Winning Space Race with Data Science

Elaheh Bagheri Zadeh
May 2023

Outline
Executive Summary
Methodology
Results
Conclusion
Appendix

Executive Summary

Summary of methodologies

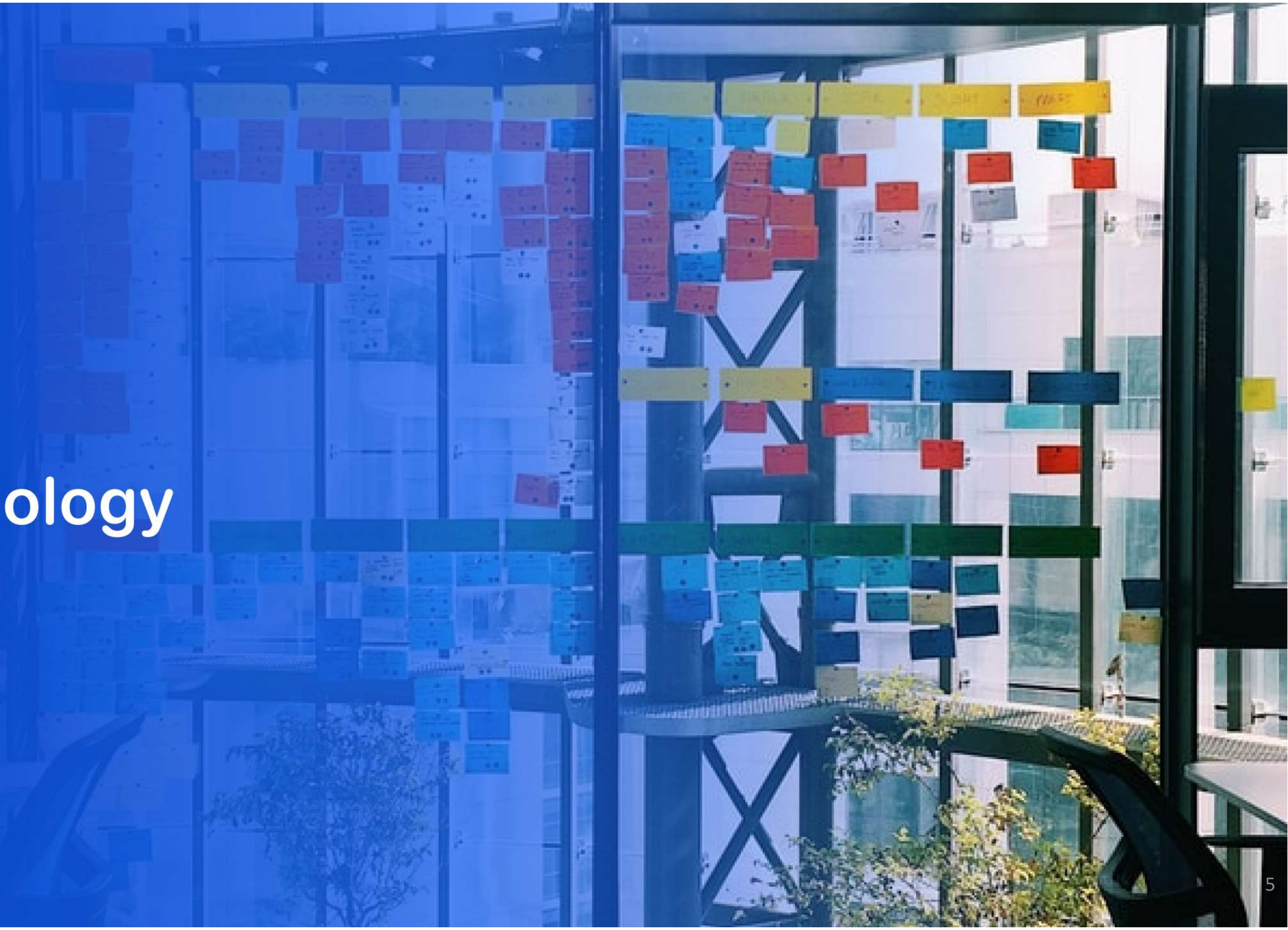
- Collecting Data (Request from SpaceX API) by Webscraping
- Cleaning the requested data
- Exploratory Data Analysis Using SQL and Visualization Packages
- Interactive Plotly Web App to visualize payload and success launch data at each Launch site
- Exploring Launch Sites Using Interactive Folium Maps
- Classification Predictive Analysis for Rocket landing Success

Summary of all results

- Exploratory Data Analysis Results
- Predictive Analysis Results

Introduction

- **Project background and context:**
- Space X claims that Falcon 9 rocket can launch with $62 M while other providers estimates are over $165 M. The key point of Space X is reusing the rocket if it landed successfully in the first stage. This claim will prove if we can predict the chance of success landing to estimate the cost of this project.
- **Problems you want to find answers:**
- We want to predict whether the falcon 9 first stage will land successfully or not. By estimation of Falcon 9 landing successfully, we can determine whether any company can bid against Space X or not.



Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Request to the SpaceX's API Website: https://api.spacexdata.com/v4
  - Webscrping data
- Perform data wrangling
  - Removing non related Falcon 9 from data (data cleaning)
  - Using One Hot Encoding to transform categorical variables to numerical ones.
  - Replacing missing numerical data with the mean value of parameter
  - Replacing categorical data with binary value (0,1)

Methodology

- Perform exploratory data analysis (EDA) using visualization and SQL
  - Visualizing results by scatter and bar plots to depict relationship between parameters
  - Using SQL queries to determine the data collecting process

- Perform interactive visual analytics using Folium and Plotly Dash
  - Interactive Plotly Web App to visualize payload and success launch data at each Launch Site
  - Exploring Launch Sites using interactive Folium Maps

- Perform predictive analysis using classification models
  - Deploying multiple classification (Logistic, SVM, Decision Tree, KNN) model to predict landing success
  - Defining top performance model

Data Collection

- Used SpaceX REST API to gather data on rocket launches:
  - https://api.spacexdata.com/v4
- API provides data on rockets used, launch dates, payload masses, launch success or failure, launch site name and location (latitude and longitude), booster version (note for this experiment we are only interested in Falcon V9 boosters), landing outcome, etc. (47 columns of data for each launch in total)
- Our Goal is to Predict the Landing outcome using the other variables
- Falcon 9 launch data was also collected via Webscraping Wikipedia using Beautiful Soup as mentioned next:

Data Collection – SpaceX API

1. Request Data From SpaceX API

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
```

```
response = requests.get(spacex_url)
```

2. Transform the Json File to a Pandas DataFrame

```
# Use json_normalize meethod to convert the json result into a dataframe
data = pd.json_normalize(response.json())
```

3. Extract required data and Make lists like as a sample

```
# Call getLaunchSite
getLaunchSite(data)
```

```
# Call getPayloadData
getPayloadData(data)
```

```
# Call getCoreData
getCoreData(data)
```

```
def getBoosterVersion(data):
    for x in data['rocket']:
        response = requests.get("https://api.spacexdata.com/v4/rockets/"+str(x)).json()
        BoosterVersion.append(response['name'])
```

Data Collection – SpaceX API

4. Make Dic. From lists & transform lists to Datafram

```
# Create a data from launch_dict
data = pd.DataFrame(launch_dict)
```

5. Extract related Falcon V9 boosters

```
# Hint data['BoosterVersion']!='Falcon 1'
data_falcon9 = data[data.BoosterVersion == 'Falcon 9']
```

6. Write Datafram into CSV

