# **Лабораторная работа №6**

Работа посвящена сбору данных в интернете. Для выполнения лабораторной работы необходимо использовать библиотеку **lxml**, для визуализации статистики по данным используется **pandas** и/или **matplotlib**.

В качестве исходных данных можно использовать следующие ресурсы:

* <https://coronavirus-monitor.ru/>
* <https://www.val.ru/valhistory.asp>
* <http://www.statdata.ru/largest_cities_russia>
* <https://visasam.ru/russia/rabotavrf/zarplaty-v-rossii.html>
* <https://riarating.ru/infografika/20201006/630183270.html> (или выбрать более свежую статистику <https://riarating.ru>)
* <https://person-agency.ru/salary.html>
* <https://hh.ru/>
* <https://avito.ru/>
* и другие

Сбор информации на примере авиакомпании Победа можно посмотреть в Colab блокноте: <https://colab.research.google.com/drive/1a9swtorB7HGkk6hyM1ATo_3VcAbik6mh#scrollTo=hWYRGxtlYg62>

Инструкция по использованию Xpath с библиотекой lxml: <https://dev-gang.ru/article/kak-ispolzovat-python-i-xpath-dlja-czistki-saitov-ht13ju9mu8/>

**Постановка задачи:**

1. Выберите интересующую Вас предметную область и проведите сбор данных, распарсив необходимые сайты. При сборе данных необходимо найти нужную информацию на странице выбранного сайта, произвести необходимую очистку данных, сохранить набор данных в файле в формате “.csv”.
2. По полученным данным необходимо провести визуальный анализ, вывести актуальную статистическую информацию с помощью графиков pandas и/или matplotlib (не менее трёх различных информативных графиков).
3. Сделать выводы.