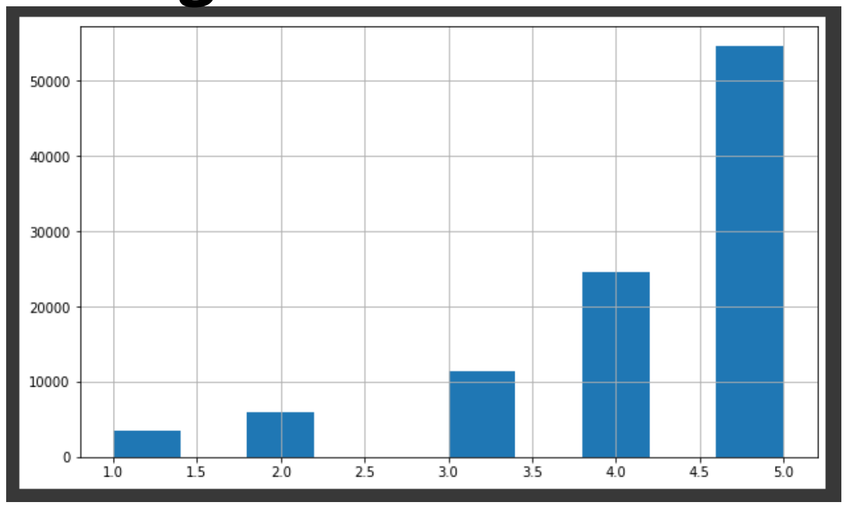
Dataset

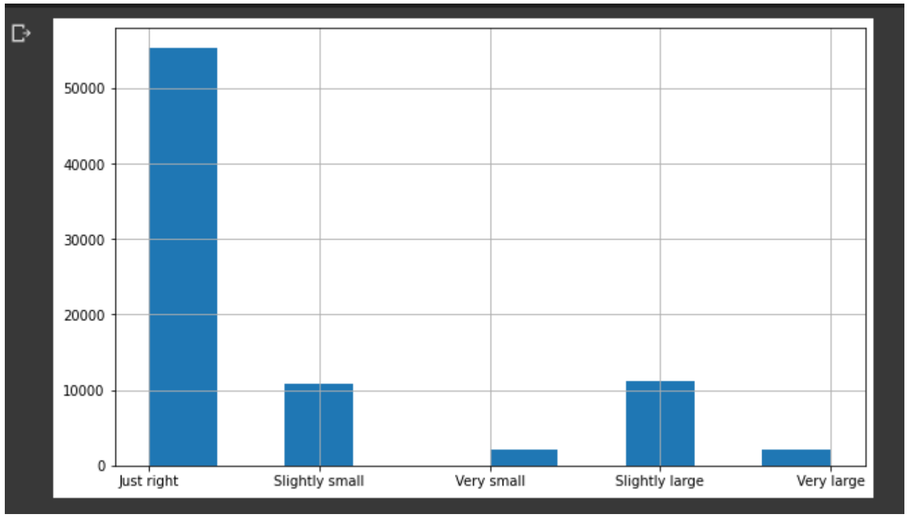
```
       item_id       user_id  rating                          timestamp  size  \
0         7443          Alex       4          2010-01-21 08:00:00+00:00   NaN
1         7443  carolyn.agan       3          2010-01-27 08:00:00+00:00   NaN
2         7443         Robyn       4          2010-01-29 08:00:00+00:00   NaN
3         7443            De       4          2010-02-13 08:00:00+00:00   NaN
4         7443         tasha       4          2010-02-18 08:00:00+00:00   NaN
...        ...           ...     ...                                ...   ...
99888   154797     BernMarie       5  2019-06-26 21:15:13.165000+00:00   6.0
99889    77949           Sam       4  2019-06-26 23:22:29.633000+00:00   4.0
99890    67194        Janice       5  2019-06-27 00:20:52.125000+00:00   NaN
99891    71607           amy       3  2019-06-27 15:45:06.250000+00:00   NaN
99892   119732         sarah       3  2019-06-29 13:55:16.542000+00:00   NaN

                  fit user_attr   model_attr   category              brand  \
0                 NaN     Small        Small    Dresses                NaN
1                 NaN       NaN        Small    Dresses                NaN
2                 NaN     Small        Small    Dresses                NaN
3                 NaN       NaN        Small    Dresses                NaN
4                 NaN     Small        Small    Dresses                NaN
...               ...       ...          ...        ...                ...
99888     Just right     Large  Small&Large    Dresses                NaN
99889  Slightly small     Small  Small&Large    Bottoms                NaN
99890     Just right     Small  Small&Large    Dresses                NaN
99891  Slightly small     Small  Small&Large  Outerwear  Jack by BB Dakota
99892     Just right     Small        Small    Dresses                NaN

       year  split
0      2012      0
1      2012      0
2      2012      0
3      2012      0
4      2012      0
...     ...    ...
99888  2017      0
99889  2014      2
99890  2013      2
99891  2016      2
99892  2016      2

[99893 rows x 12 columns]
```

