IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Van Chen
2023-02-17

Outline

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

Executive Summary

- Summary of methodologies
  - Data is collected using web scraping and API calls of data made public by SpaceX
  - Data is wrangled for summary report and predictive analyses of launch successes & failures.
- Summary of all results
  - Important variables are identified and isolated to improve future launch outcomes:
    - Payload Mass, Orbit, and Booster Version, Launch Location are determining factors.

Introduction

- The analysis is performed to evaluate and formulate how Space Y can compete with Space X with focus on each launch's first stage landing outcome.
- Questions to answer:
  - What variables are most important first-stage-rocket's landing outcomes?
  - Summary statistics of launch locations.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - SpaceX launch data is requested & parsed using the GET method form SkillsNetwork
  - From each launch record REQUEST method is used with SpaceX API for details
- Perform data wrangling
  - Categorical text values are turned into numerical values
  - Additional information is pulled
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Logistic Regression, SVM, Decision Tree, and KNN.

Data Collection

- Space X launch records:
  - https://api.spacexdata.com/v4/launches/past
- Space X API:
  - https://api.spacexdata.com/v4/rockets
  - https://api.spacexdata.com/v4/launchpads
  - https://api.spacexdata.com/v4/payloads
  - https://api.spacexdata.com/v4/cores

Data Collection – SpaceX API

Data collected using REST API

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
response = requests.get(spacex_url)
```

Data collection procedure (.ipynb) at GitHib

https://github.com/vanological/2023-02-17-Applied-Data-Science-Capstone/blob/f20d09e36b41bba9f6cafa1030fad2cad12592e5/2023-01-29%20Week1_Lab1%20Data%20Collection%20with%20Web%20Scraping.ipynb

RESP API Call

Reformat & Parse

Iterate past launches to web-scrape launch details

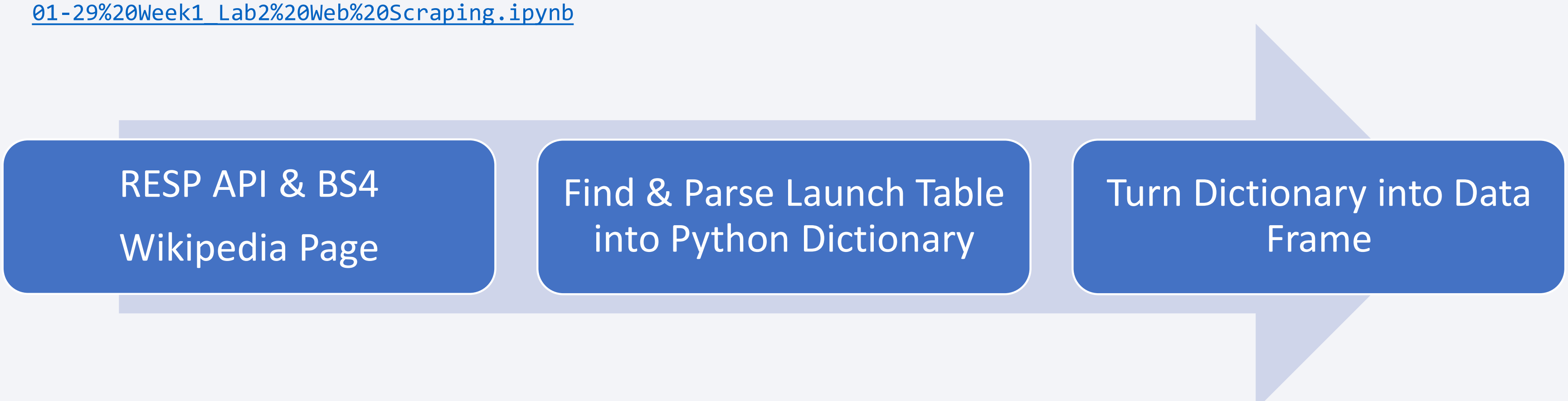
Data Collection - Scraping

Web Scraped using Beautiful Soup on Space X Wikipedia Page Table:

```
static_url =
"https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922"

response = requests.get(static_url)

soup = BeautifulSoup(response.text, 'html.parser')

