IBM Developer
SKILLS NETWORK

# Winning Space Race with Data Science

<Name>
<Date>

# Outline

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

# Executive Summary

- Summary of methodologies
- Summary of all results

# Introduction

- Project background and context

The reason SpaceX can launch rockets at a lower cost than other suppliers is that SpaceX can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch.

- Problems you want to find answers

Predicting if the Falcon 9 first stage will land successfully

Section 1

# Methodology

# Methodology

## Executive Summary

- Data collection methodology:
  - Describe how data was collected
- Perform data wrangling
  - Describe how data was processed
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

# Data Collection

- Describe how data sets were collected.
- You need to present your data collection process use key phrases and flowcharts

# Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- https://eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/62671f0c-4242-443e-9ff7-65d5cab2766b/view?access_token=f98e57357254bb0ddc32309390980b731574b42ae765977374ffb60ea1b35094

Place your flowchart of SpaceX API calls here:

1. getBoosterVersion: Rocket column
2. getLaunchSite: Payloads column
3. getPayloadData: Launchpad column
4. getCoreData: Cores column

# Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- https://eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/62671f0c-4242-443e-9ff7-65d5cab2766b/view?access_token=f98e57357254bb0ddc32309390980b731574b42ae765977374ffb60ea1b35094

Place your flowchart of web scraping here:

1. Request and parse the SpaceX launch data using the GET request
2. Combine the needed columns into a dictionary
3. Create a Pandas data frame
4. Filter the data frame to only include Falcon 9 launches

# Data Wrangling

- Describe how data were processed
- You need to present your data wrangling process using key phrases and flowcharts
- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose

# Data Wrangling

- Present your data wrangling process using key phrases and flowcharts
- https://eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/20b64490-c0ad-40f4-923f-e489b65c101a/view?access_token=c72cd4666c3928d8eb0322276981a90684361ba47324cf088b88093e70a53fef

Place your flowchart of data wrangling process here:

1. Request the Falcon9 Launch Wiki page from its URL
2. Extract all column/variable names from the HTML table header
3. Create a data frame by parsing the launch HTML tables

# EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts

- https://eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/6c463244-938a-4020-acc9-a92c82922138/view?access_token=bf727b34ba72da98877fc6841076c8f557bdbe8daad465ef02307197c8def335

Summarize what charts were plotted and why you used those charts:

1. Scatter point chart: The distribution pattern of scatter points is used to reflect the statistical relationship between features.
2. Bar chart: Visually display the distribution of one parameter on another.
3. Line chart: Viewing the trend of the dependent variable y as the independent variable x is best used to display continuous data that has changed over time

# EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
- https://eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/bd911b67-5171-486c-b4af-6253400d2065/view?access_token=3ecfdf8e7d00f7d15a8ff8892c093f801303d5f0838aed7b1f8df63175c9d7ab

Summarize the SQL queries you performed:

1. SELECT FROM WHERE
2. GROUP BY, DESC
3. LIKE
4. LIMIT
5. SUM, AVG, MIN, BETWEEN
6. COUNT
7. AS, AND, OR

# Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
- Explain why you added those objects
- https://eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/bbb4998a-f5ea-4471-892d-43088bd29b40/view?access_token=0defe5cfa878732d6657f3ba46b20f9019a4eafe99f1df18d422efb0bfc41fc1

Summarize what map objects  you performed:

1. folium.Circle: To add a highlighted circle area with a text label on a specific coordinate.
2. folium.Map: To create a folium Map object.
3. folium.Marker: To create a Marker with an icon as a text label.
4. MarkerCluster: To simplify a map containing many markers having the same coordinate.
5. folium.PolyLine: To calculate the distances between a launch site to its proximities

# Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
- Explain why you added those plots and interactions
- https://eu-gb.dataplatform.cloud.ibm.com/analytics/notebooks/v2/a3c05df6-f52d-4eff-afc6-2004aeb38c23/view?access_token=aa94887309b36b360d714a6e2b509782a4c0ba67d3872dde8bd515607e52def9

Summarize what plots/graphs and interactions you have added to a dashboard:

1. dcc.Graph(figure=px.pie()): To visualize launch success counts with selected launch site.
2. dcc.RangeSlider(): To find if variable payload is correlated to mission outcome.
3. dcc.Graph(figure=px.scatter()): To observe how payload may be correlated with mission outcomes for selected site(s).

# Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
- You need present your model development process using key phrases and flowchart
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose

Summarize how you built, evaluated, improved, and found the best performing classification model:

1. Get dataframe Y with numpy format;
2. Standardize the data in X
3. Use the function train_test_split to split the data X and Y into training and test data
4. Using GridSearchCV, score, confusion_matrix to get best logreg_cv, svm_cv, tree_cv, knn_cv
5. The best fitting results are obtained by comparison

