Predicting Falcon 9 First Stage Landing Success: A Data Science Approach to Cost Optimization in SpaceX Launches

Ghada ALamoudi

11/20/2024

OUTLINE

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

EXECUTIVE SUMMARY

- **Objective**: Predict if Falcon 9's first stage will land using machine learning.
- **Data**:
  - Cleaned and standardized.
  - Split into training (80%) and test (20%) data.
- **Models Tested**: Logistic Regression, SVM, Decision Tree, KNN.
- **Best Model**:
  - **Logistic Regression**: 83.33% accuracy.
  - **Best Hyperparameters**: `C=0.01, penalty=l2, solver=lbfgs`.
- **Conclusion**:
  - Logistic Regression performed best, but false positives remain.

INTRODUCTION

- **Objective**: Predict if Falcon 9's first stage will land successfully to assess launch costs.
- **Importance**: Accurate predictions help SpaceX and competitors optimize pricing and improve efficiency.
- **Approach**: Use machine learning models (Logistic Regression, SVM, Decision Trees, KNN) to classify landing success.
- **Data**: Historical rocket launch data including mission details and outcomes.

METHODOLOGY

- **Data Preprocessing**:
- Standardized the dataset to scale the features.
- Split the data into training (80%) and testing (20%) sets.
- **Model Selection**:
- Used **Logistic Regression**, **Support Vector Machines (SVM)**, **Decision Trees**, and **K-Nearest Neighbors (KNN)** for classification.
- **Hyperparameter Tuning**:
- Applied **GridSearchCV** to identify the best hyperparameters for each model.
- **Model Evaluation**:
- Used **accuracy** and **confusion matrix** to evaluate performance on test data.

RESULTS

Model	Best Hyperparameters	Test Accuracy	Key Observation
Logistic Regression	C=0.01, penalty='l2', solver='lbfgs'	83.33%	High accuracy with some false positives.
Support Vector Machines	C=1.0, gamma=0.0316, kernel='sigmoid	84.82%	Performed well but with slightly higher complexity.
Decision Tree Classifier	Multiple parameter optimizations	93.33%	High accuracy, but prone to overfitting.
K-Nearest Neighbors	n_neighbors=5, algorithm='auto', p=2	76.67%	Lower accuracy compared to other models.

Conclusion:

- **Decision Tree Classifier performed the best with 93.33% accuracy.**

Data Collection & Wrangling :Key Findings

- **Data Sources**: Combined `dataset_part_2.csv` and `dataset_part_3.csv`.
- **Data Cleaning**: Verified no missing values; ensured data consistency.
- **Feature Preparation**: Standardized numerical features (e.g., `Payload Mass`) and encoded categorical variables.
- **Target Variable**: Created `Class` column (1 = Landed, 0 = Not Landed).
- **Data Split**: Divided into 80% training and 20% testing sets.

In [13]:

```
df.head(5)
```

Out[13]:

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

SQL Analysis: Key Findings

- **Launch Success Rate**: Queried data to calculate the proportion of successful landings.
- **Feature Impact**: Identified factors like `Booster Version` and `Launch Site` influencing outcomes.
- **Launch Site Comparison**: SQL queries revealed variations in success rates across launch sites.
- **Payload Analysis**: Filtered payload ranges and their impact on landing success.

SQL provided insights into patterns and relationships, forming the basis for machine learning predictions.

In [82]:
```
%sql SELECT "Landing_Outcome", COUNT(*) AS "Total_Missions" FROM SPACEXTABLE GROUP BY "Landing_Outcome";
```

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

Out[82]:

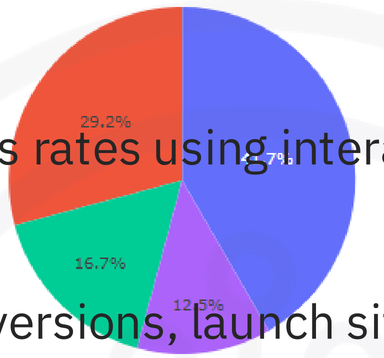
Landing_Outcome	Total_Missions
Controlled (ocean)	5
Failure	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	21
No attempt	1
Precluded (drone ship)	1
Success	38
Success (drone ship)	14
Success (ground pad)	9
Uncontrolled (ocean)	2

Exploratory Data Analysis (EDA) with Visualizations

- **Correlation Analysis**:
Heatmap visualizations highlighted relationships between features such as `Payload Mass` and success rate.
- **Landing Outcomes**:
Bar plots displayed the distribution of successful and unsuccessful landings across launch sites.
- **Payload Distribution**:
Histograms showed payload mass ranges for successful vs. unsuccessful landings.
- **Temporal Trends**:
Line charts analyzed trends over time, linking launch dates to landing success.
- **Feature Insights**:
Scatter plots provided insights into `Payload Mass` and booster version impacts.

Visualizations revealed crucial patterns, guiding feature selection for machine learning.

