SpaceX Falcon 9 Launches

Borna Zandkarimi

IBM Data Science Professional Certificate Capstone Project

08/30/2021

Outline

- ❑ Summary
- ❑ Introduction
- ❑ Methodology
- ❑ Results
- ❑ Conclusion

Summary

- Aim: SpaceX Falcon 9 rocket landing success rate
- Exploratory Data Analysis (EDA) of first stage landing
- Data was obtained from web scraping and SpaceX open REST API
- Data processing was done using python libraries such as pandas and numpy
- SQL queries used for EDA as well
- Machine Learning methods used in this study
  - Logistic Regression, Support Vector Machine, Decision Tree, and K-Nearest Neighbors
  - Accuracy achieved for the test set prediction: 83.3%
  - All 4 models perform similarly and did not show a significant difference

Introduction

- SpaceX is a space company
  - Produces reusable Falcon 9 rockets (After landing)
  - The cost of each launch: $62 million (whereas $165 million for others!)
  - The cost of launch can be predicted based on the success rate of the first stage

- Our aim is to predict the success rate of the first stage to determine the cost of each launch

Methodology

- Data Collection
  - Web scraping from Wikipedia articles about SpaceX
  - Obtained from SpaceX open REST API

- Data Wrangling
  - Dataset was cleaned by removing the irrelevant records
  - Missing values were replaced by mean values

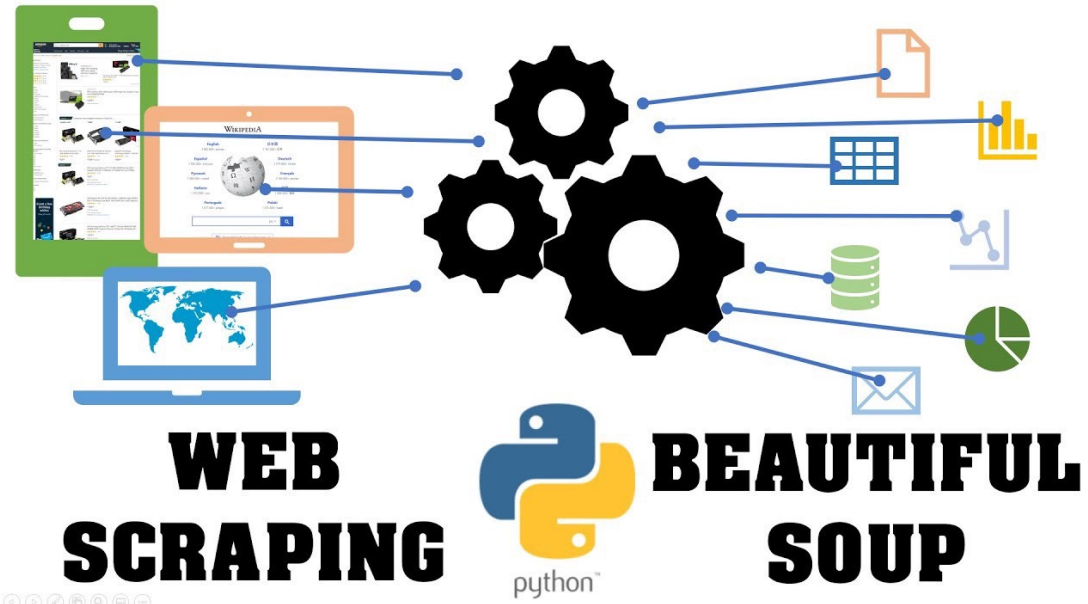
- Exploratory Data Analysis (EDA) using Python Visualization and SQL
  - Data was visualized and plotted using python visualization packages such as matplotlib and seaborn

Methodology (continued)

