SPACE X DATA ANALYSIS AND SUCCESSFUL LANDING PREDICTION ON ROCKETS FIRS STAGE

Bernardo Solorzano

08/25/2021

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Conclusion

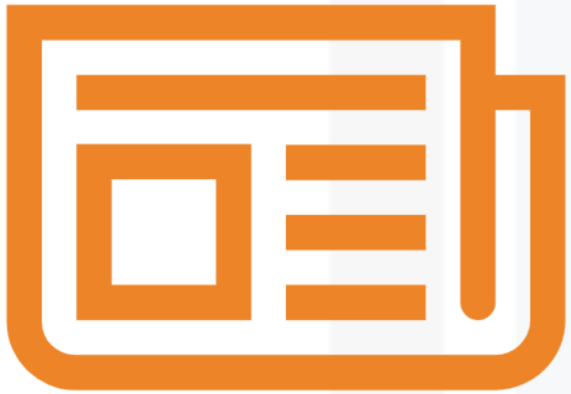
EXECUTIVE SUMMARY

- This project aims to predict if the Falcon 9 Space Rocket first stage will land successfully and to provide EDA concerning Space X rocket launches and first stage landings.
- Data on Space X rocket launches was gathered by web-scraping and API requests. It was further processed with tools available in Python programming language. EDA was provided with help of various visualization tools and SQL queries. Logistic Regression, SVM Decision Tree and KNN were trained and optimized to predict the success of first stage landing. Best performing model was chosen by accuracy score.
- EDA shows that a list of different factors affect the success of first stage landing. Over 80% accuracy was achieved for predictions. False positives rate is the point for further improvement of the predicting model based on confusion matrix.

INTRODUCTION

Space X company offers Falcon 9 rocket launches on its website with a cost of 62 million USD whereas other providers offer launches at a cost of 165 million USD. Major part of the savings is because Space X can reuse the first stage. If other company can determine if the first stage will successfully land, it can predict the cost of a launch. This information can be used to bid against Space X for a rocket launch.

METHODOLOGY (OVERVIEW)



- Data Collection methodology:
  - Data was requested from SpaceX REST API endpoints.
  - Scraped from Wikipedia Web-page.
- Data Wrangling:
  - Removing not relevant data-records.
  - Replacing missing values with average values.
  - Exploration of data-types in the given data-sets.
- Exploratory Data Analysis using visualization and SQL:
  - The dependencies between features of datasets were visualized and explored with charts and plots.
  - Additional insights in the provided datasets were made with help of SQL queries.
  - Features for prediction were prepared based on the results of visual analysis.
- Interactive visual analytics using Folium and Plotly Dash.
- Predictive analysis using classification models:
  - Features were standarized.
  - Logistic Regression, SVM, Decision tree and KNN methods were used for predictive models.

EDA WITH DATA VISUALIZATION

Name of Chart	Type of Chart	Purpose of Chart
Landing outcome for Flight Number vs. Payload Mass	Scatter	To check if landing success rate increase with for later flights and to see if success rate is higher for higher payload mass.
Landing outcome for Flight Number vs. Launch Site	Scatter	To check the distribution of launches between the launch sites in time. To check the change of success rate from earlier launches to later launches for each launch site.
Landing outcome for Payload Mass vs. Launch Site	Scatter	To check the distribution of launches with different payload mass between the launch sites. To check the payload mass range that has high and low success rates.
Success rate for Orbit Type	Bar	To see if different orbits have different success rates.
Landing outcome for Flight Number vs. Orbit Type	Scatter	To check the change of success rate from earlier launches to later launches for each orbit type.
Landing outcome for Payload Mass vs. Orbit Type	Scatter	To check the influence of payload mass on success rate for each orbit type.
Launch Success Yearly Trend	Line	To check the change of success rate from 2013 till 2020

*Charts in further slides.

EDA WITH SQL

Performed sql queries:

- Display the names of the unique launch sites.
- 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.
- List the names of the boosters which have succeded in dron 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_versiones which have carried the maximun payload mass with help of subquery.
- Rank the count of succesful landing outcomes betweeen 2010-06-04 and 2017-03-20 in descending order.

*Queries results in further slides.

INTERACTIVE MAP WITH FOLIUM

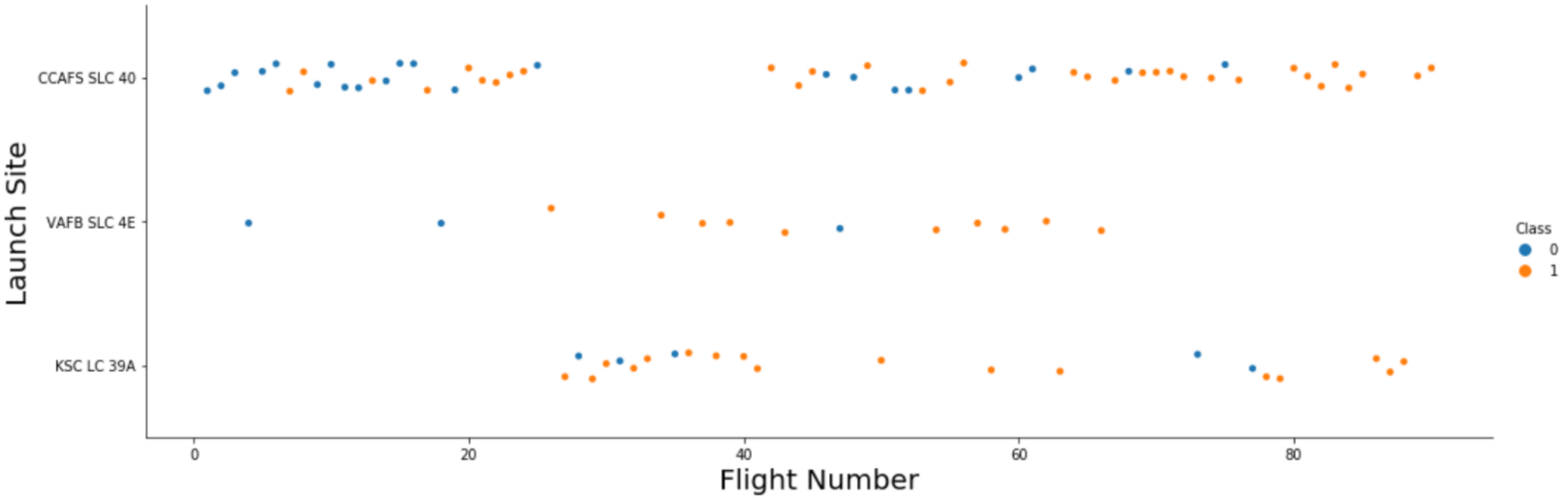
Map Objects Added to Map	Type of Map Object	Purpose of Object
NASA Johnson Space Center Markers	Circle, Popup Label, Text Label	To show the location of NASA command centre on the map.
Markers for Every Launch Site	Circle, Popup Label, Text Label	To show the location of every launch site on the map.
Markers of success/failed launches for each launch site	Color-Labeled Marker	To identify success rate for every launch site.
Lines to show distance to nearest railway, city, coast, highway	Polyline, Text Label	To measure the distance to the nearest railway, highway, city, coast.

DASHBOARD WITH PLOTLY DASH

Created Dashboard objects	Purpose of Object
Launch Site Dropdown List	To enable interactive Launch Site selection for charts
Pie Chart of Successful Launches	Shows the total successful launches count for all sites if all Launch Sites are selected. If a specific launch site is selected, the pie chart shows the Success vs. Failed counts for the site.
Slider of Payload Mass Range	To select the range of payload mass for charts where payload mass is used
Scatter Chart: Booster version for Success Rate vs. Payload Mass	To show the correlation between payload and launch success for each booster version

EDA WITH VISUALIZATION RESULTS

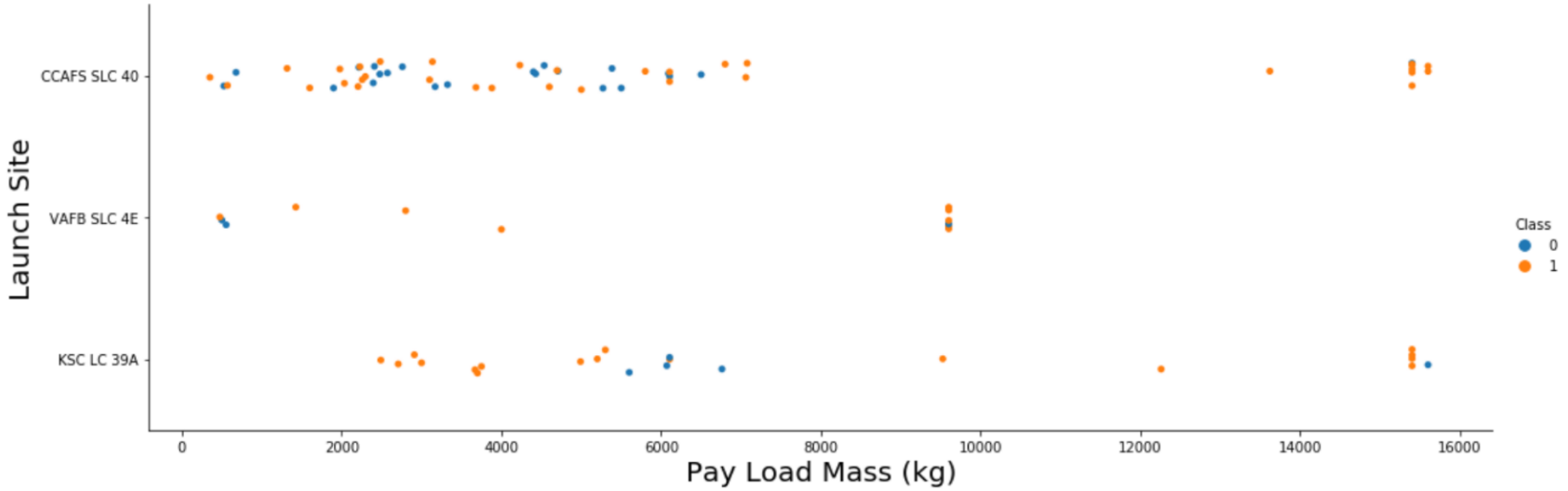
FLIGHT NUMBER VS. LAUNCH SITE



Conclusions:

- CCAFS SLC 40 launch site is used the most.
- CCAFS SLC 40 launch site has the highest number of failed launches at the beginning.
- Starting with Flight Number 78 all launches on launch sites were successful.

