IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Katarina Sekelez
30/01/2022

Outline

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

Executive Summary

- **The main problem** this capstone is addressing is to 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 it is possible to determine if the first stage will land, I can determine the cost of a launch.″ This information can be used if an alternate company wants to bid against SpaceX for a rocket launch. (Source: www.edx.org)

- **Data requirements** showed the need of data about Spacex rockets with the information about the rockets, landing outcome, mission success andconnected records.

- Mehodology that was used for **collecting data** was taking available data from SpaceX website using API data collectione and web scraping using Wiki page https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches, to collect Falcon 9 and Falcon 1 historical launch records.

- The **data wrangling** stage of research was used to find classification variable that will represent the outcome of each launch. If the value is zero, the first stage did not land successfully; one means the first stage landed successfully.

- **Techniques** such as descriptive statistics and visualization of dana are applied to the data set, to assess the content, quality, and initial insights about the data. EDA with SQL was performed using SQL magic queries and EDA with visualization using Folium and Dash.

- For **Data Modelling** was used a predictive modeling and Classification approach.

- **Results** of Capstone show that Falcon 9 rocket will successfully land and also that the success of landing is in correlation with it's payload mass, that landing sites are mostly close to equator (explanation in future slides) and ocean, but away from big cities and that there were 100 mission outcomes marked as successful (with or without additional description of mark) for Falcon 9. Best classification algorithm is Decision tree and test set showed that 12 out of 15 launches have successfully landed in reality, which represents 66,67% of success.

Introduction

- By finalizing this Capstone project, it will be predicted 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. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch. In this module, you will be provided with an overview of the problem and the tools you need to complete the course. (Source: www.edx.org)
- This is also **an interesting insight about reusable flights** from wikipedia:
- https://en.wikipedia.org/wiki/SpaceX_reusable_launch_system_development_program

Section 1

Methodology

Methodology

Executive Summary

- **Data collection** methodology:
  - performed data collection from an API: made a get request to the SpaceX API (https://api.spacexdata.com/v4/rockets/) and cleaned the requested dana to only Falcon 9 launches
  - performed data collection with webscraping to collect Falcon 9 historical launch records from the website :https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- Performed **data wrangling** using Exploratory Data Analysis and by determining Training Label: landing outcome label
- Performed **exploratory data analysis (EDA)** using visualization and SQL: visualization with scatter plots,bar chart and line chart and SQL using SQL magic queries
- Performed **interactive visual analytics** using Folium and Plotly Dash: used Folium to determine NASA location, launch sites's circles and distances from launch sites to coastline, city, ocean, railay.; performed Plotly Dash to show the relationship between different data (number of each launch site in total successful launches as percentage and the correlation between payload mass and landing outcome success)
- Performed **predictive analysis** using classification models using predictive analysis using classification models: Logistic Regression, Support Vector Machine, Decision tree and K Nearest Neighbors algorithms

Data Collection

- Data collection methodology:
- Performed data collection from an API: made a get request to the SpaceX API (https://api.spacexdata.com/v4/rockets/) and cleaned the requested dana to only Falcon 9 launches
- Performed data collection with webscraping to collect Falcon 9 historical launch records from the website https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches

DaTa collection from SPACEX API

↓

DaTa regarding FALCON 9 launches for launch dictionary

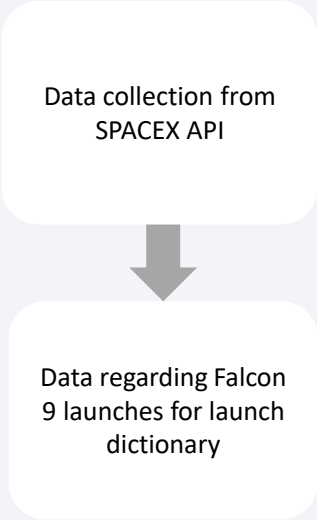
Data collection using webscraping from wiki Falcon 9 HISTORICAL LAUNCES records

↓

More DaTa regarding Falcon 9 launches for launch dictionary

Data Collection – SpaceX API

- made a get request to the SpaceX API (https://api.spacexdata.com/v4/rockets/) and cleaned the requested data to only Falcon 9 launches
- requested JSON results
- turned it into a Pandas dataframe using .json_normalize()
- used the API to get information about the launches using the IDs given for each launch: columns rocket, payloads, launchpad, and cores from which we created global variables for more information about : Booster versions, launch sites, payload dana and core data
- using this data created a Launch dictionary
- from this dictionary created a Pandas dataframe with Falcon 1 and Falcon 9 information
- filtered the dataframe to only include Falcon 9 launches
- identified and Cleaned the missing values from this dataset
- https://github.com/SKatest/Portfolio



Data Collection - Scraping

- performed webscraping to collect Falcon 9 historical launch records from the website https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- performed HTTP GET method to request the Falcon9 Launch HTML page
- created a BeautifulSoup object from the HTML response
- target table contained the actual launch records
- column names were embedded in the table header elements <th> : iteration through the <th> elements
- extracted column names to create an ampty dictionary and
- filled up the launch dictionary with new columns (append to launch dictionary) ; parsing process
- created a dataframe
- https://github.com/SKatest/Portfolio

Data collection using webscraping from wiki Falcon 9 historical launches records

More Data regarding Falcon 9 launches for launch dictionary

Data Wrangling

- data wrangling was performed using Exploratory Data Analysis, EDA:
- from previusly collected data on FALCON 9 in SpaceX dataset, after loading table:
- calculated the number of launches on each site (most CCAFS SLC 40: 55)
- as each launch aims to a dedicated orbit I calculated the number and occurrences of each orbit and occurence of mission outcome per orbit type using method value_counts() on columns Orbit and Outcome
- creating an outcome label: created a set of outcomes and a landing outcome label from Outcome column: a list where the element is 0 if the corresponding row in Outcome is in the set bad_outcome; otherwise, it's 1
- assigned the list to the variable landing_class
- landing_class is classification variable that represents the outcome of each launch. If the value is zero, the first stage did not land successfully; one means the first stage landed Successfully; this column CLASS is also important for prediction models
- using mean() determinated a Success rate (67%)
- https://github.com/SKatest/Portfolio

EDA on Falcon 9 dataset → Create a set of bad outcomes where the second stage did not land successfully → Created a landing outcome label from Outcome column, set bad_outcome used as a condition → Variable landing_class will be a classification variable

EDA with Data Visualization

- Visualization with:
- **scatter plots** for: relationship between Flight Number and Launch Site, Payload mass and Launch Site, FlightNumber and Orbit type, Payload and Orbit type
- **bar chart**: relationship between success rate of each orbit type
- **line chart** to visualize the launch success yearly trend

- https://github.com/SKatest/Portfolio

EDA with SQL

- SQL queries using SQL magic was displayed to find:
- names of the unique launch sites in the space mission
- 5 records where launch sites begin with the string 'KSC‘ and ‘CCA’ for purpose of presentation
- total payload mass carried by boosters launched by NASA (CRS)
- average payload mass carried by booster version F9 v1.1
- date where the first successful landing outcome in drone ship was achieved
- names of the boosters which have success in ground pad and have payload mass greater than 4000 but less than 6000
- total number of successful and failure mission outcomes
- names of the booster_versions which have carried the maximum payload mass
- month names, succesful landing_outcomes in ground pad, booster versions, launch_site for the months in year 2017
- ranking of the count of successful landing_outcomes between the date 2010-06-04 and 2017-03-20 in descending order
- https://github.com/SKatest/Portfolio

An Interactive Map with Folium

- used EDA with Folium to determine NASA location ,launch sites's names, markers of successful launches and circlemarkers for distance from coastline
- used mousepoint to determine latitude and longitude of closest highway, railway, city, coastline
- line distances from launch sites to highway, railway, city, coastline etc.;

- https://github.com/SKatest/Portfolio

A Dashboard with Plotly Dash

- Performed Plotly Dash to draw :
- dropdown list to enable Launch Site selection
- a **pie chart** to show the total successful launches count for all sites
- **a scatter chart** to show correlation between payload mass and outcome success

- https://github.com/SKatest/Portfolio

Predictive Analysis (Classification)

- used predictive analysis using classification models:
- built Logistic Regression, Support Vector Machine, Decision Tree and K Nearest Neighbors objects and
- output the GridSearchCV object for each of the objects
- for each algorithm fitted the object to find the best parameters from the dictionary parameters
- displayed the best parameters using the data attribute best_params_
- displayed accuracy of each algorithm on the validation data using the data attribute best_score_
- calculated the accuracy on the test data using method score
- plotted confusion matrix for each algorithm
- highest mean of train and test accuracy of algorithms checked also according in results section
- found method that performs best using max() of the best_score_ for each algorithm;

- https://github.com/SKatest/Portfolio

Results

- **EDA with SQL RESULTS:**
- launch_sites: CCAFS LC-40, CCAFS SLC-40, KSC LC-39A, VAFB SLC-4E
- booster versions for Launch sites beginning with KSC: F9 FT B1031.1, F9 FT B1030, F9 FT B1021.2, F9 FT B1032.1, F9 FT B1034
- total payload mass carried by boosters launched by NASA (CRS): 45596 kg
- average payload mass carried by booster version F9 v1.1: 2928 kg
- date where the first succesful landing outcome in drone ship was acheived: 8/4/2016
- boosters which have success in ground pad between 4000 and 6000 kg:F9 FT B1032.1,F9 B4 B1040.1.F9 B4 B1043.1
- total number of successful mission outcomes is 100 and total of failure mission outcomes is 1
- as listed the names of the booster_versions which have carried the maximum payload mass:
- F9 B5 B1048.4F9 B5 B1049.4,F9 B5 B1051.3,F9 B5 B1056.4,F9 B5 B1048.5F9 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
- records which display the month names, succesful landing_outcomes in ground pad ,booster versions, launch_site for the months in year 2017: recored on 19-02-2017, 1/5/2017, 3/6/2017, 14-08-2017, 7/9/2017, 15-12-2017
- ranking of the count of successful landing_outcomes between the date 2010-06-04 and 2017-03-20 in descending order: Success (drone ship): 5 and Success (ground pad): 3

Results

EDA visualization results:

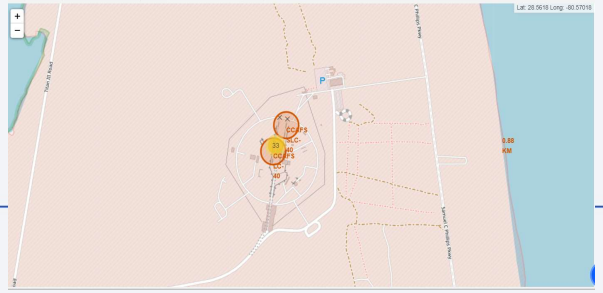
- Flight number was the biggest on launch site CCAPS SLC 40;
- for the bigger flight number, bigger is the number of positive landing outcomes
- for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000)
- orbits ES-L1,SSO HEO and GEO have the same and biggest successss rates
- 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
- with heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS.
- for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both there here
- the launch sucess rate since 2013 kept increasing till 2020

