Make Up Session Assignment for Theory 1

Name : Arjun Mankuskar

Roll No:CS7-43

PRN:202401110038

1. Print the complete content of the Summary sheet.

```python
import pandas as pd

# Load the Summary sheet without any header
summary_df = pd.read_excel("yelp_analysis.xlsx", sheet_name="Summary", header=None)

# Print the entire content of the Summary sheet
print("Summary Sheet Content:")
print(summary_df)
```

```
PS D:\VS code\EDS makeup> & C:/Users/Arjun/AppData/Local/Programs/Python/Python313/python.exe "d:/VS code/EDS makeup/1.py"
Summary Sheet Content:
               0                                   1  ...                                               5                     6
0  Average Stars                Most Useful Review Text  ...                               Longest Review Text  Longest Review Stars
1         3.7775  Love this place! Amazing Happy Hour Specials!!  ...  In our continuing quest to identify cool, loca...                     4

[2 rows x 7 columns]
PS D:\VS code\EDS makeup>
```

2. Calculate the total number of reviews from the Reviews Per Year sheet.

```python
import pandas as pd

# Load the Reviews Per Year sheet
reviews_year_df = pd.read_excel("yelp_analysis.xlsx", sheet_name="Reviews Per Year")

#Calculate the Total Number of Reviews from the Reviews Per Year Sheet

# Assuming the sheet contains "Year" and "Review Count" columns
total_reviews = reviews_year_df["Review Count"].sum()

print("Total Number of Reviews:", total_reviews)
```

```
PS D:\VS code\EDS makeup> & C:/Users/Arjun/AppData/Local/Programs/Python/Python313/python.exe "d:/VS code
Total Number of Reviews: 10000
PS D:\VS code\EDS makeup>
```

3. Display the star distribution table from the Star Distribution sheet and determine which star rating received the highest count.

```python
# Load the entire Excel file once.
xls = pd.ExcelFile("yelp_analysis.xlsx")

star_dist_df = pd.read_excel(xls, sheet_name="Star Distribution")

#Display the star distribution table from the “Star Distribution” sheet and determine which star rating received the highest count.
# Assuming the DataFrame has columns named "Stars" and "Count"
max_star = star_dist_df.loc[star_dist_df["Count"].idxmax(), "Stars"]
print("Star rating with the highest count:", max_star)
print("Full Star Distribution:")
print(star_dist_df)
```

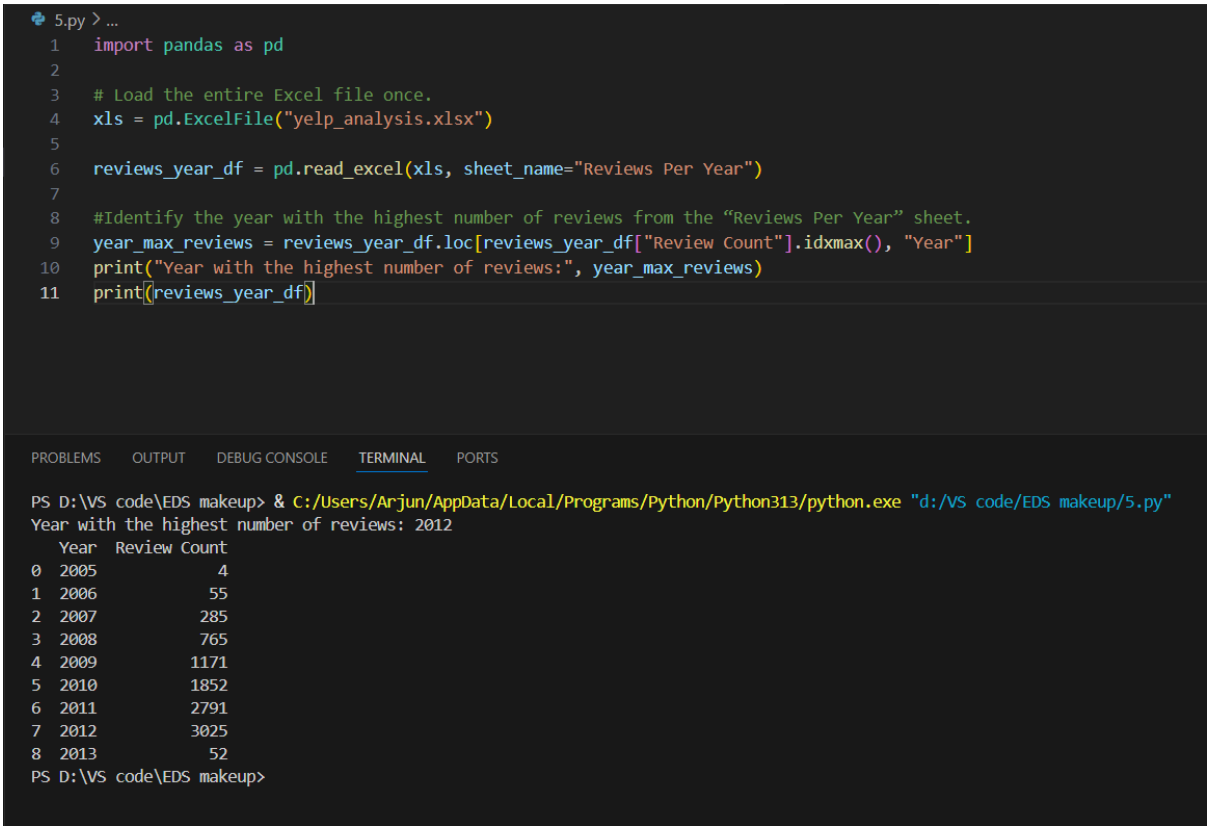
```
PS D:\VS code\EDS makeup> & C:/Users/Arjun/AppData/Local/Programs/Python/Python313/python.exe "d:/VS code/EDS makeup/3.py"
Star rating with the highest count: 4
Full Star Distribution:
   Stars  Count
0      1    749
1      2    927
2      3   1461
3      4   3526
4      5   3337
PS D:\VS code\EDS makeup>
```

4. From the Funny Votes by Stars sheet, determine which star rating has the highest average funny votes.

```python
# Load the entire Excel file once.
xls = pd.ExcelFile("yelp_analysis.xlsx")

funny_votes_df = pd.read_excel(xls, sheet_name="Funny Votes by Stars")

#From the “Funny Votes by Stars” sheet, determine which star rating has the highest average funny votes.

highest_funny = funny_votes_df.loc[funny_votes_df["Avg Funny Votes"].idxmax(), "stars"]
print("Star rating with the highest average funny votes:", highest_funny)
print("Funny Votes by Stars:")
print(funny_votes_df)
```

```
PS D:\VS code\EDS makeup> & C:/Users/Arjun/AppData/Local/Programs/Python/Python313/python.exe "d:/VS code/EDS makeup/4.py"
Star rating with the highest average funny votes: 1
Funny Votes by Stars:
   stars  Avg Funny Votes
0      1         1.056075
1      2         0.875944
2      3         0.694730
3      4         0.670448
4      5         0.608631
PS D:\VS code\EDS makeup>
```

