IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Gina Vlassi
September 2021

Outline

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

Executive Summary

Summary of methodologies

- Data Collection
- Data Wrangling
- EDA (Exploratory Data Analysis) with Data Visualization
- EDA with SQL
- Building an Interactive map with Folium
- Predictive Analysis (Classification)

Summary of all results

- Exploratory data analysis
- Interactive analysis demo
- Predictive analysis

Introduction

- Gather information about SpaceX and use them for our analysis
- Implement EDA with SQL commands and visualizations
- Use Folium and Plotly Dash for interactive analysis
- Use classification models for predictive analysis

Section 1

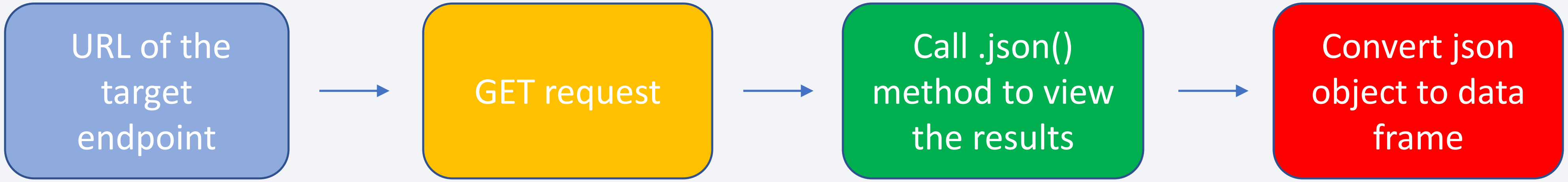
Methodology

Methodology

Executive Summary

- Data collection methodology:
  - The Space X REST API
  - Web Scarping
- Perform data wrangling
  - One hot encoding
  - Drop NA values
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
  - How to build, tune, evaluate classification models

Data Collection – SpaceX API



1. spacex_url=https://api.spacexdata.com/v4/launches/past
2. response = requests.get(spacex_url)
3. data = pd.json_normalize(response.json())

	static_fire_date_utc	static_fire_date_unix	net	window	rocket	success	failures	details	crew	ships	capsules	payloads
0	2006-03-17T00:00:00.000Z	1.142554e+09	False	0.0	5e9d0d95eda69955f709d1eb	False	[{'time': 33, 'altitude': None, 'reason': 'merlin engine failure'}]	Engine failure at 33 seconds and loss of vehicle	[]	[]	[]	[5eb0e4b5b
1	None	NaN	False	0.0	5e9d0d95eda69955f709d1eb	False	[{'time': 301, 'altitude': 289, 'reason': 'harmonic oscillation leading to premature engine shutdown'}]	Successful first stage burn and transition to second stage, maximum altitude 289 km, Premature engine shutdown at T+7 min 30 s, Failed to reach orbit, Failed to recover first stage	[]	[]	[]	[5eb0e4b6b

[Data Collection REST API]()

Data Collection - Scraping

Request the Falcon9 Launch Wiki page from its URL → Extract all column/variable names from the HTML table header → Create a data frame by parsing the launch HTML tables

1. response = requests.get(static_url)
2. soup = BeautifulSoup(response.content, 'html.parser')
3. html_tables = soup.find_all('table')
4. table_headers = first_launch_table.find_all('th')
5. launch_dict= dict.fromkeys(column_names)

```
<tr>
<th scope="col">Flight No.
</th>
<th scope="col">Date and<br/>time (<a href="/wiki/Coordinated_Universal_Time" title="Coordinated Universal Time">UTC</a>)
</th>
<th scope="col"><a href="/wiki/List_of_Falcon_9_first-stage_boosters" title="List of Falcon 9 first-stage boosters">Version,<br/>B
ooster</a> <sup class="reference" id="cite_ref-booster_11-0"><a href="#cite_note-booster-11">[b]</a></sup>
</th>
<th scope="col">Launch site
</th>
<th scope="col">Payload<sup class="reference" id="cite_ref-Dragon_12-0"><a href="#cite_note-Dragon-12">[c]</a></sup>
</th>
<th scope="col">Payload mass
</th>
<th scope="col">Orbit
</th>
<th scope="col">Customer
</th>
<th scope="col">Launch<br/>outcome
</th>
<th scope="col"><a href="/wiki/Falcon_9_first-stage_landing_tests" title="Falcon 9 first-stage landing tests">Booster<br/>landing
</a>
</th></tr>
```


[Data Collection Web Scarping](#)

Data Wrangling

Perform EDA

Calculate the number of launches

Calculate the number of occurrences of each orbit

Calculate the number and occurrence of mission outcome per orbit type

Create Class column

The values in outcome Class column is converted into classes:

0 → bad outcome (unsuccessful launch)

1 → good outcome (successful launch)

Data Wrangling

EDA with Data Visualization

Scatter Plots:

- FlightNumber vs. PayloadMass
- FlightNumber vs LaunchSite
- PayloadMass vs LaunchSite
- Orbit vs FlightNumber
- Orbit vs PayloadMass

Bar Chart:

- Orbit vs SuccessRate

Line Chart:

- Year vs Class

Explanatory Data Analysis with Visualization

EDA with SQL

The SQL queries performed:

- Task 1: Display the names of the unique launch sites in the space mission
- Task 2: Display 5 records where launch sites begin with the string 'CCA‘
- Task 3: Display the total payload mass carried by boosters launched by NASA (CRS)
- Task 4: Display average payload mass carried by booster version F9 v1.1
- Task 5: List the date when the first successful landing outcome in ground pad was achieved
- 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
- Task 7: List the total number of successful and failure mission outcomes
- Task 8: List the names of the booster_versions which have carried the maximum payload mass
- Task 9: List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015
- Task 10: 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

[Explanatory Data Analysis with SQL]()

Build an Interactive Map with Folium

The Folium package is used for the map:

- To visualize launch data into an interactive map – latitude and longitude coordinates of each launch site are used to add circle markers with a label of launch site name
- To add Green (success) vs Red (failure) markers on the map
- To calculate the distances between a launch site to its proximities (railway, highway, coastline)

[Interactive Map with Folium]

Predictive Analysis (Classification)

Building a model

- Create a numpy from a loaded dataset
- Standardize the data in X
- Train and test dataset split
- Construct models: Logistic Regression, Support Vector Machine, Decision Tree Classifier, K Nearest Neighbors

Evaluating a model

- Accuracy
- Get tuned hyperparmeters for each model
- Confusion Matrix

Finding the best performing model

[Predictive Analysis - Classification]

Results

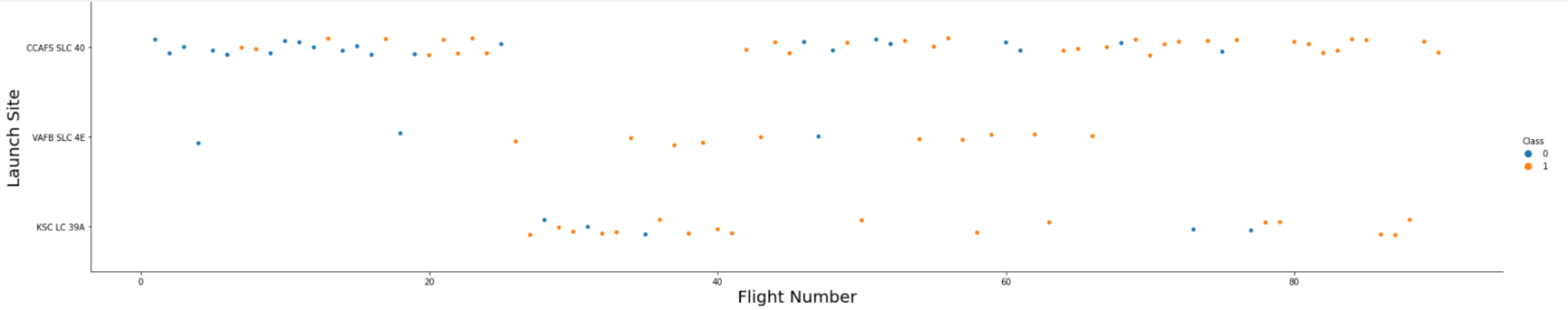
- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

Insights drawn from EDA

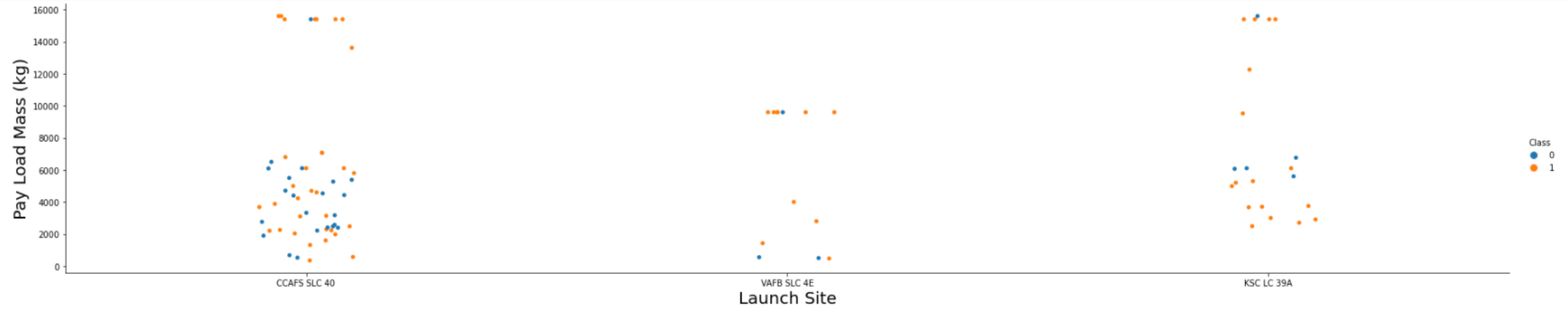
Flight Number vs. Launch Site

- Observing the data we understand that as the number of the flight increases, is more likely to land successfully on the launch site



