

```
In [1]: import pandas as pd  
import matplotlib.pyplot as plt
```

```
In [2]: space_df=pd.read_csv('space_missions.csv',encoding='latin-1')
```

```
In [3]: space_df.head()
```

```
Out[3]:
```

	Company	Location	Date	Time	Rocket	Mission	RocketStatus	Price	MissionStatus
0	RVSN USSR	Site 1/5, Baikonur Cosmodrome, Kazakhstan	1957-10-04	19:28:00	Sputnik 8K71PS	Sputnik-1	Retired	NaN	Success
1	RVSN USSR	Site 1/5, Baikonur Cosmodrome, Kazakhstan	1957-11-03	02:30:00	Sputnik 8K71PS	Sputnik-2	Retired	NaN	Success
2	US Navy	LC-18A, Cape Canaveral AFS, Florida, USA	1957-12-06	16:44:00	Vanguard	Vanguard TV3	Retired	NaN	Failure
3	AMBA	LC-26A, Cape Canaveral AFS, Florida, USA	1958-02-01	03:48:00	Juno I	Explorer 1	Retired	NaN	Success
4	US Navy	LC-18A, Cape Canaveral AFS, Florida, USA	1958-02-05	07:33:00	Vanguard	Vanguard TV3BU	Retired	NaN	Failure

```
In [4]: space_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 4630 entries, 0 to 4629  
Data columns (total 9 columns):  
 #   Column            Non-Null Count  Dtype     
---  --    
 0   Company          4630 non-null    object    
 1   Location          4630 non-null    object    
 2   Date              4630 non-null    object    
 3   Time              4503 non-null    object    
 4   Rocket             4630 non-null    object    
 5   Mission            4630 non-null    object    
 6   RocketStatus       4630 non-null    object    
 7   Price              1265 non-null    object    
 8   MissionStatus      4630 non-null    object    
dtypes: object(9)  
memory usage: 325.7+ KB
```

```
In [5]: space_df.shape
```

```
Out[5]: (4630, 9)
```

```
In [6]: space_df['Price'].isna().value_counts()
```

```
Out[6]: True    3365  
False   1265  
Name: Price, dtype: int64
```

**Question 1: How have rocket launches trended across time? Has mission success rate increased?**

```
In [7]: space_df['Date'] = pd.to_datetime(space_df['Date'], infer_datetime_format=True)
```

```
In [8]: space_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 4630 entries, 0 to 4629  
Data columns (total 9 columns):  
 #   Column      Non-Null Count  Dtype     
 ---  --          --          --  
 0   Company     4630 non-null   object    
 1   Location    4630 non-null   object    
 2   Date        4630 non-null   datetime64[ns]  
 3   Time        4503 non-null   object    
 4   Rocket       4630 non-null   object    
 5   Mission      4630 non-null   object    
 6   RocketStatus 4630 non-null   object    
 7   Price        1265 non-null   object    
 8   MissionStatus 4630 non-null   object    
dtypes: datetime64[ns](1), object(8)  
memory usage: 325.7+ KB
```

```
In [9]: space_df['year_'] = space_df['Date'].map(lambda dt: dt.strftime('%Y'))
```

```
In [10]: space_df['MissionStatus'].value_counts()
```

```
Out[10]: Success           4162  
Failure            357  
Partial Failure    107  
Prelaunch Failure 4  
Name: MissionStatus, dtype: int64
```

```
In [11]: grouped = space_df.groupby('year_')['MissionStatus'].size().to_frame('TotalStatusPerYear')
```

```
In [12]: grouped
```

```
Out[12]: TotalStatusPerYear
```

```
year_
```

```
1957
```

```
3
```

### TotalStatusPerYear

year_	
1958	28
1959	20
1960	39
1961	52
...	...
2018	117
2019	109
2020	119
2021	157
2022	93

66 rows × 1 columns

```
In [13]: groupMS = space_df.groupby(['year_', 'MissionStatus']).size().to_frame('CountMissionSta
```

```
In [14]: groupMS.head()
```

