Space_Missions_Analysis

September 29, 2024

1 Introduction

This dataset was scraped from nextspaceflight.com and includes all the space missions since the beginning of Space Race between the USA and the Soviet Union in 1957!

1.0.1 Import Statements

1.0.2 Notebook Presentation

1.0.3 Load the Data

2 Preliminary Data Exploration

- What is the shape of `df_data`?
- How many rows and columns does it have?
- What are the column names?
- Are there any NaN values or duplicates?

```
1- Number of rows: 4324
2- Number of columns: 9
3- Columns:
  - `Unnamed: 0.1`
  - `Unnamed: 0`
  - `Organisation`
  - `Location`
  - `Date`
  - `Detail`
  - `Rocket_Status`
  - `Price`
  - `Mission_Status`
4- Number of duplicates: 0
5- Missingness:
  - `Price`: 3360(77.71%)
```

2.1 Data Cleaning - Check for Missing Values and Duplicates

Consider removing columns containing junk data.

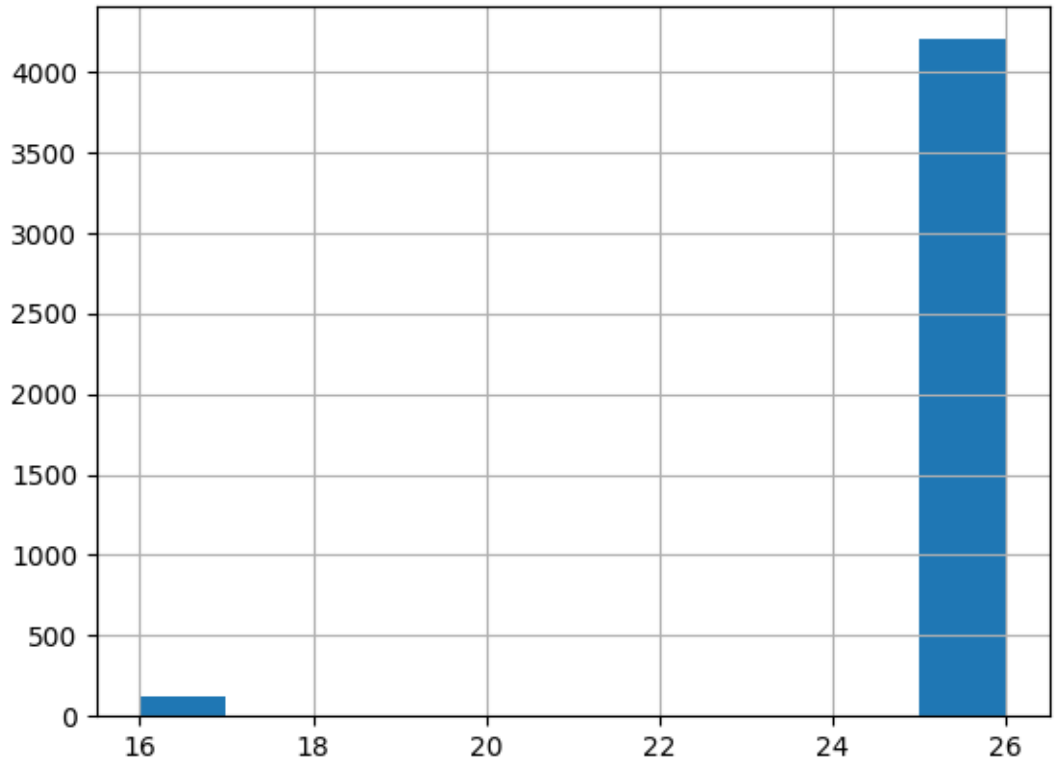
```
   Unnamed: 0.1  Unnamed: 0 Organisation  \
0             0           0       SpaceX
```

```
1             1           1         CASC
2             2           2       SpaceX
3             3           3    Roscosmos
4             4           4          ULA

                                        Location  \
0         LC-39A, Kennedy Space Center, Florida, USA
1  Site 9401 (SLS-2), Jiuquan Satellite Launch Ce…
2                      Pad A, Boca Chica, Texas, USA
3       Site 200/39, Baikonur Cosmodrome, Kazakhstan
4           SLC-41, Cape Canaveral AFS, Florida, USA

                          Date                                        Detail  \
0  Fri Aug 07, 2020 05:12 UTC  Falcon 9 Block 5 | Starlink V1 L9 & BlackSky
1  Thu Aug 06, 2020 04:01 UTC           Long March 2D | Gaofen-9 04 & Q-SAT
2  Tue Aug 04, 2020 23:57 UTC            Starship Prototype | 150 Meter Hop
3  Thu Jul 30, 2020 21:25 UTC  Proton-M/Briz-M | Ekspress-80 & Ekspress-103
4  Thu Jul 30, 2020 11:50 UTC                    Atlas V 541 | Perseverance

  Rocket_Status  Price Mission_Status
0  StatusActive  50.00        Success
1  StatusActive  29.75        Success
2  StatusActive    NaN        Success
3  StatusActive  65.00        Success
4  StatusActive 145.00        Success
```




```
     organisation                                    location  \
2556    RVSN USSR  Site 31/6, Baikonur Cosmodrome, Kazakhstan

                  date                                 detail  rocket_status  \
2556  Mon Oct 30, 1978  Molniya-M /Block SO-L | Prognoz n†7  StatusRetired

      price mission_status
2556    NaN        Success

     organisation                                   location  \
2773    RVSN USSR  Site 1/5, Baikonur Cosmodrome, Kazakhstan

                           date            detail  rocket_status  price  \
2773  Thu Oct 14, 1976 17:39 UTC  Soyuz | Soyuz 23  StatusRetired    NaN

     mission_status
2773        Success
```

- The `date` column has two formats:
  1. weekday month day of the month, year
  2. weekday month day of the month, year hour:minute UTC

```
Duplicates after removing redundant columns: 1
```

- One duplicate was found and will be removed.

```
------------------rocket_status------------------
StatusActive, StatusRetired
==================================================

-----------------mission_status-----------------
Failure, Partial Failure, Prelaunch Failure, Success
==================================================
```

