Winning Space Race with Data Science

Sardorbek Bobomurodov
20.02.2022

Outline

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

Executive Summary

- SpaceY is a new commercial rocket launch provider who wants to bid against SpaceX.
- SpaceX advertises launch services starting at $62 million for missions that allow some fuel to be reserved for landing the 1st stage rocket booster, so that it can be reused.
- Given mission parameters such as payload mass and desired orbit, the models produced in this report were able to predict the first stage rocket booster landing successfully with an accuracy level of 83.3%.

Introduction

- This report has been prepared as part of the IBM Data Science Professional Specialization course.
- With the help of the data science findings and models in this report, SpaceY will be able to make more informed bids against SpaceX for a rocket launch.

Section
1

Methodology

Methodology

Executive Summary

- Data collection methodology:
- Data wrangling
- Exploratory data analysis (EDA) using visualization and SQL
- Performing interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models

Data Collection

API

- Acquired historical launch data from Open Source REST API for SpaceX
- Requested and parsed the SpaceX launch data using the GET request
- Filtered the dataframe to only include Falcon 9 launches
- Replaced missing payload mass values from classified missions with mean

Web Scrapping

- Requested the Falcon9 Launch Wiki page from its Wikipedia URL
- Extracted all column/variable names from the HTML table header
- Parsed the table and converted it into a Pandas data frame

Data Collection – SpaceX API

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	Reused
0	1	2006-03-24	Falcon 1	None	None	Kwajalein Atoll	None None	1	False	False	False	None	None	0
1	2	2007-03-21	Falcon 1	None	None	Kwajalein Atoll	None None	1	False	False	False	None	None	0
2	4	2008-09-28	Falcon 1	None	None	Kwajalein Atoll	None None	1	False	False	False	None	None	0
3	5	2009-07-13	Falcon 1	None	None	Kwajalein Atoll	None None	1	False	False	False	None	None	0
4	6	2010-06-04	Falcon 9	None	None	CCSFS SLC 40	None None	1	False	False	False	None	1	0

This is our dataset that we get from API before filtering.

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	Reus
4	1	2010-06-04	Falcon 9	None	None	CCSFS SLC 40	None None	1	False	False	False	None	1	0
5	2	2012-05-22	Falcon 9	None	None	CCSFS SLC 40	None None	1	False	False	False	None	1	0
6	3	2013-03-01	Falcon 9	None	None	CCSFS SLC 40	None None	1	False	False	False	None	1	0
7	4	2013-09-29	Falcon 9	None	None	VAFB SLC 4E	False Ocean	1	False	False	False	None	1	0
8	5	2013-12-03	Falcon 9	None	None	CCSFS SLC 40	None None	1	False	False	False	None	1	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
183	180	None	None	None	None	KSC LC 39A	None	None	None	None	None	None	None	None
184	181	None	None	None	None	KSC LC 39A	None	None	None	None	None	None	None	None
185	182	None	None	None	None	KSC LC 39A	None	None	None	None	None	None	None	None
186	183	None	None	None	None	CCSFS SLC 40	None	None	None	None	None	None	None	None
187	184	None	None	None	None	CCSFS SLC 40	None	None	None	None	None	None	None	None

This is our after applying changes and filtering. [Link for the whole source code.]

Data Collection - Scraping

- Another way of collecting data was web scrapping. To do I used data from Wikipedia.
- Applying changes, filtering, formating.
- Link for the whole source code.

