

IBM Data Science

M11: Applied Data Science Capstone [S1]

Technical Session 44

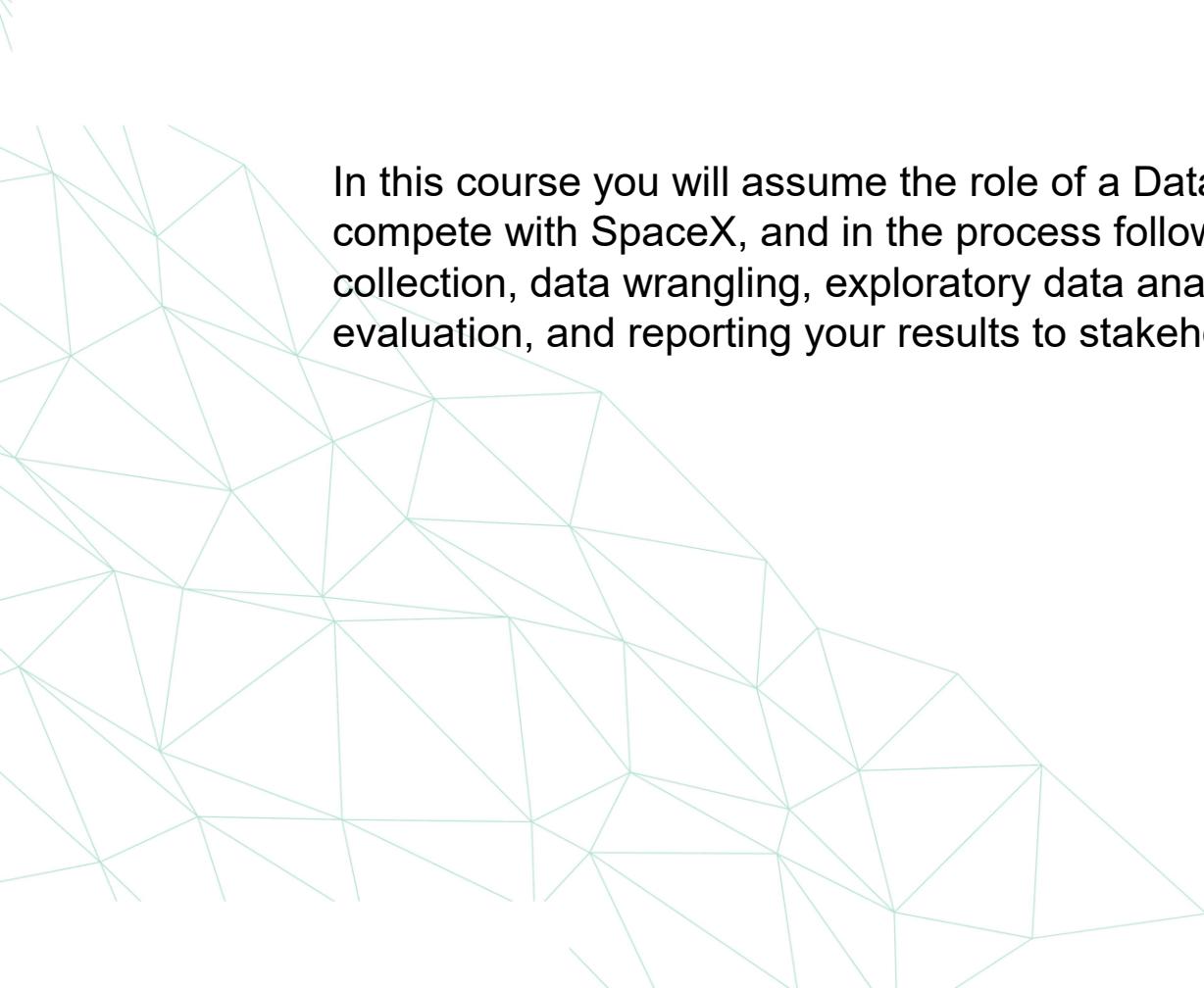


IBM Applied Data Science Capstone

07/05/2025



Course Overview



In this course you will assume the role of a Data Scientist working for a startup intending to compete with SpaceX, and in the process follow the Data Science methodology involving data collection, data wrangling, exploratory data analysis, data visualization, model development, model evaluation, and reporting your results to stakeholders.



- Benefits of Enrolling in a Course:

- Demonstrate proficiency in data science and machine learning techniques using a real-world data set and prepare a report for stakeholders.
- Apply your skills to perform data collection, data wrangling, exploratory data analysis, data visualization model development, and model evaluation.
- Write Python code to create machine learning models including support vector machines, decision tree classifiers, and k-nearest neighbours.
- Evaluate the results of machine learning models for predictive analysis, compare their strengths and weaknesses and identify the optimal model.

- Course Modules:

- Introduction
- Exploratory Data Analysis
- Interactive Visual Analytics and Dashboards
- Predictive Analysis
- Present your data-driven insights.

Module I

Introduction

Module I

Introduction

Course Introduction

Course Introduction

- Project Overview:
 - Apply data science skills for a space launch company.
- Data Collection and Wrangling:
 - Collect relevant data from various sources.
 - Improve data quality through wrangling processes.
- Data Exploration:
 - Explore real-world datasets together.
 - Practice SQL skills for querying and insights.
- Statistical Analysis and Visualization:
 - Apply basic statistical analysis and data visualization.
 - Understand variable relationships.
- Predictive Modeling and Presentation:
 - Build, evaluate, and refine predictive models.
 - Create a presentation summarizing the analysis.

Module I Introduction

Project Scenario and Overview

Project Scenario and Overview - Video



Project Overview

- Space Industry Overview:
 - Introduction to major players like Virgin Galactic, Rocket Lab, Blue Origin, and SpaceX.
 - Highlight SpaceX's notable achievements such as sending spacecraft to the ISS, Starlink satellite internet constellation, and manned space missions.
- Cost Efficiency of SpaceX:
 - Provide specific cost comparisons between SpaceX's Falcon 9 launches (\$62 million) and other providers (upwards of \$165 million).
 - Emphasize the cost-saving aspect due to the reusability of the first stage, a key factor in SpaceX's success.
- Falcon 9 Rocket Anatomy:
 - Detail the importance of the first stage in the launch process and its substantial size compared to the second stage.
 - Discuss challenges related to first stage recovery and how mission parameters affect reuse decisions.

Capstone Project Details

- Role in Capstone Project:
 - Describe the role of the data scientist for Space Y, a company competing with SpaceX, founded by industrialist Alon Musk.
 - Highlight the primary tasks: determining launch prices, analyzing first stage reuse feasibility.
- Dashboards Creation:
 - Explain the process of gathering crucial information about SpaceX's operations and market dynamics.
 - Emphasize the creation of interactive dashboards for team analysis, aiding decision-making.
- Machine Learning Approach:
 - Outline the use of machine learning to predict first stage reuse based on public data.
 - Discuss the significance of this predictive model in cost estimation and strategic planning for Space Y.

Lab 1 Introduction

Data Collection Overview



SpaceX REST API



SpaceX REST API

Open Source REST API for launch, rocket, core, capsule, starlink, launchpad, and landing pad data.

[build](#) [passing](#) [docker pulls](#) [2.1M](#) [release](#) [v4.0.0](#) [interface](#) [REST](#)

We are not affiliated, associated, authorized, endorsed by, or in any way officially connected with Space Exploration Technologies Corp (SpaceX), or any of its subsidiaries or its affiliates. The names SpaceX as well as related names, marks, emblems and images are registered trademarks of their respective owners.



<https://api.spacexdata.com/v4/>

api.spacexdata.com/v4/capsules

```
[{"reuse_count":0,"water_landings":1,"land_landings":0,"last_update":"Hanging in atrium at SpaceX HQ in Hawthorne","launches":["5eb87cdeffd86e000604b330"],"serial":"C101","status":"retired","type":"Dragon 1.0","id":"5e9e2c5bf35918ed873b2664"}, {"reuse_count":0,"water_landings":1,"land_
```

api.spacexdata.com/v4/cores

```
[{"block":null,"reuse_count":0,"rtls_attempts":0,"rtlts_landings":0,"asds_attempts":0,"asds_landings":0,"last_update":"Engine failure at T+33 seconds resulted in loss of vehicle","launches":["5eb87cd9ffd86e000604b32a"],"serial":"Merlin1A","status":"lost","id":..}
```

api.spacexdata.com/v4/launches/past

```
url="https://api.spacexdata.com/v4/launches/past"  
response =requests.get(url)
```



url="https://api.spacexdata.com/v4/launches/past"

```
response =requests.get(url)
```

```
response.json()
```

```
response.json()  
[{"fairings": {"reused": false,  
"recovery_attempt": false,  
"recovered": false,  
"ships": []},  
"links": {"patch": {"small": "https://images2.imgur.com/3c/0e/T8JcSN3_o.png",  
"large": "https://imgbox.com/40/e3/GypSkayF_o.png"},  
"reddit": {"campaign": None},  
"launch": None,  
"media": None,  
"recovery": None},  
"flickr": {"small": [], "original": []},  
"presskit": None,  
"webcast": "https://www.youtube.com/watch?v=0a_00nJ_Y88",  
"youtube_id": "0a_00nJ_Y88",  
"article": "https://www.space.com/2196-spacex-inaugural-falcon-1-rocket-lost-launch.html",  
"wikipedia": "https://en.wikipedia.org/wiki/DemoSat"},  
"static_fire_date_utc": "2006-03-24T00:00:00Z",  
"static_fire_date_unix": 1142533600,  
"tbd": False,  
"net": False,  
"window": 0,  
"rocket": "5e9d0d95eda69955f709d1eb",  
"success": false,  
"details": "Engine failure at 33 seconds and loss of vehicle",  
"crew": [],  
"ships": [],  
"capsules": [],  
"payloads": ["5eb0e4b5b6c3bb0006eeb1e1"],  
"launchpad": "Se9e4502f5090995de566f86",  
"auto_update": true,  
"failures": [{"time": 33},  
{"altitude": None},  
{"reason": "merlin engine failure"}],  
"flight_number": 1,  
"name": "FalconSat",  
"date_utc": "2006-03-24T22:30:00.000Z",  
"date_unix": 1143239400,  
"date_local": "2006-03-25T10:30:00+12:00",  
"date_precision": "hour",  
"upcoming": false,  
"cores": [{"core": "5e9e289df35918033d3b2623",  
"flight": 1,  
"gridfins": false,  
"leg": 1},  
{"reused": false,  
"landing_attempt": false,  
"landing_success": None,  
"landing_type": None,  
"landpad": None},  
{"id": "5eb87cd9fd86e000604b32a"},  
{"fairings": {"reused": false,  
"recovery_attempt": false,  
"recovered": false,  
"ships": []},  
"links": {"patch": {"small": "https://images2.imgur.com/4f/e3/I0lkuj2e_o.png",  
"large": "https://imgbox.com/be/e7/iNqsqVYh_o.png"},
```

Wrangling Data using an API

```
data = pd.json_normalize(response.json())
```

	static_fire_date_utc	static_fire_date_unix	tbd	net	window	rocket	success	details	crew	ships	capsules	payloads	launchpad	auto_update	failures	flight_number	name	date_utc	date_unix
0	2006-03-17T00:00:00.000Z	1.142554e+09	False	False	0.0	Se9d0d95eda69955f709d1eb	False	Engine failure at 33 seconds and loss of vehicle	[]	[]	[]	[Seb0e4b5b6c3bb0006eeb1e1]	Se9e4502f5090995de566f86	True	[{"time": 33, "altitude": 0, "reason": "None", "stage": 1}], [{"time": 33, "altitude": 0, "reason": "motor engine failure"}]	1	FalconSat	2006-03-17T22:30:00.000Z	1143239400
1	None	Nan	False	False	0.0	Se9d0d95eda69955f709d1eb	False	Successful first stage burn and transition to second stage maximum altitude 289 km. Failed to reach orbit. Failed to recover first stage	[]	[]	[]	[Seb0e4b6b6c3bb0006eeb1e2]	Se9e4502f5090995de566f86	True	[{"time": 301, "altitude": 289, "reason": "harmonic oscillation leading to premature engine shutdown"}]	2	DemoSat	2007-03-21T01:10:00.000Z	1174439400
2	None	Nan	False	False	0.0	Se9d0d95eda69955f709d1eb	False	Residual stage 1 thrust led to collision between stage 1 and stage 2	[]	[]	[]	[Seb0e4b6b6c3bb0006eeb1e3, Seb0e4b6b6c3bb0006eeb1e4]	Se9e4502f5090995de566f86	True	[{"time": 140, "altitude": 35, "reason": "residual stage-1 thrust led to collision between stage 1 and stage 2"}]	3	Trailblazer	2008-08-03T03:34:00.000Z	1217734440
3	20T00:00:00.000Z	1.221869e+09	False	False	0.0	Se9d0d95eda69955f709d1eb	True	Ratsat was carried to orbit on the first successful orbital launch of any privately funded and developed liquid-propelled	[]	[]	[]	[Seb0e4b7b6c3bb0006eeb1e5]	Se9e4502f5090995de566f86	True	[], [{"time": 140, "altitude": 35, "reason": "residual stage-1 thrust led to collision between stage 1 and stage 2"}]	4	RatSat	2008-09-28T23:15:00.000Z	1222643700

Web scraping Falcon 9 Launch records



Flight No.	Date and time (UTC)	Version, Booster ^[1]	Launch site	Payload ^[2]	Payload mass	Orbit	Customer	Launch outcome	Booster landing
78	7 January 2020, 02:13:21 ^[3]	F9 B1 v5.0 B1048.4	CCAFS, SLC-40	Starlink 2 v1.0 (60 satellites)	15,600 kg (34,400 lb) ^[3]	LEO	SpaceX	Success 	Success
79	19 January 2020, 15:30 ^[4]	F9 B1 v5.0 B1048.4	KSC, LC-39A	Crew Dragon In-flight abort test ^[5] (Dragon C206.1)	12,500 kg (28,070 lb)	Sub-orbital ^[6]	NASA (CTB) ^[7]	Success 	No attempt
80	29 January 2020, 14:05 ^[8]	F9 B1 v5.0 B1051.3	CCAFS, SLC-40	Starlink 3 v1.0 (60 satellites)	15,600 kg (34,400 lb) ^[3]	LEO	SpaceX	Success 	Success
81	17 February 2020, 04:50 ^[9]	F9 B1 v5.0 B1054.4	CCAFS, SLC-40	Starlink 4 v1.0 (60 satellites)	15,600 kg (34,400 lb) ^[3]	LEO	SpaceX	Success 	Failure
82	7 March 2020, 12:16 ^[10]	F9 B1 v5.0 B1059.2	CCAFS, SLC-40	SpaceX CRS-20 (Dragon C112.3 & Z)	1,977 kg (4,359 lb) ^[11]	LEO (ISS)	NASA (CRS)	Success 	Success
83	18 March 2020, 12:16 ^[12]	F9 B1 v5.0 B1048.5	KSC, LC-39A	Starlink 5 v1.0 (60 satellites)	15,600 kg (34,400 lb) ^[3]	LEO	SpaceX	Success 	Failure

In late 2019, Gwynne Shotwell stated that SpaceX hoped for as many as 24 launches for Starlink satellites in 2020^[4,13] in addition to 14 or 15 non-Starlink launches. At 26 launches, 13 of which for Starlink satellites, Falcon 9 had its most prolific year, and Falcon rockets were second most prolific rocket family of 2020, only behind China's Long March rocket family^[14].

