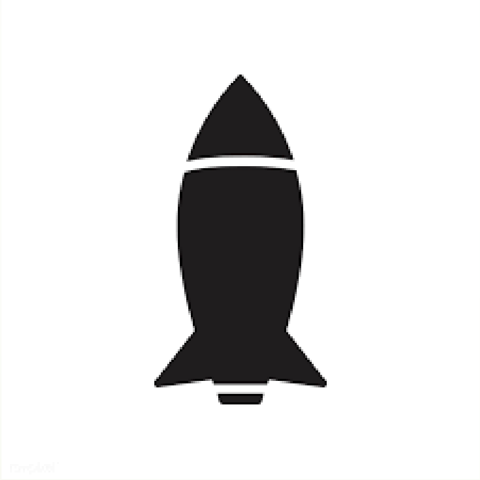
SPACE X
CAPSTONE PROJECT



EXECUTIVE SUMMARY

- Methodologies
  - Data collection and wrangling
  - EDA with Visualization
  - EDA with SQL
  - Building Dashboards and maps
- Results
  - Exploratory data analysis results
  - Screenshots and predictive analysis results

Introduction

- Background

  - SpaceX

- Problems

  - Correlations between each rocket variables and successful landing rate
  - Conditions to get best results

Data Collection with spaceX API

- Process includes combination of API requests

- Requesting data

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

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

- Converting response to JSON

```
data = pd.json_normalize(response.json())
```

- Using custom functions to clean data

```
# Call getBoosterVersion
getBoosterVersion(data)
```

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

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

- Creating data frame

```
launch_dict = {'FlightNumber': list(data['flight_number']),
'Date': list(data['date']),
'BoosterVersion':BoosterVersion,
'PayloadMass':PayloadMass,
'Orbit':Orbit,
'LaunchSite':LaunchSite,
'Outcome':Outcome,
'Flights':Flights,
'GridFins':GridFins,
'Reused':Reused,
'Legs':Legs,
'LandingPad':LandingPad,
'Block':Block,
'ReusedCount':ReusedCount,
'Serial':Serial,
'Longitude': Longitude,
'Latitude': Latitude}
```

Data Collection with Web Scraping

- Getting Response from HTML

```
html_data = requests.get(static_url).text
```

- Creating Beautiful Soup Object

```
soup = BeautifulSoup(html_data, 'html5lib')
```

- Finding tables and assigning results to a list

```
html_tables = soup.find_all('table')
```

- Extracting column names

```
for row in first_launch_table.find_all('th'):
    name = extract_column_from_header(row)
    if(name != None and len(name) > 0):
        column_names.append(name)
```

- Creating Empty data dictionaries

```
launch_dict= dict.fromkeys(column_names)

# Remove an irrelvant column
del launch_dict['Date and time ( )']

launch_dict['Flight No.'] = []
launch_dict['Launch site'] = []
launch_dict['Payload'] = []
launch_dict['Payload mass'] = []
launch_dict['Orbit'] = []
launch_dict['Customer'] = []
launch_dict['Launch outcome'] = []
# Added some new columns
launch_dict['Version Booster']=[]
launch_dict['Booster landing']=[]
launch_dict['Date']=[]
launch_dict['Time']=[]
```

- Creating a data frame and exporting to csv

```
df=pd.DataFrame(launch_dict)
```

```
df.to_csv('spacex_web_scraped.csv', index=False)
```

Data Wrangling

- Converting failed landings into training labels (1s / 0s)

- Calculating number of launches

```
df['LaunchSite'].value_counts()
```

- Calculating number and occurrence of each orbit

```
df.Orbit.value_counts()
```

- Number of occurrence in mission outcome per orbit type

```
landing_outcomes = df.Outcome.value_counts()
```

- Creating landing outcome label from outcome column

```
landing_class = []
for outcome in df.Outcome:
    if outcome in bad_outcomes:
        landing_class.append(0)
    else:
        landing_class.append(1)
```

- Calculating success rate

```
df["Class"].mean()
```

```
0.6666666666666666
```

- Exporting dataset to CSV

```
df.to_csv("dataset_part_2.csv", index=False)
```

EDA with Data Visualization

- Scatter Chart

  - Flight Number vs Launch Site

  - Payload vs Launch Site

  - Flight Number vs Orbit Type

  - Payload vs Orbit Type

- Bar Chart

  - Orbit Type vs Success Rate

- Line Chart

  - Year vs Success Rate

EDA with SQL

- **Loading the dataset into db2 database and execute queries to answer the following requests**

1. Names of unique launch sites in the space mission
2. 5 records where launch sites begin with 'CCA'
3. Total payload mass carried by boosters launched by NASA
4. Average payload mass carried by booster version F9 v1.1
5. Date when first successful landing outcome in ground pad
6. Total number of successful and failure mission outcomes

Interactive Map with Folium

- Objects created and added to a folium map

  - markers showing all launch sites on a map

  - markers showing successful and failure launches for each site

- By adding these objects , following patterns about launch sites are found

  - Are launch sites close to highways ? Yes

  - Are launch sites close to coastline ? Yes

  - Do they keep certain distance from cities ? Yes

Dashboard with Plotly Dash

- Contains Pie Chart and Scatter Point Chart

Pie Chart

- To show total success launches
- Successful landing distribution across launch sites

