IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Jared Platt
March 23rd, 2024

Table of Contents

Executive Summary

How is SpaceX so successful and how could we replicate, or even improve on, their cost-effective methodology if we started from scratch? To answer this question, SpaceX launch data was analyzed to identify any trends that may contribute to successful missions. These trends were then used to predict launch outcomes.

Data was collected from the SpaceX REST API and SpaceX Launch Wiki. Data was processed and wrangled using Pandas and Numpy. After collection and processing, Exploratory Data Analysis (EDA) was performed through visualization and SQL queries. Interactive maps and dashboards were created to allow for an intuitive evaluation of all data. Finally, all the insights collected through EDA were used to build and optimize various machine learning models that can predict the outcome of a launch from various characteristics (booster(s), launch site, payload, etc.).

Learning from SpaceX historical data, a few key areas need to be prioritized for success. KSC LC-39A was determined to be the best launch site based on success rate. SpaceY should mimic this site in terms of its distance to relevant landmarks (i.e. cities, coasts, railways, highways). Payload size can have a significant effect on outcome. Launches with payloads between 2500 kg – 5000 kg have the highest success rate. The Falcon9 booster also has a high success rate with this payload range. These two factors should be implemented into SpaceY. Finally, landings should take place on ground pads due to the high success rate (0 failures). However, SpaceY will need to ensure a large distance (>100 km) from the nearest city to prevent any mass casualties in the event of a crash landing. Boosters, especially one similar to the Falcon9, need to be reused for cost savings. The price of launching a rocket with reused boosters is 1/3 that of a new rocket. SpaceY can be 87.3% confident that the SpaceX historical data will translate into real world SpaceY events.

Introduction

- I am a data scientist working for a new rocket company, Space Y, that would like to compete with SpaceX founded by Billionaire industrialist Allon Musk. My job is to determine the cost of each launch, whether to reuse the first stage, and identify what contributes to a successful launch so that SpaceY can mimic those procedures.
- We will perform data collection, data processing, exploratory data analysis, build interactive maps and dashboards, and use machine learning to predict outcomes based on launch data.
- What do we want to know?
    - Best location for launch sites
    - Optimal payload sizes to maximize success rate
    - Most successful orbit types
    - Most successful booster versions
    - Most successful landing pad to allow booster reuse

SPACE Y

Allon Mask

Section 1

Methodology

Methodology

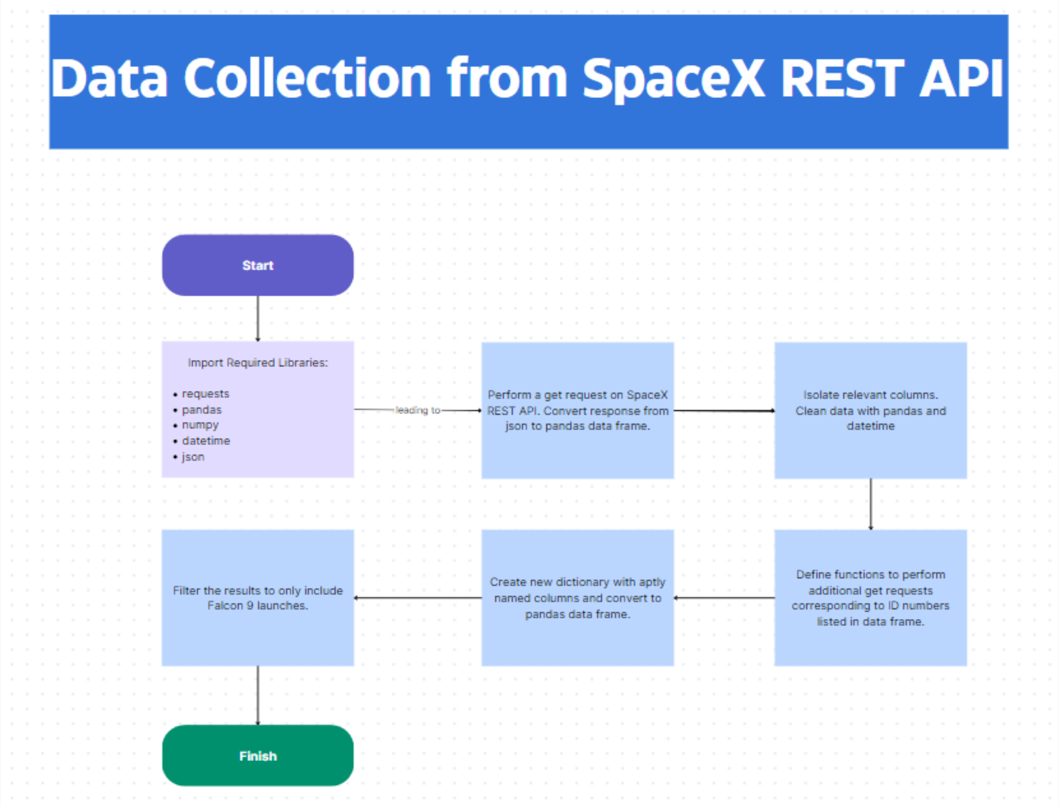
Executive Summary

- Data collection methodology
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
    - How to build, tune, evaluate classification models

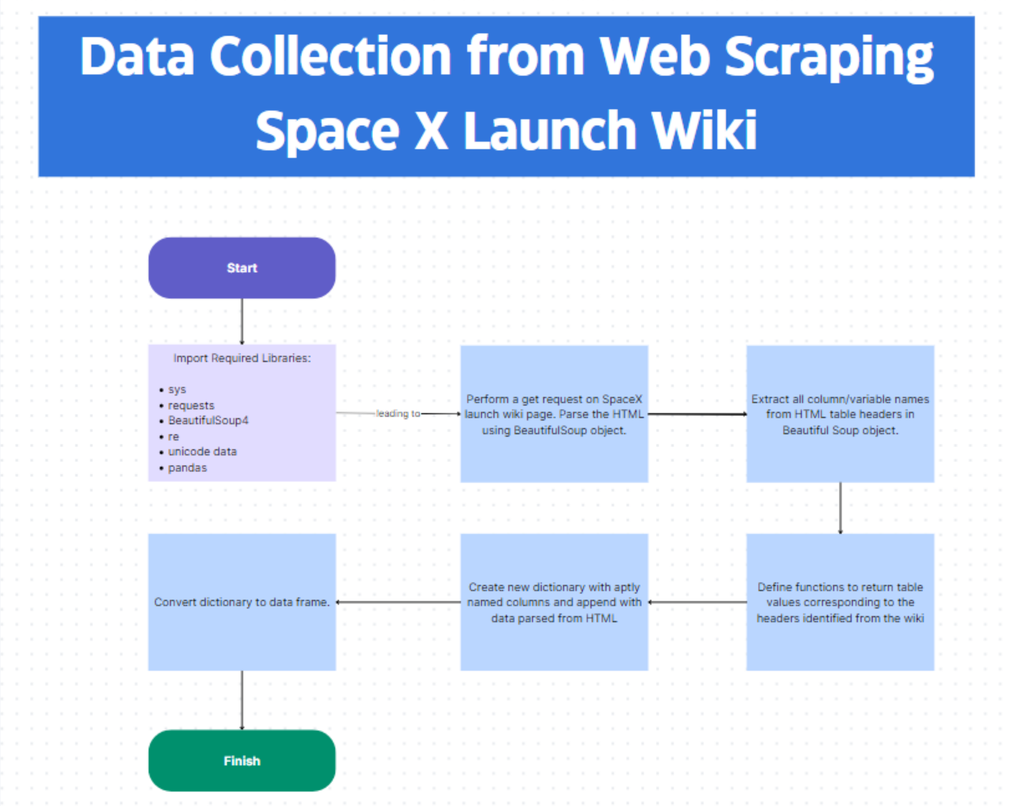
Data Collection

Collected SpaceX launch and landing data using two methods:

- SpaceX REST API

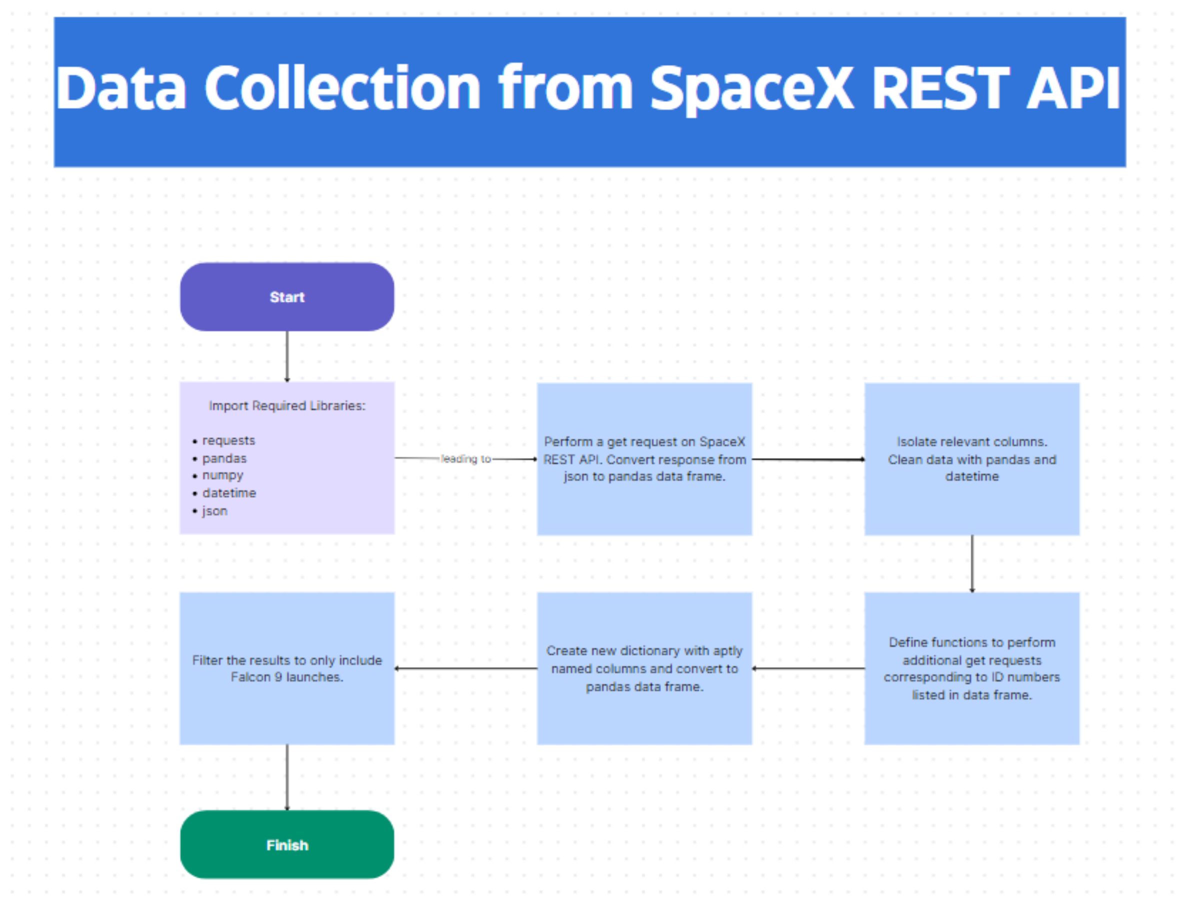


- Web Scraping



Data Collection – SpaceX API

- Get requests from SpaceX Rest API:
  - Launch and landing data for Falcon9
    - Rocket used
    - Payload delivered
    - Launch specifications
    - Landing specifications
    - Landing outcome
  - Parsed response to yield data frame of relevant columns
  - Cleaned data with pandas



https://github.com/jaredplatt/Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

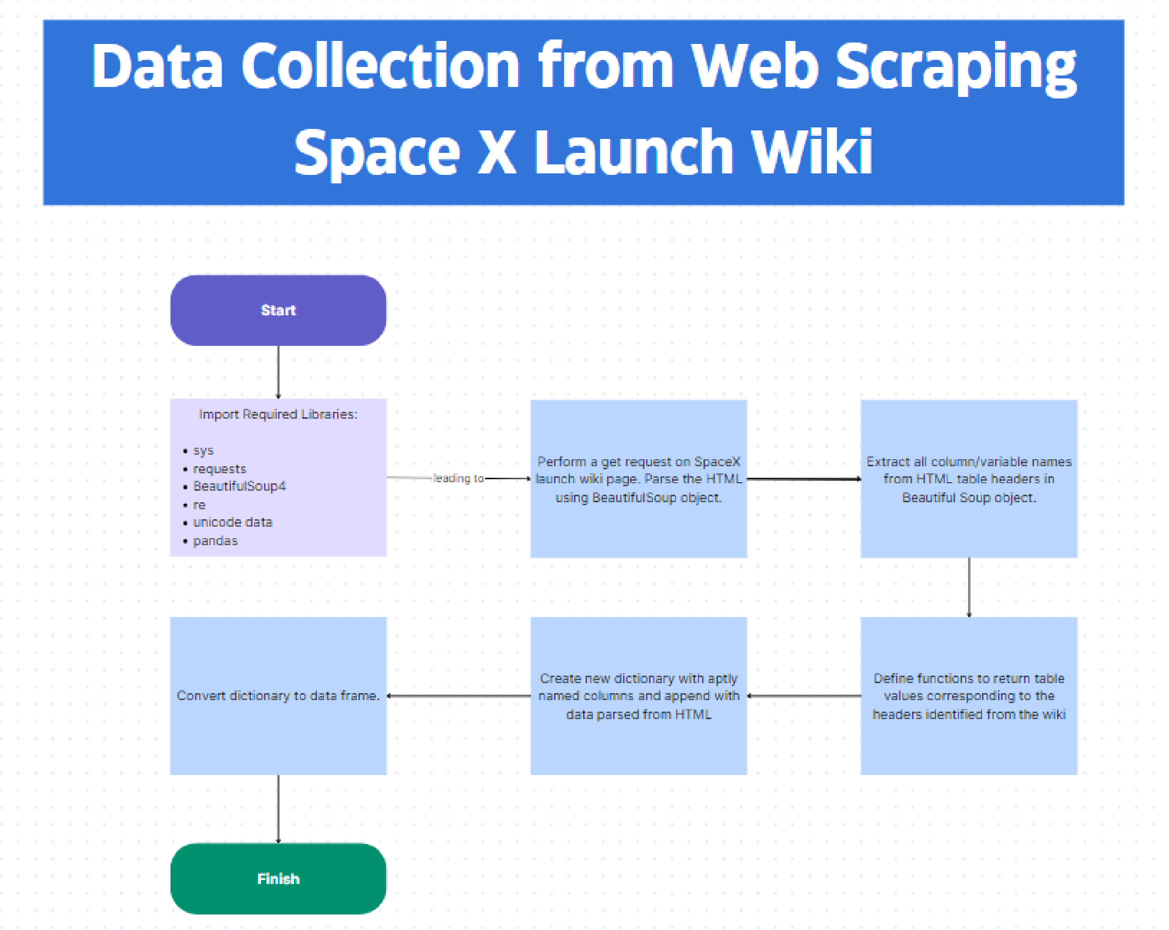
Data Collection – SpaceX API

- CSV created via SpaceX REST API

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
1	1	2010-06-04	Falcon 9	6123.547647058824	LEO	CCSFS SLC 40	None None	1	False	False	False		1.0	0	B0003	-80.577366	28.5618571
2	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False		1.0	0	B0005	-80.577366	28.5618571
3	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False		1.0	0	B0007	-80.577366	28.5618571
4	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False		1.0	0	B1003	-120.610829	34.632093
5	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False		1.0	0	B1004	-80.577366	28.5618571
6	6	2014-01-06	Falcon 9	3325.0	GTO	CCSFS SLC 40	None None	1	False	False	False		1.0	0	B1005	-80.577366	28.5618571
7	7	2014-04-18	Falcon 9	2296.0	ISS	CCSFS SLC 40	True Ocean	1	False	False	True		1.0	0	B1006	-80.577366	28.5618571
8	8	2014-07-14	Falcon 9	1316.0	LEO	CCSFS SLC 40	True Ocean	1	False	False	True		1.0	0	B1007	-80.577366	28.5618571
9	9	2014-08-05	Falcon 9	4535.0	GTO	CCSFS SLC 40	None None	1	False	False	False		1.0	0	B1008	-80.577366	28.5618571
10	10	2014-09-07	Falcon 9	4428.0	GTO	CCSFS SLC 40	None None	1	False	False	False		1.0	0	B1011	-80.577366	28.5618571
11	11	2014-09-21	Falcon 9	2216.0	ISS	CCSFS SLC 40	False Ocean	1	False	False	False		1.0	0	B1010	-80.577366	28.5618571
12	12	2015-01-10	Falcon 9	2395.0	ISS	CCSFS SLC 40	False ASDS	1	True	False	True	5e9e3032383ecb761…	1.0	0	B1012	-80.577366	28.5618571
13	13	2015-02-11	Falcon 9	570.0	ES-L1	CCSFS SLC 40	True Ocean	1	True	False	True		1.0	0	B1013	-80.577366	28.5618571
14	14	2015-04-14	Falcon 9	1898.0	ISS	CCSFS SLC 40	False ASDS	1	True	False	True	5e9e3032383ecb761…	1.0	0	B1015	-80.577366	28.5618571
15	15	2015-04-27	Falcon 9	4707.0	GTO	CCSFS SLC 40	None None	1	False	False	False		1.0	0	B1016	-80.577366	28.5618571
16	16	2015-06-28	Falcon 9	2477.0	ISS	CCSFS SLC 40	None ASDS	1	True	False	True	5e9e3032383ecb6bb…	1.0	0	B1018	-80.577366	28.5618571
17	17	2015-12-22	Falcon 9	2034.0	LEO	CCSFS SLC 40	True RTLS	1	True	False	True	5e9e3032383ecb267…	1.0	0	B1019	-80.577366	28.5618571
18	18	2016-01-17	Falcon 9	553.0	PO	VAFB SLC 4E	False ASDS	1	True	False	True	5e9e3033383ecbb9e…	1.0	0	B1017	-120.610829	34.632093
19	19	2016-03-04	Falcon 9	5271.0	GTO	CCSFS SLC 40	False ASDS	1	True	False	True	5e9e3032383ecb6bb…	1.0	0	B1020	-80.577366	28.5618571
20	20	2016-04-08	Falcon 9	3136.0	ISS	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb…	2.0	1	B1021	-80.577366	28.5618571
21	21	2016-05-06	Falcon 9	4696.0	GTO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb…	2.0	0	B1022	-80.577366	28.5618571
22	22	2016-05-27	Falcon 9	3100.0	GTO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb…	2.0	1	B1023	-80.577366	28.5618571
23	23	2016-07-18	Falcon 9	2257.0	ISS	CCSFS SLC 40	True RTLS	1	True	False	True	5e9e3032383ecb267…	2.0	1	B1025	-80.577366	28.5618571
24	24	2016-08-14	Falcon 9	4600.0	GTO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb…	2.0	0	B1026	-80.577366	28.5618571
25	25	2016-09-01	Falcon 9	5500.0	GTO	CCSFS SLC 40	None ASDS	1	True	False	True	5e9e3032383ecb6bb…	3.0	0	B1028	-80.577366	28.5618571
26	26	2017-01-14	Falcon 9	9600.0	PO	VAFB SLC 4E	True ASDS	1	True	False	True	5e9e3033383ecbb9e…	3.0	1	B1029	-120.610829	34.632093
27	27	2017-02-19	Falcon 9	2490.0	ISS	KSC LC 39A	True RTLS	1	True	False	True	5e9e3032383ecb267…	3.0	1	B1031	-80.6039558	28.6080585
28	28	2017-03-16	Falcon 9	5600.0	GTO	KSC LC 39A	None None	1	False	False	False		3.0	0	B1030	-80.6039558	28.6080585
29	29	2017-03-30	Falcon 9	5300.0	GTO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb…	2.0	1	B1021	-80.6039558	28.6080585
30	30	2017-05-01	Falcon 9	6123.547647058824	LEO	KSC LC 39A	True RTLS	1	True	False	True	5e9e3032383ecb267…	3.0	1	B1032	-80.6039558	28.6080585
31	31	2017-05-15	Falcon 9	6070.0	GTO	KSC LC 39A	None None	1	False	False	False		3.0	0	B1034	-80.6039558	28.6080585
32	32	2017-06-03	Falcon 9	2708.0	ISS	KSC LC 39A	True RTLS	1	True	False	True	5e9e3032383ecb267…	3.0	1	B1035	-80.6039558	28.6080585
33	33	2017-06-23	Falcon 9	3669.0	GTO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb…	3.0	1	B1029	-80.6039558	28.6080585
34	34	2017-06-25	Falcon 9	9600.0	PO	VAFB SLC 4E	True ASDS	1	True	False	True	5e9e3033383ecbb9e…	3.0	1	B1036	-120.610829	34.632093
35	35	2017-07-05	Falcon 9	6761.0	GTO	KSC LC 39A	None None	1	False	False	False		3.0	0	B1037	-80.6039558	28.6080585
36	36	2017-08-14	Falcon 9	2910.0	ISS	KSC LC 39A	True RTLS	1	True	False	True	5e9e3032383ecb267…	4.0	1	B1039	-80.6039558	28.6080585
37	37	2017-08-24	Falcon 9	475.0	SSO	VAFB SLC 4E	True ASDS	1	True	False	True	5e9e3033383ecbb9e…	3.0	1	B1038	-120.610829	34.632093
38	38	2017-09-07	Falcon 9	4990.0	LEO	KSC LC 39A	True RTLS	1	True	False	True	5e9e3032383ecb267…	4.0	1	B1040	-80.6039558	28.6080585
39	39	2017-10-09	Falcon 9	9600.0	PO	VAFB SLC 4E	True ASDS	1	True	False	True	5e9e3033383ecbb9e…	4.0	1	B1041	-120.610829	34.632093
40	40	2017-10-11	Falcon 9	5200.0	GTO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb…	3.0	1	B1031	-80.6039558	28.6080585
41	41	2017-10-30	Falcon 9	3700.0	GTO	KSC LC 39A	True ASDS	1	True	False	True	5e9e3032383ecb6bb…	4.0	0	B1042	-80.6039558	28.6080585
42	42	2017-12-15	Falcon 9	2205.0	ISS	CCSFS SLC 40	True RTLS	2	True	True	True	5e9e3032383ecb267…	3.0	1	B1035	-80.577366	28.5618571
43	43	2017-12-23	Falcon 9	9600.0	PO	VAFB SLC 4E	True Ocean	2	True	True	False		3.0	1	B1036	-120.610829	34.632093
44	44	2018-01-08	Falcon 9	6123.547647058824	LEO	CCSFS SLC 40	True RTLS	1	True	False	True	5e9e3032383ecb267…	4.0	1	B1043	-80.577366	28.5618571
45	45	2018-01-31	Falcon 9	4230.0	GTO	CCSFS SLC 40	True Ocean	2	True	True	True		3.0	1	B1032	-80.577366	28.5618571
46	46	2018-03-06	Falcon 9	6092.0	GTO	CCSFS SLC 40	None None	1	True	False	True		4.0	0	B1044	-80.577366	28.5618571
47	47	2018-03-30	Falcon 9	9600.0	PO	VAFB SLC 4E	None None	2	True	True	True		4.0	1	B1041	-120.610829	34.632093
48	48	2018-04-02	Falcon 9	2760.0	ISS	CCSFS SLC 40	None None	2	True	True	True		4.0	1	B1039	-80.577366	28.5618571
49	49	2018-04-18	Falcon 9	350.0	HEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb…	4.0	1	B1045	-80.577366	28.5618571
50	50	2018-05-11	Falcon 9	3750.0	GTO	KSC LC 39A	True ASDS	1	True	False	True	5e9e3032383ecb6bb…	5.0	3	B1046	-80.6039558	28.6080585

