IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Sierra K
06-23-2025

Outline

- Executive Summary
- Introduction
- Methodology
- Insights Drawn from EDA
- Launch Site Proximities Analysis
- Build Dashboard with Plotly Dash
- Predictive Analysis
- Conclusion

Executive Summary

- Summary of methodologies
  - Data Collection
  - Data Wrangling
  - Exploratory Data Analysis
  - Interactive Visual Analytics
  - Predictive Analysis
- Summary of all results
  - Exploratory Data Analysis Results
  - Geospatial Analytics
  - Interactive Dashboard
  - Predictive Analysis of Classification Models

Introduction

- In this capstone, we will predict if the Falcon 9 first stage will land successfully. 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.
- If we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Making GET requests to the SpaceX REST API
  - Web Scrapping
- Perform data wrangling
  - Removed NAN values
- Determined the following:
  - Number of launches on each site
  - Number of occurrence of each orbit
  - Number and occurrence of mission outcome per orbit type

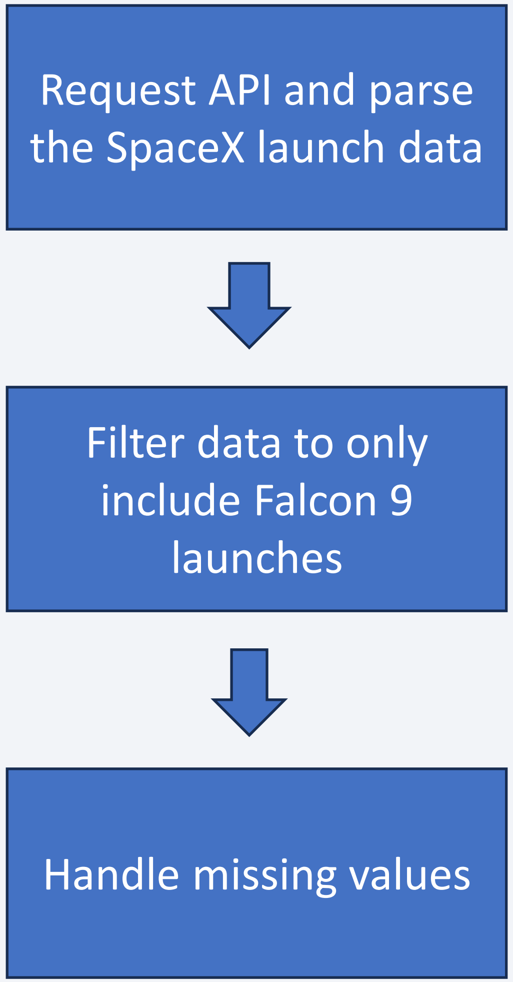
- Exploratory Data Analysis
  - Used SQL to manipulate and evaluate SpaceX dataset
  - Used Pandas and Matplotlib to visualize relationships between variables, and determine patterns
- Interactive Visual Analysis
  - Geospatial analytics using Folium
  - Creating an interactive dashboard using Plotly Dash
- Data Modeling and Evaluation
  - Used Scikit-Learn to:
  - Pre-processed (standardize) the data
  - Split the data into training and testing data
  - Train different classification models
  - Found hyperparameters
  - Plotted confusion matrices for each classification model
  - Assessed the accuracy of each classification model

Data Collection

- Used the SpaceX API to retrieve data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
- Data Sets were collected from SpaceX API
  - https://api.spacexdata.com/v4/rockets/
  - https://api.spacexdata.com/v4/payloads/
  - https://api.spacexdata.com/v4/cores/
  - https://api.spacexdata.com/v4/launches/past

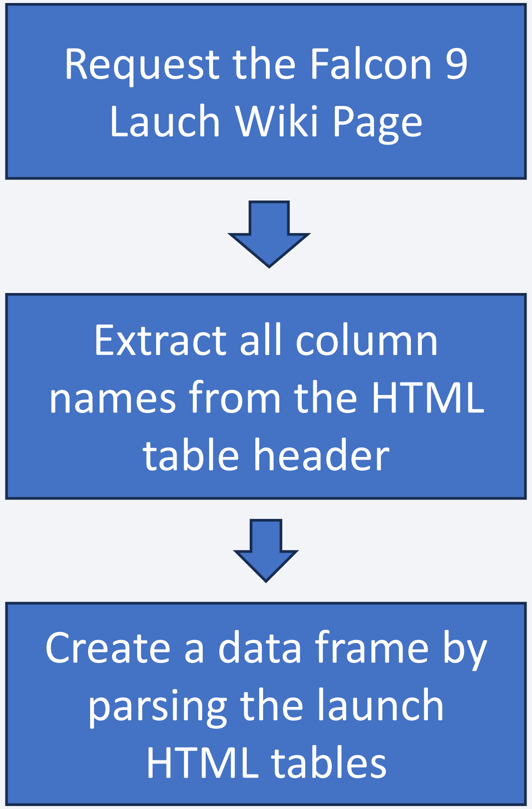
Data Collection – SpaceX API

- SpaceX offers a public API where data can be obtained and used
- This API was used according to the flowchart
- Source:
  - [SpaceX_Applied_Data_Science_Capstone/SpaceX Falcon 9 first stage Landing Prediction_Collecting the data.ipynb at main · snkty8/SpaceX_Applied_Data_Science_Capstone]