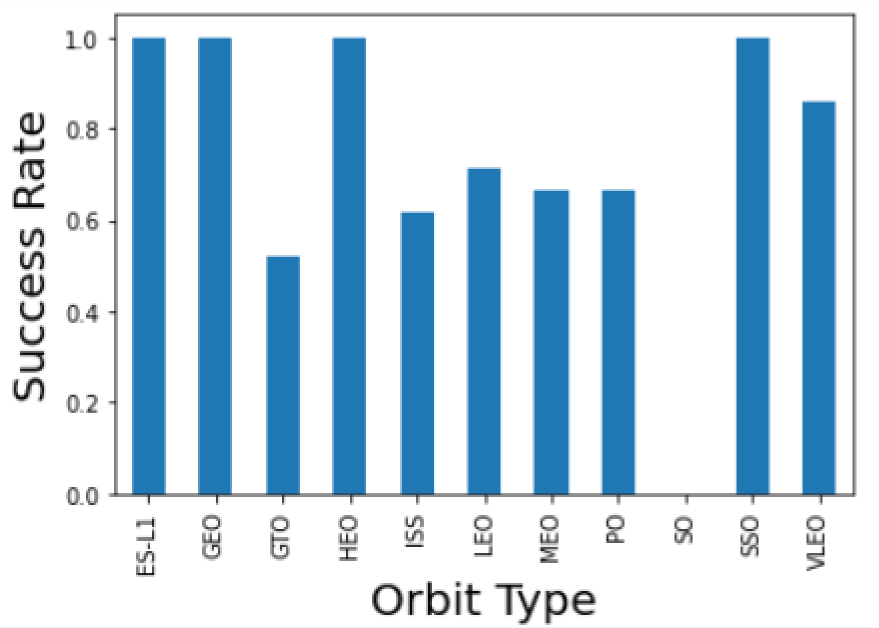
Payload vs. Launch Site

- Observing the data we understand that sites CCAFS SLC 40 and KSC LC 39A have the the successful landing of bigger pay load mass in KG.



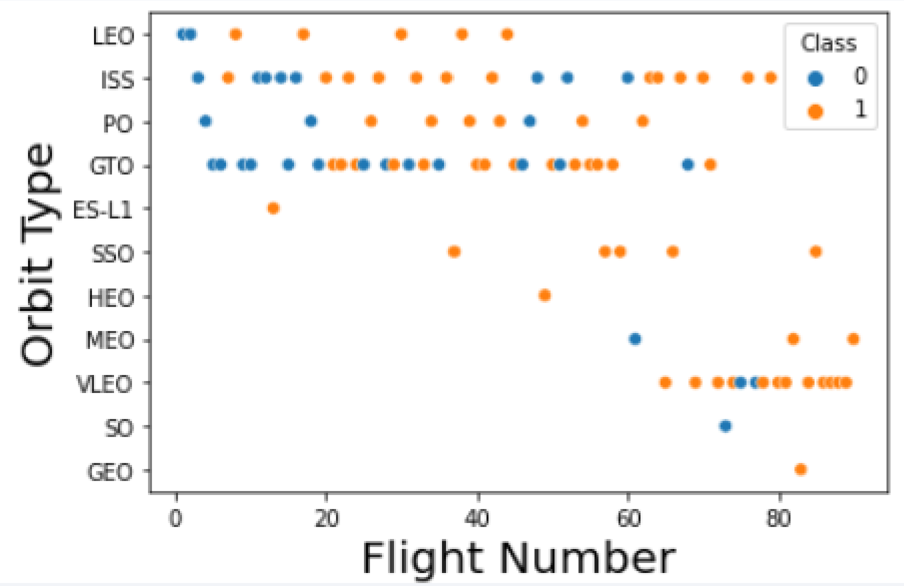
Success Rate vs. Orbit Type

- Orbit types ES-L1, GEO, HEO and SSO have the absolute landing success (100%)