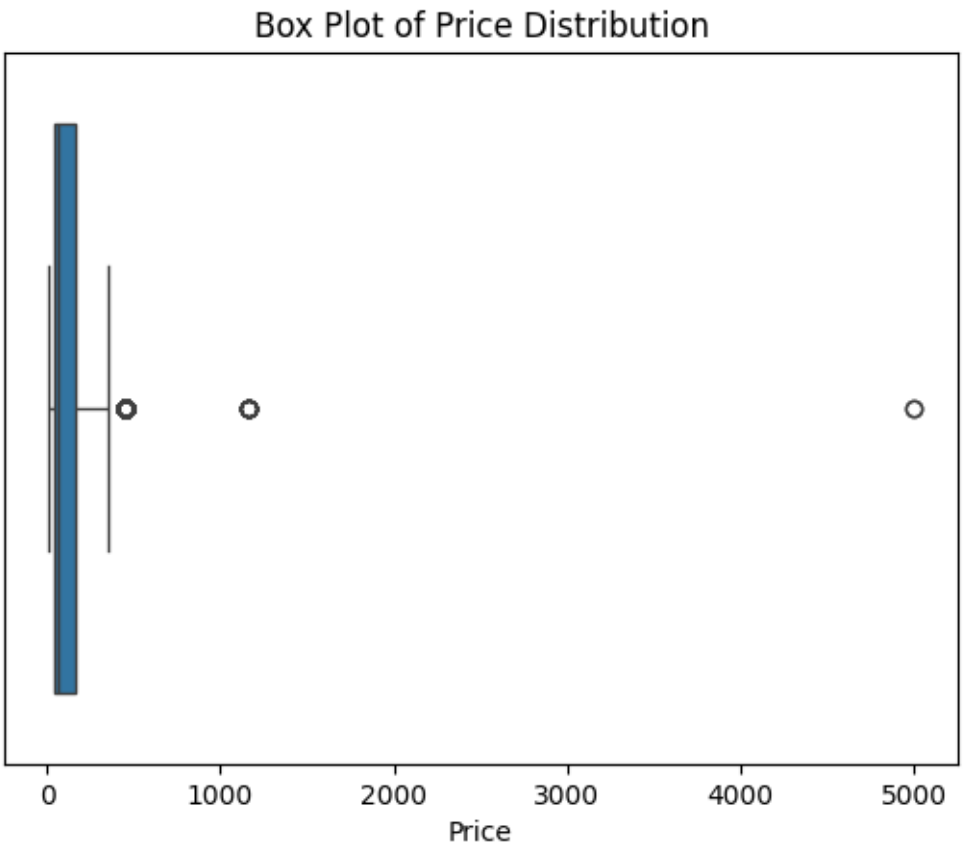
2.2 Descriptive Statistics

```
       count   mean    std  min   25%   50%    75%      max
price 963.00 153.92 288.57 5.30 40.00 62.00 164.00 5,000.00
```

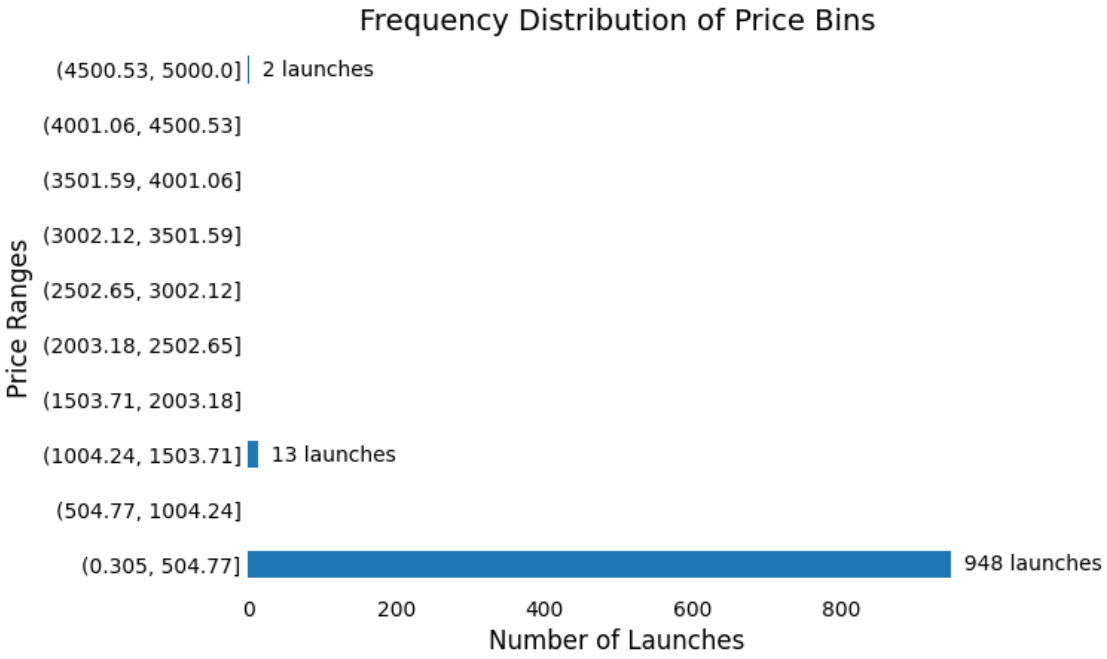
- The output indicates:
  1. Mean price: \$153.92 million
  2. Standard deviation: \$288.57 million
  3. Minimum and maximum prices show a wide range from \$5.3 to \$5,000 million
  4. There is significant variability in the price data.

```
       skew  kurtosis
price 10.56    165.56
```

- Interpretation of results:
  1. A skewness of 10.56 indicates a highly positive skew, meaning the distribution has a long right tail and most data points cluster on the left side.
  2. A kurtosis of 165.56 is extremely high, suggesting a very peaked distribution with heavy tails, indicating the presence of many outliers.
- These metrics suggest a distribution that is both asymmetrical and has extreme values, pointing to potential challenges in analysis and interpretation.

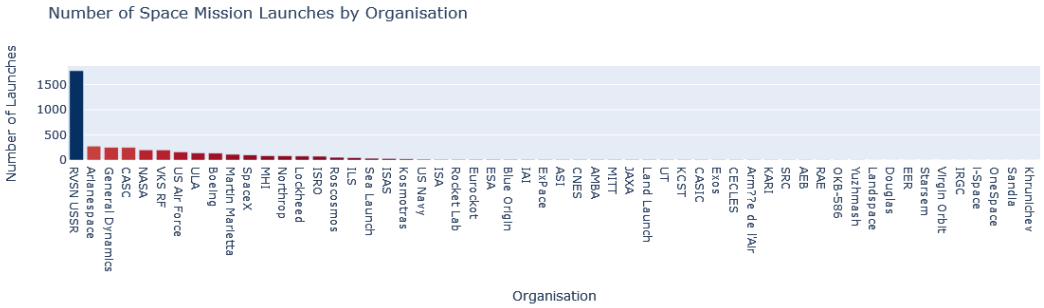


- The box plot indicates the presence of outliers in the price data.
- Let's check the distribution of the price using 10 bins.



