CAPSTONE PROJECT

Analyzing SpaceX Falcon 9 Launch Data: Predicting First Stage Reusability and Launch Costs.

Decoding SpaceX Falcon 9: Analyzing Mission Success and Cost Optimization Strategies

Project Overview:

The commercial space age has ushered in a new era of accessibility to space travel.

Overview of key players in the space industry, including Virgin Galactic, Rocket Lab, Blue Origin, and SpaceX.

Focus on SpaceX's achievements, including ISS missions, Starlink satellite constellation, and manned space missions.

Project Objective:

Aim to determine the cost-effectiveness of Falcon 9 rocket launches by predicting first stage reusability.

Role of data science in providing insights for a new rocket company, Space Y, aiming to compete with SpaceX.

Approach:

Utilizing machine learning techniques to predict the successful landing and reusability of Falcon 9 first stages.

Analysis of public data to gather insights into SpaceX launch operations and cost optimization strategies.

Key Findings:

Identification of factors influencing the success of Falcon 9 first stage landings.

Insights into the cost implications of successful first stage reusability for launch operations.

Potential opportunities for cost optimization and competitive positioning in the commercial space market.

Recommendations:

Leveraging predictive analytics to inform decision-making and pricing strategies for Space Y.

Continuous monitoring and analysis of SpaceX launch data to refine predictive models and improve accuracy over time.

Exploring the Future of Space Travel with Data Science

Welcome to our presentation on the future of space travel and the role of data science in shaping it. In today's commercial space age, companies like SpaceX are leading the charge, making space exploration more accessible and affordable than ever before. SpaceX's groundbreaking achievements, including reusable rocket technology, have transformed the industry landscape.

In this presentation, we'll delve into SpaceX's Falcon 9 launches, analyze the factors driving their success, and explore the implications for the future of space travel. Our objective is to leverage data science and machine learning to predict first stage reusability and determine launch costs.

Data Collection

Importance of data collection:

Data collection forms the foundation of any data-driven analysis or prediction task.It provides the necessary raw material for insights and decision-making.

Data Sources:

SpaceX REST API:Utilized to gather comprehensive launch data, including rocket details, payload information, and landing outcomes.

API endpoints used: **/launches/past.**

Web Scraping:

BeautifulSoup package employed to extract Falcon 9 launch records from HTML tables on Wiki pages.

Data parsed and converted into Pandas DataFrame for further analysis.

Collection Process:

SpaceX REST API:GET request made to the specified endpoint using the requests library.

Resulting JSON data normalized using json_normalize function to create a flat table.

Web Scraping:

HTML tables scraped using BeautifulSoup.Data extracted, parsed, and converted into Pandas DataFrame.

Data Wrangling

Introduction:

Data wrangling is a crucial step in preparing raw data for analysis.

It involves cleaning, transforming, and integrating data to create a structured dataset suitable for analysis.

Exploration of Attributes:

- Launch Site:
  - Contains various launch sites like Vandenberg AFB, Kennedy Space Center, and CCAFS SLC 40.
- Orbits:
  - Represents different orbits of the payload, such as LEO (Low Earth Orbit) and GTO (Geosynchronous Transfer Orbit).
- Outcome:
  - Indicates the success of the first stage landing, classified as True (successful) or False (unsuccessful).

  Conversion to Classes:
- Landing outcomes are transformed into binary classes for classification:
  - Class 0 signifies a bad outcome (booster did not land).
  - Class 1 signifies a good outcome (booster successfully landed).

Exploratory Data Analysis

Exploratory Data Analysis (EDA) is the initial step in any data science project. It involves exploring and understanding the structure, patterns, and relationships within the dataset.

Purpose of EDA:

- To understand the characteristics and properties of the dataset.
- To identify potential patterns, trends, and anomalies in the data.

Key Insights:

- Success Rate Improvement:
  - The success rate of SpaceX launches has shown improvement since 2013.
  - Incorporating launch number as a feature may provide insights into success trends over time.
- Launch Site Analysis:
  - Different launch sites exhibit varying success rates.
  - For example, CCAFS LC-40 has a success rate of 60%, while KSC LC-39A and VAFB SLC 4E have success rates around 77%.
- Feature Combination:
  - Combining attributes can provide deeper insights.
  - For instance, when overlaying landing outcomes with payload mass, CCAFS LC-40 shows a 100% success rate for payloads above 10,000 kg

  Correlation Analysis:

- Determine attributes correlated with successful landings.
- Explore relationships between variables to identify predictive features for successful first-stage landings.

Data Preparation:

- Convert categorical variables using one-hot encoding.
- Prepare the dataset for machine learning model training to predict first-stage landing success.

Conclusion:

- EDA is a critical step in understanding the dataset and informing subsequent analysis.
- Insights gained from EDA guide feature selection, model development, and decision-making processes.

Predictive Analysis Methodology

Predictive Analysis involves building machine learning models to forecast future outcomes based on historical data.

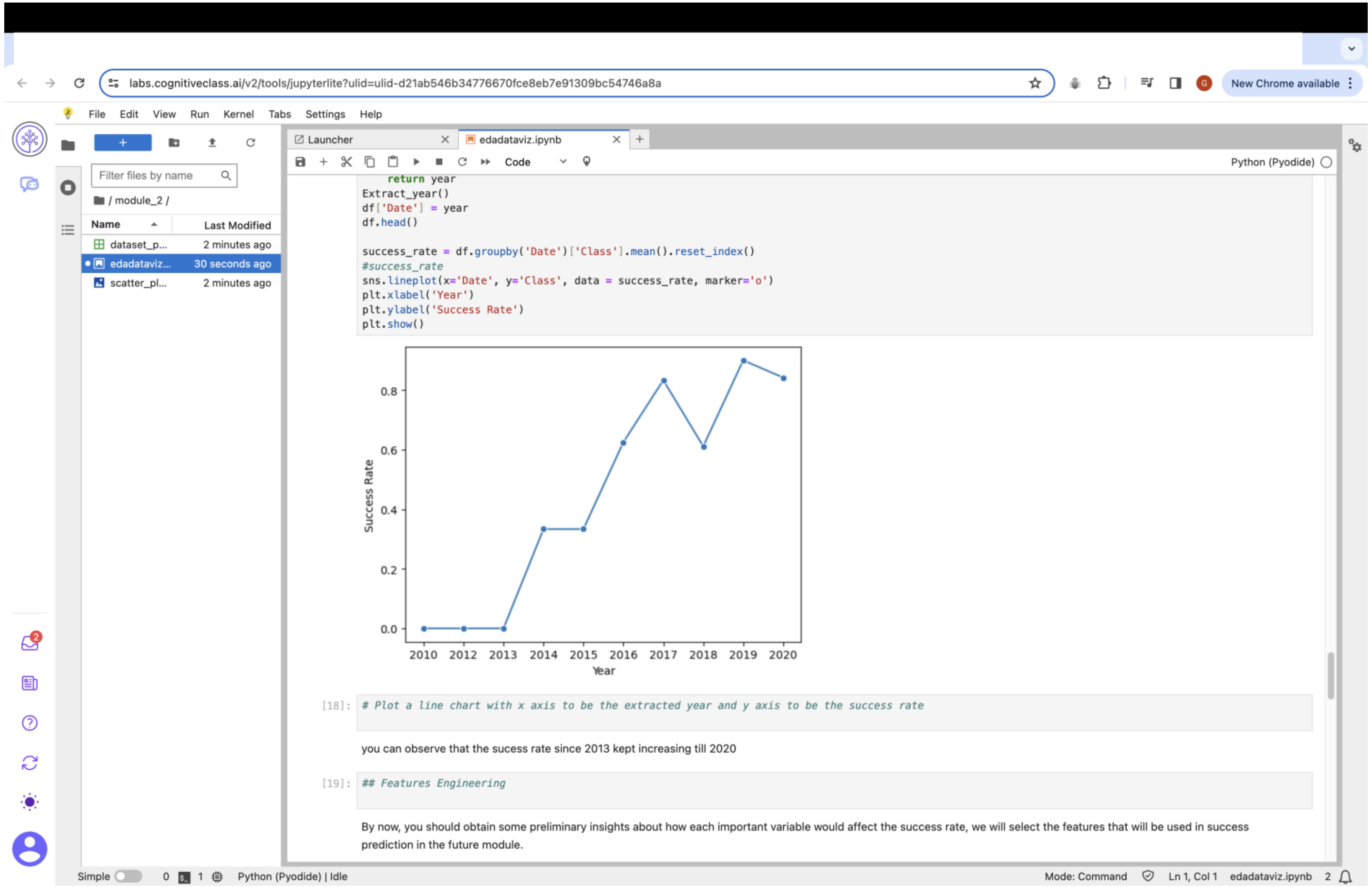
Machine Learning Pipeline:

- **Preprocessing:**
  - Standardize the data to ensure all features have the same scale.
  - Handle missing values and categorical variables through encoding or imputation.
- **Train-Test Split:**
  - Divide the dataset into training and testing subsets.
  - Training data is used to train the model, while testing data evaluates its performance.
- **Model Training:**
  - Train machine learning models using the training data.
  - Explore various algorithms to find the best-performing model for the task.
- **Grid Search:**
  - Perform hyperparameter tuning using Grid Search.
  - Find the optimal combination of hyperparameters that maximize model performance.
- **Model Evaluation:**
  - Assess model accuracy using the training data.
  - Compare the performance of different algorithms, such as Logistic Regression, Support Vector Machines, Decision Tree Classifier, and K-nearest Neighbors.
- **Confusion Matrix:**
  - Output the confusion matrix to evaluate the model's performance.
  - Provides insights into the model's predictive ability, including true positives, false positives, true negatives, and false negatives.

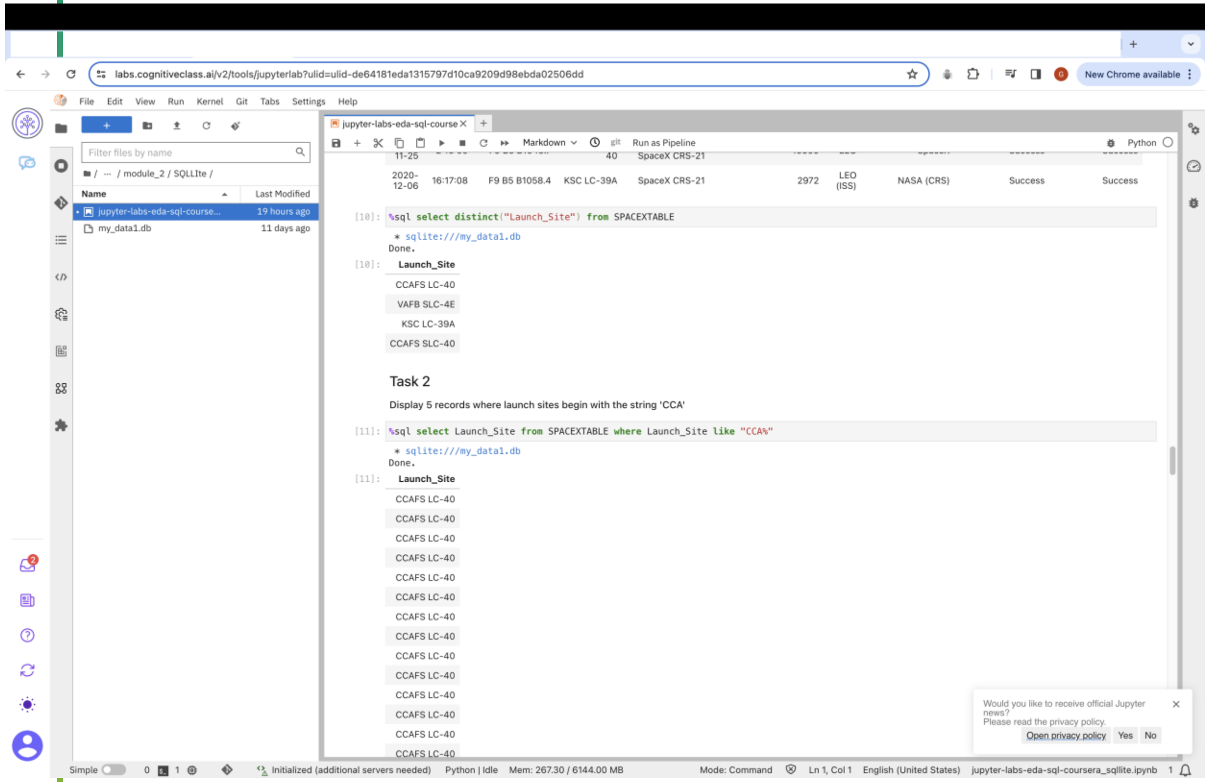
Conclusion:

- Predictive Analysis methodology enables the development of robust machine learning models for forecasting outcomes.
- Through preprocessing, model training, hyperparameter tuning, and evaluation, predictive models can provide valuable insights and inform decision-making processes.

