Path to a Successful Rocket Launch

Author: Anyue Wang

Date: 2022-11-29

OUTLINE

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

EXECUTIVE SUMMARY

- More successful rocket launch as time goes by
- Locate launch sites near coastline but far way from cities
- Successful launches may be related to:
  - Large flight numbers
  - Heavy payloads
  - Different orbits
  - Good choice of boosters
- Different classification models have similar performance with 83.3% accuracy

INTRODUCTION

- Background:
  - The results (success or not) of launching rockets are essential to the cost for rocket companies.
  - To help save potential cost, rocket companies are desired to explore useful information from previous rocket launching data.
- Objective:
  - To analyze factors affecting the results of rocket launches
  - To predict the result of a rocket launch
- Analytical questions:
  - What features are related the most to the success/failure of a rocket launch?
  - What models can accurately predict the results given the features?

METHODOLOGY – Data Collection and Wrangling

- Data Source:
  - SpaceX API: https://api.spacexdata.com/v4
  - Wikipedia page "List of Falcon 9 and Falcon Heavy Launches": https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- Data Collecting Methods:
  - Request to API (requests, pandas)
  - Web scraping (BeautifulSoup)
- Data Wrangling Methods:
  - Convert non-numeric data into numeric data
    - e.g., landing results with True Ocean, False ASDS, etc., are transformed into numeric data where 1 meaning successful landing and 0 meaning failure landing

METHODOLOGY – EDA (Visualization)

- Data visualization:
    - Pandas, Matplotlib (enabling basic analysis and feature engineering)
    - Folium (analysis related to geographical features)
    - Plotly Dash Dashboard (interactive visual analytics)

METHODOLOGY – EDA (SQL)

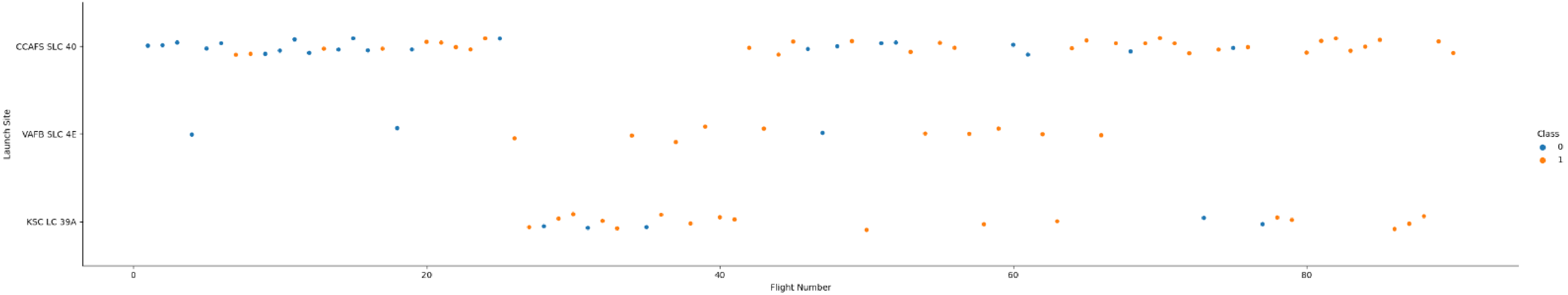
- Data analysis by SQL queries
  - Display the names of the unique launch sites in the space mission
  - Display 5 records where launch sites begin with the string 'CCA'
  - Display the total payload mass carried by boosters launched by NASA (CRS)
  - Display average payload mass carried by booster version 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 successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order

METHODOLOGY – Predictive Analysis

- Classification methods are applied to predict the landing results:
  - Classification methods (sklearn): logistic regression, SVM, decision tree, k nearest neighbors
  - Grid search method is applied to select appropriate hyperparameters
  - Cross-validation method is applied to improve the predictive performance on out-of-sample data (test_size=0.2)
  - Accuracy rate is adopted for performance evaluation
- Other information:
  - Features: FlightNumber, PayloadMass, Flights, Block, ReusedCount, Orbit, LaunchSite, LandingPad, Serial, GridFins, Reused, Legs

