

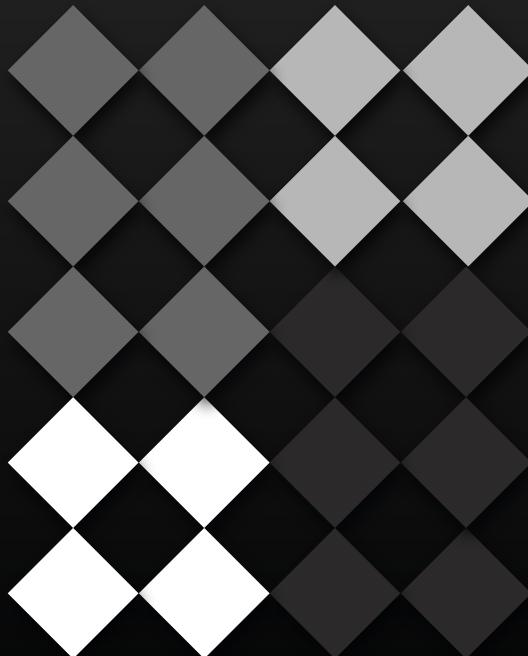
# Applied Data Science

Macarena Sagredo  
April 9 2022

# OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
- Discussion
- Findings & Implications
- Conclusion

# EXECUTIVE SUMMARY



## Methodologies

- Data Collection: API, SQL, Web Scraping
- Data Wrangling and Analysis
- Interactive Maps with Folium
- Predictive Analysis



## Results

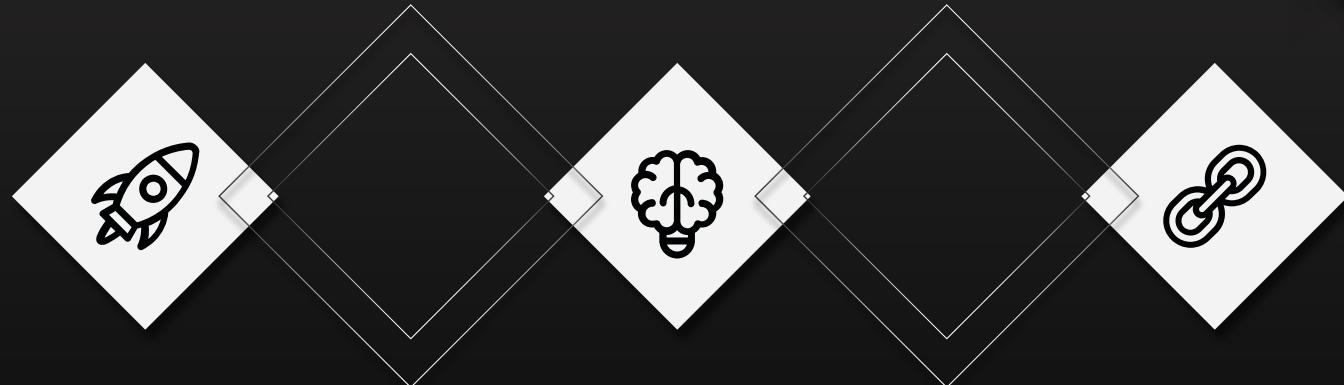
- Data Analysis with Interactive Visualizations
- Best Model for Predictive Analysis

# INTRODUCTION

- ◊ Space exploration is too expensive!
  
- ◊ Falcon 9 reduces the cost.
  - Reusing important parts.
  - Recovering Stage One.

# METHODOLOGY

# Data Collection:



API

SQL

Web  
Scraping

# Data Wrangling:

- ◆ Data was assigned a value depending on outcome.
- ◆ Was recovery successful?

True ASDS → Successful = 1

True RTLS → Successful = 1

True Ocean → Successful = 1

None None → Failed = 0

None ASDS → Failed = 0

False ASDS → Failed = 0

False RTLS → Failed = 0

False Ocean → Failed = 0

# EDA with Data Visualization:

- Flight Number vs. Payload
- Flight Number vs. Launch Site
- Launch Site vs. Payload
- Success Rate vs. Orbit type
- Orbit type vs. Flight Number
- Orbit type vs. Payload
- Success rate vs. Time in years

# EDA with SQL:

- SQL allows for complicated queries.
- Queries:
  - Launch Site
  - Payload Mass (kg)
  - Mission Outcome
  - Booster Version
  - Date

# Interactive Maps with Folium:

- Folium is a Python package.
- We built an interactive map with:
  - The Falcon 9 launch site.
  - Successful launches per location.
  - Failed launches per location.
  - Distances from nearest coastline, city, railway and highway.

# Dashboard with Plotly Dash:

- Pie chart showing the proportion of successful recoveries to unsuccessful ones for each site.
- Recovery Outcome vs. Payload Mass scatter plot with a range (0-10000kg) with bounds that can be changed by the user.
- Provides the relationships of launch sites' and payload masses' with their recovery outcomes.

# Predictive Analysis:

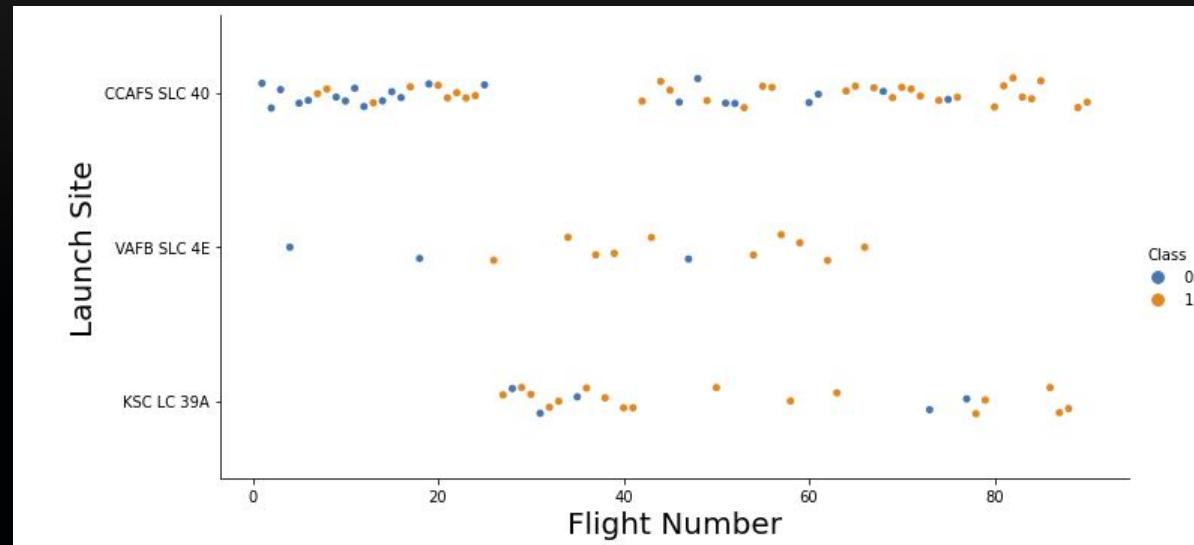
- Models:
  - Logarithmic Regression
  - Support Vector Machine
  - Decision Tree
  - KNN
- Steps:
  - Standardize data
  - Split into train/test
  - Find best parameters
  - Fit models with these parameters
  - Test models, compute accuracy and plot confusion matrix