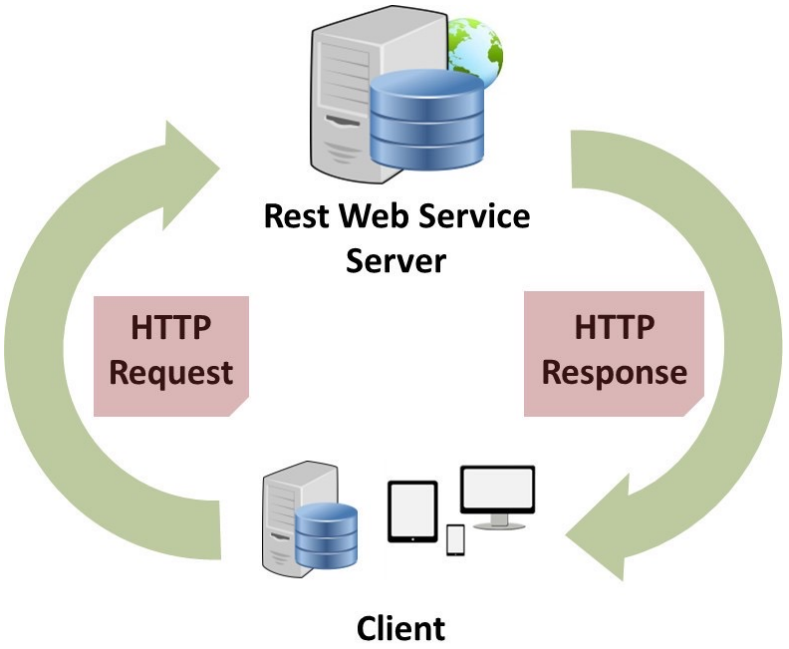
- Exploratory Data Analysis using Python Visualization and SQL (continued)
  - SQL queries were used to gain additional insight into dataset
- Interactive Visual Analytics
  - Plotly Dash was used to create a dashboard and visualize SpaceX dataset
  - Folium maps were created to plot interactive maps
- Predictive Analysis (Classification)
  - All dataset features were standardized and then fed into 4 different ML models
  - Logistic Regression, Support Vector Machine, Decision Tree, and K-Nearest Neighbors were used in order to predict the success rate