3 Number of Launches per Company

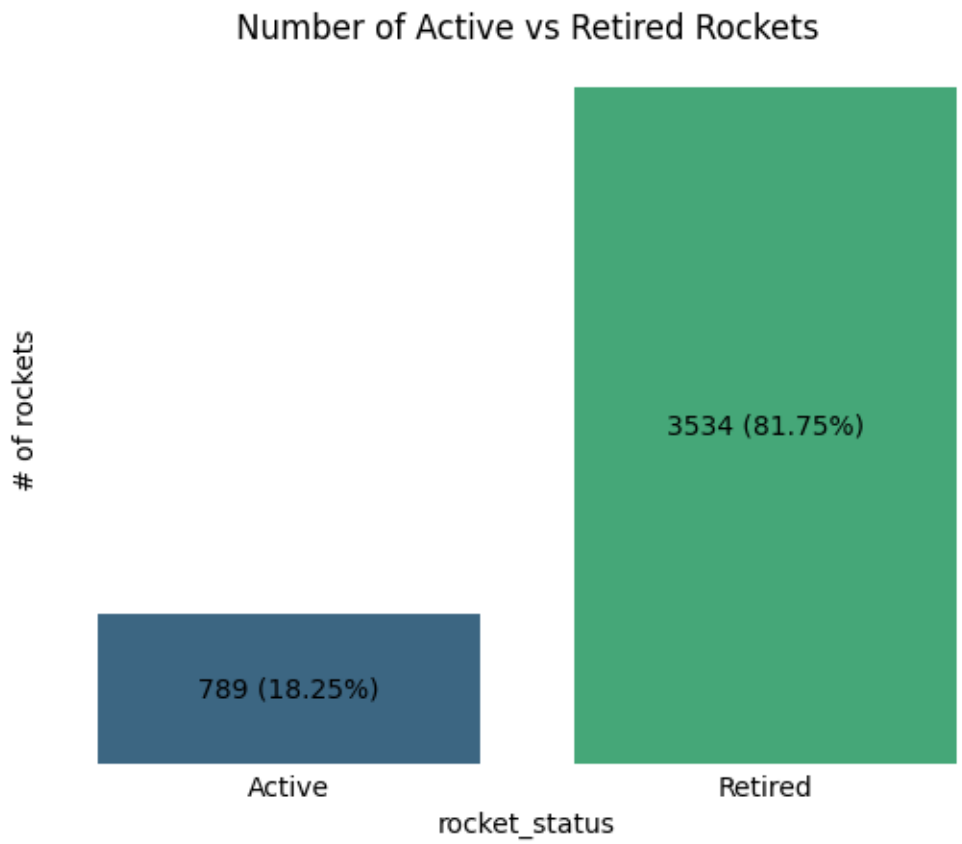
Create a chart that shows the number of space mission launches by organisation.



- The dataset shows that the Soviet Union has the most launches.
- Keep in mind that U.S. launch data cannot be solely obtained from NASA, as the U.S. has multiple organizations involved in space missions, unlike the Soviet Union, which had only one organization.

4 Number of Active versus Retired Rockets

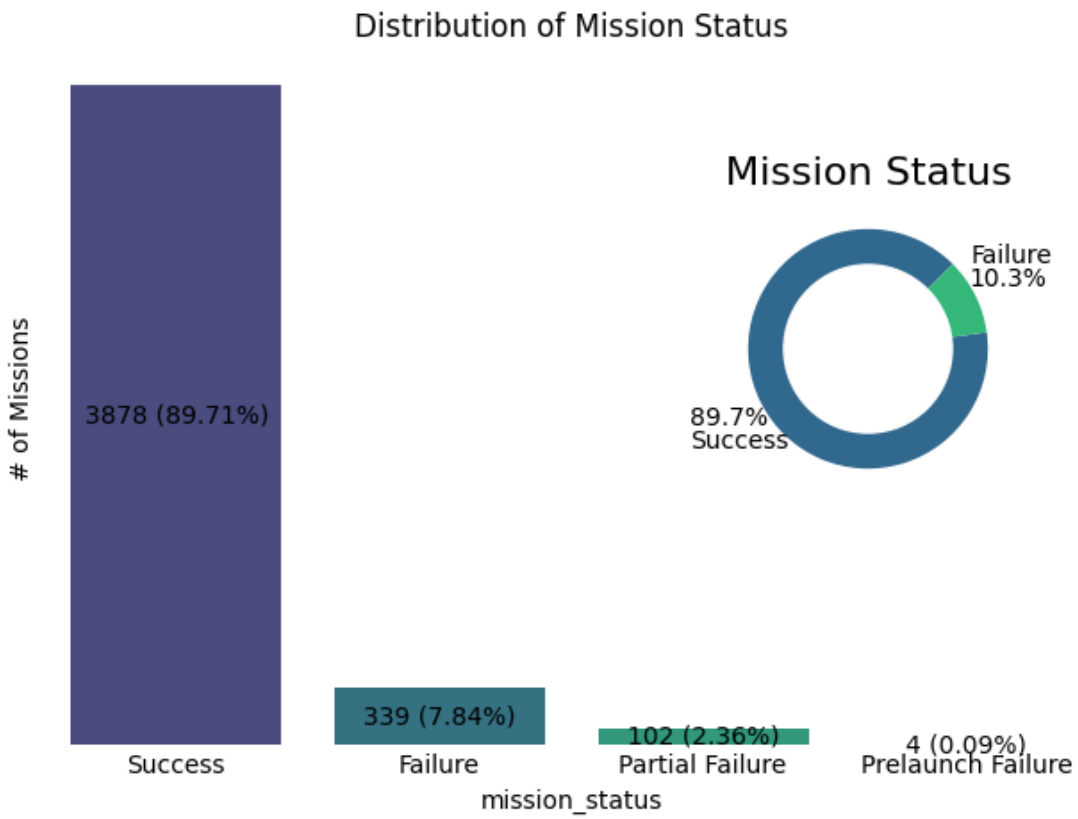
How many rockets are active compared to those that are decomissioned?



- Approximately **82%** of launched rockets have been retired.
- Most missions were retired by the time this data was collected in **2020**.

5 Distribution of Mission Status

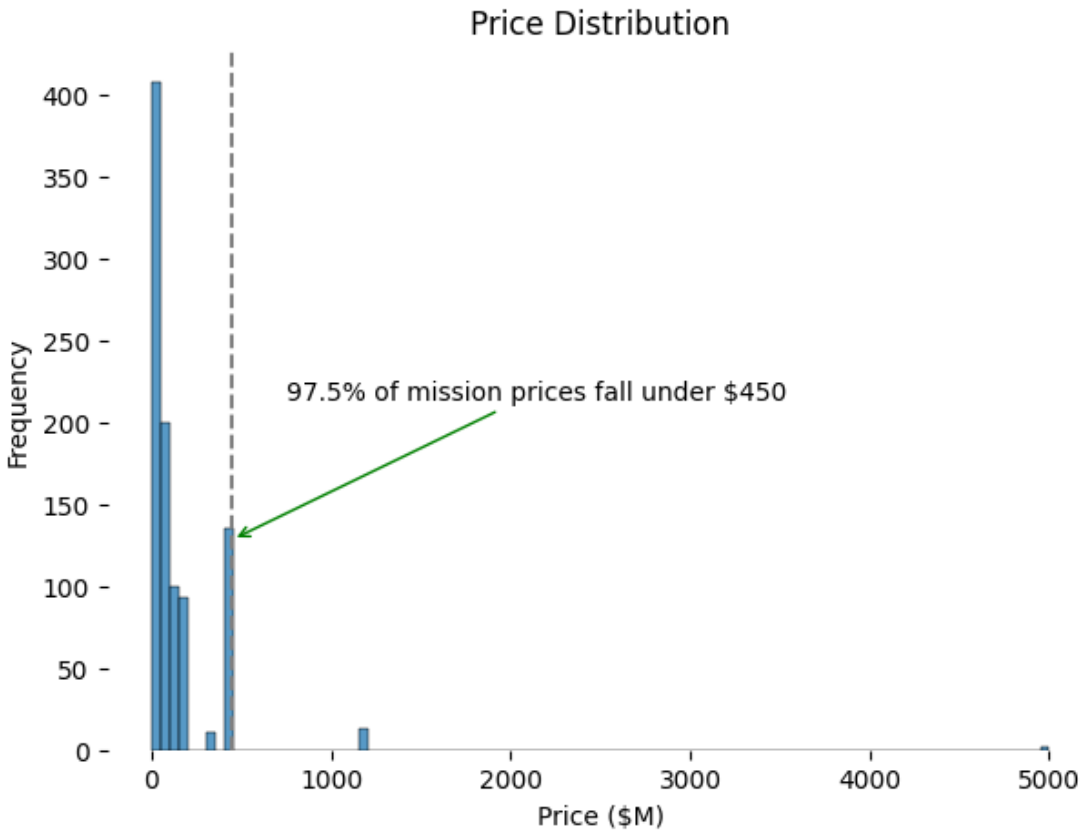
How many missions were successful? How many missions failed?



- The vast majority of missions were successful.
- Approximately **10%** of missions resulted in failure or partial failure.
- Only **4 missions** failed before launch.

6 How Expensive are the Launches?

Create a histogram and visualise the distribution. The price column is given in USD millions (careful of missing values).



- The price distribution ranges from approximately **\$5.3 million** to **\$5,000 million**. However, when narrowing the focus to include **97.5%** of prices, the maximum is limited to **\$450 million**.

7 Use a Choropleth Map to Show the Number of Launches by Country

- Create a choropleth map using the plotly documentation
- Experiment with plotly's available colours. I quite like the sequential colour `matter` on this map.
- You'll need to extract a `country` feature as well as change the country names that no longer exist.

Wrangle the Country Names

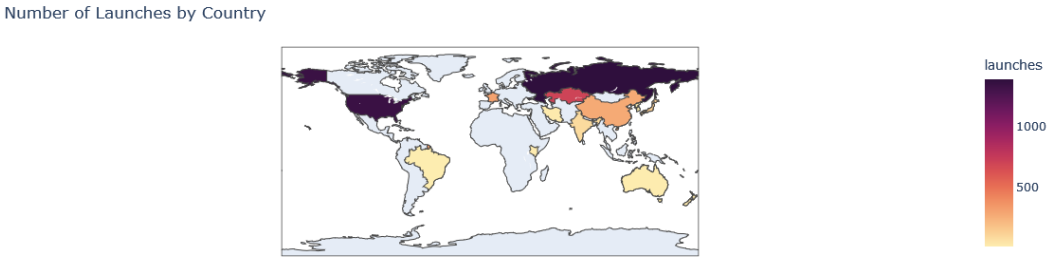
You'll need to use a 3 letter country code for each country. You might have to change some country names.

- Russia is the Russian Federation
- New Mexico should be USA
- Yellow Sea refers to China
- Shahrud Missile Test Site should be Iran

- Pacific Missile Range Facility should be USA
- Barents Sea should be Russian Federation
- Gran Canaria should be USA

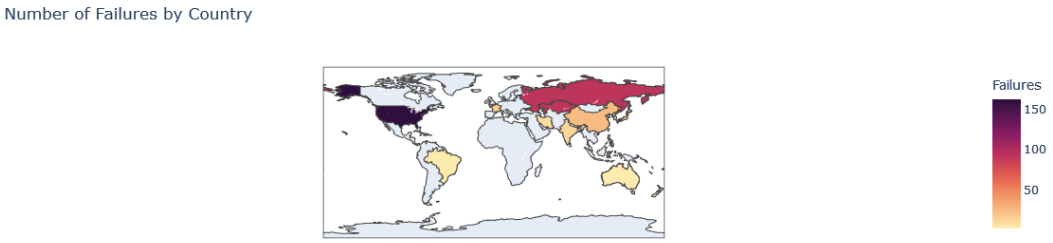
You can use the iso3166 package to convert the country names to Alpha3 format.

```
array(['USA', 'China', 'Kazakhstan', 'Japan', 'Israel', 'New Zealand',
       'Russian Federation', 'Iran, Islamic Republic of', 'France',
       'India', "Korea, Democratic People's Republic of", 'Kiribati',
       'Korea, Republic of', 'Brazil', 'Kenya', 'Australia'], dtype=object)
```



- The map indicates that the USA and the Russian Federation dominate the launch landscape, with 1,351 and 1,398 launches, respectively.

8 Use a Choropleth Map to Show the Number of Failures by Country

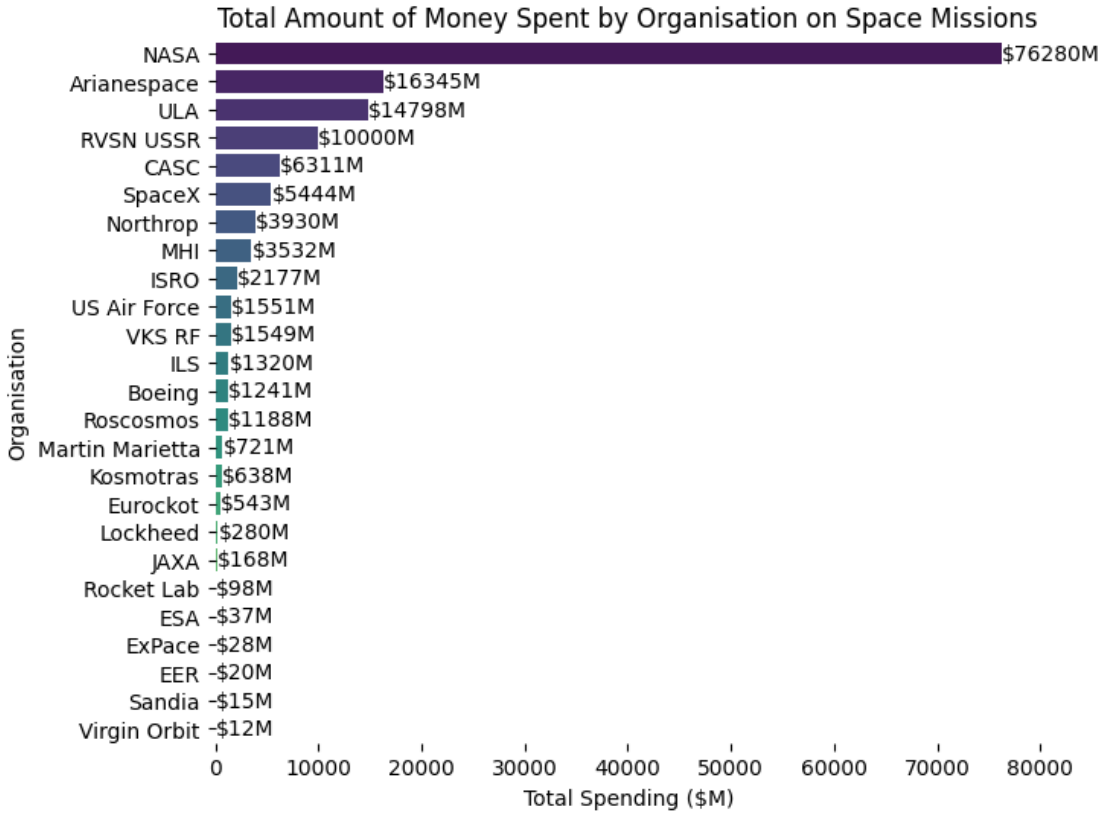


- The USA has the highest number of total failures, followed by the Russian Federation and Kazakhstan.
- This can be attributed to their large number of attempted launches.
- In contrast, the Republic of Korea and New Zealand have the lowest number of failures.

- When considering failure as a proportion of total launches, Brazil ranks first in total loss, followed by the Republic of Korea and the Islamic Republic of Iran.
- France has the lowest proportion of failures, followed by the Russian Federation.

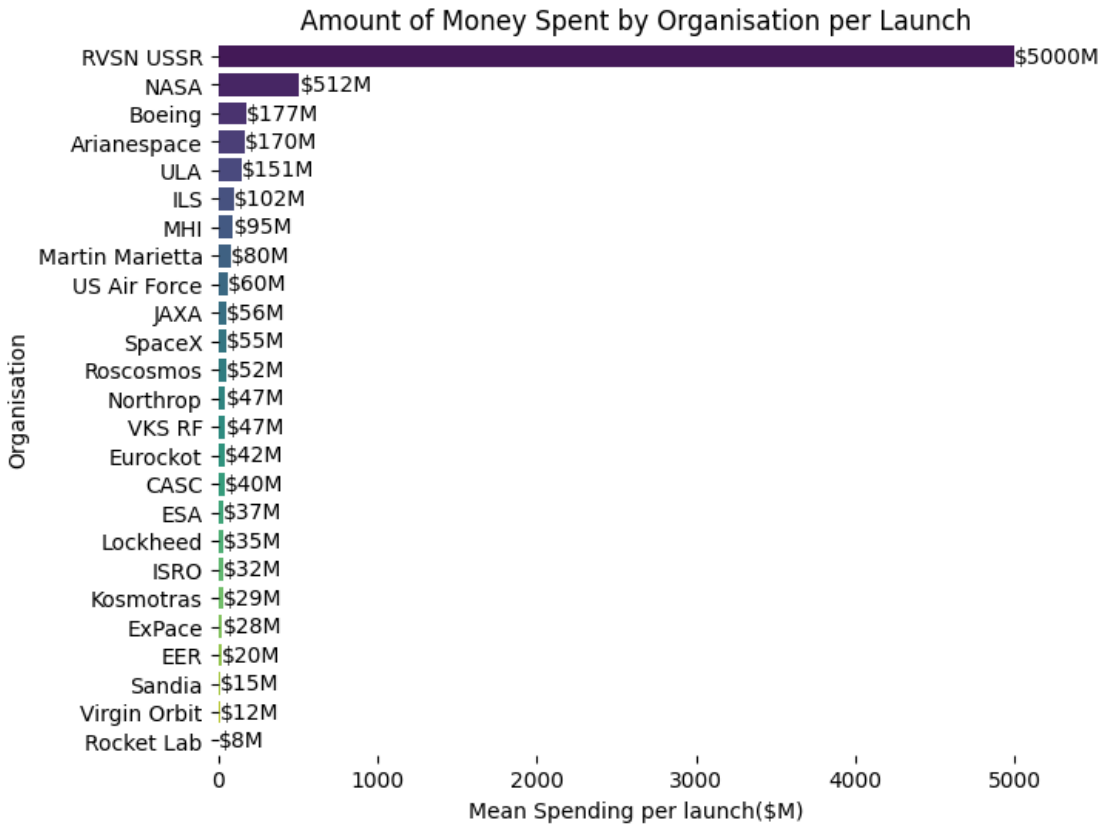
9 Create a Plotly Sunburst Chart of the countries, organisations, and mission status.

10 Analyse the Total Amount of Money Spent by Organisation on Space Missions



- NASA has spent \$76 billion, with expenditures surpassing the next organization by \$60 billion.

