In [5]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [6]:
```
df1=pd.read_csv('leads_basic_details.csv')
df2=pd.read_csv('leads_demo_watched_details.csv')
df3=pd.read_csv('leads_interaction_details.csv')
df4=pd.read_csv('leads_reasons_for_no_interest.csv')
df5=pd.read_csv('sales_managers_assigned_leads_details.csv')
```

In [7]:
```
df1
```

Out[7]:

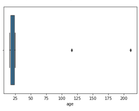
	lead_id	age	gender	current_city	current_education	parent_occupation	lead_gen_source
0	USR1001	16	FEMALE	Hyderabad	Intermediate	Private Employee	social_media
1	USR1002	20	MALE	Bengaluru	B.Tech	Business	user_referrals
2	USR1003	20	FEMALE	Visakhapatnam	B.Tech	Lawyer	user_referrals
3	USR1004	16	MALE	Mumbai	Intermediate	IT Employee	user_referrals
4	USR1005	16	MALE	Chennai	Intermediate	Government Employee	user_referrals
...	...	...	...	...	...	...	...
355	USR1356	21	MALE	Mumbai	Degree	Government Employee	user_referrals
356	USR1357	22	MALE	Chennai	Looking for Job	Government Employee	website
357	USR1358	25	MALE	Chennai	B.Tech	Government Employee	SEO
358	USR1359	18	FEMALE	Mumbai	B.Tech	Government Employee	email_marketing
359	USR1360	16	MALE	Mumbai	Intermediate	Government Employee	social_media

360 rows × 7 columns

In [45]:
```
sns.boxplot(df1['age'])
```

```
C:\Users\Usman\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

Out[45]:
```
<AxesSubplot:xlabel='age'>
```



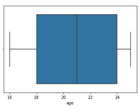
In [47]:
```
Q1 = df1.quantile(0.25)
Q3 = df1.quantile(0.75)
IQR = Q3 - Q1

df1 = df1[~((df1 < (Q1 - 1.5 * IQR)) |(df1 > (Q3 + 1.5 * IQR))).any(axis=1)]
```

```
C:\Users\Usman\AppData\Local\Temp\ipykernel_56388\775481838.py:5: FutureWarning: Automatic reindexing on DataFrame vs Series comparisons is deprecated and will raise ValueError in in a future version. Do `left, right = left.align(right, axis=1, copy=False)` before e.g. `left == right`
  df1 = df1[~((df1 < (Q1 - 1.5 * IQR)) |(df1 > (Q3 + 1.5 * IQR))).any(axis=1)]
```

In [48]:
```
sns.boxplot(df1['age'])
```

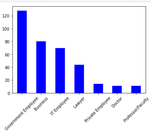
```
C:\Users\Usman\Anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

Out[48]:
```
<AxesSubplot:xlabel='age'>
```



Most role of parents occupation

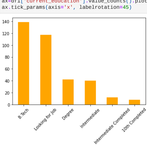
In [82]:
```
ax=df1['parent_occupation'].value_counts().plot(kind='bar',color='b')
ax.tick_params(axis='x', labelrotation=45)
```



• Most parents are are government employees and least are in educational field

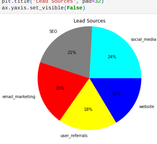
Lead's Educational background

In [81]:
```
ax=df1['current_education'].value_counts().plot(kind='bar',color='orange')
ax.tick_params(axis='x', labelrotation=45)
```



Most lead sources

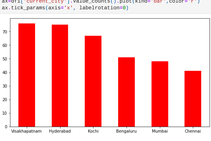
In [79]:
```
ax=df1['lead_gen_source'].value_counts().plot(kind='pie',autopct='%1.0f%%',radius=1.5,colors = ['aqua', 'grey','red','yellow','b'])
plt.title('Lead Sources', pad=32)
ax.yaxis.set_visible(False)
```



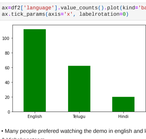
• Majority of leads are from social media and least leads are from referrals

Most residing cities

In [80]:
```
fig_dims = (9,5)
fig, ax = plt.subplots(figsize=fig_dims)

ax=df1['current_city'].value_counts().plot(kind='bar',color='r')
ax.tick_params(axis='x', labelrotation=0)
```

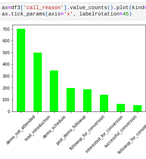


In [77]:
```
ax=df2['language'].value_counts().plot(kind='bar',color='green')
ax.tick_params(axis='x', labelrotation=0)
```



• Many people preferred watching the demo in english and least amount of people watched it in hindi and a moderate amount of people watched the demo in Telgu because we have many leads residing in Hyderabad & Vishakapatnam

In [75]:
```
ax=df3['call_reason'].value_counts().plot(kind='bar',color='lime')
ax.tick_params(axis='x', labelrotation=45)
```

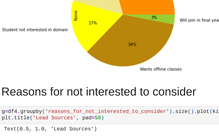


• Here we can see majority of leads did not attend the demo so in order to get follow up a call was made

Reasons for not interested in demo

In [97]:
```
df4.groupby('reasons_for_not_interested_in_demo').size().plot(kind='pie',autopct='%1.0f%%',radius=1.5,colors = ['darkorange', 'khaki','yellow','darkgoldenrod','yellowgreen'])
plt.title('Lead Sources', pad=32)
```

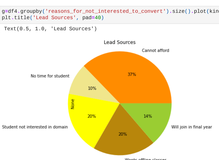
Out[97]:
```
Text(0.5, 1.0, 'Lead Sources')
```



Reasons for not interested to consider

In [123]:
```
df4.groupby('reasons_for_not_interested_to_consider').size().plot(kind='pie',autopct='%1.0f%%',radius=1.5,colors = ['darkorange', 'khaki','yellow','darkgoldenrod','yellowgreen'])
plt.title('Lead Sources', pad=50)
```

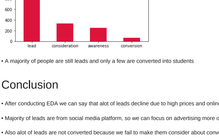
Out[123]:
```
Text(0.5, 1.0, 'Lead Sources')
```



Reasons for not interested to convert

In [119]:
```
df4.groupby('reasons_for_not_interested_to_convert').size().plot(kind='pie',autopct='%1.0f%%',radius=1.5,colors = ['darkorange', 'khaki','yellow','darkgoldenrod','yellowgreen'])
plt.title('Lead Sources', pad=40)
```

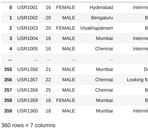
Out[119]:
```
Text(0.5, 1.0, 'Lead Sources')
```



• The main reason for not joining is 'cannot afford' it is possible the fees is too high and not everyone can afford it. Also since pandemic is over everyone wants offline class for personal attention and interaction

Stage of lead

In [122]:
```
ax=df3['lead_stage'].value_counts().plot(kind='bar',color='crimson')
ax.tick_params(axis='x', labelrotation=0)
```



• A majority of people are still leads and only a few are converted into students

Conclusion

• After conducting EDA we can say that alot of leads decline due to high prices and online classes

• Majority of leads are from social media platform, so we can focus on advertising more on social media

• Also alot of leads are not converted because we fail to make them consider about conversion, the sales team needs to be worked on

• Since majority of students do not attend demo they are not aware of the classes provided. This leads to less people looking into considering abot taking admission

• At the end offline classes should be considered so that the students who are not comfortable with online classes can attend offline institutes