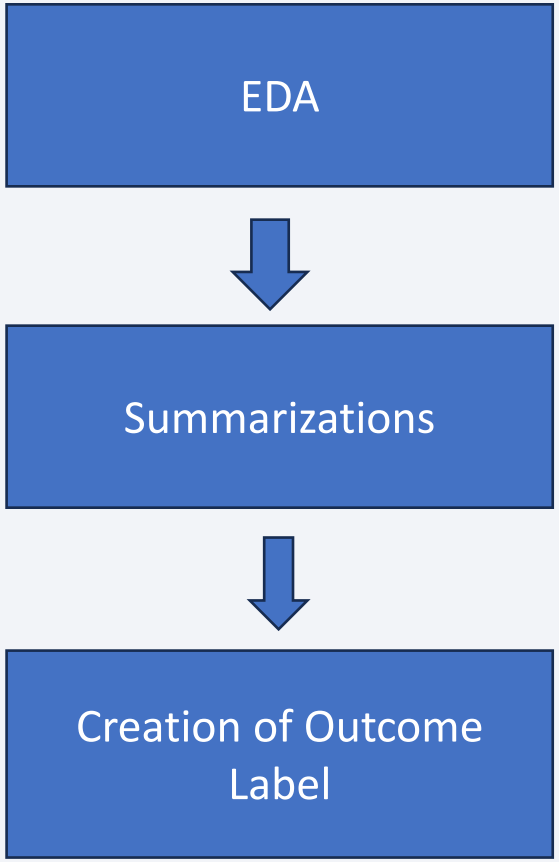
Data Collection - Scraping

- Data from SpaceX launches can also be obtained from Wikipedia
- Flow chart shows how data was obtained and used.
- Source:
  - [SpaceX_Applied_Data_Science_Capstone/SpaceX Falcon 9 first stage Landing Prediction_Web scraping.ipynb at main · snkty8/SpaceX_Applied_Data_Science_Capstone]



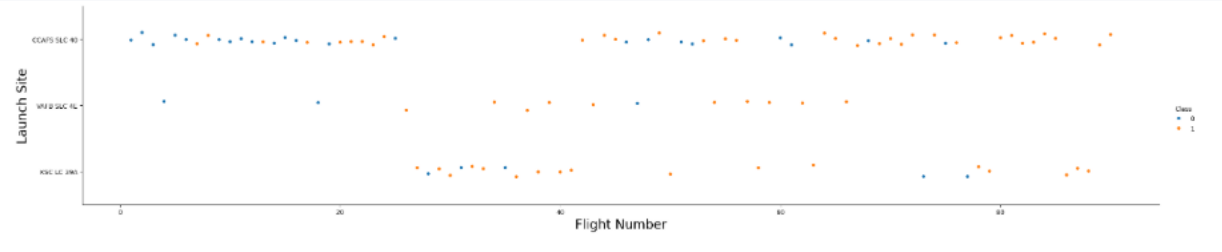
Data Wrangling

- Initial Exploratory Data Analysis was performed on the dataset
- Then the launch summaries per site, occurrences of each orbit, and mission outcome per orbit were calculated
- Source:
    - SpaceX_Applied_Data_Science_Capstone/SpaceX Falcon 9 first stage Landing Prediction_Data wrangling.ipynb at main · snkty8/SpaceX_Applied_Data_Science_Capstone



EDA with Data Visualization

- To explore the data, scatterplots, and bar plots were used to visualize the relationship between features:
  - Payload Mass X Flight Number, Lauch Site X Flight Number, Lauch Site X Payload Mass, Orbit and Flight Number, Payload and Orbit



Source:

[SpaceX_Applied_Data_Science_Capstone/SpaceX Falcon 9 first stage Landing Prediction_Visualizations.ipynb at main · snkty8/SpaceX_Applied_Data_Science_Capstone]()

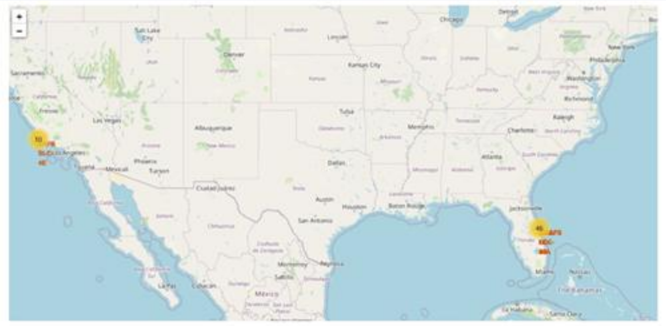
EDA with SQL

SQL Queries were used to:

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in ground pad was achieved.
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List all the booster_versions that have carried the maximum payload mass, using a subquery with a suitable aggregate function.
- List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.
- 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.
- Source:
    - SpaceX_Applied_Data_Science_Capstone/SpaceX Falcon 9 first stage Landing Prediction_EDA With SQL.ipynb at main · snkty8/SpaceX_Applied_Data_Science_Capstone

Build an Interactive Map with Folium

- Markers, circles, lines, and marker clusters were used with Folium Maps
  - Markers indicate launch sites
  - Circles indicate highlighted areas around specific coordinates, such as NASA Johnson Space Center
  - Marker clusters indicates groups of events in each coordinate, such as launches in a launch site
  - Lines are used to indicate distances between two coordinates

