IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Justin Fernandes
7/8/2024

Outline

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

Executive Summary

The Space Race of the past can be utilized to determine which factors play an important role in the success rate of launches. From the research, it was determined that the success rate increased over time, and varied based on different factors, like distance of travel, launch location, and booster versions. This was done utilizing data visualization, data analysis, and standardization.

Introduction

The main question that we intend to answer is:

What factors play the most role in the success rate of a launch and mission.

In doing so, we can nearly guarantee the success rate of a launch, providing a safer, cheaper, and easier pathway towards better space travel

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

Data was collected in two basic ways, web scraping and basic inputs. For basic inputs, we utilized User Defined Functions to take the data from the source and convert it into a data frame. With web scraping, we did the same thing but used Beautiful Soup to obtain the data.

Data Collection – SpaceX API

- [IBM-Data-Science-Final/jupyter-labs-spacex-data-collection-api.ipynb at main · Just1nFernandes/IBM-Data-Science-Final (github.com)]

Obtain Source -> Standardize Data -> Remove Clutter -> Input into Table -> Clean up non-values

Data Collection - Scraping

- [IBM-Data-Science-Final/jupyter-labs-webscraping.ipynb at main · Just1nFernandes/IBM-Data-Science-Final (github.com)]

Create Soup -> Pull Table -> Create Data Frame -> Insert Table into Data Frame

Data Wrangling

Data was wrangled by adding a “Class” column, which determined if the missions were successful or not based on the outcome. If the mission outcome was a “Bad Outcome”, it would be a failure and given a class of 0, if it was not a bad outcome, it would be given a value of 1

IBM-Data-Science-Final/labs-jupyter-spacex-Data wrangling.ipynb at main · Just1nFernandes/IBM-Data-Science-Final (github.com)

EDA with Data Visualization

Data was visualized with multiple different graphs, including bar charts, line charts, and scatter plots

IBM-Data-Science-Final/edadataviz.ipynb at main · Just1nFernandes/IBM-Data-Science-Final (github.com)

EDA with SQL

Using SQL we found

- The launch sites
- The records with CCA
- Total and average payload mass
- Earliest success date
- Booster that worked
- Successful and failed mission outcomes
- All the max boosters
- [IBM-Data-Science-Final/jupyter-labs-eda-sql-coursera_sqllite.ipynb at main · Just1nFernandes/IBM-Data-Science-Final (github.com)]

Build an Interactive Map with Folium

The interactive maps had markers with launch sites, successes and failures, and distances with launch sites

[IBM-Data-Science-Final/lab_jupyter_launch_site_location.ipynb at main · Just1nFernandes/IBM-Data-Science-Final (github.com)]

Build a Dashboard with Plotly Dash

- The dashboard had interactive charts that would enable users to visualize the data better

Predictive Analysis (Classification)

