IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Anushka Thenuwara
Feb 2022

Outline

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

Executive Summary

- Summary of methodologies and results
  - Numerous processes were followed starting from obtaining the data, cleaning data, exploratory analyses and preliminary visualizations
  - Primary outcomes were
  - An interactive map with Folium
  - An interactive Dashboard with Plotly Dash
  - Predictive analysis (Classification)

Introduction

- The background

  The focal point of the study revolved around SpaceX where they publish a wide range of information about their launces online. Out of their reusable fleet of rocket ships the Falcon 9 rocket boasts a cost saving of about 100 million dollars due to its reusability of the first stage. The study was conducted to see how certain it is that the stage will successfully and the conditions that boosts this certainty so the cost of launch can be determined

- What we are looking for

  Find the parameters that impact the success of landing such as;

  - Is it location dependent?
  - Does the payload increase the risk of failure?
  - Any other insights we can obtain

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was collected via APIs and web scraping
- Perform data wrangling
  - One Hot Encoding data fields for Machine Learning and dropping irrelevant columns
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

Data collection was done via two methods

- SpaceX REST API
    - https://api.spacexdata.com/v4/rockets/

- Web scraping Wikipedia using BeautifulSoup.

Data Collection – SpaceX API

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False
...	...	...	...	...	...	...	...	...	...
89	86	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True
90	87	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True
91	88	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True
92	89	2020-10-24	Falcon 9	15600.0	VLEO	CCSFS SLC 40	True ASDS	3	True
93	90	2020-11-05	Falcon 9	3681.0	MEO	CCSFS SLC 40	True ASDS	1	True

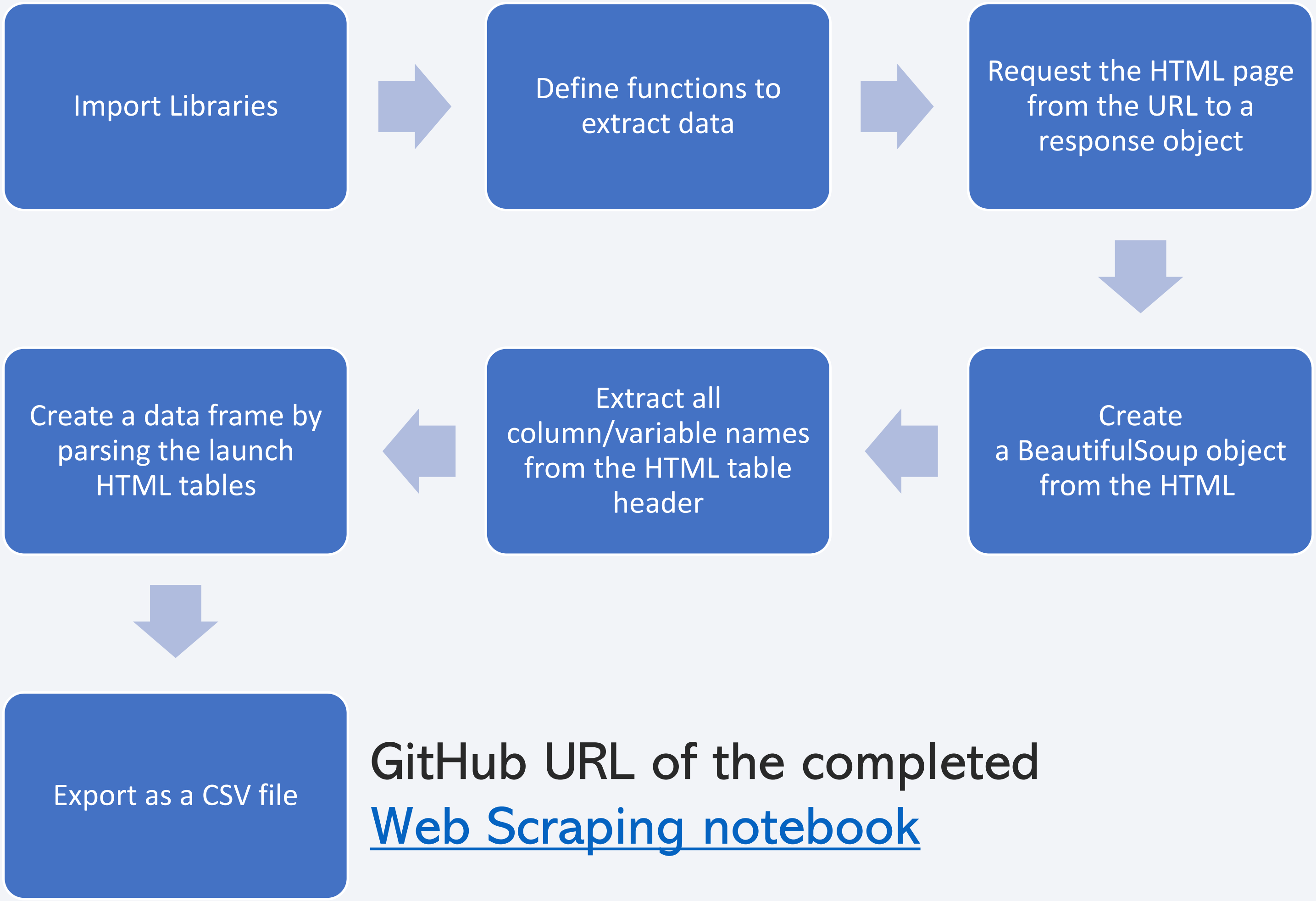
90 rows × 17 columns

1. Import Libraries
2. Define functions to extract data
3. Requesting rocket launch data from SpaceX API
4. Decode response as a Jason and convert to a Pandas data frame
5. Clean data frame, extract values via functions defined earlier and filter for Falcon 9 data
6. Check for missing values and replace payload mass with mean value
7. Export as a CSV file

GitHub URL of the completed [SpaceX API calls notebook]

Data Collection - Scraping

	Flight No.	Launch site	Payload	Payload mass	Orbit
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO
1	2	CCAFS	Dragon	0	LEO
2	3	CCAFS	Dragon	525 kg	LEO
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO



GitHub URL of the completed [Web Scraping notebook]

Data Wrangling

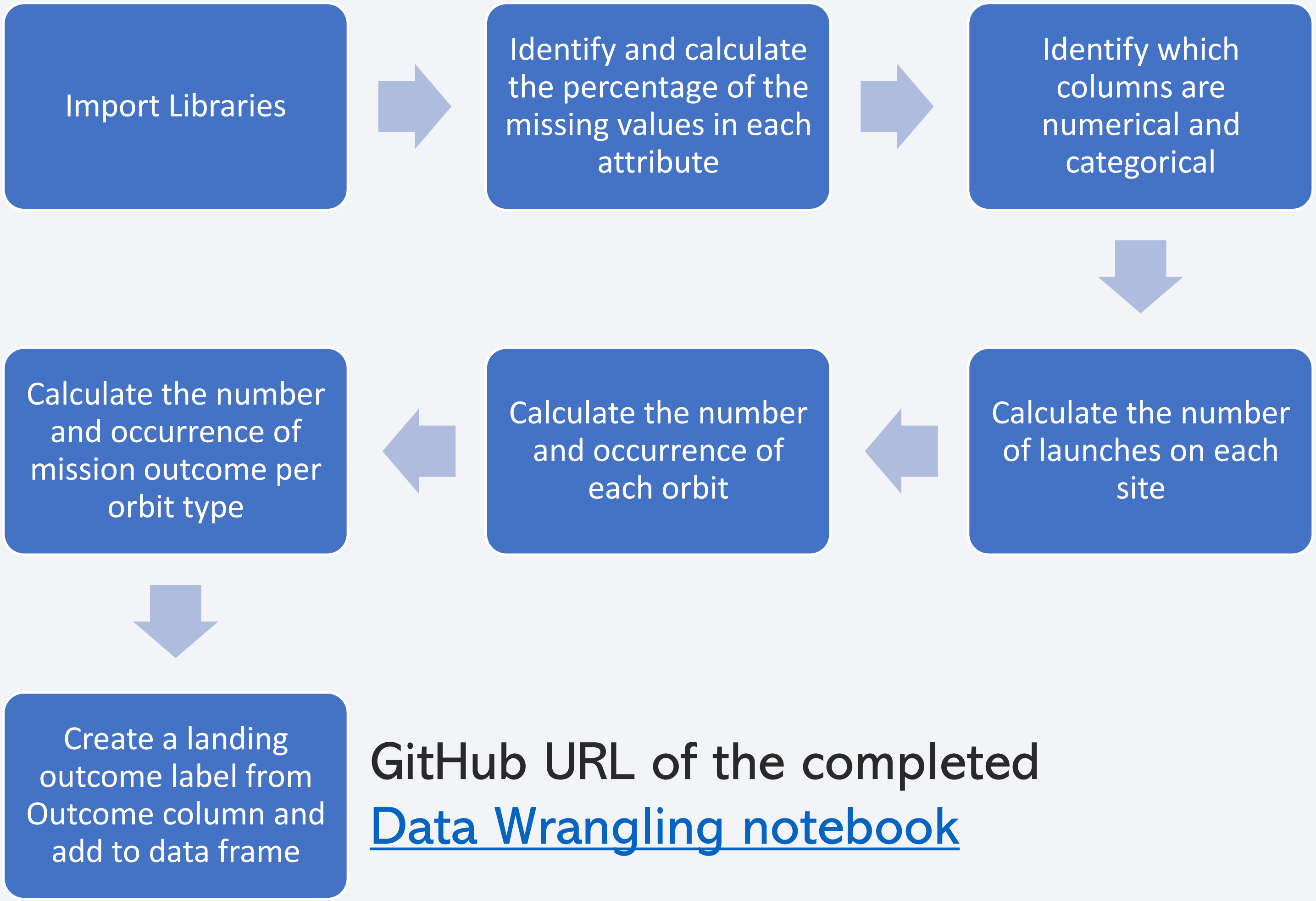
Primary objectives are;

- Performed Exploratory Data Analysis
- Determine Training Labels

CCAFS SLC 40	55
KSC LC 39A	22
VAFB SLC 4E	13

True ASDS	41
None None	19
True RTLS	14
False ASDS	6
True Ocean	5
None ASDS	2
False Ocean	2
False RTLS	1

GTO	27
ISS	21
VLEO	14
PO	9
LEO	7
SSO	5
MEO	3
ES-L1	1
HEO	1
SO	1
GEO	1



GitHub URL of the completed [Data Wrangling notebook]

EDA with Data Visualization

Data visualization shows the dependencies of parameters with others. These correlations were looked at via the following;

Scatter plots:

- Flight Number VS. Payload Mass
- Flight Number VS. Launch Site
- Payload VS. Launch Site
- Orbit VS. Flight Number
- Payload VS. Orbit Type
- Orbit VS. Payload Mass

Bar graphs

- Mean VS. Orbit

Line graphs

- Success Rate VS. Year

GitHub URL for EDA with data visualization notebook

EDA with SQL

- EDA with SQL and the following were explored via SQL queries;
  - Call the names of the unique launch sites
  - Filter with launch site names
  - Total payload mass carried by boosters launched by specific customers
  - Average payload mass carried by booster versions
  - Total number of successful and failure mission outcomes
  - Booster versions which have carried the maximum payload mass
  - Rank the count of landing outcomes
  - And several other specific cases

GitHub URL for [EDA with SQL notebook]

Build an Interactive Map with Folium

- An interactive map was developed with markers, circles and lines
- The circles were used to locate the launch sites
- The markers were used to denote the successful and unsuccessful landings from each launch site
- Lines were used to denote the nearest locations to prominent geographical locations such as the shoreline, railway and nearest city

GitHub URL for interactive map with Folium map

Build a Dashboard with Plotly Dash

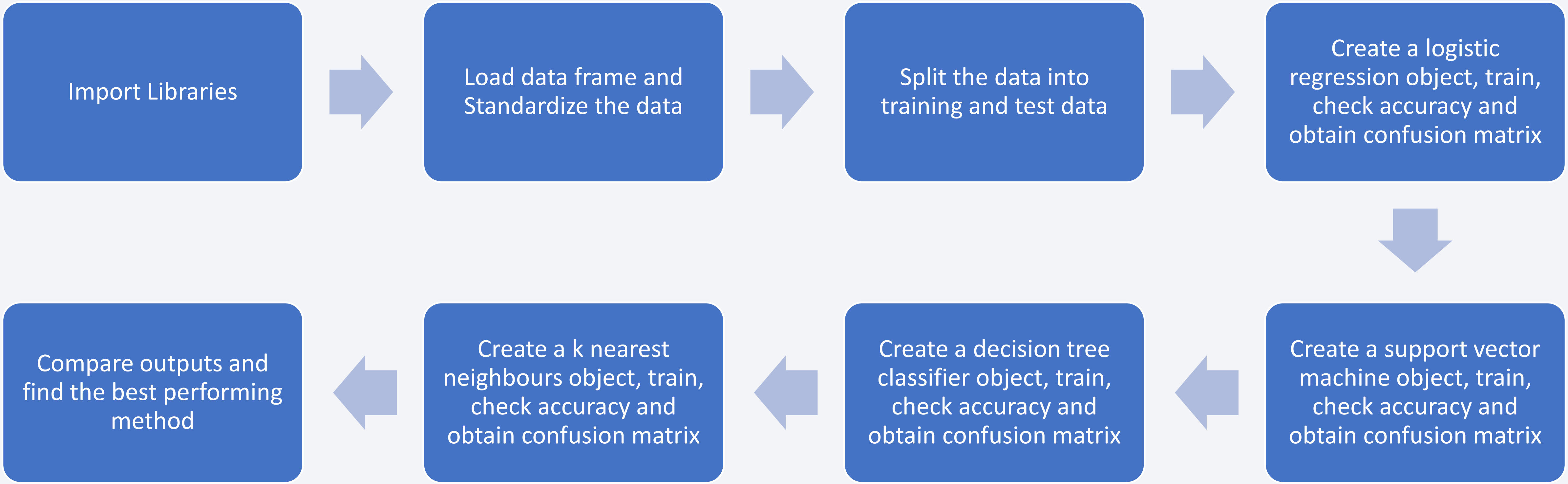
- An interactive dashboard via Plotly was developed which had the following;
    - Pie chart showing the total launches site

    This allows to gauge the portion of launches from each site from the total launches as well as the success rate as a whole

    - Scatter graph showing outcome and payload mass for different booster versions

GitHub URL for [Plotly Dash lab]

Predictive Analysis (Classification)



GitHub URL for [predictive analysis lab]

