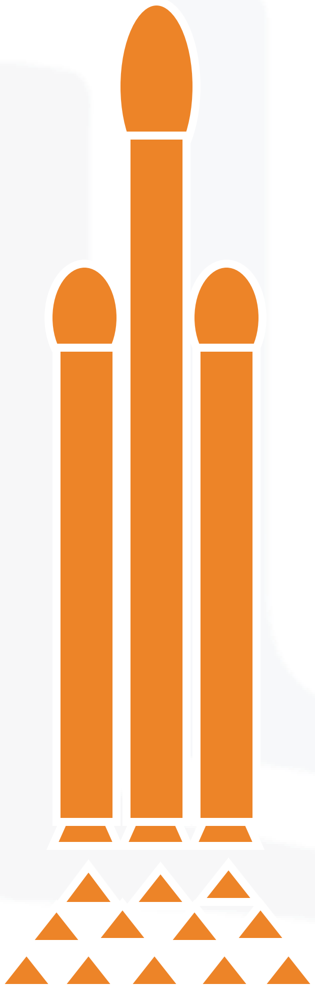


~~Rocket~~ Data Science: An Exploration of SpaceX Falcon 9 Rockets

A. Ryan Kutayiah

10/20/2023

IBM Developer

SKILLS NETWORK

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - SQL queries
  - Visualization
  - Dashboard
  - Predictive Analysis
- Conclusion

EXECUTIVE SUMMARY

- The Following Methodologies were implemented:
  - Data Collection
  - Data Wrangling (clean, organize, and transform raw data to usable data)
  - Exploratory Data Analysis (EDA for short, discovering patterns in the data)
  - Interactive Dashboards
  - Machine Learning (classification models)
- Summary of Results
  - Landing success increased with number of flights
  - Difficult to assess if landing success depended on payload mass
  - Over time the success rate of landings increased from 0% between 2010-2013 to over 50% by 2016 to over 80% by 2019
  - The most successful launches favored the following orbit types: ES-L1, GEO, and HEO. The least successful was SO followed by GTO and ISS.
  - The most successful launch site was KSC LC-39 A
  - The learning models comparable in performance

INTRODUCTION

- SpaceX achieved historic milestones by being the first private company to successfully land a rocket from low-earth orbit!
- Why is this a big deal?
  - SpaceX Falcon 9 rocket launches cost about $62 million
  - Other providers cost upward of 165 million dollars each
  - Much of the savings is because Space X can reuse the first stage.
- If we can determine if the first stage will land, we can determine the cost of a launch
  - We'll use the power of data science to predict the success of landings and use these findings to aid in the decision-making process.
- Let's explore some data!
  - …but first…methodology

METHODOLOGY

- We have to massage the insights from the data
- Like a massage therapist we need to procure a client, for us, this is data collection
  - Get Requests to REST API and Web Scraping with BeautifulSoup
- The client needs to be clean and transformed from their state of being clothed to discarding the unnecessary garb and positioned lying face down on the massage workstation, for us, we need to clean and transform data; data wrangling
  - Removed some NaN values with the .fillna() method of the Pandas DataFrame class
  - Create numerical Landing Outcome labels
- The therapist explores the client's muscle mass to find tension, likewise, we explore the data set looking for patterns
  - Use SQL, Pandas and Matplotlit to explore the data set
  - Use Feature Engineering and One-Hot Encoding to convert categorical variables to numerical ones
- The therapist might interact with the client by showing them where there is tension, so too, we build dashboards to visually interact with the data
  - Created interactive dashboards using Plotly Dash and Folium
- The therapist uses experience from past clients to predict areas of high tension and knows the best techniques to relieve the tension, analogously, we use the data to train machine learning models. Once the models are experienced, they too can make predictions
  - Used classification models such as Logistic Regression, Support Vector Machine, Decision Tree, and K-Nearest-Neighbor
  - Compare model performance using accuracy score

RESULTS: EDA with SQL (Summary)

- Launch Site Names
- Filter by Names beginning with CCA
- Total Payload Mass by NASA
- Average Payload Mass by SpaceX F9 v1.1
- First Successful Landing
- Successful Drone Ship Landings for a range of Payload masses
- Total Success and Failure Mission Outcome
- Names of Boosters carrying Maximum Payload
- Drone Ship Landing Failures for a certain Year

EDA SQL – Launch Sites

Unique Landing Sites

```
# get unique landing sites
%sql SELECT DISTINCT Launch_Site FROM SPACEXTABLE;
```

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

Landing Sites beginning with CCA

```
%%sql SELECT Launch_Site FROM SPACEXTABLE
WHERE Launch_Site LIKE 'CCA%'
LIMIT 5;
```

Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

EDA SQL – Payload Mass

Total Payload Mass - NASA

```
%%sql SELECT SUM(PAYLOAD_MASS__KG_)
AS Total_Payload_Mass FROM SPACEXTBL;
```

Total_Payload_Mass
619967

Boosters with Max Payload Mass - SpaceX

```
%%sql SELECT Booster_Version FROM SPACEXTABLE
WHERE PAYLOAD_MASS__KG_ = (SELECT MAX(PAYLOAD_MASS__KG_)
                           FROM SPACEXTABLE);
```

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

Average Payload Mass - SpaceX

```
%%sql SELECT AVG(PAYLOAD_MASS__KG_)
AS Avg_Payload_Mass_F9 FROM SPACEXTBL
WHERE Booster_Version LIKE 'F9 v1.1';
```