Scatter Chart

- Showing relationships between outcomes and payload mass

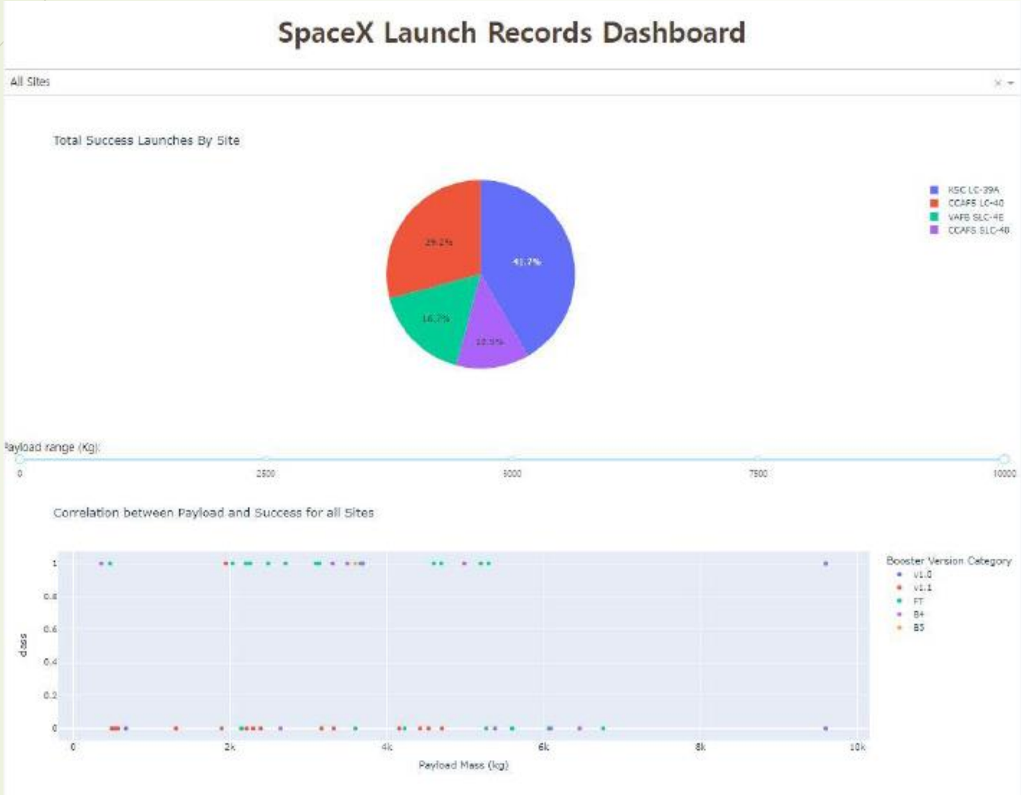
Predictive Analysis

- Perform exploratory data analysis and determine training labels

BUILDING MODEL → EVALUATING MODEL → IMPROVING MODEL → FINDING BEST CLASSIFICATION MODEL

- Find best hyperparameter for SVM, classification trees and logical regression

Results

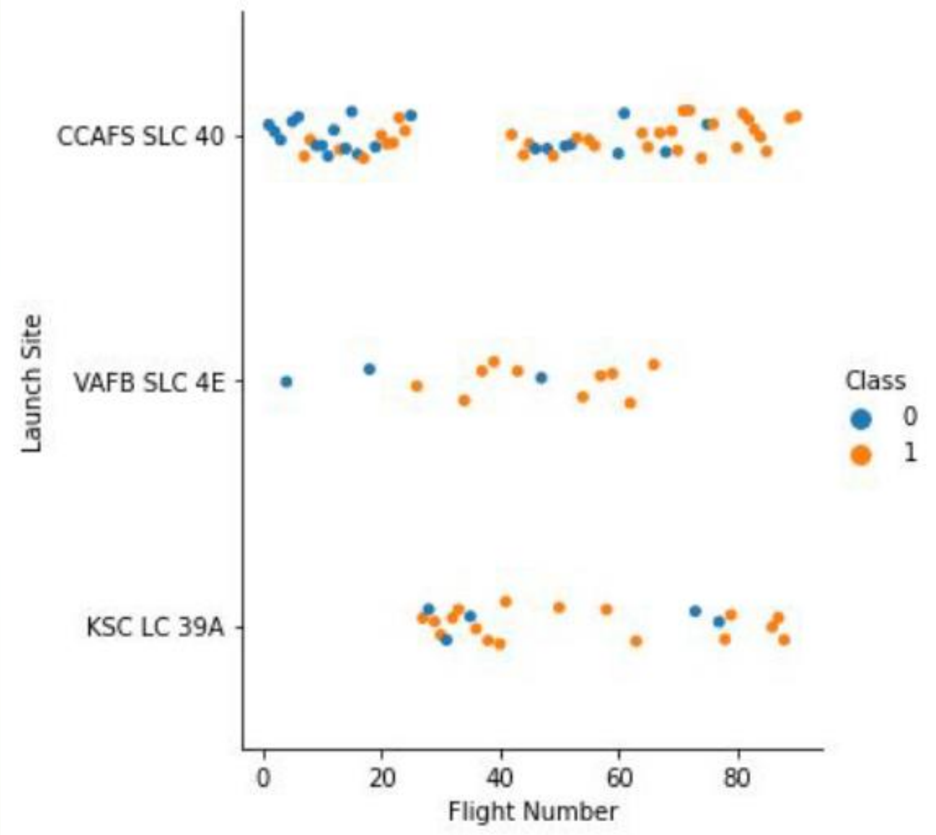


- Preview of dashboard with Plotly dash

EDA Insights

- Flight Number vs Launch Site
- Blue = unsuccessful launch
  Orange= successful launch

