IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Huong Vo
October 21,,2021

Outline

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

Executive Summary

- **Summary of methodologies**
  - Data Collection
  - Data Wrangling
  - EDA with Data Visualization
  - EDA with SQL
  - Build a Dashboard with Plotly Dash
  - Build an Interactive Map with Folium
  - Predictive Analysis (Classification)

- **Summary of all results**
  - Exploratory data analysis results
  - Interactive analytics demo in screenshots
  - Predictive analysis results

Introduction

- **Project background and context**
    - SpaceY to compete with SpaceX on the race to Space. The cost will largely depend on whether the rocket's first stage can be reused.
    - Use public data and machine learning to predict whether spaceX will be able to reused the first stage. This information is helpful in order to place a bid against SpaceX for a rocket launch.

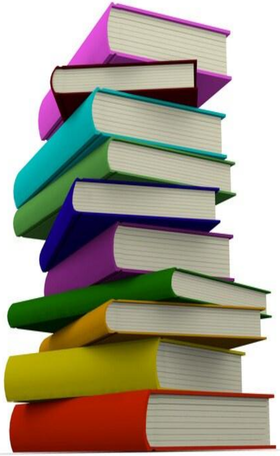
- **Project's purpose:**
    - Predict whether the first stage of Falcon 9 will land successfully.
    - Find out which factors affect the success of the launch
    - Do a location analysis of the launch sites of SpaceX in order to find factors of the perfect launch site.

Section 1

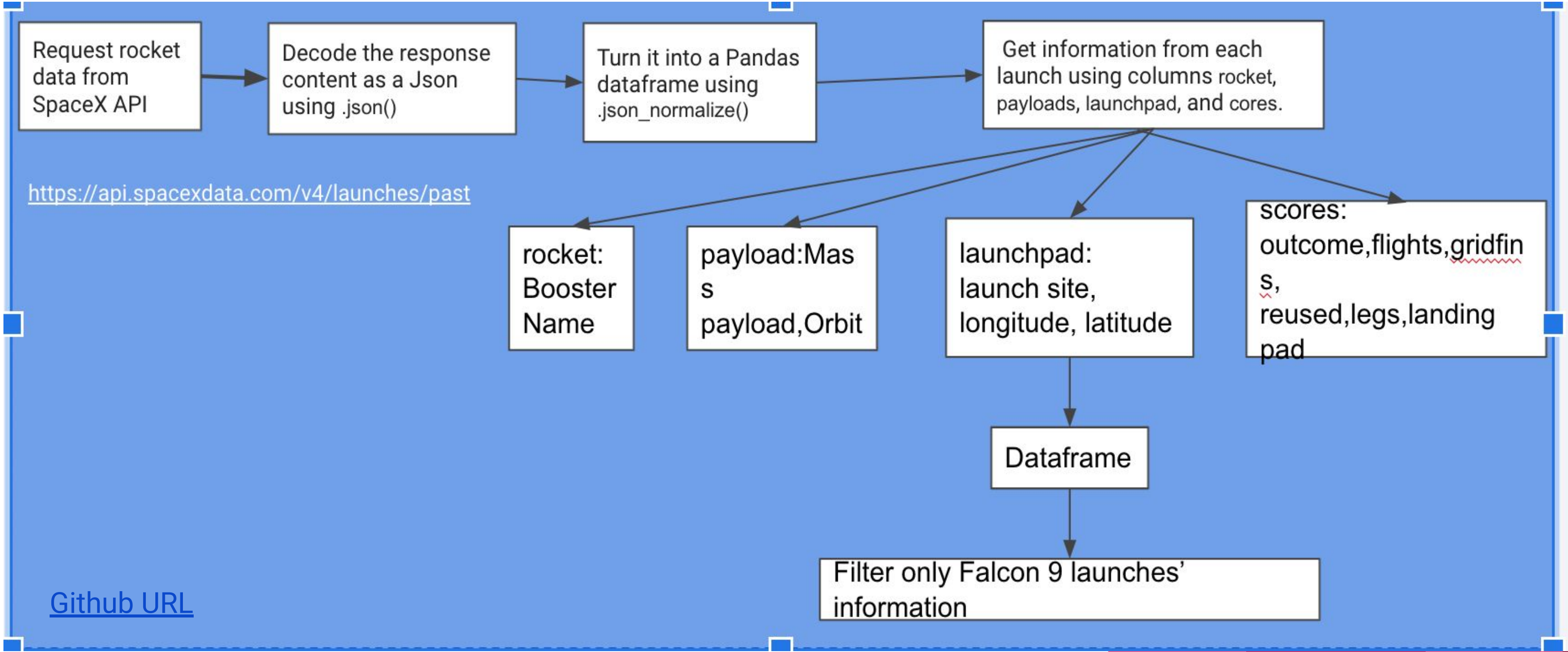
Methodology

Methodology

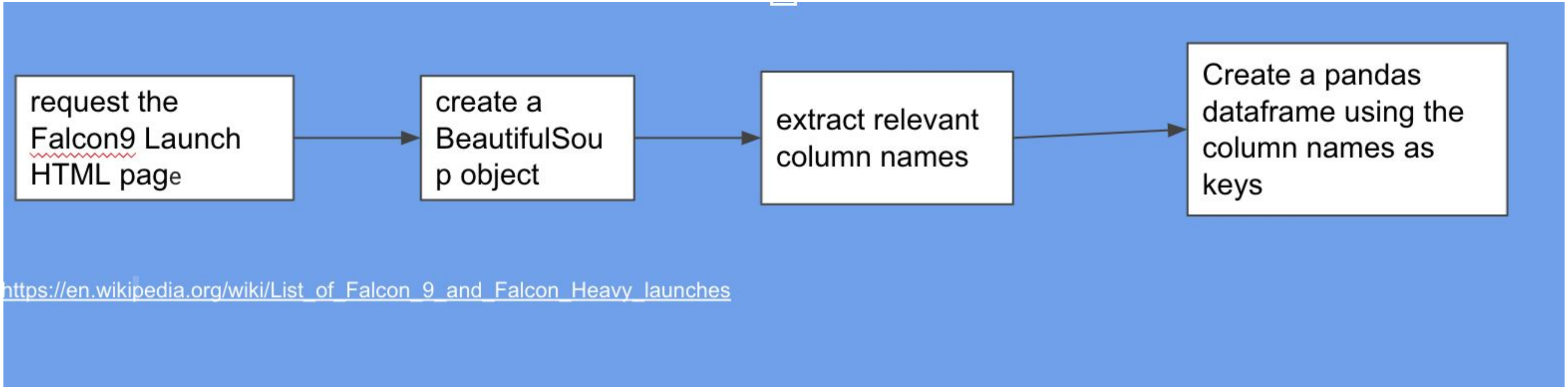
- Data collection methodology:
  - Data was collected using SpaceX REST API and Web scraping from wikipedia
- Data wrangling
  - Data was cleaned, and organize for further analysis
- Exploratory data analysis (EDA) using visualization and SQL
- Interactive visual analytics using Folium and Plotly Dash
- Predictive analysis using classification models
  - Use machine learning models to find the best models for prediction.



