IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

- SpaceY for rocket launching
  - Predict landing of first stage for Falcon 9
- Collect and clear a dataset
- Analyze the dataset for feature extracting
- Visualize and gain important insights
- Machine Learning for prediction

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaxeX API and webscraping
- Perform data wrangling
  - Perform exploratory Data Analysis and determine Training Labels
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Scikit-learn library, Grid search for tuning, Confusion matrix and accuracy

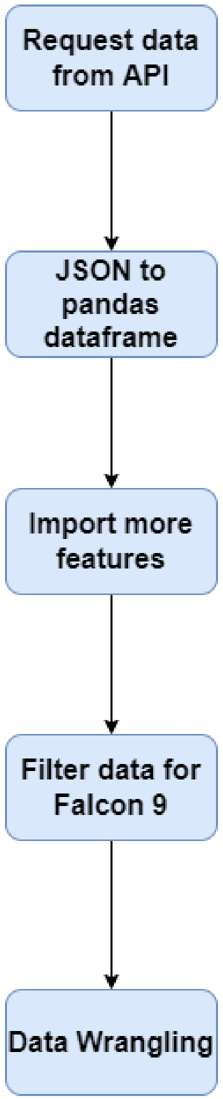
Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- Request data from SpaceX API
- Normalize JSON data to pandas DataFrame
- Import more features
- Data for Falcon 9
- Data Wrangling
    - Deal with missing values

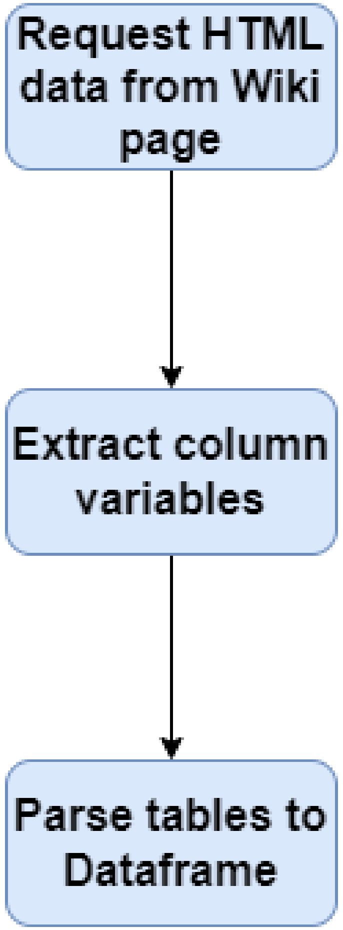
- https://github.com/tsakas96/IBM_Data_Science_Capstone_SpaceY/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Request HTML data from Wiki page for Falcon 9
- Extract column variables from table
- Parse tables into pandas DataFrame

- https://github.com/tsakas96/IBM_Data_Science_Capstone_SpaceY/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Calculate missing values
- Calculate the number of launches on each site
- Calculate the number and occurrence of each orbit
- Calculate the number and occurrence of mission outcome per orbit type
- Create a landing outcome label from Outcome column

- https://github.com/tsakas96/IBM_Data_Science_Capstone_SpaceY/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Scatter Plot
    - Relationship between two features e.g., Payload mass vs Flight number
- Bar plot
    - Success rate for orbit type
- Line plot
    - Success yearly trend

- https://github.com/tsakas96/IBM_Data_Science_Capstone_SpaceY/blob/main/jupyter-labs-eda-dataviz.ipynb

EDA with SQL

- Queries for launch site column e.g., unique names, etc.
- Queries for payload mass e.g., maximum, total
- Date of the first successful outcome
- Complex queries
    - Successful landings in drone ship in a certain range of payload mass
    - Count of the outcomes
    - Booster versions with maximum payload
    - Failed landing outcomes of 2015 and count of outcomes in specific timeline
- https://github.com/tsakas96/IBM_Data_Science_Capstone_SpaceY/blob/main/jupyter-labs-eda-sql-coursera.ipynb

Build an Interactive Map with Folium

- Circle object
  - Mark Nasa and launch site locations
- Marker object
  - Mark Nasa and launch site locations with their names
- Line Object
  - Draw a straight line between two locations
- MarkerCluster object
  - Includes mark and circle objects of close-range locations

- https://github.com/tsakas96/IBM_Data_Science_Capstone_SpaceY/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

- Dropdown
  - Select the Launch Sites
- RangeSlider
  - Select min and max range of Payload Mass
- Pie plot
  - Plot the total success landings of Launch Sites between a given Payload Mass range
- Scatter Plot
  - Plot the correlation between Payload and Success for Launch Sites

- https://github.com/tsakas96/IBM_Data_Science_Capstone_SpaceY/blob/main/spacex_dash_app.py

Predictive Analysis (Classification)

- Scikit-learn Library
  - Logistic Regression
  - Support Vector Machines (SVM)
  - Decision Tree
  - K Nearest Neighbor (KNN)
- Split dataset into train and test sets
- Grid Search for parameters of each model
- Calculate accuracy for test set

- https://github.com/tsakas96/IBM_Data_Science_Capstone_SpaceY/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