An atmospheric test of the Dragon 2 abort system after Max Q. The capsule fired its SuperDraco engines, reached an apogee of 40 km (25 mi), deployed parachutes after reentry, and splashed down in the ocean 31 km (19 mi) downrange from the launch site. The test was previously slated to be accomplished with the Crew Dragon Demo-1 capsule^[15], but that test article exploded during a ground test of SuperDraco engines on 20 April 2019^[16]. The abort test — the capsule originally intended for the first crewed flight^[17] — as expected, the booster was destroyed by aerodynamic forces after the capsule aborted^[18]. First flight of a Falcon 9 with only one functional stage — the second stage had a mass simulator in place of its engine.

Web scraping with BeautifulSoup

FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1 2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin1A	167.743129	9.047721
1	2 2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2A	167.743129	9.047721
2	4 2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2C	167.743129	9.047721
3	5 2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin3C	167.743129	9.047721
4	6 2010-06-04	Falcon 9	NaN	LEO	CCAFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857

Data Wrangling Problems

- Wrangling Data using an API
- Sampling Data
- Dealing with Nulls



Wrangling Data using an API

Function	Targets	Endpoint
getBoosterVersion	Rockets	URL: https://api.spacexdata.com/v4/rock
getLaunchSite	Launchpads	URL: https://api.spacexdata.com/v4/laun
getPayloadData	Payloads	URL: https://api.spacexdata.com/v4/payl
getCoreData	getCoreData	URL: https://api.spacexdata.com/v4/core



Sampling Data

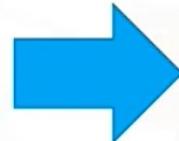
FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1 2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	Nan	0	Merlin1A	167.743129	9.047721
1	2 2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None	Nan	0	Merlin2A	167.743129	9.047721
2	4 2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	Nan	0	Merlin2C	167.743129	9.047721
3	5 2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	Nan	0	Merlin3C	167.743129	9.047721
4	6 2010-06-04	Falcon 9	NaN	LEO	CCAFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857



Dealing with Nulls

```
data_falcon9.isnull().sum()
```

```
FlightNumber      0
Date             0
BoosterVersion   0
PayloadMass      5
Orbit            0
LaunchSite       0
Outcome          0
Flights          0
GridFins         0
Reused           0
Legs              0
LandingPad      26
Block             0
ReusedCount     0
Serial            0
Longitude        0
Latitude         0
dtype: int64
```



```
data_falcon9.isnull().sum()
```

```
FlightNumber      0
Date             0
BoosterVersion   0
PayloadMass      0
Orbit            0
LaunchSite       0
Outcome          0
Flights          0
GridFins         0
Reused           0
Legs              0
LandingPad      26
Block             0
ReusedCount     0
Serial            0
Longitude        0
Latitude         0
dtype: int64
```

Lab 2 Introduction

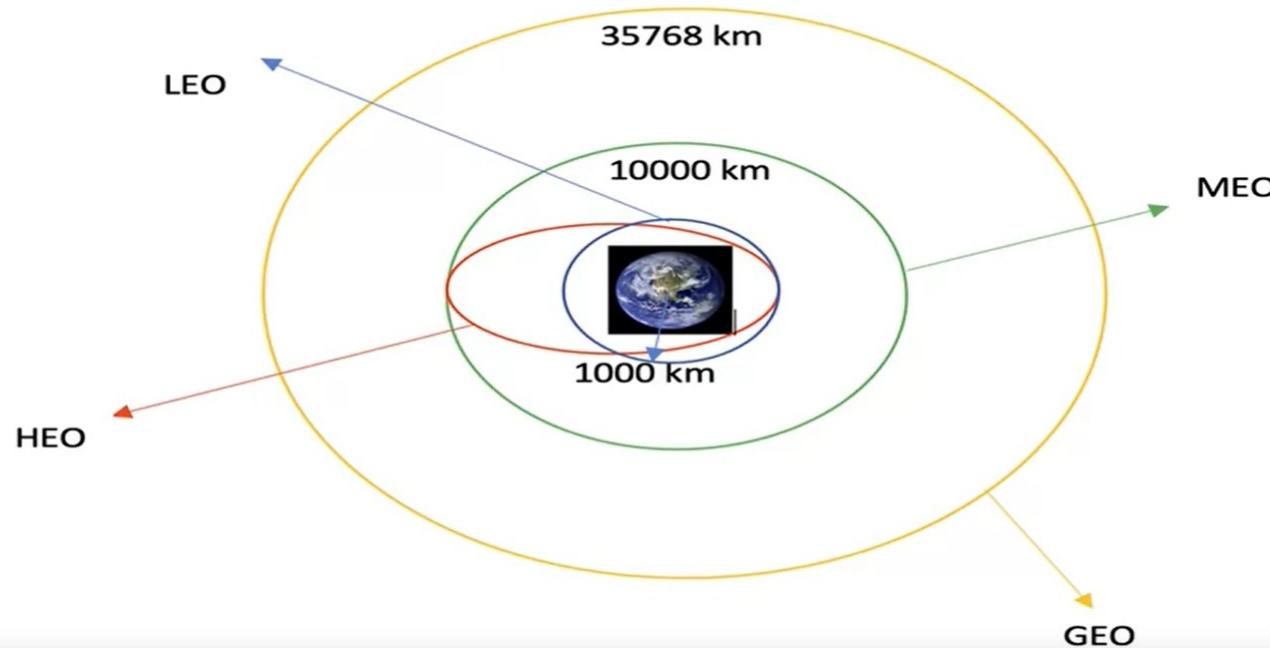
Data Wrangling Overview

Overview of Dataset

FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1 2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857
1	2 2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857
2	3 2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857
3	4 2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093
4	5 2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857
5	6 2014-01-06	Falcon 9	3325.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1005	-80.577366	28.561857
6	7 2014-04-18	Falcon 9	2296.000000	ISS	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1006	-80.577366	28.561857
7	8 2014-07-14	Falcon 9	1316.000000	LEO	CCAFS SLC 40	True Ocean	1	False	False	True	NaN	1.0	0	B1007	-80.577366	28.561857
8	9 2014-08-05	Falcon 9	4535.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1008	-80.577366	28.561857
9	10 2014-09-07	Falcon 9	4428.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1011	-80.577366	28.561857



Orbit



Outcome

- The column Outcome indicates if the first stage successfully landed

- 0 True ASDS
- 1 None None
- 2 True RTLS
- 3 False ASDS
- 4 True Ocean
- 5 None ASDS
- 6 False Ocean
- 7 False RTLS

Outcome

- We would like landing outcomes to be converted to Classes y . (either 0 or 1).
- 0 is a bad outcome i.e., the booster did not land
- 1 is a good outcome i.e., the booster did land

