Winning Space Race with Data Science

Mateus de Araújo Carvalho
12/20225

Outline

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

Executive Summary

- **Foram usados os seguintes passo durante o desenvolvimento deste trabalho:**
  - Coleta de dados por web scraping
  - EDA para adequação dos dados, visualização e análise preliminar
  - Aprendizado de máquina para criação de modelo de predição de sucesso dos lançamentos.

- **Resumo de resultados:**
  - Foi possível realizar a coleta e tratamento de dados;
  - Com a etapa de EDA é possível realizar a definição das principais variáveis de relevância e já obter informações gerais sobre o problema;
  - Os modelos de aprendizagem de máquina conseguiram realizar a predição das variáveis desejadas.

Introduction

In this capstone, we will predict if the Falcon 9 first stage will land successfully. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.

Therefore **, if we can determine if the first stage will land, we can determine the cost of a launch.** This information can be used if an alternate company wants to bid against SpaceX for a rocket launch. In this lab, you will collect and make sure the data is in the correct format from an API. The following is an example of a successful and launch.

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - We will use da Requests libraries to get data information's.
- Perform data wrangling
  - Describe how data was processed
- 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

- Usando a biblioteca “request” obteve-se as seguintes informações.:

  Ex: spacex_url=https://api.spacexdata.com/v4/launches/past

  response = requests.get(spacex_url)

From the rocket we would like to learn the booster name

From the payload we would like to learn the mass of the payload and the orbit that it is going to

From the launchpad we would like to know the name of the launch site being used, the longitude, and the latitude.

From cores we would like to learn the outcome of the landing, the type of the landing, number of flights with that core, whether gridfins were used, whether the core is reused, whether legs were used, the landing pad used, the block of the core which is a number used to seperate version of cores, the number of times this specific core has been reused, and the serial of the core.

Data Collection - Scraping

Para coleta de dados foi utilizado o commando request

- spacex_url="https://api.spacexdata.com/v4/launches/past"Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose
- response = requests.get(spacex_url)

Link para Notebook

Data Wrangling

- Foi realizado primeiramente o EDA

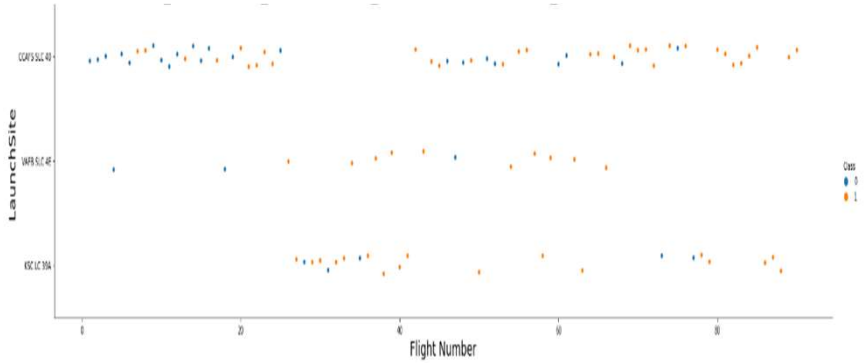
- Depois foi realizada a sumarização e por fim foi criado uma tabela contendo a classificação de sucesso das aterrisagens.
- Para a etapa de análise de dados foram usadas ferramentas tratamento e análise de dados por SQL e quereis entre outras técnicas.

EDA with Data Visualization

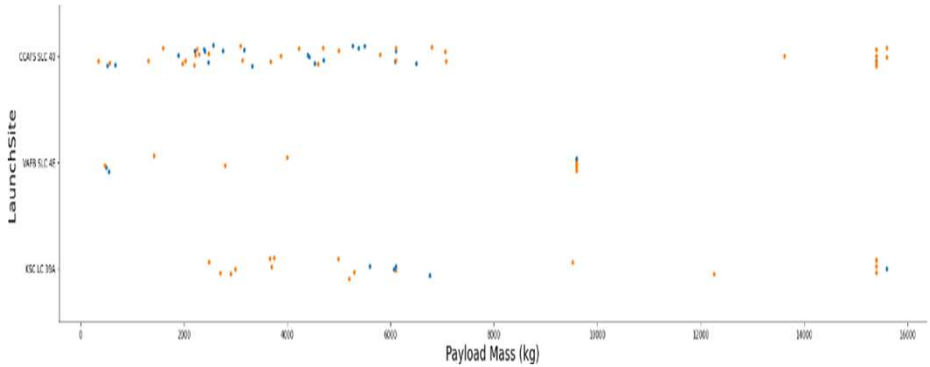
- Para visualização foram realizadas a criação de scatter e bar plots, uso do Folium para análise geográfica etc...

EDA with Data Visualization

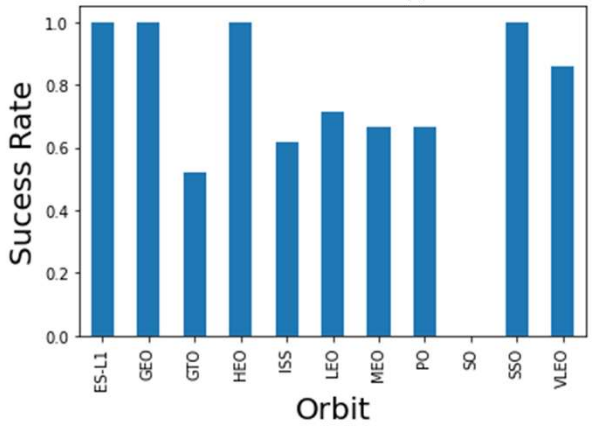
- FlightNumber vs. **Launch Site**



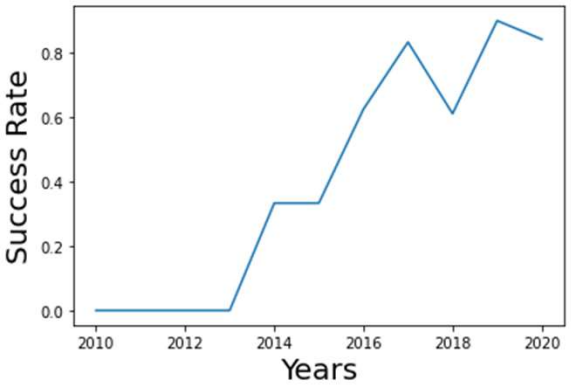
Payload and Launch Site



success rate of each orbit type



launch success yearly trend



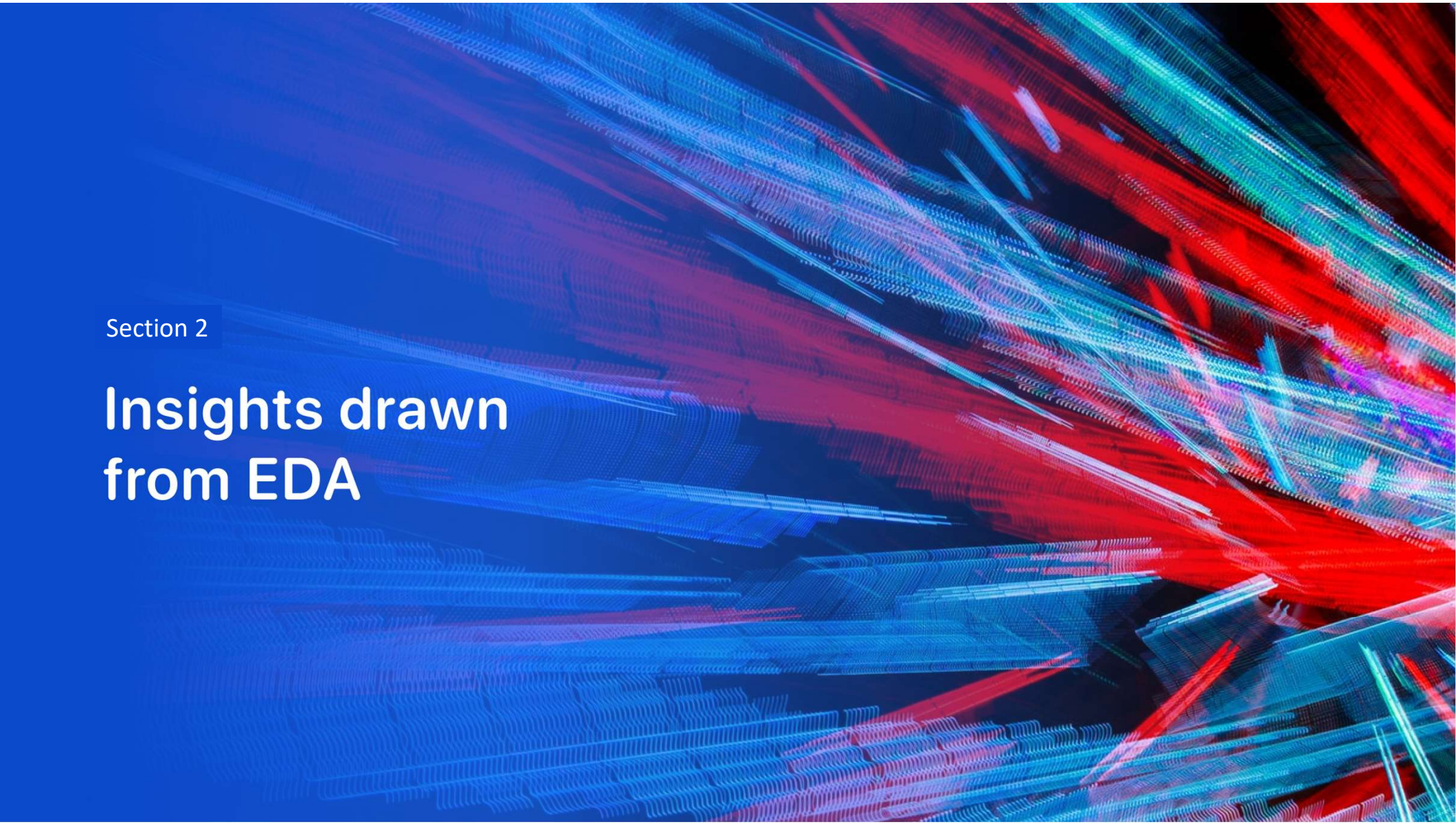
Build a Dashboard with Plotly Dash

- GitHub URL of your completed Plotly Dash lab, as an external reference and peer-review purpose

