Space X Landing Prediction

IBM Data Science Capstone

Fethi BENMAKHLOUF

October 2021

PRESNTATION OUTLINE

- □ Executive Summary
- □ Introduction
- □ Methodology
- □ Results
    - ➢ Visualization – Charts
    - ➢ Dashboard
- □ Discussion
    - ➢ Findings & Implications
- □ Conclusion
- □ Appendix

EXECUTIVE SUMMARY

- Data collection,
- Data Wrangling
- Explatory Data Analysis with SQL Pandas and PyPlot
- Dashboard with Folium Plotly & Dash
- Apply Machine learning algorithms on data cleaned

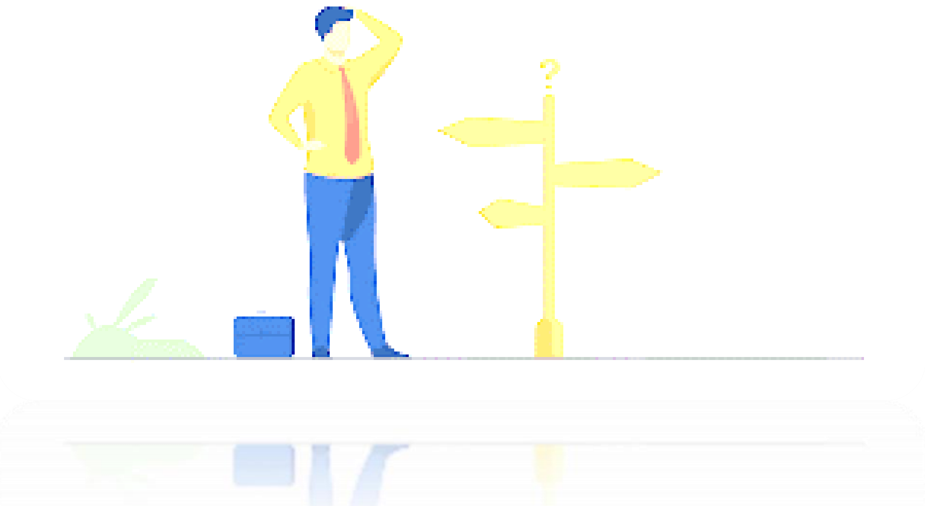
INTRODUCTION

This analysis aims to get help aero spatial researchers in their information mining process to collect informations and factors to have successful rocket landing

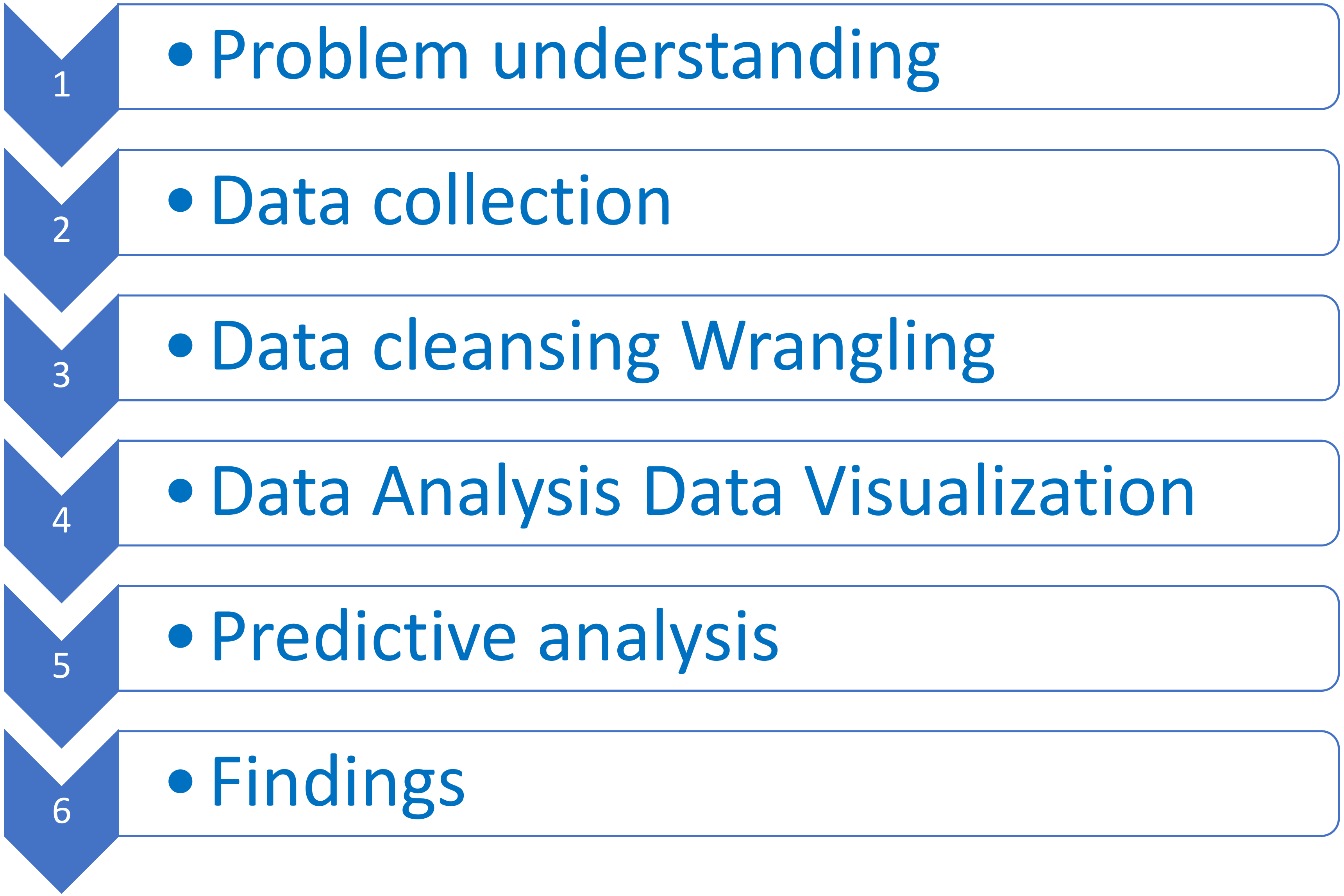
Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

What factors influences a successful rocket landing ?

