



DATA ANALYTICS

By : Rudhramyna Amshu

Under guidance of : Dr Jhansi Rani T

Need for Data Analysis

What is Data Analysis?

- Process of collecting, cleaning, organizing, and interpreting data.
- Extracts meaningful insights and identifies patterns in data.
- Supports data-driven decision-making across industries.

Why is Data Analysis Important?

- Helps in making informed decisions based on data.
- Optimizes processes and efficiency in various fields.
- Uncovers hidden patterns and trends that humans may not easily detect.

Phases of Data Analysis

We can obtain actionable insights by transforming raw data using the following key steps.



EXCEL FOR ANALYSIS

Built-in Functions for Data Analysis

- Excel provides a wide range of built-in functions to summarize and analyze data efficiently.
- Common functions include:
 - **SUM()**: Adds a range of numbers.
 - **COUNT()**: Counts the number of values in a range.
 - **MAX() / MIN()**: Finds the highest and lowest values.
 - **AVERAGE()**: Calculates the mean of selected values.

Data Import and Formatting

- Excel allows importing external data into a spreadsheet.
- It can automatically detect delimiters from text/CSV files and format the data properly.

DATABASES AND QUERYING LANGUAGES

1

Schema: Defines the logical structure of a database, including tables, columns, data types, relationships, and constraints.

2

Table Relationships: Connect data across tables using Primary Keys and Foreign Keys. Types include:

- One-to-One
- One-to-Many
- Many-to-Many

3

Relational Database: Stores data in structured tables with rows and columns, ensuring relationships (e.g., MySQL, PostgreSQL).

DATABASES AND QUERYING LANGUAGES

4

SQL (Structured Query Language): Used to create, manipulate, and manage relational databases through queries.

5

Distributed Database: Stores data across multiple servers for better scalability and availability (e.g., Google Spanner, Amazon DynamoDB).

6

NoSQL Database: Non-relational database for handling large-scale, unstructured data (e.g., MongoDB, Cassandra).

STATISTICS FOR ANALYSIS

Observations, Variables and Values

An observation is a single data point, a variable is a characteristic being measured, value is the specific data recorded for a variable.

Population vs Sample

A population includes all possible data points in a study, while a sample is a subset used to conclude the population.

Descriptive vs. Inferential Statistics

Descriptive statistics summarize and organize data, inferential statistics analyze samples to make predictions about a population.

TABLEAU

- Tableau – A data visualization & BI tool for interactive data analysis.
- Tableau Prep – Cleans, shapes, and prepares data for analysis. Tableau
- Public – Free platform for creating and sharing visualizations online.
- Tableau Desktop – Professional tool for building reports & dashboards.
- Tableau Explorer – Web-based tool for analyzing existing dashboards.
- Tableau Dashboards – Interactive visual reports combining charts, tables, and filters.

BIAS IN DATA ANALYTICS

Bias in data analysis is systematic error that distorts insights, leading to inaccurate or unfair conclusions.

● Selection Bias

When the sample is not representative of the population.

● Confirmation Bias

When analysts favor data that supports their preexisting beliefs.

● Information Bias

When data is collected, recorded, or interpreted inaccurately.

● Interpretation Bias

When data is misinterpreted due to subjective analysis or misleading visualizations.

METHODS TO AVOID BIAS

- Be aware that bias exists.
- Record your assumptions and hypotheses before beginning your analysis.
- Avoid selecting only data and methods that you believe will support your assumptions.
- Validate your data sources and the methodology used to collect the data.
- Focus on larger patterns and trends Be open-minded and impartial in your analysis.

ETHICS IN DATA ANALYSIS

The CIA Triad (Confidentiality, Integrity, Availability) ensures ethical and secure data handling

Confidentiality

Prevent unauthorized access

Integrity

Ensure data accuracy & reliability

Availability

Ensure reliable data access



PROJECT – DATA VISUALIZATION

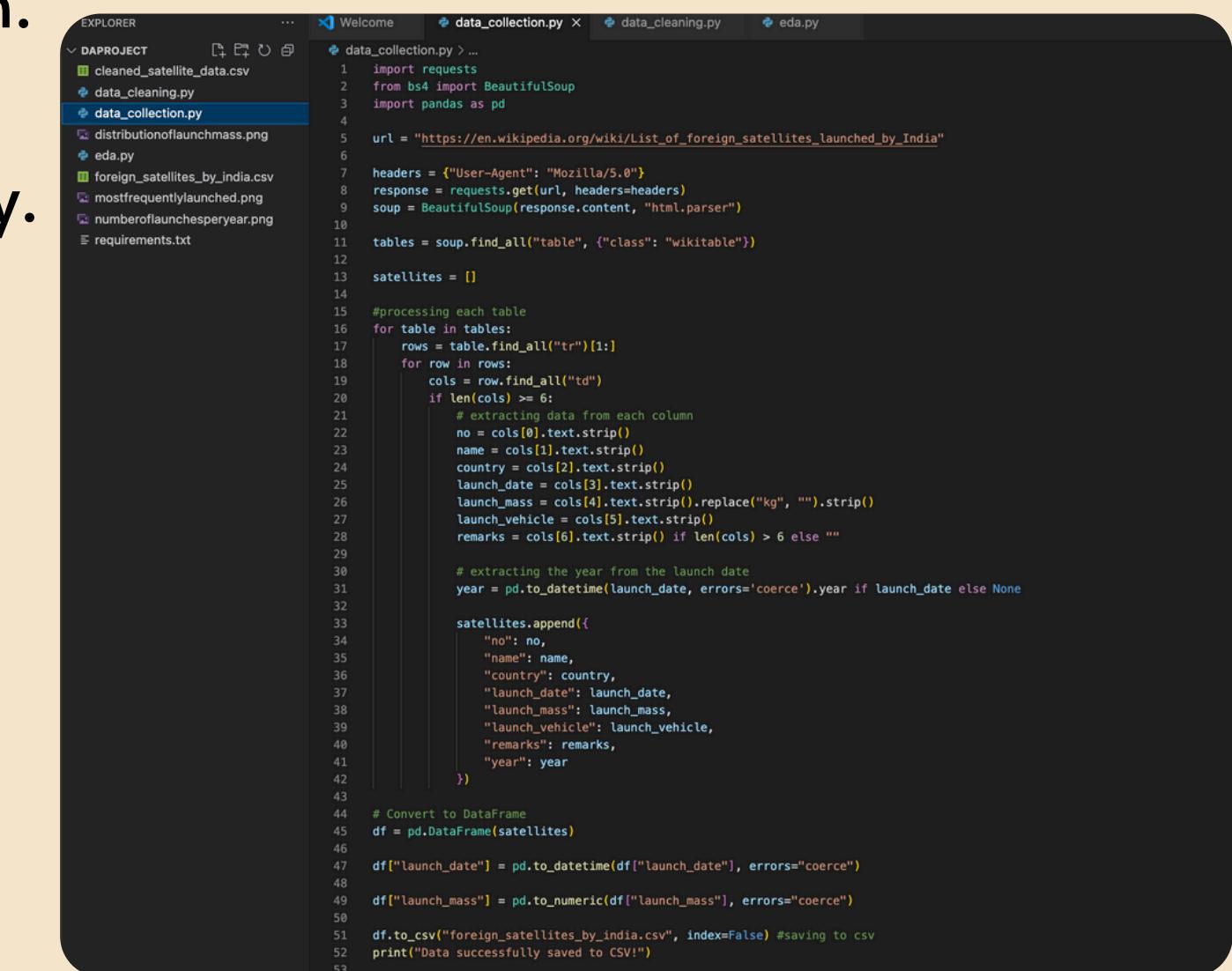
• Foreign Satellites Launched by India

I. Introduction

- Objective: Analyze trends in foreign satellite launches by India.
- Data Source: Wikipedia – List of foreign satellites launched by India.
- Tools Used: Python, BeautifulSoup, Pandas, Matplotlib, Seaborn.