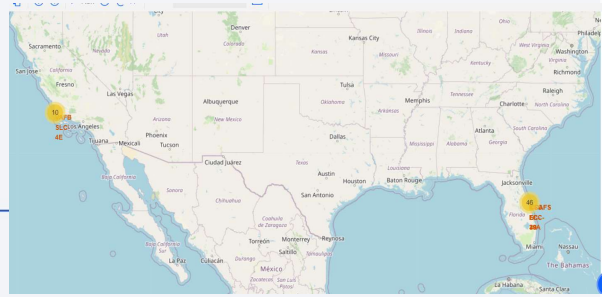
Results

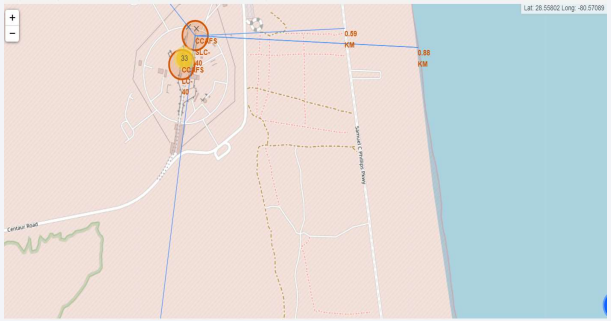
Interactive analytics demo in screenshots; sreenshots Folium

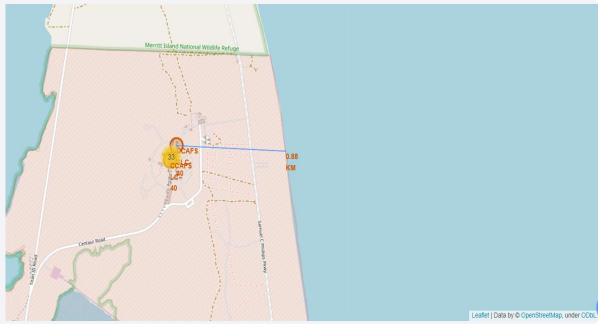
Findings:

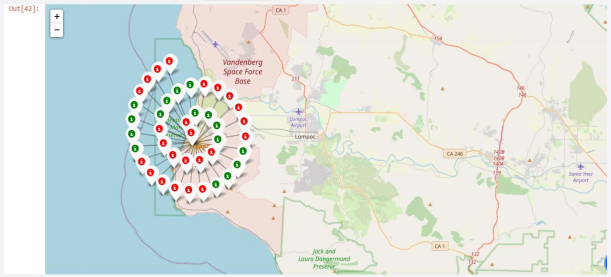
- KSC LC 39A and CCAFS LC 40 show huge number of green markers, that represent successful outcomes
- Distances from launch sites to highway, railway and coastline are not big because of the necessary conditions for launching: to be able to land in the water (ocean) and to be able to bring the necessery equipment and the Flacon 9 itself.
- #Distance from the closest city is bigger because of lower danger of hurting numerous people if somehing goes wrong with eather launhing or landing.
- Launch sites are in proximity close to the equator. I found out when searching about this, that it is because
- rockets launched from sites near the Equator get an additional natural boost that helps save the cost of putting in extra fuel and boosters.˝
- Source: https://www.scienceabc.com/eyeopeners/why-are-rockets-launched-from-areas-near-the-equator.html)

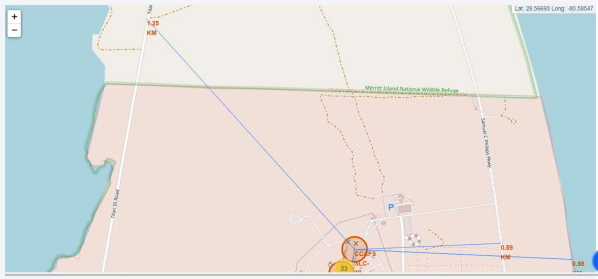


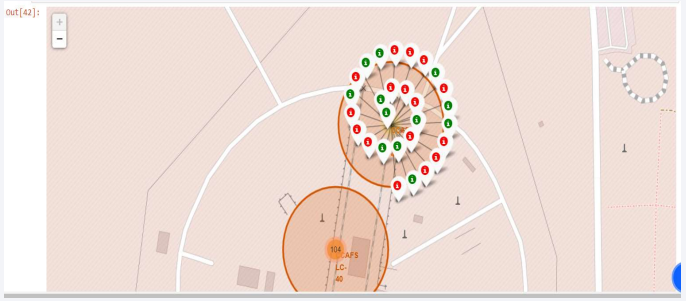


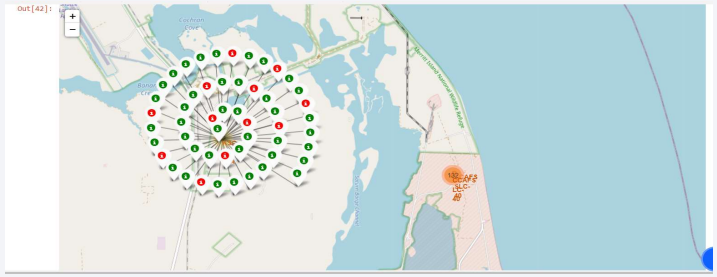






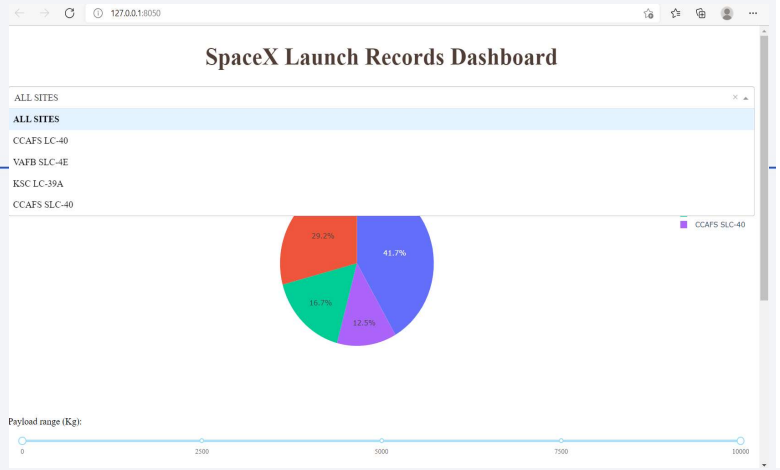


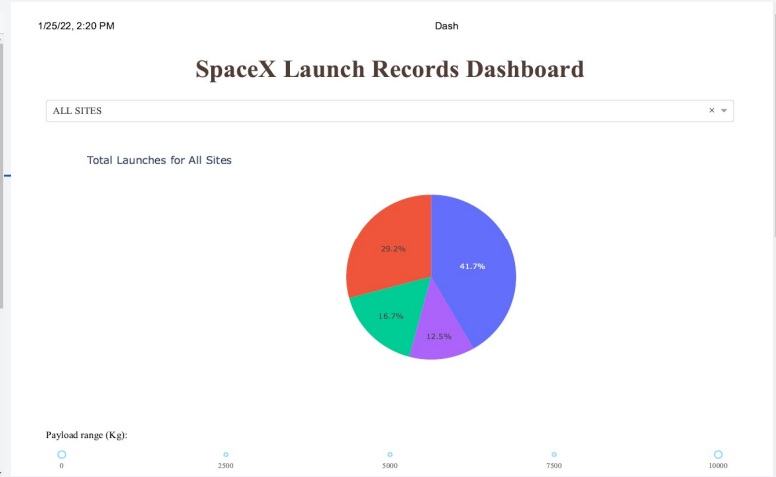


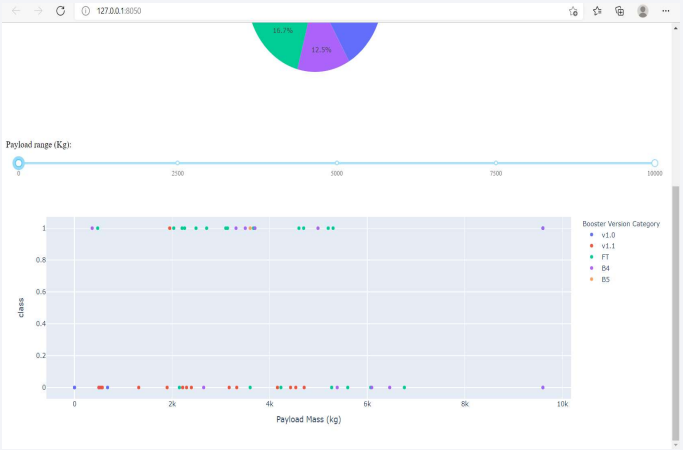


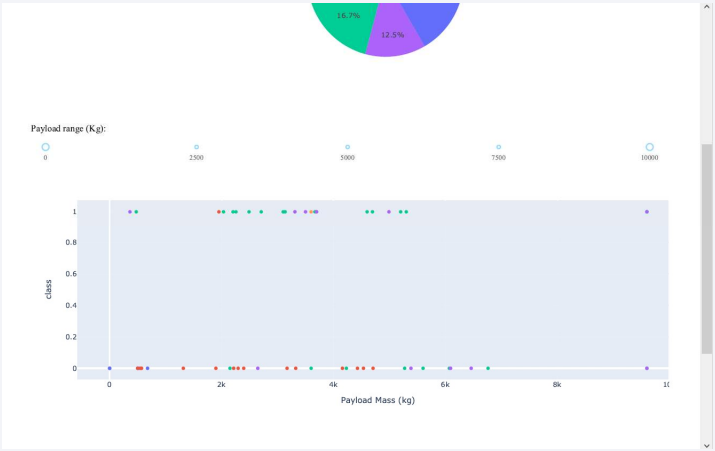
Results

- **Interactive analytics demo in screenshots; Screenshots DASH I. part**
- bigger the payload mass makes more difficult a successful landing outcome
- KSC LC 39A is a launch site with the highest launch success ratio