- Success rate increase as the number of flights increase

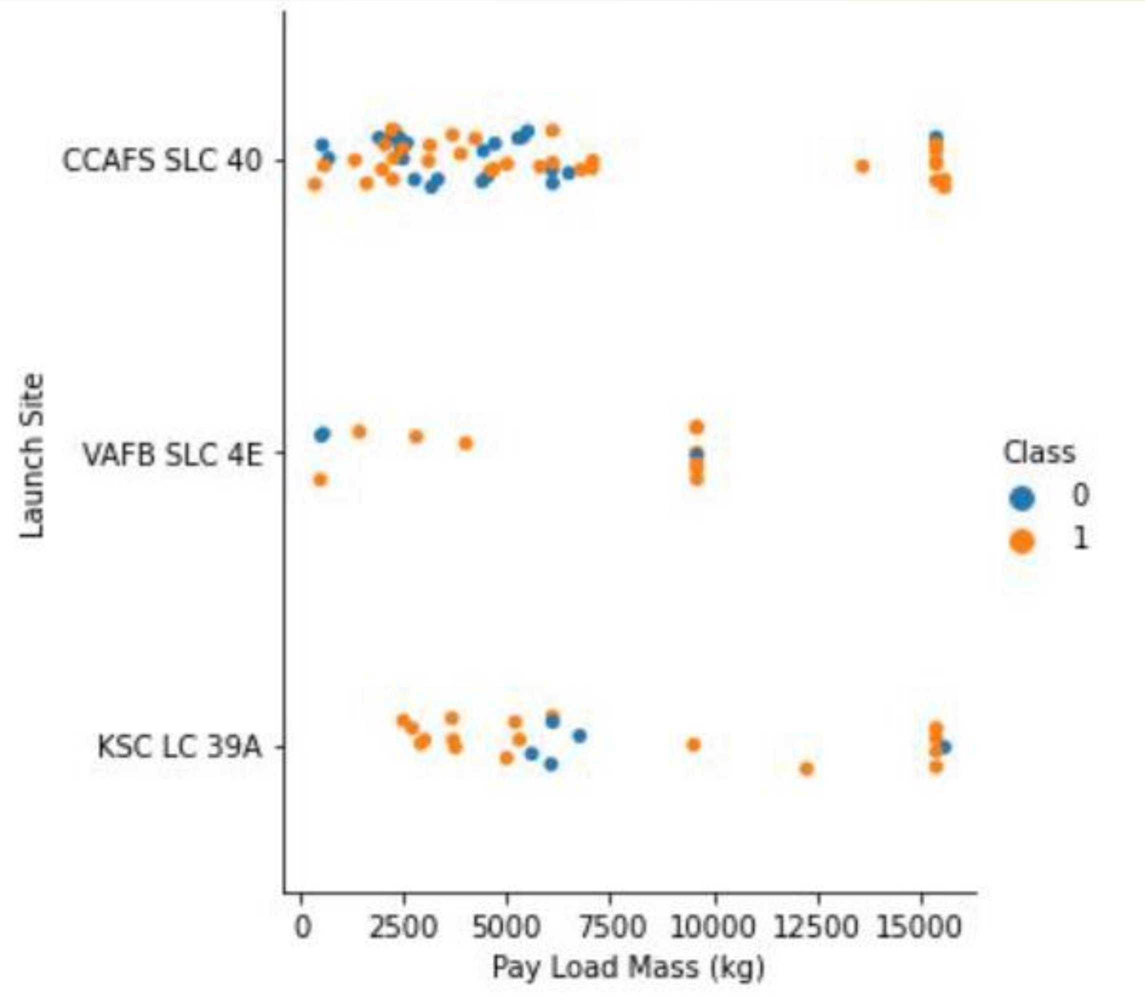


- **Payload vs Launch site**

- Blue = unsuccessful launch

  Orange = successful launch

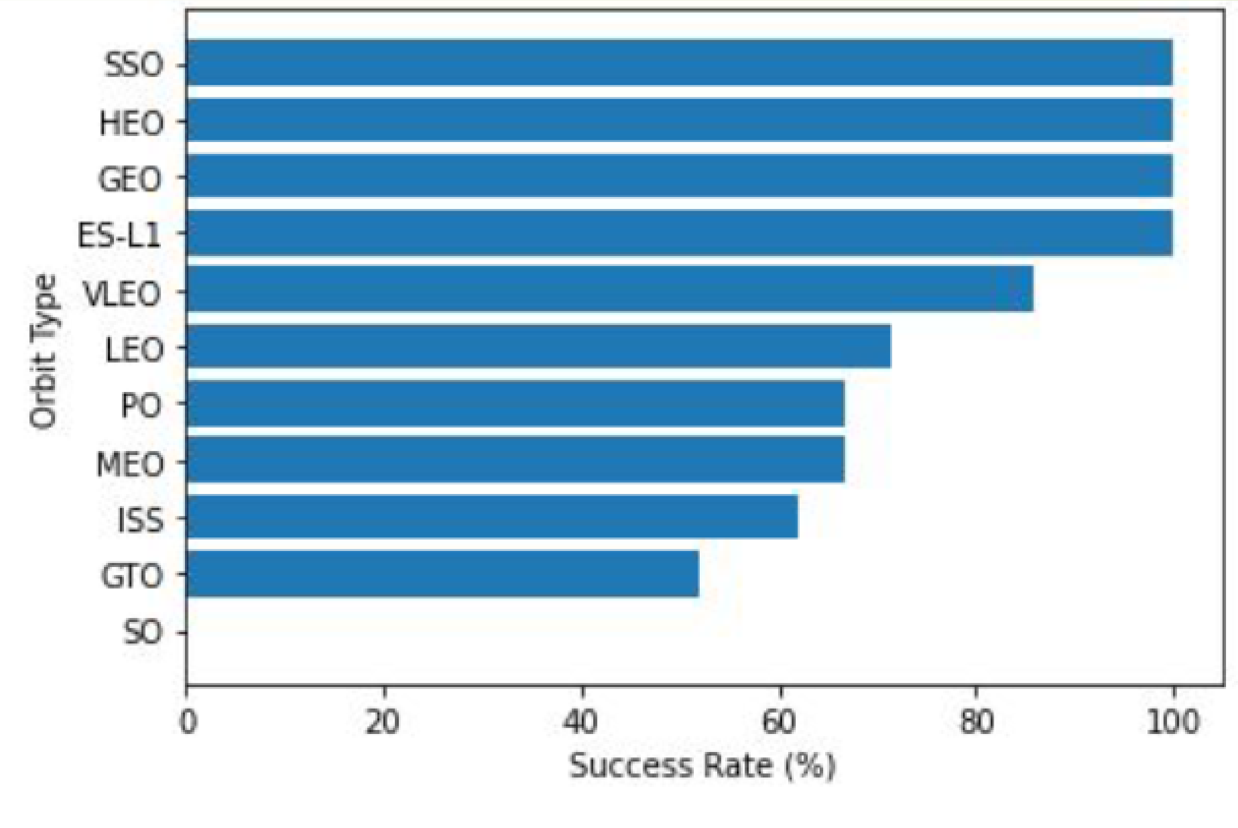
- No clear pattern can be found



- **Success rate vs Orbit type**

  - SSO, HEO, GEO has ESL-1 have the highest success rates

  -

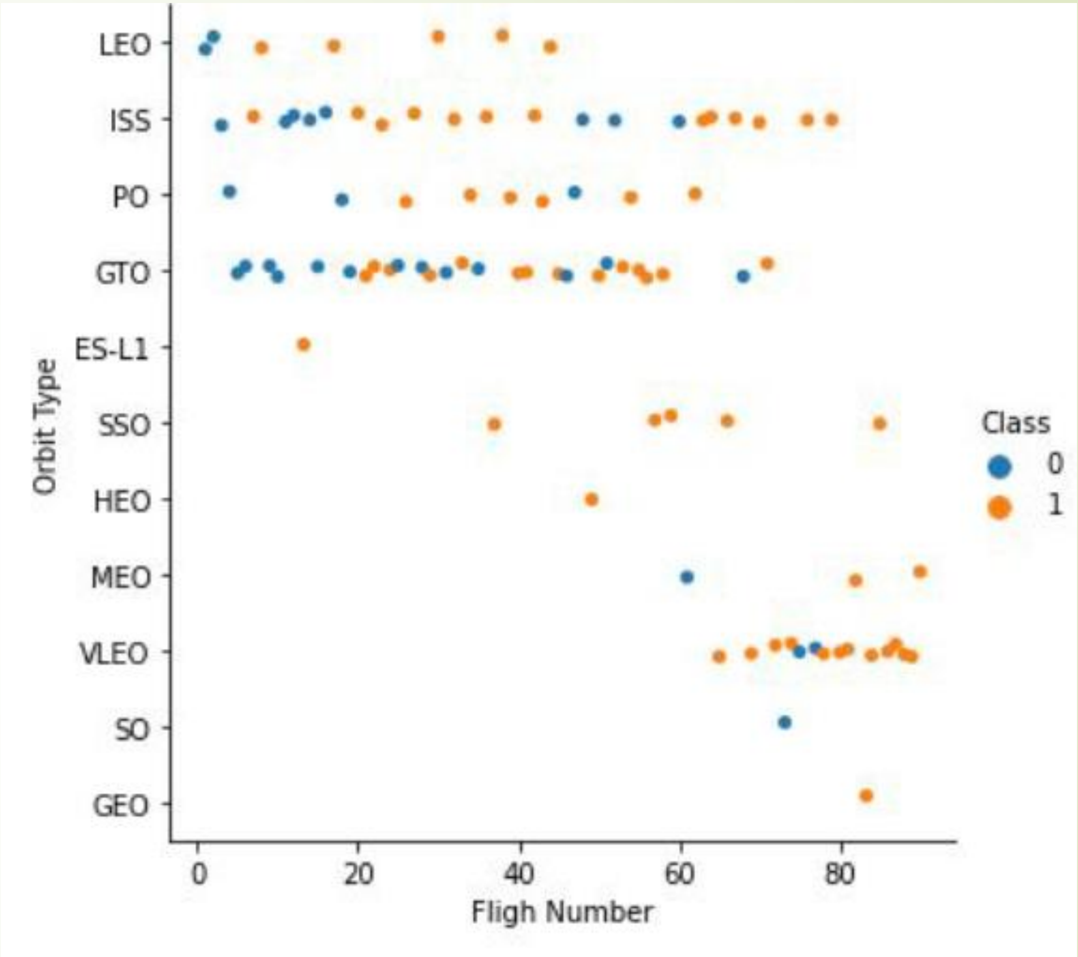


Flight Number vs Orbit Type

