IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Hikari Nakamura
2023/05/29

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection using web scraping and SpaceX API
  - Exploratory Data Analysis (EDA), including data wrangling, data visualization and interactive visual analytics
  - Machine Learning Prediction.
- Summary of all results
  - It was possible to collected valuable data from public sources
  - EDA allowed to identify which features are the best to predict success of launchings
  - Machine Learning Prediction showed the best model to predict which characteristics are important to drive this opportunity by the best way, using all collected data

Introduction

- Project background and context
  - The objective is to evaluate the viability of the new company Space Y to compete with Space X.
- Problems you want to find answers
  - The best way to estimate the total cost for launches, by predicting successful landings of the first stage of rockets
  - Where is the best place to make launches.

Section
1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data from Space X was obtained from 2 sources:
    - Space X API (https://api.spacexdata.com/v4/rockets/)
    - WebScraping(https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)
- Perform data wrangling
  - Collected data was enriched by creating a landing outcome label based on outcome data after summarizing and analyzing features

Methodology

Executive Summary

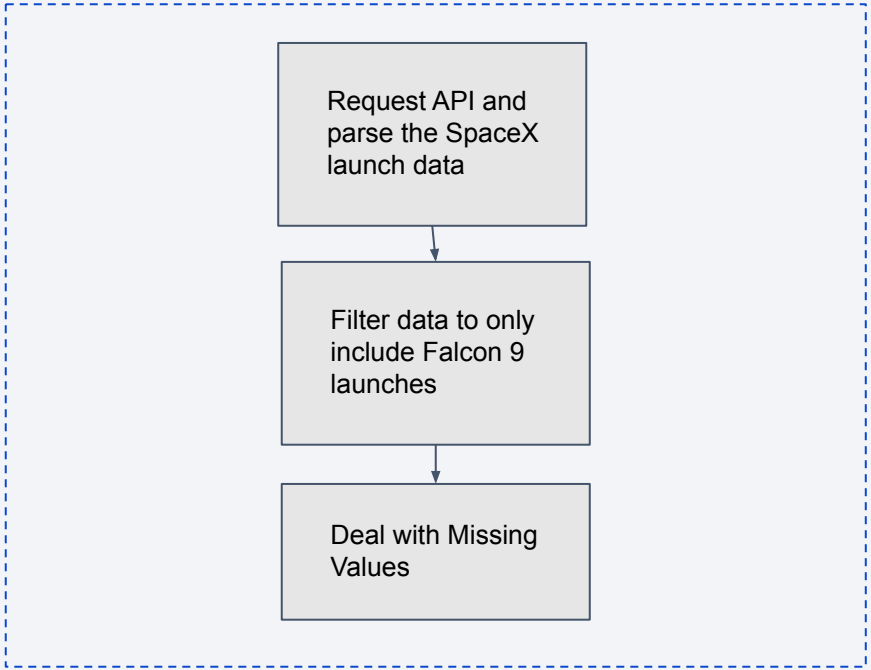
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - Data that was collected until this step were normalized, divided in training and test data sets and evaluated by four different classification models, being the accuracy of each model evaluated using different combinations of parameters.

Data Collection

- Data sets were collected from Space X API (https://api.spacexdata.com/v4/rockets/) and from Wikipedia (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches) , using web scraping technics.

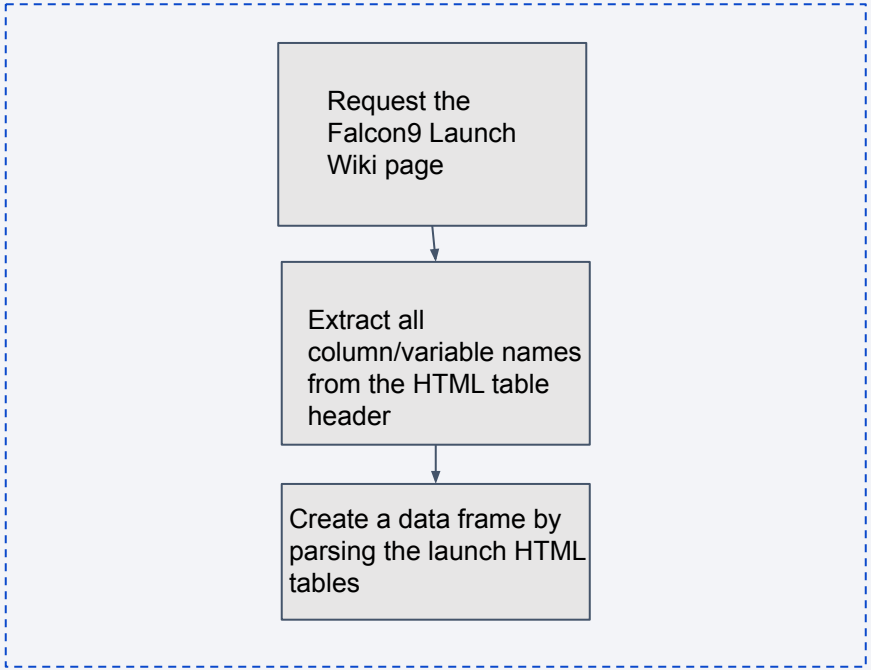
Data Collection – SpaceX API

- SpaceX offers a public API from where data can be obtained and then used
- This API was used according to the flowchart beside and then data is persisted.



Data Collection - Scraping

- Data from SpaceX launches can also be obtained from Wikipedia;
- Data are downloaded from Wikipedia according to the flowchart and then persisted.

