Winning Space Race with Data Science

Mikhail Voloshin
September 16, 2021

Outline

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

Just adding a cute little picture to apply my creativity to improve the presentation beyond the template.

Executive Summary

- Summary of methodologies
  - Obtain data using
    - Web scraping via BeautifulSoup
    - REST API
    - Loading tables into DB2
  - Use exploratory analysis to get a “feel” for the data
    - SQL
    - Pandas with Matplotlib
  - Use visual tools to perform analytics
    - Visualize launch site locations with Folium
    - Build Interactive Visual Analytics Dashboard with Plotly Dash
  - Find patterns using Machine Learning algorithms with scikit-learn
    - Logistic regression
    - Support vector machine
    - Decision tree
    - K Nearest Neighbor

- Summary of all results
  - Heavy rockets are more likely to crash
  - Rockets with high-orbit payloads are more likely to crash
  - Coastal launch sites have better recovery

Just adding a cute little picture to apply my creativity to improve the presentation beyond the template.

Introduction

- Project background and context
  - We will predict if the Falcon 9 first stage will land successfully.
  - SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because 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
  - What types of rockets land most successfully?
  - What launch sites result in more successful landings?
  - What payloads result in more successful landings?
  - What orbital targets result in more successful landings?

Just adding a cute little picture to apply my creativity to improve the presentation beyond the template.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Data was loaded via HTTP from SpaceX's REST API
  - Data was scraped from Wikipedia using BeautifulSoup
  - Data was loaded into a DB2 table from a CSV file
- Perform data wrangling
  - We determined which outcomes we consider "good" and "bad", and assigned a class for each launch

Methodology

Executive Summary

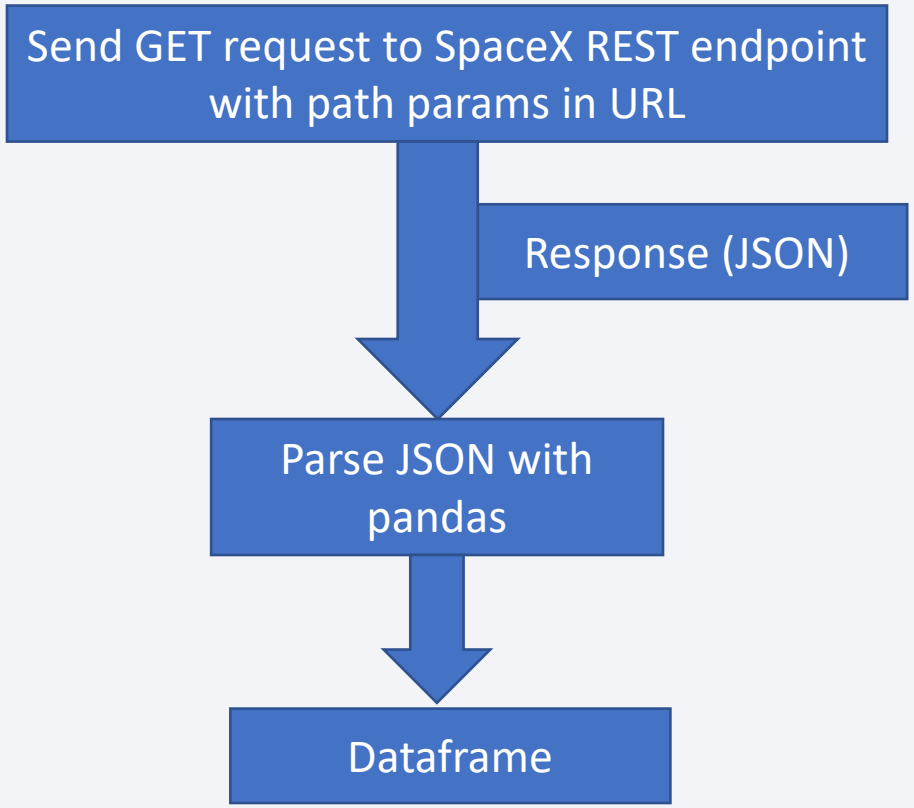
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - Using Machine Learning algorithms with scikit-learn
    - Logistic regression
    - Support vector machine
    - Decision tree
    - K Nearest Neighbor

Data Collection

- Describe how data sets were collected.
  - Data was loaded via HTTP from SpaceX's REST API
  - Data was scraped from Wikipedia using BeautifulSoup
  - Data was loaded into a DB2 table from a CSV file
- You need to present your data collection process use key phrases and flowcharts
  - REST API, BeautifulSoup, DB2 import from CSV

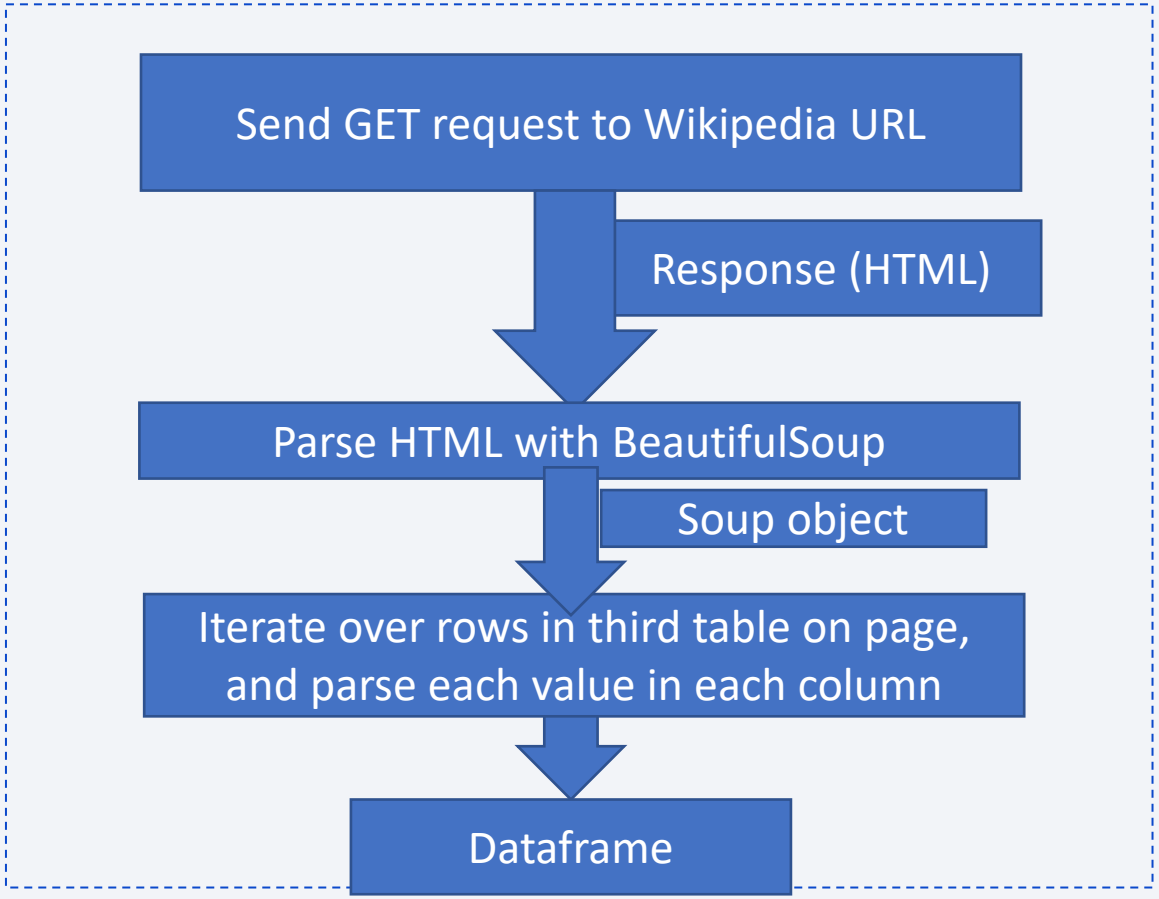
Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts
  - Used Python requests library
  - Sent GET request to SpaceX REST API endpoint with proper path params and queryparams in the URL
  - Received JSON response
  - Used pandas to parse JSON into a dataframe
- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose
  - https://github.com/omedalus/SpaceX-Capstone/blob/master/Week%201-a:%20Hands%20On%20Lab:%20Data%20Collection%20API.ipynb



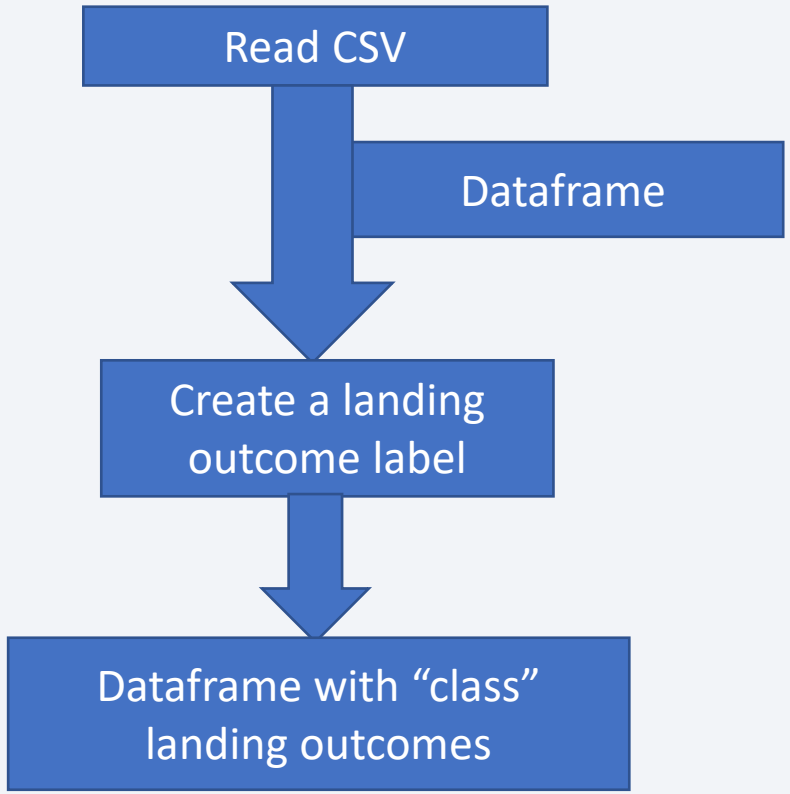
Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts
  - Send GET request to Wikipedia URL
  - Receive response as HTML
  - Parse HTML with BeautifulSoup
  - Find all tables in the HTML
  - In the list of all tables, the first launch table will be the third one
  - For each row in the first launch table, for each column, parse the value. We hard-coded how each column should be parsed, i.e. whether it should be a number, date, string, etc.
  - Store the results in a dataframe
- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose
  - https://github.com/omedalus/SpaceX-Capstone/blob/master/Week%201-b:%20Data%20Collection%20with%20Web%20Scraping.ipynb



Data Wrangling

- Describe how data were processed
  - We determined which outcomes we consider “good” and “bad”
  - We assigned a class for each launch
- You need to present your data wrangling process using key phrases and flowcharts
  - Numpy and pandas
  - Landing outcome labels
  - value_counts
- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose
  - https://github.com/omedalus/SpaceX-Capstone/blob/master/Week%201-c:%20Hands-On%20Lab:%20Data%20Wrangling.ipynb



EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
  - Charts plotted:
    - Histogram of Payload Mass vs. Flight Number
    - Histogram of Launch Site vs. Flight Number
    - Histogram of Launch Site vs. Payload Mass
    - Bar chart of successful landing (class) vs. Orbit
    - Orbit vs. Flight Number
    - Orbit vs. Payload Mass
    - Line Chart of Class vs. Year
  - Reasons why I used these charts
    - Because they are what the assignment came pre-populated with
    - Because they may reveal causative relationships between controllable factors (e.g. launch site, payload mass, etc.) and non-controllable ones (e.g. landing success)
- Add the GitHub URL of your completed EDA with data visualization notebook, as an external reference and peer-review purpose
  - https://github.com/omedalus/SpaceX-Capstone/blob/master/Week%202-b:%20EDA%20with%20Visualization.ipynb

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed
  - Distinct launch sites
  - Top 5 records whose launch site started with “CCA”
  - Total payload mass launched by “NASA (CRS)”
  - Average payload by boosters version “F9 v1.1”
  - Date of first successful landing
  - Boosters which have succeeded in a drone landing with a payload between 4k and 6k
  - Total number of successes and failures
  - Booster versions which have carried the max payload
  - Failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
  - 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¶

- Add the GitHub URL of your completed EDA with SQL notebook, as an external reference and peer-review purpose
  - https://github.com/omedalus/SpaceX-Capstone/blob/master/Week%202-a:%20EDA%20with%20SQL.ipynb

Build an Interactive Map with Folium

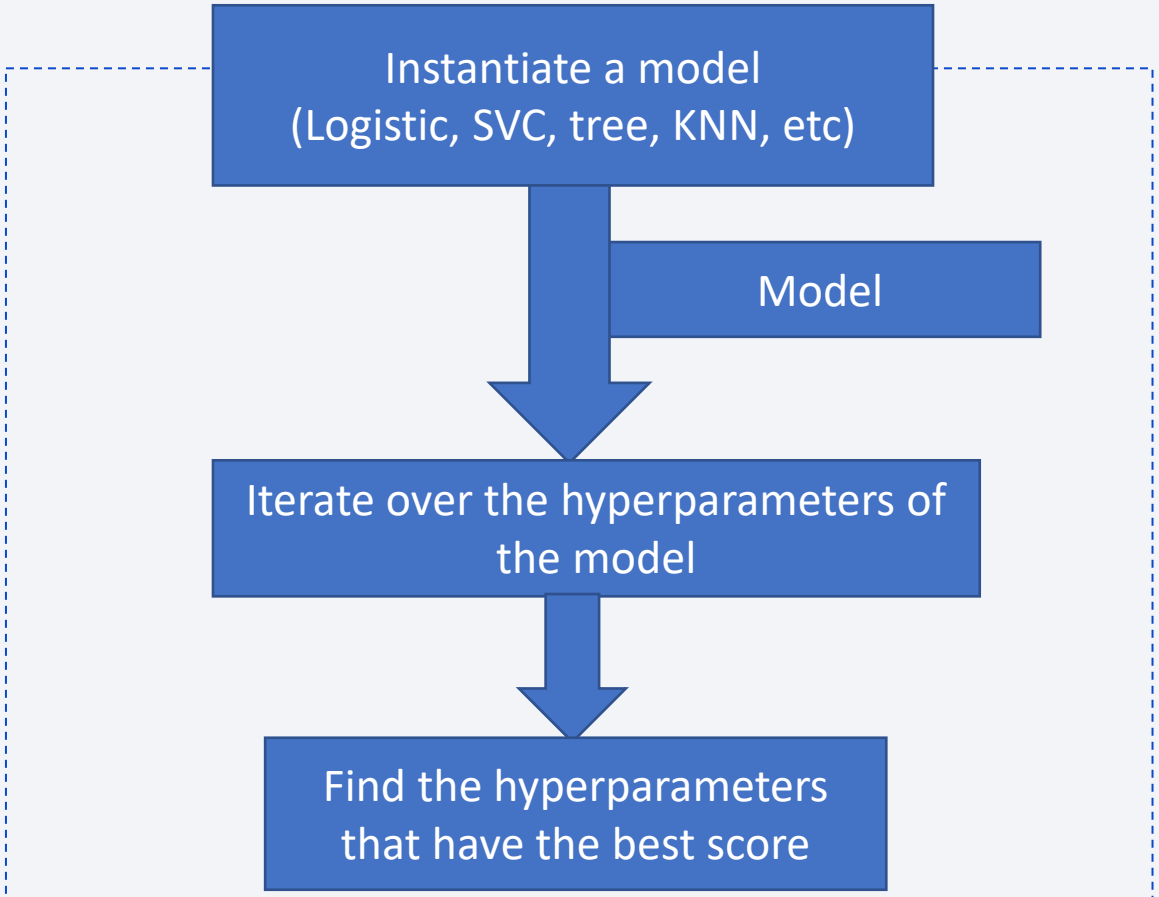
- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
  - Circles to mark launch sites
  - Markers to provide labels for launch sites
  - Lines from launch site to nearby “feature”, e.g. railway or highway or coastline
- Explain why you added those objects
  - I added them because the assignment told me to
- Add the GitHub URL of your completed interactive map with Folium map, as an external reference and peer-review purpose
  - https://github.com/omedalus/SpaceX-Capstone/blob/master/Week%203:%20Interactive%20Visual%20Analytics%20with%20Folium.ipynb

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
  - Dropdown to select launch site
  - Double-valued slider to select payload range
  - Pie chart to show the total successful launches count for all sites, or success vs. fail counts for one site if only one site is selected
  - Scatter chart to show the correlation between payload and launch success
- Explain why you added those plots and interactions
  - Because that's the requirement for the assignment?
- Add the GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose
  - https://github.com/omedalus/SpaceX-Capstone/blob/main/spacex_dash_app.py

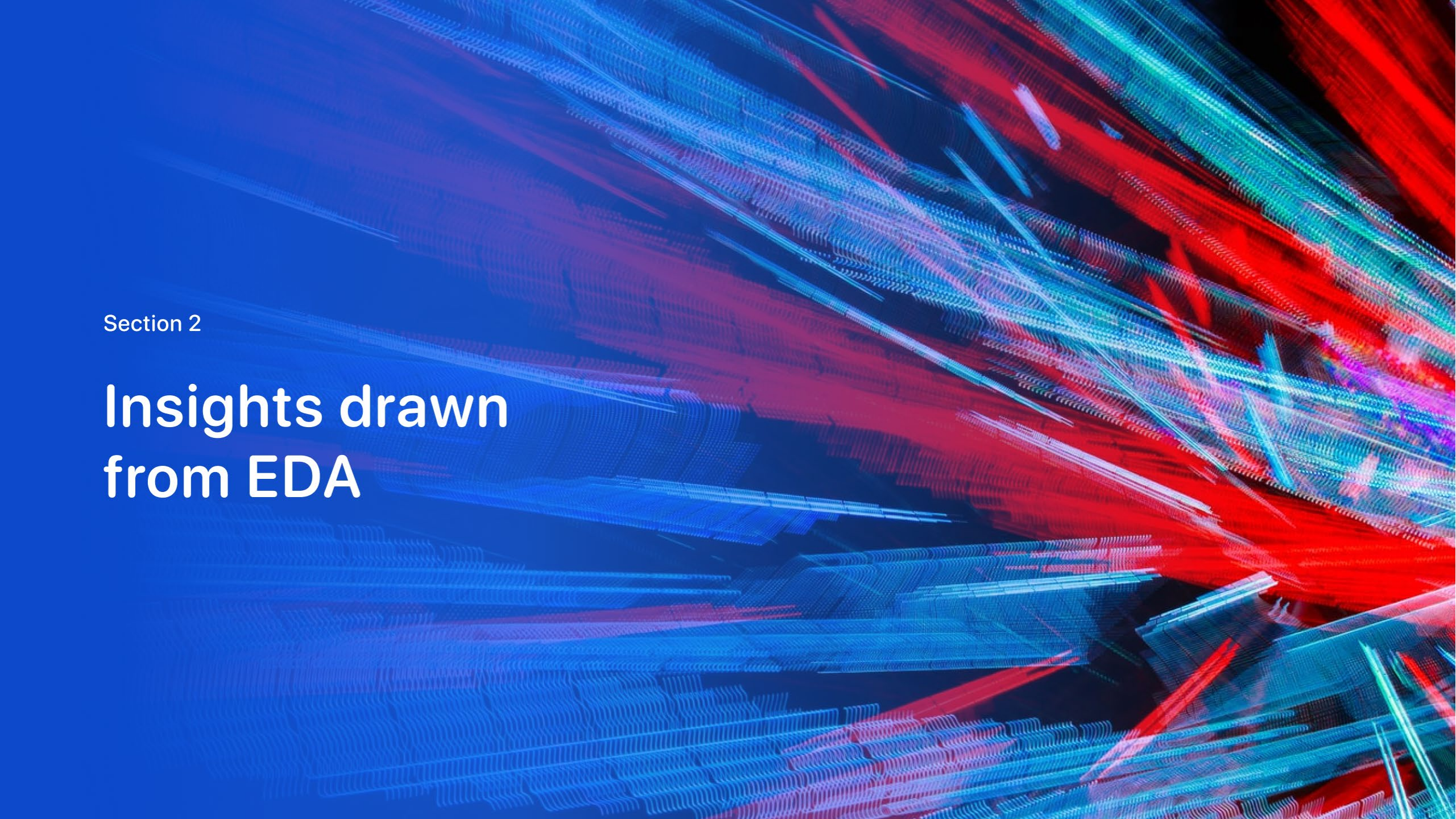
Predictive Analysis (Classification)

- Summarize how you built, evaluated, improved, and found the best performing classification model
  - Ran a GridSearchCV on…
    - Logistic Regression Classifier
    - Support Vector Classifier
    - Decision Tree Classifier
    - K Nearest Neighbors Classifier
- You need present your model development process using key phrases and flowchart
  - Training Set
  - Testing Set
- Add the GitHub URL of your completed predictive analysis lab, as an external reference and peer-review purpose
  - https://github.com/omedalus/SpaceX-Capstone/blob/master/Week%204:%20Machine%20Learning%20Prediction.ipynb



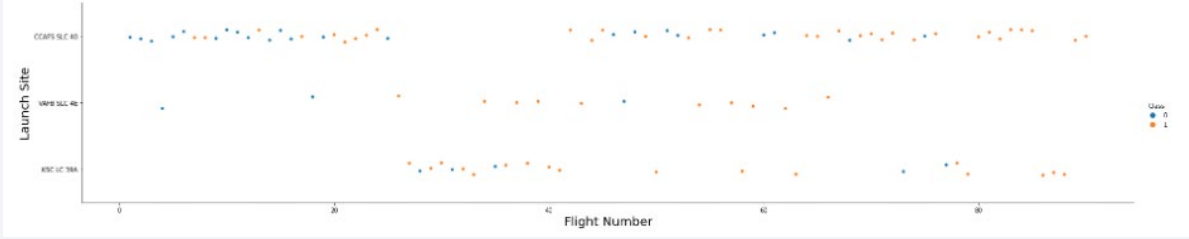
Results

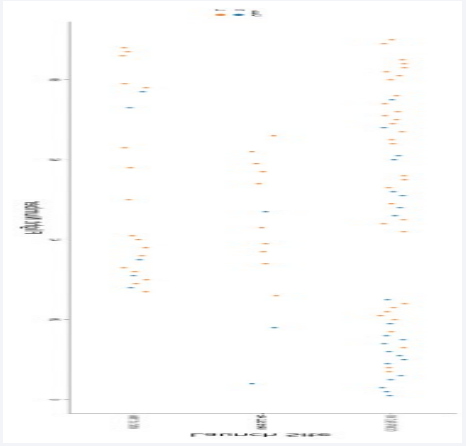
- Exploratory data analysis results
  - All models have a score of 0.83333333333…
- Interactive analytics demo in screenshots
  - Would have been nice to have been told during the assignment instructions that we should take screenshots and save them for this final project
- Predictive analysis results
  - Heavy rockets that travel farther have a lower chance of landing successfully. Duh.


Section 2
Insights drawn
from EDA

Flight Number vs. Launch Site

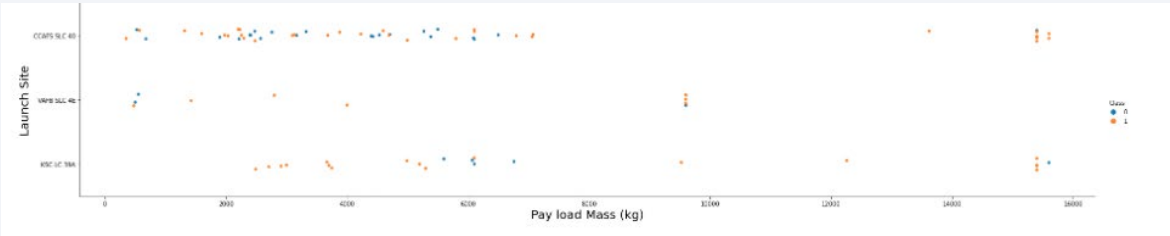
- Show a scatter plot of Flight Number vs. Launch Site
  - In standard scientific presentation, this means plotting Flight Number on the Y axis and Launch Site on the X axis, which is the *exact opposite* of how we were instructed to do it in the lab.
  - I'll plot both "Flight Number vs Launch Site" and also "Launch Site vs Flight Number" because it's pretty obvious that whoever wrote these questions is either unable or simply unwilling to clearly specify what the heck they want.
- Show the screenshot of the scatter plot with explanations
  - I think I've explained enough

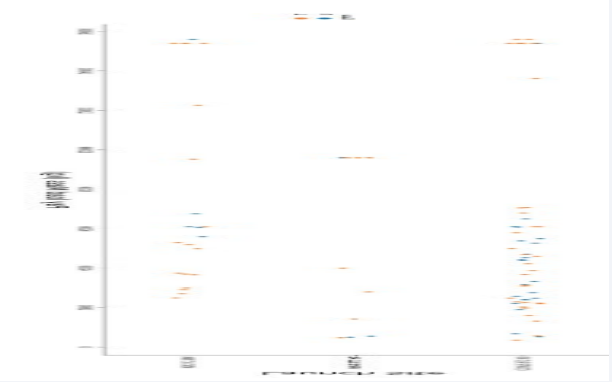




Payload vs. Launch Site

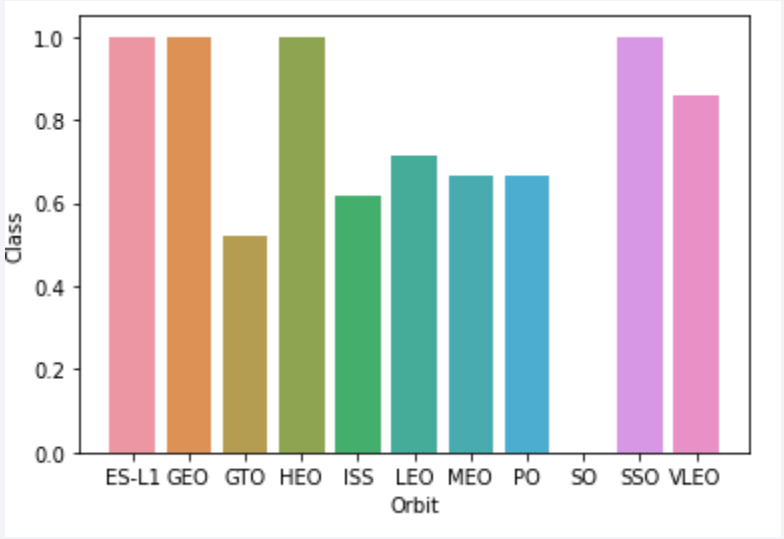
- Show a scatter plot of Payload vs. Launch Site
  - Again, “Payload vs. Launch Site” means that Payload is on the Y axis and Launch Site is on the X axis, the exact opposite of what was asked for
  - I don’t know if the instructor wants us to re-plot the graph or if they simply mis-stated this request due to a poor grasp of how to communicate with scientific rigor.
- Show the screenshot of the scatter plot with explanations
  - Done.