```
data_falcon9.to_csv('dataset_part_1.csv', index=False)
```

https://github.com/eliba12/IBM-Course/blob/main/Lab_1__EB_SpaceX_Data_Collection.ipynb

Data Wrangling

Check for Missing Values

```
data_falcon9.isnull().sum()
```

Calculate Sample Mean for Payload Mass

```
# Calculate the mean value of PayloadMass column
Mean_PayloadMass = data_falcon9.PayloadMass.mean()
```

Replace nas for Payload Mass with Sample Mean

```
# Replace the np.nan values with its mean value
data_falcon9['PayloadMass'] = data_falcon9['PayloadMass'].replace(np.nan, Mean_PayloadMass)
```


https://github.com/eliba12/IBM-Course/blob/main/Lab_2__EB___Spacex_Data___wrangling.ipynb

EDA with Data Visualization

- Visualize Orbits of Rockets and Count Launches in each orbit
  - Larger orbits require more fuel which may impact launch landing success
- Bar Graph to show landing success percentage forgiven Orbit Launches
  - By grouping our success based on orbit we can see if our intuition and combining this chart with our orbits visualization we can see the relationship between orbit radius and landing success
- Scatter Plots to show the relationships between the following Variables and visualize their correlation:
  - Flight Number vs. Payload Mass
  - Flight Number vs. Launch Site
  - Payload Mass vs Launch Site
  - Orbit vs Flight Number
  - Payload Mass vs. Orbit
- Line Graph to show landing success percentage with respect to the year of launch
  - We expect SpaceX Engineers to learn from their mistakes and improve over time but how fast are they improving

IBM-Course/Lab3_EDA_SQL.ipynb at main · eliba12/IBM-Course (github.com)

EDA with SQL

- Display the names of the unique launch sites in the space mission
- Display 5 records where launch sites begin with the string 'CCA'
- Display the total payload mass carried by boosters launched by NASA (CRS)
- Display average payload mass carried by booster version F9 v1.1
- List the date when the first successful landing outcome in ground pad was achieved
- List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000
- List the total number of successful and failure mission outcomes
- List the names of the booster versions which have carried the maximum payload mass
- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- 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

https://github.com/eliba12/IBM-Course/blob/main/Lab_4_EDA_with_Visualization.ipynb

Visualized Launch Data in an interactive Map

- Used Latitude and Longitude Coordinates of Launch Sites to Add Circle Markers with the site names labeled
- Assigned Launch Outcome (Success/Failure)from the data frame to Classes 1and 0respectively and assigned the classes Green and Red markers on the map to Marker Clusters grouped by Launch Site
- Used lines and points to measure (via Haversine's Distance Formula) and label the minimum distances of the launch sites to:
  - Cities
  - Highways
  - Coastlines
  - Railways
- Answered the following Questions:
  - Are launch sites in close proximity to railways? Yes
  - Are launch sites in close proximity to highways? Yes
  - Are launch sites in close proximity to coastline? Yes
  - Do launch sites keep certain distance away from cities? About 50 km

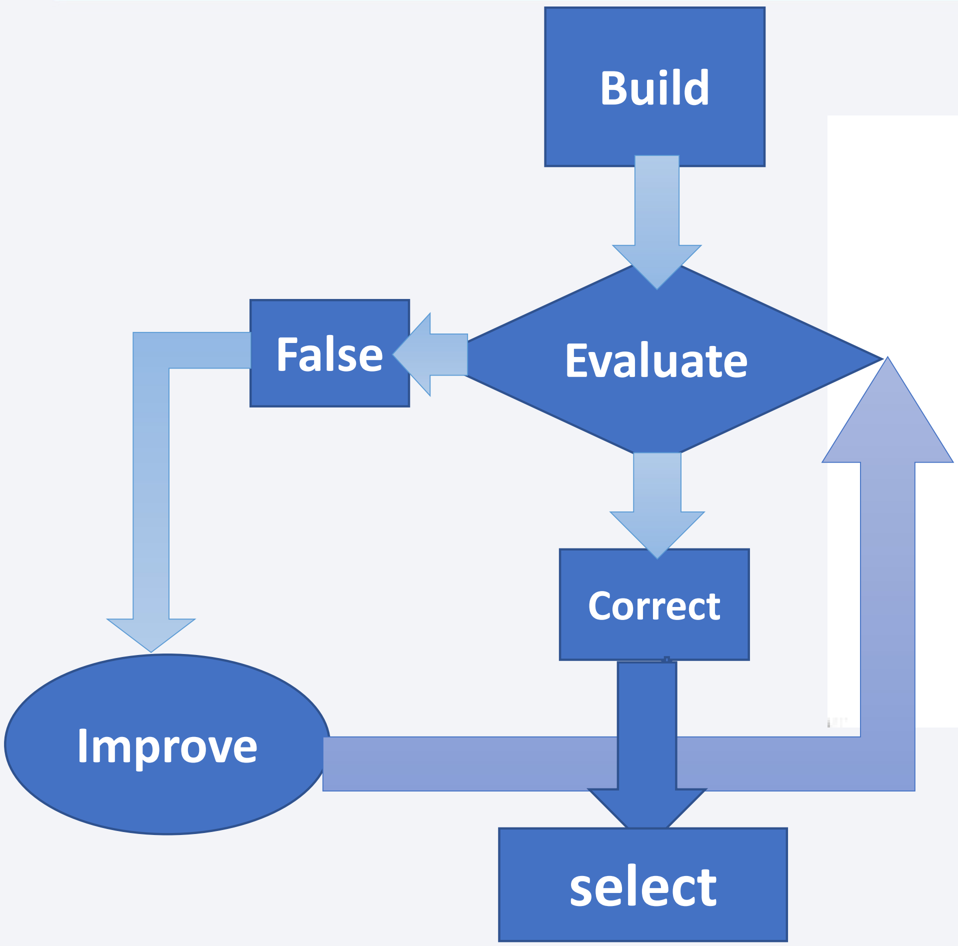
https://github.com/eliba12/IBM-Course/blob/main/Lab_5__EB_Launch_site_location.ipynb

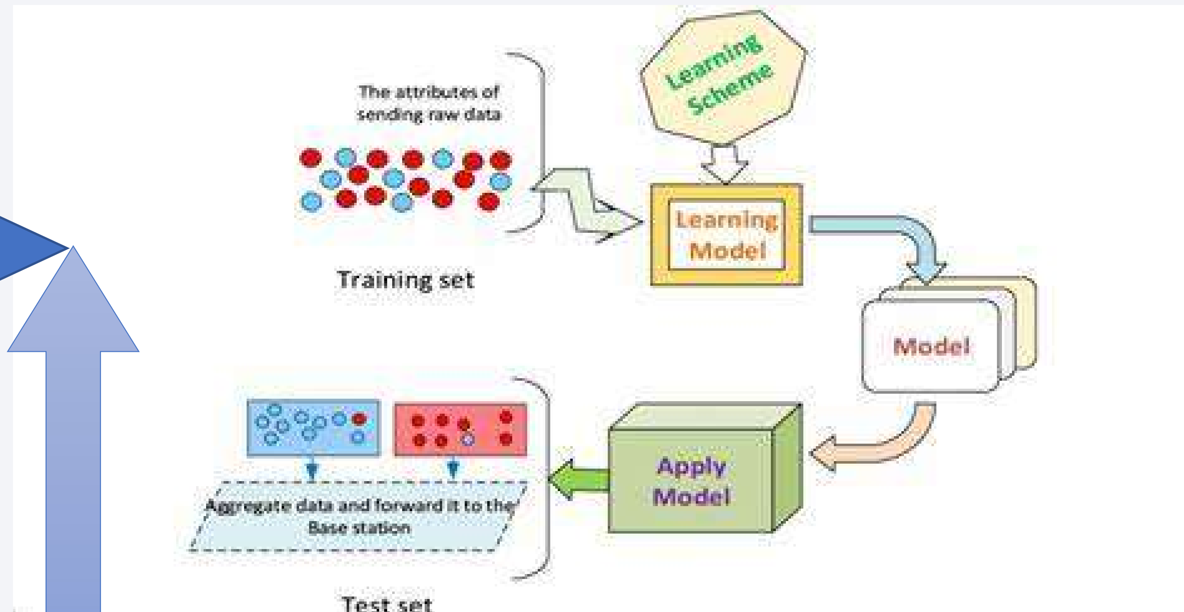
Build a Dashboard with Plotly Dash

- Built Plotly Dashboard to make an interactive web app to visualize launch data
- Includes Pie Charts to visualize launch landing success broken down by Launch Site
- If all sites are selected, we get the proportion successful launch landings each site accounts for
- If we select an individual site, we see the proportion of all launches at that site which landed successfully
- Includes Scatter Plot of Payload Mass (kg) vs Landing Success Rating (0 for Failure, 1 for Success) color coded by booster version
- Selecting a single site removes points from other sites, Selecting All includes all data
- Plot Payload Mass Range can be selected by a slider for Min and Max

https://github.com/eliba12/IBM-Course/blob/main/Ploty.ipynb

Predictive Analysis (Classification)





https://github.com/eliba12/IBM-Course/blob/main/Lab_6__EB_Machine_Learning_Prediction.ipynb

Results

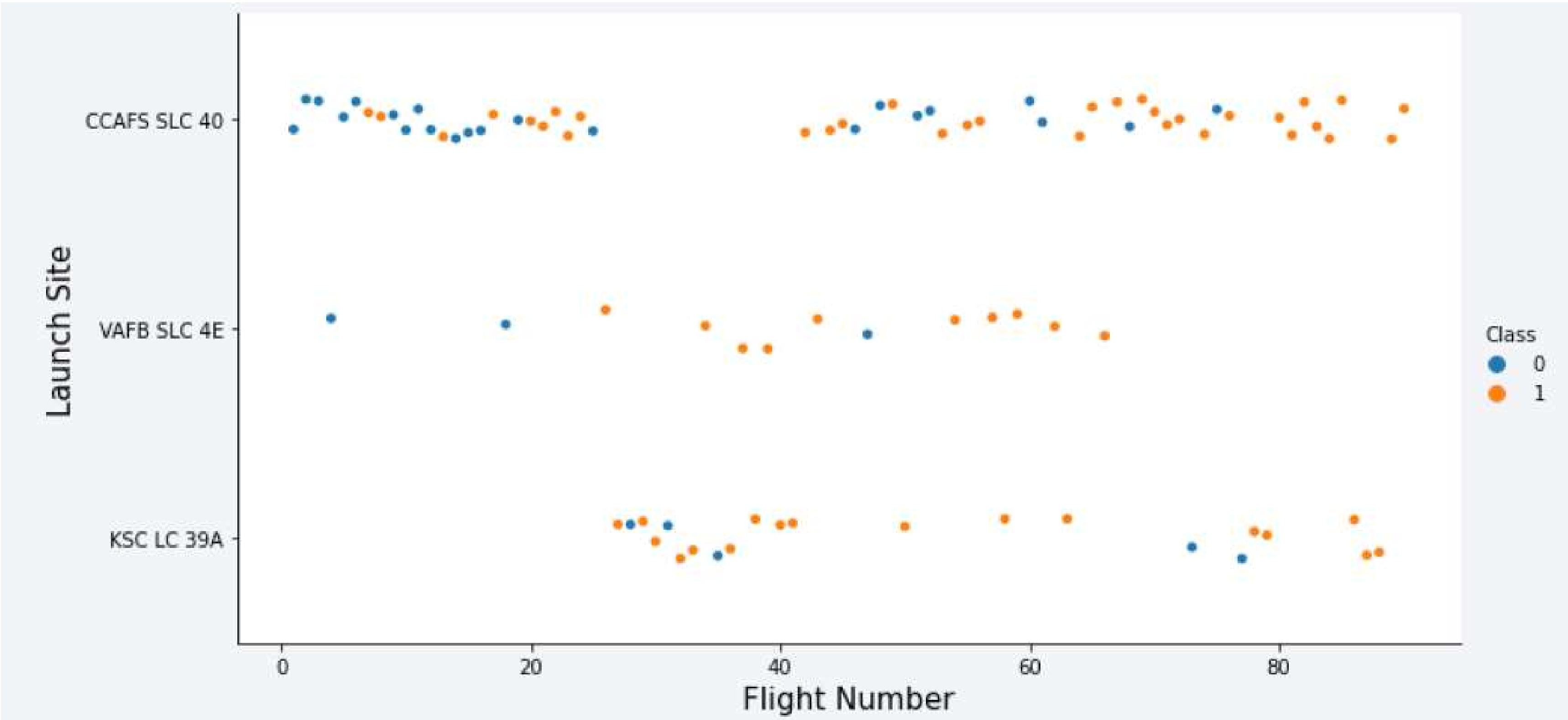
Exploratory data analysis results

Interactive analytics demo in screenshots

