IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ryan Sandner
15DEC22

Outline

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

Executive Summary

- Methodologies employed
  - Data collection including webscraping wiki pages and SpaceX Rest API
  - Data analysis
    - Data Wrangling
    - Data Visualization
    - Interactive Visual Analytics and Dashboards
  - Predictive Analysis

- Summary of all results
  - It is possible to predict a successful 1st stage Falcon 9 landing with over 83% probability utilizing any one of 4 different predictive models.

Introduction

- The commercial space age is here, companies are making space travel affordable for everyone. Virgin Galactic is providing suborbital spaceflights. Rocket Lab is a small satellite provider. Blue Origin manufactures sub-orbital and orbital reusable rockets. Perhaps the most successful is SpaceX. SpaceX's accomplishments include: Sending spacecraft to the International Space Station. Starlink, a satellite internet constellation providing satellite Internet access. Sending manned missions to Space. One reason SpaceX can do this is the rocket launches are relatively inexpensive. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.
- Can we use historical Falcon 9 data to determine if the first stage will land successfully?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was scraped from the Falcon 9 wiki page (appendix 1) and SpaceX rest API (appendix 2)
- Perform data wrangling
  - Data was analyzed with nulls removed or substituted where appropriate, and extraneous data discarded.
  - Column created for landing outcome label
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
  - Interactive maps and charts creared
- Perform predictive analysis using classification models
  - Data divided into training and test subsets and evaluated in several different applied to several different machine learning models

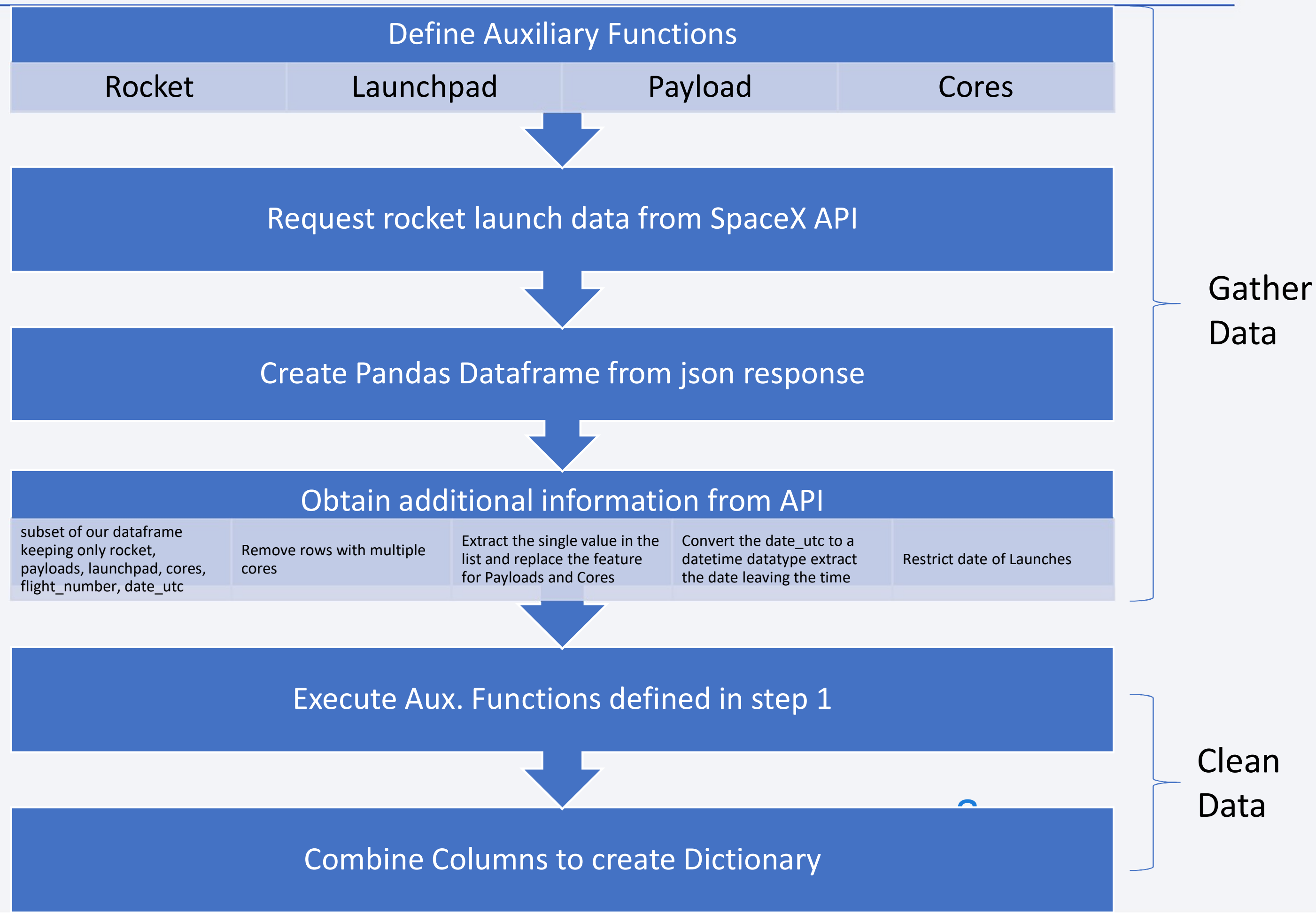
Data Collection

- Data collected from two main sources
  1. SpaceX Rest API – past launches
     - https://api.spacexdata.com/v4/launches/past
  2. Webscraping SpaceX Flacon Heavy wiki page
     - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Data Collection – SpaceX API

- Request data from SpaceX API
- Clean requested data

- [Source Code: https://github.com/RyanSandner/Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api(1).ipynb](https://github.com/RyanSandner/Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api(1).ipynb)