Avg_Payload_Mass_F9
2928.4

EDA SQL – Landing Outcomes

First Successful Ground Pad Landing

```
%%sql SELECT MIN(date) FROM SPACEXTABLE
WHERE Landing_Outcome = 'Success (ground pad)';
```

MIN(date)
2015-12-22

Boosters with Successful Drone Ship Landing and Payload Mass between 4000 and 6000 kg

```
%%sql SELECT Booster_Version FROM SPACEXTABLE
WHERE Landing_Outcome = 'Success (drone ship)'
AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
```

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

EDA SQL – Landing Outcomes

Ground Pad Failures in 2015

```
%%sql SELECT STRFTIME("%m", Date) AS 'month',
STRFTIME("%Y", Date) AS 'year',
Landing_Outcome, Booster_Version, Launch_Site
FROM SPACEXTBL
WHERE year = '2015'
AND Landing_Outcome = 'Failure (drone ship)';
```

month	year	Landing_Outcome	Booster_Version	Launch_Site
10	2015	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	2015	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Landing Outcomes between 2010-2017

```
%%sql SELECT Landing_Outcome, COUNT(Landing_Outcome)
AS 'Outcomes' FROM SPACEXTBL
WHERE STRFTIME("%d-%m-%Y", Date) BETWEEN '04-06-2010'
AND '20-03-2017'
GROUP BY Landing_Outcome
ORDER BY 'Outcomes' DESC;
```

Landing_Outcome	Outcomes
Success (ground pad)	7
Success (drone ship)	8
Success	20
No attempt	1
No attempt	9
Failure (parachute)	2
Failure (drone ship)	3
Failure	3
Controlled (ocean)	2

EDA SQL – Mission Outcomes

Total Successes and Failures

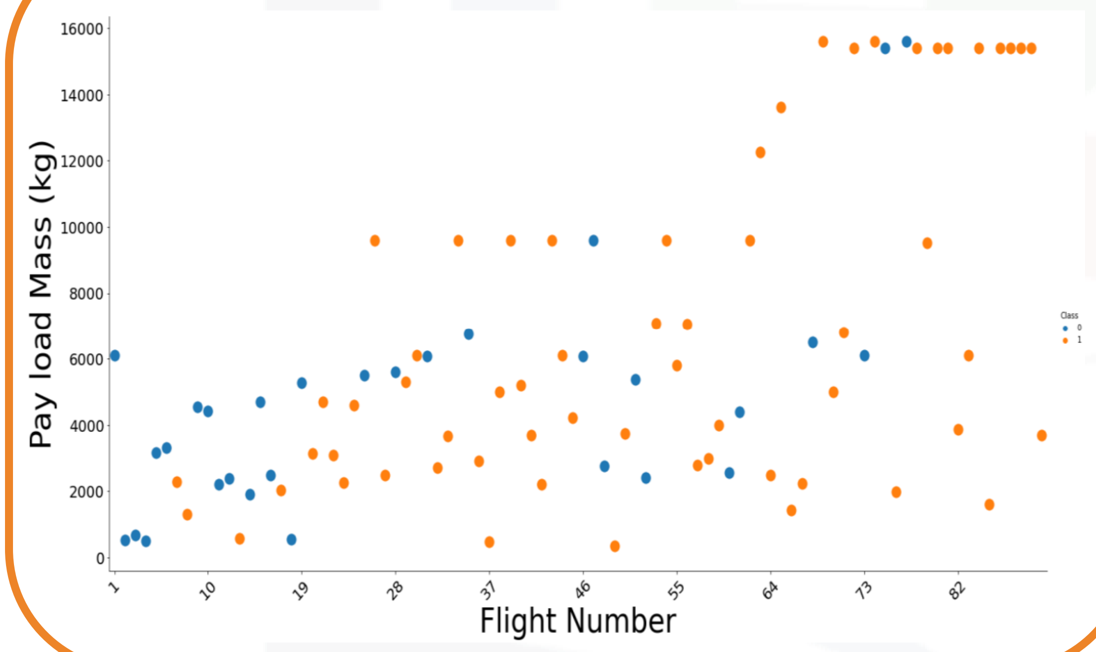
```
%%sql SELECT COUNT(Mission_Outcome), Mission_Outcome
FROM SPACEXTABLE
GROUP BY Mission_Outcome;
```

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

RESULTS: EDA with Visualization (Summary)