What conditions should be available to achieve to get the best results and ensure the best rocket success landing rate?

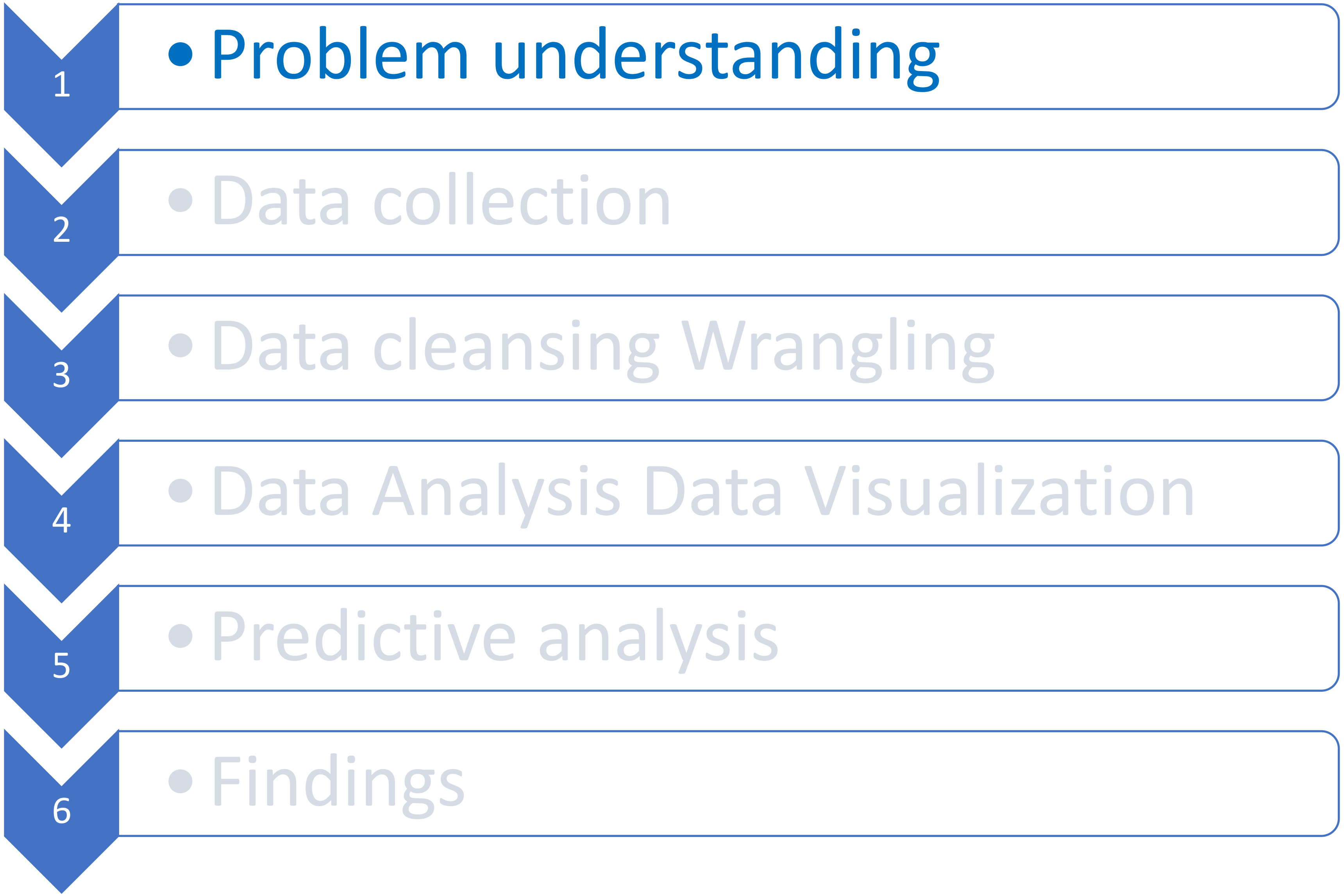


METHODOLOGY



1. Problem understanding
2. Data collection
3. Data cleansing Wrangling
4. Data Analysis Data Visualization
5. Predictive analysis
6. Findings

