

Задание на стажировку по анализу данных

Вы проходите стажировку в университете S. S занимает значимое место в экономике региона (г. Санкт-Петербург) и готовит специалистов для строительной отрасли.

Для того, чтобы правильно выстроить стратегию работы с абитуриентами и стать настоящим полюсом притяжения талантов, университет просит Вас проанализировать данные приемной кампании 2022 года.

У Вас есть файл в формате xls, который содержит следующие столбцы:

- Уникальный код – id абитуриента
- Дата подачи заявления
- Сумма баллов – за ЕГЭ и собственные вступительные испытания (абитуриенты, поступающие на архитектурный факультет сдают 4 экзамена)
- Сумма баллов за инд.дост.(конкурсные) – дополнительные баллы для поступления
- Пол
- Дата рождения
- Гражданство
- Основание поступления – бюджет, контракт или целевое обучение (т.е. обучение за счет средств будущего работодателя)
- Категория приема – имеют ли льготы при поступлении
- Форма обучения – очная, очно-заочная (вечерняя), заочная
- Уровень подготовки – бакалавриат или специалитет
- Направление\специальность - из направлений, реализуемых в университете
- Факультет
- Регион по прописке
- Вид образования – что окончил абитуриент - среднее (школа) или среднее специальное (колледж)
- Год завершения – год завершения образования предыдущей ступени
- Состояние – зачислен ли абитуриент
- Доставка документов - способ подачи документов, где веб – личный кабинет на сайте университета
- Дата согласия на зачисление – дата, когда абитуриент дал согласие на зачисление, может быть до публикации конкурсных списков (публикуются 27 июля) или в обозначенные сроки после публикации (с 27 июля по 3 августа). Если абитуриент подал согласие до публикации конкурсных списков, значит вуз для него в приоритете, если после, то рассматривал и другие варианты поступления.

В тетрадке Jupiter notebook / Collab с помощью библиотек python проведите анализ имеющихся данных только для абитуриентов, которые подавали заявление на очную форму обучения.

!Не забудьте про предобработку данных! (переименование столбцов, поиск дубликатов, корректность типов данных, вывод общей информации по таблице и т.д.)

Отчет должен содержать код, визуализацию и выводы.

Оформление текста - с помощью языка markdown

Содержание отчета:

1. Общая информация о приемной кампании:

- Дата начала и окончания приема документов
 - Общее количество заявлений, количество заявлений по факультетам и направлениям подготовки
 - Количество уникальных абитуриентов
 - Количество заявлений на одного абитуриента, среднее, медиана (используйте результаты метода describe())
 - Распределение по способам подачи документов
 - Распределение по основаниям поступления, уровням подготовки, образованию
 - Распределение по регионам, гражданству
 - Количество согласий по направлениям подготовки до момента публикации конкурсных списков
 - Промежуточный вывод
- Обратите внимание, что часть данных должна приводиться по уникальным абитуриентам!

2. Портрет абитуриента

Выберите 2 категории из предложенных и сравните их портреты:

- Зачисленный на бюджет абитуриент
- Зачисленный на контракт абитуриент
- Ушедший абитуриент (обратите внимание, что нужно оставить только тех, кто не был зачислен ни по одному направлению подготовки)

Состав портрета: пол, возраст, где учился, где живет (Санкт-Петербург или регионы, может какой-то регион в среднем более активен), какие направления подготовки выбирал (т.е. самые популярные направления подготовки), сколько подал заявлений, каким способом, какие у него баллы вступительных испытаний. Что еще важное можно сказать о нем?

3. Общий вывод