IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Name>
<Date>

Outline

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

Executive Summary

Methods :

- Data Collection
- Data Wrangling
- EDA
- Predictive Modeling
- Dashboard Creation
- Mapping

Results :

- Launch Site Analysis
- Payload and Outcome Correlations
- Orbit-Specific Success Trends
- Trends Over Time
- Predictive Modelong
- Dashboard Insights
- Mapping Results

Introduction

1. Background

- SpaceX is a leading private aerospace company revolutionizing space exploration and satellite launches.
- With the goal of reducing the cost of space travel and increasing mission efficiency, SpaceX regularly launches satellites, cargo missions, and payloads into various orbits.
- A growing dataset of SpaceX missions provides valuable insights into mission outcomes, success rates, and operational efficiency.

2. Purpose of the Project

- To analyze SpaceX's historical launch data using data science techniques to identify patterns, trends, and key factors affecting mission outcomes.
- To create predictive models that can determine the likelihood of mission success under different configurations, such as payload, orbit type, and launch site.
- To build interactive dashboards and maps for visualizing launch performance, providing actionable insights for SpaceX's stakeholders.

3. Tools and Techniques Used

- **Data Collection**: API calls, web scraping, and dataset extraction.
- **Analysis**: SQL for querying data, Pandas for manipulation, and visualizations using Matplotlib, Seaborn, and Plotly.
- **Predictive Modeling**: Machine learning models like Logistic Regression, SVM, and Decision Trees.
- **Visualization**: Folium for interactive maps and Dash for real-time data dashboards.

Problems You Want to Find Answers To

1. **Launch Success Rates**:
- What are the success rates of launches across different sites and orbits?
- How do launch sites impact mission success?

2. **Payload and Mission Outcome**:
- Is there a correlation between payload mass and mission success?
- What payload ranges are optimal for maximizing success rates?

3. **Orbit-Specific Analysis**:
- Which orbits are associated with the highest success rates?
- Are there orbits that consistently face challenges or higher failure rates?

4. **Trends Over Time**:
- How have SpaceX's success rates evolved over the years?
- What technological or operational advancements contributed to improvements?

5. **Predictive Insights**:
- What factors (e.g., launch site, payload, orbit type) are the most important predictors of mission success?
- Can we build a reliable predictive model for future mission success?

6. **Geographical Analysis**:
- How do the locations of launch sites contribute to mission planning and trajectory optimization?
- What is the relationship between launch sites and nearby coastlines for safety?

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX API and web scraping
- Perform data wrangling
  - Removing duplicate, handle missing value, formatting columns consistency
- 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

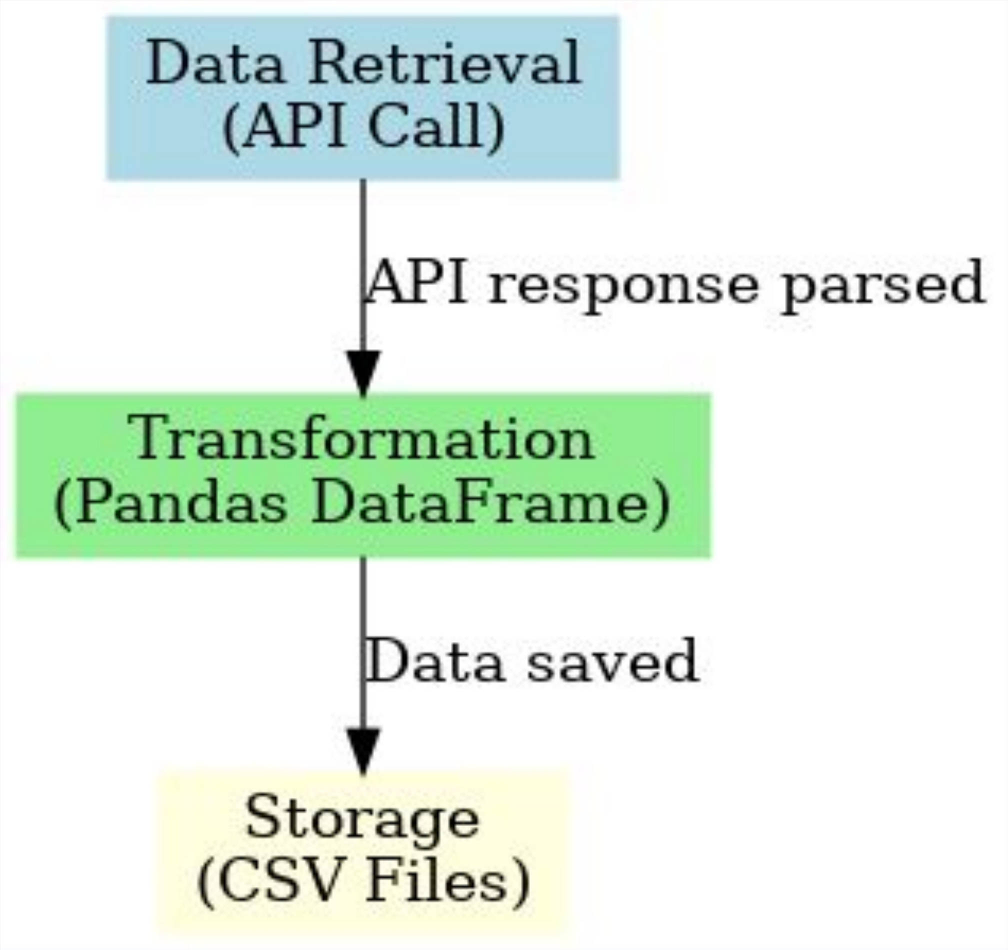
Data Collection

Sources :

- SpaceX API
- Web scraping

Types of data :

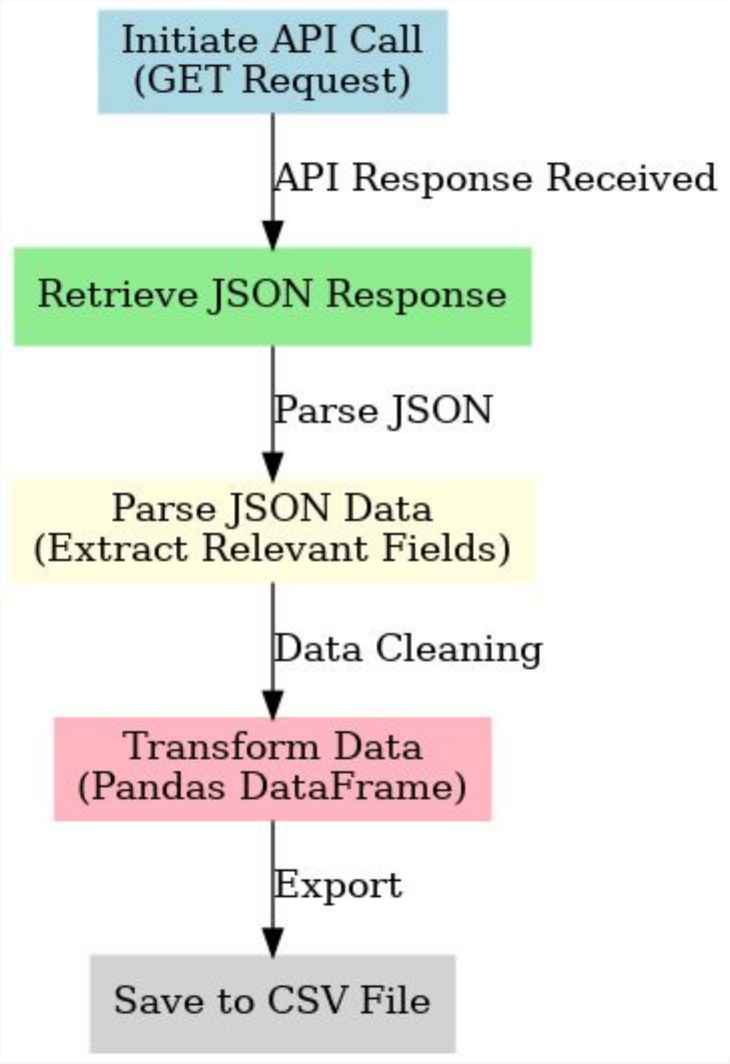
- Launch details
- Mission outcomes



Data Collection – SpaceX API

- Detailed informations about past rocket launches
- Payload details, launch outcomes and more

