

# Space-O-Analytics

Name : Aniruddha Joshi

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
## Display all the columns of the dataframe
```

In [2]:

```
dataset=pd.read_csv('dataset.csv')
print(dataset.shape)
```

(147636, 17)

In [3]:

dataset

Out[3]:

	source_id	mh_gspphot	distance_gspphot	radius_gspphot	vsini_esphs	rac
0	1892971994938385152	-0.2297	1913.2518	0.6903	NaN	
1	1892971994938390912	-1.1103	1414.8071	0.4954	NaN	
2	1892971999234781312	0.0661	1411.8239	0.7323	NaN	
3	1892972098017629696	-0.9274	1215.6301	0.5737	NaN	
4	1892972102313998080	-0.7981	1557.3232	0.6925	NaN	
...	...	...	...	...	...	...
147631	1897959173524272896	0.0041	1814.6660	0.7479	NaN	
147632	1897959173524280704	0.0436	1189.6232	0.6318	NaN	
147633	1897959207884022912	-0.6622	2112.5496	0.8272	NaN	
147634	1897959212180118400	-0.4128	1479.3003	1.4229	NaN	
147635	1897959349618945280	-0.0758	1979.1002	3.8875	NaN	

147636 rows × 17 columns

In [4]:

```
dataset=dataset.drop(['vsini_esphs', 'mass_flame', 'age_flame'], axis=1)
```

In [5]:

```
features_with_na=[features for features in dataset.columns if dataset[features].isnull().sum()>0]
features_with_na
```

Out[5]:

```
[]
```

In [6]:

```
dataset.isnull().sum()
```

Out[6]:

```
source_id          0
mh_gspphot         0
distance_gspphot   0
radius_gspphot     0
radius_flame       0
lum_flame          0
gravredshift_flame 0
parallax           0
l                  0
b                  0
ecl_lon            0
ecl_lat            0
teff_gspphot       0
logg_gspphot       0
dtype: int64
```

In [7]:

```
dataset.shape
```

Out[7]:

```
(147636, 14)
```

In [8]:

```
dataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 147636 entries, 0 to 147635
Data columns (total 14 columns):
#   Column                Non-Null Count  Dtype
---  -
0   source_id             147636 non-null  int64
1   mh_gspphot            147636 non-null  float64
2   distance_gspphot      147636 non-null  float64
3   radius_gspphot        147636 non-null  float64
4   radius_flame          147636 non-null  float64
5   lum_flame             147636 non-null  float64
6   gravredshift_flame    147636 non-null  float64
7   parallax              147636 non-null  float64
8   l                     147636 non-null  float64
9   b                     147636 non-null  float64
10  ecl_lon               147636 non-null  float64
11  ecl_lat               147636 non-null  float64
12  teff_gspphot          147636 non-null  float64
13  logg_gspphot          147636 non-null  float64
dtypes: float64(13), int64(1)
memory usage: 15.8 MB
```

In [9]:

```
dataset.describe()
```

Out[9]:

	source_id	mh_gspphot	distance_gspphot	radius_gspphot	radius_flame	lum
count	1.476360e+05	147636.000000	147636.000000	147636.000000	147636.000000	147636.0
mean	1.895640e+18	-0.547239	1479.805052	1.071527	1.074612	2.5
std	1.465586e+15	0.527875	853.941043	2.453623	1.839009	30.5
min	1.892972e+18	-4.094300	26.960800	0.164700	0.072414	0.0
25%	1.894377e+18	-0.852700	1008.858200	0.612600	0.623485	0.1
50%	1.895770e+18	-0.500350	1393.641200	0.735900	0.764071	0.5
75%	1.896933e+18	-0.163000	1804.278400	0.958200	1.019875	0.7
max	1.897959e+18	0.798800	20662.037000	383.998400	163.556800	3441.9



In [10]:

```

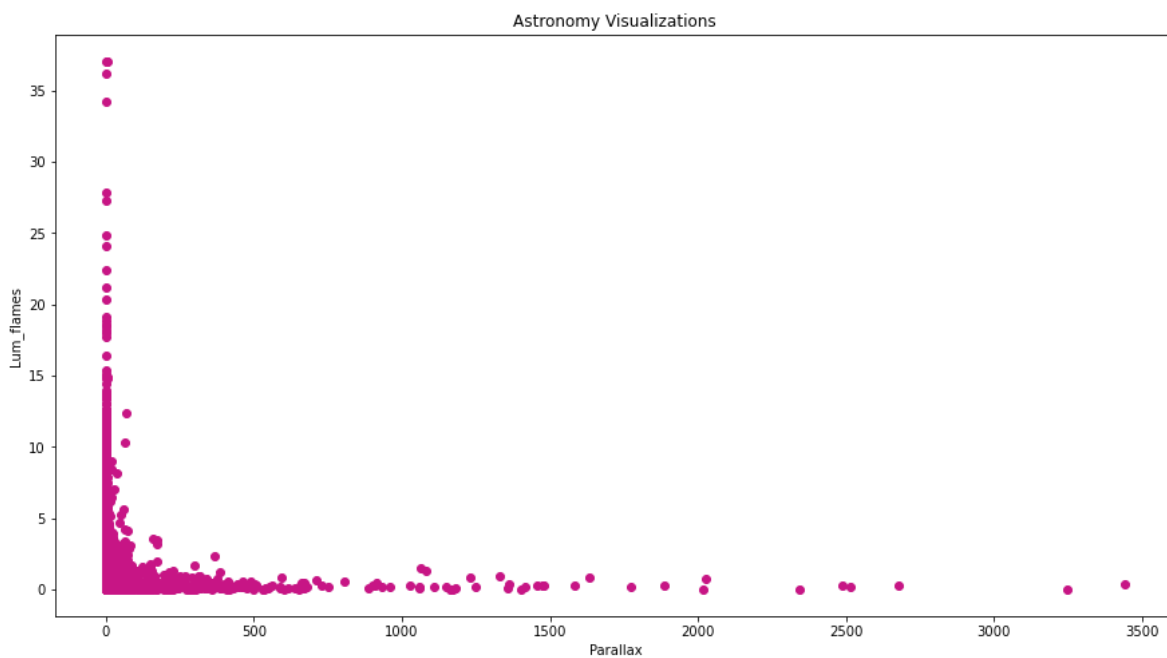
fig = plt.figure(figsize=(15,8))
plt.scatter(dataset.lum_flame, dataset.parallax,color='mediumvioletred')
plt.title("Astronomy Visualizations")
plt.ylabel('Lum_flames')
plt.xlabel('Parallax')

