Predicting successful SpaceX Falcon 9 launches

IBM Data Science Professional Certificate - Applied Capstone report

Jose Viosca Ros

August 15th, 2022

IBM Developer

SKILLS NETWORK

OUTLINE

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

EXECUTIVE SUMMARY

- Business profitability of space rocket trips:
  - Largely determined by launch cost
  - Reusing rockets provides great competitive advantage
- Data Science to predict successful landings (for later reuse) of SpaceX Falcon 9 launches:
  - Machine learning pipeline comparing 4 different classification techniques (K-Nearest-Neighbors, Decision Tree, Support Vector Machine and Logistic Regression)
  - Hyperparameter finetuning using GridSearchCV
  - Feature engineering to analyze influence of rocket reuse, rocket design, launch site position, orbit, etc.
- Prediction of launch outcome (success vs failed landing) with 83% accuracy
  - Emphasis on Launch site, Booster version and Payload Mass effect

INTRODUCTION

- High launch cost halts business profitability of space rocket trips
- Reducing launch cost by reusing rockets can yield significant competitive advantage
- A Data Science approach is used here to predict launch outcome (success vs failed landing) of individual launches of SpaceX Falcon 9 rockets
- Output/Deliverables:
  - An interactive dashboard to visualize influence of Launch Site, Payload Mass and Booster Category
  - A machine learning classification pipeline that yielded an 83% accuracy in the prediction of launch outcome

METHODOLOGY

- Step 1: Data collection via SpaceX API and Wikipedia web scrapping with Python
- Step 2: Data wrangling, including the creation of landing outcome label (‘Class’ column)
- Step 3: Exploratory Data Analysis with SQL and Data Visualization with Python’s Pandas and Matplotlib/Seaborn
- Step 4: Features engineering
  - One Hot Encoding of categorical variables (‘Orbits’, ‘Launch Site’, ‘Landing Pad’, ‘Serial’)
- Step 5: Launch site location analysis with Folium
- Step 6: Interactive Dash app creation
- Step 7: Machine learning prediction of landing outcome using 4 different classification techniques: KNN, SVM, Logistic Regression and Decision Trees
  - Data standardization, train-test data split and hyperparameter finetuning using Scikit-learn

METHODOLOGY: DATA COLLECTION AND DATA WRANGLING

- Python libraries used: requests, pandas, numpy, datetime, BeautifulSoup
- API response content converted to dataframe with pandas json_normalize method
- Webscraping of Wikipedia page done with BeautifulSoup
- Filtered data to keep only Falcon 9 launches
- Missing Payload Mass values replaced by column mean
- Created ‘Class’ column in Dataframe with value = 0 for failed launches and value = 1 for successful launches to label launch outcome of every launch

METHODOLOGY: EXPLORATORY DATA ANALYSIS AND VISUALIZATION

- Python libraries used: sqlite3, matplotlib, seaborn, dash, plotly
- Exploratory Data Analysis with SQL: CSV data loaded in a database table using DB2 service at Watson Studio and later explored with Python
- Exploratory Data Visualization with Python's Seaborn/Matplotlib to analyze correlations between several Feature variables and outcome label 'Class' column
- Interactive dashboard created with Dash to visualize effect and interactions between Launch site, Payload Mass and Booster Version Category on launch success rate

METHODOLOGY: PREDICTIVE ANALYSIS

- Libraries used: Pandas, Scikit-learn, seaborn
- Feature engineering: one hot encoding of selected feature variables done with Pandas ‘get_dummies’ method
- Data normalized with scikit-learn preprocessing ‘StandardScaler’ method
- Data split done with scikit-learn’s ‘train_test_split’ function
- Models fit with train data using 4 different classification prediction techniques (KNN, SVM, Logistic Regression, Decision trees)
- Hyperparameter finetuning using scikit-learn ‘GridSearchCV’ function
- Accuracy of each classification technique obtained with scikit-learn’s ‘score’ method
- Confusion matrix plotted using seaborn