Results

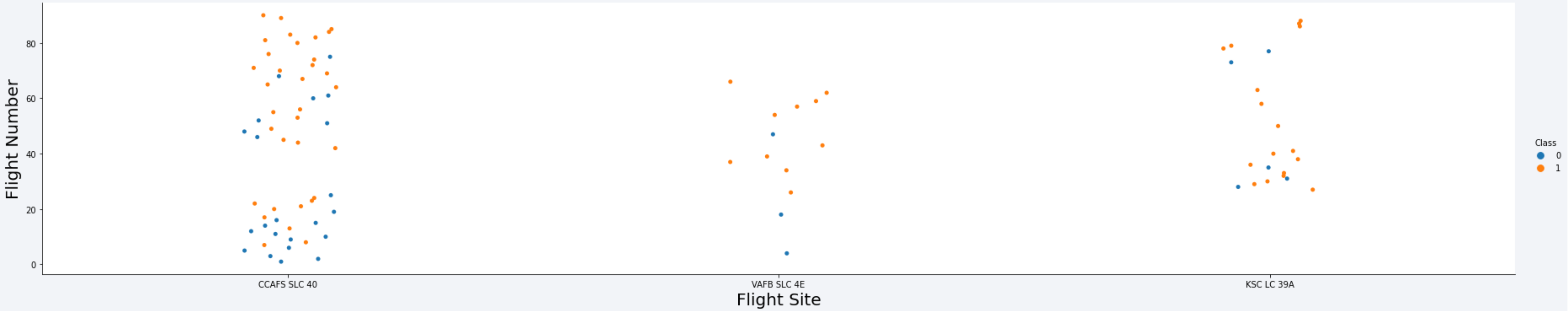
- Exploratory data analysis results
  - The success rates for SpaceX launches increases with time
- Interactive analytics demo in screenshots
  - The KSC LC 39A had the most successful launches from all the sites
- Predictive analysis results
  - The Tree Classifier Algorithm is the best for Machine Learning for this dataset

Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

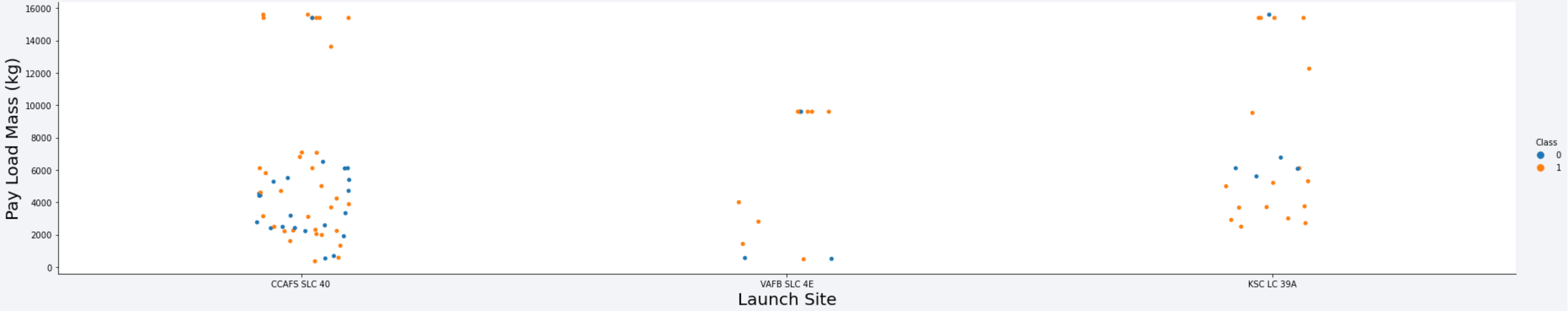
Scatter plot of Flight Number vs. Launch Site



- Shows that the earlier launches were mostly done from CCAFS SLC 40
- The success rate is much better for flights after 20 runs denoting an improvement in internal procedures
- Most launches were done from CCAFS SLC 40 but the other two sites have much better success rates

Payload vs. Launch Site

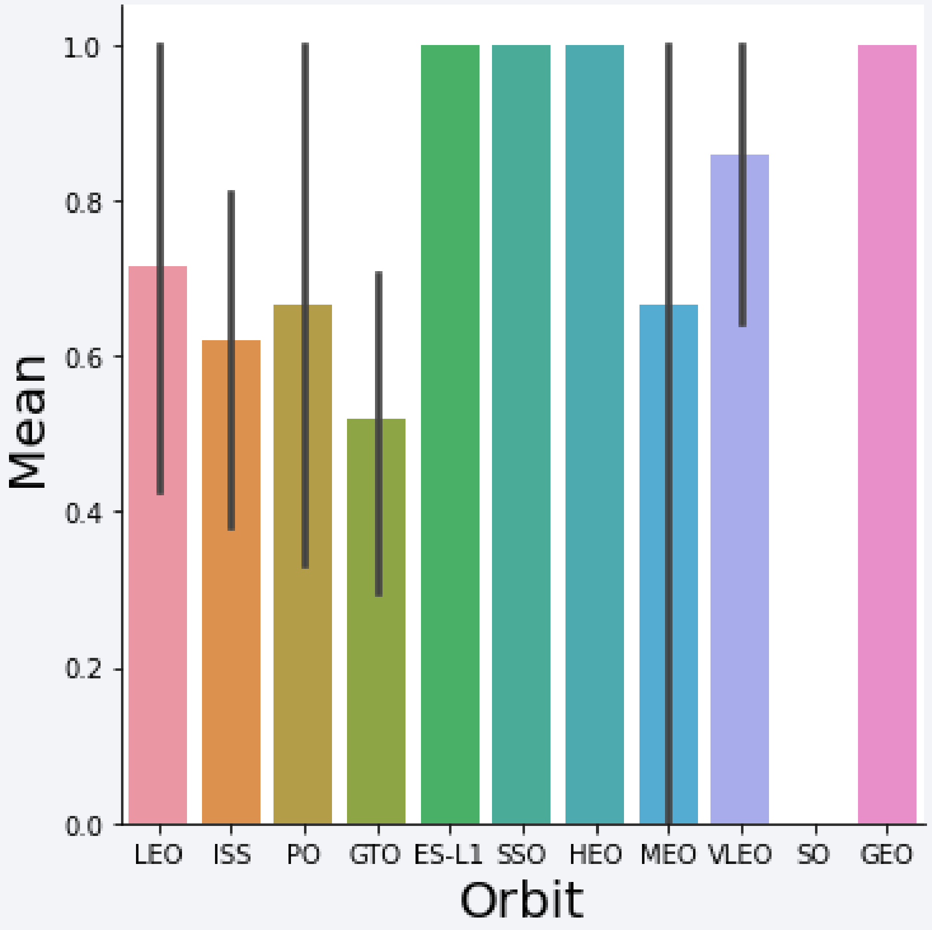
Scatter plot of Payload vs. Launch Site



- Most of the launces carry less than 8000 kg
- CCASF SLC 40 and KSC LC 39A are predominantly chosen for high payload masses of beyond 12000 kg
- VAFB SLC 4E is preferred for payloads in the 8000 – 10000 kg range