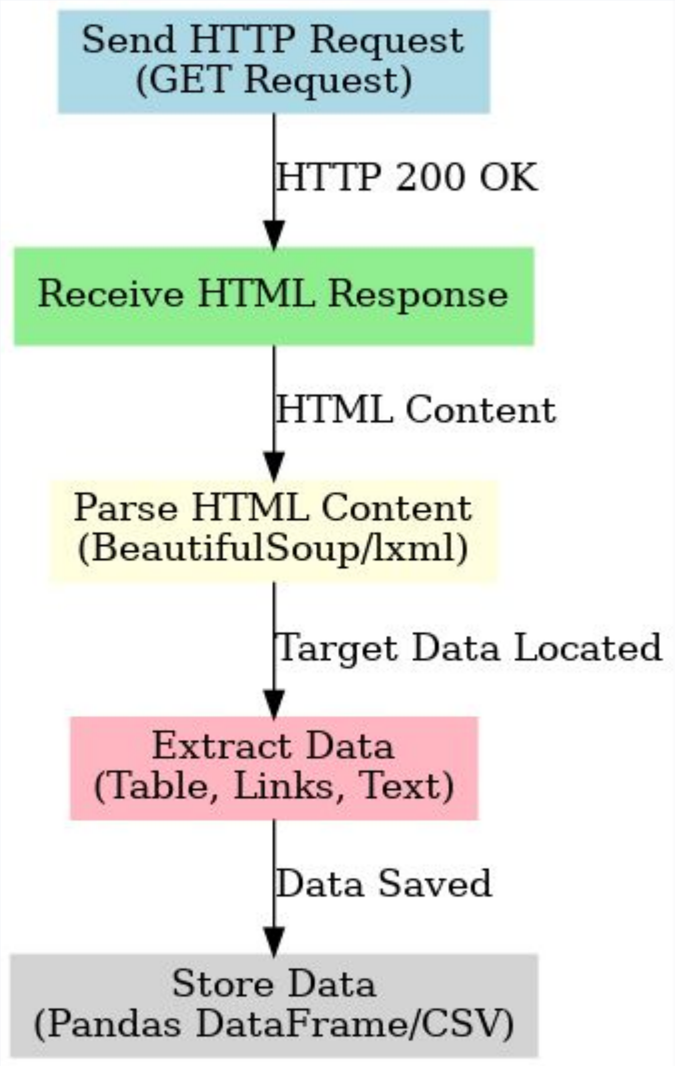
- https://github.com/Rhodham96/AppliedDataScienceCapstone/blob/main/Lab1.1_CollectingData.ipynb



Data Collection - Scraping

- Send HTTP Request
- Receive HTML Response
- Parse HTML content
- Extract Relevant Data
- Save Data

- https://github.com/Rhodham96/AppliedDataScienceCapstone/blob/main/Lab1.2_WebScrapping.ipynb



Data Wrangling

- Handling missing data
- removing duplicates
- Transformations

```
FlightNumber       int64
Date              object
File display ersion    object
PayloadMass      float64
Orbit             object
LaunchSite        object
Outcome           object
Flights            int64
GridFins            bool
Reused              bool
Legs                bool
LandingPad        object
Block            float64
ReusedCount        int64
Serial            object
Longitude        float64
Latitude         float64
dtype: object
```

```
FlightNumber       0.000000
Date               0.000000
BoosterVersion     0.000000
PayloadMass        0.000000
Orbit              0.000000
LaunchSite         0.000000
Outcome            0.000000
Flights            0.000000
GridFins           0.000000
Reused             0.000000
Legs               0.000000
LandingPad        28.888889
Block              0.000000
ReusedCount        0.000000
Serial             0.000000
Longitude          0.000000
Latitude           0.000000
dtype: float64
```

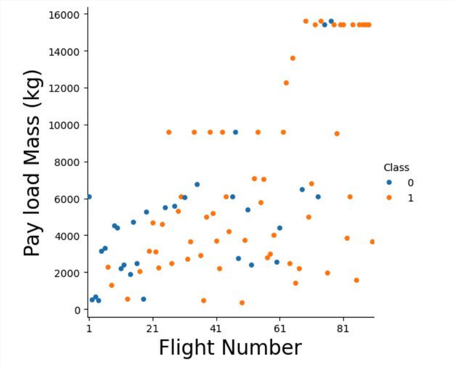
ut[3]:

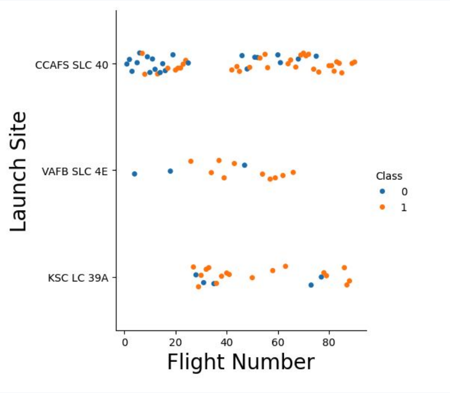
	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPac
	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN
5	6	2014-01-06	Falcon 9	3325.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN
6	7	2014-04-18	Falcon 9	2296.000000	ISS	CCAFS SLC 40	True Ocean	1	False	False	True	NaN
7	8	2014-07-14	Falcon 9	1316.000000	LEO	CCAFS SLC 40	True Ocean	1	False	False	True	NaN
8	9	2014-08-05	Falcon 9	4535.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN
9	10	2014-09-07	Falcon 9	4428.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN

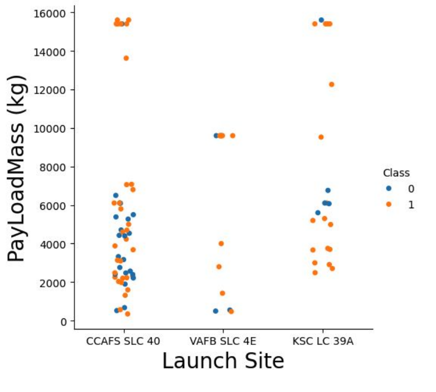
- https://github.com/Rhodham96/AppliedDataScienceCapstone/blob/main/Lab2_DataWrangling.ipynb

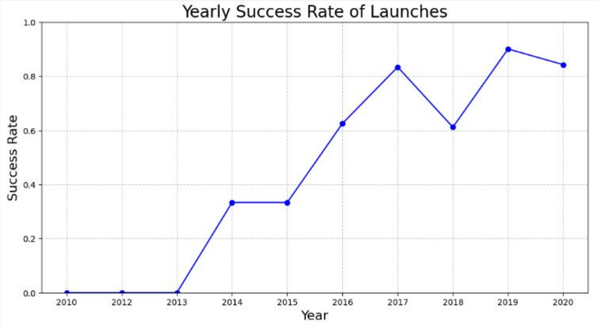
EDA with Data Visualization

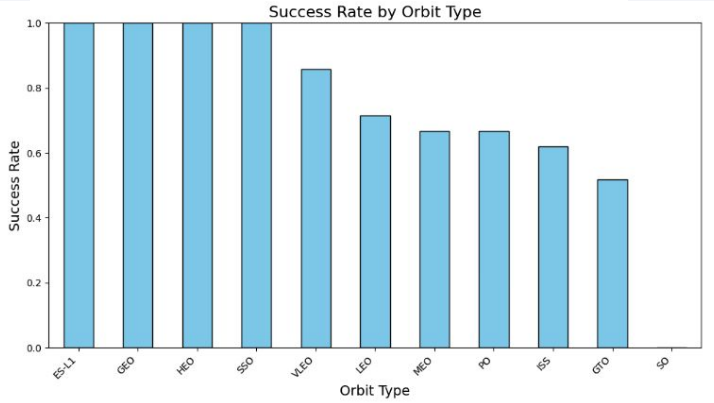
- Scatter Plot : Flight Number vs. Launch Site, Flight Number vs. Payload Mass, Launch Site vs. Payload Mass
- Bar Chart : Success rates across orbit types
- Line Chart : Success rates over the years











- https://github.com/Rhodham96/AppliedDataScienceCapstone/blob/main/Lab4_ExplorePrepare.ipynb

