How SpaceX transform space race

Lih Jen Tobin

Jan 17, 2022

Executive Summary

- One of biggest SpaceX successful factor is it can reuse its rocket and hence reduce its cost for launching rocket to space for satellite deployment, supply equipment/material and etc.
- This report demonstrates the analysis of features like launch sites, payloads, and number of flight that link to the success,
- Charts and Graphs will be used to illustrate the relationship between features.
- Machine learning models are used to simulate the success rates among models.
- Results from the models will be shown using ROC curve.

Introduction

- What to learn from SpaceX in order to compete for low cost rocket launch?
- To compete with SpaceX, we can use the SpaceX launch info to help us understand what elements go into a successful rocket recycled.
- We'd like know the detail analysis of all its rocket launches up to year 2020.
- The analysis will include the launch sites, obits, and payloads and etc for success and failure launches.
- We then can create a machine learning model to predict the success rate of a new launch.

Methodology - summary

- Use the data from IBM and SpacX web sites.
- Import the data, clean the data, and create a classification column named "Class", and limit the data to include only Falcon 9 rockets.
- Using Sql to get insights of the data.
- Creating charts and graphs to visualize the relationship between features.
- Using maps to display the locations of launch sites, distance to varies landmark like city, coastline, and railway.
- Use different machine models to predict the success rates.

Methodology – Data Source

Data sources: https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json

Fetch the necessary features data, eg: payload, launch pad and etc.

Data is cleaned and wrangled to limit to only "Falcon 9" launch data.

FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Seri
1	2010-06-04	Falcon 9	6123.547647	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B000
2	2012-05-22	Falcon 9	525.000000	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B000
3	2013-03-01	Falcon 9	677.000000	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B000
4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B100
5	2013-12-03	Falcon 9	3170.000000	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B100

Methology - Data Wrangling

Load the data into IBM DB2 database.

New column "Class" is added to the dataframe

GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	0
False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	0
False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	0
False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	0
False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	0

Methology - Database Observation

- ***What are the sites used for mission?***

- %sql select distinct(launch_site) from SPACEXTBL2

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

Methology - Database Observation

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

%sql SELECT * FROM SPACEXTBL2 where launch_site like 'CCA%' FETCH FIRST 5 ROWS ONLY

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	
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)	(CC
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NA
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NA

Methology - Database Observation

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

- %sql select sum(payload_mass__kg_) FROM SPACEXTBL2 where customer = 'NASA (CRS)'

1
45596

Methology - Database Observation

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

- %sql select AVG(payload_mass__kg_) FROM SPACEXTBL2 where booster_version like 'F9 v1.1%'

1
2534

Methology - Database Observation

- *List the "date" when the first successful landing outcome in ground pad was achieved.*

- %sql select min(DATE) FROM SPACEXTBL2 where landing__outcome like 'Success%'

1
2015-12-22

Methology - Database Observation

- ***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 SPACEXTBL2 where landing__outcome = 'Success (drone ship)' and payload_mass__kg_ > 4000 and payload_mass__kg_ < 6000

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

Methology - Database Observation

- ***List the total number of successful and failure mission outcomes***

- %sql select mission_outcome, count(*) as count FROM SPACEXTBL2 group by mission_outcome

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

Methology - Database Observation

- ***List the names of the "booster_versions" which have carried the maximum payload mass.***

- %sql select booster_version from SPACEXTBL2 where payload_mass__kg_ = (select max(payload_mass__kg_) from SPACEXTBL2)

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

Methology - Database Observation

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

- %sql select Booster_Version, launch_site from SPACEXTBL2 where landing__outcome = 'Failure (drone ship)' and year(DATE) = '2015

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

Methology - Database Observation

- ***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(*) as count , dense_rank() over(order by count(*) desc) rn from SPACEXTBL2 where DATE between '2010-06-04' and '2017-03-20' group by landing_outcome order by count

landing__outcome	COUNT	rn
Precluded (drone ship)	1	5
Uncontrolled (ocean)	2	4
Failure (parachute)	2	4
Success (ground pad)	3	3
Controlled (ocean)	3	3
Success (drone ship)	5	2
Failure (drone ship)	5	2
No attempt	10	1

