SPACE LAUNCHES EXPLORATION VISUALIZATION

- VARUN G

ABSTRACT

Space is one of the most fascinating and a field where technology is under constant development, to reduce the cost, increase the efficiency, range, and life expectancy. The dataset that we have collected has the main information of all the launches since the dawn of the space programmes all over the world (1957). This also has the When? where? And what? Of the rockets and the satellites launched.

KEYWORD

Space Launch, Rocket name, Time series visualization, Technological Development, Rockets used by the Organization, Rocket status, Rocket Mission, Location, Organization.

1. INTRODUCTION

We are a group of Three. We wanted to know how space technology influence our human life and development of technology in space exploration. We wanted to clearly visualize the space exploration which has been happened by time by time.

Spatial visualization leads to greater organizational efficiency by accelerating all phases of realworld projects that depend on design work for their success.

Spatial Visualization gives us a clear understanding about the space exploration in the human civilization. It helps in providing a clear view about the prediction model about the upcoming Space mission

2. Literature Survey

Human exploration of space has already a history of longer than 50 years (starting by the first flight of Yuri Gagarin on 12 April 1961). Since then, many short- and long-duration missions have taken place on a variety of platforms. We took a data where we can infer about the space exploration by human from 1957-2020 using different Visualization techniques.

Space exploration is the ongoing discovery and exploration of celestial structures in outer space by means of continuously evolving and growing space technology. While the study of space is carried out mainly by astronomers with telescopes, the physical exploration of space is conducted both by unmanned robotic probes and human spaceflight.

3. About The Dataset

The Dataset which we have used for this visualization is Space Exploration done from the year 1951-2020.**Unnamed:** the s.no for the file ◦ **Company** Name: name of the company that launched the rocket.

- **Location**: the location from where the rocket was launched.
- **Datum**: the date and time of the launch.
- **Detail**: the name of the rocket and the satellite it carried.
- **Status** Rocket: the status of the rocket (active/retired).
- **Rocket**: the cost of launch.
- **Status Mission**: the status of the satellite that was launched.
- **Reason**: the main reasons of the launch of satellite (earth observation, solar physics, microgravity)

4.1 Feature components for analysis

The technological development in the space department has been the main factor or the development and the scaling of other industries more importantly entertainment, communication, and information industries. The satellites are also the main factor to determine natural disaster and catastrophes early and make early measures for prevention and aftermath of the event.

To find the lead runner in the space department and to find the success rate of each organization, rocket, and to find the technological development of the rockets used by the organisations/country.

1. Design and flow of models

df.info()

RangeIndex: 4324 entries, 0 to 4323

Data columns (total 9 columns):

#	Column	Non-Null Count	Dtype
---	------	--------------	-----
0	Unnamed: 0	4324 non-null	int64
1	Unnamed: 0.1	4324 non-null	int64
2	Company Name	4324 non-null	object

3	Location	4324 non-null	object
4	Datum	4324 non-null	object
5	Detail	4324 non-null	object
6	Status Rocket	4324 non-null	object
7	Rocket	964 non-null	object
8	Status Mission	4324 non-null	object

1 design and flow of model for the analysis we have used the following

modules and analysis attributes:

5.1. module 1: data collection

The data that we are using is collected from Kaggle, this data has all the main and the meta data about all the launches that were done from 1957 to 2020 this data has also the satellite name, rocket name, cost of the mission, date of launch the satellites launched. The data we got was a crude data that had a lot of Nan values in the rocket column and needed processing.

5.2. module 2: data cleaning and dataset analysis

Once now we have the crude data, we can use the data cleaning and processing packages like pandas to clean the data either by removing the Nan values or by using methods like multiple imputation, maximum likelihood to fill the Nan values.

5.3. module 3: analysis and visualization from dataset

The attributes from the obtained data set are compared with each other to find correlations and dependencies and then these are visualized using different types of graphs. The graphs obtained are then studied in detail to obtain further observations to see to what factor each attribute influences the recommendation system and which attributes are dependent and independent.

a. Understanding the distributions of the dimensions:

The data we got is not numerical but categorical so it is important to plot graphs such as bar graphs and pie plots to understand the distribution of the categories in the data.

b. Based on time series:

The data that we have collected is a chronological dataset with all the data in a chronological order and it speaks about the history of the space exploration of all the pioneers and the new companies. And their accomplishments and loses.

7. IMPLEMENTATION

7.1 First we import modules that can help us clean and preprocess the data ready for modelling.

```
In [1]: import numpy as np
        import pandas as pd

        import os
        for dirname, _, filenames in os.walk('/kaggle/input'):
            for filename in filenames:
                print(os.path.join(dirname, filename))
```

```
In [2]: import matplotlib.pyplot as plt
        from plotly.offline import init_notebook_mode, iplot, plot
        import plotly as py
        init_notebook_mode(connected=True)
        import plotly.graph_objs as go
        import plotly_express as px
        import cufflinks as cf
        cf.go_offline()
        cf.set_config_file(offline=False, world_readable=True)
```

```
In [3]: import warnings
        warnings.filterwarnings("ignore")
```

Importing all the libraries and modules

Here NumPy and Pandas are basically used for data preprocessing, manipulation and cleaning. Plotly, matplotlib, cufflinks are used for data visualization and plotting.

```
In [4]: spacemission_data = pd.read_csv("C:/Users/rupac/Documents/SEM_4/Information visualisation/Space_Corrected.csv")
        spacemission_data.head()
```

Out[4]:

	Unnamed: 0	Unnamed: 0.1	Company Name	Location	Datum	Detail	Status Rocket	Rocket	Status Mission
0	0	0	SpaceX	LC-39A, Kennedy Space Center, Florida, USA	Fri Aug 07, 2020 05:12 UTC	Falcon 9 Block 5 | Starlink V1 L9 & BlackSky	StatusActive	50.0	Success
1	1	1	CASC	Site 9401 (SLS-2), Jiuquan Satellite Launch Ce...	Thu Aug 06, 2020 04:01 UTC	Long March 2D | Gaofen-9 04 & Q-SAT	StatusActive	29.75	Success
2	2	2	SpaceX	Pad A, Boca Chica, Texas, USA	Tue Aug 04, 2020 23:57 UTC	Starship Prototype | 150 Meter Hop	StatusActive	NaN	Success
3	3	3	Roscosmos	Site 200/39, Baikonur Cosmodrome, Kazakhstan	Thu Jul 30, 2020 21:25 UTC	Proton-M/Briz-M | Ekspress-80 & Ekspress-103	StatusActive	65.0	Success
4	4	4	ULA	SLC-41, Cape Canaveral AFS, Florida, USA	Thu Jul 30, 2020 11:50 UTC	Atlas V 541 | Perseverance	StatusActive	145.0	Success

In [14]:
```
df.describe()
```

Out[14]:

	Company Name	Location	Datum	Detail	Status Rocket	Rocket	Status Mission
count	4324	4324	4324	4324	4324	964	4324
unique	56	137	4319	4278	2	56	4
top	RVSN USSR	Site 31/6, Baikonur Cosmodrome, Kazakhstan	Tue Jun 26, 1973	Cosmos-3MRB (65MRB) | BOR-5 Shuttle	StatusRetired	450.0	Success
freq	1777	235	2	6	3534	136	3879

2.Dataset analysis

MISSING VALUES

In [5]:
```
spacemission_data = spacemission_data.drop(columns=['Unnamed: 0', 'Unnamed: 0.1'])
```

In [6]:
```
spacemission_data.columns = ['Company Name', 'Location', 'Datum', 'Detail', 'Status Rocket', 'Rocket', 'Status Mission']
```

In [7]:
```
spacemission_data['DateTime'] = pd.to_datetime(spacemission_data['Datum'])

# Extract the launch year
spacemission_data['Year'] = spacemission_data['DateTime'].apply(lambda datetime: datetime.year)

# Extract the country of launch
spacemission_data["Country"] = spacemission_data["Location"].apply(lambda location: location.split(", ")[-1])
```

In [8]:
```
# check missing data

total_of_all = spacemission_data.isnull().sum().sort_values(ascending=False)
percent_of_all = (spacemission_data.isnull().sum()/spacemission_data.isnull().count()).sort_values(ascending=False)
missing_missiondata_test = pd.concat([total_of_all, percent_of_all], axis=1, keys=['Total', 'Percent'])
missing_missiondata_test.head()
```

Out[8]:

	Total	Percent
Rocket	3360	0.777058
Company Name	0	0.000000
Location	0	0.000000
Datum	0	0.000000
Detail	0	0.000000

In [9]:
```
# handle missing values

spacemission_data['Rocket'] = spacemission_data['Rocket'].fillna(0)
spacemission_data.head()
```

Out[9]:

	Company Name	Location	Datum	Detail	Status Rocket	Rocket	Status Mission	DateTime	Year	Country
0	SpaceX	LC-39A, Kennedy Space Center, Florida, USA	Fri Aug 07, 2020 05:12 UTC	Falcon 9 Block 5 | Starlink V1 L9 & BlackSky	StatusActive	50.0	Success	2020-08-07 05:12:00+00:00	2020	USA
1	CASC	Site 9401 (SLS-2), Jiuquan Satellite Launch Ce...	Thu Aug 06, 2020 04:01 UTC	Long March 2D | Gaofen-9 04 & Q-SAT	StatusActive	29.75	Success	2020-08-06 04:01:00+00:00	2020	China
2	SpaceX	Pad A, Boca Chica, Texas, USA	Tue Aug 04, 2020 23:57 UTC	Starship Prototype | 150 Meter Hop	StatusActive	0	Success	2020-08-04 23:57:00+00:00	2020	USA
3	Roscosmos	Site 200/39, Baikonur Cosmodrome, Kazakhstan	Thu Jul 30, 2020 21:25 UTC	Proton-M/Briz-M | Ekspress-80 & Ekspress-103	StatusActive	65.0	Success	2020-07-30 21:25:00+00:00	2020	Kazakhstan
4	ULA	SLC-41, Cape Canaveral AFS, Florida, USA	Thu Jul 30, 2020 11:50 UTC	Atlas V 541 | Perseverance	StatusActive	145.0	Success	2020-07-30 11:50:00+00:00	2020	USA

ROCKET STATUS:

We also imputing missing Values of the dataset to include the fields we need for our analysis. Then analyze the correlation of each column or parameter in our dataset.

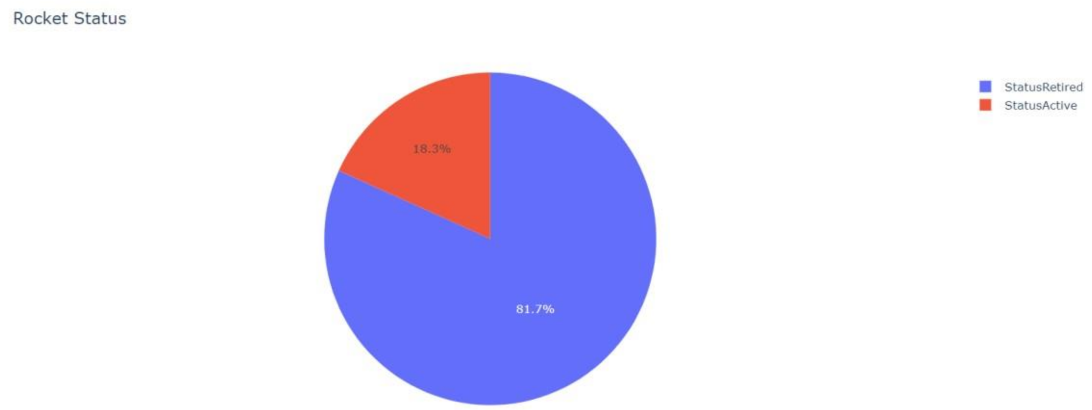
In [16]:
```
ds = df["Status Rocket"].value_counts().reset_index()
ds
```

Out[16]:

	index	Status Rocket
0	StatusRetired	3534
1	StatusActive	790

This table explains the niuber of rocket models that have retired and the rocket models that are active and are being used now.

```
In [17]: fig = px.pie(ds, values = "Status Rocket", names = "index", title = "Rocket Status")
         fig.show()
```



STATUS MISSION:

we have found that 81.7% of the rocket models have retired and only 18.3% are being used now. This is because of the increasing cost of making the rocket. there is a continuous need to create a model that is more fuel efficient and cost efficient.

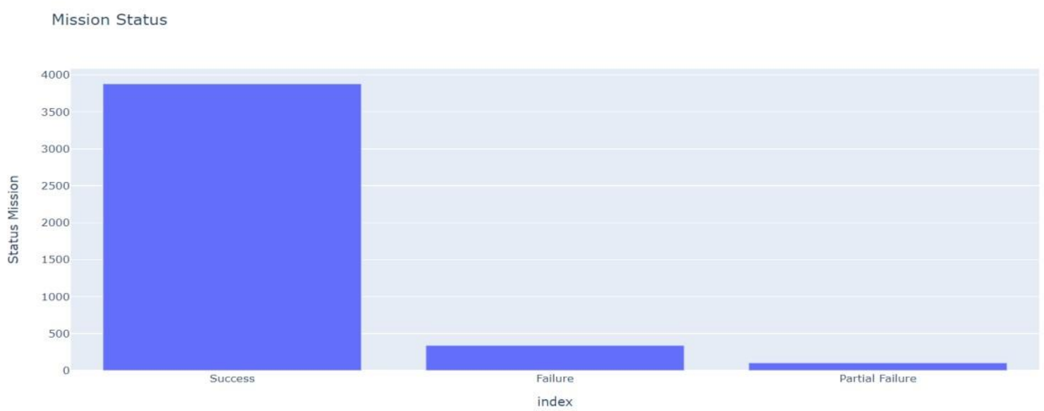
```
In [18]: ds = df["Status Mission"].value_counts().reset_index()[:3]
         ds
```