Data Collection – SpaceX API continued

Create dataframe using disctionary

Filter for only Falcon 9 launches

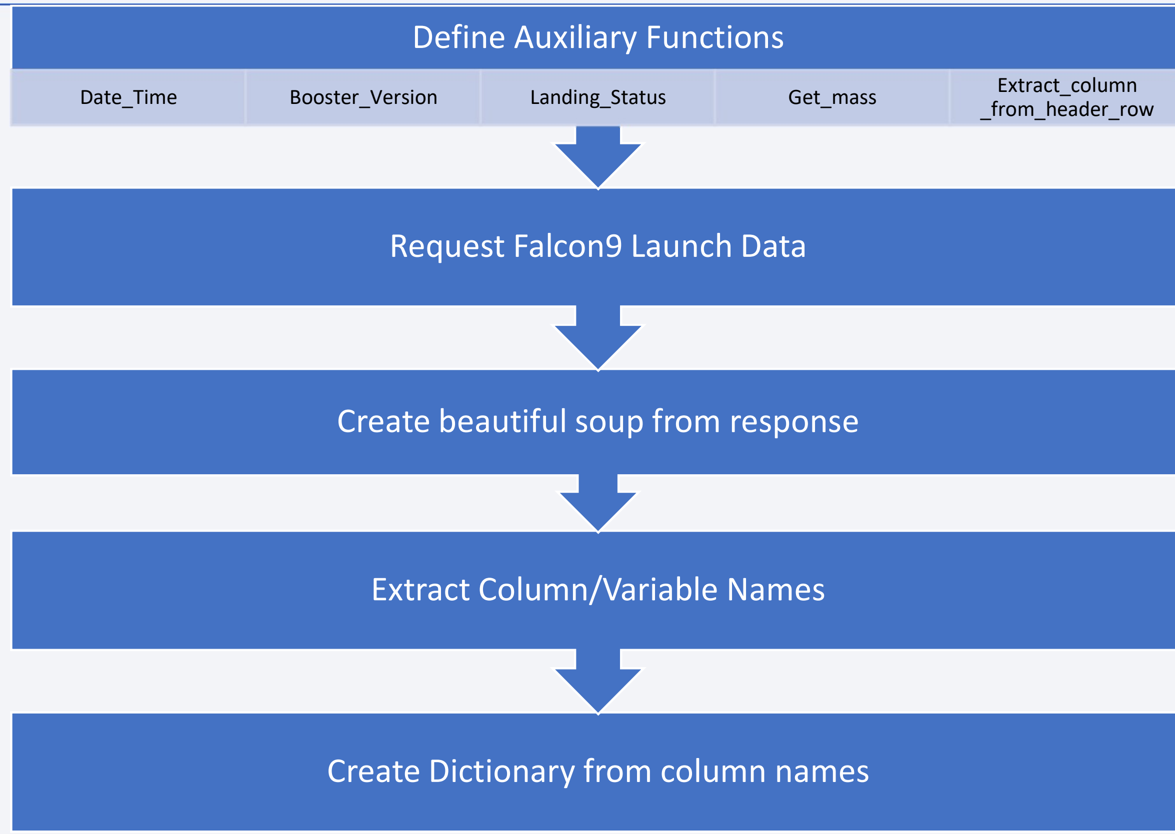
Replace nulls in PayloadMass with the average payload mass

Export Data

Clean Data
Data

Data Collection - Scraping

- Scrape data from Falcon9 wiki page
- Parse scraped data to dataframe



- [Source Code: https://github.com/RyanSandner/Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb](https://github.com/RyanSandner/Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb)

Data Collection – Scraping continued

Fill Dictionary with data extracted from table rows

Create dataframe from dictionary

Export csv file

- Source Code: https://github.com/RyanSandner/Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

- Exploratory Data Analysis (EDA) performed find some patterns in the data and determine/create suitable labels for training supervised models.
- [Source code: https://github.com/RyanSandner/Data-Science-Capstone/blob/main/labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb](https://github.com/RyanSandner/Data-Science-Capstone/blob/main/labs-jupyter-spacex-data_wrangling_jupyterlite.jupyterlite.ipynb)

Calculate Number of Launches on each site

Calculate number and occurrence of each orbit

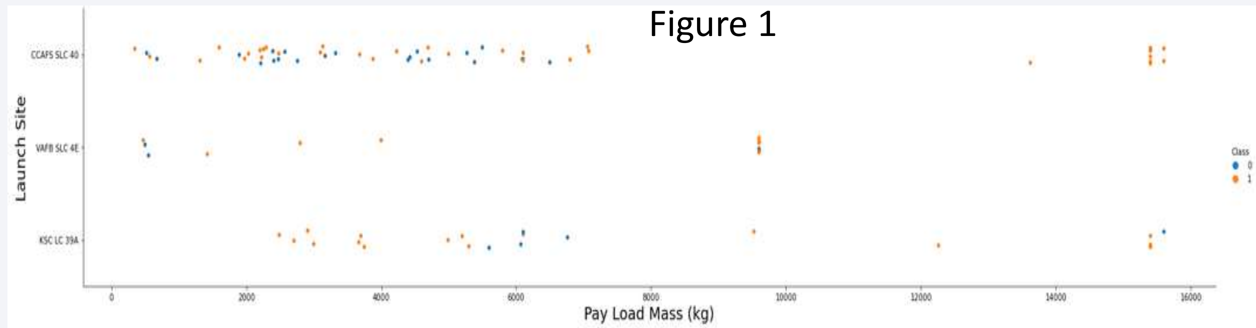
Calculate number and occurrence of mission outcome per orbit type

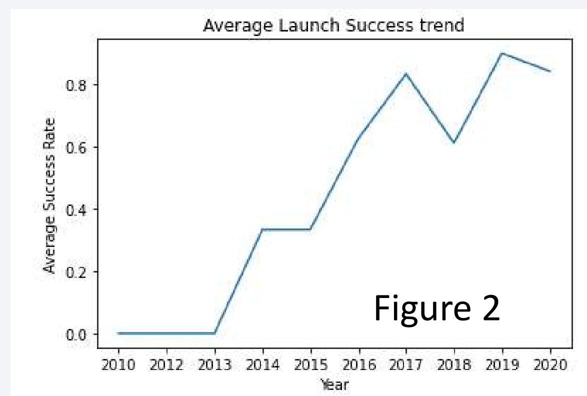
Create Landing Outcome label

Use Outcome label

EDA with Data Visualization

- Several charts were created to see if there are any visual correlations between the most likely influential variables and the landing outcome.
  - Flight Number and Launch Site
  - Payload and Launch Site (figure 1)
  - Success rate of each orbit type
  - Flight Number and Orbit type
  - Payload and Orbit type
  - Launch Success yearly trend (figure 2)
- Source code: https://github.com/RyanSandner/Data-Science-Capstone/blob/main/EDA%20with%20Visualization.ipynb



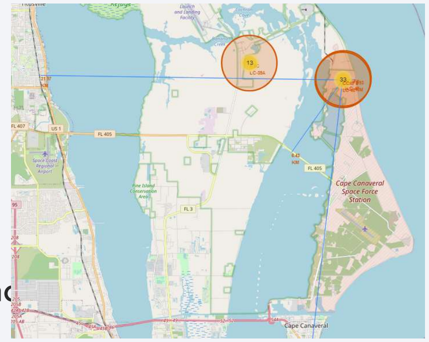


EDA with SQL

- SQL Queries performed:
  - Names of the unique launch sites in the space mission
  - 5 records where launch site begins 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
  - Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
  - Total number of successful and failure mission outcomes
  - Names of the booster versions which have crried the maximum payload mass
  - List of failed landing outcomes in drone ship, their booster versions and launch site names for 2015
  - Count of landing outcomes between June 4, 2010 and March 20, 2017 in descending order
- Source code: https://github.com/RyanSandner/Data-Science-Capstone/blob/main/Complete%20EDA%20with%20SQL%20Lab.ipynb