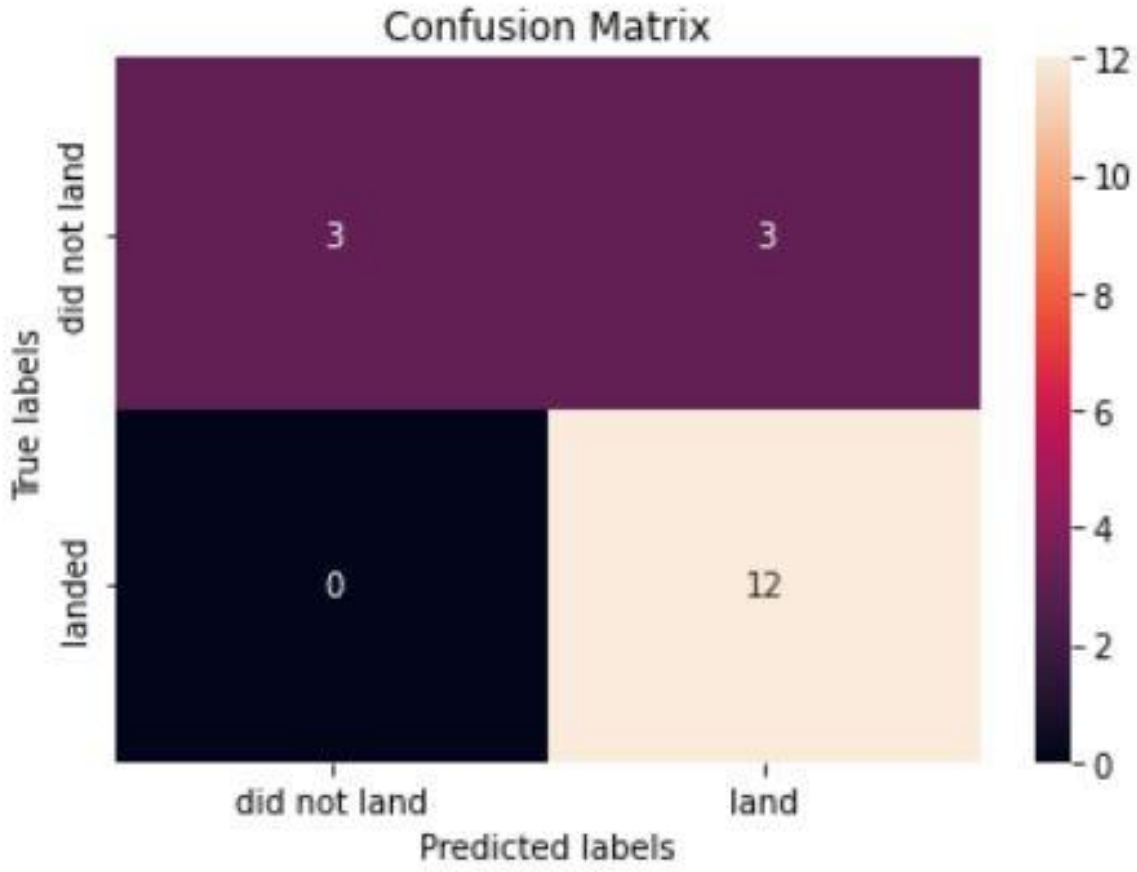
RESULTS: EXPLORATORY DATA ANALYSIS WITH SQL

Total number of successful and failure mission outcomes

Mission_Outcome	count(*)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

