IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Fred
2/23/2024

Outline

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

Executive Summary

- Summary of methodologies
  - Data Collection
  - EDA
  - ML prediction
- Summary of all results
  - ML Prediction is viable by collecting enough data and creating a useful ML model.

Introduction

- Project background and context:
  - 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.
- Problems you want to find answers:
  - Predict if the Falcon 9 first stage will land successfully

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - With SpaceX launch data that is gathered from an API, specifically the SpaceX REST API. This API will give the data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
- Perform data wrangling
  - perform some Exploratory Data Analysis (EDA) to find some patterns in the data
- 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

- Describe how data sets were collected.
  - With SpaceX launch data that is gathered from an API, specifically the SpaceX REST API. This API will give the data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
  - Perform some Exploratory Data Analysis (EDA) to find some patterns in the data
- You need to present your data collection process use key phrases and flowcharts
  - Data collection
  - Data Wrangling

Data Collection – SpaceX API

Request and parse the SpaceX launch data using the GET request

Filter the dataframe to only include `Falcon 9` launches

Dealing with Missing Values

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

Data Wrangling

Calculate the number of launches on each site

Calculate the number and occurrence of each orbit

EDA with Data Visualization

- Summarize what charts were plotted and why you used those charts
  - Scatter plot * 3
  - Bar chart * 1
  - Scatter point * 2
  - Line Chart *1

EDA with SQL

- Using bullet point format, summarize the SQL queries you performed

%sql SELECT DISTINCT LAUNCH_SITE FROM SPACEXDATASET

%sql SELECT * FROM SPACEXDATASET WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5

%sql SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PADLOAD FROM SPACEXDATASET WHERE CUSTOMER='NASA (CRS)'

%sql SELECT AVG(PAYLOAD_MASS__KG_) AS PAYLOAD_MASS__KG FROM SPACEXDATASET WHERE BOOSTER_VERSION= 'F9 v1.1'

%sql SELECT MIN(DATE) FROM SPACEXDATASET WHERE LANDING__OUTCOME = 'Success (ground pad)'

%sql SELECT DISTINCT BOOSTER_VERSION FROM SPACEXDATASET WHERE

Build an Interactive Map with Folium

- Summarize what map objects such as markers, circles, lines, etc. you created and added to a folium map
  - Makers and circles
  - Marker clusters
  - MousePosition
  - Distance-showing Line

Build a Dashboard with Plotly Dash

- Summarize what plots/graphs and interactions you have added to a dashboard
  - Pie Chart *2
  - For success launches overview
  - For determining which launch site has the highest launch success rate

Predictive Analysis (Classification)