Interactive Visualization and Dashboard

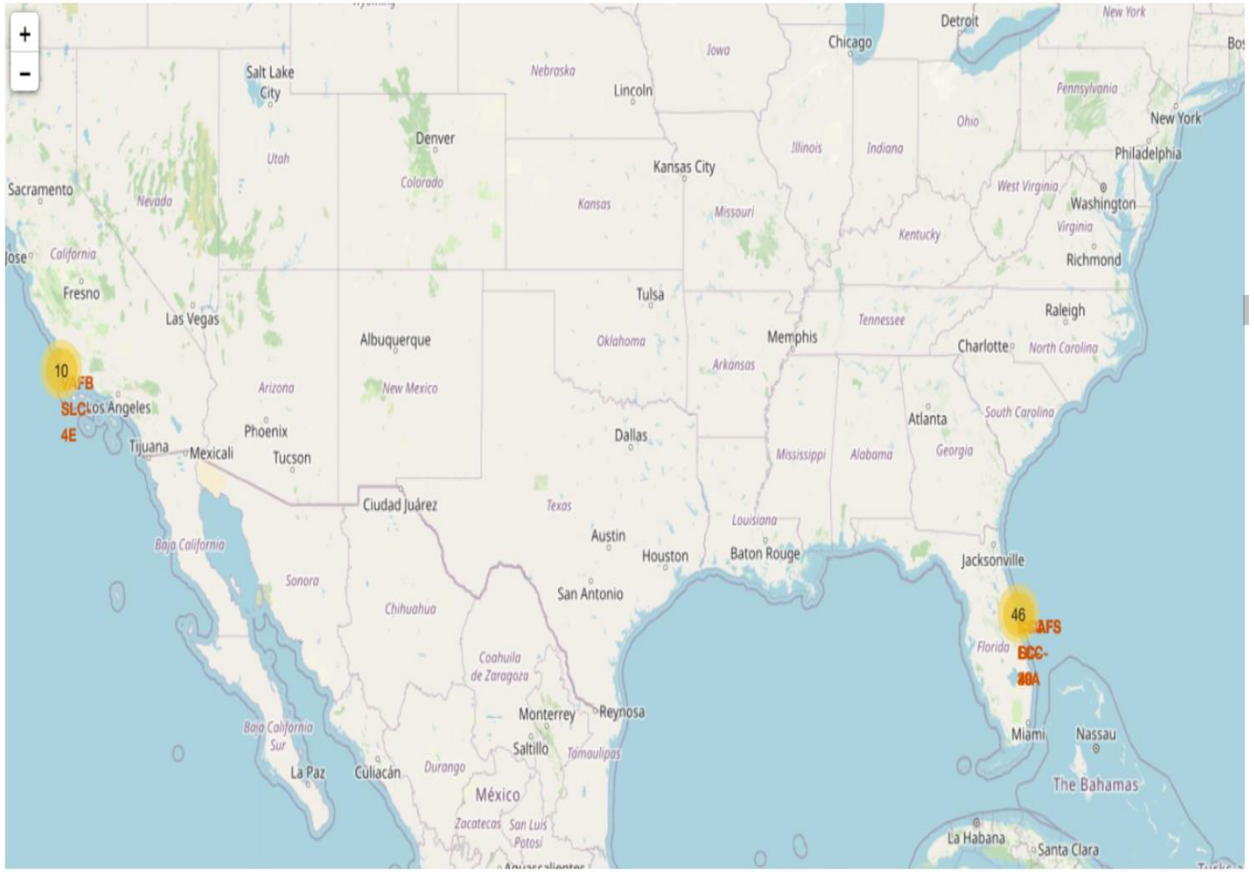


- **Interactive Maps**:
  Visualized launch sites and their success rates using interactive maps (e.g., Folium) for better spatial analysis.
- **Success Rate Filters**:
  Dashboards included filters for booster versions, launch sites, and payload mass ranges to explore success patterns dynamically.
- **Customizable Graphs**:
  Users could switch between bar charts, scatter plots, and line graphs to analyze trends interactively.
- **Payload and Outcome Correlation**:
  Interactive scatter plots allowed users to investigate the relationship between payload mass and success probability.
- **Real-Time Insights**:
  Dashboards enabled users to explore the dataset dynamically, gaining actionable insights for predictive modeling.

This feature significantly enhanced exploratory analysis, making data insights more accessible and user-friendly.

Interactive Visualization with Folium

- Mapped all SpaceX launch sites with interactive markers for success/failure rates.
- Integrated hover tooltips showing payload mass and mission outcomes.
- Added dynamic layers to categorize launches by outcome.
- Enabled zoom and navigation for spatial pattern analysis.

