IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Tim Holzmann
January 30, 2025

Outline

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

Executive Summary

- Analysis of SpaceX Launch Data
  - Data collection & Wrangling
  - Exploratory & interactive data analysis (EDA) with Plots, SQL, Folium, Plotly Dash
  - Predictive analysis using classification models

- Key Findings
  - 4 Launch Sites
  - Decision Tree model is best classifier for this data
  - Hard to launch heavy payloads

Introduction

- Space Y
  - Founded by Billionaire industrialist Allon Musk.
  - Competing with Space X
- Key Questions:
  - Price of each launch.
  - Likelihood of successful launch and first-stage reuse

SPACE Y

Section 1

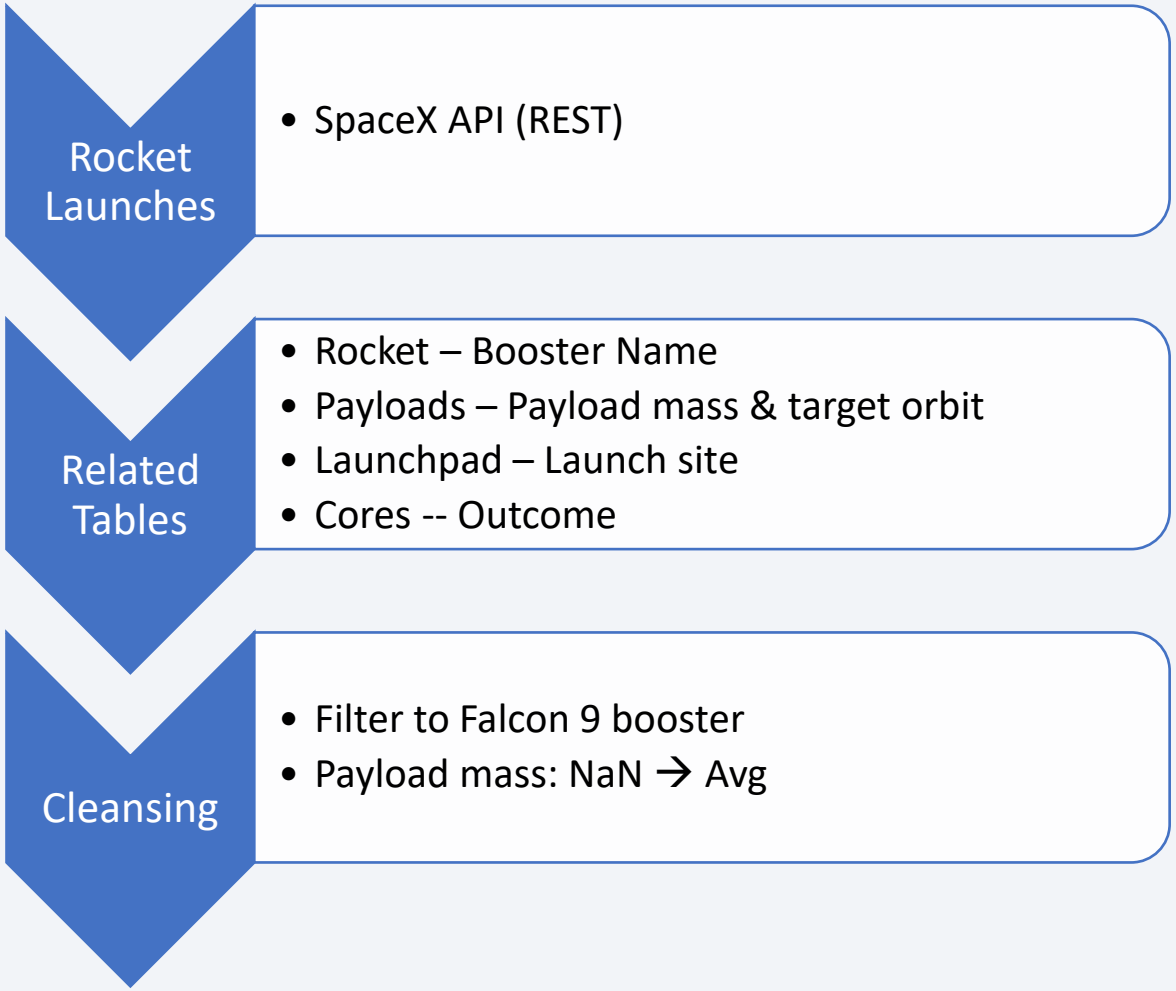
Methodology

Methodology

- Data collection & wrangling
  - Public information from SpaceX (API & WikiPedia)
  - Basic Analysis, Aliases for launches, Map to related tables
- Exploratory data analysis (EDA) with visualization and SQL
- Interactive visual analytics using Folium and Plotly Dash
- Predictive analysis using classification models
  - Logistic Regression, Support Vector Machines, Decision Tree, KNN
  - Grid-search algorithm for hyperparameter tuning

Data Collection – SpaceX API

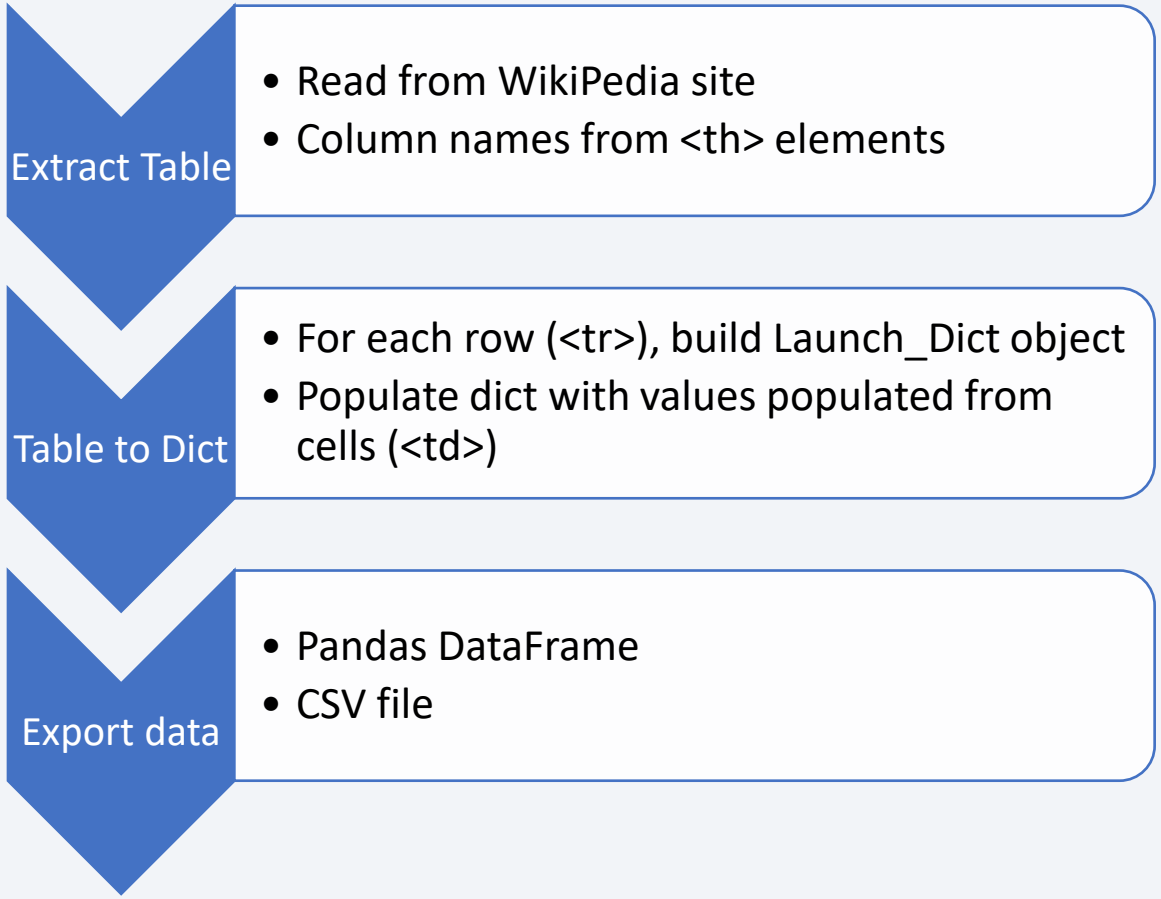
- Launch Data
  - SpaceX REST API
  - Requests library
  - Most columns contain IDs
- Collect related data
  - Use additional API calls
  - Lookup related information
- Cleansing
  - Falcon 9 launch: $62M (vs. $165M)
  - Reuse Stage 1 gives cost savings.
  - Filter Data to Falcon9
- Notebook: https://github.com/TimHolzmann/IBMClass/blob/main/jupyter-labs-spacex-data-collection-api.ipynb



Data Collection - Scraping

- Scraped Wikipedia data for Falcon 9 launches

- Notebook: https://github.com/TimHolzmann/IBMClass/blob/main/jupyter-labs-webscraping.ipynb