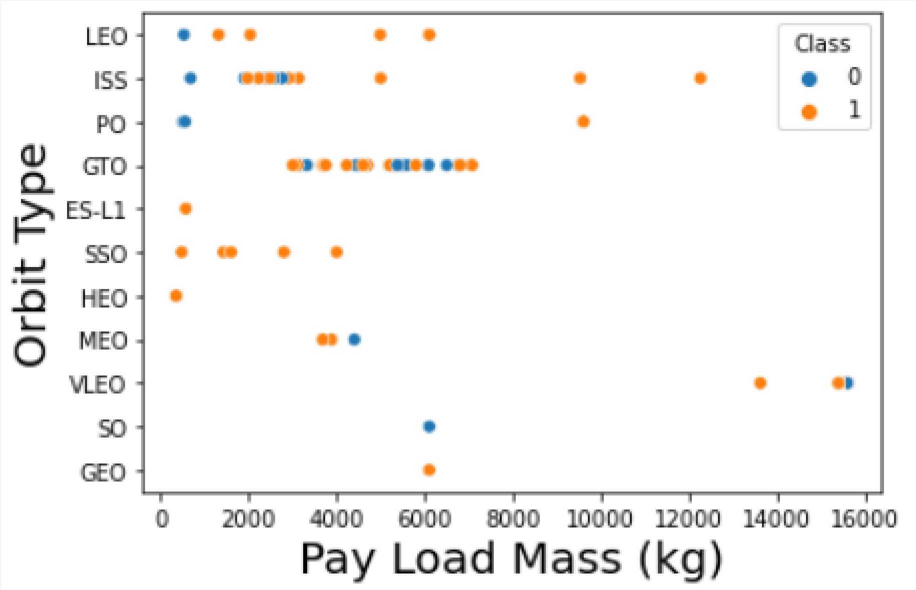
Flight Number vs. Orbit Type

- Observing the data we understand that as the flight number increases in Orbit types LEO, SSO, VLEO, we have the more successful landing outcome



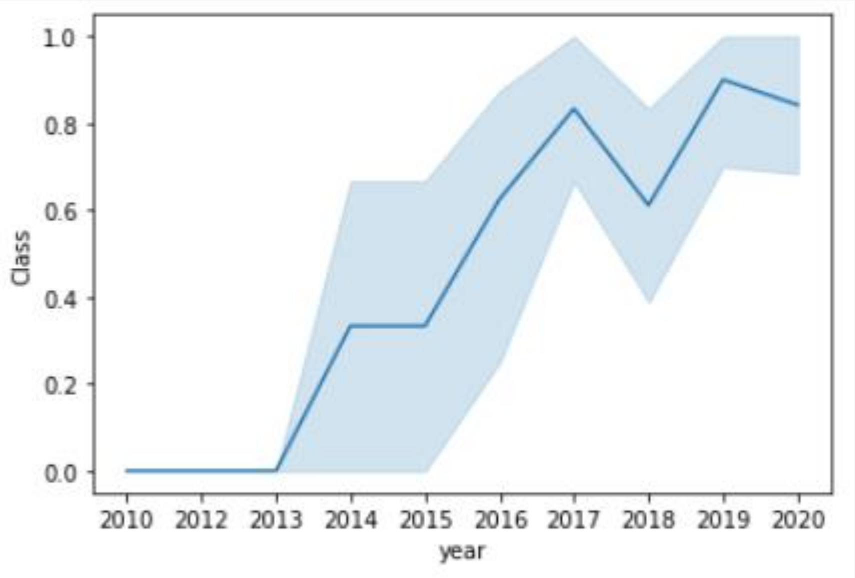
Payload vs. Orbit Type

- Observing the data we understand that the most successful lands need from 2000 to 6000 KG of pay load



Launch Success Yearly Trend

- During the years we observe that the average success rate increases



All Launch Site Names

- Use DISTINCT() to find the names of the unique launch sites

```
%sql select distinct(launch_site) from SPACEXTBL
```

```
 * ibm_db_sa://jcw72784:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

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

Launch Site Names Begin with 'CCA'

- Use LIKE to find sites begin with `CCA%`
- Use LIMIT 5 to select the first 5 records

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

```
 * ibm_db_sa://jcw72784:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- Use SUM() to count the total payload
- Use WHERE to specify the customer

```
%sql select sum(payload_mass__kg_) from SPACEXTBL where customer = 'NASA (CRS)'
```

```
 * ibm_db_sa://jcw72784:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

1
45596

Average Payload Mass by F9 v1.1

- Use AVG() to calculate the average payload mass
- Use WHERE to specify the booster version F9 v1.1

```
%sql select avg(payload_mass__kg_) from SPACEXTBL where booster_version ='F9 v1.1'
```

```
 * ibm_db_sa://jcw72784:***@fbd88901-ebdb-4a4f-a32e-9822b9fb237b.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32731/bludb
Done.
```

1
2928

First Successful Ground Landing Date

- Use MIN() to find the first date
- Use WHERE to specify successful outcome

```
%sql select min(DATE) from SPACEXTBL where mission_outcome = 'Success'
```

```
 * ibm_db_sa://mps48094:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