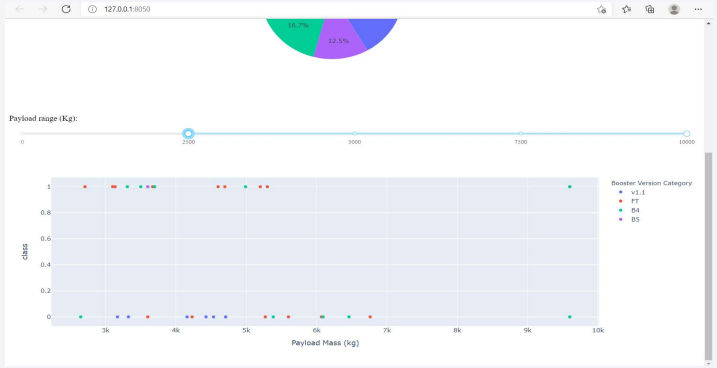


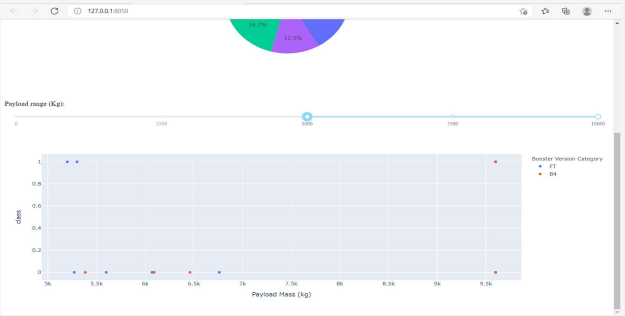




Results

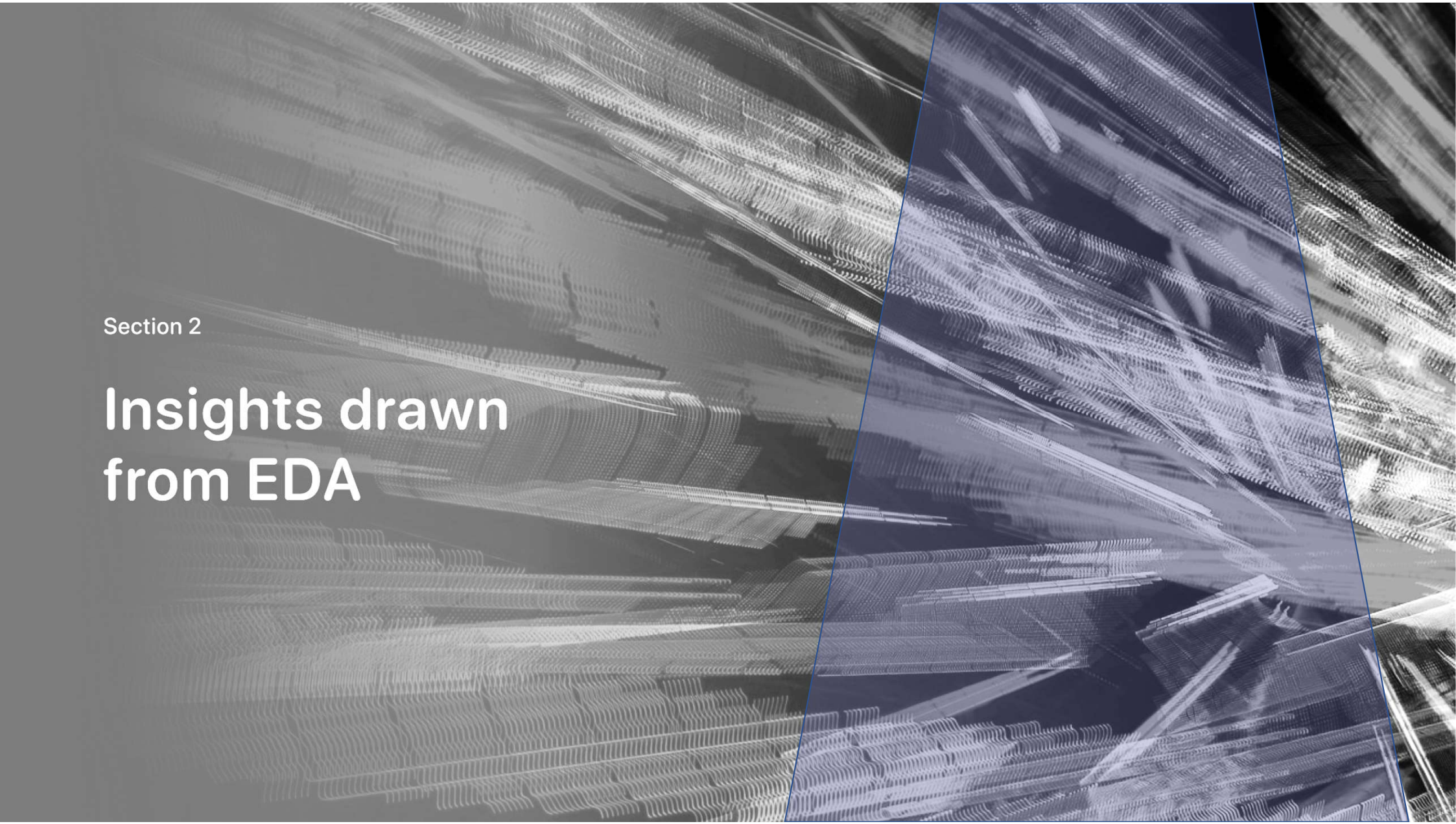
- **Interactive analytics demo in screenshots; Screenshots DASH II. part**
- bigger the payload mass makes more difficult a successful landing outcome
- KSC LC 39A is a launch site with the highest launch success ratio





Results

- **Predictive analysis results:**
- there were 18 records in the test sample
- the best algorithm to predict landing outcome is Decision tree, with the accuracy of train test of 87,5%
- the mean of train and test best accuracy of algorithms: also highest for Decision tree because of 83.33 % of test accuracy of tests and highest train test accuracy of Decision tree model
- Confusion matrix shows that Falcon 9 will successfully land with classifier that predicted correctly 12/15 successfully landed outcomes and 3/3 unsuccesful landing outcomes. Out of 18 observed outcomes, 12 were successful in reality
- after selecting the best hyperparameters, the accuracy achieved on the test daTa for Decision tree is 83,33%



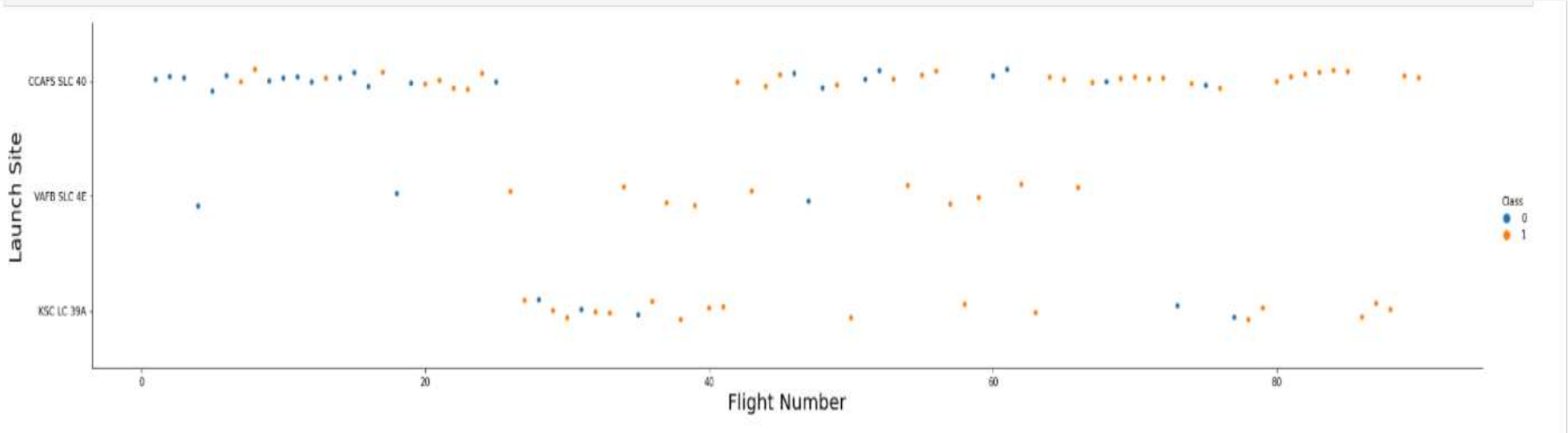
Section 2

Insights drawn from EDA

Flight Number vs. Launch Site

- flight number was the biggest on launch site CCAPS SLC 40;
- for the bigger the flight number bigger is the number of positive landing outcomes

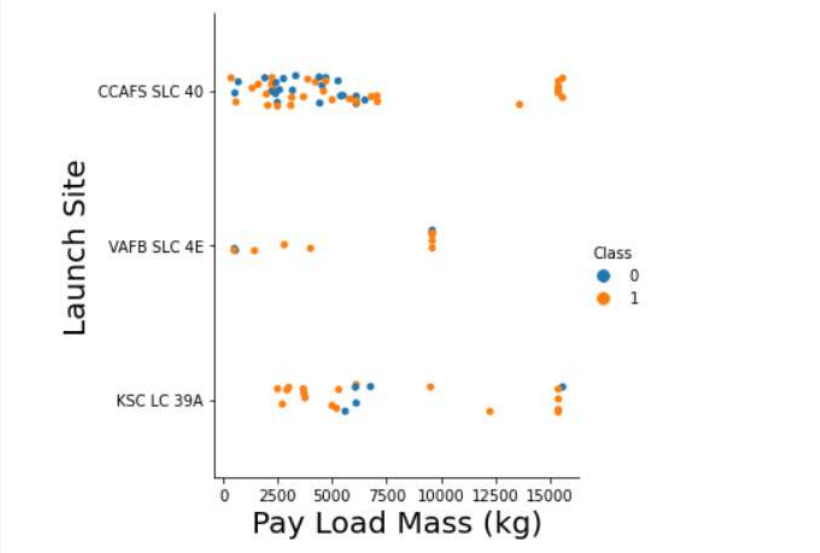
- https://github.com/SKatest/Portfolio



Payload vs. Launch Site

- from the VAFB-SLC launchsite there are no rockets launched for heavy payload mass(greater than 10000)
- from CCAFS SLC 40 launch sites there were many launches with up to 5000kg

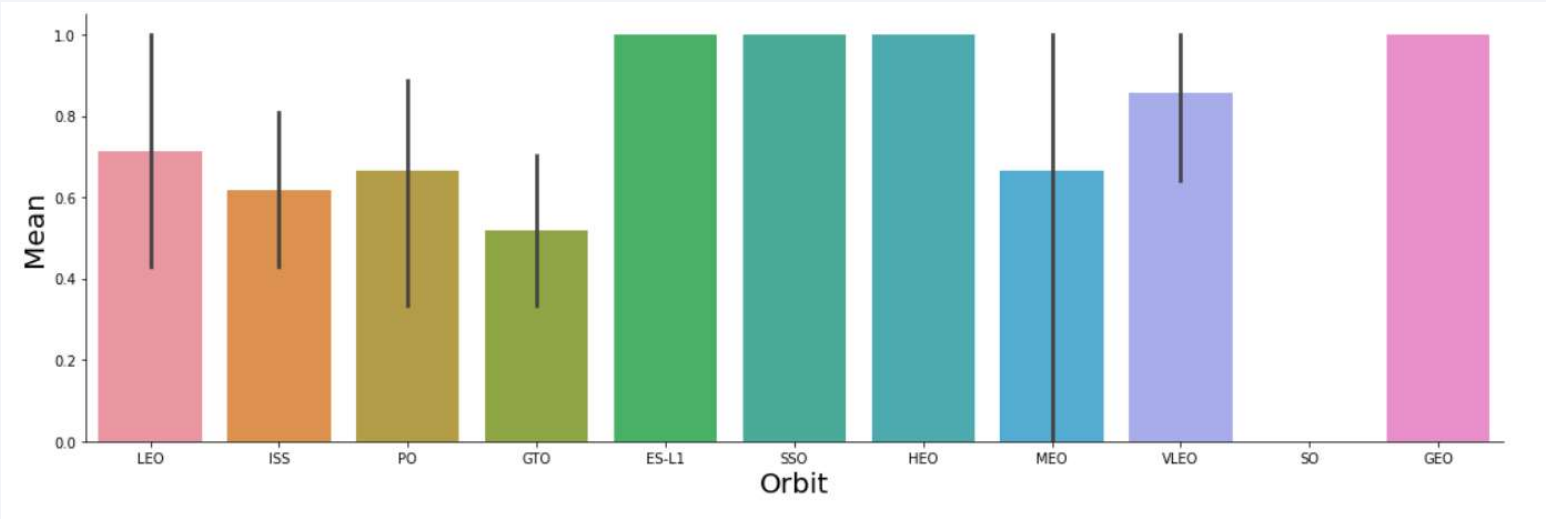
- https://github.com/SKatest/Portfolio



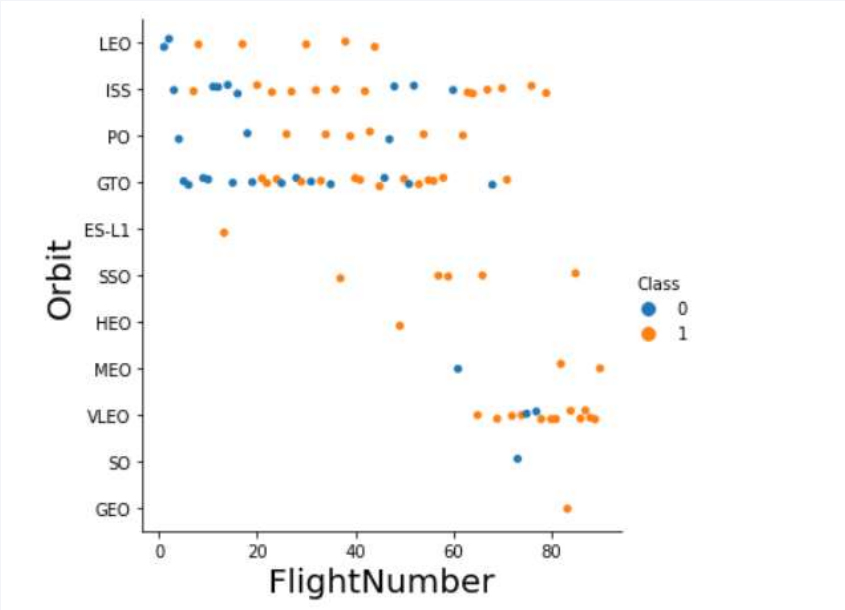
Success Rate vs. Orbit Type

- Orbits ES-L1,SSO HEO and GEO have the same and biggest success rates in the LEO orbit

