

Анализ опроса словацкой молодежи 2013

Никита Козловский

ВГУ, ПММ, 3 группа

1 Введение

Рассмотрим данные, представленные на сайте <https://www.kaggle.com/miroslavsabo/young-people-survey> Опрос был проведен факультетом социальных и экономических наук Университета имени Коменского в Братиславе в 2013 году. Было опрошено 1010 человек по 150 пунктам:

- Предпочтения в музыке (19 пунктов)
- Предпочтения в фильмах (12 пунктов)
- Хобби и интересы (32 пунктов)
- Фобии (10 пунктов)
- Здравоохранение (3 пункта)
- Черты характера, взгляды на жизнь и мнения (57 пунктов)
- На что тратите деньги (7 пунктов)
- Демографические данные (10 пунктов)

Попробуем найти ответы на следующие вопросы:

- Можно ли разделить студентов на группы по предпочтениям в музыке? (описательные, факторный)
- Что определяет бережливого человека ? (регрессия)
- Анализ страхов по группам

2 Анкета

2.1 МУЗЫКАЛЬНЫЕ ПРЕДПОЧТЕНИЯ

1. Мне нравится слушать музыку: Категорически не согласен 1-2-3-4-5 Полностью согласен (целое)
2. Я предпочитаю: Медленную музыку 1-2-3-4-5 Быструю музыку (целое)
3. Dance, Disco, Funk: Не нравится вообще 1-2-3-4-5 Нравится очень (целое)
4. Народная музыка: Не нравится вообще 1-2-3-4-5 люблю (целое)
5. Кантри: Не нравится вообще 1-2-3-4-5 Нравится очень (целое)
6. Классика: Не нравится вообще 1-2-3-4-5 Нравится очень (целое)
7. Мюзиклы: Не нравится вообще 1-2-3-4-5 Нравится очень (целое)
8. Поп: Не нравится вообще 1-2-3-4-5 Нравится очень (целое)
9. Рок: Не нравится вообще 1-2-3-4-5 Нравится очень (целое)
10. Металл, хард-рок: Не нравится вообще 1-2-3-4-5 Нравится очень (целое)
11. Панк: Не нравится вообще 1-2-3-4-5 Нравится очень (целое)
12. Хип-хоп, Рэп: Не нравится вообще 1-2-3-4-5 Нравится очень (целое)
13. Reggae, Ska: Не нравится вообще 1-2-3-4-5 Нравится очень (целое)
14. Swing, Jazz: не нравится вообще 1-2-3-4-5 Нравится очень (целое)
15. Рок-н-ролл: не нравится вообще 1-2-3-4-5 Нравится очень (целое)
16. Альтернативная музыка: не нравится вообще 1-2-3-4-5 Нравится очень (целое)
17. Латино: не нравится вообще 1-2-3-4-5 Нравится очень (целое)
18. Техно, Транс: не нравится вообще 1-2-3-4-5 Нравится очень (целое)
19. Опера: не нравится вообще 1-2-3-4-5 Нравится очень (целое)

2.2 ФИЛЬМЫ ПРЕДПОЧТЕНИЯ

1. Мне очень нравится смотреть фильмы: Сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
2. Фильмы ужасов: Не нравится вообще 1-2-3-4-5 Нравится очень (целое)
3. Триллеры: Не нравится вообще 1-2-3-4-5 Нравится очень (целое)
4. Комедии: не нравится вообще 1-2-3-4-5 Нравится очень (целое)
5. Романтические фильмы: Не нравится вообще 1-2-3-4-5 Нравится очень (целое)
6. Научно-фантастические фильмы: Не нравится вообще 1-2-3-4-5 Нравится очень (целое)

7. Военные фильмы: Не нравится вообще 1-2-3-4-5 Нравится очень (целое)
8. Сказки: Не нравится вообще 1-2-3-4-5 Нравится очень (целое)
9. Мультфильмы: Не нравится вообще 1-2-3-4-5 Наслаждайтесь очень (целое)
10. Документальные фильмы: Не нравится вообще 1-2-3-4-5 Нравится очень (целое)
11. Западные фильмы: Не нравится вообще 1-2-3-4-5 Нравится очень (целое)
12. Экшн фильмы: Не нравится вообще 1-2-3-4-5 Нравится очень (целое)