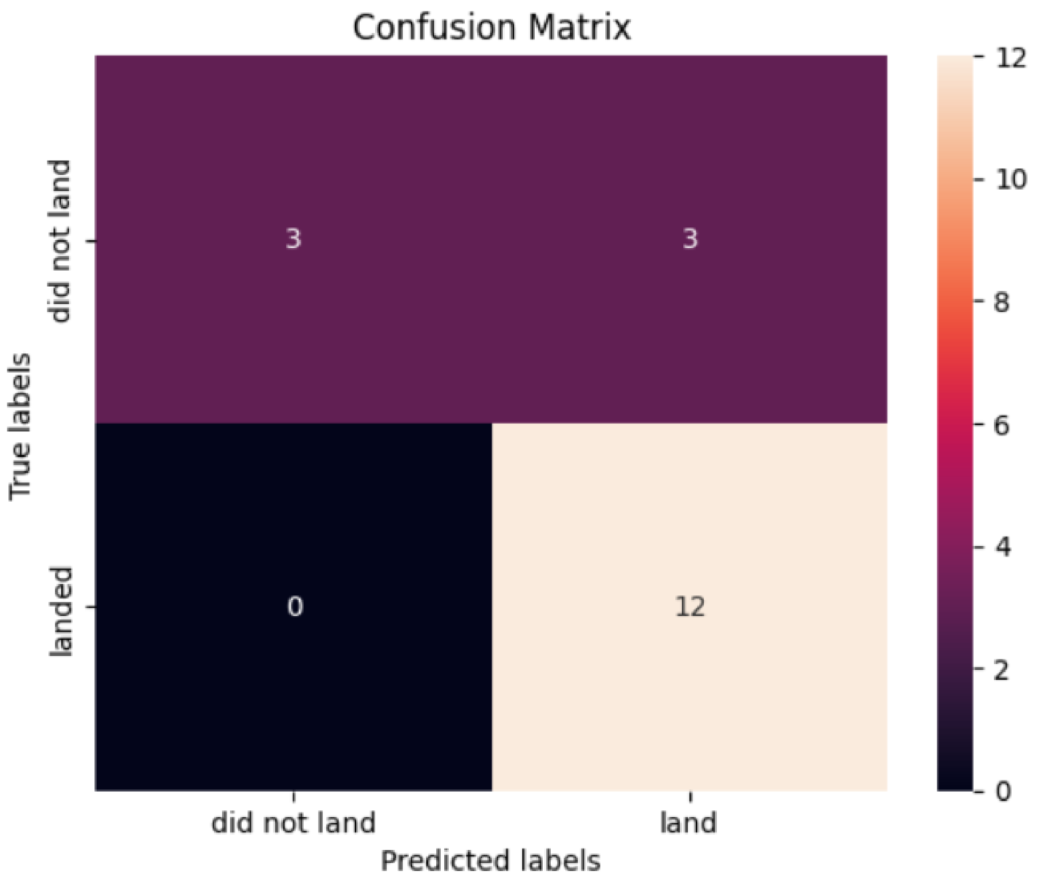


Skills Network

IBM

Machine Learning Predictions Lab

❖ **Data Preprocessing**:
&. Standardized the dataset using consistency.
&. Split data into training (80%) and test (20%) sets.
❖ **Model Training**:
&. Used **Logistic Regression**, **Support Vector Machines (SVM)**, **Decision Trees**, and **K-Nearest Neighbors (KNN)**.
❖ **GridSearchCV** :for optimal model configurations.
❖ **Best Model Performance**:
&. Logistic Regression achieved the best accuracy on test data (**83.33%**).
&. Confusion matrix analysis highlighted minimal false positives and high true positive rates.
❖ **Insights**:
&. Predictive models effectively classify whether the first stage will land, assisting in cost-saving strategies.