- https://github.com/SKatest/Portfolio



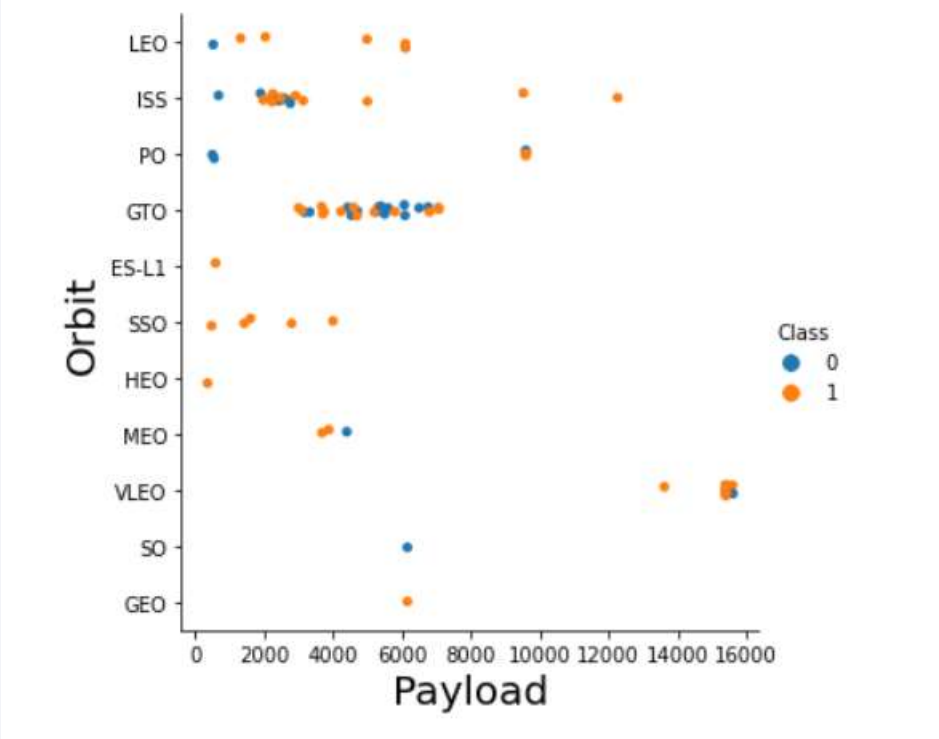
Flight Number vs. Orbit Type



- success is related to the number of flights; in the LEO orbit the Success appears related to the number of flights
- there seems to be no relationship between flight number when in GTO orbit

- https://github.com/SKatest/Portfolio

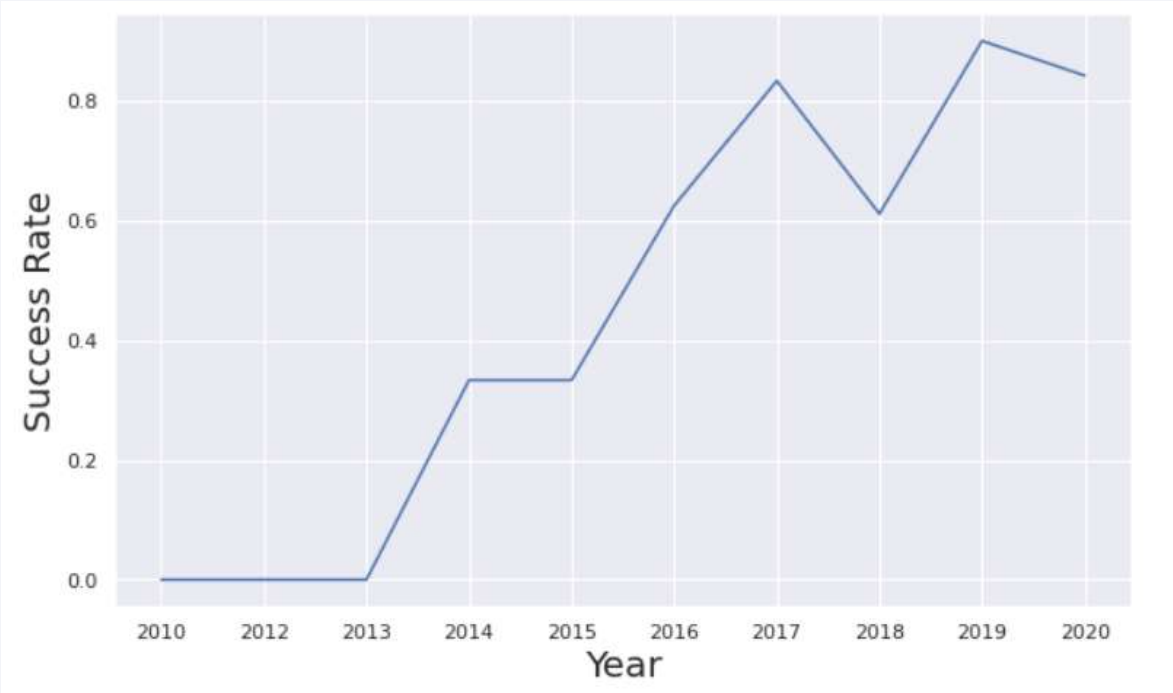
Payload vs. Orbit Type



- with heavy payloads the successful landing or positive landing rate are more for Polar,LEO and ISS
- for GTO we cannot distinguish this well as both positive landing rate and negative landing(unsuccessful mission) are both here

- https://github.com/SKatest/Portfolio

Launch Success Yearly Trend



- launch sucess rate since 2013 kept increasing till 2020
- https://github.com/SKatest/Portfolio

All Launch Site Names

- %sql select distinct(LAUNCH_SITE) from SPACEXKAT
- Results:

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

- **Explanation:** select distinct as a query to distinct unique values
- https://github.com/SKatest/Portfolio

Launch Site Names Begin with 'CCA': 5 records where launch sites begin with `CCA`

- %sql select * from SPACEXKAT where LAUNCH_SITE like 'CCA%' limit 5
- **Explanation:** selected results represent launch sites that begin with CCA, top 5 listed
- https://github.com/SKatest/Portfolio
- Results:

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome	date1
4/6/2010	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)	None
8/12/2010	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)	None
22-05-2012	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt	None
8/10/2012	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt	None
1/3/2013	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt	None

Launch Site Names Begin with 'KSC': 5 records where launch sites begin with `KSC`

- %sql select * from SPACEXKAT where LAUNCH_SITE like 'KSC%' limit 5
- **Explanation:** selected results represent launch sites that begin with KSC, top 5 listed
- https://github.com/SKatest/Portfolio
- Results:

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome	date1
19-02-2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)	None
16-03-2017	6:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attempt	None
30-03-2017	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)	None
1/5/2017	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)	None
15-05-2017	23:21:00	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070	GTO	Inmarsat	Success	No attempt	None

Total Payload Mass:
Display the total payload mass carried by boosters launched by NASA (CRS)

- %sql select sum(PAYLOAD_MASS__KG_) from SPACEXKAT where CUSTOMER = 'NASA (CRS)‘
- **Explanation: r**epresents a sum of payload mass in kg for NASA (CRS) customer
- https://github.com/SKatest/Portfolio

- Results:

1
45596

Average Payload Mass by F9 v1.1

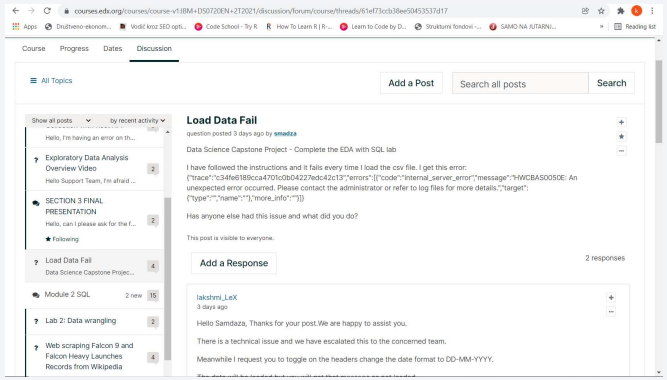
- %sql select avg(PAYLOAD_MASS__KG_) from SPACEXKAT where BOOSTER_VERSION = 'F9 v1.1‘
- **Explanation: r**epresents an average of payload mass for booster version F9 v1.1. from Booster version column
- https://github.com/SKatest/Portfolio
- Results:

1
2928

First Successful Ground Landing Date:
Listed the date where the first succesful landing outcome in drone ship was acheived

- %sql select min(DATE) from SPACEXKAT where Landing__Outcome = 'Success (drone ship)‘
- Result: 8/4/2016
- **Explanation:** it represents the oldest date when the outcome was marked as Succes (drone ship)
- https://github.com/SKatest/Portfolio(query correct, result not loaded well by column in sql, I apply excel result here)
- There was a problem to import DATE column as date, so queries with DATE column, even though I have changed the date format to DD-MM-YYYY, daTa couldn’t be loaded as DATE datatype, I improted it as a varchar to be able to use the table for queries.
- This is why I put here the query I would use in SQL, show the results in Excel and I add here a link of a discussion about the problem with date column from this course discussion: https://courses.edx.org/courses/course-v1:IBM+DS0720EN+2T2021/discussion/forum/course/threads/61ef73ccb38ee50453537d17

Date	Time (UTC	Booster_V	Launch_Sit	Payload	PAYLOAD_	Orbit	Customer	Mission_O	Landing _Outcome
8/4/2016	20:43:00	F9 FT B102	CCAFS LC-4	SpaceX CR	3136	LEO (ISS)	NASA (CRS	Success	Success (drone ship)
6/5/2016	5:21:00	F9 FT B102	CCAFS LC-4	JCSAT-14	4696	GTO	SKY Perfec	Success	Success (drone ship)
27-05-2016	21:39:00	F9 FT B102	CCAFS LC-4	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
14-08-2016	5:26:00	F9 FT B102	CCAFS LC-4	JCSAT-16	4600	GTO	SKY Perfec	Success	Success (drone ship)
14-01-2017	17:54:00	F9 FT B102	VAFB SLC-4	Iridium NE	9600	Polar LEO	Iridium Co	Success	Success (drone ship)
30-03-2017	22:27:00	F9 FT B10	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
23-06-2017	19:10:00	F9 FT B10	KSC LC-39A	BulgariaSa	3669	GTO	Bulsatcom	Success	Success (drone ship)
25-06-2017	20:25:00	F9 FT B103	VAFB SLC-4	Iridium NE	9600	LEO	Iridium Co	Success	Success (drone ship)
24-08-2017	18:51:00	F9 FT B103	VAFB SLC-4	Formosat-	475	SSO	NSPO	Success	Success (drone ship)
9/10/2017	12:37:00	F9 B4 B104	VAFB SLC-4	Iridium NE	9600	Polar LEO	Iridium Co	Success	Success (drone ship)
11/10/2017	22:53:00	F9 FT B10	KSC LC-39A	SES-11 / Ec	5200	GTO	SES EchoSt	Success	Success (drone ship)
30-10-2017	19:34:00	F9 B4 B104	KSC LC-39A	Koreasat 5	3500	GTO	KT Corpor	Success	Success (drone ship)
18-04-2018	22:51:00	F9 B4 B104	CCAFS SLC	Transiting	362	HEO	NASA (LSP	Success	Success (drone ship)
11/5/2018	20:14:00	F9 B5 B10	KSC LC-39A	Bangabanc	3600	GTO	Thales-Ale	Success	Success (drone ship)