2.3 ХОББИ ИНТЕРЕСЫ

1. История: Не интересуется 1-2-3-4-5 Очень интересно (целое)
2. Психология: не интересуется 1-2-3-4-5 Очень интересно (целое)
3. Политика: Не интересуется 1-2-3-4-5 Очень интересно (целое)
4. Математика: не интересуется 1-2-3-4-5 Очень интересно (целое)
5. Физика: не интересуется 1-2-3-4-5 Очень интересно (целое)
6. Интернет: не интересуется 1-2-3-4-5 Очень интересно (целое)
7. Программное обеспечение для ПК, Оборудование: Не интересуется 1-2-3-4-5 Очень интересно (целое)
8. Экономика, Менеджмент: Не интересуется 1-2-3-4-5 Очень интересно (целое)
9. Биология: не интересуется 1-2-3-4-5 Очень интересно (целое)
10. Химия: не интересуется 1-2-3-4-5 Очень интересно (целое)
11. Чтение стихов: Не интересуется 1-2-3-4-5 Очень интересно (целое)
12. География: Не интересуется 1-2-3-4-5 Очень интересно (целое)
13. Иностранные языки: не интересуются 1-2-3-4-5 Очень интересно (целое)
14. Медицина: не интересуется 1-2-3-4-5 Очень интересно (целое)
15. Закон: не интересуется 1-2-3-4-5 Очень интересно (целое)
16. Автомобили: Не интересуется 1-2-3-4-5 Очень интересно (целое)
17. Искусство: Не интересуется 1-2-3-4-5 Очень интересно (целое)
18. Религия: Не интересуется 1-2-3-4-5 Очень интересно (целое)
19. Мероприятия на свежем воздухе: не интересуется 1-2-3-4-5 Очень интересно (целое)
20. Танцы: не интересуются 1-2-3-4-5 Очень интересно (целое)
21. Игра на музыкальных инструментах: не интересуется 1-2-3-4-5 Очень интересно (целое)

22. Поэзия: Не интересно 1-2-3-4-5 Очень интересно (целое)
23. Занятия спортом: Не интересуюсь 1-2-3-4-5 Очень интересно (целое)
24. Спорт на конкурентном уровне: не интересует 1-2-3-4-5 Очень интересно (целое)
25. Садоводство: Не интересуется 1-2-3-4-5 Очень интересно (целое)
26. Знаменитый образ жизни: не интересует 1-2-3-4-5 Очень интересно (целое)
27. Покупки: Не интересует 1-2-3-4-5 Очень интересно (целое)
28. Наука и техника: не интересует 1-2-3-4-5 Очень интересно (целое)
29. Театр: не интересует 1-2-3-4-5 Очень интересно (целое)
30. Общение: Не интересует 1-2-3-4-5 Очень интересно (целое)
31. Адреналин спорт: не интересует 1-2-3-4-5 Очень интересно (целое)
32. Домашние животные: Не интересуются 1-2-3-4-5 Очень интересно (целое)

2.4 ФОБИИ

1. Летающий: совсем не боится 1-2-3-4-5 Очень боится (целое)
2. Гром, молния: совсем не боюсь 1-2-3-4-5 Очень боится (целое)
3. Тьма: не боится совсем 1-2-3-4-5 Очень боится (целое)
4. Высота: совсем не боится 1-2-3-4-5 Очень боится (целое)
5. Пауки: совсем не боюсь 1-2-3-4-5 Очень боится (целое)
6. Змей: совсем не боюсь 1-2-3-4-5 Очень боится (целое)
7. Крысы, мыши: совсем не боятся 1-2-3-4-5 Очень боится (целое)

2.5 ЗДОРОВЬЕ

1. Отношение к курению: Никогда не курил - Пробовал курить - Бывший курильщик - Текущий курильщик (Номинальная)
2. Пью: Никогда - Пью в компаниях - Пью много (Номинальная)
3. Я живу очень здоровым образом жизни: сильно не согласен 1-2-3-4-5 Полностью согласен (целое)

