*Задание (последовательность указанных действий определяется вами, обратите внимание на важность этого момента):*

* Собрать и визуализировать общую статистику базы данных
* Предложить оптимальные методы заполнения возможных пропусков в данных на основании статистического анализа; написать функцию, делающую это когда пропуски есть и ничего не делающую в обратном
* Построить корреляционную матрицу для числовых признаков
* Использовать критерий хи-квадрат на категориальных признаках
* Предложить возможные новые синтетические параметры, обосновать их интерпретируемость с позиций здравого смысла и показать их валидность графиками
* Предложить кандидаты на выброс и обосновать
* **(Для гиков)** Написать функцию, принимающую на вход обученную модель МО и БД, которая будет автоматически делать отбор признаков
* **(Для гиков)** Сжать данные при помощи PCA или любого другого алгоритма на ваш выбор, визуализировать сжатые данные и показать по каким признакам данные в этом сжатом пространстве формируют отдельные группы

**! Всё выполняется в коде, никаких Excel !**

*Предсказываемые свойства:* Km, Kcat, Vmax

*Результат:*

Рабочий код, выполняющий все эти функции и выводящий читаемые и понятные графики. Чем лаконичнее, читаемее и быстрее – тем лучше. Отдельный плюсик тем, кто напишет код, запускающийся у других людей без внесения каких-либо корректировок.

*По всем вопросам можно обращаться к Никите Серову, Юле Разливиной и Оле Капустиной.*