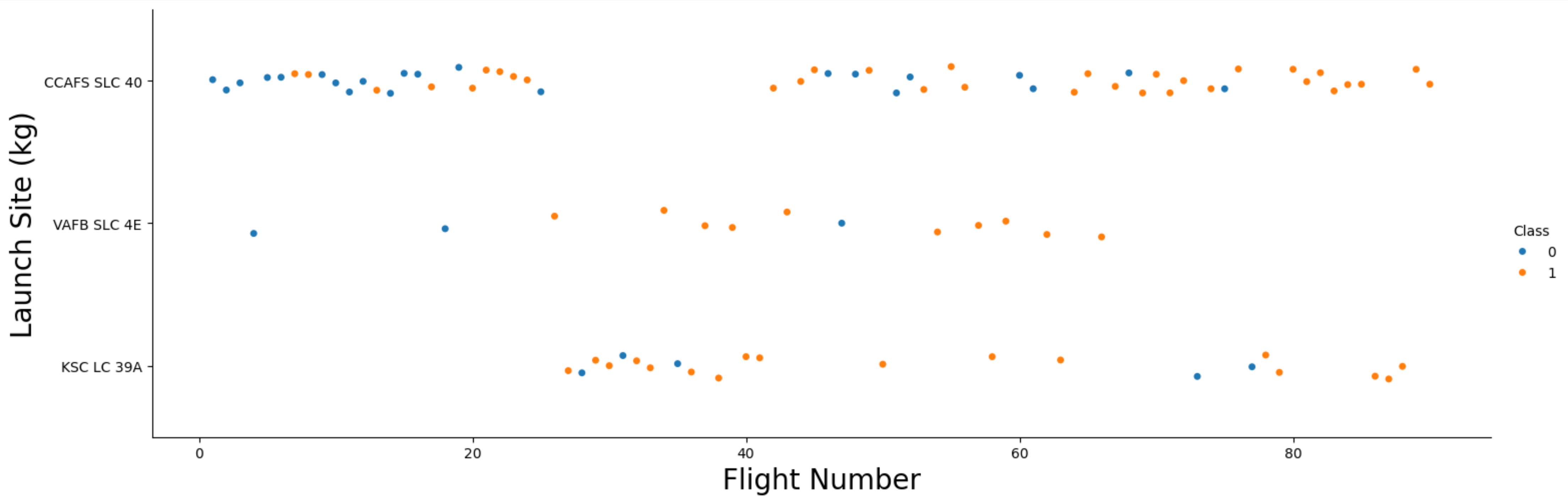
Using machine learning, we were able to standardize, train, and test it. In doing this ,we found the most accurate was K Nearest Neighbors and SVM, and the least accurate was decision tree.

[IBM-Data-Science-Final/SpaceX_Machine Learning Prediction_Part_5.ipynb at main · Just1nFernandes/IBM-Data-Science-Final (github.com)]