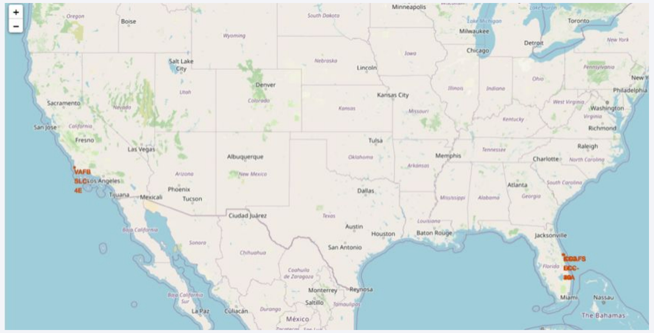
EDA with SQL

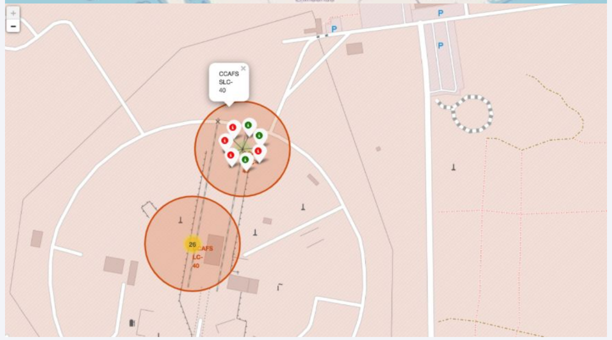
- `SELECT DISTINCT launch_site FROM SPACEXTBL;`
- `SELECT * FROM SPACEXTBL WHERE launch_site LIKE 'CCA%' LIMIT 5;`
- `SELECT SUM(PAYLOAD_MASS__KG_) AS total_payload_mass FROM SPACEXTBL WHERE Customer = 'NASA (CRS)';`
- `SELECT AVG(PAYLOAD_MASS__KG_) AS total_payload_mass FROM SPACEXTBL WHERE Booster_Version LIKE 'F9 v1.1%';`
- `SELECT MIN(Date) FROM SPACEXTBL WHERE Landing_Outcome = 'Success (ground pad)';`
- `SELECT Booster_version FROM SPACEXTBL WHERE landing_outcome = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;`
- `SELECT landing_outcome, COUNT(*) AS total FROM SPACEXTBL GROUP BY landing_outcome;`
- `SELECT landing_outcome, COUNT(*) AS total FROM SPACEXTBL WHERE landing_outcome LIKE 'Success%' OR landing_outcome LIKE 'Failure%' GROUP BY landing_outcome;`
- `SELECT Booster_version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL) ;`

- https://github.com/Rhodham96/AppliedDataScienceCapstone/blob/main/Lab3_SQL.ipynb

Build an Interactive Map with Folium

- All launch sites on a map
- Add success and failed launches markers
- Calculate distance between launch site and its proximities





- https://github.com/Rhodham96/AppliedDataScienceCapstone/blob/main/Lab5_InteractiveVisuals.ipynb

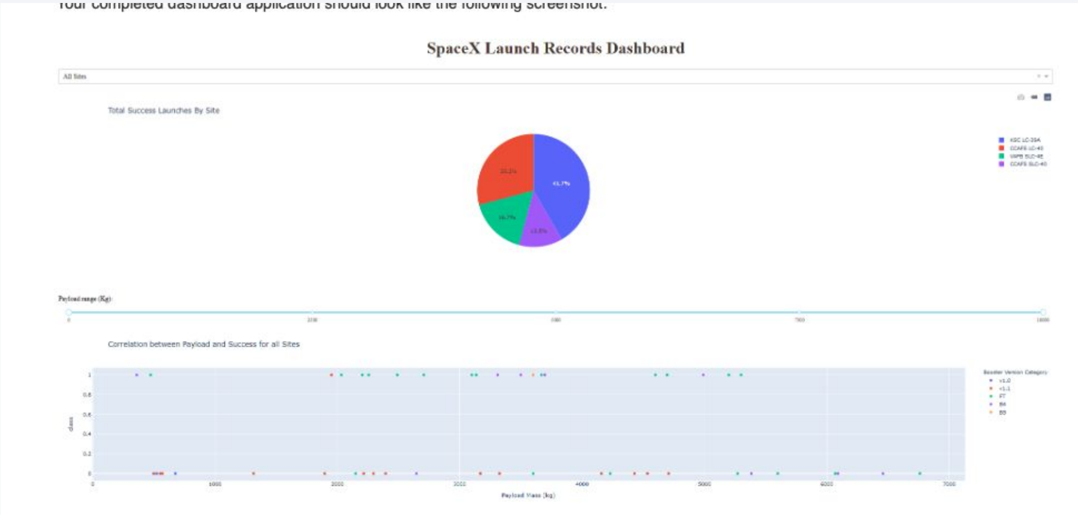
Build a Dashboard with Plotly Dash

- Total launch success counts by site (pie chart)

  (Sites like KSC LC-39A and CCAFS LC-40 have higher success rate due to more frequent launches)

- Correlation between Payload and launch outcome (Scatter Plot)

  (For lighter payloads (<5000kg), success is more consistent, For heavier payloads (>10000kg), success drop slightly, Certain booster versions perform better for heavier payloads.)

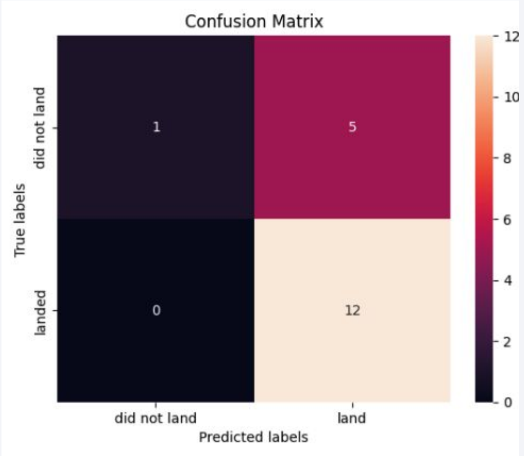


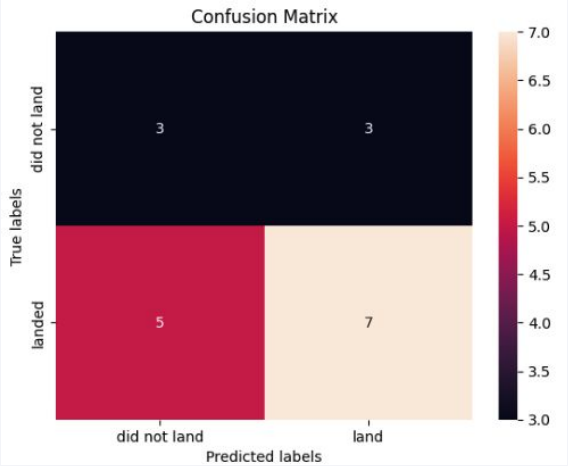
- https://github.com/Rhodham96/AppliedDataScienceCapstone/blob/main/Lab6_Dashboard.py

Predictive Analysis (Classification)

- **Logistic Regression (score = 0.888888888888…)**
- Support Vector Machine (score = ?)
- Decision Tree Classifier (score = 0.7222222222222222…)
- K Nearest Neighbors (0.555555555555…)

- LR is better. (SVM I never could have a result)





- https://github.com/Rhodham96/AppliedDataScienceCapstone/blob/main/Lab7_Predictions.ipynb

