SpaceX Falcon 9 First stage Landing Prediction

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Insights
- Conclusion
- Appendix

EXECUTIVE SUMMARY

- Over the past few years SpaceX reached an incredible progress in Falcon 9 launches success rate. Just over 4 years, from 2014 to 2017 a % of successful launch outcomes increased from ~30% to ~80%. With a slight fluctuation in recent years, it remains outstanding: around 70-80%.
- There are 2 main geographical clusters where launches are conducted: 2 of them are in Florida(CCAFS sites, KSC LC 39A) and 1 is in California(VAFB SLC 4E). KSC LC-39A has the highest success rate compared to the other launch sites.
- Launch success rate has certain influences from the orbit and payload mass.

  Overall, the higher payload mass is the higher the success rate. However, heavy payloads have a negative influence on GTO orbits and positive on Polar, LEO and ISS orbits.

  Certain orbits had 100% success Rate: ES-L1,GEO,HEO and SSO.

- Support Vector Machine model predicts with a high level of accuracy 83% if the stage one would land.

INTRODUCTION

- Since 2010, SpaceX company has performed launches of Falcon9. Over just a few years, it gathered a great amount of data to perform a predictive analysis of launches in order to estimate whether the outcome of a stage 1 would be successful or not with an established set of parameters.

- The prediction plays a crucial role in setting a launch price which has a key importance in market positioning of the company and finance management. The module can be reused after the Stage 1; therefore, the prediction directly influences the cost of the program.

SKILLS NETWORK

Methodology

Methodology: summary

- Data collection and data wrangling
- EDA and interactive visual data analytics
- Predictive analysis

Data collection and data wrangling

SPACEX LAUNCHES DATA COLLECTION

LAUNCHES DATA WRANGLING

HISTORICAL DATA COLLECTION

Launches data collection

- **SpaceX launch data collection**
  - Import libraries and define auxiliary functions*
  - Collect data from the API https://api.spacexdata.com/v4/launches/past
  - Turn the data into a dataframe using response and json_normalize methods
  - Using the API data, get the values for columns *rocket, payloads, launchpad*, and *cores*
  - Create a dictionary with the column titles
  - Create a Pandas dataframe using the dictionary
- **Data processing**
  - Filtering: filter on Falcon 9 launches
  - Data cleansing: remove missing values using mean method

** Functions are: getBoosterVersion(data), getLaunchSite(data), getPayloadData(data), getCoreData(data)*

GitHub link

SKILLS NETWORK

Data wrangling

- Using the dataset, identify and calculate the percentage of the missing values in each attribute, using isnull method. Identify which columns are numerical and categorical using dtypes method.
- Using value_counts, apply the method to the columns LaunchSite, Orbit, and Outcome in order to get the following:
  - Calculate the number of launches for each site
  - Calculate the number and occurrence of each orbit
  - Calculate the number and occurrence of mission outcome per orbit type
- Creating a landing outcome label using condition/cycle. This variable labels successful and failed outcomes

```
# landing_class = 0 if bad_outcome
# landing_class = 1 otherwise
landing_class=[]
for i in range(0,len(df['Outcome'])):
    if df['Outcome'][i] in bad_outcomes:
        landing_class.append(0)
    else:
        landing_class.append(1)
print(landing_class)
```

[GitHub link]

Historical data collection

Web scrap Falcon 9 launch records using BeautifulSoup:

- Install the libraries requests and BeautifulSoup
- Define the functions to
- Extract a Falcon 9 launch records HTML table from Wikipedia using the page URL and BeautifulSoup library. Collect all relevant column names from the HTML table header.
- Parse the table and convert it into a Pandas data frame.
  - Create an empty dictionary
  - Fill it up with the data using append method
  - Create a dataframe from the dictionary using standard pd.DataFrame method
  - Export the data as csv to the file called “spacex_web_scraped.csv”

GitHub link

EDA & Interactive data analytics

VISUAL DATA EXPLORATION

BAR CHARTS & SCATTER PLOTS DESCRIBING KEY OBSERVATIONS

LAUNCHES GEOGRAPHICAL AND QUALITATIVE FEATURES

INTERACTIVE MAP AND DASH APPLICATION

PREDICTIVE ANALYTICS

TESTING VARIOUS MODELS AND SELECTING THE MOST SUITABLE CLASSIFICATION APPROACH

EDA with Data Visualization

- Demonstrating the key relations between variables and KPIs trends using visual graphs   GitHub link

  The following KPIs were analyzed

  1. Overall Success Rate per Orbit type
  2. How Success Rate per each Orbit type varies depending on Flight Number
  3. Payload mass
  4. Number of flights

- Descriptive analysis using SQL. SQL Queries used are mentioned on respective slides.   GitHub link

Launches geographical distribution

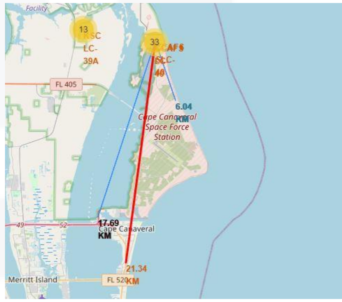
With the goal to visualize the launch sites and their respective landing outcomes, all the elements mentioned above were placed on a map.