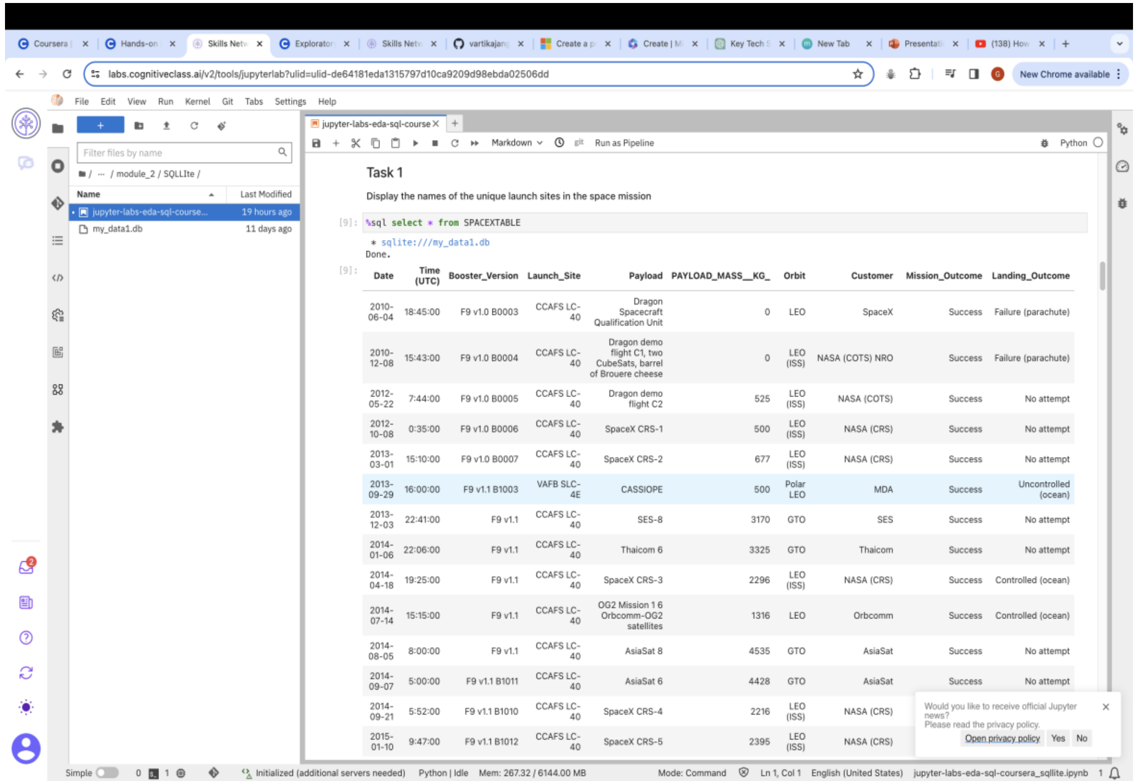
EDA with visualization results



EDA with SQL

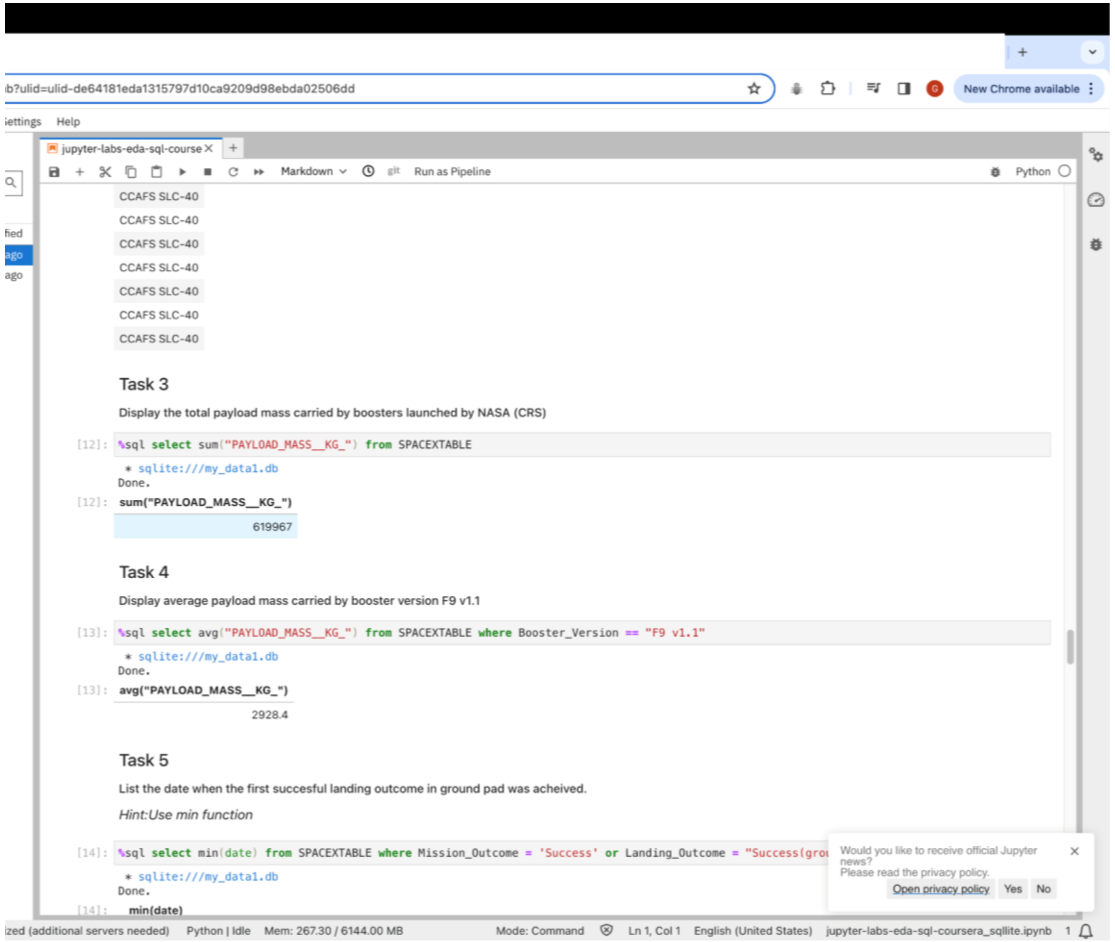


labs.cognitiveclass.ai/v2/tools/jupyterlab?ulid=ulid-de64181eda1315797d10ca9209d98ebda02506dd
File Edit View Run Kernel Git Tabs Settings Help
jupyter-labs-eda-sql-course
/ module_2 / SQLlte /
Name Last Modified
jupyter-labs-eda-sql-course... 19 hours ago
my_data1.db 11 days ago
Run as Pipeline
2020-12-06 16:17:08 F9 B5 B1058.4 KSC LC-39A SpaceX CRS-21 2972 LEO (ISS) NASA (CRS) Success Success
[10]: %sql select distinct("Launch_Site") from SPACEXTABLE
* sqlite:///my_data1.db
Done.
Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
Task 2
Display 5 records where launch sites begin with the string 'CCA'
[11]: %sql select Launch_Site from SPACEXTABLE where Launch_Site like "CCA%"
* sqlite:///my_data1.db
Done.
Launch_Site
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
CCAFS LC-40
Would you like to receive official Jupyter news?
Please read the privacy policy.
Open privacy policy Yes No
Initialized (additional servers needed) Python | Idle Mem: 267.30 / 6144.00 MB Mode: Command Ln 1, Col 1 English (United States) jupyter-labs-eda-sql-coursera_sqllite.ipynb