Results

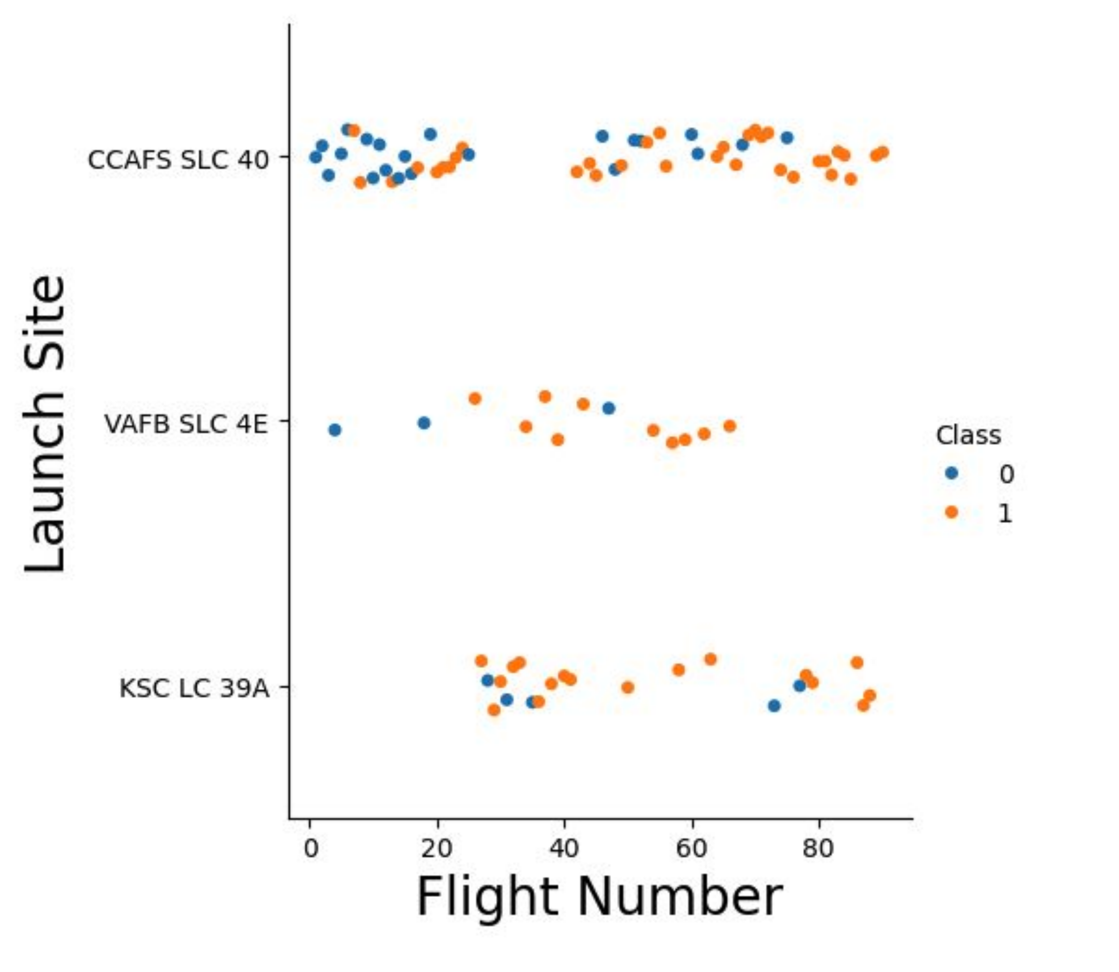
- Models used : Logistic Regression, SVM, Decision Tree, KNN
- Hyperparameter tuning with GridSearchCV
- Metrics: Accuracy, F1-score

Section 2

Insights drawn from EDA

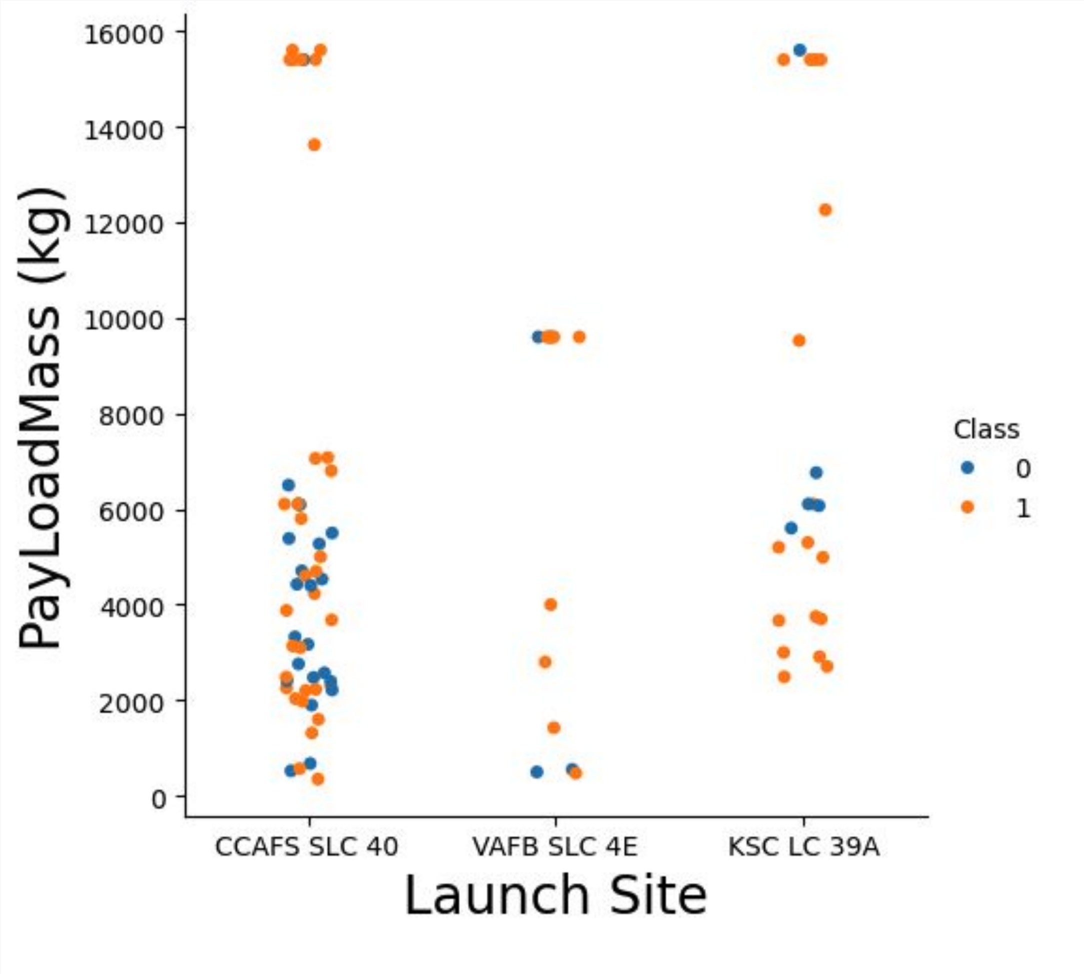
Flight Number vs. Launch Site

- When flight number gets bigger, success grow too
- Some launch sites where't user at beginning or failed but all became better



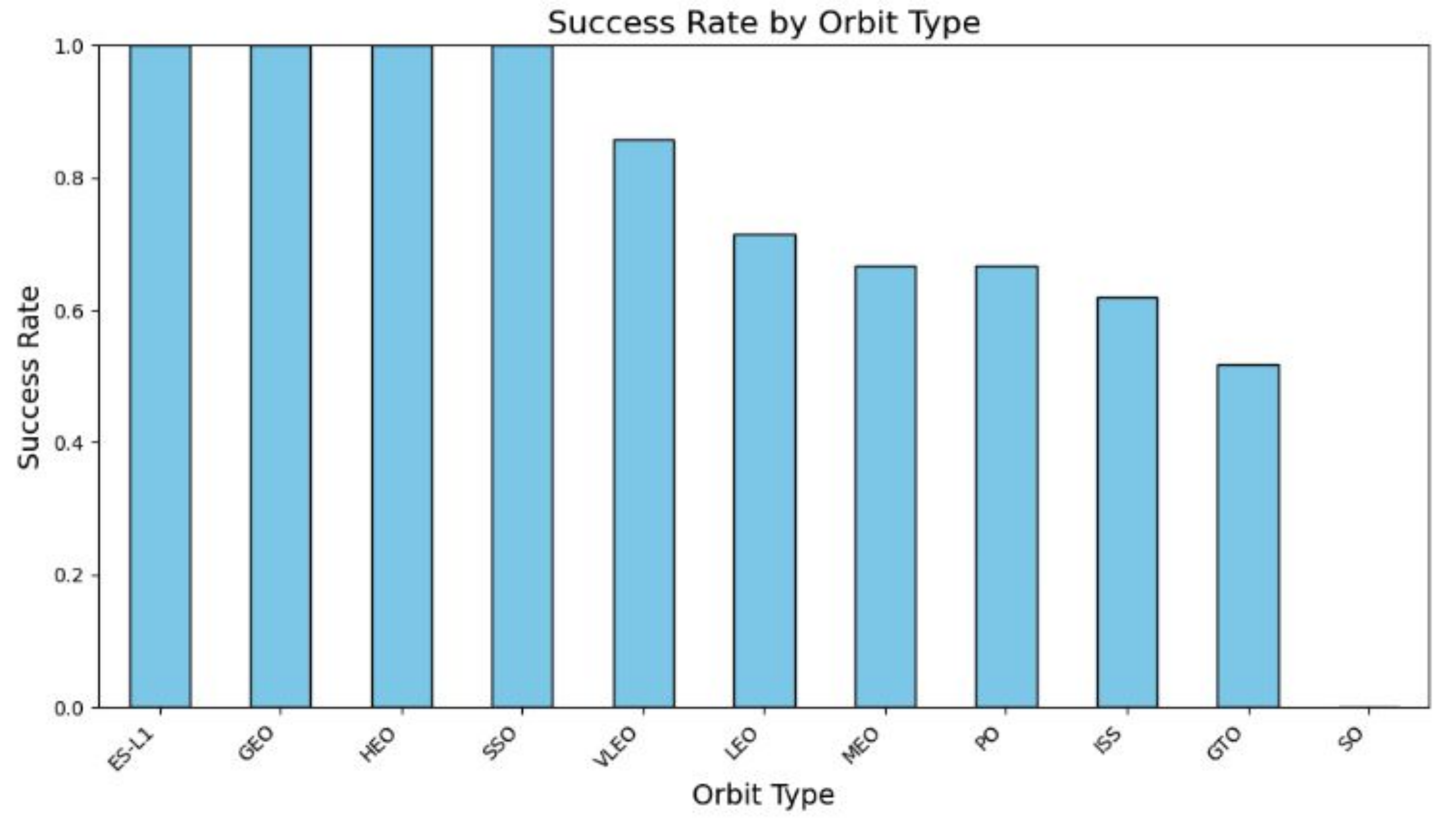
Payload vs. Launch Site

- CCAFS SLC 40 highest density of launches, various payload ranges, successful launches across payload ranges
- VAFBSLC 4E fewer launches, lighter payloads
- KSC LC 39A launches across the entire payload spectrum, success rate remain high, site is optimized for complex missions