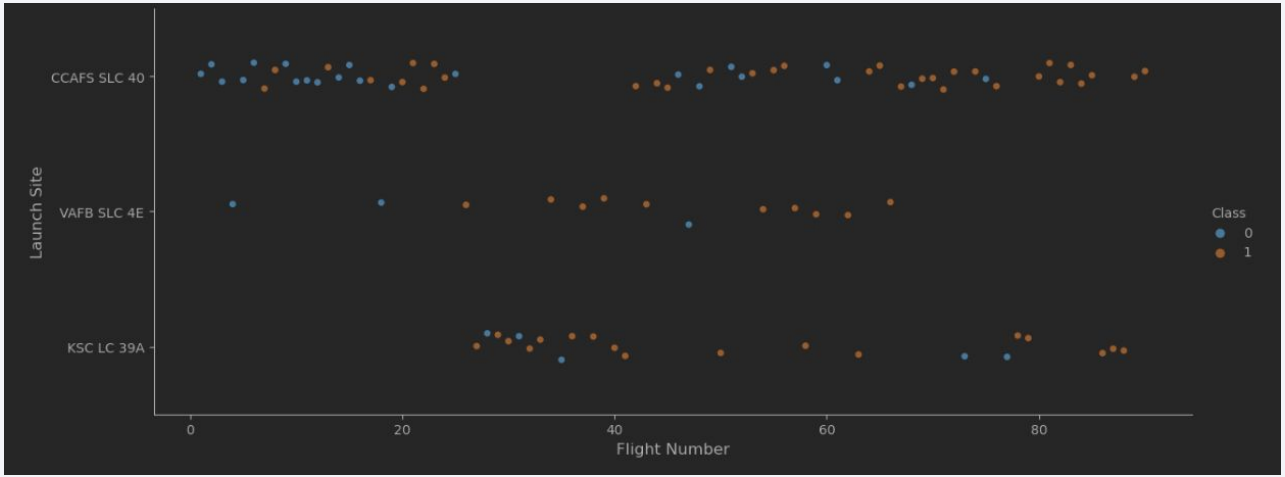


Data Wrangling

- Initially some Exploratory Data Analysis (EDA) was performed on the dataset.
- Then the summaries launches per site, occurrences of each orbit and occurrences of mission outcome per orbit type were calculated.
- The landing outcome label was created from Outcome column.

EDA with Data Visualization

- To explore data, scatterplots and barplots were used to visualize the relationship between pair of features



EDA with SQL

The following SQL queries were performed:

- Top 5 launch sites whose name begin with the string 'CCA’
- Total payload mass carried by boosters launched by NASA (CRS)
- Average payload mass carried by booster version F9 v1.1
- •Datewhenthefirstsuccessfullandingoutcomeingroundpadwasachieve
- Names of the boosters which have success in drone ship and have payload mass between 4000 and 6000 kg
- Total number of successful and failure mission outcomes;
- Names of the booster versions which have carried the maximum payload mass;
- Failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015; and

Rank of the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20.

Build an Interactive Map with Folium

Markers indicate points like launch sites

- Circles indicate highlighted areas around specific coordinates, like NASA Johnson Space Center
- Marker clusters indicates groups of events in each coordinate, like launches in a launch site and
- Lines are used to indicate distances between two coordinates.

Build a Dashboard with Plotly Dash

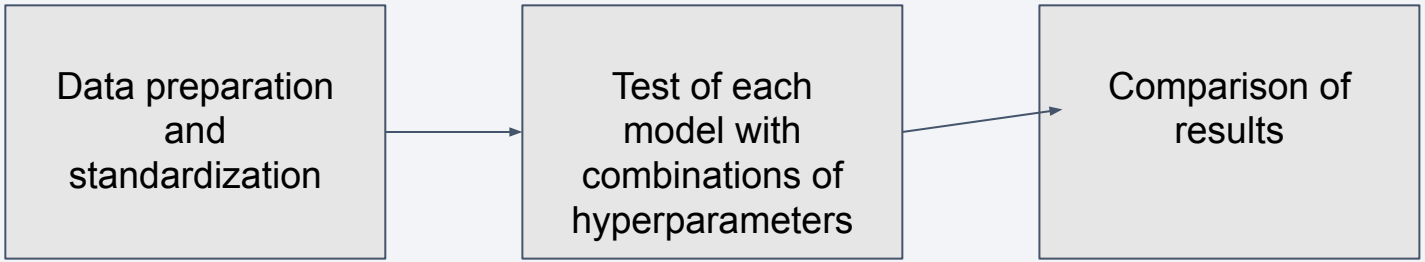
The following graphs and plots were used to visualize data • Percentage of launches by site

• Payloadrange

• This combination allowed to quickly analyze the relation between payloads and launch sites, helping to identify where is best place to launch according to payloads.

Predictive Analysis (Classification)

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



Results

- Exploratory data analysis results

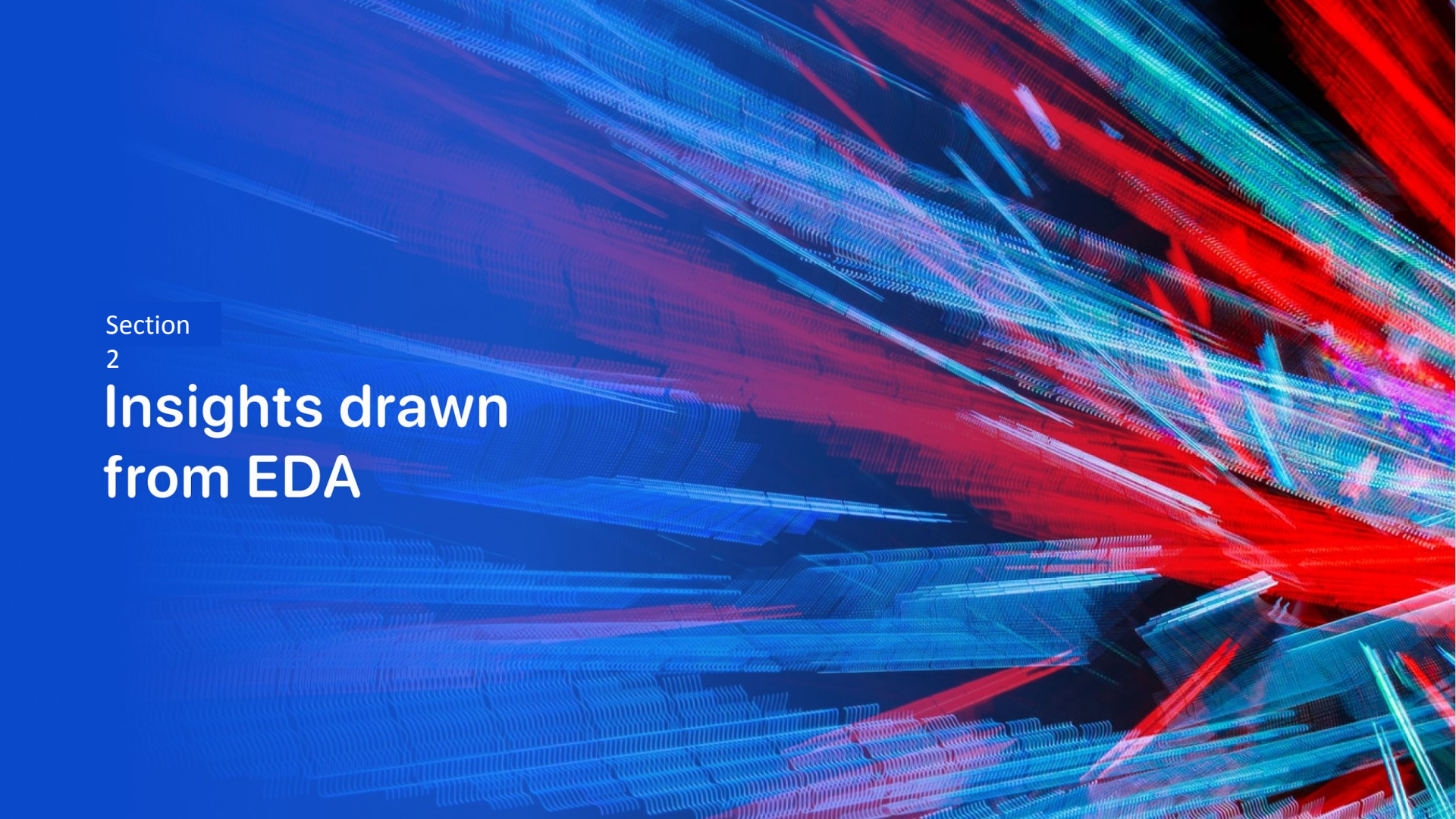
  Space X uses 4 different launch sites

  - The first launches were done to Space X itself and NASA
  - The average payload of F9 v1.1 booster is 2,928 kg
  - The first success landing outcome happened in 2015 fiver year after the first launch
  - Many Falcon 9 booster versions were successful at landing in drone ships having payload above the average
  - Almost 100% of mission outcomes were successful
  - Two booster versions failed at landing in drone ships in 2015: F9 v1.1 B1012 and F9 v1.1 B1015
  - The number of landing outcomes became as better as years passed.

Results

Using interactive analytics was possible to identify that launch sites use to be in safety places, near sea, for example and have a good logistic infrastructure around.

- Most launches happens at east cost launch sites.



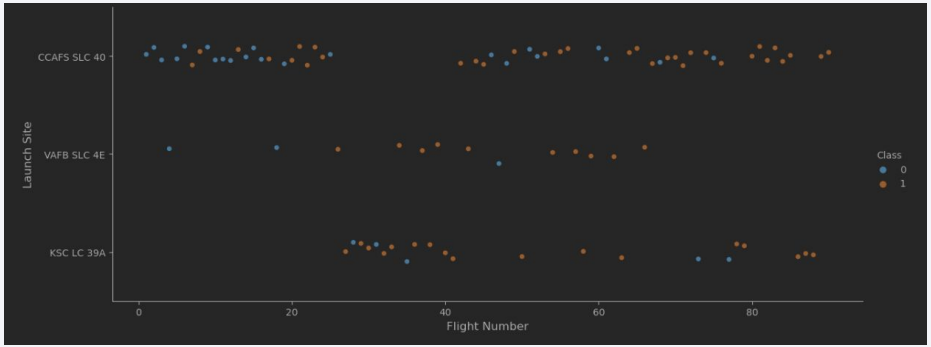
Section 2

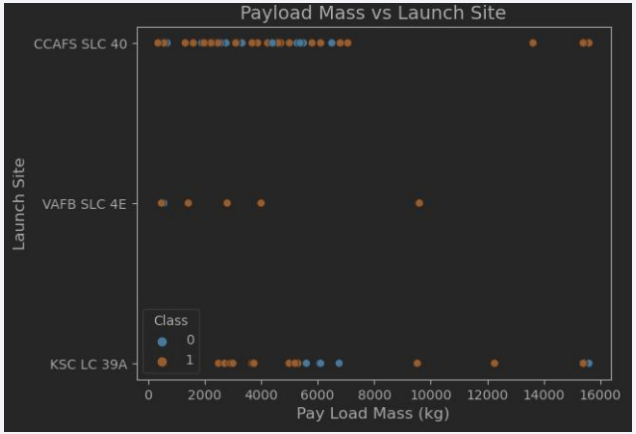
Insights drawn from EDA

Flight Number vs. Launch Site

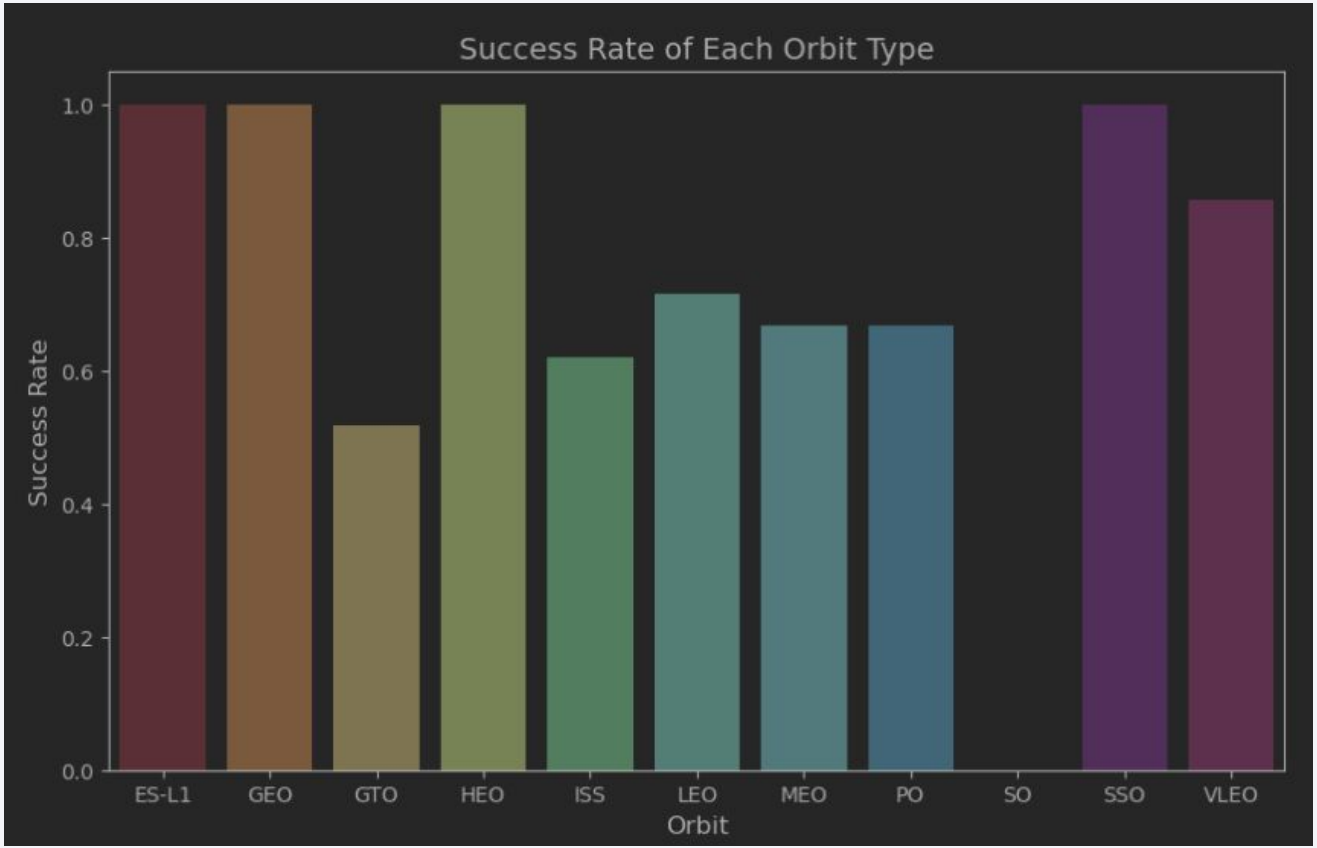
- According to the plot above, it’s possible to verify that the best launch site nowadays is CCAF5 SLC 40, where most of recent launches were successful;
- In second place VAFB SLC 4E and third place KSC LC 39A;
- It’s also possible to see that the general success rate improved over time.
- Payloads over 9,000kg (about the weight of a school bus) have excellent success rate;
- Payloads over 12,000kg seems to be possible only on CCAFS SLC 40 and KSC LC 39A launch sites.

Flight Number vs. Launch Site



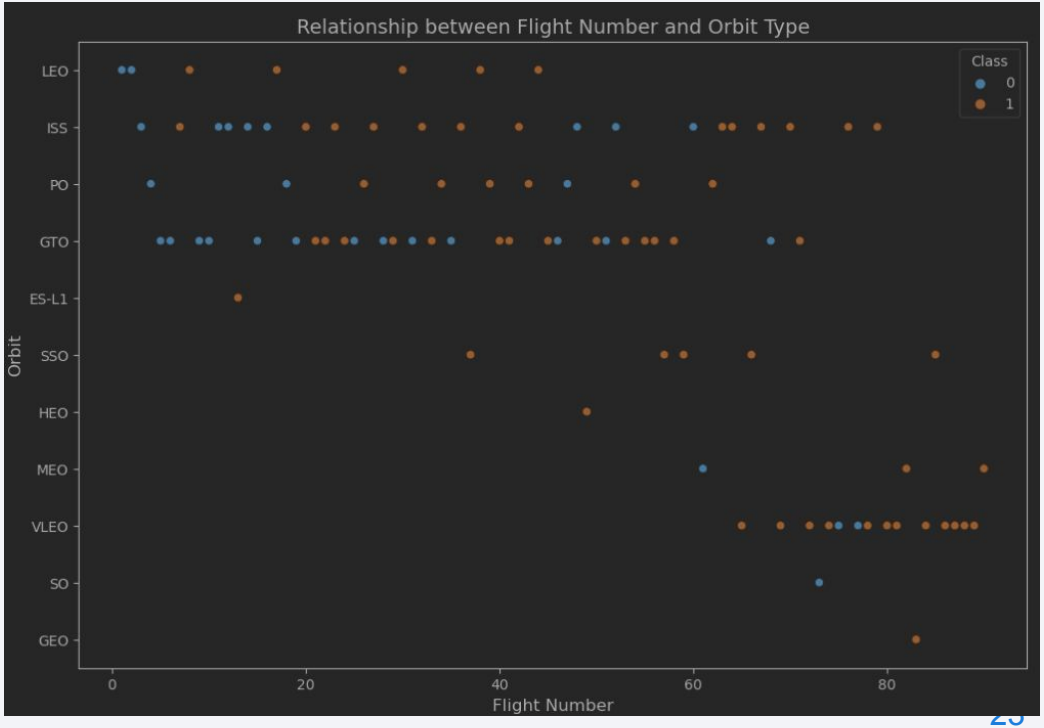


Success Rate vs. Orbit Type

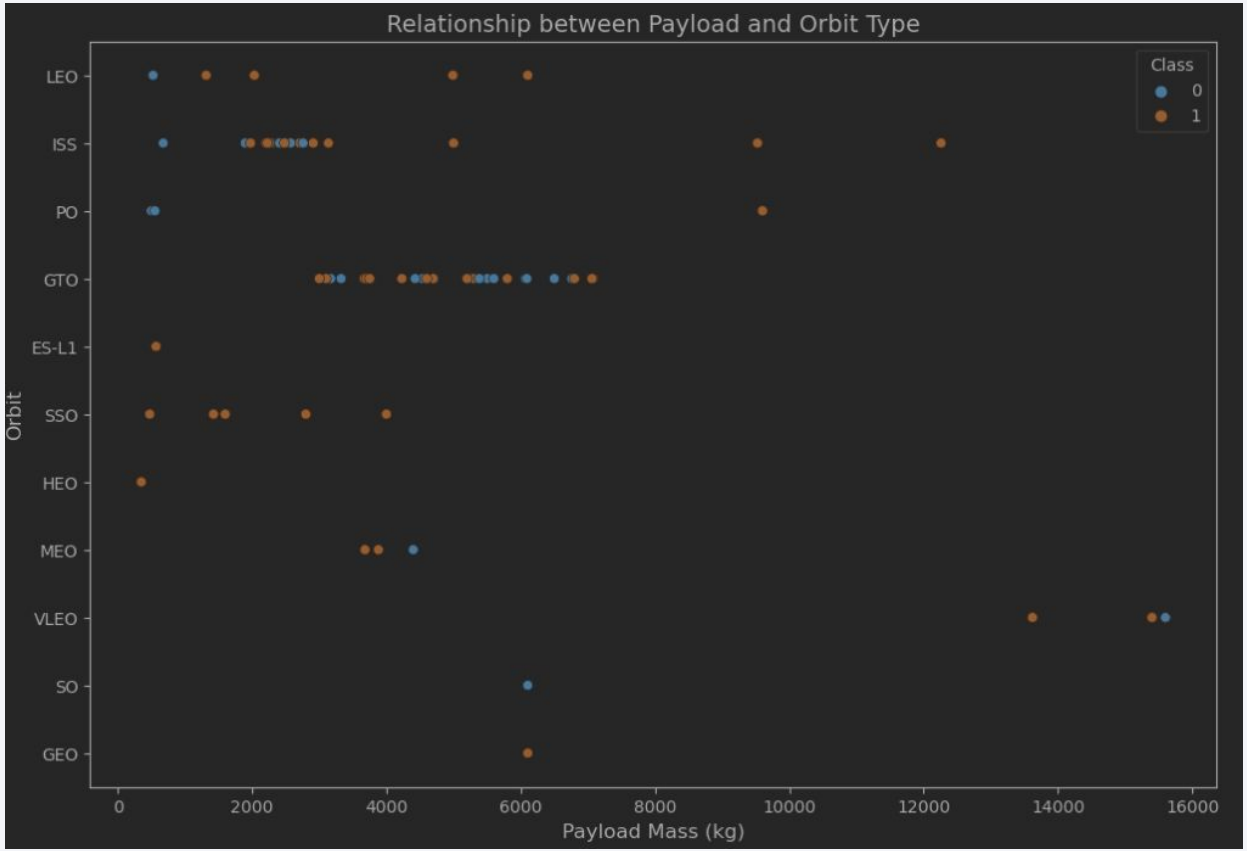


Flight Number vs. Orbit Type

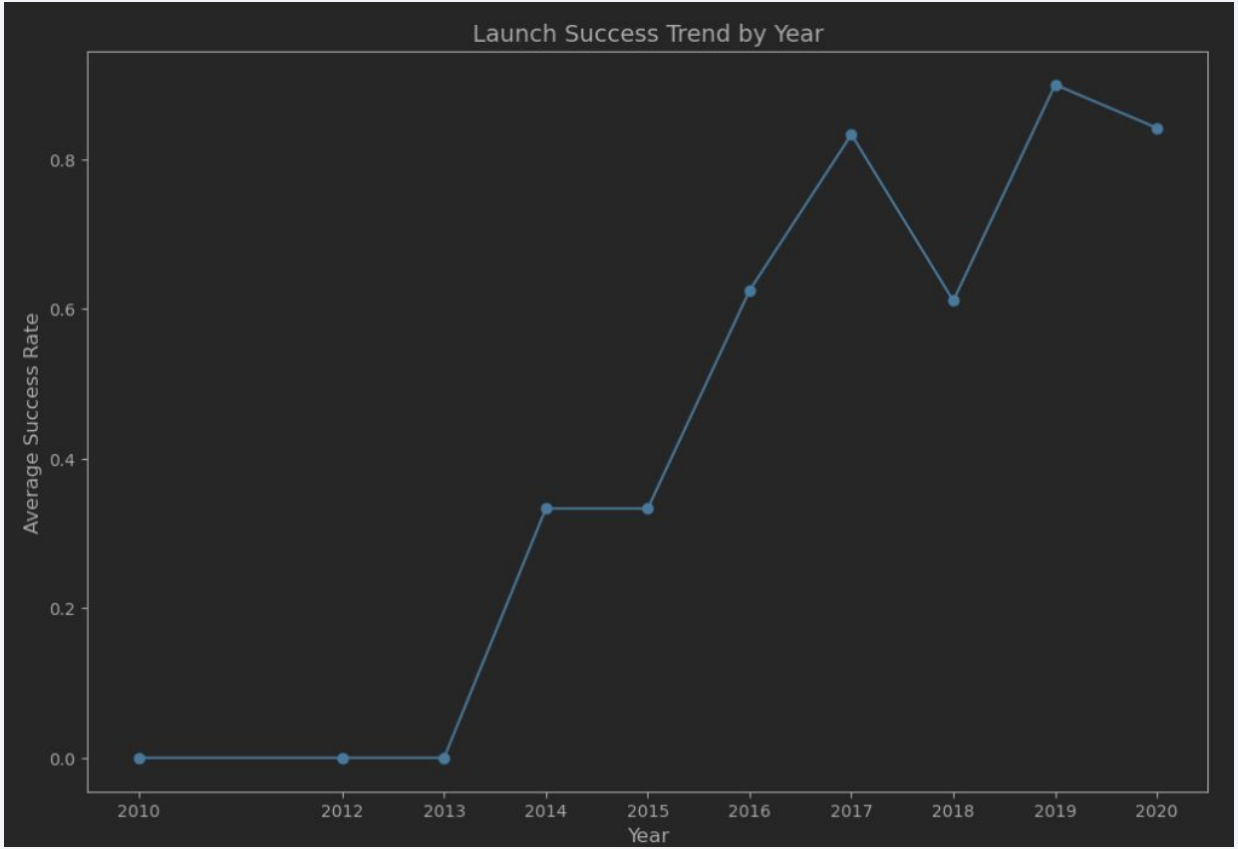
- Apparently, success rate improved over time to all orbits
- VLEO orbit seems a new business opportunity, due to recent increase of its frequency.



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

```python
# クエリを実行して結果を取得
query = "SELECT DISTINCT Launch_Site FROM SPACEX"
cursor = cnx.cursor()
cursor.execute(query)
results = cursor.fetchall()

# 結果を表示
for row in results:
    print(row[0])

# 接続をクローズ
cursor.close()
cnx.close()
```

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

Launch Site Names Begin with 'CCA'

```
# クエリを実行して結果を取得
query = "SELECT * FROM SPACEX WHERE Launch_Site LIKE 'CCA%' LIMIT 5"
cursor = cnx.cursor()
cursor.execute(query)
results = cursor.fetchall()

# 結果を表示
for row in results:
    print(row)

# 接続をクローズ
cursor.close()
cnx.close()
```

```
(datetime.date(2010, 6, 4), '18:45:00', 'F9 v1.0  B0003', 'CCAFS LC-40', 'Dragon Spacecraft Qualification Unit', 0, 'LEO', 'SpaceX', 'Success', 'Failure
 (parachute)')
(datetime.date(2010, 12, 8), '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)')
(datetime.date(2012, 5, 22), '7:44:00', 'F9 v1.0  B0005', 'CCAFS LC-40', 'Dragon demo flight C2', 525, 'LEO (ISS)', 'NASA (COTS)', 'Success', 'No
attempt')
(datetime.date(2012, 10, 8), '0:35:00', 'F9 v1.0  B0006', 'CCAFS LC-40', 'SpaceX CRS-1', 500, 'LEO (ISS)', 'NASA (CRS)', 'Success', 'No attempt')
(datetime.date(2013, 3, 1), '15:10:00', 'F9 v1.0  B0007', 'CCAFS LC-40', 'SpaceX CRS-2', 677, 'LEO (ISS)', 'NASA (CRS)', 'Success', 'No attempt')
```

Total Payload Mass

```python
import mysql.connector

# MySQL接続情報を指定
config = {
    'user': 'root',
    'password': '',
    'host': 'localhost',
    'database': 'SPACEX',
    'raise_on_warnings': True
}

# MySQLに接続
cnx = mysql.connector.connect(**config)

# クエリを実行して結果を取得
query = "SELECT SUM(PAYLOAD_MASS_KG_) FROM SPACEX WHERE Customer = 'NASA (CRS)'"
cursor = cnx.cursor()
cursor.execute(query)
result = cursor.fetchone()

# 結果を表示
total_payload_mass = result[0]
print("Total Payload Mass carried by NASA (CRS):", total_payload_mass, "kg")

# 接続をクローズ
cursor.close()
cnx.close()
```

```
Total Payload Mass carried by NASA (CRS): 45596 kg
```

Average Payload Mass by F9 v1.1

```python
import mysql.connector

# MySQL connection configuration
config = {
    'user': 'root',
    'password': '',
    'host': 'localhost',
    'database': 'SPACEX',
    'raise_on_warnings': True
}

# Connect to MySQL
cnx = mysql.connector.connect(**config)

# Execute the query
query = "SELECT AVG(PAYLOAD_MASS_KG_) FROM SPACEX WHERE Booster_Version = 'F9 v1.1'"
cursor = cnx.cursor()
cursor.execute(query)
result = cursor.fetchone()

# Display the average payload mass
print("Average Payload Mass carried by F9 v1.1 booster:", result[0])

# Close the connection
cursor.close()
cnx.close()
```

```
Average Payload Mass carried by F9 v1.1 booster: 2928.4000
```

First Successful Ground Landing Date

```
import mysql.connector

# MySQL connection configuration
config = {
    'user': 'root',
    'password': '',
    'host': 'localhost',
    'database': 'SPACEX',
    'raise_on_warnings': True
}

# Connect to MySQL
cnx = mysql.connector.connect(**config)

# Execute the query
query = "SELECT MIN(Date) FROM SPACEX WHERE Landing_Outcome LIKE '%ground pad%' AND Landing_Outcome LIKE '%Success%'"
cursor = cnx.cursor()
cursor.execute(query)
result = cursor.fetchone()

# Display the date of the first successful landing on a ground pad
print("Date of the first successful landing on a ground pad:", result[0])

# Close the connection
cursor.close()
cnx.close()
```

```
Date of the first successful landing on a ground pad: 2015-12-22
```

Successful Drone Ship Landing with Payload between 4000 and 6000

```
# Connect to MySQL
cnx = mysql.connector.connect(**config)

# Execute the query
query = "SELECT Booster_Version FROM SPACEX WHERE Landing_Outcome LIKE '%drone ship%' AND PAYLOAD_MASS_KG_ > 4000 AND PAYLOAD_MASS_KG_ < 6000 AND
 Landing_Outcome LIKE '%Success%'"
cursor = cnx.cursor()
cursor.execute(query)
results = cursor.fetchall()

# Display the names of the boosters
print("Boosters with successful landing on a drone ship and payload mass between 4000 and 6000:")
for result in results:
    print(result[0])

# Close the connection
cursor.close()
cnx.close()
```

```
Boosters with successful landing on a drone ship and payload mass between 4000 and 6000:
F9 FT B1022
F9 FT B1026
F9 FT  B1021.2
F9 FT  B1031.2
```

Total Number of Successful and Failure Mission Outcomes

```
# Connect to MySQL
cnx = mysql.connector.connect(**config)

# Execute the query for successful missions
query_success = "SELECT COUNT(*) FROM SPACEX WHERE Mission_Outcome LIKE '%Success%'"
cursor = cnx.cursor()
cursor.execute(query_success)
result_success = cursor.fetchone()[0]

# Execute the query for failed missions
query_failure = "SELECT COUNT(*) FROM SPACEX WHERE Mission_Outcome LIKE '%Failure%'"
cursor.execute(query_failure)
result_failure = cursor.fetchone()[0]

# Display the total number of successful and failure mission outcomes
print("Total number of successful missions: ", result_success)
print("Total number of failure missions: ", result_failure)

# Close the connection
cursor.close()
cnx.close()
```

```
Total number of successful missions:  100
Total number of failure missions:  1
```

Boosters Carried Maximum Payload

```
# Execute the query
query = """
SELECT booster_version
FROM SPACEX
WHERE PAYLOAD_MASS_KG_ = (
    SELECT MAX(PAYLOAD_MASS_KG_)
    FROM SPACEX
)
"""
cursor = cnx.cursor()
cursor.execute(query)
results = cursor.fetchall()

# Display the names of the booster_versions with the maximum payload mass
print("Booster Versions with Maximum Payload Mass:")
for result in results:
    print(result[0])

# Close the connection
cursor.close()
cnx.close()
```

```
Booster Versions with Maximum Payload Mass:
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

```
# Display the records
print("Records for Months in 2015 with Failure Landing Outcomes in Drone Ship:")
for result in results:
    print("Month:", result[0])
    print("Landing Outcome:", result[1])
    print("Booster Version:", result[2])
    print("Launch Site:", result[3])
    print()

# Close the connection
cursor.close()
cnx.close()
```

```
Records for Months in 2015 with Failure Landing Outcomes in Drone Ship:
Month: 01
Landing Outcome: Failure (drone ship)
Booster Version: F9 v1.1 B1012
Launch Site: CCAFS LC-40

Month: 04
Landing Outcome: Failure (drone ship)
Booster Version: F9 v1.1 B1015
Launch Site: CCAFS LC-40
```

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

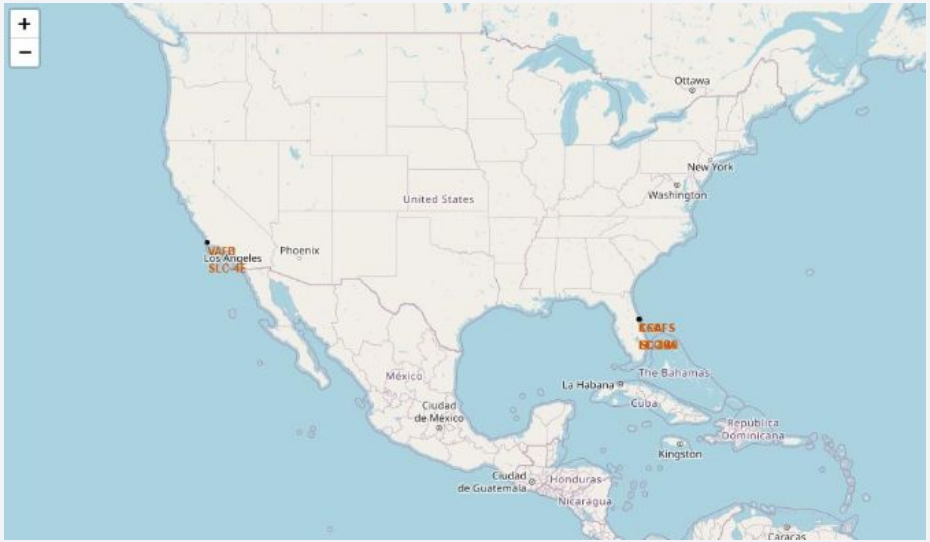
```
cursor = cnx.cursor()
cursor.execute(query)
results = cursor.fetchall()

# Display the records
print("Ranking of Successful Landing Outcomes between 04-06-2010 and 20-03-2017:")
rank = 1
for result in results:
    print("Rank:", rank)
    print("Landing Outcome:", result[0])
    print("Success Count:", result[1])
    print()
    rank += 1