Predictive analysis results

Section 2

Insights drawn from EDA

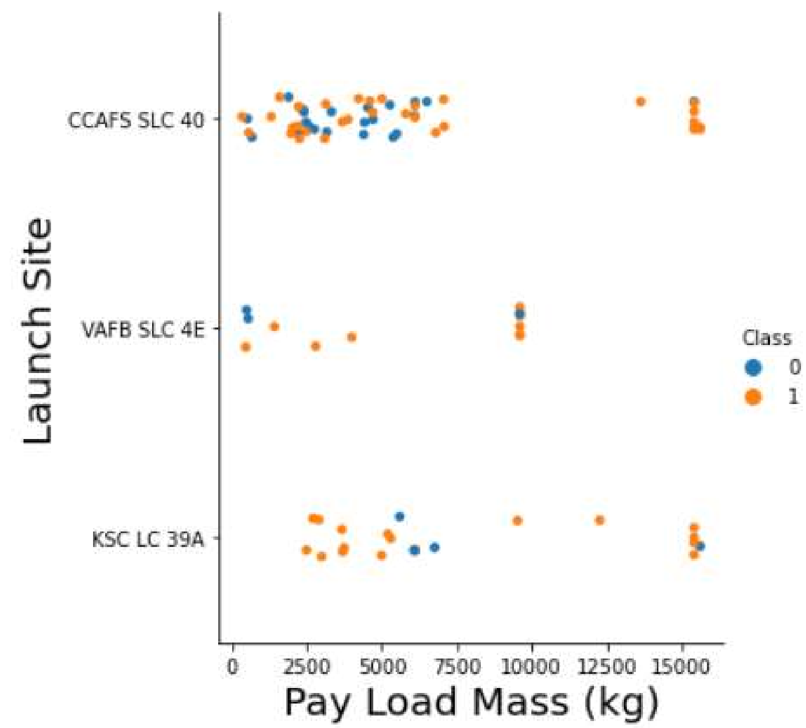


Flight Number vs. Launch Site

- At CCAFS SLC 40, flight numbers less than 20 flight numbers were more successful and flight number between 40 and 80 were more unsuccessful.
- At KSL LC 39A launch site, flight numbers between 25 and 40 placed and mostly landed successfully.
- At VAFB SLC 4E launch site, higher flight numbers landed more successfully.

Payload vs. Launch Site

- We can see that as payload mass increases for Site CCAFS SLC 40, the probability of a successful landing increases
- There is not a clear correlation between payload mass and launch success for the other two sites
- Lastly we find that middle mass launches (~9,000kgs) are usually performed at VAFB SLC 4E but the other two sites launch heavy masses (>10,000kgs) while VAFB SLC 4E does not



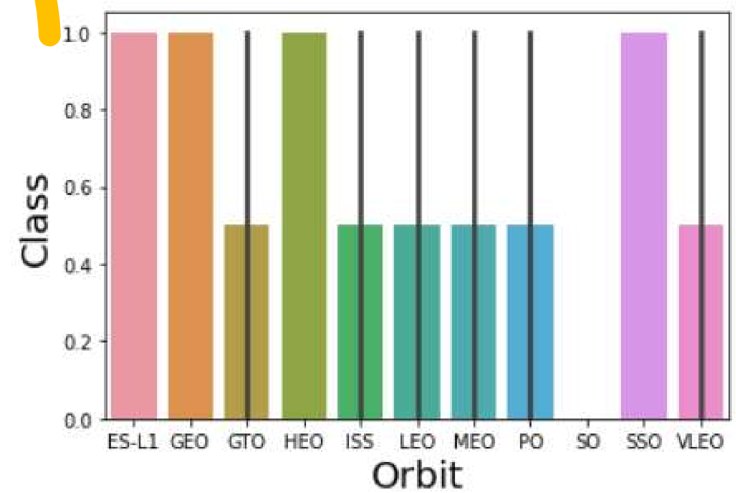
SUCCESS RATE VS. ORBIT TYPE

- ES-L1, SSO, HEO, and GEO orbits have perfect landing scores

- From the previous slide we find that orbits within 450-10000 kilometers have the highest rate of success

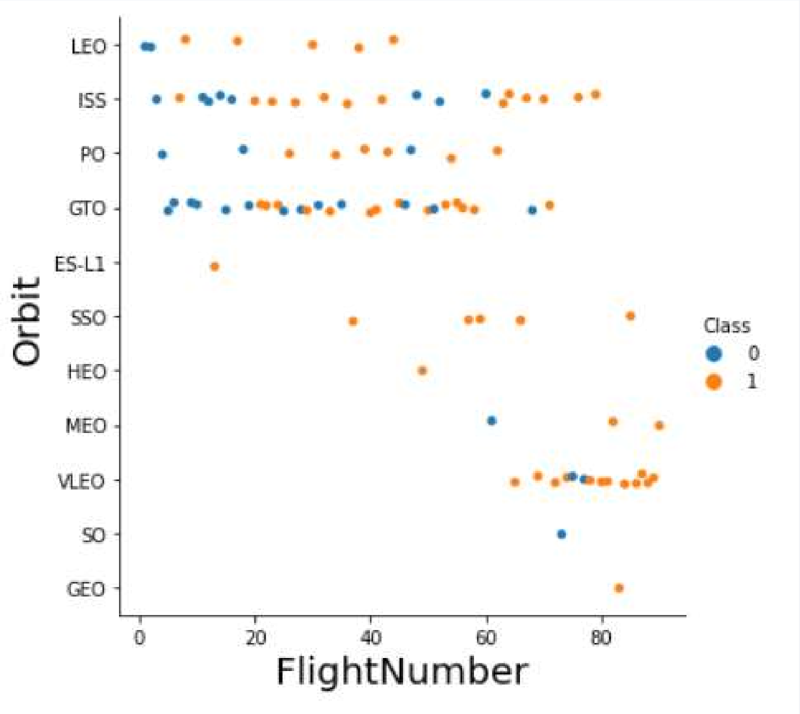
- Launches beyond 10000 km have lower success ratings

- Launches between 1000-2000 km have less success than launches between 1000-450 km (lower altitude) and launches between 2000-10000km (higher altitude)



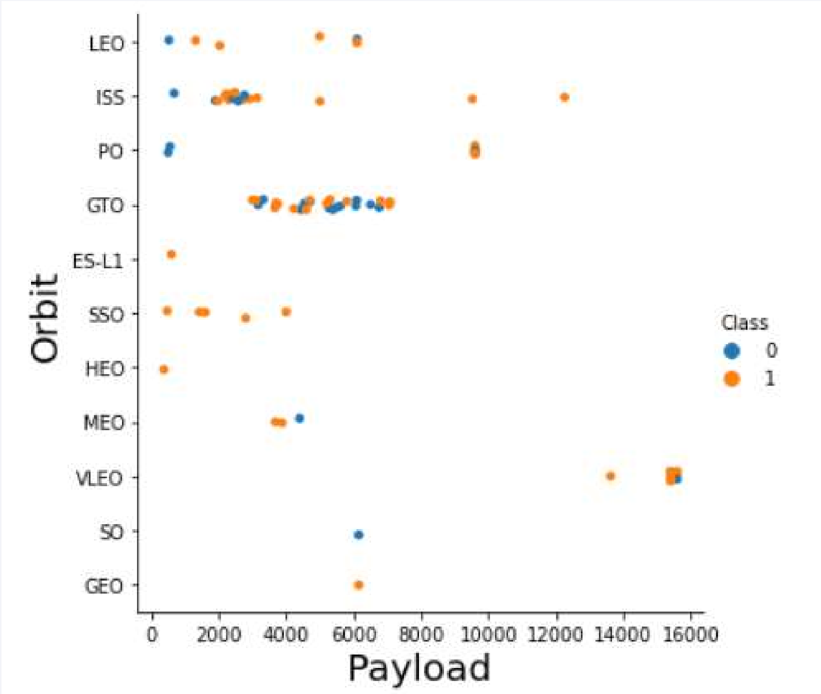
Flight Number vs. Orbit Type

- SSO, ES-L1, GEO & HEO orbits have 100% success for all flight numbers.
- ISS, GTO, LEO, MEO, & VLEO orbits show that the higher flight number has higher chance of success.



Payload vs. Orbit Type

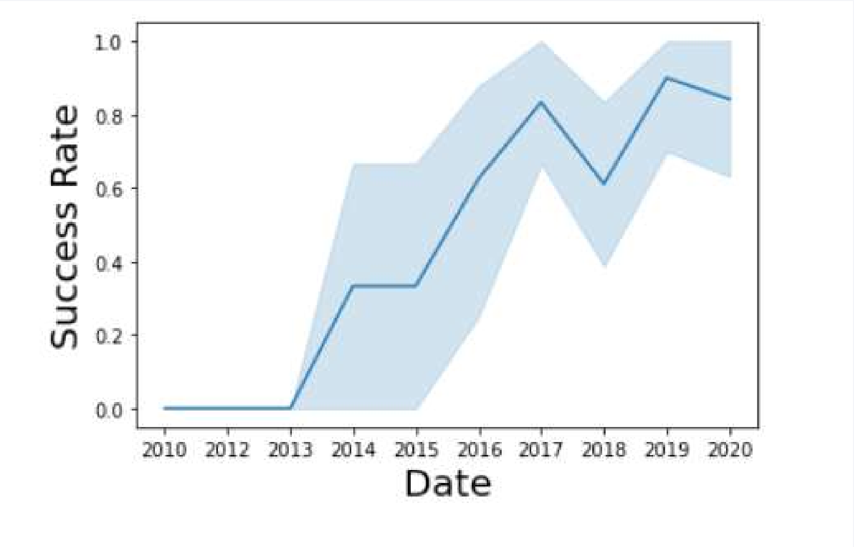
- GEO, HEO, SSO, & ES-L1 orbits present 100% success rate independent of payload
- PO, LEO and ISS Orbits have higher probability for successful landings as payload mass increases
- GTO orbits have no correlation between payload mass and landing success



Launch Success Yearly Trend

- Success rate was zero up to 2013 and it was increased every year however the figure shows that success rate has declined between 2017-2018 & 2019-2020

- Maximum success rate was 0.9 and it happened on 2019



All Launch Site Names

Using SQL magic :