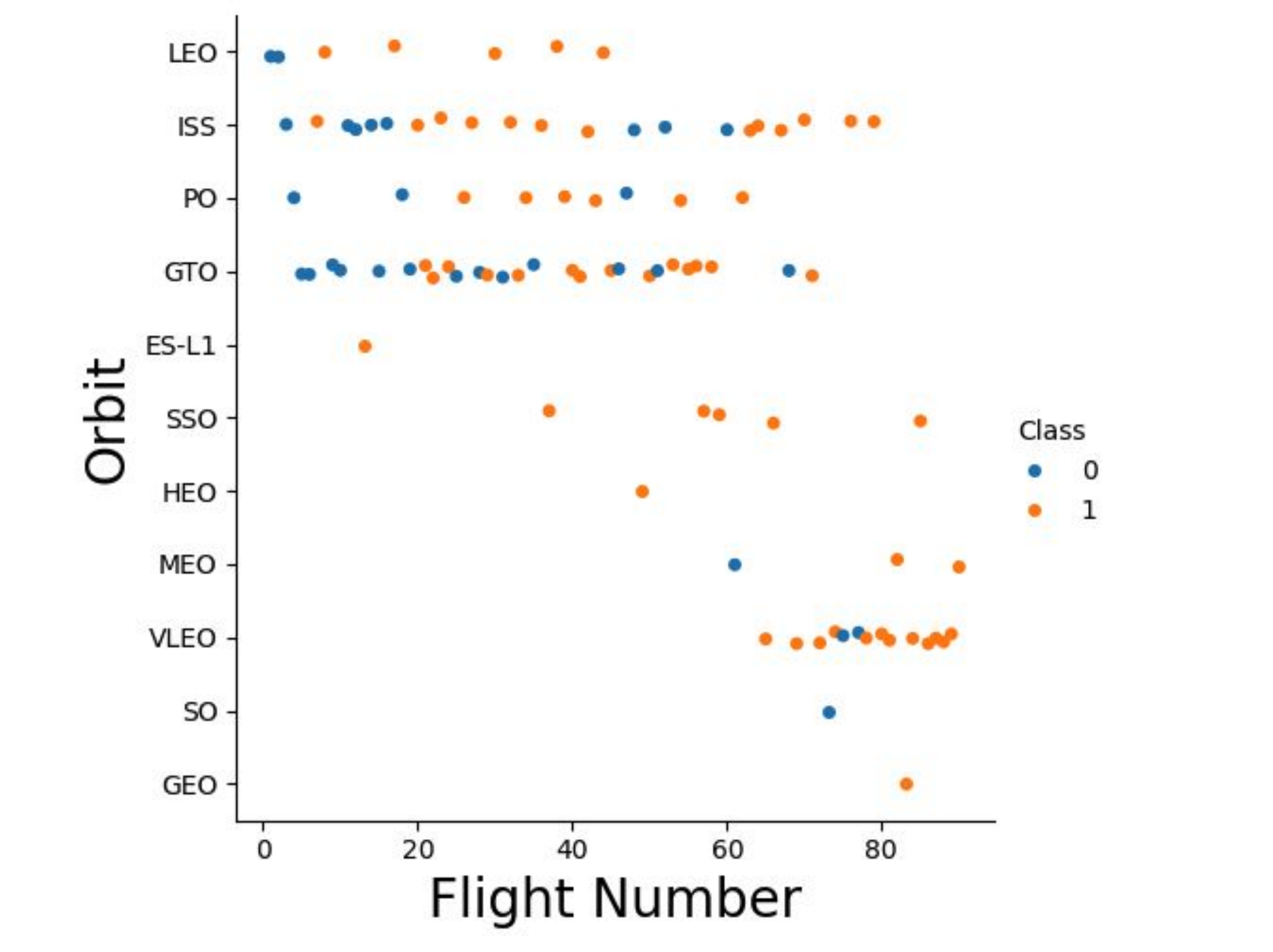
Success Rate vs. Orbit Type

- Orbits ES-L1, GEO, HEO ans SSO have perfect success rate
- VLEO, LEO, MEO, PO, success between 70%-90%
- GTO, SO lowest success rate (higher risk)



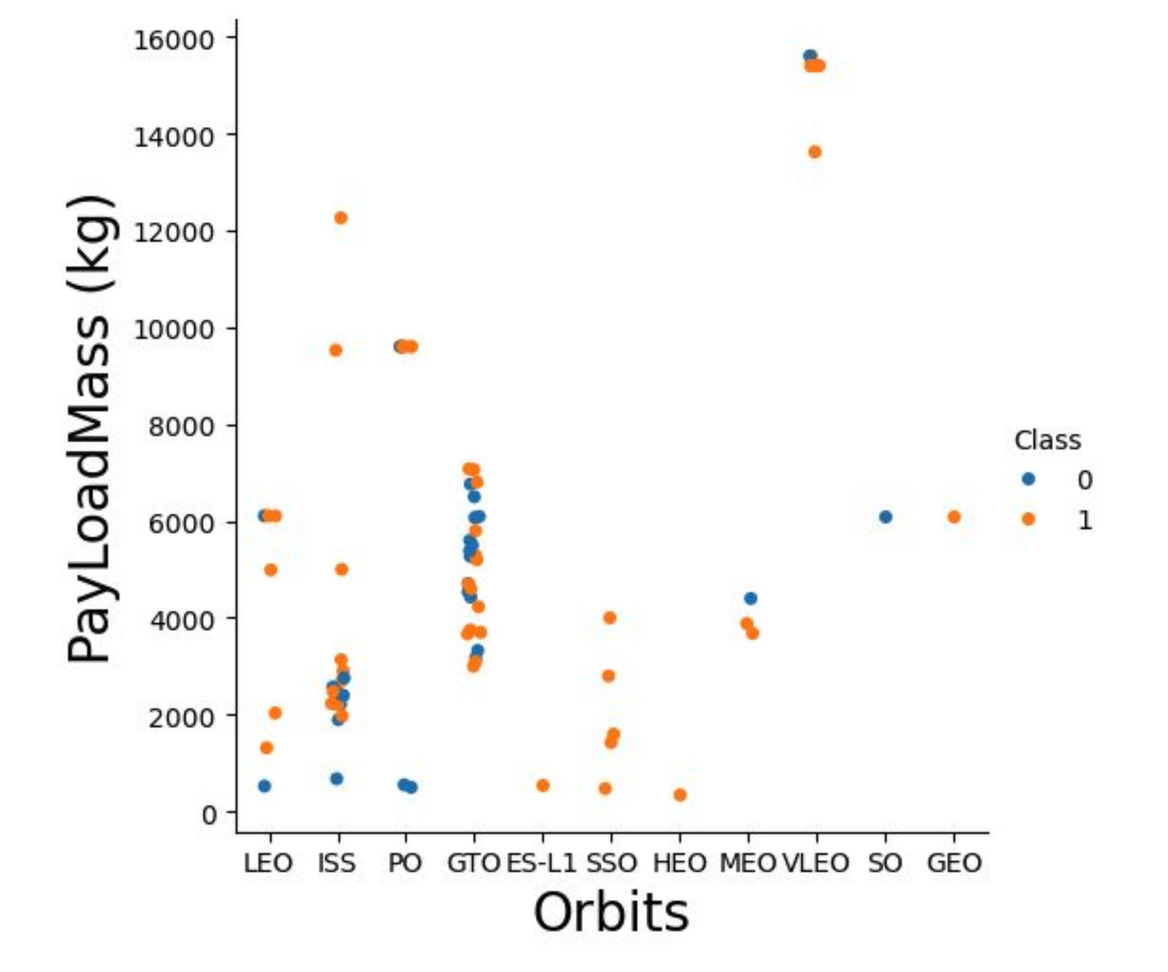
Flight Number vs. Orbit Type

- LEO and ISS frequent launches success dominate
- GTO failure rates are higher at earlier flight number, improvement over time
- SSO All successful
- For most orbit, success rate improve with higher flight number



