

YOUNG PEOPLE PROFILES

ANALYZING SURVEY DATA FOR
IMPROVED INSIGHTS

Winona Timothea – Batch 17
Final Project at Dibimbing



BACKGROUND

This dataset is taken from kaggle, which are answers to a survey in Slovakia to people aged 15 to 30. This dataset can be divided into 8 groups. The hobbies, music preferences, movie preferences, phobias, health habits, spending habits, personality and views on life, and demographics.

Understanding young people profiles can be beneficial for a marketing strategy or product development.

DATA DESCRIPTION



The dataset consists of 150 columns and 1010 rows. The values are 1, 2, 3, 4, 5 with 1 being the least related experience and 5 being the most related with young people



DATA ANALYSIS

THE MEAN AND MEDIAN FOR THE AGE COLUMN ARE AROUND 20, WHILE 15 IS THE MINIMUM AND 30 IS THE MAXIMUM. SO THE PEOPLE WHO ANSWERED THIS MOSTLY AROUND THE AGE OF 20

```
df.describe()
```

Rock	Metal or Hardrock	...	Age	Height	Weight	Number of siblings	Gender	Left - right handed	Education	Only child	Village - town	H bl
0.000000	1010.000000	...	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000
3.761952	2.361470	...	20.433699	173.514141	66.405051	1.297809	0.593069	0.900000	3.286422	0.251485	0.703960	0.000000
1.181333	1.370952	...	2.819010	9.924657	13.701713	1.010331	0.491505	0.300149	0.782758	0.434082	0.456735	0.000000
1.000000	1.000000	...	15.000000	62.000000	41.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000
3.000000	1.000000	...	19.000000	167.000000	56.000000	1.000000	0.000000	1.000000	3.000000	0.000000	0.000000	0.000000
1.000000	2.000000	...	20.000000	173.000000	65.000000	1.000000	1.000000	1.000000	3.000000	0.000000	1.000000	1.000000
5.000000	3.000000	...	22.000000	180.000000	75.000000	2.000000	1.000000	1.000000	4.000000	1.000000	1.000000	1.000000
5.000000	5.000000	...	30.000000	203.000000	165.000000	10.000000	1.000000	1.000000	6.000000	1.000000	1.000000	1.000000

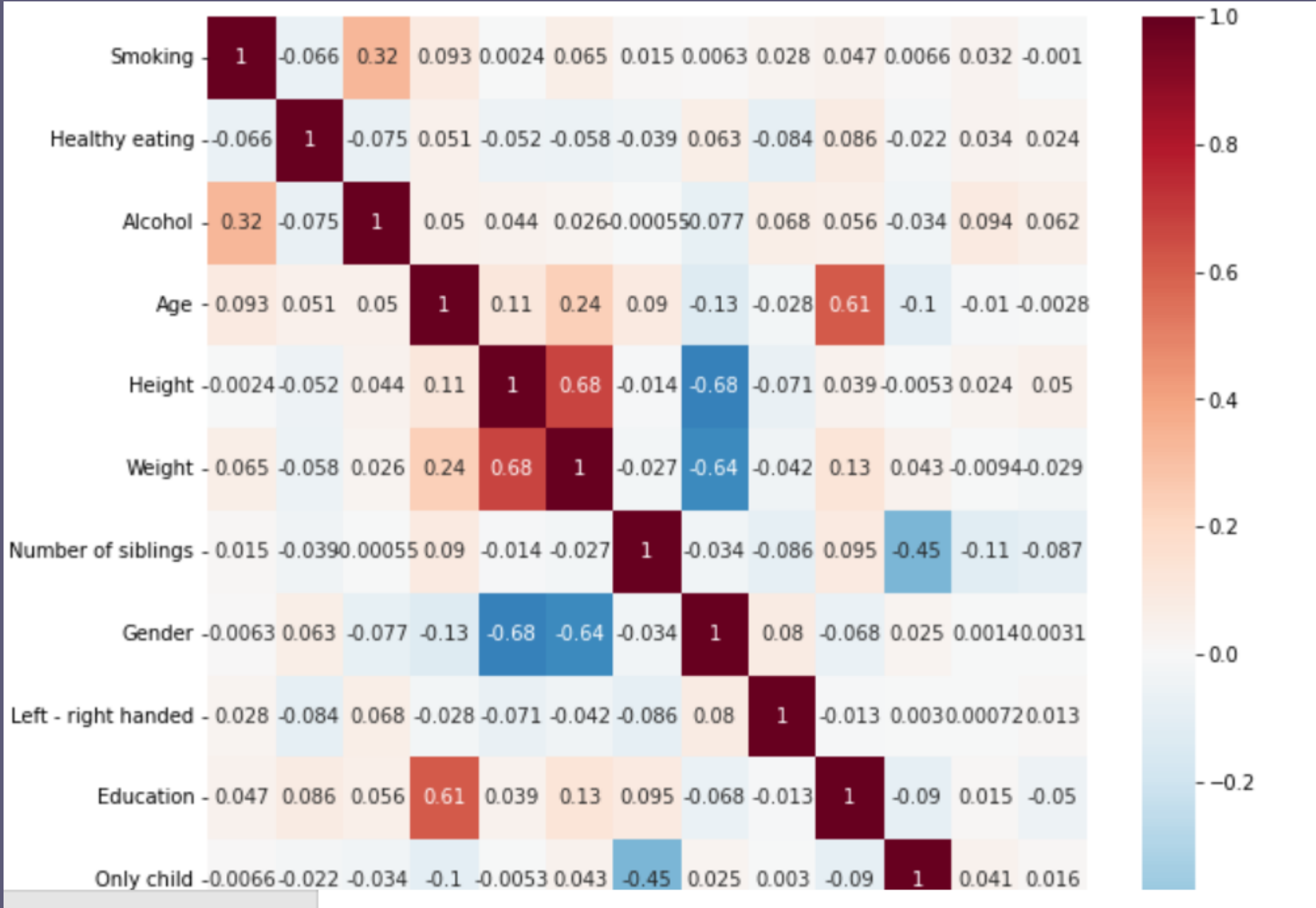
HEALTHY HABITS AND DEMOGRAPHICS

01

There are no strong correlations between healthy habits and demographic group. So generally healthy habits are not influenced by demographics (gender, age, etc)

02

But there is a weak positive correlation in the healthy habit group



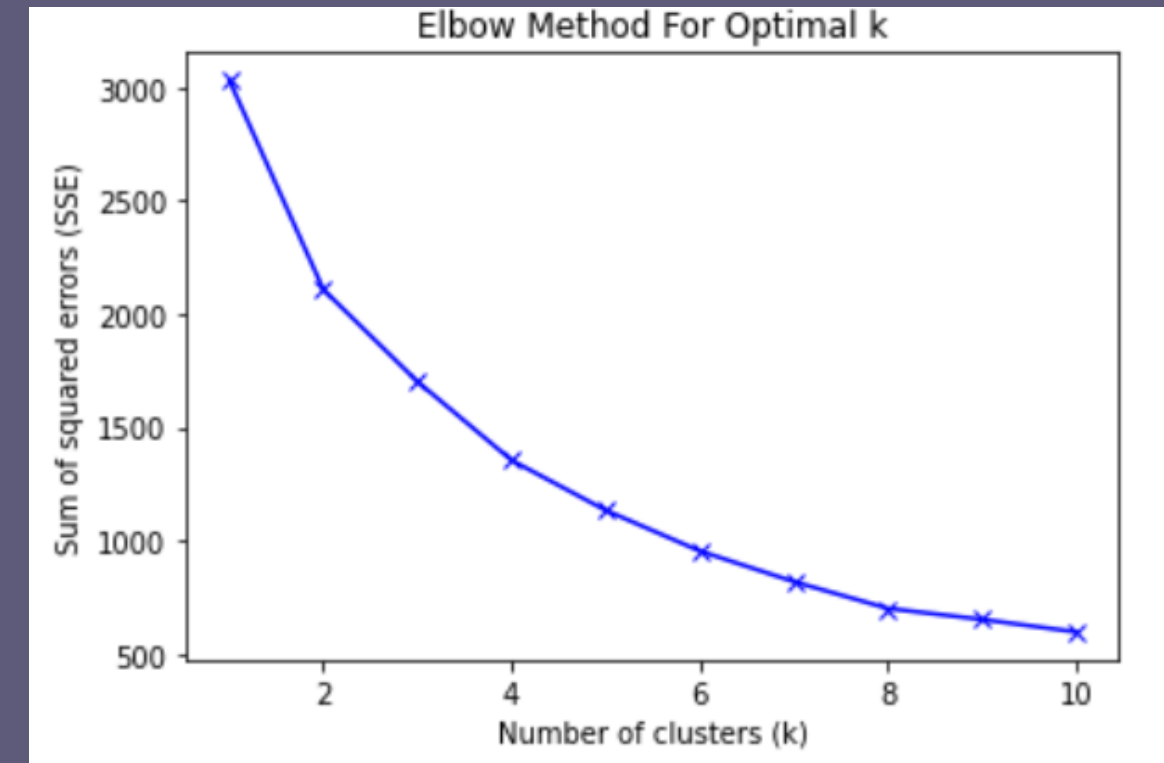
CLUSTERS IN HEALTH HABITS

Cluster 0: doesn't eat healthy at all, but they also doesn't like smoking and alcohol

Cluster 3: also doesn't like smoking and alcohol, but they are a bit better in healthy eating

Cluster 1: doesn't like smoking that much but tend to like alcohol. They did like a bit of healthy eating.

Cluster 2: Doesn't like smoking and alcohol, loves healthy eating. Also with the smallest number of people.



Using elbow method, 4 looks like an optimal number for clustering

```
Cluster 0: 352 data points, centroid = [2. 1. 3.]
Cluster 3: 258 data points, centroid = [2. 2. 3.]
Cluster 1: 255 data points, centroid = [3. 2. 4.]
Cluster 2: 145 data points, centroid = [2. 4. 3.]
```


[illegible]

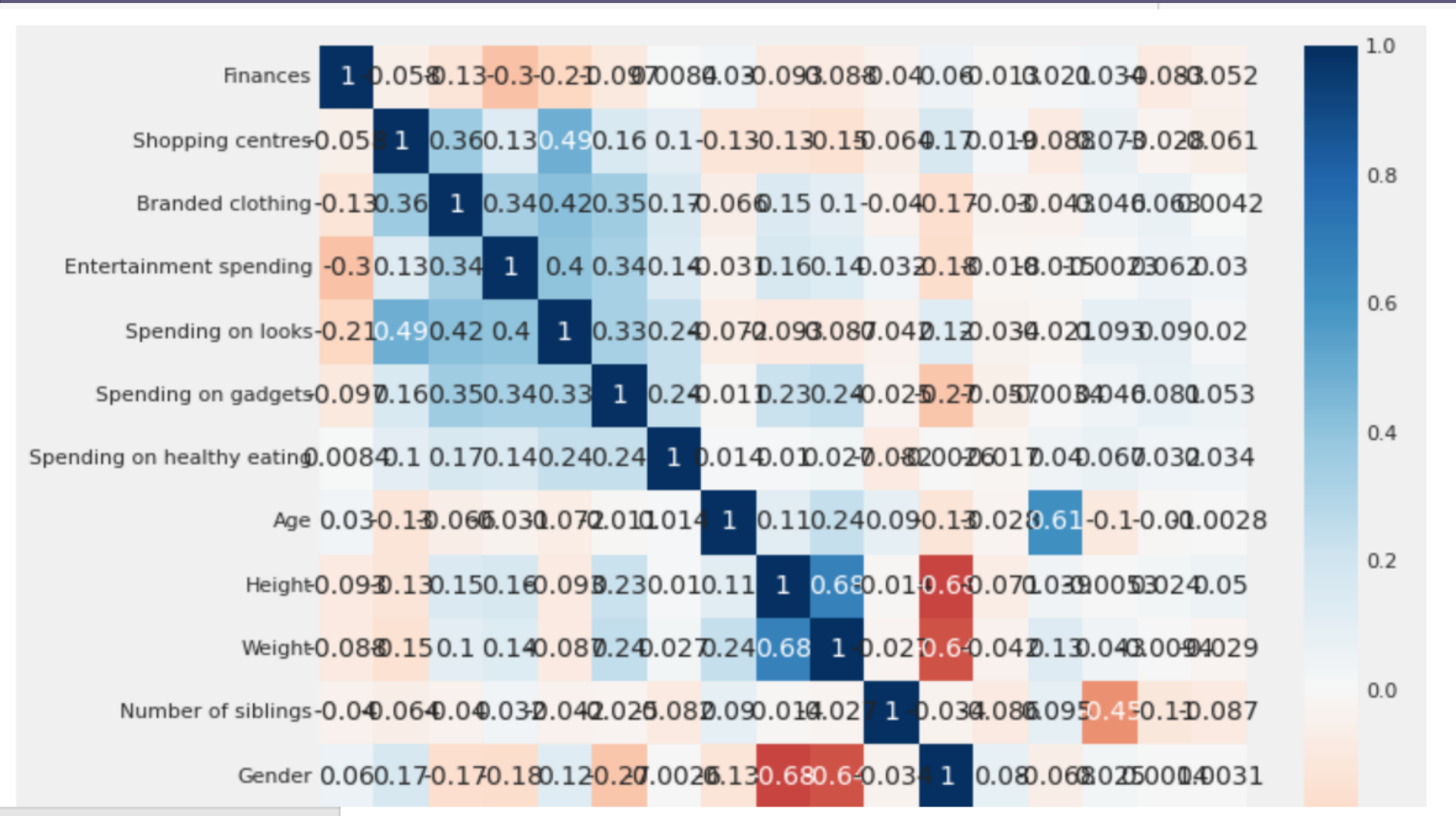
SPENDING HABITS AND DEMOGRAPHICS

01

There are no strong correlations between spending habits and demographic group too. So generally spending habits are not influenced by demographics (gender, age, etc)

02

But there is positive correlations in the spending habit group



CLUSTERS IN SPENDING HABITS

Cluster 0: normal finance management, tend to like spend money on every category (gadget, healthy eating, entertainment, etc).

Cluster 1: normal amount of spending

Cluster 2: a bit low spending generally, high finance management and high spending on healthy eating

Cluster 3: High spending on every category, no management on finance

Using elbow method, 5 looks like an optimal number for clustering.

```
Cluster 0: 318 data points, centroid = [3. 4. 4. 4. 4. 4. 4.]
Cluster 1: 284 data points, centroid = [3. 3. 3. 3. 3. 3. 3.]
Cluster 2: 168 data points, centroid = [4. 2. 2. 2. 2. 2. 4.]
Cluster 3: 137 data points, centroid = [1. 5. 5. 5. 5. 5. 5.]
Cluster 4: 103 data points, centroid = [4. 1. 1. 1. 1. 1. 2.]
```

Cluster 4: Very low in spending anything except for healthy eating, which is also not in normal amount, high finance management.

[illegible]

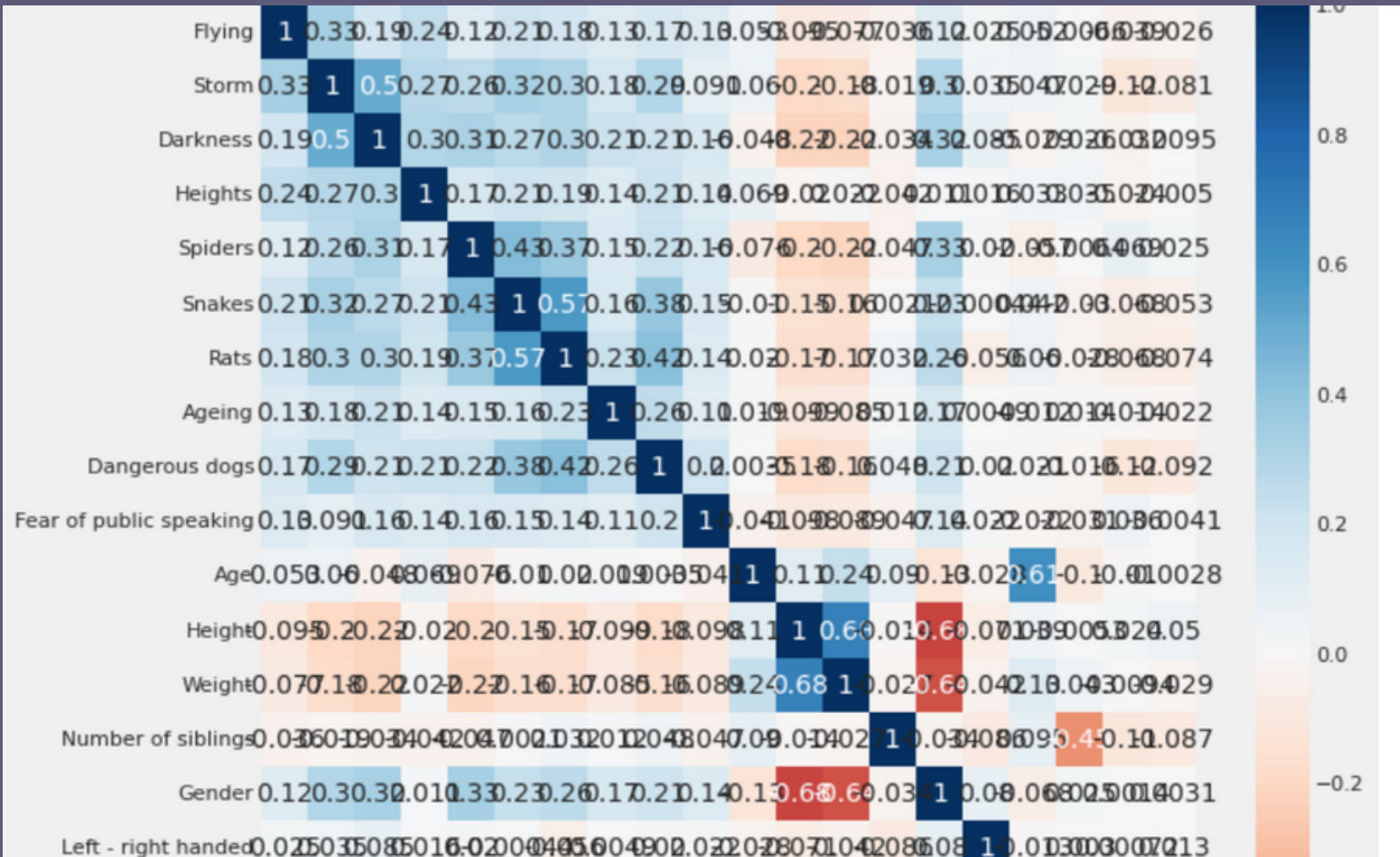
PHOBIAS AND DEMOGRAPHICS

01

There are no strong correlation between phobias and demographic group too. But there is a weak positive correlation between phobias and gender.

02

There is positive correlations in the phobias group



CLUSTERS IN PHOBIAS

Cluster 1: No fear of any category at all except public speaking

Cluster 0: There is a bit of fear on every category, except flying.

Cluster 3: No fear of flying, a normal amount of fears on storms, darkness, heights, public speaking, and ageing. Very high phobia on the rest of category.

Cluster 2: tend to have normal fear amount on every category, with no fear on storms and darkness.

Using elbow method, 4 looks like an optimal number for clustering.

```
➞ Cluster 1: 377 data points, centroid = [1. 1. 1. 1. 1. 1. 1. 1. 1. 3.]  
Cluster 0: 296 data points, centroid = [1. 2. 2. 2. 2. 2. 2. 2. 3. 3.]  
Cluster 3: 190 data points, centroid = [1. 3. 3. 3. 5. 5. 5. 3. 5. 3.]  
Cluster 2: 147 data points, centroid = [3. 1. 1. 3. 3. 3. 3. 3. 3. 4.]
```

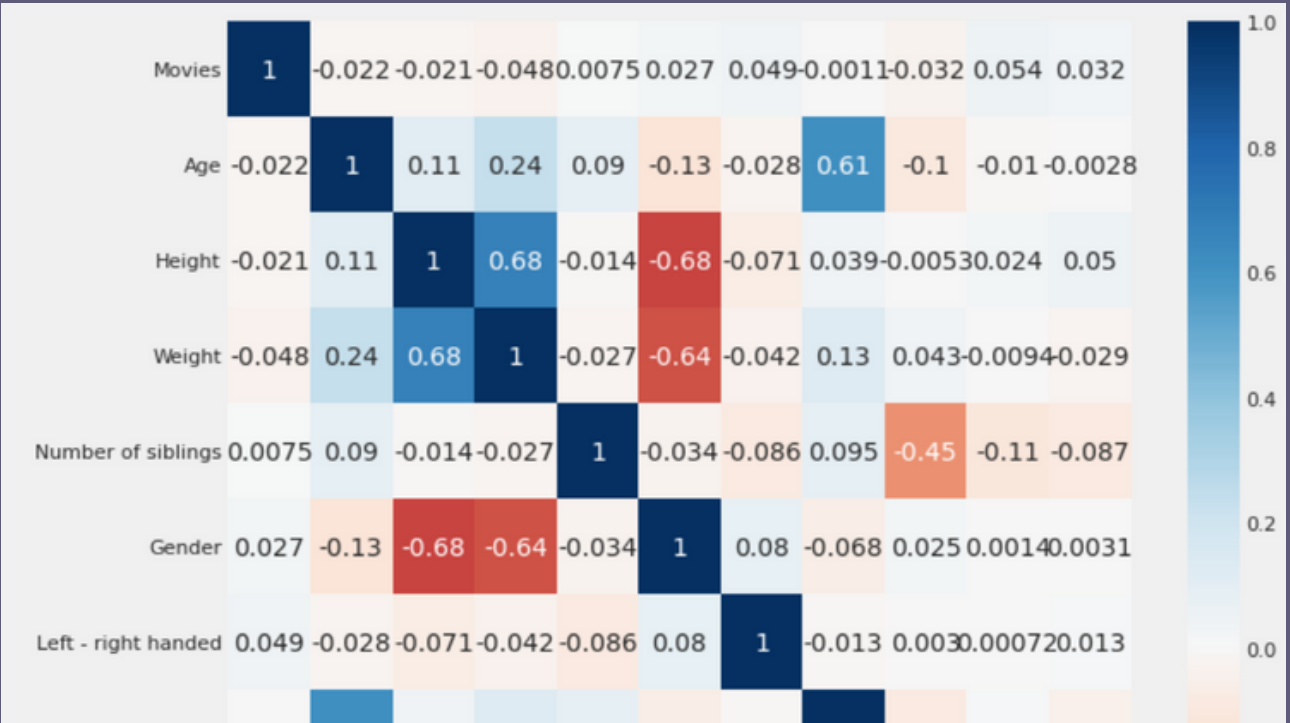
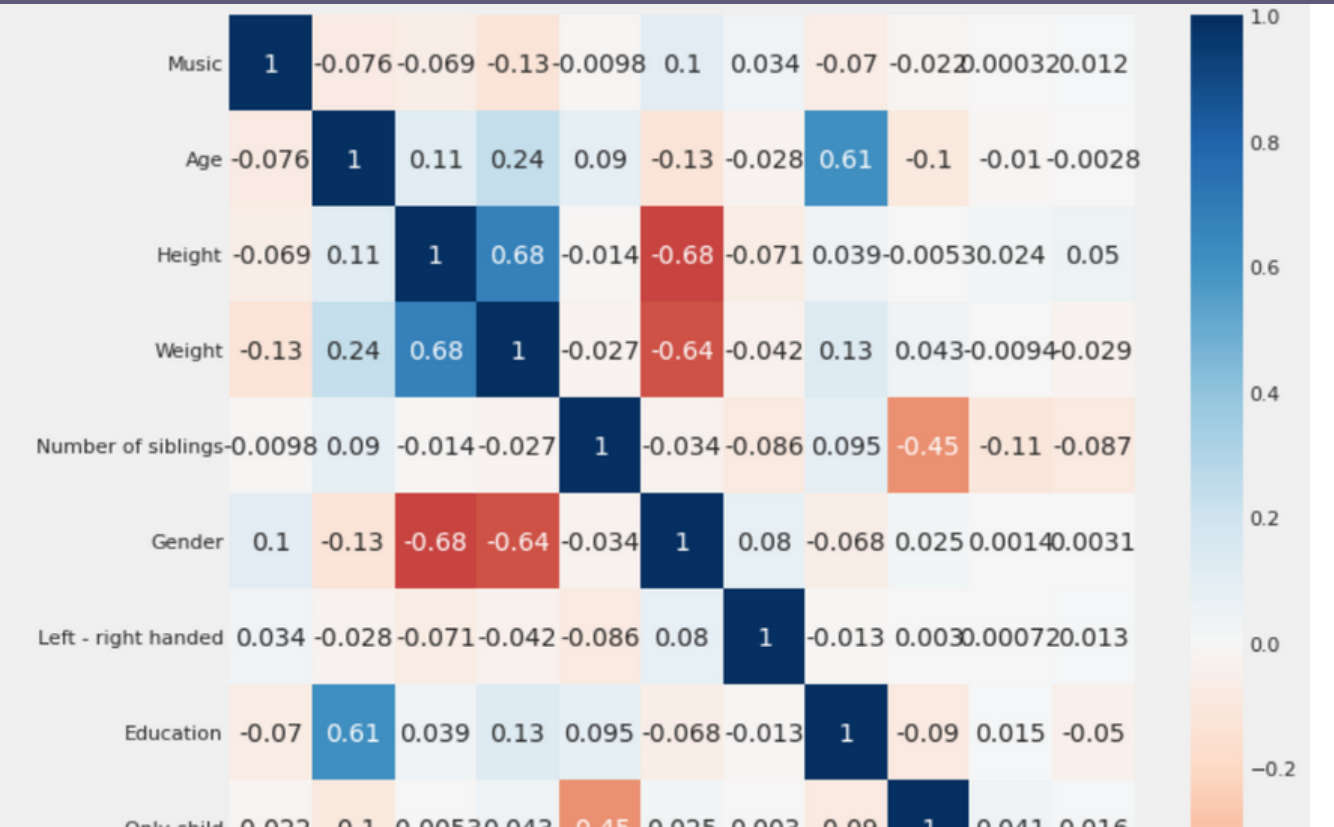
MOVIE, MUSIC AND DEMOGRAPHICS

01

There are no strong correlation between love of music and demographic group.

02

There are no strong correlation between love of movies and demographic group too.





01

Generally no correlation between demographics and other groups, but tend to have some correlation inside each groups..

02

The clusters with the highest number are people who generally dont fear anything, tend to spend on every category (gadget, entertainment, etc), and people that have low healthy habits.