first_launch_table = soup.find_all('tbody')[2]
```

Web Scraping procedure (.ipynb) at GitHib:

https://github.com/vanological/2023-02-17-Applied-Data-Science-Capstone/blob/91f7f477c53b039aca95c44063e2ceb8b65cb789/2023-01-29%20Week1_Lab2%20Web%20Scraping.ipynb



Data Wrangling

Launch data was, in turn, checked for missing values, correct data types, and with text values re-labeled to numerical values.

Data wrangling procedure (.ipynb) at GitHib

https://github.com/vanological/2023-02-17-Applied-Data-Science-Capstone/blob/3aee16dd595151569d1e87fc51a037d41f485666/2023-01-29%20Week1_Lab3%20Data%20Wrangling.ipynb

- Iterate through table columns: ***.isna()***
- ***DataFrame.info()*** & ***astype()***
- Landing outcomes turned into numerical values

EDA with Data Visualization

Data Visualization List:

- Scatter plot of ‘Payload Mass’ by ‘Flight Number’
- Scatter plot of ‘Launch site Mass’ by ‘Flight Number’
- Scatter plot of ‘Launch site Mass’ by ‘Payload Mass’
- Bar chart of ‘Number of Launches’ by ‘Orbit’
- Bar chart of ‘Success %’ by ‘Orbit’
- Scatter plot of ‘Orbit’ by ‘Flight Number’
- Line plot of ‘Success %’ by ‘Year’

Data wrangling procedure (.ipynb) at GitHib

https://github.com/vanological/2023-02-17-Applied-Data-Science-Capstone/blob/1f31fea21dc7ac5554d4dff904d39d489afd3104/2023-02-06%20Week2_Lab2%20EDA%20with%20Visualization%20Lab.ipynb

