IBM Developer
SKILLS NETWORK

SpaceY

Winning Space Race with Data Science

Paula Reitan
Data Scientist for SpaceY
September 2021

Outline

Executive Summary

- [Summary of methodologies]
- Summary of all results

Introduction - First Stage Rocket ReUse Prediction

- Project background and context
  - SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.
- Problems you want to find answers
  - If SpaceY can determine if the first stage of Falcon 9 will land, we can determine the cost of a launch. We can use this information to make a data-driven decision when we should bid against SpaceX for a rocket launch.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Gathered SpaceX launch data from the SpaceX REST API and scraped data from Falcon 9 Wiki Pages
- Perform data wrangling
  - Cleaned data, replaced PayloadMass NAN => PayloadMass.mean
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Used machine learning (ML) to determine if the first stage of Falcon 9 will land successfully. Split data into training data and test data to find the best Hyperparameter for K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Trees (DT), and Logistic Regression (LR). Determined method that performed best using test data.

Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

Data Collection – SpaceX API

- https://github.com/sweetps/datascience-capstone/blob/main/Applied%20Data%20Science%20Capstone%20-%20Data%20Collection.ipynb

https://api.spacexdata.com/v4/**rockets/**

https://api.spacexdata.com/v4/**payloads/**

https://api.spacexdata.com/v4/**lauchpads/**

https://api.spacexdata.com/v4/**cores/**

https://api.spacexdata.com/v4/**past**
/

- Filter on Falcon 9
- Remove irrelevant columns
- Replace NAN PayloadMass with Mean Payload Mass

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
4	1	2010-06-04	Falcon 9	6123.547647	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857
5	2	2012-05-22	Falcon 9	525.000000	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005	-80.577366	28.561857
6	3	2013-03-01	Falcon 9	677.000000	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007	-80.577366	28.561857
7	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003	-120.610829	34.632093
8	5	2013-12-03	Falcon 9	3170.000000	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004	-80.577366	28.561857
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
89	86	2020-09-03	Falcon 9	15600.000000	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	7	B1060	-80.603956	28.608058
90	87	2020-10-06	Falcon 9	15600.000000	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	7	B1058	-80.603956	28.608058
91	88	2020-10-18	Falcon 9	15600.000000	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	9	B1051	-80.603956	28.608058
92	89	2020-10-24	Falcon 9	15600.000000	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e3033383ecbb9e534e7cc	5.0	7	B1060	-80.577366	28.561857
93	90	2020-11-05	Falcon 9	3681.000000	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb234e7ca	5.0	2	B1062	-80.577366	28.561857

Data Collection – Web Scraping Falcon 9 Wiki

- https://github.com/sweetps/datascience-capstone/blob/master/Applied%20Data%20Science%20Capstone%20-%20Web%20Scraping.ipynb

TASK 1: Request the Falcon9 Launch Wiki page from its URL

TASK 2: Extract all column/variable names from the HTML table header

TASK 3: Create a data frame by parsing the launch HTML tables

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10

Data Wrangling

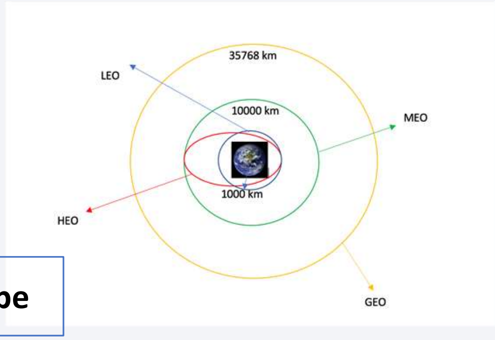
- https://github.com/sweetps/datascience-capstone/blob/master/Applied%20Data%20Science%20Capstone%20-%20Data%20Wrangling.ipynb

TASK 1: Calculate the number of launches on each site

TASK 2: Calculate the number and occurrence of each orbit

TASK 3: Calculate the number and occurrence of mission outcome per orbit type

TASK 4: Create a landing outcome label from Outcome column



	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	False
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	False
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	False
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	False
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
85	86	2020-09-03	Falcon 9	15400.000000	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	2	B1060	-80.603956	28.608058	True
86	87	2020-10-06	Falcon 9	15400.000000	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	2	B1058	-80.603956	28.608058	True
87	88	2020-10-18	Falcon 9	15400.000000	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	5	B1051	-80.603956	28.608058	True
88	89	2020-10-24	Falcon 9	15400.000000	VLEO	CCAFS SLC 40	True ASDS	3	True	True	True	5e9e3033383ecbb9e534e7cc	5.0	2	B1060	-80.577366	28.561857	True
89	90	2020-11-05	Falcon 9	3681.000000	MEO	CCAFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb234e7ca	5.0	0	B1062	-80.577366	28.561857	True

EDA with Data Visualization

- https://github.com/sweetps/datascience-capstone/blob/master/Applied%20Data%20Science%20Capstone%20-%20EDA%20with%20Viz.ipynb

- EDA – Understanding Data through Visualization
  - Scatter Plot FlightNumber vs. PayloadMass and overlay the outcome of the launch
  - Scatter Plot FlightNumber vs. Launchsite and overlay the outcome of the launch
  - Scatter Plot PayloadMass vs. Launchsite and overlay the outcome of the launch
  - Bar Chart Orbit vs Success
  - Scatter Plot FlightNumber vs. Orbit and overlay the outcome of the launch
  - Scatter Plot PayloadMass vs. Orbit and overlay the outcome of the launch
  - Line Chart Mean Success by Year
- Feature Engineering
  - Selected ['Orbit', 'LaunchSite', 'LandingPad', 'Serial'] with one hot-encoding

EDA with SQL

- https://github.com/sweetps/datascience-capstone/blob/master/Applied%20Data%20Science%20Capstone%20-%20EDA%20with%20SQL.ipynb

- Loaded Spacex DataSet (.csv) into DB2 database
- Ran SQL Queries
  - ***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 acheived.***
  - ***List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000***
  - ***List the total number of successful and failure mission outcomes***
  - ***List the names of the booster versions which have carried the maximum payload mass using subquery***
  - ***List the failed landing_outcomes in drone ship, their booster versions, and launch site names for 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***

