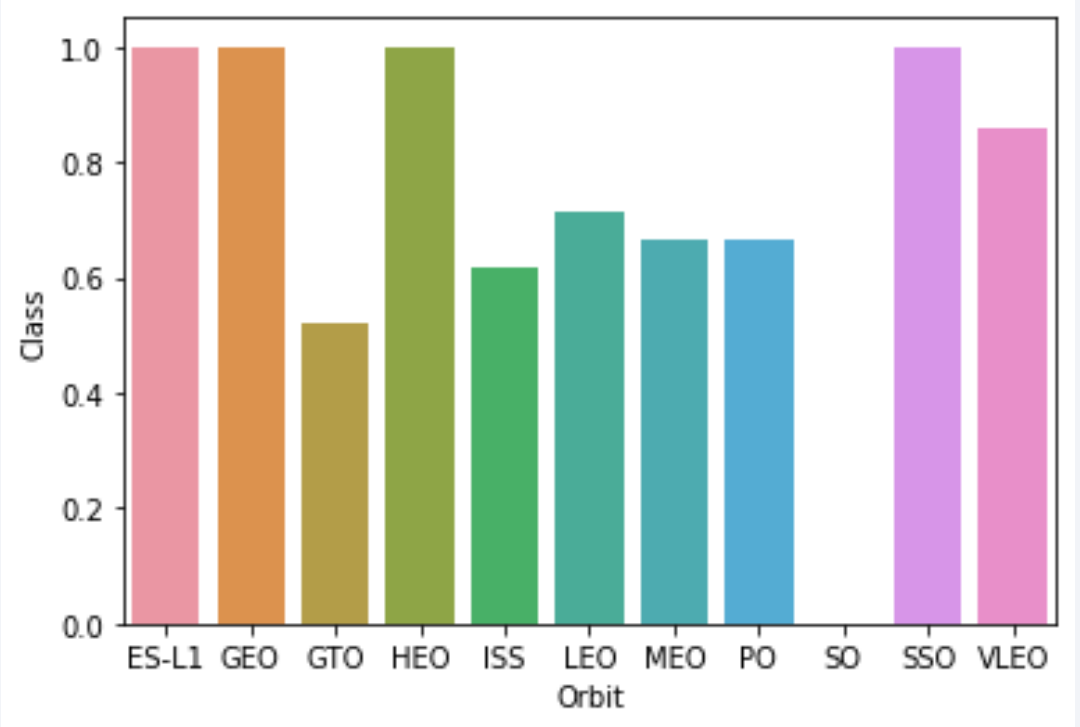
Payload vs. Launch Site

- An interesting trend with launches with more than 7500Kg of payload mass tend to be for successful in general.
- But a concentration of failed launches can be seen in both the first and third launch site.



