

Лабораторная работа 1.

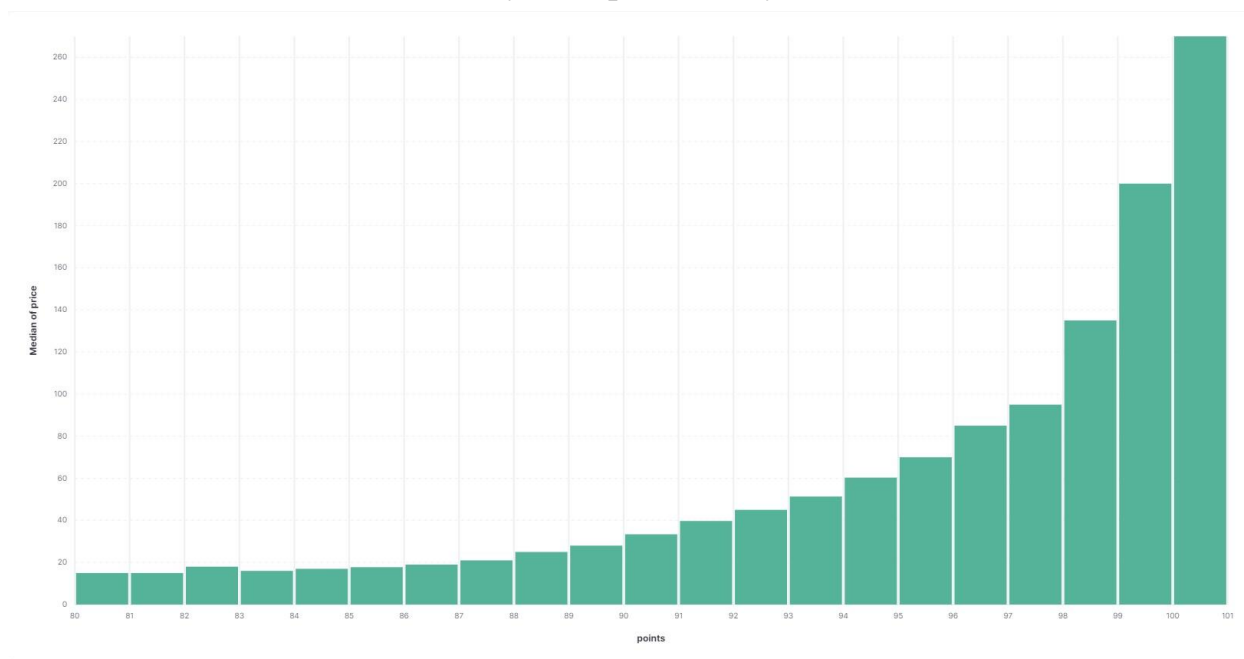
1. Apache Airflow.

Был реализован DAG файл, выполняющий чтение файла, разбиение файла, фильтрацию строк, установки значения и сохранения строки в итоговую таблицу и в ElasticSearch. В основном выполнение этого задания не составило много проблем, так как это было похоже на написание обычной лабораторной в Python, нужно было только почитать немного руководств в интернете. Одной из неприятности являлась долгая отправка данных в ElasticSearch, длившаяся около 8 минут, что заставило побеспокоиться.

2. Apache nifi.

Был реализован шаблон, выполняющий чтение файла, разбиение файла, фильтрацию строк, установки значения и сохранения строки в итоговую таблицу и в ElasticSearch. С nifi возникло много проблем, так как документация была не особо понятна и приходилось искать в интернете информацию по частям. Не очень привычное для меня визуальное программирование также сыграло свою роль. Также возникла проблема с MergeContent, из-за которой её пришлось использовать дважды. В целом nifi оказался намного сложнее для реализации, чем Airflow.

На рисунке представлена гистограмма зависимости стоимости напитка от баллов, поставленных дегустаторами, полученная с помощью Kibana.



Гистограмма зависимости