EDA with SQL

SQL Queries Summary:

- Names of unique launches
- Top 5 launch sites with names that begins with ‘CCA’ (‘CCA%’)
- Total payload mass launched by NASA (CRS)
- Average payload mass of booster version F9 v1.1
- Date of first successful landing outcome in ground pad
- Booster names, carrying payload mass in 4,000-6,000 kg range, with success in drone ship landing.
- Total number of success & failures
- Booster versions that carry maximum payload mass
- In year 2015, booster versions & launch sites of Drone ship landing failures
- Between 2016-06-04 & 2017-03-20 ranking of landing outcomes by count

SQL Queries at GitHub:

https://github.com/vanological/2023-02-17-Applied-Data-Science-Capstone/blob/e9c60eb65b11141b3ac30e68b44caaffb039563a/2023-01-31%20Week2_Lab1%20EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

- Markers, circles, and marker-clusters shown on maps using Folium:
  - Markers: Launch Sites
  - Circles: Coordinates of Launch Site
  - Marker-Cluster: Mission Success/Failures at Launch Sites

Data Visualization with Folium at GitHub:

https://github.com/vanological/2023-02-17-Applied-Data-Science-Capstone/blob/c27eb07ce2109f4cba70c37f84fb8d5c2dc7b257/2023-02-07%20Week3_Lab1%20Data%20Visualization%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

