IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Linh Phan
09/02/2023

Outline

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

Executive Summary

- Summary of methodologies
- Summary of all results

Introduction

LAUNCHING ROCKETS IS EXPENSIVE! SPACEX has publicly stated that each launch of Falcon 9 Rocket costs 62 million dollars.

Since the first stage of a Falson 9 rocket can be reused. We would like to predict if the Falson 9 first stage will land successfully. The cost of Falcon 9 rocket launching is extremely large thus knowing if the first stage will land, we can determine the cost of a launch easily.

If we want to bid against SpaceX for a Rocket laucnh, we need to estimate their launch cost to have a promising quotation.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The data was collected using REST API and Web Scrapping using BeautifulSoup
- Perform data wrangling
  - Select only useful attributes and remove/replace null values. We also classify the outcome of the landing and create a class = 1 (Sucess) and class = 0 (Failure)
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Use train_test_split to split up the data, then use GridSearCV and multiple models to fit and score to find the best classification models and parameters for ML model.

Data Collection

- The Data Collection process consists of:
  - Using Requests to SpaceX API to retrieve records of unique Launch ID such as BoosterVersion, Launch Site, Payload etc.
  - Generate a dataframe from all the data collected and filter the data to only contain Falcon 9 launches.
  - Using BEAUTIFULSOUP to scrape data from Wikipedia HTML tables, and convert it into a dataframe.

2020 [edit]

In late 2019, Gwynne Shotwell stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020,[490] in addition to 14 or 15 non-Starlink launches. At 26 launches, 13 of which for Starlink satellites, Falcon 9 had its most prolific year, and Falcon rockets were second most prolific rocket family of 2020, only behind China's Long March rocket family.[491]

[hide] Flight No.	Date and time (UTC)	Version, Booster[b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:19:21[492]	F9 B5 △ B1049.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third large batch and second operational flight of Starlink constellation. One of the 60 satellites included a test coating to make the satellite less reflective, and thus less likely to interfere with ground-based astronomical observations.[493]								
79	19 January 2020, 15:30[494]	F9 B5 △ B1046.4	KSC, LC-39A	Crew Dragon in-flight abort test[495] (Dragon C205.1)	12,050 kg (26,570 lb)	Sub-orbital[496]	NASA (CTS)[497]	Success	No attempt
	An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes after reentry, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule;[498] but that test article exploded during a ground test of SuperDraco engines on 20 April 2019.[419] The abort test used the capsule originally intended for the first crewed flight.[499] As expected, the booster was destroyed by aerodynamic forces after the capsule aborted.[500] First flight of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its engine.								
80	29 January 2020, 14:07[501]	F9 B5 △ B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)
	Third operational and fourth large batch of Starlink satellites, deployed in a circular 290 km (180 mi) orbit. One of the fairing halves was caught, while the other was fished out of the ocean.[502]								
81	17 February 2020, 15:05[503]	F9 B5 △ B1056.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fourth operational and fifth large batch of Starlink satellites. Used a new flight profile which deployed into a 212 km × 386 km (132 mi × 240 mi) elliptical orbit instead of launching into a circular orbit and firing the second stage engine twice. The first stage booster failed to land on the drone ship[504] due to incorrect wind data.[505] This was the first time a flight proven booster failed to land.								
82	7 March 2020, 04:50[506]	F9 B5 △ B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 △)	1,977 kg (4,359 lb)[507]	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
	Last launch of phase 1 of the CRS contract. Carries *Bartolomeo*, an ESA platform for hosting external payloads onto ISS.[508] Originally scheduled to launch on 2 March 2020, the launch date was pushed back due to a second stage engine failure. SpaceX decided to swap out the second stage instead of replacing the faulty part.[509] It was SpaceX's 50th successful landing of a first stage booster, the third flight of the Dragon C112 and the last launch of the cargo Dragon spacecraft.								
83	18 March 2020, 12:16[510]	F9 B5 △ B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Failure (drone ship)
	Fifth operational launch of Starlink satellites. It was the first time a first stage booster flew for a fifth time and the second time the fairings were reused (Starlink flight in May 2019).[511] Towards the end of the first stage burn, the booster suffered premature shut down of an engine, the first of a Merlin 1D variant and first since the CRS-1 mission in October 2012. However, the payload still reached the targeted orbit.[512] This was the second Starlink launch booster landing failure in a row, later revealed to be caused by residual cleaning fluid trapped inside a sensor.[513]								
84	22 April 2020, 19:30[514]	F9 B5 △ B1051.4	KSC, LC-39A	Starlink 6 v1.0 (60 satellites)	15,600 kg (34,400 lb)[5]	LEO	SpaceX	Success	Success (drone ship)

Data Collection – SpaceX API

- Github link: https://github.com/phanlinh94/datascience/blob/ee4da9255e7735e4447b7ac41b8b16a456b6a89c/API%20Data%20Collection.ipynb

- The output of Data Collection from SpaceX API:

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004

Import Supporting Libraries and Functions

Using **SpaceXData API at api.spacexdata.com** to retrieve launch data

Normalize the **JSON response** to create a dictionary with Launch ID and information such as Payload, BoosterVersion, Launch Site etc.

Mapping **the Lauch ID with information gathered using other APIs** to create a simplified, concise dictionary

Build **dictionary into a dataframe** and **filter** to contain only **Falcon 9 launches**

Data Collection - Scraping

- Github link:

https://github.com/phanlinh94/datascience/blob/f15057a4bd4505bafc44e30b421f9c09effd659f/Data%20Collection%20with%20Web%20Scraping.ipynb

- The result of Web Scrapping:

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA (COTS)\nNRO	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA (COTS)	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA (CRS)	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA (CRS)	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10
...	...	...	...	...	...	...	...	...	...	...	...
116	117	CCSFS	Starlink	15,600 kg	LEO	SpaceX	Success\n	F9 B5B1051.10	Success	9 May 2021	06:42
117	118	KSC	Starlink	~14,000 kg	LEO	SpaceX Capella Space and Tyvak	Success\n	F9 B5B1058.8	Success	15 May 2021	22:56
118	119	CCSFS	Starlink	15,600 kg	LEO	SpaceX	Success\n	F9 B5B1063.2	Success	26 May 2021	18:59
119	120	KSC	SpaceX CRS-22	3,328 kg	LEO	NASA (CRS)	Success\n	F9 B5B1067.1	Success	3 June 2021	17:29
120	121	CCSFS	SXM-8	7,000 kg	GTO	Sirius XM	Success\n	F9 B5	Success	6 June 2021	04:26

Request Falcon 9 Launch Data from Wikipedia URL

Using **BeautifulSoup** to scrape the tables containing information

Extract all the **column names**, create an empty dictionary

Iterate each row of the data table from **BeautifulSoup scrapping** and **append the result into a Dataframe**

Data Wrangling

- After Data Collection, the data was assessed to check for NaN values.
- Missing values were then replaced with the mean (PayloadMass)
- The resulting dataframe contains no missing values.
- Github link: https://github.com/phanlinh94/datascience/blob/8a4362d99cf529f2db3c36ab3949c963451725d7/Data%20Wrangling.ipynb

Data collection

Check for Null values in the dataframe

Replace the Null vales with the mean of each column

EDA with Data Visualization

- Data visualization employed to discover the following correlations with the Outcomes
  - Flight Number and Payload Mass (kg) - Scatterplot
  - Flight Number and Launch Site – Scatterplot
  - Orbit Types – Bar plot
  - Flight Number and Orbit Types – Scatterplot
  - Payload and Orbit Types – Scatterplot
  - Success Trend over the years – Line plot
- GitHub Link: https://github.com/phanlinh94/datascience/blob/474210371f6d1e9c6939a1e1982a4cc39d51940f/EDA%20with%20Visualization.ipynb