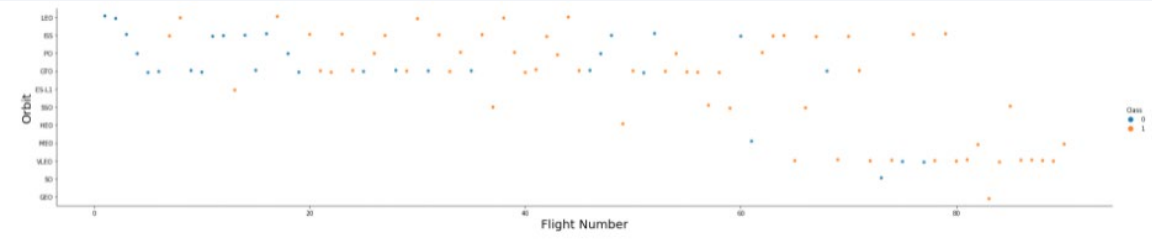
Success Rate vs. Orbit Type

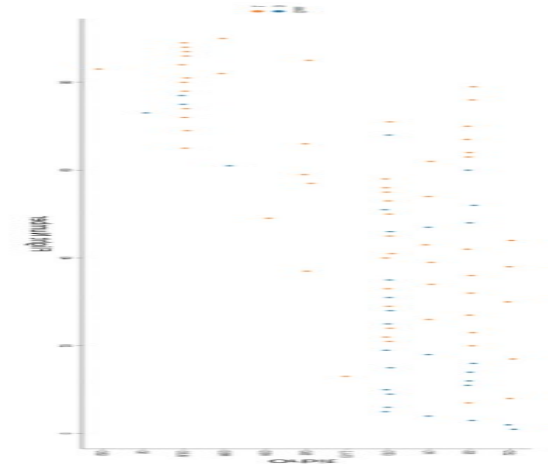
- Show a bar chart for the success rate of each orbit type
  - Okay
- Show the screenshot of the scatter plot with explanations
  - The bar chart shows the success rate for each orbit type



Flight Number vs. Orbit Type

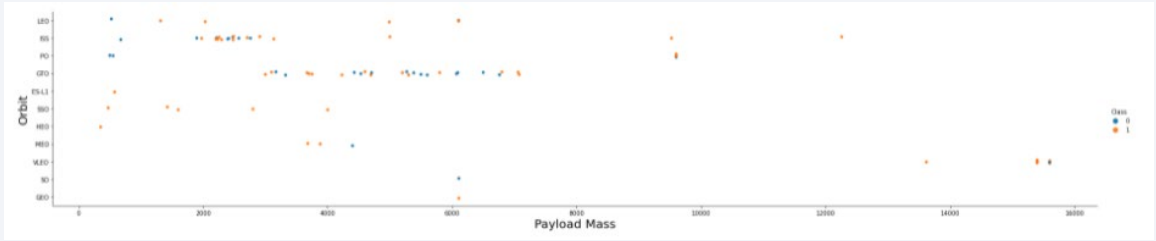
- Show a scatter point of Flight number vs. Orbit type
  - I've showed a scatter plot for both Orbit vs. Flight Number and also Flight Number vs. Orbit, because I don't think the instructor knows what he's asking for.
- Show the screenshot of the scatter plot with explanations
  - Yeah, I did that.

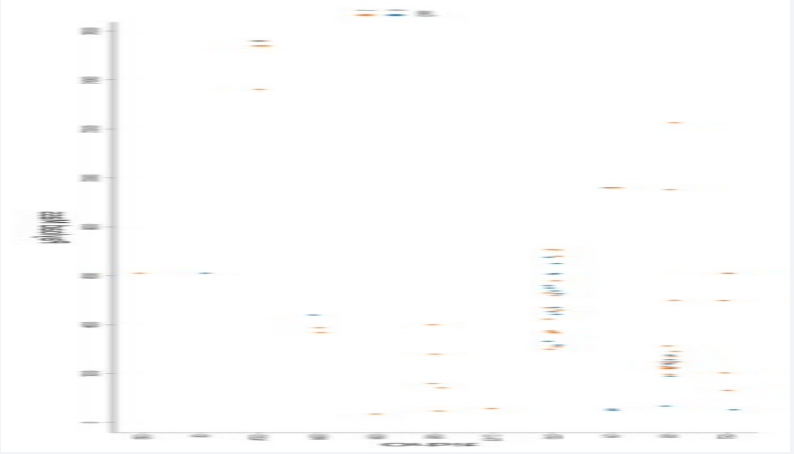




Payload vs. Orbit Type

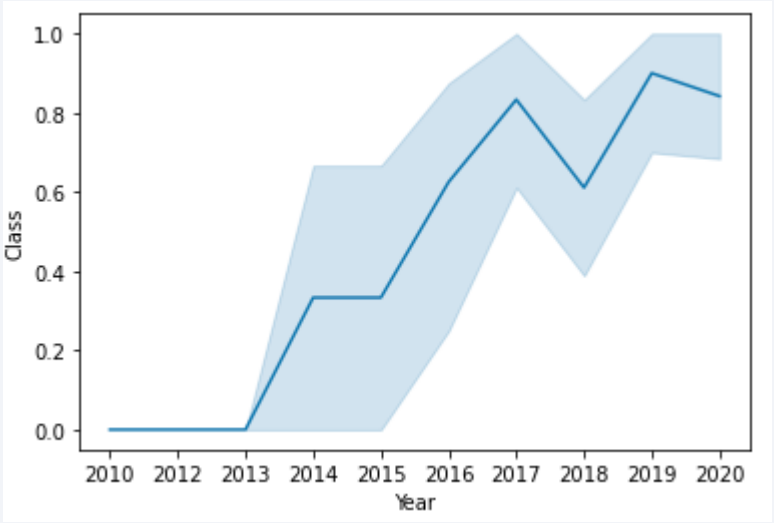
- Show a scatter point of payload vs. orbit type
  - Here's a plot of orbit vs. playload, and also a plot of payload vs. orbit. This game is getting really tiresome.
- Show the screenshot of the scatter plot with explanations
  - Done.





Launch Success Yearly Trend

- Show a line chart of yearly average success rate
  - At least he didn't ask for "Year vs. Class", which would have been the sideways version of this.
- Show the screenshot of the scatter plot with explanations
  - What does this look like?



All Launch Site Names

- Find the names of the unique launch sites
  - CCAFS LC-40
  - CCAFS SLC-40
  - KSC LC-39A
  - VAFB SLC-4E
- Present your query result with a short explanation here
  - Query: SELECT DISTINCT launch_site FROM SPACEXTBL
  - Explanation: I select distinct values of “launch_site” from table “SPACEXTBL”

```
In [4]: %sql SELECT DISTINCT launch_site FROM SPACEXTBL
 * ibm_db_sa://qqh09233:***@fbd88901-ebdb-4a4f-a32e-9822b9fb23
7b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

Out[4]:

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`
  - Records shown in screenshot
- Present your query result with a short explanation here
  - Query: SELECT * FROM SPACEXTBL WHERE launch_site LIKE 'CCA%' LIMIT 5
  - Explanation: I select the first 5 records where the value of “launch_site” starts with CCA
  - I’m deeply impressed that, for once, the question requested in this final assignment actually corresponds to the task performed in the lab

Display 5 records where launch sites begin with the string 'CCA'

```
In [5]: %sql SELECT * FROM SPACEXTBL WHERE launch_site LIKE 'CCA%' LIMIT 5
 * ibm_db_sa://qqh09233:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

Out[5]:

DATE	time_utc_	booster_version	launch_site	payload	payload_mass_kg_	orbit	customer	mission_outcome	landing_outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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)
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-12	22:41:00	F9 v1.1	CCAFS LC-40	SES-8	3170	GTO	SES	Success	No attempt

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
  - Carried by “NASA (CRS)”, as per the assignment instructions: 22,007
  - Carried by NASA in general, as per this question (which is different from what was asked for in the assignment): 36,679
- Present your query result with a short explanation here
  - QUERY (1): SELECT SUM(payload_mass__kg__) FROM SPACEXTBL WHERE CUSTOMER=‘NASA (CRS)’
  - QUERY (2): SELECT SUM(payload_mass__kg__) FROM SPACEXTBL WHERE CUSTOMER LIKE ‘NASA%’
  - Explanation:
    - The lab assignment asked us to calculate the total payload carried by boosters specifically from NASA (CRS).
    - This questions asks us to calculate the total payload carried by boosters from NASA in general, which is a superset of boosters from NASA (CRS).
    - I’ve presented both results here because the instructors seem incapable of asking for whatever it is that they actually want, and I’ll be damned if I’m docked points due to their inability to communicate.