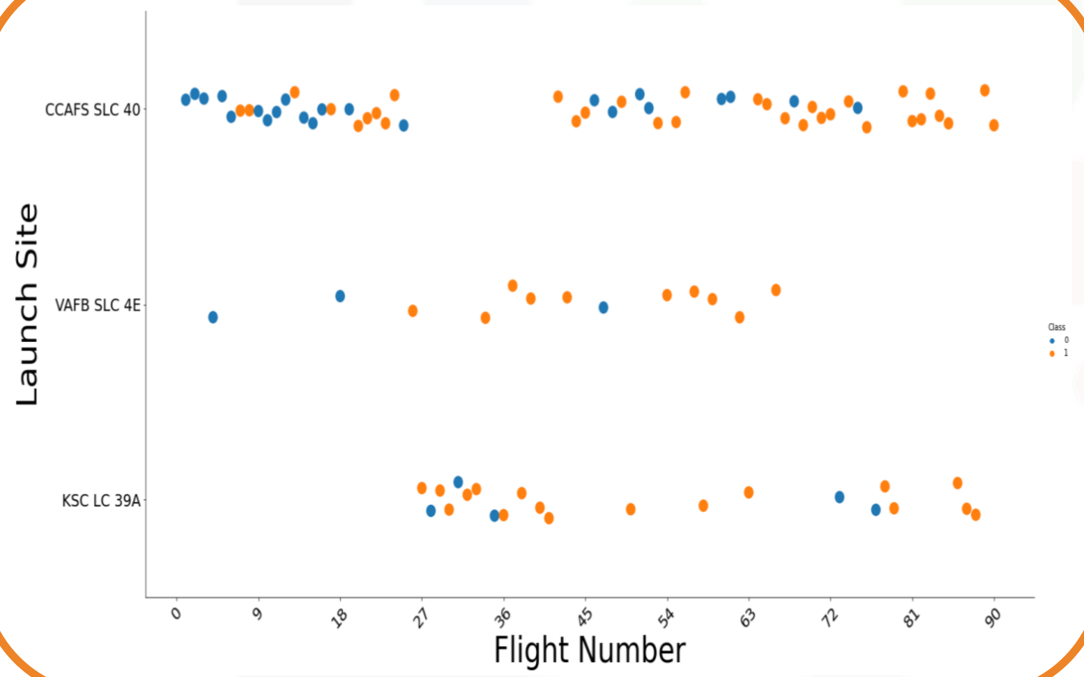
- Payload Mass vs Flight Number
  - Scatter Plot
- Launch Site vs Flight Number
  - Scatter Plot
- Launch Site vs Payload Mass
  - Scatter Plot
- Success vs Orbit Type
  - Bar Chart
- Orbit Type vs Flight Number
  - Scatter Plot
- Orbit Type vs Payload Mass
  - Scatter Plot
- Success vs Time
  - Line Plot

EDA Visual - Payload vs Flight Number



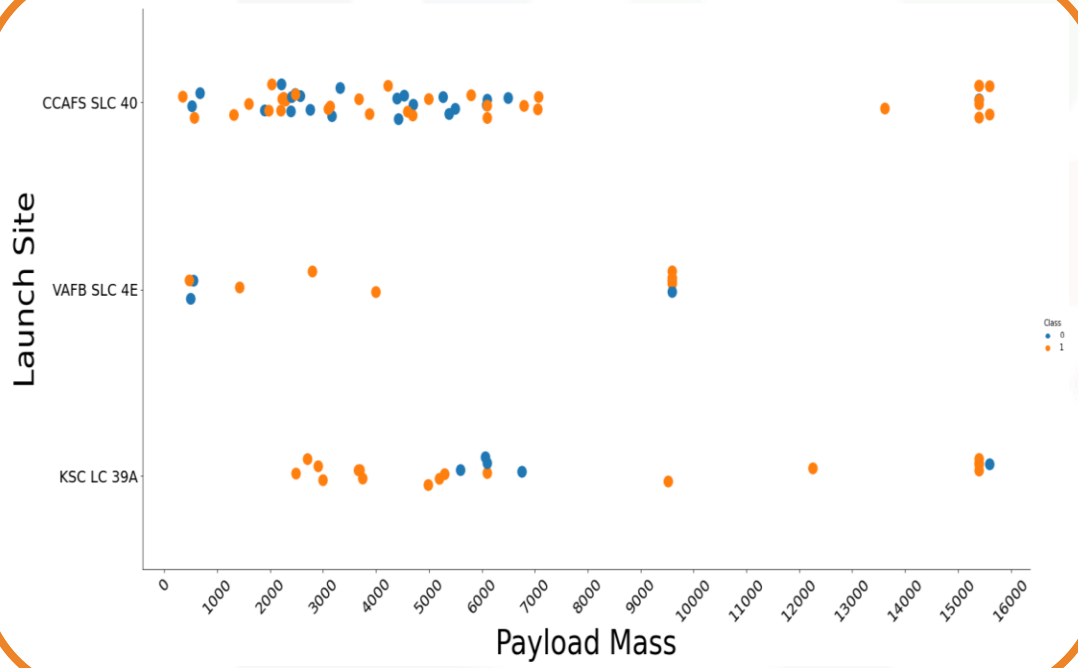
- Blue dots = failure, orange dots = success
- As the flight number increases, the first stage is more likely to land successfully.