2. Data Collection

- Method: Web scraping using BeautifulSoup and requests library.
- Steps:
 - Send HTTP request to Wikipedia URL.
 - Parse the webpage to extract tabular data.
 - Store data into a structured list.
- Key Tools:
 - Python
 - requests (for fetching webpage content)
 - BeautifulSoup (for parsing HTML tables)



```

EXPLORER
DAPROJECT
cleaned_satellite_data.csv
data_cleaning.py
data_collection.py
distributionoflauchmass.png
eda.py
foreign_satellites_by_india.csv
mostfrequentlylaunched.png
numberoflauchesperyear.png
requirements.txt

Welcome data_collection.py x data_cleaning.py eda.py

data_collection.py > ...
1 import requests
2 from bs4 import BeautifulSoup
3 import pandas as pd
4
5 url = "https://en.wikipedia.org/wiki/List_of_foreign_satellites_launched_by_India"
6
7 headers = {"User-Agent": "Mozilla/5.0"}
8 response = requests.get(url, headers=headers)
9 soup = BeautifulSoup(response.content, "html.parser")
10
11 tables = soup.find_all("table", {"class": "wikitable"})
12
13 satellites = []
14
15 #processing each table
16 for table in tables:
17     rows = table.find_all("tr")[1:]
18     for row in rows:
19         cols = row.find_all("td")
20         if len(cols) >= 6:
21             # extracting data from each column
22             no = cols[0].text.strip()
23             name = cols[1].text.strip()
24             country = cols[2].text.strip()
25             launch_date = cols[3].text.strip()
26             launch_mass = cols[4].text.strip().replace("kg", "").strip()
27             launch_vehicle = cols[5].text.strip()
28             remarks = cols[6].text.strip() if len(cols) > 6 else ""
29
30             # extracting the year from the launch date
31             year = pd.to_datetime(launch_date, errors='coerce').year if launch_date else None
32
33             satellites.append({
34                 "no": no,
35                 "name": name,
36                 "country": country,
37                 "launch_date": launch_date,
38                 "launch_mass": launch_mass,
39                 "launch_vehicle": launch_vehicle,
40                 "remarks": remarks,
41                 "year": year
42             })
43
44 # Convert to DataFrame
45 df = pd.DataFrame(satellites)
46
47 df["launch_date"] = pd.to_datetime(df["launch_date"], errors="coerce")
48
49 df["launch_mass"] = pd.to_numeric(df["launch_mass"], errors="coerce")
50
51 df.to_csv("foreign_satellites_by_india.csv", index=False) #saving to csv
52 print("Data successfully saved to CSV!")

```

PROJECT – DATA VISUALIZATION

13

3. Data Cleaning

- Steps:

- Convert column names to lowercase and replace spaces.
- Standardize launch vehicle names.
- Handle missing values (fillna method).
- Convert data types (dates, numeric values).
- Extract year from launch date.

- Key Tools:

- Pandas (.fillna(), .to_datetime(), .to_numeric())

```
data_cleaning.py > ...
import pandas as pd

df = pd.read_csv('foreign_satellites_by_india.csv')

print("Initial DataFrame:")
print(df.head())

df.columns = df.columns.str.lower().str.replace(' ', '_')

# standardizing launch vehicle names
df['launch_vehicle'] = df['launch_vehicle'].str.upper().str.strip()

df.fillna(method='ffill', inplace=True)

df['launch_mass'] = pd.to_numeric(df['launch_mass'], errors='coerce')

df['launch_date'] = pd.to_datetime(df['launch_date'], errors='coerce')

df['year'] = df['launch_date'].dt.year

print("\nCleaned DataFrame:")
print(df.head())

df.to_csv('cleaned_satellite_data.csv', index=False)

print("\nData cleaning complete. Cleaned data saved to 'cleaned_satellite_data.csv'.")
```

PROJECT – DATA VISUALIZATION

14

4. Exploratory Data Analysis (EDA)

- Steps:

- Identify frequently used launch vehicles (sns.countplot).
- Analyze launches per year (sns.countplot).
- Examine the distribution of launch mass (sns.histplot).

- Key Tools:

- Matplotlib
- Seaborn (countplot, histplot)

```
eda.py > ...
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd

# loading the cleaned data
df = pd.read_csv('cleaned_satellite_data.csv')

# analyzing trends in payload types over time
# identifying the most frequently used launch vehicles
plt.figure(figsize=(10, 6))
sns.countplot(data=df, x='launch_vehicle', order=df['launch_vehicle'].value_counts().index)
plt.title('Most Frequently Used Launch Vehicles')
plt.xticks(rotation=45)
plt.show()

# analyzing the number of launches per year
plt.figure(figsize=(10, 6))
sns.countplot(data=df, x='year', order=df['year'].sort_values().unique())
plt.title('Number of Launches Per Year')
plt.xticks(rotation=45)
plt.show()

# analyzing the distribution of launch mass
plt.figure(figsize=(10, 6))
sns.histplot(df['launch_mass'].dropna(), bins=20, kde=True)
plt.title('Distribution of Launch Mass')
plt.xlabel('Launch Mass (kg)')
plt.show()
```

PROJECT – DATA VISUALIZATION

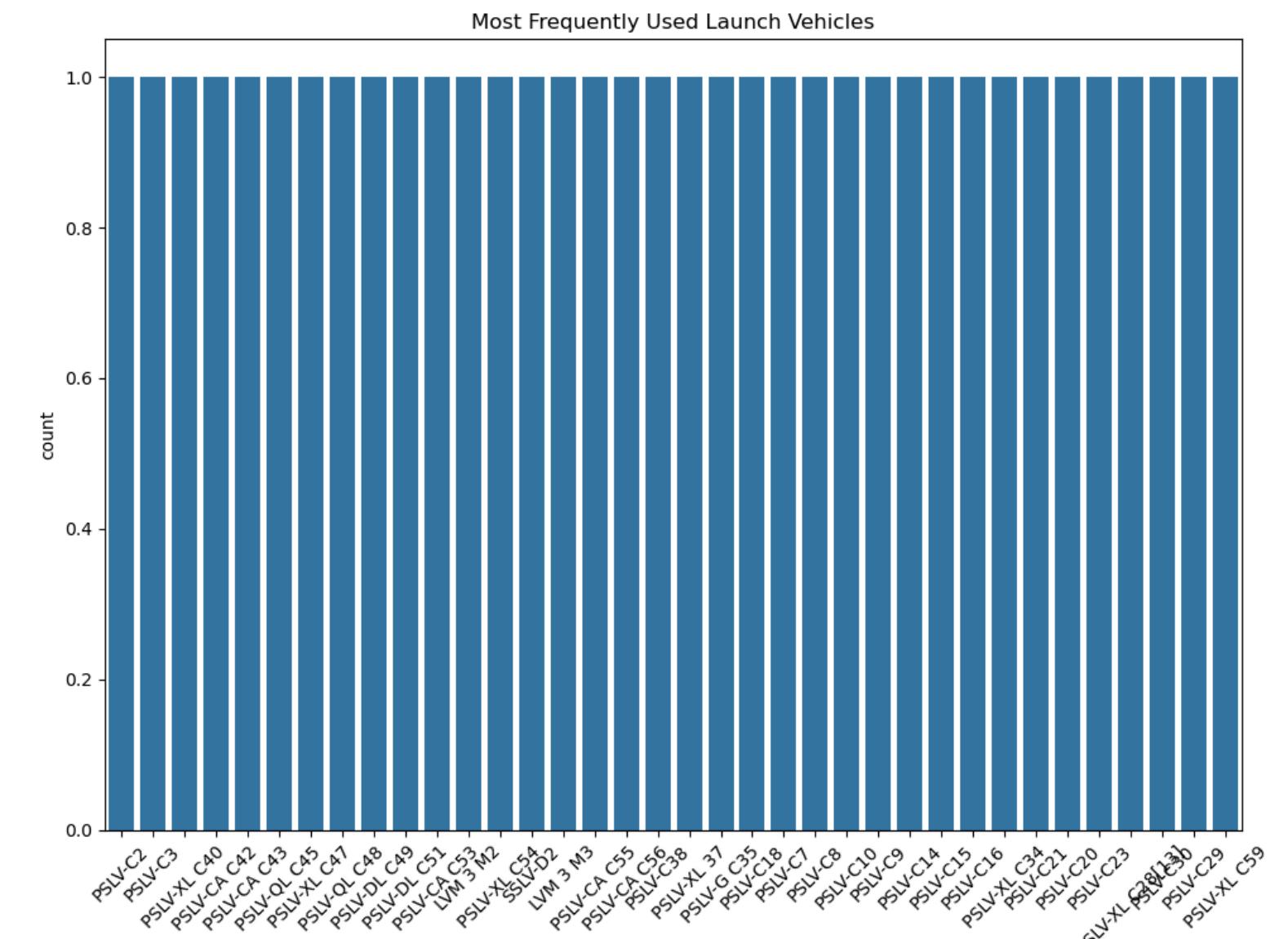
15

5. Data Storage & Export

- Steps:
 - Convert cleaned data into a Pandas DataFrame.
 - Save to CSV (`to_csv()`).
 - Use cleaned data for visualization.
- Key Tools:
 - Pandas (`DataFrame, to_csv()`)

