

Задание

Дан файл HR.csv с данными по опросу уровня удовлетворенности сотрудниками работой.

Файл доступен тут -

<https://drive.google.com/file/d/1INgo03nal-vwFJe7Lec5vOUtOwfJdUr1/view?usp=sharing>

Признаки:

1. **satisfaction_level** - Уровень удовлетворенности работой
2. **Last_evaluation** - Время с момента последней оценки в годах
3. **number_projects** - Количество проектов, выполненных за время работы
4. **average_monthly_hours** - Среднее количество часов на рабочем месте в месяц
5. **time_spend_company** - Стаж работы в компании в годах
6. **work_accident** - Происходили ли несчастные случаи на рабочем месте с сотрудником
7. **left** - уволился ли сотрудник
8. **promotion_last_5years** - повышался ли сотрудник за последние пять лет
9. **department** - отдел в котором работает сотрудник
10. **salary** - относительный уровень зарплаты

Требуется выполнить следующее задание:

#	Задание	Баллы
1	Загрузите файл HR.csv в pandas dataframe	5
2	Рассчитайте основные статистики для переменных (среднее, медиана, мода, мин/макс, сред. отклонение).	10
3	Рассчитайте и визуализировать корреляционную матрицу для количественных переменных. Определите две самые скоррелированные и две наименее скоррелированные переменные.	10
4	Рассчитайте сколько сотрудников работает в каждом департаменте.	5
5	Показать распределение сотрудников по зарплатам.	5
6	Показать распределение сотрудников по зарплатам в каждом департаменте по отдельности	5
7	Проверить гипотезу, что сотрудники с высоким окладом проводят на работе больше времени, чем сотрудники с низким окладом	10
8	Рассчитать следующие показатели среди уволившихся и не уволившихся сотрудников (по отдельности):	10

	<ul style="list-style-type: none"> • Доля сотрудников с повышением за последние 5 лет • Средняя степень удовлетворенности • Среднее количество проектов 	
9	Разделить данные на тестовую и обучающую выборки Построить модель LDA, предсказывающую уволился ли сотрудник на основе имеющихся факторов (кроме department и salary) Оценить качество модели на тестовой выборки	20
10	Загрузить jupyter notebook с решение на github и прислать ссылку	5

Итого - максимум 85 баллов

Для зачета необходимо набрать минимум 55