For additional results, see 'Exploratory Data Analysis with SQL' Jupyter Notebook in Appendix

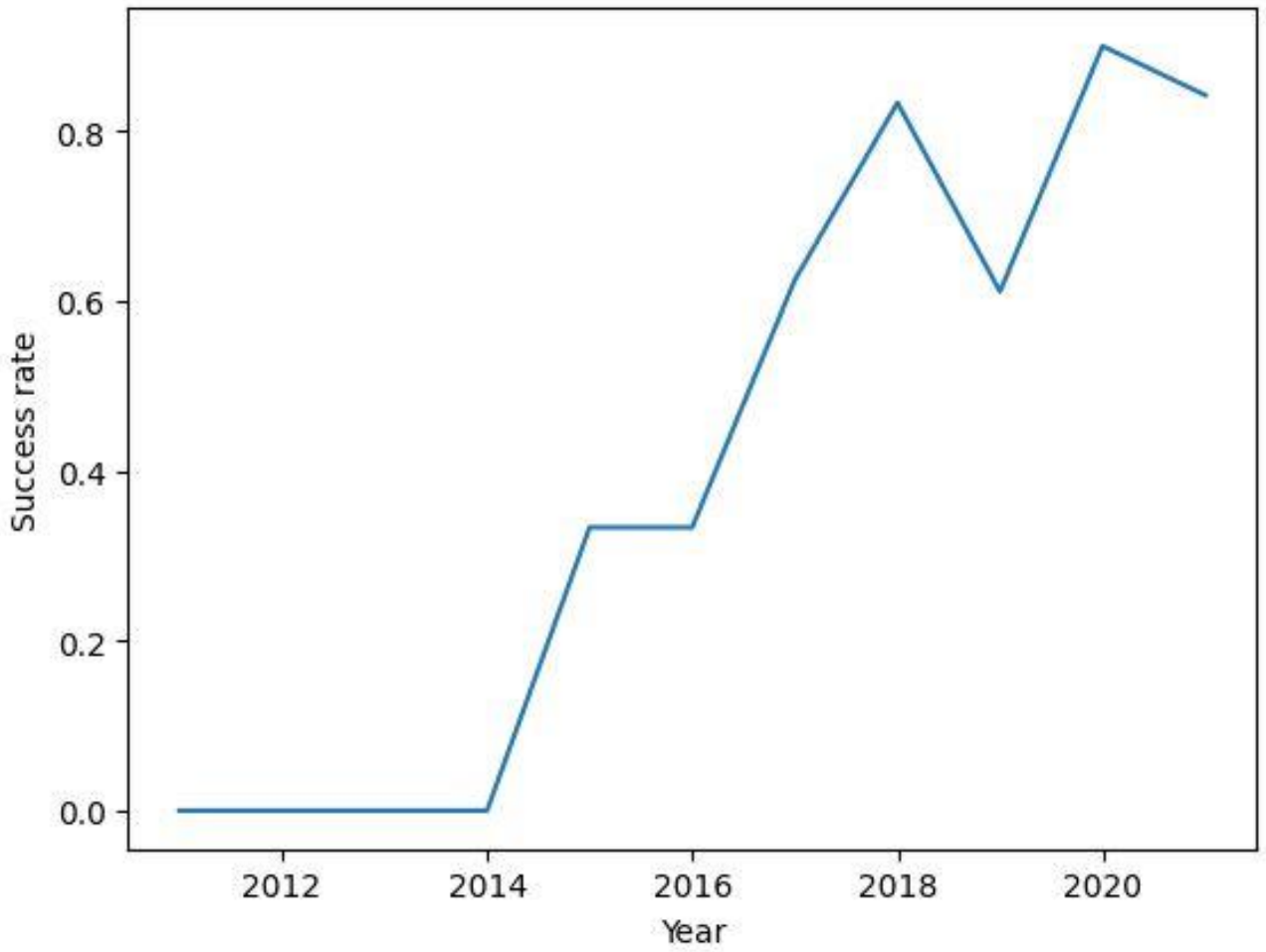
PREDICTIVE ANALYSIS: 83% LANDING OUTCOME ACCURACY



Classification method	Accuracy
K-Nearest-Neighbors	83.33%
Decision Tree	83.33%
Support Vector Machine	83.33%
Logistic Regression	83.33%

For additional predictive (classification) analysis results, see 'Machine learning prediction with Scikit-learn' Jupyter Notebook in Appendix

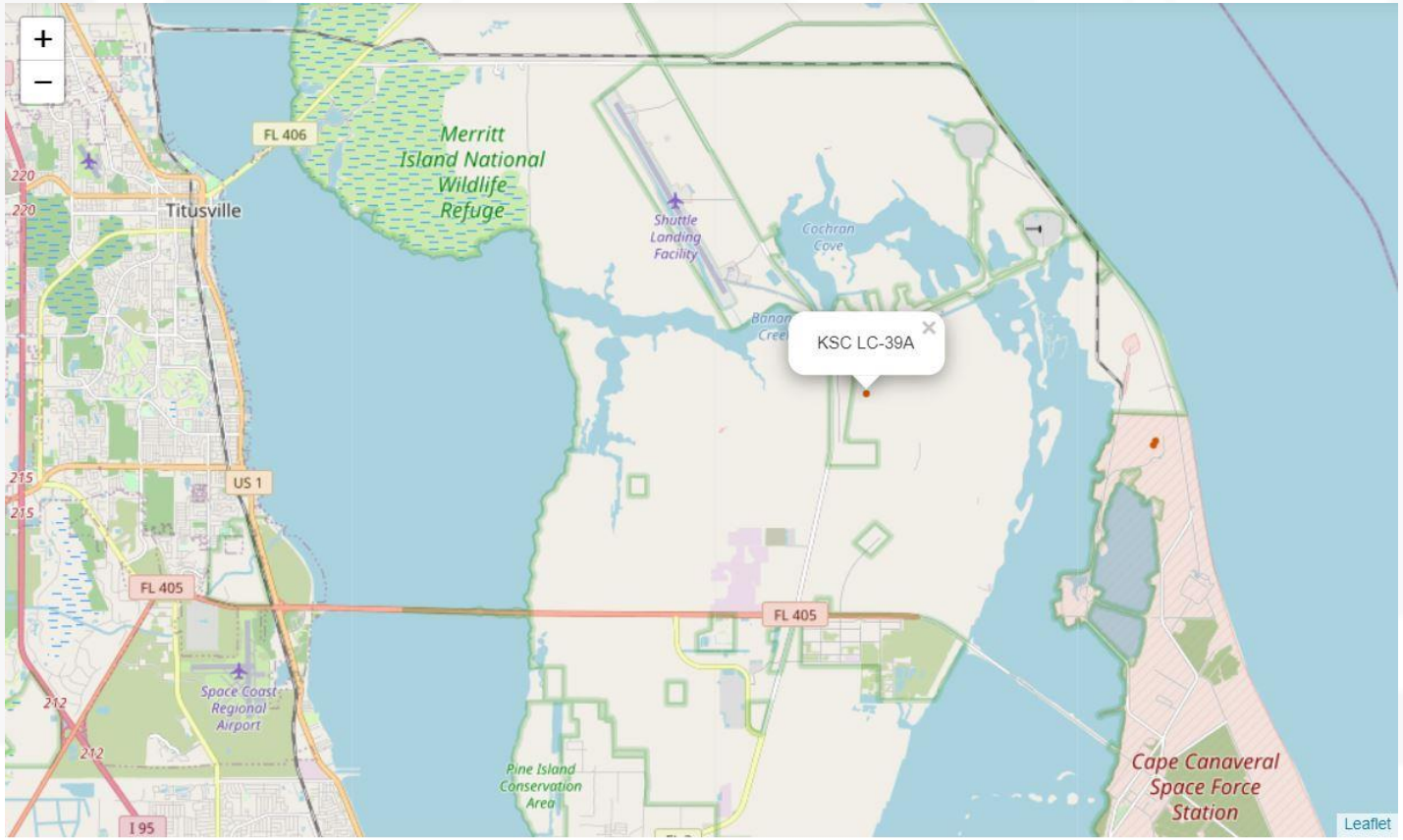
INCREASING LANDING SUCCESS OVER TIME



For additional Data visualization results, see 'Dashboard screenshots 2 and 3' and 'Data Visualization' Jupyter Notebook in Appendix

SITE WITH HIGHEST FALCON 9 LANDING SUCCESS RATE

Site name	Number of Successful landings	Success rate
KSC LC-39A	10	76.9%



For additional data on KSC LC-39 A site, see Dashboard screenshots 2 and 3

For additional spatial analysis results, see 'Site locations analysis with Folium' Jupyter Notebook in Appendix

PAYLOAD MASS RANGE WITH HIGHEST AND LOWEST LANDING SUCCESS RATES (FALCON 9 ROCKETS)

Success classification	Payload Mass range (Kg)	Number of Successful landings / total	Success rate
Lowest success	5.000 – 10.000	3/11	27%
Highest success	0 – 5.000	18/39	46%