Markers demonstrate launch outcomes, with red being unsuccessful, and 1 being green I.e., successful landing.

A polyline object was added to the map to show the distance from the launch site to its nearby objects such as nearest railway station. Haversine's formula is used to calculate the distance between objects

Positive outcome

Negative outcome



Polyline object

GitHub link

Launches qualitative features Dash application

An interactive web application using dash which includes the following features:

- Drop-down menu to select the launch site and range-slider to choose the range of payload.
- Interactive pie chart showing success rate of all launch sites by default
- Scatter plot showing launch outcomes of all sites according to their payloads in the default range

**I didn't manage to accomplish this part, as dash feature had bugs and didn't update the app. As I didn't figure out how to add things to the skeleton, I described the methodology, but missing the results*

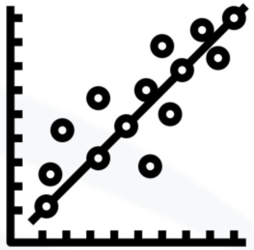
GitHub link

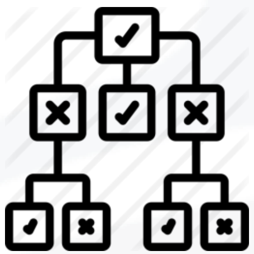
SKILLS NETWORK

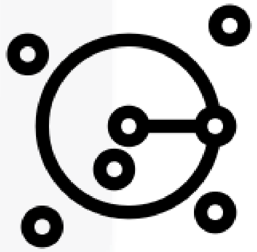
Predictive Analytics (Classification model)

- In order to fins the model best fit for the data, several models were pre-selected: logistic regression, decision tree, SVM and KNN. The KPI which determines a better fit is Accuracy.
- The data is split into a train and test subsets. Each model was trained using train data and then applied to test data. Accuracy is a measure which shows a % of accurate predictions vs all predictions.

Logistic regression	SVM	Decision tree	KNN





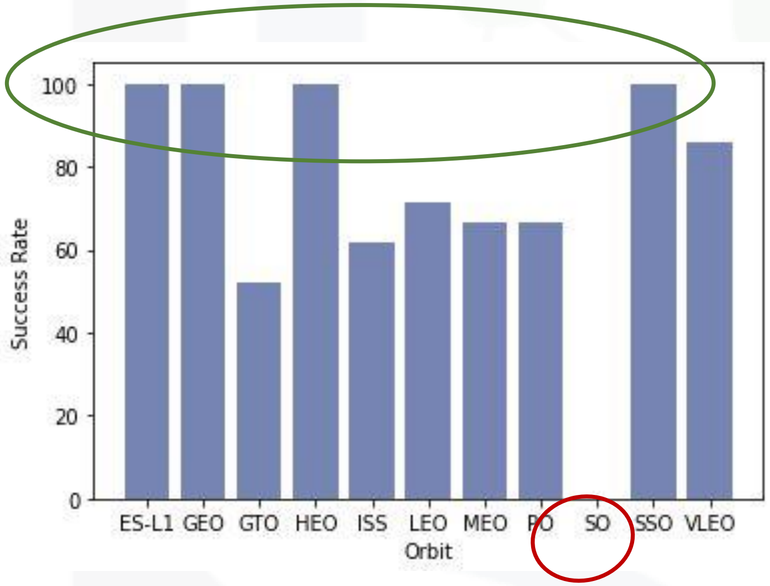


[GitHub link]

SKILLS NETWORK


Insights
SPACEX

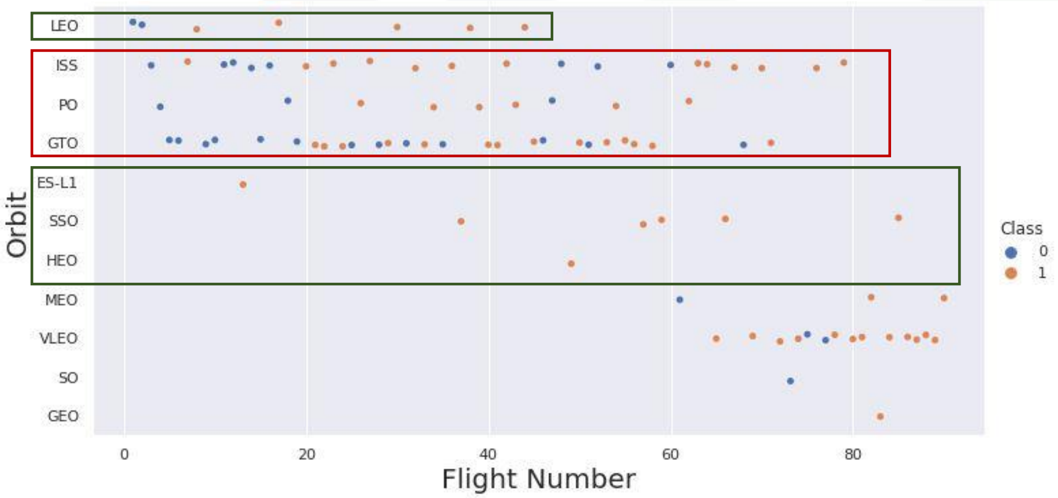
Success Rate vs. Orbit Type



- The most successful orbits were ES-L1,GEO,HEO and SSO.
- SO Orbit, in the opposite, has a zero-success rate.