Build an Interactive Map with Folium

- https://github.com/sweetps/datascience-capstone/blob/master/Applied%20Data%20Science%20Capstone%20-%20Folium.ipynb

- Created Folium maps with following objects
  - Marker Cluster– used to mark launches – red failure, green success
  - Circles – used to mark launch sites
  - Lines – used to show distances between launch sites and proximities such as railway

Build a Dashboard with Plotly Dash

- https://github.com/sweetps/datascience-capstone/blob/main/spacex_dash_appy.py

- Created Interactive Pie Chart of total number of successful launches with drop down list to select launch site
- Created a linked scatter plot of PayloadMass vs Success with point colored based on booster version category. Enabled user to select the range of payload.
- Explain why you added those plots and interactions
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

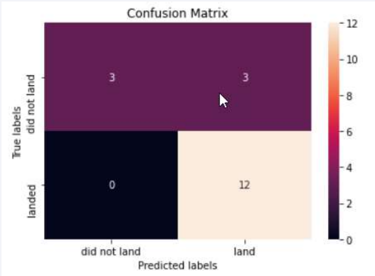
- Finding Insights Visually
  1. Which site has the largest successful launches? CFAFS LC-40
  2. Which site has the highest launch success rate? KSC LC-39A
  3. Which payload range(s) has the highest launch success rate? 2K-4K
  4. Which payload range(s) has the lowest launch success rate? 2K-6K
  5. Which F9 Booster version (v1.0, v1.1, FT, B4, B5, etc.) has the highest launch success rate? FT

Predictive Analysis (Classification)

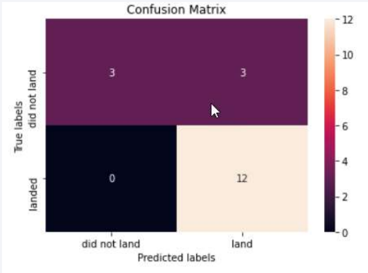
- [datascience-capstone/Applied Data Science Capstone - ML Prediction.ipynb at master · sweetps/datascience-capstone (github.com)]

- Find best Hyperparameter using **GridSearchCV**

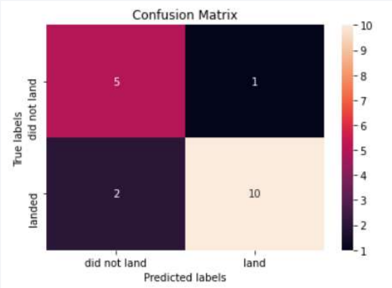
K-Nearest Neighbor (KNN) – score = 0.83



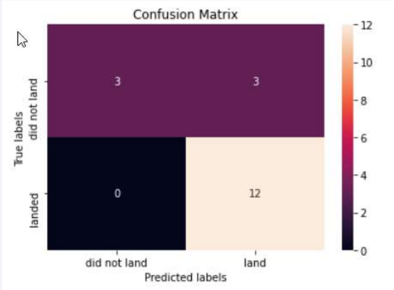
Support Vector Machine (SVM) – score = 0.83



Decision Trees (DT) – score = 0.83



Logistic Regression (LR) – score = 0.83



Results

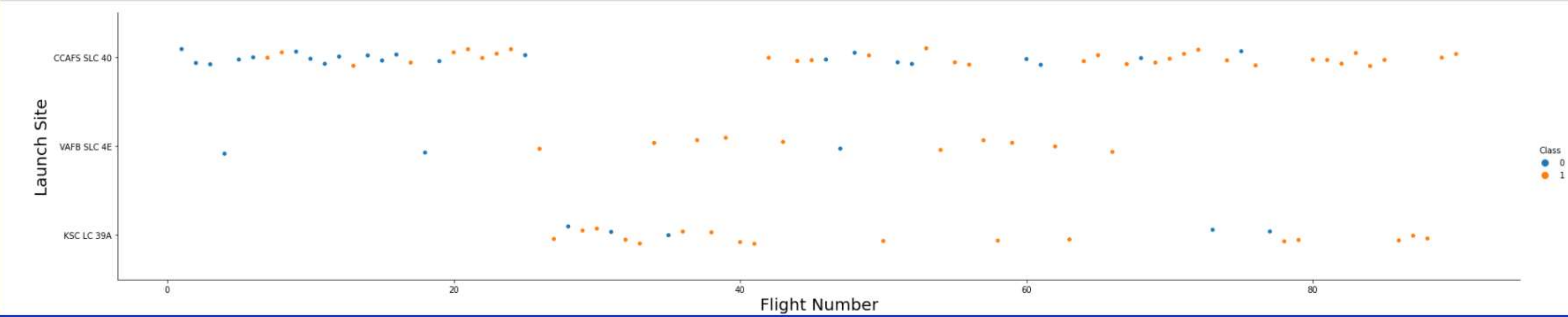
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Launch Sites Proximities Analysis
- Predictive analysis results

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

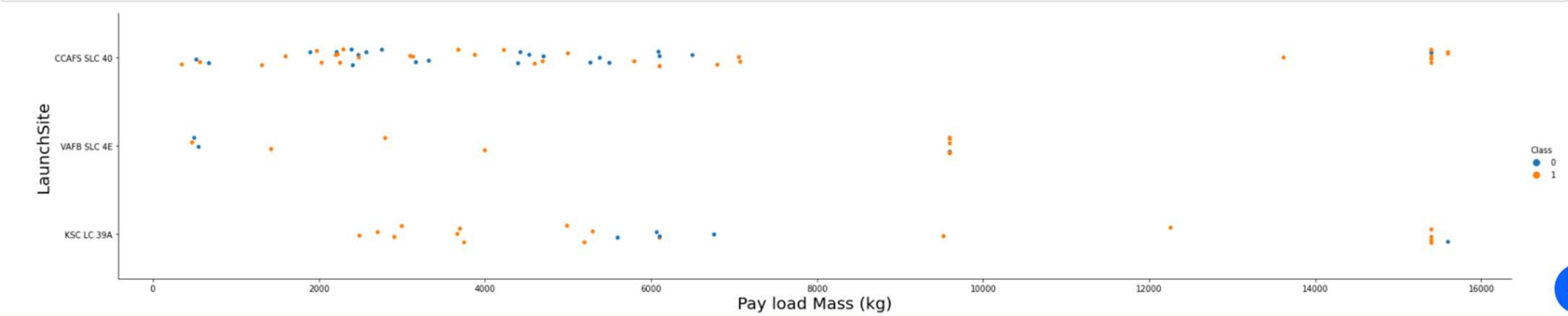
```
# Plot a scatter point chart with x axis to be Flight Number and y axis to be the launch site, and hue to be the class value
sns.catplot(y="LaunchSite", x="FlightNumber", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("Launch Site",fontsize=20)
plt.show()
```