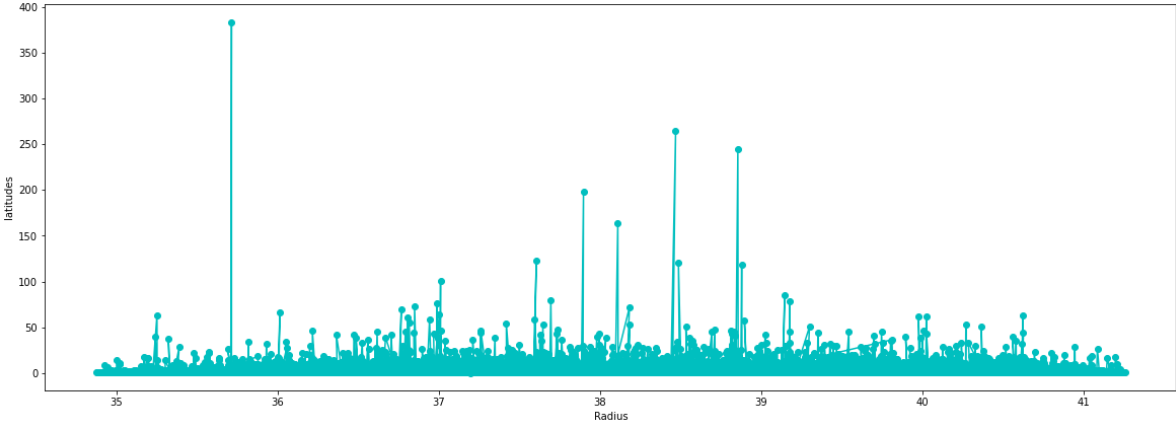
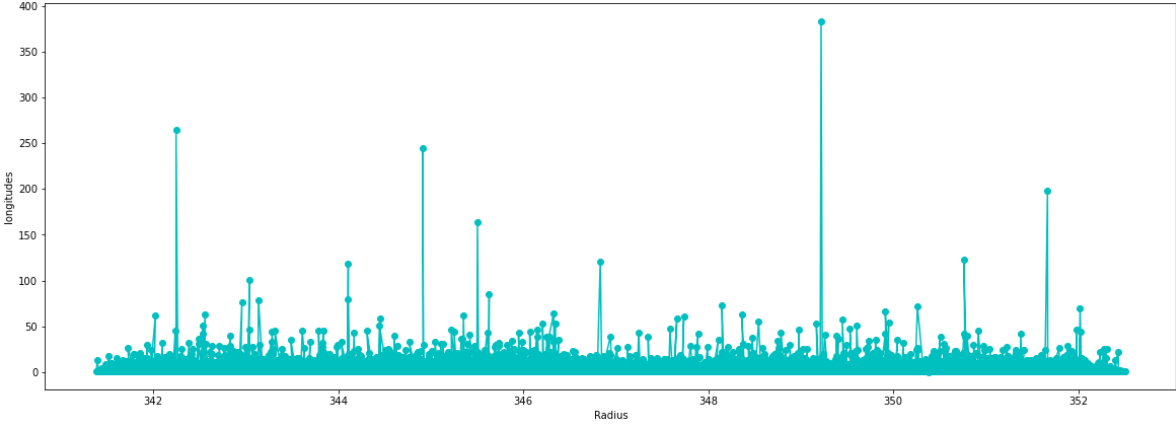
plt.figure(figsize=(20,7))
plt.plot(dataset.ecl_lon,dataset.radius_gspphot,marker='o', color='c')
# plt.bar(dataset.ecl_Lon,dataset.radius_gspphot)
plt.ylabel('longitudes')
plt.xlabel('Radius')

plt.figure(figsize=(20,7))
plt.plot(dataset.ecl_lat,dataset.radius_gspphot,marker='o', color='c')
# plt.bar(dataset.ecl_Lon,dataset.radius_gspphot)
plt.ylabel('latitudes')
plt.xlabel('Radius')
plt.show()

