Data Science with Python Project

Building user-based recommendation model for Amazon

DESCRIPTION

The dataset provided contains movie reviews given by Amazon customers. Reviews were given between May 1996 and July 2014.

analysis

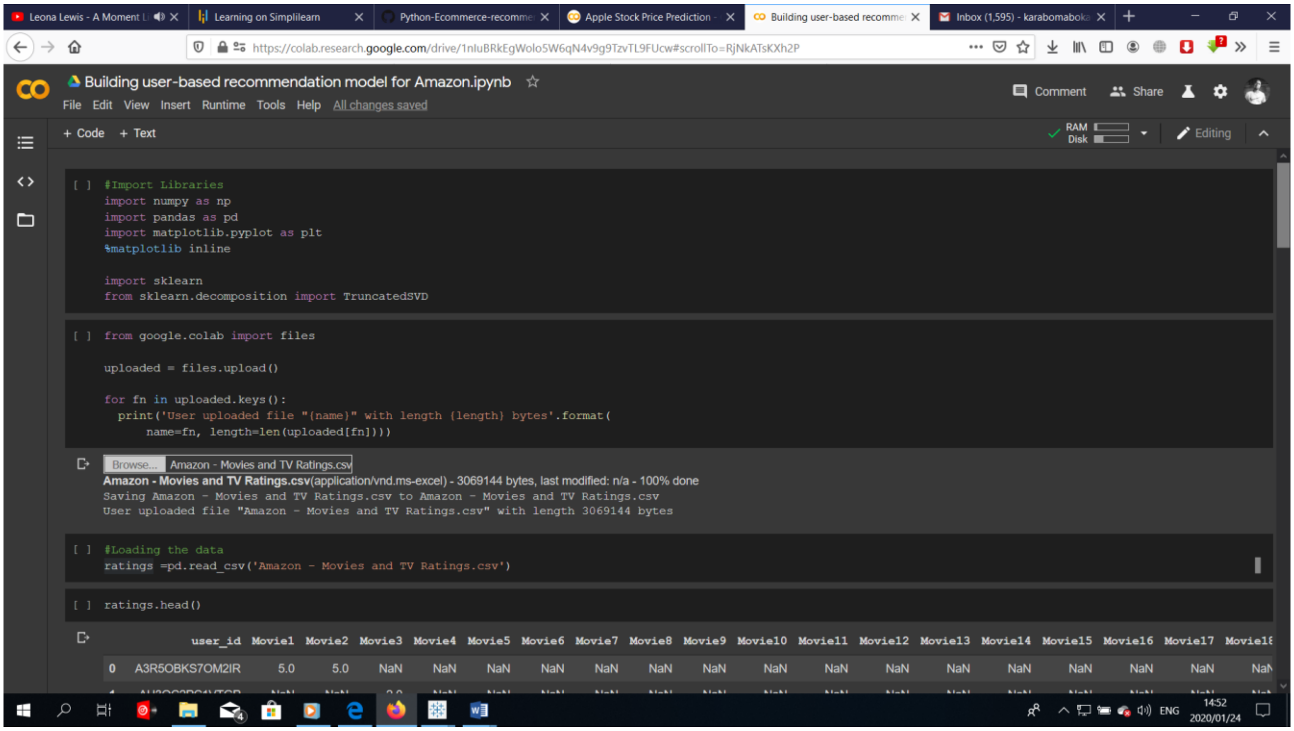
- Exploratory Data Analysis:

- Which movies have maximum views/ratings?
- What is the average rating for each movie? Define the top 5 movies with the maximum ratings.
- Define the top 5 movies with the least audience.

- Recommendation Model: Some of the movies hadn’t been watched and therefore, are not rated by the users. Netflix would like to take this as an opportunity and build a machine learning recommendation algorithm which provides the ratings for each of the users.

