МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение высшего образования

«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ   
АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ»

КАФЕДРА БИЗНЕС-ИНФОРМАТИКИ И МЕНЕДЖМЕНТА

ОЦЕНКА

РУКОВОДИТЕЛЬ

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Канд.тех.наук, доцент |  |  |  | А.Д. Державина |
| должность, уч. степень, звание |  | подпись, дата |  | инициалы, фамилия |

|  |
| --- |
| ЛАБОРАТОРНАЯ РАБОТА №1 |
| Начальное знакомство со средой RapidMiner |
| по дисциплине: Математические методы и модели принятия решений |
|  |
|  |

ВЫПОЛНИЛ

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| СТУДЕНТ ГР. № | 8426МК |  |  |  | Д. Р. Иконникова |
|  |  |  | подпись, дата |  | инициалы, фамилия |

Санкт-Петербург 2024

Цель работы: получение начальных навыков работы с программным комплексом RapidMinerStudio.

Ход выполнения работы:

Был установлен пакет RapidMiner. После чего программа была запущена и был осуществлен переход в начальное окно запуска, представленное на рисунке 1.

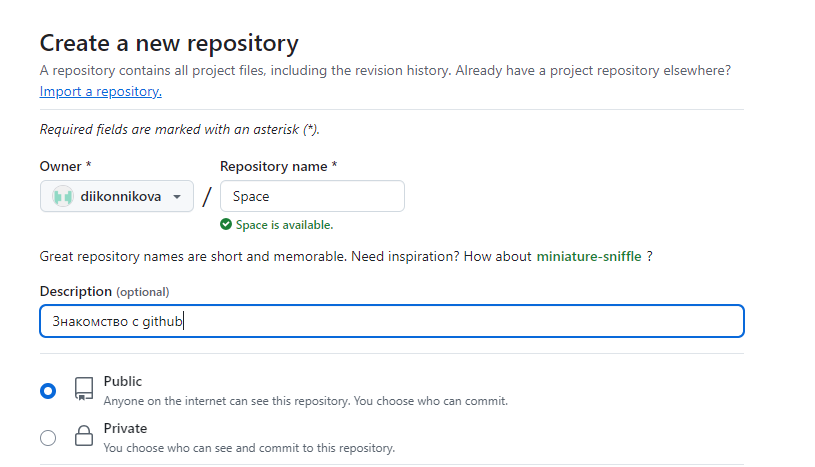


Рисунок 1 – Создание репозитория

После чего был создан новый процесс, что показано на рисунке 2.

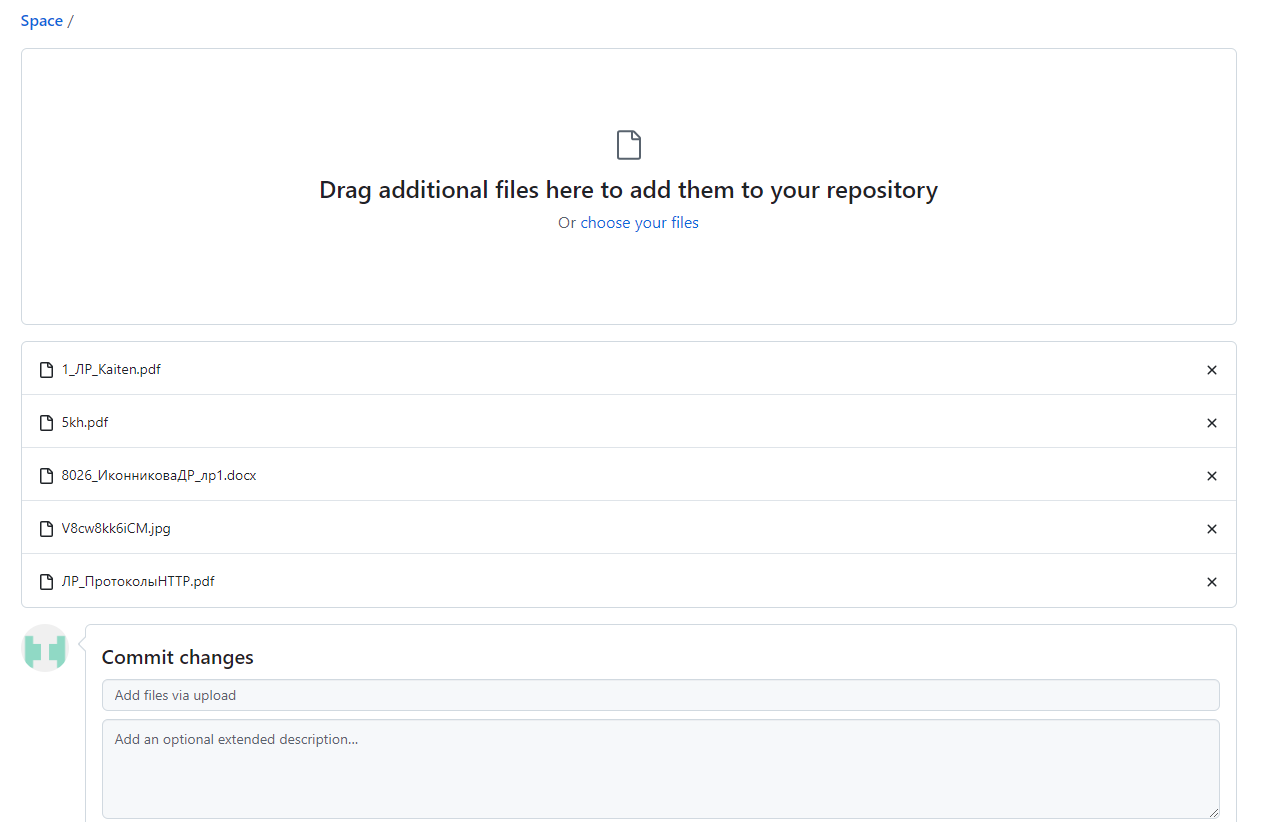


Рисунок 2 – Добавленные файлы

Далее были изучены структуры данных различных наборов данных из папки Samples и определены назначения атрибутов и полей. Для примера было проведено сравнение одноименных наборов Titanic, TitanicTraining и TitanicUnlabeled. Вкладки Data, Statistics, Vizualization для трех наборов данных представлены на рисунках 3 – 11.

