Министерство образования Республики Беларусь

Учреждение образования «Белорусский государственный технологический университет»

Кафедра информационных систем и технологий

Лабораторная Работа № 5

**Кластерный анализ**

Выполнил:

Студент 3 курса, 1 группы ф-та ИТ

Пуйша Е.В.

Минск, 2017

# Цель работы

Знакомство с кластерным анализом, дальнейшее погружение в выявление зависимостей между параметрами виртуального производственного комплекса. Разбор возможностей программного обеспечения в области кластерного анализа.

1. **Теоретическое введение**

Кластером является объединение нескольких однородных элементов, которое может рассматриваться как самостоятельная единица, обладающая определенными свойствами.

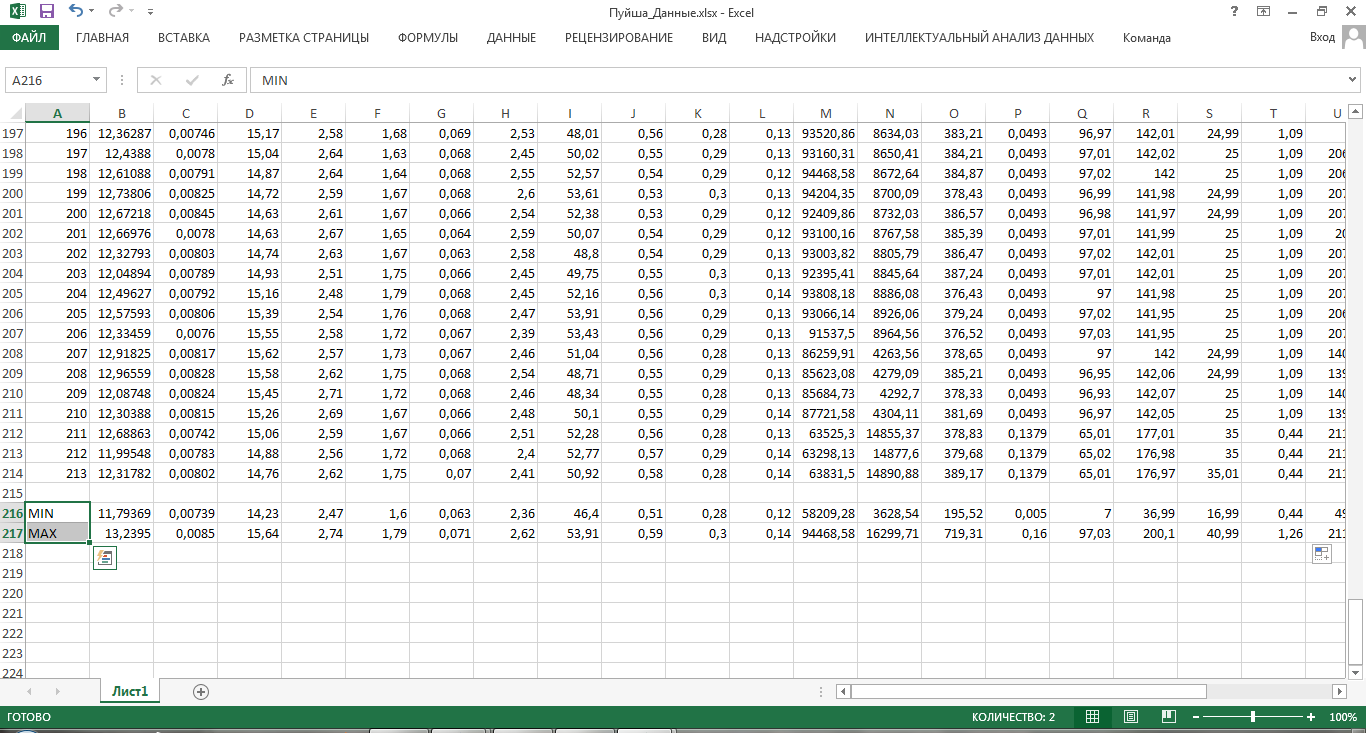
Свойства элементов, степень их сходства разнятся от задачи к задаче, но в данном случае определение подходит к условию.

Кластерный анализ позволяет проводить разбиение по набору признаков, влияние каждого из них может быть усилено или ослаблено путем внесения в формулы соответствующих коэффициентов. Кластерный анализ не накладывает ограничение на вид группируемых объектов и позволяет рассматривать множество исходных данных разной природы. Многие алгоритмы способны самостоятельно определить количество кластеров для разбиения.

