IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Erika Goto
05 May 2022

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection using API Lab and Web Scraping;
  - Exploratory Data Analysis (EDA), including SQL, data wrangling, data visualization, and interactive visual analytics with Folium and Plotly Dash; and
  - Predictive analysis using classification models.

- Summary of all results
  - Based on the results of the analyses in previous presentations, we can conclude that the best launch site is KSC LC-39A.
  - Booster version FT will provide a higher success rate.
  - Over time, space launches show positive / increasing trend, which could be based on technological advancements. Given that, we can expect companies to provide cost efficient and cost effective product offers.
  - Decision Tree Classifier is the best method to use to predict success landings that we can use to predict how much Space Y could charge its customers and the number of occurrences and probability that SpaceX will reuse the first stage.

Introduction

Background:

When we talk about Space Travel, SpaceX immediately comes to mind, with it being the best among its competitors. One reason SpaceX beats its competitors is because their rocket launches are relatively inexpensive. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.

Now, there's a new company who would like to compete with SpaceX. It is called **Space Y** and was founded by a Billionaire industrialist named **Allon Musk**.

Problem:

The objective of this project is to determine the price of each launch that would compete with SpaceX by gathering information about SpaceX and using the answers to the following questions for our analysis:

- **How much will be the price of each launch using the data for SpaceX?**
- **Will SpaceX reuse the first stage?**

Section 1

Methodology

Methodology

- **Data collection methodology:**

  Data about SpaceX was obtained through the following:

  1. SpaceX REST API (https://api.spacexdata.com/v4/rockets/); and by
  2. Web Scraping (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)

- **Perform data wrangling:**

  - after obtaining all the data from SpaceX REST API & Web Scraping Wikipedia, data were cleaned up and missing data were replaced by mean of the available values.

  - Exploratory Data Analysis (EDA) was performed to find some patterns in the data and determine what would be labelled for training supervised models.

- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

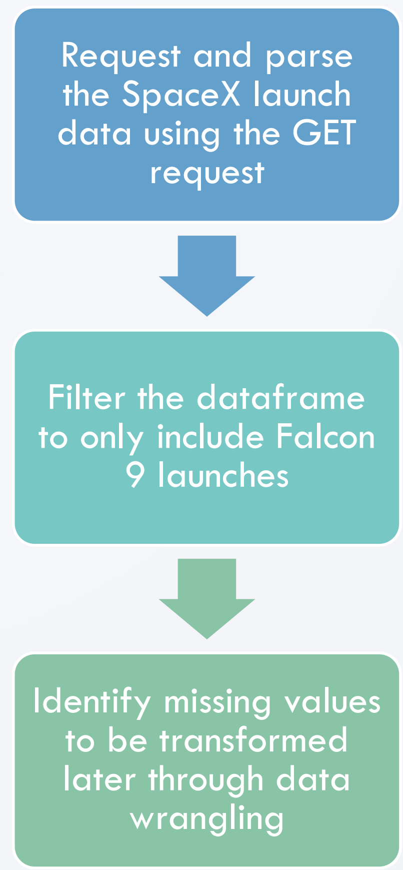
Data Collection and Data Wrangling

- The data collection process involved utilizing:
  - SpaceX REST API (https://api.spacexdata.com/v4/rockets/); and
  - Web Scraping (https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches)

Data Collection – SpaceX API

- SpaceX REST API is an open source REST API for launch, rocket, core, capsule, starlink, launchpad, and landing pad data.
- Data collection steps were performed chronologically as stated in the flowchart at the right → → → → →

- source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W101_Data%20Collection%20API%20Lab.ipynb



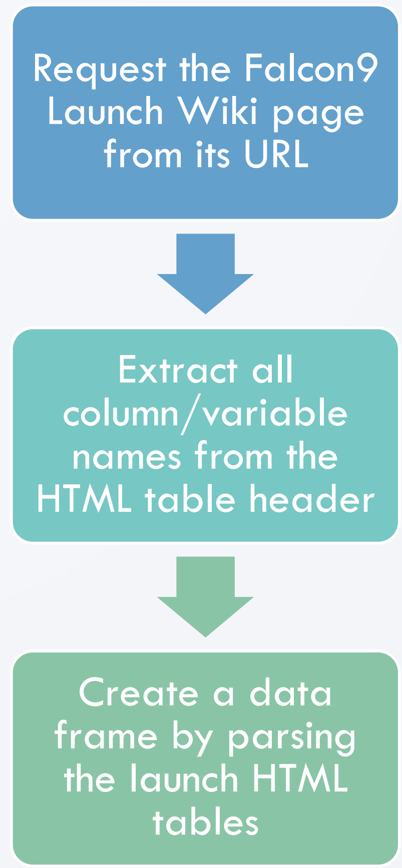
Data Collection – Web Scraping

- Data was also collected by Web Scraping Falcon 9 historical launch records from Wikipedia page titled" List of Falcon 9 and Falcon Heavy Launches" from the following link:

https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

- Data collection steps were performed chronologically as stated in the flowchart at the right → → → → →

source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W102_Data%20Collection%20with%20Web%20Scraping.ipynb



Data Wrangling

- In this process, Exploratory Data Analysis (EDA) was performed to find some patterns in the data and determine what would be labelled for training supervised models.
- Data wrangling steps performed are as follows:

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

source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W103_Data%20Wrangling.ipynb

EDA with Data Visualization

- Scatter Plot, Bar Graph, and Line Graph were used in this process:

  - ➢Scatter Plot to effectively visualize the relationship and correlations between:

• Flight Number and Launch Site	• Payload and Launch Site
• Flight Number and Orbit Type	• Payload and Orbit Type

  - ➢Bar Graph to visualize the relationship between success rate of each orbit type
  - ➢Line Graph to visualize the yearly trend for successful launches

- source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W202_EDA%20with%20Data%20Visualization.ipynb

EDA with SQL

- To further analyze the data, the following steps were performed in SQL:
    - ➢ Connect to the database
    - ➢ 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 booster launched by NAS (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 payload mass greater than 4000 but less than 6000
    - ➢ List the total number of successful and failed mission outcomes
    - ➢ List the names of the booster_versions which have carried the maximum payload mass using a subquery
    - ➢ List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
    - ➢ Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-05-04 and 2017-03-20, in descending order

- source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W201_EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Folium markers, circles, lines, and marker clusters were used in this process:

  - ➢Markers were used to indicate successful or failed launches for each site on the map to easily visualize different outcomes of launches;
  - ➢To visualize the distances between 2 launch sites, polylines were drawn;
  - ➢To highlight specific coordinates, Circles were used; and
  - ➢Marker clusters were used to show groups of events that happened in each coordinate.

source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W301_Interactive%20Visual%20Analytics%20with%20Folium.ipynb