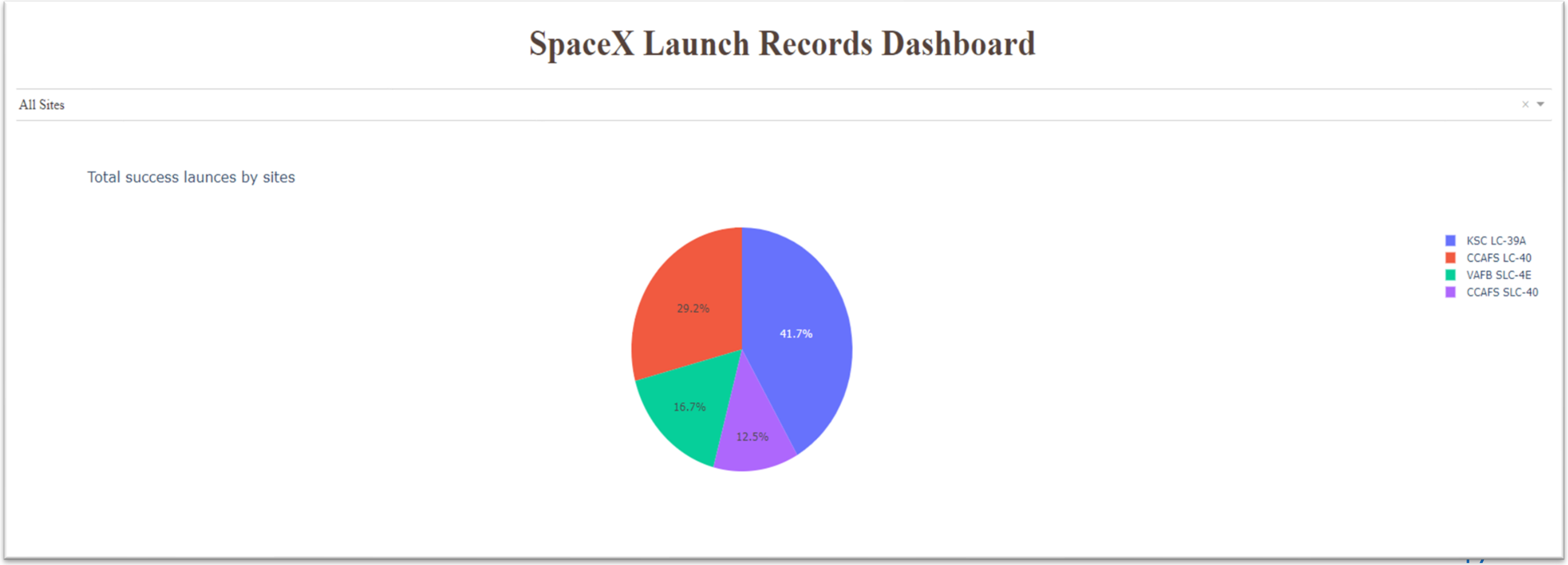
Results

- Exploratory data analysis results

	FlightNumber	PayloadMass	Orbit	LaunchSite	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial
0	1	6104.959412	LEO	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B0003
1	2	525.000000	LEO	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B0005
2	3	677.000000	ISS	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B0007
3	4	500.000000	PO	VAFB SLC 4E	1	False	False	False	NaN	1.0	0	B1003
4	5	3170.000000	GTO	CCAFS SLC 40	1	False	False	False	NaN	1.0	0	B1004

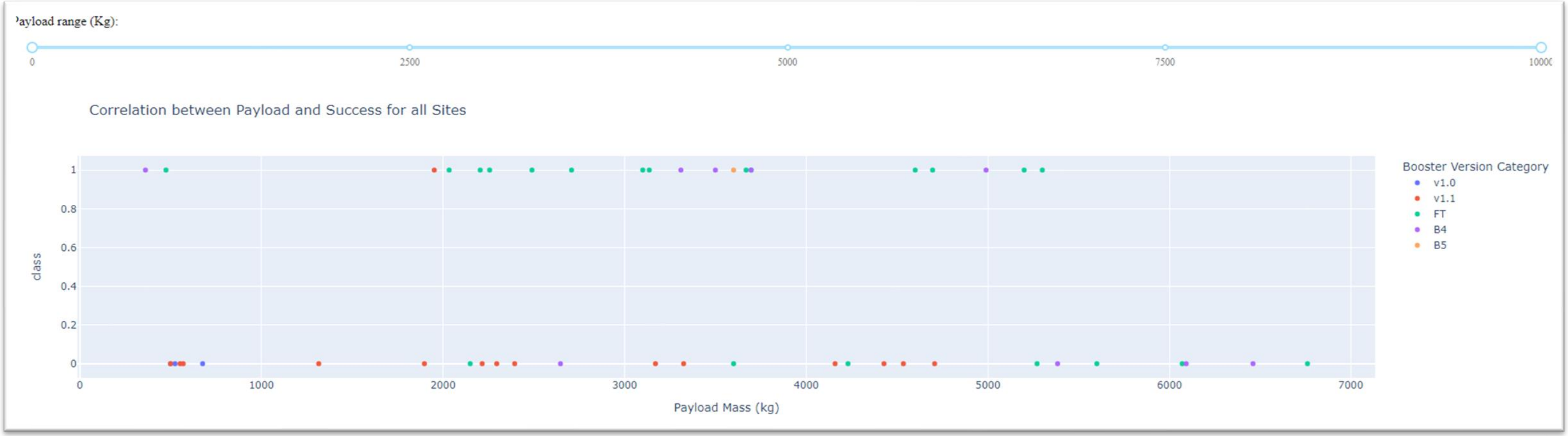
Results

- Interactive analytics demo in screenshots (1/2)



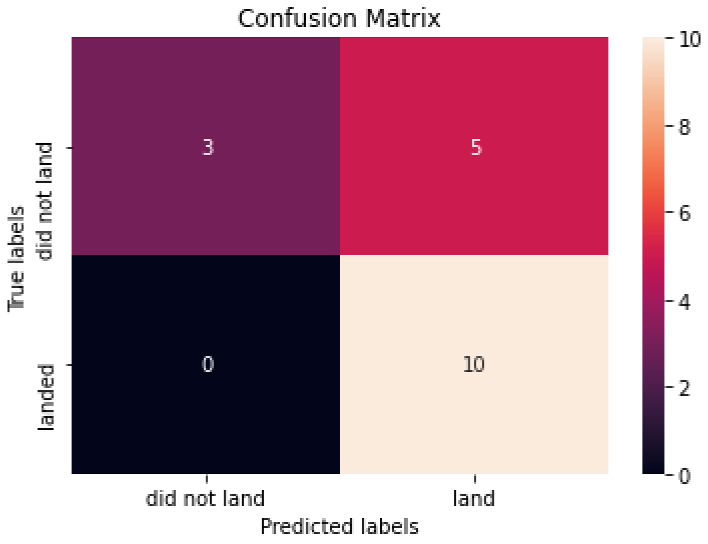
Results

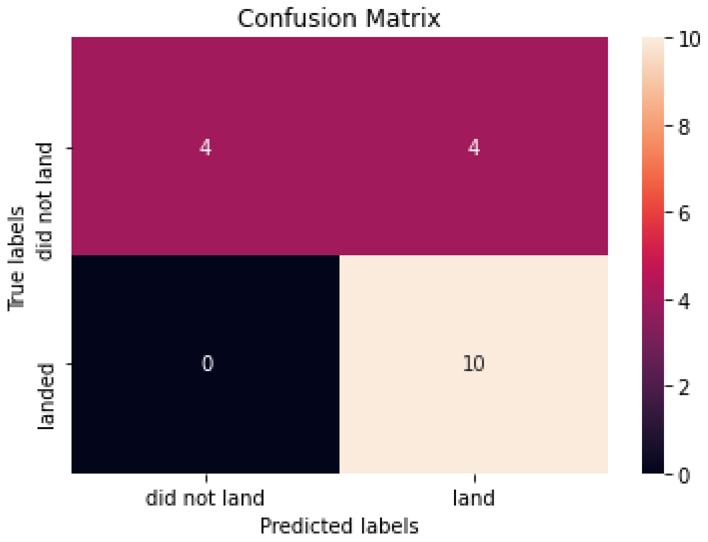
- Interactive analytics demo in screenshots (2/2)



Results

- Predictive analysis results (1/2)



