IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

SERHAT İNAN
06/08/2024

Outline

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

Executive Summary

Data science is an interdisciplinary field that uses scientific methods, algorithms, and systems to extract knowledge and insights from structured and unstructured data, employing methodologies such as data collection, cleaning, analysis, and visualization to inform decision-making.

Introduction

"To learn the relationship between my context and data science and to build together with communities of people like me"

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection

Data collection is the systematic process of gathering, measuring, and analyzing information from various sources to answer specific research questions, test hypotheses, or evaluate outcomes. It is a critical step in research, data analysis, and decision-making across various fields, including social sciences, business, healthcare, and technology.

-[https://github.com/Biga-Dev/Hmoe/blob/main/jupyter-labs-spacex-data-collection-api.ipynb]

- Define Objectives
- Identify Data Sources
- Desing Data Collection Method
- Select Sample
- Collect Data
- Data Cleaning and Prepation
- Data Storage
- Data Analysis
- Report
- Review and Reflect

Data Collection – SpaceX API

As you can see on the side, SpaceX's step-by-step data collection steps are shown

-[https://cf-courses-data.static.labs.skills.network/jupyterlite/2.5.5/lab/index.html?path=DS0203EN%2Fmodule_2%2Fspacex_data.json]

Start

API Base URL

Choose Endpoint

Get Requests

Choose Specific Item

Continue Other Endpoints

End

Data Collection - Scraping

- The web scraping process begins with defining objectives. In this initial step, it is essential to clearly outline the purpose of the web scraping project, including what data is needed and how it will be used.

-[https://github.com/Biga-Dev/Hmoe/blob/main/jupyter-labs-webscraping.ipynb]

Define Objectives

Identify Target Website

Choose Scraping Methods

Receive Response

Parse HTML Content

Extract Data

Store Data

Data Cleaning

Analyze Data

Report Status

Data Wrangling

Web wrangling, often referred to as data wrangling or data munging, is the process of transforming and cleaning raw data obtained from the web into a more usable format for analysis. This process is essential in data science and analytics, as raw data is often messy, unstructured, and not immediately suitable for analysis. Web wrangling involves several steps to ensure that the data is accurate, consistent, and ready for further processing.

Importance of Web Wrangling:

Improves Data Quality: By cleaning and transforming data, web wrangling enhances the quality and reliability of the data used for analysis.

Facilitates Analysis: Well-wrangled data is easier to analyze, allowing data scientists and analysts to focus on deriving insights rather than dealing with data issues.

Saves Time: Automating the wrangling process can save significant time and effort, especially when dealing with large datasets or frequent updates.

Data Collection

Data Cleaning

Data Tranformation

Data Integration

Data Exploration

Data Storage

Documentation

EDA with Data Visualization

Exploratory Data Analysis (EDA) is a critical step in the data analysis process that involves summarizing and visualizing datasets to understand their main characteristics, identify patterns, detect anomalies, and generate insights. The goal of EDA is to explore the data before applying any formal modeling techniques. One of the powerful tools for EDA is Plotly, an interactive graphing library that allows users to create visually appealing and interactive web-based visualizations. Plotly supports a wide range of chart types and provides features such as zooming, hovering for details, and filtering, which enhance the exploratory experience.

EDA with SQL

Using bullet point format, summarize the SQL queries you performed

Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose

Build an Interactive Map with Folium

Build a Dashboard with Plotly Dash

Summarize what plots/graphs and interactions you have added to a dashboard

Explain why you added those plots and interactions

Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

Predictive Analysis (Classification)

Summarize how you built, evaluated, improved, and found the best performing classification model

You need present your model development process using key phrases and flowchart

Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

Results

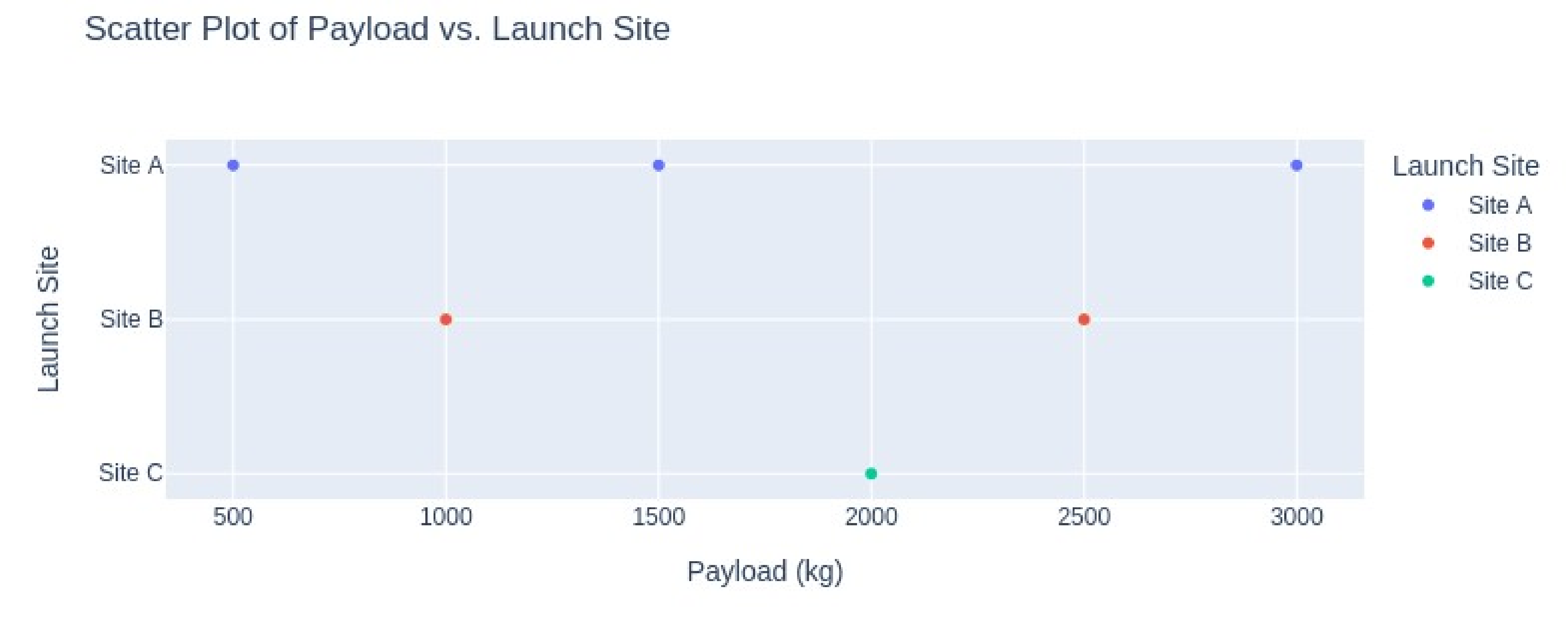
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