# Results

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

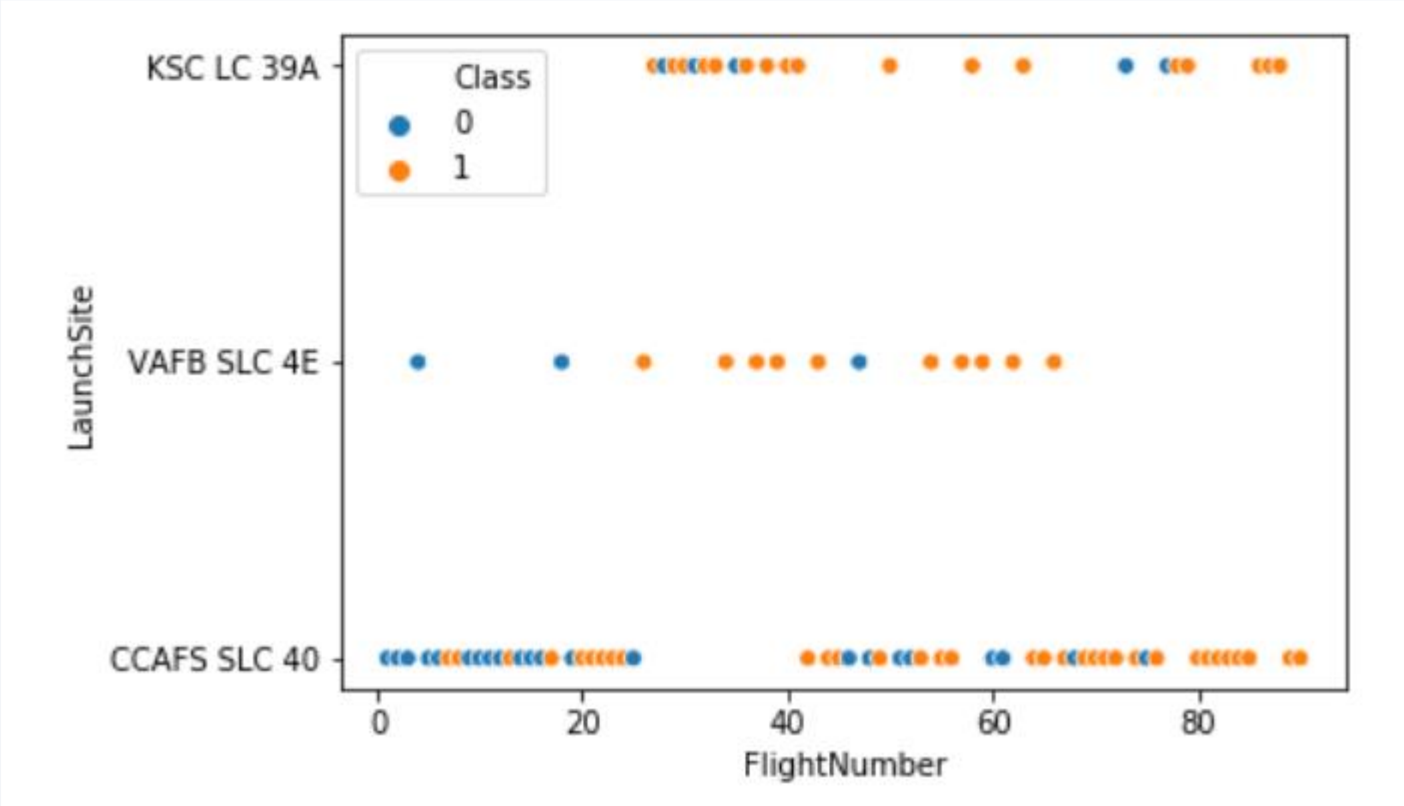
Section 2

# Insights drawn from EDA

# Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site

- Show the screenshot of the scatter plot with explanations

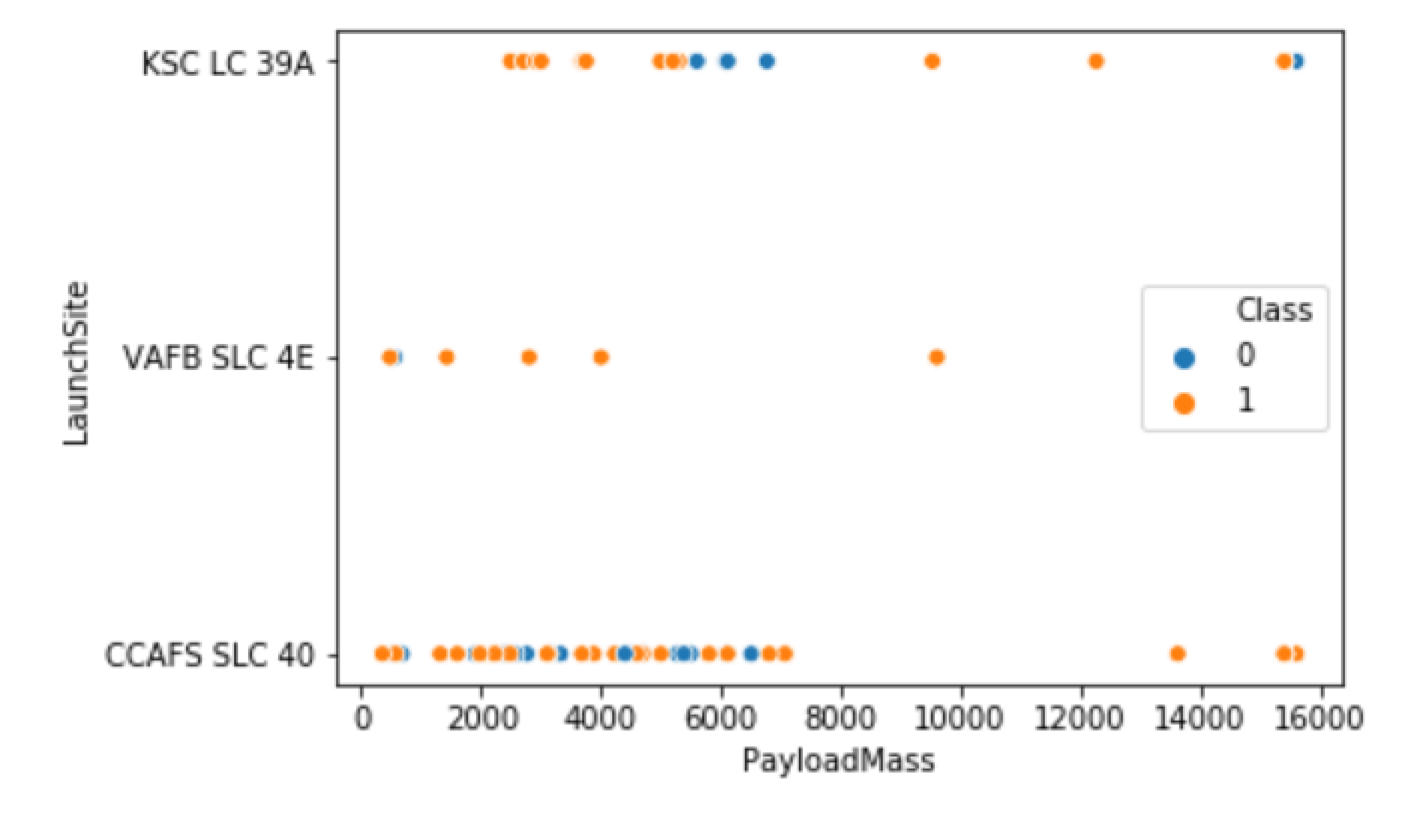



- The launch success rate at all three sites is increasing

# Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site