Success Rate vs. Orbit Type

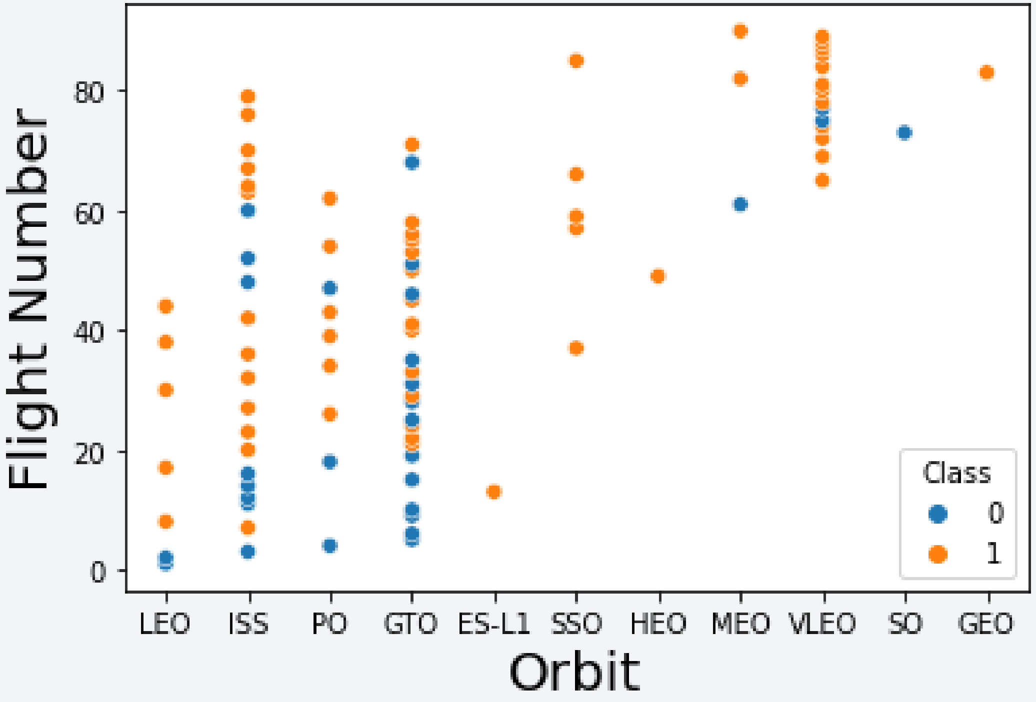
Bar chart for the success rate of each orbit type



- The orbits of ES-L1, SSO, HEO and GEO has 100% success rates
- The other orbits have a success rate of about 50% with large variability

Flight Number vs. Orbit Type

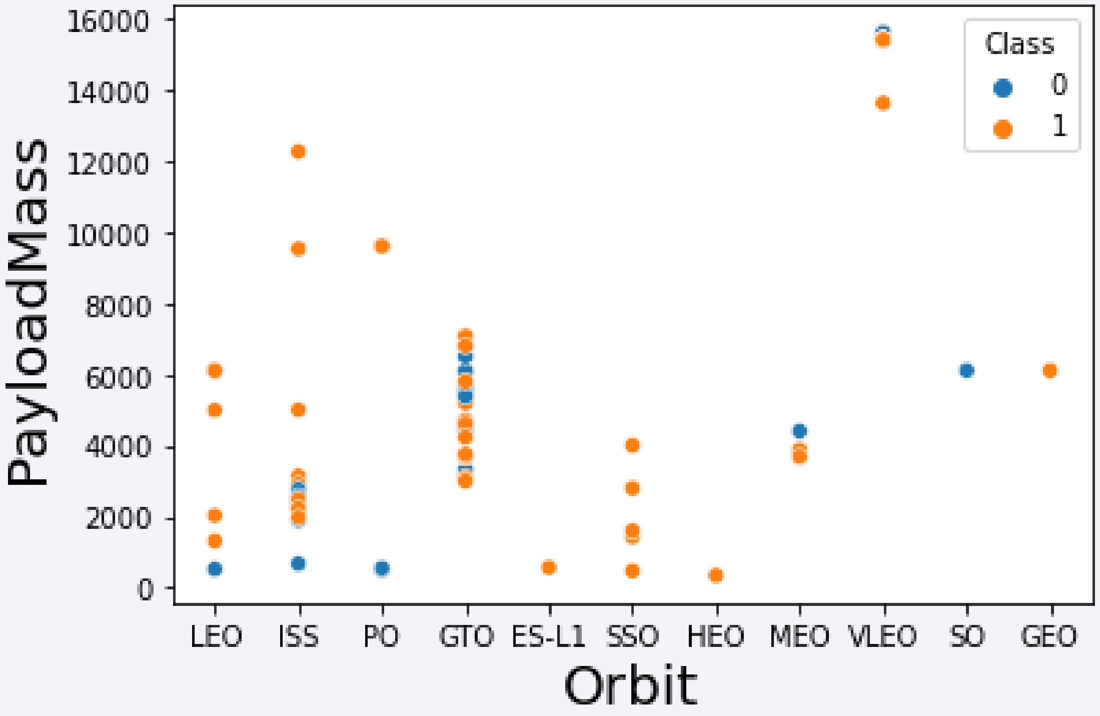
Scatter plot for Flight number vs. Orbit type



- Most of the early lunches were concentrated on LEO, ISS, PO, GTO and ES-L1 orbits
- From flight number 40 onwards the orbits of SSO, HEO, MEO, VLEO, SO and GEO were engaged