Display the total payload mass carried by boosters launched by NASA (CRS)

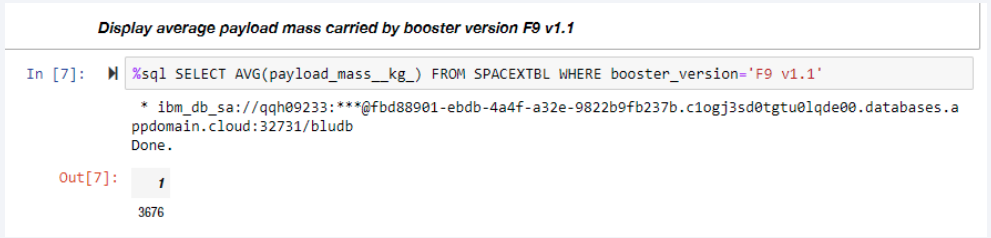
```
In [9]: %sql SELECT SUM(payload_mass__kg_) FROM SPACEXTBL WHERE CUSTOMER='NASA (CRS)'
 * ibm_db_sa://qqh09233:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
Out[9]:
1
22007
```

Display the total payload mass carried by boosters launched by NASA (CRS)

```
In [15]: %sql SELECT SUM(payload_mass__kg_) FROM SPACEXTBL WHERE CUSTOMER LIKE 'NASA%'
 * ibm_db_sa://qqh09233:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
Out[15]:
1
36679
```

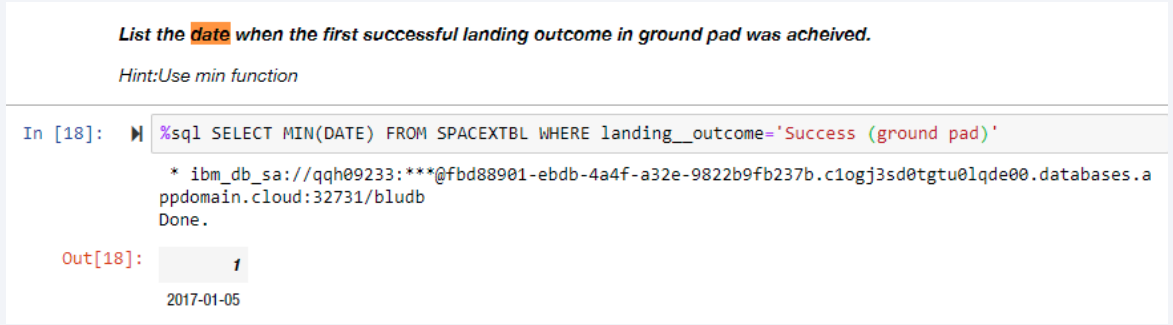
Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
  - 3,676
- Present your query result with a short explanation here
  - Query: SELECT AVG(payload_mass__kg__) FROM SPACEXTBL where booster_version='F9 v1.1'
  - Explanation: I'm extremely confused as to what's happening here. In this particular case, for once, the question presented here in the final assignment seems to be exactly what was requested in the original lab. This is highly abnormal and makes me suspect that some sort of trick is occurring.



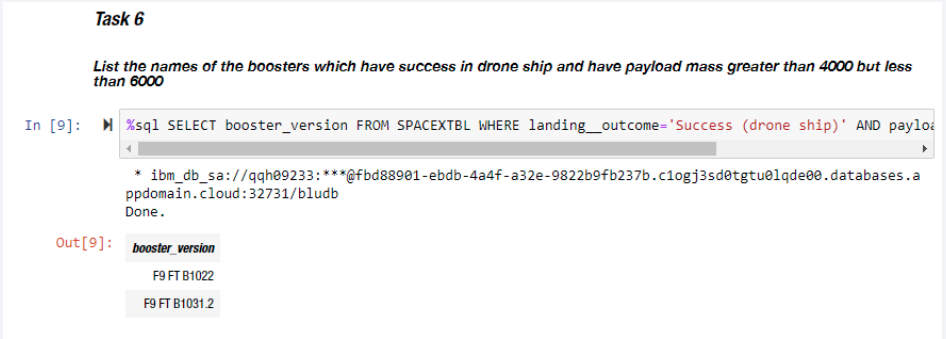
First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
  - 2017-01-05
- Present your query result with a short explanation here
  - Query: SELECT MIN(DATE) FROM SPACEXTBL WHERE landing__outcome='Success (ground pad)'
  - Explanation: I select the minimum date from the table SPACEXTBL where the value of the field “landing__outcome” is equal to the string “Success (ground pad)”. *Come on*, why are we doing this?



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
  - F9 FT B1022
  - F9 FT B1031.2
- Present your query result with a short explanation here
  - Query: SELECT booster_version FROM SPACEXTBL WHERE landing__outcome='Success (drone ship)' AND payload_mass__kg_ BETWEEN 4000 AND 6000
  - Explanation: Dude. Come on.
    Fine. I select booster_version from the table SPACEXTBL where the value of the field “landing__outcome” is “Success (drone ship)” and the value of the field “payload_mass__kg_” is between 4000 and 6000.
  - This farce is demeaning to us both. It’s pretty obvious that the instructors gave absolutely no thought into structuring these questions, and literally cut-and-pasted their way through this entire course curriculum. It’s insulting to us that they should expect us to not do likewise.



Task 6

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

```
In [9]: %sql SELECT booster_version FROM SPACEXTBL WHERE landing__outcome='Success (drone ship)' AND payloa
         * ibm_db_sa://qqh09233:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.a
        ppdomain.cloud:32731/bludb
        Done.
```

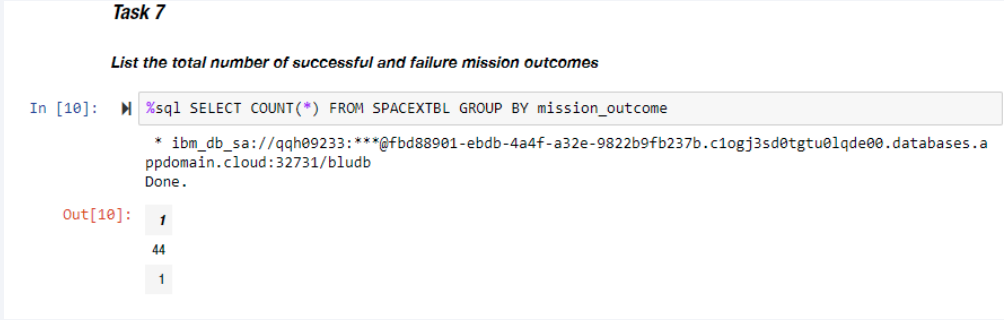
Out[9]:

booster_version
F9 FT B1022
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
  - Success: 44
  - Failure: 1

- Present your query result with a short explanation here
  - Query: SELECT COUNT(*) FROM SPACEXTBL GROUP BY mission_outcome
  - Explanation: I count the mission outcomes come on are you freaking kidding me look at the sql code it's simpler than plain English



Task 7

List the total number of successful and failure mission outcomes

```
In [10]: %sql SELECT COUNT(*) FROM SPACEXTBL GROUP BY mission_outcome
          * ibm_db_sa://qqh09233:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
         Done.
Out[10]: 1
         44
         1
```

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
  - F9 B5 B1048.4
  - F9 B5 B1049.4
  - F9 B5 B1049.5
  - F9 B5 B1058.3
  - F9 B5 B1060.2
- Present your query result with a short explanation here
  - Query: SELECT DISTINCT booster_version FROM SPACEXTBL WHERE payload_mass__kg_ = (SELECT MAX(payload_mass__kg_) AS maxpayload FROM SPACEXTBL)
  - Explanation: I'm done playing this game. Give me whatever score you feel appropriately reflects how well you believe I understand the subject matter.