Methology - Visualizations

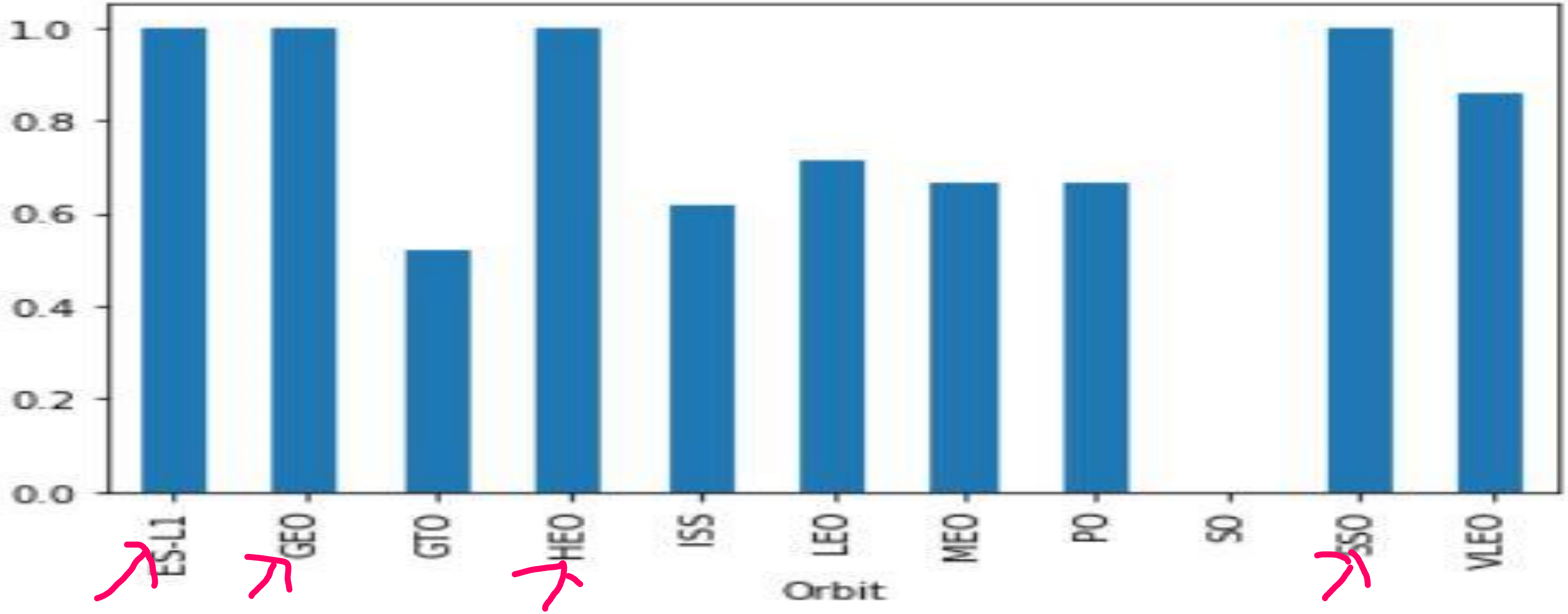
Relationship between Launch Sites and Payloads

Observation:

At "VAFB-SLC" launch site, there are no rockets launched for heavy payload mass(greater than 10000).

Methology - Visualizations

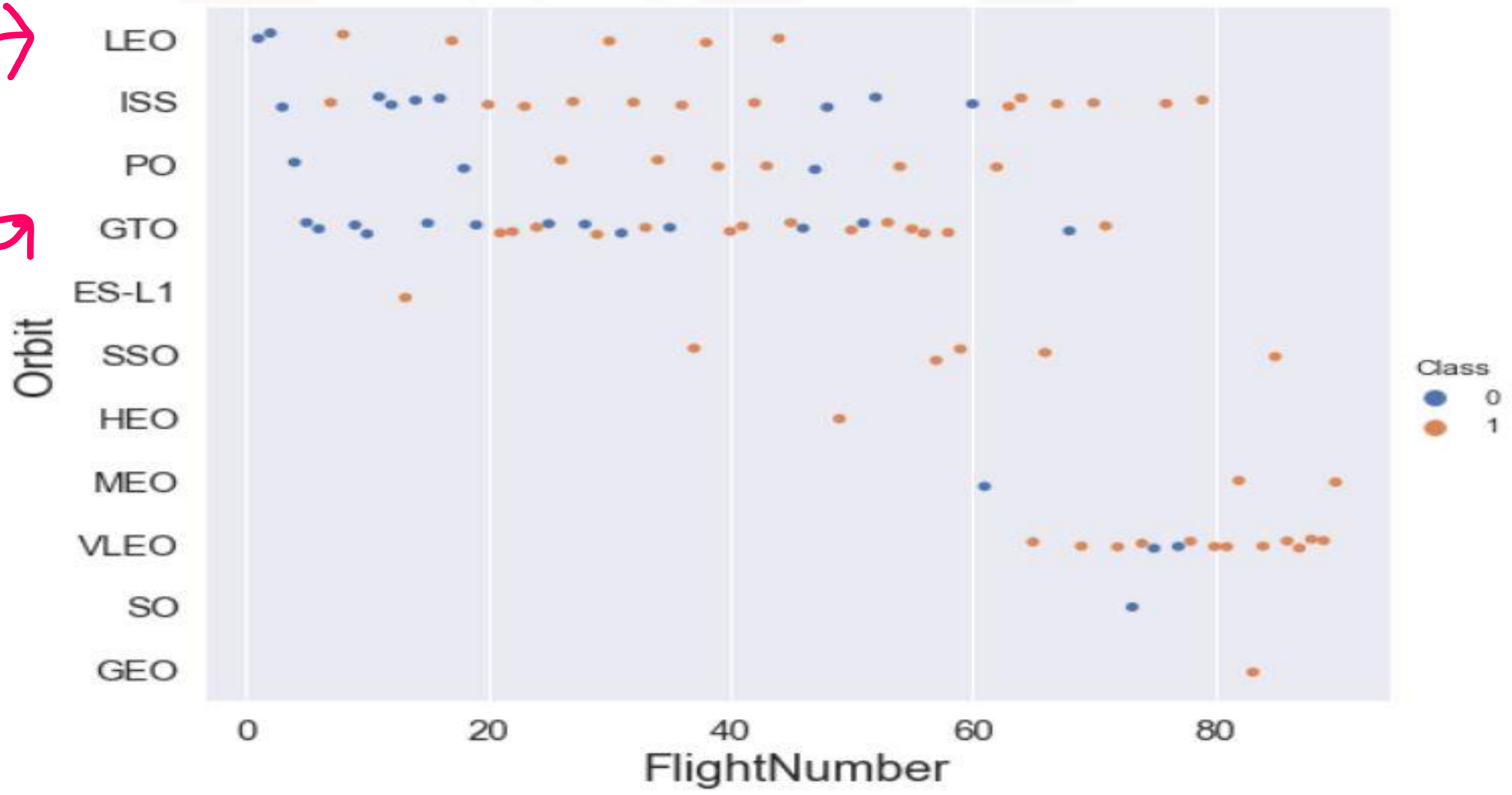
Success rate of each orbit type



Observation : ES-L1, HEO, GEO, and SSO has high sucess rate.

