

Проект по ТАБД

I.Содержание проекта

1) Подготовить файл csv по субъектам РФ, который будет состоять из 5-ти переменных (наименование федерального округа, наименования субъекта РФ и 3 переменных из статистического ежегодника). В качестве основной статистической информации можно брать, например, сборник «Регионы России. Социально-экономические показатели» (<https://rosstat.gov.ru/folder/210/document/13204>)

Для каждого отдельного проекта берутся свои социально- экономические характеристики

Считать составленный csv файл с помощью встроенных функция чтения, а также с помощью библиотеки pandas.

2) На основании полученных данных необходимо создать словарь по ключам Федеральный округ и вложенный словарь по ключам наименование субъекта РФ.

По запросу программа должна выводить значения переменных, выбранных для проекта.

Также необходимо написать команды, которые будут выводить значения, соответствующие только одному федеральному округу, сделать это с помощью ключей и с помощью циклов.

3) Написать функцию “pr”, которая будет выводить список значений выбранной переменной в процентах к итогу. Применить эту функцию к каждой переменной и сохранить в объекты списки, которые будут хранить значения только процентов к итогу.

Создать список со значениями в 100 % и проверить разницу между этим списком и списками процентов к итогу.

Сделать выводы по полученным данным.

4) Написать самостоятельно функции «mn» - среднее арифметическое, «med» - медиана, «md» - мода, «sd» - среднее квадратическое отклонение, «r» - размах, которые будут рассчитывать характеристики для выбранного списка по запросу пользователя. Запрос пользователя осуществляется через input(). Для запроса подаются названия функций.

Составить список для каждой переменной, каждый из которых будет включать значения всех функций в качестве индекса списка использовать наименование функции, в качестве наименования столбца использовать наименование переменной.

Сделать выводы по полученным данным.

5) Составить матрицу коэффициентов корреляций по исходным переменным, построить тепловую карту, построить график корреляционных полей, сделать выводы

6) Для каждой переменной построить гистограмму распределения, все три графика разместить на одном рисунке.

7) Построить линейные графики для каждой переменной, каждый график должен быть с уникальным цветом, содержать подписи осей, наименование переменной.

8) Построить круговые диаграммы по переменным из пункта 3, на которых отображаются проценты к итогу.

Сделать выводы по всем графикам

9) Создать dataframe, который будет содержать все наблюдения больше средней величины первой переменной.

10) Создать dataframe, который будет содержать все наблюдения меньше медианы по второй переменной.

11) Создать dataframe, который будет отсортирован по возрастанию переменных в следующем порядке: по возрастанию 2ой переменной, по возрастанию 1-ой, по возрастанию 3-ей

Вывести первые 10 значений отсортированной таблицы.

II. Оформление и защита проекта

На защиту представляется программа, написанная на языке Python, в формате .ipynb (Jupyter Notebook) и файл с данными в формате csv.

Файл должен содержать:

- -рабочий читаемый код,
- оформление блоков кода с помощью Markdown
- комментарии к каждому блоку заданий (что код делает);

Защита проходит устно, в свободной форме.

На защите студент демонстрирует программу и взаимодействие с ней (вывод данных по запросу), отвечает на вопросы по коду.

Полезные ссылки:

Графики в Pandas <https://pandas.pydata.org/pandas-docs/stable/visualization.html>