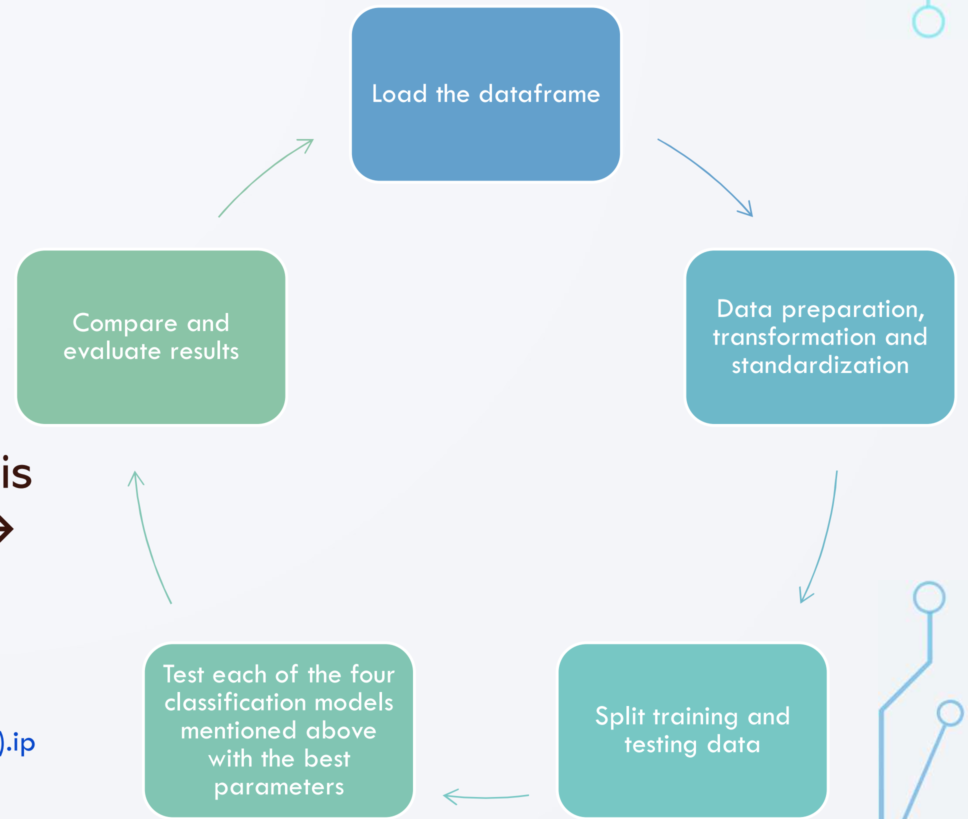
Build a Dashboard with Plotly Dash

- The plots/graphs and interactions have been added to the dashboard:
    - ➢Pie Chart with site-dropdown menu to provide interactive visualization of Success Launches by Site
    - ➢Scatter Plot with payload-slider to provide interactive visualization of the relationship between class of launches (successful or failed) and Payload Mass

- source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/spacex_dash_app.py
- Reports:
    - https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/Dash%2001_Total%20Success%20Launches%20by%20Site.png
    - https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/Dash%2002_Payload%20Mass.png

Predictive Analysis (Classification)

- Four classification models were utilized, tested, and evaluated:
  - ➢Logistic Regression
  - ➢Support Vector Machine
  - ➢Decision Tree
  - ➢K-Nearest Neighbors (KNN)
- The model development process performed is as shown in the cycle at the right → → → →



source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W401_Predictive%20Analysis%20(Classification).ipynb

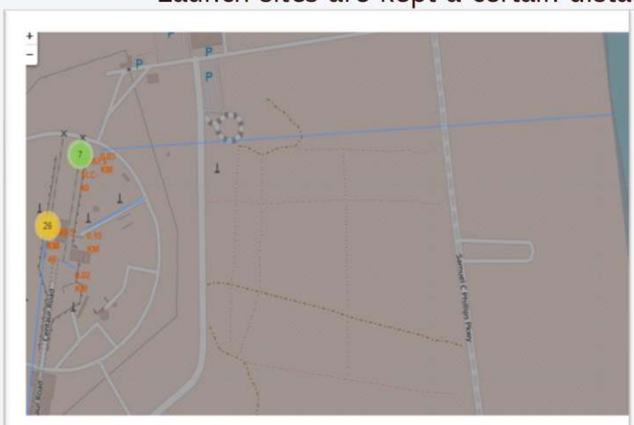
Results

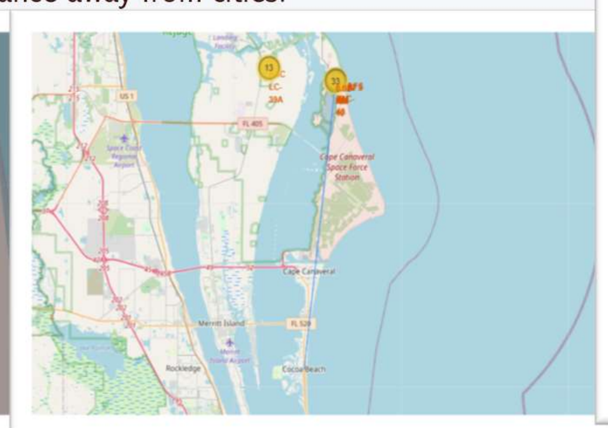
- Exploratory data analysis results:
  - There were 4 unique launch sites in the space mission: CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, and VAFB SLC-4E.
  - The total payload mass carried by boosters launched by NASA is 111,268 kgs.
  - The average payload mass carried by booster version F9 V1.1 is 2,928 kgs.
  - The first successful landing in gound pad happened on Dec. 22, 2015.
  - There were 4 F9 booster versions with payload mass above average which successfully landed on drone ships.
  - 99% of missions were successful.
  - F9 B5* boosters have carried the maximum payload mass.
  - Only CCAFS LC 40's F9 v1.1 B1012 and B1015 booster versions failed landing in 2015.
  - Landing outcomes showed a positive trend as time passes by.

source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W202_EDA%20with%20Data%20Visualization.ipynb

Results

- Interactive Analytics using Folium
  - Using Folium, launch sites, successful and failed launches on each site, and the distances between a launch site to its proximities were marked, circled, and drawn with polylines.
  - Based on the above, the following information were obtained:
    - Launch sites are in close proximity to railways, highways, and coastlines; but
    - Launch sites are kept a certain distance away from cities.

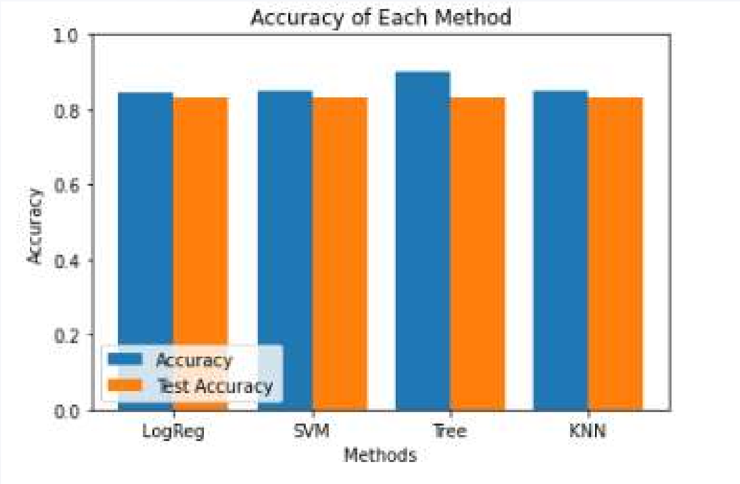




source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W301_Interactive%20Visual%20Analytics%20with%20Folium.ipynb

Results

Out of all the four methods tested, the Decision Tree Classifier is the best model to predict the successful landings, with accuracy of 90.179% and test data accuracy of 83.3333%.

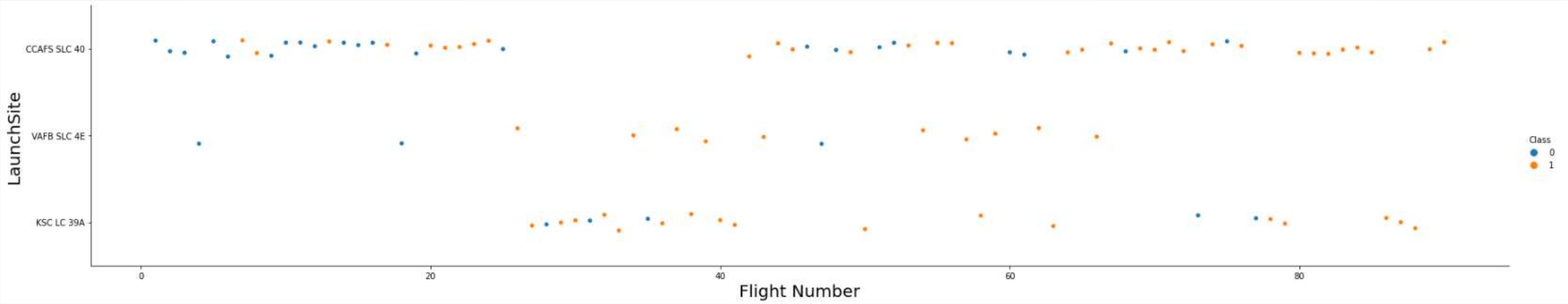


source code: source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W301_Interactive%20Visual%20Analytics%20with%20Folium.ipynb

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

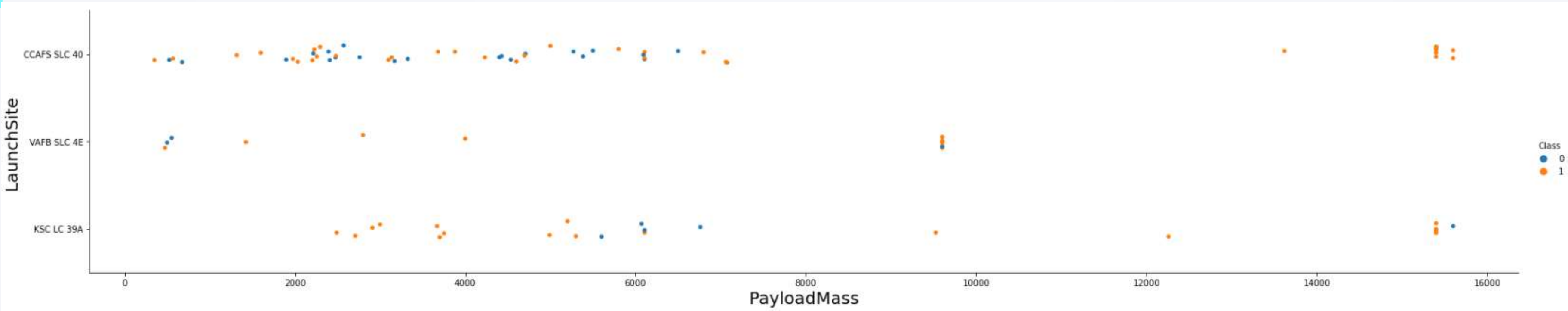


Based on the above plot:

- CCAFS SLC 40 has the most number of launches and most number of successful launches, with 100% success rate for flight numbers above 80.
- Despite fewer number of launches, VAFB SLC 4E has an average of 77% success rate. For flight numbers above 50, success rate is at a 100%.
- KSC LC 39A also has a success rate of 77% and flight numbers above 80 have success rate of 100%.

source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W202_EDA%20with%20Data%20Visualization.ipynb

Payload vs. Launch Site

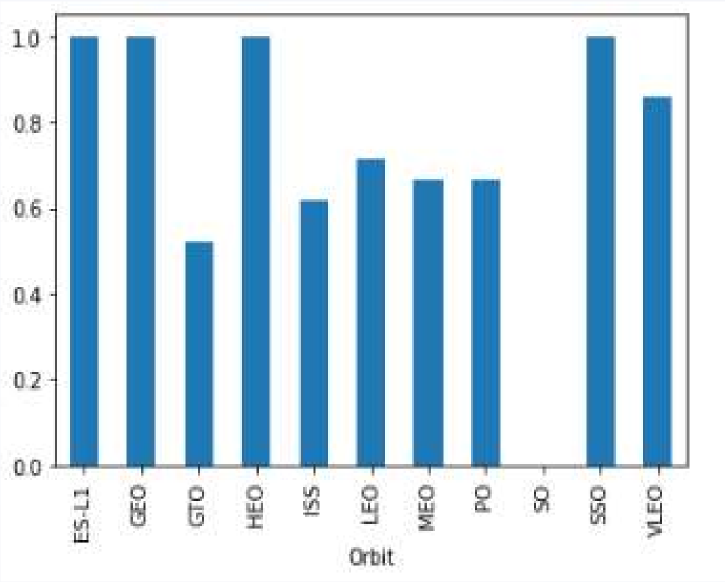


- Based on the above plot:
  - For VAFB SLC 4E, there are no rockets launched for heavy payload mass (greater than 10,000 kg).
  - Rockets with payload mass above 8,000 kgs. for CCAFS SLC 40 have success rate of 100%.
  - Rockets outside of the payload mass of 5000 to 7000 have approximately 99% success rate.

source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W202_EDA%20with%20Data%20Visualization.ipynb

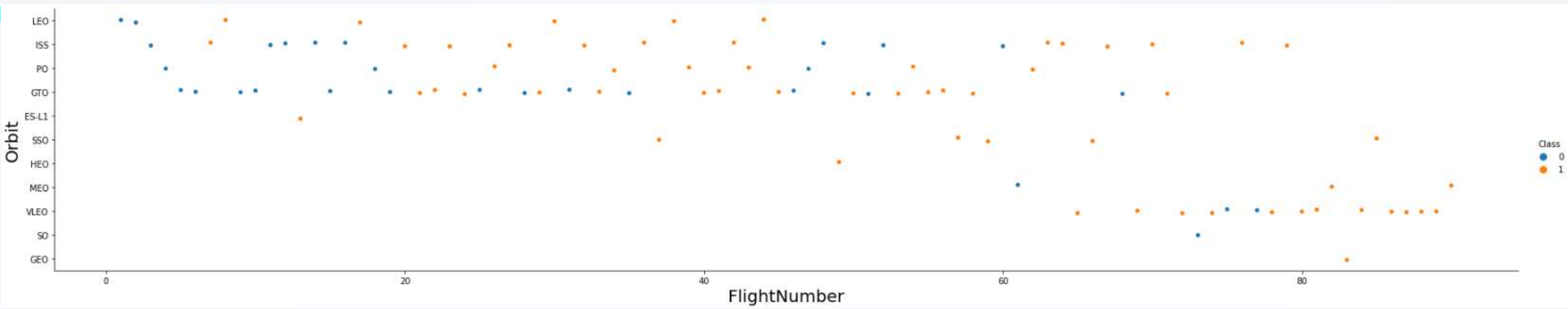
Success Rate vs. Orbit Type

- Based on the bar chart at the right:
  - The following orbits show the highest success rates:
    - ➢ ES-L1;
    - ➢ GEO;
    - ➢ HEO; and
    - ➢ SSO
  - GTO shows the lowest success rate.



source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W202_EDA%20with%20Data%20Visualization.ipynb

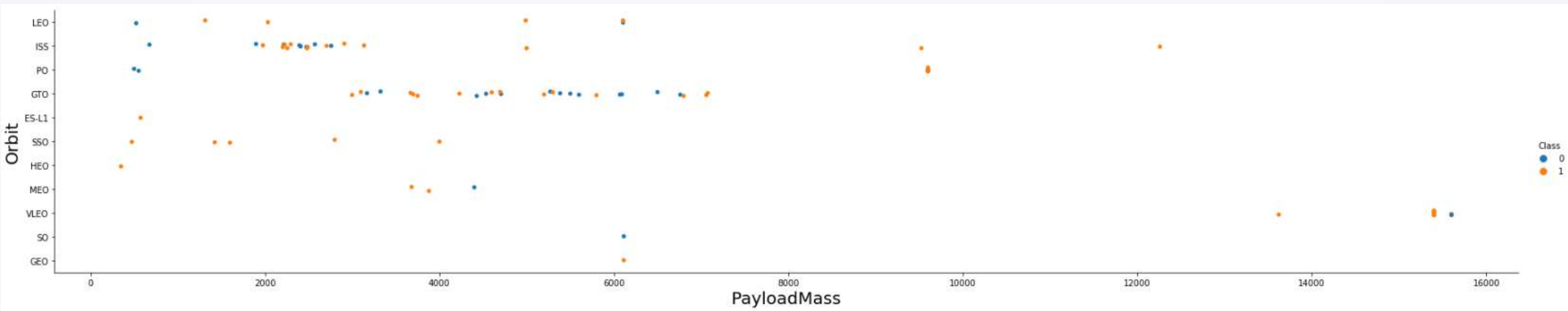
Flight Number vs. Orbit Type



- Based on the above plot:
  - ➢In 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.
  - ➢GEO has launched only 1 rocket and it was successful.

source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W202_EDA%20with%20Data%20Visualization.ipynb

Payload vs. Orbit Type

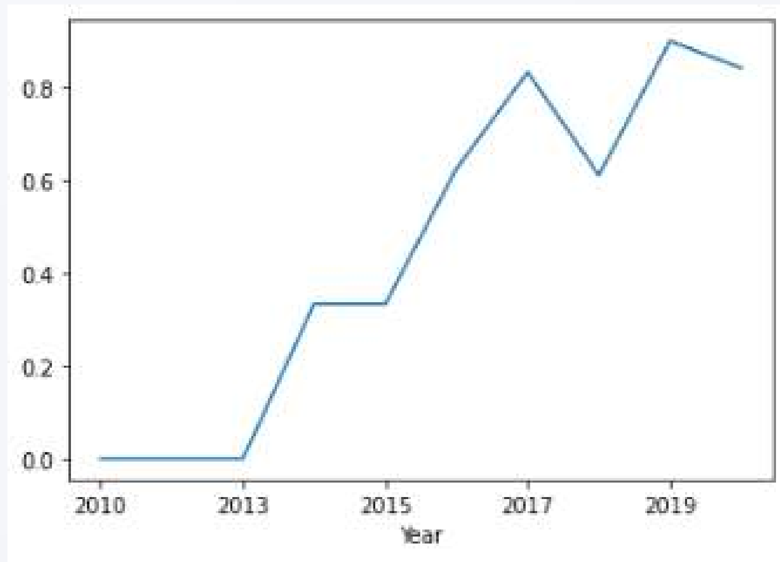


- Based on the above plot:
  - ➢With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS
  - ➢However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W202_EDA%20with%20Data%20Visualization.ipynb

Launch Success Yearly Trend

- Based on the line chart at the right:
  - the success rate since 2013 shows a positive trend that kept increasing until 2019.



source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W202_EDA%20with%20Data%20Visualization.ipynb

All Launch Site Names

- Names of the unique launch sites:
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
  - VAFB SLC-4E

```
In [4]: sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL ORDER BY 1;

 * ibm_db_sa://RZC31110:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
Done.

Out[4]:
```

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

source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W201_EDA%20with%20SQL.ipynb

Launch Site Names Begin with 'CCA'

```
In [5]: sql SELECT  * FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;

 * ibm_db_sa://RZC31110:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
Done.
```

Out[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

- Based on the first 5 items shown above, launch sites with 'CCA' have 100% mission outcome success rate but 0% landing outcome.

source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W201_EDA%20with%20SQL.ipynb

Total Payload Mass

- Total payload mass carried by boosters launched by NASA (CRS) is equal to 111,268 kgs:

```
In [7]: sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD FROM SPACEXTBL WHERE PAYLOAD LIKE '%CRS%' ;

 * ibm_db_sa://RZC31110:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
Done.

Out[7]:
```

total_payload
111268

- The total payload mass was calculated by summing all payloads whose payload column contain 'CRS'.

source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W201_EDA%20with%20SQL.ipynb

Average Payload Mass by F9 v1.1

- Average payload mass carried by booster version F9 v1.1 is equal to 2,928 kgs.:

```
In [8]: sql SELECT AVG(PAYLOAD_MASS__KG_) AS AVG_PAYLOAD FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1' ;
 * ibm_db_sa://RZC31110:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
Done.
Out[8]:
```

avg_payload
2928

- The above was calculated by getting the average of the payload mass of booster versions with 'F9 v1.1' only.

source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W201_EDA%20with%20SQL.ipynb

First Successful Ground Landing Date

- The first successful landing outcome on ground pad happened on December 22, 2015:

```
In [9]: sql SELECT MIN(DATE) AS FIRST_SUCCESSFUL_GP_LANDING FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Success (ground pad)' ;

 * ibm_db_sa://RZC31110:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
Done.

Out[9]:
```

first_successful_gp_landing
2015-12-22

- Successful landing outcome on ground pad was calculated by getting the minimum date from the table.

source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W201_EDA%20with%20SQL.ipynb

Successful Drone Ship Landing with Payload between 4000 and 6000

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

```
In [10]: sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTBL WHERE (PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000) AND (LANDING__OUTCOME = 'Succ
ess (drone ship)') ;

 * ibm_db_sa://RZC31110:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
Done.

Out[10]:
```

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

- The above resulted from filtering the unique booster version with successful (drone ship) landing outcome.

source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W201_EDA%20with%20SQL.ipynb

Total Number of Successful and Failure Mission Outcomes

- The total number of successful and failure mission outcomes is as follows:

```
In [11]: sql SELECT MISSION_OUTCOME, COUNT(*) AS QTY FROM SPACEXTBL GROUP BY MISSION_OUTCOME ORDER BY MISSION_OUTCOME;

 * ibm_db_sa://RZC31110:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
Done.

Out[11]:
```

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

- Approximately 99% of the missions had successful mission outcomes.

source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W201_EDA%20with%20SQL.ipynb

Boosters Carried Maximum Payload

- The following contains the names of the boosters which have carried the maximum payload mass:

In [12]:
```
sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL) ORDE
R BY BOOSTER_VERSION;
```

```
 * ibm_db_sa://RZC31110:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
Done.
```

Out[12]:

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

source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W201_EDA%20with%20SQL.ipynb

2015 Launch Records

- There were only two failed landing outcomes in drone ship in 2015.
- Its booster versions, and launch site names are as follows:

In [13]:
```
sql SELECT BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTBL WHERE LANDING__OUTCOME = 'Failure (drone ship)' AND DATE_PART('YEAR', DAT
E) = 2015;
```

```
 * ibm_db_sa://RZC31110:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
Done.
```

Out[13]:

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

- source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W201_EDA%20with%20SQL.ipynb

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

- The rank of the count of each landing outcome (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, is shown in descending order as follows:

```
In [14]: sql SELECT LANDING__OUTCOME, COUNT(*) AS QTY FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY LANDING__O
UTCOME ORDER BY QTY DESC;

 * ibm_db_sa://RZC31110:***@ba99a9e6-d59e-4883-8fc0-d6a8c9f7a08f.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:31321/BLUDB
Done.

Out[14]:
```

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

- This report showed us that "no attempt" landing outcome should be considered as part of the analysis.
- source code: https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W201_EDA%20with%20SQL.ipynb

Section 3

Launch Sites Proximities Analysis

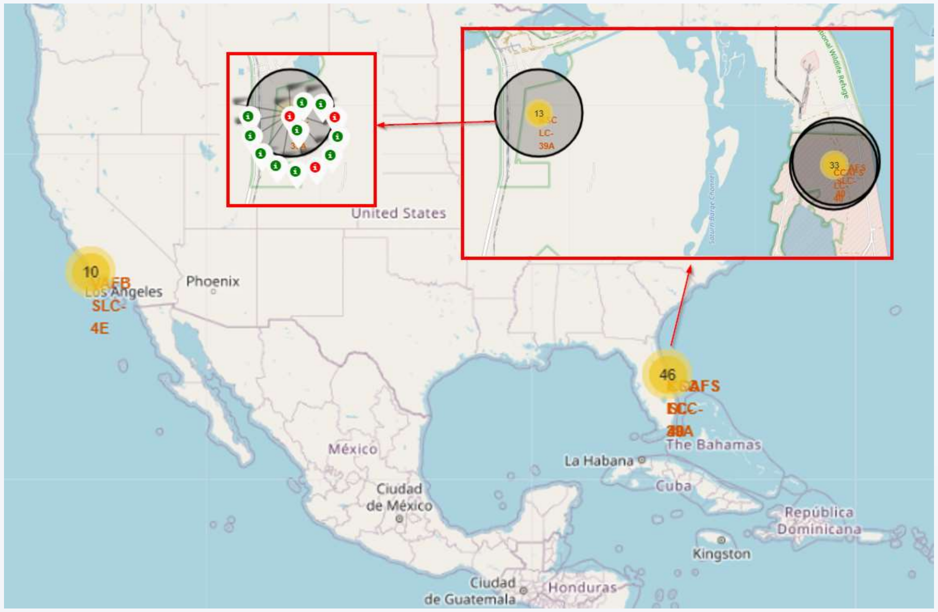
Launch Sites – Global View

- Global view of launch sites with orange labels:

- *By looking at the map, it can be concluded that the launch sites are placed near coastlines.*

Sample Launch Outcomes by Site

- Sample launch outcomes for KSC LC-39A:

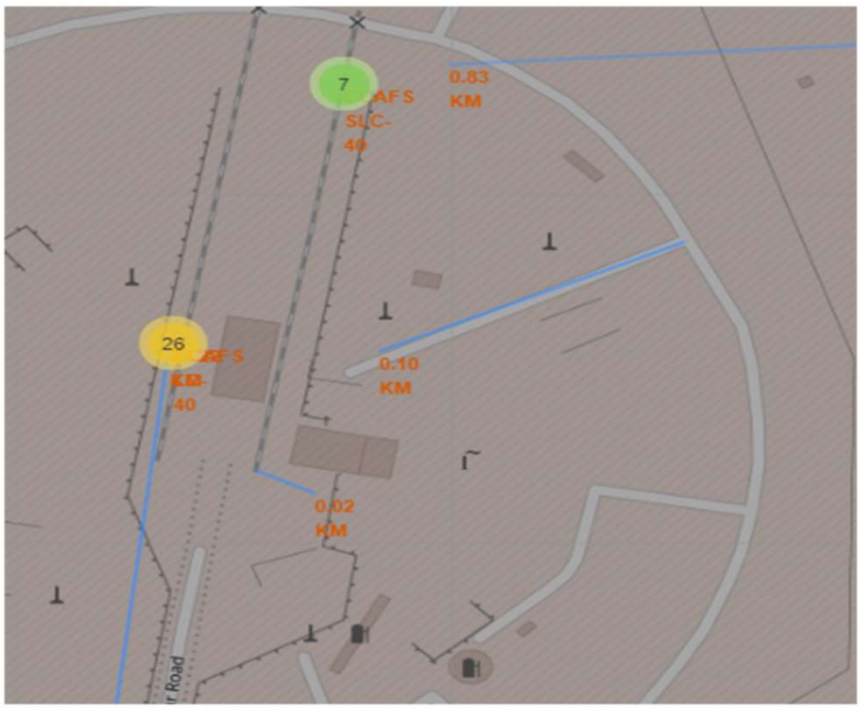


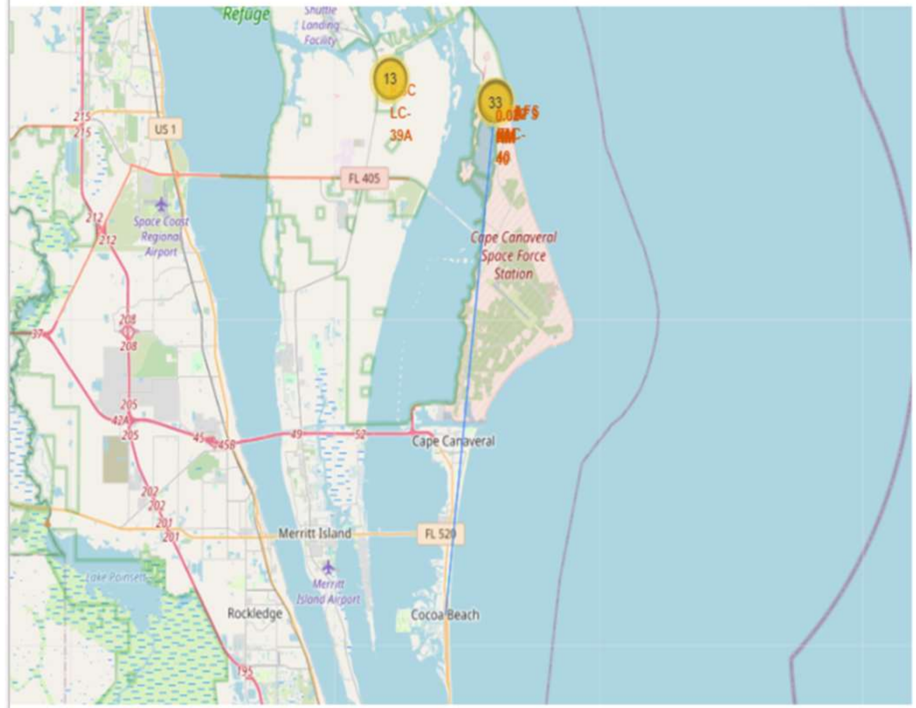
Green markers means successful launch

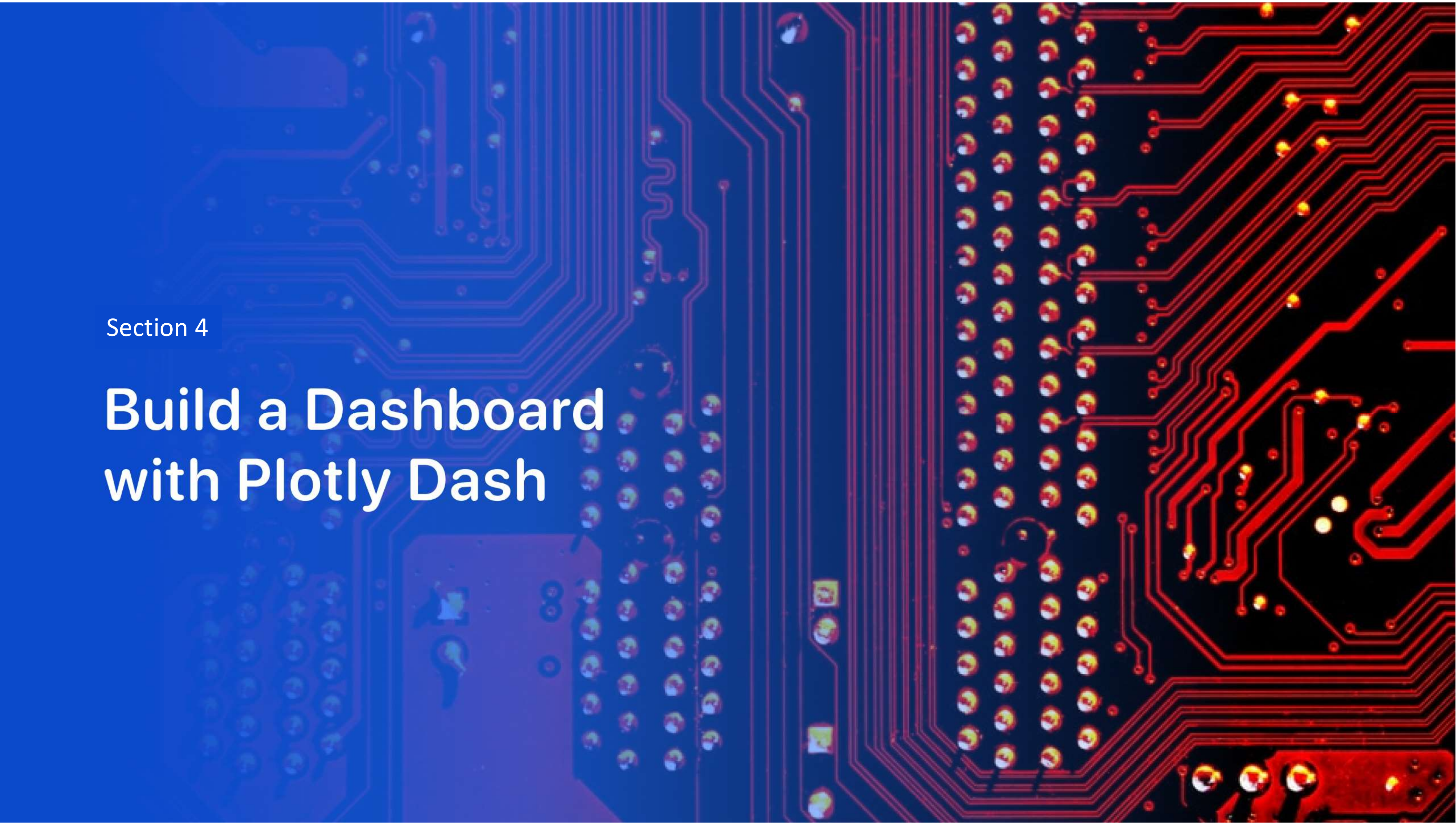
Red markers mean failed launches

Launch site proximities

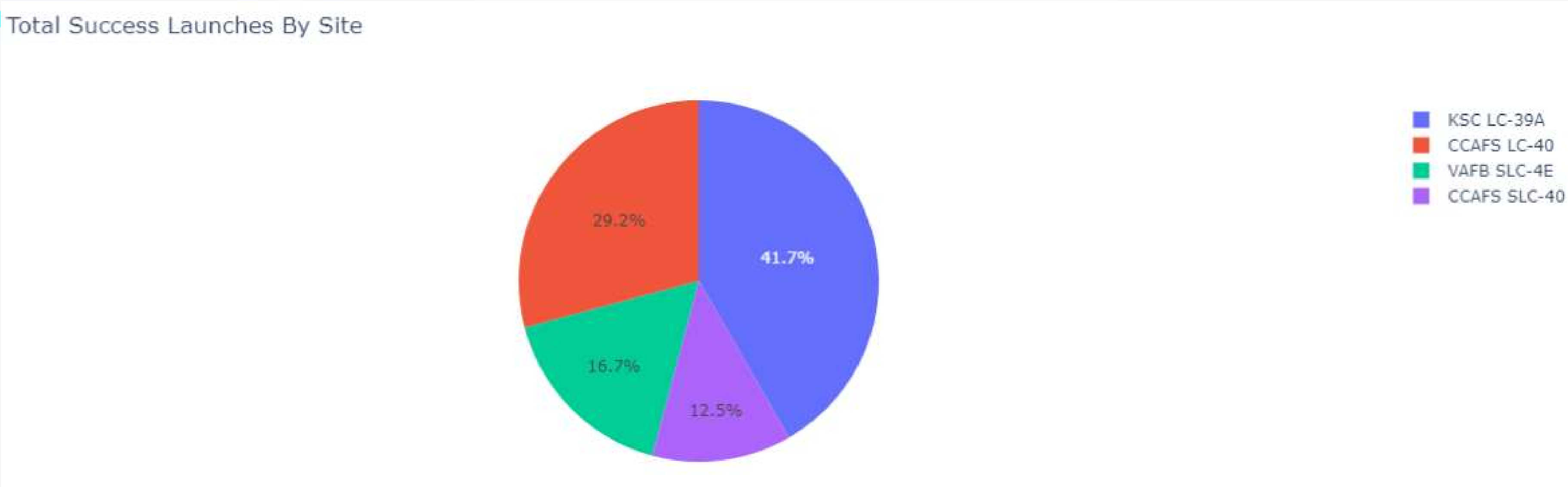
- Launch sites are situated near coastlines and are accessible through railways and highways. It is also relatively safe since it is far from inhabited cities.






Section 4
Build a Dashboard
with Plotly Dash

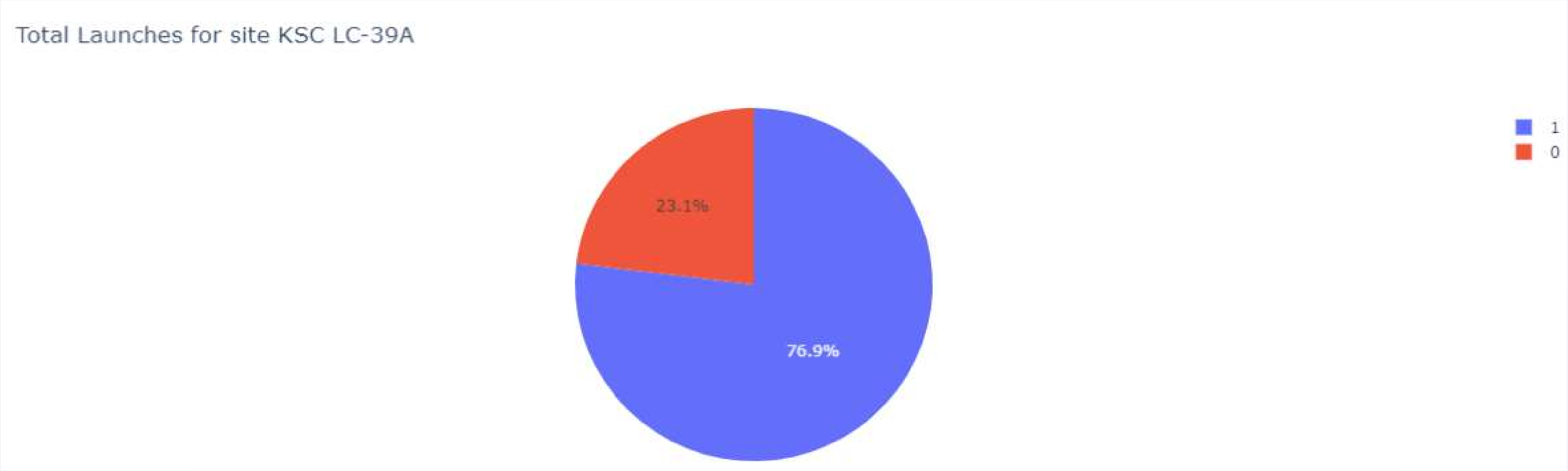
Successful Launches by Site



- *The location and proximities of the sites from railways, highways, coastlines, and cities seem to have a direct impact on the outcome of the launches.*
- *KSC LC 39A had the most successful launches out of the four sites.*