Module II

Exploratory Data Analysis

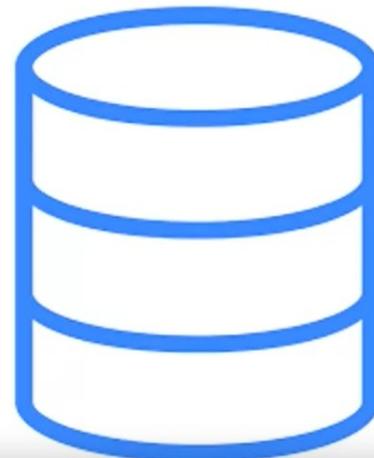
Lab 3

Exploratory Data Analysis

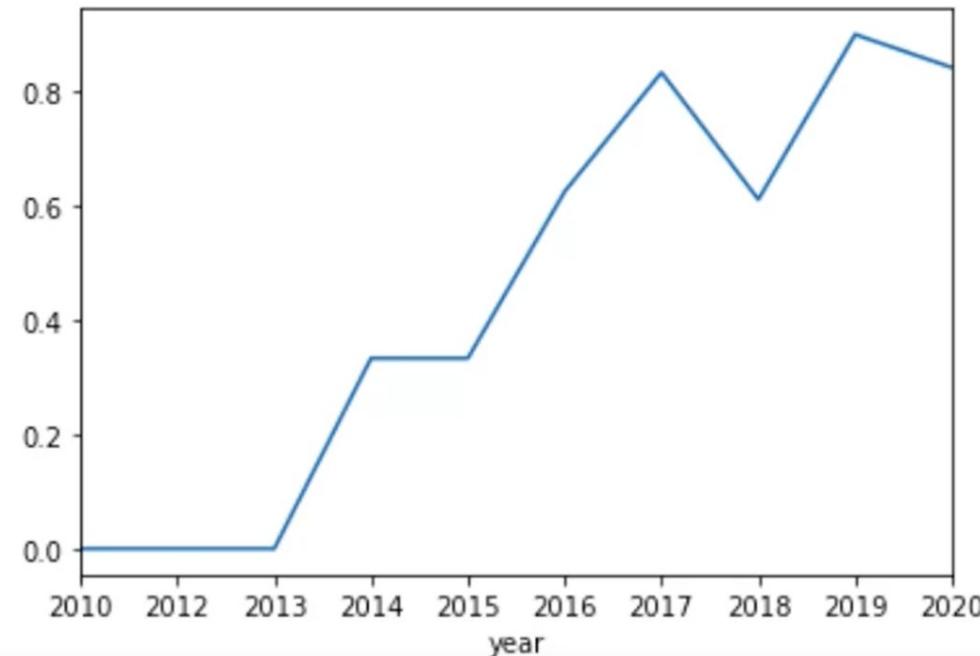
Exploratory Data Analysis Overview

Exploring the Data

- In the first lab you will perform some Exploratory Data Analysis using a database

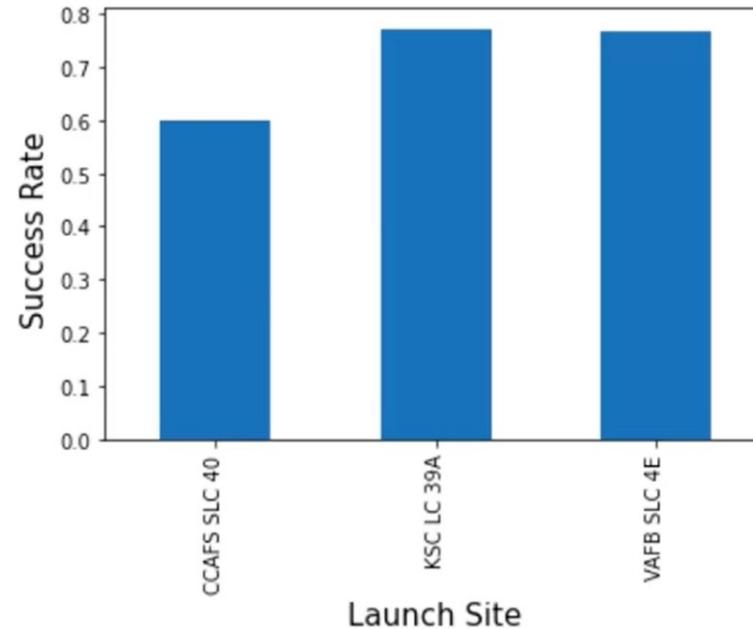


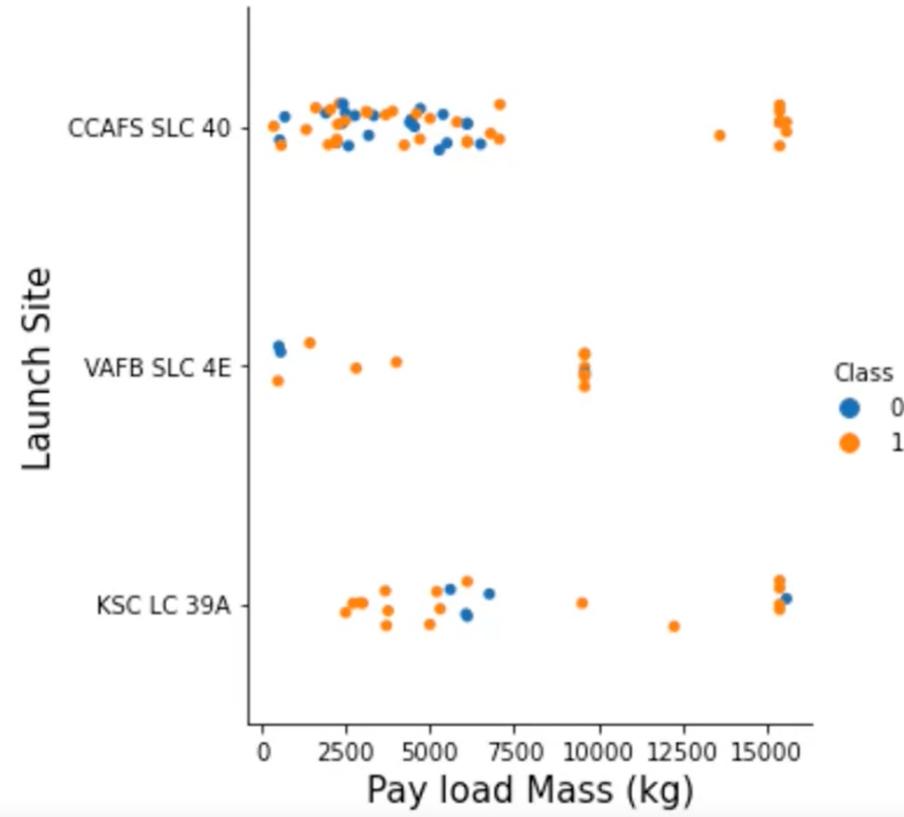
Feature Engineering





Feature Engineering





	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
1	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	0
2	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	0
3	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	0
4	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	0
5	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	0

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
1	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	0
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	0
1	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	0
1	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	0
1	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	0



	FlightNumber	PayloadMass	Flights	Block	ReusedCount	Orbit_ES-L1	Orbit_GEO	Orbit_GTO	Orbit_HEO	Orbit_ISS	...	Serial_B1058	Serial_B1059	Serial_B1060	Serial_B1062	GridFins_False	GridFins_True	Reused_False	Reused_True	Legs_False	Legs_True
1	1	6104.959412	1	1.0	0	0	0	0	0	0	...	0	0	0	0	1	0	1	0	1	0
1	2	525.000000	1	1.0	0	0	0	0	0	0	...	0	0	0	0	1	0	1	0	1	0
1	3	677.000000	1	1.0	0	0	0	0	0	1	...	0	0	0	0	1	0	1	0	1	0
1	4	500.000000	1	1.0	0	0	0	0	0	0	...	0	0	0	0	1	0	1	0	1	0
1	5	3170.000000	1	1.0	0	0	0	1	0	0	...	0	0	0	0	1	0	1	0	1	0

