



# Преместване в нов град и оптимален избор

КАКВИ СА ЖЕЛАНИЯТА НА БЪЛГАРСКИЯ СТУДЕНТ

Християн Марков ф.н. 00147 | Статистическа лаборатория 16/17 | 11.1.2017г.

## ЗАДАЧА НА ПРОЕКТА

Задачата на този проект е да се разбере какви са най-важните неща, от които зависи изборът на нов град за живеене, както и по какъв начин да се оптимизира той. Тъй като емиграцията в България сред младежите е най-голяма, изследването се занимава конкретно с техните предпочитания и интереси.

## ДАННИ

В рамките на 3 седмици бяха анкетирани 74 човека, като анкетата беше разпространена в Фейсбук, и по-специално групите на ФМИ, СУ, както и индивидуално на други студенти. Използвани са отговорите на 72 от тях.

Въпросите в анкетата бяха следните:

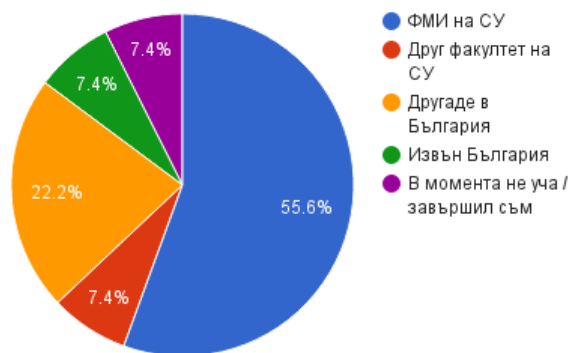
*Каква е Вашата възраст?, Къде учите?, В какво професионално направление учите/работите или сте учили/работили преди?, Къде живеете в момента?, Колко важни са за Вас следните неща, ако трябва да изберете ново място за живеене?.*

За последния въпрос бяха изброени 12 характеристики, които бяха същността на проекта: *разходи за жилище, цена за достъп до Интернет, цена в ресторант, цена за такси, цена за градски транспорт, цена за гориво, ниво на престъпност, ниво на здравеопазване, климат, чистота и природа, език, култура.*

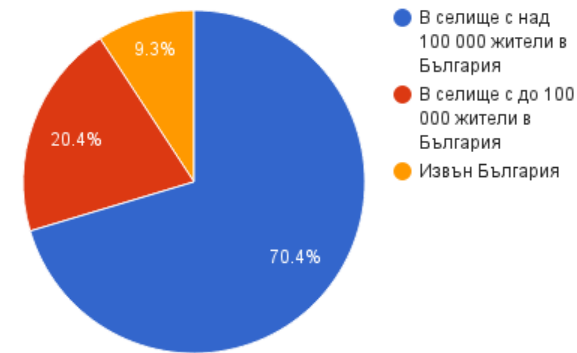
Данните бяха записани на .csv файл, върху който се работеше с **R** (графиките са направени с **Google Sheets** и **R**).

Всяка характеристика беше класирана (coarse classed), така че да получим логична разбивка на стойностите. За основната характеристика rent получихме 3 категории: not important (1), important (2), very important (3).

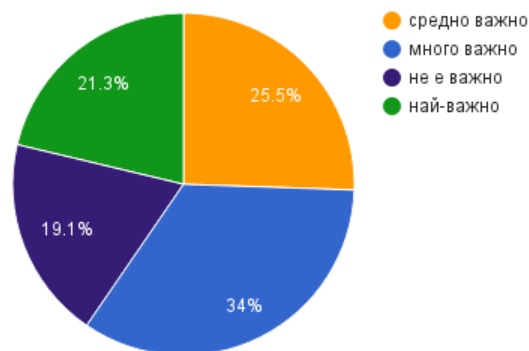
Къде учите?



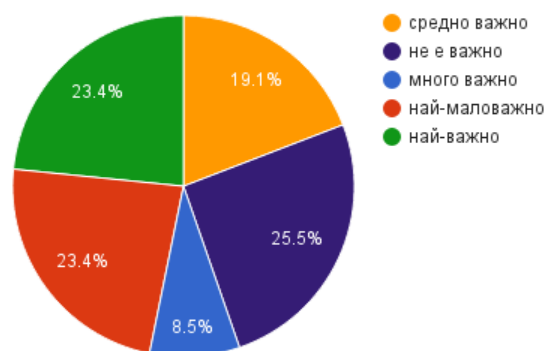
Къде живеете в момента?



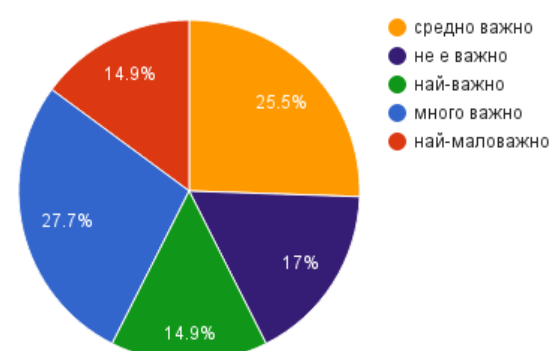
Цена за достъп до интернет



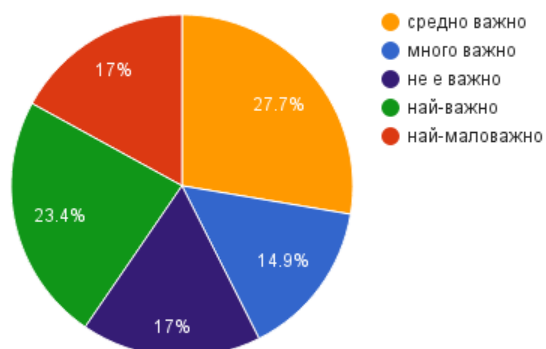
Цена в ресторант



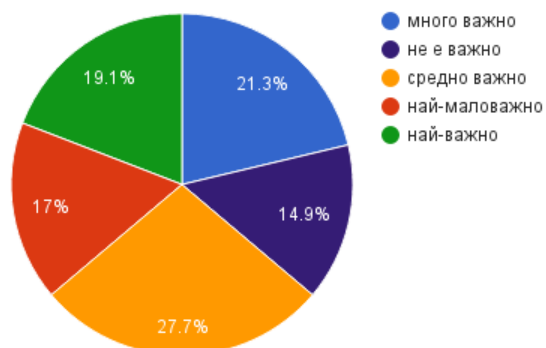
Чистота и природа



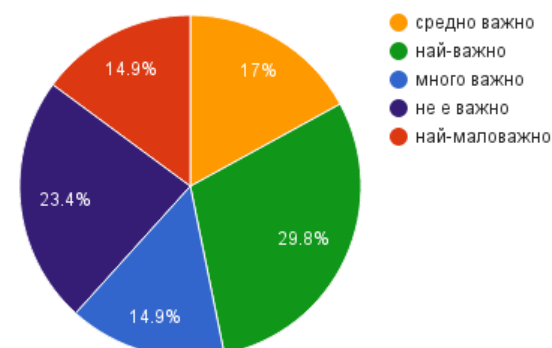
Цена за градски транспорт



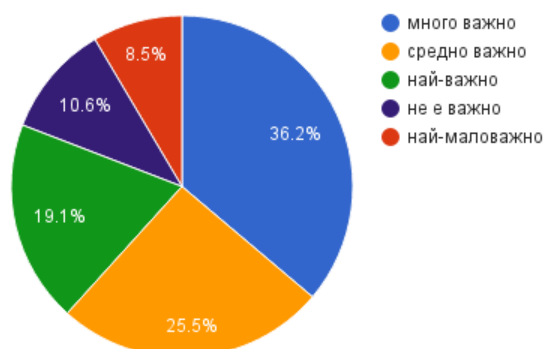
Цена за гориво



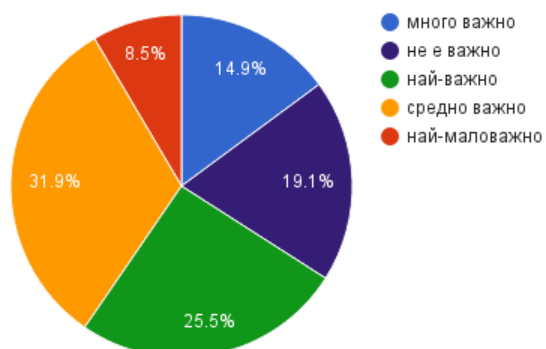
Благоприятен климат



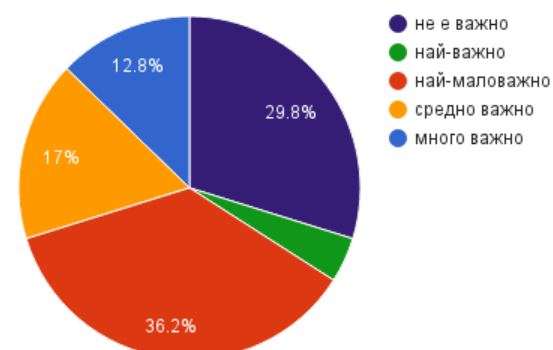
Ниво на престъпност



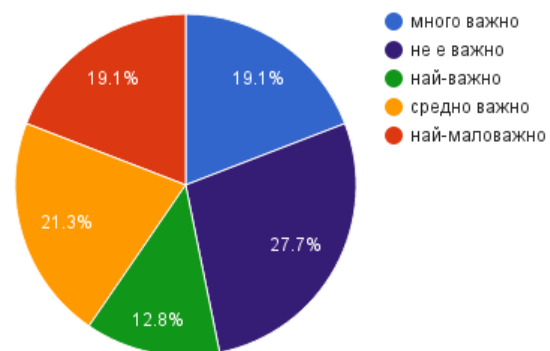
Ниво на здравеопазване



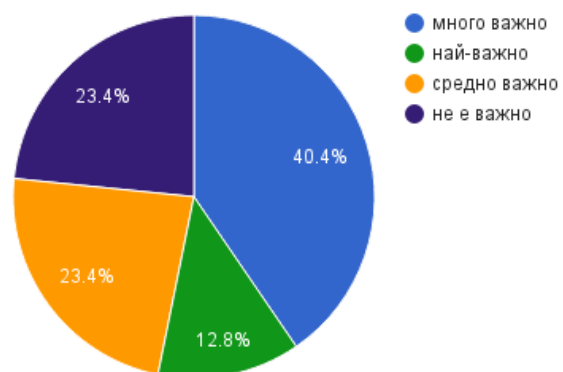
Култура



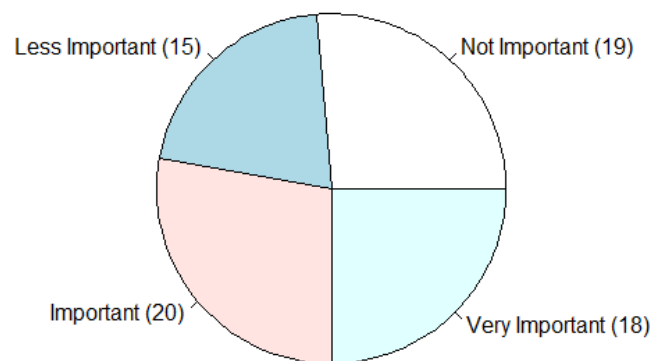
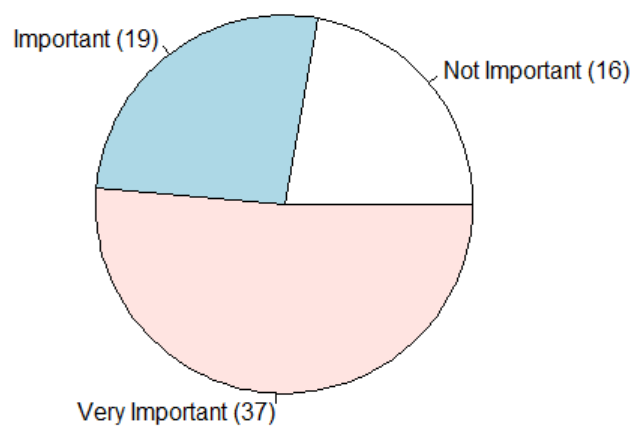
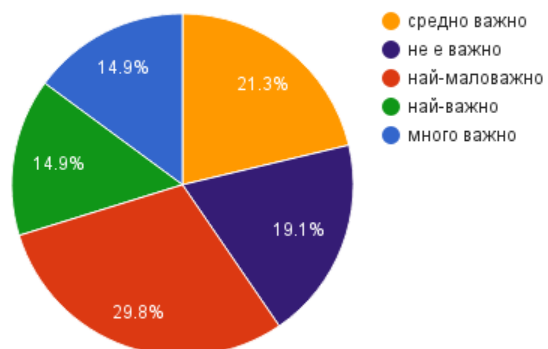
## Език