EDA with SQL

- SQL Queries that were performed:
  - Names of all the different launch sites for SpaceX
  - 5 records of Launching at CCAFS Launch site.
  - Total Payload Mass that was carried out by boosters for NASA (CRS)
  - Average Payload Mass carried by Booster Version F9 V1.1
  - Name of the ground pad where the first successful launch was achieved
  - List of Boosters that had successful drone ship launches with Payload Mass in the range between 4000 to 6000kg
  - Total number of Successful and Failed Mission Outcomes
  - Name of the Booster Version that carried that maximum Payload Mass
  - Failed Drone ship launches with information about Booster Versions, Launch Sites in 2015
  - Total number of outcomes for different types of landing between 2010/06/04 and 2017/03/20
- GitHub Link: https://github.com/phanlinh94/datascience/blob/cc3d3f8e7bbabeeb33e06f32cbd2f0b1045d49b6/EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Map objects added on the Folium Map
  - Launch Sites with Names, circle markers
  - Markers with different colors of different launch outcomes for each launch site
  - Calculate the distances between the Launch site to different proximities: coastline, closest city, railway, highway etc.
- The objects were to:
  - Visualize geographically the launch sites
  - See the clusters of launch sites with the most positive results as well as negative ones in one holistic look
  - Visualize the distances to different locations that might contribute to the success of the launch
- GitHub Link: https://github.com/phanlinh94/datascience/blob/dea25bc261681d2d856bdad042c0655e413c743e/Interactive%20Visual%20Analytics.ipynb

Build a Dashboard with Plotly Dash

- Components of the Plotly Dash:
  - A pie chart showing success counts for all sites/each site of launching based on selection from a Dropdown setting
  - A scatter plot showing launch outcomes for all launches with range of Payload Mass defined by a slider at the launch site selected from a Dropdown setting
- These graphs will provide information about:
  - Sites that have the largest number of successful launches
  - Sites that have the highest launch success rate
  - Payload ranges that give the highest launch success rate
  - Payload ranges that give the lowest launch success rate
  - F9 Booster Version that has the highest launch success rate
- GitHub Link: https://github.com/phanlinh94/datascience/blob/cc2bb37b1adf5956635fb9c75b9389890933416b/spacex_dash_app.py

Predictive Analysis (Classification)

- ML training:
  - Standardize the data and split data using train_test_split
  - Fit the training data into different ML models: Logistic Regression, SVM, Decision Tree, KNN etc. to find the best parameters, best scores
  - Determine which ML model is the best Model for this dataset
- GitHub Link: https://github.com/phanlinh94/datascience/blob/cc2bb37b1adf5956635fb9c75b9389890933416b/IBM-DS0321EN-SkillsNetwork_labs_module_4_SpaceX_Machine_Learning_Prediction_Part_5.jupyterlite.ipynb

Standardize the Data

Split the data into training set and test set using train_test_split

Fit the training data into different models: **LogReg, DecisionTree, KNN, SVM etc.**

Find the **best parameters, best scores** for each model

Determine the **best ML model** for SpaceX dataset

Results

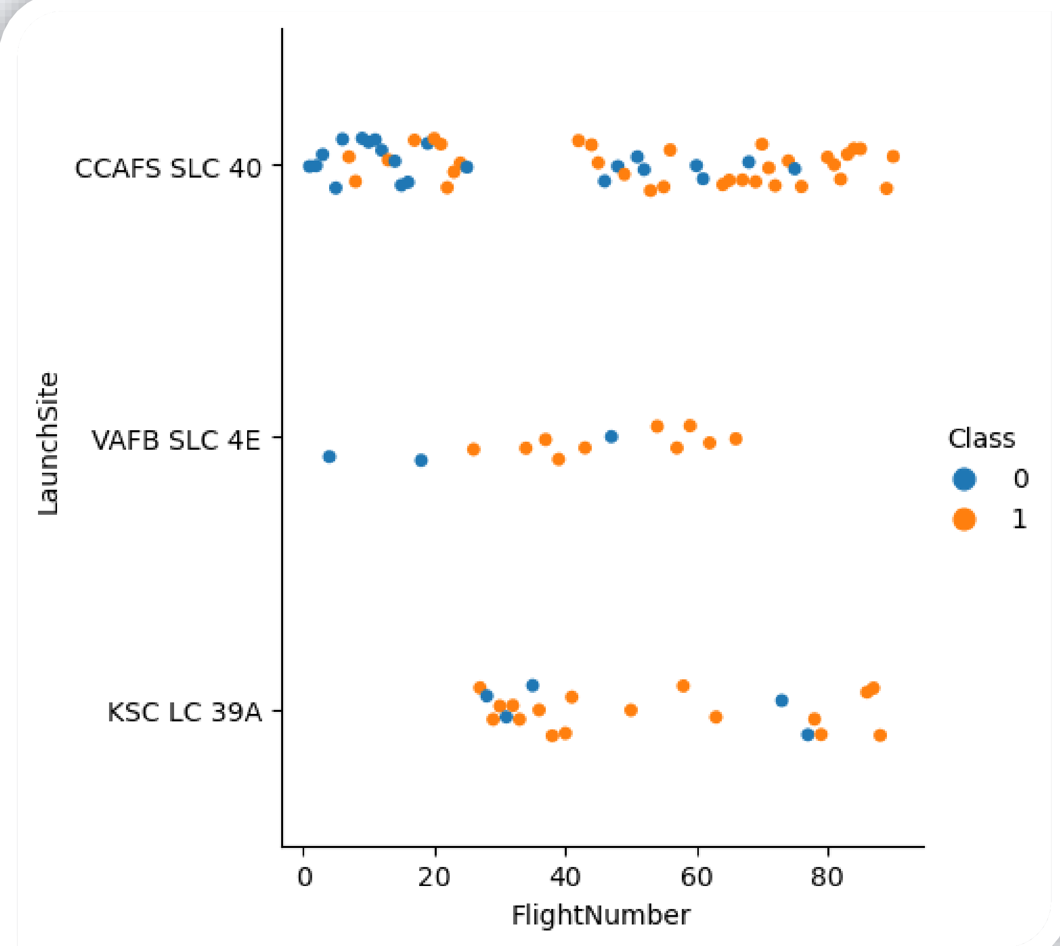
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

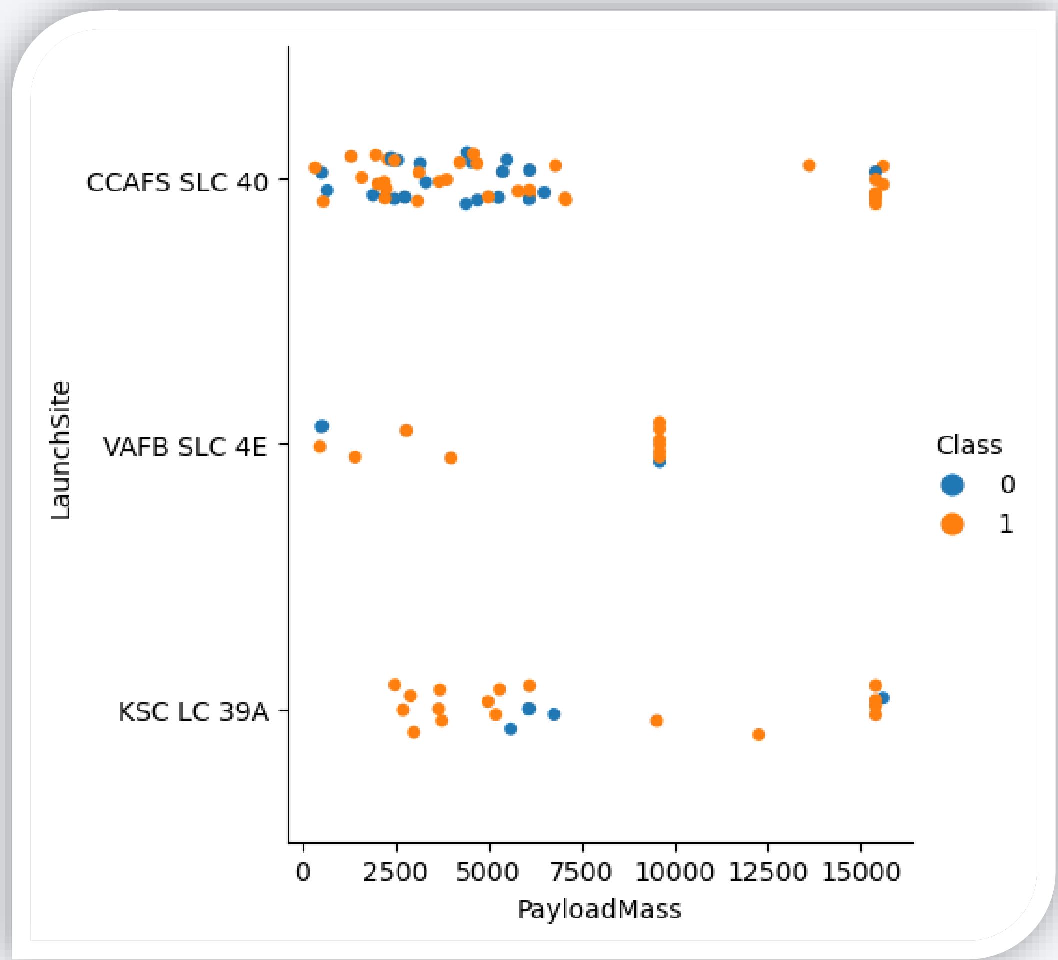
Flight Number vs. Launch Site

- The scatterplot shows the correlation between the launch sites and the success outcomes.
- Blue: Failure
- Orange: Success
- We can see that for VAFB SLC-4E and KSC LC-39A, the ratio of successful launches are quite high while CCAFS SLC-40 is a mix of both successful and failed launches.



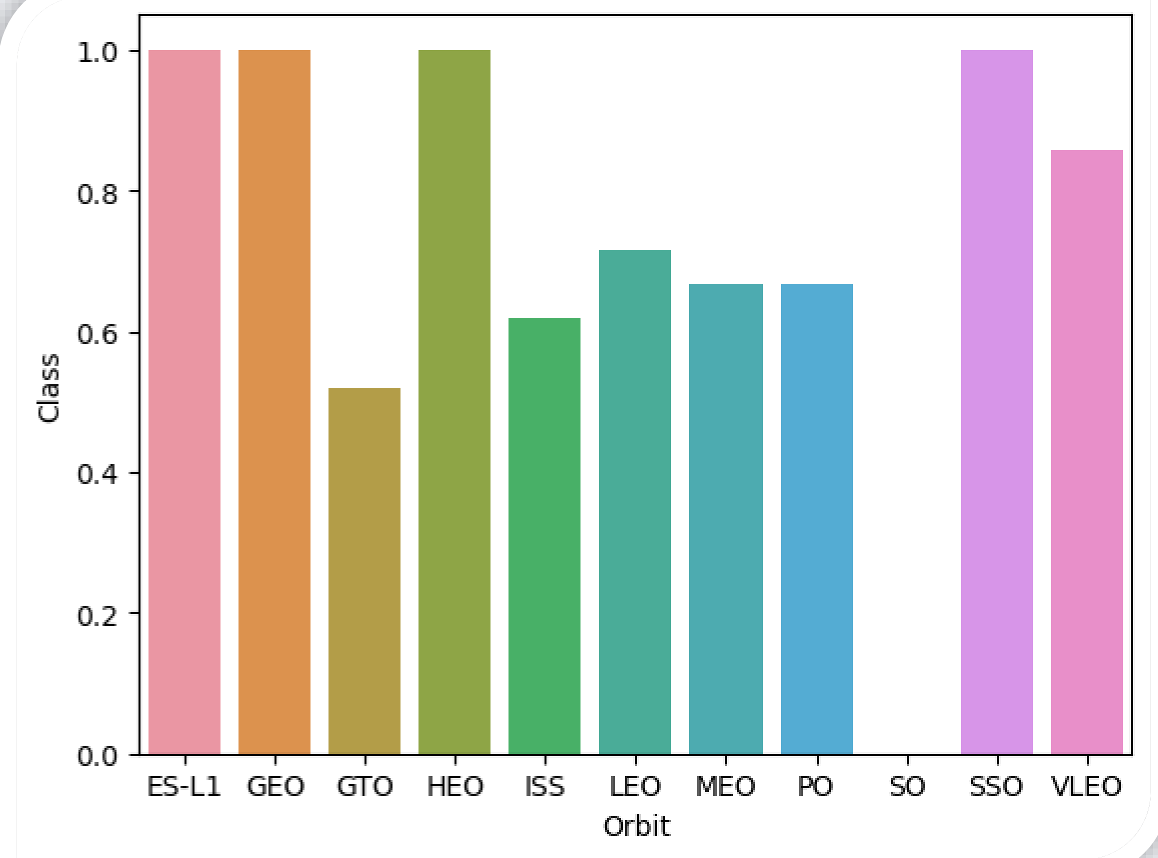
Payload vs. Launch Site

- The scatterplot shows the correlation between the payloads and the success outcomes at each site.
- Blue: Failure
- Orange: Success
- Lower payload launches share a good amount of failed trials. While higher payload launches observe a significantly better results, especially with **Payload Mass > 7500 kg at all sites.**



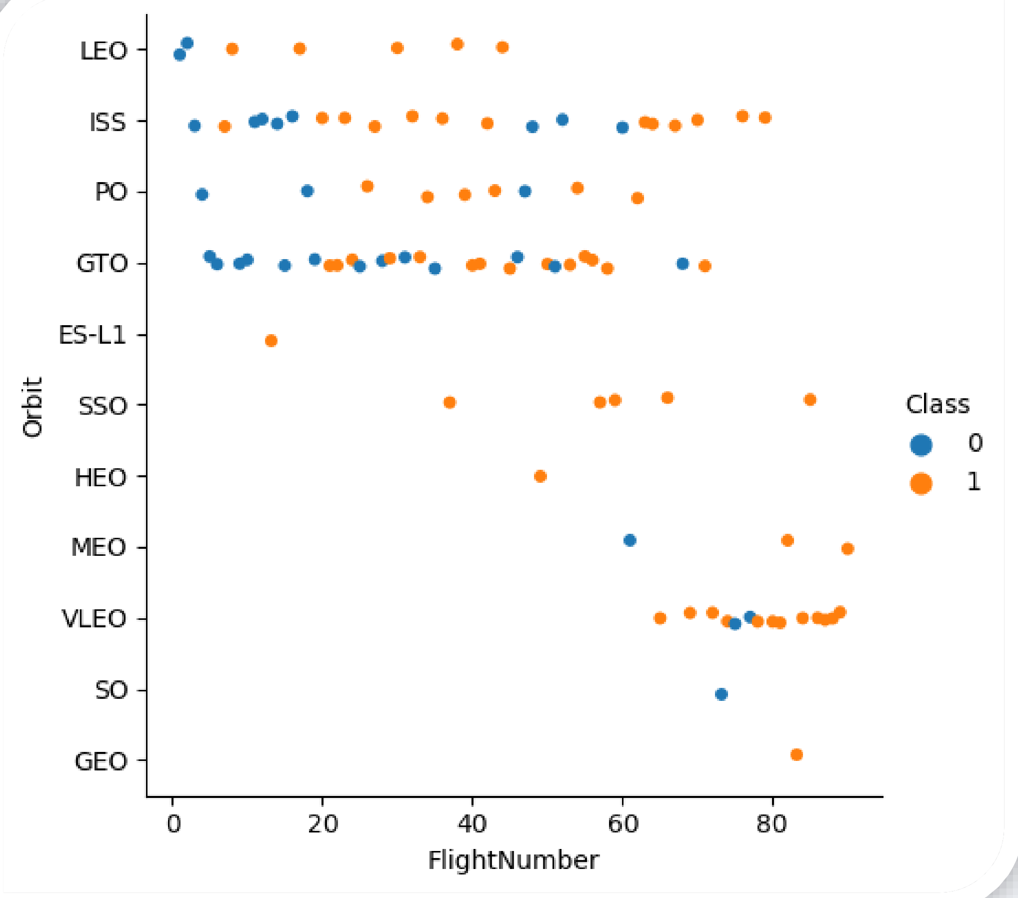
Success Rate vs. Orbit Type

- The Bar chart shows the average success rates of all Orbit types.
- Values:
  - Success = 1.0
  - Failure = .0
- The chart shows that 4 Orbit types have the highest rate of success (ES-L1, GEO, HEO and SSO) while GTO orbit yields the lowest success rate.



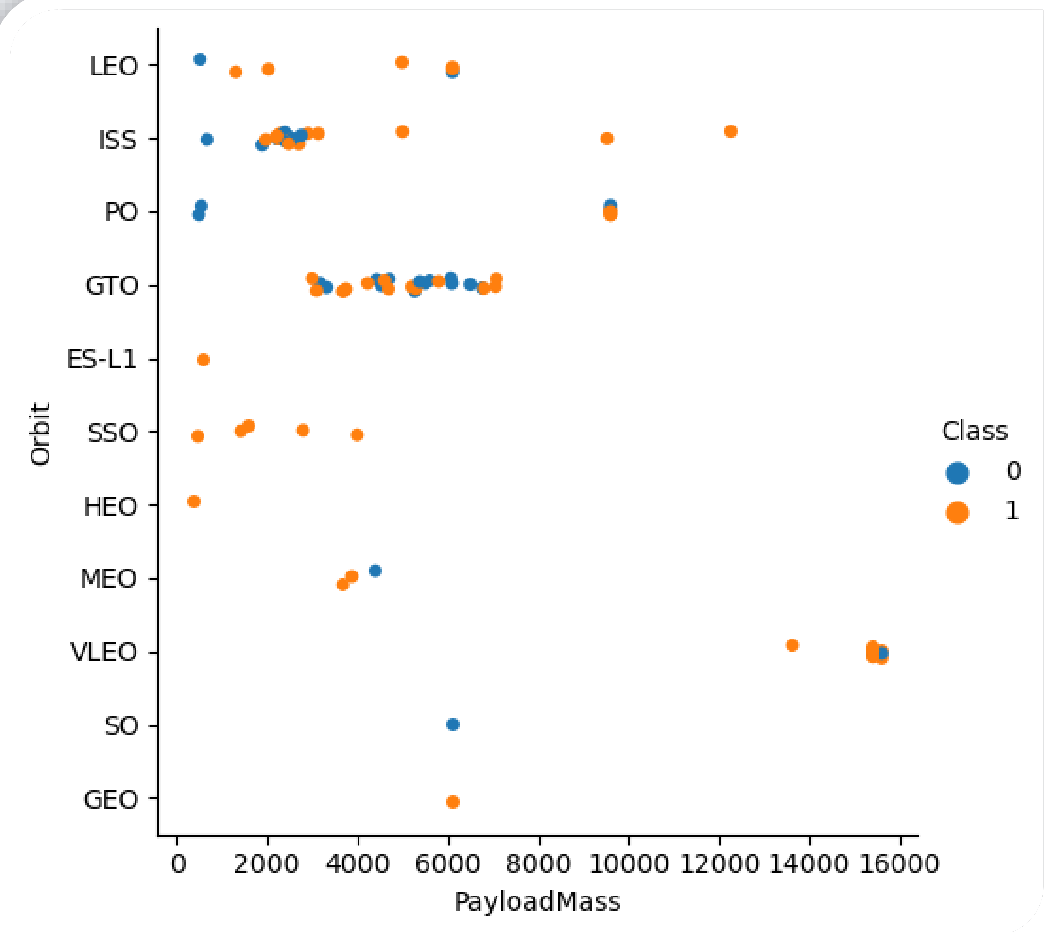
Flight Number vs. Orbit Type

- The scatterplot shows the correlation between the Flight Number and the success outcomes with each Orbit Type.
- Blue: Failure
- Orange: Success
- Most of the launches are in Orbit types: LEO, ISS, PO, GTO and VLEO. With a particularly high success rate at Orbit type SSO, LEO and VLEO. **HEO, SO, GEO only has 1 launch for each.**



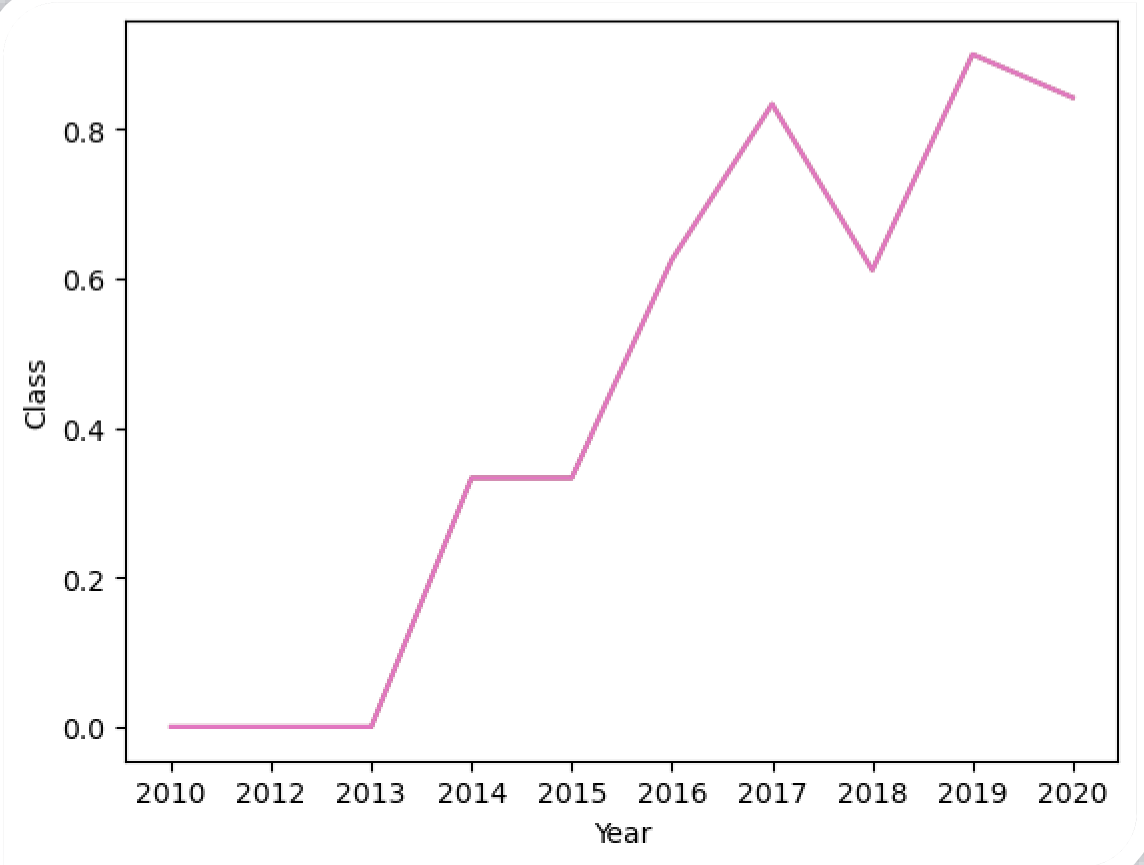
Payload vs. Orbit Type

- The scatterplot shows the correlation between the Payload Mass and the success outcomes with each Orbit Type.
- Blue: Failure
- Orange: Success
- Most of the launches are at Payload Mass lower than 8000 kg. SSO has very good success rate at low Payload Mass (<6000 kg). GTO does observe both failed and successful launches at a range of Payload from ~3000 to 7000 kg.



Launch Success Yearly Trend

- The Line Chart shows the average success rate over the years. It can be clearly seen that the average success rate has increased significantly overall, with a slight dip at 2018.



All Launch Site Names

- Names of all Unique Launch Sites

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

- *%sql* SELECT **DISTINCT (LAUNCH_SITE)** FROM **SPACEX**
  - This query selects distinct values in the column Launch_Site from the Table SpaceX

Launch Site Names Begin with 'CCA'

- 5 records where launch sites begin with `CCA`

```
* ibm_db_sa://tzf18981:***@b1bc1829-6f45-4cd4-bef4-10cf081900bf.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32304/bludb
Done.
```

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

- *%sql* SELECT * FROM **SPACEX**
- WHERE **LAUNCH_SITE LIKE 'CCA%'** LIMIT 5
  - This query selects 5 rows from SpaceX table with names in Launch_Site column beginning with ‘CCA’

Total Payload Mass

- Total Payload Mass (kg) by Boosters launched by NASA: **45 596 (kg)**
- *%sql* SELECT **SUM(PAYLOAD_MASS__KG_)** FROM **SPACEX**
- WHERE **CUSTOMER = 'NASA (CRS)'**
    - This query calculates the total of all values in Payload_Mass__Kg_ column in SpaceX table with the value in Customer column being NASA (CRS)

Average Payload Mass by F9 v1.1

- The Average Payload Mass by Booster F9 V1.1 is **2 534** kg
- *%sql* SELECT **AVG(PAYLOAD_MASS__KG_)** FROM **SPACEX**
- WHERE **BOOSTER_VERSION LIKE 'F9 v1.1%'**
    - This query calculates the mean value for all values in Payload_Mass__kg_ column in the SpaceX table with the Booster_Version containing F9 V1.1______

First Successful Ground Landing Date

- The dates of the first successful landing outcome on ground pad: **2015-12-22 (22nd of December, 2015)**
- *%sql* SELECT **MIN(DATE)** FROM **SPACEX**
- WHERE **LANDING__OUTCOME = 'Success (ground pad)'**
  - This query finds the minimum value in the Date column of SpaceX table where the Landing_Outcome specifies that it is successful on the Ground Pad.

Successful Drone Ship Landing with Payload between 4000 and 6000

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

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

- *%sql* SELECT **BOOSTER_VERSION, LANDING__OUTCOME, PAYLOAD_MASS__KG_** FROM **SPACEX**
- WHERE **LANDING__OUTCOME = 'Success (drone ship)'** AND **PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000**
  - This query selects 3 columns from SpaceX table that contain Booster Version, Outcome of the Landing and Payload Mass with 2 conditions: the landing is successful on the Drone ship and the Payload Mass is between 4000 and 6000 kg.

Total Number of Successful and Failure Mission Outcomes

- Total number of successful and failure mission outcomes

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

- *%sql* SELECT **MISSION_OUTCOME, COUNT(*)** FROM **SPACEX**
- GROUP BY **MISSION_OUTCOME**
  - This query extract the Mission Outcome values from SpaceX table and group and count the number of each unique value.

Boosters Carried Maximum Payload

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

- The names of the booster which have carried the maximum payload mass (on the right-hand side)

- %sql SELECT **BOOSTER_VERSION, PAYLOAD_MASS__KG_** FROM **SPACEX**
- WHERE **PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEX)**
  - This query uses subquery to find the Maximum Payload Mass, then selects all the rows that match the Payload Max. This only shows the Booster Version and Payload_Mass__KG_ to simplify the view.

2015 Launch Records

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

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

- *%sql* SELECT **DATE, BOOSTER_VERSION, LAUNCH_SITE, LANDING__OUTCOME** FROM **SPACEX**
- WHERE **DATE LIKE '2015%'** AND **LANDING__OUTCOME = 'Failure (drone ship)'**
  - This query selects 4 columns Date, Booster Version, Launch Site and Landing Outcome from SpaceX table with Date starting with 2015 (in 2015) and the Outcome is “Failure (Drone Ship)”.

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(*) AS COUNT_NUMBER** FROM **SPACEX**
- WHERE **DATE BETWEEN '2010-06-04' AND '2017-03-20'**
- GROUP BY **LANDING__OUTCOME**
- ORDER BY **COUNT_NUMBER DESC**
    - This query counts the number of each of Landing Outcome unique values that between 2010-06-04 and 2017-03-20, then sort the count values.

landing__outcome	count_number
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

