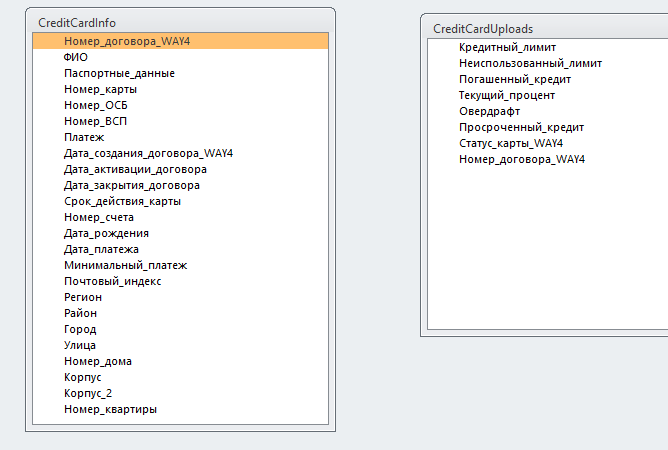
# SQL

****

1. Найдите наибольший платеж по кредитной карте в разрезе ВСП. Формат вывода: ВСП, максимальный платёж.

**Ответ:**

SELECT

Номер\_ВСП,

MAX(Платеж) AS Максимальный\_платеж

FROM

CreditCardInfo

GROUP BY

Номер\_ВСП

1. Вывести следующую информацию по всем кредитным картам: ФИО, номер карты, кредитный лимит для тех записей, где будет совпадать номер договора WAY4.

**Ответ:**

SELECT

ФИО,

Номер\_карты,

Кредитный\_лимит

FROM

CreditCardInfo

JOIN CreditCardUploads ON CreditCardInfo.Номер\_договора\_WAY4 = CreditCardInfo.Номер\_договора\_WAY4

1. Вывести все кредитные карты того ВСП, в котором обслуживается клиент с паспортными данными «12 34 567890».

**Ответ:**

SELECT

\*

*-- Если правильно поняла, что “вывести все кредитные карты” - это вывод всей   
-- информации по картам, а не только её номера*

FROM

CreditCardInfo

WHERE

Номер\_ВСП = (

SELECT

Номер\_ВСП

FROM

CreditCardInfo

WHERE

Паспортные\_данные = '12 34 567890'

)

1. Необходимо разработать код хранимой процедуры\функции, которая при передаче в нее текущей отчетной даты будет возвращать в табличном режиме значения текущей даты и дат предыдущих 3 недель.

call test ('2013-03-06')--вызов отчетной даты

**Таблица результат**

| 06.03.2013 |
| --- |
| 27.02.2013 |
| 20.02.2013 |
| 13.02.2013 |

**Ответ:**

Для MS SQL:

CREATE PROCEDURE test @value nvarchar(30)

AS

BEGIN

DECLARE @Date datetime

SET @Date=@value

SELECT \* from (values

(CONVERT(varchar,@Date,104)),

(CONVERT(varchar,@Date - 7,104)),

(CONVERT(varchar,@Date - 14,104)),

(CONVERT(varchar,@Date - 21,104))) as t(Дата)

END;

Вызов через:

EXEC test '2020-05-07'

Для MySQL:

DELIMITER //

CREATE PROCEDURE test(

IN date VARCHAR(50))

BEGIN

select DATE\_FORMAT(STR\_TO\_DATE(date, "%Y-%c-%e"), '%d.%m.%Y') Дата

union SELECT DATE\_FORMAT(STR\_TO\_DATE(date, "%Y-%c-%e") - INTERVAL 7 DAY, '%d.%m.%Y')

union SELECT DATE\_FORMAT(STR\_TO\_DATE(date, "%Y-%c-%e") - INTERVAL 14 DAY, '%d.%m.%Y')

union SELECT DATE\_FORMAT(STR\_TO\_DATE(date, "%Y-%c-%e") - INTERVAL 21 DAY, '%d.%m.%Y');

END//

DELIMITER ;

Вызов через:

call test ('2013-03-06')

## DS & DE

1. Чем логистическая регрессия отличается от линейной?

**Ответ:** Линейная регрессия предсказывает значение непрерывной переменной, логистическая – вероятность классификации результата в одну из категорий зависимой переменной.

1. Чем градиентный спуск отличается от стохастического градиентного спуска?

**Ответ:** Не могу ответить - на курсах тема не затрагивалась, для изучения в других источниках нужно больше времени.

1. Какие компоненты входят в тройное экспоненциальное сглаживание временного ряда?

**Ответ:** Уровень, тренд, сезонность.

1. Известные Вам метрики качества для классификации и регрессии?

**Ответ:** Классификация - Accuracy, Precision, Recall, F-score, ROC.  
Регрессия - MSE (среднеквадратичная ошибка), MAE (средняя абсолютная ошибка), MASE (средняя абсолютная масштабированная ошибка), R-квадрат (коэффициент детерминации).

1. Для чего нужна регуляризация?

**Ответ:** чтобы избежать переобучения, когда модель хорошо работает на данных обучения и плохо - на данных тестирования.

1. Загрузить датасет taxi.csv.

**Ответ:** анализ данных см. в приложенном к письму файле или по ссылке <https://drive.google.com/file/d/1BvReJ6-mWXdbfWnZI6ItUS19xkb2wGbB/view?usp=share_link>

* Изучить данные. Отобразить общую информацию по признакам. Описать свои наблюдения.
* Выявить пропуски, а также возможные причины их возникновения. Решить, что следует сделать с ними. Описать свои наблюдения и решение.
* Оценить и описать зависимость переменных между собой.
* Определите стратегию преобразования категориальных признаков для дальнейшего построения моделей.
* Найти признаки, которые можно разделить на другие, или преобразовать в другой тип данных. Удалить лишние, при необходимости.
* Разделить выборку на обучаемую и тестовую.
* Обучить несколько моделей, выбрать критерий оценки выбрать наилучшую.
* Визуализировать результаты.

Указать ссылки на ваш профиль GitHub, Kaggle и т.п.

**Ответ:**