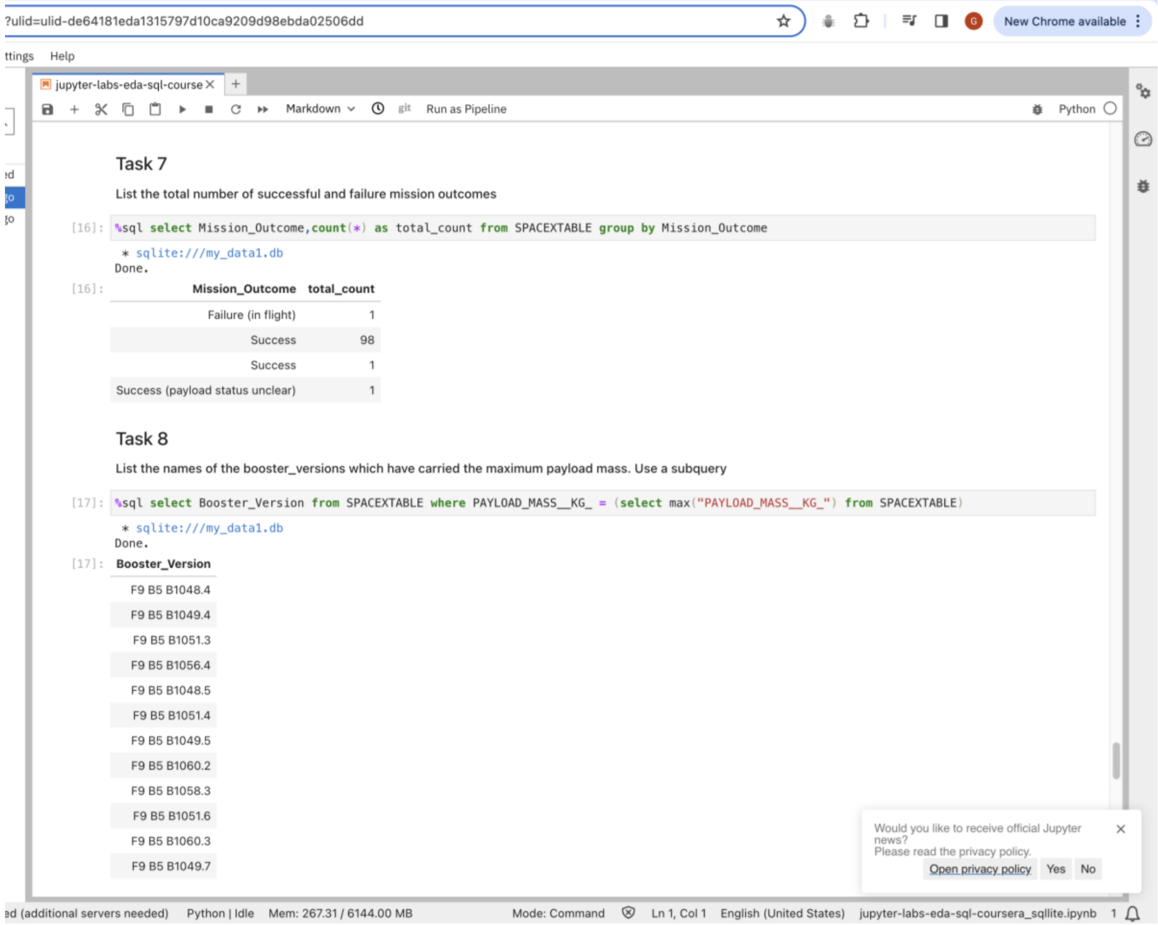


labs.cognitiveclass.ai/v2/tools/jupyterlab?ulid=ulid-de64181eda1315797d10ca9209d98ebda02506dd
File Edit View Run Kernel Git Tabs Settings Help
jupyter-labs-eda-sql-course
/ module_2 / SQLlte /
Name Last Modified
jupyter-labs-eda-sql-course... 19 hours ago
my_data1.db 11 days ago
Run as Pipeline
Task 1
Display the names of the unique launch sites in the space mission
[9]: %sql select * from SPACEXTABLE
* sqlite:///my_data1.db
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 7:44:00 F9 v1.0 B0005 CCAFS LC-40 Dragon demo flight C2 525 LEO (ISS) NASA (COTS) Success No attempt
2012-10-08 0:35:00 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
2013-09-29 16:00:00 F9 v1.1 B1003 VAFB SLC-4E CASSIOPE 500 Polar LEO MDA Success Uncontrolled (ocean)
2013-12-03 22:41:00 F9 v1.1 CCAFS LC-40 SES-8 3170 GTO SES Success No attempt
2014-01-06 22:06:00 F9 v1.1 CCAFS LC-40 Thaicom 6 3325 GTO Thaicom Success No attempt
2014-04-18 19:25:00 F9 v1.1 CCAFS LC-40 SpaceX CRS-3 2296 LEO (ISS) NASA (CRS) Success Controlled (ocean)
2014-07-14 15:15:00 F9 v1.1 CCAFS LC-40 OG2 Mission 1 6 Orbcomm-OG2 satellites 1316 LEO Orbcomm Success Controlled (ocean)
2014-08-05 8:00:00 F9 v1.1 CCAFS LC-40 AsiaSat 8 4535 GTO AsiaSat Success No attempt
2014-09-07 5:00:00 F9 v1.1 B1011 CCAFS LC-40 AsiaSat 6 4428 GTO AsiaSat Success No attempt
2014-09-21 5:52:00 F9 v1.1 B1010 CCAFS LC-40 SpaceX CRS-4 2216 LEO (ISS) NASA (CRS)
2015-01-10 9:47:00 F9 v1.1 B1012 CCAFS LC-40 SpaceX CRS-5 2395 LEO (ISS) NASA (CRS)
Would you like to receive official Jupyter news?
Please read the privacy policy.
Open privacy policy Yes No
Initialized (additional servers needed) Python | Idle Mem: 267.32 / 6144.00 MB Mode: Command Ln 1, Col 1 English (United States) jupyter-labs-eda-sql-coursera_sqllite.ipynb

EDA with SQL

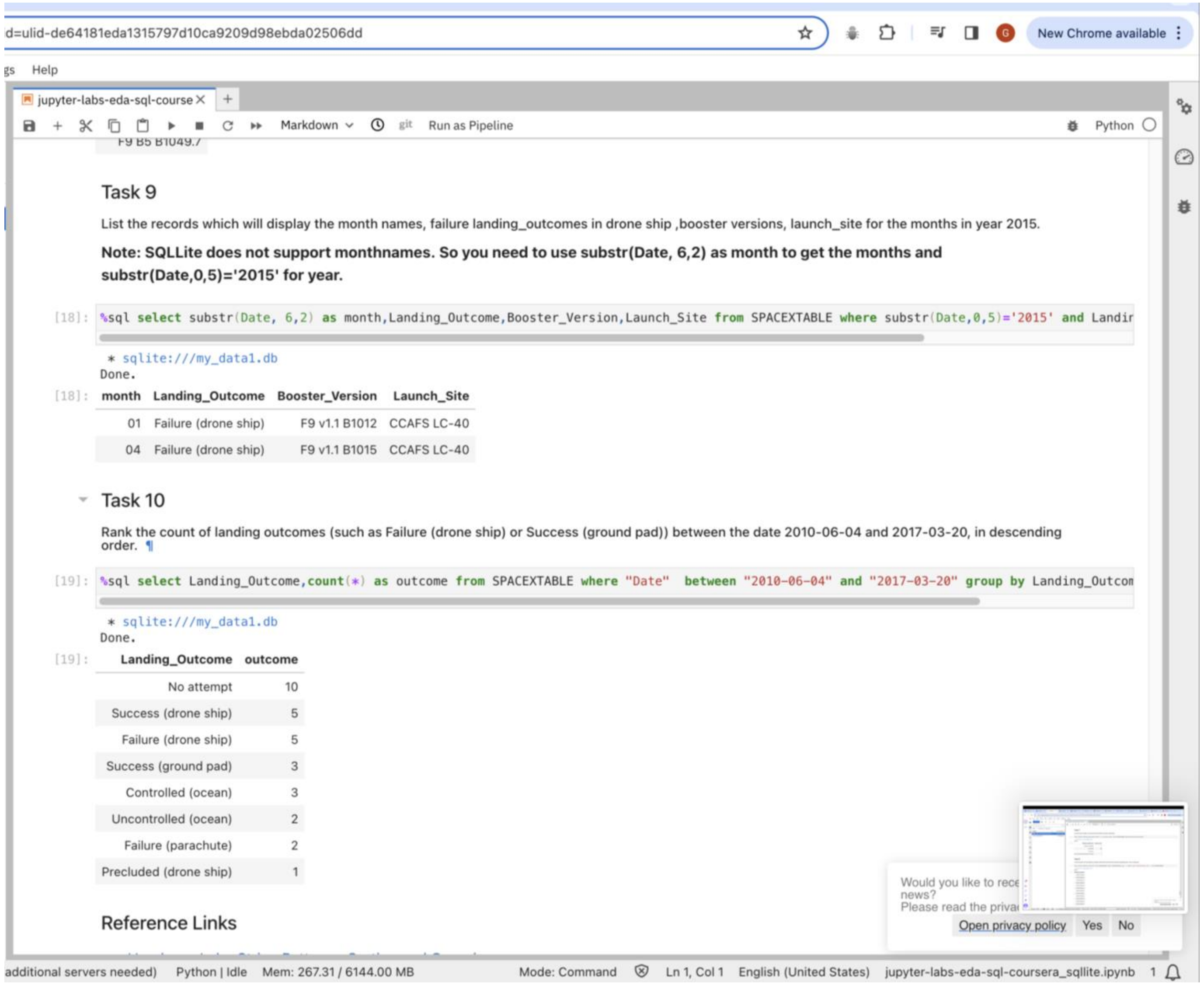


jupyter-labs-eda-sql-course
CCAFS SLC-40
CCAFS SLC-40
CCAFS SLC-40
CCAFS SLC-40
CCAFS SLC-40
CCAFS SLC-40
CCAFS SLC-40
Task 3
Display the total payload mass carried by boosters launched by NASA (CRS)
[12]: %sql select sum("PAYLOAD_MASS__KG_") from SPACEXTABLE
* sqlite:///my_data1.db
Done.
[12]: sum("PAYLOAD_MASS__KG_")
619967
Task 4
Display average payload mass carried by booster version F9 v1.1
[13]: %sql select avg("PAYLOAD_MASS__KG_") from SPACEXTABLE where Booster_Version == "F9 v1.1"
* sqlite:///my_data1.db
Done.
[13]: avg("PAYLOAD_MASS__KG_")
2928.4
Task 5
List the date when the first succesful landing outcome in ground pad was acheived.
Hint:Use min function
[14]: %sql select min(date) from SPACEXTABLE where Mission_Outcome = 'Success' or Landing_Outcome = "Success(grou
* sqlite:///my_data1.db
Done.
min(date)
Would you like to receive official Jupyter news?
Please read the privacy policy.
Open privacy policy Yes No



jupyter-labs-eda-sql-course
Task 7
List the total number of successful and failure mission outcomes
[16]: %sql select Mission_Outcome,count(*) as total_count from SPACEXTABLE group by Mission_Outcome
* sqlite:///my_data1.db
Done.
[16]: Mission_Outcome total_count
Failure (in flight) 1
Success 98
Success 1
Success (payload status unclear) 1
Task 8
List the names of the booster_versions which have carried the maximum payload mass. Use a subquery
[17]: %sql select Booster_Version from SPACEXTABLE where PAYLOAD_MASS__KG_ = (select max("PAYLOAD_MASS__KG_") from SPACEXTABLE)
* sqlite:///my_data1.db
Done.
[17]: 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
Would you like to receive official Jupyter news?
Please read the privacy policy.
Open privacy policy Yes No

EDA with SQL



Task 9

List the records which will display the month names, failure landing_outcomes in drone ship ,booster versions, launch_site for the months in year 2015.

Note: SQLLite does not support monthnames. So you need to use substr(Date, 6,2) as month to get the months and substr(Date,0,5)='2015' for year.

[18]:
```
%sql select substr(Date, 6,2) as month,Landing_Outcome,Booster_Version,Launch_Site from SPACEXTABLE where substr(Date,0,5)='2015' and Landir
```

```
 * sqlite:///my_data1.db
Done.
```

[18]:

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

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.

[19]:
```
%sql select Landing_Outcome,count(*) as outcome from SPACEXTABLE where "Date"  between "2010-06-04" and "2017-03-20" group by Landing_Outcom
```

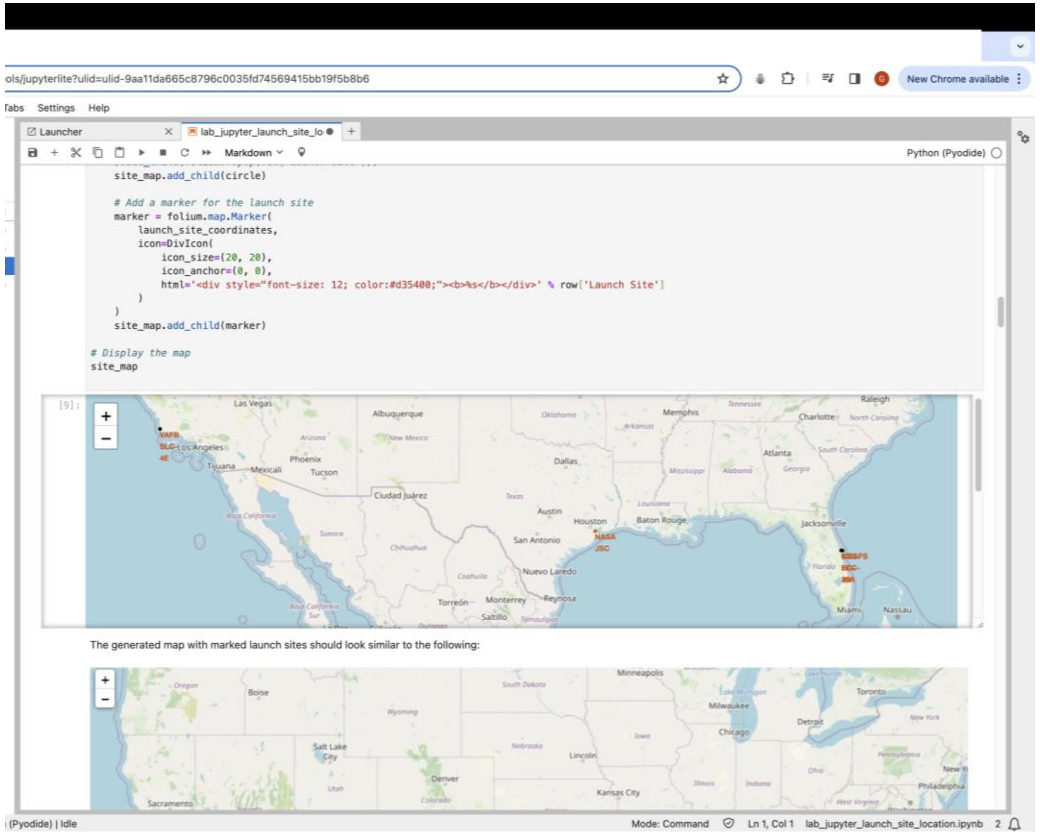
```
 * sqlite:///my_data1.db
Done.
```

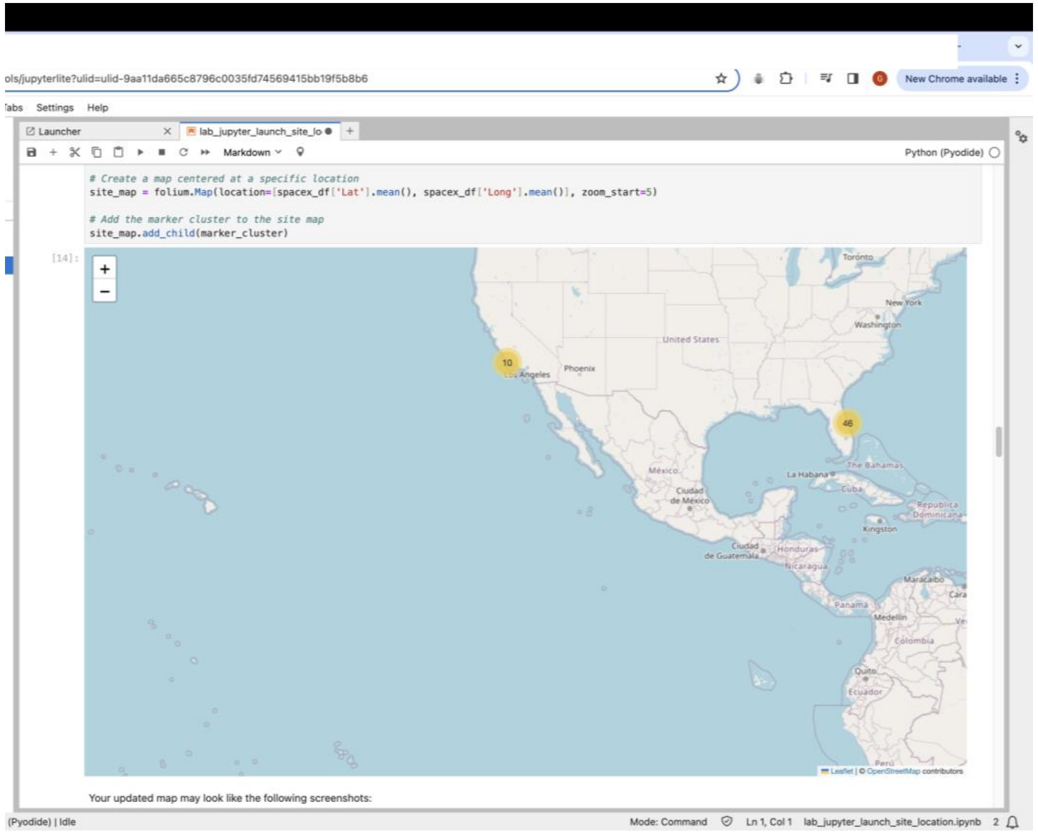
[19]:

Landing_Outcome	outcome
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

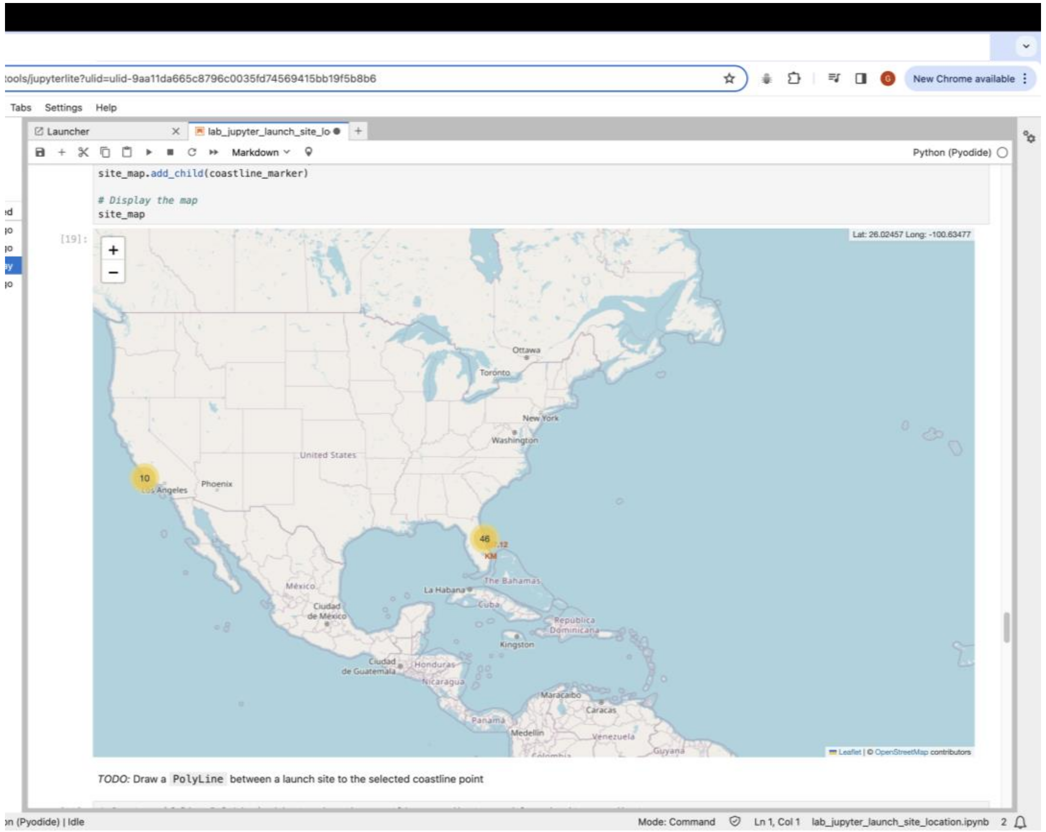
Reference Links

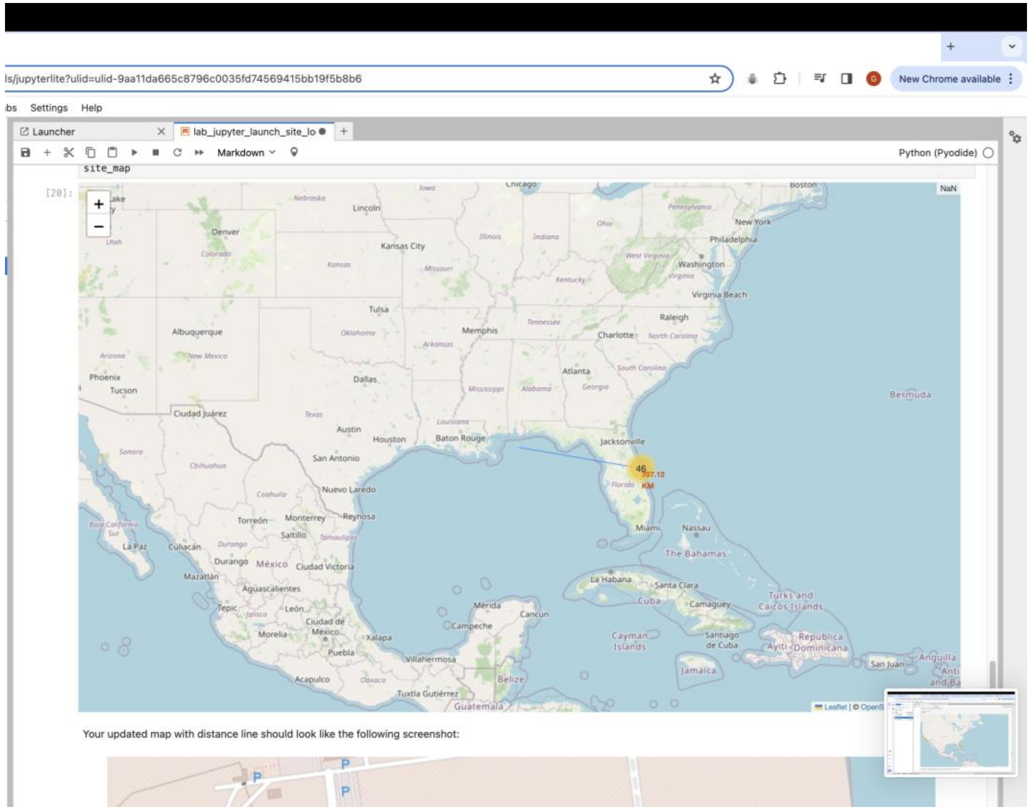
Interactive map with Folium



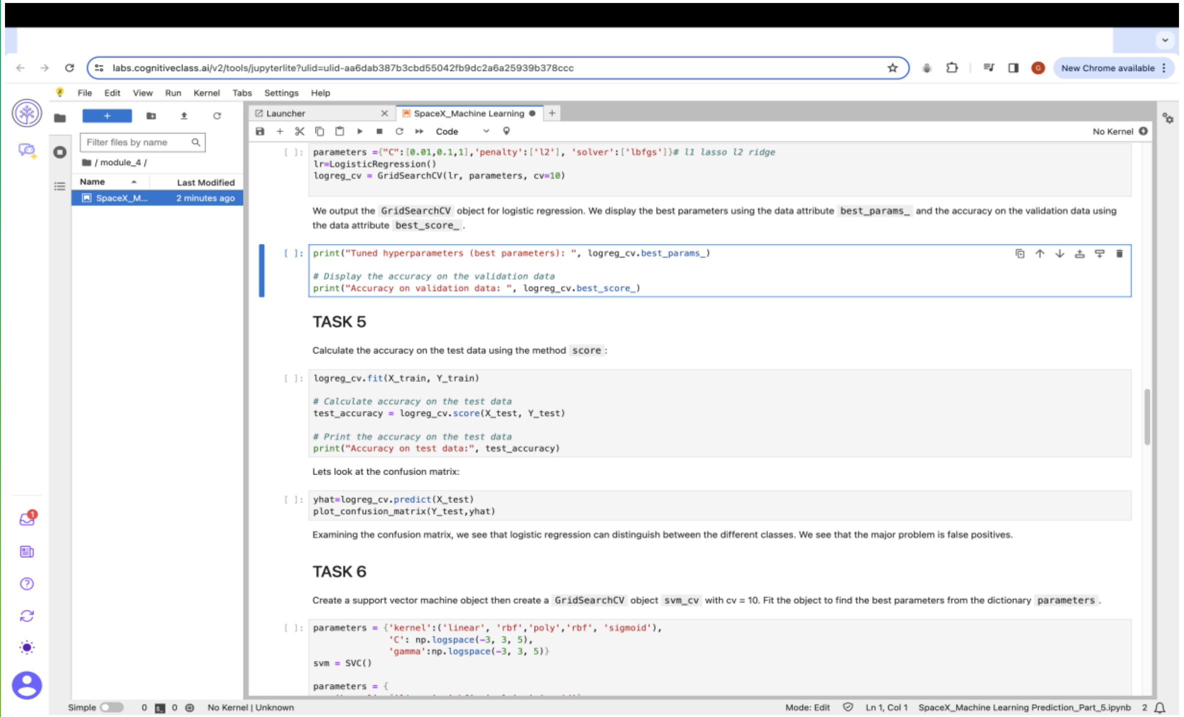


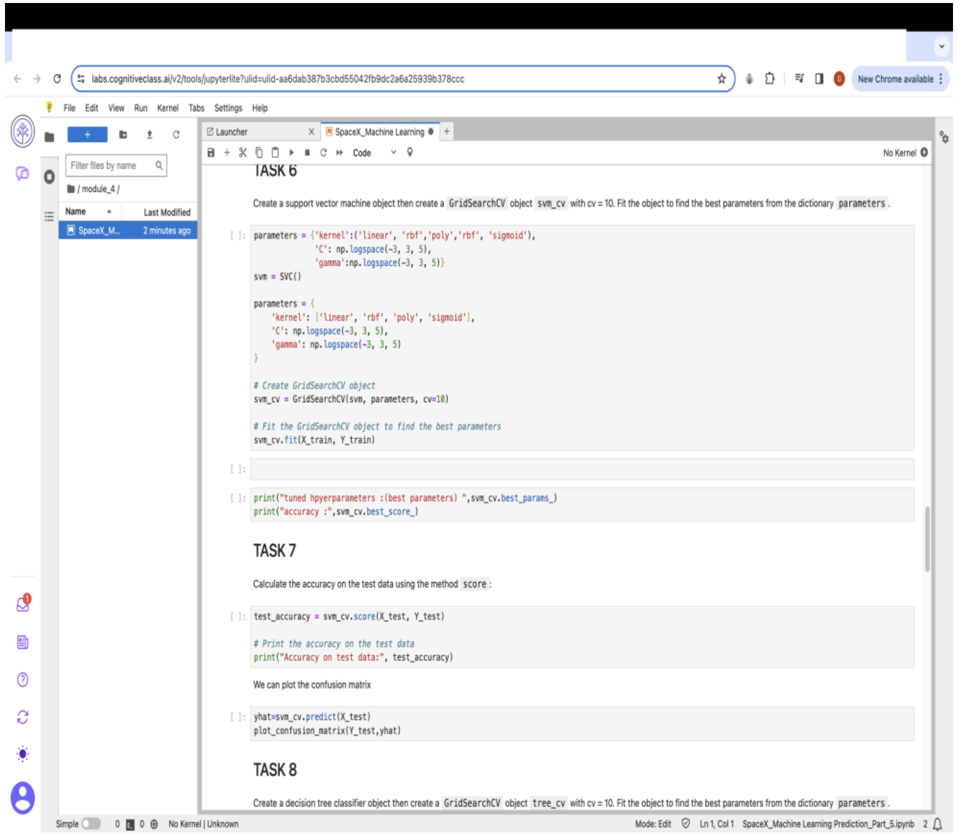
Interactive map with Folium





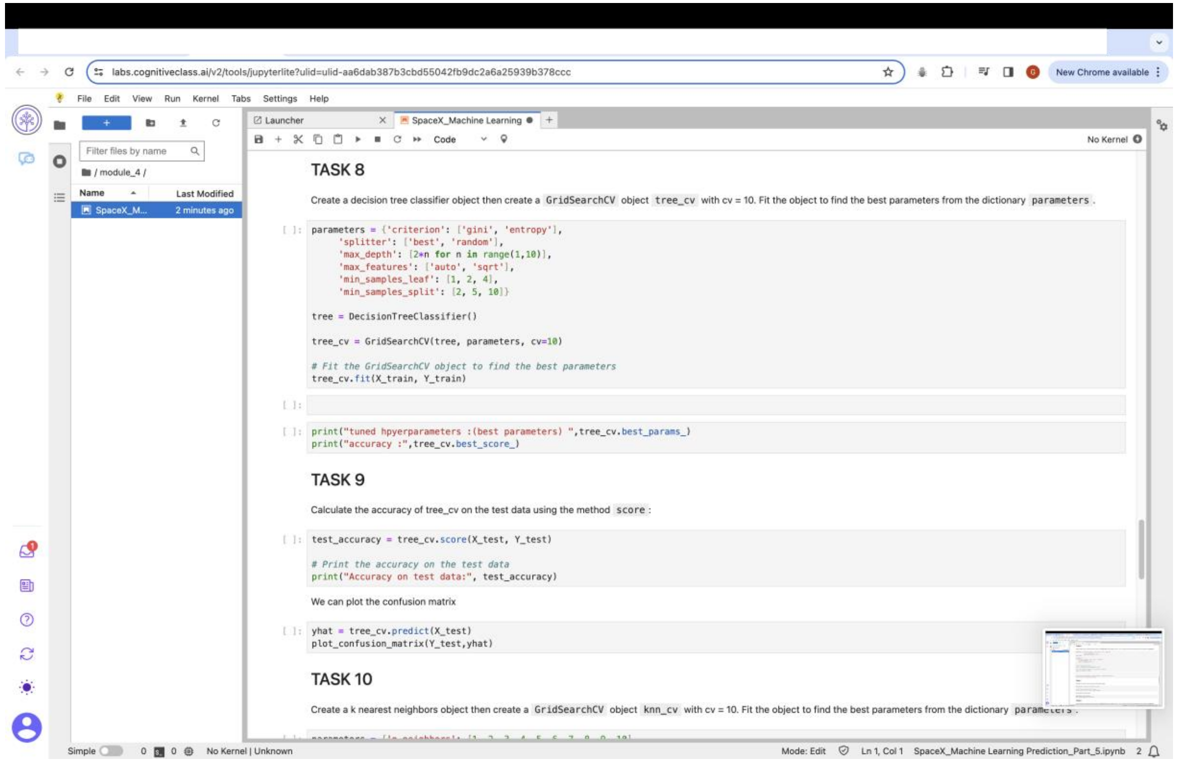
Predictive Analysis Results

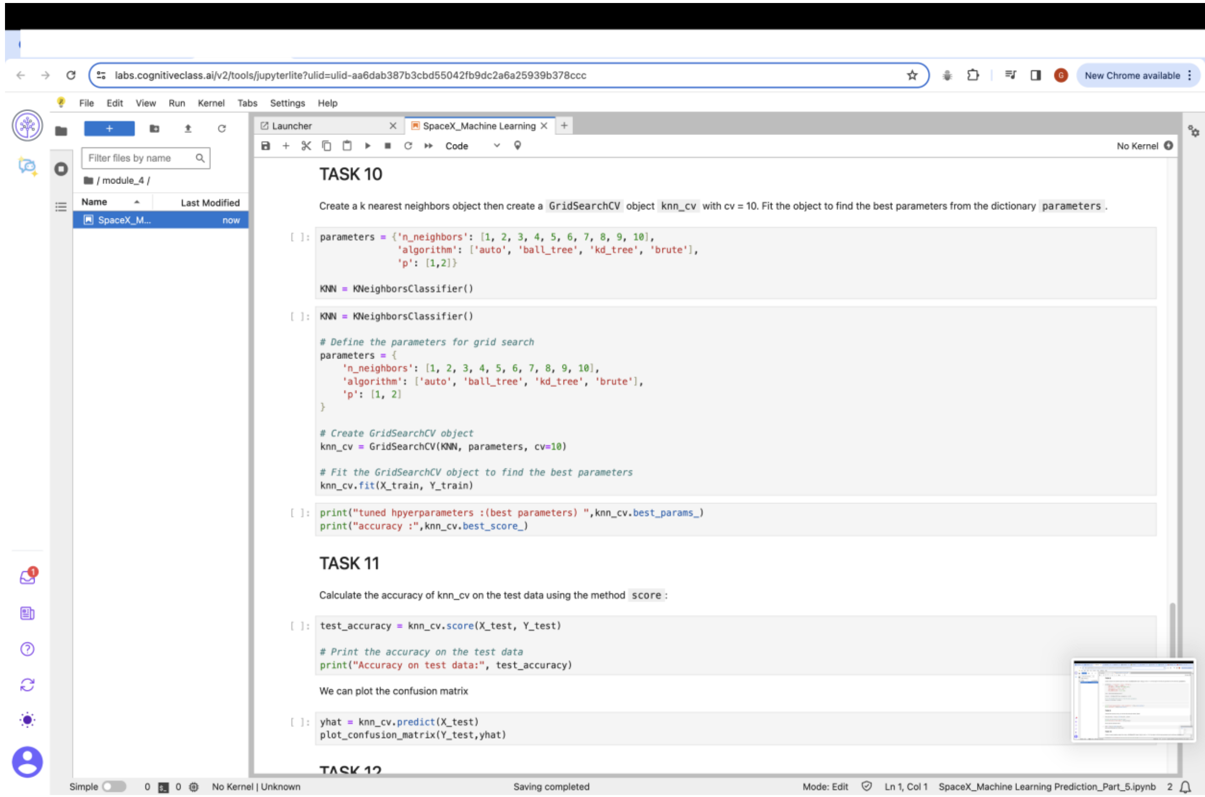




**THE LAB WAS NOT RESPONDING SO RESULTS ARE NOT THER JUST THE CODE...PLEASE FIR THIS PROBLEM OF LAB IT'S VERY HECTIC AND TIME TAKING.

Predictive Analysis Results





Conclusion

Key Takeaways:

- Explored the commercial space age and the role of SpaceX in revolutionizing space travel.
- Conducted data collection, wrangling, and exploratory analysis to understand SpaceX launch data.
- Utilized predictive analysis methodologies to train machine learning models for predicting first-stage landing success.
- Examined various algorithms including Logistic Regression, Support Vector Machines, Decision Tree Classifier, and K-nearest Neighbors.
- Evaluated model performance using techniques like Grid Search and confusion matrices.

Significance of Findings:

- Insights gained from predictive analysis can inform decision-making processes for space missions.
- Accurate predictions of first-stage landing success contribute to cost-effectiveness and efficiency in space exploration.

Future Directions:

- Further refinement of predictive models through additional feature engineering and algorithm optimization.
- Integration of real-time data streams and advanced machine learning techniques for enhanced predictive capabilities.
- Collaboration with space agencies and industry stakeholders to apply findings in practical space missions.

Acknowledgments:

- Acknowledge the contributions of team members, mentors, and collaborators throughout the project.
- Appreciation to the SpaceX REST API and other data sources for providing valuable data for analysis.

Charting a Course for Space Exploration

Introduction:

- Embark on a journey into the realm of space exploration, where innovation and technology converge to redefine the boundaries of possibility.

Exploring SpaceX's Impact:

- Venture into the realm of commercial space travel, where companies like SpaceX are democratizing access to the cosmos.
- Witness the transformative impact of SpaceX's achievements, from launching spacecraft to the International Space Station to revolutionizing satellite internet access with Starlink.

Challenges and Opportunities:

- Delve into the complexities of space travel economics, where cost-effective solutions are essential for unlocking the full potential of space exploration.
- Explore how SpaceX's pioneering reuse of rocket components, notably the Falcon 9's first stage, has reshaped the landscape of space launch affordability.

Data Journey:

- Embark on a data-driven odyssey as we traverse through SpaceX's launch data, uncovering insights and trends that illuminate the path towards predictive analysis.

Predictive Analysis Unveiled:

- Unveil the predictive analysis methodologies that power our quest to forecast the success of Falcon 9's first-stage landings.
- Peer into the intricate web of machine learning algorithms, from Logistic Regression to Support Vector Machines, as we seek the optimal model for our predictive endeavors.

Charting the Future:

- Peer through the telescope of possibility as we gaze into the future of space travel, guided by the insights gleaned from our predictive analysis.
- Witness how data-driven decision-making charts the course for cost-effective, efficient, and visionary space missions.

Conclusion:

- Set sail on the starship of innovation as we bid farewell to the present and journey into the cosmos of tomorrow.

Unveiling Innovative Insights

Introduction:

- Embark on a journey into the depths of space exploration, where innovation fuels discovery and redefines the limits of human achievement.

Exploring New Horizons:

- Venture beyond the ordinary as we unravel groundbreaking insights gleaned from the data-rich cosmos of SpaceX's launch endeavors.
- Discover the hidden gems and innovative breakthroughs that illuminate the path to a new era of space exploration.

Innovative Discoveries:

- Peer through the telescope of innovation to witness the emergence of transformative insights that challenge conventional wisdom and redefine our understanding of space travel.
- Unveil the revolutionary concepts and paradigm-shifting revelations that pave the way for a bold new frontier in human exploration.

Unlocking the Future:

- Harness the power of innovation to unlock the untapped potential of space exploration, where every discovery opens the door to a universe of infinite possibilities.
- Chart a course towards a future where innovation propels us to the farthest reaches of the cosmos, where the boundaries of possibility are limited only by the scope of our imagination.

Conclusion:

- Celebrate the spirit of innovation that drives us forward, fueling our quest for discovery and inspiring us to reach for the stars.