# Kiva Python Study

October 16, 2022

Kiva is a non-profit organization that allows users to lend money to low income recipients. The following dataset is extracted from a larger dataset from the website Kaggle.com (https://www.kaggle.com/datasets/fkosmowski/kivadhsv1).

## 1 1. Import Matplotlib, Pandas, and Seaborn modules

[24]:
```
from matplotlib import pyplot as plt
import pandas as pd
import seaborn as sns
```

## 2 2. Load data into dataframe and check it

[25]:
```
df =pd.read_csv('C:\kiva_data.csv')
df.head()
```

[25]:

| | loan_amount | activity | country | gender |
|---|---|---|---|---|
| 0 | 625 | Food Production/Sales | Pakistan | female |
| 1 | 250 | Food Production/Sales | Pakistan | female |
| 2 | 400 | Food Production/Sales | Pakistan | female |
| 3 | 400 | Food Production/Sales | Pakistan | female |
| 4 | 500 | Food Production/Sales | Pakistan | female |

## 3 3. Examine a larger overview of the dataset.

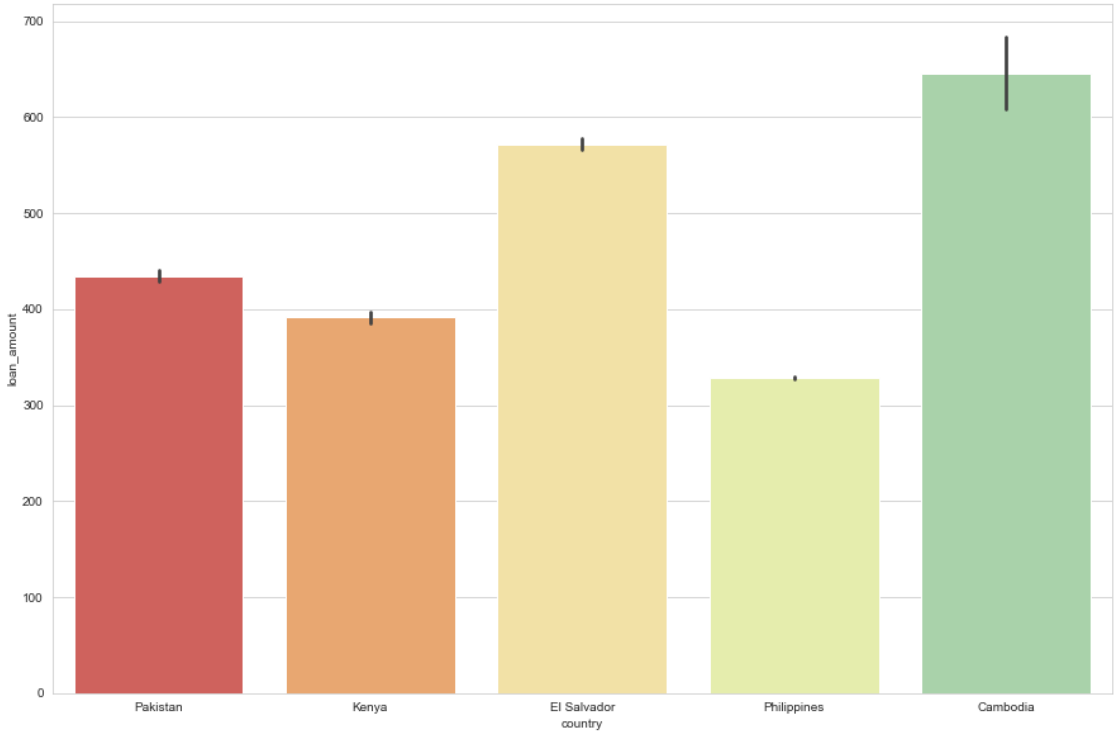
[26]:
```
print(df.head(25))
```

```
    loan_amount               activity   country  gender
0           625  Food Production/Sales  Pakistan  female
1           250  Food Production/Sales  Pakistan  female
2           400  Food Production/Sales  Pakistan  female
3           400  Food Production/Sales  Pakistan  female
4           500  Food Production/Sales  Pakistan  female
5           500  Food Production/Sales  Pakistan  female
6           400  Food Production/Sales  Pakistan  female
7           500  Food Production/Sales  Pakistan  female
8           400  Food Production/Sales  Pakistan  female
```

| | | | | |
|---|---|---|---|---|
| 9 | 450 | Food Production/Sales | Pakistan | female |
| 10 | 250 | Food Production/Sales | Pakistan | female |
| 11 | 300 | Food Production/Sales | Pakistan | female |
| 12 | 275 | Food Production/Sales | Pakistan | female |
| 13 | 425 | Food Production/Sales | Pakistan | female |
| 14 | 425 | Food Production/Sales | Pakistan | female |
| 15 | 475 | Food Production/Sales | Pakistan | female |
| 16 | 225 | Food Production/Sales | Pakistan | female |
| 17 | 475 | Food Production/Sales | Pakistan | female |
| 18 | 525 | Food Production/Sales | Pakistan | female |
| 19 | 425 | Food Production/Sales | Pakistan | female |
| 20 | 475 | Food Production/Sales | Pakistan | female |
| 21 | 550 | Food Production/Sales | Pakistan | female |
| 22 | 450 | Food Production/Sales | Pakistan | female |
| 23 | 250 | Food Production/Sales | Pakistan | female |
| 24 | 600 | Food Production/Sales | Pakistan | female |

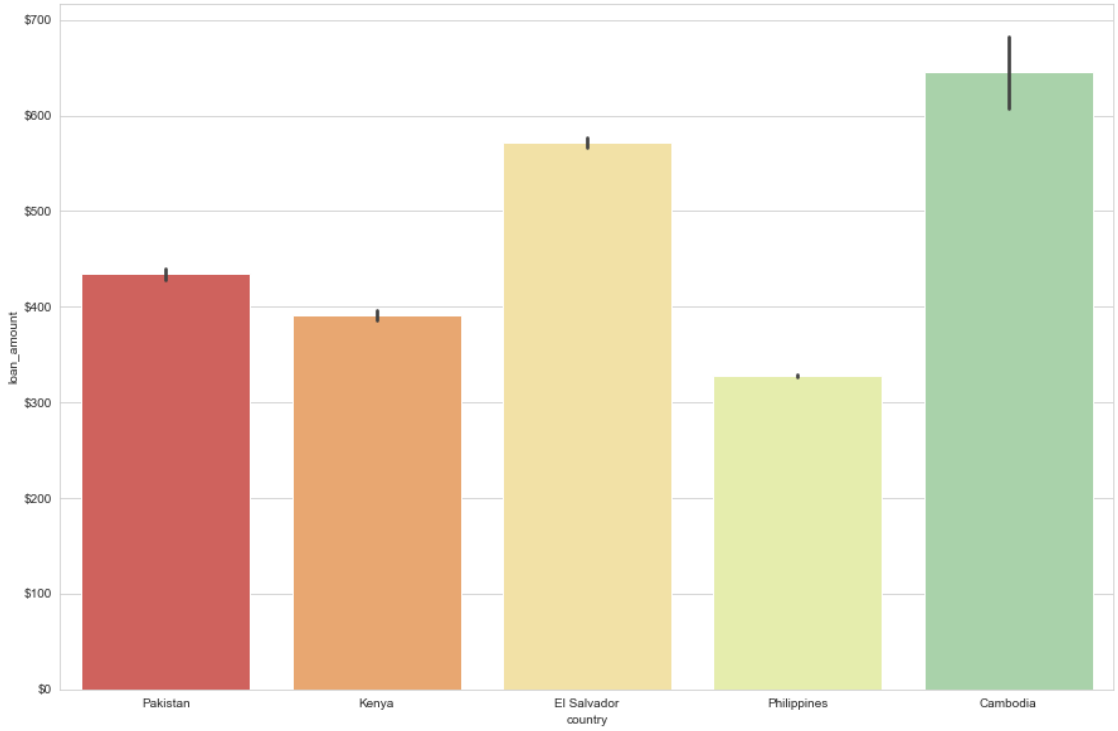
## 4 4. Examine the average loan amount by country using a Bar Chart

[27]:
```
f, ax = plt.subplots(figsize=(15, 10))

sns.barplot(data=df, x="country", y="loan_amount")
plt.show()
```



