Winning Space Race with Data Science

Tawanda
13 December 2021

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion?
  - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY

- The main purpose of the project is to predict the success of a planned space launch.
- A successfully landing means first stage can be reused lowering the cost of a launch.
- Costs become deterministic.
- Results can be applied when bidding against SpaceX for a rocket launch.
- The analysis shows that we can predict the outcome of a launch with 83.33% accuracy.

INTRODUCTION

- The project uses predictive models to determine whether a space launch will be successful or not.
- Records of past launches by SpaceX will be used in this analysis.
- Some key questions to be addressed include:
  - What factors affect the success of a launch?
  - What booster versions result in higher rates of success?
  - How does the destination orbit affect the success of a launch?
  - Does the choice of launch site affect the rate of success of a launch?
  - Do proximities affect the selection of a launch site?

Section 1

Methodology

convert the different outcomes into Training Labels with 1 means the booster successfully landed 0 means it was unsuccessful.
convert the different outcomes into Training Labels with 1 means the booster successfully landed 0 means it was unsuccessful.

Methodology

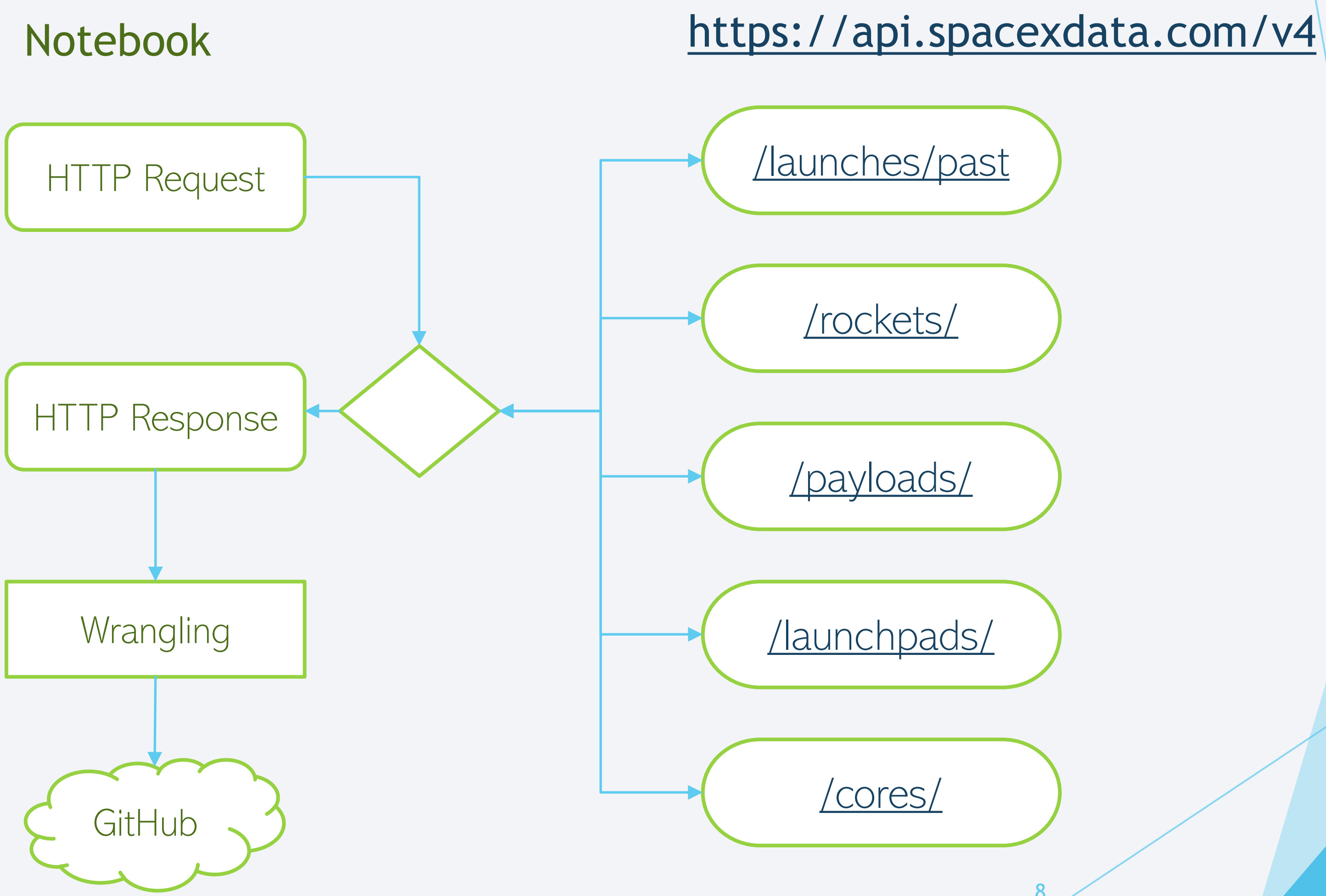
Executive Summary

- Data collection methodology
  - REST API calls.
  - Web scraping Wikipedia.
- Data wrangling
  - Convert launch outcomes to labels.
- Exploratory Data Analysis (EDA)
  - SQL querying.
  - Visualization.
- Interactive Visual Analytics
  - Folium
  - Dashboard (Plotly Dash)
- Predictive Analysis
  - Grid Search

Data Collection

- Data was collected using two methods.
- REST API calls to the following:
  - Historical launch data – JSON file: https://api.spacexdata.com/v4/launches/past
  - Rocket booster name: https://api.spacexdata.com/v4/rockets/
  - Payload mass and orbit: https://api.spacexdata.com/v4/payloads/
  - Launch site name and coordinates: https://api.spacexdata.com/v4/launchpads/
  - Cores – reuses, landing pad, outcome: https://api.spacexdata.com/v4/cores/
- Web scrapping SpaceX launch data on Wikipedia:
  - Falcon 9 & Falcon Heavy launch records: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