- Show the screenshot of the scatter plot with explanations

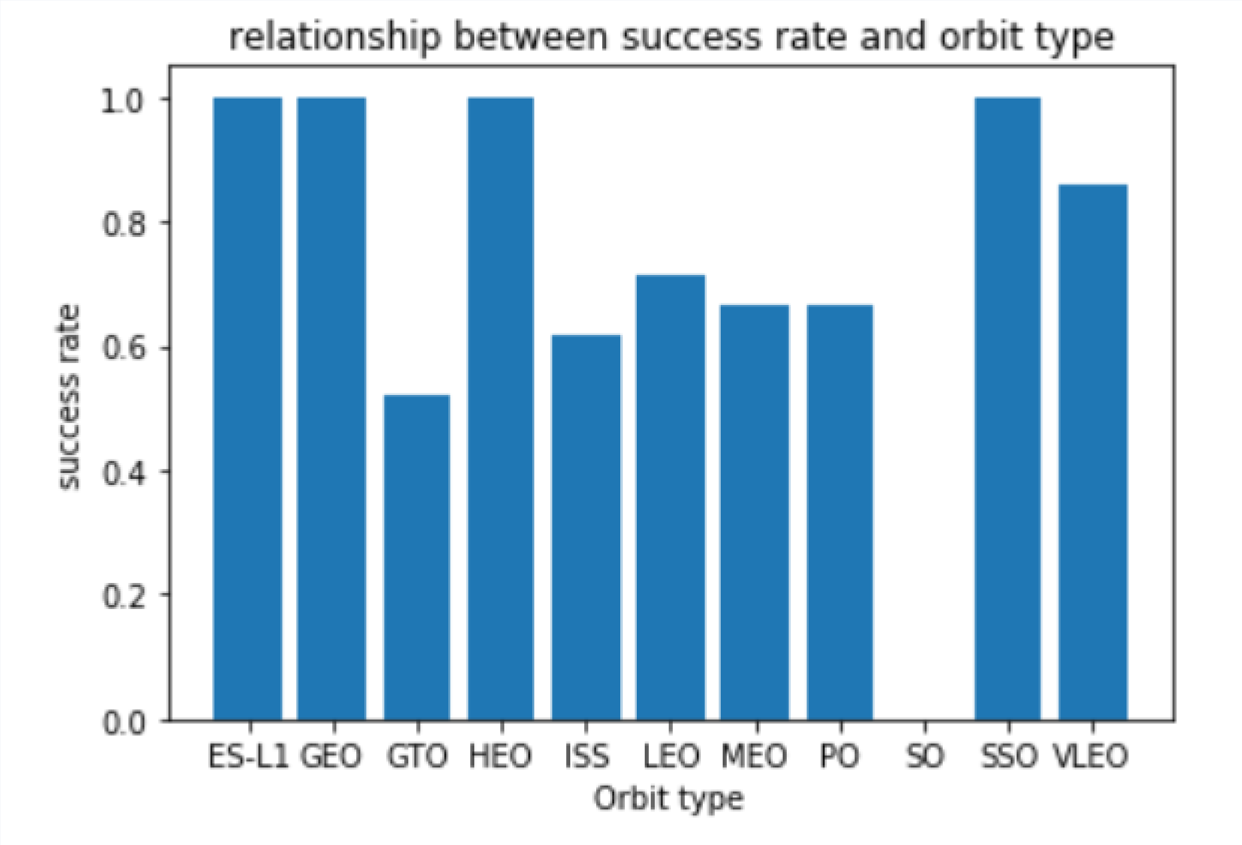



- VAFB-SLC launch site there are no rockets launched for heavy payload mass(greater than 10000)

# Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type

- Show the screenshot of the scatter plot with explanations

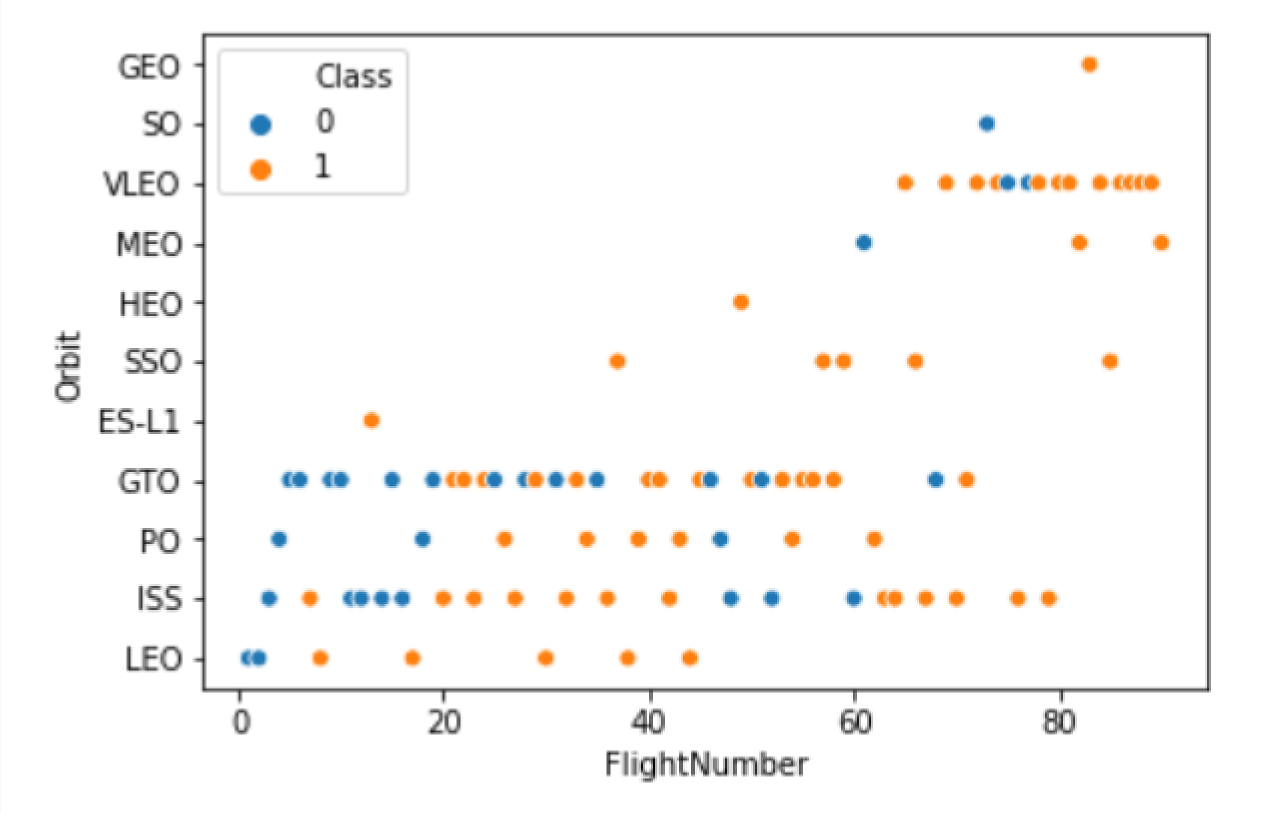



- ES-L1,GEO,HEO,SSO have high success rate.