Data Wrangling

- Basic Analysis
  - Percent missing values
  - Data types
  - # launches by site
  - # launches by target orbit
- Categorized unsuccessful recovery of Falcon9 Stage 1 booster:
  - Not landed
  - Didn’t land on drone ship
  - Didn’t land in ocean region
  - Didn’t land on ground pad
- Notebook: https://github.com/TimHolzmann/IBMClass/blob/main/jupyter-labs-Data%20wrangling.ipynb

EDA with Data Visualization

- Charts built:
  - Flight Number vs Launch Site (colored by success) – Identify trends of launch sites & whether mission was successful
  - Payload Mass vs. Launch site (colored by success) – Identify which sites work best for different payloads and which payloads are successful.
  - Orbit vs. Success rate – Identify which target orbits have highest success rates
  - Flight Number vs. Orbit (colored by success) – Identify which target orbits are best initial candidates and which are most successful
  - Payload Mass vs. Orbit (colored by success) – Inspect whether there are trends in which mass/orbit combinations are successful or unsuccessful.
  - Year vs. Success rate – Observe trends in success rates
- Notebook: https://github.com/TimHolzmann/IBMClass/blob/main/jupyter-labs-EDA_Viz.ipynb

EDA with SQL

- Queries
  - Launches for various sites (e.g., “CCAFS LC-40”)
  - Payload mass by customer, booster
  - First successful landing
  - Successful boosters within payload mass range or max payload
  - Total mission outcomes (success/failure)
  - Landing outcomes for specified year
- Notebook: https://github.com/TimHolzmann/IBMClass/blob/main/jupyter-labs-eda-sql.ipynb

Build an Interactive Map with Folium

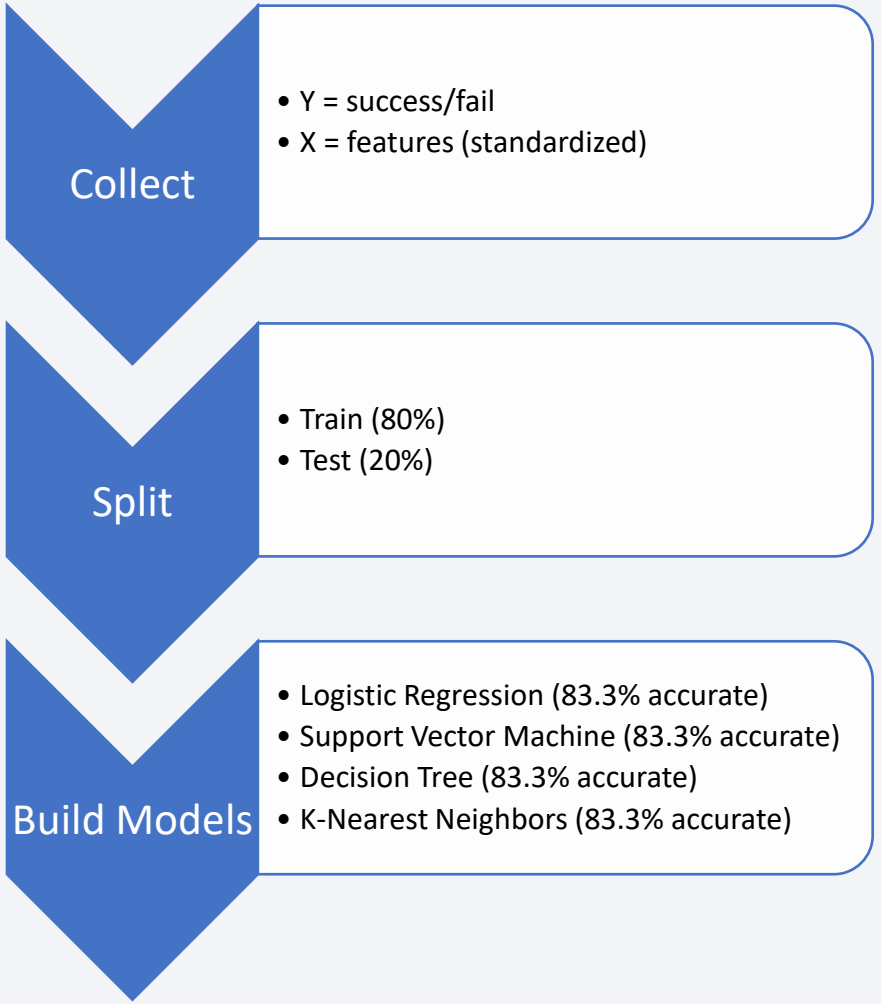
- Map objects
  - Circles on each launch site
  - Marker clusters for launches
  - Plotted nearby coastal points and calculated distance
- Mapping distance to nearby environment features
- Notebook: https://github.com/TimHolzmann/IBMClass/blob/main/jupyter-labs-Folium_Mapping.ipynb

Build a Dashboard with Plotly Dash

- Pie Charts controlled by dropdown
  - “All sites” → Shows % of total successful missions launched from each site
  - Single site → Shows % of launches from chosen site that are successful/failed
  - ID which sites give successful launches
- Scatter Plot shows success/failure (y axis) by payload (x axis)
  - Dropdown controls which launch sites to include
  - Range bar controls payloads to display
  - ID which payloads tend to succeed.
- Python file: https://github.com/TimHolzmann/IBMClass/blob/main/spacex_dash_app.py

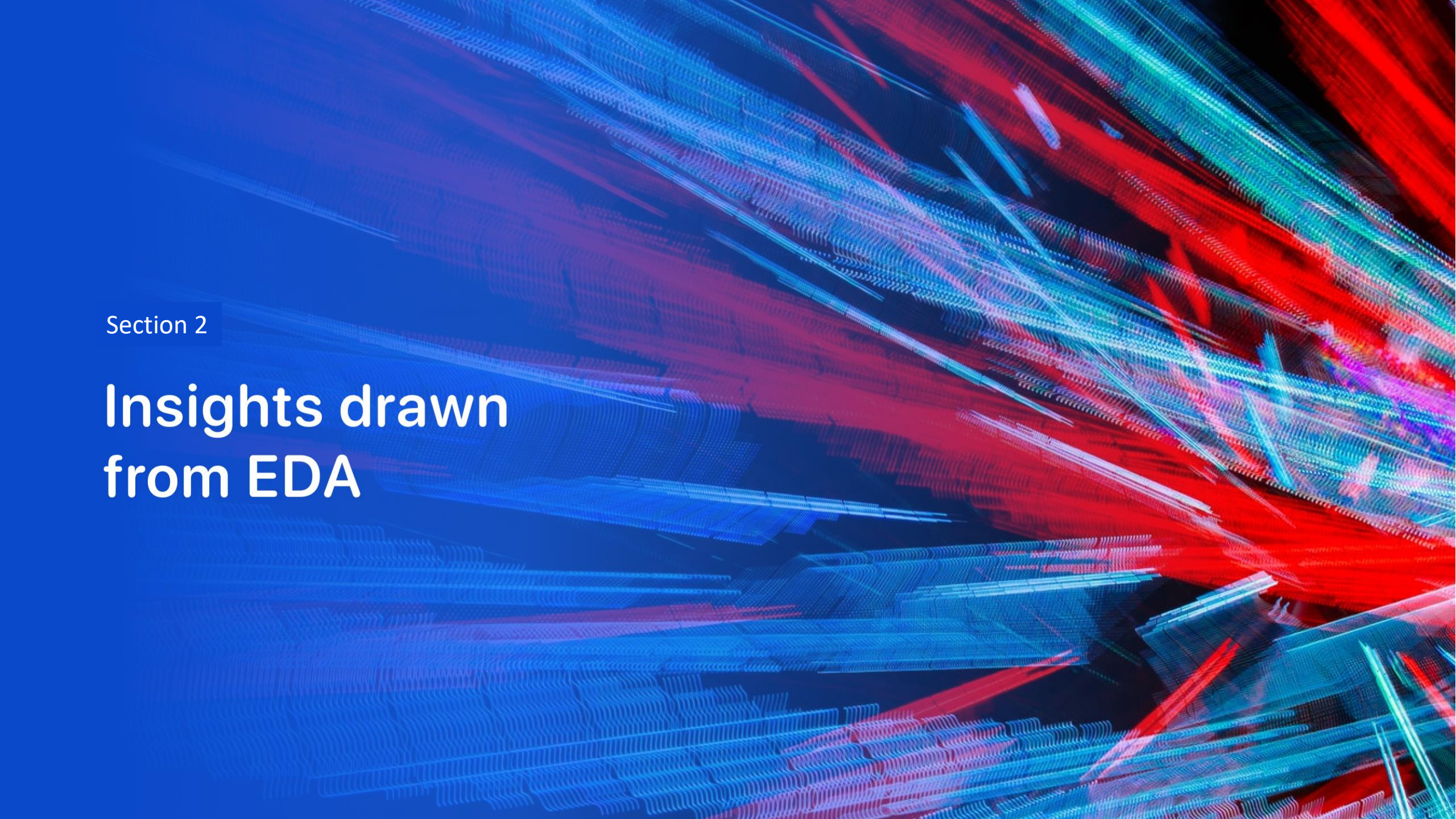
Predictive Analysis (Classification)

