IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Roaa Alzanbaqi
4 May 2024

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion

Executive Summary

- Summary of Methodologies
  - Data Collection with API And Web Scraping
  - Data Wrangling
  - Exploratory Data Analysis with SQL
  - Exploratory Data Analysis with Data Visualization
  - Interactive Visual Analytics with Folium
  - Machine Learning Prediction

- Summary of All Results
  - Exploratory Data Analysis
  - Interactive Analytics
  - Predictive Analytics

Introduction

In its quest to challenge Space X, the new rocket company, Space Y, tasked its data science team with the following objectives:

1. Gather information about Space X
2. Determine the pricing for each launch
3. Create dashboards to visually present key insights
4. Train a machine learning model to predict whether Space X will reuse the first stage

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
    - Data collection - SpaceX API
    - Data collection - Web scraping
- 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

The data sets were collected by:

- Collecting the data with an API from the SpaceX REST API, providing details about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome

- Collecting the data with Web scraping using Beautiful Soup to extract HTML tables to get Falcon 9 launch records

Data Collection – SpaceX API

- Request and parse the SpaceX launch data using the GET request
- Decode the response content as a Json and turn it into a Pandas dataframe
- create a new dataframe with columns rocket, payloads, launchpad, and cores
- Filter the dataframe to only include Falcon 9 launches
- GitHub URL: https://github.com/Roaweeh/IBM-Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Task 1: Request and parse the SpaceX launch data using the GET request

To make the requested JSON results more consistent, we will use the following static response object for this project:

```
static_json_url='https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/da
```

We should see that the request was successfull with the 200 status response code

```
response.status_code
```

```
200
```

Now we decode the response content as a Json using `.json()` and turn it into a Pandas dataframe using `.json_normalize()`

```
# Use json_normalize meethod to convert the json result into a dataframe
data= pd.json_normalize(response.json())
```

Using the dataframe `data` print the first 5 rows

```
# Get the head of the dataframe
data.head()
```

	static_fire_date_utc	static_fire_date_unix	net	window	rocket	success	failures	details	crew
0	2006-03-17T00:00:00.000Z	1.142554e+09	False	0.0	5e9d0d95eda69955f709d1eb	False	[{'time': 33, 'altitude': None, 'reason': 'merlin engine failure'}]	Engine failure at 33 seconds and loss of vehicle	[]

Data Collection - Scraping

- Web scrap Falcon 9 launch records with BeautifulSoup
- Extract a Falcon 9 launch records HTML table from Wikipedia
- Parse the table and convert it into a Pandas data frame
- GitHub URL: https://github.com/Roaweeh/IBM-Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb

TASK 1: Request the Falcon9 Launch Wiki page from its URL

First, let's perform an HTTP GET method to request the Falcon9 Launch HTML page, as an HTTP response.

```python
# use requests.get() method with the provided static_url
# assign the response to a object
data  = requests.get(static_url).text
print(data)
```


```
<!DOCTYPE html>
<html class="client-nojs vector-feature-language-in-header-enabled vector-feature-language-in-main-page-header-disabled vector-feature-sticky-header-disabled vector-feature-page-tools-pinned-disabled vector-feature-toc-pinned-clientpref-1 vector-feature-main-menu-pinned-disabled vector-feature-limited-width-clientpref-1 vector-feature-limited-width-content-enabled vector-feature-custom-font-size-clientpref-0 vector-feature-client-preferences-disabled vector-feature-client-prefs-pinned-disabled vector-toc-available" lang="en" dir="ltr">
<head>
<meta charset="UTF-8">
<title>List of Falcon 9 and Falcon Heavy launches - Wikipedia</title>
<script>(function(){var className="client-js vector-feature-language-in-header-enabled vector-feature-language-in-main-page-header-disabled vector-feature-sticky-header-disabled vector-feature-page-tools-pinned-disabled vector-feature-toc-pinned-clientpref-1 vector-feature-main-menu-pinned-disabled vector-feature-limited-width-clientpref-1 vector-feature-limited-width-content-enabled vector-feature-custom-font-size-clientpref-0 vector-feature-client-preferences-disabled vector-feature-client-prefs-pinned-disabled vector-toc-available";var cookie=document.cookie.match(/(?:^|; )enwikimwclientpreferences=([^;]+)/);if(cookie){cookie[1].split('%2C').forEach(function(pref){className=className.replace(new RegExp('(^| )'+pref.replace(/-clientpref-\w+$|[^\w-]+/g,'')+'-clientpref-\\w+( |$)'),'$1'+pref+'$2');});}document.documentElement.className=className;}());RLCONF={"wgBreakFrames":false,"wgSeparatorTransformTable":["",""],"wgDigitTransformTable":["",""],"wgDefaultDateFormat":"dmy","wgMonthNames":["","January","February","March","April","May","June","July","August","September","October","November","December"],"wgRequestId":"fa58b3ec-7e74-407b-811a-73f9c9061ed1","wgCanonicalNamespace":"","wgCanonicalSpecialPageName":false,"wgNamespaceNumber":0,"wgPageName":"List_of_Falcon_9_and_Falcon_Heavy_launches","wgTitle":"List of Falcon 9 and Falcon Heavy launches","wgCurRevisionId":1199083341,"wgRevisionId":1027686922,"wgArticleId":37574004,"wgIsArticle":
```

Data Wrangling

The raw dataset was transformed into a clean dataset and performed some Exploratory Data Analysis (EDA) to find some patterns in the data :

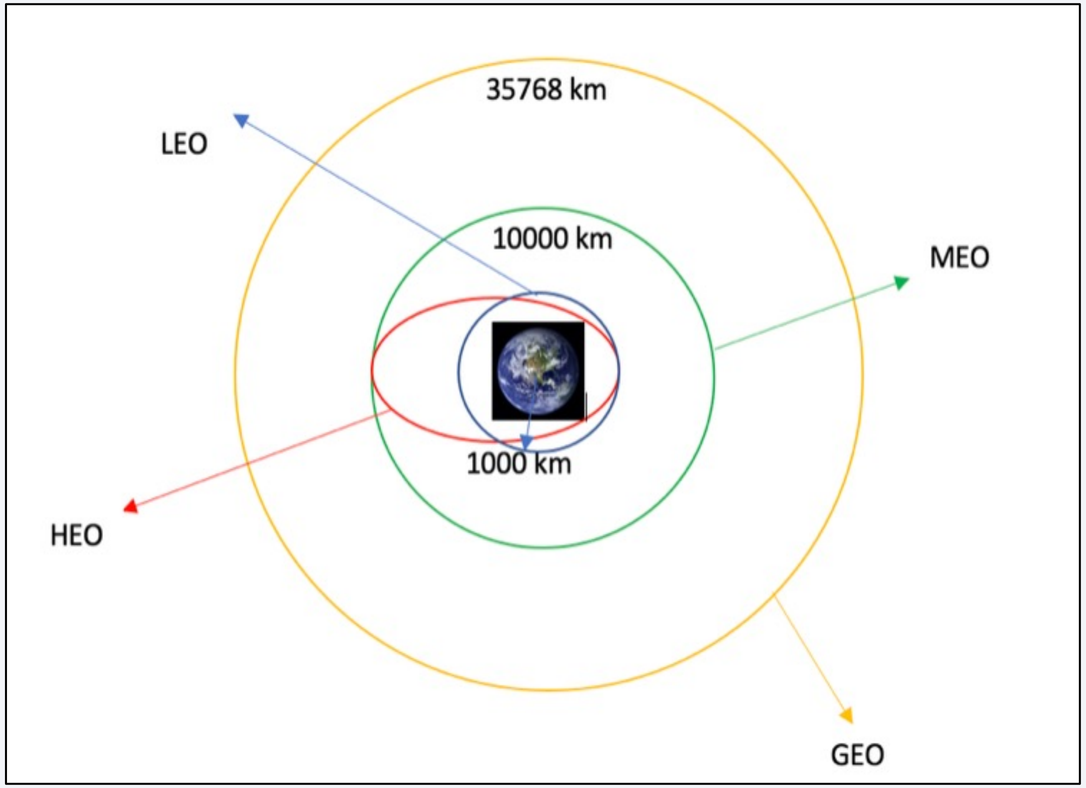
- Identify and calculate the percentage of the missing values
- Calculate the number of launches on each site
- Calculate the number and occurrence of each orbit
- Calculate the number and occurrence of mission outcome of the orbits
- Create a landing outcome label from Outcome column

- GitHub URL: https://github.com/Roaweeh/IBM-Data-Science-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

Perform Exploratory Data Analysis and Feature Engineering using Pandas and Matplotlib.
Charts plotted:

- Flight number Vs Launch site
- Payload Vs Launch site
- Success rate Vs Orbit type
- Flight number Vs Orbit type
- Payload Vs Orbit type
- The launch success yearly trend
- GitHub URL: https://github.com/Roaweeh/IBM-Data-Science-Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb



EDA with SQL

Execute the SQL queries to:

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in ground pad was achieved
- List the names of the boosters which have success in drone ship and have 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 records which will display the month names, failure landing outcomes in drone ship ,booster versions, launch site for the months in year 2015
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

- GitHub URL: https://github.com/Roaweeh/IBM-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Performed interactive visual analytics using Folium to:

- Mark all launch sites on a map
- Mark the success/failed launches for each site on the map
- Calculate the distances between a launch site to its proximities
- GitHub URL: https://github.com/Roaweeh/IBM-Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.jupyterlite.ipynb

Predictive Analysis (Classification)

Applied Machine Learning Prediction by:

- Created a column for the class
- Standardized the data
- Split the data into training and test data
- Found the best Hyperparameter for the logistic regression, support vector machine, decision tree classifier, and k-nearest neighbors methods
- Found the method performs best using test data
- GitHub URL: https://github.com/Roaweeh/IBM-Data-Science-Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results

1. The launch success rate has been progressively improving over the years
2. Factors to consider when selecting a launch site include its proximity to the Equator and its location near the coastline
3. The KSC LC-39A is the optimal launch site, boasting the highest success rate of approximately 80% for successful launches across payloads of various masses
4. The selection of the most suitable orbit depends on the payload mass. For payloads weighing over 8000 kg, it is recommended to consider the ISS, PO, and VLEO orbits. Conversely, for payloads under 8000 kg, it is advisable to use the LEO, GTO, and SSO orbits
5. The four machine learning models exhibited similar performance

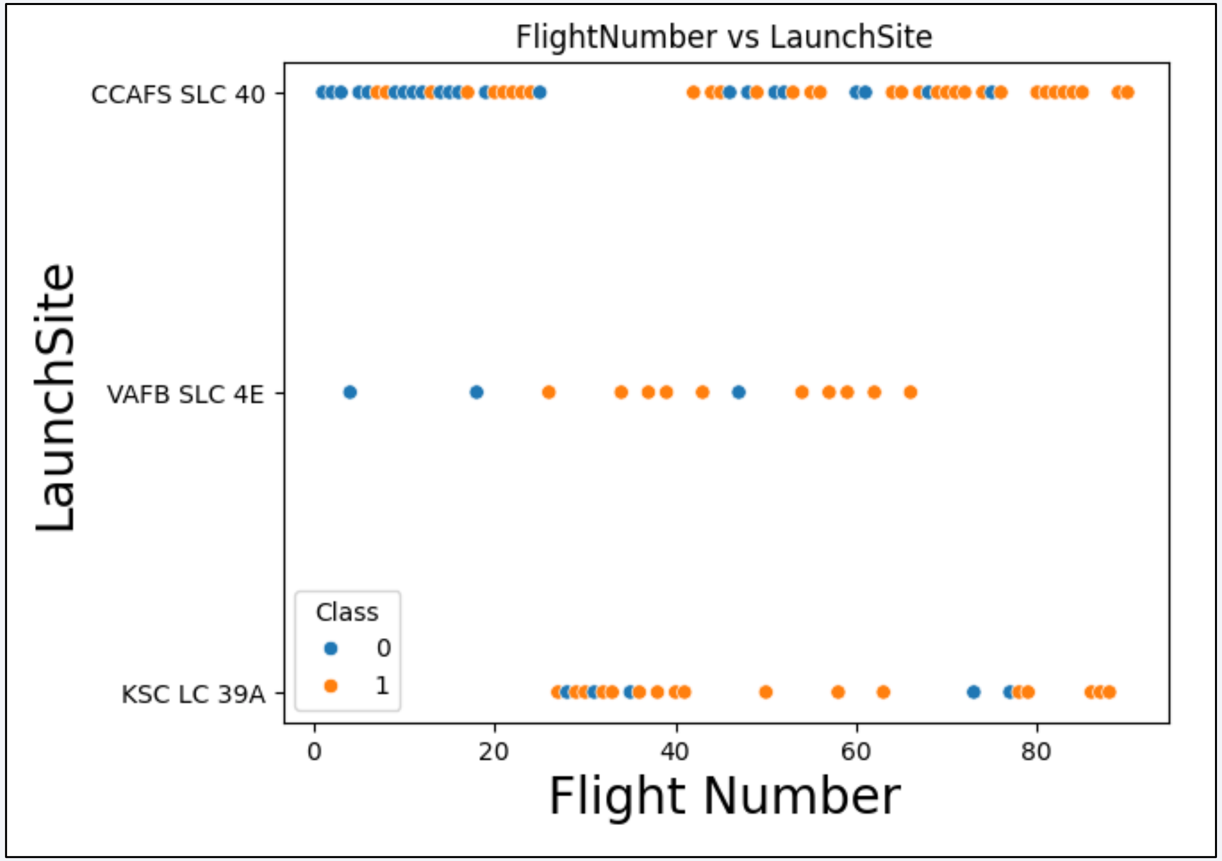
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

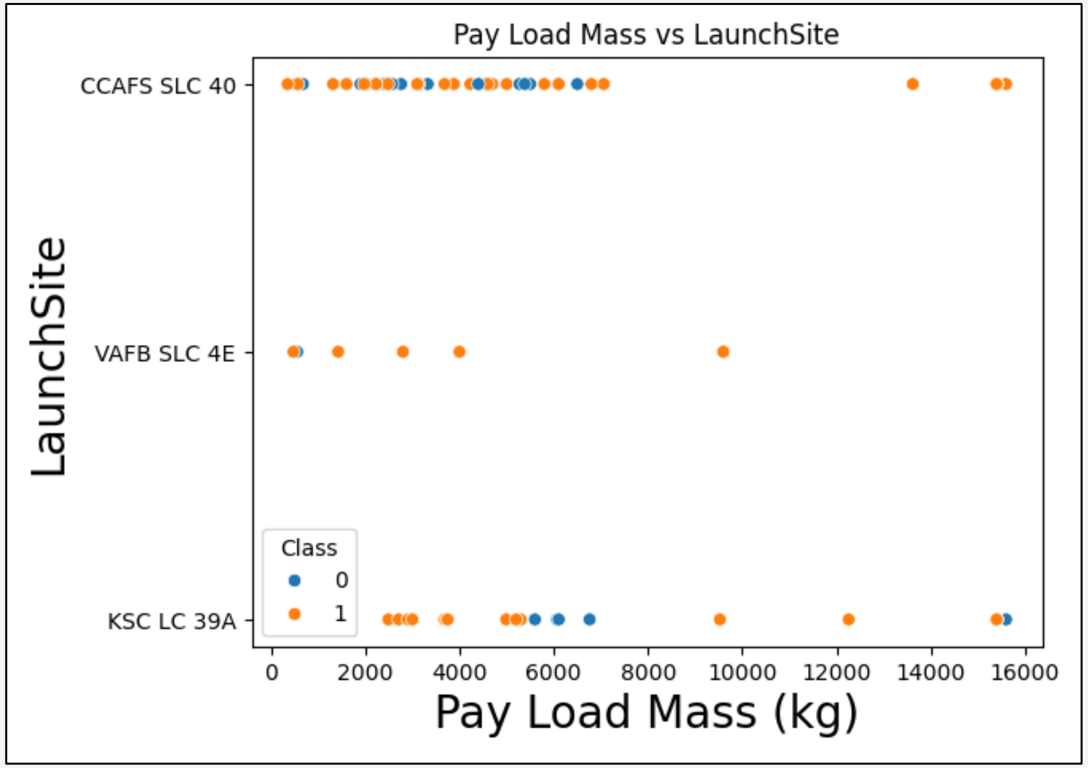
- The scatter plot shows the relation between the flight number and the launch site

- Based on the scatter plot, there does not appear to be a clear correlation between the launch site and the number of flights that have occurred