Data Collection – SpaceX API

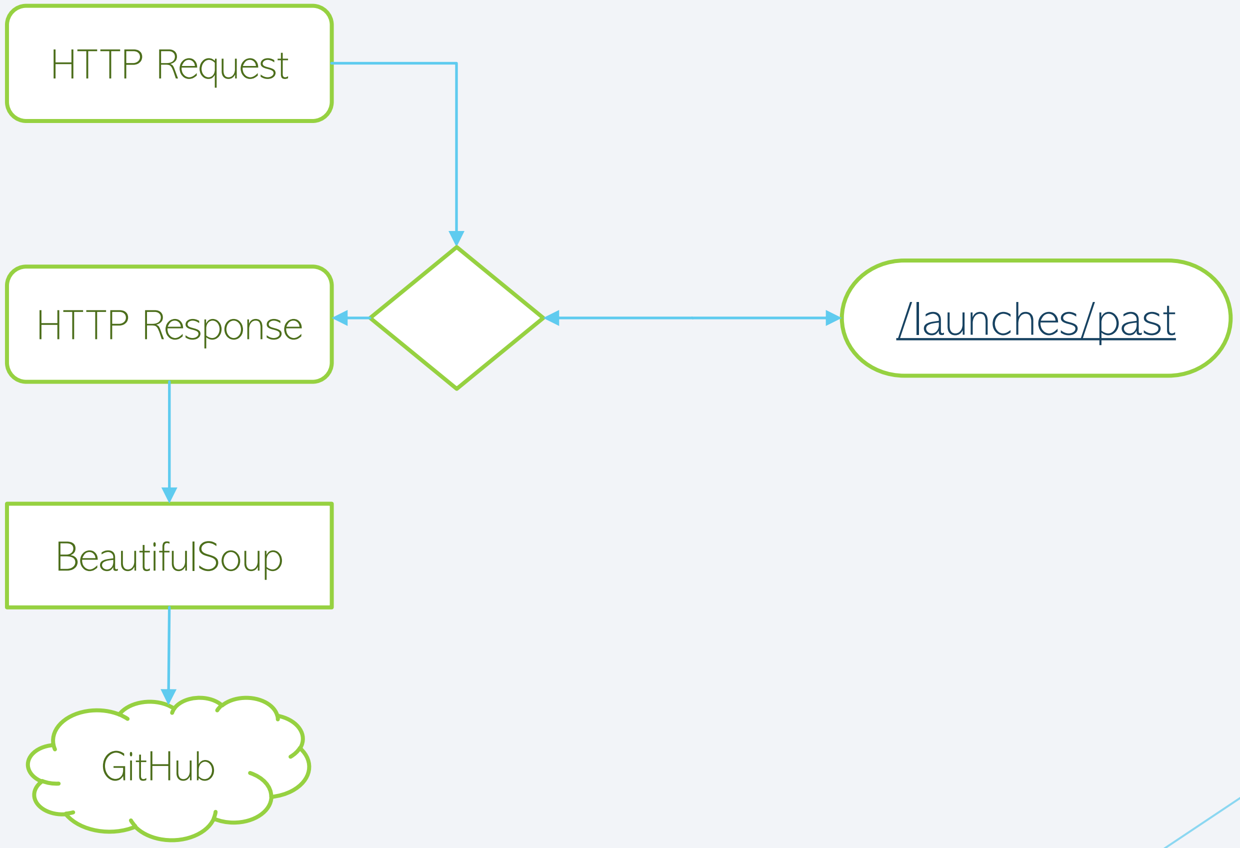


github.com/SpaceX_Landing_Prediction/API_Data_Collection

Data Collection - Scraping

Notebook

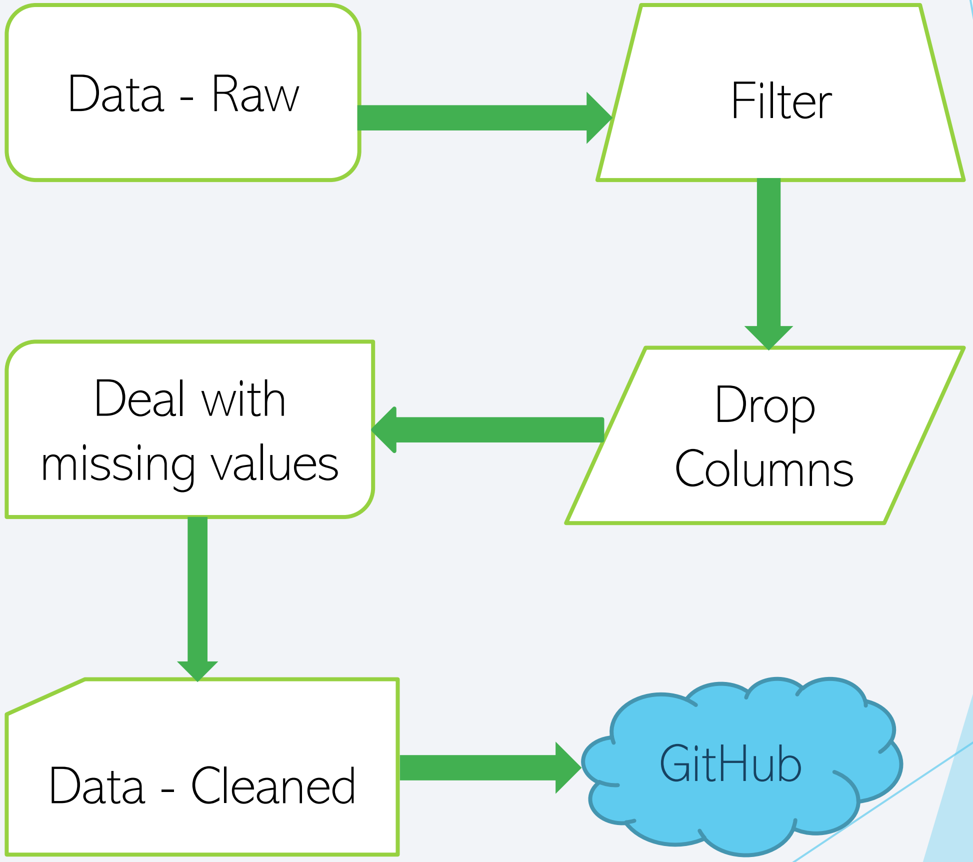
https://en.wikipedia.org/List_of_Falcon_9_Launches



//github.com/SpaceX_Landing_Prediction/Web_Scraping_Data

Data Wrangling

- Process:
  - Filtering – Falcon 9 data only.
  - Dropping columns.
  - Dealing with missing values.
  - Creating outcome labels



https://github.com/Wrangling

EDA with Data Visualization

- Scatter PlotCatplots
  - Flight Number vs Payload Mass – success increases with mass & no. of flights.
  - Flight Number vs Launch Site – success rate increased after 60th launch.
  - Payload vs Launch Site – success increased with mass only at VAFB-SLC.
- Bar Plot
  - Success Rate vs Orbit – higher success rate at higher orbitals.
- Notebook url: https://github.com/EDA_with_Data_Visualization

EDA with SQL

- SQL queries executed on the data showed the following results:
  - First successful landing outcome – December 22, 2022.
  - Average payload for booster version F9 v1.1* – 2 534 kg.
  - Total payload carried for NASA – 45 596kg
  - Medium payload booster versions – F9 FT B1022, B1026, B1021.2 & B1031.2
  - Total success missions – 100, Total failed missions – 1.
- Notebook url: //github.com/SpaceX_Landing_Prediction/EDA_with_SQL

Interactive Maps with Folium

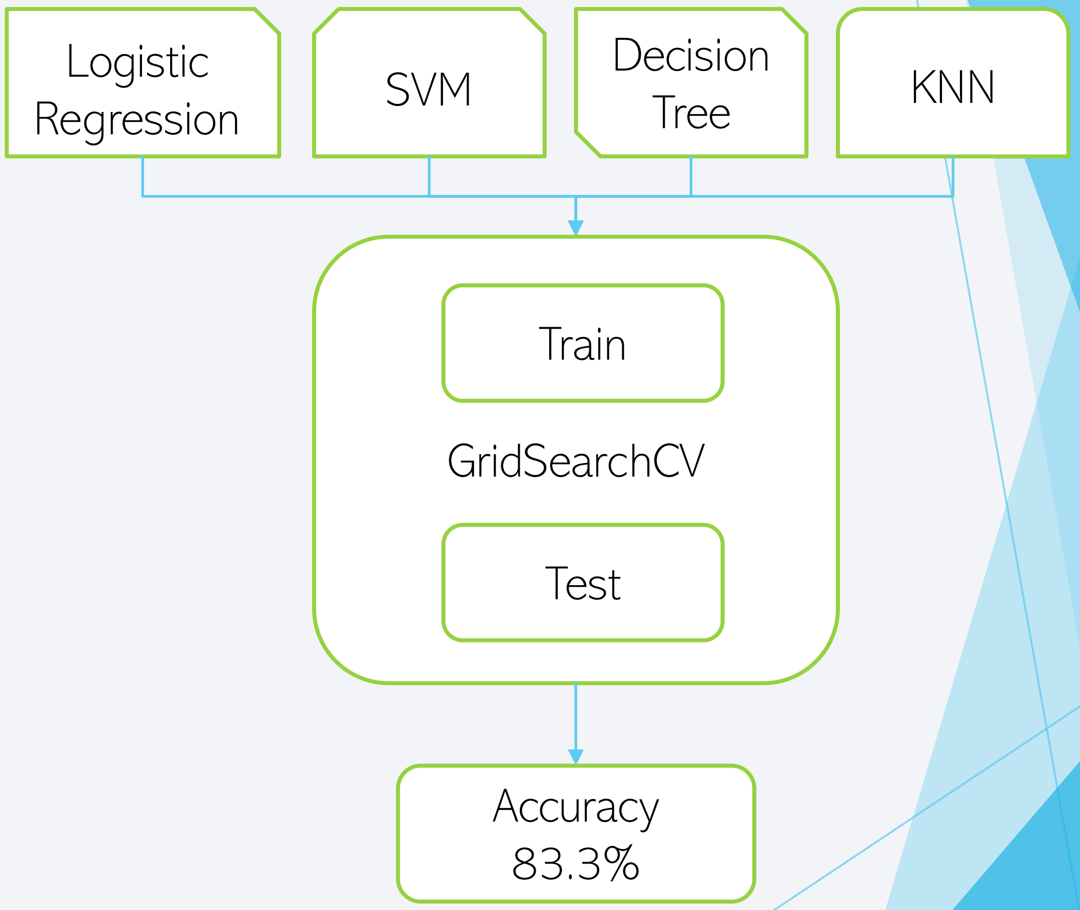
- Launch site locations analysis to identify geographical patterns.
- The following map objects were added:
  - Circles – to show all four SpaceX launch sites and NASA Johnson Space Centre.
  - Markers (clustered) – green and red indicating success or failure at each site.
  - Lines – to highlight straight line distances between two locations on the map.
  - Distance markers – to show the distance (in km) between two locations on the map.
- Proximity analysis shows that:
  - Cities are beyond a 30 km radius from the sites.
  - All launch sites are in situated within a 10km radius from the coast.
  - Rail lines are within a 5km distance from the launch sites.
- The notebook url: https://github.com/6_Launch_Site_Location_Analysis

Dashboard with Plotly Dash

- The dashboard shows:
  - Pie Chart – the launch successes by site.
  - Scatter Plot – Correlation between payload and outcome.
- Features of the dashboard include:
  - Launch Site drop-down input – to allow filtering by site.
  - Range Slider – allows easy selection of payload ranges.
- KSC LC 39A has the highest success rate – 41.7%
- The FT booster version has the highest success rate – at medium payload mass.
- Dashboard script url: https://github.com/ndarin/success_rate_dashboard.py

Predictive Analysis (Classification)

- Analysis focused on selecting the best performing classification model.
- Models trained and tested include:
  - Logistic Regression (LR)
  - Support Vector Machine (SVM)
  - Decision Tree (DT)
  - K-Nearest Neighbors (KNN)
- All models showed similar performance on test data with accuracy of 83.33%.
- You need present your model development process using key phrases and flowchart
- Notebook url: https://github.com/



Results

- Exploratory data analysis results
  - Success increases with mass & no. of flights.
  - Success rate increased after 60th launch.
  - Success rate is higher for higher orbitals.