- “Percentage of Launches by Site” & “ Payload Range” were used to visualize data.
- The above visuals were selected to highlight relations between Payloads and Launch Sites.

Dashboard with Plotly at GitHub:

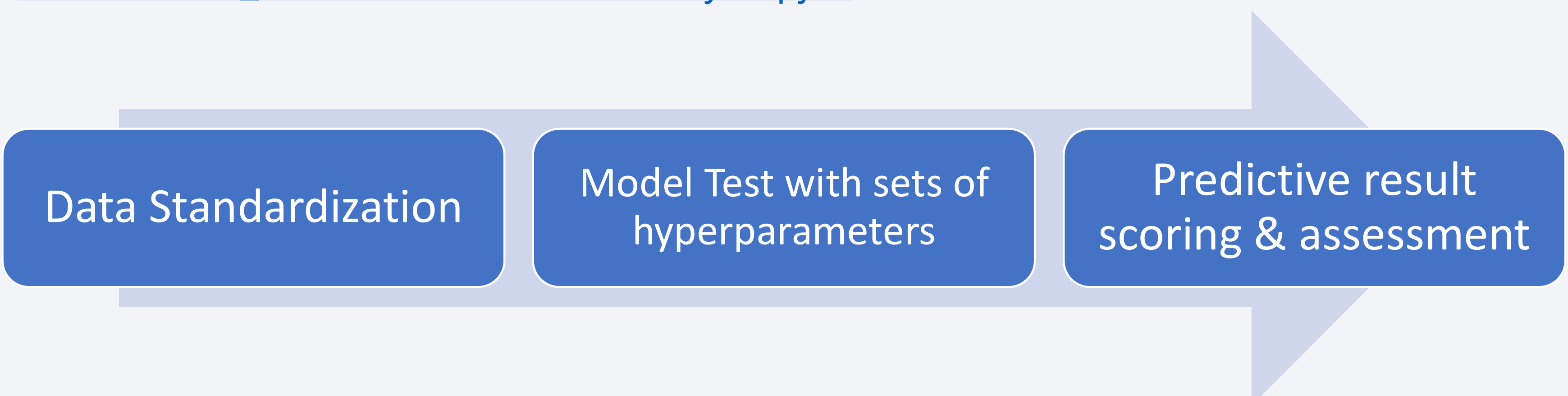
https://github.com/vanological/2023-02-17-Applied-Data-Science-Capstone/blob/f718413be59d64363a395d7759c03d44078ca26d/2023-02-13%20Week3_Lab2%20spacex_dash_app.py

Predictive Analysis (Classification)

- Logistic Regression, Support Vector Machine, Decision Tree, and K-Nearest-Neighbors were used to analyze data for predictive results.

Dashboard with Plotly at GitHub:

https://github.com/vanological/2023-02-17-Applied-Data-Science-Capstone/blob/463f825074467a3a785e72cb5b8a5af3f922a2d8/2023-02-16%20Week4_Lab1%20Predictive%20Analysis.ipynb



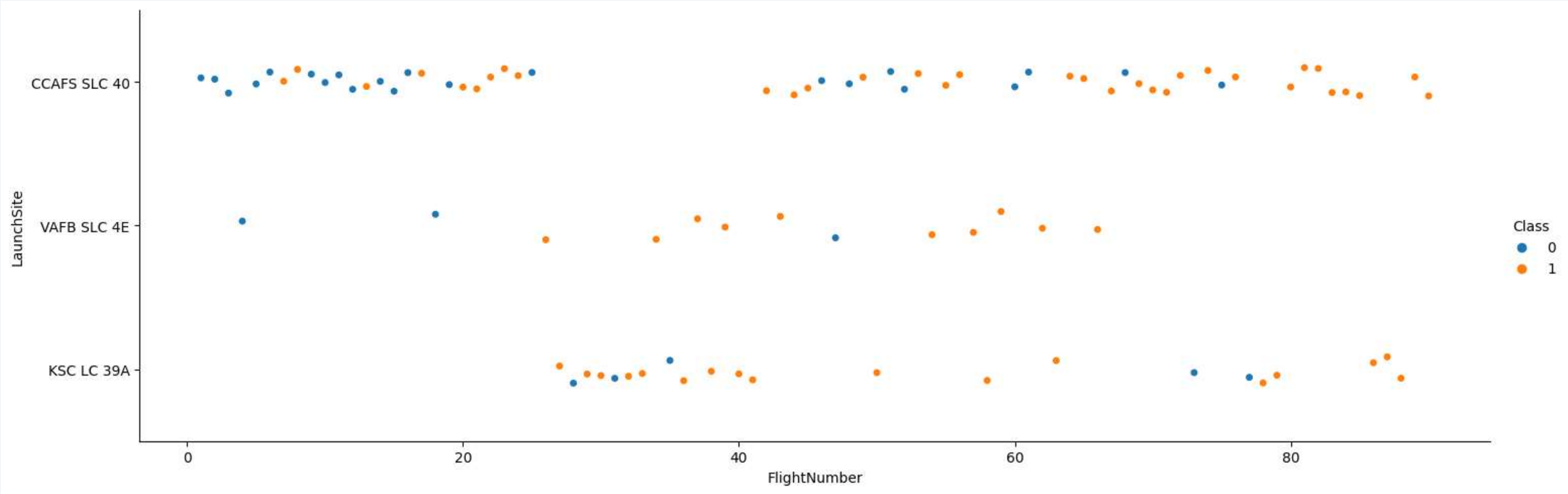
Results

- 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.

Section 2

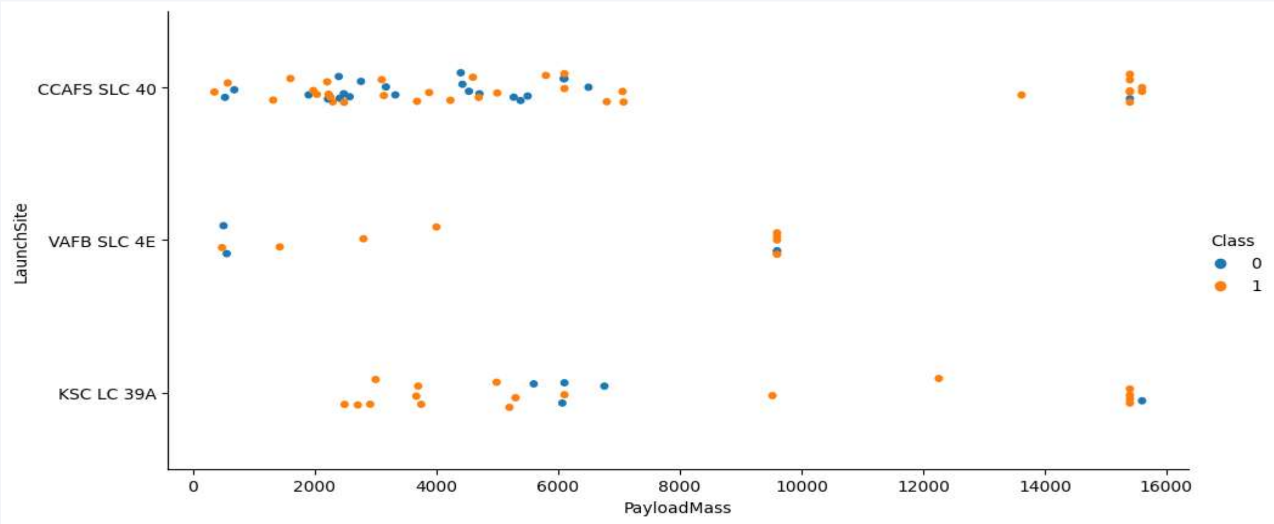
Insights drawn from EDA

Flight Number vs. Launch Site



- VAFB SLC 4E is not in use for recent launches despite high success rate.
- Launches are concentrated at CCAFS SLC 40.
- There has been visible & material increase in success rate after flight 60-70.

