Tasks

- Task1: Remove unwanted columns
- Task2: Check if there is any null entries and check duplicate records¶
- Task3: Analysis of mission_status¶
- Task4: Analysis of Rocket_Status¶
- Task5: Analysis of Organization¶
- Task 5.1: Find which organisation have highest number of rocket launches.¶
- Task 5.2: Find number of successfull Launches and Failure(Failure+PreLaunch Failure+Partial Failure) Launches for each organisation¶
- Task6: Analysis of Date Column¶
- Task 6.1: Create new column(launch_year) which contain year of corresponding records only¶
- Task 6.2: Find number of launched rockets year-wise¶

Import libraries

In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
```

Read csv file with using pandas

In [2]:

In [3]:

Out[3]:

	Unnamed: 0.1	Unnamed: 0	Organisation	Location	Date	Detail	Rocket_Status	Pri
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
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.
2	2	2	SpaceX	Pad A, Boca Chica, Texas, USA	Tue Aug 04, 2020 23:57 UTC	Starship Prototype | 150 Meter Hop	StatusActive	Na
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
4	4	4	ULA	SLC-41, Cape Canaveral AFS, Florida, USA	Thu Jul 30, 2020 11:50 UTC	Atlas V 541 | Perseverance	StatusActive	145
...	...	...	...	...	...	...	...	
4319	4319	4319	US Navy	LC-18A, Cape Canaveral AFS, Florida, USA	Wed Feb 05, 1958 07:33 UTC	Vanguard | Vanguard TV3BU	StatusRetired	Na
4320	4320	4320	AMBA	LC-26A, Cape Canaveral AFS, Florida, USA	Sat Feb 01, 1958 03:48 UTC	Juno I | Explorer 1	StatusRetired	Na
4321	4321	4321	US Navy	LC-18A, Cape Canaveral AFS, Florida, USA	Fri Dec 06, 1957 16:44 UTC	Vanguard | Vanguard TV3	StatusRetired	Na

	Unnamed: 0.1	Unnamed: 0	Organisation	Location	Date	Detail	Rocket_Status	Pri
4322	4322	4322	RVSN USSR	Site 1/5, Baikonur Cosmodrome, Kazakhstan	Sun Nov 03, 1957 02:30 UTC	Sputnik 8K71PS | Sputnik-2	StatusRetired	Na

Fri

check head,tail,shape & sample functions

In [4]:

Out[4]:

	Unnamed: 0.1	Unnamed: 0	Organisation	Location	Date	Detail	Rocket_Status	Price
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
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
2	2	2	SpaceX	Pad A, Boca Chica, Texas, USA	Tue Aug 04, 2020 23:57 UTC	Starship Prototype | 150 Meter Hop	StatusActive	NaN
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
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

In [5]:

Out[5]:

	Unnamed: 0.1	Unnamed: 0	Organisation	Location	Date	Detail	Rocket_Status	Price
4319	4319	4319	US Navy	LC-18A, Cape Canaveral AFS, Florida, USA	Wed Feb 05, 1958 07:33 UTC	Vanguard | Vanguard TV3BU	StatusRetired	NaN
4320	4320	4320	AMBA	LC-26A, Cape Canaveral AFS, Florida, USA	Sat Feb 01, 1958 03:48 UTC	Juno I | Explorer 1	StatusRetired	NaN
4321	4321	4321	US Navy	LC-18A, Cape Canaveral AFS, Florida, USA	Fri Dec 06, 1957 16:44 UTC	Vanguard | Vanguard TV3	StatusRetired	NaN
4322	4322	4322	RVSN USSR	Site 1/5, Baikonur Cosmodrome, Kazakhstan	Sun Nov 03, 1957 02:30 UTC	Sputnik 8K71PS | Sputnik-2	StatusRetired	NaN
4323	4323	4323	RVSN USSR	Site 1/5, Baikonur Cosmodrome, Kazakhstan	Fri Oct 04, 1957 19:28 UTC	Sputnik 8K71PS | Sputnik-1	StatusRetired	NaN

In [6]:

Out[6]:

	Unnamed: 0.1	Unnamed: 0	Organisation	Location	Date	Detail	Rocket_Status	Price
1945	1945	1945	RVSN USSR	Site 32/1, Plesetsk Cosmodrome, Russia	Mon May 30, 1988 07:59 UTC	Tsyklon-3 | Cosmos 1950	StatusRetired	NaN

In [7]:

Out[7]: (4324, 9)

check datatypes of column and Null values

In [8]:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4324 entries, 0 to 4323
Data columns (total 9 columns):
 #   Column          Non-Null Count  Dtype 
---  ------          --------------  ----- 
 0   Unnamed: 0.1    4324 non-null   int64 
 1   Unnamed: 0      4324 non-null   int64 
 2   Organisation    4324 non-null   object
 3   Location        4324 non-null   object
 4   Date            4324 non-null   object
 5   Detail          4324 non-null   object
 6   Rocket_Status   4324 non-null   object
 7   Price           964 non-null    object
 8   Mission_Status  4324 non-null   object
dtypes: int64(2), object(7)
memory usage: 304.2+ KB
```

Performing Exploratory Data Analysis(EDA)

Task1: Remove unwanted columns.

In [9]:

In [10]:

Out[10]:

	Unnamed: 0	Organisation	Location	Date	Detail	Rocket_Status	Price	Mission_S
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	Su
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	Su
2	2	SpaceX	Pad A, Boca Chica, Texas, USA	Tue Aug 04, 2020 23:57 UTC	Starship Prototype | 150 Meter Hop	StatusActive	NaN	Su
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	Su
4	4	ULA	SLC-41, Cape Canaveral AFS, Florida, USA	Thu Jul 30, 2020 11:50 UTC	Atlas V 541 | Perseverance	StatusActive	145.0	Su
5	5	CASC	LC-9, Taiyuan Satellite Launch Center, China	Sat Jul 25, 2020 03:13 UTC	Long March 4B | Ziyuan-3 03, Apocalypse-10 & N...	StatusActive	64.68	Su
6	6	Roscosmos	Site 31/6, Baikonur Cosmodrome, Kazakhstan	Thu Jul 23, 2020 14:26 UTC	Soyuz 2.1a | Progress MS-15	StatusActive	48.5	Su
7	7	CASC	LC-101, Wenchang Satellite Launch Center, China	Thu Jul 23, 2020 04:41 UTC	Long March 5 | Tianwen-1	StatusActive	NaN	Su
8	8	SpaceX	SLC-40, Cape Canaveral AFS, Florida, USA	Mon Jul 20, 2020 21:30 UTC	Falcon 9 Block 5 | ANASIS-II	StatusActive	50.0	Su

