Winning Space Race with Data Science

Denis Bee
September 26, 2021

Outline

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

Executive Summary

- Space X API and Web scraping was used to obtain Falcon 9 landing data.
- A somehow reliable model was constructed to predict the landing outcome based on input params such as launch site, payload, launch orbit, etc.

Introduction

- In order to build a successful competition, we are going to explore Space X historical launch data in order to make predictions of the first stage reusal for a subsequent launches and, consequently, its impact on the price.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Space X official API and Web scraping.
- 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

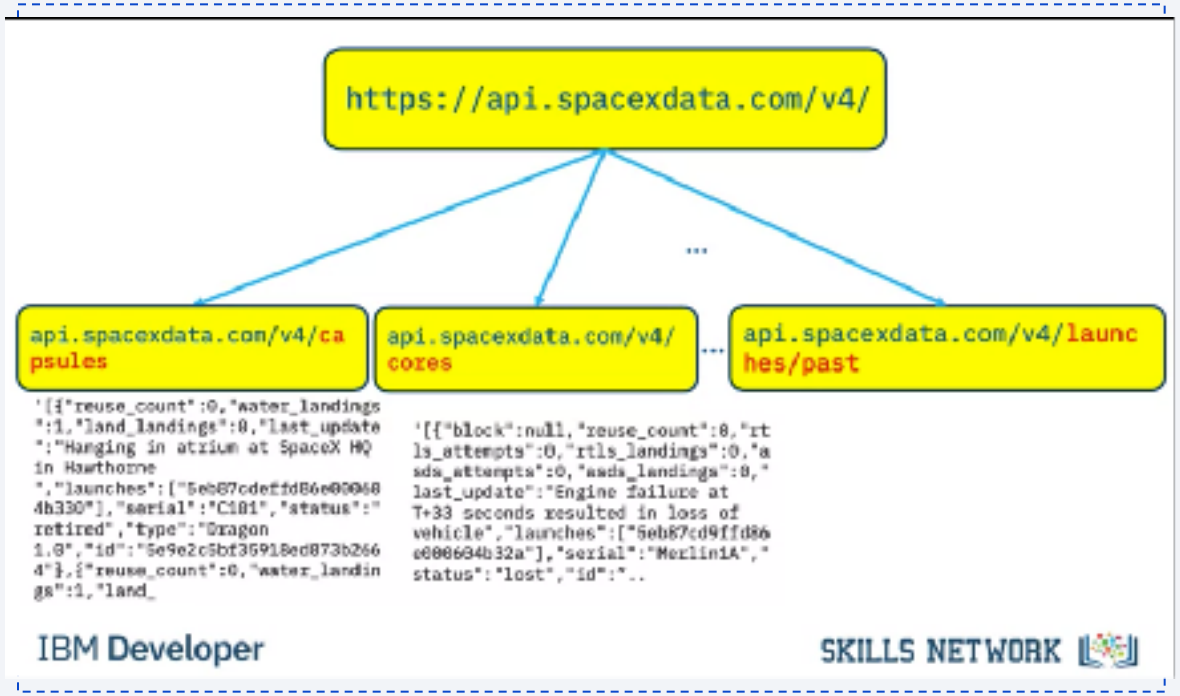
Two methods of data collection were used:

Official Space X API and Web Scraping (using data from Wikipedia).

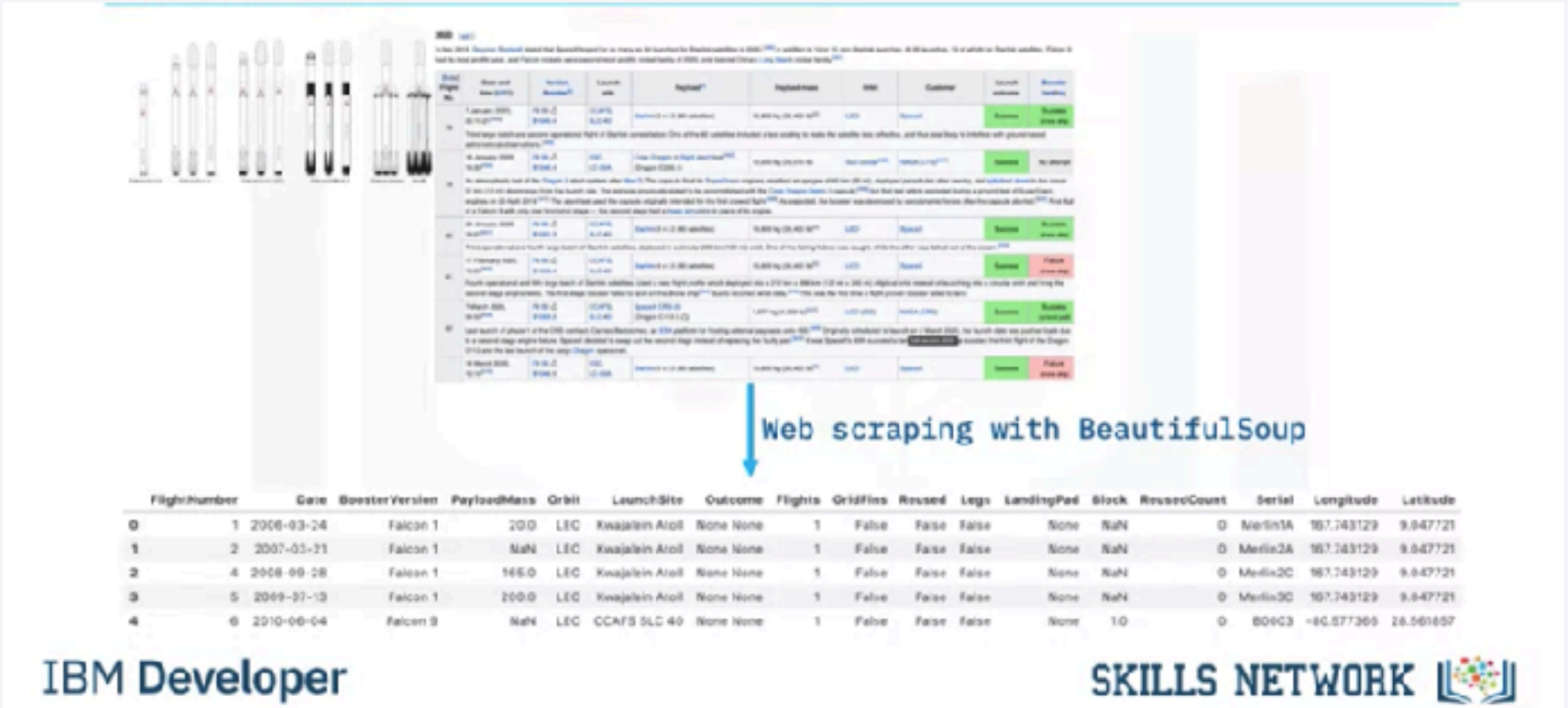
Data Collection – SpaceX API

`https://api.spacexdata.com/v4/` was used to query the SpaceX API.