rows x 83 columns

Module III

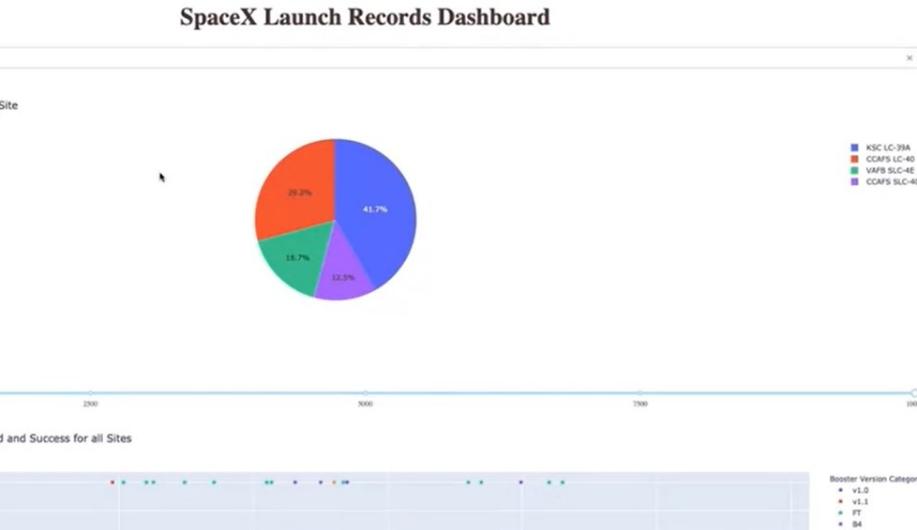
Interactive Visual Analytics and

Dashboards



Interactive Visual Analytics

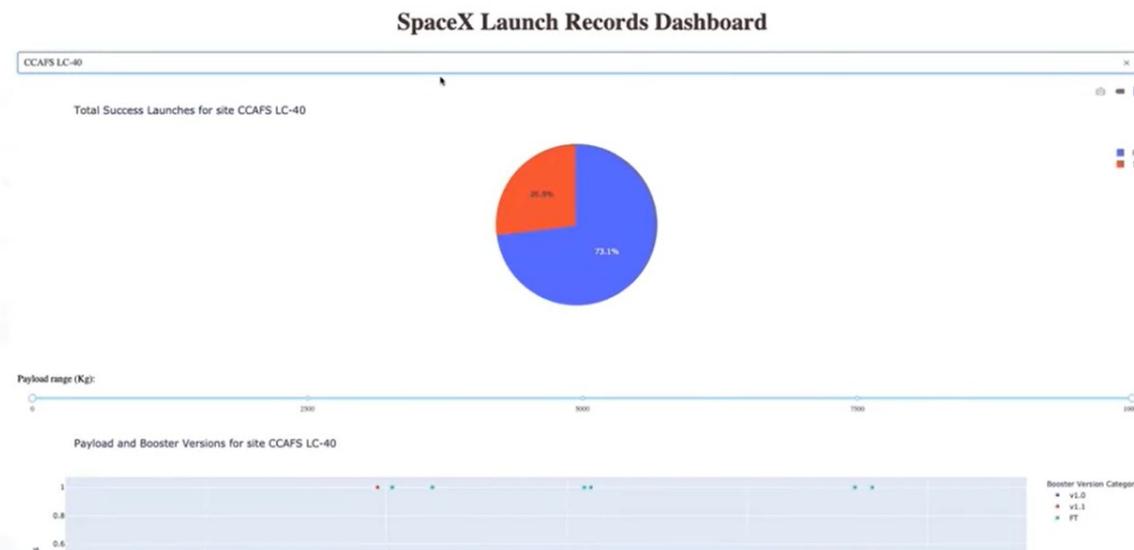
- Enable direct data exploration and analytics





Interactive Visual Analytics

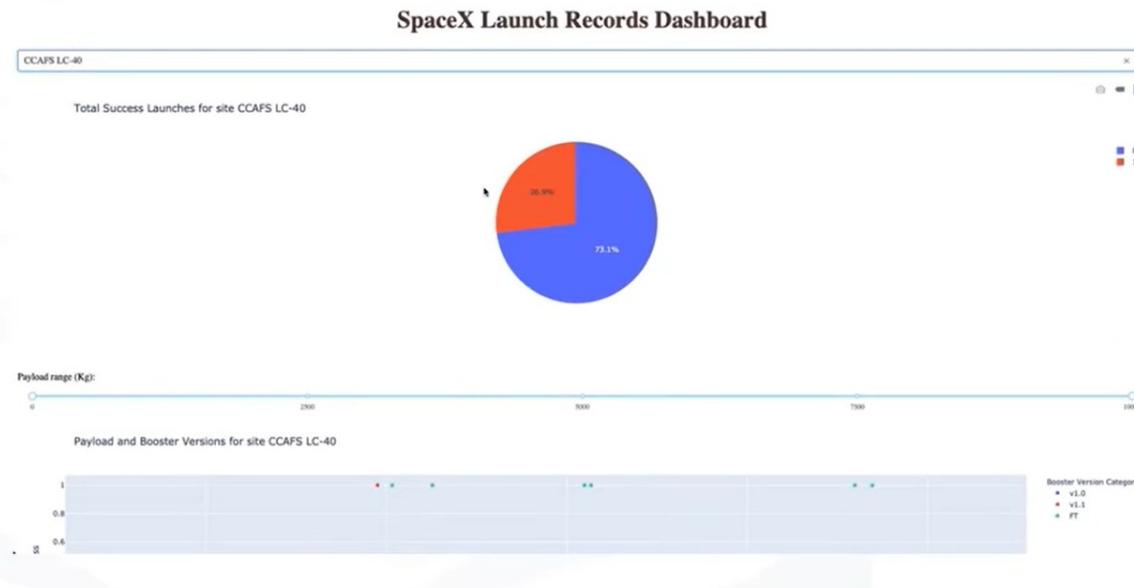
- Enable direct data exploration and analytics
- Zoom-in/out, Pan, Filter, Search, Link, etc.





Interactive Visual Analytics

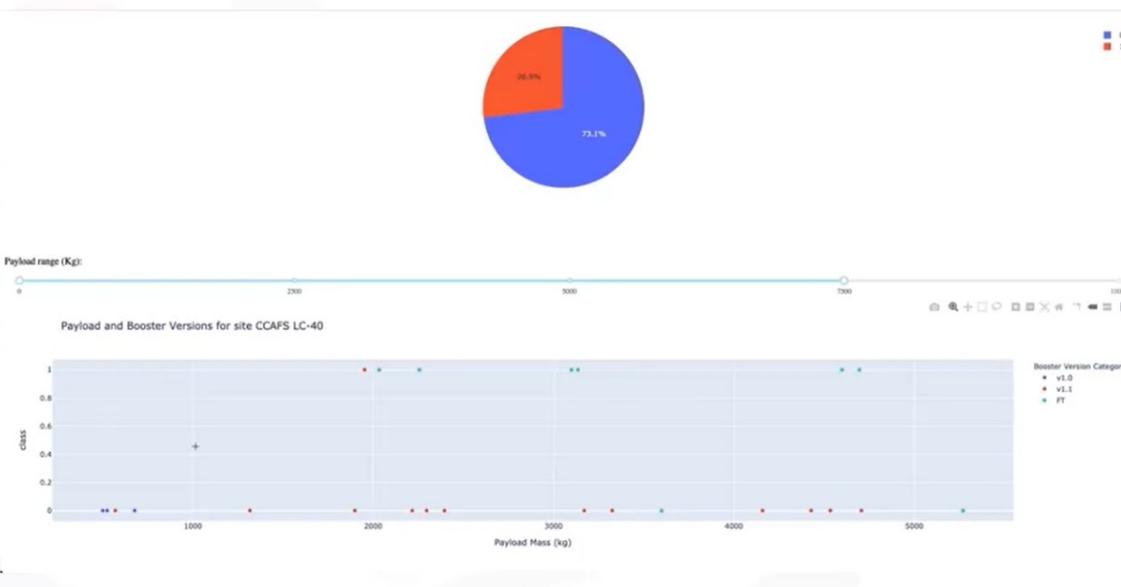
- Enable direct data exploration and analytics
- Zoom-in/out, Pan, Filter, Search, Link, etc.
- Identify patterns faster and more effectively





