IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ivana Kolorici-Livnjak
2023

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection through API
  - Data Collection with Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Data Visualization
  - Interactive Visual Analytics with Folium
  - Machine Learning Prediction
- Summary of all results
  - Exploratory Data Analysis result
  - Interactive analytics in screenshots
  - Predictive Analytics result

Introduction

Project background and context

- The main objective is to predict whether the Falcon 9 rocket will land successfully.
- SpaceX Falcon 9 rocket launches cost 62 million dollars, while other providers cost upward of 165 million dollars each. SpaceX reuses the first stage, allowing it to save more money.
- By determining the success of the first stage 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.

Problems to find answers to

- What influences if SpaceX will reuse the first stage?
- What model should be used for binary classification?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data from Space X was obtained from 2 sources:
  - Space X API (https://api.spacexdata.com/v4/rockets/)
  - Web Scraping (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)
- Perform data wrangling
  - A landing outcome label was created after analyzing different features

Methodology

Executive Summary

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Data was normalized and divided into training and testing sets
  - Different models were tested and the confusion matrix was examined

Data Collection

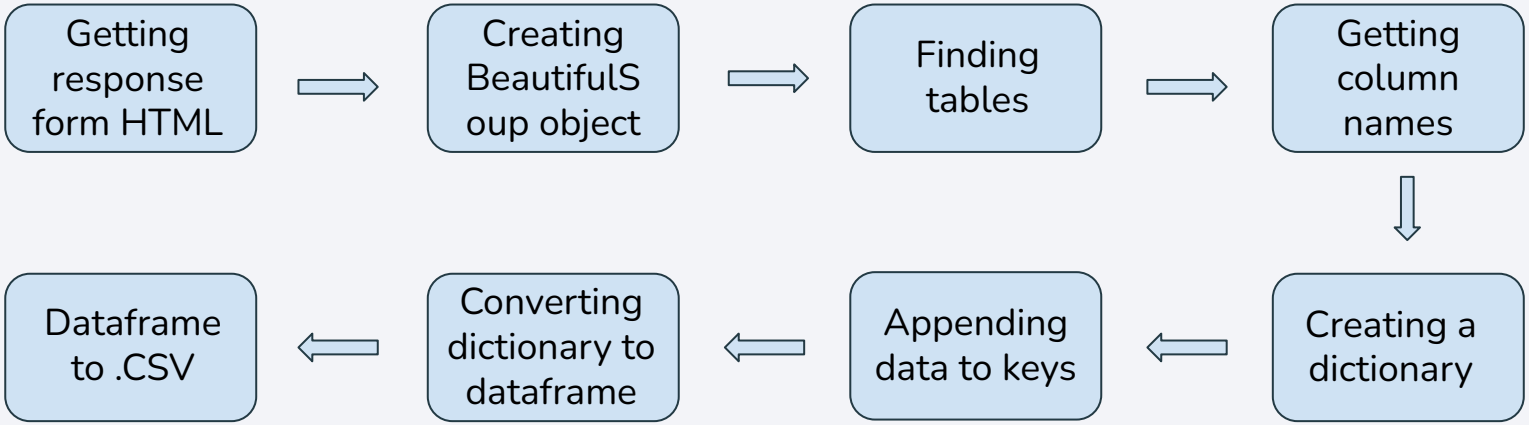
SpaceX launch data was obtained from SpaceX REST API and Wikipedia using BeautifulSoup.

Data Collection – SpaceX API

- The GitHub URL of the completed SpaceX API calls notebook:

  https://github.com/koloric/Data_Science_capstone/blob/bb24d5784633454988f85284d9bfafa2a88286a6/Data_Science_capstone/1_collecting_data/jupyter-labs-spacex-data-collection-api.ipynb

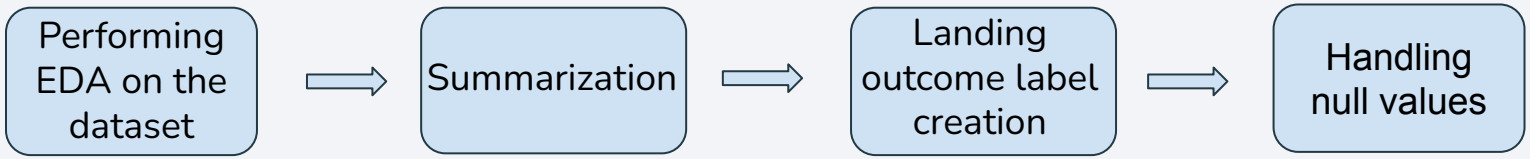
Data Collection - Scraping



- The GitHub URL of the completed web scraping notebook:

https://github.com/koloric/Data_Science_capstone/blob/bb24d5784633454988f85284d9bfafa2a88286a6/Data_Science_capstone/1_collecting_data/jupyter-labs-webscraping.ipynb

Data Wrangling

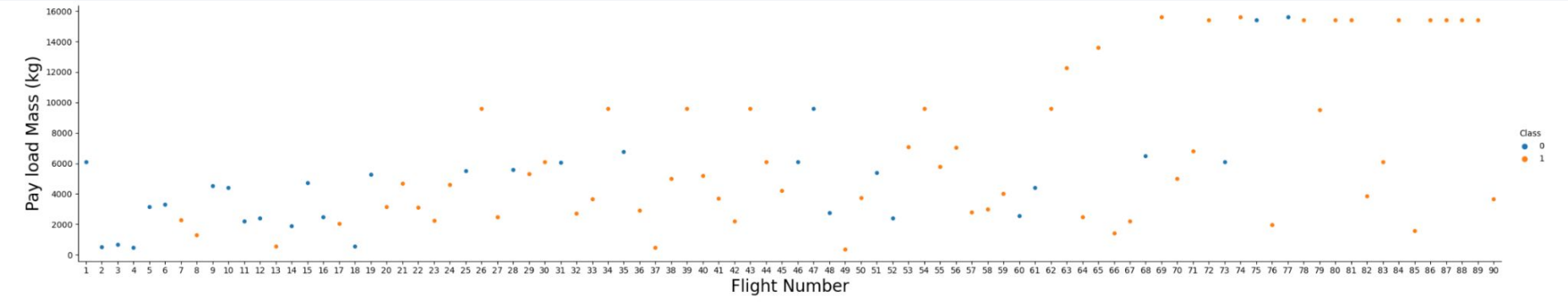


- The GitHub URL of the completed data wrangling notebook:

https://github.com/koloric/Data_Science_capstone/blob/bb24d5784633454988f85284d9bfafa2a88286a6/Data_Science_capstone/2_data_wrangling/Lab2_Data_wrangling.ipyn

EDA with Data Visualization

Scatter plots and bar plots were used to visualize relationships between variables as:
Payload mass, flight number, Launch site and orbit



- The GitHub URL of the completed EDA and Data Visualization notebook:

https://github.com/koloric/Data_Science_capstone/blob/bb24d5784633454988f85284d9bfafa2a88286a6/Data_Science_capstone/4_exploratory_analysis_pandas_matplotlib/Exploring_Preparing_Data.ipynb

EDA with SQL

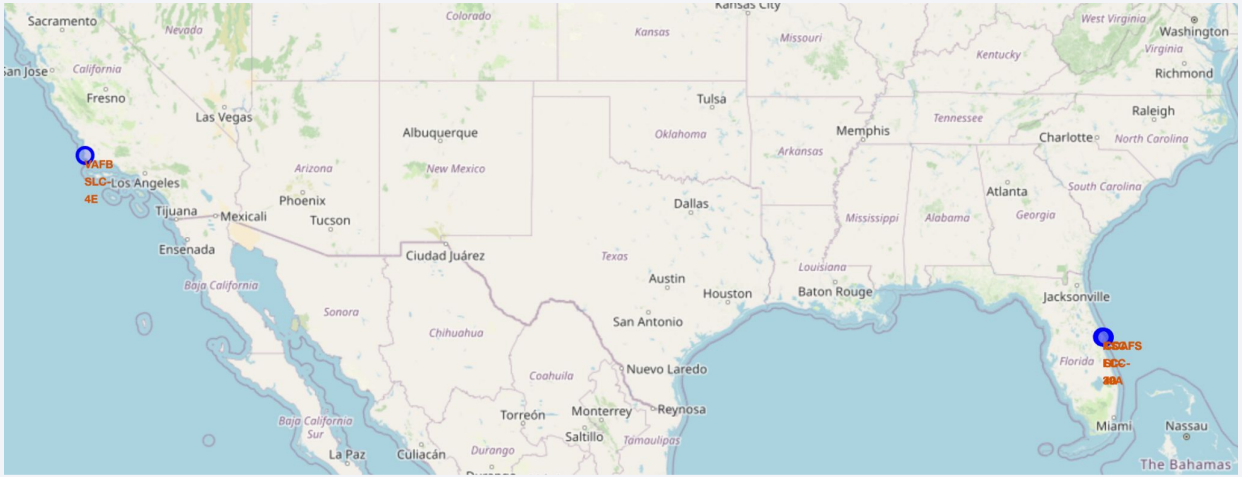
SQL queries performed include:

- Names of unique launch sites
- Top 5 records starting with string ‘CCA’
- Total payload mass carried by booster launched by NASA (CRS)
- Average payload mass carried by booster F9 v1.1
- First successful landing outcome date
- The names of the boosters which have success in drone ships and have payload mass greater than 4000 but less than 6000
- Total number of successful and unsuccessful mission outcomes
- Names of boosters that carried the max payload mass
- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for the year 2015
- Rank of the count of landing outcomes between 2010-06-04 and 2017-03-20

https://github.com/koloric/Data_Science_capstone/blob/bb24d5784633454988f85284d9bfafa2a88286a6/Data_Science_capstone/3_exploratory_analysis_sql/EDA_SQL.ipynb

Build an Interactive Map with Folium

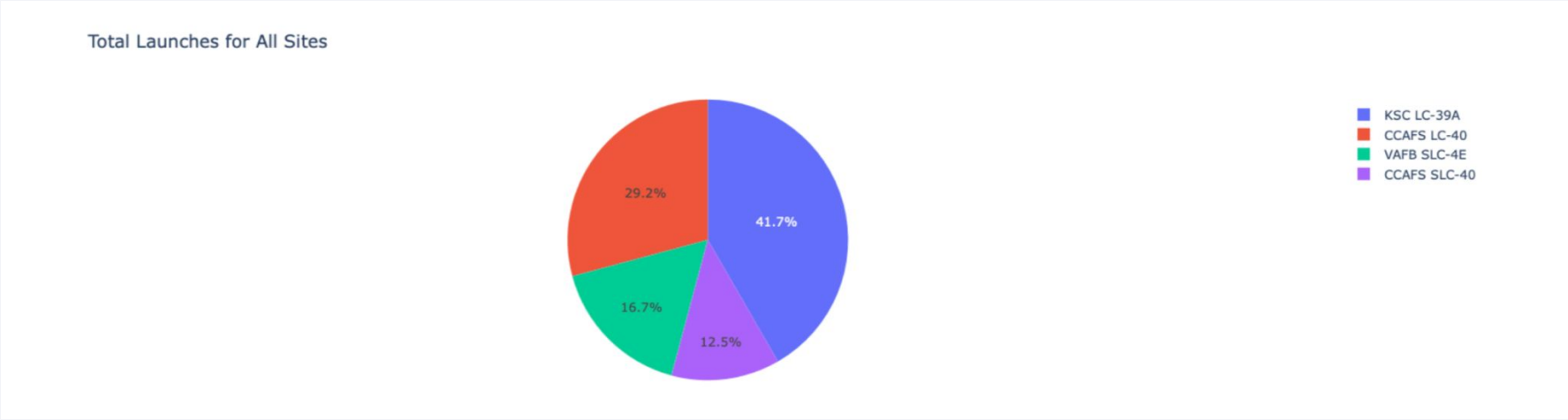
Markers, circles, lines and clusters were added to find the optimal launch site



- https://github.com/koloric/Data_Science_capstone/blob/bb24d5784633454988f85284d9bfafa2a88286a6/Data_Science_capstone/5_visual_analytics_dashboard/launch_site_location.ipynb

Build a Dashboard with Plotly Dash

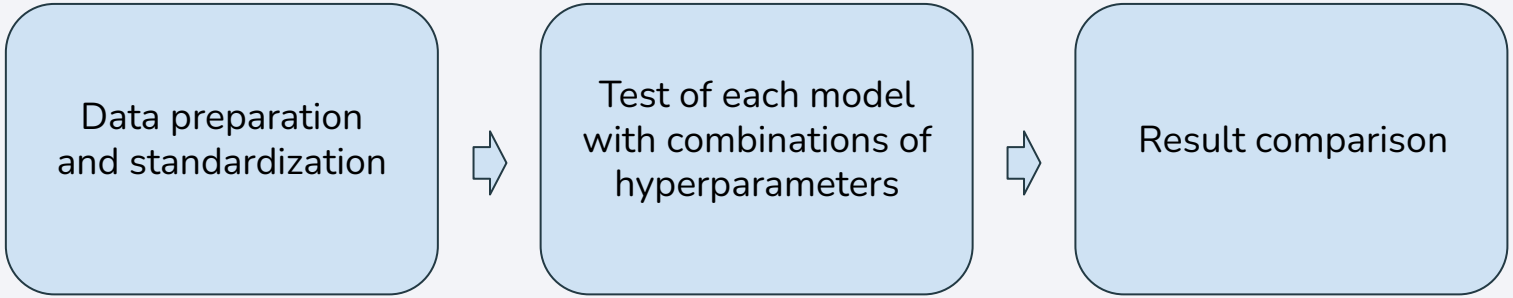
Graphs and plots help us with analyzing the relation between launch sites and payloads and thereby help with predicting the best launching site.



- https://github.com/koloric/Data_Science_capstone/blob/bb24d5784633454988f85284d9bfafa2a88286a6/Data_Science_capstone/5_visual_analytics_dashboard/spacex_dash_app.py

Predictive Analysis (Classification)

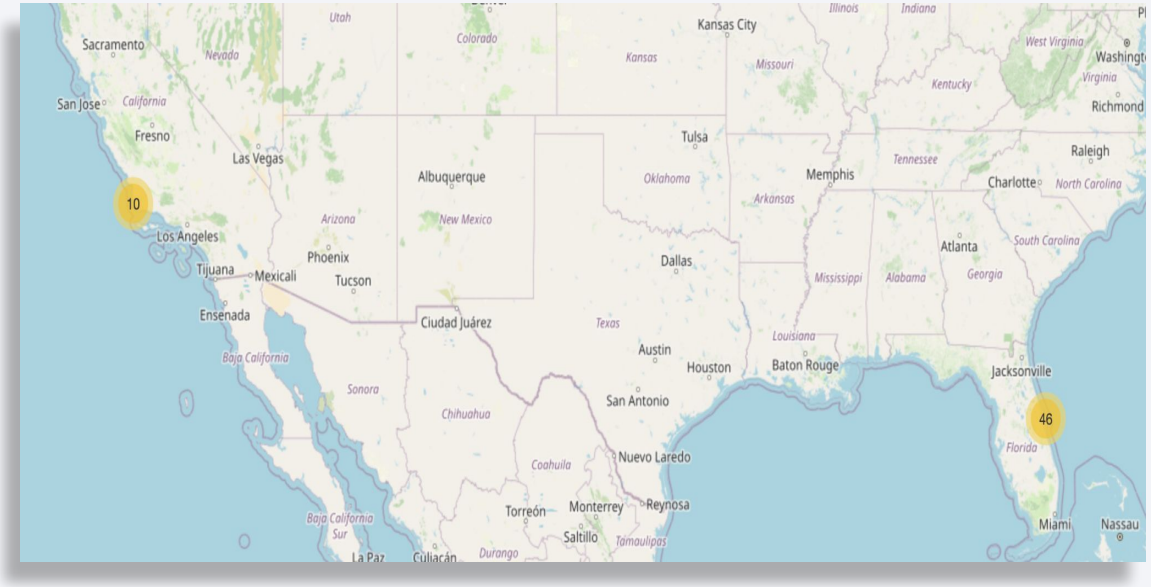
Four classification models were compared: logistic regression, support vector machine, decision tree and k nearest neighbors.



https://github.com/koloric/Data_Science_capstone/blob/bb24d5784633454988f85284d9bfafa2a88286a6/Data_Science_capstone/6_predictive_analysis/Machine%20Learning%20Prediction-2.ipynb