List the names of the booster_versions which have carried the maximum payload mass. Use a subquery

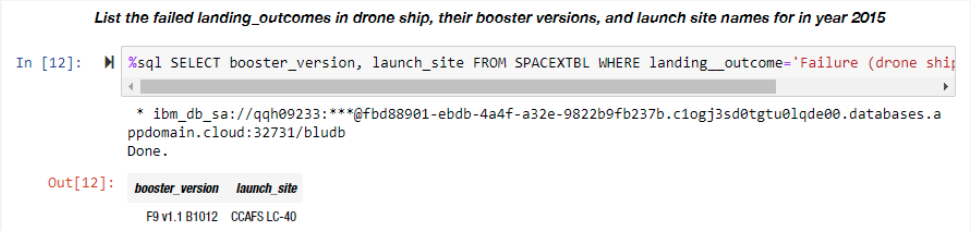
```
In [11]: %sql SELECT DISTINCT booster_version FROM SPACEXTBL WHERE payload_mass__kg_ = (SELECT MAX(payload_r
          * ibm_db_sa://qqh09233:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.a
ppdomain.cloud:32731/bludb
Done.
```

Out[11]:

booster_version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1058.3
F9 B5 B1060.2

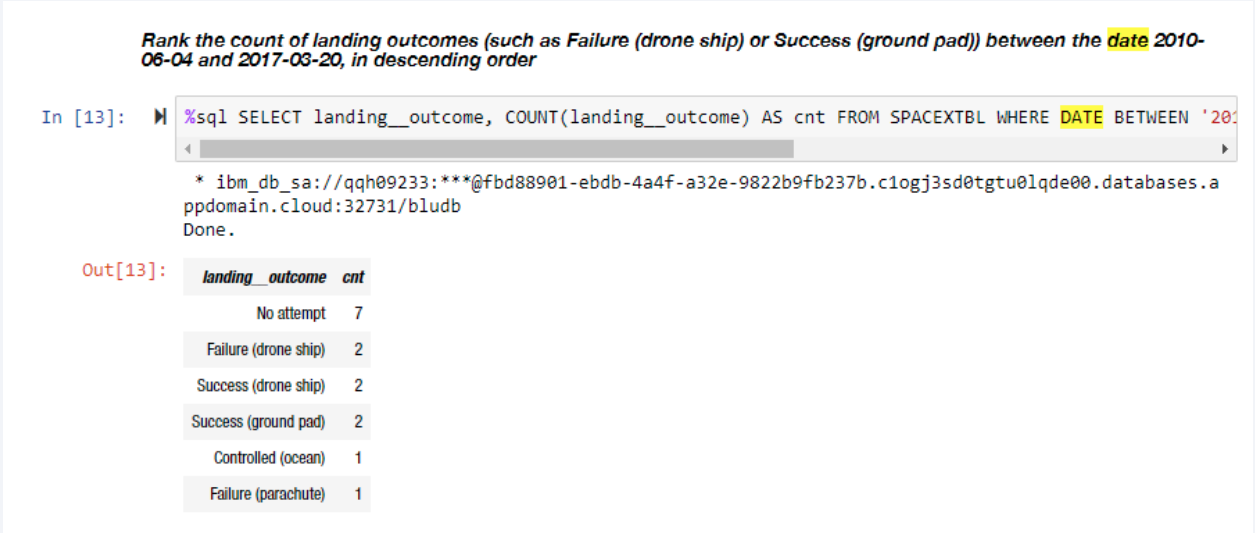
2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
  - booster_version: F9 v1.1 B1012
    launch_site: CCAFS LC-40
- Present your query result with a short explanation here
  - SELECT booster_version, launch_site FROM SPACEXTBL WHERE landing__outcome='Failure (drone ship)' AND DATE LIKE '2015%'
  - Explanation: Come on.



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
  - No attempt 7
  - Failure (drone ship) 2
  - Success (drone ship) 2
  - Success (ground pad) 2
  - Controlled (ocean) 1
  - Failure (parachute) 1
- Present your query result with a short explanation here
  - Query: SELECT landing__outcome, COUNT(landing__outcome) AS cnt FROM SPACEXTBL WHERE DATE BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY landing__outcome ORDER BY cnt DESC
  - Explanation: Provided sufficiently as needed.