Payload vs. Orbit Type

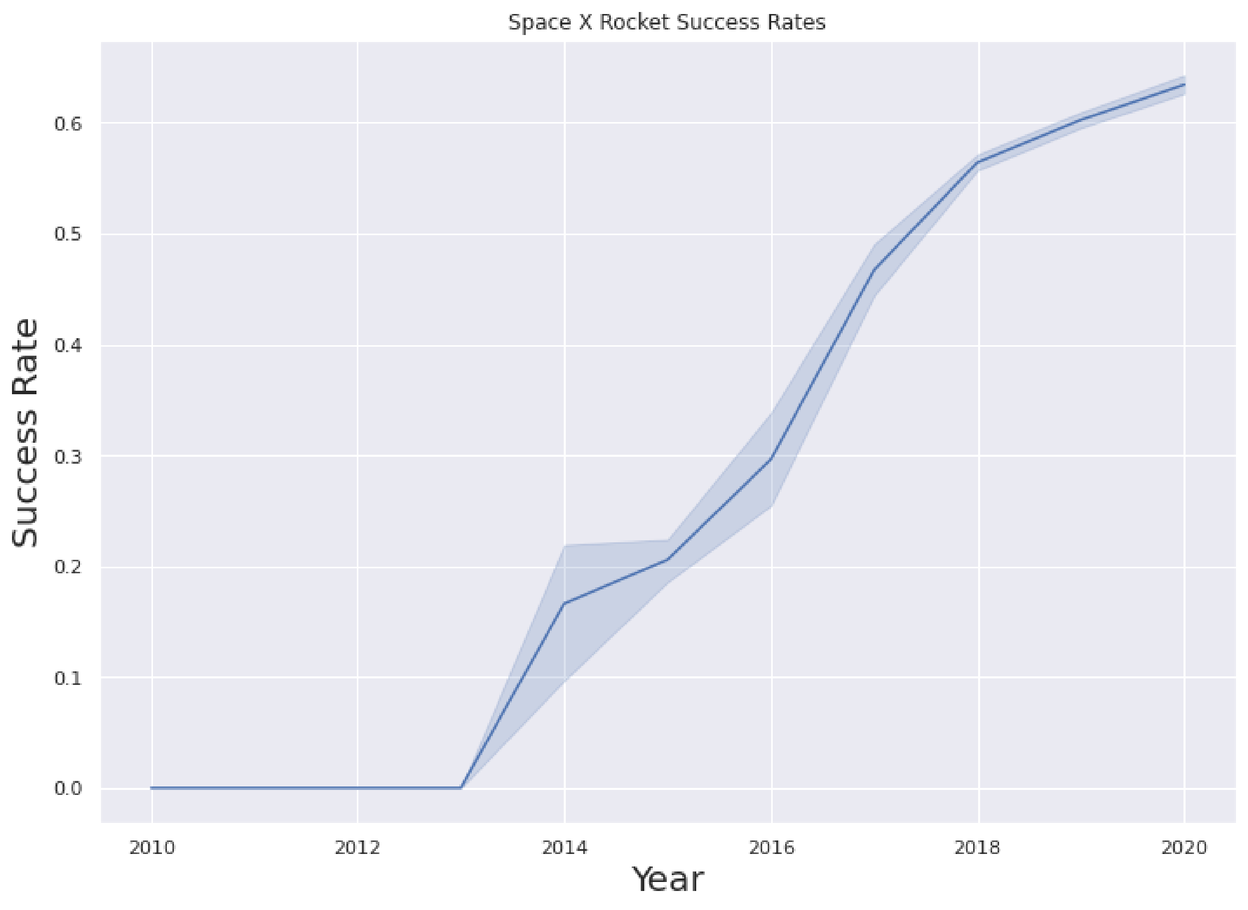
Scatter plot for Payload vs. Orbit type



- GTO orbit has a tight payload mass requirement of 3000 – 8000kg
- LEO, ES-L1, SSO, HEO MEO, SO and GEO orbits command a payload mass less than 7000kg
- ISS, PO and VLEO orbits require large payload masses beyond 9000kg

Launch Success Yearly Trend

Line chart for the yearly average success rate



- Launches starts off from 2013
- The success rate has increased during 2013 – 2017 however with large variations
- Beyond 2018 the certainty of a successful launch is the best it has been

All Launch Site Names

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

In [6]:
```
%sql select distinct LAUNCH_SITE from SX
```

```
 * ibm_db_sa://xtd33361:***/bludb
Done.
```

Out[6]:

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

Launch Site Names Begin with 'CCA'

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

In [7]:

```
%sql select * from SX where LAUNCH_SITE like 'CCA%' fetch first 5 row only
```

```
 * ibm_db_sa://xtd33361:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

Out[7]:

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	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	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

The total payload carried by boosters from NASA

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

In [8]:

```
%sql select sum(PAYLOAD_MASS__KG_) from SX where CUSTOMER = 'NASA (CRS)'
```

```
 * ibm_db_sa://xtd33361:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

Out[8]:

1
45596

Average Payload Mass by F9 v1.1

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

In [9]:

```
%sql select avg(PAYLOAD_MASS__KG_) from SX where BOOSTER_VERSION = 'F9 v1.1'
```

```
 * ibm_db_sa://xtd33361:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

Out[9]:

1
2928

First Successful Ground Landing Date

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

Hint:Use min function

In [10]:
```
%sql select min(date) from SX where LANDING__OUTCOME = 'Success (ground pad)'
```

```
 * ibm_db_sa://xtd33361:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

Out[10]:

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

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

In [11]:

```
%sql select BOOSTER_VERSION from SX where LANDING__OUTCOME ='Success (drone ship)' and PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000
```

```
 * ibm_db_sa://xtd33361:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

Out[11]:

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

Total Number of Successful and Failure Mission Outcomes

List the total number of successful and failure mission outcomes

In [12]:

```
%sql Select MISSION_OUTCOME,count(MISSION_OUTCOME) as count from SX GROUP BY MISSION_OUTCOME
```

```
 * ibm_db_sa://xtd33361:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

Out[12]:

mission_outcome	COUNT
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

In [13]:
```
%sql select distinct BOOSTER_VERSION from SX where PAYLOAD_MASS__KG_ = (select MAX(PAYLOAD_MASS__KG_) FROM SX)
```

```
 * ibm_db_sa://xtd33361:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

Out[13]:

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

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

In [14]:
```
%sql select LANDING__OUTCOME,BOOSTER_VERSION,LAUNCH_SITE from SX where YEAR(DATE) = '2015' and LANDING__OUTCOME = 'Failure (drone ship)'
```

```
 * ibm_db_sa://xtd33361:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

Out[14]:

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

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

In [50]:

```
%sql select LANDING__OUTCOME, count(LANDING__OUTCOME) as count from SX where DATE between '2010-06-04' and '2017-03-20' group by LANDING__OUTCOME order
```

```
 * ibm_db_sa://xtd33361:***@b0aebb68-94fa-46ec-a1fc-1c999edb6187.c3n41cmd0nqnrk39u98g.databases.appdomain.cloud:31249/bludb
Done.
```

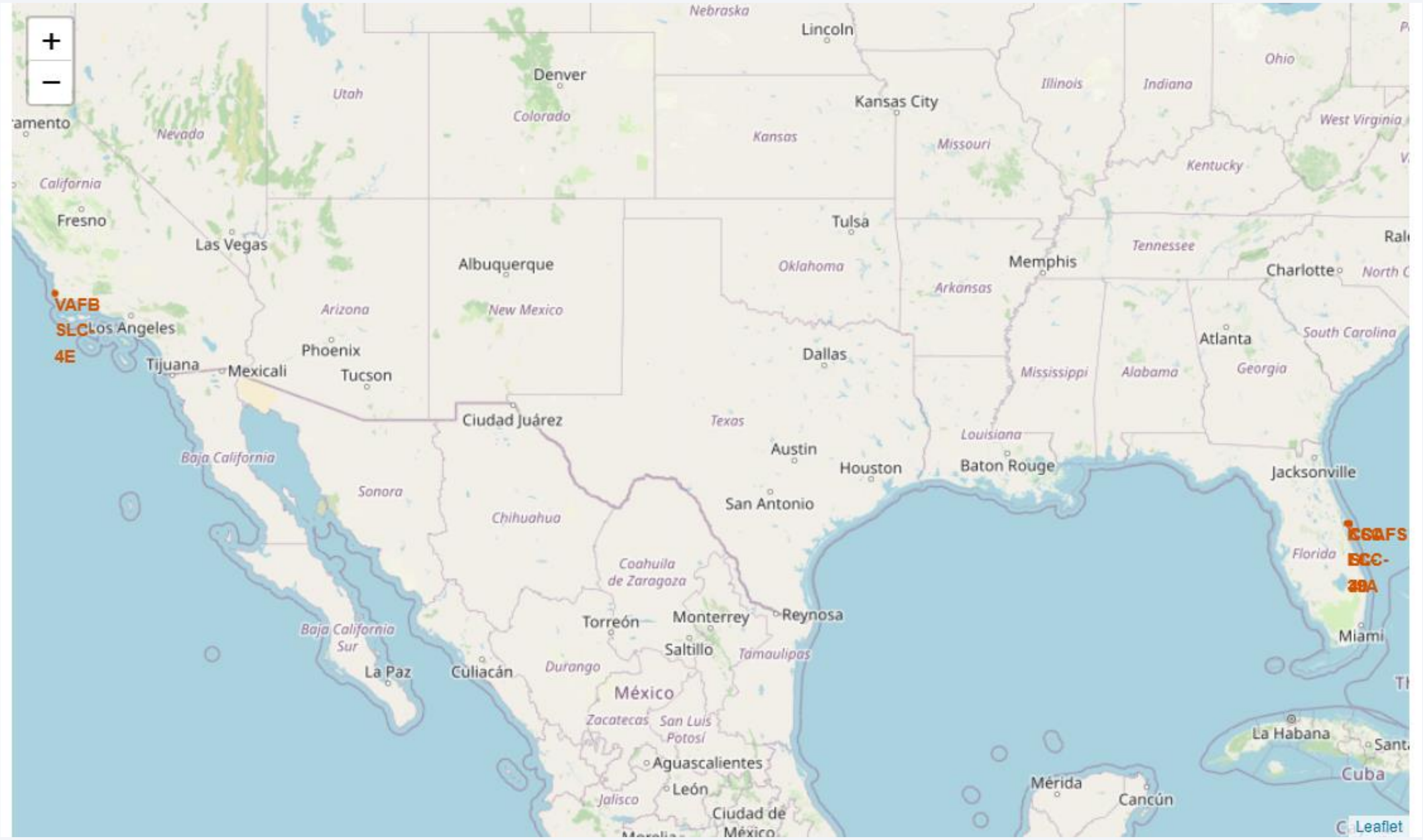
Out[50]:

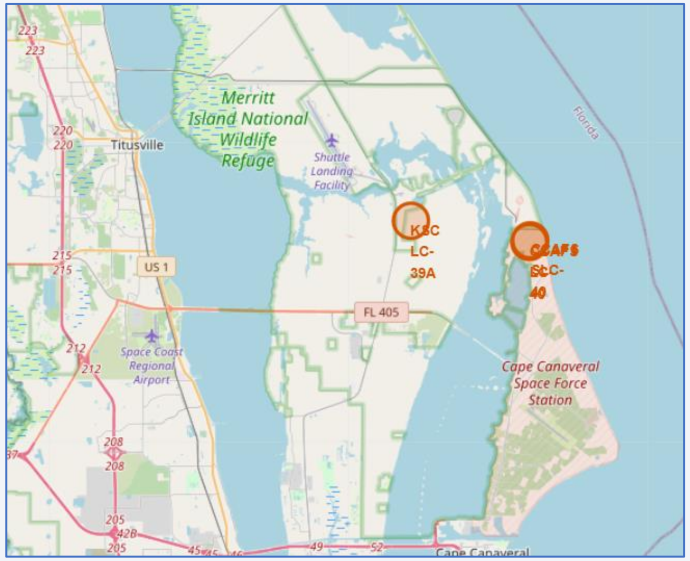
landing__outcome	COUNT
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