11 Analyse the Amount of Money Spent by Organisation per Launch

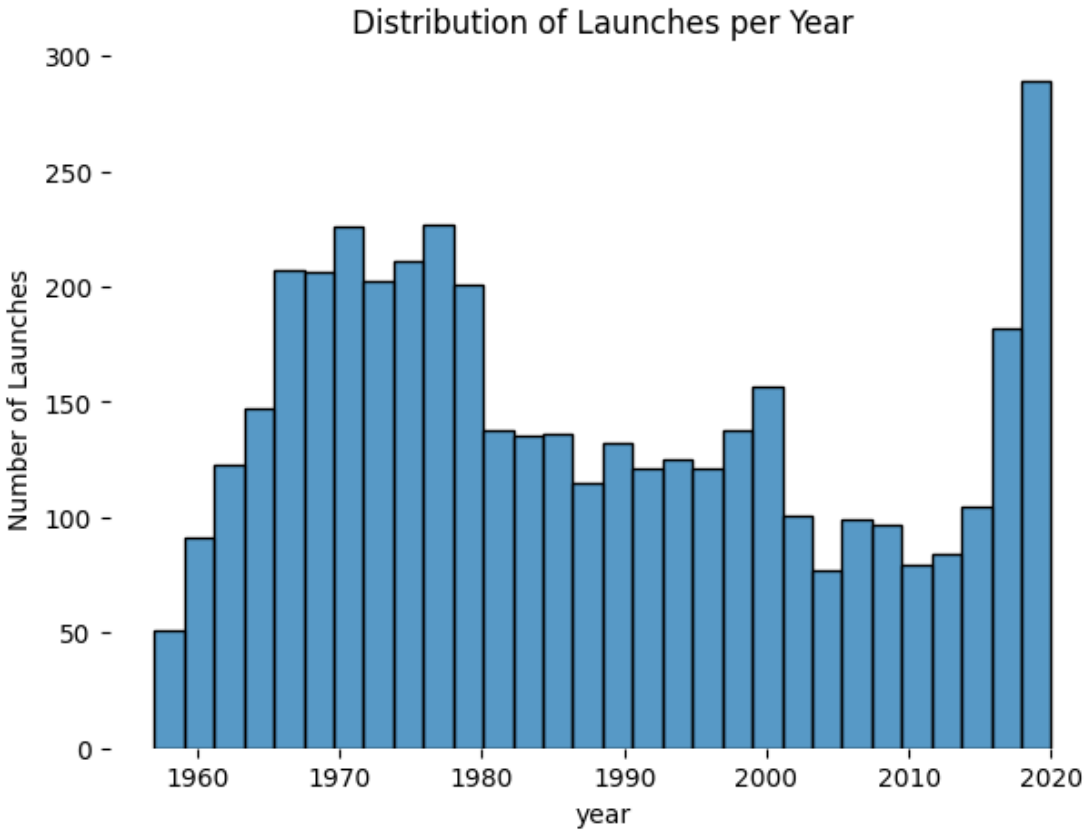


- The mean cost per launch for RVSN (USSR) is significantly higher than that of other organizations, primarily due to a large number of missing price values. The two available values, which are \$5,000 million each, heavily influence this average.

	year	organisation	detail	price
0	1988	RVSN USSR	Energiya/Buran | Buran	5,000.00
1	1987	RVSN USSR	Energiya/Polyus | Polyus Space Station	5,000.00

- NASA ranks second with an average spending of \$511 million.

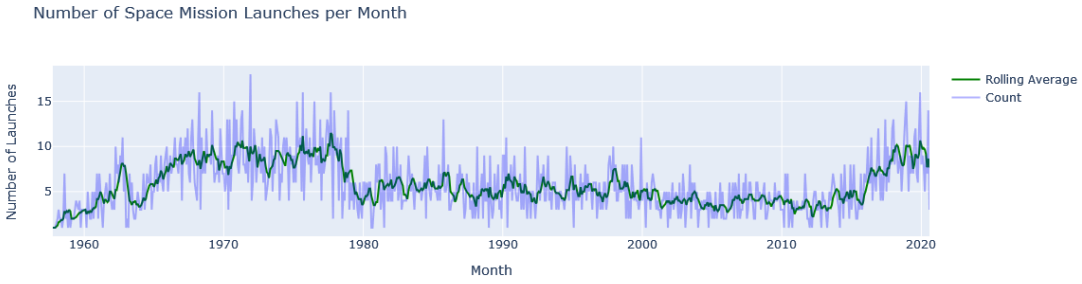
12 Chart the Number of Launches per Year



- There was a momentum of launches that began in the late 1960s, peaked in the 1970s, and then started to decline thereafter.
- However, there has been an increase in launches in recent years.

13 Chart the Number of Launches Month-on-Month until the Present

Which month has seen the highest number of launches in all time? Superimpose a rolling average on the month on month time series chart.

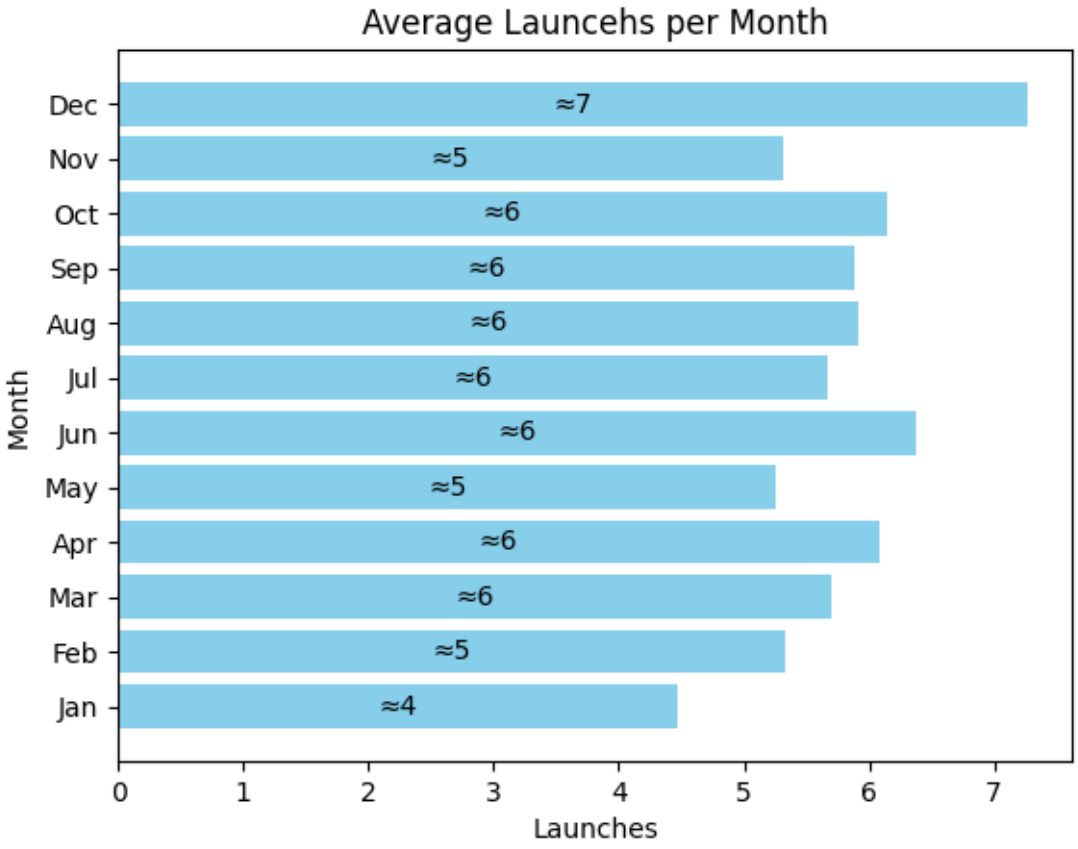



```
Month with the highest number of launches in all time:
    - 1971-12: 18 launches
```

- December 1971 recorded the highest number of launches, with a total of 18 launches.