Data Collection

- **BeautifulSoup** was used to process the HTML retrieved from Wikipedia

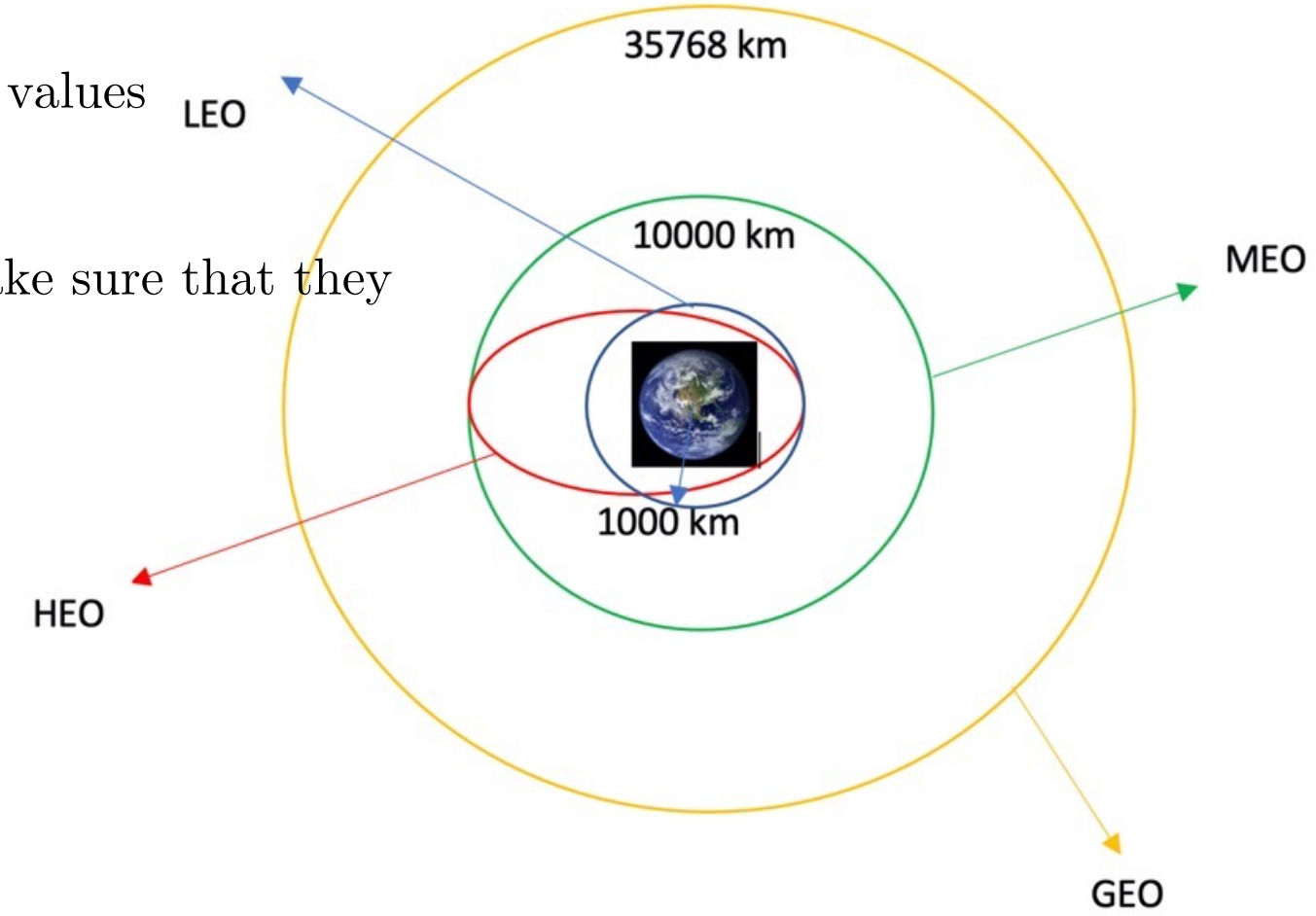


- Get requests used from SpaceX REST API to obtain data



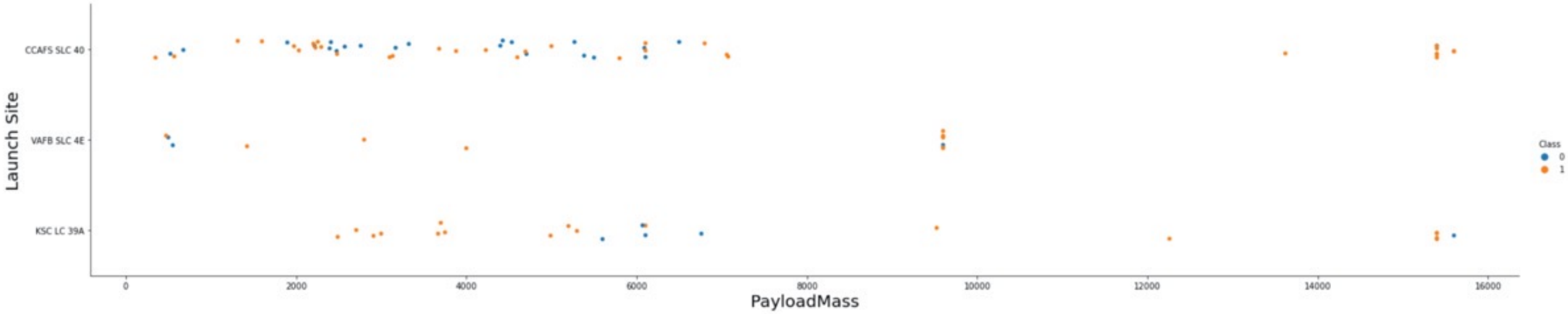
Data Wrangling

- Dataset was cleaned by removing the irrelevant records
- Missing values were replaced by mean values
- Data types were double-checked to make sure that they are consistent and easy to work with