Results

- Exploratory data analysis results
  - 4 different launch sites are in use
  - The average payload mass is 2 534kg
  - The first successful landing was in 2015 (5y after the first launch)
  - The number of landing outcomes became better with each year

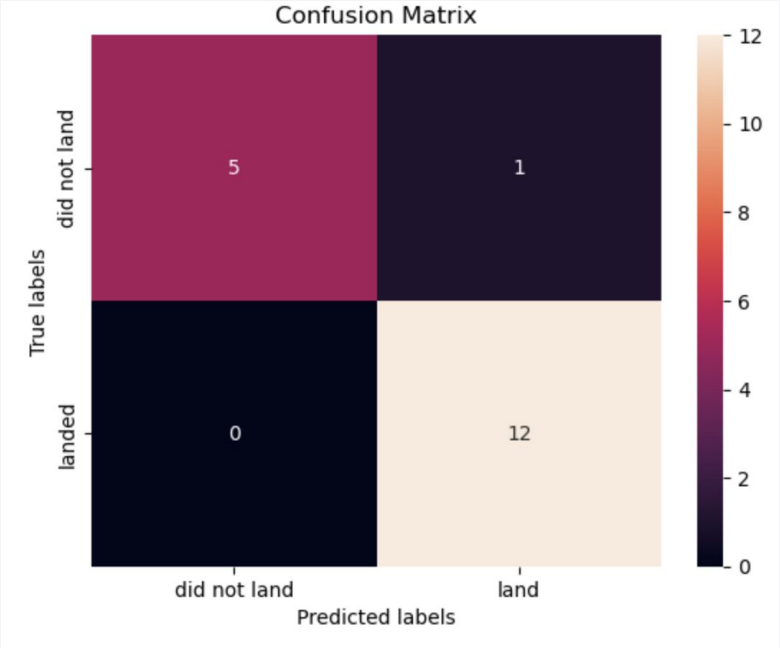


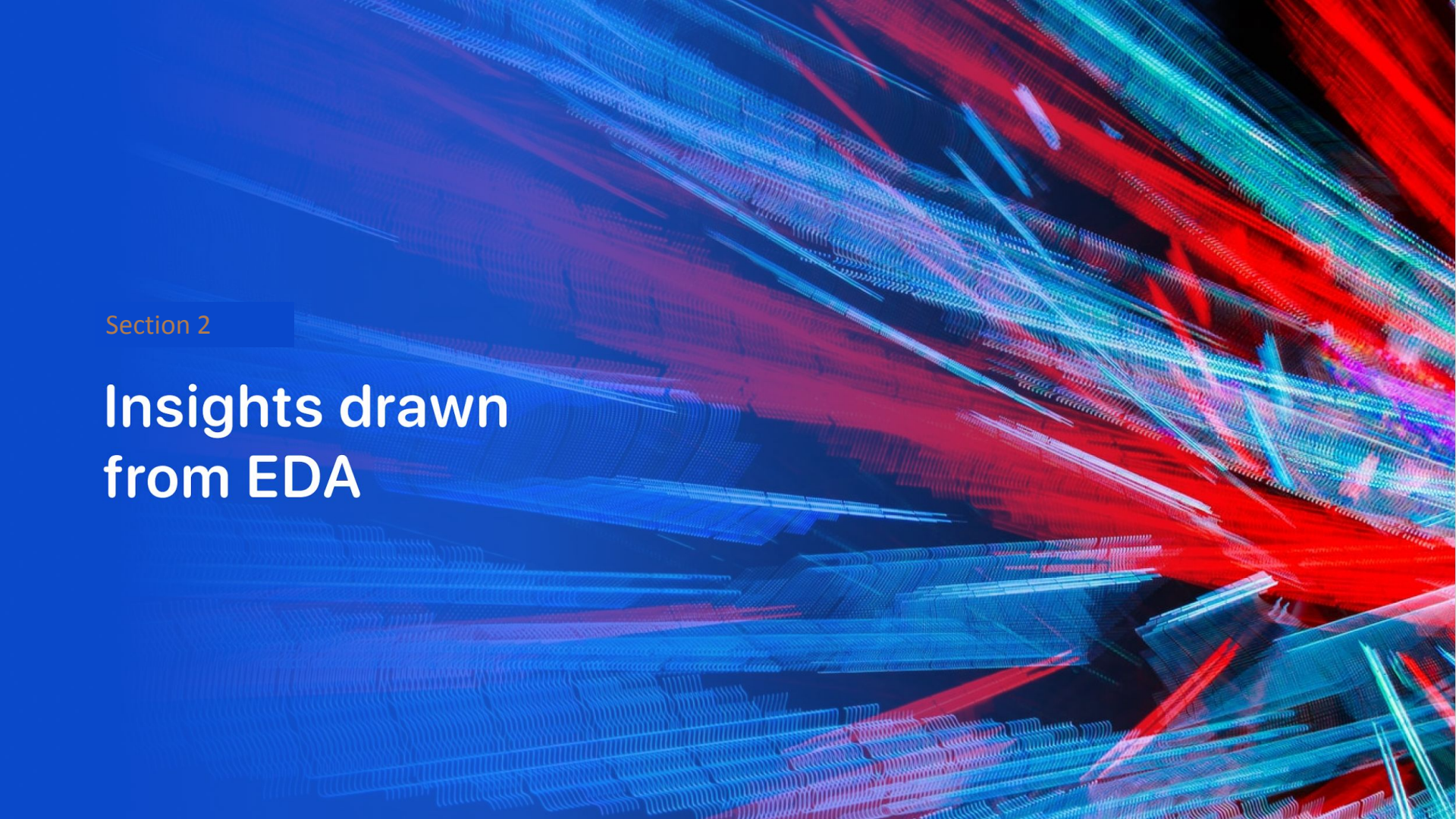
- Interactive analytics showed that launch sites are near sea in places with good infrastructure.
- Most launches happen at the east coast.

Results

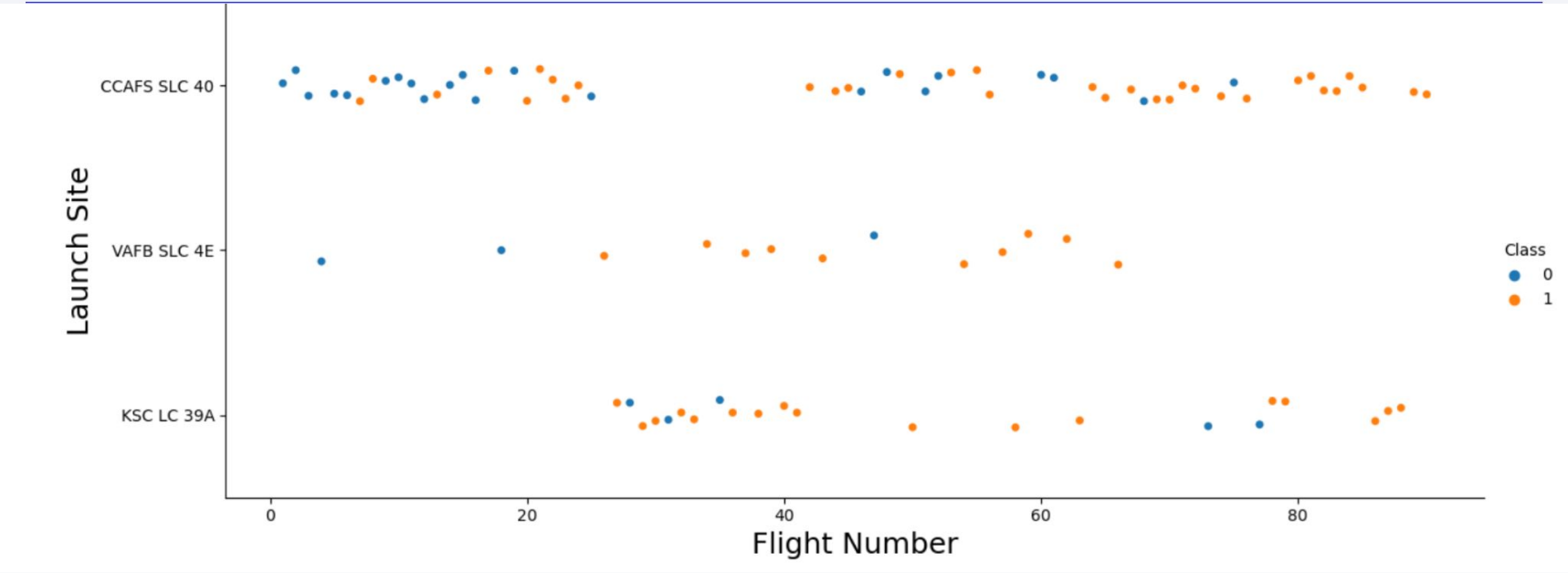
- Predictive analysis results

	Accuracy Train	Accuracy Test
Logreg	0.846429	0.833333
Svm	0.848214	0.833333
Tree	0.889286	0.944444
Knn	0.848214	0.833333