- Divide the data into training and test data
- Build a recommendation model on training data
- Make predictions on the test data

Screenshots


Leona Lewis - A Moment L
Learning on Simplilearn
Python-Ecommerce-recomme
Apple Stock Price Prediction
Building user-based recomme
Inbox (1,595) - karabomaboka
https://colab.research.google.com/drive/1nluBRkEgWolo5W6qN4v9g9TzvTL9FUcw#scrollTo=RjNkATsKXh2P
Building user-based recommendation model for Amazon.ipynb
File
Edit
View
Insert
Runtime
Tools
Help
All changes saved
Comment
Share
+ Code
+ Text
RAM
Disk
Editing
ENG
14:52
2020/01/24

```
[ ] #Import Libraries
    import numpy as np
    import pandas as pd
    import matplotlib.pyplot as plt
    %matplotlib inline

    import sklearn
    from sklearn.decomposition import TruncatedSVD
```

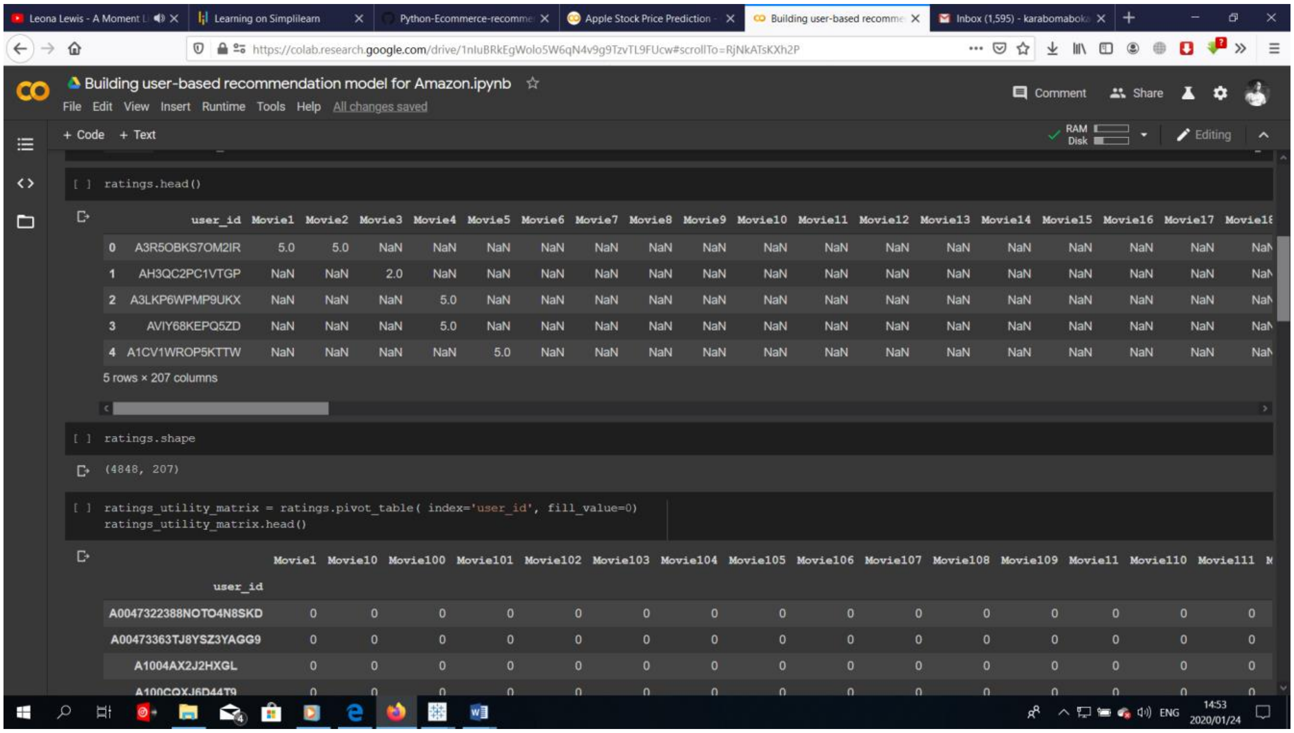
```
[ ] from google.colab import files

    uploaded = files.upload()

    for fn in uploaded.keys():
      print('User uploaded file "{name}" with length {length} bytes'.format(
          name=fn, length=len(uploaded[fn])))
```

```
Browse... Amazon - Movies and TV Ratings.csv
Amazon - Movies and TV Ratings.csv(application/vnd.ms-excel) - 3069144 bytes, last modified: n/a - 100% done
Saving Amazon - Movies and TV Ratings.csv to Amazon - Movies and TV Ratings.csv
User uploaded file "Amazon - Movies and TV Ratings.csv" with length 3069144 bytes
```

```
[ ] #Loading the data
    ratings =pd.read_csv('Amazon - Movies and TV Ratings.csv')
```

```
[ ] ratings.head()
```

	user_id	Movie1	Movie2	Movie3	Movie4	Movie5	Movie6	Movie7	Movie8	Movie9	Movie10	Movie11	Movie12	Movie13	Movie14	Movie15	Movie16	Movie17	Movie1
0	A3R5OBKS7OM2IR	5.0	5.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN


Leona Lewis - A Moment L
Learning on Simplilearn
Python-Ecommerce-recomme
Apple Stock Price Prediction
Building user-based recomme
Inbox (1,595) - karabomaboka
https://colab.research.google.com/drive/1nluBRkEgWolo5W6qN4v9g9TzvTL9FUcw#scrollTo=RjNkATsKXh2P
Building user-based recommendation model for Amazon.ipynb
File
Edit
View
Insert
Runtime
Tools
Help
All changes saved
Comment
Share
+ Code
+ Text
RAM
Disk
Editing
ENG
14:53
2020/01/24

```
[ ] ratings.head()
```

	user_id	Movie1	Movie2	Movie3	Movie4	Movie5	Movie6	Movie7	Movie8	Movie9	Movie10	Movie11	Movie12	Movie13	Movie14	Movie15	Movie16	Movie17	Movie1
0	A3R5OBKS7OM2IR	5.0	5.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	AH3QC2PC1VTGP	NaN	NaN	2.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	A3LKP6WPMP9UKX	NaN	NaN	NaN	5.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	AVIY68KEPQ5ZD	NaN	NaN	NaN	5.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	A1CV1WROP5KTTW	NaN	NaN	NaN	NaN	5.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

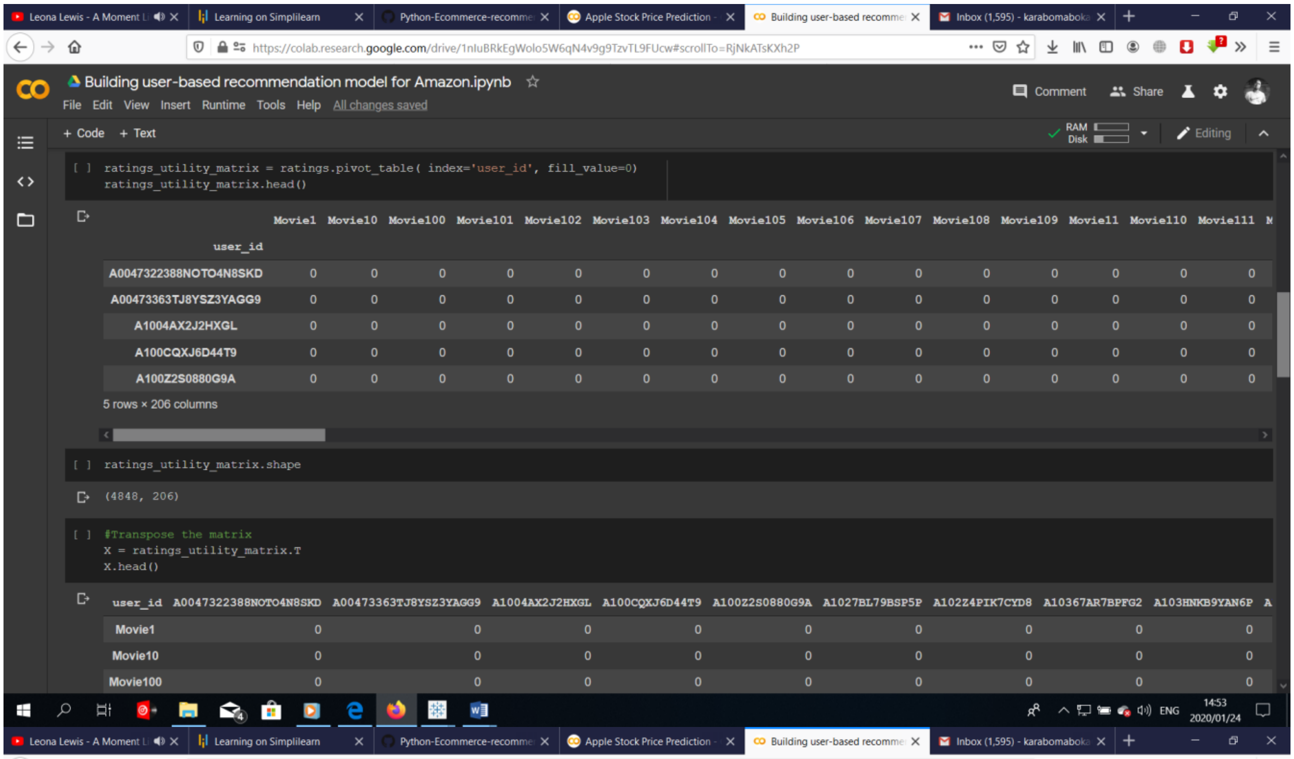
5 rows × 207 columns

```
[ ] ratings.shape
```

```
(4848, 207)
```

```
[ ] ratings_utility_matrix = ratings.pivot_table( index='user_id', fill_value=0)
    ratings_utility_matrix.head()
```