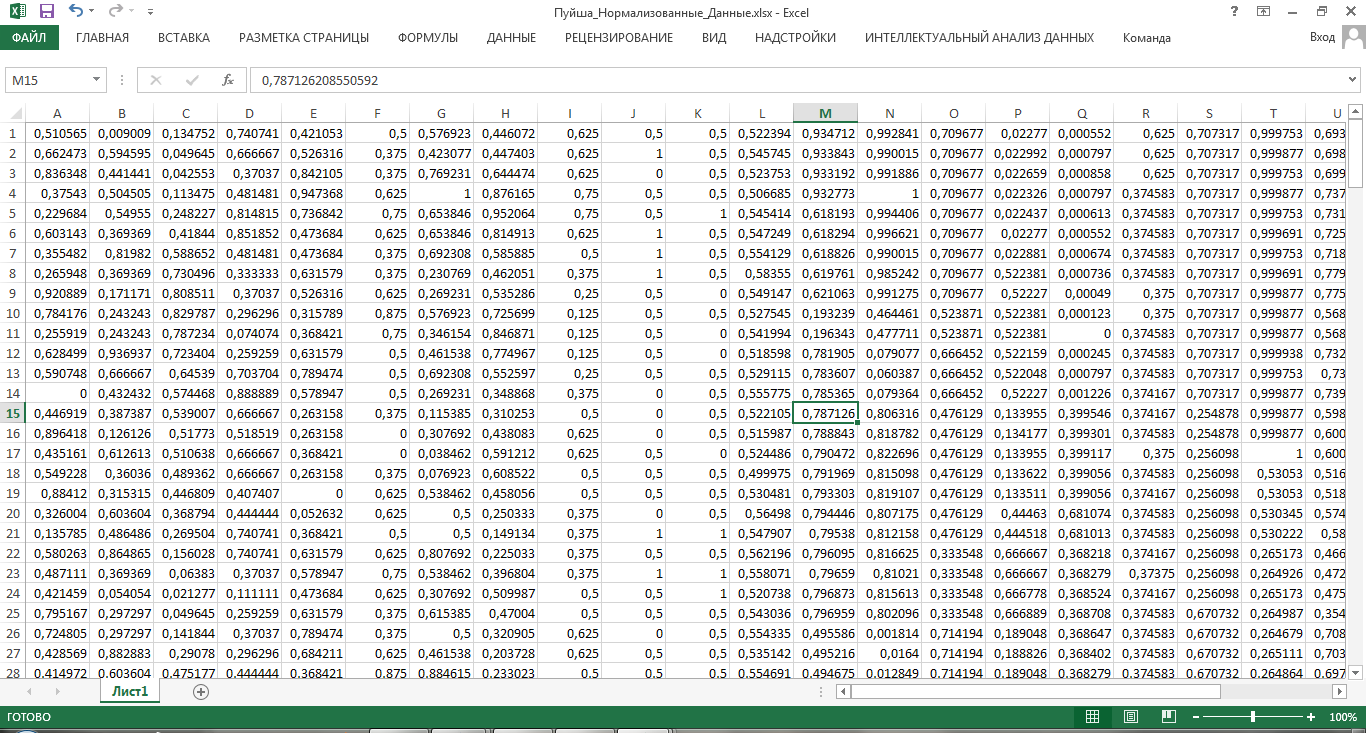
Кластеризация, как метод обучения без учителя, отличается от классификации, т.к. изначально не заданы метки для исходных объектов. Результат кластеризации зависит от метрики, выбор которой, как правило, субъективен.

# Исходные данные

В качестве исходных данных мне нужно нормализовать данные, полученные в результате выполнения лабораторной работы №1.