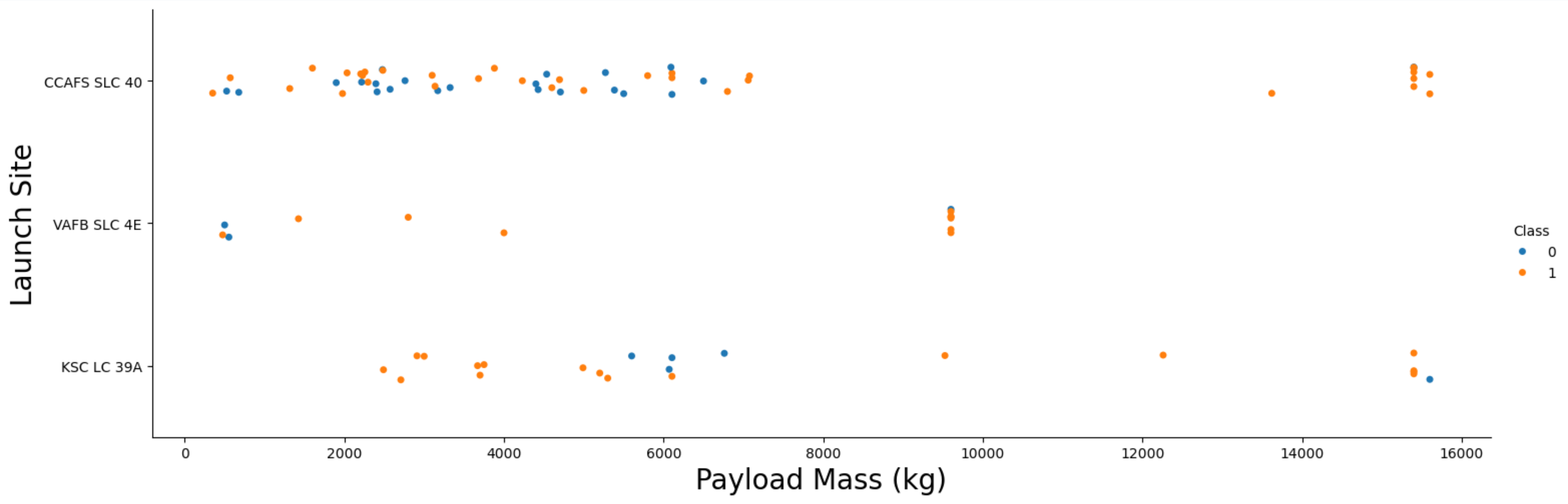
Section 2

Insights drawn from EDA

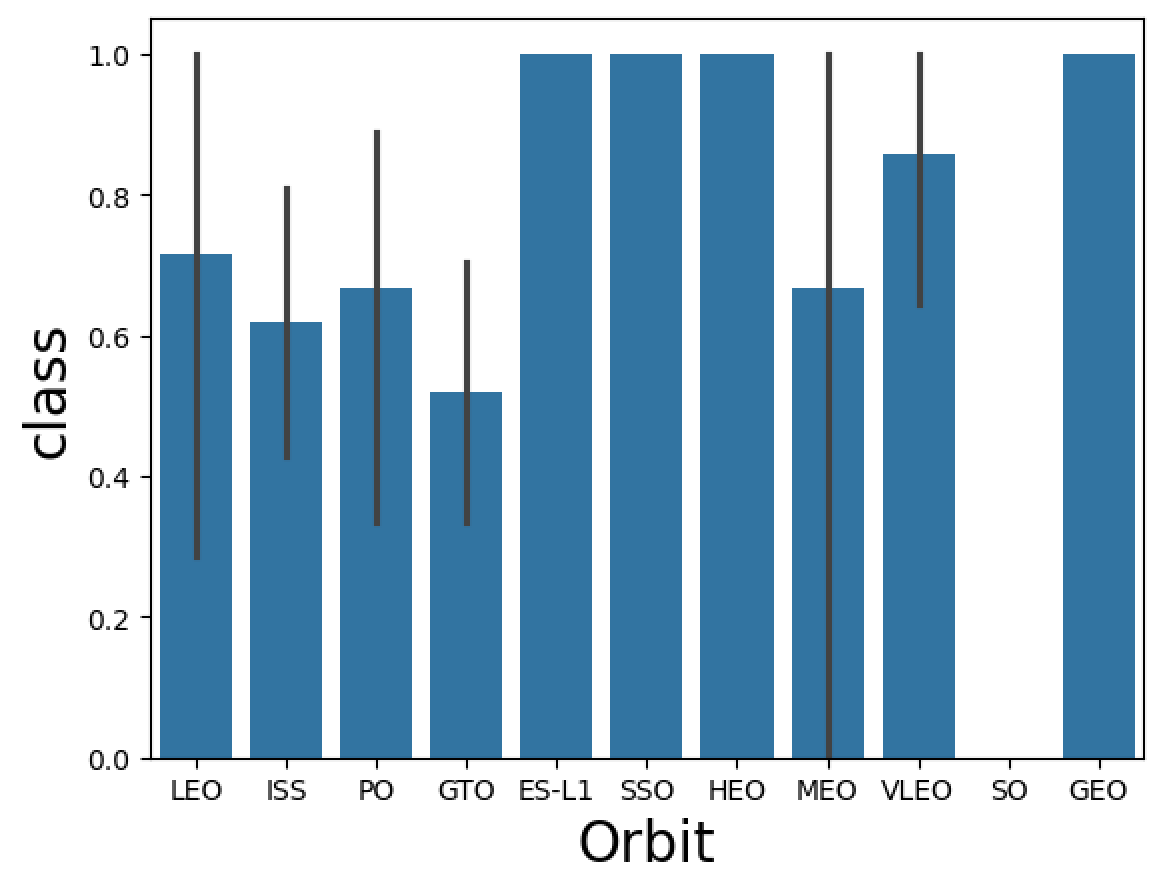
Flight Number vs. Launch Site