# Results

# EDA Data Visualization

Rate of success increases over time

Launch Site/Flight Number

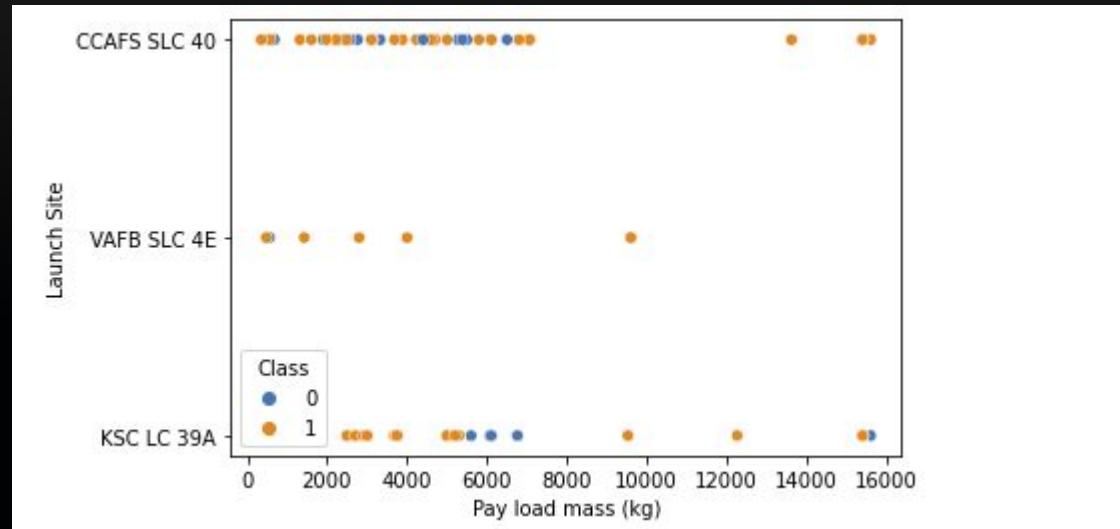


# EDA Data Visualization

VAFB SLC 4E has higher success rates, and lower payloads.