Out[18]:

	index	Status Mission
0	Success	3879
1	Failure	339
2	Partial Failure	102

this table shows the calculations of the total of success and failure and the partial failures of the missions that has happened till 2020. this is used for further visualizations.

```
In [19]: fig = px.bar(ds, x = "index", y = "Status Mission", title = "Mission Status")
         fig.show()
```



TOTAL MONEY SPENT:

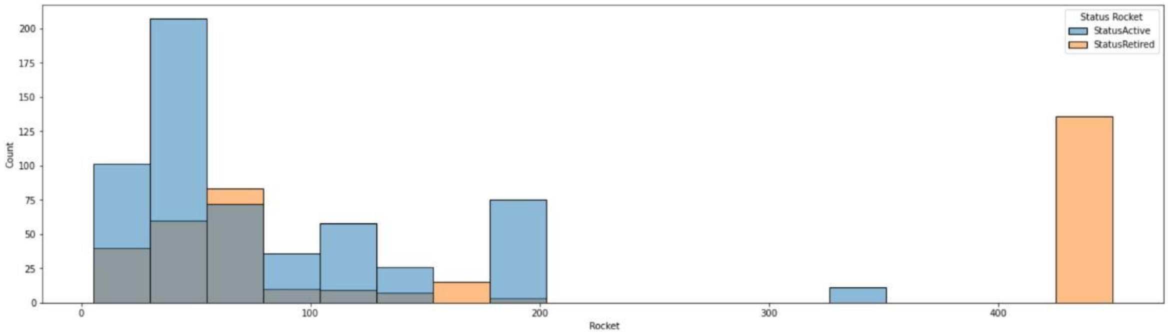
this code handles the NA values of the rocket column for further visualizations.

Rocket cost distribution with rocket status

```
In [20]: sum(pd.isna(df.loc[:, " Rocket"]))

Out[20]: 3360

In [21]: df_ = df.dropna(subset = [" Rocket"], axis = "rows")
         len(df_)

Out[21]: 964

In [22]: sum(pd.isna(df_.loc[:, " Rocket"]))

Out[22]: 0
```

This shows the total amount of money spent by organizations on the development and the total build cost of the rocket.

```
In [25]: df_d = df_[df_.loc[:, " Rocket"] < 1000]
         plt.figure(figsize = (22, 6))
         sns.histplot(data = df_d, x = " Rocket", hue = "Status Rocket")
         plt.show()
```



Total Money spent for each company

TOTAL MONEY SPENT BY COMPANIES:

Now as we have seen the distributions now let's see the money spent by each individual organization.

```
In [27]: df_money = df_.groupby(["Company Name"])[" Rocket"].sum().reset_index()
         df_money = df_money[df_money[" Rocket"] > 0]
         df_money.head()
```

Out[27]:

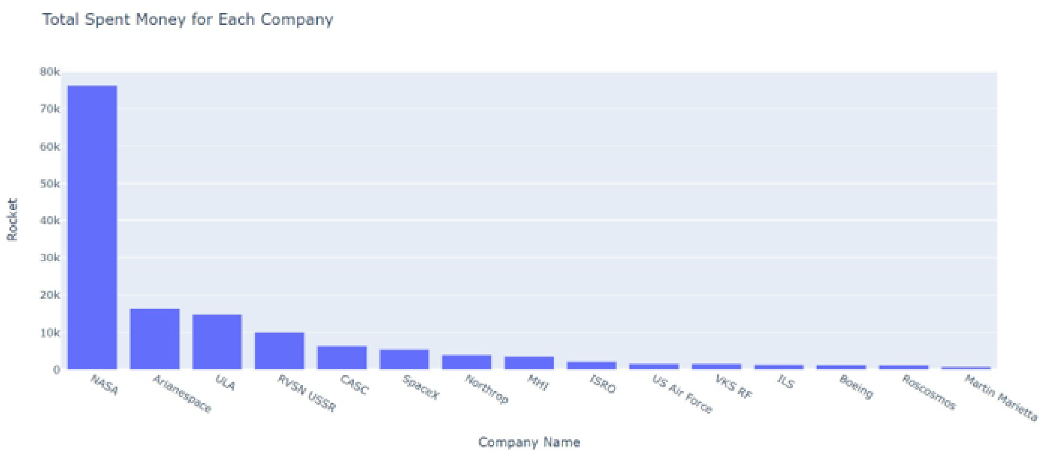
	Company Name	Rocket
0	Arianespace	16345.00
1	Boeing	1241.00
2	CASC	6340.26
3	EER	20.00
4	ESA	37.00

```
In [28]: df_money_ = df_money.sort_values(by = [" Rocket"], ascending = False)[:15]
         df_money_.head()
```

Out[28]:

	Company Name	Rocket
14	NASA	76280.00
0	Arianespace	16345.00
21	ULA	14798.00
16	RVSN USSR	10000.00
2	CASC	6340.26

```
In [29]: fig = px.bar(df_money_, x = "Company Name", y = " Rocket", title = "Total Spent Money for Each Company")
         fig.show()
```

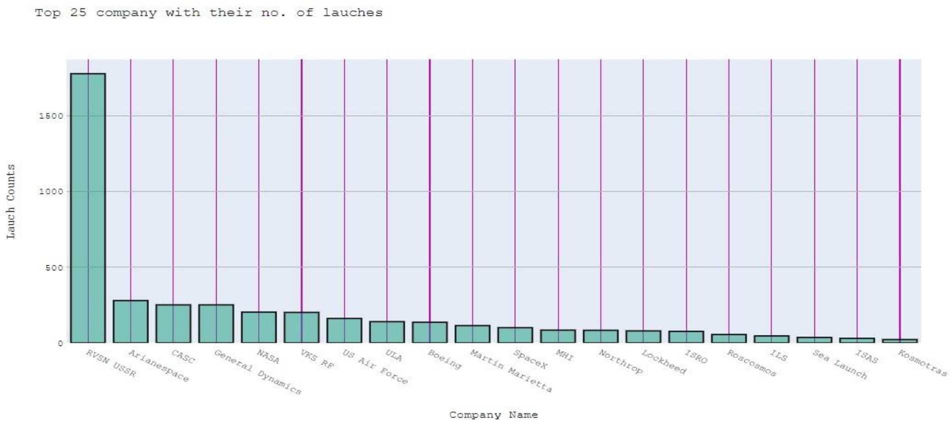


by inspection in the bar chart, we can say that Nasa has spent a lot in the development of the rockets, but also, we can see in the companies launches graph that Nasa doesn't come first. This is because that NASA develops the rockets and gives the launching in tender forms to other private companies.