	Movie1	Movie10	Movie100	Movie101	Movie102	Movie103	Movie104	Movie105	Movie106	Movie107	Movie108	Movie109	Movie11	Movie110	Movie111
user_id															
A0047322388NOTO4N8SKD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A00473363TJ8YSZ3YAGG9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A1004AX2J2HXGL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A100COXJ6D44T9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0



Leona Lewis - A Moment L
Learning on Simplilearn
Python-Ecommerce-recomme
Apple Stock Price Prediction
Building user-based recomme
Inbox (1,595) - karabomaboka
https://colab.research.google.com/drive/1nIuBRkEgWolo5W6qN4v9g9TzvTL9FUcw#scrollTo=RjNkATsKXh2P
Building user-based recommendation model for Amazon.ipynb
File
Edit
View
Insert
Runtime
Tools
Help
All changes saved
Comment
Share
+ Code
+ Text
RAM
Disk
Editing
ENG
14:53
2020/01/24

```
ratings_utility_matrix = ratings.pivot_table( index='user_id', fill_value=0)
ratings_utility_matrix.head()
```

	Movie1	Movie10	Movie100	Movie101	Movie102	Movie103	Movie104	Movie105	Movie106	Movie107	Movie108	Movie109	Movie11	Movie110	Movie111
user_id															
A0047322388NOTO4N8SKD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A00473363TJ8YSZ3YAGG9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A1004AX2J2HXGL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A100CQXJ6D44T9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A100Z2S0880G9A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