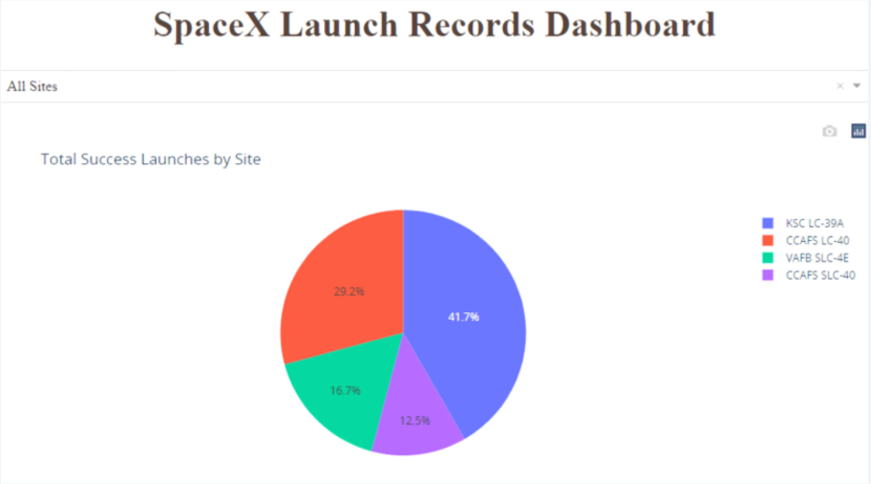


Source:

[SpaceX_Applied_Data_Science_Capstone/Interactive Visual Analytics with Folium.ipynb at main · snkty8/SpaceX_Applied_Data_Science_Capstone]

Build a Dashboard with Plotly Dash

- The following graphs and plots were to visualize data:
  - Percentage of launch site
  - Payload range
- This combination allowed quick analysis of the relation between payloads and launch sites, helping to identify where the best place was according to payloads

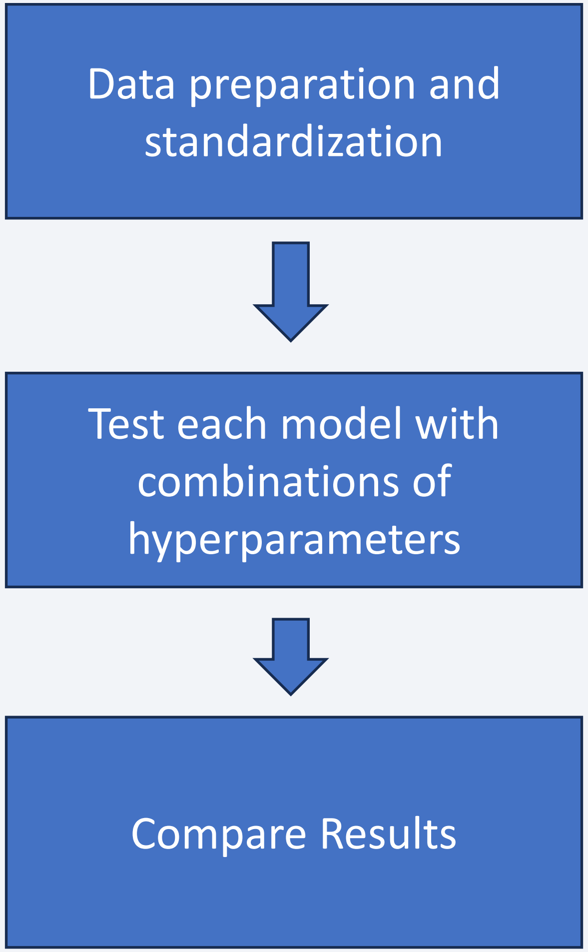


Source: [SpaceX_Applied_Data_Science_Capstone/Build an Interactive Dashboard with Plotly Dash.py at main · snkty8/SpaceX_Applied_Data_Science_Capstone]()

Predictive Analysis (Classification)

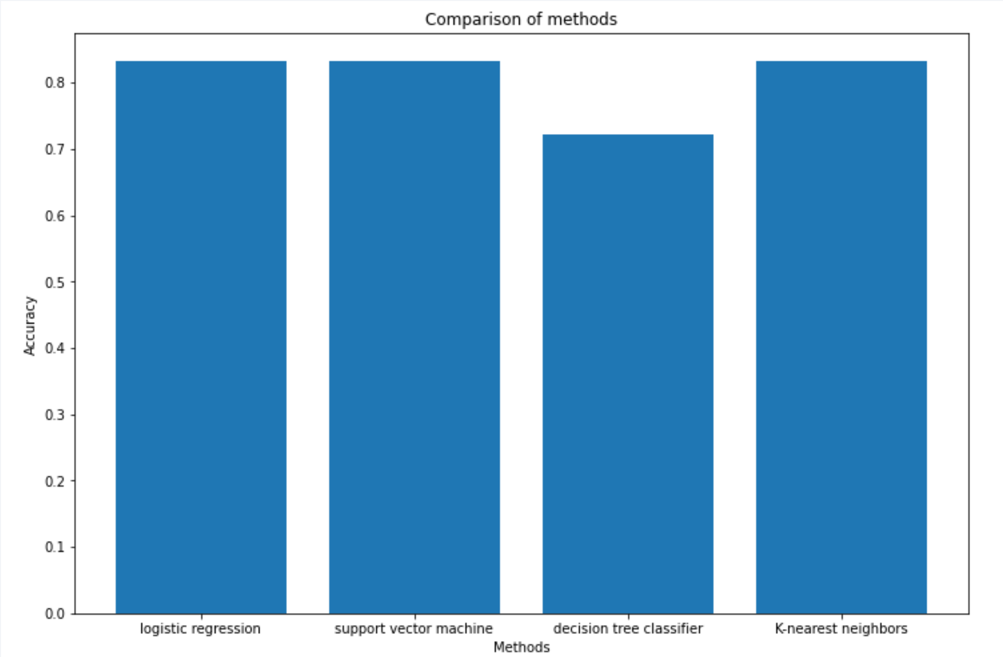
- Four classification models where compared:
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - K Nearest Neighbor

Source: [SpaceX_Applied_Data_Science_Capstone/Machine Learning Prediction.ipynb at main · snkty8/SpaceX_Applied_Data_Science_Capstone]()



Results

- **Logistic regression**, **Support vector machine** and **K-nearest neighbors**, all these three methods give the best performance, with accuracy of 0.8333333333333334.

