# Problem statement - Build a recommender system by using cosine simillarties score.

In [1]:

```python
import pandas as pd
```

In [3]:

```python
book_data = pd.read_csv('book (1).csv')
book_data
```

Out[3]:

|  | Unnamed: 0 | User.ID | Book.Title | Book.Rating |
|---|---|---|---|---|
| **0** | 1 | 276726 | Classical Mythology | 5 |
| **1** | 2 | 276729 | Clara Callan | 3 |
| **2** | 3 | 276729 | Decision in Normandy | 6 |
| **3** | 4 | 276736 | Flu: The Story of the Great Influenza Pandemic... | 8 |
| **4** | 5 | 276737 | The Mummies of Urumchi | 6 |
| **...** | ... | ... | ... | ... |
| **9995** | 9996 | 162121 | American Fried: Adventures of a Happy Eater. | 7 |
| **9996** | 9997 | 162121 | Cannibal In Manhattan | 9 |
| **9997** | 9998 | 162121 | How to Flirt: A Practical Guide | 7 |
| **9998** | 9999 | 162121 | Twilight | 8 |
| **9999** | 10000 | 162129 | Kids Say the Darndest Things | 6 |

10000 rows × 4 columns

In [4]:

```python
book_data_1 = book_data.rename({'User.ID':'User_ID',
                                'Book.Title':'Book_Title',
                                'Book.Rating':'Book_Rating'},axis=1)
book_data_1
```

Out[4]:

|  | Unnamed: 0 | User_ID | Book_Title | Book_Rating |
|---|---|---|---|---|
| **0** | 1 | 276726 | Classical Mythology | 5 |
| **1** | 2 | 276729 | Clara Callan | 3 |
| **2** | 3 | 276729 | Decision in Normandy | 6 |
| **3** | 4 | 276736 | Flu: The Story of the Great Influenza Pandemic... | 8 |
| **4** | 5 | 276737 | The Mummies of Urumchi | 6 |
| **...** | ... | ... | ... | ... |
| **9995** | 9996 | 162121 | American Fried: Adventures of a Happy Eater. | 7 |
| **9996** | 9997 | 162121 | Cannibal In Manhattan | 9 |
| **9997** | 9998 | 162121 | How to Flirt: A Practical Guide | 7 |
| **9998** | 9999 | 162121 | Twilight | 8 |
| **9999** | 10000 | 162129 | Kids Say the Darndest Things | 6 |

10000 rows × 4 columns

In [5]:

```python
book_data_1.Book_Title.unique()
```

Out[5]:

```
array(['Classical Mythology', 'Clara Callan', 'Decision in Normandy', ...,
       'How to Flirt: A Practical Guide', 'Twilight',
       'Kids Say the Darndest Things'], dtype=object)
```

In [6]:

```python
book_data_1.Book_Title.nunique()
```

Out[6]:

9659

In [7]:

```python
book_data_1.User_ID.unique()
```

Out[7]:

```
array([276726, 276729, 276736, ..., 162113, 162121, 162129], dtype=int64)
```

In [8]:

```python
book_data_1.User_ID.nunique()
```

Out[8]:

2182

In [9]:

```python
book_data_1.isna().sum()
```

Out[9]:

```
Unnamed: 0      0
User_ID         0
Book_Title      0
Book_Rating     0
dtype: int64
```

In [10]:

```python
# Conversion of long data in other formate
user_recomm = pd.pivot_table(data = book_data_1, values='Book_Rating',index='User_ID',colum
user_recomm
```

Out[10]:

| Book_Title | Jason, Madison &amp | Other Stories;Merril;1985;McClelland &amp | Repairing PC Drives &amp | '48 | 'O Au No Keia: Voices from Hawai'I's Mahu and Transgender Communities | ...AND THE HORSE HE RODE IN ON : THE PEOPLE V. KENNETH STARR | A Mi |
|---|---|---|---|---|---|---|---|
| **User_ID** | | | | | | | |
| **8** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| **9** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| **10** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| **12** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| **14** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| **...** | ... | ... | ... | ... | ... | ... | |
| **278846** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| **278849** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| **278851** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| **278852** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| **278854** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |

2182 rows × 9659 columns
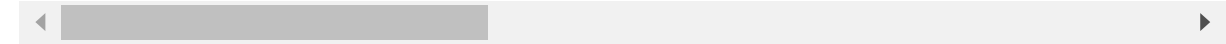
In [11]:

```python
# Replacing the index value with unique User Id
user_recomm.index=book_data_1.User_ID.unique()
user_recomm
```

Out[11]:

| Book_Title | Jason, Madison &amp | Other Stories;Merril;1985;McClelland &amp | Repairing PC Drives &amp | '48 | 'O Au No Keia: Voices from Hawai'I's Mahu and Transgender Communities | ...AND THE HORSE HE RODE IN ON : THE PEOPLE V. KENNETH STARR | A Mi |
|---|---|---|---|---|---|---|---|
| 276726 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 276729 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 276736 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 276737 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 276744 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| ... | ... | ... | ... | ... | ... | ... | |
| 162107 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 162109 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 162113 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 162121 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| 162129 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |

2182 rows × 9659 columns

In [12]:

```python
user_recomm.values
```

Out[12]:

```
array([[0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       ...,
       [0., 0., 0., ..., 7., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.]])
```

In [13]:

```python
from sklearn.metrics import pairwise_distances
```

In [14]:

```python
# Cosine Similarity for Recommendation Engine
user_to_user_cosine = 1-pairwise_distances(user_recomm.values, metric='cosine',)
```

In [15]:

```python
user_to_user_cosine_df = pd.DataFrame(data=user_to_user_cosine, index=book_data_1['User_ID'
user_to_user_cosine_df
```

Out[15]:

|        | 276726 | 276729 | 276736 | 276737 | 276744 | 276745 | 276747 | 276748 | 276751 | 276754 | .. |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|----|
| 276726 | 1.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | .. |
| 276729 | 0.0    | 1.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | .. |
| 276736 | 0.0    | 0.0    | 1.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | .. |
| 276737 | 0.0    | 0.0    | 0.0    | 1.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | .. |
| 276744 | 0.0    | 0.0    | 0.0    | 0.0    | 1.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | .. |
| ...    | ...    | ...    | ...    | ...    | ...    | ...    | ...    | ...    | ...    | ...    | .. |
| 162107 | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | .. |
| 162109 | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | .. |
| 162113 | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | .. |
| 162121 | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | .. |
| 162129 | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | 0.0    | .. |

2182 rows × 2182 columns

In [17]:

```python
#Nullify diagonal values
import numpy as np
np.fill_diagonal(user_to_user_cosine_df.to_numpy(), val = 0)
user_to_user_cosine_df
```

Out[17]:

|        | 276726 | 276729 | 276736 | 276737 | 276744 | 276745 | 276747 | 276748 | 276751 | 276754 | .. |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|----|
| **276726** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | .. |
| **276729** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | .. |
| **276736** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | .. |
| **276737** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | .. |
| **276744** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | .. |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | .. |
| **162107** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | .. |
| **162109** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | .. |
| **162113** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | .. |
| **162121** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | .. |
| **162129** | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | .. |

2182 rows × 2182 columns

In [18]:

```python
# Most similar user
user_to_user_cosine_df.idxmax()
```

Out[18]:

```
276726    276726
276729    276726
276736    276726
276737    276726
276744    276726
           ...
162107    276726
162109    276726
162113    161453
162121    276726
162129    276726
Length: 2182, dtype: int64
```

In [19]:

```python
# Extrating the user having similarity
book_data_1[(book_data_1['User_ID']==276729) | (book_data_1['User_ID']==276726)]
```

Out[19]:

| | Unnamed: 0 | User_ID | Book_Title | Book_Rating |
|---|---|---|---|---|
| 0 | 1 | 276726 | Classical Mythology | 5 |
| 1 | 2 | 276729 | Clara Callan | 3 |
| 2 | 3 | 276729 | Decision in Normandy | 6 |

In [22]:

```python
book_data_1[(book_data_1['User_ID']==162109 ) | (book_data_1['User_ID']== 276726)]
```

Out[22]:

| | Unnamed: 0 | User_ID | Book_Title | Book_Rating |
|---|---|---|---|---|
| 0 | 1 | 276726 | Classical Mythology | 5 |
| 9988 | 9989 | 162109 | The Flower in the Skull | 10 |

In [23]:

```python
book_data_1[(book_data_1['User_ID']==162113  ) | (book_data_1['User_ID']== 161453)]
```

Out[23]:

| | Unnamed: 0 | User_ID | Book_Title | Book_Rating |
|---|---|---|---|---|
| 8959 | 8960 | 161453 | Bread, Tomato, Garlic: Quick Cooking With 3 Ma... | 9 |
| 8960 | 8961 | 161453 | The Ubiquitous Shrimp: From Simple to Exotic, ... | 8 |
| 9989 | 9990 | 162113 | The Cape Ann (Contemporary American Fiction) | 8 |

In [ ]: