

Курс «Программирование на Python»

Проект

Цели проекта: разработать систему для работы с данными об отделах и сотрудниках компании, а также о кандидатах на вакантные должности. Система должна обеспечить реализацию базовых HR-функций: учет данных о сотрудниках, анализ эффективности и результативности их работы, формирование отчетов на основе имеющихся данных, первоначальный отбор кандидатов по заданным критериям. В проекте необходимо продемонстрировать умение работы со следующими структурами данных: строки, списки, кортежи, словари, множества, массивы numpy, датафреймы pandas. В ходе работы потребуется использовать условия и циклы, создавать функции, читать данные из файла и записывать их в файл.

Модуль 1. Хранение HR-данных и формирование отчетов на их основе (2 балла)

Вам необходимо хранить о сотрудниках следующую информацию:

- Id_сотрудника
- ФИО
- Должность
- Пол
- Дата приема на работу
- Список навыков
- Заработная плата
- Размер ежемесячного бонуса
- Количество подчиненных

Определите многоуровневую структуру, в которой они будут храниться. Это могут быть списки списков, списки кортежей, списки словарей, вложенные словари или более сложные структуры. В этом пункте не используйте датафреймы pandas или массивы numpy. Обоснуйте выбор формата хранения данных, в чем его преимущества и ограничения? Сохраните данные о 10 сотрудниках в выбранном формате (данные можно сгенерировать или придумать).

Напишите функции, которые обрабатывают эти данные (не менее 5 функций). Нужно придумать собственные задачи, реализуемые функциями.

Примеры задач:

- формируется персонализированное письменное уведомление сотруднику в определенном формате на основе указанного id_сотрудника;
- рассчитываются данные о сотрудниках для статистической отчетности, например доля женщин среди сотрудников, имеющих не менее 5 подчиненных.

Все функции должны быть снабжены тестовым описанием (docstring) и аннотациями типов.

Модуль 2. Чтение HR-данных из файла, их обработка и запись в файл (2 балла)

Данные о сотрудниках хранятся в txt файле в следующем формате.

Идентификатор: 463

ФИО: Иванов Иван Иванович

Должность: программист

Оклад: 150000

Дата приема: 2023-05-15

Идентификатор: 677

ФИО: Петрова Мария Сергеевна

Должность: аналитик

Оклад: 165000

Дата приема: 2021-11-20

Создайте такой файл, содержащий информацию о не менее 10 сотрудниках. Данные можно сгенерировать или придумать самостоятельно.

Необходимо прочитать эти данные из файла и сохранить в некую структуру, с которой можно дальше работать для извлечения из нее необходимой информации. Обоснуйте выбор структуры.

Сформируйте отчет, содержащий следующие данные:

- число сотрудников,
- средний оклад,
- доля сотрудников, принятых на работу за последние 3 года.

Добавьте данные отчета в конец файла с данными о сотрудниках.

Модуль 3. Расчёт показателей эффективности работы в numpy (2 балла)

Данные о количественных показателях результативности деятельности сотрудников хранятся в многомерном массиве numpy. Это могут быть данные о производственных результатах, финансовых показателях или эффективности рабочего времени. Придумайте или сгенерируйте данные о не менее, чем 10 сотрудниках. Массив должен быть хотя бы двумерным. Рассчитайте не менее 5-ти разных метрик, характеризующих эффективность работы сотрудников, используя функции по работе с массивами numpy. Можно взять готовые метрики из различных источников или придумать свои.

Модуль 4. Анализ и визуализация данных об отделах в pandas (2 балла)

В файле CSV в структурированном виде хранятся данные об отделах организации: id, название, число сотрудников, годовой бюджет, число реализуемых проектов, и пр. Добавьте еще не менее 5 признаков разного типа (метрические и категориальные). Создайте файл с данными о не менее 10 отделах. Данные можно сгенерировать или придумать. Нужно считать данные из файла в pandas DataFrame и провести анализ данных, используя инструменты библиотеки pandas. Построить сводные таблицы и не менее 5 разных графиков. Описать тенденции и взаимосвязи, наблюдаемые в таблицах и графиках.

Модуль 5. Авторский (2 балла)

Придумайте собственный модуль, опишите его и напишите код на Python для его реализации. Он должен решать одну из базовых HR-функций, перечисленных в целях проекта.

Критерии выполнения проекта:

- Все модули проекта должны сопровождаться *выполненным* фрагментом кода.
- Проект можно выполнять индивидуально или в группе из 2 человек, критерии оценивания при этом не меняются.
- В названии загружаемого в Smart LMS файла нужно указать фамилию(ии) автора(ов) работы.
- До **13:00 27 октября** любому члену группы нужно загрузить в Smart LMS следующие файлы:
 - Jupyter Notebook, содержащий выполненный код, на основе которого выполнялся проект.
 - Файлы с данными.