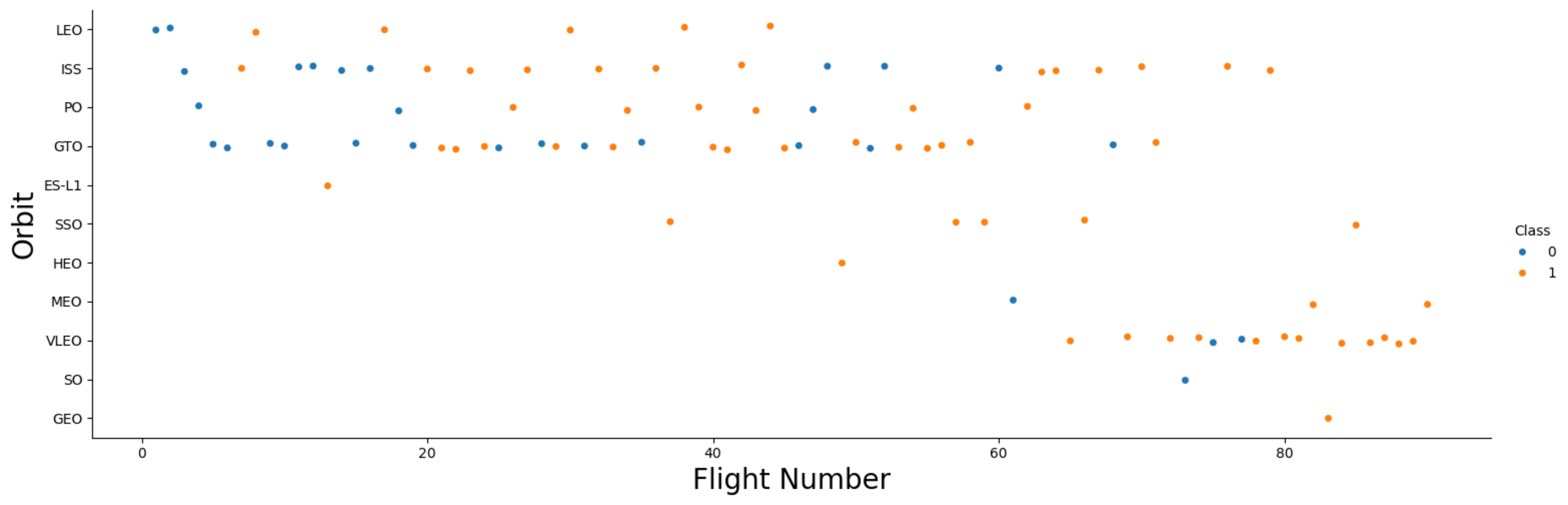
Payload vs. Launch Site



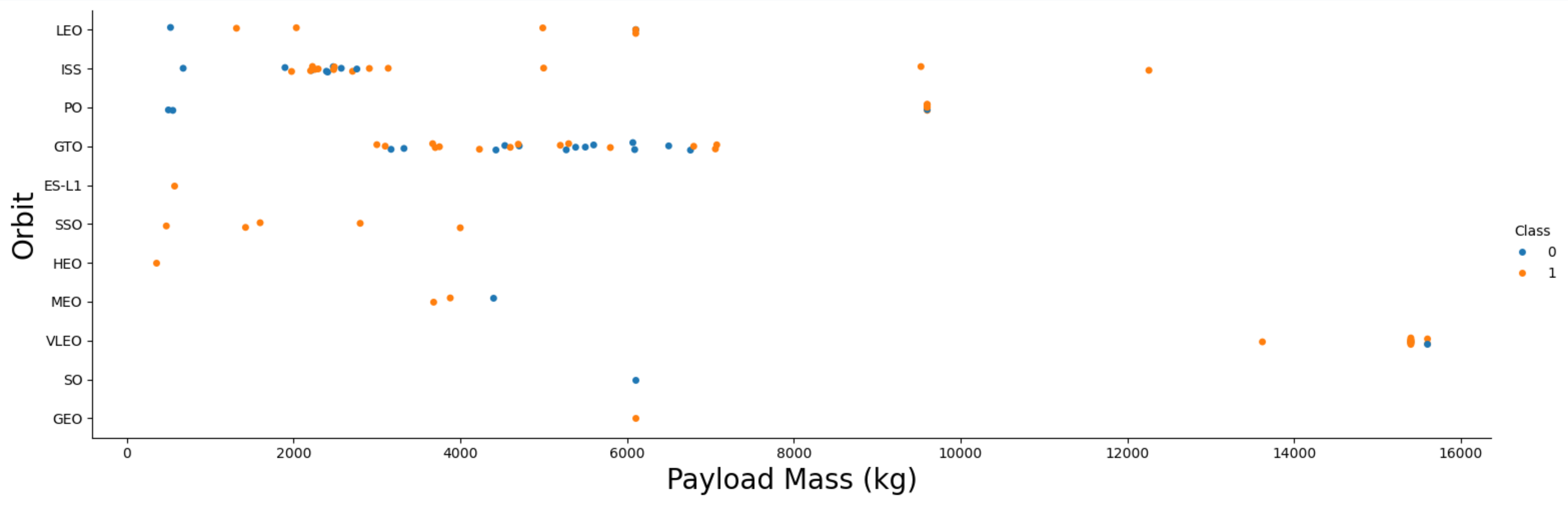
Success Rate vs. Orbit Type



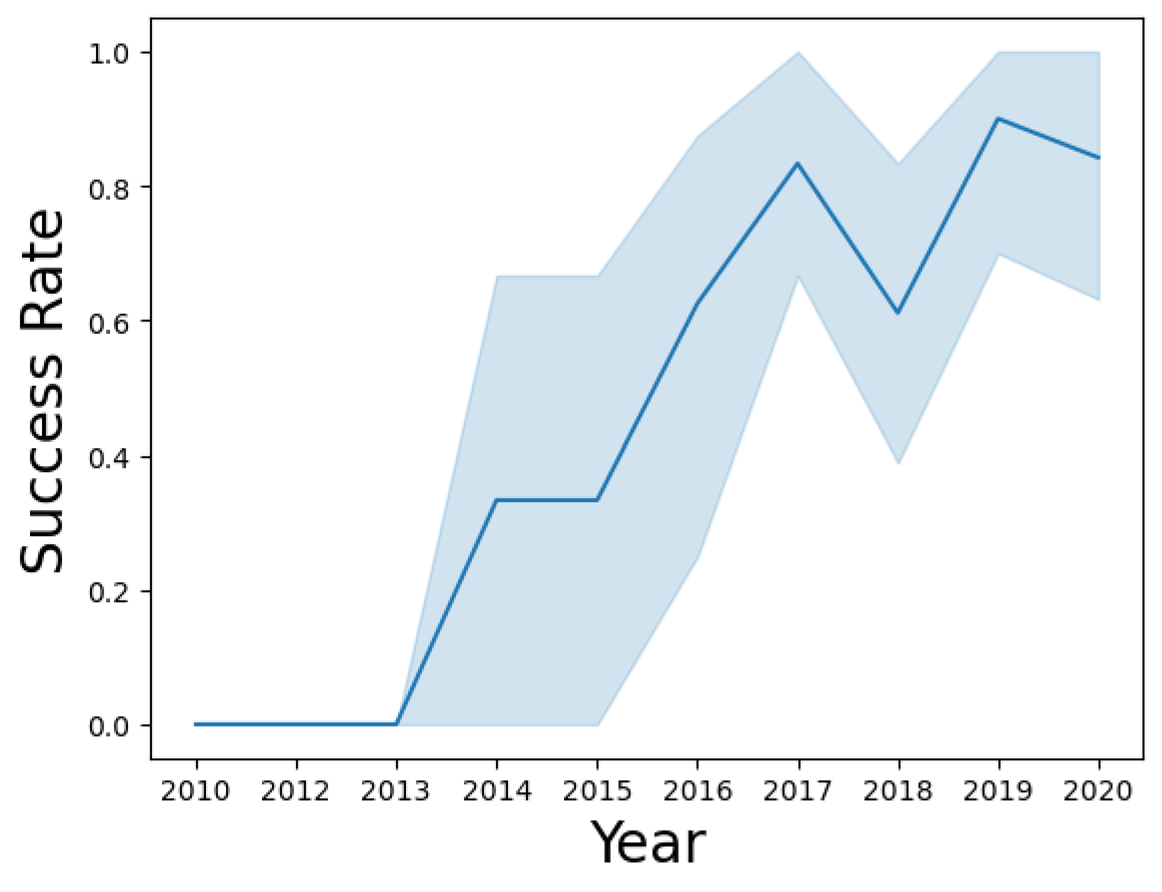
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

```
%sql select distinct launch_site from SPACEXTBL;
```

```
 * sqlite:///my_data1.db
Done.
```