- Blue = unsuccessful launch
  Orange = successful launch

- Launch outcome seems to correlate with flight number

- In GTO orbit there is no relationship between flight numbers and success rate

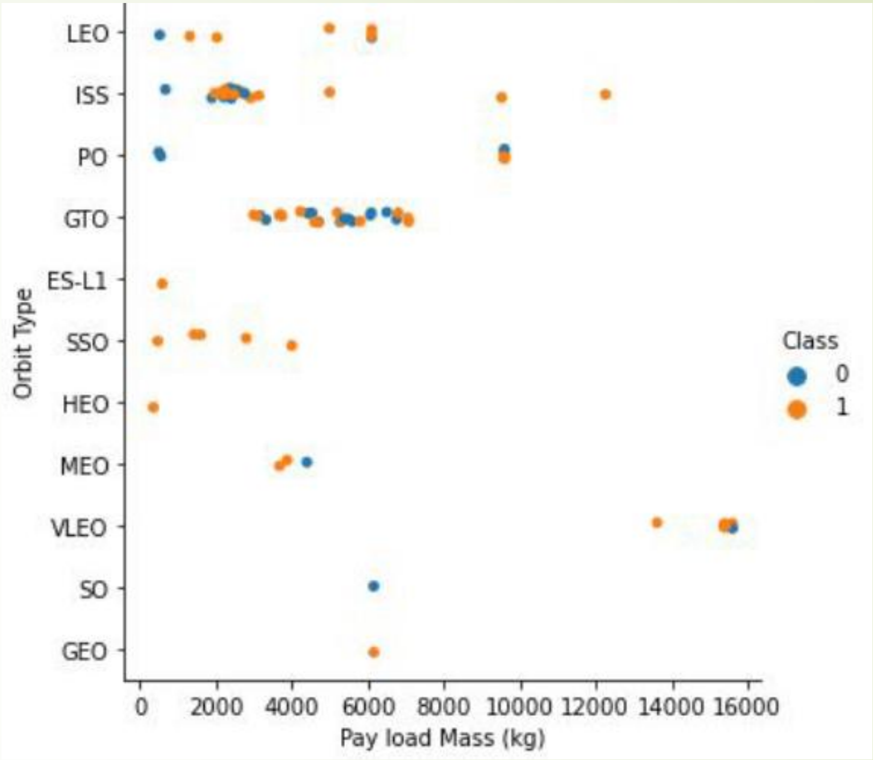


Payload vs Orbit type

- Blue = unsuccessful launch

  Orange = successful launch

- With heavy payloads the successful landing rate is more for LEO and ISS

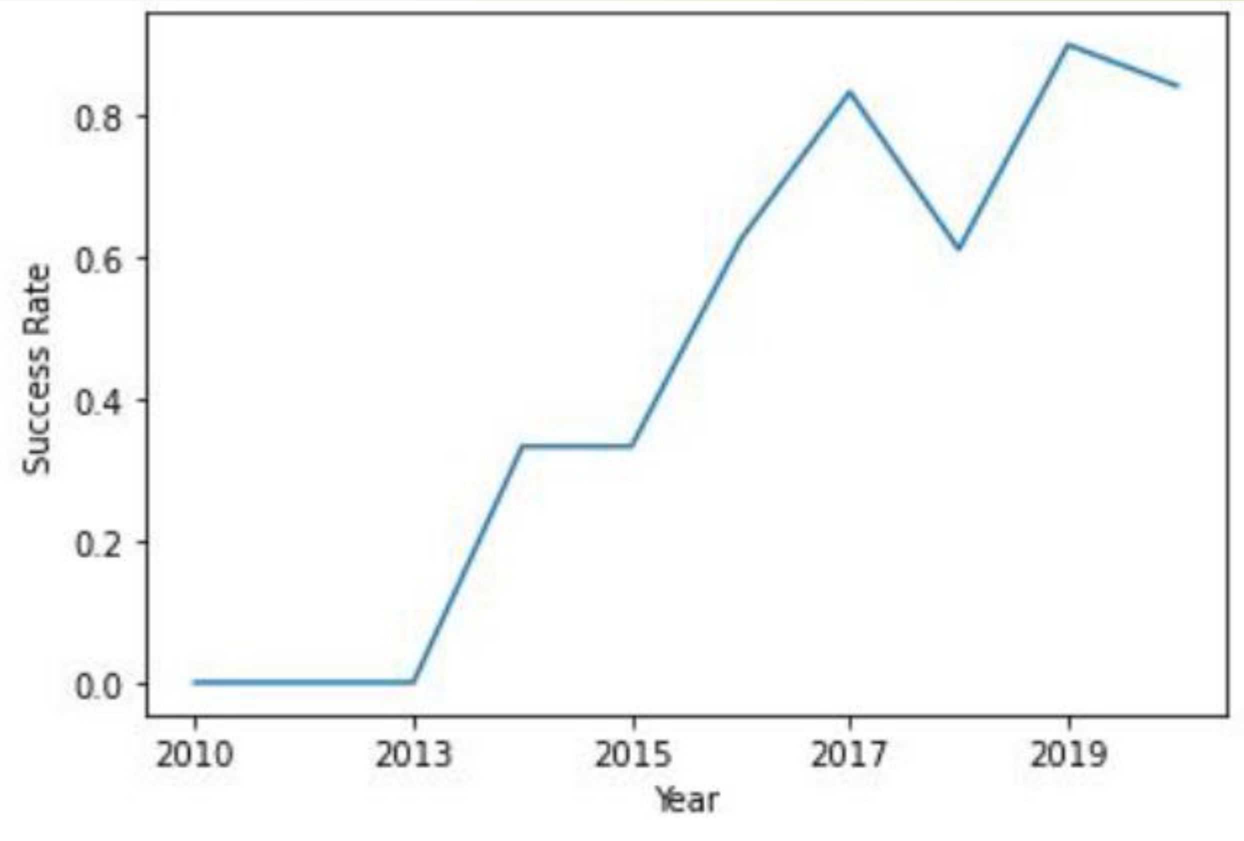


- **Launch Success Yearly trend**

- Success rate has increased from 2013 to 2017

- Decreased slightly in 2018

- Recent success rate is about 80%



EDA with SQL

- **All Launch Site Names**

```
SELECT DISTINCT LAUNCH_SITE
FROM SPACEXTBL
```

Query

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

Result

Launch site names begin with CCA

```
SELECT * FROM SPACEXTBL
WHERE LAUNCH_SITE LIKE 'CCA%'
LIMIT 5
```