RESULTS - Visualization

Launch Site vs Flight Number

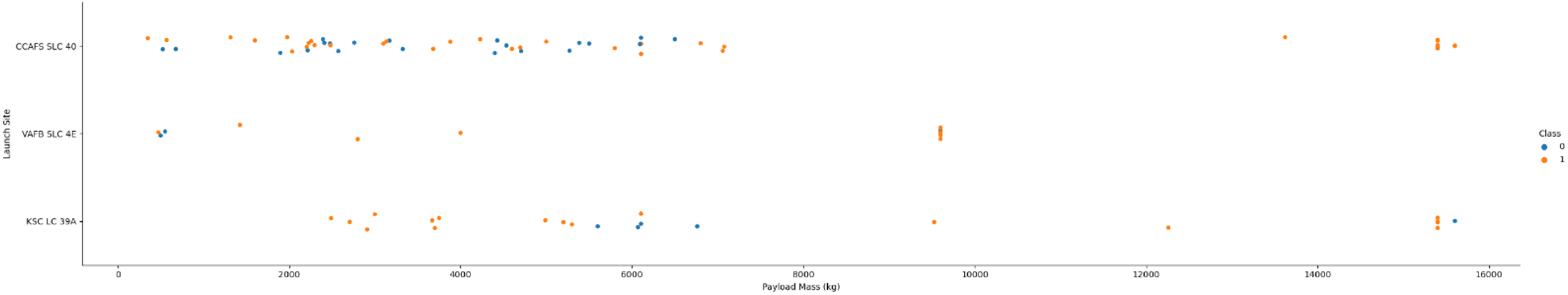


Pattern analysis:

- Compared to CCAFS SLC 40, the landing successful rates at the launch sites KSC LC 39A and VAFB SLC 4E are higher
- As the flight number increases, the landing tends to be more successful

RESULTS - Visualization

Launch Site vs Payload Mass

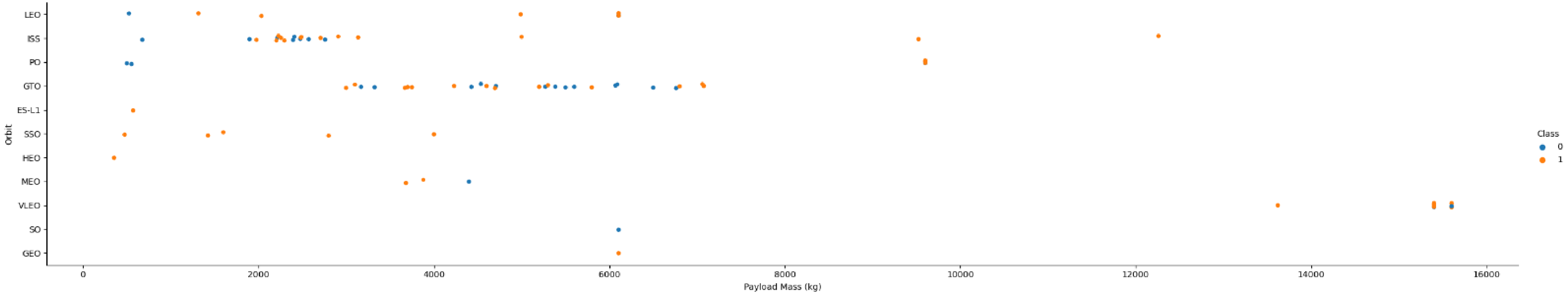


Pattern analysis:

- Generally, landing tends to be successful when the payload mass is larger than 8000 kg
- For KSC LC 39A, failures happen frequently when payload mass is between 5000 kg and 7000 kg

RESULTS - Visualization

Orbit vs Payload Mass

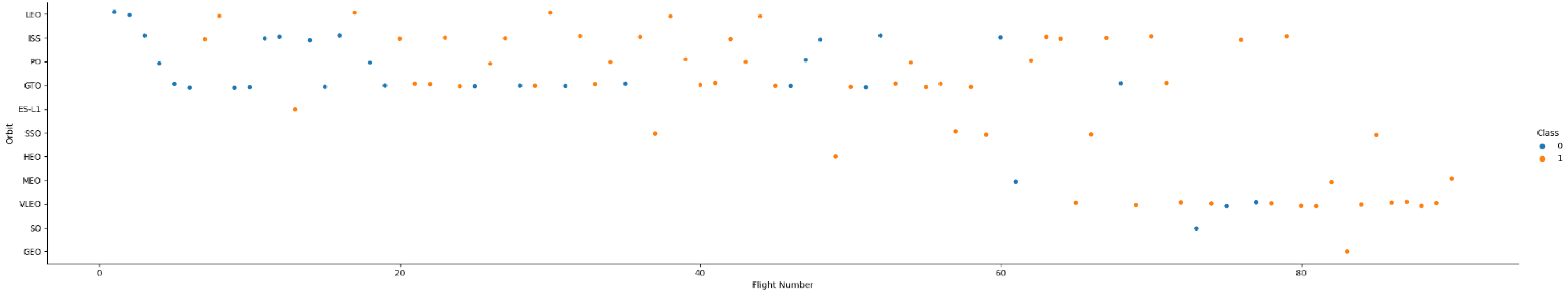


Pattern analysis:

- The relationship seems different among various orbit
- The heavier the payload is, the more successful landing is envisioned for LEO and PO
- For GTO, the relationship is not obvious

RESULTS - Visualization

Orbit vs Flight Number



Pattern analysis:

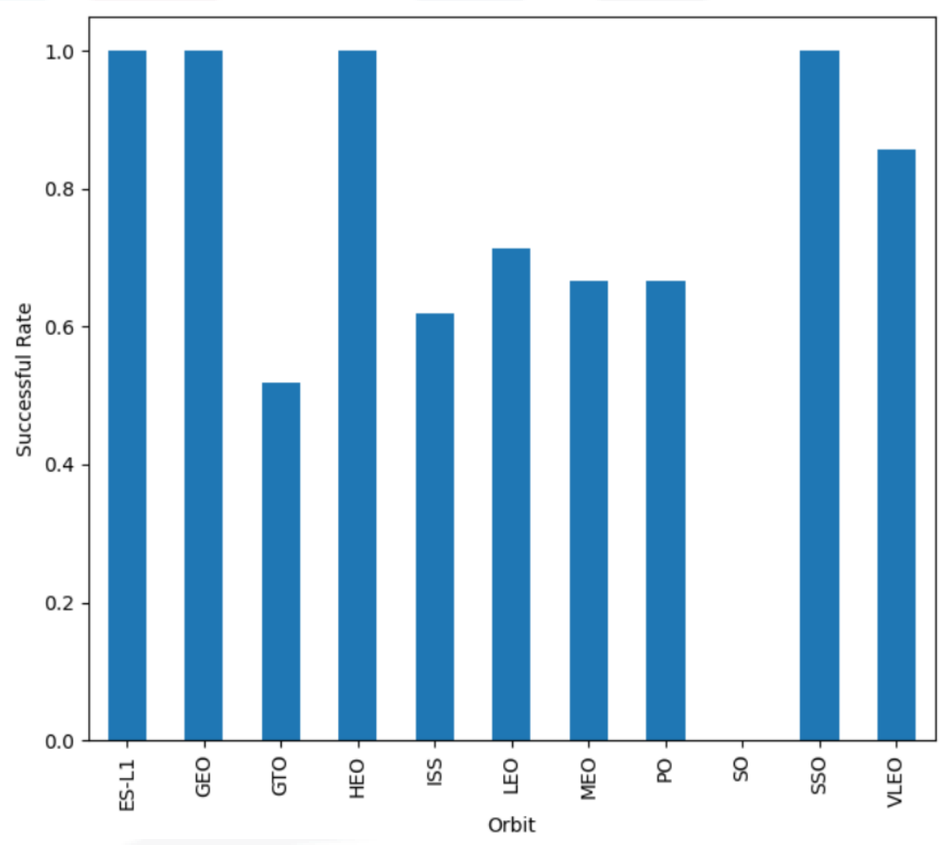
- The relationship seems different among various orbit
- For LEO, the landing tends to be successful as the flight number increases
- For ISS and GTO, the relationship is not obvious

RESULTS - Visualization

Successful Rate vs Orbit

Pattern analysis:

- The successful rates of satellites at ES-L1, GEO, HEO, SSO orbits are 100%
- Satellites at SO have no successul landing yet