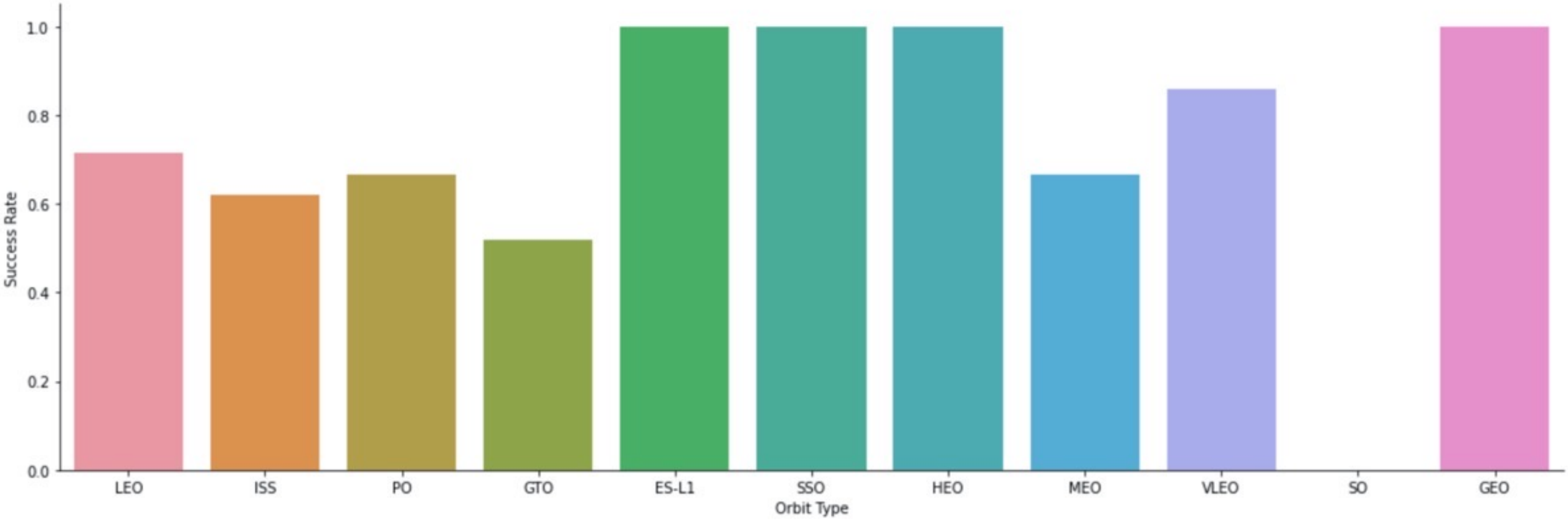
EDA with Visualization

- For KSC launch site when the payload mass is low or very high the success rate increases. For VAFB launch site as the payload mass increases the success rate increases as well. For CCAFS launch site we have a mixed success rate for low payload masses.



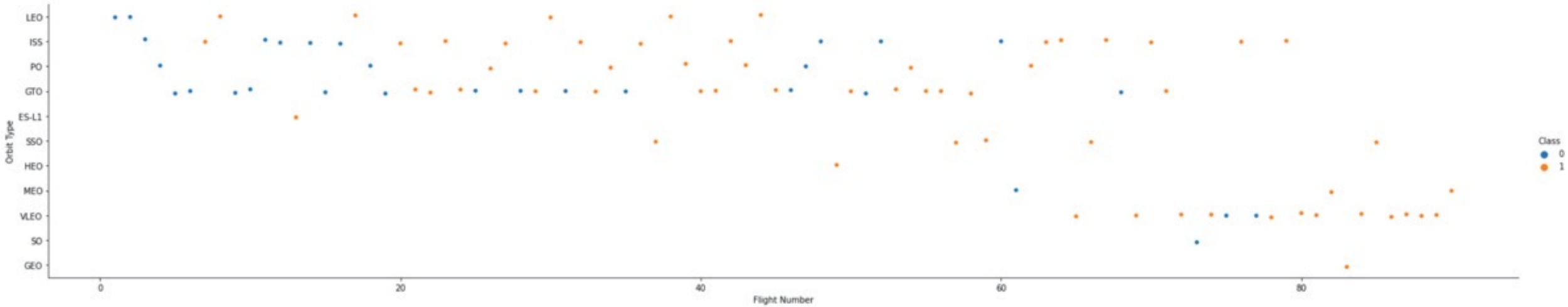
EDA with Visualization

- Orbit types ES-L1, SSO, and GEO have the highest success rate.



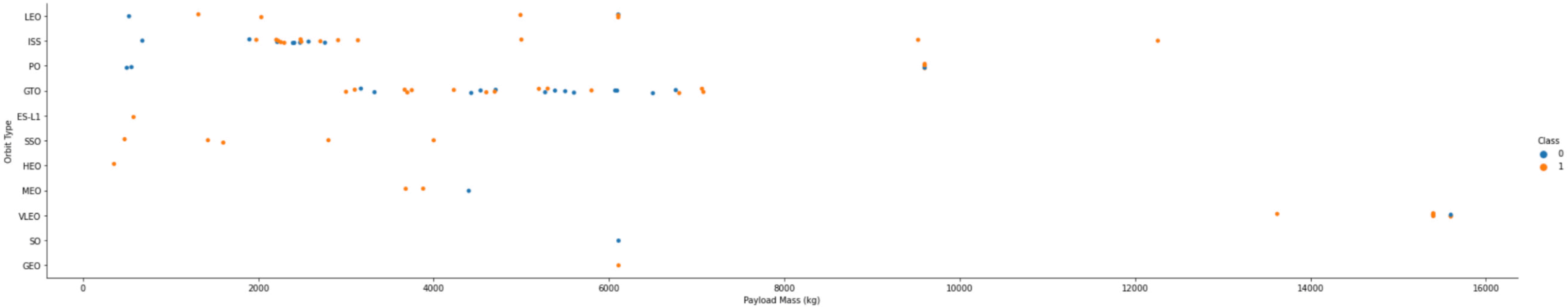
EDA with Visualization

- In the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.