# Flight Number vs. Orbit Type

- Show a scatter point of Flight number vs. Orbit type

- Show the screenshot of the scatter plot with explanations

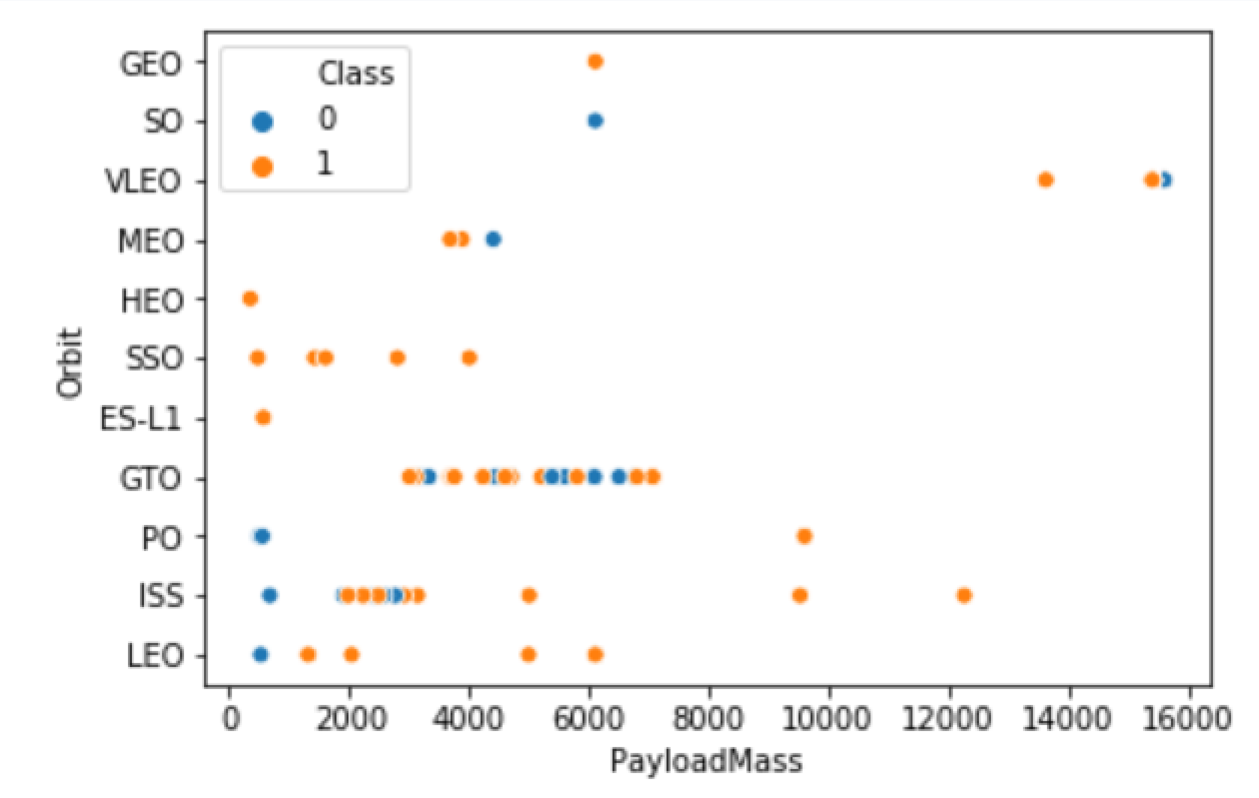



- You should see that in the LEO orbit the Success appears related to the number of flights; on the other hand, there seems to be no relationship between flight number when in GTO orbit.

# Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type

- Show the screenshot of the scatter plot with explanations



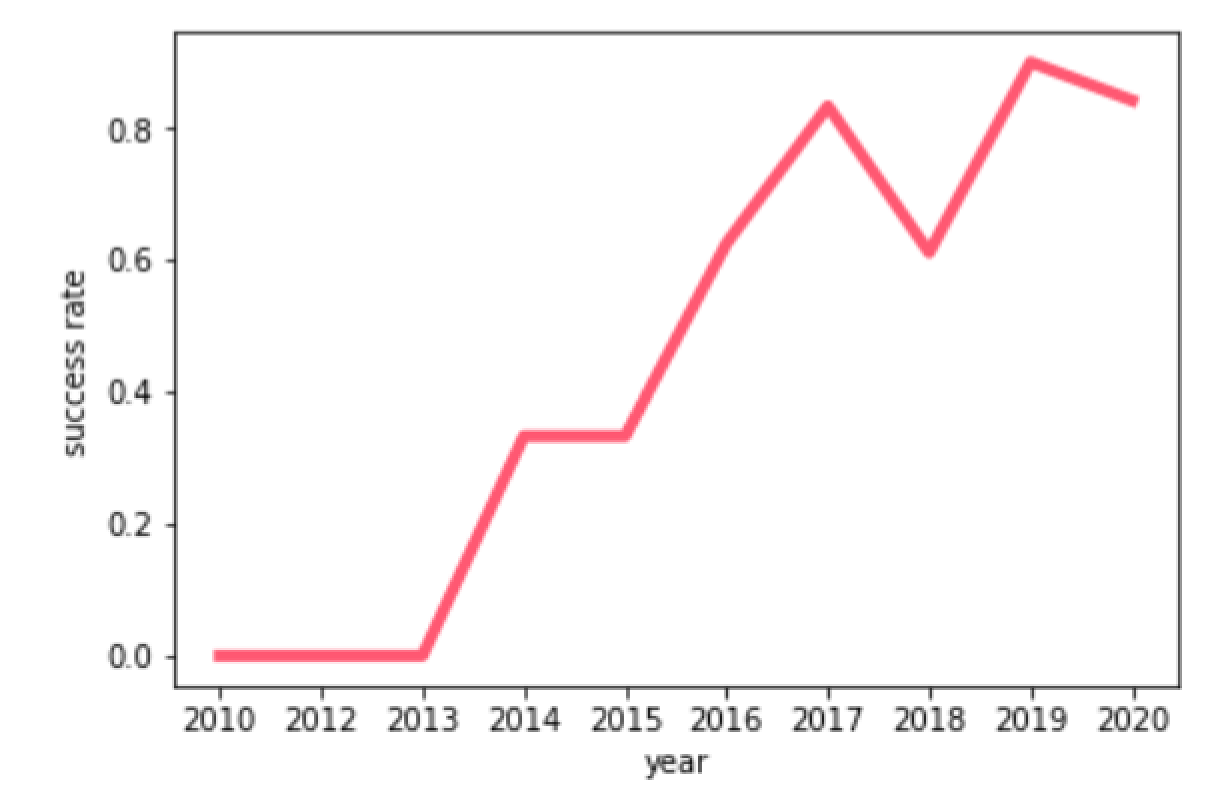

- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.