- VAFB SLC 4E had fewer launches than the other 2 sites and has had fewer failures.
- CCAFS SLC 40 had many failures early on.

Payload vs. Launch Site

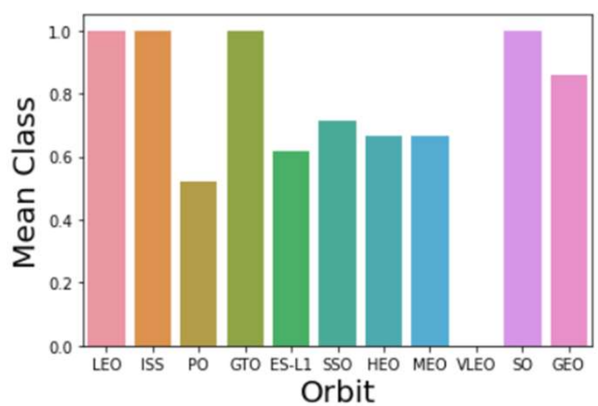
```
# Plot a scatter point chart with x axis to be Pay Load Mass (kg) and y axis to be the launch site, and hue to be the class value
sns.catplot(x="PayloadMass", y="LaunchSite", hue="Class", data=df, aspect = 5)
plt.ylabel("LaunchSite",fontsize=20)
plt.xlabel("Pay load Mass (kg)",fontsize=20)
plt.show()
```



- VAFB SLC 4E has not launched heavy loads.
- CCAFS SLC 40 and KSC LC 39A launched more lighter loads than heavy loads.

Success Rate vs. Orbit Type

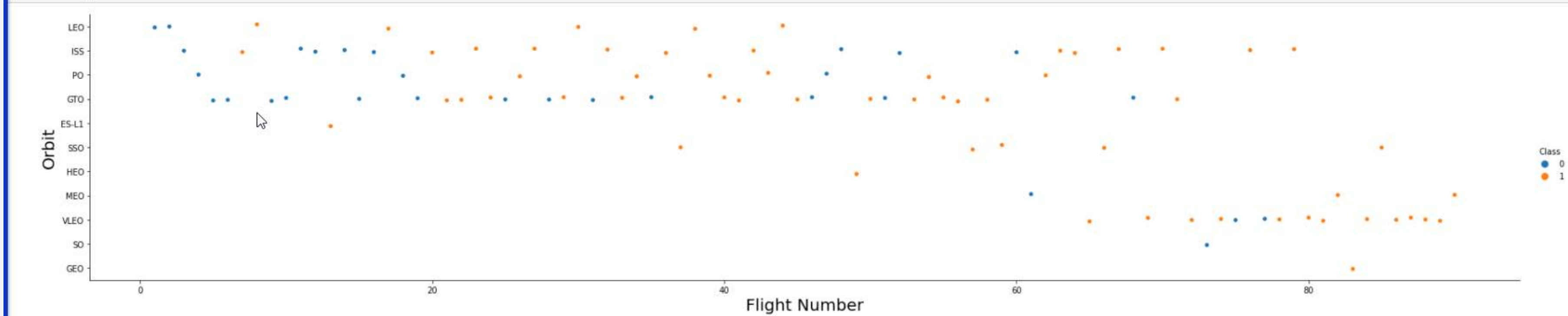
```python
# HINT use groupby method on Orbit column and get the mean of Class column
#df.groupby("Orbit").mean(['Class']).plot(kind='bar')
sns.barplot(x = df['Orbit'].unique() , y = df.groupby(['Orbit'])['Class'].mean())
plt.xlabel("Orbit",fontsize=20)
plt.ylabel("Mean Class",fontsize=20)
plt.show()
```



- LEO, ISS, GTO, SO orbits all had a mean success (class) of 1.0.
- PO orbit had the smallest mean success

Flight Number vs. Orbit Type

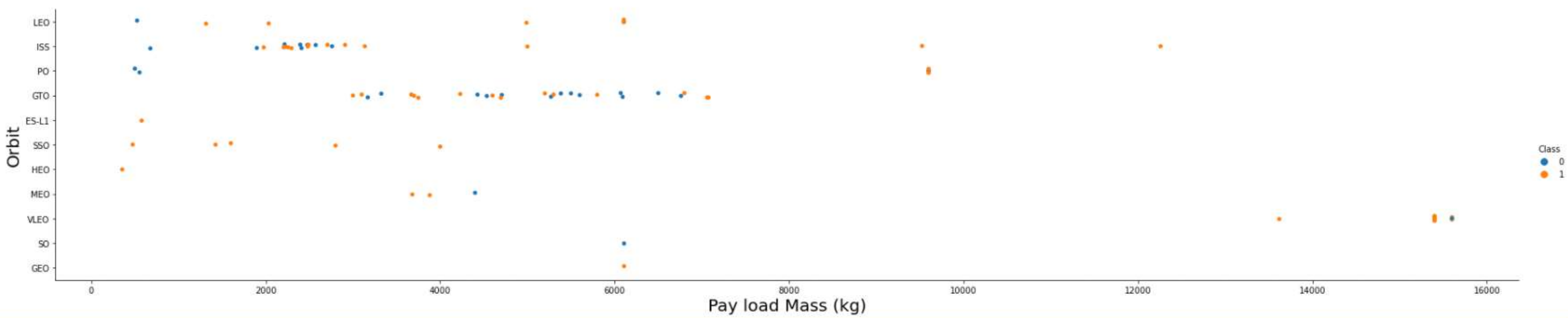
```
# Plot a scatter point chart with x axis to be FlightNumber and y axis to be the Orbit, and hue to be the class value
sns.catplot(y="Orbit", x="FlightNumber", hue="Class", data=df, aspect = 5)
plt.xlabel("Flight Number",fontsize=20)
plt.ylabel("Orbit",fontsize=20)
plt.show()
```