A correlation can be inferred

Launch Site/Payload Mass

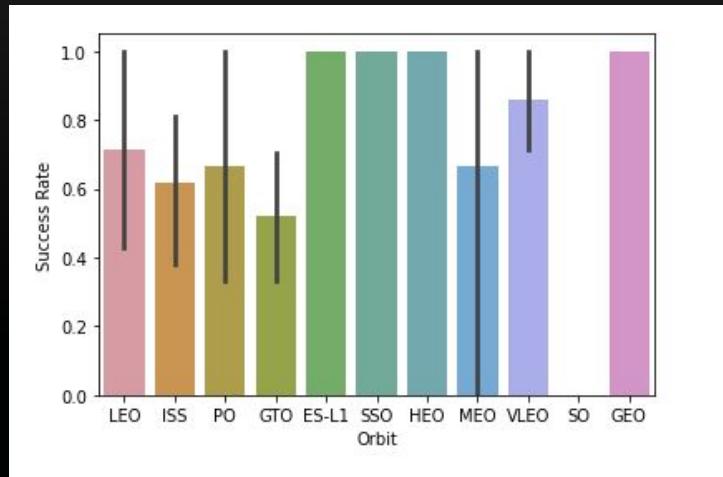


# EDA Data Visualization

Orbits with higher success rate

- ES-L1
- SSO
- HEO
- GEO

Success Rate/Orbit



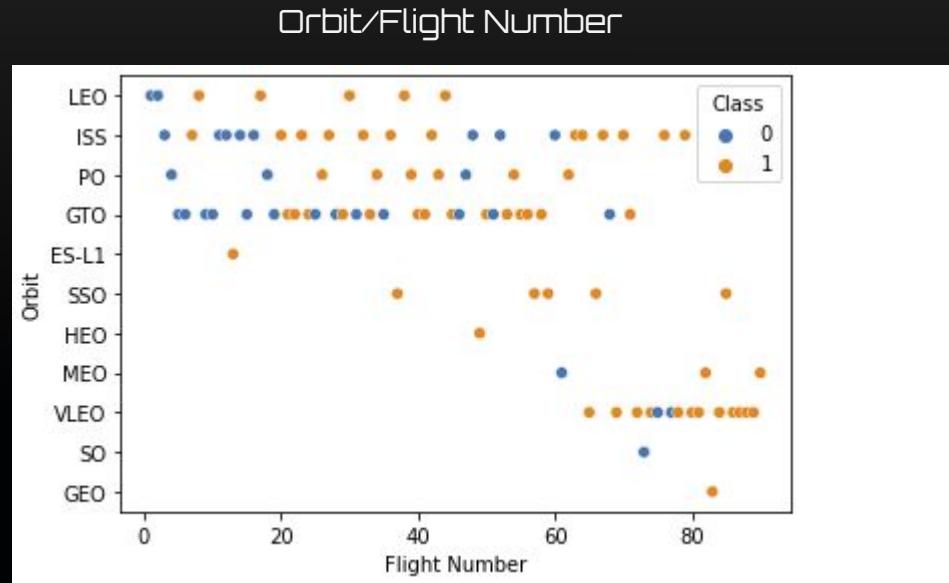
# EDA Data Visualization

Big sample size

- SSO
- LEO
- VLEO

Small Sample Size

- HEO
- GEO
- ES-I1

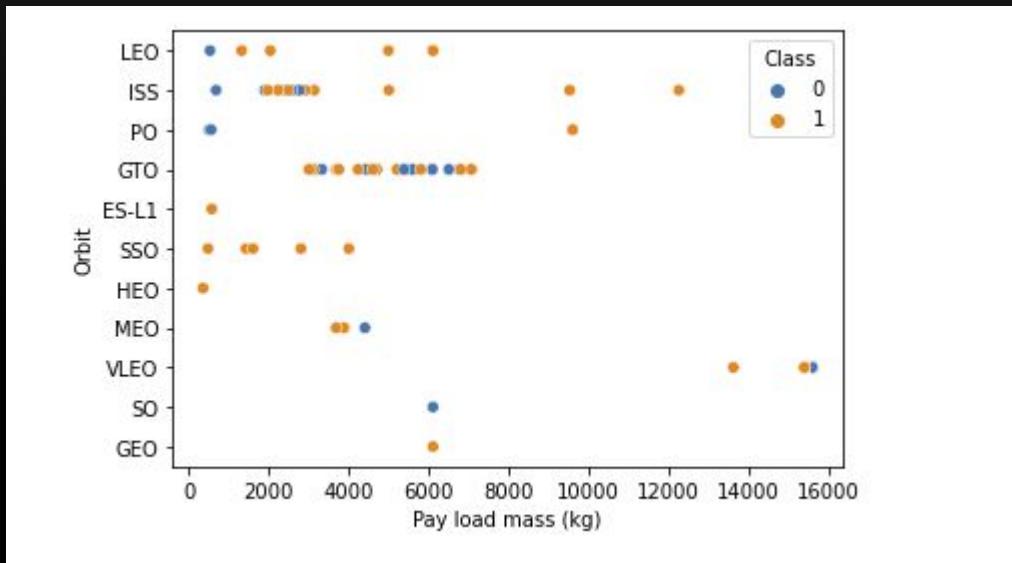


# EDA Data Visualization

Success correlated to light payload

- SSO
- LEO

Orbit/Payload Mass



# EDA Data Visualization

- Create Dummies

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

# EDA Data Visualization

- Cast all Numeric

	FlightNumber	PayloadMass	Flights	GridFins	Reused	Legs	Block	ReusedCount	Orbit_ES-L1	Orbit_GEO	...	s
0	1.0	6104.959412	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0
1	2.0	525.000000	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0
2	3.0	677.000000	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0
3	4.0	500.000000	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0
4	5.0	3170.000000	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0
...	...	...	...	...	...	...	...	...	...	...	...	...
85	86.0	15400.000000	2.0	1.0	1.0	1.0	5.0	2.0	0.0	0.0	...	0
86	87.0	15400.000000	3.0	1.0	1.0	1.0	5.0	2.0	0.0	0.0	...	0
87	88.0	15400.000000	6.0	1.0	1.0	1.0	5.0	5.0	0.0	0.0	...	0
88	89.0	15400.000000	3.0	1.0	1.0	1.0	5.0	2.0	0.0	0.0	...	0
89	90.0	3681.000000	1.0	1.0	0.0	1.0	5.0	0.0	0.0	0.0	...	0

# EDA SQL

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

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

# EDA SQL

Display 5 records where launch sites begin with the string 'CCA'

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

# EDA SQL

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

total_payload
111268

# EDA SQL

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

avg_payload
2534

# EDA SQL

List the date when the first successful landing outcome in ground pad was achieved.

first\_success

2010-06-04

# EDA SQL

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

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

# EDA SQL

List the total number of successful and failure mission outcomes

total_success
100

# EDA SQL

List the names of the booster\_versions which have carried the maximum payload mass.  
Use a subquery

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

# EDA SQL

List the failed landing\_outcomes in drone ship, their booster versions, and launch site names for in year 2015

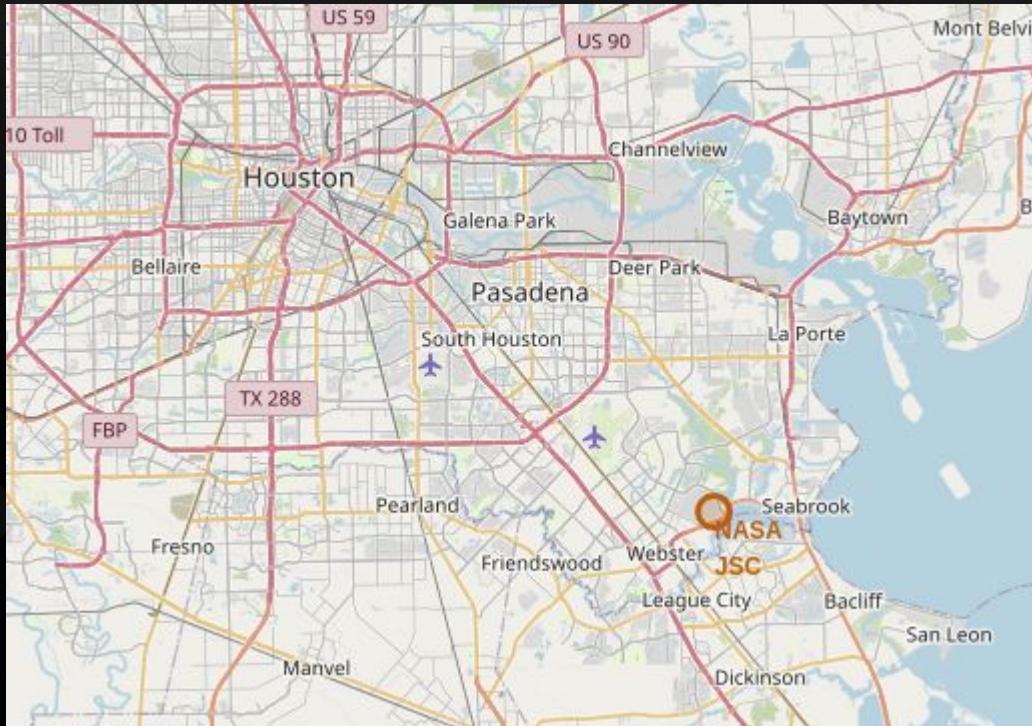
DATE	landing_outcome	booster_version	launch_site
2015-01-10	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

# EDA SQL

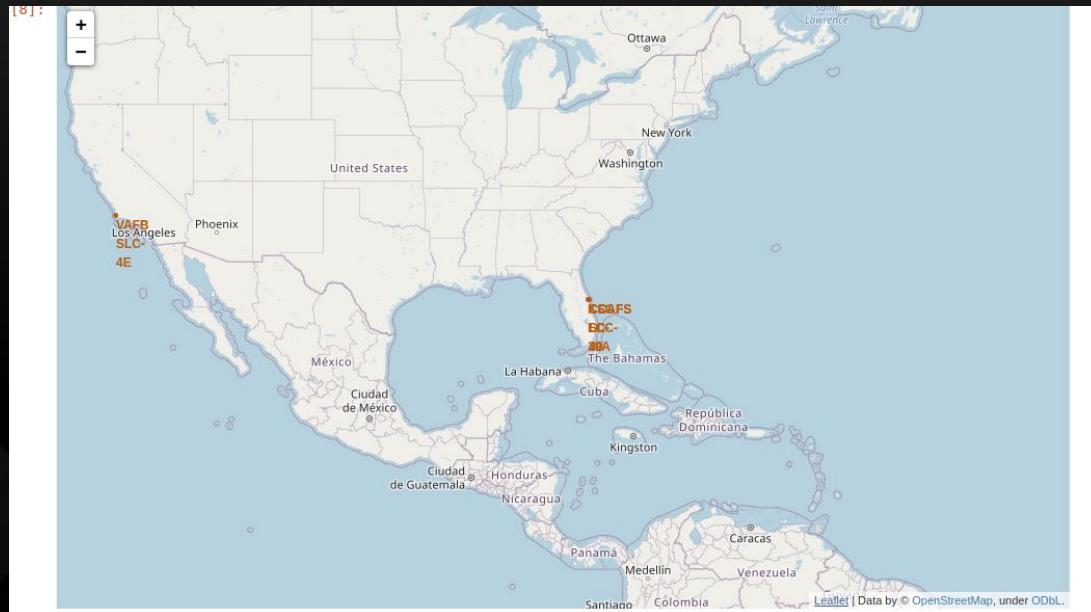
Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

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

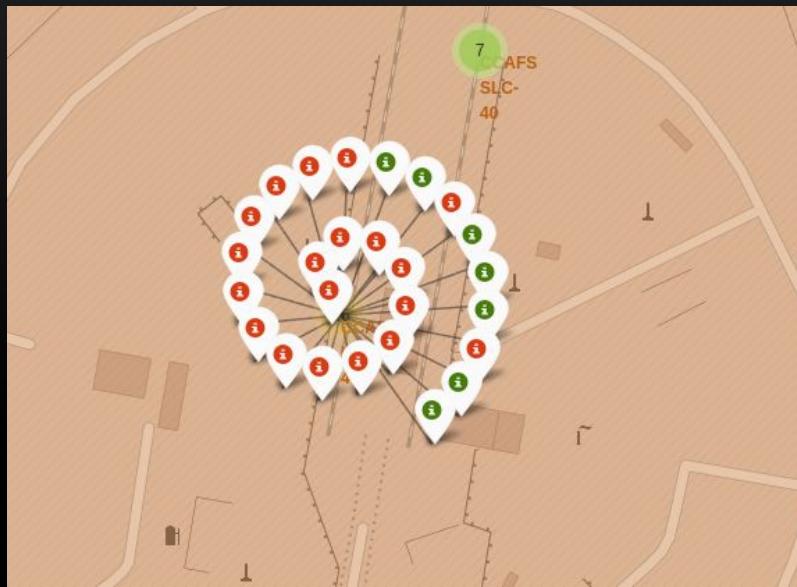
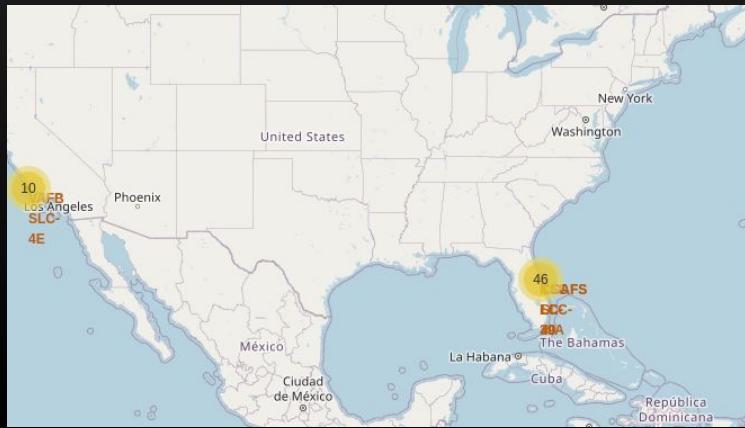
# Interactive Map



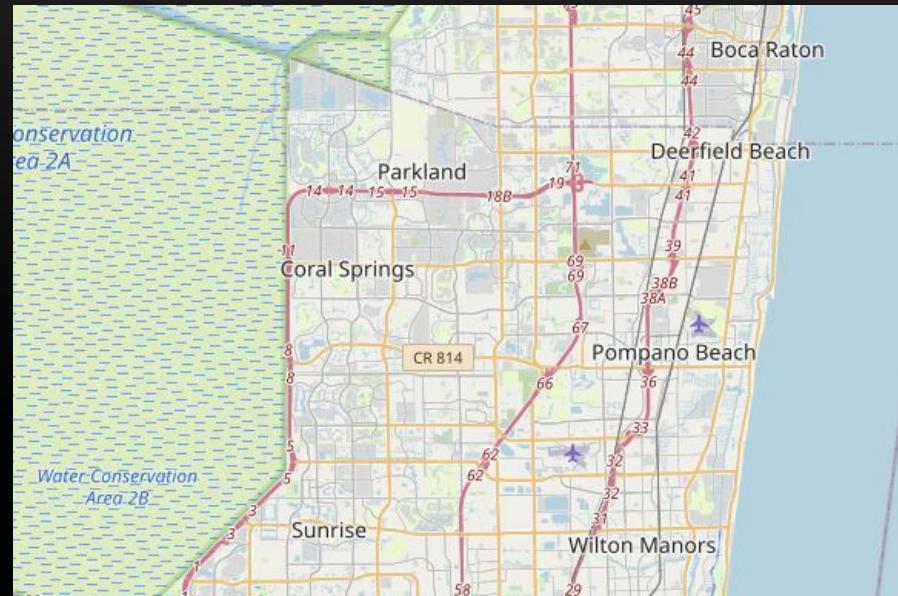
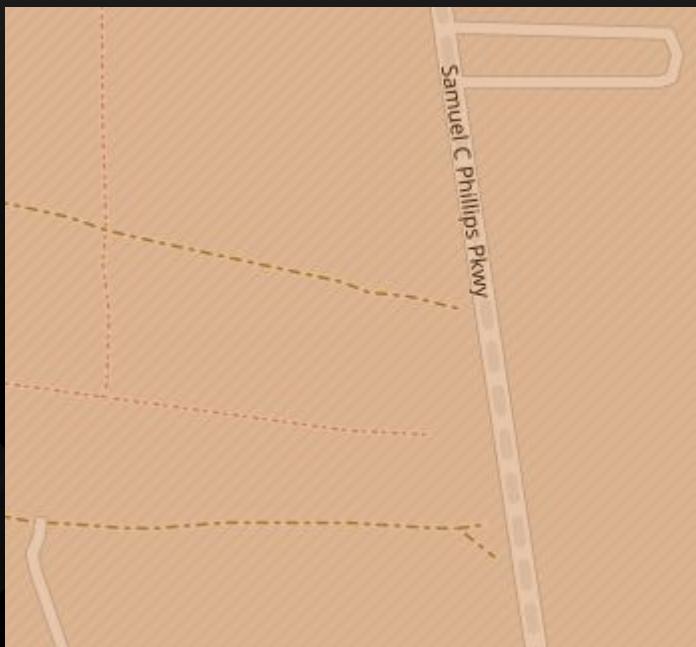
# Interactive Map



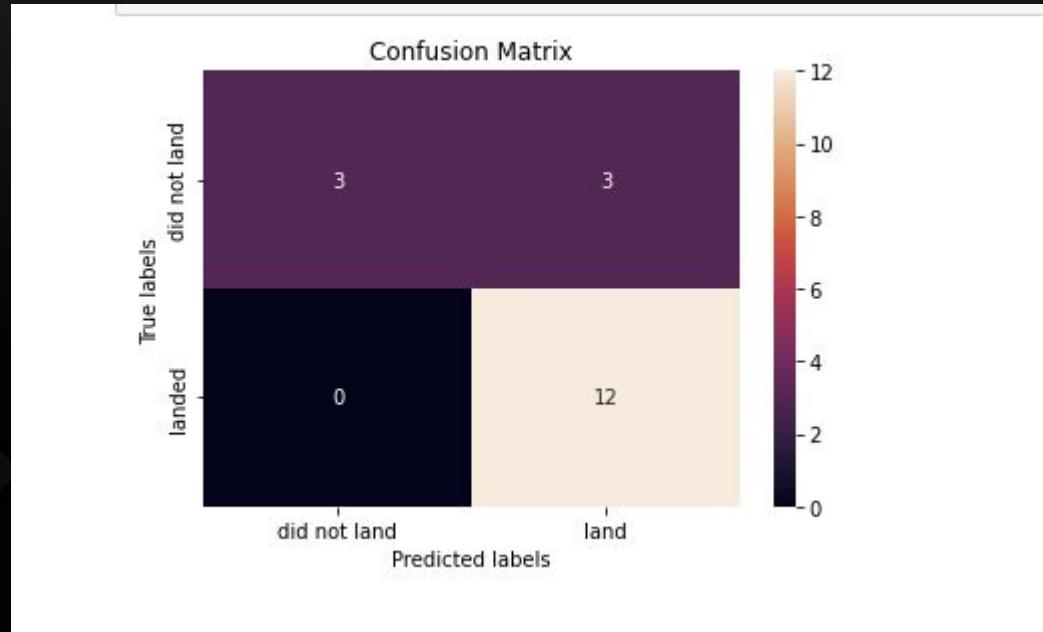
# Interactive Map



# Interactive Map



# Ploty



# Conclusions

SpaceX's successful recoveries generally have the following properties:

- A launch date in the year 2017 or later
- Light payload (in the range 2000-4000kg)
- Launched from site KSC LC-39A
- Successfully recovered via drone ship

Our model can predict the outcome of a given recovery with a reasonable degree of accuracy, 83.33%

CREDITS: This presentation template was created  
by **Slidesgo**, including icons by **Flaticon** and  
infographics & images by **Freepik**