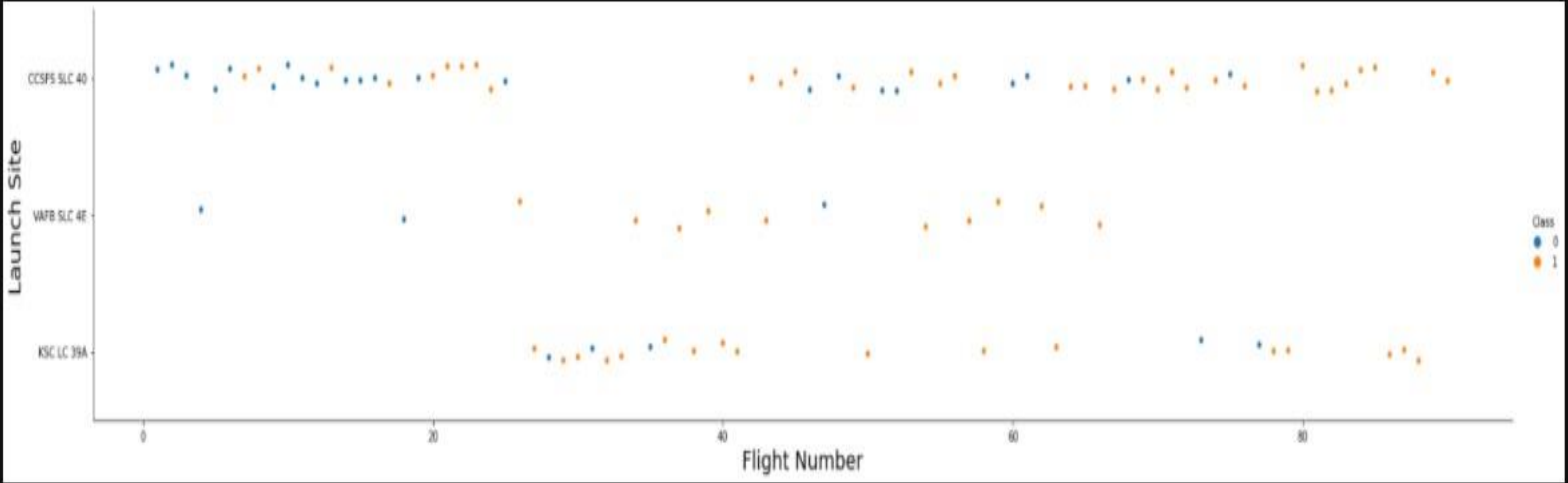
- Location Site analysis results
  - Success rate varies by site
  - Launch sites are in proximity of rail lines and coast

- Predictive Analysis results
  - All models show similar performance
  - Accuracy is 83.3%

Section 2

Insights drawn from EDA

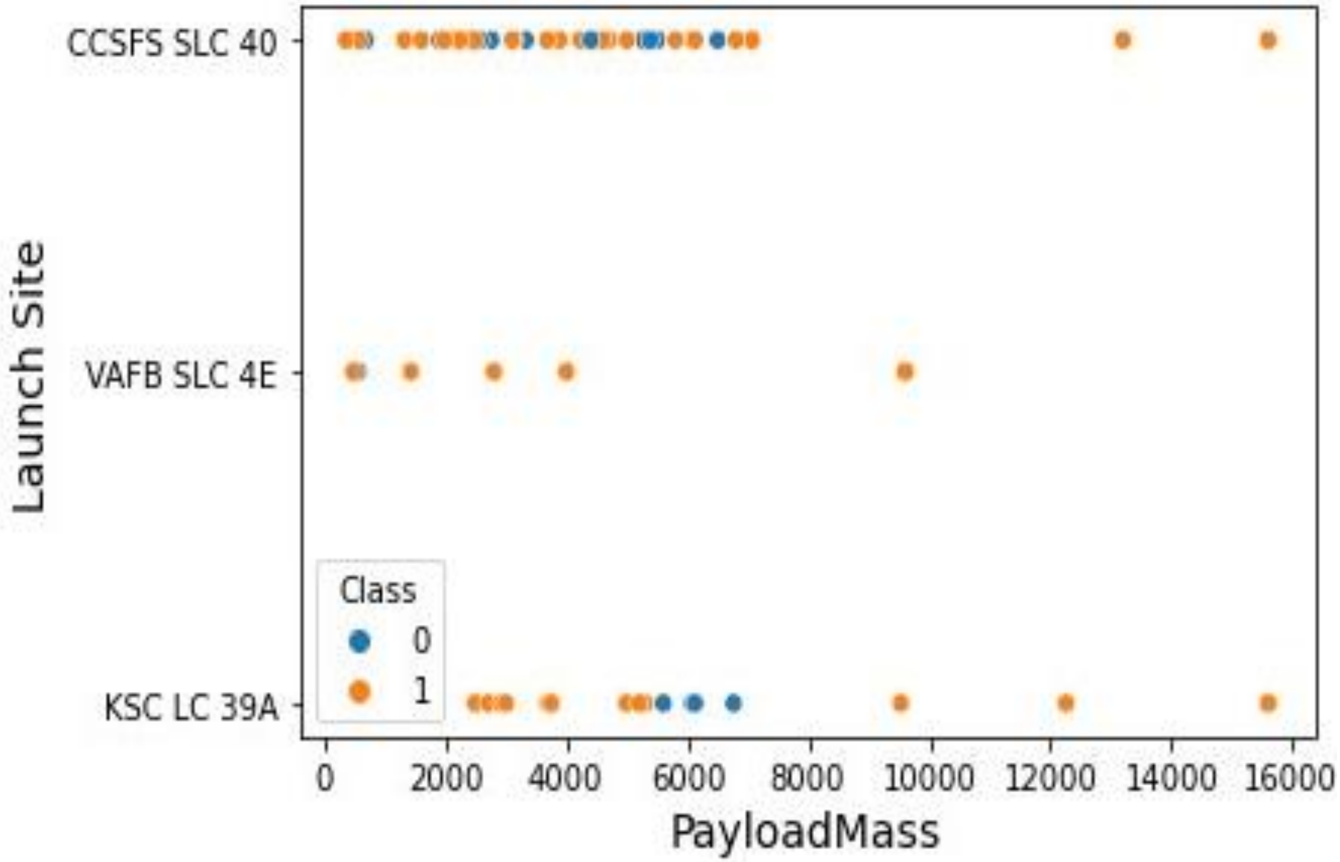
Flight Number vs. Launch Site



Most of the launches are concentrated at `CCAFS LC-40`. The success rate at this location increased with the number of launches and was relatively high after the 60th launch. For a brief period launches seem to have been moved to `KSC LC-39A`. Launches at `VAFB SLC 4E` are sparse and have a similarly high sucess rate as those at `KSC LC-39A`.

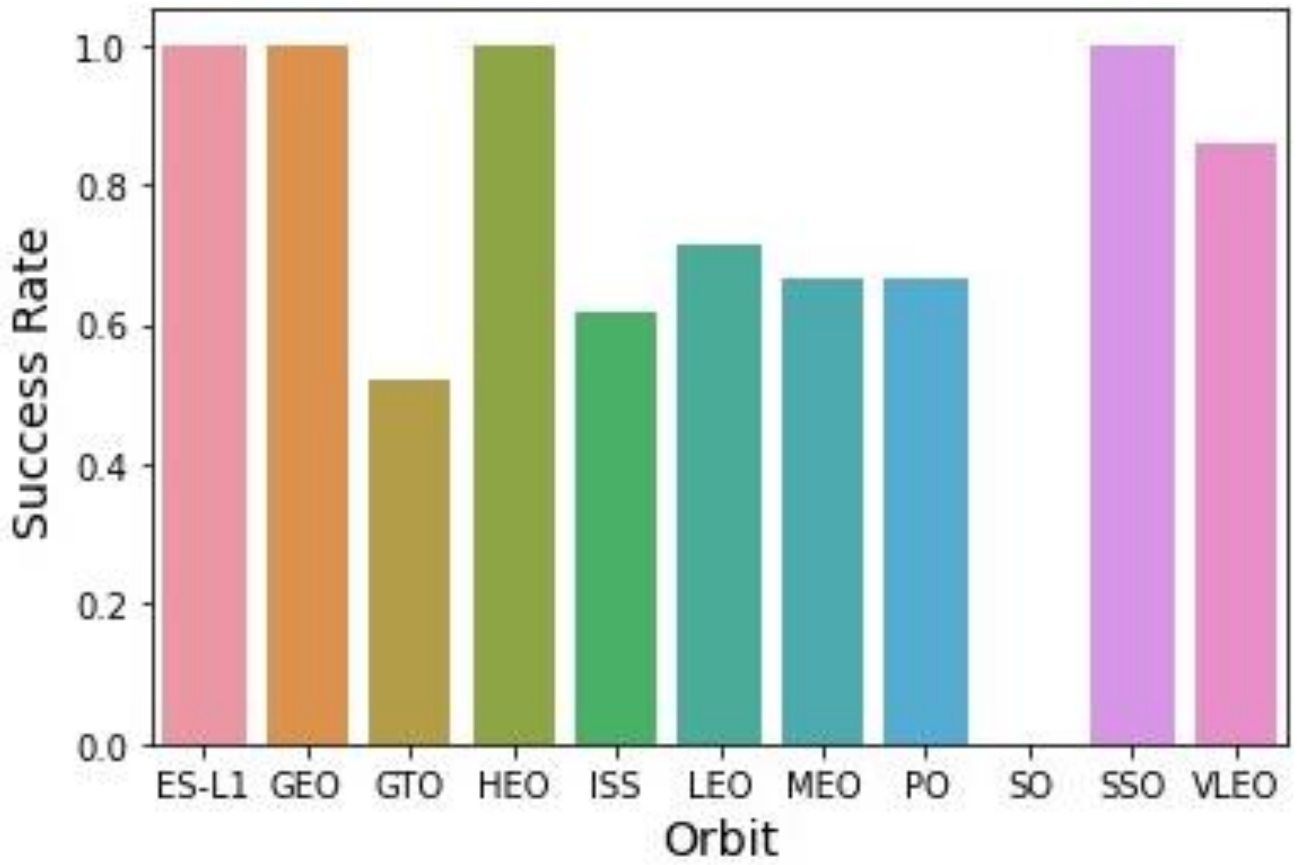
Payload vs. Launch Site

- The scatter plot shows that for launch site VAFB-SLC there are no rockets launched for heavy payload mass greater than 10 000kg.