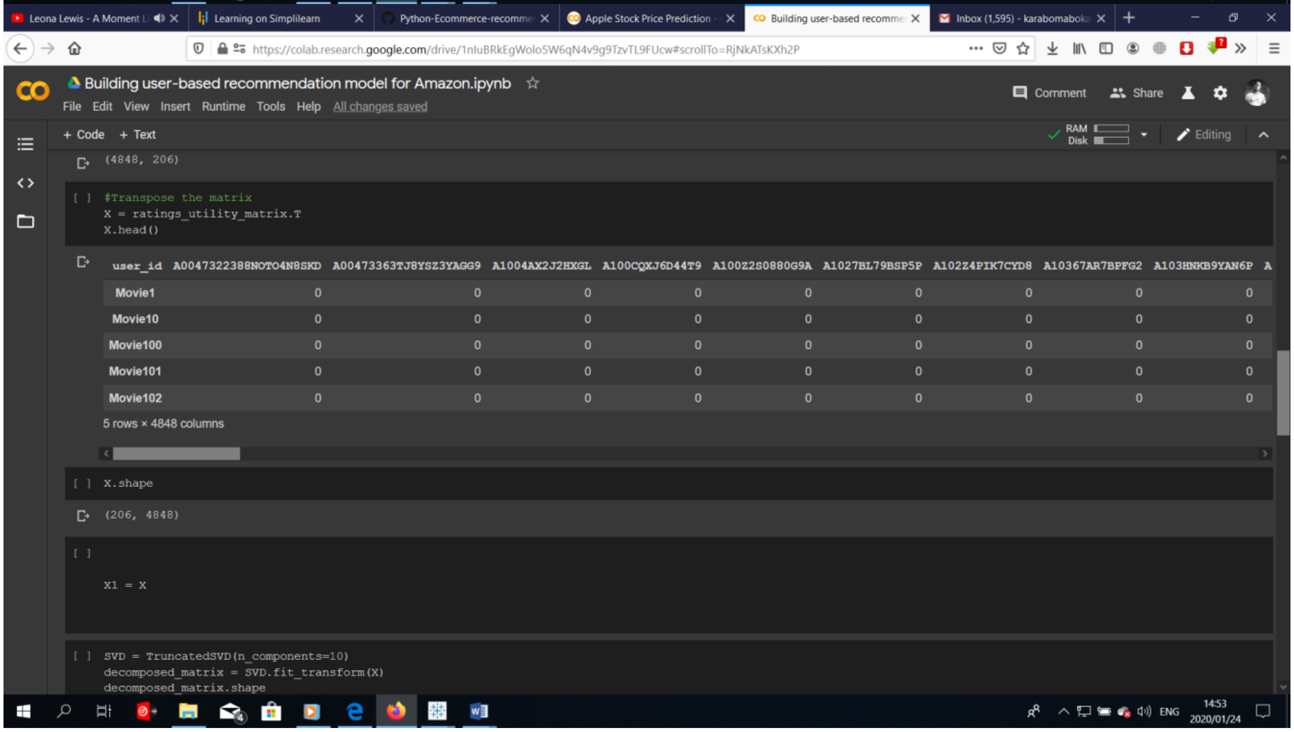
5 rows × 206 columns

```
ratings_utility_matrix.shape
```

```
(4848, 206)
```

```
#Transpose the matrix
X = ratings_utility_matrix.T
X.head()
```

user_id	A0047322388NOTO4N8SKD	A00473363TJ8YSZ3YAGG9	A1004AX2J2HXGL	A100CQXJ6D44T9	A100Z2S0880G9A	A1027BL79BSP5P	A102Z4PIK7CYD8	A10367AR7BPFG2	A103HNKB9YAN6P
Movie1	0	0	0	0	0	0	0	0	0
Movie10	0	0	0	0	0	0	0	0	0
Movie100	0	0	0	0	0	0	0	0	0



Leona Lewis - A Moment L
Learning on Simplilearn
Python-Ecommerce-recomme
Apple Stock Price Prediction
Building user-based recomme
Inbox (1,595) - karabomaboka
https://colab.research.google.com/drive/1nIuBRkEgWolo5W6qN4v9g9TzvTL9FUcw#scrollTo=RjNkATsKXh2P
Building user-based recommendation model for Amazon.ipynb
File
Edit
View
Insert
Runtime
Tools
Help
All changes saved
Comment
Share
+ Code
+ Text
RAM
Disk
Editing
ENG
14:53
2020/01/24

```
(4848, 206)
```

```
#Transpose the matrix
X = ratings_utility_matrix.T
X.head()
```

user_id	A0047322388NOTO4N8SKD	A00473363TJ8YSZ3YAGG9	A1004AX2J2HXGL	A100CQXJ6D44T9	A100Z2S0880G9A	A1027BL79BSP5P	A102Z4PIK7CYD8	A10367AR7BPFG2	A103HNKB9YAN6P
Movie1	0	0	0	0	0	0	0	0	0
Movie10	0	0	0	0	0	0	0	0	0
Movie100	0	0	0	0	0	0	0	0	0
Movie101	0	0	0	0	0	0	0	0	0
Movie102	0	0	0	0	0	0	0	0	0

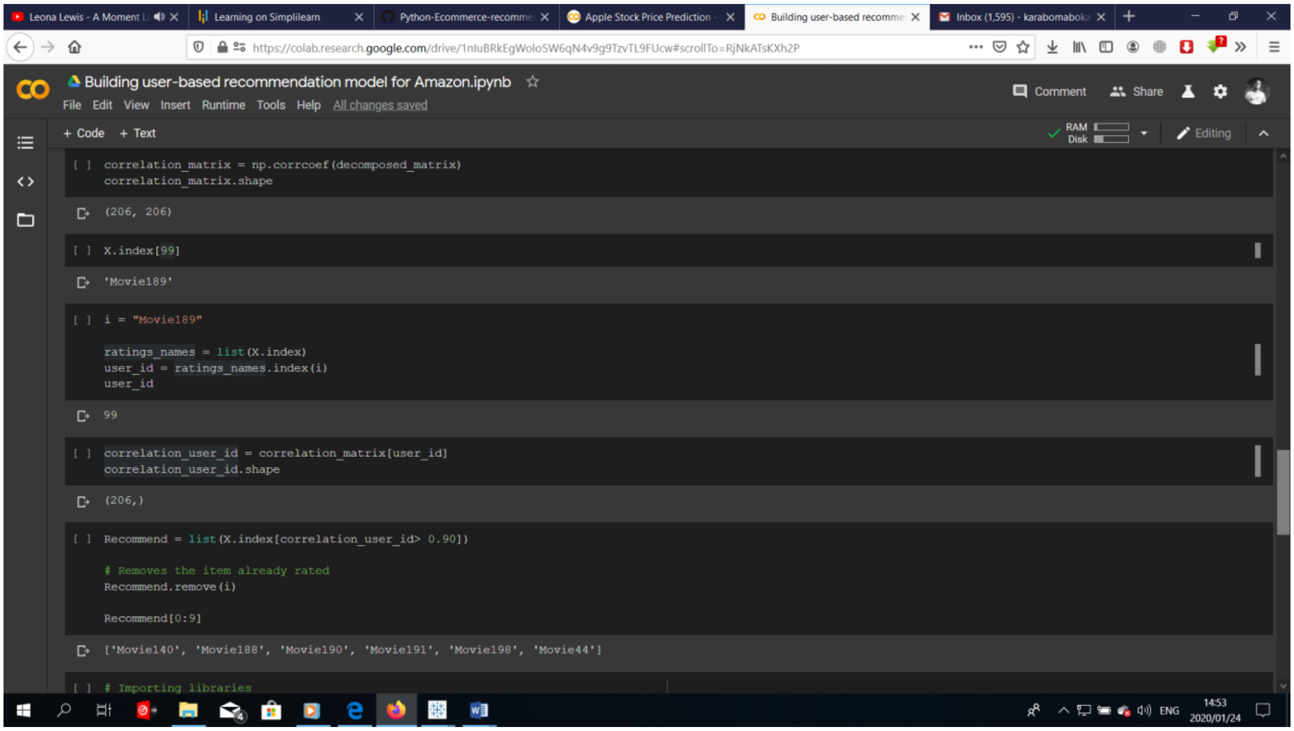
5 rows × 4848 columns

```
X.shape
```

```
(206, 4848)
```

```


X1 = X


```

```
SVD = TruncatedSVD(n_components=10)
decomposed_matrix = SVD.fit_transform(X)
decomposed_matrix.shape
```


Leona Lewis - A Moment L
Learning on Simplilearn
Python-Ecommerce-recomme
Apple Stock Price Prediction -
Building user-based recomme
Inbox (1,595) - karabomaboka
https://colab.research.google.com/drive/1nluBRkEgWolo5W6qN4v9g9TzvTL9FUcw#scrollTo=RjNkATsKXh2P
Building user-based recommendation model for Amazon.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Editing
ENG
14:53
2020/01/24

```
correlation_matrix = np.corrcoef(decomposed_matrix)
correlation_matrix.shape
```

```
(206, 206)
```

```
X.index[99]
```

```
'Movie189'
```

```
i = "Movie189"

ratings_names = list(X.index)
user_id = ratings_names.index(i)
user_id
```

```
99
```

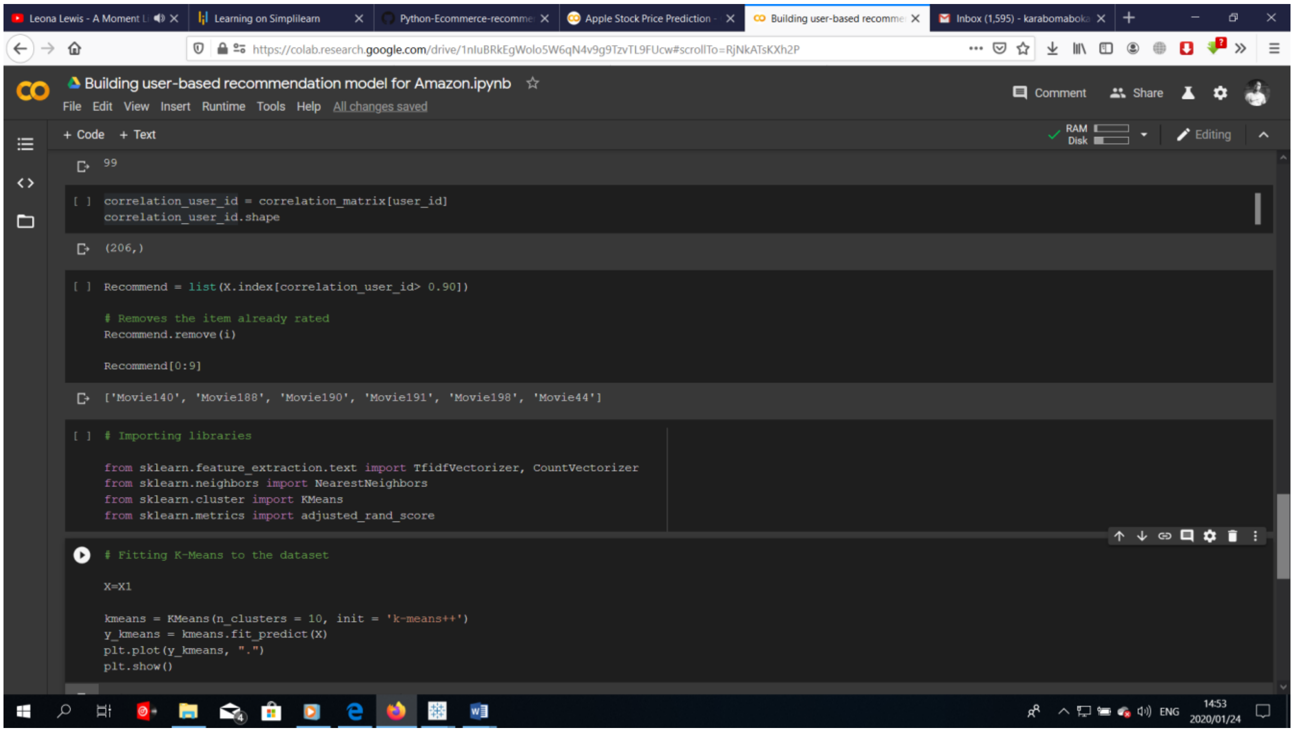
```
correlation_user_id = correlation_matrix[user_id]
correlation_user_id.shape
```

```
(206,)
```

```
Recommend = list(X.index[correlation_user_id> 0.90])

# Removes the item already rated
Recommend.remove(i)

Recommend[0:9]
```

```
['Movie140', 'Movie188', 'Movie190', 'Movie191', 'Movie198', 'Movie44']
```

```
# Importing libraries
```


Leona Lewis - A Moment L
Learning on Simplilearn
Python-Ecommerce-recomme
Apple Stock Price Prediction -
Building user-based recomme
Inbox (1,595) - karabomaboka
https://colab.research.google.com/drive/1nluBRkEgWolo5W6qN4v9g9TzvTL9FUcw#scrollTo=RjNkATsKXh2P
Building user-based recommendation model for Amazon.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
+ Code + Text
RAM
Disk
Editing
ENG
14:53
2020/01/24

```
99
```

```
correlation_user_id = correlation_matrix[user_id]
correlation_user_id.shape
```

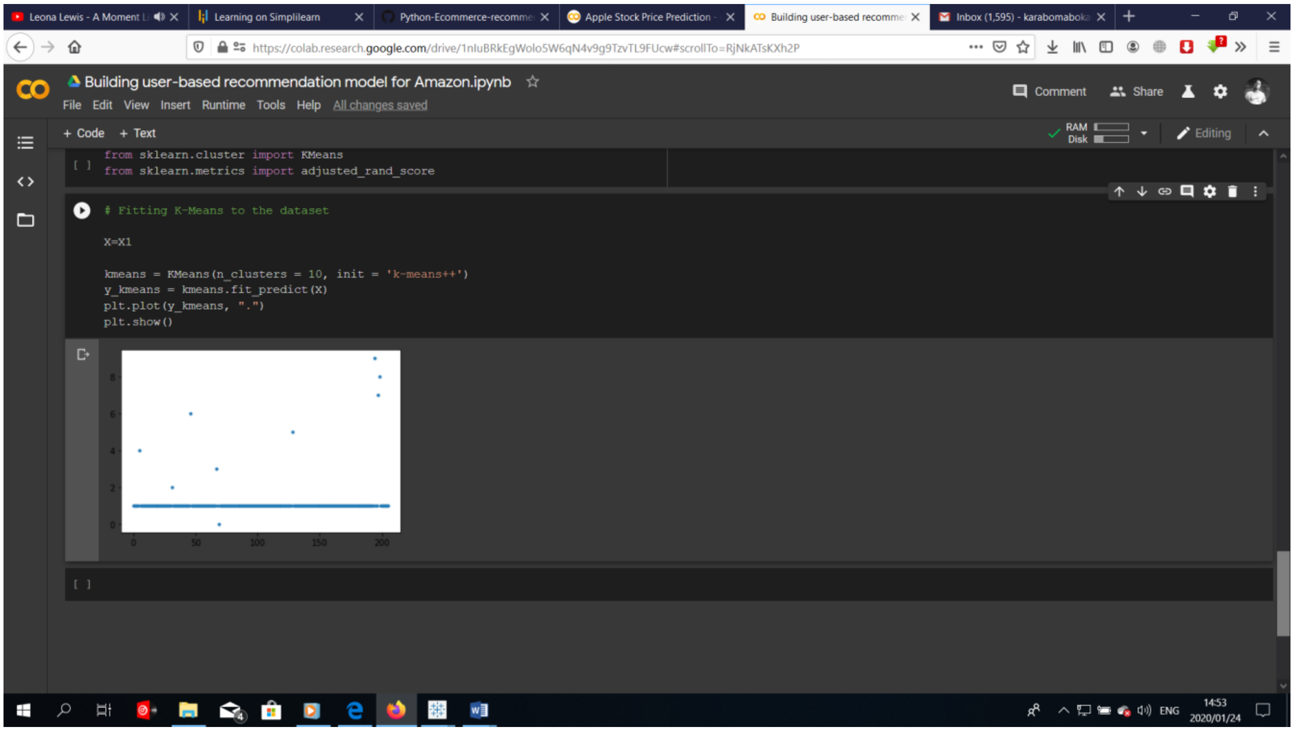
```
(206,)
```

```
Recommend = list(X.index[correlation_user_id> 0.90])

# Removes the item already rated
Recommend.remove(i)

Recommend[0:9]
```

```
['Movie140', 'Movie188', 'Movie190', 'Movie191', 'Movie198', 'Movie44']
```

```
# Importing libraries

from sklearn.feature_extraction.text import TfidfVectorizer, CountVectorizer
from sklearn.neighbors import NearestNeighbors
from sklearn.cluster import KMeans
from sklearn.metrics import adjusted_rand_score
```

```
# Fitting K-Means to the dataset

X=X1

kmeans = KMeans(n_clusters = 10, init = 'k-means++')
y_kmeans = kmeans.fit_predict(X)
plt.plot(y_kmeans, ".")
plt.show()
```


Leona Lewis - A Moment L
Learning on Simplilearn
Python-Ecommerce-recomme
Apple Stock Price Prediction
Building user-based recomme
Inbox (1,595) - karabomaboka
https://colab.research.google.com/drive/1nIuBRkEgWolo5W6qN4v9g9TzvTL9FUcw#scrollTo=RjNkATsKXh2P
Building user-based recommendation model for Amazon.ipynb
File
Edit
View
Insert
Runtime
Tools
Help
All changes saved
Comment
Share
+ Code
+ Text
RAM
Disk
Editing
8
6
4
2
0
0
50
100
150
200
ENG
14:53
2020/01/24

```
from sklearn.cluster import KMeans
from sklearn.metrics import adjusted_rand_score
```

```
# Fitting K-Means to the dataset

X=X1

kmeans = KMeans(n_clusters = 10, init = 'k-means++')
y_kmeans = kmeans.fit_predict(X)
plt.plot(y_kmeans, ".")
plt.show()
```