ANALISI DESCRITTIVA

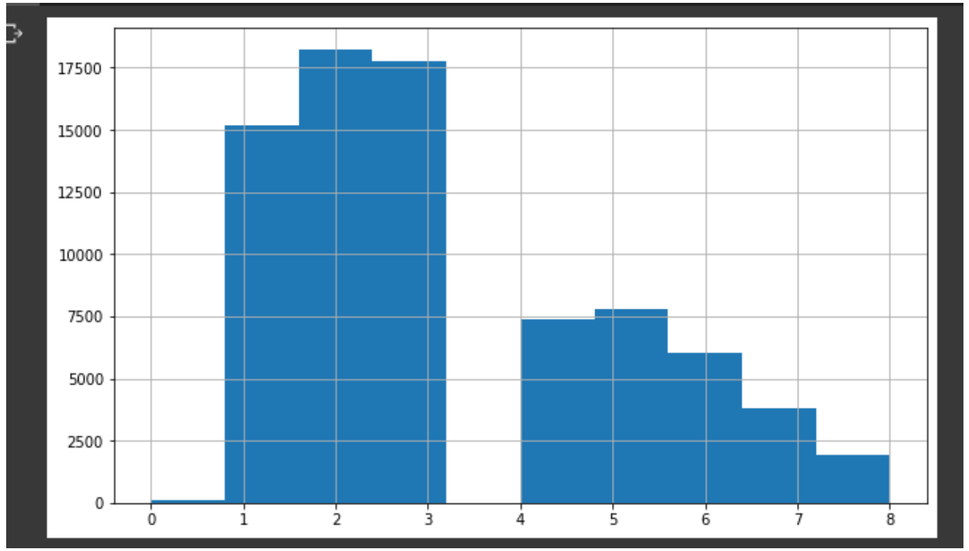
Rating



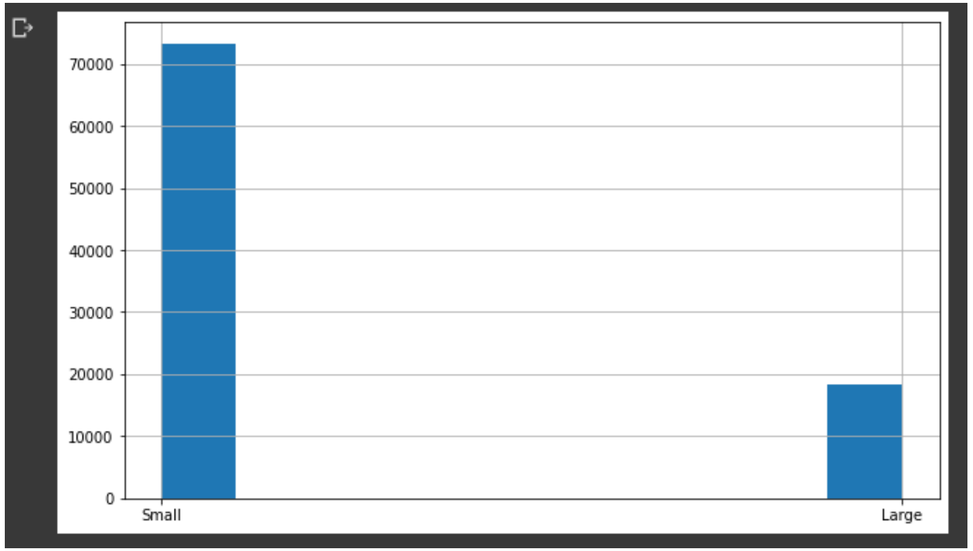
Fit



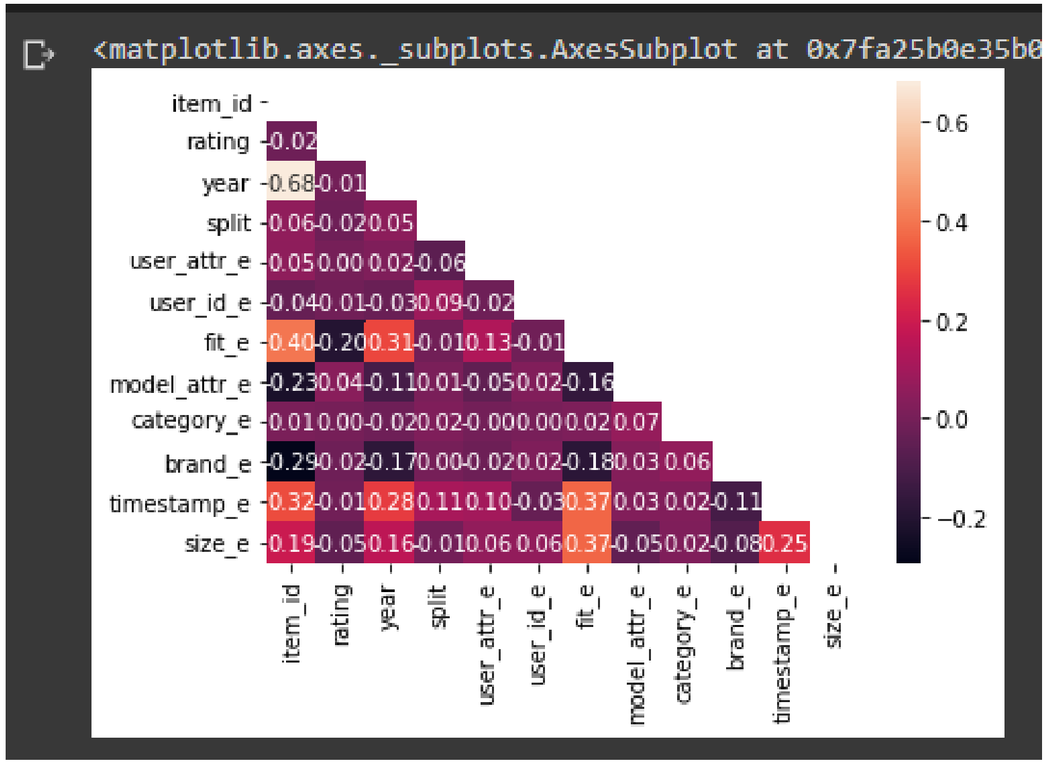
Size



User Attr



CORRELAZIONI

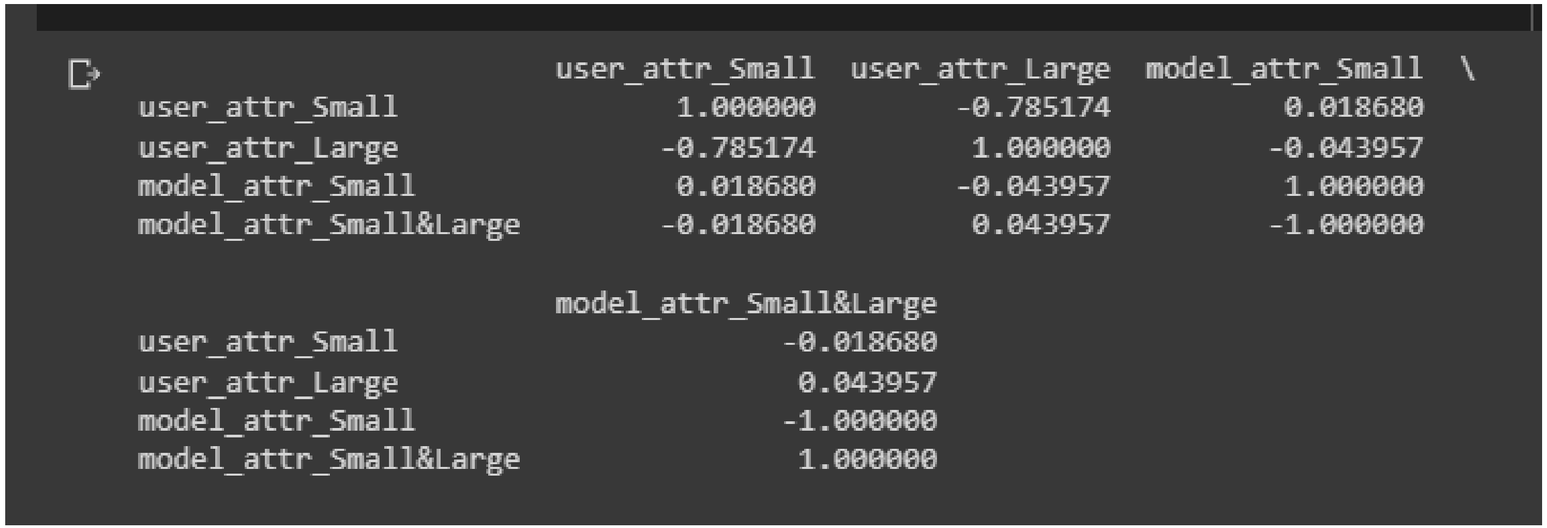


	fit_Just right	rating
fit_Just right	1.000000	0.323832
rating	0.323832	1.000000

Il rating aumenta se il capo calza "Giusto"

	user_attr_Small	user_attr_Large	size
user_attr_Small	1.000000	-0.785174	-0.69699
user_attr_Large	-0.785174	1.000000	0.69699
size	-0.696990	0.696990	1.00000

La size aumenta se lo user é Large, se Small viceversa

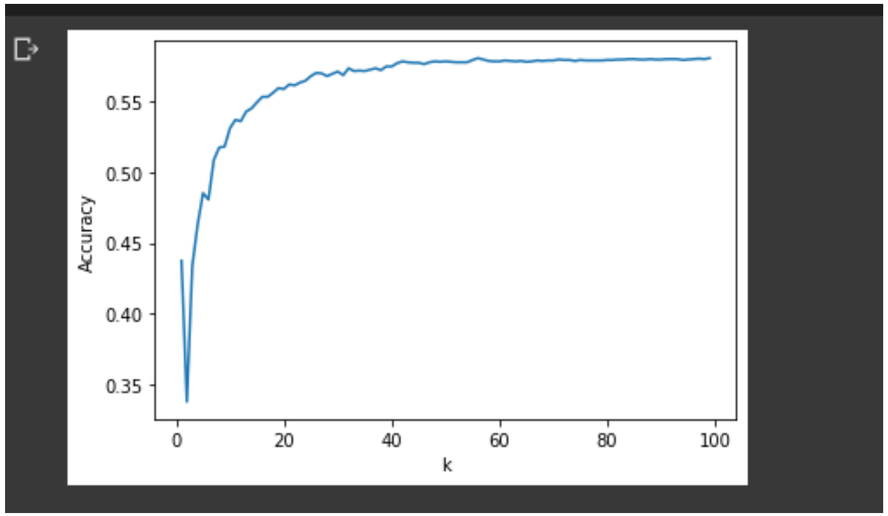


	user_attr_Small	user_attr_Large	model_attr_Small	\
user_attr_Small	1.000000	-0.785174	0.018680	
user_attr_Large	-0.785174	1.000000	-0.043957	
model_attr_Small	0.018680	-0.043957	1.000000	
model_attr_Small&Large	-0.018680	0.043957	-1.000000	

	model_attr_Small&Large
user_attr_Small	-0.018680
user_attr_Large	0.043957
model_attr_Small	-1.000000
model_attr_Small&Large	1.000000

Nessuna correlazione tra il user Small e modelle small e user Large e modelle sia Small che Large

KNN



- **Dopo il 40 i k neighbours l'accuracy migliora molto poco**
- **RMSE migliore possiamo avere circa 1**
- **Come distanza usiamo la Cosine ed é User Based**

```
Best RMSE = 1.0217
Best configuration = {'k': 70, 'sim_options': {'name': 'cosine', 'user_based': True}}
```