14 Launches per Month: Which months are most popular and least popular for launches?

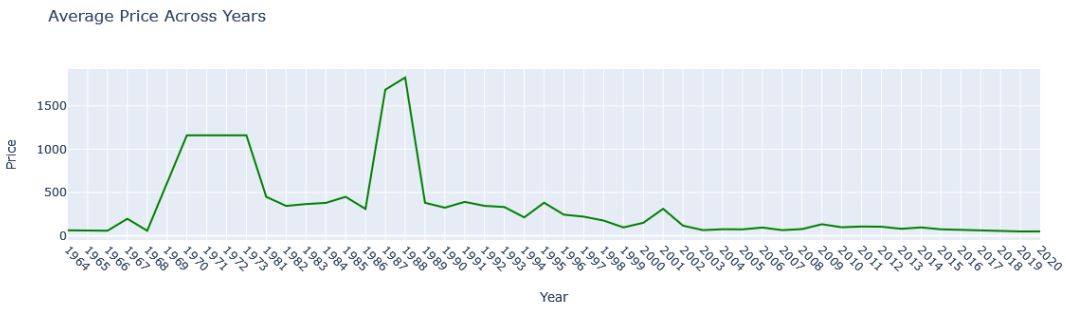
Some months have better weather than others. Which time of year seems to be best for space missions?



- December is the most popular month for space missions, while January is the least popular.
- Based on the average number of launched rockets per month, December is the best month for space missions, and January is the worst.

15 How has the Launch Price varied Over Time?

Create a line chart that shows the average price of rocket launches over time.



15.0.1 Analysis of Price Trends After Outlier Removal Within Each Year:

1. **Data Gaps (1957-1964 & 1973-1981):**
   Notable gaps in the data during these periods indicate a lack of recorded launch price information, which could affect trend analysis.

2. **Price Spike in 1987 and 1988:**
   A significant increase in launch prices during these years suggests an anomaly or a shift in market dynamics, possibly due to geopolitical or economic factors.

3. **Fluctuations and Declining Trend Post-1980s:**
   The data shows fluctuations in launch prices after the 1980s, alongside an overall declining trend, indicating potential instability in the market.

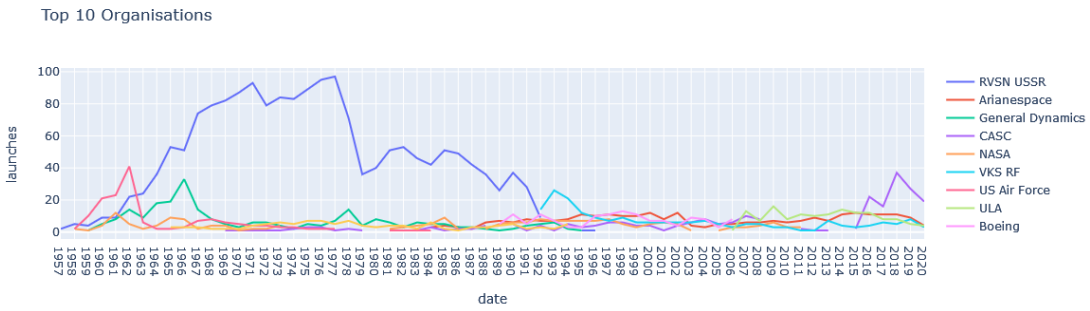
4. **Stable Decline (2010-2020):**
   The last decade exhibits a more consistent decline in prices, suggesting increased efficiency and competition in the launch market, leading to lower costs.

16 Chart the Number of Launches over Time by the Top 10 Organisations.

How has the dominance of launches changed over time between the different players?

```
Top ten organisations:
    01- RVSN USSR (1777)
    02- Arianespace (279)
```

```
03- General Dynamics (251)
04- CASC (250)
05- NASA (203)
06- VKS RF (201)
07- US Air Force (161)
08- ULA (140)
09- Boeing (136)
10- Martin Marietta (114)
```



1. **Dominance in Launch Counts by RVSN USSR (1965-1990):**
   - There was a strong dominance in launch counts from 1965 to 1990, with a notable decline in 1979 for the RVSN USSR organization.
2. **Early Dominance of the US Air Force:**
   The US Air Force was a key player in the early stages of space exploration, establishing its dominance during this formative period.
3. **CASC's Recent Leadership:**
   In recent years, particularly over the last five, the China Aerospace Science and Technology Corporation (CASC) has emerged as a leading organization in space exploration among the top ten entities.

17 Cold War Space Race: USA vs USSR

The cold war lasted from the start of the dataset up until 1991.

```
Countries:
    01- Australia
    02- Brazil
    03- China
    04- France
    05- India
    06- Islamic Republic of Iran
    07- Israel
    08- Japan
    09- Kazakhstan
```

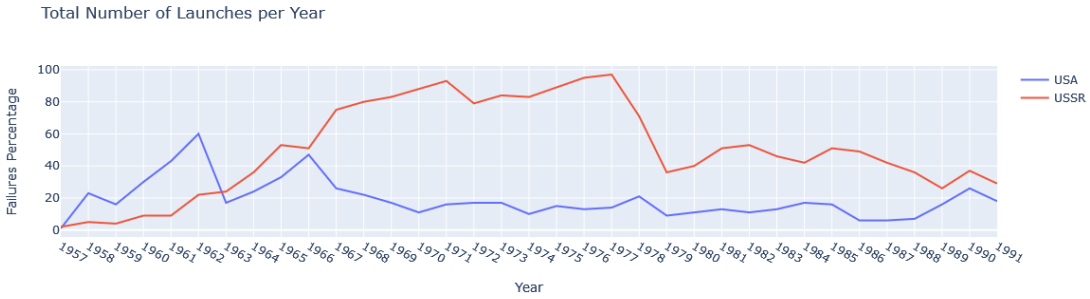
```
10- Kenya
11- Kiribati
12- Democratic People's Republic of Korea
13- Republic of Korea
14- New Zealand
15- Russian Federation
16- USA
```

- **Soviet Union During the Cold War:**
  - Both the Russian Federation and Kazakhstan were integral parts of the Soviet Union during the Cold War era.

17.1 Create a Plotly Pie Chart comparing the total number of launches of the USSR and the USA

- **Soviet Union vs. USA Launches:**
  During this era, the number of launches by the USSR was approximately three times that of the USA.