METHODOLOGY

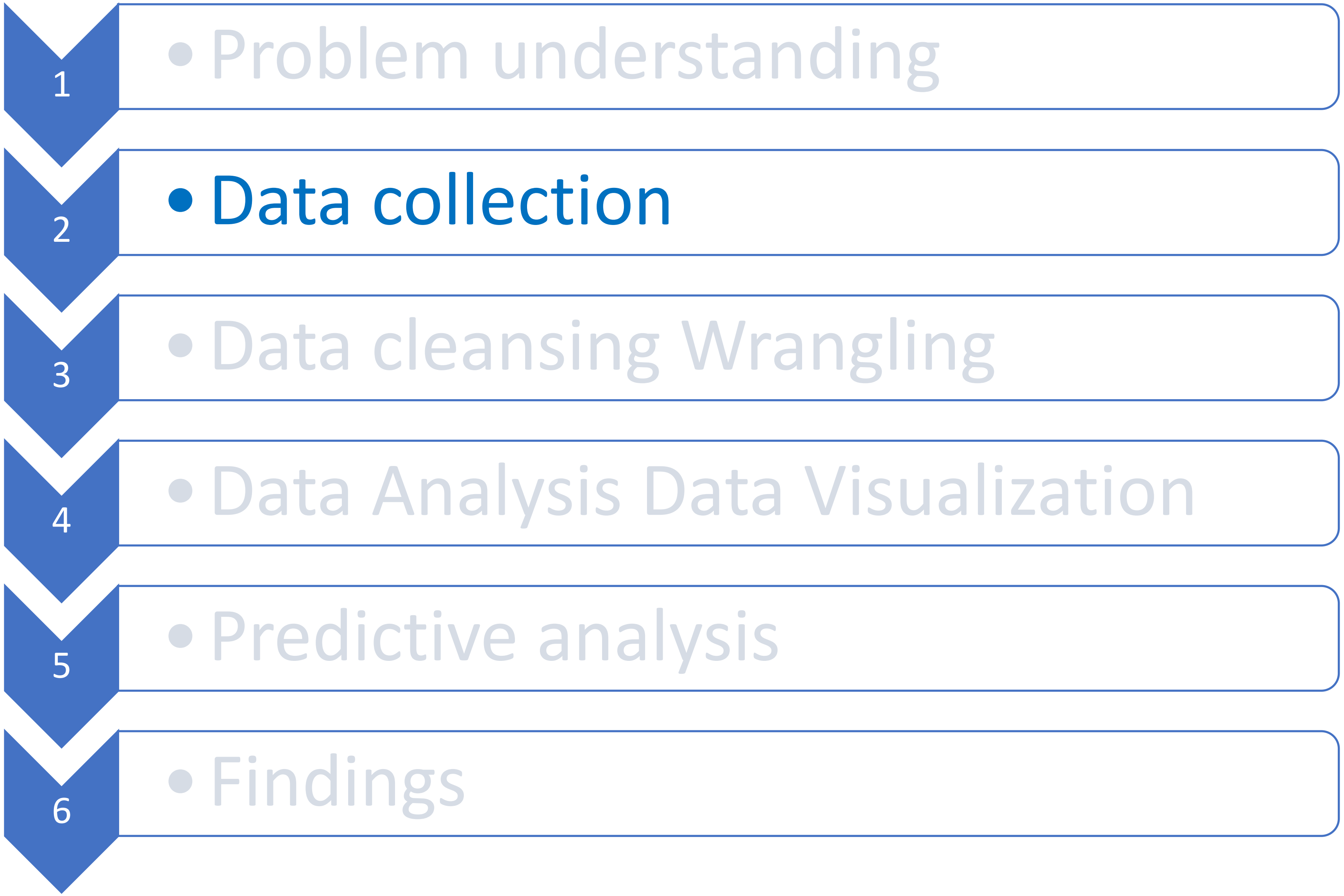


1. Problem understanding
2. Data collection
3. Data cleansing Wrangling
4. Data Analysis Data Visualization
5. Predictive analysis
6. Findings

Problem

- What cause a success?
- What cause a failure?
- What kind of data we need?
- How to obtain it?
- Data scraping it?