```
%sql select distinct(LAUNCH_SITE) from SPACEXTBL
```

```
 * ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:50000/BLUDB
Done.
```

launch_site
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`
- By selecting * we request all columns of the data frame
- Five different launch sites begin with "CCAFS LC-40" which all have success mission outcome.

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

- **Using multi-line sql magic we can query the database as follows:**
- **The sum call calculates the total**
- **The as clause renames the returned column to Total Payload mass**
- **We select from a subquery that queries the list of customers and their payload masses but limits the search to rows where the customer is Nasa (CRS)**
- **The group by clause makes sure our sum combines the payloads of all NASA (CRS) rows**

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)'
```

```
 * ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:50000/BLUDB
Done.
```

1
45596

Average Payload Mass by F9 v1.1

- **Using multi-line sql magic we can query the database as follows:**
- **The AVG call calculates the average payload mass**
- **The as clause renames the returned column to AVERAGE Payload mass**
- **We select from a subquery that queries the list of boster_versionsand their payload masses but limits the search to rows where the booster_versionis F9 v1.1**
- **The group by clause makes sure our average is calculated over all entries with the Booster_version**

```
%sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1'

 * ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:50000/BLUDB
Done.
     1
2928.400000
```

First Successful Ground Landing Date

- The “MIN” call selects the earliest date meeting the criteria
- The as clause renames the returned column to FIRST SUCCESS
- Selecting from a subquery that queries the list of “DATES” from the SpaceX table where the “LANDING_OUTCOME” is Success (Ground Pad)

```
PACEXTBL where Landing__Outcome = 'Success (ground pac
```

```
dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net
```

Successful Drone Ship Landing with Payload between 4000 and 6000 kg

- **The where clause restricts results to those with a payload mass between 4000 and 6000 kgs and a Landing Outcome of Success (drone ship)**
- **The database returns the Booster Version, Payload mass, and Landing outcome for these data**

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

booster_version	payload_mass_kg	landing_outcome
F9 FT B1022	4696	Success (drone ship)
F9 FT B1026	4600	Success (drone ship)
F9 FT B1021.2	5300	Success (drone ship)
F9 FT B1031.2	5200	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

- **Select “Mission Outcomes” and count of each type of “mission_outcome” as the Total**
- **Group by clause ensures we count the <u>success</u> and <u>failures</u> as separate groups**

```
[ ]  %sql select count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure (in flight)'

 * ibm_db_sa://sdk38546:***@dashdb-txn-sbox-yp-lon02-07.services.eu-gb.bluemix.net:50000/BLUDB
Done.
```

1
100

Boosters Carried Maximum Payload

- **Select Unique Booster Versions to ensure we get distinct results**
- **Returned Payload Mass to see what the Max Payload is**
- **Where clause checks that the payload mass equals the max payload which is found via a sub query**

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

2015 Launch Records

- **Select Landing_Outcome, Booster Versions, Launch_Site, and Date as requested**

- **Where clause restricts results to have a landing outcome of Failure (Drone Ship) and a launch date in the year 2015**

landing_outcome	booster_version	launch_site	DATE
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	2015-01-10
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	2015-04-14

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
- df["landing_class"].mean( ) = 0.67

```
True  ASDS      41
None  None      19
True  RTLS      14
False ASDS       6
True  Ocean      5
False Ocean      2
None  ASDS       2
False RTLS       1
Name: Outcome, dtype: int64
```

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-01-14	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
2016-08-14	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-07-18	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
2016-05-06	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
2015-12-22	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

Section 3

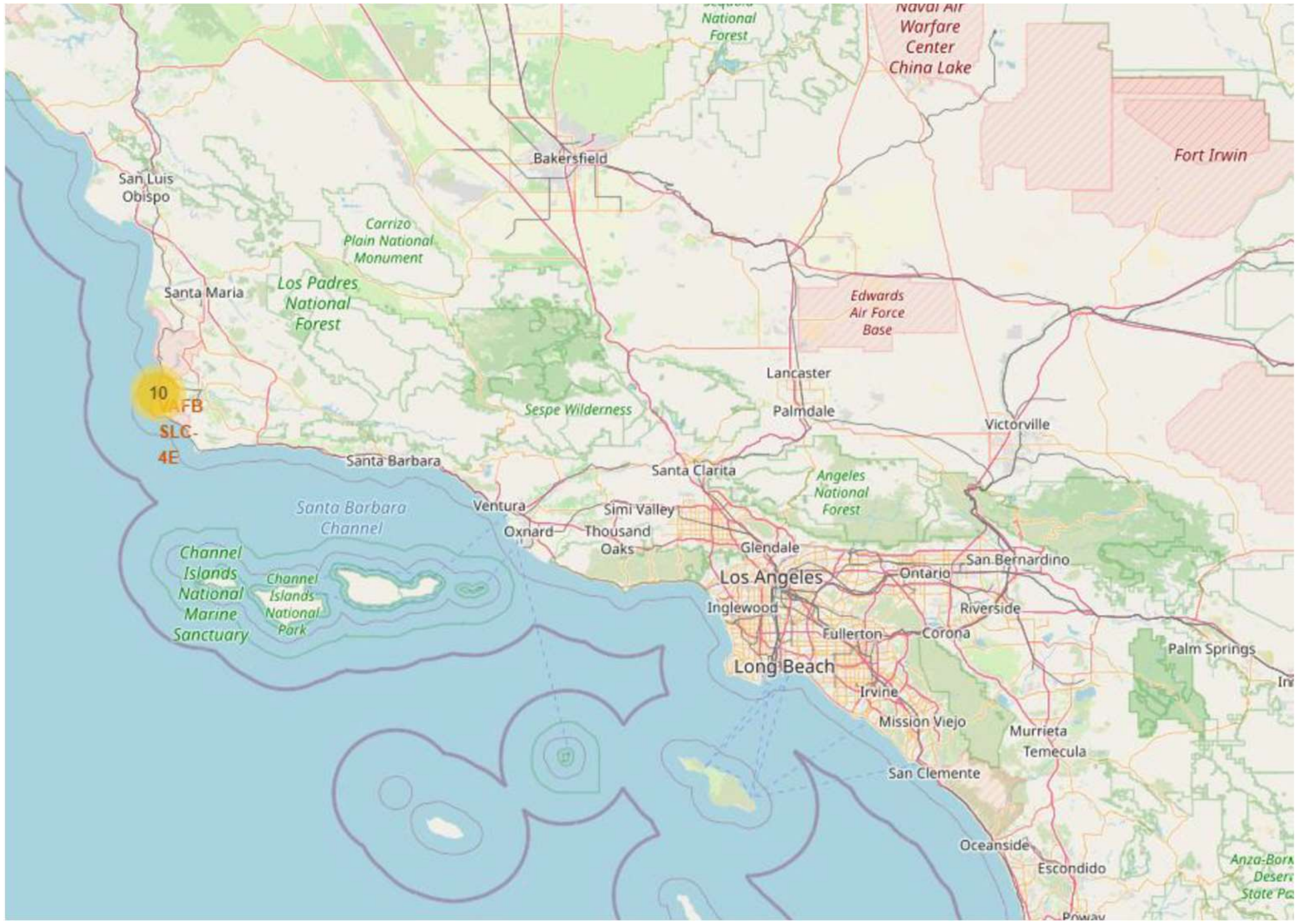
Launch Sites Proximities Analysis

Global Map of SpaceX Launch Sites

- Space X has Launch Sites Exclusively in the United States
- There are 4 total Launch Sites 1 is Located on the West Coast in California
- 3 are located on the East Coast in Florida