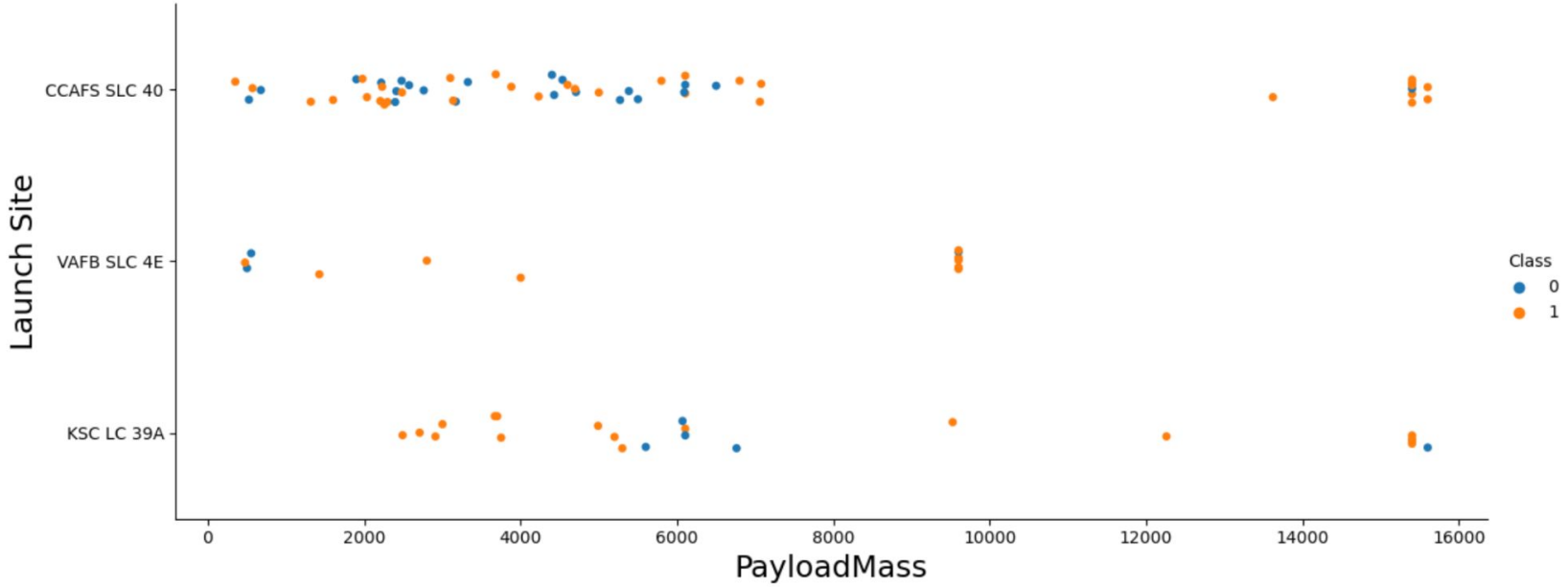

Section 2
Insights drawn
from EDA

Flight Number vs. Launch Site



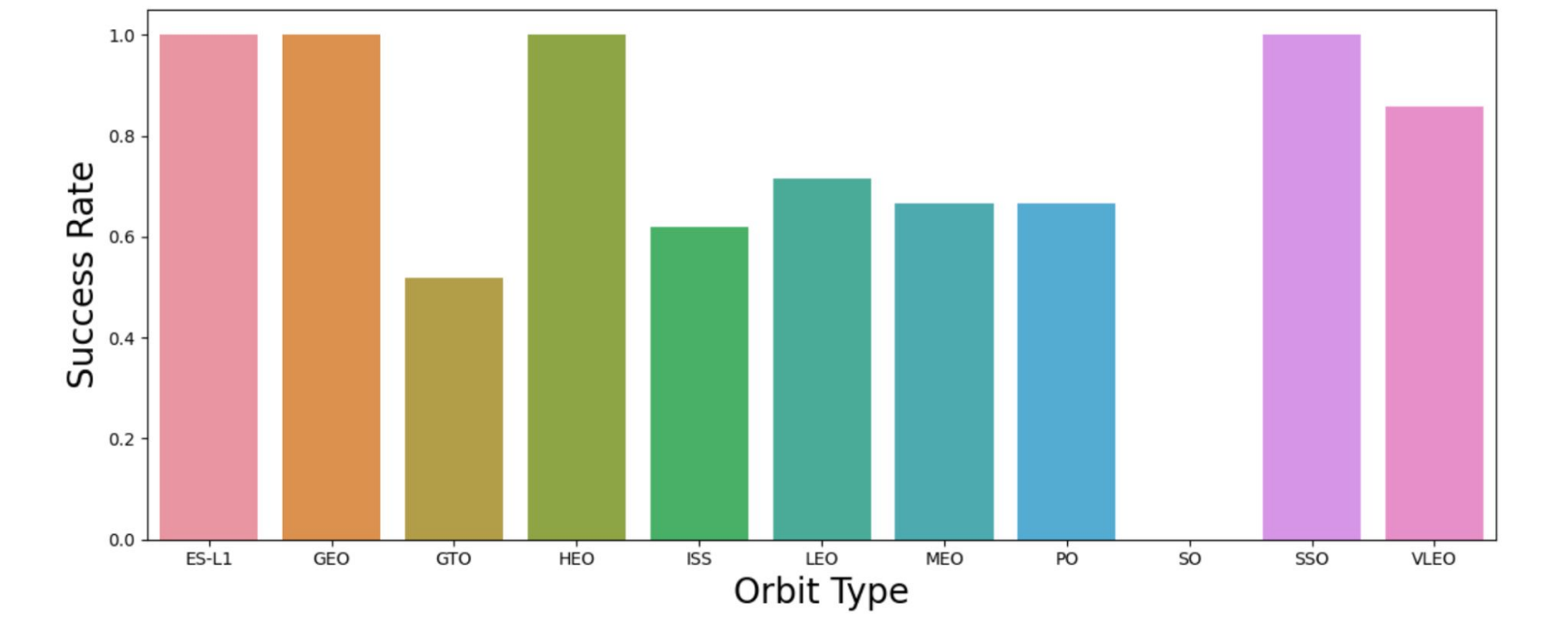
- Higher Flight Numbers have a better success rate
- Most used Lunch site is CCAFS SLC 40
- Later Flight Numbers are not present for the VAFB SLC 4E Launch Site