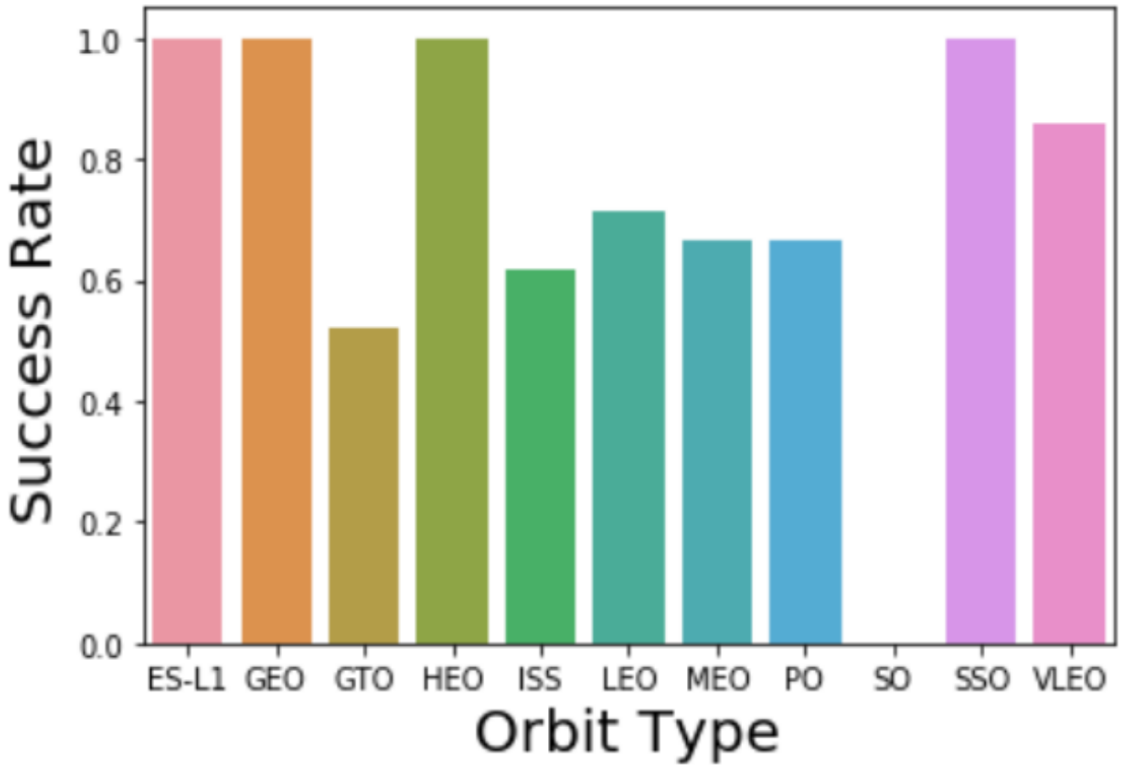
PAYLOAD VS LAUNCH SITE



Conclusions:

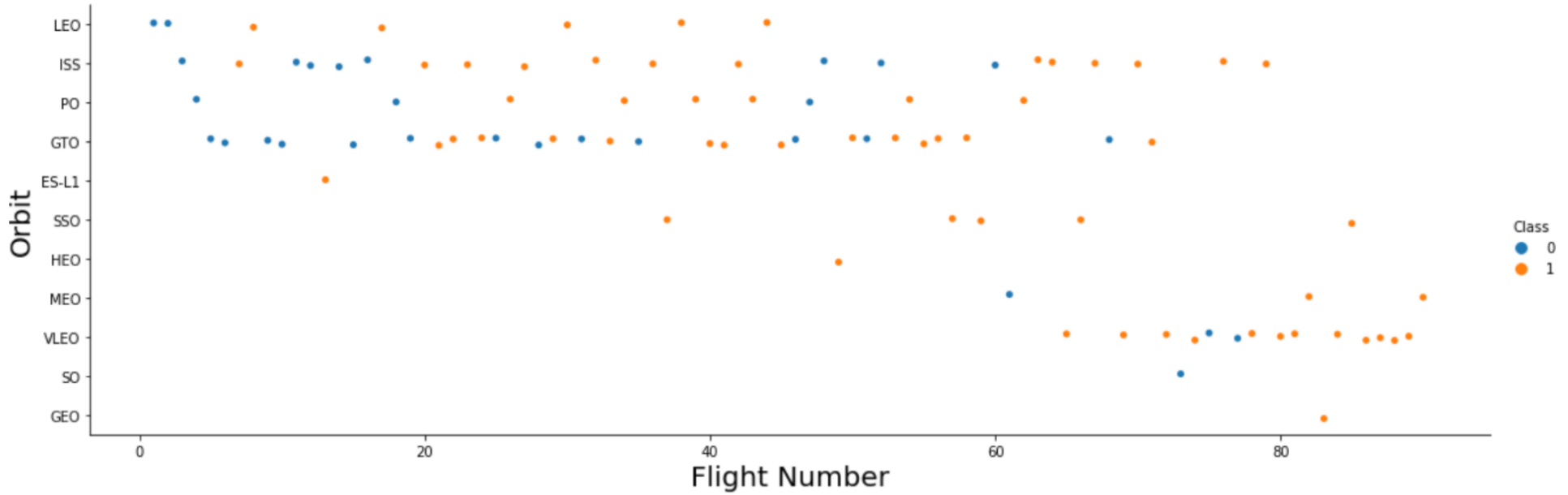
- For every launch site , the higher the payload and mass are, the higher is the success rate.
- KSC LC 39A launch site has the highest general success rate, but it has problems for payload mass in range from 5000 to 7000 kg.
- Most of unsuccessful launches had payload mass under 7000kg.