## Разходи за жилище



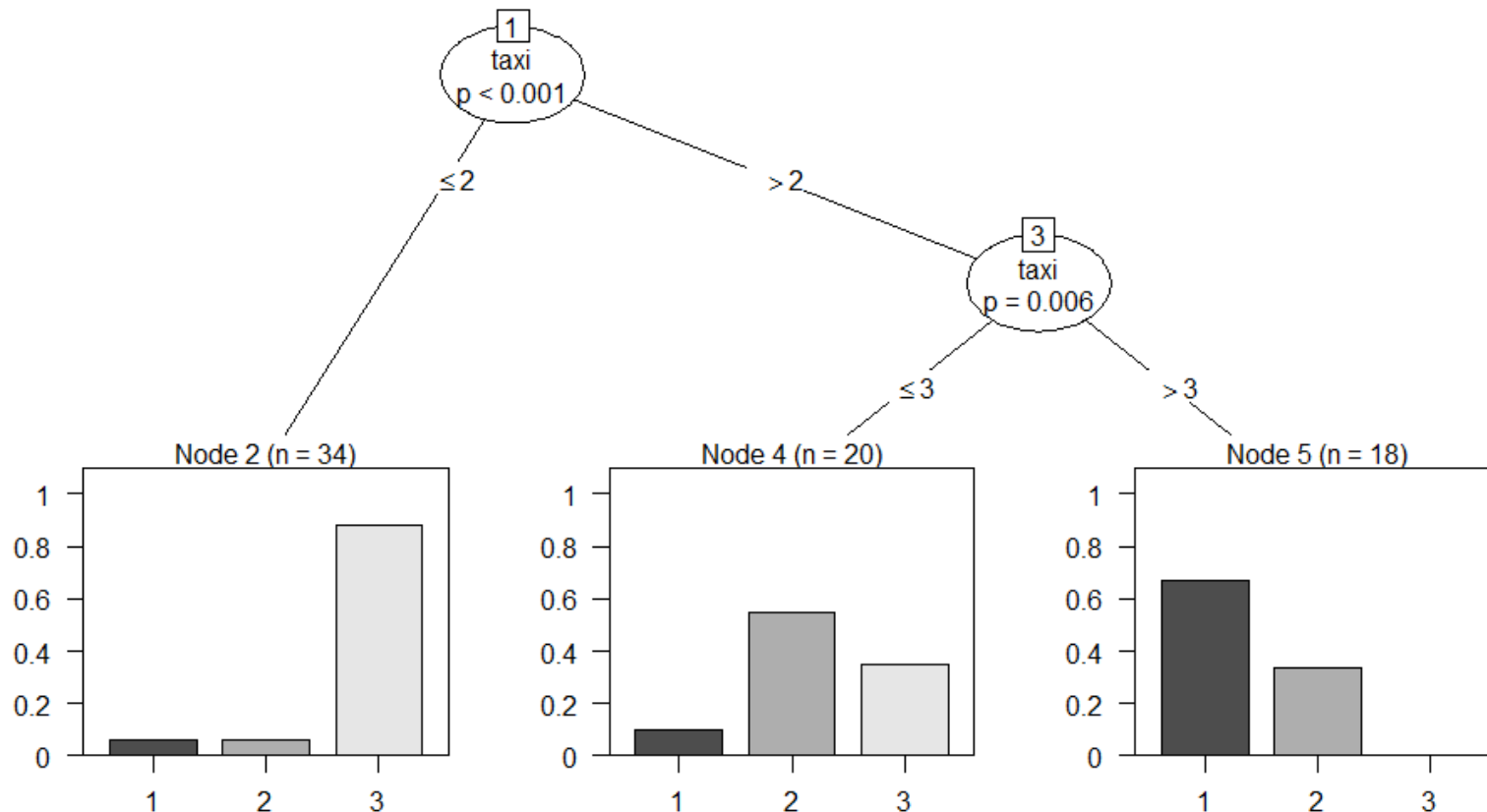
## Цена за такси



## МОДЕЛ И РЕЗУЛТАТИ

Използвано е дърво на решенията (функция `ctree` от пакет `party`, като основен фактор е `rent` – разходите за жилище). По време на работа по проекта бяха изпробвани няколко други модела (например линейна регресия), но дървото на решенията представя много по-добре графично стъпките за избора, и затова се спрях на него. От всичките 12 характеристики, 8 бяха с  $p\text{-value} < 0.05$  (проверено с `chisq.test`). Въпреки това са моделирането ползваме всички. На произволен принцип (с функцията `sample`) данните бяха разделени на `development` (~80%) и `validation` (~20%) групи.

Първоначалното разделение за модела е по характеристиката `taxi`, т.е. цената за пътуване с такси. За хората, които цената за такси не е от особено значение, разходите за жилище са много важни. Втората разбивка също е по `taxi`, там забелязваме, че хората, които държат да ползват тази услуга са склонни да плащат по-висок наем. Хората, за които такситата са от средна важност влизат в средната категория и относно наема.



## РЕШЕНИЕ

След проведеното проучване, можем да заключим, че най-важното е да се проучат цените за наем/покупка на жилище, а после и цените за такси, тъй като те разграничават най-чисто възможностите.

След проверка в сайтовете [numbeo.com](https://numbeo.com) и [theearthawaits.com](https://theearthawaits.com), открих, че повечето големи градове в Източна и Южна Европа покриват изискванията за ниски цени по желаните характеристики.

## ЛЮБОПИТНО

Най-високи коефициенти на корелация наблюдаваме между rent и taxi (-0.71) и culture и health (0.39).

Данните, получени чрез анкетата не са гарантирано изчерпателни. Възможни са грешки и/или изкривявания в предложеното решение, заради броя на отговорилите.

## ВЪПРОСИ?