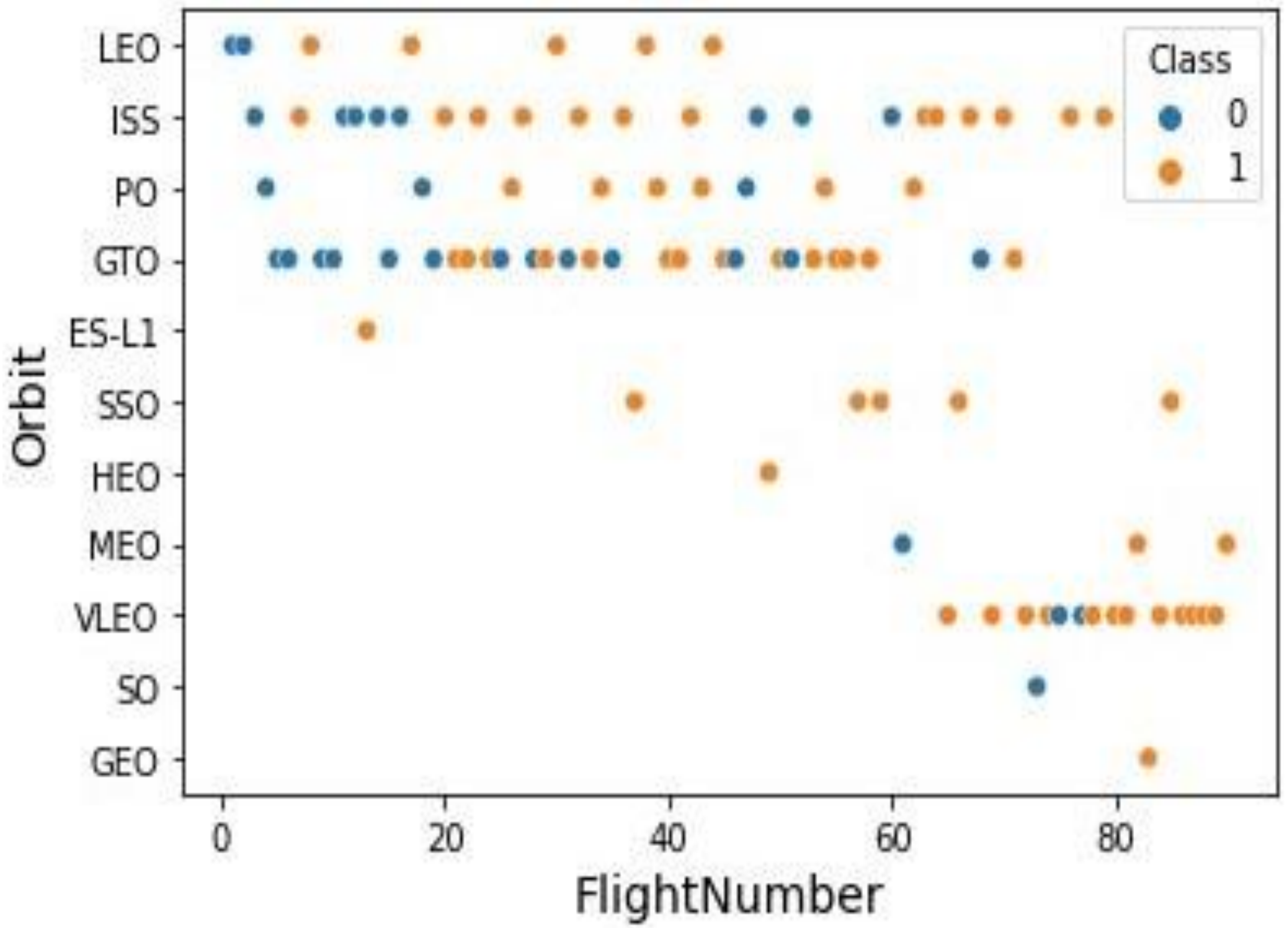
Success Rate vs. Orbit Type

- Higher orbits such as ES-L1, GEO, HEO and SSO have a higher success rate than lower orbits.
- SpaceX has only made single launches to each higher orbital.



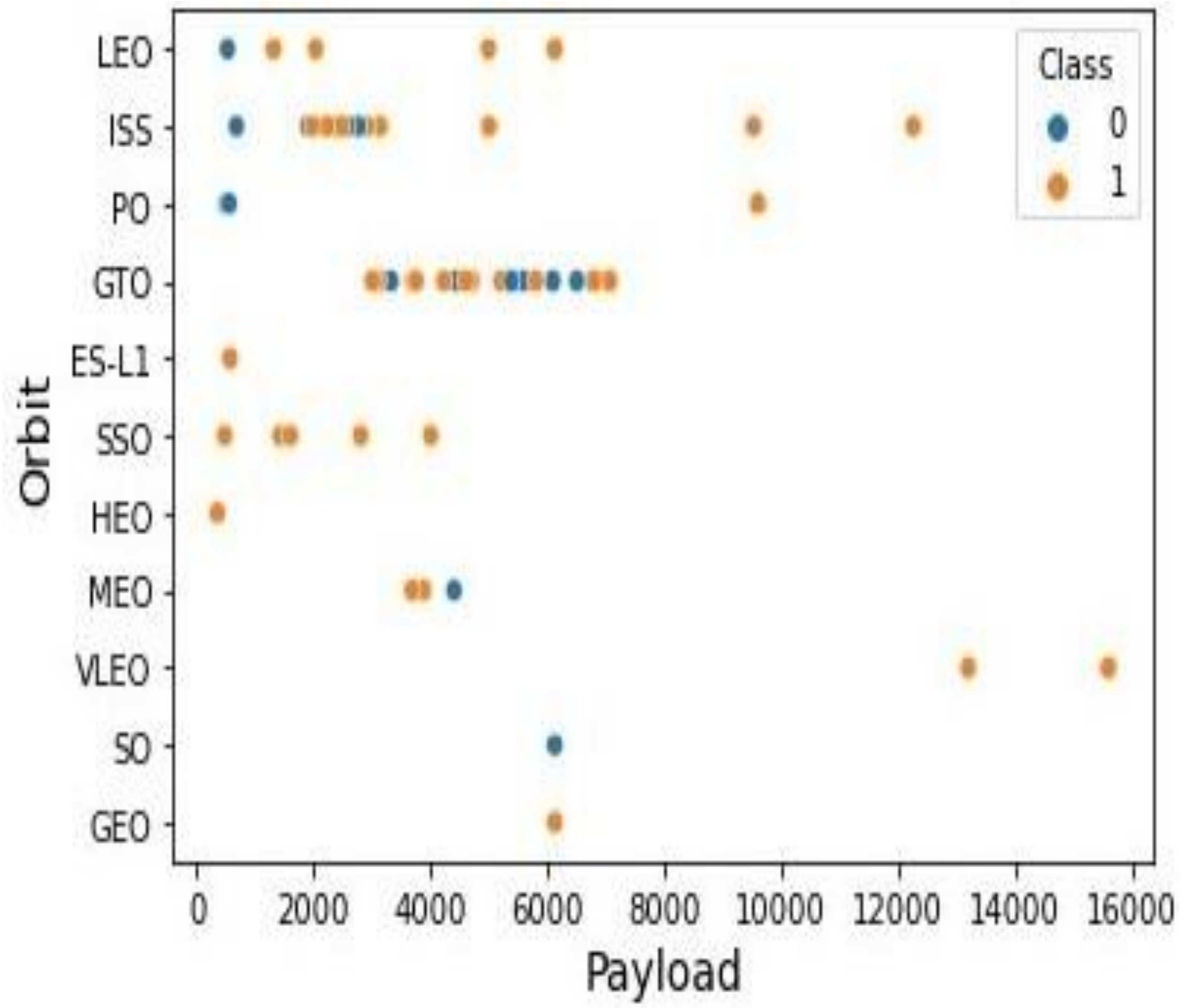
Flight Number vs. Orbit Type

- No consistent relationship between orbits and number of flights.
- For LEO orbit, success is directly related to number of flights.
- There is no relationship between ISS, GTO and the number of flights.