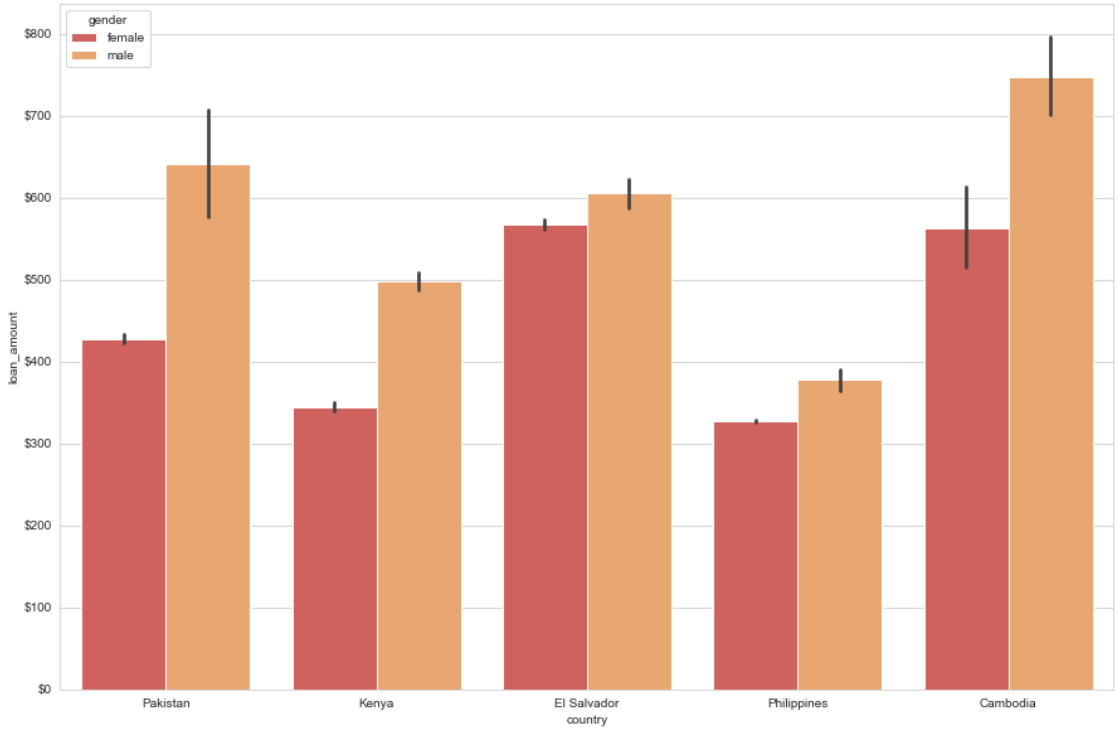
## 5 5. Adding $ symbol to the Y axis.

[28]:
```
import matplotlib.ticker as mtick
f, ax = plt.subplots(figsize=(15, 10))
sns.barplot(data=df, x="country", y = "loan_amount")
fmt = '${x:,.0f}'
tick = mtick.StrMethodFormatter(fmt)
ax.yaxis.set_major_formatter(tick)
```




## 6 6. Add the Hue parameter to the bar plot and set it to gender

[29]:
```
f, ax = plt.subplots(figsize=(15, 10))

fmt = '${x:,.0f}'
tick = mtick.StrMethodFormatter(fmt)
ax.yaxis.set_major_formatter(tick)

sns.barplot(data=df, x="country", y = "loan_amount", hue="gender")
plt.show()
```



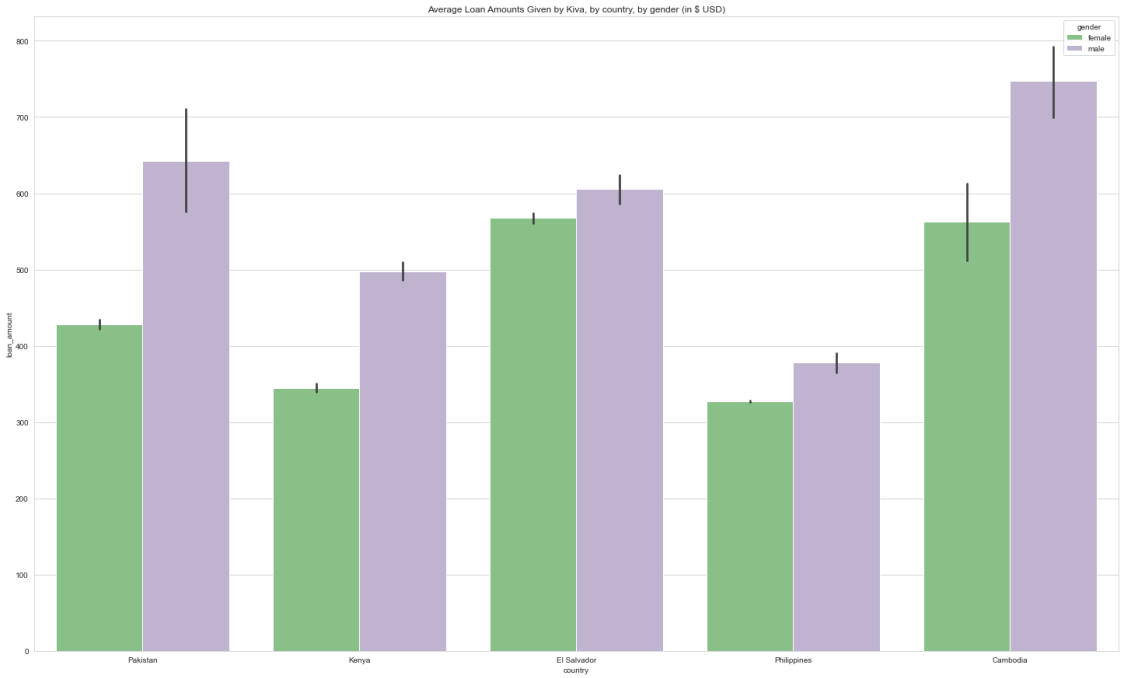

## 7 7. Bar chart data assessments

1. Males recieve larger loans from Kiva on average.
2. El Salvador has the least amount of disparity in terms of loans awarded by gender.
3. Kiva should try to decrease the gender gap in loans given to males and females.

## 8 8. Styling the chart

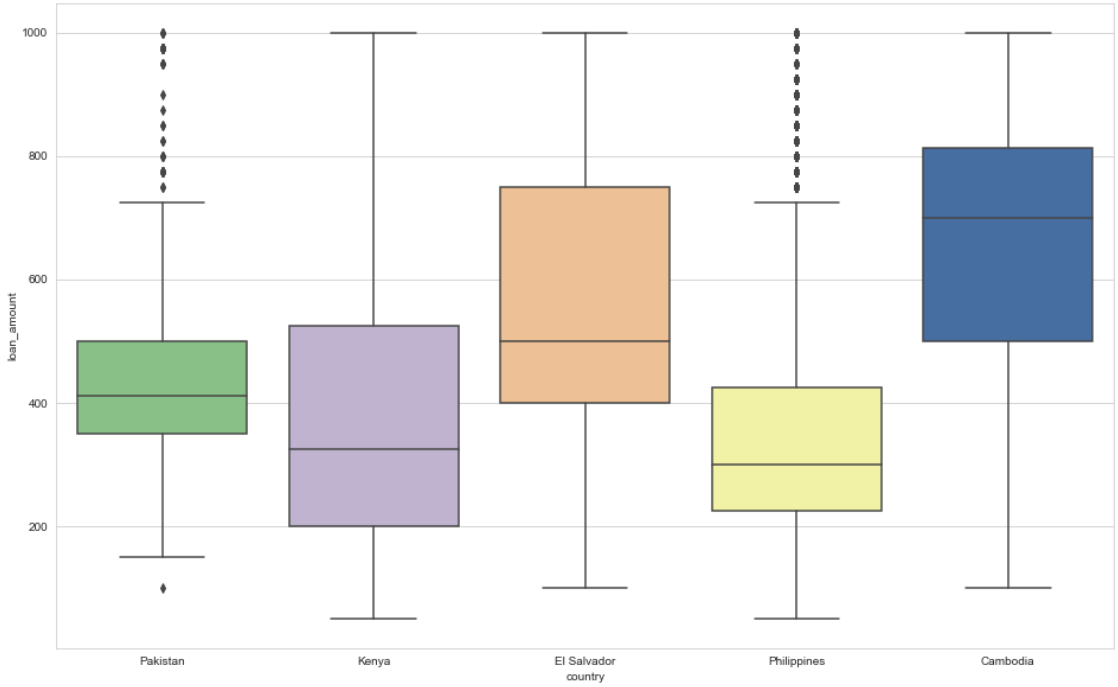
[30]:
```
sns.set_palette("Accent")
sns.set_style("whitegrid")
plt.figure(figsize=(25, 15))
plt.title("Average Loan Amounts Given by Kiva, by country, by gender (in $␣
  ↪USD)")
sns.barplot(data=df, x="country", y="loan_amount", hue="gender")
```

[30]:
```
<AxesSubplot: title={'center': 'Average Loan Amounts Given by Kiva, by country,
by gender (in $ USD)'}, xlabel='country', ylabel='loan_amount'>
```




# 9 9. A box plot to compare the distribution of loans by country

[31]:
```
plt.figure(figsize=(16, 10))
sns.boxplot(data=df, x="country", y="loan_amount")
plt.show()
```

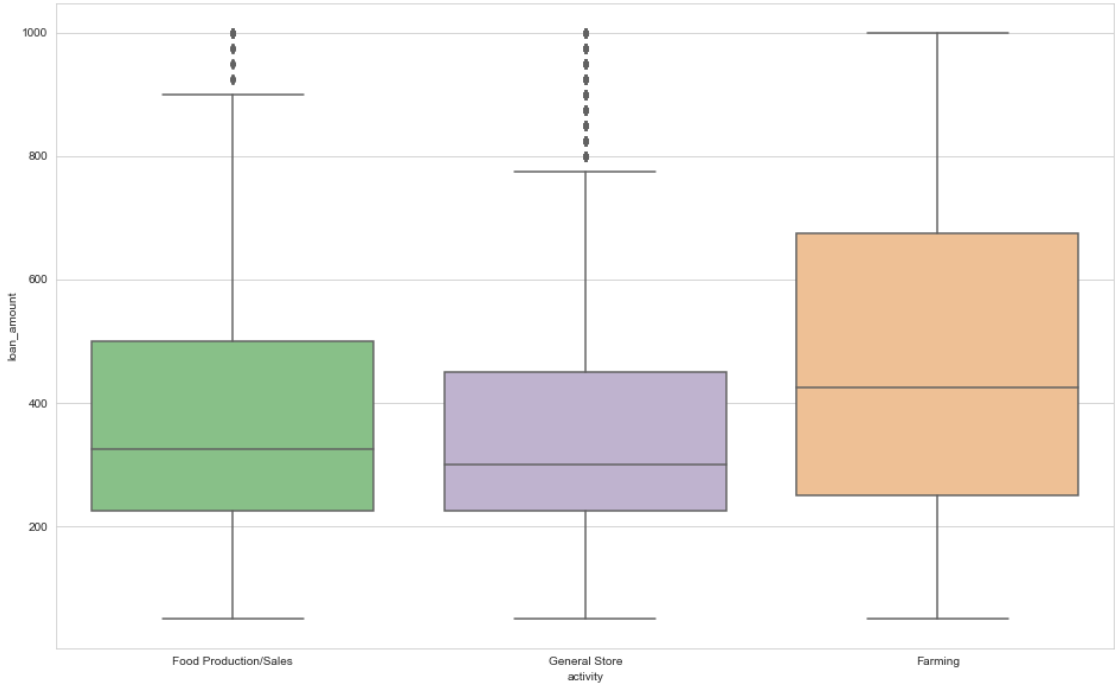



## 10 10. Box Plot by country assessments

1. Kenya has the widest distribution of loans.
2. You would most likely recieve the highest loan in Cambodia. This is due to it's place on the Y axis.

## 11 11. Visualize the loan amount by using `sns.boxplot()` to plot the loan amount by *activity*.

[32]:
```
plt.figure(figsize=(16, 10))

sns.boxplot(data=df, x="activity", y="loan_amount")
plt.show()
```