GitHub link

Flight Number vs. Orbit Type



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

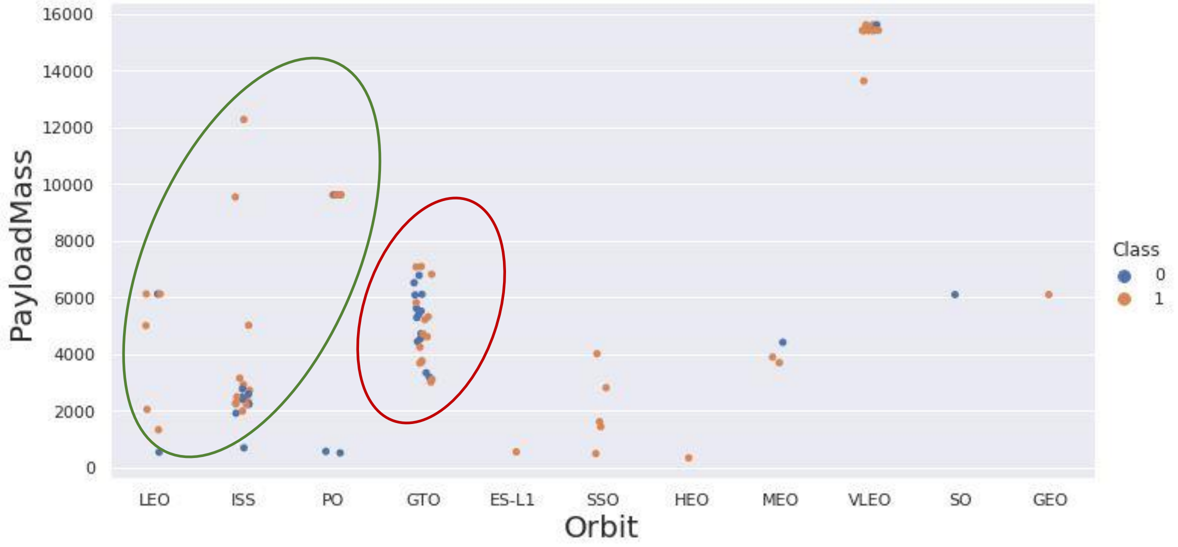
There seems to be no relationship between flight number in GTO, ISS and PO orbits

A vast majority of the orbits having 100% successful rate (see prev. slide) have a relatively low variation of Flight numbers (ES-L1, HEO, GEO), which might hold a potential caveat.

SSO Orbit shows a consistent level of success across the range.

[GitHub link]