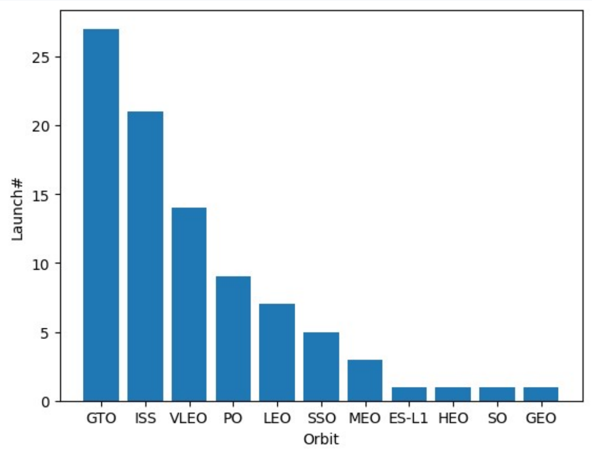
Payload vs. Launch Site

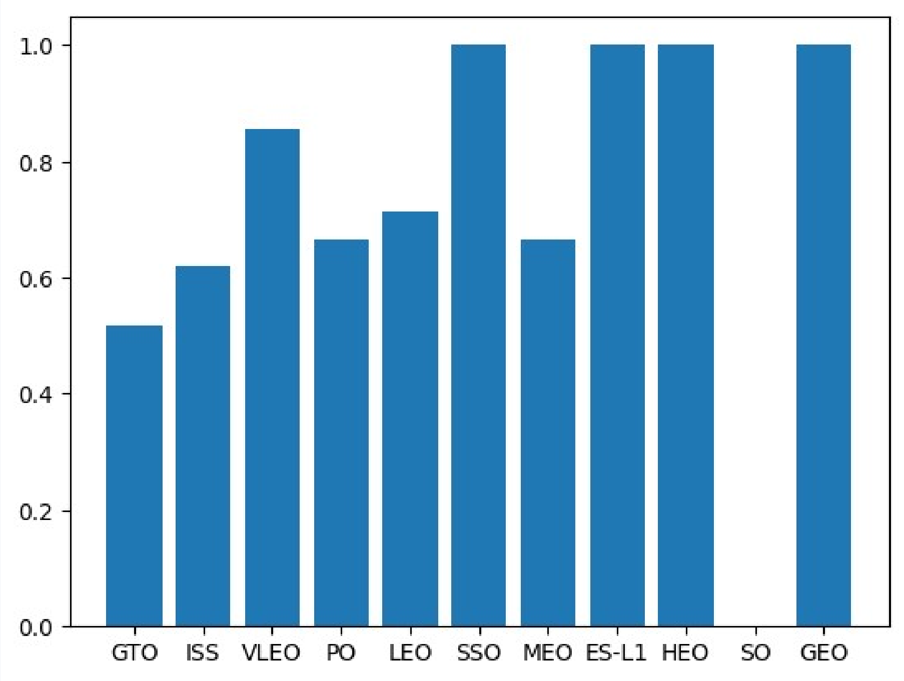


- VAFB SLC 4E is not used for launches with heavy payloads.
- Tentative correlation between payloads & success – further exploration needed to assess whether heavier payloads are associated with later launches so not to confound payloads & experience as causes of higher success rate.

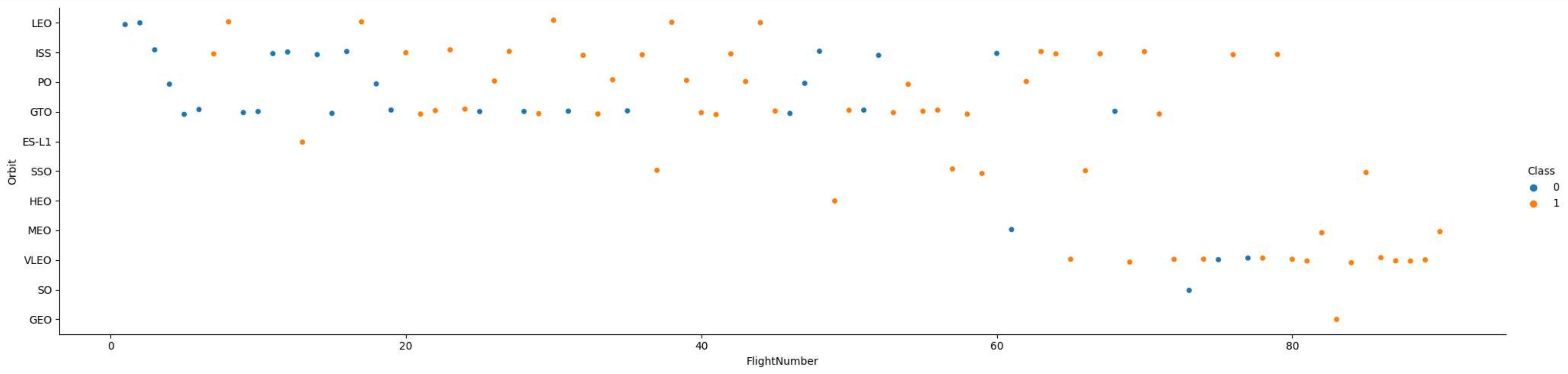
Success Rate vs. Orbit Type

- While certain orbit is associated with impressive success rate, it is more a matter of smaller number of launches.



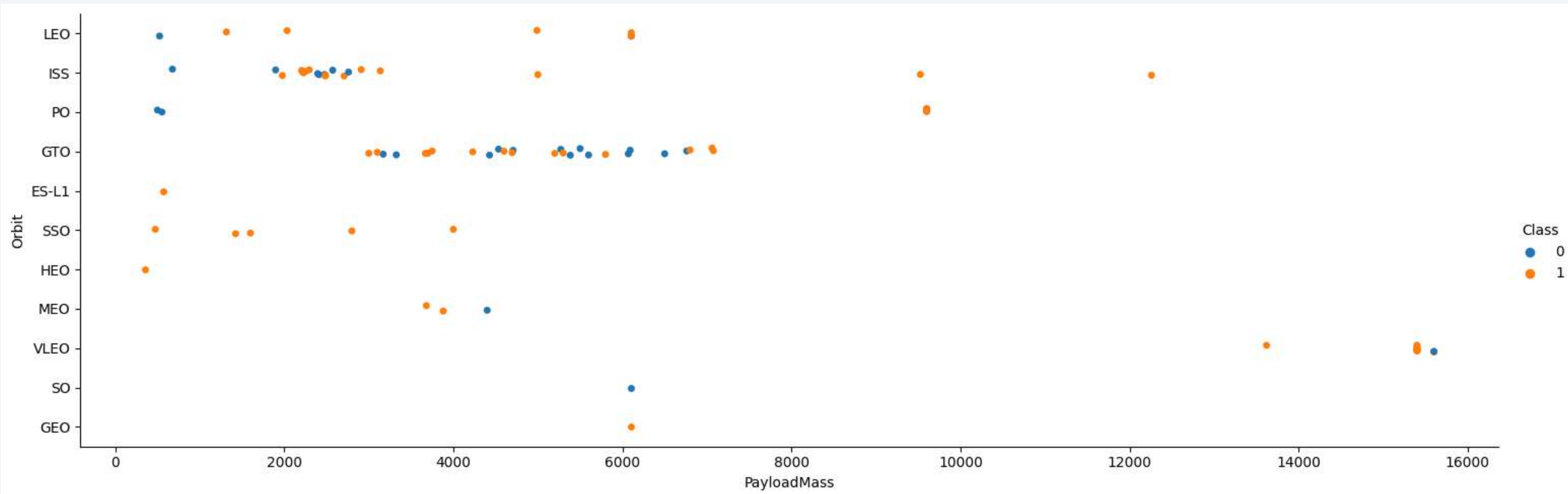


Flight Number vs. Orbit Type



- It doesn't seem that the success rate of launches to one orbit carry over to another.
- Higher success rate of one orbit (e.g. SSO) does not lead to consistent & repeated launch attempts.

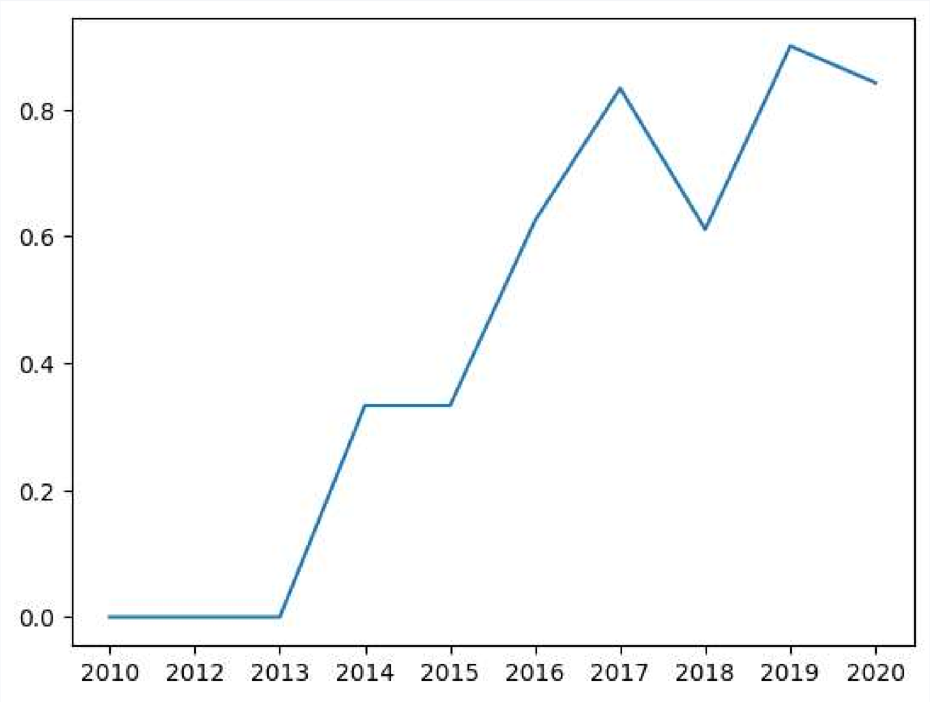
Payload vs. Orbit Type



- High success rate is seen once payload mass exceeds 8,000 kg.
- Further exploration needed so not to confound payload mass & launches' vintages as reason of higher success rate.

Launch Success Yearly Trend

- Clear indication of advancement of capabilities and assurances over time.



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;
```