*Рисунок 1 – Вычисление минимальных и максимальных значений.*

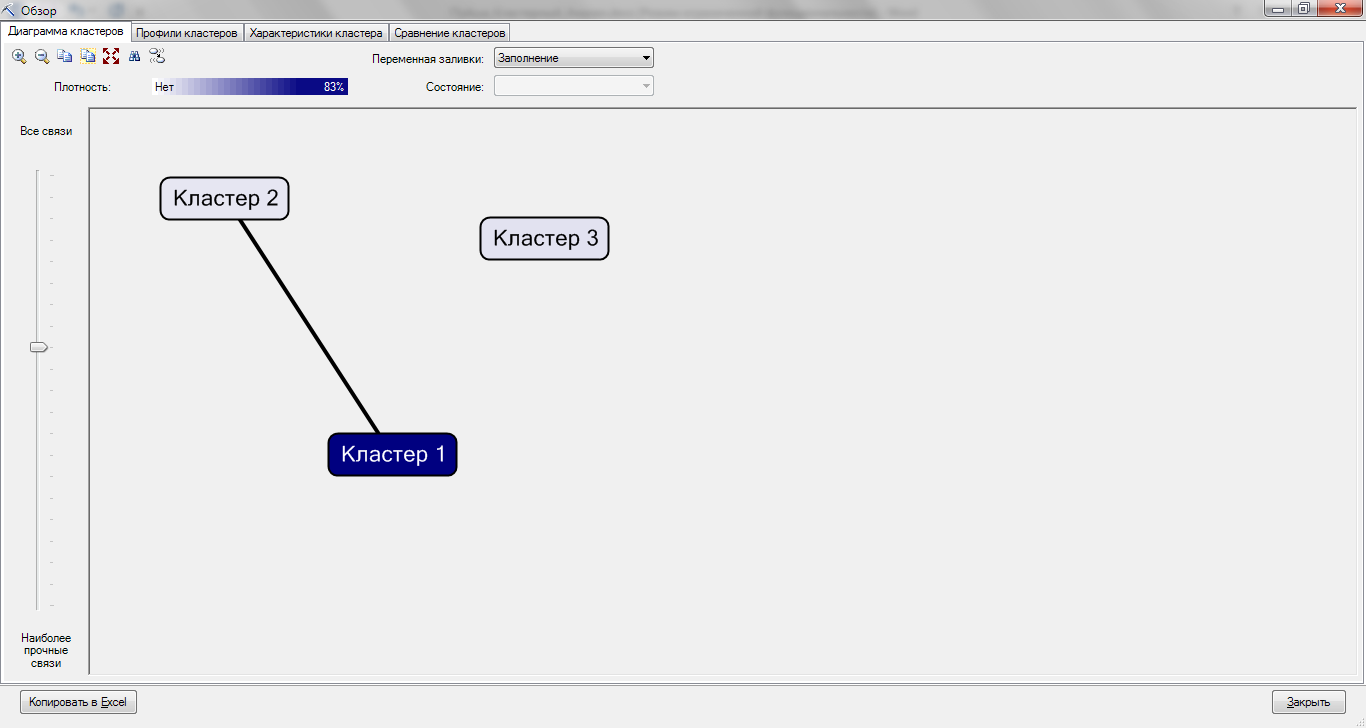


*Рисунок 2 – Нормализованные данные на новом листе Excel.*

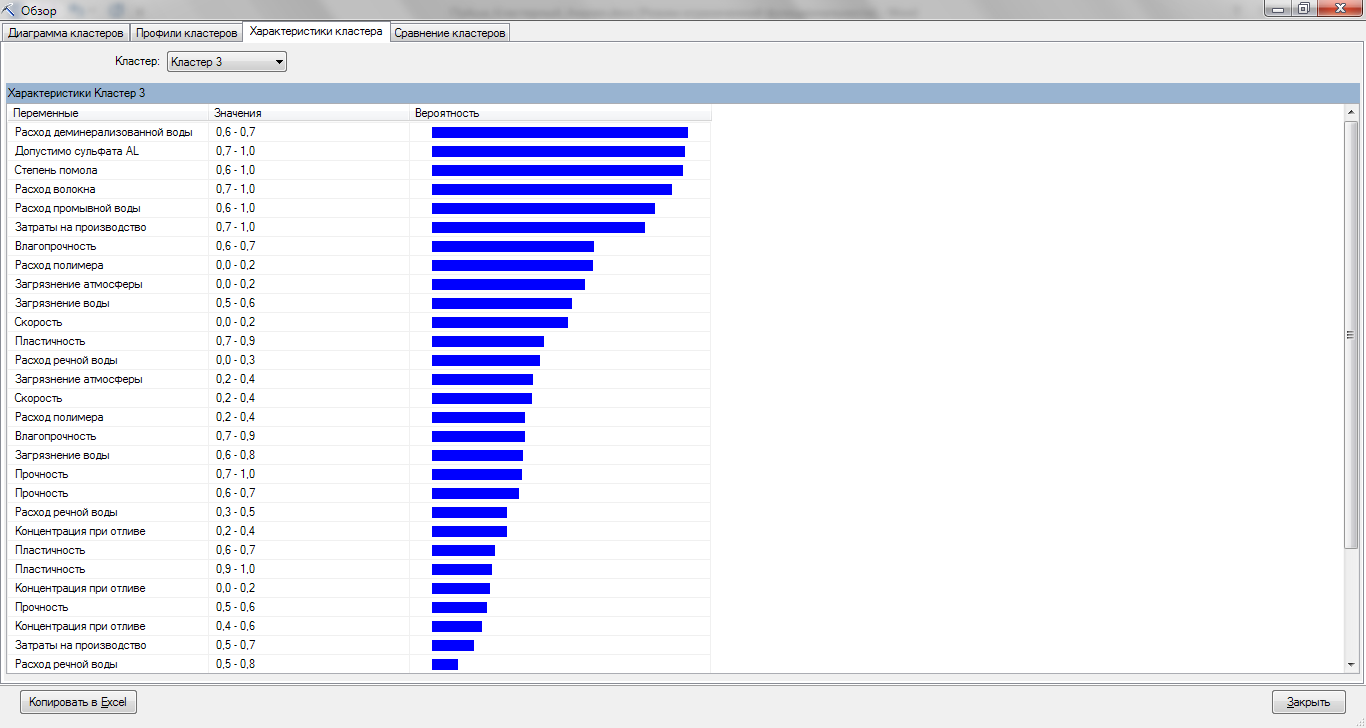
Это прекрасно, можно начинать анализ.