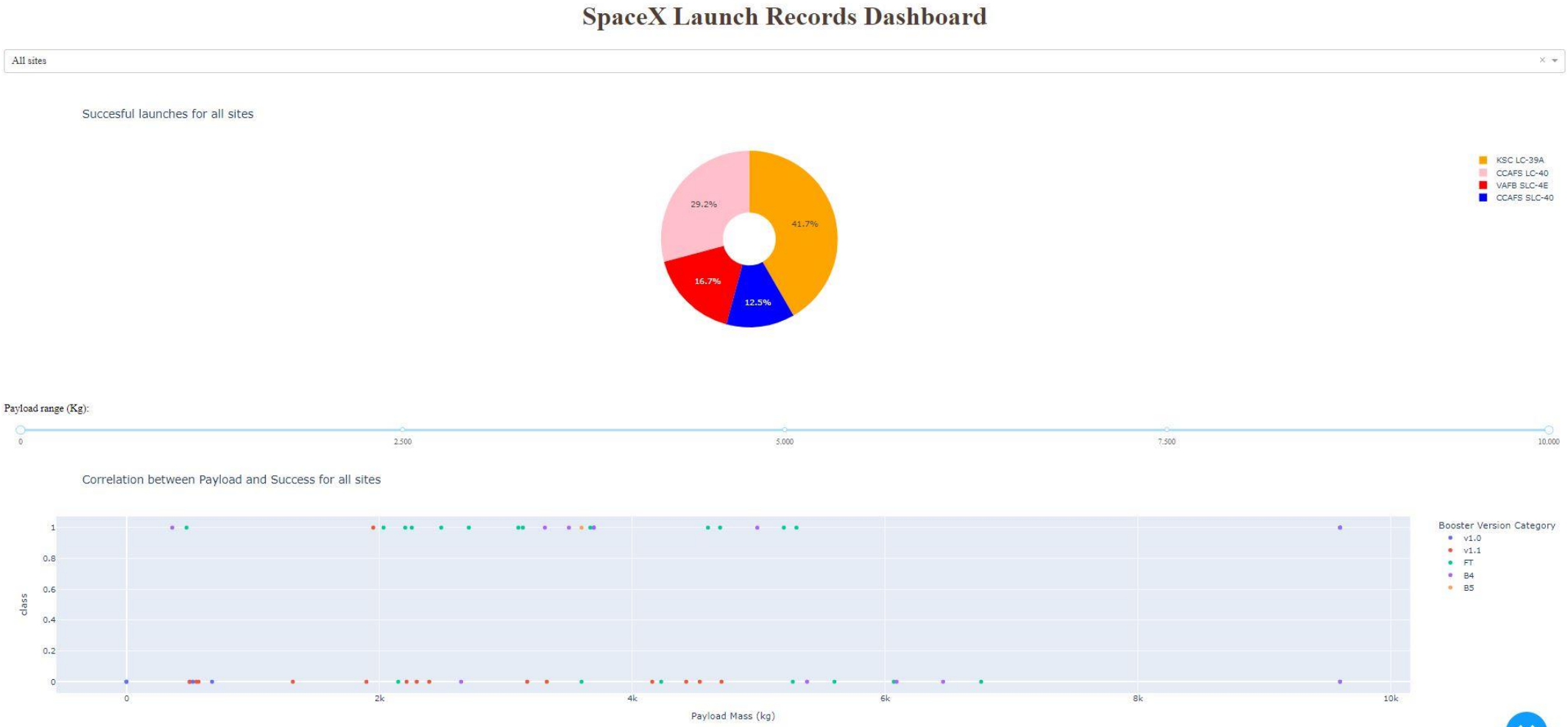
See also Dashboard screenshots 4 and 5

BOOSTER CATEGORY WITH HIGHEST LANDING SUCCESS RATE (FALCON 9 ROCKETS)

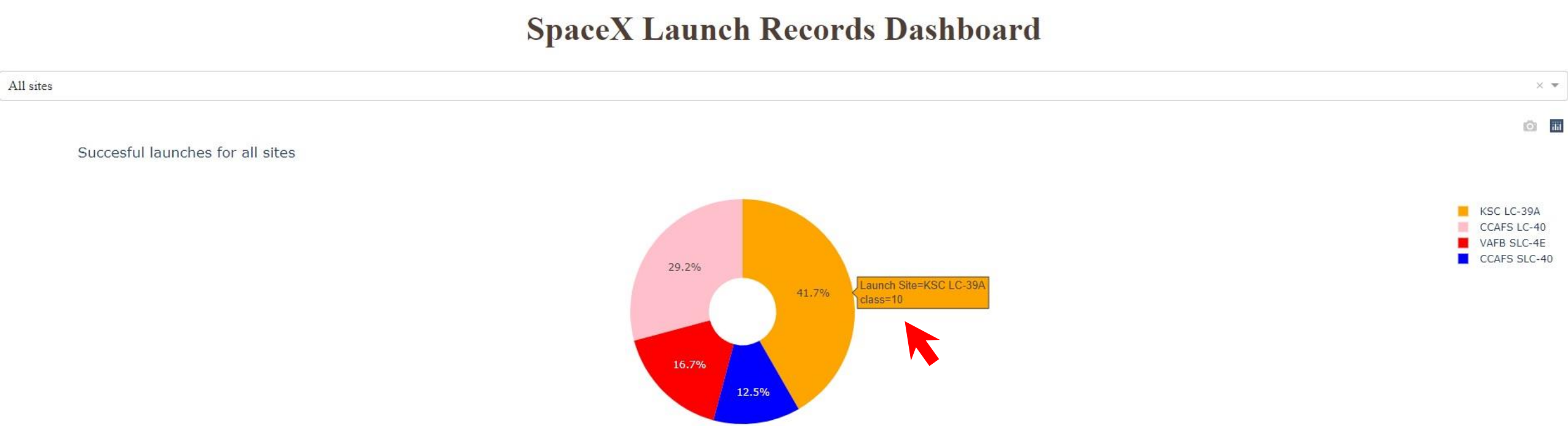
Booster Version Category	Number of Successful landings / total	Success rate
B5	1/1	100%