Methology - Visualizations

Relationship between FlightNumber and Obit



Observation :

In the LEO orbit the Success appears related to the number of flights;

There seems to be no relationship between flight number in GTO orbit.

Methology - Visualizations

Relationship between PayLoads and Obit

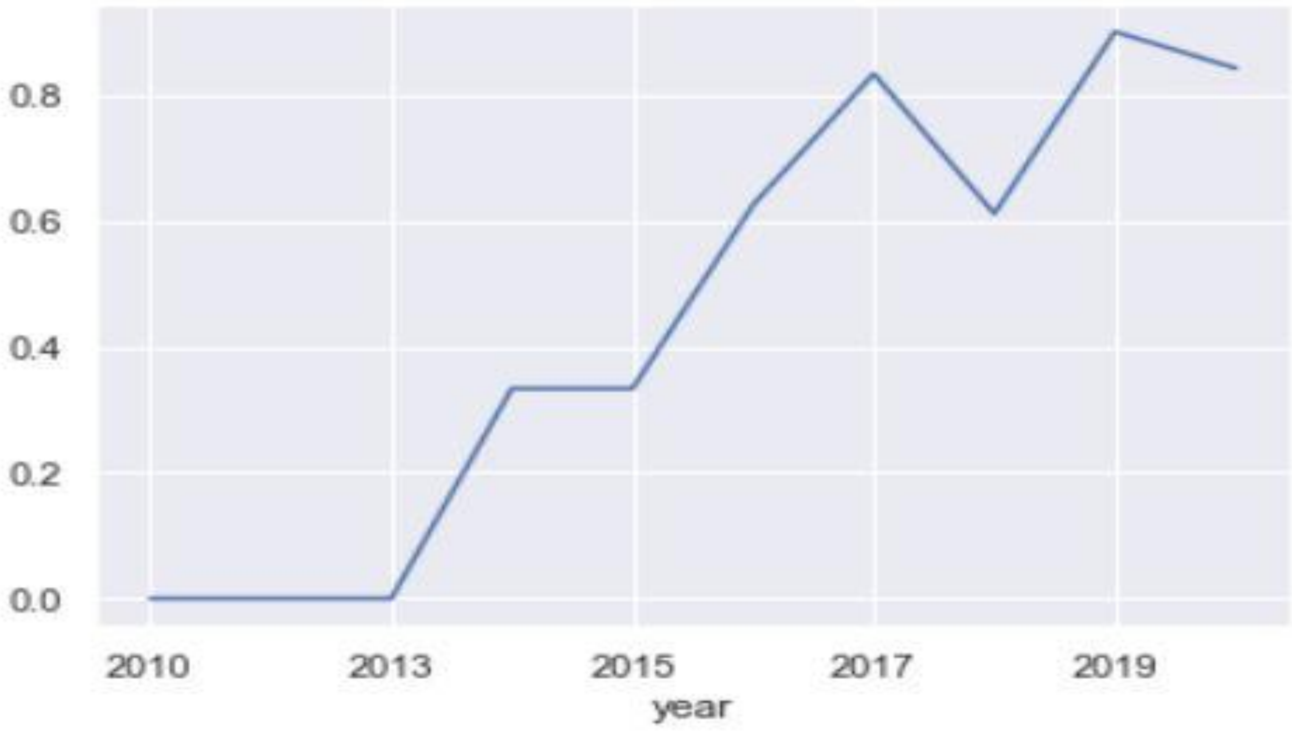
Observation :

With heavier payloads the successful landing rate are more for "LEO" and "ISS"

We cannot distinguish this well for "GTO "as both successful / fail landing rate both there here.

Methology - Visualizations

Relationship between Year and Success Rate



Observation:

Sucess rate kept increasing since 2013 till 2020

Methology - Visualizations

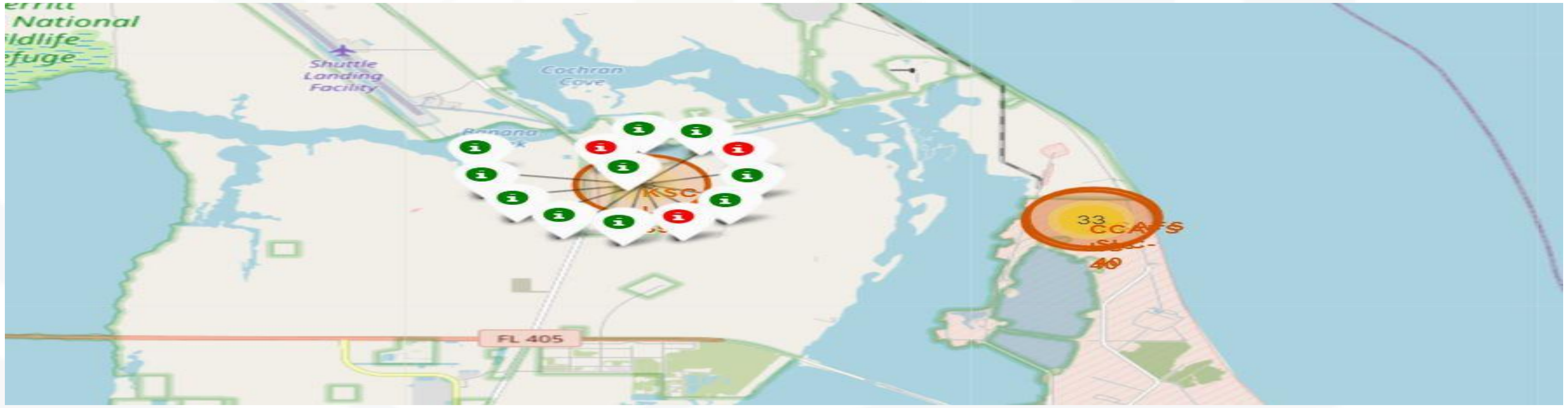
Let's look at the launch location in the United State as a whole

Total 56 launches
46 in East Coast
10 in West Coast

Methology - Visualizations

Observation:

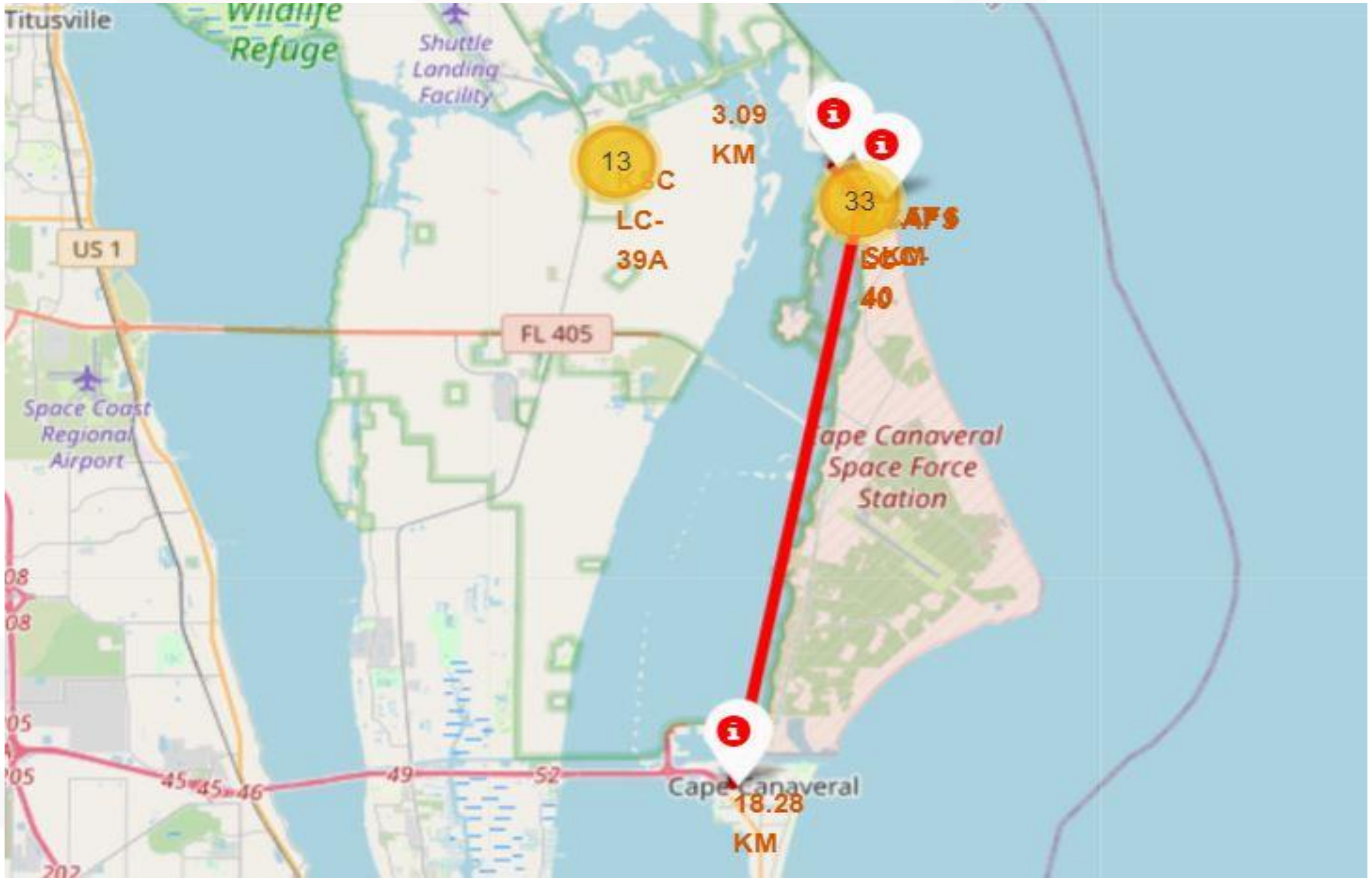
In "KSC LC-39A" launch site , we can see the high success rate with green markers in the lower graph