	Unnamed: 0	Organisation	Location	Date	Detail	Rocket_Status	Price	Mission_S
9	9	JAXA	LA-Y1, Tanegashima Space Center, Japan	Sun Jul 19, 2020 21:58 UTC	H-IIA 202 | Hope Mars Mission	StatusActive	90.0	Su
10	10	Northrop	LP-0B, Wallops Flight Facility, Virginia, USA	Wed Jul 15, 2020 13:46 UTC	Minotaur IV | NROL-129	StatusActive	46.0	Su
11	11	ExPace	Site 95, Jiuquan Satellite Launch Center, China	Fri Jul 10, 2020 04:17 UTC	Kuaizhou 11 | Jilin-1 02E, CentiSpace-1 S2	StatusActive	28.3	F
12	12	CASC	LC-3, Xichang Satellite Launch Center, China	Thu Jul 09, 2020 12:11 UTC	Long March 3B/E | Apstar-6D	StatusActive	29.15	Su
13	13	IAI	Pad 1, Palmachim Airbase, Israel	Mon Jul 06, 2020 01:00 UTC	Shavit-2 | Ofek-16	StatusActive	NaN	Su
14	14	CASC	Site 9401 (SLS-2), Jiuquan Satellite Launch Ce...	Sat Jul 04, 2020 23:44 UTC	Long March 2D | Shiyan-6 02	StatusActive	29.75	Su
15	15	Rocket Lab	Rocket Lab LC-1A, M?hia Peninsula, New Zealand	Sat Jul 04, 2020 21:19 UTC	Electron/Curie | Pics Or It Didn??¦t Happen	StatusActive	7.5	F
16	16	CASC	LC-9, Taiyuan Satellite Launch Center, China	Fri Jul 03, 2020 03:10 UTC	Long March 4B | Gaofen Duomo & BY-02	StatusActive	64.68	Su
17	17	SpaceX	SLC-40, Cape Canaveral AFS, Florida, USA	Tue Jun 30, 2020 20:10 UTC	Falcon 9 Block 5 | GPS III SV03	StatusActive	50.0	Su

	Unnamed: 0	Organisation	Location	Date	Detail	Rocket_Status	Price	Mission_S
18	18	CASC	LC-2, Xichang Satellite	Tue Jun 23, 2020	Long March 3B/E |	StatusActive	29.15	Su

Task2: Check if there is any null entries and check duplicate records.

In [11]:

```
Unnamed: 0           0
Organisation         0
Location             0
Date                 0
Detail               0
Rocket_Status        0
Price             3360
Mission_Status       0
dtype: int64
```

- It means, only Price column has null entries.

In [12]:

```
0
```

- Their are no duplicate value here

In [13]:

In [14]:

Out[14]:

	Unnamed: 0	Organisation	Location	Date	Detail	Rocket_Status	Price	Mission_Sta
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	Succ
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	Succ
2	2	SpaceX	Pad A, Boca Chica, Texas, USA	Tue Aug 04, 2020 23:57 UTC	Starship Prototype | 150 Meter Hop	StatusActive	0	Succ
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	Succ
4	4	ULA	SLC-41, Cape Canaveral AFS, Florida, USA	Thu Jul 30, 2020 11:50 UTC	Atlas V 541 | Perseverance	StatusActive	145	Succ