DASHBOARD

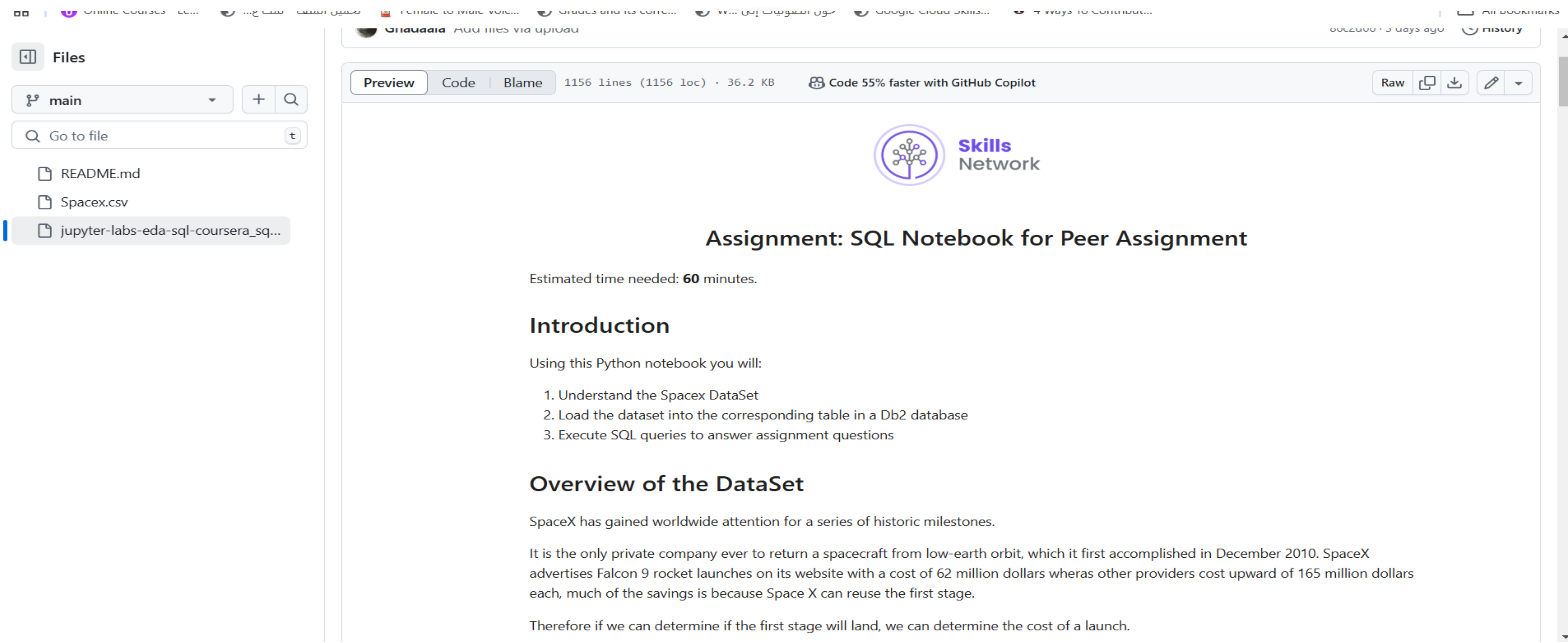
https://github.com/Ghadaala

Assignment: Exploring and Preparing Data

https://github.com/Ghadaala/Assignment-Exploring-and-Preparing-Data

• Assignment: SQL Notebook for Peer Assignment

https://github.com/Ghadaala/Assignment-SQL-Notebook-for-Peer-Assignment/blob/main/jupyter-labs-eda-sql-coursera_sqllite%20(1).ipynb



Files

main

Go to file

README.md

Spacex.csv

jupyter-labs-eda-sql-coursera_sq...

Preview | Code | Blame | 1156 lines (1156 loc) · 36.2 KB | Code 55% faster with GitHub Copilot | Raw

Skills Network

Assignment: SQL Notebook for Peer Assignment

Estimated time needed: **60** minutes.

Introduction

Using this Python notebook you will:

1. Understand the Spacex DataSet
2. Load the dataset into the corresponding table in a Db2 database
3. Execute SQL queries to answer assignment questions

Overview of the DataSet

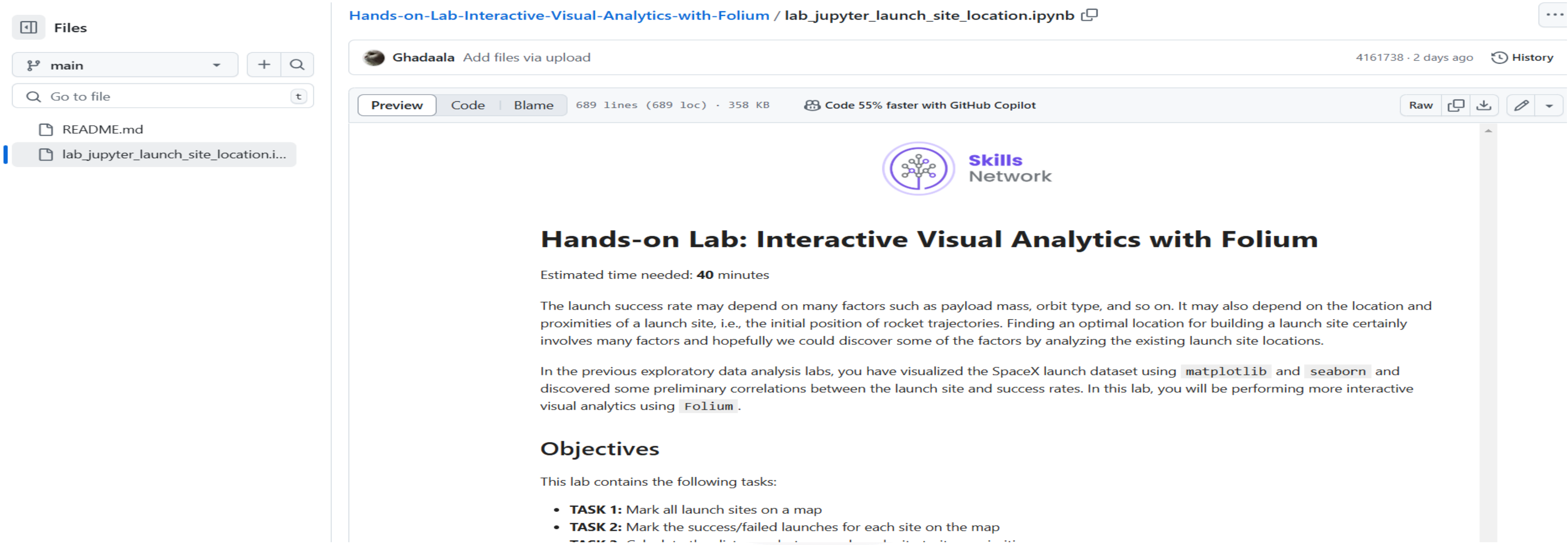
SpaceX has gained worldwide attention for a series of historic milestones.

It is the only private company ever to return a spacecraft from low-earth orbit, which it first accomplished in December 2010. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars wheras other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage.

Therefore if we can determine if the first stage will land, we can determine the cost of a launch.