Methology - Visualizations

Observation:

"CCAFS SLC-40" launch site to "cape canaveral" city is about 18.28 km



Methology - Visualizations

Observation:

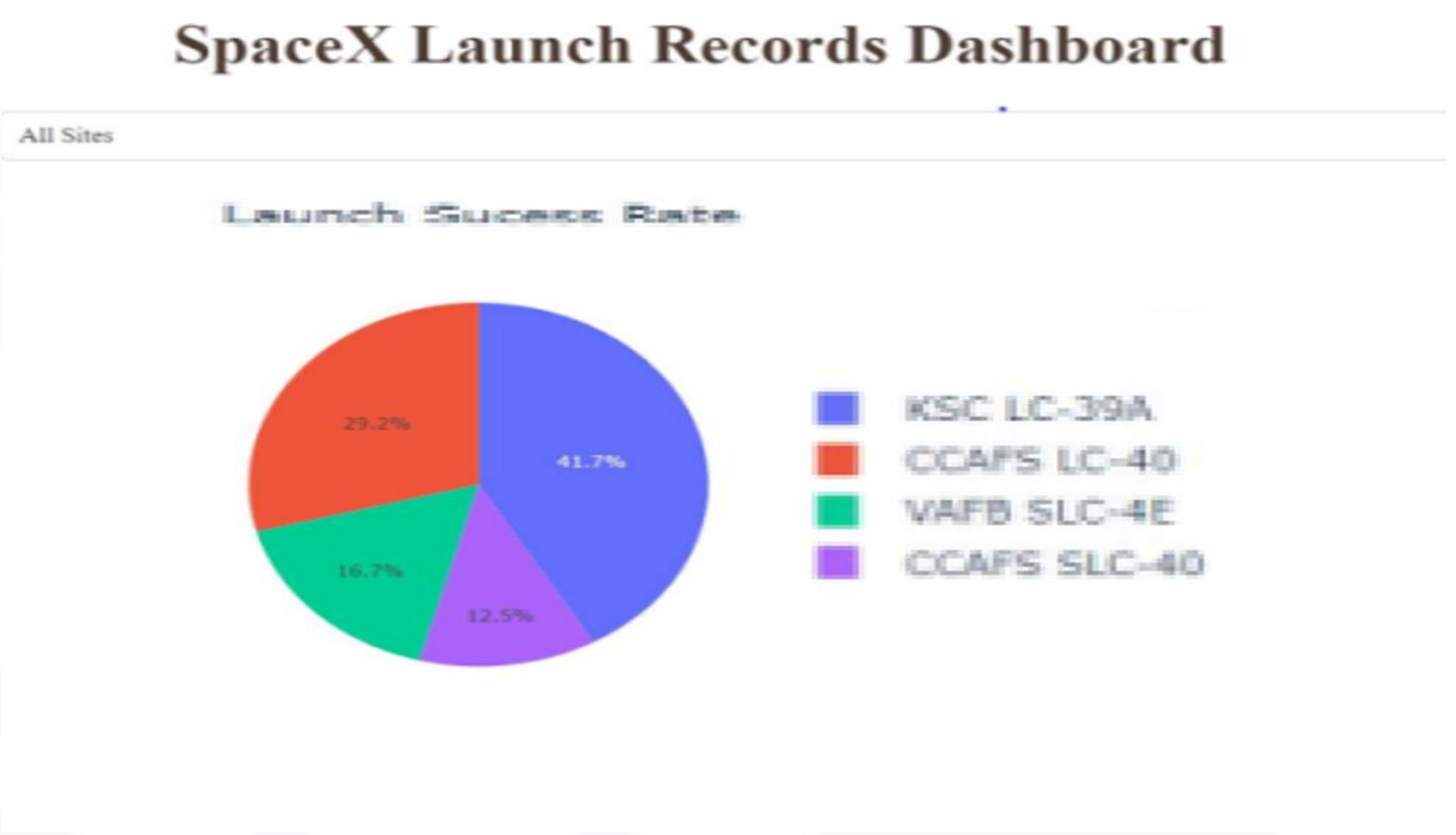
"CCAFS SLC-40" launch site to "railroad" is 1.27 km