https://github.com/hifv/ibm-ds-capstone/blob/master/M1L2%20Space%20X%20Data%20Collection%20API.ipynb



Data Collection - Scraping



	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin1A	167.743129	9.047721
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2A	167.743129	9.047721
2	4	2008-08-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2C	167.743129	9.047721
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin3C	167.743129	9.047721
4	6	2010-06-04	Falcon 9	NaN	LEO	CCAFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857

https://github.com/hifv/ibm-ds-capstone/blob/master/M2L1%20SQL%20Exploratory%20Data%20Analysis.ipynb

Data Wrangling

- An importance of orbit was explored.
- A Successful/Unsuccessful classification of outcomes was added

- https://github.com/hifv/ibm-ds-capstone/blob/master/M1L3%20Data%20Wrangling.ipynb

EDA with Data Visualization

- Scatter plots, bar plots and line plots were used to evaluate the correlation and importance of the features
- https://github.com/hifv/ibm-ds-capstone/blob/master/M2L2%20EDA%20Dataviz.ipynb

EDA with SQL

- https://github.com/hifv/ibm-ds-capstone/blob/master/M2L1%20SQL%20Exploratory%20Data%20Analysis.ipynb

Build an Interactive Map with Folium

- Launch sites were mapped with Folium
- https://github.com/hifv/ibm-ds-capstone/blob/master/M3L1%20Mapping%20locations%20with%20Folium.ipynb

Predictive Analysis (Classification)

- Logistic regression, Support Vector Machine, Decision Tree, and kNN models were explored for Launch outcome classification prediction
- https://github.com/hifv/ibm-ds-capstone/blob/master/M4L1%20%20Machine%20Learning%20Prediction.ipynb

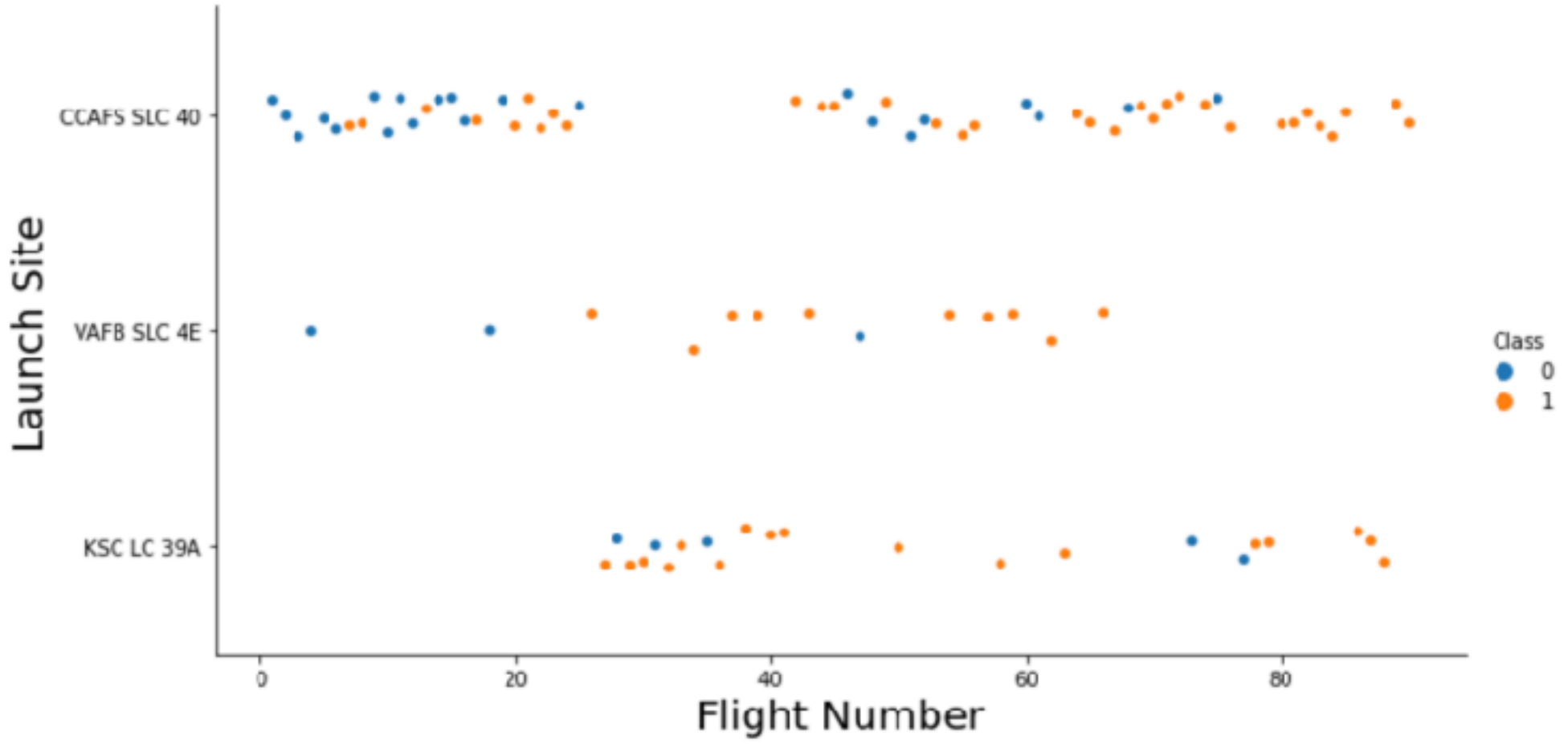
Results

Decision three showed the most promise as a classifier model, but in the end all evaluated models struggled with false positives on test data.