Out[14]: **CountMissionStatusSplit**

year_	MissionStatus	
1957	Failure	1
	Success	2
1958	Failure	20
	Partial Failure	2
	Success	6

```
In [15]: groupFinal=grouped.join(groupMS)
```

```
In [16]: groupFinal['PercentMissionStatus']=(groupFinal['CountMissionStatusSplit']/groupFinal['T
```

```
In [17]: groupFinal.sort_values(by=['PercentMissionStatus'],ascending=False).head()
```

Out[17]: **TotalStatusPerYear CountMissionStatusSplit PercentMissionStatus**

year_	MissionStatus			
1983	Success	66	65	98.484848

		TotalStatusPerYear	CountMissionStatusSplit	PercentMissionStatus
year_	MissionStatus			
1978	Success	97	94	96.907216
2022	Success	93	90	96.774194
1988	Success	59	57	96.610169
2018	Success	117	113	96.581197

```
In [18]: groupFinal.reset_index(inplace=True)
```

```
In [19]: groupFinal.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 184 entries, 0 to 183
Data columns (total 5 columns):
 #   Column           Non-Null Count  Dtype  
---  --  
 0   year_            184 non-null    object 
 1   MissionStatus    184 non-null    object 
 2   TotalStatusPerYear 184 non-null    int64  
 3   CountMissionStatusSplit 184 non-null    int64  
 4   PercentMissionStatus 184 non-null    float64
dtypes: float64(1), int64(2), object(2)
memory usage: 7.3+ KB
```

```
In [20]: groupFinal
```

	year_	MissionStatus	TotalStatusPerYear	CountMissionStatusSplit	PercentMissionStatus
0	1957	Failure	3	1	33.333333
1	1957	Success	3	2	66.666667
2	1958	Failure	28	20	71.428571
3	1958	Partial Failure	28	2	7.142857
4	1958	Success	28	6	21.428571
...	...	...	...	...	...
179	2021	Failure	157	11	7.006369
180	2021	Partial Failure	157	3	1.910828
181	2021	Success	157	143	91.082803
182	2022	Failure	93	3	3.225806
183	2022	Success	93	90	96.774194

184 rows × 5 columns

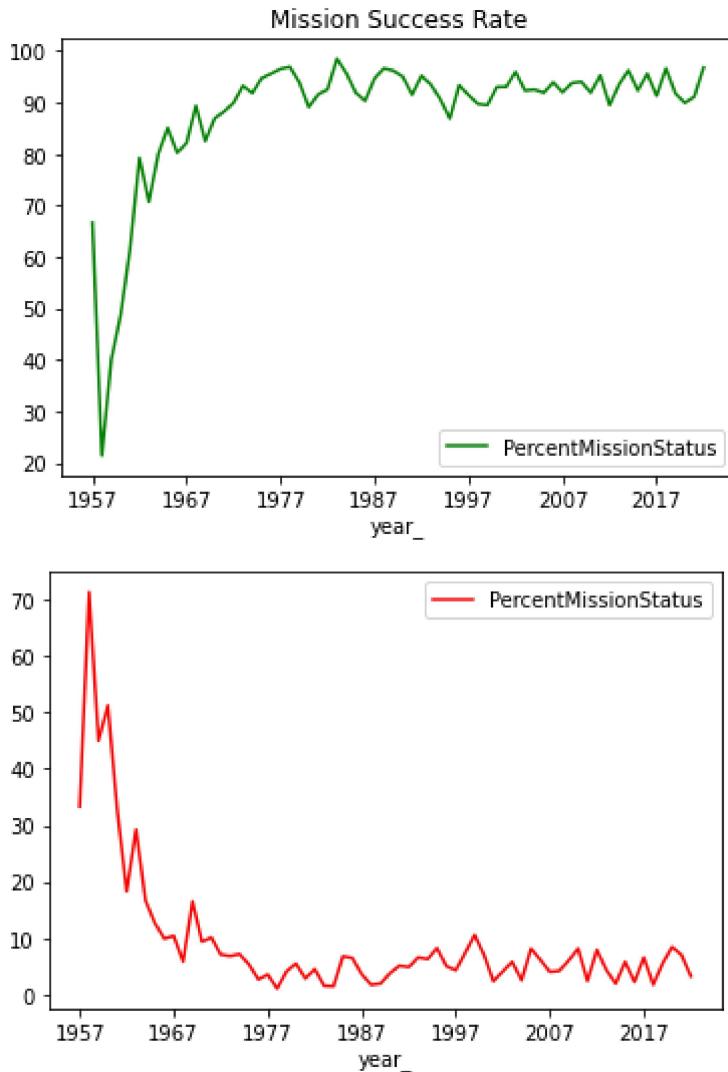
```
In [21]: #groupFinal.plot.line(x="year_",y="PercentMissionStatus",title="Mission Success Rate")
```

```

groupFinal[groupFinal['MissionStatus']=='Success'].plot.line(x="year_",y="PercentMissionStatus")
groupFinal[groupFinal['MissionStatus']=='Failure'].plot.line(x="year_",y="PercentMissionStatus")

plt.show();