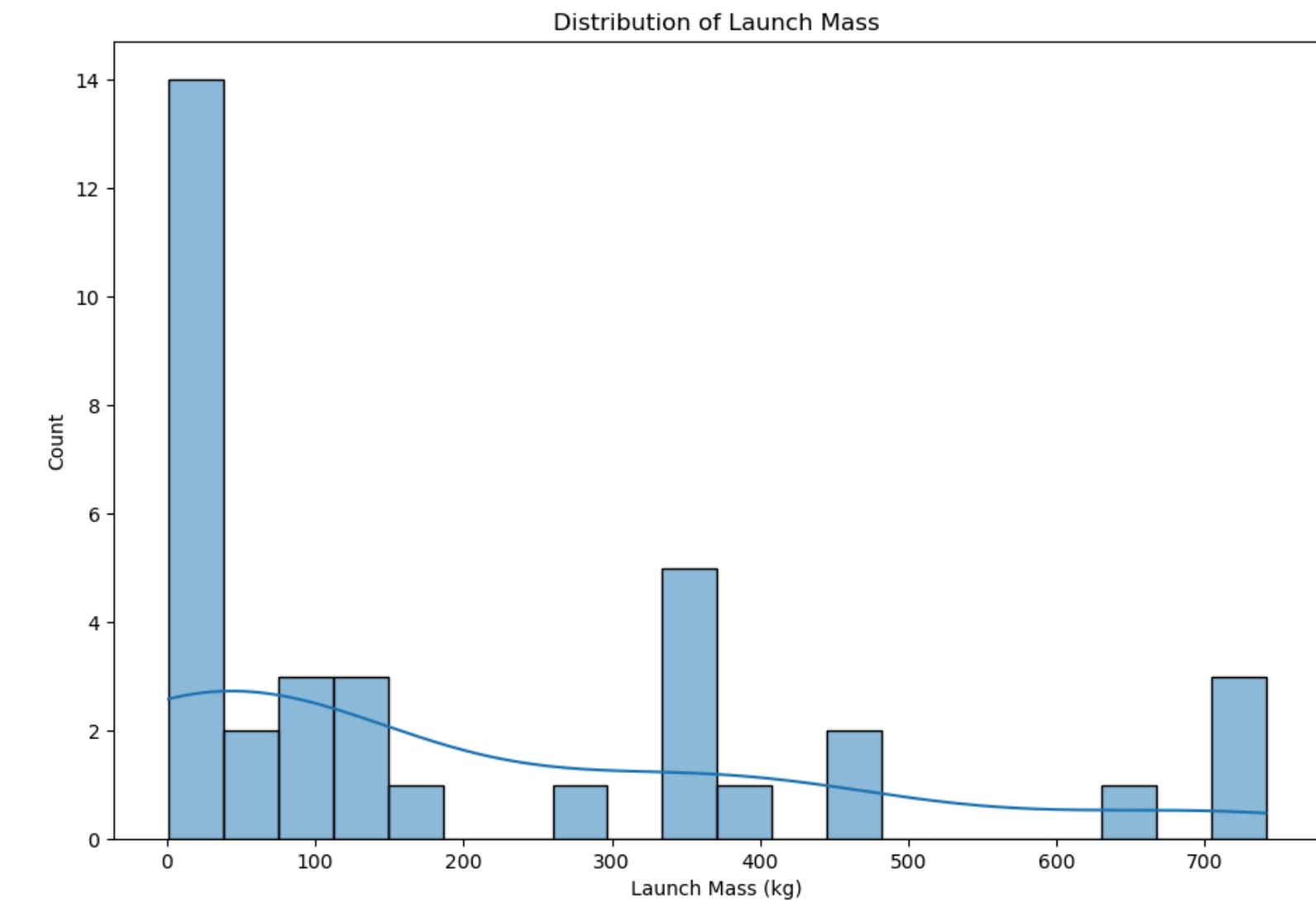
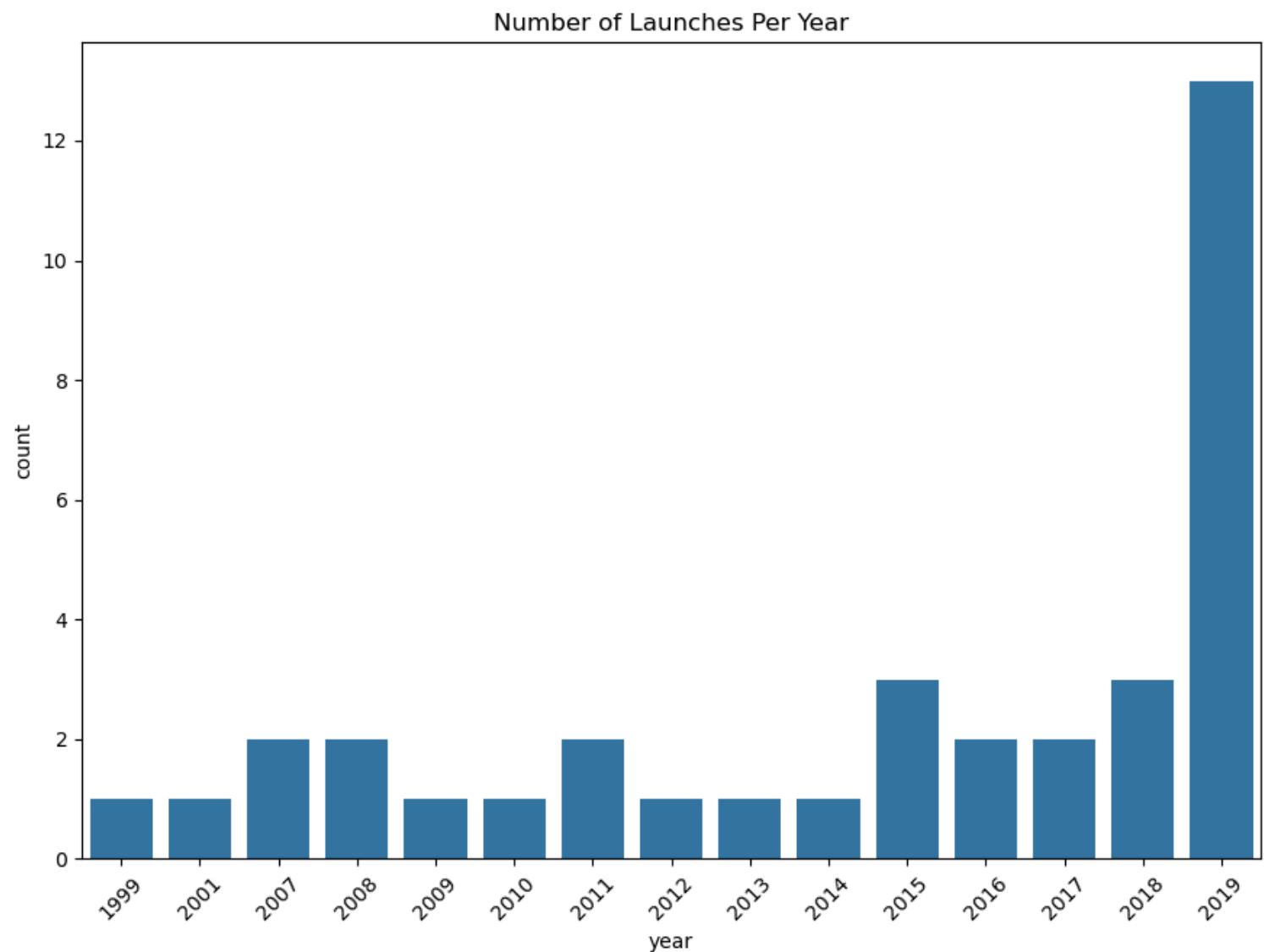
6. Visualization & Insights