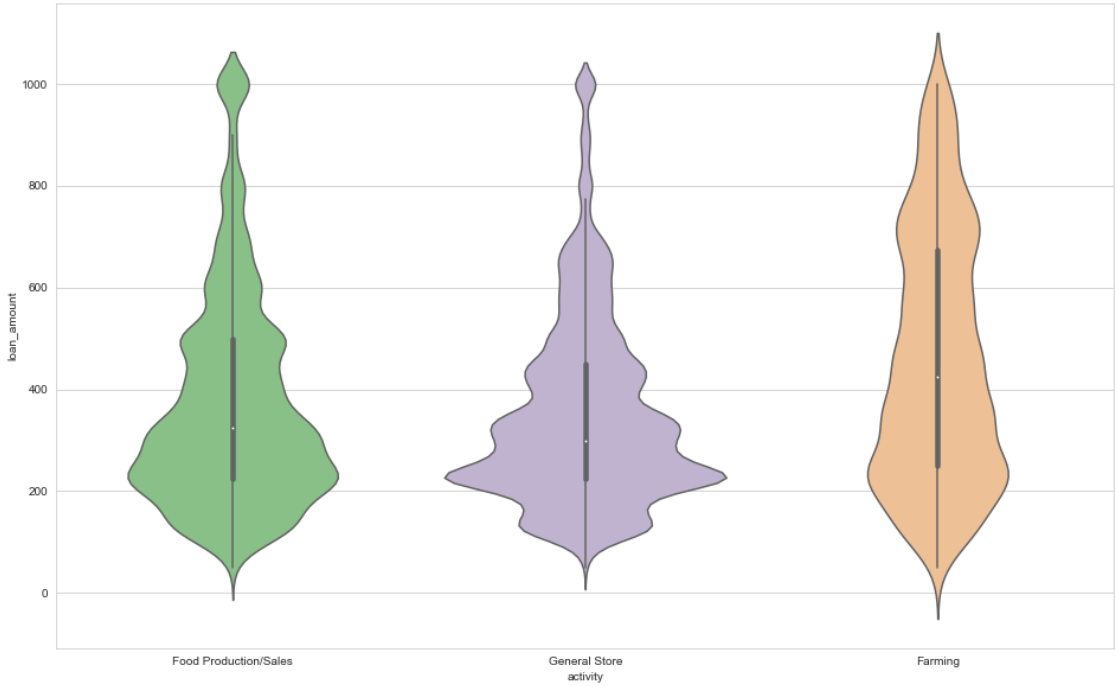



## 12 12. Activity box plot assessment

1. Farming has widest distribution of all of the activites.

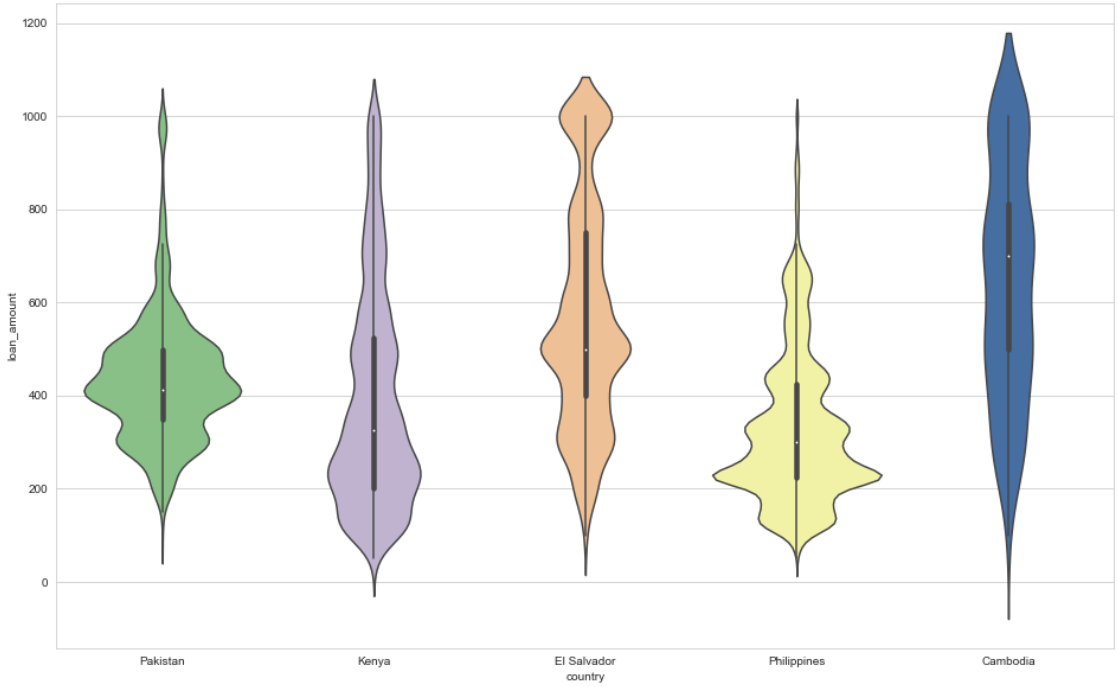
## 13 13. Violin plots by activity

[33]:
```
plt.figure(figsize=(16, 10))

sns.violinplot(data=df, x="activity", y = "loan_amount")
plt.show()
```