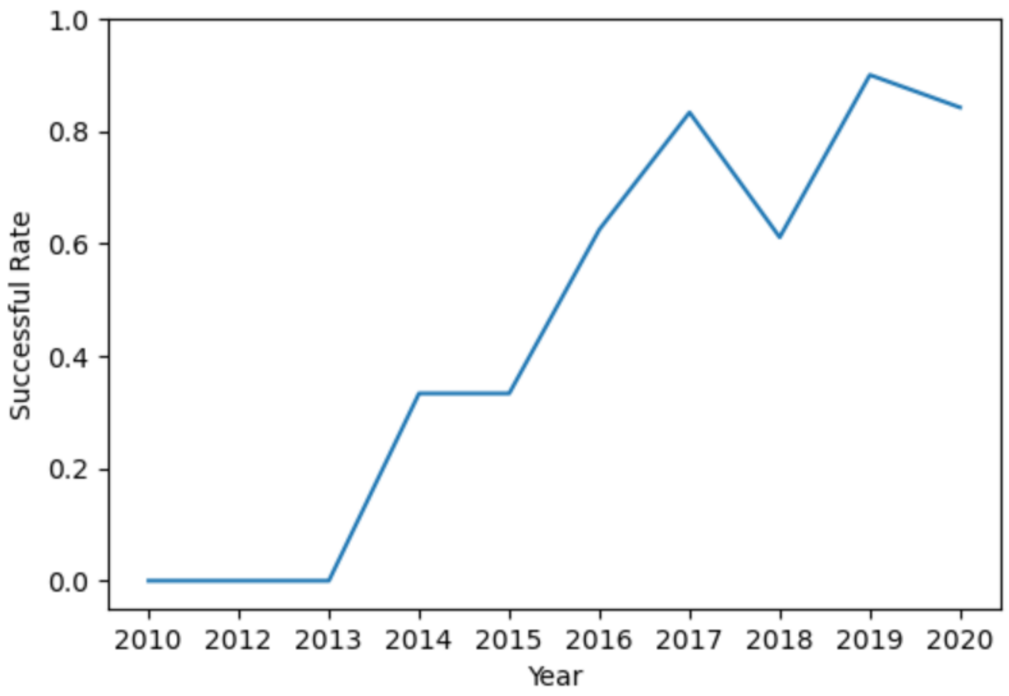


RESULTS - Visualization

Successful Rate vs Year

Pattern analysis:

- Before 2014, the successful rate is 0%
- Since 2014, the successful rate increases as the year increases



RESULTS - Visualization

Feature Engineering

Generate dummy variables of selected features, including Orbit, LaunchSite, LandingPad, Serial

	FlightNumber	PayloadMass	Flights	GridFins	Reused	Legs	Block	ReusedCount	Orbit_ES-L1	Orbit_GEO	...	Serial_B1048	Serial_B1049	Serial_B1050	Serial_B1051	Serial_E
0	1	6104.959412	1	False	False	False	1.0	0	0	0	...	0	0	0	0	
1	2	525.000000	1	False	False	False	1.0	0	0	0	...	0	0	0	0	
2	3	677.000000	1	False	False	False	1.0	0	0	0	...	0	0	0	0	
3	4	500.000000	1	False	False	False	1.0	0	0	0	...	0	0	0	0	
4	5	3170.000000	1	False	False	False	1.0	0	0	0	...	0	0	0	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
85	86	15400.000000	2	True	True	True	5.0	2	0	0	...	0	0	0	0	
86	87	15400.000000	3	True	True	True	5.0	2	0	0	...	0	0	0	0	
87	88	15400.000000	6	True	True	True	5.0	5	0	0	...	0	0	0	1	
88	89	15400.000000	3	True	True	True	5.0	2	0	0	...	0	0	0	0	
89	90	3681.000000	1	True	False	True	5.0	0	0	0	...	0	0	0	0	

90 rows × 80 columns

RESULTS - SQL

Task 1

Display the names of the unique launch sites in the space mission

```
%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

Task 2

Display 5 records where launch sites begin with the string '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
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

Task 3

Display the total payload mass carried by boosters launched by NASA (CRS)

```
%sql select sum("PAYLOAD_MASS__KG_") from SPACEXTBL where "Customer" = "NASA (CRS)"
```

* sqlite:///my_data1.db
Done.

sum("PAYLOAD_MASS__KG_")
45596

Task 4

Display average payload mass carried by booster version F9 v1.1

```
%sql select avg("PAYLOAD_MASS__KG_") from SPACEXTBL where "Booster_Version" like "%F9 v1.1%"
```

* sqlite:///my_data1.db
Done.

avg("PAYLOAD_MASS__KG_")
2534.6666666666665

RESULTS - SQL

Task 5

List the date when the first succesful landing outcome in ground pad was acheived.

Hint:Use min function