LAUNCHES PER COMPANY BAR CHART:

We wanted to know which company has the largest share in the space launches and the exploration.

```
In [10]: # lauches per company bar chart

         company_launch_analysis = pd.DataFrame(spacemission_data['Company Name'].value_counts().sort_values(ascending=False))
         company_launch_analysis = company_launch_analysis.rename(columns={'Company Name':'Count'})

         trace = go.Bar(x = company_launch_analysis.index[:20],
                       y = company_launch_analysis['Count'][:20],
                       marker = dict(color='rgba(25,155,128,0.5)',
                       line = dict(color='rgb(0,0,0)', width=1.5)))

         layout = go.Layout(title="Top 25 company with their no. of lauches",
                           xaxis=dict(title='Company Name',zeroline= False,
                                  gridcolor='rgb(183,14,183)',showline=True),
                           yaxis=dict(title='Lauch Counts',zeroline= False,
                                     gridcolor='rgb(183,183,183)',showline=True),
                           font=dict(family='Courier New, monospace', size=10, color='rgb(1,9,9)')
         )
         data = [trace]
         fig = go.Figure(data = data, layout = layout)
         iplot(fig)
```

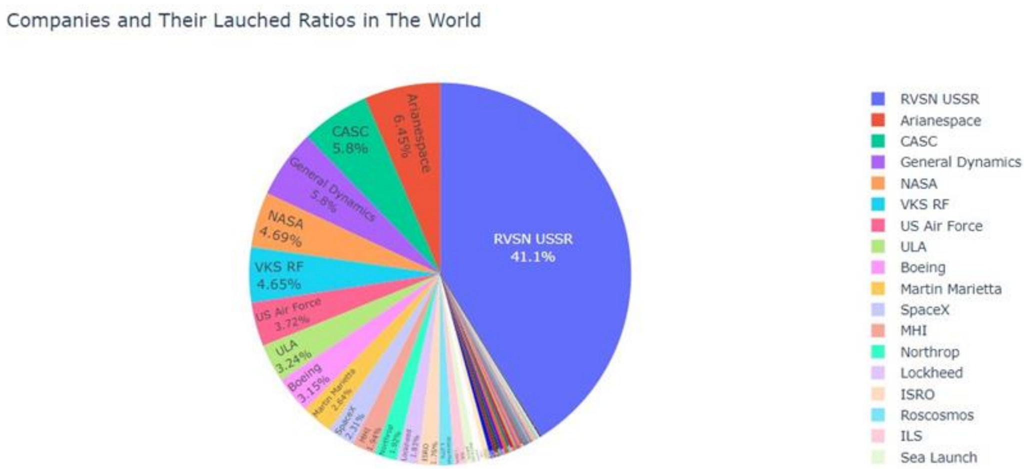


We have found that RSVN USSR has launched the largest number of the rockets in the total number of the rockets launched and next to this stands the commercial rocket launcher form the France "Ariane space" followed by the Chinese space agency CASC.

PIE CHART FOR COMPANIES AND THEIR LAUNCHES:

In this visualization we can infer about the information about the Companies which has participated in the exploration of space missions. So, Here RVSN has Contributed more to the exploration of space. We can also infer the ratio about different companies that took part in the exploration of space from 1957-2020.

```
In [11]: root = px.pie(company_launch_analysis, values=company_launch_analysis['Count'], names=company_launch_analysis.index,
                       title='Companies and Their Lauched Ratios in The World',
                      )
         root.update_traces(textposition='inside', textinfo='percent+label')
         root.update_layout(
             template='plotly_white'
         )
         root.show()
```



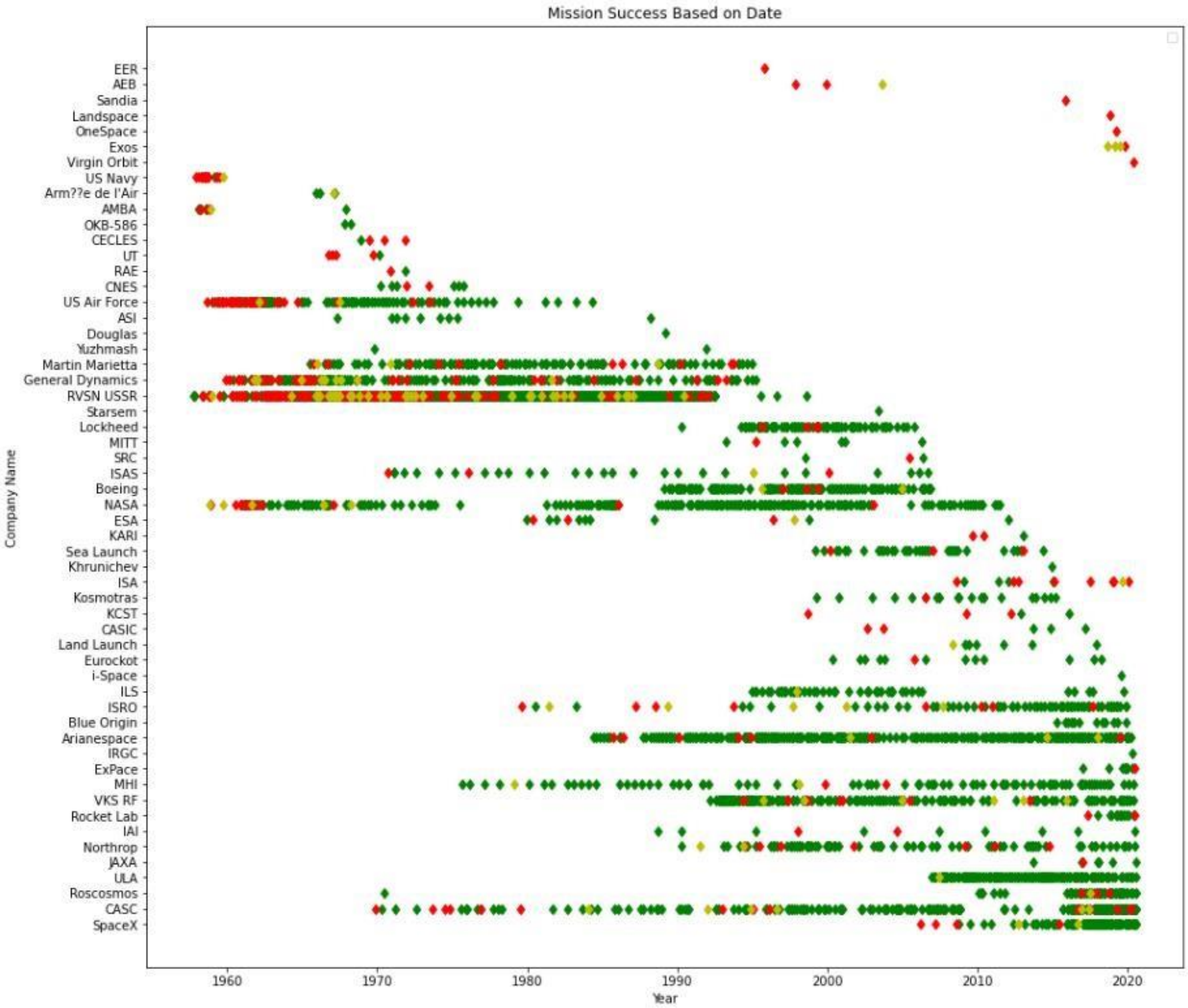
In [4]:
```
date=[]
time=[]
year=[]
for i in range(0,Space_data.shape[0]):
    str=Space_data['Datum'].iloc[i]
    day=str[:3]
    mm=str[4:7]
    dd=str[8:10]
    aaaa=str[12:16]
    hh_mm=str[17:22]
    str = dd+'/'+mm+'/'+aaaa
    date_object = datetime.strptime(str, '%d/%b/%Y')
    date.append(date_object)
    year.append(aaaa)


Space_data['Date'] = date
Space_data['Year'] = year

# Convert to numeric (the column was a string)
Space_data['Year']=pd.to_numeric(Space_data.Year)
```