[hide] Flight No.	Date and time (UTC)	Version, Booster [b]	Launch site	Payload[c]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
1	4 June 2010, 18:45	F9 v1.0[7] B0003[8]	CCAFS, SLC-40	Dragon Spacecraft Qualification Unit	No payload (excl. Dragon Mass)	LEO	SpaceX	Success	Failure[9][10] (parachute)
	First flight of Falcon 9 v1.0.[11] Used a boilerplate version of Dragon capsule which was not designed to separate from the second stage.(more details below) Attempted to recover the first stage by parachuting it into the ocean, but it burned up on reentry, before the parachutes even got to deploy.[12]								
2	8 December 2010, 15:43[13]	F9 v1.0[7] B0004[8]	CCAFS, SLC-40	Dragon demo flight C1 (Dragon C101)	Classified (excl. Dragon Mass)	LEO (ISS)	NASA (COTS) NRO	Success[9]	Failure[9][14] (parachute)
	Maiden flight of SpaceX's Dragon capsule, consisting of over 3 hours of testing thruster maneuvering and then reentry.[15] Attempted to recover the first stage by parachuting it into the ocean, but it disintegrated upon reentry, again before the parachutes were deployed.[12] (more details below) It also included two CubeSats,[16] and a wheel of Brouère cheese. Before the launch, SpaceX discovered that there was a crack in the nozzle of the 2nd stage's Merlin vacuum engine. So Elon just had them cut off the end of the nozzle with a pair of shears and launched the rocket a few days later. After SpaceX had trimmed the nozzle, NASA was notified of the change and they agreed to it.[17]								
3	22 May 2012, 07:44[18]	F9 v1.0[7] B0005[8]	CCAFS, SLC-40	Dragon demo flight C2+[19] (Dragon C102)	525 kg (1,157 lb)[20] (excl. Dragon mass)	LEO (ISS)	NASA (COTS)	Success[21]	No attempt
	The Dragon spacecraft demonstrated a series of tests before it was allowed to approach the International Space Station. Two days later, it became the first commercial spacecraft to board the ISS.[18] (more details below)								
4	8 October 2012, 00:35[22]	F9 v1.0[7] B0006[8]	CCAFS, SLC-40	SpaceX CRS-1[23] (Dragon C103)	4,700 kg (10,400 lb) (excl. Dragon mass)	LEO (ISS)	NASA (CRS)	Success	No attempt
				Orbcomm-OG2[24]	172 kg (379 lb)[25]	LEO	Orbcomm	Partial failure[26]	

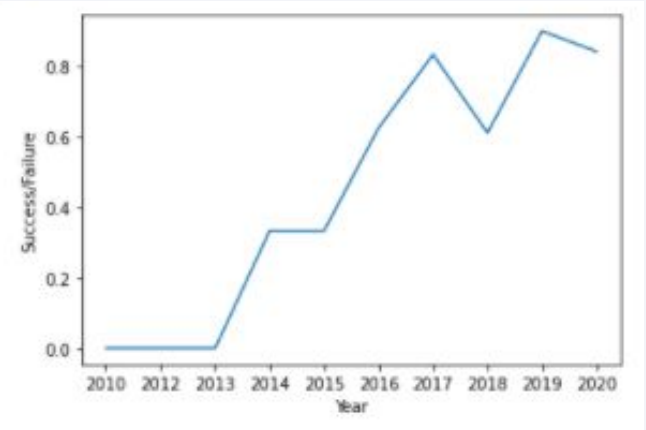
Data Wrangling

True ASDS	41
None None	19
True RTLS	14
False ASDS	6
True Ocean	5
False Ocean	2
None ASDS	2
False RTLS	1

- Explored data to determine the label for training supervised models
  - Calculated the number of launches on each site
  - Calculated the number and occurrence of each orbit
  - Calculated the number and occurrence of mission outcome per orbit type
- Created a landing outcome training label from 'Outcome' column
  - Class = 0; first stage booster did not land successfully
  - Class = 1; first stage booster landed successfully
- Link for the whole source code.

EDA with Data Visualization

- Used Matplotlib and Seaborn visualization libraries to plot
  - Relationship between Flight Number and Launch Site
  - Relationship between Payload and Launch Site
  - Relationship between success rate of each orbit type
  - Relationship between FlightNumber and Orbit type
  - Relationship between Payload and Orbit type
  - The launch success yearly trend



Link for the whole source code.

EDA with SQL

- Loaded data into an IBM DB2 instance
- Ran several queries to get new and display information about:
  - Unique launch sites in the space mission
  - Launch sites begin with the string 'CCA'
  - Total payload mass carried by boosters launched by NASA (CRS)
  - Average payload mass carried by booster version F9 v1.1
  - Date when the first successful landing outcome in ground pad was acheived.
  - Total number of successful and failure mission outcomes

Link for the whole source code.

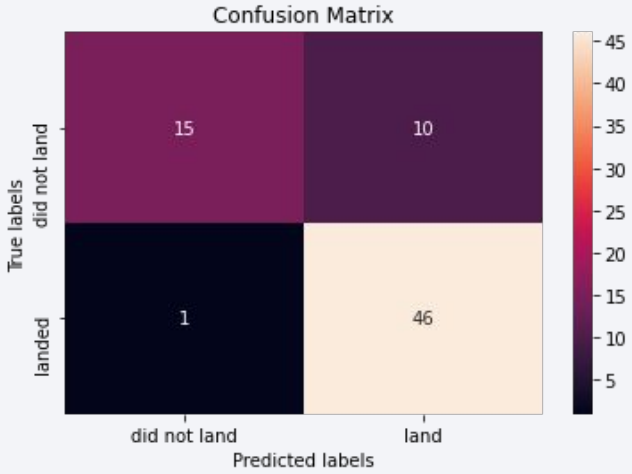
Build an Interactive Map with Folium

Above screenshot is for identifing which launch sites have relatively high success rates.