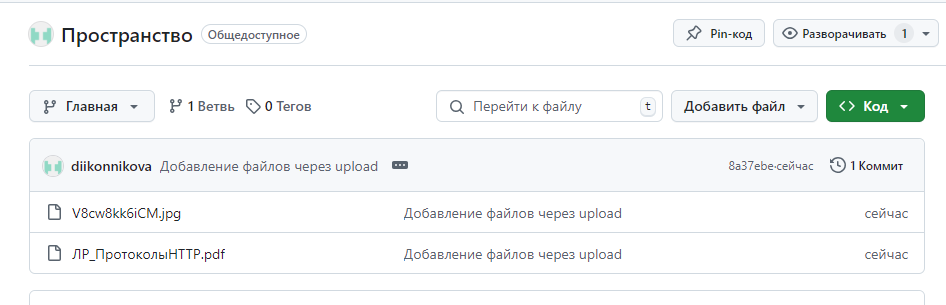


Рисунок 3 – Результат копирования файлов в облачное хранилище

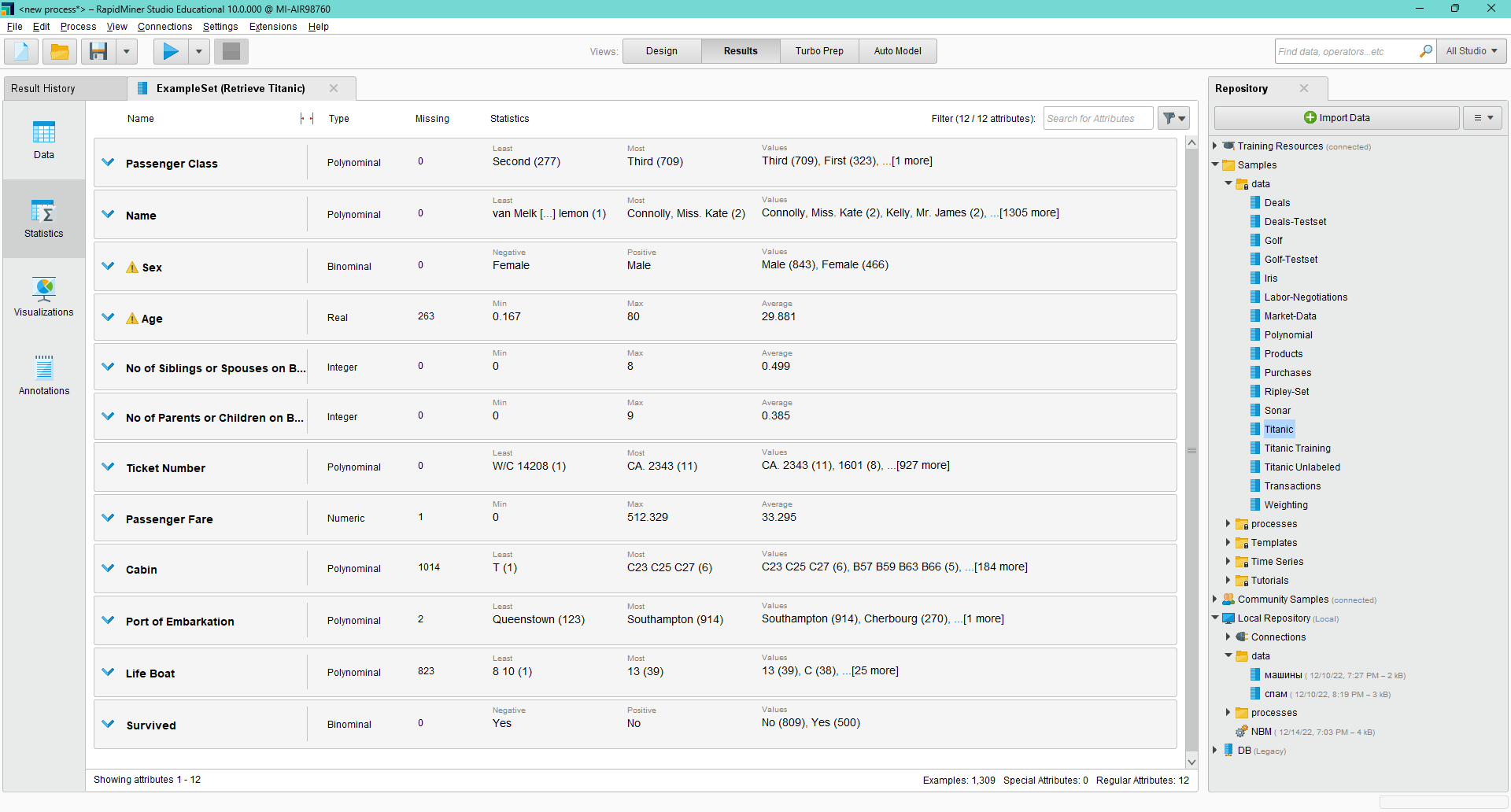


Рисунок 4 - Вкладка Statistics набора Titanic

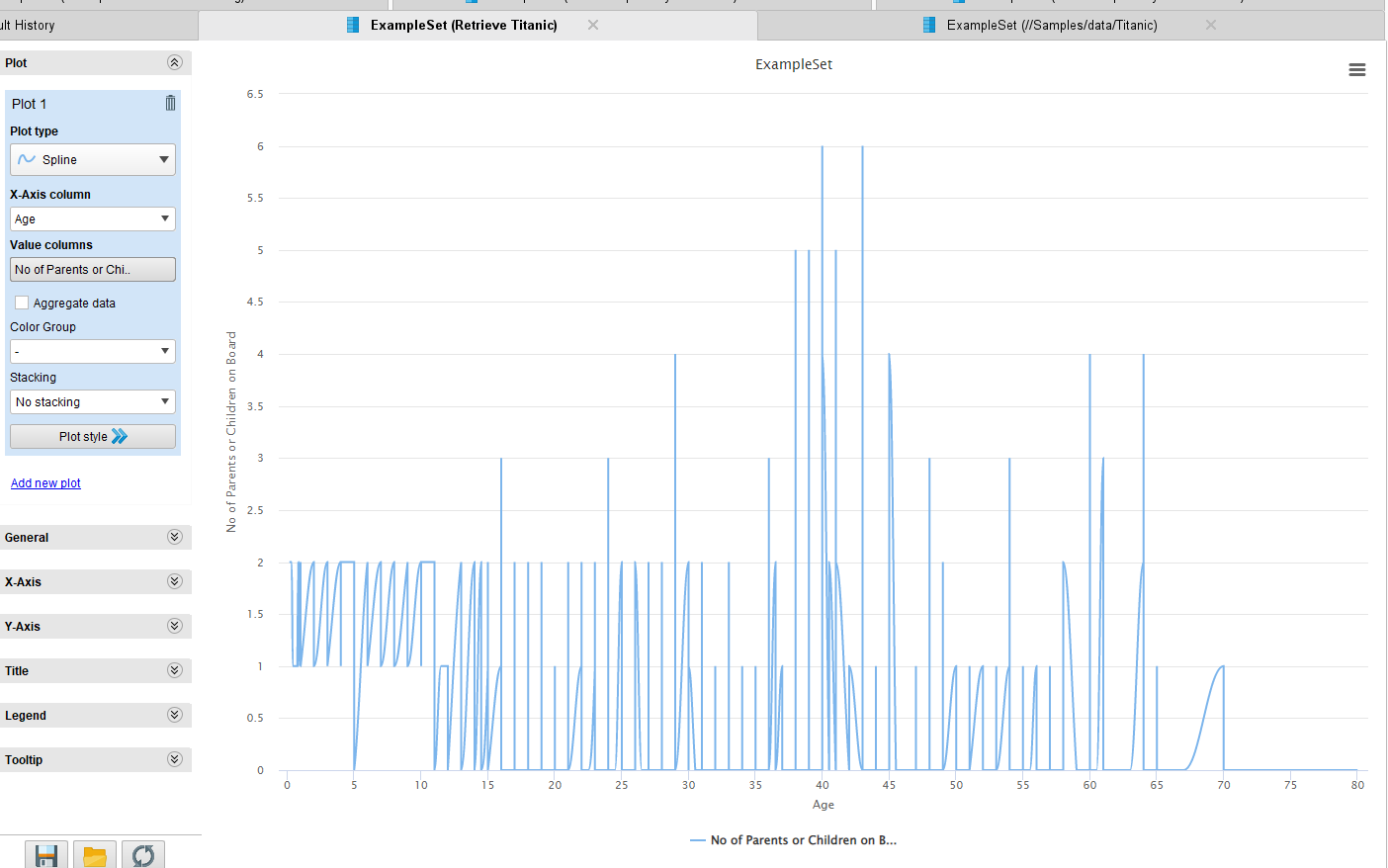


Рисунок 5 - Вкладка Vizualization набора Titanic

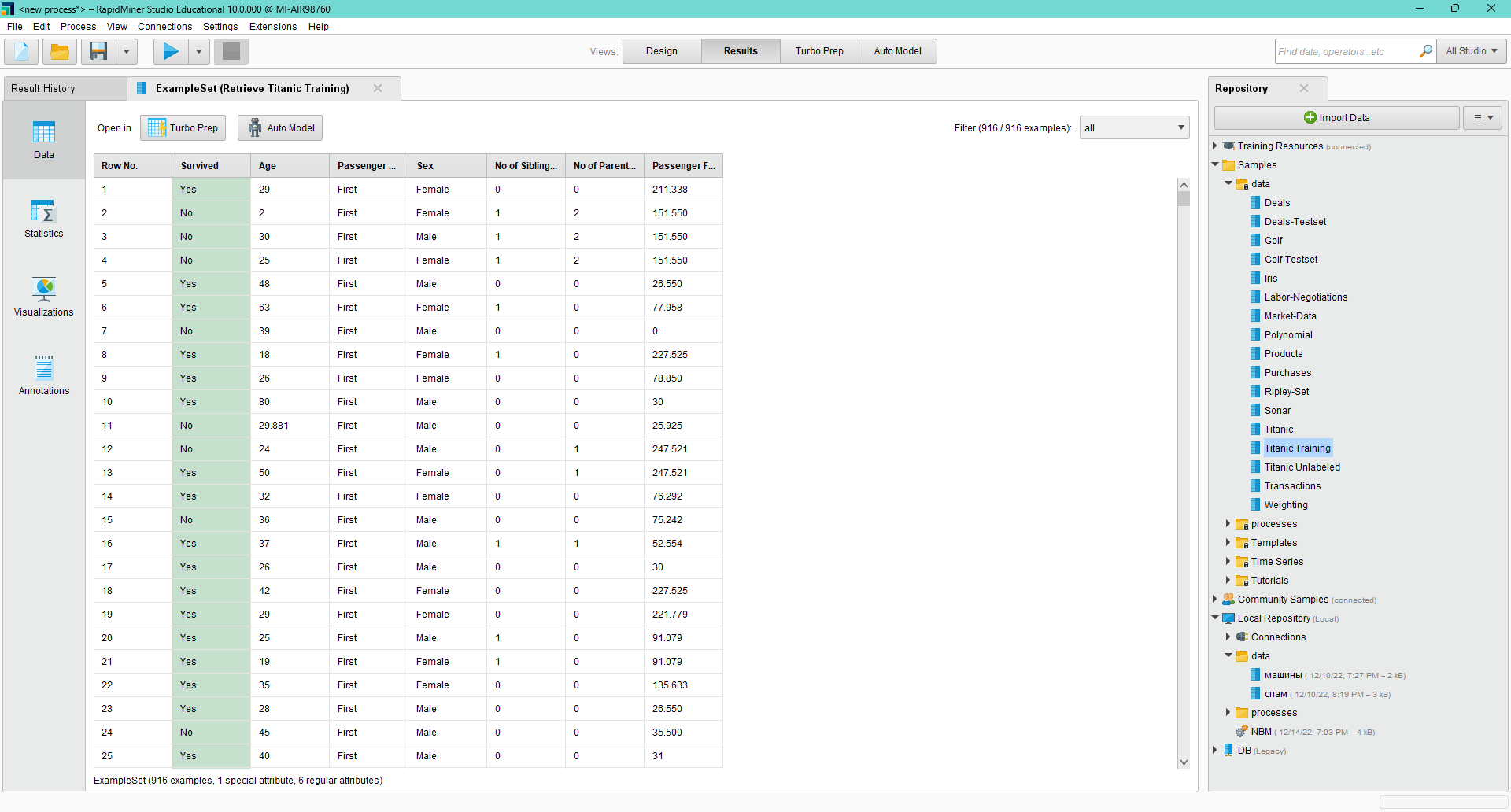


Рисунок 6 - Вкладка Data набора Titanic Training

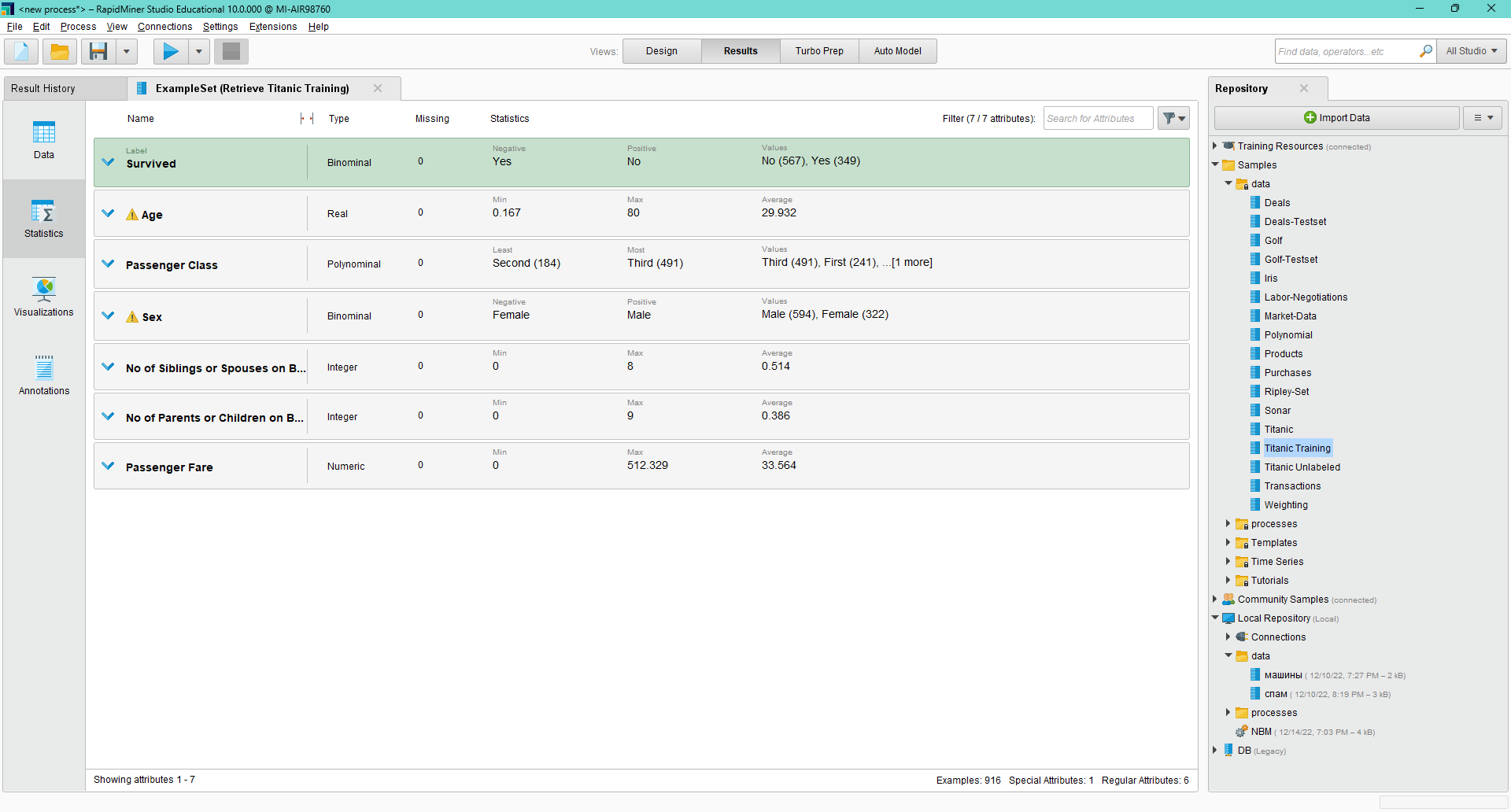