See also Dashboard screenshot 6

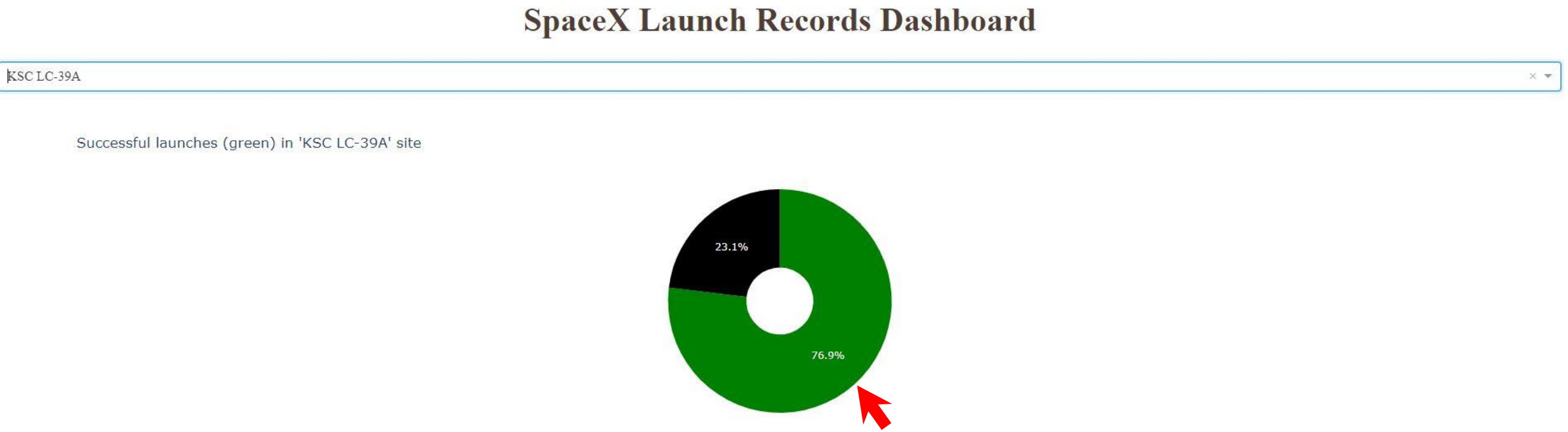
DASHBOARD SCREENSHOT 1



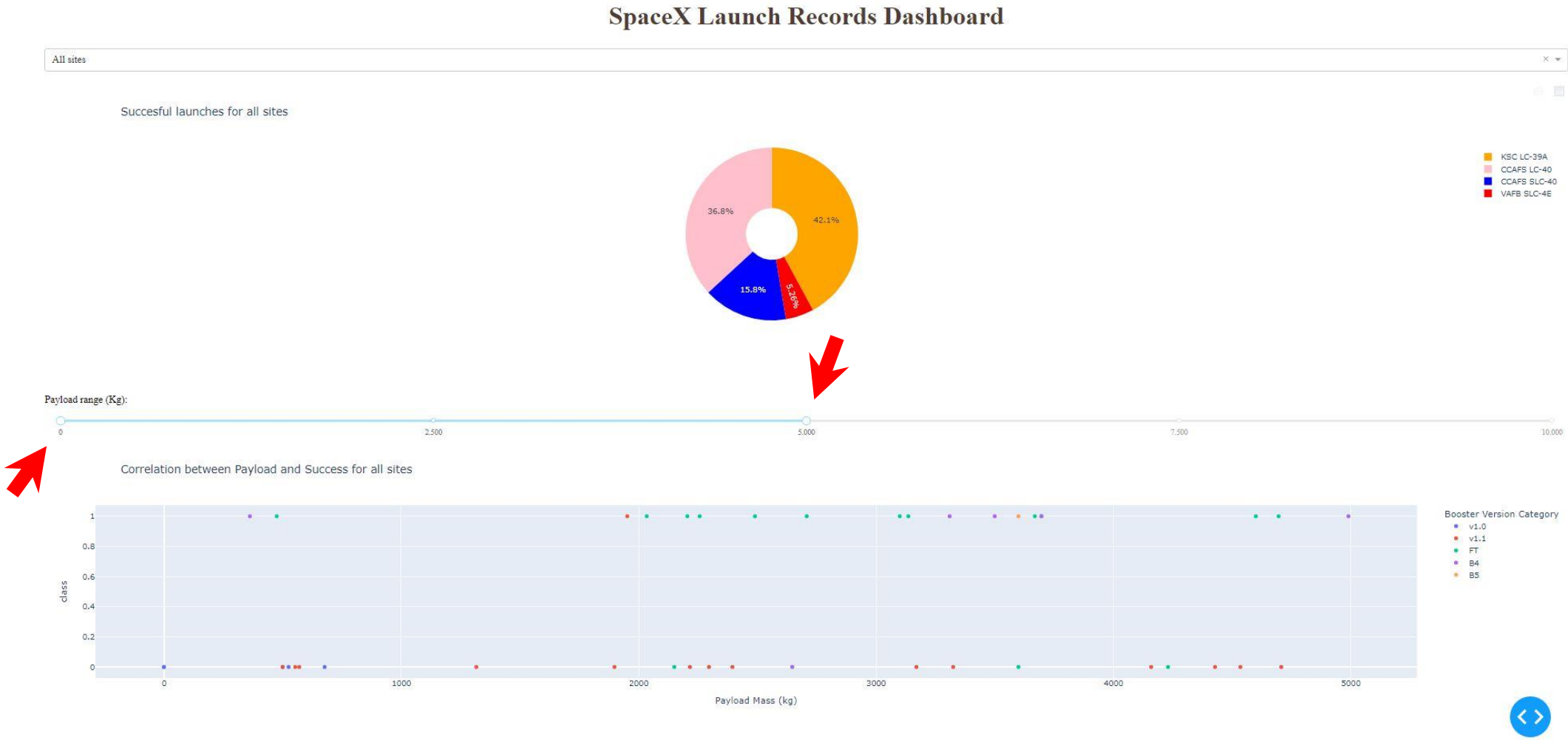
DASHBOARD SCREENSHOT 2



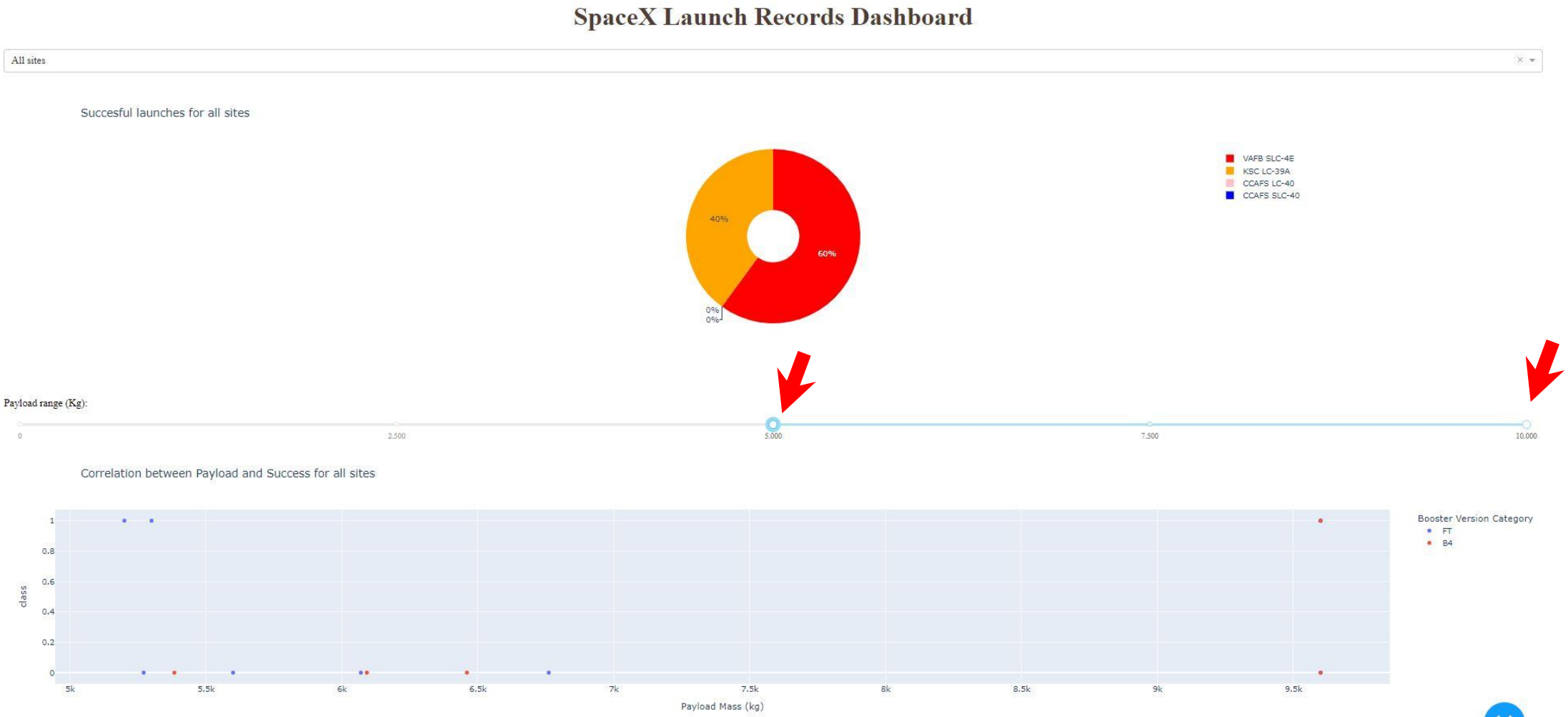
DASHBOARD SCREENSHOT 3



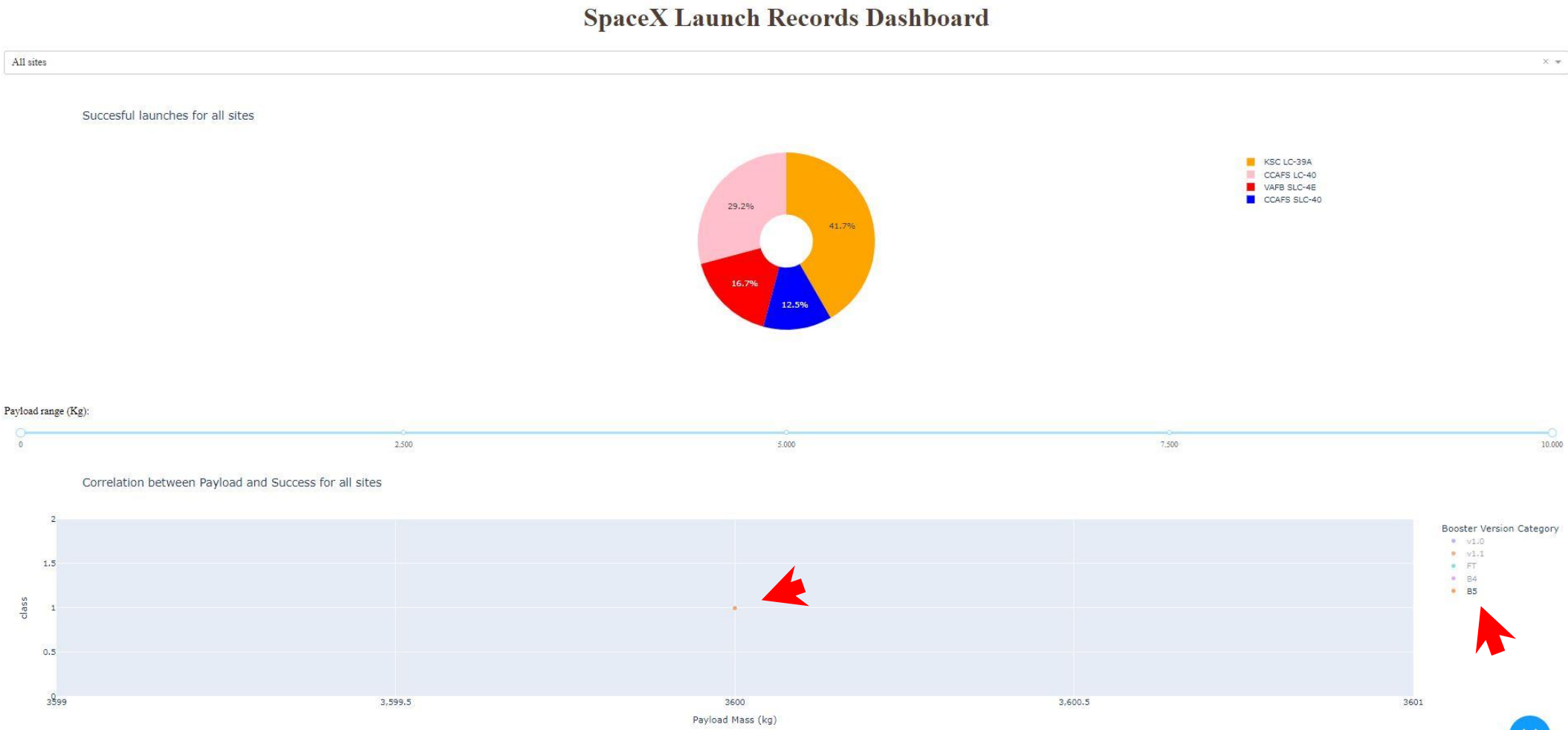
DASHBOARD SCREENSHOT 4



DASHBOARD SCREENSHOT 5



DASHBOARD SCREENSHOT 6



FALCON 9 SUCCESFUL LANDING PREDICTION - FINDINGS & IMPLICATIONS

Findings

- Increasing landing success over time
- Launch outcome prediction accuracy: 83%
- Payload Mass ranges with top (<5.000 Kg) and bottom (>5.000 Kg) success rates identified
- Booster category with top success rate: B5

Implications

- High rate of successful landing prediction, but relatively high false positive rate
- Lower Payload Mass has higher landing success rate