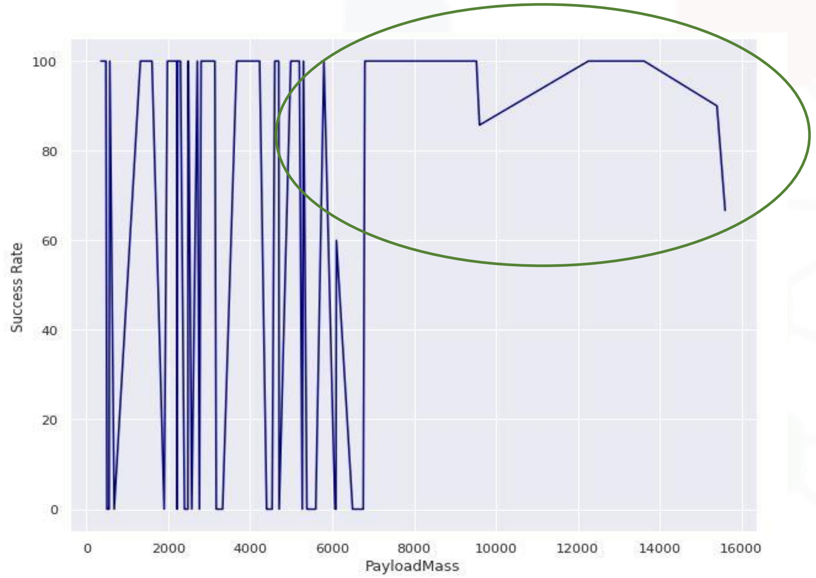
Payload vs. Orbit Type

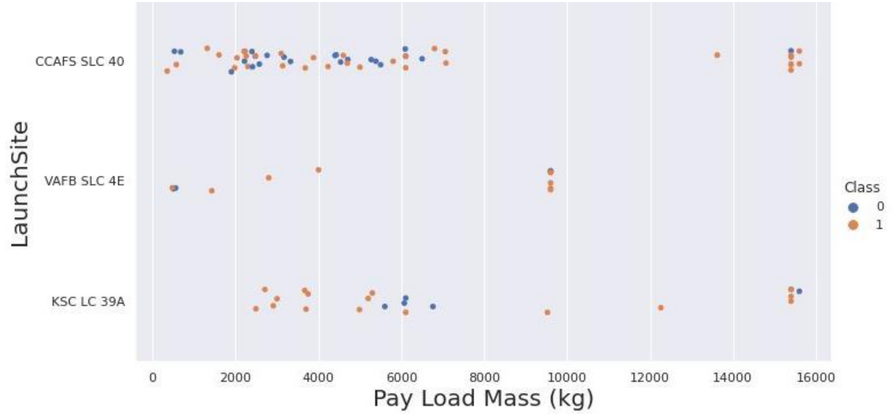


Heavy payloads have a negative influence on GTO orbits and positive on Polar, LEO and ISS orbits.

[GitHub link]

Payload vs. Launch site





The best success rate values are achieved at the high amounts of payload.

CCAFS SLC 40: success is more probable if Payload Mass is over 13000kg.

VADS SLC 4CC: success is probable if Payload Mass is over 1000kg.

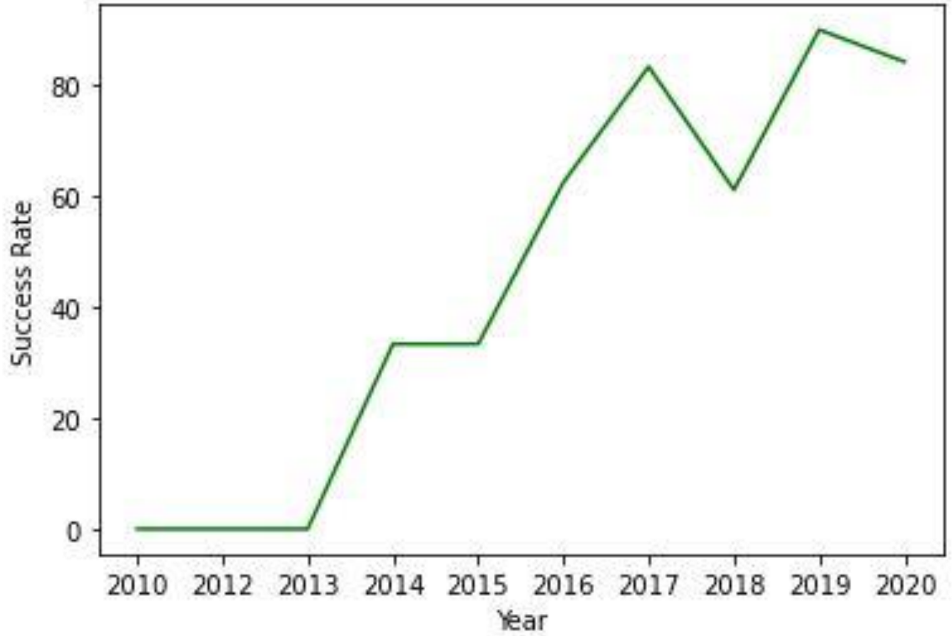
KSL LC 39A: mass interval between 5500 and 7000kg. there is a higher chance of failure in the payload

GitHub link

IBM Developer

SKILLS NETWORK

Launch Success Yearly Trend



We can observe that the success rate reached 80% mark in a relatively short span of 4 years (2014-2017).

With a slight fluctuation in recent years, it remains around 70-80% which shows an incredible progress achieved by Space X company.

GitHub link

Launch Site Names

The names of the all launch sites are

GitHub link

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

5 booster versions launched from KSC LC-39A site

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_...	Orbit	Customer	Mission_Outcome	Landing _Outcome
19-02-2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground ...
16-03-2017	06:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attempt
30-03-2017	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone s...
01-05-2017	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground ...
15-05-2017	23:21:00	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070	GTO	Inmarsat	Success	No attempt

```
SELECT DISTINCT(launch_site) FROM SPACEXTBL

SELECT * FROM SPACEXTBL WHERE launch_site like 'KSC%' limit 5
```

Overall observations(payload mass and timing)

```
The total payload mass carried by boosters launched by NASA (CRS)
Average payload mass carried by booster version F9 v1.1
The date where the successful landing outcome in drone ship was achieved
```

Total and average payload masses can be calculated using respective aggregate functions (sum() and avg()). In order to calculate to first successful launch date, I used min() function. For the simplicity of illustration, I put the first two results together using UNION.

We can see that the boosters altogether launched by Nasa, carried 45.6t of payload.

In average, one booster F9 v.1.1 carried 2.5t of payload.

The first successful landing outcome in drone ship was achieved on 2016/04/08.