SUCCESS RATE VS ORBIT TYPE



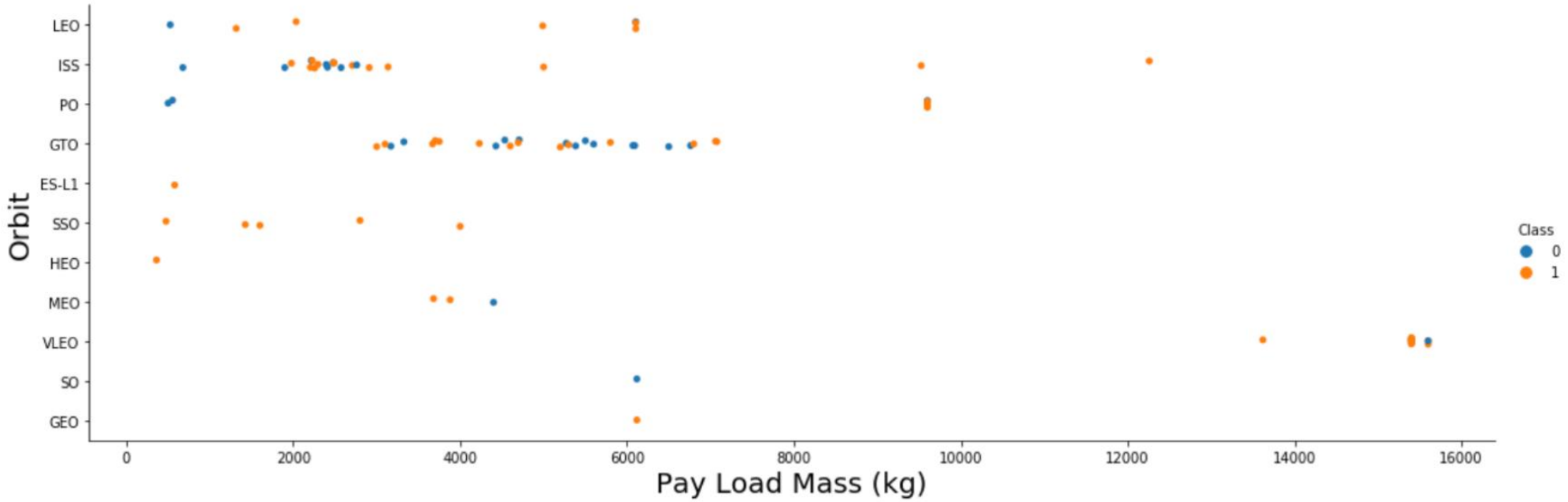
- ES-L1, GEO, HEO and SSO have 100% success rate.
- SO has 0% success rate.
- VLEO has success rate above 80%.
- GTO, ISS, LEO, MEO, PO have success rate in range from 50% to 80%.

FLIGHT NUMBER VS ORBIT TYPE



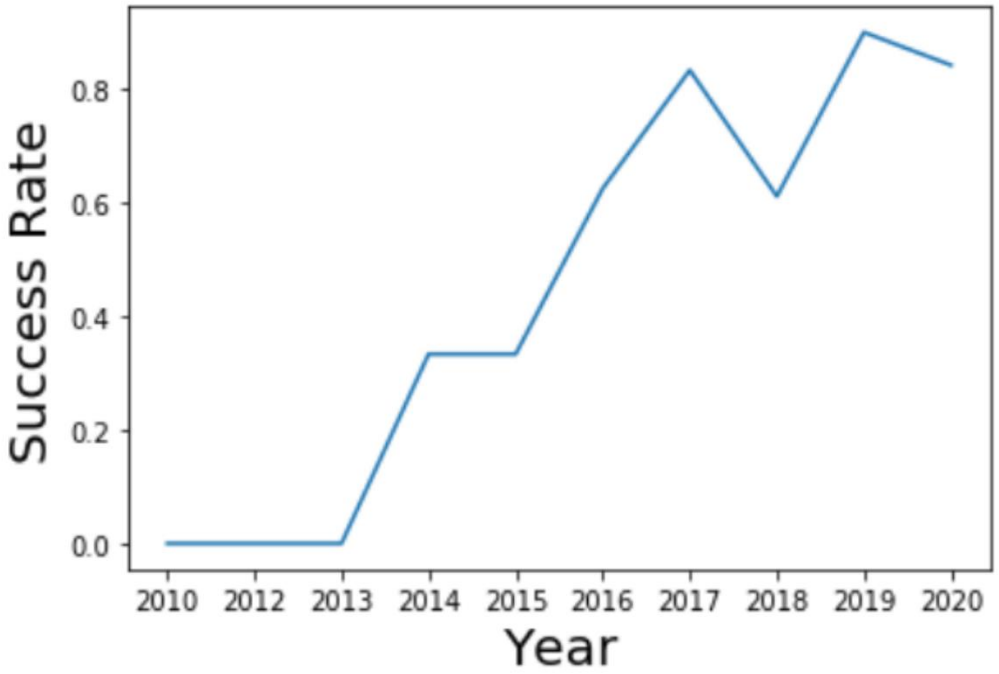
- In the LEO orbit the success is related to the number of flights.
- No relationship between flight number for GTO orbit.
- LAunches to VLEO orbit were performed late: after flight #60.

PAYLOAD VS ORBIT TYPE



- Heavy payloads have a negative influence on GTO orbits.
- Heavy payloads have a positive influnece on LEO and ISS orbits.

LAUNCH SUCCESS YEARLY TREND



The success rate since 2013 kept increasing until 2020.

EDA WITH SQL RESULTS

LAUNCHES SITE NAMES

```
In [4]: %sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXDATASET
 * ibm_db_sa://vql23019:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

Out[4]:

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

- 5 unique launch sites are used for SpaceX launches.

LAUNCHES SITE NAMES THAT BEGIN WITH "CCA"

In [14]:
```
%%sql
SELECT DISTINCT LAUNCH_SITE
FROM SPACEXDATASET
WHERE LAUNCH_SITE LIKE 'CCA%'
```

```
 * ibm_db_sa://vql23019:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

Out[14]:

launch_site
CCAFS LC-40
CCAFS SLC-40
CCAFSSLC-40

- 3 unique launch sites have names starting with "CCA".

TOTAL PAYLOAD MASS

```
In [6]: %sql SELECT SUM(payload_mass__kg_) FROM SPACEXDATASET WHERE CUSTOMER='NASA (CRS)'
 * ibm_db_sa://vql23019:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
Out[6]:
1
45596
```

- Total payload carried by boosters launched by NASA (CRS) is 45 596 kg.

AVERAGE PAYLOAD MASS BY F9 v1.1.

```
In [7]: %sql SELECT AVG(payload_mass__kg_) FROM SPACEXDATASET WHERE BOOSTER_VERSION LIKE 'F9 v1.1%'
 * ibm_db_sa://vql23019:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90108kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
Out[7]:
1
2534
```

- Total payload carried by F9 v1.1 is 2534 kg.