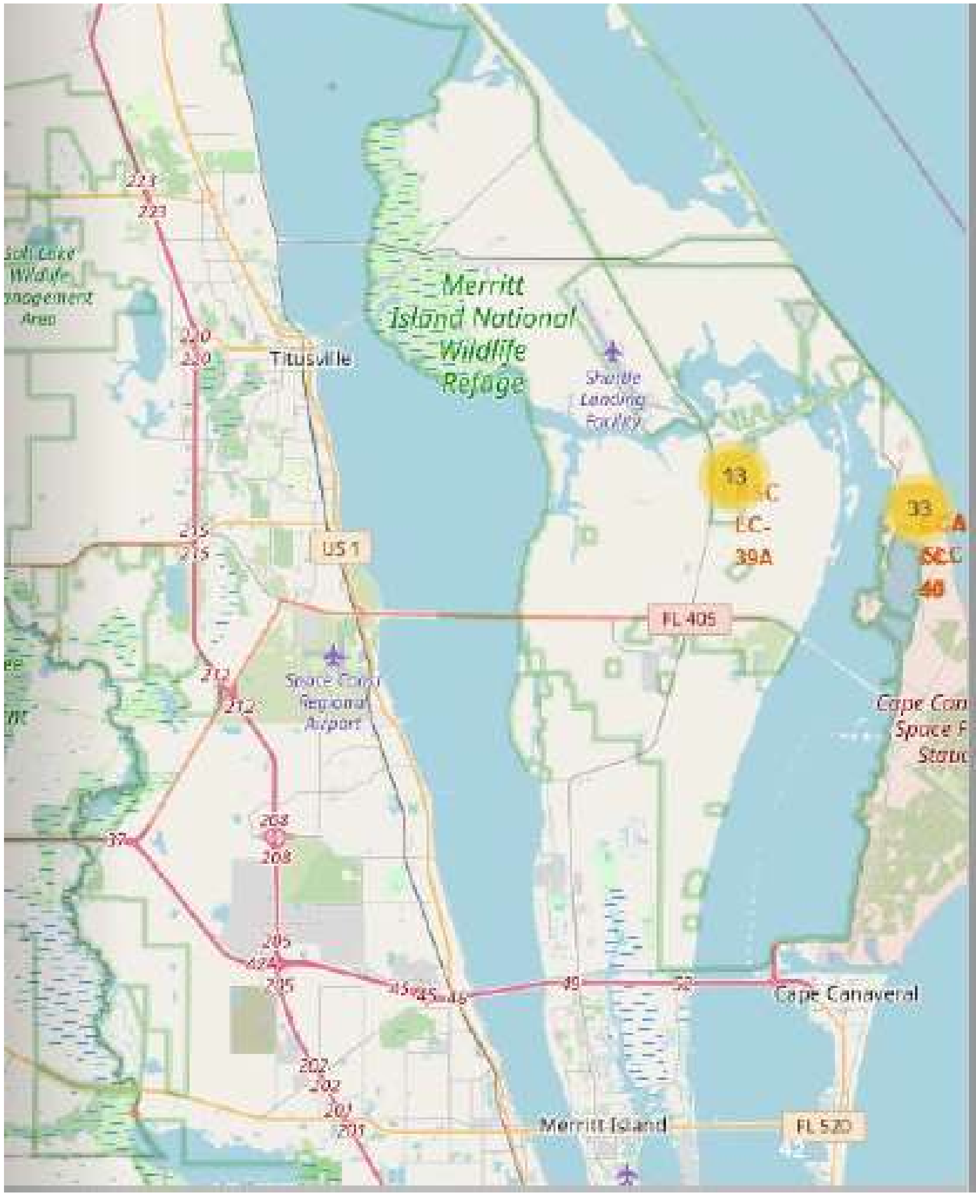
SpaceX VAFB SLC-4E Launch Site

- Space X has a single Launch Site Stationed off the Coast of Santa Mariain California
- Our Cluster Marker indicates 10 Falcon 9 launches have taken place at this site

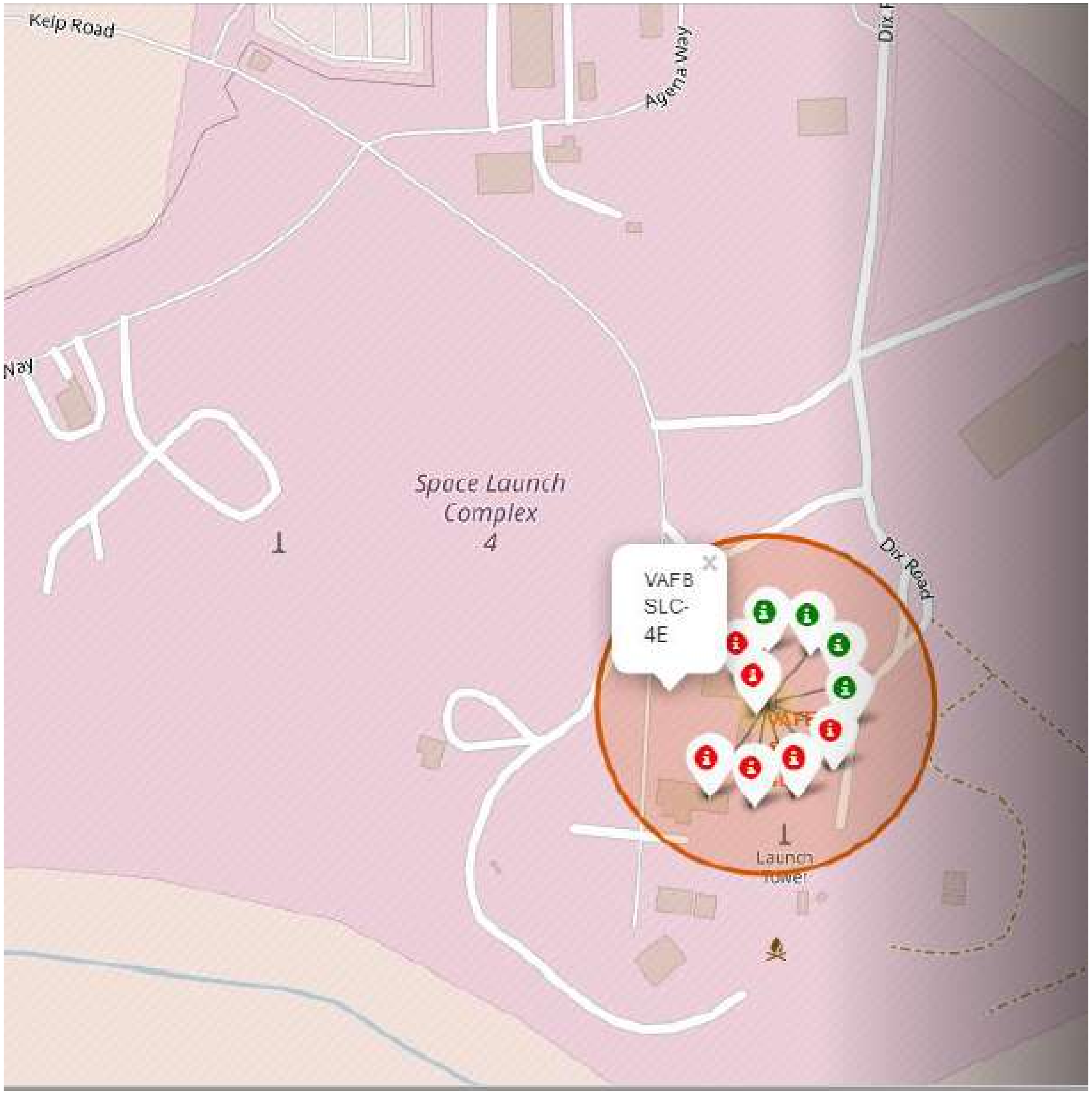


SpaceX Florida Launch Sites:

Merritt Island	***Cape Canaveral***
KSC LC-39A	**CCAFS LC40** **CCAFS SLC-40**
13 Launches	**33 Launches**



SpaceX VAFB SLC-4E Launch Site



- **Success Landing: 4 Rockets**
- **Fail Landings: 6 Rockets**
- **Total Landings: 10 @ VAFB SLC-4E**

SpaceX CCAFS LC-40 Launch Site Cape Canaveral

Success Landing: 7 Rockets

Fail Landing: 19 Rockets

Total Launches : 26 @ CCAFS LC-40

SpaceX CCAFS SLC-40 Launch Site Cape Canaveral

Success Landing: 3 Rockets

Fail landing: 4 Rockets

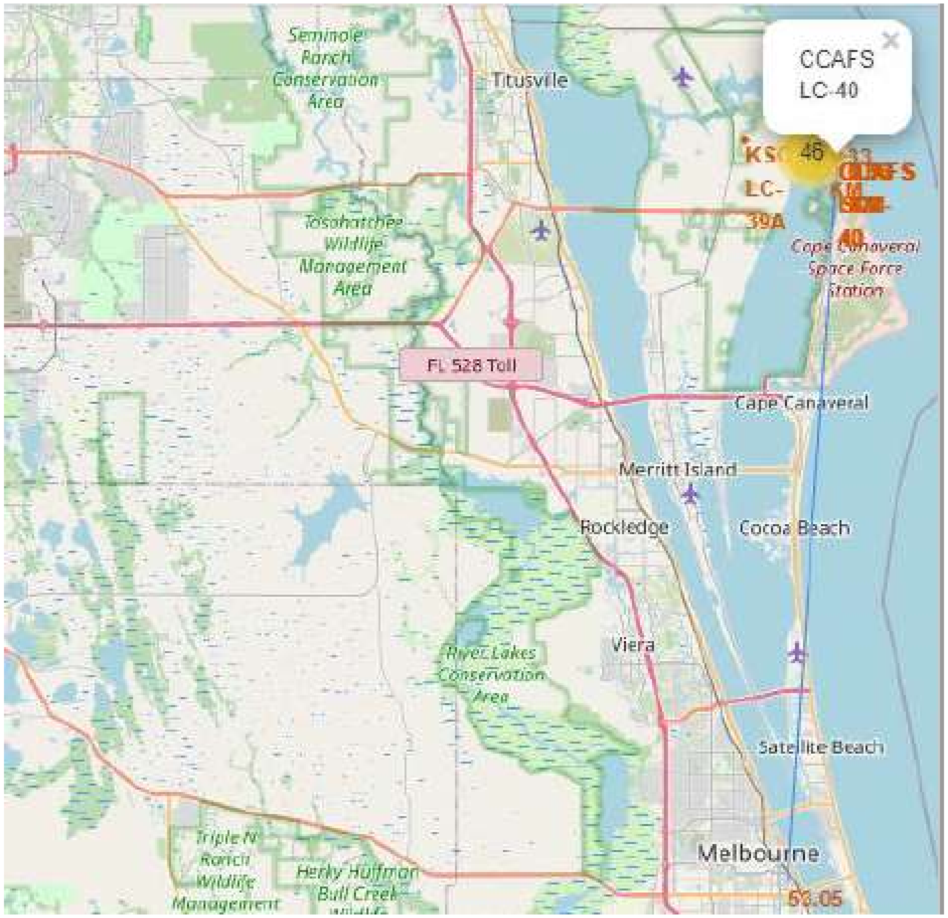
Total Launches: 10 @CCAFS SLC-40

Infrastructure near Cape Canaveral Sites

- We can see that the nearest coastline to the Cape Canaveral sites is within 1 kilometer and exactly measured to 0.93km for the LC-40 site
- We can see that the nearest highway is within 0.7 km and is exactly 0.66km from the LC-40 site
- The nearest railway is within 1.5 km to the base and exactly 1.33km from the LC-40 site
- From this we recognize that infrastructure is near launch sites for easy access to manufactured parts

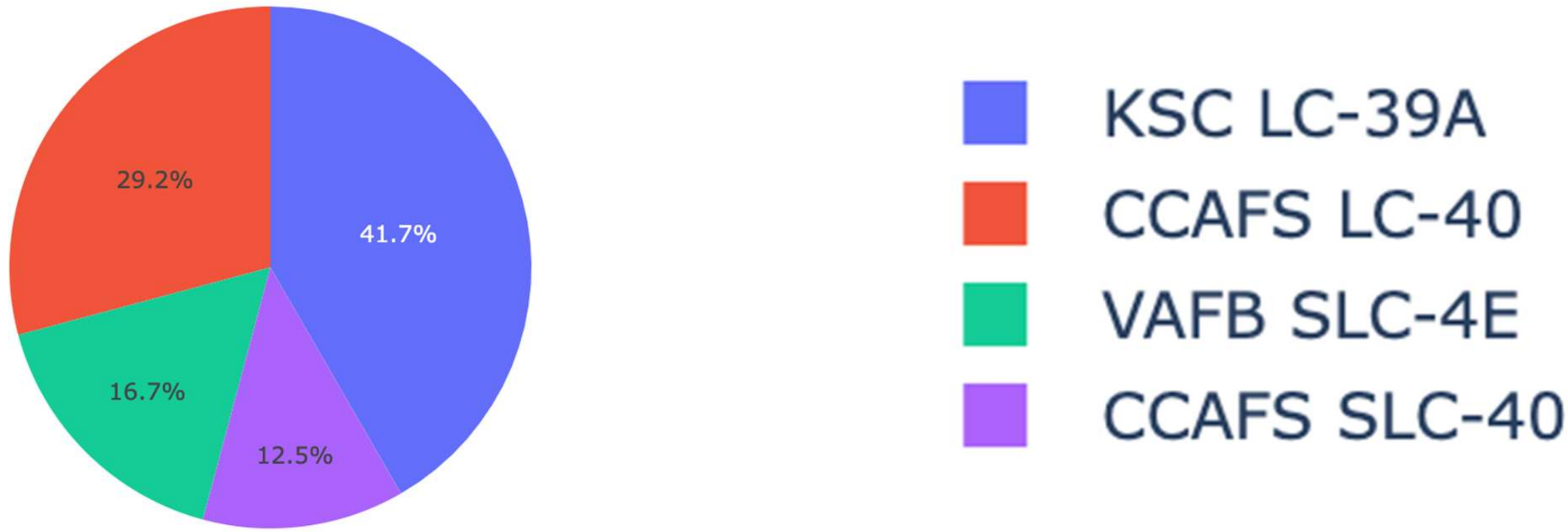
City distance to Cape Canaveral Sites

- The nearest city to the launch sites in Florida is Melbourne located 53.05Km from the Cape Canaveral LC-40 Site
- Proximity to the city is not a priority for SpaceX as we previously saw infrastructure is built near the launch sites meaning travel time for parts and employees is minimal



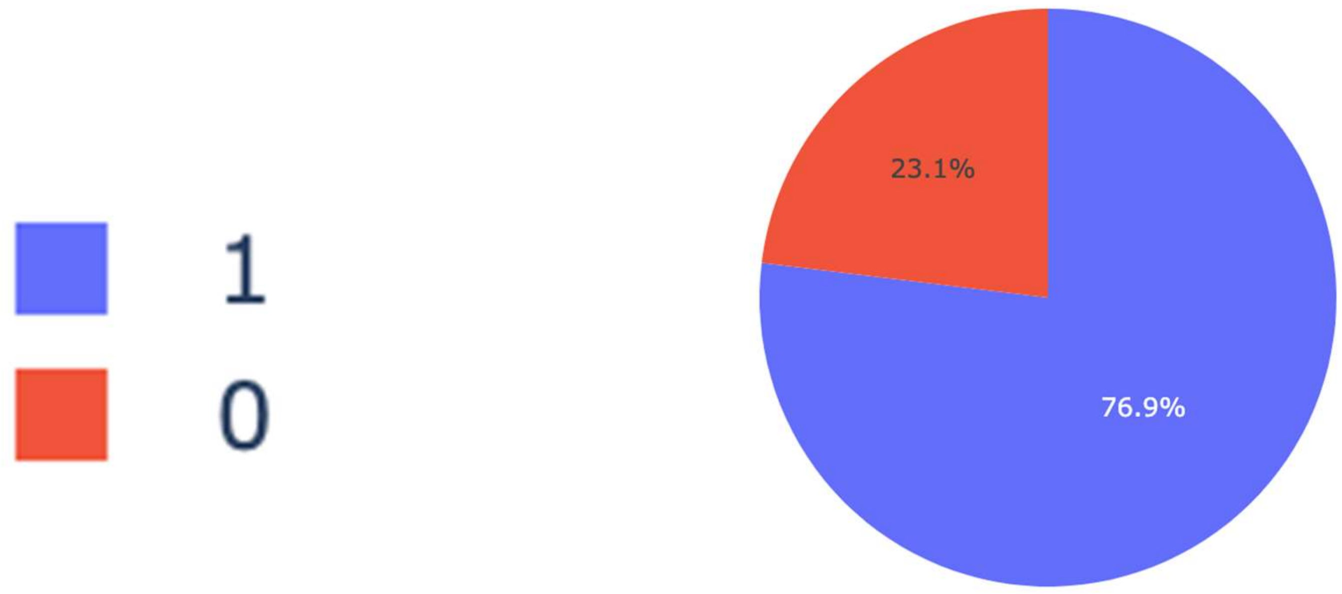
Section 4

Build a Dashboard with Plotly Dash



SpaceX Launch Dashboard
Success Count for all sites

- **KSC LC-39A Launch site accounts for the largest percentage of the total number of successful landings at 41.7%**



Launch Site with the Highest Probability of Success

Total success launches for site KSC LC-39A

- The KSC LC-39A Launch Site also has the highest probability of success per launch
- 76.9% of all launches at the KSC LC-39A Site Land Successfully
- 23.1%of all launches at the KSC LC-39A Site Fail to Land

Payload Weight vs Success Labeled by Booster Version Heavy Vs Light Weight

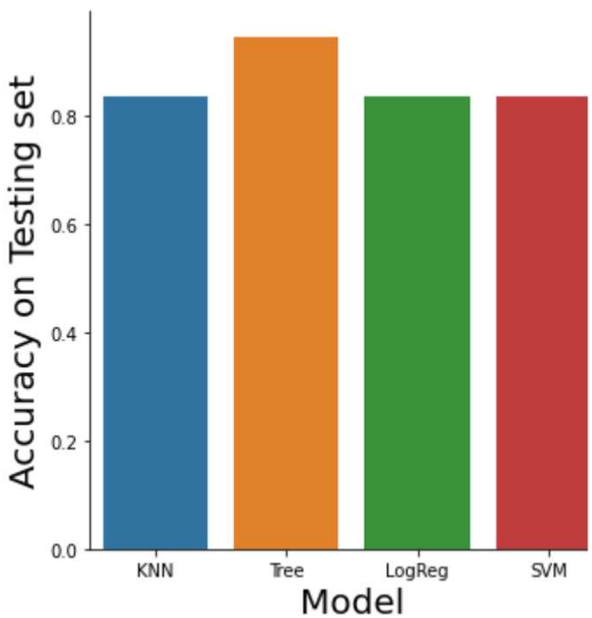
Payloads under 5000kg	Payloads over 5000kg
18 successful landings 20 failures landings	3 successful landings 8 failures landings
47.36% Success rate	27.27% Success rate

Section 5

Predictive Analysis (Classification)

Classification Accuracy

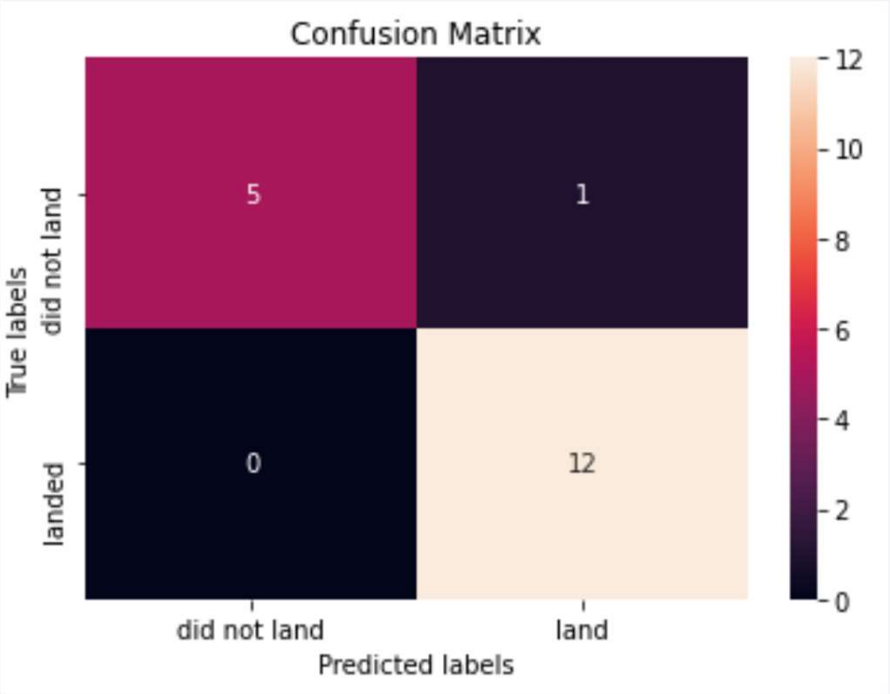
- Tree model (94%) has the highest accuracy



KNN	SVM	Log Regression	Tree
0.83	0.83	0.83	**0.94**

Confusion Matrix for Decision Tree Model

- We see that our model is able to correctly predict that 5 of the 6 testing points that fail to land
- Our model can correctly predict that all 12 rockets that land will land
- Sensitivity: 5/6=83.33%
- Specificity = 12/12 = 100%
- Accuracy: (5+12)/(5+1+12+0)=%94.44



Conclusions

- Decision Tree model shows the best accuracy %94.44 for predicting Falcon 9 Rocket landing.
- KSC LC-39A Launch Site has the highest probability of success rate.
- Payload plays a pivotal role in rocket landing successfully. Rockets with higher weight have lower chance to land successfully.
- ES-L1, SSO, HEO and GEO orbits have the highest rate of landing successfully.
- Success landing rate has improved generally during years 2013 to 2020.


Thank you!