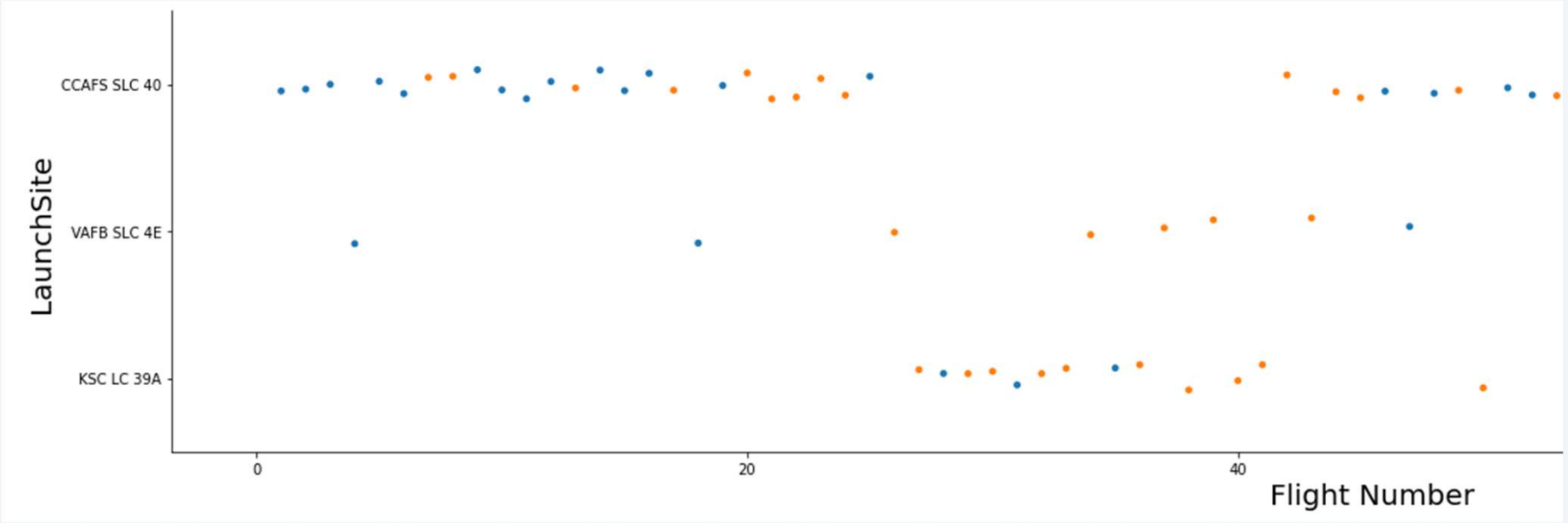
https://github.com/Mateusdearaujocarvalho/SpaceRace/blob/6ed7871f76f98d18e033dbf81758e636616cc985/Dash%20Ploty.ipynb

Predictive Analysis (Classification)

- Para realizar a modelagem, os dados foram preparados.
- Transformados em matrizes do tipo numpy.
- Separados em dados de teste e validação
- Realizados o treinamento dos modelos:
  - LR, SVM, KNN e Decision Tree;
- Avaliou-se o desempenho dos modelos Segundo a acurracia obtida e matriz de confusão.

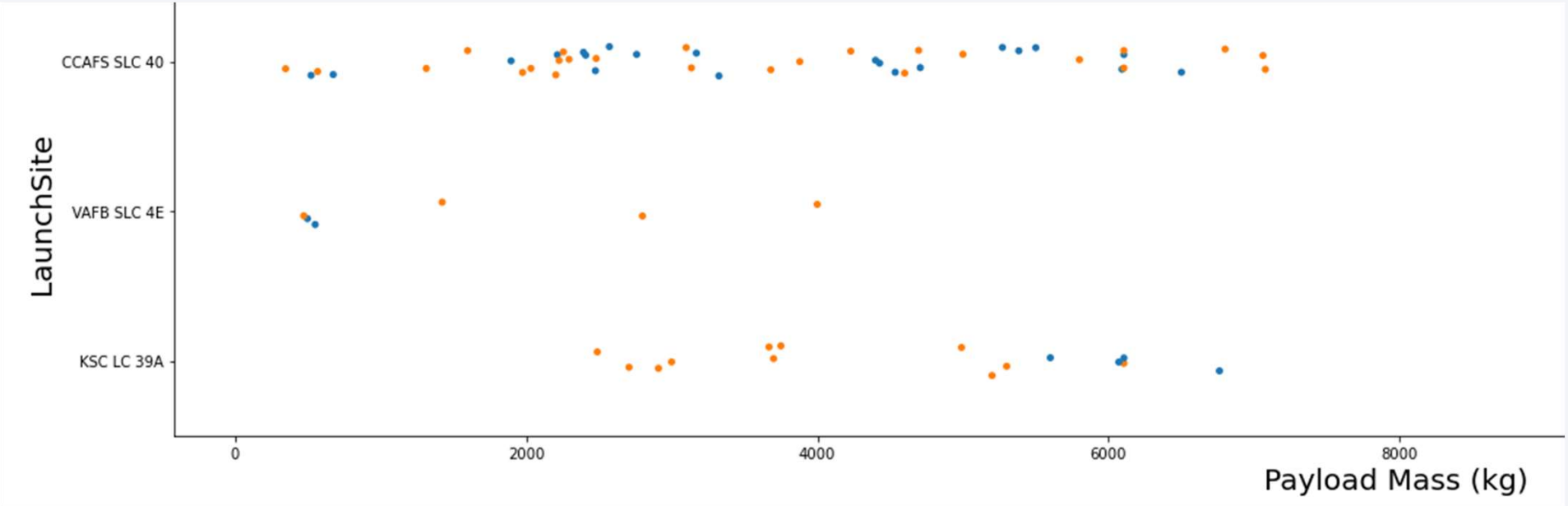

Section 2
Insights drawn
from EDA

Flight Number vs. Launch Site



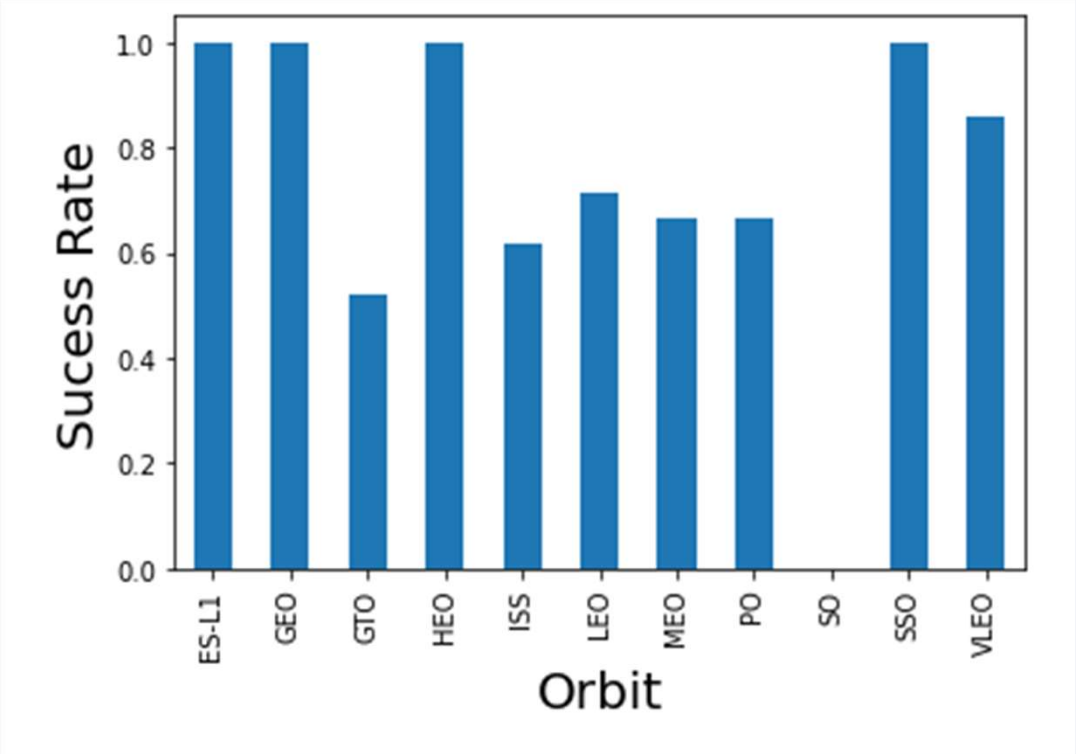
A medida que o número de voos aumenta percebe-se mais sucesso em todos os sites

Payload vs. Launch Site



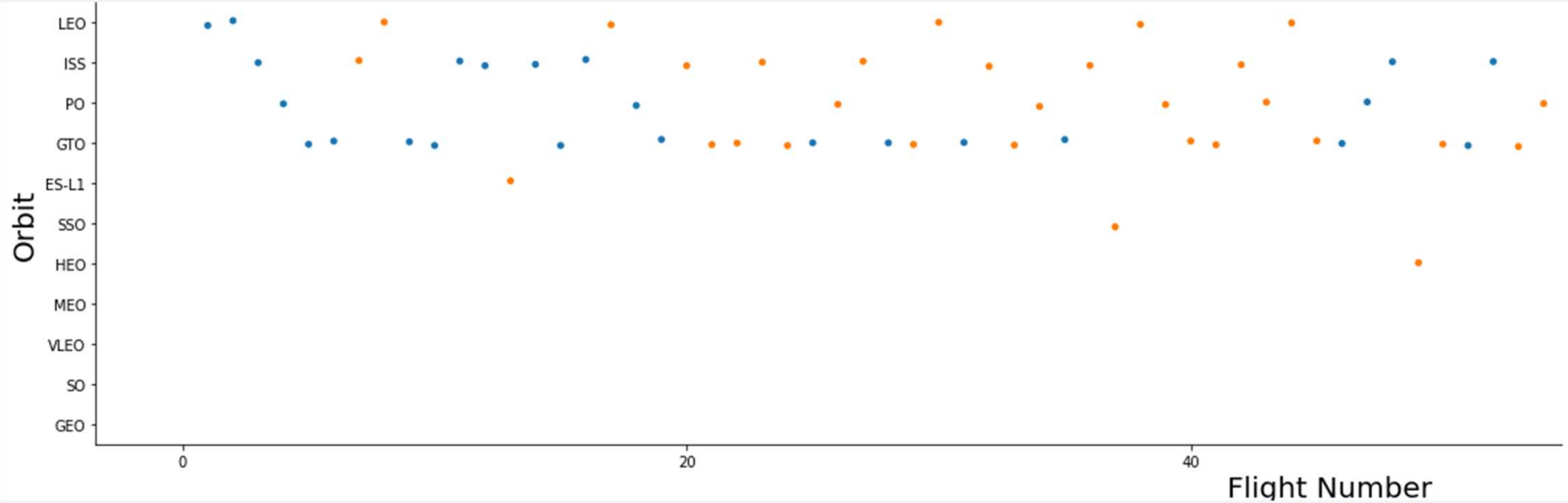
Não consegui achar uma relação direta entre carga útil e sucesso.

Success Rate vs. Orbit Type



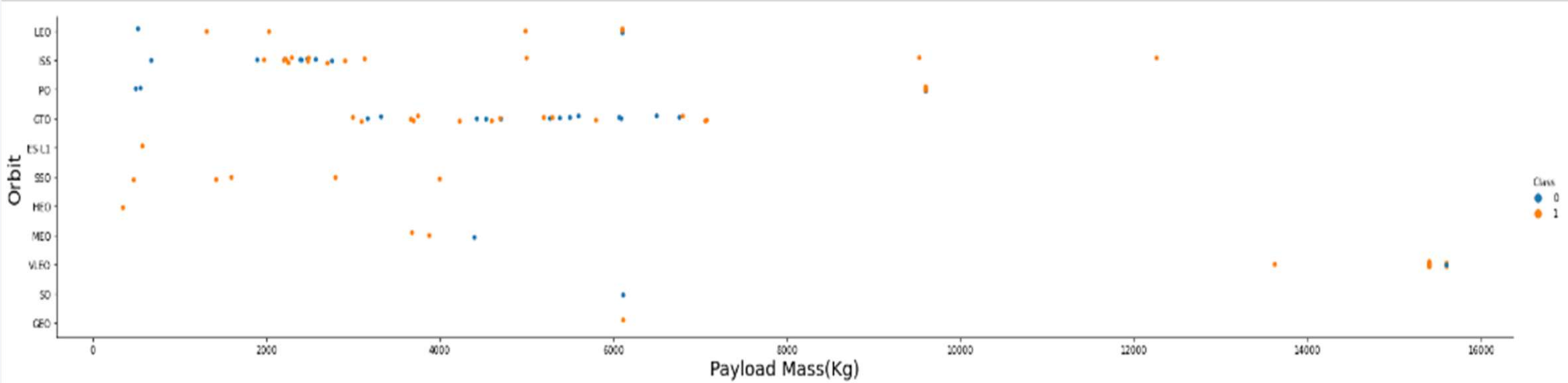
- Algumas órbitas possuem taxa de sucesso próximas a 100%

Flight Number vs. Orbit Type



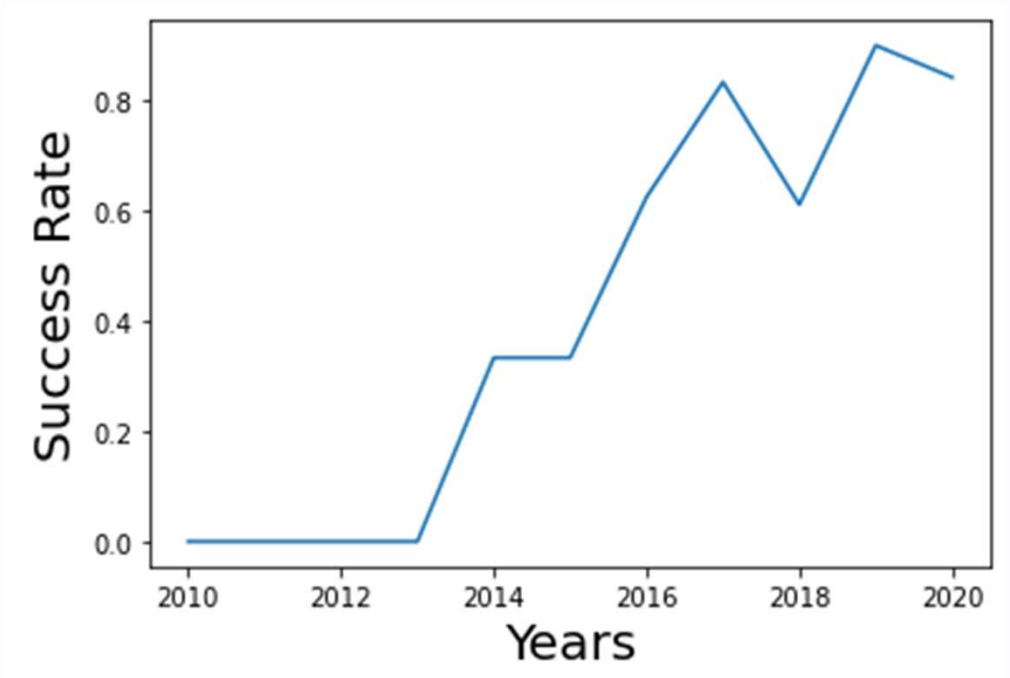
- Entretanto algumas tem baixo número de voos. A que parece ter uma boa relação de sucesso é a SSO

Payload vs. Orbit Type



- Para cargas maiores, as órbitas Polar, LEO e ISS apresentam resultados melhores

Launch Success Yearly Trend



- O resultado tem evoluído muito com o passar do tempo.