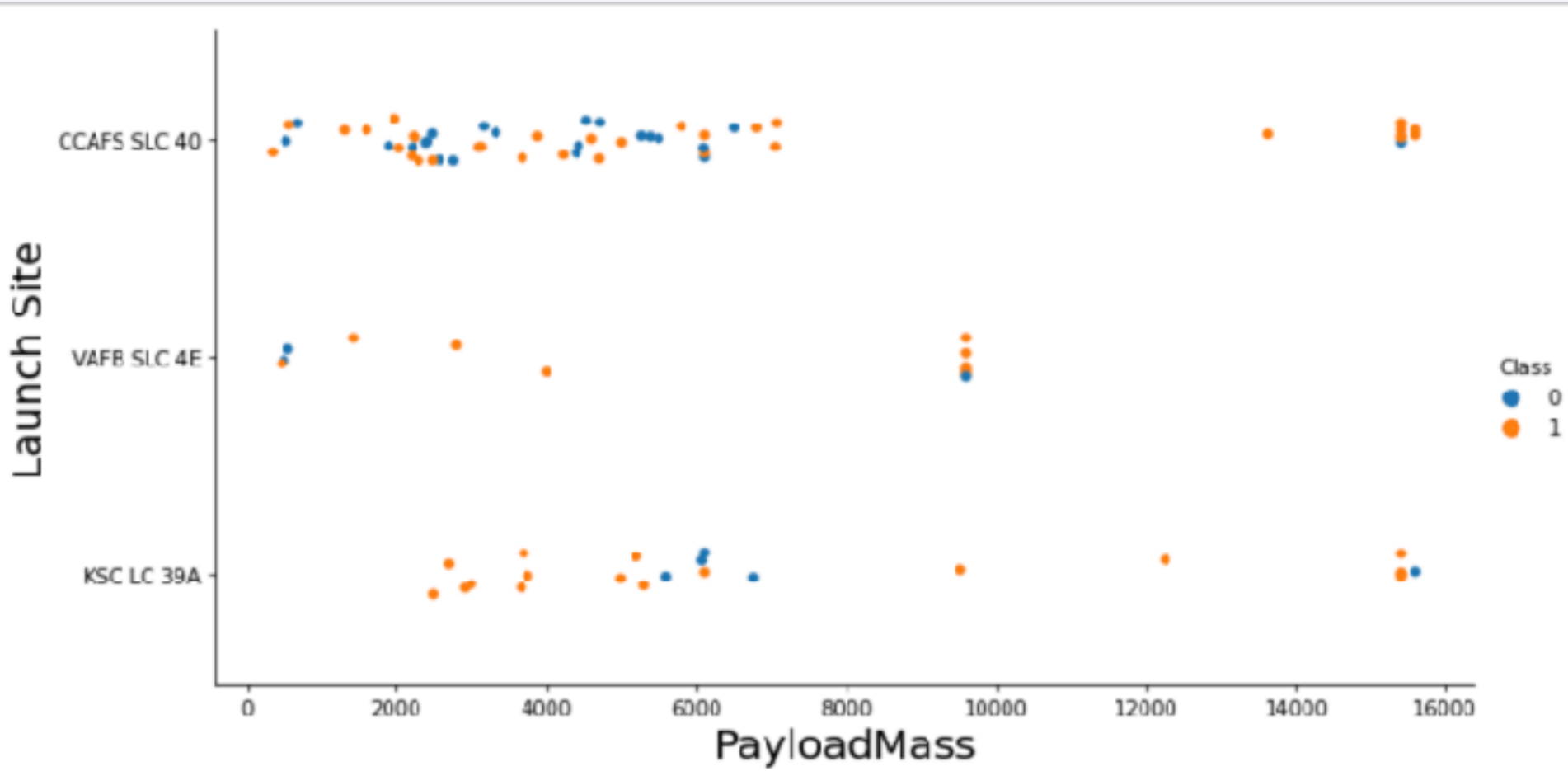
Section 2

Insights drawn from EDA

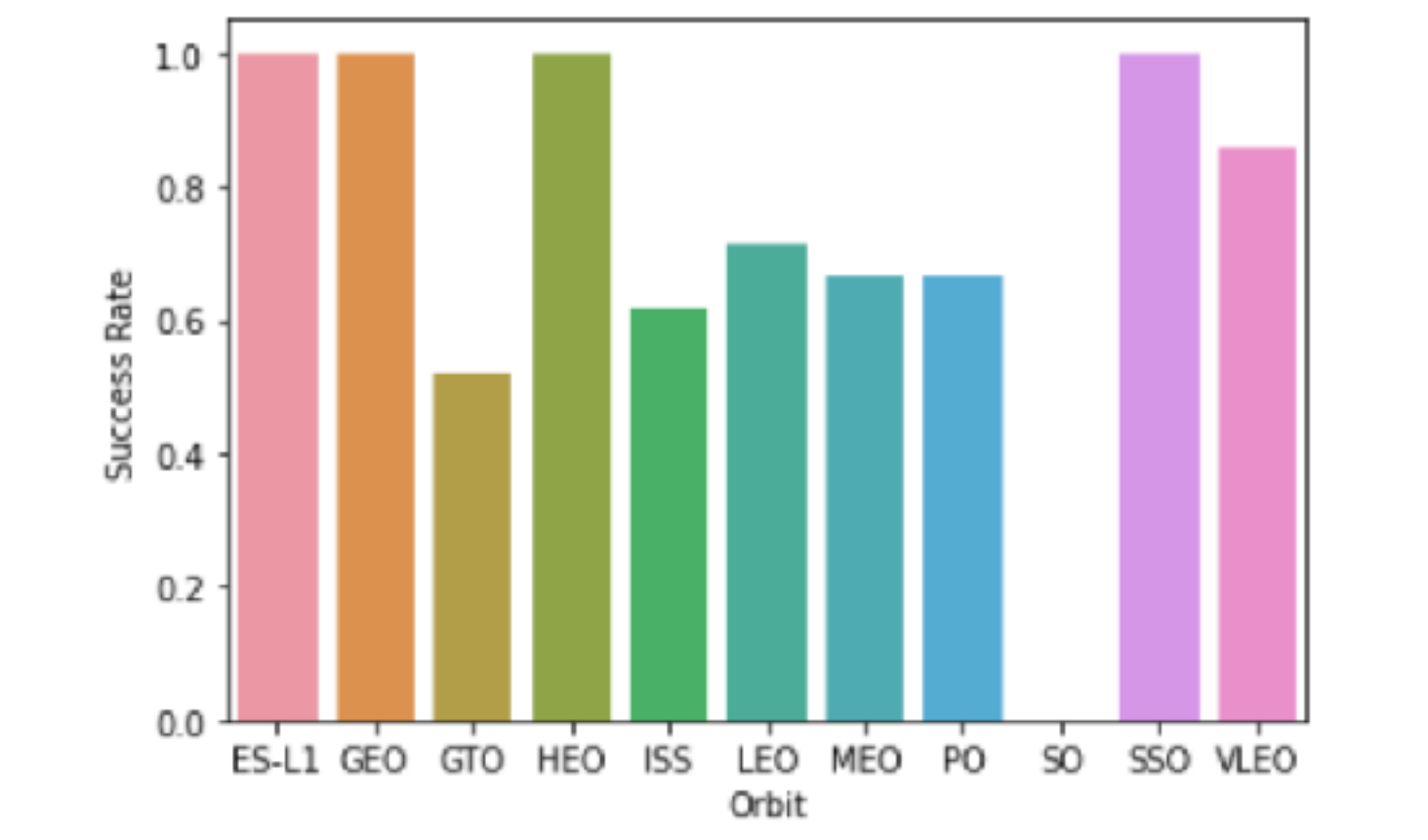
Flight Number vs. Launch Site



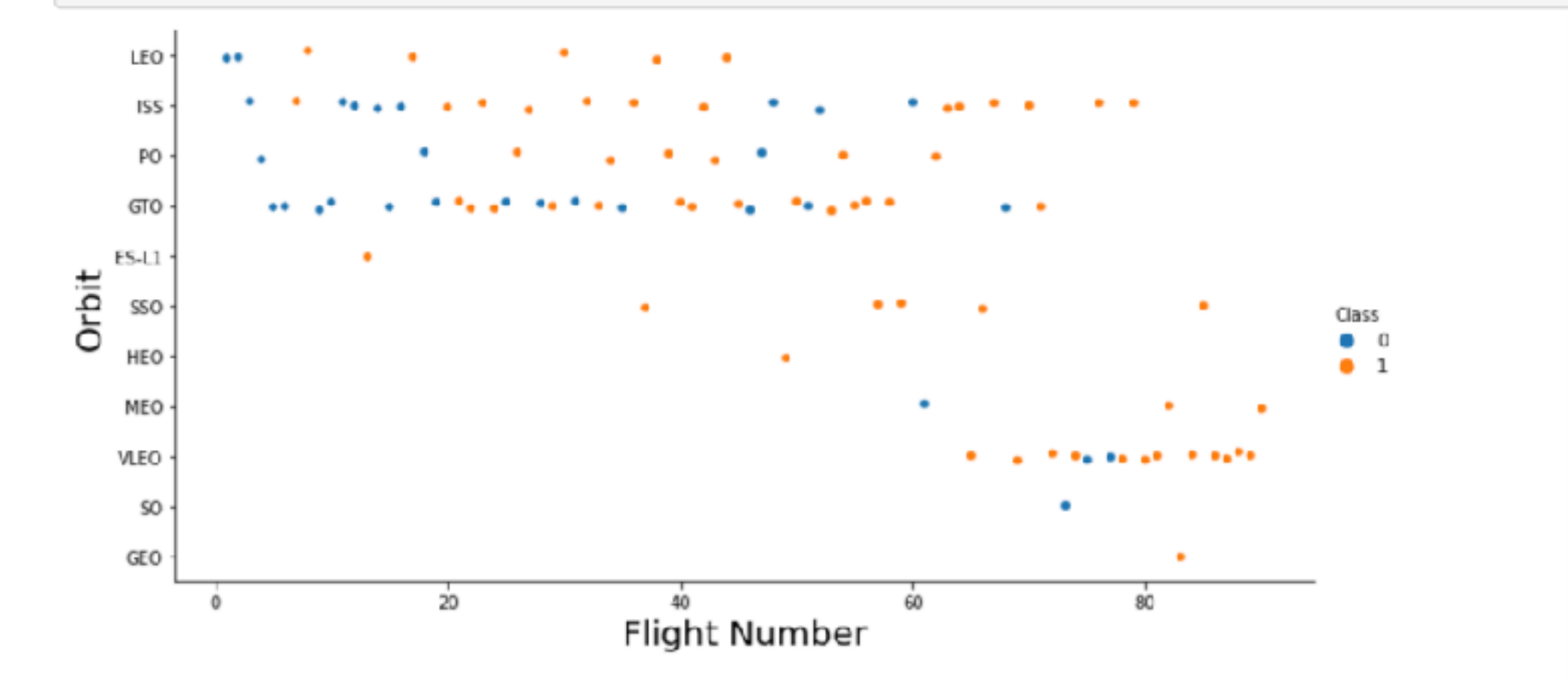
Payload vs. Launch Site



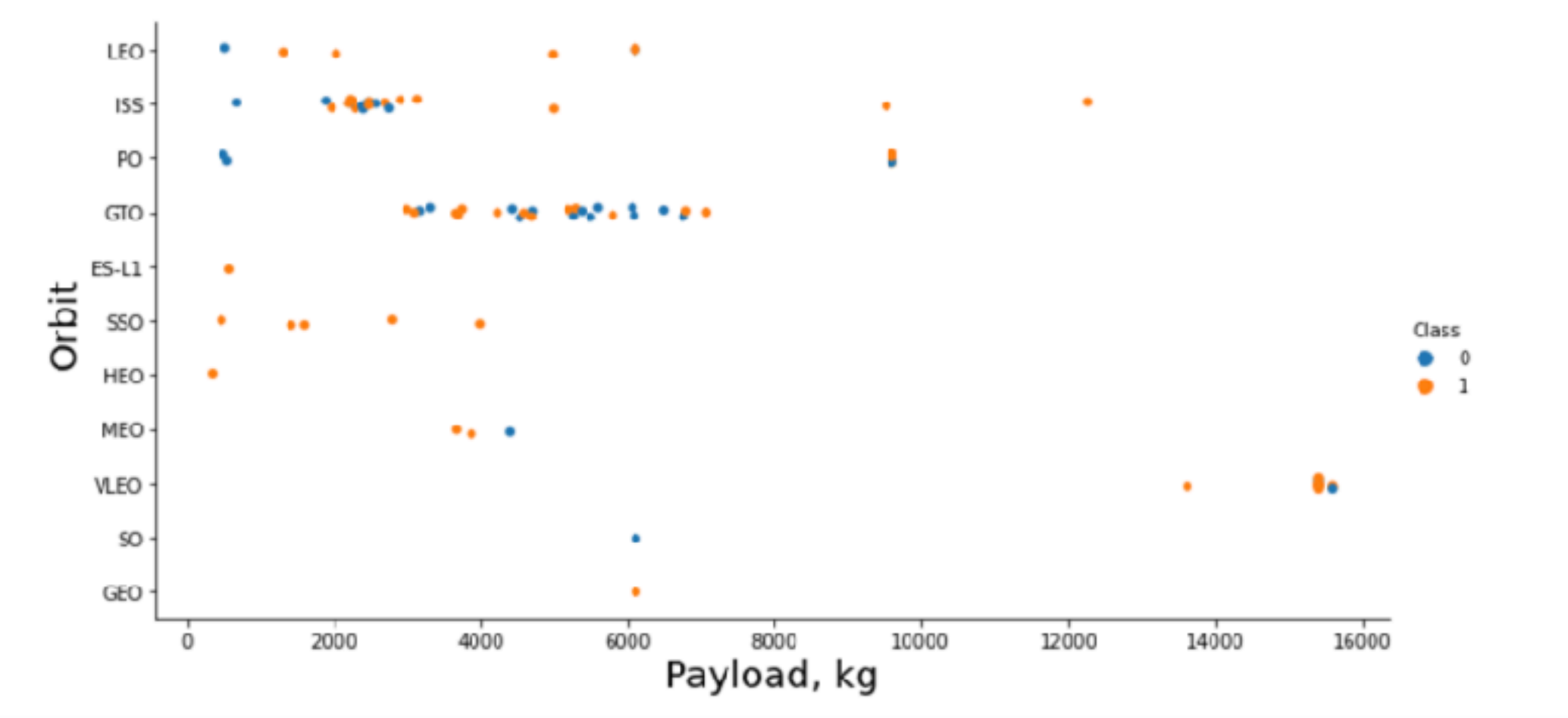
Success Rate vs. Orbit Type



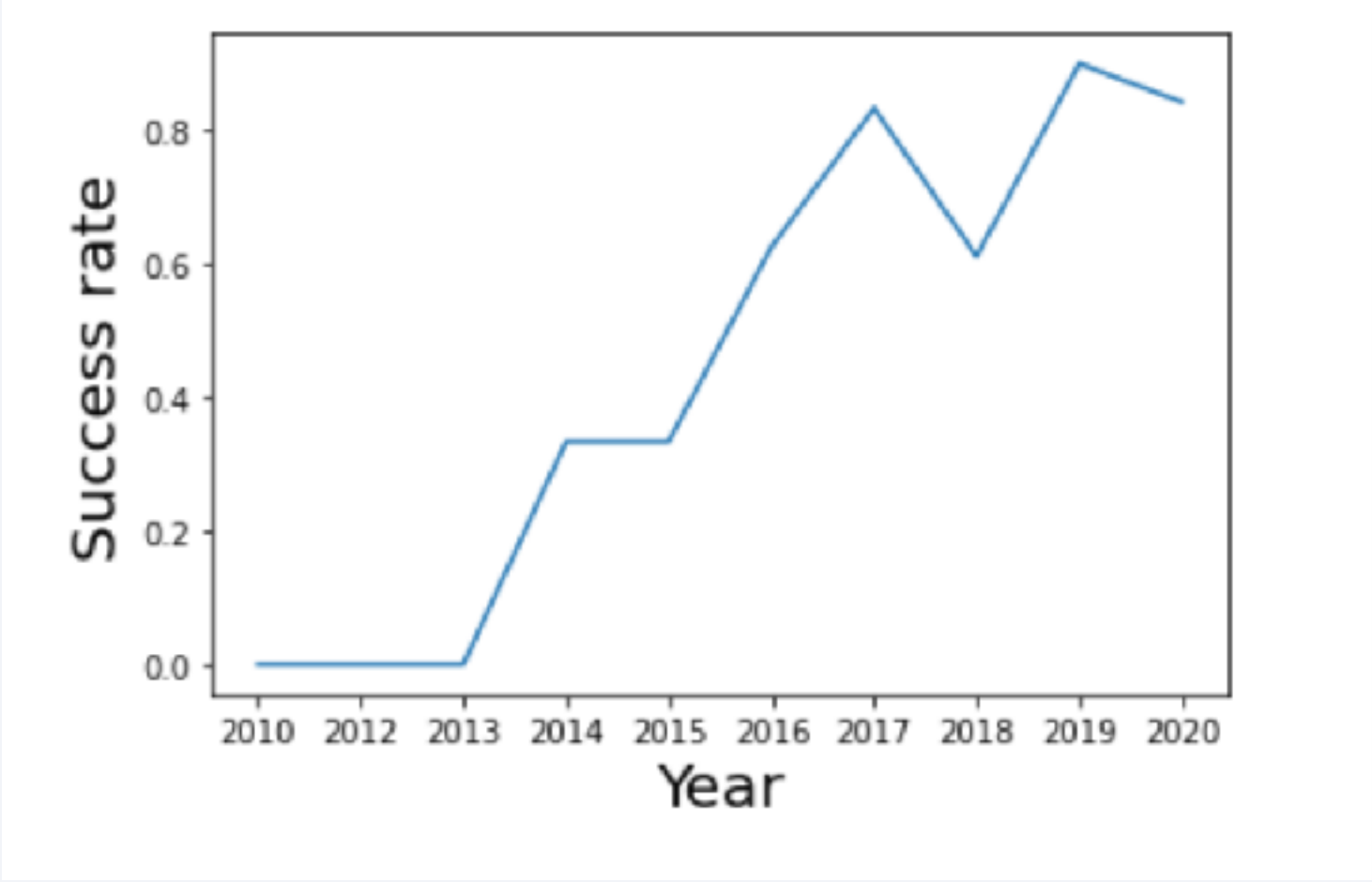
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

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

Launch Site Names Begin with 'CCA'

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

```
%sql select sum(payload_mass__kg_) from spacextbl where customer = 'NASA (CRS)';
```

```
 * ibm_db_sa://bnx14314:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1
Done.
```

]:

1
45596

Average Payload Mass by F9 v1.1

```
%sql select avg(payload_mass__kg_) from spacextbl where booster_version = 'F9 v1.1';
```

```
 * ibm_db_sa://bnx14314:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8
Done.
```

:

1
2928

First Successful Ground Landing Date

```
%sql select min("DATE") from spacextbl where landing__outcome LIKE 'Success%';
```

```
 * ibm_db_sa://bnx14314:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90108kq
Done.
```

1
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select distinct(booster_version) from spacextbl \
    where landing__outcome = 'Success (drone ship)' \
    and payload_mass__kg_ > 4000 \
    and payload_mass__kg_ < 6000;
```

```
 * ibm_db_sa://bnx14314:***@764264db-9824-4b7c-82df-40d1b
Done.
```

]:

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

Total Number of Successful and Failure Mission Outcomes