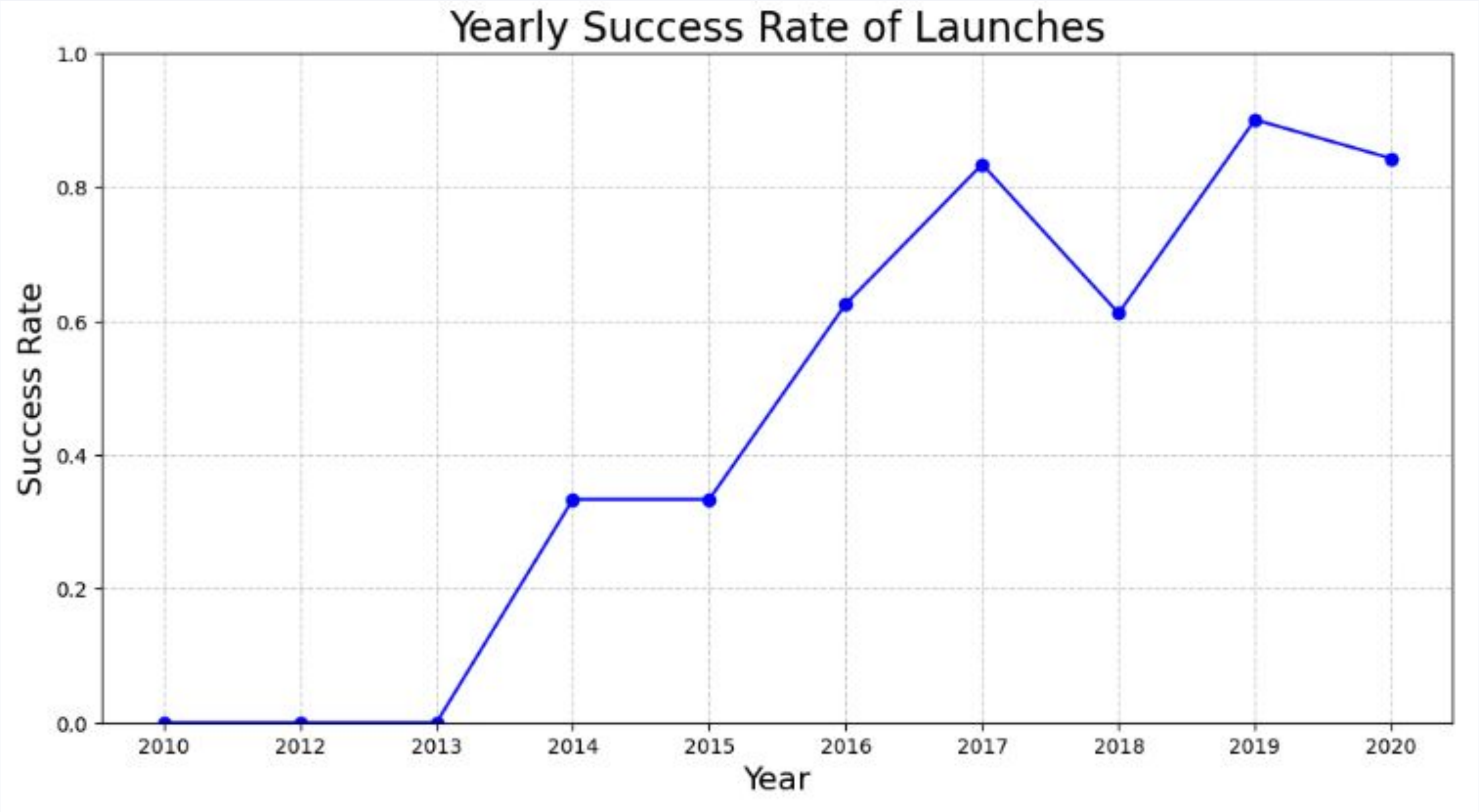
Payload vs. Orbit Type

- LEO wide range of payloads majority successful
- GTO mid range load mix success failure
- ISS lighter payloads (<8000kg)
- SSO more light (<6000kg)



Launch Success Yearly Trend

- Success increase over time
- Slight variablility in 2017-2018
- Stability in 2019-2020



All Launch Site Names

- %sql SELECT DISTINCT launch_site FROM SPACEXTBL;

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;

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)';

total_payload_mass
45596

Average Payload Mass by F9 v1.1

- %sql SELECT AVG(PAYLOAD_MASS__KG_) AS total_payload_mass FROM SPACEXTBL WHERE Booster_Version LIKE 'F9 v1.1%';
- Present your query result with a short explanation here

total_payload_mass
2534.6666666666665

First Successful Ground Landing Date

- %sql SELECT MIN(Date) FROM SPACEXTBL WHERE Landing_Outcome = 'Success (ground pad)';

MIN(Date)
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_ > 4000 AND PAYLOAD_MASS__KG_ < 6000;

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 landing_outcome, COUNT(*) AS total FROM SPACEXTBL GROUP BY landing_outcome;

Landing_Outcome	total
Controlled (ocean)	5
Failure	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	21
No attempt	1
Precluded (drone ship)	1
Success	38
Success (drone ship)	14
Success (ground pad)	9
Uncontrolled (ocean)	2

Boosters Carried Maximum Payload

- %sql SELECT Booster_version FROM SPACEXTBL WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXTBL);

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

- SELECT month_mapping.month_name, Booster_version, Launch_site, landing_outcome FROM SPACEXTBL JOIN month_mapping ON substr(Date, 6, 2) = month_mapping.month_num WHERE substr(Date, 1, 4) = '2015' AND landing_outcome = 'Failure (drone ship)';

month_name	Booster_Version	Launch_Site	Landing_Outcome
January	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
April	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

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

- %%sql SELECT landing_outcome, COUNT(*) AS outcome_count FROM SPACEXTBL WHERE Date BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY landing_outcome ORDER BY outcome_count DESC;

Landing_Outcome	outcome_count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