- Marked all launch sites on a map
- Marked the success/failed launches for each site on the map
- Calculated the distances between a launch site to its proximities

Link for the whole source code.

Predictive Analysis (Classification)



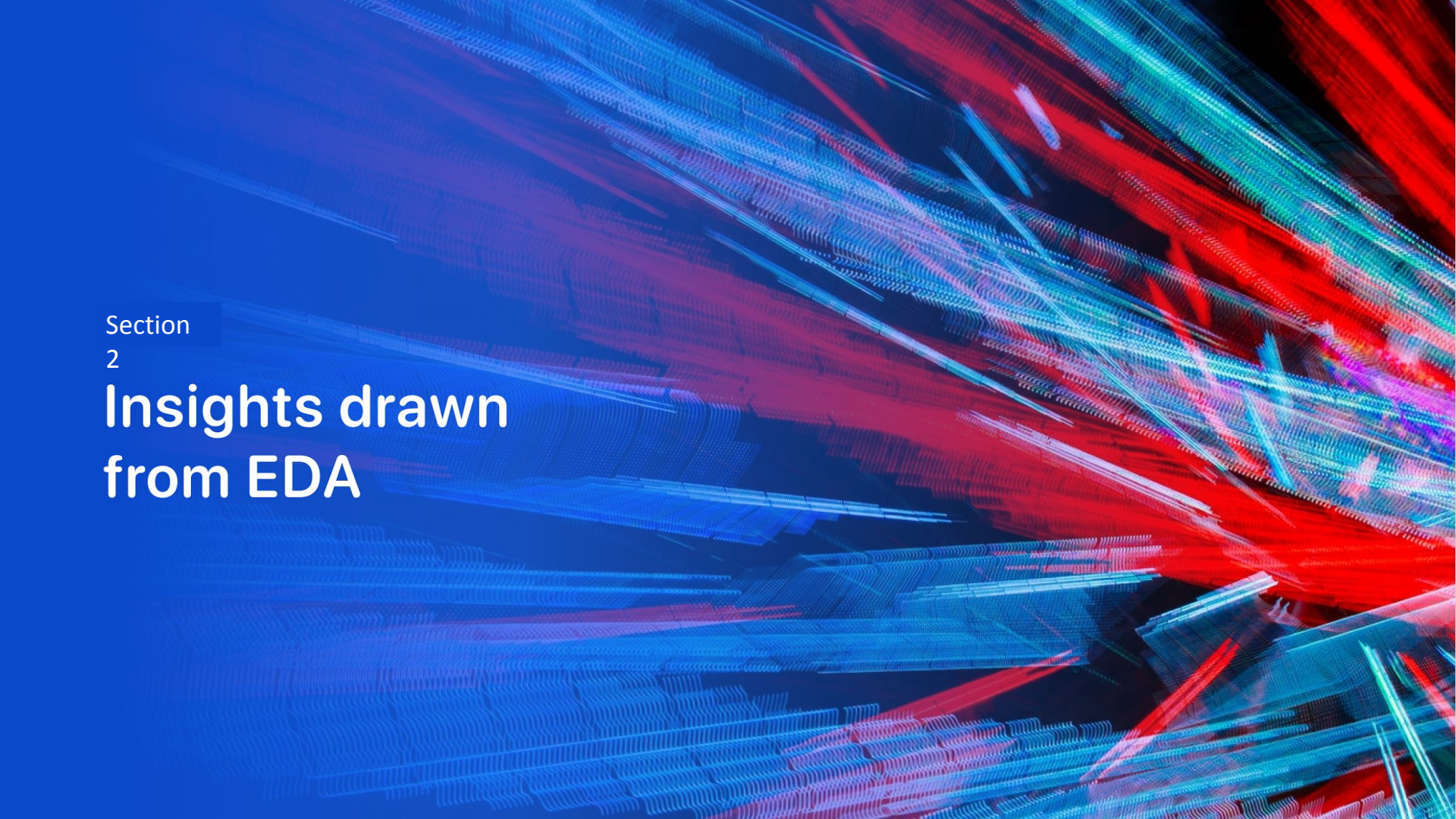
Confusion matrix

- **Model Development**
  - Imported libraries
    - Numpy
    - Pandas
    - Matplotlib
    - Seaborn
    - Sklearn
  - Created a NumPy array from the column Class in data
  - Standardized the data
  - Splited the data X and Y into training and test data.
  - Fit the training data to various model types