Data Collection – SpaceX API

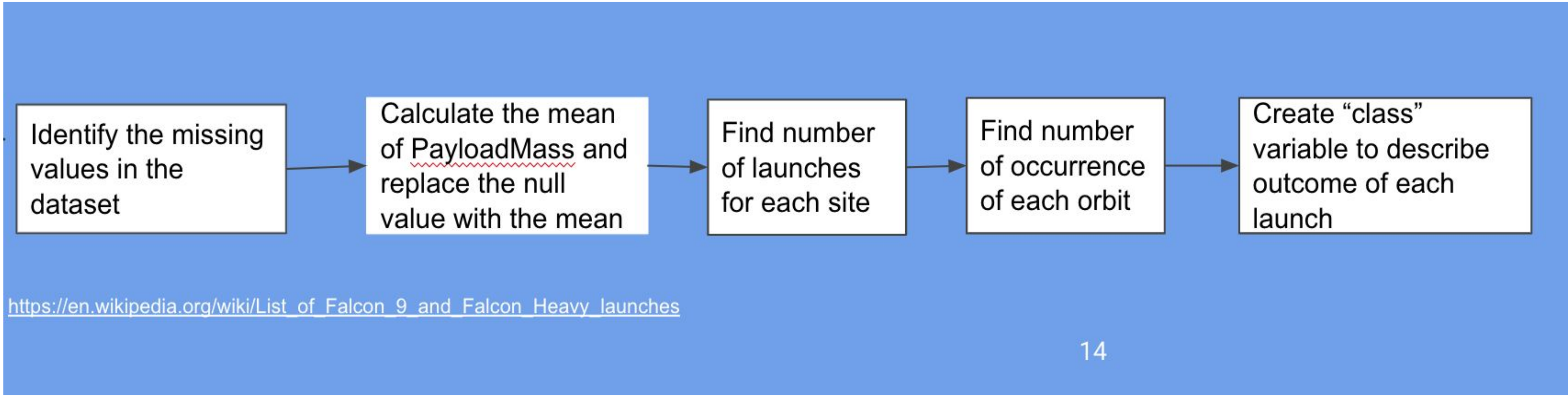


Data Collection - Scraping



[Github URL]

Data Wrangling



[Github URL]

EDA with Data Visualization

- **Scatter plots, bar graphs, line plot were used to generate graphs to demonstrate relationship between different variables:**
  - Relationship between Flight Number and PayloadMass
  - Relationship between Flight Number and Launch Site
  - Relationship between Payload and Launch Site
  - Relationship between success rate of each orbit type
  - Relationship between Flight Number and Orbit type
  - Visualize the relationship between Payload and Orbit type
  - Visualize launch success yearly trend: there is an upward trend of success rate over the years
- Github URL

EDA with SQL

- **Use SQL queries to understand data:**
  - Display the names of the unique launch sites in the space mission
  - Display 5 records where launch sites begin with “CCA”
  - Display the total payload mass carried by boosters launch by NASA (CRS)
  - Display average payload mass carried by booster version F9 v1.1
  - List the date when the first successful landing outcome in the 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 the names of the booster versions which have carried the maximum payload mass
  - List the failed landing outcomes in droneship, their booster versions and launch site in year 2015
  - Rank the count of landing outcomes between 2010-06-04 and 2017-03-20
- GitHub URL

Build a Dashboard with Plotly Dash

- Use a pie chart to display the success launch rate for each launch site and all launch sites. The pie charts shows which site has the highest success launch rate and the success rates for every launch as a whole.

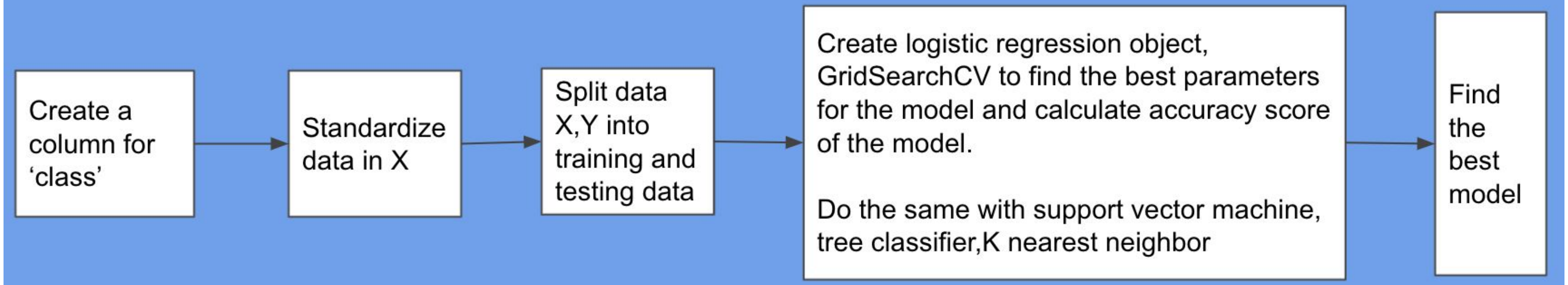
- Use a scattergraph to display the launch outcome with respect to the payload. With the payload lider, we can easily find out if there is any relationship between the payload and the success of the launch.
- [GitHub URL]

Build an Interactive Map with Folium

- We performed location analysis using Folium on the existing launch sites to find out what factors will contribute to the optimal launch site location:
  - Add one circle for each launch sites
  - Mark the success/failed launches for each site on the map with green for successful launches and red for failed launches for clear visualization.
  - Use markers clusters to simplify map with many markers on the same location
  - Draw a polyline between a launch site to select coastlines.
- Github URL

Predictive Analysis (Classification)

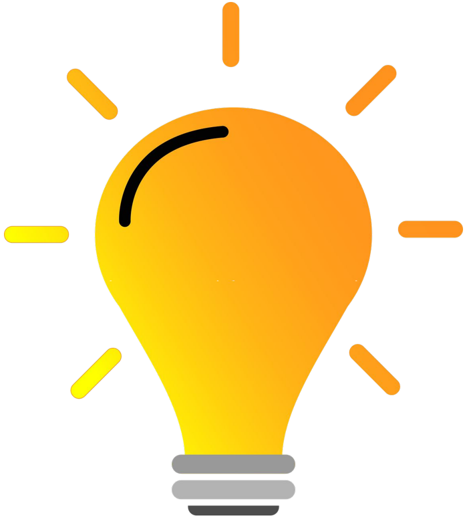
- Model development process flowchart:



[Github URL]

Results

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

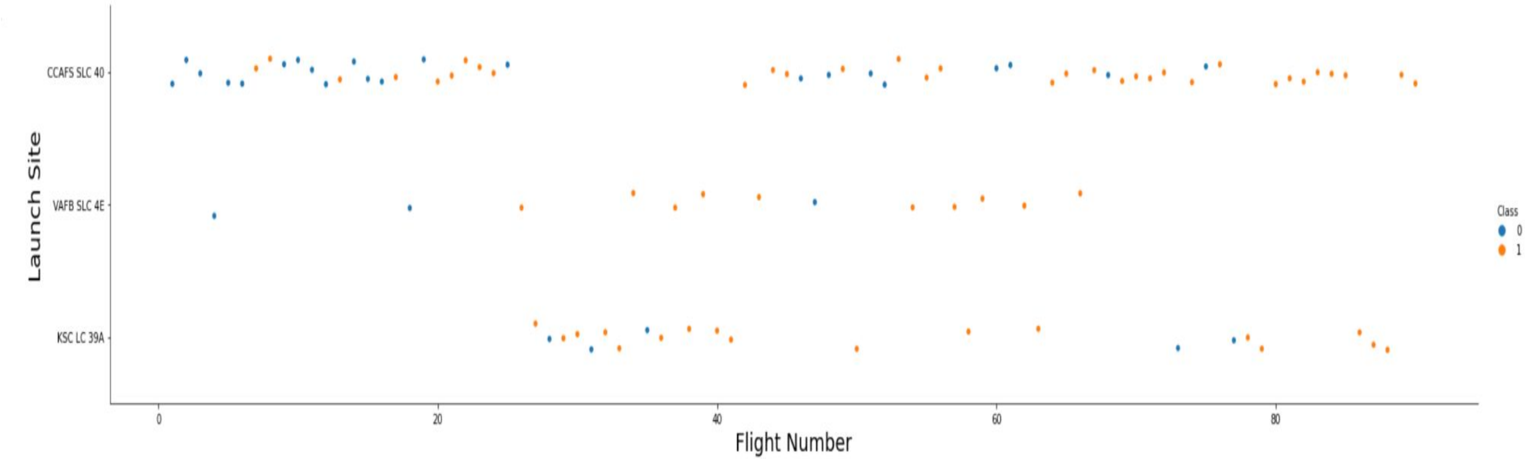


Section 2

Insights drawn from EDA

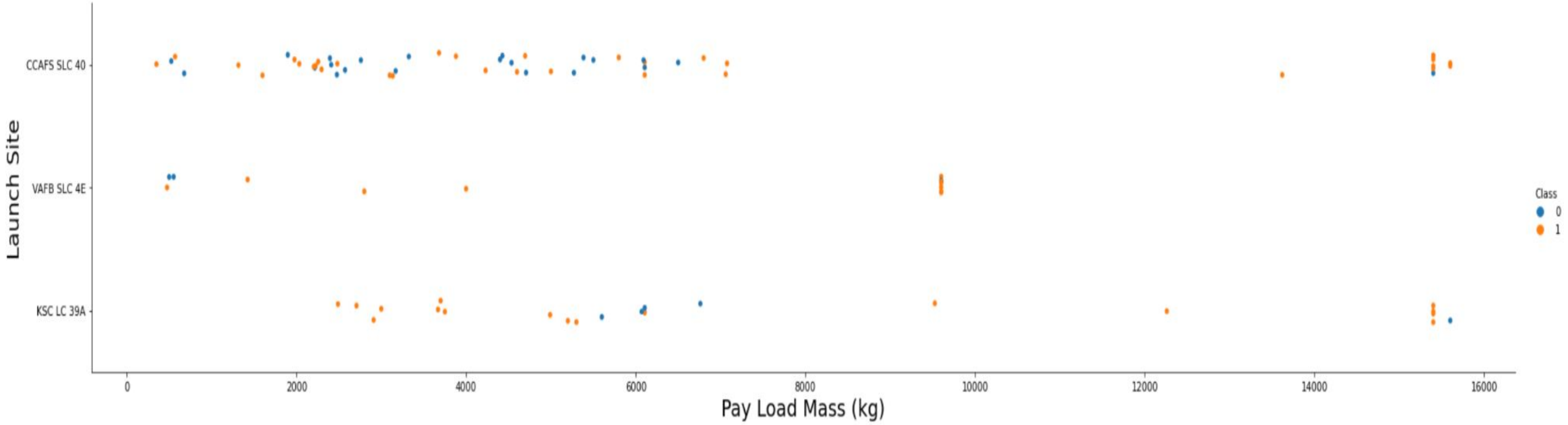
Flight Number vs. Launch Site

- Scatter plot of Flight Number vs. Launch Site: CCAFS SLC 40 has more launches than KSC LC 39A and VAFB SLC 4E.



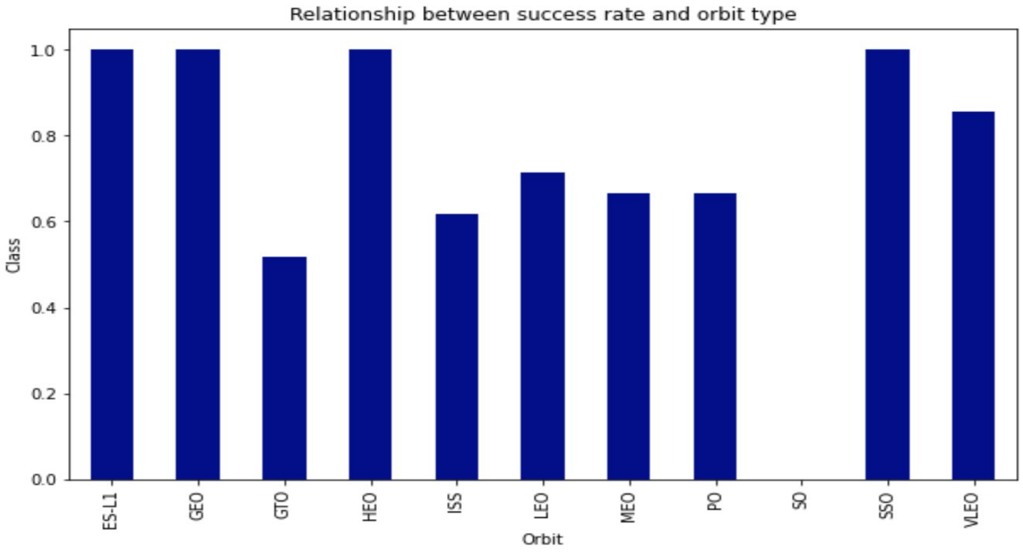
Payload vs. Launch Site

- Scatter plot of Payload vs. Launch Site:
  - Higher payload seems to have higher success rate in VAFB SLC 4E
  - Lower and higher payload seem to have higher success rate in KSC LC 39A
  - Payloads seems to have no relationship to success rate in CCAFS SLC 40.