- LEO orbit the Success appears related to the number of flights
- There seems to be no relationship between flight number when in GTO orbit

Payload vs. Orbit Type

```
# Plot a scatter point chart with x axis to be Payload and y axis to be the Orbit, and hue to be the class value
sns.catplot(y="Orbit", x="PayloadMass", hue="Class", data=df, aspect = 5)
plt.xlabel("Pay load Mass (kg)",fontsize=20)
plt.ylabel("Orbit",fontsize=20)
plt.show()
```



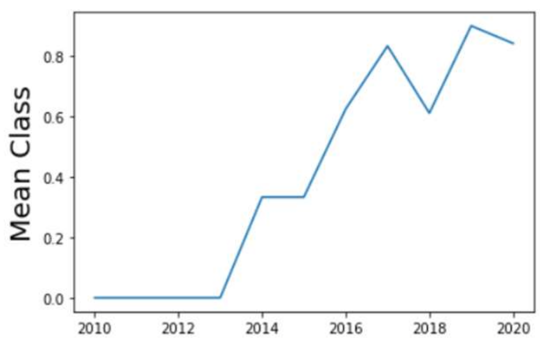
- Heavy payloads have a negative influence on GTO orbits
- Heavy payloads have a positive on GTO and Polar LEO (ISS) orbits.

Launch Success Yearly Trend

```python
# Plot a line chart with x axis to be the extracted year and y axis to be the success rate
df['Year'] = pd.DataFrame(Extract_year(df['Date'])).astype('int')
sns.lineplot(x = df['Year'].unique() , y = df.groupby(['Year'])['Class'].mean())
plt.ylabel("Mean Class",fontsize=20)
df.shape
df.head()
```

7]:

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class	Year
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	0	2010
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	0	2012
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	0	2013
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	0	2013
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	0	2013



- The success rate since 2013 rose from 2017, rising again in 2018
- In 2019 the success rate began to decrease again

All Launch Site Names

```
%%sql
select distinct launch_site from NLM13978.SPACEX
```

```
 * ibm_db_sa://nlm13978:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

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

- Use Distinct so each launch site is included only one time

Launch Site Names Begin with 'CCA'

```
%%sql
select LAUNCH_SITE from NLM13978.SPACEX WHERE launch_site like 'CCA%' limit 5
```

```
 * ibm_db_sa://nlm13978:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

launch_site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40

- Filter launch_site on ‘CCA’ in where clause

Total Payload Mass

```
%%sql
select SUM(PAYLOAD_MASS__KG_) from NLM13978.SPACEX WHERE CUSTOMER='NASA (CRS)'

 * ibm_db_sa://nlm13978:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.

   1
45596
```

- Filter CUSTOMER on 'NASA (CRS)' in where clause
- Aggregate PAYLOAD_MASS__KG_ using SUM

Average Payload Mass by F9 v1.1

```
%%sql
select SUM(PAYLOAD_MASS__KG_)/count(*) from NLM13978.SPACEX WHERE BOOSTER_VERSION='F9 v1.1'

 * ibm_db_sa://nlm13978:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

9]:

1
2928

- Filter on BOOSTER_VERSION on ‘F9 v1.1’ in where clause
- Aggregate PAYLOAD_MASS__KG_ using SUM

First Successful Ground Landing Date

```
%%sql
select MIN(DATE) from NLM13978.SPACEX WHERE LANDING__OUTCOME='Success'
```

```
 * ibm_db_sa://nlm13978:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

1
2018-07-22

- Filter LANDING__OUTCOME on 'Success' in where clause
- Find first using MIN function on DATE

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%%sql
select BOOSTER_VERSION,PAYLOAD_MASS__KG_,MISSION_OUTCOME from NLM13978.SPACEX WHERE PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000 AND MISSION_OUTCOME='Success'
```

```
 * ibm_db_sa://nlm13978:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

booster_version	payload_mass__kg_	mission_outcome
F9 v1.1	4535	Success
F9 v1.1 B1011	4428	Success
F9 v1.1 B1014	4159	Success
F9 v1.1 B1016	4707	Success
F9 FT B1020	5271	Success
F9 FT B1022	4696	Success
F9 FT B1026	4600	Success
F9 FT B1030	5600	Success
F9 FT B1021.2	5300	Success
F9 FT B1032.1	5300	Success

- Filter on PAYLOAD_MASS__KG_ on between 4000 and 6000 in where clause
- And Filter on MISSION_OUTCOME equal 'Success'

Total Number of Successful and Failure Mission Outcomes

```
%%sql
select count(MISSION_OUTCOME) from NLM13978.SPACEX WHERE MISSION_OUTCOME='Success' union select count(MISSION_OUTCOME) from NLM13978.SPACEX WHERE MISSION_OUTCOME<>'Success'
```

```
 * ibm_db_sa://nlm13978:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

1
2
99

- union of count MISSION_OUTCOME equal 'Success' and count MISSION_OUTCOME not equal 'Success'

Boosters Carried Maximum Payload

```
%%sql
select BOOSTER_VERSION from NLM13978.SPACEX WHERE PAYLOAD_MASS__KG_=(select max(PAYLOAD_MASS__KG_) from NLM13978.SPACEX)
```

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

- SELECT BOOSTER_VERSION where PAYLOAD_MASS__KG_ is the max AYLOAD_MASS__KG_ using a subquery

2015 Launch Records

```
%%sql
select LANDING__OUTCOME,BOOSTER_VERSION,LAUNCH_SITE, MISSION_OUTCOME from NLM13978.SPACEX WHERE LANDING__OUTCOME NOT LIKE 'Success%' AND YEAR(DATE)=2015
```

```
 * ibm_db_sa://nlm13978:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

landing__outcome	booster_version	launch_site	mission_outcome
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	Success
Controlled (ocean)	F9 v1.1 B1013	CCAFS LC-40	Success
No attempt	F9 v1.1 B1014	CCAFS LC-40	Success
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	Success
No attempt	F9 v1.1 B1016	CCAFS LC-40	Success
Precluded (drone ship)	F9 v1.1 B1018	CCAFS LC-40	Failure (in flight)

- Filter year on '2015' and LANDING__OUTCOME not like 'Success%' in where clause

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

```
%%sql
select LANDING__OUTCOME,COUNT(LANDING__OUTCOME) from NLM13978.SPACEX GROUP BY LANDING__OUTCOME
```

```
 * ibm_db_sa://nlm13978:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

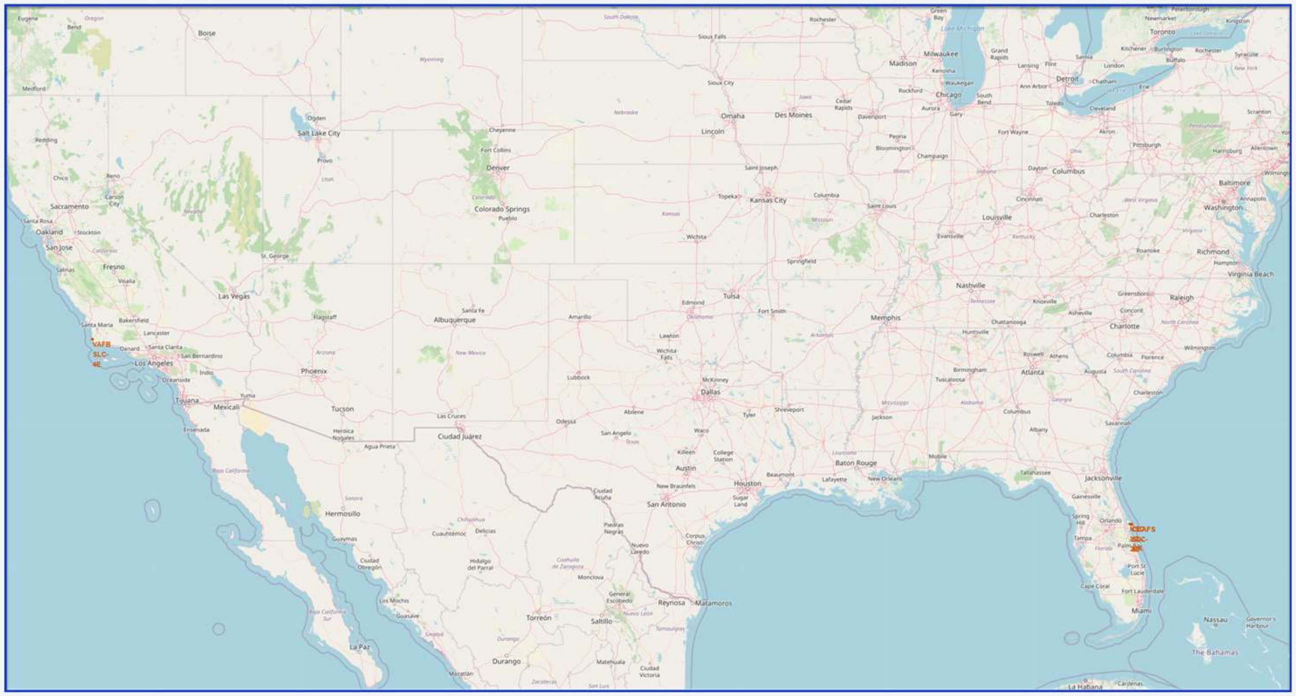
landing__outcome	2
Controlled (ocean)	5
Failure	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	22
Precluded (drone ship)	1
Success	38
Success (drone ship)	14
Success (ground pad)	9
Uncontrolled (ocean)	2

- I had to create a new account because I exceeded my monthly limit. In new account I cannot get connected to new database. I cannot correct & rerun query. Should be as follows:
    - select LANDING__OUTCOME,COUNT(LANDING__OUTCOME) from NLM13978.SPACEX WHERE DATE between '06-04-2010' and '03-20-2017' GROUP BY LANDING__OUTCOME ORDER BY COUNT(LANDING__OUTCOME) DESC

Section 4

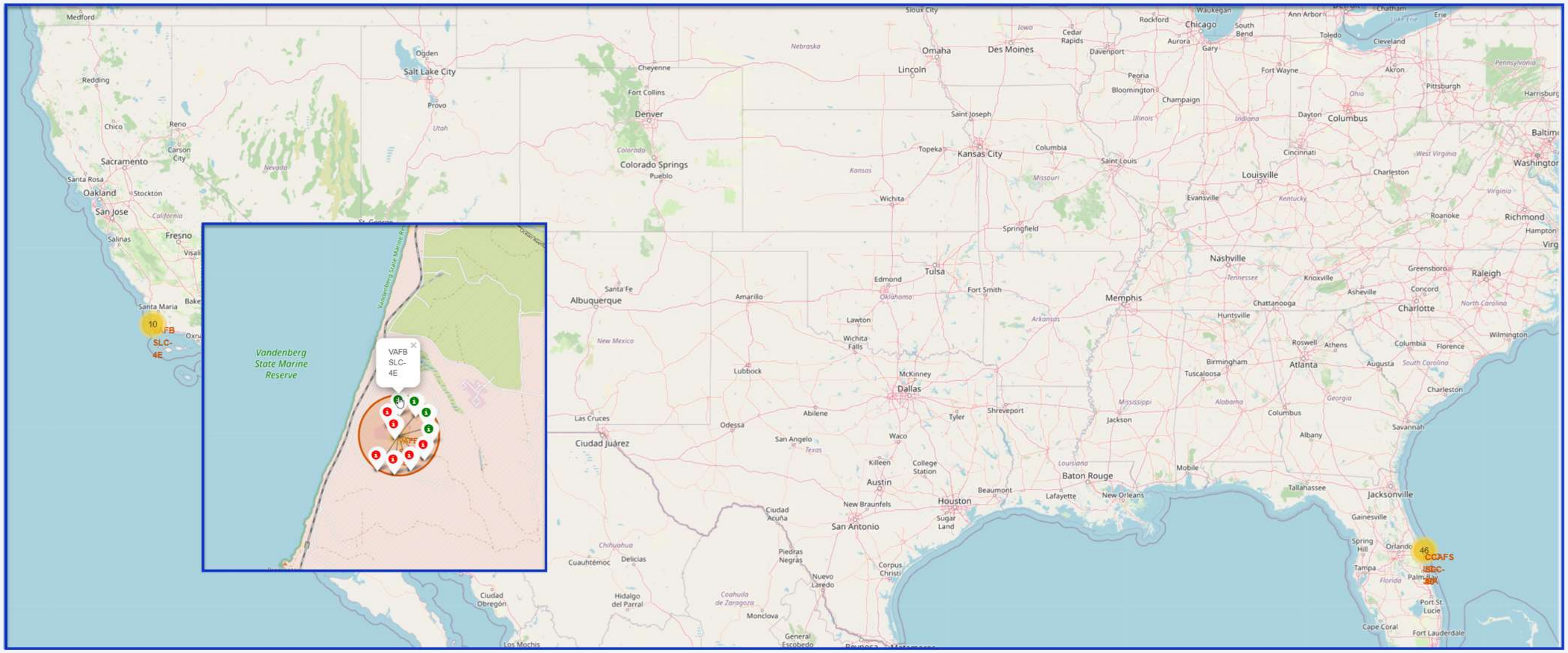
Launch Sites Proximities Analysis

Launch Site Locations



- Lanch sites have been marked on map.

Launch Clusters



- Launches are marked on map with cluster marker
- Zoom in to see each launch marked with Green (success) or Red (Fail) marker

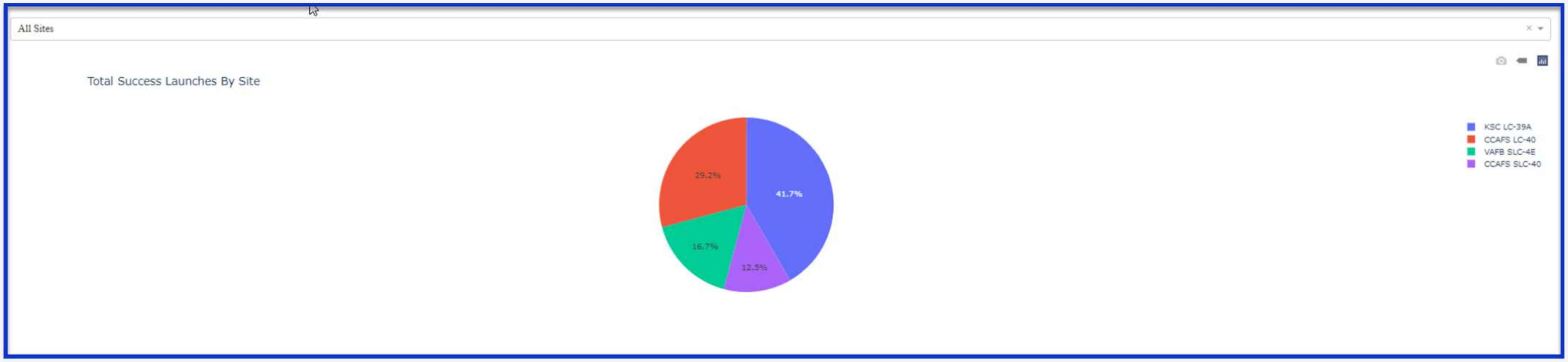
Plot Distances Between a Launch Site to its Proximities

- Lines can be added to show distances between a launch site and its proximities such as railway, highway, coastline

Section 5

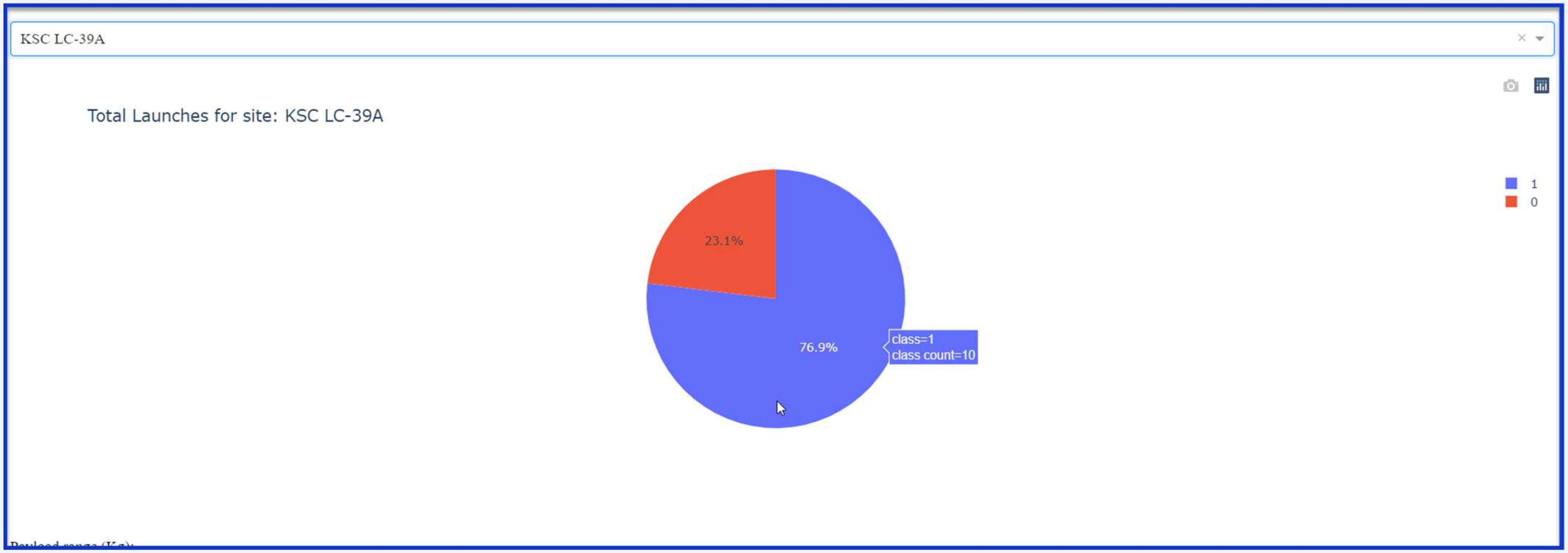
Build a Dashboard with Plotly Dash

Success Rate by Launch Site



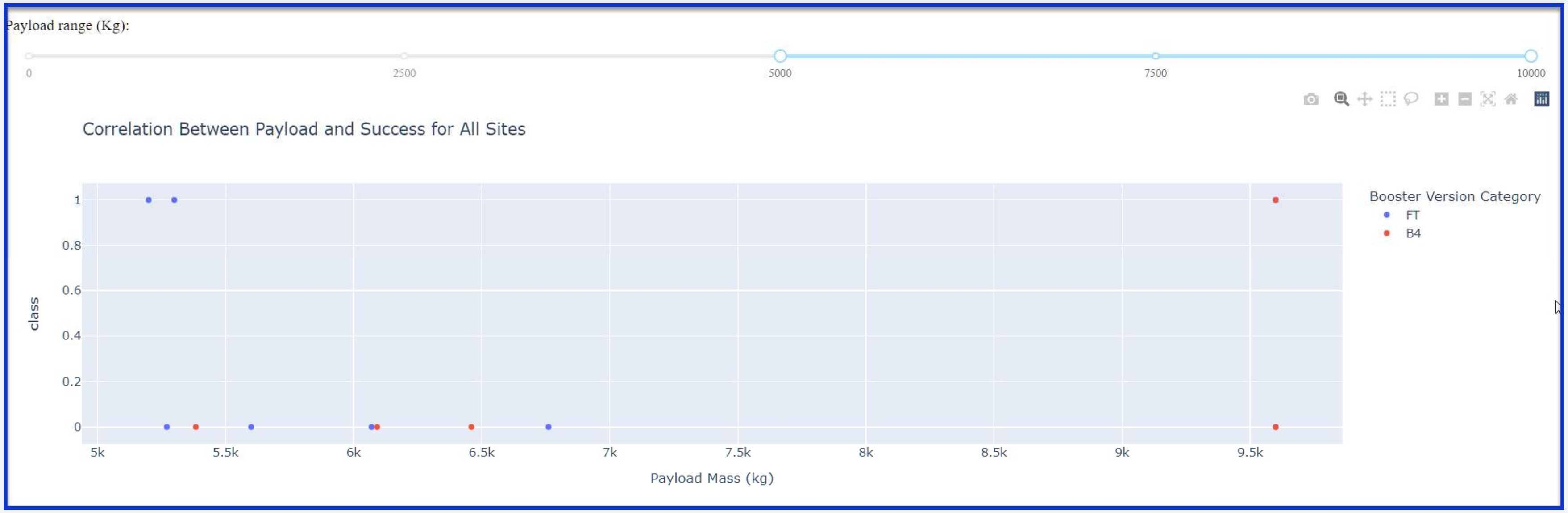
- KSC LC-39A has the highest success rate

Launch Site with Highest Success Ratio



- KSC LC-39A has the highest success ratio.
- 10 of 13 rockets launched were successful

Correlation Between Heavy Payloads and Success

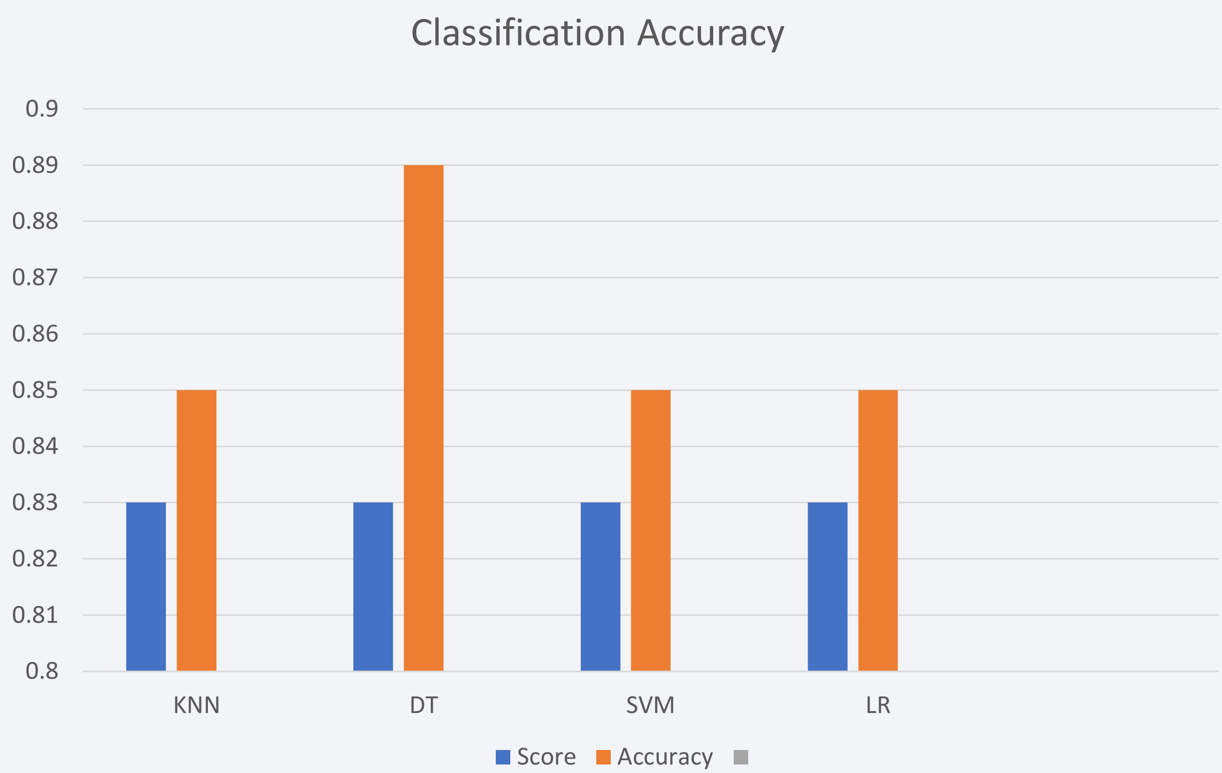


- B4 and FT are the heaviest payloads. There have only been 3 successful launches, 2 FT and 1 B4.

Section 6

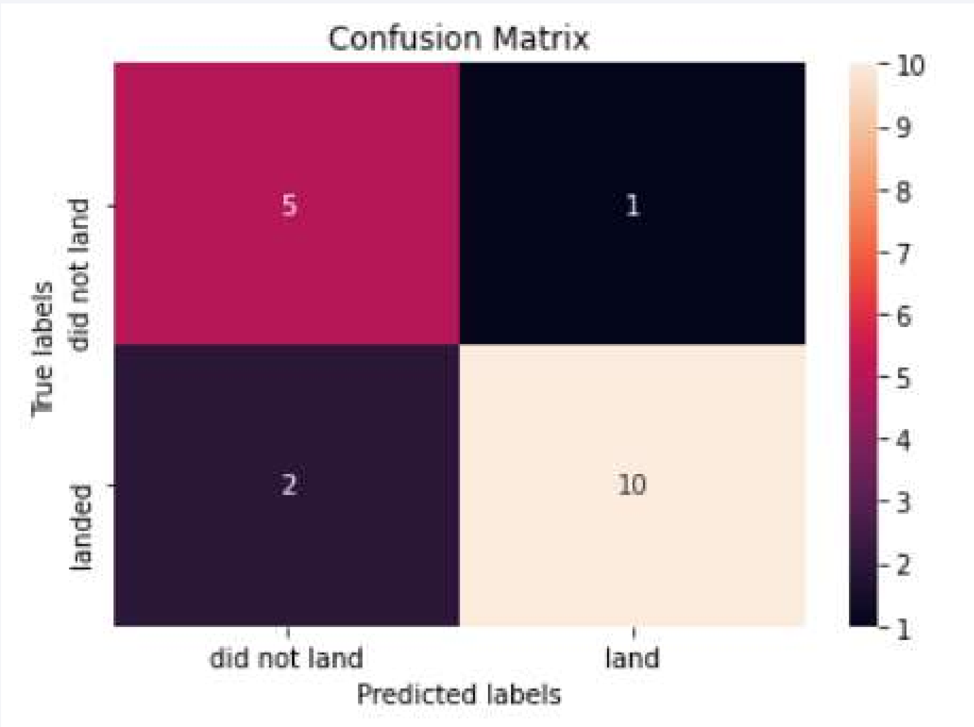
Predictive Analysis (Classification)

Classification Accuracy



- Based on the score all methods performed the same
- Using accuracy DT is slightly better than the other methods

Confusion Matrix – KNN is Best



- KNN Confusion Matrix is the best because it provides the greater number of accurate classifications for 'did not land' while still only misclassifying 2 of the landed
- KNN classified15 correctly as did the other 3 methods

Conclusions

- Using ML we can predict with ~85% accuracy whether or not a launch's first stage booster will fail.
- Given the high stakes of making a bad decision, more analysis to determine under which circumstances the models fail to predict success accurately, checking for outliers
- Heavy payloads have high failure rate and should be invested in with extreme caution

Appendix

- Data Collection and Wrangling
  - Data Collection Jupyter Notebook - https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/07f237de-c67c-4347-94a0-293a07e294f3/view?access_token=bb4416bacfe323bda1a48f18f3fe5ea38c2bc87ef2475c87ab3c92cbde002c70
    - https://api.spacexdata.com/v4/rockets
    - https://api.spacexdata.com/v4/lauchpads
    - https://api.spacexdata.com/v4/payloads
    - https://api.spacexdata.com/v4/cores
    - https://api.spacexdata.com/v4/past
  - Web Scraping Jupyter Notebook - https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/cc1600ab-4e0f-4a7d-b435-46f2f16aebdd/view?access_token=e3498a131042a00ea74093f937fc1b765ad4cce4340c19939e93c6871bac2634
    - https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
  - Data Wrangling Jupyter Notebook - https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/e6818c7f-c162-43c2-a03b-13cf13e8d9f7/view?access_token=506e0e5231d0fcf29bfb0516bbd5d3238d2f9794320cb7732205bf43e0c671f5
- EDA
  - EDA Visualization Jupyter Notebook – https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/cb361d0a-5ee9-4670-83e1-43dbbb7a87a7/view?access_token=cc5cd130699a93fb06f223906a8f854b6eb5d2774bc21846b861bffd96f8714e
  - EDA SQL Jupyter Notebook – https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/a3fc5b0f-6093-4c29-939d-82d31b0554dd/view?access_token=5e21e539e55fd2463c7a2264d7448427b8e9a23791c1e7b52b00ac45dbe64321
    - Spacex DataSet

Appendix

- **Launch Sites Locations Analysis with Folium**
  - Folium Map Jupyter Notebook - https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/8c7b5c1e-c8ed-43ff-afdf-d5d2c4703fbf/view?access_token=9f3d24228f7b2a13eae7c47b55a6b3bd1f04ede8196a0492614b0e91e93e56a8
- **Interactive Dashboard with Plotly Dash**
  - https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/spacex_launch_dash.csv
- ML Prediction
  - ML Prediction Jupyter Notebook - https://eu-de.dataplatform.cloud.ibm.com/analytics/notebooks/v2/205f3569-da11-4816-bf41-238470b02064/view?access_token=01dbc3e0948444e847974b5055b45aee2f2b70f8fbd85defaeb6cbc42dff177d


Thank you!