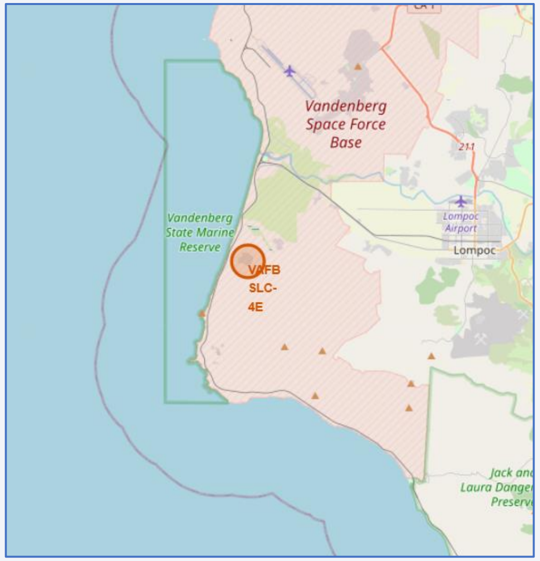
Section 3

Launch Sites Proximities Analysis

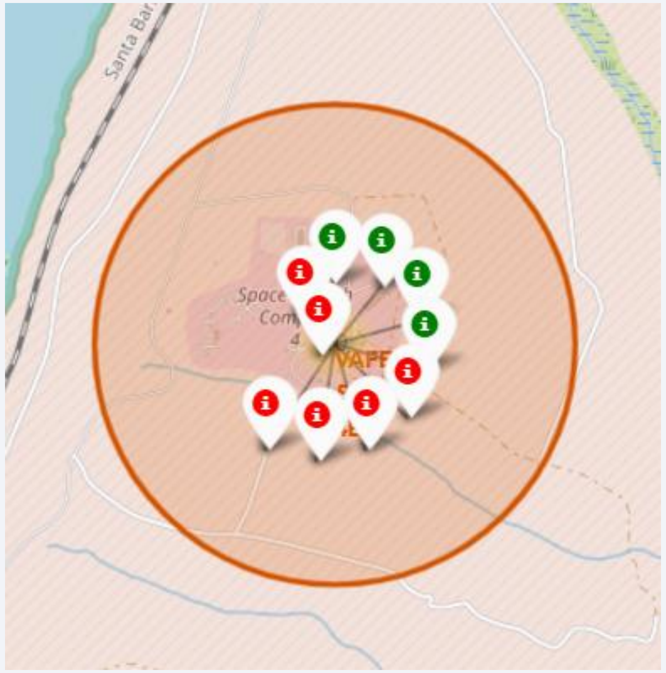
Global locations of all launch sites

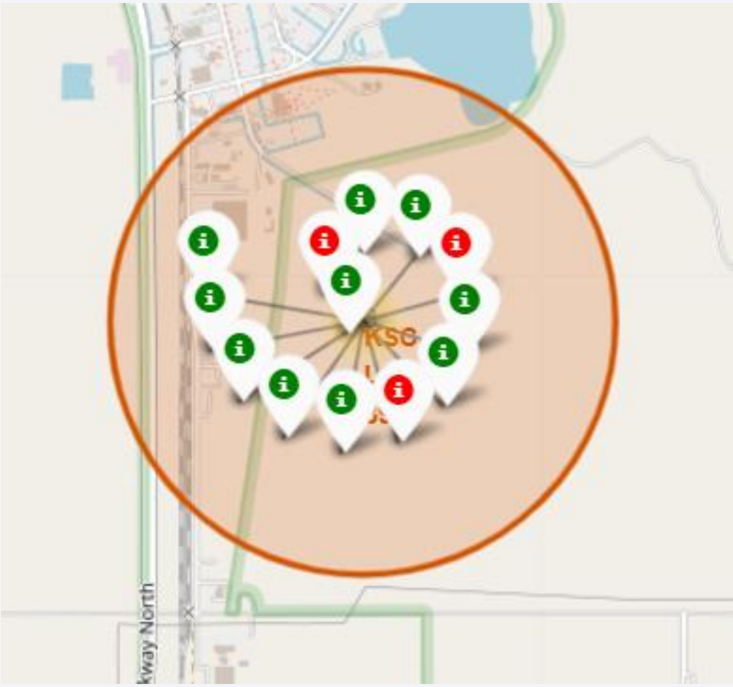


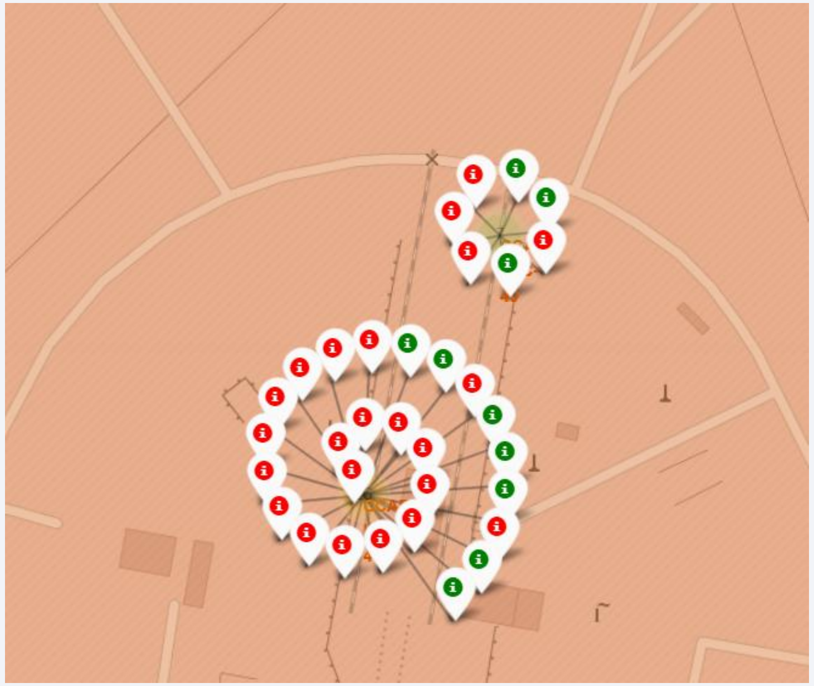




Success rates of each launch site







VAFB SLC – 4E:

10 launches

KSC LC – 39 A:

13 launches

CCAFS LC and SLC:

33 launches

Launch site to its proximities

Site:

CCAFS SLC

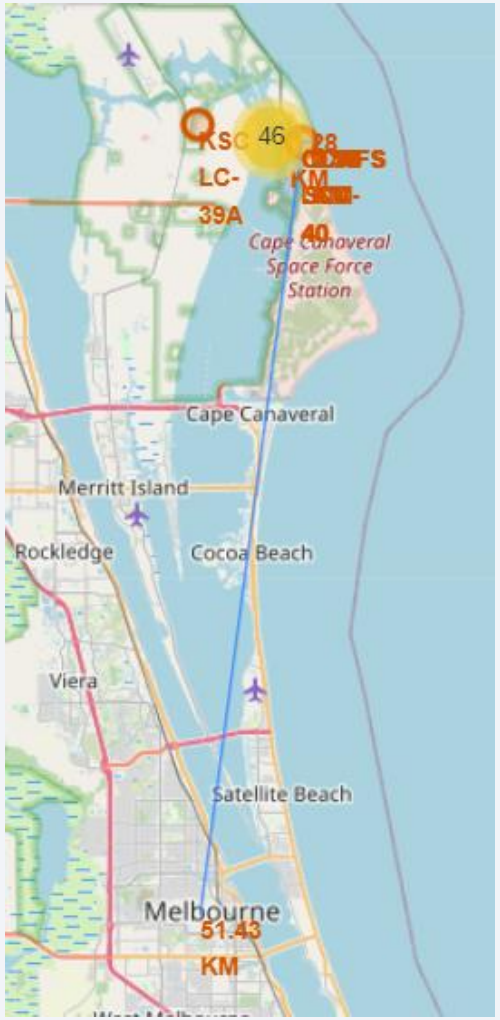
Nearest proximities:

Shoreline: 0.58km

Railway: 1.28km

City: 51.43km

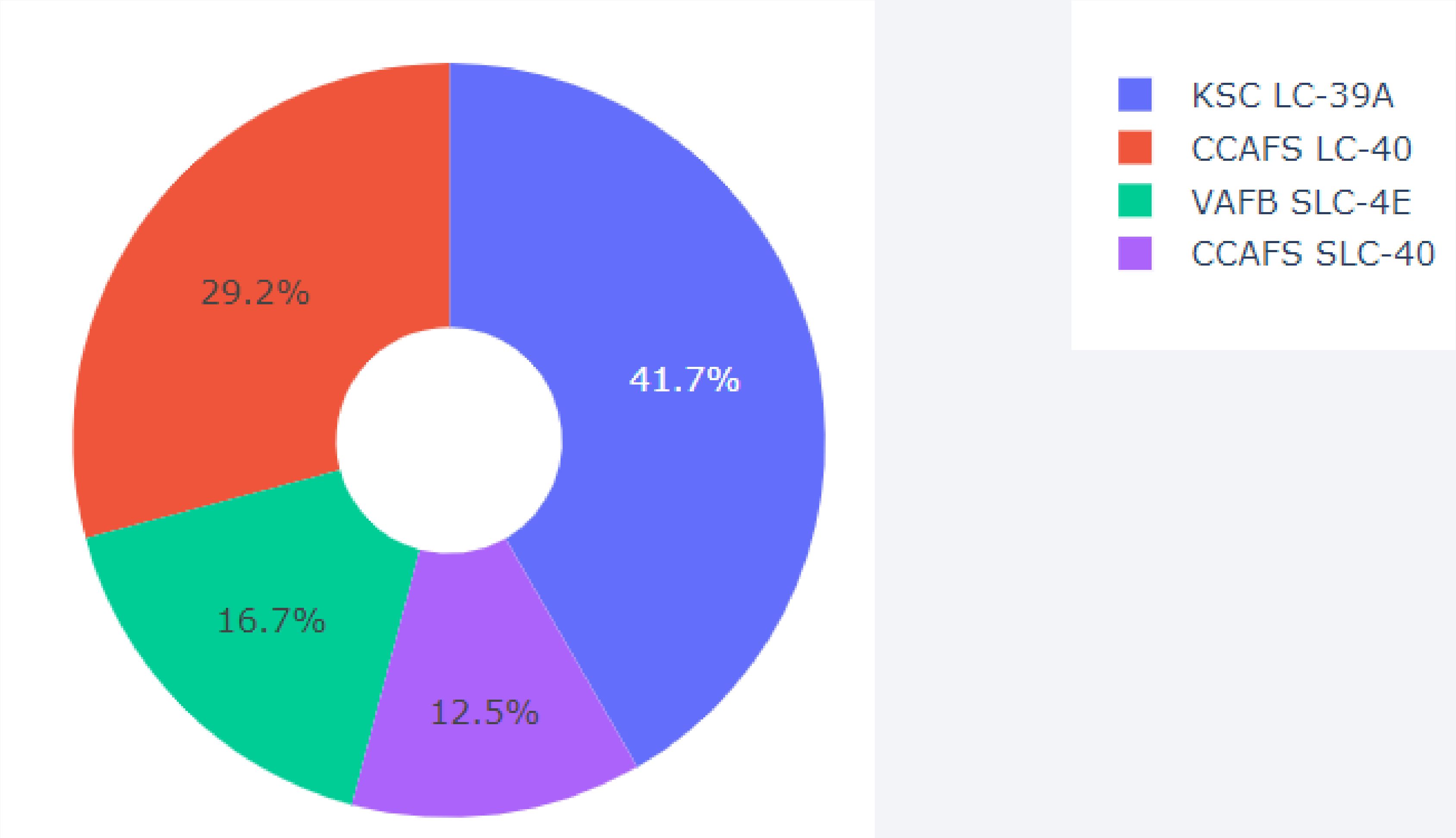
Melbourne City



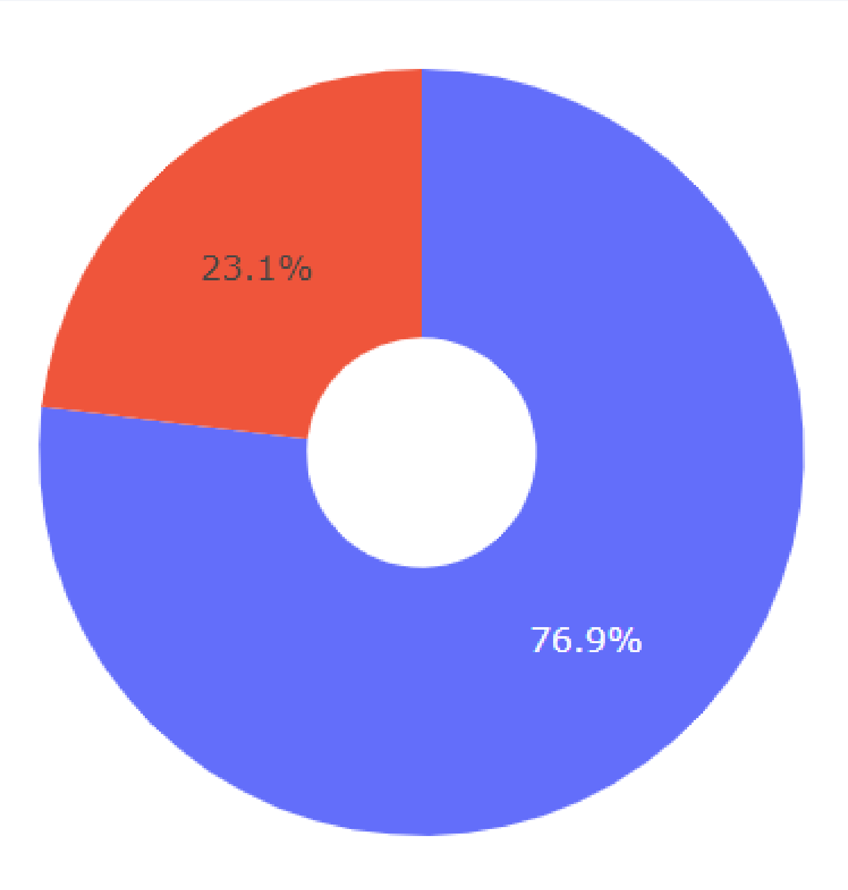
Section 4

Build a Dashboard with Plotly Dash

Launch success count for all sites



Launch site with highest launch success



Launch site:

KSC LC – 39A

Success rate 76.9%

Payload vs. Launch Outcome scatter plot for all sites

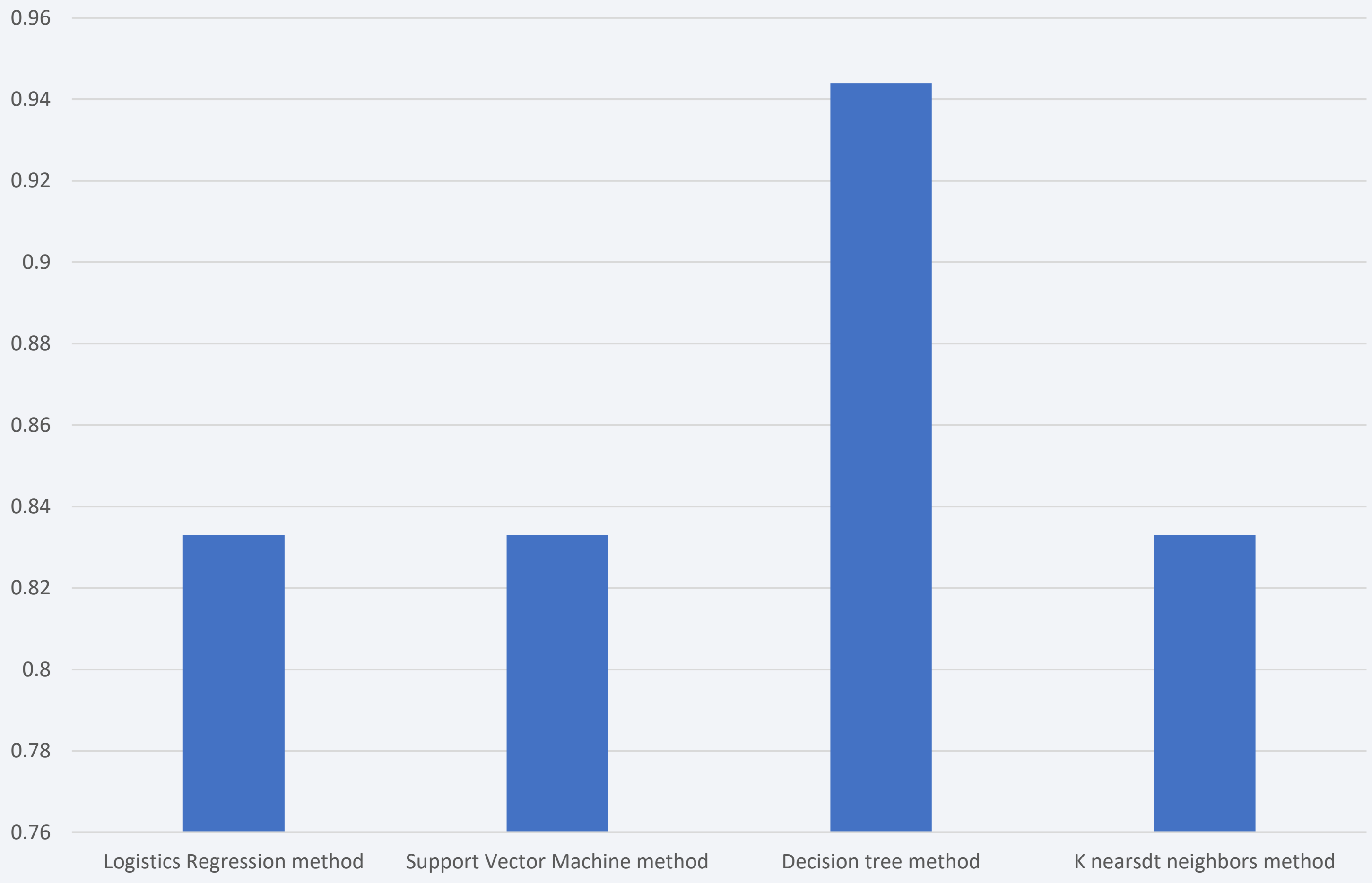
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Four methods were check for best accuracy
  - Logistics Regression method
  - Support Vector Machine method
  - Decision tree method
  - K nearest neighbors method

- Decision tree method has the highest accuracy of 94.44%



Confusion Matrix

- Confusion matrix of the decision tree method

- The only missed aspect is the three points in False Negative

Conclusions

- The success rates for SpaceX launches increases with time
- The KSC LC 39A had the most successful launches from all the sites
- The Tree Classifier Algorithm is the best for Machine Learning for this dataset
- GTO orbit has a tight payload mass requirement of 3000 – 8000kg
- Beyond 2018 the certainty of a successful launch is the best it has been

Appendix

- Please follow the link to access all Notebooks on the GitHub repository
  - https://github.com/SanthikaEvo/Python-Final-Projects.git

Thank you!