```
%sql select min(substr("Date",7,4)||substr("Date",4,2)||substr("Date",1,2)) from SPACEXTBL where "Landing _Outcome" = "Success (ground pad)"
```

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

min(substr("Date",7,4)||substr("Date",4,2)||substr("Date",1,2))
20151222

Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
%sql select "Booster_Version", "Payload" from SPACEXTBL where "Landing _Outcome" = "Success (drone ship)" and "PAYLOAD_MASS__KG_" between 4000 and 6000
```

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

Booster_Version	Payload
F9 FT B1022	JCSAT-14
F9 FT B1026	JCSAT-16
F9 FT B1021.2	SES-10
F9 FT B1031.2	SES-11 / EchoStar 105

RESULTS - SQL

Task 7

List the total number of successful and failure mission outcomes

```
%sql select "Mission_Outcome", count("MISSION_OUTCOME") from SPACEXTBL group by "MISSION_OUTCOME" like "Success%" or "Failure%"
```

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

Mission_Outcome	count("MISSION_OUTCOME")
Failure (in flight)	1
Success	100

Task 8

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

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

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

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

RESULTS - SQL

Task 9

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.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 4, 2) as month to get the months and substr(Date,7,4)='2015' for year.

```
%sql select substr(Date,4,2), "Landing _Outcome", "BOOSTER_VERSION", "LAUNCH_SITE" from SPACEXTBL where ("LANDING _Outcome" = "Failure (drone ship)") and (substr(
```

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

substr(Date,4,2)	Landing _Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Task 10

Rank the count of successful landing_outcomes between the date 04-06-2010 and 20-03-2017 in descending order.

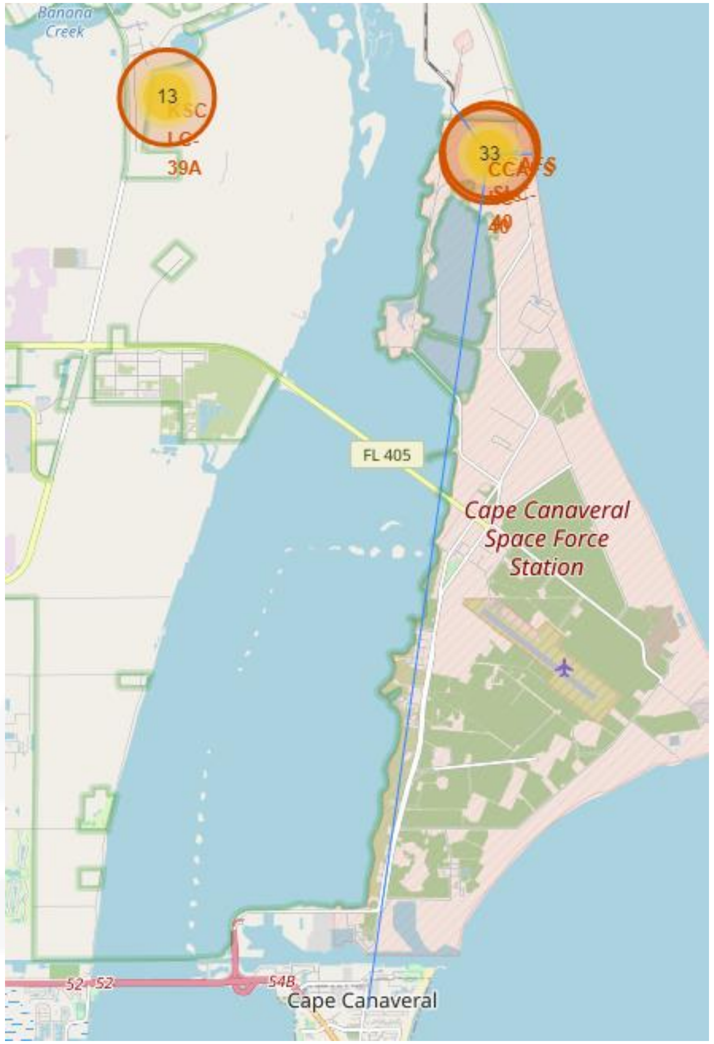
```
%sql select "Landing _Outcome", count("Landing _Outcome") as sr from (select * from SPACEXTBL where ("Landing _Outcome" like 'Success%') and (substr("Date",7,4)||
```

