In [1]:

```
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
%matplotlib inline
import seaborn as sns
sns.set()
import random
```

In [21]:

```
# Question 1:Read the dataset.
data= pd.read_csv("C:\\Users\\ADMIN\\Desktop\\LETSUPGRADE\\DATA SCIENCE\\NOT
print(data)
```

```
                    Id         groupId         matchId  assists  boosts  \
0     2f262dd9795e60  78437bcd91d40e  d5db3a49eb2955        0       0
1     a32847cf5bf34b  85b7ce5a12e10b  65223f05c7fdb4        0       0
2     1b1900a9990396  edf80d6523380a  1cadec4534f30a        0       3
3     f589dd03b60bf2  804ab5e5585558  c4a5676dc91604        0       0
4     c23c4cc5b78b35  b3e2cd169ed920  cd595700a01bfa        0       0
...              ...             ...             ...      ...     ...
9995  ef4f474acd8e85  2eca2a8391f75d  492ecdfae90b46        0       3
9996  cf0bf82fb4d80e  2eaf2765f93adb  14bffd71e96320        0       0
9997  a0a31a0b1dcbe1  8d50c64ccc5071  147e4bbb62e3bb        0       0
9998  f6874657399d69  d31843d7e62ccb  662567dcf280f5        0       0
9999  90359b0b8f8b0d  61d5b1bb8da43f  258bfa48d88014        0       0

      damageDealt  DBNOs  headshotKills  heals  killPlace  ...  roadKills  \
0             0.0      0              0      0         92  ...          0
1           163.2      1              1      0         42  ...          0
2           278.7      2              1      8         16  ...          0
3           191.9      1              0      0         31  ...          0
4           100.0      1              0      0         87  ...          0
...           ...    ...            ...    ...        ...  ...        ...
9995        204.5      1              0      0         17  ...          0
9996          0.0      0              0      0         49  ...          0
9997          0.0      0              0      0         83  ...          0
9998          0.0      0              0      0         82  ...          0
9999          0.0      0              0      0         78  ...          0

      swimDistance  teamKills  vehicleDestroys  walkDistance  weaponsAcquired  \
0              0.0          0                0          0.00                0
1              0.0          0                0        132.70                2
2              0.0          0                0       3591.00               10
3              0.0          0                0        332.70                3
4              0.0          0                0        252.70                3
...            ...        ...              ...           ...              ...
9995           0.0          0                0       1648.00                4
9996           0.0          0                0        897.10                6
9997           0.0          0                0        188.20                2
9998           0.0          0                0        108.10                5
9999           0.0          0                0         53.36                1

      winPoints  winPlacePerc  Unnamed: 29  kils
0          1470        0.0000          NaN     0
1          1531        0.2222          NaN     1
2             0        0.8571          NaN     1
3             0        0.3462          NaN     0
4          1557        0.0690          NaN     0
...         ...           ...          ...   ...
9995       1471        0.8333          NaN     0
9996       1500        0.7174          NaN     0
9997       1434        0.2083          NaN     0
9998       1534        0.2449          NaN     0
9999          0        0.1875          NaN     0

[10000 rows x 31 columns]
```

In [3]:

```
#Question 2:Check the datatype of all the columns.
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 29 columns):
Id                  10000 non-null object
groupId             10000 non-null object
matchId             10000 non-null object
assists             10000 non-null int64
boosts              10000 non-null int64
damageDealt         10000 non-null float64
DBNOs               10000 non-null int64
headshotKills       10000 non-null int64
heals               10000 non-null int64
killPlace           10000 non-null int64
killPoints          10000 non-null int64
kills               10000 non-null int64
killStreaks         10000 non-null int64
longestKill         10000 non-null float64
matchDuration       10000 non-null int64
matchType           10000 non-null object
maxPlace            10000 non-null int64
numGroups           10000 non-null int64
rankPoints          10000 non-null int64
revives             10000 non-null int64
rideDistance        10000 non-null float64
roadKills           10000 non-null int64
swimDistance        10000 non-null float64
teamKills           10000 non-null int64
vehicleDestroys     10000 non-null int64
walkDistance        10000 non-null float64
weaponsAcquired     10000 non-null int64
winPoints           10000 non-null int64
winPlacePerc        10000 non-null float64
dtypes: float64(6), int64(19), object(4)
memory usage: 2.2+ MB
```