Interactive Visual Analytics

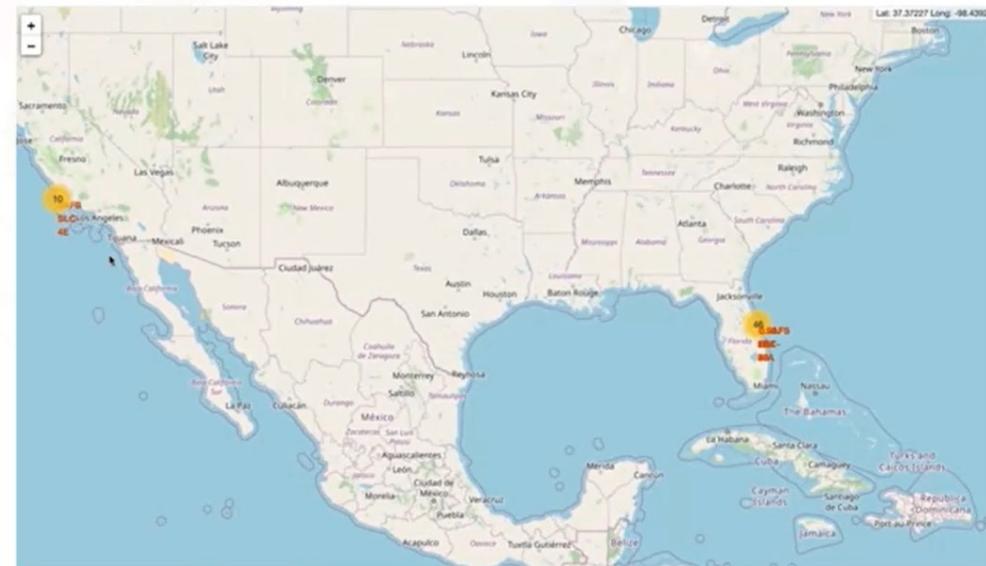
- Enable direct data exploration and analytics
- Zoom-in/out, Pan, Filter, Search, Link, etc.
- Identify patterns faster and more effectively
- More appealing stories



Folium

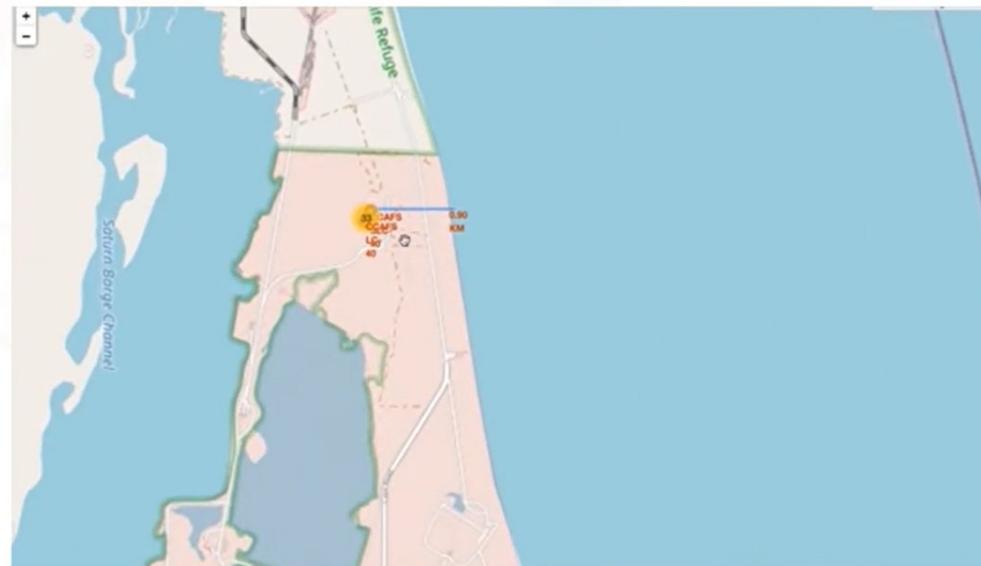
Plotly

Analyze Launch Site Geo Data with Folium



Analyze Launch Site Geo Data with Folium

- Mark the locations and proximities of launch sites





Analyze Launch Site Geo Data with Folium

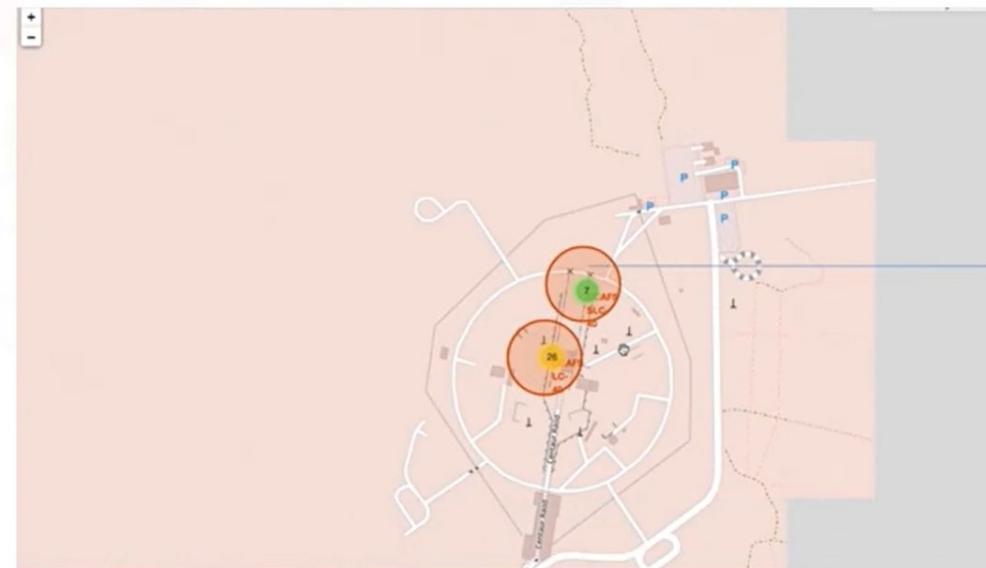
- Mark the locations and proximities of launch sites
- Discover patterns via exploring the map





Analyze Launch Site Geo Data with Folium

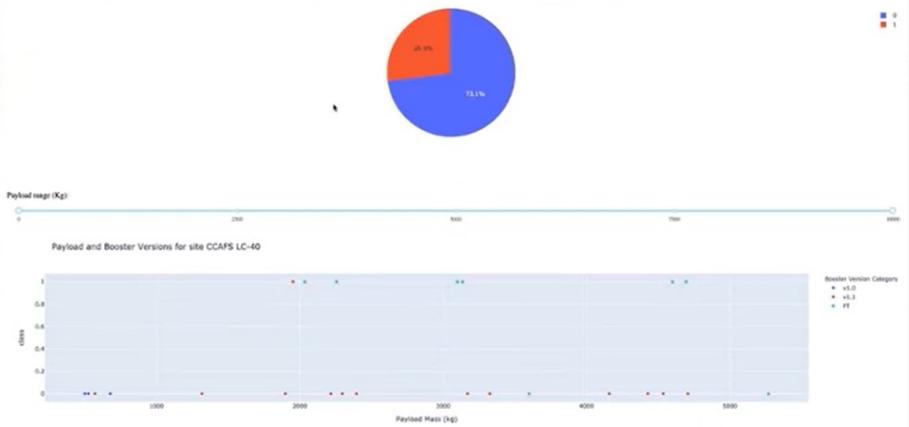
- Mark the locations and proximities of launch sites
- Discover patterns via exploring the map
- Explain how to choose an optimal launch site locations