# plt.tight_layout()

```

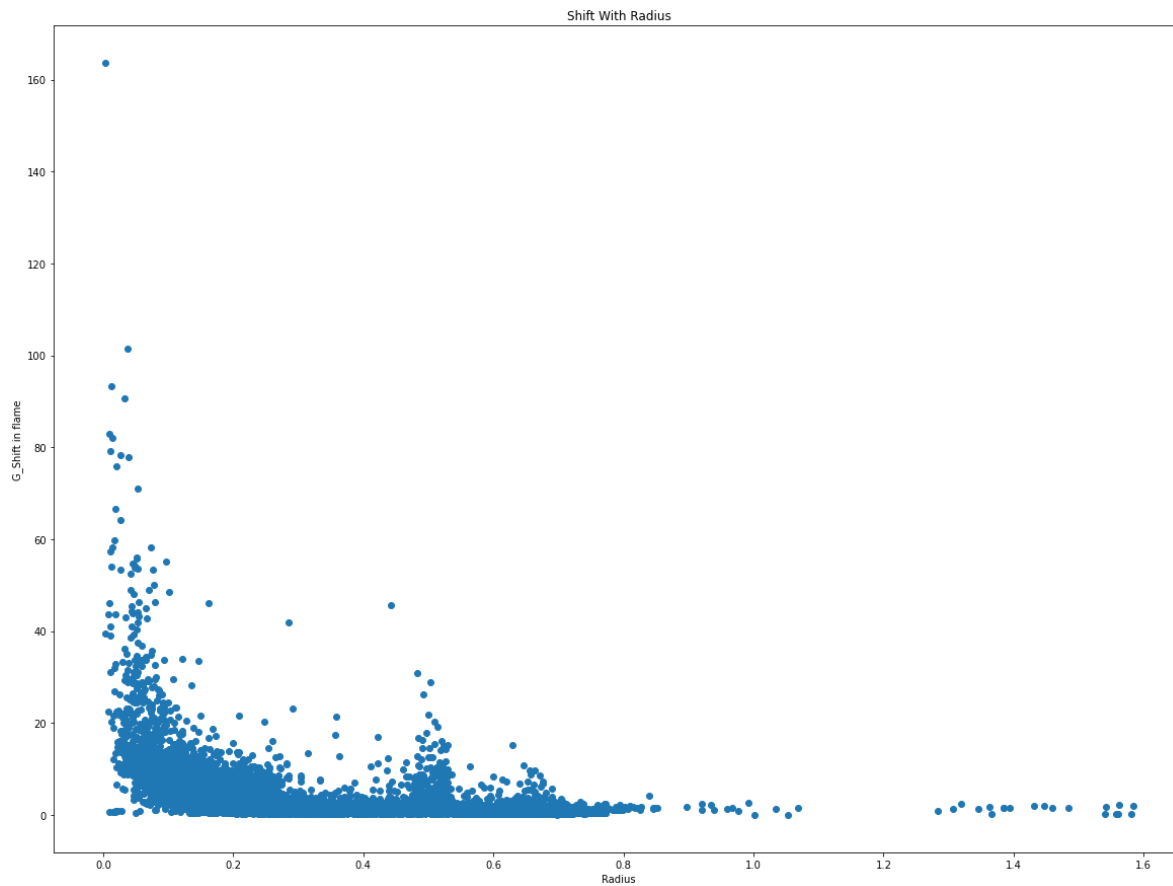




In [ ]:

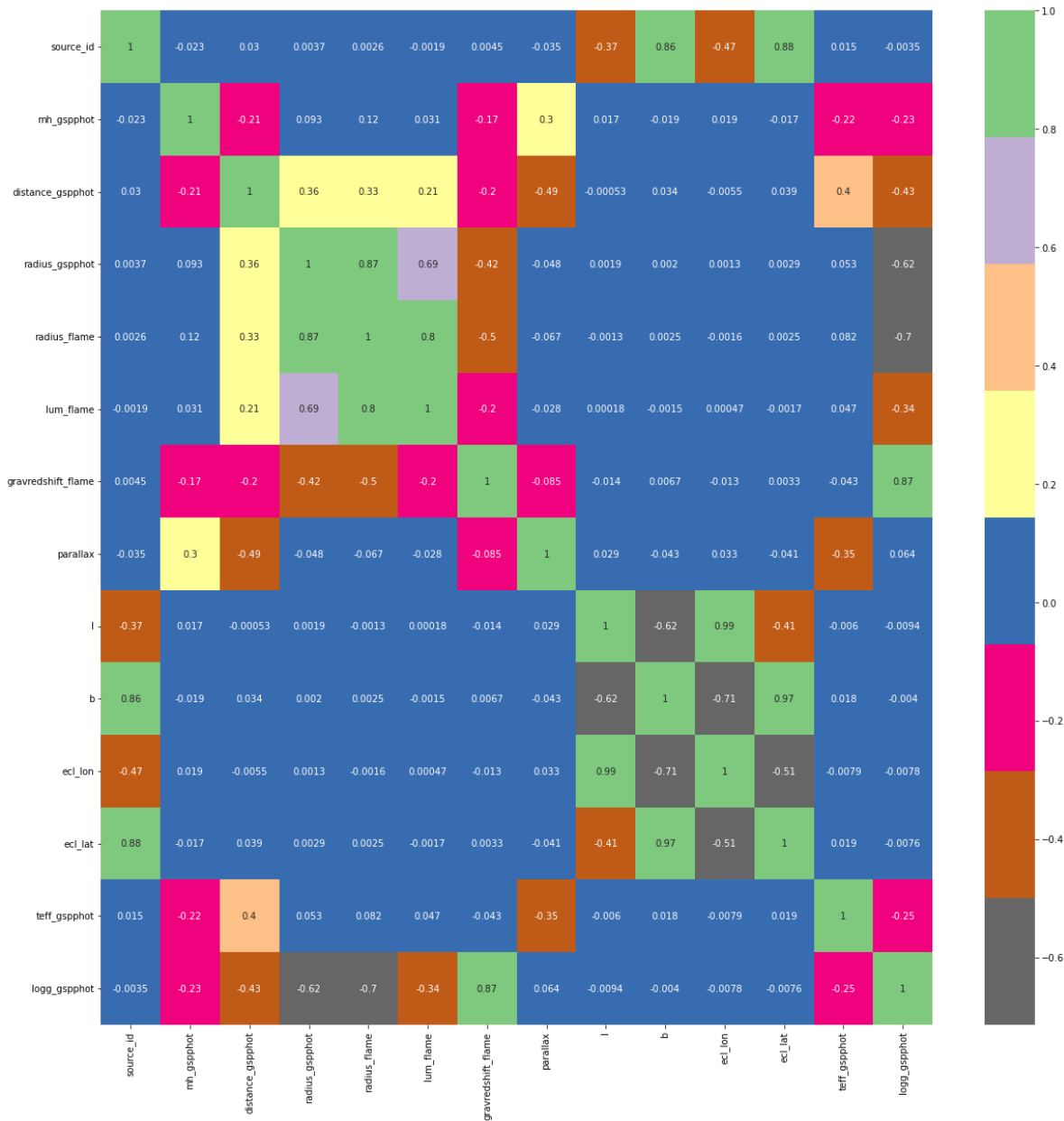
In [11]:

```
plt.figure(figsize=(20,15))  
plt.scatter(dataset.gravredshift_flame,dataset.radius_flame)  
plt.ylabel('G_Shift in flame')  
plt.xlabel('Radius')  
plt.title("Shift With Radius")  
plt.show()
```



In [12]:

```
import seaborn as sns
plt.figure(figsize=(20,20))
sns.heatmap(dataset.corr(),annot=True,cmap=plt.cm.Accent_r)
plt.show()
```

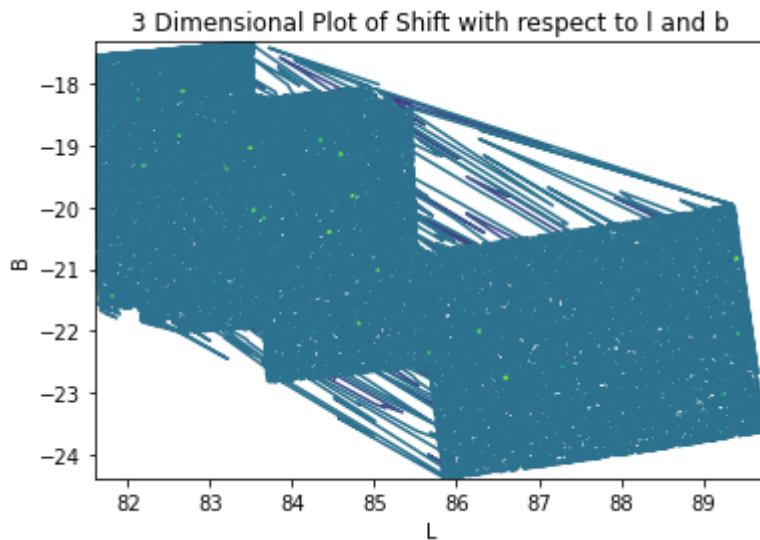


In [22]:

```
plt.title("3 Dimensional Plot of Shift with respect to l and b")  
# plt.figure(figsize=(15,9))  
# fig, ax = plt.subplots()  
plt.tricontour(dataset.l,dataset.b,dataset.gravredshift_flame)  
plt.xlabel("L")  
plt.ylabel("B")  
# plt.show()
```

Out[22]:

Text(0, 0.5, 'B')



**Conclusion:** EDA i.e. Exploratory Data Analysis is done on the dataset. Some features which had less correlation has been dropped from the analysis. Visualizations are done on a dataset. Correlation matrix is telling the correlations in between the features. Astronomy is really a vast field and the graphs are trying to tell the same.

**Thank You!**

*github:* <https://github.com/amj011/Space-O-Analytics> (<https://github.com/amj011/Space-O-Analytics>)

*linkedin:* <https://www.linkedin.com/in/aniruddha-joshi-5113b8213> (<https://www.linkedin.com/in/aniruddha-joshi-5113b8213>)

In [ ]: