IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Pierre Levy
3 December 2023

Outline

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

Executive Summary

We used Jupyter Notebooks and Pything to analyze data regarding SpaceX's rocket launchers.

Introduction

The goal is to find the probability of a satellite launch to be successful.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:

  Data was collected on SpaceX's website and Wikipedia.

- Perform data wrangling

  We processed data using various Python libraries.

- 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

- For data on SpaceX's website, we used API calls.
- For data on Wikipedia, we used web scraping techniques.

Data Collection – SpaceX API

Collect data from SpaceX Website => Import data on Jupyter Notebooks => Treat data with Python => Convert data to charts for easy understanding.

All the files can be found here:

https://github.com/sowon27/coursera-data-science-project

Data Collection - Scraping

Extract Data from the Wikipedia page => Import data on Jupyter Notebooks => Treat data with Python => Convert data to charts for easy understanding.

All the files can be found here:

https://github.com/sowon27/coursera-data-science-project

Data Wrangling

We removed empty data and normalized data. We also converted categorical data to numerical data.

https://github.com/sowon27/coursera-data-science-project

EDA with Data Visualization

We plotted bar charts, line charts, and pie charts.

https://github.com/sowon27/coursera-data-science-project

EDA with SQL

We used data query to list distinct features as well as calculates some sums and averages.

https://github.com/sowon27/coursera-data-science-project

Build an Interactive Map with Folium

We made of map of the United States of America with the various launch sites.

Build a Dashboard with Plotly Dash

We built a user-friendly dashboard with Plotly Dash.

https://github.com/sowon27/coursera-data-science-project

Predictive Analysis (Classification)

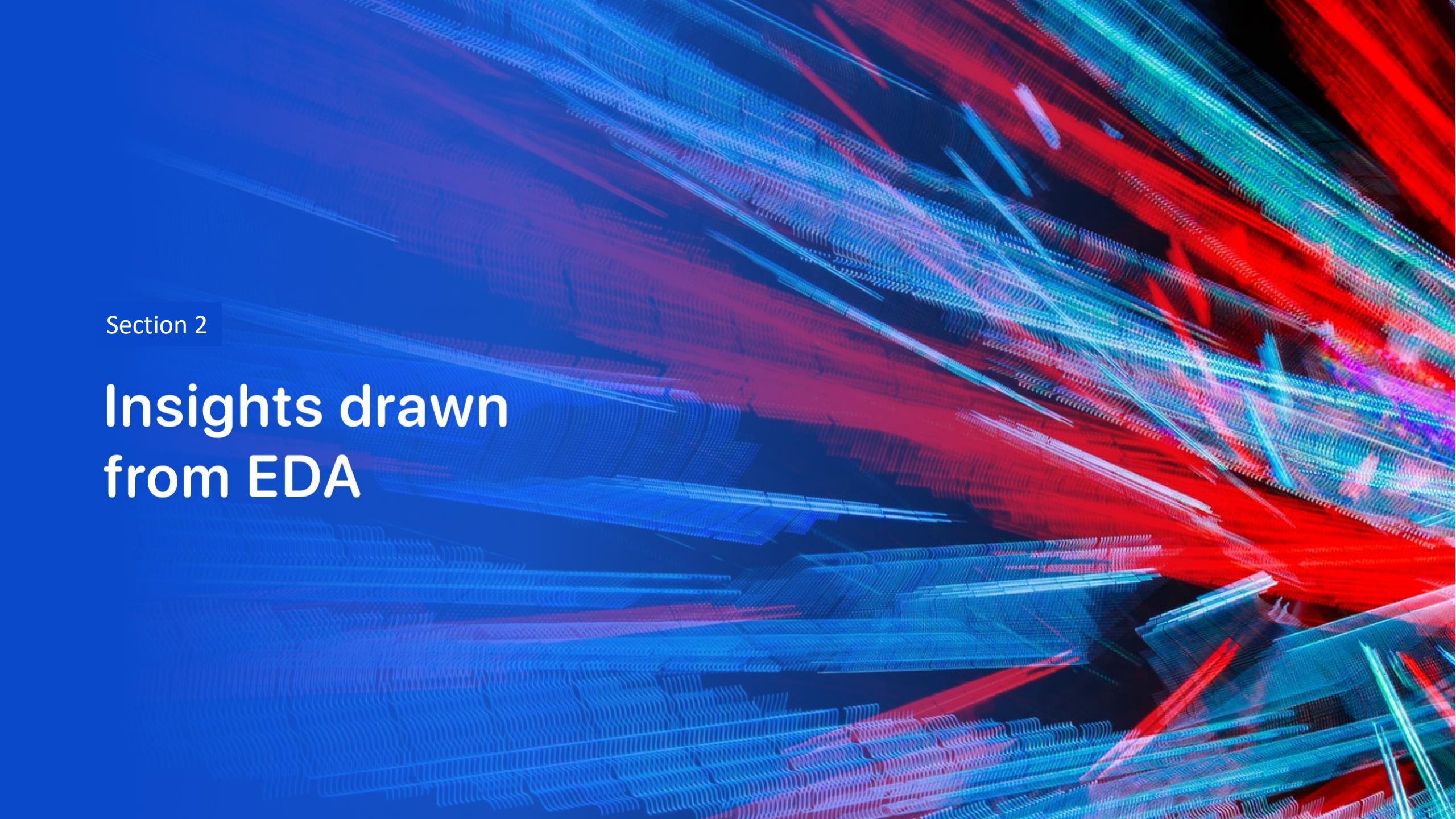
We used logistic regression, decision trees, as well as the K-nearest neighbor method to make predictions.

https://github.com/sowon27/coursera-data-science-project

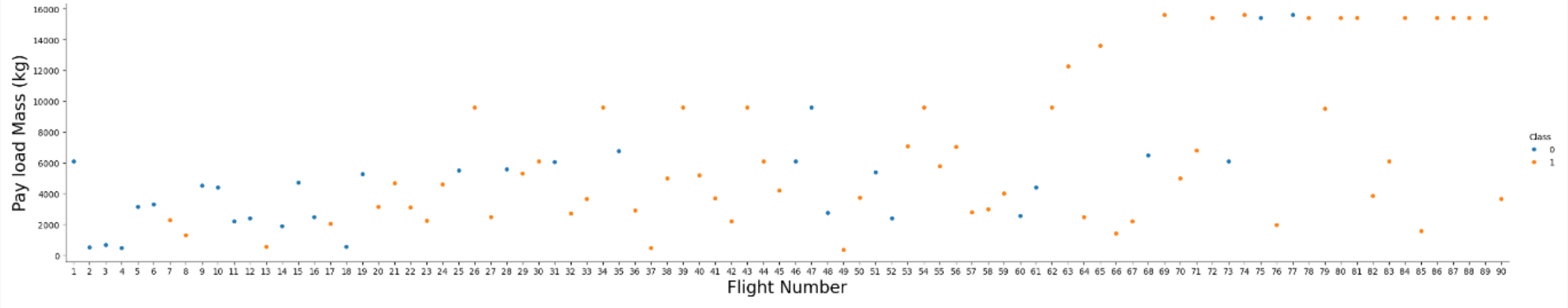
Results

We managed to predict the probability of a launch being successful.

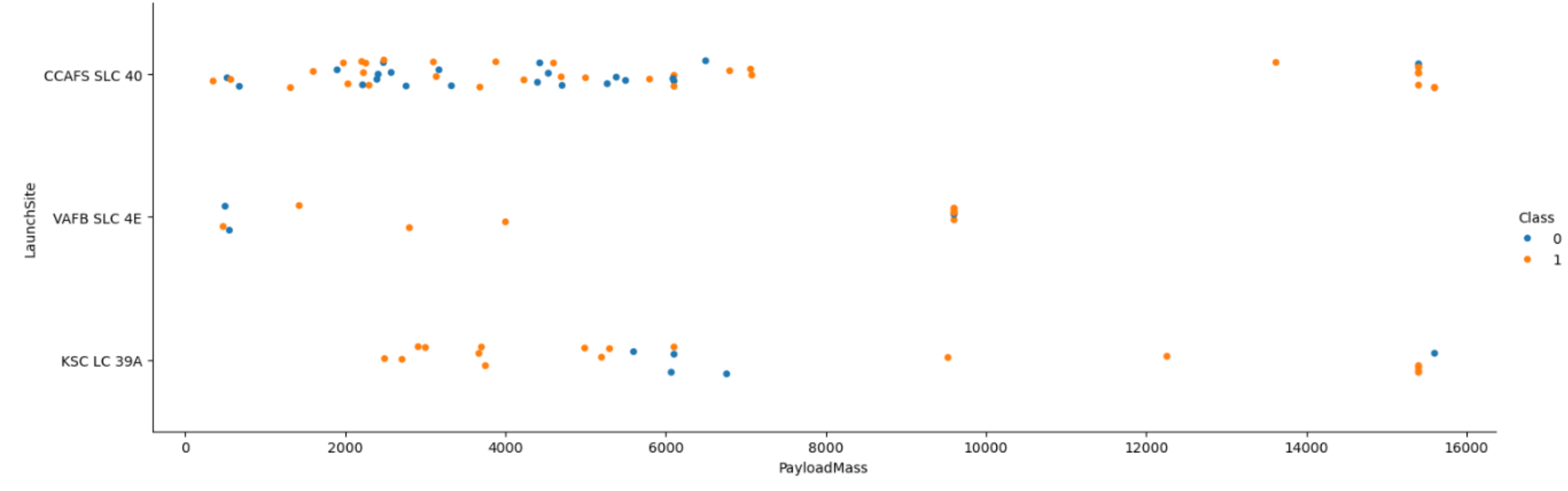
https://github.com/sowon27/coursera-data-science-project