Matrice di Rating con valori predetti

```
[ ]  print(df_predictions)

            user_id   item_id    rating
0           39041.0  153445.0  4.306708
1           39041.0  129113.0  4.306708
2           39041.0  152990.0  4.306708
3           39041.0  152868.0  4.306708
4           39041.0  138414.0  4.306708
...             ...       ...       ...
455935      36523.0   32134.0  4.306708
455936      36523.0  142398.0  4.306708
455937      36523.0  154762.0  4.306708
455938      36523.0  152544.0  4.306708
455939      36523.0   50405.0  4.306708

[455940 rows x 3 columns]
```

- **Predizioni basate sul modello KNN**
- **Predizioni fatte su 447 users e 1020 items**

Consigliare Items

	item_id	mean_rating
0	6454.0	4.327801
1	7443.0	4.306708
2	11960.0	4.306708
3	16411.0	4.306708
4	21296.0	4.296763
...	...	...
1015	155308.0	4.306708
1016	155317.0	4.306708
1017	155537.0	4.306708
1018	155597.0	4.306708
1019	155950.0	4.306708

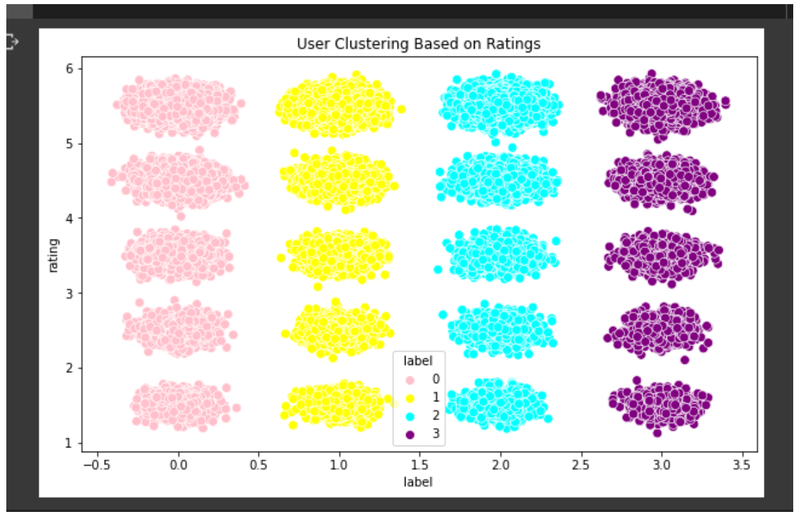
[1020 rows x 2 columns]

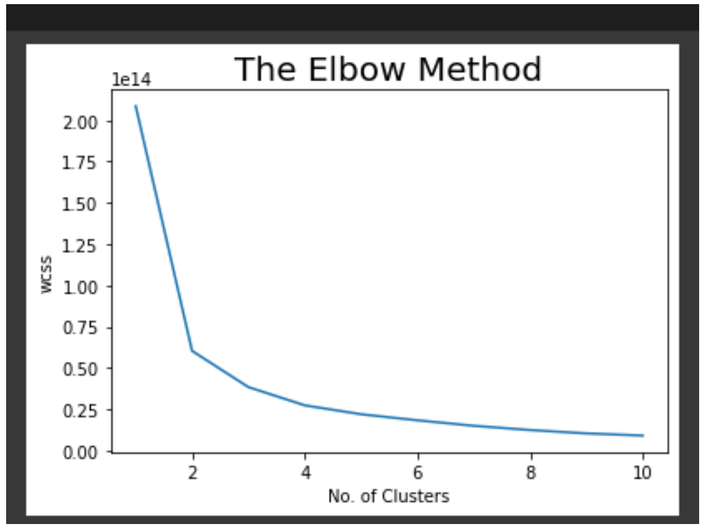
	user_id	recommended_items
0	Alex	[152835, 153228, 148277, 153543, 142082, 15353...
1	carolyn.agan	[152835, 153350, 152834, 152691, 153536, 15354...
2	Robyn	[142082, 152834, 152691, 153536, 154928, 15354...
3	De	[152835, 153350, 152834, 152691, 153536, 15354...
4	tasha	[153228, 153196, 151458, 153350, 153131, 14827...
...	...	...
44779	Marso	[152834, 153536, 148277, 153196, 153131, 15364...
44780	Foucault	[152834, 154928, 142082, 153350, 153196, 15283...
44781	nadgee	[152835, 148277, 153543, 151458, 153228, 15353...
44782	roshelle	[152834, 153536, 153196, 148277, 153131, 15145...
44783	BernMarie	[152691, 153536, 153131, 142082, 154928, 15283...

[44784 rows x 2 columns]

Gli items consigliati non sono mai stati votati dallo user e sono in ordine decrescente

K-Means



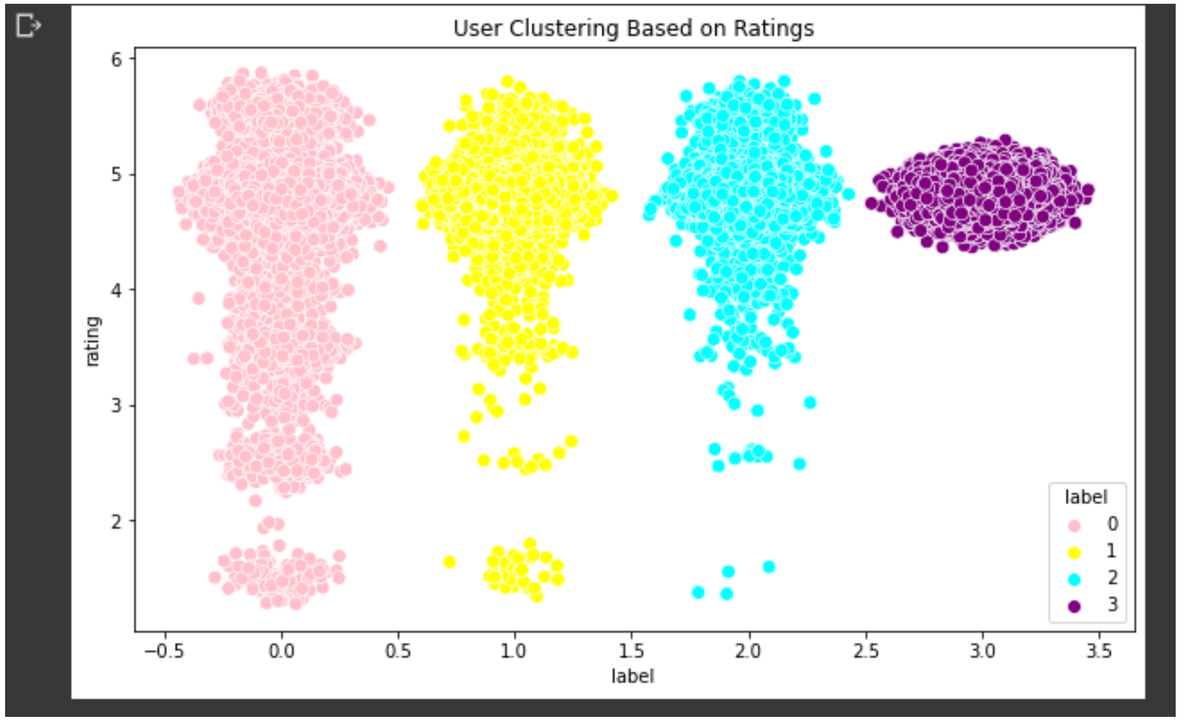


- **Distanza euclidea**
- **Siluette sore 0.45**
- **eseguito su dataframe inizia**

```
Silhouette score = 0.4543181075836887
```

- **4 clusters**

K-Means su Matrice Rating Predetto




```
Silhouette score = 0.5439173022340751
```

Matrix Factorization

	user_id	item_id	rating
0	10350.0	141853.0	3.985506
1	10350.0	144801.0	4.385892
2	10350.0	154934.0	4.350939
3	10350.0	153536.0	4.284791
4	10350.0	153223.0	4.385892
...	...	...	...
99995	31333.0	16411.0	4.302171
99996	31333.0	34931.0	4.302171
99997	31333.0	152049.0	4.370535
99998	31333.0	149829.0	4.119078
99999	31333.0	153118.0	4.355988

```
MSE: 0.9161
RMSE: 0.9571
```

- **RMSE migliorato rispetto al KNN**
- **Users e items uguali alla matrice con il KNN**
- **L'algoritmo usato é SVD dalla libreria surprise**