IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

LE Thanh Tai
10-05-2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data collection
  - Data wrangling
  - Exploratory Data Analysis with Data Visualization
  - Exploratory Data Analysis with SQL
  - Building an interactive map with Folium
  - Building a Dashboard with Plotly Dash
  - Predictive analysis (Classification)
- Summary of all results
  - Exploratory Data Analysis results
  - Interactive analytics demo in screenshots
  - Predictive analysis results

Introduction

- Project background and context:
  - SpaceX is a leader in making space travel more accessible by offering cheaper launch costs. Their Falcon 9 rockets are advertised at a fraction of the price compared to competitors (around 62 million dollars versus 165 million or more). This affordability is largely due to SpaceX's ability to reuse the first stage of the rocket. By predicting whether a specific launch will successfully recover the first stage, we can potentially estimate the launch cost. To achieve this, we can leverage publicly available information and machine learning models to forecast SpaceX's first-stage reusability.
- Problems you want to find answers:
  - How do variables such as payload mass, launch site, number of flights, and orbits affect the success of the first stage landing?
  - Does the rate of successful landings increase over the years?
  - What is the best algorithm that can be used for binary classification in this case?

Section 1

Methodology

Methodology

- Data collection methodology:
  - Using SpaceX Rest API
  - Using Web Scrapping from Wikipedia
- Perform data wrangling
  - Filtering the data
  - Dealing with missing values
  - Using One Hot Encoding to prepare the data to a binary classification
- 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

- To gather comprehensive launch information for a thorough analysis, we employed two data collection techniques: SpaceX's REST API and web scraping.
- SpaceX REST API provided valuable data points including flight number, launch date, booster version, payload mass, orbital path, launch site, mission outcome, number of flights the booster completed, grid fin presence, reusability status, landing legs details, landing pad used, booster block version, reusability count, booster serial number, and launch location coordinates (longitude and latitude).
- Web scraping a table from SpaceX's Wikipedia entry yielded additional launch details such as flight number, launch site, payload description, payload mass, orbital path, customer information, launch outcome, booster version, booster landing status, launch date, and launch time.

Data Wrangling

- Github URL: link

Data Wrangling

- Describe how data were processed
- You need to present your data wrangling process using key phrases and flowcharts
- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose

EDA with Data Visualization

- Scatter plots are our tool to explore connections between various launch parameters. If a clear pattern emerges, these variables could be used to build a machine learning model for prediction.
- Bar charts come in handy when we want to compare different launch attributes (like launch sites) and see how they relate to a specific measurement (e.g., success rate).
- Line charts are ideal for visualizing trends over time. For instance, we can use them to see if SpaceX's first-stage landing success rate has improved over the years.
- Github URL: link

EDA with SQL

- Displaying the names of the unique launch sites in the space mission
- Displaying 5 records where launch sites begin with the string ‘CCA'
- Displaying the total payload mass carried by boosters launched by NASA (CRS)
- Displaying average payload mass carried by booster version F9 v1.1
- Listing the date when the first successful landing outcome in ground pad was achieved
- Listing the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- Listing the total number of successful and failure mission outcomes
- Listing the names of the booster versions which have carried the maximum payload mass
- Listing the failed landing outcomes in drone ship, their booster versions and launch site names for the months in year 2015
- Ranking 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: Link

Build an Interactive Map with Folium

- I've created a map with a marker (circle) for NASA's Johnson Space Center, including a popup label and text label displaying its name. This marker's location is based on the center's latitude and longitude coordinates.
- Additionally, I've added markers (circles) for all launch sites using their latitude and longitude coordinates. Each marker has a popup label and text label to identify the specific launch site. This visualization helps me understand the geographical distribution of launch sites and their proximity to the equator and coastlines.
- To analyze launch success rates at different sites, I've colored the launch site markers green for successful launches and red for failed launches. Marker clusters highlight launch sites with a high concentration of either success or failure.
- Finally, to visualize the surroundings of a particular launch site (e.g., KSC LC-39A), I've added colored lines to depict its distances to nearby features like railways, highways, coastlines, and the nearest city.
- Github: [Link]