"CCAFS SLC-40" launch site to "coastline" is 0.57 km

Methology - Interactive charts

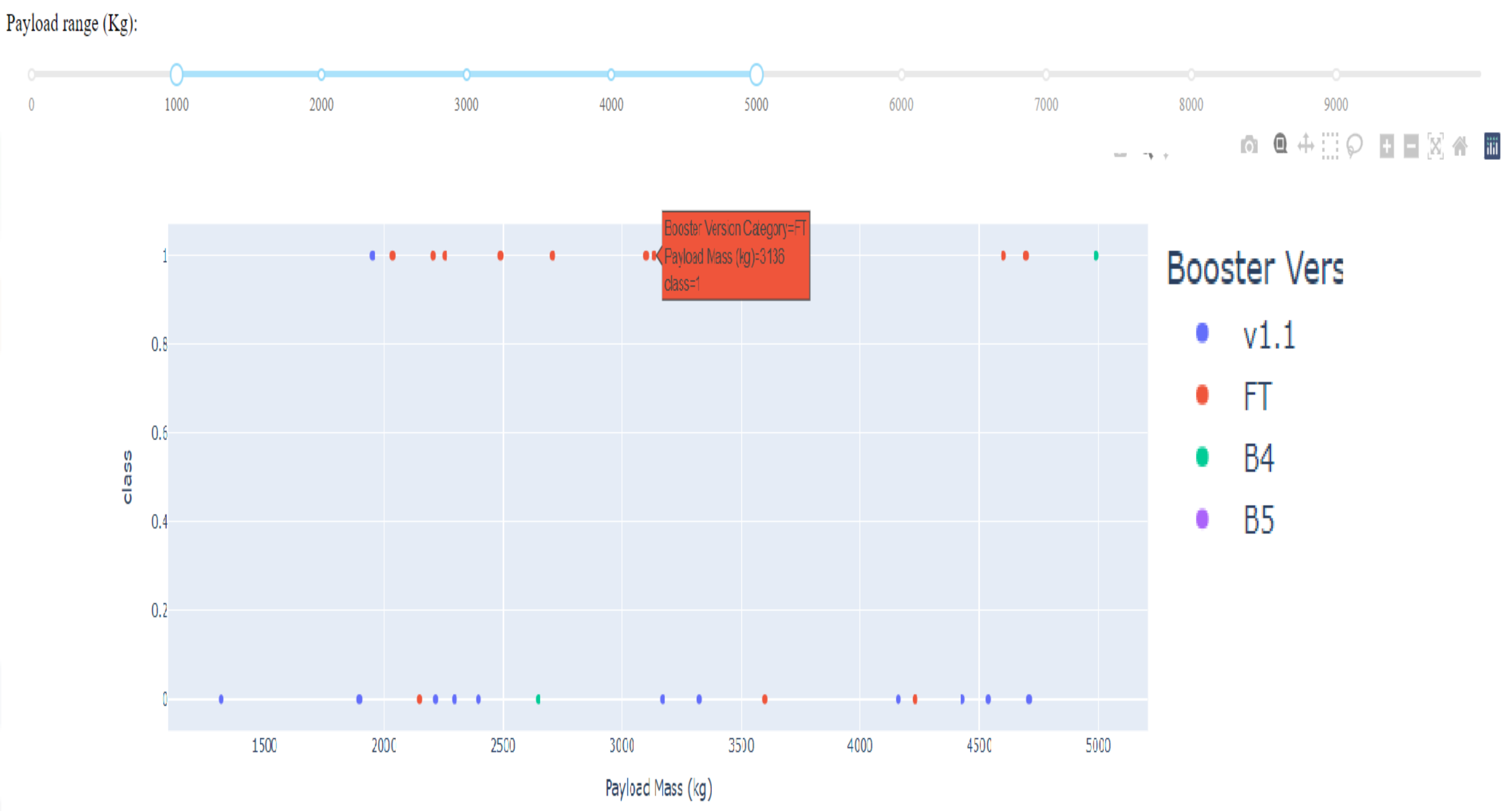
Using Plotly app, we can observe the launch success rate for all sites and each site

KSC LC-39A site has the highest successful launch rate



Methology - Interactive charts

In Plotly app, using Payload slider, we can observe success/fail launch for different boosters



Methology – Apply Machine Models to Data

- **We will continue to find the best Machine Learning Model using these data.**
- **SVM, Decision Trees, K Nearest Neighbors and Logistic Regression Classification models are used in this trail.**
- **Use Accuracy rate, ROC chart, and F1- score to determine the best performing model.**
- ***Good luck! Machines!!!***

Methology – Prepare Data for models

Assign Class column to a series as dependent variable

```
y = pd.Series(data['Class'])
```

Scale the independent variables

```
X = transform.fit_transform(X)
```

Split the data into train and test data set with ratio of 80:20.

```
X_train, X_test, Y_train, Y_test = train_test_split(X, y,
test_size=0.2, random_state=2 )
```

Photo credit: Google Pictures

Methology – Logistic Regression

Photo credit: Pixabay

Use GridSearchCV object to find the best fit of model and fit data to the model

Logistic Regression

```
logreg_cv = GridSearchCV(lr,
                         param_grid=parameters,
                         scoring='accuracy',
                          cv=10)
logreg_cv.fit(X_train, Y_train)
```

Get the Accuracy Score and F1 score

```
logreg_cv.score(X_test, Y_test)  Out : 0.8333333333333334

yhat_lr = logreg_cv.predict(X_test)
f1_lr = f1_score(Y_test, yhat_lr) Out : 0.888888888888889
```

Methology – Support Vector

Use GridSearchCV object to find the best fit of Support Vector model and fit data to the model

```
svm_cv = GridSearchCV(svm,
                      param_grid=parameters,
                      scoring='accuracy',
                      cv=10)
svm_cv.fit(X_train, Y_train)
```

Get the Accuracy Score and F1 score

svm_cv.score(X_test, Y_test) Out : 0.8333333333333334

yhat_svm = svm_cv.predict(X_test)
f1_svm = f1_score(Y_test, yhat_svm) Out : 0.8888888888888889

Methology – Decision Tree

Use GridSearchCV object to find the best fit of Decision Tree model and fit data to the model