Payload vs. Launch Site



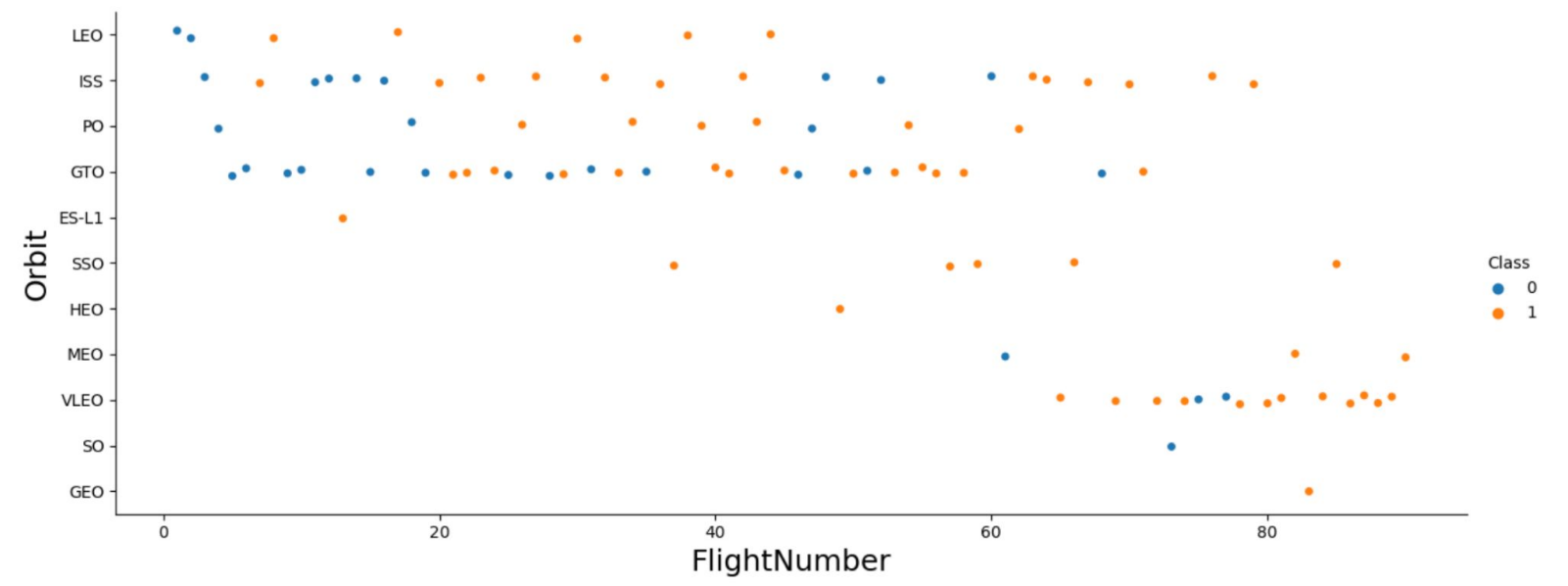
Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

Success Rate vs. Orbit Type



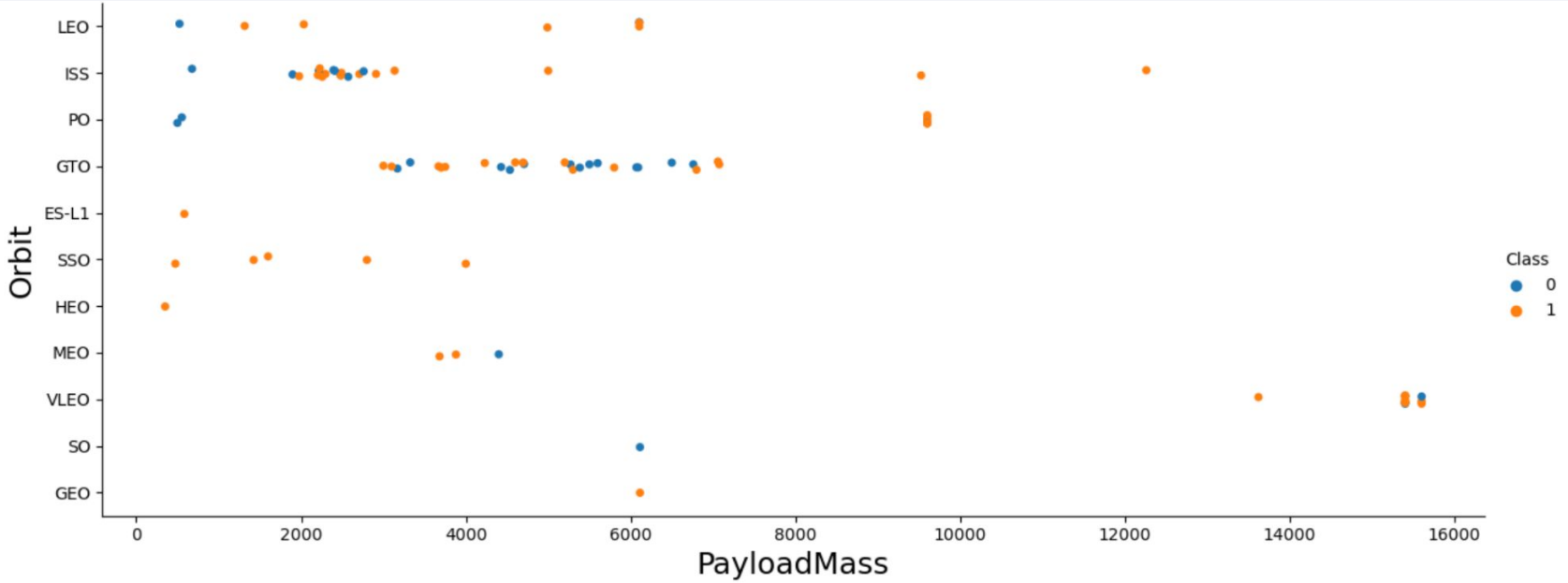
- Orbit Types related to the highest Success rate are:
  ES-L1, GEO, HEO and SSO

Flight Number vs. Orbit Type



- The LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

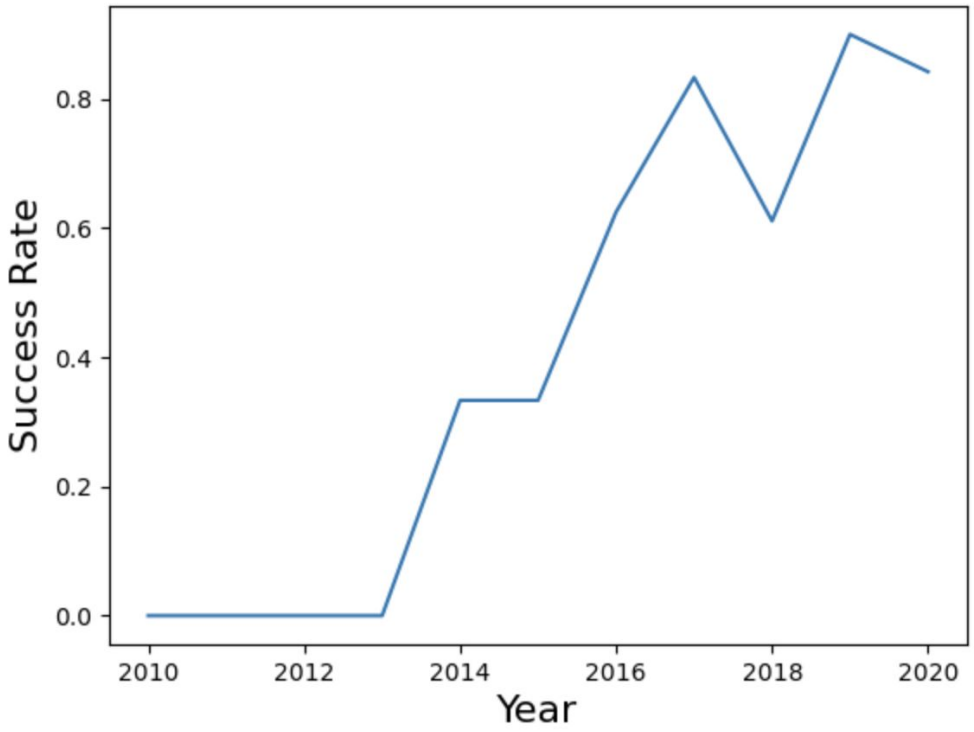
Payload vs. Orbit Type



With heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.