FIRST SUCCESSFUL GROUND LANDING DATE

```
In [8]: %sql SELECT MIN(DATE) FROM SPACEXDATASET WHERE LANDING__OUTCOME='Success (ground pad)'
 * ibm_db_sa://vql23019:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90108kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
Out[8]:
1
2015-12-22
```

- First successful landing on ground pad was achieved on December 22, 2015.

SUCCESSFULL DRONE SHIP LANDING WITH PAYLOAD BETWEEN 4000 AND 6000

In [15]:
```
%%sql
SELECT BOOSTER_VERSION
FROM SPACEXDATASET
WHERE LANDING__OUTCOME='Success (drone ship)' AND payload_mass__kg_ > 4000 AND payload_mass__kg_ < 6000
```

```
 * ibm_db_sa://vql23019:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

Out[15]:

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

- There are 4 boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000 kg.

TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES

In [16]:
```
%%sql
SELECT MISSION_OUTCOME, Count(*) AS COUNT
FROM SPACEXDATASET
GROUP BY MISSION_OUTCOME
```

```
 * ibm_db_sa://vql23019:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

Out[16]:

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

- There are 100 successful outcomes and 1 failure mission outcome.

BOOSTERS CARRYING MAXIMUN PAYLOAD

In [11]:
```
%%sql
SELECT DISTINCT BOOSTER_VERSION
FROM SPACEXDATASET
WHERE payload_mass__kg_=(SELECT MAX(payload_mass__kg_) FROM SPACEXDATASET)
```

```
 * ibm_db_sa://vql23019:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

Out[11]:

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

- 12 boosters carried maximun payload mass.

2015 LAUNCH RECORDS

In [12]:
```
%%sql
SELECT MONTHNAME(DATE) AS MONTH, LANDING__OUTCOME, BOOSTER_VERSION, LAUNCH_SITE
FROM SPACEXDATASET
WHERE EXTRACT(YEAR FROM DATE)=2015 AND LANDING__OUTCOME='Failure (drone ship)'
```

```
 * ibm_db_sa://vql23019:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

Out[12]:

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

- Total of 2 records, one in January and one in April.

RANK SUCCESS COUNT BETWEEN 2010-06-04 AND 2017-03-20

In [13]:
```
%%sql
SELECT LANDING__OUTCOME, COUNT(*) AS OUTCOME_COUNT
FROM SPACEXDATASET
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY LANDING__OUTCOME
ORDER BY OUTCOME_COUNT DESC
```

```
 * ibm_db_sa://vql23019:***@0c77d6f2-5da9-48a9-81f8-86b520b87518.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:31198/bludb
Done.
```

Out[13]:

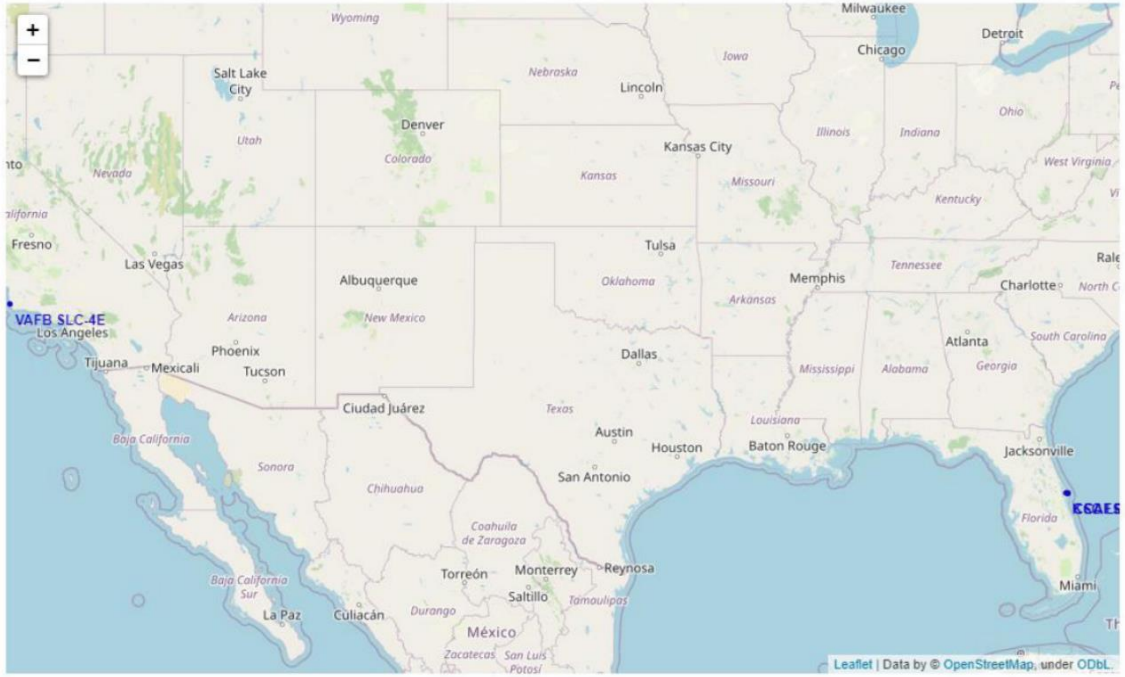
landing__outcome	outcome_count
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

- Count of success outcomes : 5 from drone ship, 3 ground pad.