```
%sql select mission_outcome, count(mission_outcome) as cnt from spacextbl \
    group by mission_outcome;
```

```
 * ibm_db_sa://bnx14314:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90108
Done.
```

]:

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

Boosters Carried Maximum Payload

```
%sql select unique(booster_version) from spacextbl \
    where payload_mass__kg_ = (select max(payload_mass__kg_) from spacextbl);
```

```
 * ibm_db_sa://bnx14314:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90108kqb1
Done.
```

]:

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

```
%sql select date, booster_version, launch_site from spacextbl \
    where landing__outcome = 'Failure (drone ship)' and year(date)=2015;
```

```
 * ibm_db_sa://bnx14314:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io9
Done.
```

]:

DATE	booster_version	launch_site
2015-01-10	F9 v1.1 B1012	CCAFS LC-40
2015-04-14	F9 v1.1 B1015	CCAFS LC-40

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

```
%sql select landing__outcome, count(landing__outcome) as cnt from spacextbl \
    where date >= '2010-06-04' and date <= '2017-03-20' \
    group by landing__outcome \
    order by cnt desc;
```

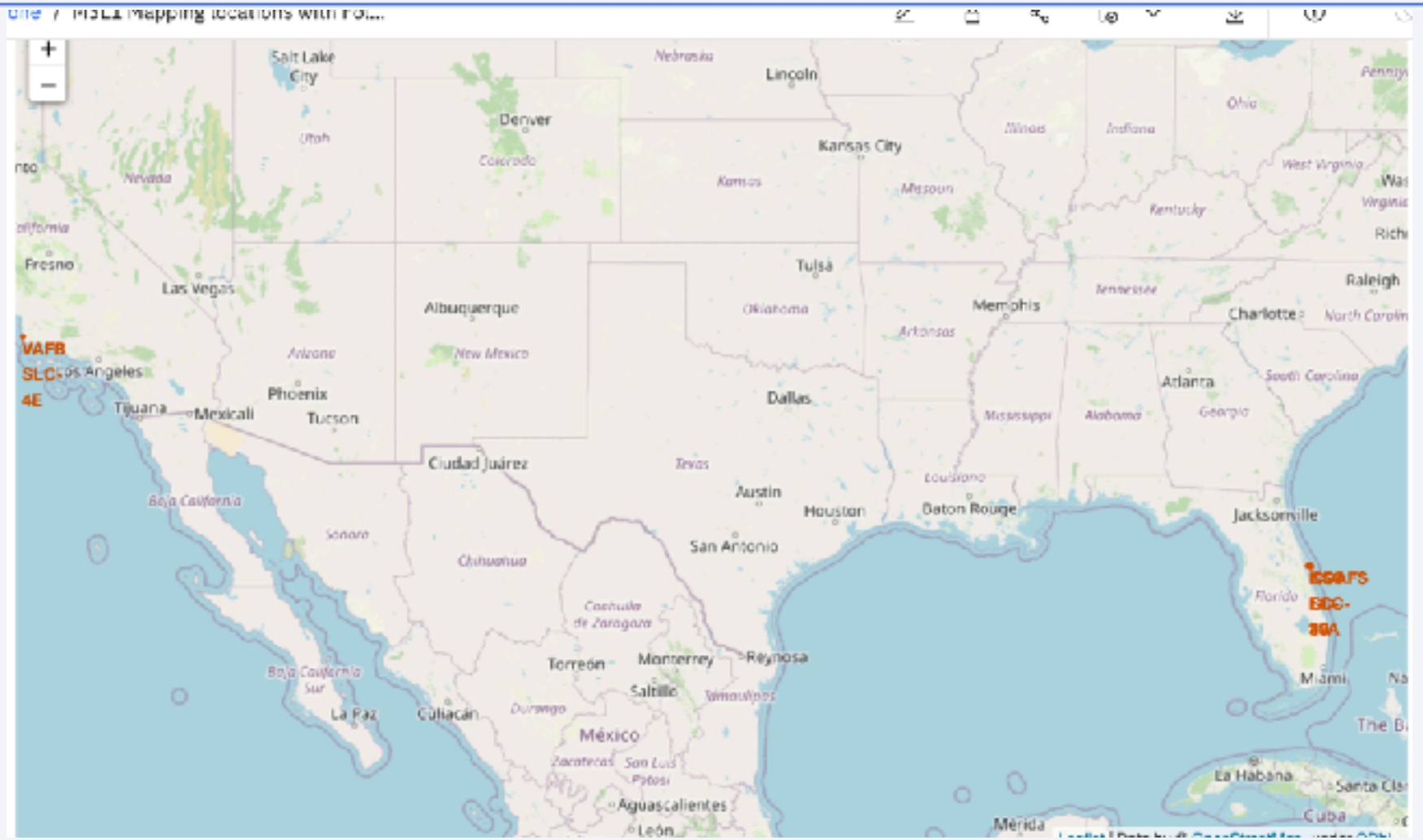
```
 * ibm_db_sa://bnx14314:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb
Done.
```

]:

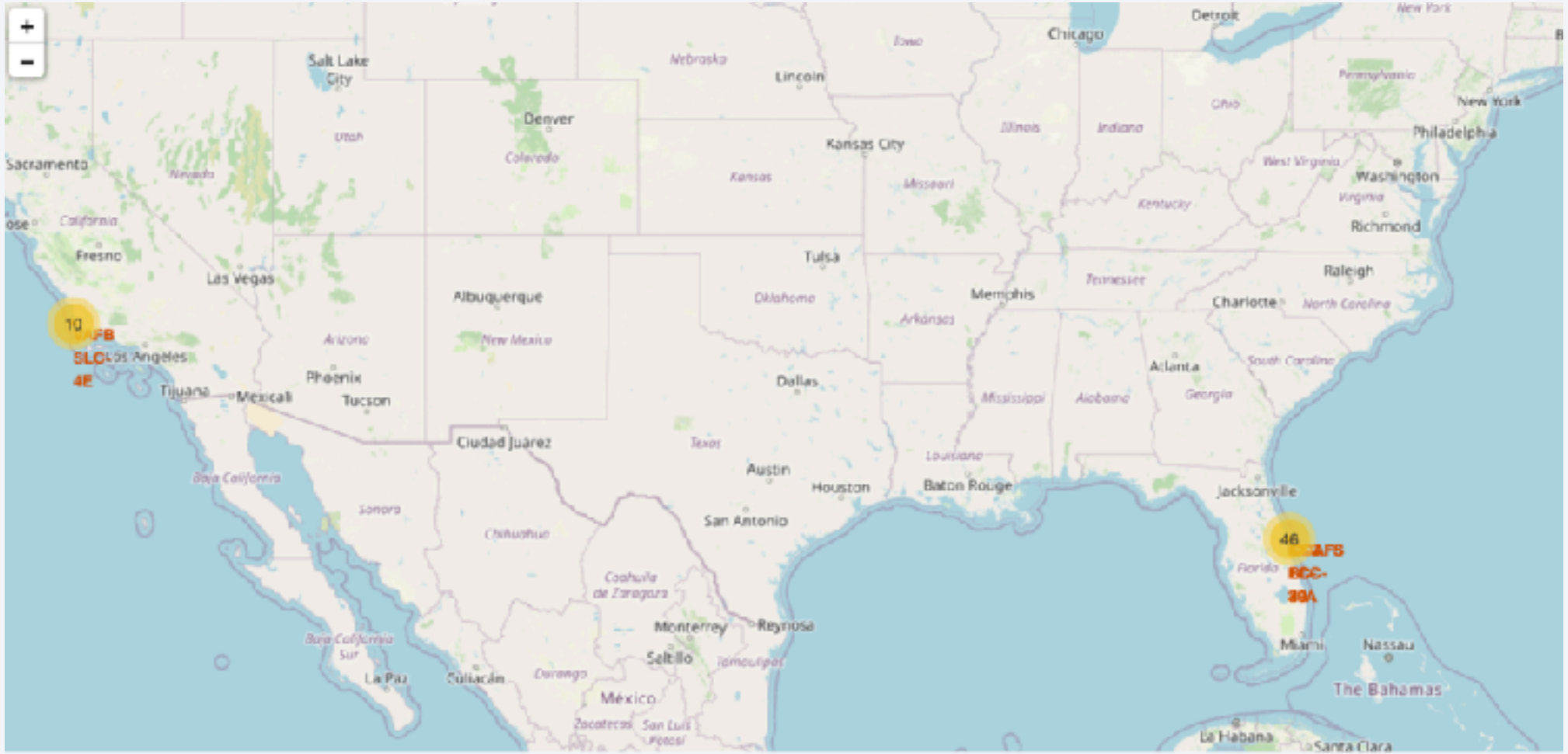
landing__outcome	cnt
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1