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

Landing _Outcome	sr
Success (drone ship)	5
Success (ground pad)	3

RESULTS – Folium

Locations of launch sites

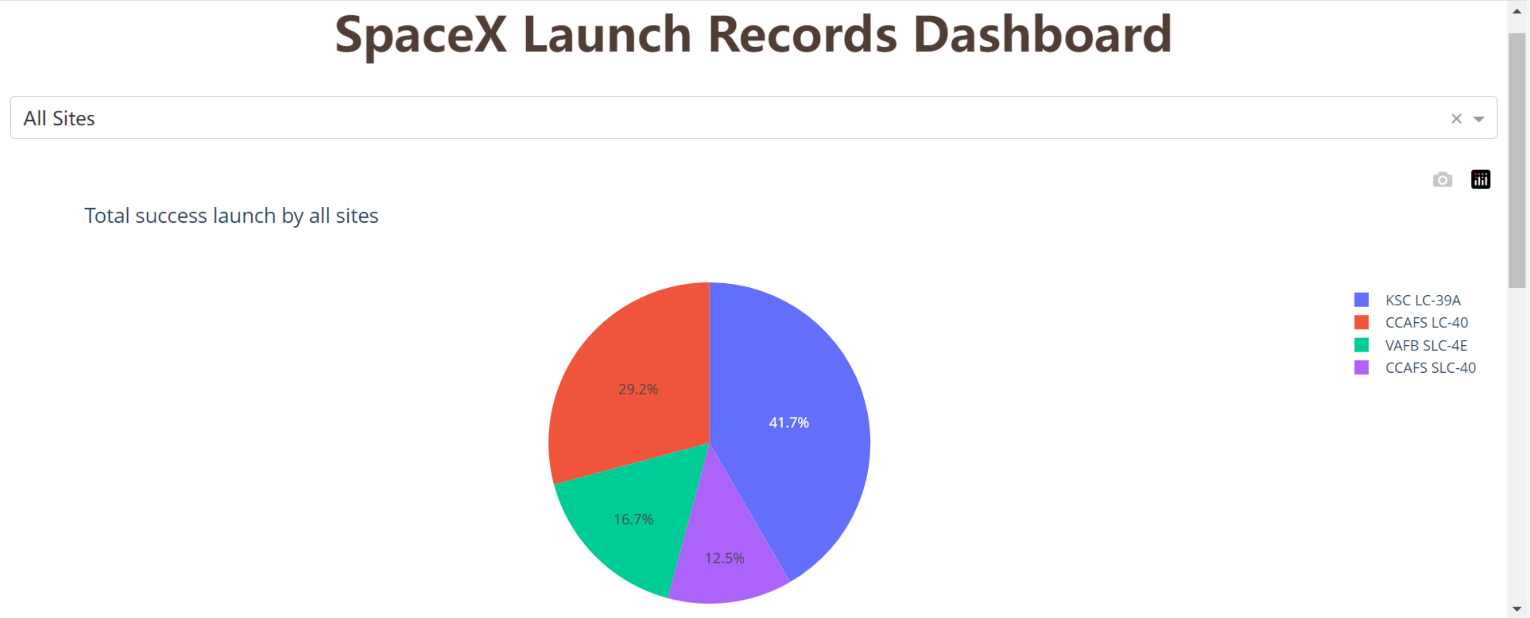


Analysis:

- Launch sites are usually located close to the coastline, railway, highway
  - potential factors: logistics convenience, collection of payloads landing unsuccessfully
- But launch sites are far away from cities
  - potential factors: consideration of safety, pollution, noise

RESULTS – Plotly Dash Dashboard

SpaceX Launch Records Dashboard



Analysis:

- KSCLC-39A occupies the largest proportion of successful launch
- CCAFS SLC-40 has the fewest successful launches