EDA with SQL

number of successful and failure mission outcomes

mission_outcome	COUNT
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

failed landing_outcomes in drone ship

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

All Launch Site Names

Out[4]:

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

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`

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, Average Payload and 1° Successful landing

total payload mass carried by boosters

```
Out[8]:    1
       45596
```

Average payload mass carried by boosters

```
1

2928
```

first successful landing

2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

booster_version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

mission_outcome	COUNT
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Boosters Carried Maximum Payload

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

2015 Launch Records

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

```
%sql select
LANDING__OUTCOME,BOOSTER_VERSION,LAUNCH_SITE from
SPACEXTBL where YEAR(DATE) = '2015' and
LANDING__OUTCOME = 'Failure (drone ship)'
```

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

- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

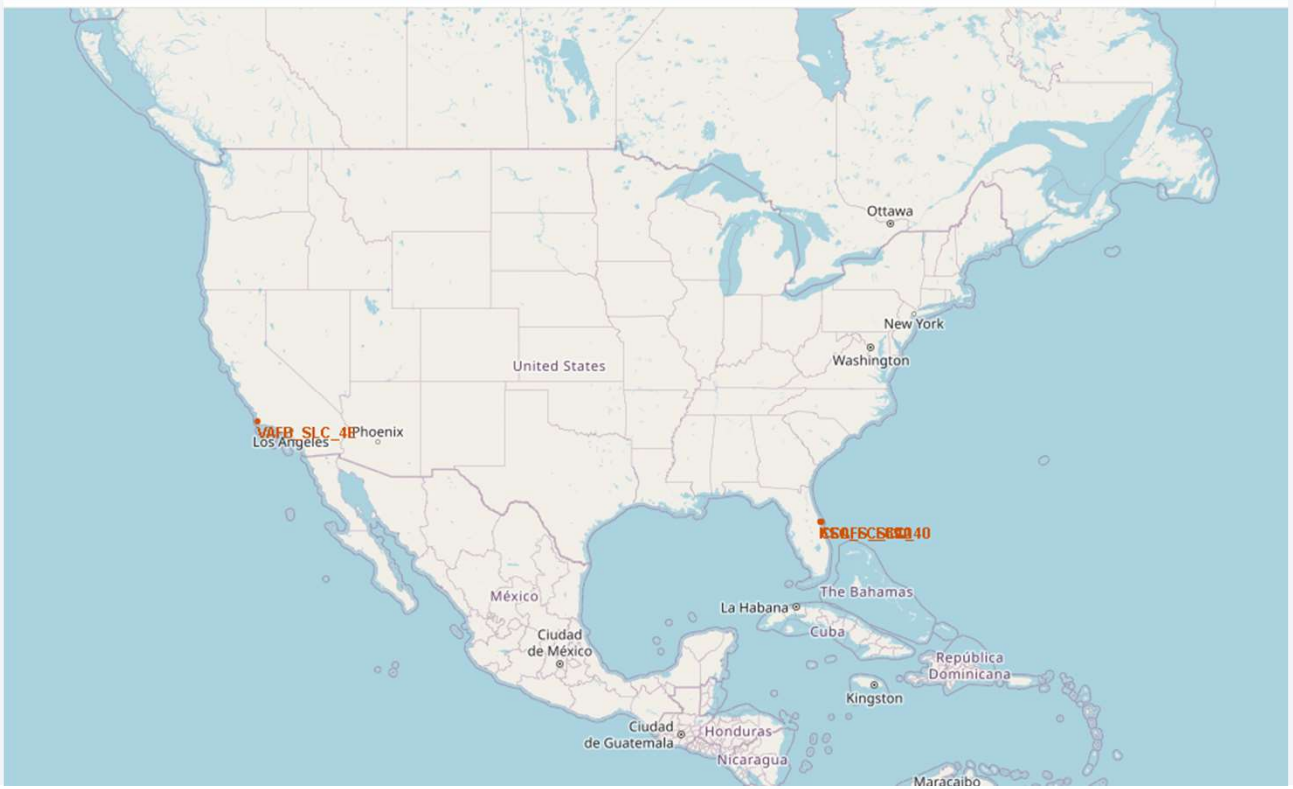
- Present your query result with a short explanation here

Section 3

Launch Sites Proximities Analysis

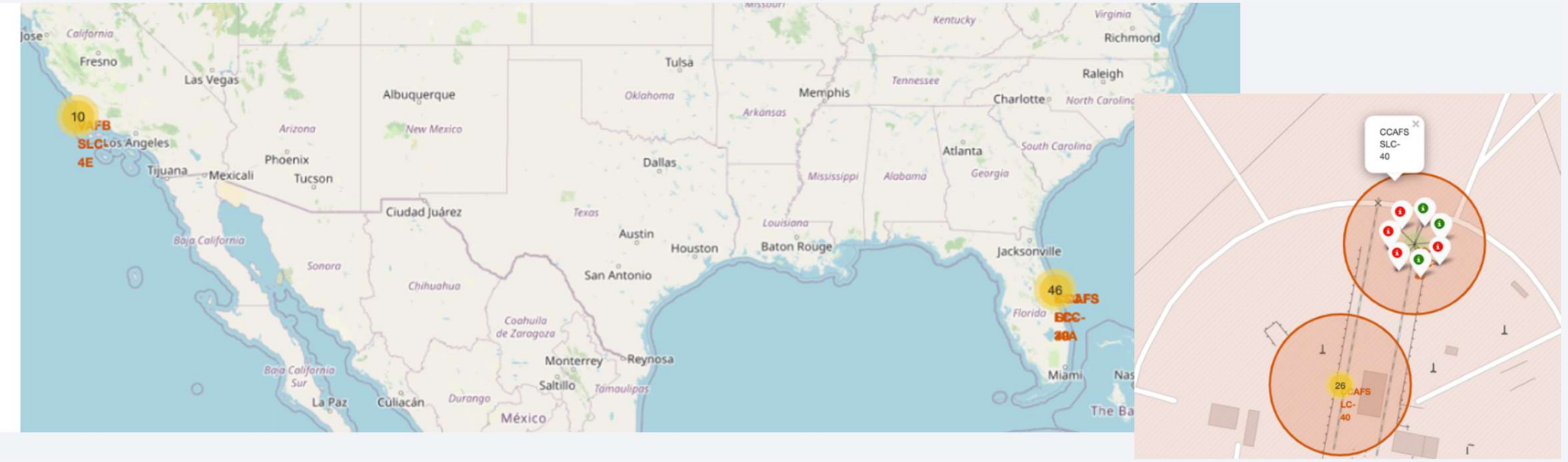
Laucn Sites

- Áreas de lançamento estão próximos ao litoral e estão nas duas costas dos EUA



Mapa de sucesso de lançamento por site

- Mapa mostrando os lançamentos e a condição de sucesso deles por site

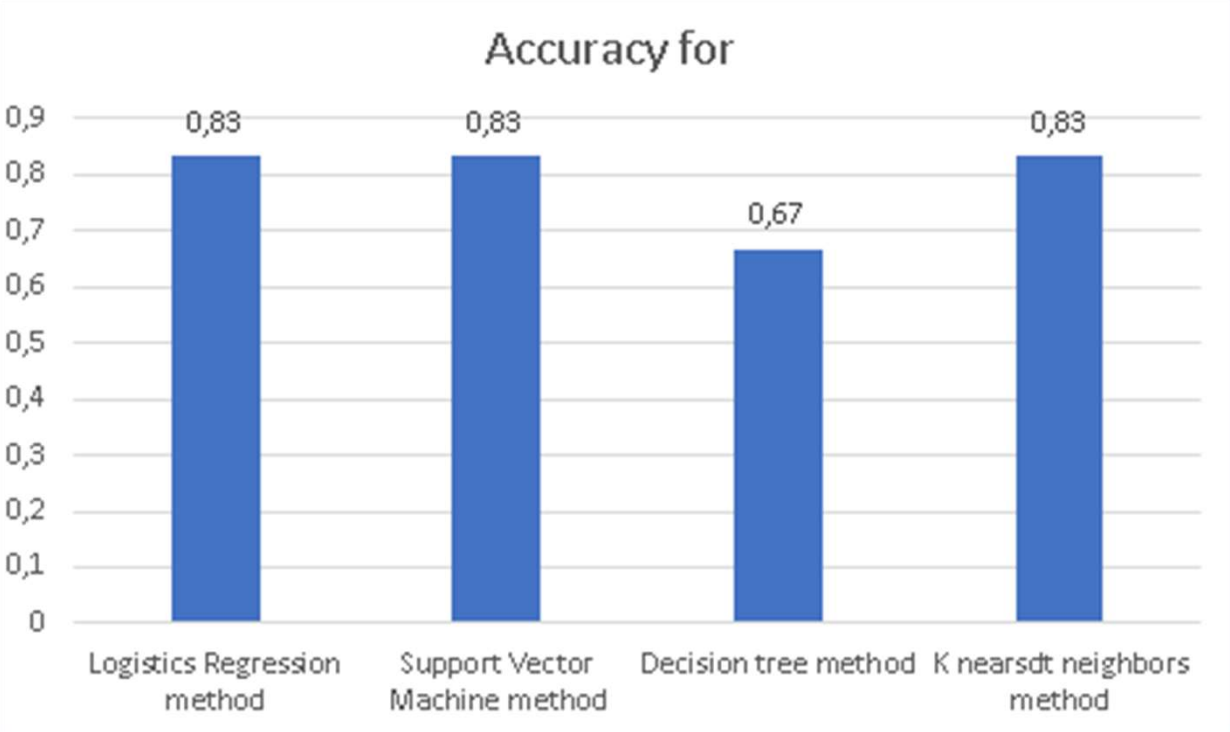


Mapa mostrando proximidade com litoral

Section 5

Predictive Analysis (Classification)

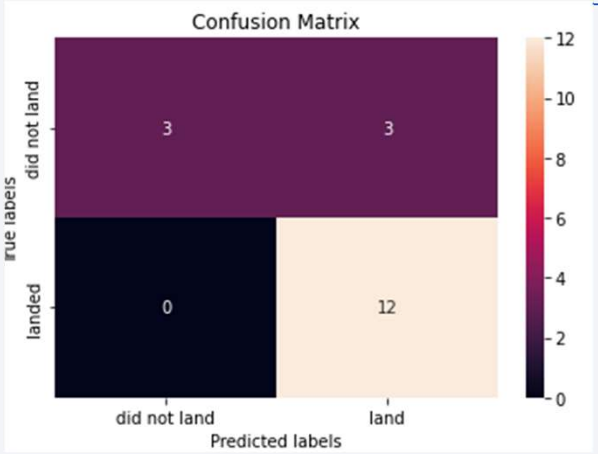
Classification Accuracy



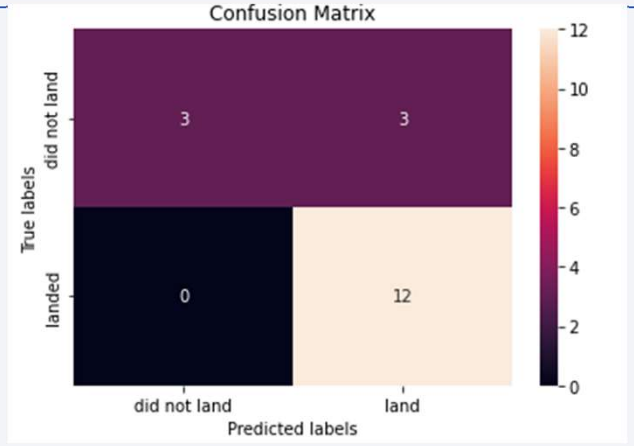
- LR, SVM and KNN has similar accurancy

Confusion Matrix

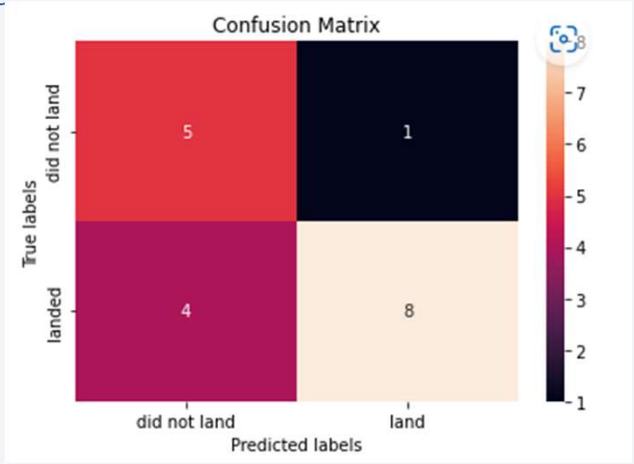
LR



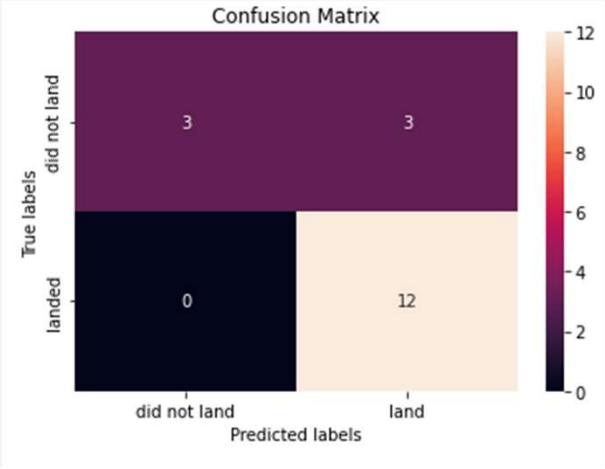
SVM



Tree



KNN



LR, SVM and KNN tiveram a mesma configuração de Matrix de confusão. O desempenho destes três modelos se mostrou superior ao do modelo de Árvore de decisão para prever a precisão do pouso bem-sucedido.

Conclusions

- Os resultados de sucesso em voos tem mellhorado consideravelmente ao longo do tempo
- O maior número de lançameentos ocorre no CCAFS SLC 40
- A órbita tem um impacto significativo no grau de sucesso e são especialmente importantes em caso de grandes cargas
- Percebe-se que existe uma preferência para lançamentos próximo ao litoral

- Os modelos de machine learning testados apresentaram uma boa acurácia, sendo que modelos de LR, SVM e KNN tiveram desempenho similar e melhor do que a árvore de decisão.

Appendix

Link para os arquivos no github

- Mateusdearaujocarvalho/SpaceRace: Winning Space Race with Data Science (github.com):

- https://github.com/Mateusdearaujocarvalho/SpaceRace

Thank you!