Описание эксперимента.  
Есть данные результатов анкетирования по 87 IT предприятиям города Ульяновска по 90 вопросов.  
Всего в таблице данных 7830 клеток , ячеек числовых данных. Среди них заполнены 1402 ячейки (18% дот общего объёма данных) и не заполнены – 6428 (82% дот общего объёма данных).  
Исходные данные с пропусками располагаются в папке EX1/Эксперимент\_.xlsx $Лист”Таблица”и”Таблица\_”.

План эксперимента. Использовать алгоритм восстановления данных для заполнения пропусков в таблице входных числовых данных (EX1/Эксперимент\_.xlsx $Лист” РезультатВосстановления” и EX1/ вОССТАНОВЛЕНИЕ.xlsx). Использовать алгоритмы кластерного анализа для получения разбиения входных данных и дальнейшего анализа динамик ключевых характеристик объектов кластеров (EX1/ КластеризацияВсехПоВсемНа3.xlsx и EX1/Эксперимент\_.xlsx $Лист Кластеизация в лоб на три класт”).

Эксперимент можно считать заверенным. Результаты эксперимента оказались не совсем ожидаемы. Мы узнали что в целях анализа IT предприятий результаты этого эксперимента не совсем применимы. Получены новые знания о характере работы алгоритмов и о их границах применимости.

Опишем проблемы возникшие при проведении эксперимента.  
Некоторые параметры таблицы EX1/Эксперимент\_.xlsx $Лист”Таблица”и”Таблица\_” обладают всего одним или несколькими заполными значениями. Восстановление подобных пареметров, представленных малым количеством заполненных значений даёт целые блоки очень похожих, иногда с точностью да совпадения данных. При их анализе целесообразно заменить блок таких совпадающих данных на один объектов их описывающий, но это так … лирическое отступление. При использовании агломеративных методов кластеризации происходит укрупнение кластеров по принципу выбора двух самых кластеров, их удаление из списка и добавление в список кластеров в конец нового “объединённого кластра”. Если есть несколько одинаковых объектов, равноблизких к какому-то кластеру, то объединяется кластер с первым попавшимся таким объектом и новый кластер (кластер – результат группировки ) опускается в конец списка. Таким образом в результатах кластеризации наблюдается попадание одинаковых объектов в совершенно разные кластеры. ***Это не поломка алгоритма, все впорядке!!!*** \_\_\_ Это нарушение условия, накладываемого на входные данные для методов кластеризации. Условие для методов кластеризации и их входных данных заключается в запрещении обработки нескольких одинаковых объектов. (Эту заглушку для методов кластеризации можно дописывать)