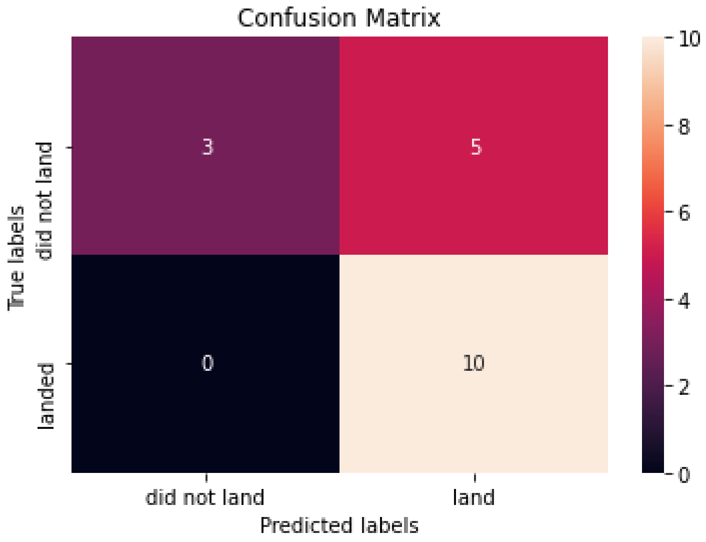


Logistic Regression accuracy: 0.722

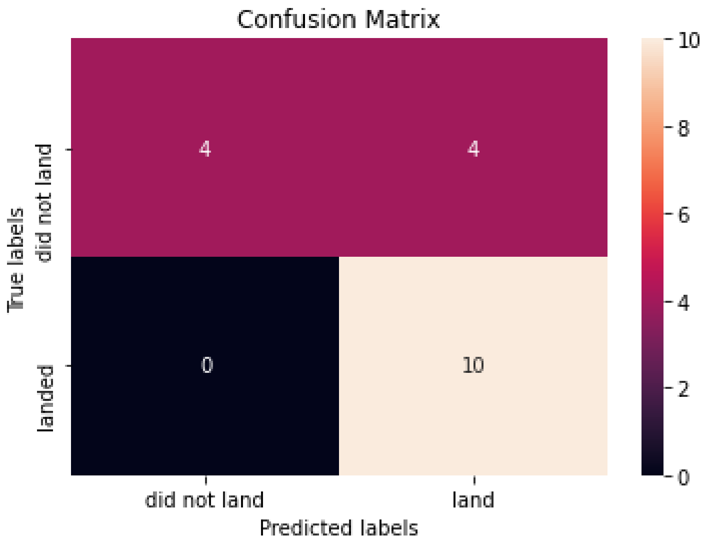
Support Vector Machines accuracy: 0.778

Results

- Predictive analysis results (2/2)



Decision Tree accuracy: 0.722



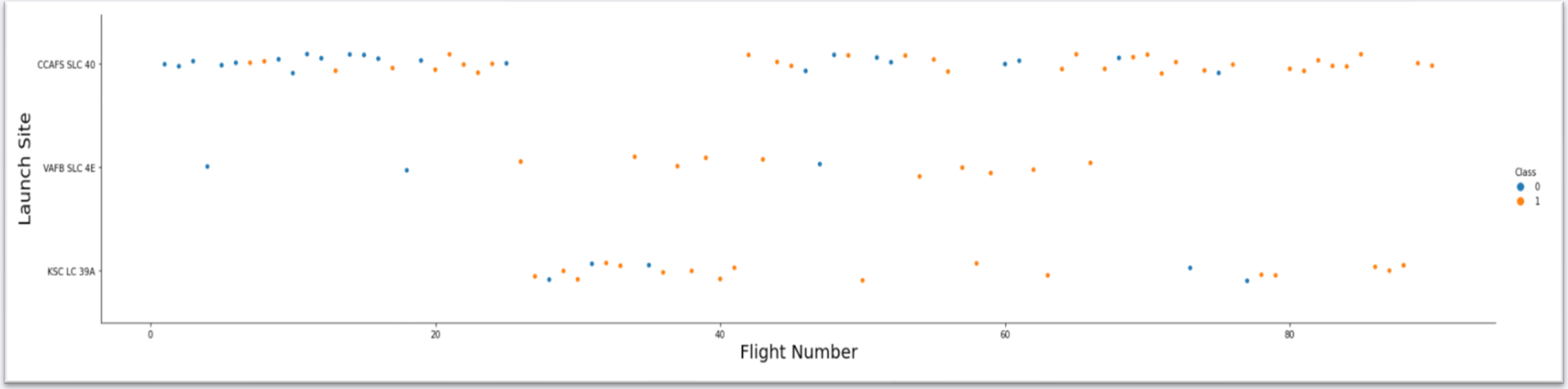
K Nearest Neighbor accuracy: 0.778

Section 2

Insights drawn from EDA

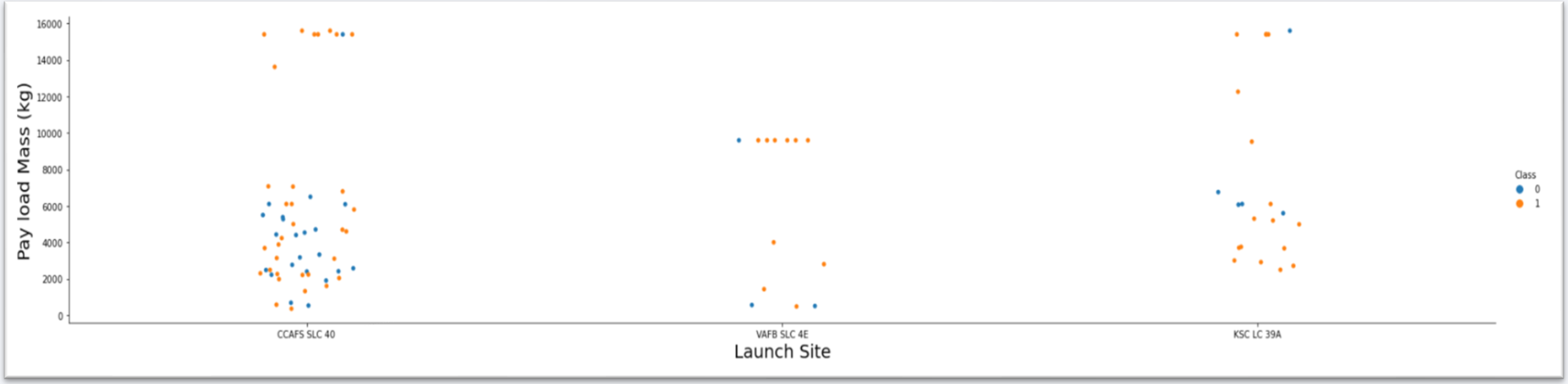
Flight Number vs. Launch Site

- CCAFS SLC 40 with higher number of flights has more successful landings
- VAFB SLC 4E has more successful landings



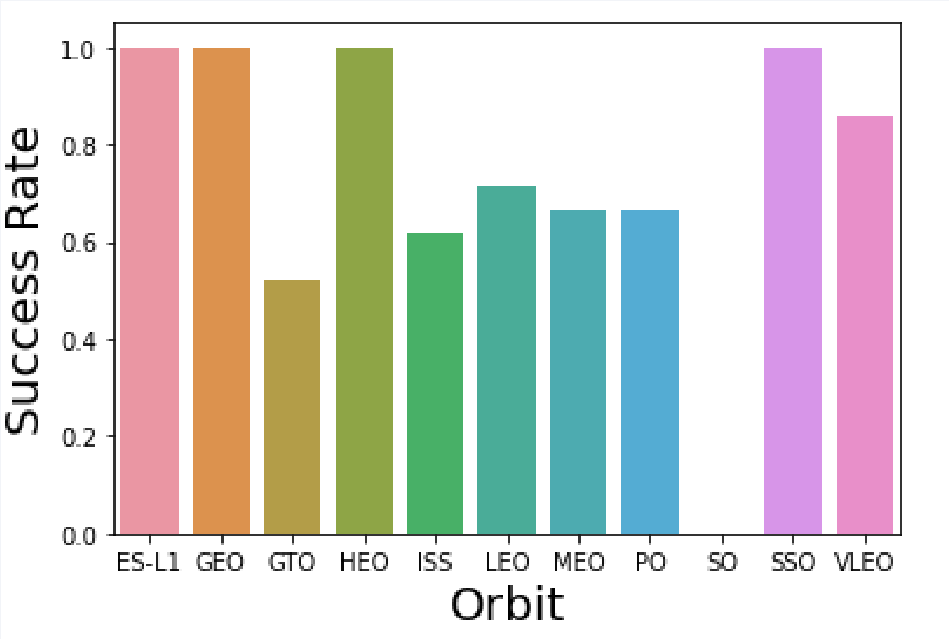
Payload vs. Launch Site

- Higher payload mass has more successful landings for all sites
- VAFB-SLC launch site there are no rockets launched for payload mass(greater than 10000)
- KSC-LC launch site unsuccessful landings for payload mass(around 6500)



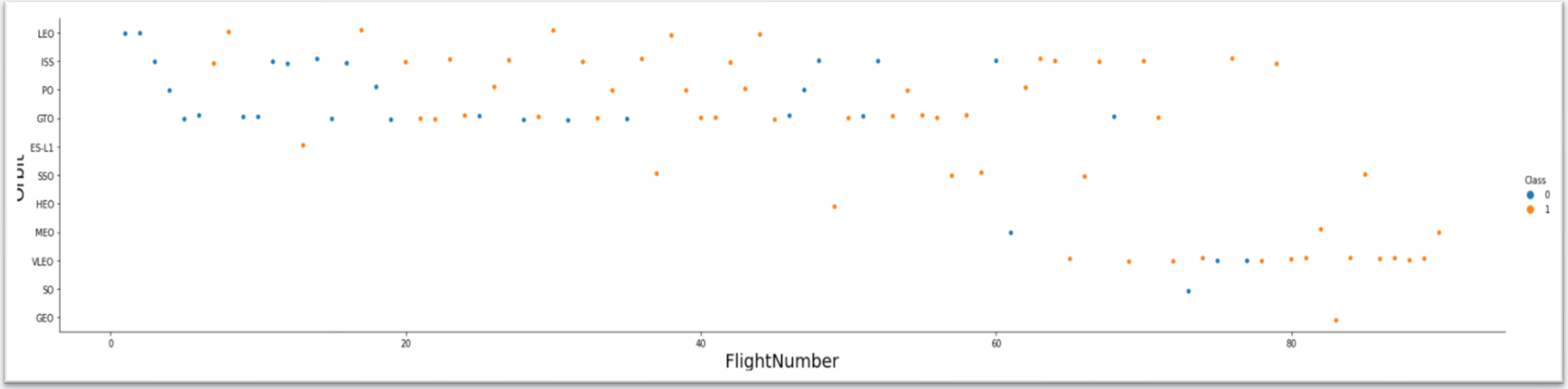
Success Rate vs. Orbit Type

- ES-L1, GEO, HEO and SSO with success rate of 1
- SO with success rate of 0
- Minimum success rate of approx. 0.52 not including SO



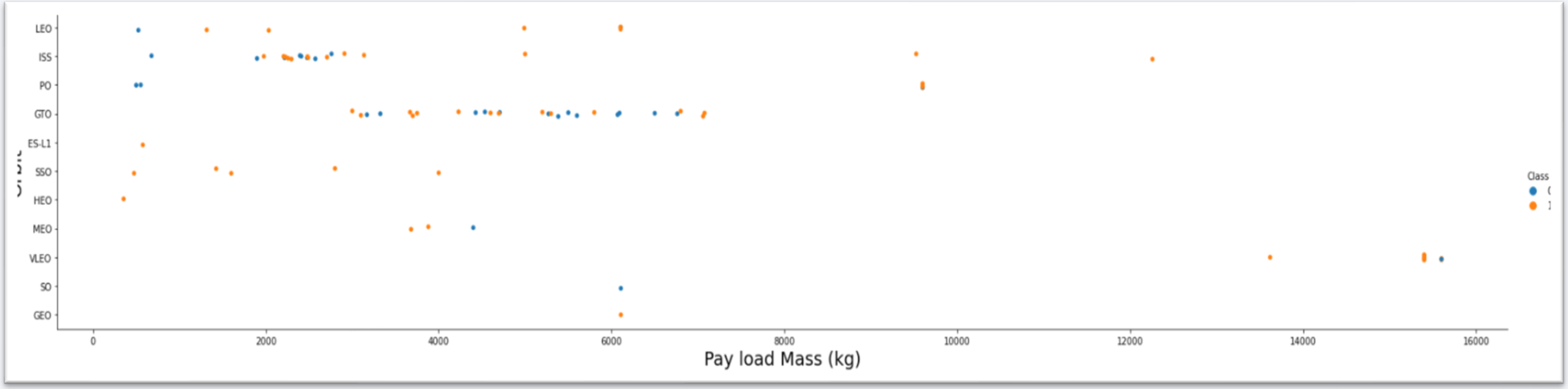
Flight Number vs. Orbit Type

- LEO orbit the Success appears related to the number of flights
- No relationship between flight number when in GTO orbit



