**Домашнее задание**

**Часть 1**

1. Найдите любой набор данных, например, на портале Kaggle.com, содержащий не менее 200 наблюдений (строк) и не менее 10 переменных (столбцов). Файл должен содержать как интервальные, так и категориальные переменные.

2. Дайте его краткое описание. Укажите количество строк и столбцов. По каждой переменной укажите:

- название и метки к переменной и к значениям, которые она принимает (если есть),

- что она измеряет или отражает, в каких единицах измерения,

- тип переменной: количественная (непрерывная/дискретная), категориальная (номинальная/порядковая/бинарная),

- тип данных в колонке (string, integer, пр.),

- число пустых значений.

3. Измените тип данных в столбцах на более подходящий (если применимо).

4. Рассчитайте новый столбец с данными на основе существующего(их).

5. Разбейте любую интервальную переменную на группы и сделайте её категориальной. Обоснуйте выбранные интервалы/группы.

6. Сформируйте срез по таблице и транспонируйте его.

7. Отсортируйте значения по нескольким полям (многоуровневая сортировка) и выведите 20 первых строк таблицы, содержащих столбцы с переменными, по которым проводилась сортировка.

8. Сформулируйте не менее пяти исследовательских гипотез, которые будут проверены в ходе анализа данных.

9. Постройте не менее 5 графиков по имеющимся данным. У графиков должны быть подписи осей, названия и легенда. Каждый график должен отражать некоторые тенденции или взаимосвязи, наблюдаемые в данных. Под каждым графиком нужно написать комментарий, описывающий тенденции или взаимосвязи, которые удалось выявить в результате его построения.

10. Постройте не менее 3 сводных таблицы. Под каждой таблицей нужно написать комментарий, описывающий тенденции или взаимосвязи, которые удалось выявить в результате её построения.

11. Рассчитайте описательные статистики, характеризующие центральную тенденцию и разброс значений относительно меры центральной тенденции по всей выборке и по отдельным группам наблюдений. Под каждым выводом напишите комментарий, описывающий тенденции или взаимосвязи, которые удалось выявить в результате расчётов.

12. Посчитайте корреляции между числовыми данными и интерпретируйте полученные результаты.

13. Сформулируйте общие выводы по результатам анализа данных. Сделайте выводы относительно того, какие исследовательские гипотезы нашли подтверждение в ходе исследования, а какие были отвергнуты и почему.

**Часть 2**

Необходимо собрать данные из Интернета и сохранить их в таблице MS Excel. Задачу нужно придумать самостоятельно. Нужно собрать данные не менее чем со 100 разных веб-страниц. Файл Excel должен содержать не менее 10 разных колонок с данными и, соответственно, не менее 100 строк, каждая из которых содержит данные из одной веб-страницы.