# Close the connection
cursor.close()
cnx.close()
```

```
Ranking of Successful Landing Outcomes between 04-06-2010 and 20-03-2017:
```

Section 3

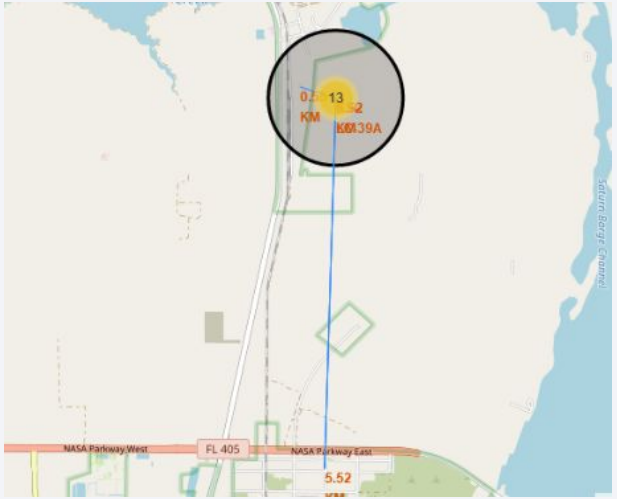
Launch Sites Proximities Analysis

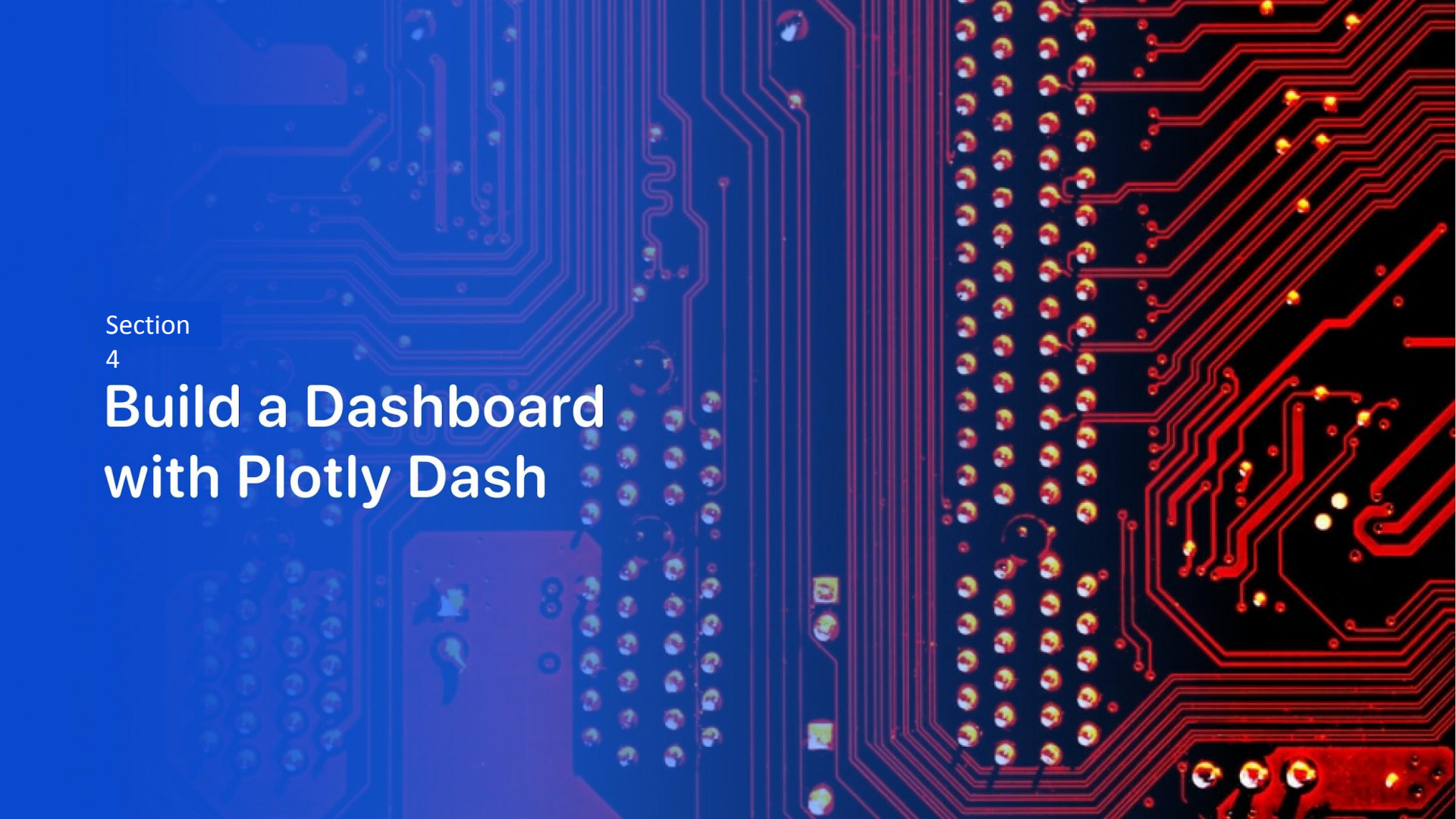
All launch sites



Launch Outcomes by Site

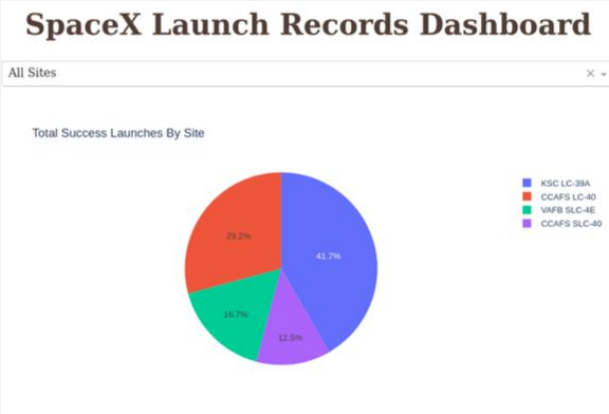
Logistics and Safety




Section
4
Build a Dashboard
with Plotly Dash

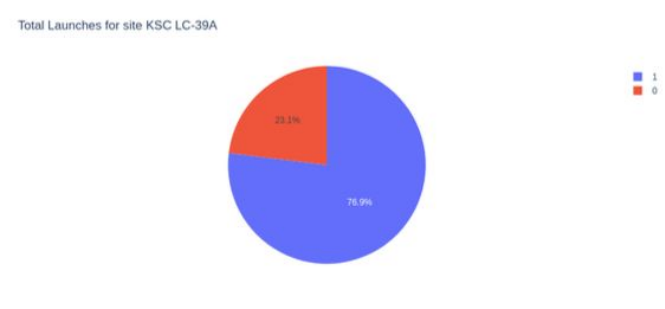
Successful Launches by Site

- The place from where launches are done seems to be a very important factor of success of missions.



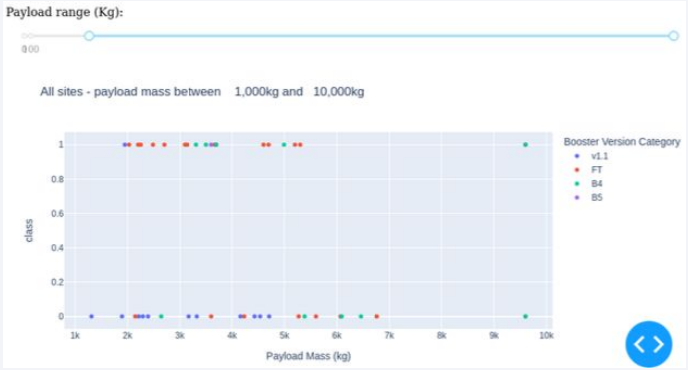
Launch Success Ratio for KSC LC-39A

- 76.9% of launches are successful in this site.



Payload vs. Launch Outcome

- Payloads under 6,000kg and FT boosters are the most successful combination.

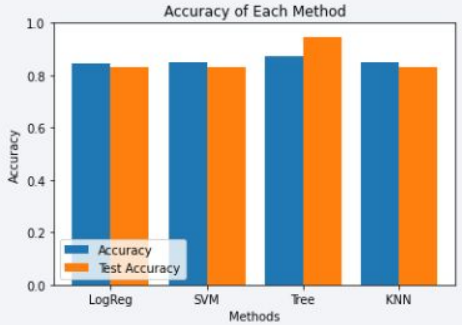


Section
5

Predictive Analysis (Classification)

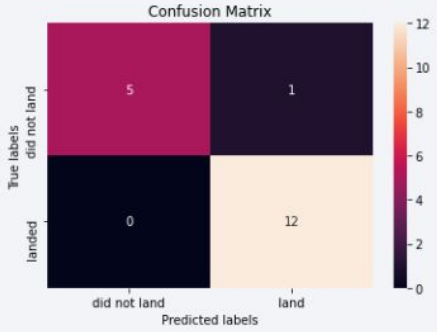
Classification Accuracy

- The model with the highest classification accuracy is Decision Tree Classifier, which has accuracies over than 87%.



Confusion Matrix

- Confusion matrix of Decision Tree Classifier proves its accuracy by showing the big numbers of true positive and true negative compared to the false ones.



Conclusions

- Different data sources were analyzed, refining conclusions along the process;
- The best launch site is KSC LC-39A;
- Launches above 7,000kg are less risky;

Thank you!