Build a Dashboard with Plotly Dash

- To allow you to focus on specific launch sites, I've added a dropdown list where you can choose the site you're interested in.
- A pie chart displays the launch success rates. When you select "All Sites," the chart shows the total successful launches across all sites. If you choose a specific launch site from the dropdown, the pie chart breaks down the success and failure rates for that particular site.
- A slider that lets you adjust the payload mass range.
- Finally, a scatter chart explores the relationship between payload mass and launch success rate, differentiated by booster versions. This visualization can help identify if specific booster versions perform better with heavier payloads
- Github: Link

Predictive Analysis (Classification)

- Github: Link

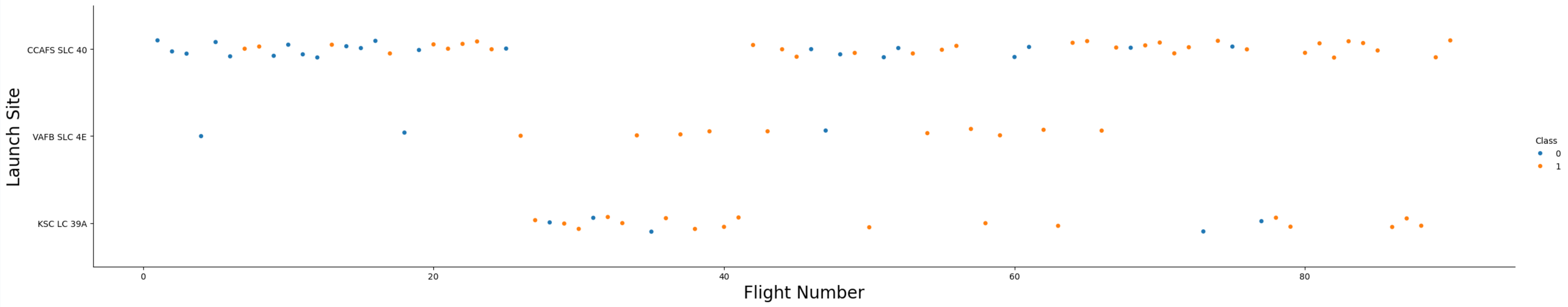
Results

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

Section 2

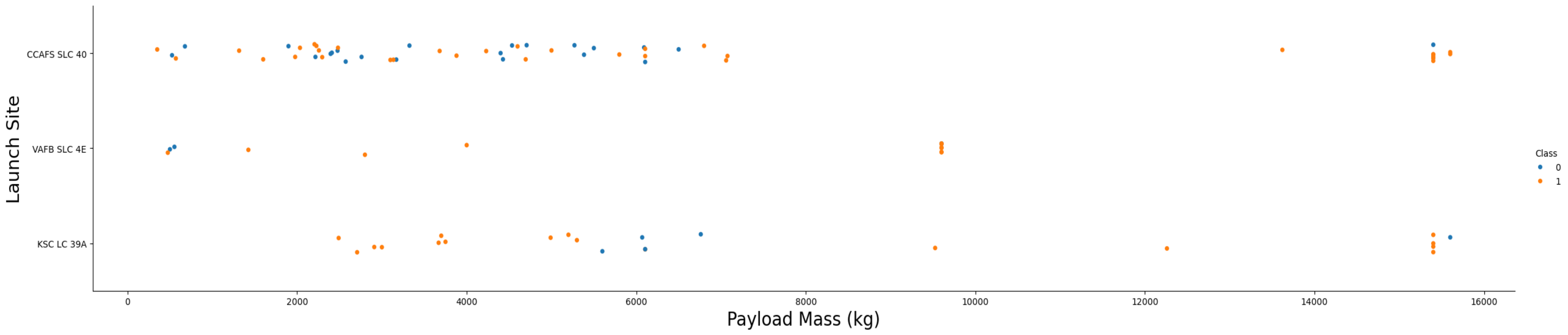
Insights drawn from EDA

Flight Number vs. Launch Site



- Early SpaceX flights faced higher failure rates, suggesting mission improvements over time.
- Cape Canaveral's SLC 40 is the workhorse launch site, handling most missions.
- Notably, VAFB SLC 4E and KSC LC-39A show higher success rates.