In [15]:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4324 entries, 0 to 4323
Data columns (total 8 columns):
 #   Column          Non-Null Count  Dtype 
---  ------          --------------  ----- 
 0   Unnamed: 0      4324 non-null   int64 
 1   Organisation    4324 non-null   object
 2   Location        4324 non-null   object
 3   Date            4324 non-null   object
 4   Detail          4324 non-null   object
 5   Rocket_Status   4324 non-null   object
 6   Price           4324 non-null   int32 
 7   Mission_Status  4324 non-null   object
dtypes: int32(1), int64(1), object(6)
memory usage: 253.5+ KB
```

In [16]:

Out[16]: (4324, 8)

Task3: Analysis of mission_status.

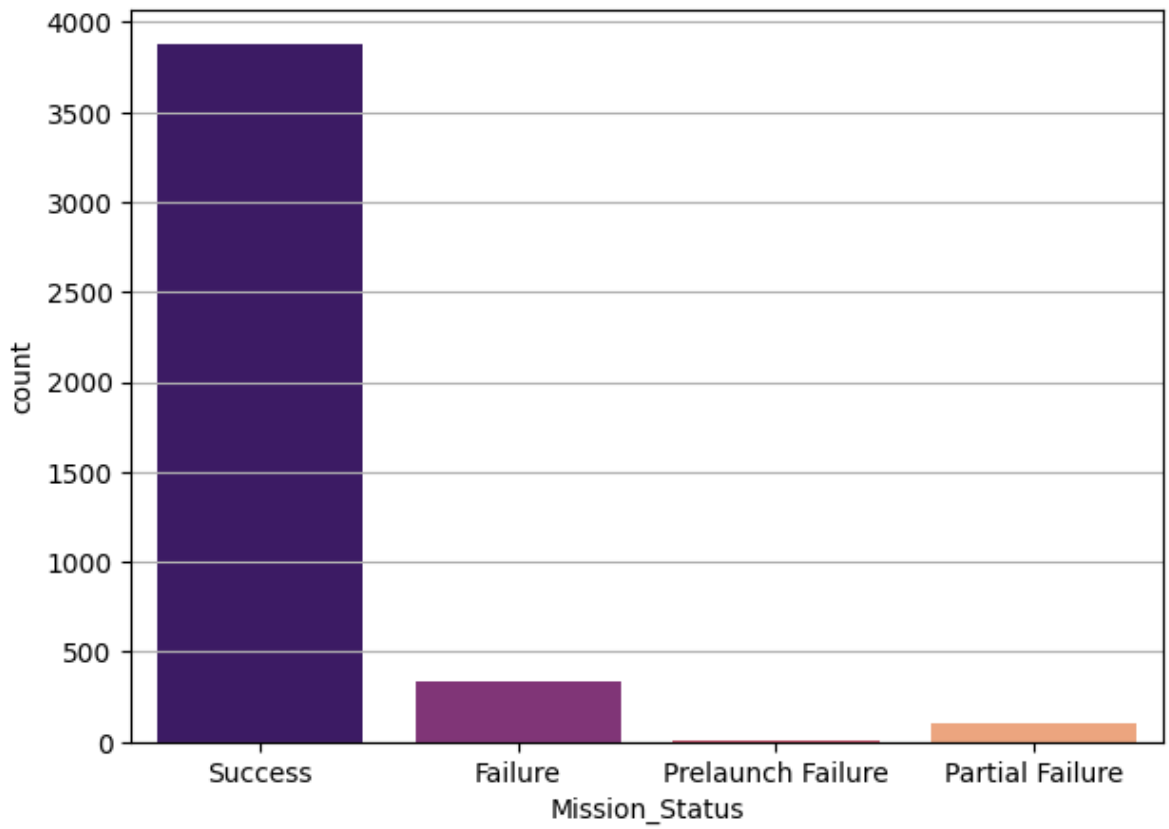
In [17]:

Out[17]:
```
array(['Success', 'Failure', 'Prelaunch Failure', 'Partial Failure'],
      dtype=object)
```

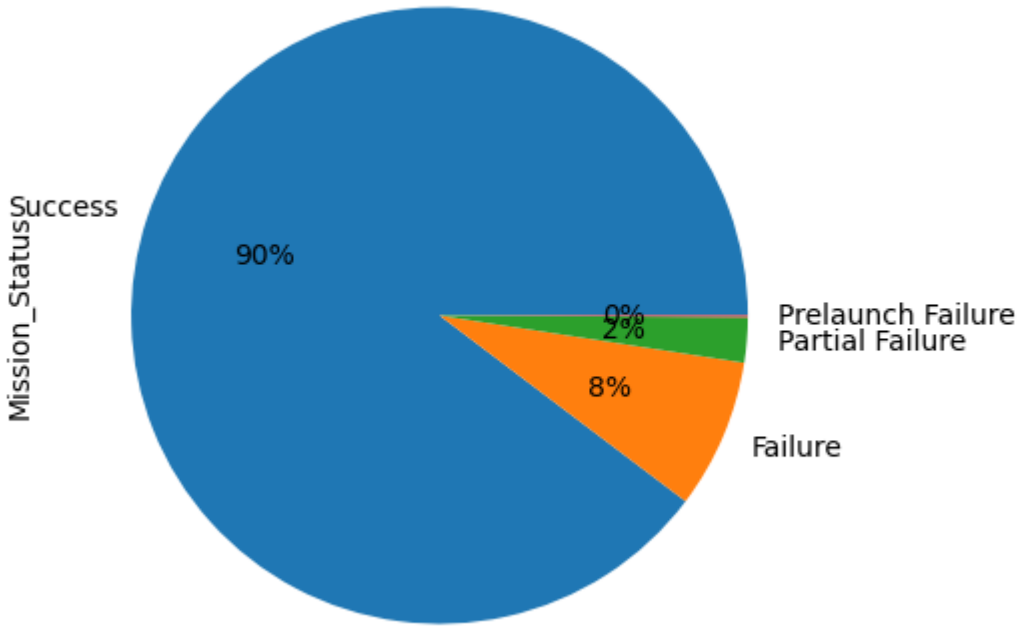
In [18]:

Out[18]:
```
Success              3879
Failure               339
Partial Failure       102
Prelaunch Failure       4
Name: Mission_Status, dtype: int64
```

In [19]:
```
plt.figure(figsize=(7,5))
sns.countplot(x="Mission_Status" ,data=df , palette="magma")
plt.xlabel=("Mission_Status")
plt.ylabel=("count")
plt.grid(axis='y')
```



In [20]:
```
plt.figure(figsize=(7,5))
df["Mission_Status"].value_counts().plot.pie(autopct='%.0f%%')
plt.xticks(rotation=90)
```



- Therefore, it means 3878 missions were successfull and rest are failure.

Task4: Analysis of Rocket_Status.

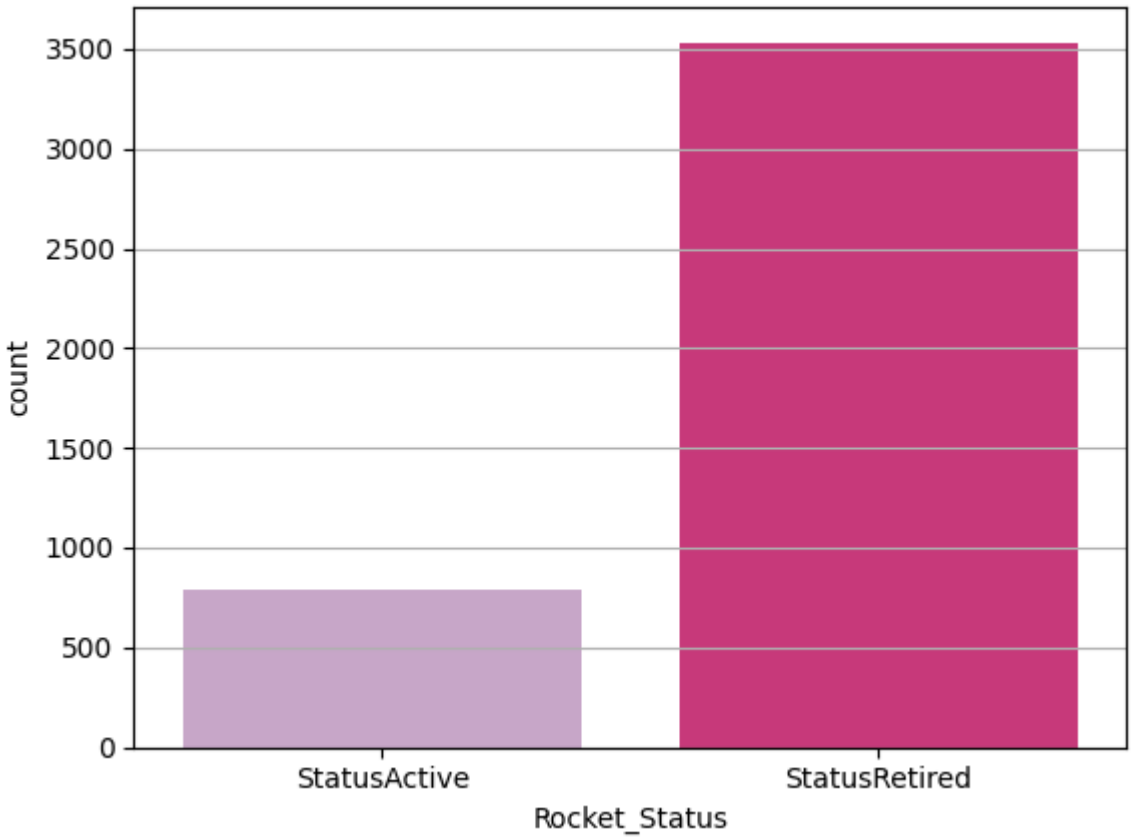
In [21]:
```
df["Rocket_Status"].unique()
```

Out[21]:
```
array(['StatusActive', 'StatusRetired'], dtype=object)
```

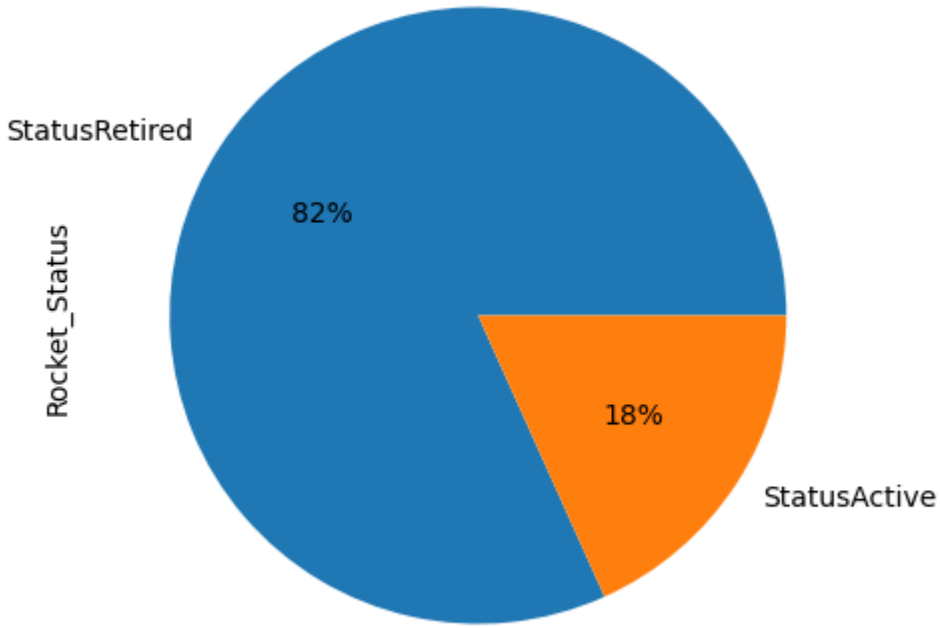
In [22]:
```
df["Rocket_Status"].value_counts()
```

Out[22]:
```
StatusRetired    3534
StatusActive      790
Name: Rocket_Status, dtype: int64
```

In [23]:
```
sns.countplot(x="Rocket_Status",data=df ,palette='PuRd')
```



In [24]:
```
plt.figure(figsize=(7,5))
df["Rocket_Status"].value_counts().plot.pie(autopct='%.0f%%')
plt.xticks(rotation=90)
```



- 790 Rocket were Active.
- 3534 rockets are retired

Task5: Analysis of Organisation.

In [25]:

Out[25]: 56

In [26]:

Out[26]:

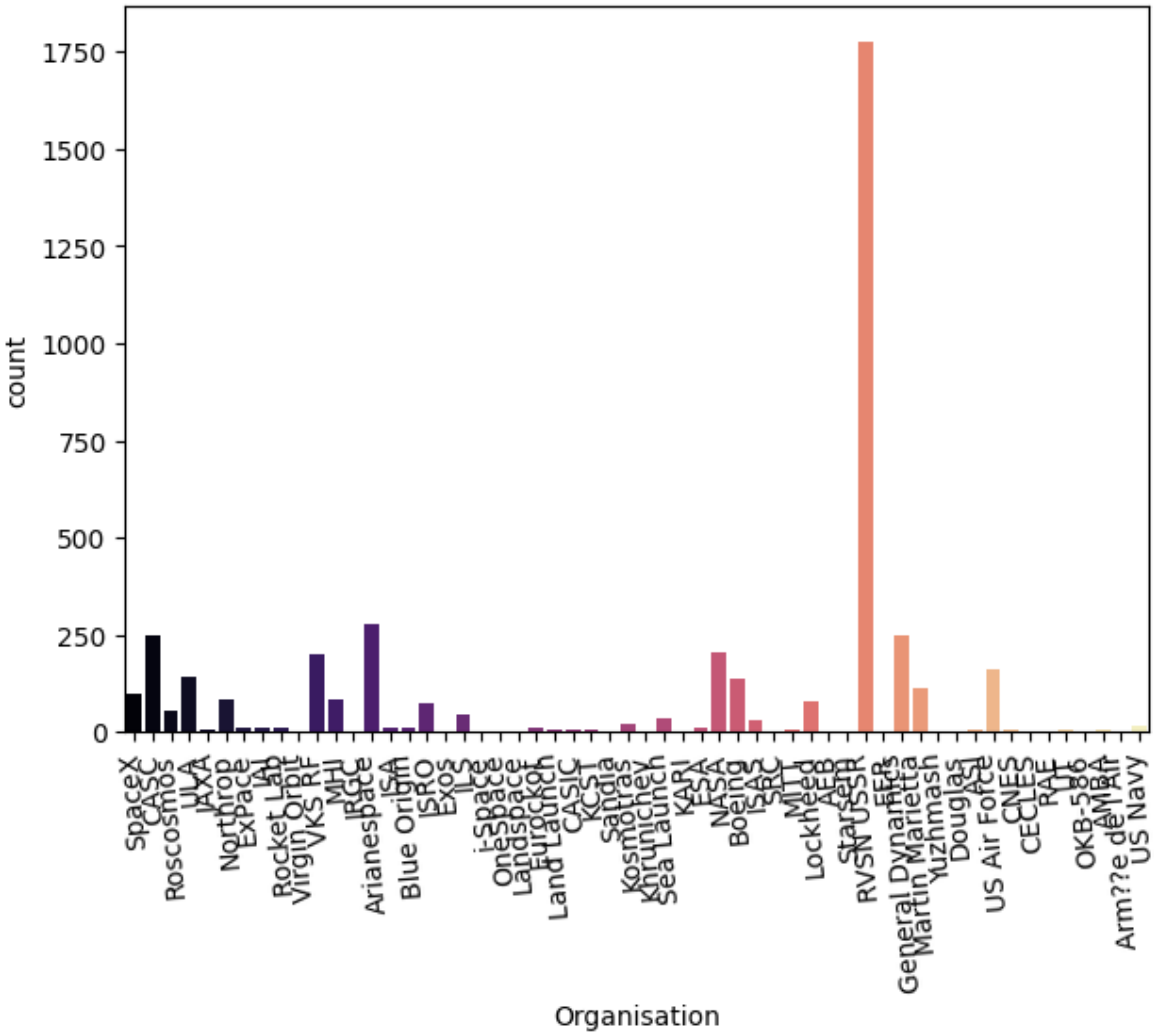
RVSN USSR	1777
Arianespace	279
CASC	251
General Dynamics	251
NASA	203
VKS RF	201
US Air Force	161
ULA	140
Boeing	136
Martin Marietta	114
SpaceX	100
MHI	84
Northrop	83
Lockheed	79
ISRO	76
Roscosmos	55
ILS	46
Sea Launch	36
ISAS	30
Kosmotras	22
US Navy	17
ISA	13
Rocket Lab	13
Eurockot	13
ESA	13
Blue Origin	12
IAI	11
ExPace	10
ASI	9
CNES	8
AMBA	8
MITT	7
JAXA	7
Land Launch	7
UT	5
KCST	5
CASIC	5
Exos	4
CECLES	4
Arm??e de l'Air	4
KARI	3
SRC	3
AEB	3
RAE	2
OKB-586	2
Yuzhmash	2
Landspace	1
Douglas	1
EER	1
Starsem	1
Virgin Orbit	1
IRGC	1
i-Space	1
OneSpace	1