Query

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

Result

Total Payload Mass

```sql
SELECT SUM(PAYLOAD_MASS__KG_)
       AS total_payload_mass_kg
FROM SPACEXTBL
WHERE CUSTOMER = 'NASA (CRS)'
```

Query

total_payload_mass_kg
45596

Result

Average Payload Mass by F9 V1.1

```
SELECT AVG(PAYLOAD_MASS__KG_)
      AS avg_payload_mass_kg
FROM SPACEXTBL
WHERE BOOSTER_VERSION = 'F9 v1.1'
```

Query

avg_payload_mass_kg
2928

Result

First Successful ground landing date

```sql
SELECT MIN(DATE)
  AS first_successful_landing_date
FROM SPACEXTBL
WHERE LANDING__OUTCOME
      = 'Success (ground pad)'
```

Query

first_successful_landing_date
2015-12-22

Result

- **Successful Drone Ship landing with payload between 4000 and 6000**

```
SELECT BOOSTER_VERSION
FROM SPACEXTBL
WHERE LANDING__OUTCOME = 'Success (drone ship)'
   AND (PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000)
```

Query

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

Result

Total Number of Successful and Failure Mission Outcomes

```
SELECT MISSION_OUTCOME,
       COUNT(*) AS total_number
FROM SPACEXTBL
GROUP BY MISSION_OUTCOME
```

Query

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

Result

Boosters carried maximum payload

```
SELECT DISTINCT BOOSTER_VERSION,
       PAYLOAD_MASS__KG_
FROM SPACEXTBL
WHERE PAYLOAD_MASS__KG_ = (
       SELECT MAX(PAYLOAD_MASS__KG_)
       FROM SPACEXTBL)
```

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

Result

2015 launch records

```
SELECT LANDING__OUTCOME,
       BOOSTER_VERSION,
       LAUNCH_SITE
FROM SPACEXTBL
WHERE LANDING__OUTCOME
          = 'Failure (drone ship)'
       AND YEAR(DATE) = '2015'
```

Query

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

Result

Rank Landing Outcomes Between 2010 and 2017

```sql
SELECT LANDING__OUTCOME,
       COUNT(LANDING__OUTCOME) AS total_number
FROM SPACEXTBL
WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20'
GROUP BY LANDING__OUTCOME
ORDER BY total_number DESC
```

Query

landing__outcome	total_number
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

Result

Launch Site proximities analysis

- **All Launch Sites Location**

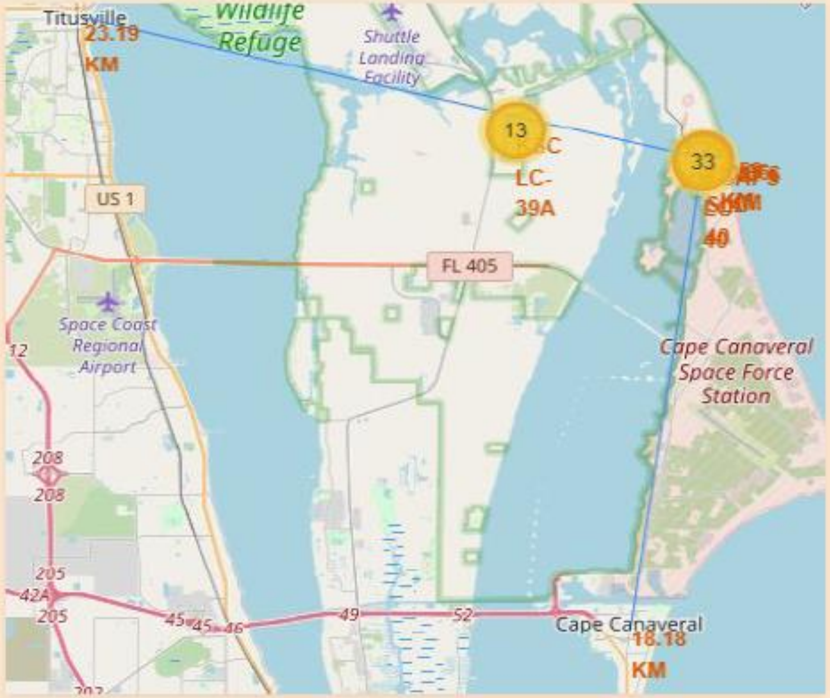
All launch sites are near coastline in the US