METHODOLOGY



1
- Problem understanding

2
- Data collection

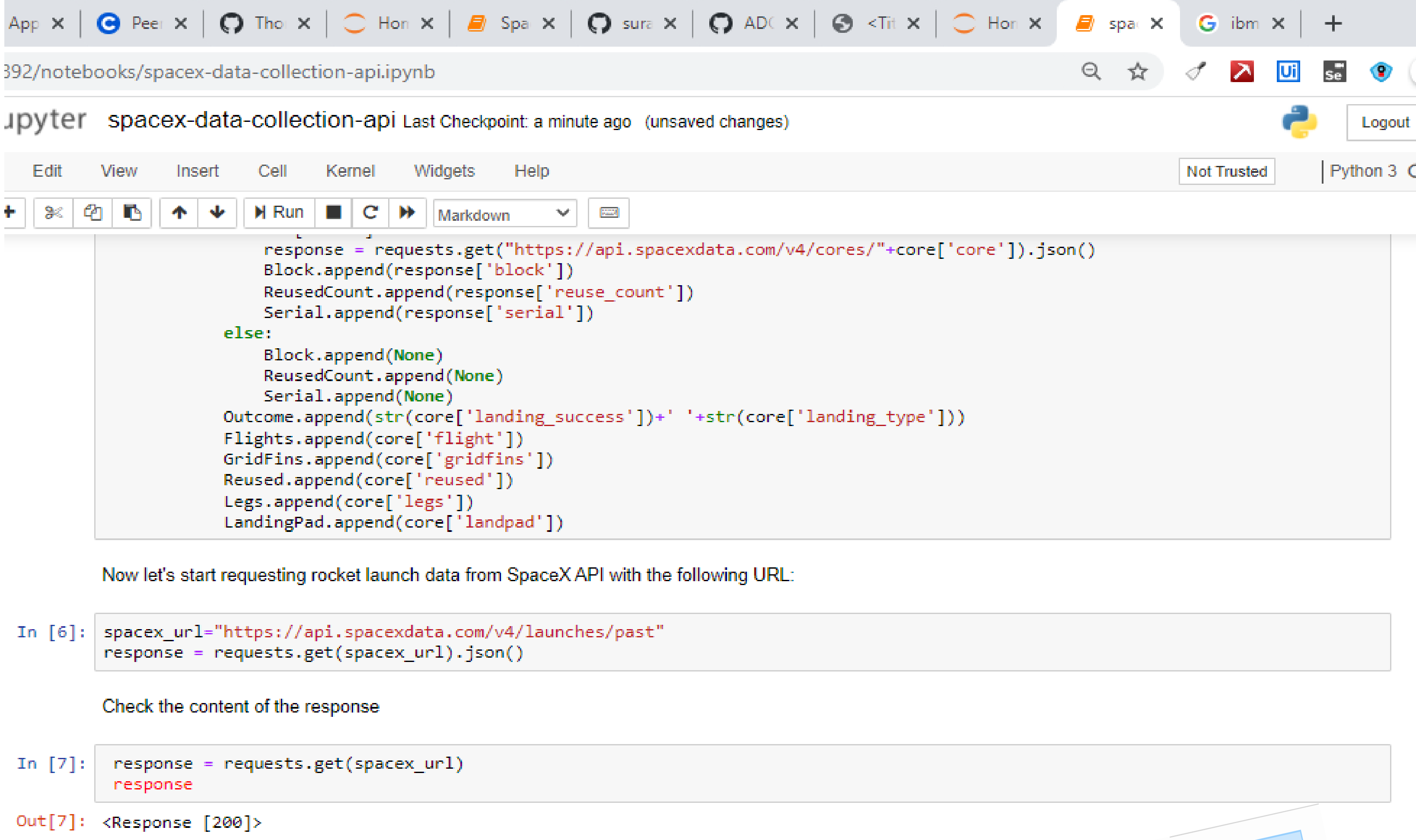
3
- Data cleansing Wrangling

4
- Data Analysis Data Visualization

5
- Predictive analysis

6
- Findings

Data collection


App
Peer
Tho
Hon
Spa
sura
ADC
<Tit
Hor
spa
ibm
392/notebooks/spacex-data-collection-api.ipynb
upyter
spacex-data-collection-api Last Checkpoint: a minute ago (unsaved changes)
Logout
Edit
View
Insert
Cell
Kernel
Widgets
Help
Not Trusted
Python 3
Run
Markdown

```
            response = requests.get("https://api.spacexdata.com/v4/cores/"+core['core']).json()
            Block.append(response['block'])
            ReusedCount.append(response['reuse_count'])
            Serial.append(response['serial'])
        else:
            Block.append(None)
            ReusedCount.append(None)
            Serial.append(None)
        Outcome.append(str(core['landing_success'])+' '+str(core['landing_type']))
        Flights.append(core['flight'])
        GridFins.append(core['gridfins'])
        Reused.append(core['reused'])
        Legs.append(core['legs'])
        LandingPad.append(core['landpad'])
```

Now let's start requesting rocket launch data from SpaceX API with the following URL:

In [6]:

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

Check the content of the response

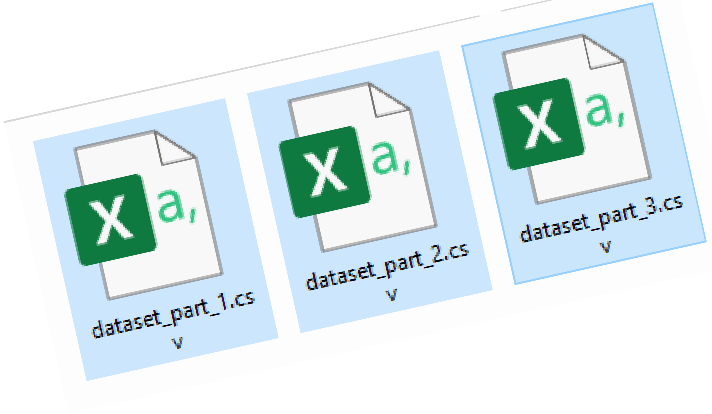
In [7]:

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

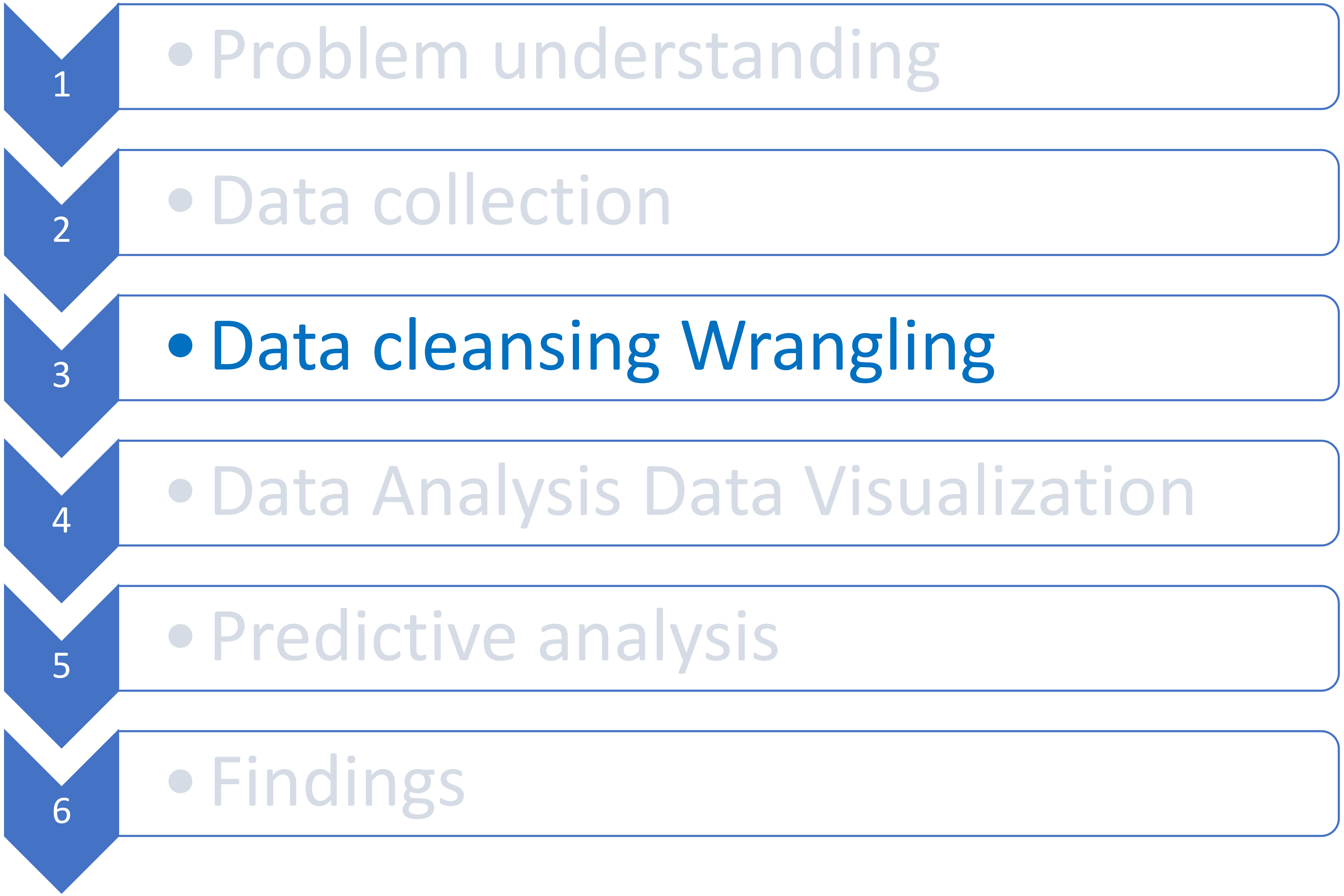
Out[7]:

```
<Response [200]>
```

- HTML API
- Data scraping with BeautifulSoup Library

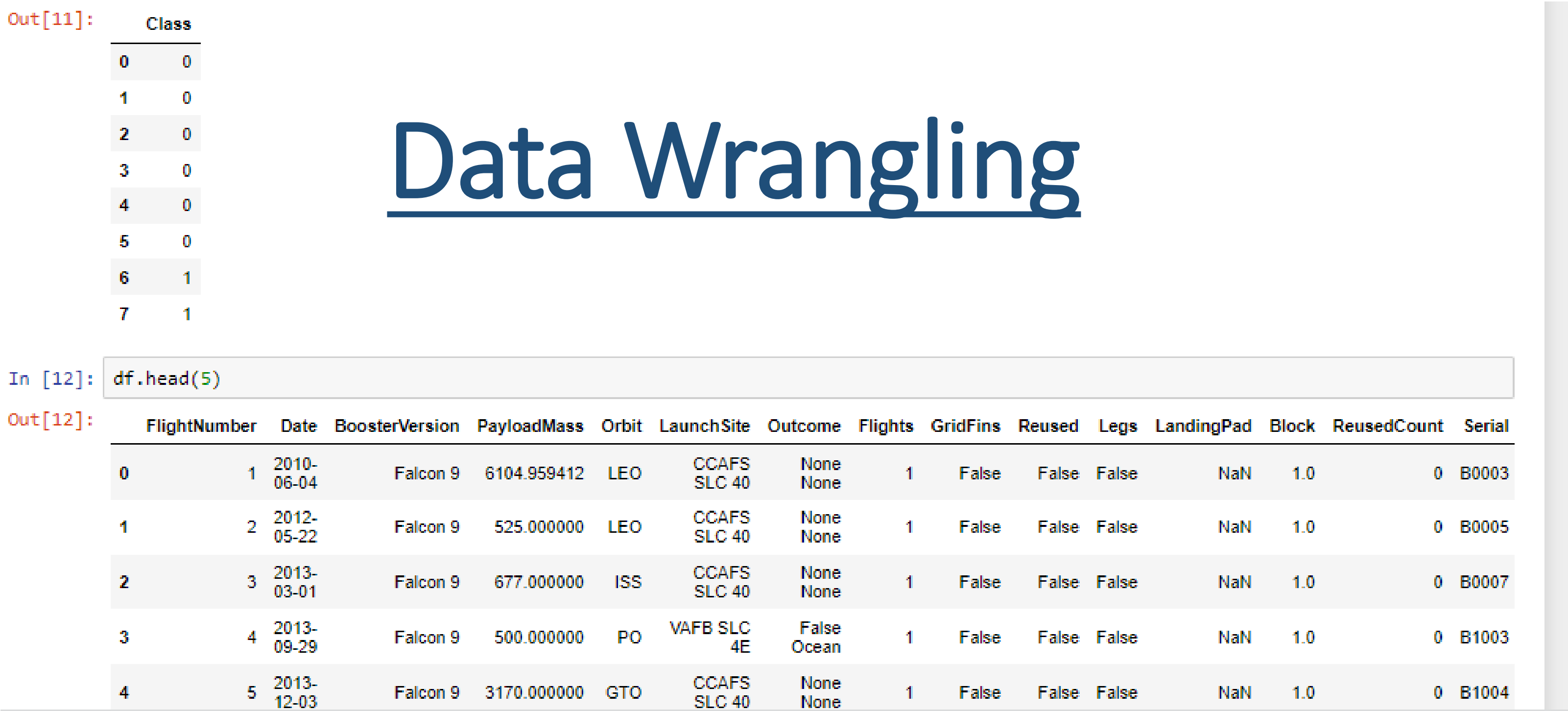


METHODOLOGY



1. Problem understanding
2. Data collection
3. Data cleansing Wrangling
4. Data Analysis Data Visualization