Section 2
Insights drawn from EDA

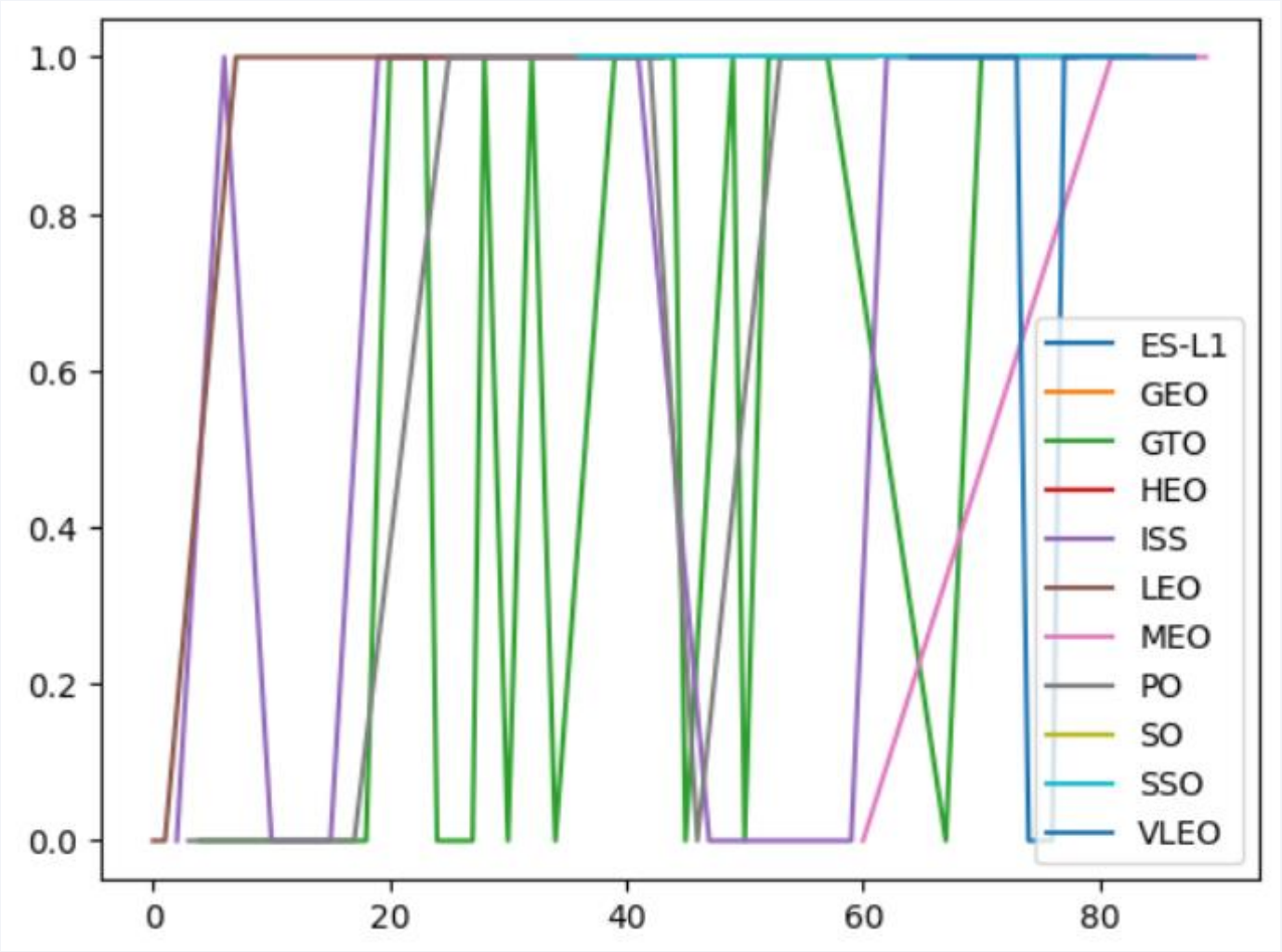
Flight Number vs. Launch Site



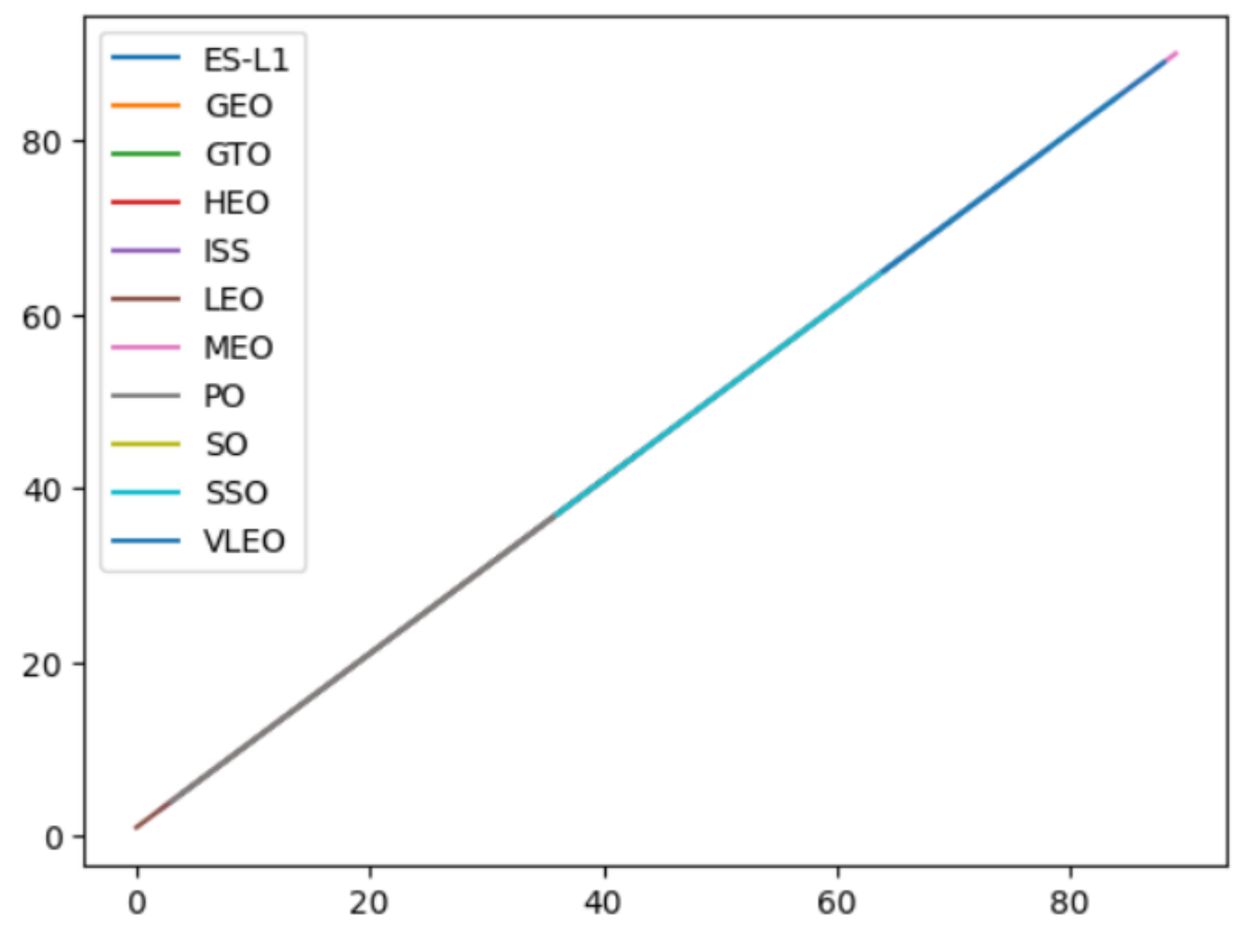
Payload vs. Launch Site



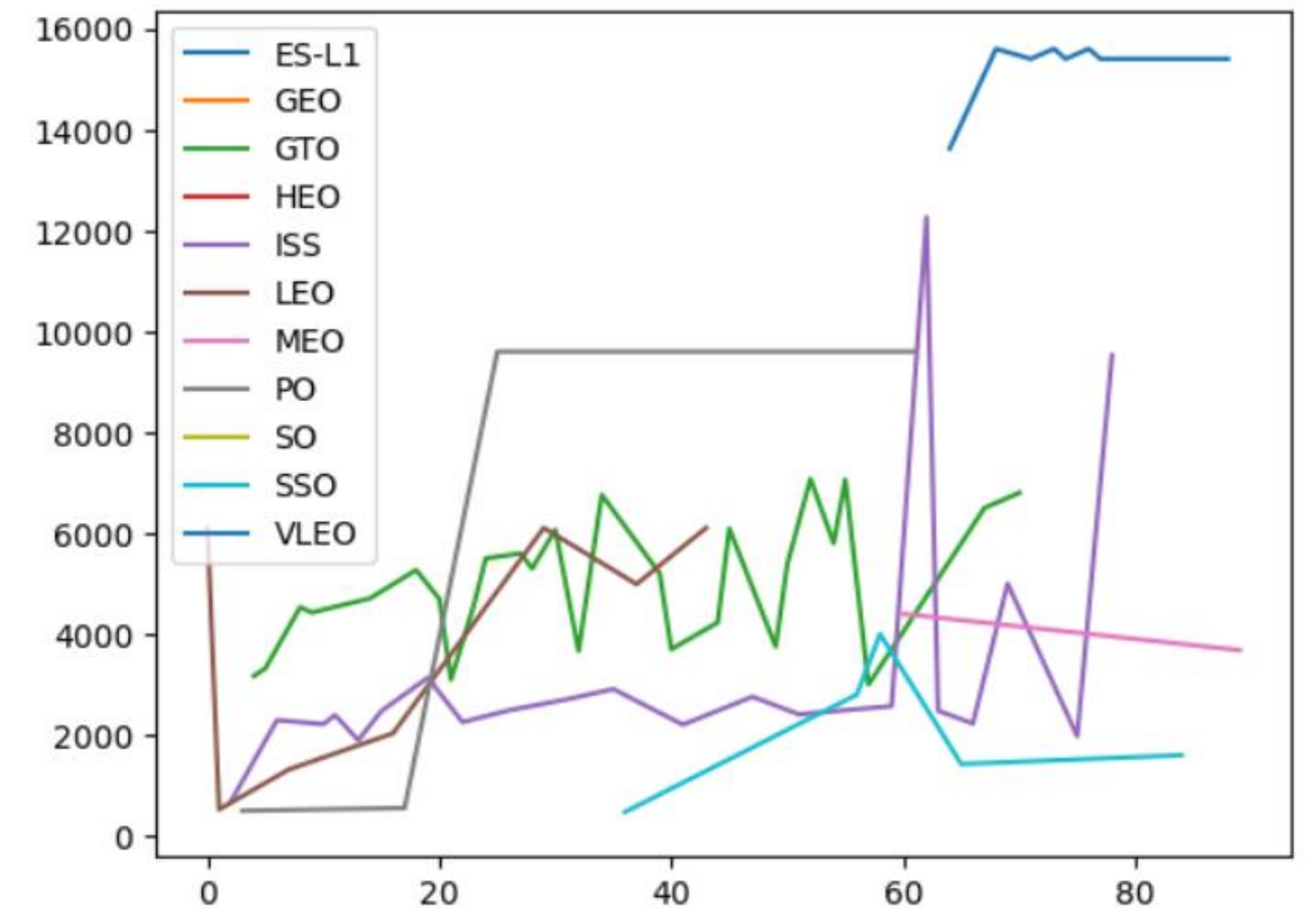
Success Rate vs. Orbit Type



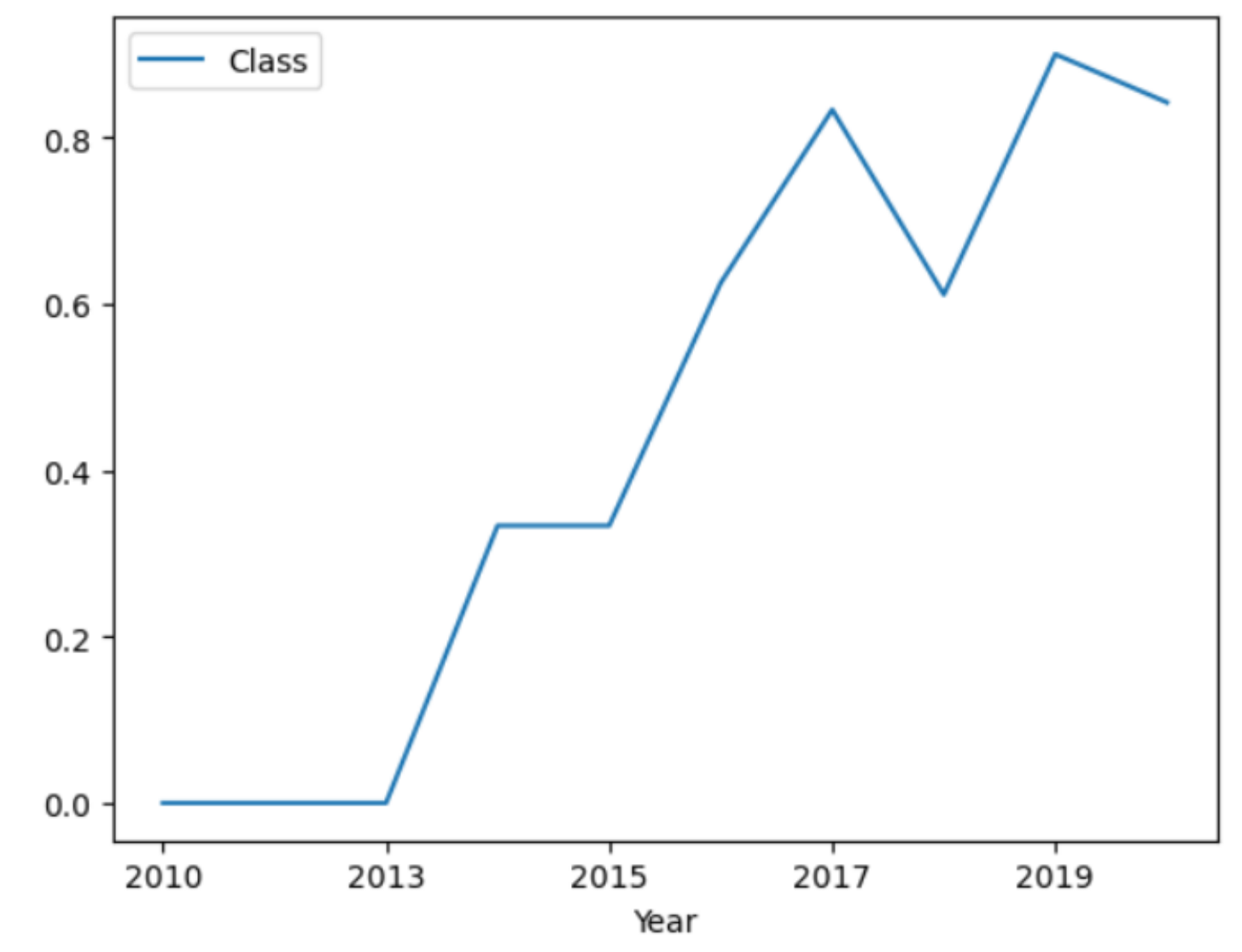
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

Here are all the launch site names:

```
[9]: %sql select distinct Launch_Site from SPACEXTBL
 * sqlite:///my_data1.db
Done.
```

[9]:

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

Launch Site Names Begin with 'CCA'

```
[10]: %sql select * from SPACEXTBL where Launch_Site like '%CCA%'limit 5
```

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

[10]:

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

```
[11]: %sql select SUM(PAYLOAD_MASS__KG_) from SPACEXTBL where Customer like '%NASA%';
```

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

[11]:

SUM(PAYLOAD_MASS__KG_)
107010

Average Payload Mass by F9 v1.1

Display average payload mass carried by booster version F9 v1.1

```
[20]: %sql select AVG(payload_mass__kg_) from SPACEXTBL where Booster_Version like '%F9 v1.1%';
```

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

[20]:

AVG(payload_mass__kg_)
2534.6666666666665

First Successful Ground Landing Date

The first successful landing was achieved on 22 December 2015.

```
[24]: %sql select * from SPACEXTBL where Landing_Outcome like '%success%' limit 1;
 * sqlite:///my_data1.db
Done.
```

[24]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2015-12-22	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

Successful Drone Ship Landing with Payload between 4000 and 6000

[31]:
```
%sql select * from SPACEXTBL where Landing_Outcome like '%success%drone%' and (PAYLOAD_MASS__KG_ between 4000 and 6000);
```

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

[31]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2016-05-06	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-08-14	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2017-03-30	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
2017-10-11	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoStar 105	5200	GTO	SES EchoStar	Success	Success (drone ship)

Total Number of Successful and Failure Mission Outcomes

```
[35]: %sql select count(Landing_Outcome) from SPACEXTBL where Landing_Outcome like '%success%';
 * sqlite:///my_data1.db
Done.
[35]: count(Landing_Outcome)
                          61
```

```
[36]: %sql select count(Landing_Outcome) from SPACEXTBL where Landing_Outcome like '%failure%';
 * sqlite:///my_data1.db
Done.
[36]: count(Landing_Outcome)
                          10
```

There were 61 successes and 10 failures. Controlled and uncontrolled landings are not included in these statistics.

Boosters Carried Maximum Payload

[39]:
```
%sql select * from SPACEXTBL where PAYLOAD_MASS__KG_ = (select max(PAYLOAD_MASS__KG_) from SPACEXTBL);
```

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

[39]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2019-11-11	14:56:00	F9 B5 B1048.4	CCAFS SLC-40	Starlink 1 v1.0, SpaceX CRS-19	15600	LEO	SpaceX	Success	Success
2020-01-07	2:33:00	F9 B5 B1049.4	CCAFS SLC-40	Starlink 2 v1.0, Crew Dragon in-flight abort test	15600	LEO	SpaceX	Success	Success
2020-01-29	14:07:00	F9 B5 B1051.3	CCAFS SLC-40	Starlink 3 v1.0, Starlink 4 v1.0	15600	LEO	SpaceX	Success	Success
2020-02-17	15:05:00	F9 B5 B1056.4	CCAFS SLC-40	Starlink 4 v1.0, SpaceX CRS-20	15600	LEO	SpaceX	Success	Failure
2020-03-18	12:16:00	F9 B5 B1048.5	KSC LC-39A	Starlink 5 v1.0, Starlink 6 v1.0	15600	LEO	SpaceX	Success	Failure
2020-04-22	19:30:00	F9 B5 B1051.4	KSC LC-39A	Starlink 6 v1.0, Crew Dragon Demo-2	15600	LEO	SpaceX	Success	Success
2020-06-04	1:25:00	F9 B5 B1049.5	CCAFS SLC-40	Starlink 7 v1.0, Starlink 8 v1.0	15600	LEO	SpaceX, Planet Labs	Success	Success
2020-09-03	12:46:14	F9 B5 B1060.2	KSC LC-39A	Starlink 11 v1.0, Starlink 12 v1.0	15600	LEO	SpaceX	Success	Success
2020-10-06	11:29:34	F9 B5 B1058.3	KSC LC-39A	Starlink 12 v1.0, Starlink 13 v1.0	15600	LEO	SpaceX	Success	Success
2020-10-18	12:25:57	F9 B5 B1051.6	KSC LC-39A	Starlink 13 v1.0, Starlink 14 v1.0	15600	LEO	SpaceX	Success	Success
2020-10-24	15:31:34	F9 B5 B1060.3	CCAFS SLC-40	Starlink 14 v1.0, GPS III-04	15600	LEO	SpaceX	Success	Success
2020-11-25	2:13:00	F9 B5 B1049.7	CCAFS SLC-40	Starlink 15 v1.0, SpaceX CRS-21	15600	LEO	SpaceX	Success	Success

2015 Launch Records

[49]:
```
%sql select substr(Date,6,2) as Month, Landing_Outcome, Booster_Version, Launch_Site from SPACEXTBL where Landing_Outcome like '%failure%drone%' and substr(Date,0,5)='2015';
```

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

[49]:

Month	Landing_Outcome	Booster_Version	Launch_Site
01	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

[76]:
```
%%sql select Landing_Outcome, count(Landing_Outcome) as countlo from SPACEXTBL
    where (Date > '2010-06-04') and (Date < '2017-03-20')
    group by Landing_Outcome order by countlo desc;
```