In [5]:
```
status_mapping={'StatusRetired':0,'StatusActive':1}

Space_data['Status Rocket']=Space_data['Status Rocket'].map(status_mapping)
```

In [6]:
```
# There are spaces after the data
Space_data['Rocket']=Space_data[' Rocket'].str.strip()

# The thousands mark is a comma
Space_data['Rocket'] = Space_data['Rocket'].str.replace(',','')

# There are nans, so I fill with 0.
Space_data['Rocket'] = Space_data['Rocket'].fillna(0)

# Convert to numeric (the column was a string)
Space_data['Rocket']=pd.to_numeric(Space_data.Rocket)
```

CHRONOLOGICAL PLOT OF ALL THE LAUNCHES:

This plot shows the missions attempted by the company as a dot in the plot if it is a success then it is a green, red denotes failure, yellow is prelaunch and partial failures. This is a time series.



In this graph as showed from ILS to US navy from bottom to the top has retired from the aerospace industry but companies such as SpaceX has initialized the launching of rocket recently.

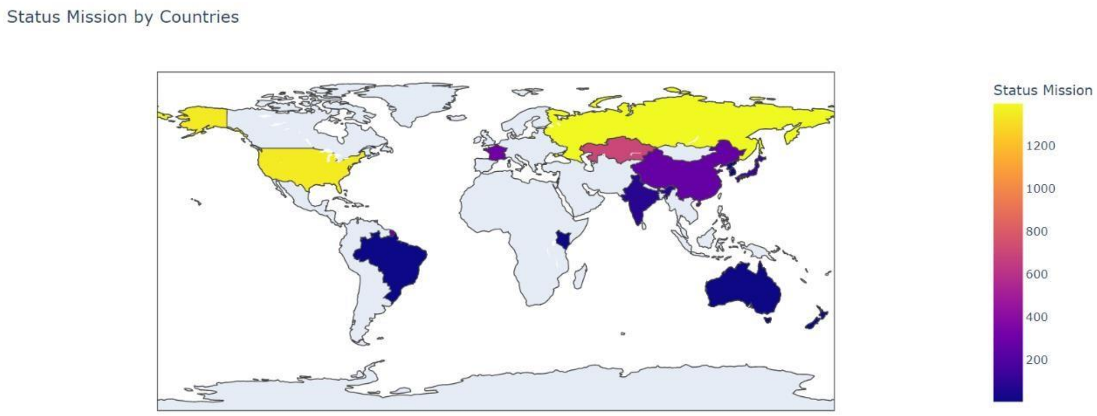
MISSIONS BY COUNTIES:

In [17]:
```
mapdf = df.groupby(["country","alpha3"])["Status Mission"].count().reset_index()
mapdf.head()
```

Out[17]:

	country	alpha3	Status Mission
0	Australia	AUS	6
1	Brazil	BRA	3
2	China	CHN	269
3	France	FRA	303
4	India	IND	76

In [18]:
```
fig = px.choropleth(mapdf, locations = "alpha3", hover_name = "country", color = "Status Mission", title ="Status Mission by Countries")
fig.show()
```



This plot shows that the Russia has launched the largest number of the missions, following that is the USA, next is Mongolia, China, and India.

In [4]:
```
import pandas as pd
import numpy as np
import plotly.graph_objs as go
import plotly.express as px
from plotly.subplots import make_subplots
from plotly.offline import init_notebook_mode, iplot, plot
import plotly as py
from pywaffle import Waffle
import matplotlib.pyplot as plt
init_notebook_mode(connected=False)

#importing modeling libraries
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
import xgboost as xgb
from sklearn.ensemble import RandomForestClassifier
```

In [3]:
```
pip install xgboost
```

```
Collecting xgboost
  Using cached xgboost-1.6.0-py3-none-win_amd64.whl (126.1 MB)
Requirement already satisfied: numpy in c:\users\kumar\anaconda3\lib\site-packages (from xgboost) (1.20.1)
Requirement already satisfied: scipy in c:\users\kumar\anaconda3\lib\site-packages (from xgboost) (1.6.2)
Installing collected packages: xgboost
Successfully installed xgboost-1.6.0
Note: you may need to restart the kernel to use updated packages.
```

In [5]:
```
df = pd.read_csv("C:/Users/kumar/OneDrive/Documents/IV_J_COMPONENT/Space_Corrected.csv")
df.drop(['Unnamed: 0.1','Unnamed: 0'], axis = 1, inplace = True)
df.head()
```

Out[5]:

	Company Name	Location	Datum	Detail	Status Rocket	Rocket	Status Mission
0	SpaceX	LC-39A, Kennedy Space Center, Florida, USA	Fri Aug 07, 2020 05:12 UTC	Falcon 9 Block 5 | Starlink V1 L9 & BlackSky	StatusActive	50.0	Success
1	CASC	Site 9401 (SLS-2), Jiuquan Satellite Launch Ce...	Thu Aug 06, 2020 04:01 UTC	Long March 2D | Gaofen-9 04 & Q-SAT	StatusActive	29.75	Success
2	SpaceX	Pad A, Boca Chica, Texas, USA	Tue Aug 04, 2020 23:57 UTC	Starship Prototype | 150 Meter Hop	StatusActive	NaN	Success
3	Roscosmos	Site 200/39, Baikonur Cosmodrome, Kazakhstan	Thu Jul 30, 2020 21:25 UTC	Proton-M/Briz-M | Ekspress-80 & Ekspress-103	StatusActive	65.0	Success
4	ULA	SLC-41, Cape Canaveral AFS, Florida, USA	Thu Jul 30, 2020 11:50 UTC	Atlas V 541 | Perseverance	StatusActive	145.0	Success