5. Predictive analysis
6. Findings

Data Wrangling



Out[11]:

	Class
0	0
1	0
2	0
3	0
4	0
5	0
6	1
7	1

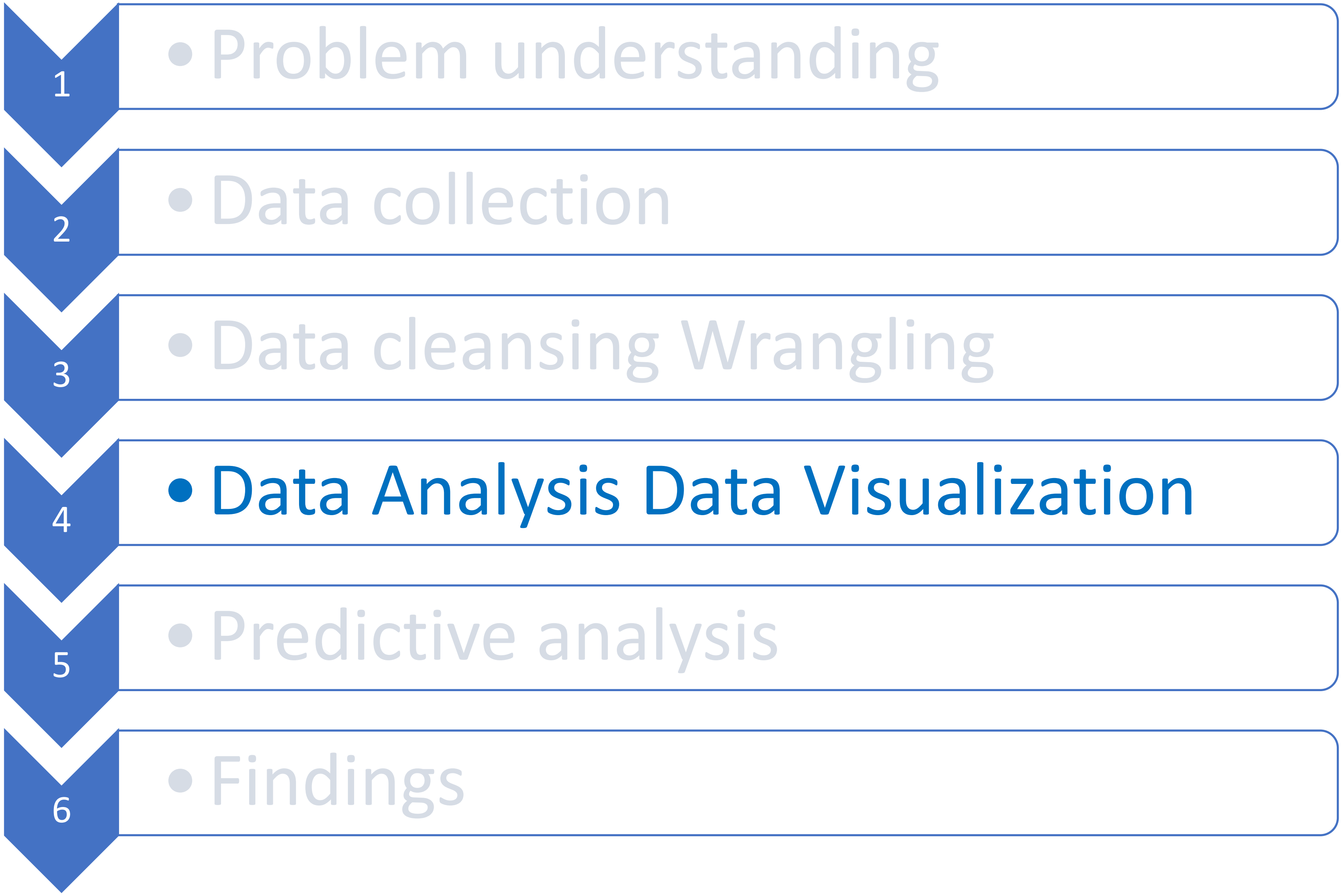
In [12]:
```
df.head(5)
```

Out[12]:

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

convert those outcomes into Training Labels with 1 means the booster successfully landed 0 means it was unsuccessful.