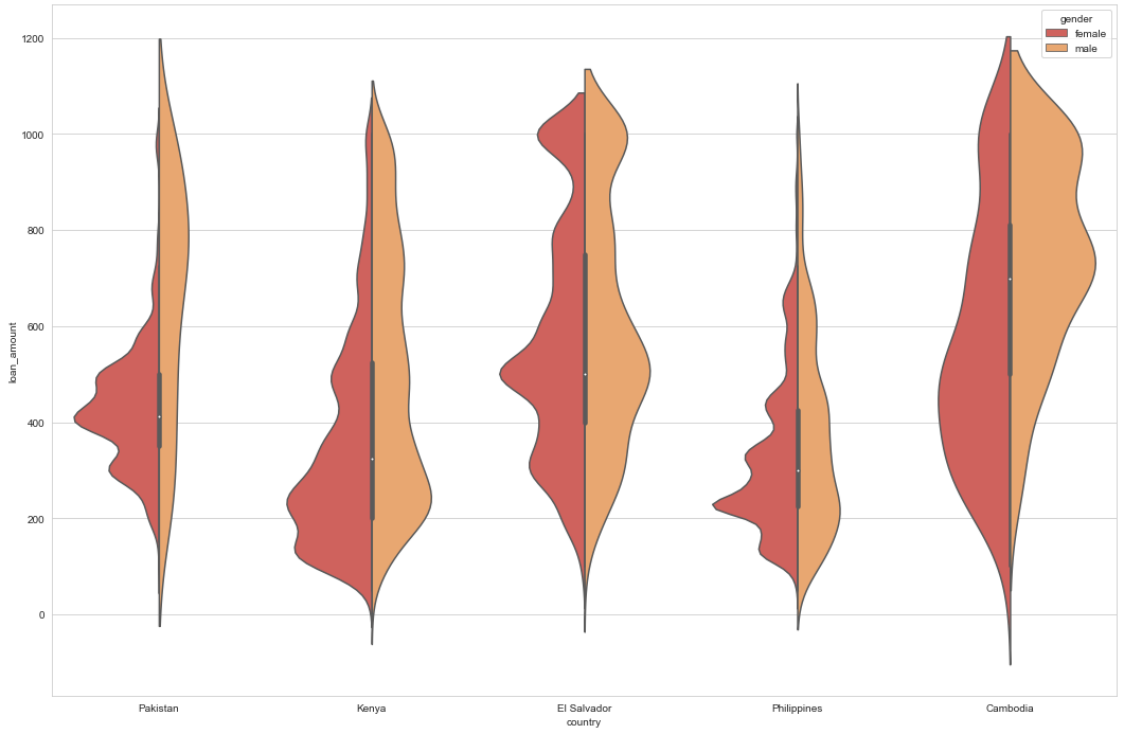

## 14 14. Violin plot that visualizes that distribution amount by country.

[34]:
```
plt.figure(figsize=(16, 10))

sns.violinplot(data=df, x="country", y = "loan_amount")
plt.show()
```




## 15 15. Using 'Hue' and 'Split' parameters to visual the loan amount by country and gender.

[35]:
```
sns.set_palette("Spectral")
plt.figure(figsize=(18, 12))

sns.violinplot(data=df, x="country", y = "loan_amount", hue="gender",␣
↪split=True)
plt.show()
```




# 16 16. Split Violin plot assessment

1. The average loan for men in Pakistan is greater than the average amount given to women.
2. The average loan for men in Cambodia is greater than the average amount given to women.
3. Loan amounts between men and women are similiar in El Salvador.

# 17 Final Assessments

## 17.1 Bar chart data

1. Males receive larger loans from Kiva on average.
2. El Salvador has the least amount of disparity in terms of loans awarded by gender.
3. Kiva should try to decrease the gender gap in loans given to males and females.

## 17.2 Box plot by country

1. Kenya has the widest distribution of loans.
2. You would most likely receive the highest loan in Cambodia. This is due to it's place on the Y axis.

## 17.3 Box plot by activity

1. Farming has the widest distribution of all of the activities.

### 17.4 Violin plot

1. The average loan for men in Pakistan is greater than the average amount given to women.
2. The average loan for men in Cambodia is greater than the average amount given to women.
3. Loan amounts between men and women are similar in El Salvador.