

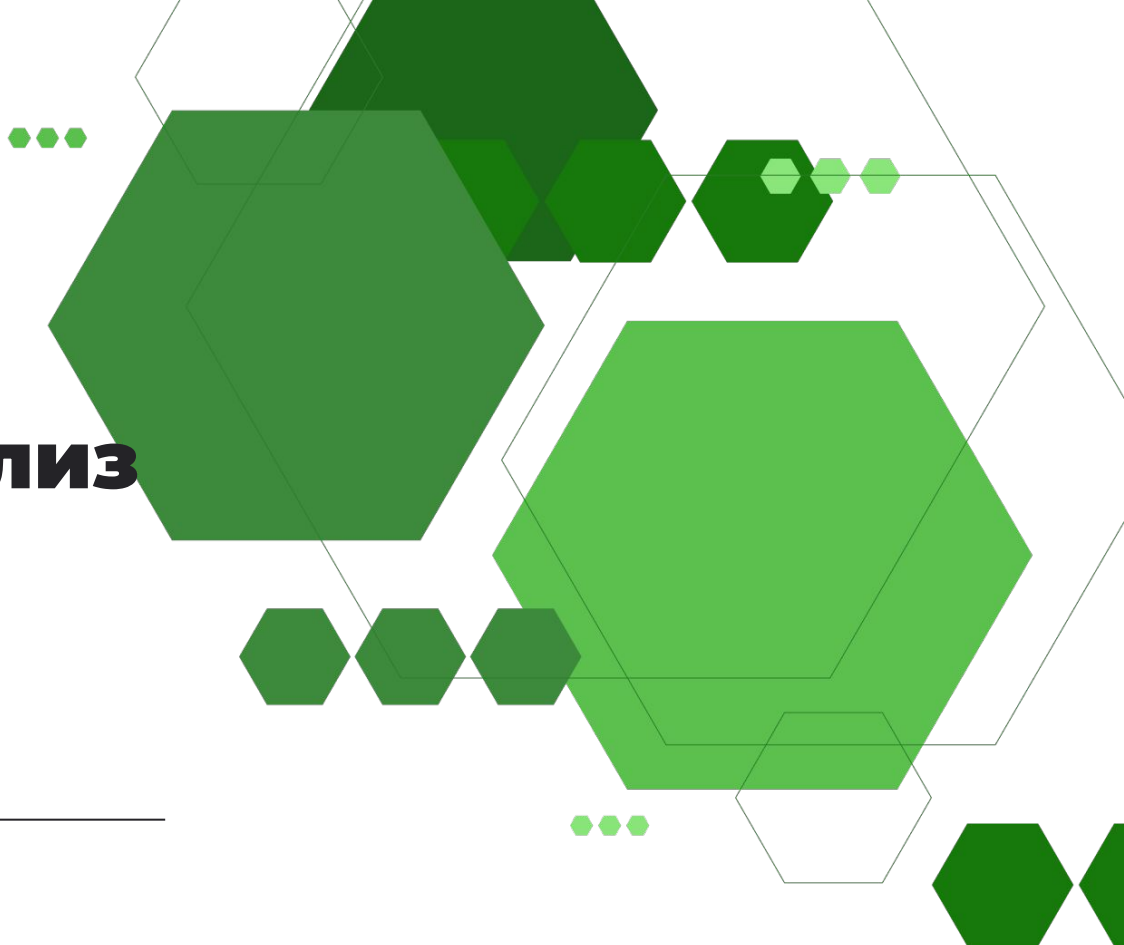
SKILLFACTORY

Вводный вебинар

ПРОЕКТ-1 Анализ вакансий из HeadHunter

Жарова Мария

эксперт по Python, Math, ментор проекта



План вебинара:

- Цели и суть проекта
- Чему нас научит этот проект?
- Знакомство с проектом
- Рекомендации по выполнению
- Подведение итогов вебинара



Обязательно задавайте
вопросы голосом или в чате в
любое время :)

Цели и суть проекта

Цель: отработать все изученные ранее приёмы работы с данными и их анализа.



Почувствуйте себя
аналитиком
HeadHunter!



Цели и суть проекта



Работа с большими данными включает в себя несколько этапов



Самый первый - обработка сырых данных, их исследование и подготовка к дальнейшей работе.



Наш проект!





Чему нас научит этот проект?

Темы:

- анализ структуры таблиц
- преобразование данных с помощью встроенных и рукописных функций
- разведывательный анализ данных с помощью визуализации
- очистка данных

Знакомство с проектом

★ Необходимо провести:

- анализ структуры  часть 1
- сделать предобработку данных  часть 2
- сделать визуальный анализ  часть 3
- очистить данные  часть 4

Часть 1 “Исследование структуры данных”

- ★ Перед началом работы необходимо понять, как устроены признаки в данных и каких они типов
- ★ Что будем делать:
 - вывод размерности таблицы
 - изучение информации в столбцах и строках
 - вывод информации о типизации столбцов
 - вывод основных статистических параметров таблиц

Часть 2 “Преобразование данных”

★ Ознакомившись с данными, сделаем их предобработку:

*некоторые данные представлены в очень неудобном “сыром”
виде - преобразуем их, чтобы можно было работать с ними на
полную*




Т.е. готовим данные к дальнейшему анализу и визуализации!

Часть 2 “Преобразование данных”

★ Что будем делать:

- преобразование исходных столбцов к новым форматам
- переименование столбцов
- выделение ключевой информации из признаков и создание новых
- преобразование типов данных, выделение дат

 **Рекомендация: оформить функции-преобразователи, которые будут принимать в качестве аргумента элемент столбца, а возвращают его преобразованную версию. Далее используем `apply()`**

Часть 3 “Исследование зависимостей в данных (EDA)”

- ★ Наконец, всё готово для проведения непосредственно первичного анализа данных!
- ★ Что будем делать:
 - выявлять связи между признаками
 - находить зависимости
 - определять распределения признаков
 - искать аномалии, выбросы и дефекты в данных
- ★ И визуализируем полученные результаты на графиках!

Часть 3 “Исследование зависимостей в данных (EDA)”



Важно!

Этот этап - логическое продолжение предыдущих, в нём мы работаем с признаками, созданными в предыдущем разделе



Рекомендация: проще всего делать визуализацию при помощи библиотеки plotly (но если хотите, можно использовать и pandas, matplotlib, seaborn:))

Часть 4 “Очистка данных”



★ После визуального анализа обнаружатся несколько несостыковок в данных:

- пропуски
- гигантские размеры желаемых зарплат
- резюме людей слишком “преклонного” возраста
- опыт работы, превышающий возраст

Всё это говорит о том, что данные подлежат очистке!

Ещё рекомендации по выполнению

- ★ Все импорты библиотек в 1-й ячейке ноутбука
- ★ Хотя бы минимальное комментирование кода
- ★ Текстовые выводы должны быть оформлены в отдельных ячейках типа Markdown
- ★ Следите, чтобы не сбивалась нумерация заданий при загрузке их на github!
- ★ Для большей наглядности можете выделить пожирнее номера заданий, а выводы оформить курсивом :)
- ★ На github обязательно добавляйте файл README.md (именно в формате .md!) с основной информацией о проекте

Ещё более полный список написан в задании на платформе курса

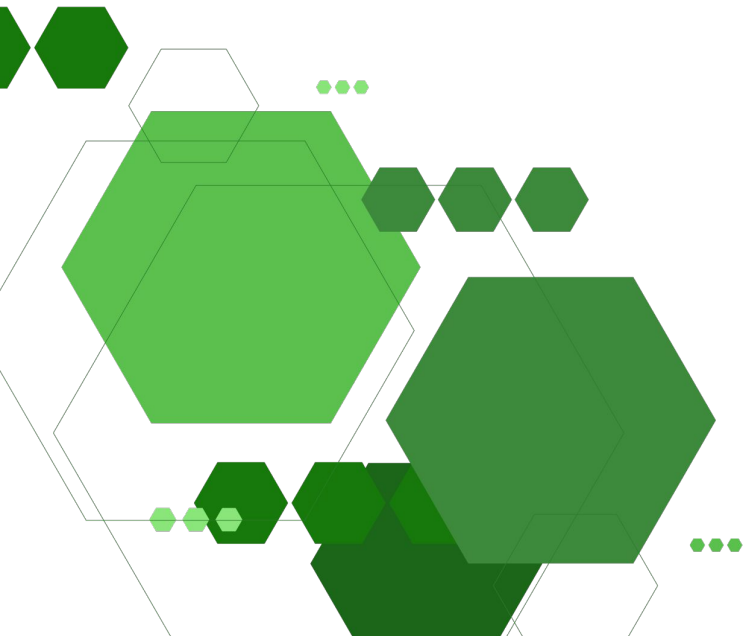
Подведение итогов:


Теперь осталось только собрать все знания и рекомендации и реализовать их на практике средствами Python 😊

По любым вопросам обращайтесь в канал #01-project-1

Чтобы получить презентацию, напишите мне в лс в Slack @Мария Жарова

УДАЧИ!!!





После вебинара заполните, пожалуйста,
форму обратной связи (будет в чате потока)



**Всем спасибо за
внимание!**