SpaceX launch analysis

L. Vidal

22.10.2023

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY

- Statistic Analysis of previous launches to understand patterns and trends influencing the launch success
- Visualisation of data from previous launches
- Model development for predicting launch success

INTRODUCTION

- Launching a space rocket is a risky and expensive initiative, data analysis can help to understand the factors that maximize the launch success
- Data from previous launches is publicly available and can be analysed with Data Science techniques
- Predictions models can be build based on historical data

METHODOLOGY

- Data collection
- Data wrangling
- Exploratory data analysis
- Prediction models: training and evaluation

Exemple of data collected

```
In [4]: df.dtypes
```

```
Out[4]: FlightNumber        int64
        Date               object
        BoosterVersion     object
        PayloadMass       float64
        Orbit              object
        LaunchSite         object
        Outcome            object
        Flights             int64
        GridFins             bool
        Reused               bool
        Legs                 bool
        LandingPad         object
        Block             float64
        ReusedCount         int64
        Serial             object
        Longitude         float64
        Latitude          float64
        dtype: object
```

```
In [2]: df=pd.read_csv("https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datase
        df.head(10)
```

Out[2]:

er	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	
1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	:
2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	:
3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	:
4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	:
5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	:
6	2014-01-06	Falcon 9	3325.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1005	-80.577366	:
7	2014-04-18	Falcon 9	2296.000000	ISS	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1006	-80.577366	:
8	2014-07-14	Falcon 9	1316.000000	LEO	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1007	-80.577366	:
9	2014-08-05	Falcon 9	4535.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1008	-80.577366	:
0	2014-09-07	Falcon 9	4428.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1011	-80.577366	:

Sources and methodology
SpaceX API: read data from public API
Wikipedia: Webscraping HTML tables

Statistics using SQL

```
In [65]: %sql select distinct "Launch_Site" from SPACEXTABLE
 * sqlite:///my_data1.db
Done.
```

Out[65]:

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

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

```
%sql select Booster_Version from SPACEXTABLE where Payload_mass__kg_ = (select max(Payl
 * sqlite:///my_data1.db
Done.
```

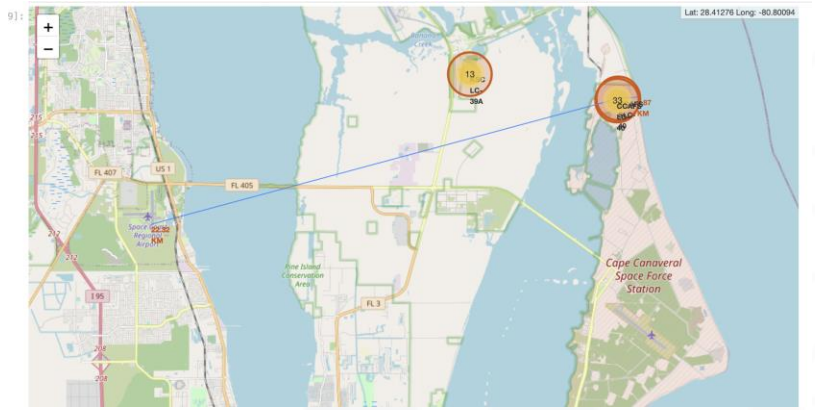
Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