Data Collection - Scraping

- HTTP GET method to request the Falcon9 Launch HTML page
- Create Beautiful Soup object and parse for:
  - Column Headers
  - Values
- Append all requested data into a dictionary with proper column names
- Convert to pandas data frame.



https://github.com/jaredplatt/Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb

Data Collection - Scraping

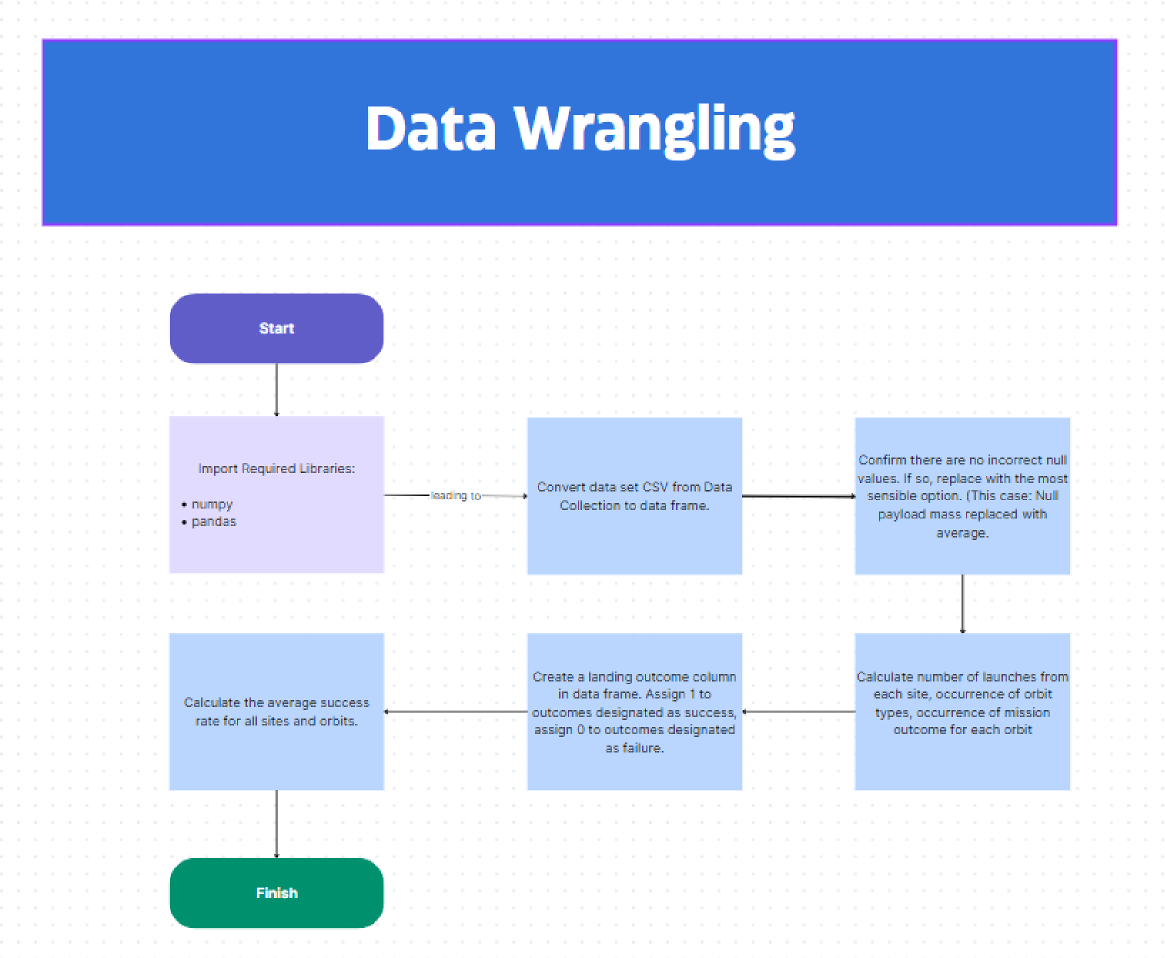
- CSV created via Web Scraping from SpaceX Launch Wiki

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
1	1	CCAFS	Dragon Spacecraft Q...	0	LEO	SpaceX	Success	F9 v1.0B0003.1	Failure	4 June 2010	18:45
2	2	CCAFS	Dragon	0	LEO	NASA (COTS) NRO	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
3	3	CCAFS	Dragon	525 kg	LEO	NASA (COTS)	Success	F9 v1.0B0005.1	No attempt	22 May 2012	07:44
4	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA (CRS)	Success	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
5	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA (CRS)	Success	F9 v1.0B0007.1	No attempt	1 March 2013	15:10
6	6	VAFB	CASSIOPE	500 kg	Polar orbit	MDA	Success	F9 v1.1B1003	Uncontrolled	29 September 2013	16:00
7	7	CCAFS	SES-8	3,170 kg	GTO	SES	Success	F9 v1.1	No attempt	3 December 2013	22:41
8	8	CCAFS	Thaicom 6	3,325 kg	GTO	Thaicom	Success	F9 v1.1	No attempt	6 January 2014	22:06
9	9	Cape Canaveral	SpaceX CRS-3	2,296 kg	LEO	NASA (CRS)	Success	F9 v1.1	Controlled	18 April 2014	19:25
10	10	Cape Canaveral	Orbcomm-OG2	1,316 kg	LEO	Orbcomm	Success	F9 v1.1	Controlled	14 July 2014	15:15
11	11	Cape Canaveral	AsiaSat 8	4,535 kg	GTO	AsiaSat	Success	F9 v1.1	No attempt	5 August 2014	08:00
12	12	Cape Canaveral	AsiaSat 6	4,428 kg	GTO	AsiaSat	Success	F9 v1.1	No attempt	7 September 2014	05:00
13	13	Cape Canaveral	SpaceX CRS-4	2,216 kg	LEO	NASA (CRS)	Success	F9 v1.1	Uncontrolled	21 September 2014	05:52
14	14	Cape Canaveral	SpaceX CRS-5	2,395 kg	LEO	NASA (CRS)	Success	F9 v1.1	Failure	10 January 2015	09:47
15	15	Cape Canaveral	DSCOVR	570 kg	HEO	USAF NASA NOAA	Success	F9 v1.1	Controlled	11 February 2015	23:03
16	16	Cape Canaveral	ABS-3A	4,159 kg	GTO	ABS Eutelsat	Success	F9 v1.1	No attempt	2 March 2015	03:50
17	17	Cape Canaveral	SpaceX CRS-6	1,898 kg	LEO	NASA (CRS)	Success	F9 v1.1	Failure	14 April 2015	20:10
18	18	Cape Canaveral	TürkmenÄlem 52°E / ...	4,707 kg	GTO	Turkmenistan Nationa...	Success	F9 v1.1	No attempt	27 April 2015	23:03
19	19	Cape Canaveral	SpaceX CRS-7	1,952 kg	LEO	NASA (CRS)	Failure	F9 v1.1	Precluded	28 June 2015	14:21
20	20	Cape Canaveral	Orbcomm-OG2	2,034 kg	LEO	Orbcomm	Success	F9 FT	Success	22 December 2015	01:29
21	21	VAFB	Jason-3	553 kg	LEO	NASA (LSP) NOAA C...	Success	F9 v1.1	Failure	17 January 2016	18:42
22	22	Cape Canaveral	SES-9	5,271 kg	GTO	SES	Success	F9 FT	Failure	4 March 2016	23:35
23	23	Cape Canaveral	SpaceX CRS-8	3,136 kg	LEO	NASA (CRS)	Success	F9 FT	Success	8 April 2016	20:43
24	24	Cape Canaveral	JCSAT-14	4,696 kg	GTO	SKY Perfect JSAT Gr...	Success	F9 FT	Success	6 May 2016	05:21
25	25	Cape Canaveral	Thaicom 8	3,100 kg	GTO	Thaicom	Success	F9 FT	Success	27 May 2016	21:39
26	26	Cape Canaveral	ABS-2A	3,600 kg	GTO	ABS Eutelsat	Success	F9 FT	Failure	15 June 2016	14:29
27	27	Cape Canaveral	SpaceX CRS-9	2,257 kg	LEO	NASA (CRS)	Success	F9 FT	Success	18 July 2016	04:45
28	28	Cape Canaveral	JCSAT-16	4,600 kg	GTO	SKY Perfect JSAT Gr...	Success	F9 FT	Success	14 August 2016	05:26
29	29	VAFB	Iridium NEXT	9,600 kg	Polar	Iridium Communications	Success	F9 FT	Success	14 January 2017	17:54
30	30	KSC	SpaceX CRS-10	2,490 kg	LEO	NASA (CRS)	Success	F9 FT	Success	19 February 2017	14:39
31	31	KSC	EchoStar 23	5,600 kg	GTO	EchoStar	Success	F9 FT	No attempt	16 March 2017	06:00
32	32	KSC	SES-10	5,300 kg	GTO	SES	Success	F9 FT△	Success	30 March 2017	22:27
33	33	KSC	NROL-76	C	LEO	NRO	Success	F9 FT	Success	1 May 2017	11:15
34	34	KSC	Inmarsat-5 F4	6,070 kg	GTO	Inmarsat	Success	F9 FT	No attempt	15 May 2017	23:21
35	35	KSC	SpaceX CRS-11	2,708 kg	LEO	NASA (CRS)	Success	F9 FT	Success	3 June 2017	21:07
36	36	KSC	BulgariaSat-1	3,669 kg	GTO	Bulsatcom	Success	F9 FTB1029.2	Success	23 June 2017	19:10
37	37	VAFB	Iridium NEXT	9,600 kg	LEO	Iridium Communications	Success	F9 FT	Success	25 June 2017	20:25
38	38	KSC	Intelsat 35e	6,761 kg	GTO	Intelsat	Success	F9 FT	No attempt	5 July 2017	23:38
39	39	KSC	SpaceX CRS-12	3,310 kg	LEO	NASA (CRS)	Success	F9 B4	Success	14 August 2017	16:31
40	40	VAFB	Formosat-5	475 kg	SSO	NSPO	Success	F9 FT	Success	24 August 2017	18:51
41	41	KSC	Boeing X-37B	4,990 kg	LEO	USAF	Success	F9 B4	Success	7 September 2017	14:00
42	42	VAFB	Iridium NEXT	9,600 kg	Polar	Iridium Communications	Success	F9 B4	Success	9 October 2017	12:37
43	43	KSC	SES-11	5,200 kg	GTO	SES S.A. EchoStar	Success	F9 FTB1031.2	Success	11 October 2017	22:53:00
44	44	KSC	Koreasat 5A	3,500 kg	GTO	KT Corporation	Success	F9 B4	Success	30 October 2017	19:34
45	45	Cape Canaveral	SpaceX CRS-13	2,205 kg	LEO	NASA (CRS)	Success	F9 FTB1035.2	Success	15 December 2017	15:36
46	46	VAFB	Iridium NEXT	9,600 kg	Polar	Iridium Communications	Success	F9 FTB1036.2	Controlled	23 December 2017	01:27
47	47	CCAFS	Zuma	C	LEO	Northrop Grumman [f]...	Success	F9 B4	Success	8 January 2018	01:00
48	48	CCAFS	GovSat-1	4,230 kg	GTO	SES	Success	F9 FTB1032.2	Controlled	31 January 2018	21:25
49	49	VAFB	Paz	2,150 kg	SSO	Hisdesat exactEarth ...	Success	F9 FTB1038.2	No attempt	22 February 2018	14:17
50	50	CCAFS	Hispasat 30W-6	6,092 kg	GTO	Hispasat[277] NovaW...	Success	F9 B4	No attempt	6 March 2018	05:33

Note: Scraping does not yield the exact data frame that the API does due to their different sources.