    - Logistic Regression
    - Support Vector Machine
    - Decision Tree Classifier
    - K Nearest Neighbors Classifier
  - Evaluated accuracy of each model using test data to select the best model

Link for the whole source code.

Results

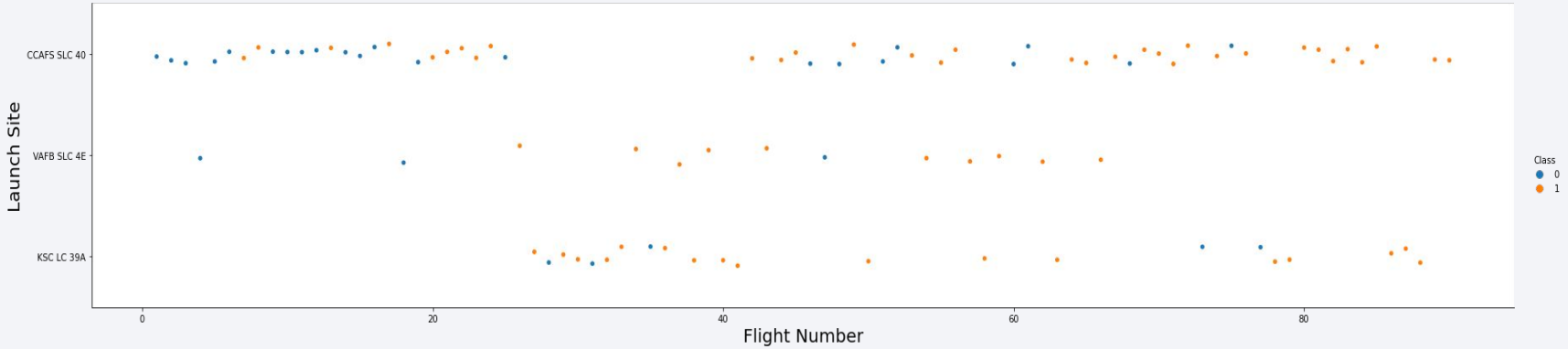
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results



Section 2

Insights drawn from EDA

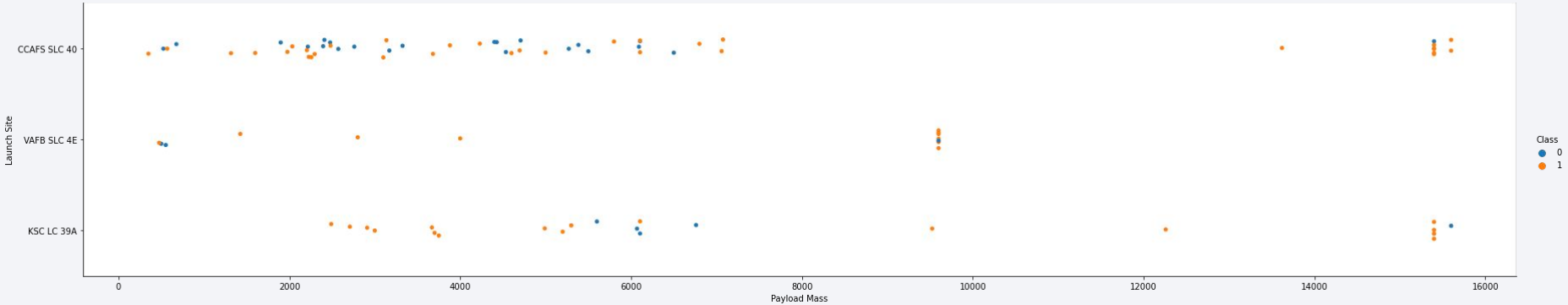
Flight Number vs. Launch Site



- Scatter plot of Flight Number vs. Launch Site

  CCAFS SLC 40 appears to have been where most of the early 1st stage landing failures took place