DISCUSSION

- The prediction system can spot all successful landings, but a few unsuccessful landings will be misclassified as successful (false positives)
- Remains unknown if landing success can further increase in the future or has reached its maximum level
- Lower payload mass and characteristics of B5 boosters and KSC LC-39A launch site could further increase success rate

CONCLUSION

- Falcon 9 successful landings can be predicted with over 80% accuracy
- Success landing rate has increased over the years
- If the same trend is kept, success rate may continue to increase, perhaps reaching near 100% success in the next few years.
- Beyond the effect of lower Payload Mass, a better understanding of B5 booster design and KSC LC-39A launch site could further contribute to increasing successful landing rate

APPENDIX

- Data collection via SpaceX API (*Jupyter notebook*)
- Wikipedia Webscrapping (*Jupyter notebook*)
- Data wrangling (*Jupyter notebook*)
- Exploratory Data Analysis with SQL (*Jupyter notebook*)
- Data Visualization and Feature Engineering (*Jupyter notebook*)
- Launch sites locations analysis with Folium (*Jupyter notebook*)
- Interactive dashboard (*Dash app script*)
- Machine learning prediction with Scikit-learn (*Jupyter notebook*)
- Github repo (*link*)

ACKNOWLEDGEMENTS

- I am grateful to the guidance of mentors at IBM and Coursera who shared their knowledge and guided students throughout this Data Science Professional Certificate with 10 extraordinary courses that helped me learn fascinating tools and techniques for Data Science.