In [4]:

```
#Question 3:Find the summary of all the numerical columns and write your fin
numerical_cols = data.describe().columns
df = data[numerical_cols]
df

```

Out[4]:

	assists	boosts	damageDealt	DBNOs	headshotKills	heals	killPlace	killPoints	kills	killSt
0	0	0	0.0	0	0	0	92	1126	0	
1	0	0	163.2	1	1	0	42	1309	1	
2	0	3	278.7	2	1	8	16	0	2	
3	0	0	191.9	1	0	0	31	0	1	
4	0	0	100.0	1	0	0	87	1332	0	
...	...	...	...	...	...	...	...	...	...	
9995	0	3	204.5	1	0	0	17	1033	2	
9996	0	0	0.0	0	0	0	49	1000	0	
9997	0	0	0.0	0	0	0	83	1334	0	
9998	0	0	0.0	0	0	0	82	1335	0	
9999	0	0	0.0	0	0	0	78	0	0	

10000 rows × 25 columns

In [7]:

```
# Question 4:The average person kills how many players?
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns
sns.set_style("darkgrid")
sns.set(color_codes=True)
df = pd.read_csv("C:\\Users\\ADMIN\\Desktop\\LETSUPGRADE\\DATA SCIENCE\\NOTE
print(df["kills"])

```

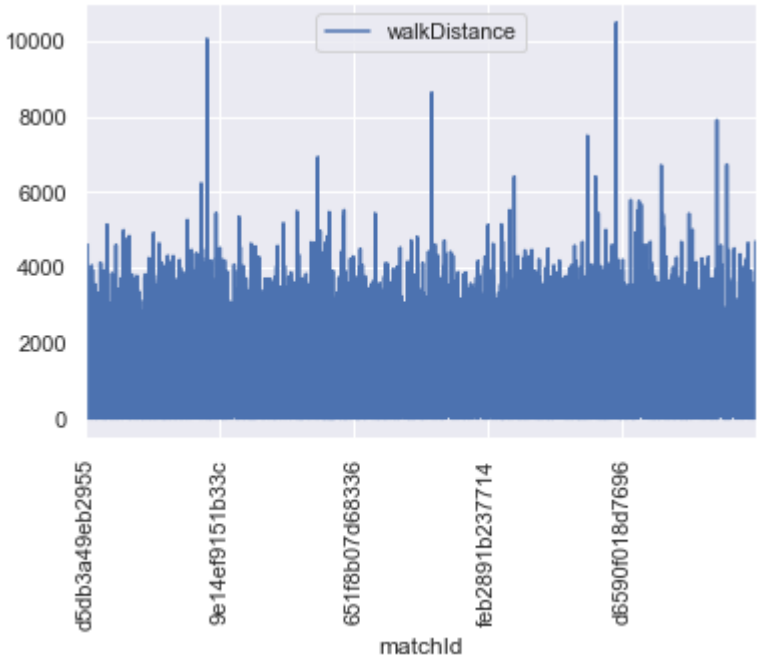
```
0       0
1       1
2       2
3       1
4       0
       ..
9995    2
9996    0
9997    0
9998    0
9999    0
Name: kills, Length: 10000, dtype: int64
```

In [8]:

```
#Question 11:Plot distribution of the match's duration vs walk distance side
import seaborn as sns
from matplotlib import pyplot as plt
import pandas as pd
df = pd.read_csv("C:\\Users\\ADMIN\\Desktop\\LETSUPGRADE\\DATA SCIENCE\\NOTE
df.plot(x='matchId',y='walkDistance')
plt.xticks(rotation=90)
```

Out[8]:

```
(array([    0.,  2000.,  4000.,  6000.,  8000., 10000.]),
 <a list of 6 Text xticklabel objects>)
```