Successful Drone Ship Landing with Payload between 4000 and 6000: Listed names of the boosters which have success in ground pad and have payload mass greater than 4000 but less than 6000

- %sql select BOOSTER_VERSION from SPACEXKAT where Landing__Outcome = 'Success (ground pad)' and PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000

- Results:

booster_version
F9 FT B1032.1
F9 B4 B1040.1
F9 B4 B1043.1

- **Explanation:** represents Booster versions which have payload mass between, but not included 4000 and 6000 kg AND have an outcome as Success (ground pad)

- https://github.com/SKatest/Portfolio

Total Number of Successful and Failure Mission Outcomes

- %sql SELECT COUNT(MISSION_OUTCOME) AS "Successful" FROM SPACEXKAT WHERE MISSION_OUTCOME LIKE '%Success%';
- Result:

Successful
100

- %sql SELECT COUNT(MISSION_OUTCOME) AS "Failure" FROM SPACEXKAT WHERE MISSION_OUTCOME LIKE '%Failure%‘
- Result:

Failure
1

- **Explanation: r**epresents the count as a number od total Mission outcomes that were names as successful with or without extension in the name and also as Failure with or without extension in a name
- https://github.com/SKatest/Portfolio

Boosters Carried Maximum Payload

- %sql SELECT BOOSTER_VERSION FROM SPACEXKAT WHERE PAYLOAD_MASS__KG_ =(SELECT MAX(PAYLOAD_MASS__KG_) FROM SPACEXKAT)
- **Explanation: r**epresents Booster versions that have all carried the same and maximum payload mass ; Results:

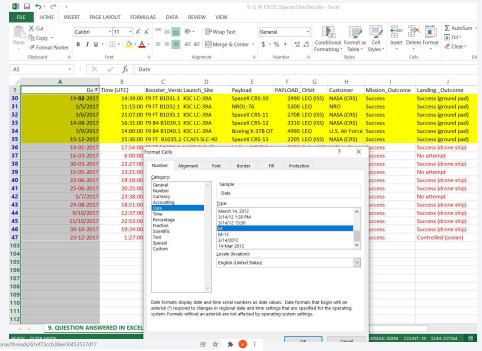
booster_version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

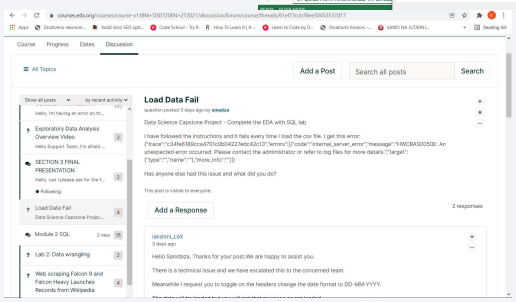
- (Interesting: it was 15600 kg)
- https://github.com/SKatest/Portfolio

2017 Launch Records:
Listed records which will display the month names, succesful landing_outcomes in ground pad ,booster versions, launch_site for the months in year 2017

- %sql SELECT month(DATE) as Month, Landing__Outcome, BOOSTER_VERSION, LAUNCH_SITE FROM SPACEXKAT WHERE year(DATE) = '2017' AND LANDING__OUTCOME = 'Success (ground pad)'
- **Explanation: r**epresents information about months, Success (ground pad), Booster version and launch sites selected by the months that belong to 2017.
- https://github.com/SKatest/Portfolio(query correct, result not loaded well by column in sql, I apply excel result here)
- There was a problem to import DATE column as date, so queries with DATE column, even though I have changed the date format to DD-MM-YYYY, daTa couldn't be loaded as DATE datatype, I improted it as a varchar to be able to use the table for queries.
- This is why I put here the query I would use in SQL, show the results in Excel and I add here a link of a discussion about the problem with date column from this course discussion: https://courses.edx.org/courses/course-v1:IBM+DS0720EN+2T2021/discussion/forum/course/threads/61ef73ccb38ee50453537d17
- Results:

Da	Time (UTC)	Booster_Versi	Launch_Site	Payload	PAYLOAD_	Orbit	Customer	Mission_Outcome	Landing _Outcome
19-**02**-2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
1/5/2017	11:15:00	F9 FT B1032.1	KSC LC-39A	NROL-76	5300	LEO	NRO	Success	Success (ground pad)
3/6/2017	21:07:00	F9 FT B1035.1	KSC LC-39A	SpaceX CRS-11	2708	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
14-08-2017	16:31:00	F9 B4 B1039.1	KSC LC-39A	SpaceX CRS-12	3310	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
7/9/2017	14:00:00	F9 B4 B1040.1	KSC LC-39A	Boeing X-37B OT	4990	LEO	U.S. Air Force	Success	Success (ground pad)
15-12-2017	15:36:00	F9 FT B1035.2	CCAFS SLC-40	SpaceX CRS-13	2205	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
14-01-2017	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Comm	Success	Success (drone ship)
16-03-2017	6:00:00	F9 FT B1030	KSC LC-39A	EchoStar 23	5600	GTO	EchoStar	Success	No attempt
30-03-2017	22:27:00	F9 FT B1021.2	KSC LC-39A	SES-10	5300	GTO	SES	Success	Success (drone ship)
15-05-2017	23:21:00	F9 FT B1034	KSC LC-39A	Inmarsat-5 F4	6070	GTO	Inmarsat	Success	No attempt
23-06-2017	19:10:00	F9 FT B1029.2	KSC LC-39A	BulgariaSat-1	3669	GTO	Bulsatcom	Success	Success (drone ship)
25-06-2017	20:25:00	F9 FT B1036.1	VAFB SLC-4E	Iridium NEXT 2	9600	LEO	Iridium Comm	Success	Success (drone ship)
5/7/2017	23:38:00	F9 FT B1037	KSC LC-39A	Intelsat 35e	6761	GTO	Intelsat	Success	No attempt
24-08-2017	18:51:00	F9 FT B1038.1	VAFB SLC-4E	Formosat-5	475	SSO	NSPO	Success	Success (drone ship)
9/10/2017	12:37:00	F9 B4 B1041.1	VAFB SLC-4E	Iridium NEXT 3	9600	Polar LEO	Iridium Comm	Success	Success (drone ship)
11/10/2017	22:53:00	F9 FT B1031.2	KSC LC-39A	SES-11 / EchoSta	5200	GTO	SES EchoStar	Success	Success (drone ship)
30-10-2017	19:34:00	F9 B4 B1042.1	KSC LC-39A	Koreasat 5A	3500	GTO	KT Corporatio	Success	Success (drone ship)
23-12-2017	1:27:00	F9 FT B1036.2	VAFB SLC-4E	Iridium NEXT 4	9600	Polar LEO	Iridium Comm	Success	Controlled (ocean)





Rank Landing Outcomes Between 2010-06-04 and 2017-03-20:
Rank the count of successful landing_outcomes between the date 2010-06-04 and 2017-03-20 in descending order.

- %sql SELECT COUNT(LANDING__OUTCOME) FROM SPACEXKAT WHERE LANDING__OUTCOME LIKE '%Success%' AND DATE > '2010-06-04' AND DATE < '2017-03-20' ORDER BY DATE DESC: 5,3

- **Explanation: r**epresents the number of landing outcomes that were marked as success with or no additional explanation and happend between 4 June 2010 and 20 March 2017 from bigger count (**5 for Succcess (drone ship) and 3 for Success (ground pad)**

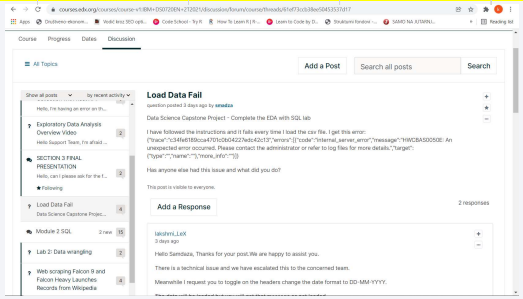
- https://github.com/SKatest/Portfolio(query correct, result not loaded well by column in sql, I apply excel result here)

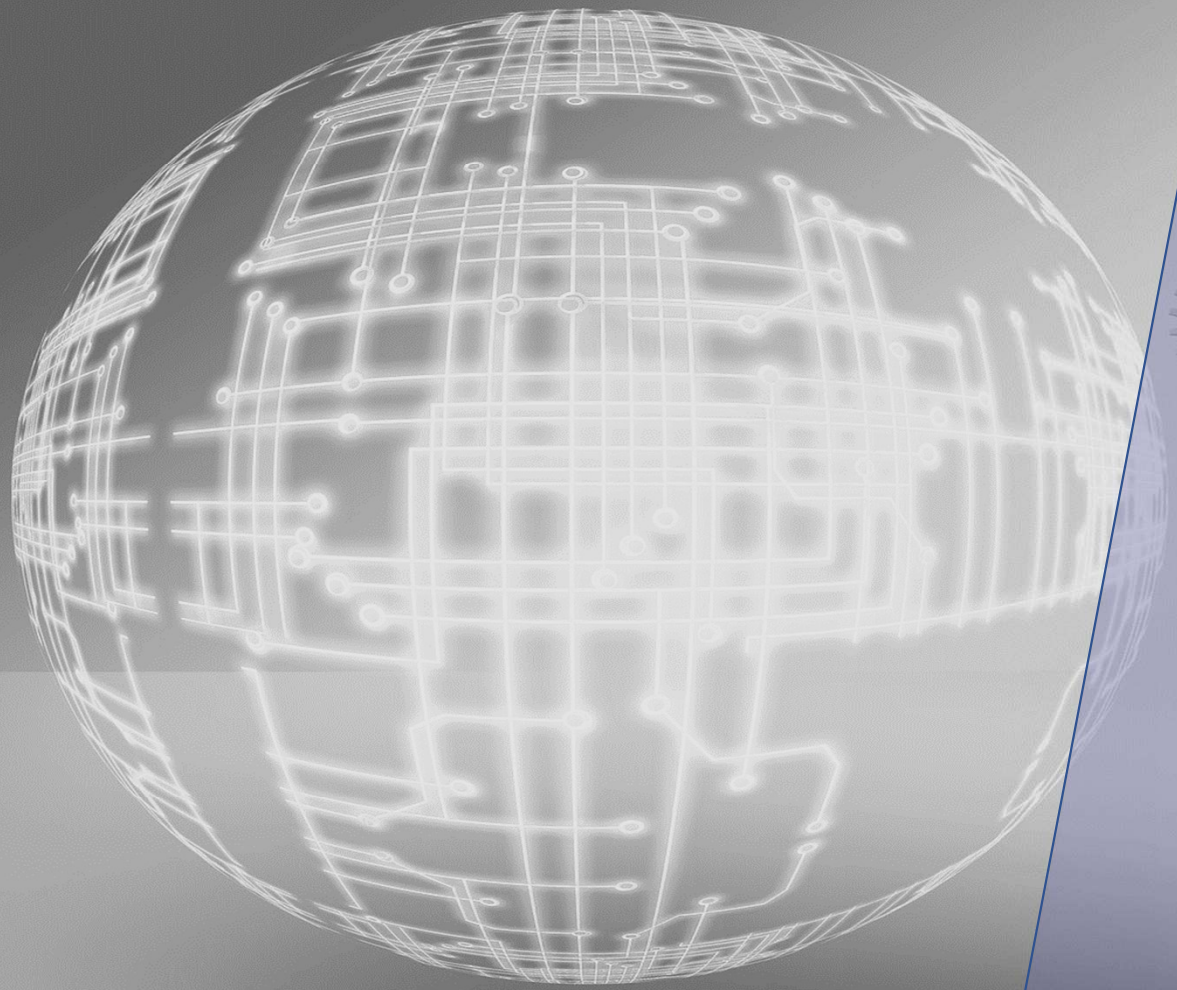
- There was a problem to import DATE column as date, so queries with DATE column, even though I have changed the date format to DD-MM-YYYY, daTa couldn't be loaded as DATE datatype, I improted it as a varchar to be able to use the table for queries.