INTERACTIVE MAP WITH FOLIUM

LAUNCH SITES

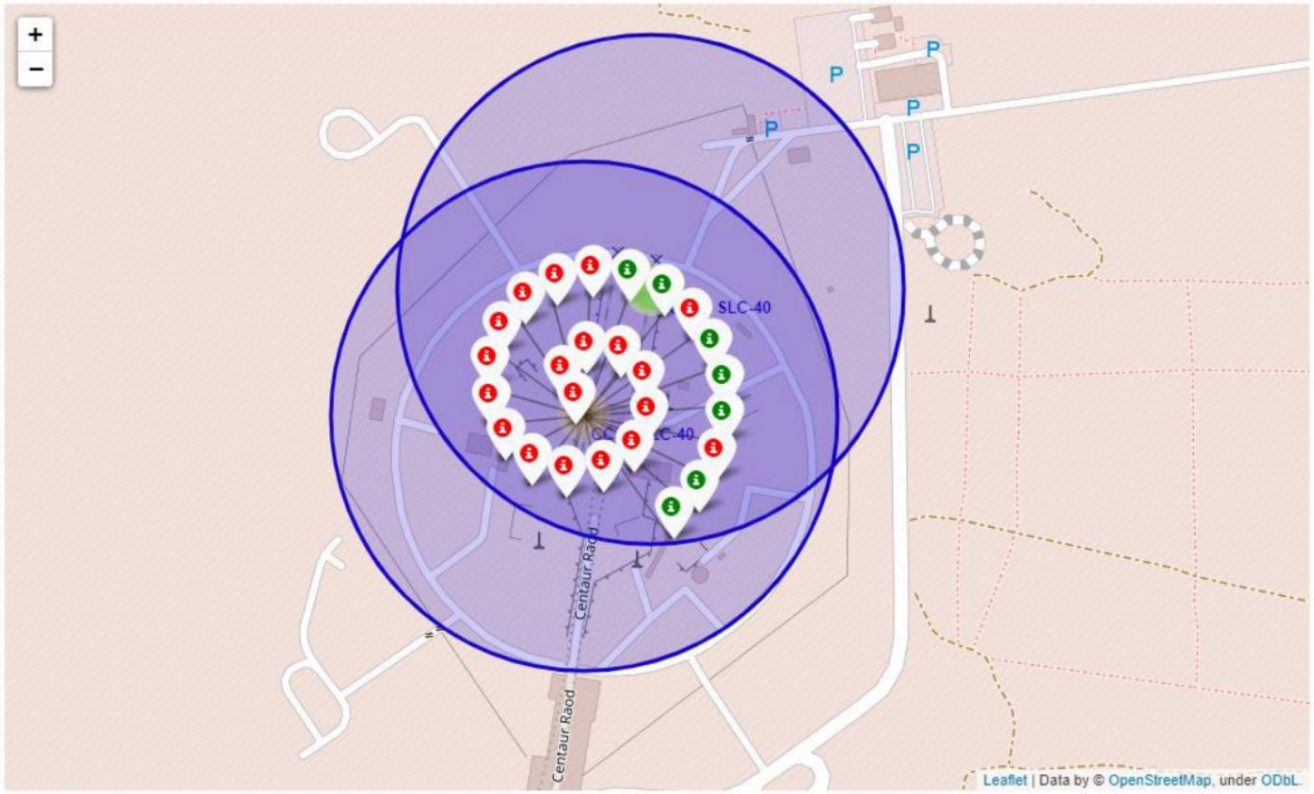
Out[76]:



- **Most of launch sites considered in this project are in proximity to the equator.** Anything on the surface of the Earth at the equator is moving faster (1670 km/hr).
- **All launch sites are in very close proximity to the coast.** Starting rockets towards the ocean helps to minimize the risk of having any debris dropping or exploding near people.

SUCCESS RATE FOR LAUNCH SITE (CCAFS LC-40)

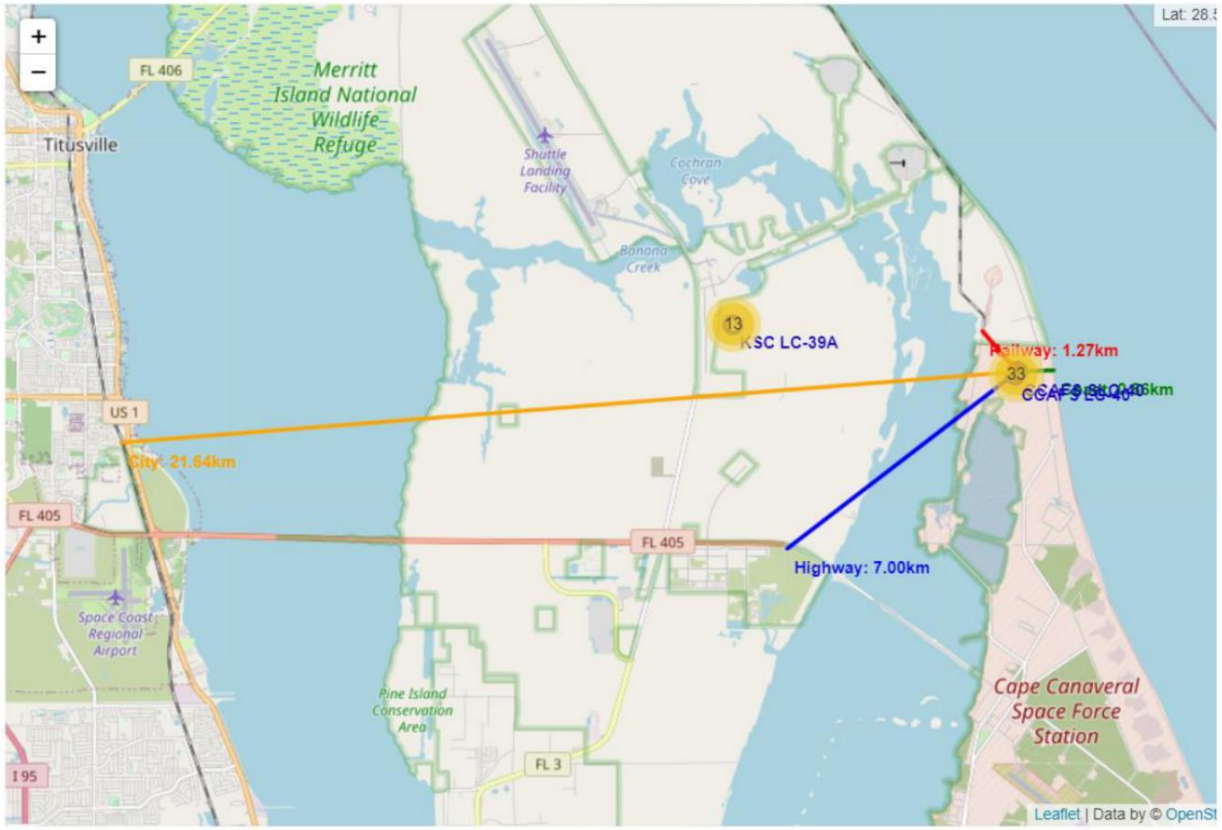
Out[82]:



- For the launch site CCAFS LC-40 the success rate is not quite good.