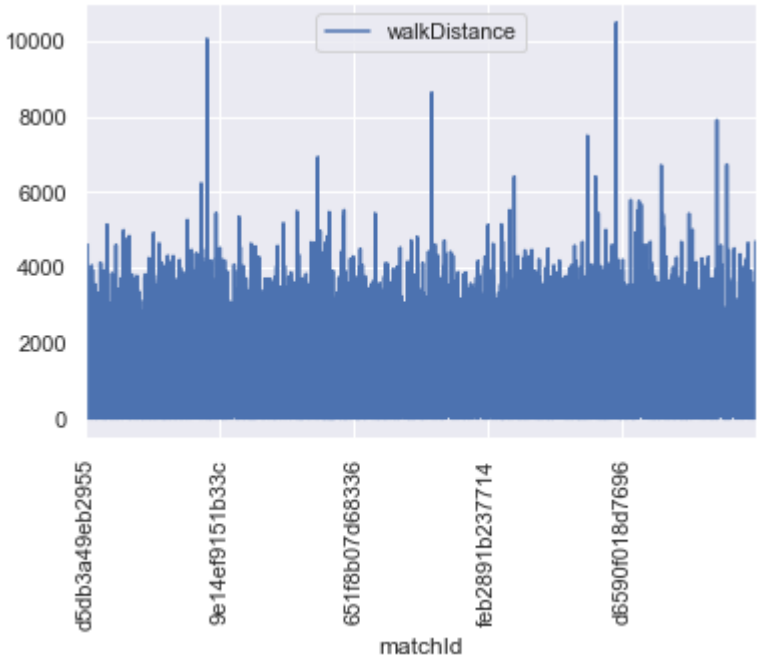


In [9]:

```
#question 10: Plot distribution of the match's duration vs walk distance one
import seaborn as sns
from matplotlib import pyplot as plt
import pandas as pd
df = pd.read_csv("C:\\Users\\ADMIN\\Desktop\\LETSUPGRADE\\DATA SCIENCE\\NOTE
df.plot(x='matchId',y='walkDistance')
plt.xticks(rotation=90)
```

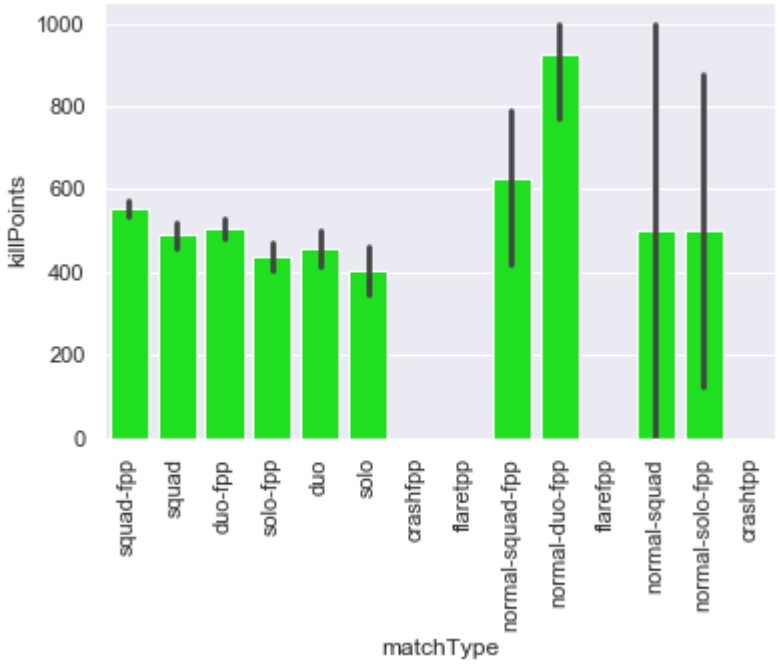
Out[9]:

```
(array([    0.,  2000.,  4000.,  6000.,  8000., 10000.]),
 <a list of 6 Text xticklabel objects>)
```