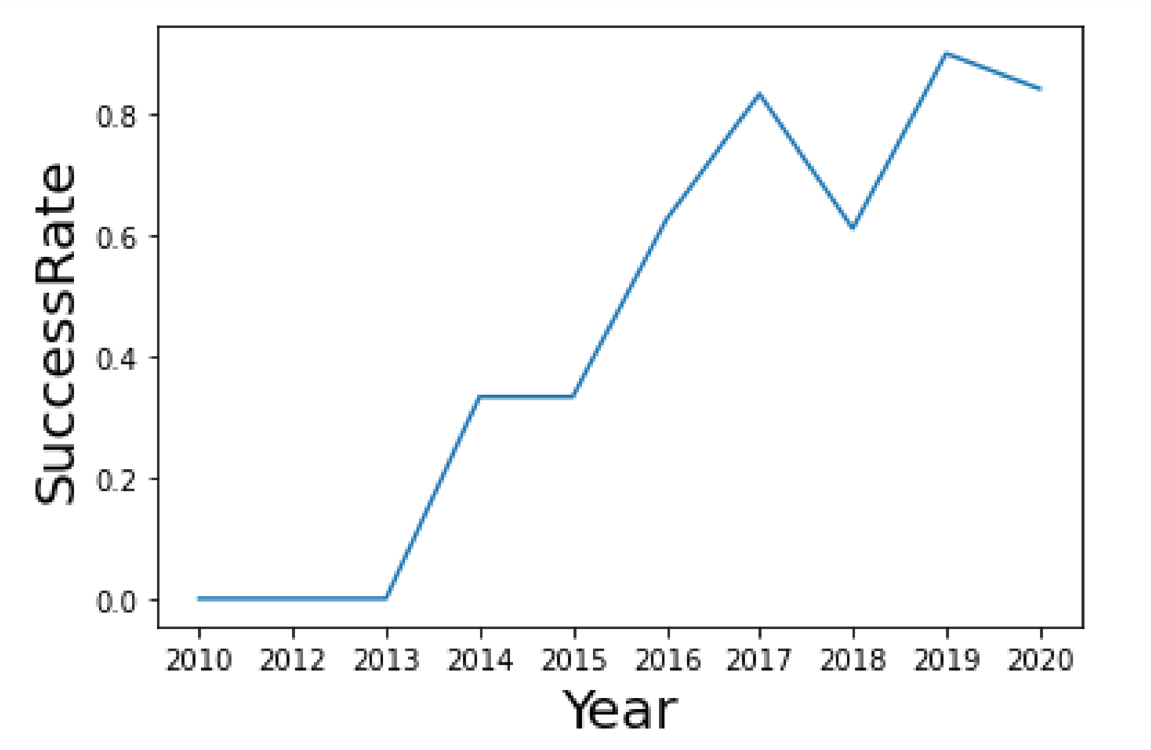
Payload vs. Orbit Type

- For heavy payloads, the successful landing or positive landing rate are more for Polar, LEO and ISS
- No relationship between payload and landing outcome in GTO orbit



Launch Success Yearly Trend

- Success rate since 2013 kept increasing till 2020



All Launch Site Names

- Find the names of the unique launch sites
  - 4 Launch Sites

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin 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

Total Payload Mass

- Calculate the total payload carried by boosters from NASA

total_payload
45596

Average Payload Mass by F9 v1.1

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

avg_payload
2928

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

first_success_date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

booster_version	payload_mass__kg_	Landing _Outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

mission_outcome	COUNT
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

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

booster_version	launch_site	Landing _Outcome	DATE
F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)	2015-01-10
F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)	2015-04-14

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

- 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

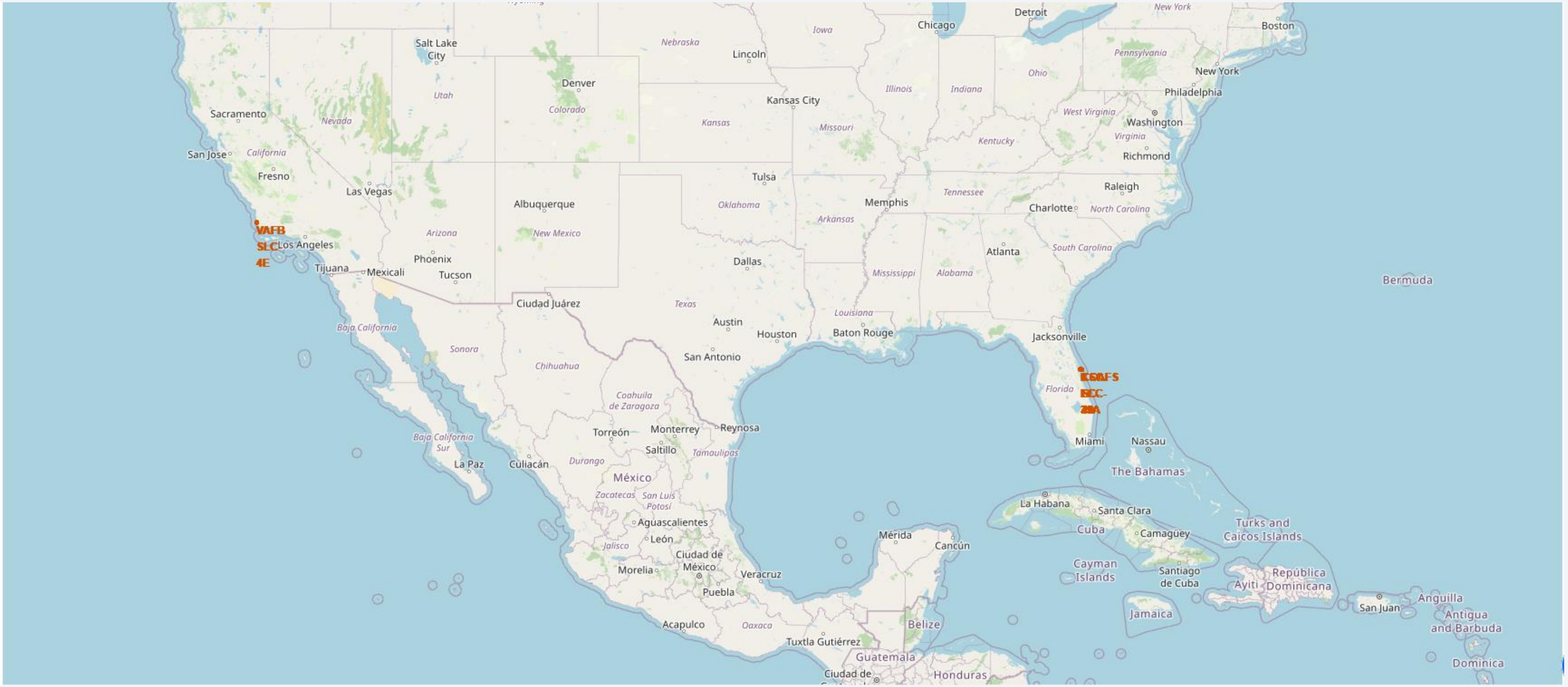
Landing _Outcome	COUNT
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 4