However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Launch Success Yearly Trend



you can observe that the sucess rate since 2013 kept increasing till 2020

All Launch Site Names

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

```
%sql select Unique(LAUNCH_SITE) from SPACEXTBL;
```

Launch Site Names Begin with 'CCA'

launch_site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

```
%sql select LAUNCH_SITE from SPACEXTBL where (LAUNCH_SITE) like 'CCA%' limit 5;
```

Total Payload Mass

```
sql select sum (PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER='NASA (CRS)'
```

```
 * ibm_db_sa://mgb67380:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk
Done.
```

1
45596

Average Payload Mass by F9 v1.1

```
sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where booster_version like 'F9 v1.1%'
```

```
 * ibm_db_sa://mgb67380:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.da
Done.
```

1
2534

First Successful Ground Landing Date

```
sql select min(Date) from SPACEXTBL where LANDING__OUTCOME = 'Success (ground pad)';
```

```
 * ibm_db_sa://mgb67380:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.
Done.
```

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

sql select distinct BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS__KG_ between 4000 and 6000 and LANDING__OUTCOME = 'Success (drone ship)';

Total Number of Successful and Failure Mission Outcomes

```
sql select MISSION_OUTCOME, count(*) from SPACEXTBL group by MISSION_OUTCOME
```

```
 * ibm_db_sa://mgb67380:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u9
Done.
```

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

```
sql select BOOSTER_VERSION,PAYLOAD_MASS__KG_ from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

```
 * ibm_db_sa://mgb67380:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

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

```
sql select BOOSTER_VERSION, LAUNCH_SITE from SPACEXTBL where LANDING__OUTCOME='Failure (drone ship)' and Date like '2015%';
```

```
 * ibm_db_sa://mgb67380:***@98538591-7217-4024-b027-8baa776ffad1.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:30875/bludb
Done.
```

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

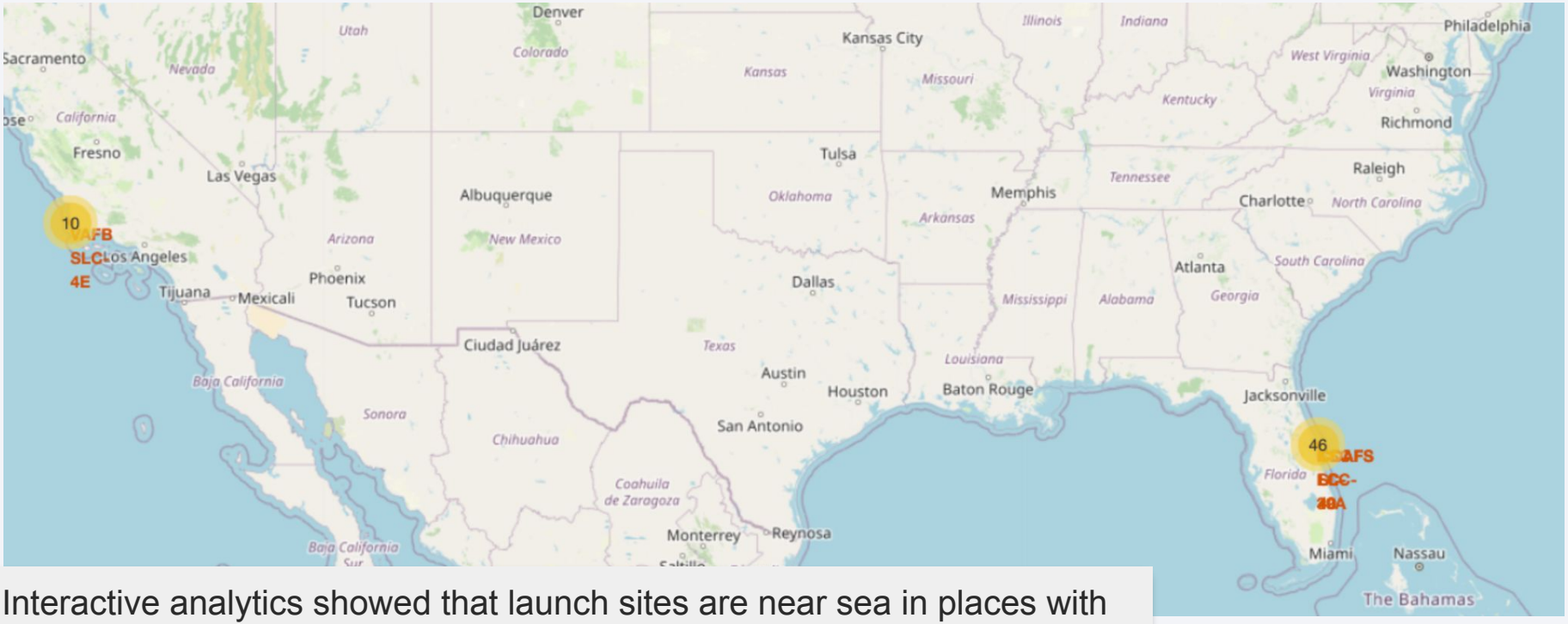
landing__outcome	qty
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

```
sql select LANDING__OUTCOME, count(*) as qty from SPACEXTBL where Date between '2010-06-04' and
'2017-03-20' group by LANDING__OUTCOME order by qty DESC;
```

Section 3

Launch Sites Proximities Analysis

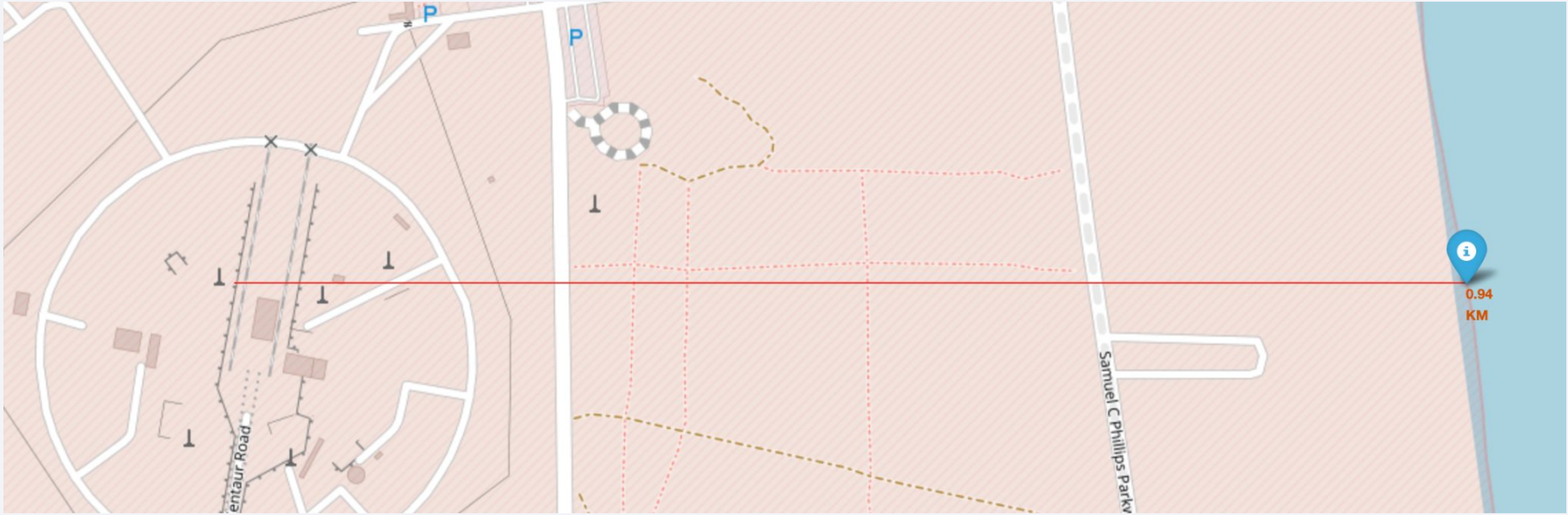
All Launch Sites



Interactive analytics showed that launch sites are near sea in places with good infrastructure. Most launches happen at the east coast.

