

Лабораторная работа 2: Обработка данных в pandas.

Описание

В задании приведена база данных `REIGN_2019_10.csv`.

Набор данных «Правители, выборы и нерегулярное управление» (REIGN) описывает политические условия в каждой стране каждый месяц. Эти условия включают срок полномочий и личные характеристики мировых лидеров, типы политических институтов и действующих политических режимов, результаты выборов и объявления о выборах, а также нерегулярные события, такие как перевороты, попытки переворота и другие насилиственные конфликты ». Последний набор данных охватывает 200 стран, с 1950 года по настоящее время, и включает в себя десятки переменных для каждого ежемесячного снимка.

Данные взяты отсюда. Описание всех полей в оригинале можно найти [тут](#).

Дисклаймер. Информация предоставлена исключительно в целях обучения студентов работе с большими наборами данных. Содержание файлов не несет в себе никакой смысловой нагрузки и не преследует никаких политических целей. Ответственность за содержимое несут исключительно авторы сайта oefresearch.org.

Поля, описывающие данные (Внимание!! Представлен приблизительный перевод. Оригинал можно найти в [документации](#)):

- `ccode` - число от 1 до 3 символов - уникальный ИД одной из 201 рассмотренных стран
- `country` - название страны
- `leader` - имя / фамилия лидера
- `year` - календарный год от 1950 до 2019
- `month` - месяц, число от 1 до 12
- `elected` - 1 для даты, когда проходили выборы президента, иначе 0
- `age` - возраст президента
- `male` - 1 если мужчина, 0 если женщина
- `militarycareer` - 1 если правитель в истории своей карьеры работал в силовых структурах (армия, полиция, спецслужбы, ...)
- `tenure_months` - количество месяцев сколько человек у власти
- `government` - один из 16 типов правления.
 - Presidential Democracy - демократия, при которой основными полномочиями наделен президент
 - Parliamentary Democracy - демократия, при которой парламент более полномочен чем президент
 - Personalist Systems - власть в руках одного человека - не монарха
 - Monarchies - монархия
 - Single-Party Systems - власть принадлежит одной партии.
 - Oligarchies - Власть принадлежит одному человеку, либо партии, в которой этот человек стоит во главе. Человек представляет интересы элиты.
 - Party-Personalist Hybrids - среднее между партией и диктаторством, когда власть принадлежит партии, но полностью подконтрольной одному человеку
 - Military Juntas - во главе стоит военный, но его действия проверяются другими военными службами
 - Indirect Military Juntas - по факту власть принадлежит военным, но по факту ими правят гражданские лица
 - Personalist-Military Hybrids - смесь военной и диктаторской системы
 - Party-Military Hybrids - смесь военной и партийной власти
 - Party-Personalist-Military Hybrids - смесь диктаторской и партийной власти
 - Warlordism - для стран, в которых военное положение разрушило основную форму правления.
 - Foreign-Occupied - когда зарубежные силы захватывают власть
 - Civilian Provisional или Military Provisional - временные режимы правления, которые закрепляются в стране на время перехода к демократии (обычно) либо следуют за переворотами и конституционными кризисами.
- `anticipation` - ожидание. Значение равно 1 если в течение следующих шести месяцев ожидаются перевыборы. Для тех стран, где выборы происходят не через регулярные промежутки времени
- `ref_ant` - 1 если в течение следующих 6 месяцев ожидается референдум
- `leg_ant` - 1 если в течение следующих 6 месяцев ожидаются выборы, предписанные основным законом страны
- `exec_ant` - 1 если в течение следующих 6 месяцев ожидаются квалификационные выборы
- `irreg_lead_ant` - 1 если в течение следующих нескольких месяцев ожидаются внеочередные выборы
- `election_now` - 1 если выборы проходят в текущем месяце

- election_recent - 1 если выборы прошли менее 6 месяцев назад
- leg_recent , exec_recent , lead_recent , ref_recent , direct_recent , indirect_recent - 1 устанавливается в течение 6 месяцев после выборов разных типов власти
- victory_recent - 1 в течение 6 месяцев после повторной победы правящей партии либо лидера
- defeat_recent - 1 в течение 6 месяцев после поражения правящей партии либо лидера (если сменилось правление)
- change_recent - 1 в течение 6 месяцев после выборов, которые повлекли за собой смену власти
- nochange_recent - 1 в течение 6 месяцев после выборов, которые не повлекли за собой смену власти
- delayed - 1 следующие 6 месяцев после выборов, которые были отменены либо задержаны на два или более месяца
- lastelection - количество месяцев с момента последних прошедших отборов (выборы на высшую должность или референдум, который расширит исполнительную власть), или, в отсутствие предыдущих выборов, количество месяцев с момента последнего изменения политической системы.
- loss - это количество месяцев с момента, когда действующая политическая партия или лидер в последний раз проиграли выборы или, в отсутствие предыдущих потерь, количество месяцев с момента последнего изменения политической системы
- irregular - аналогично loss, но с момента последнего переворота в стране
- prev_conflict -
- pt_suc - 1 если в этом месяце произошел успешный переворот с лидером
- pt_attempt - 1 если в этом месяце произошла попытка переворота против лидера
- precip - (нет в документации)
- couprisk - риск переворота в стране в текущем месяце
- pctile_risk - тот же риск, в процентах

Задача

С помощью библиотеки pandas для Python получить данные из базы и ответить на следующие вопросы:

1. Сколько форм правления (government) представлено в наборе?
2. Каков средний возраст правителей?
3. Процентное соотношение мужчин и женщин среди правителей (без учета срока правления)

**пояснение: один правитель может править 1 год, тогда в базе данных его фамилия появится 12 раз. А другой может править 10 лет, тогда будет 120 строк. Необходимо учитывать только уникальные значения имени правителя.*

**еще пояснение: в наборе есть тезки: люди с одинаковыми именами, но правящие в разных странах в разное время. Это тоже надо учесть*

4. Процентное соотношение мужчин и женщин с учетом срока правления

**пояснение: один правитель может править 1 год, тогда в базе данных его фамилия появится 12 раз. Надо учитывать его фамилию все 12 раз. А другой может править 10 лет, тогда будет 120 строк. Значит он повлияет на выборку в 10 раз больше.*

**еще пояснение: в наборе есть тезки: люди с одинаковыми именами, но правящие в разных странах в разное время. Это тоже надо учесть*

5. Каковы средние значения и среднеквадратичные отклонения возраста правителей которые правят более 10 лет (120 месяцев) и менее 120 месяцев?
6. Правда ли, что каждый 10 правитель (10% лидеров) оставались у власти более чем на 1 срок?
7. Выведите статистику возраста и пола для каждой из страны. Используйте groupby и describe. Найдите таким образом максимальный и минимальный возраст лидеров США.
8. Посчитайте среднегодовой риск переворота (pctile_risk) в стране для каждого лидера. Для кого и в каком году он был максимальным? А минимальным?
9. При какой форме правления чаще всего проходили государственные перевороты?
10. Постройте график зависимости принадлежности президента к силовым структурам (militarycareer) от формы правления (government). Для выполнения сперва сгруппируйте данные по лидерам, потом используйте crosstab.

Рекомендации и требования

Сдача лабораторной работы предполагает объяснение преподавателю кода, с помощью которого получен результат. Для выполнения работы рекомендуется использовать jupyter - графический интерфейс для работы с python. Задача и описание продублированы в файле `task_REGIN.ipynb` - это формат, с которым работает jupyter.

Для обучения в комплекте с работой предоставлен файл `lection.ipynb` - материалы лекций. Перед выполнением лабораторной работы рекомендуется ознакомиться с этими материалами.