- Most Frequently Used Launch Vehicles: Bar chart.
- Number of Launches Per Year: Trend analysis.
- Launch Mass Distribution: Histogram.



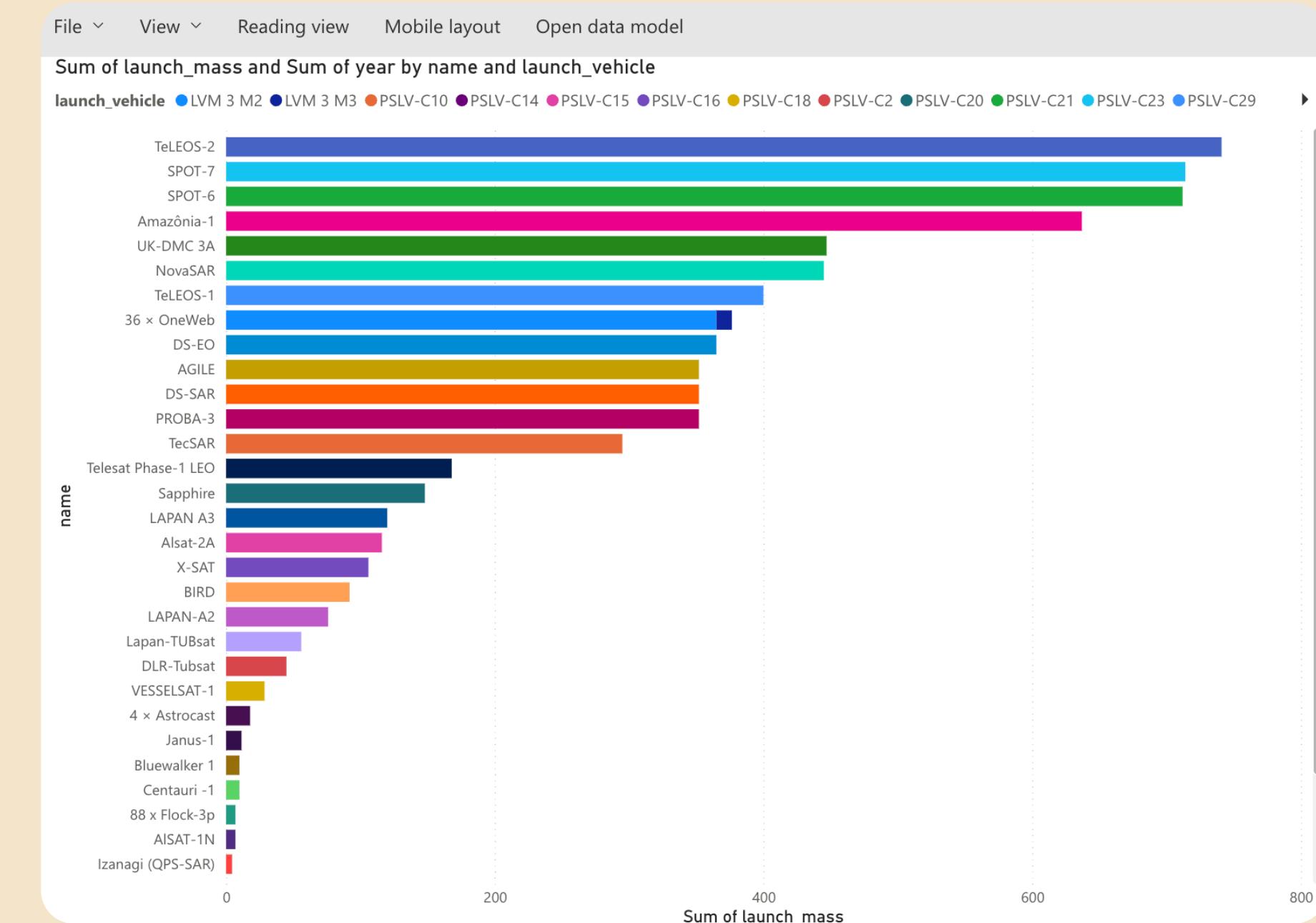
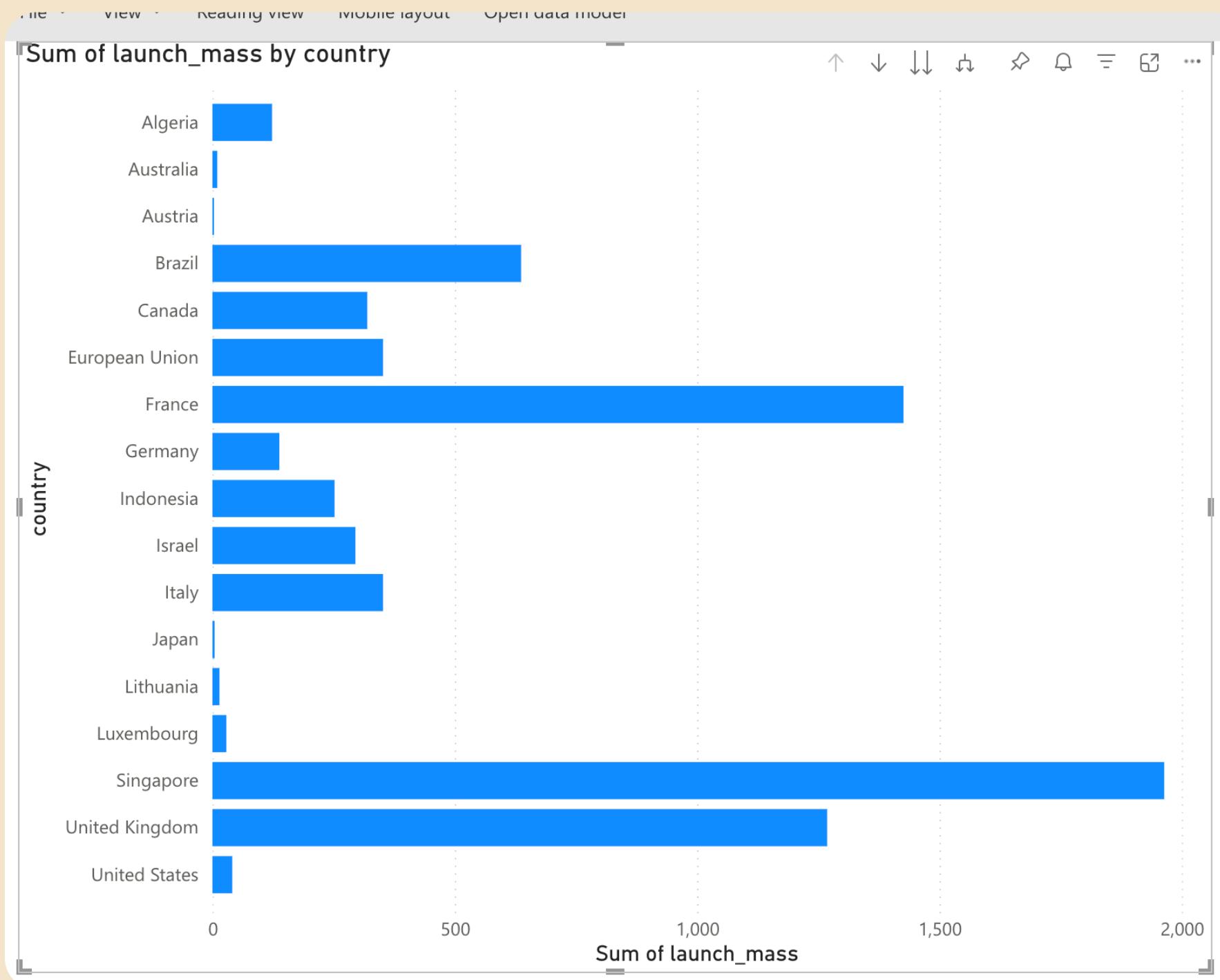
PROJECT – DATA VISUALIZATION

16



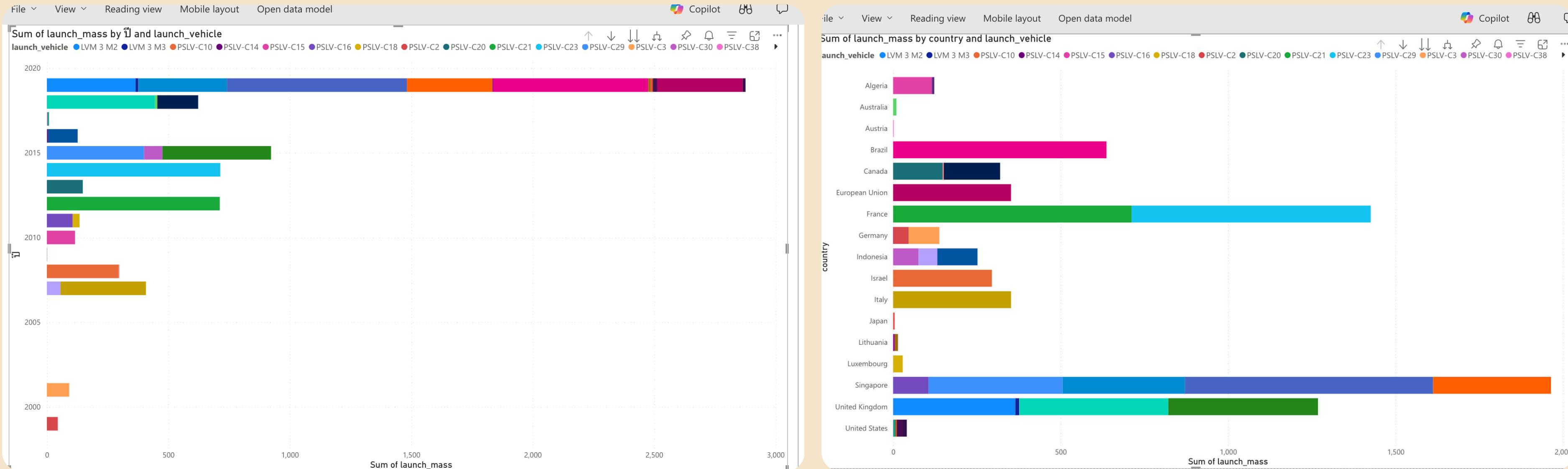
PROJECT – DATA VISUALIZATION

17



PROJECT – DATA VISUALIZATION

18



CONCLUSION

India has emerged as a global leader in satellite launches, deploying hundreds of foreign satellites with cost-effective and reliable technology.

This project explores the trends in foreign satellite launches by India using data from Wikipedia, leveraging Python for web scraping, data cleaning, and visualization.

By analyzing launch frequencies, preferred launch vehicles, and payload distributions, we uncover key insights into India's growing role in the commercial space sector.



Certificate of Course Completion

Amshu Rudhramyna

has successfully achieved student level credential for completing the *Data Analytics Essentials* course.

The student was able to proficiently:

- Explain how the data analytics process creates value from data.
- Explain the characteristics of data, including formats, availability and methods to acquire.
- Transform data using analytics tools.
- Analyze data using basic statistical and data preparation techniques.
- Complete hands-on lab using Excel, SQL, Tableau and other tools.
- Evaluate and share project portfolio.



Scan to Verify

Lynn Bloomer

Lynn Bloomer
Director, Cisco Networking Academy

GITAM HYD | IIIrd year

Internship I

THANK YOU

Presented By : R Amshu

HU22CSEN0101640