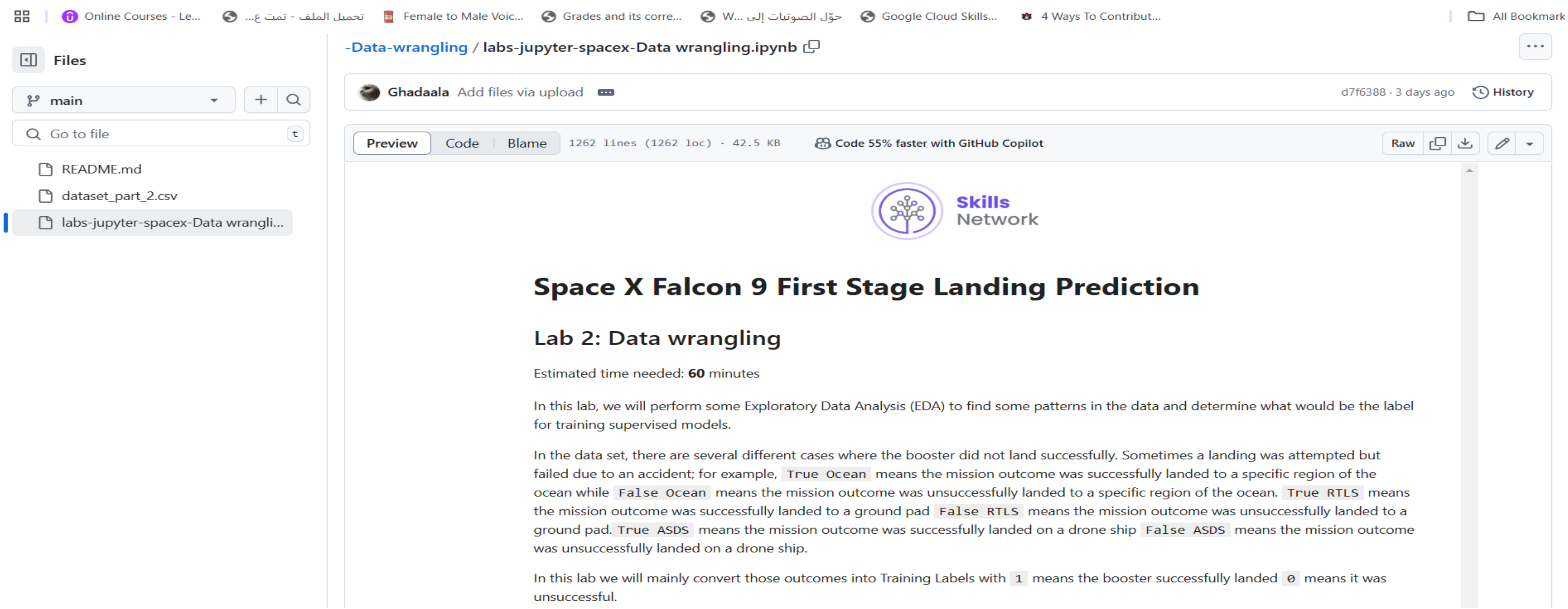
Interactive Visual Analytics with Folium

https://github.com/Ghadaala/Hands-on-Lab-Interactive-Visual-Analytics-with-Folium/blob/main/lab_jupyter_launch_site_location.ipynb



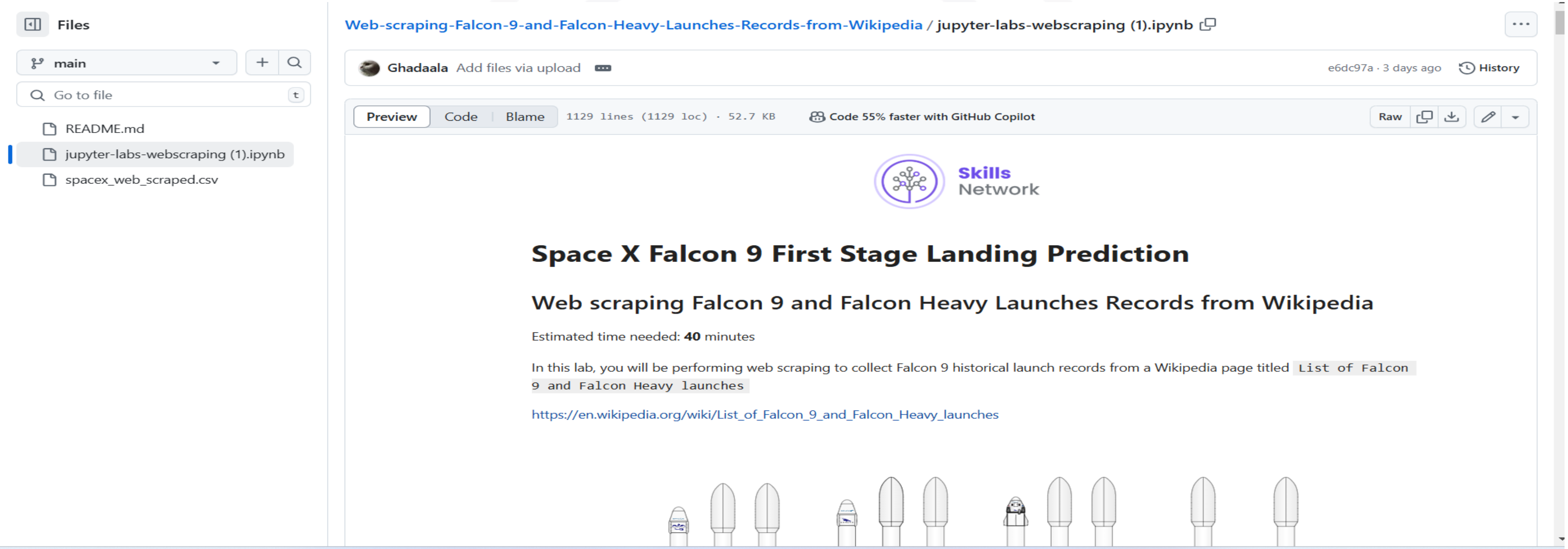
Data wrangling

- https://github.com/Ghadaala/-Data-wrangling/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb



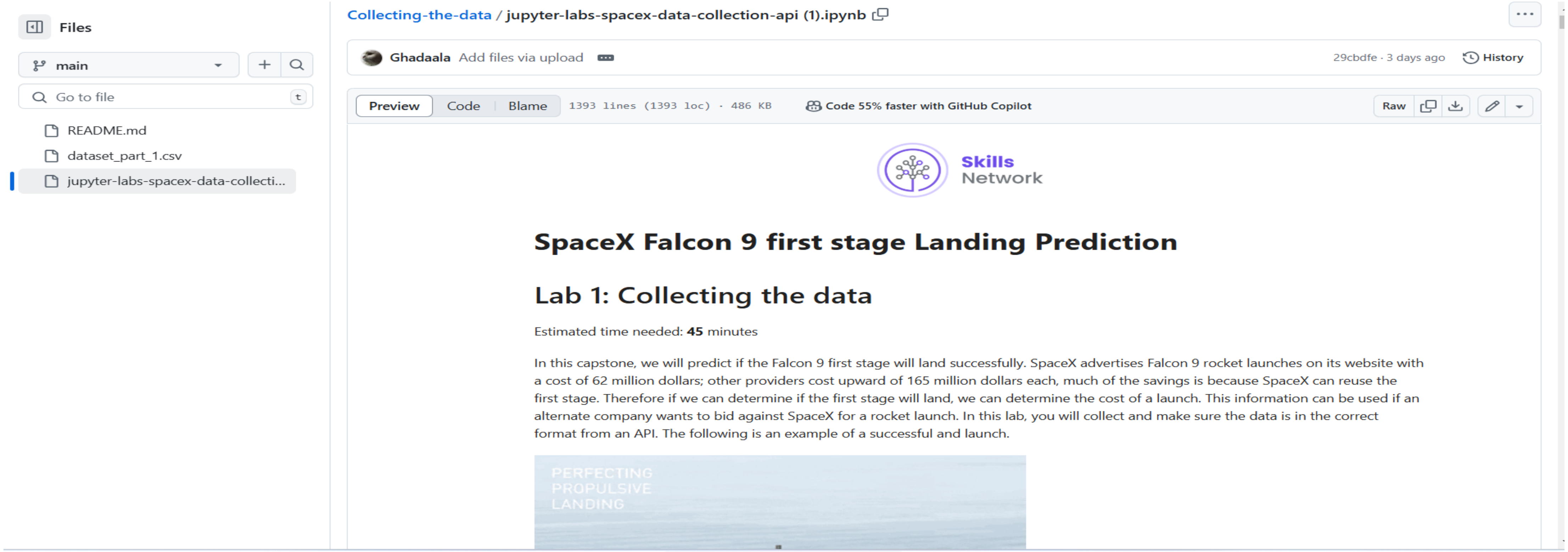
Web scraping Falcon 9 and Falcon Heavy Launches Records from Wikipedia

https://github.com/Ghadaala/-Data-wrangling/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

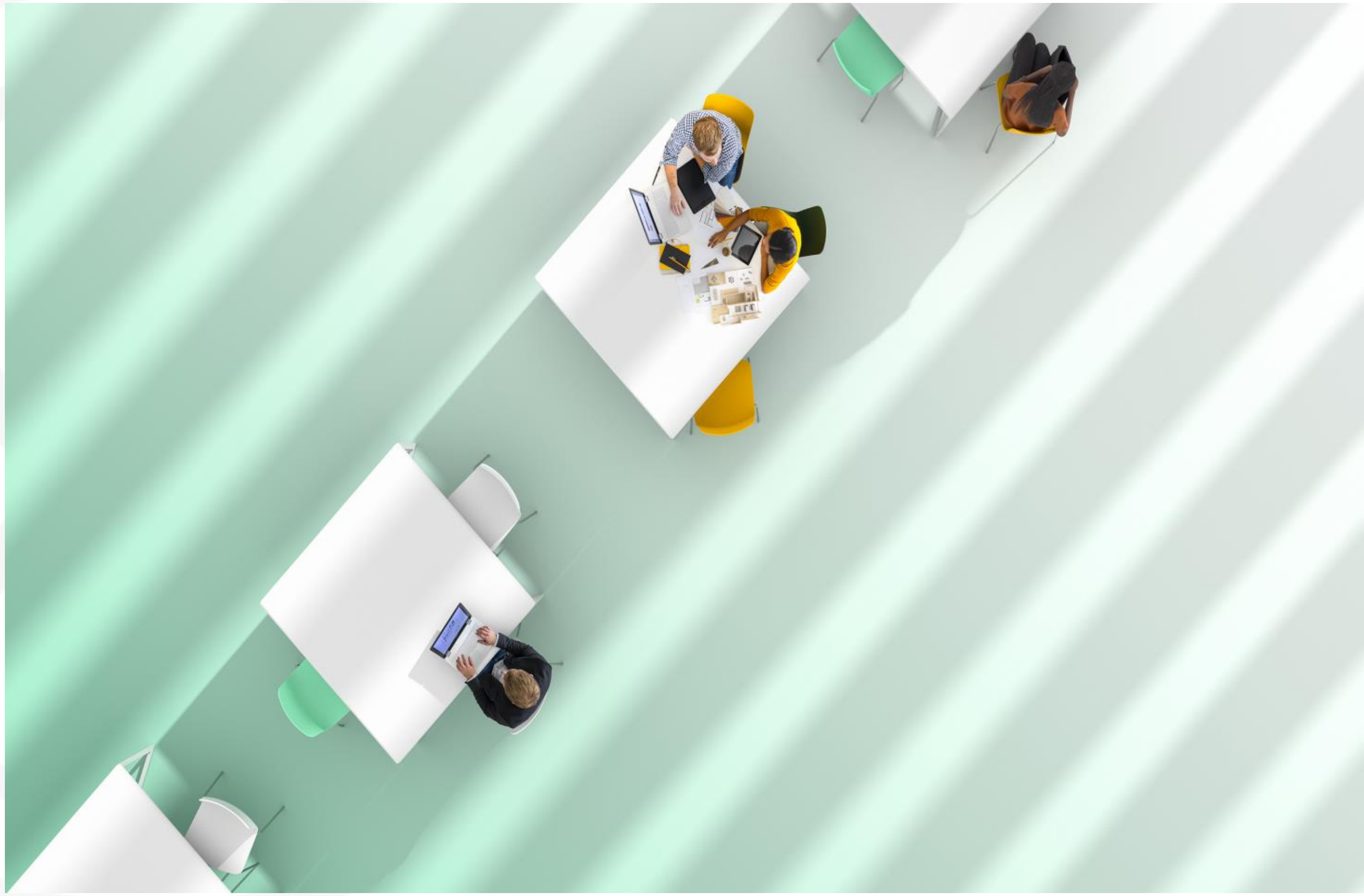


Collecting the data

https://github.com/Ghadaala/Collecting-the-data/blob/main/jupyter-labs-spacex-data-collection-api%20(1).ipynb



DISCUSSION



OVERALL FINDINGS & IMPLICATIONS

Findings

- **Decision Tree Classifier**: Best performer with **93.33% accuracy**.
- **Logistic Regression**: Reliable with **83.33% accuracy**.
- **SVM**: Competitive at **84.82% accuracy** but resource-heavy.
- **KNN**: Least effective at **76.67% accuracy**

Implications

- **Cost Prediction**: Supports pricing strategies for rocket launches.
- **Model Choice**: Decision Tree is ideal for accuracy; Logistic Regression for simplicity.
- **Business Impact**: Enhances decision-making for space providers and clients.

CONCLUSION

- **Best Model**: Decision Tree with **93.33% accuracy**.
- **Key Insight**: Models can predict rocket landing reliability, aiding cost efficiency.
- **Business Impact**: Supports competitive bidding and strategic decision-making.
- **Future Direction**: Explore more data and model refinements.

APPENDIX

- ➢Sample SQL queries or Python code snippet
- ➢References to any external sources or datasets used.
- ➢videos that can help explain key points in project

Sample SQL queries or Python code snippet

Task 1

Display the names of the unique launch sites in the space mission

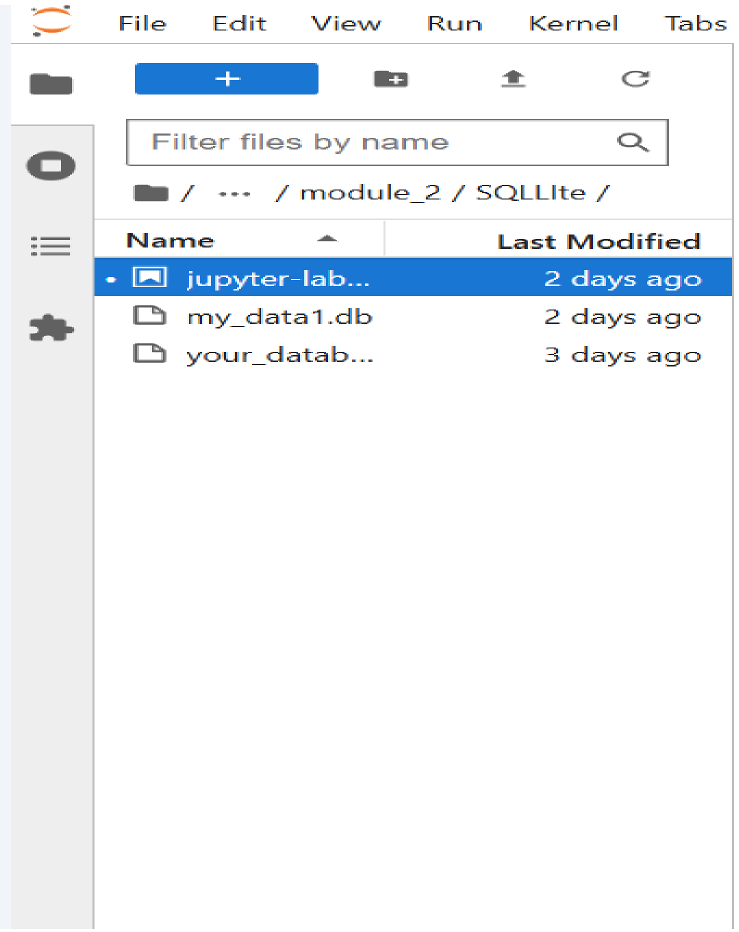
[76]:

```
# Run this if you're using SQLAlchemy connection
%sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
```