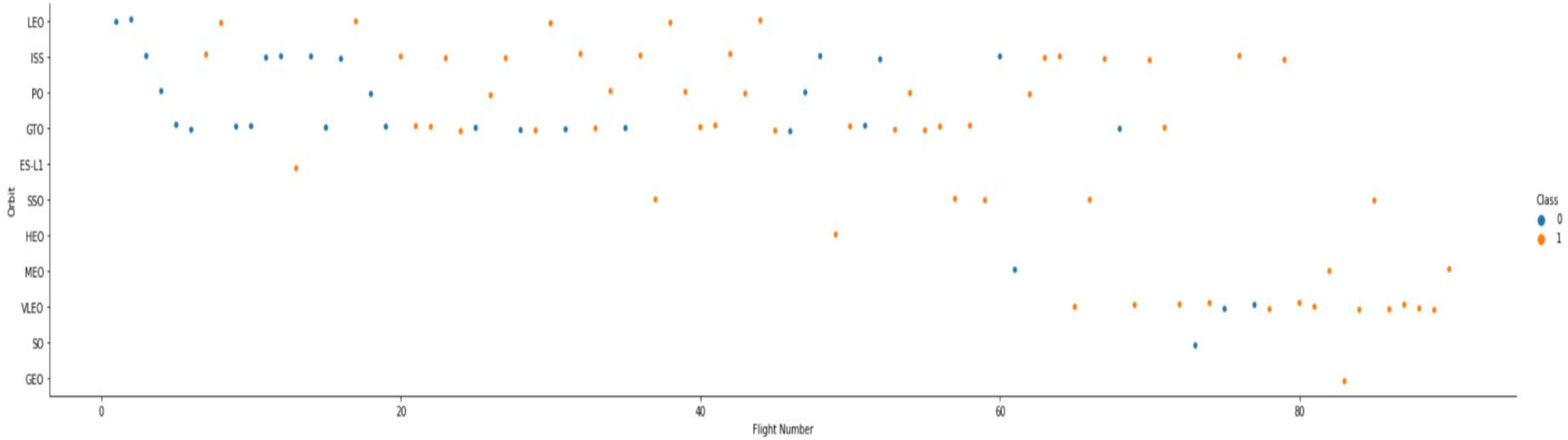
Success Rate vs. Orbit Type

- Bar chart for the success rate of each orbit type: ES-L1,GEO,HEO and SSO have the highest success rate.



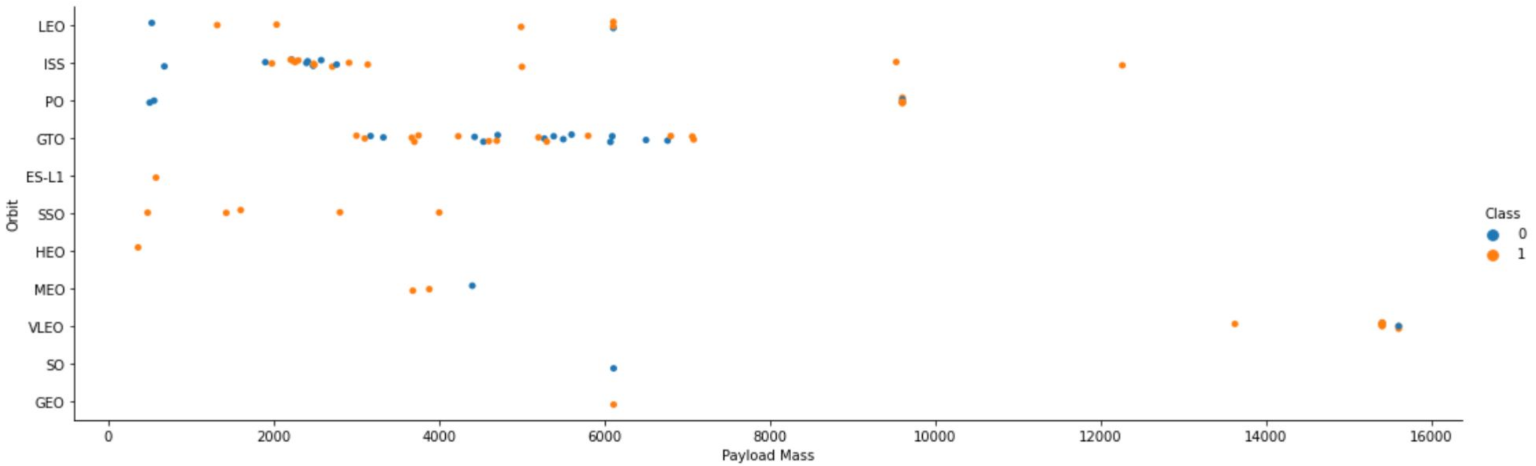
Flight Number vs. Orbit Type

- Scatter point of Flight number vs. Orbit type:
  - Success appears related to the number of flights in LEO orbit
  - There seems to be no relationship between flight number and GTO orbit.



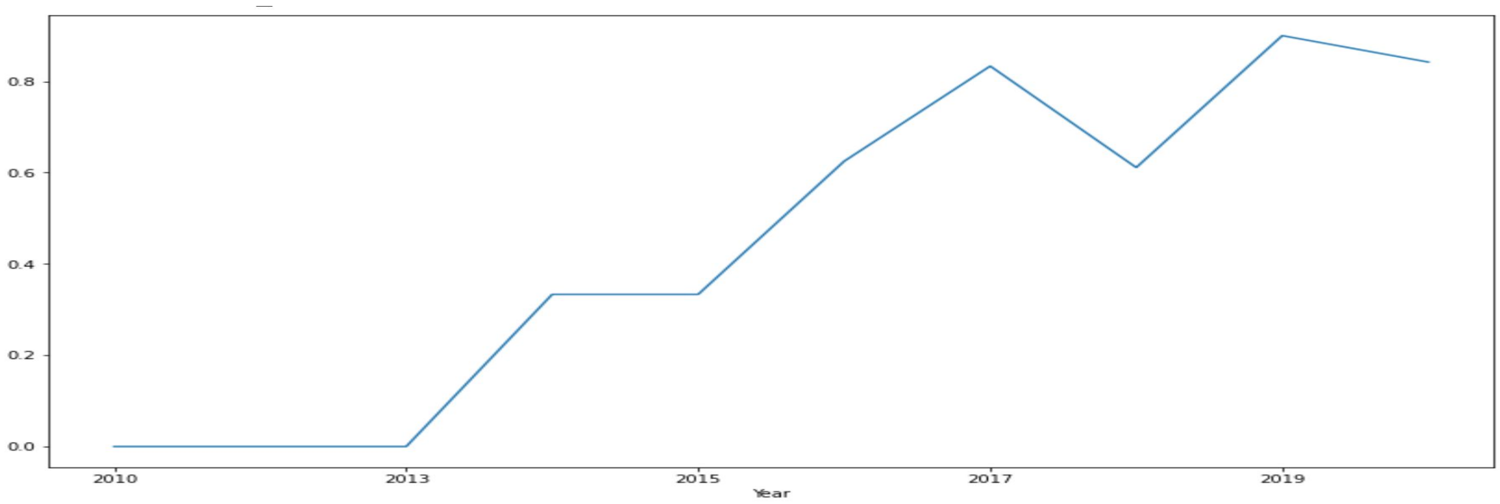
Payload vs. Orbit Type

- Scatter point of payload vs. orbit type:
  - Heavy payloads have a negative influence on GTO orbit and positive influence on LEO and ISS orbits.



Launch Success Yearly Trend

- Line chart of yearly average success rate: there is an upward trend of success rate over the years.



All Launch Site Names

- We use this query to find the unique launch sites:

```
1 %sql select distinct(launch_site) from spacextbl
```

- Result:

  There are 4 unique launch sites:
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
  - VAFB SLC-4A

Launch Site Names Begin with 'CCA'

- Query to find 5 launch site names begin with ‘CCA’:

```
%%sql select *
from spacextbl
where launch_site like 'CCA%'
limit 5
```

- Results:

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

- Query to find total payload carried by boosters from NASA

```
%%sql select sum(payload_mass__kg_) as total_payload_mass
from spacextbl
where customer = 'NASA (CRS)'
```

- Result:

Total payload mass : 45,596 kg

Average Payload Mass by F9 v1.1

- Query to find average payload mass carried by booster version F9 v1.1

```
%%sql select avg(payload_mass__kg_)
from spacextbl
where booster_version like 'F9 v1.1%';
```

- Result:

  The average payload mass carried by F9 v1.1 is 2534 kg

First Successful Ground Landing Date

- Query to find the dates of the first successful landing outcome on ground pad

```
%%sql select min(date)
from spacextbl
where landing__outcome='Success (ground pad)'
```

- Result:

  Date : 2012-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
%%sql select booster_version
from spacextbl
where landing__outcome = 'Success (drone ship)'
and (payload_mass__kg_>4000 and payload_mass__kg_<6000)
```

- Result:

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

Total Number of Successful and Failure Mission Outcomes

- Query to find total number of successful and failure mission outcomes

```
%%sql select mission_outcome,count(mission_outcome)as count
from spacextbl
group by mission_outcome
```

- Result:

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

Boosters Carried Maximum Payload

- Query to find names of the booster which have carried the maximum payload mass

```
%%sql select booster_version
from spacextbl
where payload_mass__kg_ = (select max(payload_mass__kg_) from spacextbl)
```

- Result:

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

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

```
%%sql select landing__outcome,booster_version,launch_site
from spacextbl
where date like '2015%' and landing__outcome like '%drone ship%'
```

- Result:

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40

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