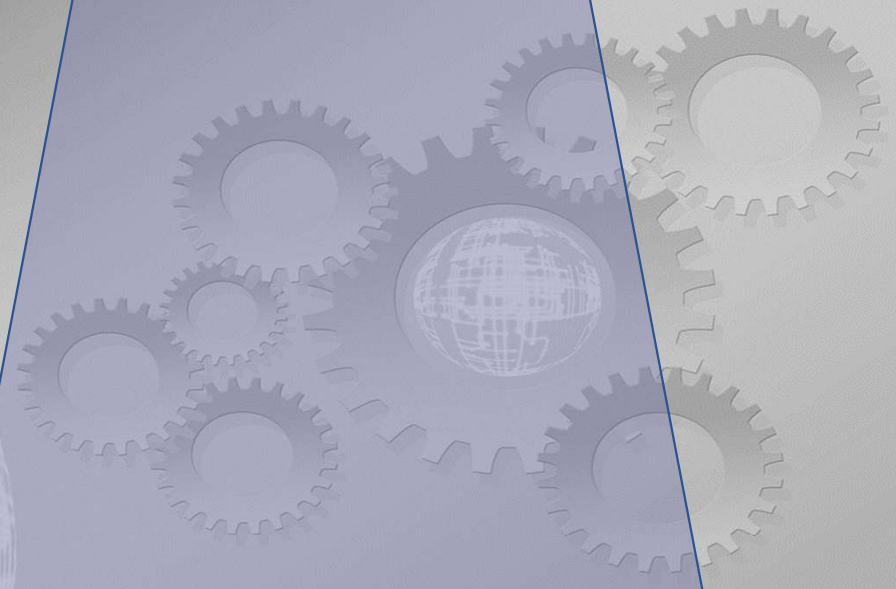
- This is why I put here the query I would use in SQL, show the results in Excel and I add here a link of a discussion about the problem with date column from this course discussion: https://courses.edx.org/courses/course-v1:IBM+DS0720EN+2T2021/discussion/forum/course/threads/61ef73ccb38ee50453537d17

- Results:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__	Orbit	Customer	Mission_Outcome	Landing _Outcome
8/4/2016	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
6/5/2016	5:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Grou	Success	Success (drone ship)
27-05-2016	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
14-08-2016	5:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Grou	Success	Success (drone ship)
14-01-2017	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communicatic	Success	Success (drone ship)
22-12-2015	1:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Ort	2034	LEO	Orbcomm	Success	Success (ground pad)
18-07-2016	4:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
19-02-2017	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)







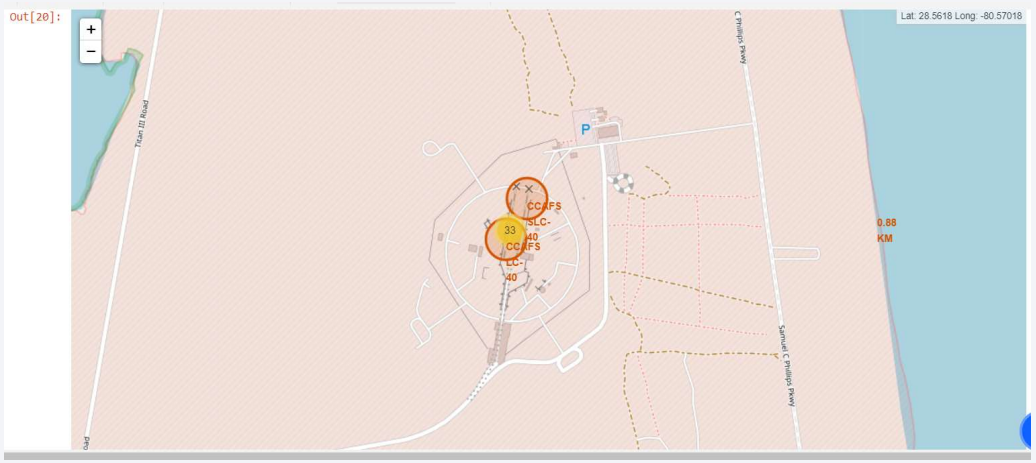
Section 3

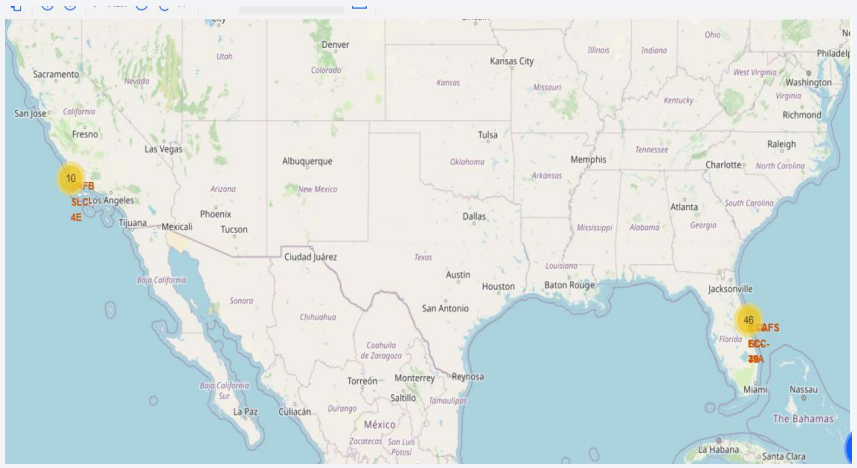
Launch Sites Proximities Analysis

Folium Map Screenshot 1: SpaceX launch sites' locations

- **Findings:**
- Launch sites are located in clusters in Florida and California, most probably because od the logistics reasons and safety, close to ocean
- Launch sites are in proximity close to the equator. I found out when searching about this, that it is because ˝ rockets launched from sites near the Equator get an additional natural boost that helps save the cost of putting in extra fuel and boosters˝ (source: https://www.scienceabc.com/eyeopeners/why-are-rockets-launched-from-areas-near-the-equator.html)

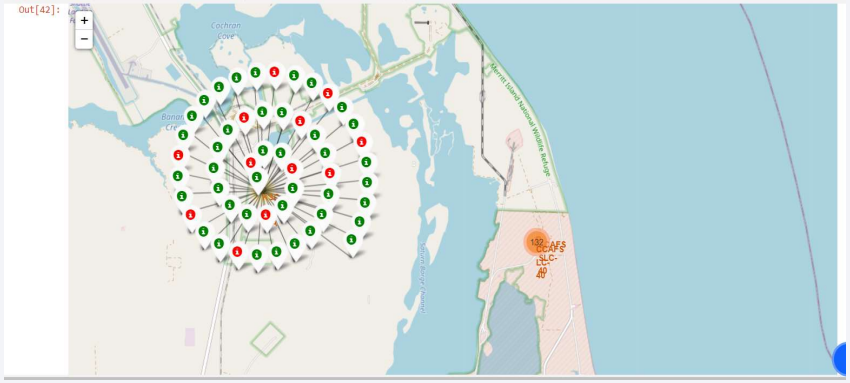
- https://github.com/SKatest/Portfolio

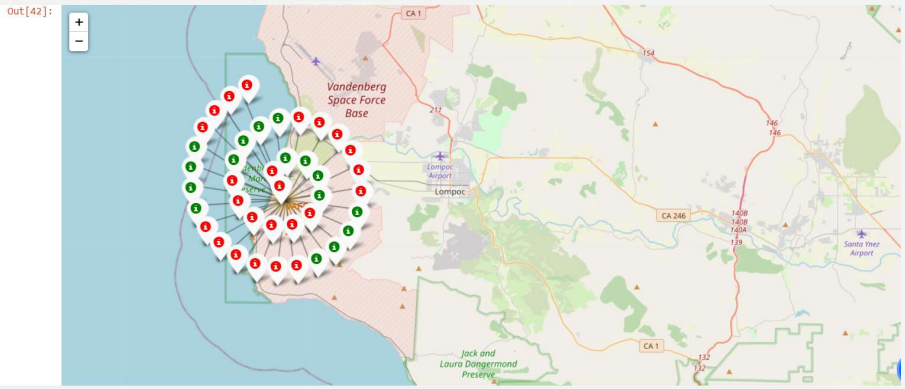


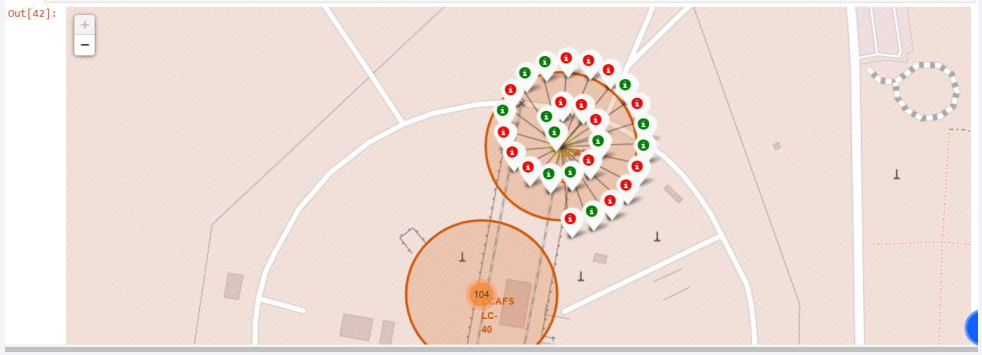


Folium Map Screenshot 2: Color-labeled launch outcomes

- KSC LC and CCAFS LC 40 show a huge number of green markers, that represent successful outcomes
- https://github.com/SKatest/Portfolio

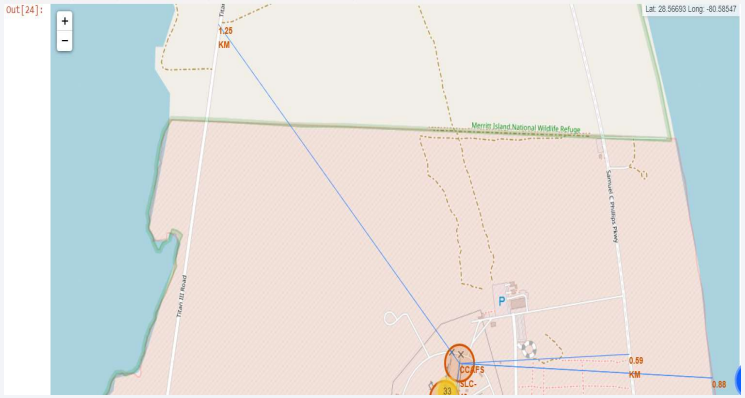


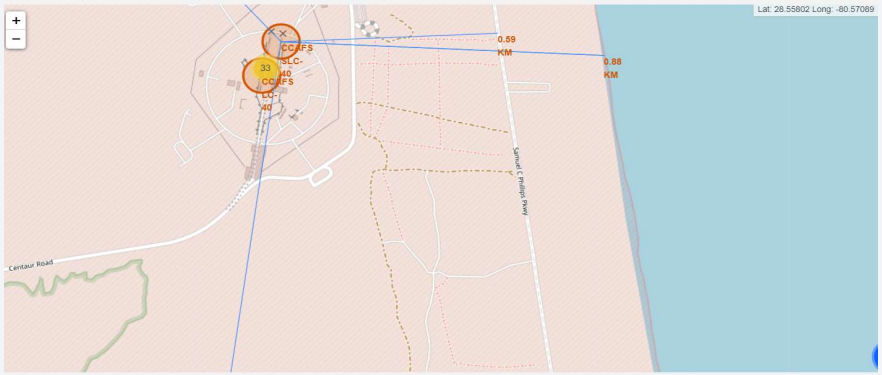




Folium Map Screenshot 3: Launch sites distance to its proximities: railway, highway, coastline

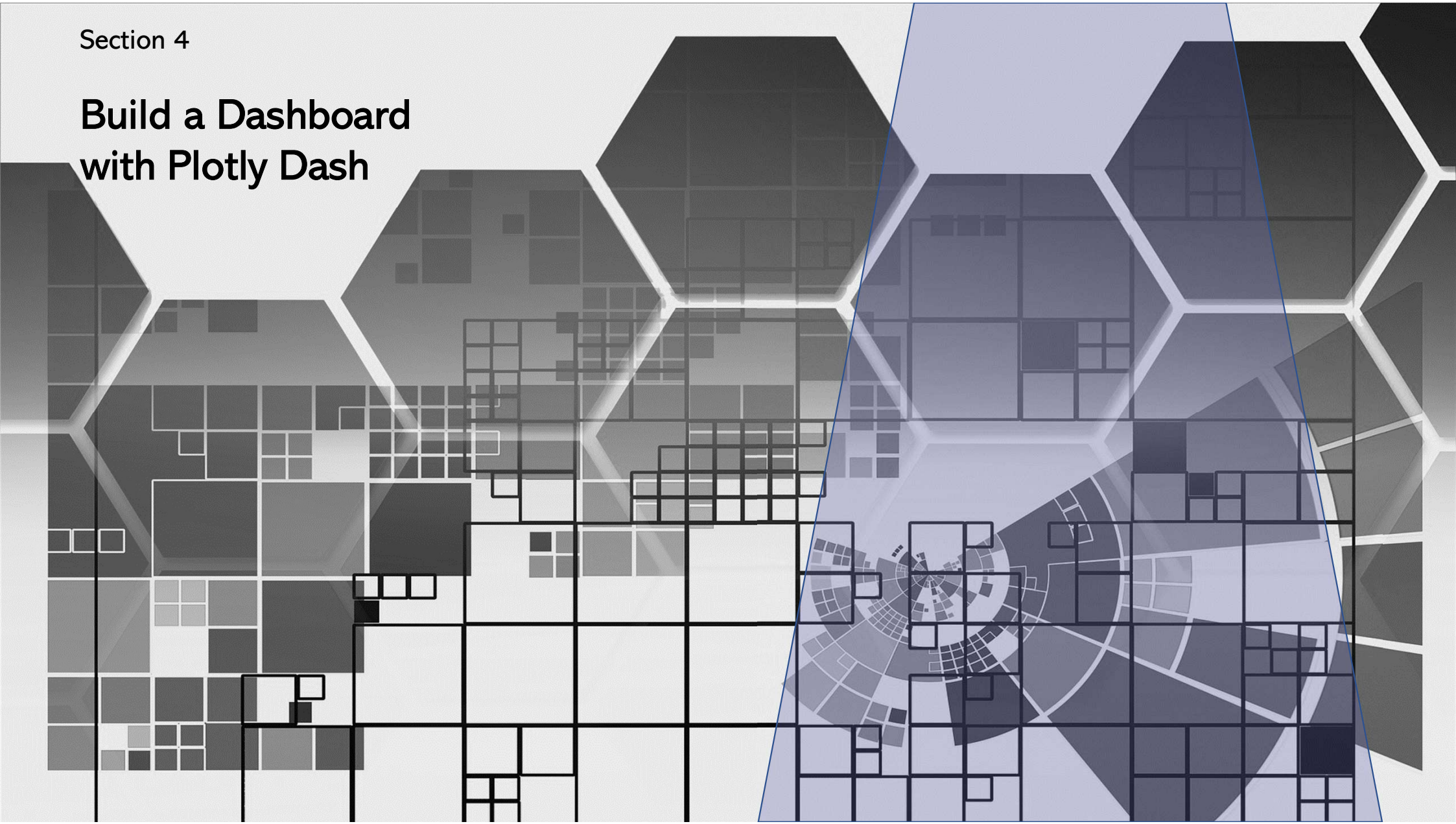
- distance_highway = 0.5876524983627341 km
- distance_railway = 1.247858808511078 km
- distance_city = 51.415077660301016 km
- closest coastline: 0.8752468484961807 km
- distances from launch sites to:
  highway, railway and coastline are not big because of the necessary
- conditions for launching: to be able to land in the water (ocean)
  and to be able to bring the necessery equipment and the Flacon 9 itself.
- distance from the closest city is bigger because of lower danger of hurting numerous people
  if something goes wrong with eather launching or landing.
- https://github.com/SKatest/Portfolio





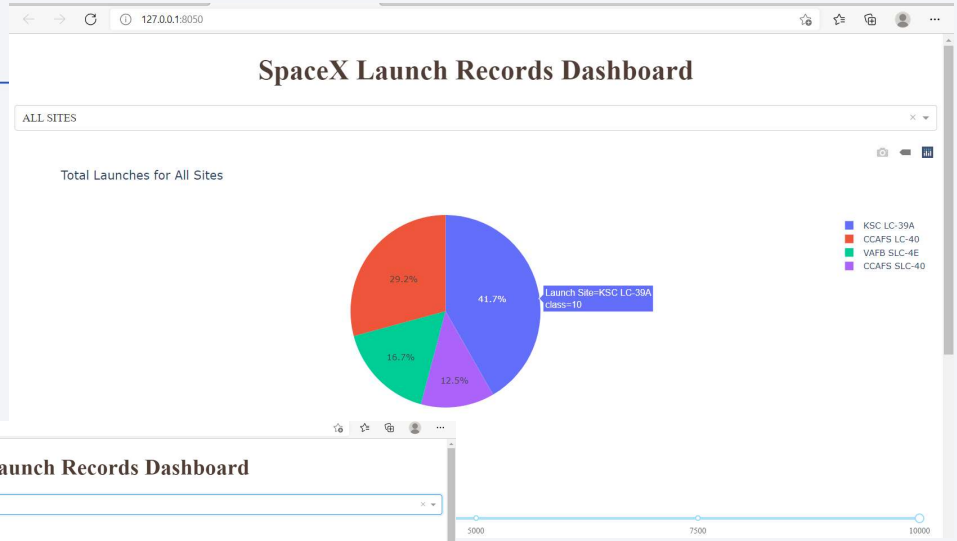
Section 4

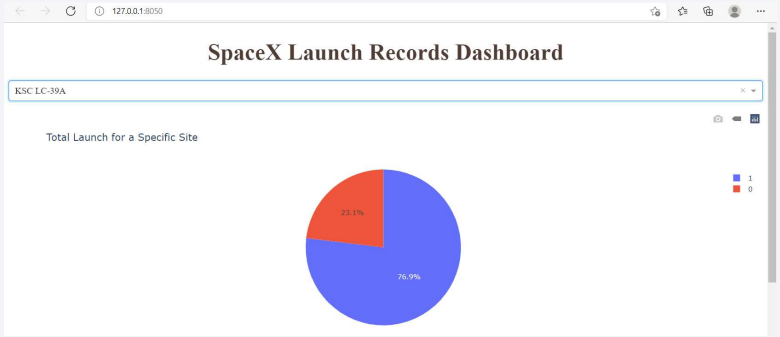
Build a Dashboard with Plotly Dash

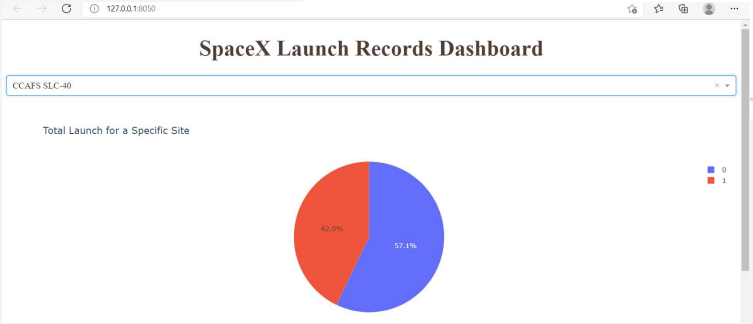


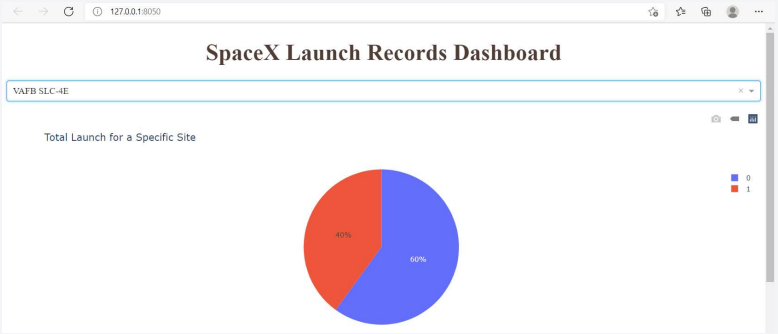
Dashboard Screenshot: Launch success count for all sites

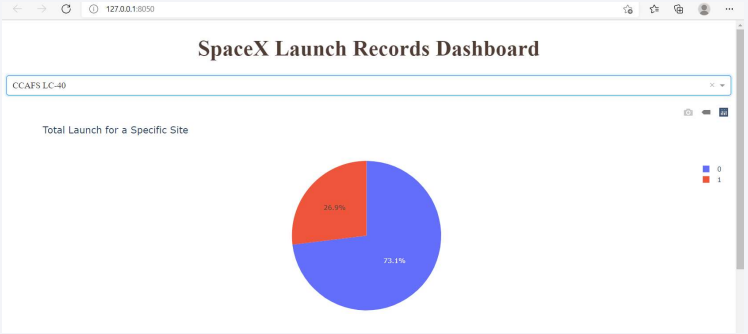
- KSC LC 39A is a launch site with the highest launch success ratio
- CCAFS LS40 has just a little smaller ratio
- VAFB SLC 4E has a smaller ratio
- CCAFS SLC40 shows the lowest ratio
- https://github.com/SKatest/Portfolio





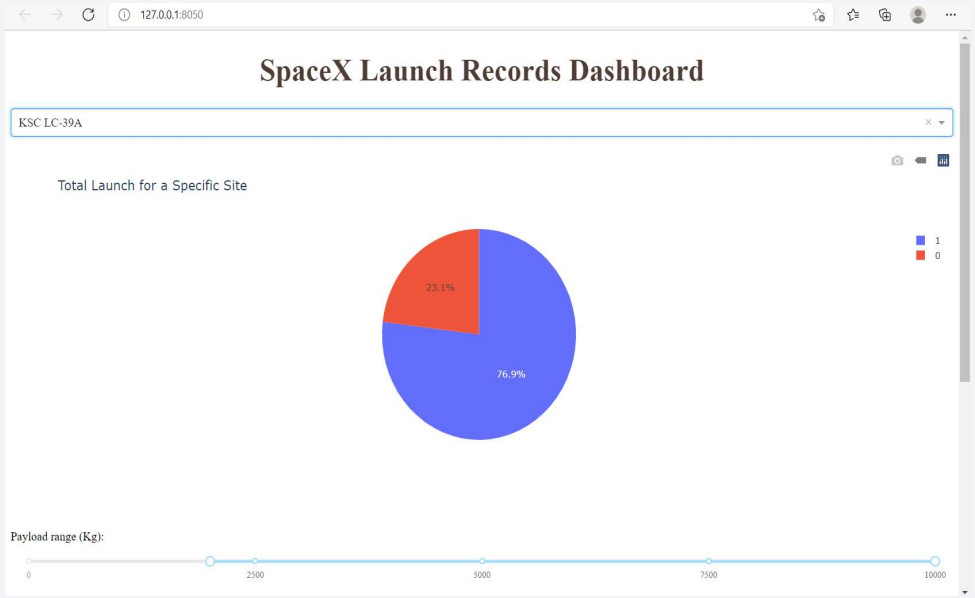






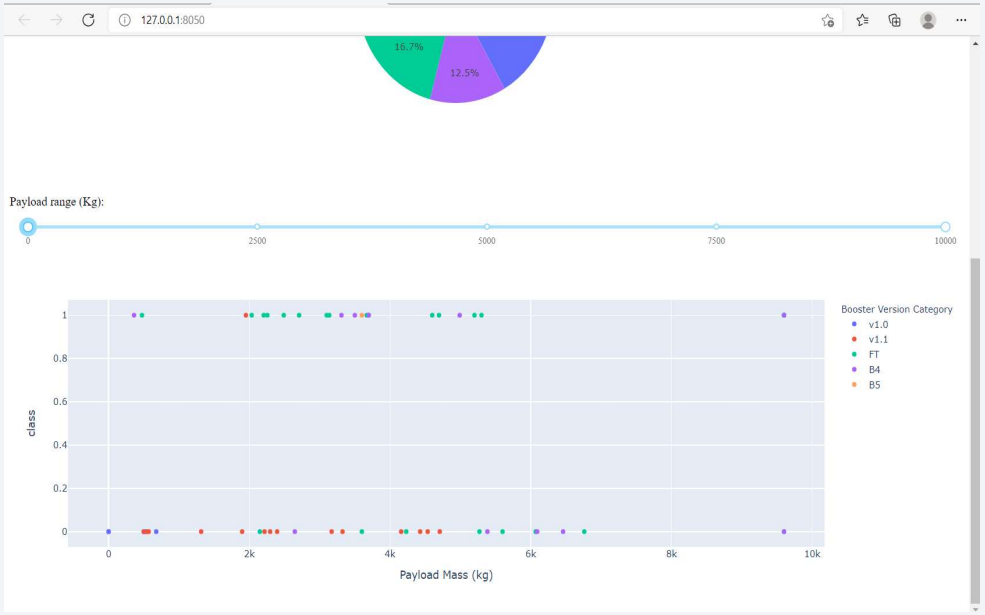
Dashboard Screenshot 2: Launch site with highest launch success ratio

- KSC LC 39A is a launch site with the highest launch success ratio
- https://github.com/SKatest/Portfolio



Dashboard Screenshot 3: Payload vs. Launch Outcome

- Dash: bigger payload mass makes it more difficult to reach a successful landing outcome
- https://github.com/SKatest/Portfolio



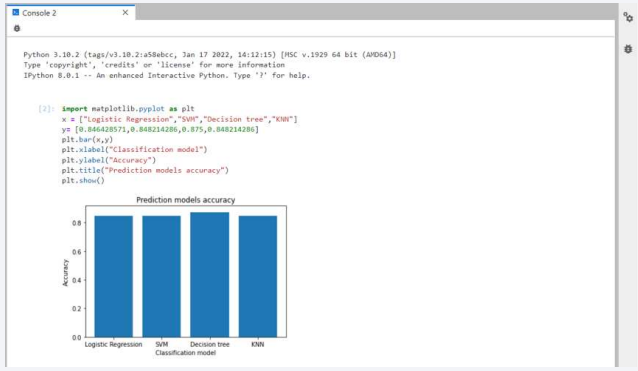
Section 5

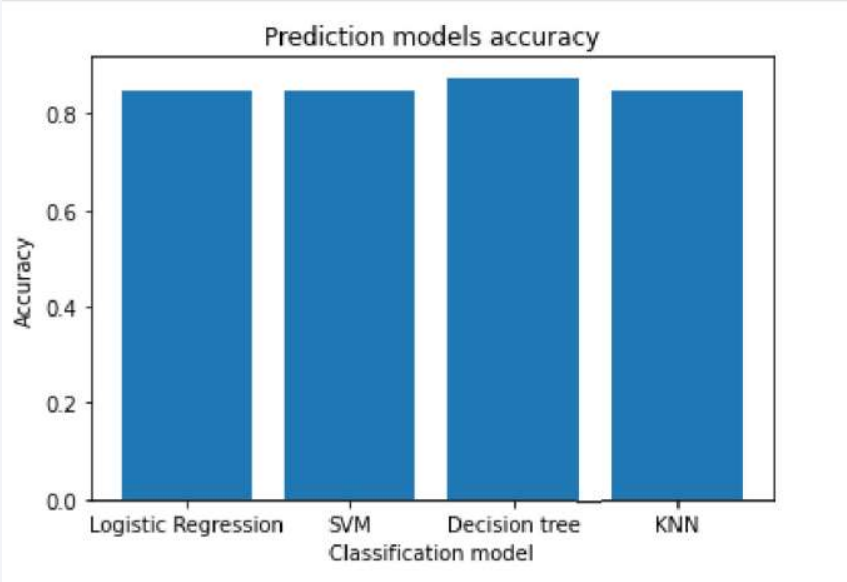
Predictive Analysis (Classification)

Classification Accuracy

- **Predictive analysis results:**
- there were 18 records in the test sample
- the best algorithm to predict landing outcome is DECISION TREE, with the accuracy of 87,5%
- the mean of train and test best accuracy of algorithms is also highest for Decision tree because of 83.33 % of test accuracy of tests and highest train test accuracy of Decision tree model
- after selecting the best hyperparameters, the accuracy achieved on the test data for Decision tree is 83,33%

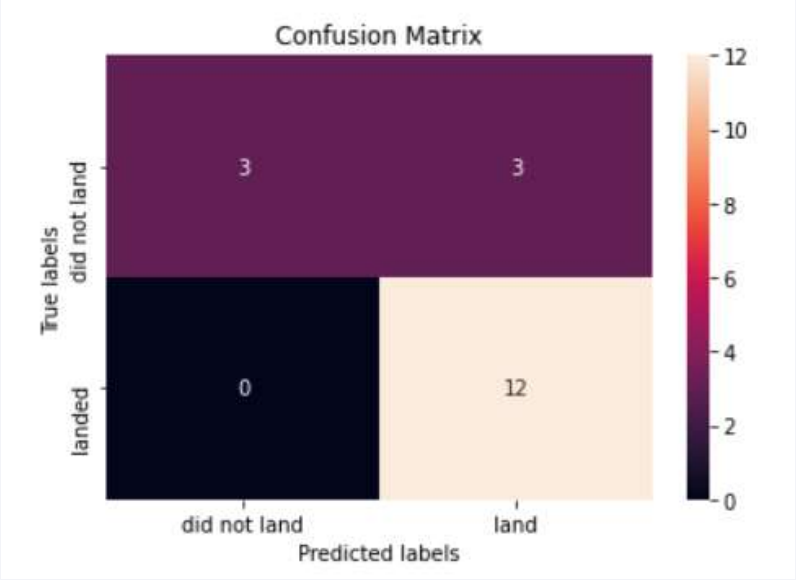
- https://github.com/SKatest/spaceXkat/blob/master/PREDICTIVE%20ANALYSIS%20CLASSIFICATION.ipynb
- https://github.com/SKatest/Portfolio





Confusion Matrix

- confusion matrix shows that Falcon 9 will successfully land with classifier that predicted correctly 12/15 successfully landed outcomes and 3/3 unsuccessful landing outcomes. Out of 18 observed outcomes, 12 were successful in reality
- https://github.com/SKatest/Portfolio



Conclusions

- It can be predicted that **Falcon 9 first stage will land successfully**
- Confusion matrix for the test set shows that 12 out of 15 launches have successfully landed in reality, which represents 66,77% of success

If you could go to space, would you?

- I have chosen to show **a fun and interesting questioning found on Quora** (source:https://www.quora.com/If-you-could-go-to-space-would-you**):**

  ˝**If you colud go to space, would you?**˝

- This is such an interesting Quora question, because while reading the answers one can see how interesting this topic is even for those with negative answers!
- It is also interesting because the price variable is put out of sight as a problem and questioning is based on the opinion about the idea of the space trip itself.
- I believe for the Company' s board would be very interesting to have an insight in the ˝public˝ opinion about this question because:
- they can find out what majority of answers currently sounds like.. yes or no..//1 or 0;
- what are the worries of people who wouldn't go; what are the reasons to go;
- what are the variables that affect a decision, for example: safety of trip, people's age, etc.;what can be done for people to choose to try if they once really could;
- what are the variables that company needs to put emphasis on already while developing the story, in order to make more people have a positive opinion and even a positive decision to go to space, if they could;
- It is possible that people who are willing to go were more attracted to this question, but since the negative answers were also very insightful, I decided just to check on the ratio of positive and negative answers.
- I took randomly 30 answers and put aside ups on every question, just for basic research and to put the same value on each positive and negative answer.
- Result: out of 30 randomly picked answers, 3 were negative and 27 positive! ; Professional research would of course need more groups of randomly interviewed people, not only people who decide to deal with this topic. Still, I find this questioning very fun and interesting and in real life situation, I believe the results of bigger, detailed research would be very interesting for marketing and financial department and management board of the company, especially if we look very fast forward to future and to promotion of commercial flights or maybe even Space Tourism.

Appendix

- https://scikit-learn.org/stable/auto_examples/model_selection/plot_grid_search_digits.html#sphx-glr-auto-examples-model-selection-plot-grid-search-digits-py
- https://plotly.com/python/getting-started/#jupyterlab-support
- https://courses.edx.org/courses/course-v1:IBM+DS0720EN+2T2021/discussion/forum/course/threads/61ef73ccb38ee50453537d17
- https://pypi.org/project/ipython-sql/ ; RDBMS access via IPython
- https://notebook.community/DB2-Samples/db2jupyter/Db2%20Jupyter%20Extensions%20Tutorial
- https://python-visualization.github.io/folium/modules.html
- https://dash.plotly.com/
- https://dash.plotly.com/dash-core-components
- https://plotly.com/python/line-and-scatter/
- https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.iterrows.html
- https://community.plotly.com/t/dash-plotly-csv-to-pie-chart-using-slider-callback/33552/2
- https://developers.google.com/chart/interactive/docs/gallery/controls
- https://plotly.com/python/line-and-scatter/
- https://dash.plotly.com/dash-core-components
- https://youtu.be/mTsZL-VmRVE
- https://python-visualization.github.io/folium/quickstart.html
- https://pixabay.com/
- https://www.quora.com/If-you-could-go-to-space-would-you
- www.edx.org

Thank you!