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

[76]:

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



References to any datasets used.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1		Date	BoosterVe	PayloadMa	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPa	Block	ReusedCo	Serial	Longitude	Latitude
2	1	6/4/2010	Falcon 9		LEO	CCSFS SLC	None None	1	FALSE	FALSE	FALSE		1	0	B0003	-80.5774	28.56186
3	2	########	Falcon 9	525	LEO	CCSFS SLC	None None	1	FALSE	FALSE	FALSE		1	0	B0005	-80.5774	28.56186
4	3	3/1/2013	Falcon 9	677	ISS	CCSFS SLC	None None	1	FALSE	FALSE	FALSE		1	0	B0007	-80.5774	28.56186
5	4	########	Falcon 9	500	PO	VAFB SLC	False Ocea	1	FALSE	FALSE	FALSE		1	0	B1003	-120.611	34.63209
6	5	########	Falcon 9	3170	GTO	CCSFS SLC	None None	1	FALSE	FALSE	FALSE		1	0	B1004	-80.5774	28.56186
7	6	1/6/2014	Falcon 9	3325	GTO	CCSFS SLC	None None	1	FALSE	FALSE	FALSE		1	0	B1005	-80.5774	28.56186
8	7	########	Falcon 9	2296	ISS	CCSFS SLC	True Ocea	1	FALSE	FALSE	TRUE		1	0	B1006	-80.5774	28.56186
9	8	########	Falcon 9	1316	LEO	CCSFS SLC	True Ocea	1	FALSE	FALSE	TRUE		1	0	B1007	-80.5774	28.56186
10	9	8/5/2014	Falcon 9	4535	GTO	CCSFS SLC	None None	1	FALSE	FALSE	FALSE		1	0	B1008	-80.5774	28.56186
11	10	9/7/2014	Falcon 9	4428	GTO	CCSFS SLC	None None	1	FALSE	FALSE	FALSE		1	0	B1011	-80.5774	28.56186
12	11	########	Falcon 9	2216	ISS	CCSFS SLC	False Ocea	1	FALSE	FALSE	FALSE		1	0	B1010	-80.5774	28.56186
13	12	########	Falcon 9	2395	ISS	CCSFS SLC	False ASDS	1	TRUE	FALSE	TRUE	5e9e3032	1	0	B1012	-80.5774	28.56186
14	13	########	Falcon 9	570	ES-L1	CCSFS SLC	True Ocea	1	TRUE	FALSE	TRUE		1	0	B1013	-80.5774	28.56186
15	14	########	Falcon 9	1898	ISS	CCSFS SLC	False ASDS	1	TRUE	FALSE	TRUE	5e9e3032	1	0	B1015	-80.5774	28.56186
16	15	########	Falcon 9	4707	GTO	CCSFS SLC	None None	1	FALSE	FALSE	FALSE		1	0	B1016	-80.5774	28.56186
17	16	########	Falcon 9	2477	ISS	CCSFS SLC	None ASDS	1	TRUE	FALSE	TRUE	5e9e3032	1	0	B1018	-80.5774	28.56186
18	17	########	Falcon 9	2034	LEO	CCSFS SLC	True RTLS	1	TRUE	FALSE	TRUE	5e9e3032	1	0	B1019	-80.5774	28.56186
19	18	########	Falcon 9	553	PO	VAFB SLC	False ASDS	1	TRUE	FALSE	TRUE	5e9e3033	1	0	B1017	-120.611	34.63209
20	19	3/4/2016	Falcon 9	5271	GTO	CCSFS SLC	False ASDS	1	TRUE	FALSE	TRUE	5e9e3032	1	0	B1020	-80.5774	28.56186
21	20	4/8/2016	Falcon 9	3136	ISS	CCSFS SLC	True ASDS	1	TRUE	FALSE	TRUE	5e9e3032	2	1	B1021	-80.5774	28.56186
22	21	5/6/2016	Falcon 9	4696	GTO	CCSFS SLC	True ASDS	1	TRUE	FALSE	TRUE	5e9e3032	2	0	B1022	-80.5774	28.56186
23	22	########	Falcon 9	3100	GTO	CCSFS SLC	True ASDS	1	TRUE	FALSE	TRUE	5e9e3032	2	1	B1023	-80.5774	28.56186
24	23	########	Falcon 9	2257	ISS	CCSFS SLC	True RTLS	1	TRUE	FALSE	TRUE	5e9e3032	2	1	B1025	-80.5774	28.56186
25	24	########	Falcon 9	4600	GTO	CCSFS SLC	True ASDS	1	TRUE	FALSE	TRUE	5e9e3032	2	0	B1026	-80.5774	28.56186
26	25	9/1/2016	Falcon 9	5500	GTO	CCSFS SLC	None ASDS	1	TRUE	FALSE	TRUE	5e9e3032	3	0	B1028	-80.5774	28.56186

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/labs/module_2/data/Spacex.csv

videos that can help explain key points in project