DISTANCE FROM LAUNCH SITE CCAFS LC-40 TO PROXIMITIES

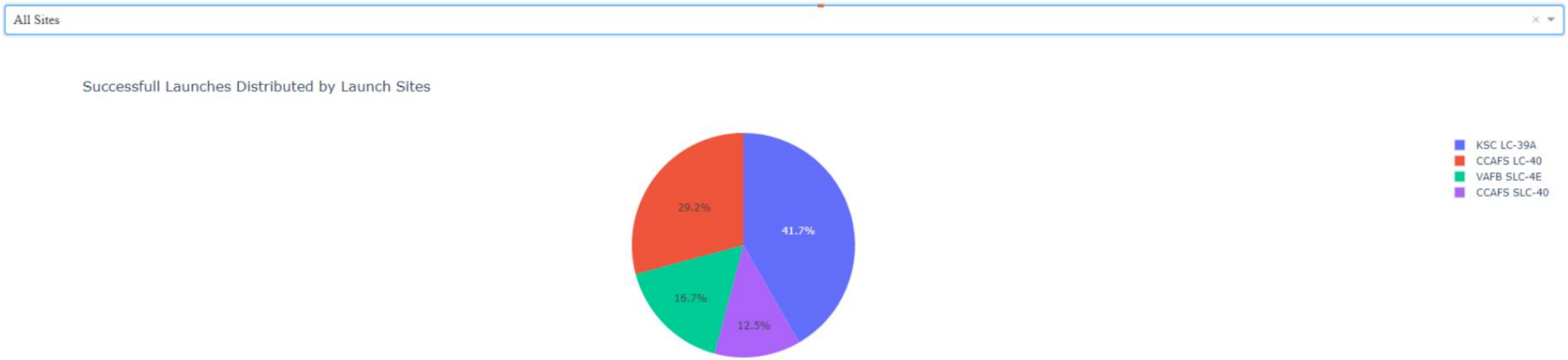
Out[88]:



- Launch sites are built close to major water bodies to ensure that no components are shed over populated areas.
- Launch site is built next to railways/highways to provide convenient transportation of space-craft part and cargos.
- A rocket launch site is built as far as possible from major population centers in order to mitigate risk to bystanders should a rocket experience a failure.

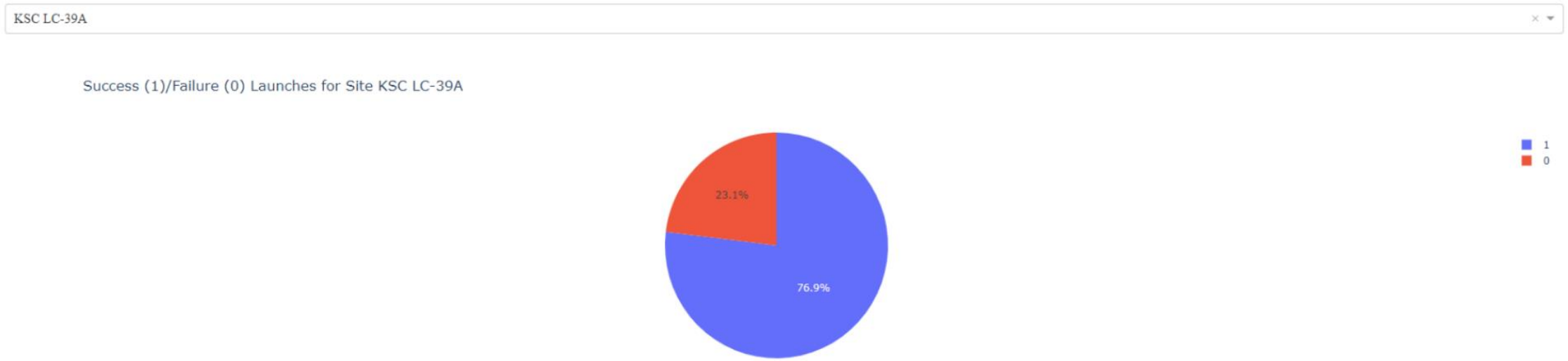
DASHBOARD WITH PLOTLY DASH

LAUNCH SUCCESS COUNT FOR ALL SITES



- Most of successful launches were made on KSC LC-39A launch site.