RESULTS – Plotly Dash Dashboard

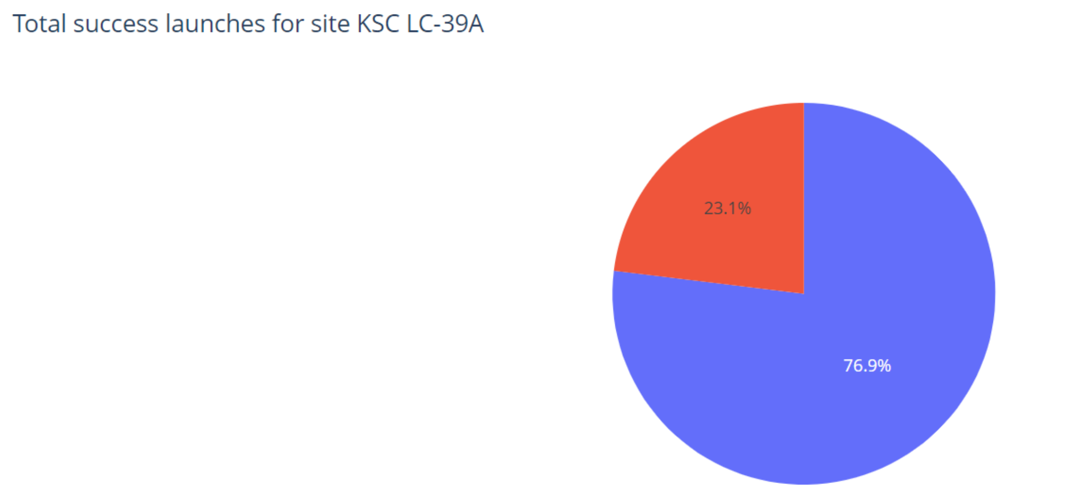
SpaceX Launch Records Dashboard

Analysis:

- Almost all payloads with v1.0 and v1.1 boosters end up with failure
- Most successes fall into the range between 2k kg and 6k kg

RESULTS – Plotly Dash Dashboard

Records for site KSC LC-39A



Analysis:

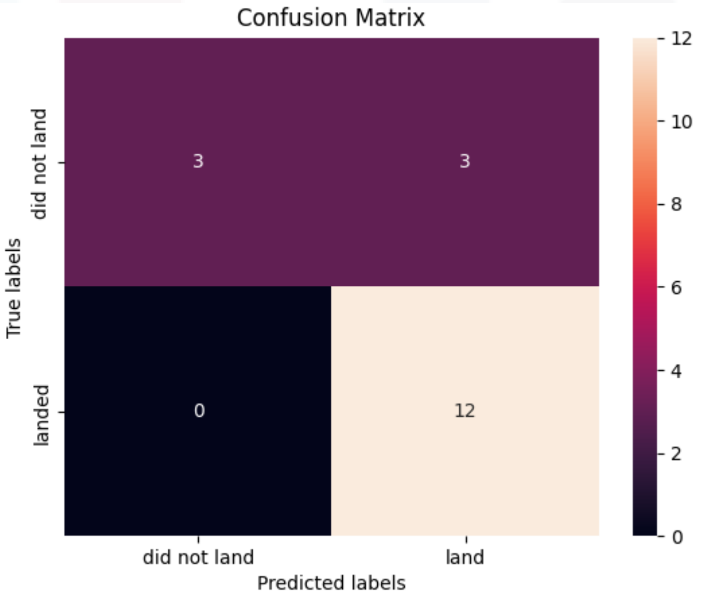
- The successful rate at site KSC LC-39A is 76.9%
- Payloads with over-5500-kg FT Boosters all failed in launching

RESULTS – Predictive Analysis

All four classification models have the same prediction performance on testing data (with accuracy of 83.3%)

```
print(logreg_cv.score(X_test,Y_test), \
      svm_cv.score(X_test,Y_test),\
      knn_cv.score(X_test,Y_test),\
      tree_cv.score(X_test,Y_test))
```

```
0.8333333333333334 0.8333333333333334 0.8333333333333334 0.8333333333333334
```



Analysis:

- We can select any of logistics regression, SVM, kNN, and decision tree methods to train the model
- The same performance among different predictors might be resulted from the fact that the amount of data sets is too small

DISCUSSION

OVERALL FINDINGS

- As time goes by, failure launches become less possible
- Launch sites are usually located near coastline, highway, and railway but distant from cities
- Launches tend to be successful with:
    - Larger flight number
    - Heavier payloads
    - Orbit at ES-L1, GEO, HEO, SSO
    - Boosters other than v1.0 and v1.1 boosters
- Logistic regression, SVM, decision tree, kNN can be applied to predict the results of a launch with similar performance

CONCLUSION

- A successful launch is related to multiple features
- Choose a right place to launch rockets
- Classification models can provide prediction with 83.3% accuracy
- Maybe more data is desired to improve the prediction accuracy

APPENDIX

- For more details of codes and figures, please kindly find below the Github repository: https://github.com/AnyueWang/Capstone