Build a Dashboard with Plotly Dash

- Build a dashboard with dropdown list, range slider, and graphs





Build a Dashboard with Plotly Dash

- Build a dashboard with dropdown list, range slider, and graphs
- Obtain insights by using the dashboard





Take a Break

Module IIII

Predictive Analysis (Classification)

Predictive Analysis Overview



Build a Machine Learning Pipeline

- Predict whether first stage of Falcon 9 will land successfully



Build a Machine Learning Pipeline

- Predict whether first stage of Falcon 9 will land successfully

Preprocessing

```
from sklearn import preprocessing
```

Build a Machine Learning Pipeline

- Predict whether first stage of Falcon 9 will land successfully



```
from sklearn.model_selection import train_test_split
```

Build a Machine Learning Pipeline

- Predict whether first stage of Falcon 9 will land successfully

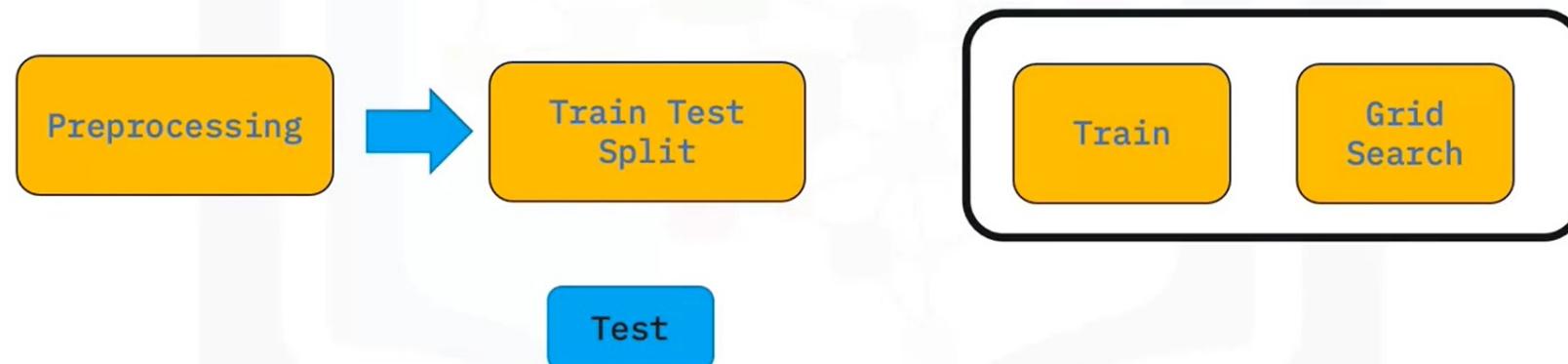


```
|from sklearn.model_selection import GridSearchCV
```



Build a Machine Learning Pipeline

- Predict whether first stage of Falcon 9 will land successfully



```
|from sklearn.model_selection import GridSearchCV
```

Determine Model with Best Accuracy



Test

Determine Model with Best Accuracy

Test



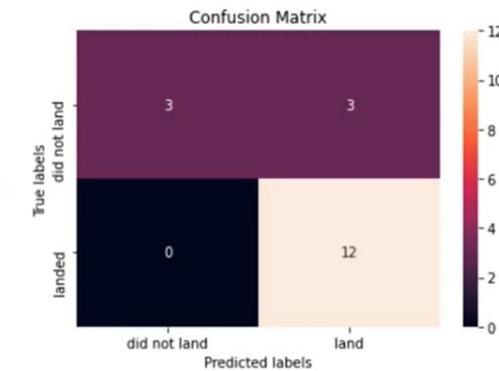
Model
Logistic Regression
Support Vector Machine
Decision Tree Classifier
K-nearest Neighbors

Determine Model with Best Accuracy

Test



Model
Logistic Regression
Support Vector Machine
Decision Tree Classifier
K-nearest Neighbors



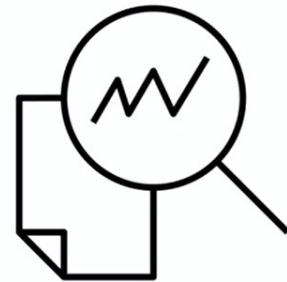
Module 5

Present your data-driven insights

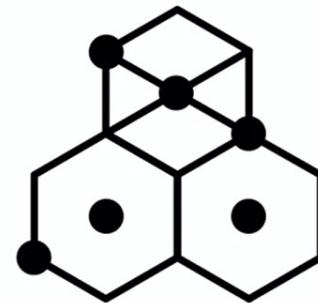
Elements of a Successful
Data Findings Report



Introduction



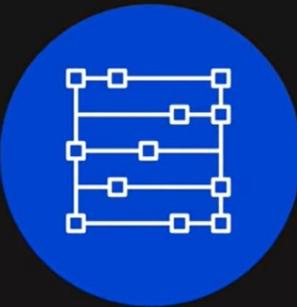
finding and
cleaning data



organize and
represent the findings

Find the answers

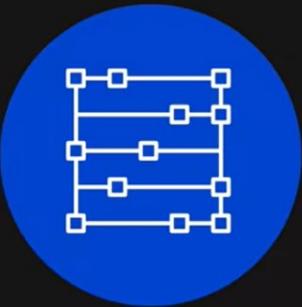
Data



collected
cleaned
organized

Find the answers

Data



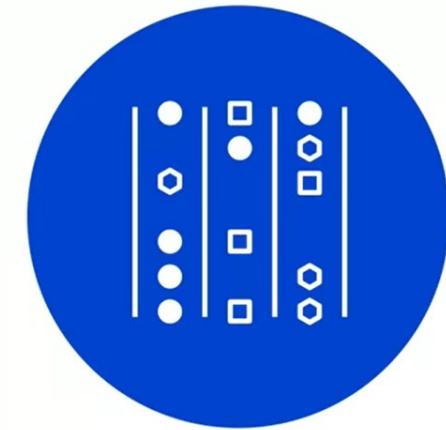
collected
cleaned
organized

Report



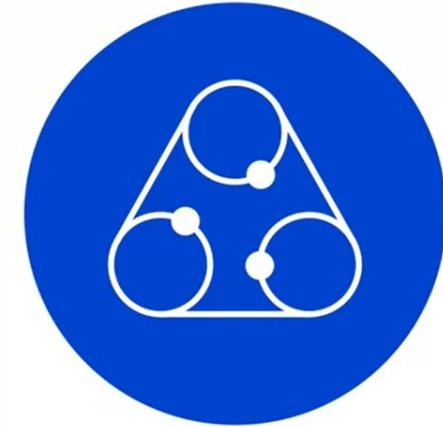
paper style report
slideshow presentation

Elements of a data findings report



Findings Report

Elements of a data findings report



Outline



Elements of a data findings report



-  Cover Page
-  Executive Summary
-  Table of Contents
-  Introduction
-  Methodology
-  Results
-  Discussion
-  Conclusion
-  Appendix