Section 4
Launch Sites
Proximities Analysis

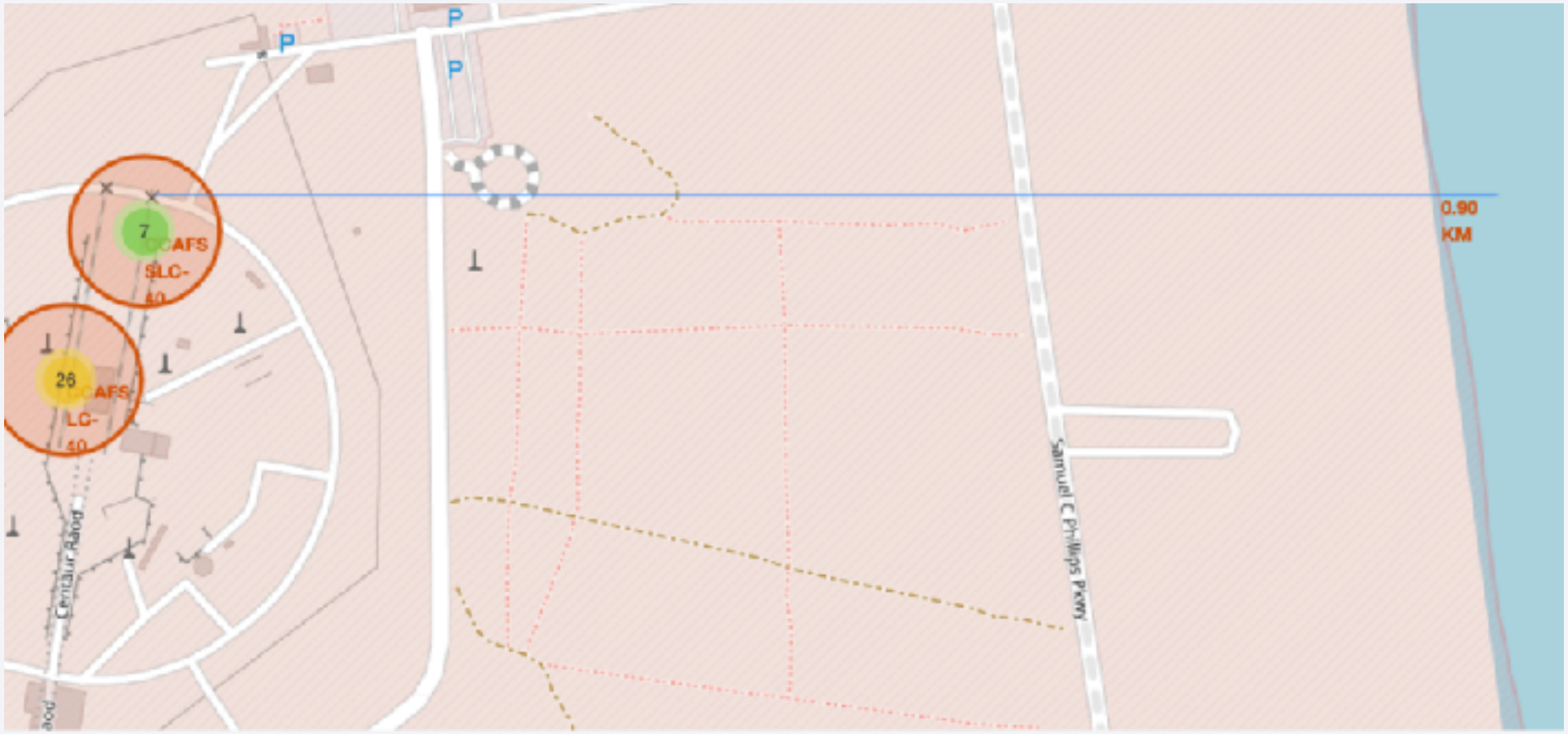
<Folium Map Screenshot 1>



<Folium Map Screenshot 2>



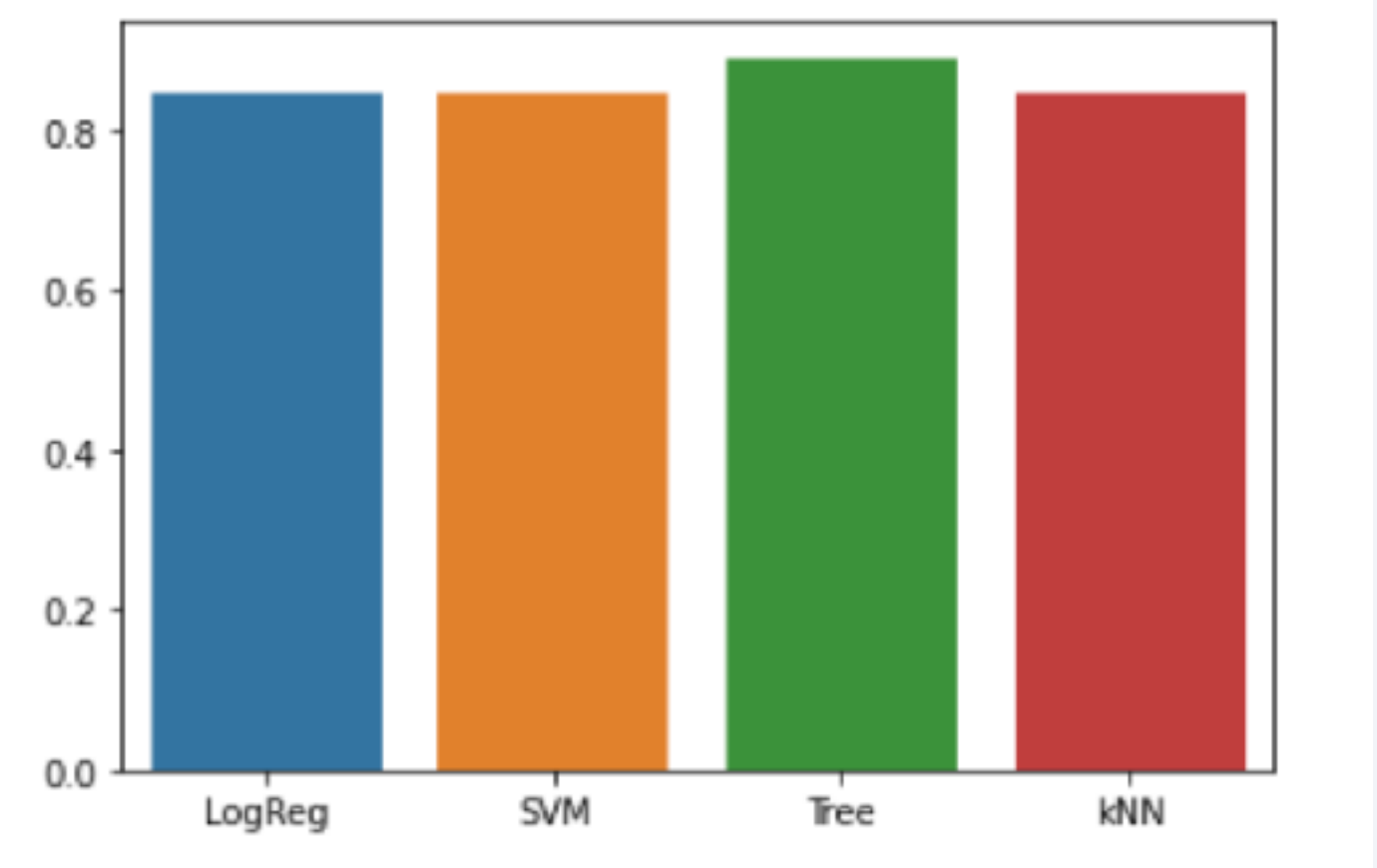
<Folium Map Screenshot 3>



Section 6

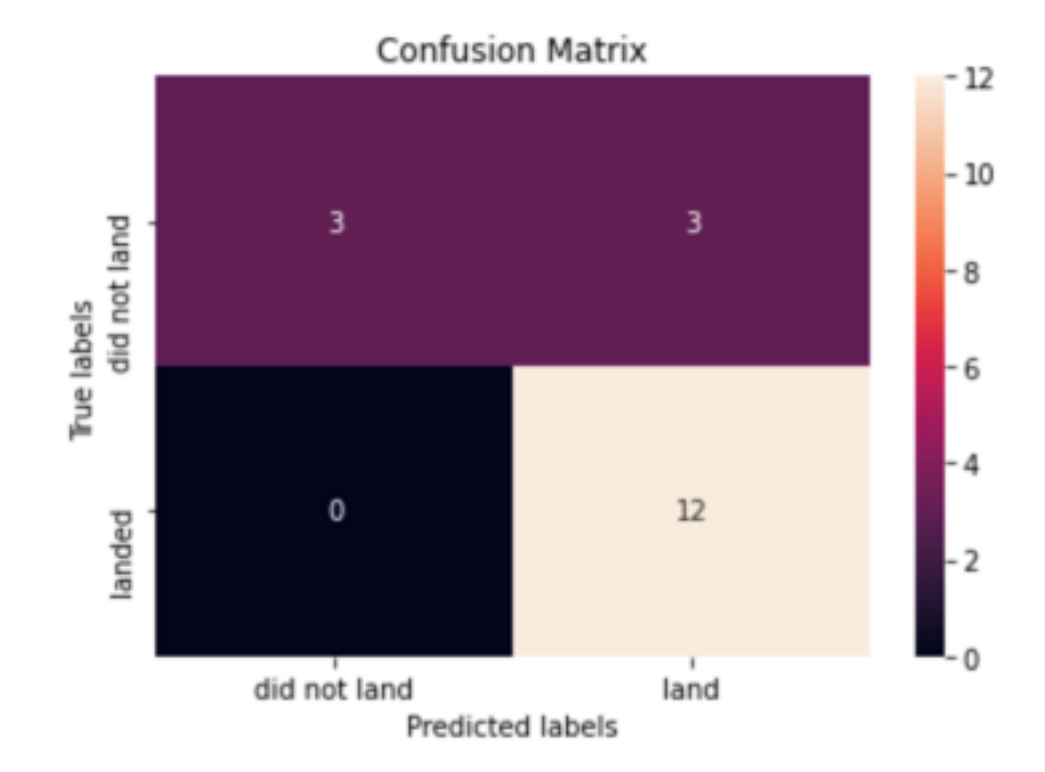
Predictive Analysis (Classification)

Classification Accuracy



Decision Tree classifier has marginally better accuracy.

Decision Tree Confusion Matrix



Conclusions

All classifiers had comparable performance, with Decision Tree model being slightly better on training set.
All classifiers struggled with false positive.

The most likely reason why Decision Tree was slightly better than others is the discrete nature of data.

Thank you!