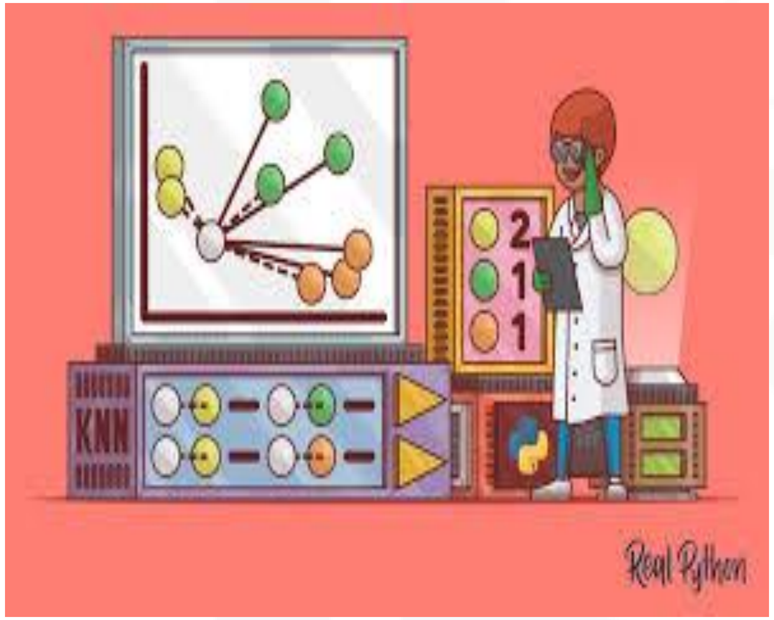
```
tree_cv = GridSearchCV(tree,
                       param_grid=parameters,
                       scoring='accuracy',
                       cv=10)
tree_cv.fit(X_train, Y_train)
```

Get the Accuracy Score and F1 score

tree_cv.score(X_test, Y_test) Out : 0.8888888888888888

yhat_tree = tree_cv.predict(X_test)
f1_tree = f1_score(Y_test, yhat_tree) Out : 0.916667

Methology – K Nearest Neighbors



Use GridSearchCV object to find the best fit of K Nearest Neighbors model and fit data to the model

```
knn_cv = GridSearchCV(KNN,
                      param_grid=parameters,
                      scoring='accuracy',
                      cv=10)
```

Get the Accuracy Score and F1 score

knn_cv.score(X_test, Y_test) Out : 0.8333333333333334

yhat_knn = knn_cv.predict(X_test)
f1_knn = f1_score(Y_test, yhat_knn) Out : 0.888889

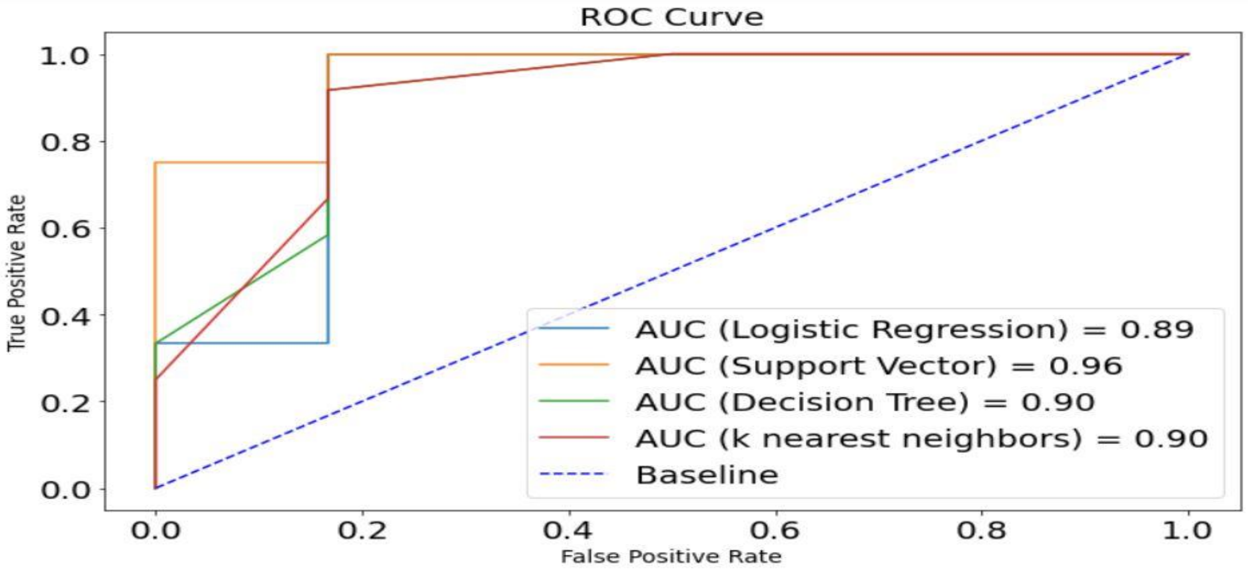
Result – Summary of the Scores

- Log reg F1 score: 0.888889
- SVM F1 score: 0.888889
- Tree F1 score: 0.916667
- KNN F1 score: 0.888889

- Decision Tree has the best F1 score of 91%. Indicates 91 % of the launches has successful landings.

Result – ROC Curve

Looking at the ROC, Support Vector perform the best with 0.96.
The dotted blue line represent random classifier. Support Vector has the biggest gap between the dotted line and its curve.