Executive Summary

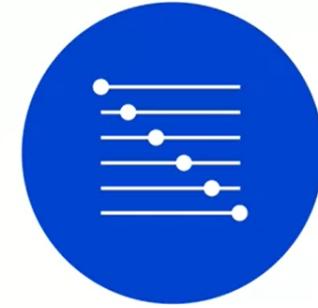


- Briefly explain the details
- Considered a stand-alone document

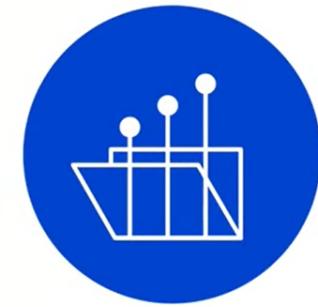
Introduction



- Nature of the analysis
- States the problem
- States questions for analysis



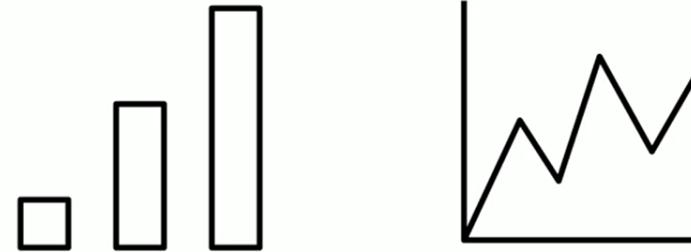
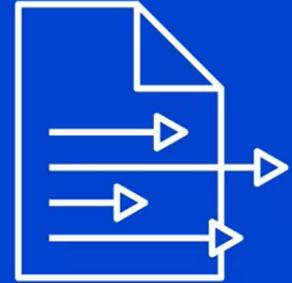
explains the data sources



outlines the plan for the collected data



Results



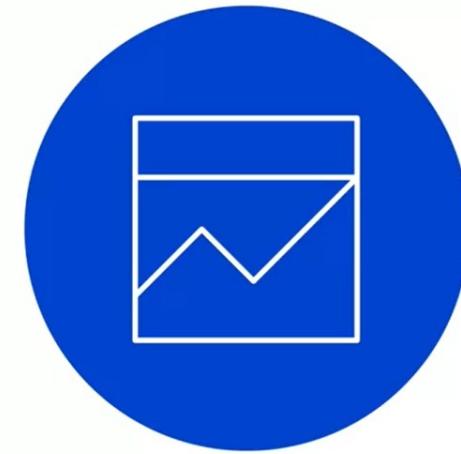
Charts and graphs

Discussion Findings and Implications



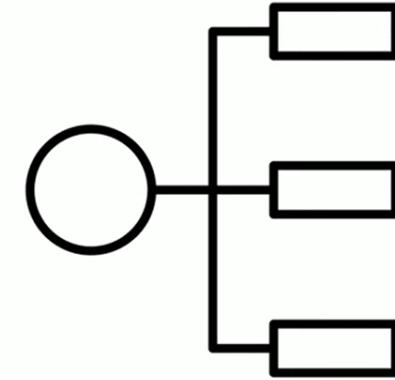
engage the audience

Conclusion



Conclusion of the report findings

Appendix References



Information that didn't fit in the report

Module 5

Present your data-driven insights

Best Practices for
Presenting Your Findings

Best Practices For Presenting Your Findings

You've spent weeks, maybe months, studying the data and the time has come to report your findings. The questions have been answered, and you feel good about the story. How will you speak to your audience so they leave with the intended message?



Delivering your message

Factors to remember in accurately conveying your message

- Make sure charts and graphs are not too small, and are clearly labeled
- Use the data only as supporting evidence
- Share only one point from each chart
- Eliminate data that does not support the key message

Delivering your message

Have you ever sat through a presentation and the information being presented was difficult to read or understand?

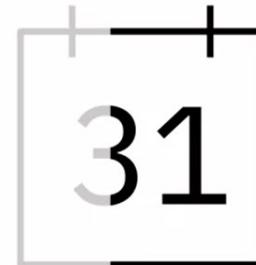


Image difficult to read

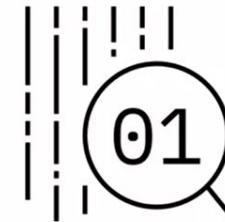


Delivering your message

Have you ever sat through a presentation and the information being presented was difficult to read or understand?



Charts

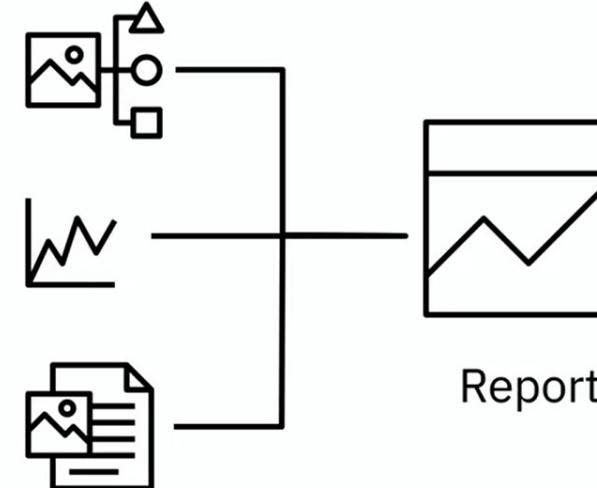


Labels

Delivering your message



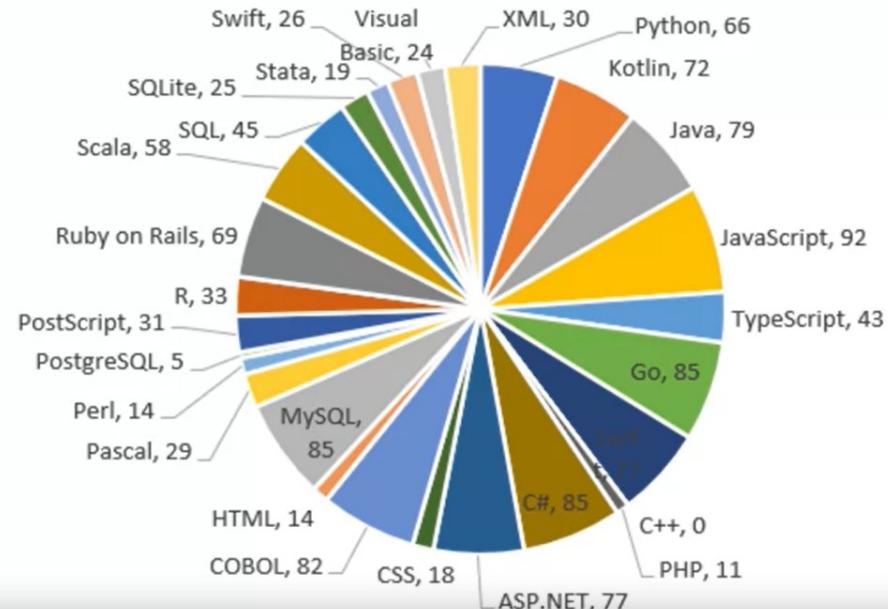
Delivering your message



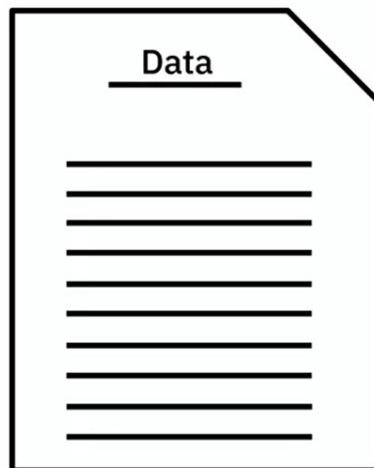
Report



Delivering your message



Delivering your message



Effective Data-Driven Presentations

- Delivering Findings:
 - Ensure clear communication of conclusions.
 - Focus on engaging the audience effectively.
- Presentation Tips:
 - Use clearly labeled and visible charts/graphs.
 - Share one point per visualization for clarity.
- Storytelling Approach:
 - Form key messages before adding supporting data.
 - Create an engaging narrative around key points.

Presentation Best Practices

- Visual Clarity:
 - Test visualizations for readability at various distances.
 - Avoid overwhelming slides with excessive data.
- Message Focus:
 - Highlight relevant data points that support key ideas.
 - Eliminate irrelevant data to maintain clarity.
- Conclusion:
 - Data-driven presentations should be engaging and concise.
 - Emphasize clear communication and message alignment with audience understanding.

Q&A

Questions and Answers



Thanks