Рисунок 7 - Вкладка Statistics набора Titanic Training

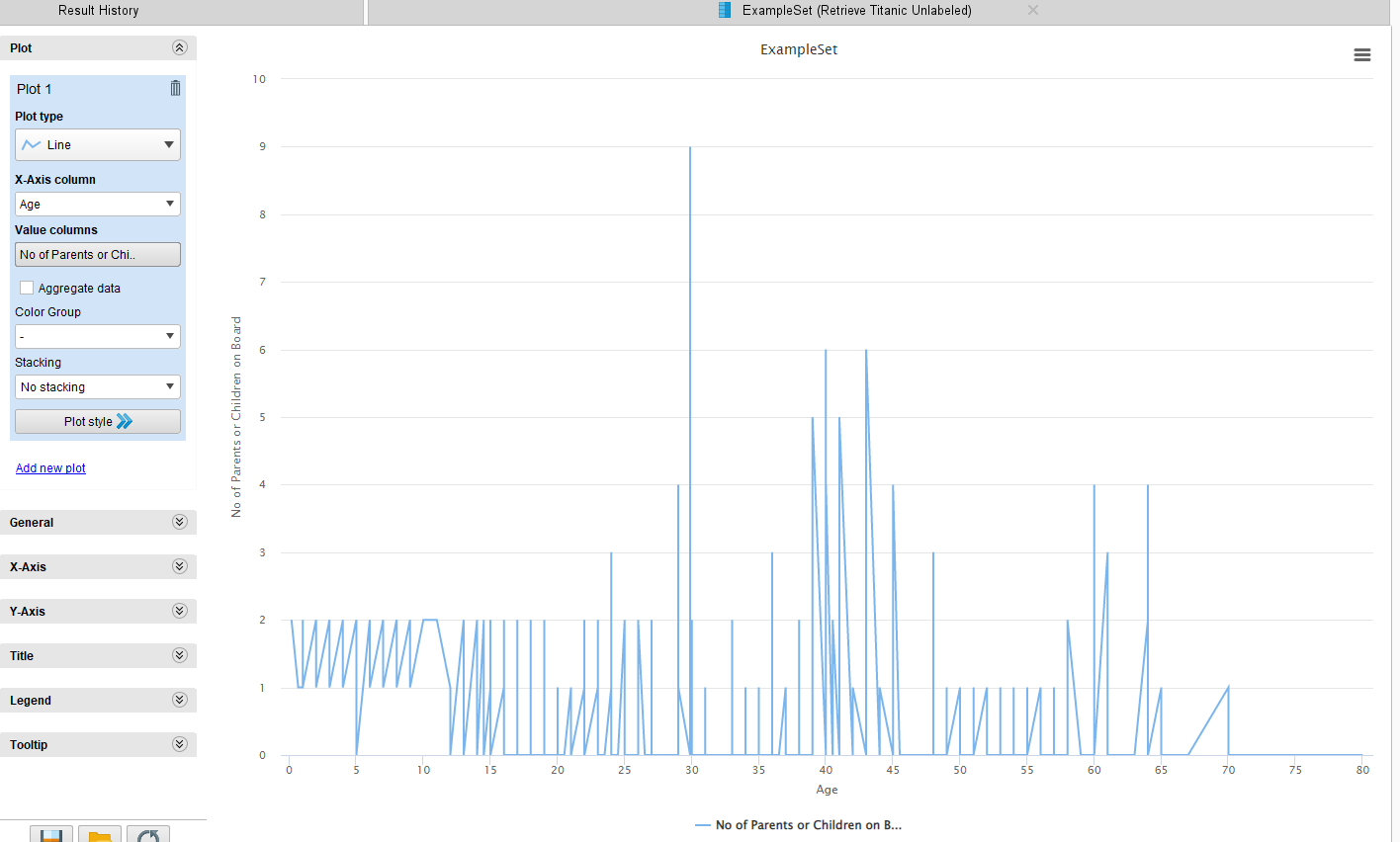


Рисунок 8 - Вкладка Vizualization набора Titanic Training

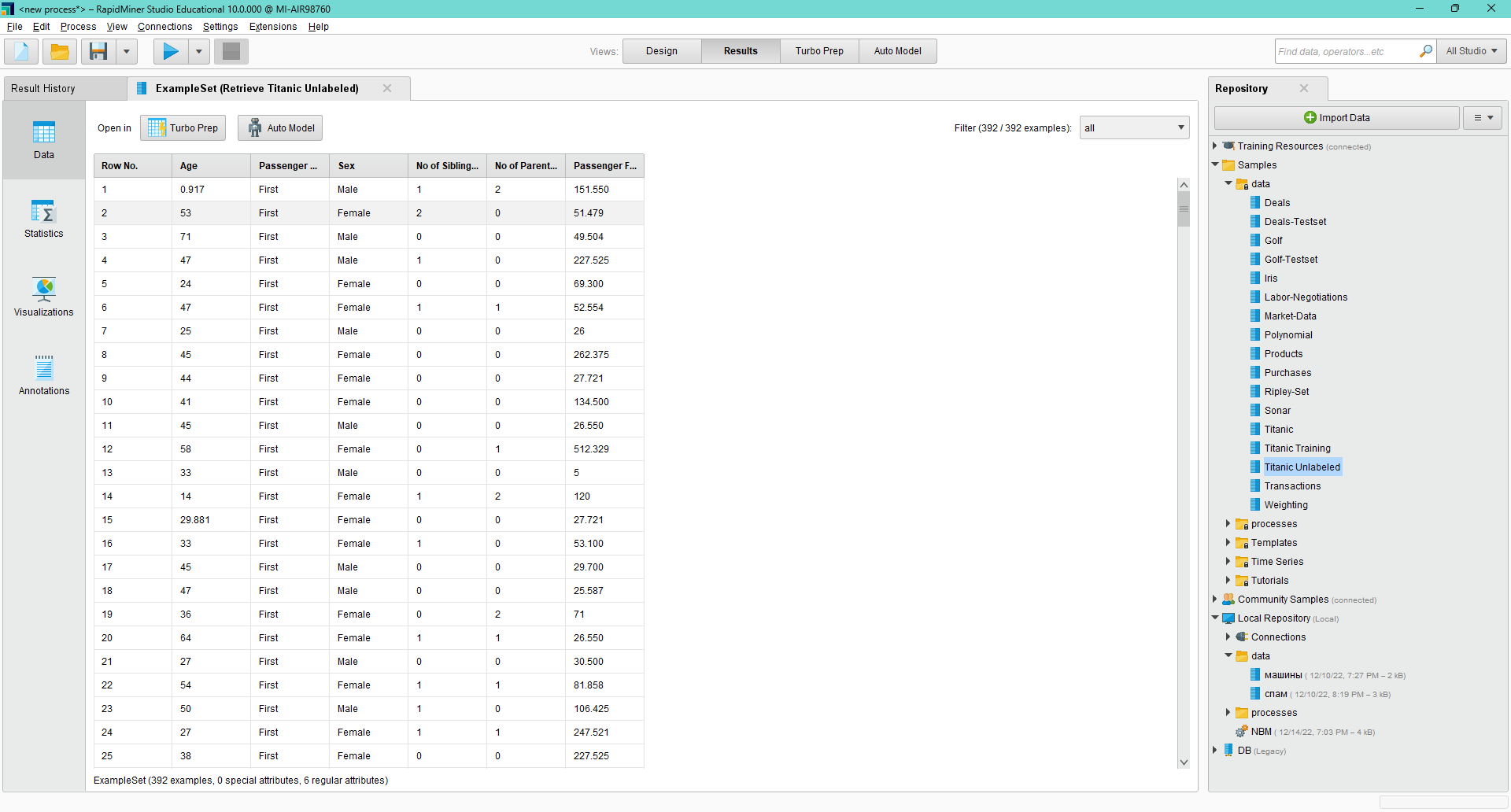


Рисунок 9 - Вкладка Data набора Titanic Unlabeled

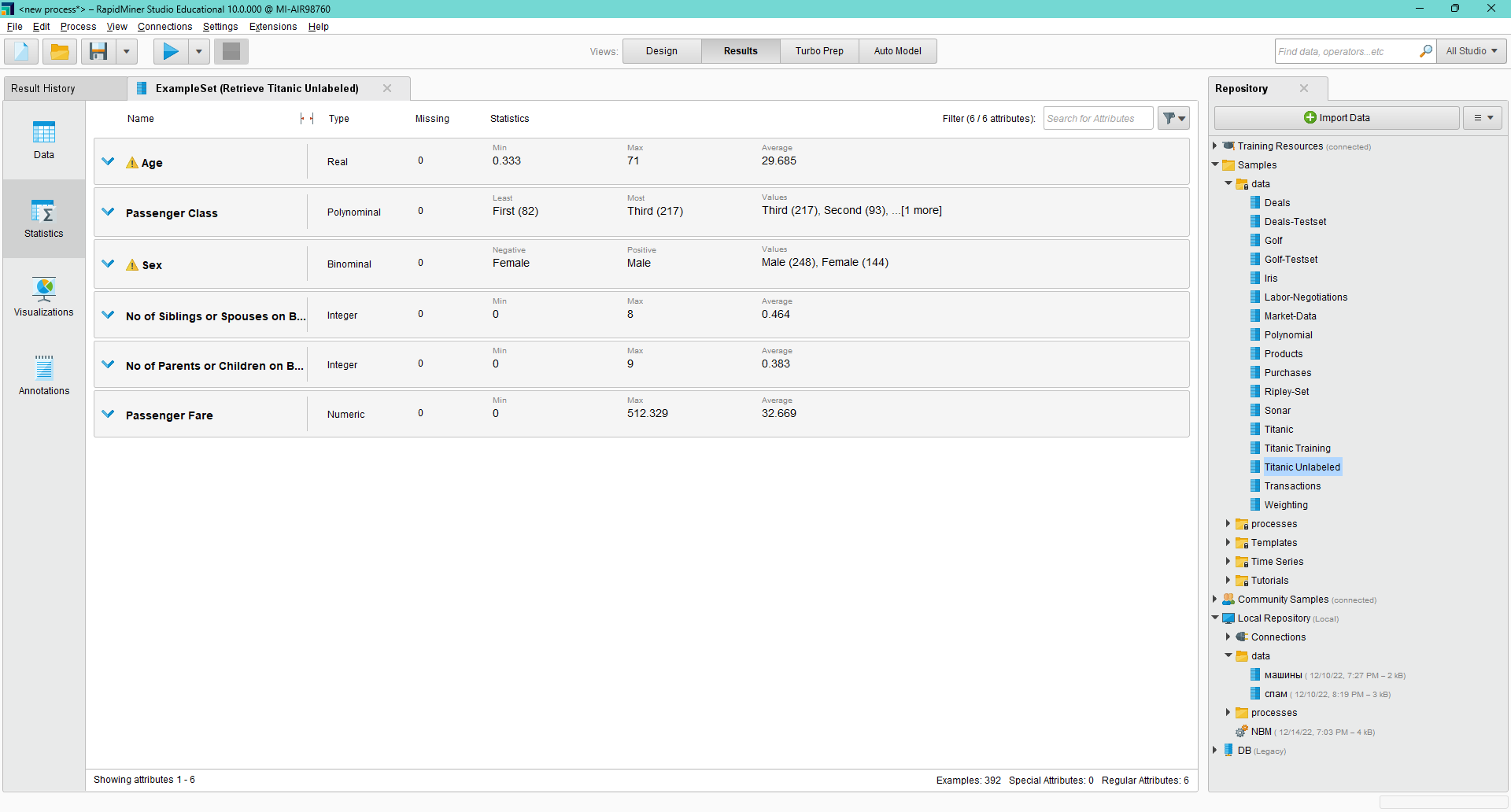


Рисунок 10 - Вкладка Statistics набора Titanic Unlabeled

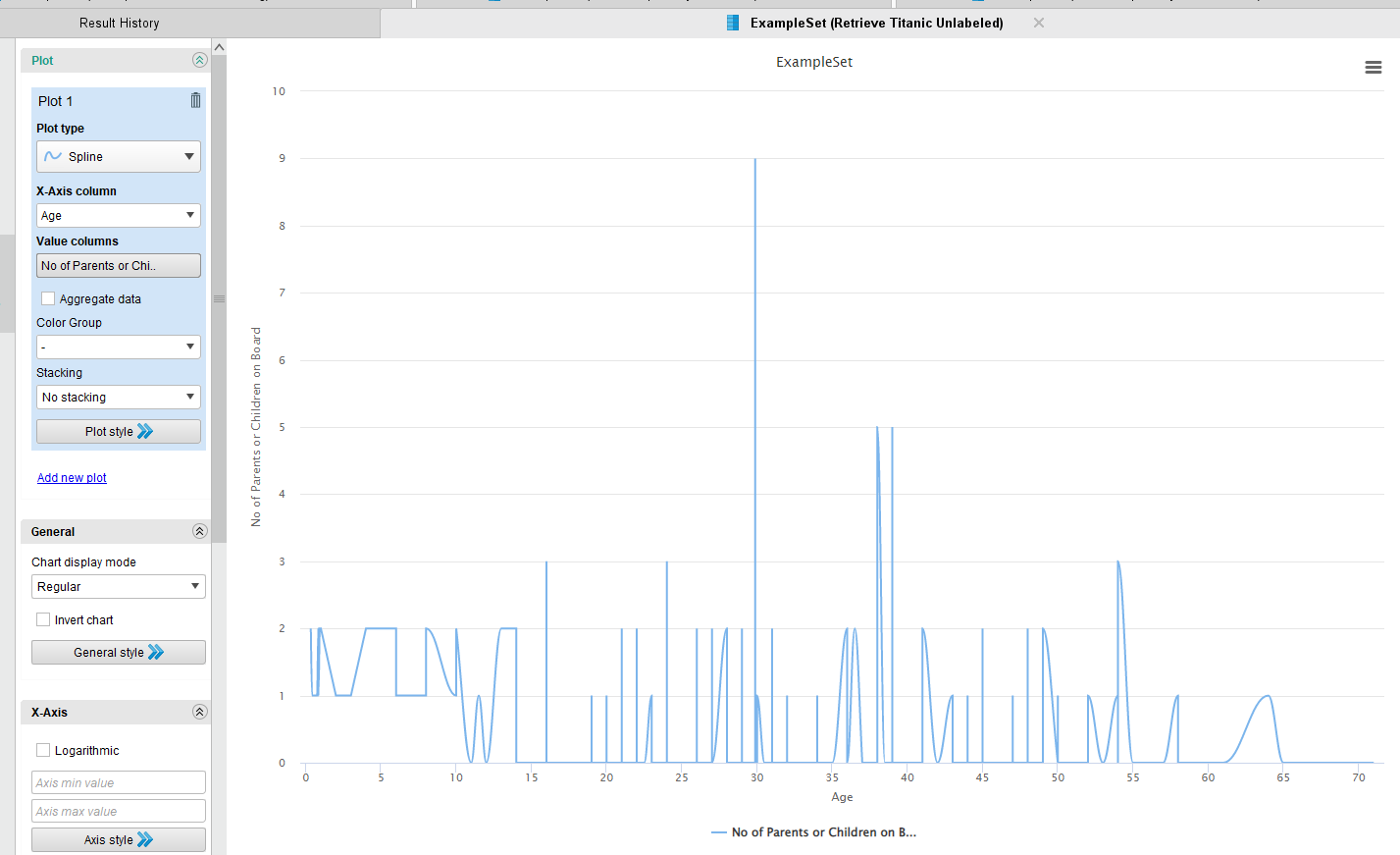