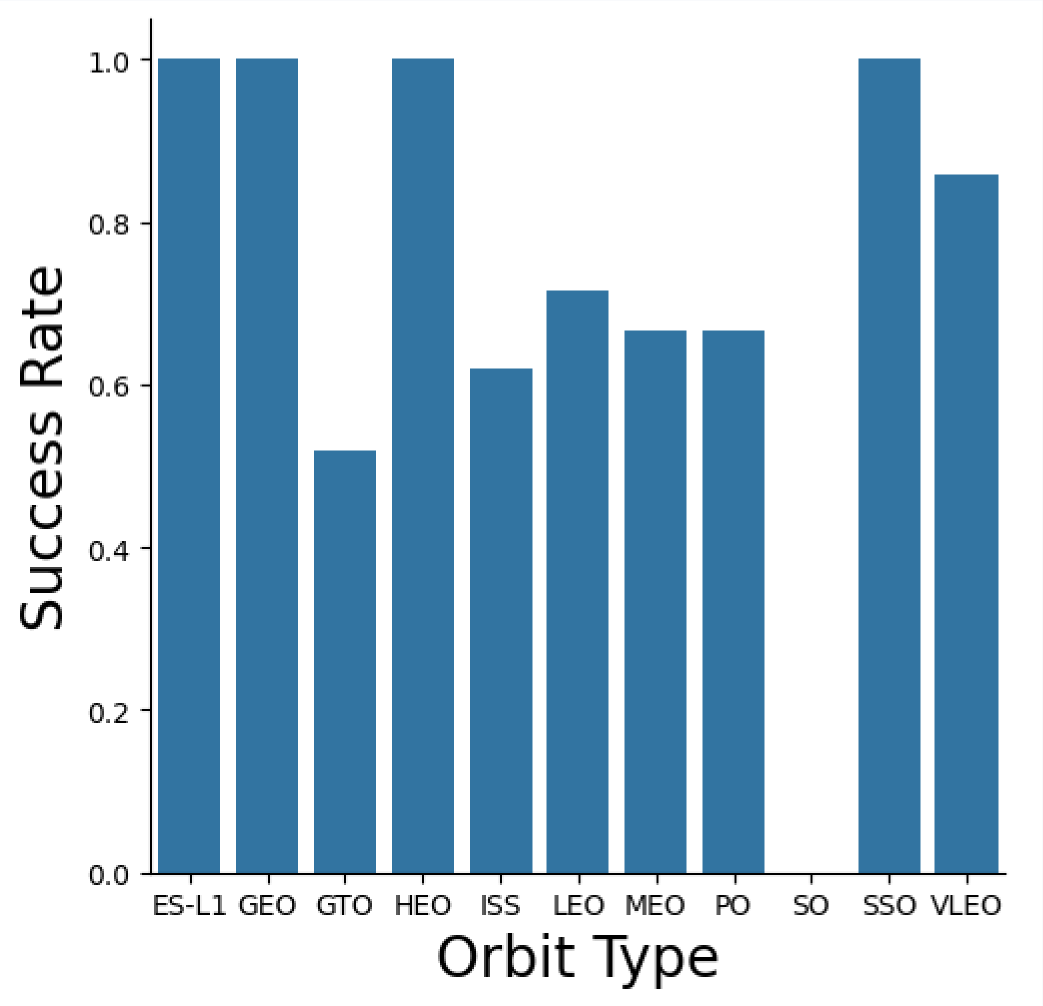
Payload vs. Launch Site



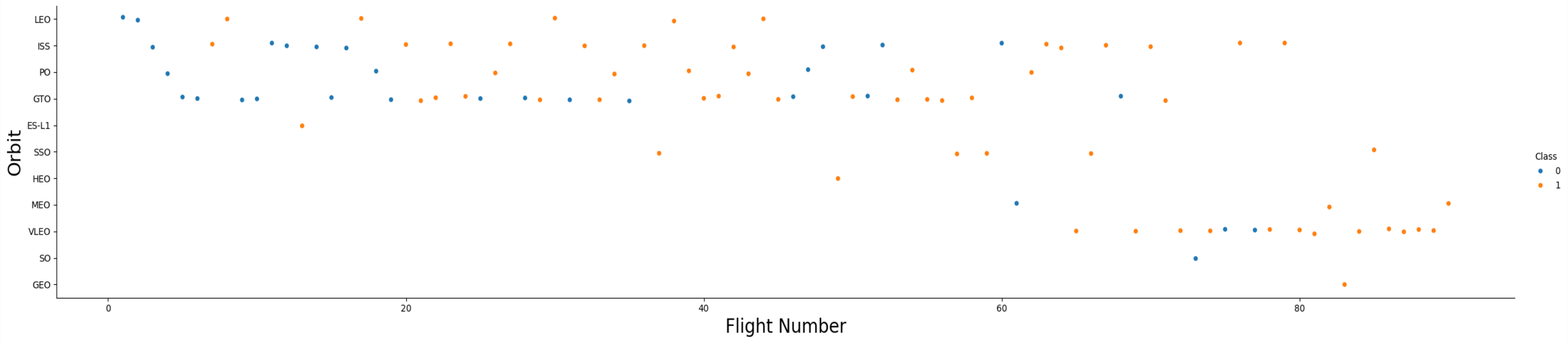
- The data suggests a possible trend where success rates might increase with payload mass across all launch sites.
- Launches exceeding 7,000 kg in payload mass appear to have a higher success rate overall.
- KSC LC-39A seems to boast a perfect success rate for payloads under 5,500 kg. It's important to note that further analysis with more data is needed to confirm these observations.