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

Launch Site Names Begin with 'CCA'

```
%sql select * from SPACEXTBL where launch_site like '%CCA%' limit 5;
```

```
 * sqlite:///my_data1.db
Done.
```

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	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	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

```
%sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL;
```

```
 * sqlite:///my_data1.db
Done.
```

sum(PAYLOAD_MASS__KG_)
619967

Average Payload Mass by F9 v1.1

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

 * sqlite:///my_data1.db
Done.
```

avg(PAYLOAD_MASS__KG_)
2928.4

First Successful Ground Landing Date

```
%sql select min(Date) from SPACEXTBL where Landing_Outcome like "%Success%"
```

```
 * sqlite:///my_data1.db
Done.
```

min(Date)
2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

```
%sql select distinct Booster_Version from SPACEXTBL where Landing_Outcome = "Success (drone ship)" and PAYLOAD_MASS__KG_ > 4
```

```
 * sqlite:///my_data1.db
Done.
```

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

Total Number of Successful and Failure Mission Outcomes

```
print('Successful')
%sql select count(*) from SPACEXTBL where Mission_Outcome like "%Success%";
```

```
Successful
 * sqlite:///my_data1.db
Done.
```

count(*)
100

```
print('Failure')
%sql select count(*) from SPACEXTBL where Mission_Outcome like "%Failure%";
```

```
Failure
 * sqlite:///my_data1.db
Done.
```

count(*)
1

Boosters Carried Maximum Payload

```
%sql select distinct Booster_Version FROM SPACEXTBL WHERE(SELECT max(PAYLOAD_MASS__KG_) FROM SPACEXTBL);
```

```
 * sqlite:///my_data1.db
Done.
```

Booster_Version
F9 v1.0 B0003
F9 v1.0 B0004
F9 v1.0 B0005
F9 v1.0 B0006
F9 v1.0 B0007
F9 v1.1 B1003
F9 v1.1
F9 v1.1 B1011
F9 v1.1 B1010
F9 v1.1 B1012
F9 v1.1 B1013
F9 v1.1 B1014

Rest of Table was Cut off

2015 Launch Records

```
%sql select * from SPACEXTBL where Date between '2015-01-01' and '2015-12-31'
```

