**Лабораторная работа № 1**

**Тема**: Статистическая обработка данных .

**Цель работы** *– импорт данных в формате csv. Знакомство с функциями статистической обработки данных модуля pandas.*

На последних слайдах лекции разобран пример с датасетом Титаник, можно рассматривать его как шаблон для своих вычислений.

1. Выберите датасет для работы (к примеру на сайте [www.kaggle.com](http://www.kaggle.com))
2. Примените различные функции библиотеки *pandas* для расчета описательной статистики, группировки, сделайте выводы по полученным результатам (*например, из датасета по зарплатам ИТ-специалистов можно узнать в какой стране каким специалистам платят больше и т.д.)*
3. Используя библиотеку *Seaborn* постройте гистограмму и график *box-plot* для визуализации отдельных признаков датасета.

Вопросы:

1. Какие метрики статистики вы вычисляли в этой работе?
2. Что выводит метод **.describe()**?
3. Что такое квантиль, процентиль?
4. Что показывает гистограмма?
5. Что показывает диаграмма разброса (box plot)?