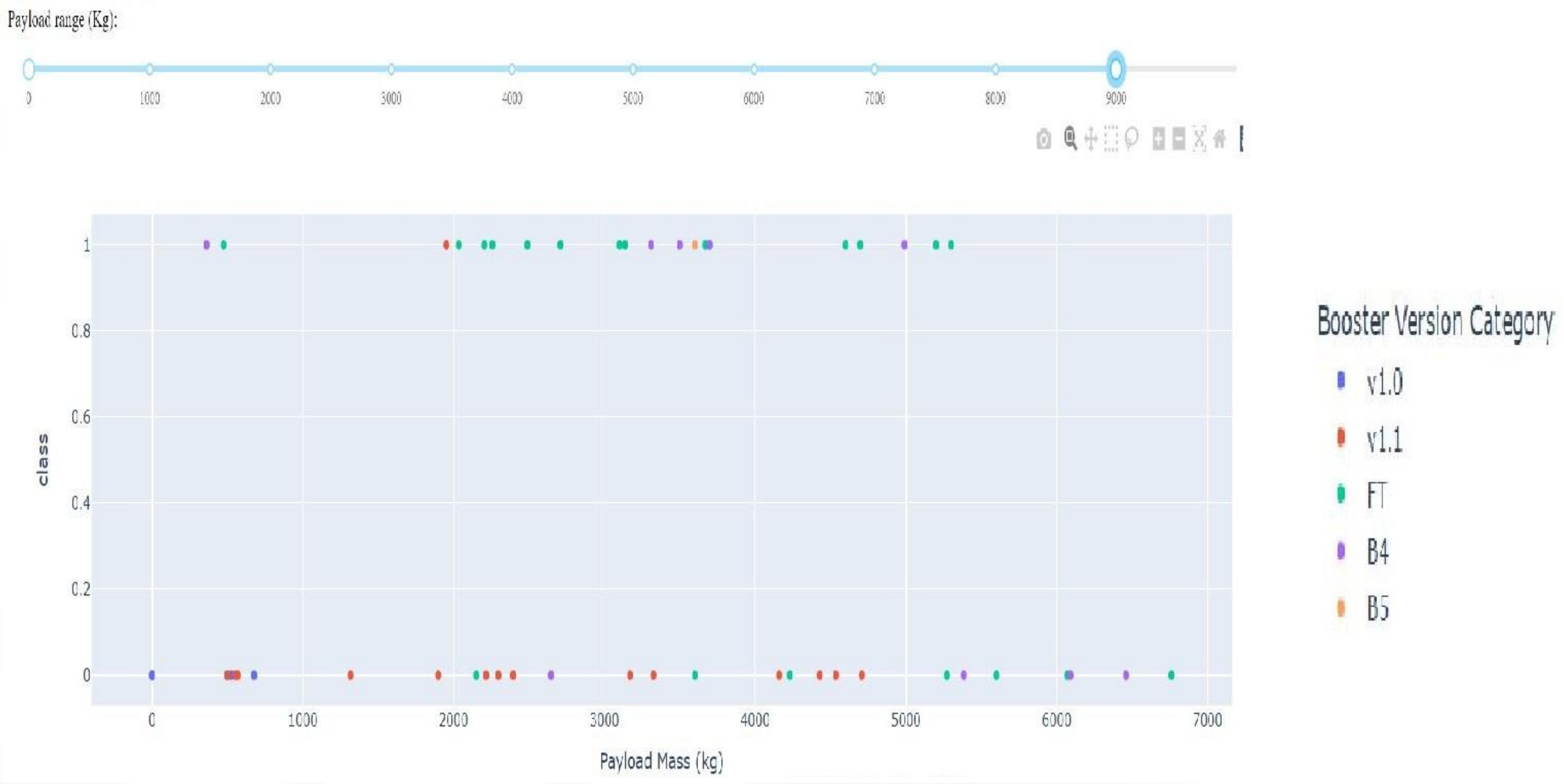


Result – Launch site significance

- To transport rocket, the railroad distance to the launch site is important since rocket is transport by railway to the launch site.
- East coast has the better environment for rocket launch in terms of equator closeness, Earth's west-to-east spin for speed and coastline closeness for safely landing.
- KSC LC-39 A has 100% success rate.

Result – Booster Version significance

- Using Plotly, between payload of 1000 and 9000, "F1" boosters has the most successful Launches sho wing as green dots on the graph.
- The "B5" boosters has carried the maximum payload mass.
- The successful drone ship with payload between 4000 and 6000 are "FT " version booster



Result – Obit significance

- In obits, ES-L1, HEO, GEO, and SSO has high sucess rate.
- With heavier payloads the successful landing rate are more for "LEO" and "ISS".
- In the LEO orbit the Success rate appears related to the number of flights.
- There seems to be no relationship between flight number in GTO orbit.

Discussion

Since Obits have different distance to the earth, From the analysis, we found that some obit and payload relationship. But are they really co-related?

With heavier payloads the successful landing rate are more for "LEO" and "ISS". Is this due to the distance between obits and earth?

What else we can find from the data sets?

Conclusion

- From the data, we concluded that the factors of the success of SpaceX has continue to improve its booster versions. From F1 to F9, and re-use parts from the booster to help save cost.
- With heavier payloads the successful landing rate are more for "LEO" and "ISS"
- SpaceX has many successful launches into the lower obit "LEO" and "ISS" but no consistent success rate with GTO.
- There seems to be no relationship between flight number in GTO orbit either

Appendix

- Data Source:

  SkillsNetwork/datasets

  Wikipedia

- Photos:

  Google images/pictures

  Pixabay