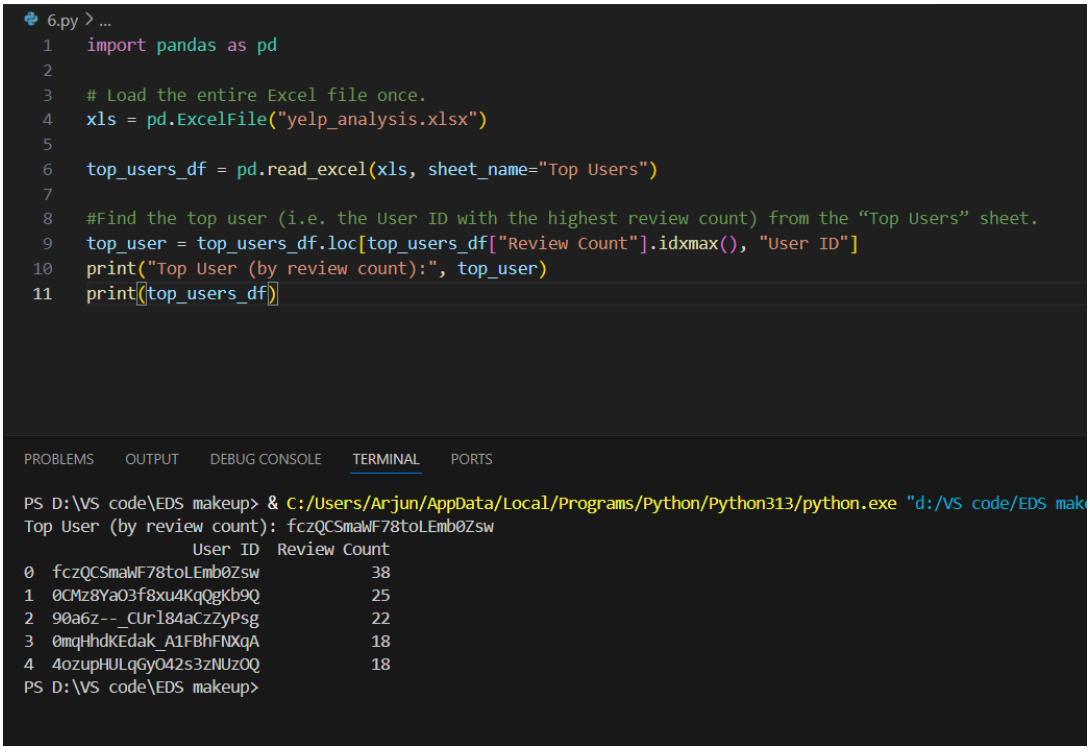
5. Identify the year with the highest number of reviews from the Reviews Per Year sheet.



```
5.py > ...
1   import pandas as pd
2
3   # Load the entire Excel file once.
4   xls = pd.ExcelFile("yelp_analysis.xlsx")
5
6   reviews_year_df = pd.read_excel(xls, sheet_name="Reviews Per Year")
7
8   #Identify the year with the highest number of reviews from the “Reviews Per Year” sheet.
9   year_max_reviews = reviews_year_df.loc[reviews_year_df["Review Count"].idxmax(), "Year"]
10  print("Year with the highest number of reviews:", year_max_reviews)
11  print(reviews_year_df)
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS D:\VS code\EDS makeup> & C:/Users/Arjun/AppData/Local/Programs/Python/Python313/python.exe "d:/VS code/EDS makeup/5.py"
Year with the highest number of reviews: 2012
   Year  Review Count
0  2005             4
1  2006            55
2  2007           285
3  2008           765
4  2009          1171
5  2010          1852
6  2011          2791
7  2012          3025
8  2013            52
PS D:\VS code\EDS makeup>
```

6. Find the top user (i.e., the User ID with the highest review count) from the Top Users sheet.



```
6.py > ...
1   import pandas as pd
2
3   # Load the entire Excel file once.
4   xls = pd.ExcelFile("yelp_analysis.xlsx")
5
6   top_users_df = pd.read_excel(xls, sheet_name="Top Users")
7
8   #Find the top user (i.e. the User ID with the highest review count) from the “Top Users” sheet.
9   top_user = top_users_df.loc[top_users_df["Review Count"].idxmax(), "User ID"]
10  print("Top User (by review count):", top_user)
11  print(top_users_df)
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS D:\VS code\EDS makeup> & C:/Users/Arjun/AppData/Local/Programs/Python/Python313/python.exe "d:/VS code/EDS mak
Top User (by review count): fczQCSmaWF78toLEmb0Zsw
                  User ID  Review Count
0  fczQCSmaWF78toLEmb0Zsw            38
1  0CMz8YaO3f8xu4KqQgKb9Q            25
2  90a6z--_CUrl84aCzZyPsg            22
3  0mqHhdKEdak_A1FBhFNXqA            18
4  4ozupHULqGyO42s3zNUzOQ            18
PS D:\VS code\EDS makeup>
```