1
2010-06-04

Successful Drone Ship Landing with Payload between 4000 and 6000

- Use <= and >= to specify payload_mass__kg_

```
%sql select payload from SPACEXTBL where (payload_mass__kg_>=4000 and payload_mass__kg_<=6000) and mission_outcome = 'Success'
```

```
 * ibm_db_sa://mps48094:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

payload
AsiaSat 8
AsiaSat 6
ABS-3A Eutelsat 115 West B
Turkmen 52 / MonacoSAT
SES-9
JCSAT-14
JCSAT-16
EchoStar 23
SES-10
NROL-76
Boeing X-37B OTV-5
SES-11 / EchoStar 105
GovSat-1 / SES-16
SES-12
Merah Putih
Es hail 2
SSO-A
GPS III-01
Nusantara Satu, Beresheet Moon lander, S5
RADARSAT Constellation, SpaceX CRS-18
GPS III-03, ANASIS-II
ANASIS-II, Starlink 9 v1.0
GPS III-04 , Crew-1

Total Number of Successful and Failure Mission Outcomes

- Use COUNT() to calculate the total number of successful and failure mission outcomes

```
%sql select count(mission_outcome) from SPACEXTBL
```

```
 * ibm_db_sa://mps48094:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

1
101

Boosters Carried Maximum Payload

- Use subquery to list the names of the booster which have carried the maximum payload mass

```
%sql select booster_version from SPACEXTBL where payload_mass__kg_=(select max(payload_mass__kg_) from SPACEXTBL)
```

```
 * ibm_db_sa://mps48094:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

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

2015 Launch Records

- Use YEAR() to cast the date

```
%sql select landing__outcome,booster_version,launch_site from SPACEXTBL where landing__outcome = 'Failure (drone ship)' and year(DATE) =
'2015'
```

```
 * ibm_db_sa://mps48094:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

landing__outcome	booster_version	launch_site
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

- Use ORDER BY to rank the count of landing outcomes

```
%sql select * from SPACEXTBL where landing__outcome like 'Success%' and (DATE between '2010-06-04' and '2017-03-20') order by date desc
```

```
 * ibm_db_sa://mps48094:***@764264db-9824-4b7c-82df-40d1b13897c2.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:32536/bludb
Done.
```

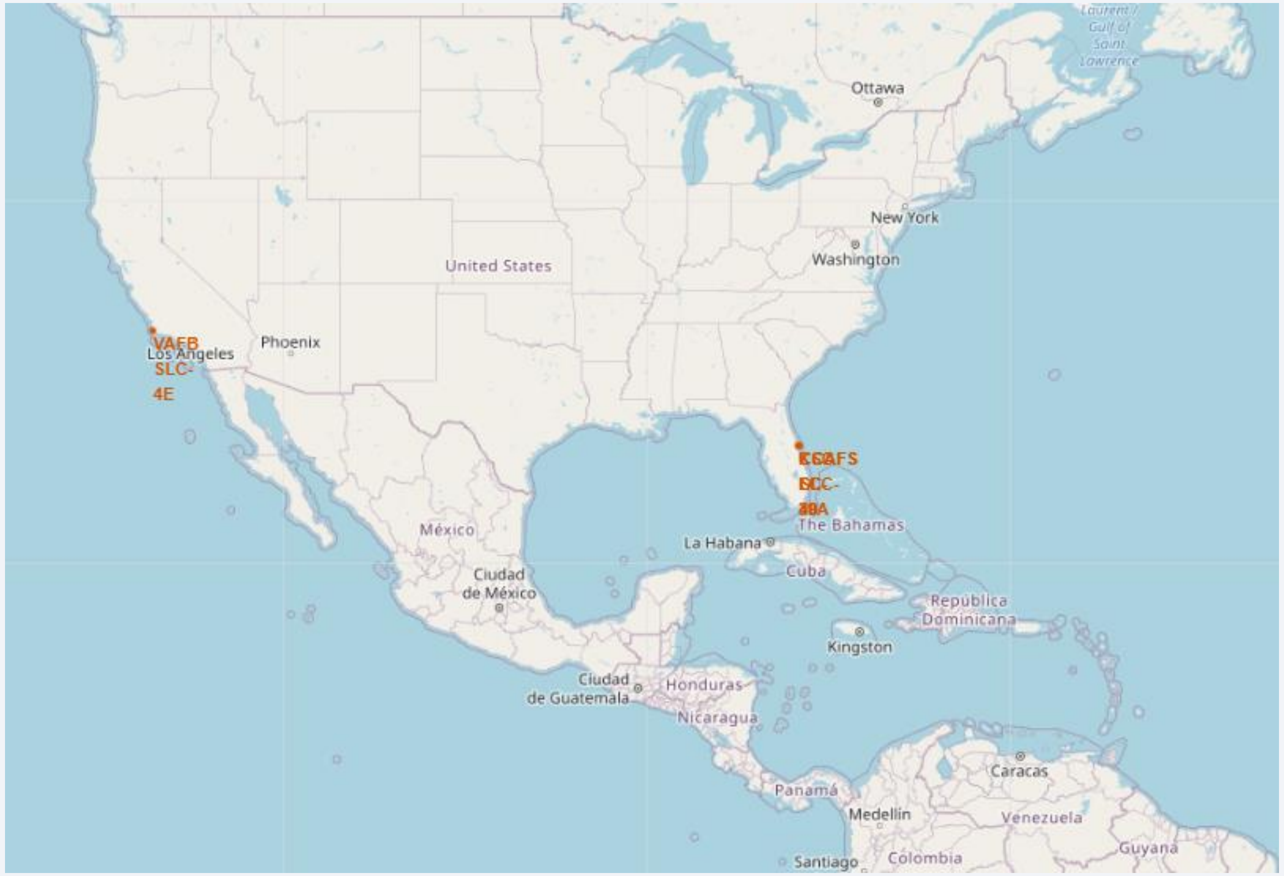
DATE	time__utc_	booster_version	launch_site	payload	payload_mass__kg_	orbit	customer	mission_outcome	landing__outcome
2017-02-19	14:39:00	F9 FT B1031.1	KSC LC-39A	SpaceX CRS-10	2490	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2017-01-14	17:54:00	F9 FT B1029.1	VAFB SLC-4E	Iridium NEXT 1	9600	Polar LEO	Iridium Communications	Success	Success (drone ship)
2016-08-14	05:26:00	F9 FT B1026	CCAFS LC-40	JCSAT-16	4600	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-07-18	04:45:00	F9 FT B1025.1	CCAFS LC-40	SpaceX CRS-9	2257	LEO (ISS)	NASA (CRS)	Success	Success (ground pad)
2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thaicom 8	3100	GTO	Thaicom	Success	Success (drone ship)
2016-05-06	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect JSAT Group	Success	Success (drone ship)
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3136	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
2015-12-22	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