Build an Interactive Map with Folium

- Map Objects added to folium map
  - Markers and Circles added to highlight each launch site
    - Show where launch sites are in relation to geography
  - Markers added at each launch site to denote success of launch landing
    - Visualize if certain launch sites are more successful than others
  - Lines added to denote distance to closest coastline, railway, highway and city to one launc
    - Visualize proximity to geographical features, population and infrastructure



- Source code: https://github.com/RyanSandner/Data-Science-Capstone/blob/main/SkillsNetwork_labs_module_3_lab_jupyter_launch_site_location.jupyterlite(1).ipynb

Predictive Analysis (Classification)

- Several different models were created and trained on traiing data set
- Each model was then tested on the test data set and an accuracy calculation and conficion matrix were performed.
- The results (presented further in the presentation) were tabulated and compared to choose the best (most accurate) model.
- Source Code: https://github.com/RyanSandner/Data-Science-Capstone/blob/main/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite(1).ipynb

Data Preparation

- Create Y data from Class
- Standardize and assign X data
- Create test and train X and Y subsets

Create Models

- Logistic Regression
- Support Vector Machine
- Decision Tree Classifier
- K Nearest Neighbors

Train Models

- Trained with X_train and Y_train data subsets

Test Models

- Tested with X_test and Y_test

Evaluate test data

- Calculate accuracy of the : Scikit-Learn Score function used
- Create and view confusion matrix

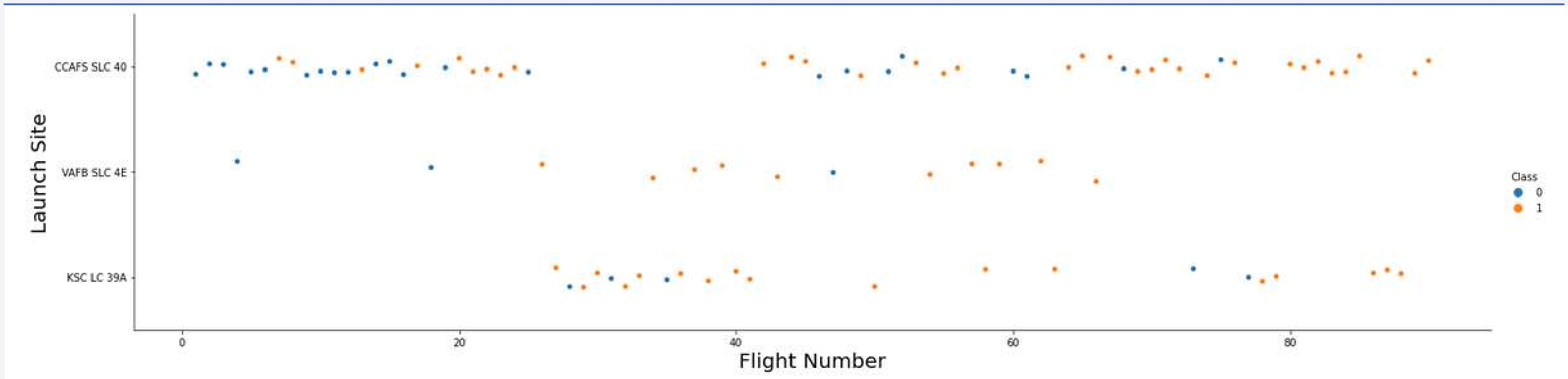
Results

- Exploratory data analysis results- the largest takeaway was the mprovement of launch success over time.
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

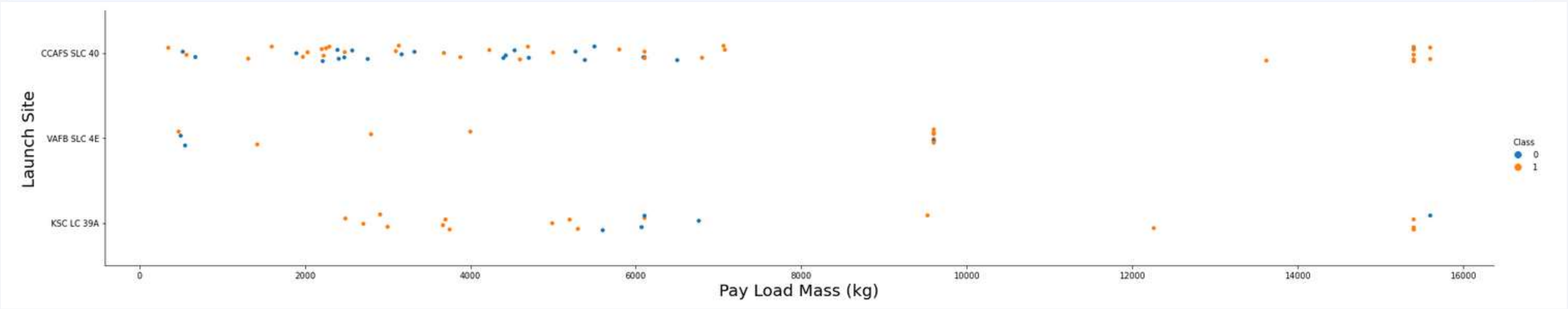
Insights drawn from EDA

Flight Number vs. Launch Site



- The data shows that each launch sites success rate improved over time. Currently the most successful site is CCAFS SLC 40 with a 100% success rate for their last 9 launches.
  - VAFB SCL 4E had one failure in the last 9 while KSC LC 39A had 2
- CCAFS SLC 40 also has the most launches of the three sites.