Success Rate vs. Orbit Type

- 00% success rate: ES-L1, GEO, HEO, SSO
- 0% success rate: SO

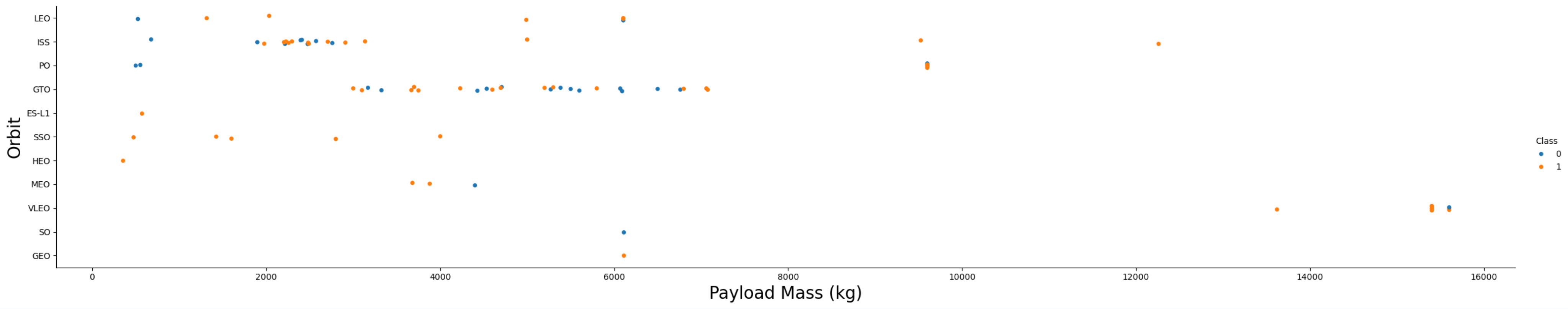


Flight Number vs. Orbit Type



- Success rates in LEO may rise with a booster's flight experience, hinting at improved performance over time.
- GTO missions, however, don't seem to exhibit the same link between flight number and success.