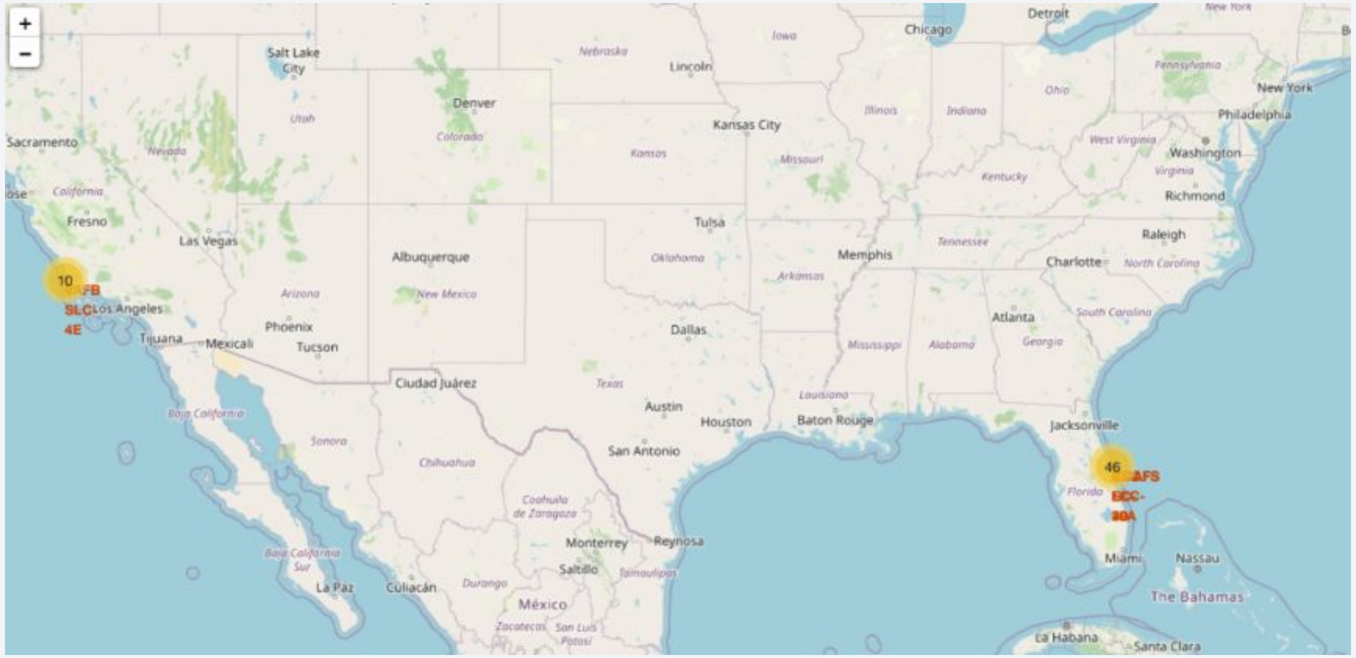
Launch Sites Proximities Analysis

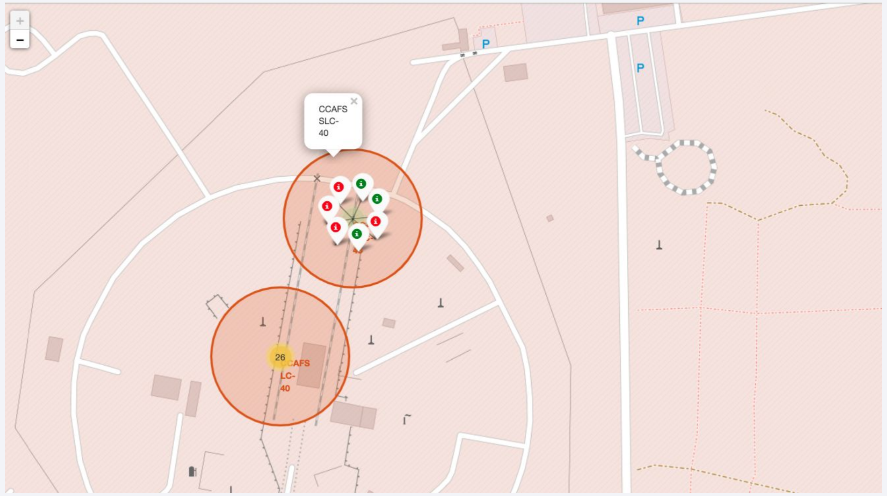
Mark all launch sites on a map

We can see where are all launch site (near the see)

Marks the success/failed launches on the map

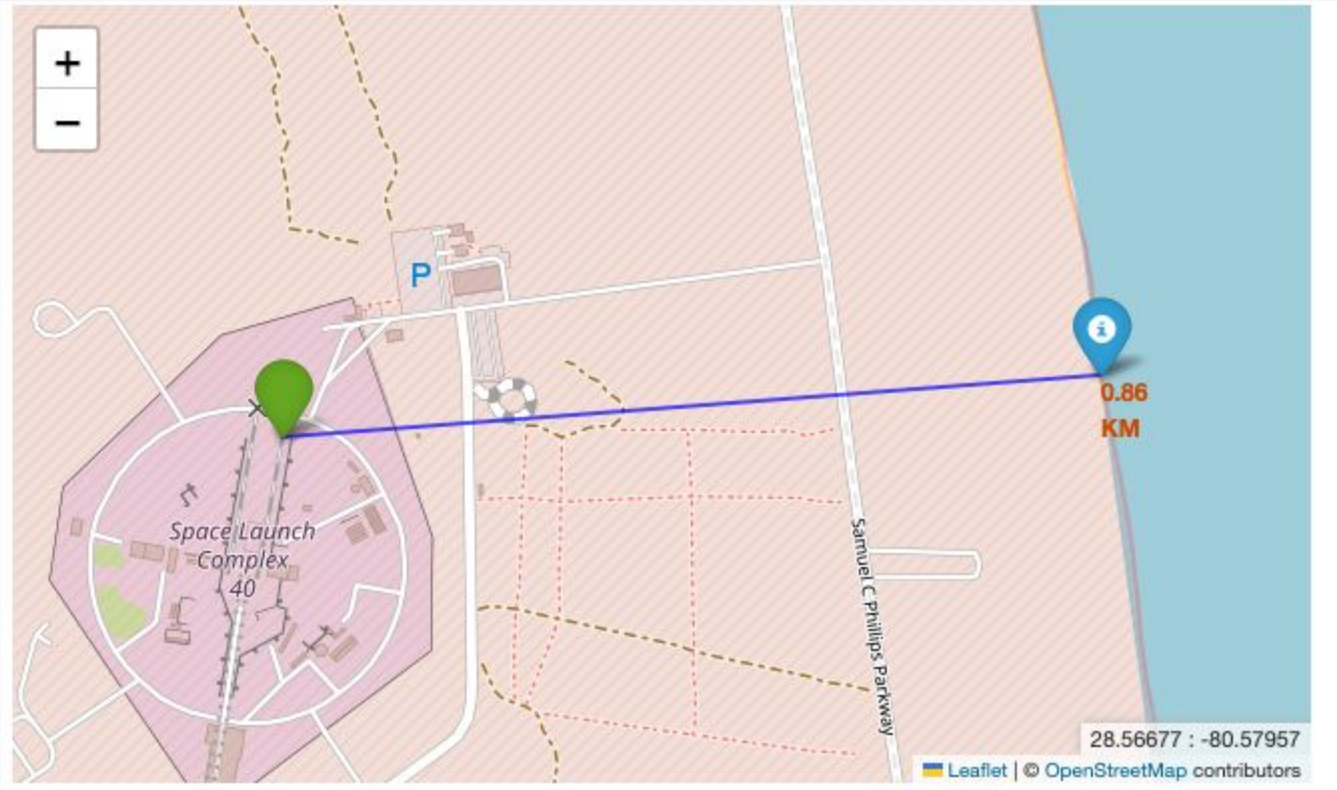
- We can see launches are on launch sites, if we zoom we can click on them and know if success or failed