- Trained four different predictive models
  - Logistic Regression
  - Support Vector Machine
  - Decision Tree
  - K-Nearest Neighbor
- Used Grid-Search Algorithm to tune hyperparamters
- All searches achieved same 83.3% accuracy on test set
- Notebook: https://github.com/TimHolzmann/IBMClass/blob/main/jupyter-labs-ML.ipynb



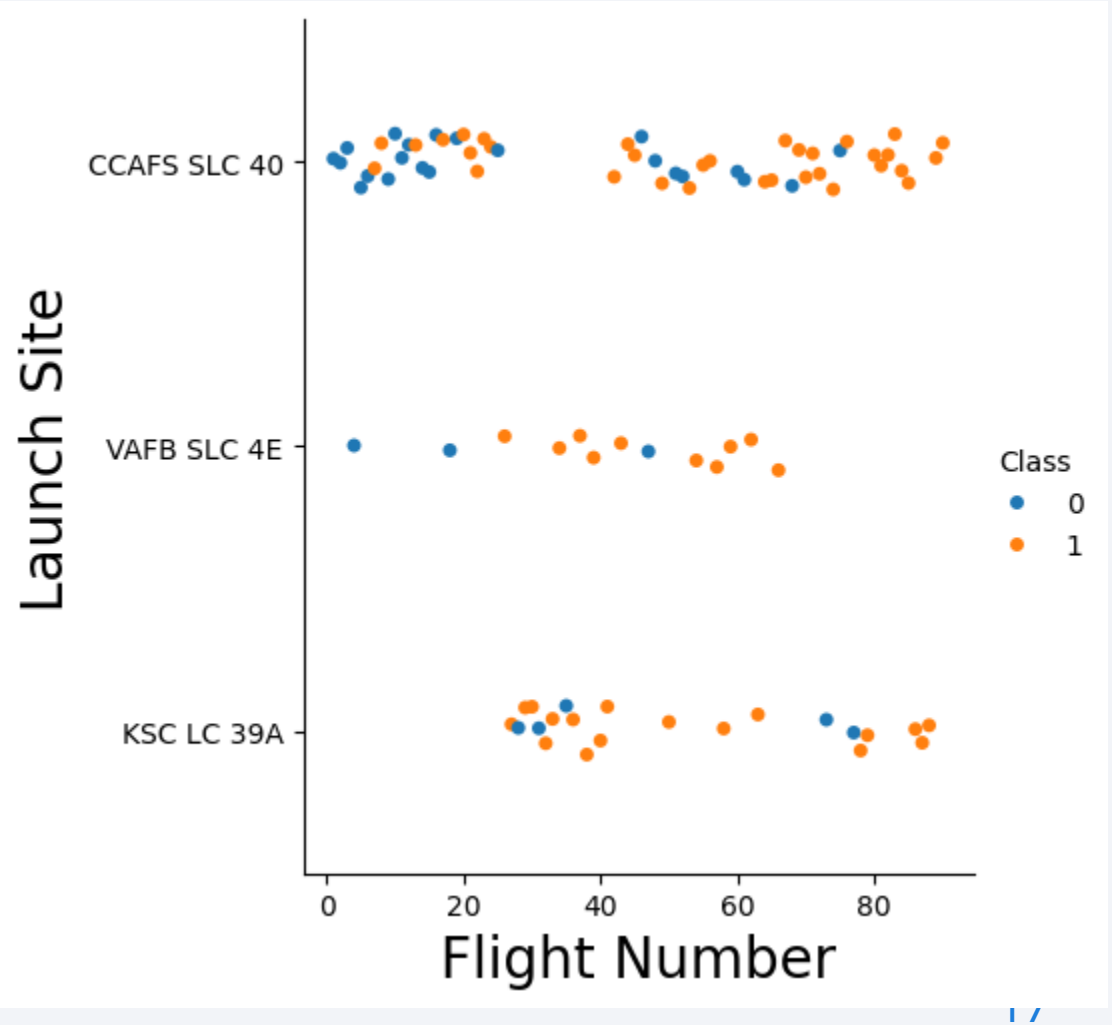
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results


Section 2
Insights drawn from EDA

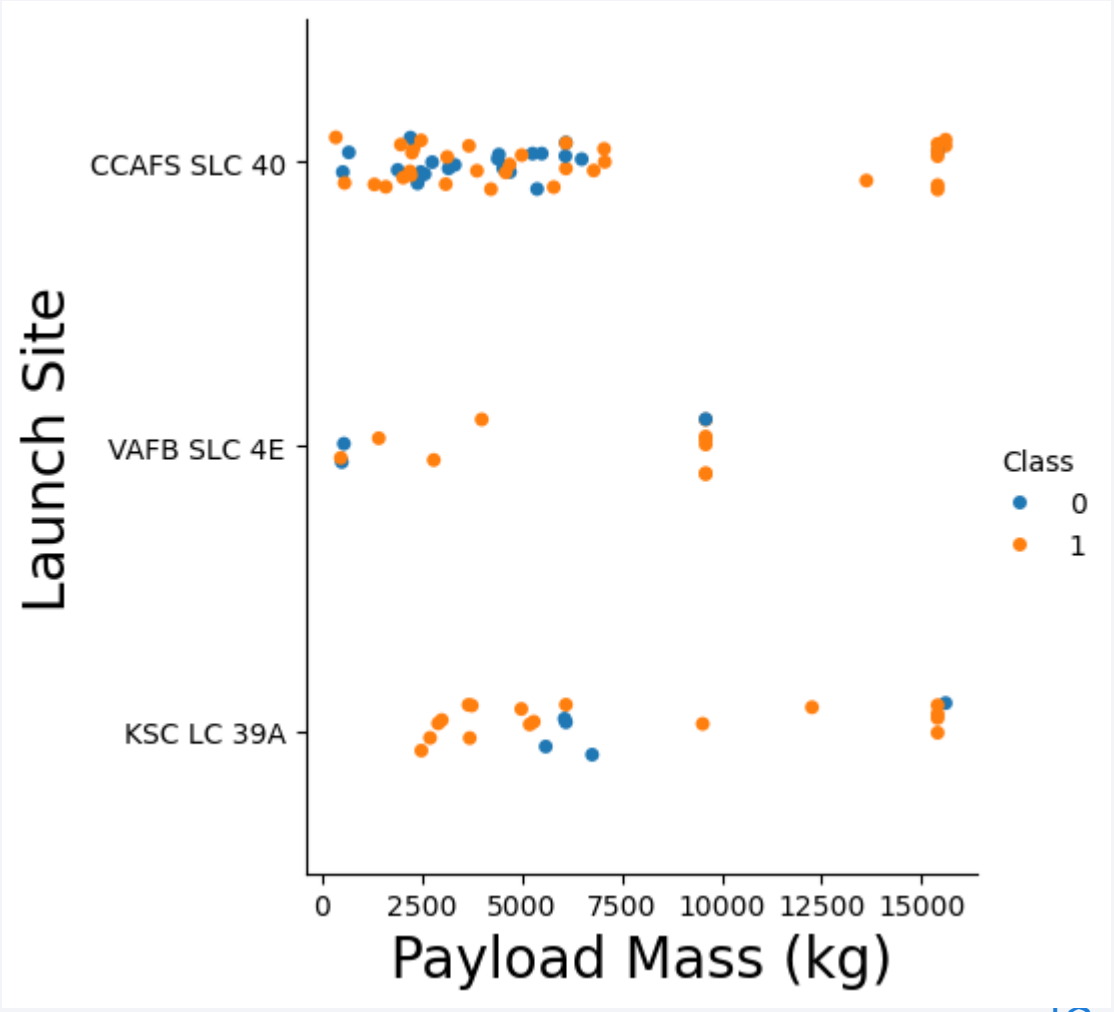
Flight Number vs. Launch Site

- We started later at KSC.
- Early launches at each site were not successful landings, but we corrected with later launches.