```
Sandia                1
Khrunichev            1
Name: Organisation, dtype: int64
```

In [27]:
```
plt.figure(figsize=(7,5))
sns.countplot(x="Organisation" , data=df ,palette='magma')
```

Out[27]:
```
(array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
        17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33,
        34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50,
        51, 52, 53, 54, 55]),
 [Text(0, 0, 'SpaceX'),
  Text(1, 0, 'CASC'),
  Text(2, 0, 'Roscosmos'),
  Text(3, 0, 'ULA'),
  Text(4, 0, 'JAXA'),
  Text(5, 0, 'Northrop'),
  Text(6, 0, 'ExPace'),
  Text(7, 0, 'IAI'),
  Text(8, 0, 'Rocket Lab'),
  Text(9, 0, 'Virgin Orbit'),
  Text(10, 0, 'VKS RF'),
  Text(11, 0, 'MHI'),
  Text(12, 0, 'IRGC'),
  Text(13, 0, 'Arianespace'),
  Text(14, 0, 'ISA'),
  Text(15, 0, 'Blue Origin'),
  Text(16, 0, 'ISRO'),
  Text(17, 0, 'Exos'),
  Text(18, 0, 'ILS'),
  Text(19, 0, 'i-Space'),
  Text(20, 0, 'OneSpace'),
  Text(21, 0, 'Landspace'),
  Text(22, 0, 'Eurockot'),
  Text(23, 0, 'Land Launch'),
  Text(24, 0, 'CASIC'),
  Text(25, 0, 'KCST'),
  Text(26, 0, 'Sandia'),
  Text(27, 0, 'Kosmotras'),
  Text(28, 0, 'Khrunichev'),
  Text(29, 0, 'Sea Launch'),
  Text(30, 0, 'KARI'),
  Text(31, 0, 'ESA'),
  Text(32, 0, 'NASA'),
  Text(33, 0, 'Boeing'),
  Text(34, 0, 'ISAS'),
  Text(35, 0, 'SRC'),
  Text(36, 0, 'MITT'),
  Text(37, 0, 'Lockheed'),
  Text(38, 0, 'AEB'),
  Text(39, 0, 'Starsem'),
  Text(40, 0, 'RVSN USSR'),
  Text(41, 0, 'EER'),
  Text(42, 0, 'General Dynamics'),
  Text(43, 0, 'Martin Marietta'),
  Text(44, 0, 'Yuzhmash'),
  Text(45, 0, 'Douglas'),
  Text(46, 0, 'ASI'),
  Text(47, 0, 'US Air Force'),
  Text(48, 0, 'CNES'),
```