KPI	value
Average Payload mass	2534.7
Total payload mass, Nasa (CRS)	45596.0

KPI	Date
Date of the first successful drone landing	2016-04-08

GitHub link

```
SELECT 'Total payload mass, Nasa (CRS)' as KPI, sum(payload_mass__kg_) as value from SPACEXTBL where customer = 'NASA (CRS)' group by KPI
UNION
SELECT 'Average Payload mass' as KPI, round(avg(payload_mass__kg_),1) as value FROM SPACEXTBL WHERE booster_version like 'F9 v1.1%' group by KPI
-----------
SELECT 'Date of the first successful drone landing' as KPI, min(to_date(date,'dd-MM-yyyy')) as value FROM SPACEXTBL WHERE landing_outcome like 'Success (drone%' group by KPI
```

Boosters carrying 4000-6000

The boosters which have success in ground pad and have payload mass greater than 4000 but less than 6000T

For selecting landing outcomes and payload masses, WHERE condition should be used. There are 3 boosters corresponding to these conditions. It is listed in the table below.

booster_version	payload_mass__kg_
F9 FT B1032.1	5300
F9 B4 B1040.1	4990
F9 B4 B1043.1	5000

GitHub link

```
SELECT booster_version , payload_mass__kg_
FROM SPACEXTBL
WHERE (landing_outcome like 'Success (ground%') AND (payload_mass__kg_ BETWEEN 4000
AND 6000)
```

Successful and Failure Mission Outcomes

Using case when condition we can label successful and failure landing outcomes and this way. We can see that in this case there is no unclassified mission outcomes.

In total, there were 101 mission outcomes, ~99% of which were successful.

mission_outcome	count
Failure	1
Success	100
TOTAL	101

GitHub link

```
(SELECT case when mission_outcome like 'Success%' then 'Success' else (case when mission_outcome like 'Failure%' then 'Failure' else 'Other' end) end as mission_outcome, count(*) as count from SPACEXTBL GROUP BY 1)
UNION
(SELECT "TOTAL" as mission_outcome, count(*) as count from SPACEXTBL GROUP BY 1)
ORDER BY 1
```

Successful and Failure Landing Outcomes

Using case when condition we can label successful and failure landing outcomes and this way. We can see that there is a certain number of unclassified launches which fell into a group “Other”. After looking through them, it is clear that these can be assigned to ‘Failures’ as well.

In total, there were 101 launches, 61% of which had successful landing outcomes.

landing_outcome	count
Controlled (ocean)	5
Failure	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	21
No attempt	1
Precluded (drone ship)	1
Success	38
Success (drone ship)	14
Success (ground pad)	9
Uncontrolled (ocean)	2
Z_TOTAL	101

landing_outcome	count
Failure	10
Other	30
Success	61
TOTAL	101

GitHub link

```
(SELECT case when landing_outcome like 'Success%' then 'Success' else (case when landing_outcome like 'Failure%' then 'Failure' else 'Other' end) end as
landing_outcome, count(*) as count from SPACEXTBL GROUP BY 1)
UNION
(SELECT "TOTAL" as landing_outcome, count(*) as count from SPACEXTBL GROUP BY 1)
ORDER BY 1
```

Boosters Carried Maximum Payload

To get the names of the booster which have carried the maximum payload mass, each value of payload mass is compared to maximum mass and the ones which individual mass is equal to the max column value are selected.

The maximum payload mass per booster type is 15.6t. There are 12 booster versions which correspond to this criteria

GitHub link

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

Select booster_version, payload_mass__kg_ from SPACEXTBL where payload_mass__kg_=(select max(payload_mass__kg_) from SPACEXTBL)

2017 Successful ground pad launches

The records which will display the month names, successful landing outcomes in ground pad ,booster versions, launch site for the months in year 2017

GitHub link

month_name	landing_outcome	booster_version	launch_site
February	Success (ground pad)	F9 FT B1031.1	KSC LC-39A
May	Success (ground pad)	F9 FT B1032.1	KSC LC-39A
June	Success (ground pad)	F9 FT B1035.1	KSC LC-39A
August	Success (ground pad)	F9 B4 B1039.1	KSC LC-39A
September	Success (ground pad)	F9 B4 B1040.1	KSC LC-39A
December	Success (ground pad)	F9 FT B1035.2	CCAFS SLC-40

```
with spacex_data as (Select right(left(date,5),2) AS month_num, right(date,4) as year, landing_outcome, booster_version, launch_site FROM SPACEXTBL)
 select case when month_num='01' then 'January' else
    (case when month_num='02' then 'February' else
     (case when month_num='03' then 'March' else
      (case when month_num='04' then 'April' else
       (case when month_num='05' then 'May' else
        (case when month_num='06' then 'June' else
         (case when month_num='07' then 'July' else
          (case when month_num='08' then 'August' else
           (case when month_num='09' then 'September' else
            (case when month_num='10' then 'October' else
             (case when month_num='11' then 'November' else
              (case when month_num='12' then 'December' else '-' end)end)end)end)end)end)end)end)end)end)end)end as month_name,
landing_outcome, booster_version, launch_site
from spacex_data
where year=2017 and landing_outcome='Success (ground pad)'
order by landing_outcome, month_name, launch_site, booster_version;
```

Successful landing outcomes over the period of 2010/06/04 – 2017/03/20

- Over the 7 years of the period mentioned there were only 8 successful launches.
- 5 of them were related to drone ship and 3 - ground pad.