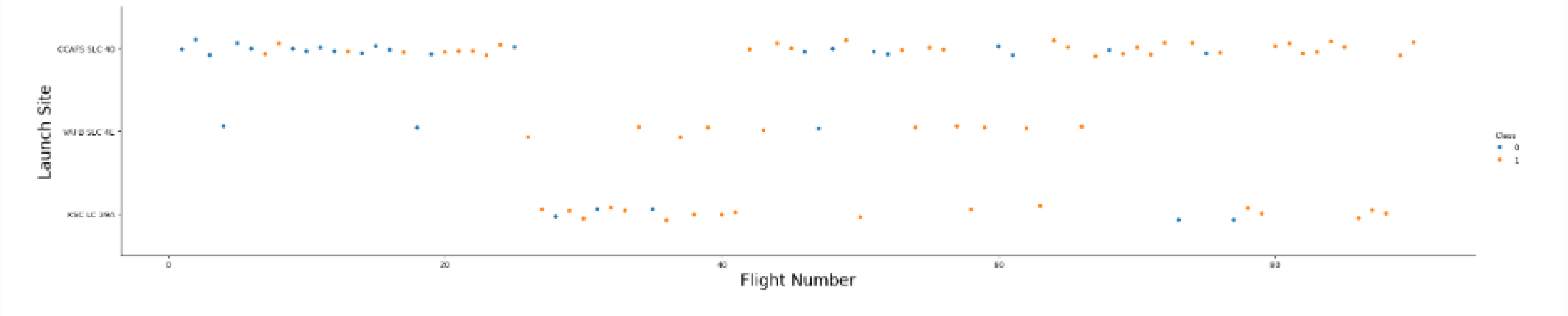


Source: [SpaceX_Applied_Data_Science_Capstone/Machine Learning Prediction.ipynb at main · snkty8/SpaceX_Applied_Data_Science_Capstone]

Section 2

Insights drawn from EDA

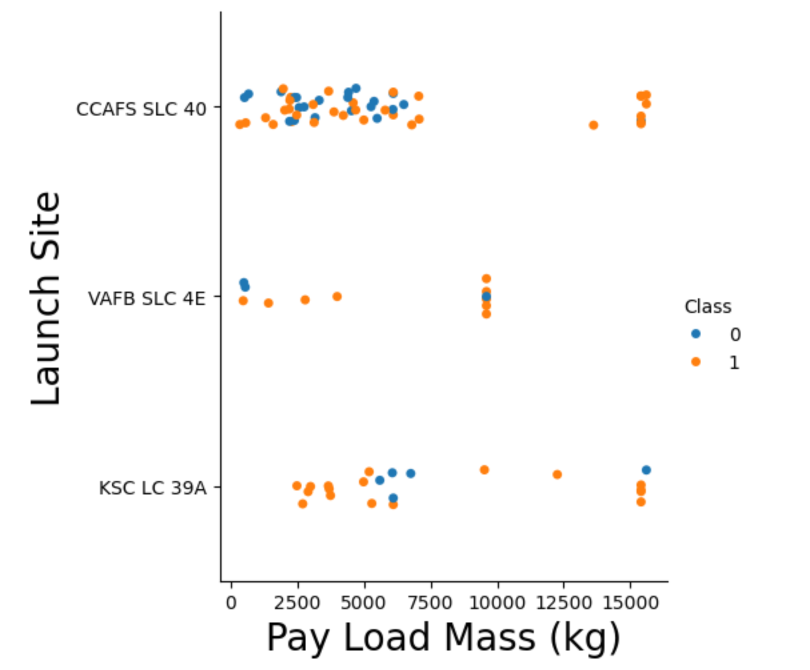
Flight Number vs. Launch Site



- According to the plot above, it is possible to verify that the best launch site is CCAF5 SLC 40, where most of the recent launches were successful
- In second place VAFB SLC 4E and third KSC LC 39A
- It is also possible to see that the general success rate improved over time

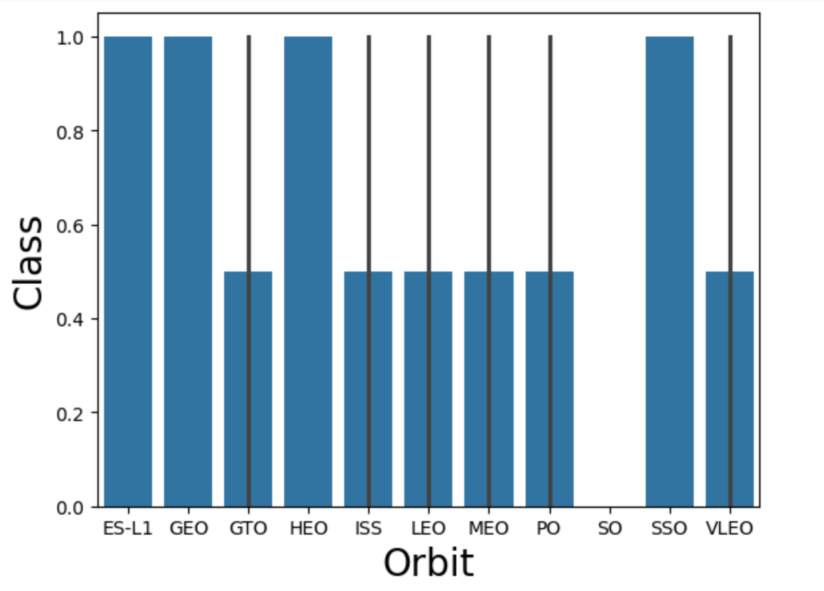
Payload vs. Launch Site

- Payloads over 9,000 kg have great a success rate
- Payloads over 12,000 kg only seems possible for CCAFS SLC 40 and KSC LC 39A launch sites