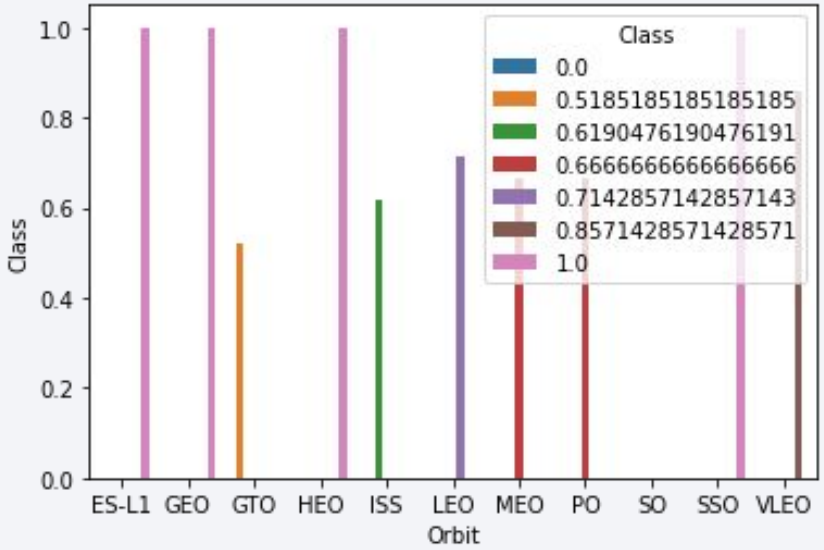
Payload vs. Launch Site



- Scatter plot of Payload vs. Launch Site

  We can see from the plot that for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

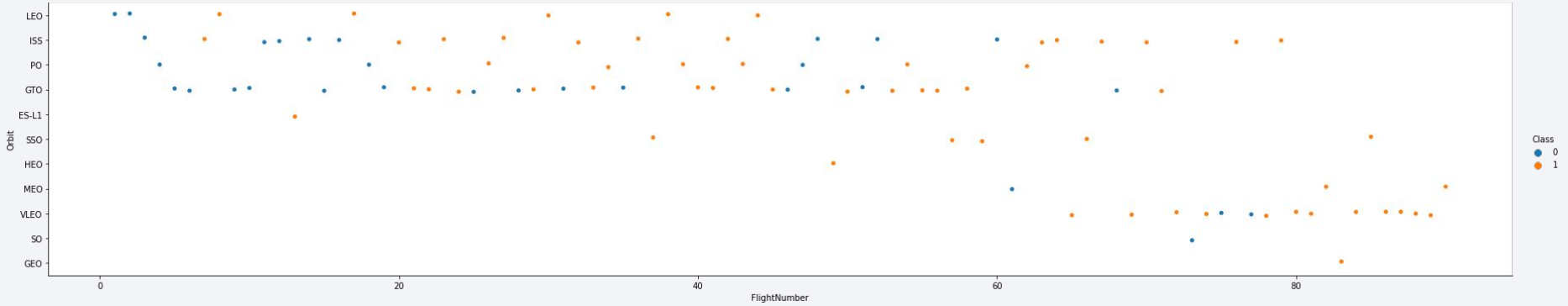
Success Rate vs. Orbit Type



- Bar chart for the success rate of each orbit type

  All orbit types except 'SO' have had successful 1st stage landings

Flight Number vs. Orbit Type

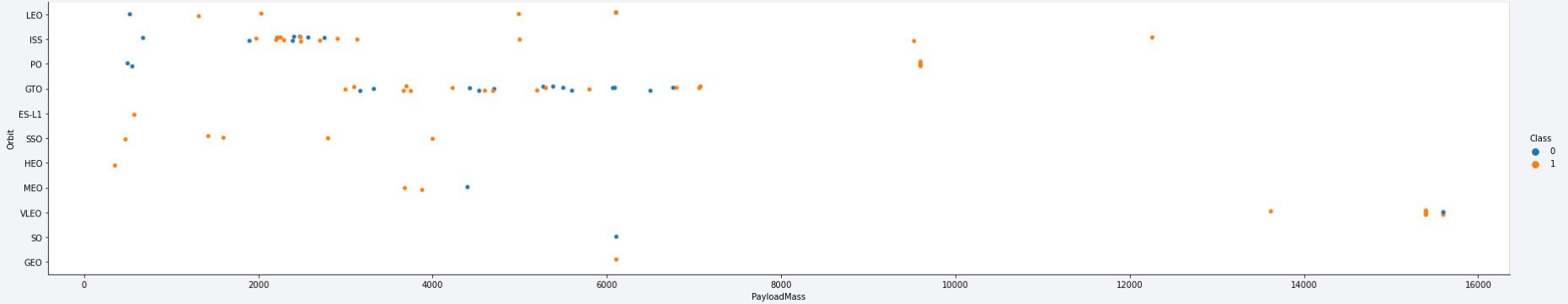


- Scatter point of Flight number vs. Orbit type

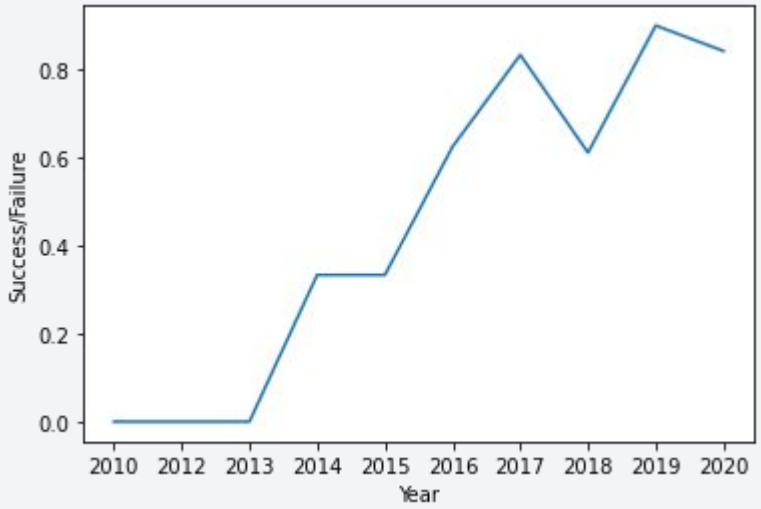
  Flight number positively correlated with 1st stage recovery for all orbit types

Payload vs. Orbit Type

- Scatter point of payload vs. orbit type



Launch Success Yearly Trend



- Show a line chart of yearly average success rate

  Success rate have experienced rapid increase since 2013, yet time by time decreasing.

All Launch Site Names

- Find the names of the unique launch sites
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
  - VAFB SLC-4E

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites 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

- Calculate the total payload carried by boosters from NASA

```
select sum(payload_mass__kg_)
from SPACEXTBL
where customer = 'NASA (CRS)'
```

45596

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1

2928

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad

2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

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

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes

mission_outcome	2
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here

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

2015 Launch Records

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

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

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

- 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

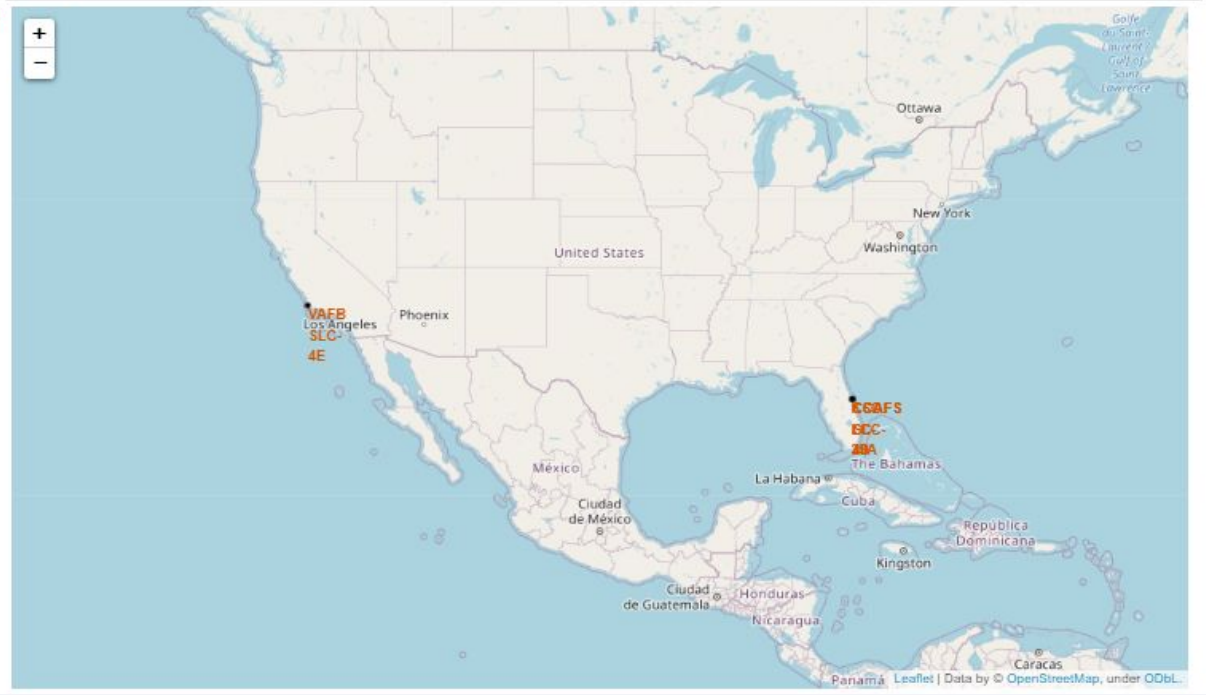
landing__outcome	2
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

Launch Site Location Analysis

Visualizing all launch sites location, overall there 4 unique cities that was used for launching

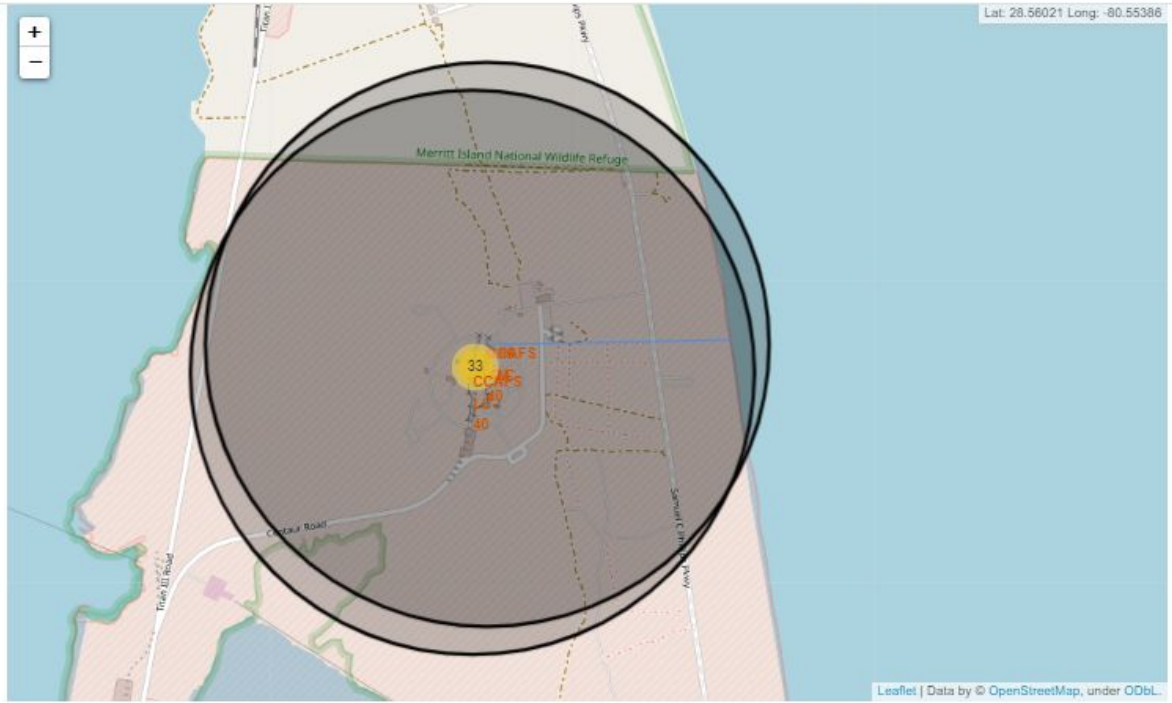


Launch Site Location Analysis

Visualizing the booster landing outcomes for each launch site highlights which launch sites have relatively high success rates

Launch Site Location Analysis

Below illustration shows distance between a launch site to the selected coastline point