```



We can notice that the Mission Success Rate has increased after 1960s and has been around 90 percent ever since.

**Question 2: Which countries have had the most successful space missions? Has it always been that way?**

```
In [22]: space_df['Country']=space_df['Location'].str.split(',').str[-1]
```

```
In [23]: space_df
```

	Company	Location	Date	Time	Rocket	Mission	RocketStatus	Price	MissionStatus
0	RVSN USSR	Site 1/5, Baikonur Cosmodrome, Kazakhstan	1957-10-04	19:28:00	Sputnik 8K71PS	Sputnik-1	Retired	NaN	Success

	Company	Location	Date	Time	Rocket	Mission	RocketStatus	Price	MissionStatus
1	RVSN USSR	Site 1/5, Baikonur Cosmodrome, Kazakhstan	1957-11-03	02:30:00	Sputnik 8K71PS	Sputnik-2	Retired	NaN	Success
2	US Navy	LC-18A, Cape Canaveral AFS, Florida, USA	1957-12-06	16:44:00	Vanguard	Vanguard TV3	Retired	NaN	Failure
3	AMBA	LC-26A, Cape Canaveral AFS, Florida, USA	1958-02-01	03:48:00	Juno I	Explorer 1	Retired	NaN	Success
4	US Navy	LC-18A, Cape Canaveral AFS, Florida, USA	1958-02-05	07:33:00	Vanguard	Vanguard TV3BU	Retired	NaN	Failure
...	...	...	...	...	...	...	...	...	...
4625	SpaceX	SLC-4E, Vandenberg SFB, California, USA	2022-07-22	17:39:00	Falcon 9 Block 5	Starlink Group 3-2	Active	67	Success
4626	CASC	LC-101, Wenchang Satellite Launch Center, China	2022-07-24	06:22:00	Long March 5B	Wentian	Active	NaN	Success
4627	SpaceX	LC-39A, Kennedy Space Center, Florida, USA	2022-07-24	13:38:00	Falcon 9 Block 5	Starlink Group 4-25	Active	67	Success
4628	CAS Space	Jiuquan Satellite Launch Center, China	2022-07-27	04:12:00	Zhongke-1A	Demo Flight	Active	NaN	Success
4629	CASC	LC-3, Xichang Satellite Launch Center, China	2022-07-29	13:28:00	Long March 2D	Yaogan 35 Group 03	Active	29.75	Success

4630 rows × 11 columns



```
In [24]: space_df.groupby(['MissionStatus', 'Country']).size().sort_values(ascending=False).head
```

```
Out[24]: MissionStatus    Country
Success        Russia      1323
              USA         1298
              Kazakhstan  625
```

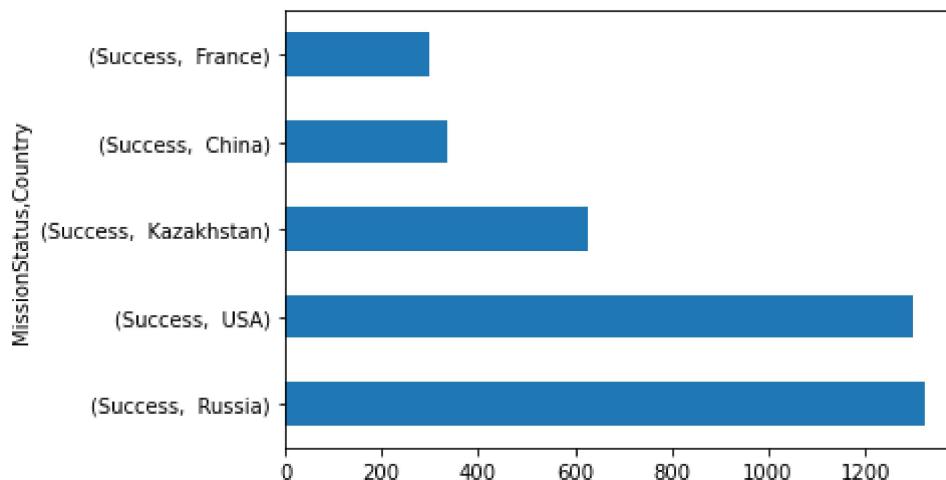
```
    China      335  
    France     299
```

dtype: int64

Russia, USA, Kazakhstan, China and France are the countries with most successful space missions

```
In [25]: space_df.groupby(['MissionStatus', 'Country']).size().sort_values(ascending=False).head
```

```
Out[25]: <AxesSubplot:ylabel='MissionStatus,Country'>
```



**Question 3 : Which rocket has been used for the most space missions? Is it still active?**

```
In [26]: space_df.groupby(['Rocket', 'RocketStatus']).size().sort_values(ascending=False).head(5)
```

```
Out[26]: Rocket          RocketStatus  
Cosmos-3M (11K65M)    Retired        446  
Voskhod                 Retired        299  
Molniya-M /Block ML   Retired        128  
Cosmos-2I (63SM)       Retired        126  
Soyuz U                  Retired        125  
dtype: int64
```

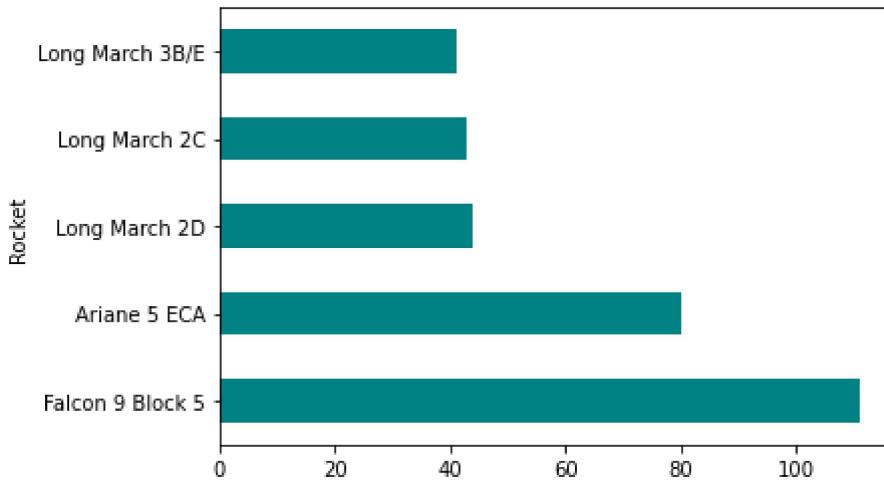
Since the above list provides Rockets whose status is Retired, applying filter to obtain only Active Rockets

```
In [27]: space_df[space_df['RocketStatus']=='Active'].groupby(['Rocket']).size().sort_values(asc
```

```
Out[27]: Rocket  
Falcon 9 Block 5      111  
Ariane 5 ECA           80  
Long March 2D            44  
Long March 2C            43  
Long March 3B/E          41  
dtype: int64
```

```
In [28]: space_df[space_df['RocketStatus']=='Active'].groupby(['Rocket']).size().sort_values(asc
```

```
Out[28]: <AxesSubplot:ylabel='Rocket'>
```



#### Question 4: Are there any patterns you can notice with the launch locations?

```
In [29]: space_df['LaunchSite']=space_df['Location'].str.split(',').str[0]
```

```
In [30]: space_df['LaunchSite'].head()
```

```
Out[30]: 0    Site 1/5
1    Site 1/5
2    LC-18A
3    LC-26A
4    LC-18A
Name: LaunchSite, dtype: object
```

```
In [31]: space_df[['LaunchSite','Country','MissionStatus']].head()
```

	LaunchSite	Country	MissionStatus
<b>0</b>	Site 1/5	Kazakhstan	Success
<b>1</b>	Site 1/5	Kazakhstan	Success
<b>2</b>	LC-18A	USA	Failure
<b>3</b>	LC-26A	USA	Success
<b>4</b>	LC-18A	USA	Failure

```
In [32]: space_df.groupby(['LaunchSite','Country','MissionStatus']).size().sort_values(ascending
```

LaunchSite	Country	MissionStatus	Missions
Site 31/6	Kazakhstan	Success	236
Site 132/1	Russia	Success	203
Site 43/4	Russia	Success	199
Site 41/1	Russia	Success	186
Site 132/2	Russia	Success	164
ELS	France	Partial Failure	1
LC-14	USA	Partial Failure	1
K-407 Submarine	Barents Sea	Success	1
Site 35/1	Russia	Failure	1

```
SLC-40          USA      Prelaunch Failure      1  
Length: 284, dtype: int64
```

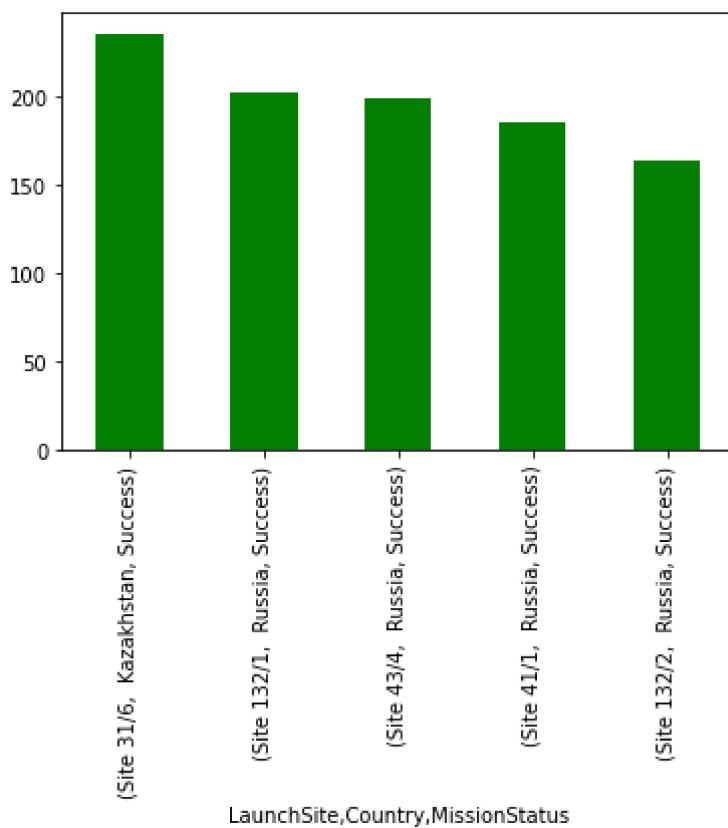
We can notice that Sites Site 31/6, Site 132/1 , Site 43/4 have maximum successful launches

```
In [33]: space_df[space_df['MissionStatus']=='Success'].groupby(['LaunchSite','Country','Mission
```

```
Out[33]: LaunchSite  Country      MissionStatus  
Site 31/6    Kazakhstan  Success        236  
Site 132/1    Russia     Success        203  
Site 43/4    Russia     Success        199  
Site 41/1    Russia     Success        186  
Site 132/2    Russia     Success        164  
dtype: int64
```

```
In [34]: space_df[space_df['MissionStatus']=='Success'].groupby(['LaunchSite','Country','Mission
```

```
Out[34]: <AxesSubplot:xlabel='LaunchSite,Country,MissionStatus'>
```



```
In [ ]:
```