

**Санкт-Петербургский национальный исследовательский университет
информационных технологий, механики и оптики**

**Отчет по практической работе №2 по дисциплине:
«Проектирование интеллектуальных систем в управлении»**

Подготовила: Лукина А.С.

Группа: К3443

Преподаватель: Бережков Андрей Вячеславович

2021

Задание: Проведение анализа датасета “Титаник” в KNIME.

Ход работы:

В ходе выполнения практической работы был построен поток работ по визуализации этапов разведочного анализа данных для датасета «Титаник». Общая схема потока представлена на рисунке 1.

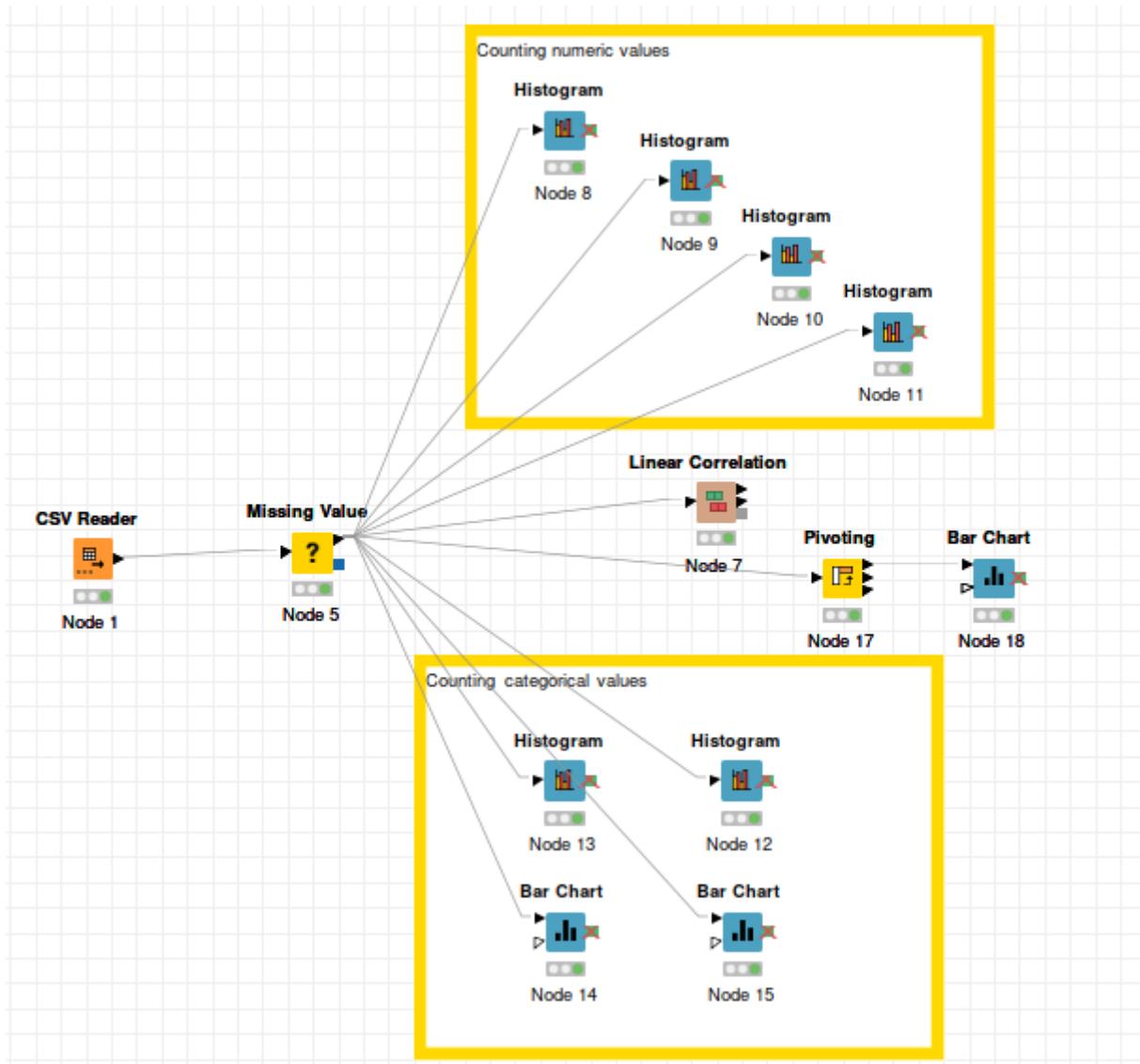


Рисунок 1 — Поток работ по визуализации выборочных признаков

В первую очередь был добавлен узел «CSV Reader» для импорта и чтения датасета. Затем был добавлен узел «Missing Value», необходимый для обработки пропущенных значений. После проделанных шагов была произведена визуализация некоторых признаков датасета.

