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 _

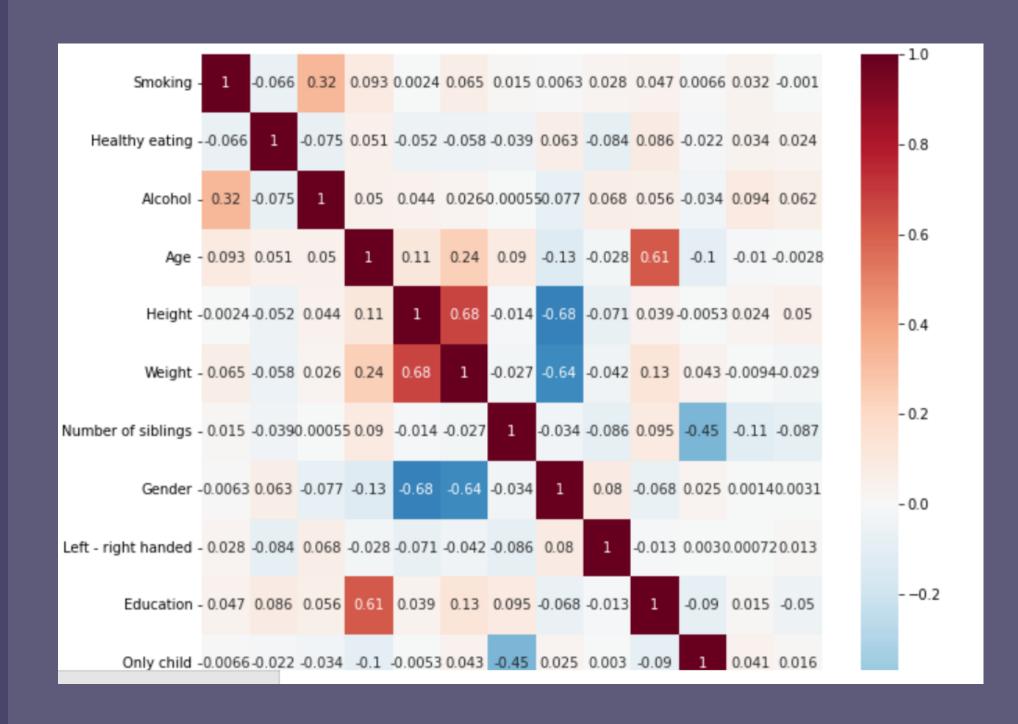
THE MEAN AND MEDIAN FOR THE AGE COLUMN ARE AROUND 20, WHILE 15 IS THE MINIMUM AND 30 IS THE MAXIMUM. SO THE PEOPLE WHO ANSWERED THIS MOSTLY AROUND THE AGE OF 20

df.desc	ribe()										
Rock	Metal or Hardrock	 Age	Height	Weight	Number of siblings	Gender	Left - right handed	Education	Only child	Village - town	Н b 1
0.000000	1010.000000	 1010.000000	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000	1010.
3.761952	2.361470	 20.433699	173.514141	66.405051	1.297809	0.593069	0.900000	3.286422	0.251485	0.703960	0.
1.181333	1.370952	 2.819010	9.924657	13.701713	1.010331	0.491505	0.300149	0.782758	0.434082	0.456735	0.
1.000000	1.000000	 15.000000	62.000000	41.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.
3.000000	1.000000	 19.000000	167.000000	56.000000	1.000000	0.000000	1.000000	3.000000	0.000000	0.000000	0.
1.000000	2.000000	 20.000000	173.000000	65.000000	1.000000	1.000000	1.000000	3.000000	0.000000	1.000000	1.
5.000000	3.000000	 22.000000	180.000000	75.000000	2.000000	1.000000	1.000000	4.000000	1.000000	1.000000	1.
5.000000	5.000000	 30.000000	203.000000	165.000000	10.000000	1.000000	1.000000	6.000000	1.000000	1.000000	1.