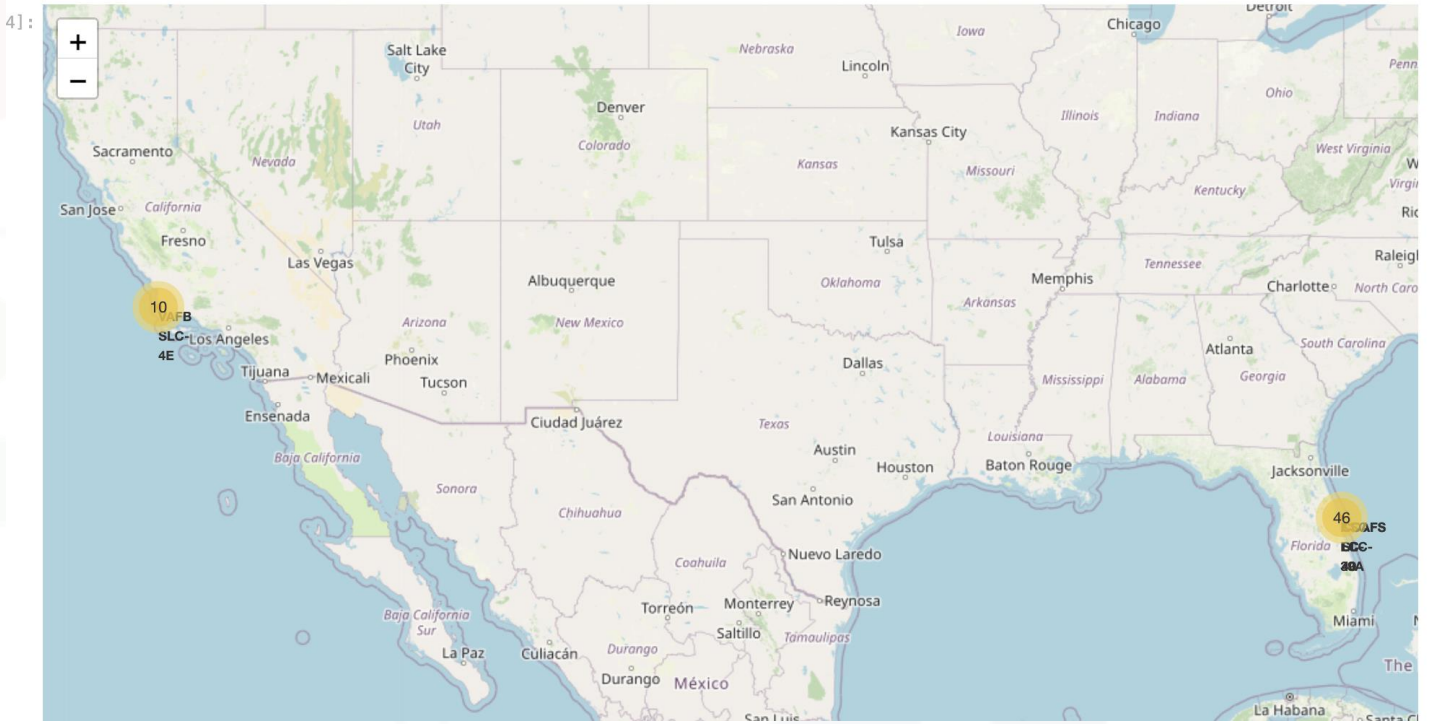
```
%%sql select Landing_Outcome, count(Landing_Outcome) as count_landing
from SPACEXTABLE
where Date between '2010-06-04' and '2017-03-20'
group by Landing_Outcome
order by count_landing desc
 * sqlite:///my_data1.db
Done.
```

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

List of launchsites
Boosters with maximum payload
Statistics for landing outcomes