Section 4

Launch Sites Proximities Analysis

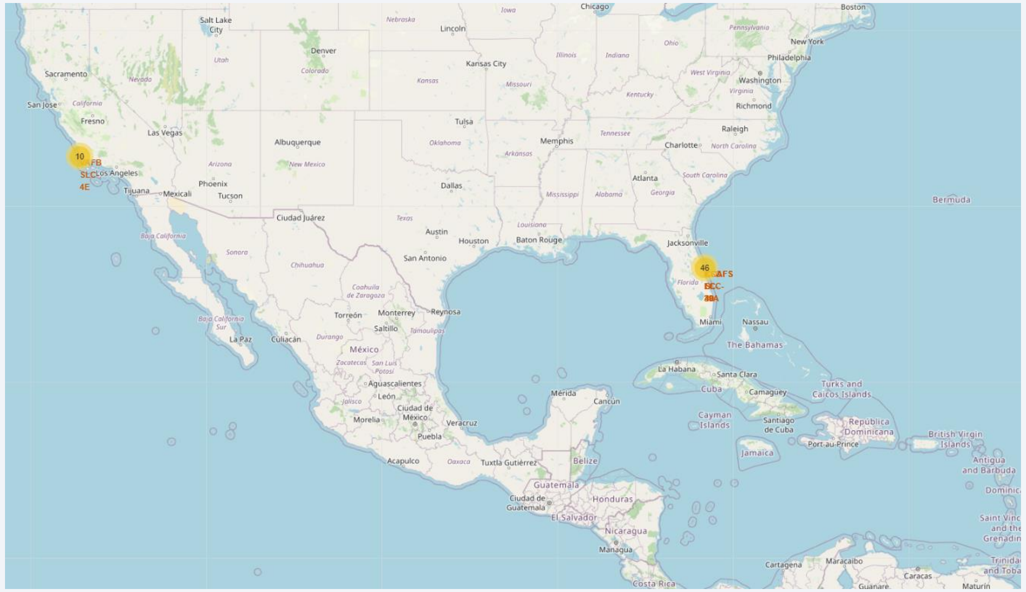
Launch all sites on global map

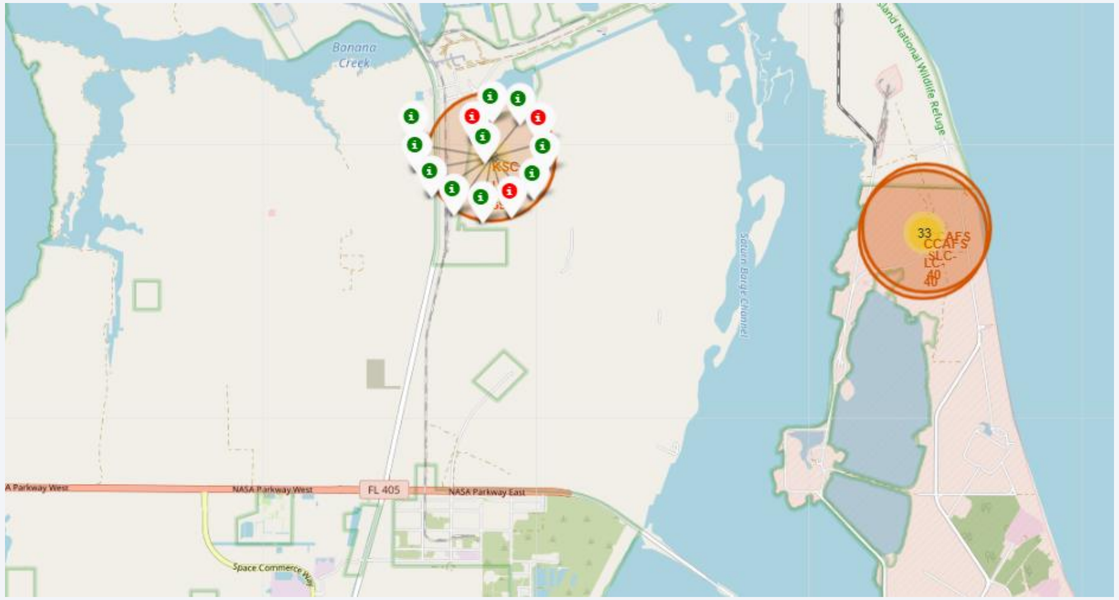
- We use the NASA coordinates to construct the map. With the red circles we highlight the all the launch sites.



Mark the success/failed launches for each site on the map