Data Wrangling

- Replaced null payload masses with average payload mass
- Calculated number of launches from each site, occurrence of orbit types, occurrence of mission outcome for each orbit
  - Identified which outcomes are classified as failures.
  - Used this information to create 'Class' column
    - (Success = 1, Fail = 0)



https://github.com/jaredplatt/Data-Science-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

EDA with Data Visualization

- Import CSV from Data Wrangling and convert to data frame.
- Cat plot of flight payload mass vs. flight number
  - Identify how the success rate changes with payload mass over time.
- Cat plot of launch site vs. flight number
  - Identify the success rates for each launch site over time
- Scatter plot of payload mass vs. launch site.
  - Identify success rates at each launch pad and how payload mass contributes to each

https://github.com/jaredplatt/Data-Science-Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with Data Visualization Continued

- Bar plot of success rate vs. orbit type
  - Identify which orbits have the highest success rates
- Scatter plot of orbit type vs. flight number
  - Identify which orbits were successful early and how successful additional orbits were over time
- Scatter plot of orbit type vs. payload mass
  - Identify which orbits can be successfully achieved based on payload mass.
- Line plot of success rate vs. year
  - Identify how the success rate has changed with more launches over time

https://github.com/jaredplatt/Data-Science-Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb

EDA with SQL

- Load SQL extension and establish connection via cursor with CSV from EDA with Data Visualization.
    - Query to create table from CSV
- 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.

https://github.com/jaredplatt/Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

EDA with SQL

- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 kg but less than 6000 kg
- 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 landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.

https://github.com/jaredplatt/Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

- Mark all launch sites on a map with circle objects and pop-up labels
  - To identify the launch sites proximity to the Equator line and coastline on an interactive map. Determine if there are similarities between the locations. . This would be useful for positioning a SpaceY launch site.
- Mark the success/failed launches for each site on the map with marker cluster object
  - To identify which sites have high success rates on an interactive map. This would be useful for positioning a SpaceY launch site.
- Calculate the distances between each launch site and various points of interest (cities, highways, railroads)
  - Add to the map using markers for the POIs and polyline to represent the distances.
  - The distances were calculated using a user defined function and were used as labels for the markers.
  - Can identify if all sites share nearby points of interest. This would be useful for positioning a SpaceY launch site.

https://github.com/jaredplatt/Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.jupyterlite.ipynb

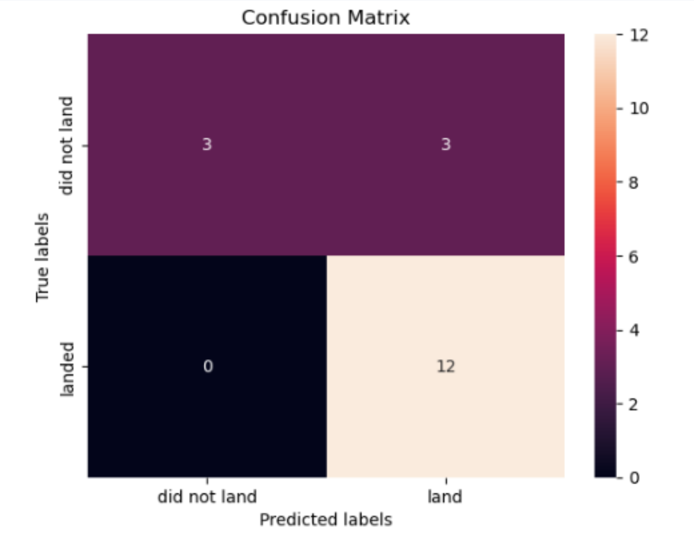
Build a Dashboard with Plotly Dash

- Title – SpaceX Launch Records Dashboard
- Dropdown – Select Launch Sites
    - Options - All Sites, CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, CCAFS SLC-40
        - Allows user to sort data on plots by launch site to identify insights about each
- Graph – Success Pie Chart
    - Uses a callback function to create success rate pie chart depending on site chosen
        - All Sites – Presents the percentage of all successful launches at each site
        - Individual site – Presents the percentage of successful and failed launches at a given site
        - Allows user to visualize which sites have the highest success rates and which are used the most
- Range Slider – Payload Mass
    - Allows user to filter data such that a range of payload masses can be analyzed.
- Graph – Success Payload Scatter Plot
    - Uses a callback function to create scatter plot of launch outcome vs. payload mass depending on launch site and selected payload range
    - Allows user to analyze the impact of payload mass on launch outcome from all sites or each individual site.

https://github.com/jaredplatt/Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.jupyterlite.ipynb

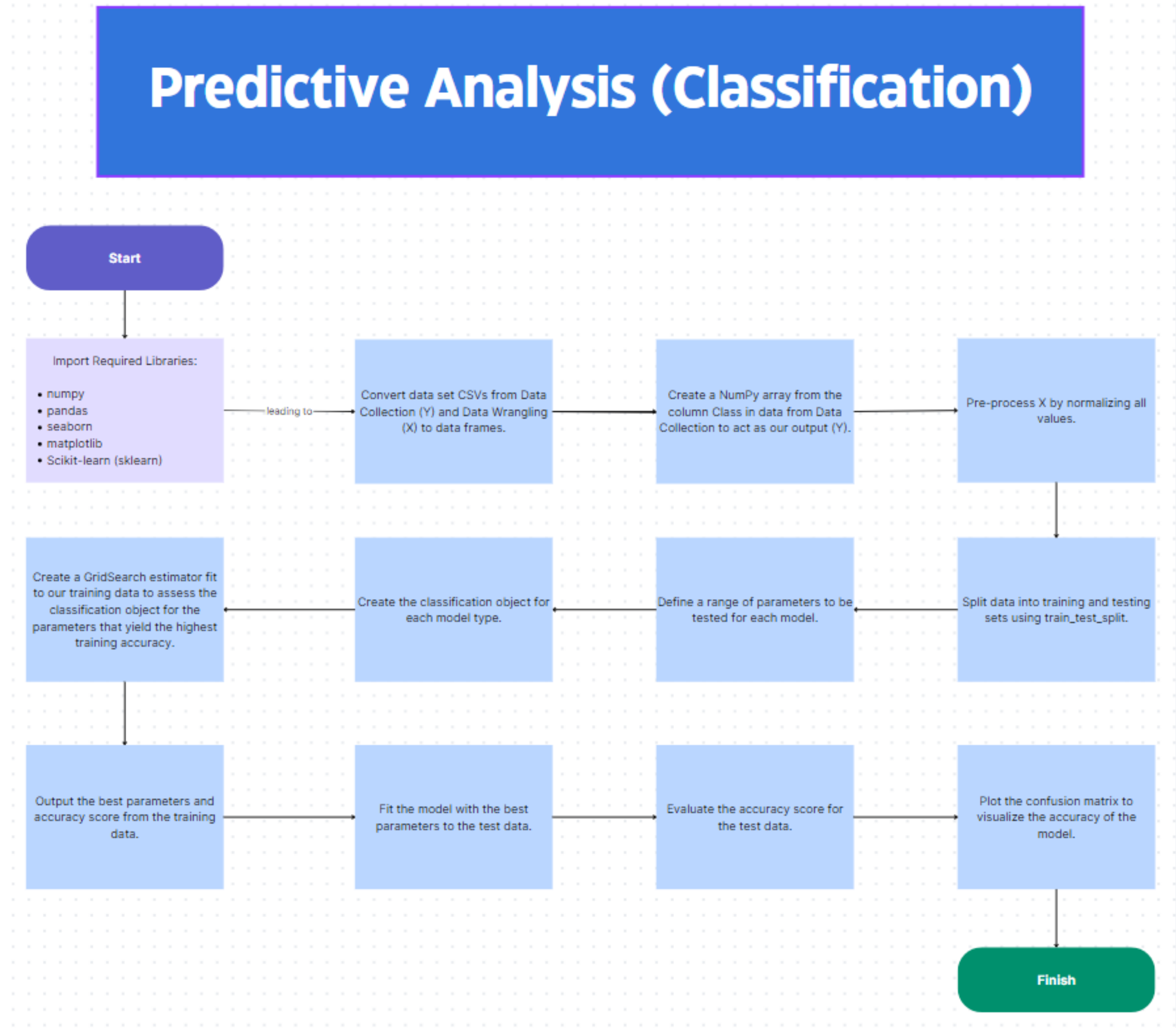
Predictive Analysis (Classification)

- Built 4 classification models to predict the outcome (class) of a launch. Optimized the parameters for each using GridSearch
  - Logistic Regression
  - Support Vector Machine (SVM)
  - Decision Tree Classifier
  - K Nearest Neighbor
- Imported CSVs from Data Collection and Data Wrangling and converted to pandas data frames
  - The output we are trying to optimize for is 'Class' (Y)
- Split data into training and testing sets using train_test_split
- Created a classification object corresponding to each of the models
- Fit the model with training data set
- Use GridSearch method to estimate the best parameters for each model
- Use best parameters to fit model to testing data and create confusion matrix to visually inspect each
- Compare the accuracy scores and confusions matrices for each model to determine which fits the test data best



https://github.com/jaredplatt/Data-Science-Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Predictive Analysis (Classification) Flow Chart



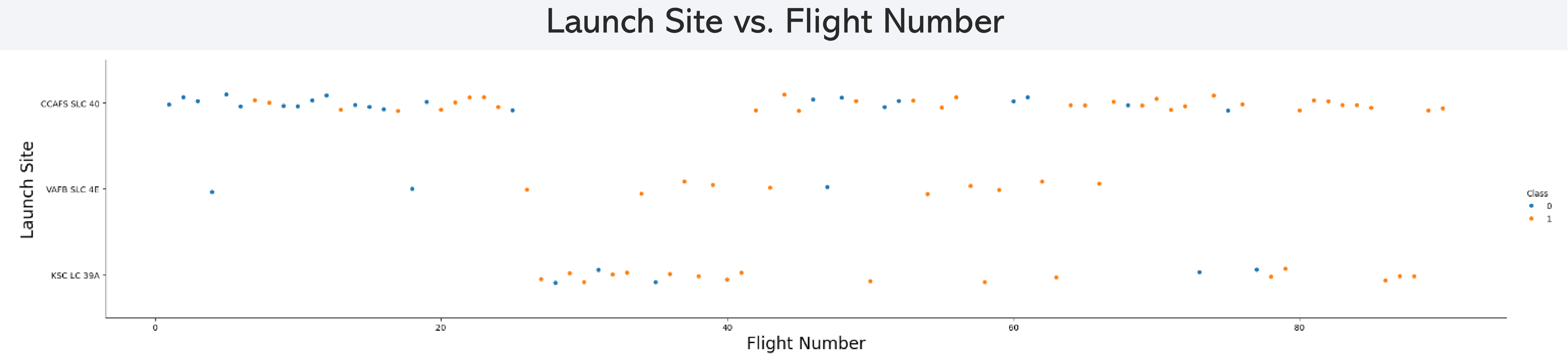
Results

- Insights drawn from exploratory data analysis (EDA)
  - Visualization
  - SQL
- Interactive analytics demo
  - Launch site proximities analysis - Folium
  - Launch records dashboard – Plotly Dash
- Predictive analysis results
  - Best classification model for predicting launch outcomes - Scikit-Learn

Section 2

Insights drawn from EDA

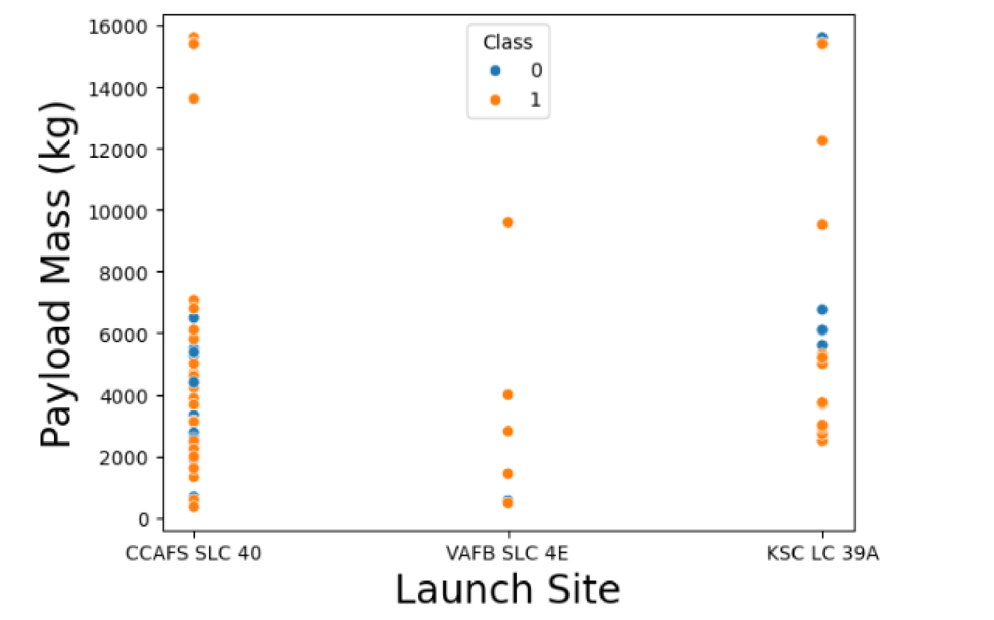
Launch Site vs. Flight Number



- Most of the first 25 launches occurred at CCAFS SLC 40. It is the most used.
- VAFB SLC 4E is the least used launch site.
- CCAFS SLC 40 has a success rate of 60%.
- KSC LC 39A and VAFB SLC 4E have matching success rates of 77%.

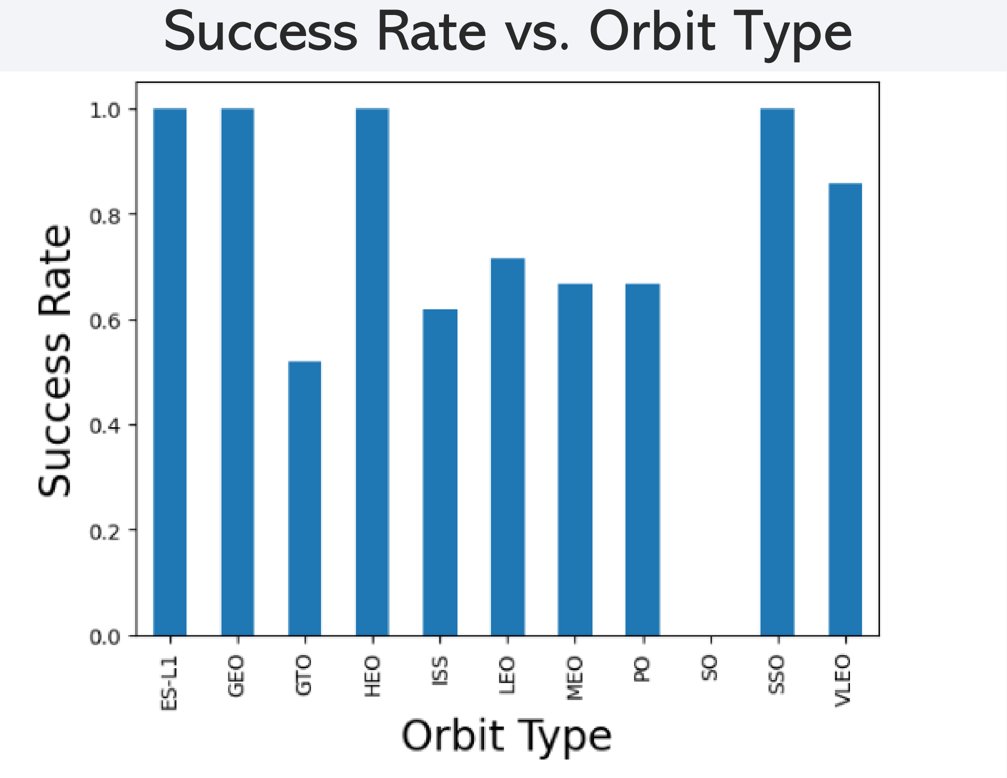
Payload Mass (kg) vs. Launch Site

Payload Mass (kg) vs. Launch Site



- VAFB SLC 4E cannot support payloads over 10,000 kg.
- CCAFS SLC 40 is only used for payloads greater than 13,000 kg and less than 8000 kg.
- KSC LC 39A is only used for payloads greater than 2000 kg.

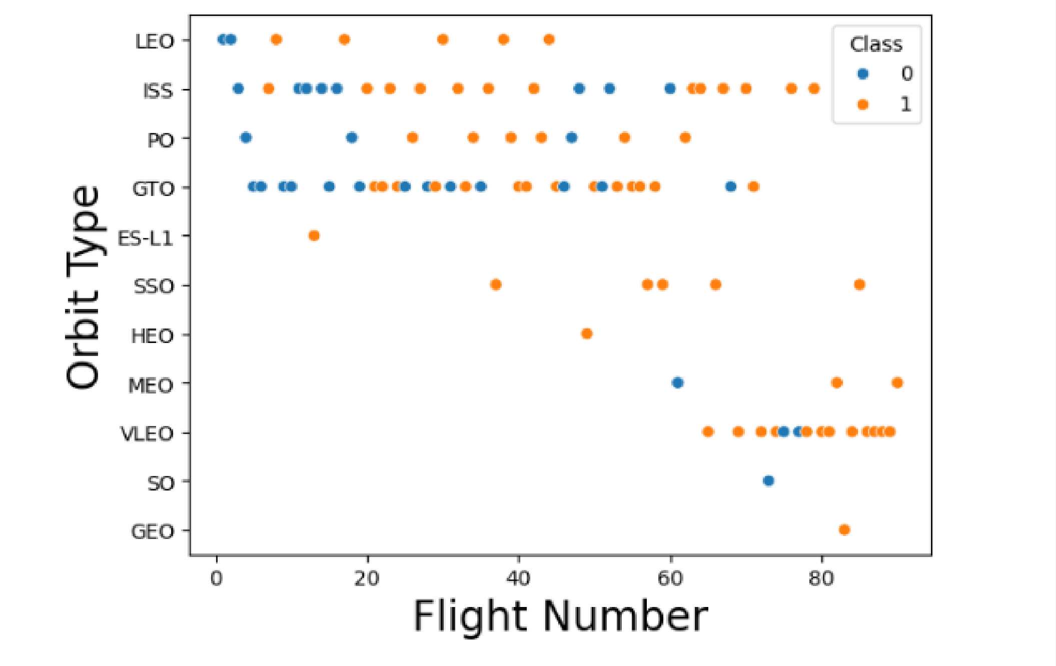
Success Rate vs. Orbit Type



- ES-L1, GEO, HEO, and SSO all have 100% success rates.
- GTO (~50%) and SO (0%) are the only orbits with success rates lower than 60%

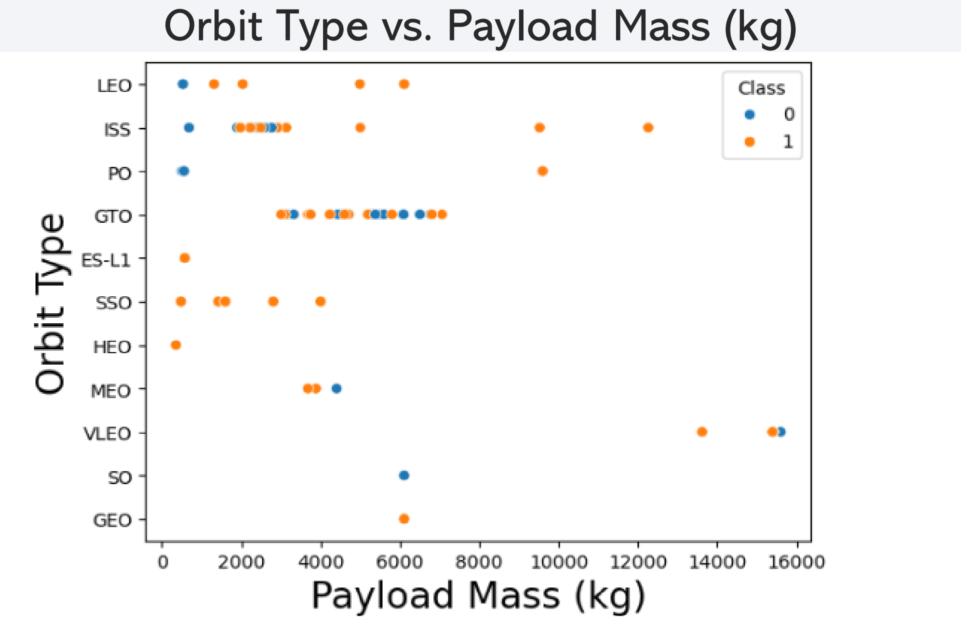
Orbit Type vs. Flight Number

Orbit Type vs. Flight Number



- LEO orbit success rate increased to 100% after first 2 LEO launches.
- ISS, PO, and GTO success rates are higher for flight numbers greater than 40.
- VLEO has a 100% success rate after the first 7 VLEO launches.

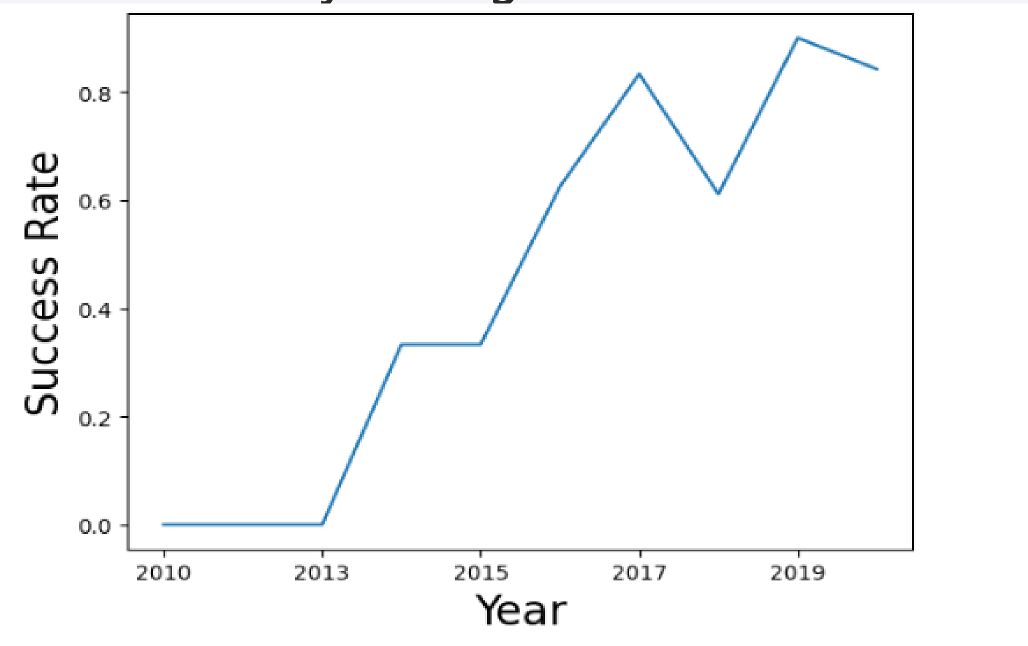
Orbit Type vs. Payload Mass (kg)



- 50% success rate for all launches with payloads of 1000 kg or less.
- GTO, ES-L1, SSO, HEO, MEO, SO, and GEO all have 0 launches with payloads greater than 8000 kg.
- VLEO has only launched with payloads greater than 13000 kg
- Smaller payloads (<8000 kg) are launched significantly more often

Launch Success Yearly Trend

Yearly Average Success Rate



- There were 0 successful launches before 2014
- Success rate sharply increased from 2013 – 2017 (0% - >80%)
- Success rate dipped in 2018 (~60%), increased in 2019 (>90%), then dipped again in 2020 (>80%)

All Launch Site Names

- Query:

```
%%sql
select DISTINCT Launch_Site
from SPACEXTBL
```

- Result:

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

- 4 Unique launch sites displayed with DISTINCT, selecting from the Launch Site column in our SpaceX Launch Data SQL table.

Launch Site Names Begin with 'CCA'

- Query:

```
%%sql
select *
from SPACEXTBL
where Launch_Site like '%CCA%' Limit 5
```

- Result:

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

- 5 launch records displayed using WHERE to select only 5 (LIMIT) launch sites beginning with ‘CCA’ (LIKE) from our SpaceX Launch Data SQL table

Total Payload Mass

- Query:

```
%%sql
select SUM(PAYLOAD_MASS__KG_) as 'Total Payload Mass Carried by Boosters Launched by NASA (CRS)'
from SPACEXTBL
where Customer like '%NASA (CRS)%'
```

- Result:

Total Payload Mass Carried by Boosters Launched by NASA (CRS)
48213

- Total payload mass from all NASA CRS launches displayed using SUM on the Payload Mass (kg) column and WHERE and LIKE on the Customer column from our SpaceX Launch Data SQL table. Used AS for a more apt column name.

Average Payload Mass by F9 v1.1

- Query:

```
%%sql
select AVG(PAYLOAD_MASS__KG_) as 'Average Payload Mass (Booster Version F9 v1.1)'
from SPACEXTBL
where Booster_Version like 'F9 v1.1'
```

- Result:

Average Payload Mass (Booster Version F9 v1.1)
2928.4

- Average payload mass from all F9 v1.1 Booster launches displayed using AVG on the Payload Mass (kg) column and WHERE and LIKE on the Booster Version column from our SpaceX Launch Data SQL table. Used AS for a more apt column name.

First Successful Ground Landing Date

- Query:

```
%%sql
select MIN(Date) as 'First Succesful Landing Date (Ground Pad)'
from SPACEXTBL
where Landing_Outcome like '%Success (ground pad)%'
```

- Result:

First Succesful Landing Date (Ground Pad)
2015-12-22

- Date of first successful landing on a ground pad displayed using MIN on the Date column and WHERE and LIKE on the Landing Outcome column from our SpaceX Launch Data SQL table. Used AS for a more apt column name.

Successful Drone Ship Landing with Payload between 4000 and 6000

- Query:

```
%%sql
select distinct Booster_Version as 'Boosters Which Have Success in Drone Ship (4000 kg < Payload Mass < 6000 kg)'
from SPACEXTBL
where Landing_Outcome like '%Success (drone ship)%' and PAYLOAD_MASS__KG_ BETWEEN 4000 and 6000
```

- Result:

Boosters Which Have Success in Drone Ship (4000 kg < Payload Mass < 6000 kg)
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

- 4 boosters that have successfully landed on a drone ship with a payload mass greater than 4000 kg and less than 6000 kg displayed using DISTINCT on the Booster Version column and WHERE, LIKE, AND, BETWEEN on the Landing Outcome column from our SpaceX Launch Data SQL table. Used AS for a more apt column name.

Total Number of Successful and Failure Mission Outcomes

- Query:

```
%%sql
select Mission_Outcome as 'Mission Outcome', count(Mission_Outcome) as 'Count'
from SPACEXTBL
GROUP BY MISSION_OUTCOME
ORDER BY count(Mission_Outcome) DESC;
```

- Result:

Mission Outcome	Count
Success	98
Success (payload status unclear)	1
Success	1
Failure (in flight)	1

- Total number of successful and failure mission outcomes displayed using COUNT on the Mission Outcome column and GROUP BY, ORDER BY, DESC on the Mission Outcome column and COUNT of Mission Outcomes from our SpaceX Launch Data SQL table. Used AS for more apt column names.

Boosters Carried Maximum Payload

- Query:

```
%%sql
select distinct Booster_Version as 'Booster Versions Which Have Carried the Maximum Payload Mass'
from SPACEXTBL
where PAYLOAD_MASS__KG_ = (select MAX(PAYLOAD_MASS__KG_) from SPACEXTBL)
```

- Result:

Booster Versions Which Have Carried the Maximum Payload Mass
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

- 12 unique F9 B5 booster versions which have carried the maximum payload displayed using DISTINCT on the Booster Version column and WHERE, =, and a sub-query with MAX on the Payload Mass column from our SpaceX Launch Data SQL table. Used AS for a more apt column name.

2015 Launch Records

- Query:

```
%%sql
select substr(Date, 6,2) as 'Month', Landing_Outcome as 'Landing Outcome', Booster_Version as 'Booster Version', Launch_Site as 'Launch Site'
from SPACEXTBL
where substr(Date,0,5)='2015' and  Landing_Outcome like '%Failure (drone ship)%'
```

- Result:

Month	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

- Listed the records which will display the month names, failure landing outcomes in drone ship , booster versions, and launch site for the months in year 2015. Displayed using substr() to replace Month names with Month numbers as well as WHERE, AND, =, LIKE on the Date and Landing Outcome columns from our SpaceX Launch Data SQL table. Used AS for more apt column names.

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

- Query:

```
%%sql
select Landing_Outcome as 'Landing Outcome (2010-06-04 through 2017-03-20)' , count(Landing_Outcome) as 'Count'
from SPACEXTBL
where Date between '2010-06-04' and '2017-03-20'
Group By Landing_Outcome
Order by count(Landing_Outcome) DESC
```

- Result:

Landing Outcome (2010-06-04 through 2017-03-20)	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

- Ranked the count of landing outcomes between the date 2010-06-04 and 2017-03-20, in descending order. Displayed using COUNT on Landing Outcome column; WHERE, AND on the Date column; GROUP BY, ORDER BY, COUNT, DESC on the Landing Outcome column from our SpaceX Launch Data SQL table. Used AS for more apt column names.

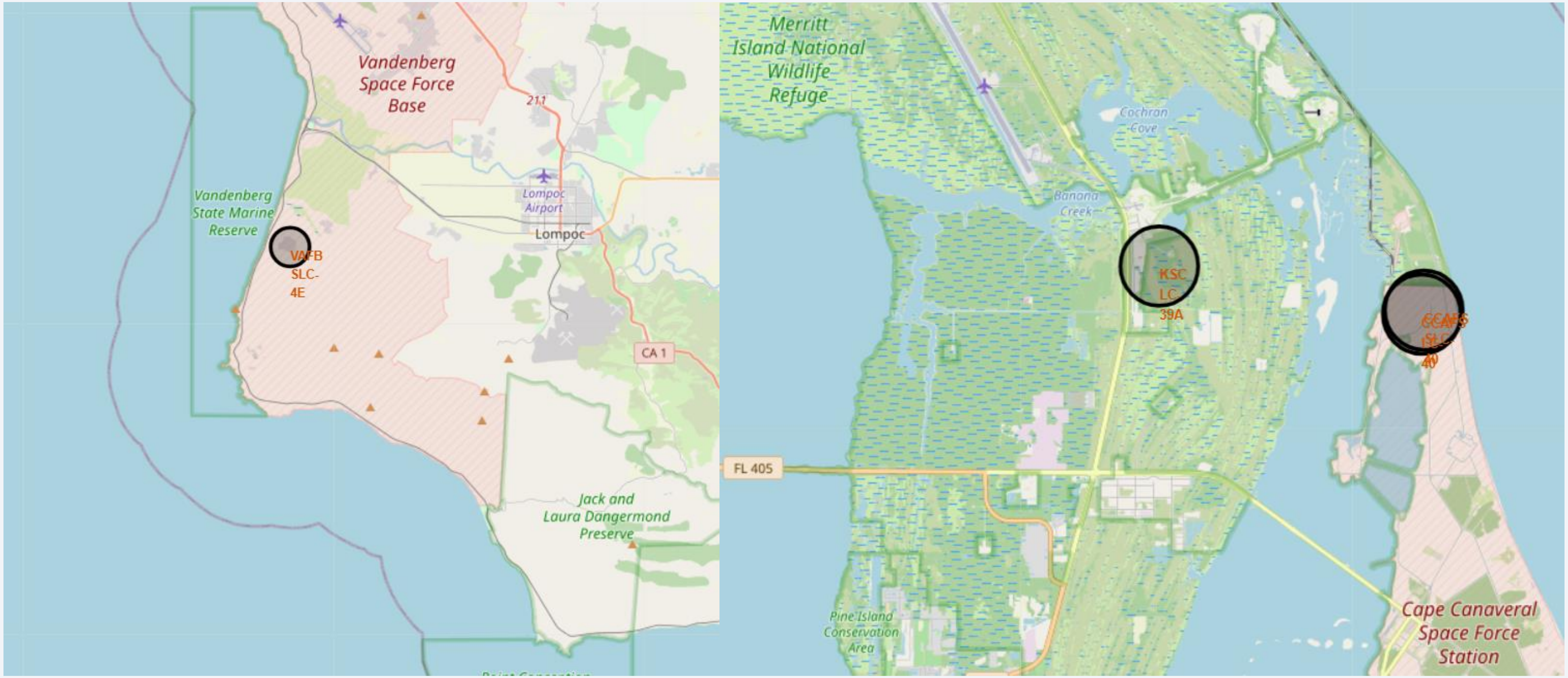
Section 3

Launch Sites Proximities Analysis

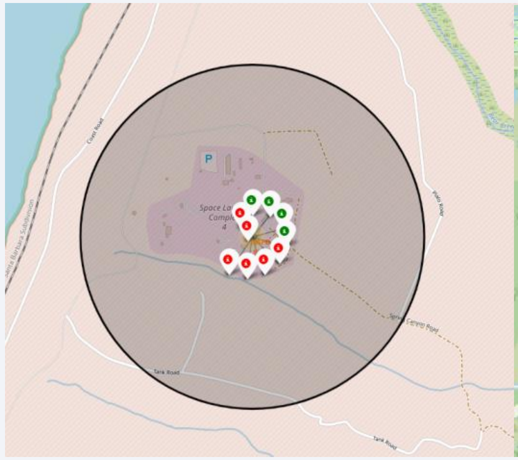
All United States SpaceX Launch Sites

- All SpaceX launches share:
  - Close to equator line – Allows rocket to maximize the use of Earth's rotation
  - Close to coastline – Prevents failed launches crashing into populated areas

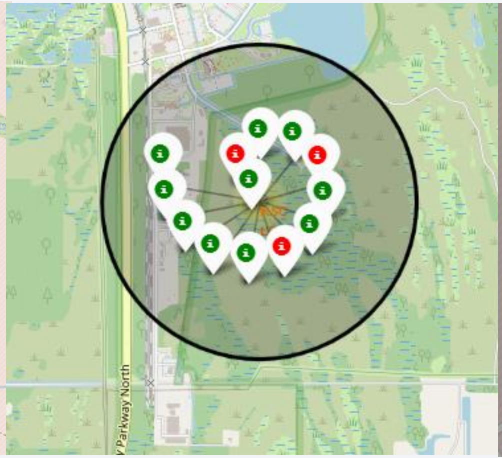
All United States SpaceX Launch Sites (Zoomed)



Launch Outcomes at Each Launch Site

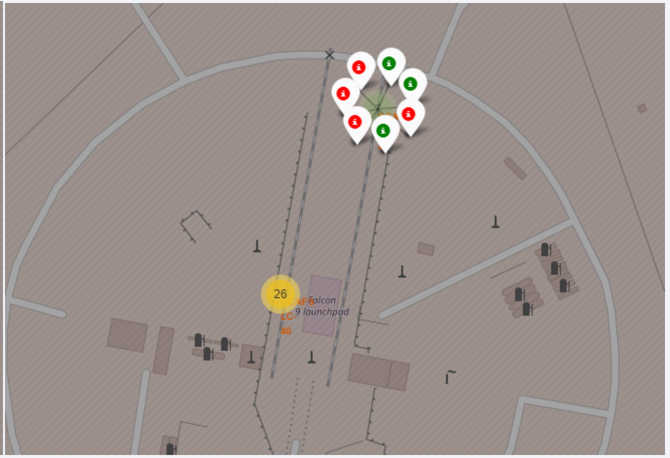


VAFB SLC-4E



KSC LC-39A

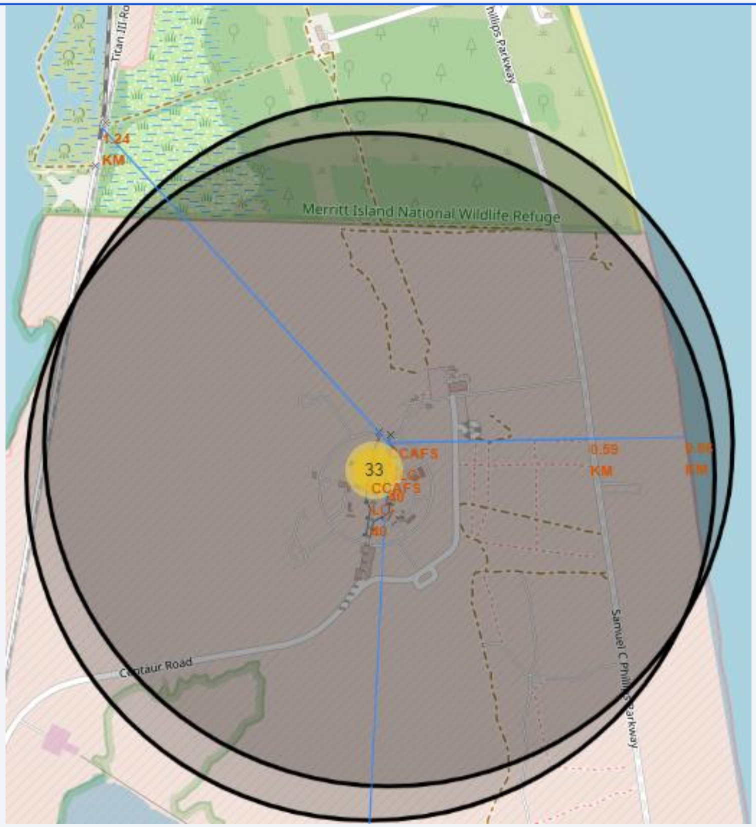
CCAFS LC-40

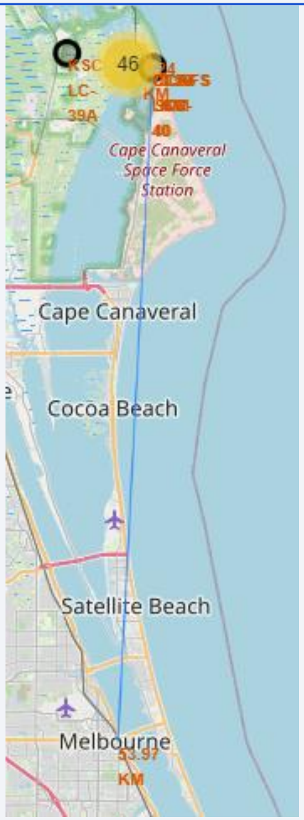


CCAFS SLC-40

- VAFB SLC-4E (West Coast) has a success rate of 40%
- KSC LC-39A (East Coast) has a success rate of ~77%
- CCAFS LC-40 (East Coast) has a success rate of ~27%
- CCAFS SLC-40 (East Coast) has a success rate of ~43%

Nearest Points of Interest (Highways, Railways, Coasts, Cities) to CCAFS SLC-40



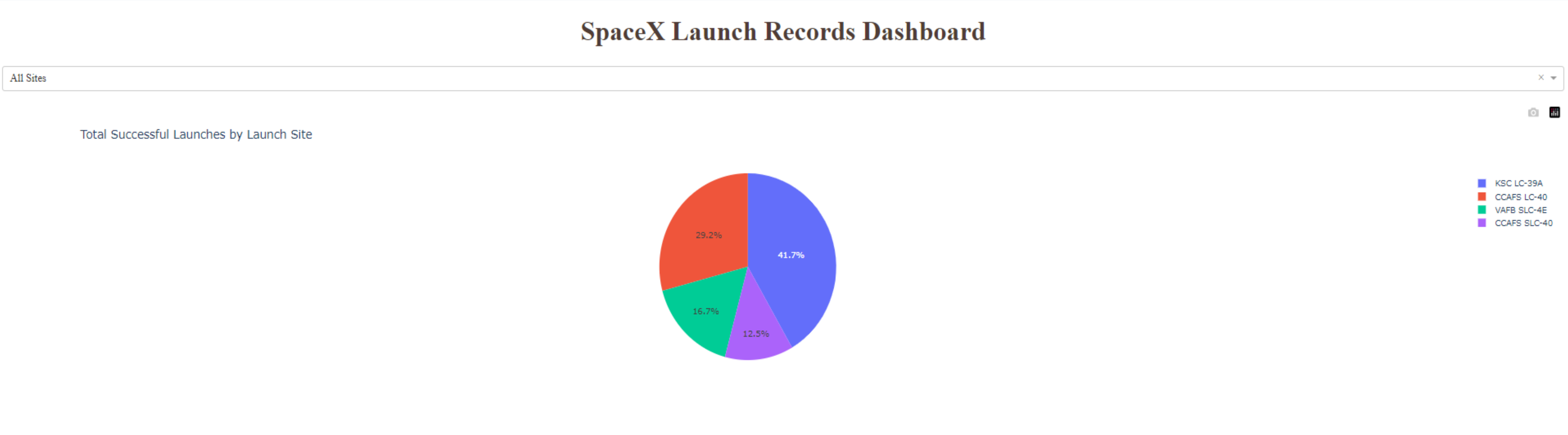


- There is a railroad, highway, and coastline all within 1.25 km of the launch site. Great for material shipments and employees to get to site.
- The nearest city is nearly 54 km away. Great for preventing mass casualty and property damage from failed launches.

Section 4

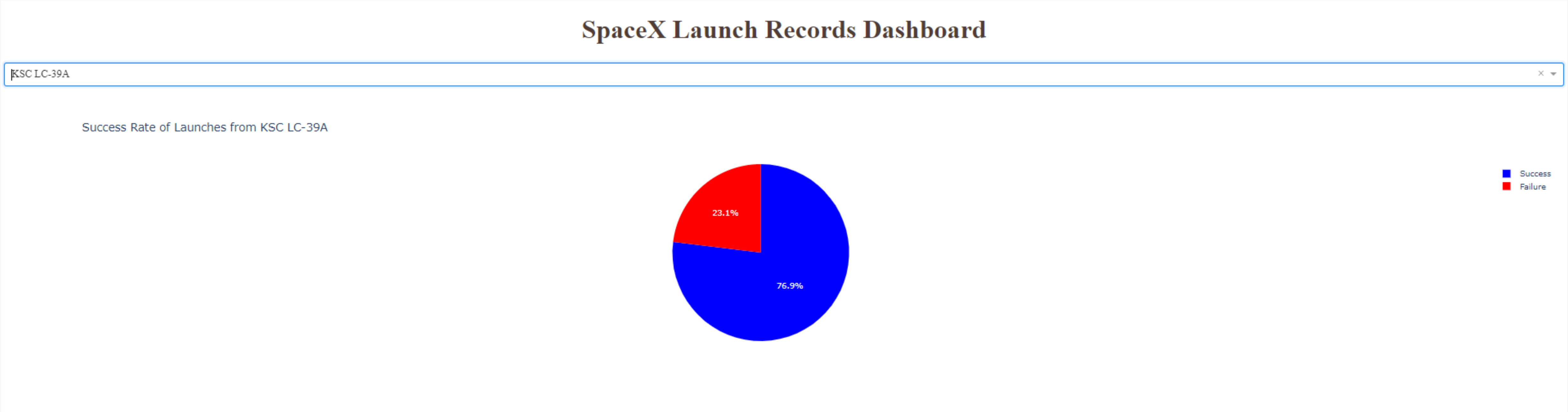
Build a Dashboard with Plotly Dash

Pie Chart of Successful Launch Count by Launch Site



- KSC LC-39A is the most used launch site for successful launches (41.7%)
- CCAFS SLC-40 is the least used launch site for successful launches (12.5%)

Success Rate for Launch Site with Highest Launch Success Ratio

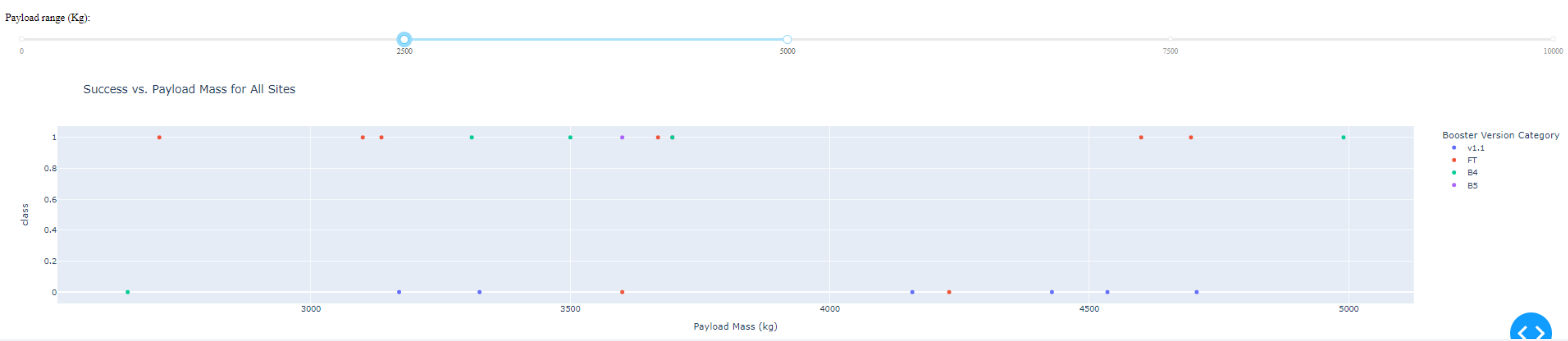


- KSC LC-39A has the highest success rate (76.9%)
- KSC LC-39A was also the most used launch pad for successful launches.
- Based on the SpaceX launch data, SpaceY would have the highest likelihood of success at a launch site similar to KSC LC-39A

Payload Mass (kg) vs. Launch Outcome for Various Payload Ranges

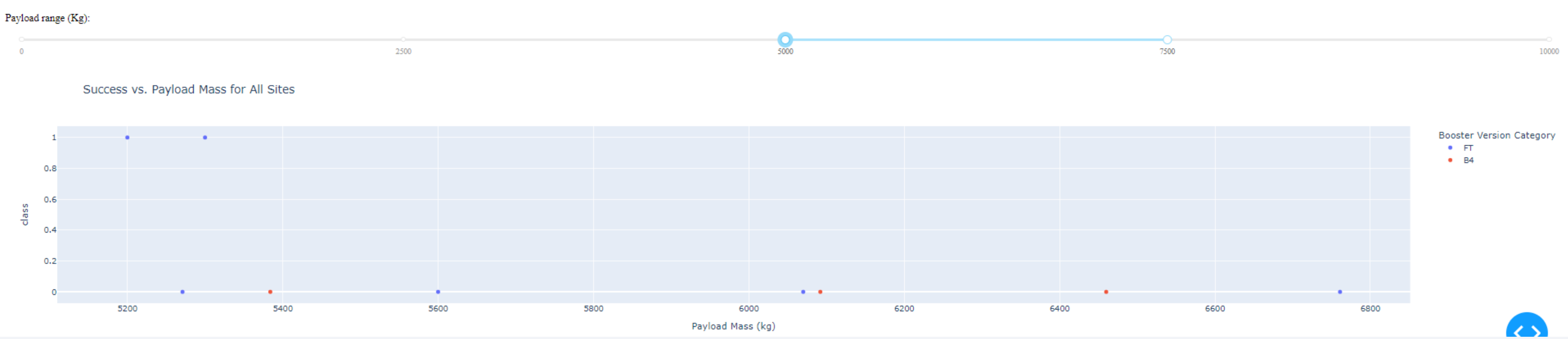
- 0 kg – 2500 kg (7/19 or ~37%), (FT Booster most successful 5/7 or ~71%)

- 2500 kg – 5000 kg (11/20 or 55%), (FT Booster most successful 6/11 or ~55%)

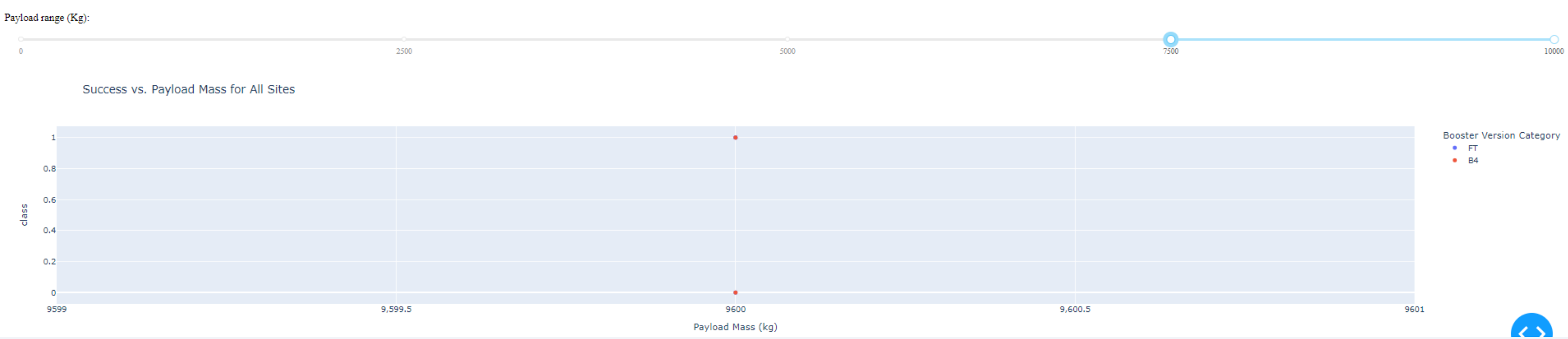


Payload Mass (kg) vs. Launch Outcome for Various Payload Ranges

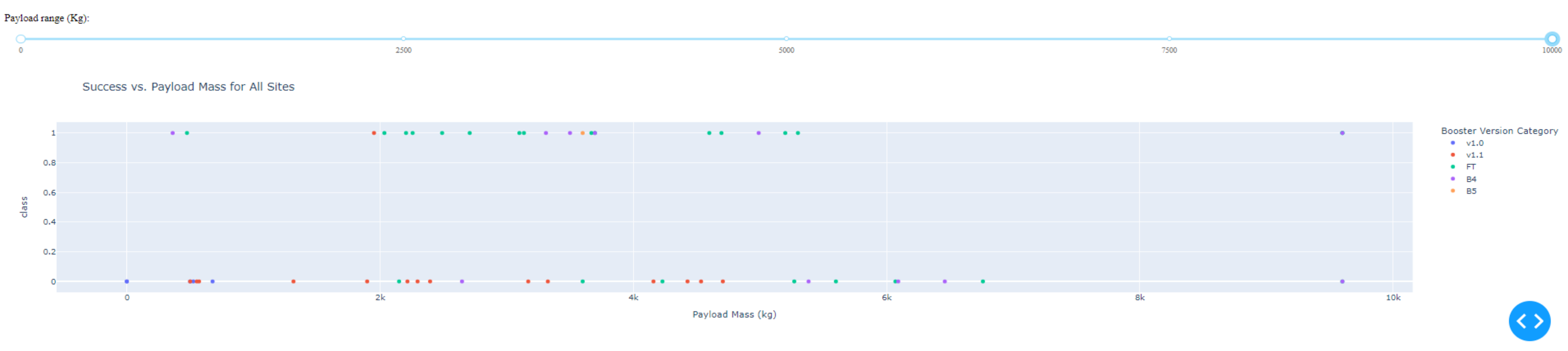
- 5000 kg – 7500 kg (2/11 or ~18%), (FT Booster Most Successful 2/2 or 100%)



- 7500 kg – 10000 kg (1/2 or 50%), (B4 only booster used)



Payload Mass (kg) vs. Launch Outcome for Various Payload Ranges



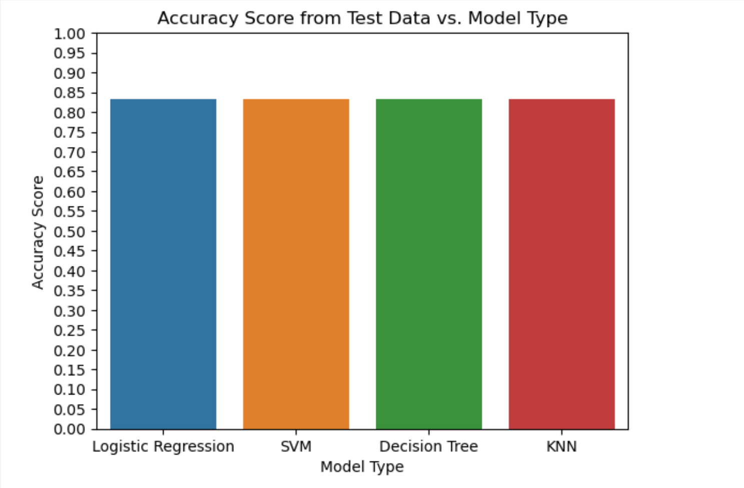
- Insights:
  - No launches between 7000 kg and 9500 kg. No launches greater than 9600 kg
  - Launches with payloads between 2500 kg – 5000 kg are the most common (20 launches) and most successful (55%)
  - Launches with payloads between 5000 kg – 7500 kg are the 2nd least common (11 launches) and overall least successful (18%)
  - FT Booster most successful for payloads between 0 kg – 7500 kg
  - B4 booster is the only booster with launches over 7000 kg.

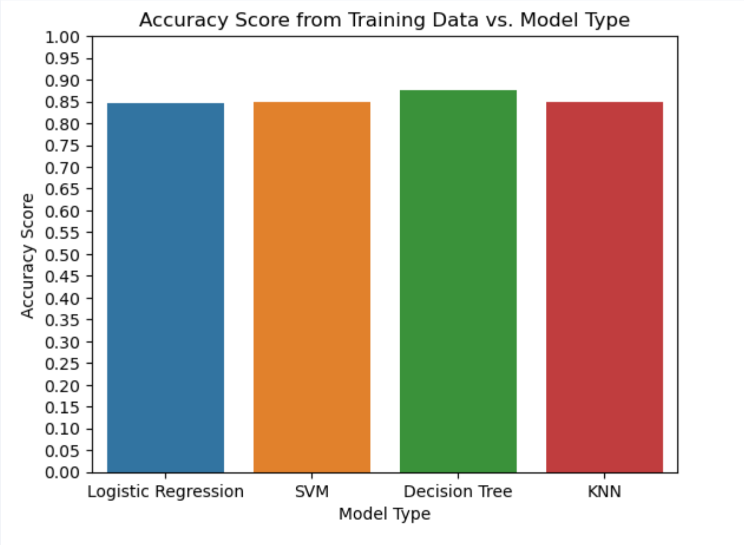
Section 5

Predictive Analysis (Classification)

Classification Accuracy

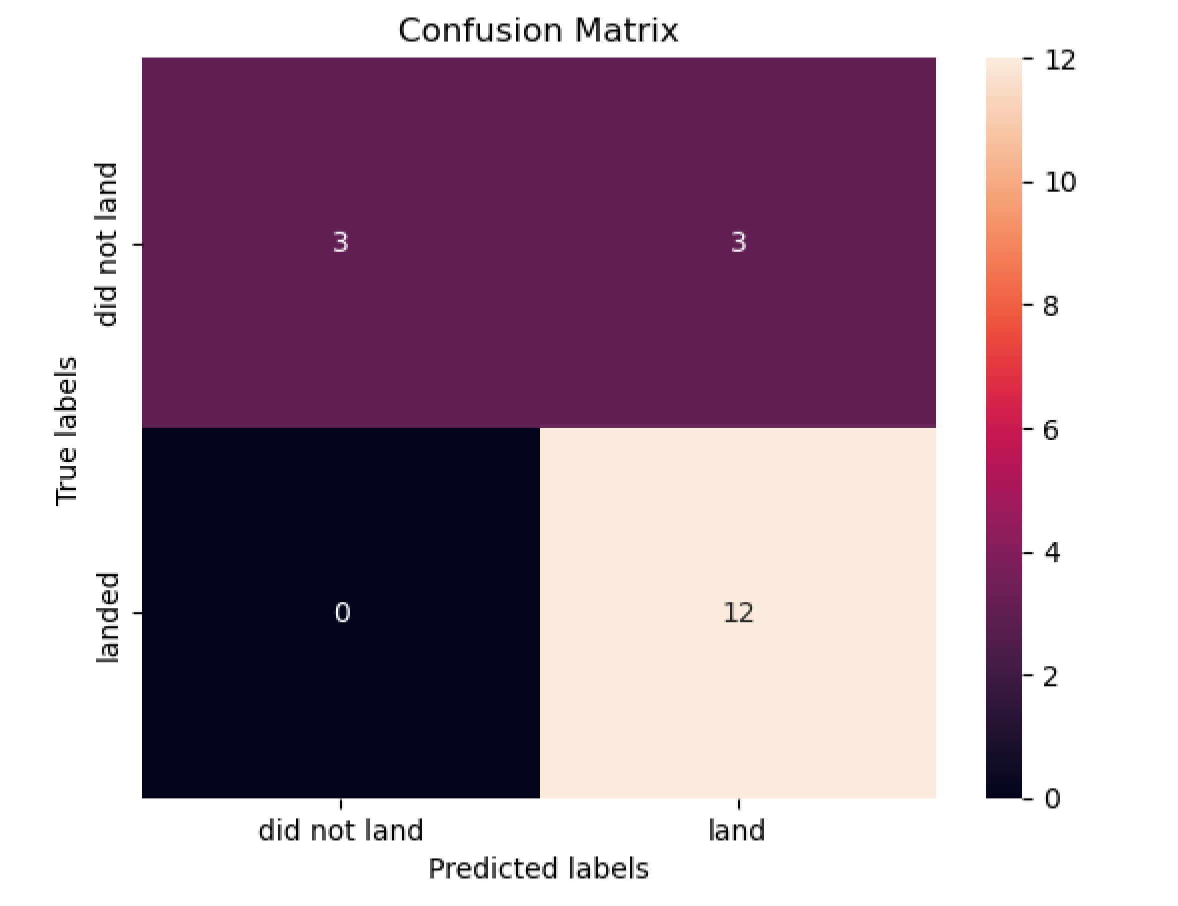
- Each model yields the same accuracy score on testing data (83.3%)
- Decision Tree yields a slightly higher accuracy score on training data (87.3%) than the other models.
- While all models are valid for predicting launch outcomes, Decision Tree Classifier is our best model.





Confusion Matrix (Decision Tree)

- Confusion Matrix from Decision Tree model
- No false negatives for successful landings
- 3 false positives for failed landing/did not land



Conclusions

- KSC LC-39A has the highest success rate for launches. A launch site like this gives the best hope for SpaceY success.
    - Needs to be close to coast, equator, railways, and highways while maintaining a large distance from a city or populated area.
- ES-L1, GEO, HEO, and SSO orbits have 100% success rates
    - Prioritize these for SpaceY whenever possible.
    - LEO is also a good option due to volume and relatively high success rate (only 2 total failures)
- Payloads less than 8000 kg are launched most often.
    - SpaceY should prioritize payloads of 2500 kg – 5000 kg as they have the highest success rate (55%)
- Falcon 9 booster has a success rate of over 98%
    - SpaceY should try to replicate this design for most launches under 5000 kg
- Ground pad landings had zero failures. Prioritize this type of landing over drone ship.
    - May need to place launch site further from city than SpaceX does
- Confidence interval: 87.3%

Appendix

- Data Collection
  - https://github.com/jaredplatt/Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb
  - https://github.com/jaredplatt/Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb
- Data Wrangling
  - https://github.com/jaredplatt/Data-Science-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb
- EDA with Visualization
  - https://github.com/jaredplatt/Data-Science-Capstone/blob/main/jupyter-labs-eda-dataviz.ipynb.jupyterlite.ipynb
- EDA with SQL
  - https://github.com/jaredplatt/Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb
- Interactive Visual Analytics with Folium
  - https://github.com/jaredplatt/Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.jupyterlite.ipynb
- Interactive Dashboard with Plotly Dash
  - https://github.com/jaredplatt/Data-Science-Capstone/blob/main/spacex_dash_app.py
- Machine Learning Prediction
  - https://github.com/jaredplatt/Data-Science-Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Thank you!