GitHub link

landing_outcome	count
Success (drone ship)	5
Success (ground pad)	3

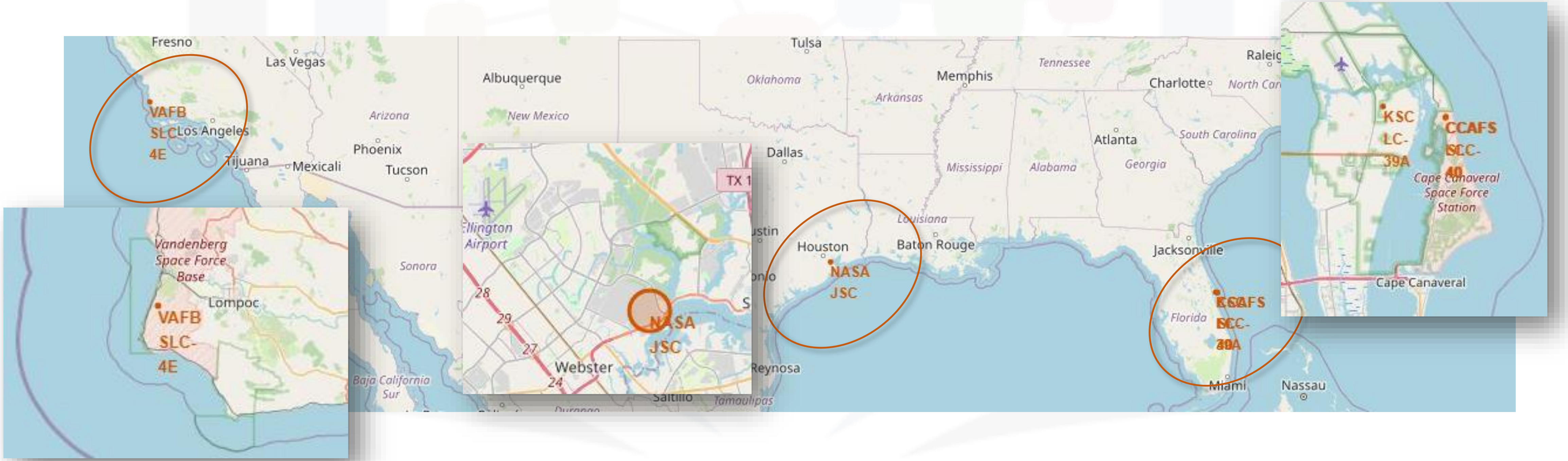
SELECT landing_outcome,count(*) as count FROM cos://us-east/nmalkovaspacex/Spacex.csv SPACEXTBL WHERE to_date(Date,'dd-MM-yyyy') between to_date('04-06-2010','dd-MM-yyyy') AND to_date('20-03-2017','dd-MM-yyyy') and landing_outcome like 'Success%’
GROUP BY landing_outcome ORDER BY count DESC;

Plotting launch sites on map

There are 3 main clusters where launch sites are situated:

- Florida (Cape Canaveral) – KSC and CCAFS
- California - VAFB

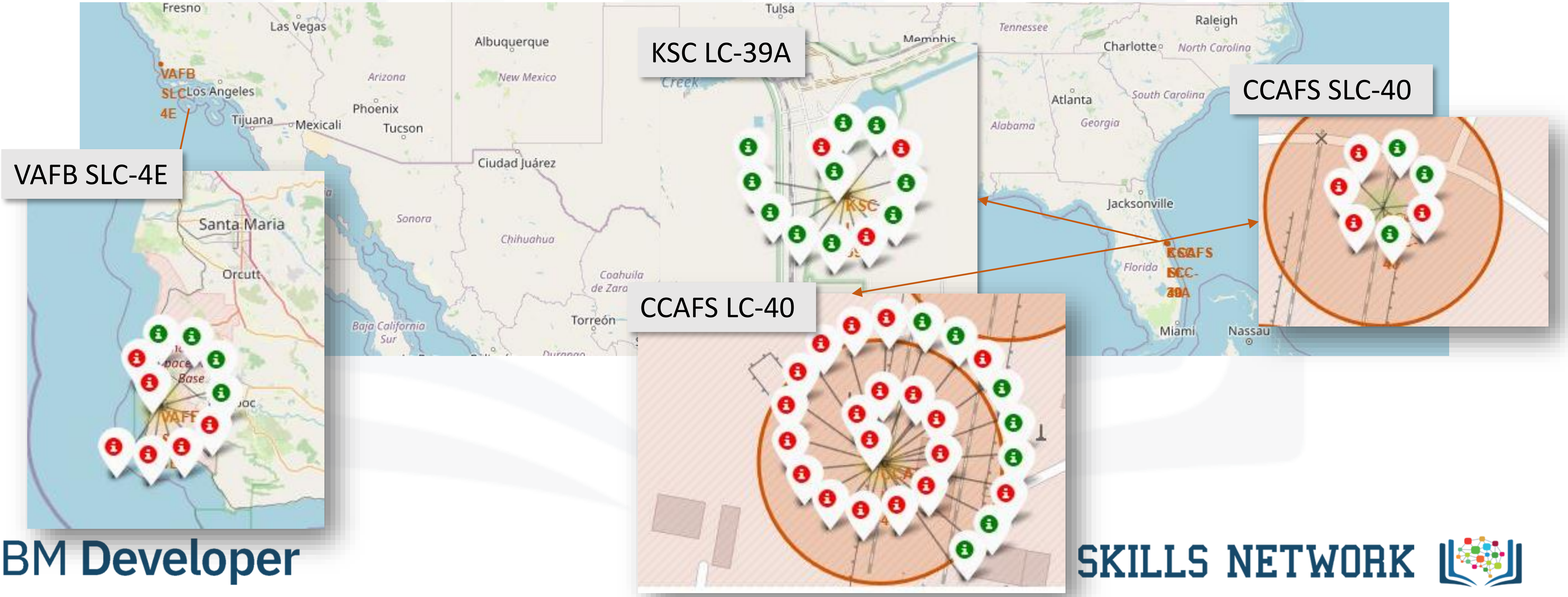
GitHub link



Marking outcomes

Green and red markers demonstrate successful and failed launches. It is clearly seen that KSL LC-39A site show a better success rate compared to other sites.

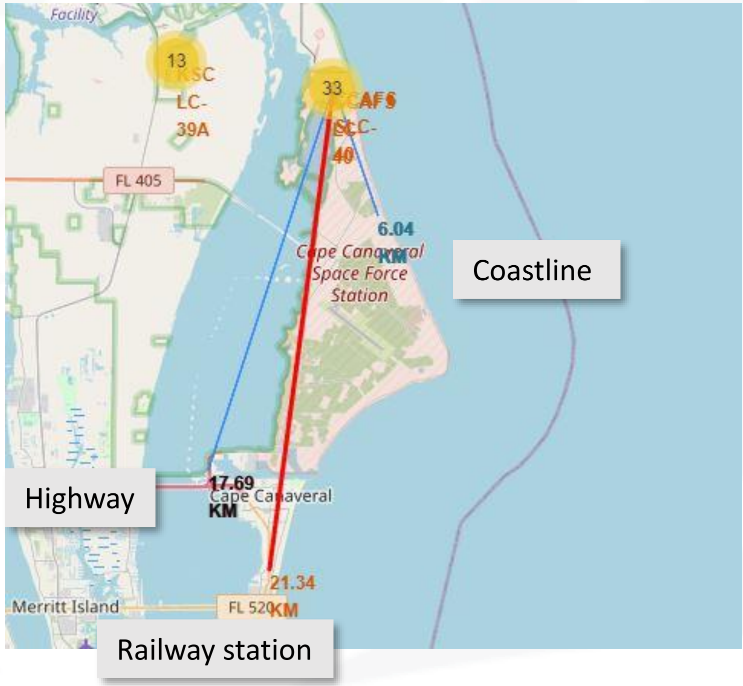
GitHub link



Distance to proximities

CCAFS sites are quite close to the coastline whereas it takes quite a bit to get to the Highways and Railways.

[GitHub link]



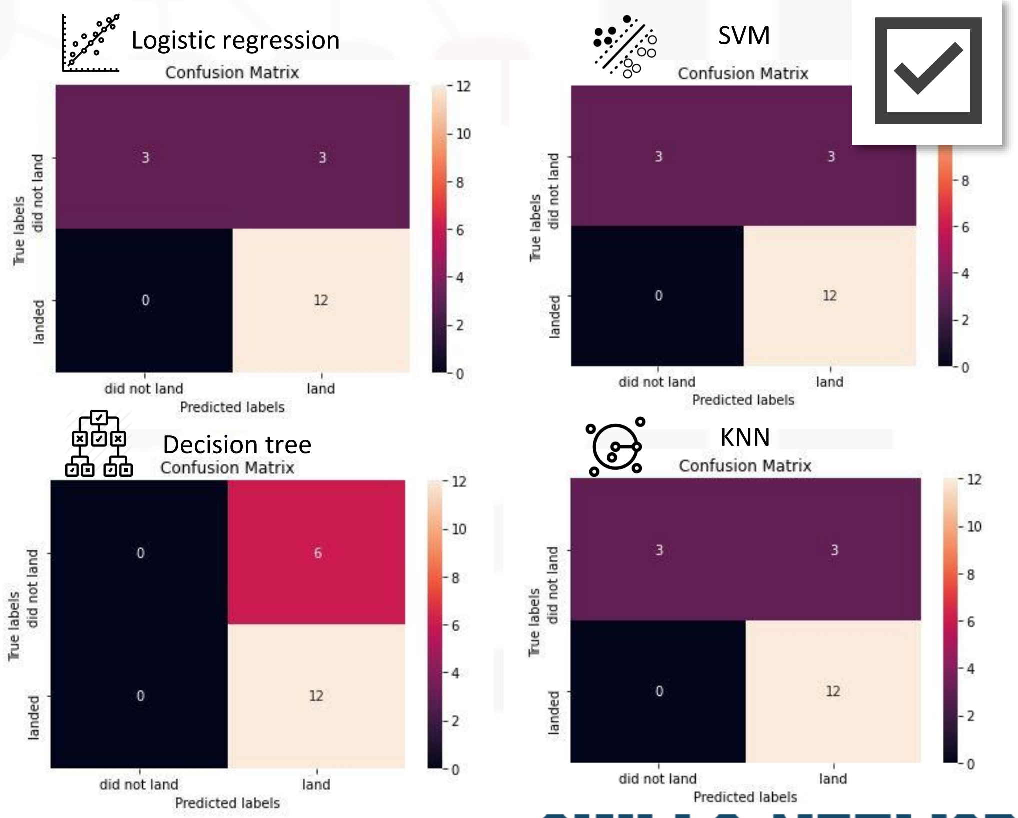
Landing prediction: classification model