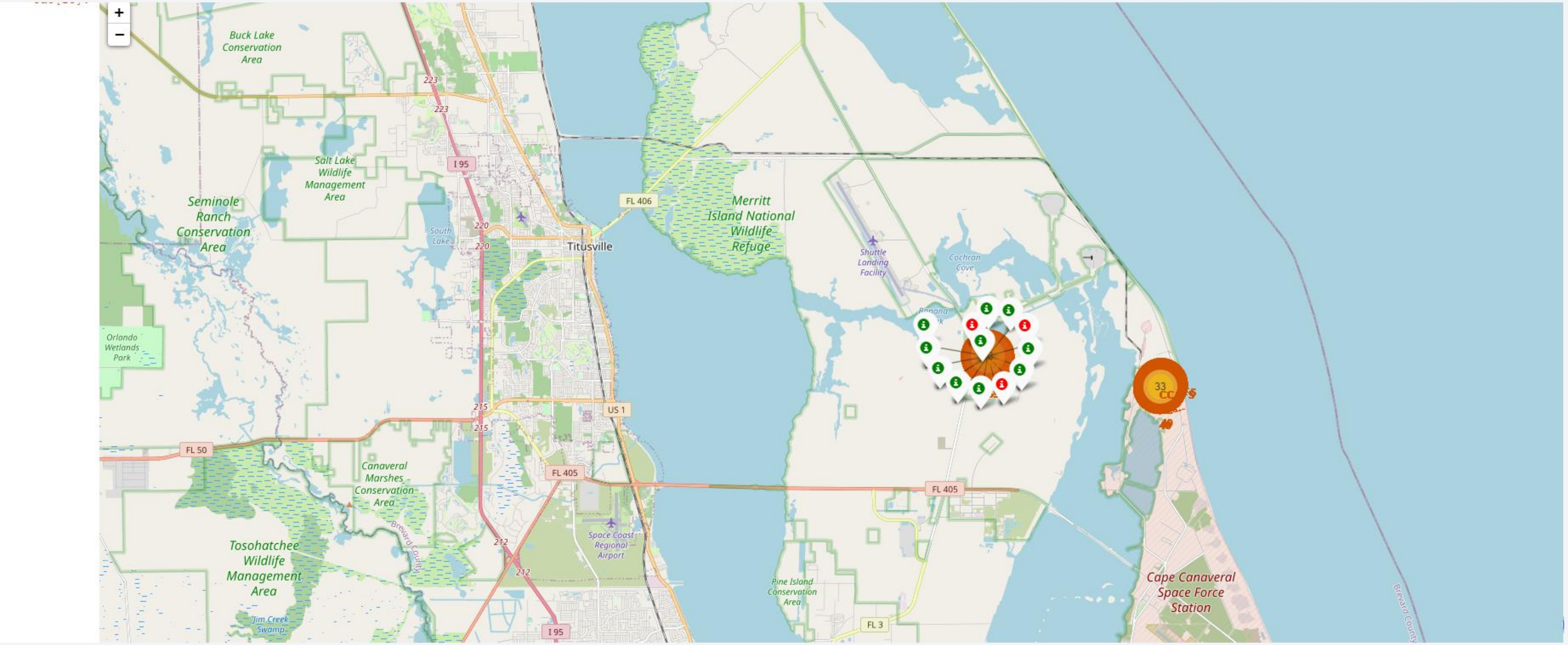
Launch Sites Proximities Analysis

Launch Site Locations

3 Launch Site locations close to each other

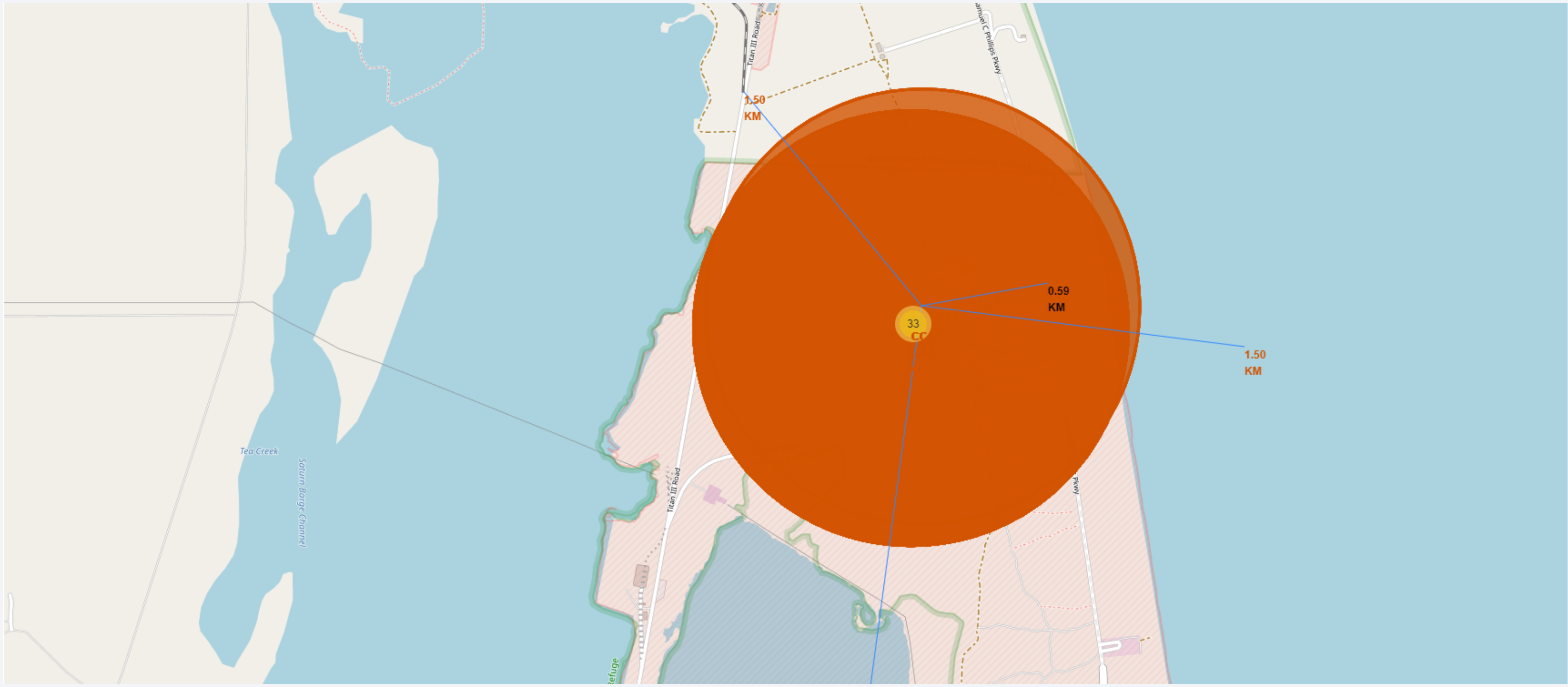


Landing Outcomes of KSC LC-39A



Location Proximities

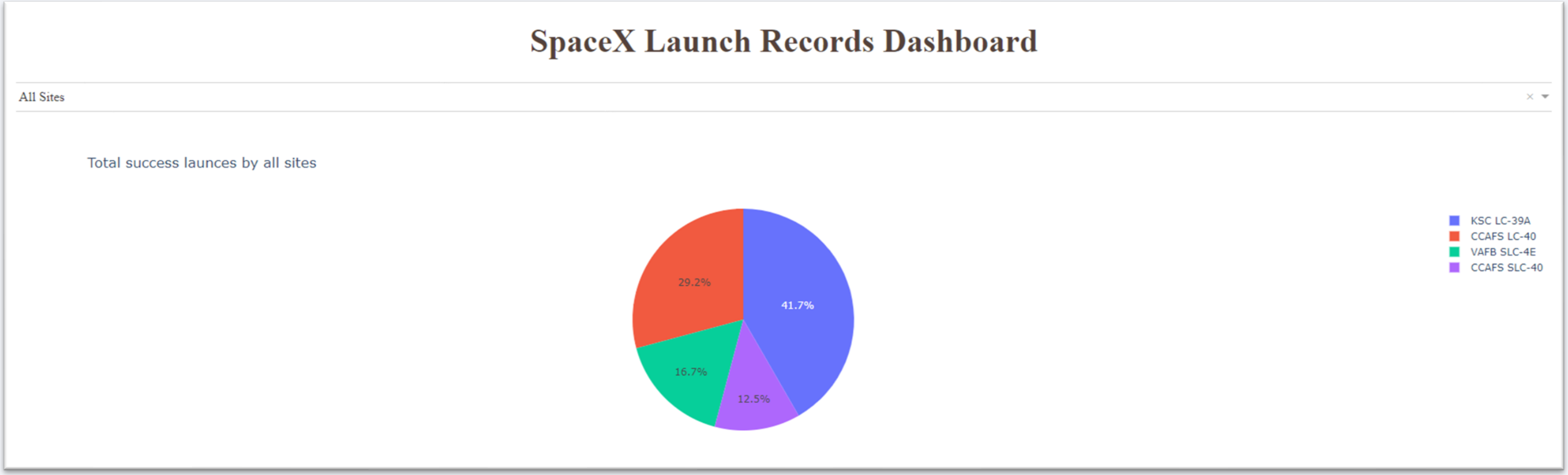