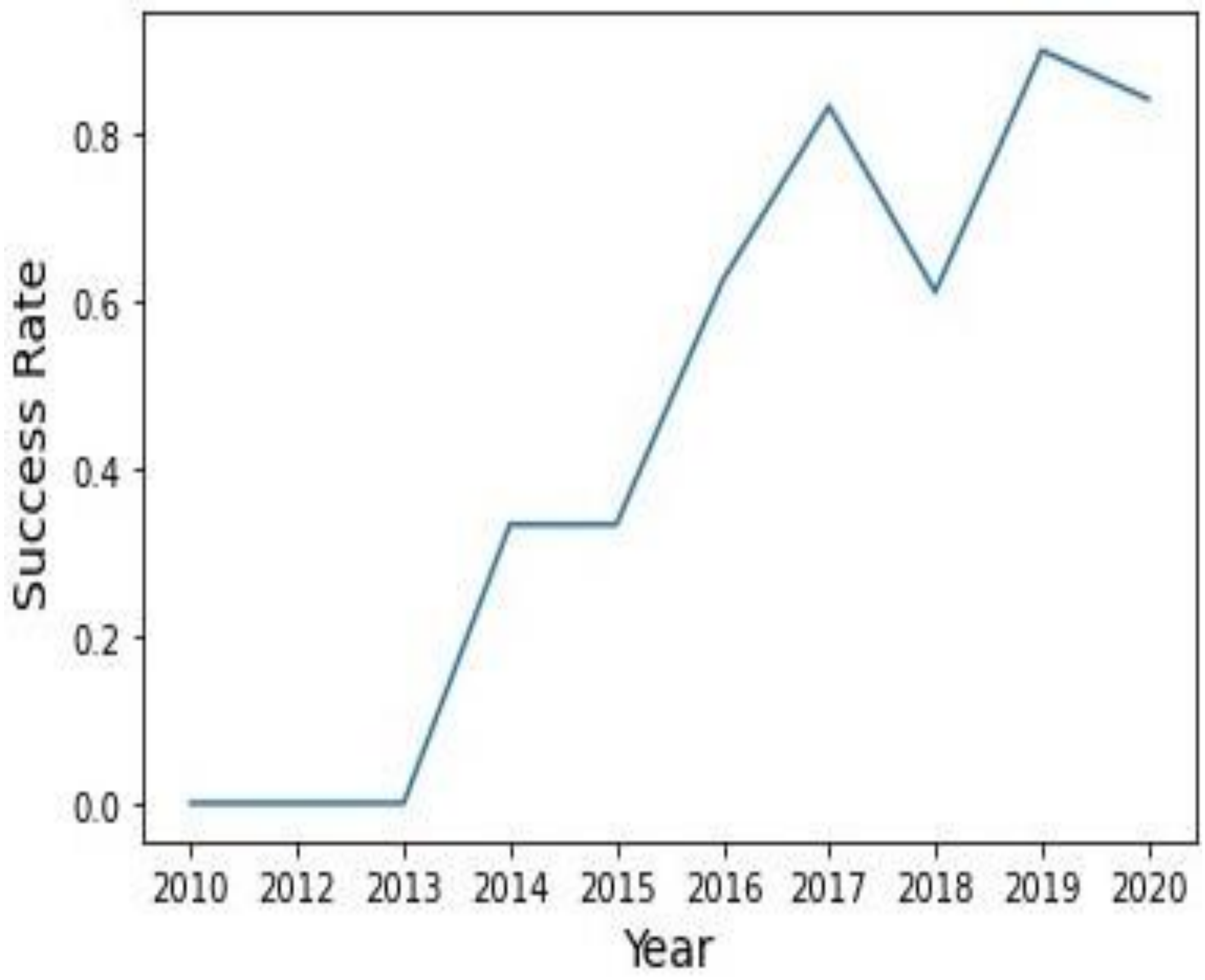
Payload vs. Orbit Type

- Heavy payloads to Polar, LEO and ISS orbits are more successful.
- No consistent relationship between payload mass and GTO.



Launch Success Yearly Trend

- Overall, success rate has been increasing since 2013.
- No growth in 2014 and a sharp drop in 2017.



All Launch Site Names

- The names of all unique launch sites

- SpaceX operates from four sites:
  - CCAFS SLC-40
  - KSC LC-39A
  - CCAFS LC-40
  - VAFB SLC-4E

```
%sql select DISTINCT (launch_site) from spacextbl;
```

```
 * postgresql+psycopg2://postgres:***@localhost:5432/mlprojects
4 rows affected.
```

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

Launch Sites beginning with 'CCA'

- A list of five records for matching site.

```
[6]: %sql select * from SPACEXTBL where launch_site like 'CCA%' limit 5

 * postgresql+psycopg2://postgres:***@localhost:5432/mlprojects
5 rows affected.
```

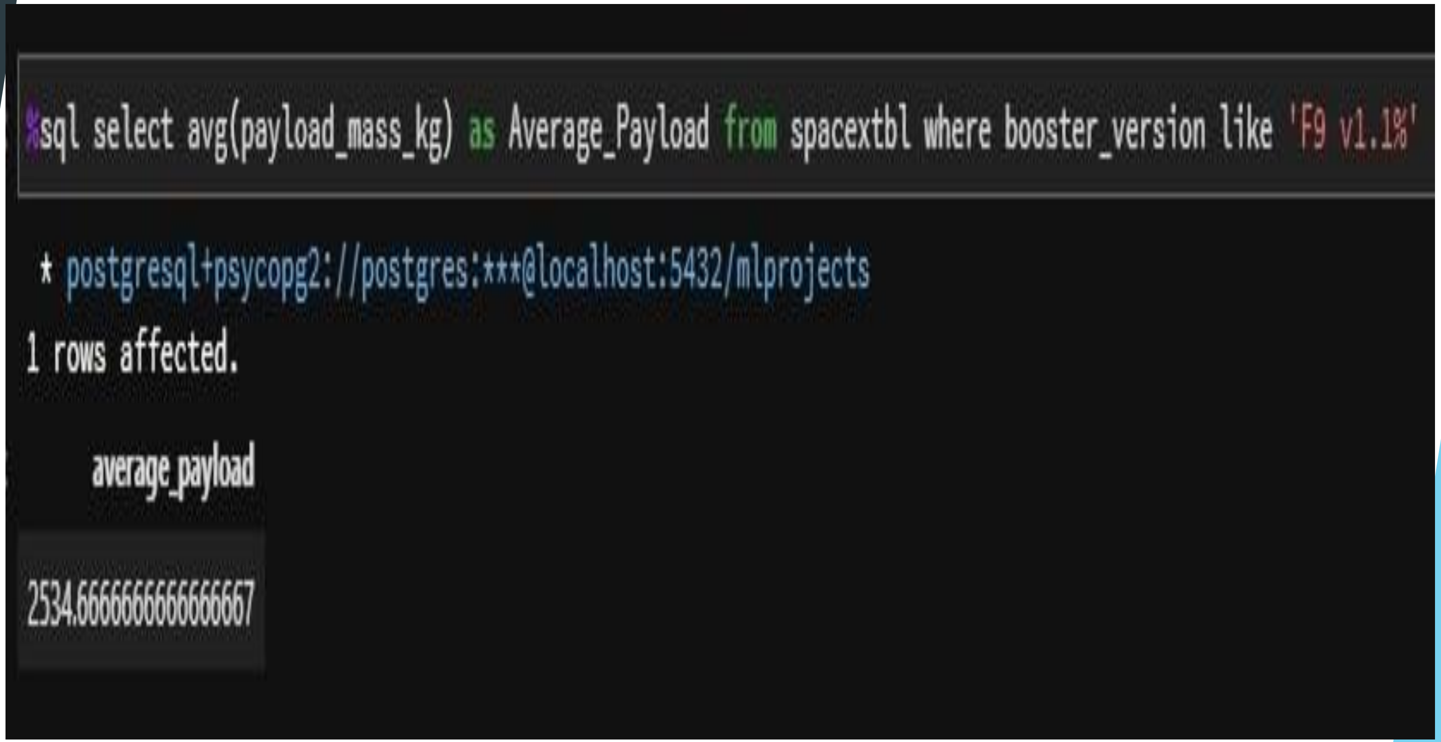
flight_id	date_	time_	booster_version	launch_site	payload	payload_mass_kg	orbit	customer	mission_outcome	landing_outcome
4	2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
5	2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
6	2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
7	2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
8	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

- The total payload carried by boosters for NASA
  - 45 596 kg

Average Payload Mass by F9 v1.1

- The average payload mass carried by booster version F9 v1.1
  - 2 534.67 kg



First Successful Ground Landing Date

- The date of the first successful landing outcome on ground pad
    - 2015-12-22

```
%sql select min(date_) from spacextbl where landing_outcome = 'Success (ground pad)'

 * postgresql+psycopg2://postgres:***@localhost:5432/mlprojects
1 rows affected.

    min
2015-12-22
```

Successful Drone Ship Landing Payload between 4000 & 6000

- Boosters which have successfully landed on drone ship
- Payload mass greater than 4000 but less than 6000.

```
%sql select booster_version from spacextbl where landing_outcome = 'Success (drone ship)' and (payload_mass_kg between 4000 and 6000)

 * postgresql+psycopg2://postgres:***@localhost:5432/mlprojects
4 rows affected.
```

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

Total Number of Successful and Failure Mission Outcomes

- Total number of outcomes:
  - Successful: 100
  - Failure: 1

```
%sql select mission_outcome, COUNT(*) as Total from spacextbl group by mission_outcome
```

```
 * postgresql+psycopg2://postgres:***@localhost:5432/mlprojects
4 rows affected.
```

mission_outcome	total
Success (payload status unclear)	1
Success	98
Success	1
Failure (in flight)	1

Boosters Carried Maximum Payload

- List of booster which have carried the maximum payload mass

- Result obtained using a subquery.

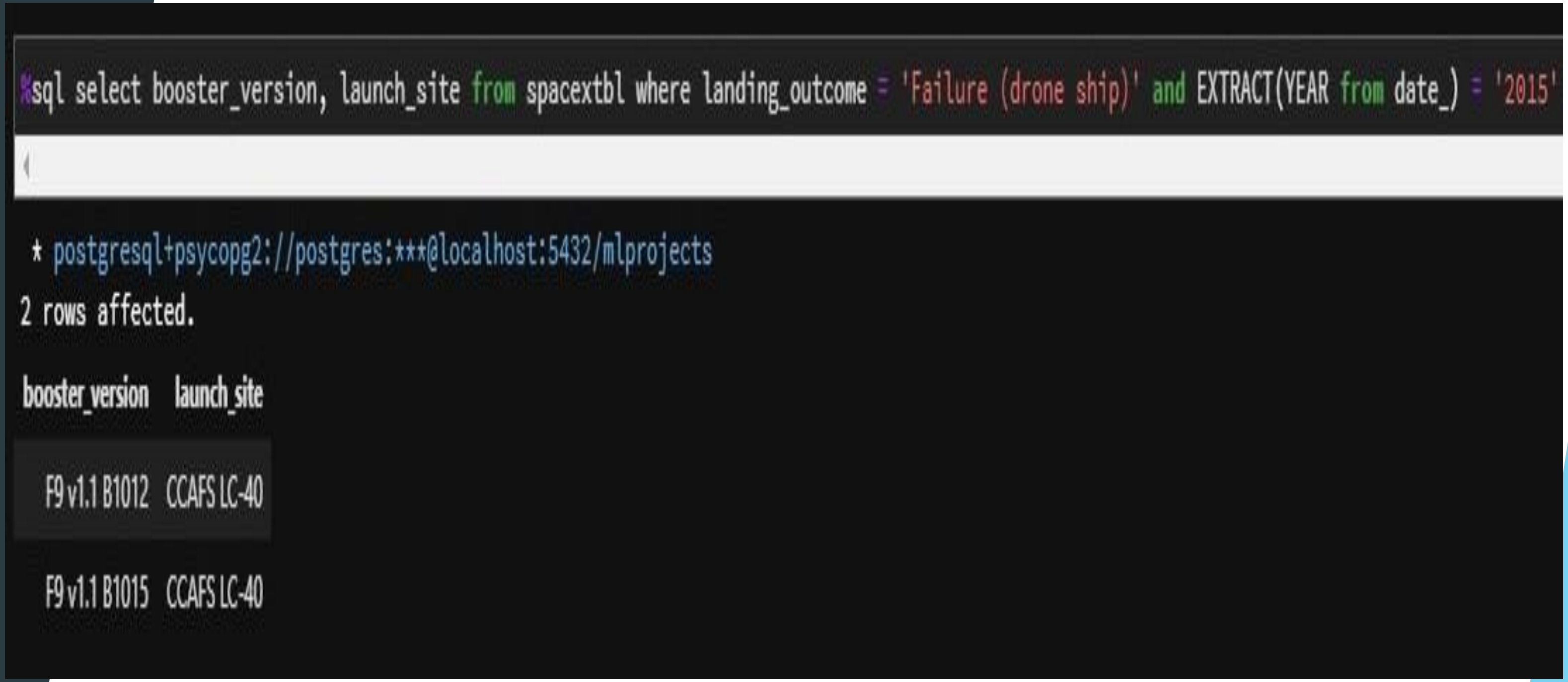
```
# Using a subquery
%sql select Distinct (booster_version) from spacextbl where payload_mass_kg = (select max(payload_mass_kg) from spacextbl)
```

```
 * postgresql+psycopg2://postgres:***@localhost:5432/mlprojects
12 rows affected.
```

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

- Failed landing outcomes on drone ship – 2015.
- Booster versions and launch site names.
  - 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

- Ranking the count of landing outcomes
- Between 2010-06-04 & 2017-03-20. Descending order

```
%sql select landing_outcome, count(*) as Count from spacextbl where date_ between '2010-06-04' and '2017-03-20' group by landing_outcome order by Cou
```

```
* postgresql+psycopg2://postgres:***@localhost:5432/mlprojects
8 rows affected.
```

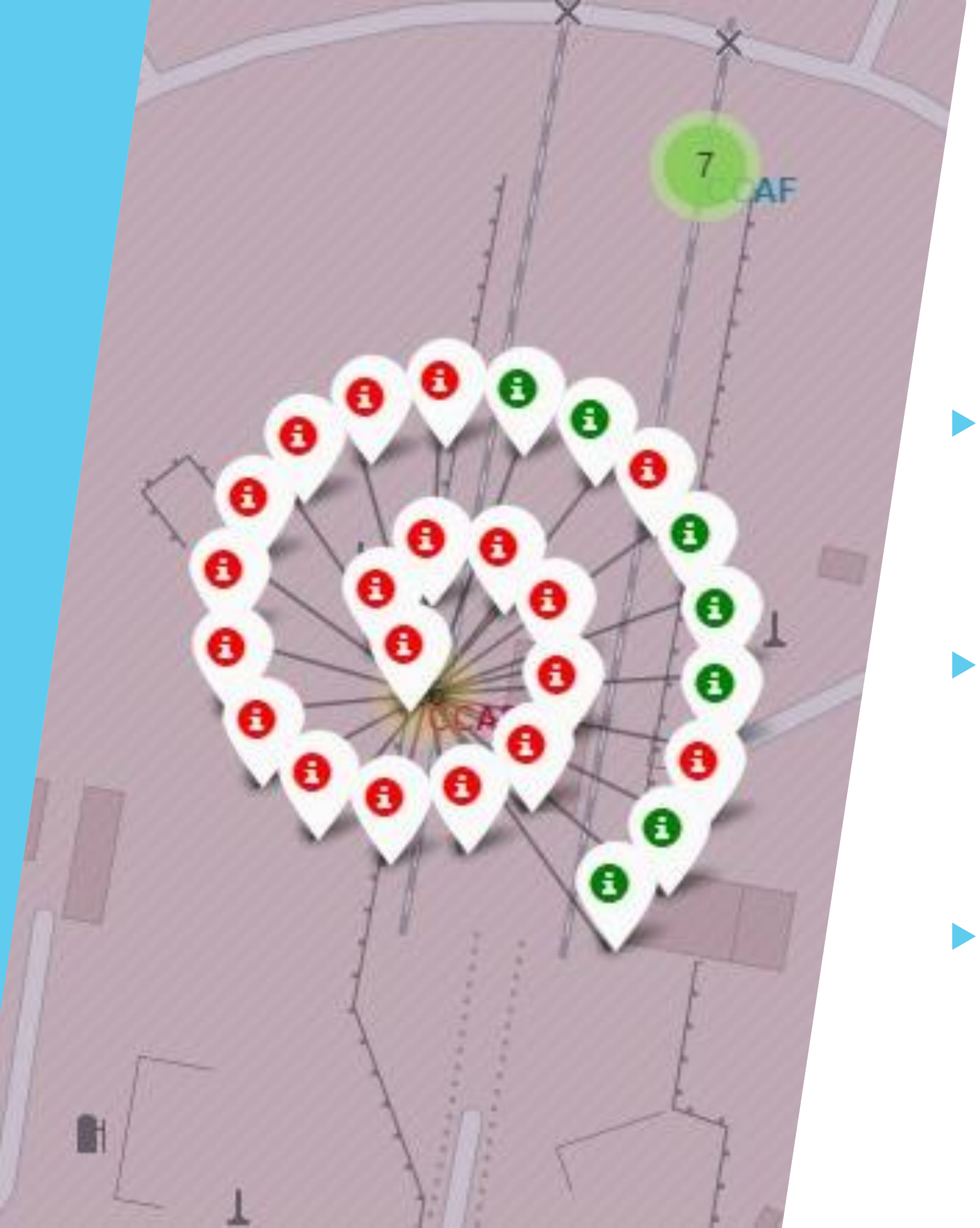
landing_outcome	count
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 4