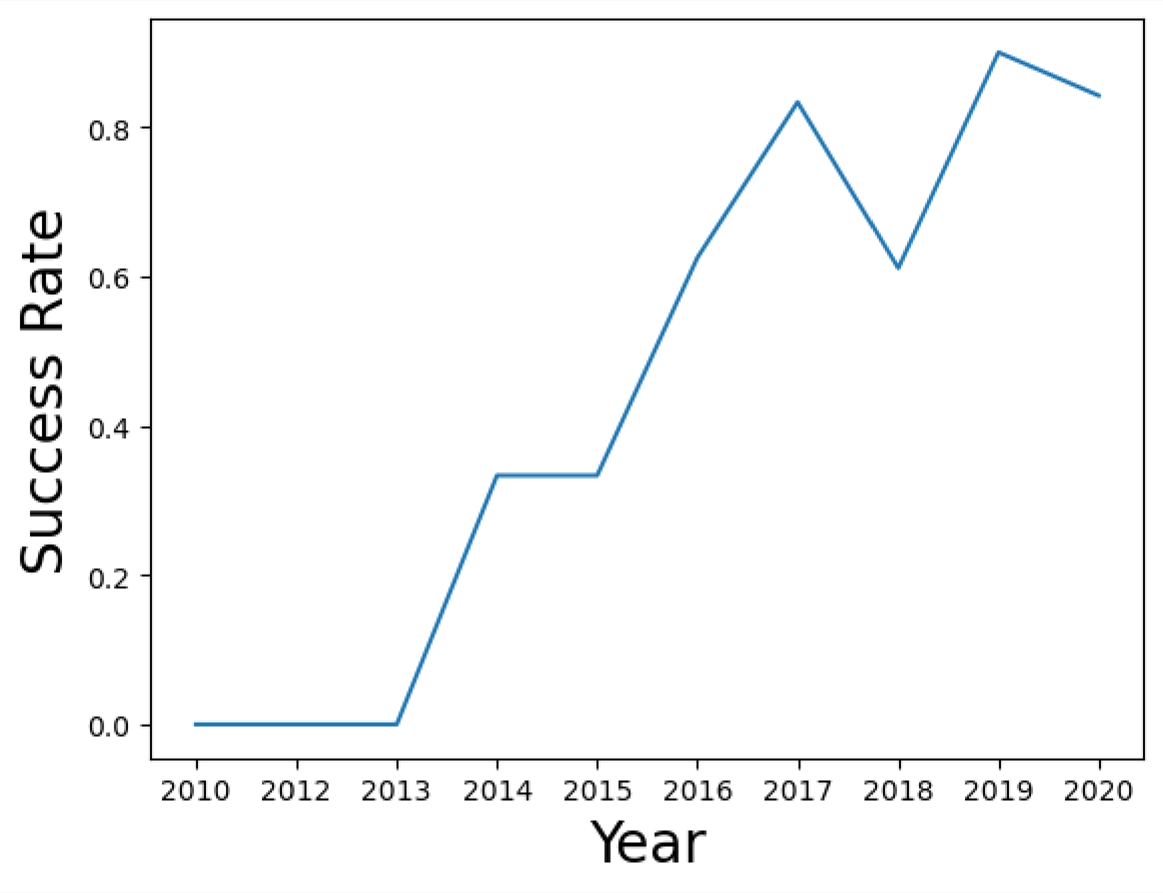
Payload vs. Orbit Type



- Heavier payloads seem to hurt success rates for GTO orbits, but help in delivering cargo to the ISS (GTO & Polar LEO).
- This suggests orbit type plays a role in how weight affects success.

Launch Success Yearly Trend

- The success rate kept increasing since 2013 until 2020



All Launch Site Names

```
%sql select distinct launch_site from SPACEXTBL;
```

```
 * sqlite:///my_data1.db
Done.
```

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

Launch Site Names Begin with 'CCA'

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

```
 * sqlite:///my_data1.db
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	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

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

```
 * sqlite:///my_data1.db
Done.
```

total_payload_mass
45596

Average Payload Mass by F9 v1.1

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

```
 * sqlite:///my_data1.db
Done.
```

average_payload_mass
2534.6666666666665

First Successful Ground Landing Date

```
%sql select min(date) as first_successful_landing from SPACEXTBL where landing_outcome = 'Success (ground pad)';
```

