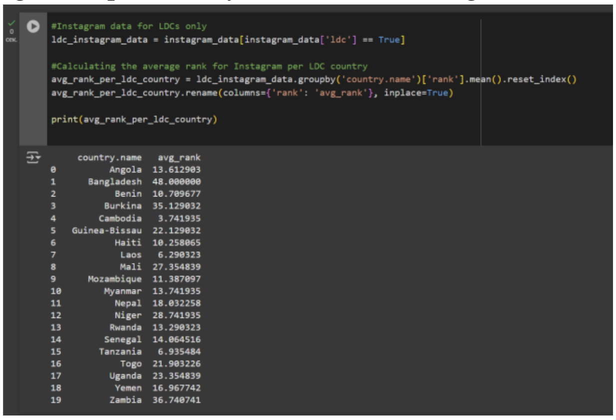
1. i) The dataset contains Instagram ranks per country over a one-month period (August 2021) and to analyze Instagram's popularity, I computed the mean rank for each country across the given period. The mean rank was chosen because it is an interpretable measure of central tendency that provides a clear representation of average popularity. Using Python and the Pandas library, I filtered the dataset to include only Instagram data and calculated the average rank per country, as illustrated in Figure 1.

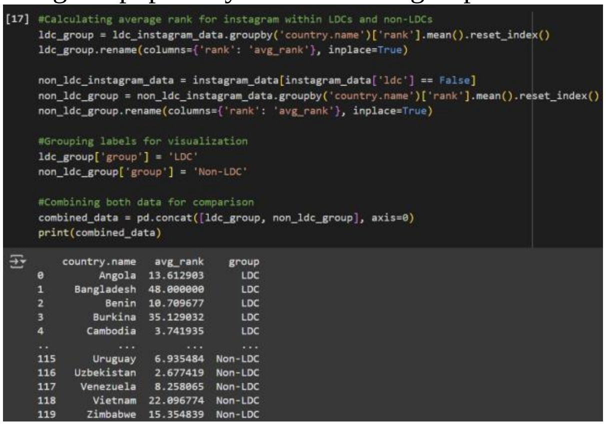


```
#Instagram data for LDCs only
ldc_instagram_data = instagram_data[instagram_data['ldc'] == True]

#Calculating the average rank for Instagram per LDC country
avg_rank_per_ldc_country = ldc_instagram_data.groupby('country.name')['rank'].mean().reset_index()
avg_rank_per_ldc_country.rename(columns={'rank': 'avg_rank'}, inplace=True)

print(avg_rank_per_ldc_country)
```

```
      country.name   avg_rank
0           Angola  13.612903
1       Bangladesh  48.000000
2            Benin  10.709677
3          Burkina  35.129032
4         Cambodia   3.741935
5    Guinea-Bissau  22.129032
6            Haiti  10.258065
7             Laos   6.290323
8             Mali  27.354839
9       Mozambique  11.387097
10         Myanmar  13.741935
11           Nepal  18.032258
12           Niger  28.741935
13          Rwanda  13.290323
14         Senegal  14.064516
15        Tanzania   6.935484
16            Togo  21.903226
17          Uganda  23.354839
18           Yemen  16.967742
19          Zambia  36.740741
```

Figure 1

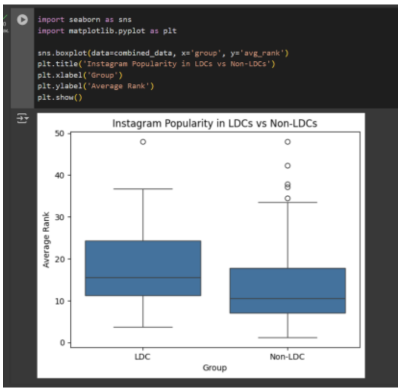
ii) a) To visualize differences in popularity between LCDs and non-LCDs, we first calculated the average Instagram popularity rank for each group as illustrated in Figure 2.



