IBM Developer
SKILLS NETWORK

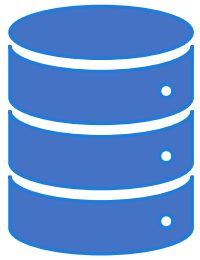
Winning Space Race with Data Science

<Thuy Tran>
<10th August 2022>

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary



Summary of methodologies

Data collection by API and Web – Scraping

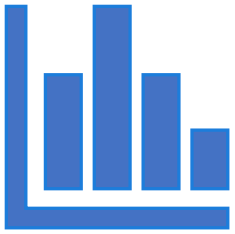
Data wrangling

EDA with Data Visualization and SQL

Building an Interactive Map with Folium

Building a Dashboard with Plotly Dash

Predictive Analysis



Summary of all results

Exploratory data analysis results

Interactive analytics demonstrations

Predictive analysis results

Introduction

- **Project background and context:**
  - SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars. If Space X can reuse the first stage, a cost of165 million dollars each can be saved. Therefore 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.
- **Problems to solve:**
  - Using a machine learning model will be trained to predict if SpaceX will reuse the first stage.

Section 1

Methodology

Methodology

- Executive Summary
- Data collection methodology:
  - Data is collect through SpaceX REST API and Web Scarping Wiki pages.
- Perform data wrangling
  - Replace the NULL values and using one-hot encoding
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Evaluate the performances of Logistic Regression, Support Vector Machine, Decision Tree and KNN models to find the best method.

Data Collection

SpaceX launch data in JSON form is collected using the GET request.

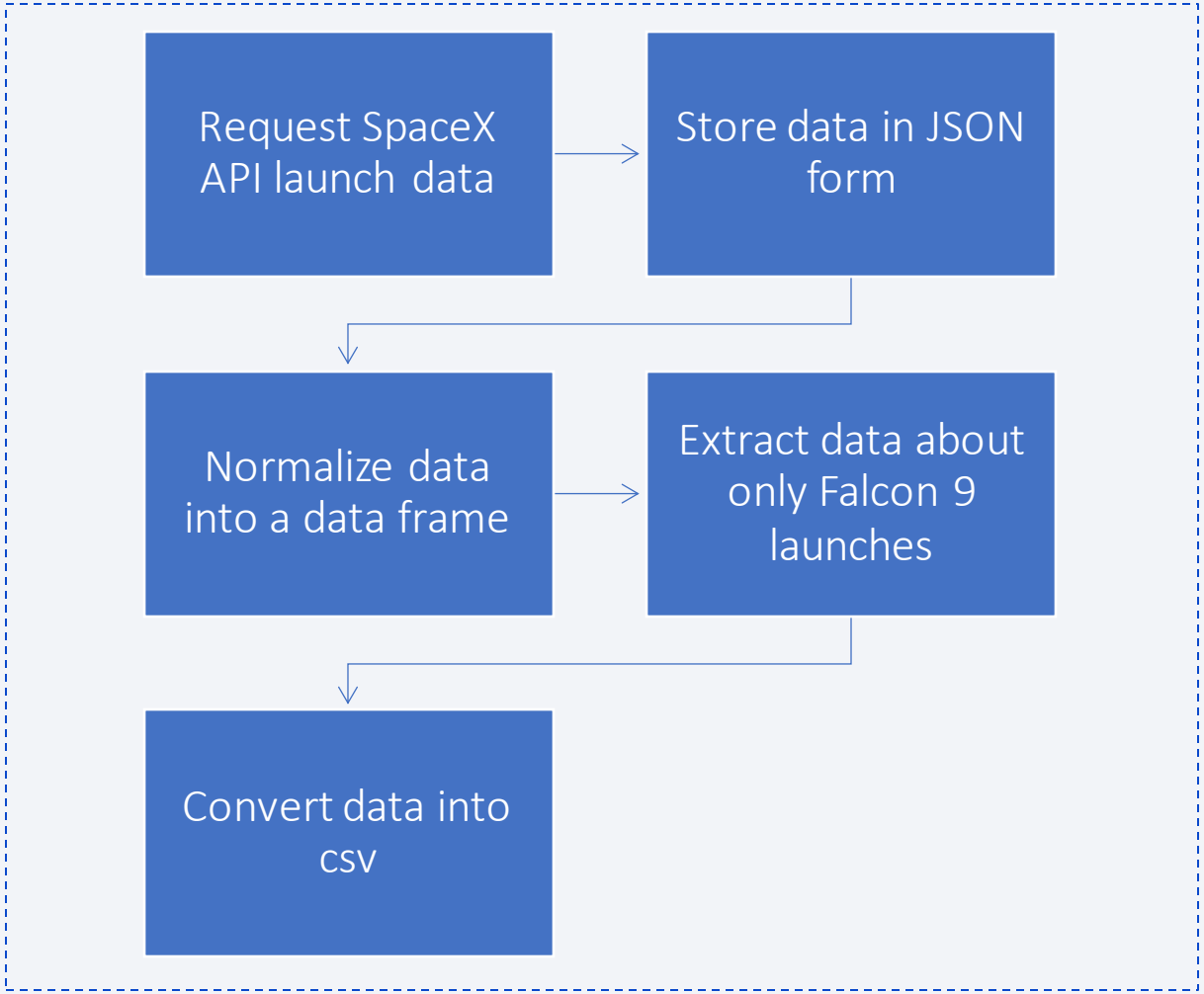
Use json_normalize method to convert the result into a data frame.

Use the API to get information of rocket, payloads, launchpad, and cores, then extract the useful data for each columns.

Filter the data frame to get only the information about the Falcon 9 launches.

Data Collection – SpaceX API

- Data collection with SpaceX REST calls using key phrases and flowcharts:

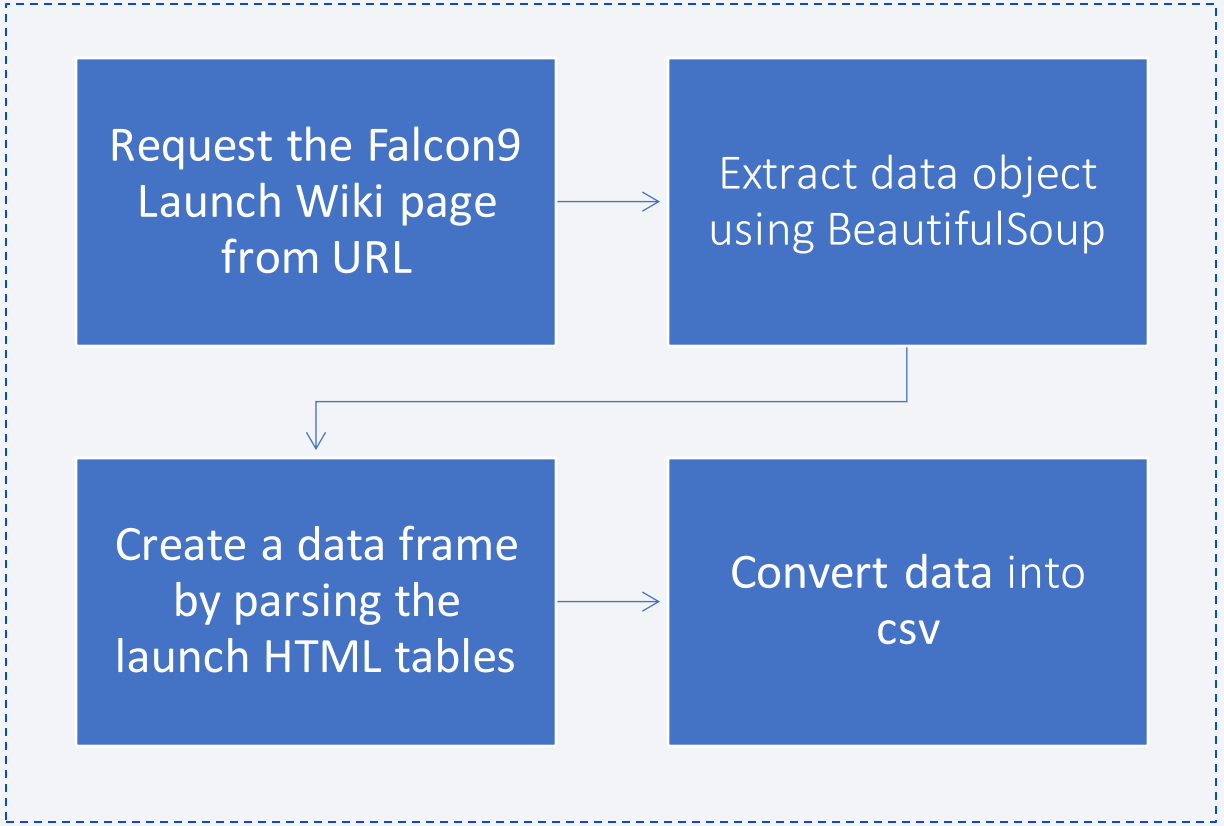


- GitHub URL of the completed SpaceX API calls notebook:

https://github.com/thuytran226/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/328d79856100a460e9761b428c4cf8157659ca9d/Data%20Collection.ipynb

Data Collection - Web Scraping

- Web scraping process using key phrases and flowcharts:

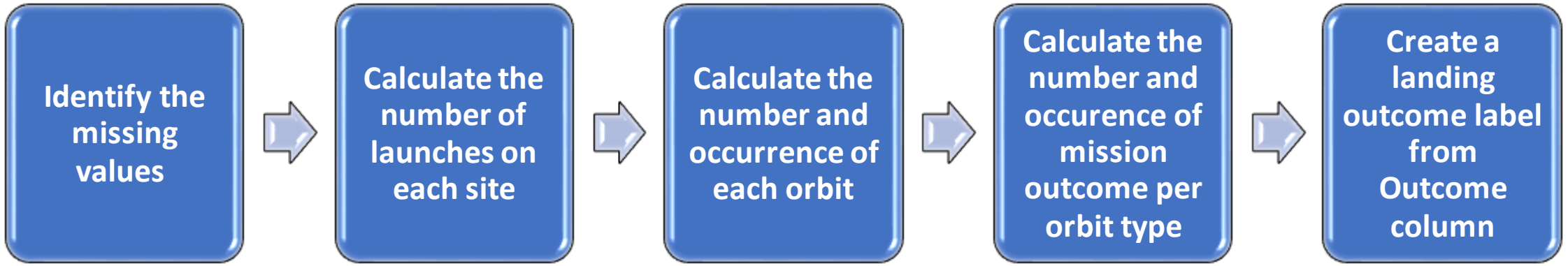


- GitHub URL of the completed web scraping notebook:

https://github.com/thuytran226/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/328d79856100a460e9761b428c4cf8157659ca9d/Data%20Collection%20with%20Web%20Scraping.ipynb

Data Wrangling

- Data Wrangling process using key phrases and flowcharts:



- GitHub URL of data wrangling related notebooks:
https://github.com/thuytran226/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/328d79856100a460e9761b428c4cf8157659ca9d/Data%20wrangling.ipynb

EDA with Data Visualization

- Charts were plotted to check the relationship between:
  - Flight Number and Launch Site
  - Payload and Launch Site
  - Success rate of each orbit type
  - Flight Number and Orbit type
  - Payload and Orbit type

  and examine The launch success yearly trend
- GitHub URL of the completed EDA with data visualization notebook:

https://github.com/thuytran226/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/14e4b84faa9f3ba42b71095cfd31c73b94ade1f6/EDA%20with%20Visualisation%20and%20Feature%20Engineering.ipynb

EDA with SQL

- The SQL performed queries:
  - 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 4000< payload mass < 6000
  - List the total number of successful and failure mission outcomes
  - List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
  - List the failed landing_outcomes in drone ship, their booster versions, and launch site names in 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
- GitHub URL of completed EDA with SQL notebook: https://github.com/thuytran226/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/7cc7e3bc01049f37bc20b2909b35787af9eeb908/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Map objects are used as follows:
  - Use circles to mark all launch sites: A blue circle at NASA Johnson Space Cente and a circle for each launch site
  - Use marker to Mark the success/failed launches for each site: a green marker for successful launch and a red one for failed launch
  - Use MousePosition to get coordinate for a mouse over a point on the map
  - Use line to visualize the distance between the launch sites to railways, highways, coastline or cities

- GitHub URL of completed interactive map with Folium map:

https://github.com/thuytran226/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/46537e3992fc43d40dfed51eb142451eb3578c6e/Interactive%20Visual%20Analytics%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

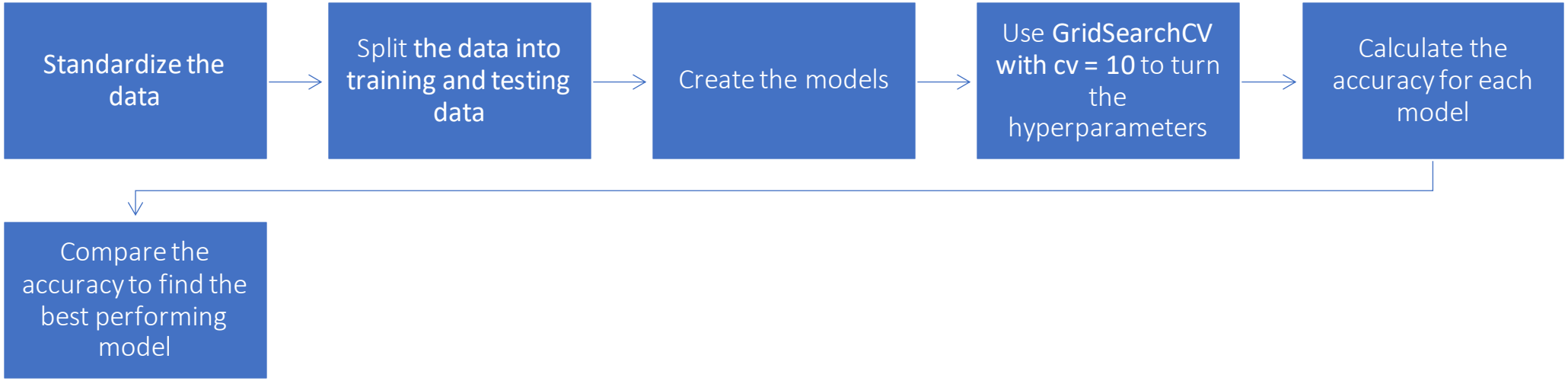
- Plots/graphs and interactions added to a dashboard:
  - Total success launches by all sites
  - Total success launches by a specific sites
  - Interaction between playload and outcome

- GitHub URL of completed Plotly Dash lab:

https://github.com/thuytran226/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/c111a1aa14da0ad16500f2ff33b819d32bf64f9a/spacex_dash_app.py

Predictive Analysis (Classification)

- Flowchart of model development :



- GitHub URL of the completed predictive analysis lab:

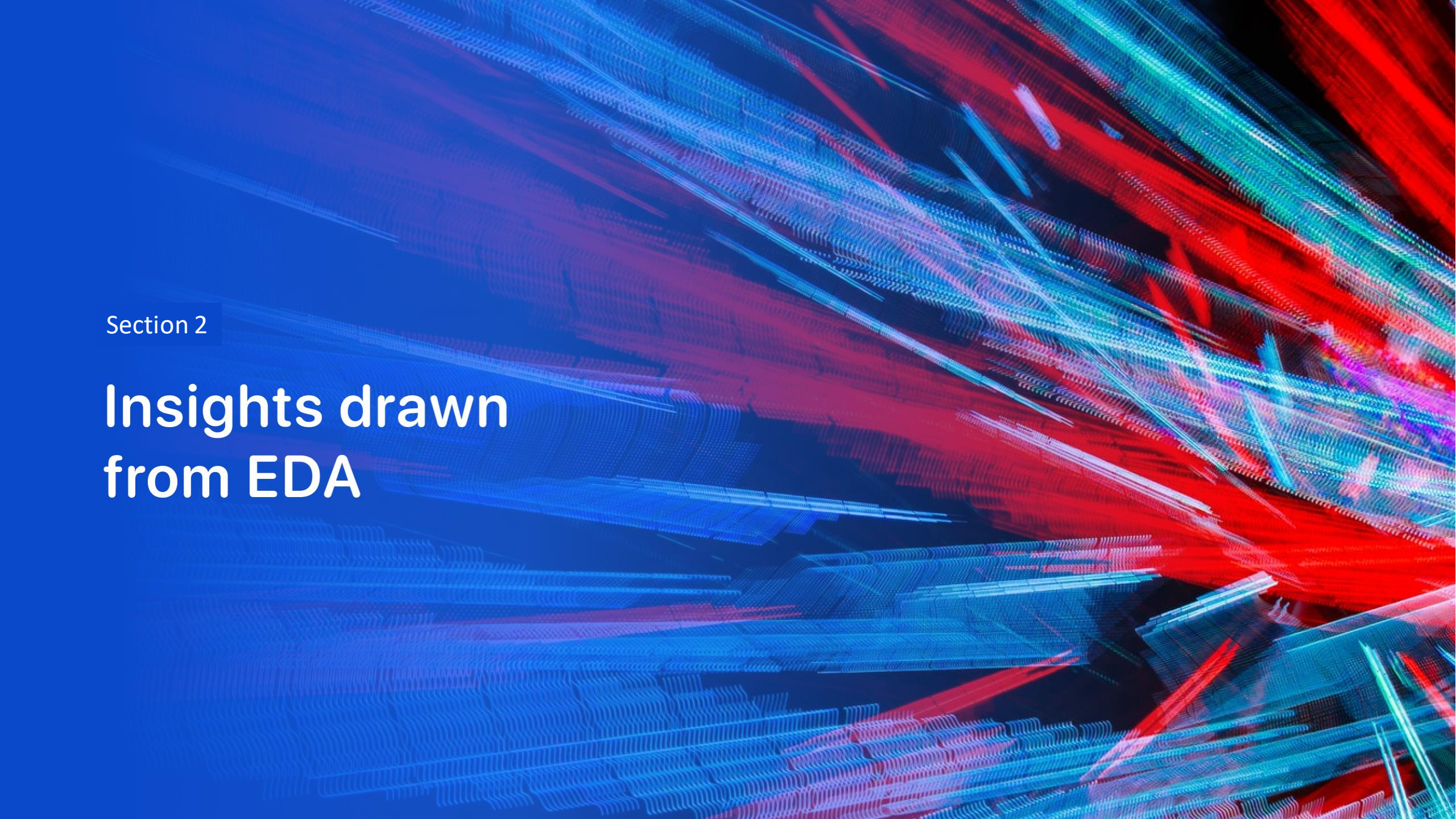
https://github.com/thuytran226/SpaceX-Falcon-9-first-stage-Landing-Prediction/blob/14e4b84faa9f3ba42b71095cfd31c73b94ade1f6/Interactive%20Visual%20Analytics%20with%20Folium.ipynb

Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

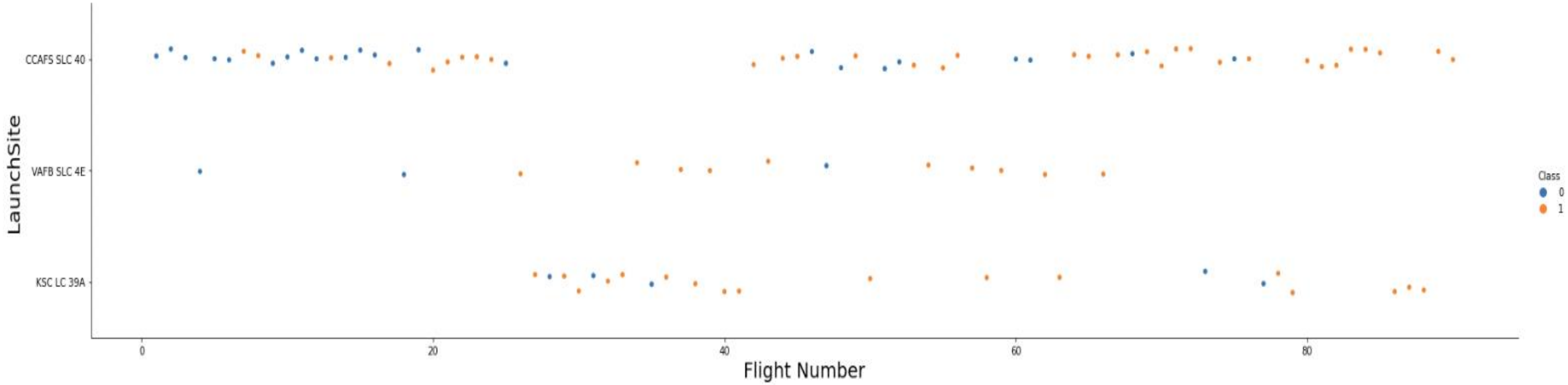
Section 2

Insights drawn from EDA



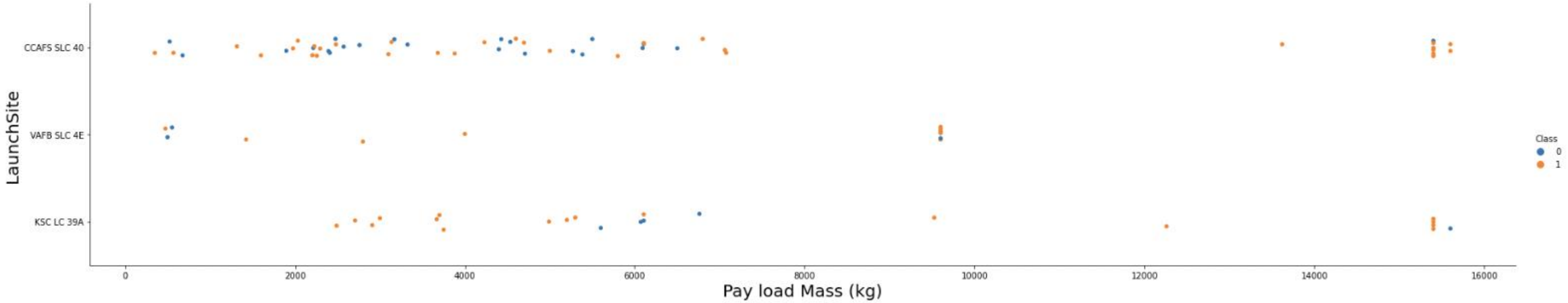
EDA with Visualization

Flight Number vs. Launch Site



- The amount of launches at CCAFS LC-40 are significantly higher than at other launch sites.

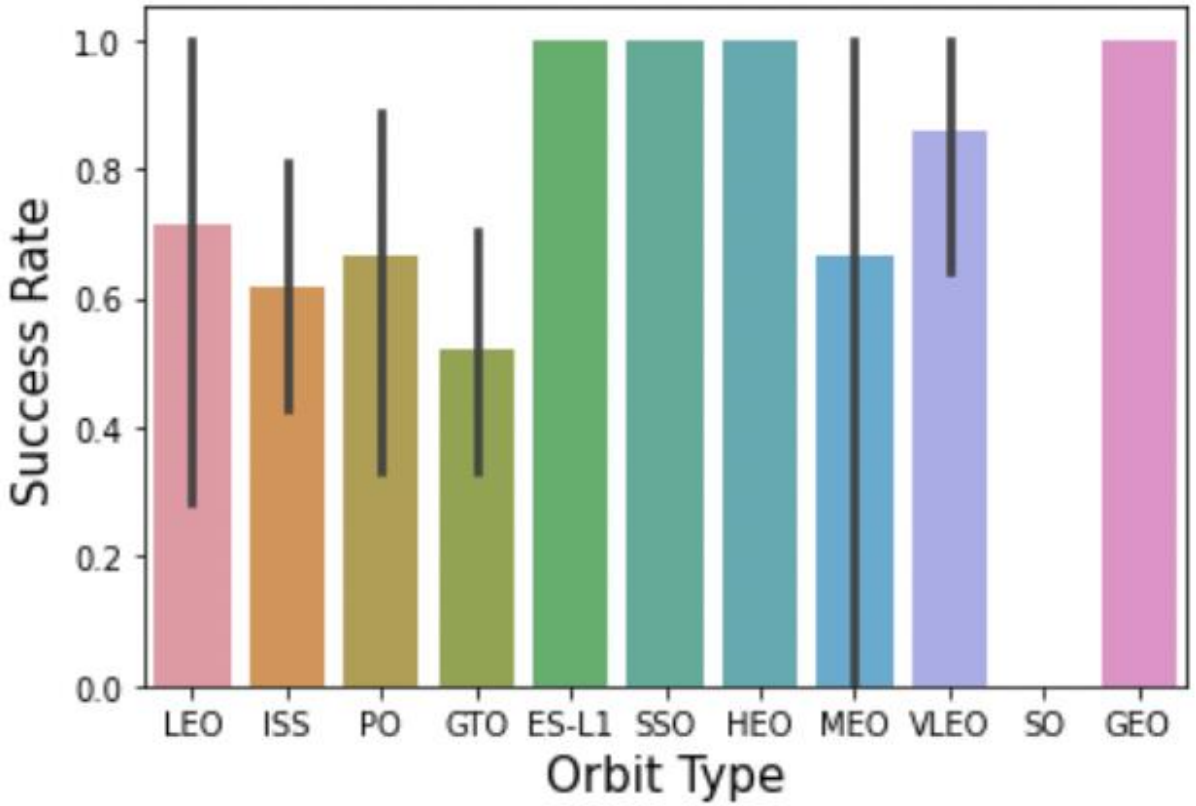
Payload vs. Launch Site



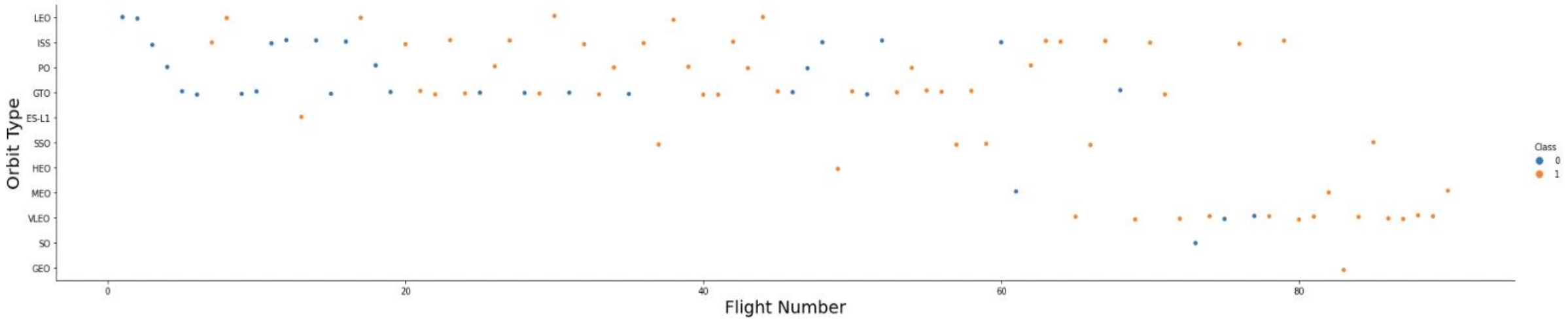
- Most of the rockets with payload mass less than 8000 kg were launched at CCAFS LC-40 launch site, then KSC LC-39A.
- There are no rockets launched for heavy payload mass(greater than 10000 kg) at the VAFB-SLC launch site.

Success Rate vs. Orbit Type

- ES-L1, SSO, HEO and GEO are the orbit types with highest success rate.

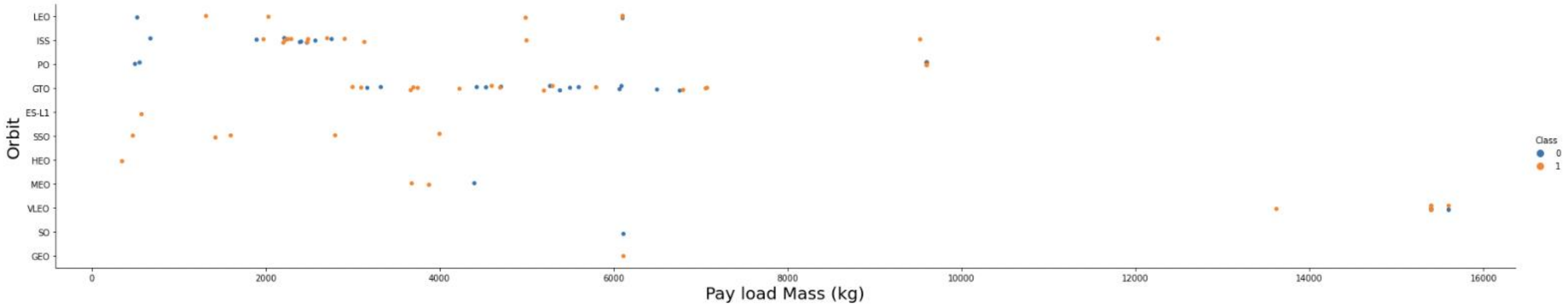


Flight Number vs. Orbit Type



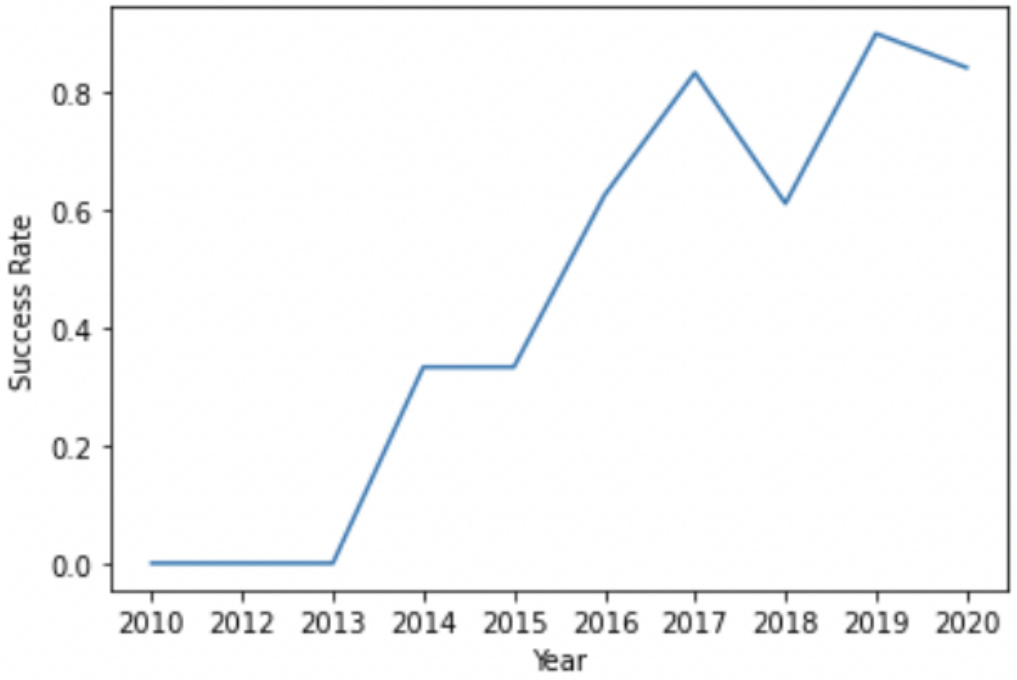
- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

Payload vs. Orbit Type



- Most of the rockets with Pay load Mass < 8000 kg are launched at Orbit Type LEO, ISS and SSO
- The rockets with low Pay load Mass has more success launches than the ones with heavy Pay load Mass.

Launch Success Yearly Trend



- ***Sucess rate is increasing from 2013 to 2020 with a slight fluctuation between 2017 to 2019.***

EDA with SQL

All Launch Site Names

%**sql** SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL

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

- ***Display the names of the unique launch sites in the space mission.***

Launch Site Names Begin with 'CCA'

%**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	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

- ***Display 5 records where launch sites begin with the string 'CCA'***

Total Payload Mass

%**sql** SELECT SUM(payload_mass__kg_) FROM SPACEXTBL

619967 kg

- ***Display the total payload mass carried by boosters launched by NASA (CRS)***

Average Payload Mass by F9 v1.1

%**sql** SELECT SUM(payload_mass__kg_) FROM SPACEXTBL WHERE booster_version LIKE 'F9 v1.1%'

38020 kg

- ***Display average payload mass carried by booster version F9 v1.1***

First Successful Ground Landing Date

%**sql** SELECT MIN(DATE) FROM SPACEXTBL WHERE landing__outcome LIKE 'Success%'

2015-12-22

- ***List the date when the first successful landing outcome in ground pad was acheived.***

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql SELECT booster_version FROM SPACEXTBL WHERE
landing__outcome LIKE 'Success%' AND payload_mass__kg_>4000
AND payload_mass__kg_<6000
```

booster_version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1032.1
F9 B4 B1040.1
F9 FT B1031.2
F9 B4 B1043.1
F9 B5 B1046.2
F9 B5 B1047.2
F9 B5 B1048.3
F9 B5 B1051.2
F9 B5B1060.1
F9 B5 B1058.2
F9 B5B1062.1

- ***List the 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

```
%sql SELECT COUNT(*) FROM SPACEXTBL WHERE mission_outcome = 'Success'OR
mission_outcome = 'Failure'
```

99

- ***List the total number of successful and failure mission outcomes***

Boosters Carried Maximum Payload

%**sql** SELECT booster_version FROM SPACEXTBL WHERE payload_mass__kg_=(SELECT max(payload_mass__kg_)FROM SPACEXTBL)

- ***List the names of the booster_versions 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

2015 Launch Records

%**sql** SELECT landing__outcome,booster_version,launch_site FROM SPACEXTBL WHERE year(DATE) = 2015

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Controlled (ocean)	F9 v1.1 B1013	CCAFS LC-40
No attempt	F9 v1.1 B1014	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
No attempt	F9 v1.1 B1016	CCAFS LC-40
Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40
Success (ground pad)	F9 FT B1019	CCAFS LC-40

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

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

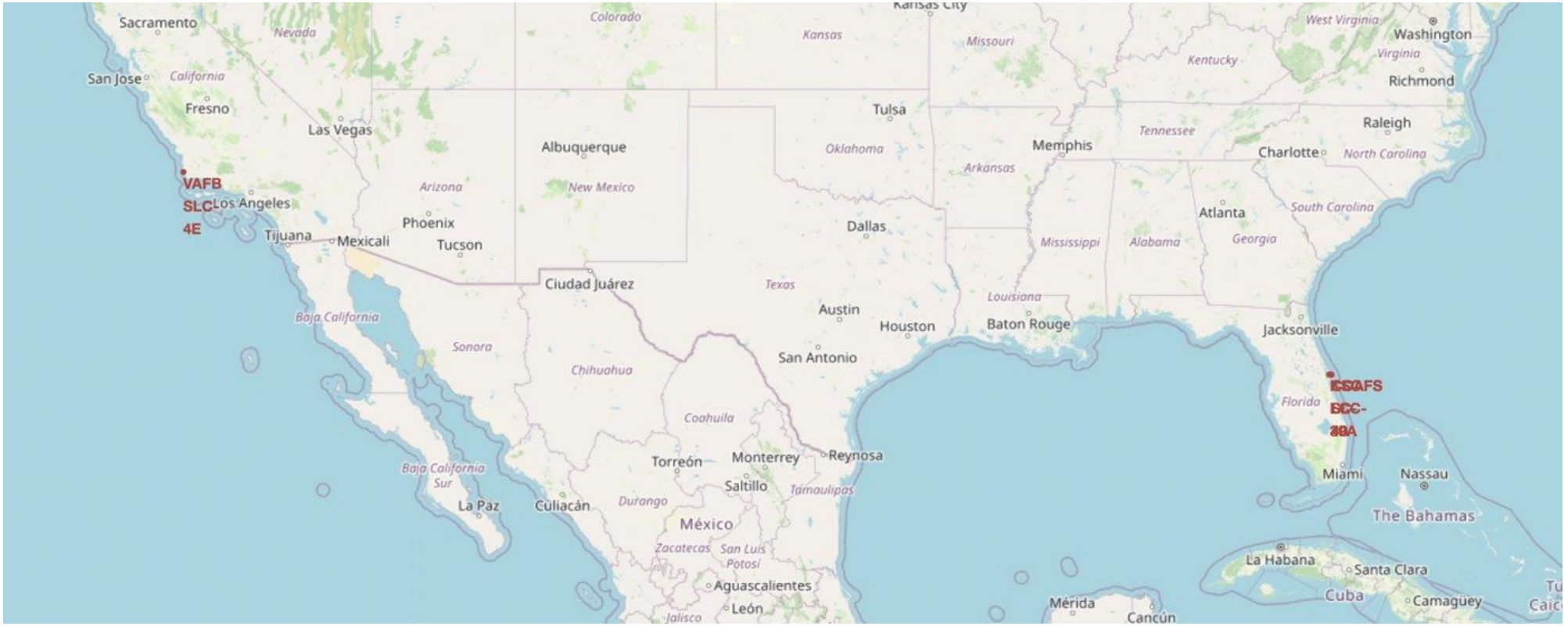