Color labeled launch outcomes

- Successful Landing = Green
- Unsuccessful landing = Red

Launch Site Proximities

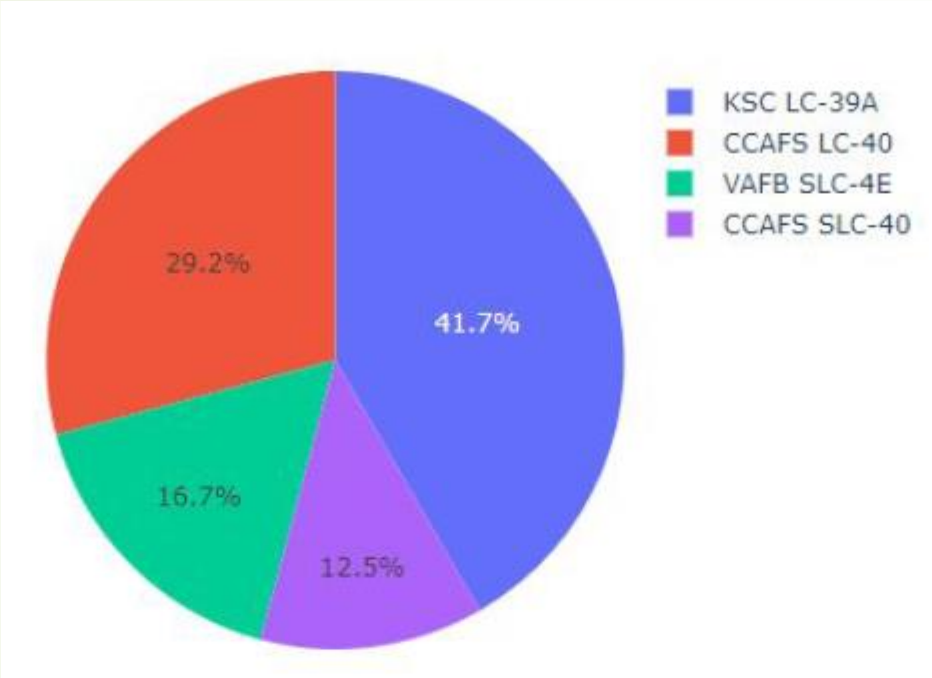
Are launch sites in close proximity to railways? ***Yes***
Are launch sites in close proximity to highways? ***Yes***
Are launch sites in close proximity to coastline? ***Yes***