In [6]:
```
#function to extract the name of the country from the Location
def extract_country_name(location):
    country = location.split(',')[-1]
    country = country.strip()
    return country

#dictionary to help in mapping to get consistent and correct Country Names
countries_dict = {
    'Russia' : 'Russian Federation',
    'New Mexico' : 'USA',
    "Yellow Sea": 'China',
    "Shahrud Missile Test Site": "Iran",
    "Pacific Missile Range Facility": 'USA',
    "Barents Sea": 'Russian Federation',
    "Gran Canaria": 'USA'
}
```

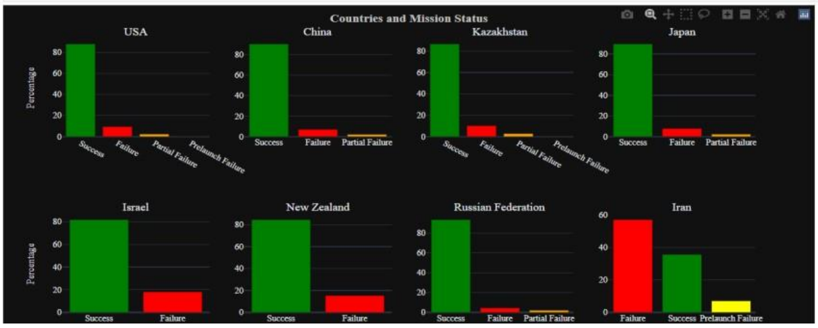
In [9]:
```
df['Country'] = df['Location'].apply(lambda x: extract_country_name(x))
df['Country'] = df['Country'].replace(countries_dict)
```

We have also plotted the success, failure, partial and prelaunch failures percentage of all the countries. We see that in terms of Success Percentages :

4. Kenya ranks first with a **100% success rate**. Kenya has made 9 Space Missions and all of them are Successful.
5. France with 303 space missions comes in second with a Success percent of 94%
6. Russia with 1398 Space Missions comes at a close 3rd with a Success rate of 93.34%. Now that's something. In comparison to Launches taking place in USA, Russia fares better as US missions have a Success rate of around 88% In terms of Failure rate:
7. Brazil and South Korea have a similar failure rate of 66.67%, i.e 2/3rd of their space missions fail.
    - It should be noted that South Korea has a Success rate of 33%, while **Brazil is yet to make a successful Space mission.**
    - These are not disheartening results,as both South Korea and Brazil have only made 3 Space Mission attempts.
8. Next, we have North Korea with a Failure rate of 60% in its 5 space missions.
9. Iran has a Failure rate of about 57%. It has however, undertaken only 14 space missions.

```
In [14]:
from sklearn.preprocessing import LabelEncoder
encoder = LabelEncoder()
encoder.fit(df['Status Mission'])
colors = {0 : 'red', 1 : 'Orange', 2 : 'Yellow', 3 : 'Green'}
```

```
In [15]:
fig = make_subplots(rows = 4 ,cols = 4,subplot_titles=df['Country'].unique())
for i, country in enumerate(df['Country'].unique()):
    counts = df[df['Country'] == country]['Status Mission'].value_counts(normalize = True) * 100
    color = [colors[x] for x in encoder.transform(counts.index)]
    trace = go.Bar(x = counts.index, y = counts.values, name = country,showlegend=False,marker={'color' : color})
    fig.add_trace(trace, row = (i//4)+1, col = (i%4)+1)
fig.update_layout(template = 'plotly_dark',margin=dict(l=80, r=80, t=50, b=10),
                  title = { 'text' : '<b>Countries and Mission Status</b>', 'x' : 0.5},
                 font_family = 'Fira Code',title_font_color= '#cacaca',height = 1000,width = 1100)
for i in range(1,5):
    fig.update_yaxes(title_text = 'Percentage',row = i, col = 1)
fig.show()
```



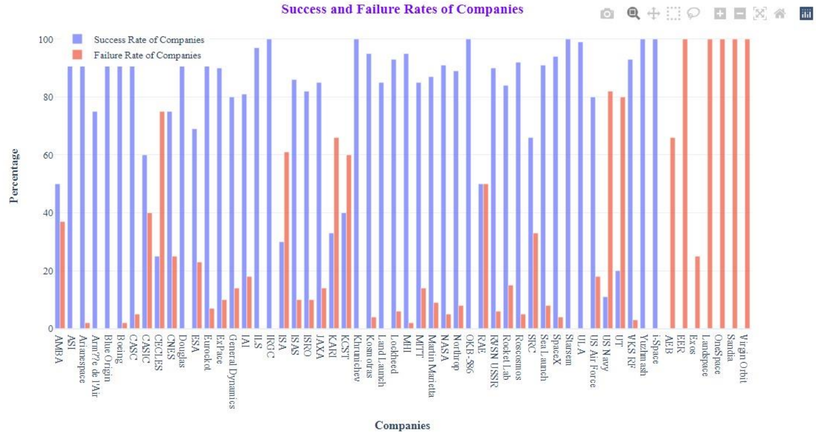

```
In [13]:
country_counts = dict(df['Country'].value_counts())
fig = go.Figure(data=[go.Table(
    header=dict(values=['<b>Country Name</b>', '<b>Number of Space Missions</b>'],
                line_color='black',
                fill_color='darkorange',
                align='left',
                font=dict(color='black', size=14)),
    cells=dict(values=[list(country_counts.keys()),
                       list(country_counts.values())],
               line_color='black',
               fill_color='white',
               align='left',
               font=dict(color='black', size=13)))
])

fig.update_layout(width=500, height=450,margin=dict(l=80, r=80, t=25, b=10),
                  title = { 'text' : '<b>Number of Space Missions Per Launch Location</b>', 'x' : 0.95},
                 font_family = 'Fira Code',title_font_color= '#ff0d00')
fig.show()
```

Number of Space Missions Per Launch Location

Country Name	Number of Space Missions
Russian Federation	1398
USA	1351
Kazakhstan	701
France	303
China	269
Japan	126
India	76
Pacific Ocean	36
Iran	14
New Zealand	13
Israel	11
Kenya	9
Australia	6
North Korea	5
Brazil	3
South Korea	3

We see that a large number of Space Missions launch from Russia and USA. A lot of this has been because of the Space Race between the two countries.

```
In [17]:
trace1 = go.Bar(x = successPerc.index, y = successPerc.values, name = 'Success Rate of Companies',opacity=0.7)
trace2 = go.Bar(x = FailurePerc.index, y = FailurePerc.values, name = 'Failure Rate of Companies',opacity=0.7)
fig = go.Figure([trace1,trace2])
fig.update_layout(template = 'plotly_white',margin=dict(l=80, r=80, t=25, b=10),
                  title = {'text' : '<b>Success and Failure Rates of Companies</b>', 'x' : 0.5},
                 font_family = 'Fira Code',title_font_color= '#8000ff',width = 1000,yaxis_title = '<b>Percentage</b>',xaxis_title = '<b>Companies</b>',
                 legend=dict(
                    yanchor="top",
                    y=0.99,
                    xanchor="left",
                    x=0.01
))

fig.show()
```



We measure the Success and Failure Rate of Each company.