```
Text(49, 0, 'CECLES'),
Text(50, 0, 'RAE'),
Text(51, 0, 'UT'),
Text(52, 0, 'OKB-586'),
Text(53, 0, 'AMBA'),
Text(54, 0, "Arm??e de l'Air"),
Text(55, 0, 'US Navy')])
```



In [28]:

Out[28]:

	Unnamed: 0	Organisation	Location	Date	Detail	Rocket_Status	Price	Mission_Sta
4319	4319	US Navy	LC-18A, Cape Canaveral AFS, Florida, USA	Wed Feb 05, 1958 07:33 UTC	Vanguard | Vanguard TV3BU	StatusRetired	0	Fail
4320	4320	AMBA	LC-26A, Cape Canaveral AFS, Florida, USA	Sat Feb 01, 1958 03:48 UTC	Juno I | Explorer 1	StatusRetired	0	Succ
4321	4321	US Navy	LC-18A, Cape Canaveral AFS, Florida, USA	Fri Dec 06, 1957 16:44 UTC	Vanguard | Vanguard TV3	StatusRetired	0	Fail
4322	4322	RVSN USSR	Site 1/5, Baikonur Cosmodrome, Kazakhstan	Sun Nov 03, 1957 02:30 UTC	Sputnik 8K71PS | Sputnik-2	StatusRetired	0	Succ
4323	4323	RVSN USSR	Site 1/5, Baikonur Cosmodrome, Kazakhstan	Fri Oct 04, 1957 19:28 UTC	Sputnik 8K71PS | Sputnik-1	StatusRetired	0	Succ

- It means, 56 organisations launches 4323 rockets.

Task 5.1: Find which organisation have highest number of rocket launches.

In [29]:

Out[29]:

RVSN USSR	1777
Arianespace	279
CASC	251
General Dynamics	251
NASA	203
VKS RF	201
US Air Force	161
ULA	140
Boeing	136
Martin Marietta	114
SpaceX	100
MHI	84
Northrop	83
Lockheed	79
ISRO	76
Roscosmos	55
ILS	46
Sea Launch	36
ISAS	30
Kosmotras	22
US Navy	17
ISA	13
Rocket Lab	13
Eurockot	13
ESA	13
Blue Origin	12
IAI	11
ExPace	10
ASI	9
CNES	8
AMBA	8
MITT	7
JAXA	7
Land Launch	7
UT	5
KCST	5
CASIC	5
Exos	4
CECLES	4
Arm??e de l'Air	4
KARI	3
SRC	3
AEB	3
RAE	2
OKB-586	2
Yuzhmash	2
Landspace	1
Douglas	1
EER	1
Starsem	1
Virgin Orbit	1
IRGC	1
i-Space	1
OneSpace	1

```
Sandia                  1
Khrunichev              1
Name: Organisation, dtype: int64
```

In [30]:
```
max_launches=df.groupby('Organisation')['Rocket_Status'].count().idxmax()
```

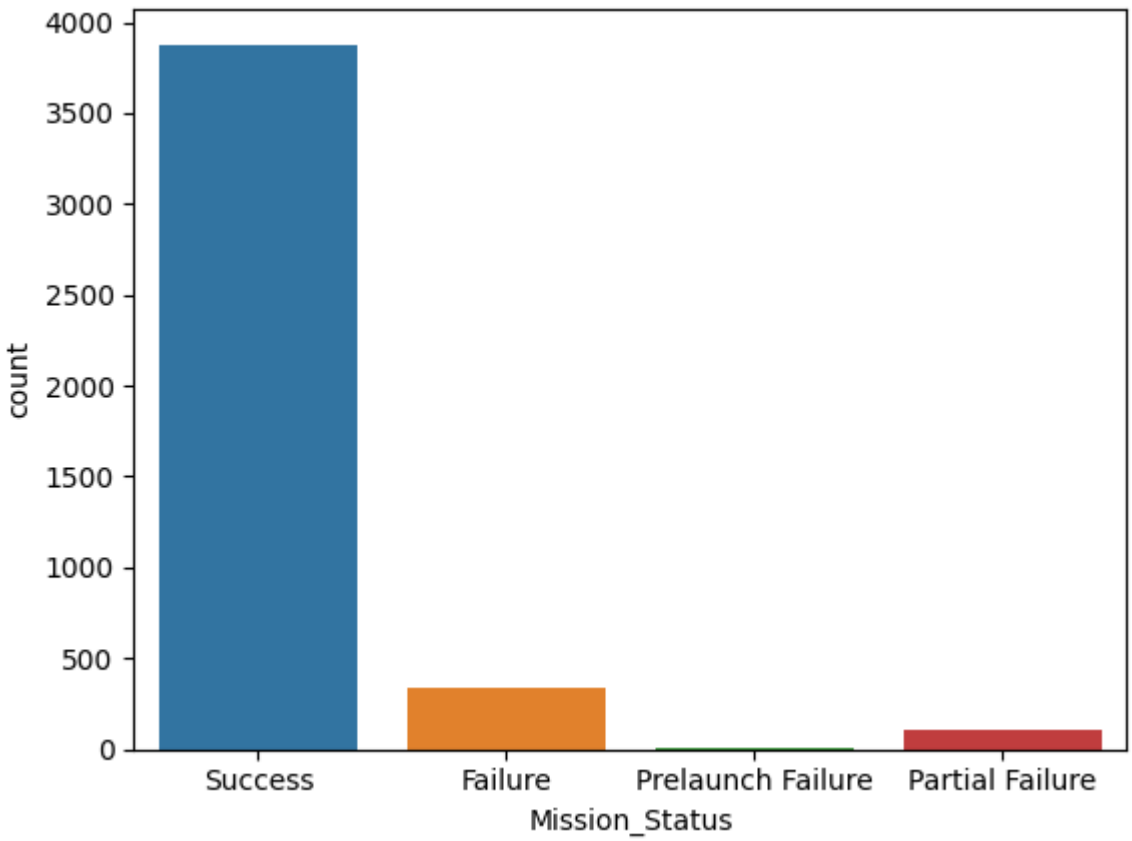
```
The organisation with the highest number of rocket launches is: RVSN USSR
```

- The organisation with the highest number of rocket launches is: RVSN USSR which is highest number-1777
- It means RVSN USSR launches 1777 rocket launches.

Task 5.2: Find number of successfull Launches and Failure(Failure+PreLaunch Failure+Partial Failure) Launches for each organisation.

In [31]:

Out[31]: <AxesSubplot: xlabel='Mission_Status', ylabel='count'>



In [32]:

```
Out[32]: Success              3879
         Failure               339
         Partial Failure       102
         Prelaunch Failure       4
         Name: Mission_Status, dtype: int64
```

- It means, 3879 are successful Launches
- 339 are failure launches
- 102 partial failure
- 4 prelaunch failure

Task6: Analysis of Date Column.

In [33]:

Out[33]:

	Unnamed: 0	Organisation	Location	Date	Detail	Rocket_Status	Price	Mission_Sta
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	Succ
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	Succ
2	2	SpaceX	Pad A, Boca Chica, Texas, USA	Tue Aug 04, 2020 23:57 UTC	Starship Prototype | 150 Meter Hop	StatusActive	0	Succ
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	Succ
4	4	ULA	SLC-41, Cape Canaveral AFS, Florida, USA	Thu Jul 30, 2020 11:50 UTC	Atlas V 541 | Perseverance	StatusActive	145	Succ

In [34]:

```
Out[34]: array(['Fri Aug 07, 2020 05:12 UTC', 'Thu Aug 06, 2020 04:01 UTC',
                'Tue Aug 04, 2020 23:57 UTC', ..., 'Fri Dec 06, 1957 16:44 UTC',
                'Sun Nov 03, 1957 02:30 UTC', 'Fri Oct 04, 1957 19:28 UTC'],
               dtype=object)
```

In [35]:

```
Out[35]: Wed Nov 05, 2008 00:15 UTC    2
         Sun Aug 25, 1991 08:40 UTC    2
         Tue Aug 28, 1990 09:05 UTC    2
         Wed Feb 07, 1990 01:33 UTC    2
         Tue Jun 26, 1973              2
                                      ..
         Thu May 16, 1996 01:56 UTC    1
         Sun May 12, 1996 21:32 UTC    1
         Tue Apr 30, 1996 04:31 UTC    1
         Wed Apr 24, 1996 23:37 UTC    1
         Fri Oct 04, 1957 19:28 UTC    1
         Name: Date, Length: 4319, dtype: int64
```

In [36]:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4324 entries, 0 to 4323
Data columns (total 8 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Unnamed: 0      4324 non-null   int64
 1   Organisation    4324 non-null   object
 2   Location        4324 non-null   object
 3   Date            4324 non-null   object
 4   Detail          4324 non-null   object
 5   Rocket_Status   4324 non-null   object
 6   Price           4324 non-null   int32
 7   Mission_Status  4324 non-null   object
dtypes: int32(1), int64(1), object(6)
memory usage: 253.5+ KB
```

- u can see Date column datatype is object

In [37]:

In [38]:

Out[38]:

	Unnamed: 0	Organisation	Location	Date	Detail	Rocket_Status	Price	Mi
0	0	SpaceX	LC-39A, Kennedy Space Center, Florida, USA	2020-08-07 05:12:00+00:00	Falcon 9 Block 5 | Starlink V1 L9 & BlackSky	StatusActive	50	
1	1	CASC	Site 9401 (SLS-2), Jiuquan Satellite Launch Ce...	2020-08-06 04:01:00+00:00	Long March 2D | Gaofen-9 04 & Q-SAT	StatusActive	29	
2	2	SpaceX	Pad A, Boca Chica, Texas, USA	2020-08-04 23:57:00+00:00	Starship Prototype | 150 Meter Hop	StatusActive	0	
3	3	Roscosmos	Site 200/39, Baikonur Cosmodrome, Kazakhstan	2020-07-30 21:25:00+00:00	Proton-M/Briz-M | Ekspress-80 & Ekspress-103	StatusActive	65	
4	4	ULA	SLC-41, Cape Canaveral AFS, Florida, USA	2020-07-30 11:50:00+00:00	Atlas V 541 | Perseverance	StatusActive	145	

In [39]:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4324 entries, 0 to 4323
Data columns (total 8 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Unnamed: 0      4324 non-null   int64
 1   Organisation    4324 non-null   object
 2   Location        4324 non-null   object
 3   Date            4324 non-null   datetime64[ns, UTC]
 4   Detail          4324 non-null   object
 5   Rocket_Status   4324 non-null   object
 6   Price           4324 non-null   int32
 7   Mission_Status  4324 non-null   object
dtypes: datetime64[ns, UTC](1), int32(1), int64(1), object(5)
memory usage: 253.5+ KB
```

- Now you can see date column its turn into datetime64[ns, UTC]

* Task 6.1: Create new column(launch_year) which contain year of corresponding records

only

In [40]:

In [41]:

Out[41]:

	Unnamed: 0	Organisation	Location	Date	Detail	Rocket_Status	Price	Mi
0	0	SpaceX	LC-39A, Kennedy Space Center, Florida, USA	2020-08-07 05:12:00+00:00	Falcon 9 Block 5 | Starlink V1 L9 & BlackSky	StatusActive	50	
1	1	CASC	Site 9401 (SLS-2), Jiuquan Satellite Launch Ce...	2020-08-06 04:01:00+00:00	Long March 2D | Gaofen-9 04 & Q-SAT	StatusActive	29	
2	2	SpaceX	Pad A, Boca Chica, Texas, USA	2020-08-04 23:57:00+00:00	Starship Prototype | 150 Meter Hop	StatusActive	0	
3	3	Roscosmos	Site 200/39, Baikonur Cosmodrome, Kazakhstan	2020-07-30 21:25:00+00:00	Proton-M/Briz-M | Ekspress-80 & Ekspress-103	StatusActive	65	
4	4	ULA	SLC-41, Cape Canaveral AFS, Florida, USA	2020-07-30 11:50:00+00:00	Atlas V 541 | Perseverance	StatusActive	145	

In [42]:

Out[42]: (4324, 9)

- You can see we create new column that is launch_year which contain year of corresponding records only

Task 6.2: Find number of launched rockets year-wise.

In [43]:
```
df.groupby('launch_year').agg(Total_rockets_lauch_year_wise = ('Organisation',
```

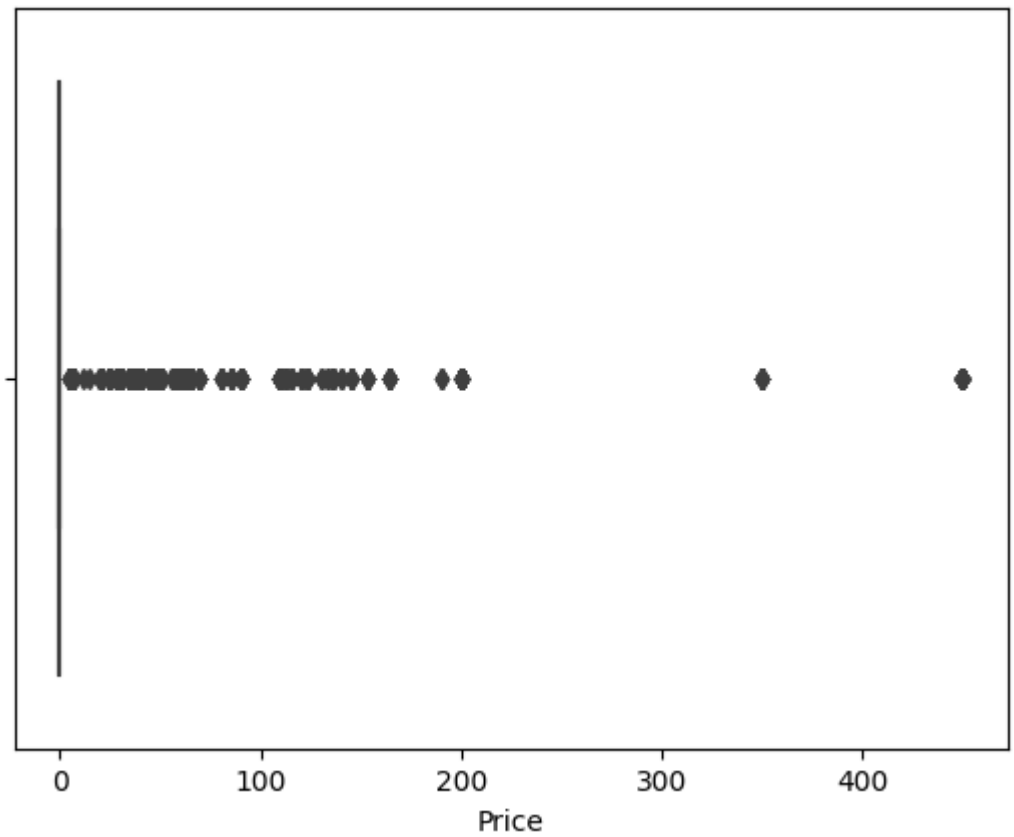
Out[43]:

	launch_year	Total_rockets_lauch_year_wise
0	1957	3
1	1958	28
2	1959	20
3	1960	39
4	1961	52
...	...	...
59	2016	90
60	2017	92
61	2018	117
62	2019	109
63	2020	63

64 rows × 2 columns

In [44]:

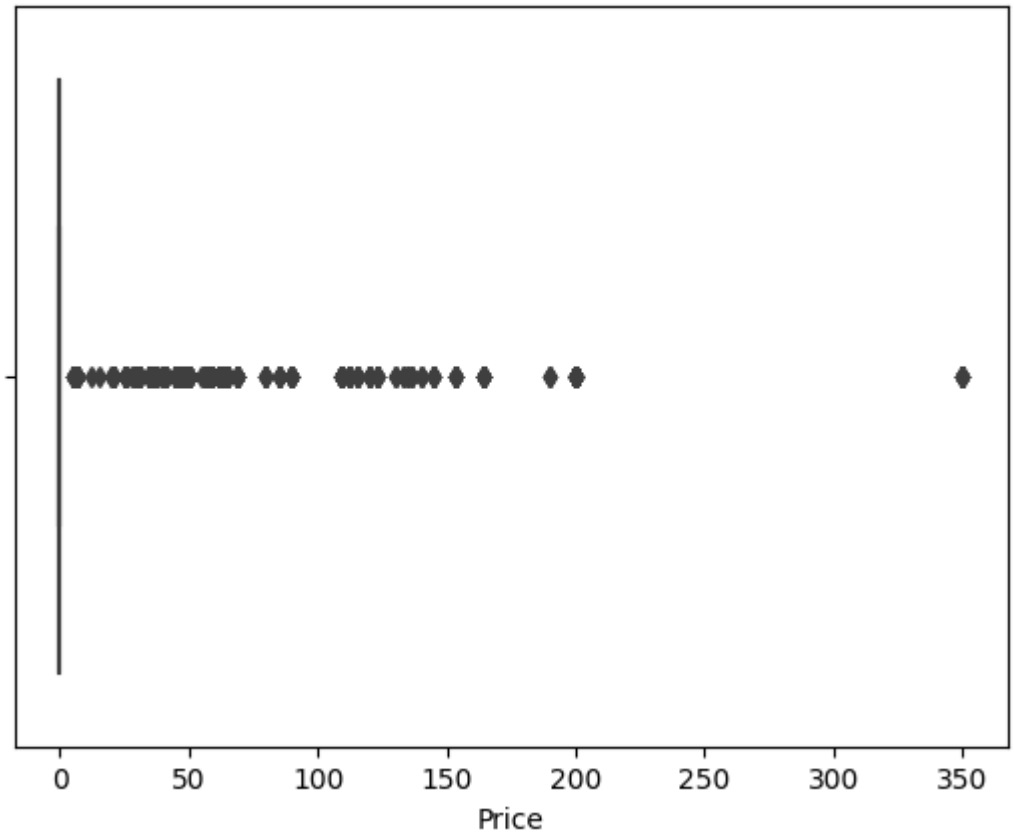
Out[44]: <AxesSubplot: xlabel='Price'>



In [45]:

In [46]:

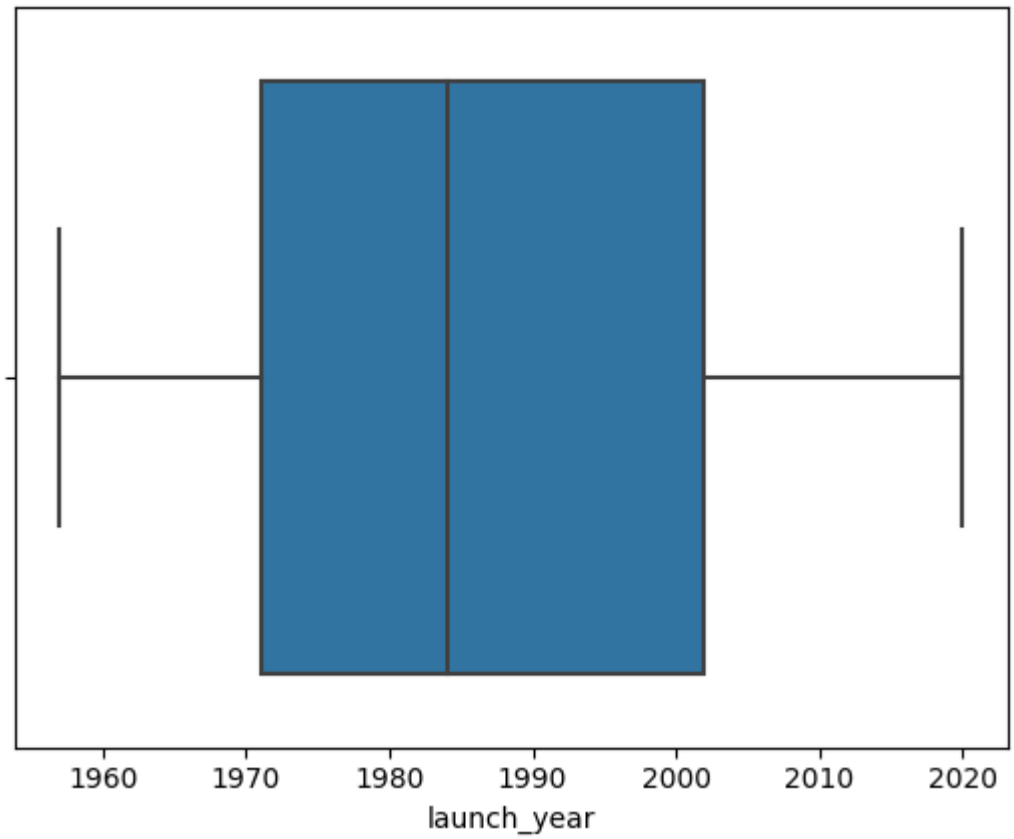
Out[46]: <AxesSubplot: xlabel='Price'>



- outlier removed

In [47]:

Out[47]: <AxesSubplot: xlabel='launch_year'>



In [48]:

Out[48]: (4188, 9)

In [49]:

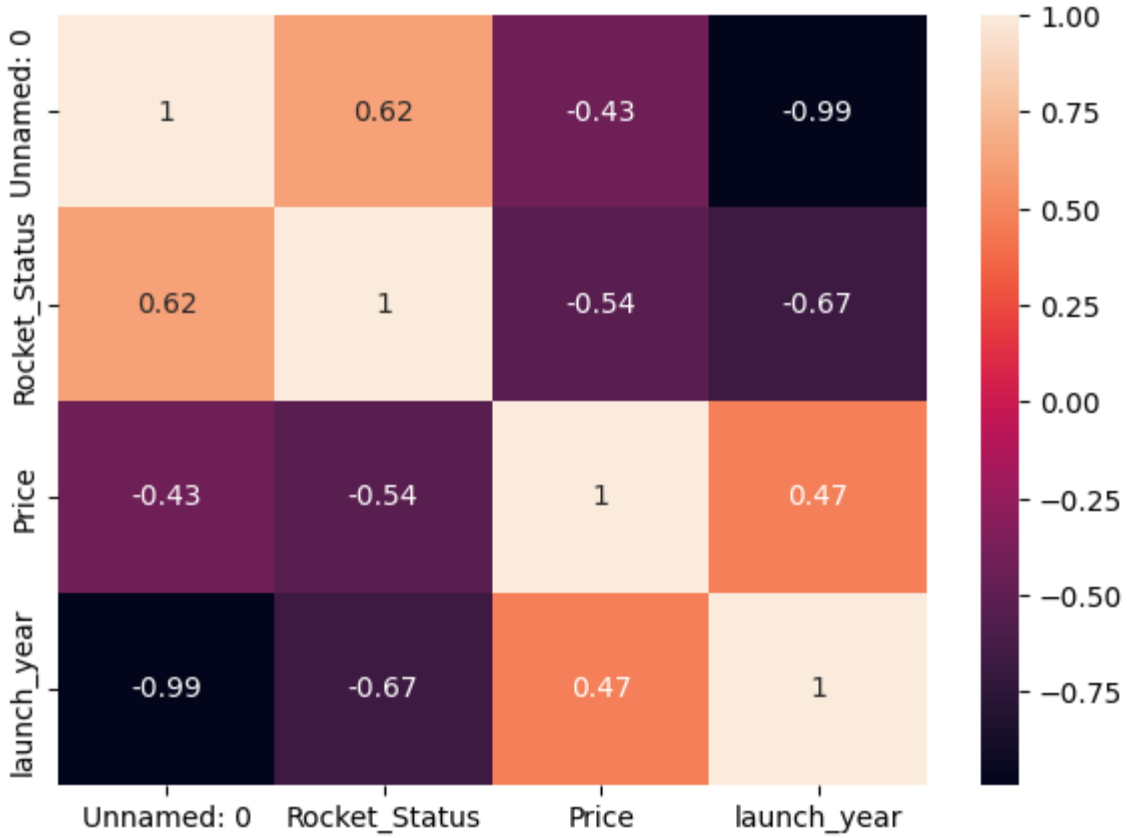
In [50]:

Out[50]:

	Unnamed: 0	Price	launch_year
Unnamed: 0	1.000000	-0.432811	-0.991237
Price	-0.432811	1.000000	0.469879
launch_year	-0.991237	0.469879	1.000000

In [95]:
```
plt.figure(figsize=(7,5))
```

Out[95]:
```
<AxesSubplot: >
```



Feature Engineering

- we used here one hot encoding.
- Because there are multiple independent column with categorical value.

In [52]:
```
df.columns
```

Out[52]:
```
Index(['Unnamed: 0', 'Location', 'Rocket_Status', 'Price', 'Mission_Status',
       'launch_year'],
      dtype='object')
```

In [53]:
```
# convert all categorical column into numeric form (0,1)


col=['Mission_Status']
dataset=pd.get_dummies(df[col],dtype=int,drop_first=True)
```

Out[53]:

	Mission_Status_Partial Failure	Mission_Status_Prelaunch Failure	Mission_Status_Success
0	0	0	1
1	0	0	1
2	0	0	1
3	0	0	1
4	0	0	1
...	...	...	...
4319	0	0	0
4320	0	0	1
4321	0	0	0
4322	0	0	1
4323	0	0	1

4188 rows × 3 columns

We use lable encoding here for output column because output has a categorical

yes=1 No=0

In [54]:
```
from sklearn.preprocessing import LabelEncoder
label_encoder = LabelEncoder()
```

In [55]:

Out[55]:
```
array(['Success', 'Failure', 'Prelaunch Failure', 'Partial Failure'],
      dtype=object)
```

In [56]:

Out[56]:

	Unnamed: 0	Location	Rocket_Status	Price	Mission_Status	launch_year
0	0	LC-39A, Kennedy Space Center, Florida, USA	0	50	Success	2020
1	1	Site 9401 (SLS-2), Jiuquan Satellite Launch Ce...	0	29	Success	2020
2	2	Pad A, Boca Chica, Texas, USA	0	0	Success	2020
3	3	Site 200/39, Baikonur Cosmodrome, Kazakhstan	0	65	Success	2020
4	4	SLC-41, Cape Canaveral AFS, Florida, USA	0	145	Success	2020
...	...	...	...	...	...	...
4319	4319	LC-18A, Cape Canaveral AFS, Florida, USA	1	0	Failure	1958
4320	4320	LC-26A, Cape Canaveral AFS, Florida, USA	1	0	Success	1958
4321	4321	LC-18A, Cape Canaveral AFS, Florida, USA	1	0	Failure	1957
4322	4322	Site 1/5, Baikonur Cosmodrome, Kazakhstan	1	0	Success	1957
4323	4323	Site 1/5, Baikonur Cosmodrome, Kazakhstan	1	0	Success	1957

4188 rows × 6 columns

Merge two dataset first is Dataset and second is le (lable encoder)

In [57]:

In [58]:

Out[58]:

	Mission_Status_Partial Failure	Mission_Status_Prelaunch Failure	Mission_Status_Success	Rocket_Status
0	0	0	1	0
1	0	0	1	0
2	0	0	1	0
3	0	0	1	0
4	0	0	1	0

Using StandardScaler

In [59]:
```
from sklearn.preprocessing import StandardScaler
scaler=StandardScaler()
X_std = scaler.fit_transform(dataset)
```

Out[59]:
```
array([[-0.15799775, -0.03091962,  0.34393467],
       [-0.15799775, -0.03091962,  0.34393467],
       [-0.15799775, -0.03091962,  0.34393467],
       ...,
       [-0.15799775, -0.03091962, -2.90752895],
       [-0.15799775, -0.03091962,  0.34393467],
       [-0.15799775, -0.03091962,  0.34393467]])
```

Using MinMaxScaler for better Accuracy and range of MinMaxScaler is 0 to 1.

In [81]:
```
from sklearn.preprocessing import MinMaxScaler
mxsc=MinMaxScaler()
X_min = mxsc.fit_transform(dataset)
```

Out[81]:
```
array([[0.00000000e+00, 0.00000000e+00, 1.42857143e-01, ...,
        0.00000000e+00, 0.00000000e+00, 1.00000000e+00],
       [2.31320842e-04, 0.00000000e+00, 8.28571429e-02, ...,
        0.00000000e+00, 0.00000000e+00, 1.00000000e+00],
       [4.62641684e-04, 0.00000000e+00, 0.00000000e+00, ...,
        0.00000000e+00, 0.00000000e+00, 1.00000000e+00],
       ...,
       [9.99537358e-01, 1.00000000e+00, 0.00000000e+00, ...,
        0.00000000e+00, 0.00000000e+00, 0.00000000e+00],
       [9.99768679e-01, 1.00000000e+00, 0.00000000e+00, ...,
        0.00000000e+00, 0.00000000e+00, 1.00000000e+00],
       [1.00000000e+00, 1.00000000e+00, 0.00000000e+00, ...,
        0.00000000e+00, 0.00000000e+00, 1.00000000e+00]])
```

Spliting data into train and test set

In [82]:
```
X=df1
```

In [83]:

In [84]:

In [85]:

Out[85]:
```
(3350, 4)
```

In [86]:

Out[86]: (3350,)

In [87]:

Out[87]: (838, 4)

In [88]:

Out[88]: (838,)

Logistic Regression Algorithm

In [89]:

In [91]:

In [92]:

Out[92]:

▾ LogisticRegression

LogisticRegression()

In [93]:

Out[93]: 99.94029850746269

In [94]:

Out[94]: 100.0