- However for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here.

# Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Show the screenshot of the scatter plot with explanations




- you can observe that the success rate since 2013 kept increasing till 2020 except 2018

# All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here

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

- There are four kinds of unique launch sites by GROUP BY

# Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Present your query result with a short explanation here

| Date | Time (UTC) | Booster_Version | Launch_Site | Payload | PAYLOAD_MASS__KG_ | Orbit | Customer | Mission_Outcome | Landing _Outcome |
|---|---|---|---|---|---|---|---|---|---|
| 04-06-2010 | 18:45:00 | F9 v1.0 B0003 | CCAFS LC-40 | Dragon Spacecraft Qualification Unit | 0 | LEO | SpaceX | Success | Failure (parachute) |
| 08-12-2010 | 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) |
| 22-05-2012 | 07:44:00 | F9 v1.0 B0005 | CCAFS LC-40 | Dragon demo flight C2 | 525 | LEO (ISS) | NASA (COTS) | Success | No attempt |
| 08-10-2012 | 00:35:00 | F9 v1.0 B0006 | CCAFS LC-40 | SpaceX CRS-1 | 500 | LEO (ISS) | NASA (CRS) | Success | No attempt |
| 01-03-2013 | 15:10:00 | F9 v1.0 B0007 | CCAFS LC-40 | SpaceX CRS-2 | 677 | LEO (ISS) | NASA (CRS) | Success | No attempt |



- Get each launch site begin with 'CCA' by LIKE and LIMIT

# Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here

| SUM(PAYLOAD_MASS__KG_) |
| --- |
| 45596 |

- Get the total payload carried by boosters from NASA by SUM and WHERE

# Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here

| AVG(PAYLOAD_MASS__KG_) |
|---|
| 2928.4 |

- Get the average payload mass carried by booster version F9 v1.1 by AVG and WHERE

# First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Present your query result with a short explanation here

| Date | first successful landing outcome in ground pad |
| --- | --- |
| 22-12-2015 | 2015122022 |

- Get the dates of the first successful landing outcome on ground pad by MIN and rewriting the date in year - month - day format

# 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

- Present your query result with a short explanation here

| Booster_Version |
|---|
| F9 FT B1022 |
| F9 FT B1026 |
| F9 FT B1021.2 |
| F9 FT B1031.2 |

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

# Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Present your query result with a short explanation here

| SUC_NUM |
|---|
| 38 |

| FAL_NUM |
|---|
| 3 |

- Using LIKE

# 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
- Using subquery

| 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

- Present your query result with a short explanation here

| Landing _Outcome | Booster_Version | Launch_Site |
|---|---|---|
| Failure | F9 B5B1050 | CCAFS SLC-40 |
| Failure | F9 B5 B1056.4 | CCAFS SLC-40 |
| Failure | F9 B5 B1048.5 | KSC LC-39A |

- Using LIKE 'Failure' OR 'Failure%' AND Date LIKE '%2015'

# 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

- Present your query result with a short explanation here

- Using ORDER BY and DESC

| Date | Landing _Outcome |
|---|---|
| 19-02-2017 | Success (ground pad) |
| 14-01-2017 | Success (drone ship) |
| 14-08-2016 | Success (drone ship) |
| 18-07-2016 | Success (ground pad) |
| 15-06-2016 | Failure (drone ship) |
| 27-05-2016 | Success (drone ship) |
| 06-05-2016 | Success (drone ship) |
| 08-04-2016 | Success (drone ship) |
| 04-03-2016 | Failure (drone ship) |
| 17-01-2016 | Failure (drone ship) |
| 22-12-2015 | Success (ground pad) |
| 28-06-2015 | Precluded (drone ship) |
| 27-04-2015 | No attempt |
| 14-04-2015 | Failure (drone ship) |
| 02-03-2015 | No attempt |
| 11-02-2015 | Controlled (ocean) |
| 10-01-2015 | Failure (drone ship) |
| 21-09-2014 | Uncontrolled (ocean) |
| 07-09-2014 | No attempt |
| 05-08-2014 | No attempt |
| 14-07-2014 | Controlled (ocean) |
| 18-04-2014 | Controlled (ocean) |
| 06-01-2014 | No attempt |
| 03-12-2013 | No attempt |
| 29-09-2013 | Uncontrolled (ocean) |
| 01-03-2013 | No attempt |
| 08-10-2012 | No attempt |
| 22-05-2012 | No attempt |
| 08-12-2010 | Failure (parachute) |
| 04-06-2010 | Failure (parachute) |

Section 3

# Launch Sites Proximities Analysis

# <Folium Map Screenshot 1>

- Replace <Folium map screenshot 1> title with an appropriate title
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map
- Explain the important elements and findings on the screenshot

**Note: Because the default OpenStreetMap form is not displayed, the surrounding city, railway, highway names are not visible in Stamen Terrain form.**

Seattle
WA
MT
ND
MN
NB
PE
NS
Montréal
OR
ID
SD
WI
MI
Toronto
NH
Boston
WY
NY
Chicago
IA
NE
PA
Denver
New York City
Philadelphia
NV
UT
IL
IN
OH
San Francisco
CO
KS
MO
KY
VA
CA
OK
AR
TN
NC
Los Angeles
AZ
NM
Atlanta
SC
VAFB SLC-4E
Dallas
MS
AL
GA
LA
TX
Houston
FL
Miami
Havana
MEXICO
CUBA
DOMINICAN REPUBLIC
Mexico City
Guadalajara
HONDURAS
NICARAGUA
Caracas
PANAMA
VENEZUELA