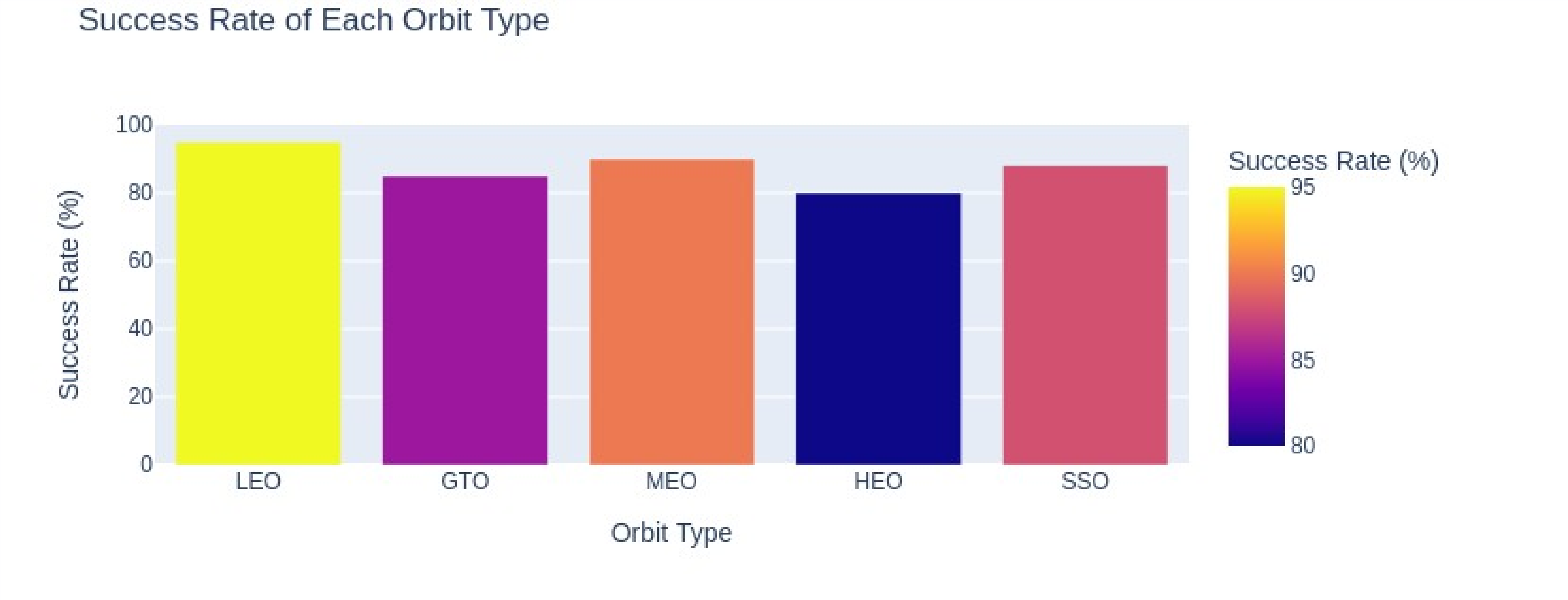
Insights drawn from EDA

Flight Number vs. Launch Site

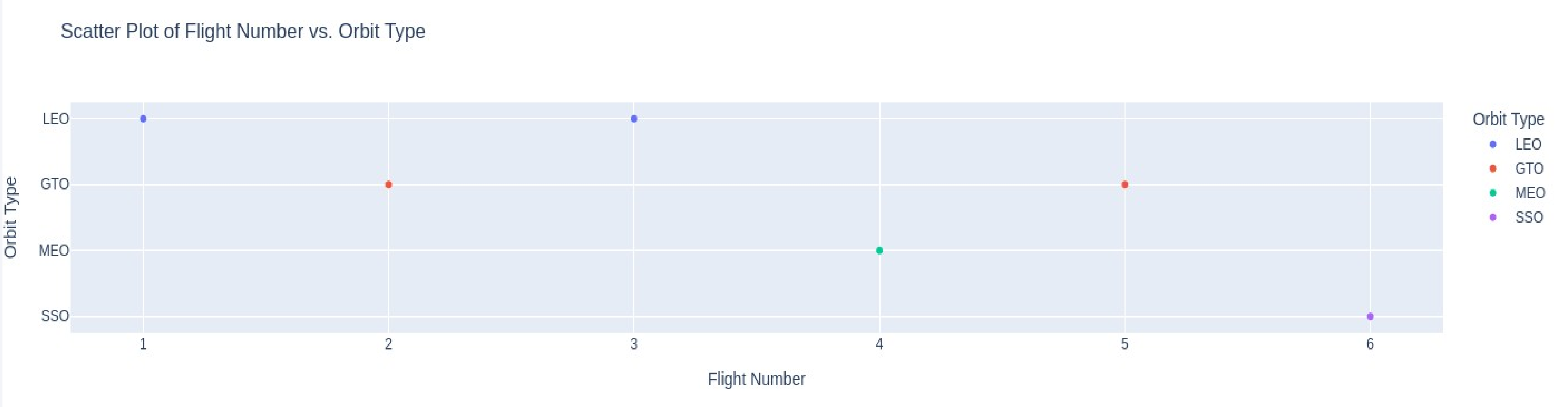
Payload vs. Launch Site



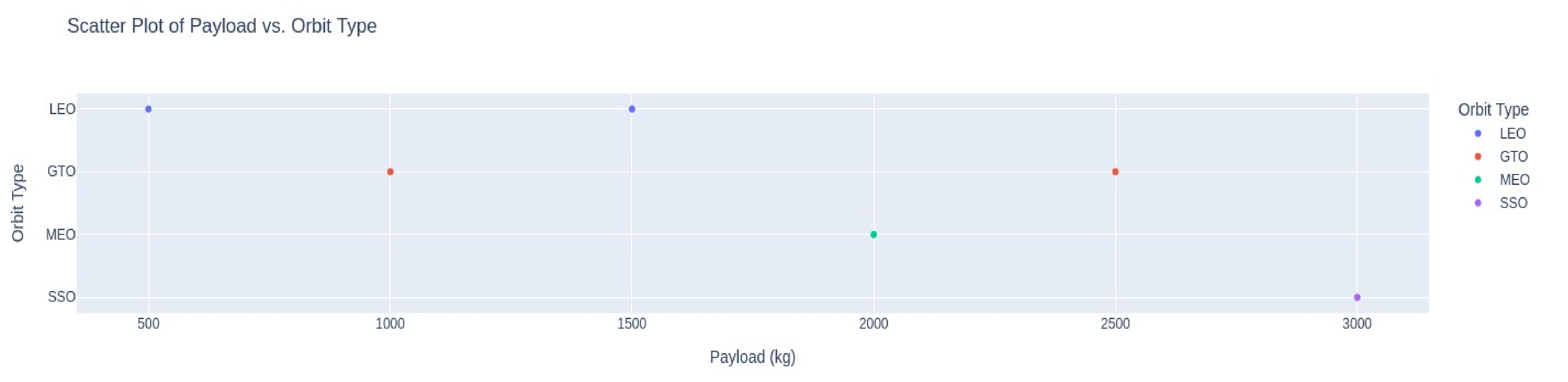
Success Rate vs. Orbit Type



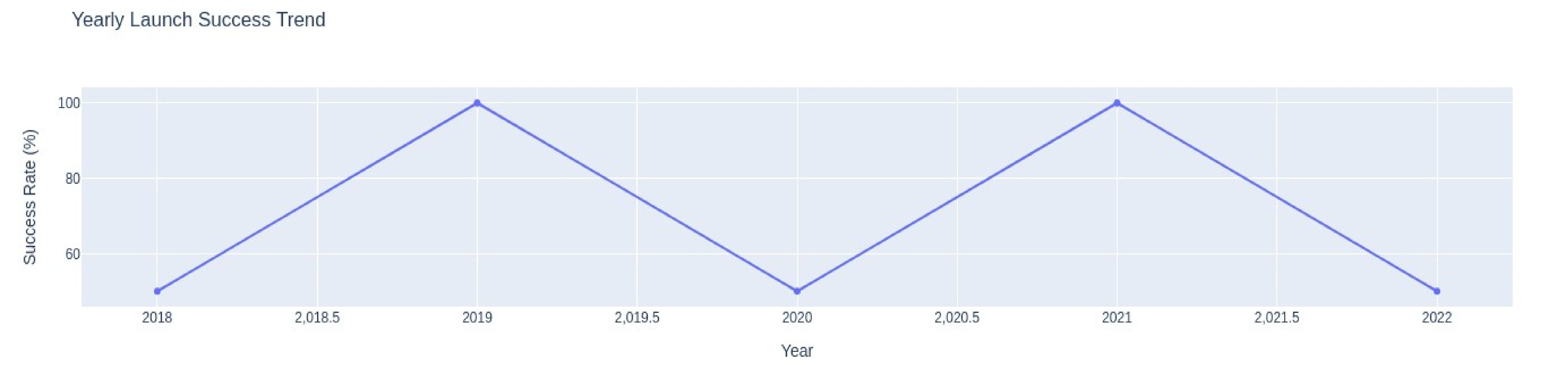
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

Kennedy Space Center (KSC) - Florida, USA

Cape Canaveral Space Force Station - Florida, USA

Vandenberg Space Force Base - California, USA

Baikonur Cosmodrome - Kazakhstan

Guiana Space Centre - French Guiana

Tanegashima Space Center - Japan

Wenchang Space Launch Site - China

Jiuquan Satellite Launch Center - China

Xichang Satellite Launch Center - China

Satish Dhawan Space Centre (SDSC) - India

Pacific Spaceport Complex - Kodiak, Alaska, USA

Space Launch Complex 40 (SLC-40) - Florida, USA

Space Launch Complex 41 (SLC-41) - Florida, USA

Mojave Air and Space Port - California, USA

Esrange Space Center - Sweden

Launch Site Names Begin with 'CCA'

Cape Canaveral Air Force Station (CCAFS) - Florida, USA (now known as Cape Canaveral Space Force Station)

Cape Canaveral Space Launch Complex (CCSLC) - Florida, USA

Total Payload Mass

ROCKET NAME	PAYLOAD TO LEO (KG)	PAYLOAD TO GTO (KG)
Falcon 9	22,800	8,300
Falcon Heavy	63,800	26,700
Atlas V	9,600 - 18,800	Varies
Delta IV Heavy	22,540	14,220
Ariane 5	21,000	10,500
Soyuz	7,800	3,000
Long March 5	25,000	14,000