EDA Visual - Launch Site vs Flight Number



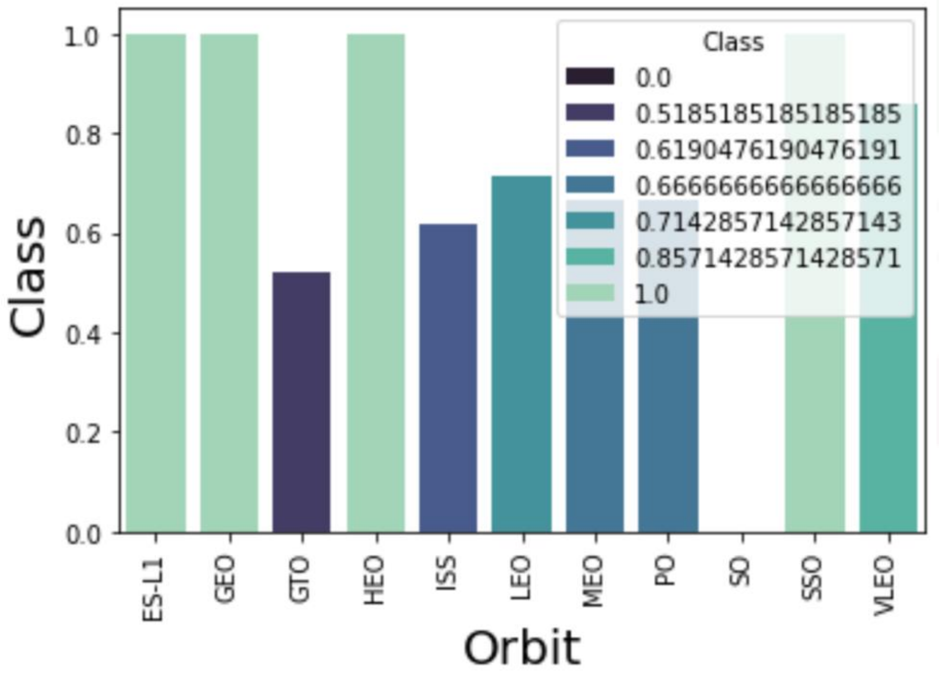
- Blue dots = failure, orange dots = success
- As the flight number increases, the first stage is more likely to land successfully at each site.
  - For CCAFS SLC 40 and VAFB SLC 4E we see more failures for flight numbers less than 35 and more success above this number
  - KSC LC 39A does not show any record of having very early flights as its flight number starts in the 20s. But we see more failures clustered around the low flight number end.

EDA Visual - Launch Site vs Payload



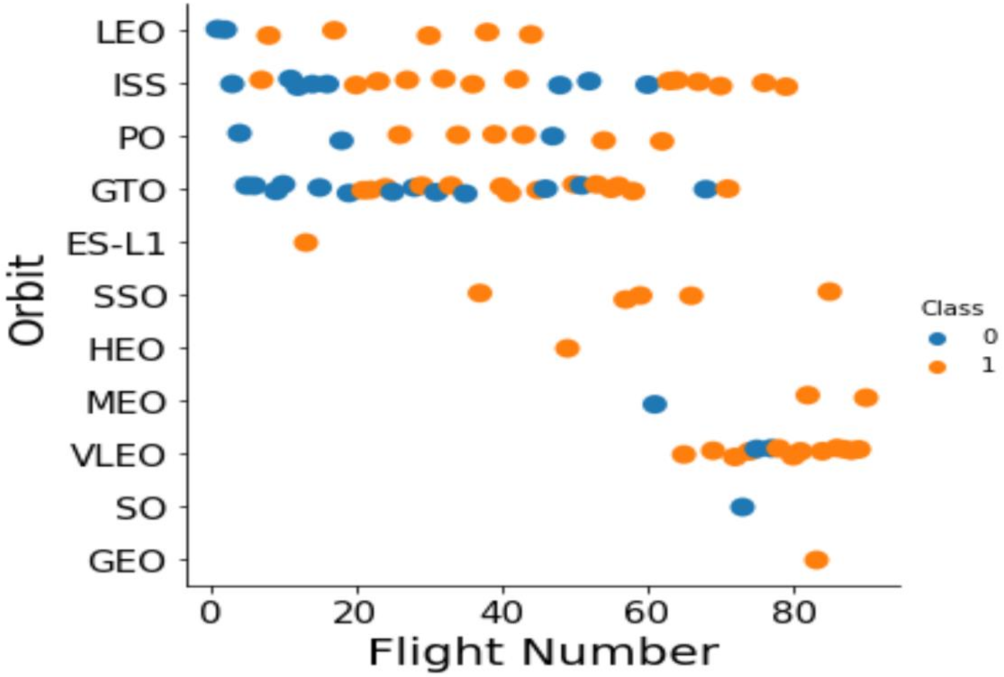
- Blue dots = failure, orange dots = success
- While there are more successes at higher masses for each launch site, there does not seem to be a strong relationship between payload mass and the success rate for each site
  - For CCAFS SLC 40 there seem to be as many failures as success below 8000 kg and more success between 12000 to 16000 kg
  - For KSC LC 39A there is more success in the low mass region between 2000 and 5000 kg with failures clustered around 6000-7000 kg region