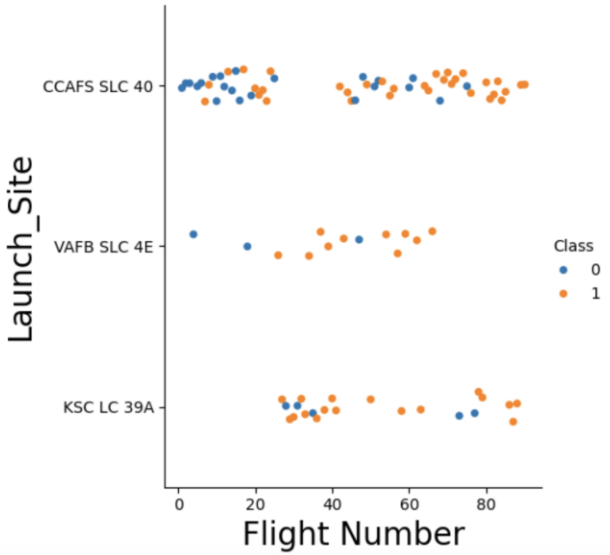
Launch locations

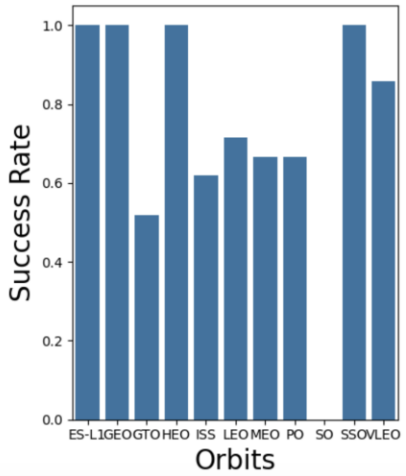


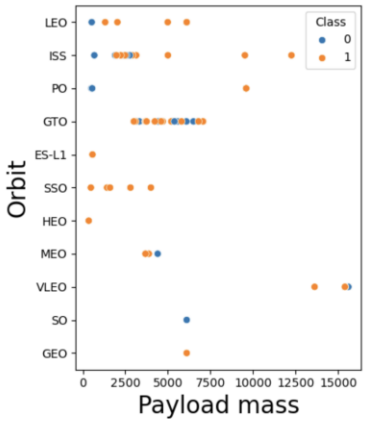


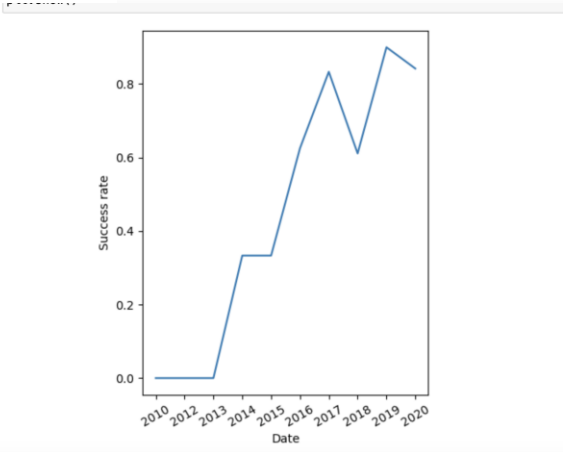
Locations near the coastline and near facilities as airports, roads, cities, etc. Launch failures can have a tremendous impact on population and infrastructure