17.2 Create a Chart that Shows the Total Number of Launches Year-On-Year by the Two Superpowers

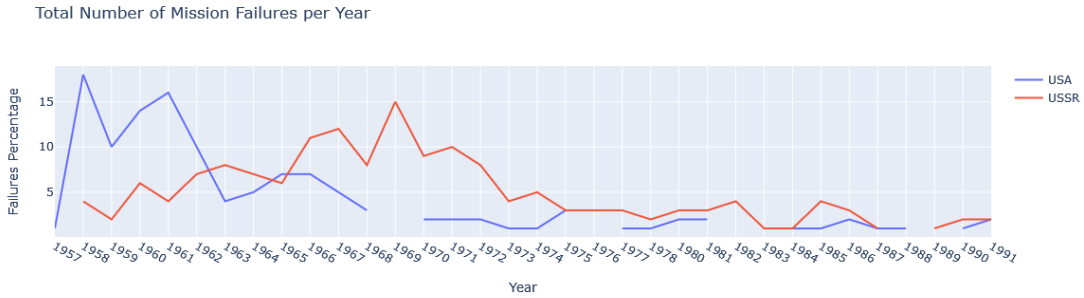


- The race peaked in 1966, with each country contributing around 50 launches. The USSR

then maintained a significant lead in launches until 1979, when the difference began to narrow.

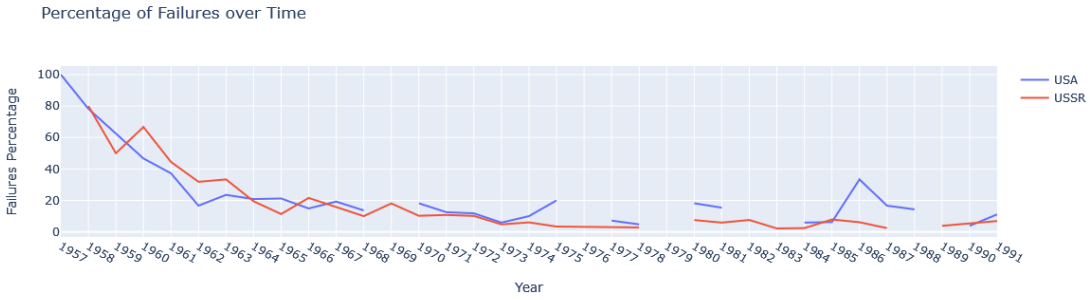
- From 1970, the USA demonstrated consistency in its launch numbers, reaching a peak in 1990.

17.3 Chart the Total Number of Mission Failures Year on Year.



17.4 Chart the Percentage of Failures over Time

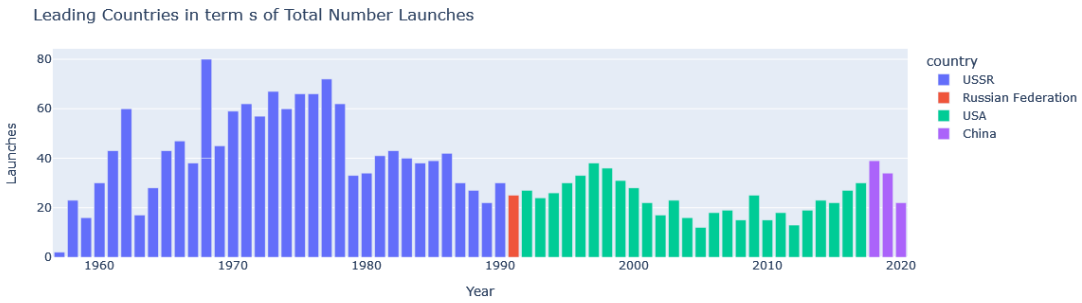
Did failures go up or down over time? Did the countries get better at minimising risk and improving their chances of success over time?

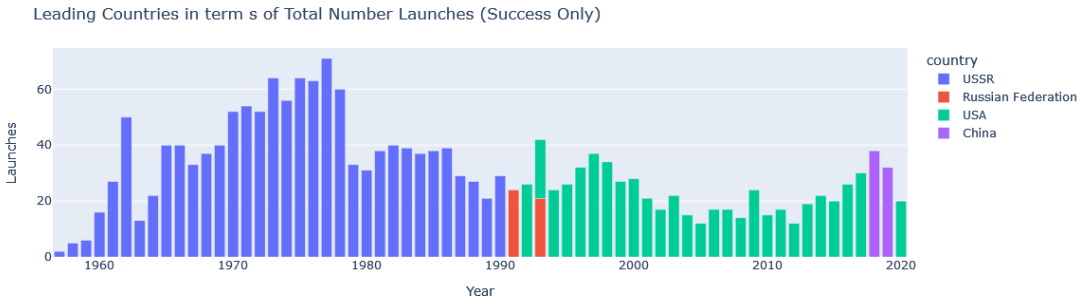


- The plot here shows that the percentage of failures decreased over time, indicating improvements in reliability and success rates in launches.

18 For Every Year Show which Country was in the Lead in terms of Total Number of Launches up to and including including 2020)

Do the results change if we only look at the number of successful launches?

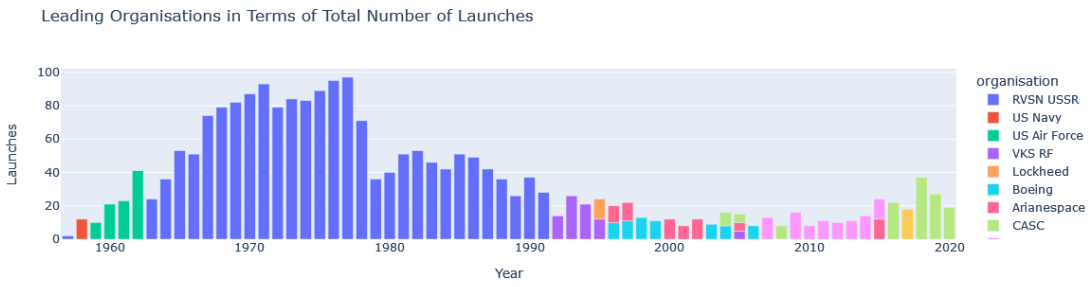




- During the Cold War, the Soviet Union dominated in total launches. After this period, the USA took the lead until 2017, when China began to emerge as the leading country in total launches.
- If we consider only successful launches, China led in 2018 and 2019, while the USA led in 1992. Additionally, there was a tie between the Russian Federation and the USA in 1993. In all other years, the dominance remained consistent with the previous trends.

19 Create a Year-on-Year Chart Showing the Organisation Doing the Most Number of Launches

Which organisation was dominant in the 1970s and 1980s? Which organisation was dominant in 2018, 2019 and 2020?



- The RVSN (Rocket Forces of the Strategic Missile Troops) of the USSR dominated during the 1970s and 1980s. In contrast, the China National Space Administration (CASC) emerged as a dominant force from 2018 to 2020.