```
%%sql select landing__outcome,count(landing__outcome)
from spacextbl
where date between '2010-06-04' and '2017-03-20'
group by landing__outcome
order by count(landing__outcome) desc
```

- Result:

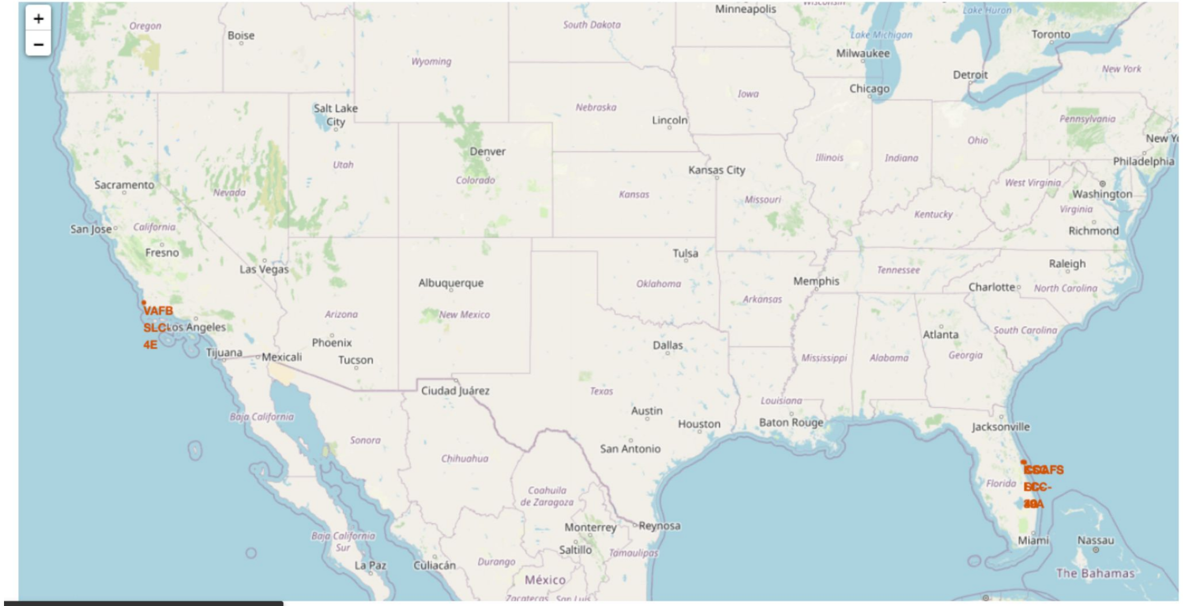
landing__outcome	2
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 4

Launch Sites Proximities Analysis

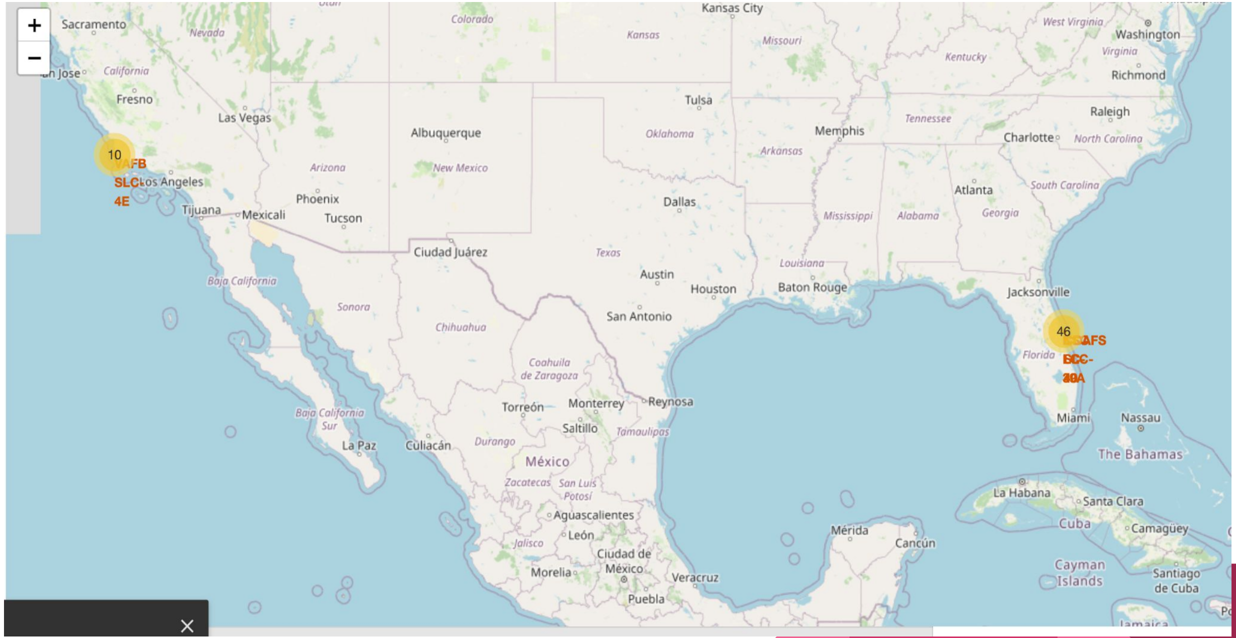
Map of launch sites

- The red circles demonstrate 4 launch site locations: The launch sites are near the equator and coastlines with 3 sites on the East Coast and 1 site on the West Coast.

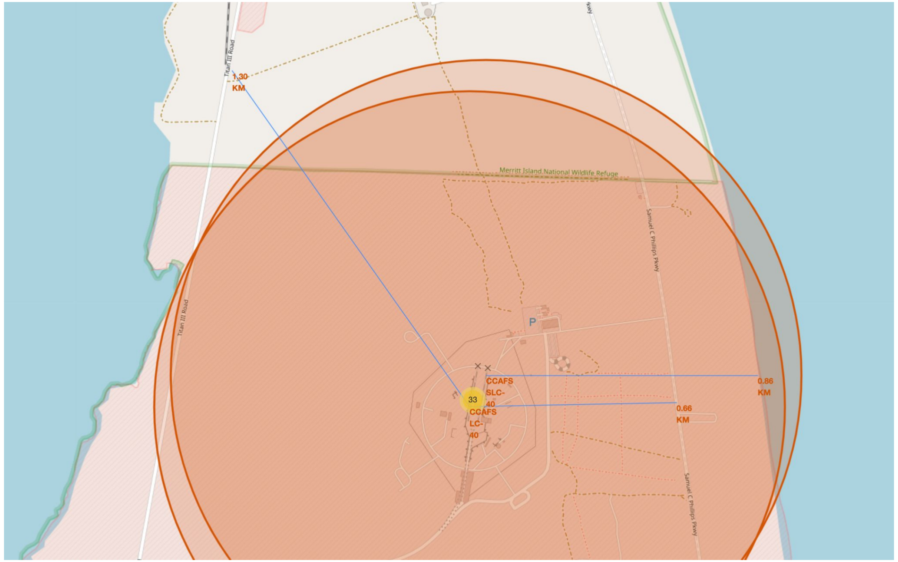


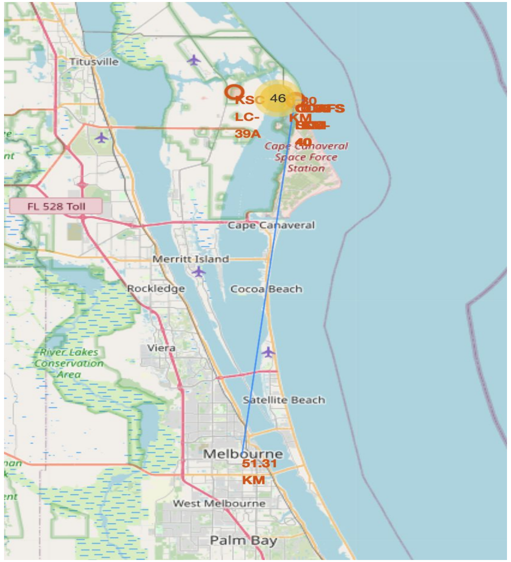
Success/failed launches map

- The maps show the launch outcomes in marker clusters: with successful launch in green and failed launches in red.



Distance from launch sites to its proximities



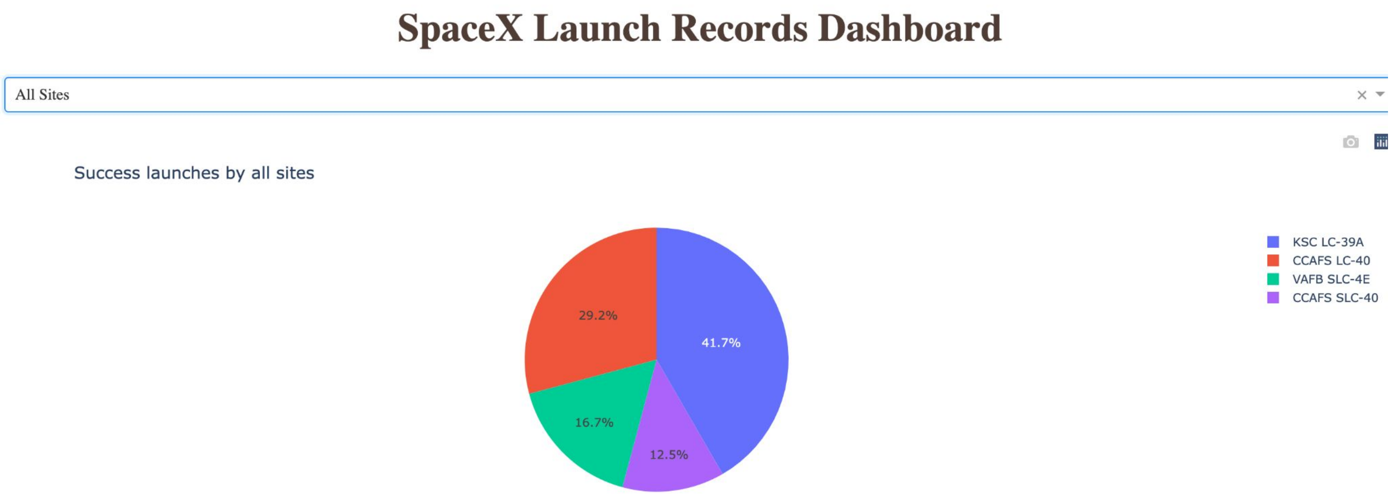


Launch sites are close to the coastline, but keep some distance from railroads, highways, and are far from cities.

Section 5

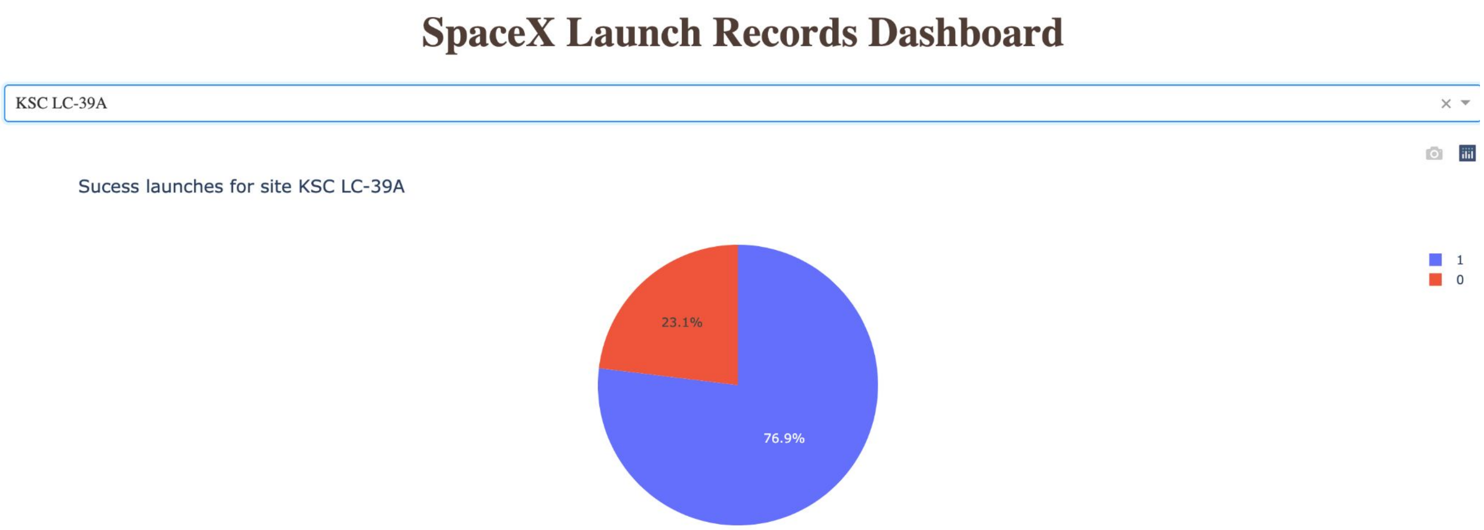
Build a Dashboard with Plotly Dash

Pie chart of successful launches of all sites



The chart shows that of the total success launches, KSC LC-39A is 41.7%, CCAFS LC-40 is 29.2%,VAFB SLC-4E is 16.7% and CCAFS SLC-40 is 12.5%

Pie chart for launch site with highest launch success ratio



The chart shows that KSC LC-39A has the highest success launch ratio of 76.9%

Payload vs. Launch Outcome scatter plot for all sites

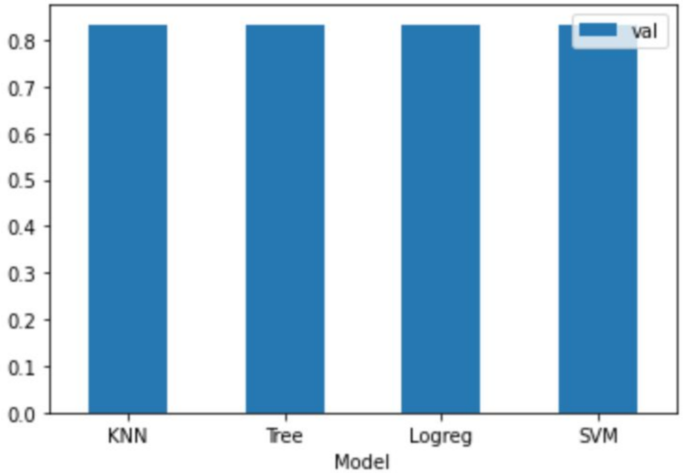
- Payload range with largest success rate: 2000-5000kg
- Payload larger than 6000 kg has the lowest success rate
- Booster version have the largest success rate: F9 v1.1

Section 6

Predictive Analysis (Classification)

Classification Accuracy

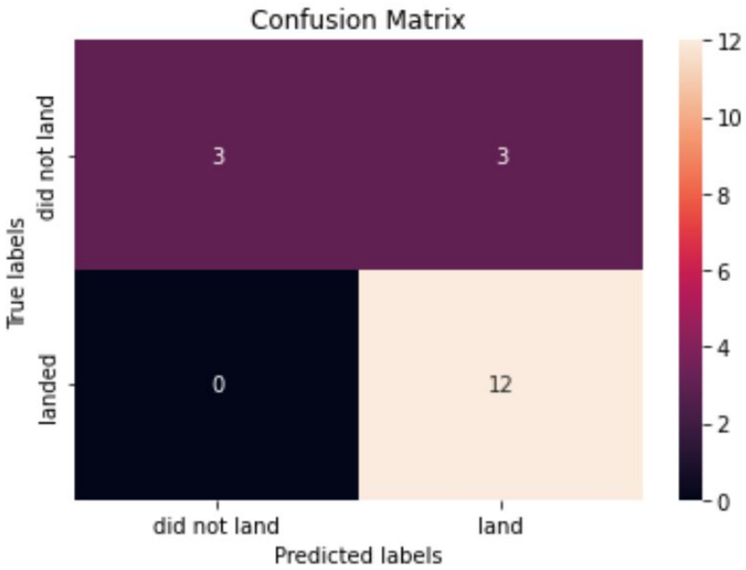
- Model accuracy for all built classification models, in a bar chart:



- Due to small data size, logistic regression, support vector machine and K nearest neighbors has the same highest classification accuracy of 83.34%

Confusion Matrix

- The confusion matrix of the best performing models show that the main problem with the models is false positive.



Project conclusion

- Factors that affect the outcome of the launch are payload, booster version, orbit type, launch sites.
- Location of a launch site must be near coastlines and far away from cities, highways and railroads.
- Using machine learning model, we can predict with 83.34% accuracy that whether the first stage of Falcon 9 will have a successful landing or not.

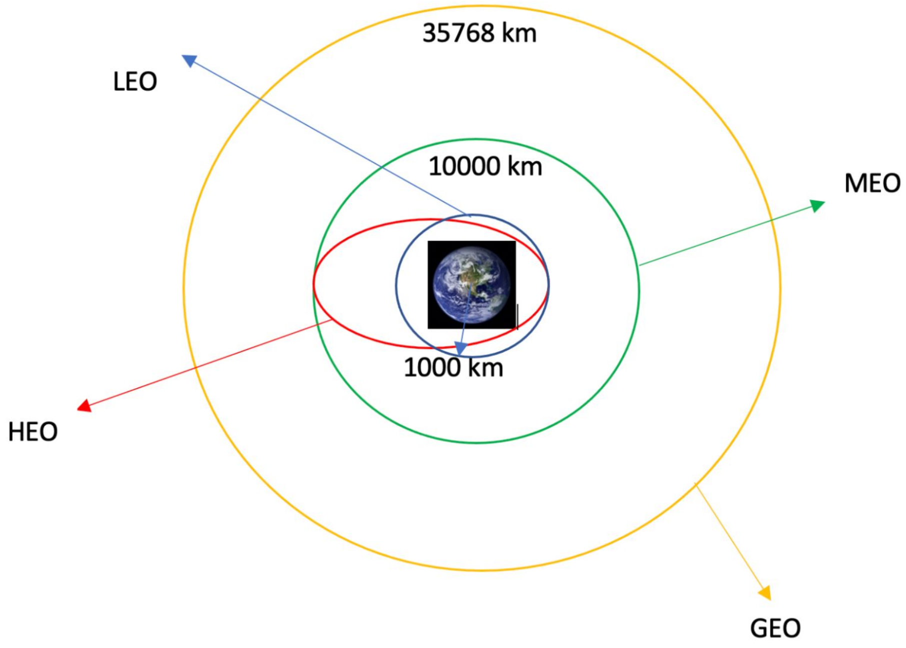
Appendix

Data set:

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitud
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.57736
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.57736
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.57736
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.61082
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.57736
5	6	2014-01-06	Falcon 9	3325.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1005	-80.57736
6	7	2014-04-18	Falcon 9	2296.000000	ISS	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1006	-80.57736
7	8	2014-07-14	Falcon 9	1316.000000	LEO	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1007	-80.57736
8	9	2014-08-05	Falcon 9	4535.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1008	-80.57736
9	10	2014-09-07	Falcon 9	4428.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1011	-80.57736

Appendix

Earth orbit:



Appendix

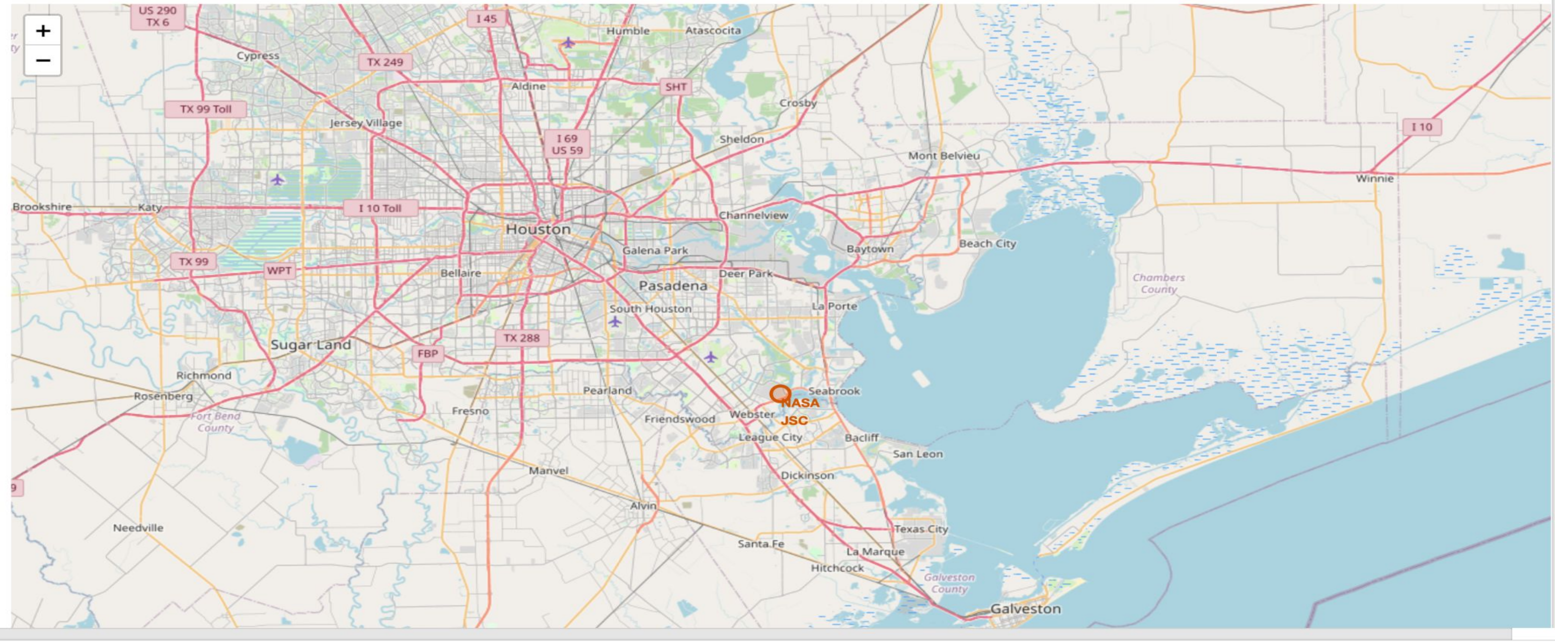
Use one-hot coding to create dummy variables for categorical data

	FlightNumber	PayloadMass	Flights	GridFins	Reused	Legs	Block	ReusedCount	Orbit_ES-L1	Orbit_GEO	Orbit_GTO	Orbit_HEO	Orbit_ISS	Orbit_LEO	Orbit_MEO	(
0	1.0	6104.959412	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	
1	2.0	525.000000	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	
2	3.0	677.000000	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	
3	4.0	500.000000	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
4	5.0	3170.000000	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
85	86.0	15400.000000	2.0	1.0	1.0	1.0	5.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
86	87.0	15400.000000	3.0	1.0	1.0	1.0	5.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
87	88.0	15400.000000	6.0	1.0	1.0	1.0	5.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
88	89.0	15400.000000	3.0	1.0	1.0	1.0	5.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
89	90.0	3681.000000	1.0	1.0	0.0	1.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	

90 rows × 80 columns

Appendix

Folium map of NASA Johnson Space Center at Houston, Texas



Thank you!