```
 * sqlite:///my_data1.db
Done.
```

first_successful_landing
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select booster_version from SPACEXTBL where landing_outcome = 'Success (drone ship)' and payload_mass__kg_ between 400
```

```
 * sqlite:///my_data1.db
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

```
%sql select mission_outcome, count(*) as total_number from SPACEXTBL group by mission_outcome;
```

```
 * sqlite:///my_data1.db
Done.
```

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

Boosters Carried Maximum Payload

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

```
 * sqlite:///my_data1.db
Done.
```

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

2015 Launch Records

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

- Present your query result with a short explanation here

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

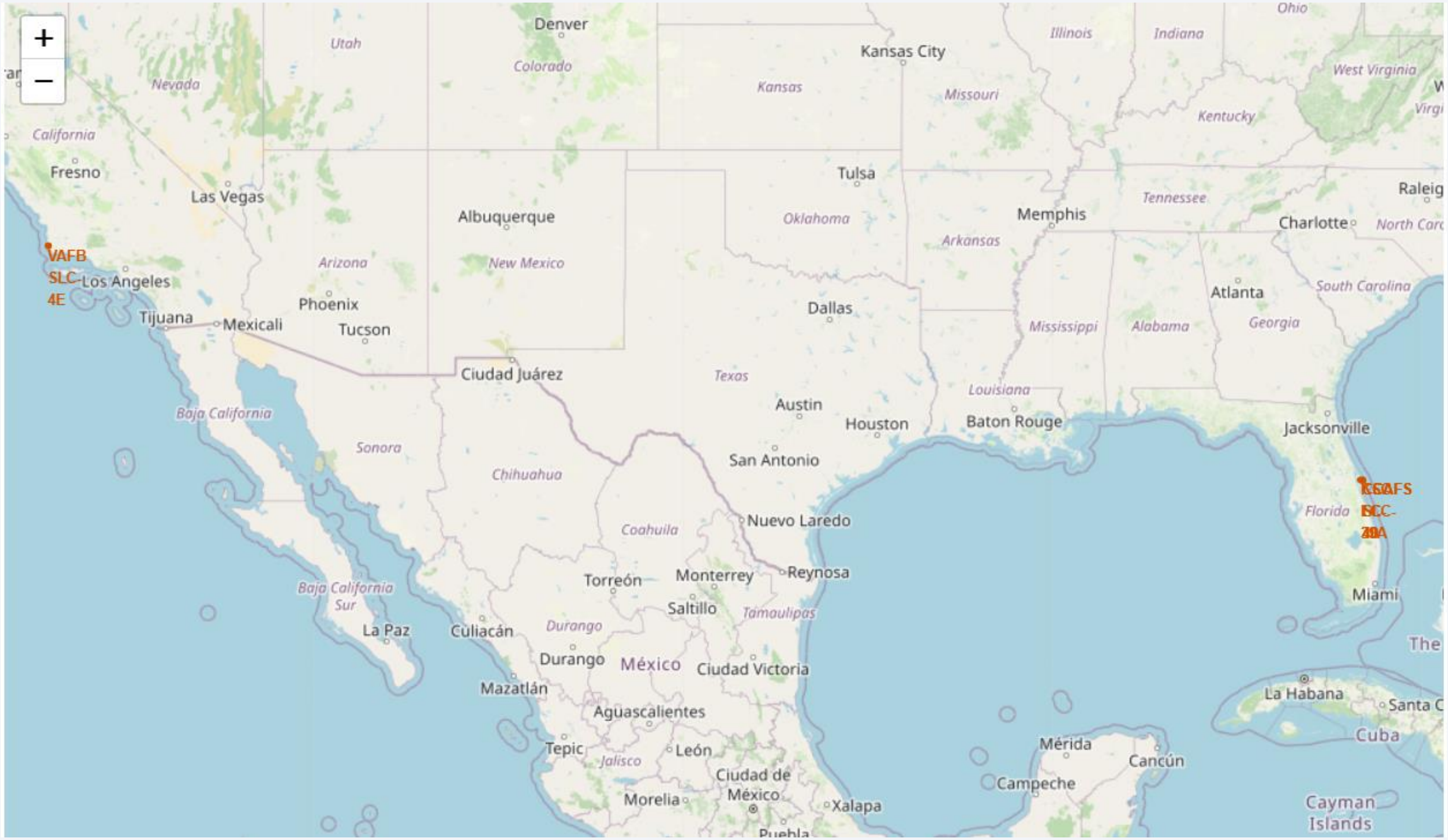
- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

- Present your query result with a short explanation here

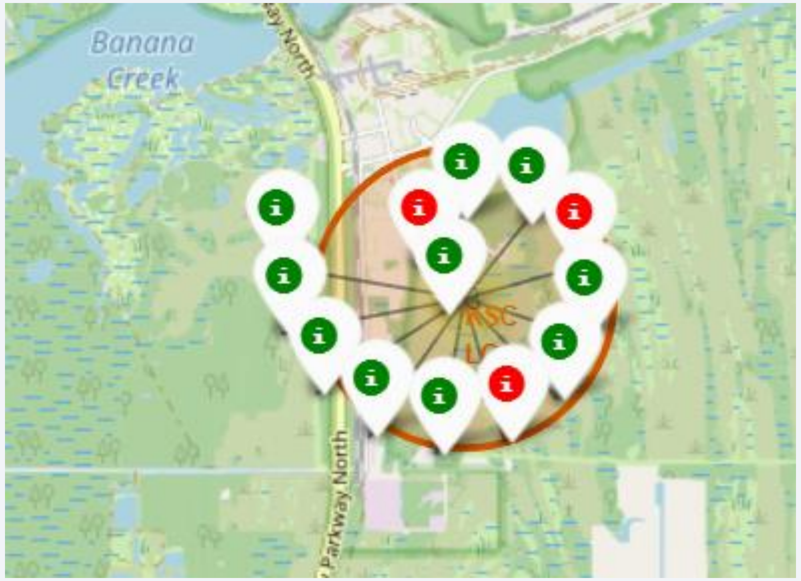
Section 3

Launch Sites Proximities Analysis

Interactive Visual Analytics with Folium



Success/failed launches



<Folium Map Screenshot 3>

- Replace <Folium map screenshot 3> title with an appropriate title

- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed

- Explain the important elements and findings on the screenshot

Section 4

Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title

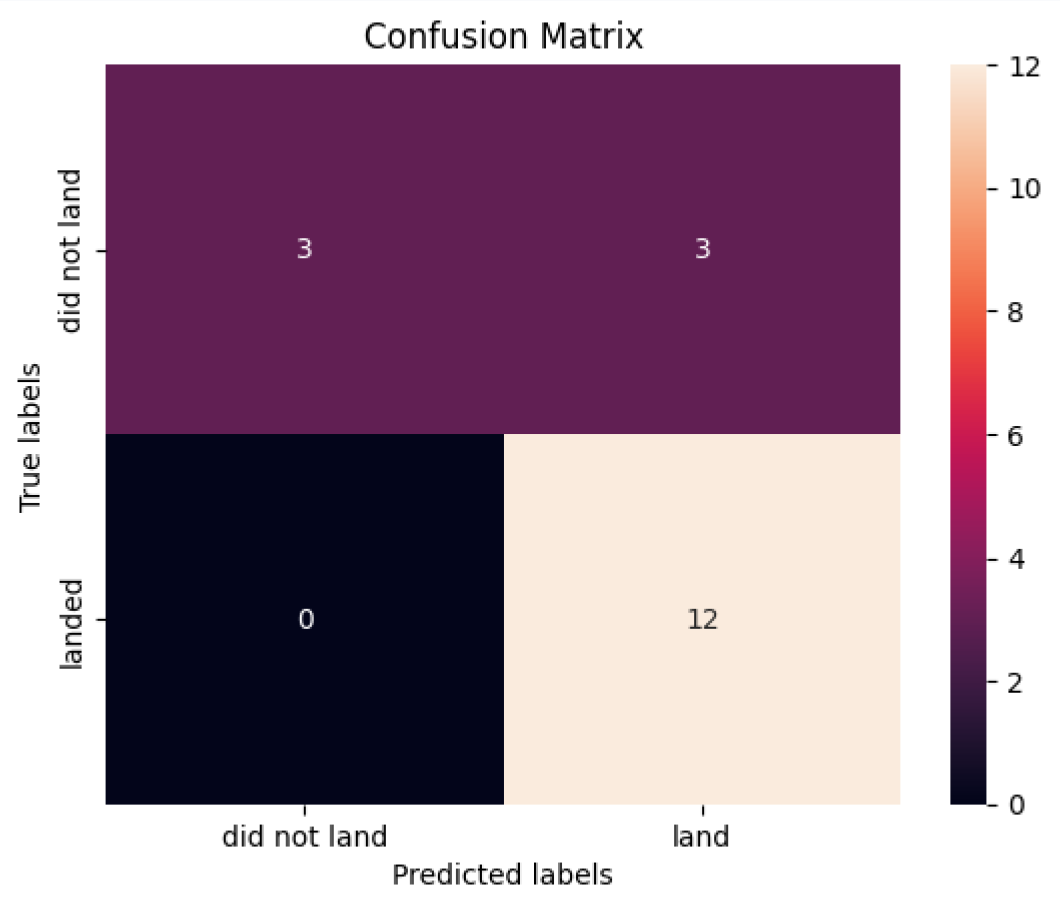
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider

- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

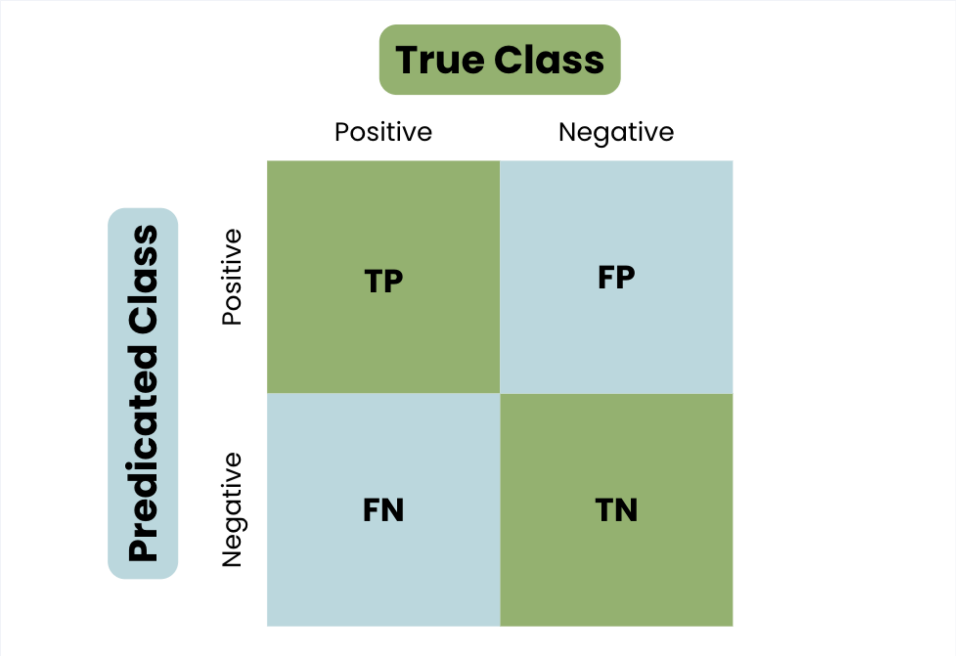
Section 5

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- Our analysis suggests the Decision Tree model may be the most effective algorithm for predicting SpaceX first-stage landing success based on this dataset.
- Launches with lower payload mass tend to have higher success rates compared to heavier payloads.
- Geographically, most launch sites are located near the equator and the coast, potentially offering launch advantages.
- There's an upward trend in the overall success rate of launches over time.
- Among all launch sites, KSC LC-39A boasts the highest success rate.
- Interestingly, orbits like ES-L1, GEO, HEO, and SSO have a perfect success rate in the data analyzed. It's important to note that further investigation might be needed to confirm this finding with a larger dataset.

Thank you!