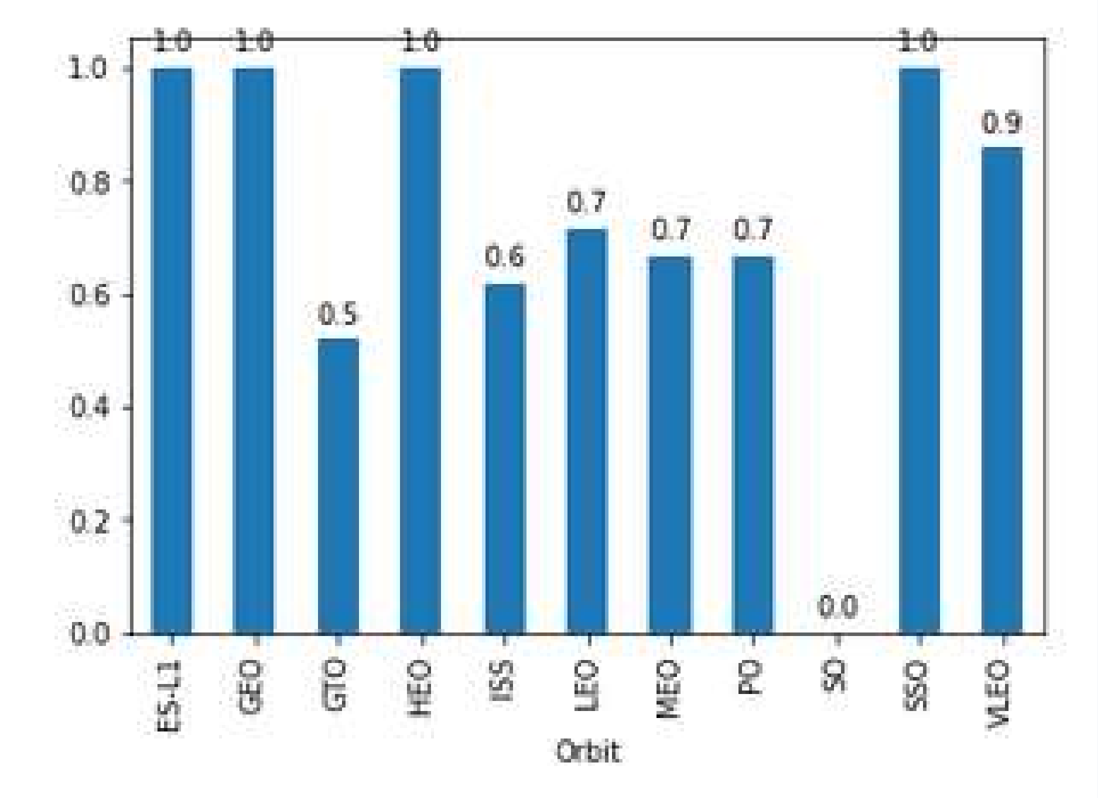
Payload vs. Launch Site



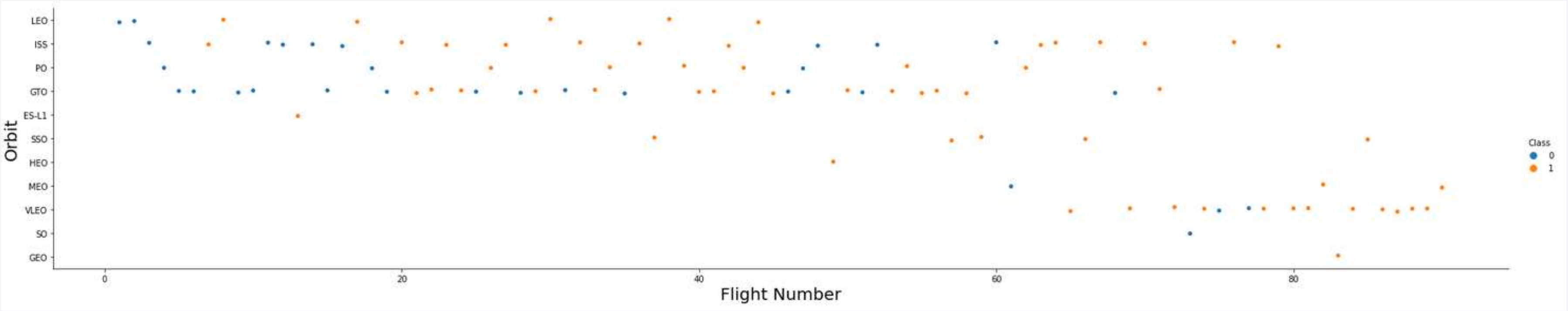
- CCAPS SLC 40 is the most successful site for heavy payloads
  - WFB SLC 4E did not launch any payloads greater than 10,000 kg
- KSC LC 39A has the highest success rate for payloads between 2,000 kg and 5,000 kg
- In general the heavier payloads (greater than 8,000kg) have a higher success rate.

Success Rate vs. Orbit Type

- Orbits Earth Sun Lagrange 1 (ES-L1), Geosynchronous (GEO), Highly Elliptical Orbit (HEO) and Sun Synchronus Orbit (SSO) have a 100% success rate

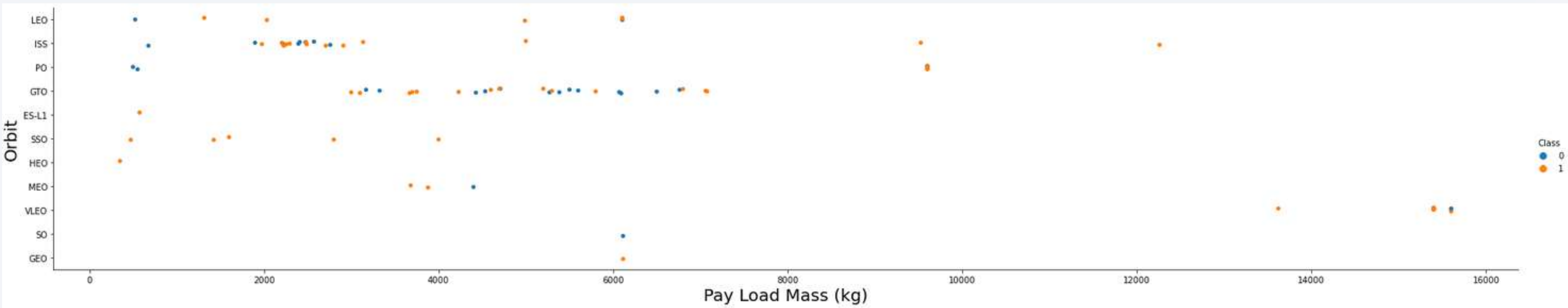


Flight Number vs. Orbit Type



- Again we see that most orbits have improved success rate over time
- Very Low Earth Orbit (VLEO) has the highest frequency of recent launches

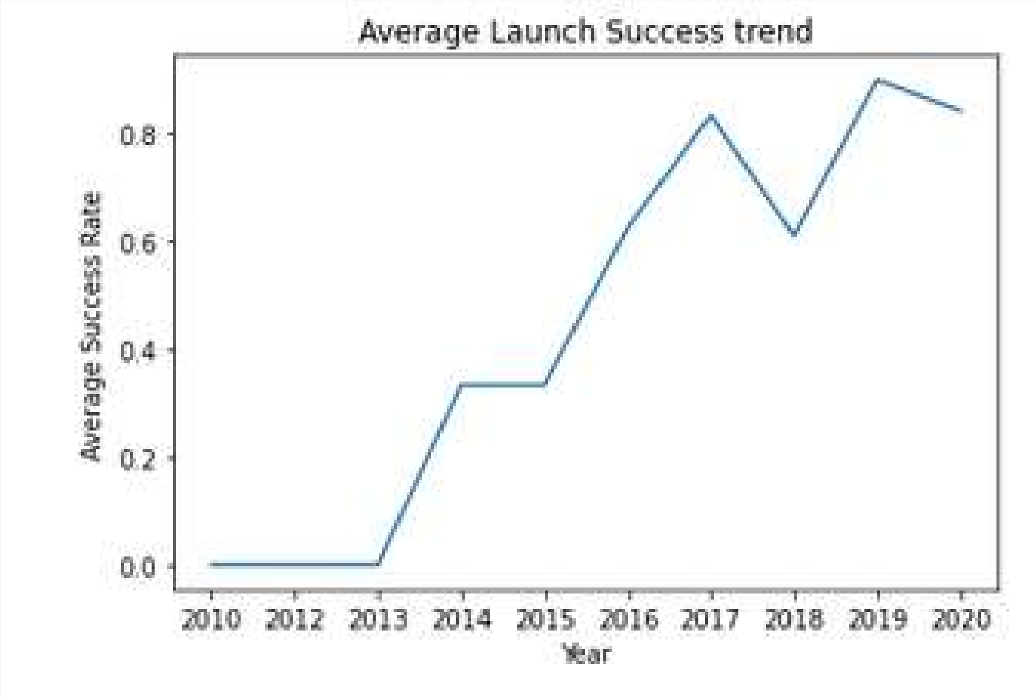
Payload vs. Orbit Type



- The orbit with the most launches is Geostationary Transfer Orbit, the payload ranges between ~3,000kg and 7,000kg.
  - The success rate is distributed throughout the payload range
- International Space Station (ISS) orbit has the widest range of payloads with a high success rate in payloads greater than ~3,000kg.
- Sun Syncronous Orbit (SSO) was successful for all launches
  - Payload range less than 4,000kg
- There is no direct correlation between payload mass and orbit to the success of the launch

Launch Success Yearly Trend

- It is easy to see here the trend of successful launches is increasing with passage of time.



All Launch Site Names

- The four unique launch sites in the data set are shown to the right
  - Results obtained by selecting distinct launch sites from the data set

launch_site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

Launch Site Names Begin with 'CCA'

- The table below shows 5 records where the launch site begins with ‘CCA’

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

- Results obtained by selecting data with launch site like ‘CCA’ and limiting selection to 5 rows

Total Payload Mass

- The total payload mass carried by boosters launched by NASA (CRS)
  - 45,596kg
  - Results obtained by summing the payload mass from the data where the customer = Nasa (CRS)

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1
  - 2,534kg
  - Results obtained by averaging the payload mass where the booster version is like ‘F9 v1.1’

First Successful Ground Landing Date

- Date of the first successful landing outcome on ground pad
  - December 22, 2015
  - Results obtained by selecting the minimum date where landing outcome was sucessful

Successful Drone Ship Landing with Payload between 4000 and 6000

- Boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

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

  - Results obtained by selecting booster versions for the data set where the outcome was successful and the range of payload mass was between 4000 and 6000 kg

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

sum_fail	sum_pass
1	100

- Results obtained by selecting and summing mission outcome and for like “Fail%’ and like ‘Suc%’

Boosters Carried Maximum Payload

- Names of the booster which have carried the maximum payload mass

booster_version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

- Results obtained by selecting booster version where payload is maximum from payload mass

2015 Launch Records

- Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

booster_version	launch_site	landing_outcome
F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

  - Results obtained by selecting booster version, launch site and landing outcome from the data set where data like ‘2015’ and landing outcome like ‘Fail’

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

- 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

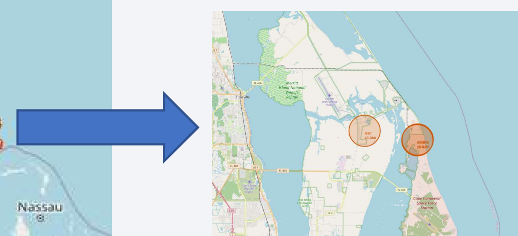
landing_outcome	num
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

  - Results obtained by selecting landing outcome, counting the outcomes between the two specified dates then ordering by the cunt descending

Section 3

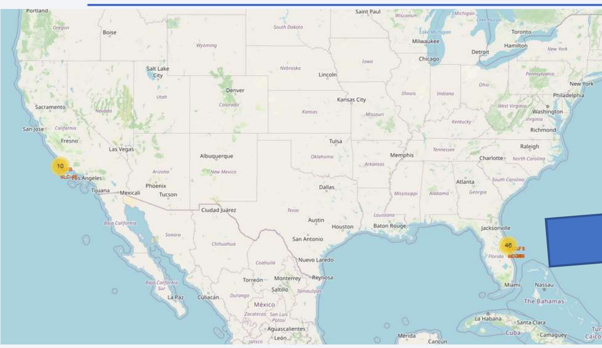
Launch Sites Proximities Analysis

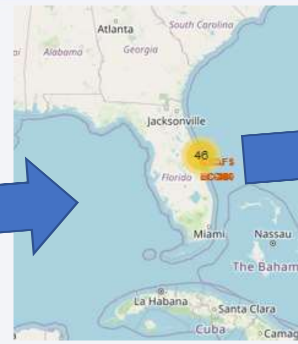
Map of All Launch Sites

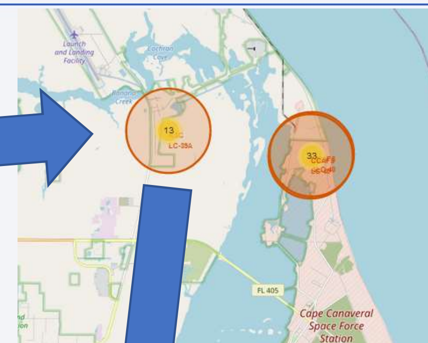


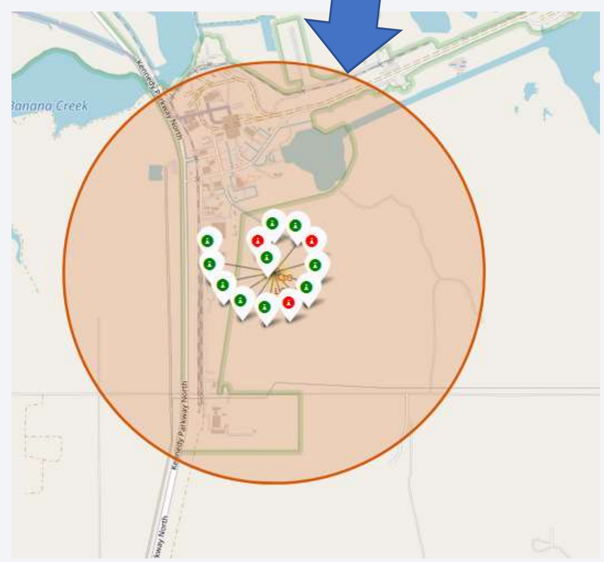
- Launch sites shown on US map
  - Launch sites are located in coastal areas.