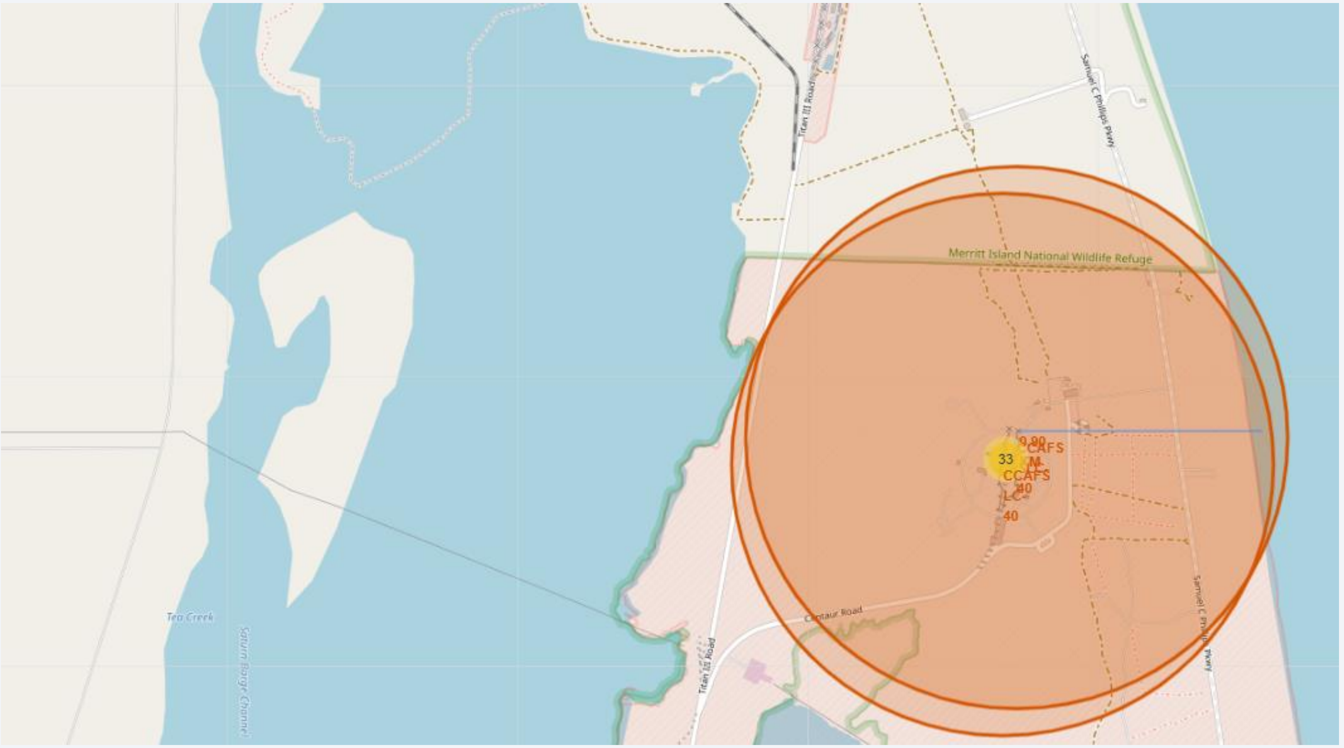
- With the yellow highlights we observe the total launches of each site
- The green spots specify successful launch in comparison with the red spots





Calculate the distances between a launch site to its proximities

- The blue line indicates the distance between the launch site and the closest coastline

