Import Library

```
In [1]: import pandas as pd
        from collections import Counter
        import mysql.connector as sql
```

Load Data from MySQL DB

```
In [3]: db_connection = sql.connect(host='remotemysql.com', database='Z17DZVttfB', user='Z17DZVttfB', password='7Uw9X4gOto')
        db_cursor = db_connection.cursor()
        db_cursor.execute('SELECT * FROM simulation')
        table_rows = db_cursor.fetchall()
```

```
In [11]: df = pd.DataFrame(table_rows)
         df.columns = df.iloc[0]
         df = df[1:]
```

Load Data from CSV

```
In [9]: dataset = "reviews_2.csv"
        DataFrame = pd.read_csv(dataset, engine='python', error_bad_lines=False, delimiter=";")
        DataFrame.head(10)
```

Out[9]:

	title	bookID	author	rating	ratingsCount	reviewsCount	reviewerName	reviewerRating
0	Unauthorized Harry Potter Book Seven News: "Ha...	9	W. Frederick Zimmerman	0,175694444	22	1	Sageri	2
1	Harry Potter Boxed Set, Books 1-5 (Harry Potte...	8	J.K. Rowling	0,220138889	34107	156	Danit	3
2	Harry Potter and the Sorcerer's Stone	3	J.K. Rowling	04.44	4911929	77741	Sageri	3
3	Harry Potter and the Half-Blood Prince (Harry ...	1	J.K. Rowling	04.54	1810404	28053	Cahyono	5
4	Harry Potter and the Order of the Phoenix (Har...	2	J.K. Rowling	04.47	1862749	29308	Teddy	3
5	Harry Potter and the Chamber of Secrets (Harry...	4	J.K. Rowling	04.38	1936698	35055	Elsa	1
6	The Harry Potter Collection (Harry Potter, #1-6)	7	J.K. Rowling	0,217361111	26702	909	Cahyono	3
7	Harry Potter and the Prisoner of Azkaban (Harr...	5	J.K. Rowling	04.53	2000827	37005	Choddy	2
8	Charlie Oink (Easy Peasy People)	8053	Roger Hargreaves	04.33	3	0	Rifki	3
9	Harry Potter and the Goblet of Fire (Harry Pot...	6	J.K. Rowling	04.53	1897821	31817	Abdul Aziz	2

PreProcessing Part 1 : Filtering Column

```
In [3]: DataFrame = DataFrame.filter(items=['reviewerName', 'title', 'reviewerRating'])
        DataFrame.head(10)
```

Out[3]:

	reviewerName	title	reviewerRating
0	Sageri	Unauthorized Harry Potter Book Seven News: "Ha...	2
1	Danit	Harry Potter Boxed Set, Books 1-5 (Harry Potte...	3
2	Sageri	Harry Potter and the Sorcerer's Stone	3
3	Cahyono	Harry Potter and the Half-Blood Prince (Harry ...	5
4	Teddy	Harry Potter and the Order of the Phoenix (Har...	3
5	Elsa	Harry Potter and the Chamber of Secrets (Harry...	1
6	Cahyono	The Harry Potter Collection (Harry Potter, #1-6)	3
7	Choddy	Harry Potter and the Prisoner of Azkaban (Harr...	2
8	Rifki	Charlie Oink (Easy Peasy People)	3
9	Abdul Aziz	Harry Potter and the Goblet of Fire (Harry Pot...	2

PreProcessing Part 2 : Check Null Value

```
In [4]: column = ['reviewerName', 'title', 'reviewerRating']
        for columns in column:
            missing = DataFrame[columns].isnull().value_counts()
            print(missing)
```

```
False    5000
Name: reviewerName, dtype: int64
False    5000
Name: title, dtype: int64
False    5000
Name: reviewerRating, dtype: int64
```

PreProcessing Part 3 : Remove non-ASCII Char

```
In [5]: DataFrame = DataFrame.dropna(how='any')

        DataFrame = DataFrame[~DataFrame.reviewerName.str.contains(r'[^\x00-\x7F]')]
        DataFrame = DataFrame[~DataFrame.title.str.contains(r'[^\x00-\x7F]')]

        DataFrame = DataFrame.reset_index(drop=True)

        DataFrame.head(10)
```

Out[5]:

	reviewerName	title	reviewerRating
0	Sageri	Unauthorized Harry Potter Book Seven News: "Ha...	2
1	Danit	Harry Potter Boxed Set, Books 1-5 (Harry Potte...	3
2	Sageri	Harry Potter and the Sorcerer's Stone	3
3	Cahyono	Harry Potter and the Half-Blood Prince (Harry ...	5
4	Teddy	Harry Potter and the Order of the Phoenix (Har...	3
5	Elsa	Harry Potter and the Chamber of Secrets (Harry...	1
6	Cahyono	The Harry Potter Collection (Harry Potter, #1-6)	3
7	Choddy	Harry Potter and the Prisoner of Azkaban (Harr...	2
8	Rifki	Charlie Oink (Easy Peasy People)	3
9	Abdul Aziz	Harry Potter and the Goblet of Fire (Harry Pot...	2

```
In [6]: Counter(DataFrame['title'].head(20))
```

```
Out[6]: Counter({'Unauthorized Harry Potter Book Seven News: "Half-Blood Prince" Analysis and Speculation': 1,
                 'Harry Potter Boxed Set, Books 1-5 (Harry Potter, #1-5)': 1,
                 "Harry Potter and the Sorcerer's Stone": 1,
                 'Harry Potter and the Half-Blood Prince (Harry Potter, #6)': 1,
                 'Harry Potter and the Order of the Phoenix (Harry Potter, #5)': 1,
                 'Harry Potter and the Chamber of Secrets (Harry Potter, #2)': 1,
                 'The Harry Potter Collection (Harry Potter, #1-6)': 1,
                 'Harry Potter and the Prisoner of Azkaban (Harry Potter, #3)': 1,
                 'Charlie Oink (Easy Peasy People)': 1,
                 'Harry Potter and the Goblet of Fire (Harry Potter, #4)': 1,
                 'My Country Right or Left: 1940-1943 (The Collected Essays, Journalism & Letters, Vol. 2)': 1,
                 'Why I Write': 1,
                 'Bella Tuscany & Under the Tuscan Sun (2 Book Set)': 1,
                 'Selected Poems of Herman Melville': 1,
                 "The Senator and the Socialite: The True Story of America's First Black Dynasty": 1,
                 'Great Short Works of Herman Melville': 1,
                 'Vance Packard & American Social Criticism': 1,
                 'Blue Like Jazz: Nonreligious Thoughts on Christian Spirituality': 1,
                 'Redburn, White-Jacket, Moby-Dick': 1,
                 'The Hidden Persuaders': 1})
```

