Name: Varad Moholkar

Roll no : ET2-81 DIV : ET-2 Batch : E-24 PRN : 202401070210

```
章 65
Untitled3.pynb
                                                                                                                                                                                                                      File Edit View Insert Runtime Tools Help
 Commands
                   + Code + Text
       Problem 1: Find the total number of reviews in the dataset.
   C Import pandas as pd
               # Create a small sample Yelp Hexieus dataset
                    a = {
    HevimeID': [1, 2, 3, 4, 5],
    'UnerID': ['U1, 'U2', 'U1', 'U3', 'UA'],
    'BusinessID': ['81', '82', '81', '82'],
    'Stars': [5, 4, 5, 2, 1],
    'Text': ['Amazing Food', 'Good service', 'sowed the ambiance', 'Foor experience', 'very bad'],
    'UnerU1': [10, 5, 8, 3, 1],
    'UnerU1': [10, 5, 8, 3, 1],
    'Good': [5, 2, 4, 2, 8, 1],
    'Good': [5, 2, 4, 2, 8, 1],

                    'Cool': [5, 2, 4, 1, 0]
              # Create Outainame
             df = pd.DataFrame(data)
              # If Date is meeded as datetim
             df['Bate'] = pd.to_datetime(df['Bate'])
              total_reviews - len(df)
              print(total_reviews)
       ⊕ 5
       Problem 2: Find the average star rating across all reviews.

∑ [3] import pandas as pd
```

```
V 100 - A
9. Commands + Code + Text
H
       Problem 2: Find the average star rating across all reviews.
0
   Z O inport pandas as pd
43
          # Create a small sample velp Meriters dataset.
               (x)
©u.
"Cool" | [%, 2, 4, 1, 0]
           of - pd.Datafrom(data)
           # If Date is monded as deterion:

df['date'] = pd.to_datelion(df['date'])

average_stars = df['diate'].menn()

print(average_stars)
       D 3.4
```



```
CO & Untitled3.jpynb $\phi \triangle \triangle
                                                                                                                                                                                                                                                                                                                                                                   9
                                                                                                                                                                                                                                                                             File Edit View Insert Runtime Tools Help
                                                                                                                                                                                                                                                                                                                                                                   0
                                                                                                                                                                                                                                                                                                                    V 100 - ^
 5, Comments + Code + Text
                                                                                                                                                                                                                                                                                                                                                                   0
                                                                                                                                                                                                                                                                                                                                                                   (2)
擅
                                                                                                                                                                                                                                                                                                                                                                   101
0
            Problem 4: Find the number of unique users who wrote reviews.
43

≤ ○ Import pandes as pd

                     # Create a small sample velp Newless dataset
data = {
    NewlessID's [3, 2, 3, 4, 5],
    'UserID's ['Us', 'Us', 'Us', 'Us', 'Us'],
    'DesimersID's ['Us', 'Us', 'Us', 'Us', 'Us'],
    'Stars' [5, 5, 2, 3],
    'Stars' [5, 5, 2, 3],
    'Stars' [5, 4, 5, 2, 3],
    'Date | ['202-01-01', '202-01-02', 'tord the adminion', 'Poor experience', 'Very had'],
    'Useful': ['202-01-01', '202-01-02', 'tord-03-01', '2025-01-01', '2025-01-00'],
    'Useful': [10, 5, 9, 2, 3],
    'Coul': [5, 2, 4, 1, 0]
{x}
60
0
                     # Cruste Bataframe

df = pd.Dataframe(data)

unique_users = df['Uner10'].munique()

print(unique_users)
              E5 4
```

```
Problem 5: Find how many reviews have a 5-star rating.

Import pandes as pd

# Create a small sample Velp Becies dataset

data = {

**Meximan(*) {1, 2, 3, 4, 4, 3},

**Out = (*, *, *, *, *, *, *, *),

**Star's {3, 4, 5, 4, 1},

**Star's {3, 4, 5, 6, 7, 8, 7, 8},

**Star's {4, 5, 5, 8, 7, 8, 1},

**Star's {5, 5, 8, 7, 8, 1},

**Star's {5, 8, 7, 8, 8, 1},

**Star's {5, 8,
```