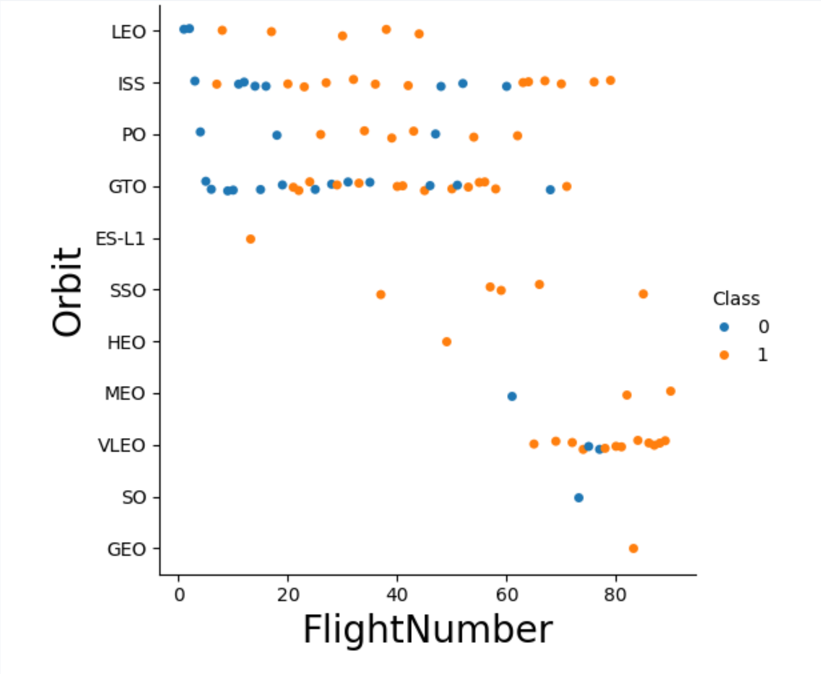
Success Rate vs. Orbit Type

- Orbits larges success rates:
  - ES-L 1
  - GEO
  - HEO
  - SSO



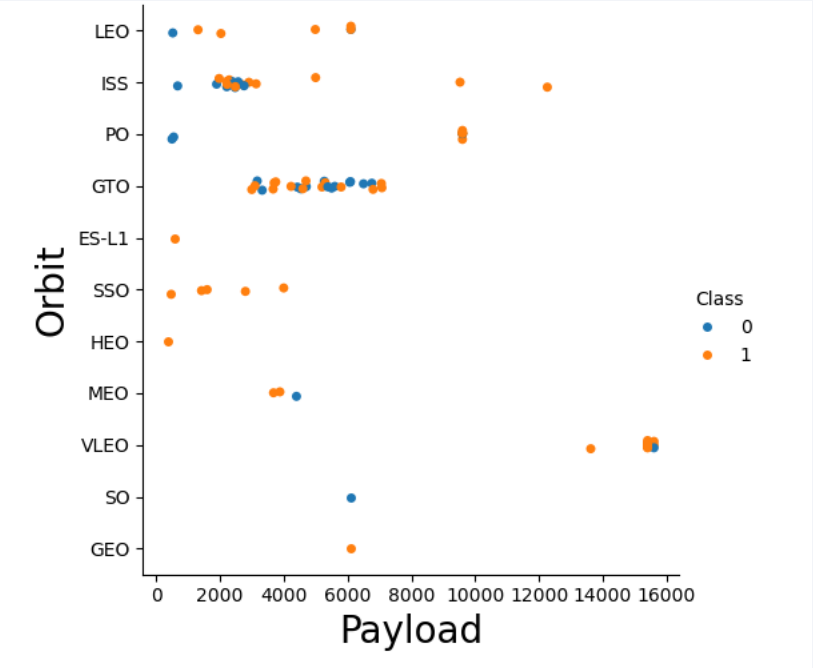
Flight Number vs. Orbit Type

- Success looks to improve over time in all orbits
- VLEO orbit seems to be a new business opportunity, due to an increase of its frequency



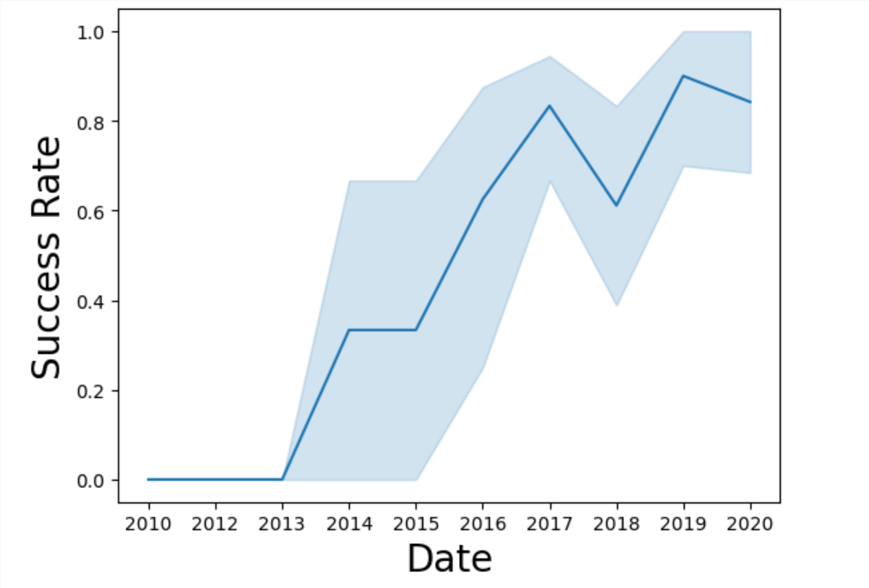
Payload vs. Orbit Type

- Does not look to be a relation between payload and success rate in GTO
- ISS orbit has the widest range of payload and a good rate of success
- There are a few launches to the orbits SO and GEO