Launch Outcomes by Site



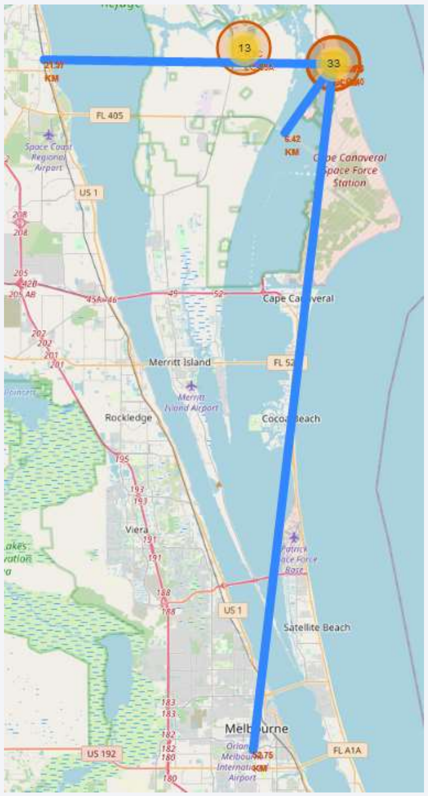




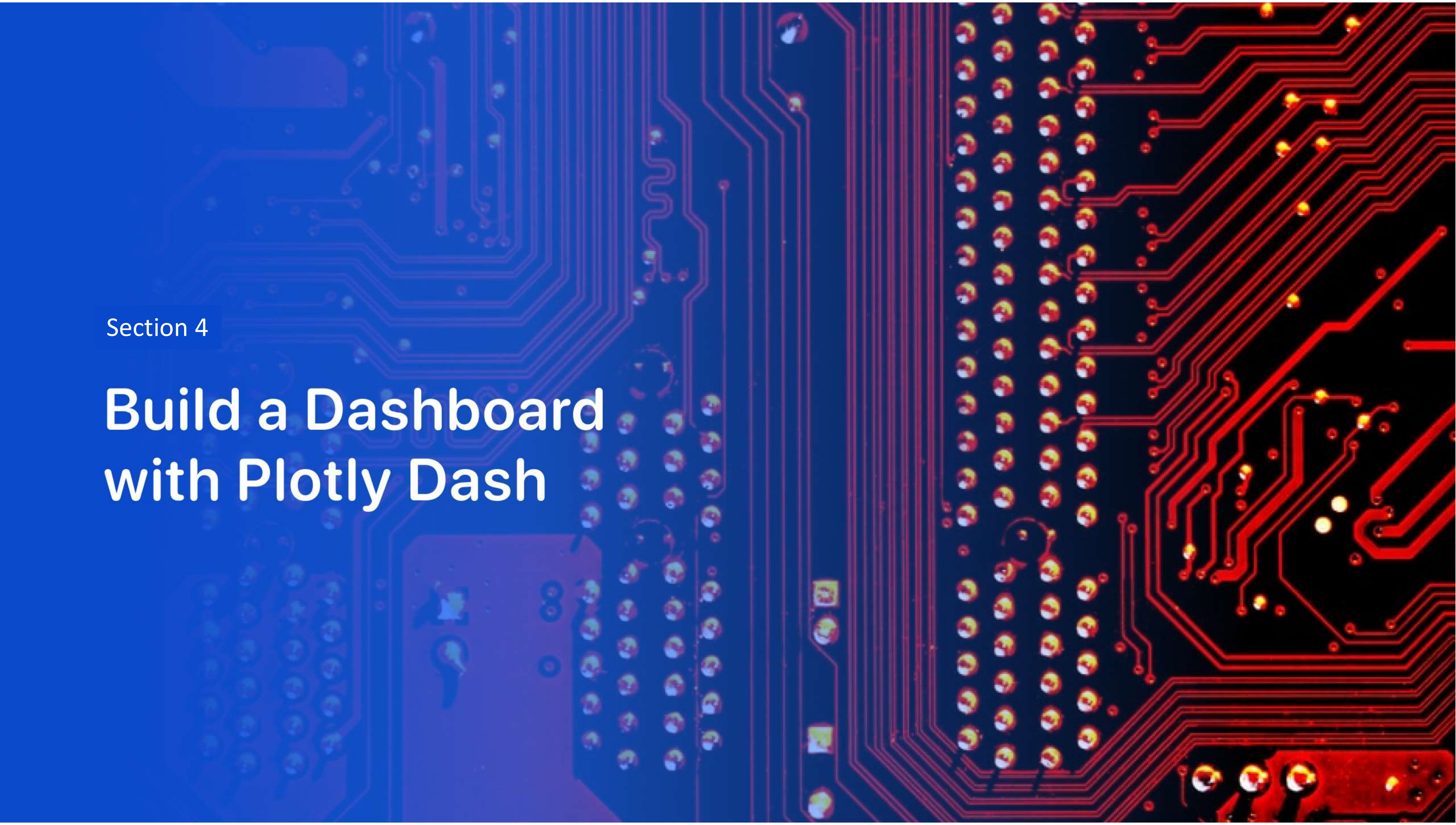


- Example of drilling down on Map – starting at US level, down to site level.
- Final detail shows number of launches color coded for outcome
    - red = failure
    - green = success)

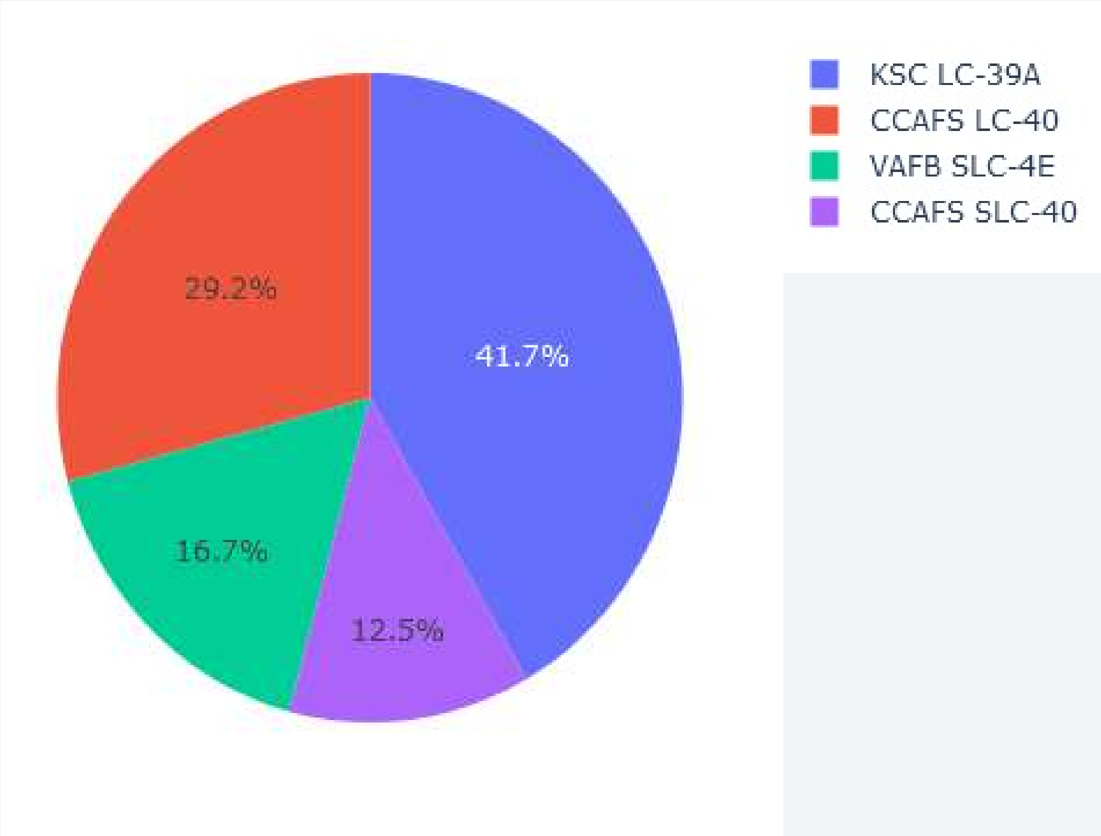
Proximity of Launch Site to Major Infrastructure



- We can see the proximity of the launch site to major infrastructure
  - ~22km to the nearest rail
  - ~6.4km to the nearest highway
  - ~52.8km to the nearest city

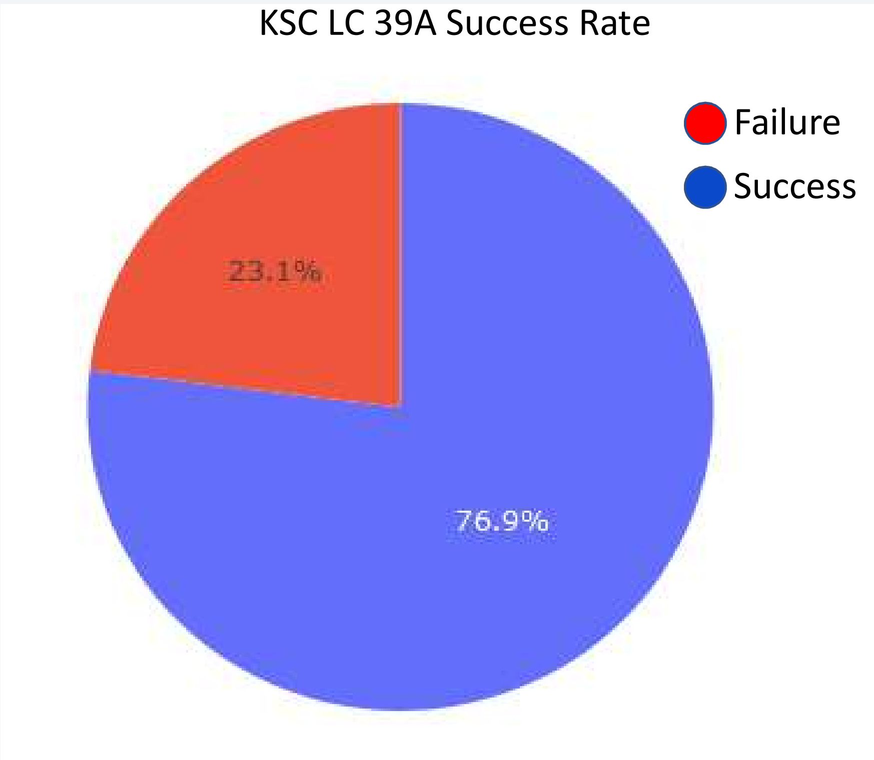

Section 4
Build a Dashboard with Plotly Dash

Launch Success for All Sites



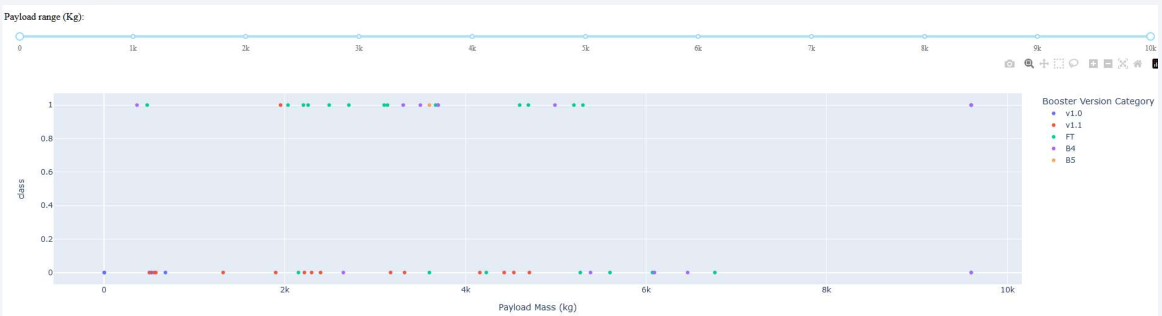
- We can see Kennedy Space Center Launch Complex 39A (KSC LC-39A) launch site has the highest success rate while Cape Canaveral Space Launch Complex 40 (CCAFS SLC-40) is the lowest