```
 * sqlite:///my_data1.db
Done.
```

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcom
2015-01-10	9:47:00	F9 v1.1 B1012	CCAFS LC-40	SpaceX CRS-5	2395	LEO (ISS)	NASA (CRS)	Success	Failure (drone shi
2015-02-11	23:03:00	F9 v1.1 B1013	CCAFS LC-40	DSCOVR	570	HEO	U.S. Air Force NASA NOAA	Success	Controlled (ocea
2015-03-02	3:50:00	F9 v1.1 B1014	CCAFS LC-40	ABS-3A Eutelsat 115 West B	4159	GTO	ABS Eutelsat	Success	No attem
2015-04-14	20:10:00	F9 v1.1 B1015	CCAFS LC-40	SpaceX CRS-6	1898	LEO (ISS)	NASA (CRS)	Success	Failure (drone shi
2015-04-27	23:03:00	F9 v1.1 B1016	CCAFS LC-40	Turkmen 52 / MonacoSAT	4707	GTO	Turkmenistan National Space Agency	Success	No attem
2015-06-28	14:21:00	F9 v1.1 B1018	CCAFS LC-40	SpaceX CRS-7	1952	LEO (ISS)	NASA (CRS)	Failure (in flight)	Precluded (dror shi
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (grour pa

Section 3

Launch Sites Proximities Analysis

Launch Locations

<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

<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

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

<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

<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.

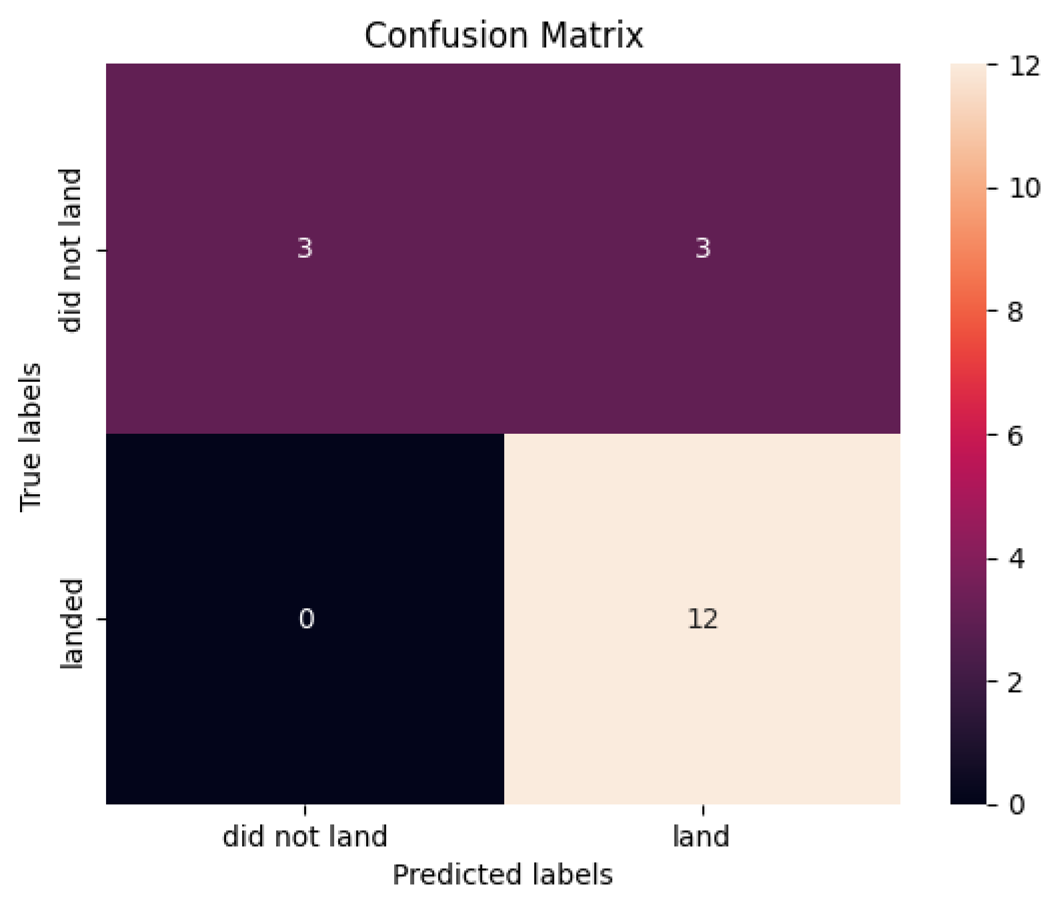
Section 5

Predictive Analysis (Classification)

Classification Accuracy

	Accuracy	Test Accuracy
Logistic regression	84.64	83.33
Support vector machine	84.82	83.33
Decision tree classifier	87.4	55.5
K nearest neighbors	84.82	83.33

Confusion Matrix



Appendix

- [Just1nFernandes/IBM-Data-Science-Final (github.com)]

Thank you!