Calculate the distance between launch sites and proximities

- There is 0.86KM between launch site and coast line

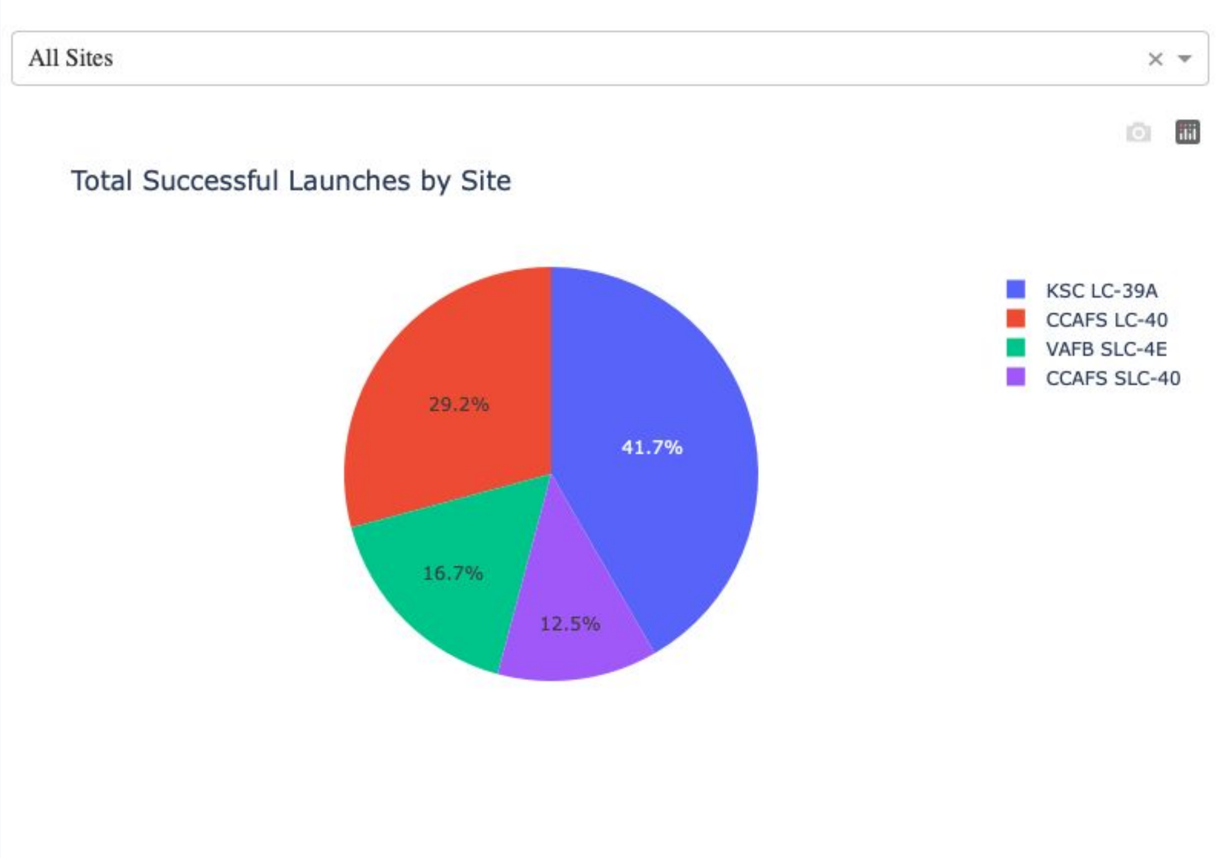


Section 4

Build a Dashboard with Plotly Dash

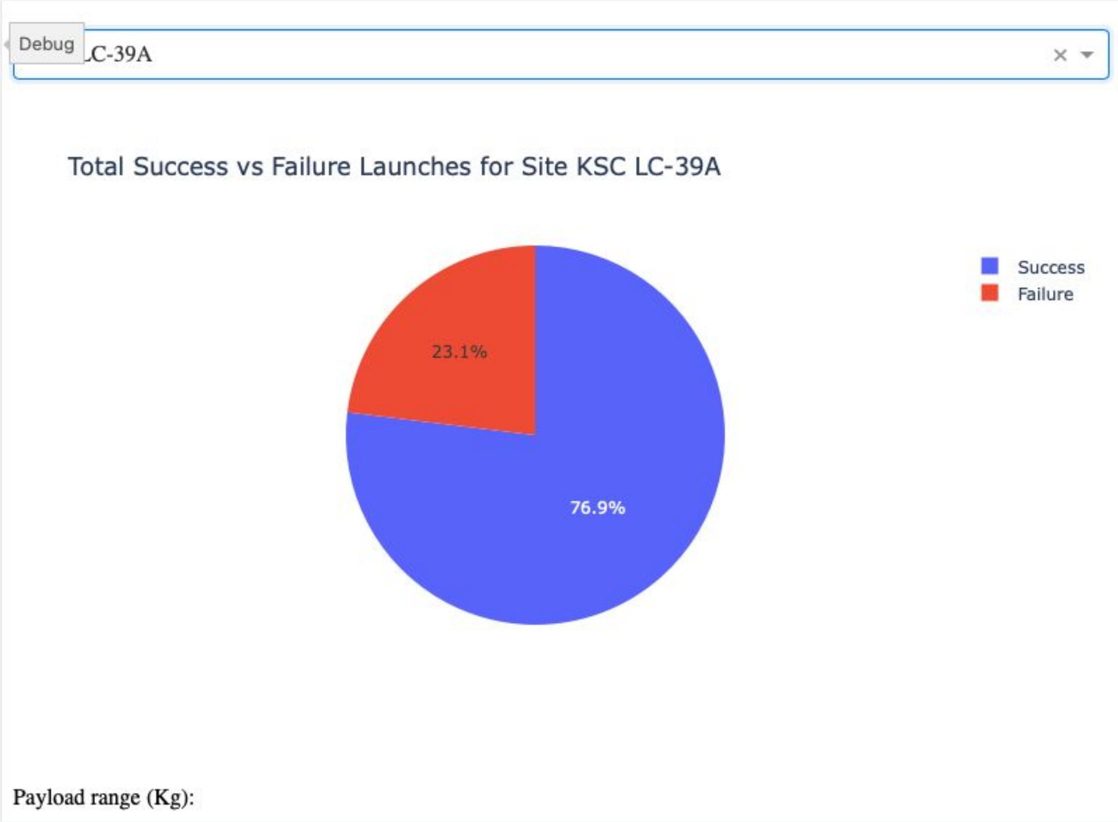
Launch sites repartition of success

- KSC Has more success



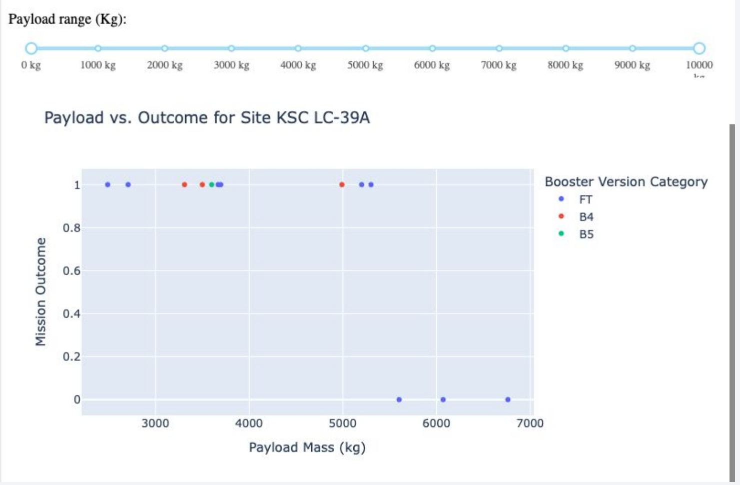
Success vs. Failure for KSC LC-39A

- Almost 72% success



Payload vs. Outcome for KSC LC-39A

- We can see differences between booster version with different payload.



Section 5

Predictive Analysis (Classification)

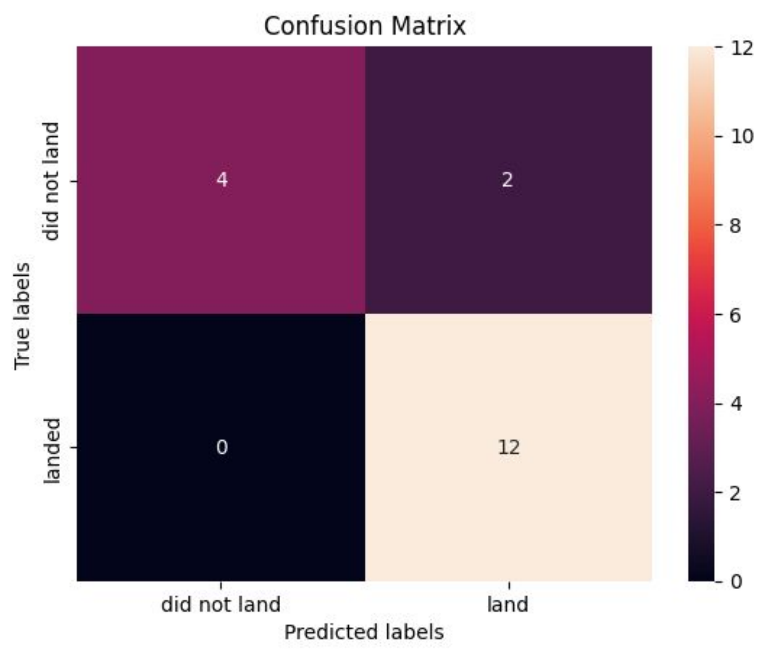
Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

- Find which model has the highest classification accuracy

Confusion Matrix

- Its almost perfect : 0 false didn’t land and 2 false land



Conclusions

1. **Launch Site Analysis**:
   - Florida-based launch sites (**KSC LC-39A** and **CCAFS LC-40**) have higher success rates, reflecting the optimization for frequent missions like geostationary and low Earth orbit launches.
   - The West Coast site (**VAFB SLC-4E**) focuses on polar orbit missions, with fewer launches but consistent success rates for lighter payloads.
2. **Payload and Outcome**:
   - Lighter payloads (<5000 kg) are generally more successful across all sites.
   - For heavier payloads (>10,000 kg), low Earth orbits (LEO) dominate and maintain high success rates, indicating advancements in handling complex missions.
   - Geostationary Transfer Orbits (GTO) show a higher risk profile with mixed success and failures, particularly in the mid-payload range (4000–6000 kg).
3. **Orbit-Specific Success Trends**:
   - Orbits like **SSO**, **GEO**, and **ES-L1** exhibit near-perfect success rates, suggesting reliability in specialized missions.
   - **GTO** and **SO** orbits experience lower success rates due to higher complexity and risk.
4. **Trends Over Time**:
   - Success rates have improved significantly since 2013, with consistent performance above 80% since 2016.
   - Early failures reflect the learning curve, while recent stability showcases operational maturity.
5. **Predictive Modeling Insights**:
   - Among classification models, Logistic Regression achieved the highest accuracy (88.89%).
   - Predictive insights indicate that payload, orbit type, and launch site are key factors in determining mission success.

Appendix

- https://github.com/Rhodham96/AppliedDataScienceCapstone

Thank you!