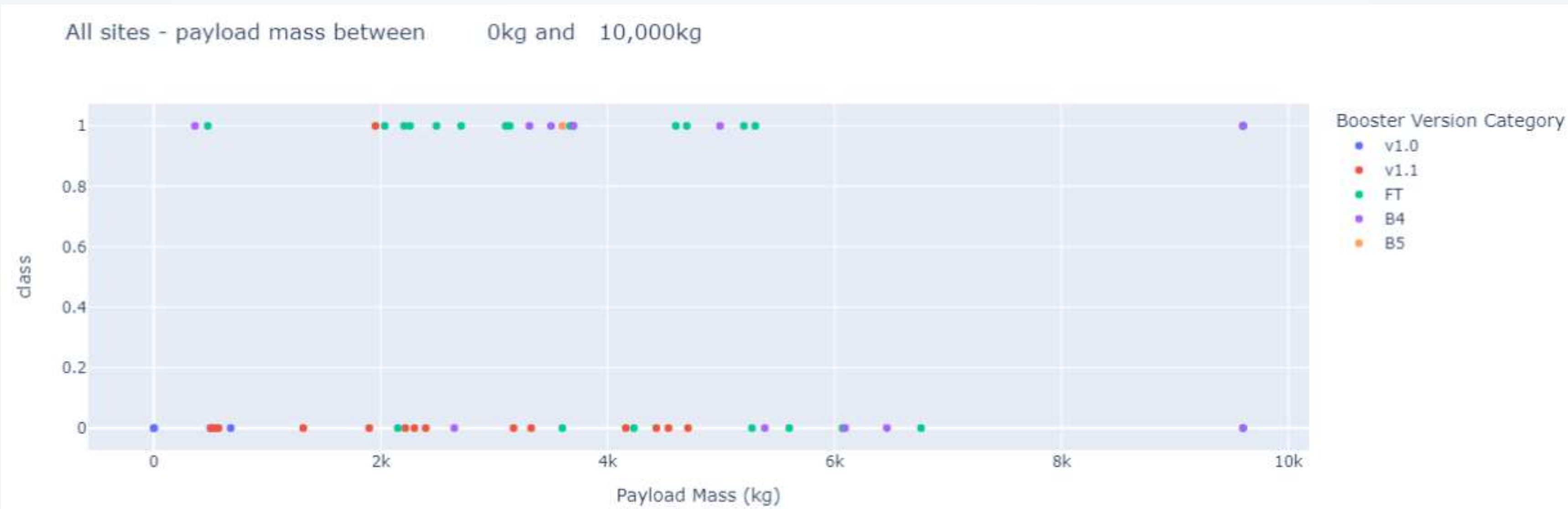
Site with Highest Launch Success Rate

- KSC LC 39A also had the highest launch success rate. 76.9% (10 successful launches) out of its 13 launches was successfully landed:



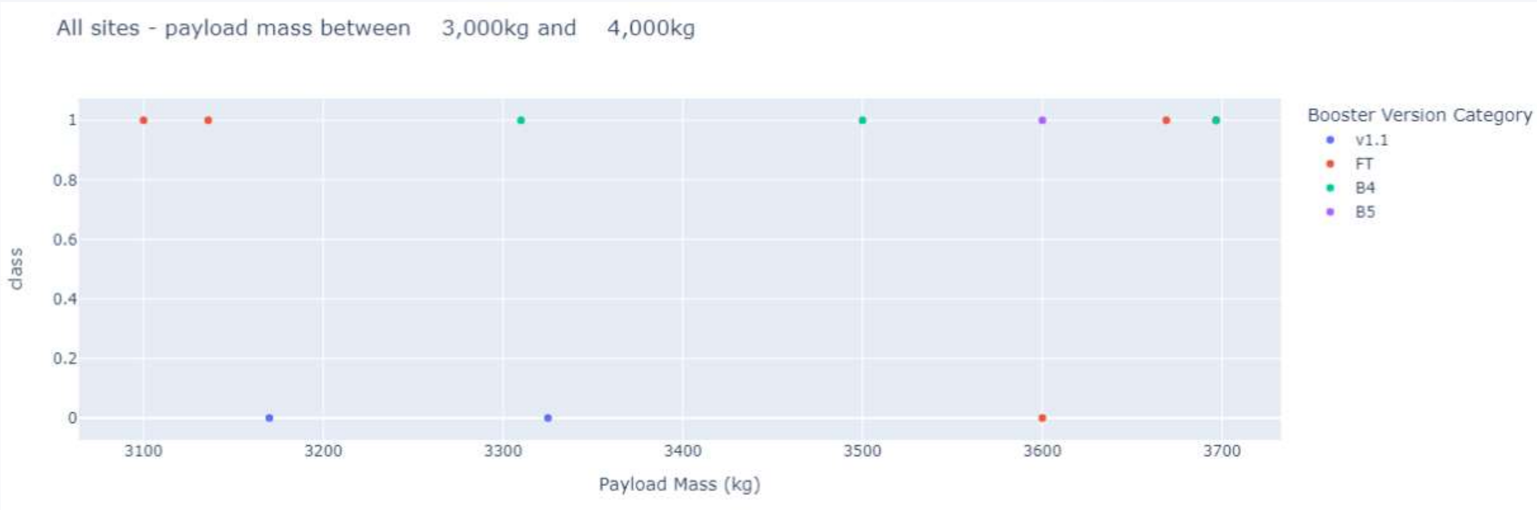
Launch Outcomes based on Payload Mass

- Booster versions with Payload Mass from 2000 to 6000 showed more successful launches:



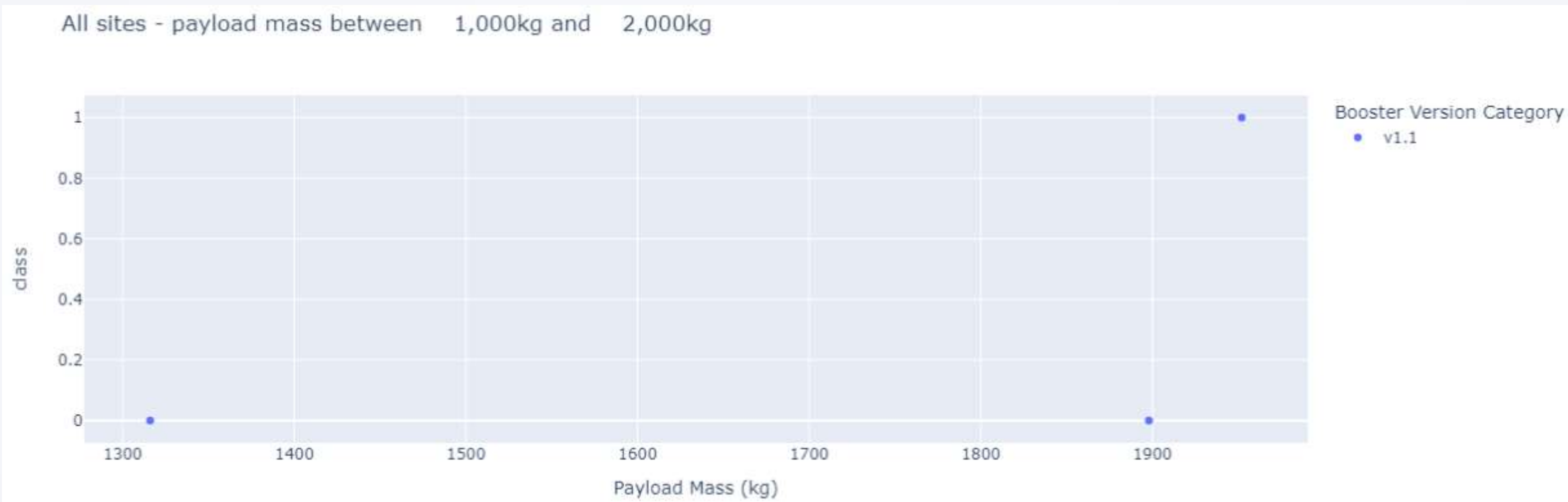
Launch Outcomes based on Payload Mass

- Booster versions with Payload Mass from 3000 to 4000 showed the highest launch success rate:



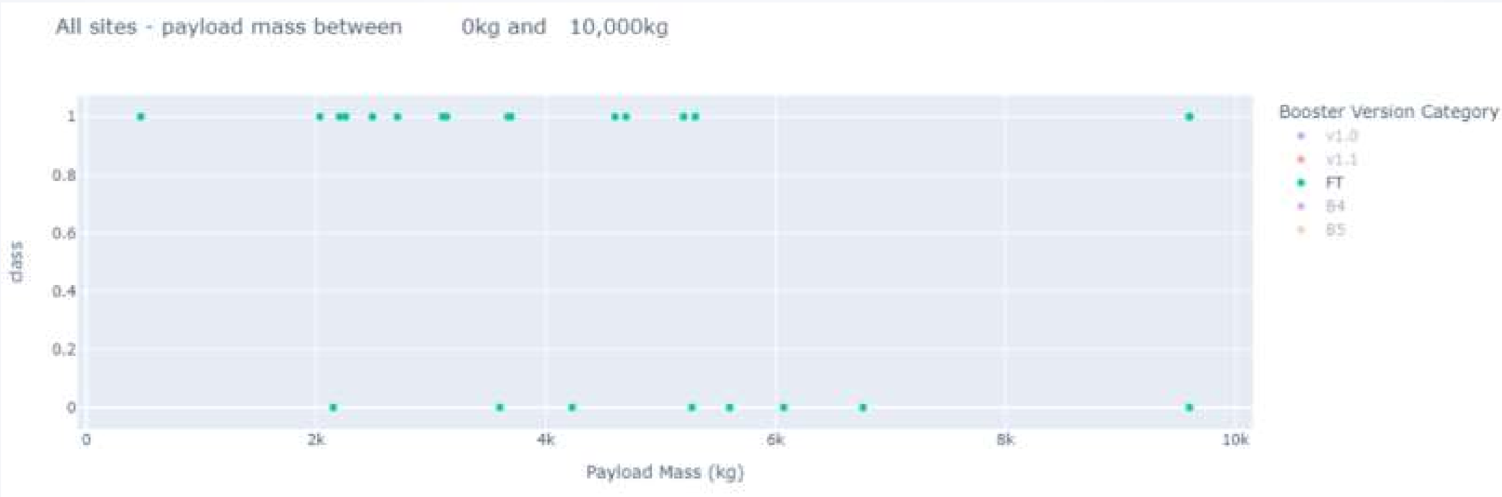
Launch Outcomes based on Payload Mass

- Booster version v1.1 with Payload Mass from 1000 to 2000 showed the lowest launch success rate:



Launch Outcomes based on Payload Mass

- Booster version FT showed the highest launch success rate:

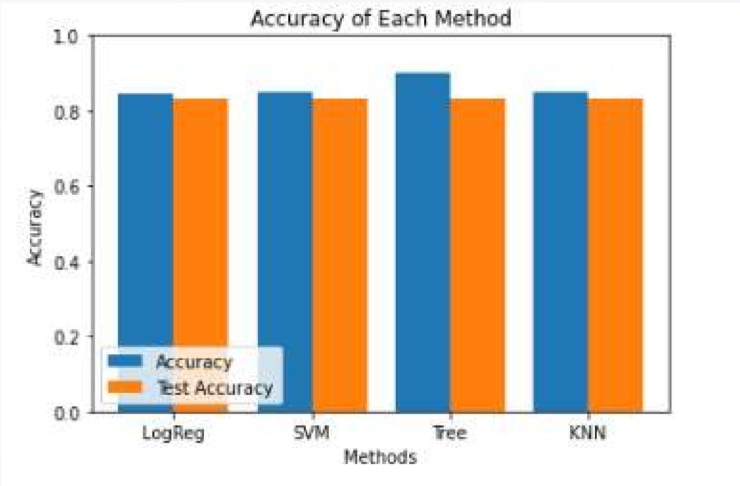


Section 5

Predictive Analysis (Classification)

Classification Accuracy

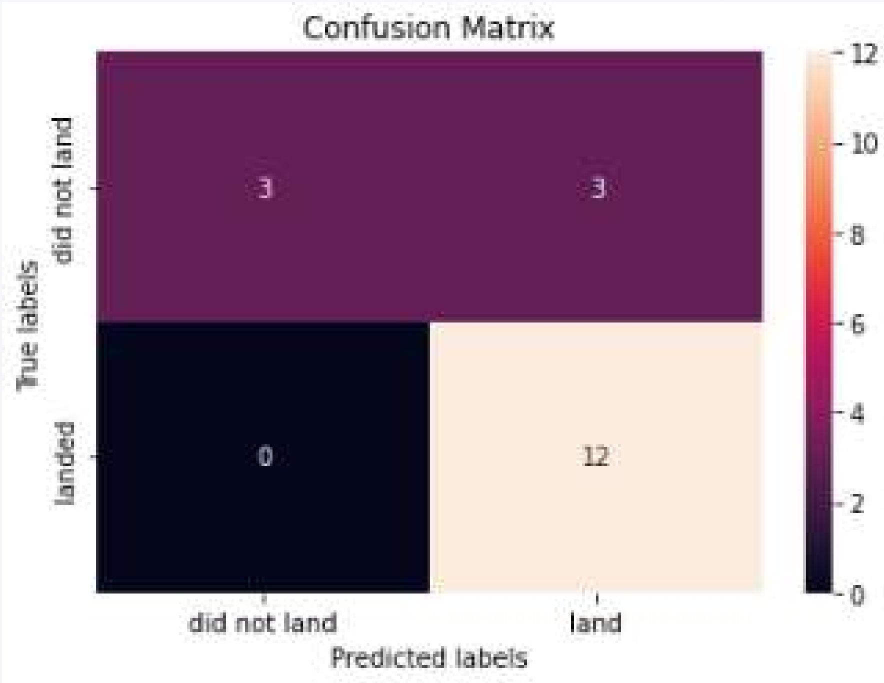
Out of all the four methods tested, the Decision Tree Classifier is the best model to predict the successful landings, with accuracy of 90.179% and test data accuracy of 83.3333%.



source code: source code:
https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE/blob/main/10_W301_Interactive%20Visual%20Analytics%20with%20Folium.ipynb

Confusion Matrix

- Confusion matrix of Decision Tree Classifier showed the same ratio of true positives and true negatives with the other three models. But Decision Tree Classifier became the ultimate / best method due to its high accuracy (90.18%) based on best / tuned hyperparameters:



Conclusions

- Based on the results of the analyses in previous presentations, we can conclude that the best launch site is KSC LC-39A.
- Booster version FT will provide a higher success rate.
- Over time, space launches show positive / increasing trend, which could be based on technological advancements. Given that, we can expect companies to provide cost efficient and cost effective product offers.
- Decision Tree Classifier is the best method to use to predict success landings that we can use to predict how much Space Y could charge its customers and the number of occurrences and probability that SpaceX will reuse the first stage.

Appendix

- For more detailed and organized grouping of the notebooks, please refer to:

https://github.com/erikabgoto14/EBG_IBM_APPLIED-DATA-SCIENCE-CAPSTONE

- If Folium maps failed to show in the notebook, please refer to the folium map screenshots file.

- For Plotly Dashboard results, please refer to 'Dash*' screenshots in github link provided above.


Thank you!