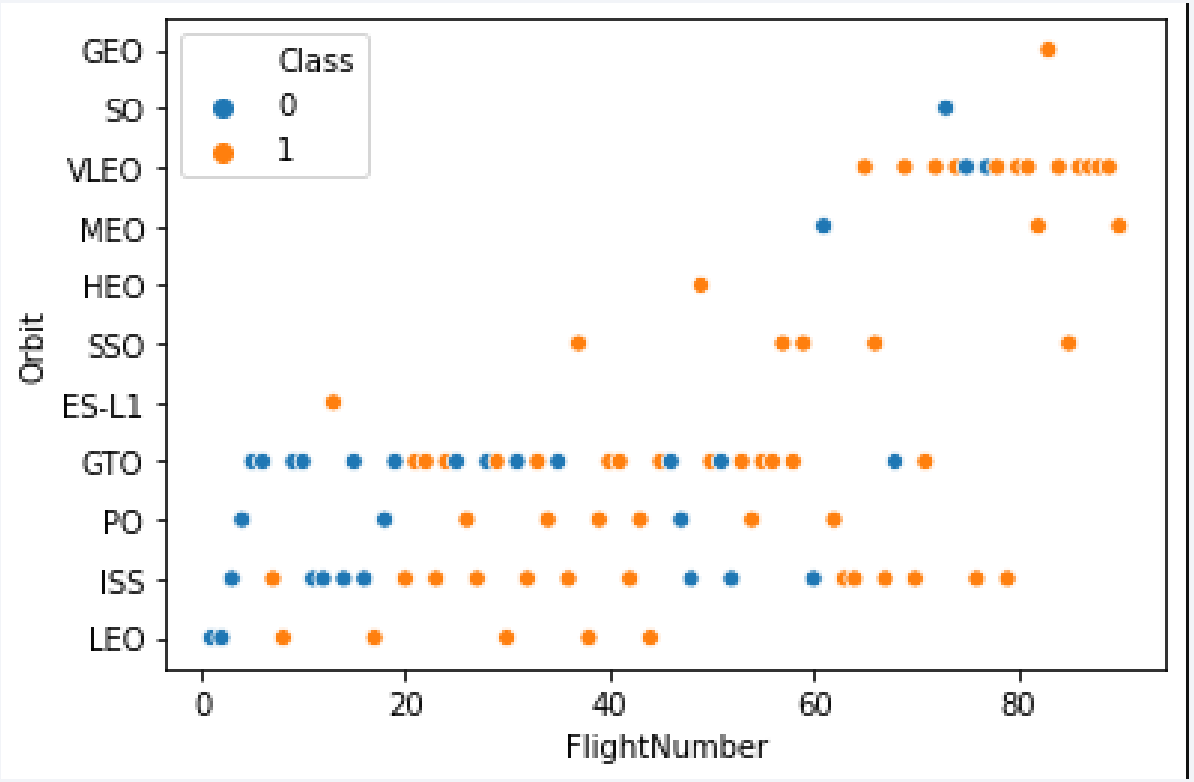
Success Rate vs. Orbit Type

- For ES-L1, GEO, HEO and SSO orbits all have a 100% success rate for launches



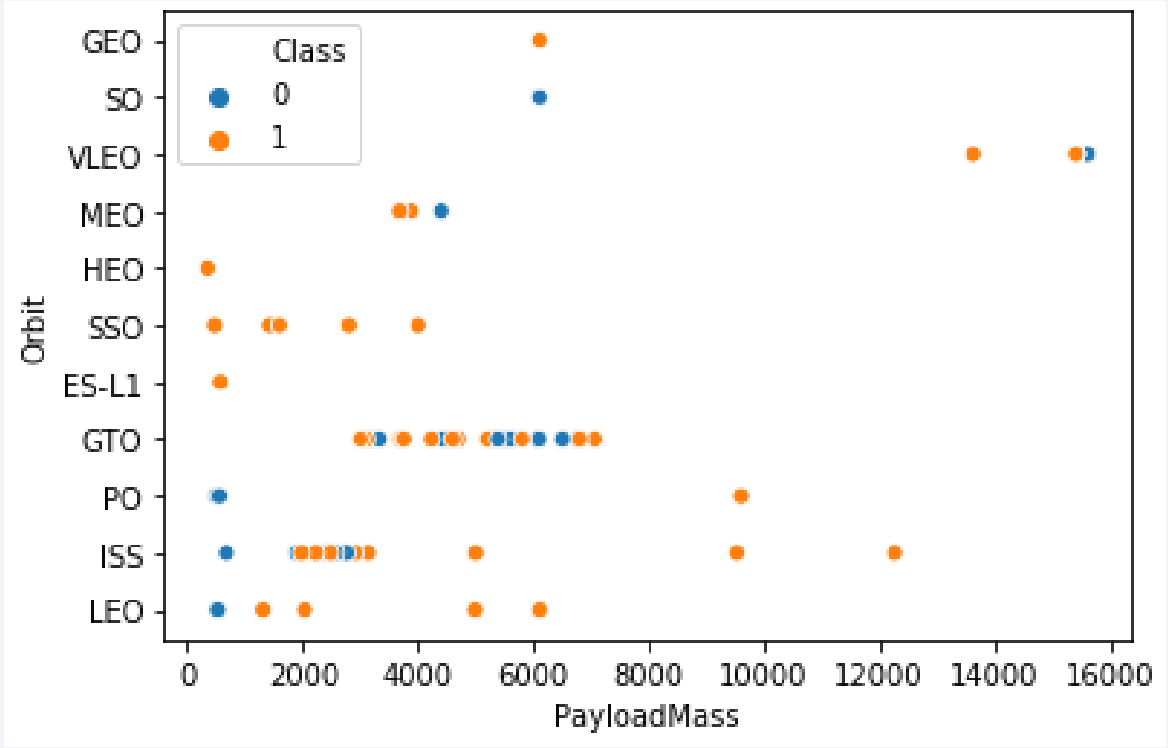
Flight Number vs. Orbit Type

- In the scatter plot you can observe that the 100% success rate also have a low number of launches as well as a tend of new orbits most recently



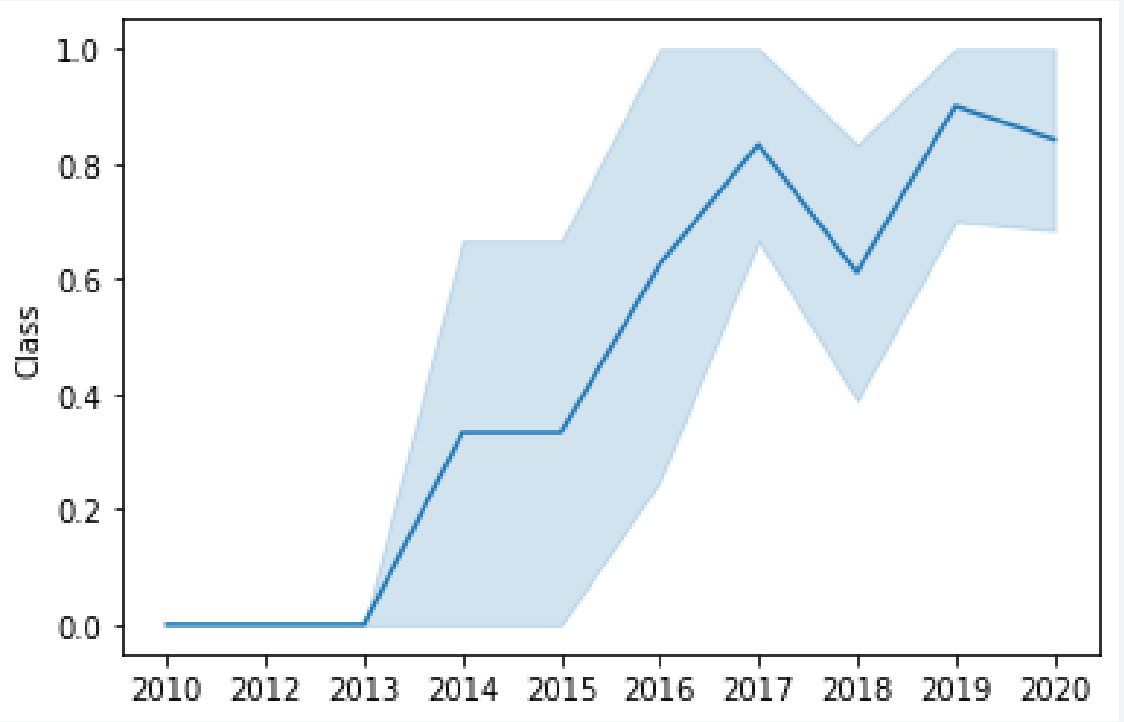
Payload vs. Orbit Type

- In the scatter plot of orbits most of the high success rate orbits have a payload mass on the lower side of the scale.