Problem & Find the business with the highest number of 5-star reviews. C O Import pandas as pêdefa = {
 NewlewED': [1, 2, 3, 4, 5],
 NewlewED': [1, 2, 3, 4, 5],
 NewlewED': ['MI, 'MI, 'MI', 'MI', 'MI'],
 NewlewEED': ['MI', 'MI', 'MI', 'MI'],
 NewlewEED': ['MI', 'MI', 'MI', 'MI'],
 NewlewEED': ['MI', 'MI', 'MI', 'MI'],
 NewlewEED': [MI', 'MI', 'MI', 'MI', 'MI'],
 NewlewEED': ['MI', 'MI', # Create a swall cample Yelp Newless dataset # Create DataFrame df = pd.OutaFrame(data) top_besiness = Of[df['stars'] == 3]['Business20'].value_counts().idemax() print(top_besiness) 225 mi Problem 7: Find the percentage of reviews rated as 1-star. C O import pundes as pd # Create a small sample Velp Reviews datases.

data = {
 "ReviewED': [1, 3, 3, 4, 5],
 "UserED': ['01', '02', '02', '02'],
 "BoolnestED': ['01', '02', '02', '02'],
 "BoolnestED': ['01', '02', '02', '02'],
 "Stare': [5, 4, 5, 2, 3],
 "Set': ['020', 01.01', '025', '100'],
 "Data': ['020', 01.01', '020', # Create a small sample Velp Nevlews dataset W Create D df = pt.bataframe(data)
une_star_percentage = (df['stars'].value_counts(normalize=frue)[z]) * ino
print(one_star_percentage) 39 20.0 Problem 8: Find the average length of review text. O import pundes as pd # Create a Small Sample Yelp Reviews dataset.

data = {
 "ReviewED': {1, 2, 3, 4, 5},
 "OwnerED': {'01, '02', '01', '02', '04'},
 "OwnerED': {'01, '02', '01', '02', '02'},
 "OwnerED': {'01, '02', '01', '02', '02'},
 "OwnerED': {'3, 4, 3, 2, 1},
 "Stars': {'Amazing food', 'base Service', 'Loved the ambience', 'Puber saperience', 'bery bad'},
 "Date': {'Amazing food', 'base Service', 'Loved the ambience', 'Puber saperience', 'bery bad'},
 "Date': {'Amazing food', 'base Service', '2025-03-03', '2025-03-04'},
 "Date': {'Amazing food', 'base Service', '2025-03-03', '2025-03-04'},
 "Owner': {'1, 0, 3, 2, 2, 1},
 "Foody': {1, 0, 3, 2, 0, 1},
 "Cool': {5, 2, 4, 3, 0} # Create a small sample velp Nevines dataset. # Create Datain df = pd.bataFrame(data)
average_text_length = df['fext'].apply(len).mean() print(average_text_length) T 13.0

```
Problem 10: Find the average number of "Funny" votes per review.
```

```
import pandes as gd

# Create a small sample Velp Hevimes dataset

data = {
    "ResimmED' [ 1, 2, 3, 4, 5],
    "Year of [ 1, 1, 2, 3, 4, 5],
    "Year of [ 1, 1, 2, 3, 4, 5],
    "Year of [ 1, 1, 2, 3, 3, 3, 3],
    "Star' | [ 1, 4, 5, 2, 4],
    "Test | [ Nancing food', Scool service', "Loved the addisons', "Poor experience', "Very bad'],
    "Nate | [ 1, 1, 2, 3, 3, 3, 4],
    "Year of [ 1, 1, 2, 3, 3, 4],
    "Year of [ 1, 2, 3, 3, 4, 5]
}

# Create Intairase

df = pd.Datarrase(data)

wersage, furny, votee = df[ Young'].mean()
print(wersage, furny, votee)

### 0.8
```

Problem 11: Find how many reviews received more than 10 useful votes.

```
Problem 14: Calculate total useful votes across all reviews.
```

```
import pandss as pd

s Create a small sample yelp Reviews dutaset
data = {
    "ReviewDo": [1, 2, 3, 4, 5],
    "UserD": ['U1', 'U2', 'U1', 'U3', 'U4'],
    "UserD": ['U1', 'U2', 'U1', 'U3', 'U4'],
    "UserD": ['Analing food', 'U2', '
```

Problem 15: Group reviews by star rating and find the average number of useful votes.