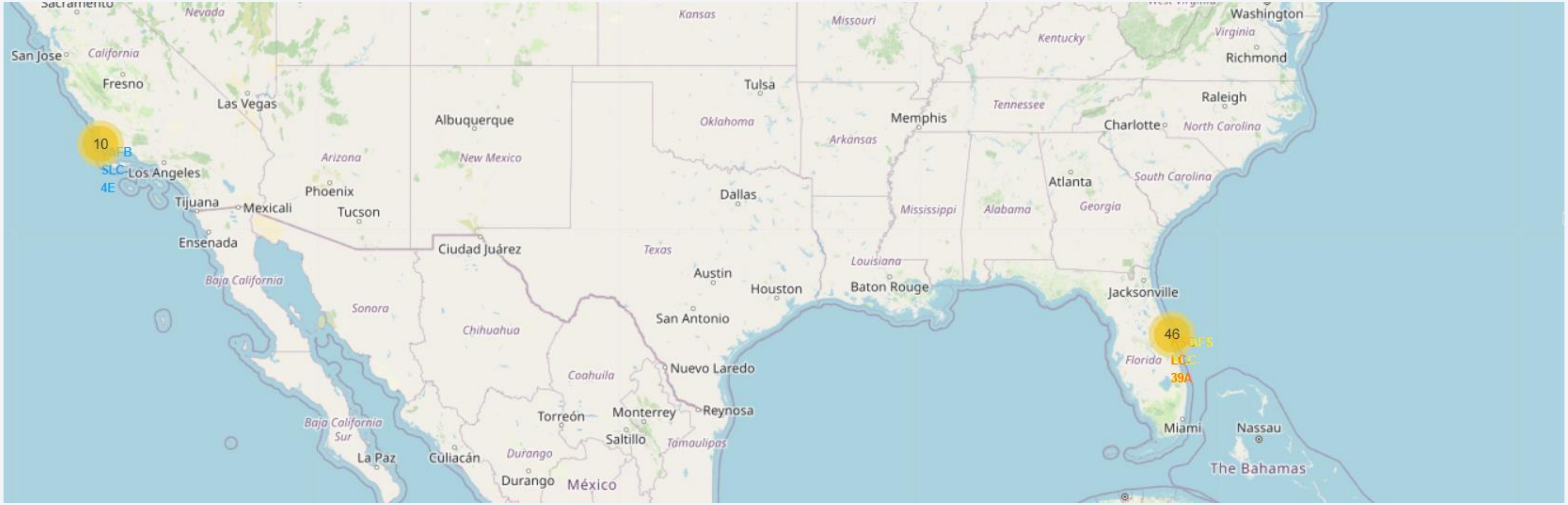
Section 3

Launch Sites Proximities Analysis

Map of All Launch Sites

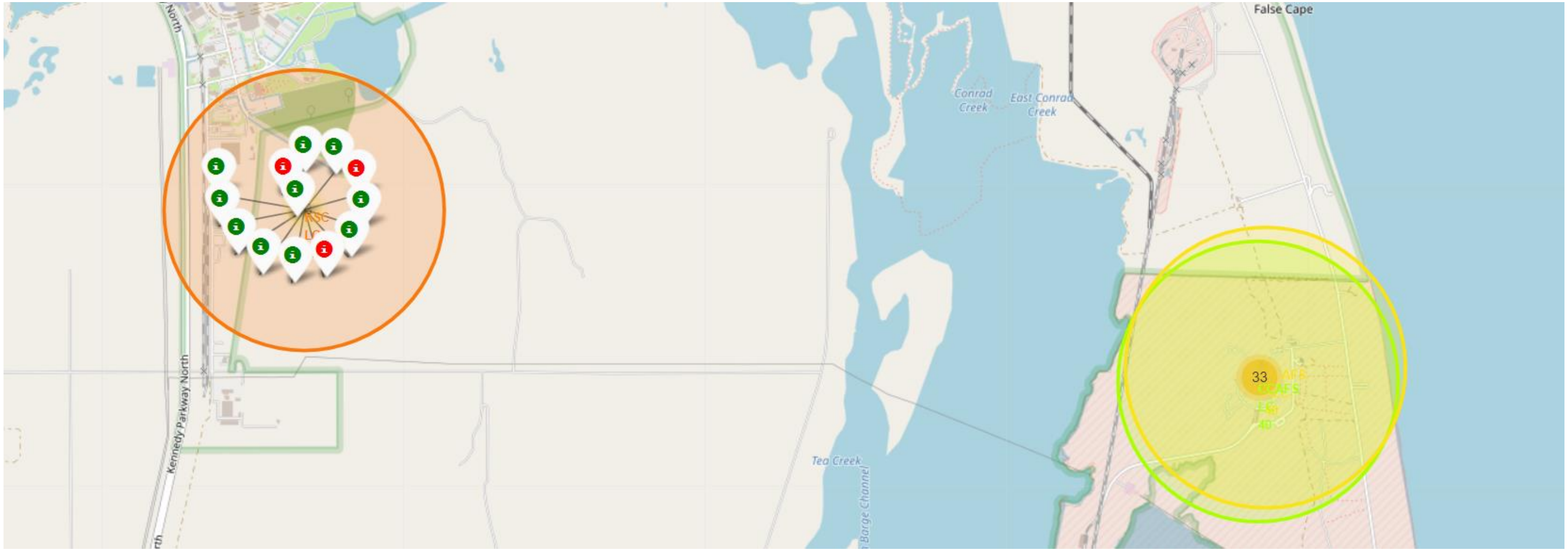
- This map shows all the launch sites of SpaceX with different colours. Blue on the West Coast and Yellow, Orange on the East Coast.
- Each has the Corresponding ID.

Success/Failed Launches for Each site

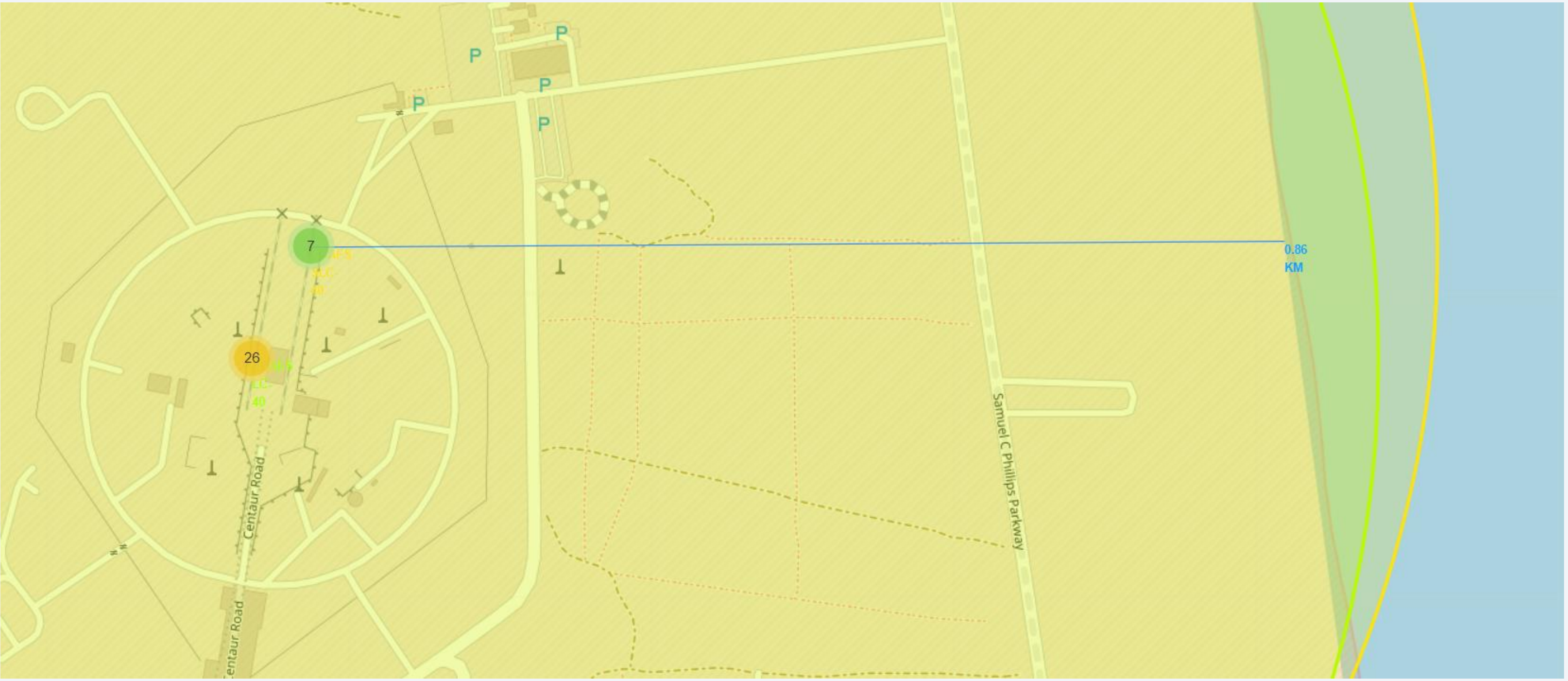


- This map shows the numbers of success/failed launches for each site on the map.

This is what each circle looks like when it is zoomed out. Green – Success, Red - Failed