2.6 ЧЕРТЫ ХАРАКТЕРА, ВЗГЛЯДЫ НА ЖИЗНЬ

1. Я обращаю внимание на то, что происходит вокруг: сильно не согласен 1-2-3-4-5. Полностью согласен (целое)
2. Я стараюсь выполнять задания как можно скорее и не оставлять их до последней минуты .: Сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
3. Я всегда составляю список, поэтому ничего не забываю: сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
4. Я часто учусь или работаю даже в свободное время .: Сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
5. Я смотрю на вещи с разных точек зрения, прежде чем идти вперед. Полностью не согласен 1-2-3-4-5. Полностью согласен (целое)
6. Я считаю, что плохие люди пострадают в один прекрасный день, и хорошие люди будут вознаграждены. Полностью не согласен 1-2-3-4-5. Полностью согласен (целое)
7. Я надёжен на работе и всегда выполняю все заданные мне задачи: Полностью не согласен 1-2-3-4-5 Полностью согласен (целое)
8. Я всегда соблюдаю свои обещания: сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
9. Я могу быстро попасть на кого-то, а затем полностью потерять интерес. Полностью не согласен 1-2-3-4-5. Полностью согласен (целое)
10. Я бы предпочел иметь много друзей, чем много денег. Полностью не согласен 1-2-3-4-5. Полностью согласен (целое)
11. Я всегда стараюсь быть самым смешным: сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
12. Иногда я могу столкнуться с двумя лицами: Полностью не согласен 1-2-3-4-5 Полностью согласен (целое)
13. Я повредил вещи в прошлом, когда сердился .: Сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
14. Я не тороплюсь, чтобы принимать решения: сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
15. Я всегда стараюсь голосовать на выборах: сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
16. Я часто думаю и сожалею о принимаемых мной решениях: Полностью не согласен 1-2-3-4-5 Полностью согласен (целое)
17. Я могу сказать, слушают ли меня люди или нет, когда я говорю с ними: Полностью не согласен 1-2-3-4-5 Полностью согласен (целое)
18. Я ипохондрик. Полностью не согласен 1-2-3-4-5. Полностью согласен (целое)

19. Я - эмфатичный человек. Сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
20. Я ем, потому что должен. Я не наслаждаюсь едой и еду так быстро, как могу: Категорически не согласен 1-2-3-4-5 Полностью согласен (целое)
21. Я стараюсь отдать столько, сколько я могу, другим людям на Рождество: сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
22. Мне не нравится встречаться с животными: Полностью не согласен 1-2-3-4-5 Полностью согласен (целое)
23. Я ухаживаю за вещами, которые я заимствовал у других: сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
24. Я чувствую себя одиноким в жизни: Сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
25. Раньше я учился в школе: Сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
26. Я беспокоюсь о своем здоровье: сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
27. Хотел бы я изменить прошлое из-за того, что я сделал: Полностью не согласен 1-2-3-4-5 Сильно соглашаюсь (целое)
28. Я верю в Бога: Полностью не согласен 1-2-3-4-5 Полностью согласен (целое)
29. У меня всегда хорошие мечты: сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
30. Я всегда отдаю благотворительность. Полностью не согласен 1-2-3-4-5. Полностью согласен (целое)
31. У меня много друзей: сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
32. Сроки: Я часто бываю на ранней стадии. - Я всегда вовремя. - Я часто опаздываю. (Категорично)
33. Вы лжете другим: Никогда. - Только чтобы не причинять кому-либо вреда. Иногда. - Каждый раз мне это подходит. (Номинальная)
34. Я очень терпелив: сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
35. Я могу быстро адаптироваться к новой среде.: Сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
36. Мое настроение меняется быстро: сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
37. Я хорошо воспитан, и я ухаживаю за внешностью: Полностью не согласен 1-2-3-4-5 Полностью согласен (целое)
38. Мне нравится встречаться с новыми людьми.: Сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
39. Я всегда позволяю другим людям узнать о моих достижениях: Полностью не согласен 1-2-3-4-5 Полностью согласен (целое)