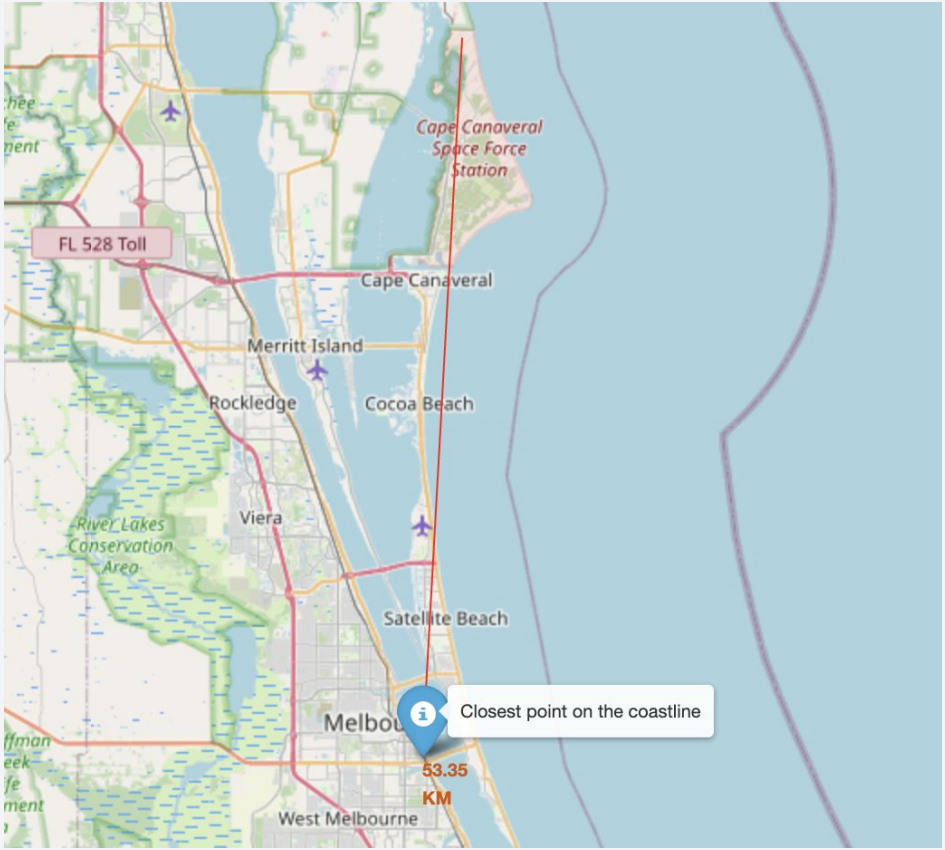
Launch Outcomes

Coast Distance

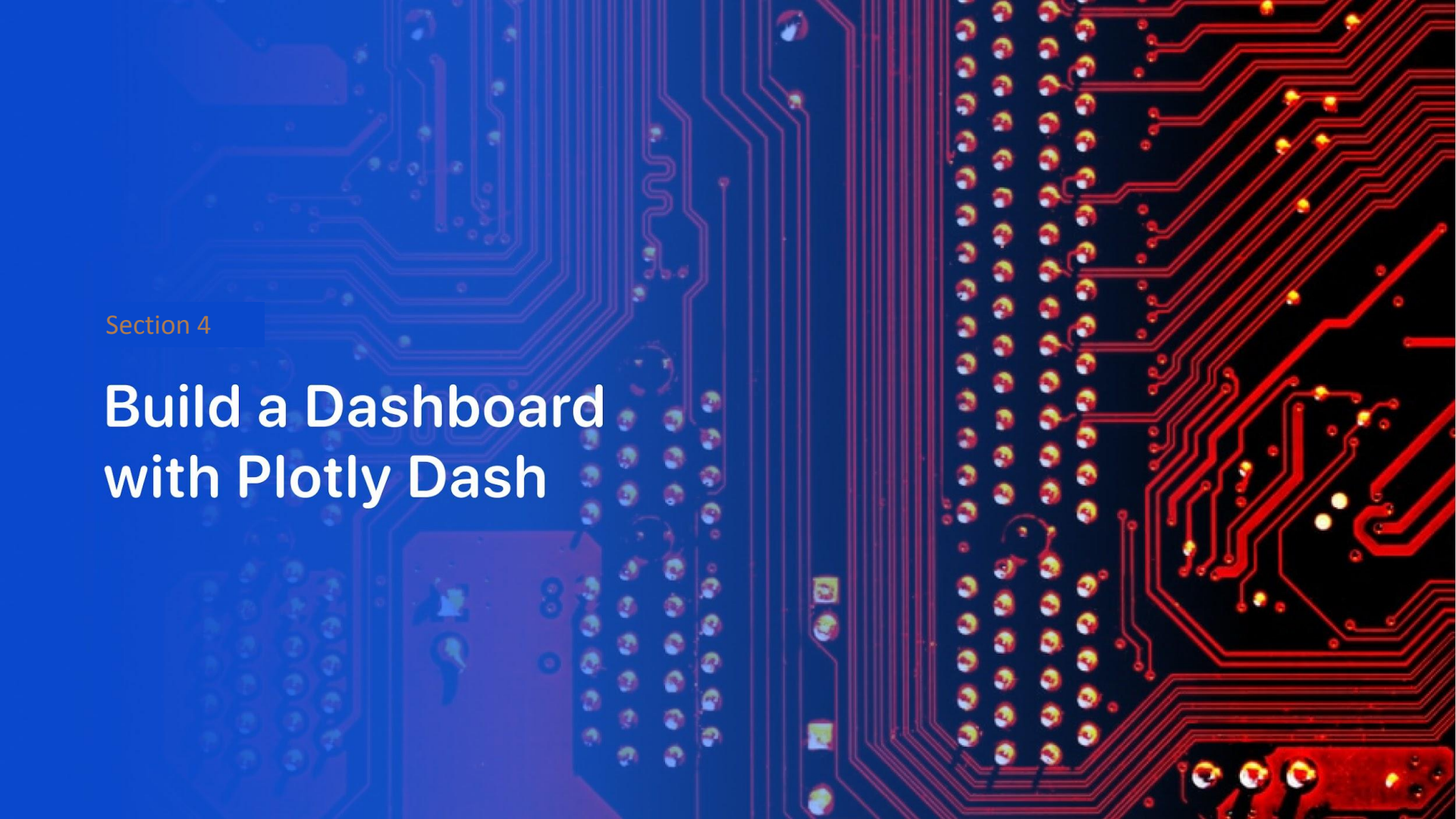


- Coast distance from the launch site is relatively close.

