**Тестовая Задача. Номинация – Big Data.**

**Основное задание будет открыто на хакатоне тем командам, которые пройдут в очный этап.**

**Первичный анализ данных**

В этом заданиии мы будем использовать данные [SOCR](http://wiki.stat.ucla.edu/socr/index.php/SOCR_Data_Dinov_020108_HeightsWeights) по росту и весу 25 тысяч подростков.

*# Необходимые библиотеки*

**[1]** Считаем данные по росту и весу (*weights\_heights.csv*, приложенный в задании) в объект DataFrame:

*# Ваш код здесь*

**[2]** Постройте 2 раздельные гистограммы распределения роста и веса подростков из выборки *data*. Выделите параметры разным цветом и подпишите оси.

*# Ваш код здесь*

**[3.1]** Добавьте третий признак ('BMI'). Создайте признак *Индекс массы тела* ([BMI](https://en.wikipedia.org/wiki/Body_mass_index)). Для этого воспользуйтесь коэффициентами перевода единиц с точностью до 5 знака после запятой.

*# Ваш код здесь*

**[3.2]** Постройте картинку, на которой будут отображены попарные зависимости признаков , 'Height', 'Weight' и 'BMI' друг от друга.

*# Ваш код здесь*

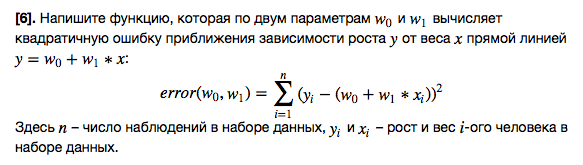
**[4]**. Создайте в DataFrame *data* новый признак *weight\_category*, который будет иметь 3 значения: 1 – если вес меньше 120 фунтов. (~ 54 кг.), 3 - если вес больше или равен 150 фунтов (~68 кг.), 2 – в остальных случаях. Постройте boxplot, демонстрирующий зависимость роста от весовой категории. Подпишите ось *y* меткой «Рост», ось *x* – меткой «Весовая категория».

*# Ваш код здесь*

**[5].** Постройте scatter plot зависимости роста от веса. Подпишите картинку.

*# Ваш код здесь*

**Минимизация квадратичной ошибки**



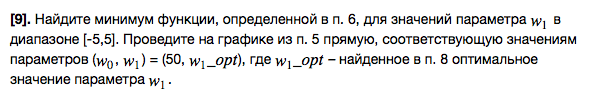
*# Ваш код здесь*

/Users/Kate/Desktop/Screen Shot 2017-04-21 at 19.02.48.png

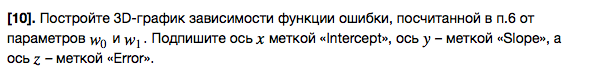
*# Ваш код здесь*

/Users/Kate/Desktop/Screen Shot 2017-04-21 at 19.02.54.png

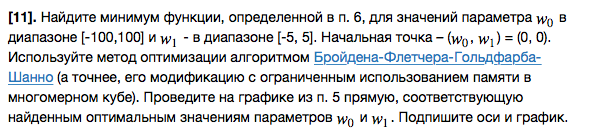
*# Ваш код здесь*



*# Ваш код здесь*



*# Ваш код здесь*



*# Ваш код здесь*