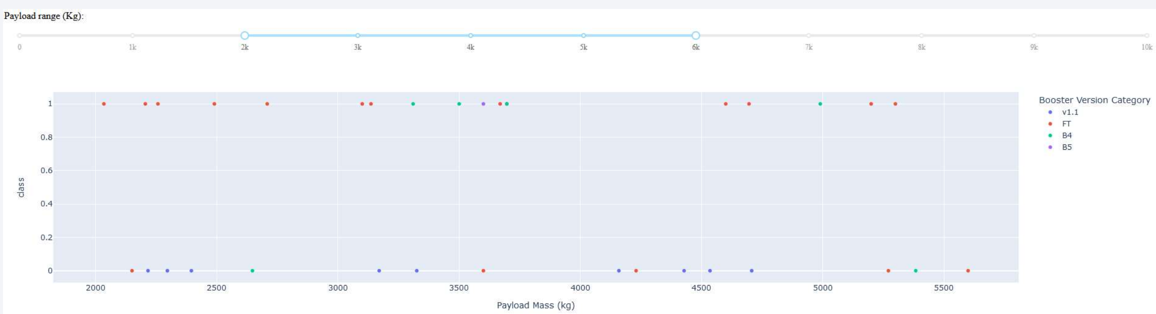
Site with the highest Launch Success

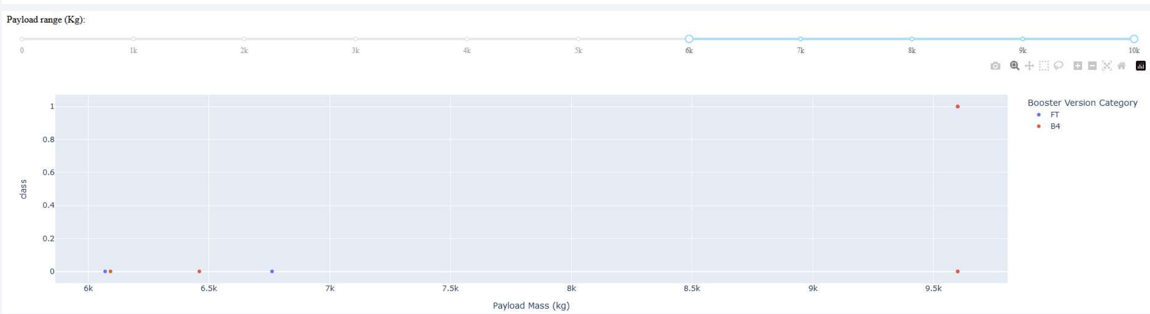


- More than ¾ of the launches from KSC LC-39A have been successful.

Payload vs. Launch Outcome







- Total outcome for all ranges: we can see the concentration of payload is 2000kg to 6000 kg

- Focusing on the 2000-6000kg payload range we can see the FT booster version has a high success rate

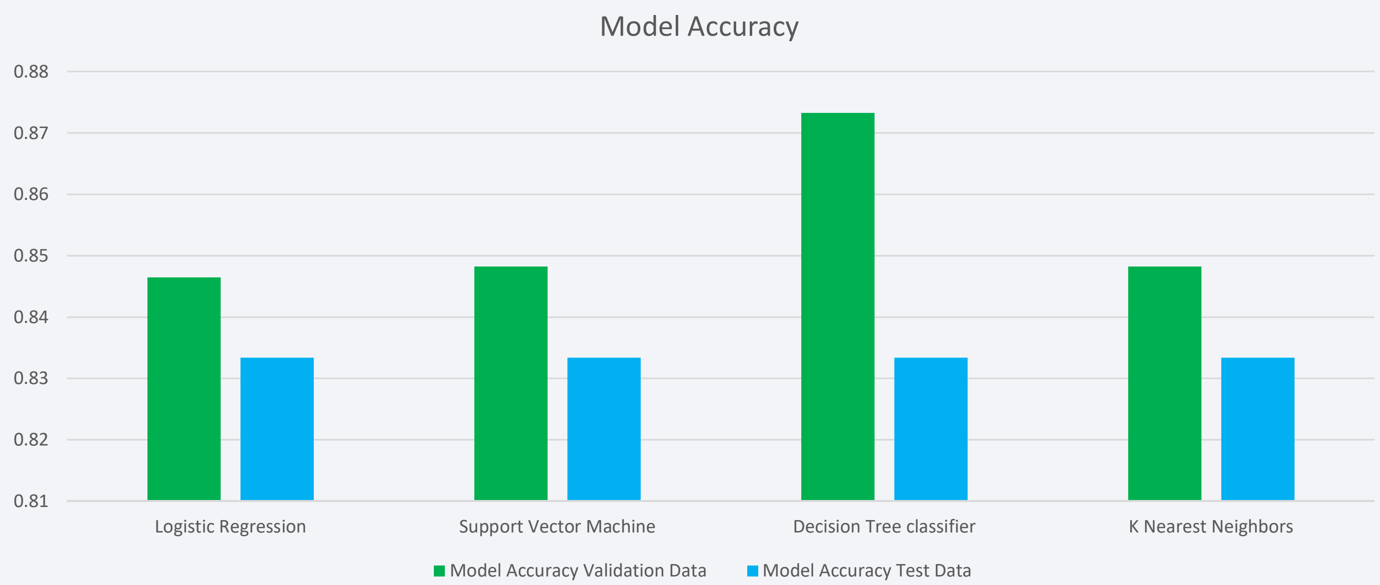
- Looking at heavy payload (>7000kg) we see a high percentage of success with no FT booster failures

- We conclude the FT is better for midrange payloads than heavy payloads

Section 5

Predictive Analysis (Classification)

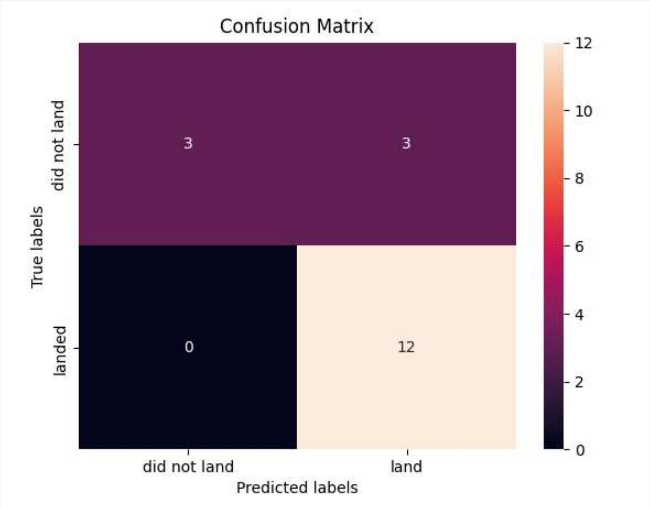
Classification Accuracy



- The decision tree classifier model has the highest accuracy on the validation data
- All models have the same accuracy on the test data

Confusion Matrix

- All models have the same accuracy on the test data therefore the confusion matrix for all is the same
- We can see the models can distinguish between the different classes and the major problem is false positives



Conclusions

- All machine learning algorithms performed with the same accuracy on the test data however the decision tree model performs slightly higher on the training data.
- The FT booster with payload range 2000-6000kg performed well compared to the other boosters and payload ranges.
- Kennedy Space Center Launch Complex 39A has the highest success rate
- Launch success has improved drastically over time
    - Intuitively this make sense, learning from previous launches and improvement in technology aid in success.
- Sun Syncronous Orbit (SSO) was successful for all launches (note, payload was less than 4000kg for all attempts)

Appendix

1. Falcon 9 Launch Wiki:
    - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
2. SpaceX Rest API
    - api.spacexdata.com/v4/launches/past

Thank you!