Launch Success Yearly Trend

- As well as the analysis we are doing during this report SpaceX is also optimizing launches year after year and that shows in the increase every year of successful launches.



All Launch Site Names

- In total SpaceX has launched from 4 unique launch sites inside the USA.

Launch Site Names Begin with 'CCA'

- It appears that the launch sites that start with CCA is the launch site that were used more frequently in early days in which landing wasn't even attempted.

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

Total Payload Mass

- NASA, one of the main clients of SpaceX, has since its inception launched more than 45 tons of payload mass into space with SpaceX's rockets.

Average Payload Mass by F9 v1.1

- The F9 v1.1, one of SpaceX's initial rockets has averaged a total payload mass per launch of 2928 Kgs

average_payload_mass
2928

First Successful Ground Landing Date

- SpaceX's first ground pad successful launch was accomplished in 2015 on launch pad CCAFS LC-40 going to a LEO orbit for the customer Orbcomm.

DATE	time_utc_	booster_version	launch_site	payload	payload_mass_kg_	orbit	customer	mission_outcome	landing_outcome
2015-12-22	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

Successful Drone Ship Landing with Payload between 4000 and 6000

- In total only 4 booster versions have successfully landed in a droneship with a payload mass between 4000 and 6000 kgs.

booster_version
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

Total Number of Successful and Failure Mission Outcomes

- Inherently much, SpaceX is either known to have a successful landing or don't attempt a landing since it is the first 2 most common results of the landing outcomes.

landing_outcome	2
Success	38
No attempt	22
Success (drone ship)	14
Success (ground pad)	9
Controlled (ocean)	5
Failure (drone ship)	5
Failure	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Boosters Carried Maximum Payload

- In total 12 booster version have been launched into space with the mas payload mass ever attempted. This is the list of booster versions.

booster_version
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

2015 Launch Records

- Only 2 failed landings occurred in 2015 and while on 1 successfully landed the other were controlled or were not attempted.

landing__outcome	booster_version	launch_site	DATE
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	2015-01-10
Controlled (ocean)	F9 v1.1 B1013	CCAFS LC-40	2015-02-11
No attempt	F9 v1.1 B1014	CCAFS LC-40	2015-03-02
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	2015-04-14
No attempt	F9 v1.1 B1016	CCAFS LC-40	2015-04-27
Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40	2015-06-28
Success (ground pad)	F9 FT B1019	CCAFS LC-40	2015-12-22

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

- Between 2010 and 2017 the success rate of SpaceX is more than 50% while having an equal 5 failed and successful attempts of landing on a drone ship.

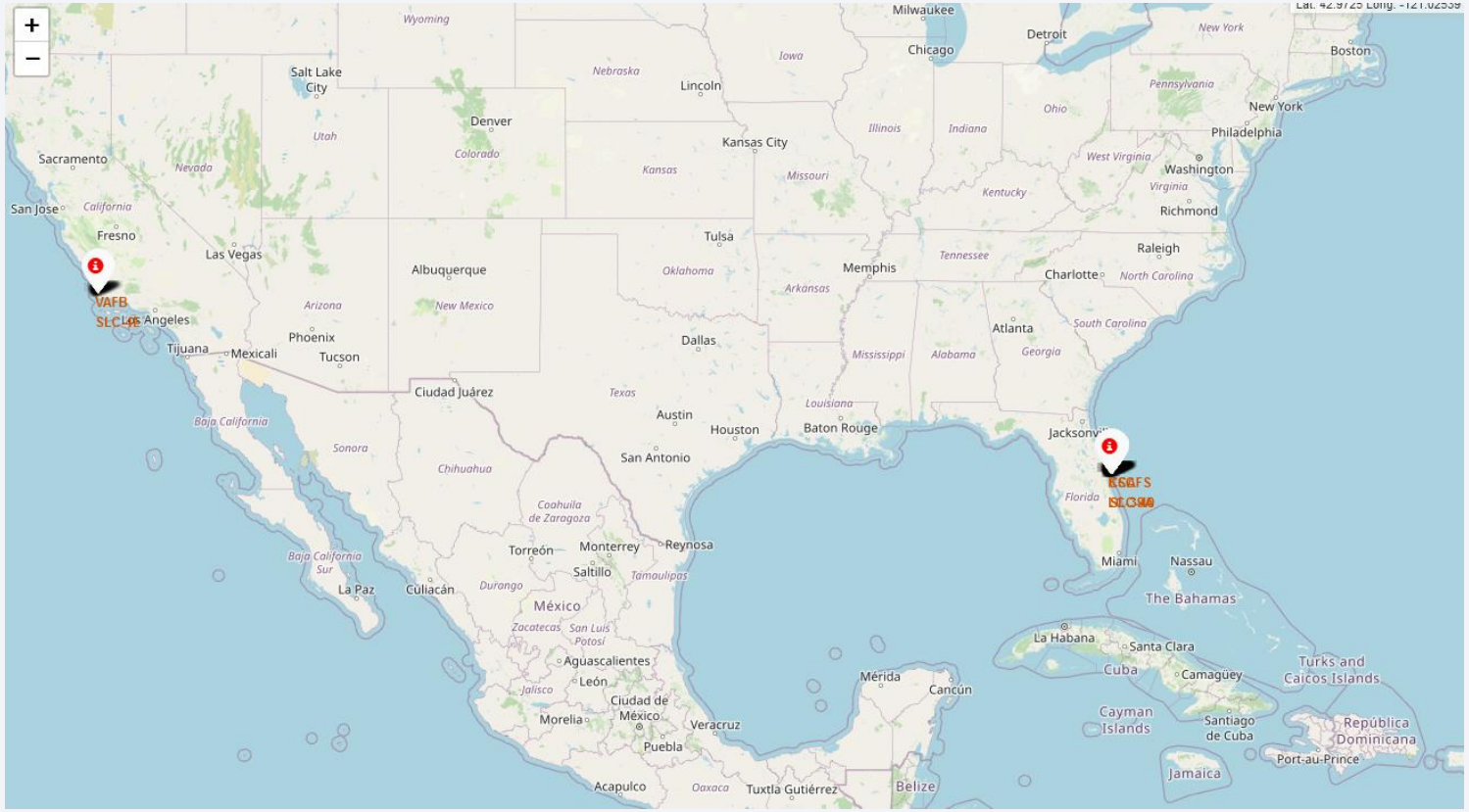
landing_outcome	2
Failure (drone ship)	5
Success (drone ship)	5
Success (ground pad)	3
Failure (parachute)	2