7. Find the top business (i.e., the Business ID with the highest review count) from the Top Businesses sheet.

```python
# 7.py
import pandas as pd

# Load the entire Excel file once.
xls = pd.ExcelFile("yelp_analysis.xlsx")

#Find the top business (i.e. the Business ID with the highest review count) from the “Top Businesses” sheet.
top_biz_df = pd.read_excel(xls, sheet_name="Top Businesses")
# Assuming columns "Business ID" and "Review Count"
top_business = top_biz_df.loc[top_biz_df["Review Count"].idxmax(), "Business ID"]
print("Top Business (by review count):", top_business)
print(top_biz_df)
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS D:\VS code\EDS makeup> & C:/Users/Arjun/AppData/Local/Programs/Python/Python313/python.exe "d:/VS code/EDS makeup/7.py"
Top Business (by review count): JokKtdXU7zXHcr20Lrk29A
              Business ID  Review Count
0  JokKtdXU7zXHcr20Lrk29A            37
1  ntN85eu27C04nwyPa8IHtw            37
2  hW0Ne_HTHEAgGF1rAdmR-g            34
3  WNy1uzcmm_UHmTyR--o5IA            32
4  VVeogjZya58oiTxK7qUjAQ            31
5  V1nEpIRmEa1768oj_tuxeQ            30
6  SDwYQ6eSu1htn8vHWv128g            25
7  uKSX1n1RoAzGq4bV8GPHVg            25
8  EWMwV5V9BxNs_U6nNVMeqw            24
9  -sC66z4SO3tR7nFCjfQwuQ            24
PS D:\VS code\EDS makeup>
```

8. Extract and display the correlation between useful votes and stars from the Useful-Stars Correlation sheet.

```python
# 8.py
# Load the entire Excel file once.
xls = pd.ExcelFile("yelp_analysis.xlsx")

# Read the Useful-Stars Correlation sheet with the first column as index.
corr_df = pd.read_excel(xls, sheet_name="Useful-Stars Correlation", index_col=0)

#Extract and display the correlation between useful votes and stars from the “Useful-Stars Correlation” sheet.

corr_useful_stars = corr_df.loc["useful", "stars"]
print("Correlation between 'useful' votes and 'stars':", corr_useful_stars)
print("Correlation Matrix:")
print(corr_df)
```