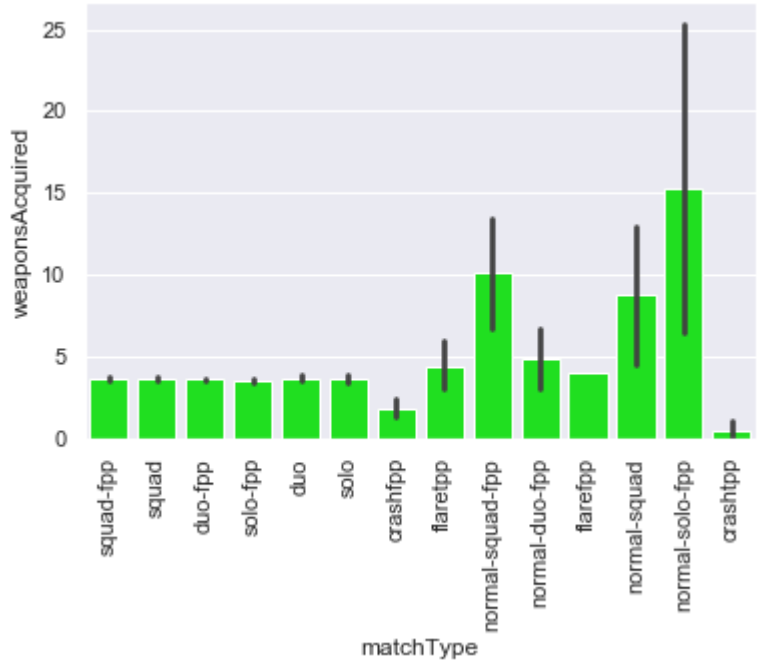
In [19]:

```
#Question 14. Plot a barplot of ‘matchType’ vs 'killPoints'. Write your infe
import seaborn as sns
from matplotlib import pyplot as plt
import pandas as pd
df = pd.read_csv("C:\\Users\\ADMIN\\Desktop\\LETSUPGRADE\\DATA SCIENCE\\NOTE
sns.barplot(x="matchType", y="killPoints", data=df,color="lime")
plt.xticks(rotation=90)
plt.show()

```



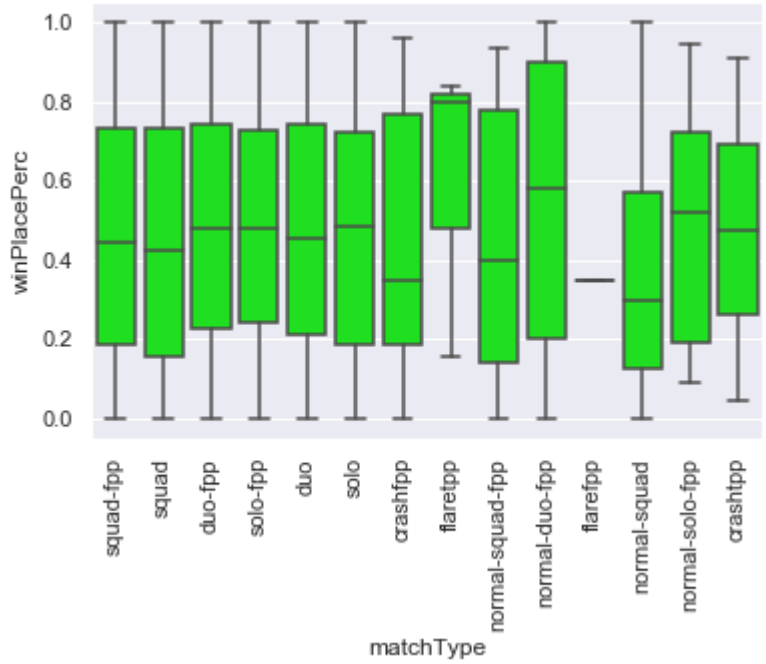
In [20]:

```
#Question 15. Plot a barplot of ‘matchType’ vs ‘weaponsAcquired’. Write your
import seaborn as sns
from matplotlib import pyplot as plt
import pandas as pd
df = pd.read_csv("C:\\Users\\ADMIN\\Desktop\\LETSUPGRADE\\DATA SCIENCE\\NOTE
sns.barplot(x="matchType", y="weaponsAcquired", data=df,color="lime")
plt.xticks(rotation=90)
plt.show()