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

In [13]:
```
%sql SELECT landing__outcome, COUNT(landing__outcome) AS cnt FROM SPACEXTBL WHERE DATE BETWEEN '201
```

```
 * ibm_db_sa://qqh09233:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

Out[13]:

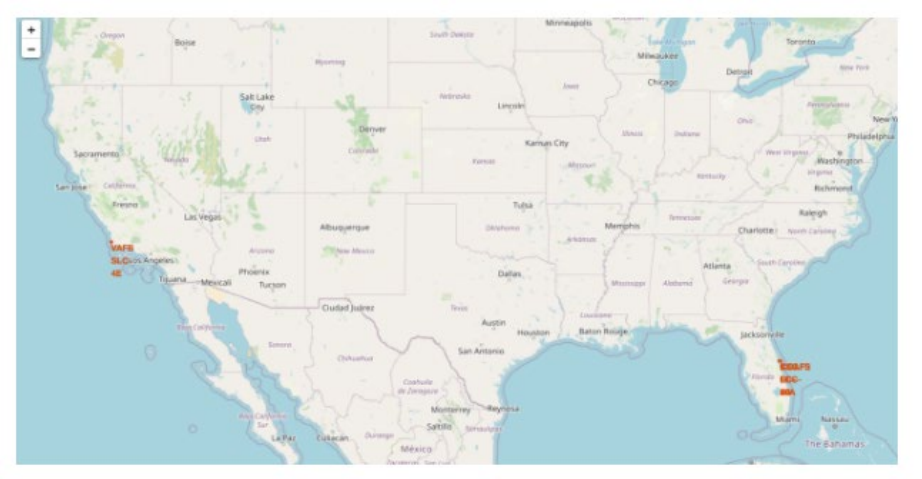
landing__outcome	cnt
No attempt	7
Failure (drone ship)	2
Success (drone ship)	2
Success (ground pad)	2
Controlled (ocean)	1
Failure (parachute)	1

Section 4

Launch Sites Proximities Analysis

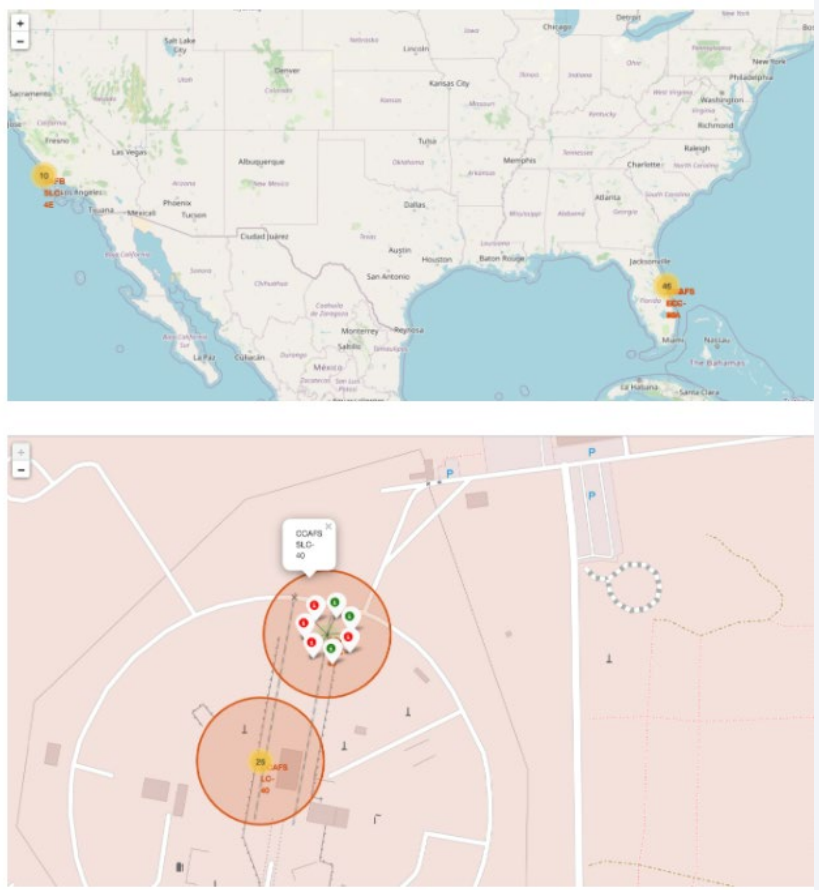
Folium plot of first or maybe second map

- Replace <Folium map screenshot 1> title with an appropriate title
  - Ok. Do you mean literally the first map, which only shows NASA JSC? Or do you mean the second map, which includes multiple launch sites?
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map
  - Ok. I have shown map markers on both the first and the second maps.
- Explain the important elements and findings on the screenshot
  - In the first map, they show the Corpus Christi (near-ish to Houston) launch site. In the second one, they show the launch sites in Florida and California.



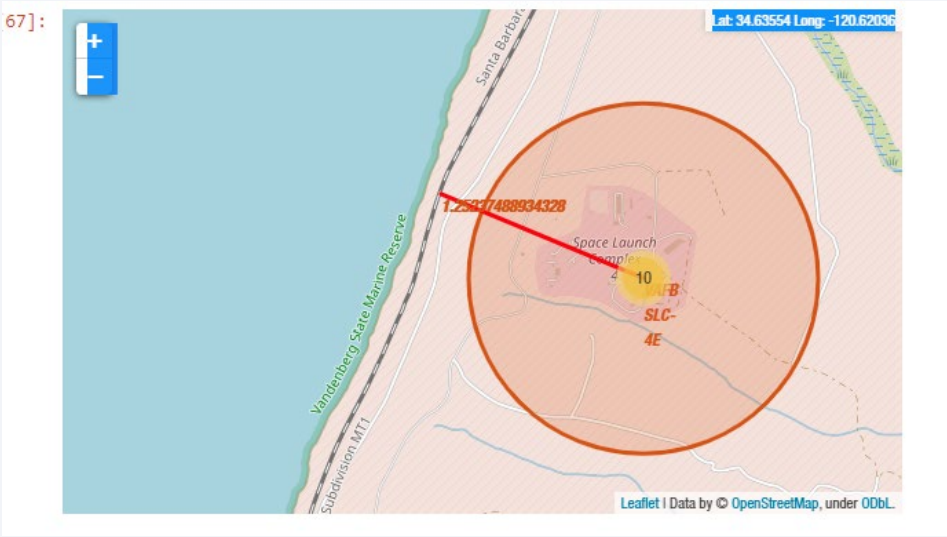
Color-labeled launch outcomes

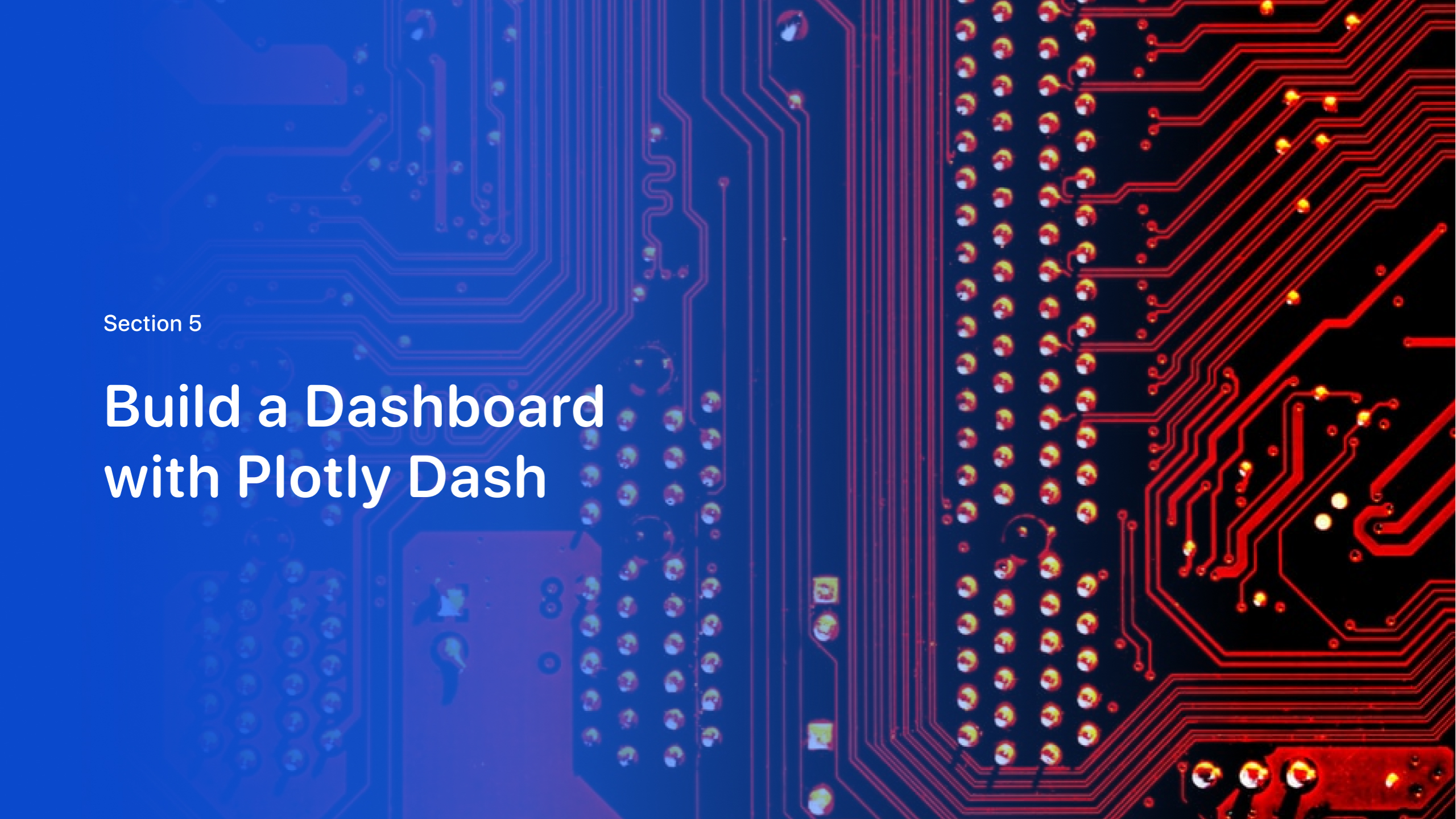
- Replace <Folium map screenshot 2> title with an appropriate title
  - Ok. It's not actually the second map; it's the third. But whatever.
- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map
  - Done.
- Explain the important elements and findings on the screenshot
  - The successful launches are green. The failures are red.



Distance from launch site to railway

- Replace <Folium map screenshot 3> title with an appropriate title
  - Ok
- 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
  - Ok
- Explain the important elements and findings on the screenshot
  - This screenshot shows very important elements because it is very important to know that the very important VAFB SLC-4E launch site is 1.2523748893432 8 km away from the nearest very important railway.



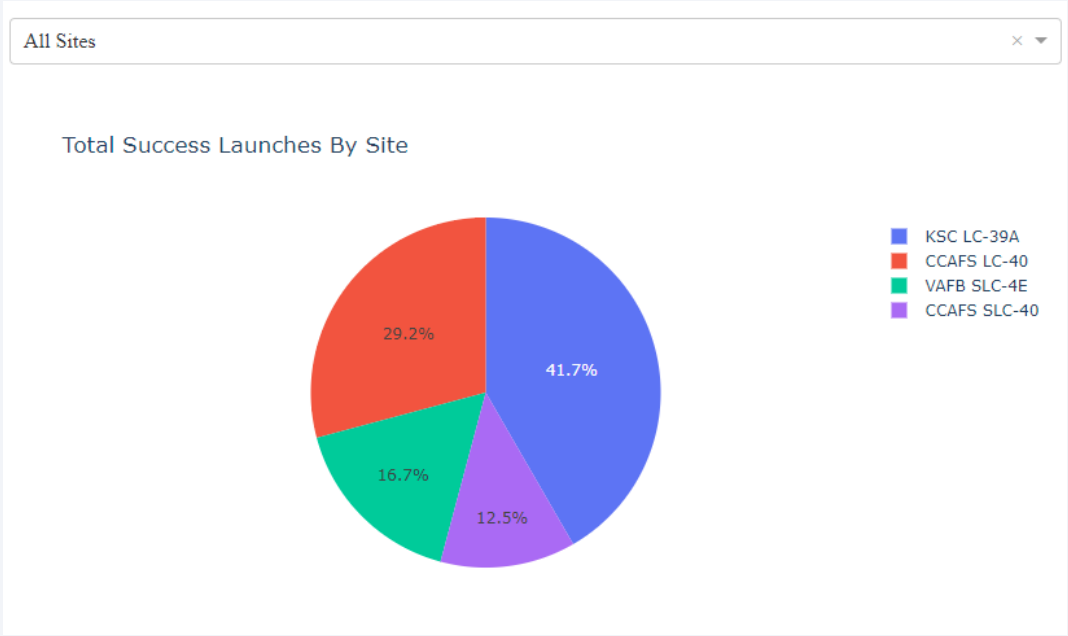


Section 5

Build a Dashboard with Plotly Dash

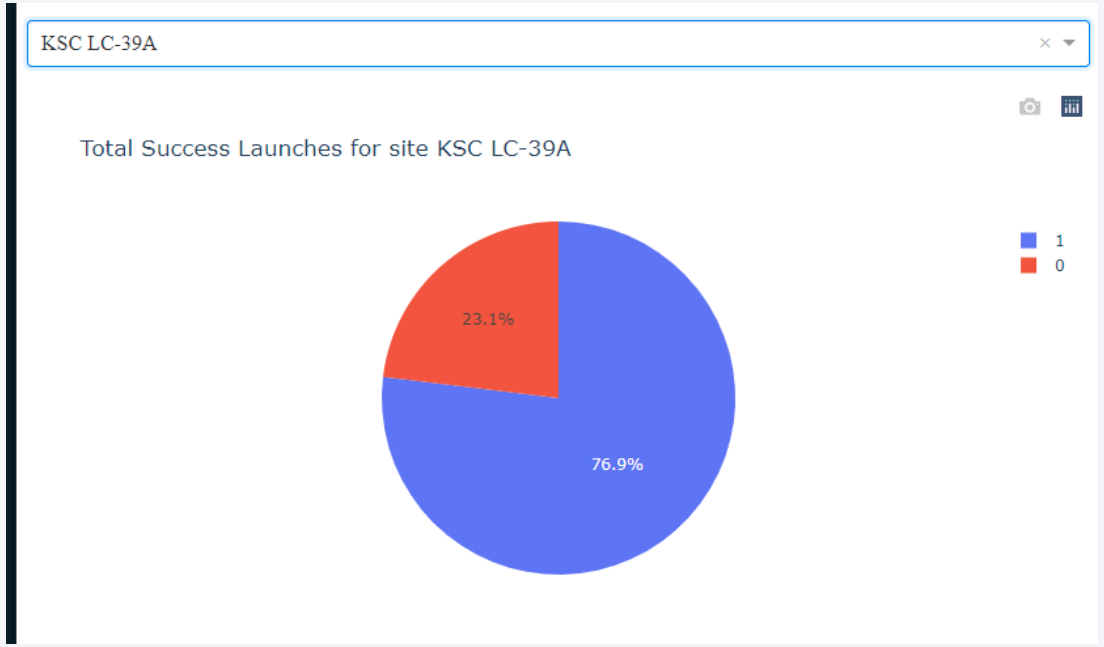
Plotly Dash Launch Success Count For All Sites

- Replace <Dashboard screenshot 1> title with an appropriate title
  - Sure, whatever
- Show the screenshot of launch success count for all sites, in a piechart
  - Shown
- Explain the important elements and findings on the screenshot
  - There are no “important” elements, because this chart is not important. But there is a pie chart, a title, a color key, and a selection dropdown.



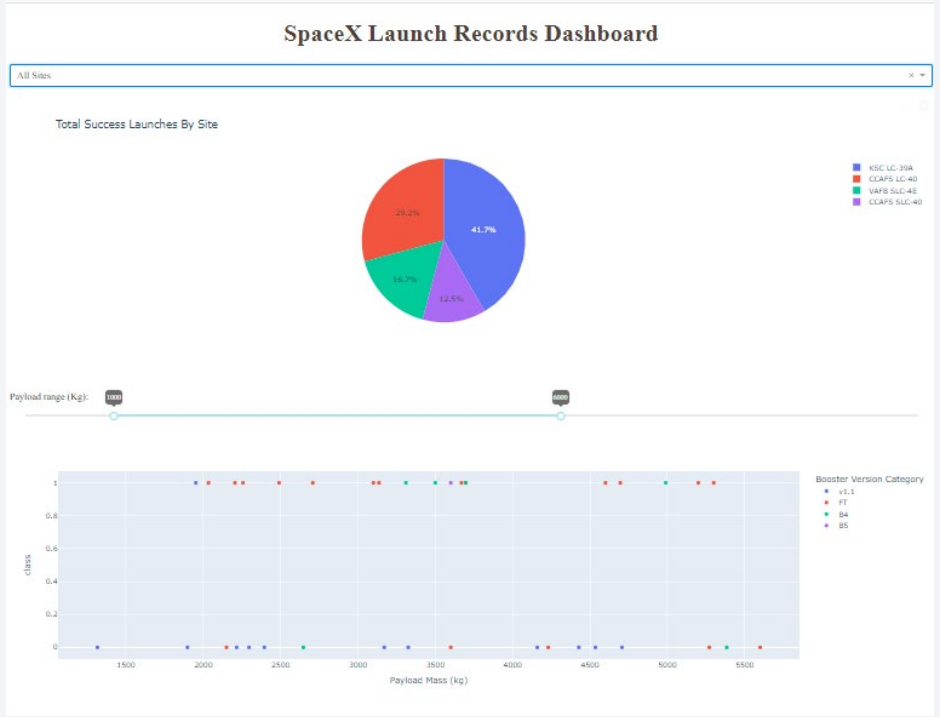
KSC LC-39A has highest success ratio

- Replace <Dashboard screenshot 2> title with an appropriate title
  - Ok
- Show the screenshot of the piechart for the launch site with highest launch success ratio
  - KSC LC-39A has the highest success ratio, with 76.9% success
- Explain the important elements and findings on the screenshot
  - There is a pie chart, a title, a color key, and a selection dropdown.



Launch Outcome vs. Payload Mass for All Launch Sites

- Replace <Dashboard screenshot 3> title with an appropriate title
  - Ok
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
  - For the last time, it's "Y vs. X", not "X vs. Y". We do not have a chart of "Payload vs. Launch Outcome", for heaven's sake. We *do* have a chart of Launch Outcome vs. Payload, which I have attached.
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.
  - There is a pie chart, a title, a color key, and a selection dropdown.
  - FT boosters seem to have a high rate of success when delivering small payloads.

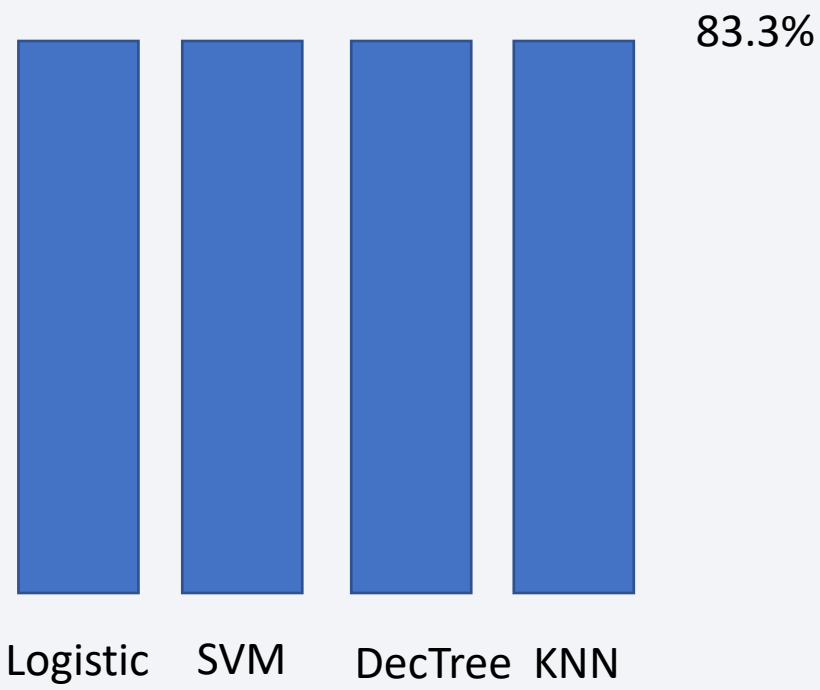


Section 6

Predictive Analysis (Classification)

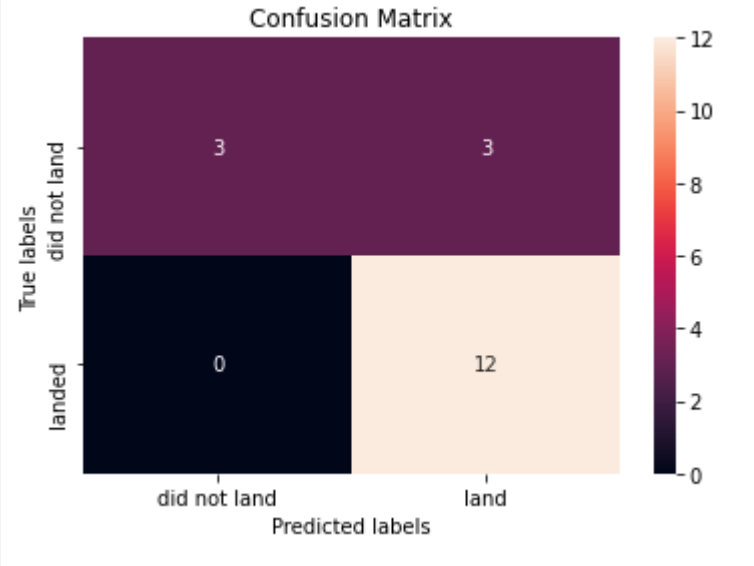
Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart
  - Is this a joke? They all scored 83.3%
- Find which model has the highest classification accuracy
  - All models scored 83.3%.



Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation
  - Of the 18 test cases…
    - 12 were predicted to land and did indeed land
    - 3 were predicted to land but did not land
    - 3 were predicted to not land and did indeed not land
    - 0 were predicted to not land but did indeed land



Conclusions

- Point 1
  - The instructors don't understand the difference between a plot of "Y vs. X" and "X vs. Y"
- Point 2
  - The instructors expected every single one of us to spend more time and do more work on this assignment than they themselves did in organizing it. The instructors copy-pasted the majority of the slides in this presentation.
- Point 3
  - Rockets with larger payloads and higher-orbit deployment targets are less likely to land successfully, which is, like, duh.
- Point 4
  - Newer rockets (higher version designations) are more likely to land successfully, which, again, we should darn well hope so.

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
  - All code, queries, charts, etc. are in my GitHub repository

    https://github.com/omedalus/SpaceX-Capstone/tree/master

Thank you!