40. Я думаю, тщательно, прежде чем отвечать на любые важные письма .: Полностью не согласен 1-2-3-4-5 Полностью согласен (целое)
41. Я люблю детей: сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
42. Я не боюсь высказать свое мнение, если я сильно против чего либо: Полностью не согласен 1-2-3-4-5 Полностью согласен (целое)
43. Я могу рассердиться очень легко: сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
44. Я всегда убеждаюсь, что я общаюсь с нужными людьми: сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
45. Я должен быть хорошо подготовлен перед публичным выступлением: Полностью не согласен 1-2-3-4-5 Полностью согласен (целое)
46. Я виноват в том, люди меня не любят. Полностью не согласен 1-2-3-4-5. Полностью согласен (целое)
47. Я плачу, когда чувствую себя, или все идет не так. Абсолютно не согласен 1-2-3-4-5. Полностью согласен (целое)
48. Я на 100% доволен своей жизнью. Полностью не согласен 1-2-3-4-5. Полностью согласен (целое)
49. Я всегда полна жизни и энергии: Полностью не согласен 1-2-3-4-5 Полностью согласен (целое)
50. Я предпочитаю крупных опасных собак более мелким, более спокойным собакам: Полностью не согласен 1-2-3-4-5 Полностью согласен (целое)
51. Я считаю, что все мои черты характера положительные: сильно не согласны 1-2-3-4-5. Полностью согласен (целое)
52. Если я найду что-то, что не принадлежит мне, я передам его. Полностью не согласен 1-2-3-4-5. Полностью согласен (целое)
53. Мне очень трудно вставать утром: сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
54. У меня много разных увлечений и интересов: Полностью не согласен 1-2-3-4-5 Полностью согласен (целое)
55. Я всегда слушаю совет моих родителей: Полностью не согласен 1-2-3-4-5. Полностью согласен (целое)
56. Мне нравится участвовать в опросах: сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
57. Сколько времени вы проводите в Интернете ?: Не провожу - Менее часа в день - Несколько часов в день - Большая часть дня (Номинальная)

2.7 НА ЧТО ВЫ ТРАТИТЕ ДЕНЬГИ

1. Я могу экономить все, что смогу. Полностью не согласен 1-2-3-4-5. Полностью согласен (целое)
2. Мне нравится ходить в крупные торговые центры .: Сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
3. Я предпочитаю фирменную одежду: сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
4. Я трачу много денег на вечеринки и общение: Сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
5. Я трачу много денег на внешний вид: сильно не согласен 1-2-3-4-5 Полностью согласен (целое)
6. Я трачу много денег на гаджеты: Полностью не согласен 1-2-3-4-5 Полностью согласен (целое)
7. Я буду с удовольствием платить больше денег за хорошее, качественное или здоровое питание: Полностью не согласен 1-2-3-4-5 Полностью согласен (целое)

2.8 ДАННЫЕ О СЕБЕ

1. Возраст: (целое)
2. Высота: (целое)
3. Вес: (целое)
4. Сколько у вас братьев и сестер?: (целое)
5. Пол: Женский - Мужской (Номинальная)
6. Я: Левая рука - Правая (Номинальная)
7. Высшее образование: В настоящее время ученик начальной школы - Начальная школа - Средняя школа - Колледж / Степень бакалавра (Номинальная)
8. Я единственный ребенок: Нет - Да (Номинальная)
9. Я провел большую часть своего детства в: Городе - деревне (Номинальная)
10. Я прожил большую часть своего детства в: доме - многоквартирном доме (Номинальная)

3 Описательные статистики

Из-за большого количества признаков, рассмотрим лишь подмножество из множества ответов, а именно рассмотрим предпочтения в музыке. Для каждой характеристики рассчитаем

- mean – средняя величина из исходных значений $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$;
- std – стандартное отклонение, мера того, насколько широко разбросаны точки данных относительно их средних $s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$
- min, max – минимальное и максимальное значение
- 25%, 50%, 75% квантили соответствующих уровней. Значение, которое заданная случайная величина не превышает с фиксированной вероятностью. Верхняя квантиль включает 25% наибольших чисел в наборе, нижняя, соответственно, 25% наименьших чисел в наборе, 50% соответствует медиане выборки.

```
In [11]: import pandas as pd
import numpy as np
```

```
In [4]: from pandas import Series
def customDescribe(x):
    data = [x.mean(), x.std(), x.min(), x.quantile(0.25), x.median(),
            x.quantile(0.75), x.max(), x.skew(), x.kurtosis(), x.mode().max(),
            x.isnull().sum()]
    names = ['mean', 'std', 'min', '25%', '50%', '75%', 'max',
            'skewness', 'kurtosis', 'mode', 'NAs']
    return Series(data, index=names)

names = pd.read_csv('columns.csv')
df = pd.read_csv('responses.csv')
music = df.iloc[:,2:18]
music.apply(customDescribe)
```

```
Out[4]:
```

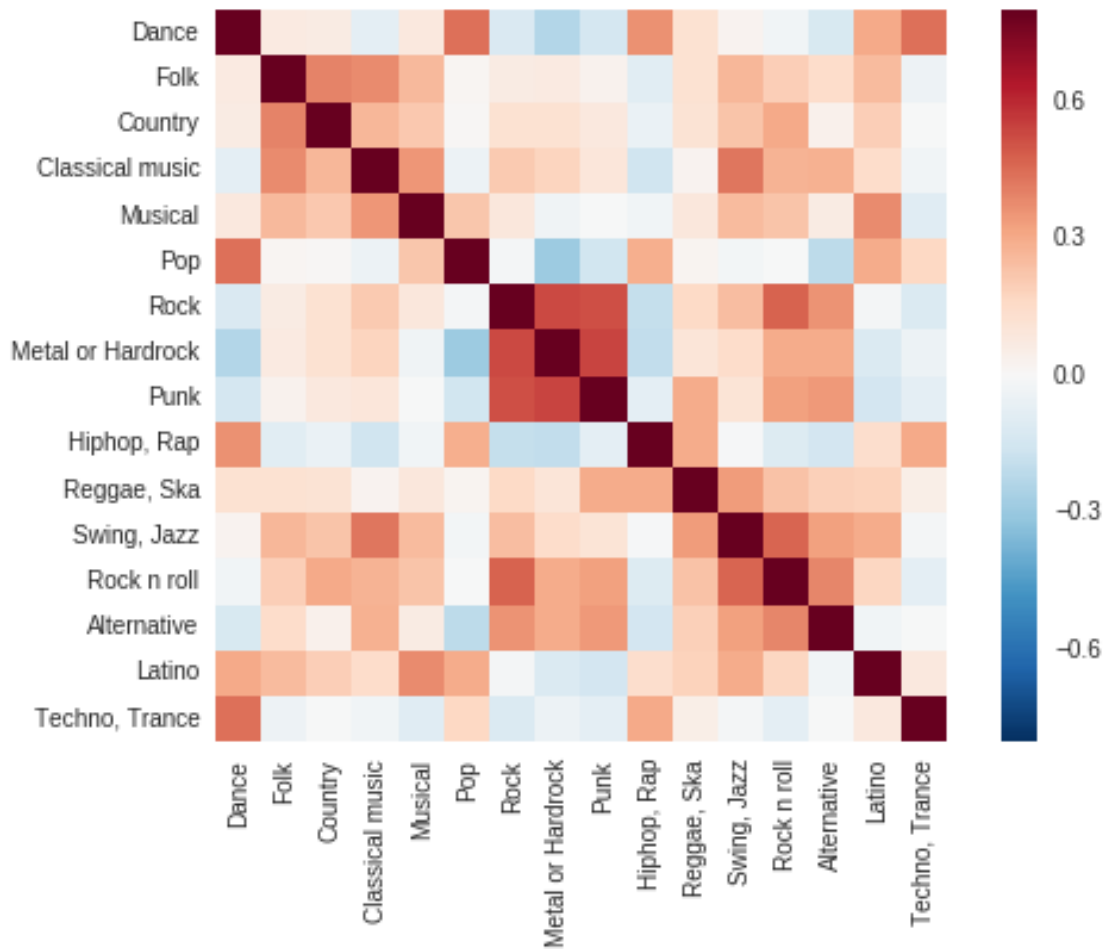
	Dance	Folk	Country	Classical music	Musical	Pop	\
mean	3.113320	2.288557	2.123383	2.956132	2.761905	3.471698	
std	1.170568	1.138916	1.076136	1.252570	1.260845	1.161400	
min	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	
25%	2.000000	1.000000	1.000000	2.000000	2.000000	3.000000	
50%	3.000000	2.000000	2.000000	3.000000	3.000000	4.000000	
75%	4.000000	3.000000	3.000000	4.000000	4.000000	4.000000	
max	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	
skewness	-0.045760	0.694783	0.795798	0.107357	0.219951	-0.383317	
kurtosis	-0.803331	-0.216416	-0.037576	-0.969287	-0.928080	-0.704309	
mode	3.000000	2.000000	2.000000	3.000000	3.000000	4.000000	
NAs	4.000000	5.000000	5.000000	7.000000	2.000000	3.000000	