EDA Visual - Success vs Orbit Type



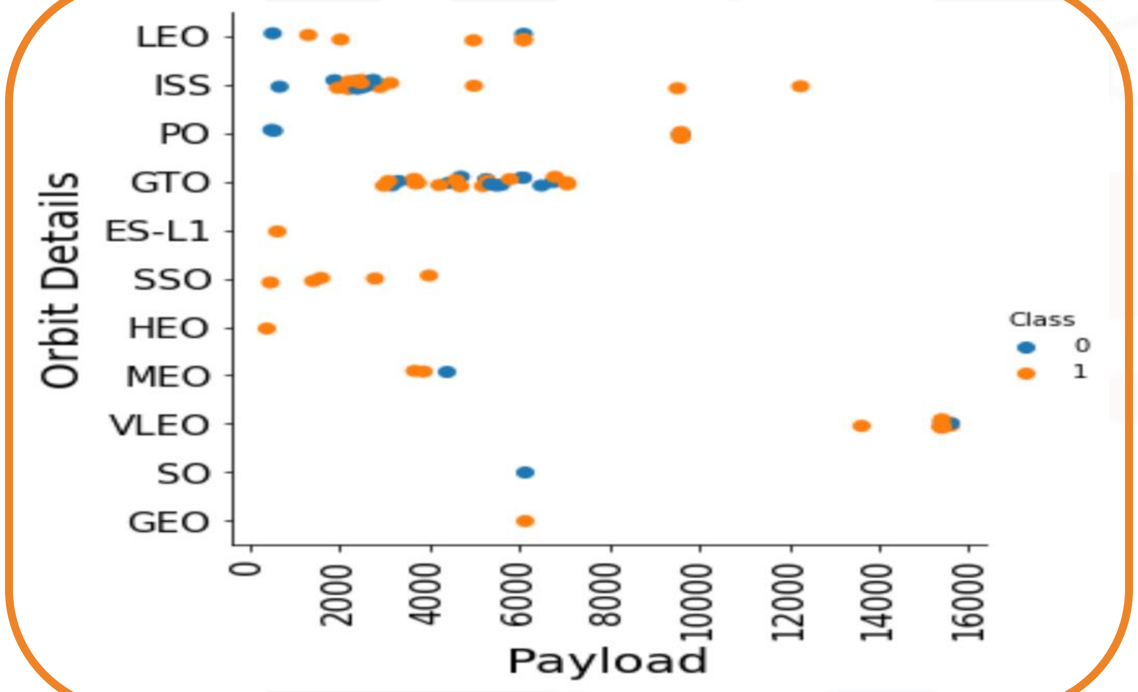
- The least successful orbits:
  - SO (0%)
  - GTO (52%)
- The most successful orbits:
  - ES-L1 (100%)
  - GEO (100%)
  - HEO (100%)
  - SSO (100%)
- The other orbits have a success rate between 51.9% to 85.7%

EDA Visual - Orbit vs Flight Number



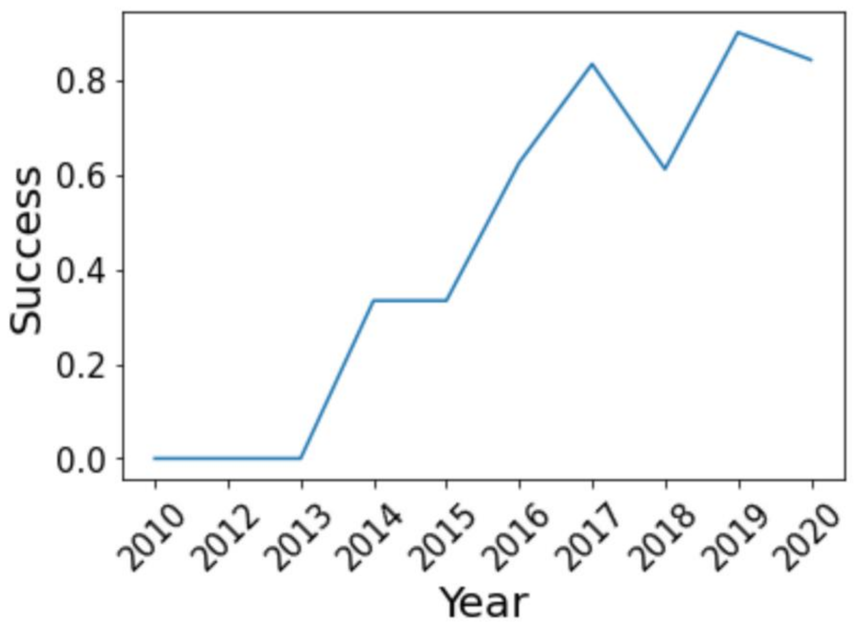
- It is not clear how flight number affects the success rate for the orbits
  - LEO orbit indicates a greater success rate with increasing flight number
  - Whereas, GTO shows about as many successes as failures across flight numbers
  - ISS has clumps of success and failures at low and high flight numbers
  - We cannot definitively conclude that flight number predicts the success or failure of a given orbit

EDA Visual - Orbit vs Payload



- It is not clear how payload mass affects the success rate for the orbits
  - Much like the Orbit vs Payload scatter plot, the results are mixed.

EDA Visual - Success over the years

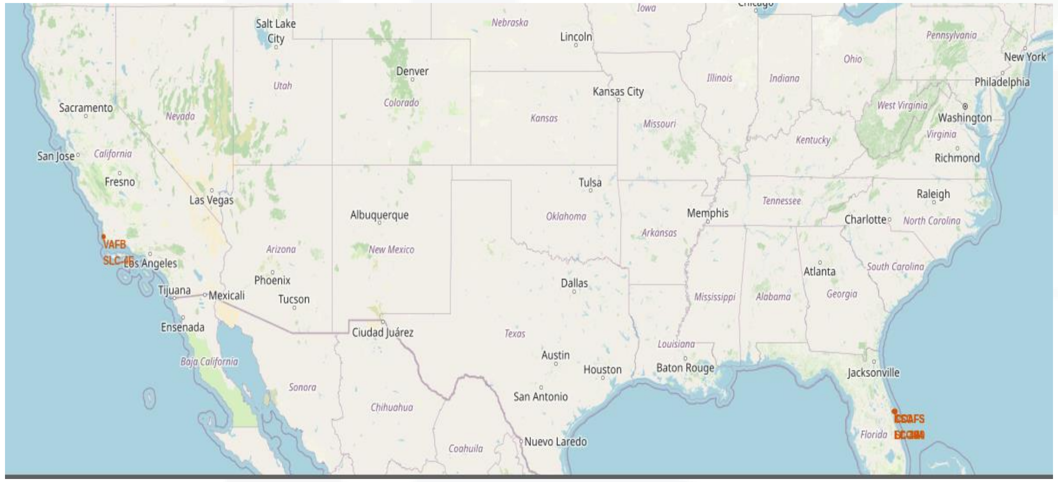


- Overall the success rate of the rocket landings increased over time
  - There were no successes between 2010-13
  - There was a sharp increase in success between 2013 and 2014 from 0% to more than 30% (but less than 40%) with a plateau between 2014-15.
  - From 2015-2017 there was an increase in success from <40% to >80%
  - There was a dip in success at 2018 where the success rate fell to approximately 60%
  - The rate increased again at 2019 to >80%

RESULTS: Interactive Dashboard

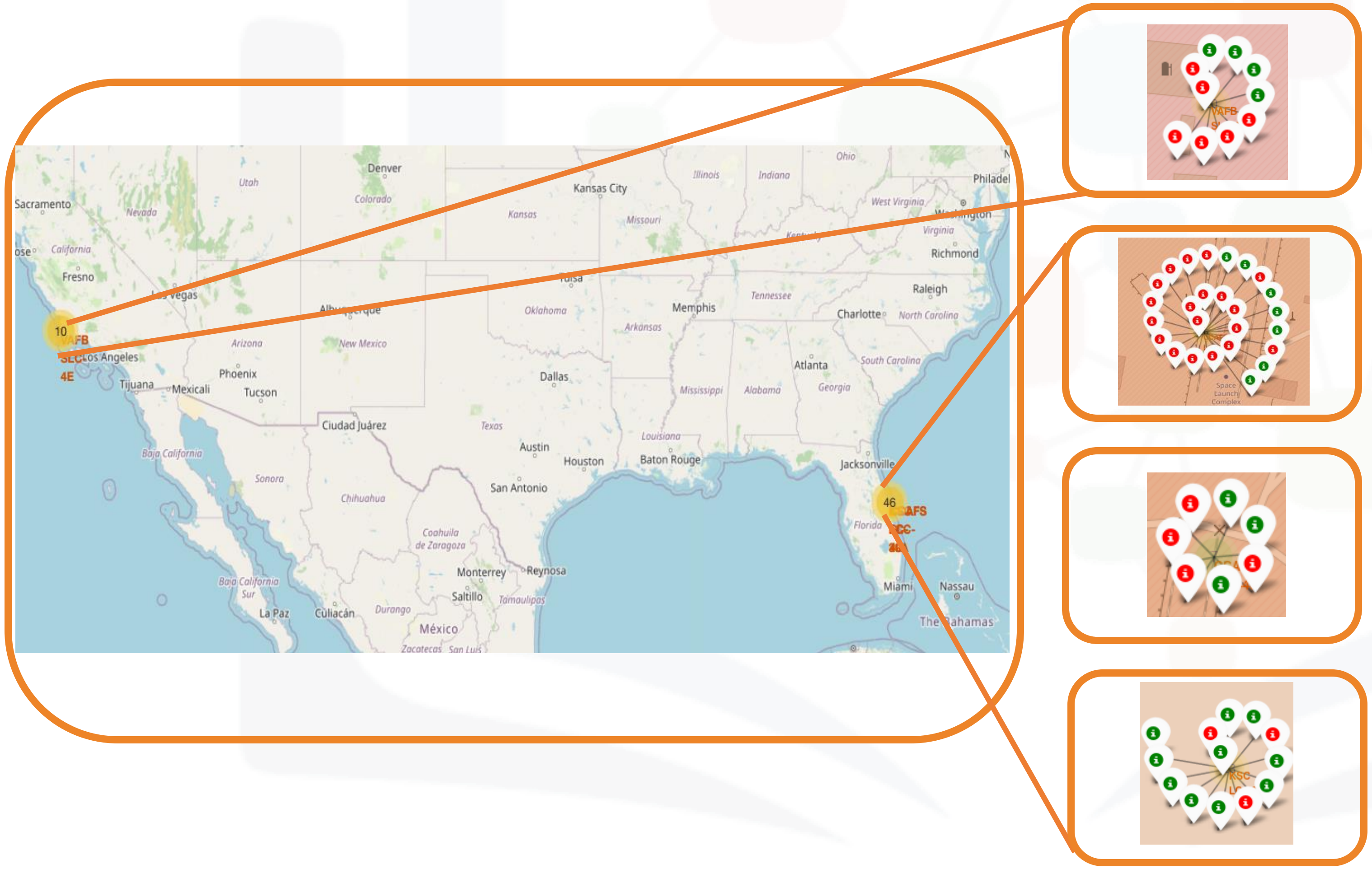
- Folium for Geospatial Data
  - Identify Launch Sites on Map
  - Indicate Success/Failure for each Launch Site
  - Calculate Launch Site Proximity to Highways, Railroads, Cities and Coastline
- Plotly Dash for Interactive Dashboard
  - Pie Chart Displaying Total Successful Launches Based on Launch Site Selection from a Dropdown menu
  - Scatter plot of Landing Outcome based on Payload Mass Selection from a Slider

Folium – Launch Sites



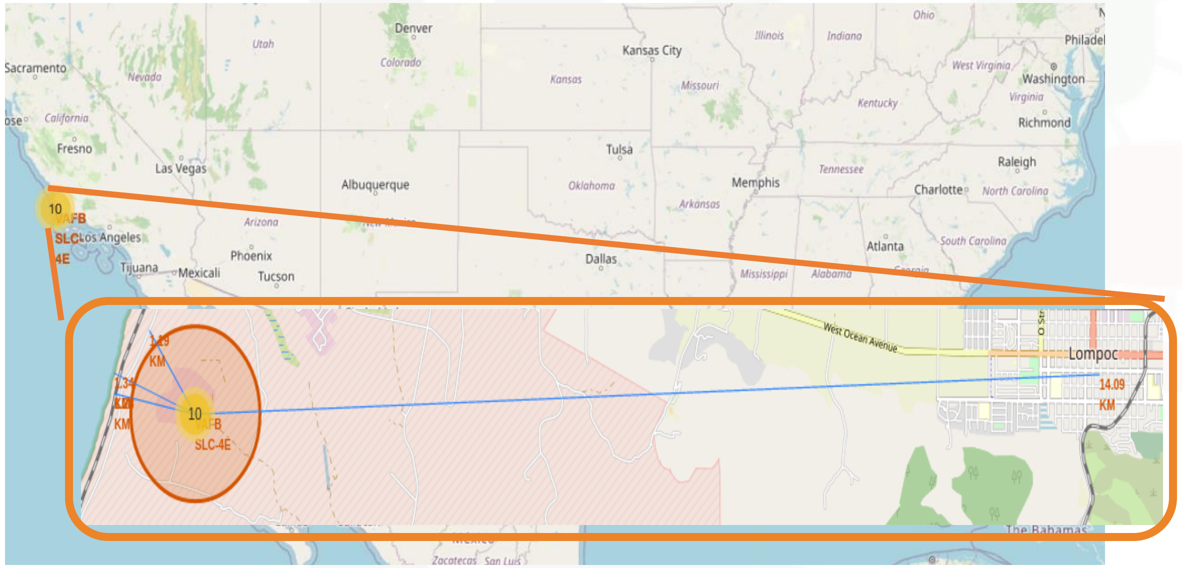
- Launch Sites in red
- On the East Coast (Florida) and West Coast (California) of the United States of America

Folium – Launch Sites Outcomes



- Red = Failure
- Green = Success

Folium – Launch Sites Proximities

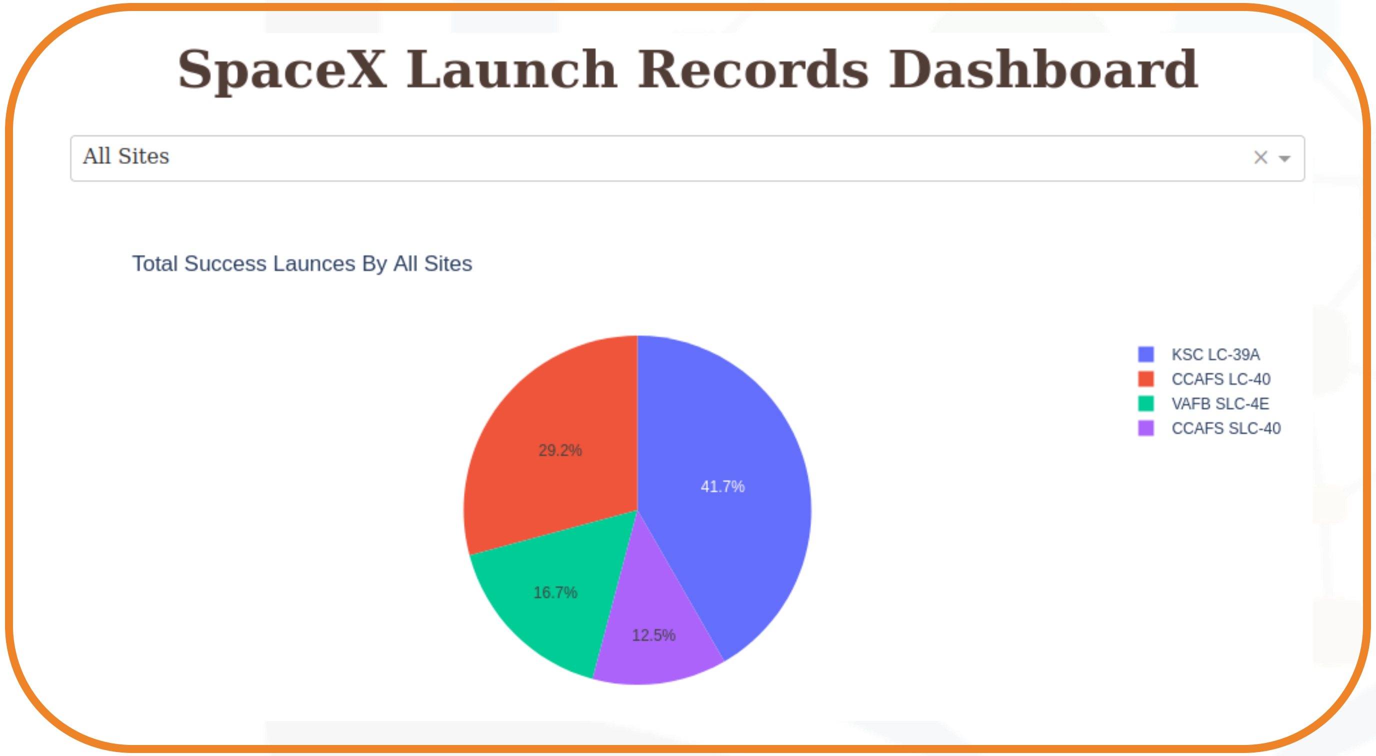


- Less than 1.5 km from
  - Coast
  - Railroad
  - Highway
- About 14 km from city