%**sql** SELECT landing__outcome,COUNT(*) FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY landing__outcome ORDER BY landing__outcome DESC

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

- ***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***

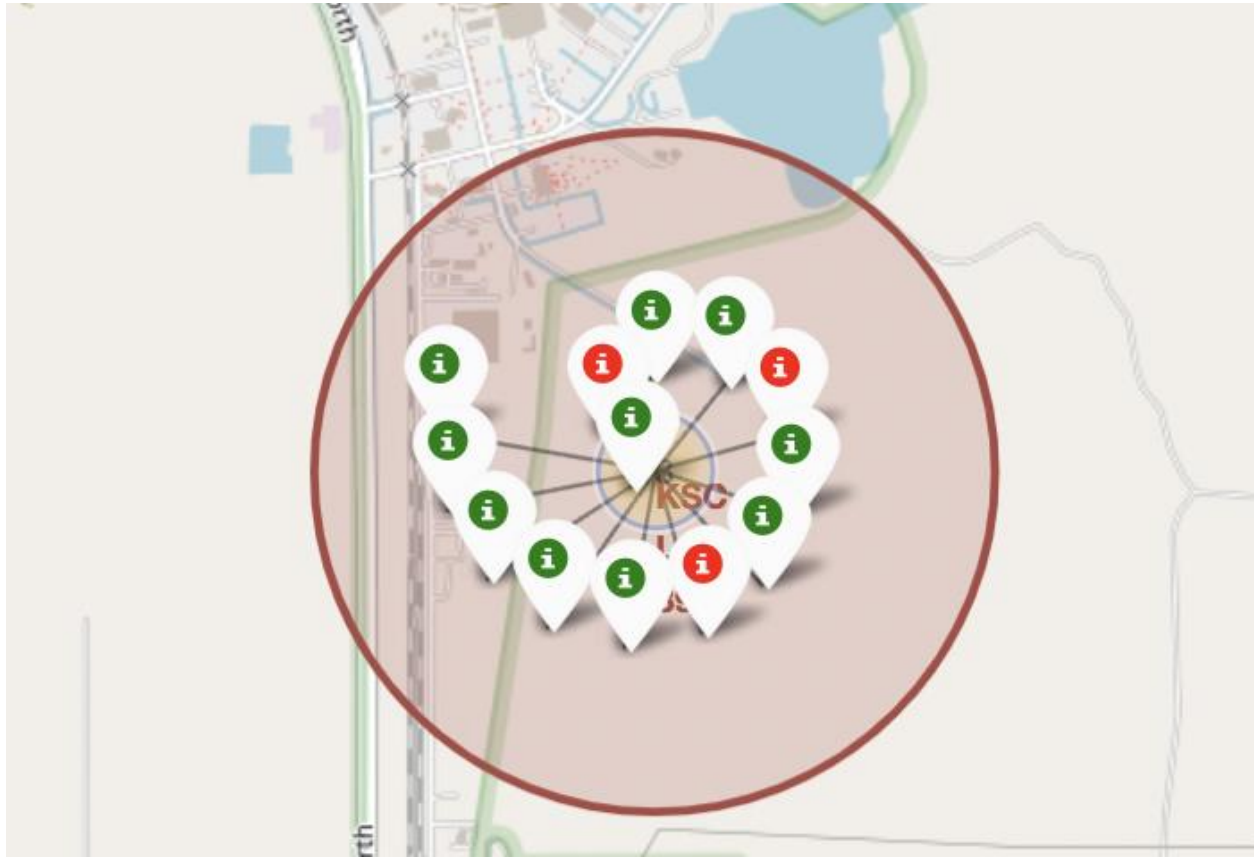

Section 3
Launch Sites
Proximities Analysis

All launch sites marked on a map



- All the launch sites stay in the south coastal areas of US.
- All launch sites in proximity to the Equator line

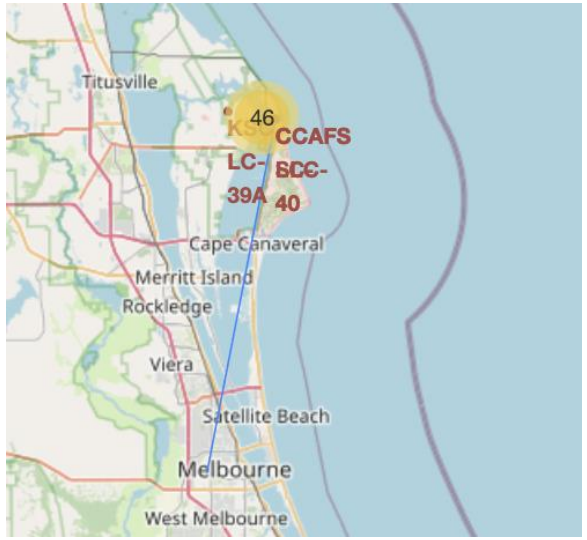
The success/failed launches for each site

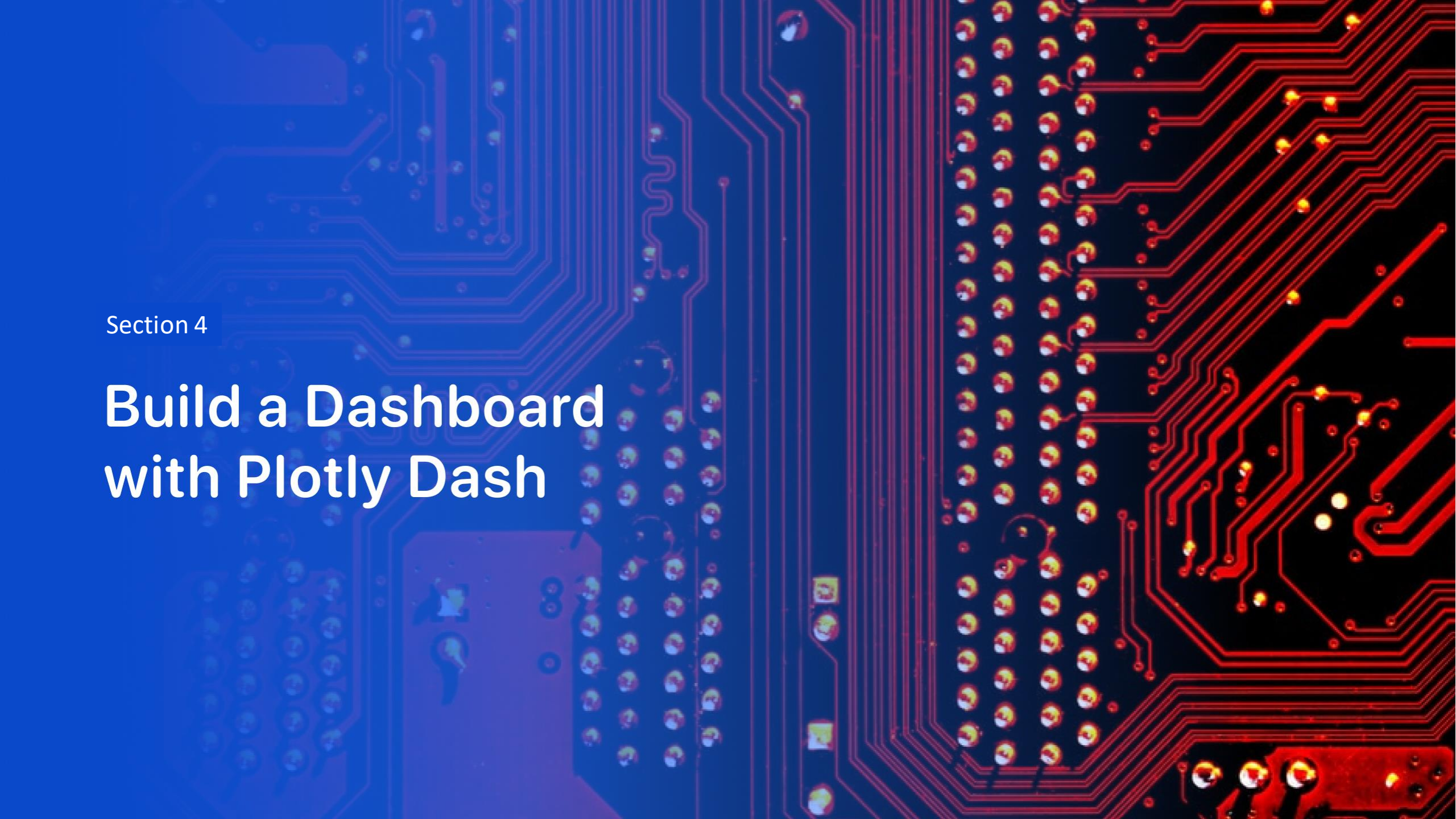


- Green mark: success launches
- Red mark: failed launches

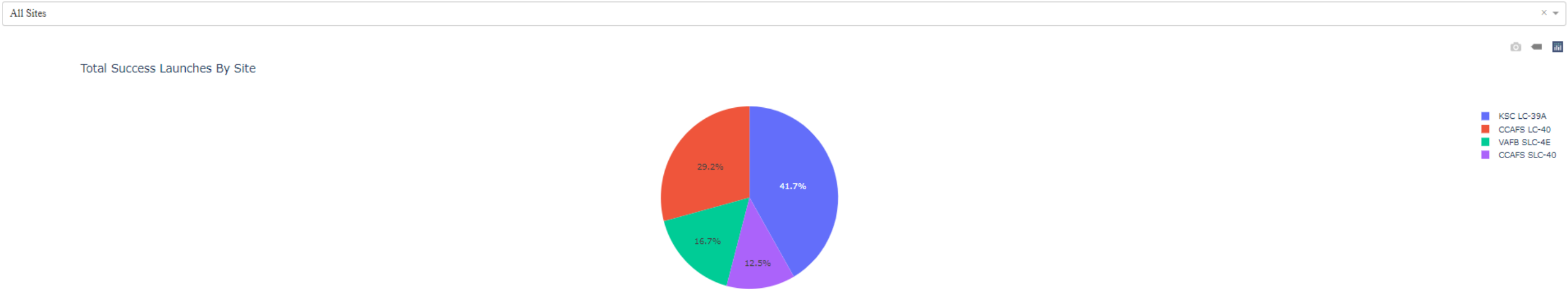
Distances between a launch site to its proximities

- Connect the launch sites to its proximities (coastlines, highways, railways, cities...) by drawing the lines.




Section 4
Build a Dashboard with Plotly Dash

Total Success Launches By Site



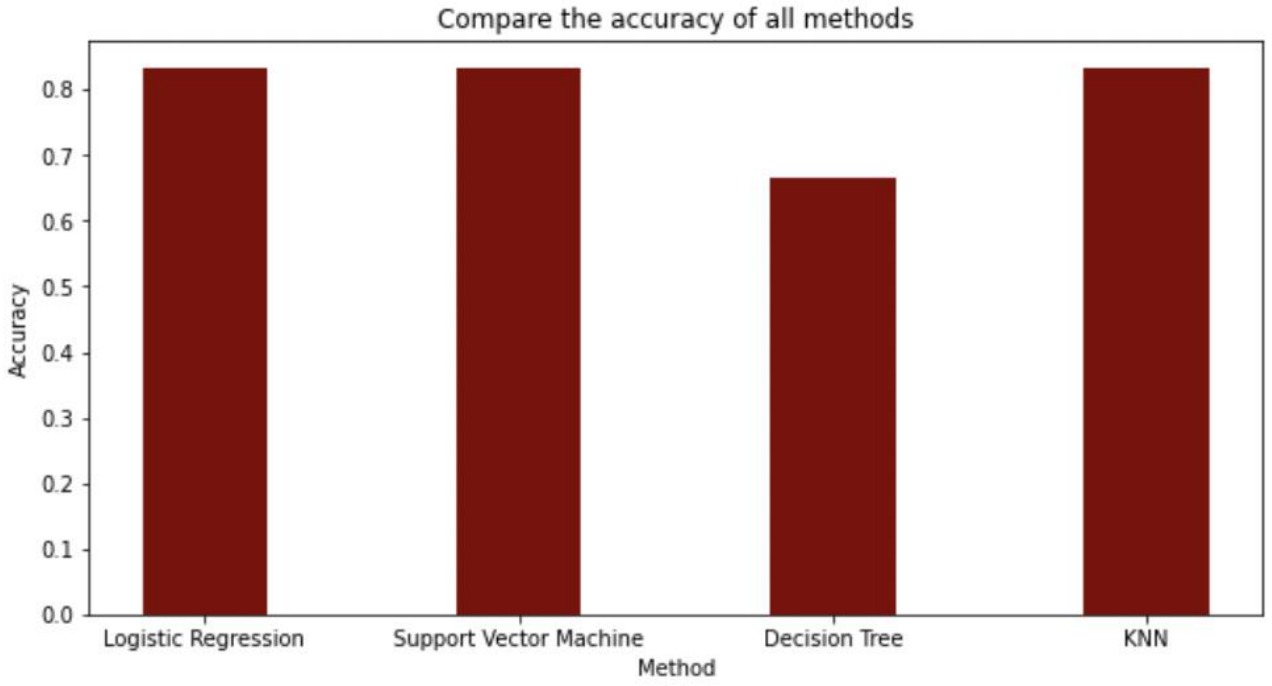
- KSC LC-39 A has the largest successful launches (47%) of all sites.
- CCAFS SLC-40 has the smallest successful launches (12.5%) of all sites.

Note: Since my terminal connection has errors for a while and I couldn't generate the graph again (that worked before), I use the graph from the lab as a reference.

Section 5

Predictive Analysis (Classification)

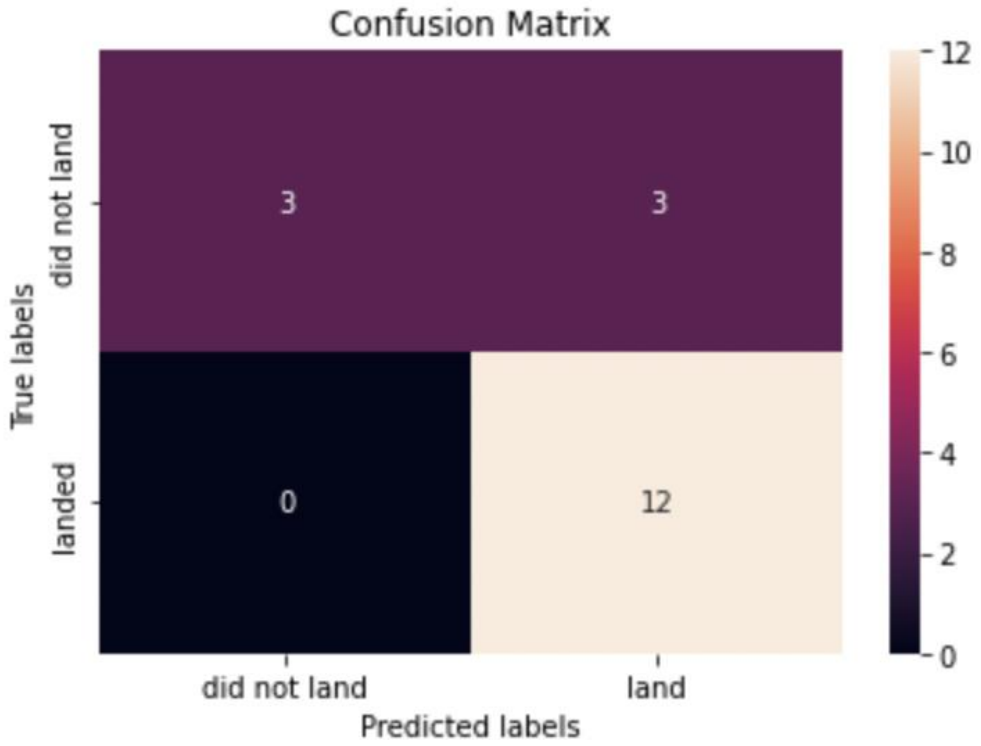
Classification Accuracy



- Support Vector Machine has the highest classification accuracy.

Confusion Matrix

The confusion matrix of the best performing model SVM:



- As we can see from the slides, SVM has the highest rate of correct prediction true positive: 12).

Conclusions

All the launch sites stay in the south coastal areas of US.

ES-L1, SSO, HEO and GEO are the orbit types with highest success rate.

KSC LC-39 A has the largest successful launches of all sites.

CCAFS LC-40 has the highest launch success rate.

Rockets with lower pay load mass performs better than heavy pay load mass

Success rate of launches is increasing from 2013 to 2020

Among 4 models (Logistic Regression, SVM, Decision Tree and KNN), SVM model has the highest classification accuracy.

Thank you!