1. **SQL Server + Excel**

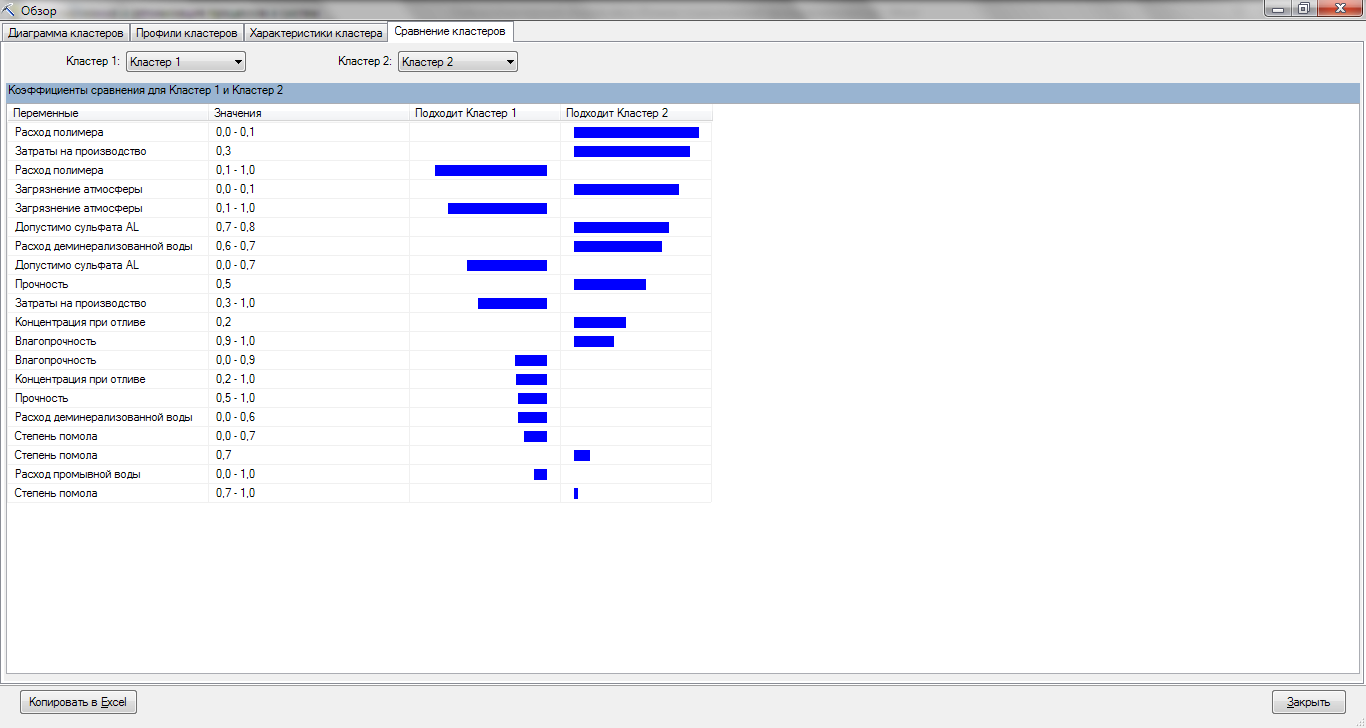
В результате исправления мелких недочетов (добавление заголовков для столбцов нормализованных данных и т.д.) и выставления рациональных настроек, я получаю такую картину:



*Рисунок 3 – Результат разбиения на три кластера.*



*Рисунок 4 – Характеристики на примере кластера №3.*



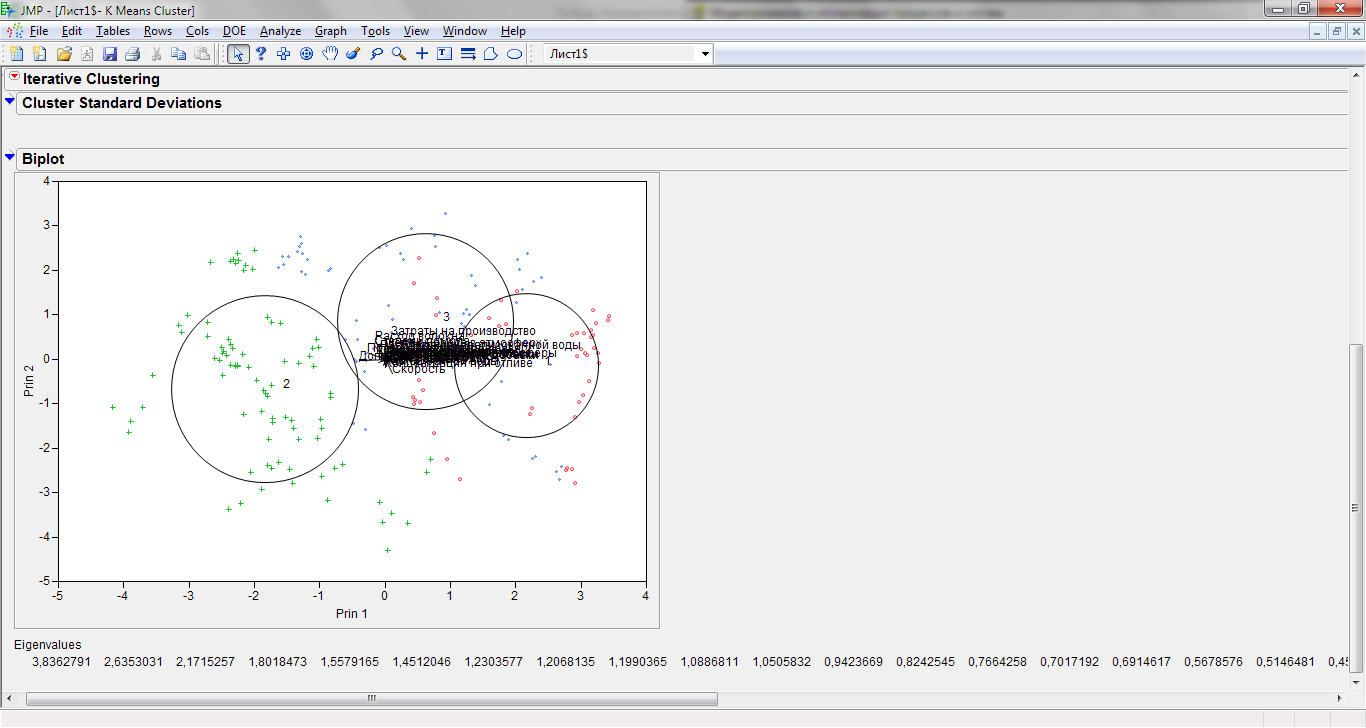
*Рисунок 5 – Сравнение кластеров.*

На данном этапе можно сказать, что некоторые параметры нельзя однозначно отнести к одному кластеру. На примере *Загрязнения атмосферы* видно, что есть параметры, которые при разных диапазонах значений одинаково хорошо подходят в несколько кластеров.

