Практика по теме 1.2. Признаки в ML-задачах

Вы будете работать с обезличенной выгрузкой данных по абитуриентам Алтайского государственного университета за 2013-2018 учебные годы – Abit_2013-18.csv.
Результат выполнения каждого пункта задания (с 1 по 9) должен быть выведен на экран.
В коде должны присутствовать комментарии и выводы.

Пункты задания:

1. *Бинаризуйте переменную* «**Статус**», используя таблицу ниже.
Предположим, мы хотим прогнозировать, отчислится ли студент из университета, не закончив успешно обучение. Подумайте, как имеет смысл объединить статусы между собой, чтобы в итоге новые значения переменной «**Статус**» были: 0 – отчислится, 1 – не отчислится.

Код	Статус	Описание
-1	Академический	Находится в академическом отпуске
1	Учащийся	Учится в университете
3	Отчислен из университета	Отчислен
4	Закончил	Закончил обучение
5	Призван	Призван в армию
6	Архив	Отчислены в прошлые годы
7	БывшАбит	Выбыл из участия в поступлении
9	Зачислен	Абитуриент, зачисленный в университет
10	Web-Абитуриент	

Выведите количество полученных значений 0 и 1.

2. *Выведите описательную статистику* для датасета.

3. *Выведите варианты значений* для каждого категориального класса в датасете.

4. *Стандартизируйте признаки, указывающие на географию* (например, «г Барнаул», «Барнаул» и «г. Барнаул» описывают одну и ту же сущность, но из-за человеческого фактора записаны по-разному), *а также другие категориальные признаки*.

5. *Сконструируйте новый признак «Год_рождения»*. Вообще говоря, в датасете остался фрагмент персональных данных в поле «Дата_Рождения». Сконструируйте новый признак «Год_рождения» и запишите туда соответствующее значение, чтобы персональных данных не осталось.

6. *Придумайте и реализуйте способы заполнения пропусков* в данных. Если это невозможно или затруднительно, поясните, почему.
Выведите на экран информацию о числе значений с столбцах по результатам всех замен.

7. *Придумайте способы кодирования имеющихся признаков*, исходя из предположения, что мы хотим прогнозировать, отчислится ли студент из университета, не закончив успешно обучение, с помощью какого-то алгоритма бинарной классификации (логистическая регрессия, древесные модели и т.п.), *обоснуйте свой выбор*. Если это невозможно или затруднительно, поясните, почему.

8. *Придумайте способы генерации новых признаков*, исходя из предположения, что мы хотим прогнозировать, отчислится ли студент из университета, не закончив успешно обучение, с помощью какого-то алгоритма бинарной классификации (логистическая регрессия, древесные модели и т.п.). Сконструируйте минимум два новых признака.

9. *Постройте матрицу корреляции* для выявления попарно зависимых признаков (таблицу значений и тепловую карту). Используйте все числовые при