IBM Developer

SKILLS NETWORK

Plotly – Launch Sites Success

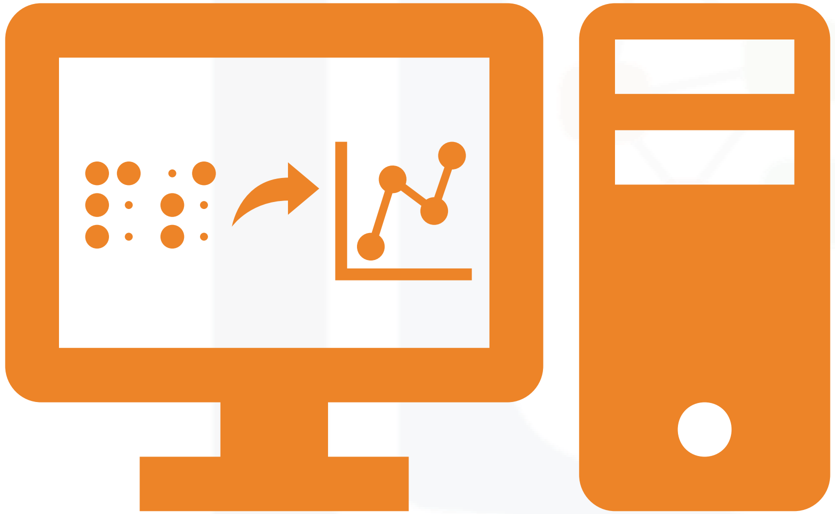


- Can select each site from dropdown
- KSC LC-39A highest success rate
- CCASF SLC-40 lowest success rate

Plotly – Launch Sites Success

- Very little data for masses beyond 7k
- More failure than successes below 2k and beyond 6k
- Above 2k but below 6k there seems to be about as much success as failure

RESULTS: Predictive Analysis (Summary)



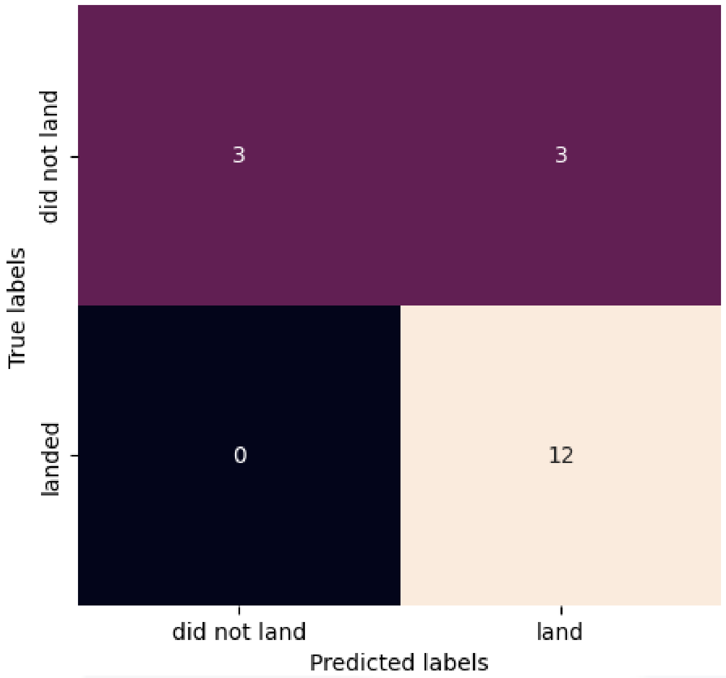
- Classification Models
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - K-Nearest-Neighbor
- Compared models using the accuracy score
- Examined Confusion Matrix for False Positives and False Negatives

Predictive Analysis – Accuracy Score

	Model	Train Set Accuracy Score	Test Set Accuracy Score
0	Logistic Regression	0.846429	0.833333
1	Support Vector Machine	0.848214	0.833333
2	Decision Tree	0.876786	0.833333
3	K-Nearest-Neighbor	0.848214	0.833333

- All models had similar accuracy scores on the test set
- Decision Tree had a higher score on the train set compared to other models
- Bases on the accuracy scores on the train and test set all models had similar performance

Predictive Analysis – Confusion Matrix



- All Models had identical confusion matrices
- 16.7% (3 of 18) false positives
  - True Label = did not land and Predicted Label = landed
- 0% (0 of 18) false negatives
  - True Label = landed and Predicted Label = did not land

CONCLUSION

- Landing success increased with number of flights
- Difficult to assess if landing success depended on payload mass
- Over time the success rate of landings increased from 0% between 2010-2013 to over 50% by 2016 to over 80% by 2019
- The most successful launches favored the following orbit types: ES-L1, GEO, and HEO. The least successful was SO followed by GTO and ISS.
- The most successful launch site was KSC LC-39 A
- The learning models comparable in performance

Thank you!

Thank you for your time and attention!