Launch Sites Proximities Analysis

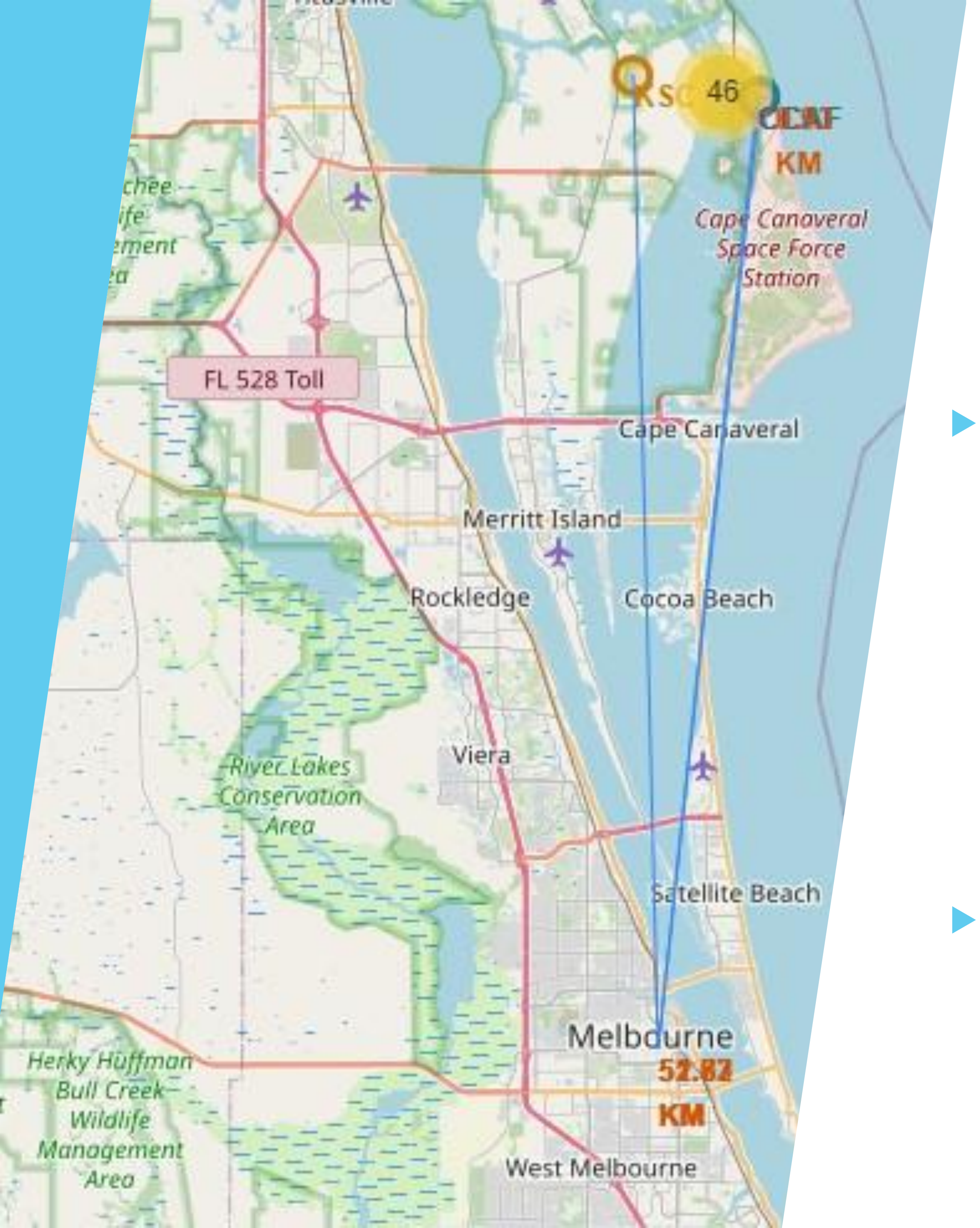
Launch Site Analysis - Locations

- All four launch sites are marked on the map.
- Sites are located near the coast.
- East Coast:
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
- West Coast:
  - VAFB SLC-4E



Launch Site Analysis - Markers

- Markers added to the allow a quick visual analysis of launch outcomes at each site.
- The colors represent:
  - Green – Successful launch
  - Red – Failed launch
- Markers are also clustered to easily show location count.



Launch Site Analysis - Proximities

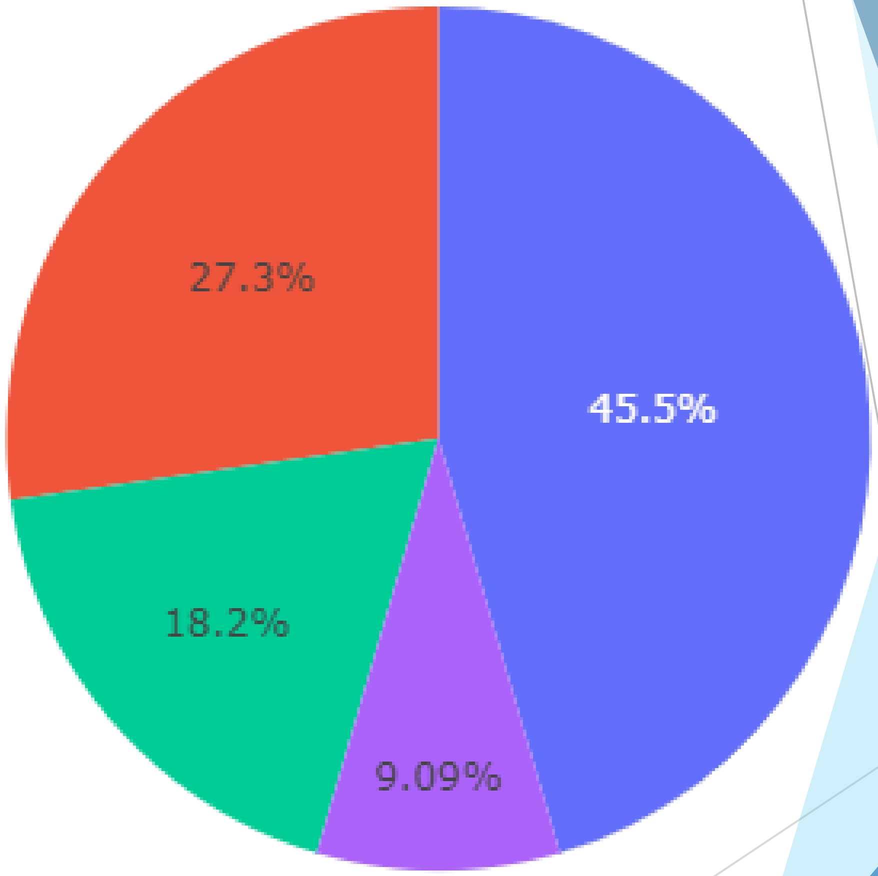
- Additional features added in the analysis include:
  - Coastline marker
  - Cities markers – Melbourne & Santa Maria
  - Distance lines
  - Distance markers
  - Rail lines
- Distances calculated are in km.

Section 5

Build a Dashboard with Plotly Dash

Dashboard – Pie Chart

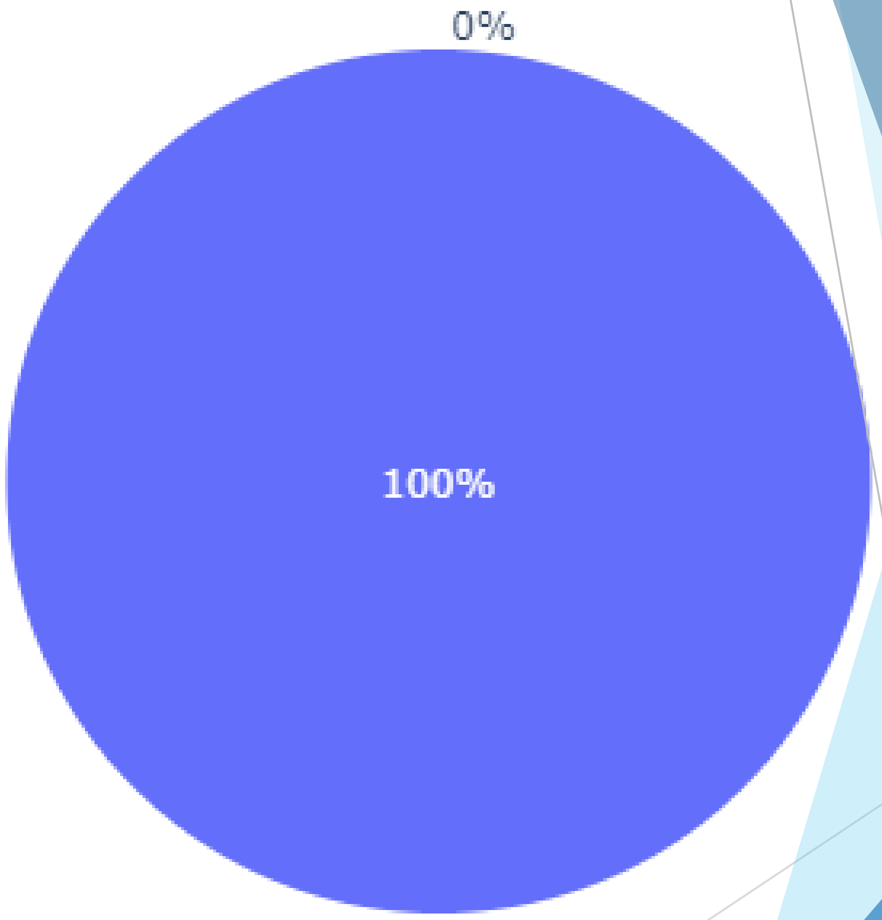
- Total Successful launches by site.
  - KSC LC-39A - 45.5%
  - CCAFS LC-40 – 27.3%
  - VAFB SLC-4E – 18.2%
  - CCACCAFS SLC-40 - 9.09%



Dashboard – Pie Chart 2

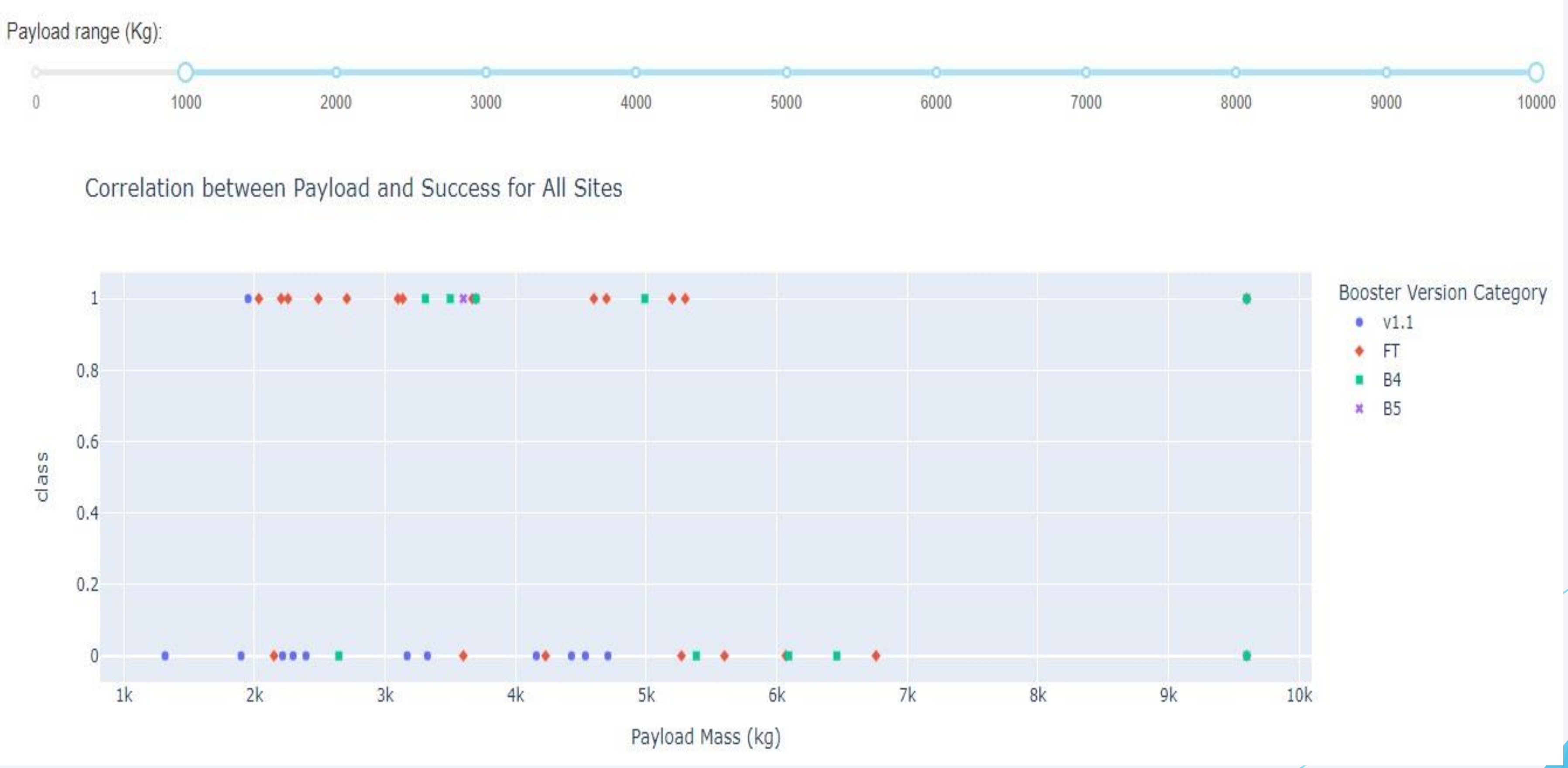
- Launch site with highest launch success ratio - KSC LC-39A.