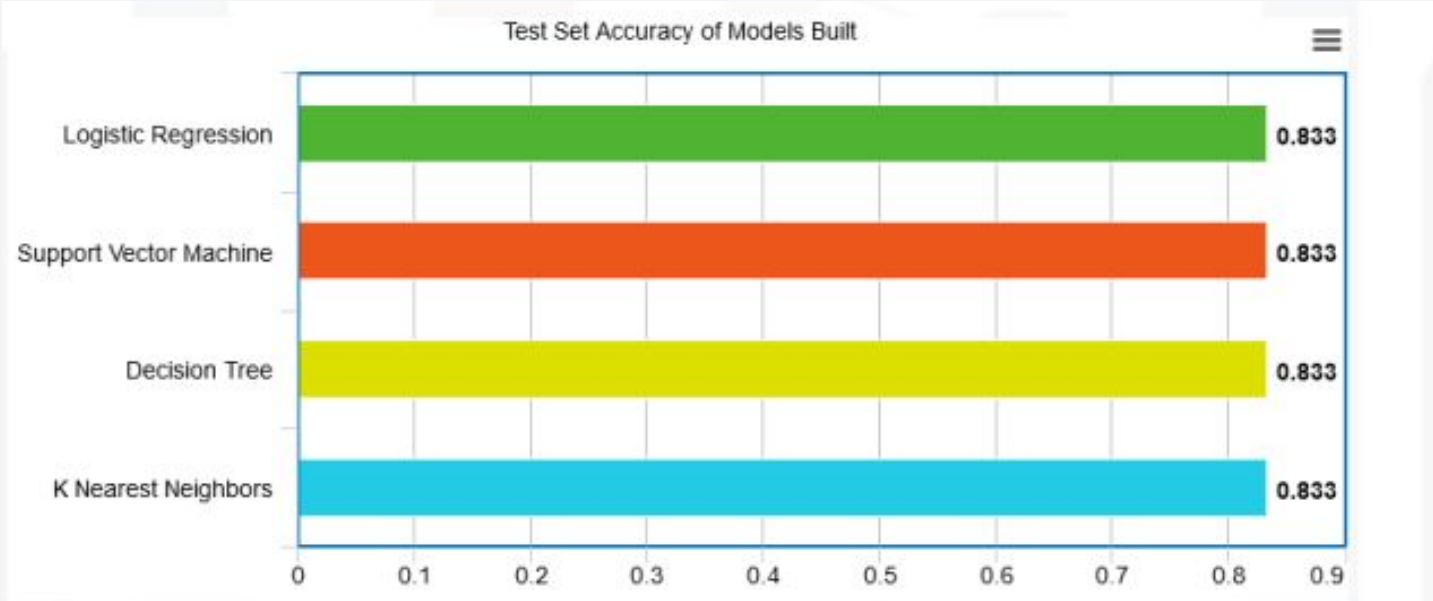


Section
5

Predictive Analysis (Classification)

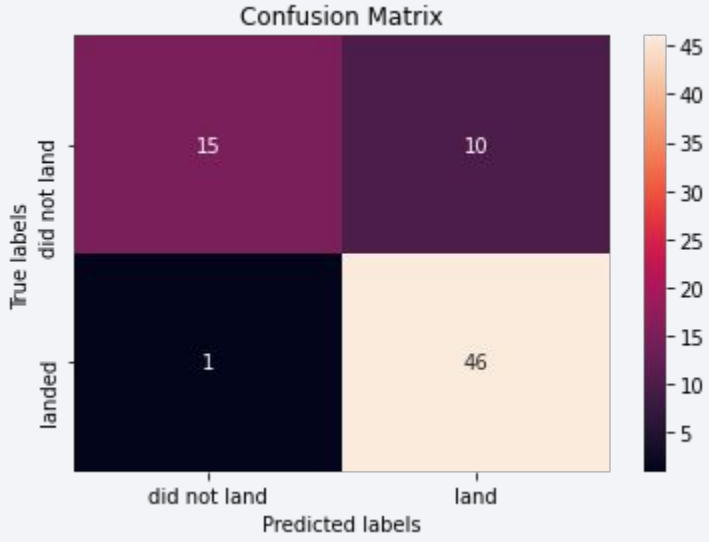
Classification Accuracy

From the below diagram we can conclude that all models have same accuracy rate



Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation



Conclusions

- Using the models from this report SpaceY can predict when SpaceX will successfully land the 1st stage booster with 83.3% accuracy
- With right strategic plans and action SpaceY can easily win race against SpaceX.

Appendix

- All source codes for our analysis are provided in their respective slide.
- Source of SpaceX's advertised $62 million launch price
  - https://www.spacex.com/vehicles/falcon-9/
- List of Falcon 9 and Falcon Heavy launches
  - https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- Useful Python resource
  - https://www.geeksforgeeks.org/python-programming-language/?ref=shm

Thank you!