Distance from CCAFS SLC-40 Launch site to the closest coastline

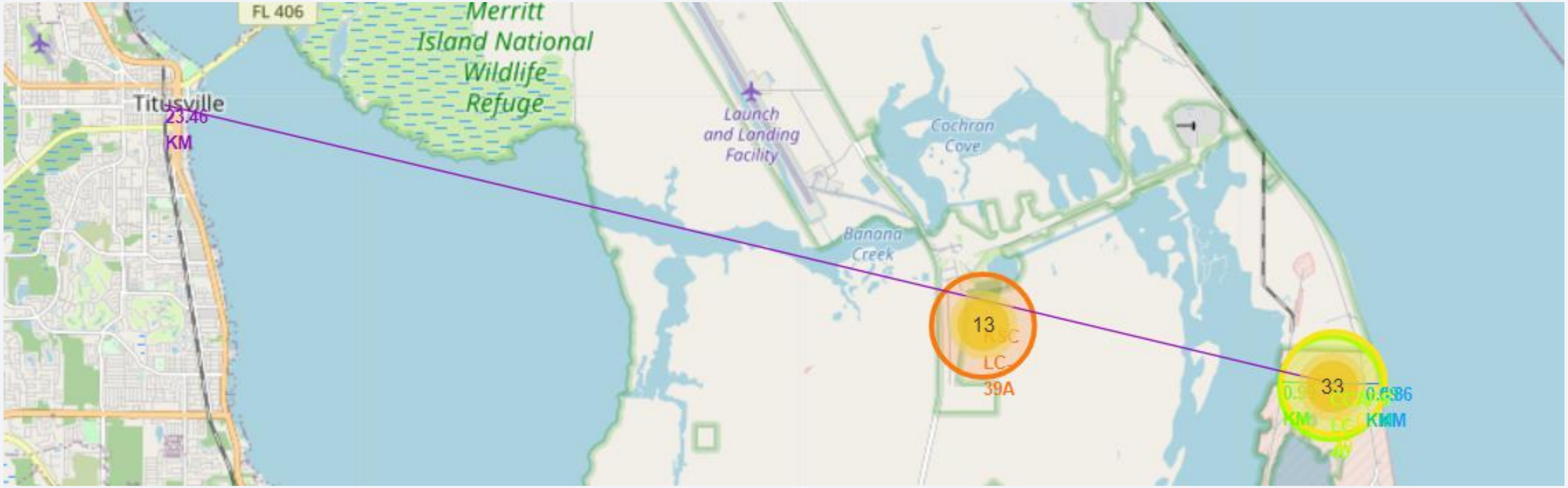


- The launch sites are very close to the coastline, only a distance of ~0.86km.

Distance from CCAFS SLC-40 Launch site to the closest Highway and Railway

- The launch sites are very close to the railway and highway, with a distance of 0.99km and 0.59km respectively.

Distance from CCAFS SLC-40 Launch site to the closest City (Titusville) – The purple line

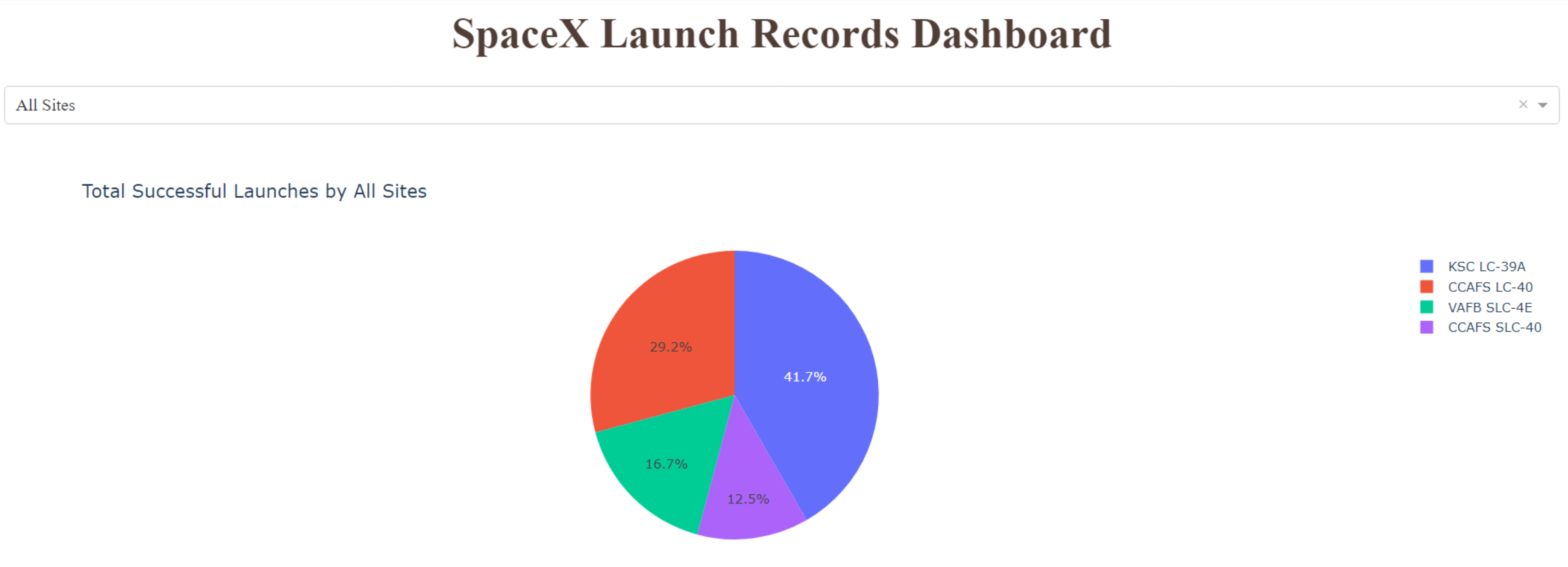


- The launch sites are quite far the city, with a distance of ~23.5km.

Section 4

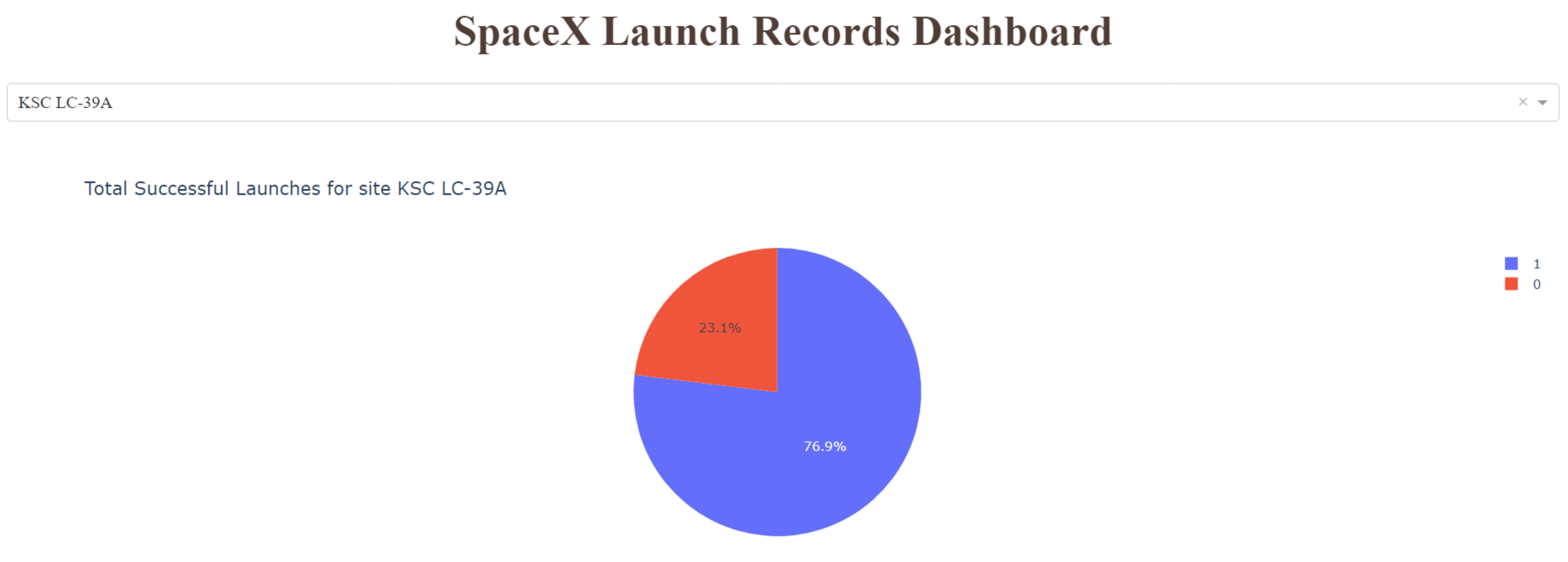
Build a Dashboard with Plotly Dash

Plotly Dashboard – Success percentage of Each Launch Site



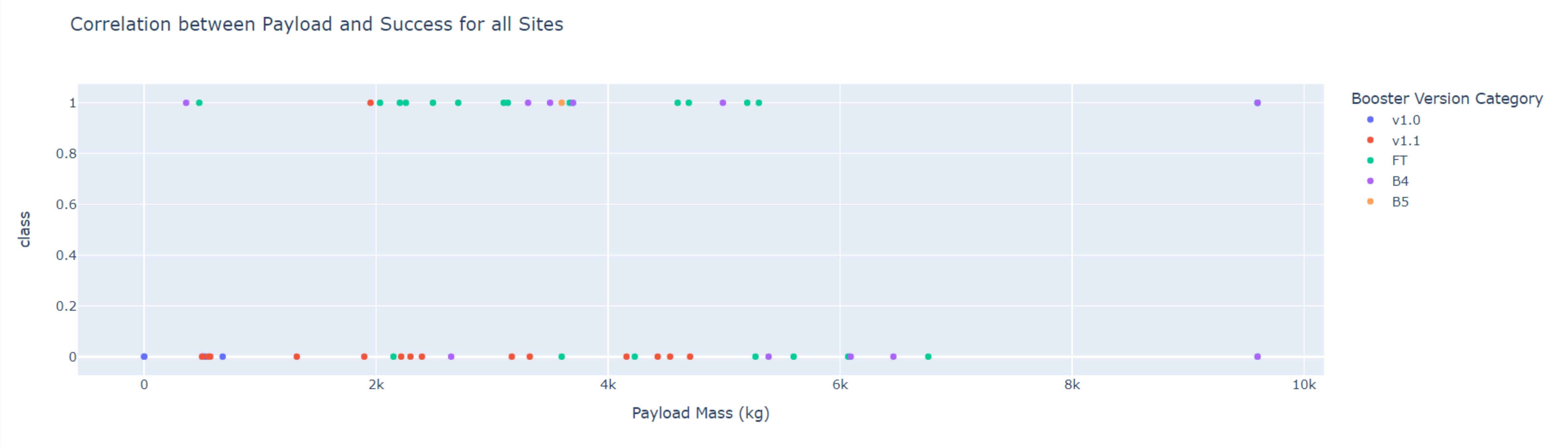
- The pie chart shows that the highest percentage of successful launches is at KSC LC-39A site (~42%) and the lowest is CCAFS SLC-40 (~12.5%)

Ratio of Success/Fail for KSC LC-39A site



- KSC LC-39A does not only have the highest number of success launches, it also has a high ratio of success/fail compared to all other launch sites. The **success ratio is over 76%**

Correlation between Payload and Success (All Sites)



- The scatterplot shows that a significant number of successful launches have **Payload between 2000 and 4000 kg.** This also shows that within this range, **FT** is the **best Booster Version** with the highest success rate.

Section 5

Predictive Analysis (Classification)

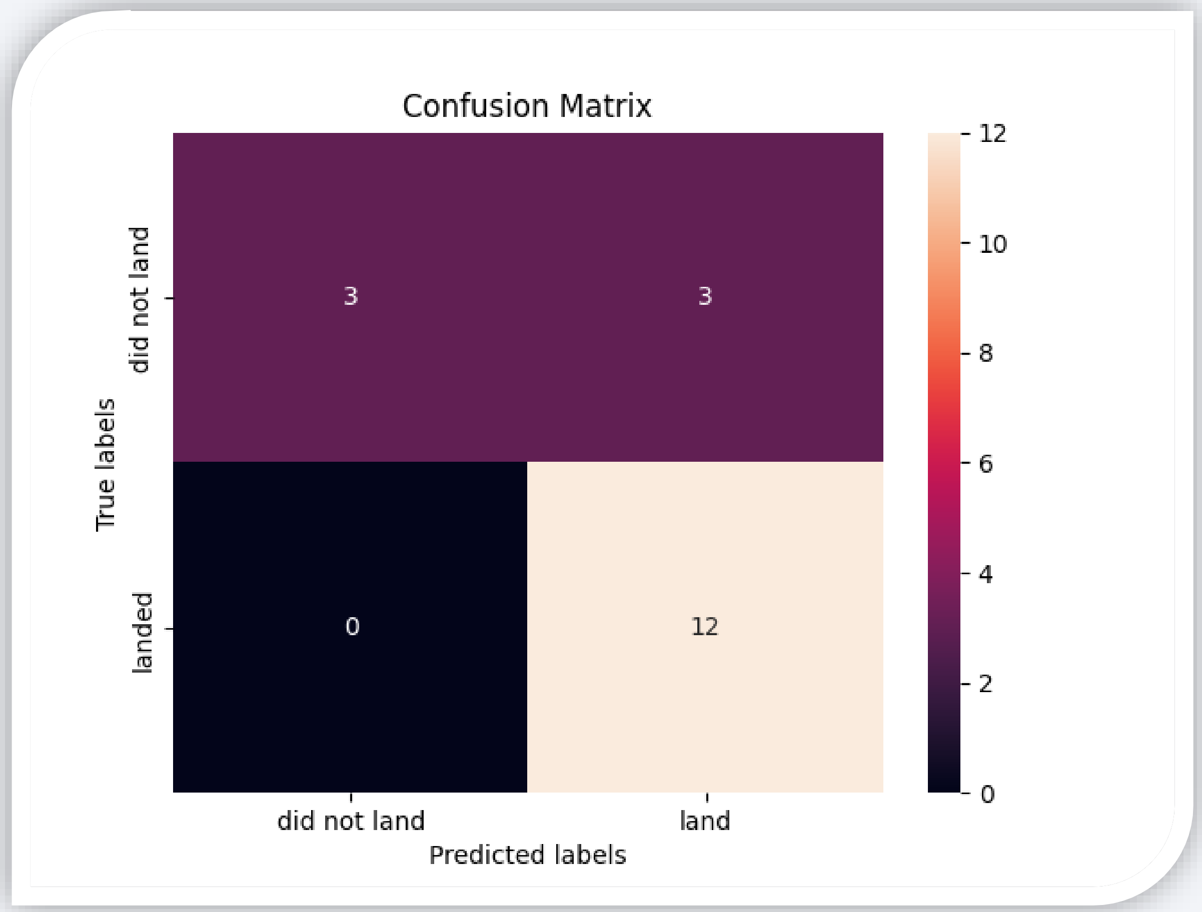
Classification Accuracy

- The built model accuracy for all built classification models

- **Decision Tree** has the best accuracy score.

Confusion Matrix for Decision Tree Model

- This Confusion Matrix shows that with the “DID NOT land” predicted results were correct 50% of the time. **Equal amount** of True Negative and False Negative
- For the positive result, ‘LANDED’, the model predicted 100% accurate with True Positive with ZERO False Positive.



Conclusions

- The launch site that has a highest success rate is KSC LC-39A
- The Booster Version that yield the highest success rate is FT
- The best Payload Range to achieve success launches is between 2000kg and 4000kg.
- The best Orbit types are ES-L1, GEO, HEO and SSO
- Most of the Launch Sites are quite far out to the City, closer to the Coastline, Railway and Highway for easy transportation, void of population, etc.
- The best Landing option is via Drone Ship
- The best ML model for predicting the result of a launch is Decision Tree Model

Appendix

spacex_launch_dash.csv	SpaceX Launch Data in CSV
spacex_launch_geo.csv	SpaceX Geo Data in CSV
API Data Collection.ipynb	Data Collection via API in Jupyter Notebook
Data Collection with Web Scraping.ipynb	Data Collection via Web Scrapping in Jupyter Notebook
Data Wrangling.ipynb	Data Cleanup in Jupyter Notebook
EDA with SQL.ipynb	EDA with SQL in Jupyter Notebook
EDA with Visualization.ipynb	EDA with Visualization in Jupyter Notebook
EDA.ipynb	EDA in Jupyter Notebook
Interactive Visual Analytics.ipynb	
SpaceX ML Prediction.ipynb	ML Model in Jupyter Notebook
ds-capstone-template-coursera	Presentation Deck
spacex_dash_app.py	Python App for Dash

Thank you!