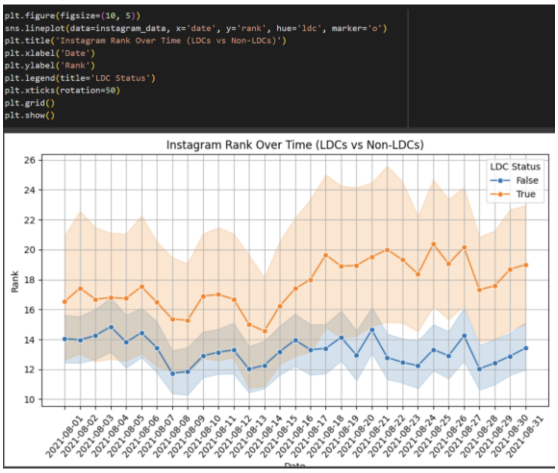
```
[17] #Calculating average rank for instagram within LDCs and non-LDCs
ldc_group = ldc_instagram_data.groupby('country.name')['rank'].mean().reset_index()
ldc_group.rename(columns={'rank': 'avg_rank'}, inplace=True)

non_ldc_instagram_data = instagram_data[instagram_data['ldc'] == False]
non_ldc_group = non_ldc_instagram_data.groupby('country.name')['rank'].mean().reset_index()
non_ldc_group.rename(columns={'rank': 'avg_rank'}, inplace=True)

#Grouping labels for visualization
ldc_group['group'] = 'LDC'
non_ldc_group['group'] = 'Non-LDC'

#Combining both data for comparison
combined_data = pd.concat([ldc_group, non_ldc_group], axis=0)
print(combined_data)
```

```
    country.name   avg_rank    group
0         Angola  13.612903      LDC
1     Bangladesh  48.000000      LDC
2          Benin  10.709677      LDC
3        Burkina  35.129032      LDC
4       Cambodia   3.741935      LDC
..           ...        ...      ...
115      Uruguay   6.935484  Non-LDC
116   Uzbekistan   2.677419  Non-LDC
117    Venezuela   8.258065  Non-LDC
118      Vietnam  22.096774  Non-LDC
119     Zimbabwe  15.354839  Non-LDC
```

Figure 2



```
import seaborn as sns
import matplotlib.pyplot as plt

sns.boxplot(data=combined_data, x='group', y='avg_rank')
plt.title('Instagram Popularity in LDCs vs Non-LDCs')
plt.xlabel('Group')
plt.ylabel('Average Rank')
plt.show()
```

```
plt.figure(figsize=(10, 5))
sns.lineplot(data=instagram_data, x='date', y='rank', hue='ldc', marker='o')
plt.title('Instagram Rank Over Time (LDCs vs Non-LDCs)')
plt.xlabel('Date')
plt.ylabel('Rank')
plt.legend(title='LDC Status')
plt.xticks(rotation=50)
plt.grid()
plt.show()
```


Figure 3

Figure 4

I used boxplot (Figure 3) and bar chart (Figure 4) to measure the popularity differences over time. Instagram is more popular in non-LDCs (lower ranks) compared to LDCs, as seen in both the boxplot and line chart. Additionally, LDC ranks show greater fluctuations, especially toward the final part of the time period.

b) To summarize differences between two groups, I used describe() method (Figure 5) since it includes all the necessary descriptive statistics. Mean indicates that Instagram is slightly less popular in LDCs compared to non-LDCs. The LDC group shows higher variability (SD = 11.42) than the non-LDC (SD = 9.15). This suggests Instagram's popularity is more consistent across non-LDC countries, while in LDC countries, the rank varies significantly.

```
ldc_stats = ldc_group['avg_rank'].describe()

non_ldc_stats = non_ldc_group['avg_rank'].describe()

print("LDC Group Descriptive Statistics:")
print(ldc_stats)

print("\nNon-LDC Group Descriptive Statistics:")
print(non_ldc_stats)
```

```
LDC Group Descriptive Statistics:
count    20.000000
mean     19.119295
std      11.422936
min       3.741935
25%      11.217742
50%      15.516129
75%      24.354839
max      48.000000
Name: avg_rank, dtype: float64

Non-LDC Group Descriptive Statistics:
count    120.000000
mean      13.549205
std        9.115151
min        1.225806
25%        7.112903
50%       10.516129
75%       17.766129
max       48.000000
Name: avg_rank, dtype: float64
```

Figure 5

iii) a) To define if Facebook is more popular, we can compute the average rank of Facebook for each country.

```
[16] #Does Facebook tend to be more popular in LDCs? Explain your reasoning
facebook_data = df[df['app.name'] == 'Facebook']
facebook_avg_rank = facebook_data.groupby('ldc')['rank'].mean()
print(facebook_avg_rank)
```

```
ldc
False    28.462559
True     24.503226
Name: rank, dtype: float64
```

Figure 6

The rank for LDCs (true, 24.5) is lower than for non-LDCs, it means Facebook tends to be more popular in LDCs as illustrated in Figure 6.

b) We can use Mann-Whitney U Test because ranks may not follow a normal distribution and this test is a non-parametric, which is suitable for comparing two independent groups when data does not meet any assumptions.

c) The observed difference in Facebook's ranks between LDCs and non-LDCs is statistically significant since P-value is low (Figure 7). We reject the null hypothesis and conclude that there is a significant difference in Facebook's mean rank between the two groups.

```
# Mann-Whitney U Test
from scipy.stats import mannwhitneyu

# Separate Facebook ranks by LDC status
ldc_ranks = facebook_data[facebook_data['ldc'] == True]['rank']
non_ldc_ranks = facebook_data[facebook_data['ldc'] == False]['rank']

# Mann-Whitney U Test
stat, p_value = mannwhitneyu(ldc_ranks, non_ldc_ranks)
print(f"Mann-Whitney U Test Statistic: {stat}")
print(f"P-Value: {p_value}")
```

```
Mann-Whitney U Test Statistic: 500058.0
P-Value: 1.0944718785937608e-12
```

Figure 7

d) The Rank-Biserial Correlation from Figure 8 represents a small-to-moderate effect size, which indicates a measurable but limited difference in ranks between two groups. This result highlights that the observed trend, while statistically significant, may not have a strong impact.

```
[19] # Report an effect size using an appropriate metric and interpret its meaning

     n1 = len(ldc_ranks)
     n2 = len(non_ldc_ranks)

     # Calculate effect size
     effect_size = 1 - (2 * stat) / (n1 * n2)
     print(f"Effect Size (Rank-Biserial Correlation): {effect_size}")

     Effect Size (Rank-Biserial Correlation): 0.1930637349231361
```

Figure 8

e) In conclusion, the analysis shows that Facebook is statistically more popular in LDCs by its lower average rank and the results of the Mann-Whitney U Test. However, the effect size is relatively small in practical terms. The results are limited by the use of rank as a proxy for popularity and the short timeframe of the dataset. Further analysis with additional variables and a longer timeframe could provide deeper understanding of the relationship.

2. i) a) To visualize changing popularity of Instagram in Pakistan, I used a line plot (Figure 9). Lower rank means more popularity. The line chart shows Instagram's popularity in Pakistan increased significantly during the first half of August 2021 by the declining rank. However, after stabilizing in mid-August, there is a slight decline towards the end of the month. This visualization illustrates both short-term fluctuations and the overall trend in popularity.

```
import matplotlib.pyplot as plt
import seaborn as sns

pakistan_instagram_data = df[(df['app.name'] == 'Instagram')
& (df['country.name'] == 'Pakistan')]

plt.figure(figsize=(12, 6))
sns.lineplot(data=pakistan_instagram_data, x='date', y='rank', marker='o')
plt.title('Rank of Instagram in Pakistan in August 2021')
plt.xlabel('Date')
plt.ylabel('Rank')
plt.xticks(rotation=45)
plt.grid()
plt.show()
```

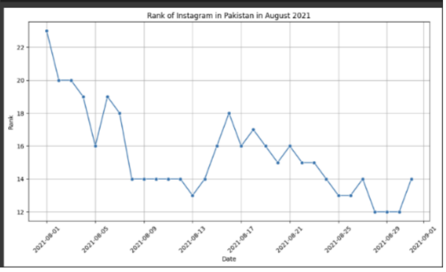


Figure 9

b) Descriptive statistics from Figure 10 show that Instagram had an average rank of M=15.4 (SD = 3.2) in Pakistan during this period. The minimum rank was 10 (highest popularity) and a maximum rank was 20 (lowest popularity).

```
# Descriptive statistics for Instagram ranks in Pakistan
desc_stats = pakistan_instagram_data['rank'].describe()
print(desc_stats)
```

```
count    31.000000
mean     15.483871
std       2.669219
min      12.000000
25%      14.000000
50%      15.000000
75%      16.500000
max      23.000000
Name: rank, dtype: float64
```

Figure 10

ii) a) During this period, Instagram's rank started at around 23 and decreased to 14 by the end of the period, rising in popularity (Figure 9).

b) The Mann-Kendall Test, non-parametric test, is appropriate because it is suitable for comparing the ordinal rank data between two independent groups (LDCs and non-LDCs) and it does not assume normal distribution, which is important for rank-based data.

c) The Mann-Kendall Test (Figure 11) showed a significant difference since P-value is much smaller than the common significance level (a=0.05). This means the observed trend is statistically significant, and we reject the null hypothesis that there is no trend in Instagram's rank over time.

```
[17] import pymannkendall as mk
     result = mk.original_test(pakistan_instagram_data['rank'])

     print(f"Trend: {result.trend}")
     print(f"P-Value: {result.p}")
     print(f"Z-Statistic: {result.z}")
```

```
Trend: decreasing
P-Value: 2.814177374221849e-05
Z-Statistic: -4.1880046619860645
```

Figure 11

d) As defined in Figure 12, Kendall's Tau test was conducted to analyze the trend in Instagram's rank over time in Pakistan. The results revealed a significant negative trend ($\tau = -0.558$) indicating that Instagram's popularity increased over time. The effect size represents a large effect and shows a meaningful improvement in Instagram's rank during the observed period.

```
[34] from scipy import stats
     n1 = pakistan_instagram_data['rank']
     n2 = pakistan_instagram_data['date']
     tau = stats.kendalltau(n1, n2)
     print(f"Kendall's Tau: {tau.statistic}")
```

```
Kendall's Tau: -0.5585764126615107
```

Figure 12