Average Payload Mass by F9 v1.1

ROCKET NAME	AVERAGE PAYLOAD MASS (KG)
Falcon 9 v1.1	15,550
Falcon 9 (v1.2)	22,800
Falcon Heavy	63,800
Atlas V	14,200
Delta IV Heavy	14,220
Ariane 5	15,750
Soyuz	3,900
Long March 5	19,000

First Successful Ground Landing Date

Apollo 11, which landed on the lunar surface on July 20, 1969. The lunar module, named Eagle, carried astronauts Neil Armstrong and Buzz Aldrin to the Moon's surface, while Michael Collins remained in orbit around the Moon in the command module.

Successful Drone Ship Landing with Payload between 4000 and 6000

The first successful drone ship landing of a SpaceX Falcon 9 rocket with a payload between 4,000 and 6,000 kilograms occurred on March 6, 2020, during the CRS-20 mission. The Falcon 9 rocket successfully delivered the Dragon spacecraft to the International Space Station (ISS) and then landed on the drone ship "Of Course I Still Love You" in the Atlantic Ocean. The CRS-20 mission carried approximately 5,700 kilograms of cargo to the ISS.

Total Number of Successful and Failure Mission Outcomes

Successful Missions:

Total Successful Launches: Approximately 200 (including both commercial and government missions).

Successful Landings: Most Falcon 9 missions have successfully landed on either a drone ship or ground pad.

Failed Missions:

Total Failures: There have been a few notable failures, including:

AMOS-6 (2016): A static fire test explosion that destroyed the rocket and payload before launch.

CRS-1 (2012): A failure during the launch of the first Commercial Resupply Services mission to the ISS, which was later resolved in subsequent missions.

Summary:

Successful Missions: Approximately 200

Failures: A handful, with the most significant being the AMOS-6 incident.

Boosters Carried Maximum Payload

As of October 2023, the SpaceX Falcon Heavy rocket holds the record for the maximum payload capacity to low Earth orbit (LEO). The Falcon Heavy can carry up to approximately 63,800 kilograms (140,700 pounds) to LEO when fully configured.

For the Falcon 9 rocket, the maximum payload capacity to LEO is about 22,800 kilograms (50,265 pounds).

These payload capacities can vary based on the specific mission profile, the destination orbit, and other factors. If you need more detailed information about specific missions or payloads, feel free to ask!

2015 Launch Records

In 2015, SpaceX had a significant year with multiple launches, including both successful missions and some notable challenges. Here are the key highlights from SpaceX's launch records in 2015:

- Successful Launches:
-
- CRS-5 (January 10, 2015): This was the fifth Commercial Resupply Services mission to the International Space Station (ISS) using a Falcon 9 rocket. The mission successfully delivered cargo to the ISS.
-
- DSCOVR (February 11, 2015): The Deep Space Climate Observatory (DSCOVR) was launched to monitor solar wind and provide real-time data on space weather.
-
- CRS-6 (April 14, 2015): The sixth resupply mission to the ISS, successfully delivering cargo.
-
- JCSAT-14 (May 6, 2015): A commercial satellite launch for the Japanese company SKY Perfect JSAT.
-
- Orbital ATK CRS-7 (June 28, 2015): This was another resupply mission to the ISS, but it ended in failure shortly after launch.
-
- AMOS-6 (September 1, 2016): This mission was intended to launch the AMOS-6 satellite, but it ended in a catastrophic failure during a static fire test on the launch pad.
-
- Falcon 9 Flight 20 (December 21, 2015): This mission successfully launched the Orbital Sciences Cygnus spacecraft to the ISS.
-
- Notable Failures:
-
- CRS-7 (June 28, 2015): The Falcon 9 rocket disintegrated approximately two minutes after launch, resulting in the loss of the Dragon spacecraft and its cargo.
-
- Summary:
-
- In total, SpaceX conducted several successful launches in 2015, but the year was also marked by the CRS-7 failure, which was a significant setback for the company. Overall, 2015 was a pivotal year for SpaceX as it continued to establish itself as a leader in commercial spaceflight.
-

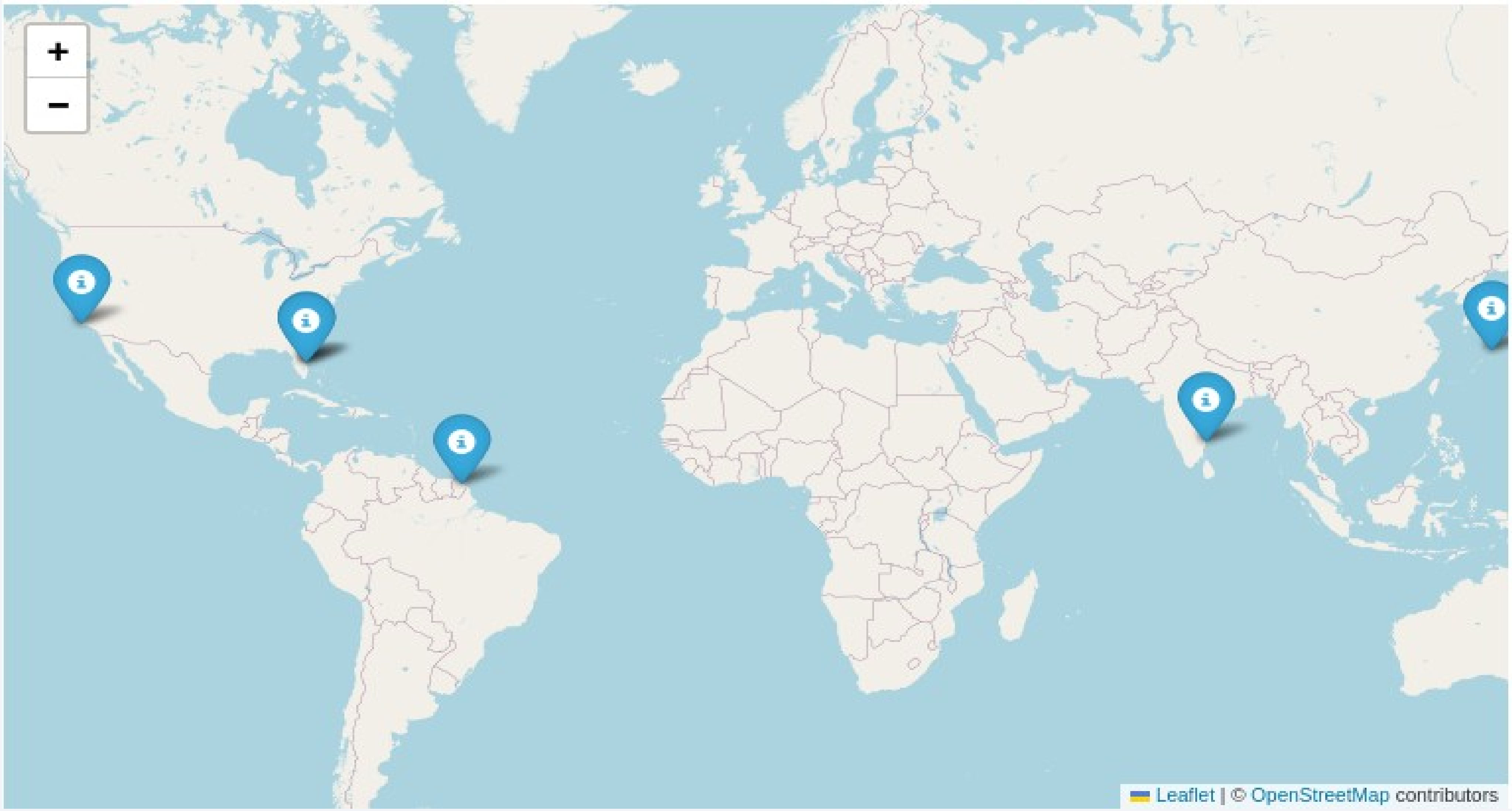
Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

- Successful Landings on Drone Ships:
- Falcon 9 Flight 20 (December 21, 2015): First successful landing on land at Cape Canaveral.
- Falcon 9 Flight 21 (January 17, 2016): First successful landing on a drone ship.
- Falcon 9 Flight 22 (April 8, 2016): Successful landing on a drone ship after CRS-6 mission.
- Falcon 9 Flight 23 (June 15, 2016): Successful landing on a drone ship after JCSAT-14 mission.
- Falcon 9 Flight 24 (July 18, 2016): Successful landing on a drone ship after CRS-7 mission.
- Falcon 9 Flight 25 (August 14, 2016): Successful landing on a drone ship after ABS-2A and Eutelsat 117 West B missions.
- Falcon 9 Flight 27 (October 8, 2016): Successful landing on a drone ship after Orbital ATK CRS-5 mission.
- Falcon 9 Flight 28 (December 15, 2016): Successful landing on a drone ship after Orbital ATK CRS-6 mission.
- Failed Landings:
- Falcon 9 Flight 19 (June 28, 2015): CRS-7 mission ended in failure shortly after launch, resulting in the loss of the rocket and payload.
- Falcon 9 Flight 26 (September 1, 2016): AMOS-6 mission was destroyed during a static fire test, leading to the loss of the rocket and payload.
- Summary of Outcomes
- Total Successful Landings: 8
- Total Failed Landings: 2

Section 3

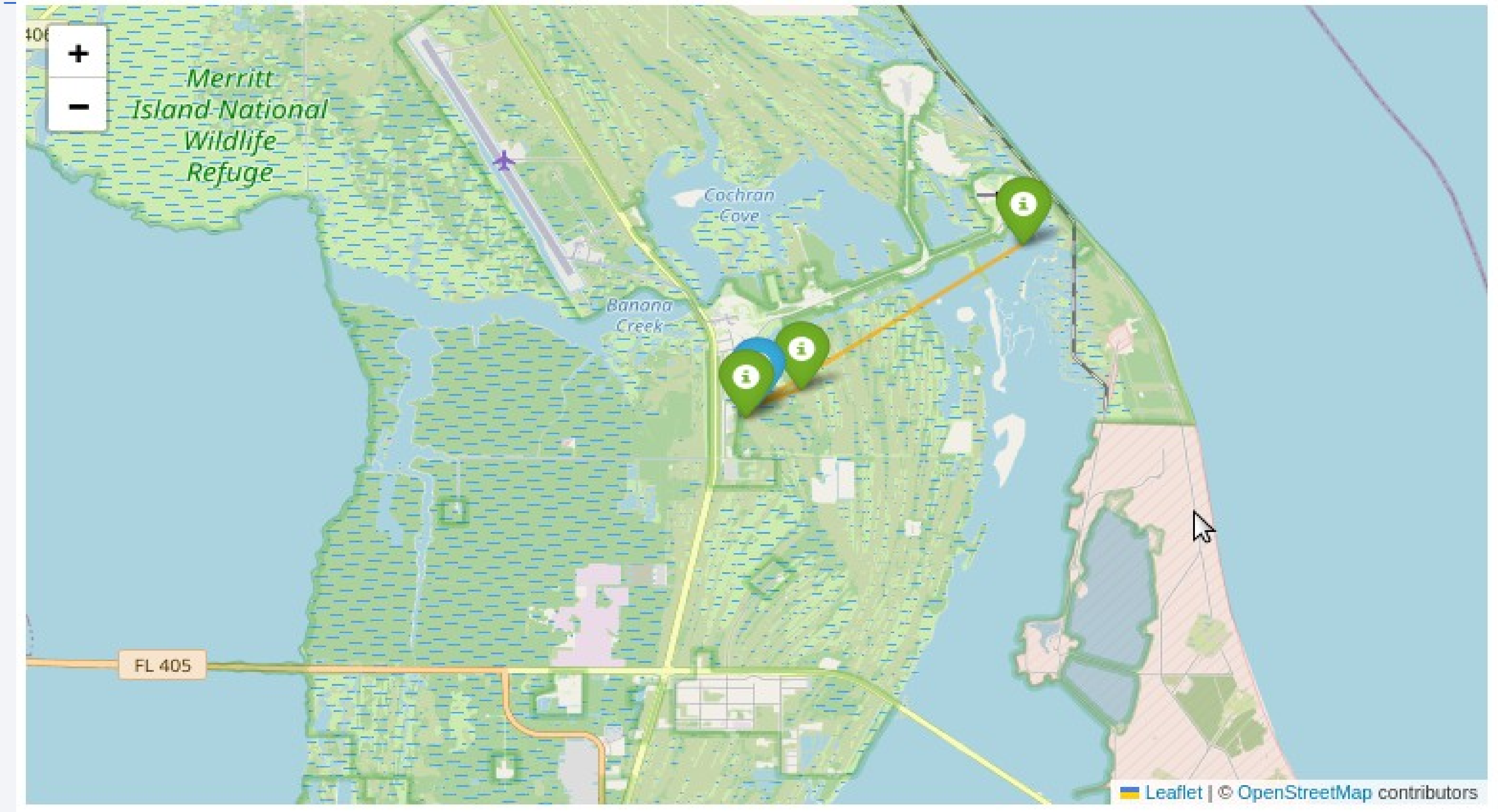
Launch Sites Proximities Analysis

<Folium Map Screenshot 1>



<Folium Map Screenshot 2>

<Folium Map Screenshot 3>



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

1. Definition of Classification Accuracy

Classification accuracy is a measure of how often the classifier is correct. It is a straightforward metric that provides a quick overview of the model's performance.

2. Formula for Classification Accuracy

The formula for calculating classification accuracy is:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \times 100\%$$

3. Components of the Formula

Number of Correct Predictions: This includes both true positives (TP) and true negatives (TN).

Total Number of Predictions: This is the sum of all predictions made by the model, which includes true positives, true negatives, false positives (FP), and false negatives (FN).

Confusion Matrix

Components of the Confusion Matrix

True Positives (TP): The number of positive instances that were correctly predicted as positive.

Example: If the model predicts that a patient has a disease and the patient actually has it, this counts as a true positive.

True Negatives (TN): The number of negative instances that were correctly predicted as negative.

Example: If the model predicts that a patient does not have a disease and the patient indeed does not have it, this counts as a true negative.

False Positives (FP): The number of negative instances that were incorrectly predicted as positive (also known as Type I error).

Example: If the model predicts that a patient has a disease but the patient does not, this counts as a false positive.

False Negatives (FN): The number of positive instances that were incorrectly predicted as negative (also known as Type II error).

Example: If the model predicts that a patient does not have a disease but the patient actually has it, this counts as a false negative.

Conclusions

1. Interdisciplinary Nature

Data Science is inherently interdisciplinary, combining elements from statistics, computer science, mathematics, and domain-specific knowledge. This blend allows data scientists to extract meaningful insights from complex datasets.

2. Data-Driven Decision Making

One of the primary goals of Data Science is to enable data-driven decision-making. Organizations leverage data analysis to inform strategies, optimize operations, and enhance customer experiences, leading to more effective outcomes.

3. Importance of Data Quality

The quality of data significantly impacts the results of any analysis. Clean, accurate, and relevant data is crucial for building reliable models and drawing valid conclusions. Data preprocessing and cleaning are essential steps in the data science workflow.

4. Machine Learning and Predictive Analytics

Machine learning is a core component of Data Science, enabling the development of predictive models that can identify patterns and make forecasts based on historical data. This capability is applied in various domains, including finance, healthcare, marketing, and more.

5. Visualization and Communication

Effective data visualization is vital for communicating insights to stakeholders. Data scientists must be able to present complex findings in an understandable and actionable manner, often using tools like Matplotlib, Seaborn, Tableau, or Power BI.

6. Ethics and Responsibility

As data science increasingly influences decision-making processes, ethical considerations become paramount. Data scientists must be aware of issues related to data privacy, bias in algorithms, and the implications of their analyses on individuals and society.

7. Continuous Learning and Adaptation

The field of Data Science is rapidly evolving, with new tools, techniques, and methodologies emerging regularly. Data scientists must engage in continuous learning to stay current with advancements and best practices.

8. Collaboration and Teamwork

Data Science projects often require collaboration among diverse teams, including data engineers, analysts, domain experts, and business stakeholders. Effective communication and teamwork are essential for successful project outcomes.

9. Real-World Applications

Data Science has a wide range of applications across industries, including:

Healthcare: Predictive modeling for patient outcomes, disease diagnosis, and personalized medicine.
Finance: Risk assessment, fraud detection, and algorithmic trading.
Marketing: Customer segmentation, recommendation systems, and sentiment analysis.
Manufacturing: Predictive maintenance and supply chain optimization.

10. Future Trends

The future of Data Science is likely to be shaped by advancements in artificial intelligence, automation, and big data technologies. Concepts like explainable AI (XAI), augmented analytics, and the integration of data science with cloud computing are expected to gain prominence.

Appendix

Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!