Section 4

Launch Sites Proximities Analysis

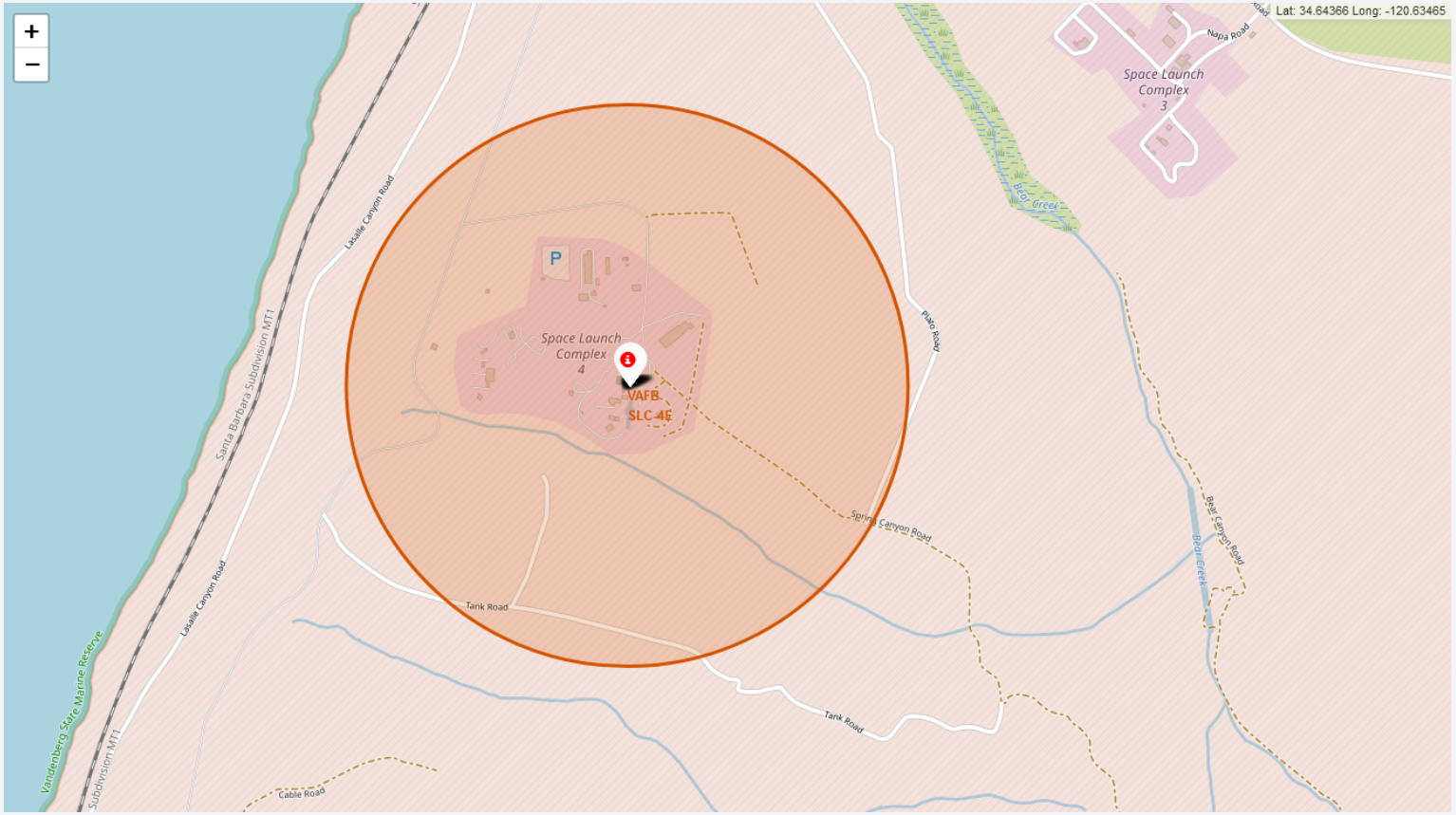
USA Map with SpaceX launch sites

- SpaceX concur with at least one launch site in either east or west coast of the united states.



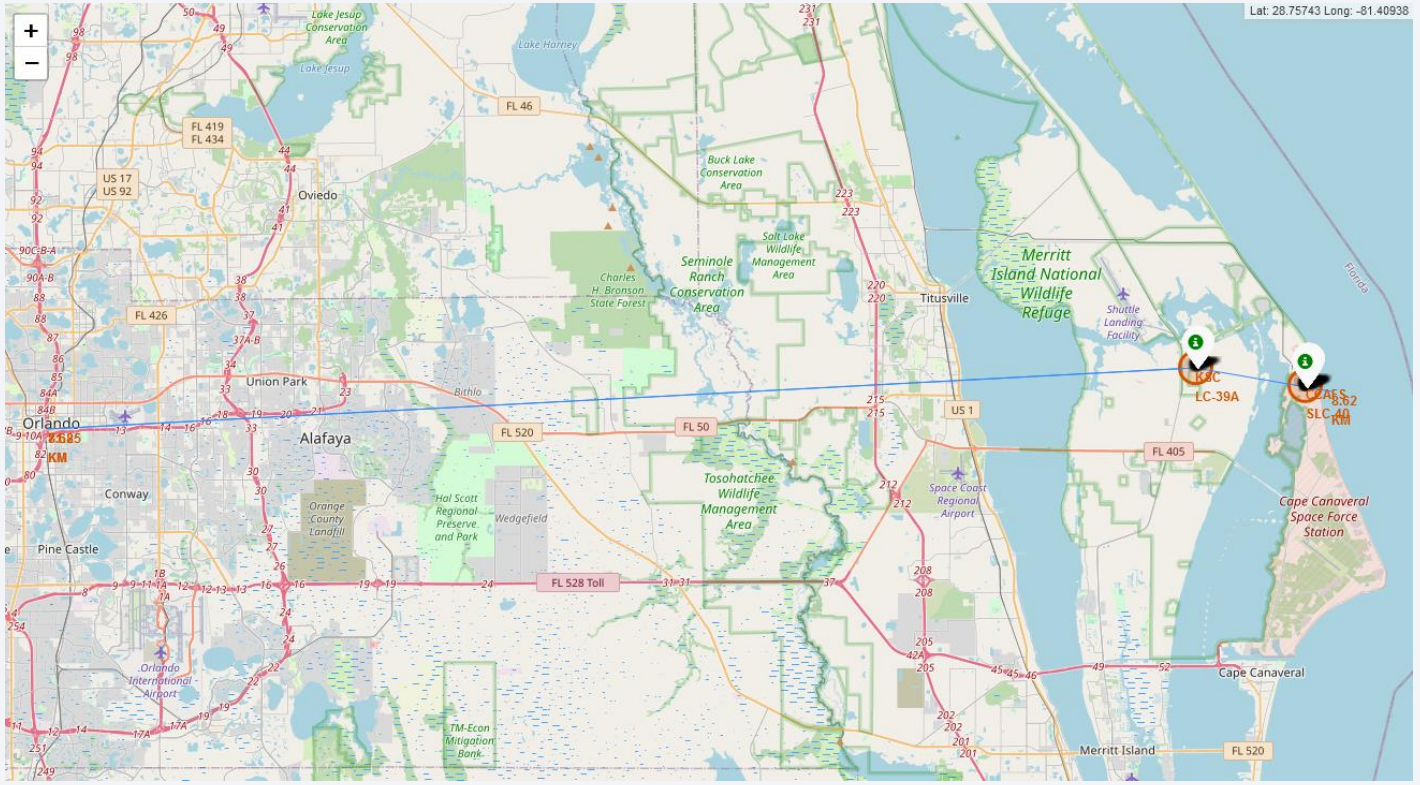
Launch Outcomes Per Launch Site

- Failed launch outcomes can be visualized in the map via the colored labeled markers.



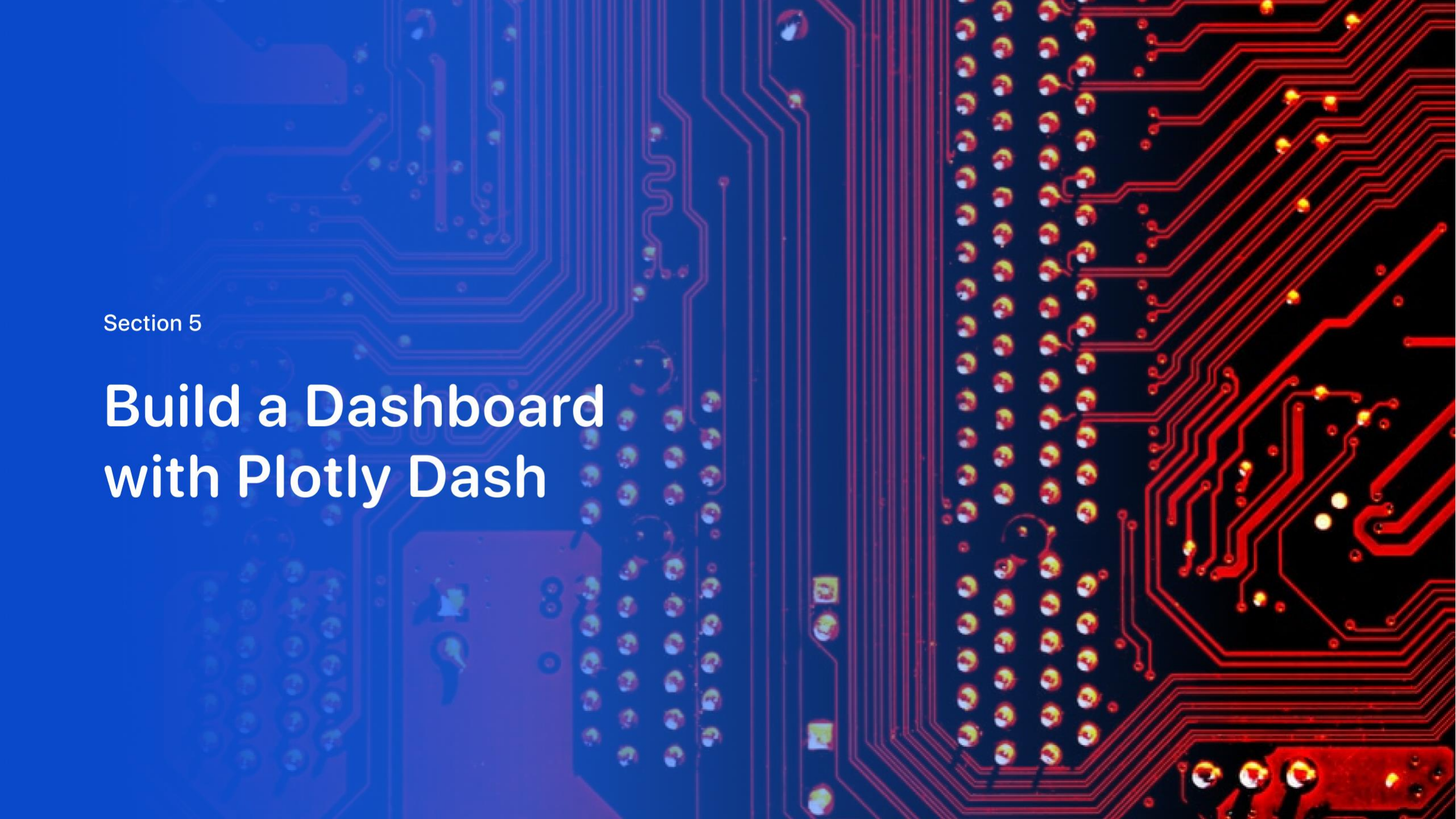
Contextual map of launch sites

- Both launch sites in Florida, USA are close to the mayor city, Orlando, aprox 72 Kms away.



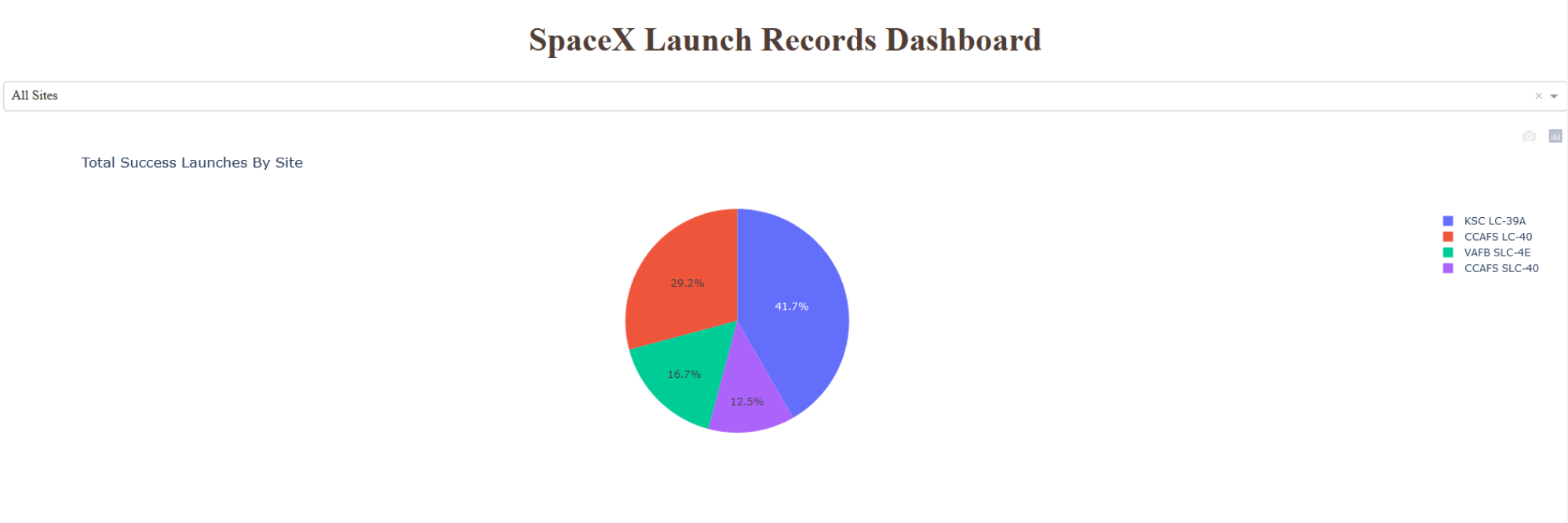
Section 5

Build a Dashboard with Plotly Dash



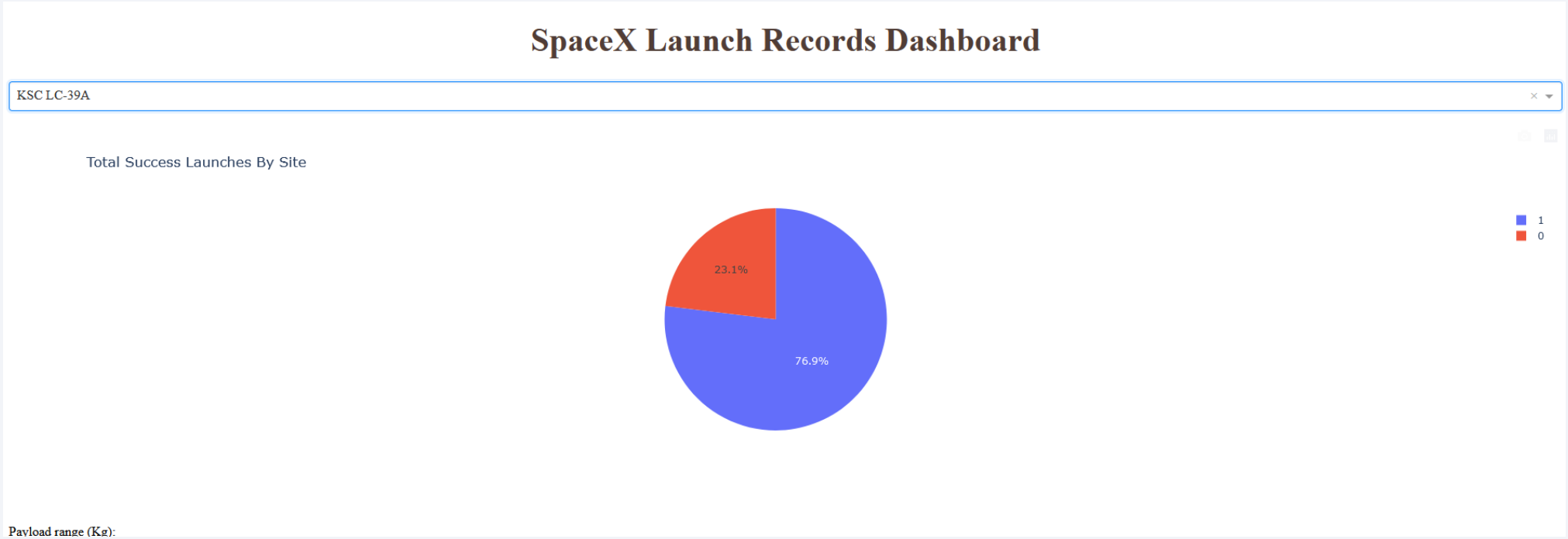
Success Distribution per Launch Site

- In the dashboard created an intuitive and interactive way to visualize data helped encounter interesting insights of data. Launch site “KSC LC-39A” is responsible of 41.7% of successful landings.



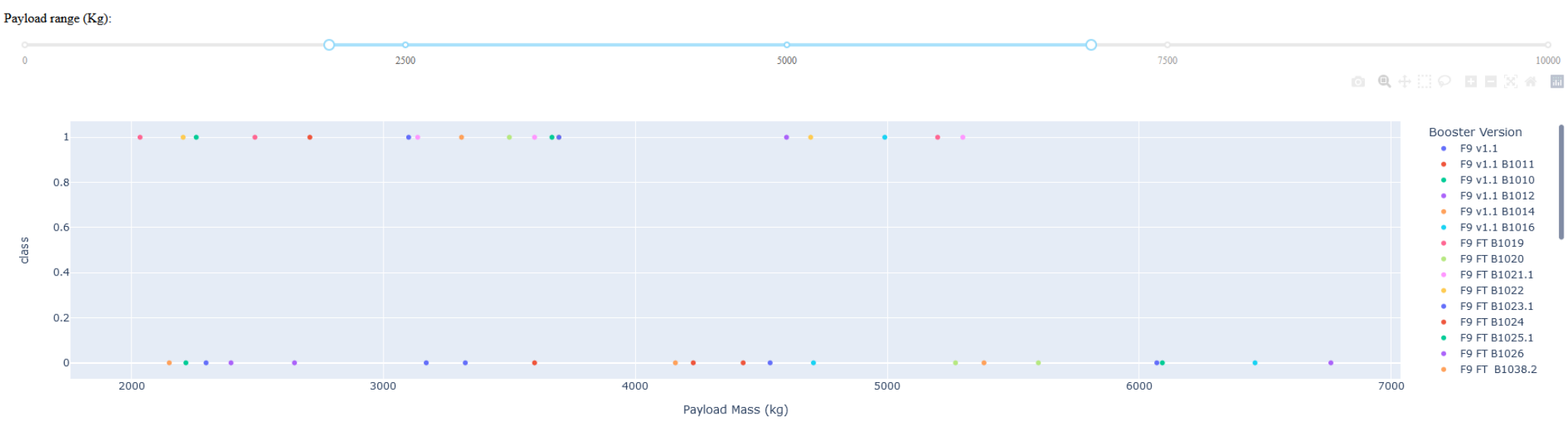
Most Successful Launch Site

- Launch site “KSC LC-39A” has the highest success rate of any launch sites with 76.9% of successful landings.



<Dashboard Screenshot 3>

- The range of 2000 kgs to roughly 6500 kgs is were most of the successful landings occur.

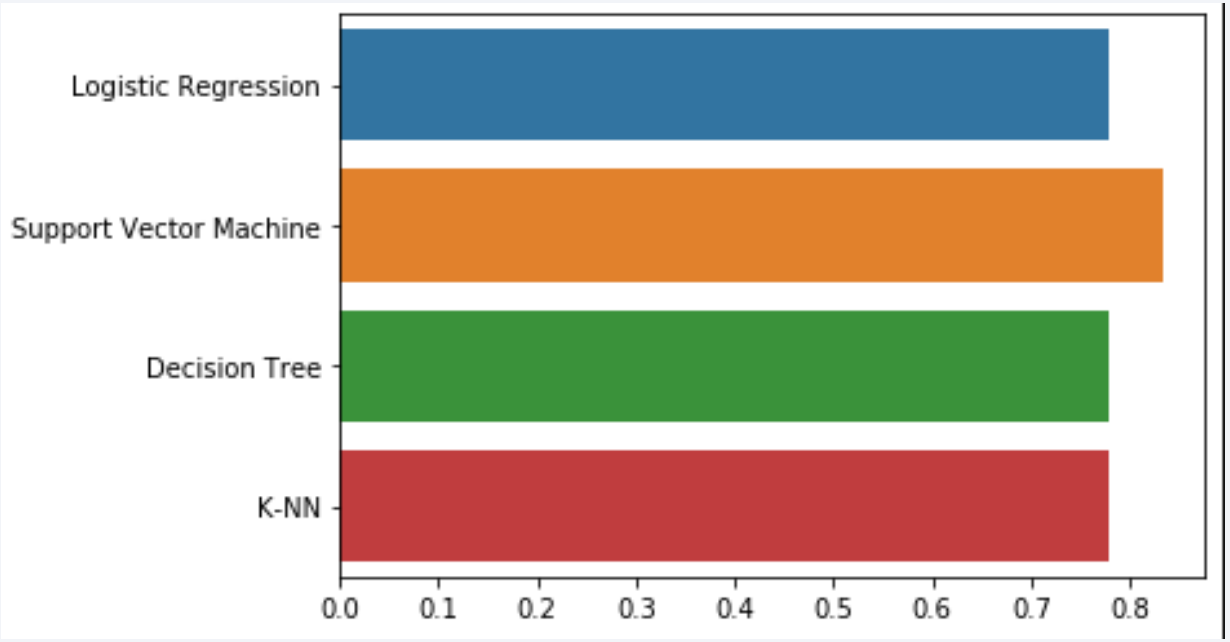


Section 6

Predictive Analysis (Classification)

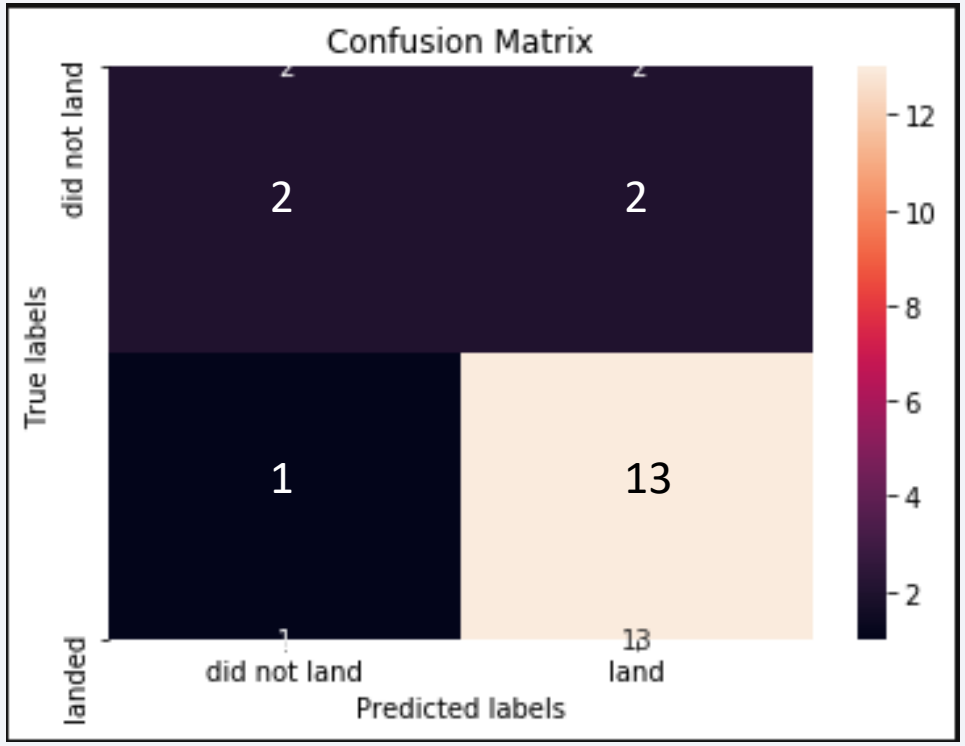
Classification Accuracy

- After the process mentioned above the algorithm with the best score using out of sample test data, the best algorithm is Support Vector Machine.



Confusion Matrix

- The diagonal line starting top left to bottom right are the correctly predicted launch outcomes. The algorithms main flaws lie in the False Negatives in the top left corner where it predicted the rocket to land but didn't.



Conclusions

- The main take away was the accomplishment of the main objective to create a model to predict the landing outcome of launches
- The best model is a support vector machine with an accuracy of 94%.
- Some interesting insights of the EDA are the discrepancies between different orbits.
- Lower payload mass launched tend to have a more successful landing outcome.

Thank you!