Logistic regression, KNN and SVM achieved the same level of test data accuracy 83.3% whereas decision tree model performed worse than other models (66.7%). For the prediction, SVM model would be a better fit as being a robust model for high-dimensional data sets, which corresponds to our case.

	ModelType	Value
0	logreg	0.833333
1	svm	0.833333
2	tree	0.666667
3	knn	0.833333

Best score is: 0.8333333333333334

[GitHub link]



Conclusion

- Over the past few years SpaceX reached an incredible progress in Falcon 9 launches success rate. Just over 4 years, a % of successful launch outcomes increased from ~30% to ~80%. With a slight fluctuation in recent years, it remains outstanding: around 70-80%.
- There are 3 main geographical clusters where launches are conducted: 3 of them are situated in Florida(CCAFS sites, KSC LC 39A) and 1 is in California(VAFB SLC 4E). KSC LC-39A has the highest success rate compared to the other launch sites.
- Launch success rate has certain influences from the orbit and payload mass. Overall, the higher payload mass is the higher the success rate. However, heavy payloads have a negative influence on GTO orbits and positive on Polar, LEO and ISS orbits. Certain orbits had 100% success Rate: ES-L1,GEO,HEO and SSO.
- Support Vector Machine was a suitable model to predict if the stage one would land. This model reached a high level of accuracy (83%) and is appropriate for high-dimensional data-sets, which makes it a good fit.

Appendix

Appendix 1. Data wrangling. Resulting table

Out[27]:

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005	-80.577366	28.561857
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007	-80.577366	28.561857
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003	-120.610829	34.632093
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004	-80.577366	28.561857
9	6	2014-01-06	Falcon 9	3325.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1005	-80.577366	28.561857
10	7	2014-04-18	Falcon 9	2296.0	ISS	CCSFS SLC 40	True Ocean	1	False	False	True	None	1.0	0	B1006	-80.577366	28.561857
11	8	2014-07-14	Falcon 9	1316.0	LEO	CCSFS SLC 40	True Ocean	1	False	False	True	None	1.0	0	B1007	-80.577366	28.561857
12	9	2014-08-05	Falcon 9	4535.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1008	-80.577366	28.561857
13	10	2014-09-07	Falcon 9	4428.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1011	-80.577366	28.561857
14	11	2014-09-21	Falcon 9	2216.0	ISS	CCSFS SLC 40	False Ocean	1	False	False	False	None	1.0	0	B1010	-80.577366	28.561857
15	12	2015-01-10	Falcon 9	2395.0	ISS	CCSFS SLC 40	False ASDS	1	True	False	True	5e9e3032383ecb761634e7cb	1.0	0	B1012	-80.577366	28.561857
16	13	2015-02-11	Falcon 9	570.0	ES-L1	CCSFS SLC 40	True Ocean	1	True	False	True	None	1.0	0	B1013	-80.577366	28.561857
17	14	2015-04-14	Falcon 9	1898.0	ISS	CCSFS SLC 40	False ASDS	1	True	False	True	5e9e3032383ecb761634e7cb	1.0	0	B1015	-80.577366	28.561857
18	15	2015-04-27	Falcon 9	4707.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1016	-80.577366	28.561857
19	16	2015-06-28	Falcon 9	2477.0	ISS	CCSFS SLC 40	None ASDS	1	True	False	True	5e9e3032383ecb6bb234e7ca	1.0	0	B1018	-80.577366	28.561857
20	17	2015-12-22	Falcon 9	2034.0	LEO	CCSFS SLC 40	True RTLS	1	True	False	True	5e9e3032383ecb267a34e7c7	1.0	0	B1019	-80.577366	28.561857

Appendix 2. Data wrangling. Resulting table

	Flight No.		Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	NaN	Version Booster	Booster landing	Date	Time
0	1	None	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	None	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	None	CCAFS	Dragon	0	LEO	\n	Success	None	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	None	CCAFS	Dragon	525 kg	LEO	NASA	Success	None	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	None	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	None	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	None	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	None	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10

Appendix 3 Links

Image:

- https://po.al/spacex-ka-480-satelite-a-do-kemi-internet-satelitor-ne-cilesine-e-rrjeteve-tokesore/nmwasupmrfzy3ayztk55ys/