Problem 16: Find the user who wrote the most reviews.

```
# (route a small sample velp mortage dataset data = {

| "Mew.LandD's [1, 3, 3, 4, 3],
| "bourtD's ["ot', 'uz', 'uz', 'uz', 'uz'],
| "bourtD's ["ot', 'uz', 'uz', 'uz', 'uz'],
| "bourtD's ["ot', 'uz', 'uz', 'uz', 'uz'],
| "bourt is ["ot', 'uz', 'uz', 'uz', 'uz'],
| "bourt is ["ot', 'uz', 'uz', 'uz', 'uz', 'uz'],
| "bourt is ["ote', 'uz', 'uz', 'uz', 'uz', 'uz'],
| "out' is ["ote', 'uz', '
```

Problem 17: Find the number of reviews where the review text length is greater than 500 characters.

```
# Create a small sample with Sovieus dataset

data v {
    "SeviceOP': {1, 2, 3, 4, 5}, ..., 'se'},
    "severo': ['Ur, 2, 3, 4, 5], ..., 'se'],
    "severo': ['Ur, 2, 3, 4, 5], ..., 'se'],
    "stars': [5, 4, 7, 2, 1],
    "stars': [5, 4, 7, 2, 1],
    "stars': [5, 4, 7, 2, 1],
    "stars': ['Assaing Stod': Good service', 'Loved the ambiance', 'Four emperience', 'Very bas'],
    "bas': ['2025-01-01', '2025-01-02', '2025-01-01', '2025-01-03', '2025-01-04'],
    "service ['1, 0, 5, 9, 2, 1],
    "cool': [5, 2, 4, 1, 0]
}
                                   on 16: Find the ar
C O import pandas as pd
                       # Criste DataFrame (data)

dF = pd.DataFrame(data)

aug_cool_by_stars = df.groupby("Stars")|"Cool"|.mean()

print(avg_cool_by_stars)
        3 0.0
2 1.0
4 2.0
5 4.5
Name: Cool, Stype: Float64
       Problem 19: Find the earliest and latest review dates.
                 # (costs a small sample velp ferices dataset data = {
    "ReviseBP's [1, 2, 3, 4, 5],
    "UserBP's ['Us', 'Us', 'Us', 'Us', 'Us'],
    "UserBP's ['Us', 'Us', 'Us', 'Us', 'Us'],
    "UserBP's ['Us', 'Us', 'Us', 'Ss', 'Bs'],
    "Stars' [5, 4, 7, 2, 3],
    "Test's [5, 4, 7, 2, 3],
    "Test's ['Assiding food', Good service', 'Lowed the ambiance', 'Wood esperience', 'very bad'],
    "UserBul's ['Assiding food', 'Good service', 'Lowed the ambiance', 'Wood esperience', 'very bad'],
    "UserBul's ['Assiding food', 'Good service', 'Lowed the ambiance', 'Wood esperience', 'very bad'],
    "UserBul's ['Assiding food', 'Good service', 'Lowed the ambiance', 'Wood esperience', 'very bad'],
    "UserBul's ['Assiding food', 'Good service', 'Lowed the ambiance', 'Wood esperience', 'very bad'],
    "UserBul's ['Assiding food', 'Good service', 'Lowed the ambiance', 'Wood esperience', 'very bad'],
    "UserBul's ['Assiding food', 'Good service', 'Lowed the ambiance', 'Wood esperience', 'very bad'],
    "UserBul's ['Assiding food', 'Good service', 'Lowed the ambiance', 'Wood esperience', 'very bad'],
    "UserBul's ['Assiding food', 'Good service', 'Lowed the ambiance', 'Wood esperience', 'very bad'],
    "UserBul's ['Assiding food', 'Good service', 'Lowed the ambiance', 'Wood esperience', 'very bad'],
    "UserBul's ['Assiding food', 'Good service', 'Lowed the ambiance', 'Wood esperience', 'very bad'],
    "UserBul's ['Assiding food', 'Good service', 'Lowed the ambiance', 'Wood esperience', 'very bad'],
    "UserBul's ['Assiding food', 'Good service', 'Lowed the ambiance', 'Wood esperience', 'Lowed the ambiance', 'Lo
        O Import pandas as pd
                  s (rests totalrame

of = pd.Ostarrame(data)

carliest = of ("osto").els()

latest = of ("osto").els()

print("carliest" (warliest), Latest) (latest)")
        ➡ tarliest: 2025-01-01, Latest: 2025-01-04
          Problem 20: Find the review(s) that received both the most Funny and Cool votes
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               ↑ + + ∞ 円 章 🛭 日 1
 O taport pandas as pd
                       0 Create Satarrame
df = pd.Catarrame(data)
df['Note'] = pd.to_datetime(df['Note'])
                     # Add raw column; Intal Votes (Funny + Cool)
df["TotalVotes"] = df["Funny"] + df["Cool"]
                         # Find review with maximum IntalVetes
most_funey_coal = df[df["totalVetes"] -- df["TotalVetes"].max[)]
print(most_funey_coal)
          Furny Caol TotalVotes
0 3 5 6
2 2 4 6
```

20 of 22

Thank You!!!