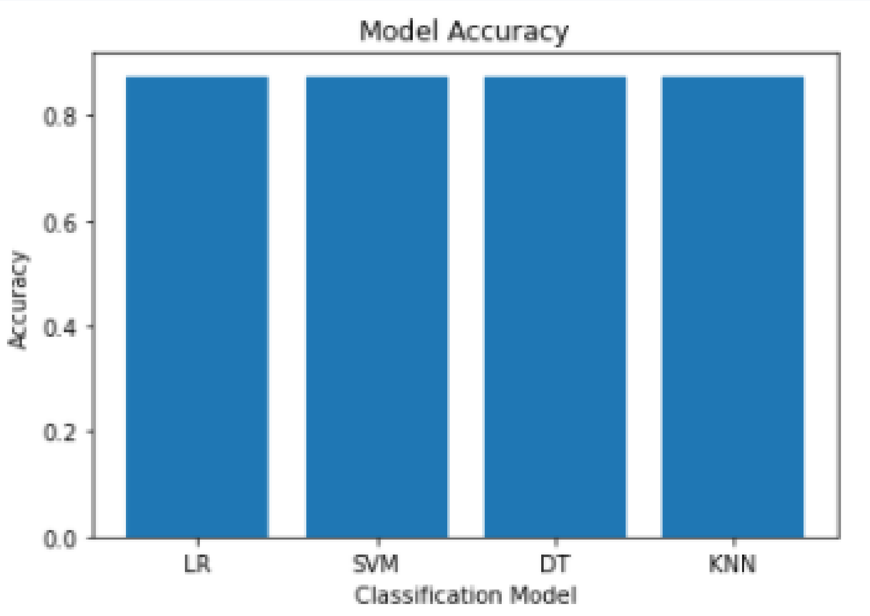


Section 6

Predictive Analysis (Classification)

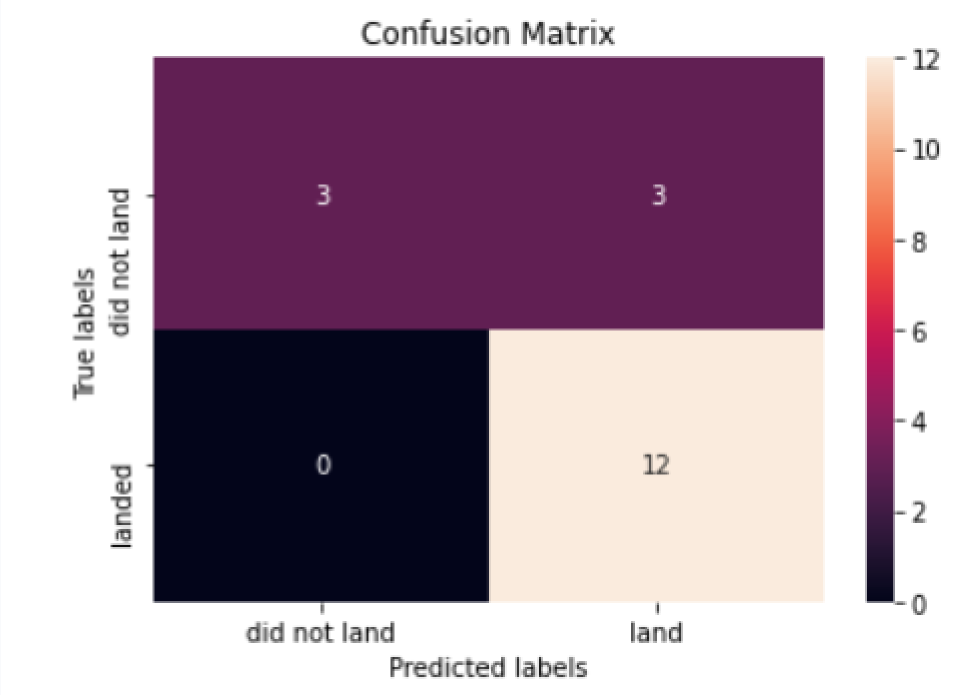
Classification Accuracy

- The accuracy results are practically the same of all the classification models



Confusion Matrix

- A k nearest neighborhood has a high True positive rate



Conclusions

Completing this analysis

we achieved to:

- Analyze real data from a REST API
- Construct explanatory data analysis with simple SQL command and visualization charts
- Use Folium package to draw a map with different interactions
- Build different classification models and calculate the accuracy of the best model

we conclude that:

- All the classification models have a high accuracy of nearly 90%
- Launch site KSC LC-39A had the highest success rate
- Orbit types GEO, HEO, SSO, ES-L1 has the highest success rates (100%)
- Heavy payloads have a negative influence on GTO orbits and positive on GTO and Polar LEO (ISS) orbits

Appendix

All the code of the labs is uploaded at the below link:

Hands-on Lab

Thank you!