- 0.6 KM from highway
- 1.5 KM from railway
- 57.71 KM from Melbourne



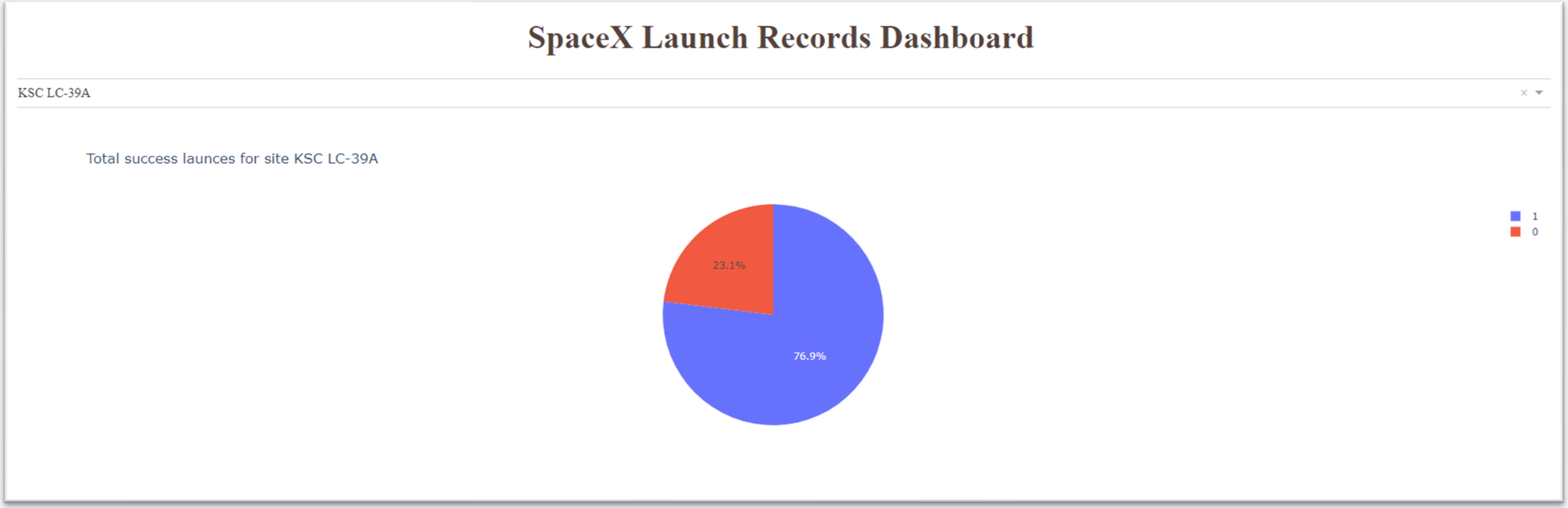
Section 5

Build a Dashboard with Plotly Dash

Total success by all launch sites

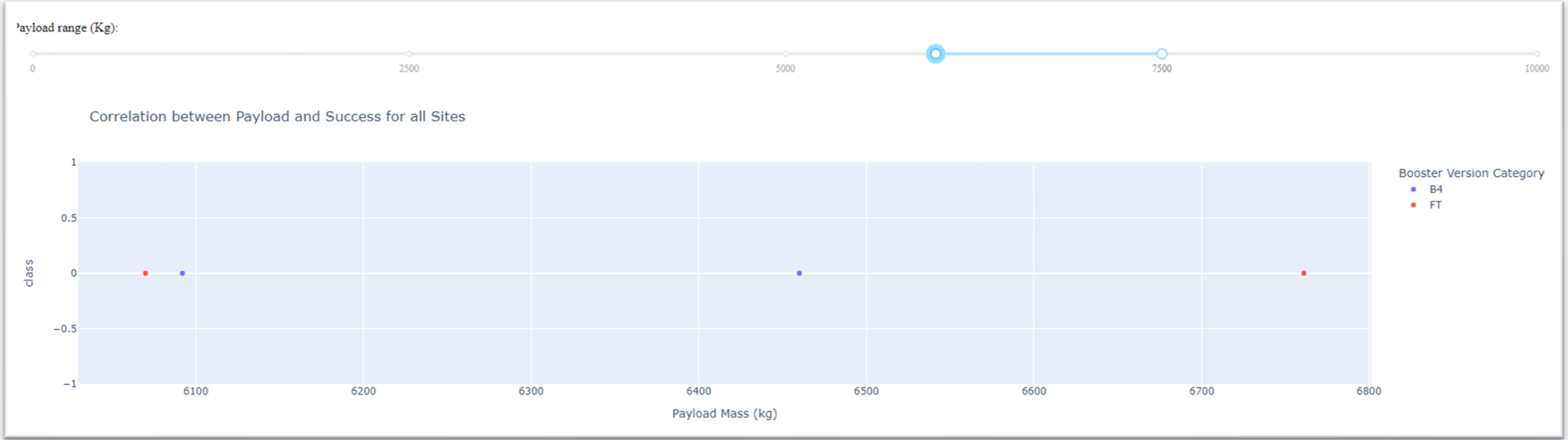


Launch site with highest launch success (KSC LC-39A)