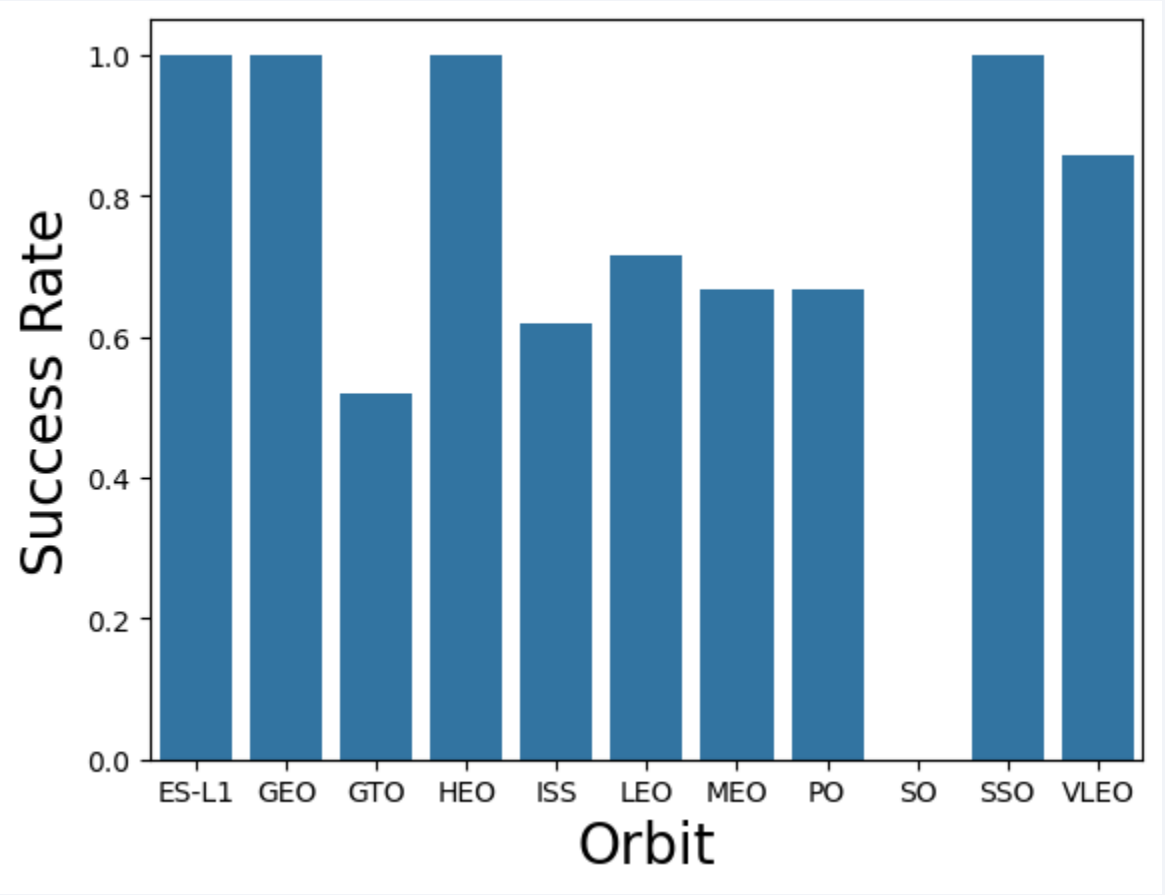
Payload vs. Launch Site

- VAFB doesn't do heavy weight launches



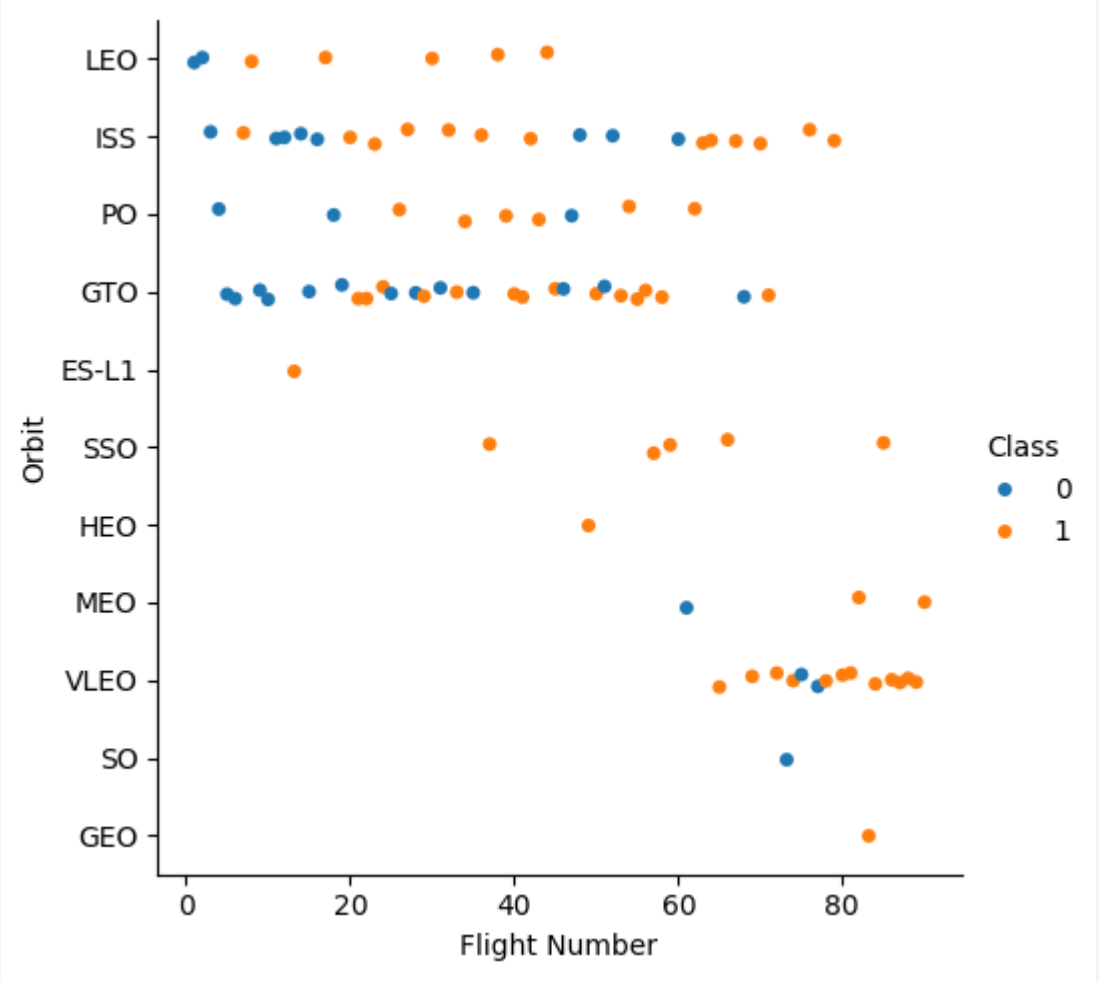
Success Rate vs. Orbit Type

- 100% success rate for
  - ES-L1
  - GEO
  - HEO
  - SSO



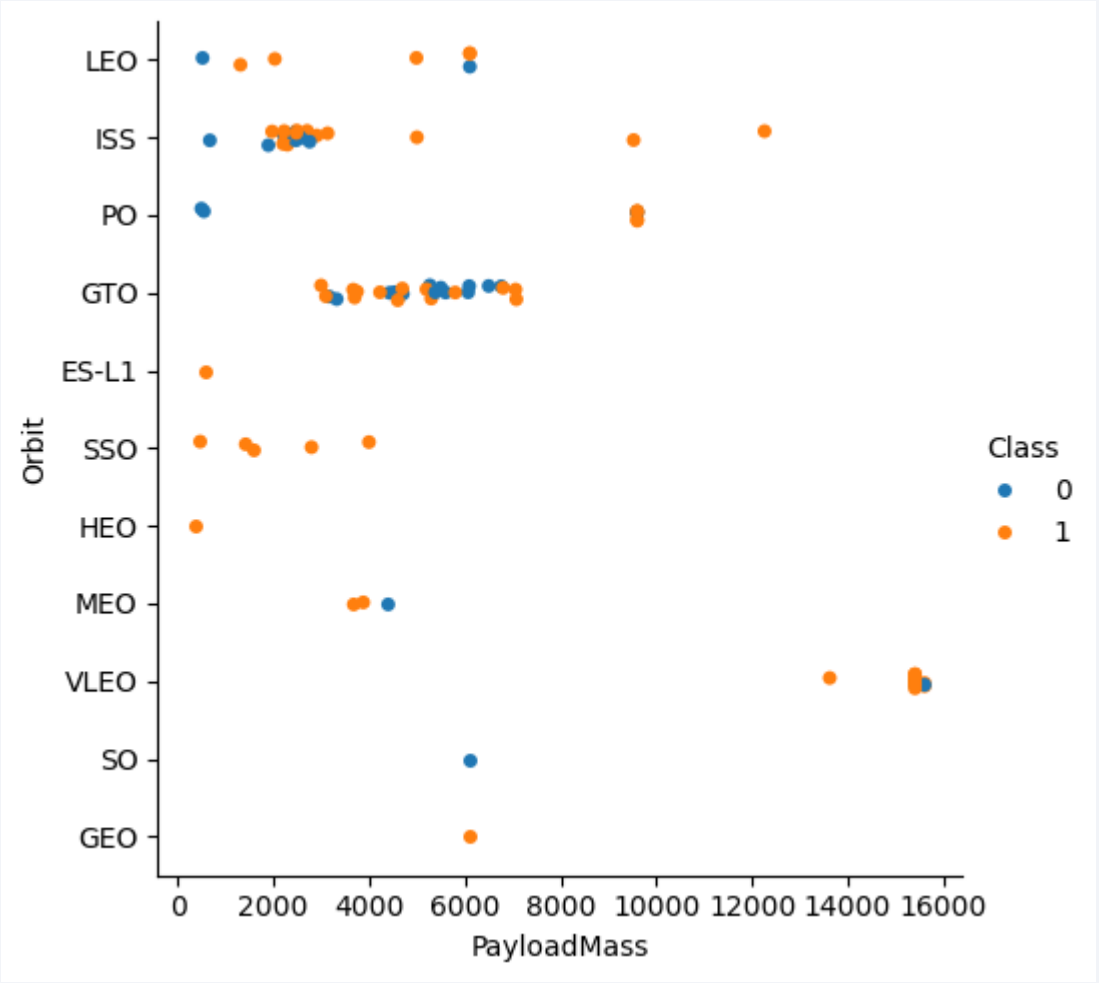
Flight Number vs. Orbit Type

- Early launches had lower success rates



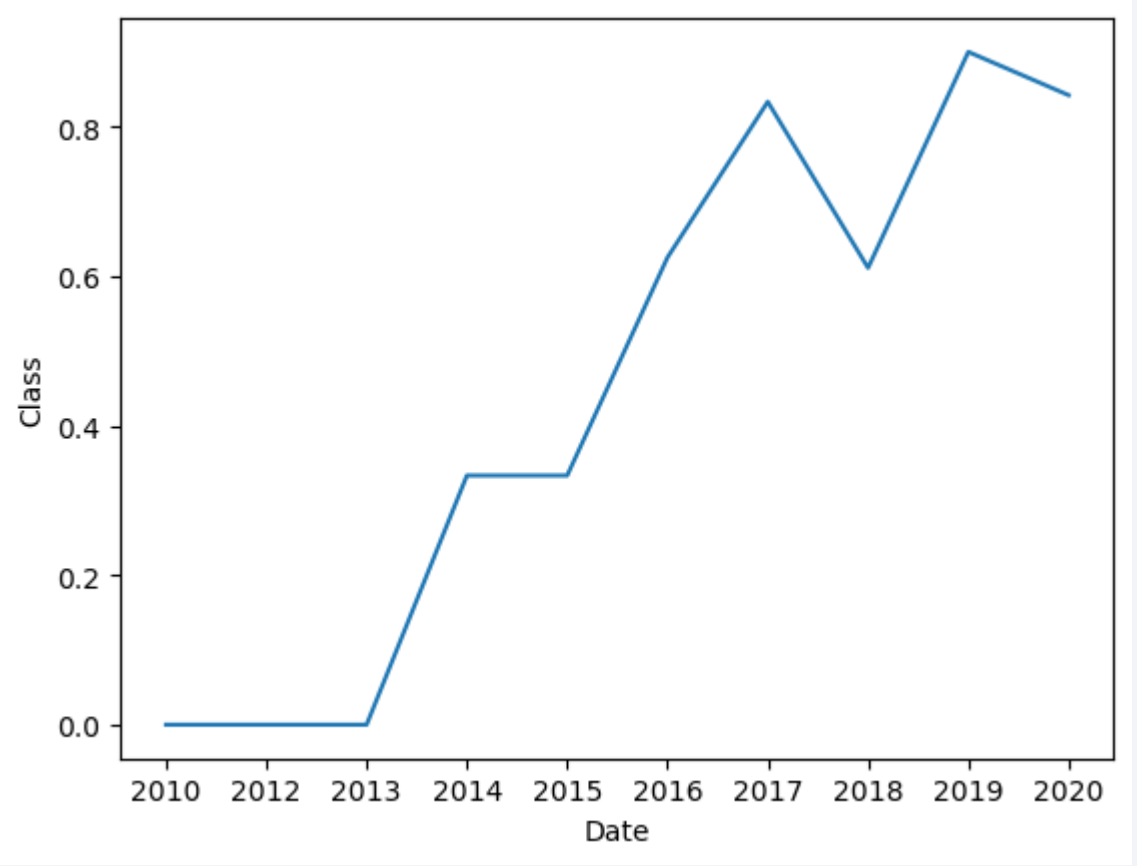
Payload vs. Orbit Type

- Elliptical orbits have high success rates but lowest payloads
- Highest payloads (>14,000 kg) can only go into Very Low Earth Orbit



Launch Success Yearly Trend

- Increasing trend in average success rate



All Launch Site Names

- 4 Launch Sites
  - Cape Canaveral (2x in FL)
  - Kenedy Space Center (in FL)
  - Vandenberg (in CA)

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

Launch Site Names Begin with 'CCA'

- NASA is the primary customer at CCAFS LC-40 launch site.

ate	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

- 45596kg
- Thank you, NASA (CRS). We love you too!

Average Payload Mass by F9 v1.1

- 2,534.666kg
- That’s a lot of kg!

First Successful Ground Landing Date

- 2015-12-22
- We were all very happy on this day!

Successful Drone Ship Landing with Payload between 4000 and 6000

- Many booster options for payloads 4000-6000kg

Booster_Version	
F9 v1.1	F9 FT B1031.2
F9 v1.1 B1011	F9 FT B1032.2
F9 v1.1 B1014	F9 B4 B1040.2
F9 v1.1 B1016	F9 B5 B1046.2
F9 FT B1020	F9 B5 B1047.2
F9 FT B1022	F9 B5 B1048.3
F9 FT B1026	F9 B5 B1051.2
F9 FT B1030	F9 B5B1060.1
F9 FT B1021.2	F9 B5 B1058.2
F9 FT B1032.1	F9 B5B1062.1
F9 B4 B1040.1	

Total Number of Successful and Failure Mission Outcomes

- Over 99% mission success rate!

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

Boosters Carried Maximum Payload

- Only certain boosters are able to carry the maximum payload

Booster_Version	
F9 B5 B1048.4	F9 B5 B1048.5
F9 B5 B1049.4	F9 B5 B1049.5
F9 B5 B1049.7	F9 B5 B1051.3
F9 B5 B1051.4	F9 B5 B1051.6
F9 B5 B1056.4	F9 B5 B1058.3
F9 B5 B1060.2	F9 B5 B1060.3

2015 Launch Records

Month	Landing_Outcome	Booster_Version	Launch_Site
January	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
February	Controlled (ocean)	F9 v1.1 B1013	CCAFS LC-40
March	No attempt	F9 v1.1 B1014	CCAFS LC-40
April	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40
04	No attempt	F9 v1.1 B1016	CCAFS LC-40
June	Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40
December	**Success (ground pad)**	F9 FT B1019	CCAFS LC-40

Only one successful landing in 2015 with the Falcon 9 booster.

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

Attempted landing on 2/3 of launches:

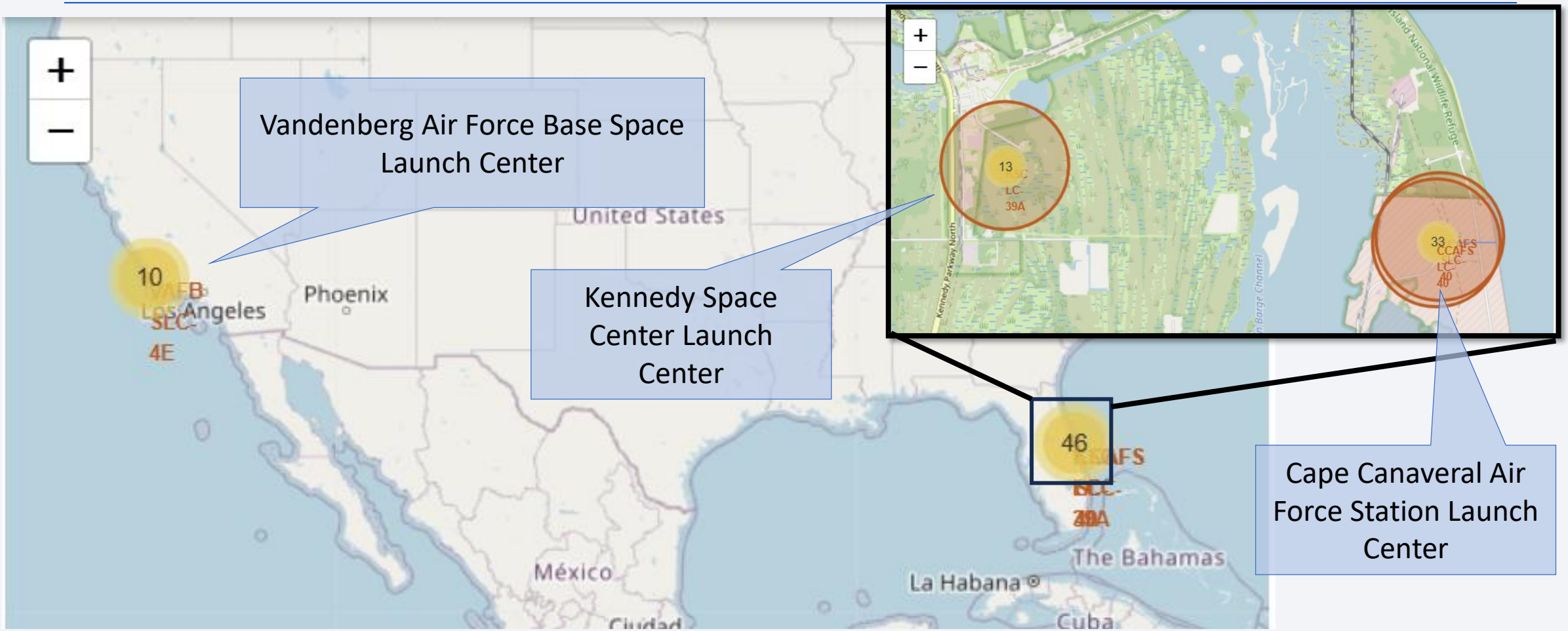
- Success
  - Drone Ship (5)
  - Group pad (1)
- Not successful
  - Drone ship (5)
  - Precluded (1)
  - Ocean-controlled (3)
  - Ocean-uncontrolled (2)
  - Parachute (1)

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

Section 3

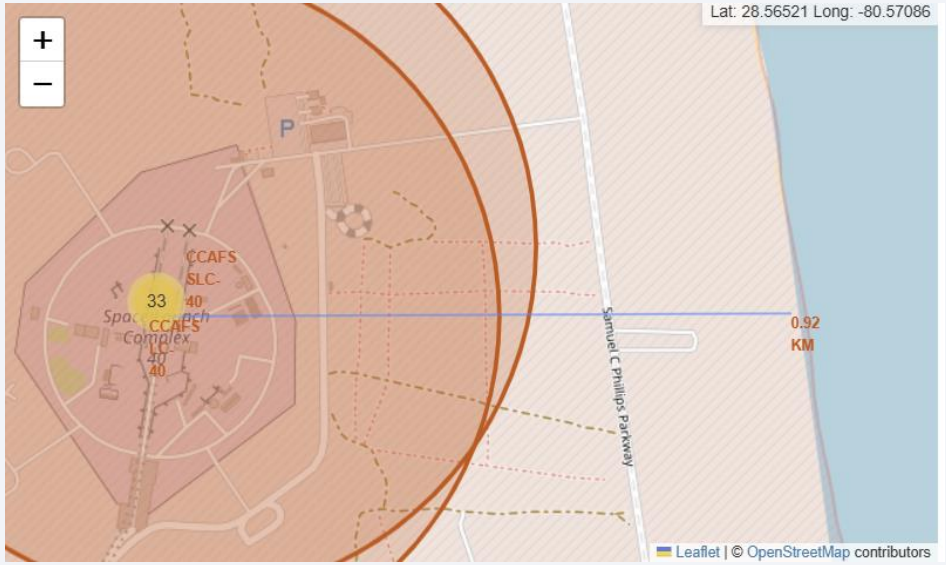
Launch Sites Proximities Analysis

All Launch Sites



Distance to Coastline

- CCAFS is just under 1km from coast

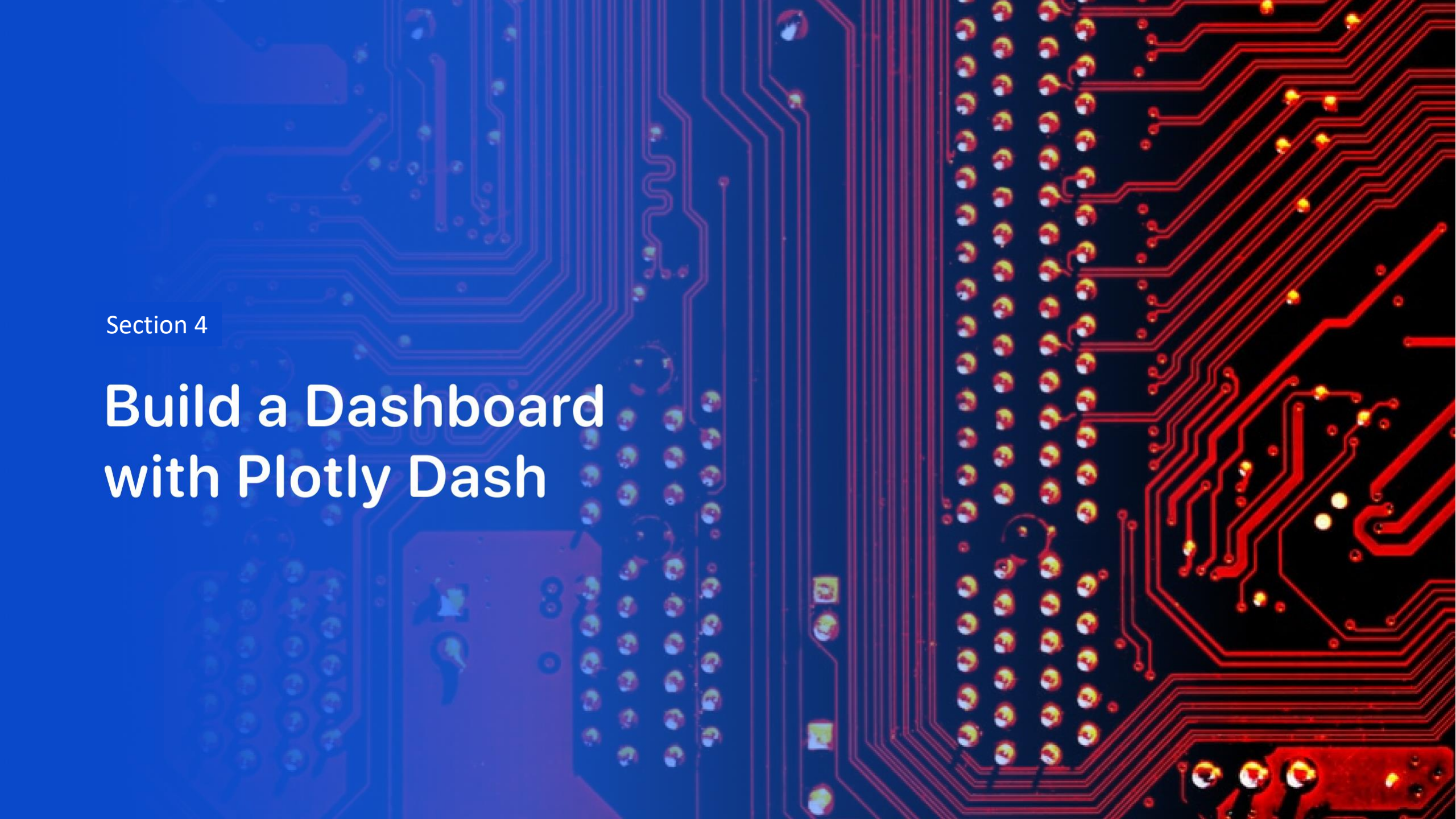


Distance from Population Centers

- 18km from Cape Canaveral, FL
- 23km from Titusville, FL
- 60km from Orlando, FL