```



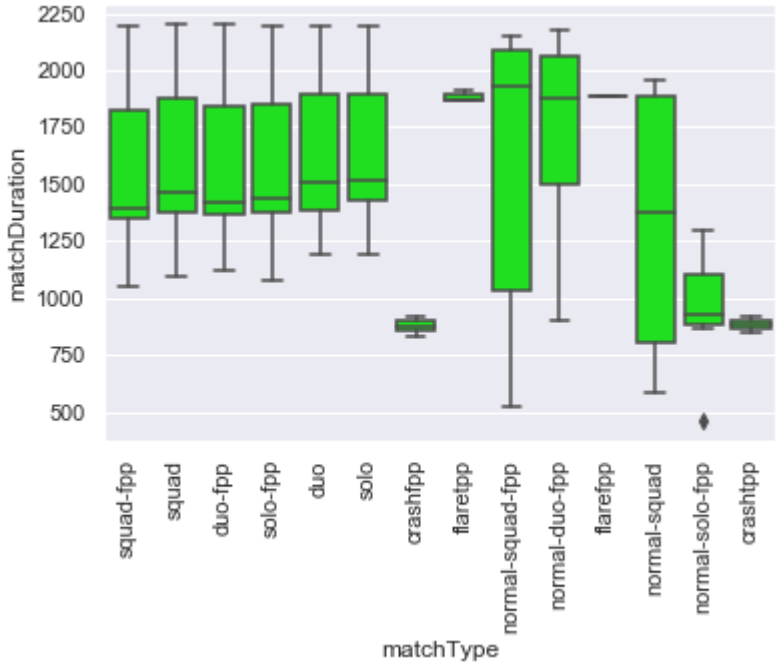
In [21]:

```
#Question 17. Plot a boxplot of 'matchType' vs 'winPlacePerc'. Write your in
import seaborn as sns
from matplotlib import pyplot as plt
import pandas as pd
df = pd.read_csv("C:\\Users\\ADMIN\\Desktop\\LETSUPGRADE\\DATA SCIENCE\\NOTE
sns.boxplot(x="matchType", y="winPlacePerc", data=df,color="lime")
plt.xticks(rotation=90)
plt.show()

```



In [22]:

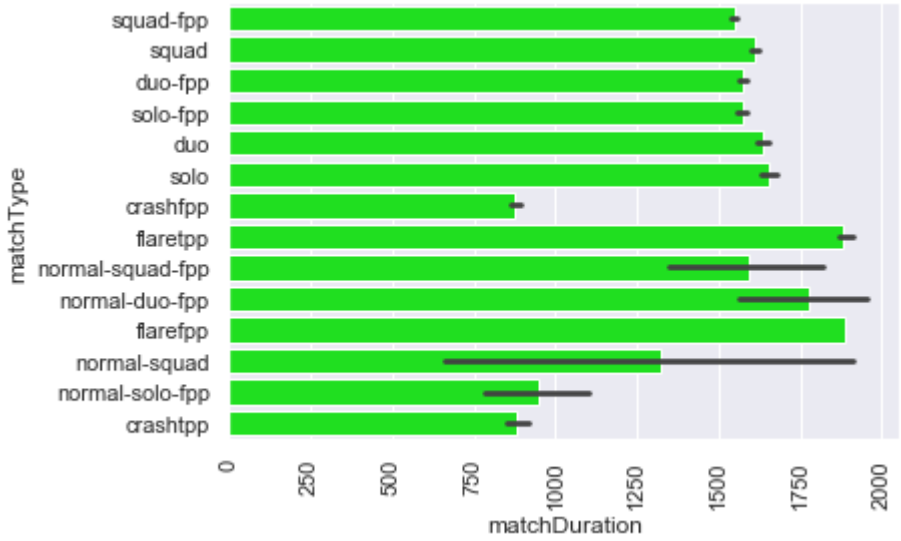
```
#Question 18. Plot a boxplot of 'matchType' vs 'matchDuration'. Write your i
import seaborn as sns
from matplotlib import pyplot as plt
import pandas as pd
df = pd.read_csv("C:\\Users\\ADMIN\\Desktop\\LETSUPGRADE\\DATA SCIENCE\\NOTE
sns.boxplot(x="matchType", y="matchDuration", data=df,color="lime")
plt.xticks(rotation=90)
plt.show()