Рисунок 11 - Вкладка Vizualization набора Titanic Unlabeled

Данные, представленные выше, отличаются друг от друга количеством атрибутов.

Набор данных Titanic содержит полную информацию о пассажирах, включая их данные результаты (выжил ли пассажир или нет). Этот набор данных обычно используется для обучения моделей машинного обучения

Набор данных TitanicTraining является частью набора данных Titanic и содержит только часть данных о пассажирах, предназначенных для обучения модели. Этот набор данных обычно используется для обучения моделей машинного обучения в RapidMiner.

Набор данных TitanicUnlabeled также является частью набора данных Titanic, но не содержит информации о результатах пассажиров (выжили они или нет). Этот набор данных содержит только данные о пассажирах и обычно используется для прогнозирования результатов на новых наблюдениях после обучения модели на наборе данных TitanicTraining.

Далее был импортирован тестовый набор данных, в котором определялась вероятность попадания в спам одного из ключевых слов. Пример данных представлен на рисунке 12.

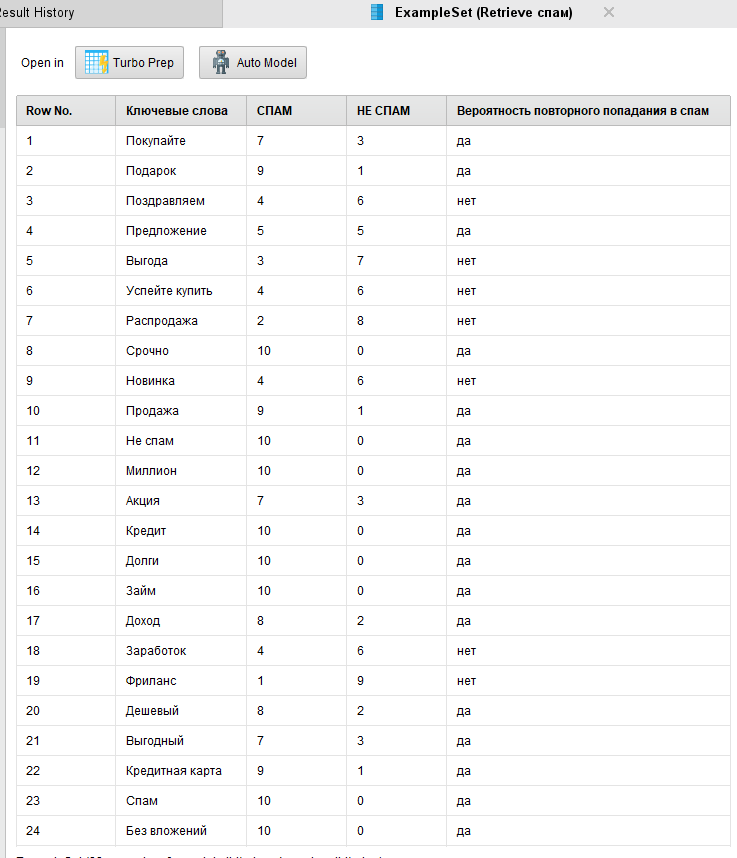


Рисунок 12 - Пример данных

Вкладка Statistics представлена на рисунке 13.

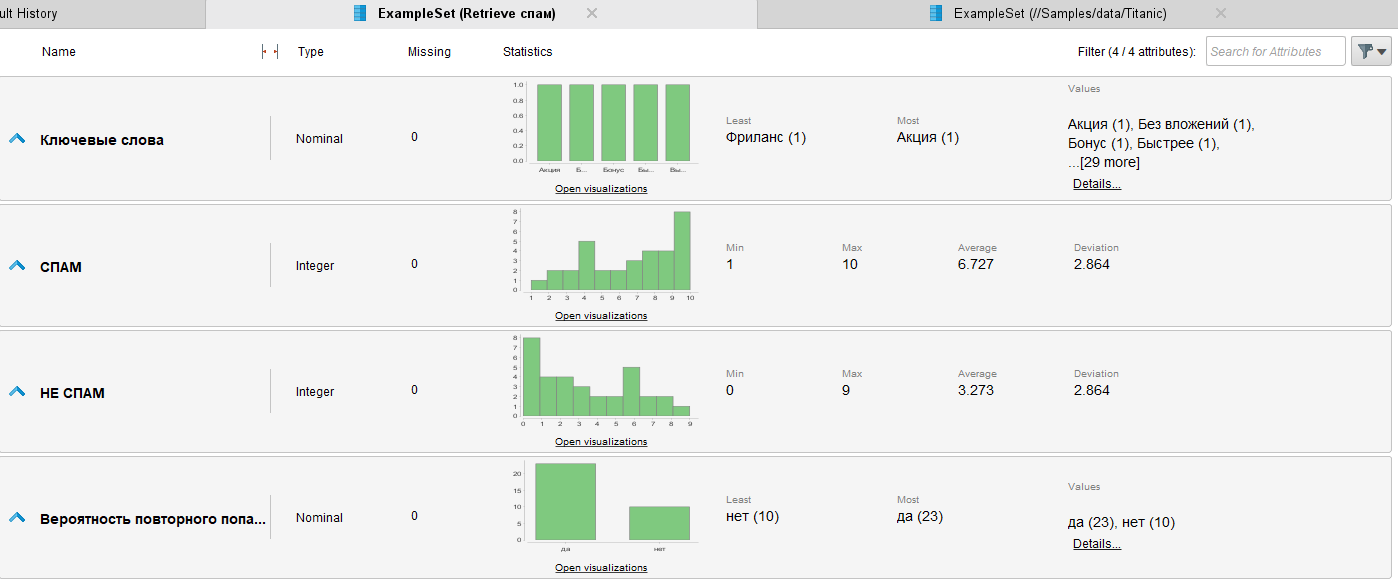


Рисунок 13 - Вкладка Statistics

Графическое представление данных представлено на рисунках 14-15, где зеленым цветом показано количество писем, которые при наличии ключевого слова не попали в спам, а синим – попали.

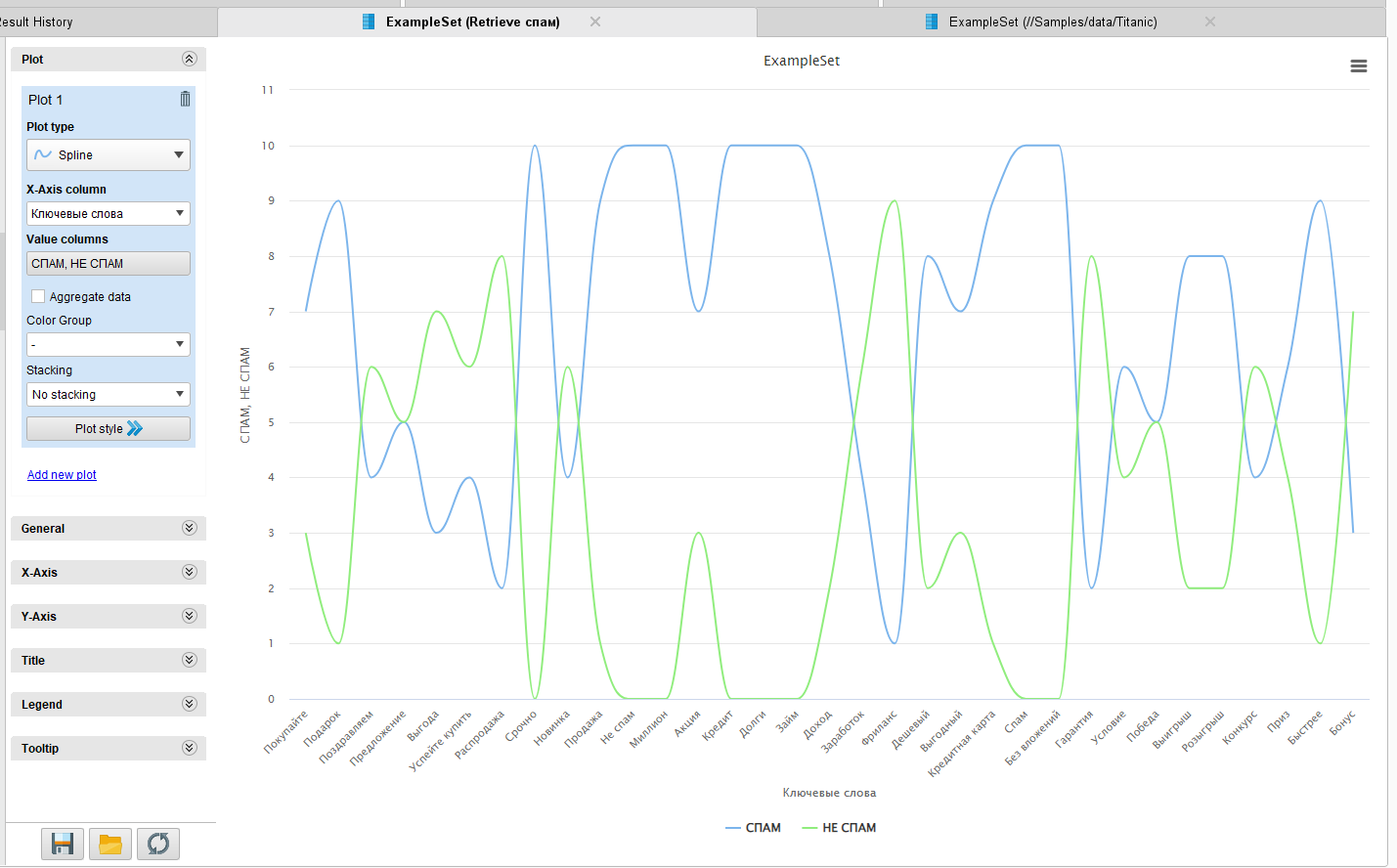


Рисунок 14 - Вкладка Vizualization график Spline

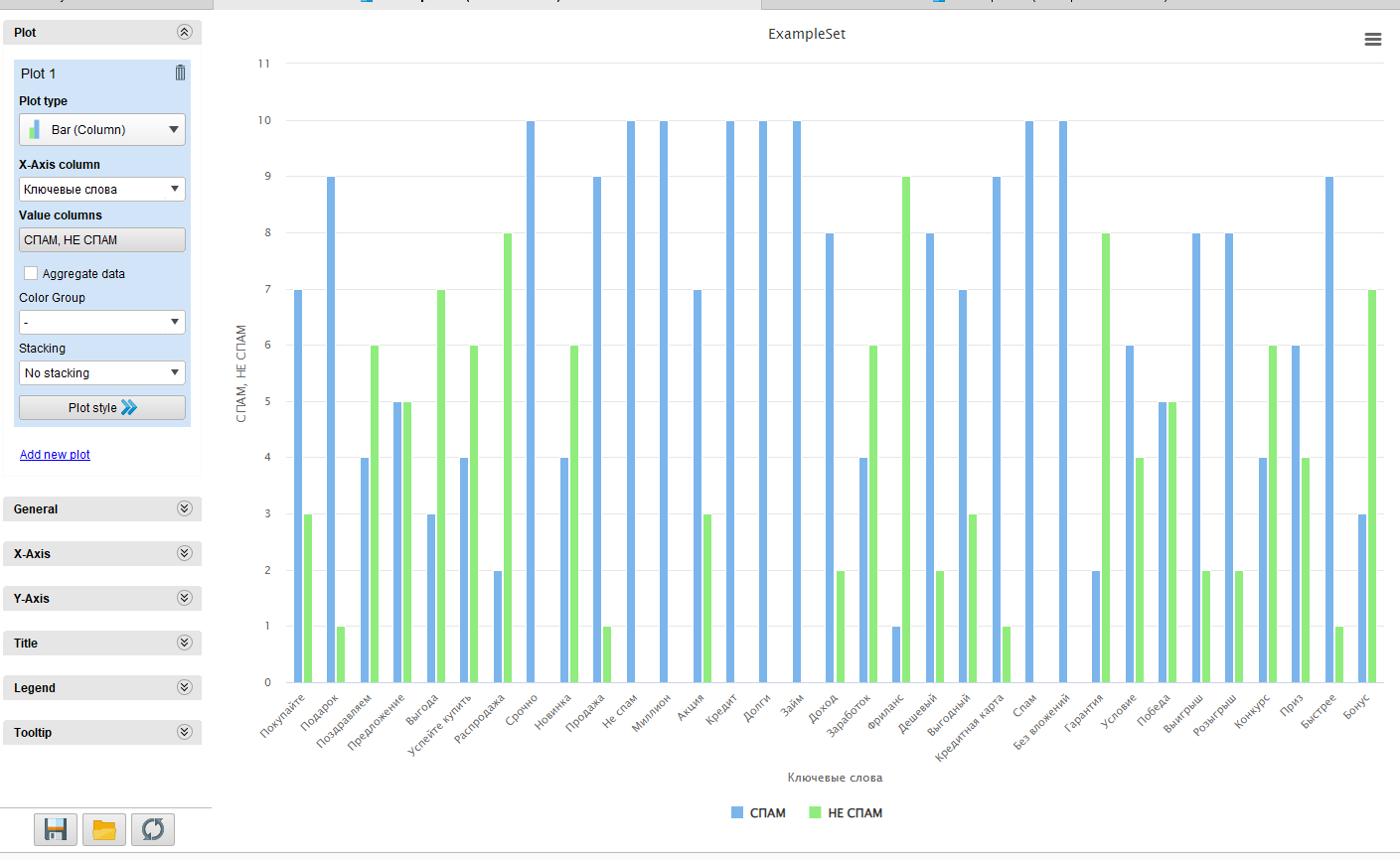


Рисунок 15 - Вкладка Vizualization гистограмма

Вывод: в ходе выполнения лабораторной работы были получены начальные навыки работы с пакетом RapidMiner Studio.

Контрольные вопросы:

1. Назначение пакета RapidMiner:

RapidMiner - это интегрированная среда для анализа данных, машинного обучения и создания моделей. Он используется для извлечения информации из данных, построения и оценки моделей машинного обучения, а также для визуализации результатов.

1. Тип лицензии и права/обязанности пользователя:

Тип лицензии для RapidMiner может различаться в зависимости от версии (Community или Commercial). Пользователи обязаны соблюдать условия лицензии, не нарушать авторские права, и, в случае коммерческой версии, оплатить необходимые лицензионные сборы.

1. Содержимое папок в разделе Repository:

В папках раздела Repository хранятся проекты, данных для анализа, преобразования данных, модели машинного обучения и другие артефакты, которые могут быть использованы в рамках проектов.

1. Технология импорта данных в репозиторий:

Для импорта данных в репозиторий пользователь может использовать различные источники данных, такие как файлы CSV, базы данных, API и другие. RapidMiner предоставляет инструменты для удобного импорта и преобразования данных.

1. Атрибуты набора данных:

Атрибуты набора данных - это характеристики, параметры или столбцы в вашем наборе данных. Они представляют собой переменные, на которых вы можете выполнять анализ, машинное обучение и другие операции.

1. Возможности режима Statistics:

Режим Statistics в RapidMiner предоставляет инструменты для проведения различных статистических анализов, включая описательную статистику, корреляцию, тесты гипотез и многие другие статистические методы.

1. Возможности режима Visualizations:

Режим Visualizations в RapidMiner предназначен для создания информативных визуализаций данных, включая графики, диаграммы, и другие графические представления для лучшего понимания данных.

1. Начало работы и шаблоны решаемых задач:

Пользователь может начать работу с RapidMiner, выбирая предложенные варианты начала работы, которые могут включать создание нового проекта, импорт данных или использование шаблонов для решения конкретных задач анализа данных.

1. Возможности главного окна:

Главное окно RapidMiner предоставляет доступ к различным инструментам и функциям программы, включая проекты, процессы анализа данных, репозиторий, визуализации и многие другие.

1. Система обучения и система помощи:

RapidMiner предоставляет систему обучения и справочную систему для пользователей, где можно найти обучающие материалы, документацию и ответы на часто задаваемые вопросы для поддержки при работе с программой.