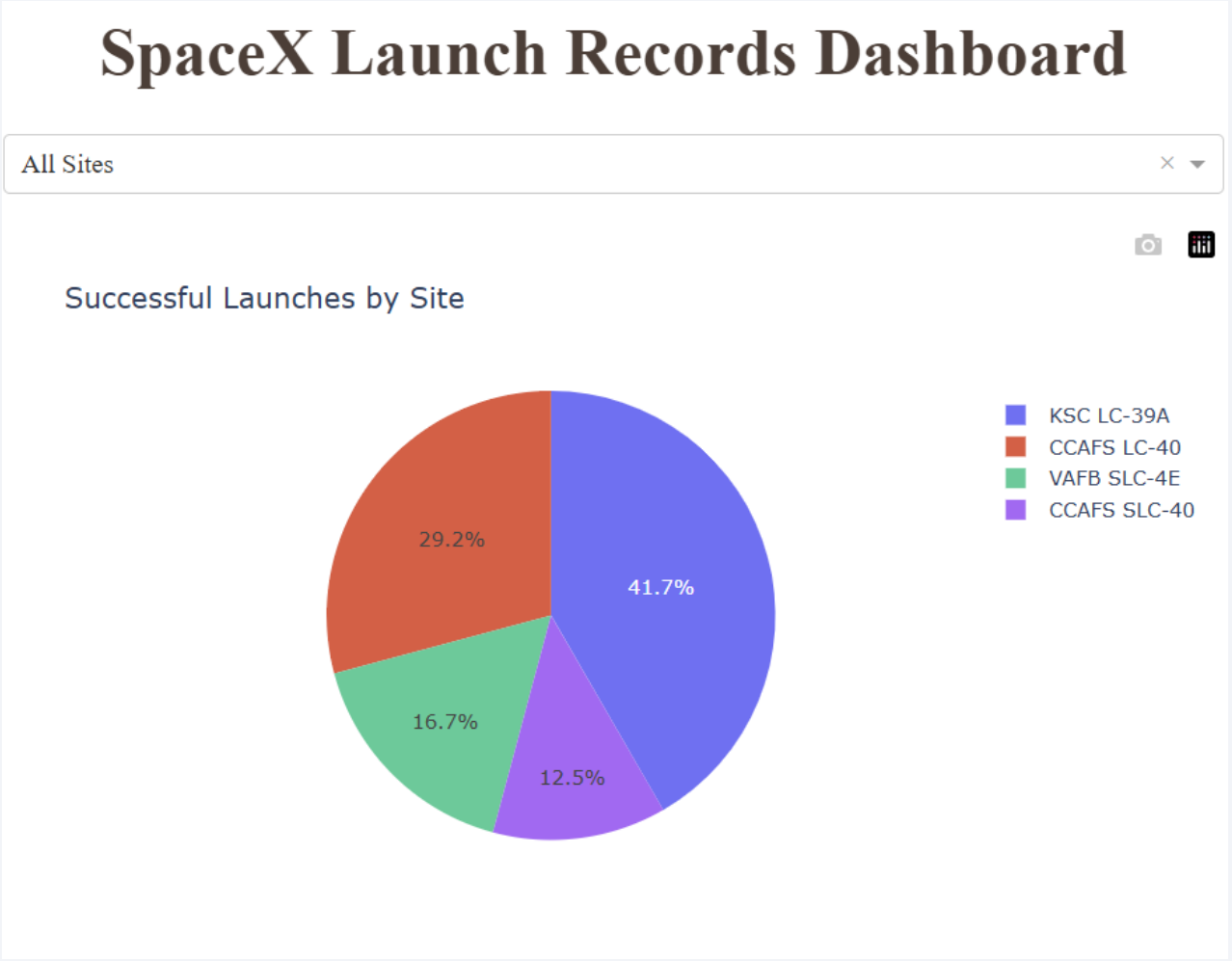
Distance from Railways

- 0.93km from Railway


Section 4
Build a Dashboard with Plotly Dash

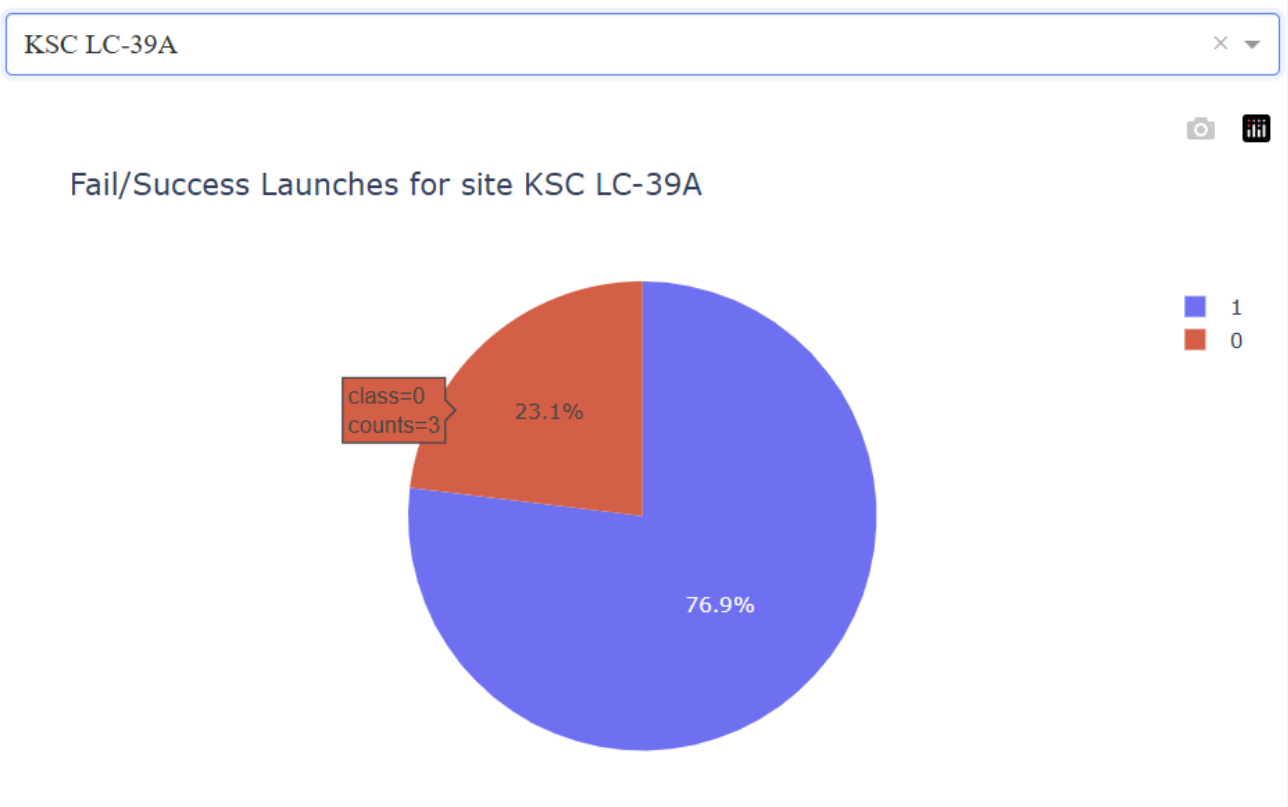
Successful Launches by Site

- 83% of Launches are from FL



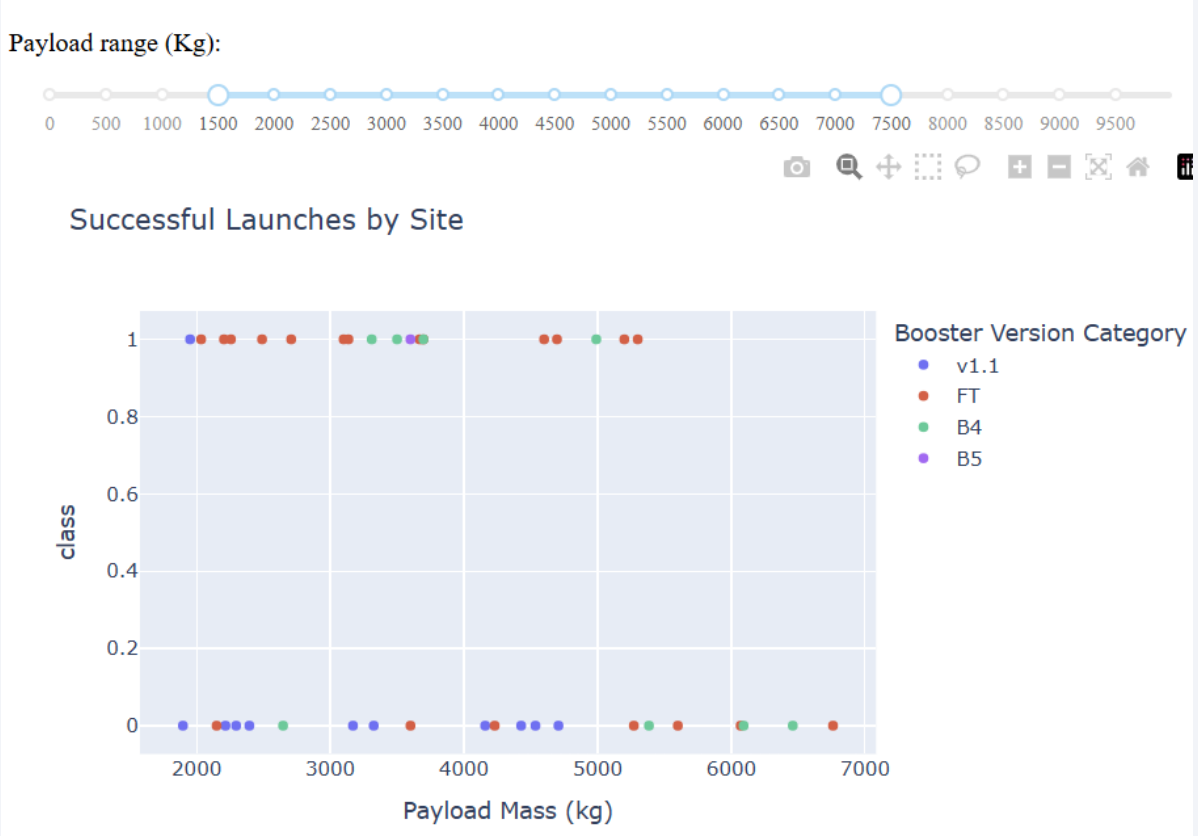
Highest Success Rate

- KSC LC-39A has highest success rate (10/13)



Launches by Booster & Success

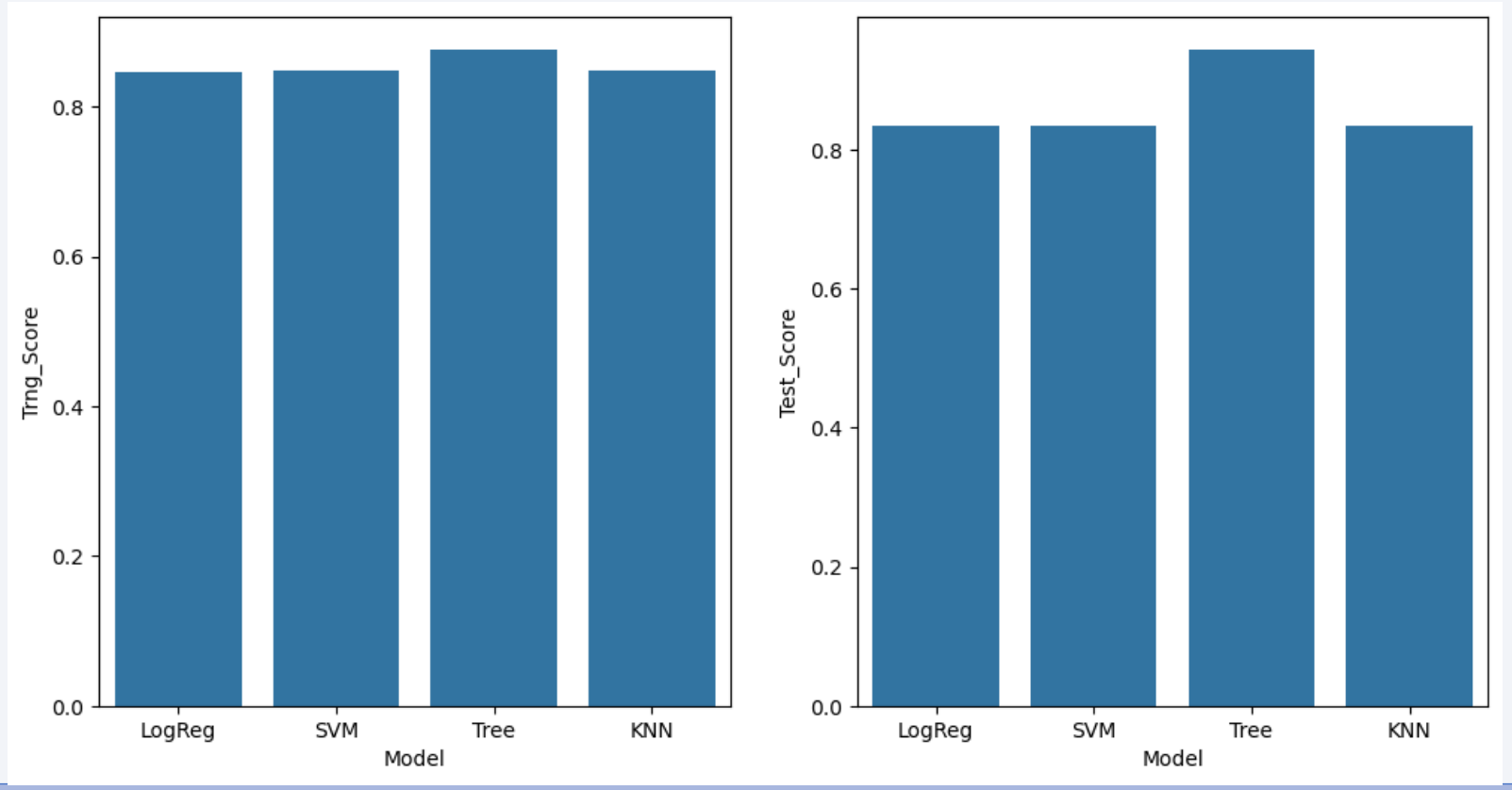
- FT has higher success with lower payload
- No boosters are successful with >5500kg payload



Section 5

Predictive Analysis (Classification)

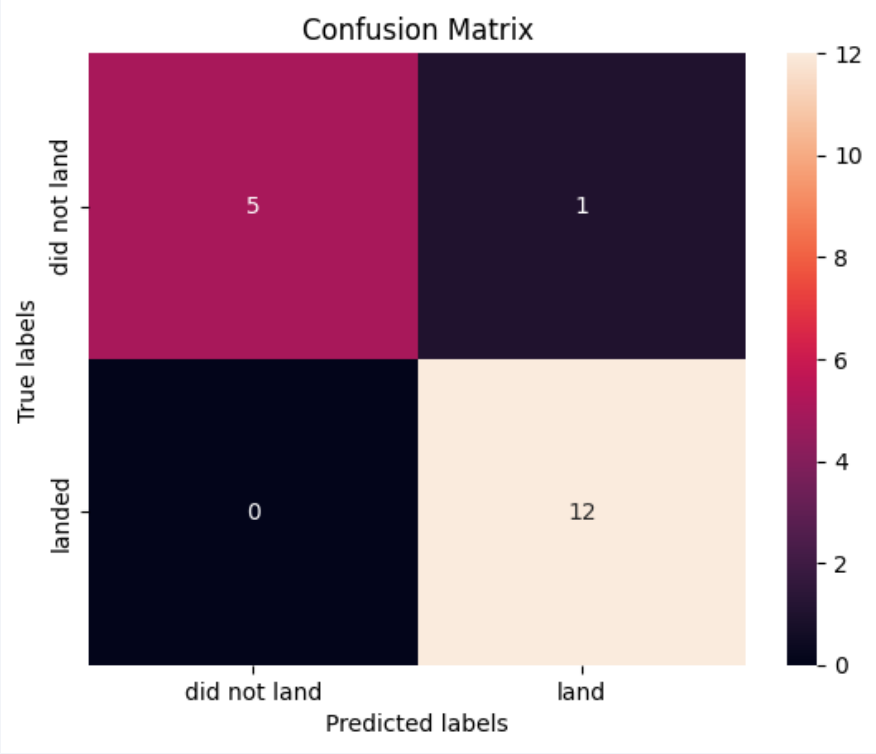
Classification Accuracy



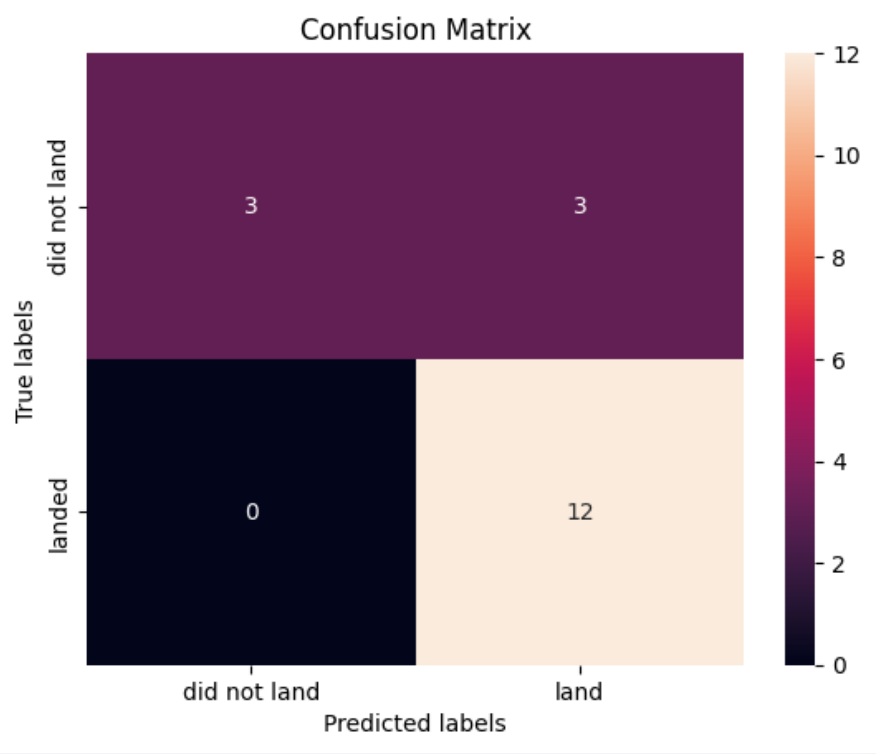
Decision Tree model has highest classification accuracy on both training and testing data sets.

Confusion Matrix

Decision Tree model



Other models



Conclusions

- 4 sites
- Decision Tree model is best predictor of launch success
- Large payload is harder to recover Stage 1 booster, but still have mission success

Thank you!