Лабораторная работа по модулю “Препроцессинг данных”

1. Загрузить исходные данные data.csv
2. Провести первоначальный анализ предложенных данных
3. Предложить стратегию по работе с пропущенными значениями в данных
4. Предложить стратегию по работе с категориальными переменными
5. Провести углубленный анализ данных (корреляции переменных, визуализация взаимосвязей)
6. Разбить данные на train/test
7. Решить задачу бинарной классификации и предсказать переменную ‘SalesCategory’ протестировав нескольких алгоритмов (Logistic Regression, Random Forest)
8. Проверить качество классификации с использованием следующих метрик: Accuracy, F1-Score, Precision, Recall
9. \*Попробовать отобрать наиболее важные переменные (наибольшая корреляция или метод на ваш выбор)

10.\*Сравнить результаты классификации при использовании изначального датасета и датасета с уменьшенной размерностью

11.Загрузить ipython notebook с результатами работы на github репозиторий