split the data into training and testing data

Used different methods to train data and find the best parameters

Calculate the accuracy on the test data

Plot the confusion matrix to assist the work

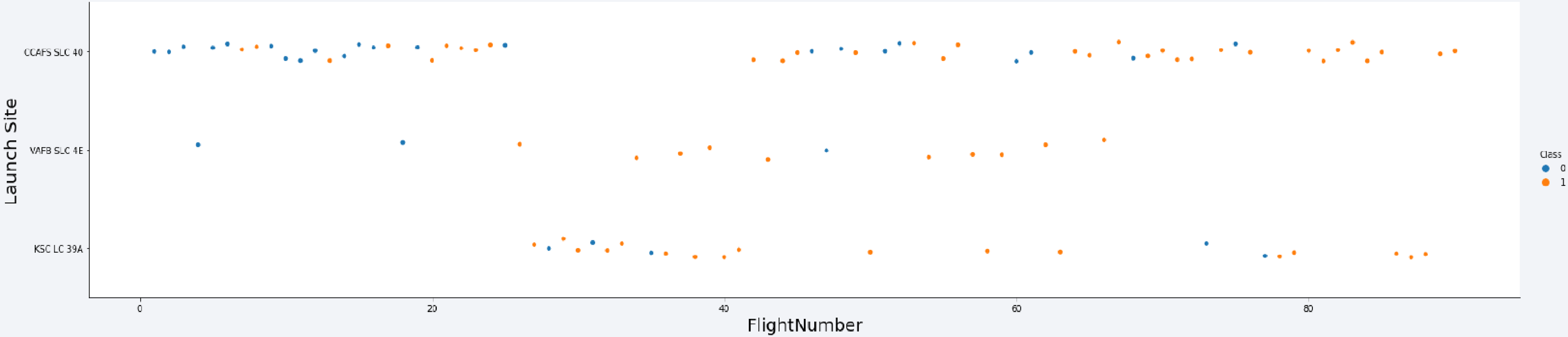
Results

- Exploratory data analysis results
  - East launch sites has more launches.
- Predictive analysis results
  - Higher the flight number, higher success rate; Higher the payload, lower the success rate

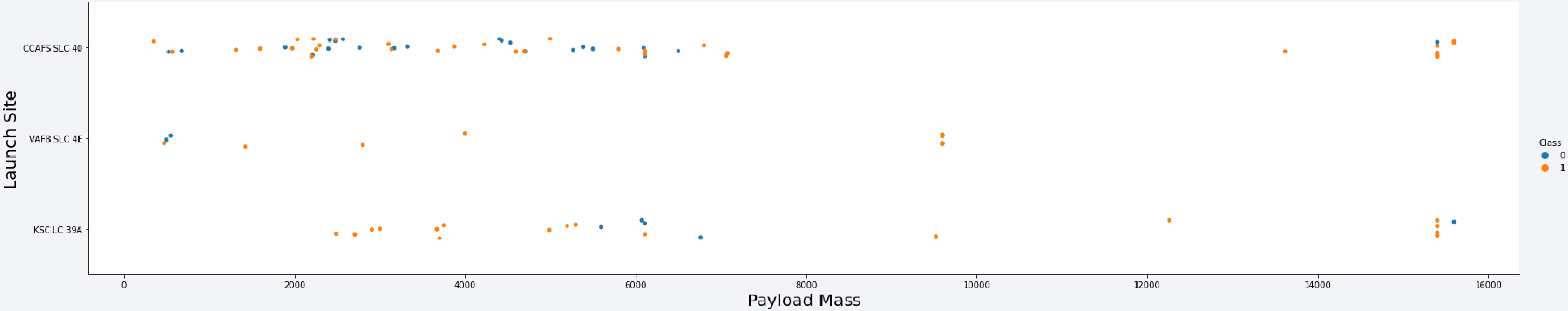
Section 2

Insights drawn from EDA

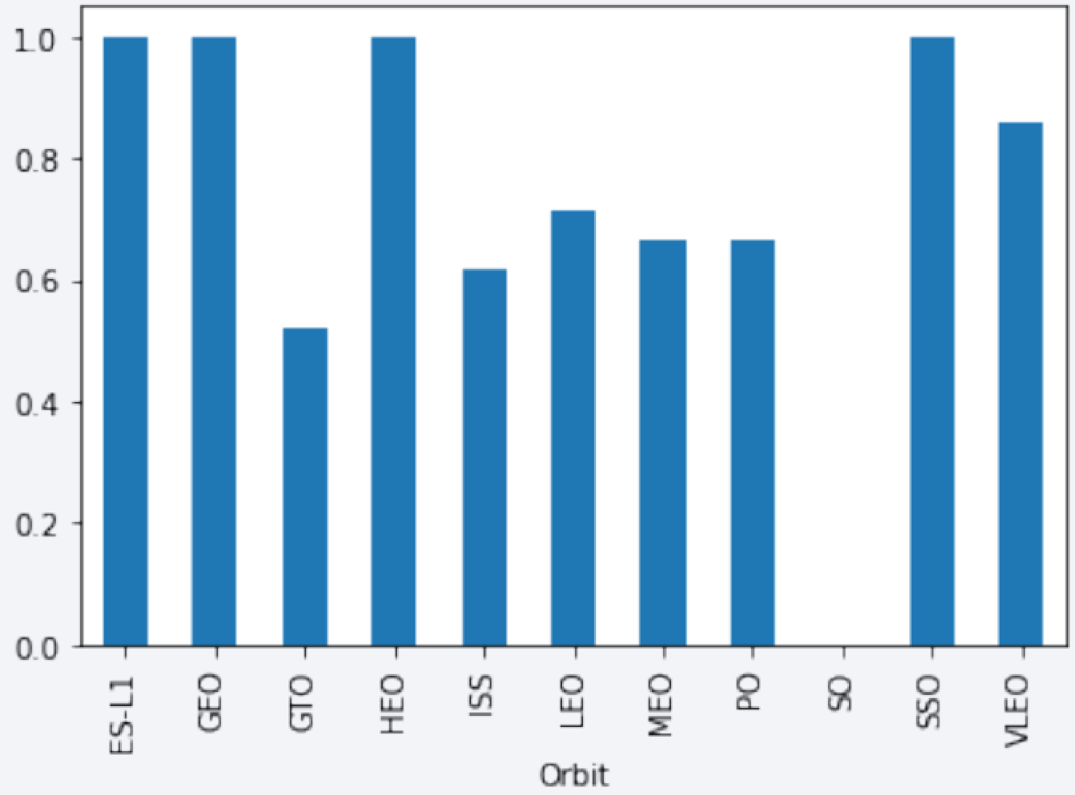
Flight Number vs. Launch Site



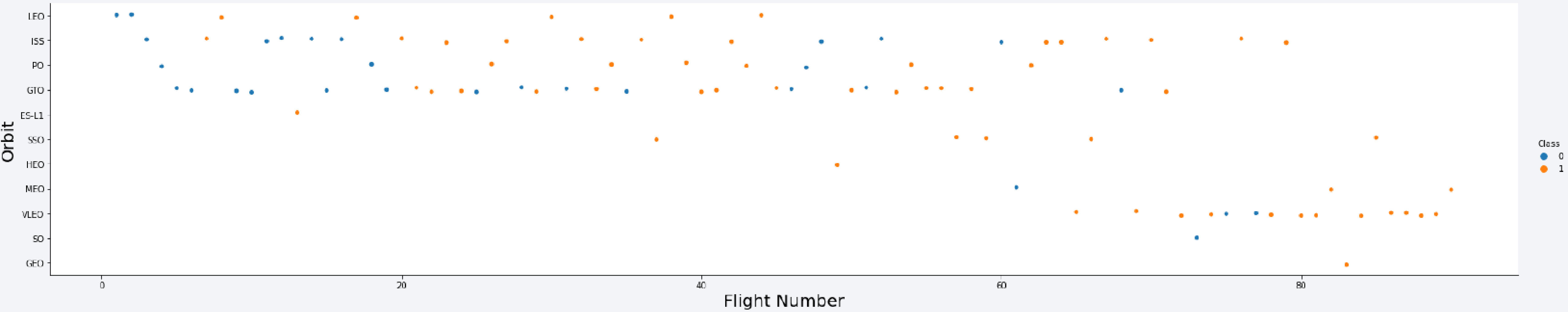
Payload vs. Launch Site



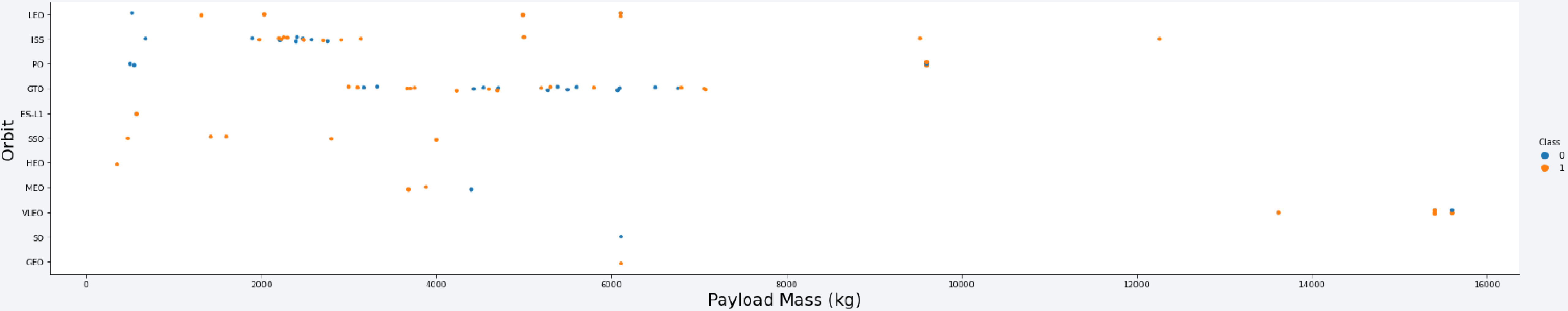
Success Rate vs. Orbit Type



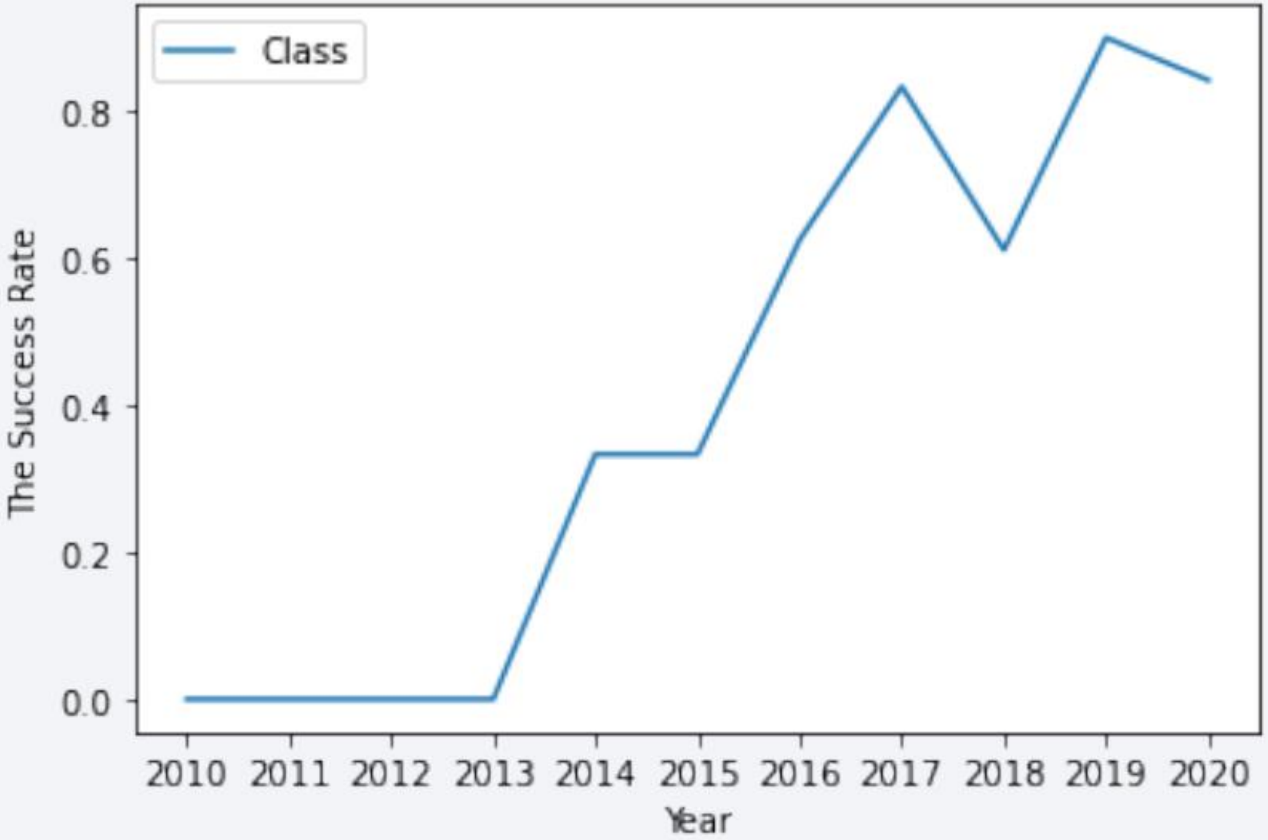
Flight Number vs. Orbit Type



Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names

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

Launch Site Names 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

- Calculate the total payload carried by boosters from NASA
    - 45596 KG

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
    - 2928 kg

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
  - 12/22/2015

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

Name of Boosters with success in drone ship
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
  - Success: 100
  - Failure: 1

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass

booster_version	payload_mass__kg_
F9 B5 B1048.4	15600
F9 B5 B1049.4	15600
F9 B5 B1051.3	15600
F9 B5 B1056.4	15600
F9 B5 B1048.5	15600
F9 B5 B1051.4	15600
F9 B5 B1049.5	15600
F9 B5 B1060.2	15600
F9 B5 B1058.3	15600
F9 B5 B1051.6	15600
F9 B5 B1060.3	15600
F9 B5 B1049.7	15600

2015 Launch Records

- List the failed landing outcomes in drone ship, their booster versions, and launch site names for in year 2015

Month	Year	Landing_Outcome	Booster_Version	Launch_Site
01	2015	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
04	2015	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

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

Landing Outcome	Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Section 3

Launch Sites Proximities Analysis

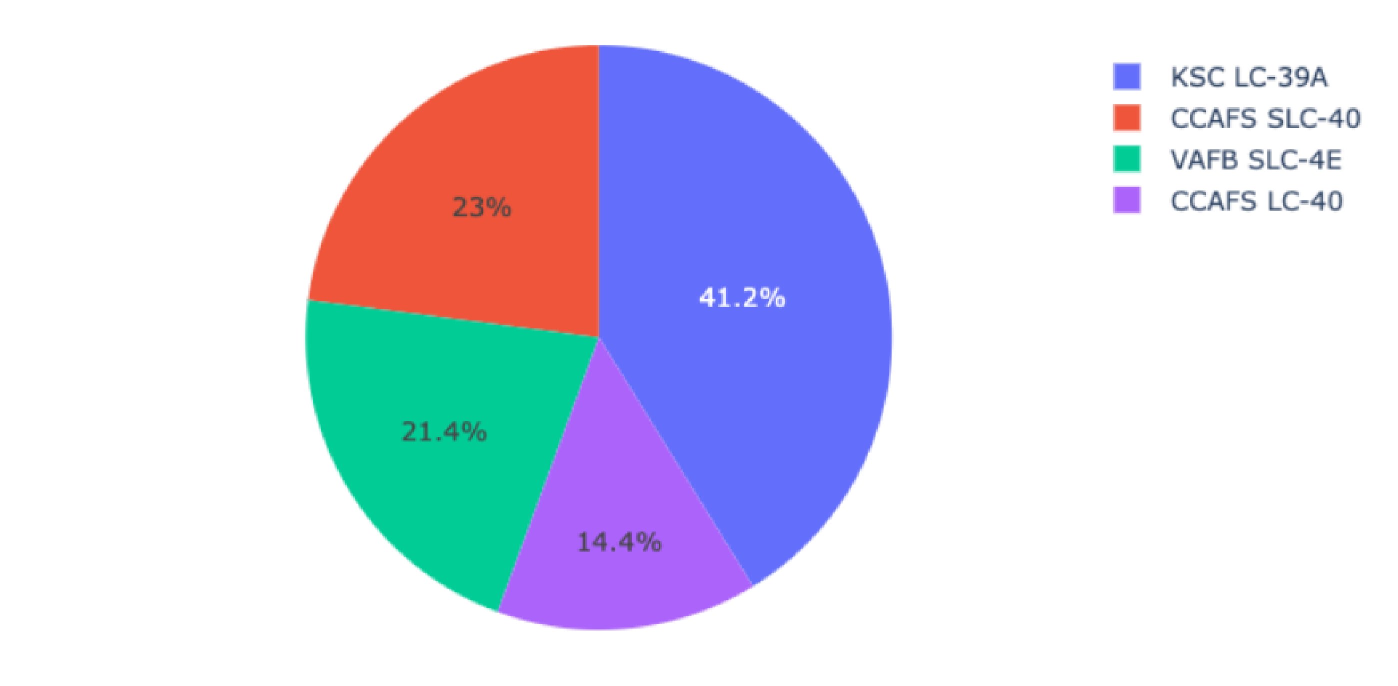
Launch Site Locations

Launch Results

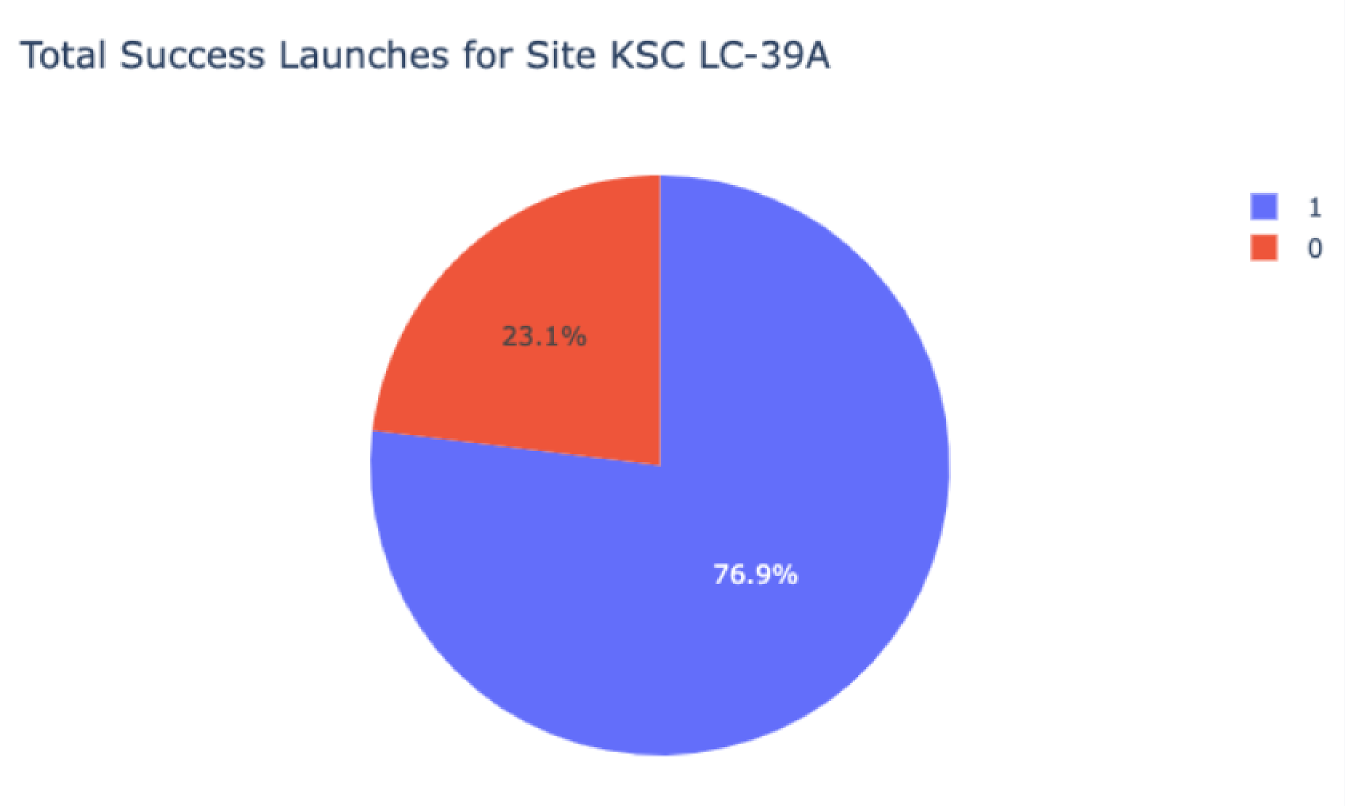
Section 4

Build a Dashboard with Plotly Dash

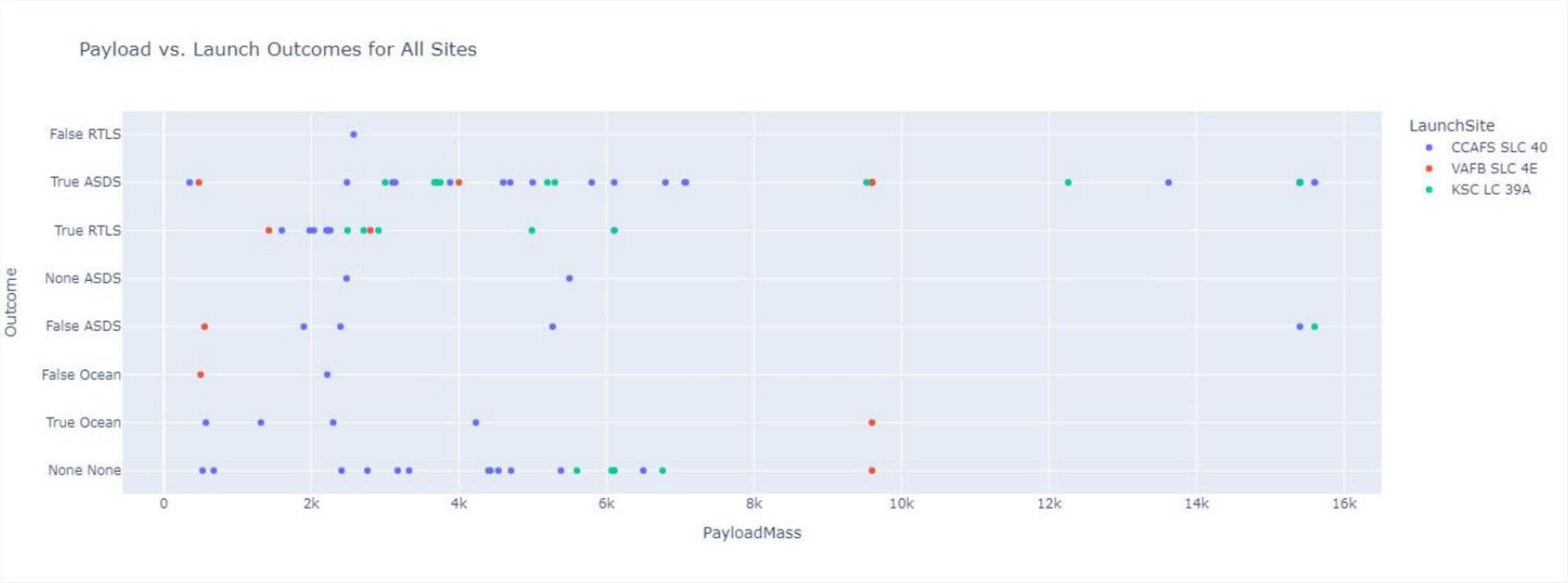
The success launches by site



The success launches for KSC LC-39A



The correlation between Payload and Launch Outcomes



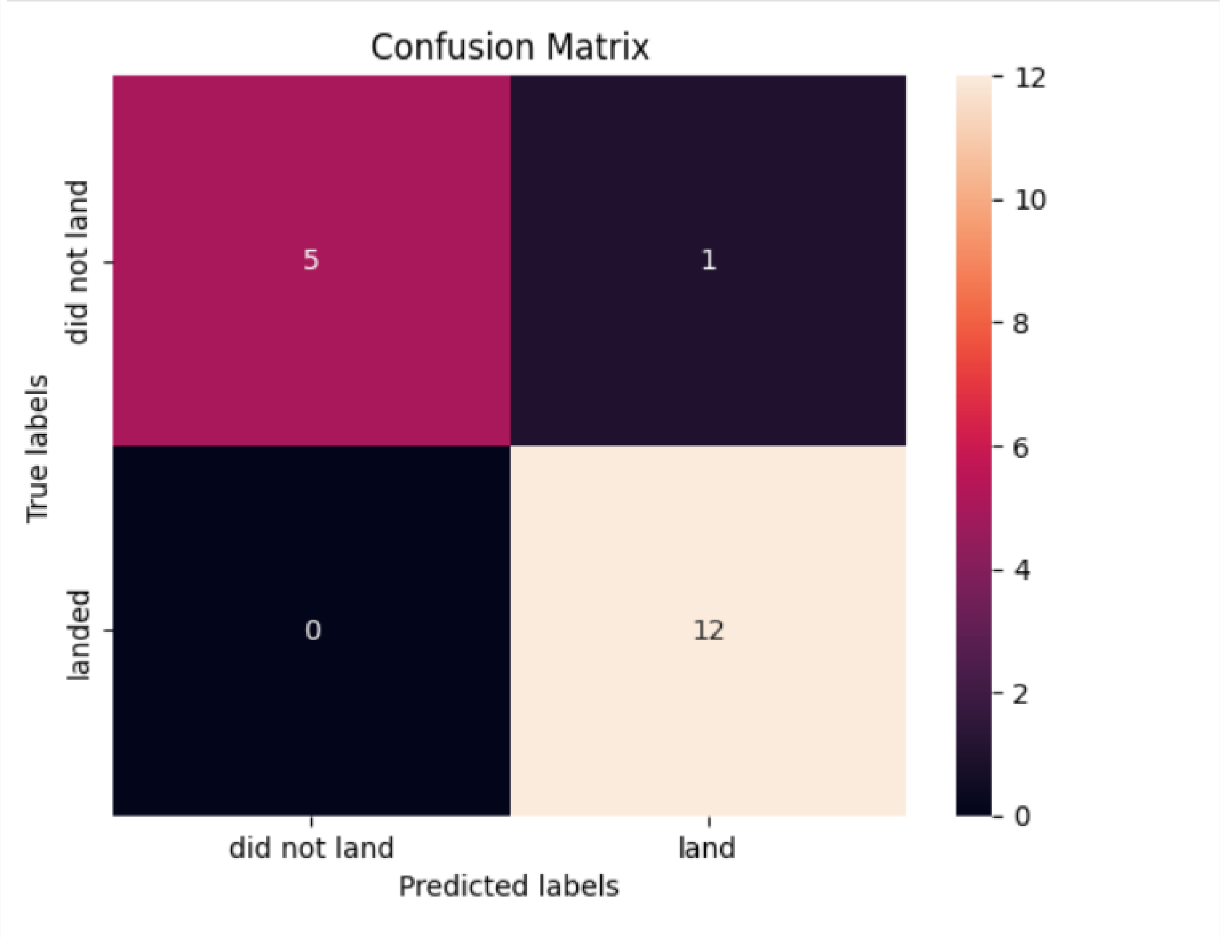
Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart
- Find which model has the highest classification accuracy

Confusion Matrix



Conclusions

- The features (booster version, payload, launch sites, etc.) can affect the success rate.
- ML prediction is useful!
- Different model evaluation methods result in different accuracy values, and are important to use multiple evaluation methods.
- Data Science is fun!

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!