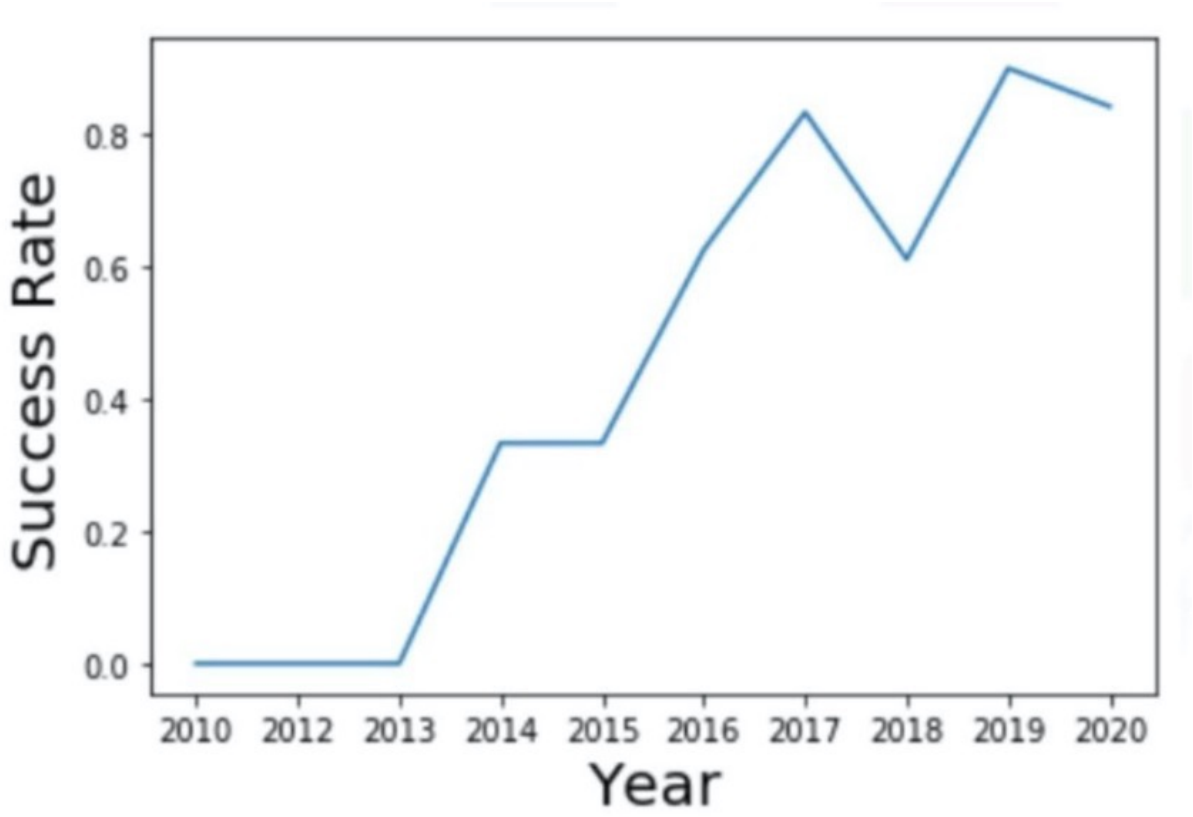
EDA with Visualization

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



EDA with Visualization

- The success rate since 2013 kept increasing till 2020



EDA with SQL

- Here are the results from the notebook

Task 1

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

```
%%sql
Select distinct Launch_Site from SPACEXTBL
```

```
 * ibm_db_sa://nmf32271:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb
Done.
```

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

EDA with SQL

Task 2

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

```
%%sql
Select * from SPACEXTBL where Launch_Site like 'CCA%' limit 5;
```

```
 * ibm_db_sa://nmf32271:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb
Done.
```

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

EDA with SQL

Task 3

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

```
%%sql
Select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where SPACEXTBL.Customer in 'NASA (CRS)' ;
```

```
 * ibm_db_sa://nmf32271:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb
Done.
```

1
45596

Task 4

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

```
%%sql
Select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where SPACEXTBL.Booster_Version in 'F9 v1.1' ;
```

```
 * ibm_db_sa://nmf32271:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb
Done.
```

1
2928

EDA with SQL

Task 5

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

Hint:Use min function

```
%%sql
Select min(Date) from SPACEXTBL where SPACEXTBL.Landing__Outcome in 'Success (ground pad)' ;
```

```
 * ibm_db_sa://nmf32271:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb
Done.
```

1
2015-12-22

EDA with SQL

Task 6

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

```
%%sql
Select booster_version from SPACEXTBL
where SPACEXTBL.Landing__Outcome in 'Success (drone ship)'  and SPACEXTBL.payload_mass__kg_ between 4000 and 6000 ;
```

```
 * ibm_db_sa://nmf32271:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb
Done.
```

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

EDA with SQL

Task 7

List the total number of successful and failure mission outcomes ¶

```
%%sql
Select count(mission_outcome) from SPACEXTBL where mission_outcome in ('Success', 'Failure')
```

```
 * ibm_db_sa://nmf32271:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb
Done.
```

1
99

EDA with SQL

Task 8

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

```
%%sql
Select booster_version from SPACEXTBL where
payload_mass__kg_ = (Select max(payload_mass__kg_) from SPACEXTBL)
```

```
 * ibm_db_sa://nmf32271:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb
Done.
```

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

EDA with SQL

Task 9

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

```
%%sql
Select SPACEXTBL.landing__outcome, SPACEXTBL.booster_version, SPACEXTBL.launch_site from SPACEXTBL
where SPACEXTBL.landing__outcome in ('Failure (drone ship)') and year(SPACEXTBL.Date) = 2015;
```

```
 * ibm_db_sa://nmf32271:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb
Done.
```

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

EDA with SQL

Task 10

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 ¶

```
%%sql
SELECT landing__outcome,COUNT(landing__outcome) as OccurenceValue FROM SPACEXTBL
where SPACEXTBL.Date between '2010-06-04' and '2017-03-20'
GROUP BY landing__outcome ORDER BY OccurenceValue desc;
```

```
 * ibm_db_sa://nmf32271:***@54a2f15b-5c0f-46df-8954-7e38e612c2bd.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32733/bludb
Done.
```

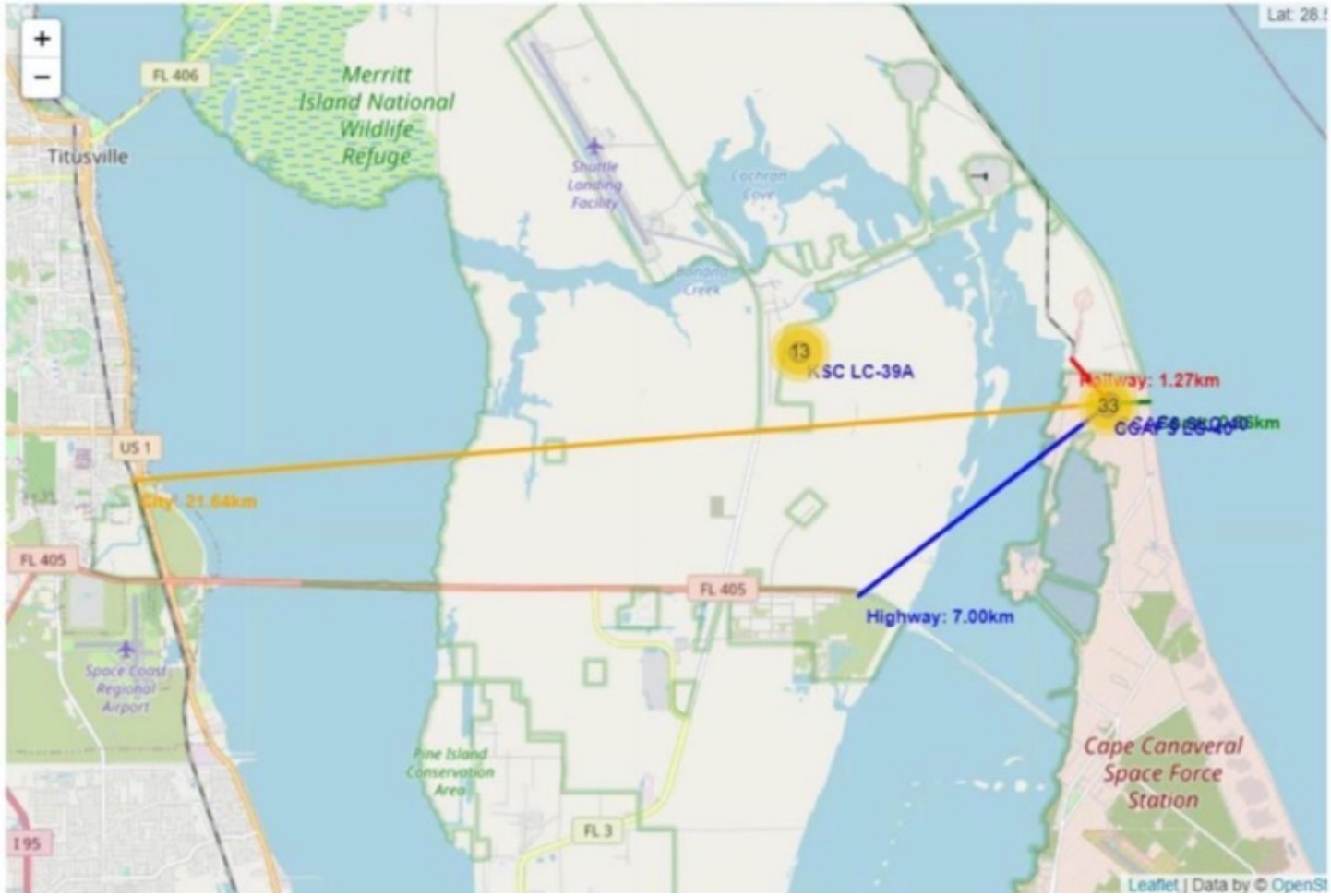
landing__outcome	occurencevalue
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Interactive Visual Analytics

- SpaceX Launch sites

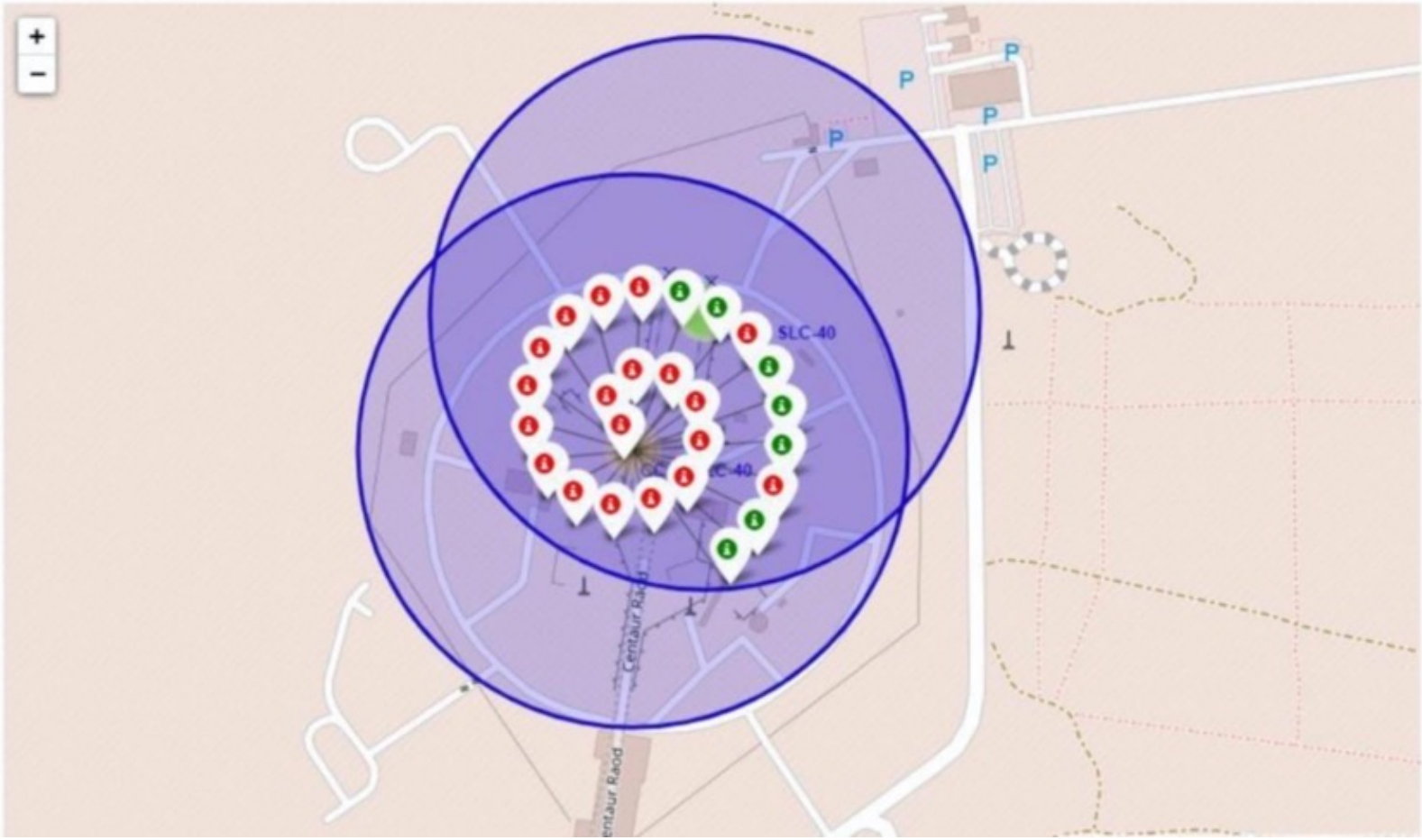
Interactive Visual Analytics

- ❑ Locations of CCAFS LC-40
- ❑ Near water (minimizes risk of destruction)
- ❑ Far from city centers



Interactive Visual Analytics

- Success rate for CCAFS LC-40 are mainly unsuccessful



Interactive Visual Analytics

- Successful launches distributed by launch sites
- Most successful rate (41.7%) belongs to the KSC LC-39A site

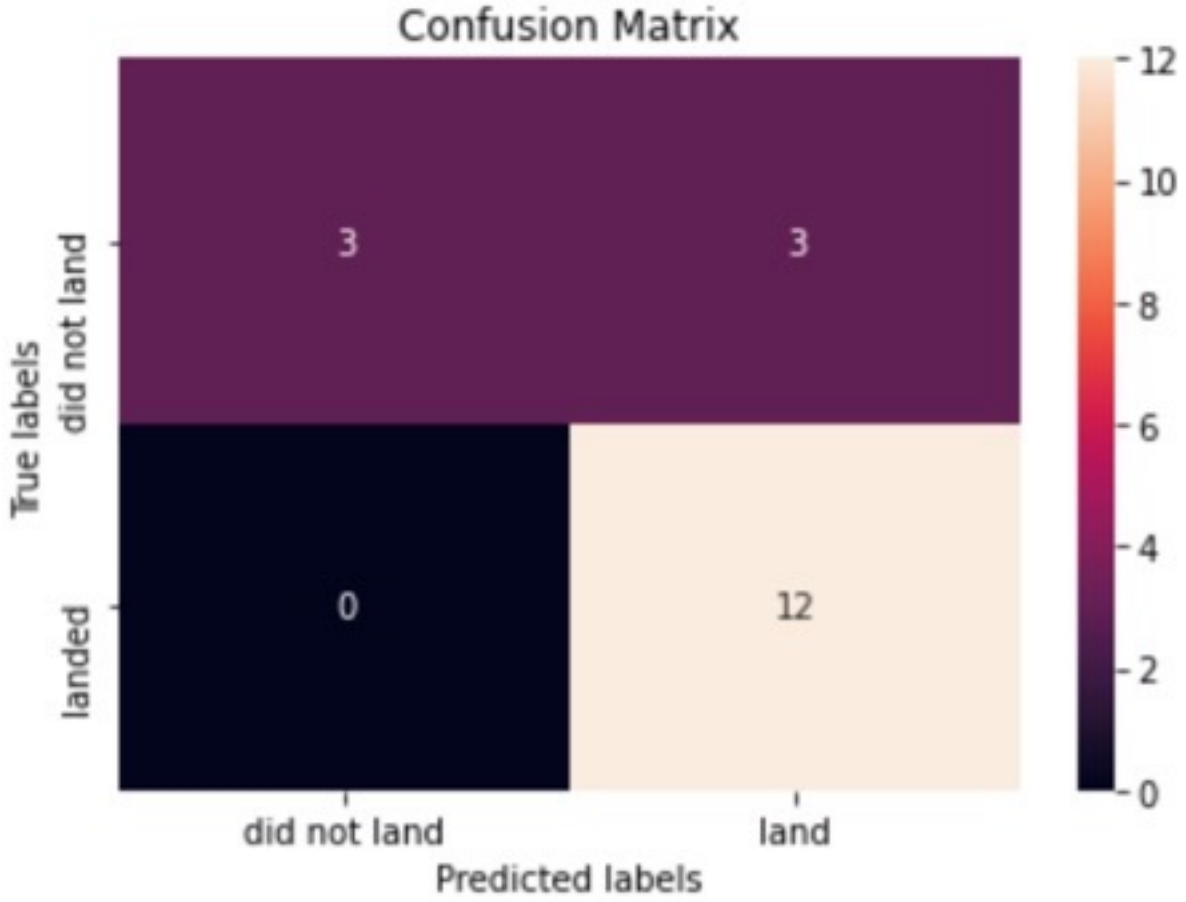
Interactive Visual Analytics

- Success(1)/Failure(0) launches for site KSC LC-39A

Predictive Analysis (Classification)

- All 4 models (LogReg, SVM, Tree, and KNN) show similar behavior and performance.
- The false positive part of the confusion matrix is the main source of error

	LogisticReg	SVM	Tree	KNN
F1_Score	0.814815	0.814815	0.814815	0.814815
Accuracy	0.833333	0.833333	0.833333	0.833333



Conclusions

- Orbit type is an important factor in determining the success rate.
- Depending on the launch site, the payload mass can positively or negatively affect the success rate.
- The success rate since 2013 kept increasing till 2020.
- Most successful rate belongs to the KSC LC-39A site.
- All 4 ML models (LogReg, SVM, Tree, and KNN) show similar accuracy and performance.
- The false positive part of the confusion matrix is the main source of error.