HEALTHY HABITS AND DEMOGRAPHICS

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

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



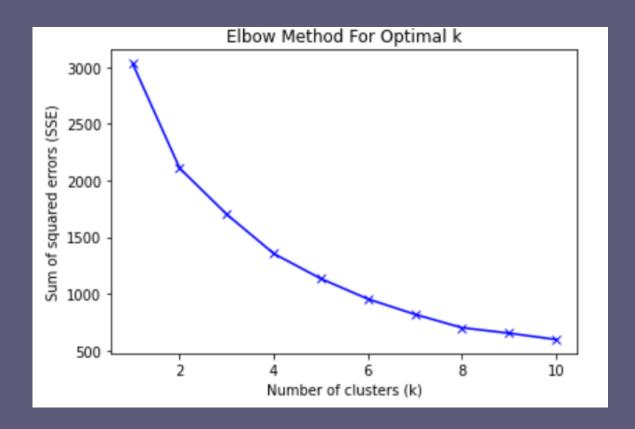
CLUSTERS IN HEALTH HABITS

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.]
```



DATA ANALYSIS

ON THIS SPENDING HABITS GROUP. THE MEAN FOR EVERY COLUMN IS AROUND 3, WHICH IS THE NORMAL AMOUNT OF SPENDING HABITS. WITH THE LOWEST IS SPENDING ON GADGETS AND THE HIGHEST IS THE SPENDING FOR HEALTHY EATING. SO GENERALY THEIR SPENDING HABITS ARE NORMAL AMOUNT AND LOOKS HEALTHY

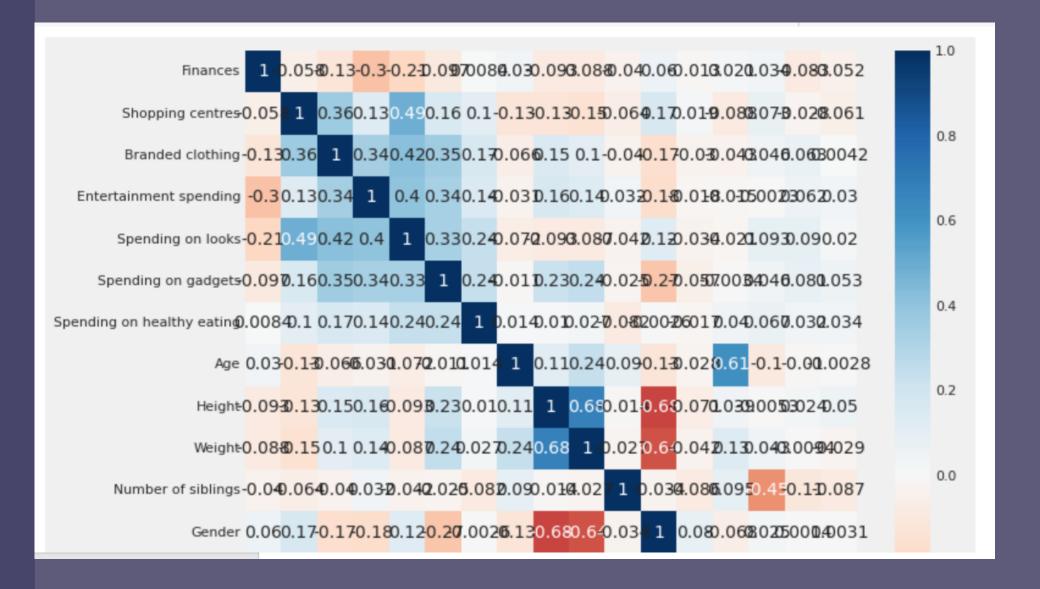
Spending

	Finances	Shopping centres	Branded clothing	Entertainment spending	Spending on looks	Spending on gadgets	on healthy eating
count	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000
mean	3.023833	3.234127	3.050595	3.201589	3.106256	2.870297	3.557540
std	1.142662	1.321750	1.305026	1.187179	1.203575	1.284970	1.092665
min	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
25%	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000	3.000000
50%	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000	4.000000
75%	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000
max	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000

SPENDING HABITS AND DEMOGRAPHICS

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

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. 2.]
```

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



DATA ANALYSIS

ON THIS PHOBIAS GROUP. THE MEAN AND MEDIAN FOR EVERY COLUMN IS AROUND 2 TO 3, WHICH MEANS GENERALLY THEY DON'T HAVE PHOBIAS FOR ANY OF THE CATEGORY. THEY ESPECIALLY DOESN'T REALLY FEAR STORMS, BUT THEY DO TEND FEAR A BIT ON SNAKES AND DANGEROUS DOGS.

	Flying	Storm	Darkness	Heights	Spiders	Snakes	Rats	Ageing	Dangerous dogs	Fear of public speaking	
count	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000	1010.000000	1010.0
mean	2.061569	1.973241	2.250992	2.615690	2.825871	3.027723	2.409136	2.580773	3.042616	2.803766	20.4
std	1.209552	1.163654	1.253538	1.293527	1.539866	1.500982	1.398511	1.385779	1.366152	1.214035	2.8
min	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	15.0
25%	1.000000	1.000000	1.000000	2.000000	1.000000	2.000000	1.000000	1.000000	2.000000	2.000000	19.0
50%	2.000000	2.000000	2.000000	2.000000	3.000000	3.000000	2.000000	2.000000	3.000000	3.000000	20.0
75%	3.000000	3.000000	3.000000	4.000000	4.000000	4.000000	3.000000	4.000000	4.000000	4.000000	22.0
max	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	30.0

PHOBIAS AND DEMOGRAPHICS

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

There is positive correlations in the phobias group

Flying	1	0.33	0.19	0.24	0.12).2 1	0.18	.13	0.17	0.10	.050	R 0 -9 050	7 07.03	6 10	.02050-192	00 6 60	3 09026	Т	1.0
Storm (0.33	1	0.5	0.27	0.26).32	0.30	.18	0.20	.090	1.06	0.20.1	8 .01	9.30	.03050407	02 <mark>9.</mark>	1 <mark>2</mark> .081		
Darkness (0.19	0.5	1	0.3	0.31).27	0.30	.21	0.21	0.1-6	.04	8.2-20.2	20 .03	@130	.0805.0209	202060	B 20095		0.8
Heights (0.24	D.27	0.3	1	0.17	0.21	0.190	.14	0.21	0.10	.069	9.020	202.04	2010	101060308	0325.0	204005		
Spiders (0.12	0.26	0.31	0.17	1	0.43	0.37	.15	0.22	0.1-6	.07	6 .20.2	20 .04	07.33	0.0 12 .0 15 .0	DO 5.4	609025		0.6
Snakes (0.21	0.32	0.27	0.21	0.43	1 ().570	.16	0.38	0.19	0.0	0.150.1	06002	01-103	00 0 494412	.9 <mark>3</mark> .0	6 8053		
Rats (0.18	80.3	0.30	0.19	D.37).57	1	.23	0.42	0.14	0.02	0.1-70.1	07.03	0.2-6	0.05060-0	.0-2080	6 8074		
Ageing (0.13	0.18	0.21	0.14	0.15	0.16	0.23	1	0.26	0.10	.010	9.0 9 090	8 501	2 . 107	00 -0 90 D2	2012040	-D4022		0.4
Dangerous dogs (0.17	0.29	0.21	0.21	0.22	38.0	0.420	.26	1	0.0.	003	55.1-8 0.1	0 504	0.21	0.02.020	.01 6 .:	1 <mark>0</mark> .092		
Fear of public speaking (0.10	.090	1.16	0.14	0.16	0.15	0.140	.11	0.2	1-0	.040	109080	3 904	07.10	.0202.0-202	2010	B 60041		0.2
AgeO	.05	3.06	.04	8060	907	5 .0 1	0.00	200.	DOB	504	11	0.110.2	40.09	0.10	3.02 <mark>0.61</mark> -	0.10.6	010028		
Heigh ŧ 0	.09	50.20	0.22	D.02	20.2	0.15	0.10	.09	9.10	8.090	8.11	1 0.6	0.01	0.68	0.01010-1899.	E.0 00	20.05		0.0
Weigh ŧ 0	.07	7 .18	0.20	202	D .2-2	0.1€	0.10	.08	5.10	5.08	9.24	0.68 1	0.02	0.64	.040210.	0.030	9 9.4 029		0.0
Number of siblings0	£0.)60-ID	90-3	404€	204.	Ø0Q	.D30	2010	2040	B040	7.09	.0-1040	2 <mark>71-</mark> (0.0 3	40 8 6095	.4-0.:	1 0 .087		
Gender (0.12	0.30	0.30	.01	1 .33	0.23	0.260	.17	0.21	0.14	0.13	0.680.6	0.03	41	0.08.06	302.50	014031		-0.2
Left - right handed0	.020	5030	5080	501	6.0 2	000	495.0	004	900	.030	2.0-20	807010	4 0208	6 08	1-0.0 D	300030	007213		

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. 1. 3.]

Cluster 0: 296 data points, centroid = [1. 2. 2. 2. 2. 2. 2. 2. 3. 3.]

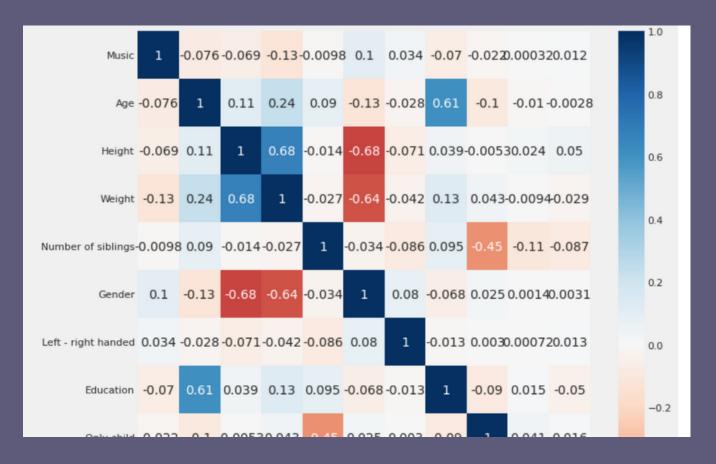
Cluster 3: 190 data points, centroid = [1. 3. 3. 3. 5. 5. 5. 3. 5. 3.]

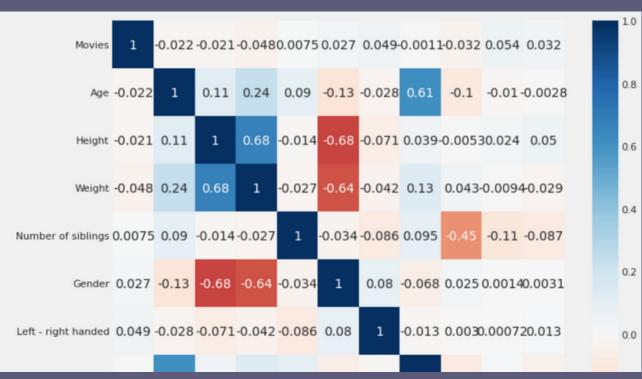
Cluster 2: 147 data points, centroid = [3. 1. 1. 3. 3. 3. 3. 3. 3. 4.]
```

MOVIE, MUSIC AND DEMOGRAPHICS

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

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







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

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.