Success Rate highlights:

- Companies like ASI,Blue Origin,Douglas, IRGC,Khrunichev, OKB-586,Starsem, Yuzhmash and iSpace have a **100% success rate.**

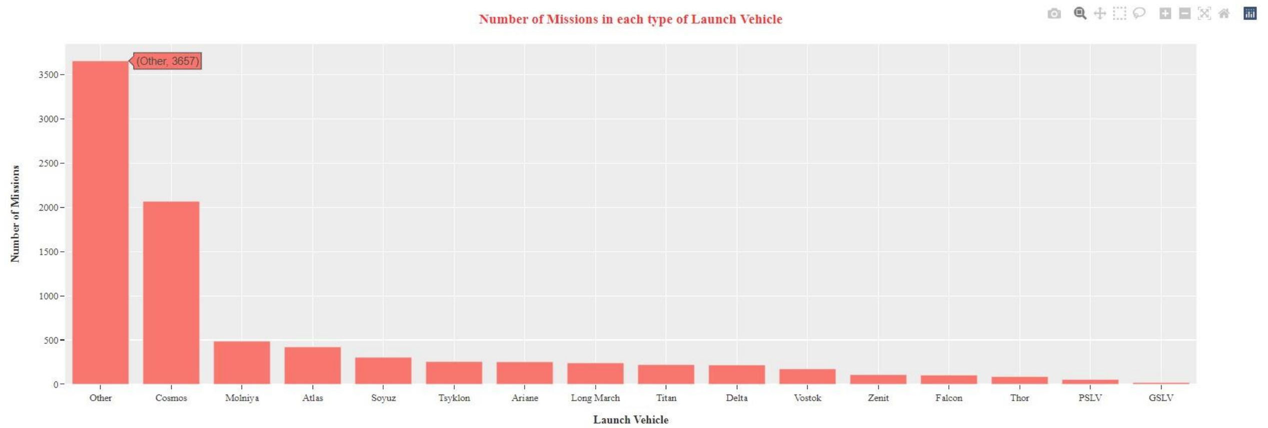
Failure Rate highlights

- Companies like EER, Landspace, OneSpace, Sandia and Virgin Orbit have a **100% failure rate**, i.e they haven't had any successful space mission.

In [19]:
```
fig = px.treemap(df,path = ['Status Mission','Country','Company Name'])
fig.update_layout(template = 'ggplot2',margin=dict(l=80, r=80, t=50, b=10),
                  title = { 'text' : '<b>Mission Status,Countries and Companies</b>', 'x' : 0.5},
                 font_family = 'Fira Code',title_font_color= '#ff6767')
fig.show()
```

From the above treemap, I was immediately interested in seeing **'RVSN USSR' having Launch Locations in both Kazakhstan and Russia.**

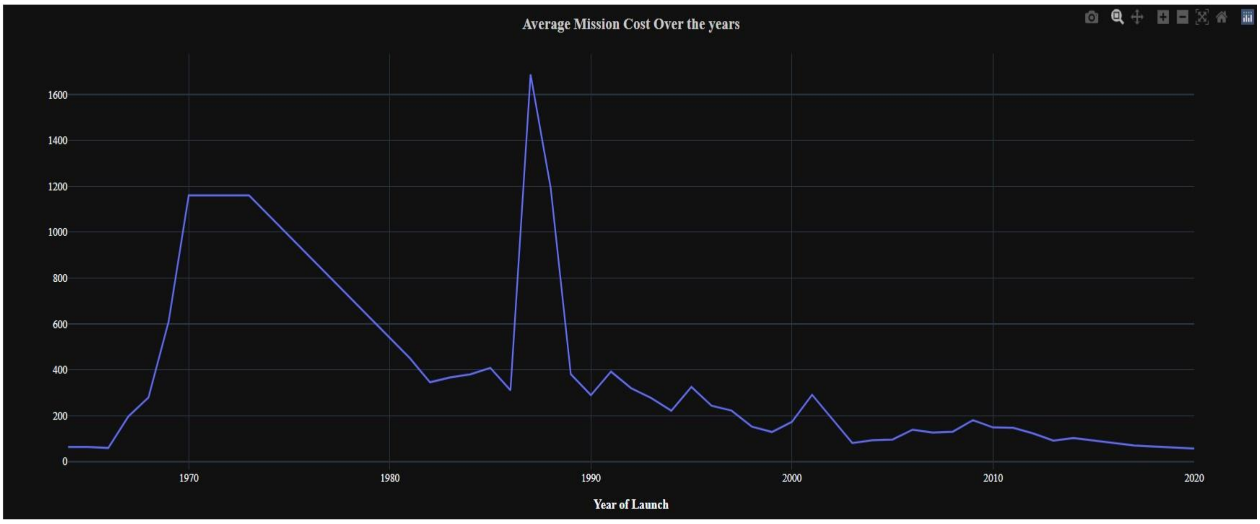
In [22]:
```
counts = dict(pd.Series(details).value_counts(sort = True))
fig = go.Figure(go.Bar(x = list(counts.keys()), y = list(counts.values())))
fig.update_layout(template = 'ggplot2',margin=dict(l=80, r=80, t=50, b=10),
                  title = { 'text' : '<b>Number of Missions in each type of Launch Vehicle</b>', 'x' : 0.5},
                 font_family = 'Fira Code',title_font_color= '#ff3434',
                 yaxis_title = '<b>Number of Missions</b>',xaxis_title = '<b>Launch Vehicle</b>',)
fig.show()
```



The previous result of USA reducing the average mission cost over years is also confirmed here from the results shown from NASA.

Another thing that I found particularly interesting is that the earlier SpaceX missions were unsuccessful and they had markedly low mission cost as compared to their later Space Missions. Thus, increasing the budget that they allocate for each space mission helped them become more Successful.

```
In [27]: costDict = dict(df[df[' Rocket'] > 0].groupby('year')[' Rocket'].mean())
fig = go.Figure(go.Scatter(x = list(costDict.keys()), y = list(costDict.values()), yaxis = 'y2',mode = 'lines',showlegend=False,name = 'Average Mission Cost Over the years'))
fig.update_layout(template = 'plotly_dark',margin=dict(l=80, r=80, t=50, b=10),
                  title = { 'text' : '<b>Average Mission Cost Over the years</b>', 'x' : 0.5},
                 font_family = 'Fira Code',title_font_color= '#cacaca',
                 yaxis_title = '<b>Cost of Mission in Million Dollars</b>',xaxis_title = '<b>Year of Launch</b>',)
fig.show()
```

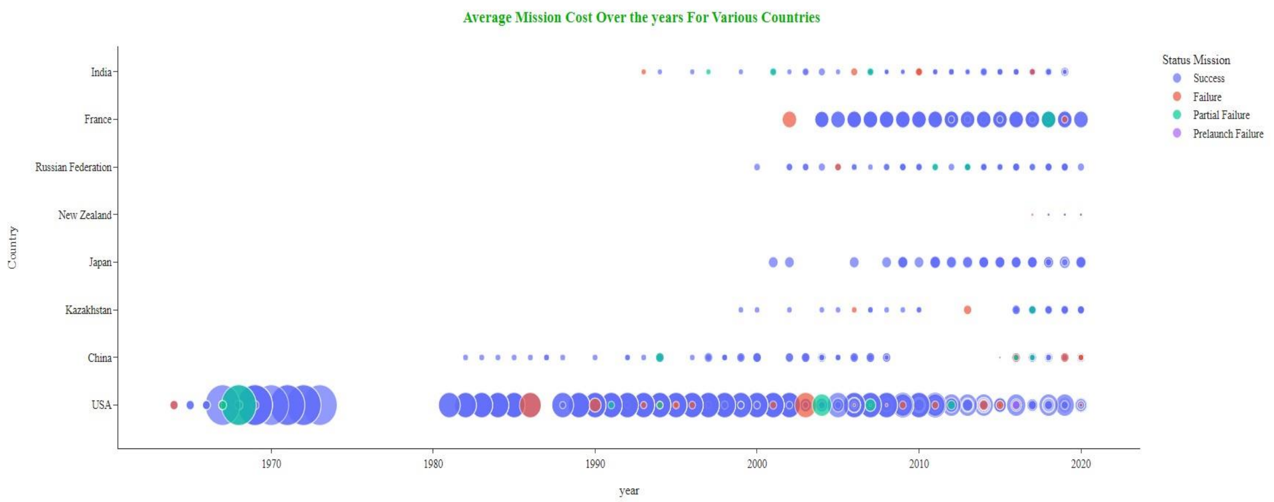


We see that overtime, **the average mission cost has decreased** since 1987. However, as seen earlier, the failure rate of Space Missions has decreased over time. Thus, one thing in clear, **as the technology has advanced, we have been able to do space missions at a lower cost and with lower failure rates.**

The reason why we see a very high spike in 1987 is because of a **Space Mission by RVSN USSR that has an estimated cost of 5000 million dollars.** Now, that's a lotta money! This skewed the data a lot as most other space missions of that year did not have any Mission Cost listed in the dataset.

AVERAGE MISSION COST OVER THE YEARS FOR VARIOUS COUNTRIES:

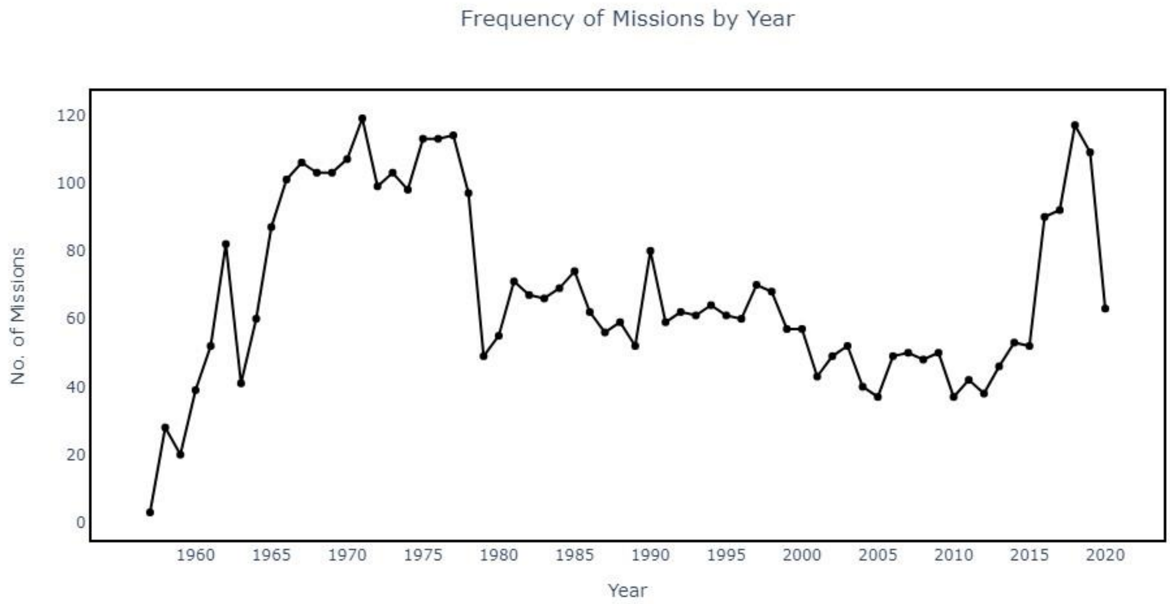
```
In [28]: fig = px.scatter(df[df[' Rocket'].between(1,4999)],x = 'year', y = 'Country', color = 'Status Mission',size = ' Rocket', size_max=30)
fig.update_layout(template = 'simple_white',margin=dict(l=80, r=80, t=50, b=10),
                  title = { 'text' : '<b>Average Mission Cost Over the years For Various Countries</b>', 'x' : 0.5},
                 font_family = 'Fira Code',title_font_color= '#00b300')
fig.show()
```



One important trend that we see is that the USA has managed to lower it's cost of Space Missions over time The previous result of USA reducing the average mission cost over years is also confirmed here from the results shown from NASA.

Another thing that I found particularly interesting is that the earlier SpaceX missions were unsuccessful and they had markedly low mission cost as compared to their later Space Missions. Thus, **increasing the budget that they allocate for each space mission helped them become more Successful.**

FREQUENCY OF MISSIONS BY YEAR:



In this Visualization, we can infer the Frequency of the mission and the number of missions performed in a year. In the 1960s Space exploration has started growing up massively, by 1970's it reached up to 120 space missions and by 2005-2010 it is decreased by 40 and again after 20 the Century it has seen an all-time` high rate of space missions performed.

Final inferences:

- The success ratio of the pioneer organizations have come down a lot due to constant technology development and reseach for new techniques.
- the no of launches has become low from 1976 to 2015 as the versatility of the satellites havbe increased, explaining that a single launch can put multiplle satetllites to the orbit.
- United States of America has launched a the largest number of rockets in the space race and the second is arianespace, ULCA and comes RSVN USSR.
- There is large sum spent on the space race to find the unknowns of the space and the origins of the human race.
- At this rate the cost of the missions are increasing linearly and there is a need to reduce the cost to reach the future flagship missions of colonisation of mars and other projects.