	Rock	Metal or	Hardrock	Punk	Hiphop, Rap	Reggae, Ska	\
mean	3.761952		2.361470	2.456088	2.910537	2.769691	

std	1.184861	1.372995	1.301105	1.375677	1.214434
min	1.000000	1.000000	1.000000	1.000000	1.000000
25%	3.000000	1.000000	1.000000	2.000000	2.000000
50%	4.000000	2.000000	2.000000	3.000000	3.000000
75%	5.000000	3.000000	3.000000	4.000000	4.000000
max	5.000000	5.000000	5.000000	5.000000	5.000000
skewness	-0.702586	0.604915	0.441427	0.037217	0.156497
kurtosis	-0.419187	-0.934732	-0.959379	-1.250059	-0.900509
mode	5.000000	1.000000	1.000000	4.000000	3.000000
NAs	6.000000	3.000000	8.000000	4.000000	7.000000

	Swing, Jazz	Rock n roll	Alternative	Latino	Techno, Trance
mean	2.759960	3.141575	2.828514	2.842315	2.338983
std	1.257936	1.237269	1.347173	1.327902	1.324099
min	1.000000	1.000000	1.000000	1.000000	1.000000
25%	2.000000	2.000000	2.000000	2.000000	1.000000
50%	3.000000	3.000000	3.000000	3.000000	2.000000
75%	4.000000	4.000000	4.000000	4.000000	3.000000
max	5.000000	5.000000	5.000000	5.000000	5.000000
skewness	0.146457	-0.108936	0.162211	0.188489	0.569644
kurtosis	-0.997739	-0.917436	-1.129404	-1.099347	-0.906037
mode	3.000000	3.000000	3.000000	2.000000	1.000000
NAs	6.000000	7.000000	7.000000	8.000000	7.000000

```
In [15]: import seaborn as sbn
import matplotlib.pyplot as plt
corr = music.corr() #попарная корреляция Пирсона
sbn.heatmap(corr,
             xticklabels=corr.columns.values,
             yticklabels=corr.columns.values,
             vmax=.8, square=True)
plt.show()
```



```
In [69]: import numpy as np
import matplotlib
import matplotlib.pyplot as plt
from sklearn.model_selection import cross_val_score

n_features = len(music.columns)
n_components = np.arange(0, 4)
fa_scores = []
fa = FactorAnalysis()
fa.fit(music)
for n in n_components:
    fa.n_components = n
    fa_scores.append(np.mean(cross_val_score(fa, music)))
n_components_fa = n_components[np.argmax(fa_scores)]
print(n_components_fa)
```

```
In [72]: from sklearn.decomposition import FactorAnalysis
         factor = FactorAnalysis(n_components=3)
         factor.fit(music)
         print ((pd.DataFrame(factor.components_, columns=music.columns)).transpose())
```

	0	1	2
Dance	-0.303149	0.665696	-0.425720
Folk	0.311782	0.402950	0.319220
Country	0.313558	0.329835	0.181988
Classical music	0.553627	0.341860	0.451575
Musical	0.262464	0.568232	0.329080
Pop	-0.271719	0.560852	-0.217120
Rock	0.801508	-0.061497	-0.227127
Metal or Hardrock	0.876026	-0.348435	-0.235428
Punk	0.816995	-0.247060	-0.472734
Hiphop, Rap	-0.374550	0.465859	-0.570731
Reggae, Ska	0.342663	0.340363	-0.349031
Swing, Jazz	0.604564	0.515972	0.191347
Rock n roll	0.780514	0.306515	-0.001735
Alternative	0.737268	-0.003769	-0.025753
Latino	0.043111	0.813650	0.124829
Techno, Trance	-0.234808	0.353415	-0.452851

```
In [ ]:
```