```



In [23]:
```
#Question16. Find the Categorical columns.
list(data.columns.values)
```

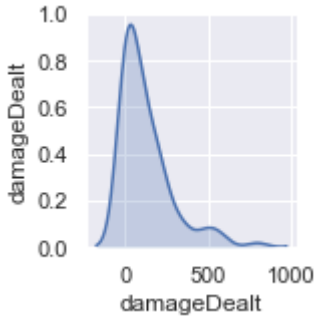
Out[23]:
```
['Id',
 'groupId',
 'matchId',
 'assists',
 'boosts',
 'damageDealt',
 'DBNOs',
 'headshotKills',
 'heals',
 'killPlace',
 'killPoints',
 'kills',
 'killStreaks',
 'longestKill',
 'matchDuration',
 'matchType',
 'maxPlace',
 'numGroups',
 'rankPoints',
 'revives',
 'rideDistance',
 'roadKills',
 'swimDistance',
 'teamKills',
 'vehicleDestroys',
 'walkDistance',
 'weaponsAcquired',
 'winPoints',
 'winPlacePerc']
```

In [39]:
```
#Question 19. Change the orientation of the above plot to horizontal.
import seaborn as sns
from matplotlib import pyplot as plt
import pandas as pd
df = pd.read_csv("C:\\Users\\ADMIN\\Desktop\\LETSUPGRADE\\DATA SCIENCE\\NOTE
sns.barplot(x="matchDuration",y="matchType",data=df,color="lime")
plt.xticks(rotation=90)
plt.show()

```



In [43]:
```
#Question   22. Take a sample of size 50 from the column damageDealt for 100
col=[ "damageDealt"]
sns.pairplot(data[col].sample(100), size =2.5, kind="reg", diag_kind="kde")
```

Out[43]: <seaborn.axisgrid.PairGrid at 0xb83be33248>



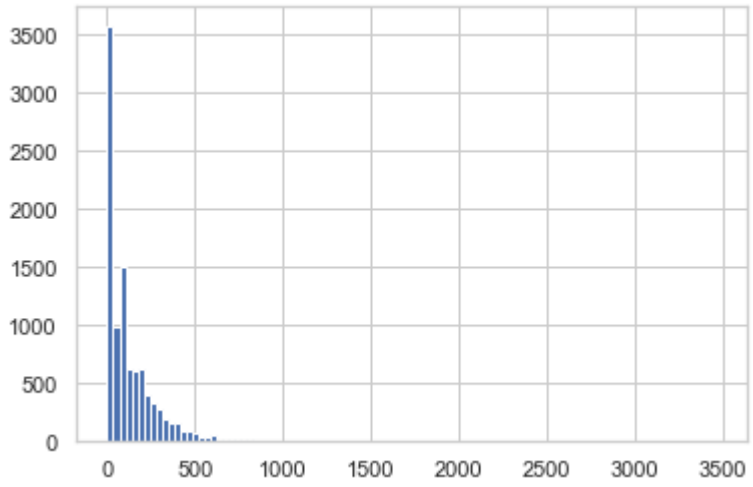
In [44]:
```
average = df['damageDealt'].mean()
print(average)
```

```
129.2112641000002
```

