**Статистика 1**

Допустим, нам нужно проверить, какой из двух шаблонов рекламного письма лучше. Мы разослали каждый двум непересекающимся случайным выборкам из нашей БД клиентов. Размер каждой выборки составил 250 пользователей. В случае одной мы получили совершение целевого действия от 58 пользователей, в случае второй от 71.

Необходимо рассчитать, влияет ли шаблон на действия пользователей. Рассчитать двумя методами.

**Статистика 2**

В кармане у Флоренции 5 монет, причем 4 монеты правильные, а у пятой вероятность выпадения “Орла” при случайном подбрасывании равна 0.9. Флоренция случайным образом достает из кармана монету и подбрасывает ее.

1. Какова вероятность того, что выпадет “Орел“ ?
2. Какова вероятность того, что монета правильная, если известно, что выпал “Орел”

**BASH**

Написать bash-скрипт, который делает следующее:

1) Принимает на вход адрес директории, переданной параметром

1) Находит в ней файлы с разрешением .txt, сортируя их по размеру в порядке убывания (гарантированно, что файлов одинакового размера не будет)

2) Выделяет каждый третий файл начиная с самого большого (т.е. взять первый, четвертый, седьмой и т.д. файл по размеру)

**Python-1**

Проиллюстрировать работу следующих функций библиотеки pandas на каком-либо публичном датасете:

 1) функций df.groupby()

2) pd.pivot\_table()

3) добавление и удаление мультииндексов

4) Использование matplotlib (можно функции, встроенные в pandas, а можно использовать и саму библиотеку. Использование их комбинации, т.е. создание графика с помощью встроенных функций pandas, но на основе контекста, сформированного matplotlib, будет плюсом)

**ML1**

Описать основные метрики, которыми можно описывать качество моделей регрессии и классификации, к каждой описать их достоинства и недостатки (в каких случаях их не стоит применять), 1-2 пункта на то и на другое.

**ML2**

Взять небольшой датасет (например, Iris) и решить задачу классификации с использованием любого из методов (нейросети, деревья решений, SVM), написать короткий отчет о решении с визуализацией. Привести в отчете метрики качества модели. Ход решения подробно расписывать не надо, но код приложить к отчету необходимо.

**ML3**

Взять какой-нибудь небольшой датасет  (кошки-собаки, например) , построить и обучить нейросеть MobileNetV2 применив метод трансфер-лернинга на классификацию изображений. Реализовать в Keras или в Pytorch.

**SQL1**

Есть таблица с колонками a и b, обе колонки типа INT. Дан запрос "select a, count(\*) from t group by a". Как изменить этот запрос, чтобы вывелись уникальные значения “a” которые встречаются в таблице более 2х раз

**SQL2**

Существует таблица, в которой хранятся записи о неких событиях (например, выставки или фестивали).

CREATE TABLE events (

   id INTEGER PRIMARY KEY NOT NULL,

   name CHARACTER VARYING(255),

   begin\_date TIMESTAMP(0) WITHOUT TIME ZONE,

   end\_date TIMESTAMP(0) WITHOUT TIME ZONE

);

Необходимо написать код, который выводил бы на экран события, которые проходят на этой неделе.