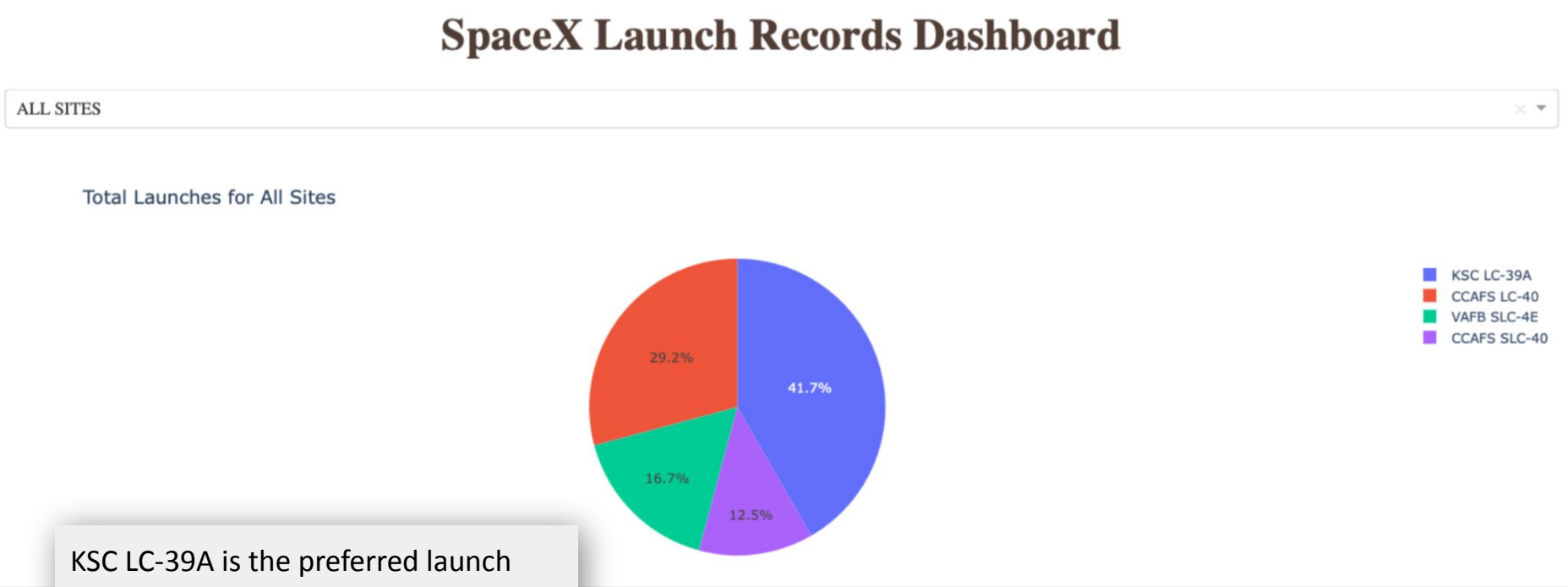
Closest city



- Launch site is relatively far from inhabited places

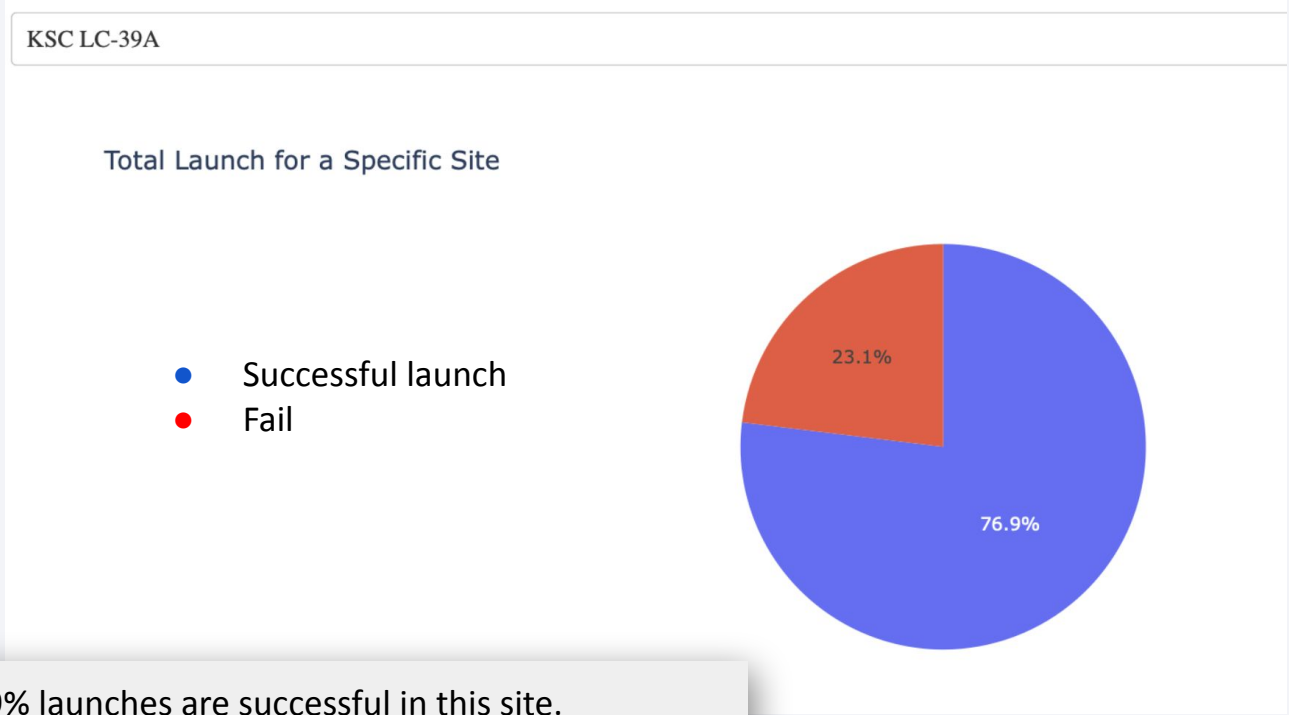

Section 4
Build a Dashboard with Plotly Dash

Total Launches for All Sites



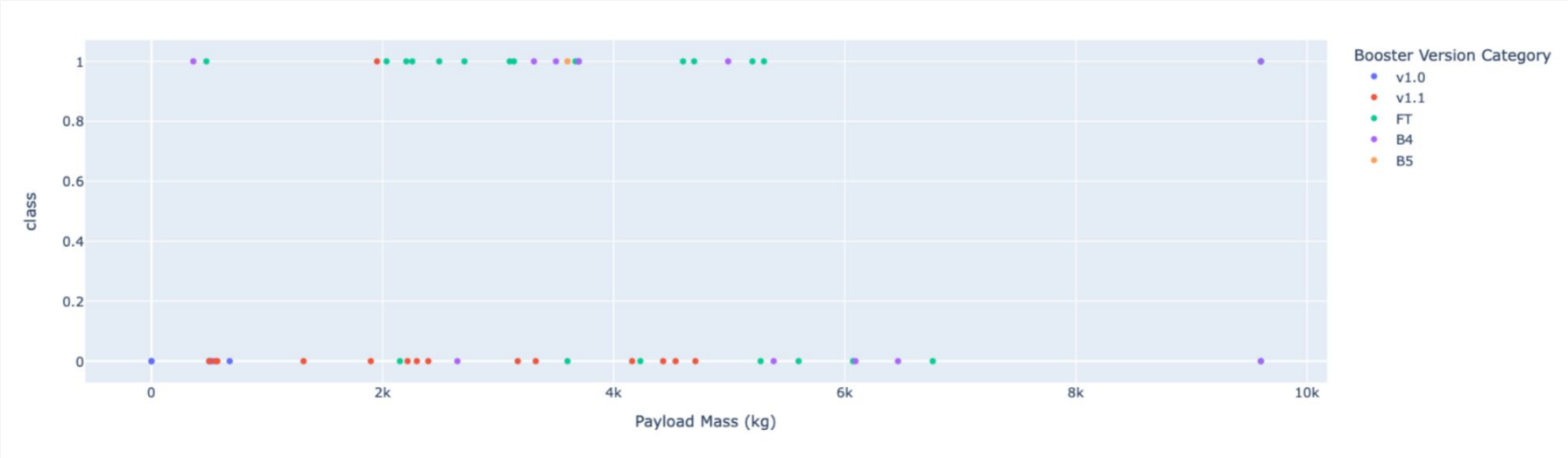
KSC LC-39A is the preferred launch site when it comes SpaceX.

KSC LC-39A launch success



76.9% launches are successful in this site.

Payload vs. Launch Outcome



More successful launches are those of a mass between 2Mg and 6Mg and booster version category FT

Section 5

Predictive Analysis (Classification)

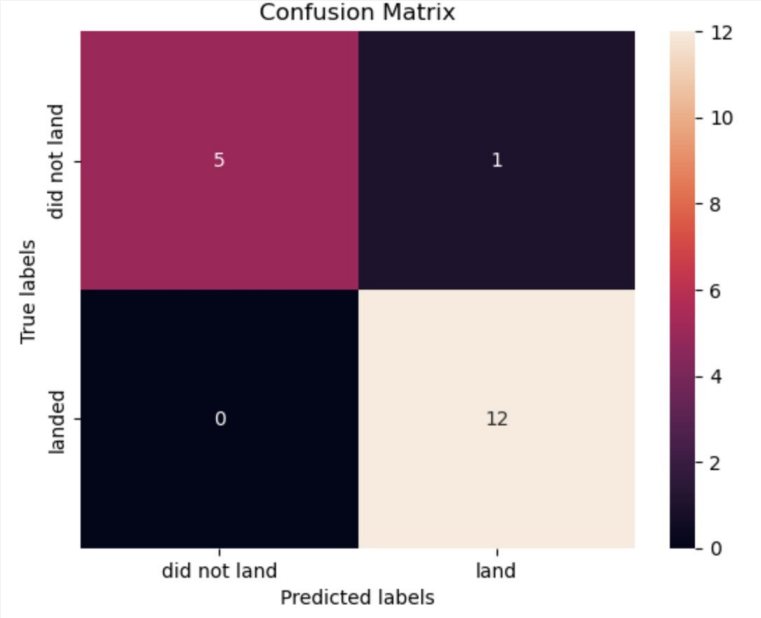
Classification Accuracy

	Accuracy Train	Accuracy Test
Logreg	0.846429	0.833333
Svm	0.848214	0.833333
Tree	0.889286	0.944444
Knn	0.848214	0.833333

- Out of the four tested classification models, the Decision Tree Classifier is the one with the highest accuracy in both data sets.

Confusion Matrix

- Confusion matrix of the best performing model (Tree)



- Examining the confusion matrix, we see that logistic regression can distinguish between the different classes.
- We only have 1 false positive outcome which is better in comparison to other confusion matrices that resulted in 3 false negative results.

Conclusions

- KSC LC-39A is the best launch site.
- Launches between 2 000 kg and 6 000 kg are less risky.
- Orbits ES-L1, GEO, HEO and SSO have a higher success rate.
- Although most of mission outcomes are successful, successful landing outcomes seem to improve over time, according the evolution of processes and rockets.
- Decision Tree Classifier can be used to predict successful landings and increase profits.

Appendix

- https://github.com/koloric/Data_Science_capstone.git
  - Folium didn’t show maps on Github, but for ipynb file there is a pdf version.

Thank you!