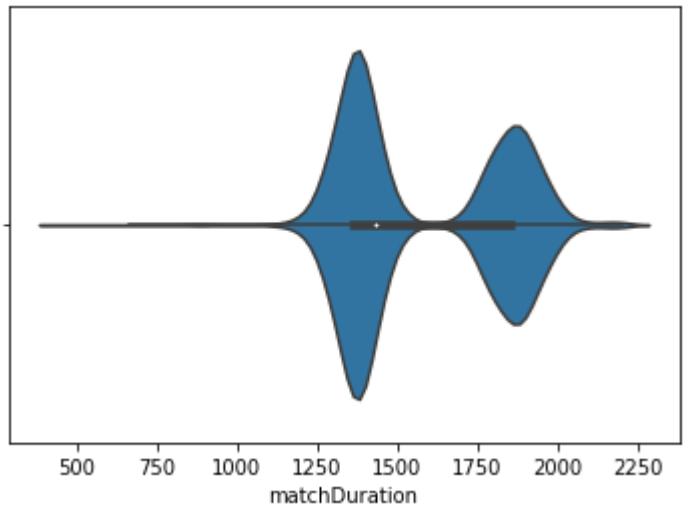
In [51]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [53]:
```
data_url = 'C:\\Users\\ADMIN\\Desktop\\LETSUPGRADE\\DATA SCIENCE\\NOTES\\pub
gapminder = pd.read_csv("C:\\Users\\ADMIN\\Desktop\\LETSUPGRADE\\DATA SCIENC
gapminder.head(n=3)
gapminder['damageDealt'].hist(bins=100)
```

Out[53]: <matplotlib.axes._subplots.AxesSubplot at 0xb83656abc8>

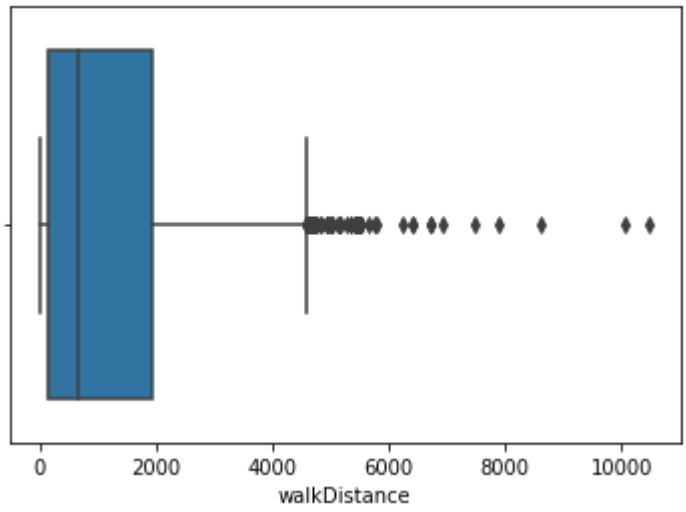


In [6]:
```
#Question8. Comment on distribution of the match's duration. Use seaborn.
import seaborn
import pandas
import matplotlib.pyplot as plt
csv = pandas.read_csv("C:\\Users\\ADMIN\\Desktop\\LETSUPGRADE\\DATA SCIENCE\
res = seaborn.violinplot(x=csv['matchDuration'])
plt.show()
```



In [7]:

```
#Question 9  Comment on distribution of the walk distance. Use seaborn.
import seaborn
import pandas
import matplotlib.pyplot as plt
csv = pandas.read_csv("C:\\Users\\ADMIN\\Desktop\\LETSUPGRADE\\DATA SCIENCE\
res = seaborn.boxplot(x=csv['walkDistance'])
plt.show()
```



In [9]:

```
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt
%matplotlib inline
import seaborn as sns
sns.set()
import random
```

In [10]:

```
#Question 20 Add a new column called 'KILL' which contains the sum of follow
teamKills, roadKills.
data= pd.read_csv("C:\\Users\\ADMIN\\Desktop\\LETSUPGRADE\\DATA SCIENCE\\NOT
data.head()
```

Out[10]:

	Id	groupId	matchId	assists	boosts	damageDealt	DBNOs	heads
0	2f262dd9795e60	78437bcd91d40e	d5db3a49eb2955	0	0	0.0	0	
1	a32847cf5bf34b	85b7ce5a12e10b	65223f05c7fdb4	0	0	163.2	1	
2	1b1900a9990396	edf80d6523380a	1cadec4534f30a	0	3	278.7	2	
3	f589dd03b60bf2	804ab5e5585558	c4a5676dc91604	0	0	191.9	1	
4	c23c4cc5b78b35	b3e2cd169ed920	cd595700a01bfa	0	0	100.0	1	

5 rows × 31 columns

In [19]:

```
#Question21. Round off column 'winPlacePerc' to 2 decimals.
df = pd.read_csv("C:\\Users\\ADMIN\\Desktop\\LETSUPGRADE\\DATA SCIENCE\\NOTE
print(df["winPlacePerc"])
df.round()
```

```
0       0.0000
1       0.2222
2       0.8571
3       0.3462
4       0.0690
         ...
9995    0.8333
9996    0.7174
9997    0.2083
9998    0.2449
9999    0.1875
Name: winPlacePerc, Length: 10000, dtype: float64
```

Out[19]:

	Id	groupId	matchId	assists	boosts	damageDealt	DBNOs	h
0	2f262dd9795e60	78437bcd91d40e	d5db3a49eb2955	0	0	0.0	0	
1	a32847cf5bf34b	85b7ce5a12e10b	65223f05c7fdb4	0	0	163.0	1	
2	1b1900a9990396	edf80d6523380a	1cadec4534f30a	0	3	279.0	2	
3	f589dd03b60bf2	804ab5e5585558	c4a5676dc91604	0	0	192.0	1	
4	c23c4cc5b78b35	b3e2cd169ed920	cd595700a01bfa	0	0	100.0	1	
...	...	...	...	...	...	...	...	
9995	ef4f474acd8e85	2eca2a8391f75d	492ecdfae90b46	0	3	204.0	1	
9996	cf0bf82fb4d80e	2eaf2765f93adb	14bffd71e96320	0	0	0.0	0	
9997	a0a31a0b1dcbe1	8d50c64ccc5071	147e4bbb62e3bb	0	0	0.0	0	
9998	f6874657399d69	d31843d7e62ccb	662567dcf280f5	0	0	0.0	0	
9999	90359b0b8f8b0d	61d5b1bb8da43f	258bfa48d88014	0	0	0.0	0	

10000 rows × 31 columns

In [20]:
```
# Question 7:  Print all the columns of the dataframe.
numerical_cols = data.describe().columns
df = data[numerical_cols]
df
```

Out[20]:

	assists	boosts	damageDealt	DBNOs	headshotKills	heals	killPlace	killPoints	kills	killSt
0	0	0	0.0	0	0	0	92	1126	0	
1	0	0	163.2	1	1	0	42	1309	1	
2	0	3	278.7	2	1	8	16	0	2	
3	0	0	191.9	1	0	0	31	0	1	
4	0	0	100.0	1	0	0	87	1332	0	
...	...	...	...	...	...	...	...	...	...	
9995	0	3	204.5	1	0	0	17	1033	2	
9996	0	0	0.0	0	0	0	49	1000	0	
9997	0	0	0.0	0	0	0	83	1334	0	
9998	0	0	0.0	0	0	0	82	1335	0	
9999	0	0	0.0	0	0	0	78	0	0	

10000 rows × 27 columns

In [22]:
```
# Question 13. How many unique values are there in 'matchType' and what are
import pandas as pd
data = pd.read_csv("C:\\Users\\ADMIN\\Desktop\\LETSUPGRADE\\DATA SCIENCE\\NO
print(data['matchType'].unique())
```

```
['squad-fpp' 'squad' 'duo-fpp' 'solo-fpp' 'duo' 'solo' 'crashfpp'
 'flaretpp' 'normal-squad-fpp' 'normal-duo-fpp' 'flarefpp' 'normal-squad'
 'normal-solo-fpp' 'crashtpp']
```

In [ ]:
```
1
```