Launch Success Yearly Trend

- Success rate starts to increase in 2013, slightly down in 2018, but rises again in 2019
- The first few year seems to be a period of adjustments and improvements



All Launch Site Names

- Selected the names of the unique launch sites in the space mission
- Query Used:
  - %sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL;

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

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with the string 'CCA’
- Using query:
  - %sql SELECT * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

- Total Payload obtained by summing all payload whose code contained ‘CRS’
- Query Used:
  - sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD FROM SPACEXTBL WHERE PAYLOAD LIKE '%CRS%';

TOTAL_PAYLOAD
111268

Average Payload Mass by F9 v1.1

- Filtered data by booster version and calculating the average payload by mass
- Query Used:
  - %sql SELECT AVG(PAYLOAD_MASS__KG_) FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1'

AVG(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

- Filtered data by successful landing outcome on ground pad and getting the minimum date
- Query Used:
    - %sql select min(DATE) from SPACEXTBL where Landing_Outcome = 'Success (ground pad)';

min(DATE)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- Selected booster version according to values between 4000 and 6000, and successful landing drop ship
- Query Used:
  - %sql SELECT BOOSTER_VERSION from SPACEXTBL WHERE LANDING_OUTCOME = 'Success (drone ship)' and PAYLOAD_MASS__KG_ >4000 and PAYLOAD_MASS__KG_ <6000;

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

Total Number of Successful and Failure Mission Outcomes

- Selected the count of successful and failed mission outcomes
- Query Used:
  - %sql select count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure (in flight)'

count(MISSION_OUTCOME)
99

Boosters Carried Maximum Payload

- Selected each booster version with maximum payload mass
- Query Used:
  - %sql SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT max(PAYLOAD_MASS__KG_) FROM SPACEXTBL);

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

2015 Launch Records

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2015-01-10	9:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)
2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone ship)

- Selected the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

- Query Used: %sql SELECT * FROM SPACEXTBL where Landing_Outcome = 'Failure (drone ship)' and substr(Date,0,5)='2015';

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

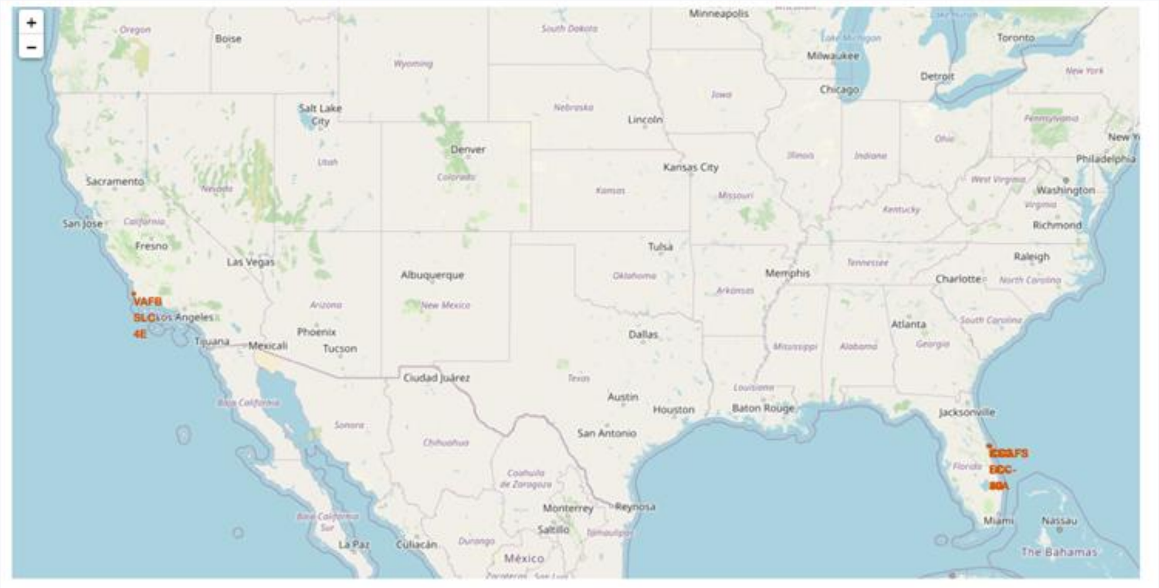
Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2016-07-18	4:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

- Ranked 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.
- Query Used:
  - %sql select * from SPACEXTBL where Landing_Outcome = 'Success (ground pad)' and (DATE between '2010-06-04' and '2017-03-20') order by date desc;

Section 3

Launch Sites Proximities Analysis

Lauch Sites

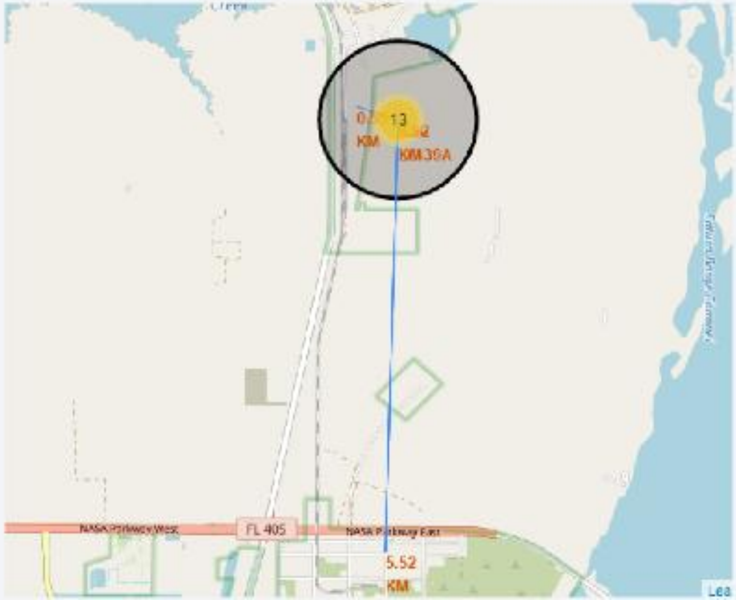


- Launch sites are near the sea, but not too far from roads and railroads

Lauch OutComes

- KSC LC-39A launch site outcomes
- Green markers show successful and red show failures

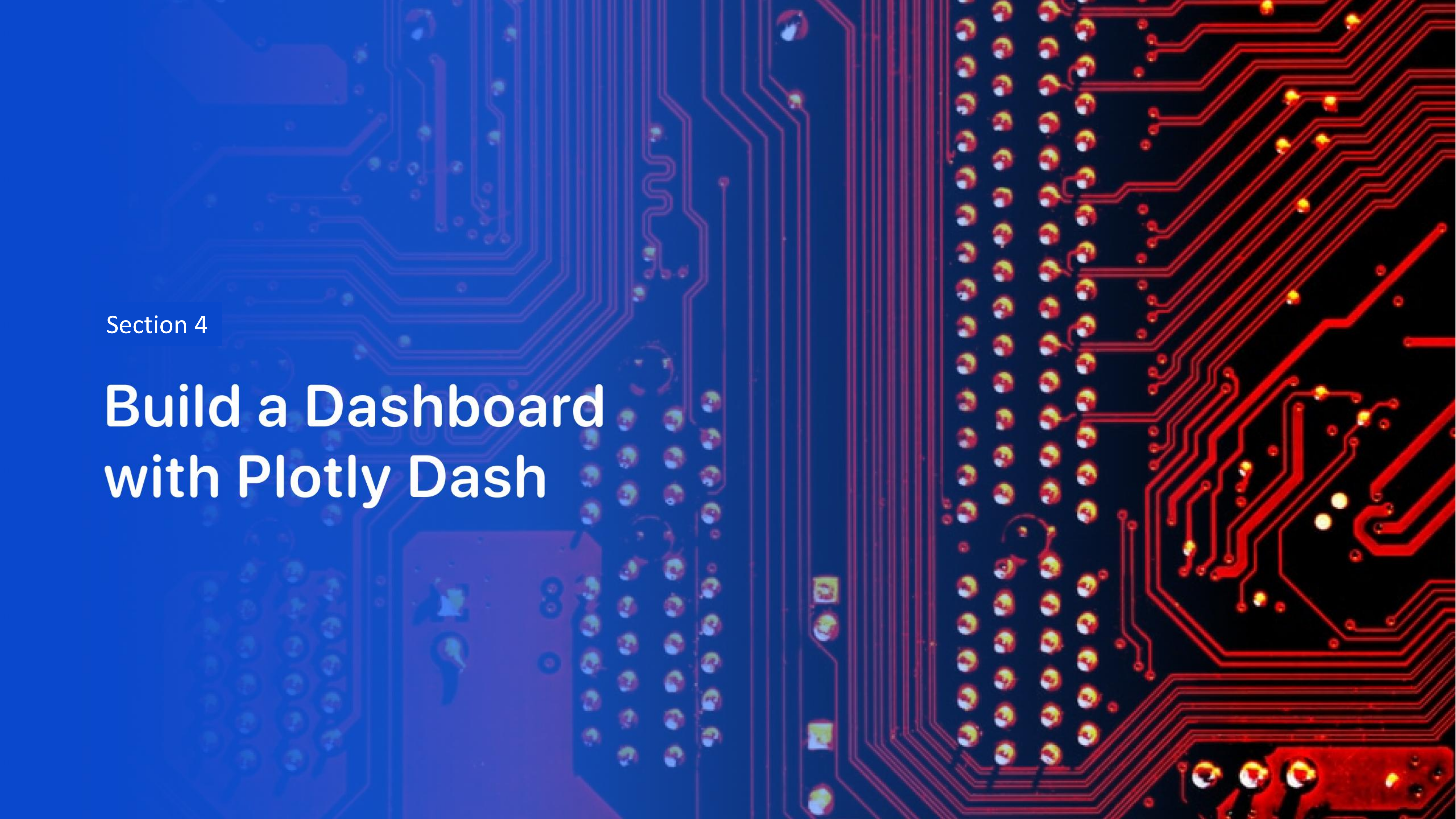
Distance from Railway, Highway, Coastline



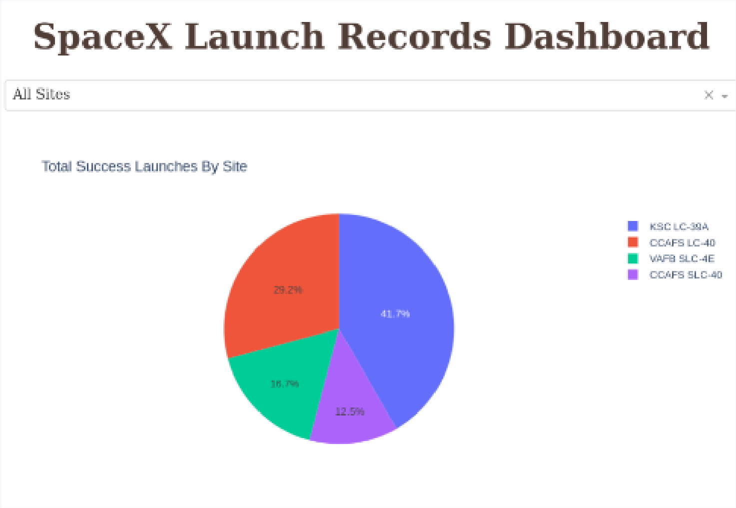
- KSC LC-39A is near railroad and road, relatively far from inhabited areas for safety

Section 4

Build a Dashboard with Plotly Dash

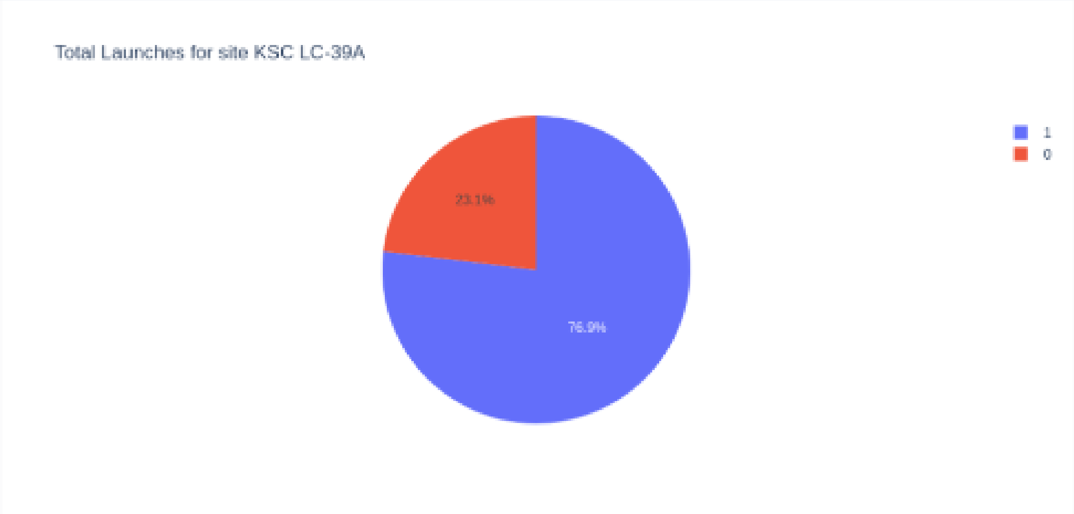


Successful Launches



- Pie charts shows the percentage a successful launches per site

Launches for KSC LC-39A



- KSC LC-39A has 76.9% successful launches

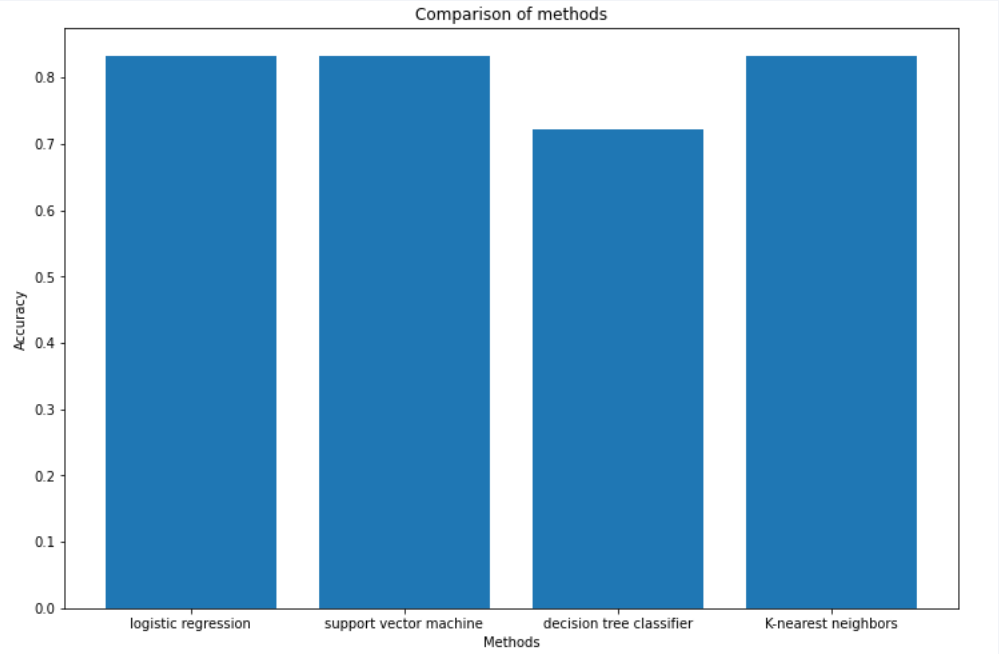
Payload vs. Launch Outcome

- Payloads under 6000 kg and FT boosters are the most successful combination

Section 5

Predictive Analysis (Classification)

Classification Accuracy



- **Logistic regression**, **Support vector machine** and **K-nearest neighbors**, all of these three methods give the best performance, with accuracy of 0.8333333333333334.

Confusion Matrix

- K Nearest Neighbors
- Has a high numbers for true positives and true negatives. Does has an issue with false positives

Conclusions

- Different data sources were analyzed
- Best launch is KSC LC-39A
- Launches above 7000 kg are less risky
- Most of the mission outcomes are successful, landing outcomes seem to improve over time according to the evolution processes and rocket development
- **Logistic regression**, **Support vector machine** and **K-nearest neighbors**, all of these three methods give the best performance, with accuracy of 0.8333333333333334.

Thank you!