Landing outcomes for payload from 6000 to 7500

For payload from 6000 to 7500 all landings failed

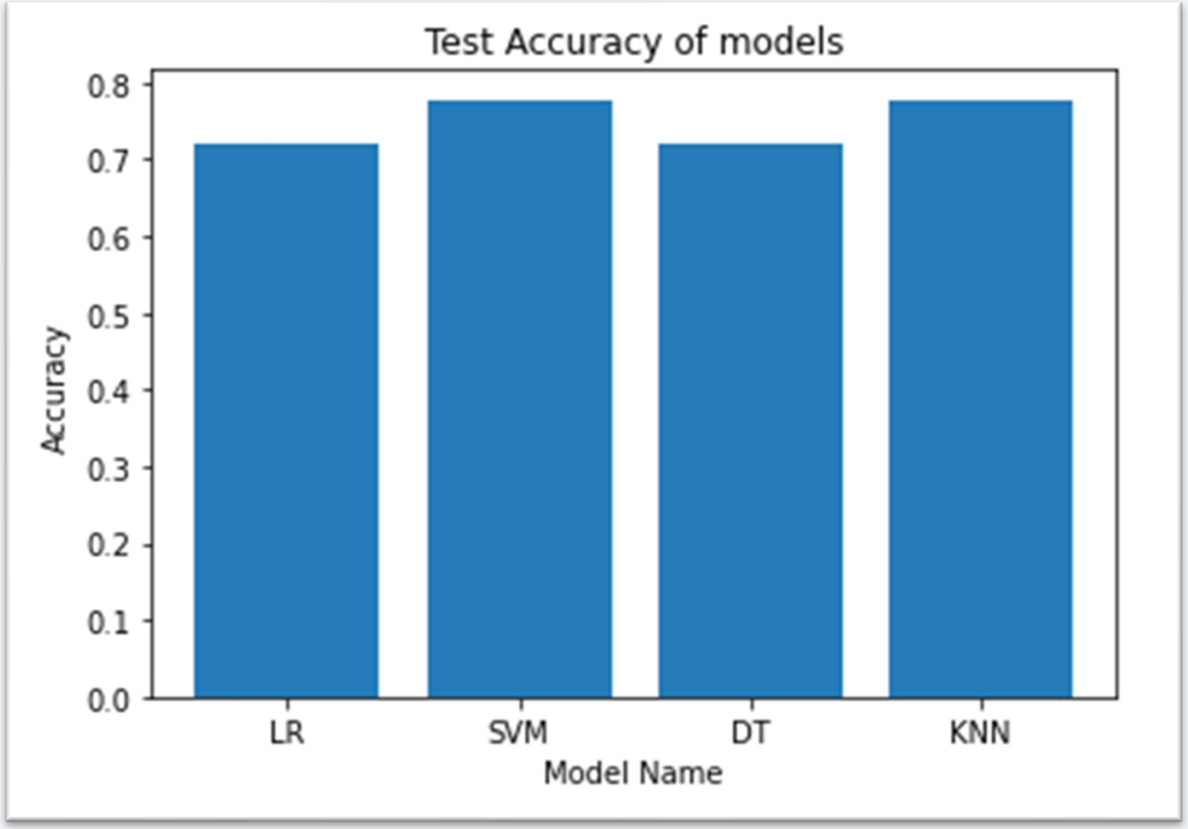


Section 6

Predictive Analysis (Classification)

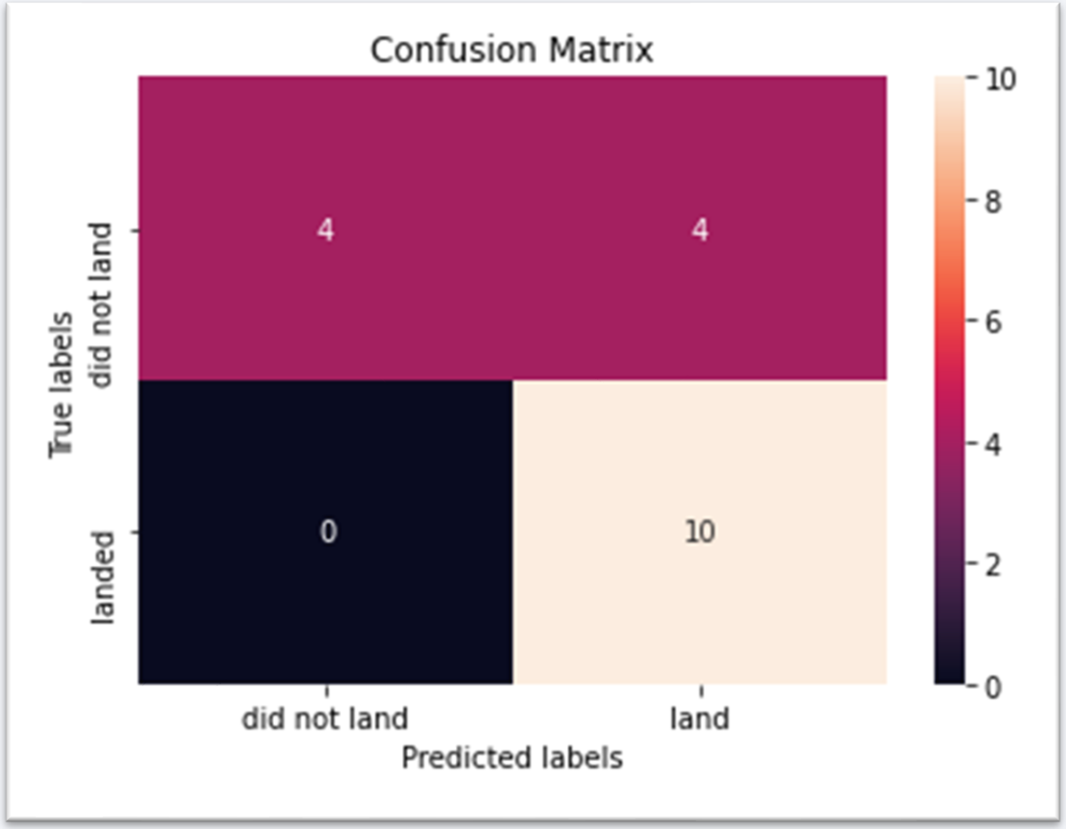
Classification Accuracy

- SVM and KNN have the same accuracy of 0.778



Confusion Matrix

- 4 out of 8 correct predictions for positive outcome
- 10 out of 10 correct predictions for negative outcome



Conclusions

- Data are limited
- Data analysis is important for feature extraction and prediction
- Models perform relatively well
- Important insights for placing launch sites
- Improvement of accuracy with state-of-the-art algorithms

Appendix

- Repo link
  - https://github.com/tsakas96/IBM_Data_Science_Capstone_SpaceY

Thank you!