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

[76]:

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

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

<Folium Map Screenshot 2>

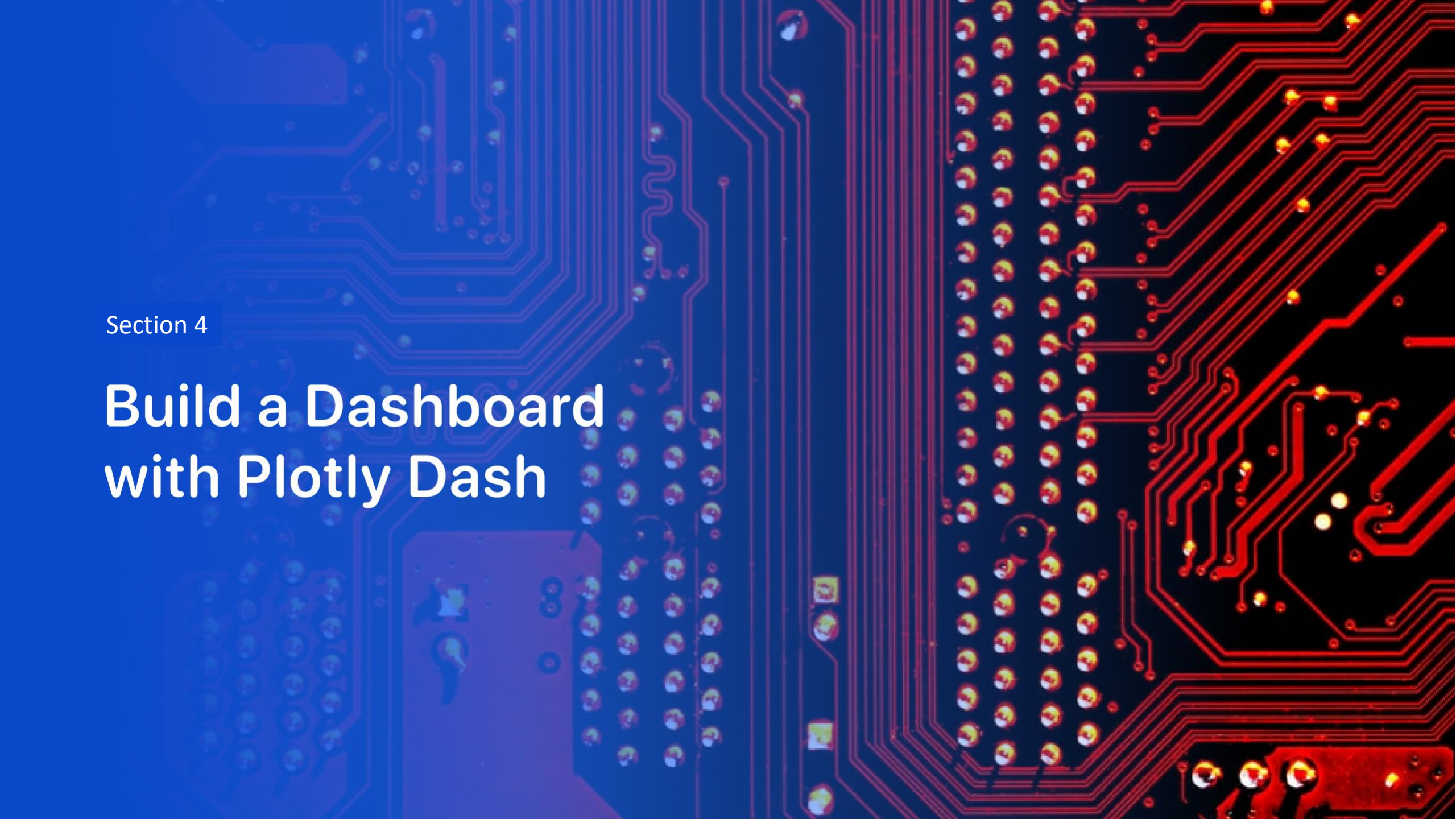
- Replace <Folium map screenshot 2> title with an appropriate title
- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map
- Explain the important elements and findings on the screenshot

<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

- Visualize the built model accuracy for all built classification models, in a bar chart

- Find which model has the highest classification accuracy

Confusion Matrix

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

Conclusions

- Point 1
- Point 2
- Point 3
- Point 4
- …

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!