Visualize

```
In [7]: df1 = DataFrame.set_index(['reviewerName', 'title']).sort_index()
        df1.head(10)
```

Out[7]:

reviewerName	title	reviewerRating
Abdul Aziz	8 Minutes in the Morning Kit	2
	8 Minutos Por La Manana Para Reducir Caderas Y Muslos/ 8 Minutes in the Morning For Lean Hips and Thin Thighs: Pierde Hasta 10 Cm En Menos De 4 Semanas, ... / Lose up to 10cm in less than 4 weeks	1
	A Catholic Perspective on the Purpose Driven Life	5
	A Guide for Using a Christmas Carol in the Classroom	2
	A History of the County of Middlesex: Volume IV: Elthorne Hundred (Continued) and Gore Hundred (Part)	1
	A People and a Nation: A History of the United States, Dolphin Edition - Complete	3
	A Taste for Death (Adam Dalgliesh #7)	3
	A Visit to Vanity Fair: Moral Essays on the Present Age	3
	A Voyage for Madmen	2
	Across the River: Walter & Oliver's Amazing Adventure	1

Preparation : Conver DataTable to Dict

```
In [8]: d = (DataFrame.groupby('reviewerName')['title','reviewerRating'].apply(lambda x: dict(x.values)).to_dict())
```

Modelling : Euclidean Distance

```
In [9]:  def jarak_similarity(prefs,person1,person2):
             si = {}
             for item in prefs[person1]:
                 if item in prefs[person2]:
                     si[item]=1
             if len(si) == 0:
                 return 0
             sum_of_squares = sum([pow(prefs[person1][item] - prefs[person2][item],2)
                             for item in prefs[person1] if item in prefs[person2]])
             return 1/(1+sum_of_squares)
```

```
In [10]:  def preferensi_mirip(prefs, person, n=10, similarity = jarak_similarity):
              scores = [(similarity(prefs,person,other), other)
                      for other in prefs if other!=person]
              scores.sort()
              scores.reverse()
              return scores[0:n]
```

```
In [11]:  def get_recommendations(prefs, person, n=10, similarity = jarak_similarity):
              totals = {}
              simSums = {}
              for other in prefs:
                  if other == person:
                      continue
                  sim = similarity(prefs, person, other)
                  if sim <= 0:
                      continue
                  for item in prefs[other]:
                      if item not in prefs[person] or prefs[person][item] == 0:
                          totals.setdefault(item,0)
                          totals[item] += prefs[other][item] * sim
                          simSums.setdefault(item,0)
                          simSums[item] += sim
              rankings = [(total/simSums[item], item) for item, total in totals.items()]
              rankings.sort()
              rankings.reverse()
              return rankings[0:n]
```

Result

```
In [12]:  preferensi_mirip(d, 'Witsqadianto')
```

```
Out[12]:  [(1.0, 'Reyhan R'),
          (0.5, 'Teddy'),
          (0.5, 'Abdul Aziz'),
          (0.14285714285714285, 'Randi'),
          (0.14285714285714285, 'Dirga'),
          (0.09090909090909091, 'Riyan'),
          (0.05555555555555555, 'Firdaus B'),
          (0.041666666666666664, 'Elsa'),
          (0.037037037037037035, 'Youga P'),
          (0.014285714285714285, 'Danit')]
```

```
In [13]:  get_recommendations(d, 'Witsqadianto')
```

```
Out[13]:  [(5.000000000000001,
            "Writing Plain Visions: The Mystic Migrant In Joy Kogawa's Obasan"),
          (5.000000000000001,
            'World History in Brief: Major Patterns of Change and Continuity, Volume II: Since 1450'),
          (5.000000000000001, 'V for Vendetta'),
          (5.000000000000001, 'Trinity (Works of Saint Augustine)'),
          (5.000000000000001, 'Timbuktu / Leviathan / Moon Palace'),
          (5.000000000000001,
            'Three Complete Novels: Tim/An Indecent Obsession/The Ladies of Missalonghi'),
          (5.000000000000001, 'Three Complete Novels Deluxe Edition'),
          (5.000000000000001, "This Side of Paradise (World's Best Reading)"),
          (5.000000000000001, 'Theatre: A Way of Seeing (with InfoTrac)'),
          (5.000000000000001, 'Theatre, a Way of Seeing: A Way of Seeing')]
```