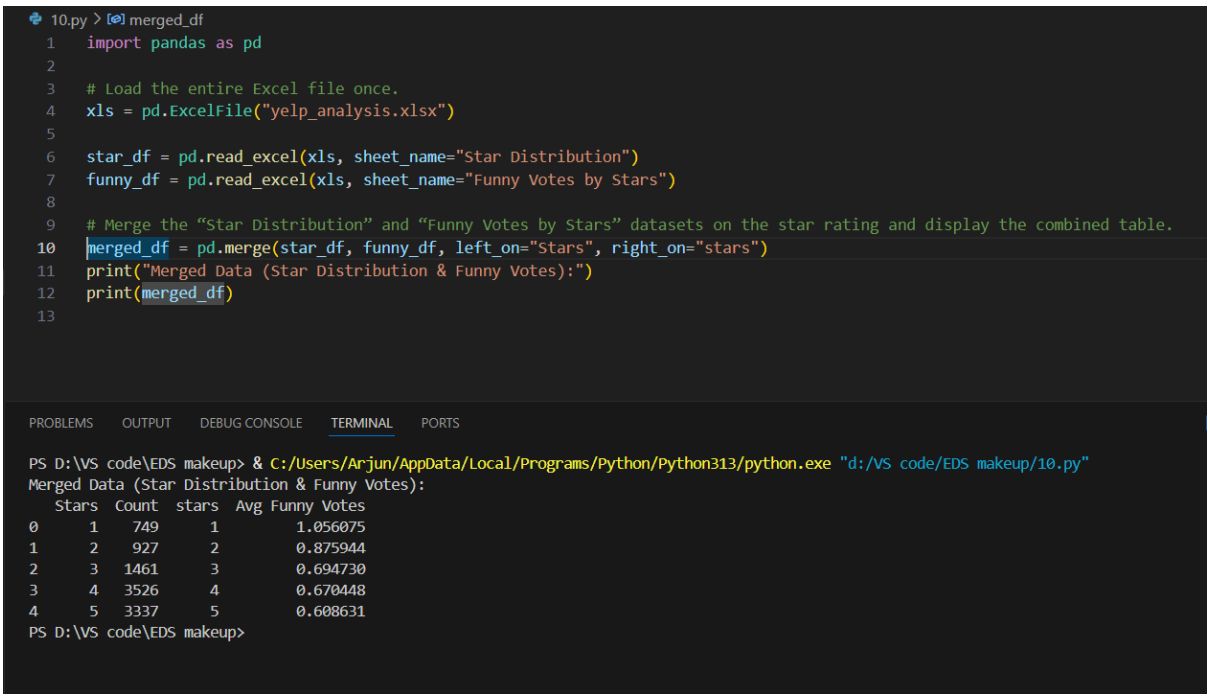
```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

PS D:\VS code\EDS makeup> & C:/Users/Arjun/AppData/Local/Programs/Python/Python313/python.exe "d:/VS code/EDS makeup/8.py"
Correlation between 'useful' votes and 'stars': -0.0234789551276939
Correlation Matrix:
          useful     stars
useful  1.000000 -0.023479
stars  -0.023479  1.000000
PS D:\VS code\EDS makeup>
```

9. Using the Star Distribution data, calculate the percentage contribution of each star rating toward the total number of reviews.

```python
import pandas as pd

# Load the entire Excel file once.
xls = pd.ExcelFile("yelp_analysis.xlsx")

# Re-read the star distribution in case you need to refresh it
star_dist_df = pd.read_excel(xls, sheet_name="Star Distribution")

#Using the “Star Distribution” data, calculate the percentage contribution of each star rating toward the total number of reviews.

total_reviews = star_dist_df["Count"].sum()
star_dist_df["Percentage"] = (star_dist_df["Count"] / total_reviews) * 100

print("Star Distribution with Percentage Contribution:")
print(star_dist_df)
```

```
PS D:\VS code\EDS makeup> & C:/Users/Arjun/AppData/Local/Programs/Python/Python313/python.exe "d:/VS code/EDS makeup/9.py"
Star Distribution with Percentage Contribution:
   Stars  Count  Percentage
0      1    749        7.49
1      2    927        9.27
2      3   1461       14.61
3      4   3526       35.26
4      5   3337       33.37
PS D:\VS code\EDS makeup>
```

10. Merge the Star Distribution and Funny Votes by Stars datasets on the star rating and display the combined table.



```python
import pandas as pd

# Load the entire Excel file once.
xls = pd.ExcelFile("yelp_analysis.xlsx")

star_df = pd.read_excel(xls, sheet_name="Star Distribution")
funny_df = pd.read_excel(xls, sheet_name="Funny Votes by Stars")

# Merge the “Star Distribution” and “Funny Votes by Stars” datasets on the star rating and display the combined table.
merged_df = pd.merge(star_df, funny_df, left_on="Stars", right_on="stars")
print("Merged Data (Star Distribution & Funny Votes):")
print(merged_df)
```

```
PS D:\VS code\EDS makeup> & C:/Users/Arjun/AppData/Local/Programs/Python/Python313/python.exe "d:/VS code/EDS makeup/10.py"
Merged Data (Star Distribution & Funny Votes):
   Stars  Count  stars  Avg Funny Votes
0      1    749      1         1.056075
1      2    927      2         0.875944
2      3   1461      3         0.694730
3      4   3526      4         0.670448
4      5   3337      5         0.608631
PS D:\VS code\EDS makeup>
```