- All launches at this location have been successful.



Dashboard – Payload and Booster

- KSC LC 39A has the highest success rate – 41.7%
- The FT booster version has the highest success rate – at medium payload mass.



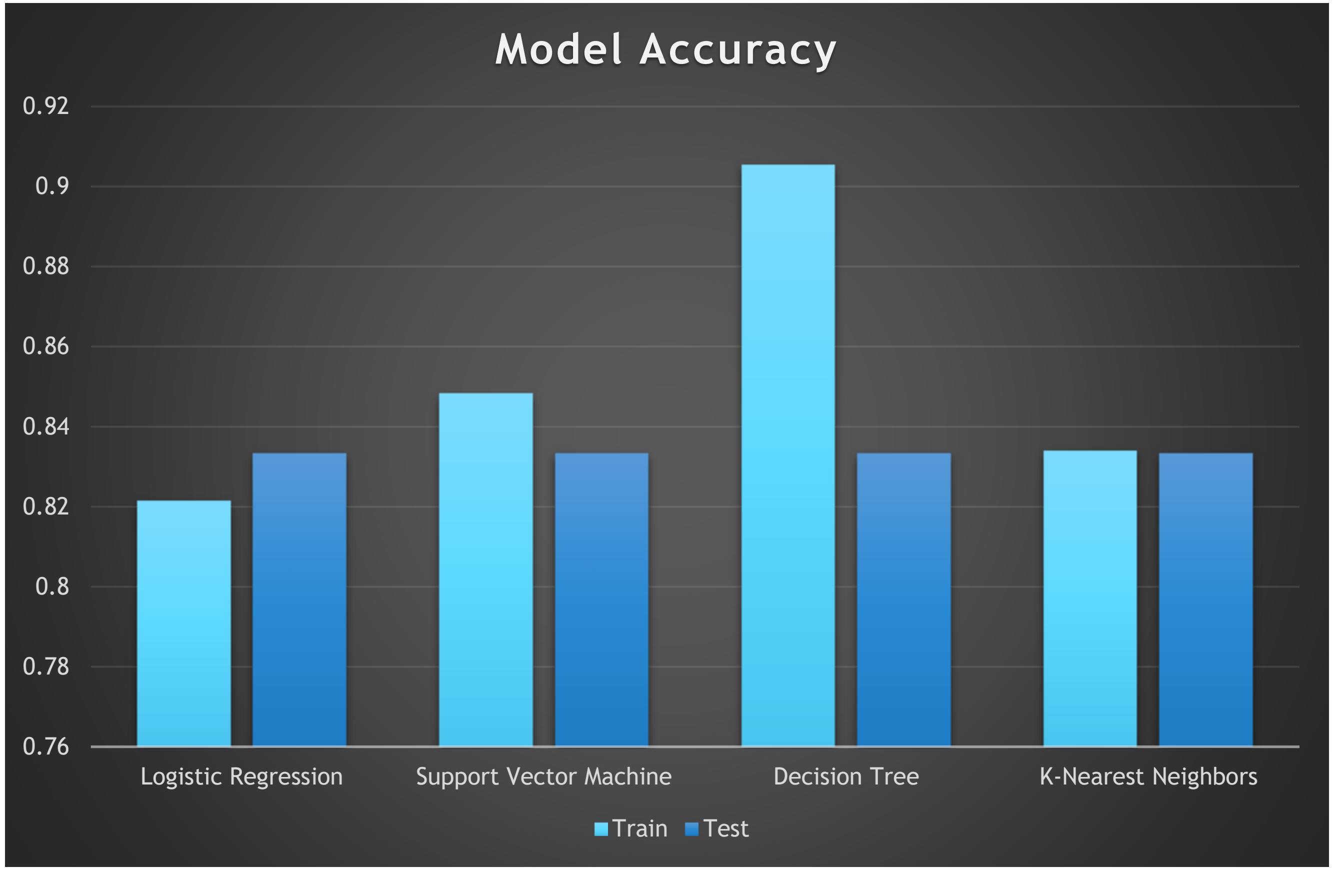
Section 6

Predictive Analysis (Classification)

Classification Accuracy

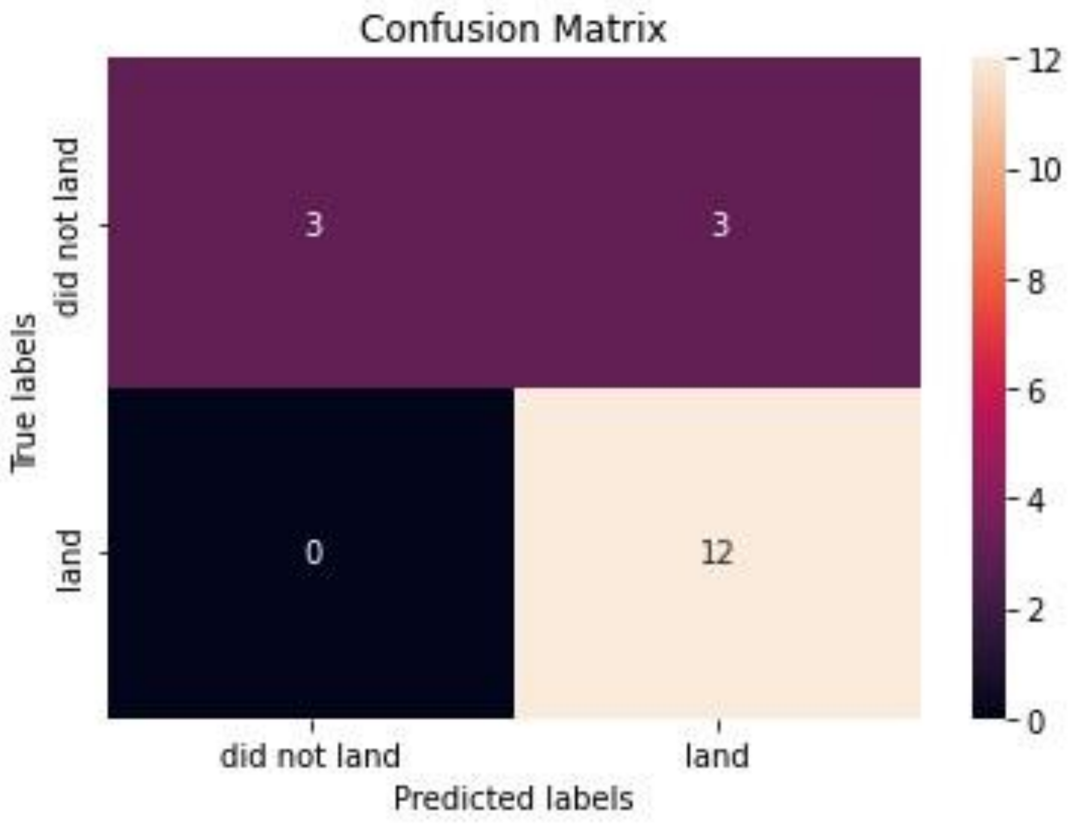
- Decision Tree model shows a very accuracy on the training set.

- All models show similar performance on the test set.
  - Accuracy – 83.33%



Confusion Matrix

- All the models show the same performance on test data.
- The accuracy for all models – 83.33%



Conclusions

- We can predict landing outcome with 83.3% accuracy.
- FT booster versions offer the highest success rate for medium payload mass.
- Higher orbital destinations contribute to higher success rate.
- Success increases with both mass and number of flights.
- Chosen launch site must be in proximity of coast and rail lines.

Appendix

- List of project files.
- Notebooks:
  - API Data Collection
  - Web Scraping Data
  - Data Wrangling
  - EDA with SQL
  - EDA with Data Visualisation
  - Launch Site Location Analysis
  - Predictive Analysis (Classification)

Thank you!