Do launch sites keep certain distance away from cities? ***Yes***

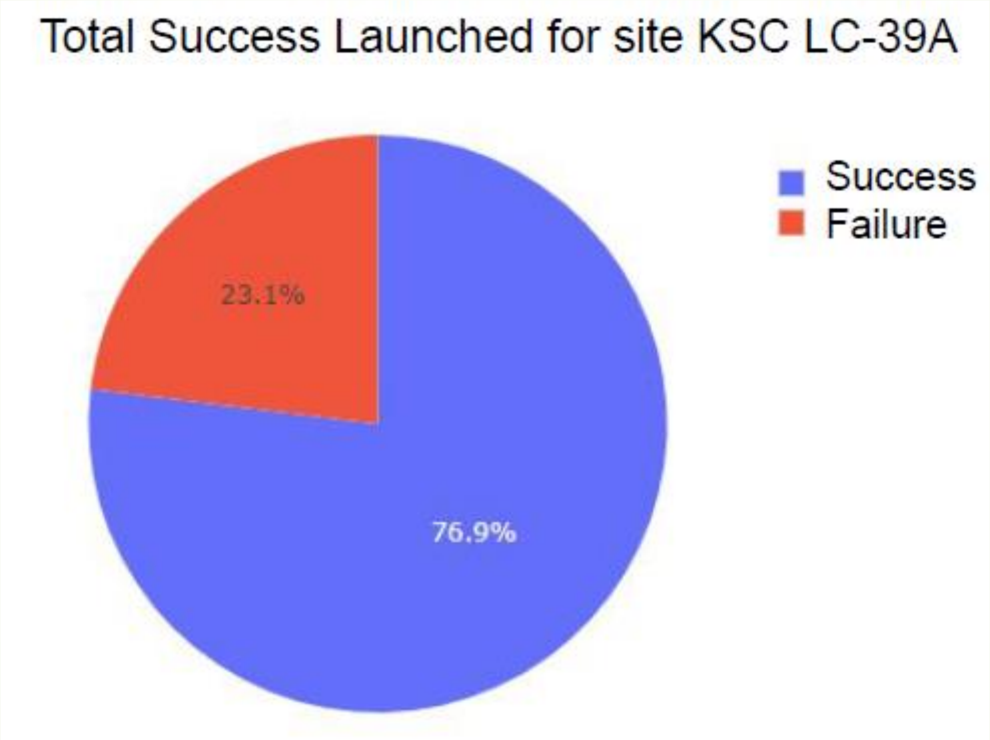
Dashboard with Plotly Dash

- Total Success Launches by all sites

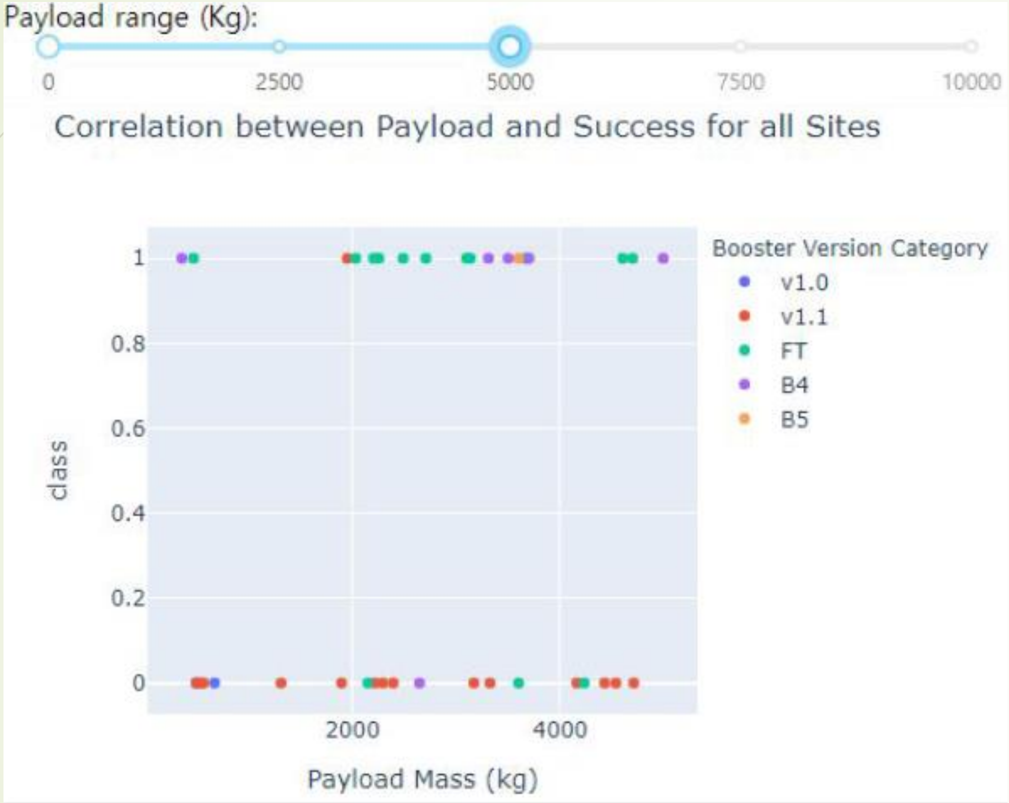


KSC LC-39A records most success among launch sites

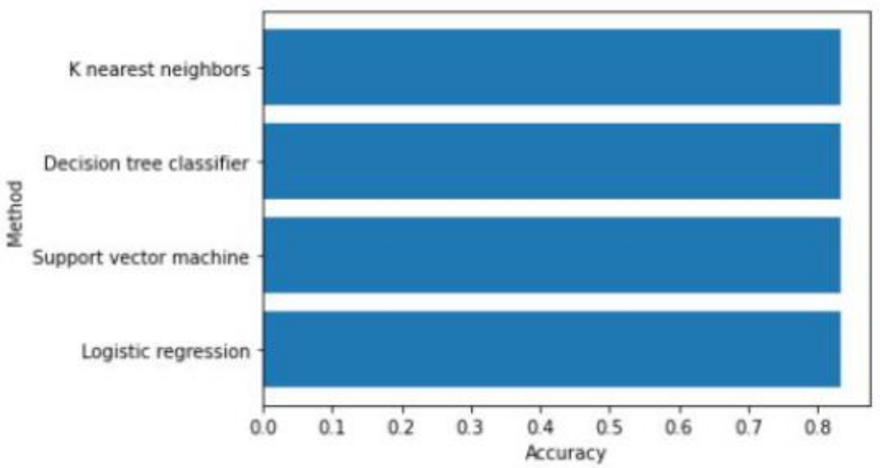
- **Launch Site with Highest Launch Success Ratio**



Payload with Launch Outcome Scatter Plot for all sites



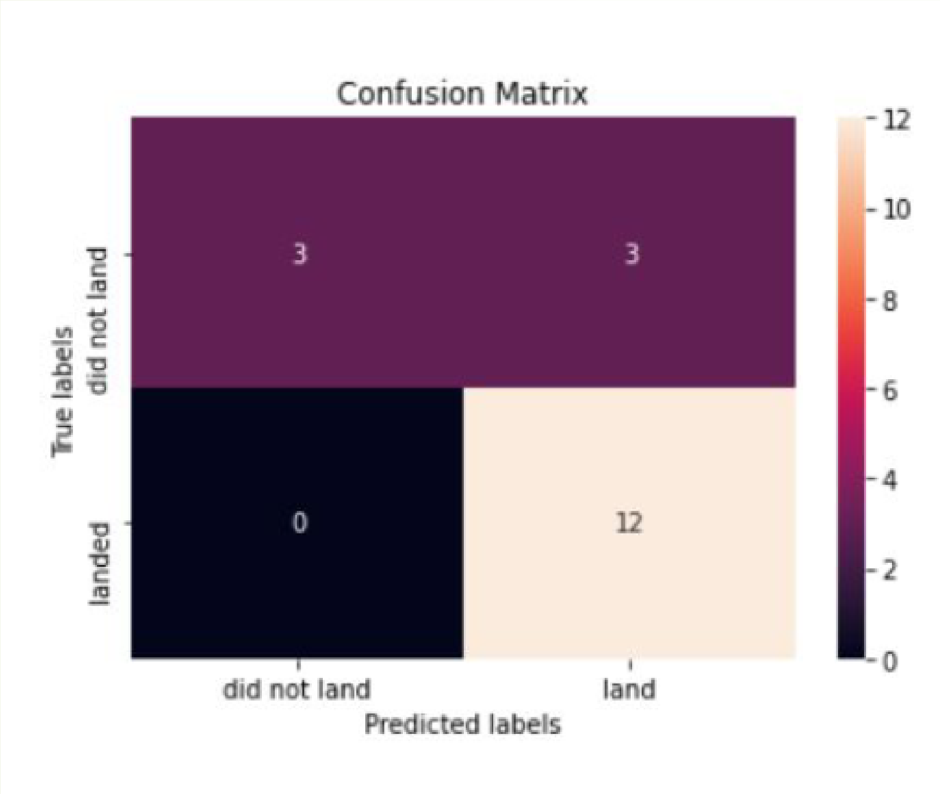
Predictive Analysis



	Method	Accuracy
0	Logistic regression	0.833333
1	Support vector machine	0.833333
2	Decision tree classifier	0.833333
3	K nearest neighbors	0.833333

Accuracy of all models were same

Confusion Matrix



Conclusions

- Number of flights increased → success rate increased
- SSO, HEO, GEO and ES-L1 orbital types have the highest success rate
- Launch site is far from cities but close to railways and coastline
- All models have the same accuracy