Python

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

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
04-06-2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
08-12-2010	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)
22-05-2012	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
08-10-2012	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
01-03-2013	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_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)';
```

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

sum(PAYLOAD_MASS__KG_)
45596

Average Payload Mass by F9 v1.1

```
1  %sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSION like 'F9 v1.1';
```

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

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

```
%sql select min(Date) from SPACEXTBL where MISSION_OUTCOME='Success';
```

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

min(Date)
01-03-2013

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 4000 and 6000;

```

```
 * 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 QTY FROM SPACEXTBL GROUP BY MISSION_OUTCOME ORDER BY MISSION_OUTCOME;
```

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

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

Boosters Carried Maximum Payload

```
%sql select distinct(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

```
%sql SELECT BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXTBL WHERE "Landing _Outcome" = 'Failure (drone ship)' AND substr(Date,7,4)='2015';
```

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

Booster_Version	Launch_Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

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

```
%sql SELECT "DATE","Landing _Outcome",count("Landing _Outcome")as LANDING_OUTCOME_COUNT,DATE \
        from SPACEXTBL where substr(Date,7,4) || substr(Date,4,2) || substr(Date,1,2) between '20100604' \
        and '20170320' group by "Landing _Outcome" order by count("Landing _Outcome") desc

```

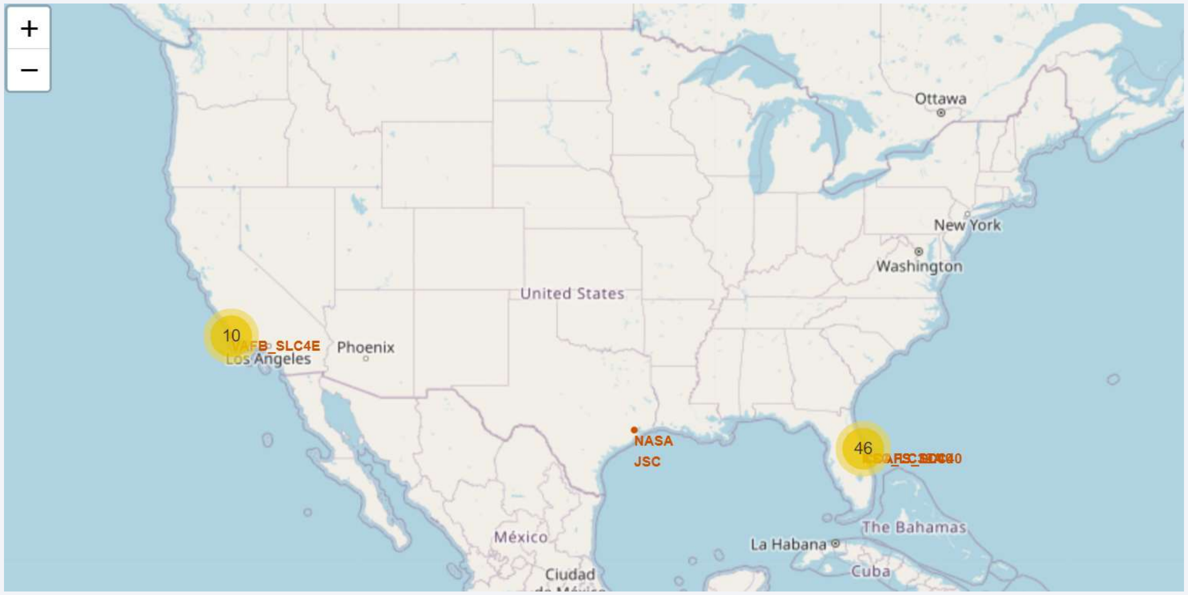
```
 * sqlite:///my_data1.db
Done.
```

Date	Landing _Outcome	LANDING_OUTCOME_COUNT	Date_1
22-05-2012	No attempt	10	22-05-2012
08-04-2016	Success (drone ship)	5	08-04-2016
10-01-2015	Failure (drone ship)	5	10-01-2015
22-12-2015	Success (ground pad)	3	22-12-2015
18-04-2014	Controlled (ocean)	3	18-04-2014
29-09-2013	Uncontrolled (ocean)	2	29-09-2013
04-06-2010	Failure (parachute)	2	04-06-2010
28-06-2015	Precluded (drone ship)	1	28-06-2015

Section 3

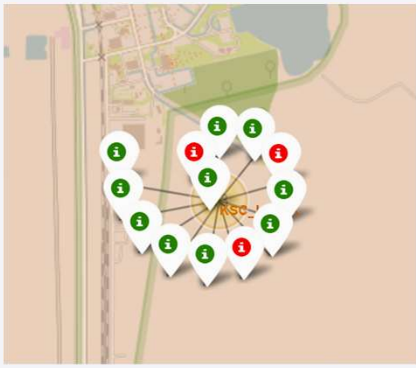
Launch Sites Proximities Analysis

All Launch Sites: Global View



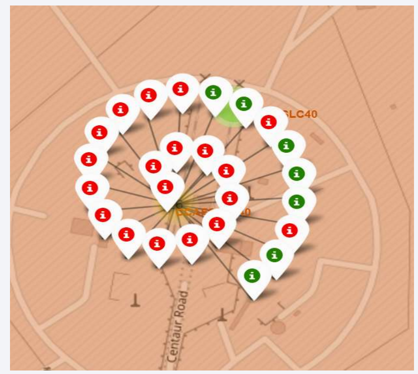
- Launch sites are at coastal region with launches concentrated on the East Coast

Launch Site Details with Outcomes

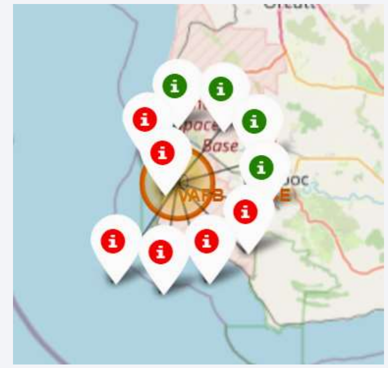


KSC_LC39A

CCAFS_SLC40



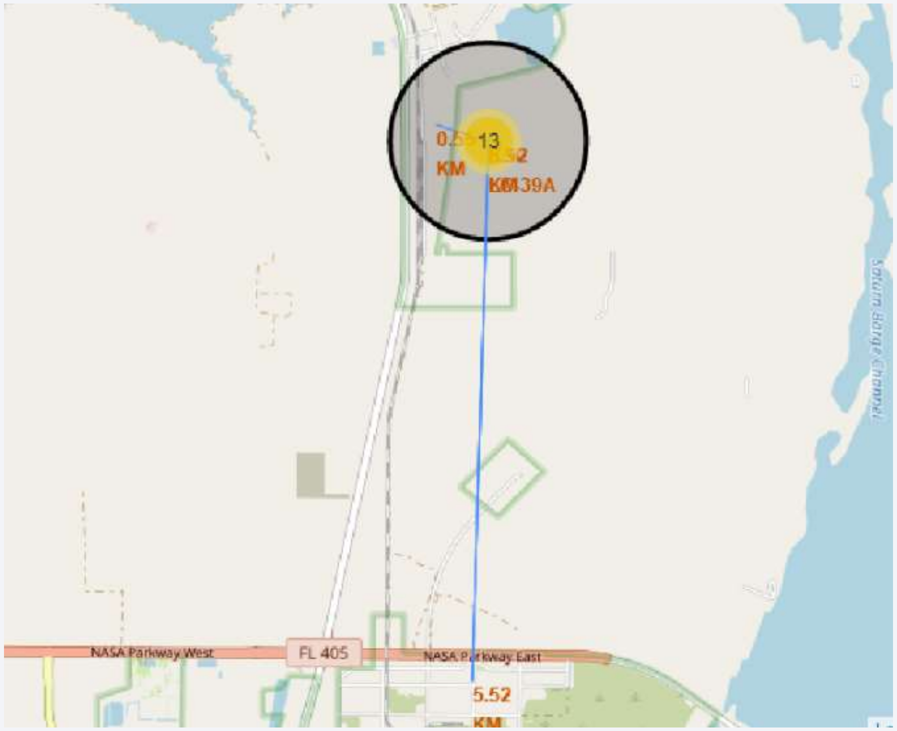
CCAFS_LC40



VAFB-SLC4E

- Green: Success / Red: Failure

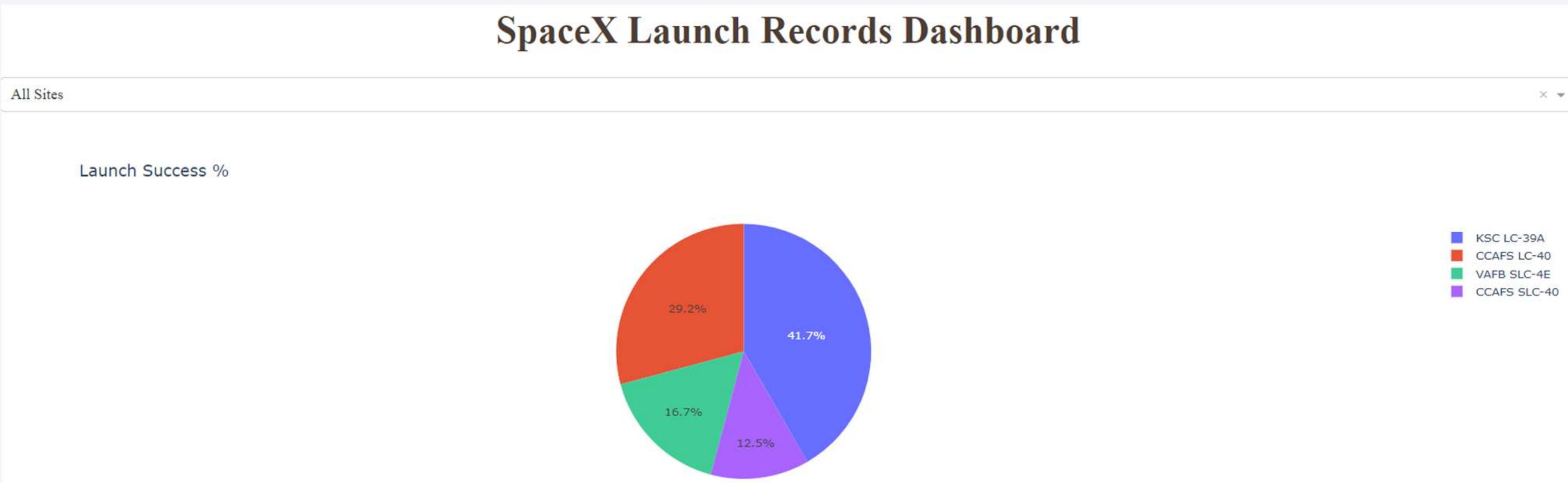
Distance to Nearby Rail Tracks & Roads



Section 4

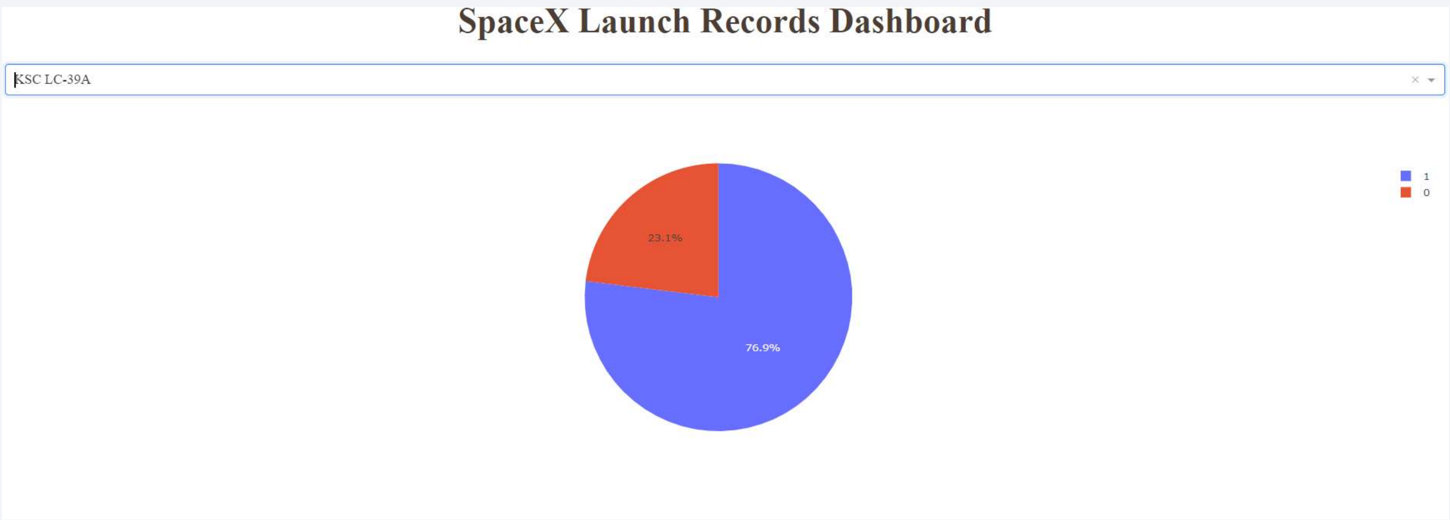
Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>



- Launch site itself is an important determinant of outcomes

KSC LC-39A



- KSC LC-39A has the highest success rate at 76.9%

Payload & Launch Site to Outcome

Section 5

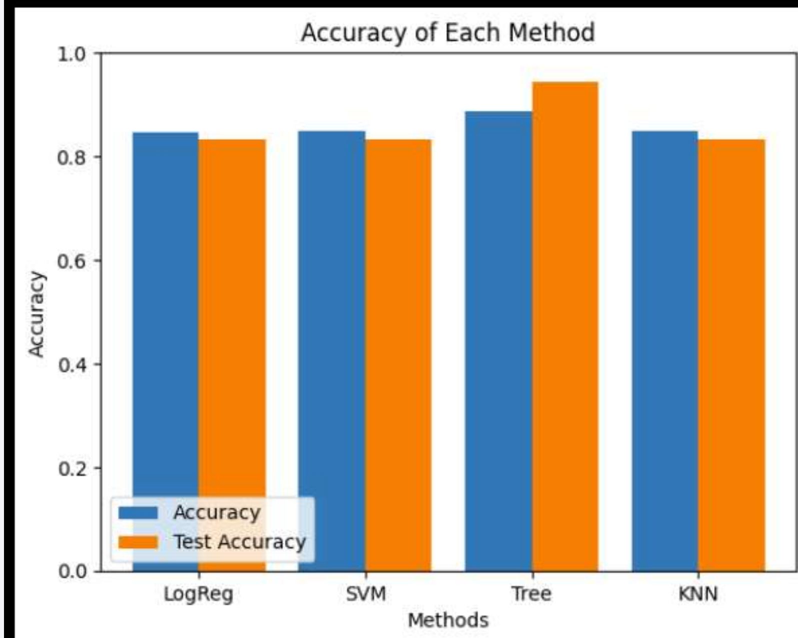
Predictive Analysis (Classification)

Classification Accuracy

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

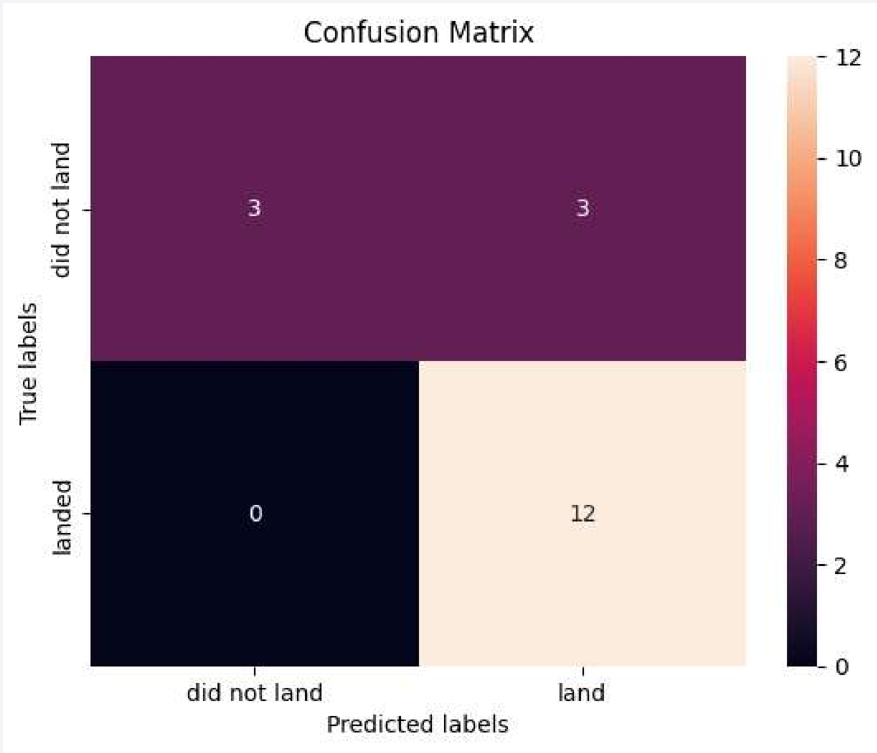
- Decision Tree has the highest accuracy score

Model	Accuracy	TestAccuracy
LogReg	0.84643	0.83333
SVM	0.84821	0.83333
Tree	0.88571	0.94444
KNN	0.84821	0.83333



Confusion Matrix

- True positive: 12
- False positive: 0
- True negative: 3
- False negative: 0
- Decision Tree performs well in both Precision (12 out of 15) and Recal (12 out of 12)

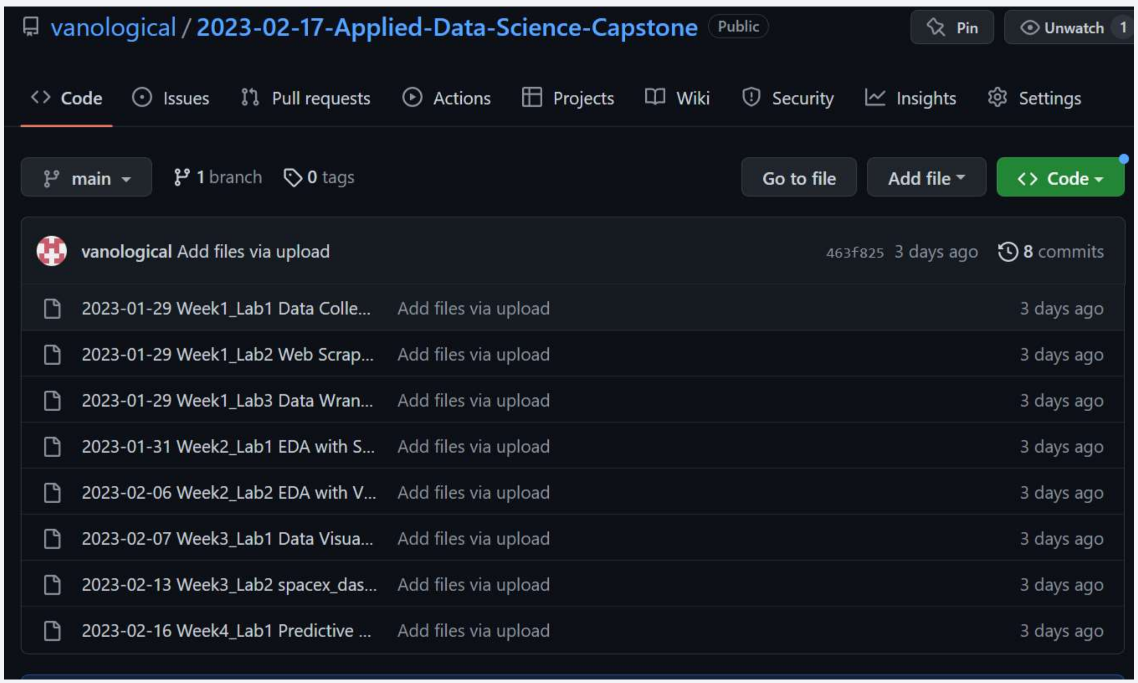


Conclusions

- Landing success has significantly improved over time.
- The more recent launches tend to:
  - Carry heavier payload (<8,000kg before flight#25, exceeding 12,000kg after flight#64)
  - Favor CCAFS SLC 40 over KSC LC 39A, with VAFB SLC 4E unused for the most recent 20 launches.
  - Favor orbit VLEO, ISS, and GTO (in sequence). VLEO launches see rockets with heavier payloads.
- Payloads, orbit, and launch sites are likely not the contributing factors of recent successes but the outcome of Space X's growing confidence, sophistication, and capability to handle bigger rockets with more ambitious (farther out) orbits that demand specific launch site configuration (CCAFS SLC 40).

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!