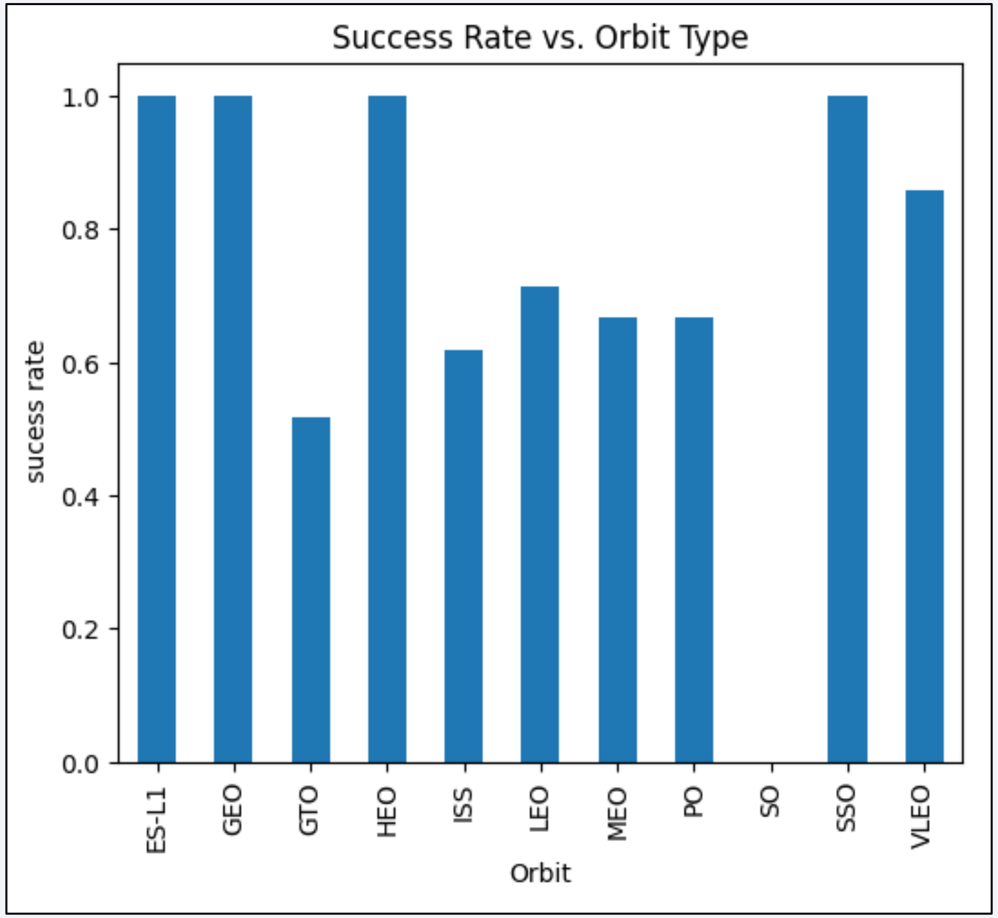
Payload vs. Launch Site

- The scatter plot shows the relation between the payload mass and the different launch sites
- At the VAFB-SLC launch site, there have been no launches of rockets carrying payloads exceeding 10,000 kg
- At the CCAFS-SLC launch site, there have been no launches of rockets carrying payloads between 8,000 and 13,000 kg
- Among the three launch sites, KSC-LC was the one that conducted launches with payloads of different masses



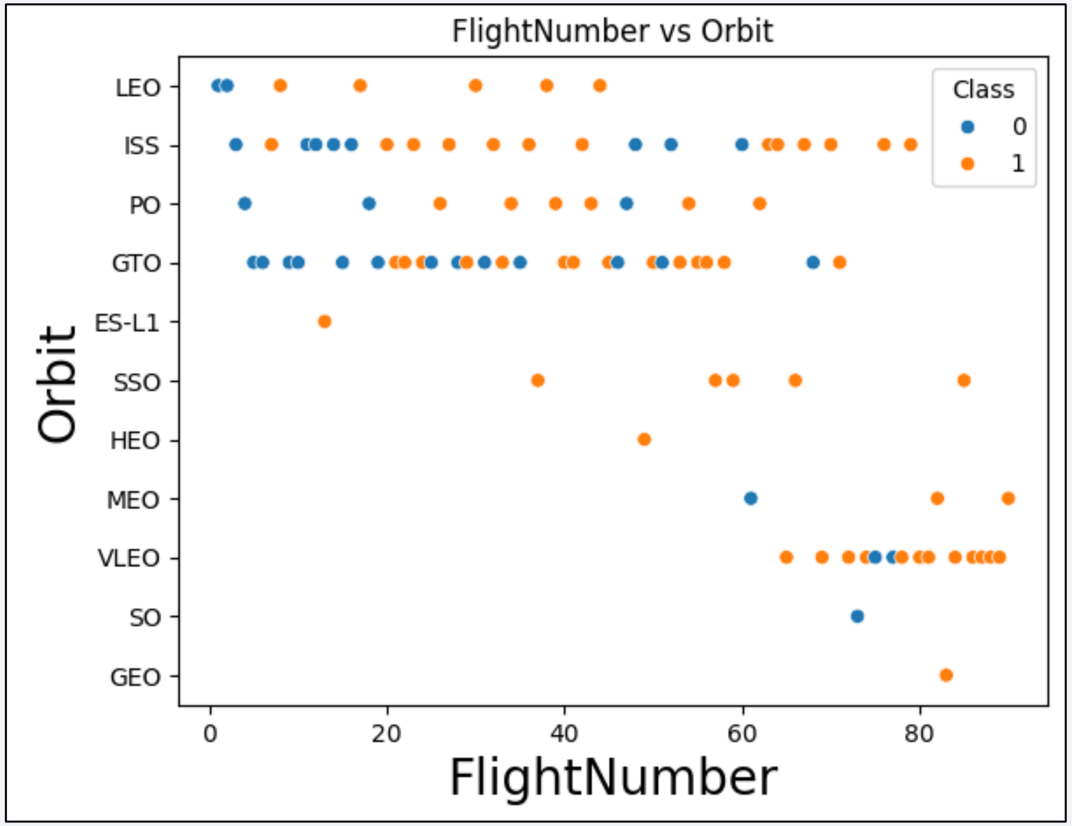
Success Rate vs. Orbit Type

- The bar chart shows the relation between success rate and the orbit type
- Based on the data shown in the figure, it can be inferred that the orbits ES-L1, GEO, HEO, and SSO exhibit the highest success rates among all orbit types with 100% success rate
- Other orbits have a success rates between 50% up to 80%
- Among all orbit types, the SO orbit achieved a 0% success rate, representing the lowest success rate



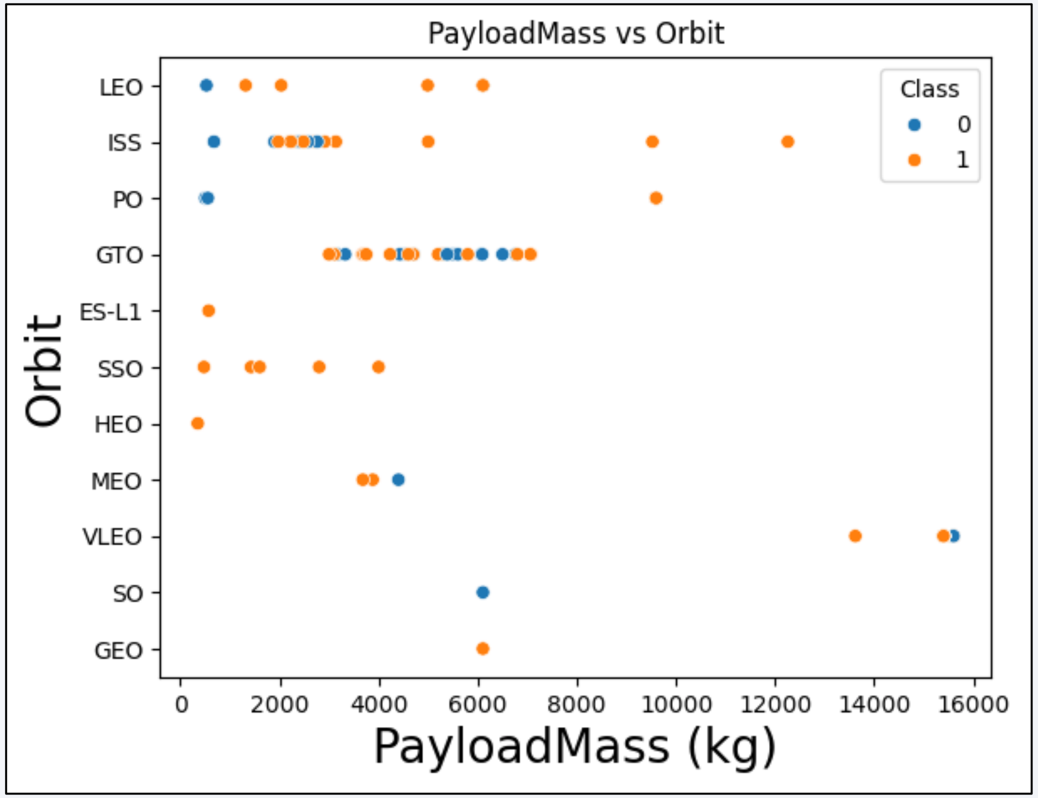
Flight Number vs. Orbit Type

- The scatter plot shows the relation between flight number and the orbit type
- Comparing the scatter plot of Flight Number vs Orbit Type with the Success Rate vs Orbit Type bar plot, we conclude that the reported 100% success rate for orbits ES-L1, GEO, and HEO is based on misleading data from a single flight occurrence and should not be considered valid
- Similarly, the SO orbit achieved a 0% success rate due to a single failed flight occurrence
- For a more reliable analysis, it is advisable to exclude any orbit with fewer than 5 total flight numbers



Payload vs. Orbit Type

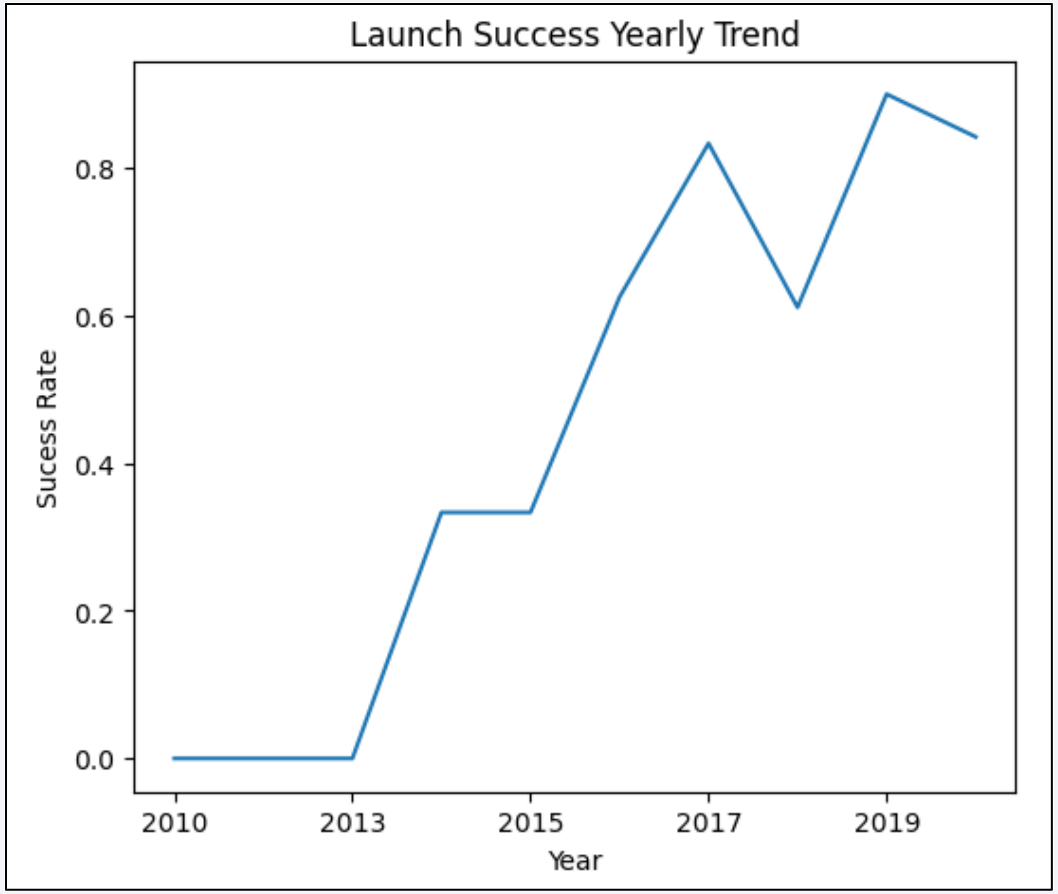
- The scatter plot shows the relation between flight number and the orbit type
- For payloads over 8000 kg, consider using the ISS, PO and VLEO orbits for their proven success
- For payloads under 8000 kg, consider using the LEO, GTO and SSO orbits for their higher success rates compared to other orbits



Launch Success Yearly Trend

- The line chart shows the relation between launch success and the yearly trend

- The chart shows that the overall success rate of launches has consistently increased from 2013 to 2020



All Launch Site Names

- Find the names of the unique launch sites:

- Query:

```
%sql select Distinct(Launch_Site) from SPACEXTBL
```

- There are four launch sites, they are:
- CCAFC-LC-40
- VAFB-SLC-4E
- KSC-LC-39A
- CCAFS-SLC-40

- Result:

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`:

- Query:

```
%sql select * from SPACEXTBL where Launch_Site like'CCA%' limit 5
```

- The table shows five launch sites names that begin with `CCA`
- 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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

- Calculate the total payload carried by boosters from NASA (CRS):

- Query:

```
%sql select Customer, Booster_Version, sum(PAYLOAD_MASS__KG_) as total from SPACEXTBL
where Customer='NASA (CRS)' group by Booster_Version order by total DESC
```

- The table displays the total payload masses in kilograms for each NASA (CRS) booster, with the booster version F9 B4 B1039.1 achieving the highest total payload mass of 3,310 kg

- The combined total payload mass for all boosters is 45,596 kg

total
45596

- Result:

Customer	Booster_Version	total
NASA (CRS)	F9 B4 B1039.1	3310
NASA (CRS)	F9 FT B1021.1	3136
NASA (CRS)	F9 B5 B1058.4	2972
NASA (CRS)	F9 FT B1035.1	2708
NASA (CRS)	F9 B4 B1045.2	2697
NASA (CRS)	F9 B4 B1039.2	2647
NASA (CRS)	F9 B5B1050	2500
NASA (CRS)	F9 B5B1056.1	2495
NASA (CRS)	F9 FT B1031.1	2490
NASA (CRS)	F9 v1.1 B1012	2395
NASA (CRS)	F9 v1.1	2296
NASA (CRS)	F9 B5 B1056.2	2268
NASA (CRS)	F9 FT B1025.1	2257
NASA (CRS)	F9 v1.1 B1010	2216
NASA (CRS)	F9 FT B1035.2	2205
NASA (CRS)	F9 B5 B1059.2	1977
NASA (CRS)	F9 v1.1 B1018	1952
NASA (CRS)	F9 v1.1 B1015	1898
NASA (CRS)	F9 v1.0 B0007	677
NASA (CRS)	F9 v1.0 B0006	500

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1:

- Query:

```
%sql select avg(PAYLOAD_MASS__KG_) as avg from SPACEXTBL where Booster_Version='F9 v1.1'
```

- The average payload mass carried by booster version F9 v1.1 is about 2928.4 kg

- Result:

avg
2928.4

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad:

- Query:

```
%sql select min(Date) as date from SPACEXTBL where Landing_Outcome='Success (ground pad)'
```

- The first successful landing on a ground pad was achieved on December 22, 2015

- Result:

date
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

- Query:

```
%sql select Booster_Version,Landing_Outcome,PAYLOAD_MASS__KG_ from SPACEXTBL where Landing_Outcome='Success (drone ship)'
and PAYLOAD_MASS__KG_ between '4000' and '6000'
```

- The boosters that have successfully landed on a drone ship with a payload mass greater than 4,000 kg but less than 6,000 kg are:
- F9 FT B1022
- F9 FT B1026
- F9 FT B1021.2
- F9 FT B1031.2

- Result:

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes:

- Query:

```
%sql select Mission_Outcome, count(Mission_Outcome) as total from SPACEXTBL group by(Mission_Outcome)
```

- There have been a total of 100 successful missions, with only one mission that ended in failure

- Result:

Mission_Outcome	total
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass:

- Query:

```
%sql select Booster_Version from SPACEXTBL where PAYLOAD_MASS__KG_ =(select max(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

- The table shows the names of boosters that have carried a maximum payload mass of approximately 15,600 kg

- Result:

Booster_Version	PAYLOAD_MASS__KG_
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

2015 Launch Records

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

- Query:

```
%sql select Booster_Version, substr(Date, 6,2) as month, substr(Date,0,5) as year, Landing_Outcome as failure_landing_outcomes, Launch_Site
from SPACEXTBL where Landing_Outcome='Failure (drone ship)' and year = '2015'
```

- There were two unsuccessful landing attempts on a drone ship that occurred in 2015

- Result:

Booster_Version	month	year	failure_landing_outcomes	Launch_Site
F9 v1.1 B1012	01	2015	Failure (drone ship)	CCAFS LC-40
F9 v1.1 B1015	04	2015	Failure (drone ship)	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:

- Query:

```
%sql select Landing_Outcome,Count(Landing_Outcome) as No_Landing_Outcome, date as Date from SPACEXTBL
where date between '2010-06-0' and '2017-03-2' group by Landing_Outcome order by No_Landing_Outcome DESC
```

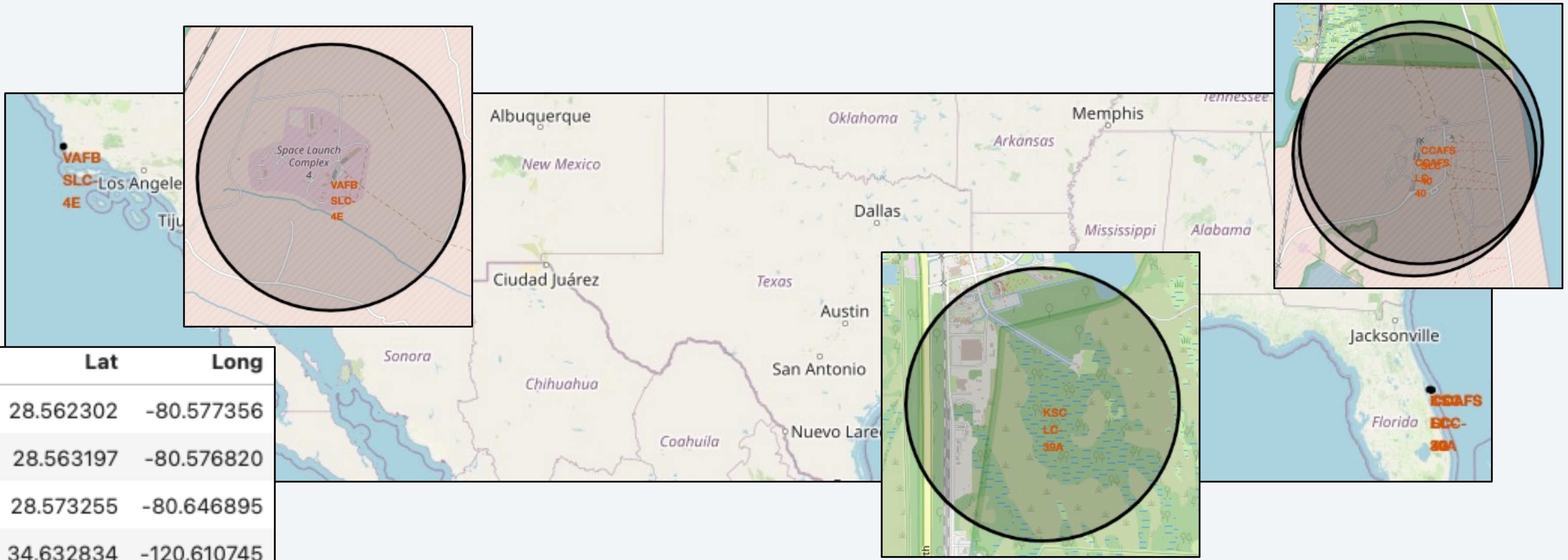
- Result:

Landing_Outcome	No_Landing_Outcome	Date
No attempt	10	2012-05-22
Success (drone ship)	5	2016-04-08
Failure (drone ship)	5	2015-01-10
Success (ground pad)	3	2015-12-22
Controlled (ocean)	3	2014-04-18
Uncontrolled (ocean)	2	2013-09-29
Failure (parachute)	2	2010-06-04
Precluded (drone ship)	1	2015-06-28

Section 3

Launch Sites Proximities Analysis

All Launch Sites

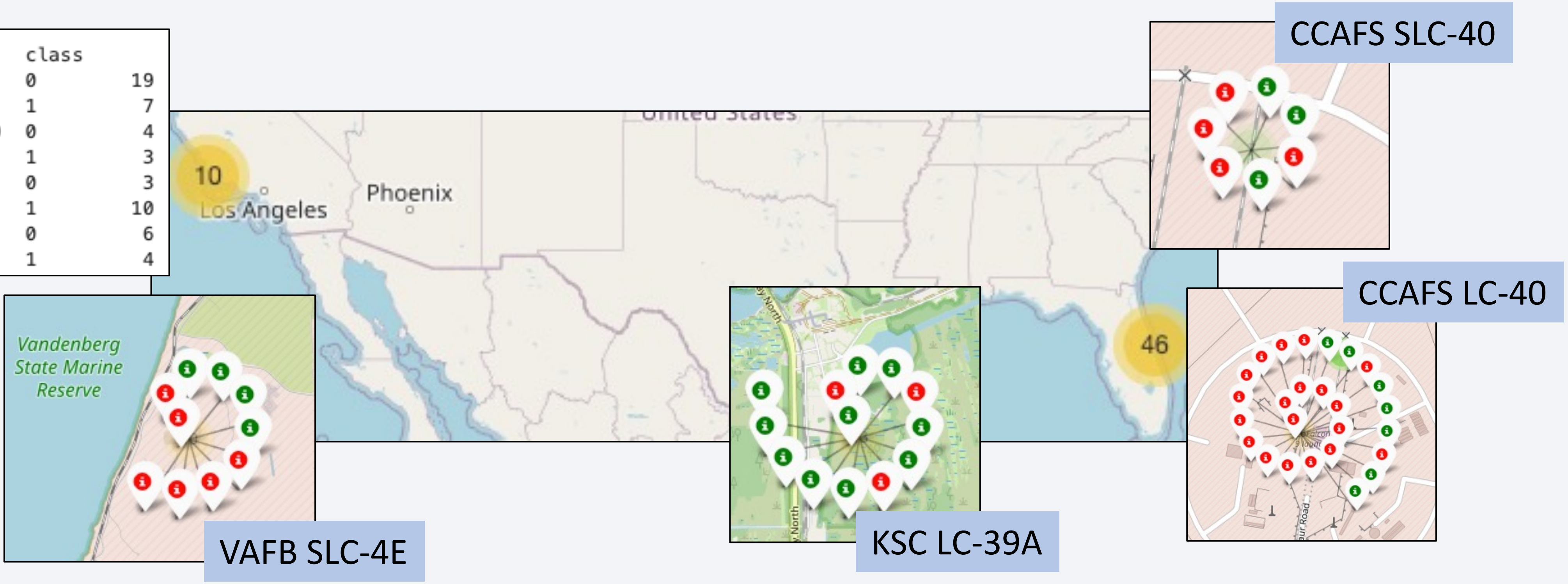


Launch Site	Lat	Long
CCAFS LC-40	28.562302	-80.577356
CCAFS SLC-40	28.563197	-80.576820
KSC LC-39A	28.573255	-80.646895
VAFB SLC-4E	34.632834	-120.610745

- Based on the map screenshot, it's evident that the four launch sites are:
- Located near the Equator
- Positioned close to the coastline

Success and Failed Launches for Each Site

Launch Site	class	
CCAFS LC-40	0	19
	1	7
CCAFS SLC-40	0	4
	1	3
KSC LC-39A	0	3
	1	10
VAFB SLC-4E	0	6
	1	4

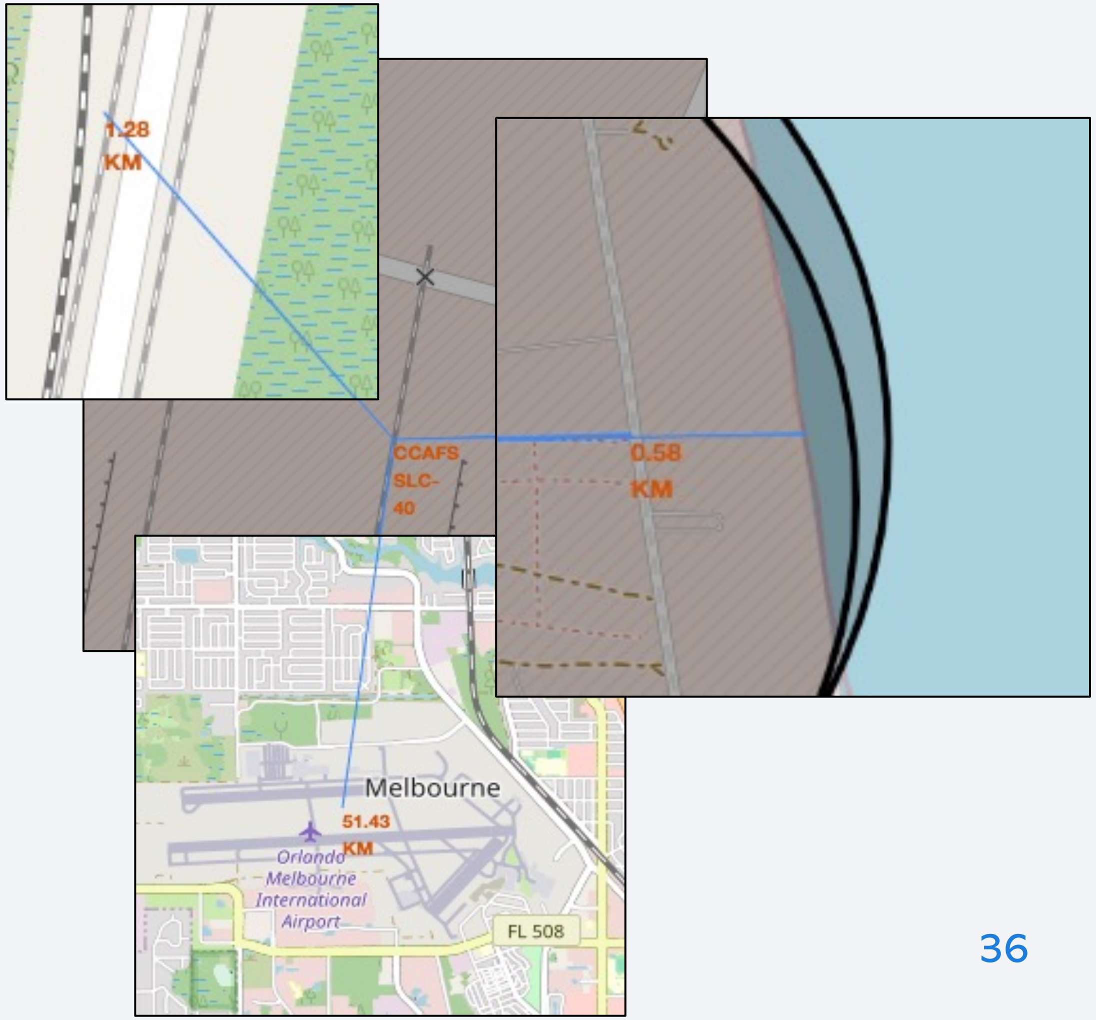


- The launch site KSC LC-39A maintains the highest success rate, estimated at approximately 80% for successful launches, while other launch sites have success rates of roughly less than 50%

The Distance Between a Launch Site to its Proximities

- All launch sites are situated at a distance from nearby cities but are conveniently located near highways, railways, and airport

- The launch site KSC LC-39A is located slightly farther away from the coast compared to the other launch sites



Section 5

Predictive Analysis (Classification)

Classification Accuracy

- The accuracy for all built classification models:

Logistic Regression:

```
logreg_cv.score(X_test, Y_test)
```

0.8333333333333334

Support Vector Machine:

```
svm_cv.score(X_test, Y_test)
```

0.8333333333333334

Decision Tree Classifier:

```
tree_cv.score(X_test, Y_test)
```

0.8333333333333334

K-Nearest Neighbors

```
knn_cv.score(X_test, Y_test)
```

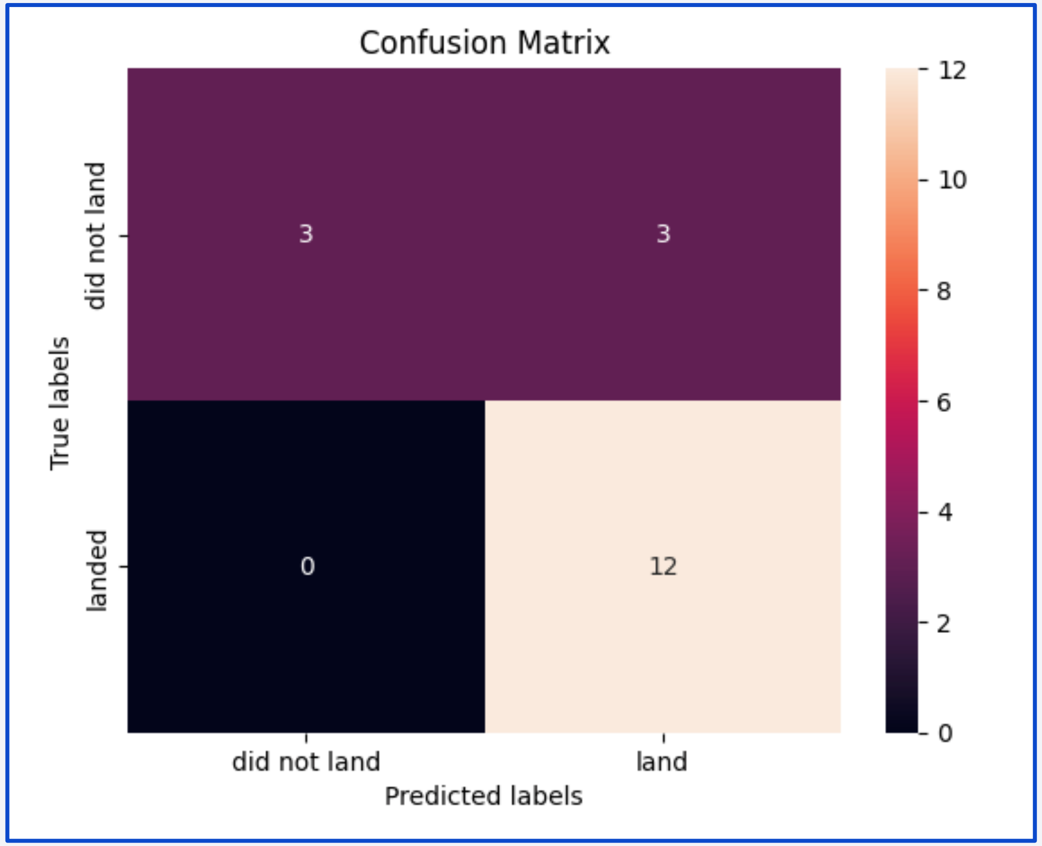
0.8333333333333334

- The logistic regression, support vector machine, decision tree classifier, and k-nearest neighbors models all produce similar accuracy results

Confusion Matrix

- The confusion matrix provides a summary of a classification algorithm's performance

- The confusion matrices for logistic regression, support vector machine, decision tree classifier, and k-nearest neighbors models were all identical having:
- 12 True Positive
- 3 True Negative
- 3 False Positive
- 0 False Negative



Conclusions

1. The launch success rate has been progressively improving over the years
2. Factors to consider when selecting a launch site include its proximity to the Equator and its location near the coastline
3. The KSC LC-39A is the optimal launch site, boasting the highest success rate of approximately 80% for successful launches across payloads of various masses
4. The selection of the most suitable orbit depends on the payload mass. For payloads weighing over 8000 kg, it is recommended to consider the ISS, PO, and VLEO orbits. Conversely, for payloads under 8000 kg, it is advisable to use the LEO, GTO, and SSO orbits
5. The four machine learning models exhibited similar performance

Thank you!