# <Folium Map Screenshot 2>

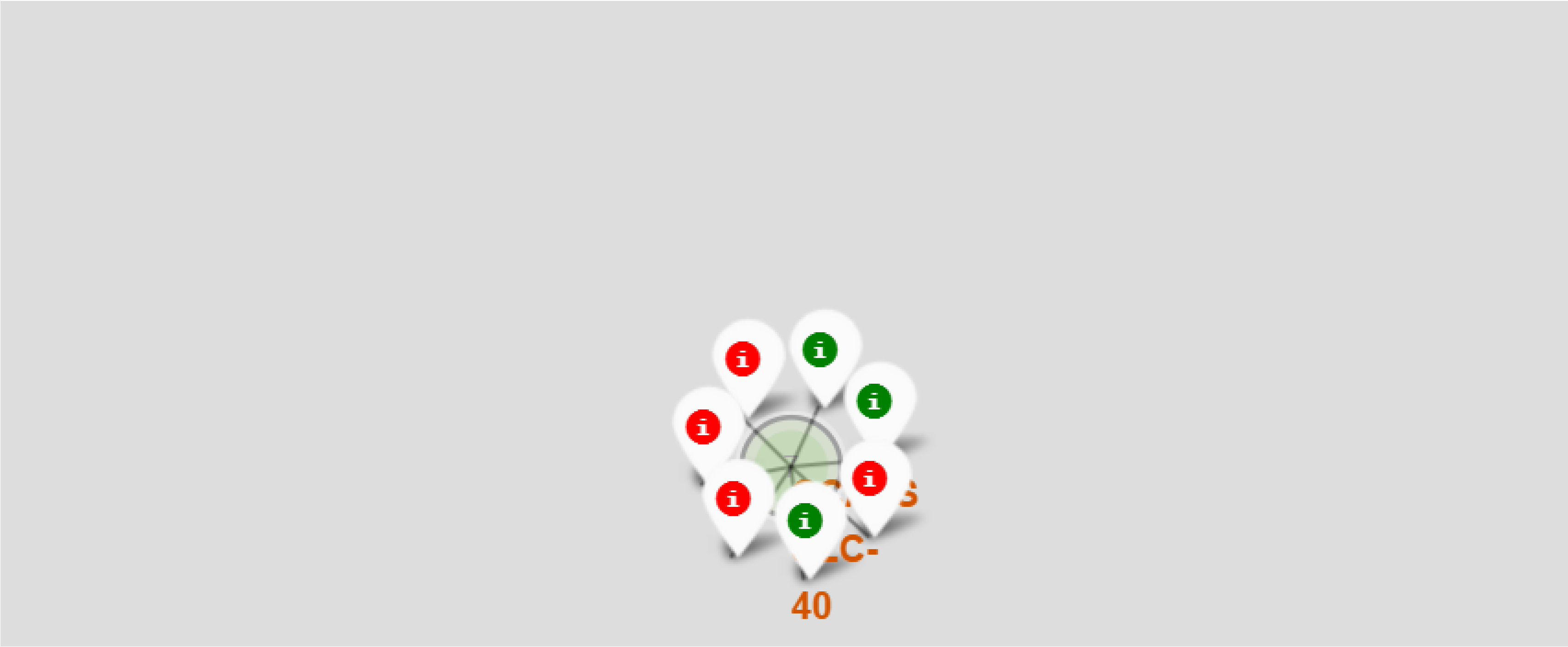
- Replace <Folium map screenshot 2> title with an appropriate title

- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map

- Explain the important elements and findings on the screenshot


OR
ID
WY
SD
MN
WI
MI
IA
NE
NV
UT
CO
KS
MO
IL
IN
OH
PA
NY
NH
NS
PE
KY
VA
TN
NC
SC
GA
AL
MS
LA
AR
OK
TX
NM
AZ
CA
FL
Montréal
Toronto
Boston
Chicago
New York City
Philadelphia
Denver
San Francisco
Los Angeles
Dallas
Houston
Atlanta
Miami
Havana
Guadalajara
Mexico City
Caracas
MEXICO
CUBA
DOMINICAN REPUBLIC
HONDURAS
NICARAGUA
10
VAFB
SLC-4E
46

**High resolution is not displayed due to network problems**



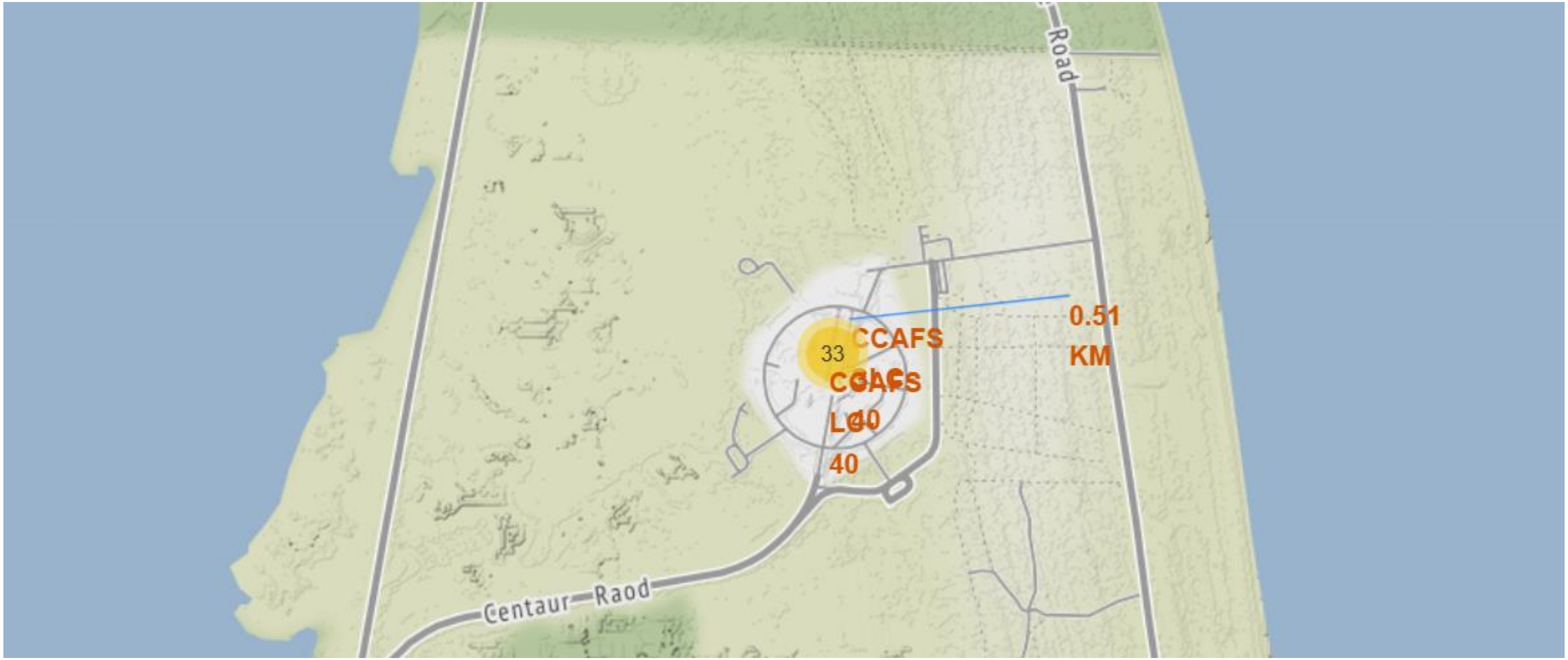
# <Folium Map Screenshot 3>

- Replace <Folium map screenshot 3> title with an appropriate title

- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed

- Explain the important elements and findings on the screenshot

**coastline_coordinate = [28.56367, -80.57163]**



Section 4

# Build a Dashboard with Plotly Dash

# <Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot

# SpaceX Launch Records Dashboard

## Report Type:

All Sites

Total Success Launches By Sites


**KSC LC-39A > CCAFS LC-40 > VAFB SLC-4E > CCAFS SLC-40**

# <Dashboard Screenshot 2>

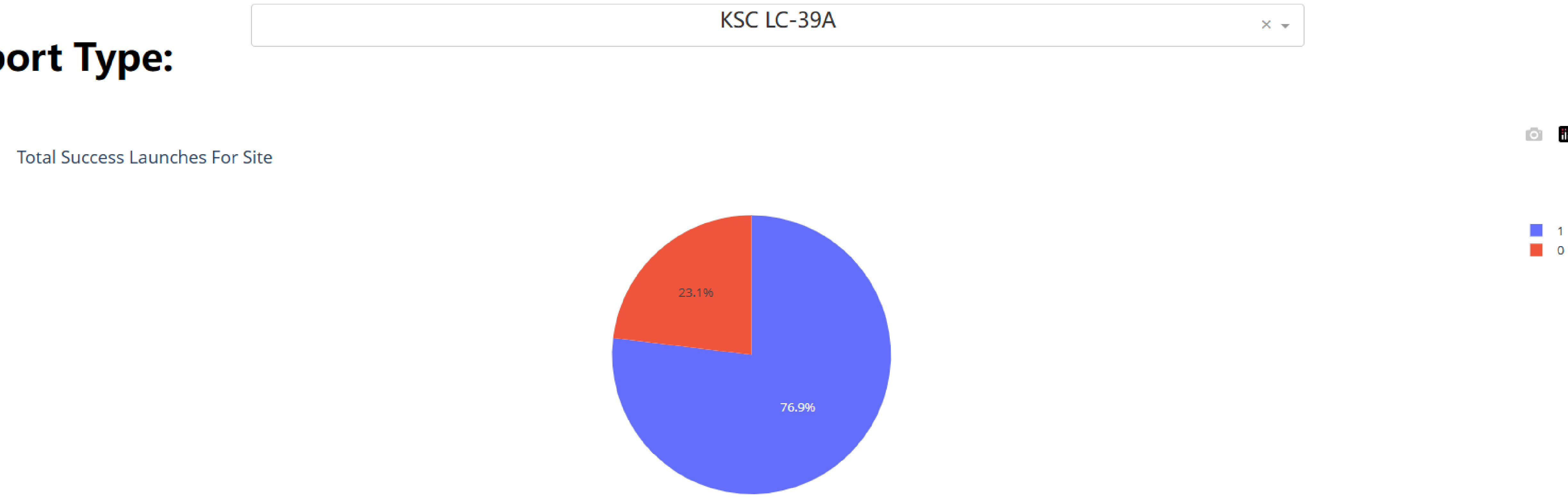
- Replace <Dashboard screenshot 2> title with an appropriate title

- Show the screenshot of the piechart for the launch site with highest launch success ratio

- Explain the important elements and findings on the screenshot

# SpaceX Launch Records Dashboard

## Report Type:

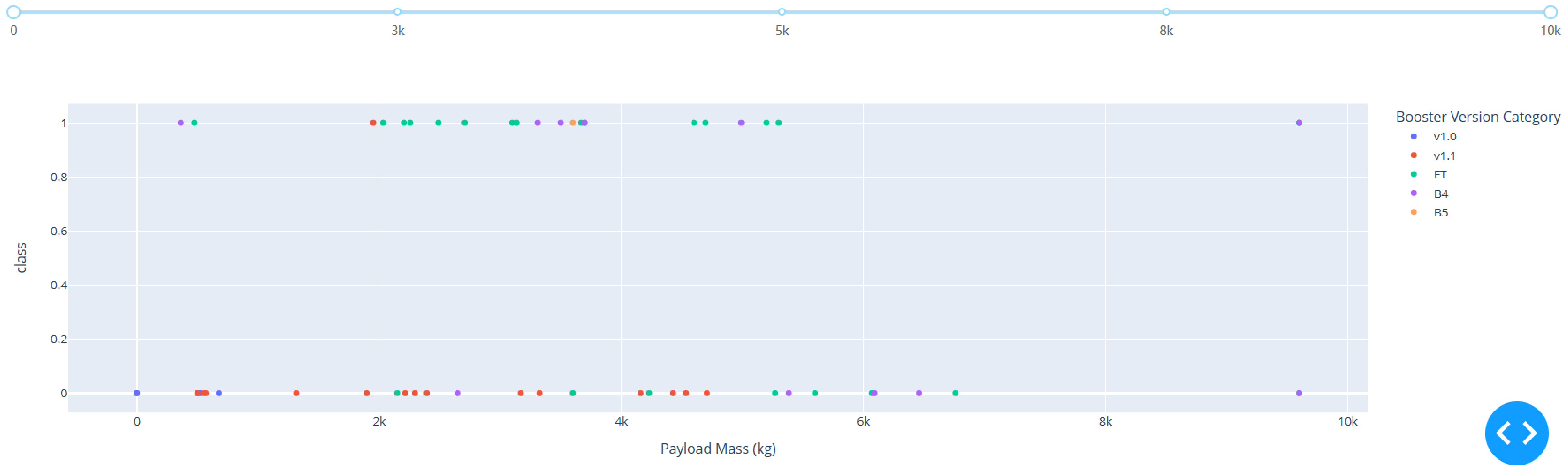



**The highest is KSC LC-39A**

# <Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

# Payload Mass (kg):




**The highest is FT with green**

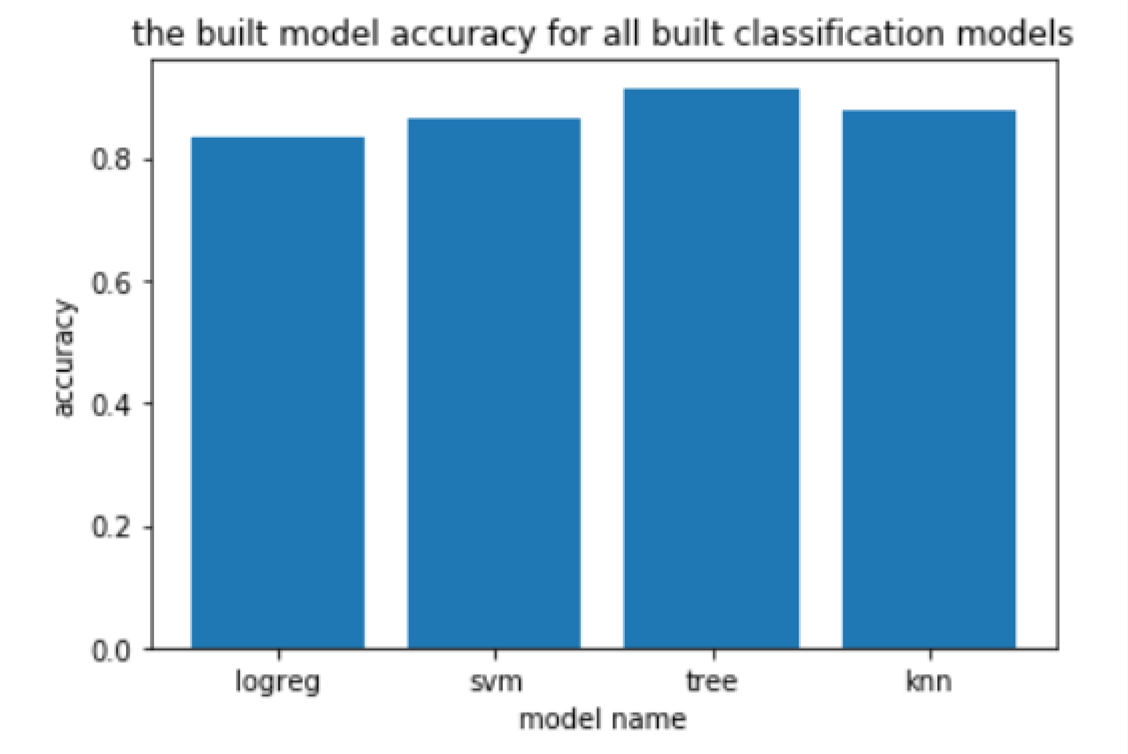
Section 5

# Predictive Analysis (Classification)

# Classification Accuracy

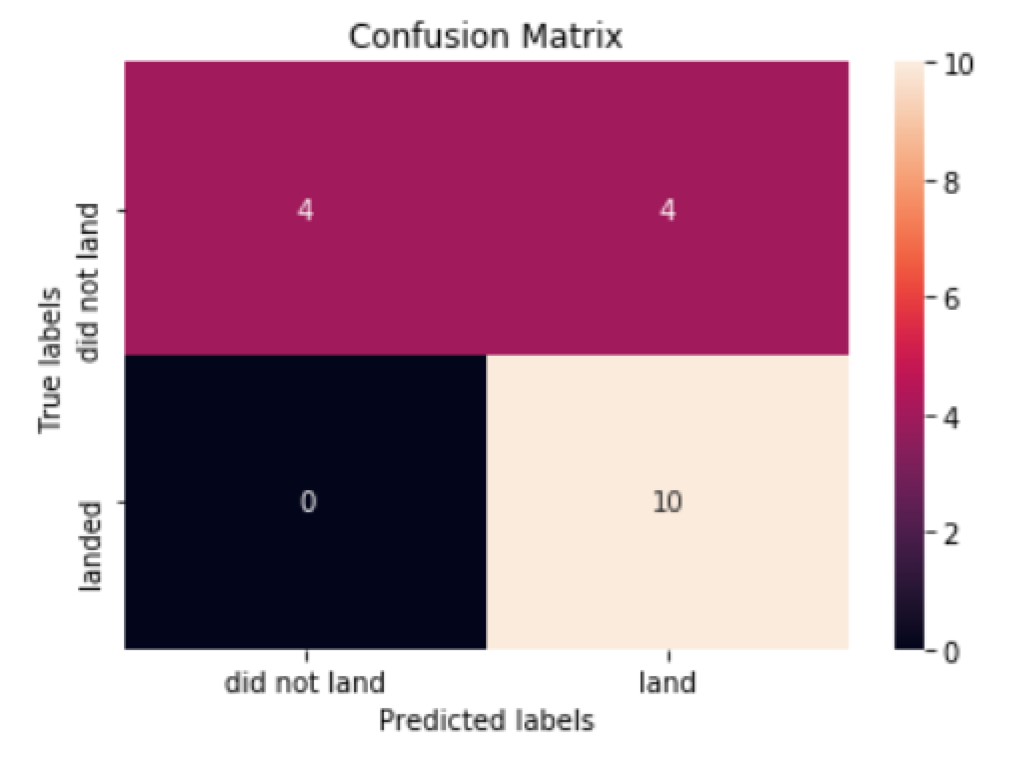
- Visualize the built model accuracy for all built classification models, in a bar chart
- Find which model has the highest classification accuracy

**The best model is the tree model**



# Confusion Matrix

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




**The best model is the tree model**

# Conclusions

- With heavy payloads the successful landing or positive landing rate are more for Polar, LEO and ISS.
- The success rate since 2013 kept increasing till 2020
- KSC LC-39A has a highest success rate of 77%
- Launch sites in close proximity to railways, highways, coastline, but keep certain distance away from cities.
- The best model is the tree model.

# Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

# Thank you!