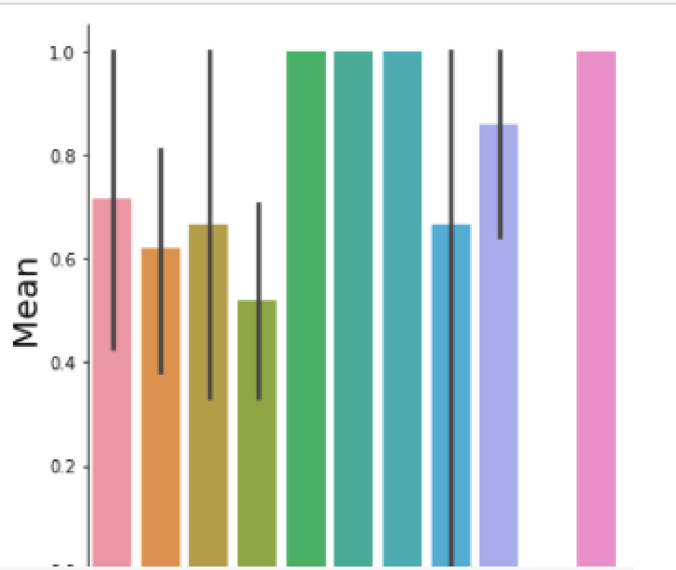
METHODOLOGY

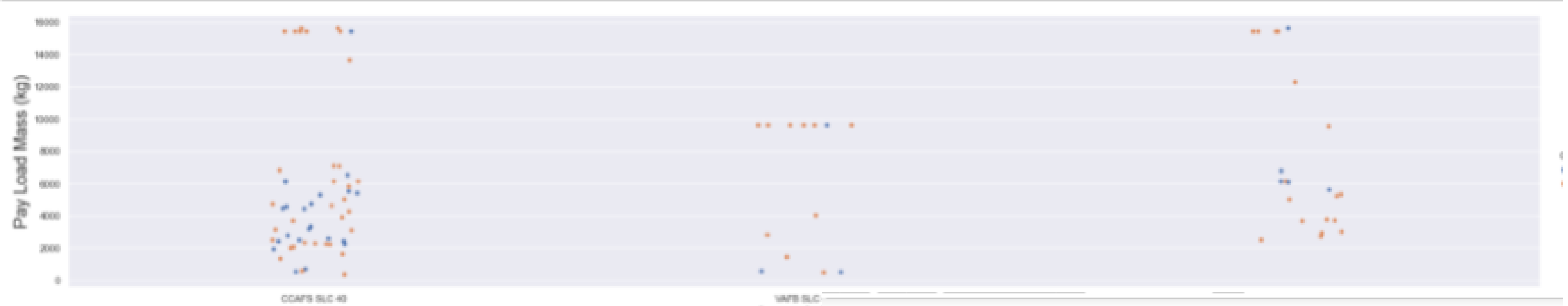


1. Problem understanding
2. Data collection
3. Data cleansing Wrangling
4. Data Analysis Data Visualization
5. Predictive analysis
6. Findings

Analysis and Visualization

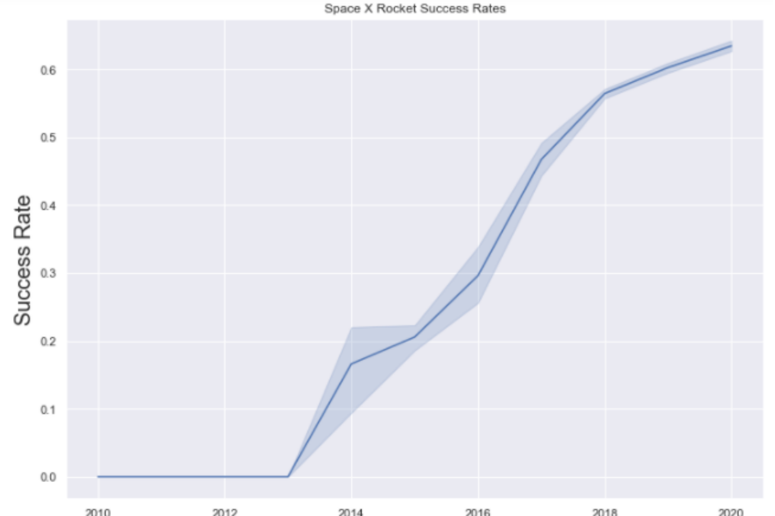
- Collectig inferential statistics with SQL & Py Pandqs
- Visualizing Data using Pyplot