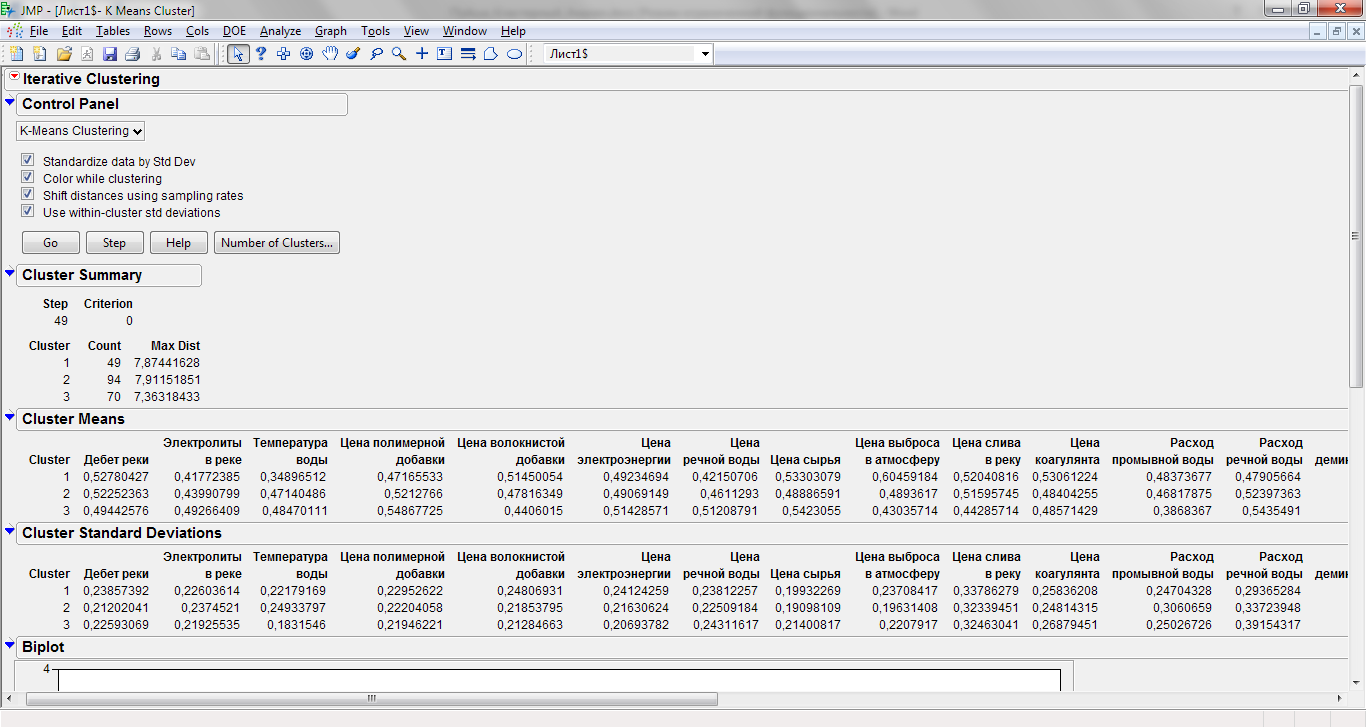
Пока что нет необходимой наглядности. Для этого я приступаю к следующему этапу.

1. **JMP**

Был проведен анализ данных в JMP, данные были разделены на три кластера.



*Рисунок 6 – Визуализация кластеров.*



*Рисунок 7 – Параметры кластеров.*

# Ну и что?

А вот так. Кластерный анализ – весьма своеобразный инструмент, который не всегда постоянен в своих ответах. По сути, человек, проводящий анализ, выступает в качестве «гадалки», главная задача которой – по символам и знамениям рассказать всю судьбу изучаемого объекта.

На основании прошлых лабораторных работ я подразделил параметры производственного комплекса на 3 вида – *финансовая сторона, экология, качество продукции.*

В качестве эксперимента при проведении анализа я делил данные на 3 кластера, чтобы посмотреть, сможет ли компьютер подтвердить мою догадку и разнести параметры так, чтобы я мог однозначно назвать кластер каким-нибудь именем. Однозначности не получилось, но намеки на это есть, из чего следует, что кластерный анализ в таком виде – достаточно полезный инструмент.

1. **Вывод**

Я на практике познакомился с кластерным анализом. Кластерный анализ попытался разбить параметры виртуального предприятия на три кластера, которые примерно одинаково зависят от внешних факторов. Главное отличие кластеризации от классификации состоит в том, что перечень групп четко не задан и определяется в процессе работы алгоритма.