Overview of past success launches





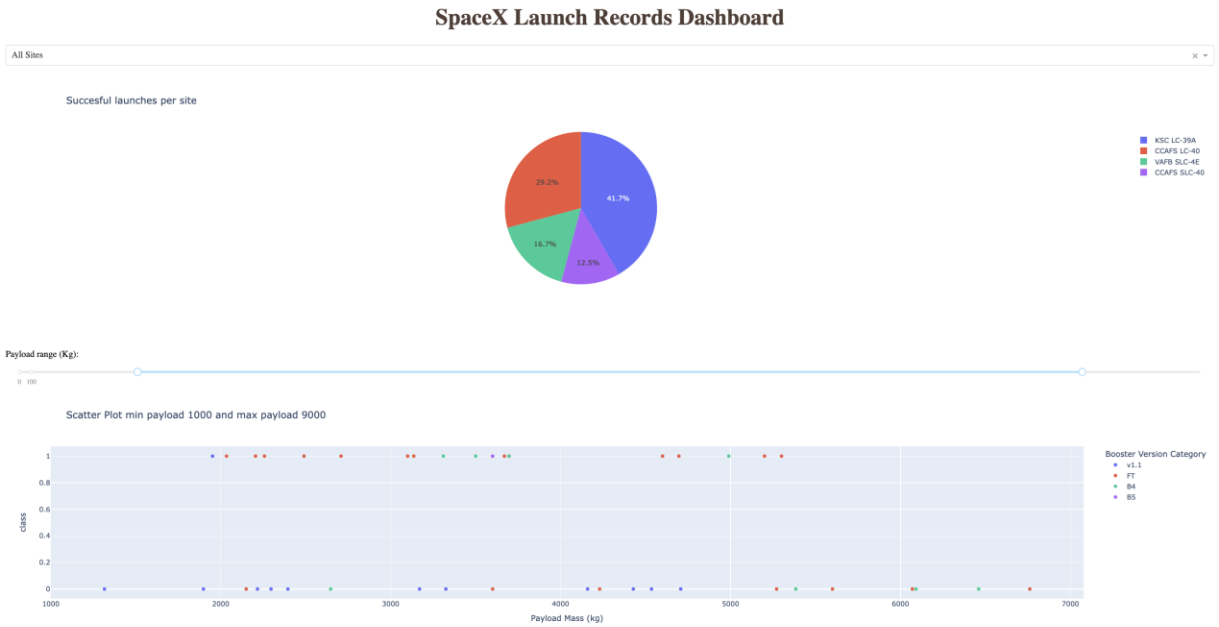




Launch success have improved since 2013

All launch sites have a history of succesful and unsuccesful launches

Application for interactive analysis

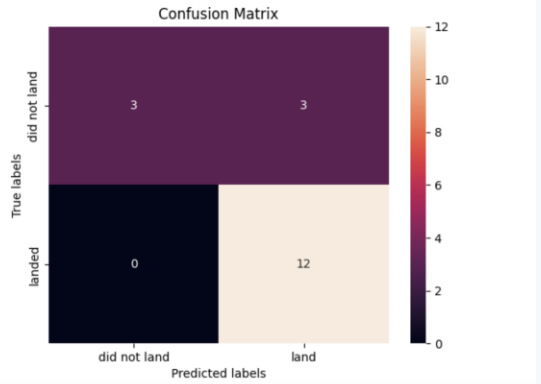


Launch success per site

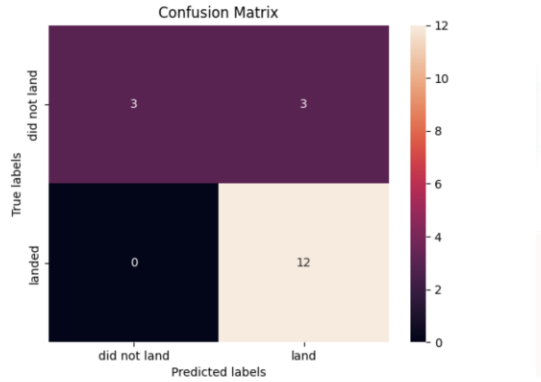
Payload range selection for class statistics

Predictive analysis

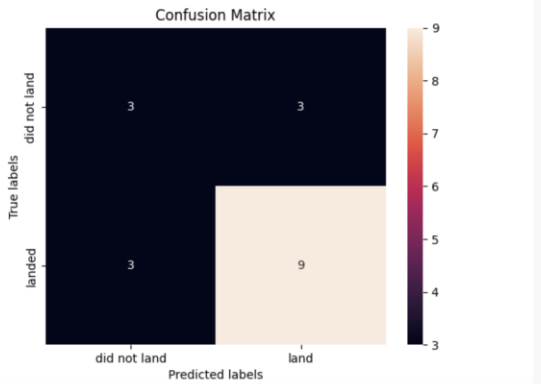
```
yhat_log=logreg_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat_log)
```



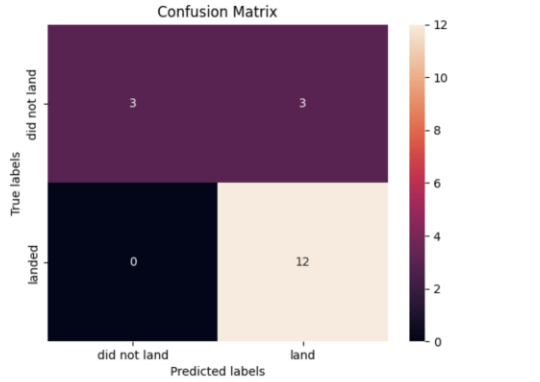

```
yhat_svm=svm_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat_svm)
```




```
yhat_tree = tree_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat_tree)
```




```
yhat = knn_cv.predict(X_test)
plot_confusion_matrix(Y_test,yhat)
```




```
logr_acc_train = logreg_cv.best_score_
svm_acc_train = svm_cv.best_score_
tree_acc_train = tree_cv.best_score_
knn_acc_train = knn_cv.best_score_

log_acc_test = logreg_cv.score(X_test,Y_test)
svm_acc_test = svm_cv.score(X_test,Y_test)
tree_acc_test = tree_cv.score(X_test,Y_test)
knn_acc_test = knn_cv.score(X_test,Y_test)

results_dict = {"LogR":[logr_acc_train,log_acc_test], "SVM":[svm_acc_train, s
                "Tree":[tree_acc_train,tree_acc_test], "KNN":[knn_acc_train,
pd.DataFrame(results_dict, index = ['Train','Test'])
```

	LogR	SVM	Tree	KNN
Train	0.846429	0.848214	0.875000	0.848214
Test	0.833333	0.833333	0.666667	0.833333

The tree model is the best fit for training data, but it gives the lowest performance with the test data

Specially costly are the false positives: launch predicted as succesful when it is not

Conclusions

- Data Science can contribute to the preparation work for rocket launch, analysing pattern behind the data collected from the previous launches
- Predictions models can be built with a high accuracy, nevertheless the cost behind false positives should be integrated in future versions of the models

Special thanks

- Teacher staff for valuable inputs in the discussion forums
- Peers that have shared questions and answers in the forums
- Peers for the promptly and constructive peer review