Во-первых была построена диаграмма распределения количества выживших в зависимости от пола. Для этого необходимо было добавить узел «Pivoting», который составил сводную таблицу для признаков «Sex» и «Survived». Сводная таблица представлена на рисунке 2.

Row ID	\$ Sex	I 0	I 1
Row1	male	468	109
Row0	female	81	233

Рисунок 2 — Сводная таблица для признаков "Sex" и "Survived"

После составления сводной таблицы с помощью узла «Bar Chart» была построена необходимая диаграмма, представленная на рисунке 3.

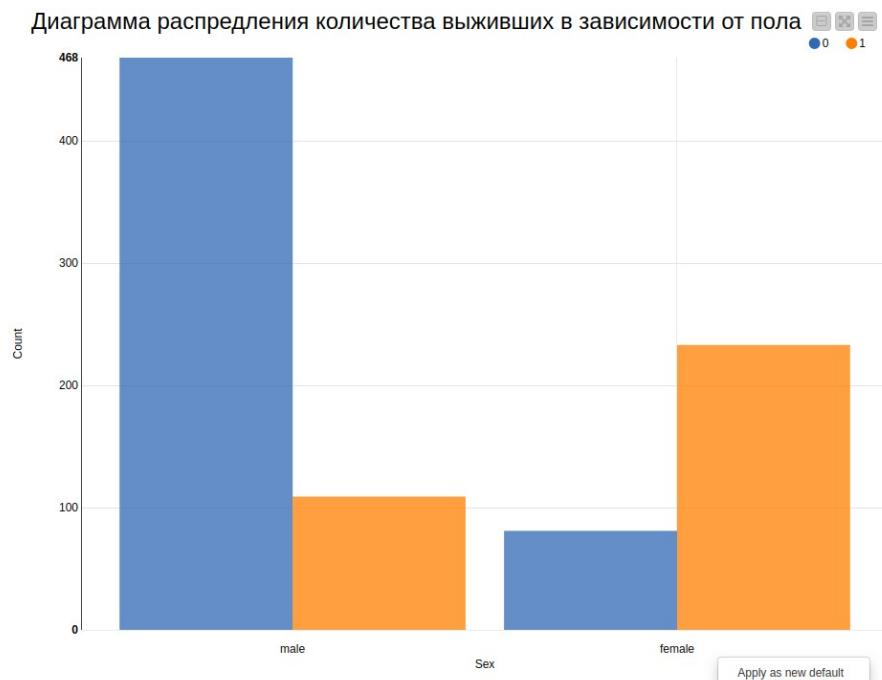


Рисунок 3 — Диаграмма распределения количества выживших в зависимости от пола

Кроме этого были построены диаграммы частотности для числовых признаков — «Age», «SibSp», «Parch», «Fare». Для этого были выбраны узлы типа «Histogram». Полученные диаграммы частотности представлены на рисунках 4-7.

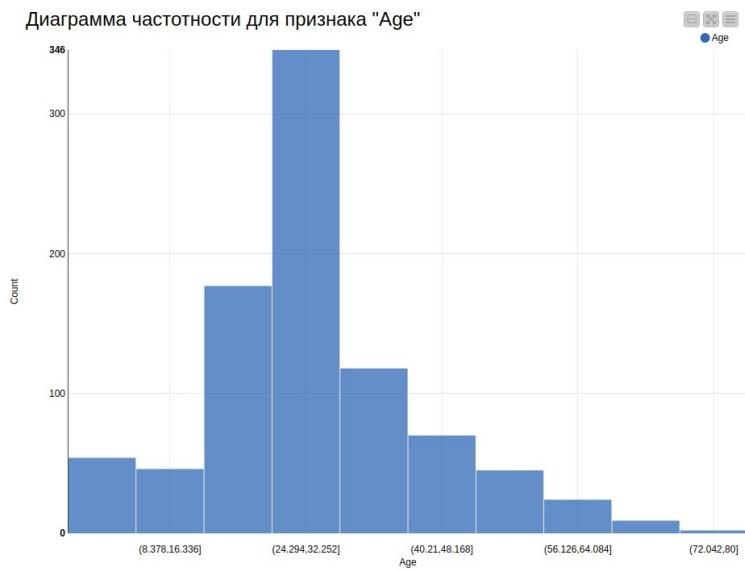


Рисунок 4 — Диаграмма частотности для признака
"Age"

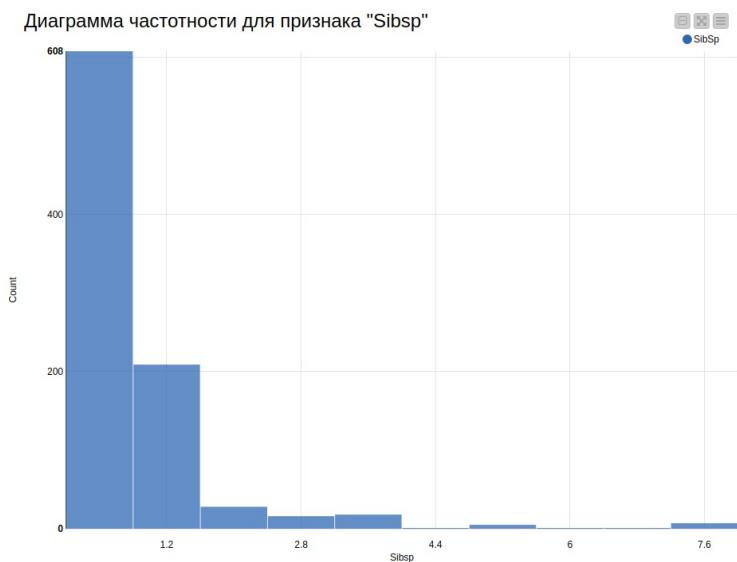


Рисунок 5 — Диаграмма частотности для признака
"SibSp"

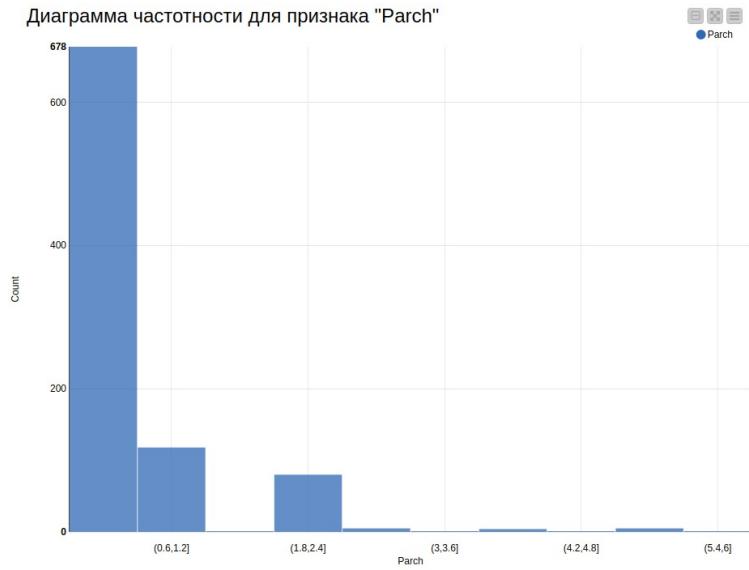


Рисунок 6 — Диаграмма частотности для признака "Parch"

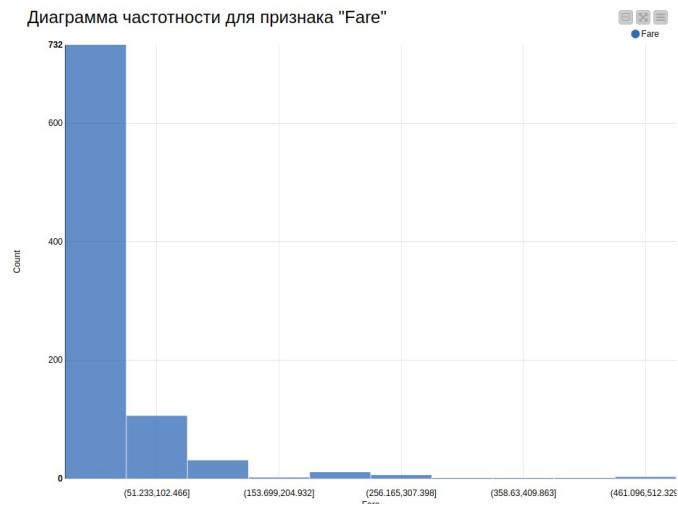


Рисунок 7 — Диаграмма частотности для признака "Fare"

Помимо диаграмм частотности была построена таблица корреляции и тепловая карта для числовых признаков. Корреляционная таблица представлена на рисунке 8. Тепловая карта представлена на рисунке 9.

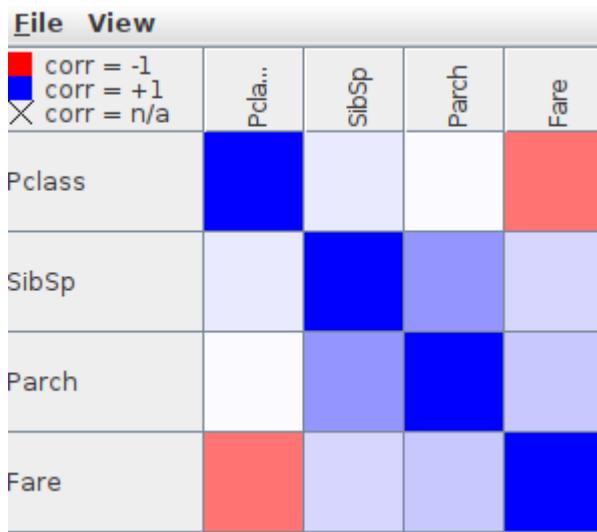


Рисунок 8 — Тепловая карта корреляции численных признаков

Row ID	Pclass	SibSp	Parch	Fare
Pclass	1	0.083	0.018	-0.549
SibSp	0.083	1	0.415	0.16
Parch	0.018	0.415	1	0.216
Fare	-0.549	0.16	0.216	1

Рисунок 9 — Корреляционная таблица для численных признаков

На заключительном шаге были построены диаграммы распределения частоты для категориальных признаков «Survived», «Pclass», «Sex» и «Embarked». Для этого были использованы узлы типа «Histogram» и «Bar Chart». Результаты представлены на рисунках 10-13.

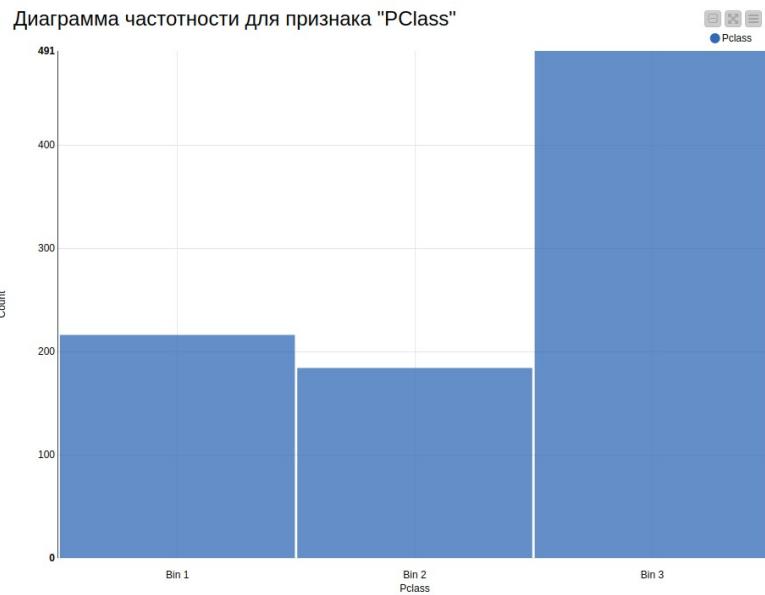


Рисунок 10 — Диаграмма частотности для признака "PClass"

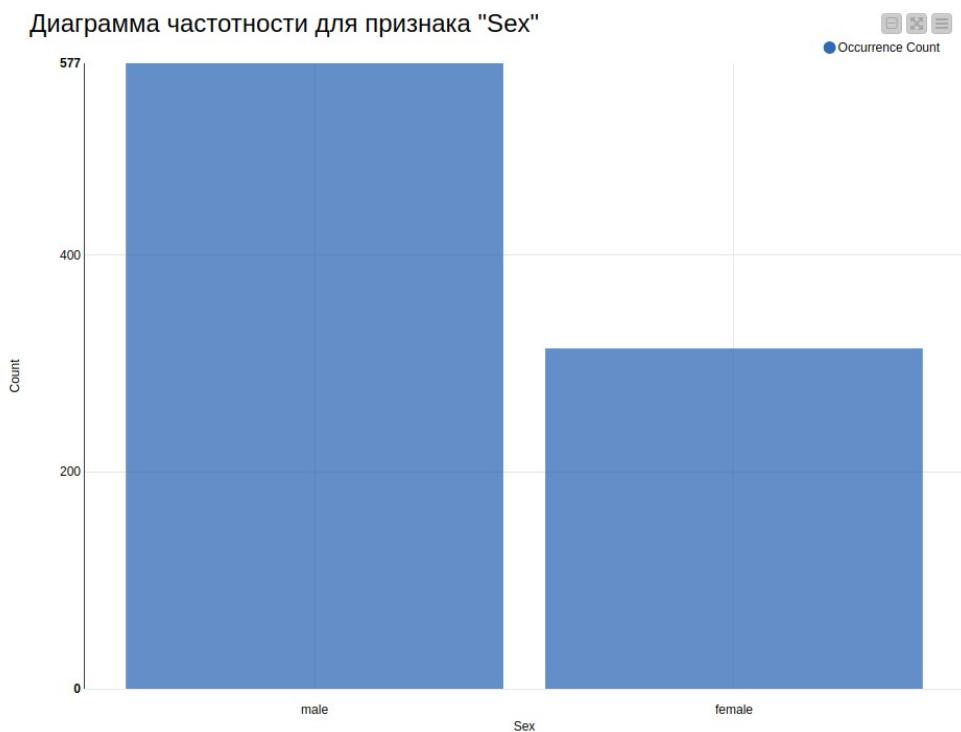


Рисунок 11 — Диаграмма частотности для признака "Sex"

Диаграмма частотности для признака "Embarked"

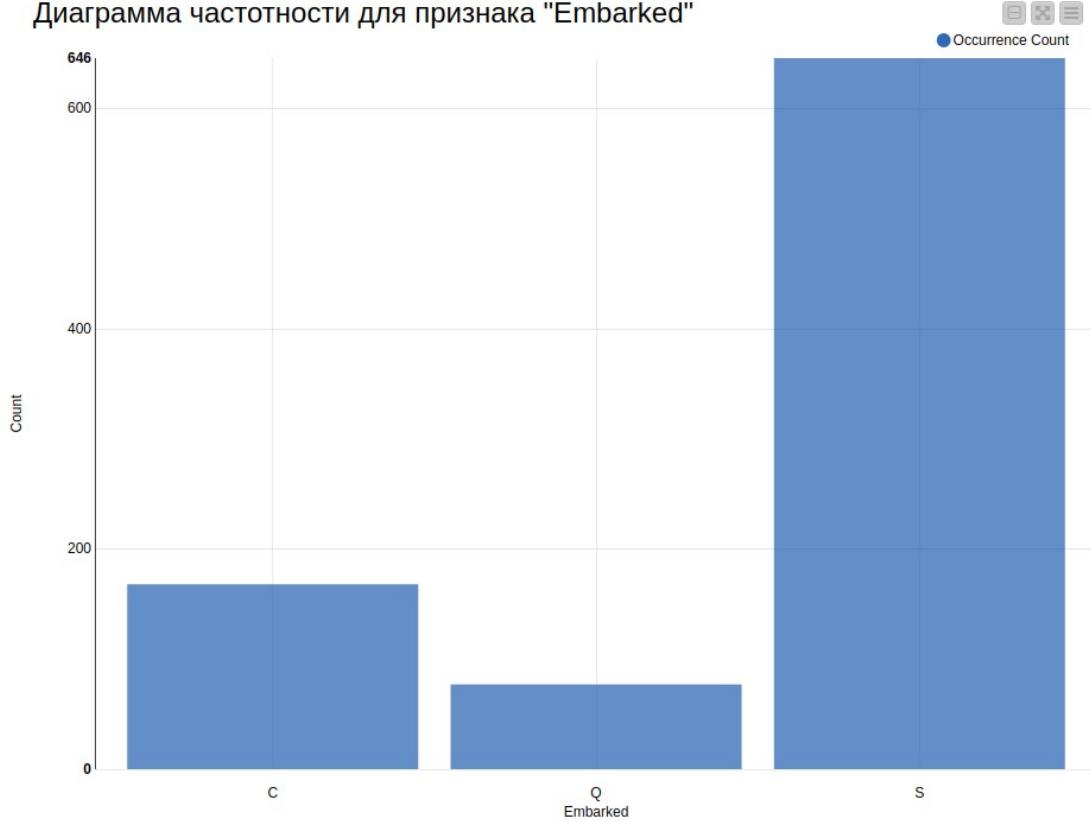


Рисунок 12 — Диаграмма частотности для признака "Embarked"

Диаграмма частотности для признака "Survived"

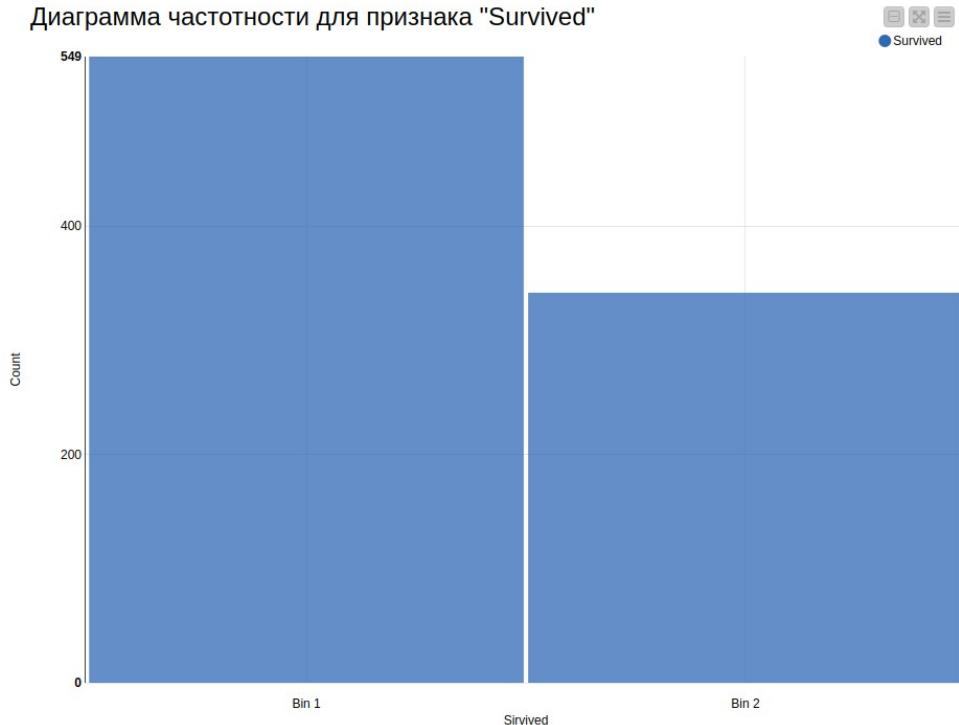


Рисунок 13 — Диаграмма частотности для признака "Survived"