SUCCESS/FAILURE RATE FOR KSC LC-39A LAUNCH SITE

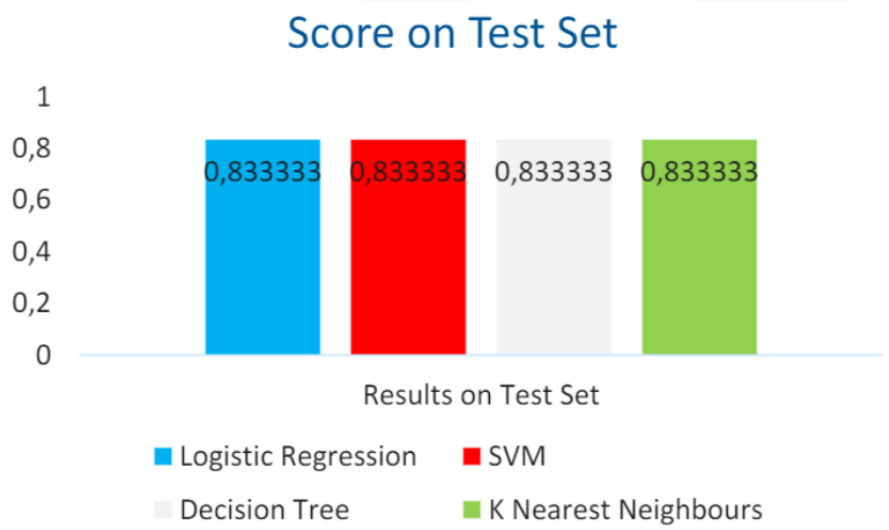


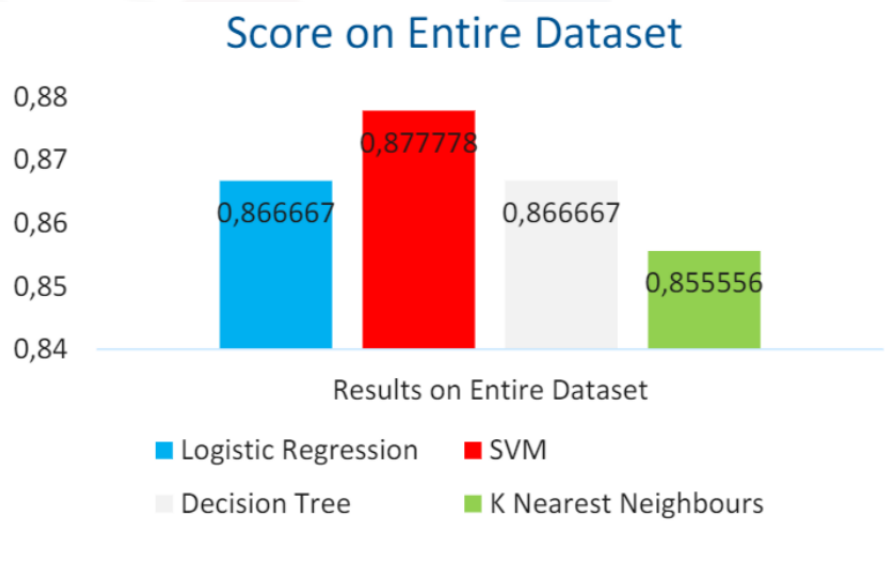
- KSC LC-39A launch site has the highest success rate (76.9%).

PAYLOAD VS LAUNCH OUTCOME (DIFFERENT PAYLOAD RANGES), ALL LAUNCH SITES

PREDICTIVE ANALYSIS (CLASSIFICATION)

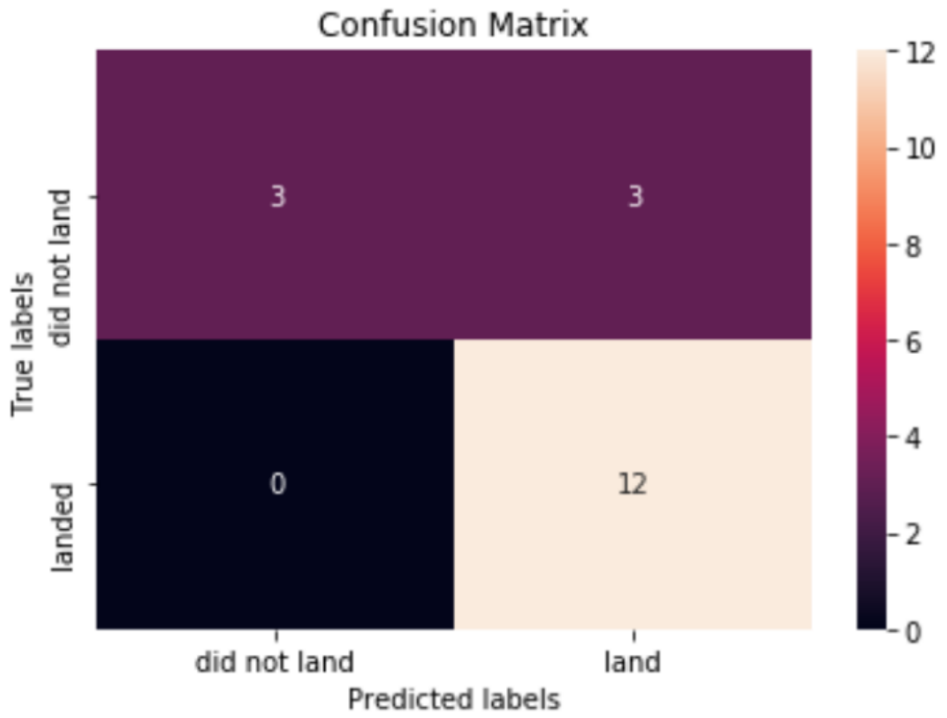
SUCCESS/FAILURE RATE FOR KSC LC-39A LAUNCH SITE





- All prediction methods showed pretty high accuracy score (over 80%).
- All prediction methods showed equal accuracy score 83.33% on test set.
- SVM method performed best when using the entire dataset.

CONFUSION MATRIX FOR SVM



Examining the confusion matrix, we see that logistic regression can distinguish between the differnet classes. We see that the major problem is false positives.

- SVM method con distinguish between the different classes.
- False positives is the point for improvement of the prediction accuracy.

CONCLUSION

- Gathered datasets provide a good bias both for predicting if the Falcon 9 Space Rocket first stage will land successfully and for EDA concerning Space X rocket launches.
- Various EDA techniques show that a list of different factors affect the success of first stage landing.
- SVM method provides best accuracy score result (on entire dataset).
- False positives rate is the point for further improvement of the predicting model based on confusion matrix analysis.