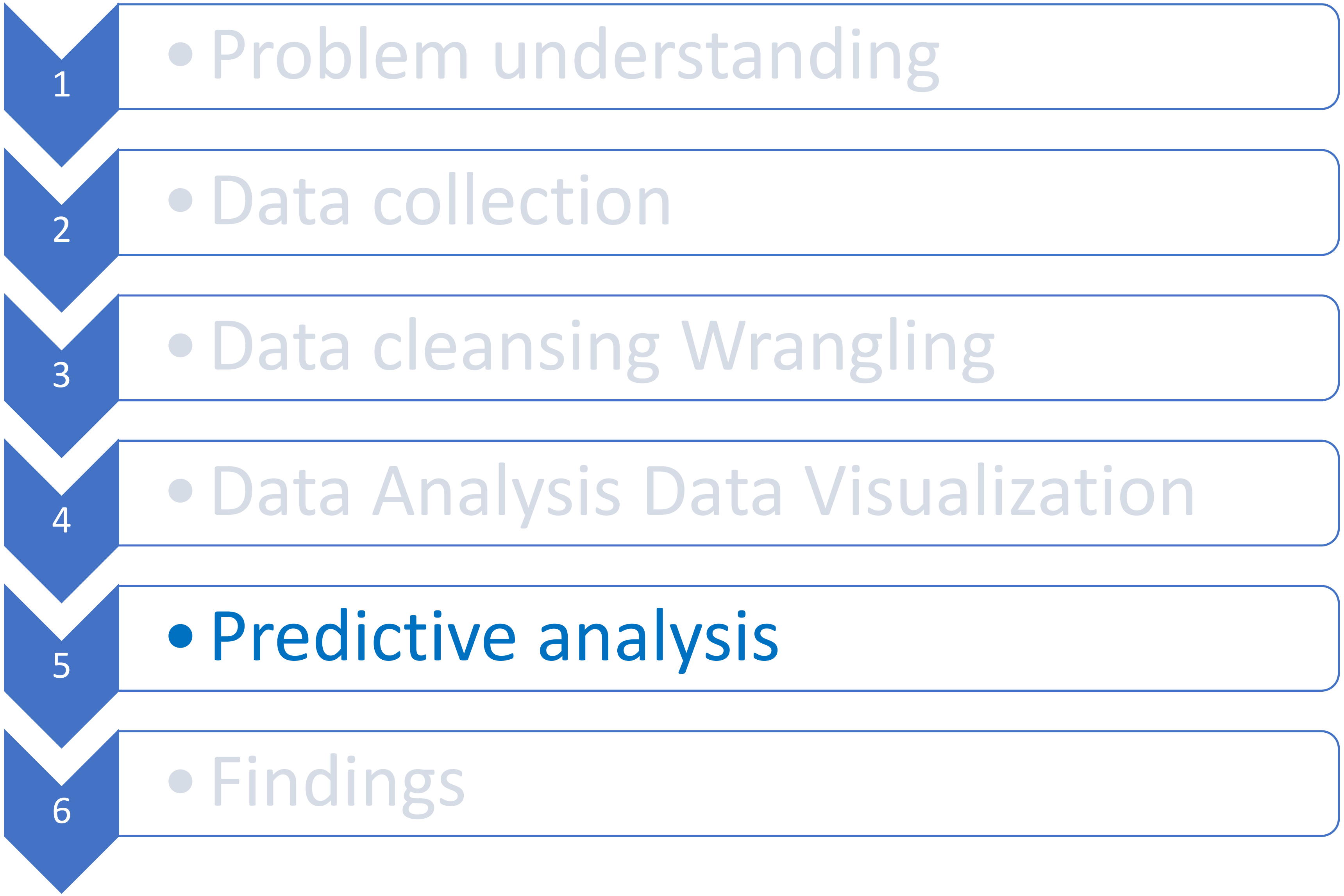


Identify correlations

Eliminate redundancy



METHODOLOGY



1
- Problem understanding

2
- Data collection

3
- Data cleansing Wrangling

4
- Data Analysis Data Visualization

5
- **Predictive analysis**

6
- Findings

Predictive Analysis

BUILDING MODEL

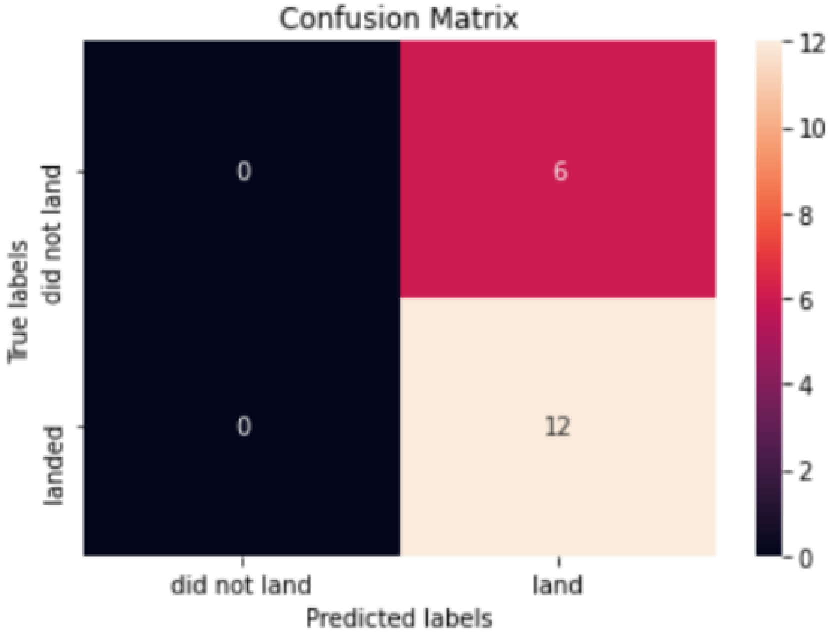
- Load our dataset into NumPy and Pandas
- Transform Data (scalar)
- Split data into training and test data sets
- Check the number of test samples
- Decide which type of machine learning algorithms to run
- Set parameters and algorithms to GridSearchCV
- Fit datasets into the GridSearchCV objects and train our dataset.

EVALUATING MODEL

- compute the accuracy for each model
- Get tuned hyperparameters for each type of algorithms
- Plot Confusion Matrix IMPROVING MODEL
- Feature Engineering
- Algorithm Tuning

FINDING THE BEST MODEL

- The model with the highest accuracy score is retained



METHODOLOGY

1
- Problem understanding

2
- Data collection

3
- Data cleansing Wrangling

4
- Data Analysis Data Visualization

5
- Predictive analysis

6
- Findings

DISCUSSION

- Best Algorithms: Tree Classifier
- Best launch site : KSC LC-39A
- Best Orbite : GEO,HEO,SSO,ES-L1
- Best payloads : with low weight
- Success rate increase by year after year.